aws

Guia de gerenciamento

Amazon EMR

Amazon EMR: Guia de gerenciamento

Table of Contents

O que é o Amazon EMR?

O Amazon EMR (anteriormente chamado de Amazon Elastic MapReduce) é uma plataforma de cluster gerenciada que simplifica a execução de estruturas de big data, como Apache Hadoop e Apache Spark, para processar e analisar grandes quantidades de dados. AWS Ao usar essas estruturas e projetos de código aberto relacionados, é possível processar dados para finalidades analíticas e workloads de inteligência de negócios. O Amazon EMR também permite transformar e mover grandes quantidades de dados de e para outros armazenamentos de dados e bancos de dados da AWS , como o Amazon Simple Storage Service (Amazon S3) e o Amazon DynamoDB.

Se você for um usuário iniciante do Amazon EMR, recomendamos começar com a leitura do seguinte material, além desta seção:

- Amazon EMR: esta página do serviço fornece destaques, detalhes do produto e informações sobre preços para o Amazon EMR.
- Tutorial: conceitos básicos do Amazon EMR: este tutorial permite que você comece a usar o Amazon EMR rapidamente.

Nesta seção

Visão geral do Amazon EMR

Este tópico fornece uma visão geral dos clusters do Amazon EMR, incluindo como enviar trabalho para um cluster, como esses dados são processados e quais são os diversos estados pelos quais o cluster passa durante o processamento.

Neste tópico

Noções básicas sobre clusters e nós

O componente central do Amazon EMR é o cluster. Um cluster é um conjunto de instâncias do Amazon Elastic Compute Cloud (Amazon EC2). Cada instância do cluster é chamada de nó. Cada nó tem um perfil dentro do cluster, conhecido como tipo de nó. O Amazon EMR também instala diferentes componentes de software em cada tipo de nó, atribuindo a cada nó um perfil em uma aplicação distribuída, como o Apache Hadoop.

Os tipos de nós no Amazon EMR são os seguintes:

- Nó primário: um nó que gerencia o cluster ao executar componentes de software para coordenar a distribuição de dados e de tarefas entre outros nós para processamento. O nó primário rastreia o status das tarefas e monitora a integridade do cluster. Cada cluster tem um nó primário e é possível criar um cluster de nó único apenas com o nó primário.
- Nó core: nó com componentes de software que executam tarefas e armazenam dados no Hadoop Distributed File System (HDFS) do cluster. Clusters de vários nós têm pelo menos um nó core.
- Nó de tarefa: nó com componentes de software que apenas executa tarefas e não armazena dados no HDFS. Nós de tarefa são opcionais.

Envio de trabalhos para um cluster

Ao executar um cluster no Amazon EMR, você tem diversas opções sobre como especificar o trabalho que precisa ser feito.

- Forneça a definição completa do trabalho a ser feito nas funções que você especifica como etapas ao criar um cluster. Isto é normalmente feito para clusters que processam uma quantidade definida de dados e, em seguida, são encerrados quando o processamento é concluído.
- Crie um cluster de longa duração e use o console do Amazon EMR, a API do Amazon EMR ou AWS CLI o para enviar etapas, que podem conter um ou mais trabalhos. Para ter mais informações, consulte Enviar trabalhos a um cluster.
- Crie um cluster, conecte-se ao nó primário e a outros nós conforme necessário usando o SSH e use as interfaces que as aplicações instaladas fornecem para executar tarefas e enviar consultas, com scripts ou de forma interativa. Para obter mais informações, consulte o Guia de versão do Amazon EMR.

Processar dados

Ao executar o cluster, você escolhe as estruturas e os aplicativos a serem instalados para as suas necessidades de processamento de dados. Para processar dados no cluster do Amazon EMR, é possível enviar trabalhos ou consultas diretamente para aplicações instaladas ou executar etapas no cluster.

Envio de trabalhos diretamente para aplicações

Você pode enviar trabalhos e interagir diretamente com os softwares instalados no cluster do Amazon EMR. Para fazer isso, normalmente você se conecta ao nó primário usando uma conexão segura e acessa as interfaces e ferramentas que estão disponíveis para os softwares que são executados diretamente em seu cluster. Para ter mais informações, consulte Conectar-se a um cluster.

Execução de etapas para processar dados

Você pode enviar uma ou mais etapas ordenadas para um cluster do Amazon EMR. Cada etapa é uma unidade de trabalho que contém instruções para manipular dados para processamento pelos softwares instalados no cluster.

Veja a seguir um exemplo de processo utilizando quatro etapas:

1. Enviar um conjunto de dados de entrada para processamento.
2. Processar a saída da primeira etapa usando um programa Pig.
3. Processar um segundo conjunto de dados de entrada usando um programa Hive.
4. Gravar um conjunto de dados de saída.

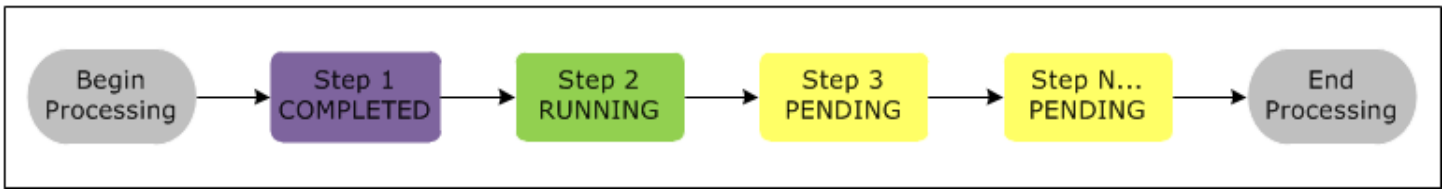
Geralmente, quando você processa dados no Amazon EMR, a entrada corresponde a dados armazenados como arquivos no sistema de arquivos subjacente escolhido, como o Amazon S3 ou o HDFS. Esses dados passam de uma etapa para a próxima na sequência de processamento. A etapa final grava os dados de saída em um local especificado, como um bucket do Amazon S3.

As etapas são executadas na seguinte sequência:

1. É enviada uma solicitação para iniciar as etapas de processamento.
2. O estado de todas as etapas é definido como PENDING (Pendente).

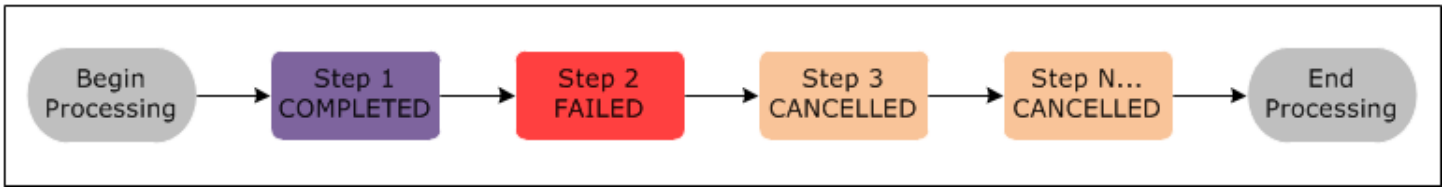
3. Quando a primeira etapa da sequência é iniciada, seu estado muda para RUNNING (Em execução). As outras etapas permanecem no estado PENDING (Pendente).
4. Após a conclusão da primeira etapa, seu estado muda para COMPLETED (Concluído).
5. A próxima etapa da sequência é iniciada, e seu estado muda para RUNNING (Em execução). Após a conclusão, seu estado muda para COMPLETED (Concluído).
6. Esse padrão repete-se para cada etapa, até todas elas estejam concluídas e o processamento seja encerrado.

O diagrama a seguir representa a sequência de etapas e mudança de estado para as etapas conforme elas são processadas.



Se uma etapa falhar durante o processamento, seu estado será alterado para FAILED. Você pode determinar o que acontece a seguir em cada etapa. Por padrão, todas as etapas restantes na sequência são definidas como CANCELLED e não são executadas se uma etapa anterior falhar. Você também pode optar por ignorar a falha e permitir que as etapas restantes continuem ou encerrem o cluster imediatamente.

O diagrama a seguir representa a sequência de etapas e a mudança de estado padrão quando uma etapa falha durante o processamento.



Noções básicas sobre o ciclo de vida do cluster

Um cluster do Amazon EMR com êxito segue este processo:

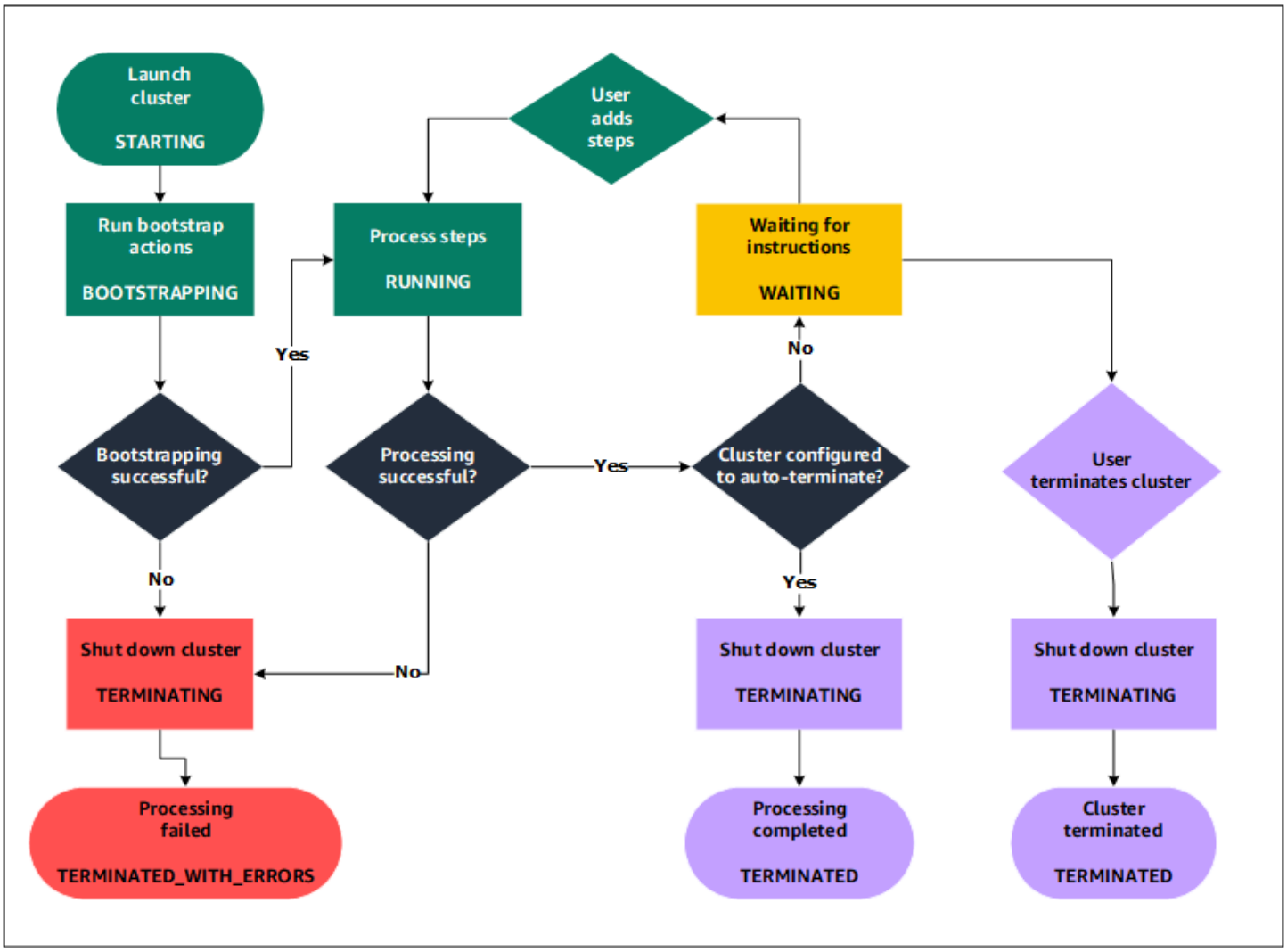
1. Primeiro, o Amazon EMR provisiona instâncias do EC2 no cluster para cada instância, de acordo com as suas especificações. Para ter mais informações, consulte Configurar o hardware e as redes do cluster. Para todas as instâncias, o Amazon EMR usa a AMI padrão para o Amazon EMR ou uma AMI personalizada do Amazon Linux especificada por você. Para ter mais

informações, consulte Usar uma AMI personalizada. Durante essa fase, o estado do cluster é `STARTING`.

2. O Amazon EMR executa ações de bootstrap especificadas em cada instância. Você pode usar as ações de bootstrap para instalar aplicativos personalizados e executar as personalizações necessárias. Para ter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais. Durante essa fase, o estado do cluster é `BOOTSTRAPPING`.
3. O Amazon EMR instala as aplicações nativas especificadas ao criar o cluster, como o Hive, o Hadoop, o Spark e outros.
4. Depois que as ações de bootstrap forem concluídas com êxito e que os aplicativos nativos forem instalados, o estado do cluster será `RUNNING`. Nesse momento, você pode se conectar às instâncias de cluster, e o cluster executará de maneira sequencial todas as etapas que você especificou ao criar o cluster. Você pode enviar etapas adicionais, que serão executadas depois que as etapas anteriores forem concluídas. Para ter mais informações, consulte Enviar trabalhos a um cluster.
5. Depois que as etapas forem executadas com êxito, o cluster entrará no estado `WAITING`. Se um cluster estiver configurado para o encerramento automático após a conclusão da última etapa, ele entrará em um estado `TERMINATING` e, em seguida, no estado `TERMINATED`. Se o cluster estiver configurado para aguardar, você deverá encerrá-lo manualmente quando não precisar mais dele. Após encerrar manualmente o cluster, ele entrará no estado `TERMINATING` e, em seguida, no estado `TERMINATED`.

Uma falha durante o ciclo de vida do cluster faz com que o Amazon EMR encerre o cluster e todas as suas instâncias, a menos que você habilite a proteção contra encerramento. Se um cluster for encerrado devido a uma falha, todos os dados armazenados no cluster serão excluídos, e o estado do cluster será definido como `TERMINATED_WITH_ERRORS`. Se você tiver habilitado a proteção contra encerramento, você poderá recuperar os dados do seu cluster e, em seguida, remover a proteção contra encerramento e encerrá-lo. Para ter mais informações, consulte Usar a proteção contra término.

O diagrama a seguir representa o ciclo de vida de um cluster e como cada estágio desse ciclo é mapeado para um estado de cluster específico.



Benefícios do uso do Amazon EMR

Há muitos benefícios em usar o Amazon EMR. Esta seção fornece uma visão geral desses benefícios, além de links para informações adicionais para ajudá-lo a explorar ainda mais.

Tópicos

- Redução de custos
- AWS integração
- Implantação
- Escalabilidade e flexibilidade
- Confiabilidade
- Segurança

- Monitoramento
- Interfaces de gerenciamento

Redução de custos

Os preços do Amazon EMR dependem do tipo de instância e do número de instâncias do Amazon EC2 implantadas, bem como da região em que o seu cluster é iniciado. A definição de preço sob demanda oferece tarifas baixas, mas você pode reduzir os custos ainda mais comprando instâncias reservadas ou instâncias spot. As instâncias spot podem oferecer economias significativas. Em alguns casos, até um décimo dos preços sob demanda.

> **Note**
>
> Se você usar o Amazon S3, o Amazon Kinesis ou o DynamoDB com o cluster do EMR, haverá cobranças adicionais para os serviços faturados separadamente do uso do Amazon EMR.

> **Note**
>
> Ao configurar um cluster do Amazon EMR em uma sub-rede privada, recomendamos configurar também endpoints da VPC para o Amazon S3. Se o cluster do EMR estiver em uma sub-rede privada sem endpoints da VPC para o Amazon S3, você incorrerá em cobranças adicionais de gateway NAT associadas ao tráfego do S3, pois o tráfego entre o cluster do EMR e o S3 não permanecerá na VPC.

Para obter mais informações sobre as opções e os detalhes dos preços, consulte Preço do Amazon EMR.

AWS integração

O Amazon EMR se integra a outros AWS serviços para fornecer recursos e funcionalidades relacionados à rede, armazenamento, segurança, etc., para seu cluster. A lista a seguir fornece vários exemplos dessa integração:

- Amazon EC2 para as instâncias que compõem os nós do cluster.

- Amazon Virtual Private Cloud (Amazon VPC) para configurar a rede virtual na qual você inicia as instâncias.
- Amazon S3 para armazenar dados de entrada e de saída.
- Amazon CloudWatch monitorará o desempenho do cluster e configurará alarmes
- AWS Identity and Access Management (IAM) para configurar permissões
- AWS CloudTrail para auditar solicitações feitas ao serviço
- AWS Data Pipeline para programar e iniciar seus clusters
- AWS Lake Formation para descobrir, catalogar e proteger dados em um data lake do Amazon S3

Implantação

O cluster do EMR consiste de instâncias do EC2, que realizam o trabalho que você envia ao seu cluster. Ao executar o seu cluster, o Amazon EMR configura as instâncias com as aplicações que você escolher, como Apache Hadoop ou Spark. Escolha o tamanho de instância e o tipo que melhor se adequa às necessidades de processamento do seu cluster: processamento em lotes, consultas de baixa latência, dados de streaming ou armazenamento físico de dados grandes. Para obter mais informações sobre os tipos de instâncias disponíveis para o Amazon EMR, consulte Configurar o hardware e as redes do cluster.

O Amazon EMR oferece diversas maneiras de configurar softwares em seu cluster. Por exemplo, você pode instalar uma versão do Amazon EMR com um conjunto de aplicações escolhidas que pode incluir estruturas versáteis, como o Hadoop, e aplicações, como o Hive, o Pig ou o Spark. Também é possível instalar uma das diversas distribuições do MapR. O Amazon EMR usa o Amazon Linux, portanto, você também pode instalar softwares no cluster de forma manual ao usar o gerenciador de pacotes YUM ou a partir da origem. Para ter mais informações, consulte Configuração de software do cluster.

Escalabilidade e flexibilidade

O Amazon EMR oferece flexibilidade para aumentar ou reduzir a escala verticalmente do seu cluster conforme as necessidades de computação são alteradas. Você pode redimensionar seu cluster para adicionar instâncias para cargas de trabalho de pico e remover instâncias para controlar custos quando as cargas de pico diminuírem. Para ter mais informações, consulte Redimensionar manualmente um cluster em execução.

O Amazon EMR também oferece a opção de executar vários grupos de instâncias para que você possa usar instâncias sob demanda em um grupo para garantir a capacidade de processamento

em conjunto com instâncias spot em outro grupo para concluir os trabalhos com mais rapidez e custos mais baixos. Você também pode combinar diferentes tipos de instâncias para tirar proveito dos melhores preços por um tipo de instância spot sobre o outro. Para ter mais informações, consulte Quando você deve usar instâncias spot?.

Além disso, o Amazon EMR oferece flexibilidade para usar vários sistemas de arquivos para dados de entrada, de saída e intermediários. Por exemplo, você pode escolher o Sistema de Arquivos Distribuído do Hadoop (HDFS), que é executado nos nós primários e centrais do cluster para o processamento de dados que não precisam ser armazenados além do ciclo de vida do cluster. Você pode escolher o Sistema de Arquivos do EMR (EMRFS) para usar o Amazon S3 como uma camada de dados para aplicações em execução no cluster, com a finalidade de separar a computação e o armazenamento, e manter os dados persistentes de forma externa ao ciclo de vida do cluster. O EMRFS fornece o benefício adicional de permitir que você aumente ou diminua a escalabilidade independentemente, de acordo com as suas necessidades de computação e armazenamento. Você pode escalar suas necessidades de computação ao redimensionar o cluster e escalar as necessidades de armazenamento ao usar o Amazon S3. Para ter mais informações, consulte Trabalhar com armazenamento e sistemas de arquivos.

Confiabilidade

O Amazon EMR monitora nós no cluster e encerra e substitui automaticamente uma instância em caso de falha.

O Amazon EMR oferece opções de configuração que controlam se o cluster será encerrado automática ou manualmente. Se você configurar o cluster para ser automaticamente encerrado, isso acontecerá após a conclusão de todas as etapas. Ele é conhecido como cluster transitório. No entanto, você pode configurar o cluster para continuar a ser executado após o processamento, para poder optar por terminá-lo manualmente quando não precisar mais dele. Outra opção é criar um cluster, interagir diretamente com os aplicativos instalados e então terminá-lo manualmente quando você não precisar mais dele. Os clusters nestes exemplos são chamados de clusters de longa execução.

Além disso, você pode configurar a proteção contra encerramento para impedir que instâncias do seu cluster sejam terminadas devido a erros ou problemas durante o processamento. Quando a proteção contra encerramento está habilitada, você pode recuperar dados de instâncias antes do encerramento. As configurações padrão para essas opções são diferentes dependendo de você executar o cluster usando o console, a CLI ou a API. Para ter mais informações, consulte Usar a proteção contra término.

Segurança

O Amazon EMR utiliza outros AWS serviços, como IAM e Amazon VPC, e recursos como pares de chaves do Amazon EC2, para ajudar você a proteger seus clusters e dados.

IAM

O Amazon EMR se integra ao IAM para gerenciar permissões. Você define permissões usando políticas do IAM, que você anexa a usuários ou grupos do IAM. As permissões que você definir na política determinam as ações que esses usuários ou membros do grupo podem realizar, bem como os recursos que eles podem acessar. Para ter mais informações, consulte Como o Amazon EMR funciona com o IAM.

Além disso, o Amazon EMR usa perfis do IAM para o próprio serviço do Amazon EMR e o perfil de instância do EC2 para as instâncias. Essas funções concedem permissões para que o serviço e as instâncias acessem outros AWS serviços em seu nome. Há um perfil padrão para o serviço do Amazon EMR e um perfil padrão para o perfil de instância do EC2. As funções padrão usam políticas AWS gerenciadas, que são criadas automaticamente para você na primeira vez que você inicia um cluster do EMR a partir do console e escolhe as permissões padrão. Você também pode criar os perfis do IAM padrão usando a AWS CLI. Se quiser gerenciar as permissões em vez de AWS, você pode escolher funções personalizadas para o perfil do serviço e da instância. Para ter mais informações, consulte Configurar perfis de serviço do IAM para permissões do Amazon EMR aos serviços e recursos da AWS.

Grupos de segurança

O Amazon EMR usa grupos de segurança para controlar o tráfego de entrada e de saída para as instâncias do EC2. Ao iniciar seu cluster, o Amazon EMR usa um grupo de segurança para a instância primária e um grupo de segurança para ser compartilhado pelas instâncias centrais e de tarefas. O Amazon EMR configura as regras do grupo de segurança para garantir a comunicação entre as instâncias do cluster. Como opção, é possível configurar grupos de segurança adicionais e atribuí-los às instâncias primárias, centrais e de tarefas para obter regras mais avançadas. Para ter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

Criptografia

O Amazon EMR oferece suporte à opção de criptografia do lado do cliente e do servidor do Amazon S3 com EMRFS para ajudar a proteger os dados armazenados no Amazon S3. Com a criptografia do lado do servidor, o Amazon S3 criptografa seus dados após o upload.

Com a criptografia no lado do cliente, o processo de criptografia e descriptografia ocorre no cliente EMRFS, no seu cluster do EMR. Você gerencia a chave raiz para criptografia do lado do cliente usando o AWS Key Management Service (AWS KMS) ou seu próprio sistema de gerenciamento de chaves.

Para obter mais informações, consulte Specifying Amazon S3 encryption using EMRFS properties.

Amazon VPC

O Amazon EMR oferece suporte à execução de clusters em uma nuvem privada virtual (VPC) na Amazon VPC. Uma VPC é uma rede virtual isolada AWS que fornece a capacidade de controlar aspectos avançados da configuração e do acesso à rede. Para ter mais informações, consulte Configurar redes.

AWS CloudTrail

O Amazon EMR se integra CloudTrail para registrar informações sobre solicitações feitas por ou em nome de sua conta. AWS Com essas informações, você pode manter o controle de quem está acessando seu cluster, quando isso é feito e o endereço IP do qual a solicitação foi feita. Para ter mais informações, consulte Registro de chamadas de API do Amazon EMR em AWS CloudTrail.

Pares de chaves do Amazon EC2

Você pode monitorar e interagir com o seu cluster ao criar uma conexão segura entre o computador remoto e o nó primário. Você usa o protocolo de rede Secure Shell (SSH) nesta conexão ou usar o Kerberos para autenticação. Se você usar o SSH, um par de chaves do Amazon EC2 será necessário. Para ter mais informações, consulte Usar um par de chaves do EC2 para credenciais SSH.

Monitoramento

Você pode usar as interfaces de gerenciamento e os arquivos de log do Amazon EMR para solucionar problemas de cluster, como falhas ou erros. O Amazon EMR oferece a capacidade de arquivar arquivos de log no Amazon S3 para que você possa armazenar logs e solucionar problemas mesmo após o encerramento do cluster. O Amazon EMR também fornece uma ferramenta de depuração opcional no console do Amazon EMR para navegar nos arquivos de log com base em etapas, trabalhos e tarefas. Para ter mais informações, consulte Configurar registro em log e depuração do cluster.

O Amazon EMR se integra CloudWatch para monitorar métricas de desempenho do cluster e dos trabalhos dentro do cluster. Você pode configurar alarmes com base em várias métricas, por exemplo, se o cluster está ocioso ou a porcentagem de armazenamento usado. Para ter mais informações, consulte Monitorando métricas do Amazon EMR com CloudWatch.

Interfaces de gerenciamento

Existem diversas maneiras de interagir com o Amazon EMR:

- Console: uma interface gráfica do usuário que você pode usar para iniciar e gerenciar clusters. Com ela, você preenche formulários da Web para especificar os detalhes dos clusters a serem executados, visualizar os detalhes de clusters existentes, depurar e encerrar clusters. Usar o console é a maneira mais fácil de começar a usar o Amazon EMR e nenhum conhecimento de programação é necessário. O console está disponível on-line em https://console.aws.amazon.com/elasticmapreduce/home.
- AWS Command Line Interface (AWS CLI) — Um aplicativo cliente que você executa em sua máquina local para se conectar ao Amazon EMR e criar e gerenciar clusters. O AWS CLI contém um conjunto rico em recursos de comandos específicos para o Amazon EMR. Com isso, você pode escrever scripts que automatizam o processo de execução e gerenciamento de clusters. Se você preferir trabalhar em uma linha de comando, usar o AWS CLI é a melhor opção. Para obter mais informações, consulte Amazon EMR em AWS CLI Command Reference.
- Kit de desenvolvimento de software (SDK): os SDKs fornecem funções que chamam o Amazon EMR para criar e gerenciar clusters. Com eles, você pode escrever aplicativos que automatizam o processo de criação e gerenciamento de clusters. Usar o SDK é a melhor opção para ampliar ou personalizar a funcionalidade do Amazon EMR. No momento, o Amazon EMR está disponível nos seguintes SDKs: Go, Java, .NET (C# e VB.NET), Node.js, PHP, Python e Ruby. Para obter mais informações sobre esses SDKs, consulte Ferramentas para criar com a AWS e códigos de exemplo e bibliotecas do Amazon EMR.
- API do serviço Web: uma interface de baixo nível que você pode usar para chamar o serviço Web diretamente, usando JSON. Usar a API é a melhor opção para criar um SDK personalizado que chame o Amazon EMR. Para obter mais informações, consulte a Referência da API do Amazon EMR.

Visão geral da arquitetura do Amazon EMR

A arquitetura de serviços do Amazon EMR consiste em várias camadas, cada uma delas fornecendo determinados recursos e funcionalidades ao cluster. Esta seção fornece uma visão geral das camadas e dos componentes de cada uma.

Neste tópico

Armazenamento

A camada de armazenamento inclui os diferentes sistemas de arquivos que são usados com o cluster. Existem vários tipos diferentes de opções de armazenamento, da seguinte maneira.

Hadoop Distributed File System (HDFS)

O Hadoop Distributed File System (HDFS) é um sistema de arquivos distribuído e escalável para o Hadoop. O HDFS distribui os dados armazenados entre as instâncias do cluster, armazenando várias cópias dos dados em instâncias diferentes para garantir que nenhum dos dados se perca caso uma das instâncias falhe. O HDFS é um armazenamento temporário que é reivindicado quando um cluster é encerrado. O HDFS é útil para armazenar em cache resultados intermediários durante o MapReduce processamento ou para cargas de trabalho com E/S aleatória significativa.

Para obter mais informações, consulte Armazenamento de instâncias neste guia ou acesse o HDFS User Guide no site do Apache Hadoop.

Sistema de arquivos do EMR (EMRFS)

Ao usar o Sistema de Arquivos do EMR (EMRFS), o Amazon EMR amplia o Hadoop para adicionar a capacidade de acessar diretamente os dados armazenados no Amazon S3, como se ele fosse um sistema de arquivos como o HDFS. Você pode usar o HDFS ou o Amazon S3 como o sistema de arquivos em seu cluster. Na maioria das vezes, o Amazon S3 é usado para armazenar dados de entrada e de saída, e os resultados intermediários são armazenados no HDFS.

Sistema de arquivos local

O sistema de arquivos local é a um disco conectado localmente. Quando você cria um cluster do Hadoop, cada nó é criado a partir de uma instância do Amazon EC2 que tem componentes configurados previamente para o armazenamento em disco anexado previamente, que é chamado de armazenamento de instância. Os dados nos volumes de armazenamento de instância persistem somente durante o ciclo de vida da instância do Amazon EC2.

Gerenciamento de recursos de cluster

A camada de gerenciamento de recursos é responsável por gerenciar os recursos de cluster e agendar os trabalhos para o processamento de dados.

Por padrão, o Amazon EMR usa o YARN (Yet Another Resource Negotiator), que é um componente introduzido no Apache Hadoop 2.0 para gerenciar centralmente os recursos de cluster para várias estruturas de processamento de dados. No entanto, existem outras estruturas e aplicações disponibilizadas no Amazon EMR que não usam o YARN como gerenciador de recursos. O Amazon EMR também tem um atendente em cada nó que administra os componentes do YARN, mantém o cluster íntegro e se comunica com o Amazon EMR.

Como as instâncias spot são frequentemente usadas para executar nós de tarefas, o Amazon EMR tem a funcionalidade padrão para programar trabalhos do YARN para que os trabalhos em execução não falhem quando os nós de tarefas em execução nas instâncias spot forem encerrados. O Amazon EMR faz isso ao permitir que processos principais de aplicações sejam executados somente em nós centrais. O processo principal da aplicação controla os trabalhos em execução e precisa permanecer ativo durante a vida útil do trabalho.

A versão 5.19.0 e as versões posteriores do Amazon EMR usam o recurso de rótulos de nós do YARN integrado para conseguir isso. (As versões anteriores usavam um patch de código). As propriedades nas classificações de configuração `yarn-site` e `capacity-scheduler` são configuradas por padrão para que o programador de capacidade e o programador justo do YARN aproveitem os rótulos de nós. O Amazon EMR rotula automaticamente os nós centrais com o rótulo `CORE` e define propriedades para que as aplicações principais sejam programadas somente em nós com o rótulo CORE. Modificar manualmente as propriedades relacionadas nas classificações de configuração yarn-site e docapacity-scheduler, ou diretamente nos arquivos XML associados, pode interromper esse recurso ou modificar essa funcionalidade.

Estruturas de processamento de dados

A camada de estruturas de processamento de dados é o mecanismo usado para processar e analisar dados. Existem muitas estruturas disponíveis que são executadas no YARN ou têm seu próprio gerenciamento de recursos. Estruturas diferentes estão disponíveis para tipos distintos de necessidades de processamento, como e lote, interativo, na memória, streaming e assim por diante. A estrutura que você escolhe depende do seu caso de uso. A escolha afeta as linguagens e interfaces disponíveis na camada de aplicativo, que é a camada usada para interagir com os dados que você deseja processar. As principais estruturas de processamento disponíveis para o Amazon EMR são o MapReduce Hadoop e o Spark.

Hadoop MapReduce

O Hadoop MapReduce é um modelo de programação de código aberto para computação distribuída. Ele simplifica o processo de gravação de aplicativos distribuídos em paralelo, manipulando toda a lógica, enquanto você fornece as funções Map e Reduce. A função Map mapeia dados para conjuntos de pares de chave/valor chamados de resultados intermediários. A função Reduce combina os resultados intermediários, aplica algoritmos adicionais e produz o resultado final. Existem várias estruturas disponíveis para MapReduce, como o Hive, que gera automaticamente os programas Map e Reduce.

Para obter mais informações, acesse How map and reduce operations are actually carried out no site Wiki do Apache Hadoop.

Apache Spark

O Spark é um modelo de programação e estrutura de cluster para o processamento de cargas de trabalho de Big Data. Como o Hadoop MapReduce, o Spark é um sistema de processamento distribuído de código aberto, mas usa gráficos acíclicos direcionados para planos de execução e armazenamento em cache na memória para conjuntos de dados. Ao executar o Spark no Amazon EMR, você pode usar o EMRFS para acessar diretamente os dados no Amazon S3. O Spark oferece suporte a vários módulos de consulta interativos, como o SparkSQL.

Para obter mais informações, consulte Apache Spark on Amazon EMR clusters no Guia de lançamento do Amazon EMR.

Aplicações e programas

O Amazon EMR oferece suporte para muitas aplicações, como o Hive, o Pig e a biblioteca Spark Streaming, para fornecer funcionalidades como o uso de linguagens de nível superior para criar

workloads de processamento, o uso de algoritmos de machine learning, a criação de aplicações de processamento de fluxos e o desenvolvimento de data warehouses. Além disso, o Amazon EMR também oferece suporte a projetos de código aberto que têm suas próprias funcionalidades de gerenciamento de cluster em vez de usarem o YARN.

Você usa várias bibliotecas e linguagens para interagir com as aplicações executadas no Amazon EMR. Por exemplo, você pode usar Java, Hive ou Pig com MapReduce Spark Streaming, Spark SQL, MLlib e GraphX com o Spark.

Para obter mais informações, consulte o Guia de versão do Amazon EMR.

Configuração do Amazon EMR

Conclua as tarefas desta seção antes de iniciar um cluster do Amazon EMR pela primeira vez:

Antes de usar o Amazon EMR pela primeira vez, conclua as seguintes tarefas:

Inscreva-se para um Conta da AWS

Se você não tiver um Conta da AWS, conclua as etapas a seguir para criar um.

Para se inscrever em um Conta da AWS

1. Abra https://portal.aws.amazon.com/billing/signup.
2. Siga as instruções on-line.

   Parte do procedimento de inscrição envolve receber uma chamada telefônica e digitar um código de verificação no teclado do telefone.

   Quando você se inscreve em um Conta da AWS, um Usuário raiz da conta da AWSé criado. O usuário raiz tem acesso a todos os Serviços da AWS e recursos na conta. Como prática recomendada de segurança, atribua acesso administrativo a um usuário e use somente o usuário raiz para realizar tarefas que exijam acesso do usuário raiz.

AWS envia um e-mail de confirmação após a conclusão do processo de inscrição. A qualquer momento, é possível visualizar as atividades da conta atual e gerenciar sua conta acessando https://aws.amazon.com/ e selecionando Minha conta.

Crie um usuário com acesso administrativo

Depois de se inscrever em um Conta da AWS, proteja seu Usuário raiz da conta da AWS AWS IAM Identity Center, habilite e crie um usuário administrativo para que você não use o usuário root nas tarefas diárias.

Proteja seu Usuário raiz da conta da AWS

1. Faça login AWS Management Consolecomo proprietário da conta escolhendo Usuário raiz e inserindo seu endereço de Conta da AWS e-mail. Na próxima página, digite sua senha.

Para obter ajuda ao fazer login usando o usuário raiz, consulte Fazer login como usuário raiz no Guia do usuário do Início de Sessão da AWS .

2. Ative a autenticação multifator (MFA) para o usuário raiz.c

   Para obter instruções, consulte Habilitar um dispositivo de MFA virtual para seu usuário Conta da AWS raiz (console) no Guia do usuário do IAM.

Crie um usuário com acesso administrativo

1. Habilitar o IAM Identity Center.

   Para obter instruções, consulte Habilitar AWS IAM Identity Center no Guia do usuário do AWS IAM Identity Center .

2. No IAM Identity Center, conceda acesso administrativo a um usuário.

   Para ver um tutorial sobre como usar o Diretório do Centro de Identidade do IAM como fonte de identidade, consulte Configurar o acesso do usuário com o padrão Diretório do Centro de Identidade do IAM no Guia AWS IAM Identity Center do usuário.

Faça login como usuário com acesso administrativo

- Para fazer login com seu usuário do Centro de Identidade do IAM, use a URL de login que foi enviada ao seu endereço de e-mail quando você criou o usuário do Centro do Usuário do IAM.

  Para obter ajuda para fazer login usando um usuário do IAM Identity Center, consulte Como fazer login no portal de AWS acesso no Guia Início de Sessão da AWS do usuário.

Atribua acesso a usuários adicionais

1. No IAM Identity Center, crie um conjunto de permissões que siga as melhores práticas de aplicação de permissões com privilégios mínimos.

   Para obter instruções, consulte Criar um conjunto de permissões no Guia AWS IAM Identity Center do usuário.

2. Atribua usuários a um grupo e, em seguida, atribua acesso de login único ao grupo.

   Para obter instruções, consulte Adicionar grupos no Guia AWS IAM Identity Center do usuário.

Crie um par de chaves do Amazon EC2 para SSH

> **Note**
>
> Com as versões 5.10.0 ou posteriores do Amazon EMR, é possível configurar o Kerberos para autenticar usuários e conexões SSH com um cluster. Para ter mais informações, consulte Usar o Kerberos para autenticação com o Amazon EMR.

Para realizar a autenticação e se conectar aos nós em um cluster por meio de um canal seguro usando o protocolo Secure Shell (SSH), crie um par de chaves do Amazon Elastic Compute Cloud (Amazon EC2) antes de iniciar o cluster. Também é possível criar um cluster sem par de chaves. Isso geralmente é feito com clusters transitórios que são iniciados, executam etapas e são encerrados automaticamente.

Se...	Então...
Você já tem um par de chaves do Amazon EC2 que deseja usar ou não tem a necessidade de se autenticar no cluster.	Pule esta etapa.
Você precisa criar um par de chaves.	Consulte Creating your key pair using Amazon EC2.

Próximas etapas

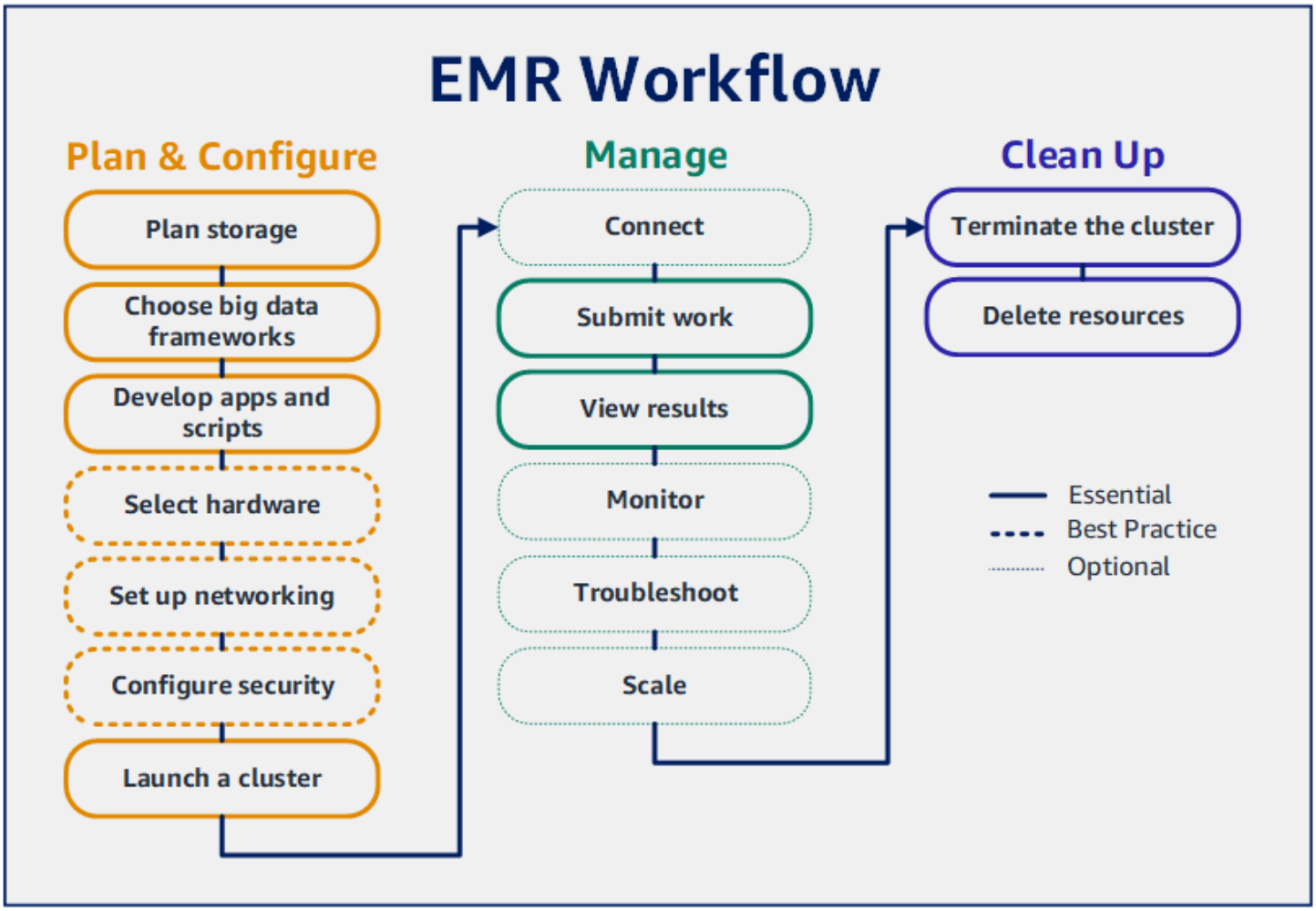
- Para obter orientação sobre como criar um cluster de exemplo, consulte Tutorial: conceitos básicos do Amazon EMR.
- Para obter mais informações sobre como configurar um cluster personalizado e controlar o acesso a ele, consulte Planejar e configurar clusters e Segurança no Amazon EMR.

Tutorial: conceitos básicos do Amazon EMR

Visão geral

Com o Amazon EMR, é possível configurar um cluster para processar e analisar dados com estruturas de big data em apenas alguns minutos. Este tutorial mostra como iniciar um cluster de amostra usando o Spark e como executar um PySpark script simples armazenado em um bucket do Amazon S3. Ele abrange tarefas essenciais do Amazon EMR em três categorias principais de fluxo de trabalho: planejar e configurar, gerenciar e limpar.

Você encontrará links para tópicos mais detalhados à medida que avança no tutorial e obterá ideias para etapas adicionais na seção Próximas etapas. Se você tiver perguntas ou dúvidas, entre em contato com a equipe do Amazon EMR em nosso fórum de discussão.



Pré-requisitos

- Antes de iniciar um cluster do Amazon EMR, certifique-se de concluir as tarefas em Configuração do Amazon EMR.

Custo

- O exemplo de cluster que você criar será executado em um ambiente dinâmico. O cluster acumula cobranças mínimas. Para evitar cobranças adicionais, certifique-se de concluir as tarefas de limpeza na última etapa deste tutorial. As cobranças são acumuladas à taxa por segundo, de acordo com os preços do Amazon EMR. As cobranças também variam com base na região. Para obter mais informações, consulte Preço do Amazon EMR.
- Cobranças mínimas podem ser acumuladas para arquivos pequenos armazenados no Amazon S3. Algumas ou todas as cobranças do Amazon S3 podem ser dispensadas se você estiver dentro dos limites de uso do AWS nível gratuito. Para obter mais informações, consulte Preço do Amazon S3 e nível gratuito da AWS.

Etapa 1: planejar e configurar um cluster do Amazon EMR

Preparação do armazenamento para o Amazon EMR

Ao usar o Amazon EMR, é possível escolher entre uma variedade de sistemas de arquivos para armazenar dados de entrada, dados de saída e arquivos de log. Neste tutorial, você usa o EMRFS para armazenar dados em um bucket do S3. O EMRFS é uma implementação do sistema de arquivos do Hadoop que permite a leitura e a gravação de arquivos regulares no Amazon S3. Para ter mais informações, consulte Trabalhar com armazenamento e sistemas de arquivos.

Para criar um bucket para este tutorial, siga as instruções em How do I create an S3 bucket? no Guia do usuário do console do Amazon Simple Storage Service. Crie o bucket na mesma AWS região em que você planeja lançar seu cluster do Amazon EMR. Por exemplo, Oeste dos EUA (Oregon) us-west-2.

Os buckets e as pastas usados com o Amazon EMR têm as seguintes limitações:

- Os nomes podem consistir em letras minúsculas, números, pontos (.) e hifens (-).
- Os nomes não podem terminar em números.
- O nome do bucket deve ser exclusivo em todas as contas da AWS .

- Uma pasta de saída deve estar vazia.

Preparação de uma aplicação com dados de entrada para o Amazon EMR

A maneira mais comum de preparar uma aplicação para o Amazon EMR é fazer o upload da aplicação e de seus dados de entrada para o Amazon S3. Em seguida, ao enviar o trabalho para o cluster, você especifica os locais do Amazon S3 para o script e para os dados.

Nesta etapa, você carrega um PySpark script de amostra no seu bucket do Amazon S3. Fornecemos um PySpark script para você usar. O script processa os dados de inspeção de estabelecimentos alimentícios e retorna um arquivo de resultados em seu bucket do S3. O arquivo de resultados lista os dez principais estabelecimentos com mais violações do tipo “vermelho”.

Você também carrega dados de entrada de amostra no Amazon S3 para que o PySpark script seja processado. Os dados de entrada correspondem a uma versão modificada dos resultados de inspeções do Departamento de Saúde no Condado de King, em Washington, de 2006 a 2020. Para obter mais informações, consulte King County Open Data: Food Establishment Inspection Data. Confira a seguir exemplos de linhas do conjunto de dados.

```
name, inspection_result, inspection_closed_business, violation_type, violation_points
100 LB CLAM, Unsatisfactory, FALSE, BLUE, 5
100 PERCENT NUTRICION, Unsatisfactory, FALSE, BLUE, 5
7-ELEVEN #2361-39423A, Complete, FALSE, , 0
```

Para preparar o PySpark script de exemplo para o EMR

1. Copie o código de exemplo abaixo em um novo arquivo no editor de sua preferência.

```
import argparse

from pyspark.sql import SparkSession

def calculate_red_violations(data_source, output_uri):
    """
    Processes sample food establishment inspection data and queries the data to
 find the top 10 establishments
    with the most Red violations from 2006 to 2020.

    :param data_source: The URI of your food establishment data CSV, such as 's3://
DOC-EXAMPLE-BUCKET/food-establishment-data.csv'.
```

```
    :param output_uri: The URI where output is written, such as 's3://DOC-EXAMPLE-
BUCKET/restaurant_violation_results'.
    """
    with SparkSession.builder.appName("Calculate Red Health
 Violations").getOrCreate() as spark:
        # Load the restaurant violation CSV data
        if data_source is not None:
            restaurants_df = spark.read.option("header", "true").csv(data_source)

        # Create an in-memory DataFrame to query
        restaurants_df.createOrReplaceTempView("restaurant_violations")

        # Create a DataFrame of the top 10 restaurants with the most Red violations
        top_red_violation_restaurants = spark.sql("""SELECT name, count(*) AS
 total_red_violations
          FROM restaurant_violations
          WHERE violation_type = 'RED'
          GROUP BY name
          ORDER BY total_red_violations DESC LIMIT 10""")

        # Write the results to the specified output URI
        top_red_violation_restaurants.write.option("header",
 "true").mode("overwrite").csv(output_uri)

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument(
        '--data_source', help="The URI for you CSV restaurant data, like an S3
 bucket location.")
    parser.add_argument(
        '--output_uri', help="The URI where output is saved, like an S3 bucket
 location.")
    args = parser.parse_args()

    calculate_red_violations(args.data_source, args.output_uri)
```

2. Salve o arquivo como `health_violations.py`.
3. Faça o upload de `health_violations.py` para o Amazon S3 no bucket criado para este tutorial. Para obter instruções, consulte Fazer upload de um objeto para o bucket no Guia de conceitos básicos do Amazon Simple Storage Service.

Preparar os dados de entrada de exemplo para o EMR

1. Faça o download do arquivo zip food_establishment_data.zip.
2. Descompacte e salve `food_establishment_data.zip` como `food_establishment_data.csv` em sua máquina.
3. Faça o upload do arquivo CSV para o bucket do S3 criado para este tutorial. Para obter instruções, consulte Fazer upload de um objeto para o bucket no Guia de conceitos básicos do Amazon Simple Storage Service.

Para obter mais informações sobre como configurar dados para o EMR, consulte Preparar dados de entrada.

Inicialização de um cluster do Amazon EMR

Após preparar um local de armazenamento e a aplicação, você poderá iniciar um cluster do Amazon EMR de exemplo. Nesta etapa, você inicia um cluster do Apache Spark usando a versão mais recente do Amazon EMR.

Console

Para iniciar um cluster com o Spark instalado com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Na página Criar cluster, observe os valores padrão para Versão, Tipo de instância, Número de instâncias e Permissões. Esses campos são preenchidos automaticamente com valores que funcionam para clusters de uso geral.
4. No campo Nome do cluster, insira um nome exclusivo para o cluster para ajudar você a identificá-lo, como *`Meu primeiro cluster`*. O nome do cluster não pode conter os caracteres <, >, $, | ou `(crase).
5. Em Aplicações, escolha a opção Spark para instalar o Spark em seu cluster.

> **Note**
>
> Escolha as aplicações que você deseja em seu cluster do Amazon EMR antes de iniciar o cluster. Não é possível adicionar ou remover aplicações de um cluster após a inicialização.

6. Em Logs do cluster, marque a caixa de seleção Publicar logs específicos do cluster no Amazon S3. Substitua o valor do local do Amazon S3 pelo bucket do Amazon S3 criado, seguido por **`/logs`**. Por exemplo, **`s3://DOC-EXAMPLE-BUCKET/logs`**. A adição de **`/logs`** cria uma nova pasta chamada “logs” em seu bucket, na qual o Amazon EMR pode copiar os arquivos de log do seu cluster.
7. Em Configuração e permissões de segurança, escolha seu par de chaves do EC2. Na mesma seção, selecione o menu suspenso Função de serviço para o Amazon EMR e escolha EMR_. DefaultRole Em seguida, selecione a função do IAM no menu suspenso do perfil da instância e escolha EMR_EC2_. DefaultRole
8. Escolha Criar cluster para iniciar o cluster e abrir a página de detalhes do cluster.
9. Veja o Status do cluster próximo ao nome do cluster. O status é alterado de Iniciando para Em execução e, em seguida, para Aguardando, conforme o Amazon EMR provisiona o cluster. Pode ser necessário escolher o ícone de atualização à direita ou atualizar seu navegador para visualizar as atualizações de status.

O status do cluster é alterado para Aguardando quando o cluster está ativo, em execução e pronto para aceitar trabalhos. Para obter mais informações sobre como ler o resumo do cluster, consulte Visualizar o status e os detalhes do cluster. Para obter informações sobre o status do cluster, consulte Noções básicas sobre o ciclo de vida do cluster.

CLI

Para iniciar um cluster com o Spark instalado com o AWS CLI

1. Crie perfis do IAM padrão que podem ser usados para criar seu cluster ao usar o comando apresentado a seguir.

```
aws emr create-default-roles
```

Para obter mais informações sobre `create-default-roles`, consulte AWS CLI Command Reference.

2. Crie um cluster do Spark com o comando a seguir. Insira um nome para o seu cluster com a opção `--name` e especifique o nome do seu par de chaves do EC2 com a opção `--ec2-attributes`.

```
aws emr create-cluster \
--name "<My First EMR Cluster>" \
--release-label <emr-5.36.2> \
--applications Name=Spark \
--ec2-attributes KeyName=<myEMRKeyPairName> \
--instance-type m5.xlarge \
--instance-count 3 \
--use-default-roles
```

Observe os outros valores necessários para `--instance-type`, `--instance-count` e `--use-default-roles`. Esses valores foram escolhidos para clusters de uso geral. Para obter mais informações sobre `create-cluster`, consulte AWS CLI Command Reference.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

O resultado deverá ser parecido com o que segue. A saída mostra o `ClusterId` e o `ClusterArn` do seu novo cluster. Anote o seu `ClusterId`. Você usa o `ClusterId` para verificar o status do cluster e enviar trabalhos.

```
{
    "ClusterId": "myClusterId",
    "ClusterArn": "myClusterArn"
}
```

3. Verifique o status do seu cluster com o comando a seguir.

```
aws emr describe-cluster --cluster-id <myClusterId>
```

Você deverá visualizar uma saída semelhante à apresentada a seguir com o objeto `Status` para o seu novo cluster.

```
{
    "Cluster": {
        "Id": "myClusterId",
        "Name": "My First EMR Cluster",
        "Status": {
            "State": "STARTING",
            "StateChangeReason": {
                "Message": "Configuring cluster software"
            }
        }
     }
}
```

O valor de `State` é alterado de `STARTING` para `RUNNING` e, em seguida, para `WAITING`, conforme o Amazon EMR provisiona o cluster.

O status do cluster é alterado para **`WAITING`** quando um cluster está ativo, em execução e pronto para aceitar trabalhos. Para obter informações sobre o status do cluster, consulte Noções básicas sobre o ciclo de vida do cluster.

Etapa 2: gerenciar o cluster do Amazon EMR

Envio do trabalho para o Amazon EMR

Após iniciar um cluster, você poderá enviar trabalhos ao cluster em execução para processar e analisar dados. Você envia os trabalhos para um cluster do Amazon EMR como uma etapa. Uma etapa é uma unidade de trabalho composta por uma ou mais ações. Por exemplo, você pode enviar uma etapa para calcular valores ou para transferir e processar dados. É possível enviar etapas ao criar um cluster ou para um cluster em execução. Nesta parte do tutorial, você envia `health_violations.py` como uma etapa para o cluster em execução. Para saber mais sobre as etapas, consulte Enviar trabalhos a um cluster.

Console

Para enviar um aplicativo Spark como uma etapa com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e, em seguida, selecione o cluster para o qual deseja enviar o trabalho. O estado do cluster deve ser Aguardando.
3. Escolha a guia Etapas e, em seguida, escolha Adicionar etapa.
4. Configure a etapa de acordo com as seguintes diretrizes:

   - Para Tipo, escolha Aplicação do Spark. Você deverá visualizar campos adicionais para Modo de implantação, Local da aplicação e Opções de spark-submit.
   - Para Nome, insira um novo nome. Se você tiver muitas etapas em um cluster, nomear cada etapa ajudará a controlá-las.
   - Para Modo de implantação, deixe o valor padrão Modo de cluster. Para obter mais informações sobre os modos de implantação do Spark, consulte Cluster mode overview na documentação do Apache Spark.
   - Para Local da aplicação, insira o local do script `health_violations.py` no Amazon S3, como *`s3://DOC-EXAMPLE-BUCKET/health_violations.py`*.
   - Deixe o campo Opções de spark-submit vazio. Para obter mais informações sobre as opções de `spark-submit`, consulte Launching applications with spark-submit.
   - No campo Argumentos, insira os seguintes argumentos e valores:

     ```
     --data_source s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv
     --output_uri s3://DOC-EXAMPLE-BUCKET/myOutputFolder
     ```

     Substitua *`s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv`* pelo URI do bucket do S3 dos dados de entrada preparados em Preparação de uma aplicação com dados de entrada para o Amazon EMR.

     Substitua *`DOC-EXAMPLE-BUCKET`* pelo nome do bucket que você criou para este tutorial e substitua por um nome para a pasta *`myOutputFolder`*de saída do cluster.
   - Para Ação se a etapa falhar, aceite a opção padrão Continuar. Dessa forma, se a etapa falhar, o cluster continuará em execução.

5. Escolha Adicionar para enviar a etapa. A etapa deve ser exibida no console com o status Pendente.
6. Monitore o status da etapa. Ele deve ser alterado de Pendente para Em execução e, por fim, para Concluído. Para atualizar o status no console, escolha o ícone de atualização à direita de Filtrar. O script demora cerca de um minuto para ser executado. Quando o status for alterado para Concluído, a etapa será concluída com êxito.

CLI

Para enviar uma inscrição no Spark como uma etapa com o AWS CLI

1. Certifique-se de ter o `ClusterId` do cluster iniciado em Inicialização de um cluster do Amazon EMR. Também é possível recuperar o ID do cluster com o comando apresentado a seguir.

```
aws emr list-clusters --cluster-states WAITING
```

2. Envie `health_violations.py` como uma etapa com o comando `add-steps` e seu `ClusterId`.

   - Você pode especificar um nome para sua etapa ao substituir *"Minha aplicação do Spark"*. Na matriz `Args`, substitua *s3://DOC-EXAMPLE-BUCKET/health_violations.py* pelo local da aplicação `health_violations.py`.
   - Substitua *s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv* pelo local do S3 do seu conjunto de dados `food_establishment_data.csv`.
   - Substitua *s3://DOC-EXAMPLE-BUCKET/ MyOutputFolder* pelo caminho S3 do bucket designado e por um nome para a pasta de saída do cluster.
   - `ActionOnFailure=CONTINUE` significa que o cluster continuará em execução se a etapa falhar.

```
aws emr add-steps \
--cluster-id <myClusterId> \
--steps Type=Spark,Name="<My Spark
 Application>",ActionOnFailure=CONTINUE,Args=[<s3://DOC-EXAMPLE-
BUCKET/health_violations.py>,--data_source,<s3://DOC-EXAMPLE-BUCKET/
food_establishment_data.csv>,--output_uri,<s3://DOC-EXAMPLE-BUCKET/
MyOutputFolder>]
```

Para obter mais informações sobre o envio de etapas usando a CLI, consulte AWS CLI Command Reference.

Após enviar a etapa, você deverá visualizar uma saída como a apresentada a seguir, com uma lista de `StepIds`. Como você enviou uma etapa, verá somente um ID na lista. Copie o ID da etapa. Você usa o ID da etapa para verificar o status da etapa.

```
{
    "StepIds": [
        "s-1XXXXXXXXXXA"
    ]
}
```

3. Consulte o status da sua etapa com o comando `describe-step`.

```
aws emr describe-step --cluster-id <myClusterId> --step-id <s-1XXXXXXXXXXA>
```

Você deverá visualizar uma saída como a apresentada a seguir com informações sobre a etapa.

```
{
    "Step": {
        "Id": "s-1XXXXXXXXXXA",
        "Name": "My Spark Application",
        "Config": {
            "Jar": "command-runner.jar",
            "Properties": {},
            "Args": [
                "spark-submit",
                "s3://DOC-EXAMPLE-BUCKET/health_violations.py",
                "--data_source",
                "s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv",
                "--output_uri",
                "s3://DOC-EXAMPLE-BUCKET/myOutputFolder"
            ]
        },
        "ActionOnFailure": "CONTINUE",
        "Status": {
            "State": "COMPLETED"
        }
```

```
    }
}
```

O `State` da etapa é alterado de `PENDING` para `RUNNING` e para `COMPLETED`, conforme a etapa é executada. A etapa demora cerca de um minuto para ser executada, então pode ser necessário verificar o status algumas vezes.

Você saberá que a etapa obteve êxito quando o `State` for alterado para **`COMPLETED`**.

Para obter mais informações sobre o ciclo de vida da etapa, consulte Execução de etapas para processar dados.

Visualização dos resultados

Após a execução com êxito de uma etapa, você poderá visualizar os resultados de saída na pasta de saída do Amazon S3.

Visualizar os resultados de **`health_violations.py`**

1. Abra o console do Amazon S3 em https://console.aws.amazon.com/s3/.
2. Escolha o nome do bucket e, em seguida, a pasta de saída que você especificou ao enviar a etapa. Por exemplo, *`DOC-EXAMPLE-BUCKET`* e depois. *`myOutputFolder`*
3. Verifique se os seguintes itens aparecem na sua pasta de saída:
   - Um objeto de tamanho pequeno chamado `_SUCCESS`.
   - Um arquivo CSV que começa com o prefixo `part-`, que contém seus resultados.
4. Escolha o objeto com seus resultados e, em seguida, escolha Fazer download para salvar os resultados em seu sistema de arquivos local.
5. Abra os resultados no editor de sua preferência. O arquivo de saída lista os dez principais estabelecimentos de alimentação com o maior número de violações vermelhas. O arquivo de saída também mostra o número total de violações vermelhas para cada estabelecimento.

   Confira a seguir um exemplo de resultados para `health_violations.py`.

```
name, total_red_violations
SUBWAY, 322
T-MOBILE PARK, 315
```

```
WHOLE FOODS MARKET, 299
PCC COMMUNITY MARKETS, 251
TACO TIME, 240
MCDONALD'S, 177
THAI GINGER, 153
SAFEWAY INC #1508, 143
TAQUERIA EL RINCONSITO, 134
HIMITSU TERIYAKI, 128
```

Para obter mais informações sobre a saída do cluster do Amazon EMR, consulte Configurar um local de saída.

(Opcional) Conexão com o cluster do Amazon EMR em execução

Ao usar o Amazon EMR, você pode desejar se conectar a um cluster em execução para ler arquivos de log, depurar o cluster ou usar ferramentas da CLI, como o shell do Spark. O Amazon EMR permite que você se conecte a um cluster usando o protocolo Secure Shell (SSH). Esta seção abrange como configurar o SSH, conectar-se ao cluster e visualizar arquivos de log do Spark. Para obter mais informações sobre como se conectar a um cluster, consulte Autenticação em nós de cluster do Amazon EMR.

Autorização de conexões SSH para o cluster

Antes de se conectar ao cluster, é necessário modificar os grupos de segurança do cluster para autorizar conexões SSH de entrada. Os grupos de segurança do Amazon EC2 atuam como firewalls virtuais para controlar o tráfego de entrada e de saída do cluster. Quando você criou o cluster para este tutorial, o Amazon EMR criou os seguintes grupos de segurança em seu nome:

ElasticMapReduce-mestre

: O grupo de segurança gerenciado padrão do Amazon EMR associado ao nó primário. Em um cluster do Amazon EMR, o nó primário corresponde a uma instância do Amazon EC2 que gerencia o cluster.

ElasticMapReduce-escravo

: O grupo de segurança padrão associado aos nós centrais e de tarefa.

Console

Para permitir o acesso SSH a fontes confiáveis para o grupo de segurança primário com o console

Para editar seus grupos de segurança, você deve ter permissão para gerenciar os grupos de segurança para a VPC na qual o cluster está localizado. Para obter mais informações, consulte Alteração de permissões de um usuário e o exemplo de política que permite o gerenciamento de grupos de segurança do EC2 no Guia do usuário do IAM.

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha o cluster que você deseja atualizar. Isso abre a página de detalhes do cluster. A guia Propriedades nesta página deve estar pré-selecionada.
3. Em Redes na guia Propriedades, selecione a seta ao lado de Grupos de segurança do EC2 (firewall) para expandir esta seção. Em Nó primário, selecione o link do grupo de segurança. Ao concluir as etapas apresentadas a seguir, como opção, você poderá voltar a esta etapa, escolher Nós centrais e de tarefa e repetir as etapas a seguir para permitir o acesso do cliente SSH aos nós centrais e de tarefas.
4. Isso abre o console do EC2. Escolha a guia Regras de entrada e, em seguida, Editar regras de entrada.
5. Verifique se há uma regra de entrada que permita acesso público com as configurações a seguir. Se existir, escolha Excluir para removê-la.

   - Tipo

     SSH
   - Porta

     22
   - Origem

     Personalizado 0.0.0.0/0

> **Warning**
>
> Antes de dezembro de 2020, o grupo de segurança ElasticMapReduce -master tinha uma regra pré-configurada para permitir tráfego de entrada na Porta 22 de todas as fontes. Essa regra foi criada para simplificar as conexões SSH iniciais com o nó principal. Recomendamos fortemente remover esta regra de entrada e restringir o tráfego para origens confiáveis.

6. Role até o final da lista de regras e escolha Adicionar regra.
7. Em Type (Tipo), selecione SSH. Selecionar SSH insere automaticamente TCP para Protocolo e 22 para Intervalo de portas.
8. Para a origem, selecione Meu IP para adicionar automaticamente seu endereço IP como o endereço de origem. Você também pode adicionar um intervalo personalizado de endereços IP de clientes confiáveis ou criar regras adicionais para outros clientes. Diversos ambientes de rede alocam endereços IP dinamicamente, portanto, pode ser necessário atualizar os endereços IP para clientes confiáveis no futuro.
9. Escolha Salvar.
10. Como opção, escolha Nós centrais e de tarefa na lista e repita as etapas acima para permitir o acesso do cliente SSH aos nós centrais e de tarefa.

Old console

Para conceder acesso SSH a fontes confiáveis ao grupo de segurança primário com o console

Para editar seus grupos de segurança, você deve ter permissão para gerenciar os grupos de segurança para a VPC na qual o cluster está localizado. Para obter mais informações, consulte Alteração de permissões de um usuário e o exemplo de política que permite o gerenciamento de grupos de segurança do EC2 no Guia do usuário do IAM.

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Escolha Clusters. Escolha o ID do cluster que você deseja modificar.
3. No painel Rede e segurança, expanda o menu suspenso Grupos de segurança (firewall) do EC2.
4. Em Nó primário, escolha seu grupo de segurança.

5. Escolha Editar regras de entrada.
6. Verifique se há uma regra de entrada que permita acesso público com as configurações a seguir. Se existir, escolha Excluir para removê-la.

   - Tipo

     SSH
   - Porta

     22
   - Origem

     Personalizado 0.0.0.0/0

   > **Warning**
   >
   > Antes de dezembro de 2020, havia uma regra pré-configurada para permitir tráfego de entrada na porta 22 de todas as fontes. Esta regra foi criada para simplificar as conexões SSH iniciais com o nó primário. Recomendamos fortemente remover esta regra de entrada e restringir o tráfego para origens confiáveis.

7. Role até o final da lista de regras e escolha Adicionar regra.
8. Em Type (Tipo), selecione SSH.

   Selecionar SSH insere automaticamente TCP para Protocolo e 22 para Intervalo de portas.
9. Para a origem, selecione Meu IP para adicionar automaticamente seu endereço IP como o endereço de origem. Você também pode adicionar um intervalo personalizado de endereços IP de clientes confiáveis ou criar regras adicionais para outros clientes. Diversos ambientes de rede alocam endereços IP dinamicamente, portanto, pode ser necessário atualizar os endereços IP para clientes confiáveis no futuro.
10. Escolha Salvar.
11. Opcionalmente, escolha o outro grupo de segurança em Nós principais e de tarefas no painel Rede e segurança e repita as etapas acima para permitir que o cliente SSH acesse os nós principais e de tarefas.

Conecte-se ao seu cluster usando o AWS CLI

Independentemente do seu sistema operacional, é possível criar uma conexão SSH com o cluster usando a AWS CLI.

Para se conectar ao seu cluster e visualizar arquivos de log usando o AWS CLI

1. Use o comando a seguir para abrir uma conexão SSH com o cluster. Substitua *<mykeypair.key>* pelo caminho completo e pelo nome do arquivo do seu par de chaves. Por exemplo, `C:\Users\<username>\.ssh\mykeypair.pem`.

   ```
   aws emr ssh --cluster-id <j-2AL4XXXXXX5T9> --key-pair-file <~/mykeypair.key>
   ```

2. Navegue até `/mnt/var/log/spark` para acessar os logs do Spark no nó principal do cluster. Em seguida, visualize os arquivos nesse local. Para obter uma lista de arquivos de log adicionais no nó principal, consulte Visualizar arquivos de log no nó primário.

   ```
   cd /mnt/var/log/spark
   ls
   ```

Etapa 3: limpar os recursos do Amazon EMR

Encerramento do cluster

Agora que você enviou o trabalho para seu cluster e visualizou os resultados do seu PySpark aplicativo, você pode encerrar o cluster. O encerramento de um cluster interrompe todas as cobranças do Amazon EMR e das instâncias do Amazon EC2 associadas ao cluster.

Ao encerrar um cluster, o Amazon EMR retém os metadados relacionados ao cluster por dois meses gratuitamente. Os metadados arquivados ajudam a clonar o cluster para um novo trabalho ou a revisitar a configuração do cluster para finalidades de referência. Os metadados não incluem os dados que o cluster grava no S3 ou os dados armazenados no HDFS no cluster.

> **Note**
>
> O console do Amazon EMR não permite que você exclua um cluster da visualização de lista após o encerramento do cluster. Um cluster encerrado desaparecerá do console quando o Amazon EMR limpar os metadados.

Console

Para encerrar o cluster com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Escolha Clusters e, em seguida, selecione o cluster que você deseja encerrar.
3. No menu suspenso Ações, escolha Encerrar cluster.
4. Escolha Encerrar na caixa de diálogo. Dependendo da configuração do cluster, o encerramento pode demorar de cinco a dez minutos. Para obter mais informações sobre os clusters do Amazon EMR, consulte Terminar um cluster.

CLI

Para encerrar o cluster com o AWS CLI

1. Inicie o processo de encerramento do cluster com o comando a seguir. Substitua *< myClusterId >* pelo ID do seu cluster de amostra. O comando não retorna uma saída.

```
aws emr terminate-clusters --cluster-ids <myClusterId>
```

2. Para verificar se o processo de encerramento do cluster está em andamento, verifique o status do cluster com o comando a seguir.

```
aws emr describe-cluster --cluster-id <myClusterId>
```

A seguir é apresentado uma saída de exemplo no formato JSON. O `Status` do cluster deve ser alterado de **`TERMINATING`** para **`TERMINATED`**. O encerramento pode demorar de cinco a dez minutos, dependendo da configuração do cluster. Para obter mais informações sobre como encerrar um cluster do Amazon EMR, consulte Terminar um cluster.

```
{
    "Cluster": {
        "Id": "j-xxxxxxxxxxxxx",
        "Name": "My Cluster Name",
        "Status": {
            "State": "TERMINATED",
            "StateChangeReason": {
                "Code": "USER_REQUEST",
```

```
                "Message": "Terminated by user request"
            }
        }
    }
}
```

Exclusão de recursos do S3

Para evitar cobranças adicionais, você deve excluir o bucket do Amazon S3. Excluir o bucket remove todos os recursos do Amazon S3 deste tutorial. O bucket deve conter:

- O PySpark roteiro
- O conjunto de dados de entrada.
- Sua pasta de resultados de saída.
- Sua pasta de arquivos de log.

Talvez seja necessário tomar medidas adicionais para excluir os arquivos armazenados se você salvou o PySpark script ou a saída em um local diferente.

> **Note**
>
> O cluster deve ser encerrado antes de você excluir o bucket. Caso contrário, pode não ser possível esvaziar o bucket.

Para excluir seu bucket, siga as instruções apresentadas em How do I delete an S3 bucket? no Guia do usuário do Amazon Simple Storage Service.

Próximas etapas

Você iniciou seu primeiro cluster do Amazon EMR do início ao fim. Você também concluiu tarefas essenciais do EMR, como preparar e enviar aplicações de big data, visualizar os resultados e encerrar um cluster.

Use os tópicos apresentados a seguir para saber mais sobre como personalizar seu fluxo de trabalho do Amazon EMR.

Exploração de aplicações de big data para o Amazon EMR

Descubra e compare as aplicações de big data que podem ser instaladas em um cluster no Guia de versão do Amazon EMR. O guia de lançamento detalha cada versão lançada do EMR e inclui dicas para usar estruturas como o Spark e o Hadoop no Amazon EMR.

Planejamento do hardware, das redes e da segurança do cluster

Neste tutorial, você criou um cluster do EMR simples, sem configurar as opções avançadas. As opções avançadas permitem especificar os tipos de instância do Amazon EC2, as redes do cluster e a segurança do cluster. Para obter mais informações sobre como planejar e iniciar um cluster que atenda aos seus requisitos, consulte Planejar e configurar clusters e Segurança no Amazon EMR.

Gerenciar clusters

Aprofunde-se no trabalho com clusters em execução em Gerenciar clusters. Para gerenciar um cluster, é possível se conectar ao cluster, depurar etapas e rastrear as atividades e a integridade do cluster. Você também pode ajustar os recursos do cluster em resposta às demandas da workload com o ajuste de escala gerenciado do EMR.

Uso de uma interface diferente

Além do console do Amazon EMR, você pode gerenciar o Amazon EMR usando a API do AWS Command Line Interface serviço web ou um dos muitos SDKs compatíveis. AWS Para ter mais informações, consulte Interfaces de gerenciamento.

Você também pode interagir com aplicações instaladas em clusters do Amazon EMR de diversas maneiras. Algumas aplicações, como o Apache Hadoop, publicam interfaces da Web que você pode visualizar. Para ter mais informações, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

Navegação pelo blog técnico do EMR

Para obter exemplos de orientações e discussões técnicas aprofundadas sobre os novos recursos do Amazon EMR, consulte o blog de big data da AWS.

Console do Amazon EMR

O console oferece uma interface atualizada que fornece uma maneira intuitiva de gerenciar seu ambiente Amazon EMR e fornece acesso conveniente à documentação, informações sobre produtos e outros recursos.

Capacidades do console

O console do Amazon EMR está disponível no seguinte URL:

- URL do console — https://console.aws.amazon.com/emr

A tabela a seguir lista o status dos principais componentes do console do Amazon EMR.

Componente do console do Amazon EMR	Console	
EMR Studio	✓	
Criar e gerenciar clusters	✓	
Bloqueio de acesso público	✓	
Monitore CloudWatch eventos da Amazon	✓	
Configurações de segurança	✓	
Clusters virtuais (Amazon EMR no EKS)	✓	
Visualize e gerencie suas sub-redes da Amazon Virtual Private Cloud [1]	✓	
Cadernos [2]	✓	

[1] No console, você pode visualizar e gerenciar suas sub-redes da Amazon VPC na seção Rede ao criar um cluster.

[2] Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Resumo das diferenças

Esta seção descreve os recursos da experiência do console do Amazon EMR. Esses recursos se enquadram nas seguintes categorias:

- Compatibilidade de clusters no console
- Criar clusters
- Visualizando ou editando detalhes do cluster
- Visualizando e pesquisando clusters
- Diferenças no trabalho com configurações de segurança

Compatibilidade de clusters no console

Em alguns casos, um cluster que você criou pode não ser compatível com o console. A lista a seguir descreve os requisitos de compatibilidade para o console do Amazon EMR.

- O console oferece suporte a clusters criados nas versões 5.20.1 e posteriores do Amazon EMR.
- Você pode clonar clusters que usam escalabilidade automática no console, mas você só pode criar novos clusters se quiser escalá-los manualmente ou usar escalabilidade gerenciada.

Para criar e trabalhar com clusters da versão 5.20.1 e anteriores, você pode usar o AWS Command Line Interface (AWS CLI) ou o AWS SDK.

Criar clusters

Recurso	Console	
Terminologia: tipos de nós de cluster do Amazon EMR	Primário, central e de tarefa	

Recurso	Console	
Versões do Amazon EMR com suporte[1]	Versão 5.20.1 e posterior do Amazon EMR	
Início rápido de um cluster	Use o botão Criar cluster no painel Resumo. O nome do cluster não pode conter os caractere s <, >, $, | ou `(crase).	
Configuração de um tempo limite para o provisionamento spot	Defina um período de tempo limite para o provision amento de instâncias para cada frota no cluster.	
Perfis de serviço e perfil para o perfil de instância do Amazon EC2	O console não cria funções padrão; você deve criar funções com o console do IAM ou selecionar uma função do IAM já criada	
Visibilidade do cluster	No console do Amazon EMR, não é possível tornar um cluster visível para todos os usuários. Sua política do IAM determina o acesso ao cluster.	
Redes: configuração de sub-redes privadas	Você deve configurar endpoints do Amazon S3 e gateways NAT usando os respectivos consoles do Amazon S3 e do Amazon VPC.	

Recurso	Console	
Visualização consistente do Sistema de Arquivos do EMR (EMRFS CV)	Com o lançamento da read-after-write consistência forte do Amazon S3 em 1º de dezembro de 2020, você não precisa usar o EMRFS CV com seus clusters do EMR	
Depuração	Você pode depurar trabalhos usando a interface do usuário da aplicação na página de detalhes do cluster.	

[1] Você não pode criar ou editar clusters usando versões anteriores ao Amazon EMR 5.20.1 no console, mas todos os clusters existentes criados usando versões anteriores à 5.20.1 continuarão funcionando. Para criar e editar clusters com versões do Amazon EMR anteriores à 5.20.1, use a API ou a CLI. Você pode visualizar todos os clusters usando o console, mas os consoles criados antes da versão 5.20.1 podem não ser compatíveis com os recursos mais recentes.

Visualizando e pesquisando clusters

A tabela a seguir destaca como você pode usar o console do Amazon EMR para visualizar, visualizar e pesquisar clusters.

> **Note**
>
> A aplicação de um filtro de dados à lista de clusters consulta todo o banco de dados. Entretanto, ao inserir uma string de texto na caixa de pesquisa, a pesquisa se aplica somente aos resultados que a lista carregou no lado do cliente.

Recurso	Console	
Visualização de detalhes do cluster	Você pode selecionar o ID do cluster para visualizar	

Recurso	Console	
	os detalhes completos do cluster, como as opções de configuração, as interfaces do usuário de aplicações persistentes e os logs.	
Pesquisa de clusters	Use um único campo de pesquisa para inserir consultas de pesquisa de texto e para criar e aplicar filtros de dados como “Status = qualquer status ativo”.	
Descoberta de clusters com falha	Para pesquisar clusters com falha, aplique o filtro Status = Encerrado com erros.	

Visualizando ou editando detalhes do cluster

Recurso	Console	
Visualização das instâncias em seus grupos de instâncias e frotas de instâncias, em conjunto com opções de escalabilidade, provisionamento, redimensionamento e encerramento.	Veja as opções e os detalhes da instância na guia Instâncias. Veja as opções de encerramento na guia Propriedades.	
Visualização de interfaces do usuário, logs e configurações de aplicações	Veja as configurações do cluster na guia Configurações. Inicie uma interface	

Recurso	Console	
(interface do usuário do Apache Spark, servidor de histórico do Spark, interface do usuário do Tez, servidor de linha do tempo do YARN)	do usuário da aplicação dinâmica e persistente para visualizar os logs de uma aplicação na guia Aplicações.	
Exportação de um cluster para a CLI	Opção disponível nos menus de detalhes e de visualização da listagem de Ações do cluster como “Visualizar comando para clonar cluster”.	

Diferenças no trabalho com configurações de segurança

Recurso	Console	
Clonagem de configurações de segurança	✓	
Governança federada ao usar Trino e Apache Ranger	✓	
Uso de um perfil de runtime para envio de trabalhos a um cluster[1]	✓	
Autorização de acesso aos dados do Sistema de Arquivos do EMR (EMRFS)	Pontos de acesso Amazon S3	
	Perfis de runtime	

Recurso	Console	
AWS Lake Formation controles de acesso		

[1] Para transmitir um perfil durante o envio da etapa, o cluster deve usar uma configuração de segurança com uma política de permissões do IAM anexada para que o usuário possa transmitir somente os perfis aprovados e os trabalhos possam acessar os recursos do Amazon EMR. Para ter mais informações, consulte Perfis de runtime para etapas ao Amazon EMR.

Amazon EMR Studio

O Amazon EMR Studio é um ambiente de desenvolvimento integrado (IDE) baseado na Web para cadernos Jupyter totalmente gerenciados que são executados em clusters do Amazon EMR. Você pode configurar um EMR Studio para que sua equipe desenvolva, visualize e depure aplicativos escritos em R, Python, Scala e. PySpark O EMR Studio é integrado ao AWS Identity and Access Management (IAM) e ao Centro de Identidade do IAM para que os usuários possam fazer login usando suas credenciais corporativas.

É possível criar um EMR Studio gratuitamente. As cobranças aplicáveis para o armazenamento do Amazon S3 e para os clusters do Amazon EMR se aplicam quando você usa o EMR Studio. Para obter detalhes e destaques do produto, consulte a página de serviços do Amazon EMR Studio.

Principais recursos do EMR Studio

O Amazon EMR Studio oferece os seguintes recursos:

- Autentique usuários com o AWS Identity and Access Management (IAM) ou o AWS IAM Identity Center com ou sem a propagação de identidade confiável e seu provedor de identidade empresarial.
- Acesse e execute clusters do Amazon EMR sob demanda para executar trabalhos do caderno Jupyter.
- Conexão aos clusters do Amazon EMR no EKS para enviar trabalhos à medida que o trabalho é executado.
- Navegação e salvamento de cadernos de exemplo. Para obter mais informações sobre exemplos de notebooks, consulte o repositório de exemplos de notebooks GitHub do EMR Studio.
- Analise dados usando Python, Spark Scala PySpark, Spark R ou SparkSQL e instale kernels e bibliotecas personalizados.
- Colaboração em tempo real com outros usuários no mesmo Workspace. Para ter mais informações, consulte Configuração da colaboração no Workspace.
- Uso do SQL Explorer do EMR Studio para navegar em seu catálogo de dados, executar consultas SQL e fazer download de resultados antes do trabalho com os dados em um caderno.
- Execução de cadernos parametrizados como parte dos fluxos de trabalho programados com uma ferramenta de orquestração, como o Apache Airflow ou o Amazon Managed Workflows for Apache

Airflow. Para obter mais informações, consulte Orchestrating analytics jobs on EMR Notebooks using MWAA no blog de Big Data da AWS.

- Vincule repositórios de código, como GitHub e. BitBucket
- Rastreamento e depuração de trabalhos usando o servidor de histórico do Spark, a interface do usuário do Tez ou o servidor de linha do tempo do YARN.

O EMR Studio também é elegível para a HIPAA e é certificado pela HITRUST CSF e pelo SOC 2. Para obter mais informações sobre a conformidade com a HIPAA para serviços da AWS, consulte https://aws.amazon.com/compliance/hipaa-compliance/. Para saber mais sobre a conformidade da HITRUST CSF para serviços da AWS, consulte https://aws.amazon.com/compliance/hitrust/. Para obter mais informações sobre outros programas de conformidade para serviços da AWS, consulte Serviços da AWS no escopo por programa de conformidade.

Histórico de recursos do Amazon EMR Studio

Esta tabela lista as atualizações na funcionalidade de ajuste de escala gerenciado do Amazon EMR.

Data de lançamento	Recurso
5 de janeiro de 2024	Foi adicionado suporte para o EMR Studio em AWS GovCloud (Leste dos EUA) e AWS GovCloud (Oeste dos EUA).
26 de novembro de 2023	Foi adicionado suporte à propagação de identidade confiável para o EMR Studio com a autenticação do Centro de Identidade do IAM.
26 de outubro de 2023	Capacidade adicional de criar uma aplicação do EMR Serverless com capacidade interativa.
28 de fevereiro de 2023	Adição de suporte para chaves gerenciadas pelo cliente do AWS KMS para o armazenamento de logs de aplicações para aplicações do EMR Sem Servidor.
23 de fevereiro de 2023	Adição da criação de perfil do IAM com um clique para envio de trabalhos do EMR Sem Servidor. Adição de pesquisa do ECR para quando você seleciona uma imagem personalizada para aplicações do EMR Sem Servidor.

Data de lançamento	Recurso
27 de janeiro de 2023	Os cadernos de execução descentralizados podem rastrear o progresso da execução de cada célula com a mágica `%execute_notebook` .
23 de janeiro de 2023	As aplicações persistentes foram otimizadas para a obtenção de tempos de inicialização mais rápidos.

Como o Amazon EMR Studio funciona

Um Amazon EMR Studio é um recurso do Amazon EMR criado para uma equipe de usuários. Cada Studio corresponde a um ambiente de desenvolvimento integrado que é independente e baseado na Web para cadernos Jupyter executados em clusters do Amazon EMR. Os usuários fazem login em um Studio usando credenciais corporativas.

Cada EMR Studio criado usa os seguintes recursos da AWS:

- Uma Amazon Virtual Private Cloud (VPC) com sub-redes: os usuários executam kernels e aplicações do Studio no Amazon EMR e clusters do Amazon EMR no EKS na VPC especificada. Um EMR Studio pode se conectar a qualquer cluster nas sub-redes especificadas na criação do Studio.
- Políticas de permissões e perfis do IAM: para gerenciar as permissões de usuários, você cria políticas de permissões do IAM que são anexadas à identidade do IAM de um usuário ou a um perfil de usuário. O EMR Studio também usa um perfil de serviço do IAM e grupos de segurança para interoperar com outros serviços da AWS. Para obter mais informações, consulte Controle de acesso e Definição de grupos de segurança para controlar o tráfego de rede do EMR Studio.
- Grupos de segurança: o EMR Studio usa grupos de segurança para estabelecer um canal de rede seguro entre o Studio e um cluster do EMR.
- Um local de backup do Amazon S3: o EMR Studio salva o trabalho do caderno em um local do Amazon S3.

As seguintes etapas descrevem como criar e administrar um EMR Studio:

1. Crie um Studio em sua Conta da AWS com a autenticação do IAM ou do Centro de Identidade do IAM. Para obter instruções, consulte Configuração de um Amazon EMR Studio.

2. Atribua usuários e grupos ao seu Studio. Use políticas de permissões para definir permissões detalhadas para cada usuário. Para obter mais informações, consulte o tópico Atribua e gerencie usuários do EMR Studio
3. Comece a monitorar as ações do EMR Studio com eventos do AWS CloudTrail. Para obter mais informações, consulte Monitoramento das ações do Amazon EMR Studio.
4. Forneça mais opções de cluster aos usuários do Studio com modelos de cluster e endpoints gerenciados do Amazon EMR no EKS.

Autenticação e login do usuário

O Amazon EMR Studio oferece suporte a dois modos de autenticação: o modo de autenticação do IAM e o modo de autenticação do Centro de Identidade do IAM. O modo do IAM usa o AWS Identity and Access Management (IAM), enquanto o modo do Centro de Identidade do IAM usa o AWS IAM Identity Center. Ao criar um EMR Studio, você escolhe o modo de autenticação para todos os usuários desse Studio.

Modo de autenticação do IAM

Com o modo de autenticação do IAM, você pode usar a autenticação do IAM ou a federação do IAM.

A autenticação do IAM permite gerenciar identidades do IAM, como usuários, grupos e perfis no IAM. Você concede aos usuários acesso a um Studio com as políticas de permissões do IAM e o controle de acesso por atributos (ABAC).

A federação do IAM permite estabelecer confiança entre um provedor de identidades (IdP) terceirizado e a AWS para que você possa gerenciar identidades de usuários por meio do seu IdP.

Modo de autenticação do Centro de Identidade do IAM

O modo de autenticação do Centro de Identidade do IAM permite conceder aos usuários o acesso federado a um EMR Studio. Você pode usar o Centro de Identidade do IAM para autenticar usuários e grupos do diretório do Centro de Identidade do IAM, do diretório corporativo existente ou de um IdP externo, como o Azure Active Directory (AD). Em seguida, você gerencia os usuários com o seu provedor de identidades (IdP).

O EMR Studio oferece suporte ao uso dos seguintes provedores de identidades para o Centro de Identidade do IAM:

- AWS Managed Microsoft AD e Active Directory autogerenciado: para obter mais informações, consulte Connect to your Microsoft AD directory.
- Provedores baseados em SAML: para obter uma lista completa, consulte Supported identity providers.
- O diretório do Centro de Identidade do IAM: para obter mais informações, consulte Gerenciamento de identidades no Centro de Identidade do IAM e Trusted identity propagation across applications no Guia do usuário do AWS IAM Identity Center.

Como a autenticação afeta o login e a atribuição de usuários

O modo de autenticação escolhido para o EMR Studio afeta como os usuários fazem login em um Studio, como você atribui um usuário a um Studio e como você autoriza (concede permissões) aos usuários para executar ações, como a criação de novos clusters do Amazon EMR.

A tabela a seguir resume os métodos de login do EMR Studio de acordo com o modo de autenticação.

Opções de login do EMR Studio por modo de autenticação

Modo de autenticação	Método de login	Descrição
• IAM (autenticação e federação) • IAM Identity Center	URL do EMR Studio	Os usuários fazem login em um Studio usando o URL de acesso ao Studio. Por exemplo, `https://xxxxxxxxxxxxxxxxxxxxxxxx.emrstudio-prod.us-east-1.amazonaws.com` . Os usuários inserem as credenciais do IAM quando você usa a autenticação do IAM. Quando você usa a federação do IAM ou o Centro de Identidade do IAM, o EMR Studio redireciona os usuários para o URL de login do seu provedor de identidades para a inserção das credenciais. No contexto da federação de identidades, esta opção de login é chamada de login iniciado com base no provedor de serviços (SP).

Modo de autenticação	Método de login	Descrição
• IAM (federação) • IAM Identity Center	Portal do provedor de identidades (IdP)	Os usuários fazem login no portal do seu provedor de identidades, como o portal do Azure, e iniciam o console do Amazon EMR. Após iniciarem o console do Amazon EMR, os usuários selecionam e abrem um Studio pela lista Studios. Você também pode configurar o EMR Studio como uma aplicação da SAML para que os usuários possam fazer login em um Studio específico usando o portal do seu provedor de identidades. Para obter instruções, consulte Para configurar um EMR Studio como uma aplicação da SAML em seu portal do IdP. No contexto da federação de identidades, esta opção de login é chamada de login iniciado com base no provedor de identidades (IdP).
• IAM (autenticação)	AWS Management Console	Os usuários fazem login no AWS Management Console usando as credenciais do IAM e abrem um Studio pela lista Studios no console do Amazon EMR.

A tabela a seguir descreve a atribuição e a autorização de usuários para o EMR Studio pelo modo de autenticação.

Atribuição e autorização de usuários do EMR Studio pelo modo de autenticação

Modo de autenticação	Atribuição de usuários	Autorização de usuários
IAM (autenticação e federação)	Permita a ação `CreateStudioPresignedUrl` em uma política de permissões do IAM	Defina políticas de permissões do IAM que permitem determinadas ações do EMR Studio.

Modo de autenticação	Atribuição de usuários	Autorização de usuários
	anexada a uma identidade do IAM (usuário, grupo ou perfil). Para usuários federados, permita a ação `CreateStudioPresignedUrl` em um IAM na política de permissões configurada para o perfil do IAM que é usado para a federação. Use o controle de acesso por atributo (ABAC) para especificar o Studio ou os Studios que o usuário pode acessar. Para obter instruções, consulte Atribuir um usuário ou um grupo a um EMR Studio.	Para usuários nativos, anexe a política de permissões do IAM a uma identidade do IAM (usuário, grupo ou perfil). Para usuários federados, permita as ações do Studio na política de permissões configurada para o perfil do IAM que é usado para a federação. Para obter mais informações, consulte Configurar permissões de usuário do EMR Studio para Amazon EC2 ou Amazon EKS.
IAM Identity Center	Para Studios criados com `IdCUserAssignment` definido como `REQUIRED`, mapeie os usuários para o Studio com uma política de sessão especificada. Para obter mais informações, consulte Atribuir um usuário ou um grupo a um EMR Studio. Para Studios criados com `IdCUserAssignment` definido como `OPTIONAL`, qualquer usuário ou grupo do Centro de Identidade pode acessar o Studio.	Opcional: defina políticas de sessão do IAM que permitam determinadas ações do EMR Studio. Mapeie uma política de sessão para um usuário ao atribuir o usuário a um Studio. Para obter mais informações, consulte Permissões de usuários para o modo de autenticação do Centro de Identidade do IAM.

Controle de acesso

No Amazon EMR Studio, você configura a autorização (permissões) de usuários com as políticas baseadas em identidade do AWS Identity and Access Management (IAM). Nessas políticas, você especifica as ações e os recursos permitidos, bem como as condições sob as quais as ações são permitidas.

Permissões de usuários para o modo de autenticação do IAM

Para definir as permissões de usuários ao usar a autenticação do IAM para o EMR Studio, você permite ações, como `elasticmapreduce:RunJobFlow`, em uma política de permissões do IAM. Você pode criar uma ou mais políticas de permissões para usar. Por exemplo, é possível criar uma política básica, que não permita que um usuário crie novos clusters do Amazon EMR, e outra política que permita a criação de clusters. Para obter uma lista de todas as ações do Studio, consulte Permissões do AWS Identity and Access Management para usuários do EMR Studio.

Permissões de usuários para o modo de autenticação do Centro de Identidade do IAM

Ao usar a autenticação do Centro de Identidade do IAM, você cria um único perfil de usuário do EMR Studio. O perfil de usuário corresponde a um perfil do IAM dedicado que um Studio assume quando um usuário faz login.

Você anexa políticas de sessão do IAM ao perfil de usuário do EMR Studio. Uma política de sessão é um tipo especial de política de permissões do IAM que limita o que um usuário federado pode fazer durante uma sessão de login do Studio. As políticas de sessão possibilitam definir permissões específicas para um usuário ou para um grupo sem a necessidade de criar diversos perfil de usuário para o EMR Studio.

Ao atribuir usuários e grupos a um Studio, você mapeia uma política de sessão para esse usuário ou grupo para a aplicação de permissões detalhadas. Você também pode atualizar a política de sessão de um usuário ou de um grupo a qualquer momento. O Amazon EMR armazena cada mapeamento de política de sessão criado.

Para obter mais informações sobre as políticas de sessão, consulte Políticas e permissões no Guia do usuário do AWS Identity and Access Management.

Workspaces

Os Workspaces são os principais componentes básicos do Amazon EMR Studio. Para organizar os cadernos, os usuários criam um ou mais Workspaces em um Studio. Para obter mais informações, consulte Compreensão das noções básicas do Workspace.

Semelhante aos espaços de trabalho no JupyterLab, um Workspace preserva o estado de trabalho do caderno. No entanto, a interface do usuário do Workspace amplia a interface do JupyterLab de código aberto com ferramentas adicionais para permitir que você crie e anexe clusters do EMR, execute trabalhos, explore cadernos de exemplo e vincule repositórios Git.

A seguinte lista inclui os principais recursos dos Workspaces do EMR Studio:

- A visibilidade do Workspace é baseada no Studio. Os Workspaces criados em um Studio não são visíveis em outros Studios.
- Por padrão, um Workspace é compartilhado e pode ser visualizado por todos os usuários do Studio. No entanto, somente um usuário pode abrir e trabalhar em um Workspace por vez. Para trabalhar simultaneamente com outros usuários, é possível realizar a Configuração da colaboração no Workspace.
- Você pode colaborar simultaneamente com outros usuários em um Workspace ao habilitar a colaboração no Workspace. Para obter mais informações, consulte Configuração da colaboração no Workspace.
- Os cadernos em um Workspace compartilham o mesmo cluster do EMR para a execução de comandos. Você pode anexar um Workspace a um cluster do Amazon EMR em execução no Amazon EC2 ou a um cluster virtual e a um endpoint gerenciado do Amazon EMR no EKS.
- Os Workspaces podem ser alternados para outra zona de disponibilidade associada às sub-redes de um Studio. Você pode interromper e reiniciar um Workspace para solicitar o processo de failover. Ao reiniciar um Workspace, o EMR Studio inicia o Workspace em uma zona de disponibilidade diferente na VPC do Studio quando o Studio está configurado com acesso a diversas zonas de disponibilidade. Se o Studio tiver somente uma zona de disponibilidade, o EMR Studio tentará iniciar o Workspace em uma sub-rede diferente. Para obter mais informações, consulte Resolução de problemas de conectividade do Workspace.
- Um Workspace pode se conectar a clusters em qualquer uma das sub-redes associadas a um Studio.

Para obter mais informações sobre como criar e configurar Workspaces do EMR Studio, consulte Compreensão das noções básicas do Workspace.

Armazenamento de cadernos no Amazon EMR Studio

Quando você usa um Workspace, o EMR Studio salva automaticamente as células em arquivos de cadernos em uma cadência regular no local do Amazon S3 associado ao seu Studio. Esse processo de backup preserva o trabalho entre as sessões para que você possa voltar a ele mais tarde sem a necessidade de confirmar as alterações em um repositório Git. Para obter mais informações, consulte Salvamento de conteúdo do Workspace.

Quando você exclui um arquivo de caderno de um Workspace, o EMR Studio exclui a versão de backup do Amazon S3 para você. No entanto, se você excluir um Workspace sem primeiro excluir os arquivos do cadernos, estes arquivos permanecerão no Amazon S3 e continuarão a acumular cobranças de armazenamento. Para saber mais, consulte Exclusão de um Workspace e de arquivos de cadernos.

Considerações sobre o EMR Studio

Considerações

Considere o seguinte ao trabalhar com o EMR Studio:

- O EMR Studio está disponível da seguinte forma: Regiões da AWS
  - Leste dos EUA (Ohio) (us-east-2)
  - Leste dos EUA (Norte da Virgínia) (us-east-1)
  - Oeste dos EUA (Norte da Califórnia) (us-west-1)
  - Oeste dos EUA (Oregon) (us-west-2)
  - África (Cidade do Cabo) (af-south-1)
  - Ásia-Pacífico (Hong Kong) (ap-east-1)
  - Ásia-Pacífico (Jacarta) (ap-southeast-3) *
  - Ásia-Pacífico (Melbourne) (ap-southeast-4) *
  - Ásia-Pacífico (Mumbai) (ap-south-1)
  - Ásia-Pacífico (Osaka) (ap-northeast-3) *
  - Ásia-Pacífico (Seul) (ap-northeast-2)
  - Ásia-Pacífico (Singapura) (ap-southeast-1)
  - Ásia-Pacífico (Sydney) (ap-southeast-2)
  - Ásia Pacific (Tóquio) (ap-northeast-1)

  - Canadá (Central) (ca-central-1)
  - Europa (Frankfurt) (eu-central-1)
  - Europa (Irlanda) (eu-west-1)
  - Europa (Londres) (eu-west-2)
  - UE (Milão) (eu-south-1)
  - Europa (Paris) (eu-west-3)
  - Europa (Espanha) (eu-south-2)
  - UE (Estocolmo) (eu-north-1)
  - Europa (Zurique) (eu-central-2) *
  - Israel (Tel Aviv) (il-central-1)*
  - Oriente Médio (EAU) (me-central-1) *
  - América do Sul (São Paulo) (sa-east-1)
  - AWS GovCloud (Leste dos EUA) (gov-us-east-1)
  - AWS GovCloud (Oeste dos EUA) (gov-us-west-1)

  * A interface ativa do Spark não é compatível com essas regiões.
- Para permitir que os usuários provisionem novos clusters do EMR em execução no Amazon EC2 para um Workspace, você pode associar um EMR Studio a um conjunto de modelos de cluster. Os administradores podem definir modelos de cluster com o Service Catalog e escolher se um usuário ou um grupo pode acessar os modelos de cluster, ou nenhum modelo de cluster, em um Studio.
- Ao definir permissões de acesso aos arquivos do notebook armazenados no Amazon S3 ou ler segredos AWS Secrets Manager, use a função de serviço do Amazon EMR. As políticas de sessão não são compatíveis com estas permissões.
- Você pode criar diversos EMR Studios para controlar o acesso a clusters do EMR em diferentes VPCs.
- Use o AWS CLI para configurar o Amazon EMR em clusters EKS. Em seguida, é possível usar a interface do Studio para anexar clusters a Workspaces com um endpoint gerenciado para executar trabalhos de cadernos.
- Há outras considerações ao usar a propagação de identidade confiável com o Amazon EMR que também se aplicam ao EMR Studio. Para ter mais informações, consulte Considerações e limitações do Amazon EMR com a integração do Centro de Identidade.
- O EMR Studio não oferece suporte aos seguintes comandos mágicos do Python:
  - `%alias`

  - `%alias_magic`
  - `%automagic`
  - `%macro`
  - `%%js`
  - `%%javascript`
  - Modificar `proxy_user` usando `%configure`
  - Modificar `KERNEL_USERNAME` usando `%env` ou `%set_env`
- O Amazon EMR em clusters EKS não oferece suporte a SparkMagic comandos para o EMR Studio.
- Para escrever instruções do Scala com várias linhas em células de cadernos, certifique-se de que todas as linhas, exceto a última, terminem com um ponto final. O exemplo a seguir usa a sintaxe adequada para instruções do Scala com várias linhas.

```
val df = spark.sql("SELECT * from table_name).
        filter("col1=='value'").
        limit(50)
```

- Para aumentar a segurança das aplicações fora do console que podem ser usadas com o Amazon EMR, os domínios de hospedagem das aplicações são registrados na Public Suffix List (PSL). Exemplos desses domínios de hospedagem incluem os seguintes: `emrstudio-prod.us-east-1.amazonaws.com`, `emrnotebooks-prod.us-east-1.amazonaws.com`, `emrappui-prod.us-east-1.amazonaws.com`. Para maior segurança, se precisar definir cookies confidenciais no nome de domínio padrão, recomendamos que você use cookies com um prefixo `__Host-`. Isso ajuda a defender seu domínio contra tentativas de falsificação de solicitação entre sites (CSRF). Para obter mais informações, consulte a página Set-Cookie em Mozilla Developer Network.

Problemas conhecidos

- Um EMR Studio que usa o Centro de Identidade do IAM com a propagação de identidade confiável habilitada só pode se associar a clusters do EMR que também usam a propagação de identidade confiável.
- Certifique-se de desativar as ferramentas de gerenciamento de proxy, como FoxyProxy ou SwitchyOmega, no navegador antes de criar um Studio. Os proxies ativos podem causar erros quando você escolhe Criar Studio e resultar em uma mensagem de erro de falha de rede.

- Os kernels executados em clusters do Amazon EMR no EKS podem falhar ao iniciar devido a problemas de tempo limite. Se você encontrar um erro ou problema ao iniciar o kernel, feche o arquivo de caderno, encerre o kernel e reabra o arquivo de caderno.
- A operação Reiniciar kernel não funciona conforme o esperado quando você usa um cluster do Amazon EMR no EKS. Após selecionar Reiniciar kernel, atualize o Workspace para que a reinicialização entre em vigor.
- Se um Workspace não estiver anexado a um cluster, uma mensagem de erro será exibida quando um usuário do Studio abrir um arquivo de caderno e tentar selecionar um kernel. Você pode ignorar essa mensagem de erro ao escolher OK, mas deve anexar o Workspace a um cluster e selecionar um kernel antes de poder executar o código do caderno.
- Ao usar o Amazon EMR 6.2.0 com uma configuração de segurança para definir a segurança do cluster, a interface do Workspace aparece em branco e não funciona conforme o esperado. Recomendamos usar uma versão diferente do Amazon EMR com suporte, se desejar configurar a criptografia de dados ou a autorização do Amazon S3 para o EMRFS em um cluster. O EMR Studio funciona com as versões 5.32.0 (série 5.x) e 6.2.0 (série 6.x) e superiores do Amazon EMR.
- Ao realizar a Depuração do Amazon EMR em execução em trabalhos do Amazon EC2, os links para a interface do usuário do Spark no cluster podem não funcionar ou não aparecer. Para gerar os links novamente, crie uma nova célula de caderno e execute o comando `%%info`.
- O Jupyter Enterprise Gateway não limpa os kernels ociosos no nó primário de um cluster nas seguintes versões de liberação do Amazon EMR: 5.32.0, 5.33.0, 6.2.0 e 6.3.0. Os kernels ociosos consomem recursos de computação e podem causar falhas em clusters de longa execução. Você pode configurar a limpeza de kernels ociosos para o Jupyter Enterprise Gateway usando o script de exemplo a seguir. É possível Conectar-se ao nó primário usando SSH ou enviar o script como uma etapa. Para obter mais informações, consulte Run commands and scripts on an Amazon EMR cluster.

```
#!/bin/bash
sudo tee -a /emr/notebook-env/conf/jupyter_enterprise_gateway_config.py << EOF
c.MappingKernelManager.cull_connected = True
c.MappingKernelManager.cull_idle_timeout = 10800
c.MappingKernelManager.cull_interval = 300
EOF
sudo systemctl daemon-reload
sudo systemctl restart jupyter_enterprise_gateway
```

- Quando você usa uma política de encerramento automático com as versões 5.32.0, 5.33.0, 6.2.0 ou 6.3.0 do Amazon EMR, o Amazon EMR marca um cluster como ocioso e pode encerrá-lo

automaticamente mesmo se você tiver um kernel do Python3 ativo. Isso ocorre porque a execução de um kernel do Python3 não envia um trabalho do Spark no cluster. Para usar o encerramento automático com um kernel do Python3, recomendamos usar a versão 6.4.0 ou as versões posteriores do Amazon EMR. Para obter mais informações sobre o encerramento automático, consulte Usar uma política de término automático.

- Quando você usa `%%display` para exibir um Spark DataFrame em uma tabela, tabelas muito largas podem ficar truncadas. Você pode clicar com o botão direito do mouse na saída e selecionar Criar nova visualização para a saída para obter uma visualização da saída com rolagem.
- Iniciar um kernel baseado em Spark, como PySpark Spark ou SparkR, inicia uma sessão do Spark, e executar uma célula em um notebook coloca as tarefas do Spark em fila nessa sessão. Quando você interrompe uma célula em execução, o trabalho do Spark continua a ser executado. Para interromper o trabalho do Spark, você deve usar a interface do usuário do Spark no cluster. Para obter instruções sobre como se conectar à interface do usuário do Spark, consulte Depuração de aplicações e trabalhos com o EMR Studio.

Limitações de recursos

O Amazon EMR Studio não oferece suporte aos seguintes recursos do Amazon EMR:

- Anexação e execução de trabalhos em clusters do EMR com uma configuração de segurança que especifica a autenticação do Kerberos.
- Clusters com vários nós primários.
- Clusters que usam instâncias do Amazon EC2 com base no AWS Graviton2 para versões 6.x do Amazon EMR inferiores a 6.9.0 e versões 5.x inferiores a 5.36.1

Os recursos a seguir não são compatíveis com um Studio que usa a propagação de identidade confiável:

- Criação de clusters do EMR sem um modelo.
- Uso de aplicações do EMR Sem Servidor.
- Execução de clusters do Amazon EMR no EKS.
- Uso de um perfil de runtime.
- Ativação da colaboração do SQL Explorer ou do Workspace.

Limites de serviço para o EMR Studio

A tabela a seguir exibe os limites de serviço para o EMR Studio.

Item	Limite
EMR Studios	Máximo de 100 por AWS conta
Subredes	Máximo de cinco associações para cada EMR Studio
Grupos do Centro de Identidade do IAM	Máximo de cinco atribuições para cada EMR Studio
Usuários do Centro de Identidade do IAM	Máximo de cem atribuições para cada EMR Studio

Práticas recomendadas para VPC e para sub-rede

Use as seguintes melhores práticas para configurar uma Amazon Virtual Private Cloud (Amazon VPC) com sub-redes para o EMR Studio:

- Você pode especificar, no máximo, cinco sub-redes em sua VPC para serem associadas ao Studio. Recomendamos fornecer várias sub-redes em diferentes zonas de disponibilidade para oferecer suporte à disponibilidade do Workspace e disponibilizar aos usuários do Studio o acesso a clusters em diferentes zonas de disponibilidade. Para saber mais sobre como trabalhar com VPCs, sub-redes e zonas de disponibilidade, consulte VPCs e sub-redes no Guia do usuário da Amazon Virtual Private Cloud .
- As sub-redes especificadas deverão ser capazes de se comunicar entre si.
- Para permitir que os usuários vinculem um Workspace a repositórios Git hospedados publicamente, você deve especificar somente sub-redes privadas que tenham acesso à Internet através da conversão de endereços de rede (NAT). Para obter mais informações sobre como configurar uma sub-rede privada para o Amazon EMR, consulte Sub-redes privadas.
- Ao usar o Amazon EMR no EKS com o EMR Studio, deve haver, no mínimo, uma sub-rede em comum entre o Studio e o cluster do Amazon EKS usado para registrar um cluster virtual. Caso contrário, o endpoint gerenciado não aparecerá como uma opção nos Workspaces do Studio. Você pode criar um cluster do Amazon EKS e associá-lo a uma sub-rede que pertence ao Studio ou criar um Studio e especificar as sub-redes do seu cluster do EKS.

- Se você planeja usar o Amazon EMR no EKS com o EMR Studio, escolha a mesma VPC dos nós de processamento do cluster do Amazon EKS.

Requisitos de cluster para o Amazon EMR Studio

Clusters do Amazon EMR em execução no Amazon EC2

Todos os clusters do Amazon EMR em execução no Amazon EC2 criados para um Workspace do EMR Studio devem atender aos requisitos apresentados a seguir. Os clusters criados usando a interface do EMR Studio atendem automaticamente a esses requisitos.

- O cluster deve usar as versões 5.32.0 (Amazon EMR de série 5.x) ou 6.2.0 (Amazon EMR de série 6.x) ou posteriores do Amazon EMR. Você pode criar um cluster usando o console do Amazon EMR, ou SDK AWS Command Line Interface, e depois anexá-lo a um espaço de trabalho do EMR Studio. Os usuários do Studio também podem provisionar e anexar clusters ao criar ou trabalhar em um Workspace do Amazon EMR. Para ter mais informações, consulte Anexar uma computação a um Workspace do EMR Studio.
- O cluster deve estar em uma Amazon Virtual Private Cloud. A plataforma EC2-Classic não é compatível.
- O cluster deve ter o Spark, o Livy e o Jupyter Enterprise Gateway instalados. Se você planeja usar o cluster para o SQL Explorer, deverá instalar o Presto e o Spark.
- Para usar o SQL Explorer, o cluster deve usar a versão 5.34.0, ou versões posteriores, ou a versão 6.4.0, ou versões posteriores, do Amazon EMR e ter o Presto instalado. Se você quiser especificar o AWS Glue Data Catalog como o metastore do Hive para o Presto, você deve configurá-lo no cluster. Para obter mais informações, consulte Using Presto with the AWS Glue Data Catalog.
- O cluster deve estar em uma sub-rede privada com conversão de endereços de rede (NAT) para usar repositórios Git hospedados publicamente com o EMR Studio.

Recomendamos as configurações de cluster apresentadas a seguir ao trabalhar com o EMR Studio.

- Defina o modo de implantação das sessões do Spark para o modo de cluster. O modo de cluster coloca os processos principais de aplicações nos nós centrais e não no nó primário de um cluster. Isso alivia o nó primário de possíveis pressões de memória. Para obter mais informações, consulte o tópico de Visão geral do modo de cluster na documentação do Apache Spark.

- Altere o tempo limite do Livy do padrão de uma hora para seis horas, como no exemplo de configuração apresentado a seguir.

```
{
    "classification":"livy-conf",
        "Properties":{
            "livy.server.session.timeout":"6h",
            "livy.spark.deploy-mode":"cluster"
        }
}
```

- Crie diversas frotas de instâncias com até 30 instâncias e selecione vários tipos de instâncias em sua frota de instâncias spot. Por exemplo, é possível especificar os seguintes tipos de instâncias otimizadas para memória para workloads do Spark: r5.2x, r5.4x, r5.8x, r5.12x, r5.16x, r4.2x, r4.4x, r4.8x, r4.12 etc. Para ter mais informações, consulte Configurar frotas de instâncias.
- Use a estratégia de alocação com capacidade otimizada para instâncias spot com a finalidade de ajudar o Amazon EMR a fazer seleções de instâncias eficazes com base em insights de capacidade em tempo real do Amazon EC2. Para ter mais informações, consulte Estratégia de alocação para frotas de instâncias.
- Habilite o ajuste de escala gerenciado em seu cluster. Defina o parâmetro máximo de nós centrais para a capacidade persistente mínima que você planeja usar, e configure a escalabilidade em uma frota de tarefas bem diversificada que é executada em instâncias spot para economizar custos. Para ter mais informações, consulte Usar o ajuste de escala gerenciado no Amazon EMR.

Também recomendamos manter o bloqueio de acesso público do Amazon EMR habilitado e restringir o tráfego SSH de entrada para origens confiáveis. O acesso de entrada a um cluster permite que os usuários executem cadernos no cluster. Para obter mais informações, consulte Usar o bloqueio de acesso público do Amazon EMR e Controle do tráfego de rede com grupos de segurança.

Clusters do Amazon EMR no EKS

Além dos clusters do EMR em execução no Amazon EC2, você pode configurar e gerenciar clusters do Amazon EMR no EKS para o EMR Studio usando a AWS CLI. Configure os clusters do Amazon EMR no EKS usando as seguintes diretrizes:

- Crie um endpoint HTTPS gerenciado para o cluster do Amazon EMR no EKS. Os usuários anexam um Workspace a um endpoint gerenciado. O cluster do Amazon Elastic Kubernetes Service (EKS)

usado para registrar um cluster virtual deve ter uma sub-rede privada para oferecer suporte a endpoints gerenciados.

- Use um cluster do Amazon EKS com, no mínimo, uma sub-rede privada e com conversão de endereços de rede (NAT) quando desejar usar repositórios Git hospedados publicamente.
- Evite usar ARM de AMIs do Amazon Linux otimizadas para o Amazon EKS, que não são compatíveis com os endpoints gerenciados pelo Amazon EMR no EKS.
- Evite usar clusters AWS Fargate somente do Amazon EKS, que não são compatíveis.

Configuração do Amazon EMR Studio

Esta seção é destinada aos administradores do EMR Studio. Ela aborda como configurar um EMR Studio para sua equipe e fornece instruções para tarefas, como a atribuição de usuários e grupos, a configuração de modelos de cluster e a otimização do Apache Spark para o EMR Studio.

Tópicos

Permissões de administrador para criar e gerenciar um EMR Studio

As permissões do IAM descritas nesta página permitem criar e gerenciar um EMR Studio. Para obter informações detalhadas sobre cada permissão obrigatória, consulte Permissões obrigatórias para gerenciar um EMR Studio.

Permissões obrigatórias para gerenciar um EMR Studio

A tabela a seguir lista as operações relacionadas à criação e ao gerenciamento de um EMR Studio. A tabela também exibe as permissões necessárias para cada operação.

> **Note**
>
> Você precisa somente de ações `SessionMapping` do Centro de Identidade do IAM e do Studio ao usar o modo de autenticação do Centro de Identidade do IAM.

Permissões para criar e gerenciar um EMR Studio

Operation	Permissões
Criar um Studio	`"elasticmapreduce:CreateStudio", "sso:CreateApplication", "sso:PutApplicationAuthentic ationMethod", "sso:PutApplicationGrant", "sso:PutApplicationAccessScope", "sso:PutApplicationAssignmentConfi guration", "iam:PassRole"`
Descrever um Studio	`"elasticmapreduce:DescribeStudio", "sso:GetManagedApplicationInstance"`
Listar Studios	`"elasticmapreduce:ListStudios"`
Excluir um Studio	`"elasticmapreduce:DeleteStudio", "sso:DeleteApplication", "sso:DeleteApplicationAuthentica tionMethod", "sso:DeleteApplicationAccessScope", "sso:DeleteApplicationGrant"`
Additional permissions required when you use IAM Identity Center mode	
Atribuir usuários ou grupos a um Studio	`"elasticmapreduce:CreateStudioSessio nMapping", "sso:GetProfile", "sso:ListDirectoryAssociations", "sso:ListProfiles",`

Operation	Permissões
	"sso:AssociateProfile", "sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:ListInstances", "sso:CreateApplicationAssignment", "sso:DescribeInstance", "organizations:DescribeOrga nization", "organizations:ListDelegatedAdmini strators", "sso:CreateInstance", "sso:DescribeRegisteredRegions", "sso:GetSharedSsoConfiguration", "iam:ListPolicies"
Recuperar detalhes de atribuição do Studio para um usuário ou um grupo específico	"sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:DescribeApplication", "elasticmapreduce:GetStudioSessio nMapping"
Listar todos os usuários e os grupos atribuídos a um Studio	"elasticmapreduce:ListStudioSessionM appings"
Atualizar a política de sessão anexada a um usuário ou a um grupo atribuído a um Studio	"sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:DescribeApplication", "sso:DescribeInstance", "elasticmapreduce:UpdateStu dioSessionMapping"

Operation	Permissões
Remover um usuário ou um grupo de um Studio	`"elasticmapreduce:DeleteStudioSessionMapping", "sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:ListDirectoryAssociations", "sso:GetProfile", "sso:DescribeApplication", "sso:DescribeInstance", "sso:ListProfiles", "sso:DisassociateProfile", "sso:DeleteApplicationAssignment", "sso:ListApplicationAssignments"`

Criar uma política com permissões de administrador para o EMR Studio

1. Siga as instruções em Criação de políticas do IAM para criar uma política usando um dos exemplos apresentados a seguir. As permissões necessárias dependem do seu modo de autenticação para o EMR Studio.

   Insira seus próprios valores para estes itens:

   - Substitua *<your-resource-ARN>* para especificar o nome do recurso da Amazon (ARN) do objeto ou dos objetos que a instrução abrange para seus casos de uso.
   - Substitua *<region>* pelo código da Região da AWS em que você planeja criar o Studio.
   - Substitua *<aws-account_id>* pelo ID da conta AWS para o Studio.
   - Substitua *<EMRStudio-Service-Role>* e *<EMRStudio-User-Role>* pelos nomes do perfil de serviço do EMR Studio e do perfil de usuário do EMR Studio.

   Example Política de exemplo: permissões de administrador ao usar o modo de autenticação do IAM

   ```
   {
       "Version": "2012-10-17",
       "Statement": [
   ```

```
        {
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:<region>:<aws-account-id>:studio/
*",
            "Action": [
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": "<your-resource-ARN>",
            "Action": [
                "elasticmapreduce:ListStudios"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": [
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-Service-Role>"
            ],
            "Action": "iam:PassRole"
        }
    ]
}
```

Example Política de exemplo: permissões de administrador ao usar o modo de autenticação do Centro de Identidade do IAM

> **Note**
>
> As APIs do Centro de Identidade e do diretório do Centro de Identidade não oferecem suporte à especificação de um ARN no elemento de recurso de uma instrução de política do IAM. Para permitir o acesso ao Centro de Identidade do IAM e ao diretório do Centro de Identidade do IAM, as permissões apresentadas a seguir especificam todos os recursos, “Resource”:“*”, para as ações do Centro de Identidade do IAM. Para obter mais informações, consulte Actions, resources, and condition keys for IAM Identity Center Directory.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:<region>:<aws-account-id>:studio/
*",
            "Action": [
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio",
                "elasticmapreduce:CreateStudioSessionMapping",
                "elasticmapreduce:GetStudioSessionMapping",
                "elasticmapreduce:UpdateStudioSessionMapping",
                "elasticmapreduce:DeleteStudioSessionMapping"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": "<your-resource-ARN>",
            "Action": [
                "elasticmapreduce:ListStudios",
                "elasticmapreduce:ListStudioSessionMappings"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": [
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-Service-Role>",
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-User-Role>"
            ],
            "Action": "iam:PassRole"
        },
        {
            "Effect": "Allow",
            "Resource": "*",
            "Action": [
                "sso:CreateApplication",
                "sso:PutApplicationAuthenticationMethod",
                "sso:PutApplicationGrant",
                "sso:PutApplicationAccessScope",
                "sso:PutApplicationAssignmentConfiguration",
```

```
                "sso:DescribeApplication",
                "sso:DeleteApplication",
                "sso:DeleteApplicationAuthenticationMethod",
                "sso:DeleteApplicationAccessScope",
                "sso:DeleteApplicationGrant",
                "sso:ListInstances",
                "sso:CreateApplicationAssignment",
                "sso:DeleteApplicationAssignment",
                "sso:ListApplicationAssignments",
                "sso:DescribeInstance",
                "sso:AssociateProfile",
                "sso:DisassociateProfile",
                "sso:GetProfile",
                "sso:ListDirectoryAssociations",
                "sso:ListProfiles",
                "sso-directory:SearchUsers",
                "sso-directory:SearchGroups",
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup",
                "organizations:DescribeOrganization",
                "organizations:ListDelegatedAdministrators",
                "sso:CreateInstance",
                "sso:DescribeRegisteredRegions",
                "sso:GetSharedSsoConfiguration",
                "iam:ListPolicies"
            ]
        }
    ]
}
```

2. Anexe a política à sua identidade do IAM (usuário, perfil ou grupo). Para obter instruções, consulte Adicionar e remover permissões de identidade do IAM.

Configuração de um Amazon EMR Studio

Conclua as etapas apresentadas a seguir para configurar um Amazon EMR Studio.

Antes de começar

> **Note**
>
> Se você planeja usar o EMR Studio com o Amazon EMR no EKS, recomendamos configurar o Amazon EMR no EKS para o EMR Studio antes de configurar um Studio.

Antes de configurar um EMR Studio, certifique-se de ter os seguintes itens:

- Uma Conta da AWS. Para obter instruções, consulte Configuração do Amazon EMR.
- As permissões para criar e gerenciar um EMR Studio. Para ter mais informações, consulte the section called “Permissões de administrador para criar um EMR Studio”.
- Um bucket do Amazon S3 no qual o EMR Studio pode fazer backup dos Workspaces e dos arquivos de cadernos do seu Studio. Para obter instruções, consulte Criação de um bucket no Guia do usuário do Amazon Simple Storage Service (S3).
- Para se conectar a um cluster do Amazon EMR no EC2 ou do Amazon EMR no EKS, ou usar repositórios Git, é necessário uma nuvem privada virtual (VPC) da Amazon para o Studio e, no máximo, cinco sub-redes. Você não precisa de uma VPC para usar o EMR Studio com o EMR Serverless. Para obter dicas sobre como configurar as redes, consulte Práticas recomendadas para VPC e para sub-rede.

Configurar um EMR Studio

1. Escolha um modo de autenticação para o Amazon EMR Studio
2. Crie os recursos do Studio a seguir.
   - Crie um perfil de serviço do EMR Studio
   - Configurar permissões de usuário do EMR Studio para Amazon EC2 ou Amazon EKS
   - (Opcional) Definição de grupos de segurança para controlar o tráfego de rede do EMR Studio.
3. Crie um EMR Studio
4. Atribuir um usuário ou um grupo a um EMR Studio

Após concluir as etapas de configuração, você poderá Uso de um Amazon EMR Studio.

Escolha um modo de autenticação para o Amazon EMR Studio

O EMR Studio oferece suporte a dois modos de autenticação: o modo de autenticação do IAM e o modo de autenticação do Centro de Identidade do IAM. O modo do IAM usa o AWS Identity and Access Management (IAM), enquanto o modo do Centro de Identidade do IAM usa o AWS IAM Identity Center. Ao criar um EMR Studio, você escolhe o modo de autenticação para todos os usuários desse Studio. Para obter mais informações sobre os diferentes modos de autenticação, consulte Autenticação e login do usuário.

Use a tabela apresentada a seguir para escolher um modo de autenticação para o EMR Studio.

Se você...	Recomendamos…
Já está familiarizado ou configurou anteriormente uma autenticação ou uma federação do IAM	O Modo de autenticação do IAM, que oferece os seguintes benefícios: • Disponibilização de uma configuração rápida para o EMR Studio, se você já gerencia identidades como usuários e grupos no IAM. • Funcionamento com provedores de identidade que são compatíveis com o OpenID Connect (OIDC) ou com a Security Assertion Markup Language 2.0 (SAML 2.0). • Oferecimento de suporte ao uso de diversos provedores de identidade com a mesma Conta da AWS. • Disponibilidade em um grande número de Regiões da AWS. • Compatibilidade com SOC 2.
É iniciante na AWS ou no Amazon EMR	O Modo de autenticação do Centro de Identidade do IAM, que fornece os seguintes recursos: • Oferece suporte à atribuição fácil de usuários e grupos aos recursos da AWS.

Se você...	Recomendamos…
	• Funcionamento com provedores de identidad e do Microsoft Active Directory e da SAML 2.0. • Facilitação da configuração da federação de várias contas para que você não tenha necessidade de configurar separadamente a federação para cada Conta da AWS na sua organização.

Configuração do modo de autenticação do IAM para o Amazon EMR Studio

Com o modo de autenticação do IAM, você pode usar a autenticação do IAM ou a federação do IAM. A autenticação do IAM permite gerenciar identidades do IAM, como usuários, grupos e perfis no IAM. Você concede aos usuários acesso a um Studio com as políticas de permissões do IAM e o controle de acesso por atributos (ABAC). A federação do IAM permite estabelecer confiança entre um provedor de identidades (IdP) terceirizado e a AWS para que você possa gerenciar identidades de usuários por meio do seu IdP.

> **Note**
>
> Se você já usa o IAM para controlar o acesso aos recursos da AWS ou se já configurou seu provedor de identidades (IdP) para o IAM, consulte Permissões de usuários para o modo de autenticação do IAM para definir permissões de usuário ao usar o modo de autenticação do IAM para o EMR Studio.

Uso da federação do IAM para o Amazon EMR Studio

Para usar a federação do IAM para o EMR Studio, você cria uma relação de confiança entre sua Conta da AWS e seu provedor de identidades (IdP) e possibilita que usuários federados acessem o AWS Management Console. As etapas executadas para criar essa relação de confiança variam com base no padrão de federação do seu IdP.

Em geral, você conclui as tarefas a seguir para configurar a federação com um IdP externo. Para obter instruções completas, consulte Habilitar o acesso de usuários federados SAML 2.0 ao AWS

Management Console e Habilitar o acesso do agente de identidades personalizado ao AWS Management Console no Guia do usuário do AWS Identity and Access Management.

1. Reúna informações do seu IdP. Geralmente, isso significa gerar um documento de metadados para validar as solicitações de autenticação SAML do seu IdP.
2. Crie uma entidade do IAM do provedor de identidades para armazenar as informações sobre seu IdP. Para obter instruções, consulte Criação de provedores de identidades do IAM.
3. Crie um ou mais perfis do IAM para seu IdP. O EMR Studio atribui um perfil a um usuário federado quando o usuário realiza o login. O perfil permite que seu IdP solicite credenciais de segurança temporárias para obter acesso à AWS. Para obter instruções, consulte Criar um perfil para um provedor de identidades de terceiros (federação). As políticas de permissões atribuídas ao perfil determinam o que os usuários federados podem fazer na AWS e em um EMR Studio. Para ter mais informações, consulte Permissões de usuários para o modo de autenticação do IAM.
4. (Para provedores SAML) Conclua o estabelecimento de confiança da SAML ao configurar seu IdP com informações sobre a AWS e sobre os perfis que você deseja que os usuários federados assumam. Esse processo de configuração cria confiança de terceira parte confiável entre seu IdP e a AWS. Para obter mais informações, consulte Configurar o IdP do SAML 2.0 com confiança da parte dependente e incluir declarações.

Para configurar um EMR Studio como uma aplicação da SAML em seu portal do IdP

Você pode configurar um EMR Studio específico como uma aplicação da SAML usando um link direto para o Studio. Isso permite que os usuários façam login no seu portal do IdP e iniciem um Studio específico em vez de navegar pelo console do Amazon EMR.

- Use o formato apresentado a seguir para configurar um link direto para seu EMR Studio como um URL de destino após a verificação de declaração da SAML.

```
https://console.aws.amazon.com/emr/home?region=<aws-region>#studio/<your-studio-id>/start
```

Configuração do modo de autenticação do Centro de Identidade do IAM para o Amazon EMR Studio

Para preparar o AWS IAM Identity Center para o EMR Studio, você deve configurar sua origem de identidade e provisionar usuários e grupos. O provisionamento é o processo de disponibilização de informações de usuários e de grupos para o uso pelo Centro de Identidade do IAM e por aplicações

que usam o Centro de Identidade do IAM. Para obter mais informações, consulte User and group provisioning.

O EMR Studio oferece suporte ao uso dos seguintes provedores de identidades para o Centro de Identidade do IAM:

- AWS Managed Microsoft AD e Active Directory autogerenciado: para obter mais informações, consulte Connect to your Microsoft AD directory.
- Provedores baseados em SAML: para obter uma lista completa, consulte Supported identity providers.
- Diretório do Centro de Identidade do IAM: para obter mais informações, consulte Manage identities in IAM Identity Center.

Configurar o Centro de Identidade do IAM para o EMR Studio

1. Para configurar o Centro de Identidade do IAM para o EMR Studio, você precisa do seguinte:
   - Uma conta de gerenciamento na sua organização da AWS, se você usar várias contas na organização.

     > **Note**
     >
     > Você deve usar a conta de gerenciamento somente para habilitar o Centro de Identidade do IAM e provisionar usuários e grupos. Após configurar o Centro de Identidade do IAM, use uma conta de membro para criar um EMR Studio e atribuir usuários e grupos. Para saber mais sobre a terminologia da AWS, consulte Terminologia e conceitos do AWS Organizations.

   - Se você habilitou o Centro de Identidade do IAM antes de 25 de novembro de 2019, talvez seja necessário habilitar aplicações que usam o Centro de Identidade do IAM para as contas em sua organização da AWS. Para obter mais informações, consulte Enable IAM Identity Center-integrated applications in AWS accounts.
   - Certifique-se de ter os pré-requisitos listados na página de pré-requisitos do Centro de Identidade do IAM.
2. Siga as instruções em Enable IAM Identity Center para habilitar o Centro de Identidade do IAM na Região da AWS em que você deseja criar o EMR Studio.

3. Conecte o Centro de Identidade do IAM ao seu provedor de identidades e provisione os usuários e os grupos que você deseja atribuir ao Studio.

Se você usar...	Fazer isso...
Um diretório do Microsoft AD	1. Siga as instruções em Connect to your Microsoft AD directory para conectar o Active Directory autogerenciado ou o diretório do AWS Managed Microsoft AD usando o AWS Directory Service. 2. Para provisionar usuários e grupos para o Centro de Identidade do IAM, você pode sincronizar dados de identidade do seu AD de origem para o Centro de Identidade do IAM. É possível sincroniz ar as identidades do seu AD de origem de várias maneiras. Uma das maneiras é atribuir usuários ou grupos do AD a uma conta da AWS na sua organização. Para obter instruções, consulte Single sign-on. A sincronização pode demorar até duas horas. Depois de concluir esta etapa, os usuários e os grupos sincronizados aparecerão no seu repositório de identidad es. **Note** Usuários e grupos não aparecem no Identity Store até que você sincronize as informações do usuário e do grupo ou use o provisionamento de usuários just-in-time (JIT). Para obter mais informações, consulte Provision ing when users come from Active Directory.

Se você usar...	Fazer isso...
	3. (Opcional) Após sincronizar usuários e grupos do AD, é possível remover o acesso deles à sua conta da AWS configurada na etapa anterior. Para obter instruções, consulte Remove user access.
Um provedor de identidades externo	Siga as instruções em Connect to your external identity provider.
O diretório do Centro de Identidade do IAM	Ao criar usuários e grupos no Centro de Identidade do IAM, o provisionamento é automático. Para obter mais informações, consulte Manage identities in IAM Identity Center.

Agora você pode atribuir usuários e grupos do seu repositório de identidades a um EMR Studio. Para obter instruções, consulte Atribuir um usuário ou um grupo a um EMR Studio.

Crie um perfil de serviço do EMR Studio

Sobre o perfil de serviço do EMR Studio

Cada EMR Studio usa um perfil do IAM com permissões que possibilitam ao Studio interagir com outros serviços da AWS. Esse perfil de serviço deve incluir permissões que possibilitam ao EMR Studio estabelecer um canal de rede seguro entre os Workspaces e os clusters, armazenar arquivos de cadernos no Amazon S3 Control e acessar o AWS Secrets Manager ao vincular um Workspace a um repositório Git.

Use o perfil de serviço do Studio (em vez das políticas de sessão) para definir todas as permissões de acesso do Amazon S3 para armazenar arquivos de cadernos e para definir as permissões de acesso do AWS Secrets Manager.

Como criar um perfil de serviço para o EMR Studio no Amazon EC2 ou no Amazon EKS

1. Siga as instruções em Criação de um perfil para delegar permissões a um serviço da AWS para criar o perfil de serviço usando a política de confiança apresentada a seguir.

> **Important**
>
> A política de confiança a seguir inclui as chaves de condição globais `aws:SourceArn` e `aws:SourceAccount` para limitar as permissões que você concede ao EMR Studio para recursos específicos em sua conta. Fazer isso pode proteger você contra o problema de "confused deputy".

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "elasticmapreduce.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "<account-id>"
        },
        "ArnLike": {
          "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
        }
      }
    }
  ]
}
```

2. Remova as permissões de perfil padrão. Em seguida, inclua as permissões do exemplo de política de permissões do IAM a seguir. Como alternativa, você pode criar uma política personalizada que use as Permissões de perfil de serviço do EMR Studio.

> **Important**
>
> - Para que o controle de acesso baseado em tags do Amazon EC2 funcione com o EMR Studio, você deve definir o acesso à API `ModifyNetworkInterfaceAttribute` conforme mostrado na política a seguir.

- Para que o EMR Studio funcione com o perfil de serviço, você não deve alterar as seguintes instruções: `AllowAddingEMRTagsDuringDefaultSecurityGroupCreation` e `AllowAddingTagsDuringEC2ENICreation`.
- Para usar a política de exemplo, você deve etiquetar os recursos apresentados a seguir com a chave "**`for-use-with-amazon-emr-managed-policies`**" e o valor "**`true`**".
  - Sua Amazon Virtual Private Cloud (VPC) para o EMR Studio.
  - Cada sub-rede que deseja usar com o Studio.
  - Qualquer grupo de segurança personalizado do EMR Studio. Você deve etiquetar todos os grupos de segurança criados durante o período de pré-visualização do EMR Studio se desejar continuar a usá-los.
  - Os segredos mantidos no AWS Secrets Manager que os usuários do Studio usam para vincular repositórios Git a um Workspace.

  Você pode aplicar etiquetas aos recursos usando a guia Etiquetas na tela de recursos relevantes no AWS Management Console.

Quando aplicável, altere * em "`Resource`":"*" na política apresentada a seguir para especificar o nome do recurso da Amazon (ARN) dos recursos abrangidos pela instrução para o seu caso de uso.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowEMRReadOnlyActions",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:ListInstances",
        "elasticmapreduce:DescribeCluster",
        "elasticmapreduce:ListSteps"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowEC2ENIActionsWithEMRTags",
```

```
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterfacePermission",
        "ec2:DeleteNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:network-interface/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowEC2ENIAttributeAction",
      "Effect": "Allow",
      "Action": [
        "ec2:ModifyNetworkInterfaceAttribute"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:instance/*",
        "arn:aws:ec2:*:*:network-interface/*",
        "arn:aws:ec2:*:*:security-group/*"
      ]
    },
    {
      "Sid": "AllowEC2SecurityGroupActionsWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:AuthorizeSecurityGroupEgress",
        "ec2:AuthorizeSecurityGroupIngress",
        "ec2:RevokeSecurityGroupEgress",
        "ec2:RevokeSecurityGroupIngress",
        "ec2:DeleteNetworkInterfacePermission"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
```

```
      "Sid": "AllowDefaultEC2SecurityGroupsCreationWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateSecurityGroup"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:security-group/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowDefaultEC2SecurityGroupsCreationInVPCWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateSecurityGroup"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:vpc/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowAddingEMRTagsDuringDefaultSecurityGroupCreation",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:*:*:security-group/*",
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true",
          "ec2:CreateAction": "CreateSecurityGroup"
        }
      }
    },
    {
```

```
      "Sid": "AllowEC2ENICreationWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:network-interface/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowEC2ENICreationInSubnetAndSecurityGroupWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:subnet/*",
        "arn:aws:ec2:*:*:security-group/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowAddingTagsDuringEC2ENICreation",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:*:*:network-interface/*",
      "Condition": {
        "StringEquals": {
          "ec2:CreateAction": "CreateNetworkInterface"
        }
      }
    },
    {
```

```
      "Sid": "AllowEC2ReadOnlyActions",
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeSecurityGroups",
        "ec2:DescribeNetworkInterfaces",
        "ec2:DescribeTags",
        "ec2:DescribeInstances",
        "ec2:DescribeSubnets",
        "ec2:DescribeVpcs"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowSecretsManagerReadOnlyActionsWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "secretsmanager:GetSecretValue"
      ],
      "Resource": "arn:aws:secretsmanager:*:*:secret:*",
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowWorkspaceCollaboration",
      "Effect": "Allow",
      "Action": [
        "iam:GetUser",
        "iam:GetRole",
        "iam:ListUsers",
        "iam:ListRoles",
        "sso:GetManagedApplicationInstance",
        "sso-directory:SearchUsers"
      ],
      "Resource": "*"
    }
  ]
}
```

3. Conceda ao seu perfil de serviço o acesso de leitura e de gravação ao local do Amazon S3 para o EMR Studio. Use o conjunto mínimo de permissões apresentado a seguir. Para obter mais

informações, consulte o exemplo Amazon S3: permite acesso de leitura e gravação a objetos em um bucket do S3 de forma programática e no console.

```
"s3:PutObject",
"s3:GetObject",
"s3:GetEncryptionConfiguration",
"s3:ListBucket",
"s3:DeleteObject"
```

Se você criptografar seu bucket do Amazon S3, inclua as permissões a seguir para o AWS Key Management Service.

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
"kms:DescribeKey"
```

4. Se você quiser controlar o acesso aos segredos do Git no nível do usuário, adicione permissões baseadas em tags a `secretsmanager:GetSecretValue` na política de perfil de usuário do EMR Studio e remova as permissões da política `secretsmanager:GetSecretValue` da política de perfil de serviço do EMR Studio. Para obter mais informações sobre como configurar as permissões refinadas de usuário, consulte Criação de políticas de permissões para usuários do EMR Studio.

Perfil de serviço mínimo para o EMR Serverless

Se quiser executar workloads interativas com o EMR Serverless por meio de cadernos do EMR Studio, use a mesma política de confiança usada para configurar o EMR Studio na seção anterior, Como criar um perfil de serviço para o EMR Studio no Amazon EC2 ou no Amazon EKS.

Para sua política do IAM, a política mínima viável tem as permissões a seguir. Atualize *bucket-name* com o nome do bucket que planeja usar ao configurar o EMR Studio e o Workspace. O EMR Studio usa o bucket para fazer backup dos Workspaces e dos arquivos de caderno no seu Studio.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ObjectActions",
```

```
            "Effect": "Allow",
            "Action": [
               "s3:PutObject",
               "s3:GetObject",
               "s3:DeleteObject"
            ],
            "Resource": ["arn:aws:s3:::bucket-name/*"]
        },
        {
            "Sid": "BucketActions",
            "Effect": "Allow",
            "Action": [
               "s3:ListBucket",
               "s3:GetEncryptionConfiguration"
            ],
            "Resource": ["arn:aws:s3:::bucket-name"]
        }
    ]
}
```

Se usar um bucket criptografado do Amazon S3, inclua as seguintes permissões na sua política:

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
"kms:DescribeKey"
```

Permissões de perfil de serviço do EMR Studio

A tabela a seguir lista as operações que o EMR Studio executa usando o perfil de serviço, em conjunto com as ações do IAM obrigatórias para cada operação.

Operation	Ações

Estabelecimento de um canal de rede seguro entre um Workspace e um cluster do EMR e execução das ações de limpeza necessárias.	`"ec2:CreateNetworkInterface",`
	`"ec2:CreateNetworkInterfacePermission",`
	`"ec2:DeleteNetworkInterface",`
	`"ec2:DeleteNetworkInterfacePermission",`
	`"ec2:DescribeNetworkInterfaces",`
	`"ec2:ModifyNetworkInterfaceAttribute",`
	`"ec2:AuthorizeSecurityGroupEgress",`

Operation	Ações

	`"ec2:AuthorizeSecurityGroupIngress",`
	`"ec2:CreateSecurityGroup",`
	`"ec2:DescribeSecurityGroups",`
	`"ec2:RevokeSecurityGroupEgress",`
	`"ec2:DescribeTags",`
	`"ec2:DescribeInstances",`
	`"ec2:DescribeSubnets",`
	`"ec2:DescribeVpcs",`
	`"elasticmapreduce:ListInstances",`
	`"elasticmapreduce:DescribeCluster",`
	`"elasticmapreduce:ListSteps"`

Uso das credenciais do Git armazenadas no AWS Secrets Manager para vincular repositórios Git a um Workspace.	`"secretsmanager:GetSecretValue"`
Aplicação de etiquetas da AWS à interface de rede e aos grupos de segurança padrão que o EMR Studio cria ao configurar o canal de rede seguro. Para obter mais informaçõ es, consulte Etiquetar recursos da AWS.	`"ec2:CreateTags"`

Operation	Ações

Acesso ou upload de arquivos e metadados de cadernos para o Amazon S3.	`"s3:PutObject",`
	`"s3:GetObject",`
	`"s3:GetEncryptionConfiguration",`
	`"s3:ListBucket",`
	`"s3:DeleteObject"`
	
	Se você usar um bucket criptografado do Amazon S3, inclua as permissões a seguir.
	
	`"kms:Decrypt",`
	`"kms:GenerateDataKey",`
	`"kms:ReEncryptFrom",`
	`"kms:ReEncryptTo",`
	`"kms:DescribeKey"`

Habilitação e configuração da colaboração no Workspace.	`"iam:GetUser",`
	`"iam:GetRole",`
	`"iam:ListUsers",`
	`"iam:ListRoles",`
	`"sso:GetManagedApplicationInstance",`
	`"sso-directory:SearchUsers"`

Criptografe cadernos e arquivos do espaço de trabalho do EMR Studio usando chaves gerenciadas pelo cliente (CMK) com AWS Key Management Service	`"kms:Decrypt",`
	`"kms:GenerateDataKey",`
	`"kms:ReEncryptFrom",`
	`"kms:ReEncryptTo",`
	`"kms:DescribeKey"`

Configurar permissões de usuário do EMR Studio para Amazon EC2 ou Amazon EKS

Você deve configurar políticas de permissões de usuário para o Amazon EMR Studio para definir permissões detalhadas de usuários e de grupos. Para obter informações sobre como as permissões de usuário funcionam no EMR Studio, consulte Controle de acesso em Como o Amazon EMR Studio funciona.

> **Note**
>
> As permissões abordadas nesta seção não impõem controle de acesso a dados. Para gerenciar o acesso aos conjuntos de dados de entrada, você deve configurar permissões para os clusters que seu Studio usa. Para ter mais informações, consulte Segurança no Amazon EMR.

Criação de um perfil de usuário do EMR Studio para o modo de autenticação do Centro de Identidade do IAM

Você deve criar um perfil de usuário do EMR Studio ao usar o modo de autenticação do Centro de Identidade do IAM.

Criar um perfil de usuário para o EMR Studio

1. Siga as instruções em Criar um perfil para delegar permissões a um serviço da AWS no Guia do usuário do AWS Identity and Access Management para criar um perfil de usuário.

   Ao criar o perfil, use a política de relação de confiança apresentada a seguir.

   ```
   {
     "Version": "2008-10-17",
     "Statement": [
       {
         "Effect": "Allow",
         "Principal": {
           "Service": "elasticmapreduce.amazonaws.com"
         },
         "Action": [
                 "sts:AssumeRole",
                 "sts:SetContext"
                 ]
       }
     ]
   }
   ```

2. Remova as permissões e as políticas de perfil padrão.
3. Antes de atribuir usuários e grupos a um Studio, anexe as políticas de sessão do EMR Studio ao perfil de usuário. Para obter instruções sobre como criar políticas de sessão, consulte Criação de políticas de permissões para usuários do EMR Studio.

Criação de políticas de permissões para usuários do EMR Studio

Consulte as seções a seguir para criar políticas de permissões do EMR Studio.

Tópicos

> **Note**
>
> Para estabelecer permissões de acesso do Amazon S3 para o armazenamento de arquivos de cadernos e permissões de acesso do AWS Secrets Manager para a leitura de segredos ao vincular Workspaces a repositórios Git, use o perfil de serviço do EMR Studio.

Criação das políticas de permissões

Crie uma ou mais políticas de permissões do IAM que especifiquem quais ações um usuário pode realizar no seu Studio. Por exemplo, é possível criar três políticas separadas para tipos de usuários básicos, intermediários e avançados do Studio com os exemplos de políticas nesta página.

Para obter um detalhamento de cada operação do Studio que um usuário pode executar e as ações mínimas do IAM necessárias para executar cada operação, consulte Permissões do AWS Identity and Access Management para usuários do EMR Studio. Para ver as etapas de criação das políticas, consulte Criação de políticas do IAM no Guia do usuário do IAM.

Sua política de permissões deve incluir as instruções apresentadas a seguir.

```
{
            "Sid": "AllowAddingTagsOnSecretsWithEMRStudioPrefix",
            "Effect": "Allow",
            "Action": "secretsmanager:TagResource",
            "Resource": "arn:aws:secretsmanager:*:*:secret:emr-studio-*"
},
{
            "Sid": "AllowPassingServiceRoleForWorkspaceCreation",
            "Action": "iam:PassRole",
```

```
            "Resource": [
                "arn:aws:iam::*:role/your-emr-studio-service-role"
            ],
            "Effect": "Allow"
}
```

Definição de propriedade para colaboração no Workspace

A colaboração no Workspace permite que vários usuários trabalhem simultaneamente no mesmo Workspace e pode ser configurada com o painel Colaboração na interface do usuário do Workspace. Para visualizar e usar o painel Colaboração, o usuário deve ter as permissões apresentadas a seguir. Qualquer usuário com essas permissões poderá visualizar e usar o painel Colaboração.

```
"elasticmapreduce:UpdateEditor",
"elasticmapreduce:PutWorkspaceAccess",
"elasticmapreduce:DeleteWorkspaceAccess",
"elasticmapreduce:ListWorkspaceAccessIdentities"
```

Para restringir o acesso ao painel Colaboração, é possível usar o controle de acesso por etiquetas. Quando um usuário cria um Workspace, o EMR Studio aplica uma etiqueta padrão com uma chave `creatorUserId` cujo valor é o ID do usuário que cria o Workspace.

> **Note**
>
> O EMR Studio adiciona a tag `creatorUserId` aos Workspaces criados após 16 de novembro de 2021. Para restringir quem pode configurar a colaboração dos espaços de trabalhos criados antes dessa data, recomendamos adicionar manualmente a tag `creatorUserId` ao seu Workspace e, em seguida, usar o controle de acesso por tags nas suas políticas de permissões de usuários.

A instrução de exemplo a seguir permite que um usuário configure a colaboração para qualquer Workspace com a chave de etiqueta `creatorUserId` cujo valor corresponde ao ID do usuário (indicado pela variável de política `aws:userId`). Em outras palavras, a instrução permite que um usuário configure a colaboração para os Workspaces criados por ele. Para saber mais sobre as variáveis de política, consulte Elementos de política do IAM: variáveis e tags no Guia do usuário do IAM.

```
    {
        "Sid": "UserRolePermissionsForCollaboration",
```

```
        "Action": [
            "elasticmapreduce:UpdateEditor",
            "elasticmapreduce:PutWorkspaceAccess",
            "elasticmapreduce:DeleteWorkspaceAccess",
            "elasticmapreduce:ListWorkspaceAccessIdentities"
        ],
        "Resource": "*",
        "Effect": "Allow",
        "Condition": {
            "StringEquals": {
                "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userid}"
            }
        }
    }
```

Criação de uma política de segredos do Git no nível de usuário

Tópicos

- Para usar permissões no nível de usuário
- Para fazer a transição de permissões do nível de serviço para permissões do nível de usuário
- Para usar permissões no nível de serviço

Para usar permissões no nível de usuário

O EMR Studio adiciona automaticamente a tag `for-use-with-amazon-emr-managed-user-policies` ao criar segredos do Git. Se você quiser controlar o acesso aos segredos do Git no nível do usuário, adicione permissões baseadas em tags à política de perfil de usuário do EMR Studio com `secretsmanager:GetSecretValue`, conforme mostrado na seção Para fazer a transição de permissões do nível de serviço para permissões do nível de usuário abaixo.

Se você tiver permissões existentes para `secretsmanager:GetSecretValue` na política de perfil de serviço do EMR Studio, deverá remover essas permissões.

Para fazer a transição de permissões do nível de serviço para permissões do nível de usuário

> **Note**
>
> A tag `for-use-with-amazon-emr-managed-user-policies` garante que as permissões da Etapa 1 abaixo concedam ao criador do espaço de trabalho acesso ao segredo do Git. No entanto, se você vinculou repositórios Git antes de 1.º de setembro

de 2023, os segredos do Git correspondentes terão o acesso negado por não terem a tag `for-use-with-amazon-emr-managed-user-policies` aplicada. Para aplicar permissões em nível de usuário, você deve recriar os segredos antigos JupyterLab e vincular os repositórios Git apropriados novamente.
Para obter mais informações sobre as variáveis de política, consulte Elementos de política do IAM: variáveis e tags no Guia do usuário do IAM.

1. Adicione as permissões a seguir à política de perfil de usuário do EMR Studio. A política usa a chave `for-use-with-amazon-emr-managed-user-policies` com valor `"${aws:userid}"`.

```
{
   "Sid": "AllowSecretsManagerReadOnlyActionsWithEMRTags",
    "Effect": "Allow",
    "Action": "secretsmanager:GetSecretValue",
    "Resource": "arn:aws:secretsmanager:*:*:secret:*",
    "Condition": {
        "StringEquals": {
            "secretsmanager:ResourceTag/for-use-with-amazon-emr-managed-user-
policies": "${aws:userid}"
        }
    }
}
```

2. Se presente, remova a permissão a seguir da política de perfil de serviço do EMR Studio. Como a política de perfil de serviço se aplica a todos os segredos definidos por cada usuário, você só precisa fazer isso uma vez.

```
{
    "Sid": "AllowSecretsManagerReadOnlyActionsWithEMRTags",
    "Effect": "Allow",
    "Action": [
        "secretsmanager:GetSecretValue"
     ],
    "Resource": "arn:aws:secretsmanager:*:*:secret:*",
    "Condition": {
        "StringEquals": {
            "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
    }
```

```
}
```

Para usar permissões no nível de serviço

A partir de 1.º de setembro de 2023, o EMR Studio adiciona automaticamente a tag `for-use-with-amazon-emr-managed-user-policies` para controle de acesso no nível de usuário. Como esse é um recurso adicional, você pode continuar usando o acesso no nível de serviço disponível por meio da permissão `GetSecretValue` no perfil de serviço do EMR Studio.

Para segredos criados antes de 1.º de setembro de 2023, o EMR Studio não adicionou a tag `for-use-with-amazon-emr-managed-user-policies`. Para continuar usando as permissões no nível de serviço, basta reter as permissões existentes de perfil de usuário e de perfil de serviço do EMR Studio. No entanto, para restringir quem pode acessar um segredo individual, recomendamos seguir as etapas em Para usar permissões no nível de usuário para adicionar manualmente a tag `for-use-with-amazon-emr-managed-user-policies` aos seus segredos e, em seguida, usar o controle de acesso por tags nas suas políticas de permissões de usuários.

Para obter mais informações sobre as variáveis de política, consulte Elementos de política do IAM: variáveis e tags no Guia do usuário do IAM.

Anexe a política de permissões à sua identidade do IAM.

A tabela a seguir resume a qual identidade do IAM você deve anexar uma política de permissões, dependendo do modo de autenticação do EMR Studio. Para obter instruções sobre como anexar uma política, consulte Adicionar e remover permissões de identidade do IAM.

Se você usar...	Anexe a política para...
Autenticação do IAM	As suas identidades do IAM (usuários, grupos de usuários ou perfis). Por exemplo, é possível anexar uma política de permissões a um usuário em sua Conta da AWS.
Federação do IAM com um provedor de identidades (IdP) externo	O perfil ou os perfis do IAM que você cria para seu IdP externo. Por exemplo, uma federação do IAM para SAML 2.0.

Se você usar...	Anexe a política para...
	O EMR Studio usa as permissões que você atribui aos perfis do IAM para os usuários com acesso federado a um Studio.
IAM Identity Center	O seu perfil de usuário do Amazon EMR Studio.

Exemplo de políticas de usuário

A política de usuário básico apresentada a seguir permite a maioria das ações do EMR Studio, mas não permite que um usuário crie novos clusters do Amazon EMR.

Política básica

> **Important**
>
> A política de exemplo não inclui a permissão `CreateStudioPresignedUrl`, que você deve conceder a um usuário ao usar o modo de autenticação do IAM. Para ter mais informações, consulte Atribuir um usuário ou um grupo a um EMR Studio.

A política de exemplo inclui elementos `Condition` para impor o controle de acesso por etiquetas (TBAC) com a finalidade de que você possa usar a política com o perfil de serviço de exemplo para o EMR Studio. Para ter mais informações, consulte Crie um perfil de serviço do EMR Studio.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"AllowDefaultEC2SecurityGroupsCreationInVPCWithEMRTags",
         "Effect":"Allow",
         "Action":[
            "ec2:CreateSecurityGroup"
         ],
         "Resource":[
            "arn:aws:ec2:*:*:vpc/*"
         ],
         "Condition":{
```

```
            "StringEquals":{
               "aws:ResourceTag/for-use-with-amazon-emr-managed-policies":"true"
            }
         }
      },
      {
         "Sid":"AllowAddingEMRTagsDuringDefaultSecurityGroupCreation",
         "Effect":"Allow",
         "Action":[
            "ec2:CreateTags"
         ],
         "Resource":"arn:aws:ec2:*:*:security-group/*",
         "Condition":{
            "StringEquals":{
               "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true",
               "ec2:CreateAction":"CreateSecurityGroup"
            }
         }
      },
      {
         "Sid":"AllowSecretManagerListSecrets",
         "Action":[
            "secretsmanager:ListSecrets"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
         "Effect":"Allow",
         "Action":"secretsmanager:CreateSecret",
         "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
         "Condition":{
            "StringEquals":{
               "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
            }
         }
      },
      {
         "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
         "Effect":"Allow",
         "Action":"secretsmanager:TagResource",
         "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
      },
```

```
    {
       "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
       "Action":"iam:PassRole",
       "Resource":[
          "arn:aws:iam::*:role/<your-emr-studio-service-role>"
       ],
       "Effect":"Allow"
    },
    {
       "Sid":"AllowS3ListAndLocationPermissions",
       "Action":[
          "s3:ListAllMyBuckets",
          "s3:ListBucket",
          "s3:GetBucketLocation"
       ],
       "Resource":"arn:aws:s3:::*",
       "Effect":"Allow"
    },
    {
       "Sid":"AllowS3ReadOnlyAccessToLogs",
       "Action":[
          "s3:GetObject"
       ],
       "Resource":[
          "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
       ],
       "Effect":"Allow"
    },
    {
       "Sid":"AllowConfigurationForWorkspaceCollaboration",
       "Action":[
          "elasticmapreduce:UpdateEditor",
          "elasticmapreduce:PutWorkspaceAccess",
          "elasticmapreduce:DeleteWorkspaceAccess",
          "elasticmapreduce:ListWorkspaceAccessIdentities"
       ],
       "Resource":"*",
       "Effect":"Allow",
       "Condition":{
          "StringEquals":{
             "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
          }
       }
    },
```

```
      {
         "Sid":"DescribeNetwork",
         "Effect":"Allow",
         "Action":[
            "ec2:DescribeVpcs",
            "ec2:DescribeSubnets",
            "ec2:DescribeSecurityGroups"
         ],
         "Resource":"*"
      },
      {
         "Sid":"ListIAMRoles",
         "Effect":"Allow",
         "Action":[
            "iam:ListRoles"
         ],
         "Resource":"*"
      }
   ]
}
```

A política de usuário intermediário apresentada a seguir permite a maioria das ações do EMR Studio e possibilita que um usuário crie novos clusters do Amazon EMR usando um modelo de cluster.

Política intermediária

> **Important**
>
> A política de exemplo não inclui a permissão `CreateStudioPresignedUrl`, que você deve conceder a um usuário ao usar o modo de autenticação do IAM. Para ter mais informações, consulte Atribuir um usuário ou um grupo a um EMR Studio.

A política de exemplo inclui elementos `Condition` para impor o controle de acesso por etiquetas (TBAC) com a finalidade de que você possa usar a política com o perfil de serviço de exemplo para o EMR Studio. Para ter mais informações, consulte Crie um perfil de serviço do EMR Studio.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"AllowEMRBasicActions",
```

```
        "Action":[
           "elasticmapreduce:CreateEditor",
           "elasticmapreduce:DescribeEditor",
           "elasticmapreduce:ListEditors",
           "elasticmapreduce:StartEditor",
           "elasticmapreduce:StopEditor",
           "elasticmapreduce:DeleteEditor",
           "elasticmapreduce:OpenEditorInConsole",
           "elasticmapreduce:AttachEditor",
           "elasticmapreduce:DetachEditor",
           "elasticmapreduce:CreateRepository",
           "elasticmapreduce:DescribeRepository",
           "elasticmapreduce:DeleteRepository",
           "elasticmapreduce:ListRepositories",
           "elasticmapreduce:LinkRepository",
           "elasticmapreduce:UnlinkRepository",
           "elasticmapreduce:DescribeCluster",
           "elasticmapreduce:ListInstanceGroups",
           "elasticmapreduce:ListBootstrapActions",
           "elasticmapreduce:ListClusters",
           "elasticmapreduce:ListSteps",
           "elasticmapreduce:CreatePersistentAppUI",
           "elasticmapreduce:DescribePersistentAppUI",
           "elasticmapreduce:GetPersistentAppUIPresignedURL",
           "elasticmapreduce:GetOnClusterAppUIPresignedURL"
        ],
        "Resource":"*",
        "Effect":"Allow"
     },
     {
        "Sid":"AllowEMRContainersBasicActions",
        "Action":[
           "emr-containers:DescribeVirtualCluster",
           "emr-containers:ListVirtualClusters",
           "emr-containers:DescribeManagedEndpoint",
           "emr-containers:ListManagedEndpoints",
           "emr-containers:DescribeJobRun",
           "emr-containers:ListJobRuns"
        ],
        "Resource":"*",
        "Effect":"Allow"
     },
     {
         "Sid": "AllowRetrievingManagedEndpointCredentials",
```

```
            "Effect": "Allow",
            "Action": [
                "emr-containers:GetManagedEndpointSessionCredentials"
            ],
            "Resource": [
                "arn:aws:emr-containers:<region>:<account-id>:/virtualclusters/<virtual-
cluster-id>/endpoints/<managed-endpoint-id>"
            ],
            "Condition": {
                "StringEquals": {
                    "emr-containers:ExecutionRoleArn": [
                        "arn:aws:iam::<account-id>:role/<emr-on-eks-execution-role>"
                    ]
                }
            }
        },
        {
           "Sid":"AllowSecretManagerListSecrets",
           "Action":[
              "secretsmanager:ListSecrets"
           ],
           "Resource":"*",
           "Effect":"Allow"
        },
        {
           "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
           "Effect":"Allow",
           "Action":"secretsmanager:CreateSecret",
           "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
           "Condition":{
              "StringEquals":{
                 "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
              }
           }
        },
        {
           "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
           "Effect":"Allow",
           "Action":"secretsmanager:TagResource",
           "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
        },
        {
           "Sid":"AllowClusterTemplateRelatedIntermediateActions",
           "Action":[
```

```
            "servicecatalog:DescribeProduct",
            "servicecatalog:DescribeProductView",
            "servicecatalog:DescribeProvisioningParameters",
            "servicecatalog:ProvisionProduct",
            "servicecatalog:SearchProducts",
            "servicecatalog:UpdateProvisionedProduct",
            "servicecatalog:ListProvisioningArtifacts",
            "servicecatalog:ListLaunchPaths",
            "servicecatalog:DescribeRecord",
            "cloudformation:DescribeStackResources"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
         "Action":"iam:PassRole",
         "Resource":[
            "arn:aws:iam::*:role/<your-emr-studio-service-role>"
         ],
         "Effect":"Allow"
      },
      {
         "Sid":"AllowS3ListAndLocationPermissions",
         "Action":[
            "s3:ListAllMyBuckets",
            "s3:ListBucket",
            "s3:GetBucketLocation"
         ],
         "Resource":"arn:aws:s3:::*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowS3ReadOnlyAccessToLogs",
         "Action":[
            "s3:GetObject"
         ],
         "Resource":[
            "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
         ],
         "Effect":"Allow"
      },
      {
         "Sid":"AllowConfigurationForWorkspaceCollaboration",
```

```
            "Action":[
               "elasticmapreduce:UpdateEditor",
               "elasticmapreduce:PutWorkspaceAccess",
               "elasticmapreduce:DeleteWorkspaceAccess",
               "elasticmapreduce:ListWorkspaceAccessIdentities"
            ],
            "Resource":"*",
            "Effect":"Allow",
            "Condition":{
               "StringEquals":{
                  "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
               }
            }
         },
         {
            "Sid":"DescribeNetwork",
            "Effect":"Allow",
            "Action":[
               "ec2:DescribeVpcs",
               "ec2:DescribeSubnets",
               "ec2:DescribeSecurityGroups"
            ],
            "Resource":"*"
         },
         {
            "Sid":"ListIAMRoles",
            "Effect":"Allow",
            "Action":[
               "iam:ListRoles"
            ],
            "Resource":"*"
         },
         {
            "Sid": "AllowServerlessActions",
            "Action": [
               "emr-serverless:CreateApplication",
               "emr-serverless:UpdateApplication",
               "emr-serverless:DeleteApplication",
               "emr-serverless:ListApplications",
               "emr-serverless:GetApplication",
               "emr-serverless:StartApplication",
               "emr-serverless:StopApplication",
               "emr-serverless:StartJobRun",
               "emr-serverless:CancelJobRun",
```

```
            "emr-serverless:ListJobRuns",
            "emr-serverless:GetJobRun",
            "emr-serverless:GetDashboardForJobRun",
            "emr-serverless:AccessInteractiveEndpoints"
         ],
         "Resource": "*",
         "Effect": "Allow"
        },
        {
           "Sid": "AllowPassingRuntimeRoleForRunningServerlessJob",
           "Action": "iam:PassRole",
           "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
           "Effect": "Allow"
        }
    ]
}
```

A política de usuário avançada apresentada a seguir permite todas as ações do EMR Studio e possibilita que um usuário crie novos clusters do Amazon EMR ao usar um modelo de cluster ou ao fornecer uma configuração de cluster.

Política avançada

> **Important**
>
> A política de exemplo não inclui a permissão `CreateStudioPresignedUrl`, que você deve conceder a um usuário ao usar o modo de autenticação do IAM. Para ter mais informações, consulte Atribuir um usuário ou um grupo a um EMR Studio.

A política de exemplo inclui elementos `Condition` para impor o controle de acesso por etiquetas (TBAC) com a finalidade de que você possa usar a política com o perfil de serviço de exemplo para o EMR Studio. Para ter mais informações, consulte Crie um perfil de serviço do EMR Studio.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"AllowEMRBasicActions",
         "Action":[
            "elasticmapreduce:CreateEditor",
            "elasticmapreduce:DescribeEditor",
```

```
            "elasticmapreduce:ListEditors",
            "elasticmapreduce:StartEditor",
            "elasticmapreduce:StopEditor",
            "elasticmapreduce:DeleteEditor",
            "elasticmapreduce:OpenEditorInConsole",
            "elasticmapreduce:AttachEditor",
            "elasticmapreduce:DetachEditor",
            "elasticmapreduce:CreateRepository",
            "elasticmapreduce:DescribeRepository",
            "elasticmapreduce:DeleteRepository",
            "elasticmapreduce:ListRepositories",
            "elasticmapreduce:LinkRepository",
            "elasticmapreduce:UnlinkRepository",
            "elasticmapreduce:DescribeCluster",
            "elasticmapreduce:ListInstanceGroups",
            "elasticmapreduce:ListBootstrapActions",
            "elasticmapreduce:ListClusters",
            "elasticmapreduce:ListSteps",
            "elasticmapreduce:CreatePersistentAppUI",
            "elasticmapreduce:DescribePersistentAppUI",
            "elasticmapreduce:GetPersistentAppUIPresignedURL",
            "elasticmapreduce:GetOnClusterAppUIPresignedURL"
        ],
        "Resource":"*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowEMRContainersBasicActions",
        "Action":[
            "emr-containers:DescribeVirtualCluster",
            "emr-containers:ListVirtualClusters",
            "emr-containers:DescribeManagedEndpoint",
            "emr-containers:ListManagedEndpoints",
            "emr-containers:DescribeJobRun",
            "emr-containers:ListJobRuns"
        ],
        "Resource":"*",
        "Effect":"Allow"
    },
    {
        "Sid": "AllowRetrievingManagedEndpointCredentials",
        "Effect": "Allow",
        "Action": [
            "emr-containers:GetManagedEndpointSessionCredentials"
```

```
            ],
            "Resource": [
                "arn:aws:emr-containers:<region>:<account-id>:/virtualclusters/<virtual-
cluster-id>/endpoints/<managed-endpoint-id>"
            ],
            "Condition": {
                "StringEquals": {
                    "emr-containers:ExecutionRoleArn": [
                        "arn:aws:iam::<account-id>:role/<emr-on-eks-execution-role>"
                    ]
                }
            }
        },
        {
            "Sid":"AllowSecretManagerListSecrets",
            "Action":[
               "secretsmanager:ListSecrets"
            ],
            "Resource":"*",
            "Effect":"Allow"
        },
        {
            "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
            "Effect":"Allow",
            "Action":"secretsmanager:CreateSecret",
            "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
            "Condition":{
               "StringEquals":{
                  "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
               }
            }
        },
        {
            "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
            "Effect":"Allow",
            "Action":"secretsmanager:TagResource",
            "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
        },
        {
            "Sid":"AllowClusterTemplateRelatedIntermediateActions",
            "Action":[
               "servicecatalog:DescribeProduct",
               "servicecatalog:DescribeProductView",
               "servicecatalog:DescribeProvisioningParameters",
```

```
            "servicecatalog:ProvisionProduct",
            "servicecatalog:SearchProducts",
            "servicecatalog:UpdateProvisionedProduct",
            "servicecatalog:ListProvisioningArtifacts",
            "servicecatalog:ListLaunchPaths",
            "servicecatalog:DescribeRecord",
            "cloudformation:DescribeStackResources"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowEMRCreateClusterAdvancedActions",
         "Action":[
            "elasticmapreduce:RunJobFlow"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
         "Action":"iam:PassRole",
         "Resource":[
            "arn:aws:iam::*:role/<your-emr-studio-service-role>",
            "arn:aws:iam::*:role/EMR_DefaultRole_V2",
            "arn:aws:iam::*:role/EMR_EC2_DefaultRole"
         ],
         "Effect":"Allow"
      },
      {
         "Sid":"AllowS3ListAndLocationPermissions",
         "Action":[
            "s3:ListAllMyBuckets",
            "s3:ListBucket",
            "s3:GetBucketLocation"
         ],
         "Resource":"arn:aws:s3:::*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowS3ReadOnlyAccessToLogs",
         "Action":[
            "s3:GetObject"
         ],
```

```
            "Resource":[
               "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
            ],
            "Effect":"Allow"
         },
         {
            "Sid":"AllowConfigurationForWorkspaceCollaboration",
            "Action":[
               "elasticmapreduce:UpdateEditor",
               "elasticmapreduce:PutWorkspaceAccess",
               "elasticmapreduce:DeleteWorkspaceAccess",
               "elasticmapreduce:ListWorkspaceAccessIdentities"
            ],
            "Resource":"*",
            "Effect":"Allow",
            "Condition":{
               "StringEquals":{
                  "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
               }
            }
         },
         {
            "Sid" : "SageMakerDataWranglerForEMRStudio",
            "Effect" : "Allow",
            "Action" : [
               "sagemaker:CreatePresignedDomainUrl",
               "sagemaker:DescribeDomain",
               "sagemaker:ListDomains",
               "sagemaker:ListUserProfiles"
            ],
            "Resource":"*"
         },
         {
            "Sid":"DescribeNetwork",
            "Effect":"Allow",
            "Action":[
               "ec2:DescribeVpcs",
               "ec2:DescribeSubnets",
               "ec2:DescribeSecurityGroups"
            ],
            "Resource":"*"
         },
         {
            "Sid":"ListIAMRoles",
```

```
            "Effect":"Allow",
            "Action":[
               "iam:ListRoles"
            ],
            "Resource":"*"
         },
         {
            "Sid": "AllowServerlessActions",
            "Action": [
               "emr-serverless:CreateApplication",
               "emr-serverless:UpdateApplication",
               "emr-serverless:DeleteApplication",
               "emr-serverless:ListApplications",
               "emr-serverless:GetApplication",
               "emr-serverless:StartApplication",
               "emr-serverless:StopApplication",
               "emr-serverless:StartJobRun",
               "emr-serverless:CancelJobRun",
               "emr-serverless:ListJobRuns",
               "emr-serverless:GetJobRun",
               "emr-serverless:GetDashboardForJobRun",
               "emr-serverless:AccessInteractiveEndpoints"
            ],
            "Resource": "*",
            "Effect": "Allow"
         },
         {
            "Sid": "AllowPassingRuntimeRoleForRunningServerlessJob",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
            "Effect": "Allow"
         },
         {
            "Sid": "AllowCodeWhisperer",
            "Effect": "Allow",
            "Action": [ "codewhisperer:GenerateRecommendations" ],
            "Resource": "*"
         },
         {
            "Sid": "AllowAthenaSQL",
            "Action": [
               "athena:StartQueryExecution",
               "athena:StopQueryExecution",
               "athena:GetQueryExecution",
```

```
            "athena:GetQueryRuntimeStatistics",
            "athena:GetQueryResults",
            "athena:ListQueryExecutions",
            "athena:BatchGetQueryExecution",
            "athena:GetNamedQuery",
            "athena:ListNamedQueries",
            "athena:BatchGetNamedQuery",
            "athena:UpdateNamedQuery",
            "athena:DeleteNamedQuery",
            "athena:ListDataCatalogs",
            "athena:GetDataCatalog",
            "athena:ListDatabases",
            "athena:GetDatabase",
            "athena:ListTableMetadata",
            "athena:GetTableMetadata",
            "athena:ListWorkGroups",
            "athena:GetWorkGroup",
            "athena:CreateNamedQuery",
            "athena:GetPreparedStatement",
            "glue:CreateDatabase",
            "glue:DeleteDatabase",
            "glue:GetDatabase",
            "glue:GetDatabases",
            "glue:UpdateDatabase",
            "glue:CreateTable",
            "glue:DeleteTable",
            "glue:BatchDeleteTable",
            "glue:UpdateTable",
            "glue:GetTable",
            "glue:GetTables",
            "glue:BatchCreatePartition",
            "glue:CreatePartition",
            "glue:DeletePartition",
            "glue:BatchDeletePartition",
            "glue:UpdatePartition",
            "glue:GetPartition",
            "glue:GetPartitions",
            "glue:BatchGetPartition",
            "kms:ListAliases",
            "kms:ListKeys",
            "kms:DescribeKey",
            "lakeformation:GetDataAccess",
            "s3:GetBucketLocation",
            "s3:GetBucketLocation",
```

```
            "s3:GetObject",
            "s3:ListBucket",
            "s3:ListBucketMultipartUploads",
            "s3:ListMultipartUploadParts",
            "s3:AbortMultipartUpload",
            "s3:PutObject",
            "s3:PutBucketPublicAccessBlock",
            "s3:ListAllMyBuckets"
        ],
        "Resource": "*",
        "Effect": "Allow"
    }
  ]
}
```

A política de usuário a seguir contém as permissões mínimas de usuário necessárias para usar uma aplicação interativa do EMR Serverless com os Workspaces do EMR Studio.

Política interativa do EMR Serverless

Neste exemplo de política que tem permissões de usuário para aplicativos interativos do EMR Serverless com o EMR Studio, substitua os espaços reservados para e por *emr-studio-service-role*sua função de serviço do EMR Studio e função de *serverless-runtime-role*tempo de execução do EMR Serverless corretas.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowServerlessActions",
            "Action": [
                "emr-serverless:CreateApplication",
                "emr-serverless:UpdateApplication",
                "emr-serverless:DeleteApplication",
                "emr-serverless:ListApplications",
                "emr-serverless:GetApplication",
                "emr-serverless:StartApplication",
                "emr-serverless:StopApplication",
                "emr-serverless:StartJobRun",
                "emr-serverless:CancelJobRun",
                "emr-serverless:ListJobRuns",
                "emr-serverless:GetJobRun",
                "emr-serverless:GetDashboardForJobRun",
```

```
                "emr-serverless:AccessInteractiveEndpoints"
            ],
            "Resource": "*",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowEMRBasicActions",
            "Action": [
                "elasticmapreduce:CreateEditor",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:UpdateStudio",
                "elasticmapreduce:StartEditor",
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:OpenEditorInConsole",
                "elasticmapreduce:AttachEditor",
                "elasticmapreduce:DetachEditor",
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio",
                "elasticmapreduce:ListStudios",
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "*",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowPassingRuntimeRoleForRunningEMRServerlessJob",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowPassingServiceRoleForWorkspaceCreation",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/emr-studio-service-role",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowS3ListAndGetPermissions",
            "Action": [
                "s3:ListAllMyBuckets",
                "s3:ListBucket",
```

```
                "s3:GetBucketLocation",
                "s3:GetObject"
            ],
            "Resource": "arn:aws:s3:::*",
            "Effect": "Allow"
        },
        {
            "Sid":"DescribeNetwork",
            "Effect":"Allow",
            "Action":[
                "ec2:DescribeVpcs",
                "ec2:DescribeSubnets",
                "ec2:DescribeSecurityGroups"
            ],
            "Resource":"*"
        },
        {
            "Sid": "ListIAMRoles",
            "Effect": "Allow",
            "Action": [
                "iam:ListRoles"
            ],
            "Resource": "*"
        }
    ]
}
```

Permissões do AWS Identity and Access Management para usuários do EMR Studio

A tabela a seguir inclui cada operação do Amazon EMR Studio que um usuário pode realizar e lista as ações mínimas do IAM que são necessárias para executar essa operação. Você permite essas ações nas suas políticas de permissões do IAM (ao usar a autenticação do IAM) ou nas políticas de sessão do perfil de usuário (ao usar a autenticação do Centro de Identidade do IAM) para o EMR Studio.

A tabela também exibe as operações permitidas em cada exemplo de política de permissões para o EMR Studio. Para obter mais informações sobre exemplos de políticas de permissões, consulte Criação de políticas de permissões para usuários do EMR Studio.

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Criação e exclusão de Workspaces	Sim	Sim	Sim	`"elasticmapreduce:CreateEditor", "elasticmapreduce:DescribeEditor", "elasticmapreduce:ListEditors", "elasticmapreduce:DeleteEditor"`
Visualização do painel Colaboração, habilitação da colaboração no Workspace e adição de colaboradores. Para obter mais informações, consulte Definição de propriedade para colaboração no Workspace.	Sim	Sim	Sim	`"elasticmapreduce:UpdateEditor", "elasticmapreduce:PutWorkspaceAccess", "elasticmapreduce:DeleteWorkspaceAccess", "elasticmapreduce:ListWorkspaceAccessIdentities"`
Visualização de uma lista de buckets de armazenamento do Amazon S3 Control na mesma conta do Studio ao criar um novo cluster do EMR e acesso ao logs de contêiner ao usar uma interface do usuário da Web para depurar aplicações.	Sim	Sim	Sim	`"s3:ListAllMyBuckets", "s3:ListBucket", "s3:GetBucketLocation", "s3:GetObject"`
Acesso aos Workspaces.	Sim	Sim	Sim	`"elasticmapreduce:DescribeEditor",`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
				`"elasticmapreduce:ListEditors", "elasticmapreduce:StartEditor", "elasticmapreduce:StopEditor", "elasticmapreduce:OpenEditorInConsole"`
Anexo ou remoção do anexo para clusters existentes do Amazon EMR associados ao Workspace.	Sim	Sim	Sim	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster", "elasticmapreduce:ListInstanceGroups", "elasticmapreduce:ListBootstrapActions"`
Anexo ou remoção do anexo para clusters do Amazon EMR no EKS.	Sim	Sim	Sim	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "emr-containers:ListVirtualClusters", "emr-containers:DescribeVirtualCluster", "emr-containers:ListManagedEndpoints", "emr-containers:DescribeManagedEndpoint", "emr-containers:GetManagedEndpointSessionCredentials"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Anexar ou desanexar aplicações do EMR Serverless associadas ao Workspace	Não	Sim	Sim	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "emr-serverless:GetApplication", "emr-serverless:StartApplication", "emr-serverless:ListApplications", "emr-serverless:GetDashboardForJobRun", "emr-serverless:AccessInteractiveEndpoints", "iam:PassRole"` A permissão `PassRole` é necessária para passar a função de runtime de tarefas do EMR Serverless. Para obter mais informações, consulte Funções de runtime de trabalho no Guia do usuário do Amazon EMR Serverless.

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Depuração do Amazon EMR em trabalhos do EC2 com interfaces do usuário de aplicações persistentes.	Sim	Sim	Sim	`"elasticmapreduce:CreatePersistentAppUI", "elasticmapreduce:DescribePersistentAppUI", "elasticmapreduce:GetPersistentAppUIPresignedURL", "elasticmapreduce:ListClusters", "elasticmapreduce:ListSteps", "elasticmapreduce:DescribeCluster", "s3:ListBucket", "s3:GetObject"`
Depuração do Amazon EMR em trabalhos do EC2 com interfaces do usuário de aplicações no cluster.	Sim	Sim	Sim	`"elasticmapreduce:GetOnClusterAppUIPresignedURL"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Depuração de execuções de trabalhos do Amazon EMR no EKS usando o servidor de histórico do Spark.	Sim	Sim	Sim	`"elasticmapreduce:CreatePersistentAppUI", "elasticmapreduce:DescribePersistentAppUI", "elasticmapreduce:GetPersistentAppUIPresignedURL", "emr-containers:ListVirtualClusters", "emr-containers:DescribeVirtualCluster", "emr-containers:ListJobRuns", "emr-containers:DescribeJobRun", "s3:ListBucket", "s3:GetObject"`
Criação e exclusão de repositórios Git.	Sim	Sim	Sim	`"elasticmapreduce:CreateRepository", "elasticmapreduce:DeleteRepository", "elasticmapreduce:ListRepositories", "elasticmapreduce:DescribeRepository", "secretsmanager:CreateSecret", "secretsmanager:ListSecrets", "secretsmanager:TagResource"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Vinculação e desvinculação de repositórios Git.	Sim	Sim	Sim	`"elasticmapreduce:LinkRepository", "elasticmapreduce:UnlinkRepository", "elasticmapreduce:ListRepositories", "elasticmapreduce:DescribeRepository"`
Criação de novos clusters a partir de modelos de cluster definidos previamente.	Não	Sim	Sim	`"servicecatalog:SearchProducts", "servicecatalog:DescribeProduct", "servicecatalog:DescribeProductView", "servicecatalog:DescribeProvisioningParameters", "servicecatalog:ProvisionProduct", "servicecatalog:UpdateProvisionedProduct", "servicecatalog:ListProvisioningArtifacts", "servicecatalog:DescribeRecord", "servicecatalog:ListLaunchPaths", "cloudformation:DescribeStackResources", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Forneça uma configuração de cluster para criar clusters.	Não	Não	Sim	`"elasticmapreduce:RunJobFlow", "iam:PassRole", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster"`
Atribuição de um usuário a um Studio ao usar o modo de autenticação do IAM.	Não	Não	Não	`"elasticmapreduce:CreateStudioPresignedUrl"`
Descrição dos objetos das redes.	Sim	Sim	Sim	`{ "Version": "2012-10-17", "Statement": [ { "Sid": "DescribeNetwork", "Effect": "Allow", "Action": [ "ec2:DescribeVpcs", "ec2:DescribeSubnets", "ec2:DescribeSecurityGroups" ], "Resource": "*" } ] }`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Listagem dos perfis do IAM.	Sim	Sim	Sim	`{ "Version": "2012-10-17", "Statement": [ { "Sid": "ListIAMRoles", "Effect": "Allow", "Action": [ "iam:ListRoles" ], "Resource": "*" } ] }`
Conecte-se ao EMR Studio a partir do Amazon SageMaker Studio e use a interface visual do Data Wrangler.	Não	Não	Sim	`"sagemaker:CreatePresignedDomainUrl", "sagemaker:DescribeDomain", "sagemaker:ListDomains", "sagemaker:ListUserProfiles"`
Use a Amazon CodeWhisperer em seu EMR Studio.	Não	Não	Sim	`"codewhisperer:GenerateRecommendations"`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
Acesso ao editor SQL do Amazon Athena por meio do EMR Studio. Essa lista pode não incluir todas as permissões necessárias para usar todos os recursos do Athena. Para ver a up-to-date lista completa, consulte a política de acesso total do Athena.	Não	Não	Sim	`"athena:StartQueryExecution", "athena:StopQueryExecution", "athena:GetQueryExecution", "athena:GetQueryRuntimeStatistics", "athena:GetQueryResults", "athena:ListQueryExecutions", "athena:BatchGetQueryExecution", "athena:GetNamedQuery", "athena:ListNamedQueries", "athena:BatchGetNamedQuery", "athena:UpdateNamedQuery", "athena:DeleteNamedQuery", "athena:ListDataCatalogs", "athena:GetDataCatalog", "athena:ListDatabases", "athena:GetDatabase", "athena:ListTableMetadata", "athena:GetTableMetadata", "athena:ListWorkGroups", "athena:GetWorkGroup", "athena:CreateNamedQuery", "athena:GetPreparedStatement", "glue:CreateDatabase",`

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
				"glue:DeleteDatabase", "glue:GetDatabase", "glue:GetDatabases", "glue:UpdateDatabase", "glue:CreateTable", "glue:DeleteTable", "glue:BatchDeleteTable", "glue:UpdateTable", "glue:GetTable", "glue:GetTables", "glue:BatchCreatePartition", "glue:CreatePartition", "glue:DeletePartition", "glue:BatchDeletePartition", "glue:UpdatePartition", "glue:GetPartition", "glue:GetPartitions", "glue:BatchGetPartition", "kms:ListAliases", "kms:ListKeys", "kms:DescribeKey", "lakeformation:GetDataAccess", "s3:GetBucketLocation", "s3:GetBucketLocation", "s3:GetObject", "s3:ListBucket", "s3:ListBucketMultipartUploads", "s3:ListMultipartUploadParts", "s3:AbortMultipartUpload", "s3:PutObject", "s3:PutBucketPublicAccessBlock",

Ação	Basic	Intermediário	Advanced (Avançado)	Ações associadas
				`"s3:ListAllMyBuckets"`

Crie um EMR Studio

É possível criar um EMR Studio para a sua equipe com o console do Amazon EMR ou a AWS CLI. A criação de uma instância do Studio faz parte da configuração do Amazon EMR Studio.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

Pré-requisitos

Antes de criar um Studio, certifique-se de ter concluído as tarefas anteriores em Configuração de um Amazon EMR Studio.

Para criar um Studio usando a AWS CLI, você deve ter a versão mais recente instalada. Para obter mais informações, consulte Instalar ou atualizar a versão mais recente da AWS CLI.

> **Important**
>
> Desative as ferramentas de gerenciamento de proxy, como FoxyProxy ou SwitchyOmega no navegador, antes de criar um Studio. Os proxies ativos podem resultar em uma mensagem de erro de falha de rede quando você escolhe Criar Studio.

O Amazon EMR oferece uma experiência de console simples para criar um Studio, para que você possa começar rapidamente com as configurações padrão para executar cargas de trabalho interativas ou trabalhos em lotes com as configurações padrão. A criação de um EMR Studio também cria um aplicativo EMR Serverless pronto para seus trabalhos interativos.

Se quiser ter controle total sobre as configurações do seu Studio, você pode escolher Personalizado, que permite definir todas as configurações adicionais.

Interactive workloads

Para criar um EMR Studio para cargas de trabalho interativas

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR Studio, no painel de navegação à esquerda, escolha Conceitos básicos. Você também pode criar um novo Studio na página Studios.
3. O Amazon EMR fornece configurações padrão para você se você estiver criando um EMR Studio para cargas de trabalho interativas, mas você pode editar essas configurações. As configurações configuráveis incluem o nome do EMR Studio, a localização do seu espaço de trabalho no S3, a função de serviço a ser usada, os espaços de trabalho que você deseja usar, o nome do aplicativo EMR Serverless e a função de tempo de execução associada.
4. Escolha Create Studio e inicie o Workspace para finalizar e navegar até a página Studios. Seu novo Studio aparece na lista com detalhes como o Nome do Studio, Data de criação e URL de acesso do Studio. Seu espaço de trabalho é aberto em uma nova guia no seu navegador.

Batch jobs

Para criar um EMR Studio para cargas de trabalho interativas

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR Studio, no painel de navegação à esquerda, escolha Conceitos básicos. Você também pode criar um novo Studio na página Studios.
3. O Amazon EMR fornece configurações padrão para você se você estiver criando um EMR Studio para trabalhos em lotes, mas você pode editar essas configurações. As configurações configuráveis incluem o nome do EMR Studio, o nome do aplicativo EMR Serverless e a função de tempo de execução associada.
4. Escolha Create Studio e inicie o Workspace para finalizar e navegar até a página Studios. Seu novo Studio aparece na lista com detalhes como o Nome do Studio, Data de criação e URL de acesso do Studio. Seu EMR Studio abre em uma nova guia no seu navegador.

Custom settings

Para criar um EMR Studio com configurações personalizadas

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR Studio, no painel de navegação à esquerda, escolha Conceitos básicos. Você também pode criar um novo Studio na página Studios.
3. Escolha Criar um Studio para abrir a página Criar um Studio.
4. Insira o nome do estúdio.
5. Escolha criar um novo bucket do S3 ou usar um local existente.
6. Escolha o espaço de trabalho a ser adicionado ao Studio. Você pode adicionar até 3 espaços de trabalho.
7. Em Autenticação, escolha um modo de autenticação para o Studio e forneça informações de acordo com a tabela a seguir. Para saber mais sobre a autenticação para o EMR Studio, consulte Escolha um modo de autenticação para o Amazon EMR Studio.

Se você usar...	Fazer isso...
A autenticação ou a federação do IAM	O método de autenticação padrão é o AWS Identity and Access Management (IAM). Na parte inferior da tela, você também pode adicionar tags para dar acesso ao Studio para usuários específicos, conforme descrito em Atribuir um usuário ou um grupo a um EMR Studio. Se você quiser que os usuários federados façam login usando o URL do Studio e as credenciais do seu provedor de identidade (IdP), selecione seu IdP na lista suspensa e insira o URL de login do provedor de identidade (IdP) e o nome do parâmetro. RelayState Para obter uma lista de URLs e RelayStat e nomes de autenticação de IdP, consulte.

Se você usar...	Fazer isso...
	RelayState Parâmetros do provedor de identidade e URLs de autenticação
Autenticação do IAM Identity Center	Selecione seu Perfil de serviço e Perfil de usuário do EMR Studio. Para obter mais informações, consulte Criação de um perfil de usuário do EMR Studio para o modo de autenticação do Centro de Identidade do IAM e Crie um perfil de serviço do EMR Studio. Ao usar a autenticação do Centro de Identidade do IAM (antigo AWS Single Sign On) para o Studio, você pode optar por simplificar a experiência de login dos usuários com a opção Habilitar propagação de identidade confiável. Com a propagaçã o de identidade confiável, os usuários podem fazer login com as credenciais do Centro de Identidade e ter suas identidades propagadas para serviços de downstream da AWS ao usarem o Studio. Na seção Acesso à aplicação, você também pode especificar se todos os usuários e grupos no seu Centro de Identidade devem ter acesso ao Studio ou se somente os usuários e grupos atribuídos que você escolher podem acessá-lo. Para obter mais informações, consulte Integre o Amazon EMR com AWS IAM Identity Center e Trusted identity propagati on across applications no Guia do usuário do Centro de Identidade do AWS IAM.

8. Para VPC, escolha uma Amazon Virtual Private Cloud (VPC) para o Studio na lista suspensa.

9. Em Sub-redes, selecione, no máximo, cinco sub-redes em sua VPC para associar ao Studio. Você tem a opção de adicionar mais sub-redes após a criação do Studio.
10. Em Grupos de segurança, escolha os grupos de segurança padrão ou os grupos de segurança personalizados. Para ter mais informações, consulte Definição de grupos de segurança para controlar o tráfego de rede do EMR Studio.

Se você escolher...	Fazer isso...
Os grupos de segurança padrão do EMR Studio	Para habilitar a vinculação de repositórios baseados em Git para o Studio, escolha Habilitar clusters, endpoints e o repositór io Git. Caso contrário, escolha Habilitar clusters e endpoints.
Os grupos de segurança personalizados para seu Studio	• Em Grupo de segurança de cluster e endpoint, selecione o grupo de segurança do mecanismo que você configurou usando a lista suspensa. Seu Studio usa esse grupo de segurança para permitir o acesso de entrada de Workspaces anexados. • Em Grupo de segurança do Workspace , selecione o grupo de segurança do Workspace que você configurou usando a lista suspensa. Seu Studio usa esse grupo de segurança com os Workspace s para fornecer o acesso de saída aos clusters anexados do Amazon EMR e aos repositórios Git hospedados publicamente.

11. Adicione tags ao seu Studio e a outros recursos. Para obter mais informações sobre tags, consulte Clusters de tags.
12. Escolha Create Studio e inicie o Workspace para finalizar e navegar até a página Studios. Seu novo Studio aparece na lista com detalhes como o Nome do Studio, Data de criação e URL de acesso do Studio.

Depois de criar um Studio, siga as instruções em Atribuir um usuário ou um grupo a um EMR Studio.

CLI

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Example – Crie um EMR Studio que use o IAM para autenticação

O exemplo de comando da AWS CLI a seguir cria um EMR Studio com o modo de autenticação do IAM. Ao usar a autenticação ou a federação do IAM para o Studio, você não especifica um `--user-role`.

Para permitir que os usuários federados façam login usando o URL do Studio e as credenciais do seu provedor de identidades (IdP), especifique seu `--idp-auth-url` e seu `--idp-relay-state-parameter-name`. Para obter uma lista de URLs e RelayState nomes de autenticação de IdP, consulte. RelayState Parâmetros do provedor de identidade e URLs de autenticação

```
aws emr create-studio \
--name <example-studio-name> \
--auth-mode IAM \
--vpc-id <example-vpc-id> \
--subnet-ids <subnet-id-1> <subnet-id-2>... <subnet-id-5>  \
--service-role <example-studio-service-role-name> \
--user-role studio-user-role-name \
--workspace-security-group-id <example-workspace-sg-id> \
--engine-security-group-id <example-engine-sg-id> \
--default-s3-location <example-s3-location> \
--idp-auth-url <https://EXAMPLE/login/> \
--idp-relay-state-parameter-name <example-RelayState>
```

Example – Crie um EMR Studio que use o Centro de Identidade para autenticação

O exemplo de comando da AWS CLI a seguir cria um EMR Studio que usa o modo de autenticação do Centro de Identidade do IAM. Ao usar a autenticação do Centro de Identidade do IAM, você deve especificar um `--user-role`.

Para obter mais informações sobre o modo de autenticação do Centro de Identidade do IAM, consulte Configuração do modo de autenticação do Centro de Identidade do IAM para o Amazon EMR Studio.

```
aws emr create-studio \
--name <example-studio-name> \
--auth-mode SSO \
--vpc-id <example-vpc-id> \
--subnet-ids <subnet-id-1> <subnet-id-2>... <subnet-id-5>  \
--service-role <example-studio-service-role-name> \
--user-role <example-studio-user-role-name> \
--workspace-security-group-id <example-workspace-sg-id> \
--engine-security-group-id <example-engine-sg-id> \
--default-s3-location <example-s3-location>
--trusted-identity-propagation-enabled \
--idc-user-assignment OPTIONAL \
--idc-instance-arn <iam-identity-center-instance-arn>
```

Example – Saída da CLI para **`aws emr create-studio`**

A seguir, é apresentado um exemplo da saída que aparece após a criação de um Studio.

```
{
    StudioId: "es-123XXXXXXXXX",
    Url: "https://es-123XXXXXXXXX.emrstudio-prod.us-east-1.amazonaws.com"
}
```

Para obter mais informações sobre o comando `create-studio`, consulte AWS CLI Command Reference.

RelayState Parâmetros do provedor de identidade e URLs de autenticação

Ao usar a federação do IAM e quiser que os usuários façam login usando o URL do Studio e as credenciais do seu provedor de identidade (IdP), você pode especificar o URL de login do provedor de identidade (IdP) RelayStatee o nome do parâmetro quando quiser. Crie um EMR Studio

A tabela a seguir mostra o URL de autenticação padrão e o nome do RelayState parâmetro para alguns provedores de identidade populares.

Provedor de identidades	Parâmetro	URL de autenticação
Auth0	`RelayState`	`https://<sub_domain> .auth0.com/samlp/<app_id>`
Contas do Google	`RelayState`	`https://accounts.google.com/o/saml2/initsso?idpid= <idp_id>&spid=<sp_id>&forceauthn=false`
Microsoft Azure	`RelayState`	`https://myapps.microsoft.com/signin/ <app_name> /<app_id>?tenantId= <tenant_id>`
Okta	`RelayState`	`https://<sub_domain> .okta.com/app/<app_name> /<app_id>/sso/saml`
PingFederate	`TargetResource`	`https://<host>/idp/<idp_id>/startSSO.ping?PartnerSpId= <sp_id>`
PingOne	`TargetResource`	`https://sso.connect.pingidentity.com/sso/sp/initsso?saasid= <app_id>&idpid=<idp_id>`

Atribua e gerencie usuários do EMR Studio

Após criar um EMR Studio, você poderá atribuir usuários e grupos a ele. O método usado para atribuir, atualizar e remover usuários depende do modo de autenticação do Studio.

- Ao usar o modo de autenticação do IAM, você configura a atribuição e as permissões de usuários do EMR Studio no IAM ou usando o IAM e seu provedor de identidades.
- Com o modo de autenticação do Centro de Identidade do IAM, você usa o console de gerenciamento do Amazon EMR ou a AWS CLI para gerenciar usuários.

Para saber mais sobre a autenticação para o Amazon EMR Studio, consulte Escolha um modo de autenticação para o Amazon EMR Studio.

Atribuir um usuário ou um grupo a um EMR Studio

IAM

Ao usar Configuração do modo de autenticação do IAM para o Amazon EMR Studio, você deve permitir a ação `CreateStudioPresignedUrl` na política de permissões do IAM para um usuário e restringir o usuário a um Studio específico. Você pode incluir `CreateStudioPresignedUrl` em suas Permissões de usuários para o modo de autenticação do IAM ou usar uma política separada.

Para restringir um usuário a um Studio (ou a um conjunto de Studios), você pode usar o controle de acesso por atributo (ABAC) ou especificar o nome do recurso da Amazon (ARN) de um Studio no elemento `Resource` da política de permissões.

Example Atribuição de um usuário a um Studio usando um ARN do Studio

O exemplo de política a seguir fornece ao usuário o acesso a um EMR Studio específico ao permitir a ação `CreateStudioPresignedUrl` e especificar o nome do recurso da Amazon (ARN) do Studio no elemento `Resource`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-
id>:studio/<studio-id>"
        }
    ]
}
```

Example Atribuição de um usuário a um Studio com ABAC para a autenticação do IAM

Há diversas maneiras de configurar o controle de acesso por atributo (ABAC) para um Studio. Por exemplo, é possível anexar uma ou mais etiquetas a um EMR Studio e, em seguida, criar uma política do IAM que restrinja a ação `CreateStudioPresignedUrl` a um determinado Studio ou a um conjunto de Studios com essas etiquetas.

Você pode adicionar etiquetas durante ou após a criação do Studio. Para adicionar etiquetas a um Studio existente, você pode usar o comando AWS CLI `emr add-tags`. O exemplo apresentado a seguir adiciona uma etiqueta com o par de valores-chave `Team = Data Analytics` a um EMR Studio.

```
aws emr add-tags --resource-id <example-studio-id> --tags Team="Data Analytics"
```

O exemplo de política de permissões a seguir permite a ação `CreateStudioPresignedUrl` para EMR Studios com o par de valores-chave `Team = DataAnalytics` na etiqueta. Para obter mais informações sobre como usar etiquetas para controlar o acesso, consulte Controle de acesso para usuários e funções do IAM usando etiquetas ou Controlar o acesso a recursos da AWS usando etiquetas.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-id>:studio/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/Team": "Data Analytics"
                }
            }
        }
    ]
}
```

Example Atribua um usuário a um Studio usando a chave de condição SourceIdentity global aws:

Ao usar a federação do IAM, é possível usar a chave de condição global `aws:SourceIdentity` em uma política de permissões para conceder aos usuários o acesso ao Studio quando eles assumirem seu perfil do IAM para a federação.

Primeiro, você deve configurar o provedor de identidades (IdP) para retornar uma string de identificação, como um endereço de e-mail ou um nome de usuário, quando um usuário se

autenticar e assumir seu perfil do IAM para a federação. O IAM define a chave de condição global `aws:SourceIdentity` para a string de identificação retornada pelo seu IdP.

Para obter mais informações, consulte a postagem do blog Como relacionar a atividade da função do IAM à identidade corporativa no AWS Security Blog e a SourceIdentity entrada aws: na referência global de chaves de condição.

O exemplo de política a seguir permite a `CreateStudioPresignedUrl` ação e fornece aos usuários um acesso `aws:SourceIdentity` que corresponda ao *< example-source-identity >* ao EMR Studio especificado por *< example-studio-arn >*.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "elasticmapreduce:CreateStudioPresignedUrl",
      "Resource": "<example-studio-arn>",
      "Condition": {
        "StringLike": {
          "aws:SourceIdentity": "<example-source-identity>"
        }
      }
    }
  ]
}
```

IAM Identity Center

Ao atribuir um usuário ou um grupo a um EMR Studio, você especifica uma política de sessão que define permissões detalhadas, como a capacidade de criar um novo cluster do EMR, para esse usuário ou para esse grupo. O Amazon EMR armazena esses mapeamentos de políticas de sessão. É possível atualizar a política de sessão de um usuário ou de um grupo após a atribuição.

> **Note**
>
> As permissões finais para um usuário ou para um grupo são uma interseção entre as permissões definidas em seu perfil de usuário do EMR Studio e as permissões definidas na política de sessão desse usuário ou desse grupo. Se um usuário pertencer a mais de

um grupo atribuído ao Studio, o EMR Studio usará uma união de permissões para esse usuário.

Atribuir usuários ou grupos a um EMR Studio usando o console do Amazon EMR

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha EMR Studio no painel de navegação à esquerda.
3. Escolha o nome do seu Studio na lista Studios ou selecione o Studio e escolha Visualizar detalhes para abrir a página de detalhes do Studio.
4. Escolha Adicionar usuários para visualizar a tabela de pesquisa para Usuários e Grupos.
5. Selecione a guia Usuários ou a guia Grupos e insira um termo de pesquisa na barra de pesquisa para localizar um usuário ou um grupo.
6. Selecione um ou mais usuários ou grupos na lista de resultados da pesquisa. É possível alternar entre as guias Usuários e Grupos.
7. Após selecionar os usuários e os grupos a serem adicionados ao Studio, escolha Adicionar. Você deve visualizar os usuários e os grupos na lista Usuários do Studio. Pode demorar alguns segundos para que a lista seja atualizada.
8. Siga as instruções em Atualizar permissões para um usuário ou um grupo atribuído a um Studio para aprimorar as permissões do Studio para um usuário ou um grupo.

Atribuir um usuário ou um grupo a um EMR Studio usando a AWS CLI

Insira seus próprios valores para os argumentos `create-studio-session-mapping` a seguir. Para obter mais informações sobre o comando `create-studio-session-mapping`, consulte AWS CLI Command Reference.

- **`--studio-id`**: o ID do Studio ao qual você deseja atribuir o usuário ou o grupo. Para obter instruções sobre como recuperar um ID do Studio, consulte Visualização de detalhes do Studio.
- **`--identity-name`**: o nome do usuário ou do grupo no repositório de identidades. Para obter mais informações, consulte UserNamepara usuários e DisplayNamegrupos na Referência da API Identity Store.
- **`--identity-type`**: use `USER` ou `GROUP` para especificar o tipo de identidade.

- **`--session-policy-arn`**: o nome do recurso da Amazon (ARN) da política de sessão que você deseja associar ao usuário ou ao grupo. Por exemplo, **`arn:aws:iam::`*`<aws-account-id>`*`:policy/`*`EMRStudio_Advanced_User_Policy`**. Para ter mais informações, consulte Criação de políticas de permissões para usuários do EMR Studio.

```
aws emr create-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-name <example-identity-name> \
 --identity-type <USER-or-GROUP> \
 --session-policy-arn <example-session-policy-arn>
```

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Use o comando `get-studio-session-mapping` para verificar a nova atribuição. Substitua *`< example-identity-name >`* pelo nome do IAM Identity Center do usuário ou grupo que você atualizou.

```
aws emr get-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
 --identity-name <user-or-group-name> \
```

Atualizar permissões para um usuário ou um grupo atribuído a um Studio

IAM

Para atualizar as permissões de usuários ou de grupos ao usar o modo de autenticação do IAM, use o IAM para alterar as políticas de permissões do IAM anexadas às suas identidades do IAM (usuários, grupos ou perfis).

Para ter mais informações, consulte Permissões de usuários para o modo de autenticação do IAM.

IAM Identity Center

Atualizar as permissões do EMR Studio para um usuário ou um grupo usando o console

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha EMR Studio no painel de navegação à esquerda.
3. Escolha o nome do seu Studio na lista Studios ou selecione o Studio e escolha Visualizar detalhes para abrir a página de detalhes do Studio.
4. Na lista de Usuários do Studio na página de detalhes do Studio, pesquise o usuário ou o grupo que você deseja atualizar. É possível pesquisar por nome ou por tipo de identidade.
5. Selecione o usuário ou o grupo que deseja atualizar e escolha Atribuir política para abrir a caixa de diálogo Política de sessão.
6. Selecione uma política para aplicar ao usuário ou ao grupo escolhido na etapa 5 e escolha Aplicar política. A lista Usuários do Studio deve exibir o nome da política na coluna Política de sessão para o usuário ou para o grupo que você atualizou.

Para atualizar as permissões do EMR Studio para um usuário ou um grupo usando a AWS CLI

Insira seus próprios valores para os argumentos `update-studio-session-mappings` a seguir. Para obter mais informações sobre o comando `update-studio-session-mappings`, consulte AWS CLI Command Reference.

```
aws emr update-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-name <name-of-user-or-group-to-update> \
 --session-policy-arn <new-session-policy-arn-to-apply> \
 --identity-type <USER-or-GROUP> \
```

Use o comando `get-studio-session-mapping` para verificar a nova atribuição de política de sessão. Substitua *< example-identity-name >* pelo nome do IAM Identity Center do usuário ou grupo que você atualizou.

```
aws emr get-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
```

```
  --identity-name <user-or-group-name> \
```

Remover um usuário ou um grupo de um Studio

IAM

Para remover um usuário ou um grupo de um EMR Studio ao usar o modo de autenticação do IAM, você deve revogar o acesso do usuário ao Studio ao configurar novamente a política de permissões do IAM para o usuário.

Na política de exemplo apresentada a seguir, suponha que você tenha um EMR Studio com o par de valores-chave `Team = Quality Assurance` na etiqueta. De acordo com a política, o usuário pode acessar os Studios etiquetados com a chave `Team` cujo valor é igual a `Data Analytics` ou `Quality Assurance`. Para remover o usuário do Studio etiquetado com `Team = Quality Assurance`, remova `Quality Assurance` da lista de valores de etiqueta.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-id>:studio/*",
            "Condition": {
                "StringEquals": {
                    "emr:ResourceTag/Team": [
                        "Data Analytics",
                        "Quality Assurance"
                  ]
                }
            }
        }
    ]
}
```

IAM Identity Center

Remover um usuário ou um grupo de um EMR Studio usando o console

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha EMR Studio no painel de navegação à esquerda.
3. Escolha o nome do seu Studio na lista Studios ou selecione o Studio e escolha Visualizar detalhes para abrir a página de detalhes do Studio.
4. Na lista de Usuários do Studio na página de detalhes do Studio, localize o usuário ou o grupo que você deseja remover do Studio. É possível pesquisar por nome ou por tipo de identidade.
5. Selecione o usuário ou o grupo que deseja excluir, escolha Excluir e confirme. O usuário ou o grupo excluído desaparecerá da lista Usuários do Studio.

Remover um usuário ou um grupo de um EMR Studio usando a AWS CLI

Insira seus próprios valores para os argumentos `delete-studio-session-mapping` a seguir. Para obter mais informações sobre o comando `delete-studio-session-mapping`, consulte AWS CLI Command Reference.

```
aws emr delete-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
 --identity-name <name-of-user-or-group-to-delete> \
```

Gerenciamento de um Amazon EMR Studio

Esta seção inclui instruções para ajudar você a monitorar, atualizar ou excluir um recurso do EMR Studio. Para obter informações sobre como atribuir usuários ou atualizar as permissões de usuários, consulte Atribua e gerencie usuários do EMR Studio.

Visualização de detalhes do Studio

New console

Visualizar detalhes sobre um EMR Studio com o novo console

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR Studio, no painel de navegação à esquerda, escolha Studios.
3. Selecione o Studio na lista Studios para abrir a página de detalhes do Studio. A página de detalhes do Studio inclui informações de Configurações do Studio, como a Descrição, a VPC e as Sub-redes do Studio.

Old console

Visualizar detalhes sobre um EMR Studio com o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/home.
2. Escolha EMR Studio no painel de navegação à esquerda.
3. Selecione o Studio na lista Studios para abrir a página de detalhes do Studio. A página de detalhes do Studio inclui informações de Configurações do Studio, como a Descrição, a VPC e as Sub-redes do Studio.

CLI

Recuperar detalhes de um EMR Studio pelo ID do Studio usando a AWS CLI

Use o comando `describe-studio` da AWS CLI, apresentado a seguir, para buscar informações detalhadas sobre um EMR Studio específico. Para obter mais informações, consulte Referência de comandos da AWS CLI.

```
aws emr describe-studio \
 --studio-id <id-of-studio-to-describe> \
```

Recuperar uma lista de EMR Studios usando a AWS CLI

Use o comando `list-studios` da AWS CLI, apresentado a seguir. Para obter mais informações, consulte Referência de comandos da AWS CLI.

```
aws emr list-studios
```

A seguir, é apresentado um exemplo do valor de retorno do comando `list-studios` no formato JSON.

```
{
    "Studios": [
        {
            "AuthMode": "IAM",
            "VpcId": "vpc-b21XXXXX",
            "Name": "example-studio-name",
            "Url": "https://es-7HWP74SNGDXXXXXXXXXXXXXX.emrstudio-prod.us-
east-1.amazonaws.com",
            "CreationTime": 1605672582.781,
            "StudioId": "es-7HWP74SNGDXXXXXXXXXXXXXX",
            "Description": "example studio description"
        }
    ]
}
```

Monitoramento das ações do Amazon EMR Studio

Visualização da atividade do EMR Studio e da API

O EMR Studio está integrado ao AWS CloudTrail, um serviço que fornece um registro das ações realizadas por um usuário, por um perfil do IAM ou por outro serviço da AWS no EMR Studio. CloudTrail captura chamadas de API para o EMR Studio como eventos. Você pode ver os eventos usando o CloudTrail console em https://console.aws.amazon.com/cloudtrail/.

Os eventos do EMR Studio fornecem informações sobre qual usuário do IAM ou do Studio realiza uma solicitação e qual é o tipo de solicitação.

> **Note**
>
> As ações no cluster, como execução de trabalhos de cadernos, não são emitidas para o AWS CloudTrail.

Você também pode criar uma trilha para entrega contínua de CloudTrail eventos do EMR Studio em um bucket do Amazon S3. Para obter mais informações, consulte o AWS CloudTrailGuia do Usuário.

Exemplo de CloudTrail evento: um usuário chama a DescribeStudio API

Veja a seguir um exemplo de AWS CloudTrail evento criado quando um usuário,`admin`, chama a DescribeStudioAPI. CloudTrail registra o nome do usuário como`admin`.

> **Note**
>
> Para proteger os detalhes do Studio, o evento da API do EMR Studio para DescribeStudio exclui um valor para. `responseElements`


```
{
   "eventVersion":"1.08",
   "userIdentity":{
      "type":"IAMUser",
      "principalId":"AIDXXXXXXXXXXXXXXXXX",
      "arn":"arn:aws:iam::653XXXXXXXX:user/admin",
      "accountId":"653XXXXXXXX",
      "accessKeyId":"AKIAIOSFODNN7EXAMPLE",
      "userName":"admin"
   },
   "eventTime":"2021-01-07T19:13:58Z",
   "eventSource":"elasticmapreduce.amazonaws.com",
   "eventName":"DescribeStudio",
   "awsRegion":"us-east-1",
   "sourceIPAddress":"72.XX.XXX.XX",
   "userAgent":"aws-cli/1.18.188 Python/3.8.5 Darwin/18.7.0 botocore/1.19.28",
   "requestParameters":{
      "studioId":"es-9O5XXXXXXXXXXXXXXXXXXX"
   },
   "responseElements":null,
   "requestID":"0fxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
   "eventID":"b0xxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
   "readOnly":true,
   "eventType":"AwsApiCall",
   "managementEvent":true,
   "eventCategory":"Management",
   "recipientAccountId":"653XXXXXXXX"
}
```

Visualização de atividades de usuários e de trabalhos do Spark

Para visualizar a atividade de trabalho do Spark que é realizada por usuários do Amazon EMR Studio, é possível configurar a representação do usuário em um cluster. Com a representação do usuário, cada trabalho do Spark enviado de um Workspace é associado ao usuário do Studio que executou o código.

Quando a representação do usuário está habilitada, o Amazon EMR cria um diretório de usuários do HDFS no nó primário do cluster para cada usuário que executa códigos no Workspace. Por exemplo, se o usuário `studio-user-1@example.com` executar um código, você poderá se conectar ao nó primário e visualizar que `hadoop fs -ls /user` tem um diretório para `studio-user-1@example.com`.

Para configurar a representação do usuário do Spark, defina as propriedades abaixo nas seguintes classificações de configuração:

- `core-site`
- `livy-conf`

```
[
    {
        "Classification": "core-site",
        "Properties": {
          "hadoop.proxyuser.livy.groups": "*",
          "hadoop.proxyuser.livy.hosts": "*"
        }
    },
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.impersonation.enabled": "true"
        }
    }
]
```

Para visualizar as páginas do servidor de histórico, consulte Depuração de aplicações e trabalhos com o EMR Studio. Você também pode se conectar ao nó primário do cluster usando SSH para visualizar as interfaces da Web da aplicação. Para ter mais informações, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

Atualização de um Amazon EMR Studio

Após criar um EMR Studio, você poderá atualizar os seguintes atributos usando a AWS CLI:

- Nome
- Descrição
- Local do S3 padrão
- Subredes

Atualizar um EMR Studio usando a AWS CLI

Use o comando `update-studio` da AWS CLI para atualizar um EMR Studio. Para obter mais informações, consulte Referência de comandos da AWS CLI.

> **Note**
>
> Você pode associar um Studio a, no máximo, cinco sub-redes. Essas sub-redes devem pertencer à mesma VPC do Studio. A lista de IDs de sub-rede enviada ao comando `update-studio` pode incluir novos IDs de sub-rede, mas também deve incluir todos os IDs de sub-rede que você já associou ao Studio. Não é possível remover sub-redes de um Studio.

```
aws emr update-studio \
 --studio-id <example-studio-id-to-update> \
 --name <example-new-studio-name> \
 --subnet-ids <old-subnet-id-1 old-subnet-id-2 old-subnet-id-3 new-subnet-id> \
```

Para verificar as alterações, use o comando `describe-studio` da AWS CLI e especifique seu ID do Studio. Para obter mais informações, consulte Referência de comandos da AWS CLI.

```
aws emr describe-studio \
 --studio-id <id-of-updated-studio> \
```

Exclusão de um Amazon EMR Studio e de Workspaces

Quando você exclui um Studio, o EMR Studio exclui todas as atribuições de usuários e de grupos do Centro de Identidade do IAM associadas ao Studio.

> **Note**
>
> Quando você exclui um Studio, o Amazon EMR não exclui os Workspaces associados a esse Studio. Você deve excluir os Workspaces do seu Studio separadamente.

Exclusão de Workspaces

Console

Como cada Workspace do EMR Studio corresponde a uma instância de caderno do EMR, você pode usar o console de gerenciamento do Amazon EMR para excluir Workspaces. É possível excluir Workspaces usando o console do Amazon EMR antes ou depois de excluir o Studio.

Excluir um Workspace usando o console do Amazon EMR

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha Cadernos.
3. Selecione os Workspaces que você deseja excluir.
4. Escolha Excluir e, em seguida, selecione Excluir novamente para confirmar.
5. Siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service para remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3.

EMR Studio UI

From the Workspace UI

Exclusão de um Workspace e dos arquivos de backup associados do EMR Studio

1. Faça login no EMR Studio com o URL de acesso do Studio e escolha Workspaces no painel de navegação à esquerda.
2. Localize seu Workspace na lista e, em seguida, marque a caixa de seleção ao lado do nome. É possível selecionar vários Workspaces a serem excluídos ao mesmo tempo.
3. Escolha Excluir no canto superior direito da lista Workspaces e confirme que deseja excluir os Workspaces selecionados. Escolha Delete para confirmar.

4. Se você desejar remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3, siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service. Se não foi você quem criou o Studio, consulte o administrador do Studio para determinar o local de backup do Amazon S3 para o Workspace excluído.

From the Workspaces list

Exclusão de um Workspace e dos arquivos de backup associados da lista Workspaces

1. Navegue até a lista Workspaces no console.
2. Selecione o Workspace que deseja excluir da lista e, em seguida, escolha Ações.
3. Escolha Excluir.
4. Se você desejar remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3, siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service. Se não foi você quem criou o Studio, consulte o administrador do Studio para determinar o local de backup do Amazon S3 para o Workspace excluído.

Exclusão de um EMR Studio

New console

Excluir um EMR Studio com o novo console

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR Studio, no painel de navegação à esquerda, escolha Studios.
3. Selecione o Studio na lista Studios com o botão de alternância à esquerda do nome do Studio. Escolha Excluir.

Old console

Excluir um EMR Studio com o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/home.
2. Escolha EMR Studio no painel de navegação à esquerda.

3. Selecione o Studio na lista Studios e escolha Excluir.

CLI

Excluir um EMR Studio com a AWS CLI

Use o comando `delete-studio` da AWS CLI para excluir um EMR Studio. Para obter mais informações, consulte Referência de comandos da AWS CLI.

```
aws emr delete-studio --studio-id <id-of-studio-to-delete>
```

Criptografando cadernos e arquivos do espaço de trabalho do EMR Studio

No EMR Studio, você pode criar e configurar diferentes áreas de trabalho para organizar e executar notebooks. Esses espaços de trabalho armazenam cadernos e arquivos relacionados em seu bucket Amazon S3 especificado. Por padrão, esses arquivos são criptografados com chaves gerenciadas pelo Amazon S3 (SSE-S3) com criptografia do lado do servidor como nível básico de criptografia. Você também pode optar por usar chaves KMS gerenciadas pelo cliente (SSE-KMS) para criptografar seus arquivos. Você pode fazer isso usando o console de gerenciamento do Amazon EMR ou por meio do AWS CLI AWS SDK ao criar um EMR Studio.

A criptografia de armazenamento do espaço de trabalho do EMR Studio está disponível em todas as regiões em que o EMR Studio está disponível.

Pré-requisitos

Antes de criptografar o caderno e os arquivos do espaço de trabalho do EMR Studio, você deve AWS Key Management Service usar para criar uma chave simétrica do gerenciador de clientes (CMK) na mesma região do Conta da AWS seu EMR Studio.

Sua política de recursos AWS KMS deve ter as permissões de acesso necessárias para a função de serviço do EMR Studio. Veja a seguir um exemplo de política do IAM que concede permissões mínimas de acesso à criptografia de armazenamento do EMR Studio Workspace:

```
{
    "Sid": "AllowEMRStudioServiceRoleAccess",
    "Effect": "Allow",
    "Principal": {
```

```
        "AWS": "arn:aws:iam::<ACCOUNT_ID>:role/<ROLE_NAME>"
    },
    "Action": [
        "kms:Decrypt",
        "kms:GenerateDataKey",
        "kms:ReEncryptFrom",
        "kms:ReEncryptTo",
        "kms:DescribeKey"
    ],
    "Resource": "*",
    "Condition": {
        "StringEquals": {
            "kms:CallerAccount": "<ACCOUNT_ID>",
            "kms:EncryptionContext:aws:s3:arn": "arn:aws:s3:::<S3_BUCKET_NAME>",
            "kms:ViaService": "s3.<AWS_REGION>.amazonaws.com"
        }
    }
}
```

Sua função de serviço do EMR Studio também deve ter as permissões de acesso para usar sua AWS KMS chave. Veja a seguir um exemplo de política do IAM que concede as permissões mínimas de acesso à criptografia de armazenamento do EMR Studio Workspace:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowEMRStudioWorkspaceStorageEncryptionAccess",
            "Effect": "Allow",
            "Action": [
               "kms:Decrypt",
               "kms:GenerateDataKey",
               "kms:ReEncryptFrom",
               "kms:ReEncryptTo",
               "kms:DescribeKey"
            ],
            "Resource": ["arn:aws:kms:<REGION>:<ACCOUNT_ID>:key/<KEY_IDENTIFIER>"]
        }
    ]
}
```

Configuração

Siga estas etapas para criar um novo EMR Studio que usa criptografia de armazenamento do espaço de trabalho.

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha Estúdios e, em seguida, escolha Criar estúdio.
3. Para a localização do S3 para armazenamento, insira ou escolha um caminho do Amazon S3. Esse é o local do Amazon S3 onde o Amazon EMR armazena cadernos e arquivos do espaço de trabalho.
4. Em Função de serviço, insira ou escolha uma função do IAM. Essa é a função do IAM que o Amazon EMR assume.
5. Escolha Criptografar arquivos do Workspace com sua própria chave. AWS KMS
6. Insira ou escolha uma AWS KMS chave para usar para criptografar cadernos e arquivos do espaço de trabalho no Amazon S3.
7. Escolha Create Studio ou Create Studio e inicie espaços de trabalho.
8. Escolha Criptografar arquivos do Workspace com sua própria chave. AWS KMS
9. Insira ou escolha um AWS KMS para usar para criptografar cadernos e arquivos do espaço de trabalho no Amazon S3.
10. Escolha Salvar alterações.

As etapas a seguir demonstram como atualizar um EMR Studio e configurar a criptografia de armazenamento do espaço de trabalho.

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha um EMR Studio existente na lista e, em seguida, escolha Editar.
3. Escolha Criptografar arquivos do Workspace com sua própria chave. AWS KMS
4. Insira ou escolha um AWS KMS para usar para criptografar cadernos e arquivos do espaço de trabalho no Amazon S3.
5. Escolha Salvar alterações.

Definição de grupos de segurança para controlar o tráfego de rede do EMR Studio

Sobre os grupos de segurança do EMR Studio

O Amazon EMR Studio usa dois grupos de segurança para controlar o tráfego de rede entre os Workspaces no Studio e um cluster anexado do Amazon EMR em execução no Amazon EC2:

- Um grupo de segurança do mecanismo que usa a porta 18888 para se comunicar com um cluster anexado do Amazon EMR em execução no Amazon EC2.
- Um grupo de segurança do Workspace associado aos Workspaces em um Studio. Este grupo de segurança inclui uma regra HTTPS de saída para possibilitar que o Workspace roteie o tráfego para a Internet e deve permitir o tráfego de saída para a Internet na porta 443 para habilitar a vinculação de repositórios Git a um Workspace.

O EMR Studio usa esses grupos de segurança em conjunto com quaisquer grupos de segurança associados a um cluster do EMR anexado a um Workspace.

Você deve criar esses grupos de segurança ao usar a AWS CLI para criar um Studio.

> **Note**
>
> Você pode personalizar os grupos de segurança do EMR Studio com regras adaptadas para seu ambiente, mas deve incluir as regras indicadas nesta página. O grupo de segurança do Workspace não pode permitir tráfego de entrada, e o grupo de segurança do mecanismo deve permitir o tráfego de entrada do grupo de segurança do Workspace.

Uso dos grupos de segurança padrão do EMR Studio

Ao usar o console do Amazon EMR, é possível escolher os grupos de segurança padrão apresentados a seguir. Os grupos de segurança padrão são criados pelo EMR Studio em seu nome e incluem as regras mínimas de entrada e de saída obrigatórias para os Workspaces em um EMR Studio.

- `DefaultEngineSecurityGroup`
- `DefaultWorkspaceSecurityGroupGit` ou `DefaultWorkspaceSecurityGroupWithoutGit`

Pré-requisitos

Para criar os grupos de segurança para o EMR Studio, você precisa de uma Amazon Virtual Private Cloud (VPC) para o Studio. Você escolhe essa VPC ao criar os grupos de segurança. Esta deve ser a mesma VPC que você especificou ao criar o Studio. Se você planeja usar o Amazon EMR no EKS com o EMR Studio, escolha a VPC para os nós de processamento do cluster do Amazon EKS.

Instruções

Siga as instruções em Criar um grupo de segurança no Guia do usuário do Amazon EC2 para instâncias do Linux para criar um grupo de segurança do mecanismo e um grupo de segurança do Workspace em sua VPC. Os grupos de segurança devem incluir as regras resumidas nas tabelas a seguir.

Ao criar grupos de segurança para o EMR Studio, anote os IDs de ambos os grupos. Você especifica cada grupo de segurança usando o ID ao criar um Studio.

Grupo de segurança do mecanismo

O EMR Studio usa a porta 18888 para se comunicar com um cluster anexado.

Regras de entrada

Type	Protocolo	Port	Destino	Descrição
TCP	TCP	18888	Seu grupo de segurança do Workspace do EMR Studio.	Permite o tráfego de quaisquer recursos no grupo de segurança do Workspace para o EMR Studio.

Grupo de segurança do Workspace

Este grupo de segurança está associado aos Workspaces em um EMR Studio.

Regras de saída

Type	Protocolo	Port	Destino	Descrição
TCP	TCP	18888	Seu grupo de segurança do	Permite o tráfego para quaisquer recursos no

Type	Protocolo	Port	Destino	Descrição
			mecanismo do EMR Studio.	grupo de segurança do mecanismo para o EMR Studio.
HTTPS	TCP	443	0.0.0.0/0	Permite que o tráfego da Internet vincule repositór ios Git hospedado s publicamente aos Workspaces.

Criação de modelos do AWS CloudFormation para o Amazon EMR Studio

Sobre os modelos de cluster do EMR Studio

Você pode criar AWS CloudFormation modelos para ajudar os usuários do EMR Studio a lançar novos clusters do Amazon EMR em um espaço de trabalho. CloudFormation modelos são arquivos de texto formatados em JSON ou YAML. Em um modelo, você descreve uma pilha de AWS recursos e explica CloudFormation como provisionar esses recursos para você. Para o EMR Studio, você pode criar um ou mais modelos que descrevem um cluster do Amazon EMR.

Você organiza os modelos no AWS Service Catalog. O AWS Service Catalog permite criar e gerenciar serviços de TI comumente implantados, chamados de produtos, na AWS. Você coleta os modelos como produtos em um portfólio que compartilha com os seus usuários do EMR Studio. Após criar modelos de cluster, os usuários do Studio poderão iniciar um novo cluster para um Workspace com um de seus modelos. Os usuários devem ter permissão para criar novos clusters usando os modelos. Você pode definir as permissões de usuário em suas políticas de permissões do EMR Studio.

Para saber mais sobre CloudFormation modelos, consulte Modelos no Guia do AWS CloudFormation usuário. Para obter mais informações sobre o AWS Service Catalog, consulte O que é o AWS Service Catalog.

O vídeo a seguir demonstra como configurar os modelos de cluster no AWS Service Catalog para o EMR Studio. Você também pode aprender mais na publicação Build a self-service environment for each line of business using Amazon EMR and Service Catalog do blog.

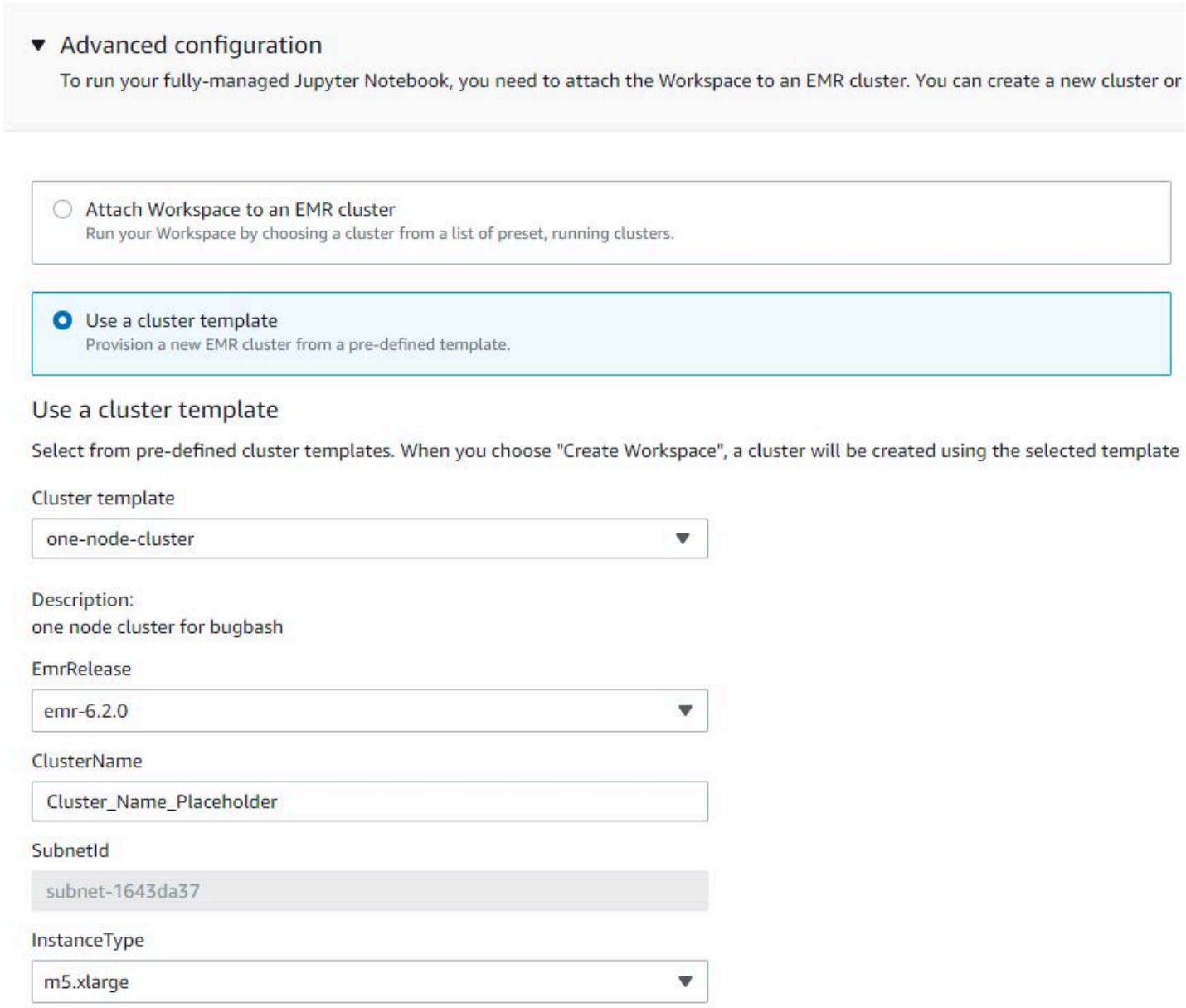
Parâmetros de modelo opcionais

Você pode incluir opções adicionais na seção `Parameters` do seu modelo. Os parâmetros permitem que os usuários do Studio insiram ou selecionem valores personalizados para um cluster. Por exemplo, é possível adicionar um parâmetro que permita aos usuários selecionar uma versão específica do Amazon EMR. Para obter mais informações, consulte Parâmetros no Guia do usuário do AWS CloudFormation.

O exemplo da seção `Parameters` a seguir define parâmetros de entrada adicionais, como o `ClusterName`, a versão de `EmrRelease` e o `ClusterInstanceType`.

```
Parameters:
  ClusterName:
    Type: "String"
    Default: "Cluster_Name_Placeholder"
  EmrRelease:
    Type: "String"
    Default: "emr-6.2.0"
    AllowedValues:
    - "emr-6.2.0"
    - "emr-5.32.0"
  ClusterInstanceType:
    Type: "String"
    Default: "m5.xlarge"
    AllowedValues:
    - "m5.xlarge"
    - "m5.2xlarge"
```

Ao adicionar parâmetros, os usuários do Studio visualizam opções adicionais de formulário após selecionar um modelo de cluster. A imagem a seguir mostra opções adicionais de formulário para a EmrReleaseversão ClusterName, InstanceTypee.



Pré-requisitos

Antes de criar um modelo de cluster, certifique-se de ter permissões do IAM para acessar a visualização do console do administrador do Service Catalog. Você também precisa das permissões do IAM que são obrigatórias para a execução de tarefas administrativas do Service Catalog. Para obter mais informações, consulte Grant permissions to Service Catalog administrators.

Instruções

Criar modelos de cluster do EMR usando o Service Catalog

1. Crie um ou mais CloudFormation modelos. O local em que você armazenará os modelos fica a seu critério. Como os modelos são arquivos de texto formatados, você pode fazer upload deles no Amazon S3 ou mantê-los em seu sistema de arquivos local. Para saber mais sobre CloudFormation modelos, consulte Modelos no Guia do AWS CloudFormation usuário.

   Use as regras apresentadas a seguir para nomear os modelos ou comparar os nomes em relação ao padrão `[a-zA-Z0-9][a-zA-Z0-9._-]*`.

   - Os nomes dos modelos devem começar com uma letra ou com um número.
   - Os nomes dos modelos podem consistir somente em letras, números, pontos (.), sublinhados (_) e hifens (-).

   Cada modelo de cluster criado deve incluir as seguintes opções:

   Parâmetros de entrada

   - ClusterName — Um nome para o cluster para ajudar os usuários a identificá-lo após o provisionamento.

   Saída

   - `ClusterId`: o ID do cluster do EMR provisionado recentemente.

   A seguir, é apresentado um exemplo de modelo do AWS CloudFormation no formato YAML para um cluster com dois nós. O modelo de exemplo inclui as opções de modelo obrigatórias e define parâmetros de entrada adicionais para `EmrRelease` e `ClusterInstanceType`.

```
awsTemplateFormatVersion: 2010-09-09

Parameters:
  ClusterName:
    Type: "String"
    Default: "Example_Two_Node_Cluster"
  EmrRelease:
    Type: "String"
```

```
    Default: "emr-6.2.0"
    AllowedValues:
    - "emr-6.2.0"
    - "emr-5.32.0"
  ClusterInstanceType:
    Type: "String"
    Default: "m5.xlarge"
    AllowedValues:
    - "m5.xlarge"
    - "m5.2xlarge"

Resources:
  EmrCluster:
    Type: AWS::EMR::Cluster
    Properties:
      Applications:
      - Name: Spark
      - Name: Livy
      - Name: JupyterEnterpriseGateway
      - Name: Hive
      EbsRootVolumeSize: '10'
      Name: !Ref ClusterName
      JobFlowRole: EMR_EC2_DefaultRole
      ServiceRole: EMR_DefaultRole_V2
      ReleaseLabel: !Ref EmrRelease
      VisibleToAllUsers: true
      LogUri:
        Fn::Sub: 's3://aws-logs-${AWS::AccountId}-${AWS::Region}/elasticmapreduce/'
      Instances:
        TerminationProtected: false
        Ec2SubnetId: 'subnet-ab12345c'
        MasterInstanceGroup:
          InstanceCount: 1
          InstanceType: !Ref ClusterInstanceType
        CoreInstanceGroup:
          InstanceCount: 1
          InstanceType: !Ref ClusterInstanceType
          Market: ON_DEMAND
          Name: Core

Outputs:
  ClusterId:
    Value:
      Ref: EmrCluster
```

```
    Description: The ID of the  EMR cluster
```

2. Crie um portfólio para seus modelos de cluster na mesma conta da AWS de seu Studio.

   a. Abra o AWS Service Catalog console em https://console.aws.amazon.com/servicecatalog/.

   b. Escolha Portfólios no menu de navegação à esquerda.

   c. Insira as informações solicitadas na página Criar portfólio.

   d. Escolha Criar. O AWS Service Catalog cria o portfólio e exibe os detalhes do portfólio.

3. Use as etapas a seguir para adicionar seus modelos de cluster como produtos do AWS Service Catalog.

   a. Navegue até a página Produtos em Administração no console de gerenciamento do AWS Service Catalog.

   b. Escolha Fazer upload de novo produto.

   c. Insira um Nome do produto e um Proprietário.

   d. Especifique seu arquivo de modelo em Detalhes da versão.

   e. Escolha Analisar para analisar as configurações do produto e, em seguida, selecione Criar produto.

4. Conclua as etapas a seguir para adicionar os produtos ao seu portfólio.

   a. Navegue até a página Produtos no console de gerenciamento do AWS Service Catalog.

   b. Escolha seu produto, selecione Ações e, em seguida, clique em Adicionar produto ao portfólio.

   c. Escolha seu portfólio e, em seguida, escolha Adicionar produto ao portfólio.

5. Crie uma restrição de inicialização para seus produtos. Uma restrição de inicialização corresponde a um perfil do IAM que especifica as permissões de usuário para iniciar um produto. Você pode personalizar suas restrições de lançamento, mas deve permitir permissões de uso, CloudFormation Amazon EMR e. AWS Service Catalog Para obter mais informações e instruções, consulte Service Catalog launch constraints.

6. Aplique a restrição de inicialização a cada produto do seu portfólio. Você deve aplicar a restrição de inicialização a cada produto individualmente.

   a. Selecione seu portfólio na página Portfólios no console de gerenciamento do AWS Service Catalog.

   b. Escolha a guia Constraints (Restrições) e Create constraint (Criar restrição).

c. Escolha seu produto e selecione Inicialização em Tipo de restrição. Escolha Continuar.

d. Selecione seu perfil de restrição de inicialização na seção Restrição de inicialização e, em seguida, escolha Criar.

7. Conceda acesso ao seu portfólio.

a. Selecione seu portfólio na página Portfólios no console de gerenciamento do AWS Service Catalog.

b. Expanda a guia Grupos, perfis e usuários e escolha Adicionar grupos, perfis e usuários.

c. Pesquise seu perfil do IAM do EMR Studio na guia Perfis, selecione o perfil e escolha Adicionar acesso.

Se você usar...	Conceda acesso a...
Autenticação do IAM	Seus usuários nativos
Federação do IAM	Seu perfil do IAM para a federação
Federação do Centro de Identidade do IAM	Seu perfil de usuário do EMR Studio

Estabelecimento de acesso e de permissões para repositórios baseados em Git

O EMR Studio oferece suporte aos seguintes serviços baseados em Git:

- AWS CodeCommit
- GitHub
- Bitbucket
- GitLab

Para permitir que os usuários do EMR Studio associem um repositório Git a um Workspace, configure os requisitos de acesso e as permissões apresentados a seguir. Você também pode configurar repositórios baseados em Git hospedados em uma rede privada ao seguir as instruções em Configuração de um repositório Git hospedado de forma privada para o EMR Studio.

Cluster com acesso à Internet

Os clusters do Amazon EMR em execução no Amazon EC2 e os clusters do Amazon EMR no EKS anexados aos Workspaces do Studio devem estar em uma sub-rede privada que usa um gateway de conversão de endereços de rede (NAT) ou devem ser capazes de acessar a Internet usando um gateway privado virtual. Para ter mais informações, consulte Opções da Amazon VPC.

Os grupos de segurança usados com o EMR Studio também devem incluir uma regra de saída que permita que os Workspaces roteiem o tráfego para a Internet usando um cluster do EMR anexado. Para ter mais informações, consulte Definição de grupos de segurança para controlar o tráfego de rede do EMR Studio.

> **Important**
>
> Se a interface de rede estiver em uma sub-rede pública, não será possível ter uma comunicação com a Internet através de um gateway da Internet (IGW).

Permissões para AWS Secrets Manager

Para permitir que os usuários do EMR Studio acessem repositórios Git com segredos armazenados no AWS Secrets Manager, adicione uma política de permissões ao perfil de serviço do EMR Studio que permite a operação `secretsmanager:GetSecretValue`.

Para obter informações sobre como vincular repositórios baseados em Git a Workspaces, consulte Vinculação de repositórios baseados em Git a um Workspace do EMR Studio.

Configuração de um repositório Git hospedado de forma privada para o EMR Studio

Use as instruções a seguir para configurar repositórios hospedados de forma privada para o Amazon EMR Studio. Forneça um arquivo de configuração com informações sobre os servidores DNS e Git. O EMR Studio usa essas informações para configurar Workspaces que podem rotear o tráfego para os repositórios com hospedagem própria.

> **Note**
>
> Se você configurar o `DnsServerIpV4`, o EMR Studio usará o servidor DNS para resolver o `GitServerDnsName` e o endpoint do Amazon EMR, como `elasticmapreduce.us-`

`east-1.amazonaws.com`. Para configurar um endpoint para o Amazon EMR, conecte-se ao endpoint através da VPC que você está usando com o Studio. Isso garante que o endpoint do Amazon EMR seja resolvido para um IP privado. Para ter mais informações, consulte Conectar-se ao Amazon EMR usando um endpoint da VPC de interface.

Pré-requisitos

Antes de configurar um repositório Git hospedado de forma privada para o EMR Studio, você precisa de um local de armazenamento do Amazon S3 no qual o EMR Studio possa fazer backup dos Workspaces e dos arquivos de cadernos no Studio. Use o mesmo bucket do S3 especificado ao criar um Studio.

Configurar um ou mais repositórios Git hospedados de forma privada para o EMR Studio

1. Crie um arquivo de configuração usando o modelo apresentado a seguir. Inclua os seguintes valores para cada servidor Git que deseja especificar em sua configuração:

   - **`DnsServerIpV4`**: o endereço IPv4 do seu servidor DNS. Se você fornecer valores para `DnsServerIpV4` e `GitServerIpV4List`, o valor para `DnsServerIpV4` terá precedência e o EMR Studio usará `DnsServerIpV4` para resolver seu `GitServerDnsName`.

     > **Note**
     >
     > Para usar repositórios Git hospedados de forma privada, seu servidor DNS deve permitir o acesso de entrada do EMR Studio. Recomendamos proteger o servidor DNS contra outros acessos não autorizados.

   - **`GitServerDnsName`**: o nome DNS do seu servidor Git. Por exemplo, `"git.example.com"`.
   - **`GitServerIpV4List`**: uma lista de endereços IPv4 que pertencem aos seus servidores Git.

```
[
    {
        "Type": "PrivatelyHostedGitConfig",
        "Value": [
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<enterprise.git.com>",
                "GitServerIpV4List": [
```

```
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            },
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<git.example.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            }
        ]
    }
]
```

2. Salve seu arquivo de configuração como `configuration.json`.
3. Faça o upload do arquivo de configuração no local de armazenamento do Amazon S3 em uma pasta chamada `life-cycle-configuration`. Por exemplo, se o local padrão do S3 for `s3://DOC-EXAMPLE-BUCKET/studios`, seu arquivo de configuração estará em `s3://DOC-EXAMPLE-BUCKET/studios/`**`life-cycle-configuration/configuration.json`**.

> **Important**
>
> Recomendamos que você restrinja o acesso à sua pasta `life-cycle-configuration` para os administradores do Studio e para o perfil de serviço do EMR Studio, e que proteja o arquivo `configuration.json` contra acessos não autorizados. Para obter instruções, consulte Controlar o acesso a um bucket com políticas de usuário ou Práticas recomendadas de segurança para o Amazon S3.

Para obter instruções sobre como fazer o upload, consulte Criar uma pasta e Fazer upload de objetos no Guia do usuário do Amazon Simple Storage Service. Para aplicar sua configuração a um Workspace, feche e reinicie o Workspace após fazer o upload do arquivo de configuração para o Amazon S3.

Otimização de trabalhos do Spark no EMR Studio

Ao executar um trabalho do Spark usando o EMR Studio, há algumas etapas que você pode realizar para ajudar a garantir que você está otimizando os recursos do cluster do Amazon EMR.

Prolongamento da sessão do Livy

Se você usar o Apache Livy em conjunto com o Spark no cluster do Amazon EMR, recomendamos aumentar o tempo limite da sessão do Livy seguindo um destes procedimentos:

- Ao criar um cluster do Amazon EMR, defina essa classificação de configuração no campo Inserir configuração.

```
[
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.server.session.timeout": "8h"
        }
    }
]
```

- Para um cluster EMR já em execução, conecte-se ao cluster usando `ssh` e defina a classificação de configuração `livy-conf` em `/etc/livy/conf/livy.conf`.

```
[
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.server.session.timeout": "8h"
        }
    }
]
```

  Pode ser necessário reiniciar o Livy após alterar a configuração.

- Se você não deseja que sua sessão do Livy expire, defina a propriedade `livy.server.session.timeout-check` como `false` em `/etc/livy/conf/livy.conf`.

Execução do Spark no modo de cluster

No modo de cluster, o driver do Spark é executado em um nó central em vez de no nó primário, melhorando a utilização de recursos no nó primário.

Para executar sua aplicação do Spark no modo de cluster em vez de no modo de cliente padrão, escolha o modo de Cluster ao definir o Modo de implantação ao configurar a etapa do Spark em seu novo cluster do Amazon EMR. Para obter mais informações, consulte Cluster mode overview na documentação do Apache Spark.

Aumento da memória do driver do Spark

Para aumentar a memória do driver do Spark, configure a sessão do Spark usando o comando mágico `%%configure` em seu Caderno do EMR, como no exemplo apresentado a seguir.

```
%%configure -f
{"driverMemory": "6000M"}
```

Uso de um Amazon EMR Studio

Esta seção contém tópicos que ajudam a configurar e interagir com um Amazon EMR Studio.

O vídeo a seguir aborda informações práticas, como, por exemplo, como criar um novo Workspace e como iniciar um novo cluster do Amazon EMR com um modelo de cluster. Além disso, o vídeo executa um caderno de exemplo.

Esta seção inclui os seguintes tópicos para ajudar você a trabalhar em um EMR Studio:

- Compreensão das noções básicas do Workspace
- Configuração da colaboração no Workspace
- Execução de um Workspace do EMR Studio com um perfil de runtime
- Execução de cadernos do Workspace de forma programática
- Navegar pelos dados com o SQL Explorer
- Anexar uma computação a um Workspace do EMR Studio
- Vinculação de repositórios baseados em Git a um Workspace do EMR Studio
- Uso do editor do SQL com Amazon Athena no EMR Studio

- CodeWhisperer Integração da Amazon com o EMR Studio Workspaces
- Depuração de aplicações e trabalhos com o EMR Studio
- Instalação de kernels e de bibliotecas em um Workspace do EMR Studio
- Aprimoramento de kernels com comandos magic
- Use cadernos em várias linguagens com kernels do Spark

Compreensão das noções básicas do Workspace

Ao usar um EMR Studio, é possível criar e configurar diferentes Workspaces para organizar e executar cadernos. Esta seção aborda como criar e trabalhar com Workspaces. Para obter uma visão geral conceitual, consulte Workspaces na página Como o Amazon EMR Studio funciona.

Esta seção aborda os seguintes tópicos para ajudar você a usar os Workspaces do EMR Studio:

- Criação de um Workspace do EMR Studio
- Inicialização de um WorkSpace
- Compreensão da interface do usuário do Workspace
- Exploração de exemplos de cadernos
- Salvamento de conteúdo do Workspace
- Exclusão de um Workspace e de arquivos de cadernos
- Compreensão do status do Workspace
- Resolução de problemas de conectividade do Workspace

Criação de um Workspace do EMR Studio

Você pode criar Workspaces do EMR Studio para executar códigos de cadernos usando a interface do EMR Studio.

Criar um Workspace em um EMR Studio

1. Faça login no seu EMR Studio.
2. Escolha Criar um Workspace.
3. Insira um Nome do WorkSpace e uma Descrição. Nomear um Workspace ajuda a identificá-lo na página Workspaces.

4. Se desejar trabalhar com outros usuários do Studio neste Workspace em tempo real, habilite a colaboração no Workspace. Você pode configurar colaboradores depois de iniciar o Workspace.
5. Se desejar anexar um cluster a um Workspace, expanda a seção Configuração avançada. Você pode anexar um cluster posteriormente, se preferir. Para ter mais informações, consulte Anexar uma computação a um Workspace do EMR Studio.

   > **Note**
   >
   > Para provisionar um novo cluster, você precisa receber permissões de acesso por parte do administrador.

   Escolha uma das opções de cluster para o Workspace e anexe o cluster. Para obter mais informações sobre o provisionamento de um cluster ao criar um Workspace, consulte Criar e anexar um novo cluster do EMR a um Workspace do EMR Studio.
6. Escolha Criar um Workspace no canto inferior direito da página.

Após a criação de um Workspace, o EMR Studio abrirá a página Workspaces. Você visualizará um banner verde representando o êxito na parte superior da página e poderá encontrar o Workspace recém-criado na lista.

Por padrão, um Workspace é compartilhado e pode ser visualizado por todos os usuários do Studio. No entanto, somente um usuário pode abrir e trabalhar em um Workspace por vez. Para trabalhar simultaneamente com outros usuários, é possível realizar a Configuração da colaboração no Workspace.

Inicialização de um WorkSpace

Para começar a trabalhar com arquivos de cadernos, inicie um Workspace para acessar o editor de caderno. A página Workspaces em um Studio lista todos os Workspaces aos quais você tem acesso com detalhes, incluindo Nome, Status, Horário de criação e Última modificação.

> **Note**
>
> Se você tinha Cadernos do EMR no console antigo do Amazon EMR, poderá localizá-los no novo console como Workspaces do EMR Studio. Os usuários de Cadernos do EMR precisam de permissões adicionais de perfil do IAM para acessar ou criar Workspaces. Se você criou recentemente um caderno no console antigo, talvez seja necessário atualizar a

lista Workspaces para visualizá-lo no novo console. Para obter mais informações sobre a transição, consulte Os notebooks Amazon EMR estão disponíveis como Amazon EMR Studio Workspaces no console. e Console do Amazon EMR.

Iniciar um Workspace para edição e execução de cadernos

1. Na página Workspaces do seu Studio, localize o Workspace. Você pode filtrar a lista por palavra-chave ou por valor de coluna.
2. Escolha o nome do Workspace para iniciá-lo em uma nova guia do navegador. Pode demorar alguns minutos para o Workspace abrir, se ele estiver Ocioso. Como alternativa, selecione a linha para o Workspace e, em seguida, escolha Iniciar o Workspace. É possível escolher entre as seguintes opções de inicialização:

   - Início rápido: inicie rapidamente seu Workspace com as opções padrão. Escolha Início rápido se quiser anexar clusters ao espaço de trabalho em JupyterLab.
   - Início com opções: inicie seu Workspace com opções personalizadas. Você pode optar por iniciar no Jupyter ou JupyterLab anexar seu espaço de trabalho a um cluster do EMR e selecionar seus grupos de segurança.

> **Note**
>
> Somente um usuário pode abrir e trabalhar em um Workspace por vez. Se você selecionar um Workspace que já esteja em uso, o EMR Studio exibirá uma notificação quando você tentar abri-lo. A coluna Usuário na página Workspaces mostra o usuário que está trabalhando no Workspace.

Compreensão da interface do usuário do Workspace

A interface do usuário do EMR Studio Workspace é baseada na JupyterLabinterface com guias indicadas por ícones na barra lateral esquerda. Ao colocar o cursor do mouse sobre um ícone, você visualizará uma descrição que mostra o nome da guia. Escolha as guias na barra lateral à esquerda para acessar os painéis apresentados a seguir.

- Navegador de arquivos: exibe os arquivos e diretórios no Workspace, bem como os arquivos e diretórios de repositórios Git vinculados.

- Kernels e terminais em execução: lista todos os kernels e os terminais em execução no Workspace. Para obter mais informações, consulte Gerenciando kernels e terminais na documentação oficial JupyterLab .
- Git: fornece uma interface gráfica do usuário para a execução de comandos nos repositórios Git anexados ao Workspace. Esse painel é uma JupyterLab extensão chamada jupyterlab-git. Para obter mais informações, consulte jupyterlab-git.
- Clusters do EMR: permitem anexar ou desanexar um cluster do Workspace para executar o código do caderno. O painel de configuração do cluster do EMR também fornece opções de configurações avançadas para ajudar você a criar e anexar um novo cluster ao Workspace. Para ter mais informações, consulte Criar e anexar um novo cluster do EMR a um Workspace do EMR Studio.
- Repositório Git do Amazon EMR: ajuda você a vincular o Workspace a até três repositórios Git. Para obter detalhes e instruções, consulte Vinculação de repositórios baseados em Git a um Workspace do EMR Studio.
- Exemplos de cadernos: fornece uma lista de exemplos de cadernos que você pode salvar no Workspace. Você também pode acessar os exemplos ao escolher Exemplos de cadernos na página Inicializador do Workspace.
- Comandos — Oferece uma forma orientada pelo teclado de pesquisar e executar comandos. JupyterLab Para obter mais informações, consulte a página da paleta Command na JupyterLab documentação.
- Ferramentas do caderno: permite selecionar e definir opções, como o tipo de deslizamento da célula e os metadados. A opção Ferramentas do caderno aparece na barra lateral à esquerda depois que você abre um arquivo de caderno.
- Guias abertas: lista os documentos e as atividades abertos na área de trabalho principal para que você possa acessar uma guia aberta. Para obter mais informações, consulte a página do modo Tabulações e documento único na JupyterLab documentação.
- Colaboração: permite habilitar ou desabilitar a colaboração no Workspace e gerenciar colaboradores. Para visualizar o painel Colaboração, você deve ter as permissões necessárias. Para obter mais informações, consulte Definição de propriedade para colaboração no Workspace.

Exploração de exemplos de cadernos

Cada Workspace do EMR Studio inclui um conjunto de exemplos de cadernos que você pode usar para explorar os recursos do EMR Studio. Para editar ou executar um exemplo de caderno, você pode salvá-lo no Workspace.

Salvar um exemplo de caderno em um Workspace

1. Na barra lateral à esquerda, escolha a guia Exemplos de cadernos para abrir o painel Exemplos de cadernos. Você também pode acessar os exemplos ao escolher Exemplos de cadernos na página Inicializador do Workspace.
2. Escolha um exemplo de caderno para visualizá-lo previamente na área de trabalho principal. O exemplo é somente para leitura.
3. Para salvar o exemplo de caderno no Workspace, escolha Salvar no Workspace. O EMR Studio salva o exemplo em seu diretório inicial. Depois de salvar um exemplo de caderno no Workspace, você poderá renomeá-lo, editá-lo e executá-lo.

Para obter mais informações sobre os exemplos de notebooks, consulte o repositório de exemplos GitHub de notebooks do EMR Studio.

Salvamento de conteúdo do Workspace

Quando você trabalha no editor de caderno de um Workspace, o EMR Studio salva o conteúdo das células de cadernos e a saída para você no local do Amazon S3 associado ao Studio. Este processo de backup preserva o trabalho entre as sessões.

Você também pode salvar um caderno ao pressionar CTRL+S na guia do caderno que está aberta ou ao usar uma das opções de salvamento em Arquivo.

Outra maneira de fazer backup dos arquivos de cadernos em um Workspace é associar o Workspace a um repositório baseado em Git e sincronizar suas alterações com o repositório remoto. Isso também permite salvar e compartilhar cadernos com membros da equipe que usam um Workspace ou um Studio diferente. Para obter instruções, consulte Vinculação de repositórios baseados em Git a um Workspace do EMR Studio.

Exclusão de um Workspace e de arquivos de cadernos

Ao excluir um arquivo de caderno de um Workspace do EMR Studio, você exclui o arquivo do Navegador de arquivos e o EMR Studio remove a cópia de backup no Amazon S3. Você não precisa tomar nenhuma medida adicional para evitar cobranças de armazenamento ao excluir um arquivo de um Workspace.

Quando você exclui um Workspace inteiro, seus arquivos e suas pastas de cadernos permanecerão no local de armazenamento do Amazon S3. Os arquivos continuam a acumular cobranças de

armazenamento. Para evitar cobranças de armazenamento, remova todos os arquivos e as pastas de backup associados ao Workspace excluído do Amazon S3.

Excluir um arquivo de cadernos de um Workspace do EMR Studio

1. Selecione o painel Navegador de arquivos na barra lateral à esquerda do Workspace.
2. Selecione o arquivo ou a pasta que deseja excluir. Clique com o botão direito do mouse na sua seleção e escolha Excluir. O arquivo desaparecerá da lista. O EMR Studio removerá o arquivo ou a pasta do Amazon S3 para você.

From the Workspace UI

Exclusão de um Workspace e dos arquivos de backup associados do EMR Studio

1. Faça login no EMR Studio com o URL de acesso do Studio e escolha Workspaces no painel de navegação à esquerda.
2. Localize seu Workspace na lista e, em seguida, marque a caixa de seleção ao lado do nome. É possível selecionar vários Workspaces a serem excluídos ao mesmo tempo.
3. Escolha Excluir no canto superior direito da lista Workspaces e confirme que deseja excluir os Workspaces selecionados. Escolha Delete para confirmar.
4. Se você desejar remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3, siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service. Se não foi você quem criou o Studio, consulte o administrador do Studio para determinar o local de backup do Amazon S3 para o Workspace excluído.

From the Workspaces list

Exclusão de um Workspace e dos arquivos de backup associados da lista Workspaces

1. Navegue até a lista Workspaces no console.
2. Selecione o Workspace que deseja excluir da lista e, em seguida, escolha Ações.
3. Escolha Excluir.
4. Se você desejar remover os arquivos de cadernos associados ao Workspace excluído do Amazon S3, siga as instruções para Excluir objetos no Guia do usuário do console do Amazon Simple Storage Service. Se não foi você quem criou o Studio, consulte o

administrador do Studio para determinar o local de backup do Amazon S3 para o Workspace excluído.

Compreensão do status do Workspace

Após a criação de um Workspace do EMR Studio, ele aparecerá como uma linha na lista Workspaces em seu Studio com o nome, o status, o horário de criação e o carimbo de data/hora da última modificação. A tabela a seguir descreve os status do Workspace.

Status	Descrição
Starting	O Workspace está sendo preparado, mas ainda não está pronto para uso. Não é possível abrir um Workspace quando o status for Iniciando.
Ready	É possível abrir o Workspace para usar o editor de caderno, mas você deve anexar o Workspace a um cluster do EMR antes de executar o código do caderno.
Anexando	O Workspace está sendo anexado a um cluster.
Attached	O Workspace está anexado a um cluster do EMR e pronto para que você escreva e execute o código do caderno. Se o status de um Workspace não for Anexado, você deverá anexá-lo a um cluster antes de executar o código do caderno.
Ocioso	O Workspace foi interrompido. Para reativar um Workspace ocioso, selecione-o na lista Workspaces. O status é alterado de Ocioso para Iniciando e, em seguida, para Pronto quando você seleciona o Workspace.

Status	Descrição
Stopping	O Workspace está sendo encerrado e será definido como Ocioso. Quando você interromp e um Workspace, ele encerra todos os kernels de cadernos correspondentes. O EMR Studio interrompe cadernos que estão ociosos há muito tempo.
Deleting	Quando você exclui um Workspace, o EMR Studio o marca para exclusão e inicia o processo de exclusão. Após a conclusão do processo de exclusão, o Workspace desaparec erá da lista. Quando você exclui um Workspace , os arquivos de cadernos permanecerão no local de armazenamento do Amazon S3.

Resolução de problemas de conectividade do Workspace

Para resolver problemas de conectividade do Workspace, você pode interromper e reiniciar um Workspace. Quando você reinicia um Workspace, o EMR Studio inicia o Workspace em uma zona de disponibilidade diferente ou em uma sub-rede diferente associada ao seu Studio.

Interromper e reiniciar um Workspace do EMR Studio

1. Feche o Workspace no seu navegador.
2. Navegue até a lista Workspace no console.
3. Selecione seu Workspace na lista e escolha Ações.
4. Escolha Interromper e aguarde até que o status do Workspace seja alterado de Interrompendo para Ocioso.
5. Escolha Ações novamente e, em seguida, selecione Iniciar para reiniciar o Workspace.
6. Aguarde até que o status do Workspace seja alterado de Iniciando para Pronto e, em seguida, escolha o nome do Workspace para abri-lo novamente em uma nova guia do navegador.

Configuração da colaboração no Workspace

A colaboração no Workspace permite escrever e executar códigos de cadernos simultaneamente com outros membros da sua equipe. Ao trabalhar no mesmo arquivo de caderno, você visualizará as alterações à medida que seus colaboradores as fizerem. É possível habilitar a colaboração ao criar um Workspace, ou habilitar e desabilitar a colaboração em um Workspace existente.

> **Note**
>
> A colaboração do Workspace com o EMR Studio não é compatível com aplicações interativas do EMR Sem Servidor ou se a propagação de identidade confiável estiver habilitada.

Pré-requisitos

Antes de configurar a colaboração para um Workspace, certifique-se de concluir as seguintes tarefas:

- Verifique se o administrador do EMR Studio concedeu as permissões necessárias. Por exemplo, a instrução apresentada a seguir permite que um usuário configure a colaboração para qualquer Workspace com a chave de etiqueta `creatorUserId` cujo valor corresponde ao ID do usuário (indicado pela variável de política `aws:userId`).

```
{
    "Sid": "UserRolePermissionsForCollaboration",
    "Action": [
        "elasticmapreduce:UpdateEditor",
        "elasticmapreduce:PutWorkspaceAccess",
        "elasticmapreduce:DeleteWorkspaceAccess",
        "elasticmapreduce:ListWorkspaceAccessIdentities"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Condition": {
        "StringEquals": {
            "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userid}"
        }
    }
}
```

- Certifique-se de que o perfil de serviço associado ao EMR Studio tenha as permissões obrigatórias para habilitar e configurar a colaboração no Workspace, como no exemplo de instrução apresentado a seguir.

```
{
    "Sid": "AllowWorkspaceCollaboration",
    "Effect": "Allow",
    "Action": [
        "iam:GetUser",
        "iam:GetRole",
        "iam:ListUsers",
        "iam:ListRoles",
        "sso:GetManagedApplicationInstance",
        "sso-directory:SearchUsers"
    ],
    "Resource": "*"
}
```

  Para ter mais informações, consulte Crie um perfil de serviço do EMR Studio.

Habilitar a colaboração no Workspace e adicionar colaboradores

1. No seu Workspace, escolha o ícone Colaboração na tela do Inicializador ou na parte inferior do painel à esquerda.

   > **Note**
   >
   > Não será possível visualizar o painel Colaboração, a menos que o administrador do Studio tenha concedido a você permissão para configurar a colaboração para o Workspace. Para obter mais informações, consulte Definição de propriedade para colaboração no Workspace.

2. Certifique-se de que o botão de alternância Permitir a colaboração no Workspace esteja na posição ativada. Ao habilitar a colaboração, somente você e os colaboradores adicionados poderão visualizar o Workspace na lista da página Workspaces do Studio.
3. Insira um Nome do colaborador. Seu Workspace pode ter, no máximo, cinco colaboradores, incluindo você. Um colaborador pode ser qualquer usuário com acesso ao seu EMR Studio. Se você não inserir um colaborador, o Workspace será considerado um Workspace privado acessível somente por você.

A tabela a seguir especifica os valores aplicáveis a serem inseridos para colaboradores com base no tipo de identidade do proprietário.

> **Note**
>
> Um proprietário pode convidar somente colaboradores com o mesmo tipo de identidade. Por exemplo, um usuário pode adicionar somente outros usuários, e um usuário do Centro de Identidade do IAM pode adicionar somente outros usuários do Centro de Identidade do IAM.

Modo de autenticação	Valor a ser inserido para o Nome do colaborador
Autenticação do IAM	Um nome de usuário. Este é o nome que um usuário visualiza quando faz login no AWS Management Console.
Federação do IAM	O nome de um perfil do IAM e um nome de sessão opcional. Para adicionar todos os usuários federados que assumem o mesmo perfil do IAM, especifique o nome de um perfil do IAM para a federação. Para adicionar um único usuário como colaborador, especifique um perfil e um nome de sessão. Por exemplo, `MyRoleName:MySessionName` .
SSO	Um nome de usuário do Centro de Identidade do IAM, como `user@example.com.` .

4. Escolha Adicionar. Agora, o colaborador poderá visualizar o Workspace em sua página Workspaces do EMR Studio e iniciar o Workspace para usá-lo em tempo real com você.

> **Note**
>
> Se você desabilitar a colaboração do Workspace, o Workspace retornará ao estado compartilhado e poderá ser visualizado por todos os usuários do Studio. No estado compartilhado, somente um usuário do Studio poderá abrir e trabalhar no Workspace por vez.

Execução de um Workspace do EMR Studio com um perfil de runtime

> **Note**
>
> A funcionalidade de função de runtime descrita nesta página se aplica somente ao Amazon EMR executado no Amazon EC2 e não se refere à funcionalidade de função de runtime em aplicações interativas do EMR Serverless. Para saber mais sobre como usar funções de runtime no EMR Serverless, consulte Funções de runtime de trabalho no Amazon Guia do usuário do Amazon EMR Serverless.

Uma função de tempo de execução é uma função AWS Identity and Access Management (IAM) que você pode especificar ao enviar um trabalho ou uma consulta para um cluster do Amazon EMR. O trabalho ou consulta que você envia ao seu cluster do EMR usa a função de tempo de execução para acessar AWS recursos, como objetos no Amazon S3.

Ao anexar um espaço de trabalho do EMR Studio a um cluster do EMR que usa o Amazon EMR 6.11 ou superior, você pode selecionar uma função de tempo de execução para o trabalho ou consulta que você envia para uso quando ele acessa recursos. AWS No entanto, se o cluster do EMR não suportar funções de tempo de execução, o cluster do EMR não assumirá a função ao acessar os recursos. AWS

Antes de usar um perfil de runtime com um Workspace do Amazon EMR Studio, um administrador deve configurar as permissões de usuário para que o usuário do Studio possa chamar a API `elasticmapreduce:GetClusterSessionCredentials` no perfil de runtime. Em seguida, inicie um novo cluster com um perfil de runtime que você possa usar com o Workspace do Amazon EMR Studio.

Nesta página

- Inicialização de um novo cluster com um perfil de runtime
- Uso do cluster do EMR com um perfil de runtime no Workspaces
- Considerações

Configuração de permissões de usuários para o perfil de runtime

Configure as permissões de usuários para que o usuário do Studio possa chamar a API `elasticmapreduce:GetClusterSessionCredentials` no perfil de runtime que deseja usar. Você também deve configurar as the section called “Permissões de usuário do Studio (EC2, EKS)” antes que o usuário possa começar a usar o Studio.

> **Warning**
>
> Para conceder essa permissão, crie uma condição com base na chave de contexto `elasticmapreduce:ExecutionRoleArn` ao conceder acesso a um chamador para chamar as APIs `GetClusterSessionCredentials`. Os exemplos a seguir demonstram como fazer isso.

```
{
      "Sid": "AllowSpecificExecRoleArn",
      "Effect": "Allow",
      "Action": [
          "elasticmapreduce:GetClusterSessionCredentials"
      ],
      "Resource": "*",
      "Condition": {
          "StringEquals": {
              "elasticmapreduce:ExecutionRoleArn": [
                  "arn:aws:iam::111122223333:role/test-emr-demo1",
                  "arn:aws:iam::111122223333:role/test-emr-demo2"
              ]
          }
      }
  }
```

O exemplo a seguir demonstra como permitir que uma entidade principal do IAM use um perfil do IAM, chamado `test-emr-demo3`, como perfil de runtime. Além disso, o titular da política poderá acessar somente os clusters do Amazon EMR com o ID de cluster `j-123456789`.

```
{
    "Sid":"AllowSpecificExecRoleArn",
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:GetClusterSessionCredentials"
    ],
    "Resource": [
          "arn:aws:elasticmapreduce:<region>:111122223333:cluster/j-123456789"
     ],
    "Condition":{
        "StringEquals":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::111122223333:role/test-emr-demo3"
            ]
        }
    }
}
```

O exemplo apresentado a seguir permite que uma entidade principal do IAM use qualquer perfil do IAM com um nome começando com a string `test-emr-demo4` como o perfil de runtime. Além disso, o titular da política poderá acessar somente os clusters do Amazon EMR marcados com o par de valores-chave `tagKey: tagValue`.

```
{
    "Sid":"AllowSpecificExecRoleArn",
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:GetClusterSessionCredentials"
    ],
    "Resource": "*",
    "Condition":{
        "StringEquals":{
             "elasticmapreduce:ResourceTag/tagKey": "tagValue"
        },
        "StringLike":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::111122223333:role/test-emr-demo4*"
            ]
        }
    }
}
```

Inicialização de um novo cluster com um perfil de runtime

Agora que você tem as permissões obrigatórias, inicie um novo cluster com um perfil de runtime que pode ser usado com o Workspace do Amazon EMR Studio.

Se você já iniciou um novo cluster com um perfil de runtime, poderá pular para a seção the section called “Uso do cluster com seu Workspace”.

1. Primeiro, conclua os pré-requisitos apresentados na seção Perfis de runtime para etapas ao Amazon EMR.
2. Em seguida, inicie um cluster com as configurações apresentadas a seguir para usar perfis de runtime com os Workspaces do Amazon EMR Studio. Para obter instruções sobre como iniciar seu cluster, consulte Especificar uma configuração de segurança para um cluster.
   - Escolha o rótulo de versão emr-6.11.0 ou posterior.
   - Selecione o Spark, o Livy e o Jupyter Enterprise Gateway como suas aplicações de cluster.
   - Use a configuração de segurança criada na etapa anterior.
   - Como opção, você pode habilitar o Lake Formation para seu cluster do EMR. Para ter mais informações, consulte Habilitar o Lake Formation com o Amazon EMR.

Depois de iniciar seu cluster, você estará com tudo pronto para usar o cluster habilitado para perfis de runtime com um Workspace do EMR Studio.

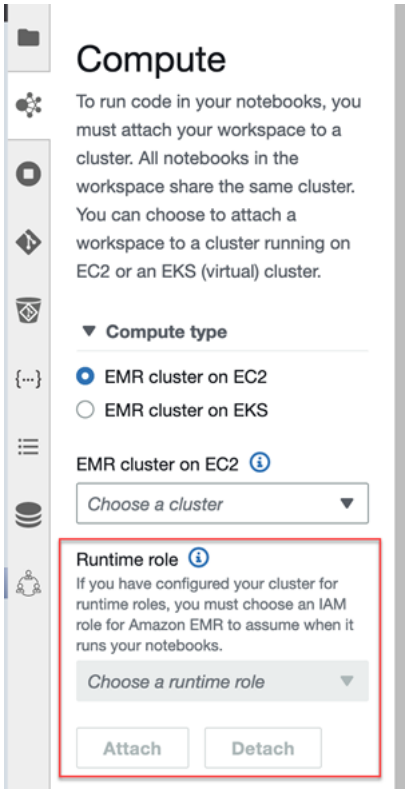
> **Note**
>
> No momento, o ExecutionRoleArnvalor não é compatível com a operação StartNotebookExecutionda API quando o `ExecutionEngineConfig.Type` valor é`EMR`.

Uso do cluster do EMR com um perfil de runtime no Workspaces

Depois de configurar e iniciar o cluster, você poderá usar o cluster habilitado para perfis de runtime com o Workspace do EMR Studio.

1. Crie um novo Workspace ou inicie um Workspace existente. Para ter mais informações, consulte Criação de um Workspace do EMR Studio.

2. Escolha a guia Clusters do EMR na barra lateral à esquerda do seu Workspace aberto, expanda a seção Tipo de computação e selecione seu cluster no menu Cluster do EMR no EC2 e o perfil de runtime no menu Perfil de runtime.



3. Escolha Anexar para anexar o cluster com um perfil de runtime ao seu Workspace.

Considerações

Tenha em mente as seguintes considerações ao usar um cluster habilitado para perfis de runtime com o Workspace do Amazon EMR Studio:

- Você pode selecionar somente um perfil de runtime ao anexar um Workspace do EMR Studio a um cluster do EMR que usa a versão 6.11 ou versões superiores do Amazon EMR.
- A funcionalidade de função de runtime descrita nesta página só é compatível com o Amazon EMR executado no Amazon EC2, e não é compatível com aplicações interativas do EMR Serverless. Para saber mais sobre as funções de runtime do EMR Serverless, consulte Funções de runtime de trabalho no Guia do usuário do Amazon EMR Serverless.
- Embora seja necessário configurar permissões adicionais antes de especificar um perfil de runtime ao enviar um trabalho para um cluster, você não precisa de permissões adicionais para acessar

os arquivos gerados por um Workspace do EMR Studio. As permissões para esses arquivos são semelhantes as dos arquivos gerados de clusters sem perfis de runtime.

- Não é possível usar o SQL Explorer em um Workspace do EMR Studio com um cluster que tenha um perfil de runtime. O Amazon EMR desabilita o SQL Explorer na interface do usuário quando um Workspace é anexado a um cluster do EMR habilitado para perfis de runtime.
- Não é possível usar o modo de colaboração em um Workspace do EMR Studio com um cluster que tenha um perfil de runtime. O Amazon EMR desabilita as funcionalidades de colaboração do Workspace quando um Workspace é anexado a um cluster do EMR habilitado para perfis de runtime. O Workspace permanecerá acessível somente ao usuário que o anexou.
- Você não pode usar perfis de runtime em um Studio com a propagação de identidade confiável do Centro de Identidade do IAM habilitada.
- Você pode encontrar um aviso “A página pode não ser segura!” da interface do usuário do Spark para um cluster habilitado para perfis de runtime. Se isso acontecer, ignore o alerta para continuar a visualizar a interface do usuário do Spark.

Execução de cadernos do Workspace de forma programática

> **Note**
>
> A execução programática de cadernos não é compatível com aplicações interativas do Amazon EMR Serverless.

Você pode executar cadernos do Workspace do Amazon EMR Studio de forma programática com um script ou na AWS CLI. Para saber como executar seu caderno de forma programática, consulte Exemplos de comandos para executar Cadernos do EMR programaticamente.

Navegar pelos dados com o SQL Explorer

> **Note**
>
> O SQL Explorer para EMR Studio não é compatível com aplicações interativas do Amazon EMR Sem Servidor ou em um Studio com a propagação de identidade confiável do Centro de Identidade do IAM habilitada.

Este tópico fornece informações para ajudar a começar a usar o SQL Explorer do Amazon EMR Studio. O SQL Explorer é uma ferramenta de página única no seu Workspace que ajuda você a entender as fontes de dados no catálogo de dados do cluster do EMR. É possível usar o SQL Explorer para navegar pelos seus dados, executar consultas SQL para recuperar dados e fazer download dos resultados da consulta.

O SQL Explorer oferece suporte ao Presto. Antes de usar o SQL Explorer, certifique-se de ter um cluster que use a versão 5.34.0 ou posterior, ou versão 6.4.0 ou posterior, do Amazon EMR com o Presto instalado. O SQL Explorer do Amazon EMR Studio não oferece suporte a clusters do Presto configurados com criptografia em trânsito. Isso ocorre porque o Presto é executado no modo de TLS nesses clusters.

Navegação pelo catálogo de dados do seu cluster

O SQL Explorer fornece uma interface para o navegador do catálogo que você pode usar para explorar e compreender como seus dados são organizados. Por exemplo, você pode usar o navegador do catálogo de dados para verificar nomes de tabelas e de colunas antes de escrever uma consulta SQL.

Navegar em seu catálogo de dados

1. Abra o SQL Explorer em seu Workspace.
2. Certifique-se de que o Workspace esteja anexado a um cluster do EMR em execução no EC2 que usa a versão 6.4.0 ou versões posteriores do Amazon EMR com o Presto instalado. Você pode escolher um cluster existente ou criar um novo. Para ter mais informações, consulte Anexar uma computação a um Workspace do EMR Studio.
3. Selecione um Banco de dados na lista suspensa para navegar.
4. Expanda uma tabela no seu banco de dados para visualizar os nomes das colunas da tabela. Também é possível inserir uma palavra-chave na barra de pesquisa para filtrar os resultados da tabela.

Execução de uma consulta SQL para recuperar dados

Recuperar dados com uma consulta SQL e fazer download dos resultados

1. Abra o SQL Explorer em seu Workspace.

2. Certifique-se de que seu Workspace esteja anexado a um cluster do EMR em execução no EC2 com o Presto e o Spark instalados. Você pode escolher um cluster existente ou criar um novo. Para ter mais informações, consulte Anexar uma computação a um Workspace do EMR Studio.
3. Selecione Abrir editor para abrir uma nova guia do editor em seu Workspace.
4. Escreva sua consulta SQL na guia do editor.
5. Escolha Executar.
6. Visualize os resultados da consulta em Visualização do resultado. Por padrão, o SQL Explorer exibe os cem primeiros resultados. Você pode escolher um número diferente de resultados a serem exibidos (até mil) usando o menu suspenso Visualizar os cem primeiros resultados da consulta.
7. Escolha Fazer download dos resultados para fazer download dos seus resultados no formato CSV. Você pode fazer download de até mil linhas de resultados.

Anexar uma computação a um Workspace do EMR Studio

O Amazon EMR Studio executa comandos de cadernos usando um kernel em um cluster do EMR. Antes de selecionar um kernel, você deve anexar o Workspace a um cluster que usa as instâncias do Amazon EC2 a um cluster do Amazon EMR no EKS ou a uma aplicação do EMR Serverless. O EMR Studio permite anexar Workspaces a clusters novos ou existentes, e oferece flexibilidade para alterar clusters sem a necessidade de fechar o Workspace.

Esta seção aborda os seguintes tópicos para ajudar você a trabalhar e provisionar clusters para o EMR Studio:

- Anexar um cluster do Amazon EC2 a um Workspace do EMR Studio
- Anexar um cluster do Amazon EMR no EKS a um Workspace do EMR Studio
- Anexar uma aplicação do Amazon EMR Serverless a um Workspace do EMR Studio
- Criar e anexar um novo cluster do EMR a um Workspace do EMR Studio
- Desanexar uma computação de um Workspace do EMR Studio

Anexar um cluster do Amazon EC2 a um Workspace do EMR Studio

Você pode anexar um cluster do EMR em execução no Amazon EC2 a um Workspace ao criar o Workspace ou anexar um cluster a um Workspace existente. Se você desejar criar e anexar um novo cluster, consulte Criar e anexar um novo cluster do EMR a um Workspace do EMR Studio.

> **Note**
>
> Um espaço de trabalho em um Studio que tenha a propagação de identidade confiável do Centro de Identidade do IAM habilitada só pode ser anexado a um cluster do EMR com uma configuração de segurança que tenha o Centro de Identidade habilitado.

On create

Anexação a um cluster de computação do Amazon EMR ao criar um Workspace

1. Na caixa de diálogo Criar um Workspace, certifique-se de já ter selecionado uma sub-rede para o novo Workspace. Expanda a seção Configuração avançada.
2. Escolha Anexar Workspace a um cluster do EMR.
3. Na lista suspensa Cluster do EMR, selecione um cluster do EMR existente para anexar ao Workspace.

Depois de anexar um cluster, conclua a criação do Workspace. Ao abrir o novo Workspace pela primeira vez e escolher o painel Clusters do EMR, você deverá visualizar o cluster selecionado anexado.

On launch

Anexação a um cluster de computação do Amazon EMR ao iniciar o Workspace

1. Navegue até a lista Workspaces e selecione a linha do Workspace que você deseja iniciar. Em seguida, selecione Iniciar o Workspace > Iniciar com opções.
2. Escolha um cluster do EMR para anexar ao seu Workspace.

Depois de anexar um cluster, conclua a criação do Workspace. Ao abrir o novo Workspace pela primeira vez e escolher o painel Clusters do EMR, você deverá visualizar o cluster selecionado anexado.

In JupyterLab

Anexe um espaço de trabalho a um cluster computacional do Amazon EMR em JupyterLab

1. Selecione seu Workspace e, em seguida, escolha Iniciar o Workspace > Início rápido.
2. Dentro JupyterLab, abra a guia Cluster na barra lateral esquerda.

3. Selecione o menu suspenso Cluster do EMR no EC2 ou selecione um cluster do Amazon EMR no EKS.
4. Selecione Anexar para anexar o cluster ao seu Workspace.

Depois de anexar o cluster, conclua a criação do Workspace. Ao abrir o novo Workspace pela primeira vez e escolher o painel Clusters do EMR, você deverá visualizar o cluster selecionado anexado.

In the Workspace UI

Anexação de um Workspace a um cluster de computação do Amazon EMR usando a interface do usuário do Workspace

1. No Workspace que você deseja anexar a um cluster, escolha o ícone Clusters do EMR na barra lateral à esquerda para abrir o painel Cluster.
2. Em Tipo de cluster, expanda o menu suspenso e selecione Cluster do EMR no EC2.
3. Escolha um cluster na lista suspensa. Pode ser necessário desanexar um cluster existente primeiro para habilitar a lista suspensa de seleção de cluster.
4. Escolha Anexar. Quando o cluster for anexado, você visualizará uma mensagem de êxito.

Anexar um cluster do Amazon EMR no EKS a um Workspace do EMR Studio

Além de usar clusters do Amazon EMR em execução no Amazon EC2, é possível anexar um Workspace a um cluster do Amazon EMR no EKS para executar códigos de cadernos. Para obter mais informações sobre o Amazon EMR no EKS, consulte O que é o Amazon EMR no EKS?.

Antes de conectar um Workspace a um cluster do Amazon EMR no EKS, o administrador do Studio deve conceder a você as permissões de acesso.

> **Note**
>
> Não é possível executar um cluster do Amazon EMR no EKS em um EMR Studio que usa a propagação de identidade confiável do Centro de Identidade do IAM.

On create

Anexar um cluster do Amazon EMR no EKS ao criar um Workspace

1. Na caixa de diálogo Criar um Workspace, expanda a seção Configuração avançada.
2. Escolha Anexar Workspace a um cluster do Amazon EMR no EKS.
3. Em Cluster do Amazon EMR no EKS, escolha um cluster na lista suspensa.
4. Em Selecionar um endpoint, escolha um endpoint gerenciado para anexar ao Workspace. Um endpoint gerenciado corresponde a um gateway que permite que o EMR Studio se comunique com o cluster escolhido.
5. Escolha Criar um Workspace para concluir o processo de criação do Workspace e anexar o cluster selecionado.

Depois de anexar um cluster, você poderá concluir o processo de criação do Workspace. Ao abrir o novo Workspace pela primeira vez e escolher o painel Clusters do EMR, você visualizará que o cluster selecionado está anexado.

In the Workspace UI

Anexar um cluster do Amazon EMR no EKS usando a interface do usuário do Workspace

1. No Workspace que você deseja anexar a um cluster, escolha o ícone Clusters do EMR na barra lateral à esquerda para abrir o painel Cluster.
2. Expanda o menu suspenso Tipo de cluster e escolha Clusters do EMR no EKS.
3. Em Cluster do EMR no EKS, escolha um cluster na lista suspensa.
4. Em Endpoint, escolha um endpoint gerenciado para anexar ao Workspace. Um endpoint gerenciado corresponde a um gateway que permite que o EMR Studio se comunique com o cluster escolhido.
5. Escolha Anexar. Quando o cluster for anexado, você visualizará uma mensagem de êxito.

Anexar uma aplicação do Amazon EMR Serverless a um Workspace do EMR Studio

É possível conectar um Workspace a uma aplicação do EMR Serverless para executar workloads interativas. Para obter mais informações, consulte Usar cadernos para executar workloads interativas com o EMR Serverless por meio do EMR Studio.

> **Note**
>
> Não é possível anexar uma aplicação do EMR Sem Servidor ao EMR Studio que usa propagação de identidade confiável do Centro de Identidade do IAM.

Example Conecte um espaço de trabalho a um aplicativo EMR Serverless no JupyterLab

Antes de conectar um Workspace a uma aplicação do EMR Serverless, o administrador da conta deve conceder permissões de acesso conforme descrito em Permissões obrigatórias para workloads interativas.

1. Navegue para o EMR Studio, selecione seu Workspace e, em seguida, escolha Iniciar o Workspace > Início rápido.
2. Dentro JupyterLab, abra a guia Cluster na barra lateral esquerda.
3. Selecione EMR Serverless como opção de computação e, em seguida, selecione uma aplicação do EMR Serverless e uma função de runtime.
4. Selecione Anexar para anexar o cluster ao seu Workspace.

Agora, ao abrir esse Workspace, você deverá ver a aplicação selecionada anexada.

Criar e anexar um novo cluster do EMR a um Workspace do EMR Studio

Os usuários avançados do EMR Studio podem provisionar novos clusters do EMR em execução no Amazon EC2 para uso com um Workspace. O novo cluster tem todas as aplicações de big data obrigatórias para o EMR Studio instaladas por padrão.

Para criar clusters, primeiro é necessário que o administrador do Studio conceda permissão a você usando uma política de sessão. Para ter mais informações, consulte Criação de políticas de permissões para usuários do EMR Studio.

Você pode criar um novo cluster na caixa de diálogo Criar um Workspace ou no painel Cluster na interface do usuário do Workspace. De qualquer forma, você tem duas opções de criação de cluster:

1. Criar um cluster do EMR: crie um cluster do EMR ao escolher o tipo e a contagem da instância do Amazon EC2.
2. Usar um modelo de cluster: provisione um cluster ao selecionar um modelo de cluster definido previamente. Esta opção aparece se você tiver permissão para usar os modelos de cluster.

> **Note**
>
> Se você habilitou a propagação de identidade confiável com o Centro de Identidade do IAM para o seu Studio, deverá usar um modelo para criar um cluster.

Criar um cluster do EMR ao fornecer uma configuração de cluster

1. Escolha um ponto de partida.

Para...	Fazer isso...
Criar o cluster ao criar um Workspace com a caixa de diálogo Criar um Workspace.	Expanda a seção Configuração avançada na caixa de diálogo Criar um Workspace e selecione Criar um cluster do EMR.
Criar o cluster usando o painel Cluster do EMR na interface do usuário do Workspace após criar um Workspace.	Escolha a guia Clusters do EMR na barra lateral à esquerda de um Workspace, expanda a seção Configuração avançada e escolha Criar cluster.

2. Insira um Nome de cluster. Nomear o cluster ajuda você a encontrá-lo posteriormente na lista Clusters do EMR Studio.
3. Na Versão do Amazon EMR, escolha uma versão de liberação do Amazon EMR para o cluster.
4. Em Instância, selecione o tipo e o número de instâncias do Amazon EC2 para o cluster. Para obter mais informações sobre como selecionar os tipos de instância, consulte Configurar instâncias do Amazon EC2. Uma instância será usada como nó primário.
5. Selecione uma Sub-rede na qual o EMR Studio possa iniciar o novo cluster. Cada opção de sub-rede é aprovada previamente pelo administrador do Studio, e seu Workspace deve ser capaz de se conectar a um cluster em qualquer sub-rede listada.
6. Escolha um URI do S3 para o armazenamento de log.
7. Escolha Criar cluster do EMR para provisionar o cluster. Se você usar a caixa de diálogo Criar um Workspace, escolha Criar um Workspace para criar o Workspace e provisionar o cluster. Depois que o EMR Studio provisiona o novo cluster, ele anexa o cluster ao Workspace.

Criar um cluster usando um modelo de cluster

1. Escolha um ponto de partida.

Para...	Fazer isso...
Criar o cluster ao criar um Workspace com a caixa de diálogo Criar um Workspace.	Expanda a seção Configuração avançada na caixa de diálogo Criar um Workspace e selecione Usar um modelo de cluster.
Criar o cluster usando o painel Cluster do EMR na interface do usuário do Workspace.	Escolha a guia Clusters do EMR na barra lateral à esquerda de um Workspace, expanda a seção Configuração avançada e, em seguida, selecione Modelo de cluster.

2. Selecione um modelo de cluster na lista suspensa. Cada modelo de cluster disponível inclui uma breve descrição para ajudar você a fazer uma seleção.
3. O modelo de cluster escolhido pode ter parâmetros adicionais, como a versão de liberação do Amazon EMR ou o nome do cluster. Você pode escolher ou inserir valores, ou usar os valores padrão selecionados pelo administrador.
4. Selecione uma Sub-rede na qual o EMR Studio possa iniciar o novo cluster. Cada opção de sub-rede é aprovada previamente pelo administrador do Studio, e seu Workspace deve ser capaz de se conectar a um cluster em qualquer sub-rede.
5. Escolha Usar modelo de cluster para provisionar o cluster e anexá-lo ao Workspace. O EMR Studio demorará alguns minutos para criar o cluster. Se você usar a caixa de diálogo Criar um Workspace, escolha Criar um Workspace para criar o Workspace e provisionar o cluster. Depois que o EMR Studio provisiona o novo cluster, ele anexa o cluster ao seu Workspace.

Desanexar uma computação de um Workspace do EMR Studio

Para trocar o cluster anexado a um Workspace, é possível desanexar um cluster da interface do usuário do Workspace.

Desanexar um cluster de um Workspace

1. No Workspace que você deseja desanexar de um cluster, escolha o ícone Clusters do EMR na barra lateral à esquerda para abrir o painel Cluster.

2. Em Selecionar cluster, escolha Desanexar e aguarde até que o EMR Studio desanexe o cluster. Quando o cluster for desanexado, você visualizará uma mensagem de êxito.

Para desanexar uma aplicação do EMR Serverless de um Workspace do EMR Studio

Para trocar a computação anexada a um Workspace, é possível desanexar a aplicação da interface do usuário do Workspace.

1. No Workspace que você deseja desanexar de um cluster, escolha o ícone Computação do Amazon EMR na barra lateral à esquerda para abrir o painel Computação.
2. Em Selecionar computação, escolha Desanexar e aguarde até que o EMR Studio desanexe a aplicação. Quando a aplicação for desanexada, você visualizará uma mensagem de êxito.

Vinculação de repositórios baseados em Git a um Workspace do EMR Studio

Sobre os repositórios Git para o EMR Studio

Você pode associar, no máximo, três repositórios Git a um Workspace do EMR Studio. Por padrão, cada espaço de trabalho permite que você escolha em uma lista de repositórios Git associados à AWS mesma conta do Studio. Também é possível criar um novo repositório Git como um recurso para um Workspace.

Você pode executar comandos do Git, como os apresentados a seguir, usando um comando de terminal enquanto estiver conectado ao nó primário de um cluster.

```
!git pull origin <branch-name>
```

Como alternativa, você pode usar a extensão jupyterlab-git. Abra-o na barra lateral à esquerda ao escolher o ícone Git. Para obter informações sobre a extensão jupyterlab-git para, consulte jupyterlab-git. JupyterLab

Pré-requisitos

- Para associar um repositório Git a um Workspace, o Studio deve ser configurado para permitir a vinculação do repositório Git. O administrador do Studio deve tomar medidas para o Estabelecimento de acesso e de permissões para repositórios baseados em Git.

- Se você usa um CodeCommit repositório, deve usar as credenciais do Git e o HTTPS. As chaves SSH e HTTPS com o auxiliar de AWS Command Line Interface credenciais não são suportadas. CodeCommit também não oferece suporte a tokens de acesso pessoal (PATs). Para obter mais informações, consulte Como usar o IAM com CodeCommit no Guia do usuário do IAM e Configuração para usuários HTTPS usando credenciais do Git no Guia do AWS CodeCommit usuário.

Instruções

Vincular um repositório Git associado a um Workspace

1. Abra o Workspace que você deseja vincular a um repositório na lista Workspaces no Studio.
2. Na barra lateral à esquerda, escolha o ícone Repositório Git do Amazon EMR para abrir o painel de ferramentas Repositório Git.
3. Em Repositórios Git, expanda a lista suspensa e selecione, no máximo, três repositórios para vincular ao Workspace. O EMR Studio registra sua seleção e começa a vincular cada repositório.

Pode demorar algum tempo para que o processo de vinculação seja concluído. Você pode visualizar o status de cada repositório selecionado no painel de ferramentas Repositório Git. Depois que o EMR Studio vincular um repositório a um Workspace, você deverá visualizar os arquivos que pertencem a esse repositório no painel Navegador de arquivos.

Adicionar um novo repositório Git a um Workspace como um recurso

1. Abra o Workspace que você deseja vincular a um repositório na lista Workspaces em seu Studio.
2. Na barra lateral à esquerda, escolha o ícone Repositório Git do Amazon EMR para abrir o painel de ferramentas Repositório Git.
3. Escolha Adicionar novo repositório Git.
4. Em Nome do repositório, insira um nome descritivo para o repositório no EMR Studio. Os nomes podem conter somente caracteres alfanuméricos, hifens e sublinhados.
5. Em Git repository URL (URL do repositório do Git), insira o URL do repositório. Quando você usa um CodeCommit repositório, essa é a URL que é copiada quando você escolhe Clonar URL e, em seguida, Clonar HTTPS. Por exemplo, `https://git-codecommit.us-west-2.amazonaws.com/v1/repos/[MyCodeCommitRepoName]`.

6. Em Filial, insira o nome de uma filial existente que você deseja conferir.
7. Em Credenciais do Git, escolha uma opção de acordo com as diretrizes apresentadas a seguir. O EMR Studio acessa suas credenciais do Git usando os segredos armazenados no Secrets Manager.

> **Note**
>
> Se você usa um GitHub repositório, recomendamos que você use um token de acesso pessoal (PAT) para autenticar. A partir de 13 de agosto de 2021, GitHub exigirá autenticação baseada em tokens e não aceitará mais senhas ao autenticar operações do Git. Para obter mais informações, consulte a publicação Requisitos de autenticação de token para operações do Git no The GitHub Blog.

Opção	Descrição
Criar um novo segredo	Escolha essa opção para associar as credenciais existentes do Git a um novo segredo que será criado para você. AWS Secrets Manager Execute um dos seguintes procedimentos com base nas credenciais do Git que você usar para o repositório. Se você usar um nome de usuário e uma senha do Git para acessar o repositório, selecione Nome de usuário e senha, insira o Nome do segredo a ser usado no Secrets Manager e, em seguida, insira o Nome de usuário e a Senha a serem associados ao segredo. OU Se você usar um token de acesso pessoal para acessar o repositório, selecione Token de acesso pessoal (PAT), insira o Nome do segredo a ser usado no Secrets Manager

Opção	Descrição
	e, em seguida, insira seu token de acesso pessoal. Para obter mais informações, consulte Criação de um token de acesso pessoal para a linha de comando GitHub e Tokens de acesso pessoal para o Bitbucket. CodeCommit os repositórios não oferecem suporte a essa opção.
Usar um repositório público sem credenciais	Escolha esta opção para acessar um repositório público.
Use um AWS segredo existente	Escolha esta opção se você já salvou suas credenciais como um segredo no Secrets Manager e, em seguida, selecione o nome do segredo na lista. Se você selecionar um segredo associado a um nome de usuário e senha do Git, o segredo deverá estar no formato `{"gitUsername": " MyUserName ", "gitPassword": " MyPassword "}`.

8. Escolha Adicionar repositório para criar o novo repositório. Depois que o EMR Studio criar o novo repositório, você visualizará uma mensagem de êxito. O novo repositório aparece na lista suspensa em Repositórios Git.
9. Para vincular o novo repositório ao seu Workspace, escolha-o na lista suspensa em Repositórios Git.

Pode demorar algum tempo para que o processo de vinculação seja concluído. Depois que o EMR Studio vincular o novo repositório ao Workspace, você deverá visualizar uma nova pasta com o mesmo nome do seu repositório no painel Navegador de arquivos.

Para abrir um repositório vinculado diferente, navegue até a pasta dele no Navegador de arquivos.

Uso do editor do SQL com Amazon Athena no EMR Studio

Visão geral

Você pode usar o Amazon EMR Studio para desenvolver e executar consultas interativas no Amazon Athena. Isso significa que você pode realizar análises SQL no Athena na mesma interface do EMR Studio usada para executar suas workloads do Spark, Scala e outras. Com essa integração, você pode usar o preenchimento automático para desenvolver consultas rapidamente, pesquisar dados em seu AWS Glue Data Catalog, criar consultas salvas, visualizar seu histórico de consultas e muito mais.

Para obter mais informações sobre como usar o Amazon Athena, consulte Como usar o SQL do Athena no Guia do usuário do Amazon Athena.

Uso do editor SQL do Athena no EMR Studio

Use as etapas a seguir para desenvolver e executar consultas interativas no Amazon Athena usando o EMR Studio.

1. Adicione as permissões necessárias ao perfil de usuário para os usuários que acessam o Workspaces nesse Studio. As permissões estão listadas na tabela Permissões do AWS Identity and Access Management para usuários do EMR Studio na coluna Acesse o editor SQL do Amazon Athena no EMR Studio. Como alternativa, você pode copiar o conteúdo da política Avançada das Exemplo de políticas de usuário para conceder aos usuários permissões completas para os recursos do EMR Studio, incluindo este.
2. Configure e crie um EMR Studio.
3. Navegue até seu Studio e selecione Editor de consultas na barra lateral.

Agora, você deve ver a interface do usuário familiar do editor do Athena. Para obter informações sobre conceitos básicos e como usar o SQL do Athena para executar consultas interativas, consulte Conceitos básicos e Como usar o SQL do Athena no Guia do usuário do Amazon Athena.

> **Note**
>
> Se tiver habilitado a propagação de identidade confiável por meio do Centro de Identidade do IAM para o EMR Studio, você deve usar os grupos de trabalho do Athena para controlar o acesso às consultas, e o grupo de trabalho usado também deve usar a propagação de

identidade confiável. Para ver as etapas de configuração do Centro de Identidade e habilitar a propagação de identidade confiável para o grupo de trabalho, consulte Uso de grupos de trabalho do Athena habilitados para o Centro de Identidade do IAM no Guia do usuário do Amazon Athena.

Considerações sobre o uso do editor SQL do Athena no EMR Studio

- A integração com o Athena está disponível em todas as regiões comerciais onde o EMR Studio e o Athena estão disponíveis.
- Os seguintes recursos do Athena não estão disponíveis no EMR Studio:
  - Recursos administrativos, como a criação ou a atualização de grupos de trabalho, fontes de dados ou reservas de capacidade do Athena.
  - Athena para o Spark ou para cadernos do Spark
  - DataZone Integração com a Amazon
  - Otimizador baseado em custos (CBO)
  - Step Functions

CodeWhisperer Integração da Amazon com o EMR Studio Workspaces

Visão geral

Você pode usar a Amazon CodeWhisperer com o Amazon EMR Studio para obter recomendações em tempo real à medida que escreve código. JupyterLab CodeWhispererpode concluir seus comentários, concluir linhas únicas de código, fazer line-by-line recomendações e gerar funções totalmente formadas.

> **Note**
>
> Quando você usa o Amazon EMR Studio, AWS pode armazenar dados sobre seu uso e conteúdo para fins de melhoria do serviço. Para obter mais informações e instruções sobre como cancelar o compartilhamento de dados, consulte Compartilhando seus dados AWS no Guia CodeWhisperer do usuário da Amazon.

Considerações sobre o uso CodeWhisperer com espaços de trabalho

- CodeWhisperer a integração está disponível no mesmo Regiões da AWS local em que o EMR Studio está disponível, conforme documentado nas considerações do EMR Studio.
- O Amazon EMR Studio usa automaticamente o CodeWhisperer endpoint no Leste dos EUA (Norte da Virgínia) (us-east-1) para recomendações, independentemente da região em que seu estúdio esteja.
- CodeWhisperer suporta somente a linguagem Python para codificar scripts ETL para trabalhos do Spark no EMR Studio.
- Uma opção de telemetria do lado do cliente quantifica seu uso de. CodeWhisperer Não há suporte para essa funcionalidade no EMR Studio.

Permissões necessárias para CodeWhisperer

Para usar CodeWhisperer, você deve anexar a seguinte política à sua função de usuário do IAM para o Amazon EMR Studio:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
          "Sid": "CodeWhispererPermissions",
          "Effect": "Allow",
          "Action": [ "codewhisperer:GenerateRecommendations" ],
          "Resource": "*"
        }
     ]
}
```

Use CodeWhisperer com espaços de trabalho

Para exibir o registro de CodeWhisperer referência JupyterLab, abra o CodeWhispererpainel na parte inferior da JupyterLab janela e escolha Abrir registro de referência de código.

A lista a seguir contém atalhos que você pode usar para interagir com CodeWhisperer sugestões:

- Recomendações de pausa — Use as sugestões automáticas de pausa nas configurações. CodeWhisperer
- Aceitar uma recomendação: pressione Tab no teclado.

- Rejeitar uma recomendação: pressione Esc no teclado.
- Navegar pelas recomendações: use as setas para cima e para baixo no teclado.
- Invocação manual: pressione Alt e C no teclado. Se estiver usando um Mac, pressione Cmd e C.

Você também pode usar CodeWhisperer para alterar configurações, como nível de registro, e obter sugestões de referências de código. Para obter mais informações, consulte Configuração CodeWhisperer JupyterLab e recursos no Guia do CodeWhisperer usuário da Amazon.

Depuração de aplicações e trabalhos com o EMR Studio

Com o Amazon EMR Studio, você pode iniciar interfaces de aplicações de dados para analisar aplicações e execuções de trabalhos no navegador.

Você também pode iniciar interfaces do usuário persistentes externas ao cluster para o Amazon EMR em execução em clusters do EC2 no console do Amazon EMR. Para ter mais informações, consulte Visualizar interfaces do usuário de aplicações persistentes.

> **Note**
>
> Com base nas configurações do seu navegador, pode ser necessário habilitar pop-ups para a abertura da interface do usuário de uma aplicação.

Para obter informações sobre como configurar e usar as interfaces da aplicação, consulte The YARN Timeline Server, Monitoring and Instrumentation ou Tez UI Overview.

Depuração do Amazon EMR em execução em trabalhos do Amazon EC2

Workspace UI

Inicialização de uma interface do usuário no cluster usando um arquivo de caderno

Ao usar as versões 5.33.0 e posteriores do Amazon EMR, você pode iniciar a interface do usuário da Web do Spark (a interface do usuário do Spark ou o servidor de histórico do Spark) de um caderno no seu Workspace.

As interfaces de usuário no cluster funcionam com os PySpark kernels, Spark ou SparkR. O tamanho máximo de arquivo visível para logs de eventos ou para logs de contêineres do Spark

é de 10 MB. Se seus arquivos de log excederem 10 MB, recomendamos usar o servidor de histórico do Spark persistente em vez da interface do usuário do Spark no cluster para depurar trabalhos.

> **Important**
>
> Para que o EMR Studio inicie interfaces do usuário de aplicações no cluster usando um Workspace, um cluster deve ser capaz de se comunicar com o Amazon API Gateway. Você deve configurar o cluster do EMR para permitir o tráfego de rede de saída para o Amazon API Gateway e certificar-se de que o Amazon API Gateway possa ser acessado pelo cluster.
>
> A interface do usuário do Spark acessa os logs de contêineres ao resolver nomes de host. Se você usar um nome de domínio personalizado, deverá se certificar de que os nomes de host dos nós do cluster possam ser resolvidos pelo DNS da Amazon ou pelo servidor DNS especificado. Para fazer isso, defina as opções do Protocolo de Configuração Dinâmica de Host (DHCP) para a Amazon Virtual Private Cloud (VPC) associada ao seu cluster. Para obter mais informações sobre as opções do DHCP, consulte Conjuntos de opções DHCP no Guia do usuário da Amazon Virtual Private Cloud.

1. No EMR Studio, abra o Workspace que você deseja usar e certifique-se de que ele esteja conectado a um cluster do Amazon EMR em execução no EC2. Para obter instruções, consulte Anexar uma computação a um Workspace do EMR Studio.
2. Abra um arquivo de notebook e use o PySpark kernel, Spark ou SparkR. Para selecionar um kernel, escolha o nome do kernel no canto superior direito da barra de ferramentas do caderno para abrir a caixa de diálogo Selecionar kernel. O nome aparecerá como Nenhum Kernel! se nenhum kernel tiver sido selecionado.
3. Execute o código do seu caderno. Quando você inicia o Spark Context, o apresentado a seguir aparece como a saída no caderno. Pode demorar alguns segundos para que a aparição ocorra. Se você iniciou o Spark Context, poderá executar o comando `%%info` para acessar um link para a interface do usuário do Spark a qualquer momento.

> **Note**
>
> Se os links da interface do usuário do Spark não funcionarem ou não aparecerem após alguns segundos, crie uma nova célula de caderno e execute o comando `%%info` para gerar os links novamente.

4. Para iniciar a interface do usuário do Spark, escolha Link em IU do Spark. Se a aplicação do Spark estiver em execução, a interface do usuário do Spark será aberta em uma nova guia. Se aplicação estiver sido concluída, o servidor de histórico do Spark será aberto.

   Depois de iniciar a interface do usuário do Spark, você pode modificar a URL no navegador para abrir o YARN ResourceManager ou o Yarn Timeline Server. Adicione um dos caminhos apresentados a seguir depois de `amazonaws.com`.

Interface do usuário da Web	Path	Exemplo de URL modificado
FIO ResourceManager	/rm	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*rm*
Servidor de linha do tempo do YARN	/yts	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*yts*

Interface do usuário da Web	Path	Exemplo de URL modificado
Servidor de histórico do Spark	/shs	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com*/shs*

Studio UI

Inicialização do servidor de linha do tempo do YARN persistente, do servidor de histórico do Spark ou da interface do usuário do Tez usando a interface do usuário do EMR Studio

1. No EMR Studio, selecione Amazon EMR no EC2 no lado esquerdo da página para abrir a lista de clusters do Amazon EMR no EC2.
2. Filtre a lista de clusters por nome, estado ou ID ao inserir valores na caixa de pesquisa. Você também pode pesquisar por intervalo de tempo de criação.
3. Selecione um cluster e, em seguida, escolha Iniciar as interfaces do usuário da aplicação para selecionar uma interface do usuário da aplicação. A interface do usuário da aplicação abre em uma nova guia do navegador e pode demorar algum tempo para carregar.

Depure o EMR Studio em execução no EMR Serverless

Semelhante ao Amazon EMR executado no Amazon EC2, você pode usar a interface de usuário do Workspace para analisar suas aplicações do EMR Serverless. Na interface do usuário do Workspace, ao usar as versões 6.14.0 e posteriores do Amazon EMR, você pode iniciar a interface do usuário da Web do Spark (a interface do usuário do Spark ou o servidor de histórico do Spark) de um caderno no seu Workspace. Para sua conveniência, também fornecemos um link para o log do driver para acesso rápido aos logs do driver do Spark.

Depuração de execuções de trabalhos do Amazon EMR no EKS com o servidor de histórico do Spark

Ao enviar uma execução de trabalho para um cluster do Amazon EMR no EKS, você pode acessar os logs dessa execução de trabalho usando o servidor de histórico do Spark. O servidor de histórico do Spark fornece ferramentas para o monitoramento de aplicações do Spark, como uma lista de estágios e de tarefas do programador, um resumo dos tamanhos de RDD e do uso de memória e

informações sobre o ambiente. Você pode iniciar o servidor de histórico do Spark para as execuções de trabalhos do Amazon EMR no EKS das seguintes maneiras:

- Ao enviar uma execução de trabalho usando o EMR Studio com um endpoint gerenciado do Amazon EMR no EKS, é possível iniciar o servidor de histórico do Spark usando um arquivo de caderno em seu Workspace.
- Ao enviar uma execução de trabalho usando o AWS SDK AWS CLI ou para Amazon EMR no EKS, você pode iniciar o Spark History Server a partir da interface do EMR Studio.

Para obter informações sobre como usar o servidor de histórico do Spark, consulte Monitoring and Instrumentation na documentação do Apache Spark. Para obter mais informações sobre as execuções de trabalhos, consulte Conceitos e componentes no Guia de desenvolvimento do Amazon EMR no EKS.

Iniciar o servidor de histórico do Spark usando um arquivo de caderno no Workspace do EMR Studio

1. Abra um Workspace conectado a um cluster do Amazon EMR no EKS.
2. Selecione e abra seu arquivo de caderno no Workspace.
3. Escolha IU do Spark na parte superior do arquivo de caderno para abrir o servidor de histórico do Spark persistente em uma nova guia.

Iniciar o servidor de histórico do Spark usando a interface do usuário do EMR Studio

> **Note**
>
> A lista de trabalhos na interface do usuário do EMR Studio exibe somente as execuções de trabalhos que você envia usando o AWS SDK AWS CLI ou o SDK para Amazon EMR no EKS.

1. No EMR Studio, selecione Amazon EMR no EKS no lado esquerdo da página.
2. Pesquise o cluster virtual do Amazon EMR no EKS que você usou para enviar a execução de trabalho. É possível filtrar a lista de clusters por status ou ID ao inserir valores na caixa de pesquisa.

3. Selecione o cluster para abrir a página de detalhes dele. A página de detalhes exibe informações sobre o cluster, como o ID, o namespace e o status. A página também mostra uma lista com todas as execuções de trabalhos enviadas para esse cluster.
4. Na página de detalhes do cluster, selecione uma execução de trabalho para depurar.
5. No canto superior à direita da lista Trabalhos, escolha Iniciar servidor de histórico do Spark para abrir a interface da aplicação em uma nova guia do navegador.

Instalação de kernels e de bibliotecas em um Workspace do EMR Studio

Cada Workspace do Amazon EMR Studio tem um conjunto de bibliotecas e kernels instalados previamente.

Kernels e bibliotecas em clusters executados no Amazon EC2

Você também pode personalizar o ambiente do EMR Studio das seguintes maneiras ao usar clusters do EMR em execução no Amazon EC2:

- Instalar kernels do caderno Jupyter e bibliotecas Python em um nó primário do cluster: ao instalar bibliotecas usando esta opção, todos os Workspaces anexados ao mesmo cluster compartilham essas bibliotecas. Você pode instalar kernels ou bibliotecas a partir de uma célula de caderno ou enquanto estiver conectado ao usar SSH para o nó primário de um cluster.
- Usar bibliotecas com escopo de cadernos: quando os usuários do Workspace instalam e usam bibliotecas a partir de uma célula de caderno, essas bibliotecas ficam disponíveis somente para esse caderno. Esta opção permite que diferentes cadernos que usam o mesmo cluster funcionem sem se preocupar com versões conflitantes da biblioteca.

Os Workspaces do EMR Studio têm a mesma arquitetura subjacente dos Cadernos do EMR. Você pode instalar e usar kernels do caderno Jupyter e bibliotecas Python com o EMR Studio da mesma forma que faria com os Cadernos do EMR. Para obter instruções, consulte Instalação e uso de kernels e bibliotecas.

Kernels e bibliotecas em clusters do Amazon EMR no EKS

Os clusters do Amazon EMR no EKS incluem os kernels e PySpark Python 3.7 com um conjunto de bibliotecas pré-instaladas. O Amazon EMR no EKS não oferece suporte à instalação de bibliotecas ou de clusters adicionais.

Cada cluster do Amazon EMR no EKS vem com os seguintes Python e bibliotecas instaladas: PySpark

- Python – boto3, cffi, future, ggplot, jupyter, kubernetes, matplotlib, numpy, pandas, plotly, pycryptodomex, py4j, requests, scikit-learn, scipy, seaborn
- PySpark – ggplot, jupyter, matplotlib, numpy, pandas, plotly, pycryptodomex, py4j, requests, scikit-learn, scipy, seaborn

Kernels e bibliotecas em aplicações do EMR Serverless

Cada aplicativo EMR Serverless vem com o seguinte Python e bibliotecas instaladas: PySpark

- Python – ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, seaborn
- PySpark – ggplot, matplotlib,numpy, pandas, plotly, bokeh, scikit-learn, scipy, seaborn

Aprimoramento de kernels com comandos magic

Visão geral

O EMR Studio e os Cadernos do EMR oferecem suporte para comandos magic. Comandos Magic, ou magics, são aprimoramentos fornecidos pelo kernel do IPython para ajudar na execução e na análise dos dados. O IPython é um ambiente de shell interativo desenvolvido com Python.

O Amazon EMR também oferece suporte Sparkmagic a um pacote que fornece comandos magic específicos aos kernels relacionados ao Spark (PySparkkernels SparkR e Scala) e que usa o Livy no cluster para enviar trabalhos do Spark.

Você pode usar comandos magic, desde que tenha um kernel do Python em seu Caderno do EMR. De forma semelhante, qualquer kernel relacionado ao Spark oferece suporte aos comandos do Sparkmagic.

Os comandos Magic, também chamados de magics, têm duas variedades:

- Linhas magics: esses comandos magic são indicados por um prefixo % único e operam em uma única linha de código.
- Células magics: esses comandos magic são indicados por um prefixo %% duplo e operam em várias linhas de código.

Para saber todos os magics disponíveis, consulte Listar os comandos magic e Sparkmagic.

Considerações e limitações

- O EMR Serverless não oferece suporte `%%sh` para execução de `spark-submit`. Ele não é compatível com os magics de Cadernos EMR.
- Os clusters do Amazon EMR no EKS não oferecem suporte a comandos Sparkmagicpara o EMR Studio. Isso ocorre porque os kernels Spark que você usa com endpoints gerenciados são integrados ao Kubernetes e não são suportados pelo Sparkmagic e pelo Livy. Você pode definir a configuração do Spark diretamente no SparkContext objeto como uma solução alternativa, conforme demonstra o exemplo a seguir.

```
spark.conf.set("spark.driver.maxResultSize", '6g')
```

- Os seguintes magic comandos e ações são proibidos por AWS:
  - `%alias`
  - `%alias_magic`
  - `%automagic`
  - `%macro`
  - Modificar `proxy_user` com `%configure`
  - Modificar `KERNEL_USERNAME` com `%env` ou `%set_env`

Listar os comandos magic e Sparkmagic

Use os seguintes comandos para listar os comandos magic disponíveis:

- `%lsmagic` lista todas as funções magic disponíveis no momento.
- `%%help` lista as funções magic relacionadas ao Spark disponíveis no momento e fornecidas pelo pacote Sparkmagic.

Use `%%configure` para configurar o Spark

Um dos comandos mais úteis do Sparkmagic é o comando `%%configure`, que configura os parâmetros de criação da sessão. Ao usar as configurações `conf`, você pode definir qualquer configuração do Spark mencionada na documentação de configuração do Apache Spark.

Example Adição de arquivo em JAR externo aos Cadernos do EMR do repositório do Maven ou do Amazon S3

Você pode usar a abordagem apresentada a seguir para adicionar uma dependência de arquivo em JAR externo a qualquer kernel relacionado ao Spark compatível com Sparkmagic.

```
%%configure -f
{"conf": {
    "spark.jars.packages": "com.jsuereth:scala-arm_2.11:2.0,ml.combust.bundle:bundle-
ml_2.11:0.13.0,com.databricks:dbutils-api_2.11:0.0.3",
    "spark.jars": "s3://DOC-EXAMPLE-BUCKET/my-jar.jar"
    }
}
```

Example : Configuração do Hudi

Você pode usar o editor de caderno para configurar seu Caderno do EMR para usar o Hudi.

```
%%configure
{ "conf": {
     "spark.jars": "hdfs://apps/hudi/lib/hudi-spark-bundle.jar,hdfs:///apps/hudi/lib/
spark-spark-avro.jar",
     "spark.serializer": "org.apache.spark.serializer.KryoSerializer",
     "spark.sql.hive.convertMetastoreParquet":"false"
     }
}
```

Use o **%%sh** para executar o **spark-submit**

A `%%sh` magic executa comandos de shell em um subprocesso em uma instância do cluster anexado. Normalmente, você usaria um dos kernels relacionados ao Spark para executar aplicações do Spark em seu cluster anexado. No entanto, se desejar usar um kernel do Python para enviar uma aplicação do Spark, você pode usar a magic apresentada a seguir, substituindo o nome do bucket pelo nome do seu bucket em letras minúsculas.

```
%%sh
spark-submit --master yarn --deploy-mode cluster s3://DOC-EXAMPLE-BUCKET/test.py
```

Neste exemplo, o cluster precisa de acesso ao local `s3://DOC-EXAMPLE-BUCKET/test.py` ou o comando falhará.

Você pode usar qualquer comando do Linux com a `%%sh` magic. Se você desejar executar qualquer comando do Spark ou do YARN, use uma das seguintes opções para criar um usuário do Hadoop `emr-notebook` e conceder permissões ao usuário para executar os comandos:

- Você pode criar explicitamente um novo usuário ao executar os comandos a seguir.

```
hadoop fs -mkdir /user/emr-notebook
hadoop fs -chown emr-notebook /user/emr-notebook
```

- Você pode ativar a representação do usuário no Livy, que cria o usuário automaticamente. Consulte Habilitação da representação do usuário para monitorar a atividade de usuários e trabalhos do Spark Para mais informações.

Use **`%%display`** para visualizar dataframes do Spark

Você pode usar o `%%display` magic para visualizar um dataframe do Spark. Para usar essa magic, execute o comando apresentado a seguir.

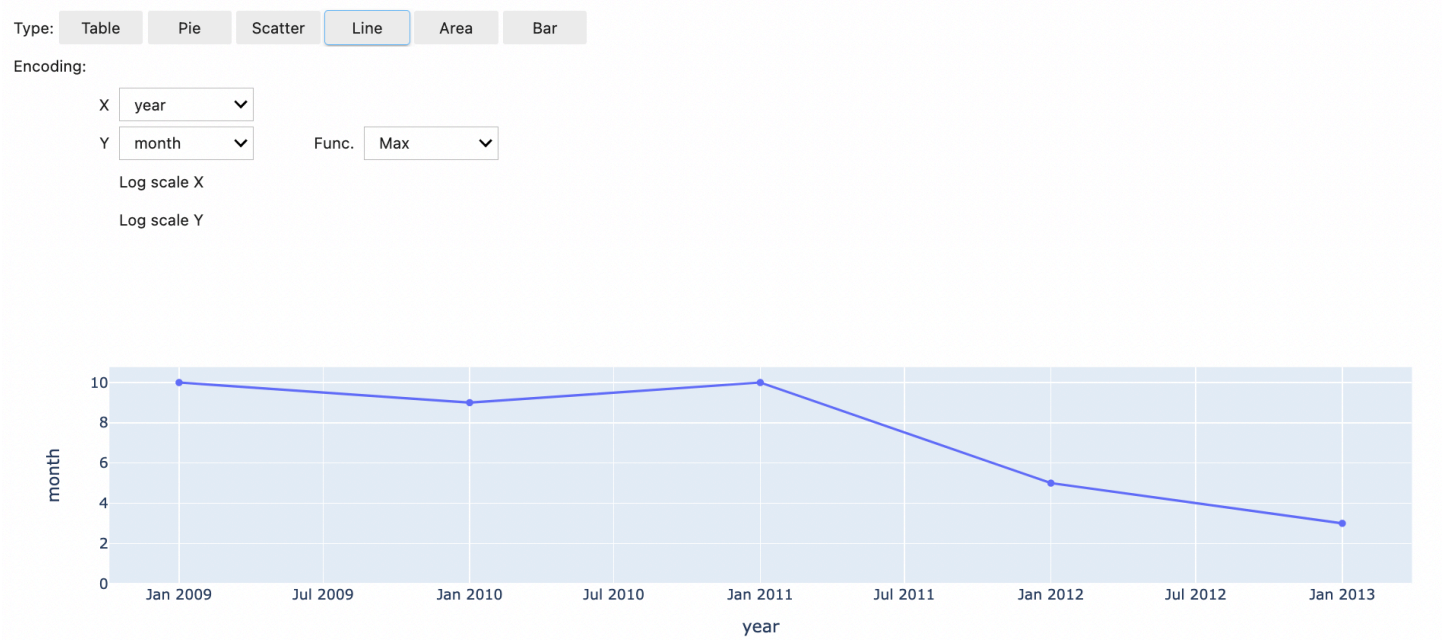
```
%%display df
```

Escolha visualizar os resultados em formato de tabela, como mostra a imagem a seguir.

Type: Table Pie Scatter Line Area Bar

year	month	total_passengers	total_trips
2012-01-01	3	26866837	16146923
2011-01-01	3	26091246	16066350
2013-01-01	3	26965079	15749228
2011-01-01	10	26287953	15707756
2009-01-01	10	26202049	15604551
2012-01-01	5	26278817	15567525
2011-01-01	5	25508952	15554868
2010-01-01	9	25533166	15540209
2010-01-01	5	26002858	15481351
2012-01-01	4	25900645	15477914

Você também pode optar por visualizar os dados com cinco tipos de gráficos. Suas opções incluem gráficos circular, de dispersão, de linha, de área e de barras.



Uso da magic dos Cadernos do EMR

O Amazon EMR fornece as seguintes magics dos Cadernos do EMR, que você pode usar com kernels baseados em Python3 e em Spark:

- `%mount_workspace_dir`: monta seu diretório do Workspace em seu cluster para que você possa importar e executar códigos de outras aplicações em seu Workspace.

> **Note**
>
> Com `%mount_workspace_dir`, somente o kernel do Python 3 pode acessar seus sistemas de arquivos locais. Os executores do Spark não terão acesso ao diretório montado com este kernel.

- `%umount_workspace_dir`: desmonta seu diretório do Workspace do seu cluster.
- `%generate_s3_download_url`: gera um link de download temporário na saída do seu caderno para um objeto do Amazon S3.

Pré-requisitos

Antes de instalar as magics dos Cadernos do EMR, conclua as seguintes tarefas:

- Certifique-se de que seu Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2) tenha acesso de leitura para o Amazon S3. O `EMR_EC2_DefaultRole` com a política gerenciada `AmazonElasticMapReduceforEC2Role` atende a esse requisito. Se você usar uma política ou um perfil personalizado, certifique-se de que ele tenha as permissões do S3 necessárias.

> **Note**
>
> As magics dos Cadernos do EMR é executada em um cluster como o usuário do caderno e usa o perfil de instância do EC2 para interagir com o Amazon S3. Quando você monta um diretório do Workspace em um cluster do EMR, todos os Workspaces e Cadernos do EMR com permissão para serem anexados a esse cluster podem acessar o diretório montado. Por padrão, os diretórios são montados somente para leitura. Embora `s3fs-fuse` e `goofys` permitam montagens de leitura e de gravação, recomendamos fortemente que você não modifique os parâmetros de montagem para montar diretórios no modo de leitura e de gravação. Se você permitir o acesso de gravação, todas as alterações realizadas

no diretório serão gravadas no bucket do S3. Para evitar a exclusão ou a substituição acidentais, você pode habilitar o versionamento para seu bucket do S3. Para saber mais, consulte Usando o versionamento em buckets do S3.

- Execute um dos scripts apresentados a seguir em seu cluster para instalar as dependências para as magics dos Cadernos do EMR. Para executar um script, é possível Usar ações de bootstrap personalizadas ou seguir as instruções em Run commands and scripts on an Amazon EMR cluster quando já tiver um cluster em execução.

  Você pode escolher qual dependência instalar. Tanto o s3fs-fuse quanto o goofys são ferramentas do FUSE (Filesystem in Userspace) que permitem montar um bucket do Amazon S3 como um sistema de arquivos local em um cluster. A ferramenta `s3fs` oferece uma experiência semelhante à POSIX. A ferramenta `goofys` é uma boa opção quando você prefere performance em vez de um sistema de arquivos compatível com a POSIX.

  A série Amazon EMR 7.x usa o Amazon Linux 2023, que não é compatível com repositórios EPEL. Se você estiver executando o Amazon EMR 7.x, siga as instruções do GitHubs3fs-fuse para instalar. `s3fs-fuse` Se você usa as séries 5.x ou 6.x, use os seguintes comandos para instalar. `s3fs-fuse`

  ```
  #!/bin/sh

  # Install the s3fs dependency for EMR Notebooks magics
  sudo amazon-linux-extras install epel -y
  sudo yum install s3fs-fuse -y
  ```

  OU

  ```
  #!/bin/sh

  # Install the goofys dependency for EMR Notebooks magics
  sudo wget https://github.com/kahing/goofys/releases/latest/download/goofys -P /usr/
  bin/
  sudo chmod ugo+x /usr/bin/goofys
  ```

Instale as magics dos Cadernos do EMR

> **Note**
>
> Com as versões 6.0 a 6.9.0 e 5.0 a 5.36.0 do Amazon EMR, somente as versões 0.2.0 e superiores do pacote `emr-notebooks-magics` oferecem suporte à magic do `%mount_workspace_dir`.

Conclua as etapas a seguir para instalar as magics dos Cadernos do EMR.

1. Em seu caderno, execute os comandos apresentados a seguir para instalar o pacote `emr-notebooks-magics`.

   ```
   %pip install boto3 --upgrade
   %pip install botocore --upgrade
   %pip install emr-notebooks-magics --upgrade
   ```

2. Reinicie o kernel para carregar as magics dos Cadernos do EMR.
3. Verifique a instalação com o comando a seguir, que deve exibir o texto de ajuda de saída para `%mount_workspace_dir`.

   ```
   %mount_workspace_dir?
   ```

Montagem de um diretório do Workspace com **`%mount_workspace_dir`**

A magic do `%mount_workspace_dir` permite montar o diretório do Workspace em seu cluster do EMR para que você possa importar e executar outros arquivos, módulos ou pacotes armazenados no diretório.

O exemplo apresentado a seguir monta todo o diretório do Workspace em um cluster e especifica o argumento opcional *`<--fuse-type>`* para usar goofys para a montagem do diretório.

```
%mount_workspace_dir . <--fuse-type goofys>
```

Para verificar se o diretório do Workspace está montado, use o exemplo a seguir para exibir o diretório de trabalho atual com o comando `ls`. A saída deve exibir todos os arquivos em seu Workspace.

```
%%sh
ls
```

Quando você terminar de fazer as alterações no Workspace, poderá desmontar o diretório do Workspace com o seguinte comando:

> **Note**
>
> O diretório do Workspace permanece montado em seu cluster mesmo quando o Workspace é interrompido ou desanexado. Você deve desmontar explicitamente o diretório do Workspace.

```
%umount_workspace_dir
```

Download de um objeto do Amazon S3 com **`%generate_s3_download_url`**

O comando `generate_s3_download_url` cria um URL assinado previamente para um objeto armazenado no Amazon S3. Você pode usar o URL assinado previamente para fazer download do objeto em sua máquina local. Por exemplo, você pode executar `generate_s3_download_url` para fazer download do resultado de uma consulta SQL que seu código grava no Amazon S3.

Por padrão, o URL assinado previamente é válido por 60 minutos. Você pode alterar o tempo de expiração ao especificar um número de segundos para o sinalizador `--expires-in`. Por exemplo, `--expires-in 1800` cria um URL válido por 30 minutos.

O exemplo apresentado a seguir gera um link de download para um objeto ao especificar o caminho completo do Amazon S3: *`s3://EXAMPLE-DOC-BUCKET/path/to/my/object`*.

```
%generate_s3_download_url s3://EXAMPLE-DOC-BUCKET/path/to/my/object
```

Para saber mais sobre como usar `generate_s3_download_url`, execute o comando a seguir para exibir o texto de ajuda.

```
%generate_s3_download_url?
```

Execução de um caderno no modo descentralizado com `%execute_notebook`

Com a magic do `%execute_notebook`, você pode executar outro caderno no modo descentralizado e visualizar a saída de cada célula executada. Essa magic requer permissões adicionais para o perfil de instância que o Amazon EMR e o Amazon EC2 compartilham. Para obter mais detalhes sobre como conceder permissões adicionais, execute o comando `%execute_notebook?`.

Durante um trabalho de execução prolongada, seu sistema pode entrar em repouso devido à inatividade ou pode perder temporariamente a conectividade com a Internet. Isso pode interromper a conexão entre o seu navegador e o servidor Jupyter. Nesse caso, você poderá perder a saída das células que executou e enviou usando o servidor Jupyter.

Se você executar o caderno no modo descentralizado com a magic do `%execute_notebook`, os Cadernos do EMR capturarão a saída das células que foram executadas, mesmo se a rede local sofrer interrupções. Os Cadernos do EMR salvam a saída de forma incremental em um novo caderno com o mesmo nome do caderno que você executou. Em seguida, os Cadernos do EMR colocam o caderno em uma nova pasta no Workspace. As execuções descentralizadas ocorrem no mesmo cluster e usam o perfil de serviço `EMR_Notebook_DefaultRole`, mas argumentos adicionais podem alterar os valores padrão.

Para executar um caderno no modo descentralizado, use o seguinte comando:

```
%execute_notebook <relative-file-path>
```

Para especificar um ID de cluster e um perfil de serviço para uma execução descentralizada, use o seguinte comando:

```
%execute_notebook <notebook_name>.ipynb --cluster-id <emr-cluster-id> --service-role
 <emr-notebook-service-role>
```

Quando o Amazon EMR e o Amazon EC2 compartilham um perfil de instância, o perfil requer as seguintes permissões adicionais:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
```

```
                "elasticmapreduce:StartNotebookExecution",
                "elasticmapreduce:DescribeNotebookExecution",
                "ec2:DescribeInstances"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "arn:aws:iam::<AccoundId>:role/EMR_Notebooks_DefaultRole"
        }
    ]
}
```

> **Note**
>
> Para usar a magic do `%execute_notebook`, instale a versão 0.2.3 ou superior do pacote `emr-notebooks-magics`.

Use cadernos em várias linguagens com kernels do Spark

Cada kernel de caderno Jupyter tem uma linguagem padrão. Por exemplo, o idioma padrão do kernel do Spark é o Scala e o idioma padrão do PySpark kernel é o Python. Com a versão 6.4.0 e versões posteriores do Amazon EMR, o EMR Studio oferece suporte a cadernos em várias linguagens. Isso significa que cada kernel no EMR Studio pode oferecer suporte às seguintes linguagens, além da linguagem padrão: Python, Spark, R e Spark SQL.

Para ativar este atributo, especifique um dos comandos magic apresentados a seguir no início de qualquer célula.

Idioma	Command
Python	`%%pyspark`
Scala	`%%scalaspark`
R	`%%rspark`

Idioma	Command
	Não há suporte para workloads interativas com o EMR Serverless.
Spark SQL	`%%sql`

Quando invocados, esses comandos executam a célula inteira na mesma sessão do Spark usando o interpretador da linguagem correspondente.

A `%%pyspark` célula magic permite que os usuários escrevam PySpark código em todos os kernels do Spark.

```
%%pyspark
a = 1
```

A célula magic do `%%sql` permite que os usuários executem código Spark SQL em todos os kernels do Spark.

```
%%sql
SHOW TABLES
```

A célula magic do `%%rspark` permite que os usuários executem código SparkR em todos os kernels do Spark.

```
%%rspark
a <- 1
```

A célula magic do `%%scalaspark` permite que os usuários executem código Spark Scala em todos os kernels do Spark.

```
%%scalaspark
val a = 1
```

Compartilhamento de dados entre interpretadores de linguagens com tabelas temporárias

Você também pode compartilhar dados entre interpretadores de linguagem usando tabelas temporárias. O exemplo a seguir usa `%%pyspark` em uma célula para criar uma tabela temporária

em Python e usa `%%scalaspark` na célula a seguir para realizar a leitura de dados dessa tabela em Scala.

```
%%pyspark
df=spark.sql("SELECT * from nyc_top_trips_report LIMIT 20")
# create a temporary table called nyc_top_trips_report_view in python
df.createOrReplaceTempView("nyc_top_trips_report_view")
```

```
%%scalaspark
// read the temp table in scala
val df=spark.sql("SELECT * from nyc_top_trips_report_view")
df.show(5)
```

Visão geral dos Cadernos do Amazon EMR

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Você pode usar os Notebooks do Amazon EMR junto com os clusters do Amazon EMR executando o Apache Spark para criar e abrir o notebook Jupyter e interfaces dentro do console do Amazon EMR. JupyterLab Um Caderno do EMR é um caderno com “tecnologia sem servidor” que você pode usar para executar consultas e códigos. Ao contrário de um caderno tradicional, o conteúdo de um Caderno do EMR, nomeadamente, as equações, as consultas, os modelos, o código e o texto narrativo das células de cadernos, é executado em um cliente. Os comandos são executados usando um kernel no cluster do EMR. O conteúdo do caderno também é salvo no Amazon S3 separadamente dos dados do cluster para maior durabilidade e reutilização flexível.

É possível iniciar um cluster, anexar um Caderno do EMR para análise e, em seguida, encerrar o cluster. Você também pode fechar um bloco de anotações anexado a um cluster em execução e alternar para outro. Diversos usuários podem anexar cadernos ao mesmo cluster simultaneamente e compartilhar arquivos de cadernos no Amazon S3 entre si. Esses recursos permitem executar clusters sob demanda para economizar custos e reduzir o tempo gasto reconfigurando blocos de anotações para diferentes clusters e conjuntos de dados.

Você também pode executar um Caderno do EMR programaticamente usando a API do Amazon EMR, sem a necessidade de interagir com o console do Amazon EMR (“execução descentralizada”). É necessário incluir uma célula no Caderno do EMR que tenha uma etiqueta de parâmetros. Essa célula permite que um script transfira novos valores de entrada para o caderno. Cadernos parametrizados podem ser reutilizados com diferentes conjuntos de valores de entrada. Não há necessidade de fazer cópias do mesmo caderno para editar e executar com novos valores de entrada. O Amazon EMR cria e salva o caderno de saída no S3 para cada execução do caderno parametrizado. Para obter exemplos de código da API do Caderno do EMR, consulte Exemplos de comandos para executar Cadernos do EMR programaticamente.

> **Important**
>
> A funcionalidade de Cadernos do EMR oferece suporte a clusters que usam versões 5.18.0 e superiores do Amazon EMR. Recomendamos usar os Cadernos do EMR com clusters que usam a versão mais recente do Amazon EMR ou, no mínimo, as versões 5.30.0, 5.32.0 ou 6.2.0. Com essas versões, os kernels do Jupyter são executados no cluster anexado, em vez de em uma instância do Jupyter. Isso melhora a performance e aprimora sua capacidade de personalizar kernels e bibliotecas. Para ter mais informações, consulte Diferenças nas funcionalidades por versão de liberação do cluster.

Cobranças são aplicáveis ao armazenamento do Amazon S3 e aos clusters do Amazon EMR.

Os notebooks Amazon EMR estão disponíveis como Amazon EMR Studio Workspaces no console.

Como realizar a transição dos Cadernos do EMR para os Workspaces

No novo console do Amazon EMR, mesclamos os Cadernos do EMR com os Workspaces do Amazon EMR Studio em uma única experiência. Ao usar um EMR Studio, é possível criar e configurar diferentes Workspaces para organizar e executar cadernos. Se você tinha Cadernos do Amazon EMR no console antigo, eles estarão disponíveis como Workspaces do EMR Studio no novo console.

O Amazon EMR criou esses novos Workspaces do EMR Studio para você. O número de Studios que criamos corresponde ao número de VPCs distintas que você usa nos Cadernos do EMR. Por exemplo, se você se conectar a clusters do EMR em duas VPCs diferentes de Cadernos do EMR, criaremos dois novos EMR Studios. Seus cadernos serão distribuídos entre os novos Studios.

> **Important**
>
> Desativamos a opção de criar novos cadernos no console antigo do Amazon EMR. Em vez disso, use Criar Workspace no novo console do Amazon EMR.

Para obter mais informações sobre os Workspaces do Amazon EMR Studio, consulte Compreensão das noções básicas do Workspace. Para obter uma visão geral conceitual do EMR Studio, consulte Workspaces na página Como o Amazon EMR Studio funciona.

O que você precisa fazer?

Embora você ainda possa usar os cadernos existentes no console antigo, recomendamos usar os Workspaces do Amazon EMR Studio no novo console. Você deve configurar permissões de perfis adicionais para ativar as funcionalidades do EMR Studio que não estão disponíveis nos Cadernos do EMR.

> **Note**
>
> No mínimo, para visualizar os Cadernos do EMR existentes como Workspaces do EMR Studio e para criar novos Workspaces, os usuários devem ter permissões `elasticmapreduce:ListStudios` e `elasticmapreduce:CreateStudioPresignedUrl` em seus perfis. Para acessar todos os recursos do EMR Studio, consulte Como habilitar recursos do EMR Studio para usuários de Cadernos do EMR para obter a lista completa de permissões adicionais que os usuários de Cadernos do EMR precisarão.

Funcionalidades aprimoradas no EMR Studio além dos Cadernos do EMR

Com o Amazon EMR Studio, você pode configurar e usar os seguintes recursos que não estão disponíveis nos Cadernos do EMR:

- Navegar e anexar clusters do EMR internamente no JupyterLab
- Navegar e anexar clusters virtuais de Cadernos do EMR internamente no JupyterLab
- Conectar-se aos repositórios Git internamente no JupyterLab
- Colaborar com outros membros da sua equipe para escrever e executar códigos de cadernos
- Navegar pelos dados com o SQL Explorer
- Provisionar clusters do EMR com o Service Catalog

Para obter uma lista completa das funcionalidades do Amazon EMR Studio, consulte Principais recursos do EMR Studio.

Como habilitar recursos do EMR Studio para usuários de Cadernos do EMR

Os novos EMR Studios que criaremos como parte dessa mesclagem usam o perfil do IAM `EMR_Notebooks_DefaultRole` existente como perfil de serviço do EMR Studio.

Os usuários que fazem a transição dos Cadernos do EMR para o EMR Studio e desejam usar os recursos adicionais do EMR Studio precisam de diversas novas permissões de perfil. Adicione as permissões apresentadas a seguir aos perfis dos usuários dos Cadernos do EMR que planejam usar o EMR Studio.

> **Note**
>
> No mínimo, para visualizar os Cadernos do EMR existentes como Workspaces do EMR Studio e para criar novos Workspaces, os usuários devem ter permissões `elasticmapreduce:ListStudios` e `elasticmapreduce:CreateStudioPresignedUrl` em seus perfis. Para usar todos os recursos do EMR Studio, adicione todas as permissões listadas abaixo. Os usuários administradores também precisam de permissão para criar e gerenciar um EMR Studio. Para ter mais informações, consulte Permissões de administrador para criar e gerenciar um EMR Studio.

```
"elasticmapreduce:DescribeStudio",
"elasticmapreduce:ListStudios",
"elasticmapreduce:CreateStudioPresignedUrl",
"elasticmapreduce:UpdateEditor",
"elasticmapreduce:PutWorkspaceAccess",
"elasticmapreduce:DeleteWorkspaceAccess",
"elasticmapreduce:ListWorkspaceAccessIdentities",
"emr-containers:ListVirtualClusters",
"emr-containers:DescribeVirtualCluster",
"emr-containers:ListManagedEndpoints",
"emr-containers:DescribeManagedEndpoint",
"emr-containers:CreateAccessTokenForManagedEndpoint",
"emr-containers:ListJobRuns",
"emr-containers:DescribeJobRun",
"servicecatalog:SearchProducts",
"servicecatalog:DescribeProduct",
"servicecatalog:DescribeProductView",
```

```
"servicecatalog:DescribeProvisioningParameters",
"servicecatalog:ProvisionProduct",
"servicecatalog:UpdateProvisionedProduct",
"servicecatalog:ListProvisioningArtifacts",
"servicecatalog:DescribeRecord",
"servicecatalog:ListLaunchPaths",
"cloudformation:DescribeStackResources"
```

As permissões apresentadas a seguir também são necessárias para usar as funcionalidades de colaboração no EMR Studio, mas não eram obrigatórias nos Cadernos do EMR.

```
"sso-directory:SearchUsers",
"iam:GetUser",
"iam:GetRole",
"iam:ListUsers",
"iam:ListRoles",
"sso:GetManagedApplicationInstance"
```

Considerações sobre o uso de Cadernos do EMR

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Considere os requisitos apresentados a seguir ao criar clusters e desenvolver soluções usando os Cadernos do EMR.

Requisitos de cluster

- Habilitar o bloqueio de acesso público do Amazon EMR: o acesso de entrada a um cluster possibilita que os usuários do cluster executem kernels de caderno. Garanta que somente usuários autorizados possam acessar o cluster. Recomendamos que você deixe o acesso público ao bloco habilitado e limite o tráfego SSH de entrada apenas a fontes confiáveis. Para obter mais

informações, consulte Usar o bloqueio de acesso público do Amazon EMR e Controle do tráfego de rede com grupos de segurança.

- Use um cluster compatível: um cluster conectado a um caderno deve atender aos seguintes requisitos:
  - Somente clusters criados usando o Amazon EMR são compatíveis. Você pode criar um cluster de forma independente no Amazon EMR e, em seguida, anexar um Caderno do EMR ou pode criar um cluster compatível ao criar um Caderno do EMR.
  - Somente clusters criados usando o Amazon EMR versão 5.18.0 e com versões posteriores são compatíveis. Consulte the section called “Diferenças nas funcionalidades por versão de liberação do cluster”.
  - Clusters criados usando instâncias do Amazon EC2 com processadores AMD EPYC (por exemplo, tipos de instância m5a.* e r5a.*) não são compatíveis.
  - Os Cadernos do EMR funciona somente com clusters criados com `VisibleToAllUsers` definidos como `true`. `VisibleToAllUsers` é `true`, por padrão.
  - O cluster deve ser executado em uma EC2-VPC. Sub-redes públicas e privadas têm suporte. A plataforma EC2-Classic solicitada não é compatível.
  - O cluster deve ser iniciado com o Hadoop, Spark e Livy instalados. Outras aplicações podem ser instaladas, mas, no momento, os Cadernos do EMR oferecem suporte somente para clusters do Spark.

    > **Important**
    >
    > Para versões 5.32.0 e posteriores, ou 6.2.0 e posteriores, do Amazon EMR seu cluster também deve estar executando a aplicação Jupyter Enterprise Gateway para funcionar com Cadernos do EMR.

  - Clusters que usam a autenticação do Kerberos não são compatíveis.
  - Clusters integrados AWS Lake Formation oferecem suporte somente à instalação de bibliotecas com escopo de notebook. A instalação de kernels e bibliotecas no cluster não é permitida.
  - Clusters com vários nós primários não são compatíveis.
  - Não há suporte para clusters que usam instâncias do Amazon EC2 baseadas em AWS Graviton2.

Diferenças nas funcionalidades por versão de liberação do cluster

É altamente recomendável usar Cadernos do EMR com clusters criados usando as versões 5.30.0, 5.32.0 ou posteriores, ou 6.2.0 ou posteriores, do Amazon EMR. Com essas versões, os Cadernos do EMR executam kernels no cluster do Amazon EMR anexado. Os kernels e as bibliotecas podem ser instalados diretamente no nó primário do cluster. O uso de Cadernos do EMR com essas versões de cluster fornece os seguintes benefícios:

- Performance aprimorada: os kernels de cadernos são executados em clusters com tipos de instância do EC2 selecionados. As versões anteriores executam kernels em uma instância especializada que não pode ser redimensionada, acessada ou personalizada.
- Capacidade de adicionar e personalizar kernels: você pode se conectar ao cluster para instalar pacotes de kernel usando `conda` e `pip`. Além disso, a instalação de `pip` é compatível com o uso de comandos de terminal dentro de células do bloco de anotações. Nas versões anteriores, somente kernels pré-instalados estavam disponíveis (Python, PySpark Spark e SparkR). Para ter mais informações, consulte Instalação de kernels e de bibliotecas Python em um nó primário do cluster.
- Capacidade de instalar bibliotecas Python: você pode instalar bibliotecas Python no nó primário do cluster usando `conda` e `pip`. Recomendamos usar `conda`. Nas versões anteriores, somente bibliotecas com escopo de notebook são suportadas. PySpark

Recursos de Cadernos do EMR compatíveis por versão de cluster

Versão do cluster	Bibliotecas com escopo de notebooks para PySpark	Instalação do kernel no cluster	Instalação da bibliotec a Python no nó primário
Antes da versão 5.18.0	Sem suporte para Cadernos do EMR		
5.18.0 a 5.25.0	Não	Não	Não
5.26.0 a 5.29.0	Sim	Não	Não
5.30.0	Sim	Sim	Sim
6.0.0	Não	Não	Não

Versão do cluster	Bibliotecas com escopo de notebooks para PySpark	Instalação do kernel no cluster	Instalação da bibliotec a Python no nó primário
5.32.0 e posteriores e 6.2.0 e posteriores	Sim	Sim	Sim

Limites para Cadernos do EMR anexados simultaneamente

Ao criar um cluster compatível com cadernos, considere o tipo de instância do EC2 do nó primário do cluster. As restrições de memória dessa instância do EC2 determinam o número de blocos de anotações que podem estar prontos simultaneamente para executar código e consultas no cluster.

Tipo de instância do EC2 do nó primário	Número de Cadernos do EMR
*.medium	2
*.large	4
*.xlarge	8
*.2xlarge	16
*.4xlarge	24
*.8xlarge	24
*.16xlarge	24

Versões do caderno Jupyter e Python

Os Cadernos do EMR executam o Caderno Jupyter versão 6.0.2 e o Python 3.6.5, independentemente da versão do Amazon EMR do cluster anexado.

Considerações sobre segurança

Usar locais criptografados do S3

Se você especificar um local criptografado no Amazon S3 para armazenar arquivos de cadernos, deverá configurar o Perfil de serviço para Cadernos do EMR como usuário da chave. A função de serviço padrão é `EMR_Notebooks_DefaultRole`. Se você estiver usando uma AWS KMS chave para criptografia, consulte Usando políticas de chaves no AWS KMS no Guia do AWS Key Management Service desenvolvedor e no artigo de suporte para adicionar usuários de chaves.

Uso de cookies com domínios de hospedagem

Para aumentar a segurança das aplicações fora do console que podem ser usadas com o Amazon EMR, os domínios de hospedagem das aplicações são registrados na Public Suffix List (PSL). Exemplos desses domínios de hospedagem incluem os seguintes: `emrstudio-prod.us-east-1.amazonaws.com`, `emrnotebooks-prod.us-east-1.amazonaws.com`, `emrappui-prod.us-east-1.amazonaws.com`. Para maior segurança, se precisar definir cookies confidenciais no nome de domínio padrão, recomendamos que você use cookies com um prefixo `__Host-`. Isso ajuda a defender seu domínio contra tentativas de falsificação de solicitação entre sites (CSRF). Para obter mais informações, consulte a página Set-Cookie em Mozilla Developer Network.

Criação de um bloco de anotações

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Você cria um Caderno do EMR usando o console antigo do Amazon EMR. A criação de notebooks usando a API do Amazon EMR AWS CLI ou a API do Amazon EMR não é suportada.

Para criar um notebook do EMR

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/.

2. Escolha Notebooks (Blocos de anotações, Create notebook (Criar bloco de anotações).
3. Insira um Notebook name (Nome do bloco de anotações) e uma Notebook description (Descrição do bloco de anotações) adicional.
4. Se você tiver um cluster ativo ao qual deseja anexar o caderno, deixe o padrão Escolher um cluster existente selecionado, clique em Escolher, selecione um cluster na lista e, em seguida, clique em Escolher cluster. Para obter informações sobre os requisitos de cluster para Cadernos do EMR, consulte Considerações sobre o uso de Cadernos do EMR.

   —ou—

   Escolha Criar um cluster, insira um Nome de cluster e escolha opções de acordo com as diretrizes a seguir. O cluster é criado na VPC padrão para a conta usando instâncias sob demanda.

Configuração	Descrição
Nome do cluster	O nome amigável usado para identificar o cluster.
Versão	Não pode ser modificado. O padrão é a versão mais recente do Amazon EMR (5.36.2).
Aplicativos	Não pode ser modificado. Lista os aplicativos instalados no cluster.
Instância	Insira o número de instâncias e selecione o tipo de instância do EC2. Uma instância é usada para o nó primário. O resto é usado para nós core. O tipo de instância determina o número de blocos de anotações que podem ser anexados ao cluster simultane amente. Para ter mais informações, consulte Limites para Cadernos do EMR anexados simultaneamente.
Função do EMR	Deixe o padrão ou escolha o link para especificar um perfil de serviço personali

Configuração	Descrição
	zado para o Amazon EMR. Para ter mais informações, consulte Perfil de serviço para Amazon EMR (perfil do EMR).
Perfil de instância do EC2	Deixe o padrão ou escolha o link para especificar uma função de serviço personali zada para instâncias do EC2. Para ter mais informações, consulte Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2).
EC2 key pair	Escolha um par de chaves do EC2 para poder se conectar a instâncias de cluster. Para ter mais informações, consulte Conectar-se ao nó primário usando SSH.
Encerramento automático	O encerramento automático é compatível com as versões 5.30.0 e 6.1.0 e posteriores do Amazon EMR. Marque a caixa de seleção para habilitar o encerramento automático e, em seguida, especifique o tempo de inatividade após o qual o cluster deverá ser desligado automatic amente. Para ter mais informações, consulte Usar uma política de término automático.

5. Em Security groups (Grupos de segurança), escolha Use default security groups (Usar grupos de segurança padrão). Como alternativa, escolha Escolher grupos de segurança e selecione grupos de segurança personalizados que estão disponíveis na VPC do cluster. Selecione um grupo para a instância primária e outro para a instância do cliente do caderno. Para ter mais informações, consulte the section called “Grupos de segurança para Cadernos do EMR”.
6. Em Perfil de serviço da AWS , deixe o padrão ou escolha um perfil personalizado na lista. A instância do cliente do bloco de anotações usa essa função. Para ter mais informações, consulte Perfil de serviço para Cadernos do EMR.

7. Em Local do caderno, escolha o local no Amazon S3 no qual o arquivo de caderno será salvo ou especifique seu próprio local. Se o bucket e a pasta não existirem, o Amazon EMR os criará.

   O Amazon EMR cria uma pasta com o ID do caderno como nome da pasta e salva o caderno em um arquivo chamado *NotebookName*`.ipynb`. Por exemplo, se você especificar o local do Amazon S3 `s3://MyBucket/MyNotebooks` para um caderno chamado `MyFirstEMRManagedNotebook`, o arquivo de caderno será salvo em `s3://MyBucket/MyNotebooks/`*NotebookID*`/MyFirstEMRManagedNotebook.ipynb`.

   Se você especificar um local criptografado no Amazon S3, deverá configurar o Perfil de serviço para Cadernos do EMR como um usuário da chave. A função de serviço padrão é `EMR_Notebooks_DefaultRole`. Se você estiver usando uma AWS KMS chave para criptografia, consulte Usando políticas de chaves no AWS KMS no Guia do AWS Key Management Service desenvolvedor e no artigo de suporte para adicionar usuários de chaves.
8. Como opção, se você adicionou um repositório baseado em Git ao Amazon EMR que deseja associar a este caderno, escolha Repositório Git, selecione Escolher repositório e, em seguida, escolha um repositório na lista. Para ter mais informações, consulte Associação de repositórios baseados em Git a Cadernos do EMR.
9. Opcionalmente, selecione Tags e, em seguida, adicione as tags de chave-valor adicionais para o bloco de anotações.

   > **Important**
   >
   > Uma tag padrão com a string Key (Chave) definida como `creatorUserID` e o valor definido como o ID de usuário do IAM são aplicados para fins de acesso. Recomendamos que você não altere nem remova essa tag, pois ela pode ser usada para controlar o acesso. Para ter mais informações, consulte Usar etiquetas de caderno e cluster com as políticas de controle de acesso do IAM.
10. Selecione Criar bloco de anotações.

Como trabalhar com Cadernos do EMR

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou

criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Depois de criar um Caderno do EMR, o caderno demora um curto período para ser iniciado. O Status na lista Notebooks (Blocos de anotações) mostra Starting (Iniciando). Você pode abrir um bloco de anotações quando seu status for Ready (Pronto). Pode demorar um pouco mais para um bloco de anotações entrar no status Ready (Pronto) se você tiver criado um cluster com ele.

> **Tip**
>
> Atualize o navegador ou escolha o ícone de atualização acima da lista de blocos de anotações para atualizar o status do bloco de anotações.

Noções básicas sobre o status do caderno

Um Caderno do EMR pode ter um dos Status apresentados a seguir na lista Cadernos.

Status	Significado
Ready	Você pode abrir o bloco de anotações usando o editor de blocos de anotações. Enquanto um bloco de anotações estiver no status Ready (Pronto), você poderá interrompê-lo ou excluí-lo. Para alterar clusters, você deve interromper o bloco de anotações primeiro. Se um bloco de anotações no status Ready (Pronto) ficar ocioso por muito tempo, ele será interrompido automaticamente.
Starting	O bloco de anotações está sendo criado e conectado ao cluster. Enquanto um bloco de anotações estiver sendo iniciado, você não poderá abrir o editor de blocos de anotações, interromper, excluir nem alterar clusters.

Status	Significado
Pendente	O bloco de anotações foi criado e está aguardando a integração com o cluster para ser concluído. O cluster ainda pode estar provisionamento recursos ou respondendo a outras solicitações. Você pode abrir o editor de blocos de anotações com o bloco de anotações no modo local. Qualquer código que se baseie em processos de cluster não será executado e falhará.
Parando	O bloco de anotações está sendo desligado ou o cluster ao qual o bloco de anotações está sendo anexado está sendo encerrado. Enquanto um bloco de anotações estiver sendo interrompido, você não poderá abrir o editor de blocos de anotações, interromper, excluir nem alterar clusters.
Interrompido	O bloco de anotações foi encerrado. Você pode iniciar o bloco de anotações no mesmo cluster, desde que o cluster ainda esteja em execução. Você pode alterar os clusters e excluir o cluster.
Excluindo	O cluster é removido da lista de clusters disponíveis. O arquivo de caderno *NotebookName* `.ipynb` permanece no Amazon S3 e continua acumulando as cobranças de armazenamento aplicáveis.

Como trabalhar com o editor de cadernos

Uma vantagem de usar um notebook EMR é que você pode iniciar o notebook no Jupyter ou JupyterLab diretamente do console.

Com o EMR Notebooks, o editor de notebook que você acessa do console do Amazon EMR é o conhecido editor de notebook Jupyter de código aberto ou. JupyterLab Como o editor de cadernos é iniciado no console do Amazon EMR, é mais eficiente para configurar o acesso do que com um caderno hospedado em um cluster do Amazon EMR. Você não precisa configurar um cliente do usuário para ter acesso à web por meio de SSH, regras de grupo de segurança e configurações de proxy. Se um usuário tiver permissões suficientes, ele poderá simplesmente abrir o editor de cadernos no console do Amazon EMR.

Somente um usuário pode ter um Caderno do EMR aberto por vez no Amazon EMR. Se outro usuário tentar abrir um Caderno do EMR que já esteja aberto, ocorrerá um erro.

> **Important**
>
> O Amazon EMR cria um URL assinado previamente exclusivo para cada sessão do editor de cadernos, que é válido somente por um curto período. Recomendamos que você não compartilhe o URL do editor de bloco de anotações. Isso cria um risco à segurança porque os destinatários do URL adotam suas permissões para editar e executar o código do bloco de anotações durante a vida útil do URL. Se outras pessoas precisarem de acesso a um caderno, forneça permissões ao usuário por meio de políticas de permissões e garanta que o perfil de serviço dos Cadernos do EMR tenha acesso ao local do Amazon S3. Para obter mais informações, consulte the section called “Segurança” e Perfil de serviço para Cadernos do EMR.

Abrir o editor de cadernos para um Caderno do EMR

1. Selecione um bloco de anotações com um Status de Ready (Pronto) ou Pending (Pendente) na lista Notebooks (Blocos de anotações).
2. Escolha Abrir no JupyterLab ou Abrir no Jupyter.

   Uma nova guia do navegador é aberta para o editor JupyterLab ou o editor do Jupyter Notebook.
3. No menu Kernel, escolha Change kernel (Alterar kernel) e, em seguida, selecione o kernel para sua linguagem de programação.

   Agora você está pronto para gravar e executar código de dentro do editor de blocos de anotações.

Como salvar o conteúdo de um caderno

Ao trabalhar no editor de cadernos, o conteúdo das células e as saídas do caderno são salvos automaticamente no arquivo de caderno no Amazon S3, de forma periódica. Um bloco de anotações que não tem alterações desde a última vez em que uma célula foi editada mostrará (autosaved) ao lado do nome do bloco de anotações no editor. Se as alterações ainda não foram salvas, unsaved changes (alterações não salvas) será exibido.

Você pode salvar um bloco de anotações manualmente. No menu Arquivo, escolha Salvar e ponto de verificação ou pressione CTRL+S. Isso cria um arquivo chamado *NotebookName*`.ipynb` em uma pasta de pontos de verificação dentro da pasta do caderno no Amazon S3. Por exemplo, `s3://`*MyBucket*`/`*MyNotebookFolder*`/`*NotebookID*`/checkpoints/`*NotebookName*`.ipynb`. Somente o arquivo de pontos de verificação mais recente é salvo nesse local.

Como alterar clusters

Você pode alterar o cluster ao qual um Caderno do EMR está anexado sem alterar o conteúdo do próprio caderno. Você pode alterar clusters apenas para os blocos de anotações que têm o status Stopped (Interrompido).

Alterar o cluster de um Caderno do EMR

1. Se o bloco de anotações que você deseja alterar estiver em execução, selecione-o na lista Notebooks (Blocos de anotações) e escolha Stop (Interromper).
2. Quando o status do bloco de anotações for Stopped (interrompido), selecione o bloco de anotações na lista Notebooks (Blocos de anotações) e, em seguida, escolha View details (Exibir detalhes).
3. Escolha Change cluster (Alterar cluster).
4. Se você tiver um cluster ativo com o Hadoop, Spark e Livy em execução ao qual você deseje anexar o bloco de anotações, deixe o padrão e selecione um cluster na lista. Somente clusters que atendam aos requisitos listados.

   —ou—

   Selecione Create a cluster (Criar um cluster) e escolha as opções de cluster. Para ter mais informações, consulte Requisitos de cluster.
5. Escolha uma opção para Security groups (Grupos de segurança) e, em seguida, escolha Change cluster and start notebook (Alterar cluster e iniciar bloco de anotações).

Como excluir cadernos e arquivos de cadernos

Ao excluir um Caderno do EMR usando o console do Amazon EMR, você exclui o caderno da lista de cadernos disponíveis. No entanto, os arquivos de cadernos permanecem no Amazon S3 e continuam a acumular as cobranças de armazenamento.

Para excluir um bloco de anotações e remover arquivos associados

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha Notebooks (Blocos de anotações), selecione seu bloco de anotações na lista e, em seguida, escolha View details (Exibir detalhes).
3. Escolha o ícone de pasta ao lado de Notebook location (Local do bloco de anotações) e copie o URL, que está no padrão `s3://`*`MyNotebookLocationPath`*`/`*`NotebookID`*`/`.
4. Escolha Excluir.

   O bloco de anotações é removido da lista e os detalhes de bloco de anotações deixam de aparecer.
5. Siga as instruções fornecidas em How do I delete folders from an S3 bucket? no Guia do usuário do Amazon Simple Storage Service. Navegue até o bucket e a pasta na etapa 3.

   —ou—

   Se você tiver o AWS CLI instalado, abra um prompt de comando e digite o comando no final deste parágrafo. Substitua o local do Amazon S3 pelo local que você copiou acima. Certifique-se de que AWS CLI esteja configurado com as chaves de acesso de um usuário com permissões para excluir a localização do Amazon S3. Para obter mais informações, consulte Configuração da AWS CLI no Guia do usuário da AWS Command Line Interface .

   ```
   aws s3 rm s3://MyNotebookLocationPath/NotebookID
   ```

Como compartilhar arquivos de cadernos

Cada Caderno do EMR é salvo no Amazon S3 como um arquivo chamado *`NotebookName`*`.ipynb`. Contanto que um arquivo de caderno seja compatível com a mesma versão do caderno Jupyter em que os Cadernos do EMR se baseiam, você poderá abrir o caderno como um Caderno do EMR.

A maneira mais fácil de abrir um arquivo de notebook de outro usuário é salvar o arquivo*.ipynb de outro usuário no sistema de arquivos local e, em seguida, usar o recurso de upload no Jupyter e nos editores. JupyterLab

É possível recorrer a esse processo para usar blocos de anotações do EMR compartilhados por outros, blocos de anotações compartilhados na comunidade do Jupyter ou para restaurar um bloco de anotações que foi excluído do console quando você ainda tinha o arquivo de bloco de anotações.

Usar um arquivo de caderno diferente como base para um Caderno do EMR

1. Antes de continuar, feche o editor de cadernos de todos os cadernos com os quais você trabalhará e, em seguida, interrompa o caderno se for um Caderno do EMR.
2. Crie um Caderno do EMR e insira um nome para ele. O nome que você inserir para o bloco de anotações será o nome do arquivo que você precisará substituir. O novo nome de arquivo deve corresponder exatamente ao nome desse arquivo.
3. Anote o local no Amazon S3 que você escolheu para o caderno. O arquivo que você substituir está em uma pasta com um caminho e nome de arquivo como o padrão a seguir: `s3://MyNotebookLocation/NotebookID/MyNotebookName.ipynb`.
4. Interrompa o bloco de anotações.
5. Substitua o antigo arquivo de caderno no local do Amazon S3 pelo novo, usando exatamente o mesmo nome.

   O AWS CLI comando a seguir para o Amazon S3 substitui um arquivo salvo em uma máquina local chamada para `SharedNotebook.ipynb` um notebook EMR pelo nome MyNotebook, um ID de `e-12A3BCDEFJHIJKLMNO45PQRST` e criado com o especificado no `MyBucket/MyNotebooksFolder` Amazon S3. Para obter informações sobre como usar o console do Amazon S3 para copiar e substituir arquivos, consulte Fazer upload, fazer download e trabalhar com objetos no Guia do usuário do Amazon Simple Storage Service.

   ```
   aws s3 cp SharedNotebook.ipynb s3://MyBucket/
   MyNotebooksFolder/-12A3BCDEFJHIJKLMNO45PQRST/MyNotebook.ipynb
   ```

Exemplos de comandos para executar Cadernos do EMR programaticamente

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Visão geral

Você pode executar Cadernos do EMR com APIs de execução usando um script ou a linha de comando. Quando você inicia, interrompe, lista e descreve as execuções do notebook EMR fora do AWS console, você pode controlar programaticamente um notebook EMR. É possível transferir valores de parâmetros diferentes para um caderno com uma célula de caderno parametrizada. Isto elimina a necessidade de criar uma cópia do caderno para cada novo conjunto de valores de parâmetros. Para obter mais informações, consulte Amazon EMR API actions.

Você pode agendar ou agrupar execuções de notebooks EMR com eventos da Amazon e. CloudWatch AWS Lambda Para obter mais informações, consulte Usando AWS Lambda com Amazon CloudWatch Events.

Permissões de perfil para a execução programática

Para usar a execução programática com os Cadernos do EMR, você deve configurar as permissões de usuário com as seguintes políticas:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowExecutionActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:StartNotebookExecution",
```

```
                "elasticmapreduce:DescribeNotebookExecution",
                "elasticmapreduce:ListNotebookExecutions"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AllowPassingServiceRole",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "arn:aws:iam::account-id:role/EMR_Notebooks_DefaultRole"
        }
    ]
}
```

Ao executar Cadernos do EMR programaticamente em um cluster de Cadernos do EMR, você deve adicionar estas permissões adicionais:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowRetrievingManagedEndpointCredentials",
            "Effect": "Allow",
            "Action": [
                "emr-containers:GetManagedEndpointSessionCredentials"
            ],
            "Resource": [
                "arn:aws:emr-containers:region:account-id:/virtualclusters/virtual-
cluster-id/endpoints/managed-endpoint-id"
            ],
            "Condition": {
                "StringEquals": {
                    "emr-containers:ExecutionRoleArn": [
                        "arn:aws:iam::account-id:role/emr-on-eks-execution-role"
                    ]
                }
            }
        },
        {
            "Sid": "AllowDescribingManagedEndpoint",
            "Effect": "Allow",
```

```
            "Action": [
                "emr-containers:DescribeManagedEndpoint"
            ],
            "Resource": [
                "arn:aws:emr-containers:region:account-id:/virtualclusters/virtual-
cluster-id/endpoints/managed-endpoint-id"
            ]
        }
    ]
}
```

Limitações da execução programática

- Há suporte para um máximo de 100 execuções simultâneas Região da AWS por conta.
- Uma execução será encerrada se for executada por mais de 30 dias.
- A execução programática de cadernos não é compatível com as aplicações interativas do Amazon EMR Serverless.

Exemplos de execução programática para Cadernos do EMR

As seções a seguir fornecem vários exemplos de execução programática de notebooks EMR com AWS CLI o Boto3 SDK (Python) e Ruby:

- Exemplos de comandos da CLI de execução de cadernos
- Exemplos de Python de execução de cadernos
- Exemplos de Ruby de execução de cadernos

Você também pode executar cadernos parametrizados como parte dos fluxos de trabalho programados com uma ferramenta de orquestração, como o Apache Airflow ou o Amazon Managed Workflows for Apache Airflow (MWAA). Para obter mais informações, consulte Orchestrating analytics jobs on EMR Notebooks using MWAA no blog de Big Data da AWS .

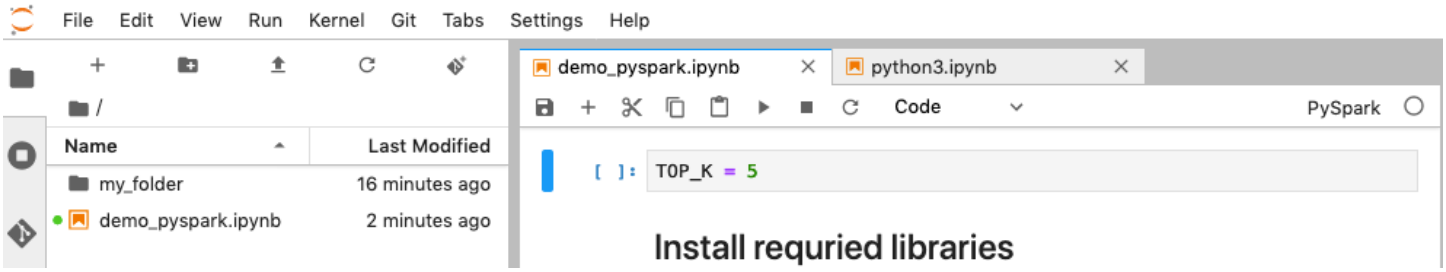
Exemplos de comandos da CLI de execução de cadernos

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou

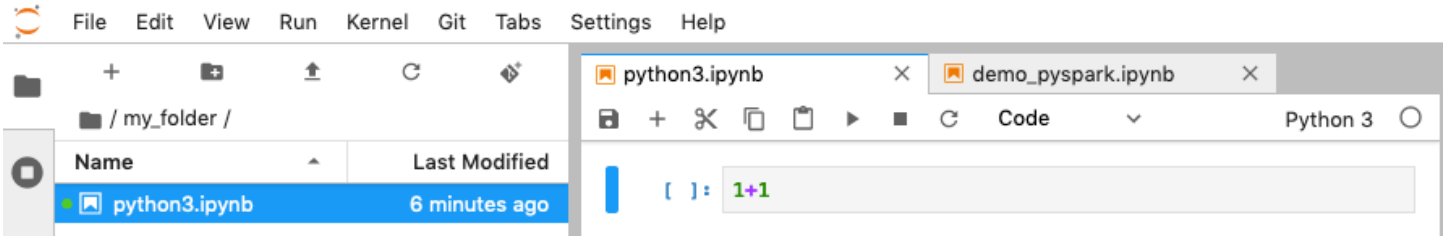
criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

O exemplo a seguir usa o caderno de demonstração do console de Cadernos do EMR. Para localizar o caderno, use o caminho do arquivo relativo ao diretório inicial. Neste exemplo, há dois arquivos de cadernos que você pode executar: `demo_pyspark.ipynb` e `my_folder/python3.ipynb`.

O caminho relativo para o arquivo `demo_pyspark.ipynb` é `demo_pyspark.ipynb`, como apresentado abaixo.



O caminho relativo para `python3.ipynb` é `my_folder/python3.ipynb`, como apresentado abaixo.



Para obter informações sobre as ações da API `NotebookExecution` do Amazon EMR, consulte Amazon EMR API actions.

Execução de um caderno

Você pode usar o AWS CLI para executar seu notebook com a `start-notebook-execution` ação, conforme demonstrado nos exemplos a seguir.

Example : execução de um Caderno do EMR em um Workspace do EMR Studio com um cluster do Amazon EMR (em execução no Amazon EC2)

```
aws emr --region us-east-1 \
start-notebook-execution \
```

```
--editor-id e-ABCDEFG123456 \
--notebook-params '{"input_param":"my-value", "good_superhero":["superman", "batman"]}'
 \
--relative-path test.ipynb \
--notebook-execution-name my-execution \
--execution-engine '{"Id" : "j-1234ABCD123"}' \
--service-role EMR_Notebooks_DefaultRole

{
    "NotebookExecutionId": "ex-ABCDEFGHIJ1234ABCD"
}
```

Example : execução de um Caderno do EMR em um Workspace do EMR Studio com um cluster dos Cadernos do EMR

```
aws emr start-notebook-execution \
    --region us-east-1 \
    --service-role EMR_Notebooks_DefaultRole \
    --environment-variables '{"KERNEL_EXTRA_SPARK_OPTS": "--conf
 spark.executor.instances=1", "KERNEL_LAUNCH_TIMEOUT": "350"}' \
    --output-notebook-format HTML \
    --execution-engine Id=arn:aws:emr-containers:us-west-2:account-id:/
virtualclusters/ABCDEFG/
endpoints/ABCDEF,Type=EMR_ON_EKS,ExecutionRoleArn=arn:aws:iam::account-
id:role/execution-role \
    --editor-id e-ABCDEFG \
    --relative-path EMRonEKS-spark_python.ipynb
```

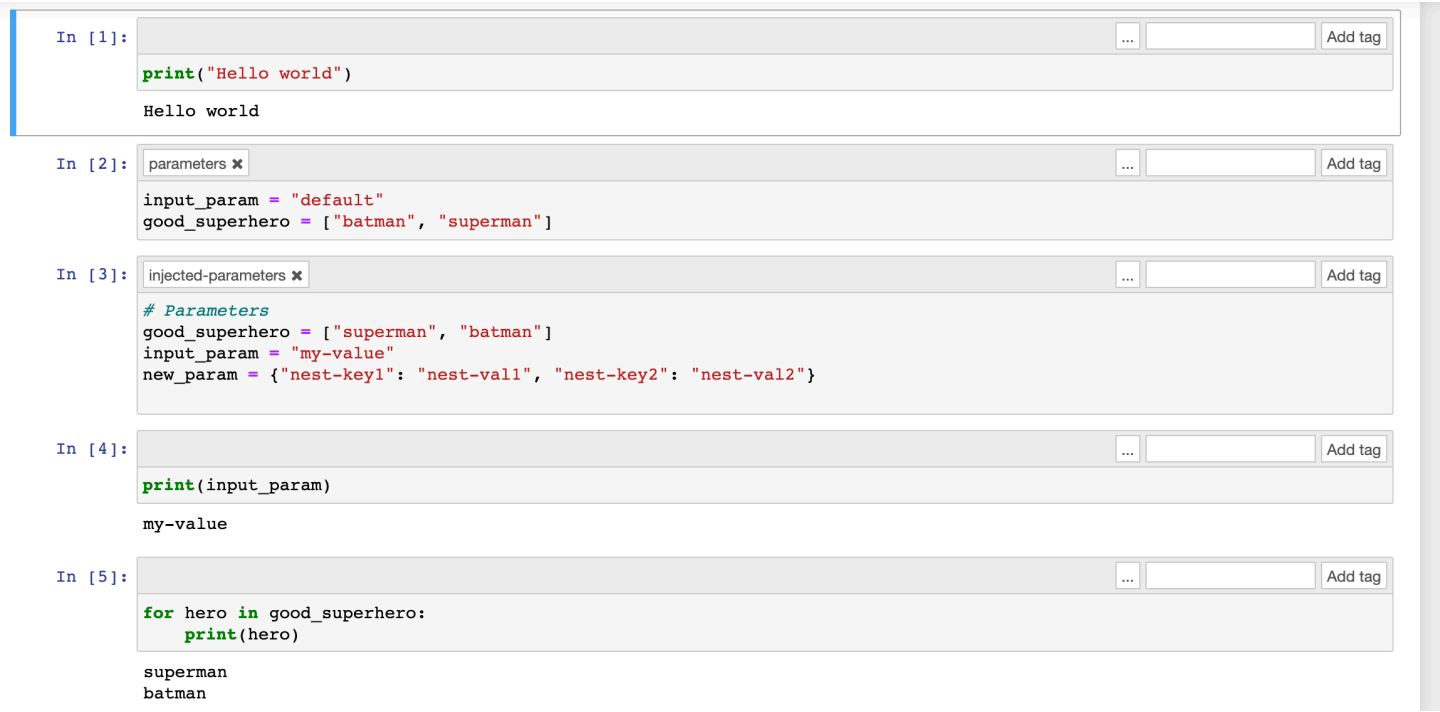
Example : execução de um Caderno do EMR com o local do Amazon S3 especificado

```
aws emr start-notebook-execution \
    --region us-east-1 \
    --notebook-execution-name my-execution-on-emr-on-eks-cluster \
    --service-role EMR_Notebooks_DefaultRole \
    --environment-variables '{"KERNEL_EXTRA_SPARK_OPTS": "--conf
 spark.executor.instances=1", "KERNEL_LAUNCH_TIMEOUT": "350"}' \
    --output-notebook-format HTML \
    --execution-engine Id=arn:aws:emr-containers:us-west-2:account-id:/
virtualclusters/ABCDEF/
endpoints/ABCDEF,Type=EMR_ON_EKS,ExecutionRoleArn=arn:aws:iam::account-
id:role/execution-role \
    --notebook-s3-location '{"Bucket": "your-s3-bucket","Key": "s3-prefix-to-notebook-
location/EMRonEKS-spark_python.ipynb"}' \
```

```
    --output-notebook-s3-location '{"Bucket": "your-s3-bucket","Key": "s3-prefix-for-storing-output-notebook"}'
```

Saída de bloco de anotações

Confira o resultado de um caderno de exemplo. A célula 3 mostra os valores dos parâmetros injetados recentemente.



Descrição de um caderno

Você pode usar a ação `describe-notebook-execution` para acessar informações sobre a execução de um caderno específico.

```
aws emr --region us-east-1 \
describe-notebook-execution --notebook-execution-id ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE

{
    "NotebookExecution": {
        "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
        "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
        "ExecutionEngine": {
            "Id": "j-2QMOV6JAX1TS2",
            "Type": "EMR",
            "MasterInstanceSecurityGroupId": "sg-05ce12e58cd4f715e"
```

```
        },
        "NotebookExecutionName": "my-execution",
        "NotebookParams": "{\"input_param\":\"my-value\", \"good_superhero\":
[\"superman\", \"batman\"]}",
        "Status": "FINISHED",
        "StartTime": 1593490857.009,
        "Arn": "arn:aws:elasticmapreduce:us-east-1:123456789012:notebook-execution/ex-
IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
        "LastStateChangeReason": "Execution is finished for cluster j-2QMOV6JAX1TS2.",
        "NotebookInstanceSecurityGroupId": "sg-0683b0a39966d4a6a",
        "Tags": []
    }
}
```

Interrupção de um caderno

Se o seu caderno estiver executando uma execução que você gostaria de interromper, poderá fazer isso com o comando `stop-notebook-execution`.

```
# stop a running execution
aws emr --region us-east-1 \
stop-notebook-execution --notebook-execution-id ex-IZWZX78UVPAATC8LHJR129B1RBN4T


# describe it
aws emr --region us-east-1 \
describe-notebook-execution --notebook-execution-id ex-IZWZX78UVPAATC8LHJR129B1RBN4T

{
    "NotebookExecution": {
        "NotebookExecutionId": "ex-IZWZX78UVPAATC8LHJR129B1RBN4T",
        "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
        "ExecutionEngine": {
            "Id": "j-2QMOV6JAX1TS2",
            "Type": "EMR"
        },
        "NotebookExecutionName": "my-execution",
        "NotebookParams": "{\"input_param\":\"my-value\", \"good_superhero\":
[\"superman\", \"batman\"]}",
        "Status": "STOPPED",
        "StartTime": 1593490876.241,
        "Arn": "arn:aws:elasticmapreduce:us-east-1:123456789012:editor-execution/ex-
IZWZX78UVPAATC8LHJR129B1RBN4T",
```

```
        "LastStateChangeReason": "Execution is stopped for cluster j-2QMOV6JAX1TS2.
 Internal error",
        "Tags": []
    }
}
```

Listagem das execuções de um caderno por horário de início

Você pode transferir um parâmetro `--from` para `list-notebook-executions` com a finalidade de listar as execuções do caderno por horário de início.

```
# filter by start time
aws emr --region us-east-1 \
list-notebook-executions --from 1593400000.000

{
    "NotebookExecutions": [
        {
            "NotebookExecutionId": "ex-IZWZX78UVPAATC8LHJR129B1RBN4T",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "STOPPED",
            "StartTime": 1593490876.241
        },
        {
            "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "RUNNING",
            "StartTime": 1593490857.009
        },
        {
            "NotebookExecutionId": "ex-IZWZYRS0M14L5V95WZ9OQ399SKMNW",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "STOPPED",
            "StartTime": 1593490292.995
        },
        {
            "NotebookExecutionId": "ex-IZX009ZK83IVY5E33VH8MDMELVK8K",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
```

```
            "StartTime": 1593489834.765
        },
        {
            "NotebookExecutionId": "ex-IZWZXOZF88JWDF9J09GJ91R57VI0N",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FAILED",
            "StartTime": 1593488934.688
        }
    ]
}
```

Listagem das execuções de um caderno por horário de início e status

O comando `list-notebook-executions` também pode usar um parâmetro `--status` para filtrar os resultados.

```
# filter by start time and status
aws emr --region us-east-1 \
list-notebook-executions --from 1593400000.000 --status FINISHED
{
    "NotebookExecutions": [
        {
            "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
            "StartTime": 1593490857.009
        },
        {
            "NotebookExecutionId": "ex-IZX009ZK83IVY5E33VH8MDMELVK8K",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
            "StartTime": 1593489834.765
        }
    ]
}
```

Exemplos de Python de execução de cadernos

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

O exemplo de código a seguir corresponde a um arquivo SDK para Python (Boto3) chamado `demo.py` que mostra as APIs de execução do caderno.

Para obter informações sobre as ações da API `NotebookExecution` do Amazon EMR, consulte Amazon EMR API actions.

```
import boto3,time

emr = boto3.client(
    'emr',
    region_name='us-west-1'
)

start_resp = emr.start_notebook_execution(
    EditorId='e-40AC8ZO6EGGCPJ4DLO48KGGGI',
    RelativePath='boto3_demo.ipynb',
    ExecutionEngine={'Id':'j-1HYZS6JQKV11Q'},
    ServiceRole='EMR_Notebooks_DefaultRole'
)

execution_id = start_resp["NotebookExecutionId"]
print(execution_id)
print("\n")

describe_response = emr.describe_notebook_execution(NotebookExecutionId=execution_id)

print(describe_response)
print("\n")

list_response = emr.list_notebook_executions()
```

```
print("Existing notebook executions:\n")
for execution in list_response['NotebookExecutions']:
    print(execution)
    print("\n")

print("Sleeping for 5 sec...")
time.sleep(5)

print("Stop execution " + execution_id)
emr.stop_notebook_execution(NotebookExecutionId=execution_id)
describe_response = emr.describe_notebook_execution(NotebookExecutionId=execution_id)
print(describe_response)
print("\n")
```

Confira o resultado da execução do arquivo `demo.py`.

```
ex-IZX56YJDW1D29Q1PHR32WABU2SAPK

{'NotebookExecution': {'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK',
 'EditorId': 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'ExecutionEngine': {'Id':
 'j-1HYZS6JQKV11Q', 'Type': 'EMR'}, 'NotebookExecutionName': '', 'Status': 'STARTING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal()),
 'Arn': 'arn:aws:elasticmapreduce:us-west-1:123456789012:notebook-execution/ex-
IZX56YJDW1D29Q1PHR32WABU2SAPK', 'LastStateChangeReason': 'Execution is starting
 for cluster j-1HYZS6JQKV11Q.', 'Tags': []}, 'ResponseMetadata': {'RequestId':
 '70f12c5f-1dda-45b7-adf6-964987d373b7', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-
amzn-requestid': '70f12c5f-1dda-45b7-adf6-964987d373b7', 'content-type': 'application/
x-amz-json-1.1', 'content-length': '448', 'date': 'Wed, 19 Aug 2020 00:49:22 GMT'},
 'RetryAttempts': 0}}

Existing notebook executions:

{'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'STARTING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5ABS5PR1E5AHMFYEMX3JJIORRB', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'RUNNING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 48, 36, 373000, tzinfo=tzlocal())}
```

```
{'NotebookExecutionId': 'ex-IZX5GLVXIU1HNI8BWVW057F6MF4VE', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 45, 14, 646000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 46, 26, 543000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5CV8YDUO8JAIWMXN2VH32RUIT1', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 43, 5, 807000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 44, 31, 632000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5AS0PPW55CEDEURZ9NSOWSUJZ6', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 42, 29, 265000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 43, 48, 320000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX57YF5Q53BKWLR4I5QZ14HJ7DRS', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 38, 37, 81000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 40, 39, 646000, tzinfo=tzlocal())}

Sleeping for 5 sec...
Stop execution ex-IZX56YJDW1D29Q1PHR32WABU2SAPK
{'NotebookExecution': {'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK',
 'EditorId': 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'ExecutionEngine': {'Id':
 'j-1HYZS6JQKV11Q', 'Type': 'EMR'}, 'NotebookExecutionName': '', 'Status': 'STOPPING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal()),
 'Arn': 'arn:aws:elasticmapreduce:us-west-1:123456789012:notebook-execution/ex-
IZX56YJDW1D29Q1PHR32WABU2SAPK', 'LastStateChangeReason': 'Execution is being stopped
 for cluster j-1HYZS6JQKV11Q.', 'Tags': []}, 'ResponseMetadata': {'RequestId':
 '2a77ef73-c1c6-467c-a1d1-7204ab2f6a53', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-
amzn-requestid': '2a77ef73-c1c6-467c-a1d1-7204ab2f6a53', 'content-type': 'application/
x-amz-json-1.1', 'content-length': '453', 'date': 'Wed, 19 Aug 2020 00:49:30 GMT'},
 'RetryAttempts': 0}}
```

Exemplos de Ruby de execução de cadernos

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou

criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

A seguir são apresentados exemplos de código do Ruby que demonstram o uso da API de execução do caderno.

```
# prepare an Amazon EMR client

emr = Aws::EMR::Client.new(
  region: 'us-east-1',
  access_key_id: 'AKIA...JKPKA',
  secret_access_key: 'rLMeu...vU0OLrAC1',
)
```

Início da execução do caderno e obtenção do ID de execução

Neste exemplo, o editor do Amazon S3 e o Caderno do EMR são `s3://mybucket/notebooks/e-EA8VGAA429FEQTC8HC9ZHWISK/test.ipynb`.

Para obter informações sobre as ações da API `NotebookExecution` do Amazon EMR, consulte Amazon EMR API actions.

```
start_response = emr.start_notebook_execution({
    editor_id: "e-EA8VGAA429FEQTC8HC9ZHWISK",
    relative_path: "test.ipynb",

    execution_engine: {id: "j-3U82I95AMALGE"},

    service_role: "EMR_Notebooks_DefaultRole",
})


notebook_execution_id = start_resp.notebook_execution_id
```

Descrição da execução do caderno e impressão dos detalhes

```
describe_resp = emr.describe_notebook_execution({
    notebook_execution_id: notebook_execution_id
```

```
})
puts describe_resp.notebook_execution
```

A saída dos comandos definidos acima será semelhante a apresentada a seguir.

```
{
:notebook_execution_id=>"ex-IZX3VTVZWVWPP27KUB90BZ7V9IEDG",
:editor_id=>"e-EA8VGAA429FEQTC8HC9ZHWISK",
:execution_engine=>{:id=>"j-3U82I95AMALGE", :type=>"EMR", :master_instance_security_group_id=>n
:notebook_execution_name=>"",
:notebook_params=>nil,
:status=>"STARTING",
:start_time=>2020-07-23 15:07:07 -0700,
:end_time=>nil,
:arn=>"arn:aws:elasticmapreduce:us-east-1:123456789012:notebook-execution/ex-
IZX3VTVZWVWPP27KUB90BZ7V9IEDG",
:output_notebook_uri=>nil,
:last_state_change_reason=>"Execution is starting for cluster
 j-3U82I95AMALGE.", :notebook_instance_security_group_id=>nil,
:tags=>[]
}
```

Filtros para cadernos

```
"EditorId": "e-XXXX",           [Optional]
"From" : "1593400000.000",    [Optional]
"To" :
```

Interrupção da execução do caderno

```
stop_resp = emr.stop_notebook_execution({
    notebook_execution_id: notebook_execution_id
})
```

Habilitação da representação do usuário para monitorar a atividade de usuários e trabalhos do Spark

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Os Cadernos do EMR permitem configurar a representação do usuário em um cluster do Spark. Esse recurso ajuda a rastrear a atividade do trabalho iniciado no editor de blocos de anotações. Além disso, os Cadernos do EMR têm um widget do caderno Jupyter integrado para visualizar detalhes do trabalho do Spark junto com a saída da consulta no editor de cadernos. O widget está disponível por padrão e não requer configuração especial. No entanto, para visualizar os servidores de históricos, o cliente deve estar configurado para visualizar interfaces Web do Amazon EMR hospedadas no nó primário.

Configuração da representação do usuário do Spark

Por padrão, os trabalhos do Spark que os usuários enviam usando o editor de blocos de anotações parecem se originar de uma identidade de usuário `livy` indiscriminada. Você pode configurar a representação do usuário para o cluster para que esses trabalhos sejam associados à identidade de usuário que executou o código. Os diretórios de usuários do HDFS no nó primário são criados para cada identidade de usuário que executa códigos no caderno. Por exemplo, se o usuário `NbUser1` executar o código do editor de cadernos, é possível se conectar ao nó primário e ver que `hadoop fs -ls /user` mostra o diretório `/user/user_NbUser1`.

Para habilitar esse recurso, configure as propriedades nas classificações de configuração `livy-conf` e `core-site`. Esse recurso não está disponível por padrão quando o Amazon EMR cria um cluster em conjunto com um caderno. Para obter mais informações sobre como usar classificações de configuração para personalizar aplicações, consulte Configuring applications no Guia de lançamento do Amazon EMR.

Use as seguintes classificações e valores de configuração para habilitar a representação do usuário para os Cadernos do EMR:

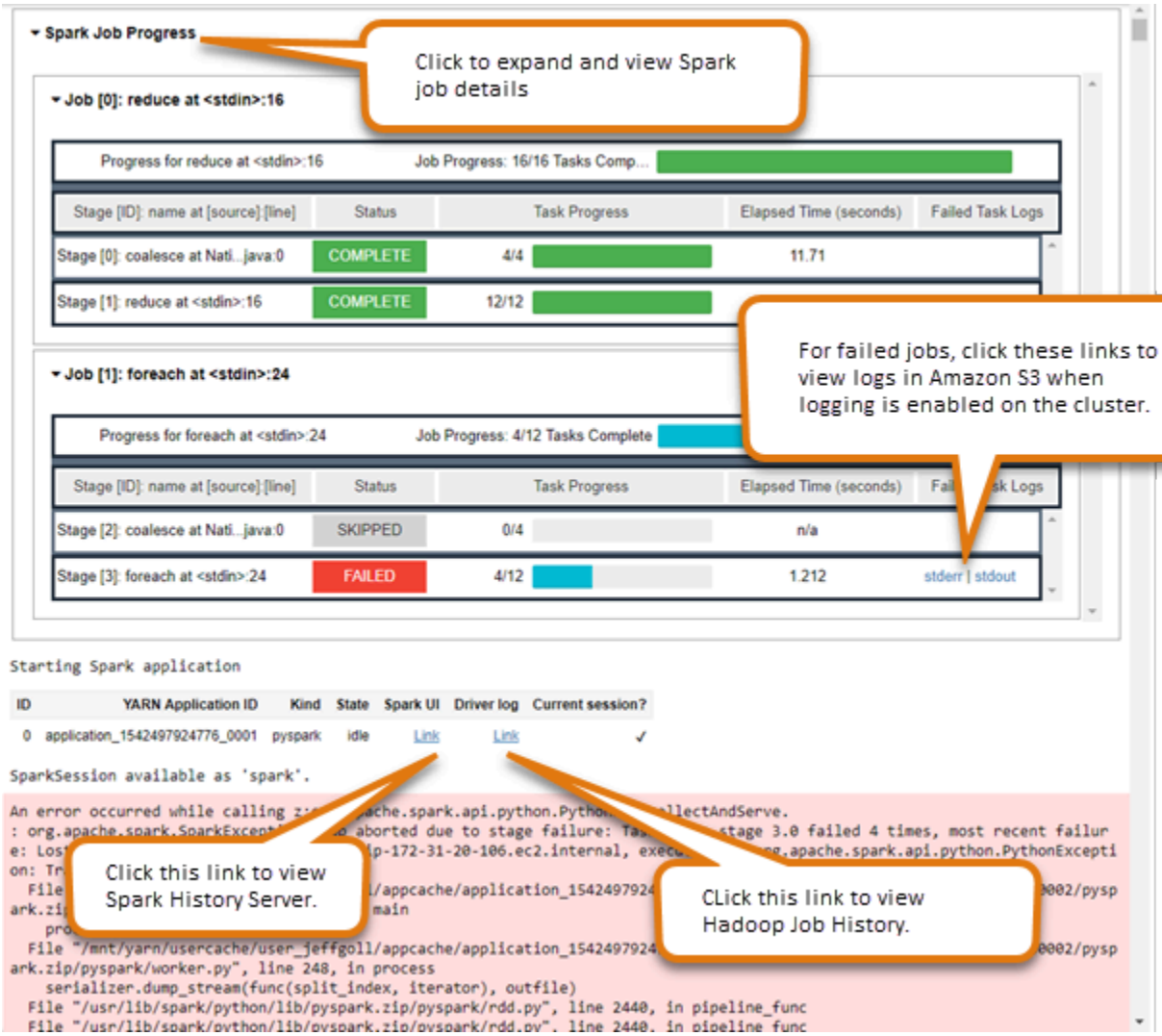
```
[
    {
        "Classification": "core-site",
        "Properties": {
          "hadoop.proxyuser.livy.groups": "*",
          "hadoop.proxyuser.livy.hosts": "*"
        }
    },
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.impersonation.enabled": "true"
        }
    }
]
```

Uso do widget de monitoramento de trabalhos do Spark

Quando você executa código no editor de blocos de anotações que executam trabalhos do Spark no cluster do EMR, a saída inclui um widget de bloco de anotações Jupyter para monitoramento de trabalhos do Spark. O widget fornece detalhes do trabalho e links úteis para a página do servidor de histórico do Spark e para a página de histórico de trabalhos do Hadoop, além de links convenientes para logs de trabalho no Amazon S3 para todos os trabalhos com falha.

Para visualizar as páginas do servidor de histórico no nó primário do cluster, você deve configurar um cliente SSH e um proxy, conforme apropriado. Para ter mais informações, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR. Para visualizar os logs no Amazon S3, o registro em log do cluster deve estar habilitado, que é o padrão para os novos clusters. Para ter mais informações, consulte Visualizar arquivos de log arquivados no Amazon S3.

A seguir é apresentado um exemplo de monitoramento de trabalhos do Spark.



Segurança e controle de acesso para Cadernos do EMR

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Diversos recursos estão disponíveis para ajudar você a personalizar a postura de segurança dos Cadernos do EMR. Isso ajuda a garantir que somente usuários autorizados tenham acesso

a um Caderno do EMR, e possam trabalhar com cadernos e usar o editor de cadernos para executar códigos no cluster. Esses recursos funcionam em conjunto com os recursos de segurança disponíveis para o Amazon EMR e para os clusters do Amazon EMR. Para ter mais informações, consulte Segurança no Amazon EMR.

- Você pode usar declarações AWS Identity and Access Management de política junto com etiquetas de caderno para limitar o acesso. Para obter mais informações, consulte Como o Amazon EMR funciona com o IAM e Exemplo de instruções de políticas baseadas em identidade para Cadernos do EMR.
- Os grupos de segurança do Amazon EC2 atuam como firewalls virtuais que controlam o tráfego de rede entre a instância primária do cluster e o editor de cadernos. Você pode usar valores padrão ou personalizar esses grupos de segurança. Para ter mais informações, consulte Especificar grupos de segurança do EC2 para Cadernos do EMR.
- Você especifica uma função de AWS serviço que determina quais permissões um notebook EMR tem ao interagir com outros serviços. AWS Para ter mais informações, consulte Perfil de serviço para Cadernos do EMR.

Instalação e uso de kernels e bibliotecas

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Cada Caderno do EMR vem com um conjunto de bibliotecas e kernels instalados previamente. Você poderá instalar bibliotecas e kernels adicionais em um cluster do EMR, se o cluster tiver acesso ao repositório no qual os kernels e as bibliotecas estão localizados. Por exemplo, para clusters em sub-redes privadas, talvez seja necessário configurar a conversão de endereços de rede (NAT) e fornecer um caminho para o cluster acessar o repositório PyPI público para instalar uma biblioteca. Para obter mais informações sobre como configurar o acesso externo para diferentes configurações de rede, consulte Cenários e exemplos no Guia do usuário da Amazon VPC.

Os aplicativos EMR Serverless vêm com as seguintes bibliotecas pré-instaladas para Python e: PySpark

- Bibliotecas Python – ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, scipy
- PySpark bibliotecas — ggplotmatplotlib,numpy,pandas,plotly,bokeh,scikit-learn,scipy, scipy

Instalação de kernels e de bibliotecas Python em um nó primário do cluster

Com o Amazon EMR versão 5.30.0 e posteriores, exceto a versão 6.0.0, é possível instalar bibliotecas e kernels Python adicionais no nó primário do cluster. Após a instalação, os kernels e as bibliotecas ficam disponíveis para qualquer usuário que execute um Caderno do EMR anexado ao cluster. As bibliotecas Python instaladas dessa forma estão disponíveis somente para processos em execução no nó primário. As bibliotecas não são instaladas nos nós principais ou de tarefas e não estão disponíveis para executores em execução nesses nós.

> **Note**
>
> Para as versões 5.30.1, 5.31.0 e 6.1.0 do Amazon EMR, você deve executar etapas adicionais para instalar kernels e bibliotecas no nó primário de um cluster.
> Para habilitar o recurso, faça o seguinte:
>
> 1. Certifique-se de que a política de permissões anexada ao perfil de serviço para os Cadernos do EMR permite a seguinte ação:
>
>    ```
>    elasticmapreduce:ListSteps
>    ```
>
>   Para obter mais informações, consulte Service role for EMR Notebooks.
> 2. Use o AWS CLI para executar uma etapa no cluster que configura os Notebooks EMR, conforme mostrado no exemplo a seguir. Você deve usar o nome da etapa `EMRNotebooksSetup`. Substitua *`us-east-1`* pela região em que seu cluster reside. Para obter mais informações, consulte Adding steps to a cluster using the AWS CLI.
>
>    ```
>    aws emr add-steps --cluster-id MyClusterID --steps
>     Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
>    east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
>    awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-
>    setup.sh"]
>    ```

Você pode instalar kernels e bibliotecas usando `pip` ou `conda` no diretório `/emr/notebook-env/bin` no nó primário.

Example : instalação de bibliotecas Python

No kernel do Python3, execute a mágica `%pip` como um comando de dentro de uma célula de caderno para instalar bibliotecas Python.

```
%pip install pmdarima
```

Pode ser necessário reiniciar o kernel para usar os pacotes atualizados. Você também pode usar a mágica `%%sh` do Spark para invocar `pip`.

```
%%sh
/emr/notebook-env/bin/pip install -U matplotlib
/emr/notebook-env/bin/pip install -U pmdarima
```

Ao usar um PySpark kernel, você pode instalar bibliotecas no cluster usando `pip` comandos ou usar bibliotecas com escopo de notebook de dentro de um notebook. PySpark

Para executar comandos `pip` no cluster do terminal, primeiro conecte-se ao nó primário usando SSH, como demonstram os comandos a seguir.

```
sudo pip3 install -U matplotlib
sudo pip3 install -U pmdarima
```

Como alternativa, você pode usar bibliotecas com escopo de cadernos. Com bibliotecas com escopo de cadernos, a instalação da sua biblioteca é limitada ao escopo da sua sessão e ocorre em todos os executores do Spark. Para obter mais informações, consulte Uso de bibliotecas com escopo de cadernos.

Se você quiser empacotar várias bibliotecas Python em um PySpark kernel, você também pode criar um ambiente virtual Python isolado. Para obter exemplos, consulte Uso de Virtualenv.

Para criar um ambiente virtual Python em uma sessão, use a propriedade `spark.yarn.dist.archives` do Spark do comando mágico `%%configure` na primeira célula de um caderno, como demonstra o exemplo a seguir.

```
%%configure -f
```

```
{
   "conf": {
   "spark.yarn.appMasterEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.appMasterEnv.PYSPARK_DRIVER_PYTHON":"./environment/bin/python",
   "spark.yarn.dist.archives":"s3://DOC-EXAMPLE-BUCKET/prefix/
my_pyspark_venv.tar.gz#environment",
   "spark.submit.deployMode":"cluster"
   }
}
```

De forma semelhante, você pode criar um ambiente de executor do Spark.

```
%%configure -f
{
   "conf": {
   "spark.yarn.appMasterEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.appMasterEnv.PYSPARK_DRIVER_PYTHON":"./environment/bin/python",
   "spark.executorEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.dist.archives":"s3://DOC-EXAMPLE-BUCKET/prefix/
my_pyspark_venv.tar.gz#environment",
   "spark.submit.deployMode":"cluster"
   }
}
```

Você também pode usar `conda` para instalar bibliotecas Python. Você não precisa de acesso ao sudo para usar `conda`. Você deve se conectar ao nó primário com SSH e, em seguida, executar `conda` do terminal. Para ter mais informações, consulte Conectar-se ao nó primário usando SSH.

Example : instalação de kernels

O seguinte exemplo demonstra a instalação do kernel do Kotlin usando um comando de terminal enquanto estiver conectado ao nó primário de um cluster:

```
sudo /emr/notebook-env/bin/conda install kotlin-jupyter-kernel -c jetbrains
```

> **Note**
>
> Estas instruções não instalam as dependências do kernel. Se o seu kernel tiver dependências de terceiros, talvez seja necessário realizar etapas adicionais de configuração antes de poder usar o kernel com o seu caderno.

Considerações e limitações com bibliotecas com escopo de cadernos

Considere o seguinte ao usar bibliotecas com escopo de cadernos:

- Bibliotecas com escopo de cadernos estão disponíveis para clusters criados com as versões 5.26.0 e superiores do Amazon EMR.
- Bibliotecas com escopo de notebook devem ser usadas somente com o kernel. PySpark
- Qualquer usuário pode instalar bibliotecas adicionais com escopo de bloco de anotações de dentro de uma célula de bloco de anotações. Essas bibliotecas só estão disponíveis para esse usuário do bloco de anotações durante uma única sessão do bloco de anotações. Se outros usuários precisarem das mesmas bibliotecas ou o mesmo usuário precisar das mesmas bibliotecas em uma sessão diferente, a biblioteca deverá ser reinstalada.
- Você pode desinstalar apenas as bibliotecas instaladas usando a API `install_pypi_package`. Não é possível desinstalar nenhuma biblioteca pré-instalada no cluster.
- Se as mesmas bibliotecas com versões diferentes estiverem instaladas no cluster e como bibliotecas com escopo de bloco de anotações, a versão da biblioteca com escopo de bloco de anotações substituirá a versão da biblioteca do cluster.

Como trabalhar com bibliotecas com escopo de cadernos

Para instalar bibliotecas, o cluster do Amazon EMR deve ter acesso ao repositório PyPI em que as bibliotecas estão localizadas.

Os exemplos a seguir demonstram comandos simples para listar, instalar e desinstalar bibliotecas de dentro de uma célula do notebook usando o PySpark kernel e as APIs. Para ver exemplos adicionais, consulte a postagem Instalar bibliotecas Python em um cluster em execução com Notebooks EMR no Big Data Blog. AWS

Example : listagem de bibliotecas atuais

O comando a seguir lista os pacotes Python disponíveis para a sessão de bloco de anotações Spark atual. Isso lista as bibliotecas instaladas no cluster e as bibliotecas com escopo de bloco de anotações.

```
sc.list_packages()
```

Example : instalação da biblioteca Celery

O comando a seguir instala a biblioteca Celery como uma biblioteca com escopo de bloco de anotações.

```
sc.install_pypi_package("celery")
```

Depois de instalar a biblioteca, o comando a seguir confirma que a biblioteca está disponível no driver e nos executores Spark.

```
import celery
sc.range(1,10000,1,100).map(lambda x: celery.__version__).collect()
```

Example : instalação da biblioteca Arrow com especificação de versão e de repositório

O comando a seguir instala a biblioteca Arrow como uma biblioteca com escopo de bloco de anotações, com uma especificação da versão da biblioteca e do URL do repositório.

```
sc.install_pypi_package("arrow==0.14.0", "https://pypi.org/simple")
```

Example : desinstalação de uma biblioteca

O comando a seguir desinstala a biblioteca Arrow, removendo-a como uma biblioteca com escopo de bloco de anotações da sessão atual.

```
sc.uninstall_package("arrow")
```

Associação de repositórios baseados em Git a Cadernos do EMR

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Você pode associar repositórios baseados em Git aos Cadernos do Amazon EMR para salvá-los em um ambiente com versão controlada. É possível associar até três repositórios a um bloco de anotações. Os seguintes serviços baseados em GIT são compatíveis:

- AWS CodeCommit
- GitHub
- Bitbucket
- GitLab

Associar repositórios baseados em GIT ao bloco de anotações tem os seguintes benefícios.

- Controle de versão: é possível registrar alterações de código em um sistema com controle de versão para poder analisar o histórico de alterações e revertê-las seletivamente.
- Colaboração: colegas que trabalham em diferentes cadernos podem compartilhar códigos por meio de repositórios remotos baseados em Git. Os blocos de anotações podem clonar ou mesclar código de repositórios remotos e retornar as alterações para esses repositórios remotos.
- Reutilização de código — Muitos notebooks Jupyter que demonstram técnicas de análise de dados ou aprendizado de máquina estão disponíveis em repositórios hospedados publicamente, como. GitHub É possível associar os blocos de anotações a um repositório para reutilizar os blocos de anotações Jupyter contidos em um repositório.

Para usar repositórios baseados em Git com Cadernos do EMR, adicione os repositórios como recursos no console do Amazon EMR, associe credenciais para os repositórios que requerem autenticação e vincule-os aos seus cadernos. É possível visualizar uma lista de repositórios armazenados em sua conta e detalhes sobre cada repositório no console do Amazon EMR. Você pode associar um repositório baseado em GIT existente a um bloco de anotações ao criá-lo.

Tópicos

Pré-requisitos e considerações

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Considere o apresentado a seguir ao planejar a integração de um repositório baseado em Git com os Cadernos do EMR.

AWS CodeCommit

Se você usa um CodeCommit repositório, deve usar as credenciais do Git e o HTTPS com. CodeCommit Chaves SSH e HTTPS com o auxiliar de AWS CLI credenciais não são compatíveis. CodeCommit não suporta tokens de acesso pessoal (PATs). Para obter mais informações, consulte Como usar o IAM com CodeCommit: credenciais do Git, chaves SSH e chaves de AWS acesso no Guia do usuário do IAM e Configuração para usuários de HTTPS usando credenciais do Git no Guia do usuário.AWS CodeCommit

Considerações sobre acesso e permissão

Antes de associar um repositório ao seu caderno, certifique-se de que o cluster, o perfil do IAM para Cadernos do EMR e os grupos de segurança tenham as configurações e as permissões corretas. Você também pode configurar repositórios baseados em Git hospedados em uma rede privada ao seguir as instruções em Configuração de um repositório Git hospedado de forma privada para Cadernos do EMR.

- Acesso à Internet do cluster: a interface de rede iniciada tem somente um endereço IP privado. Isso significa que o cluster ao qual o bloco de anotações se conecta deve estar em uma sub-rede privada com um gateway de conversão de endereço de rede (NAT) ou deve ser capaz de acessar a Internet por um gateway privado virtual. Para obter mais informações, consulte Amazon VPC options.

  Os grupos de segurança do bloco de anotações devem incluir uma regra de saída que permita ao bloco de anotações rotear tráfego para a Internet por meio do cluster. Recomendamos que você

crie seus próprios grupos de segurança. Para obter mais informações, consulte Specifying EC2 security groups for EMR Notebooks.

> **Important**
>
> Se a interface de rede for inicializada em uma sub-rede pública, não será possível ter uma comunicação com a Internet através de um gateway da Internet (IGW).

- Permissões para AWS Secrets Manager — Se você usa o Secrets Manager para armazenar segredos usados para acessar um repositório, eles the section called "Perfil de Cadernos do EMR" devem ter uma política de permissões anexada que permita a `secretsmanager:GetSecretValue` ação.

Configuração de um repositório Git hospedado de forma privada para Cadernos do EMR

Use as instruções apresentadas a seguir para configurar repositórios hospedados de forma privada para Cadernos do EMR. Você deve fornecer um arquivo de configuração com informações sobre os servidores DNS e Git. O Amazon EMR usa essas informações para configurar Cadernos do EMR que podem rotear o tráfego para seus repositórios hospedados de forma privada.

Pré-requisitos

Antes de configurar um repositório Git hospedado de forma privada para Cadernos do EMR, você deve ter o seguinte:

- Um Amazon S3 Control local onde os arquivos do seu notebook EMR serão salvos.

Configurar um ou mais repositórios Git hospedados de forma privada para Cadernos do EMR

1. Crie um arquivo de configuração usando o modelo fornecido. Inclua os seguintes valores para cada servidor Git que deseja especificar em sua configuração:

   - **`DnsServerIpV4`**: o endereço IPv4 do seu servidor DNS. Se você fornecer valores para `DnsServerIpV4` e `GitServerIpV4List`, o valor para `DnsServerIpV4` terá precedência e será usado para resolver seu `GitServerDnsName`.

> **Note**
>
> Para usar repositórios Git hospedados de forma privada, seu servidor DNS deve permitir o acesso de entrada de Cadernos do EMR. Recomendamos fortemente proteger o servidor DNS contra outros acessos não autorizados.

- **`GitServerDnsName`**: o nome DNS do seu servidor Git. Por exemplo, `"git.example.com"`.
- **`GitServerIpV4List`**: uma lista de endereços IPv4 que pertencem aos seus servidores Git.

```
[
    {
        "Type": "PrivatelyHostedGitConfig",
        "Value": [
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<enterprise.git.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            },
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<git.example.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            }
        ]
    }
]
```

2. Salve seu arquivo de configuração como `configuration.json`.
3. Faça o upload do arquivo de configuração no local de armazenamento designado do Amazon S3 em uma pasta chamada `life-cycle-configuration`. Por exemplo, se o local padrão do S3 for `s3://DOC-EXAMPLE-BUCKET/notebooks`, seu arquivo de configuração deverá estar localizado em `s3://DOC-EXAMPLE-BUCKET/notebooks/`**`life-cycle-configuration/configuration.json`**.

> **Important**
>
> Recomendamos fortemente restringir o acesso à pasta `life-cycle-configuration` somente para os administradores dos Cadernos do EMR e para o perfil de serviço dos Cadernos do EMR. Você também deve proteger `configuration.json` contra acesso não autorizado. Para obter instruções, consulte Controlar o acesso a um bucket com políticas de usuário ou Práticas recomendadas de segurança para o Amazon S3.

Para obter instruções sobre como fazer o upload, consulte Criar uma pasta e Fazer upload de objetos no Guia do usuário do Amazon Simple Storage Service.

Adição de um repositório baseado em Git ao Amazon EMR

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Consulte as seções apresentadas a seguir para obter informações sobre como adicionar um repositório baseado em Git a um Caderno do EMR no console antigo ou a um Workspace do EMR Studio no novo console.

New console

Como os Cadernos do EMR são Workspaces do EMR Studio no novo console, você pode seguir as instruções em Vinculação de repositórios baseados em Git a um Workspace do EMR Studio para associar, no máximo, três repositórios Git ao seu Workspace.

Como alternativa, você pode usar a extensão JupyterLab Git. Escolha o ícone Git na barra lateral esquerda do seu caderno JupyterLab para acessar a extensão. Para obter informações sobre a extensão, consulte o repositório GitHub jupyterlab-git.

Para associar um repositório Git a um Workspace, o administrador do Studio deve seguir etapas para configurar o Studio para permitir a vinculação do repositório Git. Para ter mais informações, consulte Estabelecimento de acesso e de permissões para repositórios baseados em Git.

Old console

Adicionar um repositório baseado em Git como um recurso na sua conta do Amazon EMR com o console antigo

1. Abra o console antigo do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce.
2. Escolha Git repositories (Repositórios Git) e depois Add repository (Adicionar repositório).
3. Em Nome do repositório, insira um nome a ser usado para o repositório no Amazon EMR.

   Os nomes só podem conter caracteres alfanuméricos, hifens (-) ou sublinhados (_).
4. Em Git repository URL (URL do repositório do Git), insira o URL do repositório. Ao usar um CodeCommit repositório, essa é a URL que é copiada quando você escolhe Clonar URL e depois Clonar HTTPS, por exemplo,. `https://git-codecommit.us-west-2.amazonaws.com/v1/repos/MyCodeCommitRepoName`
5. Em Branch (Ramificação), insira o nome de uma ramificação.
6. Em Git credentials (Credenciais do Git), escolha opções de acordo com as seguintes diretrizes: Você pode usar um nome de usuário e senha do Git ou um token de acesso pessoal (PAT) para a autenticação em seu repositório. Os Cadernos do EMR acessam suas credenciais do Git usando segredos armazenados no Secrets Manager.

> **Note**
>
> Se você usa um GitHub repositório, recomendamos que você use um token de acesso pessoal (PAT) para autenticar. A partir de 13 de agosto de 2021, não GitHub aceitaremos mais senhas ao autenticar operações do Git. Para obter mais informações, consulte a publicação Requisitos de autenticação de token para operações do Git no The GitHub Blog.

Opção	Descrição
Usar um segredo da AWS existente	Escolha esta opção se você já salvou suas credenciais como um segredo no Secrets Manager e, em seguida, selecione o nome do segredo na lista. Se você selecionar um segredo associado a um nome de usuário e senha do Git, o segredo deverá estar no formato `{"gitUsername": " MyUserName ", "gitPassword": " MyPassword "}`.

Opção	Descrição
Criar um novo segredo	Escolha esta opção para associar credenciais do Git existentes a um novo segredo criado no Secrets Manager. Execute um dos seguintes procedimentos com base nas credenciais do Git que você usar para o repositório. Se você usar um nome de usuário e uma senha do Git para acessar o repositório, selecione Nome de usuário e senha, insira o Nome do segredo a ser usado no Secrets Manager e, em seguida, insira o Nome de usuário e a Senha a serem associados ao segredo. OU Se você usar um token de acesso pessoal para acessar o repositório, selecione Token de acesso pessoal (PAT), insira o Nome do segredo a ser usado no Secrets Manager e, em seguida, insira seu token de acesso pessoal. Para obter mais informações, consulte Criação de um token de acesso pessoal para a linha de comando GitHub e Tokens de acesso pessoal para o Bitbucket. CodeCommit os repositórios não oferecem suporte a essa opção.
Usar um repositório público sem credencia is	Escolha esta opção para acessar um repositório público.

7. Escolha Adicionar repositório.

Atualização ou exclusão de um repositório baseado em Git

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Consulte as seções apresentadas a seguir para obter informações sobre como excluir um repositório baseado em Git de um Caderno do EMR no console antigo ou de um Workspace do EMR Studio no novo console.

New console

Como os Cadernos do EMR são Workspaces do EMR Studio no novo console, você pode consultar Vinculação de repositórios baseados em Git a um Workspace do EMR Studio para obter mais informações sobre como trabalhar com repositórios do Git em seu Workspace. Mas, neste momento, não é possível excluir repositórios do Git do Workspaces.

Old console

Para atualizar um repositório baseado em Git no console antigo

1. Na página Git repositories (Repositórios do Git), escolha o repositório que deseja atualizar.
2. Na página do repositório, selecione Edit repository (Editar repositório).
3. Atualize as Git credentials (Credenciais do Git) na página do repositório.

Excluir um repositório do Git no console antigo

1. Na página Git repositories (Repositórios do Git), escolha o repositório que deseja excluir.
2. Na página do repositório, escolha todos os blocos de anotações que estão vinculados ao repositório no momento. Selecione Unlink notebook (Desvincular bloco de anotações).
3. Na página do repositório, selecione Delete (Excluir).

> **Note**
>
> Para excluir o repositório local do Git do Amazon EMR, primeiro você deve desvincular todos os cadernos desse repositório. Para ter mais informações, consulte Vinculação ou desvinculação de um repositório baseado em Git. Excluir um repositório do Git não excluirá nenhum segredo criado para o repositório. Você pode excluir o segredo no AWS Secrets Manager.

Vinculação ou desvinculação de um repositório baseado em Git

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Use as etapas apresentadas a seguir para vincular ou desvincular um repositório baseado em Git a um Caderno do EMR no console antigo ou a um Workspace do EMR Studio no novo console.

New console

Como os Cadernos do EMR são Workspaces do EMR Studio no novo console, você pode consultar Vinculação de repositórios baseados em Git a um Workspace do EMR Studio para obter mais informações sobre como trabalhar com repositórios do Git em seu Workspace. Mas, neste momento, não é possível excluir repositórios do Git do Workspaces.

Old console

Como vincular um repositório baseado em Git a um bloco de anotações do EMR

O repositório poderá ser vinculado a um bloco de anotações assim que o bloco de anotações estiver Ready (Pronto).

1. Na lista Notebooks (Blocos de anotações), escolha o bloco de anotações que deseja atualizar.

2. Na seção Git repositories (Repositórios do Git), na página Notebook (Bloco de anotações), selecione Link new repository (Vincular novo repositório).
3. Na lista de repositórios da janela Link Git repository to notebook (Vincular repositório do Git ao bloco de anotações), selecione um ou mais repositórios que você deseja vincular ao bloco de anotações e selecione Link repository (Vincular repositório).

Ou

1. Na página Git repositories (Repositórios do Git), escolha o repositório que você deseja vincular ao bloco de anotações.
2. Na lista de EMR notebooks (Blocos de anotações do EMR), selecione Link new notebook (Vincular novo bloco de anotações) para vincular este repositório a um bloco de anotações existente.

Como desvincular um repositório do Git de um bloco de anotações do EMR

1. Na lista Notebooks (Blocos de anotações), escolha o bloco de anotações que deseja atualizar.
2. Na lista de Git repositories (Repositórios do Git), escolha o repositório que deseja desvincular do bloco de anotações e selecione Unlink repository (Desvincular repositório).

Ou

1. Na página Git repositories (Repositórios do Git), escolha o repositório no qual deseja fazer atualizações.
2. Na lista de EMR notebooks (Blocos de anotações do EMR), escolha o bloco de anotações que deseja desvincular do repositório e selecione Unlink notebook (Desvincular bloco de anotações).

> **Note**
>
> Vincular um repositório do Git a um bloco de anotações clona o repositório remoto no bloco de anotações Jupyter local. Desvincular o repositório do Git de um caderno somente desconecta o caderno do repositório remoto, mas não exclui o repositório local do Git.

Noções básicas sobre o status do repositório

Um repositório Git pode ter qualquer um dos status a seguir na lista de repositórios. Para obter mais informações sobre como vincular o EMR Notebooks a repositórios do Git, consulte Vinculação ou desvinculação de um repositório baseado em Git.

Status	Significado
Linking (Vinculando)	O repositório do Git está sendo vinculado ao bloco de anotações. Enquanto o repositório estiver Linking (Vinculando), não será possível interromper o bloco de anotações.
Linked (Vinculado)	O repositório do Git está vinculado ao bloco de anotações. Enquanto o repositório tiver um status Linked (Vinculado) ele estará conectado ao repositório remoto.
Link Failed (Falha ao vincular)	Ocorreu uma falha ao vincular o repositório do Git ao bloco de anotações. Você pode tentar vinculá-los novamente.
Unlinking (Desvinculando)	O repositório do Git está sendo desvinculado do bloco de anotações. Enquanto o repositór io estiver Unlinking (Desvinculando), não será possível interromper o bloco de anotações . Desvincular um repositório do Git de um bloco de anotações apenas o desconecta do repositório remoto; isso não exclui nenhum código do bloco de anotações.
Unlink Failed (Falha ao desvincular)	Ocorreu uma falha ao desvincular o repositório do Git do bloco de anotações. Você pode tentar desvinculá-los novamente.

Criação de um novo Caderno com um repositório do Git associado

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Criar um caderno e associá-lo a repositórios do Git no console antigo do Amazon EMR

1. Siga as instruções em Criação de um bloco de anotações.
2. Em Security group (Grupo de segurança), selecione Use your own security group (Usar seu próprio grupo de segurança).

   > **Note**
   >
   > Os grupos de segurança do bloco de anotações devem incluir uma regra de saída que permita ao bloco de anotações rotear tráfego para a Internet por meio do cluster. Recomendamos que você crie seus próprios grupos de segurança. Para obter mais informações, consulte Specifying EC2 security groups for EMR Notebooks.

3. Em Git repositories (Repositórios do Git), selecione Choose repository (Escolher repositório) para escolher qual repositório associar ao bloco de anotações.

   1. Escolha um repositório armazenado como um recurso na sua conta e selecione Save (Salvar).
   2. Para adicionar um novo repositório como um recurso em sua conta, selecione add a new repository (adicionar um novo repositório). Conclua o fluxo de trabalho Add repository (Adicionar repositório) em uma nova janela.

Uso de repositórios do Git em um Caderno

> **Note**
>
> Os Notebooks EMR estão disponíveis como espaços de trabalho do EMR Studio no console. O botão Criar espaço de trabalho no console permite criar novos cadernos. Para acessar ou criar Workspaces, os usuários dos Cadernos do EMR precisam de permissões adicionais de perfil do IAM. Para obter mais informações, consulte Notebooks do Amazon EMR são espaços de trabalho do Amazon EMR Studio no console e no console do Amazon EMR.

Você pode escolher Abrir no Jupyter JupyterLab ou Abrir no Jupyter ao abrir um caderno.

Se você optar por abrir o bloco de anotações no Jupyter, será exibida uma lista de pastas e arquivos expansíveis dentro do bloco de anotações. É possível executar comandos do Git manualmente, como os apresentados a seguir em uma célula do bloco de anotações.

```
!git pull origin primary
```

Para abrir qualquer um dos outros repositórios, navegue até as outras pastas.

Se você optar por abrir o notebook com uma JupyterLab interface, poderá usar a extensão JupyterLab Git pré-instalada. Para obter informações sobre a extensão, consulte jupyterlab-git.

Planejar e configurar clusters

Esta seção explica as opções de configuração e instruções para o planejamento, a configuração e a execução de clusters usando o Amazon EMR. Antes de executar um cluster, você faz escolhas sobre o seu sistema com base nos dados que está processando e nos seus requisitos de custo, velocidade, capacidade, disponibilidade, segurança e gerenciabilidade. Suas opções incluem:

- Em qual região executar um cluster, onde e como armazenar dados e como gerar a saída dos resultados. Consulte Configurar o armazenamento de dados físico e o local do cluster.
- Se você está executando clusters do Amazon EMR no Outposts ou em zonas locais. Consulte Clusters EMR em AWS Outposts ou Clusters EMR em Locais Zones AWS.
- Se um cluster é transitório ou de longa execução, e quais softwares ele executa. Consulte Configurar um cluster para continuar ou terminar após a execução da etapa e Configuração de software do cluster.
- Se um cluster tem um único nó primário ou três nós primários. Consulte Planejar e configurar nós primários.
- As opções de hardware e rede que otimizam o custo, o desempenho e a disponibilidade do seu aplicativo. Consulte Configurar o hardware e as redes do cluster.
- Como configurar clusters, para que você possa gerenciá-los com mais facilidade e monitorar as atividades, o desempenho e a integridade. Consulte Configurar registro em log e depuração do cluster e Clusters de etiqueta.
- Como autenticar e autorizar o acesso aos recursos do cluster e como criptografar os dados. Consulte Segurança no Amazon EMR.
- Como integrar-se com outros softwares e serviços. Consulte Integração de drivers e aplicações de terceiros.

Iniciar um cluster rapidamente

Para iniciar rapidamente um cluster com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr/clusters.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.

3. Na página Criar cluster, insira ou selecione os valores dos campos fornecidos. O painel de resumo persistente exibe uma visualização em tempo real das opções de cluster que estão selecionadas atualmente. Selecione um título no painel de resumo para navegar até a seção correspondente e fazer os ajustes. O nome do cluster não pode conter os caracteres <, >, $, | ou `(crase). Você deve concluir todas as configurações necessárias antes de escolher Criar cluster.
4. Escolha Criar cluster para aceitar a configuração, conforme o exemplo.
5. A página de detalhes do cluster é exibida. Veja o Status do cluster próximo ao nome do cluster. O status deve ser alterado de Starting para Running para Waiting durante o processo de criação do cluster. Talvez você precise escolher o ícone de atualização na parte superior à direita ou atualizar o navegador para receber atualizações.

   Quando o status é alterado para Waiting, o cluster está ativo, em execução e pronto para aceitar etapas e conexões SSH.

Configurar o armazenamento de dados físico e o local do cluster

Esta seção descreve como configurar a região de um cluster, os diferentes sistemas de arquivos disponíveis quando você usa o Amazon EMR e como usá-los. Ela também discute como preparar ou carregar dados no Amazon EMR, se necessário, além de como preparar um local de saída para arquivos de log e outros arquivos de dados de saída que você configure.

Tópicos

Escolha uma AWS região

A Amazon Web Services é executada em servidores distribuídos em datacenters ao redor do mundo. Os datacenters são organizados por região geográfica. Ao executar um cluster do Amazon EMR, você deve especificar uma região. É possível escolher a região para reduzir a latência, minimizar custos ou atender a exigências regulamentares. Para obter uma lista de todas as regiões e endpoints compatíveis com o Amazon EMR, consulte Regions and endpoints no Referência geral da Amazon Web Services.

Para obter a melhor performance, você deve executar o cluster na mesma região que os seus dados. Por exemplo, se o bucket do Amazon S3 que armazena seus dados de entrada estiver na região Oeste dos EUA (Oregon), você deverá executar seu cluster na região Oeste dos EUA (Oregon) para evitar taxas de transferência de dados entre regiões. Se você usar um bucket do Amazon S3 para receber a saída do cluster, também deverá criá-lo na região Oeste dos EUA (Oregon).

Se você pretende associar um par de chaves do Amazon EC2 com o cluster (necessário para usar o SSH para logon no nó principal), esse par deverá ser criado na mesma região do cluster. Da mesma forma, os grupos de segurança que o Amazon EMR cria para gerenciar o cluster são criados na mesma região do cluster.

Se você se inscreveu Conta da AWS em ou depois de 17 de maio de 2017, a região padrão ao acessar um recurso do AWS Management Console é Leste dos EUA (Ohio) (us-east-2); para contas mais antigas, a região padrão é Oeste dos EUA (Oregon) (us-west-2) ou Leste dos EUA (Norte da Virgínia) (us-east-1). Para mais informações, consulte Regiões e endpoints da .

Alguns AWS recursos estão disponíveis somente em regiões limitadas. Por exemplo, instâncias de computação em cluster estão disponíveis apenas na região Leste dos EUA (Norte da Virgínia), e a região Ásia-Pacífico (Sydney) apenas oferece suporte ao Hadoop 1.0.3 e versões posteriores. Ao escolher uma região, verifique se ela oferece suporte aos atributos que você deseja usar.

Para obter o melhor desempenho, use a mesma região para todos os AWS recursos que serão usados com o cluster. A tabela a seguir mapeia os nomes de regiões entre serviços. Para conferir a lista de regiões do Amazon EMR, consulte Regiões da AWS and endpoints na Referência geral da Amazon Web Services.

Escolher uma região usando o console

A região padrão é exibida à esquerda das informações da conta na barra de navegação. Para trocar de região no console novo ou no antigo, escolha o menu suspenso Região e selecione uma nova opção.

Especifique uma região com o AWS CLI

Especifique uma região padrão AWS CLI usando o aws configure comando ou a variável de `AWS_DEFAULT_REGION` ambiente. Para obter mais informações, consulte Configurando a AWS região no Guia do AWS Command Line Interface usuário.

Escolher uma região usando um SDK ou a API

Para escolher uma região usando um SDK, configure sua aplicação para usar o endpoint dessa região. Se estiver criando uma aplicação cliente usando um AWS SDK, você poderá alterar o endpoint do cliente chamando `setEndpoint`, como mostra o exemplo a seguir:

```
client.setEndpoint("elasticmapreduce.us-west-2.amazonaws.com");
```

Depois que a aplicação tiver especificado uma região definindo o endpoint, você poderá definir a zona de disponibilidade para instâncias do EC2 do cluster. As zonas de disponibilidade são as localizações geográficas distintas que são criadas para serem isoladas das falhas em outras zonas da disponibilidade e fornecem rede de conectividade acessível e de baixa latência a outras zonas de disponibilidade da mesma região. Uma região contém uma ou mais zonas de disponibilidade. Para otimizar o desempenho e reduzir a latência, todos os recursos devem estar localizados na mesma zona de disponibilidade do cluster que os utiliza.

Trabalhar com armazenamento e sistemas de arquivos

O Amazon EMR e o Hadoop oferecem a você uma variedade de sistemas de arquivos para processar as etapas do cluster. Você especifica o sistema de arquivos a ser usado pelo prefixo do URI que acessa os dados. Por exemplo, `s3://DOC-EXAMPLE-BUCKET1/path` referencia um bucket do Amazon S3 usando o sistema EMRFS. A tabela a seguir lista os sistemas de arquivos disponíveis e inclui as recomendações sobre quando é melhor usar cada um deles.

O Amazon EMR e o Hadoop normalmente usam dois ou mais dos seguintes sistemas de arquivos no processamento de clusters. O HDFS e o EMRFS são os dois principais sistemas de arquivos usados com o Amazon EMR.

> **Important**
>
> A partir da versão 5.22.0 do Amazon EMR, o Amazon EMR AWS usa o Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. As versões anteriores do Amazon EMR usam o AWS Signature versão 2 em alguns casos, a menos que as notas de lançamento indiquem que o Signature versão 4 é usado exclusivamente. Para obter mais informações, consulte Autenticação de solicitações (AWS assinatura versão 4) e Solicitações de autenticação (AWS assinatura versão 2) no Guia do desenvolvedor do Amazon Simple Storage Service.

Sistema de arquivos	Prefixo	Descrição
HDFS	`hdfs://` (ou sem prefixo)	O HDFS é o sistema de arquivos distribuído, escalável e portátil do Hadoop. Uma vantagem do HDFS é o reconhecimento de dados entre os nós de clusters do Hadoop que gerenciam os clusters e os nós de cluster do Hadoop que gerenciam as etapas individuais. Para obter mais informações, consulte a documentação do Hadoop. O HDFS é usado pelos nós principais e core. Uma de suas vantagens é a velocidade; uma desvantagem é ser um armazenamento temporário que é reivindicado quando o cluster é encerrado. É melhor usado para armazenamento em cache dos resultados intermedi ários produzidos pelas etapas de um fluxo de trabalho.
EMRFS	`s3://`	O EMRFS é uma implementação do sistema de arquivos do Hadoop usado para a leitura e gravação de arquivos comuns do Amazon EMR diretamente no Amazon S3. O EMRFS oferece a conveniência de armazenar dados persistentes no Amazon S3 para uso com o Hadoop, além de fornecer recursos como criptografia, consistência e consistência de listas no servidor do Amazon S3. read-after-write **Note** Antes, o Amazon EMR usava os sistemas de arquivos `s3n` e `s3a`. Embora ambos ainda funcionem, recomendamos que você use o esquema do URI `s3` para obter melhor performance, segurança e confiabilidade.
sistema de arquivos local		O sistema de arquivos local é a um disco conectado localmente. Quando um cluster do Hadoop é criado, cada nó é criado a partir de uma instância do

Sistema de arquivos	Prefixo	Descrição
		EC2 que é acompanhada de um bloco pré-confi gurado de armazenamento de disco denominado armazenamento da instância. Os dados nos volumes de armazenamento da instância persistem apenas durante a vida da sua instância do EC2. Os volumes de armazenamento da instância são ideais para armazenar dados temporários que mudam continuam ente, tais como buffers, caches, dados temporários e outros conteúdos temporários. Para obter mais informações, consulte Armazenamento de instância do Amazon EC2. O HDFS usa o sistema de arquivos local, mas o Python também é executado com base no sistema de arquivos local. Você pode optar por armazenar outros arquivos de aplicações em volumes de armazenam ento de instância.
(Herdado) Sistema de arquivos de bloco do Amazon S3	`s3bfs://`	O sistema de arquivos de bloco do Amazon S3 é um sistema de armazenamento de arquivos herdado. Não recomendamos em hipótese alguma o uso de deste sistema. **Important** Não recomendamos que você use esse sistema de arquivos pois ele pode acionar um comportamento de disputa que pode causar uma falha no cluster. No entanto, ele pode ser necessário para aplicativos herdados.

Acessar sistemas de arquivo

Você especifica o sistema de arquivos a ser usado com o prefixo do Uniform Resource Identifier (URI) que acessa os dados. Os procedimentos a seguir ilustram como fazer referência a vários tipos diferentes de sistemas de arquivos.

Para acessar um HDFS local

- Especifique o prefixo `hdfs:///` no URI. O Amazon EMR substitui os caminhos que não especificam um prefixo no URI pelo HDFS local. Por exemplo, ambos os URIs a seguir seriam substituídos pelo mesmo local no HDFS.

```
hdfs:///path-to-data

/path-to-data
```

Para acessar um HDFS remoto

- Inclua o endereço IP do nó principal no URI, como mostrado nos exemplos a seguir.

```
hdfs://master-ip-address/path-to-data

master-ip-address/path-to-data
```

Acessar o Amazon S3

- Use o prefixo `s3://`.

```
s3://bucket-name/path-to-file-in-bucket
```

Acessar o sistema de arquivos de bloco do Amazon S3

- Use apenas para aplicações herdadas que exigem o sistema de arquivos de bloco do Amazon S3. Para acessar ou armazenar dados com este sistema de arquivos, use o prefixo `s3bfs://` no URI.

  O sistema de arquivos de bloco do Amazon S3 é um sistema de arquivos antigo que foi usado para oferecer suporte a carregamentos maiores do que 5 GB de tamanho para o Amazon S3. Com a funcionalidade de upload em várias partes que o Amazon EMR fornece por meio AWS do Java SDK, você pode fazer upload de arquivos de até 5 TB para o sistema de arquivos nativo do Amazon S3, e o sistema de arquivos em blocos do Amazon S3 está obsoleto.

  > **Warning**
  >
  > Como esse sistema de arquivos herdado pode criar um comportamento de disputa que pode corromper o sistema de arquivos, você deve evitar esse formato e usar o EMRFS em seu lugar.

```
s3bfs://bucket-name/path-to-file-in-bucket
```

Preparar dados de entrada

A maioria dos clusters carrega dados de entrada e depois processa esses dados. Para carregar dados, eles precisam estar em um local que o cluster possa acessar e ter um formato que o cluster possa processar. O cenário mais comum é carregar dados de entrada no Amazon S3. O Amazon EMR fornece ferramentas para o seu cluster importar ou ler dados do Amazon S3.

O formato de entrada padrão no Hadoop é um arquivo de texto, embora você possa personalizar o Hadoop e usar ferramentas para importar dados armazenados em outros formatos.

Tópicos

- Tipos de entrada que o Amazon EMR pode aceitar
- Como inserir dados no Amazon EMR

Tipos de entrada que o Amazon EMR pode aceitar

O formato de entrada padrão para um cluster é um arquivo de texto, com cada linha separada por um caractere de nova linha (\n), que é o formato de entrada mais usado.

Se os seus dados de entrada estiverem em um formato diferente do formato de arquivo de texto padrão, você poderá usar a interface do Hadoop `InputFormat` para especificar outros tipos de entradas. Você pode até mesmo criar uma subclasse da classe `FileInputFormat` para tratar os tipos de dados personalizados. Para obter mais informações, consulte http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapred/ InputFormat .html.

Se você estiver usando o Hive, poderá usar um serializador/desserializador (SerDe) para ler dados de um determinado formato no HDFS. Para obter mais informações, consulte https://cwiki.apache.org/confluence/display/Hive/ SerDe.

Como inserir dados no Amazon EMR

O Amazon EMR fornece várias maneiras de colocar dados em um cluster. A mais comum é carregar os dados para o Amazon S3 e usar os recursos internos do Amazon EMR para carregar os dados no cluster. Você também pode usar o recurso DistributedCache do Hadoop para transferir arquivos de um sistema de arquivos distribuído para o sistema de arquivos local. A implementação do Hive fornecida pelo Amazon EMR (Hive versão 0.7.1.1 e posteriores) inclui a funcionalidade que você pode usar para importar e exportar dados entre o DynamoDB e um cluster do Amazon EMR. Se tiver grandes quantidades de dados on-premises para processar, talvez considere o serviço AWS Direct Connect útil.

Tópicos

- Carregar dados no Amazon S3
- Carregar dados usando o AWS DataSync
- Importar arquivos com o cache distribuído
- Como processar arquivos compactados
- Importar dados do DynamoDB para o Hive
- Conectar-se aos dados usando o AWS Direct Connect
- Carregar grandes quantidades de dados usando o AWS Snowball

Carregar dados no Amazon S3

Para obter instruções sobre como carregar objetos no Amazon S3, consulte Add an object to your bucket no Guia do usuário do Amazon Simple Storage Service. Para obter mais informações sobre como usar o Amazon S3 com o Hadoop, consulte http://wiki.apache.org/hadoop/AmazonS3.

Tópicos

- Criar e configurar um bucket do Amazon S3
- Configurar o carregamento multiparte para o Amazon S3
- Práticas recomendadas
- Upload de dados no Amazon S3 Express One Zone

Criar e configurar um bucket do Amazon S3

O Amazon EMR usa o AWS SDK for Java com o Amazon S3 para armazenar dados de entrada, arquivos de log e dados de saída. O Amazon S3 se refere a esses locais de armazenamento como bucket. Os buckets têm algumas restrições e limitações para estar em conformidade com os requisitos do Amazon S3 e do DNS. Para obter mais informações, consulte Restrições e limitações de bucket no Manual do usuário do Amazon Simple Storage Service.

Esta seção mostra como usar o Amazon S3 AWS Management Console para criar e depois definir permissões para um bucket do Amazon S3. Você também pode criar e definir permissões para um bucket do Amazon S3 usando a API do Amazon S3 ou a AWS CLI. Você também pode usar curl junto com uma modificação para transmitir os parâmetros de autenticação apropriados para o Amazon S3.

Consulte os recursos a seguir:

- Para criar um bucket usando o console, consulte Criação de um bucket, no Guia do usuário do Amazon S3.
- Para criar e trabalhar com buckets usando o AWS CLI, consulte Como usar comandos de alto nível do S3 com o AWS Command Line Interface no Guia do usuário do Amazon S3.
- Para criar um bucket usando um SDK, veja exemplos de criação de um bucket no Guia do usuário do Amazon Simple Storage Service.
- Para trabalhar com buckets usando Curl, consulte Amazon S3 authentication tool for curl.
- Para obter mais informações sobre buckets específicos para regiões, consulte Acesso a um bucket no Guia do usuário do Amazon Simple Storage Service.

- Para trabalhar com buckets usando Pontos de Acesso Amazon S3, consulte Usar um alias em estilo de bucket para seu ponto de acesso de bucket do S3 no Guia do usuário do Amazon S3. Você facilmente pode usar os Pontos de Acesso Amazon S3 com o alias do Ponto de Acesso Amazon S3 em vez do nome do bucket do Amazon S3. Use o Ponto de Acesso Amazon S3 para aplicações novas e já existentes, inclusive Spark, Hive, Presto e outros.

> **Note**
>
> Se você ativar o registro em log para um bucket, ele só permitirá logs de acesso do bucket, e não logs de cluster do Amazon EMR.

Durante a criação do bucket ou depois, você pode definir as permissões apropriadas para acessar o bucket, dependendo de seu aplicativo. Normalmente, você atribui acesso de leitura e gravação para si mesmo (o proprietário) e atribui acesso de leitura para os usuários autenticados.

Os buckets do Amazon S3 obrigatórios devem existir para que você possa criar um cluster. Você deve carregar todos os scripts necessários ou dados referenciados no cluster no Amazon S3. A tabela a seguir descreve dados de exemplo, scripts e locais de arquivo de log.

Configurar o carregamento multiparte para o Amazon S3

O Amazon EMR oferece suporte ao upload de várias partes do Amazon S3 por meio do SDK AWS for Java. O multipart upload permite que você faça upload de um único objeto como um conjunto de partes. O upload dessas partes de objetos pode ser feito de maneira independente e em qualquer ordem. Se a transmissão de alguma parte falhar, você poderá retransmitir essa parte sem afetar outras partes. Depois que todas as partes do objeto forem carregadas, o Amazon S3 montará as partes e criará o objeto.

Para obter mais informações, consulte Visão geral do carregamento fracionado no Manual do usuário do Amazon Simple Storage Service.

Além disso, o Amazon EMR oferece propriedades que permitem controlar mais precisamente a limpeza de partes de carregamentos multiparte com falha.

A tabela a seguir descreve as propriedades de configuração do Amazon EMR para o carregamento multiparte. Você pode configurar esses valores usando a classificação de configuração `core-site`. Para obter mais informações, consulte Configure applications no Guia de lançamento do Amazon EMR.

Nome do parâmetro de configuração	Valor padrão	Descrição
`fs.s3n.multipart.uploads.enabled`	`true`	Um tipo booleano que indica se os carregamentos multiparte devem ou não ser habilitados. Quando a visualização consistente do EMRFS está habilitada, os carregamentos multiparte são habilitados por padrão, e a definição desse valor como `false` é ignorada.
`fs.s3n.multipart.uploads.split.size`	`134217728`	Especifica o tamanho máximo de uma parte, em bytes, antes que o EMRFS inicie o upload de uma nova parte quando os multipart uploads estão habilitados. O valor mínimo é `5242880` (5 MB). Se um valor menor for especificado, `5242880` será usado. O máximo é `5368709120` (5 GB). Se um valor maior for especificado, `5368709120` será usado. Se a criptografia do lado do cliente do EMRFS estiver desabilitada e o confirmador otimizado para Amazon S3 também estiver desabilitado, esse valor também controlará o tamanho máximo que um arquivo de dados pode crescer até que o EMRFS use carregamentos multiparte em vez de uma solicitação `PutObject` para carregar o arquivo. Para obter mais informações, consulte
`fs.s3n.ssl.enabled`	`true`	Um tipo booleano que indica se o http ou o https deve ser usado.
`fs.s3.buckets.create.enabled`	`false`	Um tipo booleano que indica se um bucket deve ser criado caso ele não exista.

Nome do parâmetro de configuração	Valor padrão	Descrição
		Configurar como `false` gera uma exceção em operações `CreateBucket` .
`fs.s3.multipart.clean.enabled`	`false`	Um tipo booliano que indica se deseja habilitar a limpeza periódica em segundo plano de carregamentos multiparte incompletos.
`fs.s3.multipart.clean.age.threshold`	`604800`	Um tipo longo que especifica a idade mínima de um multipart upload, em segundos, antes de ser considerado para limpeza. O padrão é uma semana.
`fs.s3.multipart.clean.jitter.max`	`10000`	Um tipo inteiro que especifica o valor máximo de atraso de oscilação aleatória em segundos adicionado ao atraso fixo de 15 minutos antes de programar a próxima execução de limpeza.

Desabilitar carregamentos multiparte

Console

Para desativar os carregamentos de várias partes com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Configurações do software, insira a seguinte configuração: `classification=core-site,properties=[fs.s3n.multipart.uploads.enabled=false]`.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

CLI

Para desativar o upload de várias partes usando o AWS CLI

Esse procedimento explica como desabilitar o multipart upload usando a AWS CLI. Para desabilitar o multipart upload, digite o comando `create-cluster` com o parâmetro `--bootstrap-actions`.

1. Crie um arquivo, `myConfig.json`, com o seguinte conteúdo e salve-o no mesmo diretório onde você executa o comando:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "fs.s3n.multipart.uploads.enabled": "false"
    }
  }
]
```

2. Digite o seguinte comando, substituindo *myKey* pelo nome do seu par de chaves do EC2.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "Test cluster" \
--release-label emr-7.1.0 --applications Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3 --configurations file://myConfig.json
```

API

Desabilitar carregamento multiparte usando a API

- Para obter informações sobre o uso programático de carregamentos multiparte do Amazon S3, consulte Using the AWS SDK for Java for multipart upload no Guia do usuário do Amazon Simple Storage Service.

  Para obter mais informações sobre o AWS SDK para Java, AWS consulte SDK for Java.

Práticas recomendadas

Veja a seguir as recomendações para o uso de buckets do Amazon S3 com clusters do EMR.

Habilitar o versionamento

O versionamento é uma configuração recomendada para o seu bucket do Amazon S3. Habilitando o versionamento, você garante que, mesmo que os dados sejam excluídos ou substituídos sem querer, eles possam ser recuperados. Para obter mais informações, consulte Usando versionamento no Guia do usuário do Amazon Simple Storage Service.

Limpar carregamentos multiparte com falha

Os componentes de cluster do EMR usam uploads de várias partes por meio do SDK for AWS Java com as APIs do Amazon S3 para gravar arquivos de log e enviar dados para o Amazon S3 por padrão. Para obter informações sobre como alterar as propriedades relacionadas a essa configuração usando o Amazon EMR, consulte Configurar o carregamento multiparte para o Amazon S3. Às vezes, carregar um arquivo grande pode resultar em um carregamento multiparte do Amazon S3 incompleto. Quando não é possível concluir com êxito um multipart upload em andamento, este continua a ocupar seu bucket e resulta em cobranças de armazenamento. Recomendamos as seguintes opções para evitar excesso de armazenamento de arquivos:

- Para buckets usados com o Amazon EMR, use uma regra de configuração de ciclo de vida no Amazon S3 para remover carregamentos multiparte incompletos três dias após a data de início do carregamento. As regras de configuração de ciclo de vida permitem que você controle a classe de armazenamento e o tempo de vida dos objetos. Para obter mais informações, consulte Object lifecycle management e Aborting incomplete multipart uploads using a bucket lifecycle policy.
- Habilite o atributo de limpeza de multiparte do Amazon EMR definindo `fs.s3.multipart.clean.enabled` como `true` e ajustando outros parâmetros de limpeza.

Esse recurso é útil em alto volume, grande escala e com clusters que tenham tempo limitado. Nesse caso, o parâmetro `DaysAfterIntitiation` de uma regra de configuração do ciclo de vida pode ser muito longo, mesmo se definido como o mínimo, causando picos no armazenamento do Amazon S3. A limpeza multiparte do Amazon EMR possibilita um controle mais preciso. Para ter mais informações, consulte Configurar o carregamento multiparte para o Amazon S3.

Gerenciar marcadores de versão

Recomendamos habilitar a regra de configuração do ciclo de vida no Amazon S3 para remover marcadores de exclusão de objetos expirados para buckets com versionamento que você usa no Amazon EMR. Ao excluir um objeto em um bucket com versionamento, um marcador de exclusão é criado. Se todas as versões anteriores do objeto expirarem posteriormente, um marcador de exclusão de objeto expirado será deixado no bucket. Embora você não seja cobrado por marcadores de exclusão, a remoção dos marcadores expirados pode melhorar o desempenho das solicitações LIST. Para obter mais informações, consulte Lifecycle configuration for a bucket with versioning no Guia do usuário do Amazon Simple Storage Service.

Práticas recomendadas de desempenho

Dependendo das suas cargas de trabalho, tipos específicos de uso de clusters do EMR e de aplicativos nesses clusters podem resultar em um alto número de solicitações em um bucket. Para obter mais informações, consulte Request rate and performance considerations no Guia do usuário do Amazon Simple Storage Service.

Upload de dados no Amazon S3 Express One Zone

Visão geral

Com o Amazon EMR 6.15.0 e versões superiores, você pode usar o Amazon EMR com o Apache Spark e a classe de armazenamento Amazon S3 Express One Zone para melhorar a performance nos trabalhos do Spark. O S3 Express One Zone é uma classe de armazenamento do S3 para aplicações que acessam dados frequentemente com centenas de milhares de solicitações por segundo. Na hora da execução, o S3 Express One Zone oferece o armazenamento de objetos na nuvem com a menor latência e a maior performance do Amazon S3.

Pré-requisitos

- Permissões do S3 Express One Zone: quando o S3 Express One Zone inicialmente executa uma ação como `GET`, `LIST` ou `PUT` em um objeto do S3, a classe de armazenamento chama `CreateSession` em seu nome. Sua política do IAM deve conceder a permissão

`s3express:CreateSession` para que o conector S3A possa invocar a API `CreateSession`. Para obter um exemplo de política com essa permissão, consulte Conceitos básicos da classe Amazon S3 Express One Zone.

- Conector S3A: para configurar o cluster do Spark para acessar dados de um bucket do Amazon S3 que usa a classe de armazenamento S3 Express One Zone, você deve usar o conector S3A do Apache Hadoop. Para usar o conector, certifique-se de que todos os URIs do S3 usem o esquema do `s3a`. Caso contrário, você pode alterar a implementação do sistema de arquivos usado para os esquemas do `s3` e do `s3n`.

Para alterar o esquema do `s3`, especifique as seguintes configurações de cluster:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "fs.s3.impl": "org.apache.hadoop.fs.s3a.S3AFileSystem",
      "fs.AbstractFileSystem.s3.impl": "org.apache.hadoop.fs.s3a.S3A"
    }
  }
]
```

Para alterar o esquema do `s3n`, especifique as seguintes configurações de cluster:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "fs.s3n.impl": "org.apache.hadoop.fs.s3a.S3AFileSystem",
      "fs.AbstractFileSystem.s3n.impl": "org.apache.hadoop.fs.s3a.S3A"
    }
  }
]
```

Conceitos básicos da classe Amazon S3 Express One Zone

Tópicos

- Criação de uma política de permissões
- Criação e configuração de um cluster
- Visão geral das configurações

Criação de uma política de permissões

Antes de criar um cluster que use o Amazon S3 Express One Zone, você deve criar uma política do IAM para anexar ao perfil de instância do Amazon EC2 para o cluster. A política deve ter permissões para acessar a classe de armazenamento S3 Express One Zone. O exemplo de política a seguir mostra como conceder a permissão necessária. Após criar a política, anexe-a à função do perfil de instância usada para criar seu cluster do EMR, conforme descrito na seção Criação e configuração de um cluster.

```
{
  "Version":"2012-10-17",
  "Statement": [
        {
            "Effect": "Allow",
            "Resource": "arn:aws:s3express:region-code:account-id:bucket/DOC-EXAMPLE-
BUCKET",
            "Action": [
                "s3express:CreateSession"
            ]
        }
    ]

 }
```

Criação e configuração de um cluster

Em seguida, crie um cluster que execute o Spark com o S3 Express One Zone. As seguinte etapas descrevem uma visão geral de alto nível para criar um cluster no AWS Management Console:

1. Navegue até o console do Amazon EMR e selecione Clusters na barra lateral. Depois, selecione Criar cluster.
2. Selecione a versão `emr-6.15.0` ou superiores do Amazon EMR.
3. Selecione o pacote de aplicações interativas do Spark e quaisquer outras aplicações que você queira incluir no cluster. É necessário incluir pelo menos o Spark e o Hadoop no cluster.
4. Para habilitar o Amazon S3 Express One Zone, insira uma configuração semelhante ao exemplo a seguir na seção Configurações de software. As configurações e os valores recomendados estão descritos na seção Visão geral das configurações após esse procedimento.

   ```
   [
     {
   ```

```
    "Classification": "core-site",
    "Properties": {
      "fs.s3a.aws.credentials.provider":
 "software.amazon.awssdk.auth.credentials.InstanceProfileCredentialsProvider",
      "fs.s3a.change.detection.mode": "none",
      "fs.s3a.endpoint.region": "aa-example-1",
      "fs.s3a.select.enabled": "false"
    }
  },
  {
    "Classification": "spark-defaults",
    "Properties": {
      "spark.sql.sources.fastS3PartitionDiscovery.enabled": "false"
    }
  }
]
```

5. Na seção Perfil de instância do EC2 para o Amazon EMR, escolha usar um perfil existente e use um perfil com a política anexada que você criou na seção Criação de uma política de permissões acima.
6. Defina o restante das configurações do cluster conforme apropriado para a sua aplicação e selecione Criar cluster.

Visão geral das configurações

As tabelas a seguir descrevem as configurações e os valores sugeridos que você deve especificar ao configurar um cluster que usa o S3 Express One Zone com o Amazon EMR, conforme descrito na seção Criação e configuração de um cluster.

Configurações do S3A

Parâmetro	Valor padrão	Valor sugerido	Explicação
`fs.s3a.aws.credentials.provider`	Se não for especificado, usa `AWSCredentialProviderList` na seguinte ordem: `Temporary`	`software.amazon.awssdk.auth.credentials.InstanceProfil`	A função do perfil de instância do Amazon EMR deve ter a política que permita ao sistema de arquivos do S3A

Parâmetro	Valor padrão	Valor sugerido	Explicação
	`AWSCredentialsProvider`, `SimpleAWSCredentialsProvider`, `EnvironmentVariableCredentialsProvider`, `IAMInstanceCredentialsProvider`.	`eCredentialsProvider`	chamar `s3express:CreateSession`. Outros provedores de credenciais também funcionam se tiverem as permissões do S3 Express One Zone.
`fs.s3a.endpoint.region`	nulo	O Região da AWS local em que você criou o bucket.	A lógica de resolução da região não funciona com a classe de armazenamento S3 Express One Zone.
`fs.s3a.select.enabled`	`true`	`false`	O valor `select` do Amazon S3 não é compatível com a classe de armazenamento S3 Express One Zone.

Parâmetro	Valor padrão	Valor sugerido	Explicação
`fs.s3a.change.detection.mode`	`server`	nenhuma	A detecção de alterações pelo S3A funciona verificando `etags` baseadas em MD5. A classe de armazenamento S3 Express One Zone não é compatível com `checksums` de MD5.

Configurações do Spark

Parâmetro	Valor padrão	Valor sugerido	Explicação
`spark.sql.sources.fastS3PartitionDiscovery.enabled`	`true`	`false`	A otimização interna usa um parâmetro de API do S3 que não é compatível com a classe de armazenamento S3 Express One Zone.

Considerações

Considere os seguintes pontos ao integrar o Apache Spark no Amazon EMR à classe de armazenamento S3 Express One Zone:

- O Amazon S3 Express One Zone é compatível com as versões 6.15.0 e superiores do Amazon EMR.
- O conector S3A é necessário para usar o S3 Express One Zone com o Amazon EMR. Somente o S3A tem os recursos e as classes de armazenamento necessários para interagir com o S3

Express One Zone. Para ver as etapas de configuração do conector, consulte the section called "Pré-requisitos".

- A classe de armazenamento Amazon S3 Express One Zone só é compatível com o Spark em um cluster do Amazon EMR executado no Amazon EC2.
- A classe de armazenamento Amazon S3 Express One Zone só oferece suporte à criptografia SSE-S3. Para obter mais informações, consulte Criptografia do lado do servidor com chaves gerenciadas pelo Amazon S3 (SSE-S3).
- A classe de armazenamento Amazon S3 Express One Zone não oferece suporte a gravações com o `FileOutputCommitter` do S3A. As gravações com o `FileOutputCommitter` do S3A em buckets do S3 Express One Zone resultam em um erro: InvalidStorageClass: The storage class you specified is not valid.
- A classe de armazenamento Amazon S3 Express One Zone não é compatível com o Amazon EMR Sem Servidor ou com o Amazon EMR no EKS.

Carregar dados usando o AWS DataSync

AWS DataSync é um serviço de transferência de dados on-line que simplifica, automatiza e acelera o processo de movimentação de dados entre seus serviços de armazenamento e armazenamento locais ou entre serviços AWS de armazenamento. AWS DataSync oferece suporte a uma variedade de sistemas de armazenamento local, como Hadoop Distributed File System (HDFS), servidores de arquivos NAS e armazenamento autogerenciado de objetos.

A maneira mais comum de colocar dados em um cluster é carregar os dados no Amazon S3 e usar os atributos integrados do Amazon EMR para carregar os dados no cluster.

DataSync pode ajudá-lo a realizar as seguintes tarefas:

- Replicar o HDFS no cluster do Hadoop para o Amazon S3 para continuidade dos negócios
- Copiar o HDFS no Amazon S3 para preencher data lakes
- Transferir dados entre o HDFS do cluster do Hadoop e o Amazon S3 para análise e processamento

Para fazer upload de dados para seu bucket do S3, primeiro você implanta um ou mais DataSync agentes na mesma rede do seu armazenamento local. O agente é uma máquina virtual (VM) usada para ler ou gravar dados em um local autogerenciado. Em seguida, você ativa seus agentes no bucket do S3 Conta da AWS e Região da AWS onde ele está localizado.

Depois que o agente é ativado, crie um local de origem para o armazenamento on-premises, um local de destino para o bucket do S3 e uma tarefa. Uma tarefa é um conjunto de dois locais (origem e destino) e um conjunto de opções padrão que você usa para controlar o comportamento da tarefa.

Finalmente, você executa sua DataSync tarefa de transferir dados da origem para o destino.

Para obter mais informações, consulte Conceitos básicos do AWS DataSync.

Importar arquivos com o cache distribuído

Tópicos

- Tipos de arquivos compatíveis
- Local dos arquivos em cache
- Acessar arquivos em cache de aplicações de transmissão
- Acessar arquivos em cache de aplicações de transmissão

O DistributedCache é um recurso do Hadoop que pode aumentar a eficiência quando uma tarefa map ou reduce precisa acessar dados comuns. Se o cluster depender de aplicações já existentes ou binárias que não estão instaladas quando o cluster é criado, você poderá usar o DistributedCache para importar esses arquivos. Esse recurso permite que um nó de cluster leia os arquivos importados do seu sistema de arquivos local, em vez de recuperar os arquivos de outros nós do cluster.

Para obter mais informações, acesse http://hadoop.apache.org/docs/stable/api/org/apache/hadoop/filecache/ DistributedCache .html.

Você invoca o DistributedCache ao criar o cluster. Os arquivos são armazenados em cache logo antes do início do trabalho do Hadoop e permanecem no cache pela duração do trabalho. Você pode armazenar em cache os arquivos armazenados em qualquer sistema de arquivos compatível com o Hadoop, por exemplo, o HDFS ou o Amazon S3. O tamanho padrão do cache de arquivo é 10 GB. Para alterar o tamanho do cache, reconfigure o parâmetro Hadoop `local.cache.size` usando a ação de bootstrap. Para ter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais.

Tipos de arquivos compatíveis

O DistributedCache permite arquivos únicos e arquivamentos. Arquivos individuais são armazenados em cache como somente leitura. Executáveis e arquivos binários têm permissões de execução definidas.

Os arquivamentos são um ou mais arquivos empacotados por meio de um utilitário, como o `gzip`. O DistributedCache passa os arquivos compactados para cada nó central e descompacta o arquivo como parte do armazenamento em cache. O DistributedCache é compatível com os seguintes formatos de compactação:

- zip
- tgz
- tar.gz
- tar
- jar

Local dos arquivos em cache

O DistributedCache copia arquivos apenas para nós centrais. Se não houver um nó central no cluster, o DistributedCache copiará os arquivos para o nó primário.

O DistributedCache associa os arquivos de cache ao diretório de trabalho atual do mapper e do reducer usando links simbólicos. Um link simbólico é um alias para uma localização de arquivo, e não essa localização propriamente dita. O valor do parâmetro, `yarn.nodemanager.local-dirs` em `yarn-site.xml`, especifica a localização dos arquivos temporários. O Amazon EMR define esse parâmetro como `/mnt/mapred` ou alguma variação com base no tipo de instância e na versão do EMR. Por exemplo, uma configuração pode ter `/mnt/mapred` e `/mnt1/mapred` porque o tipo de instância tem dois volumes temporários. Arquivos de cache estão localizados em um subdiretório da localização de arquivo temporária em `/mnt/mapred/taskTracker/archive`.

Se você armazenar um único arquivo em cache, o DistributedCache colocará o arquivo no diretório `archive`. Se você armazenar um arquivamento em cache, o DistributedCache descompactará o arquivo e criará um subdiretório em `/archive` com o mesmo nome do arquivamento. Os arquivos individuais estão localizados no novo subdiretório.

Você pode usar o DistributedCache somente ao utilizar o Streaming.

Acessar arquivos em cache de aplicações de transmissão

Para acessar os arquivos em cache dos seus aplicativos de mapeador ou redutor, certifique-se de ter adicionado o diretório de trabalho atual (. /) ao caminho do aplicativo e referenciado os arquivos em cache como se estivessem presentes no diretório de trabalho atual.

Acessar arquivos em cache de aplicações de transmissão

Você pode usar o AWS Management Console e o AWS CLI para criar clusters que usam o cache distribuído.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Especificar arquivos de cache distribuído usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Etapas, escolha Adicionar etapa. Isso abrirá a caixa de diálogo Adicionar etapa. No campo Argumentos, inclua os arquivos e arquivamentos para salvar no cache. O tamanho do arquivo (ou o tamanho total dos arquivos em um arquivamento) deve ser menor que o tamanho do cache alocado.

   Para adicionar um arquivo individual ao cache distribuído, especifique `-cacheFile` seguido do nome e do local do arquivo, do sinal de cerquilha (#) e do nome que você deseja dar ao arquivo quando ele for inserido no cache local. O exemplo a seguir demonstra como adicionar um arquivo individual ao cache distribuído.

   ```
   -cacheFile \
   s3://DOC-EXAMPLE-BUCKET/file-name#cache-file-name
   ```

   Para adicionar um arquivo ao cache distribuído, insira `-cacheArchive` seguido da localização dos arquivos no Amazon S3, do sinal de cerquilha (#) e do nome que você deseja dar à coleção de arquivos no cache local. O exemplo a seguir demonstra como adicionar um arquivo de arquivamento ao cache distribuído.

   ```
   -cacheArchive \
   ```

```
s3://DOC-EXAMPLE-BUCKET/archive-name#cache-archive-name
```

Insira os valores correspondentes nos outros campos da caixa de diálogo. As opções diferem dependendo do tipo de etapa. Para adicionar a etapa e sair da caixa de diálogo, escolha Adicionar etapa.

4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Old console

Especificar arquivos de cache distribuído usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Escolha Step execution (Execução de etapa) como o modo de execução.
4. Na seção Steps (Etapas), no campo Add step (Adicionar etapa), escolha Streaming program (Programa de streaming) na lista e clique em Configure and add (Configurar e adicionar).
5. No campo Argumentos, inclua os arquivos e arquivamentos para salvar no cache e clique em Adicionar. O tamanho do arquivo (ou o tamanho total dos arquivos em um arquivamento) deve ser menor que o tamanho do cache alocado.

   Para adicionar um arquivo individual ao cache distribuído, especifique `-cacheFile` seguido do nome e do local do arquivo, do sinal de cerquilha (#) e do nome que você deseja dar ao arquivo quando ele for inserido no cache local. O exemplo a seguir demonstra como adicionar um arquivo individual ao cache distribuído.

   ```
   -cacheFile \
   s3://DOC-EXAMPLE-BUCKET/file_name#cache_file_name
   ```

   Para adicionar um arquivo ao cache distribuído, insira `-cacheArchive` seguido da localização dos arquivos no Amazon S3, do sinal de cerquilha (#) e do nome que você deseja dar à coleção de arquivos no cache local. O exemplo a seguir demonstra como adicionar um arquivo de arquivamento ao cache distribuído.

```
-cacheArchive \
s3://DOC-EXAMPLE-BUCKET/archive_name#cache_archive_name
```

6. Continue com a configuração e a execução do seu cluster. O cluster copia os arquivos para o local do cache antes de processar quaisquer etapas de cluster.

CLI

Para especificar arquivos de cache distribuídos com o AWS CLI

- Para enviar uma etapa de Streaming quando um cluster é criado, digite o comando `create-cluster` com o parâmetro `--steps`. Para especificar arquivos de cache distribuídos usando o AWS CLI, especifique os argumentos apropriados ao enviar uma etapa de streaming.

  Para adicionar um arquivo individual ao cache distribuído, especifique `-cacheFile` seguido do nome e do local do arquivo, do sinal de cerquilha (#) e do nome que você deseja dar ao arquivo quando ele for inserido no cache local.

  Para adicionar um arquivo ao cache distribuído, insira `-cacheArchive` seguido da localização dos arquivos no Amazon S3, do sinal de cerquilha (#) e do nome que você deseja dar à coleção de arquivos no cache local. O exemplo a seguir demonstra como adicionar um arquivo de arquivamento ao cache distribuído.

  Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Example 1

Digite o seguinte comando para executar um cluster e envie um etapa de Streaming que use `-cacheFile` para adicionar um arquivo, `sample_dataset_cached.dat`, ao cache.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications Name=Hive Name=Pig --use-default-roles --ec2-attributes KeyName=myKey
 --instance-type m5.xlarge --instance-count 3 --steps Type=STREAMING,Name="Streaming
 program",ActionOnFailure=CONTINUE,Args=["--files","s3://my_bucket/my_mapper.py
 s3://my_bucket/my_reducer.py","-mapper","my_mapper.py","-reducer","my_reducer.py,"-
```

```
input","s3://my_bucket/my_input","-output","s3://my_bucket/my_output", "-
cacheFile","s3://my_bucket/sample_dataset.dat#sample_dataset_cached.dat"]
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

Se você não tiver criado o perfil de serviço padrão do EMR e o perfil de instância do EC2, digite `aws emr create-default-roles` para criá-los antes de digitar o subcomando `create-cluster`.

Example 2

O comando a seguir mostra a criação de um cluster de streaming e usa `-cacheArchive` para adicionar um arquivamento de arquivos ao cache.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications Name=Hive Name=Pig --use-default-roles --ec2-attributes KeyName=myKey
 --instance-type m5.xlarge --instance-count 3 --steps Type=STREAMING,Name="Streaming
 program",ActionOnFailure=CONTINUE,Args=["--files","s3://my_bucket/my_mapper.py
 s3://my_bucket/my_reducer.py","-mapper","my_mapper.py","-reducer","my_reducer.py","-
input","s3://my_bucket/my_input","-output","s3://my_bucket/my_output", "-
cacheArchive","s3://my_bucket/sample_dataset.tgz#sample_dataset_cached"]
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

Se você não tiver criado o perfil de serviço padrão do EMR e o perfil de instância do EC2, digite `aws emr create-default-roles` para criá-los antes de digitar o subcomando `create-cluster`.

Como processar arquivos compactados

O Hadoop verifica a extensão do arquivo para detectar arquivos compactados. Os tipos de compactação com suporte pelo Hadoop são: gzip, bzip2 e LZO. Você não precisa tomar medidas adicionais para extrair arquivos usando esses tipos de compactação; o Hadoop manipula o processo para você.

Para indexar arquivos LZO, você pode usar a biblioteca hadoop-lzo, que pode ser baixada em https://github.com/kevinweil/hadoop-lzo. Como esta é uma biblioteca de terceiros, o Amazon EMR não oferece suporte de desenvolvedor sobre como utilizar essa ferramenta. Para obter informações de uso, consulte o arquivo leiame da hadoop-lzo.

Importar dados do DynamoDB para o Hive

A implementação do Hive fornecida pelo Amazon EMR inclui a funcionalidade que você pode usar para importar e exportar dados entre o DynamoDB e um cluster do Amazon EMR. Isso é útil quando seus dados de entrada estão armazenados no DynamoDB. Para obter mais informações, consulte Export, import, query, and join tables in DynamoDB using Amazon EMR.

Conectar-se aos dados usando o AWS Direct Connect

AWS Direct Connect é um serviço que você pode usar para estabelecer uma conexão de rede privada dedicada com a Amazon Web Services a partir do seu data center, escritório ou ambiente de colocation. Se você tiver grandes quantidades de dados de entrada, o uso AWS Direct Connect pode reduzir seus custos de rede, aumentar a taxa de transferência da largura de banda e fornecer uma experiência de rede mais consistente do que as conexões baseadas na Internet. Para obter mais informações, consulte o Guia do usuário do AWS Direct Connect.

Carregar grandes quantidades de dados usando o AWS Snowball

AWS Snowball é um serviço que você pode usar para transferir grandes quantidades de dados entre o Amazon Simple Storage Service (Amazon S3) e seu local de armazenamento de dados no local em alta velocidade. faster-than-internet O Snowball oferece suporte a dois tipos de trabalho: trabalhos de importação e trabalhos de exportação. Os trabalhos de importação envolvem a transferência de dados de uma fonte on-premises para um bucket do Amazon S3. Os trabalhos de exportação envolvem a transferência de dados de um bucket do Amazon S3 para uma fonte on-premises. Para ambos os tipos de trabalho, os dispositivos Snowball protegem seus dados, enquanto as transportadoras regionais os transportam entre o Amazon S3 e a local do armazenamento de dados. Os dispositivos Snowball são fisicamente robustos e protegidos pelo (). AWS Key Management Service AWS KMS Para obter mais informações, consulte o Guia do desenvolvedor do AWS Snowball Edge.

Configurar um local de saída

O formato de saída mais comum de um cluster do Amazon EMR é um arquivo de texto, compactado ou não. Normalmente, esse arquivo é gravado em um bucket do Amazon S3. Esse bucket deve ser

criado antes de você iniciar o cluster. Você especifica o bucket do S3 como o local de saída quando inicia o cluster.

Para obter mais informações, consulte os tópicos a seguir.

Tópicos

- Criar e configurar um bucket do Amazon S3
- Que formatos o Amazon EMR pode gerar?
- Como gravar dados em um bucket do Amazon S3 do qual você não é proprietário
- Compactar a saída do cluster

Criar e configurar um bucket do Amazon S3

O Amazon EMR (Amazon EMR) usa o Amazon S3 para armazenar dados de entrada, arquivos de log e dados de saída. O Amazon S3 se refere a esses locais de armazenamento como bucket. Os buckets têm algumas restrições e limitações para estar em conformidade com os requisitos do Amazon S3 e do DNS. Para obter mais informações, acesse Restrições e limitações de bucket no Guia do desenvolvedor do Amazon Simple Storage Service.

Para criar um bucket do Amazon S3, siga as instruções da página Criação de um bucket no Guia do desenvolverdor do Amazon Simple Storage Service.

> **Note**
>
> Se você habilitar o registro em log no assistente Create a Bucket (Criar um bucket), ele só permitirá logs de acesso do bucket, e não logs de cluster.

> **Note**
>
> Para obter mais informações sobre a especificação de buckets específicos da região, consulte Buckets and Regions no Amazon Simple Storage Service Developer Guide e os endpoints regionais disponíveis para os SDKs. AWS

Depois de criar o bucket, você poderá definir as permissões apropriadas. Normalmente, você atribui a si (o proprietário) acesso de leitura e gravação. É altamente recomendável seguir as Práticas recomendadas de segurança para o Amazon S3 ao configurar o bucket.

Os buckets do Amazon S3 obrigatórios devem existir para que você possa criar um cluster. Você deve carregar todos os scripts necessários ou dados referenciados no cluster no Amazon S3. A tabela a seguir descreve dados de exemplo, scripts e locais de arquivo de log.

Informações	Exemplo de local no Amazon S3
script ou programa	`s3://DOC-EXAMPLE-BUCKET1/script/MapperScript.py`
arquivos de log	`s3://DOC-EXAMPLE-BUCKET1/logs`
dados de entrada	`s3://DOC-EXAMPLE-BUCKET1/input`
dados de saída	`s3://DOC-EXAMPLE-BUCKET1/output`

Que formatos o Amazon EMR pode gerar?

O formato de saída padrão para um cluster é texto com pares de chave e valor gravados nas linhas individuais dos arquivos de texto. Este é o formato de saída mais comumente usado.

Se os dados de saída precisam ser gravados em um formato que não seja o de arquivos de texto padrão, você pode usar a interface do Hadoop `OutputFormat` para especificar outros tipos de saída. Você pode até mesmo criar uma subclasse da classe `FileOutputFormat` para tratar os tipos de dados personalizados. Para obter mais informações, consulte http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapred/ OutputFormat .html.

Se você estiver iniciando um cluster do Hive, poderá usar um serializador/desserializador (SerDe) para gerar dados do HDFS em um determinado formato. Para obter mais informações, consulte https://cwiki.apache.org/confluence/display/Hive/ SerDe.

Como gravar dados em um bucket do Amazon S3 do qual você não é proprietário

Ao gravar um arquivo em um bucket do Amazon Simple Storage Service (Amazon S3), por padrão, você é o único usuário capaz de ler esse arquivo. A suposição é a de que você gravará arquivos em seus próprio buckets, e essa configuração padrão protege a privacidade desses arquivos.

No entanto, se você estiver executando um cluster e quiser que a saída seja gravada no bucket Amazon S3 de outro AWS usuário e quiser que esse outro AWS usuário possa ler essa saída, você deve fazer duas coisas:

- Faça com que o outro AWS usuário conceda a você permissões de gravação para o bucket do Amazon S3. O cluster que você executa é executado sob suas AWS credenciais, portanto, qualquer cluster que você iniciar também poderá gravar no bucket desse outro AWS usuário.
- Defina permissões de leitura para o outro AWS usuário nos arquivos que você ou o cluster gravam no bucket do Amazon S3. A maneira mais fácil de definir essas permissões de leitura é usar listas de controle de acesso (ACLs) pré-configuradas, um conjunto predefinido de políticas de acesso definidas pelo Amazon S3.

Para obter informações sobre como o outro AWS usuário pode conceder a você permissões para gravar arquivos no bucket do Amazon S3 do outro usuário, consulte Editando permissões do bucket no Guia do usuário do Amazon Simple Storage Service.

Para o seu cluster usar ACLs pré-configuradas ao gravar arquivos no Amazon S3, defina a opção de configuração de cluster `fs.s3.canned.acl` para a ACL pré-configurada usar. A tabela a seguir lista as ACLs pré-configuradas atualmente definidas.

ACL pré-configurada	Descrição
`AuthenticatedRead`	Especifica que o proprietário recebe `Permission.FullControl` e o favorecido do grupo `GroupGrantee.AuthenticatedUsers` recebe o acesso `Permission.Read`.
`BucketOwnerFullControl`	Especifica que o proprietário do bucket recebe `Permission.FullControl`. O proprietário do bucket não é necessariamente o proprietário do objeto.
`BucketOwnerRead`	Especifica que o proprietário do bucket recebe `Permission.Read`. O proprietário do bucket não é necessariamente o proprietário do objeto.
`LogDeliveryWrite`	Especifica que o proprietário recebe `Permission.FullControl` e o favorecido do grupo `GroupGrantee.LogDelivery` recebe o acesso `Permissio`

ACL pré-configurada	Descrição
	`n.Write` , permitindo que logs de acesso sejam fornecidos.
`Private`	Especifica que o proprietário recebe `Permissio n.FullControl` .
`PublicRead`	Especifica que o proprietário recebe `Permissio n.FullControl` e o favorecido do grupo `GroupGran tee.AllUsers` recebe o acesso `Permissio n.Read` .
`PublicReadWrite`	Especifica que o proprietário recebe `Permissio n.FullControl` e o favorecido do grupo `GroupGran tee.AllUsers` recebe os acessos `Permissio n.Read` e `Permission.Write` .

Há muitas maneiras de definir opções de configuração do cluster, dependendo do tipo de cluster que você está executando. Os procedimentos a seguir mostram como definir a opção para casos comuns.

Para gravar arquivos usando ACLs pré-configuradas no Hive

- No prompt de comando do Hive, defina a opção de configuração `fs.s3.canned.acl` como a ACL pré-configurada desejada na qual você deseja que o cluster defina os arquivos que ele grava no Amazon S3. Para acessar o prompt de comando do Hive, conecte-se ao nó principal usando o SSH e digite Hive no prompt de comando do Hadoop. Para ter mais informações, consulte Conectar-se ao nó primário usando SSH.

  O exemplo a seguir define a configuração de opção `fs.s3.canned.acl` como `BucketOwnerFullControl`, que dá ao proprietário do bucket do Amazon S3 controle total sobre o arquivo. Observe que o comando definido faz distinção entre maiúsculas e minúsculas e não contém aspas ou espaços.

```
hive> set fs.s3.canned.acl=BucketOwnerFullControl;
create table acl (n int) location 's3://acltestbucket/acl/';
insert overwrite table acl select count(*) from acl;
```

As duas últimas linhas do exemplo criam uma tabela que é armazenada no Amazon S3 e gravam dados nessa tabela.

Para gravar arquivos usando ACLs pré-configuradas no Pig

- No prompt de comando do Pig, defina a opção de configuração `fs.s3.canned.acl` como a ACL pré-configurada na qual você deseja que o cluster defina os arquivos que gravará no Amazon S3. Para acessar o prompt de comando do Pig, conecte-se ao nó principal usando o SSH e digite Pig no prompt de comando do Hadoop. Para ter mais informações, consulte Conectar-se ao nó primário usando SSH.

  O exemplo a seguir define a opção de `fs.s3.canned.acl` configuração como BucketOwnerFullControl, o que dá ao proprietário do bucket do Amazon S3 controle total sobre o arquivo. Observe que o comando definido inclui um espaço antes do nome da ACL pré-configurada e não contém aspas.

```
pig> set fs.s3.canned.acl BucketOwnerFullControl;
store some data into 's3://acltestbucket/pig/acl';
```

Para gravar arquivos usando ACLs pré-configuradas em um JAR personalizado

- Defina a opção de configuração `fs.s3.canned.acl` usando o Hadoop com o sinalizador -D. Isso é mostrado no exemplo a seguir.

```
hadoop jar hadoop-examples.jar wordcount
-Dfs.s3.canned.acl=BucketOwnerFullControl s3://mybucket/input s3://mybucket/output
```

Compactar a saída do cluster

Tópicos

- Compactação de dados de saída

- Compactação de dados intermediária
- Usar a biblioteca Snappy com o Amazon EMR

Compactação de dados de saída

Isso comprime a saída de seu trabalho do Hadoop. Se você estiver usando TextOutputFormat o resultado é um arquivo de texto compactado com gzip. Se você estiver escrevendo para SequenceFiles , o resultado SequenceFile será comprimido internamente. Para habilitar isso, defina a configuração mapred.output.compress como true.

Se você estiver executando um trabalho de streaming, poderá habilitar isso transmitido esses argumentos ao trabalho em questão.

```
-jobconf mapred.output.compress=true
```

Você também pode usar uma ação de bootstrap para compactar automaticamente todas as saídas do trabalho. Veja a seguir como fazer isso com o cliente Ruby.

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-hadoop \
--args "-s,mapred.output.compress=true"
```

Por fim, se você estiver escrevendo um JAR personalizado, poderá habilitar a compactação de saída com a seguinte linha ao criar seu trabalho.

```
FileOutputFormat.setCompressOutput(conf, true);
```

Compactação de dados intermediária

Se o seu trabalho embaralhar uma quantidade significativa de dados dos mapeadores para os reducers, você poderá ver uma melhoria de desempenho ao habilitar a compactação intermediária. Compacta a saída de mapa e a descompacta quando ela chega ao nó core. A definição de configuração é mapred.compress.map.output. Você pode habilitar isso da mesma forma que a compactação de saída.

Ao escrever um JAR personalizado, use o seguinte comando:

```
conf.setCompressMapOutput(true);
```

Usar a biblioteca Snappy com o Amazon EMR

A Snappy é uma biblioteca de compactação e descompactação otimizada para velocidade. Ela está disponível em AMIs do Amazon EMR 2.0 e versões posteriores e é usada como o padrão para a compactação intermediária. Para obter mais informações sobre a Snappy, acesse http://code.google.com/p/snappy/.

Planejar e configurar nós primários

Ao iniciar um cluster do Amazon EMR, você pode optar por ter um ou três nós primários no cluster. A alta disponibilidade, por exemplo, de frotas é suportada pelas versões 5.36.1, 5.36.2, 6.8.1, 6.9.1, 6.10.1, 6.11.1, 6.12.0 e superiores do Amazon EMR. Nos grupos de instâncias, a alta disponibilidade é compatível com as versões 5.23.0 e superiores do Amazon EMR. Para melhorar ainda mais a disponibilidade do cluster, o Amazon EMR pode usar os grupos de posicionamento do Amazon EC2 para garantir que os nós primários sejam colocados em outro hardware subjacente. Para ter mais informações, consulte Integração do Amazon EMR com grupos de posicionamento do EC2.

Um cluster do Amazon EMR com vários nós primários oferece os seguintes benefícios:

- O nó primário não é mais um ponto único de falha. Se um nó primário falhar, o cluster usará os outros dois nós primários e executará sem interrupção. Enquanto isso, o Amazon EMR substitui automaticamente o nó primário com falha por um novo que é provisionado com a mesma configuração e ações de bootstrap.
- O Amazon EMR habilita os recursos de alta disponibilidade do Hadoop do HDFS NameNode e do YARN ResourceManager e oferece suporte à alta disponibilidade para alguns outros aplicativos de código aberto.

  Para obter mais informações sobre como um cluster do Amazon EMR com múltiplos nós primários é compatível com aplicações de código aberto e atributos do EMR, consulte Aplicações e atributo compatíveis.

> **Note**
>
> O cluster pode residir apenas em uma zona de disponibilidade ou sub-rede.

Esta seção fornece informações sobre aplicações e atributos compatíveis de um cluster do Amazon EMR com múltiplos nós primários, bem como os detalhes de configuração, práticas recomendadas e considerações para iniciar o cluster.

Tópicos

Aplicações e atributo compatíveis

Este tópico fornece informações sobre os recursos de alta disponibilidade do Hadoop do HDFS NameNode e do YARN em ResourceManager um cluster do Amazon EMR e como os recursos de alta disponibilidade funcionam com aplicativos de código aberto e outros recursos do Amazon EMR.

Alta disponibilidade do HDFS

Um cluster do Amazon EMR com vários nós primários habilita o recurso de alta disponibilidade do HDFS NameNode no Hadoop. Para obter mais informações, consulte HDFS high availability.

Em um cluster do Amazon EMR, dois ou mais nós separados são configurados como. NameNodes Um NameNode está em um `active` estado e os outros estão em um `standby` estado. Se o nó com `active` NameNode falha, o Amazon EMR inicia um processo automático de failover do HDFS. Um nó `standby` NameNode se torna `active` e assume todas as operações do cliente no cluster. O Amazon EMR substitui o nó que falhou por um novo, que, por sua vez, reintegra-se como `standby`.

> **Note**
>
> Nas versões 5.23.0 do Amazon EMR até 5.30.1, inclusive, apenas dois dos três nós principais executam o HDFS. NameNode

Se precisar descobrir qual NameNode é`active`, você pode usar o SSH para se conectar a qualquer nó primário no cluster e executar o seguinte comando:

```
hdfs haadmin -getAllServiceState
```

A saída lista os nós em que NameNode está instalado e seu status. Por exemplo,

```
ip-##-#-#-##1.ec2.internal:8020 active
ip-##-#-#-##2.ec2.internal:8020 standby
ip-##-#-#-##3.ec2.internal:8020 standby
```

YARN de alta disponibilidade ResourceManager

Um cluster do Amazon EMR com vários nós primários habilita o recurso de ResourceManager alta disponibilidade do YARN no Hadoop. Para obter mais informações, consulte ResourceManager Alta disponibilidade.

Em um cluster do Amazon EMR com vários nós primários, o YARN ResourceManager é executado em todos os três nós primários. Um ResourceManager está no `active` estado e os outros dois estão no `standby` estado. Se o nó primário `active` ResourceManager falhar, o Amazon EMR iniciará um processo de failover automático. Um nó primário com `standby` ResourceManager a assume todas as operações. O Amazon EMR substitui o nó primário que falhou por um novo, que então volta ao quórum como um. ResourceManager `standby`

Você pode se conectar a “http://*`master-public-dns-name`*:8088/cluster” para qualquer nó primário, que automaticamente direciona você para o gerenciador de recursos `active`. Para descobrir qual gerenciador de recursos está `active`, use o SSH para se conectar a qualquer nó primário no cluster. Depois, execute o seguinte comando para obter uma lista com os três nós primários e os status deles:

```
yarn rmadmin -getAllServiceState
```

Aplicações compatíveis com um cluster do Amazon EMR com múltiplos nós primários

Você pode instalar e executar as aplicações a seguir em um cluster do Amazon EMR com múltiplos nós primários. Para cada aplicação, o processo de failover do nó primário varia.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Flink	Disponibilidade não afetada pelo failover do nó primário	Os trabalhos do Flink no Amazon EMR são executados como aplicações YARN. O Flink é JobManagers executado como YARN ApplicationMasters nos nós principais. O não JobManager é afetado pelo processo de failover do nó primário. Se você usa o Amazon EMR versão 5.27.0 ou anterior, esse JobManager é um único ponto de falha. Quando o JobManager falha, ele perde todos os estados de trabalho e não retoma os trabalhos em execução. Você pode habilitar a JobManager alta disponibi lidade configurando a contagem de tentativa s de aplicativos, o checkpoint e ativando o armazenamento ZooKeeper como estado para o Flink. Para obter mais informações, consulte Configuring Flink on an Amazon EMR Cluster with multiple primary nodes. A partir da versão 5.28.0 do Amazon EMR, nenhuma configuração manual é necessária para permitir a alta disponibilidade. JobManage r
Ganglia	Disponibilidade não afetada pelo failover do nó primário	O Ganglia está disponível em todos os nós primários e, portanto, continua em execução durante o processo de failover do nó primário.
Hadoop	Alta disponibilidade	O HDFS NameNode e o YARN ResourceM anager passam automaticamente para o nó em espera quando o nó primário ativo falha.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
HBase	Alta disponibilidade	O HBase faz o failover automático do nó em espera quando o nó primário ativo falha. Se você estiver se conectando ao HBase pelo servidor Thrift ou REST, é necessário alternar para outro nó primário quando o nó primário ativo falha.
HCatalog	Disponibilidade não afetada pelo failover do nó primário	O HCatalog é baseado no metastore do Hive, existente fora do cluster. O HCatalog permanece disponível durante o processo de failover do nó primário.
JupyterHub	Alta disponibilidade	JupyterHub está instalado em todas as três instâncias principais. É altamente recomendável configurar a persistência do caderno para evitar a perda do caderno após uma falha do nó primário. Para obter mais informações, consulte Configuring persistence for notebooks in Amazon S3.
Livy	Alta disponibilidade	O Livy é instalado em todos os três nós primários. Quando o nó primário ativo falha, você perde o acesso à sessão Livy atual e precisa criar uma nova sessão em outro nó primário ou no novo nó de substituição.
Mahout	Disponibilidade não afetada pelo failover do nó primário	Como o Mahout não tem daemons, ele não é afetado pelo processo de failover do nó primário.
MXNet	Disponibilidade não afetada pelo failover do nó primário	Como o MXNet não tem daemons, ele não é afetado pelo processo de failover do nó primário.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Phoenix	Alta disponibilidade	Phoenix QueryServer funciona apenas em um dos três nós primários. O Phoenix em todos os três mestres está configurado para conectar o Phoenix QueryServer. É possível encontrar o IP privado do servidor de consulta do Phoenix usando o arquivo `/etc/phoenix/conf/phoenix-env.sh`
Pig	Disponibilidade não afetada pelo failover do nó primário	Como o Pig não tem daemons, ele não é afetado pelo processo de failover do nó primário.
Spark	Alta disponibilidade	Todas as aplicações do Spark são executadas em contêineres do YARN e podem reagir ao failover do nó primário do mesmo modo que os recursos de alta disponibilidade do YARN.
Sqoop	Alta disponibilidade	Por padrão, sqoop-job e sqoop-metastore armazenam dados (descrições de trabalhos) no disco local do mestre que executa o comando. Se você deseja salvar dados do metastore no banco de dados externo, consulte a documentação do Apache Sqoop.
Tez	Alta disponibilidade	Como o contêineres do Tez são executados no YARN, o Tez se comporta da mesma forma que o YARN durante o processo de failover do nó primário.
TensorFlow	Disponibilidade não afetada pelo failover do nó primário	Como não TensorFlow tem daemon, ele não é afetado pelo processo de failover do nó primário.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Zeppelin	Alta disponibilidade	O Zeppelin é instalado em todos os três nós primários. O Zeppelin armazena notas e configurações de intérprete no HDFS por padrão para evitar a perda de dados. As sessões de intérprete são completamente isoladas em todas as três instâncias primárias. Os dados da sessão serão perdidos após uma falha da instância mestra. Recomenda-se não modificar a mesma nota simultaneamente em instâncias primárias diferentes.
ZooKeeper	Alta disponibilidade	ZooKeeper é a base do recurso de failover automático do HDFS. ZooKeeper fornece um serviço altamente disponível para manter dados de coordenação, notificar os clientes sobre alterações nesses dados e monitorar falhas nos clientes. Para obter mais informaçõ es, consulte HDFS automatic failover.

Para executar as seguintes aplicações em um cluster do Amazon EMR com múltiplos nós primários, é necessário configurar um banco de dados externo. O banco de dados externo existe fora do cluster e torna os dados persistentes durante o processo de failover do nó primário. Para as aplicações a seguir, os componentes de serviço serão recuperados automaticamente durante o processo de failover do nó primário, mas os trabalhos ativos podem falhar e precisam ser repetidos.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
Hive	Alta disponibilidade somente para componentes de serviço	É necessário um metastore externo para o Hive. Deve ser um metastore externo do MySQL, pois o PostgreSQL não é compatíve l com clusters multiprincipais. Para obter mais

Aplicativo	Disponibilidade durante failover do nó primário	Observações
		informações, consulte Configuring an external metastore for Hive.
Hue	Alta disponibilidade somente para componentes de serviço	É necessário um banco de dados externo para o Hue. Para obter mais informações, consulte Using Hue with a remote database in Amazon RDS.
Oozie	Alta disponibilidade somente para componentes de serviço	É necessário um banco de dados externo para o Oozie. Para obter mais informações, consulte Using Oozie with a remote database in Amazon RDS. O Oozie-server e o oozie-client são instalados nos três nós primários. Os oozie-clients são configurados para se conectar ao oozie-server correto por padrão.

Aplicativo	Disponibilidade durante failover do nó primário	Observações
PrestoDB ou PrestoSQL/ Trino	Alta disponibilidade somente para componentes de serviço	É necessário ter um metastore externo do Hive para PrestoDB (PrestoSQL no Amazon EMR 6.1.0-6.3.0 ou Trino no Amazon EMR 6.4.0 e versões posteriores). Você pode usar o Presto com o AWS Glue Data Catalog ou usar um banco de dados MySQL externo para o Hive. A CLI do Presto está instalada nos três nós primários, e assim você pode usá-la para acessar o coordenador do Presto por qualquer um dos nós primários. O coordenador do Presto é instalado em apenas um nó primário. Você encontra o nome DNS do nó primário em que o coordenador do Presto está instalado chamando a API `describe-cluster` do Amazon EMR e lendo o valor retornado do campo `MasterPublicDnsName` na resposta.

> **Note**
>
> Quando um nó primário falha, o Java Database Connectivity (JDBC) ou o Open Database Connectivity (ODBC) termina a conexão com o nó primário. Você pode se conectar a qualquer um dos nós primários restantes para continuar o trabalho, pois o daemon do Hive Metastore é executado em todos os nós primários. Ou você pode esperar a substituição do nó primário com falha.

Como os atributos do Amazon EMR funcionam em um cluster com múltiplos nós principais

Conectar aos nós primários usando SSH

Você pode se conectar a qualquer um dos três nós primários de um cluster do Amazon EMR usando o SSH da mesma maneira que conecta a um único nó primário. Para obter mais informações, consulte Connect to the primary node using SSH.

Se um nó primário falha, sua conexão SSH ao nó primário encerra. Para que ela continue funcionando, você pode se conectar a um dos outros dois nós primários. Como alternativa, você pode acessar o novo nó primário após o Amazon EMR substituir o que falhou por um novo.

> **Note**
>
> O endereço IP privado para o nó primário de substituição permanece o mesmo que o anterior. O endereço IP público para o nó primário de substituição pode mudar. É possível recuperar os novos endereços IP no console ou usando o comando `describe-cluster` na AWS CLI.
> NameNode só funciona em dois dos nós primários. No entanto, você pode executar comandos `hdfs` da CLI e operar trabalhos para acessar o HDFS em todos os três nós primários.

Trabalhar com etapas em um cluster do Amazon EMR com múltiplos nós primários

Você pode enviar etapas a um cluster do Amazon EMR com múltiplos nós primários da mesma maneira que você trabalha com as etapas em um cluster com um único nó primário. Para obter mais informações, consulte Submit work to a cluster.

Veja aqui algumas considerações para trabalhar com etapas em um cluster do Amazon EMR com múltiplos nós primários:

- Se um nó primário falha, as etapas sendo executadas no nó primário são marcadas como FAILED. Todos os dados que foram gravados localmente são perdidos. No entanto, o status FAILED pode não refletir o estado real das etapas.
- Se uma etapa em execução iniciou uma aplicação do YARN antes da falha do nó primário, ela pode continuar e ter êxito devido ao failover automático do nó primário.

- Recomendamos que você verifique os status das etapas consultando a saída dos trabalhos. Por exemplo, os MapReduce trabalhos usam um `_SUCCESS` arquivo para determinar se o trabalho foi concluído com êxito.
- É recomendável definir o ActionOnFailure parâmetro como CONTINUE ou CANCEL_AND_WAIT, em vez de TERMINATE_JOB_FLOW ou TERMINATE_CLUSTER.

Proteção automática de término

O Amazon EMR habilita automaticamente a proteção contra término para todos os clusters com múltiplos nós primários e substitui as configurações de execução de etapas fornecidas na criação do cluster. É possível desabilitar a proteção contra término depois que o cluster é iniciado. Consulte Configurar a proteção contra término para clusters em execução. Para desligar um cluster com múltiplos nós primários, primeiro é necessário modificar os atributos do cluster para desabilitar a proteção contra término. Para obter instruções, consulte Terminar um cluster do Amazon EMR com múltiplos nós primários.

Para obter mais informações sobre proteção contra término, consulte Usar a proteção contra término.

Atributos incompatíveis em um cluster do Amazon EMR com múltiplos nós primários

No momento, os seguintes recursos do Amazon EMR não estão disponíveis em clusters do Amazon EMR com múltiplos nós primários:

- Cadernos do EMR
- Acesso com um clique ao servidor de histórico persistente do Spark
- Interfaces do usuário de aplicações persistentes
- No momento, o acesso com um clique às interfaces de usuário de aplicativos persistentes não está disponível para clusters do Amazon EMR com vários nós primários ou para clusters do Amazon EMR integrados ao Lake Formation. AWS

> **Note**
>
> Para usar a autenticação do Kerberos no seu cluster, é necessário configurar um KDC externo.
>
> A partir do Amazon EMR 5.27.0, é possível configurar a criptografia transparente do HDFS em um cluster do Amazon EMR com múltiplos nós primários. Para obter mais informações, consulte Transparent encryption in HDFS on Amazon EMR.

Iniciar um cluster do Amazon EMR com múltiplos nós primários

Este tópico fornece detalhes de configuração e exemplos para iniciar um cluster do Amazon EMR com múltiplos nós primários.

> **Note**
>
> O Amazon EMR habilita automaticamente a proteção contra encerramento para todos os clusters com vários nós primários e substitui as configurações de encerramento automático fornecidas na criação do cluster. Para desligar um cluster com múltiplos nós primários, primeiro é necessário modificar os atributos do cluster para desabilitar a proteção contra término. Para obter instruções, consulte Terminar um cluster do Amazon EMR com múltiplos nós primários.

Pré-requisitos

- Você pode iniciar um cluster do Amazon EMR com múltiplos nós primários em sub-redes públicas e privadas da VPC. O EC2-Classic não é compatível. Para iniciar um cluster do Amazon EMR com múltiplos nós primários em uma sub-rede pública, é necessário ativar as instâncias nessa sub-rede para receber um endereço IP público, selecionando Atribuir IPv4 automaticamente no console ou executando o comando a seguir. Substitua *22XXXX01* pelo ID da sua sub-rede.

  ```
  aws ec2 modify-subnet-attribute --subnet-id subnet-22XXXX01 --map-public-ip-on-launch
  ```

- Para executar o Hive, o Hue ou o Oozie em um cluster do Amazon EMR com múltiplos nós primários, é necessário criar um metastore externo. Para obter mais informações, consulte Configuring an external metastore for Hive, Using Hue with a remote database in Amazon RDS ou Apache Oozie.
- Para usar a autenticação do Kerberos no seu cluster, é necessário configurar um KDC externo. Para obter mais informações, consulte Configuring Kerberos on Amazon EMR.

Iniciar um cluster do Amazon EMR com múltiplos nós primários

Você pode executar um cluster com vários nós primários ao usar grupos ou frotas de instâncias. Ao usar os grupos de instâncias com vários nós primários, é preciso especificar um valor 3 de contagem de instâncias para o grupo de instâncias do nó primário. Ao usar frotas de instâncias com vários nós

primários, você deve especificar a `TargetOnDemandCapacity` de `3`, a `TargetSpotCapacity` de `0` para a frota de instâncias primária e a `WeightedCapacity` de `1` para cada tipo de instância que configurar para a frota principal.

Os seguintes exemplos demonstram como executar o cluster usando a AMI padrão ou uma AMI personalizada com grupos e frotas de instâncias.

> **Note**
>
> É necessário especificar o ID da sub-rede ao iniciar um cluster do Amazon EMR com múltiplos nós primários usando a AWS CLI. Substitua *22XXXX01* e *22XXXX02* pelo ID da sua sub-rede nos exemplos a seguir.

Default AMI, instance groups

Example Exemplo: executar um cluster de grupo de instâncias do Amazon EMR com vários nós primários usando uma AMI padrão

```
aws emr create-cluster \
--name "ha-cluster" \
--release-label emr-6.15.0 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

Default AMI, instance fleets

Example Exemplo: executar um cluster de frota de instâncias do Amazon EMR com vários nós primários usando uma AMI padrão

```
aws emr create-cluster \
--name "ha-cluster" \
--release-label emr-6.15.0 \
--instance-fleets '[
    {
        "InstanceFleetType": "MASTER",
        "TargetOnDemandCapacity": 3,
```

```
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Master - 1"
    },
    {
        "InstanceFleetType": "CORE",
        "TargetOnDemandCapacity": 5,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 2,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
```

```
            },
            {
                "WeightedCapacity": 4,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Core - 2"
    }
]' \
 --ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":
["subnet-22XXXX01", "subnet-22XXXX02"]}' \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

Custom AMI, instance groups

Example Exemplo: executar um cluster de grupo de instâncias do Amazon EMR com vários nós primários usando uma AMI personalizada

```
aws emr create-cluster \
--name "custom-ami-ha-cluster" \
--release-label emr-6.15.0 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark \
--custom-ami-id ami-MyAmiID
```

Custom AMI, instance fleets

Example Exemplo: executar um cluster de frota de instâncias do Amazon EMR com vários nós primários usando uma AMI personalizada

```
aws emr create-cluster \
--name "ha-cluster" \
--release-label emr-6.15.0 \
--instance-fleets '[
    {
        "InstanceFleetType": "MASTER",
```

```
        "TargetOnDemandCapacity": 3,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Master - 1"
    },
    {
        "InstanceFleetType": "CORE",
        "TargetOnDemandCapacity": 5,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 2,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
```

```
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 4,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Core - 2"
    }
]' \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":
["subnet-22XXXX01", "subnet-22XXXX02"]}' \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark \
--custom-ami-id ami-MyAmiID
```

Terminar um cluster do Amazon EMR com múltiplos nós primários

Para terminar um cluster do Amazon EMR com múltiplos nós primários, é preciso desabilitar a proteção contra término o antes de finalizar o cluster, como mostra o exemplo a seguir. Substitua *j-3KVTXXXXXX7UG* pelo ID do seu cluster.

```
aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-
protected
aws emr terminate-clusters --cluster-id j-3KVTXXXXXX7UG
```

Integração do Amazon EMR com grupos de posicionamento do EC2

Ao iniciar um cluster de múltiplos nós primários do Amazon EMR no Amazon EC2, você tem a opção de usar estratégias de grupos de posicionamento para especificar como deseja que as instâncias do nó primário sejam implantadas para se protegerem contra falhas de hardware.

Há suporte para estratégias de grupos de posicionamento a partir do Amazon EMR versão 5.23.0 como uma opção para clusters com múltiplos nós primários. Atualmente, somente os tipos de nós primários são compatíveis com a estratégia de grupo de posicionamento, e a estratégia `SPREAD` é aplicada a estes nós primários. A estratégia `SPREAD` posiciona um pequeno grupo de instâncias em um hardware subjacente separado para evitar a perda de múltiplos nós primários em caso de falha de hardware. Uma solicitação de inicialização de instância poderá falhar se não houver hardware exclusivo suficiente para atender à solicitação. Para obter mais informações sobre as estratégias e

limitações de posicionamento do EC2, consulte Grupos de posicionamento no Guia do usuário do EC2 para instâncias do Linux.

Há um limite inicial do Amazon EC2 de 500 clusters habilitados para estratégias de grupos de posicionamento que podem ser lançados por região. AWS Entre em contato com o AWS suporte para solicitar um aumento no número de grupos de colocação permitidos. É possível identificar os grupos de posicionamento do EC2 que o Amazon EMR cria rastreando o par de chave-valor que o Amazon EMR associa à estratégia de grupos de posicionamento do Amazon EMR. Para obter mais informações sobre etiquetas de instância de cluster do EC2, consulte Visualizar instâncias de cluster no Amazon EC2.

Anexar a política gerenciada de grupo de posicionamento ao Amazon EMRrole

A estratégia de grupos de posicionamento exige uma política gerenciada chamada `AmazonElasticMapReducePlacementGroupPolicy`. Com ela, o Amazon EMR pode criar, excluir e descrever grupos de posicionamento no Amazon EC2. É necessário anexar `AmazonElasticMapReducePlacementGroupPolicy` ao perfil de serviço do Amazon EMR antes de executar um cluster do Amazon EMR com vários nós primários.

Como alternativa, você pode anexar a política gerenciada `AmazonEMRServicePolicy_v2` ao perfil de serviço do Amazon EMR em vez da política gerenciada de grupo de posicionamento. `AmazonEMRServicePolicy_v2` e `AmazonElasticMapReducePlacementGroupPolicy` permitem os mesmos acessos a grupos de posicionamento no Amazon EC2. Para ter mais informações, consulte Perfil de serviço para Amazon EMR (perfil do EMR).

A política gerenciada `AmazonElasticMapReducePlacementGroupPolicy` é o texto JSON a seguir criado e administrado pelo Amazon EMR.

> **Note**
>
> Como a política `AmazonElasticMapReducePlacementGroupPolicy` gerenciada é atualizada automaticamente, a política mostrada aqui pode ser out-of-date. Use o AWS Management Console para visualizar a política atual.

```
{
   "Version":"2012-10-17",
   "Statement":[
```

```
      {
         "Resource":"*",
         "Effect":"Allow",
         "Action":[
            "ec2:DeletePlacementGroup",
            "ec2:DescribePlacementGroups"
         ]
      },
      {
         "Resource":"arn:aws:ec2:*:*:placement-group/pg-*",
         "Effect":"Allow",
         "Action":[
            "ec2:CreatePlacementGroup"
         ]
      }
   ]
}
```

Execução de um cluster do Amazon EMR com vários nós primários usando estratégia de grupos de posicionamento

Para executar um cluster do Amazon EMR com vários nós primários usando uma estratégia de grupos de posicionamento, anexe a política gerenciada de grupos de posicionamento `AmazonElasticMapReducePlacementGroupPolicy` ao perfil do Amazon EMR. Para ter mais informações, consulte Anexar a política gerenciada de grupo de posicionamento ao Amazon EMRrole.

Toda vez que você usa esse perfil para iniciar um cluster do Amazon EMR com vários nós primários, o Amazon EMR tenta executar um cluster com a estratégia `SPREAD` aplicada aos nós primários. Se você usar um perfil que não tenha a política gerenciada de grupos de posicionamento `AmazonElasticMapReducePlacementGroupPolicy` anexada, o Amazon EMR tentará executar um cluster do Amazon EMR com vários nós primários sem uma estratégia de grupos de posicionamento.

Se você executar um cluster do Amazon EMR que tenha vários nós primários com o parâmetro `placement-group-configs` usando a API ou a CLI do Amazon EMR, o Amazon EMR só executará o cluster se o perfil do Amazon EMR tiver a política gerenciada de grupos de posicionamento `AmazonElasticMapReducePlacementGroupPolicy` anexada. Se o perfil do Amazon EMR não tiver a política anexada, o cluster do Amazon EMR com vários nós primários não será executado.

Amazon EMR API

Example Exemplo: usar uma estratégia de grupos de posicionamento para executar um cluster de grupos de instâncias com vários nós primários da API do Amazon EMR

Ao usar a RunJobFlow ação para criar um cluster do Amazon EMR com vários nós primários, defina a `PlacementGroupConfigs` propriedade da seguinte forma. Atualmente, o perfil de instância `MASTER` usa automaticamente `SPREAD` como estratégia de grupo de posicionamento.

```
{
   "Name":"ha-cluster",
   "PlacementGroupConfigs":[
      {
         "InstanceRole":"MASTER"
      }
   ],
   "ReleaseLabel": emr-6.15.0,
   "Instances":{
      "ec2SubnetId":"subnet-22XXXX01",
      "ec2KeyName":"ec2_key_pair_name",
      "InstanceGroups":[
         {
            "InstanceCount":3,
            "InstanceRole":"MASTER",
            "InstanceType":"m5.xlarge"
         },
         {
            "InstanceCount":4,
            "InstanceRole":"CORE",
            "InstanceType":"m5.xlarge"
         }
      ]
   },
   "JobFlowRole":"EMR_EC2_DefaultRole",
   "ServiceRole":"EMR_DefaultRole"
}
```

- Substitua *`ha-cluster`* pelo nome de seu cluster de alta disponibilidade.
- Substitua *`subnet-22XXXX01`* pelo ID da sub-rede.

- Substitua *ec2_key_pair_name* pelo nome do par de chaves do EC2 para esse cluster. O par de chaves do EC2 é opcional e necessário somente se você quiser usar SSH para acessar seu cluster.

AWS CLI

Example Exemplo: usar uma estratégia de grupos de posicionamento para executar um cluster de frotas de instâncias com vários nós primários da AWS Command Line Interface

Ao usar a RunJobFlow ação para criar um cluster do Amazon EMR com vários nós primários, defina a `PlacementGroupConfigs` propriedade da seguinte forma. Atualmente, o perfil de instância `MASTER` usa automaticamente `SPREAD` como estratégia de grupo de posicionamento.

```
aws emr create-cluster \
--name "ha-cluster" \
--placement-group-configs InstanceRole=MASTER \
--release-label emr-6.15.0 \
--instance-fleets '[
    {
        "InstanceFleetType": "MASTER",
        "TargetOnDemandCapacity": 3,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
```

```
            }
        ],
        "Name": "Master - 1"
    },
    {
        "InstanceFleetType": "CORE",
        "TargetOnDemandCapacity": 5,
        "TargetSpotCapacity": 0,
        "LaunchSpecifications": {
            "OnDemandSpecification": {
                "AllocationStrategy": "lowest-price"
            }
        },
        "InstanceTypeConfigs": [
            {
                "WeightedCapacity": 1,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.xlarge"
            },
            {
                "WeightedCapacity": 2,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.2xlarge"
            },
            {
                "WeightedCapacity": 4,
                "BidPriceAsPercentageOfOnDemandPrice": 100,
                "InstanceType": "m5.4xlarge"
            }
        ],
        "Name": "Core - 2"
    }
]' \
--ec2-attributes '{
    "KeyName": "ec2_key_pair_name",
    "InstanceProfile": "EMR_EC2_DefaultRole",
    "SubnetIds": [
        "subnet-22XXXX01",
        "subnet-22XXXX02"
    ]
}' \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

- Substitua *ha-cluster* pelo nome de seu cluster de alta disponibilidade.
- Substitua *ec2_key_pair_name* pelo nome do par de chaves do EC2 para esse cluster. O par de chaves do EC2 é opcional e necessário somente se você quiser usar SSH para acessar seu cluster.
- Substitua *subnet-22XXXX01* e *subnet-22XXXX02* pelos IDs da sua sub-rede.

Iniciar um cluster com múltiplos nós primários sem uma estratégia de grupos de posicionamento

Para que um cluster com múltiplos nós primários inicie nós primários sem a estratégia de grupos de posicionamento, é necessário:

- Remover a política gerenciada de grupo de posicionamento `AmazonElasticMapReducePlacementGroupPolicy` do Amazon EMRrole ou
- Inicie um cluster com múltiplos nós primários com o parâmetro `placement-group-configs` usando a Amazon EMRAPI ou a CLI, escolhendo `NONE` como estratégia de grupos de posicionamento.

Amazon EMR API

Example — Iniciar um cluster com múltiplos nós primários sem estratégia de grupos de posicionamento usando a Amazon EMRAPI.

Ao usar a RunJobFlow ação para criar um cluster com vários nós primários, defina a `PlacementGroupConfigs` propriedade da seguinte forma.

```
{
   "Name":"ha-cluster",
   "PlacementGroupConfigs":[
      {
         "InstanceRole":"MASTER",
         "PlacementStrategy":"NONE"
      }
   ],
   "ReleaseLabel":"emr-5.30.1",
   "Instances":{
      "ec2SubnetId":"subnet-22XXXX01",
      "ec2KeyName":"ec2_key_pair_name",
```

```
      "InstanceGroups":[
         {
            "InstanceCount":3,
            "InstanceRole":"MASTER",
            "InstanceType":"m5.xlarge"
         },
         {
            "InstanceCount":4,
            "InstanceRole":"CORE",
            "InstanceType":"m5.xlarge"
         }
      ]
   },
   "JobFlowRole":"EMR_EC2_DefaultRole",
   "ServiceRole":"EMR_DefaultRole"
}
```

- Substitua *ha-cluster* pelo nome de seu cluster de alta disponibilidade.
- Substitua *subnet-22XXXX01* pelo ID da sub-rede.
- Substitua *ec2_key_pair_name* pelo nome do par de chaves do EC2 para esse cluster. O par de chaves do EC2 é opcional e necessário somente se você quiser usar SSH para acessar seu cluster.

Amazon EMR CLI

Example — Iniciar um cluster com múltiplos nós primários sem uma estratégia de grupos de posicionamento usando a Amazon EMRCLI.

Ao usar a RunJobFlow ação para criar um cluster com vários nós primários, defina a `PlacementGroupConfigs` propriedade da seguinte forma.

```
aws emr create-cluster \
--name "ha-cluster" \
--placement-group-configs InstanceRole=MASTER,PlacementStrategy=NONE \
--release-label emr-5.30.1 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
```

```
--applications Name=Hadoop Name=Spark
```

- Substitua *ha-cluster* pelo nome de seu cluster de alta disponibilidade.
- Substitua *subnet-22XXXX01* pelo ID da sub-rede.
- Substitua *ec2_key_pair_name* pelo nome do par de chaves do EC2 para esse cluster. O par de chaves do EC2 é opcional e necessário somente se você quiser usar SSH para acessar seu cluster.

Verificar a configuração da estratégia de grupos de posicionamento anexada ao cluster com múltiplos nós primários

Use a API de descrição do cluster do Amazon EMR para ver a configuração da estratégia de grupos de posicionamento anexada ao cluster com múltiplos nós primários.

Example

```
aws emr describe-cluster --cluster-id "j-xxxxx"
{
   "Cluster":{
      "Id":"j-xxxxx",
      ...
      ...
      "PlacementGroups":[
         {
            "InstanceRole":"MASTER",
            "PlacementStrategy":"SPREAD"
         }
      ]
   }
}
```

Considerações e práticas recomendadas

Considere os seguintes pontos ao criar um cluster do Amazon EMR com vários nós primários:

> **Important**
>
> Para executar clusters de alta disponibilidade do EMR com vários nós primários, é altamente recomendável usar a versão mais recente do Amazon EMR. Isso garante que você obtenha o mais alto nível de resiliência e estabilidade para os seus clusters de alta disponibilidade.

- A alta disponibilidade, por exemplo, de frotas é suportada pelas versões 5.36.1, 5.36.2, 6.8.1, 6.9.1, 6.10.1, 6.11.1, 6.12.0 e superiores do Amazon EMR. Nos grupos de instâncias, a alta disponibilidade é compatível com as versões 5.23.0 e superiores do Amazon EMR. Para saber mais, consulte Sobre as versões do Amazon EMR.
- Em clusters de alta disponibilidade, o Amazon EMR só oferece suporte à execução de nós primários com instâncias sob demanda. Isso garante a maior disponibilidade para o seu cluster.
- Você ainda pode especificar vários tipos de instância para a frota primária, mas todos os nós primários de clusters de alta disponibilidade são executados com o mesmo tipo de instância, incluindo substituições de nós primários não íntegros.
- Para continuar as operações, um cluster de alta disponibilidade com vários nós primários exige que dois dos três nós primários estejam íntegros. Como resultado, se dois nós primários falharem simultaneamente, o cluster do EMR falhará.
- Todos os clusters do EMR, incluindo os de alta disponibilidade, são executados em uma única zona de disponibilidade. Portanto, eles não são tolerantes a falhas na zona de disponibilidade. Se houver uma interrupção na zona de disponibilidade, você perde o acesso ao cluster.
- O Amazon EMR não garante alta disponibilidade para aplicações de código aberto que não estejam especificadas em Aplicações compatíveis com um cluster do Amazon EMR com múltiplos nós primários.
- Nas versões 5.23.0 a 5.30.1 do Amazon EMR, somente dois dos três nós primários de um cluster de grupos de instâncias executam o HDFS NameNode.

Considerações para configurar a sub-rede:

- Um cluster do Amazon EMR com múltiplos nós primários pode residir somente em uma zona de disponibilidade ou sub-rede. O Amazon EMR não conseguirá substituir um nó primário com falha se a sub-rede estiver sendo utilizada totalmente ou em excesso no caso de failover. Para evitar esse cenário, é recomendável dedicar uma sub-rede a um cluster do Amazon EMR. Além disso, certifique-se de que há endereços IP privados suficientes disponíveis na sub-rede.

Considerações para configurar nós core:

- Para também garantir a alta disponibilidade dos nós centrais, é recomendável executar pelo menos quatro nós centrais. Se você decidir iniciar um cluster com três nós centrais ou menos, configure `dfs.replication parameter` como, no mínimo, 2 para o HDFS ter replicação DFS suficiente. Para obter mais informações, consulte HDFS configuration.

> ⚠ Warning
>
> 1. Definir `dfs.replication` como 1 em clusters com menos de quatro nós poderá causar perda de dados do HDFS se um único nó ficar inativo. É recomendável usar um cluster com pelo menos quatro nós centrais para workloads de produção.
> 2. O Amazon EMR não permitirá que os clusters escalem os nós principais abaixo de `dfs.replication`. Por exemplo, se `dfs.replication = 2`, o número mínimo de nós central será 2.
> 3. Ao usar o Ajuste de Escala Gerenciado, o Auto Scaling ou optar por redimensionar manualmente o cluster, é recomendável definir `dfs.replication` como 2 ou mais.

Considerações para configurar alarmes em métricas:

- O Amazon EMR não fornece métricas específicas da aplicação sobre o HDFS ou YARN. É recomendável definir alarmes para monitorar a contagem de instâncias para nós primários. Configure os alarmes usando as seguintes CloudWatch métricas da Amazon:`MultiMasterInstanceGroupNodesRunning`,`MultiMasterInstanceGroupNodesRunning` ou`MultiMasterInstanceGroupNodesRequested`. CloudWatch notificará você em caso de falha e substituição do nó primário.
  - Se `MultiMasterInstanceGroupNodesRunningPercentage` for menor que 1,0 e maior que 0,5, o cluster pode ter perdido um nó primário. Nesse caso, o Amazon EMR tenta substituir um nó primário.
  - Se a `MultiMasterInstanceGroupNodesRunningPercentage` ficar abaixo de 0,5, dois nós primários podem ter falhado. Nesse caso, o quórum do cluster é perdido e não é possível recuperá-lo. É necessário migrar os dados desse cluster manualmente.

  Para obter mais informações, consulte Setting alarms on metrics.

Clusters EMR em AWS Outposts

A partir do Amazon EMR 5.28.0, você pode criar e executar clusters do EMR no. AWS Outposts AWS Outposts habilita AWS serviços, infraestrutura e modelos operacionais nativos em instalações locais. Em AWS Outposts ambientes, você pode usar as mesmas AWS APIs, ferramentas e infraestrutura que usa na AWS nuvem. O Amazon EMR on AWS Outposts é ideal para cargas de trabalho de baixa latência que precisam ser executadas nas proximidades de dados e aplicativos locais. Para obter mais informações sobre AWS Outposts, consulte o Guia AWS Outposts do usuário.

Pré-requisitos

Veja a seguir os pré-requisitos para usar o Amazon EMR no AWS Outposts:

- Você deve ter instalado e configurado AWS Outposts em seu data center local.
- Você deve ter uma conexão de rede confiável entre seu ambiente Outpost e uma AWS região.
- Você deve ter capacidade suficiente para os tipos de instância compatíveis com o Amazon EMR disponíveis em seu Outpost.

Limitações

Veja a seguir as limitações de usar o Amazon EMR no AWS Outposts:

- As instâncias sob demanda são a única opção compatível com as instâncias do Amazon EC2. As instâncias spot não estão disponíveis para o Amazon EMR no AWS Outposts.
- Se você precisar de volumes de armazenamento adicionais do Amazon EBS, somente SSD de uso geral (GP2) é compatível.
- Quando você usa AWS Outposts com as versões 5.28 a 6.x do Amazon EMR, você só pode usar buckets do S3 que armazenam objetos em um que você especificar. Região da AWS Com o Amazon EMR 7.0.0 e versões posteriores, o Amazon EMR on também AWS Outposts é compatível com o prefixo do cliente do S3A sistema de arquivos. `s3a://`
- Somente os seguintes tipos de instância são compatíveis com o Amazon EMR no AWS Outposts:

Classe de instância	Tipos de instância
Propósito geral	`m5.xlarge | m5.2xlarge | m5.4xlarge | m5.12xlarge | m5.24xlarge | m5d.xlarge | m5d.2xlarge | m5d.4xlarge | m5d.12xlarge | m5d.24xlarge`
Otimizada para computação	`c5.xlarge | c5.2xlarge | c5.4xlarge | c5.18xlarge | c5d.xlarge | c5d.2xlarge | c5d.4xlarge | c5d.18xlarge`
Otimizada para memória	`r5.xlarge | r5.2xlarge | r5.4xlarge | r5.12xlarge | r5d.xlarge | r5d.2xlarge | r5d.4xlarge | r5d.12xlarge | r5d.24xlarge`
Otimizada para armazenamento	`i3en.xlarge | i3en.2xlarge | i3en.3xlarge | i3en.6xlarge | i3en.12xlarge | i3en.24xlarge`

Considerações sobre a conectividade de rede

- Se a conectividade de rede entre seu Posto Avançado e sua AWS região for perdida, seus clusters continuarão funcionando. No entanto, você não pode criar novos clusters ou executar novas ações em clusters existentes enquanto que a conectividade não for restaurada. Em caso de falhas na instância, a instância não será substituída automaticamente. Além disso, ações como adicionar etapas a um cluster em execução, verificar o status de execução das etapas e enviar CloudWatch métricas e eventos serão adiadas.
- Recomendamos que você forneça conectividade de rede confiável e altamente disponível entre seu Posto Avançado e a AWS Região. Se a conectividade de rede entre seu Posto Avançado e sua AWS região for perdida por mais de algumas horas, os clusters que ativaram a proteção de encerramento continuarão funcionando e os clusters que desativaram a proteção de encerramento poderão ser encerrados.
- Se a conectividade da rede for afetada por uma manutenção de rotina, recomendamos que a proteção contra encerramento seja ativada proativamente. De modo geral, a interrupção da conectividade significa que quaisquer dependências externas que não sejam locais para a rede do

cliente ou o para Outpost ficarão inacessíveis. Isso inclui o Amazon S3, o DynamoDB usado com a visualização de consistência do EMRFS e o Amazon RDS, se uma instância na região for usada para um cluster do Amazon EMR com múltiplos nós primários.

Criação de um cluster do Amazon EMR em AWS Outposts

Criar um cluster do Amazon EMR no AWS Outposts é semelhante à criação de um cluster do Amazon EMR na nuvem. AWS Ao criar um cluster do Amazon EMR no AWS Outposts, você deve especificar uma sub-rede do Amazon EC2 associada ao seu Outpost.

Uma Amazon VPC pode abranger todas as zonas de disponibilidade em uma AWS região. AWS Outposts são extensões das zonas de disponibilidade, e você pode estender uma Amazon VPC em uma conta para abranger várias zonas de disponibilidade e locais associados do Outpost. Ao configurar o Outpost, você associa uma sub-rede a ele para estender o ambiente regional da VPC à instalação on-premises. As instâncias do Outpost e serviços relacionados aparecem como parte de sua VPC regional, semelhante a uma zona de disponibilidade com sub-redes associadas. Para obter mais informações, consulte o Guia do usuário do AWS Outposts.

Console

Para criar um novo cluster do Amazon EMR AWS Outposts com o AWS Management Console, especifique uma sub-rede do Amazon EC2 associada ao seu Outpost.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Para criar um cluster AWS Outposts com o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Configuração do cluster, selecione Grupos de instâncias ou Frotas de instâncias. Em seguida, escolha um tipo de instância no menu suspenso Escolher tipo de instância do EC2

ou selecione Ações e escolha Adicionar volumes do EBS. O Amazon EMR on AWS Outposts oferece suporte a tipos limitados de volume e instância do Amazon EBS.

4. Em Redes, selecione uma sub-rede EC2 com um ID do Outpost neste formato: op-123456789.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

Old console

Para criar um cluster AWS Outposts com o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Escolha Go to advanced options (Ir para opções avançadas).
4. Em Software Configuration (Configuração de software), Release (Versão), escolha 5.28.0 ou posterior.
5. Em Configuração de hardware, para EC2 Subnet, selecione uma sub-rede Amazon EC2 com um Outpost ID neste formato: op-123456789.
6. Escolha o tipo de instância ou adicione volumes de armazenamento do Amazon EBS aos grupos de instâncias uniformes ou frotas de instâncias. Os tipos limitados de volumes e instâncias do Amazon EBS são compatíveis com o Amazon EMR no AWS Outposts.

CLI

Para criar um cluster AWS Outposts com o AWS CLI

- Para criar um novo cluster do Amazon EMR AWS Outposts com o AWS CLI, especifique uma sub-rede EC2 associada ao seu Outpost, como no exemplo a seguir. Substitua a *sub-net-22xxxx01* pelo seu próprio ID de sub-rede do Amazon EC2.

```
aws emr create-cluster \
--name "Outpost cluster" \
--release-label emr-7.1.0 \
--applications Name=Spark \
```

```
--ec2-attributes KeyName=myKey SubnetId=subnet-22XXXX01 \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Clusters EMR em Locais Zones AWS

A partir da versão 5.28.0 do Amazon EMR, você pode criar e executar clusters do Amazon EMR em uma sub-rede de Zonas AWS Locais como uma extensão lógica de uma região que suporta Zonas Locais. AWS Uma zona local permite que os recursos do Amazon EMR e um subconjunto de AWS serviços, como serviços de computação e armazenamento, estejam localizados mais perto dos usuários para fornecer acesso de latência muito baixa aos aplicativos executados localmente. Para obter uma lista das zonas locais disponíveis, consulte Zonas locais da AWS. Para obter informações sobre como acessar as Zonas AWS Locais disponíveis, consulte Regiões, Zonas de Disponibilidade e zonas locais.

Tipos de instâncias compatíveis

Os tipos de instância a seguir estão disponíveis para clusters do Amazon EMR em zonas locais. A disponibilidade do tipo de instância pode variar de acordo com a região.

Classe de instância	Tipos de instância
Propósito geral	m5.xlarge | m5.2xlarge | m5.4xlarge | m5.12xlarge | m5.24xlarge | m5d.xlarge | m5d.2xlarge | m5d.4xlarge | m5d.12xlarge | m5d.24xlarge
Otimizada para computação	c5.xlarge | c5.2xlarge | c5.4xlarge | c5.9xlarge | c5.18xlarge | c5d.xlarge | c5d.2xlarge | c5d.4xlarge| c5d.9xlarge | c5d.18xlarge
Otimizada para memória	r5.xlarge | r5.2xlarge | r5.4xlarge | r5.12xlarge | r5d.xlarge | r5d.2xlarge | r5d.4xlarge | r5d.12xlarge | r5d.24xlarge
Otimizada para armazenamento	i3en.xlarge | i3en.2xlarge | i3en.3xlarge | i3en.6xlarge | i3en.12xlarge | i3en.24xlarge

Criar um cluster do Amazon EMR em zonas locais

Crie um cluster do Amazon EMR em Zonas AWS Locais lançando o cluster do Amazon EMR em uma sub-rede do Amazon VPC associada a uma Zona Local. É possível acessar o cluster usando o nome da zona local, como us-west-2-lax-1a, no console da região Oeste dos EUA (Oregon).

Atualmente, as Zonas Locais não oferecem suporte a notebooks Amazon EMR ou conexões diretamente ao Amazon EMR usando a interface VPC endpoint ().AWS PrivateLink

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Criar um cluster em uma zona local usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Rede, selecione uma sub-rede do EC2 com um ID de zona local neste formato: 123abc | us-west-2-lax-1a.
4. Escolha o tipo de instância ou adicione volumes de armazenamento do Amazon EBS aos grupos de instâncias uniformes ou frotas de instâncias.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

Old console

Criar um cluster em uma zona local usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.

2. Selecione Criar cluster.
3. Escolha Go to advanced options (Ir para opções avançadas).
4. Em Software Configuration (Configuração de software), Release (Versão), escolha 5.28.0 ou posterior.
5. Em Configuração do hardware, em Sub-rede do EC2, selecione uma sub-rede do EC2 com um ID de zona local neste formato: subnet 123abc | us-west-2-lax-1a.
6. Adicione volumes de armazenamento do Amazon EBS aos grupos de instâncias uniformes ou a frotas de instâncias e escolha um tipo de instância.

CLI

Para criar um cluster em uma zona local com o AWS CLI

- Use o comando create-cluster, junto com o SubnetId para a zona local, conforme mostrado no exemplo a seguir. Substitua a sub-rede-22xxxx1234567 pela zona local e substitua outras opções conforme necessário. SubnetId Para ter mais informações, consulte https://docs.aws.amazon.com/cli/latest/reference/emr/create-cluster.html.

```
aws emr create-cluster \
--name "Local Zones cluster" \
--release-label emr-5.29.0 \
--applications Name=Spark \
--ec2-attributes KeyName=myKey,SubnetId=subnet-22XXXX1234567 \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Configurar o Docker

O Amazon EMR 6.x é compatível com o Hadoop 3, o que permite que o YARN lance contêineres NodeManager diretamente no cluster do Amazon EMR ou dentro de um contêiner Docker. Os contêineres do Docker fornecem ambientes de execução personalizados nos quais o código do aplicativo é executado. O ambiente de execução personalizado é isolado do ambiente de execução do YARN NodeManager e de outros aplicativos.

Os contêineres do Docker podem incluir bibliotecas especiais usadas pelo aplicativo e podem fornecer diferentes versões de ferramentas e bibliotecas nativas, como R e Python. É possível usar ferramentas familiares do Docker para definir bibliotecas e dependências de runtime para as aplicações.

Os clusters do Amazon EMR 6.x são configurados por padrão para permitir que aplicações do YARN, como o Spark, sejam executadas usando contêineres do Docker. Para personalizar a configuração do contêiner, edite as opções de suporte do Docker definidas nos arquivos `yarn-site.xml` e `container-executor.cfg` disponíveis no diretório `/etc/hadoop/conf`. Para obter detalhes sobre cada opção de configuração e como ela é usada, consulte Launching applications using Docker containers.

É possível optar por usar o Docker ao enviar um trabalho. Use as variáveis a seguir para especificar o runtime do Docker e a imagem do Docker.

- `YARN_CONTAINER_RUNTIME_TYPE=docker`
- `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`

Quando você usa contêineres do Docker para executar os aplicativos do YARN, o YARN faz download da imagem do Docker especificada ao enviar o trabalho. Para que o YARN resolva essa imagem do Docker, ela deve ser configurada com um registro do Docker. As opções de configuração de um registro do Docker dependem se você implanta o cluster usando uma sub-rede pública ou privada.

Registros do Docker

Um registro do Docker é um sistema de armazenamento e distribuição de imagens do Docker. Para o Amazon EMR, é recomendável usar o Amazon ECR, que é um registro de contêiner do Docker totalmente gerenciado que permite criar suas próprias imagens personalizadas e hospedá-las em uma arquitetura altamente disponível e escalável.

Considerações de implantação

Os registros do Docker exigem acesso à rede de cada host no cluster. Isso ocorre porque cada host faz download de imagens do registro do Docker quando o aplicativo do YARN está sendo executado no cluster. Esses requisitos de conectividade de rede podem limitar sua escolha de registro do Docker, dependendo se você implanta o cluster do Amazon EMR em uma sub-rede pública ou privada.

Public subnet (Sub-rede pública)

Quando os clusters do EMR são implantados em uma sub-rede pública, os nós que executam o YARN NodeManager podem acessar diretamente qualquer registro disponível na Internet.

Sub-rede privada

Quando os clusters do EMR são implantados em uma sub-rede privada, os nós que executam o YARN NodeManager não têm acesso direto à Internet. As imagens do Docker podem ser hospedadas no Amazon ECR e acessadas por meio de. AWS PrivateLink

Para obter mais informações sobre como usar AWS PrivateLink para permitir o acesso ao Amazon ECR em um cenário de sub-rede privada, consulte Configuração do AWS PrivateLink Amazon ECS e do Amazon ECR.

Configurar registros do Docker

Para usar registros do Docker com o Amazon EMR, é necessário configurar o Docker para confiar no registro específico que você deseja usar para resolver imagens do Docker. Os registros de confiança padrão são locais (privados) e centOS. Para usar outros repositórios públicos ou o Amazon ECR, é possível substituir as configurações `docker.trusted.registries` em `/etc/hadoop/conf/container-executor.cfg` usando a API de classificação do EMR com a chave de classificação `container-executor`.

O exemplo a seguir mostra como configurar o cluster para confiar em um repositório público, chamado `your-public-repo`, e um endpoint de registro do ECR, `123456789123.dkr.ecr.us-east-1.amazonaws.com`. Se você usar o ECR, substitua esse endpoint pelo seu endpoint do ECR específico.

```
[
  {
    "Classification": "container-executor",
    "Configurations": [
        {
            "Classification": "docker",
            "Properties": {
                "docker.trusted.registries": "local,centos,your-public-
repo,123456789123.dkr.ecr.us-east-1.amazonaws.com",
                "docker.privileged-containers.registries": "local,centos,your-public-
repo,123456789123.dkr.ecr.us-east-1.amazonaws.com"
            }
        }
    ]
  }
]
```

Para iniciar um cluster do Amazon EMR 6.0.0 com essa configuração usando o AWS Command Line Interface (AWS CLI), crie um arquivo nomeado `container-executor.json` com o conteúdo da

configuração JSON anterior do executor de contêineres. Depois, use os comandos a seguir para executar o cluster.

```
export KEYPAIR=<Name of your Amazon EC2 key-pair>
export SUBNET_ID=<ID of the subnet to which to deploy the cluster>
export INSTANCE_TYPE=<Name of the instance type to use>
export REGION=<Region to which to deploy the cluster>

aws emr create-cluster \
    --name "EMR-6.0.0" \
    --region $REGION \
    --release-label emr-6.0.0 \
    --applications Name=Hadoop Name=Spark \
    --service-role EMR_DefaultRole \
    --ec2-attributes KeyName=$KEYPAIR,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=
$SUBNET_ID \
    --instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=
$INSTANCE_TYPE InstanceGroupType=CORE,InstanceCount=2,InstanceType=$INSTANCE_TYPE \
    --configuration file://container-executor.json
```

Configurar o YARN para acessar o Amazon ECR no EMR 6.0.0 e versões anteriores

Se você está começando a usar o Amazon ECR agora, siga as instruções em Como começar a usar o Amazon ECR e verifique se você tem acesso ao Amazon ECR de cada instância no cluster do Amazon EMR.

No EMR 6.0.0 e versões anteriores, para acessar o Amazon ECR usando o comando do Docker, é necessário primeiro gerar credenciais. Para verificar se o YARN pode acessar imagens do Amazon ECR, use a variável de ambiente do contêiner `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` para transmitir uma referência às credenciais geradas.

Execute o comando a seguir em um dos nós core para obter a linha de login da conta do ECR.

```
aws ecr get-login --region us-east-1 --no-include-email
```

O comando `get-login` gera o comando correto da CLI do Docker que deve ser executado para criar credenciais. Copie e execute a saída de `get-login`.

```
sudo docker login -u AWS -p <password> https://<account-id>.dkr.ecr.us-
east-1.amazonaws.com
```

Esse comando gera um arquivo `config.json` na pasta `/root/.docker`. Copie esse arquivo para o HDFS para que os trabalhos enviados ao cluster possam usá-lo a fim de fazer a autenticação no Amazon ECR.

Execute os comandos a seguir para copiar o arquivo `config.json` no diretório inicial.

```
mkdir -p ~/.docker
sudo cp /root/.docker/config.json ~/.docker/config.json
sudo chmod 644 ~/.docker/config.json
```

Execute os comandos a seguir para colocar o config.json no HDFS para que ele possa ser usado por trabalhos em execução no cluster.

```
hadoop fs -put ~/.docker/config.json /user/hadoop/
```

O YARN pode acessar o ECR como um registro de imagem do Docker e extrair contêineres durante a execução do trabalho.

Depois de configurar os registros do Docker e o YARN, é possível executar aplicativos do YARN usando contêineres do Docker. Para obter mais informações, consulte Run Spark applications with Docker using Amazon EMR 6.0.0.

No EMR 6.1.0 e versões posteriores, não é necessário configurar a autenticação no Amazon ECR manualmente. Se um registro do Amazon ECR for detectado na chave de classificação `container-executor`, o atributo de autenticação automática do Amazon ECR será ativado, e o YARN gerenciará o processo de autenticação quando você enviar um trabalho do Spark com uma imagem do ECR. Você pode confirmar se a autenticação automática está habilitada verificando `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` em yarn-site. A autenticação automática está habilitada e a configuração de autenticação do YARN está definida como `true` se `docker.trusted.registries` contém um URL de registro do ECR.

Pré-requisitos para usar a autenticação automática no Amazon ECR

- EMR versão 6.1.0 ou posterior
- O registro ECR incluído na configuração está na mesma região do cluster
- Perfil do IAM com permissões para obter o token de autorização e extrair qualquer imagem

Para obter mais informações, consulte Configuração com o Amazon ECR.

Como habilitar a autenticação automática

Siga Configurar registros do Docker para definir um registro do Amazon ECR como registro confiável e garantir que o repositório do Amazon ECR e o cluster estejam na mesma região.

Para habilitar esse atributo mesmo quando o registro ECR não estiver definido no registro confiável, use a classificação de configuração para definir `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` como `true`.

Como desabilitar a autenticação automática

Por padrão, a autenticação automática é desabilitada se nenhum registro do Amazon ECR for detectado no registro confiável.

Para desabilitar a autenticação automática, mesmo quando o registro do Amazon ECR estiver definido no registro confiável, use a classificação de configuração para definir `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` como `false`.

Como verificar se a autenticação automática está habilitada em um cluster

No nó principal, use um editor de texto, como `vi`, para visualizar o conteúdo do arquivo de log: `vi /etc/hadoop/conf.empty/yarn-site.xml`. Verifique o valor de `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled`.

Controle de término do cluster

Esta seção descreve suas opções para desativar clusters do Amazon EMR. Ele abrange o término automático e a proteção contra término e como interagem com outros atributos do Amazon EMR.

Você pode desligar um cluster do Amazon EMR das seguintes formas:

- Término após a execução da última etapa: crie um cluster transitório que será terminado após a conclusão de todas as etapas.
- Término automático (após tempo ocioso): crie um cluster com uma política de término automático que é desativado após um tempo ocioso especificado. Para ter mais informações, consulte Usar uma política de término automático.

- Término manual: crie um cluster de execução prolongada que continue em execução até você terminá-lo deliberadamente. Para obter mais informações sobre como encerrar um cluster manualmente, consulte Terminar um cluster.

Também é possível definir a proteção contra término em um cluster para evitar a desativação de instâncias do EC2 por acidente ou por erro.

Quando o Amazon EMR desliga seu cluster, todas as instâncias do Amazon EC2 no cluster são desativadas. Os dados no armazenamento de instância e nos volumes do EBS não estão mais disponíveis e não é possível recuperá-los. Entender e gerenciar o término do cluster é essencial para desenvolver uma estratégia para gerenciar e preservar dados gravando no Amazon S3 e equilibrando o custo.

Tópicos

Configurar um cluster para continuar ou terminar após a execução da etapa

Este tópico explica as diferenças entre usar um cluster de execução prolongada e criar um cluster transitório que é desativado após a execução da última etapa. Também aborda como configurar a execução de etapas em um cluster.

Criar um cluster de execução prolongada

Por padrão, os clusters que você cria com o console ou com o AWS CLI são de longa duração. Os clusters de execução prolongada continuam funcionando, aceitando trabalho e acumulando cobranças até você tomar medidas para desativá-los.

Um cluster de execução prolongada tem efeito nas seguintes situações:

- Quando você precisa consultar dados de forma interativa ou automática.
- Quando você precisa interagir continuamente com aplicações de big data hospedadas no cluster.
- Quando você processa periodicamente um conjunto de dados tão grande ou com tanta frequência que é ineficiente iniciar novos clusters e carregar dados todas as vezes.

Também é possível definir a proteção contra término em um cluster de execução prolongada para evitar a desativação de instâncias do EC2 por acidente ou por erro. Para ter mais informações, consulte Usar a proteção contra término.

> **Note**
>
> O Amazon EMR habilita automaticamente a proteção contra término para todos os clusters com múltiplos nós primários e substitui as configurações de execução de etapas fornecidas na criação do cluster. É possível desabilitar a proteção contra término depois que o cluster é iniciado. Consulte Configurar a proteção contra término para clusters em execução. Para desligar um cluster com múltiplos nós primários, primeiro é necessário modificar os atributos do cluster para desabilitar a proteção contra término. Para obter instruções, consulte Terminar um cluster do Amazon EMR com múltiplos nós primários.

Configurar um cluster para terminar após a execução da etapa

Quando você configura o término após a execução da etapa, o cluster é iniciado, executa ações de bootstrap e executa as etapas especificadas. Assim que a última etapa for concluída, o Amazon EMR terminará as instâncias do Amazon EC2 do cluster. Os clusters que você executa com a API do Amazon EMR têm a execução em etapas habilitada por padrão.

O término após a execução da etapa é eficaz para clusters que realizam uma tarefa de processamento periódico, como uma execução diária do processamento de dados. A execução de etapas também ajuda a garantir que você pague somente pelo tempo necessário para processar seus dados. Para mais informações sobre as etapas, consulte Enviar trabalhos a um cluster.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

Console

Para ativar o encerramento após a execução da etapa com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Etapas, escolha Adicionar etapa. Na caixa de diálogo Adicionar etapa, insira os valores apropriados dos campos. As opções diferem dependendo do tipo de etapa. Para adicionar a etapa e sair da caixa de diálogo, escolha Adicionar etapa.
4. Em Término do cluster, marque a caixa de seleção Terminar cluster após a conclusão da última etapa.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para ativar a rescisão após a execução da etapa com o AWS CLI

- Especifique o parâmetro `--auto-terminate` quando usar o comando `create-cluster` para criar um cluster transitório.

  O exemplo a seguir demonstra com usar o parâmetro `--auto-terminate`. Você pode digitar o comando a seguir e substituir *myKey* pelo nome do seu par de chaves do EC2.

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 \
--applications Name=Hive Name=Pig --use-default-roles --ec2-attributes
 KeyName=myKey \
--steps Type=PIG,Name="Pig Program",ActionOnFailure=CONTINUE,\
Args=[-f,s3://mybucket/scripts/pigscript.pig,-p,\
INPUT=s3://mybucket/inputdata/,-p,OUTPUT=s3://mybucket/outputdata/,\
```

```
$INPUT=s3://mybucket/inputdata/,$OUTPUT=s3://mybucket/outputdata/]
--instance-type m5.xlarge --instance-count 3 --auto-terminate
```

API

Para desativar o encerramento após a execução da etapa com a API do Amazon EMR na inicialização do cluster

1. Ao usar a ação RunJobFluxo para criar um cluster, defina a propriedade KeepJobFlowAliveWhenNoSteps como`false`.
2. Para alterar sua configuração de encerramento após a execução da etapa com a API do Amazon EMR após o lançamento do cluster:

   Use SetKeepJobFlowAliveWhenNoSteps a ação.

Usar uma política de término automático

Uma política de término automático permite orquestrar a limpeza do cluster sem a necessidade de monitorar e terminar manualmente os clusters não utilizados. Ao adicionar uma política de término automático a um cluster, especifique a quantidade de tempo ocioso após o qual o cluster deverá ser desligado automaticamente.

Dependendo da versão, o Amazon EMR usa critérios diferentes para marcar um cluster como ocioso. A tabela a seguir descreve como o Amazon EMR determina a ociosidade do cluster.

Quando você usa...	O cluster é considerado ocioso quando...
Amazon EMR versões 5.34.0 e posteriores e 6.4.0 e posteriores	• Não há aplicações YARN ativas • A utilização do HDFS está abaixo de 10% • Não há conexões ativas de caderno do EMR ou do EMR Studio • Não há interfaces de usuário de aplicações no cluster em uso •

Quando você usa...	O cluster é considerado ocioso quando...
	Não há etapas pendentes
Amazon EMR versões 5.30.0 a 5.33.0 e 6.1.0 a 6.3.0	• Não há aplicações YARN ativas • O cluster não tem trabalhos do Spark ativos **Note** O Amazon EMR marca um cluster como ocioso e pode terminá-lo automaticamente mesmo se você tiver um kernel Python3 ativo. Isso ocorre porque a execução de um kernel do Python3 não envia um trabalho do Spark no cluster. Para usar o encerramento automático com um kernel do Python3, recomendamos usar a versão 6.4.0 ou as versões posteriores do Amazon EMR.

Note

O Amazon EMR versões 6.4.0 e posteriores oferecem suporte a um arquivo no cluster para detectar atividades no nó primário: `/emr/metricscollector/isbusy`. Ao usar um cluster para executar scripts de shell ou aplicações que não sejam do YARN, você pode tocar ou atualizar `isbusy` periodicamente para informar ao Amazon EMR que o cluster não está ocioso.

É possível anexar uma política de término automático ao criar um cluster ou adicionar uma política a um cluster atual. Para alterar ou desabilitar o término automático, é possível atualizar ou remover a política.

Considerações

Leve em consideração os atributos e as limitações a seguir antes de usar uma política de término automático:

- A seguir Regiões da AWS, a terminação automática do Amazon EMR está disponível com o Amazon EMR 6.14.0 e superior:
  - Ásia-Pacífico (Hyderabad) (ap-south-2)
  - Ásia-Pacífico (Jacarta) (ap-southeast-3)
  - Europa (Espanha) (eu-south-2)
- A seguir Regiões da AWS, a terminação automática do Amazon EMR está disponível com o Amazon EMR 5.30.0 e 6.1.0 e versões superiores:
  - Leste dos EUA (Norte da Virgínia) (us-east-1)
  - Leste dos EUA (Ohio) (us-east-2)
  - Oeste dos EUA (Oregon) (us-west-2)
  - Oeste dos EUA (Norte da Califórnia) (us-west-1)
  - África (Cidade do Cabo) (af-south-1)
  - Ásia-Pacífico (Hong Kong) (ap-east-1)
  - Ásia-Pacífico (Mumbai) (ap-south-1)
  - Ásia-Pacífico (Seul) (ap-northeast-2)
  - Ásia-Pacífico (Singapura) (ap-southeast-1)
  - Ásia-Pacífico (Sydney) (ap-southeast-2)
  - Ásia Pacific (Tóquio) (ap-northeast-1)
  - Canadá (Central) (ca-central-1)
  - América do Sul (São Paulo) (sa-east-1)
  - Europa (Frankfurt) (eu-central-1)
  - Europa (Irlanda) (eu-west-1)
  - Europa (Londres) (eu-west-2)
  - UE (Milão) (eu-south-1)
  - Europa (Paris) (eu-west-3)
  - UE (Estocolmo) (eu-north-1)
  - China (Pequim) (cn-north-1)

  - China (Ningxia) (cn-northwest-1)
  - AWS GovCloud (Leste dos EUA) (us-gov-east-1)
  - AWS GovCloud (Oeste dos EUA) (us-gov-west-1)
- O tempo limite ocioso é padronizado para 60 minutos (uma hora) quando não há um valor especificado. Você pode especificar um tempo limite ocioso mínimo de um minuto e um tempo limite ocioso máximo de sete dias.
- Com o Amazon EMR versões 6.4.0 e posteriores, o término automático é habilitado por padrão quando você cria um novo cluster usando o console do Amazon EMR.
- O Amazon EMR publica Amazon CloudWatch métricas de alta resolução quando você ativa o encerramento automático de um cluster. Use essas métricas para monitorar a atividade e a ociosidade do cluster. Para ter mais informações, consulte Métricas de capacidade de cluster.
- Não há suporte para término automático ao usar aplicações que não são baseadas em Yarn, como Presto, Trino ou HBase.
- Para usar o término automático, o processo coletor de métricas deve ser capaz de se conectar ao endpoint público da API para o término automático no API Gateway. Se você usar um nome DNS privado com Amazon Virtual Private Cloud, o encerramento automático não funcionará corretamente. Para garantir que o término automático funcione, é recomendável executar uma das seguintes ações:
  - Remova o endpoint da VPC de interface do API Gateway da Amazon VPC.
  - Siga as instruções em Por que ocorre um erro HTTP 403 Proibido ao conectar APIs do API Gateway de uma VPC? para desabilitar a configuração de nome DNS privado.
  - Em vez disso, inicie o cluster em sua sub-rede privada. Para obter mais informações, consulte o tópico em Sub-redes privadas.
- (Amazon EMR 5.30.0 e versões posteriores) Se você remover a regra de saída Permitir tudo padrão para 0.0.0.0/ para o grupo de segurança primário, deverá adicionar uma regra que permita a conectividade TCP de saída ao grupo de segurança para acesso ao serviço na porta 9443. O grupo de segurança para acesso ao serviço também deve permitir tráfego TCP de entrada na porta 9443 do grupo de segurança primário. Para obter mais informações sobre como configurar grupos de segurança, consulte Amazon EMR-managed security group for the primary instance (private subnets).

Permissões para usar o término automático

Antes de aplicar e gerenciar políticas de término automático para o Amazon EMR, é necessário anexar as permissões listadas no exemplo a seguir da política de permissões do IAM aos recursos do IAM que gerenciam o cluster do EMR.

```
{
  "Version": "2012-10-17",
  "Statement": {
      "Sid": "AllowAutoTerminationPolicyActions",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:PutAutoTerminationPolicy",
        "elasticmapreduce:GetAutoTerminationPolicy",
        "elasticmapreduce:RemoveAutoTerminationPolicy"
      ],
      "Resource": "<your-resources>"
    }
}
```

Anexar, atualizar ou remover uma política de término automático

Esta seção contém instruções que ajudam a anexar, atualizar ou remover uma política de término automático de um cluster do Amazon EMR. Antes de trabalhar com políticas de término automático, verifique se você tem as permissões do IAM necessárias. Consulte Permissões para usar o término automático.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Anexar uma política de término automático ao criar um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.

2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Término do cluster, selecione Terminar cluster após tempo ocioso.
4. Especifique o número de horas e minutos ociosos que podem decorrer antes que o cluster seja terminado automaticamente. O tempo ocioso padrão é de uma hora.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

Anexar, atualizar ou remover uma política de término automático de um cluster em execução usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Propriedades da página de detalhes do cluster, localize Término do cluster e selecione Editar.
4. Selecione ou desmarque Habilitar término automático para ativar ou desativar o atributo. Se você ativar o término automático, especifique o número de horas e minutos ociosos que podem decorrer antes que o cluster seja terminado automaticamente. Depois selecione Salvar alterações para confirmar.

Old console

Anexar uma política de término automático ao criar um cluster usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Em Configuração de hardware, selecione Término automático.
4. Especifique o número de horas e minutos ociosos após os quais o cluster deverá ser terminado automaticamente. O tempo ocioso padrão é de uma hora.
5. Escolha outras configurações conforme apropriado para seu aplicativo e, em seguida, escolha Create cluster (Criar cluster).

Anexar, atualizar ou remover uma política de término automático de um cluster em execução usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Clusters e escolha o cluster que deseja atualizar.
3. Escolha a guia Hardware na página de detalhes do cluster.
4. Selecione ou desmarque Habilitar término automático para ativar ou desativar o atributo. Se você ativar o término automático, especifique o número de horas e minutos ociosos após os quais o cluster deverá ser terminado automaticamente.

AWS CLI

Antes de começar

Antes de trabalhar com políticas de término automático, é recomendável atualizar para a versão mais recente da AWS CLI. Para obter instruções, consulte Installing, updating, and uninstalling the AWS CLI.

Anexar ou atualizar uma política de término automático usando a AWS CLI

- Use o comando `aws emr put-auto-termination-policy` para anexar ou atualizar uma política de término automático em um cluster.

  O exemplo a seguir especifica 3600 segundos para. *`IdleTimeout`* Se você não especificar *`IdleTimeout`*, o valor padrão será uma hora.

```
aws emr put-auto-termination-policy \
--cluster-id <your-cluster-id> \
--auto-termination-policy IdleTimeout=3600
```

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Também é possível especificar um valor para `--auto-termination-policy` ao usar o comando `aws emr create-cluster`. Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte a Referência de AWS CLI comandos.

Para remover uma política de encerramento automático com o AWS CLI

- Use o comando `aws emr remove-auto-termination-policy` para remover uma política de término automático de um cluster. Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte a Referência de AWS CLI comandos.

```
aws emr remove-auto-termination-policy --cluster-id <your-cluster-id>
```

Usar a proteção contra término

A proteção contra encerramento protege seus clusters contra o encerramento acidental, o que pode ser especialmente útil para clusters de longa execução que processam cargas de trabalho críticas. Quando a proteção contra encerramento está habilitada em um cluster de longa execução, você ainda poderá encerrar o cluster, mas deverá removê-la explicitamente do cluster primeiro. Isso ajuda a garantir que as instâncias do EC2 não sejam encerradas por acidente ou erro. Você pode habilitar a proteção contra encerramento ao criar um cluster e alterar a configuração em um cluster em execução.

Com a proteção contra término habilitada, a ação `TerminateJobFlows` na API do Amazon EMR não funciona. Os usuários não podem encerrar o cluster usando essa API nem o comando `terminate-clusters` da AWS CLI. A API retornará um erro, e a CLI será encerrada com um código de retorno diferente de zero. Quando você usar o console do Amazon EMR para encerrar um cluster, será solicitado que você execute uma etapa adicional para desativar a proteção contra término.

> **Warning**
>
> A proteção contra término não garante que os dados sejam retidos em caso de erro humano ou de solução alternativa. Por exemplo, se um comando de reinicialização for emitido pela linha de comando enquanto estiver conectado à instância usando SSH, se uma aplicação ou um script em execução na instância emitir um comando de reinicialização ou se a API do Amazon EC2 ou do Amazon EMR for usada para desabilitar a proteção contra término.

Isso também vale se você estiver executando as versões 7.1 e superiores do Amazon EMR e uma instância ficar insalubre e irrecuperável. Mesmo com a proteção contra término habilitada, os dados salvos no armazenamento da instância, inclusive dados do HDFS, poderão ser perdidos. Grave a saída de dados nos locais do Amazon S3 e crie estratégias de backup conforme a necessidade de seus requisitos de continuidade de negócios.

A proteção contra encerramento não afeta sua capacidade de dimensionar recursos de cluster usando qualquer uma das seguintes ações:

- Redimensionando um cluster manualmente com o AWS Management Console ou AWS CLI. Para ter mais informações, consulte Redimensionar manualmente um cluster em execução.
- Removendo instâncias de um grupo de instâncias core ou de tarefa usando uma política de redução com a escalabilidade automática. Para ter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.
- Removendo instâncias de uma frota de instâncias, reduzindo a capacidade de destino. Para ter mais informações, consulte Opções de frotas de instâncias.

Proteção contra término e o Amazon EC2

A configuração de proteção contra encerramento em um cluster do Amazon EMR corresponde ao `DisableApiTermination` atributo de todas as instâncias do Amazon EC2 no cluster. Por exemplo, se você habilitar a proteção contra rescisão em um cluster do EMR, o Amazon EMR definirá automaticamente como verdadeiro `DisableApiTermination` para todas as instâncias do EC2 dentro do cluster do EMR. O mesmo se aplica se você desativar a proteção contra rescisão. O Amazon EMR define automaticamente como false `DisableApiTermination` para todas as instâncias do EC2 dentro do cluster do EMR. Se você encerrar ou reduzir um cluster do Amazon EMR e as configurações do Amazon EC2 entrarem em conflito para uma instância do EC2, o Amazon EMR priorizará a configuração do Amazon EMR sobre as `DisableApiTermination` configurações `DisableApiStop` e no Amazon EC2 e continuará a encerrar a instância do EC2.

Por exemplo, você pode usar o console do Amazon EC2 para habilitar a proteção contra encerramento em uma instância do Amazon EC2 em um cluster do EMR com a proteção contra encerramento desativada. Se você encerrar ou reduzir o cluster com o console do Amazon EMR, o ou AWS CLI a API do Amazon EMR, o Amazon EMR `DisableApiTermination` substituirá a configuração, a definirá como falsa e encerrará a instância junto com outras instâncias.

Você também pode usar o console do Amazon EC2 para habilitar a proteção de parada em uma instância do Amazon EC2 em um cluster do EMR com a proteção de terminação desativada. Se você encerrar ou reduzir a escala do cluster, o Amazon EMR `DisableApiStop` define como false no Amazon EC2 e encerra a instância junto com outras instâncias.

O Amazon EMR substitui a `DisableApiStop` configuração somente quando você encerra ou reduz a escala de um cluster. Quando você ativa ou desativa a proteção contra rescisão em um cluster do EMR, o Amazon EMR não altera a `disableApiStop` configuração de nenhuma das instâncias do EC2 no respectivo cluster do EMR.

> **Important**
>
> Se você criar uma instância como parte de um cluster do Amazon EMR com proteção contra encerramento e usar a API ou os comandos do Amazon EC2 para modificar a instância de tal forma, e então a API AWS CLI ou `DisableApiTermination` os comandos do Amazon EC2 executarem a operação`TerminateInstances`, a instância do Amazon EC2 `false` será encerrada. AWS CLI

Proteção contra término e nós não íntegros do YARN

O Amazon EMR verifica periodicamente o status do Apache Hadoop YARN de nós em execução nas instâncias centrais e de tarefa do Amazon EC2 em um cluster. O estado de saúde é relatado pelo serviço NodeManager de verificação de saúde. Se um nó reportar`UNHEALTHY`, o controlador de instância do Amazon EMR adiciona o nó a uma lista de negação e não aloca contêineres do YARN a ela até que ele fique saudável novamente. Dependendo do status da proteção contra rescisão, da substituição de nós com problemas de integridade e da versão de lançamento do Amazon EMR, o Amazon EMR substituirá a instância não íntegra ou interromperá a alocação de controladores para a instância.

Proteção de rescisão e rescisão após a execução da etapa

Quando você ativa a rescisão após a execução da etapa e também ativa a proteção contra rescisão, o Amazon EMR ignora a proteção contra rescisão.

Ao enviar etapas para um cluster, você pode definir a propriedade `ActionOnFailure` para determinar o que acontecerá se não for possível executar a etapa devido a um erro. Os valores possíveis para essa configuração são `TERMINATE_CLUSTER` (`TERMINATE_JOB_FLOW` com versões

anteriores) `CANCEL_AND_WAIT` e `CONTINUE`. Para ter mais informações, consulte Enviar trabalhos a um cluster.

Se falhar uma etapa configurada com `ActionOnFailure` set to`CANCEL_AND_WAIT`, se a terminação após a execução da etapa for ativada, o cluster será encerrado sem executar as etapas subsequentes.

Se ocorrer uma falha em uma etapa configurada com `ActionOnFailure` definida como `TERMINATE_CLUSTER`, use a tabela de configurações abaixo para determinar o resultado.

ActionOnFalha	Rescisão após a execução da etapa	Termination protection	Resultado
`TERMINATE_CLUSTER`	Habilitado	Desabilitado	O cluster é encerrado
	Habilitado	Habilitado	O cluster é encerrado
	Desabilitado	Habilitado	O cluster continua
	Desabilitado	Desabilitado	O cluster é encerrado

Proteção contra término e instâncias spot

A proteção contra término do Amazon EMR não impede que uma instância spot do Amazon EC2 seja terminada quando o preço spot ultrapassa o preço spot máximo.

Configurar a proteção contra término ao iniciar um cluster

Você pode ativar ou desativar a proteção contra encerramento ao iniciar um cluster usando o console AWS CLI, o ou a API.

Para clusters de nó único, as configurações padrão de proteção contra encerramento são as seguintes:

- Lançamento de um cluster pelo console do Amazon EMR — a Proteção de encerramento está desativada por padrão.

- A inicialização de um cluster por meio da AWS CLI `aws emr create-cluster` —Termination Protection está desativada, a menos que `--termination-protected` seja especificada.
- Lançamento de um cluster pelo comando de RunJobfluxo de API do Amazon EMR — a proteção de terminação é desativada, a menos que o valor `TerminationProtected` booleano esteja definido como. `true`

Para clusters de alta disponibilidade, as configurações padrão de proteção contra encerramento são as seguintes:

- Lançamento de um cluster pelo console do Amazon EMR — A proteção de terminação é ativada por padrão.
- A inicialização de um cluster por meio da AWS CLI `aws emr create-cluster` —Termination Protection está desativada, a menos que `--termination-protected` seja especificada.
- Lançamento de um cluster pelo comando de RunJobfluxo de API do Amazon EMR — a proteção de terminação é desativada, a menos que o valor `TerminationProtected` booleano esteja definido como. `true`

Console

Para ativar ou desativar a proteção contra encerramento ao criar um cluster com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Versão do EMR, escolha emr-6.6.0 ou posterior.
4. Em Encerramento de cluster e substituição de nós, verifique se a opção Usar proteção de encerramento está pré-selecionada ou desmarque a seleção para desativá-la.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para ativar ou desativar a proteção contra encerramento ao criar um cluster usando o AWS CLI

- Com o AWS CLI, você pode iniciar um cluster com a proteção de encerramento ativada com o `create-cluster` comando com o `--termination-protected` parâmetro. Por padrão, a proteção contra encerramento é desativada.

  O exemplo a seguir cria um cluster com proteção contra encerramento habilitada:

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

  ```
  aws emr create-cluster --name "TerminationProtectedCluster" --release-
  label emr-7.1.0 \
  --applications Name=Hadoop Name=Hive Name=Pig \
  --use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge \
  --instance-count 3 --termination-protected
  ```

  Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Configurar a proteção contra término para clusters em execução

Você pode configurar a proteção contra término para um cluster em execução usando o console ou a AWS CLI.

Console

Para ativar ou desativar a proteção contra encerramento de um cluster em execução com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.

2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Propriedades na página de detalhes do cluster, localize Término do cluster e selecione Editar.
4. Marque ou desmarque a caixa de seleção Usar proteção contra término para ativar ou desativar o atributo. Depois selecione Salvar alterações para confirmar.

AWS CLI

Para ativar ou desativar a proteção contra encerramento de um cluster em execução usando o AWS CLI

- Para habilitar a proteção contra término em um cluster em execução usando a AWS CLI, digite o comando `modify-cluster-attributes` com o parâmetro `--termination-protected`. Para desabilitá-la, use o parâmetro `--no-termination-protected`.

  O exemplo a seguir habilita a proteção contra encerramento no cluster com o ID *j-3KVTXXXXXX7UG*:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --termination-protected
  ```

  O exemplo a seguir desabilita a proteção contra encerramento no mesmo cluster:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-protected
  ```

Substituindo nós não íntegros

O Amazon EMR usa periodicamente o serviço de verificação de NodeManager saúde no Apache Hadoop para monitorar o status dos nós principais em seu Amazon EMR em clusters do Amazon EC2. Se um nó não estiver funcionando de forma ideal, o verificador de saúde reporta esse nó ao controlador do Amazon EMR. O controlador do Amazon EMR adiciona o nó a uma lista de negação, impedindo que o nó receba novos aplicativos do YARN até que o status do nó melhore. Um motivo comum pelo qual um nó pode ficar insalubre é a utilização excessiva do disco. Para obter mais informações sobre a identificação de nós não íntegros e a recuperação, consulte Erros de recursos.

Você pode escolher se o Amazon EMR deve encerrar os nós não íntegros ou mantê-los no cluster. Se você desativar a substituição de nós não íntegros, os nós não íntegros permanecerão na lista de rejeição e continuarão a contar para a capacidade do cluster. Você ainda pode se conectar à sua instância principal do Amazon EC2 para configuração e recuperação, para que você possa redimensionar seu cluster para aumentar a capacidade. Observe que o Amazon EMR substituirá os nós não íntegros mesmo se a proteção de encerramento estiver ativada.

Se a substituição de nós não íntegros estiver ativada, o Amazon EMR encerrará o nó principal não íntegro e provisionará uma nova instância com base no número de instâncias no grupo de instâncias ou na capacidade alvo das frotas de instâncias. Se vários ou todos os nós principais ficarem insalubres por mais de 45 minutos, o Amazon EMR substituirá os nós normalmente.

> **Important**
>
> Para evitar a possibilidade de perda permanente de dados do HDFS, já que o Amazon EMR substitui normalmente uma instância central não íntegra, recomendamos que você sempre faça backup de seus dados.

O Amazon EMR publica o CloudWatch Amazon Events para substituição de nós com problemas de integridade, para que você possa acompanhar o que está acontecendo com suas instâncias principais não íntegras. Para obter mais informações, consulte eventos de substituição de nós não íntegros.

Configurações padrão de substituição e proteção de terminação de nós

A substituição de nós não íntegros está disponível para todas as versões do Amazon EMR, mas as configurações padrão dependem da etiqueta de lançamento que você escolher. Você pode alterar qualquer uma dessas configurações configurando a substituição de nós não íntegra ao criar um novo cluster ou acessando a configuração do cluster a qualquer momento.

Se você estiver criando um cluster de nó único ou de alta disponibilidade que esteja executando o Amazon EMR versão 7.0 ou inferior, a configuração padrão de substituição de nó não íntegra depende da proteção contra encerramento:

- Ativar a proteção de terminação desativa a substituição não íntegra do nó.
- A desativação da proteção de terminação permite a substituição de nós não íntegra.

Configurando a substituição de nós não íntegra ao iniciar um cluster

Você pode ativar ou desativar a substituição não íntegra de nós ao iniciar um cluster usando o console AWS CLI, o ou a API.

A configuração padrão de substituição de nós não íntegros depende de como você executa o cluster:

- Console do Amazon EMR — a substituição de nós não íntegros é ativada por padrão.
- AWS CLI `aws emr create-cluster`— a substituição de nós não íntegros é ativada por padrão, a menos que você especifique`--no-unhealthy-node-replacement`.
- Comando da RunJobFlow API do Amazon EMR — a substituição de nós não íntegros é ativada por padrão, a menos que você defina o valor `UnhealthyNodeReplacement` booleano como ou. `True` `False`

Console

Para ativar ou desativar a substituição de nós não íntegros ao criar um cluster com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Para a versão de lançamento do EMR, escolha a etiqueta de lançamento do Amazon EMR que você deseja.
4. Em Encerramento do cluster e substituição de nós, verifique se a substituição de nó não íntegra (recomendada) está pré-selecionada ou desmarque a seleção para desativá-la.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para ativar ou desativar a substituição de nós não íntegros ao criar um cluster usando o AWS CLI

- Com o AWS CLI, você pode iniciar um cluster com a substituição de nós não íntegros ativada com o `create-cluster` comando com o `--unhealthy-node-replacement` parâmetro. A substituição de nós não íntegros está ativada por padrão.

  O exemplo a seguir cria um cluster com a substituição de nós não íntegros ativada:

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "SampleCluster" --release-label emr-7.1.0 \
--applications Name=Hadoop Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3 --unhealthy-node-replacement
```

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte Comandos do Amazon EMR. AWS CLI

Configurando a substituição de nós não íntegra em um cluster em execução

Você pode ativar ou desativar a substituição de nós não íntegros em um cluster em execução usando o console AWS CLI, o ou a API.

Console

Para ativar ou desativar a substituição de nós não íntegros em um cluster em execução com o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Propriedades da página de detalhes do cluster, encontre Terminação do cluster e substituição do nó e selecione Editar.
4. Marque ou desmarque a caixa de seleção substituição de nó não íntegra para ativar ou desativar o recurso. Depois selecione Salvar alterações para confirmar.

AWS CLI

Para ativar ou desativar a substituição de nós não íntegros em um cluster em execução usando o AWS CLI

- Para ativar a substituição de nós não íntegros em um cluster em execução com o AWS CLI, use o `modify-cluster-attributes` comando com o `--unhealthy-node-replacement` parâmetro. Para desabilitá-la, use o parâmetro `--no-unhealthy-node-replacement`.

  O exemplo a seguir ativa a substituição de nós não íntegros no cluster com o ID *`j-3KVTXXXXXX7UG`*:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --unhealthy-node-replacement
  ```

  O exemplo a seguir desativa a substituição de nós não íntegros no mesmo cluster:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-unhealthy-node-replacement
  ```

Trabalhar com AMIs do Amazon Linux no Amazon EMR

Imagens de máquina da Amazon (AMIs) do Amazon Linux

O Amazon EMR usa uma imagem de máquina da Amazon (AMI) do Amazon Linux para inicializar instâncias do Amazon EC2 quando você cria e inicia um cluster. A AMI contém o sistema operacional Amazon Linux, outros softwares e as configurações necessárias para cada instância hospedar suas aplicações de cluster.

Por padrão, quando você cria um cluster, o Amazon EMR usa uma AMI do Amazon Linux padrão que é criada especificamente para a versão do Amazon EMR que você usa. Para obter mais informações sobre a AMI do Amazon Linux padrão, consulte Usar a AMI padrão do Amazon Linux para Amazon EMR. Ao usar o Amazon EMR 5.7.0 ou posterior, você pode optar por especificar uma AMI personalizada do Amazon Linux em vez da AMI padrão do Amazon Linux para o Amazon EMR. Uma AMI personalizada permite criptografar o volume do dispositivo raiz e personalizar os aplicativos e configurações como uma alternativa ao uso de ações de bootstrap. É possível especificar uma AMI personalizada para cada tipo de instância na configuração do grupo de instâncias ou da frota

de instâncias de um cluster do Amazon EMR. O suporte a múltiplas AMIs personalizadas oferece a flexibilidade de usar mais de um tipo de arquitetura em um cluster. Consulte Usar uma AMI personalizada.

O Amazon EMR anexa automaticamente um volume SSD de uso geral do Amazon EBS como dispositivo raiz para todas as AMIs. AMIs com suporte do EBS melhoram a performance. Para obter mais informações sobre AMIs do Amazon Linux, consulte Imagens de máquina da Amazon (AMIs). Para obter mais informações sobre armazenamento de instância do Amazon EMR, consulte Armazenamento de instâncias.

Usar a AMI padrão do Amazon Linux para Amazon EMR

Cada versão do Amazon EMR usa uma AMI padrão do Amazon Linux para o Amazon EMR, a menos que você especifique uma AMI personalizada. Começando com as versões do Amazon EMR 5.36, Amazon EMR 6.6 e Amazon EMR 7.0, o comportamento padrão para atualizar o Amazon Linux 2 (AL2 para EMR 5.x e 6.x, AL2023 para EMR 7.x) em uma AMI padrão do Amazon EMR é aplicar automaticamente a versão mais recente do Amazon Linux para a AMI padrão do Amazon EMR.

Atualizações automáticas do Amazon Linux para versões do Amazon EMR

Quando você executa um cluster com a versão de patch mais recente do Amazon EMR 7.0 ou superior, 6.6 ou superior ou 5.36 ou superior, o Amazon EMR usa a versão mais recente do Amazon Linux para a AMI padrão do Amazon EMR. Por exemplo: .

- Onde há uma versão `x.x.0` e `x.x.1`, a versão `x.x.0` deixa de receber atualizações da AMI quando `x.x.1` é iniciada.
- Da mesma forma, a `x.x.1` para de receber atualizações da AMI quando `x.x.2` é iniciada.
- Posteriormente, quando a versão `x.y.0` é lançada, `x.x.`*`[latest]` continua recebendo atualizações da AMI junto com `x.y.`*`[latest]`.

Para ver se você está usando a versão de patch mais recente, conforme indicado pelo número após o segundo ponto decimal (`6.8.`*`1`*) de uma versão do Amazon EMR, consulte as versões disponíveis no Guia de lançamento do Amazon EMR, verifique o menu suspenso Versão do Amazon EMR ao criar um cluster no console ou use a ação a API `ListReleaseLabels` ou ação da CLI `list-release-labels`. Para receber atualizações quando lançarmos uma nova versão do Amazon EMR, assine o feed RSS da página What's new? no Guia de lançamento.

Se quiser, você pode optar por iniciar o cluster com a versão do Amazon Linux com a qual a versão do Amazon EMR foi enviada pela primeira vez. Para obter informações sobre como especificar a versão do Amazon Linux para o cluster, consulte Alteração da versão do Amazon Linux ao criar um cluster do EMR.

Versões padrão do Amazon Linux

Tópicos

AMIs padrão para o Amazon EMR 7.0 e superior

A tabela a seguir lista as informações do Amazon Linux para a versão de patch mais recente do Amazon EMR, versões 7.0 e superiores.

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240304.(	6.1.79-99.164.amzn2023	12 de março de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240219.(	6.1.77-99.164.amzn2023	1º de março de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240205.(	6.1.75-99.163.amzn2023	19 de fevereiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240122.(	6.1.72-96.166.amzn2023	5 de fevereiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 240108.(	6.1.72-96.166.amzn2023	24 de janeiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • ca-west-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2023.3.2 231211.4	6.1.66-91.160.amzn2023	19 de dezembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-southeast-4 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• ca-central-1 • il-central-1 • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

AMIs padrão para as versões 6.6 e superiores do Amazon EMR

A tabela a seguir lista as informações do Amazon Linux para a versão de patch mais recente do Amazon EMR 6.6.x e versões superiores.

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2024 223.0	4.14.336	8 de março de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-centra l-2 (6.10.1+) • eu-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southeast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20240131.0	4.14.336	14 de fevereiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10.1+) • eu-south-1 • eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southeast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-centra l-2 (6.10.1+) • eu-south-1 • eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southe ast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20240109.0	4.14.334	24 de janeiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10.1+) • eu-south-1 • eu-south-2 (6.10.1+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10.1+) • ap-southeast-3 • ap-southeast-4 (6.8.1+ e 5.36,1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10.1+) • me-south-1 • ca-central-1 • il-central-1 (6.8.1+ e 5.36,1) • ca-west-1 (6.9.1+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231218.0	4.14.330	2 de janeiro de 2024	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231206.0	4.14.330	22 de dezembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231116.0	4.14.328	11 de dezembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231101.0	4.14.327	17 de novembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231020.1	4.14.326	7 de novembro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231012.1	4.14.326	26 de outubro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230926.0	4.14.322	19 de outubro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.8+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20238906.0	4.14.322	4 de outubro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `me-central-1` (6.10+) • `me-south-1` • `ca-central-1` • `il-central-1` (6.9+ e 5.36.1)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230822.0	4.14.322	30 de agosto de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `me-central-1` (6.10+) • `me-south-1` • `ca-central-1` • `il-central-1` (6.9+ e 5.36.1)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230808.0	4.14.320	24 de agosto de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-central-1 (6.10+) • me-south-1 • ca-central-1 • il-central-1 (6.9+ e 5.36.1) • us-gov-east-1 • us-gov-west-1 • cn-north-1 • cn-northeast-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230727.0	4.14.320	14 de agosto de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `me-central-1` (6.10+) • `me-south-1` • `ca-central-1` • `il-central-1` (6.9+ e 5.36.1)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230719.0	4.14.320	2 de agosto de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-southeast-4 (6.8+ e 5.36.1) • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `me-central-1` (6.10+) • `me-south-1` • `ca-central-1` • `il-central-1` (6.9+ e 5.36.1)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230628.0	4.14.318	12 de julho de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 (6.10+)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230612.0	4.14.314	23 de junho de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 (6.10+)

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230504.1	4.14.313	16 de maio de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10+) • eu-south-1 • eu-south-2 (6.10+) • ap-east-1 • ap-south-1 • ap-south-2 (6.10+) • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• `ca-central-1`

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230418.0	4.14.311	3 de maio de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 (6.10 somente) • eu-south-1 • eu-south-2 (6.10 somente) • ap-east-1 • ap-south-1 • ap-south-2 (6.10 somente) • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 404.1	4.14.311	18 de abril de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 404.0	4.14.311	10 de abril de 2023	• `us-east-1` • `eu-west-3`

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230320.0	4.14.309	30 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230307.0	4.14.305	15 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230207.0	4.14.304	3 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-central-2 • eu-south-1 • eu-south-2 • ap-east-1 • ap-south-1 • ap-south-2 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-central-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20231119.1	4.14.301	9 de fevereiro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20221210.1	4.14.301	12 de janeiro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20221103.3	4.14.296	5 de dezembro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220004.0	4.14.294	2 de novembro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220912.1	4.14.291	7 de outubro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1
2.0.20220805.0	4.14.287	30 de agosto de 2022	• us-west-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220719.0	4.14.287	10 de agosto de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220426.0	4.14.281	10 de junho de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220406.1	4.14.275	2 de maio de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

AMIs padrão para o Amazon EMR 5.x

A tabela a seguir lista as informações do Amazon Linux para a versão de patch mais recente do Amazon EMR 5.x, versões 5.36 e posteriores.

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230504.1	4.14.313	16 de maio de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • me-central-1
2.0.20230418.0	4.14.311	3 de maio de 2023	• us-east-1 • us-east-2

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
			• us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • me-central-1

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 404.1	4.14.311	18 de abril de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1
2.0.2023 404.0	4.14.311	10 de abril de 2023	• us-east-1 • eu-west-3

OsRelea Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.2023 320.0	4.14.309	30 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230307.0	4.14.305	15 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • ca-central-1 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20230207.0	4.14.304	3 de março de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20221210.1	4.14.301	12 de janeiro de 2023	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20221103.3	4.14.296	5 de dezembro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220004.0	4.14.294	2 de novembro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220912.1	4.14.291	7 de outubro de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220719.0	4.14.287	10 de agosto de 2022	• us-west-1 • eu-west-3 • eu-north-1 • eu-central-1 • ap-south-1 • me-south-1

OsRelease Label (Versão AL)	Versão do kernel do AL	Data disponível	Regiões da AWS
2.0.20220426.0	4.14.281	14 de junho de 2022	• us-east-1 • us-east-2 • us-west-1 • us-west-2 • eu-north-1 • eu-west-1 • eu-west-2 • eu-west-3 • eu-central-1 • eu-south-1 • ap-east-1 • ap-south-1 • ap-southeast-3 • ap-northeast-1 • ap-northeast-2 • ap-northeast-3 • ap-southeast-1 • ap-southeast-2 • af-south-1 • sa-east-1 • me-south-1 • ca-central-1

Considerações sobre a atualização de software

Observe os comportamentos padrão de atualização de software a seguir.

Amazon EMR 7.x com Amazon Linux 2023

As versões 7.0 e superiores do Amazon EMR são executadas no Amazon Linux 2023 (AL2023). O comportamento padrão do AL2023 é bloquear AMIs em uma versão específica do repositório de software do Amazon Linux. Portanto, as atualizações de segurança não são aplicadas toda vez que você executa um cluster. Em vez disso, o comportamento padrão das versões 7.x do Amazon EMR é aplicar automaticamente a versão mais recente do AL2023 à AMI padrão do Amazon EMR somente ao criar o cluster. Para receber as atualizações de segurança mais recentes, recomendamos que você recrie seu cluster periodicamente.

Amazon EMR 5.x e 6.x com Amazon Linux e Amazon Linux 2

Em versões do Amazon EMR inferiores à 7.0, quando uma instância do Amazon EC2 é inicializada pela primeira vez em um cluster baseado na AMI padrão do Amazon Linux (AL) ou do Amazon Linux 2 (AL2) para Amazon EMR, ela verifica se há atualizações de software que se aplicam à versão nos repositórios de pacotes habilitados para o AL e o Amazon EMR. Assim como ocorre com outras instâncias do AL e do AL2, as atualizações de segurança críticas e importantes desses repositórios são instaladas automaticamente.

Observe também que, na configuração da sua rede, você deve permitir a saída de HTTP e HTTPS para repositórios do Amazon Linux no Amazon S3. Caso contrário, as atualizações de segurança falharão. Para obter mais informações, consulte Amazon Linux - Package repository no Amazon EC2 User Guide. Por padrão, outros pacotes de software e atualizações do kernel que exigem uma reinicialização, incluindo NVIDIA e CUDA, são excluídos do download automático na primeira inicialização.

Amazon EMR 5.35.0 e inferiores e 6.5.0 e inferiores: AMI do Amazon Linux bloqueada para a versão do Amazon EMR

Para o Amazon EMR 5.35.0 e inferior e 6.5.0 e inferior, a AMI padrão é baseada na maior parte da Amazon up-to-date Linux AMI disponível no momento do lançamento do Amazon EMR. A AMI é testada quanto à compatibilidade com as aplicações de big data e os atributos do Amazon EMR incluídos na versão.

Toda versão do Amazon EMR 5.35.0 e anteriores e do Amazon EMR 6.5.0 e anteriores está “bloqueada” na respectiva versão atribuída da AMI do Amazon Linux para manter a compatibilidade. Por esse motivo, é recomendável usar a versão mais recente do Amazon EMR, a menos que você precise de uma versão anterior para compatibilidade e não consiga fazer a migração. Se você precisar usar uma versão anterior do Amazon EMR para fins de compatibilidade, recomendamos

usar a versão mais recente de uma série. Por exemplo, se você precisar usar a série 5.12, use a 5.12.0 em vez da 5.12.2 5.12.1. Se uma nova versão se tornar disponível em uma série, considere a migração de seus aplicativos para a nova versão.

Para obter mais informações sobre o comportamento de atualização automática introduzido com o Amazon EMR 5.36.0 e versões posteriores e 6.6.0 e versões posteriores, consulte Atualizações automáticas do Amazon Linux para versões do Amazon EMR.

O comportamento de inicialização padrão exclui atualizações de kernel

Quando uma instância do Amazon EC2 em um cluster baseado na AMI padrão do Amazon Linux para o Amazon EMR é inicializada pela primeira vez, ela verifica os repositórios do pacote habilitado para Amazon Linux e Amazon EMR quanto a atualizações de software que se aplicam à versão da AMI. Assim como ocorre com outras instâncias do Amazon EC2, as atualizações de segurança críticas e importantes desses repositórios são instaladas automaticamente.

No entanto, se você estiver usando uma versão mais antiga da AMI do Amazon Linux, a atualização de segurança mais recente poderá não ser instalada automaticamente. Isso ocorre porque os repositórios referenciados pelo cluster do EMR são corrigidos para cada versão da AMI do Amazon Linux.

Observe também que, na configuração da sua rede, você deve permitir a saída de HTTP e HTTPS para repositórios do Amazon Linux no Amazon S3. Caso contrário, as atualizações de segurança falharão. Para obter mais informações, consulte Amazon Linux - Package repository no Amazon EC2 User Guide. Por padrão, outros pacotes de software e atualizações do kernel que exigem uma reinicialização, incluindo NVIDIA e CUDA, são excluídos do download automático na primeira inicialização.

> **Important**
>
> Os clusters do EMR que executam o AL2023 usam o comportamento padrão do Amazon Linux, e as imagens de máquina da Amazon (AMIs) estão bloqueadas em uma versão específica do repositório do Amazon Linux. Por padrão, seus clusters não receberão automaticamente as atualizações de segurança do software na execução. Seus clusters contêm apenas as atualizações que estavam disponíveis na versão da AMI do AL2023 que você escolheu ao criar o cluster. Para obter mais informações, consulte Atualização do Amazon Linux 2023 no Guia do usuário do Amazon Linux 2023.

> **Important**
>
> Os clusters do EMR que executam imagens de máquina da Amazon (AMIs) do Amazon Linux ou do Amazon Linux 2 usam o comportamento padrão do Amazon Linux e não baixam nem instalam automaticamente atualizações importantes e críticas do kernel que exigem reinicialização. É o mesmo comportamento de outras instâncias do Amazon EC2 que executam a AMI padrão do Amazon Linux. Se novas atualizações de software do Amazon Linux que exigem reinicialização (como atualizações do kernel, NVIDIA e CUDA) forem disponibilizadas após o lançamento de uma versão do Amazon EMR, as instâncias de cluster do Amazon EMR que executam a AMI padrão não baixarão nem instalarão essas atualizações automaticamente. Para obter atualizações do kernel, você pode personalizar sua AMI do Amazon EMR para usar a AMI do Amazon Linux mais recente.

O cluster é iniciado com ou sem atualizações

Lembre-se de que, se não foi possível instalar atualizações de software porque os repositórios de pacotes estão inacessíveis na primeira inicialização do cluster, a instância do cluster ainda concluirá sua execução. Por exemplo, os repositórios podem estar inacessíveis porque o S3 está temporariamente indisponível, ou pode haver regras da VPC ou do firewall configuradas para bloquear o acesso.

Não executar **`sudo yum update`**

Quando você se conectar a uma instância do cluster usando SSH, as primeiras linhas de saída da tela fornecerão um link para as notas de versão da AMI do Amazon Linux que a instância usa, um aviso da versão mais recente da AMI do Amazon Linux, um aviso do número de pacotes disponíveis para atualização dos repositórios habilitados e uma diretiva para executar `sudo yum update`.

> **Important**
>
> É altamente recomendável que você não execute `sudo yum update` em instâncias do cluster, enquanto estiver conectado usando SSH ou ao usar uma ação de bootstrap. Isso pode causar incompatibilidades porque todos os pacotes são instalados indiscriminadamente.

Práticas recomendadas de atualização de software

Práticas recomendadas de gerenciamento de atualizações de software

- Se você usar uma versão anterior do Amazon EMR, considere testar uma migração para a versão mais recente antes de atualizar pacotes de software.
- Se você migrar para uma versão posterior ou atualizar pacotes de software, teste primeiro a implementação em um ambiente que não seja de produção. A opção para clonar clusters usando o console do Amazon EMR é útil para isso.
- Avalie as atualizações de software para suas aplicações e para a versão da AMI do Amazon Linux individualmente. Somente teste e instale pacotes em ambientes de produção que você determinar que são totalmente necessários para sua postura de segurança, funcionalidade do aplicativo ou desempenho.
- Acompanhe o Amazon Linux Security Center para verificar se há atualizações.
- Evite a instalação de pacotes ao conectar-se a instâncias de cluster usando SSH. Em vez disso, use uma ação de bootstrap para instalar e atualizar pacotes em todas as instâncias de cluster conforme necessário. Isso requer que você encerre um cluster e reinicie-o. Para ter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais.

Usar uma AMI personalizada

Ao usar o Amazon EMR 5.7.0 ou posterior, você pode optar por especificar uma AMI personalizada do Amazon Linux em vez da AMI padrão do Amazon Linux para o Amazon EMR. Uma AMI personalizada é útil se você deseja fazer o seguinte:

- Pré-instale aplicativos e realize outras personalizações em vez de usar ações de bootstrap. Isso pode melhorar o tempo de início do cluster e agilizar o fluxo de trabalho de inicialização. Para obter mais informações e um exemplo, consulte Criar uma AMI do Amazon Linux personalizada com base em uma instância pré-configurada.
- Implementar configurações de cluster e o nó mais sofisticadas do que o permitido por ações de bootstrap.
- Criptografar os volumes do dispositivo raiz (volumes de inicialização) do EBS de instâncias do EC2 no cluster se você estiver usando uma versão do Amazon EMR anterior à 5.24.0. Assim como ocorre com a AMI padrão, o tamanho mínimo do volume raiz para uma AMI personalizada é de 10 GiB nas versões 6.9 e inferiores do Amazon EMR e de 15 GiB nas versões 6.10 e superiores

do Amazon EMR. Para ter mais informações, consulte Criar uma AMI personalizada com o volume do dispositivo raiz do Amazon EBS criptografado.

> **Note**
>
> A partir da versão 5.24.0 do Amazon EMR, você pode usar uma opção de configuração de segurança para criptografar o dispositivo raiz e os volumes de armazenamento do EBS ao especificar como seu provedor de chaves. AWS KMS Para ter mais informações, consulte Criptografia de disco local.

Uma AMI personalizada deve existir na mesma AWS região em que você cria o cluster. Também deve corresponder à arquitetura da instância do EC2. Por exemplo, uma instância m5.xlarge tem a arquitetura x86_64. Portanto, para provisionar uma instância m5.xlarge usando uma AMI personalizada, a AMI personalizada também deverá ter a arquitetura x86_64. Da mesma forma, para provisionar uma instância m6g.xlarge, que tem a arquitetura arm64, a AMI personalizada deverá ter a arquitetura arm64. Para obter mais informações sobre como identificar uma AMI Linux para seu tipo de instância, consulte Encontre uma AMI Linux no Guia do usuário do Amazon EC2.

> **Important**
>
> Os clusters do EMR que executam imagens de máquina da Amazon (AMIs) do Amazon Linux ou do Amazon Linux 2 usam o comportamento padrão do Amazon Linux e não baixam nem instalam automaticamente atualizações importantes e críticas do kernel que exigem reinicialização. É o mesmo comportamento de outras instâncias do Amazon EC2 que executam a AMI padrão do Amazon Linux. Se novas atualizações de software do Amazon Linux que exigem reinicialização (como atualizações do kernel, NVIDIA e CUDA) forem disponibilizadas após o lançamento de uma versão do Amazon EMR, as instâncias de cluster do Amazon EMR que executam a AMI padrão não baixarão nem instalarão essas atualizações automaticamente. Para obter atualizações do kernel, você pode personalizar sua AMI do Amazon EMR para usar a AMI do Amazon Linux mais recente.

Criar uma AMI do Amazon Linux personalizada com base em uma instância pré-configurada

As etapas básicas para pré-instalar softwares e realizar outras configurações para criar uma AMI personalizada do Amazon Linux para o Amazon EMR são as seguintes:

- Execute uma instância a partir da AMI base do Amazon Linux.
- Conecte-se à instância para instalar o software e realizar outras personalizações.
- Crie uma nova imagem (snapshot da AMI) da instância configurada.

Depois de criar a imagem com base na sua instância personalizada, você pode copiar essa imagem para um destino criptografado conforme descrito em Criar uma AMI personalizada com o volume do dispositivo raiz do Amazon EBS criptografado.

Tutorial: criar uma AMI de uma instância com softwares personalizados instalados

Para executar uma instância do EC2 baseada na AMI Amazon Linux mais recente

1. Use o AWS CLI para executar o comando a seguir, que cria uma instância a partir de uma AMI existente. *MyKeyName*Substitua pelo par de chaves que você usa para se conectar à instância e o *MyAmiID* pelo ID de um Amazon Linux AMI apropriado. Para os IDs de AMI mais recentes, consulte Amazon Linux AMI.

   > **Note**
   >
   > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

   ```
   aws ec2 run-instances --image-id MyAmiID \
   --count 1 --instance-type m5.xlarge \
   --key-name MyKeyName --region us-west-2
   ```

   O valor de saída `InstanceId` é usado como *MyInstanceId* na próxima etapa.

2. Execute o seguinte comando:

   ```
   aws ec2 describe-instances --instance-ids MyInstanceId
   ```

   O valor de saída `PublicDnsName` é usado para se conectar à instância na próxima etapa.

Para se conectar à instância e instalar o software

1. Use uma conexão SSH que permite executar comandos shell na sua instância Linux. Para obter mais informações, consulte Conectando-se à sua instância Linux usando SSH no Guia do usuário do Amazon EC2.
2. Realize todas as personalizações necessárias. Por exemplo: .

```
sudo yum install MySoftwarePackage
sudo pip install MySoftwarePackage
```

Para criar um snapshot da imagem personalizada

- Depois de personalizar a instância, use o comando `create-image` para criar uma AMI a partir da instância.

```
aws ec2 create-image --no-dry-run --instance-id MyInstanceId --name MyEmrCustomAmi
```

O valor de saída `imageID` é usado quando você executa o cluster ou cria um snapshot criptografados. Para obter mais informações, consulte Usar uma única AMI personalizada em um cluster do EMR e Criar uma AMI personalizada com o volume do dispositivo raiz do Amazon EBS criptografado.

Como usar uma AMI personalizada em um cluster do Amazon EMR

É possível usar uma AMI personalizada para provisionar um cluster do Amazon EMR de duas formas:

- Use uma única AMI personalizada para todas as instâncias do EC2 que estão no cluster.
- Use AMIs personalizadas diferentes para os diferentes tipos de instâncias do EC2 usadas no cluster.

É possível usar somente uma das duas opções ao provisionar um cluster do EMR e não é possível alterar depois que o cluster é iniciado.

Considerações sobre o uso de uma ou mais AMIs personalizadas em um cluster do Amazon EMR

Consideração	AMI personalizada única	Múltiplas AMIs personalizadas
Usar x86 e processadores Graviton2 com AMIs personalizadas no mesmo cluster	✗ Sem suporte	✓ Há suporte
A personalização da AMI varia conforme os tipos de instância	✗ Sem suporte	✓ Há suporte
Altere as AMIs personalizadas ao adicionar novos grupos ou frotas de instâncias de tarefa a um cluster em execução. Observação: não é possível alterar a AMI personalizada de grupos ou frotas de instâncias já existentes.	✗ Sem suporte	✓ Há suporte
Use o AWS console para iniciar um cluster	✓ Há suporte	✗ Sem suporte
Use AWS CloudFormation para iniciar um cluster	✓ Há suporte	✓ Há suporte

Usar uma única AMI personalizada em um cluster do EMR

Para especificar um ID de AMI personalizada ao criar um cluster, use uma destas opções:

- AWS Management Console
- AWS CLI
- SDK do Amazon EMR
- Fluxo da API do Amazon EMR RunJob
- AWS CloudFormation (veja a `CustomAmiID` propriedade em Cluster InstanceGroupConfig, Cluster InstanceTypeConfig InstanceGroupConfig, Resource ou Resource InstanceFleetConfig - InstanceType Config)

Amazon EMR console

Para especificar uma única AMI personalizada usando o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Nome e aplicações, localize Opções do sistema operacional. Escolha AMI personalizada e insira o ID da AMI no campo AMI personalizada.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Para especificar uma única AMI personalizada com o AWS CLI

- Use o parâmetro `--custom-ami-id` para especificar o ID da AMI ao executar o comando `aws emr create-cluster`.

  O exemplo a seguir especifica um cluster que usa uma AMI personalizada com um único volume de inicialização de 20 GiB. Para ter mais informações, consulte Personalização do volume raiz do dispositivo do Amazon EBS.

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "Cluster with My Custom AMI" \
--custom-ami-id MyAmiID --ebs-root-volume-size 20 \
--release-label emr-5.7.0 --use-default-roles \
--instance-count 2 --instance-type m5.xlarge
```

Usar múltiplas AMIs personalizadas em um cluster do Amazon EMR

Para criar um cluster usando múltiplas AMIs personalizadas, use uma destas opções:

- AWS CLI versão 1.20.21 ou superior
- AWS SDK
- RunJobFluxo do Amazon EMR na referência da API do Amazon EMR
- AWS CloudFormation (veja a `CustomAmiID` propriedade em Cluster InstanceGroupConfig, Cluster InstanceTypeConfig InstanceGroupConfig, Resource ou Resource InstanceFleetConfig - InstanceType Config)

Atualmente, o AWS Management Console não oferece suporte à criação de um cluster usando várias AMIs personalizadas.

Example - Use a AWS CLI para criar um cluster de grupos de instâncias usando várias AMIs personalizadas

Usando a AWS CLI versão 1.20.21 ou superior, você pode atribuir uma única AMI personalizada a todo o cluster ou pode atribuir várias AMIs personalizadas a cada nó da instância em seu cluster.

O exemplo a seguir mostra um cluster uniforme de grupos de instâncias criado com dois tipos de instância (m5.xlarge) usados em todos os tipos de nós (primário, central, de tarefa). Cada nó tem múltiplas AMIs personalizadas. O exemplo ilustra vários atributos das múltiplas configurações personalizadas de AMI:

- Nenhuma AMI personalizada foi atribuída no nível do cluster. Isso evita conflitos entre múltiplas AMIs personalizadas e uma única AMI personalizada, o que faria com que a inicialização do cluster falhasse.
- O cluster pode ter várias AMIs personalizadas em nós primários, centrais e individuais de tarefa. Isso permite personalizações individuais de AMI, como aplicações pré-instaladas, configurações sofisticadas de cluster e volumes de dispositivos raiz criptografados do Amazon EBS.
- O nó central do grupo de instâncias pode ter somente um tipo de instância e a AMI personalizada correspondente. Da mesma forma, o nó primário pode ter somente um tipo de instância e a AMI personalizada correspondente.
- O cluster pode ter múltiplos nós de tarefa.

```
aws emr create-cluster --instance-groups
```

```
InstanceGroupType=PRIMARY,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-234567
InstanceGroupType=TASK,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-345678
InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-456789
```

Example - Use a AWS CLI versão 1.20.21 ou superior para adicionar um nó de tarefa a um cluster de grupos de instâncias em execução com vários tipos de instância e várias AMIs personalizadas

Usando a AWS CLI versão 1.20.21 ou superior, você pode adicionar várias AMIs personalizadas a um grupo de instâncias que você adiciona a um cluster em execução. O argumento `CustomAmiId` pode ser usado com o comando `add-instance-groups`, conforme mostrado no exemplo a seguir. O mesmo ID de múltiplas AMIs personalizadas (ami-123456) é usado em mais de um nó.

```
aws emr create-cluster --instance-groups
InstanceGroupType=PRIMARY,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-234567

{
    "ClusterId": "j-123456",
    ...
}

aws emr add-instance-groups --cluster-id j-123456 --instance-groups
 InstanceGroupType=Task,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-345678
```

Example - Use a AWS CLI versão 1.20.21 ou superior para criar um cluster de frota de instâncias, várias AMIs personalizadas, vários tipos de instância, primária sob demanda, núcleo sob demanda, vários núcleos e nós de tarefas

```
aws emr create-cluster --instance-fleets
InstanceFleetType=PRIMARY,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 CustomAmiId=ami-123456}']
InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,C
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']
InstanceFleetType=TASK,TargetSpotCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,Custo
{InstanceType=m6g.xlarge, CustomAmiId=ami-567890}']
```

Example - Use a AWS CLI versão 1.20.21 ou superior para adicionar nós de tarefas a um cluster em execução com vários tipos de instância e várias AMIs personalizadas

```
aws emr create-cluster --instance-fleets
InstanceFleetType=PRIMARY,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 CustomAmiId=ami-123456}']
InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,C
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']

{
    "ClusterId": "j-123456",
    ...
}

aws emr add-instance-fleet --cluster-id j-123456 --instance-fleet
InstanceFleetType=TASK,TargetSpotCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,Custo
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']
```

Gerenciar atualizações do repositório de pacotes de AMI

Na primeira inicialização, por padrão, as AMIs do Amazon Linux conectam-se a repositórios de pacotes para instalar atualizações de segurança antes da inicialização de outros serviços. Dependendo dos seus requisitos, você tem a opção de desabilitar essas atualizações ao especificar uma AMI personalizada para o Amazon EMR. A opção para desabilitar esse recurso está disponível somente quando você usa uma AMI personalizada. Por padrão, as atualizações de kernel do Amazon Linux e de outros pacotes de software que exigem uma reinicialização não são atualizados. A configuração de rede deve permitir a saída de HTTP e HTTPS para repositórios do Amazon Linux no Amazon S3, senão as atualizações de segurança não terão êxito.

> **Warning**
>
> Recomendamos que você opte por atualizar todos os pacotes instalados na reinicialização, ao especificar uma AMI personalizada. Se os pacotes de atualização não forem atualizados, poderá haver riscos de segurança adicionais.

Com o AWS Management Console, você pode selecionar a opção de desativar as atualizações ao escolher a AMI personalizada.

Com o AWS CLI, você pode especificar `--repo-upgrade-on-boot NONE` junto com `--custom-ami-id` ao usar o create-cluster comando.

Com a API do Amazon EMR, você pode especificar `NONE` o RepoUpgradeOnBootparâmetro.

Criar uma AMI personalizada com o volume do dispositivo raiz do Amazon EBS criptografado

Para criptografar o volume do dispositivo raiz do Amazon EBS de uma AMI do Amazon LInux para o Amazon EMR, copie uma imagem de snapshot de uma AMI não criptografada em um destino criptografado. Para obter informações sobre a criação de volumes criptografados do EBS, consulte a criptografia do Amazon EBS no Guia do usuário do Amazon EC2. A AMI de origem para o snapshot pode ser a AMI base do Amazon Linux, ou é possível copiar um snapshot de uma AMI derivada da AMI base do Amazon Linux que você personalizou.

> **Note**
>
> A partir da versão 5.24.0 do Amazon EMR, você pode usar uma opção de configuração de segurança para criptografar o dispositivo raiz e os volumes de armazenamento do EBS ao especificar como seu provedor de chaves. AWS KMS Para ter mais informações, consulte Criptografia de disco local.

Você pode usar um provedor de chave externo ou uma chave AWS KMS para criptografar o volume raiz do EBS. O perfil de serviço usado pelo Amazon EMR (geralmente o `EMR_DefaultRole` padrão) deve ter permissão para criptografar e descriptografar o volume, pelo menos, para o Amazon EMR criar um cluster usando a AMI. Ao usar AWS KMS como provedor de chaves, isso significa que as seguintes ações devem ser permitidas:

- `kms:encrypt`
- `kms:decrypt`
- `kms:ReEncrypt*`
- `kms:CreateGrant`
- `kms:GenerateDataKeyWithoutPlaintext"`
- `kms:DescribeKey"`

A maneira mais simples de fazer isso é adicionar o perfil como um usuário de chave, conforme descrito no seguinte tutorial. O exemplo a seguir de declaração de política é fornecido caso você precise personalizar as políticas de função.

```
{
"Version": "2012-10-17",
"Statement": [
   {
   "Sid": "EmrDiskEncryptionPolicy",
   "Effect": "Allow",
   "Action": [
      "kms:Encrypt",
      "kms:Decrypt",
      "kms:ReEncrypt*",
      "kms:CreateGrant",
      "kms:GenerateDataKeyWithoutPlaintext",
      "kms:DescribeKey"
      ],
   "Resource": [
      "*"
      ]
   }
]
}
```

Tutorial: criar uma AMI personalizada com o volume do dispositivo raiz criptografado usando uma chave do KMS

A primeira etapa deste exemplo é localizar o ARN de uma chave do KMS ou criar uma nova. Para obter mais informações sobre como criar chaves, consulte Creating keys no Guia do desenvolvedor do AWS Key Management Service . O procedimento a seguir mostra como adicionar a função de serviço padrão, `EMR_DefaultRole`, como um usuário e chave à política de chave. Anote o valor do ARN (Nome de recurso da Amazon) para a chave ao criá-lo ou editá-lo. Você usará o ARN posteriormente, quando criar a AMI.

Adicionar o perfil de serviço do Amazon EC2 à lista de usuários de chave de criptografia usando o console

1. Faça login AWS Management Console e abra o console AWS Key Management Service (AWS KMS) em https://console.aws.amazon.com/kms.

2. Para alterar o Região da AWS, use o seletor de região no canto superior direito da página.
3. Escolha o alias da chave do KMS a ser usada.
4. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
5. Na caixa de diálogo Anexar, escolha o perfil de serviço do Amazon EMR. O nome da função padrão é `EMR_DefaultRole`.
6. Escolha Anexar.

Para criar uma AMI criptografada com o AWS CLI

- Use o `aws ec2 copy-image` comando do AWS CLI para criar uma AMI com um volume de dispositivo raiz do EBS criptografado e a chave que você modificou. Substitua o valor `--kms-key-id` especificado com o ARN completo da chave que você criou ou modificou anteriormente.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws ec2 copy-image --source-image-id MyAmiId \
--source-region us-west-2 --name MyEncryptedEMRAmi \
--encrypted --kms-key-id arn:aws:kms:us-west-2:12345678910:key/xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
```

A saída do comando fornece o ID da AMI que você criou, que você pode especificar quando ao criar um cluster. Para ter mais informações, consulte Usar uma única AMI personalizada em um cluster do EMR. Você também pode optar por personalizar essa AMI instalando softwares e realizando outras configurações. Para ter mais informações, consulte Criar uma AMI do Amazon Linux personalizada com base em uma instância pré-configurada.

Práticas recomendadas e considerações

Ao criar uma AMI personalizada para o Amazon EMR, considere o seguinte:

- A série 7.x do Amazon EMR é baseada no Amazon Linux 2023. Para essas versões do Amazon EMR, você precisa usar imagens baseadas no Amazon Linux 2023 para AMIs personalizadas. Para localizar uma AMI personalizada básica, consulte Localizar AMI do Linux.
- Para versões do Amazon EMR inferiores a 7.x, as AMIs do Amazon Linux 2023 não são suportadas.
- O Amazon EMR 5.30.0, e posteriores, e a série Amazon EMR 6.x são baseados no Amazon Linux 2. Para essas versões do Amazon EMR, é preciso usar imagens baseadas no Amazon Linux 2 para AMIs personalizadas. Para localizar uma AMI personalizada básica, consulte Localizar AMI do Linux.
- Em versões anteriores a 5.30.0 e 6.x do Amazon EMR, não há suporte para AMIs do Amazon Linux 2.
- É necessário usar uma AMI do Amazon Linux de 64 bits. Não há suporte para AMI de 32 bits.
- Não há suporte para AMIs do Amazon Linux com múltiplos volumes do Amazon EBS.
- Baseie sua personalização na AMI Amazon Linux mais recente com suporte do EBS. Para obter uma lista das Amazon Linux AMIs e os IDs das AMIs correspondentes, consulte AMI do Amazon Linux.
- Não copie um snapshot de uma instância existente do Amazon EMR para criar uma AMI personalizada. Isto provoca erros.
- Há suporte apenas para o tipo de virtualização de HVM e instâncias compatíveis com o Amazon EMR. Não se esqueça de selecionar a imagem HVM e um tipo de instância compatível com o Amazon EMR conforme você percorre o processo de personalização da AMI. Para conhecer os tipos de virtualização e instâncias compatíveis, consulte Tipos de instâncias compatíveis.
- Sua função de serviço deve ter permissão de execução na AMI e, portanto, a AMI deve ser pública, ou você deve ser o proprietário da AMI ou ter recebido direito de compartilhamento do respectivo proprietário.
- Criar usuários na AMI com o mesmo nome que aplicativos provoca erros (por exemplo, `hadoop`, `hdfs`, `yarn` ou `spark`).
- O conteúdo de `/tmp`, `/var` e `/emr` (se existirem na AMI) é movido para `/mnt/tmp`, `/mnt/var` e `/mnt/emr`, respectivamente, durante a inicialização. Os arquivos são preservados, mas, se houver uma grande quantidade de dados, a inicialização poderá demorar mais do que o esperado.
- Se você usar uma AMI do Amazon Linux personalizada com base em uma AMI do Amazon Linux com data de criação 11/8/2018, o servidor Oozie falhará ao iniciar. Se você usar o Oozie, crie uma AMI personalizada com base em um ID de AMI do Amazon Linux com uma data de criação diferente. Você pode usar o AWS CLI comando a seguir para retornar uma lista de IDs de

imagem para todas as AMIs HVM Amazon Linux com uma versão 2018.03, junto com a data de lançamento, para que você possa escolher uma Amazon Linux AMI apropriada como sua base. MyRegion Substitua pelo seu identificador de região, como us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`]|[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

- Nos casos em que você usa uma VPC com um nome de domínio e AmazonProvided DNS não padrão, você não deve usar a `rotate` opção na configuração de DNS do sistema operacional.

Para obter mais informações, consulte Criação de uma AMI Linux baseada no Amazon EBS no Guia do usuário do Amazon EC2.

Alteração da versão do Amazon Linux ao criar um cluster do EMR

Quando você executa um cluster usando o Amazon EMR 6.6.0 ou versões posteriores, ele usa automaticamente a versão mais recente do Amazon Linux 2 que foi validada para a AMI padrão do Amazon EMR. Você pode especificar uma versão diferente do Amazon Linux para o cluster usando o console do Amazon EMR ou a AWS CLI.

Amazon EMR console

Para alterar a versão do Amazon Linux ao criar um cluster usando o console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Versão do EMR, escolha emr-6.6.0 ou posterior.
4. Em Opções do sistema operacional, escolha Versão do Amazon Linux e marque a caixa de seleção Aplicar automaticamente as últimas atualizações do Amazon Linux.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

AWS CLI

Alterar a versão do Amazon Linux ao criar um cluster usando a AWS CLI

- Use o parâmetro `--os-release-label` para especificar a versão do Amazon Linux ao executar o comando `aws emr` create-cluster.

```
aws emr create-cluster --name "Cluster with Different Amazon Linux Release" \
--os-release-label 2.0.20210312.1 \
--release-label emr-6.6.0 --use-default-roles \
--instance-count 2 --instance-type m5.xlarge
```

Personalização do volume raiz do dispositivo do Amazon EBS

Padrões do volume raiz do EBS

Com o Amazon EMR 4.x e versões superiores, você pode especificar o tamanho do volume raiz ao criar um cluster. Com as versões 6.15.0 e superiores do Amazon EMR, você também pode especificar as IOPS e o throughput do volume raiz. Os atributos se aplicam somente ao volume raiz do dispositivo do Amazon EBS e a todas as instâncias no cluster. Os atributos não se aplicam a volumes de armazenamento, que você especifica separadamente para cada tipo de instância ao criar o cluster.

- O tamanho padrão do volume raiz é de 15 GiB nas versões 6.10.0 e superiores do Amazon EMR. O tamanho padrão do volume raiz das versões anteriores é de 10 GiB. Você pode ajustá-lo para até 100 GiB.
- O volume raiz padrão tem 3.000 IOPS. Você pode ajustá-las para até 16.000.
- O volume raiz padrão tem 125 MiB/s de throughput. Você pode ajustá-lo para até 1000 Mib/s.

> **Note**
>
> O tamanho e as IOPS do volume raiz não podem ter uma proporção maior do que 1 volume para 500 IOPS (1:500), enquanto as IOPS e o throughput do volume raiz não podem ter uma proporção maior do que 1 IOPS para 0,25 throughput (1:0,25).

Para obter mais informações sobre o Amazon EBS, consulte Volume do dispositivo raiz da instância do Amazon EC2.

Tipo de volume raiz do dispositivo com a AMI padrão

Ao usar a AMI padrão, o tipo de volume raiz do dispositivo é determinado pela versão do Amazon EMR usada.

- Com as versões 6.15.0 e superiores, o Amazon EMR anexa um armazenamento SSD de uso geral (gp3) como o tipo de volume raiz do dispositivo.
- Com as versões inferiores à 6.15.0, o Amazon EMR anexa um armazenamento SSD de uso geral (gp2) como o tipo de volume raiz do dispositivo.

Tipo de volume raiz do dispositivo com a AMI personalizada

Uma AMI personalizada pode ter tipos diferentes de volume raiz do dispositivo. O Amazon EMR sempre usa seu tipo de volume da AMI personalizada.

- Com as versões 6.15.0 e superiores do Amazon EMR, você pode configurar o tamanho do volume raiz, as IOPS e o throughput da AMI personalizada, desde que esses atributos sejam aplicáveis ao tipo de volume da AMI personalizada.
- Com versões do Amazon EMR inferiores à 6.15.0, você pode configurar apenas o tamanho do volume raiz da AMI personalizada.

Se você não configurar o tamanho do volume raiz, as IOPS ou o throughput ao criar o cluster, o Amazon EMR usa os valores da AMI personalizada, se aplicável. Se decidir configurar esses valores ao criar o cluster, o Amazon EMR usa os valores que você especificar, desde que sejam compatíveis e tenham suporte do volume raiz da AMI personalizada. Para ter mais informações, consulte Usar uma AMI personalizada.

Definição de preços do tamanho do volume raiz do dispositivo

O custo do volume do dispositivo raiz do EBS é proporcional à hora, com base nas cobranças mensais do EBS para esse tipo de volume na região em que o cluster é executado. O mesmo é verdadeiro para volumes de armazenamento. As cobranças são feitas em GB, mas você especifica o tamanho do volume raiz em GiB, portanto, convém considerar isso nas suas estimativas (1 GB é igual a 0.931323 GiB).

Os volumes SSD de uso geral gp2 e gp3 são cobrados de forma diferente. Para estimar as cobranças associadas aos volumes raiz do dispositivo do EBS no seu cluster, use as seguintes fórmulas:

SSD de uso geral gp2

O custo do gp2 inclui somente o tamanho do volume do EBS em GB.

```
($EBS size in GB/month) * 0.931323 / 30 / 24 * EMR_EBSRootVolumesizeInGiB *
 InstanceCount
```

Por exemplo, considere um cluster que tenha um nó primário, um nó central e use a AMI base do Amazon Linux com o volume raiz do dispositivo de 10 GiB padrão. Se o custo do EBS na região for de 0,10 USD por GB ao mês, que acaba somando cerca de 0,00129 USD por instância à hora e 0,00258 USD por hora para o cluster (0,10 USD por GB ao mês dividido por 30 dias, dividido por 24 horas, multiplicado por 10 GB, multiplicado por 2 instâncias de cluster).

SSD de uso geral gp3

O custo do gp3 inclui o tamanho do volume do EBS em GB, as IOPS acima de 3.000 (3.000 IOPS gratuitas) e throughput acima de 125 MB/s (125 MB/s gratuitos).

```
($EBS size in GB/month) * 0.931323 / 30 / 24 * EMR_EBSRootVolumesizeInGiB *
 InstanceCount
+
($EBS IOPS/Month)/30/24* (EMR_EBSRootVolumeIops - 3000) * InstanceCount
+
($EBS throughput/Month)/30/24* (EMR_EBSRootVolumeThroughputInMb/s - 125) *
 InstanceCount
```

Por exemplo, considere um cluster que tenha um nó primário, um nó central e use a AMI base do Amazon Linux com o tamanho do volume raiz do dispositivo de 15 GiB padrão, 4.000 IOPS e 140 throughput. Se o custo do EBS na região for de 0,10 USD por GB ao mês, 0,005 USD por IOPS provisionadas ao mês acima de 3.000 e 0,040 USD por MB provisionados/s ao mês acima de 125. Isso representa aproximadamente 0,009293 USD por instância à hora e 0,018586 USD por hora para o cluster.

Especificação de configurações personalizadas do volume raiz do dispositivo

> **Note**
>
> O tamanho e as IOPS do volume raiz não podem ter uma proporção maior do que 1 volume para 500 IOPS (1:500), enquanto as IOPS e o throughput do volume raiz não podem ter uma proporção maior do que 1 IOPS para 0,25 throughput (1:0,25).

Console

Para especificar os atributos do volume raiz do dispositivo do Amazon EBS usando o console do Amazon EMR

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Selecione Amazon EMR versão 6.15.0 ou superior.
4. Em Configuração do cluster, navegue até a seção Volume raiz do EBS e insira um valor para qualquer um dos atributos que deseja configurar.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

CLI

Para especificar os atributos do volume raiz do dispositivo do Amazon EBS usando a AWS CLI

- Use os parâmetros `--ebs-root-volume-size`, `--ebs-root-volume-iops` e `--ebs-root-volume-throughput` do comando create-cluster conforme mostrado no exemplo a seguir.

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --release-label emr-6.15.0\
--ebs-root-volume-size 20 \
--ebs-root-volume-iops 3000\
--ebs-root-volume-throughput 135\
--instance-groups InstanceGroupType=MASTER,\
InstanceCount=1,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=2,InstanceType=m5.xlarge
```

Configuração de software do cluster

Quando você seleciona uma versão de software, o Amazon EMR usa uma Imagem de máquina da Amazon (AMI) com o Amazon Linux para instalar o software que você escolhe ao iniciar seu cluster, como o Hadoop, o Spark e o Hive. O Amazon EMR fornece novas versões regularmente, adicionando novos atributos, novas aplicações e atualizações gerais. Recomendamos que você use a versão mais recente para executar seu cluster, sempre que possível. A versão mais recente é a opção padrão quando você executa um cluster a partir do console.

Para obter mais informações sobre as versões do Amazon EMR e as versões de software disponíveis em cada versão, consulte o Guia de lançamento do Amazon EMR. Para obter mais informações sobre como editar as configurações padrão de aplicações e softwares instalados no cluster, consulte Configuring applications no Guia de lançamento do Amazon EMR. Algumas versões dos componentes do ecossistema de código aberto do Hadoop e do Spark incluídas nas versões do Amazon EMR têm patches e melhorias, que estão documentados no Guia de lançamento do Amazon EMR.

Além do software padrão e dos aplicativos que estão disponíveis para instalação no seu cluster, você pode usar ações de bootstrap para instalar softwares personalizados. As ações de bootstrap são scripts executados nas instâncias quando o cluster é executado, e que são executados nos novos nós adicionados ao seu cluster quando eles são criados. As ações de bootstrap também são úteis para invocar AWS CLI comandos em cada nó para copiar objetos do Amazon S3 para cada nó em seu cluster.

> **Note**
>
> As ações de bootstrap são usadas de forma diferente no Amazon EMR versão 4.x e posteriores. Para obter mais informações sobre as diferenças das versões 2.x e 3.x da AMI

do Amazon EMR, consulte Differences introduced in 4.x no Guia de lançamento do Amazon EMR.

Criar ações de bootstrap para instalar softwares adicionais

Você pode usar uma ação de bootstrap para instalar softwares adicionais ou personalizar a configuração de instâncias de cluster. As ações de bootstrap são scripts que são executados no cluster depois que o Amazon EMR inicia a instância usando a imagem de máquina da Amazon (AMI) do Amazon Linux. As ações de bootstrap são executadas antes que o Amazon EMR instale as aplicações que você especifica ao criar o cluster e antes que os nós de cluster comecem o processamento de dados. Se você adicionar nós a um cluster em execução, as ações de bootstrap também serão executadas nesses nós da mesma forma. É possível criar ações de bootstrap personalizadas e especificá-las ao criar seu cluster.

A maioria das ações de bootstrap predefinidas para a AMI do Amazon EMR versões 2.x e 3.x não tem suporte no Amazon EMR versões 4.x. Por exemplo, `configure-Hadoop` e `configure-daemons` não são compatíveis com o Amazon EMR versão 4.x. Em vez disso, o Amazon EMR versão 4.x fornece essa funcionalidade nativamente. Para obter mais informações sobre como migrar ações de bootstrap das versões 2.x e 3.x da AMI do Amazon EMR para a versão 4.x do Amazon EMR, acesse Personalizar configuração de clusters e aplicações com versões anteriores da AMI do Amazon EMR no Guia de lançamento do Amazon EMR.

Noções básicas sobre ações de bootstrap

Ações de bootstrap são executadas como o usuário do Hadoop por padrão. Você pode executar uma ação de bootstrap com privilégios de root usando `sudo`.

Todas as interfaces de gerenciamento do Amazon EMR dão suporte a ações de bootstrap. Você pode especificar até 16 ações de bootstrap por cluster fornecendo vários `bootstrap-actions` parâmetros do console ou da API. AWS CLI

No console do Amazon EMR, existe a opção de especificar uma ação de bootstrap ao criar um cluster.

Ao usar a CLI, você pode transmitir referências a scripts de ação de bootstrap ao Amazon EMR, adicionando o parâmetro `--bootstrap-actions` ao criar o cluster usando o comando `create-cluster`.

```
--bootstrap-actions Path="s3://mybucket/filename",Args=[arg1,arg2]
```

Se a ação de bootstrap retornar um código de erro diferente de zero, o Amazon EMR a tratará como uma falha e terminará a instância. Se muitas instâncias falharem em suas ações de bootstrap, o Amazon EMR terminará o cluster. Se apenas algumas instâncias falharem, o Amazon EMR tentará realocar as instâncias com falha e continuar. Use o código de erro de cluster `lastStateChangeReason` para identificar falhas causadas por uma ação de bootstrap.

Executar uma ação de bootstrap condicionalmente

Para executar apenas ações de bootstrap no nó principal, você pode usar uma ação de bootstrap personalizada com um pouco de lógica para determinar se o nó é principal.

```
#!/bin/bash
if grep isMaster /mnt/var/lib/info/instance.json | grep false;
then
    echo "This is not master node, do nothing,exiting"
    exit 0
fi
echo "This is master, continuing to execute script"
# continue with code logic for master node below
```

A saída a seguir será impressa de um nó central.

```
This is not master node, do nothing, exiting
```

A saída a seguir será impressa de um nó principal.

```
This is master, continuing to execute script
```

Para usar essa lógica, carregue a ação de bootstrap, incluindo o código acima, no bucket do Amazon S3. No AWS CLI, adicione o `--bootstrap-actions` parâmetro à chamada da `aws emr create-cluster` API e especifique a localização do script de bootstrap como o valor de`Path`.

Ações de desligamento

Um script de ação de bootstrap pode criar uma ou mais ações de desligamento, escrevendo scripts no diretório `/mnt/var/lib/instance-controller/public/shutdown-actions/`. Quando um cluster é encerrado, todos os scripts nesse diretório são executados em paralelo. Cada script deve ser executado e concluído em até 60 segundos.

Scripts de ação de desligamento não terão garantia de execução se o nó for encerrado com um erro.

> **Note**
>
> Ao usar o Amazon EMR 4.0 e versões posteriores, você deve criar manualmente o diretório `/mnt/var/lib/instance-controller/public/shutdown-actions/` no nó principal. Ele não existe por padrão. No entanto, depois de serem criados, os scripts nesse diretório são executados antes do desligamento. Para obter mais informações sobre como conectar-se ao nó principal para criar diretórios, consulte Conectar-se ao nó primário usando SSH.

Usar ações de bootstrap personalizadas

Você pode criar um script personalizado para executar uma ação de bootstrap personalizada. Qualquer uma das interfaces do Amazon EMR pode referenciar uma ação de bootstrap personalizada.

> **Note**
>
> Para obter o melhor desempenho, recomendamos que você armazene ações de bootstrap, scripts e outros arquivos personalizados que você deseja usar com o Amazon EMR em um bucket do Amazon S3 que esteja no Região da AWS mesmo que seu cluster.

Conteúdo

Adicionar ações de bootstrap personalizadas

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Criar um cluster com uma ação de bootstrap usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Ações de bootstrap, escolha Adicionar para especificar um nome, um local do script e os argumentos opcionais para a ação. Selecione Adicionar ação de bootstrap.
4. Opcionalmente, adicione mais ações de bootstrap.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

Old console

Criar um cluster com uma ação de bootstrap personalizada usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Clique em Go to advanced options (Ir para opções avançadas).
4. Em Create Cluster (Criar cluster) - Advanced Options (Opções avançadas), Etapas 1 e 2, escolha as opções conforme desejado e prossiga para Step 3: General Cluster Settings (Etapa 3: configurações gerais do cluster).
5. Em ações Bootstrap Actions (Ações de bootstrap), selecione Configure and add (Configurar e adicionar) para especificar o nome, o local do JAR e os argumentos para sua ação de bootstrap. Escolha Adicionar.
6. Opcionalmente, adicione mais ações de bootstrap conforme desejar.
7. Proceda para criar o cluster. Suas ações de bootstrap serão executada depois que o cluster tiver sido provisionado e inicializado.

   Enquanto o nó primário do cluster estiver em execução, você poderá conectar-se ao nó primário e visualizar os arquivos de log que o script da ação de bootstrap gerou no diretório `/mnt/var/log/bootstrap-actions/1`.

CLI

Para criar um cluster com uma ação de bootstrap personalizada com o AWS CLI

Ao usar a ação AWS CLI para incluir uma ação de bootstrap, especifique `Path` e `Args` como uma lista separada por vírgulas. O exemplo a seguir não usa uma lista de argumentos.

- Para executar um cluster com uma ação de bootstrap personalizada, digite o comando a seguir e substitua *myKey* pelo nome de seu par de chaves do EC2. Inclua `--bootstrap-actions` como parâmetro e especifique o local do script de bootstrap como o valor de `Path`.
  - Usuários do Linux, do UNIX e do Mac OS X:

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 \
--use-default-roles --ec2-attributes KeyName=myKey \
--applications Name=Hive Name=Pig \
--instance-count 3 --instance-type m5.xlarge \
--bootstrap-actions Path="s3://elasticmapreduce/bootstrap-actions/download.sh"
```

  - Usuários do Windows:

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.2.0 --use-
default-roles --ec2-attributes KeyName=myKey --applications Name=Hive Name=Pig
 --instance-count 3 --instance-type m5.xlarge --bootstrap-actions Path="s3://
elasticmapreduce/bootstrap-actions/download.sh"
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você não tiver criado anteriormente o perfil de serviço do Amazon EMR padrão e o perfil de instância do EC2, digite `aws emr create-default-roles` para criá-los antes de digitar o subcomando `create-cluster`.

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Usar uma ação de bootstrap personalizada para copiar um objeto do Amazon S3 para cada nó

Você pode usar uma ação de bootstrap para copiar objetos do Amazon S3 para cada nó no cluster antes que suas aplicações sejam instaladas. O AWS CLI é instalado em cada nó de um cluster, para que sua ação de bootstrap possa chamar AWS CLI comandos.

O exemplo a seguir demonstra um script de ação de bootstrap simples que copia um arquivo, `myfile.jar`, do Amazon S3 para uma pasta local, `/mnt1/myfolder`, em cada nó do cluster. O script é salvo no Amazon S3 com o nome de arquivo `copymyfile.sh` com os conteúdos a seguir.

```
#!/bin/bash
aws s3 cp s3://mybucket/myfilefolder/myfile.jar /mnt1/myfolder
```

Ao iniciar o cluster, você especifica o script. O AWS CLI exemplo a seguir demonstra isso:

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 \
--use-default-roles --ec2-attributes KeyName=myKey \
--applications Name=Hive Name=Pig \
--instance-count 3 --instance-type m5.xlarge \
--bootstrap-actions Path="s3://mybucket/myscriptfolder/copymyfile.sh"
```

Configurar o hardware e as redes do cluster

Uma consideração importante ao criar um cluster do Amazon EMR é como configurar instâncias do Amazon EC2 e opções de rede. Este capítulo aborda as opções a seguir e vincula todos eles em conjunto com as práticas recomendadas e diretrizes.

- Tipos de nós: as instâncias do Amazon EC2 em um cluster do EMR são organizadas por tipos de nós. Existem três: nós primários, nós centrais e nós de tarefa. Cada tipo de nó realiza um conjunto de funções definidas pelos aplicativos distribuídos que você instala no cluster. Durante um trabalho do Hadoop MapReduce ou do Spark, por exemplo, componentes nos nós principais e de tarefas processam dados, transferem a saída para o Amazon S3 ou o HDFS e fornecem metadados de status de volta ao nó primário. Com um cluster de nó único, todos os componentes são executados no nó primário. Para ter mais informações, consulte Noções básicas sobre tipos de nó: nós primários, centrais e de tarefa.
- Instâncias EC2: ao criar um cluster, você faz escolhas sobre as instâncias do Amazon EC2 nas quais cada tipo de nó será executado. O tipo de instância do EC2 determina o perfil de processamento e armazenamento do nó. A escolha da instância do Amazon EC2 para os nós é

importante porque determina o perfil de performance dos tipos de nós individuais do cluster. Para ter mais informações, consulte Configurar instâncias do Amazon EC2.

- Redes: é possível iniciar o cluster do Amazon EMR em uma VPC usando uma sub-rede pública, uma sub-rede privada ou uma sub-rede compartilhada. A configuração de redes determina como clientes e serviços podem se conectar aos clusters para realizar o trabalho, como os clusters se conectam aos armazenamentos de dados e outros recursos da AWS e as opções que você tem para controlar o tráfego nessas conexões. Para ter mais informações, consulte Configurar redes.
- Agrupamento de instâncias: a coleção de instâncias do EC2 que hospedam cada tipo de nó é chamada de frota de instâncias ou grupo de instâncias uniforme. A configuração de agrupamento de instâncias é uma escolha que deve ser feita ao criar um cluster. Essa escolha determina como você poderá adicionar nós ao cluster enquanto ele estiver em execução. A configuração se aplica a todos os tipos de nó. Não é possível alterá-lo mais tarde. Para ter mais informações, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes.

> **Note**
>
> A configuração de frotas de instância só está disponível em versões do Amazon EMR 4.8.0 e posteriores, exceto versões 5.0.0 e 5.0.3.

Noções básicas sobre tipos de nó: nós primários, centrais e de tarefa

Use esta seção para entender como o Amazon EMR usa cada um desses tipos de nó e como base para planejamento de capacidade do cluster.

Nó primário

O nó primário gerencia o cluster e normalmente executa os componentes primários de aplicações distribuídas. Por exemplo, o nó primário executa o ResourceManager serviço YARN para gerenciar recursos para aplicativos. Ele também executa o NameNode serviço HDFS, rastreia o status dos trabalhos enviados ao cluster e monitora a integridade dos grupos de instâncias.

Para monitorar o progresso de um cluster e interagir diretamente com aplicações, você pode se conectar ao nó primário por SSH como usuário do Hadoop. Para ter mais informações, consulte Conectar-se ao nó primário usando SSH. Conectar-se ao nó primário que você acesse diretórios e arquivos, como os arquivos de log do Hadoop, diretamente. Para ter mais informações, consulte Exibir arquivos de log do . Você também pode visualizar interfaces de usuário que as aplicações

publicam como sites em execução no nó primário. Para ter mais informações, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

> **Note**
>
> Com o Amazon EMR 5.23.0 e versões posteriores, você pode iniciar um cluster com três nós principais para oferecer suporte à alta disponibilidade de aplicativos como YARN Resource Manager, HDFS, Spark, Hive e NameNode Ganglia. O nó primário não é mais um possível ponto de falha único com esse recurso. Se um dos nós primários apresenta falha, o Amazon EMR executa failover automaticamente para um nó primário em espera e substitui o nó primário com falha por um novo com as mesmas ações de configuração e bootstrap. Para obter mais informações, consulte Plan and Configure Primary Nodes.

Nós centrais

Os nós centrais são gerenciados pelo nó primário. Nós core executam o daemon Data Node para coordenar o armazenamento físico de dados como parte do Hadoop Distributed File System (HDFS). Eles também executam o daemon Task Tracker e realizam outras tarefas de computação paralelas nos dados necessários pelos aplicativos instalados. Por exemplo, um nó principal executa NodeManager daemons do YARN, MapReduce tarefas do Hadoop e executores do Spark.

Há apenas um grupo de instâncias centrais ou uma frota de instâncias por cluster, mas pode haver múltiplos nós em execução em múltiplas instâncias do Amazon EC2 no grupo de instâncias ou na frota de instâncias. Com grupos de instâncias, você pode adicionar e remover instâncias do Amazon EC2 enquanto o cluster estiver em execução. Também é possível configurar o ajuste de escala automático para adicionar instâncias com base no valor de uma métrica. Para obter mais informações sobre como adicionar e remover instâncias do Amazon EC2 com a configuração de grupos de instâncias, consulte Usar ajuste de escala de clusters.

Com frotas de instâncias, você pode adicionar e remover instâncias efetivamente, modificando as capacidades de destino da frota de instâncias para sob demanda e spot, conforme necessário. Para obter mais informações sobre capacidades alvo, consulte Opções de frotas de instâncias.

> **Warning**
>
> Há risco de perda de dados ao remover daemons do HDFS de um nó core em execução ou nós core em encerramento. Tenha cuidado ao configurar nós core para usar instâncias spot. Para ter mais informações, consulte Quando você deve usar instâncias spot?.

Nós de tarefa

Você pode usar nós de tarefas para aumentar a potência de realizar tarefas de computação paralela em dados, como tarefas do Hadoop e executores do MapReduce Spark. Nós de tarefa não executam o daemon Data Node, nem armazenam dados no HDFS. Assim como acontece com os nós centrais, você pode adicionar nós de tarefa a um cluster, adicionando instâncias do Amazon EC2 a um grupo de instâncias uniforme existente ou modificando as capacidades alvo para uma frota de instâncias de tarefa.

Com a configuração de grupo de instâncias uniforme, você pode ter um total de 48 grupos de instâncias de tarefa. A capacidade de adicionar grupos de instâncias dessa forma permite que você combine tipos de instância do Amazon EC2 e opções de preços, como instâncias sob demanda e instâncias spot. Isso proporciona a flexibilidade necessária para atender aos requisitos de workload de uma maneira econômica.

Com a configuração de frota de instâncias, a capacidade de combinar tipos de instâncias e opções de compra está integrada e, portanto, há apenas uma frota de instâncias de tarefa.

Como as instâncias spot são frequentemente usadas para executar nós de tarefas, o Amazon EMR tem a funcionalidade padrão para programar trabalhos do YARN para que os trabalhos em execução não falhem quando os nós de tarefas em execução nas instâncias spot forem encerrados. O Amazon EMR faz isso ao permitir que processos principais de aplicações sejam executados somente em nós centrais. O processo principal da aplicação controla os trabalhos em execução e precisa permanecer ativo durante a vida útil do trabalho.

A versão 5.19.0 e as versões posteriores do Amazon EMR usam o recurso de rótulos de nós do YARN integrado para conseguir isso. (As versões anteriores usavam um patch de código). As propriedades nas classificações de configuração `yarn-site` e `capacity-scheduler` são configuradas por padrão para que o programador de capacidade e o programador justo do YARN aproveitem os rótulos de nós. O Amazon EMR rotula automaticamente os nós centrais com o rótulo `CORE` e define propriedades para que as aplicações principais sejam programadas somente em nós com o rótulo CORE. Modificar manualmente as propriedades relacionadas nas classificações de

configuração yarn-site e docapacity-scheduler, ou diretamente nos arquivos XML associados, pode interromper esse recurso ou modificar essa funcionalidade.

A partir do Amazon EMR série 6.x, o recurso de rótulos de nó do YARN é desabilitado por padrão. Os processos primários da aplicação podem ser executados tanto nos nós centrais como nos nós de tarefa por padrão. É possível habilitar o recurso de rótulos de nó do YARN configurando as seguintes propriedades:

- `yarn.node-labels.enabled: true`
- `yarn.node-labels.am.default-node-label-expression: 'CORE'`

Para obter informações sobre as propriedades específicas, consulte Configurações do Amazon EMR para evitar falhas em trabalhos causado pelo término de instâncias spot de nós de tarefa.

Configurar instâncias do Amazon EC2

Instâncias do EC2 acompanham diferentes configurações conhecidas como tipos de instâncias. Os tipos de instância têm capacidades diferentes de CPU, entrada/saída e armazenamento. Além do tipo de instância, você pode escolher diferentes opções de compra para instâncias do Amazon EC2. Você pode especificar diferentes tipos de instâncias e opções de compra em grupos de instâncias uniformes ou frotas de instâncias. Para ter mais informações, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes. Para obter orientação sobre como escolher tipos de instância e opções de compra para sua aplicação, consulte Práticas recomendadas para configuração de clusters.

> **Important**
>
> Quando você escolhe um tipo de instância usando o AWS Management Console, o número de vCPUs mostrado para cada tipo de instância é o número de vcores YARN para esse tipo de instância, não o número de vCPUs EC2 para esse tipo de instância. Para obter mais informações sobre o número de vCPUs para o seu tipo de instância, consulte os tipos de instância do Amazon EC2.

Tópicos

- Tipos de instâncias compatíveis
- Configurar redes

- Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes

Tipos de instâncias compatíveis

Esta seção descreve os tipos de instância compatíveis o Amazon EMR, organizados por Região da AWS. Para saber mais sobre os tipos de instância, consulte Instâncias do Amazon EC2 e Matriz de tipo de instância da Amazon Linux AMI.

Nem todos os tipos de instância estão disponíveis em todas as regiões. A disponibilidade da instância está sujeita à disponibilidade e à demanda na região e zona de disponibilidade especificadas. A zona de disponibilidade da instância é determinada pela sub-rede usada para iniciar o cluster.

Considerações

Considere o seguinte ao escolher os tipos de instância do cluster do Amazon EMR.

> **Important**
>
> Quando você escolhe um tipo de instância usando o AWS Management Console, o número de vCPUs mostrado para cada tipo de instância é o número de vcores YARN para esse tipo de instância, não o número de vCPUs EC2 para esse tipo de instância. Para obter mais informações sobre o número de vCPUs para o seu tipo de instância, consulte os tipos de instância do Amazon EC2.

- Se você criar um cluster usando um tipo de instância que não está disponível na região e na zona de disponibilidade especificadas, o cluster poderá falhar ao tentar provisionar ou pode ficar preso no estado de provisionamento. Para obter informações sobre a disponibilidade de instâncias, consulte a página de preços do Amazon EMR ou veja as tabelas Tipos de instância compatíveis com Região da AWS nesta página.
- Começando com a versão 5.13.0 do Amazon EMR, todas as instâncias usam a virtualização de HVM e armazenamento baseado em EBS para volumes raiz. Ao usar versões do Amazon EMR anteriores à 5.13.0, algumas instâncias de gerações anteriores usam a virtualização de PVM. Para obter mais informações, consulte Tipos de virtualização da AMI em Linux.
- Alguns tipos de instâncias oferecem suporte a redes avançadas. Para obter mais informações, consulte Redes avançadas no Linux.
- Os drivers NVIDIA e CUDA são instalados em tipos de instância GPU por padrão.

Tipos de instância compatíveis com Região da AWS

As tabelas a seguir listam os tipos de instância do Amazon EC2 que o Amazon EMR suporta, organizados por. Região da AWS As tabelas também listam as primeiras versões do Amazon EMR nas séries 5.x, 6.x e 7.x que oferecem suporte a cada tipo de instância.

Leste dos EUA (Norte da Virgínia) – us-east-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g6.xlarge	emr-7.3.0
	g6.2xlarge	emr-7.3.0
	g6.4xlarge	emr-7.3.0
	g6.8xlarge	emr-7.3.0
	g6.12xlarge	emr-7.3.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g6.16xlarge	emr-7.3.0
	g6.24xlarge	emr-7.3.0
	g6.48xlarge	emr-7.3.0
	gr6.4xlarge	emr-7.3.0
	gr6.8xlarge	emr-7.3.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p5.48xlarge	emr-6.14.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Leste dos EUA (Ohio): us-east-2

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.6.0, emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.6.0, emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.6.0, emr-5.0.1, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.6.0, emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7a.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g6.xlarge	emr-7.3.0
	g6.2xlarge	emr-7.3.0
	g6.4xlarge	emr-7.3.0
	g6.8xlarge	emr-7.3.0
	g6.12xlarge	emr-7.3.0
	g6.16xlarge	emr-7.3.0
	g6.24xlarge	emr-7.3.0
	g6.48xlarge	emr-7.3.0
	gr6.4xlarge	emr-7.3.0
	gr6.8xlarge	emr-7.3.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p5.48xlarge	emr-6.14.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7a.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	h1.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.1, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Oeste dos EUA (Norte da Califórnia): us-west-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Oeste dos EUA (Oregon): us-west-2

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g6.xlarge	emr-7.3.0
	g6.2xlarge	emr-7.3.0
	g6.4xlarge	emr-7.3.0
	g6.8xlarge	emr-7.3.0
	g6.12xlarge	emr-7.3.0
	g6.16xlarge	emr-7.3.0
	g6.24xlarge	emr-7.3.0
	g6.48xlarge	emr-7.3.0
	gr6.4xlarge	emr-7.3.0
	gr6.8xlarge	emr-7.3.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p5.48xlarge	emr-6.14.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

AWS GovCloud (Oeste dos EUA) - -1 us-gov-west

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

AWS GovCloud (Leste dos EUA) - -1 us-gov-east

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para computação	c5.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.16.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

África (Cidade do Cabo): af-south-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para memória	r5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizada para armazenamento	i3.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Hong Kong): ap-east-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Jacarta): ap-southeast-3

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m5d.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	m6g.xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6g.2xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6g.4xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6g.8xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6g.12xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.16xlarge	emr-5.30.2, emr-6.1.1, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.9xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.18xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.9xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.18xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5d.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.9xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c5n.18xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	c6g.xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6g.2xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.4xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6g.8xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6g.12xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6g.16xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Computação acelerada	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r5d.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r6g.xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6g.2xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.4xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6g.8xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6g.12xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6g.16xlarge	emr-5.31.1, emr-6.1.1, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r7i.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	r7i.48xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3.4xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3.8xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3.16xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.2xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.3xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.6xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i3en.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.30.2, emr-6.0.1, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Mumbai): ap-south-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Ásia-Pacífico (Hyderabad): ap-south-2

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para memória	r5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Asia Pacific (Osaka): ap-northeast-3

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Seul): ap-northeast-2

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Ásia-Pacífico (Singapura): ap-southeast-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Ásia-Pacífico (Sydney) – ap-southeast-2

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	r6a.32xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	r6a.48xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Ásia-Pacífico (Tóquio) – ap-northeast-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	r6a.32xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	r6a.48xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Canadá (Central): ca-central-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.8.2, emr-5.0.2, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i-flex.2xlarge	emr-4.8.2, emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.8.2, emr-5.0.2, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.8.2, emr-5.0.2, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7i.xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.24xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.24xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.2, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Oeste do Canadá (Calgary): ca-west-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6i.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.9xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.18xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6gn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6i.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6g.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6i.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3en.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.3xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

China (Ningxia): cn-northwest-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

China (Pequim): cn-north-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Europa (Frankfurt): eu-central-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Europa (Zurique): eu-central-2

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	d3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	d3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	d3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Europa (Irlanda): eu-west-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7iz.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7iz.32xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	d3en.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Europa (Londres): eu-west-2

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7i.xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.2xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
Computação acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0, emr-7.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0, emr-7.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.12xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Otimizada para armazenamento	d3.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	d3.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.3, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Europa (Milão): eu-south-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.29.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Europa (Espanha): eu-south-2

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i-flex.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i-flex.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i-flex.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m7i-flex.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para computação	c5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7a.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7a.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7a.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c7i.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7a.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7a.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r7i.48xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Europa (Paris): eu-west-3

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.4xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.5.3, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Europa (Estocolmo): eu-north-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7a.48xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i-flex.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para computação	c5.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7a.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.32xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7a.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7g.8xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0, emr-7.0.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	p5.48xlarge	emr-6.14.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7a.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7a.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.32xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7a.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	r7i.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.4xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.17.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Oriente Médio (Bahrein): me-south-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizadas para computação	c5.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5d.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.9xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.8xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3.16xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.12xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.24.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Oriente Médio (EAU): me-central-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	m6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
Computação acelerada	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Otimizada para armazenamento	i3.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.36.0, emr-6.7.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

América do Sul (São Paulo): sa-east-1

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
Finalidade geral	m5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0, emr-7.0.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m6id.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7g.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	m7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.2xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	m7i-flex.4xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	m7i-flex.8xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
Otimizadas para computação	c5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5a.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6a.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c6in.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	c7i.xlarge	emr-4.2.0, emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	c7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	c7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
Computação acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0, emr-7.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	g5.12xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizadas para memória	r5.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5a.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0, emr-7.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r5d.12xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0, emr-7.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6g.4xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0, emr-7.0.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0, emr-7.0.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r6i.4xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0, emr-7.0.0
	r7i.xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.2xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.4xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.8xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.16xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	r7i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	r7i.48xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0, emr-7.0.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0, emr-7.0.0
Otimizada para armazenamento	i3.xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i3.16xlarge	emr-5.9.0, emr-6.0.0, emr-7.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0, emr-7.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.12xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0

Classe de instância	Tipo de instância	Versão mínima do Amazon EMR com suporte (5.x, 6.x, 7.x)
	i4i.16xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0
	i4i.24xlarge	emr-5.0.0, emr-6.0.0, emr-7.0.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0, emr-7.0.0

Instâncias da geração anterior

O Amazon EMR oferece suporte a instâncias de gerações anteriores para oferecer suporte a aplicações que são otimizadas para essas instâncias e ainda não foram atualizadas. Para obter mais informações sobre esses tipos de instâncias e caminhos de atualização, consulte Instâncias de gerações anteriores.

Classe de instância	Tipos de instância
General Purpose	m1.small[1] | m1.medium[1] | m1.large[1] | m1.xlarge[1] | m3.xlarge[1] | m3.2xlarge[1] | m4.large | m4.xlarge | m4.2xlarge | m4.4xlarge | m4.10xlarge | m4.16xlarge
Compute Optimized	c1.medium[1] [2] | c1.xlarge[1] | c3.xlarge[1] | c3.2xlarge[1] | c3.4xlarge[1] | c3.8xlarge[1] | c4.large | c4.xlarge | c4.2xlarge | c4.4xlarge | c4.8xlarge
Memory Optimized	m2.xlarge[1] | m2.2xlarge[1] | m2.4xlarge[1] | r3.xlarge | r3.2xlarge | r3.4xlarge | r3.8xlarge | r4.xlarge | r4.2xlarge | r4.4xlarge | r4.8xlarge | r4.16xlarge
Storage Optimized	d2.xlarge | d2.2xlarge | d2.4xlarge | d2.8xlarge | i2.xlarge | i2.2xlarge | i2.4xlarge | i2.8xlarge

¹ Usa AMI de virtualização PVM com versões anteriores ao Amazon EMR 5.13.0. Para obter mais informações, consulte Tipos de virtualização de AMI no Linux.

² Sem suporte na versão 5.15.0.

Opções de compra de instância

Ao configurar um cluster, você escolhe uma opção de compra para instâncias do Amazon EC2. É possível escolher instâncias sob demanda, instâncias spot ou ambas. Os preços variam com base no tipo de instância e na região. O preço do Amazon EMR é um acréscimo ao preço do Amazon EC2 (o preço dos servidores subjacentes) e ao preço do Amazon EBS (ao anexar volumes do Amazon EBS). Para obter os preços atuais, consulte Preço do Amazon EMR.

Sua opção para usar grupos de instâncias ou frotas de instâncias no cluster determina como você pode alterar opções de compra de instância enquanto um cluster está em execução. Ao escolher grupos de instâncias uniformes, você só poderá especificar a opção de compra para um grupo de instâncias ao criá-lo, e o tipo de instância e a opção de compra se aplicarão a todas as instâncias do Amazon EC2 em cada grupo de instâncias. Se você optar por usar frotas de instâncias, poderá alterar as opções de compra após criar a frota de instância, e poderá combinar opções de compra para preencher uma capacidade alvo especificada por você. Para obter mais informações sobre essas configurações, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes.

Instâncias sob demanda

Com instâncias sob demanda, você paga pela capacidade computacional por segundo. Opcionalmente, você pode fazer com que essas instâncias sob demanda usem as opções de compra de instâncias reservadas ou dedicadas. Com instâncias reservadas, você faz um pagamento único por uma instância para reservar capacidade. As instâncias dedicadas são fisicamente isoladas no nível do hardware do host das instâncias que pertencem a outras AWS contas. Para obter mais informações sobre as opções de compra, consulte Opções de compra de instâncias no Guia do usuário do Amazon EC2.

Usar instâncias reservadas

Para usar instâncias reservadas no Amazon EMR, você usa o Amazon EC2 para adquirir a instância reservada e especifica os parâmetros da reserva, incluindo o escopo da reserva aplicável a uma região ou a uma zona de disponibilidade. Para obter mais informações, consulte Instâncias reservadas do Amazon EC2 e Comprar instâncias reservadas no Guia do usuário do Amazon EC2.

Após você comprar uma instância reservada, ela será usada pelo Amazon EMR quando um cluster for iniciado e se todas as condições a seguir forem verdadeiras:

- Uma instância sob demanda é especificada na configuração do cluster que corresponde à especificação da instância reservada.
- O cluster é executado no escopo da reserva de instância (a zona de disponibilidade ou região).
- A capacidade da Instância reservada ainda está disponível

Por exemplo, digamos que você compre uma instância reservada `m5.xlarge` com a reserva de instância direcionada à região US-East. Em seguida, inicie um cluster do Amazon EMR em US-Leste que use duas instâncias `m5.xlarge`. A primeira instância é cobrada de acordo com a taxa da Instância reservada, e a outra de acordo com a taxa Sob demanda. A capacidade da Instância reservada é usada antes que as Instâncias sob demanda sejam criadas.

Usar instâncias dedicadas

Para usar Instâncias dedicadas, você as compra usando o Amazon EC2 e depois cria uma VPC com o atributo de locação Dedicated. Em seguida, no Amazon EMR, você especifica que um cluster deve ser executado nessa VPC. Todas as instâncias sob demanda no cluster que correspondem com a especificação de instâncias dedicadas usam as instâncias dedicadas disponíveis quando o cluster é executado.

> **Note**
>
> O Amazon EMR não oferece suporte à configuração do atributo `dedicated` em instâncias individuais.

Instâncias spot

Instâncias spot no Amazon EMR fornecem uma opção para você comprar capacidade de instâncias do Amazon EC2 a um custo reduzido em comparação à compra sob demanda. A desvantagem de usar instâncias spot é que as instâncias podem ser terminadas se a capacidade spot ficar indisponível para o tipo de instância que você está executando. Para obter mais informações sobre quando usar instâncias spot pode ser apropriado para seu aplicativo, consulte Quando você deve usar instâncias spot?.

Quando o Amazon EC2 tem capacidade não utilizada, ele oferece instâncias do EC2 a um custo reduzido, chamado de preço spot. Esse preço flutua com base na disponibilidade e na demanda

e é estabelecido por região e zona de disponibilidade. Quando você escolhe instâncias spot, você especifica o preço spot máximo que está disposto a pagar por cada tipo de instância do EC2. Quando o preço spot na zona de disponibilidade do cluster estiver abaixo do preço máximo especificado para esse tipo de instância, as instâncias serão executadas. Enquanto as instâncias forem executadas, você será cobrado de acordo com o preço spot atual e não o preço spot máximo.

> **Note**
>
> As instâncias spot com duração definida (também conhecidas como blocos spot) não estarão mais disponíveis para novos clientes a partir de 1.º de julho de 2021. Aos clientes que utilizaram o recurso anteriormente, continuaremos a oferecer suporte a instâncias spot com duração definida até 31 de dezembro de 2022.

Para obter os preços atuais, consulte Preço das instâncias spot do Amazon EC2. Para obter mais informações, consulte Instâncias spot no Guia do usuário do Amazon EC2. Ao criar e configurar um cluster, você especifica as opções de rede que, em última análise, determinam a Zona de disponibilidade na qual seu cluster é executado. Para ter mais informações, consulte Configurar redes.

> **Tip**
>
> Você pode ver o preço spot em tempo real no console ao passar o mouse sobre a dica de ferramenta de informações ao lado da opção de compra de Spot quando criar um cluster usando as Advanced Options (Opções avançadas). Os preços de cada zona de disponibilidade na região selecionada são exibidos. Os preços mais baixos estão nas linhas de cor verde. Devido à flutuação dos preços Spot entre as Zonas de disponibilidade, selecionar a Zona de disponibilidade com o menor preço inicial pode não resultar no menor preço durante a vigência do cluster. Para obter os melhores resultados, estude o histórico de preços da Zona de disponibilidade antes de escolher. Para obter mais informações, consulte o histórico de preços de instâncias spot no Guia do usuário do Amazon EC2.

As opções de instâncias Spot dependem de você usar grupos de instâncias uniformes ou frotas de instâncias na sua configuração de cluster.

Instâncias Spot em grupos de instâncias uniformes

Quando você usar instâncias Spot em um grupo de instâncias uniforme, todas as instâncias desse grupo devem ser instâncias Spot. Você especifica uma única sub-rede ou Zona de disponibilidade para o cluster. Para cada grupo de instâncias, você especifica uma única instância spot e um preço spot máximo. As instâncias spot desse tipo serão executadas se o preço spot na região e na zona de disponibilidade do cluster estiver abaixo do preço spot máximo. As instâncias serão encerradas se o preço spot estiver acima do preço spot máximo. Você define o preço spot máximo somente ao configurar um grupo de instâncias. Não é possível alterá-lo mais tarde. Para ter mais informações, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes.

Instâncias Spot em frotas de instâncias

Quando você usa a configuração de frotas de instâncias, opções adicionais dão maior controle sobre como as instâncias Spot são executadas e encerradas. Fundamentalmente, frotas de instâncias usam um método diferente daquele de grupos de instâncias uniformes para executar instâncias. Isso funciona porque estabelecer uma capacidade alvo para instâncias Spot (e instâncias sob demanda) e até cinco tipos de instâncias. Você também pode especificar uma capacidade ponderada para cada tipo de instância ou usar a vCPU (vcores YARN) do tipo de instância como capacidade ponderada. Essa capacidade ponderada conta para a capacidade de destino quando uma instância desse tipo é provisionada. O Amazon EMR provisiona instâncias com ambas as opções de compra, até que a capacidade de cada de destino seja preenchida. Além disso, é possível definir um intervalo de zonas de disponibilidade para que o Amazon EMR escolha ao executar instâncias. Você também fornece opções spot adicionais para cada frota, incluindo um tempo limite de provisionamento. Para ter mais informações, consulte Configurar frotas de instâncias.

Armazenamento de instâncias

Visão geral

O armazenamento de instância e de volumes do Amazon EBS é usado para dados do HDFS e para buffers, caches, dados transitórios e outros conteúdos temporários que algumas aplicações podem “vazar” para o sistema de arquivos local.

O Amazon EBS funciona de forma diferente dentro do Amazon EMR do que com instâncias do Amazon EC2 regulares. Os volumes do Amazon EBS anexados aos clusters do Amazon EMR são temporários: os volumes são excluídos após o término do cluster e da instância (por exemplo, ao reduzir grupos de instâncias), portanto, não espere a persistência dos dados. Embora os dados sejam temporários, é possível que os dados no HDFS sejam replicados dependendo do número e da especialização dos nós no cluster. Quando você adiciona volumes de armazenamento do Amazon EBS, eles são montados como volumes adicionais. Eles não fazem parte do volume de inicialização.

O YARN está configurado para usar todos os volumes adicionais, mas você é responsável por alocá-los como armazenamento local (para arquivos de log locais, por exemplo).

Considerações

Lembre-se destas considerações adicionais ao usar o Amazon EBS com clusters do EMR:

- Você não pode fazer snapshot de um volume do Amazon EBS e restaurá-lo no Amazon EMR. Para criar configurações personalizadas reutilizáveis, use uma AMI personalizada (disponível no Amazon EMR 5.7.0 e versões posteriores). Para ter mais informações, consulte Usar uma AMI personalizada.
- Um volume de armazenamento raiz do Amazon EBS criptografado tem suporte apenas ao usar uma AMI personalizada. Para ter mais informações, consulte Criar uma AMI personalizada com o volume do dispositivo raiz do Amazon EBS criptografado.
- Se você aplicar etiquetas usando a API do Amazon EMR, essas operações serão aplicadas a volumes do EBS.
- Existe um limite de 25 volumes por instância.
- Os volumes do Amazon EBS nos nós centrais não podem ter menos de 5 GB.

Armazenamento padrão do Amazon EBS para instâncias

Para instâncias do EC2 com armazenamento exclusivo do EBS, o Amazon EMR aloca volumes de armazenamento gp2 ou gp3 do Amazon EBS a instâncias. Ao criar um cluster usando o Amazon EMR 5.22.0 e versões superiores, a quantidade padrão de armazenamento do Amazon EBS aumenta de acordo com o tamanho da instância.

Dividimos qualquer aumento de armazenamento em vários volumes. Isso aumenta a performance de IOPS e, por sua vez, a performance de algumas workloads padronizadas. Se quiser usar uma configuração de armazenamento de instância diferente do Amazon EBS, isso poderá ser especificado ao criar um cluster do EMR ou adicionar nós a um cluster existente. É possível usar volumes gp2 ou gp3 do Amazon EBS como volumes raiz e adicionar volumes gp2 ou gp3 como volumes adicionais. Para ter mais informações, consulte Especificar volumes de armazenamento adicionais do EBS.

A tabela a seguir identifica o número padrão de volumes do Amazon EBS, tamanhos e tamanhos totais de armazenamento gp2 por tipo de instância. Para obter informações sobre volumes gp2 comparados aos gp3, consulte Comparar os tipos de volume gp2 e gp3 do Amazon EBS.

Tamanho e volumes de armazenamento padrão gp2 do Amazon EBS por tipo de instância para o Amazon EMR 5.22.0 e versões superiores

Tamanho da instância	Número de volumes	Tamanho do volume (GiB)	Tamanho total (GiB)
*.large	1	32	32
*.xlarge	2	32	64
*.2xlarge	4	32	128
*.4xlarge	4	64	256
*.8xlarge	4	128	512
*.9xlarge	4	144	576
*.10xlarge	4	160	640
*.12xlarge	4	192	768
*.16xlarge	4	256	1024
*.18xlarge	4	288	1152
*.24xlarge	4	384	1536

Volume raiz padrão do Amazon EBS para instâncias

Nas versões 6.15 e superiores, o Amazon EMR anexa automaticamente um volume SSD de uso geral do Amazon EBS (gp3) como dispositivo raiz às AMIs para melhorar a performance. Nas versões anteriores, o Amazon EMR anexa um volume SSD de uso geral do EBS (gp2) como dispositivo raiz.

	6.15 e superior	6.14 e inferior
Tipo de volume raiz padrão		
Tamanho padrão		

	6.15 e superior	6.14 e inferior
IOPS padrão		
Throughput padrão		

Para obter informações sobre como personalizar o volume raiz do dispositivo do Amazon EBS, consulte Especificar volumes de armazenamento adicionais do EBS.

Especificar volumes de armazenamento adicionais do EBS

Ao configurar tipos de instâncias no Amazon EMR, você pode especificar volumes do EBS adicionais para aumentar a capacidade além do armazenamento de instância (se houver) e do volume do EBS padrão. O Amazon EBS fornece os seguintes tipos de volumes: uso geral (SSD), IOPS provisionadas (SSD), otimizado para throughput (HDD), a frio (HDD) e Magnético. Eles diferem em características de performance e preço, para que você possa adaptar seu armazenamento às necessidades analíticas e comerciais das suas aplicações. Por exemplo, algumas aplicações podem precisar ser transferidas para o disco, enquanto outras podem trabalhar com segurança na memória ou usando o Amazon S3.

Você só pode anexar volumes do Amazon EBS a instâncias na inicialização do cluster e ao adicionar um grupo de instâncias de nós de tarefa. Se uma instância em um cluster do Amazon EMR falhar, tanto ela quanto os volumes do Amazon EBS anexados serão substituídos pelos novos volumes. Consequentemente, se você separar manualmente um volume do Amazon EBS, o Amazon EMR o tratará como uma falha e substituirá os armazenamentos de instância (se aplicável) e de volume.

Com o Amazon EMR, não é possível modificar o tipo de volume de gp2 para gp3 para um cluster do EMR já existente. Para usar o gp3 nas suas workloads, execute um novo cluster do EMR. Além disso, não é recomendável atualizar o throughput e as IOPS de um cluster que esteja em uso ou que esteja sendo provisionado, pois o Amazon EMR usa os valores de throughput e de IOPS especificados no momento de execução do cluster para qualquer nova instância adicionada durante o aumento vertical da escala do cluster. Para obter mais informações, consulte Comparar os tipos de volume gp2 e gp3 do Amazon EBS e Selecionar IOPS e throughput ao migrar para gp3.

> **Important**
>
> Para utilizar um volume gp3 com o cluster do EMR, execute um novo cluster.

Comparar os tipos de volume gp2 e gp3 do Amazon EBS

Veja aqui uma comparação dos custos entre os volumes gp2 e gp3 na região Leste dos EUA (Norte da Virgínia). Para obter as informações mais atualizadas, consulte a página do produto Volumes de uso geral do Amazon EBS e a página de preços do Amazon EBS.

Tipo de volume	gp3	gp2
Tamanho do volume	1 GiB – 16 TiB	1 GiB – 16 TiB
IOPS padrão/de referência	3000	3 IOPs/GiB (mínimo 100 IOPS) até um máximo de 16 mil IOPS. Volumes menores que 1 TiB também podem se expandir até 3 mil IOPS.
IOPS máxima/volume	16.000	16.000
Throughput padrão/de referência	125 MiB/s	O limite de throughput é entre 128 MiB/s e 250 MiB/s, dependendo do tamanho do volume.
Throughput máximo/volume	1.000 MiB/s	250 MiB/s
Preço	USD 0,08/GiB por mês 3 mil IOPS gratuitas e USD 0,005/IOPS provisionadas por mês acima de 3 mil; 125 MiB/s gratuitos e USD 0,04/MiB/ss provisionados por mês acima de 125 MiB/s	USD 0,10 USD/Gib por mês

Selecionar IOPS e throughput ao migrar para gp3

Ao provisionar um volume gp2, é necessário descobrir o tamanho do volume para obter as IOPS e o throughput proporcionais. Com o gp3, não necessário provisionar um volume maior para aumentar a performance. Você pode escolher o tamanho e a performance desejados de acordo com a necessidade da aplicação. Selecionar o tamanho certo e os parâmetros de performance certos (IOPS, throughput) pode proporcionar a máxima redução de custos, sem afetar a performance.

Aqui está uma tabela para ajudar você a selecionar as opções de configuração do gp3:

Tamanho do volume	IOPS	Throughput
1-170 GiB	3000	125 MiB/s
170-334 GiB	3000	125 MiB/s; se o tipo de instância do EC2 escolhido oferecer suporte para 125 MiB/s ou menos, use mais conforme o uso, máx. 250 MiB/s*.
334-1000 GiB	3000	125 MiB/s; se o tipo de instância do EC2 escolhido oferecer suporte para 125 MiB/s ou menos, use mais conforme o uso, máx. 250 MiB/s*.
1000+ GiB	Combine IOPS gp2 (tamanho em GiB x 3) ou máximo de IOPS determinado pelo volume gp2 atual	125 MiB/s; se o tipo de instância do EC2 escolhido oferecer suporte para 125 MiB/s ou menos, use mais conforme o uso, máx. 250 MiB/s*.

*O Gp3 tem a capacidade de fornecer throughput de até 1000 MiB/s. Como o gp2 fornece throughput máximo de 250 MiB/s, talvez não seja necessário ultrapassar esse limite ao usar o gp3.

Configurar redes

A maioria dos clusters é iniciada na rede virtual usando a Amazon Virtual Private Cloud (Amazon VPC). Uma VPC é uma rede virtual isolada AWS que está logicamente isolada em sua conta. AWS É possível configurar aspectos como intervalos de endereços IP privados, sub-redes, tabelas de roteamento e gateways de rede. Para obter mais informações, consulte o Manual do usuário da Amazon VPC.

A VPC oferece os seguintes recursos:

- Processamento de dados confidenciais

  Executar um cluster em uma VPC é semelhante a executá-lo em uma rede privada com ferramentas adicionais, como tabelas de roteamento e ACLs de rede, para definir quem tem acesso à rede. Se você estiver processando dados confidenciais no seu cluster, talvez queira o controle de acesso adicional que a execução do seu cluster em uma VPC é capaz de fornecer. Além disso, você pode optar por executar seus recursos em uma sub-rede privada, em que nenhum deles tem conectividade direta com a Internet.
- Acesso a recursos em uma rede interna

  Se sua fonte de dados estiver localizada em uma rede privada, pode ser impraticável ou indesejável carregar esses dados AWS para importação no Amazon EMR, seja devido à quantidade de dados a serem transferidos ou devido à natureza confidencial dos dados. Em vez disso, você pode executar o cluster em uma VPC e conectar seu datacenter à VPC por meio de conexão VPN, permitindo que o cluster acesse recursos na sua rede interna. Por exemplo, se você tiver um banco de dados Oracle no seu datacenter, o lançamento do seu cluster em uma VPC conectada a essa rede pela VPN torna possível que o cluster acesse o banco de dados Oracle.

Sub-redes públicas e privadas

Você pode executar clusters do Amazon EMR em sub-redes VPC públicas e privadas. Isso significa que você não precisa de conectividade com a Internet para executar um cluster do Amazon EMR; no entanto, talvez seja necessário configurar a conversão de endereços de rede (NAT) e gateways de VPN para acessar serviços ou recursos localizados fora da VPC, por exemplo, em uma intranet corporativa ou em endpoints de serviço público, como. AWS AWS Key Management Service

> **Important**
>
> O Amazon EMR só oferece suporte à inicialização de clusters em sub-redes privadas nas versões 4.2 ou posteriores.

Para obter mais informações sobre o Amazon VPC, consulte o Guia do usuário da Amazon VPC.

Tópicos

- Opções da Amazon VPC
- Configurar uma VPC para hospedar clusters
- Iniciar clusters em uma VPC
- Política mínima do Amazon S3 para uma sub-rede privada
- Mais recursos para saber mais sobre VPCs

Opções da Amazon VPC

Ao executar um cluster do Amazon EMR em uma VPC, é possível executá-lo em uma sub-rede pública, privada ou compartilhada. Existem pequenas diferenças, porém significativas, na configuração, dependendo do tipo de sub-rede escolhido para um cluster.

Sub-redes públicas

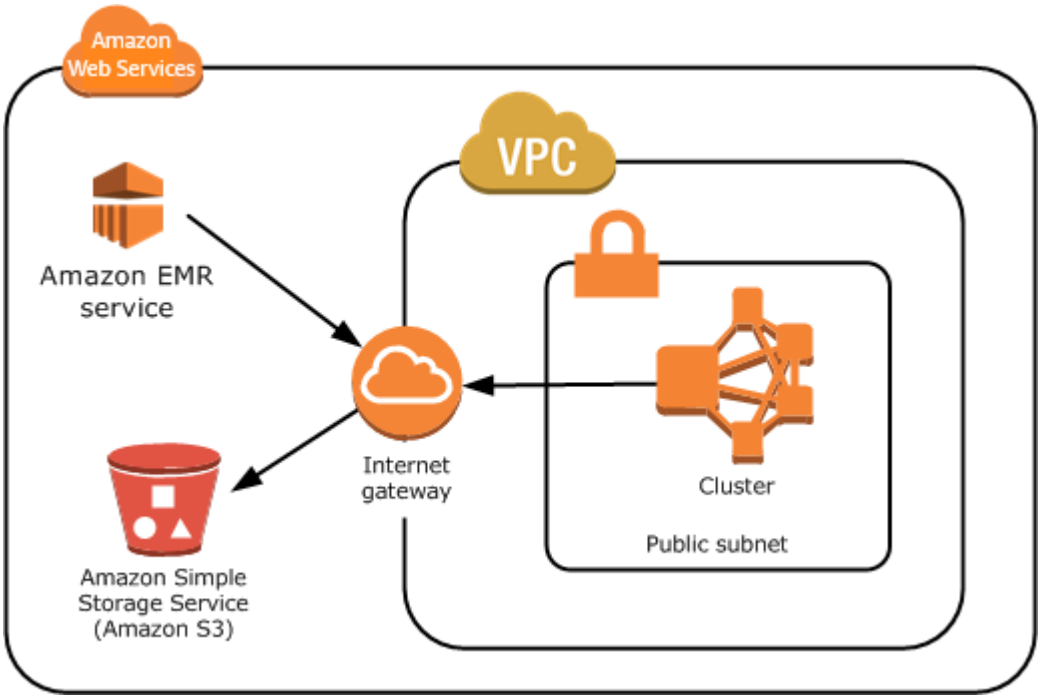
Os clusters do EMR em uma sub-rede pública exigem um gateway da Internet conectado. Isso ocorre porque os clusters do Amazon EMR devem acessar os AWS serviços e o Amazon EMR. Se um serviço, como o Amazon S3, oferecer a capacidade de criar um endpoint da VPC, você poderá acessar esses serviços usando o endpoint em vez de acessar um endpoint público por meio de um gateway da Internet. Além disso, o Amazon EMR não pode se comunicar com clusters em sub-redes públicas por meio de um dispositivo de conversão de endereços de rede (NAT). É necessário um gateway da Internet para essa finalidade, mas você ainda pode usar uma instância NAT ou um gateway para outros tipos de tráfego em cenários mais complexos.

Todas as instâncias em um cluster se conectam ao Amazon S3 por meio de um endpoint da VPC ou de um gateway da Internet. Outros AWS serviços que atualmente não oferecem suporte a endpoints VPC usam somente um gateway de Internet.

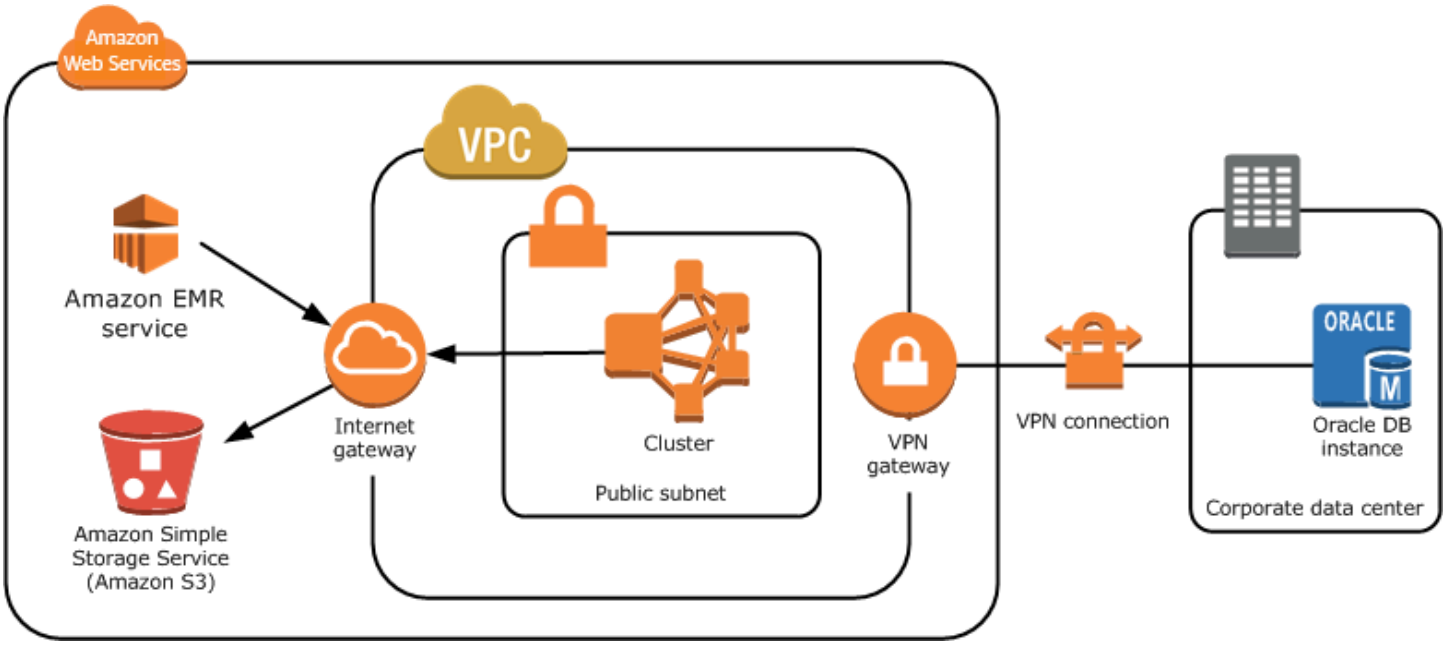
Se você tiver AWS recursos adicionais que não deseja conectar ao gateway da Internet, poderá iniciar esses componentes em uma sub-rede privada que você cria na sua VPC.

Clusters em execução em uma sub-rede pública usam dois grupos de segurança: um para o nó primário e outro para os nós centrais e de tarefa. Para ter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

O diagrama a seguir mostra como um cluster do Amazon EMR é executado em uma VPC usando uma sub-rede pública. O cluster é capaz de se conectar a outros AWS recursos, como buckets do Amazon S3, por meio do gateway da Internet.



O diagrama a seguir mostra como configurar uma VPC para que um cluster na VPC possa acessar recursos em sua própria rede, como um banco de dados Oracle.



Sub-redes privadas

Uma sub-rede privada permite que você inicie AWS recursos sem exigir que a sub-rede tenha um gateway de internet conectado. O Amazon EMR oferece suporte à inicialização de clusters em sub-redes privadas nas versões 4.2.0 ou posteriores.

> **Note**
>
> Ao configurar um cluster do Amazon EMR em uma sub-rede privada, recomendamos configurar também endpoints da VPC para o Amazon S3. Se o cluster do EMR estiver em uma sub-rede privada sem endpoints da VPC para o Amazon S3, você incorrerá em cobranças adicionais de gateway NAT associadas ao tráfego do S3, pois o tráfego entre o cluster do EMR e o S3 não permanecerá na VPC.

As sub-redes privadas diferem das sub-redes do seguinte modo:

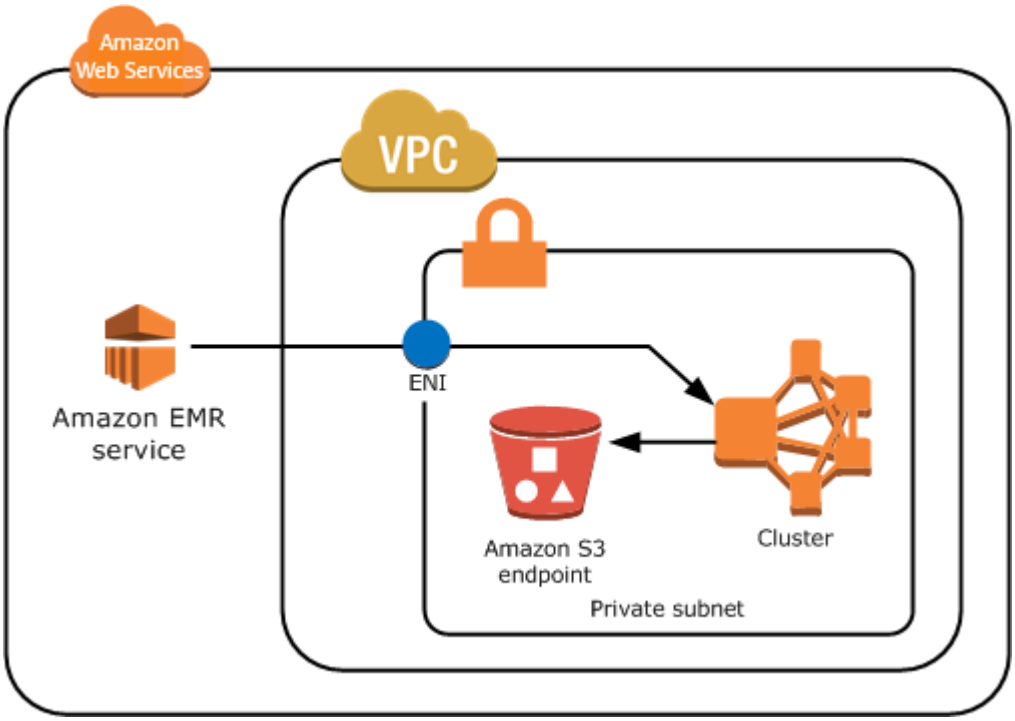
- Para acessar AWS serviços que não fornecem um VPC endpoint, você ainda precisa usar uma instância NAT ou um gateway de internet.
- Você deve fornecer pelo menos uma rota para o bucket de logs de serviço do Amazon EMR e o repositório do Amazon Linux no Amazon S3. Para mais informações, consulte Política mínima do Amazon S3 para uma sub-rede privada.

- Se você usa atributos do EMRFS, precisa ter um endpoint da VPC do Amazon S3 e uma rota da sua sub-rede privada para o DynamoDB.
- A depuração só funcionará se você fornecer uma rota da sua sub-rede privada para um endpoint do Amazon SQS público.
- A criação de uma configuração de sub-rede privada com uma instância NAT ou um gateway em uma sub-rede pública apenas tem suporte usando o AWS Management Console. A maneira mais fácil de adicionar e configurar instâncias do NAT e endpoints da VPC do Amazon S3 para clusters do Amazon EMR é usar a página Lista de sub-redes da VPC no console do Amazon EMR. Para configurar gateways NAT, consulte Gateways NAT no Guia do usuário da Amazon VPC.
- Não é possível alterar uma sub-rede com um cluster do Amazon EMR existente de pública para privada, ou vice-versa. Para localizar um cluster do Amazon EMR dentro de uma sub-rede privada, o cluster deve ser iniciado nessa sub-rede privada.

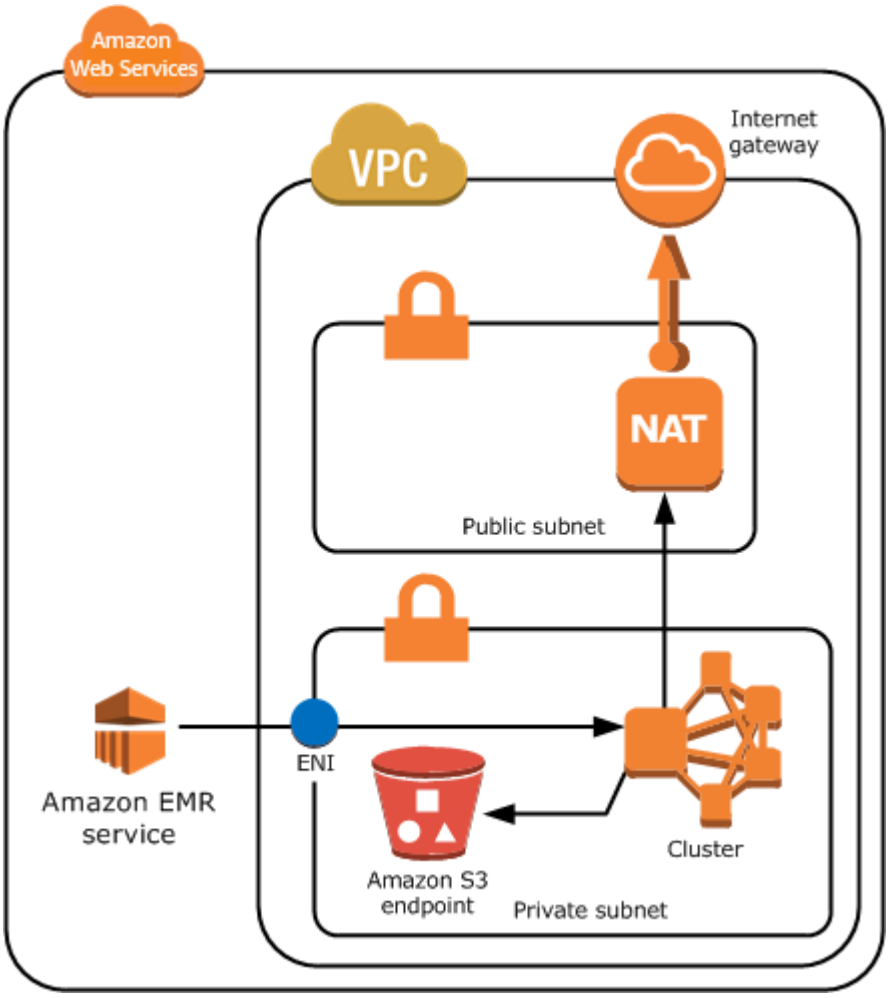
O Amazon EMR cria e usa diferentes grupos de segurança padrão para os clusters em uma sub-rede privada: ElasticMapReduce -Master-Private, Reduce-Slave-Private e -. ElasticMap ElasticMapReduce ServiceAccess Para ter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

Para obter uma listagem completa de NACLs do cluster, escolha Grupos de segurança para principais e Grupos de segurança para centrais e de tarefa na página Detalhes do cluster do console do Amazon EMR.

A imagem a seguir mostra como um cluster do Amazon EMR está configurado dentro de uma sub-rede privada. A única comunicação fora da sub-rede é com o Amazon EMR.



A imagem a seguir mostra uma configuração de exemplo para um cluster do Amazon EMR em uma sub-rede privada conectada a uma instância NAT que reside em uma sub-rede pública.



Sub-redes compartilhadas

O compartilhamento de VPC permite que os clientes compartilhem sub-redes com outras AWS contas na mesma organização. AWS É possível executar clusters do Amazon EMR nas sub-redes compartilhadas públicas e privadas, com as advertências a seguir.

O proprietário da sub-rede deve compartilhar uma sub-rede com você para que você possa executar um cluster do Amazon EMR nela. No entanto, as sub-redes compartilhadas podem deixar de ser compartilhadas posteriormente. Para obter mais informações, consulte Trabalhar com VPCs compartilhadas. Quando um cluster é iniciado em uma sub-rede compartilhada e, então, essa sub-rede deixa de ser compartilhada, é possível observar comportamentos específicos com base no estado do cluster do Amazon EMR quando a sub-rede deixa de ser compartilhada.

- A sub-rede deixa de ser compartilhada antes de o cluster ser executado com êxito. Se o proprietário para de compartilhar a Amazon VPC ou sub-rede enquanto o participante estiver

executando um cluster, o cluster pode falhar ao iniciar ou ser parcialmente inicializado sem provisionar todas as instâncias solicitadas.

- A sub-rede deixa de ser compartilhada depois de o cluster ser executado com êxito. Quando o proprietário para de compartilhar uma sub-rede ou Amazon VPC com o participante, os clusters do participante não poderão ser redimensionados para adicionar novas instâncias ou substituir instâncias com problemas de integridade.

Ao executar um cluster do Amazon EMR, são criados vários grupos de segurança. Em uma sub-rede compartilhada, o participante da sub-rede controla esses grupos de segurança. O proprietário da sub-rede pode visualizar esses grupos de segurança, mas não pode executar nenhuma ação neles. Se o proprietário da sub-rede deseja remover ou modificar o grupo de segurança, o participante que criou o grupo de segurança deve realizar a ação.

Controlar permissões da VPC com o IAM

Por padrão, todos os usuários podem ver todas as sub-redes da conta, e qualquer usuário pode iniciar um cluster em qualquer sub-rede.

Ao iniciar um cluster em uma VPC, você pode usar o AWS Identity and Access Management (IAM) para controlar o acesso aos clusters e restringir ações usando políticas, assim como faria com clusters lançados no Amazon EC2 Classic. Para obter mais informações sobre o IAM, consulte o Guia do usuário do IAM.

Você também pode usar o IAM para controlar quem pode criar e administrar sub-redes. Por exemplo, você pode criar uma conta para administrar sub-redes e uma segunda conta que pode iniciar clusters, mas não poderá modificar as configurações da Amazon VPC. Para obter mais informações sobre como administrar políticas e ações no Amazon EC2 e no Amazon VPC, consulte Políticas do IAM para o Amazon EC2 no Guia do usuário do Amazon EC2.

Configurar uma VPC para hospedar clusters

Antes de poder iniciar clusters em uma VPC, você deve criar uma VPC e uma sub-rede. Para sub-redes públicas, é necessário criar um gateway da Internet e anexá-lo à sub-rede. As instruções a seguir descrevem como criar uma VPC capaz de hospedar clusters do Amazon EMR.

Criar uma VPC com sub-redes para um cluster do Amazon EMR

1. Abra o console do Amazon VPC em https://console.aws.amazon.com/vpc/.
2. No canto superior direito da página, escolha a Região da AWS da VPC.

3. Escolha Criar VPC.
4. Na página Configurações da VPC, escolha VPC e mais.
5. Em Geração automática de etiquetas de nome, habilite Gerar automaticamente e insira um nome para a VPC. Isso ajuda você a identificar a VPC e a sub-rede no console da Amazon VPC depois que são criadas.
6. No campo Bloco CIDR IPv4, insira um espaço de endereço IP privado para a VPC para garantir a resolução adequada do nome do host DNS. Caso contrário, você poderá passar por falhas no cluster do Amazon EMR. Isso inclui os seguintes intervalos de endereços IP:
   - 10.0.0.0 - 10.255.255.255
   - 172.16.0.0 - 172.31.255.255
   - 192.168.0.0 - 192.168.255.255
7. Em Número de zonas de disponibilidade (ZAs), escolha o número de zonas de disponibilidade nas quais iniciar suas sub-redes.
8. Em Número de sub-redes públicas, escolha uma única sub-rede pública para adicionar à VPC. Se os dados usados pelo cluster estiverem disponíveis na Internet (por exemplo, no Amazon S3 ou no Amazon RDS), você só precisará usar uma sub-rede pública e não será necessário adicionar uma sub-rede privada.
9. Em Número de sub-redes privadas, escolha o número de sub-redes públicas que você deseja adicionar à sua VPC. Selecione um ou mais se os dados da aplicação estiverem armazenados em sua própria rede (por exemplo, em um banco de dados Oracle). Para uma VPC em uma sub-rede privada, todas as instâncias do Amazon EC2 devem ter pelo menos uma rota para o Amazon EMR pela interface de rede elástica. No console, isso é configurado automaticamente para você.
10. Em Gateways NAT, opcionalmente, escolha adicionar gateways NAT. Eles só são necessários se houver sub-redes privadas que precisam se comunicar com a Internet.
11. Opcionalmente, em Endpoints da VPC, escolha adicionar endpoints ao Amazon S3 para as sub-redes.
12. Verifique se as opções Habilitar nomes de host DNS e Habilitar resolução DNS estão marcadas. Para obter mais informações, consulte Como usar o DNS com sua VPC.
13. Escolha Criar VPC.
14. Uma janela de status mostra o trabalho em andamento. Quando o trabalho for concluído, escolha Visualizar VPC para navegar até a página Suas VPCs, que exibe sua VPC padrão e a

VPC que você acabou de criar. A VPC que você criou é uma VPC não padrão, portanto a coluna Default VPC exibe Não.

15. Se você quiser associar sua VPC a uma entrada de DNS que não inclua um nome de domínio, navegue até os Conjuntos de opções DHCP, escolha Criar conjunto de opções DHCP e omita um nome de domínio. Depois de criar o conjunto de opções, navegue até a nova VPC, escolha Editar conjunto de opções DHCP no menu Ações e selecione o novo conjunto de opções. Você não pode editar o nome do domínio usando o console após a criação do conjunto de opções DNS.

    É uma prática recomendada com o Hadoop e aplicativos relacionados garantir a resolução do nome de domínio totalmente qualificado (FQDN) dos nós. Para garantir uma resolução de DNS adequada, configure uma VPC que inclua um conjunto de opções de DHCP cujos parâmetros estejam definidos como os seguintes valores:

    - domain-name = **`ec2.internal`**

      Use **`ec2.internal`**, se a região for Leste dos EUA (Norte da Virgínia). Para outras regiões, use *`region-name`***`.compute.internal`**. Para exemplos em `us-west-2`, use **`us-west-2.compute.internal`**. Para a região AWS GovCloud (Oeste dos EUA), use**`us-gov-west-1.compute.internal`**.
    - domain-name-servers = **`AmazonProvidedDNS`**

    Para obter mais informações, consulte Conjuntos de opções DHCP no Guia do usuário da Amazon VPC.
16. Após a criação da VPC, acesse a página Sub-redes e anote o ID de sub-rede de uma das sub-redes da nova VPC. Você usará essas informações quando você iniciar o cluster do Amazon EMR para a VPC.

Iniciar clusters em uma VPC

Depois de ter uma sub-rede configurada para hospedar clusters do Amazon EMR, inicie o cluster nessa sub-rede especificando o identificador de sub-rede associado ao criar o cluster.

> **Note**
>
> O Amazon EMR oferece suporte a sub-redes privadas nas versões 4.2 e superiores.

Quando o cluster é iniciado, o Amazon EMR adiciona grupos de segurança conforme o tipo de sub-redes da VPC (públicas ou privadas) em que o cluster é iniciado. Todos os grupos de segurança permitem a entrada na porta 8443 para comunicação com o serviço do Amazon EMR, mas os intervalos de endereços IP variam para sub-redes públicas e privadas. O Amazon EMR gerencia todos esses grupos de segurança e pode precisar adicionar endereços IP adicionais ao AWS intervalo ao longo do tempo. Para ter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

Para gerenciar o cluster em uma VPC, o Amazon EMR anexa um dispositivo de rede ao nó primário e o administra nesse dispositivo. Você pode visualizar este dispositivo usando a ação de API do Amazon EC2 `DescribeInstances`. Se esse dispositivo for modificado de qualquer maneira, o cluster poderá falhar.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Iniciar um cluster em uma VPC usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Redes, acesse o campo Nuvem privada virtual (VPC). Insira o nome da VPC ou escolha Procurar para selecionar a VPC. Como alternativa, escolha Criar VPC para criar uma VPC que você possa usar com o cluster.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Old console

Iniciar um cluster em uma VPC usando o novo console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Escolha Go to advanced options (Ir para opções avançadas).
4. Na seção Hardware Configuration (Configuração do hardware), para Network (Rede), selecione o ID de uma rede VPC que você criou anteriormente.
5. Para EC2 Subnet (Sub-rede do EC2), selecione o ID de uma sub-rede que você criou anteriormente.

   a. Se a sua sub-rede privada estiver configurada corretamente com instâncias NAT e opções de endpoint do S3, ela exibirá (EMR Ready) (Habilitado para EMR) acima dos nomes e identificadores de sub-redes.

   b. Se a sua sub-rede privada não tiver uma instância NAT e/ou um endpoint S3, você poderá configurar isso escolhendo Add S3 endpoint and NAT instance (Adicionar endpoint do S3 e instância NAT), Add S3 endpoint (Adicionar endpoint do S3) ou Add NAT instance (Adicionar instância NAT). Selecione as opções desejadas para a instância NAT e o endpoint do S3 e escolha Configure (Configurar).

   > **Important**
   >
   > Para criar uma instância NAT a partir do Amazon EMR, você precisa de ec2CreateRoute:,`ec2:RevokeSecurityGroupEgress`,`ec2:AuthorizeSecurityGro`
   > `cloudformation:DescribeStackEvents` e permissões.
   > `cloudformation:CreateStack`

   > **Note**
   >
   > Há um custo adicional para iniciar uma instância do Amazon EC2 para o dispositivo NAT.

6. Continue com a criação do cluster.

AWS CLI

Para iniciar um cluster em uma VPC com o AWS CLI

> **Note**
>
> O AWS CLI não fornece uma maneira de criar uma instância NAT automaticamente e conectá-la à sua sub-rede privada. No entanto, para criar um endpoint do S3 na sua sub-rede, você pode usar os comandos da CLI da Amazon VPC. Use o console para criar instâncias NAT e executar clusters em uma sub-rede privada.

Depois que a VPC estiver configurada, você poderá criar clusters do Amazon EMR usando o subcomando `create-cluster` com o parâmetro `--ec2-attributes`. Use o parâmetro `--ec2-attributes` para especificar a sub-rede VPC do seu cluster.

- Para criar um cluster em uma sub-rede específica, digite o comando a seguir, substitua *myKey* pelo nome do par de chaves do Amazon EC2 e substitua *77XXXX03* pelo ID da sub-rede.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.2.0 --
applications Name=Hadoop Name=Hive Name=Pig --use-default-roles --ec2-attributes
 KeyName=myKey,SubnetId=subnet-77XXXX03 --instance-type m5.xlarge --instance-
count 3
```

  Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usam o tipo de instância especificado no comando.

  > **Note**
  >
  > Se você não tiver criado anteriormente o perfil de serviço do Amazon EMR padrão e o perfil de instância do EC2, digite `aws emr create-default-roles` para criá-los antes de digitar o subcomando `create-cluster`.

Política mínima do Amazon S3 para uma sub-rede privada

Para sub-redes privadas, você precisará ao menos fornecer a capacidade para o Amazon EMR acessar repositórios do Amazon Linux. A política de sub-rede privada faz parte das políticas de endpoint da VPC para acessar o Amazon S3. Com o Amazon EMR 5.25.0 ou versões posteriores, para habilitar o acesso com um clique ao servidor de histórico persistente do Spark, você deve permitir o acesso do Amazon EMR ao bucket do sistema que coleta logs de eventos do Spark. Se você habilitar o registro em log, forneça permissões PUT para um bucket `aws157-logs-*`. Para obter mais informações, consulte One-click access to persistent Spark History Server.

Cabe a você determinar as restrições da política que atendam às suas necessidades comerciais. Por exemplo, é possível especificar a região `packages.us-east-1.amazonaws.com` para evitar um nome ambíguo de bucket do Amazon S3. A política de exemplo a seguir fornece permissões para acessar repositórios do Amazon Linux e o bucket do sistema Amazon EMR para coleta de logs de eventos do Spark. *`MyRegion`*Substitua pela região em que seus buckets de log residem, por exemplo`us-east-1`.

Para obter mais informações sobre o uso de políticas do IAM com endpoints da Amazon VPC, consulte Endpoint policies for Amazon S3.

```
{
   "Version": "2008-10-17",
   "Statement": [
       {
           "Sid": "AmazonLinuxAMIRepositoryAccess",
           "Effect": "Allow",
           "Principal": "*",
           "Action": "s3:GetObject",
           "Resource": [
            "arn:aws:s3:::packages.MyRegion.amazonaws.com/*",
            "arn:aws:s3:::repo.MyRegion.amazonaws.com/*",
            "arn:aws:s3:::repo.MyRegion.emr.amazonaws.com/*"
           ]
       },
       {
           "Sid": "EnableApplicationHistory",
           "Effect": "Allow",
           "Principal": "*",
           "Action": [
               "s3:Put*",
               "s3:Get*",
```

```
                "s3:Create*",
                "s3:Abort*",
                "s3:List*"
            ],
            "Resource": [
             "arn:aws:s3:::prod.MyRegion.appinfo.src/*"
            ]
        }
    ]
}
```

O exemplo de política a seguir fornece as permissões necessárias para acessar repositórios do Amazon Linux 2. A AMI do Amazon Linux 2 é o padrão.

```
{
    "Statement": [
        {
            "Sid": "AmazonLinux2AMIRepositoryAccess",
            "Effect": "Allow",
            "Principal": "*",
            "Action": "s3:GetObject",
            "Resource": [
             "arn:aws:s3:::amazonlinux.MyRegion.amazonaws.com/*",
             "arn:aws:s3:::amazonlinux-2-repos-MyRegion/*"
            ]
        }
    ]
}
```

Mais recursos para saber mais sobre VPCs

Use os tópicos a seguir para saber mais sobre VPCs e sub-redes.

- Sub-redes privadas em uma VPC
  - Cenário 2: VPC com sub-redes pública e privada (NAT)
  - Instâncias NAT
  - Alta disponibilidade para instâncias NAT da Amazon VPC: um exemplo
- Sub-redes públicas em uma VPC
  - Cenário 1: VPC com uma sub-rede pública única
- Informações gerais da VPC

- Guia do usuário da Amazon VPC
- Emparelhamento de VPC
- Usar interfaces de rede elástica com sua VPC
- Conexão segura com instâncias do Linux executadas em uma VPC privada da

Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes

Quando você cria um cluster e especifica a configuração do nó primário, dos nós centrais e dos nós de tarefa, existem opções de configuração. Você pode usar frotas de instâncias ou grupos de instâncias uniformes. A opção de configuração escolhida se aplica a todos os nós e pelo tempo de vida do cluster, e frotas de instâncias e grupos de instâncias não podem coexistir em um cluster. A configuração de frotas de instâncias está disponível no Amazon EMR versão 4.8.0 e posteriores, exceto nas versões 5.0.x.

Você pode usar o console do Amazon EMR AWS CLI, o ou a API do Amazon EMR para criar clusters com qualquer configuração. Ao usar o comando `create-cluster` a partir da AWS CLI, você usar ambos os parâmetros `--instance-fleets` para criar o cluster usando frotas de instâncias ou, como alternativa, pode usar os parâmetros `--instance-groups` para criá-los usando grupos de instâncias uniformes.

O mesmo é válido com o uso da API do Amazon EMR. Você usa também a configuração `InstanceGroups` para especificar uma matriz de objetos `InstanceGroupConfig` ou usa a configuração `InstanceFleets` para especificar uma matriz de objetos `InstanceFleetConfig`.

No novo console do Amazon EMR, é possível escolher usar grupos de instâncias ou frotas de instâncias ao criar um cluster, e você pode de usar instâncias spot com cada opção. No console antigo do Amazon EMR, se você usar as configurações de Opções rápidas ao criar um cluster, o Amazon EMR aplicará a configuração de grupos de instâncias uniformes a esse cluster e usará instâncias sob demanda. Para instâncias spot com grupos de instâncias uniformes ou configurar frotas de instâncias e fazer outras personalizações, escolha Advanced Options (Opções avançadas).

Frotas de instâncias

A configuração de frotas de instâncias oferece a mais ampla variedade de opções de provisionamento para instâncias do Amazon EC2. Cada tipo de nó tem uma única frota de instâncias, e a frota de instâncias de tarefa é opcional. Você pode especificar até cinco tipos de instância EC2 por frota ou 30 tipos de instância EC2 por frota ao criar um cluster usando a API do AWS CLI Amazon EMR e uma estratégia de alocação para instâncias sob demanda e spot. Para as frotas

de instâncias centrais e de tarefa, você atribui uma capacidade de destino para instâncias sob demanda e outra para instâncias spot. O Amazon EMR escolhe qualquer combinação dos tipos de instâncias especificados para preencher as capacidades de destino, provisionando tanto instâncias sob demanda como instâncias spot.

Para o tipo de nó primário, o Amazon EMR escolhe um único tipo de instância da lista de instâncias, e você especifica se esse tipo é configurado como uma instância sob demanda ou spot. As frotas de instâncias também oferecem outras opções para compras de instâncias spot e sob demanda. As opções de instância spot incluem um tempo limite que especifica uma ação a ser tomada, caso não seja possível provisionar a capacidade spot, e uma estratégia de alocação preferencial (otimizada para capacidade) para iniciar frotas de instâncias spot. Também é possível iniciar frotas de instâncias sob demanda usando a opção de estratégia de alocação (menor preço). Se você usar um perfil de serviço que não seja o perfil de serviço padrão do EMR ou usar uma política gerenciada do EMR no perfil de serviço, será necessário adicionar outras permissões ao perfil de serviço de cluster personalizado para habilitar a opção de estratégia de alocação. Para ter mais informações, consulte Perfil de serviço para Amazon EMR (perfil do EMR).

Para obter mais informações sobre como configurar frotas de instâncias, consulte Configurar frotas de instâncias.

Grupos de instâncias uniformes

Os grupos de instâncias uniformes oferecem uma configuração mais simples do que as frotas de instâncias. Cada cluster do Amazon EMR pode ter até 50 grupos de instâncias: um grupo de instâncias primário, que contém uma única instância do Amazon EC2, um grupo de instâncias centrais, que contém uma ou mais instâncias do EC2, e até 48 grupos de instâncias de tarefa opcionais. Cada grupo de instâncias central e de tarefa pode conter qualquer número de instâncias do Amazon EC2. Você pode escalar cada grupo de instâncias adicionando e removendo instâncias do Amazon EC2 manualmente ou pode configurar o ajuste de escala automático. Para obter informações sobre como adicionar e remover instâncias, consulte Usar ajuste de escala de clusters.

Para obter mais informações sobre como configurar grupos de instâncias uniformes, consulte Configurar grupos de instâncias uniformes.

Trabalhar com frotas de instâncias e grupos de instâncias

Tópicos

- Configurar frotas de instâncias
- Usar reservas de capacidade com a frotas de instância

- Configurar grupos de instâncias uniformes
- Práticas recomendadas para flexibilidade de instâncias e de zona de disponibilidade
- Práticas recomendadas para configuração de clusters

Configurar frotas de instâncias

> **Note**
>
> A configuração de frotas de instância só está disponível em versões do Amazon EMR 4.8.0 e posteriores, exceto versões 5.0.0 e 5.0.3.

A configuração da frota de instâncias para clusters do Amazon EMR permite selecionar uma grande variedade de opções de provisionamento para instâncias do Amazon EC2 e ajuda a desenvolver uma estratégia de recursos flexível e elástica para cada tipo de nó do cluster.

Em uma configuração de frota de instância, especifique uma capacidade de destino para instâncias sob demanda e instâncias spot em cada frota. Quando o cluster é iniciado, o Amazon EMR provisiona instâncias até que os destinos sejam atendidos. Quando o Amazon EC2 recupera uma instância spot em um cluster em execução por causa de um aumento de preço ou falha de instância, o Amazon EMR tenta substituir a instância por qualquer um dos tipos de instância especificados. Isso facilita recuperar a capacidade durante um pico nos preços Spot.

Você pode especificar no máximo cinco tipos de instância do Amazon EC2 por frota para o Amazon EMR usar ao cumprir as metas, ou no máximo 30 tipos de instância do Amazon EC2 por frota ao criar um cluster usando a API do AWS CLI Amazon EMR e uma estratégia de alocação para instâncias sob demanda e spot.

Você também pode selecionar várias sub-redes em diferentes zonas de disponibilidade. Quando o Amazon EMR executa o cluster, ele procura entre as sub-redes para encontrar as instâncias e opções de compra que você especificar. Se o Amazon EMR detectar um evento de AWS grande escala em uma ou mais zonas de disponibilidade, o Amazon EMR tentará automaticamente direcionar o tráfego para fora das zonas de disponibilidade afetadas e tentará lançar novos clusters que você cria em zonas de disponibilidade alternativas de acordo com suas seleções. A seleção da zona de disponibilidade do cluster ocorre somente na criação do cluster. Os nós de cluster já existentes não são reiniciados automaticamente em uma nova zona de disponibilidade em caso de interrupção na zona de disponibilidade.

Considerações

Considere os itens a seguir ao usar as frotas de instâncias com o Amazon EMR.

- Você pode ter apenas uma frota de instância por tipo de nó (primário, central, de tarefa). Você pode especificar até cinco tipos de instância do Amazon EC2 para cada frota no AWS Management Console (ou um máximo de 30 tipos por frota de instância ao criar um cluster usando a API do AWS CLI Amazon EMR e uma). Estratégia de alocação para frotas de instâncias
- O Amazon EMR escolhe qualquer um ou todos os tipos de instâncias do Amazon EC2 especificados para provisionar com opções de compra spot e sob demanda.
- Você pode estabelecer capacidades de destino para instâncias spot e sob demanda para frota central e frota de tarefa. Use vCPU ou uma unidade genérica atribuída a cada instância do Amazon EC2 que é considerada para os destinos. O Amazon EMR provisiona instâncias até que cada capacidade de destino seja totalmente preenchida. Para a frota primária, o destino é sempre um.
- Você pode escolher uma sub-rede (zona de disponibilidade) ou um intervalo. Se você escolher um intervalo, o Amazon EMR provisionará a capacidade na zona de disponibilidade mais apropriada.
- Quando você especificar uma capacidade alvo para instâncias Spot:
  - Para cada tipo de instância, especifique um preço spot máximo. O Amazon EMR provisionará instâncias spot se o preço estiver abaixo do preço spot máximo. Você paga o preço spot e não necessariamente o preço spot máximo.
  - Para cada frota, defina um tempo limite para o provisionamento de instâncias Spot. Se o Amazon EMR não puder provisionar a capacidade spot, você pode terminar o cluster ou alternar para capacidade sob demanda de provisionamento. Isso se aplica somente ao provisionamento de clusters, não ao redimensionamento deles. Se o período de tempo limite terminar durante o processo de redimensionamento do cluster, as solicitações spot não provisionadas serão anuladas sem serem transferidas para capacidade sob demanda.
- Para cada frota, é possível especificar uma das seguintes estratégias de alocação para instâncias spot: otimizada para preço-capacidade, otimizada para capacidade, menor preço ou diversificada em todos os grupos.
- Para cada frota, é possível aplicar a estratégia de alocação de menor preço para instâncias sob demanda; não é possível personalizar a estratégia de alocação para instâncias sob demanda.
- Para cada frota com `allocation strategy - lowest-price` sob demanda, você pode optar por aplicar as opções de reserva de capacidade.

- Verifique o tamanho da sub-rede antes de iniciar o cluster. Quando você provisiona um cluster com uma frota de tarefa e não há endereços IP suficientes disponíveis na sub-rede correspondente, a frota entra em estado suspenso em vez de terminar o cluster com um erro. Para evitar esse problema, recomenda-se aumentar o número de endereços IP das sub-redes.

Opções de frotas de instâncias

Use as seguintes diretrizes para compreender as opções de frota de instância.

Tópicos

Definir capacidades de destino

Especifique as capacidades alvo que deseja para a frota de núcleo e de tarefa. Quando você fizer isso, ele determina o número de instâncias sob demanda e instâncias spot que o Amazon EMR provisiona. Quando você especifica uma instância, você decide o quanto cada instância é considerada para o destino. Quando uma instância sob demanda é provisionada, ela é considerada para o destino sob demanda. O mesmo aplica-se para instâncias spot. Ao contrário de frotas centrais e de tarefa, a frota primária é sempre uma instância. Portanto, a capacidade de destino desta frota é sempre um.

Ao usar o console, as vCPUs do tipo de instância do Amazon EC2 são usadas como a contagem para capacidades de destino por padrão. Você pode alterar isso para Unidades genéricas e, em seguida, especificar a contagem para cada tipo de instância do EC2. Ao usar o AWS CLI, você atribui manualmente unidades genéricas para cada tipo de instância.

> **Important**
>
> Quando você escolhe um tipo de instância usando o AWS Management Console, o número de vCPUs mostrado para cada tipo de instância é o número de vcores YARN para esse tipo de instância, não o número de vCPUs EC2 para esse tipo de instância. Para obter mais informações sobre o número de vCPUs para o seu tipo de instância, consulte os tipos de instância do Amazon EC2.

Até cinco tipos de instância do Amazon EC2 especificados para cada frota. Se você usa um Estratégia de alocação para frotas de instâncias e cria um cluster usando a API do Amazon EMR AWS CLI ou a API do Amazon EMR, você pode especificar até 30 tipos de instância EC2 por frota de instâncias. O Amazon EMR escolhe qualquer combinação desses tipos de instância do EC2 para preencher as capacidades de destino. Como o Amazon EMR deseja atender a capacidade de destino completamente, pode haver um excedente. Por exemplo, se houver duas unidades não atendidas e o Amazon EMR puder apenas provisionar uma instância com uma contagem de cinco unidades, a instância ainda é provisionada, o que significa que a capacidade de destino será excedida por três unidades.

Se você reduzir a capacidade pretendida para redimensionar um cluster em execução, o Amazon EMR tentará concluir as tarefas de aplicação e encerrará as instâncias para atender o novo destino. Para ter mais informações, consulte Terminar na conclusão de tarefas.

Opções de inicialização

Para instâncias spot, você pode especificar Preço spot máximo para cada tipo de instância da frota. Você pode definir esse preço como uma porcentagem do preço sob demanda ou como uma quantia em dólar. O Amazon EMR provisiona instâncias spot caso o preço spot atual em uma zona de disponibilidade esteja abaixo do preço spot máximo. Você paga o preço spot e não necessariamente o preço spot máximo.

> **Note**
>
> As instâncias spot com duração definida (também conhecidas como blocos spot) não estarão mais disponíveis para novos clientes a partir de 1.º de julho de 2021. Aos clientes que utilizaram o recurso anteriormente, continuaremos a oferecer suporte a instâncias spot com duração definida até 31 de dezembro de 2022.

Disponível no Amazon EMR 5.12.1 e versões posteriores, você tem a opção de iniciar frotas de instâncias spot e sob demanda com alocação de capacidade otimizada. Essa opção de estratégia de alocação pode ser definida na antiga AWS Management Console ou usando a API`RunJobFlow`. Não é possível personalizar a estratégia de alocação no novo console. Usar a opção de estratégia de alocação requer outras permissões de perfil de serviço. Se você usar o perfil de serviço padrão do Amazon EMR e a política gerenciada (`EMR_DefaultRole` e `AmazonEMRServicePolicy_v2`) para o cluster, as permissões para a opção de estratégia de alocação já estarão incluídas. Caso

não esteja usando o perfil de serviço e a política gerenciada padrão do Amazon EMR, você deverá adicioná-las para usar essa opção. Consulte Perfil de serviço para Amazon EMR (perfil do EMR).

Para obter mais informações sobre instâncias spot, consulte Instâncias spot no Guia do usuário do Amazon EC2. Para obter mais informações sobre instâncias sob demanda, consulte Instâncias sob demanda no Guia do usuário do Amazon EC2.

Ao escolher iniciar frotas de instâncias sob demanda com a estratégia de alocação de menor preço, você terá a opção de usar reservas de capacidade. As opções de reserva de capacidade podem ser definidas usando a API `RunJobFlow` do Amazon EMR. As reservas de capacidade exigem outras permissões de perfil de serviço que você deve adicionar para usar essas opções. Consulte Permissões da estratégia de alocação. Não é possível personalizar as reservas de capacidade no novo console.

Várias opções de sub-rede (zonas de disponibilidade)

Ao usar frotas de instância, você pode especificar várias sub-redes do Amazon EC2 em uma VPC, cada uma correspondendo a uma zona de disponibilidade diferente. Se você usar o EC2-Classic, especifique zonas de disponibilidade explicitamente. O Amazon EMR identifica a melhor zona de disponibilidade para iniciar instâncias de acordo com suas especificações de frota. Instâncias são sempre provisionadas em apenas uma Zona de disponibilidade. Você pode selecionar sub-redes privadas ou públicas, mas não pode combinar as duas. Além disso, as sub-redes que você especificar devem estar na mesma VPC.

Configuração do nó principal

Como a frota de instância primária é somente uma única instância, sua configuração é um pouco diferente de frotas de instâncias centrais e de tarefa. Você seleciona apenas sob demanda ou spot para a frota de instâncias primária, pois ela é formada por somente uma instância. Se você usar o console para criar a frota de instâncias, a capacidade alvo para a opção de compra que você selecionar será definida como 1. Se você usar o AWS CLI, sempre `TargetOnDemandCapacity` defina um `TargetSpotCapacity` ou como 1, conforme apropriado. Ainda é possível escolher até cinco tipos de instância para a frota de instâncias primárias (ou no máximo 30 ao usar a opção de estratégia de alocação para instâncias sob demanda ou spot). No entanto, ao contrário de frotas de instâncias centrais e de tarefa, nas quais o Amazon EMR pode provisionar várias instâncias de tipos diferentes, o Amazon EMR seleciona um único tipo de instância a ser provisionado para a frota de instâncias primária.

Estratégia de alocação para frotas de instâncias

Com o Amazon EMR versões 5.12.1 e posteriores, você pode usar a opção de estratégia de alocação com instâncias sob demanda e spot para cada nó do cluster. Ao criar um cluster usando a AWS CLI, a API do Amazon EMR ou o console do Amazon EMR com uma estratégia de alocação, você pode especificar até 30 tipos de instância do Amazon EC2 por frota. Com a configuração padrão da frota de instâncias de cluster do Amazon EMR, é possível ter até cinco tipos de instância por frota. É recomendável usar a opção de estratégia de alocação para obter provisionamento mais rápido do cluster, alocação mais precisa de instâncias spot e menos interrupções de instâncias spot.

Tópicos

Estratégia de alocação para instâncias sob demanda

Quando você usa a estratégia de alocação, as instâncias sob demanda usam a estratégia de menor preço. Assim, as instâncias de menor preço são iniciadas primeiro. Ao iniciar instâncias sob demanda, você pode usar reservas de capacidade abertas ou direcionadas em suas contas. É possível usar reservas de capacidade aberta para nós primários, centrais e de tarefa. Você pode ter capacidade insuficiente com instâncias sob demanda com estratégia de alocação para frotas de instâncias. É recomendável especificar um número maior de tipos de instância para diversificar e reduzir a chance de ter capacidade insuficiente. Para ter mais informações, consulte Usar reservas de capacidade com a frotas de instância.

Estratégia de alocação com instâncias spot

Em Instâncias spot, você escolher uma destas estratégias de alocação:

`price-capacity-optimized` (recomendado)

A estratégia de alocação otimizada para preço-capacidade inicia instâncias spot com base nos grupos de instâncias spot que têm a maior capacidade disponível e o menor preço para o número de instâncias que estão sendo iniciadas. Como resultado, a estratégia otimizada para preço-capacidade normalmente tem uma chance maior de obter capacidade spot e oferece taxas de interrupção mais baixas.

`capacity-optimized`

A estratégia de alocação otimizada para capacidade inicia instâncias spot nos grupos mais disponíveis com a menor chance de interrupção no curto prazo. Essa é uma boa opção para workloads que podem ter um custo maior de interrupção associado ao trabalho que é reiniciado. Essa é a estratégia padrão para as versões 6.9.0 e anteriores do Amazon EMR.

`diversified`

Com a estratégia de alocação diversificada, o Amazon EC2 distribui instâncias spot em todos os grupos de capacidade spot.

`lowest-price`

A estratégia de alocação de menor preço inicia instâncias spot pelo grupo de menor preço que tenha capacidade disponível. Se o grupo com menor preço não tiver capacidade disponível, as instâncias spot virão do próximo grupo com menor preço que tiver capacidade disponível. Se um grupo esgotar capacidade antes de atender à capacidade solicitada, a frota do Amazon EC2 utilizará o próximo grupo com preço mais baixo para continuar atendendo à solicitação. Para garantir que a capacidade desejada seja atendida, é possível receber instâncias spot de vários grupos. Como essa estratégia considera apenas o preço da instância e não considera a disponibilidade de capacidade, ela pode resultar em altas taxas de interrupção.

Permissões da estratégia de alocação

A opção de estratégia de alocação requer várias permissões do IAM que são incluídas automaticamente no perfil de serviço padrão do Amazon EMR e na política gerenciada do Amazon EMR (`EMR_DefaultRole` e `AmazonEMRServicePolicy_v2`). Ao usar um perfil de serviço personalizado ou uma política gerenciada para o cluster, você deverá adicionar essas permissões antes de criar o cluster. Para ter mais informações, consulte Permissões da estratégia de alocação.

As reservas de capacidade sob demanda (ODCRs) opcionais estão disponíveis quando você usa a opção de estratégia de alocação sob demanda. Com as opções de reserva de capacidade, você pode especificar uma preferência para usar primeiro a capacidade reservada para clusters do Amazon EMR. Use-as para garantir que suas workloads críticas utilizarão a capacidade que você já reservou usando ODCRs abertos ou direcionados. Para workloads não essenciais, as preferências de reserva de capacidade permitem especificar se a capacidade reservada deverá ser consumida.

As reservas de capacidade só podem ser usadas por instâncias que correspondam a seus atributos (tipo de instância, plataforma e zona de disponibilidade). Por padrão, o Amazon EMR usa automaticamente as reservas de capacidade aberta ao provisionar instâncias sob demanda que

correspondam aos atributos da instância. Se você não tiver nenhuma instância em execução que corresponda aos atributos das reservas de capacidade, elas permanecerão não utilizadas até você iniciar uma instância com atributos correspondentes. Se você não quiser usar nenhuma reserva de capacidade ao iniciar o cluster, defina a preferência de reserva de capacidade como nenhuma nas opções de inicialização.

No entanto, também é possível destinar uma reserva de capacidade para workloads específicas. Isso permite que você controle explicitamente quais instâncias têm permissão para executar na capacidade reservada. Para obter mais informações sobre reservas de capacidade sob demanda, consulte Usar reservas de capacidade com a frotas de instância.

Permissões do IAM necessárias para uma estratégia de alocação

O Perfil de serviço para Amazon EMR (perfil do EMR) precisa de outras permissões para criar um cluster que use a opção de estratégia de alocação para frotas de instâncias sob demanda ou spot.

Incluímos automaticamente essas permissões no perfil de serviço padrão do Amazon EMR `EMR_DefaultRole` e na política gerenciada do Amazon EMR `AmazonEMRServicePolicy_v2`.

Ao usar um perfil de serviço personalizado ou uma política gerenciada para o cluster, você deverá adicionar as seguintes permissões:

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Effect": "Allow",
   "Action": [
    "ec2:DeleteLaunchTemplate",
    "ec2:CreateLaunchTemplate",
    "ec2:DescribeLaunchTemplates",
    "ec2:CreateLaunchTemplateVersion",
    "ec2:CreateFleet"
   ],
   "Resource": "*"
  }
}
```

As permissões de perfil de serviço a seguir são necessárias para criar um cluster que usa reservas de capacidade abertas ou direcionadas. É necessário incluir essas permissões além das permissões necessárias para usar a opção de estratégia de alocação.

Example Documento de política para reservas de capacidade de perfil de serviço

Para usar reservas de capacidade aberta, é necessário incluir as permissões adicionais a seguir.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeCapacityReservations",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DeleteLaunchTemplateVersions"
            ],
            "Resource": "*"
        }
    ]
}
```

Example

Para usar reservas de capacidade direcionada, é necessário incluir as permissões adicionais a seguir.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeCapacityReservations",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DeleteLaunchTemplateVersions",
                "resource-groups:ListGroupResources"
            ],
            "Resource": "*"
        }
    ]
}
```

Configurar frotas de instâncias para o cluster

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Criar um cluster com frotas de instâncias usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e Criar cluster.
3. Em Configuração do cluster, escolha Frotas de instâncias.
4. Em cada Grupo de nós, selecione Adicionar tipo de instância e escolha até cinco tipos de instância para frotas de instâncias primárias e centrais e até quinze tipos de instância para frotas de instâncias de tarefa. O Amazon EMR poderá provisionar qualquer combinação desses tipos de instância ao executar o cluster.
5. Para alterar essas configurações, em cada tipo de grupo de nós, escolha o menu suspenso Ações ao lado de cada instância:

   Adicionar volumes do EBS

   Especifique os volumes do EBS a serem anexados ao tipo de instância após o provisionamento do Amazon EMR.

   Editar capacidade ponderada

   Para o grupo de nós centrais, altere esse valor para qualquer número de unidades adequado a suas aplicações. O número de vCores do YARN para cada tipo de instância de frota será usado como as unidade padrão de capacidade ponderada. Não é possível editar a capacidade ponderada do nó primário.

   Editar preço máximo spot

   Especifique um preço spot máximo para cada tipo de instância da frota. Você pode definir esse preço como uma porcentagem do preço sob demanda ou como uma quantia em

dólar. Caso o preço spot atual em uma zona de disponibilidade esteja abaixo do preço spot máximo, o Amazon EMR provisiona instâncias spot. Você paga o preço spot e não necessariamente o preço spot máximo.

6. Opcionalmente, para adicionar grupos de segurança aos nós, expanda Grupos de segurança do EC2 (firewall) na seção Redes e selecione o grupo de segurança para cada tipo de nó.
7. Opcionalmente, marque a caixa de seleção ao lado de Aplicar estratégia de alocação, se quiser usar a opção de estratégia de alocação, e selecione a estratégia de alocação que deseja especificar para as instâncias spot. Não selecione essa opção se seu perfil de serviço do Amazon EMR não tiver as permissões necessárias. Para ter mais informações, consulte Estratégia de alocação para frotas de instâncias.
8. Escolha qualquer outra opção que se aplique ao cluster.
9. Para iniciar o cluster, escolha Criar cluster.

Old console

Criar um cluster com frotas de instâncias usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Na parte superior da janela do console, escolha Ir para opções avançadas, insira as opções de Configuração de software e escolha Próximo.
4. Em Composição do cluster, escolha Frotas de instâncias. Ao selecionar a opção de frotas de instâncias, você deverá ver as opções para especificar a Capacidade de destino de instâncias sob demanda e spot na tabela Nós e instâncias do cluster.
5. Para Network (Rede), insira um valor. Se você escolher uma VPC para Rede, escolha uma única Sub-rede do EC2 ou clique com a tecla CTRL pressionada para escolher mais de uma sub-rede do Amazon EC2. As sub-redes selecionadas devem ser do mesmo tipo (públicas ou privadas). Se você escolher apenas uma, seu cluster será executado nessa sub-rede. Se você escolher um grupo, a sub-rede com o melhor ajuste será selecionada no grupo quando o cluster for executado.

> **Note**
>
> Sua conta e região podem lhe dar a opção de escolher Iniciar no EC2-Classic para Rede. Se você escolher essa opção, escolha uma ou mais EC2 Availability Zones (Zonas de disponibilidade do EC2) em vez de EC2 Subnets (Sub-redes do EC2). Para obter mais informações, consulte Amazon EC2 e Amazon VPC no Guia do usuário do Amazon EC2.

6. Em Estratégia de alocação, marque a caixa de seleção para aplicar estratégias de alocação, se quiser usar a opção de estratégia de alocação. Para ter mais informações, consulte Estratégia de alocação para frotas de instâncias.
7. Em cada Tipo de nó, se você quiser alterar o nome padrão de uma frota de instâncias, escolha o ícone de lápis e insira um nome amigável. Para remover a frota de instâncias de Tarefa, escolha o ícone X no lado direito da linha Tarefa.
8. Escolha Adicionar/remover tipos de instância para frota e escolha até cinco tipos de instância da lista para frotas de instâncias primárias e centrais; adicione até quinze tipos de instância para frotas de instâncias de tarefa. O Amazon EMR pode optar por provisionar qualquer combinação desses tipos de instância ao iniciar o cluster.
9. Para cada tipo de instância central e de tarefa, escolha como deseja definir a capacidade ponderada (cada instância conta como X unidades) daquela instância. O número de YARN vCores para cada tipo de instância de frota é usado como as unidades de capacidade ponderada padrão, mas é possível alterar o valor para qualquer unidade conveniente para suas aplicações.
10. Em Capacidade de destino, defina o número total de instâncias sob demanda e spot desejadas por frota. O EMR garante que as instâncias da frota atendam às unidades solicitadas para a capacidade sob demanda e spot de destino. Se nenhuma unidade sob demanda ou spot estiver especificada para uma frota, nenhuma capacidade será provisionada para a frota.
11. Se a frota estiver configurada com a capacidade de destino para spot, você poderá inserir seu preço spot máximo como um % de preço sob demanda ou poderá inserir um valor em dólares ($) em USD.
12. Para ter volumes do EBS associados ao tipo de instância quando ela for provisionada, escolha o ícone de lápis ao lado de Armazenamento do EBS e insira as opções de configuração do EBS.

13. Se você estabeleceu uma contagem instantânea para Unidades spot, defina Opções spot avançadas de acordo com as seguintes diretrizes:

    - Tempo limite de provisionamento: use essas configurações para controlar o que o Amazon EMR faz quando não consegue provisionar instâncias spot entre os valores que você especificou para Tipos de instância de frota. Você insere um tempo limite em minutos e escolhe Terminate the cluster (Encerrar o cluster) ou Switch to provisioning On-Demand Instances (Alternar para o provisionamento de instâncias sob demanda). Se você optar por mudar para instâncias sob demanda, a capacidade atribuída dessas instâncias sob demanda contará para a capacidade de destino de instâncias spot, e o Amazon EMR provisionará instâncias Sob demanda até que essa capacidade alvo para instâncias Spot seja preenchida.
14. Escolha Próximo, modifique outras configurações de cluster e escolha Próximo.
15. Se você optou por aplicar a nova opção de estratégia de alocação, nas configurações de Opções de segurança, selecione um perfil o EMR e um perfil de instância do EC2 que contenham as permissões necessárias para a opção de estratégia de alocação. Caso contrário, a criação do cluster falhará.
16. Selecione Create Cluster (Criar cluster).

AWS CLI

Para criar e executar um cluster com frotas de instâncias com o AWS CLI, siga estas diretrizes:

- Para criar e executar um cluster com frotas de instâncias, use o comando `create-cluster` com parâmetros `--instance-fleet`.
- Para obter detalhes sobre a configuração das frotas de instâncias em um cluster, use o comando `list-instance-fleets`.
- Para adicionar várias AMIs personalizadas do Amazon Linux a um cluster que você está criando, use a opção `CustomAmiId` com cada especificação `InstanceType`. Você pode configurar nós de frota de instâncias com múltiplos tipos de instância e múltiplas AMIs personalizadas para atender a suas necessidades. Consulte Exemplos: criar um cluster com a configuração de frotas de instâncias.
- Para fazer alterações na capacidade alvo para uma frota de instâncias, use o comando `modify-instance-fleet`.
- Para adicionar uma frota de instância de tarefas a um cluster que ainda não tem uma, use o comando `add-instance-fleet`.

- Várias AMIs personalizadas podem ser adicionadas à frota de instâncias de tarefas usando o CustomAmiId argumento com o add-instance-fleet comando. Consulte Exemplos: criar um cluster com a configuração de frotas de instâncias.
- Para usar a opção de estratégia de alocação ao criar uma frota de instâncias, atualize o perfil de serviço de modo a incluir o exemplo de documento de política na seção a seguir.
- Para usar as opções de reservas de capacidade ao criar uma frota de instâncias com a estratégia de alocação sob demanda, atualize o perfil de serviço de modo a incluir o exemplo de documento de política na seção a seguir.
- As frotas de instâncias são incluídas automaticamente no perfil de serviço padrão do EMR e na política gerenciada do Amazon EMR (`EMR_DefaultRole` e `AmazonEMRServicePolicy_v2`). Ao usar um perfil de serviço personalizada ou uma política gerenciada personalizada para o cluster, você deverá adicionar as novas permissões para a estratégia de alocação na seção a seguir.

Exemplos: criar um cluster com a configuração de frotas de instâncias

Os exemplos a seguir demonstram comandos `create-cluster` com uma variedade de opções que você pode combinar.

> **Note**
>
> Se você não criou o perfil de serviço do Amazon EMR padrão e o perfil de instância do EC2, use `aws emr create-default-roles` para criá-los antes de usar o comando `create-cluster`.

Example Exemplo: primária sob demanda, central sob demanda com tipo de instância única, VPC padrão

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge}
```

Example Exemplo: principal spot, central spot com tipo de instância única, VPC padrão

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetSpotCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}'] \
    InstanceFleetType=CORE,TargetSpotCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}']
```

Example Exemplo: primária sob demanda, central mista com tipo de instância única, sub-rede do EC2 única

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-ab12345c'] \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge}'] \
    InstanceFleetType=CORE,TargetOnDemandCapacity=2,TargetSpotCapacity=6,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=2}']
```

Example Exemplo: primária sob demanda, central spot com múltiplos tipos de instâncias ponderadas, tempo limite para spot, intervalo de sub-redes do EC2

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-
ab12345c','subnet-de67890f'] \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge}'] \
    InstanceFleetType=CORE,TargetSpotCapacity=11,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}',\
'{InstanceType=m4.2xlarge,BidPrice=0.9,WeightedCapacity=5}'],\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=120,TimeoutAction=SWITCH_TO_ON
```

Example Exemplo: primário sob demanda, central misto e de tarefa com múltiplos tipos de instâncias ponderadas, tempo limite para instâncias spot centrais, intervalo de sub-redes do EC2

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-
ab12345c','subnet-de67890f'] \
```

```
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 \
    InstanceFleetType=CORE,TargetOnDemandCapacity=8,TargetSpotCapacity=6,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}',\
'{InstanceType=m4.2xlarge,BidPrice=0.9,WeightedCapacity=5}'],\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=120,TimeoutAction=SWITCH_TO_ON
 \
    InstanceFleetType=TASK,TargetOnDemandCapacity=3,TargetSpotCapacity=3,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}']
```

Example Exemplo: primária spot, que não é central nem de tarefa, configuração do Amazon EBS, VPC padrão

```
aws emr create-cluster --release-label Amazon EMR 5.3.1 --service-role EMR_DefaultRole
 \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetSpotCapacity=1,\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=60,TimeoutAction=TERMINATE_CLU
\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,\
EbsConfiguration={EbsOptimized=true,EbsBlockDeviceConfigs=[{VolumeSpecification={VolumeType=gp2
\
SizeIn GB=100}},{VolumeSpecification={VolumeType=io1,SizeInGB=100,Iop
 s=100},VolumesPerInstance=4}]}}']
```

Example Exemplo: múltiplas AMIs personalizadas, múltiplos tipos de instância, primária sob demanda, central sob demanda

```
aws emr create-cluster --release-label Amazon EMR 5.3.1 --service-role EMR_DefaultRole
 \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456},
{InstanceType=m6g.xlarge, CustomAmiId=ami-234567}'] \
    InstanceFleetType=CORE,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456},
{InstanceType=m6g.xlarge, CustomAmiId=ami-234567}']
```

Example Exemplo: adicionar um nó de tarefa a um cluster em execução com múltiplos tipos de instância e múltiplas AMIs personalizadas

```
aws emr add-instance-fleet --cluster-id j-123456 --release-label Amazon EMR 5.3.1 \
  --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleet \
    InstanceFleetType=Task,TargetSpotCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456}',\
'{InstanceType=m6g.xlarge,CustomAmiId=ami-234567}']
```

Example Exemplo: usar um arquivo de configuração JSON

Você pode configurar parâmetros de frota de instância em um arquivo JSON e fazer referência a esse arquivo JSON como o único parâmetro para frotas de instâncias. Por exemplo, o seguinte comando referencia um arquivo de configuração JSON, *my-fleet-config.json*:

```
aws emr create-cluster --release-label emr-5.30.0 --service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
--instance-fleets file://my-fleet-config.json
```

O arquivo *my-fleet-config.json* especifica frotas de instâncias primárias, centrais e de tarefa, como mostra o exemplo a seguir. A frota de instâncias principais usa um preço spot máximo (`BidPrice`) como uma porcentagem do sob demanda, enquanto as frotas de tarefas e instâncias primárias usam um preço spot máximo (BidPriceAsPercentageofOnDemandPrice) como uma string em USD.

```
[
    {
        "Name": "Masterfleet",
        "InstanceFleetType": "MASTER",
        "TargetSpotCapacity": 1,
        "LaunchSpecifications": {
            "SpotSpecification": {
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "SWITCH_TO_ON_DEMAND"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
```

```
                "BidPrice": "0.89"
            }
        ]
    },
    {
        "Name": "Corefleet",
        "InstanceFleetType": "CORE",
        "TargetSpotCapacity": 1,
        "TargetOnDemandCapacity": 1,
        "LaunchSpecifications": {
          "OnDemandSpecification": {
            "AllocationStrategy": "lowest-price",
            "CapacityReservationOptions":
            {
                "UsageStrategy": "use-capacity-reservations-first",
                "CapacityReservationResourceGroupArn": "String"
            }
        },
            "SpotSpecification": {
                "AllocationStrategy": "capacity-optimized",
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "TERMINATE_CLUSTER"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
                "BidPriceAsPercentageOfOnDemandPrice": 100
            }
        ]
    },
    {
        "Name": "Taskfleet",
        "InstanceFleetType": "TASK",
        "TargetSpotCapacity": 1,
        "LaunchSpecifications": {
          "OnDemandSpecification": {
            "AllocationStrategy": "lowest-price",
            "CapacityReservationOptions":
            {
                "CapacityReservationPreference": "none"
            }
        },
            "SpotSpecification": {
```

```
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "TERMINATE_CLUSTER"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
                "BidPrice": "0.89"
            }
        ]
    }
]
```

Modificar capacidades de destino para uma frota de instâncias

Use o comando `modify-instance-fleet` para especificar novas capacidades alvo para uma frota de instâncias. Você deve especificar o ID de cluster e o ID de frota de instância. Use o comando `list-instance-fleets` para recuperar IDs de frotas de instâncias.

```
aws emr modify-instance-fleet --cluster-id <cluster-id> \
  --instance-fleet \
    InstanceFleetId='<instance-fleet-id>',TargetOnDemandCapacity=1,TargetSpotCapacity=1
```

Adicionar uma frota de instâncias de tarefa a um cluster

Se um cluster tiver apenas frotas de instâncias primárias e centrais, você poderá usar o comando `add-instance-fleet` para adicionar uma frota de instâncias de tarefa. Isso só pode ser usado para adicionar frotas de instância de tarefa.

```
aws emr add-instance-fleet --cluster-id <cluster-id>
  --instance-fleet \
    InstanceFleetType=TASK,TargetSpotCapacity=1,\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=20,TimeoutAction=TERMINATE_CLU
\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}']
```

Obter detalhes da configuração de frotas de instâncias em um cluster

Use o comando `list-instance-fleets` para obter detalhes de configuração das frotas de instâncias em um cluster. O comando utiliza um ID de cluster como entrada. O exemplo a seguir

demonstra o comando e sua saída para um cluster que contém um grupo de instâncias de tarefa primárias e um grupo de instâncias de tarefa centrais. Para obter a sintaxe de resposta completa, consulte ListInstanceFleets na referência da API do Amazon EMR.

```
list-instance-fleets --cluster-id <cluster-id>
```


```
{
    "InstanceFleets": [
        {
            "Status": {
                "Timeline": {
                    "ReadyDateTime": 1488759094.637,
                    "CreationDateTime": 1488758719.817
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "ProvisionedSpotCapacity": 6,
            "Name": "CORE",
            "InstanceFleetType": "CORE",
            "LaunchSpecifications": {
                "SpotSpecification": {
                    "TimeoutDurationMinutes": 60,
                    "TimeoutAction": "TERMINATE_CLUSTER"
                }
            },
            "ProvisionedOnDemandCapacity": 2,
            "InstanceTypeSpecifications": [
                {
                    "BidPrice": "0.5",
                    "InstanceType": "m5.xlarge",
                    "WeightedCapacity": 2
                }
            ],
            "Id": "if-1ABC2DEFGHIJ3"
        },
        {
            "Status": {
                "Timeline": {
                    "ReadyDateTime": 1488759058.598,
                    "CreationDateTime": 1488758719.811
```

```
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "ProvisionedSpotCapacity": 0,
            "Name": "MASTER",
            "InstanceFleetType": "MASTER",
            "ProvisionedOnDemandCapacity": 1,
            "InstanceTypeSpecifications": [
                {
                    "BidPriceAsPercentageOfOnDemandPrice": 100.0,
                    "InstanceType": "m5.xlarge",
                    "WeightedCapacity": 1
                }
            ],
           "Id": "if-2ABC4DEFGHIJ4"
        }
    ]
}
```

Usar reservas de capacidade com a frotas de instância

Para iniciar frotas de instâncias sob demanda com opções de reserva de capacidade, anexe outras permissões de perfil de serviço que são necessárias para usar as opções de reserva de capacidade. Como as opções de reserva de capacidade devem ser usadas junto com a estratégia de alocação sob demanda, também é necessário incluir as permissões necessárias para a estratégia de alocação em no perfil de serviço e na política gerenciada. Para ter mais informações, consulte Permissões da estratégia de alocação.

O Amazon EMR oferece suporte a reservas de capacidade abertas e direcionadas. Os tópicos a seguir mostram as configurações de frotas de instâncias que você pode usar com a ação `RunJobFlow` ou o comando `create-cluster` para iniciar frotas de instâncias usando reservas de capacidade sob demanda.

Usar reservas de capacidade aberta com base no melhor esforço

Se as instâncias sob demanda do cluster corresponderem aos atributos das reservas de capacidade aberta (tipo de instância, plataforma, locação e zona de disponibilidade) disponíveis na conta, as

reservas de capacidade serão aplicadas automaticamente. No entanto, o uso das reservas de capacidade não é garantido. Para provisionar o cluster, o Amazon EMR avalia todos os grupos de instâncias especificados na solicitação de inicialização e usa aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós principais solicitados. As reservas de capacidade aberta disponíveis que correspondem ao grupo de instâncias são aplicadas automaticamente. Se as reservas de capacidade aberta disponíveis não corresponderem ao grupos de instâncias, elas permanecerão inutilizadas.

Depois que os nós centrais são provisionados, a zona de disponibilidade é selecionada e corrigida. O Amazon EMR provisiona nós de tarefa em grupos de instâncias, começando pelos de menor preço, na zona de disponibilidade selecionada, até que todos os nós de tarefa sejam provisionados. As reservas de capacidade aberta disponíveis que correspondem aos grupos de instâncias são aplicadas automaticamente.

Veja estes casos de uso da lógica de alocação de capacidade do Amazon EMR para usar reservas de capacidade aberta com base no melhor esforço.

Exemplo 1: o grupo de instâncias de menor preço na solicitação de inicialização tem reservas de capacidade abertas disponíveis

Nesse caso, o Amazon EMR inicia capacidade no grupo de instâncias de menor preço com instâncias sob demanda. Suas reservas de capacidade aberta disponíveis nesse grupo de instâncias são usadas automaticamente.

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade aberta disponíveis	150	100	100
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	100	-	-

Reserva de capacidade aberta utilizada	100	-	-
Reservas de capacidade aberta disponíveis	50	100	100

Depois que a frota de instâncias for iniciada, você poderá executar `describe-capacity-reservations` para ver quantas reservas de capacidade não utilizadas restam.

Exemplo 2: o grupo de instâncias de menor preço na solicitação de execução não tem reservas de capacidade abertas disponíveis

Nesse caso, o Amazon EMR inicia capacidade no grupo de instâncias de menor preço com instâncias sob demanda. No entanto, suas reservas de capacidade aberta permanecem inutilizadas.

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade aberta disponíveis	-	-	100
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	100	-	-
Reserva de capacidade aberta utilizada	-	-	-
Reservas de capacidade aberta disponíveis	-	-	100

Configurar frotas de instâncias para usar reservas de capacidade aberta com base no melhor esforço

Ao usar a ação `RunJobFlow` para criar um cluster baseado em frota de instâncias, defina a estratégia de alocação sob demanda para `lowest-price` e `CapacityReservationPreference` para as opções de reservas de capacidade como `open`. Como alternativa, se você deixar esse campo em branco, o Amazon EMR padronizará a preferência de reserva de capacidade da instância sob demanda como `open`.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "CapacityReservationPreference": "open"
         }
        }
    }
```

Também é possível usar a CLI do Amazon EMR para criar um cluster baseado em frota de instâncias usando reservas de capacidade aberta.

```
aws emr create-cluster \
 --name 'open-ODCR-cluster' \
 --release-label emr-5.30.0 \
 --service-role EMR_DefaultRole \
 --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
 --instance-fleets
 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
   LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={CapacityReservationPreference=open}}'}
```

Onde,

- Substitui-se `open-ODCR-cluster` pelo nome do cluster usando reservas de capacidade abertas.
- Substitui-se `subnet-22XXXX01` pelo ID da sub-rede.

Usar primeiro as reservas de capacidade aberta

Você pode optar por substituir a estratégia de alocação de menor preço e priorizar usar primeiro as reservas de capacidade aberta disponíveis ao provisionar um cluster do Amazon EMR. Nesse caso, o Amazon EMR avalia todos os grupos de instâncias com reservas de capacidade especificadas na solicitação de inicialização e usa aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós centrais solicitados. Se nenhum dos grupos de instâncias com reservas de capacidade tiver capacidade suficiente para os nós centrais solicitados, o Amazon EMR retornará para o caso de melhor esforço descrito no tópico anterior. Ou seja, o Amazon EMR reavalia todos os grupos de instâncias especificados na solicitação de inicialização e usa aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós centrais solicitados. As reservas de capacidade aberta disponíveis que correspondem ao grupo de instâncias são aplicadas automaticamente. Se as reservas de capacidade aberta disponíveis não corresponderem ao grupos de instâncias, elas permanecerão inutilizadas.

Depois que os nós centrais são provisionados, a zona de disponibilidade é selecionada e corrigida. O Amazon EMR provisiona nós de tarefa em grupos de instâncias com reservas de capacidade, começando pelos de menor preço, na zona de disponibilidade selecionada, até que todos os nós de tarefa sejam provisionados. O Amazon EMR usa primeiro as reservas de capacidade aberta disponíveis em cada grupo de instâncias na zona de disponibilidade selecionada e, somente se necessário, usa a estratégia de menor preço para provisionar os nós de tarefa restantes.

Veja estes casos de uso da lógica de alocação de capacidade do Amazon EMR para usar primeiro reservas de capacidade aberta.

Exemplo 1: o grupo de instâncias com reservas de capacidade aberta disponíveis na solicitação de inicialização tem capacidade suficiente para os nós centrais

Nesse caso, o Amazon EMR inicia a capacidade no grupo de instâncias com reservas de capacidade aberta disponíveis, independentemente do preço do grupo de instâncias. Como resultado, as reservas de capacidade aberta são usadas sempre que possível, até que todos os nós centrais sejam provisionados.

Estratégia sob demanda	preço mais baixo
Capacidade solicitada	100
Estratégia de uso	use primeiro as reservas de capacidade

Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade aberta disponíveis	-	-	150
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	-	-	100
Reserva de capacidade aberta utilizada	-	-	100
Reservas de capacidade aberta disponíveis	-	-	50

Exemplo 2: o grupo de instâncias com reservas de capacidade aberta disponíveis na solicitação de inicialização não tem capacidade suficiente para os nós centrais

Nesse caso, o Amazon EMR retorna para iniciar os nós centrais usando a estratégia de menor preço com base no melhor esforço para usar as reservas de capacidade.

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Estratégia de uso	use primeiro as reservas de capacidade		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade aberta disponíveis	10	50	50
Preço sob demanda	$	$$	$$$

Instâncias provisionadas	100	-	-
Reserva de capacidade aberta utilizada	10	-	-
Reservas de capacidade aberta disponíveis	-	50	50

Depois que a frota de instâncias for iniciada, você poderá executar `describe-capacity-reservations` para ver quantas reservas de capacidade não utilizadas restam.

Configurar frotas de instâncias para usar primeiro reservas de capacidade aberta

Ao usar a ação `RunJobFlow` para criar um cluster baseado em frota de instâncias, defina a estratégia de alocação sob demanda para `lowest-price` e `UsageStrategy` para `CapacityReservationOptions` como `use-capacity-reservations-first`.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "UsageStrategy": "use-capacity-reservations-first"
         }
       }
    }
```

Também é possível usar a CLI do Amazon EMR para criar um cluster baseado em frota de instâncias usando primeiro as reservas de capacidade.

```
aws emr create-cluster \
  --name 'use-CR-first-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
```

```
 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={UsageStrategy=use-capacity-reservations-first}}'}
```

Onde,

- Substitui-se `use-CR-first-cluster` pelo nome do cluster usando reservas de capacidade abertas.
- Substitui-se `subnet-22XXXX01` pelo ID da sub-rede.

Usar primeiro as reservas de capacidade direcionadas

Ao provisionar um cluster do Amazon EMR, você pode optar por substituir a estratégia de alocação de menor preço e priorizar usar primeiro as reservas de capacidade direcionada disponíveis. Nesse caso, o Amazon EMR avalia todos os grupos de instâncias com reservas de capacidade direcionadas especificadas na solicitação de inicialização e escolhe aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós centrais solicitados. Se nenhum dos grupos de instâncias com reservas de capacidade direcionada tiver capacidade suficiente para os nós centrais, o Amazon EMR retornará ao caso de melhor esforço descrito anteriormente. Ou seja, o Amazon EMR reavalia todos os grupos de instâncias especificados na solicitação de inicialização e seleciona aquele com o menor preço que tenha capacidade suficiente para iniciar todos os nós centrais solicitados. As reservas de capacidade aberta disponíveis que correspondem ao grupo de instâncias são aplicadas automaticamente. No entanto, as reservas de capacidade direcionadas permanecem inutilizadas.

Depois que os nós centrais são provisionados, a zona de disponibilidade é selecionada e corrigida. O Amazon EMR provisiona nós de tarefa em grupos de instâncias com reservas de capacidade direcionadas, começando pelos de menor preço, na zona de disponibilidade selecionada, até que todos os nós de tarefa sejam provisionados. O Amazon EMR tenta primeiro usar as reservas de capacidade direcionada disponíveis em cada grupo de instâncias na zona de disponibilidade selecionada. Então, somente se necessário, o Amazon EMR usa a estratégia de menor preço para provisionar os nós de tarefa restantes.

Veja estes casos de uso da lógica de alocação de capacidade do Amazon EMR para usar primeiro reservas de capacidade direcionada.

Exemplo 1: O grupo de instâncias com reservas de capacidade direcionada disponíveis na solicitação de inicialização tem capacidade suficiente para os nós centrais

Nesse caso, o Amazon EMR inicia a capacidade no grupo de instâncias com reservas de capacidade direcionada disponíveis, independentemente do preço do grupo de instâncias. Como resultado, as reservas de capacidade direcionada são usadas sempre que possível, até que todos os nós centrais sejam provisionados.

Estratégia sob demanda	preço mais baixo		
Estratégia de uso	use primeiro as reservas de capacidade		
Capacidade solicitada	100		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade direcionada disponíveis	-	-	150
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	-	-	100
Reserva de capacidade direcionada utilizada	-	-	100
Reservas de capacidade direcionada disponíveis	-	-	50

Example Exemplo 2: o grupo de instâncias com reservas de capacidade direcionada disponíveis na solicitação de inicialização não tem capacidade suficiente para os nós centrais

Estratégia sob demanda	preço mais baixo		
Capacidade solicitada	100		
Estratégia de uso	use primeiro as reservas de capacidade		
Tipo de instância	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidade direcionada disponíveis	10	50	50
Preço sob demanda	$	$$	$$$
Instâncias provisionadas	100	-	-
Reservas de capacidade direcionada utilizadas	10	-	-
Reservas de capacidade direcionada disponíveis	-	50	50

Depois que a frota de instâncias for iniciada, você poderá executar `describe-capacity-reservations` para ver quantas reservas de capacidade não utilizadas restam.

Configurar frotas de instâncias para usar primeiro reservas de capacidade direcionada

Ao usar a ação `RunJobFlow` para criar um cluster baseado em frota de instâncias, defina a estratégia de alocação sob demanda como `lowest-price`, `UsageStrategy` de `CapacityReservationOptions` como `use-capacity-reservations-first` e `CapacityReservationResourceGroupArn` de `CapacityReservationOptions` como <your

`resource group ARN>`. Para obter mais informações, consulte Trabalhar com reservas de capacidade no Guia do usuário do Amazon EC2.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "UsageStrategy": "use-capacity-reservations-first",
            "CapacityReservationResourceGroupArn": "arn:aws:resource-groups:sa-
east-1:123456789012:group/MyCRGroup"
         }
       }
    }
```

Onde `arn:aws:resource-groups:sa-east-1:123456789012:group/MyCRGroup` é substituído pelo ARN do grupo de recursos.

Também é possível usar a CLI do Amazon EMR para criar um cluster baseado em frota de instâncias usando reservas de capacidade direcionada.

```
aws emr create-cluster \
  --name 'targeted-CR-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets
 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \
    InstanceFleetType=CORE,TargetOnDemandCapacity=100,\
InstanceTypeConfigs=['{InstanceType=c5.xlarge},{InstanceType=m5.xlarge},
{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={UsageStrategy=use-capacity-reservations-
first,CapacityReservationResourceGroupArn=arn:aws:resource-groups:sa-
east-1:123456789012:group/MyCRGroup}}'}
```

Onde,

- Substitui-se `targeted-CR-cluster` pelo nome do cluster usando reservas de capacidade direcionadas.
- Substitui-se `subnet-22XXXX01` pelo ID da sub-rede.

- Substitui-se `arn:aws:resource-groups:sa-east-1:123456789012:group/MyCRGroup` pelo ARN do grupo de recursos.

Evitar usar reservas de capacidade aberta disponíveis

Example

Para evitar o uso inesperado de qualquer uma de suas reservas de capacidade abertas ao iniciar um cluster do Amazon EMR, defina a estratégia de alocação sob demanda como `lowest-price` e `CapacityReservationPreference` para `CapacityReservationOptions` como `none`. Caso contrário, o Amazon EMR definirá `open` como padrão a preferência de reserva de capacidade da instância sob demanda e tentará usar as reservas de capacidade abertas disponíveis com base no melhor esforço.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "CapacityReservationPreference": "none"
         }
       }
    }
```

Também é possível usar a CLI do Amazon EMR para criar um cluster baseado em frota de instâncias sem usar reservas de capacidade aberta.

```
aws emr create-cluster \
  --name 'none-CR-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={CapacityReservationPreference=none}}'}
```

Onde,

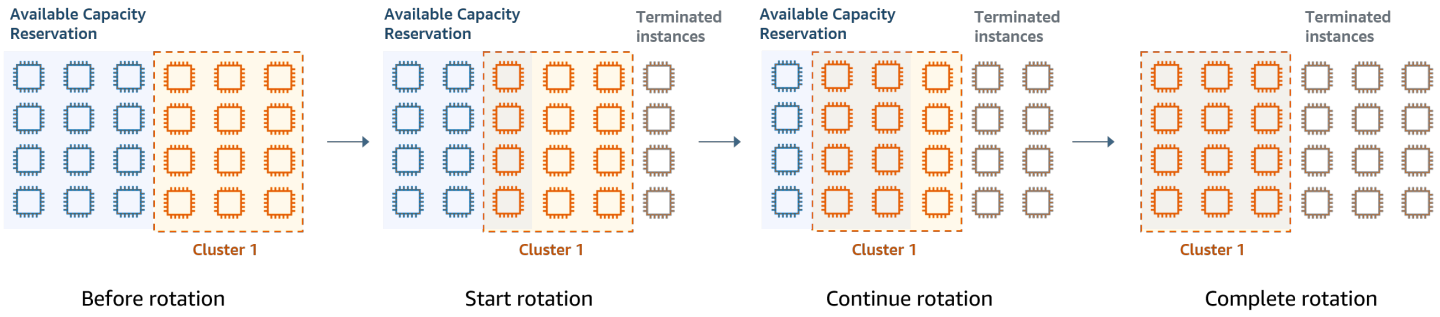
- Substitui-se `none-CR-cluster` pelo nome do cluster que não está usando reservas de capacidade abertas.
- Substitui-se `subnet-22XXXX01` pelo ID da sub-rede.

Cenários para o uso de reservas de capacidade

Você pode se beneficiar do uso de reservas de capacidade nos cenários a seguir.

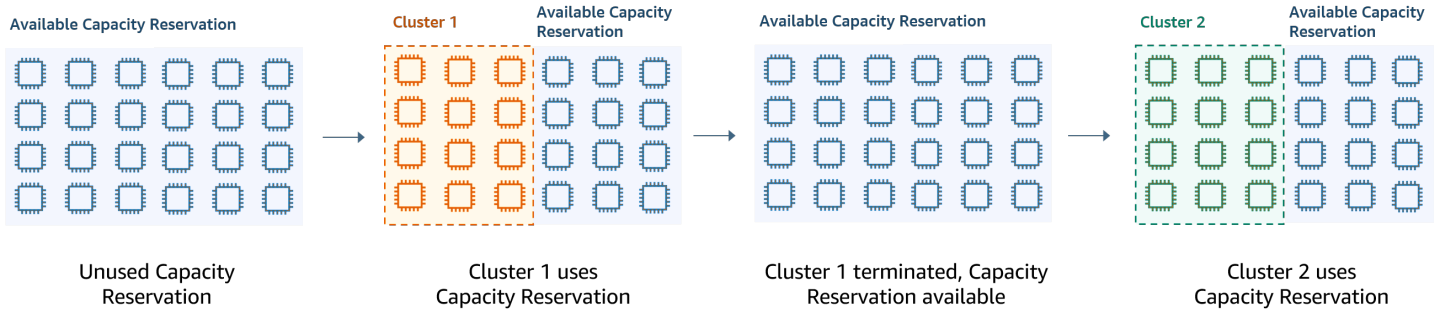
Cenário 1: alternar um cluster de execução prolongada usando reservas de capacidade

Ao alternar um cluster de execução prolongada, você poderá ter requisitos rígidos sobre os tipos de instância e as zonas de disponibilidade das novas instâncias provisionadas. Com as reservas de capacidade, você pode usar a garantia de capacidade para concluir a alternância do cluster sem interrupções.



Cenário 2: provisionar clusters sucessivos de curta duração usando reservas de capacidade

Também é possível usar reservas de capacidade para provisionar um grupo de clusters sucessivos e de curta duração para workloads individuais, de forma que, ao encerrar um cluster, o próximo cluster possa usar as reservas de capacidade. Você pode usar reservas de capacidade direcionadas para garantir que apenas os clusters pretendidos usem as reservas de capacidade.



Configurar grupos de instâncias uniformes

Com a configuração de grupos de instâncias, cada tipo de nó (principal, core ou tarefa) consiste no mesmo tipo de instância e na mesma opção de compra para instâncias: Sob demanda ou Spot. Você especifica essas configurações ao criar um grupo de instâncias. Não é possível alterá-las depois. No entanto, você pode adicionar instâncias do mesmo tipo e opção de compra a grupos de instâncias core e de tarefas. Você também pode remover instâncias.

Se as instâncias sob demanda do cluster corresponderem aos atributos das reservas de capacidade aberta (tipo de instância, plataforma, locação e zona de disponibilidade) disponíveis na conta, as reservas de capacidade serão aplicadas automaticamente. É possível usar reservas de capacidade aberta para nós primários, centrais e de tarefa. No entanto, você não poderá usar reservas de capacidade direcionadas nem impedir que instâncias sejam iniciadas em reservas de capacidade abertas com atributos correspondentes ao provisionar clusters usando grupos de instâncias. Para usar reservas de capacidade direcionadas ou evitar que instâncias sejam iniciadas em reservas de capacidade abertas, use frotas de instâncias. Para ter mais informações, consulte Usar reservas de capacidade com a frotas de instância.

Para adicionar tipos de instâncias diferentes depois que um cluster for criado, é possível adicionar outros grupos de instâncias de tarefas. Você pode escolher diferentes tipos de instância e opções de compra para cada grupo de instância. Para ter mais informações, consulte Usar ajuste de escala de clusters.

Ao iniciar instâncias, a preferência da reserva de capacidade da instância sob demanda será padronizada como `open`, o que permitirá que ela seja executada em qualquer reserva de capacidade em aberto que tenha atributos correspondentes (tipo de instância, plataforma, zona de disponibilidade). Para obter mais informações sobre reservas de capacidade sob demanda, consulte Usar reservas de capacidade com a frotas de instância.

Esta seção discute a criação de um cluster com grupos de instâncias uniformes. Para obter mais informações sobre como modificar um grupo de instâncias existente, adicionando ou removendo instâncias manualmente ou com escalabilidade automática, consulte Gerenciar clusters.

Usar o console para configurar grupos de instâncias uniformes

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Criar um cluster com grupos de instâncias usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e Criar cluster.
3. Em Configuração do cluster, escolha Grupos de instâncias.
4. Em Grupos de nós, há uma seção para cada tipo de grupo de nós. Para o grupo de nós primários, marque a caixa de seleção Usar múltiplos nós primários se quiser ter três nós primários. Marque a caixa de seleção Usar a opção de compra spot se quiser usar a compra spot.
5. Para os grupos de nós primários e centrais, selecione Adicionar tipo de instância e escolha até cinco tipos de instância. Para o grupo de tarefa, selecione Adicionar tipo de instância e escolha até 15 tipos de instância. O Amazon EMR poderá provisionar qualquer combinação desses tipos de instância ao executar o cluster.
6. Para alterar essas configurações, em cada tipo de grupo de nós, escolha o menu suspenso Ações ao lado de cada instância:

   Adicionar volumes do EBS

   Especifique os volumes do EBS a serem anexados ao tipo de instância após o provisionamento do Amazon EMR.

   Editar preço máximo spot

   Especifique um preço spot máximo para cada tipo de instância da frota. Você pode definir esse preço como uma porcentagem do preço sob demanda ou como uma quantia em dólar. Caso o preço spot atual em uma zona de disponibilidade esteja abaixo do preço

spot máximo, o Amazon EMR provisiona instâncias spot. Você paga o preço spot e não necessariamente o preço spot máximo.

7. Opcionalmente, expanda a Configuração do nó para inserir uma configuração JSON ou carregar o JSON do Amazon S3.
8. Escolha qualquer outra opção que se aplique ao cluster.
9. Para iniciar o cluster, escolha Criar cluster.

Old console

O procedimento a seguir discute Advanced options (Opções avançadas) quando você cria um cluster. Usar Quick options (Opções rápidas) também cria um cluster com a configuração de grupos de instâncias.

Criar um cluster com grupos de instâncias uniformes usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Escolha Go to advanced options (Ir para opções avançadas), insira opções de Software Configuration (Configuração do software) e, em seguida, escolha Next (Próximo).
4. Na tela Hardware Configuration (Configuração do hardware), deixe a opção Uniform instance groups (Grupos de instâncias uniformes) selecionada.
5. Escolha Network (Rede) e depois escolha a opção de EC2 Subnet (Sub-rede de EC2) para executar seu cluster. A sub-rede que você escolher é associada a um grupo de disponibilidade, que é listado em cada sub-rede. Para ter mais informações, consulte Configurar redes.

> **Note**
>
> Sua conta e região podem lhe dar a opção de escolher Iniciar no EC2-Classic para Rede. Se você escolher essa opção, escolha uma zona de disponibilidade do EC2 em EC2 Availability Zone (Zona de disponibilidade do EC2), ao invés de uma sub-rede do EC2 em EC2 Subnet (Sub-rede do EC2). Para obter mais informações, consulte Amazon EC2 e Amazon VPC no Guia do usuário do Amazon EC2.

6. Em cada linha de Node type (Tipo de nó):

   - Em Tipo de nó, se você quiser alterar o nome padrão de um grupo de instâncias, escolha o ícone de lápis e insira um nome amigável. Se quiser remover o grupo de instâncias Tarefa, escolha o ícone X. Escolha Add task instance group (Adicionar grupo de instâncias de tarefa) para adicionar outros grupos de instâncias Task (Tarefa).
   - Em Tipo de instância, escolha o ícone de lápis e escolha o tipo de instância que você deseja usar para esse tipo de nó.

     > **Important**
     >
     > Quando você escolhe um tipo de instância usando o AWS Management Console, o número de vCPUs mostrado para cada tipo de instância é o número de vcores YARN para esse tipo de instância, não o número de vCPUs EC2 para esse tipo de instância. Para obter mais informações sobre o número de vCPUs para o seu tipo de instância, consulte os tipos de instância do Amazon EC2.

   - Em Tipo de instância, escolha o ícone de lápis de Configurações e edite as configurações de aplicações para cada grupo de instâncias.
   - Em Instance count (Contagem de instâncias), digite o número de instâncias em que você deseja executar esse tipo de nó.
   - Em Opção de compra, escolha Sob demand ou Spot. Se você escolher Spot, selecione uma opção para o preço máximo de instâncias spot. Por padrão, a opção Usar sob demanda como o preço máximo é selecionada. Você pode selecionar Set max $/hr (Definir max $/hr) e, em seguida, inserir o preço máximo. Zona de disponibilidade da EC2 Subnet (Sub-rede de EC2) que você escolheu está abaixo do Maximum Spot price (Preço spot máximo).

     > **Tip**
     >
     > Pare na dica de ferramenta de Spot para ver o preço spot atual para as zonas de disponibilidade na região atual. O menor preço Spot está em verde. Talvez você queira usar essas informações para alterar sua seleção de EC2 Subnet (Sub-rede do EC2).

   - Em Auto Scaling for Core and Task node types (Escalabilidade automática para tipos de nós core e de tarefa), escolha o ícone de lápis e configure as opções de escalabilidade

automática. Para ter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.

7. Escolha Add task instance group (Adicionar grupo de instâncias de tarefa) conforme desejado e defina as configurações conforme descrito na etapa anterior.
8. Escolha Next (Próximo), modifique outras configurações de cluster e execute o cluster.

Use o AWS CLI para criar um cluster com grupos de instâncias uniformes

Para especificar a configuração de grupos de instâncias para um cluster usando a AWS CLI, use o comando `create-cluster` junto com o parâmetro `--instance-groups`. O Amazon EMR supõe a opção de instância sob demanda, a menos que você especifique o argumento `BidPrice` para um grupo de instâncias. Para obter exemplos de comandos `create-cluster` que executam grupos de instâncias uniformes com instâncias sob demanda e uma variedade de opções de cluster, digite `aws emr create-cluster help` na linha de comando, ou consulte create-cluster na AWS CLI Command Reference.

Você pode usar o AWS CLI para criar grupos de instâncias uniformes em um cluster que usa instâncias spot. O preço Spot oferecido depende da zona de disponibilidade. Ao usar a CLI ou a API, você pode especificar a zona de disponibilidade com o argumento `AvailabilityZone` (se estiver usando uma rede EC2-classic) ou o `SubnetID` argumento do parâmetro `--ec2-attributes` . A zona de disponibilidade ou sub-rede selecionada se aplica ao cluster e, portanto, é usada para todos os grupos de instâncias. Se você não especificar uma zona de disponibilidade ou sub-rede explicitamente, o Amazon EMR selecionará a zona de disponibilidade com o menor preço spot quando iniciar o cluster.

O exemplo a seguir demonstra um comando `create-cluster` que cria um grupo de instâncias primárias, um grupo de instâncias centrais e dois grupos de instâncias de tarefa, todos usando instâncias spot. Substitua *myKey* pelo nome do par de chaves do Amazon EC2.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "MySpotCluster" \
```

```
  --release-label emr-7.1.0 \
  --use-default-roles \
  --ec2-attributes KeyName=myKey \
  --instance-groups \
    InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1,BidPrice=0.25 \
    InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=2,BidPrice=0.03 \
    InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=4,BidPrice=0.03 \
    InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=2,BidPrice=0.04
```

Usando a CLI, é possível criar clusters de grupos de instâncias uniformes que especificam uma AMI personalizada exclusiva para cada tipo de instância do grupo de instâncias. Assim, você pode usar arquiteturas de instância diferentes no mesmo grupo de instâncias. Todo tipo de instância deve usar uma AMI personalizada com uma arquitetura correspondente. Por exemplo, você configuraria um tipo de instância m5.xlarge com uma AMI personalizada de arquitetura x86_64 e um tipo de instância m6g.xlarge com uma AMI personalizada de arquitetura `AWS  AARCH64` (ARM) correspondente.

O exemplo a seguir mostra um cluster uniforme de grupos de instâncias criado com dois tipos de instância, cada um com a própria AMI personalizada. As AMIs personalizadas são especificadas somente no nível do tipo de instância, não no nível do cluster. Isso evita conflitos entre AMIs do tipo de instância e uma AMI no nível do cluster, o que faria com que a inicialização do cluster falhasse.

```
aws emr create-cluster
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-groups \

 InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
 \

 InstanceGroupType=CORE,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-234567
```

É possível adicionar múltiplas AMIs personalizadas a um grupo de instâncias que você adiciona a um cluster em execução. O argumento `CustomAmiId` pode ser usado com o comando `add-instance-groups`, conforme mostrado no exemplo a seguir.

```
aws emr add-instance-groups --cluster-id j-123456 \
  --instance-groups \

 InstanceGroupType=Task,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
```

Usar o SDK para Java para criar um grupo de instâncias

Você instancia um objeto `InstanceGroupConfig` que especifica a configuração de um grupo de instâncias para um cluster. Para usar instâncias Spot, defina as propriedades `withBidPrice` e `withMarket` no objeto `InstanceGroupConfig`. O código a seguir mostra como definir grupos de instância primários, centrais e de tarefa que executam instâncias Spot.

```
InstanceGroupConfig instanceGroupConfigMaster = new InstanceGroupConfig()
 .withInstanceCount(1)
 .withInstanceRole("MASTER")
 .withInstanceType("m4.large")
 .withMarket("SPOT")
 .withBidPrice("0.25");

InstanceGroupConfig instanceGroupConfigCore = new InstanceGroupConfig()
 .withInstanceCount(4)
 .withInstanceRole("CORE")
 .withInstanceType("m4.large")
 .withMarket("SPOT")
 .withBidPrice("0.03");

InstanceGroupConfig instanceGroupConfigTask = new InstanceGroupConfig()
 .withInstanceCount(2)
 .withInstanceRole("TASK")
 .withInstanceType("m4.large")
 .withMarket("SPOT")
 .withBidPrice("0.10");
```

Práticas recomendadas para flexibilidade de instâncias e de zona de disponibilidade

Cada um Região da AWS tem vários locais isolados, conhecidos como zonas de disponibilidade. Ao iniciar uma instância, é possível especificar, opcionalmente, uma zona de disponibilidade (AZ) na Região da AWS utilizada. A flexibilidade da zona de disponibilidade é a distribuição de instâncias em diversas AZs. Se houver falha em uma instância, você poderá projetar sua aplicação para que uma instância em outra AZ possa lidar com as solicitações. Para obter mais informações sobre zonas de disponibilidade, consulte a documentação sobre regiões e zonas de disponibilidade no Guia do usuário da Amazon EC2.

A flexibilidade da instância é o uso de múltiplos tipos de instância para atender aos requisitos de capacidade. Ao expressar flexibilidade com instâncias, é possível usar a capacidade agregada em todos os tamanhos, famílias e gerações de instâncias. Uma flexibilidade maior aumenta a chance de

encontrar e alocar a quantidade necessária de capacidade computacional comparado a um cluster que usa um único tipo de instância.

A flexibilidade da instância e da zona de disponibilidade reduz erros de capacidade insuficientes (ICE) e interrupções spot quando comparada a um cluster com um único tipo de instância ou AZ. Use as práticas recomendadas abordadas aqui para determinar quais instâncias diversificar depois de conhecer a família e o tamanho iniciais da instância. Essa abordagem maximiza a disponibilidade dos grupos de capacidade do Amazon EC2 com o mínimo de performance e variação de custo.

Ser flexível em relação às zonas de disponibilidade

É recomendável configurar todas as zonas de disponibilidade para uso em sua nuvem privada virtual (VPC) e selecioná-las para o cluster do EMR. Os clusters devem existir em apenas uma zona de disponibilidade, mas com frotas de instâncias do Amazon EMR, é possível selecionar múltiplas sub-redes para diferentes zonas de disponibilidade. Quando o Amazon EMR executa o cluster, ele procura entre as sub-redes para encontrar as instâncias e opções de compra que você especificar. Quando você provisiona um cluster do EMR para múltiplas sub-redes, o cluster pode acessar um grupo de capacidade mais profundo do Amazon EC2 quando comparado aos clusters de uma única sub-rede.

Se você precisar priorizar certo número de zonas de disponibilidade para uso em sua nuvem privada virtual (VPC) para o cluster do EMR, é possível aproveitar o recurso de pontuação de posicionamento de spot com o Amazon EC2. Com a pontuação de posicionamento spot, você especifica os requisitos de computação para suas instâncias spot e, em seguida, o EC2 retorna as dez Regiões da AWS melhores zonas de disponibilidade pontuadas em uma escala de 1 a 10. Uma pontuação de 10 indica que a solicitação spot tem alta probabilidade de êxito; uma pontuação de 1 indica que a solicitação spot provavelmente não terá êxito. Para obter mais informações sobre como usar a pontuação de posicionamento spot, consulte Pontuação de posicionamento spot no Guia do usuário do Amazon EC2.

Ser flexível em relação aos tipos de instância

A flexibilidade da instância é o uso de múltiplos tipos de instância para atender aos requisitos de capacidade. A flexibilidade da instância beneficia o uso de instâncias spot e sob demanda do Amazon EC2. Com as instâncias spot, a flexibilidade da instância permite que o Amazon EC2 inicie instâncias com base em de capacidade mais profundos usando dados de capacidade em tempo real. Também prevê quais instâncias estão mais disponíveis. Isso oferece menos interrupções e pode reduzir o custo geral da workload. Com instâncias sob demanda, a flexibilidade da instância reduz

os erros de capacidade insuficiente (ICE) quando a capacidade total é provisionada em um número maior de grupos de instâncias.

Para clusters de grupos de instâncias, é possível especificar até 50 tipos de instância do EC2. Para frotas de instâncias com estratégia de alocação, você pode especificar até 30 tipos de instância do EC2 para cada grupo de nós primário, central e de tarefa. Uma variedade maior de instâncias melhora os benefícios da flexibilidade da instância.

Expressar a flexibilidade da instância

Considere as práticas recomendadas a seguir para expressar a flexibilidade de instância da aplicação.

Tópicos

- Determinar a família e o tamanho da instância
- Incluir instâncias adicionais

Determinar a família e o tamanho da instância

O Amazon EMR oferece suporte a diversos tipos de instância para diferentes casos de uso. Esses tipos de instância estão listados na documentação Tipos de instâncias compatíveis. Cada tipo de instância pertence a uma família de instâncias que descreve para qual aplicação o tipo é otimizado.

Para novas workloads, compare com os tipos de instância da família de uso geral, como `m5` ou `c5`. Em seguida, monitore as métricas do sistema operacional e do YARN do Ganglia e do Amazon CloudWatch determine os gargalos do sistema no pico de carga. Os gargalos incluem operações de CPU, de memória, de armazenamento e de E/S. Após identificar os gargalos, escolha otimizado para computação, otimizado para memória, otimizado para armazenamento ou outra família de instâncias apropriada para seus tipos de instância. Para obter mais detalhes, consulte a página Determine a infraestrutura certa para suas cargas de trabalho do Spark no guia de melhores práticas do Amazon EMR em. GitHub

Em seguida, identifique o menor contêiner do YARN ou executor do Spark que a aplicação exige. Esse é o menor tamanho de instância adequado ao contêiner e o tamanho mínimo de instância para o cluster. Use essa métrica para determinar instâncias com as quais você poderá diversificar ainda mais. Uma instância menor permitirá maior flexibilidade da instância.

Para obter a máxima flexibilidade da instância, você deve aproveitar o maior número possível de instâncias. É recomendável diversificar com instâncias que tenham especificações de hardware

semelhantes. Isso maximiza o acesso aos grupos de capacidade do EC2 com variação mínima de custo e performance. Diversifique em vários tamanhos. Para fazer isso, priorize antes o AWS Graviton e as gerações anteriores. Uma boa regra geral é tentar ser flexível para pelo menos 15 tipos de instância para cada workload. É recomendável começar com instâncias de uso geral, otimizadas para computação ou para memória. Esses tipos de instância fornecerão a maior flexibilidade.

Incluir instâncias adicionais

Para ter o máximo de diversidade, inclua outros tipos de instância. Priorize antes o tamanho da instância, o Graviton e a flexibilidade da geração. Isso permite o acesso a grupos adicionais de capacidade do EC2 com perfis de custo e performance semelhantes. Se você precisar de mais flexibilidade devido ao ICE ou a interrupções pontuais, considere a flexibilidade de variantes e de famílias. Cada abordagem tem vantagens e desvantagens conforme o caso de uso e os requisitos.

- Flexibilidade de tamanho: primeiro, diversifique com instâncias de tamanhos diferentes dentro da mesma família. As instâncias da mesma família oferecem o mesmo custo e performance, mas podem iniciar um número diferente de contêineres em cada host. Por exemplo, se o tamanho mínimo do executor necessário for de 2 vCPU e 8 Gb de memória, o tamanho mínimo da instância será `m5.xlarge`. Para flexibilidade de tamanho, inclua `m5.xlarge`, `m5.2xlarge`, `m5.4xlarge`, `m5.8xlarge`, `m5.12xlarge`, `m5.16xlarge` e `m5.24xlarge`.
- Flexibilidade do Graviton: além do tamanho, você pode diversificar com instâncias do Graviton. As instâncias Graviton são alimentadas por processadores AWS Graviton2 que oferecem a melhor relação preço/desempenho para cargas de trabalho em nuvem no Amazon EC2. Por exemplo, com o tamanho mínimo de instância de `m5.xlarge`, você pode incluir `m6g.xlarge`, `m6g.2xlarge`, `m6g.4xlarge`, `m6g.8xlarge` e `m6g.16xlarge` para flexibilidade do Graviton.
- Flexibilidade de geração: semelhante ao Graviton e à flexibilidade de tamanho, as instâncias das famílias da geração anterior compartilham as mesmas especificações de hardware. Isso resulta em um perfil de custo e performance semelhante, com um aumento no grupo total acessível do Amazon EC2. Para flexibilidade de geração, inclua `m4.xlarge`, `m4.2xlarge`, `m4.10xlarge` e `m4.16xlarge`.
- Flexibilidade de famílias e variantes
  - Capacidade: para otimizar a capacidade, recomendamos a flexibilidade de instâncias em todas as famílias de instâncias. Instâncias comuns de diferentes famílias de instâncias têm grupos de instâncias mais profundos que ajudam a atender aos requisitos de capacidade. Porém, instâncias de famílias diferentes terão diferentes proporções de vCPU em relação à memória. Isso resultará em subutilização se o contêiner da aplicação esperado for dimensionado para uma instância diferente. Por exemplo, com `m5.xlarge`, inclua instâncias otimizadas para

computação, como `c5`, ou instâncias otimizadas para memória, como `r5` para obter flexibilidade de família de instâncias.

- Custo: para otimizar o custo, recomenda-se a flexibilidade da instância em todas as variantes. Essas instâncias têm a mesma proporção de memória e de vCPU da instância inicial. A desvantagem com a flexibilidade de variantes é que essas instâncias têm grupos de capacidade menores, o que pode resultar em capacidade adicional limitada ou maiores interrupções spot. Com `m5.xlarge`, por exemplo, inclua instâncias baseadas em AMD (`m5a`), instâncias baseadas em SSD (`m5d`) ou instâncias otimizadas para rede (`m5n`) para obter flexibilidade de variantes de instância.

Práticas recomendadas para configuração de clusters

Use a orientação nesta seção para ajudá-lo a determinar os tipos de instâncias, as opções de compra e a quantidade de armazenamento a ser provisionado para cada tipo de nó em um cluster do EMR.

Que tipo de instância você deve usar?

Existem várias maneiras de adicionar instâncias do Amazon EC2 ao cluster. O método a ser escolhido depende se você usará a configuração de grupos de instâncias ou a configuração de frotas de instâncias para o cluster.

- Grupos de instâncias
  - Adicione manualmente instâncias do mesmo tipo a grupos de instâncias core e de tarefa existentes.
  - Adicione manualmente um grupo de instâncias de tarefa, que pode usar um tipo de instância diferente.
  - Configure a escalabilidade automática no Amazon EMR para um grupo de instâncias, adicionando e removendo instâncias automaticamente com base no valor de uma métrica da CloudWatch Amazon que você especificar. Para ter mais informações, consulte Usar ajuste de escala de clusters.
- Frotas de instâncias
  - Adicione uma única frota de instâncias de tarefa.
  - Altere a capacidade de destino para instâncias sob demanda e instâncias spot para as frotas de instâncias core e de tarefas existentes. Para ter mais informações, consulte Configurar frotas de instâncias.

Uma maneira de planejar as instâncias do seu cluster é executar um cluster de teste com um conjunto de dados de amostra representativo e monitorar a utilização dos nós nesse cluster. Para ter mais informações, consulte Visualizar e monitorar um cluster. Outra maneira é calcular a capacidade das instâncias que você está considerando e comparar esse valor com o tamanho dos seus dados.

Em geral, o tipo de nó primário, que atribui tarefas, não requer uma instância do EC2 com muita capacidade de processamento. Instâncias do Amazon EC2 para o tipo de nó central, que processam tarefas e armazenam dados no HDFS, precisam tanto capacidade de processamento como de capacidade de armazenamento. Instâncias do Amazon EC2 para o tipo de nó de tarefa, que não armazenam dados, precisam apenas de poder de processamento. Para conhecer diretrizes sobre instâncias do Amazon EC2 disponíveis e sua configuração, consulte Configurar instâncias do Amazon EC2.

As diretrizes a seguir se aplicam à maioria dos clusters do Amazon EMR.

- Há um limite de vCPU para o número total de instâncias sob demanda do Amazon EC2 que você executa em uma conta por. AWS Região da AWS Para obter mais informações sobre o limite de vCPUs e como solicitar um aumento de limite para sua conta, consulte Instâncias sob demanda no Guia do usuário do Amazon EC2 para instâncias do Linux.
- O nó primário normalmente não tem grandes requisitos de computação. Para clusters com um grande número de nós ou para clusters com aplicativos que são implantados especificamente no nó primário (JupyterHub, Hue etc.), um nó primário maior pode ser necessário e pode ajudar a melhorar o desempenho do cluster. Por exemplo, considere usar uma instância m5.xlarge para clusters pequenos (até 50 nós) e aumentar para um tipo de instância maior para clusters maiores.
- As necessidades de computação dos nós core e de tarefas dependem do tipo de processamento realizado pelo seu aplicativo. Muitos trabalhos podem ser executados em tipos de instâncias de uso geral, que oferecem uma performance equilibrada em termos de CPU, espaço em disco e entrada/saída. Clusters que usam muitos recursos de computação podem se beneficiar com a execução em instâncias com CPU de alta performance, que possuem proporcionalmente mais CPU do que RAM. Aplicativos de banco de dados e de cache de memória podem se beneficiar com a execução em instâncias com mais memória. Aplicações que fazem uso intenso da rede e da CPU, como análise, NLP e machine learning, podem se beneficiar com a execução de instância de computação em cluster, que fornecem recursos de CPU proporcionalmente altos e maior performance de rede.
- Se diferentes fases do seu cluster tiverem necessidades de capacidade diferentes, você pode começar com um pequeno número de nós core e aumentar ou diminuir o número de nós de tarefas para atender aos requisitos de capacidade variáveis do seu fluxo de trabalho.

- A quantidade de dados que você pode processar depende da capacidade de nós core e do tamanho dos seus dados como entrada, durante o processamento, e como saída. Os conjuntos de dados de entrada, intermediários e de saída residem todos no cluster durante o processamento.

Quando você deve usar instâncias spot?

Ao executar um cluster no Amazon EMR, você pode optar por executar instâncias primárias, centrais e de tarefa em instâncias spot. Como cada tipo de grupo de instâncias desempenha um papel diferente no cluster, há implicações na execução de cada tipo de nó em instâncias spot. Você não pode alterar uma opção de compra de instância enquanto um cluster está em execução. Para alterar um grupo de instâncias sob demanda para instâncias spot, ou vice-versa, para nós primários e centrais, você deve terminar o cluster e iniciar um novo. Para nós de tarefa, você pode iniciar um novo grupo de instâncias de tarefa ou frota de instâncias e remover o antigo.

Tópicos

Configurações do Amazon EMR para evitar falhas em trabalhos causado pelo término de instâncias spot de nós de tarefa

Como as instâncias spot são frequentemente usadas para executar nós de tarefas, o Amazon EMR tem a funcionalidade padrão para programar trabalhos do YARN para que os trabalhos em execução não falhem quando os nós de tarefas em execução nas instâncias spot forem encerrados. O Amazon EMR faz isso ao permitir que processos principais de aplicações sejam executados somente em nós centrais. O processo principal da aplicação controla os trabalhos em execução e precisa permanecer ativo durante a vida útil do trabalho.

A versão 5.19.0 e as versões posteriores do Amazon EMR usam o recurso de rótulos de nós do YARN integrado para conseguir isso. (As versões anteriores usavam um patch de código). As propriedades nas classificações de configuração `yarn-site` e `capacity-scheduler` são configuradas por padrão para que o programador de capacidade e o programador justo do YARN

aproveitem os rótulos de nós. O Amazon EMR rotula automaticamente os nós centrais com o rótulo `CORE` e define propriedades para que as aplicações principais sejam programadas somente em nós com o rótulo CORE. Modificar manualmente as propriedades relacionadas nas classificações de configuração yarn-site e docapacity-scheduler, ou diretamente nos arquivos XML associados, pode interromper esse recurso ou modificar essa funcionalidade.

O Amazon EMR configura as seguintes propriedades e valores por padrão. Tenha cuidado ao configurar essas propriedades.

> **Note**
>
> A partir do Amazon EMR série 6.x, o recurso de rótulos de nó do YARN é desabilitado por padrão. Os processos primários da aplicação podem ser executados tanto nos nós centrais como nos nós de tarefa por padrão. É possível habilitar o recurso de rótulos de nó do YARN configurando as seguintes propriedades:
>
> - `yarn.node-labels.enabled: true`
> - `yarn.node-labels.am.default-node-label-expression: 'CORE'`

- yarn-site (yarn-site.xml) Em todos os nós
  - `yarn.node-labels.enabled: true`
  - `yarn.node-labels.am.default-node-label-expression: 'CORE'`
  - `yarn.node-labels.fs-store.root-dir: '/apps/yarn/nodelabels'`
  - `yarn.node-labels.configuration-type: 'distributed'`
- yarn-site (yarn-site.xml) em nós primários e centrais
  - `yarn.nodemanager.node-labels.provider: 'config'`
  - `yarn.nodemanager.node-labels.provider.configured-node-partition: 'CORE'`
- capacity-scheduler (capacity-scheduler.xml) Em todos os nós
  - `yarn.scheduler.capacity.root.accessible-node-labels: '*'`
  - `yarn.scheduler.capacity.root.accessible-node-labels.CORE.capacity: 100`
  - `yarn.scheduler.capacity.root.default.accessible-node-labels: '*'`
  - `yarn.scheduler.capacity.root.default.accessible-node-labels.CORE.capacity: 100`

Nó primário como uma instância spot

O nó primário controla e direciona o cluster. Quando ela for terminada, o cluster será encerrado. Portanto, você só deve iniciar o nó primário como uma instância spot se você estiver executando um cluster em que o término repentino seja aceitável. Este pode ser o caso se você está testando uma nova aplicação, tem um cluster que periodicamente mantém a persistência de dados em um armazenamento externo, como o Amazon S3, ou está executando um cluster em que o custo é mais importante do que garantir a conclusão do cluster.

Quando você executa o grupo de instâncias primárias como uma instância spot, o cluster não é iniciado até que essa solicitação de instância spot seja atendida. Isso é algo a considerar ao selecionar seu preço spot máximo.

Você só pode adicionar um nó primário de instância spot ao iniciar o cluster. Não é possível adicionar ou remover nós primários de um cluster em execução.

Normalmente, você só executaria o nó primário como uma instância spot se estivesse executando o cluster inteiro (todos os grupos de instâncias) como instâncias spot.

Nós centrais em instâncias spot

Nós core processam dados e armazenam informações usando o HDFS. O encerramento de uma instância core representa risco de perda de dados. Por esse motivo, você só deve executar nós core em instâncias spot quando a perda de dados HDFS parcial é aceitável.

Quando você executa o grupo de instâncias centrais como instâncias spot, o Amazon EMR aguarda até poder provisionar todas as instâncias centrais solicitadas antes de iniciar o grupo de instâncias. Em outras palavras, se você solicitar seis instâncias do Amazon EC2 e apenas cinco estiverem disponíveis no preço spot máximo ou abaixo dele, o grupo de instâncias não será iniciado. O Amazon EMR continuará esperando até que todas as seis instâncias do Amazon EC2 estejam disponíveis ou até que você encerre o cluster. Você pode alterar o número de instâncias spot em um grupo de instâncias core para adicionar capacidade a um cluster em execução. Para obter mais informações sobre como trabalhar com grupos de instâncias, e como as instâncias spot funcionam com frotas de instâncias, consulte the section called “Configurar frotas de instâncias ou grupos de instâncias”.

Nós de tarefa em instâncias spot

Os nós de tarefa processam dados, mas não retêm dados persistentes no HDFS. Se eles forem encerrados porque o preço spot ultrapassou seu preço spot máximo, não haverá perda de dados, e o efeito no seu cluster será mínimo.

Quando você executa um ou mais grupos de instâncias de tarefa como instâncias spot, o Amazon EMR provisiona o número possível de nós de tarefa usando o preço spot máximo. Isso significa que, se você solicitar um grupo de instâncias de tarefa com seis nós, e apenas cinco instâncias spot estiverem disponíveis até seu preço spot máximo, o Amazon EMR executará o grupo de instâncias com cinco nós, adicionando o sexto posteriormente, se possível.

A execução de grupos de instâncias de tarefas como instâncias Spot é uma maneira estratégica de expandir a capacidade do seu cluster e, ao mesmo tempo, minimizar os custos. Se você executar os grupos de instâncias primárias e centrais como instâncias sob demanda, a capacidade será garantida para a execução do cluster. Você pode adicionar instâncias de tarefa aos grupos de instâncias da tarefa conforme necessário, para processar picos de tráfego ou agilizar processamento de dados.

Você pode adicionar ou remover nós de tarefas usando o console ou a API. AWS CLI Você também pode acrescentar grupos de tarefas adicionais, mas não poderá remover um grupo de tarefas depois de criado.

Configurações de instâncias para cenários de aplicações

A tabela a seguir é uma referência rápida às opções de compras de tipos de nó e configurações que são geralmente apropriadas para vários cenários de aplicativos. Escolha o link para exibir mais informações sobre cada tipo de cenário.

Cenário de aplicações	Opção de compra do nó primário	Opção de compra de nós centrais	Opção de compra de nós de tarefa
Clusters de execução prolongada e data warehouses	Sob demanda	Combinação de frotas de instâncias ou Sob demanda	Combinação de spot ou frota de instâncias
Cargas de trabalho com base no custo	Spot	Spot	Spot
Cargas de trabalho críticas para dados	Sob demanda	Sob demanda	Combinação de spot ou frota de instâncias
Testes de aplicativos	Spot	Spot	Spot

Há vários cenários em que instâncias spot são úteis para executar um cluster do Amazon EMR.

Clusters de execução prolongada e data warehouses

Se você estiver executando um cluster do Amazon EMR persistente que tem uma variação previsível de capacidade computacional, como um data warehouse, pode lidar com a demanda de pico com um custo menor usando instâncias spot. Você pode iniciar seus grupos de instâncias primárias e central como instâncias sob demanda para lidar com a capacidade normal e iniciar o grupo de instâncias de tarefa como instâncias spot para lidar com requisitos de carga de pico.

Cargas de trabalho com base no custo

Ao executar clusters transitórios para os quais um custo menor é mais importante do que o tempo para conclusão, e uma perda parcial do trabalho é aceitável, você pode executar o cluster inteiro (grupos de instâncias primárias, centrais e de tarefa) como instâncias spot para se beneficiar com a maior redução dos custos.

Cargas de trabalho críticas para dados

Se você estiver executando um cluster para o qual o menor custo é mais importante que o tempo para conclusão, mas uma perda parcial do trabalho não é aceitável, inicie os grupos de instâncias primárias e centrais como instâncias sob demanda e complemente-as com um ou mais grupos de instâncias de tarefa de instâncias spot. Executar grupos de instâncias primárias e centrais como instâncias sob demanda garante que seus dados sejam mantidos no HDFS e que o cluster fique protegido contra término devido a flutuações do mercado spot, proporcionando ao mesmo tempo redução de custos decorrentes da execução de grupos de instâncias de tarefa como instâncias spot.

Testes de aplicativos

Ao testar uma nova aplicação a fim de prepará-la para inicialização em um ambiente de produção, você pode executar o cluster inteiro (grupos de instâncias primárias, centrais e de tarefa) como instâncias spot para reduzir os custos de testes.

Calcular a capacidade necessária do HDFS de um cluster

A quantidade de armazenamento no HDFS disponível para o cluster depende dos seguintes fatores:

- O número de instâncias do Amazon EC2 usadas para nós centrais.
- A capacidade de armazenamento de instância do Amazon EC2 para o tipo de instância usado. Para obter mais informações sobre volumes de armazenamento de instâncias, consulte Armazenamento de instâncias do Amazon Amazon EC2 no Guia do usuário do Amazon EC2.

- Do número e do tamanho dos volumes do Amazon EBS anexados a nós centrais.
- De um fator de replicação, que explica como cada bloco de dados é armazenado no HDFS para redundância semelhante ao RAID. Por padrão, o fator de replicação é de três para um cluster de 10 ou mais nós core, dois para um cluster com 4 a 9 nós core e um para um cluster de três nós ou menos.

Para calcular a capacidade do HDFS de um cluster, para cada nó central, adicione a capacidade do volume de armazenamento de instância à capacidade de armazenamento do Amazon EBS (se usado). Multiplique o resultado pelo número de nós core e, em seguida, divida o total pelo fator de replicação com base no número de nós core. Por exemplo, um cluster com 10 nós centrais do tipo i2.xlarge que tem 800 GB de armazenamento de instância, sem nenhum volume do Amazon EBS anexado, tem um total de aproximadamente 2.666 GB disponíveis para o HDFS (10 nós x 800 GB ÷ 3, que é o fator de replicação).

Se o valor calculado de capacidade do HDFS for menor que os seus dados, você poderá aumentar a quantidade de armazenamento do HDFS das seguintes maneiras:

- Criando um cluster com volumes do Amazon EBS adicionais ou adicionando grupos de instâncias com volumes do Amazon EBS anexados a um cluster atual
- Adicionando mais nós core
- Escolhendo um tipo de instância do Amazon EC2 com maior capacidade de armazenamento
- Usando a compactação de dados
- Alterando as definições de configuração do Hadoop para reduzir o fator de replicação

A redução do fator de replicação deve ser usada com cautela, pois ela reduz a redundância dos dados do HDFS e a capacidade do cluster de se recuperar de blocos do HDFS perdidos ou corrompidos.

Configurar registro em log e depuração do cluster

Uma das questões a ser decidida quando você planeja o seu cluster é quanto de suporte à depuração você deseja disponibilizar. Quando você está desenvolvendo um aplicativo de processamento de dados pela primeira vez, recomendamos testar o aplicativo em um cluster processando um subconjunto pequeno, mas representativo, dos seus dados. Ao fazer isso, você provavelmente vai aproveitar todas as ferramentas de depuração que o Amazon EMR oferece, tais como o arquivamento de arquivos de log para o Amazon S3.

Uma vez concluído o desenvolvimento, e com o aplicativo de processamento de dados totalmente em produção, você pode optar por reduzir a depuração. Dessa forma, você pode economizar no custo do armazenamento de arquivos de log no Amazon S3 e reduzir a carga de processamento no cluster, pois ele não precisa mais gravar o estado no Amazon S3. O risco, obviamente, é que se ocorrer algum problema, você terá menos ferramentas disponíveis para investigar o erro.

Arquivos de log padrão

Por padrão, cada cluster grava os arquivos de log no nó primário. Esses são gravados no diretório `/mnt/var/log/`. Você pode acessá-los usando o SSH para se conectar ao nó primário, como descrito em Conectar-se ao nó primário usando SSH.

> **Note**
>
> Se você usa o Amazon EMR versão 6.8.0 ou anterior, os arquivos de log são salvos no Amazon S3 durante o término do cluster, então não é possível acessar os arquivos de log após o término do nó primário. O Amazon EMR libera a versão 6.9.0 e versões posteriores arquiva os registros no Amazon S3 durante a redução da escala verticalmente do cluster, de forma que os arquivos de log gerados no cluster persistam mesmo após o término do nó.

Não é necessário habilitar nada para fazer com que os arquivos de log sejam gravados no nó primário. Esse é o comportamento padrão do Amazon EMR e do Hadoop.

Um cluster gera vários tipos de arquivos de log, incluindo:

- Logs de etapa: esses logs são gerados pelo serviço do Amazon EMR e contêm informações sobre o cluster e os resultados de cada etapa. Os arquivos de log são armazenados no diretório `/mnt/var/log/hadoop/steps/` no nó primário. Cada etapa registra seus resultados em um subdiretório numerado separado: `/mnt/var/log/hadoop/steps/s-`*`stepId1`*`/` para a primeira etapa, `/mnt/var/log/hadoop/steps/s-`*`stepId2`*`/` para a segunda etapa, e assim por diante. Os identificadores de etapa de 13 caracteres (por exemplo, IdEtapa1, IdEtapa2) são exclusivos dos clusters.
- Registros de componentes do Hadoop e do YARN — Os registros dos componentes associados ao Apache YARN e MapReduce, por exemplo, estão contidos em pastas separadas em. `/mnt/var/log` Os locais dos arquivos de log para os componentes do Hadoop sob `/mnt/var/log` são os seguintes: hadoop-hdfs, hadoop-mapreduce, hadoop-httpfs e hadoop-yarn. O hadoop-state-pusher diretório é para a saída do processo de envio de estado do Hadoop.

- Logs de ação de bootstrap: se seu trabalho utiliza ações de bootstrap, os resultados dessas ações são registrados em logs. Os arquivos de log são armazenados em /mnt/var/log/bootstrap-actions no nó primário. Cada ação de bootstrap registra seus resultados em um subdiretório numerado separado: `/mnt/var/log/bootstrap-actions/1/` para a primeira ação de bootstrap, `/mnt/var/log/bootstrap-actions/2/` para a segunda ação de bootstrap, e assim por diante.
- Logs de estado de instância: esses logs fornecem informações sobre a CPU, o estado da memória e os threads do coletor de lixo do nó. Os arquivos de log são armazenados em `/mnt/var/log/instance-state/` no nó primário.

Arquivamento dos arquivos de log no Amazon S3

> **Note**
>
> Atualmente não é possível usar a agregação de logs para o Amazon S3 com o utilitário `yarn logs`.

O Amazon EMR libera a versão 6.9.0 e versões posteriores arquiva os registros no Amazon S3 durante a redução da escala verticalmente do cluster, de forma que os arquivos de log gerados no cluster persistam mesmo após o término do nó. Esse comportamento é habilitado automaticamente, então não é necessário ativá-lo. Para as versões 6.8.0 e anteriores do Amazon EMR, é possível configurar um cluster para arquivar periodicamente os arquivos de log armazenados no nó primário no Amazon S3. Isso garante que os arquivos de log estarão disponíveis depois que o cluster for terminado, seja por meio de desligamento normal, seja devido a um erro. O Amazon EMR arquiva os arquivos de log no Amazon S3 em intervalos de cinco minutos.

Para que os arquivos de log sejam arquivados no Amazon S3 para o Amazon EMR 6.8.0 e versões anteriores, você deve habilitar esse recurso ao executar o cluster. Você pode fazer isso usando o console, a CLI ou a API. Por padrão, os clusters executados por meio do console têm a funcionalidade de registro em log habilitada. Para clusters executados usando a CLI ou a API, o registro em log no Amazon S3 deve ser habilitado manualmente.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Arquivar arquivos de log no Amazon S3 usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Logs do cluster, marque a caixa de seleção Publicar logs específicos do cluster no Amazon S3.
4. No campo Local do Amazon S3, digite (ou navegue até) o caminho do Amazon S3 onde os logs serão armazenados. Se você digitar o nome de uma pasta que não existe no bucket, o Amazon S3 a criará.

   Quando esse valor é definido, o Amazon EMR copia os arquivos de log das instâncias do EC2 no cluster para o Amazon S3. Isso evita que os arquivos de log sejam perdidos quando o cluster é encerrado e om EC2 termina as instâncias que hospedam o cluster. Esses logs são úteis para auxiliar na solução de problemas. Para obter mais informações, consulte View log files.
5. Opcionalmente, marque a caixa de seleção Criptografar logs específicos do cluster. Em seguida, selecione uma AWS KMS chave na lista, insira o ARN da chave ou crie uma nova chave. Essa opção só está disponível no Amazon EMR versão 5.30.0 e posteriores, excluindo a versão 6.0.0. Para usar essa opção, adicione permissão AWS KMS para seu perfil de instância do EC2 e função do Amazon EMR. Para ter mais informações, consulte Como criptografar arquivos de log armazenados no Amazon S3 com uma chave do AWS KMS gerenciada pelo cliente.
6. Escolha qualquer outra opção que se aplique ao cluster.
7. Para iniciar o cluster, escolha Criar cluster.

Old console

Arquivar arquivos de log no Amazon S3 usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Escolha Go to advanced options (Ir para opções avançadas).
4. Na seção Opções gerais, no campo Registro em log, aceite a opção padrão: Ativado.

   Isso determina se o Amazon EMR captura dados de log detalhados para o Amazon S3. Você só pode definir esta opção quando o cluster é criado. Para ter mais informações, consulte Exibir arquivos de log do .
5. No campo Pasta do S3, digite (ou navegue até) um caminho do Amazon S3 para armazenar seus logs. Você também pode permitir que o console gere um caminho do Amazon S3. Se você digitar o nome de uma pasta que não existe no bucket, ela será criada.

   Quando esse valor é definido, o Amazon EMR copia os arquivos de log das instâncias do EC2 no cluster para o Amazon S3. Isso evita que os arquivos de log sejam perdidos quando o cluster é encerrado e as instâncias do EC2 que hospedam o cluster são encerradas. Esses logs são úteis para auxiliar na solução de problemas.

   Para obter mais informações, consulte View log files.
6. No campo Criptografia de registros, selecione Criptografar registros armazenados no S3 com uma chave gerenciada pelo cliente AWS KMS. Em seguida, selecione uma chave AWS KMS na lista ou insira um ARN da chave. Você também pode criar uma nova AWS KMS chave.

   Essa opção só está disponível no Amazon EMR versão 5.30.0 e posteriores, excluindo a versão 6.0.0. Para usar essa opção, adicione permissão ao AWS KMS para seu perfil de instância do EC2 e perfil do Amazon EMR. Para ter mais informações, consulte Como criptografar arquivos de log armazenados no Amazon S3 com uma chave do AWS KMS gerenciada pelo cliente.
7. Prossiga com a criação do cluster conforme descrito em Planejar e configurar clusters.

CLI

Para arquivar arquivos de log no Amazon S3 com o AWS CLI

Para arquivar arquivos de log no Amazon S3 usando o AWS CLI, digite o `create-cluster` comando e especifique o caminho de log do Amazon S3 usando o parâmetro. `--log-uri`

1. Para arquivar os arquivos de log no Amazon S3, digite o seguinte comando e substitua *myKey* pelo nome do par de chaves do EC2.

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 --log-
uri s3://DOC-EXAMPLE-BUCKET/logs --applications Name=Hadoop Name=Hive Name=Pig
 --use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge --
instance-count 3
```

2. Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você não criou a perfil de serviço do Amazon EMR padrão e o perfil de instância do EC2, insira `aws emr create-default-roles` para criá-los antes de digitar o subcomando `create-cluster`.

Como criptografar arquivos de log armazenados no Amazon S3 com uma chave do AWS KMS gerenciada pelo cliente

Com o Amazon EMR versão 5.30.0 e posterior (exceto o Amazon EMR 6.0.0), você pode criptografar arquivos de log armazenados no Amazon S3 com uma chave gerenciada pelo cliente KMS. AWS Para habilitar essa opção no console, siga as etapas em Arquivamento dos arquivos de log no Amazon S3. O perfil de instância do Amazon EC2 e o perfil do Amazon EMR devem atender aos seguintes pré-requisitos:

- O perfil de instância do Amazon EC2 usado para o cluster deve ter permissão para usar `kms:GenerateDataKey`.
- O perfil do Amazon EMR usada para o cluster deve ter permissão para usar `kms:DescribeKey`.

- O perfil da instância do Amazon EC2 e a função do Amazon EMR devem ser adicionados à lista de usuários-chave da chave gerenciada pelo cliente AWS KMS especificada, conforme demonstrado nas etapas a seguir:
  1. Abra o console AWS Key Management Service (AWS KMS) em https://console.aws.amazon.com/kms.
  2. Para alterar a AWS região, use o seletor de região no canto superior direito da página.
  3. Selecione o alias da chave do KMS a ser modificada.
  4. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
  5. Na caixa de diálogo Adicionar usuários de chave, selecione o perfil de instância do Amazon EC2 e o perfil do Amazon EMR.
  6. Escolha Adicionar.

Para obter mais informações, consulte Funções de serviço do IAM usadas pelo Amazon EMR e Como usar políticas de chaves no guia do desenvolvedor do AWS Key Management Service.

Agregar logs no Amazon S3 usando a AWS CLI

> **Note**
>
> Atualmente não é possível usar a agregação de logs com o utilitário `yarn logs`. Você só pode usar a agregação compatível com esse procedimento.

A agregação de logs (Hadoop 2.x) compila os logs de todos os contêineres de um aplicativo individual em um único arquivo. Para habilitar a agregação de logs para o Amazon S3 usando AWS CLI o, você usa uma ação de bootstrap na inicialização do cluster para habilitar a agregação de logs e especificar o bucket para armazenar os logs.

- Para habilitar a agregação de logs, crie o arquivo de configuração chamado `myConfig.json`, que contém o seguinte:

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.log-aggregation-enable": "true",
      "yarn.log-aggregation.retain-seconds": "-1",
```

```
      "yarn.nodemanager.remote-app-log-dir": "s3:\/\/DOC-EXAMPLE-BUCKET\/logs"
    }
  }
]
```

Digite o seguinte comando, substituindo *myKey* pelo nome do par de chaves do EC2. Além disso, você pode substituir os textos em vermelho por suas próprias configurações.

```
aws emr create-cluster --name "Test cluster" \
--release-label emr-7.1.0 \
--applications Name=Hadoop \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-type m5.xlarge \
--instance-count 3 \
--configurations file://./myConfig.json
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você não tiver criado o perfil de serviço padrão do EMR e o perfil de instância do EC2, execute `aws emr create-default-roles` para criá-los antes de executar o subcomando `create-cluster`.

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte Referência de AWS CLI comandos.

Locais de log

A lista a seguir inclui todos os tipos de log e seus locais no Amazon S3. Use-os para solucionar problemas do Amazon EMR.

Logs de etapa

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/steps/<step-id>/
```

Logs de aplicações

`s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/containers/`

Esse local inclui contêiner `stderr` e `stdout`, `directory.info`, `prelaunch.out` e logs `launch_container.sh`.

Logs do gerenciador de recursos

`s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/applications/hadoop-yarn/`

HDFS do Hadoop

`s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/applications/hadoop-hdfs/`

Esse local inclui NameNode, DataNode, e TimelineServer registros do YARN.

Logs do gerenciador de nós

`s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/applications/hadoop-yarn/`

Logs de estado de instância

`s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/daemons/instance-state/`

Logs de provisionamento do Amazon EMR

`s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/provision-node/*`

Logs do Hive

`s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/applications/hive/*`

- Para encontrar logs do Hive no cluster, remova o asterisco (*) e anexe `/var/log/hive/` ao link acima.
- Para encontrar HiveServer 2 registros, remova o asterisco (*) e anexe `var/log/hive/hiveserver2.log` ao link acima.
- Para encontrar logs do HiveCLI, remova o asterisco (*) e anexe `/var/log/hive/user/hadoop/hive.log` ao link acima.

- Para encontrar os logs do Hive Metastore Server, remova o asterisco (`*`) e anexe `/var/log/hive/user/hive/hive.log` ao link acima.

Se sua falha estiver no nó primário ou no nó de tarefa da aplicação Tez, forneça logs do contêiner Hadoop apropriado.

Habilitar ferramenta de depuração

A ferramenta de depuração permite procurar mais facilmente os arquivos de log do console do Amazon EMR. Para ter mais informações, consulte Visualizar arquivos de log na ferramenta de depuração. Quando a depuração é habilitada em um cluster, o Amazon EMR arquiva os arquivos de log no Amazon S3 e, em seguida, indexa esses arquivos. Você pode usar o console para pesquisar nos logs das etapas, trabalhos, tarefas e tentativas de tarefas do cluster de uma maneira intuitiva.

Para usar a ferramenta de depuração no console do Amazon EMR, você deve habilitar a depuração ao iniciar o cluster usando o console, a CLI ou a API. O novo console do Amazon EMR não oferece a ferramenta de depuração.

Old console

Ativar a ferramenta de depuração usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Escolha Go to advanced options (Ir para opções avançadas).
4. Na seção Cluster Configuration (Configuração do cluster), no campo Logging (Registro), escolha Enabled (Habilitado). Não é possível habilitar a depuração sem habilitar o registro em log.
5. No campo Local do S3 da pasta de logs, digite o caminho do Amazon S3 onde os logs serão armazenados.
6. No campo Debugging (Depuração), escolha Enabled (Habilitado). A opção de depuração cria uma troca no Amazon SQS para publicar mensagens de depuração no serviço de back-end do Amazon EMR. Podem ser cobrados encargos pela publicação de mensagens nessa troca. Para obter mais informações, consulte a página do produto do Amazon SQS.
7. Prossiga com a criação do cluster conforme descrito em Planejar e configurar clusters.

AWS CLI

Para ativar a ferramenta de depuração com o AWS CLI

Para habilitar a depuração usando o AWS CLI, digite o `create-cluster` subcomando com o parâmetro. `--enable-debugging` Especifique também o parâmetro `--log-uri` ao habilitar a depuração.

- Para ativar a depuração usando o AWS CLI, digite o comando a seguir e substitua *myKey* pelo nome do seu par de chaves do EC2.

```
aws emr create-cluster --name "Test cluster" \
--release-label emr-7.1.0 \
--log-uri s3://DOC-EXAMPLE-BUCKET/logs \
--enable-debugging \
--applications Name=Hadoop Name=Hive Name=Pig \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-type m5.xlarge \
--instance-count 3
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é executado, e as instâncias restantes são executadas como nós centrais. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você não criou a função de serviço do EMR padrão e o perfil de instância do EC2, digite aws `emr create-default-roles` para criá-los antes de digitar o subcomando `create-cluster`.

API

Ativar a ferramenta de depuração usando a API do Amazon EMR

- Habilite a depuração usando a configuração do Java SDK a seguir.

```
    StepFactory stepFactory = new StepFactory();
StepConfig enabledebugging = new StepConfig()
```

```
    .withName("Enable debugging")
    .withActionOnFailure("TERMINATE_JOB_FLOW")
    .withHadoopJarStep(stepFactory.newEnableDebuggingStep());
```

Neste exemplo, `new StepFactory()` usa `us-east-1` como a região padrão. Se o cluster for iniciado em uma região diferente, será necessário especificar a região usando `new StepFactory("region.elasticmapreduce")`, como `new StepFactory("ap-northeast-2.elasticmapreduce")`.

Informações sobre as opções de depuração

O Amazon EMR versões 4.1.0 a 5.27.0 oferece suporte à depuração em todas as regiões. Outras versões do Amazon EMR não oferecem suporte à opção de depuração. A partir de 23 de janeiro de 2023, o Amazon EMR descontinuará a ferramenta de depuração em todas as versões.

O Amazon EMR cria uma fila do Amazon SQS para processar os dados de depuração. Podem ser cobrados encargos pelas mensagens. No entanto, o Amazon SQS disponibiliza um nível gratuito de até 1 milhão de solicitações. Para ter mais informações, consulte https://aws.amazon.com/sqs.

A depuração requer o uso de perfis. É necessário que o perfil de serviço e o perfil de instância permitam utilizar todas as operações de API do Amazon SQS. Se seus perfis estão anexados às políticas gerenciadas do Amazon EMR, você não precisará modificar os perfis. Se você tiver funções personalizadas, precisará adicionar as permissões `sqs:*`. Para ter mais informações, consulte Configurar perfis de serviço do IAM para permissões do Amazon EMR aos serviços e recursos da AWS.

Clusters de etiqueta

Pode ser conveniente categorizar seus AWS recursos de maneiras diferentes; por exemplo, por finalidade, proprietário ou ambiente. Você pode conseguir isso no Amazon EMR atribuindo metadados personalizados aos clusters do Amazon EMR com o uso de etiquetas. Uma tag consiste em uma chave e um valor, ambos definidos por você. Para o Amazon EMR, o cluster é o nível de recursos que você pode marcar. Por exemplo, é possível definir um conjunto de tags para os clusters da sua conta que ajuda a rastrear o proprietário de cada cluster ou identificar um cluster de produção versus um cluster de teste. Recomendamos criar um conjunto consistente de tags para atender às necessidades da sua organização.

Ao adicionar uma etiqueta a um cluster do Amazon EMR, essa etiqueta também é propagada para cada instância do Amazon EC2 ativa associada ao cluster. Da mesma forma, quando você remove uma etiqueta de um cluster do Amazon EMR, ela é removida de cada instância do Amazon EC2 ativa associada.

> **Important**
>
> Use o console ou a CLI do Amazon EMR para gerenciar etiquetas em instâncias do Amazon EC2 que fazem parte de um cluster em vez do console ou da CLI do Amazon EC2, pois as alterações feitas no Amazon EC2 não são sincronizadas de volta com o sistema de marcação do Amazon EMR.

Você pode identificar uma instância do Amazon EC2 que faz parte de um cluster do Amazon EMR examinando as etiquetas de sistema a seguir. Neste exemplo, *CORE* é o valor para a função do grupo de instâncias e *j-12345678* é um valor de identificador de fluxo de trabalho (cluster) de exemplo:

- aws:elasticmapreduce:instance-group-role=*CORE*
- aws:elasticmapreduce:job-flow-id=*j-12345678*

> **Note**
>
> O Amazon EMR e o Amazon EC2 interpretam suas etiquetas como uma string de caracteres sem significado semântico.

Você pode trabalhar com tags usando o AWS Management Console, a CLI e a API.

Você pode adicionar etiquetas ao criar um novo cluster do Amazon EMR e pode adicionar, editar ou remover etiquetas de um cluster do Amazon EMR em execução. A edição de uma etiqueta é um conceito que se aplica ao console do Amazon EMR. Porém, usando a CLI e a API para editar uma etiqueta, você remove a etiqueta antiga e adiciona uma nova. Você pode editar chaves de tags e valores e pode remover tags de um recurso a qualquer momento durante a execução do cluster. No entanto, não é possível adicionar, editar ou remover tags de um cluster encerrado ou de instâncias encerradas que estavam anteriormente associadas a um cluster que ainda está ativo. Além disso, você pode definir o valor de uma tag como a string vazia, mas não pode definir valor de uma tag como nulo.

Se você estiver usando AWS Identity and Access Management (IAM) com suas instâncias do Amazon EC2 para permissões baseadas em recursos por tag, suas políticas do IAM são aplicadas às tags que o Amazon EMR propaga para as instâncias do Amazon EC2 de um cluster. Para que as tags do Amazon EMR se propaguem para suas instâncias do Amazon EC2, sua política do IAM para o Amazon EC2 precisa permitir permissões para chamar o Amazon EC2 e as APIs. CreateTags DeleteTags Além disso, etiquetas propagadas podem afetar as permissões baseadas em recursos do Amazon EC2. As etiquetas propagadas para o Amazon EC2 podem ser lidas como condições na sua política do IAM, exatamente como outras etiquetas do Amazon EC2. Mantenha sua política do IAM em mente ao adicionar etiquetas aos clusters do Amazon EMR, para evitar que os usuários tenham permissões incorretas para um cluster. Para evitar problemas, certifique-se de que as suas políticas do IAM não incluam condições sobre etiqueta que você também planeja usar nos seus clusters do Amazon EMR. Para obter mais informações, consulte Controlling access to Amazon EC2 resources.

Restrições de tags

As restrições básicas a seguir se aplicam a tags:

- As restrições que se aplicam aos recursos do Amazon EC2 também se aplicam ao Amazon EMR. Para ter mais informações, consulte https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Using_Tags.html#tag-restrictions.
- Não use o `aws:` prefixo nos nomes e valores das tags porque ele está reservado para AWS uso. Além disso, você não pode editar nem excluir nomes ou valores de tags com esse prefixo.
- Você não pode alterar ou editar tags em um cluster encerrado.
- Um valor de tag pode ser uma string vazia, mas não nula. Além disso, uma chave de tag não pode ser uma sequência vazia.
- Chaves e valores podem conter qualquer caractere alfabético em qualquer idioma, qualquer caractere numérico, espaço em branco, separadores invisíveis e os seguintes símbolos: _ . : / = + - @

Para obter mais informações sobre a marcação usando o AWS Management Console, consulte Como trabalhar com tags no console no Guia do usuário do Amazon EC2. Para obter mais informações sobre a marcação usando a API do Amazon EC2 ou a linha de comando, consulte a visão geral da API e da CLI no Guia do usuário do Amazon EC2.

Recursos de tag para faturamento

Você pode usar etiquetas para organizar sua AWS fatura para refletir sua própria estrutura de custos. Para fazer isso, inscreva-se para receber a fatura AWS da sua conta com os valores-chave da tag incluídos. Dessa forma, você pode organizar suas informações de faturamento por valores de chave de tag, para ver o custo dos seus recursos combinados. Embora o Amazon EMR e o Amazon EC2 tenham faturas separadas, as etiquetas de cada cluster também são colocadas em cada instância associada e, portanto, você pode usar etiquetas para vincular custos relacionados do Amazon EMR e do Amazon EC2.

Por exemplo, é possível marcar vários recursos com um nome de aplicação específico, e depois organizar suas informações de faturamento para ver o custo total daquela aplicação em vários serviços. Veja mais informações sobre alocação de recursos e uso de etiquetas no Guia do usuário do AWS Billing .

Adicionar etiquetas a um cluster

Você pode adicionar etiquetas a um cluster ao criá-lo.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Adicionar etiquetas ao criar um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Etiquetas, escolha Adicionar nova etiqueta. Especifique uma etiqueta no campo Chave. Opcionalmente, especifique uma etiqueta no campo Valor.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Old console

Adicionar etiquetas ao criar um cluster usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Na página Step 3: General Cluster Settings (Etapa 3: Configurações gerais do cluster), na seção Tags, digite uma Key (Chave) para a tag.

   Quando você começar a digitar a Key (Chave) uma nova linha aparecerá automaticamente para fornecer espaço para a próxima nova tag.
4. Opcionalmente, digite um Value (Valor) para a tag.
5. Repita as etapas anteriores para cada par de chave/valor de tag a ser adicionado ao cluster. Quando o cluster for iniciado, qualquer tag que você inserir será automaticamente associada a ele.

AWS CLI

Para adicionar tags ao criar um cluster com o AWS CLI

O exemplo a seguir demonstra como adicionar uma tag a um novo cluster usando a AWS CLI. Para adicionar tags ao criar um cluster, digite o subcomando `create-cluster` com o parâmetro `--tags`.

- Para adicionar uma tag chamada *costCenter* com um valor de chave *marketing* quando você criar um cluster, digite o comando a seguir e substitua *myKey* pelo nome do seu par de chaves do EC2.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications Name=Hadoop Name=Hive Name=Pig --tags "costCenter=marketing" --
use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge --
instance-count 3
```

  Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó principal é executado, e as instâncias restantes são executadas como nós core. Todos os nós usarão o tipo de instância especificado no comando.

> **Note**
>
> Se você não criou a função de serviço do EMR padrão e o perfil de instância do EC2, digite aws `emr create-default-roles` para criá-los antes de digitar o subcomando `create-cluster`.

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Você também pode adicionar tags a um cluster existente.

New console

Adicionar etiquetas a um cluster existente usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Etiquetas na página de detalhes do cluster, selecione Gerenciar etiquetas. Especifique uma etiqueta no campo Chave. Opcionalmente, especifique uma etiqueta no campo Valor.
4. Selecione Save Changes (Salvar alterações). A guia Etiquetas é atualizada com o novo número de etiquetas que você tem no cluster. Por exemplo, se você agora tem duas etiquetas, o rótulo da sua guia é Etiquetas (2).

Old console

Adicionar etiquetas a um cluster existente usando o console antigo

1. No console do Amazon EMR, selecione a página Lista de clusters e clique em um cluster para adicionar etiquetas.
2. Na página Cluster Details (Detalhes do cluster), no campo Tags, clique em View All/Edit (Exibir todos/Editar).
3. Na página View All/Edit (Exibir todos/Editar), clique em Add (Adicionar).

4. Clique no campo vazio da coluna Key (Chave) e digite o nome da sua chave.
5. Opcionalmente, clique no campo vazio da coluna Value (Valor) e digite o nome do seu valor.
6. Com cada nova tag iniciada, outra linha de tag vazia aparece abaixo da tag que você está editando no momento. Repita as etapas anteriores na nova linha de tag para cada tag a ser adicionada.

AWS CLI

Para adicionar tags a um cluster em execução com o AWS CLI

- Digite o subcomando `add-tags` com o parâmetro `--tag` para atribuir etiquetas a um ID do cluster. Você pode localizar o ID do cluster usando o console ou o comando `list-clusters`. Atualmente, o subcomando `add-tags` aceita apenas um ID de recurso.

  Para adicionar duas etiquetas a um cluster em execução, uma com uma chave denominada *`costCenter`* com o valor *`marketing`* e a outra com uma chave denominada *`outros`* com o valor *`contabilidade`*, digite o comando a seguir e substitua *`j-KT4XXXXXXXX1NM`* pelo ID do cluster.

  ```
  aws emr add-tags --resource-id j-KT4XXXXXXXX1NM --tag "costCenter=marketing" --
  tag "other=accounting"
  ```

  Observe que quando as tags são adicionadas usando a AWS CLI, não há saída do comando. Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Visualizar etiquetas em um cluster

Para visualizar todas as etiquetas associadas a um cluster, você pode visualizá-las no console ou na AWS CLI.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Visualizar etiquetas em um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Para visualizar todas as suas etiquetas, selecione a guia Etiquetas na página de detalhes do cluster.

Old console

Visualizar etiquetas de um cluster usando o console antigo

1. No console do Amazon EMR, selecione a página Lista de clustes e clique em um cluster para visualizar as etiquetas.
2. Na página Cluster Details (Detalhes do cluster), no campo Tags, algumas tags são exibidas aqui. Clique em View All/Edit (Exibir todos/Editar) para exibir todas as tags disponíveis no cluster.

AWS CLI

Para visualizar as tags em um cluster com o AWS CLI

Para visualizar as tags em um cluster usando o AWS CLI, digite o `describe-cluster` subcomando com o `--query` parâmetro.

- Para visualizar as tags de um cluster, digite o seguinte comando e substitua *`j-KT4XXXXXXX1NM`* pelo ID do cluster.

```
aws emr describe-cluster --cluster-id j-KT4XXXXXX1NM --query Cluster.Tags
```

A saída exibe todas as informações de tag sobre o cluster, semelhantes às seguintes:

```
Value: accounting     Value: marketing
Key: other            Key: costCenter
```

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Remover etiquetas de um cluster

Se não precisa mais de uma tag, pode removê-la do cluster.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Remover etiquetas de um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Etiquetas na página de detalhes do cluster, selecione Gerenciar etiquetas.
4. Escolha Remover para cada par de chave-valor que você deseja remover.
5. Escolha Salvar alterações.

Old console

Remover etiquetas em um cluster usando o console antigo

1. No console do Amazon EMR, selecione a página Lista de clusters e clique em um cluster para remover etiquetas.
2. Na página Cluster Details (Detalhes do cluster), no campo Tags, clique em View All/Edit (Exibir todos/Editar).
3. Na caixa de diálogo View All/Edit (Exibir todos/Editar), clique no ícone X próximo à tag a ser excluída e clique em Save (Salvar).

4. (Opcional) Repita a etapa anterior para cada par de chave-valor de etiqueta a ser removido do cluster.

AWS CLI

Para remover tags em um cluster com o AWS CLI

Digite o subcomando `remove-tags` com o parâmetro `--tag-keys`. Ao remover uma tag, apenas o nome da chave é necessário.

- Para remover uma tag de um cluster, digite o seguinte comando e substitua *j-KT4XXXXXXXX1NM* pelo ID do cluster.

```
aws emr remove-tags --resource-id j-KT4XXXXXX1NM --tag-keys "costCenter"
```

> **Note**
>
> No momento, não é possível remover várias tags usando um único comando.

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Integração de drivers e aplicações de terceiros

Você pode executar várias aplicações conhecidas de big data no Amazon EMR com preços de utilitários. Isso significa que você paga uma taxa por hora nominal adicional pelo aplicativo de terceiros enquanto seu cluster está em execução. Isso permite que você use o aplicativo sem precisar adquirir uma licença anual. As seções a seguir descrevem algumas das ferramentas que você pode usar com o EMR.

Tópicos

- Usar ferramentas de business intelligence com o Amazon EMR

Usar ferramentas de business intelligence com o Amazon EMR

Você pode usar ferramentas populares de business intelligence, como Microsoft Excel, MicroStrategyQlikView, e Tableau, com o Amazon EMR para explorar e visualizar seus dados. Muitas dessas ferramentas exigem um driver do ODBC (Open Database Connectivity) ou do JDBC (Java Database Connectivity). Para baixar e instalar os drivers mais recentes, consulte http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/.

Para encontrar versões mais antigas dos drivers, consulte http://awssupportdatasvcs.com/bootstrap-actions/Simba/.

Segurança no Amazon EMR

Segurança e conformidade são uma responsabilidade com a qual você compartilha AWS. Esse modelo de responsabilidade compartilhada pode ajudar a aliviar sua carga operacional, pois AWS opera, gerencia e controla os componentes do sistema operacional host e da camada de virtualização até a segurança física das instalações nas quais os clusters do EMR operam. Você assume a responsabilidade, o gerenciamento e a atualização dos clusters do Amazon EMR, além de configurar o software do aplicativo e os controles de segurança AWS fornecidos. Essa diferenciação de responsabilidade é comumente chamada de segurança na nuvem versus segurança na nuvem.

- Segurança da nuvem — AWS é responsável por proteger a infraestrutura que é Serviços da AWS executada AWS. AWS também fornece serviços que você pode usar com segurança. Auditores de terceiros testam e verificam regularmente a eficácia da nossa segurança como parte dos compliance programs AWS. Para saber mais sobre os programas de conformidade que se aplicam ao Amazon EMR, consulte Serviços da AWS o escopo por programa de conformidade.
- Segurança na nuvem — você também é responsável por realizar todas as tarefas de configuração e gerenciamento de segurança necessárias para proteger um cluster do Amazon EMR. Os clientes que implantam um cluster do Amazon EMR são responsáveis pelo gerenciamento do software aplicativo instalado nas instâncias e pela configuração dos recursos AWS fornecidos, como grupos de segurança, criptografia e controle de acesso, de acordo com seus requisitos, leis e regulamentos aplicáveis.

Esta documentação ajuda você a entender como aplicar o modelo de responsabilidade compartilhada ao usar o Amazon EMR. Os tópicos deste capítulo mostram como configurar o Amazon EMR e usar outros Serviços da AWS para atender aos seus objetivos de segurança e conformidade.

Segurança de rede e infraestrutura

Como um serviço gerenciado, o Amazon EMR é protegido pelos procedimentos AWS globais de segurança de rede descritos no whitepaper Amazon Web Services: Visão geral dos processos de segurança. AWS os serviços de proteção de rede e infraestrutura oferecem proteções refinadas nos limites do host e da rede. O Amazon EMR oferece suporte Serviços da AWS e recursos de aplicativos que atendem aos requisitos de conformidade e proteção de rede.

- Os grupos de segurança do Amazon EC2 atuam como um firewall virtual para instâncias de cluster do Amazon EMR, limitando o tráfego de entrada e saída da rede. Para obter mais informações, consulte Controlar o tráfego de rede com grupos de segurança.
- O bloqueio de acesso público (BPA) do Amazon EMR impede que você lance um cluster em uma sub-rede pública se o cluster tiver uma configuração de segurança que permita tráfego de entrada de endereços IP públicos em uma porta. Para obter mais informações, consulte Como usar o Amazon EMR para bloquear o acesso público.
- O Secure Shell (SSH) ajuda a fornecer uma forma segura para os usuários se conectarem à linha de comando em instâncias de cluster. Você também pode usar o SSH para visualizar as interfaces da web que os aplicativos hospedam no nó principal de um cluster. Para obter mais informações, consulte Usar um key pair do EC2 para credenciais SSH e Conectar-se a um cluster.

Atualizações da AMI padrão do Amazon Linux para Amazon EMR

> **Important**
>
> Os clusters do EMR que executam imagens de máquina da Amazon (AMIs) do Amazon Linux ou do Amazon Linux 2 usam o comportamento padrão do Amazon Linux e não baixam nem instalam automaticamente atualizações importantes e críticas do kernel que exigem reinicialização. É o mesmo comportamento de outras instâncias do Amazon EC2 que executam a AMI padrão do Amazon Linux. Se novas atualizações de software do Amazon Linux que exigem reinicialização (como atualizações do kernel, NVIDIA e CUDA) forem disponibilizadas após o lançamento de uma versão do Amazon EMR, as instâncias de cluster do Amazon EMR que executam a AMI padrão não baixarão nem instalarão essas atualizações automaticamente. Para obter atualizações do kernel, você pode personalizar sua AMI do Amazon EMR para usar a AMI do Amazon Linux mais recente.

Dependendo da postura de segurança de seu aplicativo e o período em que um cluster é executado, você pode optar por reinicializar periodicamente seu cluster para aplicar atualizações de segurança, ou criar uma ação de bootstrap para personalizar a instalação de pacotes e atualizações. Você também pode escolher testar e, em seguida, instalar determinadas atualizações de segurança nas instâncias de cluster em execução. Para ter mais informações, consulte Usar a AMI padrão do Amazon Linux para Amazon EMR. Observe que sua configuração de rede deve permitir a saída de HTTP e HTTPS para repositórios Linux no Amazon S3, caso contrário, as atualizações de segurança não serão bem-sucedidas.

AWS Identity and Access Management com o Amazon EMR

AWS Identity and Access Management (IAM) é um AWS serviço que ajuda o administrador a controlar com segurança o acesso aos AWS recursos. Os administradores do IAM controlam quem pode ser autenticado (conectado) e autorizado (ter permissões) para utilizar os recursos do Amazon EMR. As identidades do IAM incluem usuários, grupos e funções. Uma função do IAM é semelhante à de um usuário do IAM, mas não está associada a uma pessoa específica e deve ser assumida por qualquer usuário que precise de permissões. Para obter mais informações, consulte AWS Identity and Access Management para o Amazon EMR. O Amazon EMR usa várias funções do IAM para ajudá-lo a implementar controles de acesso para clusters do Amazon EMR. O IAM é um AWS serviço que você pode usar sem custo adicional.

- Função do IAM para o Amazon EMR (função do EMR) — controla como o serviço Amazon EMR é capaz de acessar outros Serviços da AWS em seu nome, como provisionar instâncias do Amazon EC2 quando o cluster do Amazon EMR é lançado. Para obter mais informações, consulte Configurar funções de serviço do IAM para permissões Serviços da AWS e recursos do Amazon EMR.
- Função do IAM para instâncias EC2 de cluster (perfil de instância EC2) — uma função que é atribuída a cada instância EC2 no cluster do Amazon EMR quando a instância é iniciada. Os processos de aplicativos executados no cluster usam essa função para interagir com outros Serviços da AWS, como o Amazon S3. Para obter mais informações, consulte Função do IAM para instâncias EC2 do cluster.
- Função do IAM para aplicativos (função de tempo de execução) — uma função do IAM que você pode especificar ao enviar um trabalho ou uma consulta para um cluster do Amazon EMR. O trabalho ou consulta que você envia ao seu cluster do Amazon EMR usa a função de tempo de execução para acessar AWS recursos, como objetos no Amazon S3. Você pode especificar perfis de runtime com o Amazon EMR para trabalhos do Spark e do Hive. Ao usar funções de tempo de execução, você pode isolar trabalhos em execução no mesmo cluster usando diferentes funções do IAM. Para obter mais informações, consulte Usando a função do IAM como função de tempo de execução com o Amazon EMR.

As identidades da força de trabalho referem-se aos usuários que criam ou operam cargas de trabalho em. AWS O Amazon EMR fornece suporte para identidades da força de trabalho com o seguinte:

- AWS O centro de identidade do IAM (Idc) é o recomendado AWS service (Serviço da AWS) para gerenciar o acesso do usuário aos AWS recursos. É um único local onde você pode atribuir identidades à sua força de trabalho e acesso consistente a várias AWS contas e aplicativos. O Amazon EMR oferece suporte às identidades da força de trabalho por meio da propagação confiável de identidades. Com um recurso confiável de propagação de identidade, um usuário pode entrar no aplicativo e esse aplicativo pode passar a identidade do usuário Serviços da AWS para outra pessoa para autorizar o acesso a dados ou recursos. Para obter mais informações, consulte Habilitando o suporte para o centro de identidade do AWS IAM com o Amazon EMR.

  O Lightweight Directory Access Protocol (LDAP) é um protocolo de aplicativo padrão do setor aberto, independente do fornecedor, para acessar e manter informações sobre usuários, sistemas, serviços e aplicativos na rede. O LDAP é comumente usado para autenticação de usuários em servidores de identidade corporativa, como o Active Directory (AD) e o OpenLDAP. Ao habilitar o LDAP com clusters do EMR, você permite que os usuários usem suas credenciais existentes para autenticar e acessar clusters. Para obter mais informações, consulte Habilitar o suporte para LDAP com o Amazon EMR.

  O Kerberos é um protocolo de autenticação de rede projetado para fornecer autenticação forte para aplicativos cliente/servidor usando criptografia de chave secreta. Quando você usa o Kerberos, o Amazon EMR configura o Kerberos para os aplicativos, componentes e subsistemas que ele instala no cluster para que sejam autenticados entre si. Para acessar um cluster com o Kerberos configurado, um kerberos principal deve estar presente no Kerberos Domain Controller (KDC). Para obter mais informações, consulte Como ativar o suporte para Kerberos com o Amazon EMR.

Clusters de inquilino único e multilocatário

Por padrão, um cluster é configurado para uma única locação com o perfil da Instância EC2 como a identidade do IAM. Em um cluster de inquilino único, cada trabalho tem acesso total e completo ao cluster e o acesso a todos os Serviços da AWS recursos é feito com base no perfil da instância do EC2. Em um cluster multilocatário, os inquilinos são isolados uns dos outros e não têm acesso total e completo aos clusters e às instâncias EC2 do cluster. A identidade em clusters multilocatários são as funções de tempo de execução ou as identificações da força de trabalho. Em um cluster multilocatário, você também pode ativar o suporte para controle de acesso refinado (FGAC) por meio do Apache Ranger. AWS Lake Formation Em um cluster com funções de tempo de execução ou FGAC habilitadas, o acesso ao perfil da instância EC2 também é desabilitado via iptables.

> **Important**
>
> Qualquer usuário que tenha acesso a um cluster de locatário único pode instalar qualquer software no sistema operacional (SO) Linux, alterar ou remover componentes de software instalados pelo Amazon EMR e impactar as instâncias do EC2 que fazem parte do cluster. Se você quiser garantir que os usuários não possam instalar ou alterar as configurações de um cluster do Amazon EMR, recomendamos que você habilite a multilocação para o cluster. Você pode habilitar a multilocação em um cluster ativando o suporte para a função de tempo de execução, a central de identidade AWS do IAM, o Kerberos ou o LDAP.

Proteção de dados

Com AWS, você controla seus dados usando Serviços da AWS ferramentas para determinar como os dados são protegidos e quem tem acesso a eles. Serviços como AWS Identity and Access Management (IAM) permitem que você gerencie com segurança o acesso Serviços da AWS e os recursos. AWS CloudTrail permite detecção e auditoria. O Amazon EMR facilita a criptografia de dados em repouso no Amazon S3 usando chaves gerenciadas por você AWS ou totalmente gerenciadas por você. O Amazon EMR também oferece suporte para habilitar a criptografia de dados em trânsito. Para obter mais informações, consulte criptografar dados em repouso e em trânsito.

Controle de acesso a dados

Com o controle de acesso aos dados, você pode controlar quais dados uma identidade do IAM ou uma identidade da força de trabalho pode acessar. O Amazon EMR oferece suporte aos seguintes controles de acesso:

- Políticas baseadas em identidade do IAM — gerencie permissões para funções do IAM que você usa com o Amazon EMR. As políticas do IAM podem ser combinadas com a marcação para controlar o acesso em uma cluster-by-cluster base. Para obter mais informações, consulte AWS Identity and Access Management para o Amazon EMR.
- AWS Lake Formationcentraliza o gerenciamento de permissões de seus dados e facilita o compartilhamento em toda a organização e externamente. Você pode usar o Lake Formation para permitir acesso refinado em nível de coluna a bancos de dados e tabelas no Glue Data Catalog. AWS Para obter mais informações, consulte Como usar AWS Lake Formation com o Amazon EMR.

- O acesso ao Amazon S3 concede identidades de mapas e identidades de mapas em diretórios como o Active Directory ou AWS Identity and Access Management (IAM) principals para conjuntos de dados no S3. Além disso, o acesso ao S3 concede ao log a identidade do usuário final e o aplicativo usado para acessar os dados do S3. AWS CloudTrail Para obter mais informações, consulte Uso de concessões de acesso do Amazon S3 com o Amazon EMR.
- O Apache Ranger é uma estrutura para habilitar, monitorar e gerenciar a segurança abrangente de dados em toda a plataforma Hadoop. O Amazon EMR oferece suporte ao controle de acesso refinado baseado no Apache Ranger para o Apache Hive Metastore e o Amazon S3. Para obter mais informações, consulte Integrar o Apache Ranger com o Amazon EMR.

Usar configurações de segurança para definir a segurança do cluster

Você pode usar as configurações de segurança do Amazon EMR para configurar a criptografia de dados, a autenticação Kerberos e a autorização do Amazon S3 para o EMRFS nos clusters. Primeiro, crie uma configuração de segurança. Em seguida, a configuração de segurança fica disponível para uso e reutilização ao criar clusters.

Você pode usar o AWS Management Console, o AWS Command Line Interface (AWS CLI) ou os AWS SDKs para criar configurações de segurança. Você também pode usar um AWS CloudFormation modelo para criar uma configuração de segurança. Para obter mais informações, consulte o Guia AWS CloudFormation do usuário e a referência do modelo para AWS::EMR::SecurityConfiguration.

Tópicos

- Criar uma configuração de segurança
- Especificar uma configuração de segurança para um cluster

Criar uma configuração de segurança

Este tópico aborda os procedimentos gerais para criar uma configuração de segurança com o console do Amazon EMR e o AWS CLI, seguido por uma referência para os parâmetros que incluem criptografia, autenticação e funções do IAM para o EMRFS. Para obter mais informações sobre esses recursos, consulte os tópicos a seguir:

- Criptografar dados em repouso e em trânsito

- Usar o Kerberos para autenticação com o Amazon EMR
- Configurar perfis do IAM para solicitações do EMRFS para o Amazon S3

Para criar uma configuração de segurança usando o console

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. No painel de navegação, escolha Security Configurations (Configurações de segurança), Create security configuration (Criar configuração de segurança).
3. Digite um nome em Name (Nome) para a configuração de segurança.
4. Escolha opções Criptografia e Autenticação conforme descrito nas seções abaixo e escolha Criar.

Para criar uma configuração de segurança usando o AWS CLI

- Use o comando `create-security-configuration` conforme mostrado no exemplo a seguir.
  - Em *`SecConfigNome`*, especifique o nome da configuração de segurança. Trata-se do nome especificado por você ao criar um cluster que usa essa configuração de segurança.
  - Para *`SecConfigDef`*, especifique uma estrutura JSON em linha ou o caminho para um arquivo JSON local, como *`file://MySecConfig.json`*. Os parâmetros JSON definem opções de Criptografia, Perfis do IAM para acesso do EMRFS ao Amazon S3 e Autenticação conforme descrito nas seções abaixo.

```
aws emr create-security-configuration --name "SecConfigName" --security-
configuration SecConfigDef
```

Configurar criptografia de dados

Antes de configurar a criptografia em uma configuração de segurança, crie as chaves e os certificados usados na criptografia. Para obter mais informações, consulte Fornecer chaves para criptografia de dados em repouso com o Amazon EMR e Fornecer certificados para criptografia de dados em trânsito com a criptografia do Amazon EMR.

Ao criar uma configuração de segurança, você especifica dois conjuntos de opções de criptografia: a criptografia de dados em repouso e a criptografia de dados em trânsito. As opções de criptografia de dados em repouso incluem o Amazon S3 com EMRFS e a criptografia do disco local. As opções de criptografia em trânsito habilitam os recursos de criptografia de código-fonte aberto para determinados aplicativos que oferecem suporte para Transport Layer Security (TLS). Opções em repouso e opções em trânsito podem ser habilitadas juntas ou separadamente. Para ter mais informações, consulte Criptografar dados em repouso e em trânsito.

> **Note**
>
> Quando você usa AWS KMS, cobranças são cobradas pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Preços do AWS KMS.

Especificar opções de criptografia usando o console

Escolha as opções em Encryption (Criptografia) de acordo com as diretrizes a seguir.

- Escolha opções em At rest encryption (Criptografia em repouso) para criptografar os dados armazenados no sistema de arquivos.

  Você pode optar por criptografar dados no Amazon S3, em discos locais ou em ambos.
- Em Criptografia de dados do S3, para Modo de criptografia, escolha um valor para determinar como o Amazon EMR criptografa dados do Amazon S3 com o EMRFS.

  O que fazer em seguida depende do modo de criptografia escolhido:

  - SSE-S3

    Especifica a criptografia do lado do servidor com chaves de criptografia gerenciadas pelo Amazon S3. Você não precisa fazer mais nada, pois o Amazon S3 manipula as chaves para você.
  - SSE-KMS ou CSE-KMS

    Especifica a criptografia do lado do servidor com chaves AWS KMS gerenciadas (SSE-KMS) ou a criptografia do lado do cliente com chaves gerenciadas (CSE-KMS). AWS KMS Em AWS KMS key, selecione uma chave. A chave deve existir na mesma região que o seu cluster do EMR. Para conhecer os requisitos de chaves, consulte Usando AWS KMS keys para criptografia.
  - CSE-Custom

Especifica a criptografia do lado do cliente usando uma chave raiz personalizada do lado do cliente (CSE-Custom). Em Objeto do S3, insira o local no Amazon S3, ou o ARN do Amazon S3, do arquivo JAR de provedor de chaves personalizado. Em seguida, em Key provider class, insira o nome completo da classe declarada em seu aplicativo que implementa a EncryptionMaterialsProvider interface.

- Em Local disk encryption (Criptografia de disco local), escolha um valor para Key provider type (Tipo de provedor de chave).
  - AWS KMS key

    Selecione essa opção para especificar uma AWS KMS key. Em AWS KMS key, selecione uma chave. A chave deve existir na mesma região que o seu cluster do EMR. Para obter mais informações sobre requisitos de chaves, consulte Usando AWS KMS keys para criptografia.

    Criptografia do EBS

    Ao especificar AWS KMS como seu provedor de chaves, você pode ativar a criptografia do EBS para criptografar o dispositivo raiz e os volumes de armazenamento do EBS. Para habilitar essa opção, você deve conceder ao perfil de serviço do Amazon EMR `EMR_DefaultRole` permissões para usar a AWS KMS key especificada. Para obter mais informações sobre requisitos de chaves, consulte Habilitar a criptografia do EBS fornecendo permissões adicionais para chaves do KMS.
  - Custom (Personalizado)

    Selecione essa opção para especificar um provedor de chaves personalizado. Em Objeto do S3, insira o local no Amazon S3, ou o ARN do Amazon S3, do arquivo JAR de provedor de chaves personalizado. Em Key provider class, insira o nome completo da classe declarada em seu aplicativo que implementa a EncryptionMaterialsProvider interface. O nome de classe fornecido aqui deve ser diferente do nome de classe fornecido ao CSE-Custom.
- Escolha In-transit encryption (Criptografia em trânsito) para habilitar os recursos de criptografia TLS de código-fonte aberto para dados em trânsito. Escolha um tipo de provedor certificado em Certificate provider type (Tipo de provedor de certificados), de acordo com as seguintes diretrizes:
  - PEM

    Selecione essa opção para usar arquivos PEM que você fornece dentro de um arquivo zip. Dois artefatos são necessários dentro do arquivo zip: privateKey.pem e certificateChain.pem. Um terceiro arquivo, trustedCertificates.pem, é opcional. Para mais detalhes, consulte Fornecer

certificados para criptografia de dados em trânsito com a criptografia do Amazon EMR. Em Objeto do S3, especifique o local no Amazon S3, ou o ARN do Amazon S3, do campo do arquivo zip.

- Custom (Personalizado)

  Selecione essa opção para especificar um provedor de certificados personalizado. Em Objeto do S3, insira o local do Amazon S3, ou o ARN do Amazon S3, do seu arquivo JAR de provedor de certificados personalizado. Em Key provider class, insira o nome completo da classe declarada em seu aplicativo que implementa a interface TLSArtifactsProvider .

Especificando opções de criptografia usando o AWS CLI

As seções a seguir usam os exemplos de cenários para ilustrar um JSON --security-configuration bem-formado para configurações e provedores de chaves diferentes, seguido de uma referência dos parâmetros JSON e valores apropriados.

Exemplo de opções de criptografia de dados em trânsito

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada, e a criptografia de dados em repouso está desabilitada.
- Um arquivo zip com certificados no Amazon S3 é usado como o provedor de chaves (consulte Fornecer certificados para criptografia de dados em trânsito com a criptografia do Amazon EMR para conhecer os requisitos de certificados).

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": false,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "s3://MyConfigStore/artifacts/MyCerts.zip"
   }
  }
 }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada, e a criptografia de dados em repouso está desabilitada.
- Um provedor de chaves personalizado é usado (consulte Fornecer certificados para criptografia de dados em trânsito com a criptografia do Amazon EMR para conhecer os requisitos de certificados).

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": false,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "CertificateProviderClass": "com.mycompany.MyCertProvider"
   }
  }
  }
}'
```

Exemplo de opções de criptografia de dados em repouso

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A SSE-S3 é usada para criptografia do Amazon S3.
- A criptografia de disco local é usada AWS KMS como provedor de chaves.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": false,
  "EnableAtRestEncryption": true,
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
```

```
    "EncryptionMode": "SSE-S3"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "AwsKms",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   }
  }
  }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada e referencia um arquivo zip com certificados PEM no Amazon S3, usando o ARN.
- A SSE-KMS é usada para criptografia do Amazon S3.
- A criptografia de disco local é usada AWS KMS como provedor de chaves.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": true,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "arn:aws:s3:::MyConfigStore/artifacts/MyCerts.zip"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "SSE-KMS",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "AwsKms",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   }
  }
```

```
 }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada e referencia um arquivo zip com certificados PEM no Amazon S3.
- A CSE-KMS é usada para criptografia do Amazon S3.
- A criptografia do disco local usa um provedor de chaves personalizado referenciado por seu ARN.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": true,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "s3://MyConfigStore/artifacts/MyCerts.zip"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "CSE-KMS",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "Custom",
    "S3Object": "arn:aws:s3:::artifacts/MyKeyProvider.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
   }
  }
 }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está habilitada com um provedor de chaves personalizado.
- A CSE-Custom é usada para dados do Amazon S3.

- A criptografia do disco local usa um provedor de chaves personalizado.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": "true",
  "EnableAtRestEncryption": "true",
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "CertificateProviderClass": "com.mycompany.MyCertProvider"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "CSE-Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
    },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
   }
  }
 }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A criptografia do Amazon S3 é habilitada com SSE-KMS.
- Várias AWS KMS chaves são usadas, uma por cada bucket do S3, e exceções de criptografia são aplicadas a esses buckets individuais do S3.
- A criptografia do disco local está desabilitada.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
```

```
   "EncryptionConfiguration": {
     "AtRestEncryptionConfiguration": {
             "S3EncryptionConfiguration": {
           "EncryptionMode": "SSE-KMS",
           "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012",
           "Overrides": [
              {
                "BucketName": "sse-s3-bucket-name",
                "EncryptionMode": "SSE-S3"
              },
              {
                "BucketName": "cse-kms-bucket-name",
                "EncryptionMode": "CSE-KMS",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
              },
              {
                "BucketName": "sse-kms-bucket-name",
                "EncryptionMode": "SSE-KMS",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
              }
             ]
             }
           },
     "EnableInTransitEncryption": false,
      "EnableAtRestEncryption": true
  }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A criptografia do Amazon S3 está habilitada com SSE-S3, e a criptografia do disco local está desabilitada.

```
aws emr create-security-configuration --name "MyS3EncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
```

```
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "S3EncryptionConfiguration": {
                "EncryptionMode": "SSE-S3"
            }
        }
    }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A criptografia de disco local é ativada AWS KMS como provedor de chaves e a criptografia do Amazon S3 está desativada.

```
aws emr create-security-configuration --name "MyLocalDiskEncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "LocalDiskEncryptionConfiguration": {
                "EncryptionKeyProviderType": "AwsKms",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
            }
        }
    }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

- A criptografia de dados em trânsito está desabilitada, e a criptografia de dados em repouso está habilitada.
- A criptografia de disco local é ativada AWS KMS como provedor de chaves e a criptografia do Amazon S3 está desativada.
- A criptografia do EBS está habilitada.

```
aws emr create-security-configuration --name "MyLocalDiskEncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "LocalDiskEncryptionConfiguration": {
                "EnableEbsEncryption": true,
                "EncryptionKeyProviderType": "AwsKms",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
            }
        }
     }
}'
```

O exemplo abaixo ilustra o seguinte cenário:

O SSE-EMR-WAL é usado para criptografia EMR WAL

```
aws emr create-security-configuration --name "MySecConfig" \
    --security-configuration '{
        "EncryptionConfiguration": {
            "EMRWALEncryptionConfiguration":{ },
            "EnableInTransitEncryption":false, "EnableAtRestEncryption":false
        }
    }'
```

`EnableInTransitEncryption`e `EnableAtRestEncryption` ainda pode ser verdade, se quiser habilitar a criptografia relacionada.

O exemplo abaixo ilustra o seguinte cenário:

- O SSE-KMS-WAL é usado para criptografia EMR WAL
- A criptografia do lado do servidor é usada AWS Key Management Service como provedor principal

```
aws emr create-security-configuration --name "MySecConfig" \
    --security-configuration '{
        "EncryptionConfiguration": {
            "EMRWALEncryptionConfiguration":{
```

```
                "AwsKmsKey":"arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
                },
            "EnableInTransitEncryption":false, "EnableAtRestEncryption":false
        }
    }'
```

`EnableInTransitEncryption`e `EnableAtRestEncryption` ainda pode ser verdade, se quiser habilitar a criptografia relacionada.

Referência JSON para configurações de criptografia

A tabela a seguir lista os parâmetros JSON para configurações de criptografia e fornece uma descrição dos valores aceitáveis para cada parâmetro.

Parâmetro	Descrição
`"EnableInTransitEncryption" : true | false`	Especifique `true` para habilitar a criptografia em trânsito e `false` para desabilitá-la. Se omitido, `false` é assumido, e a criptografia em trânsito é desabilitada.
`"EnableAtRestEncryption": true | false`	Especifique `true` para habilitar a criptografia em repouso e `false` para desabilitá-la. Se omitido, `false` é assumido, e a criptografia em repouso é desabilitada.
Parâmetros de criptografia em trânsito	
`"InTransitEncryptionConfiguration" :`	Especifica uma coleção de valores usados para configurar a criptografia em trânsito quando `EnableInTransitEncryption` é `true`.
`"CertificateProviderType": "PEM" | "Custom"`	Especifica se é necessário usar `PEM`certificados referenciados com um arquivo compactado em zip ou um provedor de certificados `Custom`. Se `PEM` for especificado, `S3Object` deve ser uma referência à localização no Amazon S3 de um arquivo zip contendo os certificados. Se Personalizado for especificado, `S3Object`

Parâmetro	Descrição
	deverá ser uma referência à localização no Amazon S3 de um arquivo JAR, seguida por uma `CertificateProviderClass` entrada.
`"S3Object" : " ZipLocation "| "JarLocation "`	Fornece a localização no Amazon S3 para um arquivo zip quando `PEM` especificado ou para um arquivo JAR quando `Custom` especificado. O formato pode ser um caminho (por exemplo, `s3://MyConfig/artifacts/CertFiles.zip`) ou um ARN (por exemplo, `arn:aws:s3:::Code/MyCertProvider.jar)`). Se for especificado um arquivo zip, ele deverá conter arquivos exatamente denominados `privateKey.pem` e `certificateChain.pem` . Um arquivo denominado `trustedCertificates.pem` é opcional.
`"CertificateProviderClass" : "MyClassID "`	Obrigatório somente se `Custom` for especificado para`CertificateProviderType` . *`MyClassID`* especifica um nome de classe completo declarado no arquivo JAR, que implementa a interface ArtifactsProvider TLS. Por exemplo, `com.mycompany.MyCertProvider` .
Parâmetros de criptografia em repouso	
`"AtRestEncryptionConfiguration" :`	Especifica uma coleção de valores para criptografia em repouso quando `EnableAtRestEncryption` estiver`true`, incluindo criptografia Amazon S3 e criptografia de disco local.
Parâmetros de criptografia do Amazon S3	

Parâmetro	Descrição
`"S3EncryptionConfiguration" :`	Especifica uma coleção de valores usados para a criptografia do Amazon S3 com o Amazon EMR File System (EMRFS).
`"EncryptionMode" : "SSE-S3" | "SSE-KMS" | "CSE-KMS" | "CSE-Custom"`	Especifica o tipo de criptografia do Amazon S3 a ser usada. Se `SSE-S3` for especificado, nenhum valor adicional de criptografia do Amazon S3 será necessário. Se um `SSE-KMS` ou `CSE-KMS` for especificado, um AWS KMS key ARN deverá ser especificado como o `AwsKmsKey` valor. Se `CSE-Custom` for especificado, os valores `S3Object` e `EncryptionKeyProviderClass` deverão ser especificados.
`"AwsKmsKey" : " MyKeyARN"`	Necessário apenas quando `SSE-KMS` ou `CSE-KMS` é especificado para `EncryptionMode`. *`MyKeyARN`* deve ser um ARN totalmente especificado para uma chave (por exemplo, `arn:aws:kms:us-east-1:123456789012:key/12345678-1234-1234-1234-123456789012`).
`"S3Object" : "JarLocation "`	Obrigatório somente quando `CSE-Custom` é especificado para`CertificateProviderType`. *`JarLocation`* fornece a localização no Amazon S3 para um arquivo JAR. O formato pode ser um caminho (por exemplo, `s3://MyConfig/artifacts/MyKeyProvider.jar`) ou um ARN (por exemplo, `arn:aws:s3:::Code/MyKeyProvider.jar)`).

Parâmetro	Descrição
`"EncryptionKeyProviderClass" : "`*`MyS3KeyClassID`*` "`	Obrigatório somente quando `CSE-Custom` é especificado para `EncryptionMode`. *`MyS3KeyClassID`* especifica o nome completo de uma classe declarada no aplicativo que implementa a EncryptionMaterialsProvider interface; por exemplo,. *`com.mycompany.MyS3KeyProvider`*
Parâmetros de criptografia do disco local	
`"LocalDiskEncryptionConfiguration"`	Especifica o provedor de chaves e os valores correspondentes a serem usados para criptografia do disco local.
`"EnableEbsEncryption": true | false`	Especifique `true` para ativar a criptografia do EBS. A criptografia do EBS criptografa o volume do dispositivo raiz do EBS e os volumes de armazenamento conectados. Para usar a criptografia EBS, você deve especificar `AwsKms` como seu `EncryptionKeyProviderType`.
`"EncryptionKeyProviderType": "AwsKms" | "Custom"`	Especifica o provedor de chaves. Se `AwsKms` for especificado, um ARN da chave KMS deverá ser especificado como `AwsKmsKey` o valor. Se **`Custom`** for especificado, os valores `S3Object` e `EncryptionKeyProviderClass` deverão ser especificados.
`"AwsKmsKey : " `*`MyKeyARN`*`"`	Obrigatório somente quando `AwsKms` é especificado para `Type`. *`MyKeyARN`* deve ser um ARN totalmente especificado para uma chave (por exemplo, `arn:aws:kms:us-east-1:123456789012:key/12345678-1234-1234-1234-456789012123`).

Parâmetro	Descrição
`"S3Object" : "JarLocation "`	Obrigatório somente quando `CSE-Custom` é especificado para`CertificateProvideerType` . *`JarLocation`* fornece a localização no Amazon S3 para um arquivo JAR. O formato pode ser um caminho (por exemplo, `s3://MyConfig/artifacts/MyKeyProvider.jar`) ou um ARN (por exemplo, `arn:aws:s3:::Code/MyKeyProvider.jar)`).
`"EncryptionKeyProviderClass" : "`*`MyLocalDiskKeyClassID`*` "`	Obrigatório somente quando `Custom` é especificado para`Type`. *`MyLocalDiskKeyClassID`* especifica o nome completo de uma classe declarada no aplicativo que implementa a EncryptionMaterialsProvider interface; por exemplo,. *`com.mycompany.MyLocalDiskKeyProvider`*
Parâmetros de criptografia EMR WAL	
`"EMRWALEncryptionConfiguration"`	Especifica o valor da criptografia EMR WAL.
`"AwsKmsKey"`	Especifica a ID da chave CMK Arn.

Configurar a autenticação Kerberos

Uma configuração de segurança com definições Kerberos só pode ser usada por um cluster criado com atributos Kerberos, ou ocorrerá um erro. Para ter mais informações, consulte Usar o Kerberos para autenticação com o Amazon EMR. O Kerberos somente está disponível no Amazon EMR 5.10.0 e versões posteriores.

Especificar configurações do Kerberos usando o console

Escolha opções em Kerberos authentication (Autenticação Kerberos) de acordo com as diretrizes a seguir.

Parâmetro		Descrição
Kerberos		Especifica que o Kerberos está habilitado em clusters que usam essa configuração de segurança. Ao usar essa configuração de segurança, o cluster também deverá ter configurações Kerberos especificadas ou ocorrerá um erro.
Provedor	KDC dedicado ao cluster	Especifica que o Amazon EMR criará um KDC no nó primário de qualquer cluster que usar essa configura ção de segurança. Você especifica o nome do realm e a senha de administrador do KDC ao criar o cluster. Você pode referenciar esse KDC por outros clusters, se necessário. Crie esses clusters usando outra configuração de segurança, especifique um KDC externo e use o nome do território e a senha de administrador do KDC que você especificar para o KDC dedicado ao cluster.
	KDC externo	Disponível apenas no Amazon EMR 5.20.0 e versões posteriores. Especifica que os clusters que usam essa configuração de segurança autenticarão as entidades principais do Kerberos usando um servidor do KDC fora do cluster. O KDC não é criado no cluster. Ao criar o cluster, especifique o nome do realm e a senha de administrador do KDC para o KDC externo.
Vida útil do tíquete		Opcional. Especifica o período de validade de um tíquete do Kerberos emitido pelo KDC em clusters que usam essa configuração de segurança. Os ciclos de vida do tíquete são limitados por motivos de segurança. As aplicações e os serviços de cluster renovarão automaticamente os tíquetes quando perderem a validade. Os usuários que se conectam ao cluster via SSH usando credenciais do Kerberos

Parâmetro		Descrição
		precisam executar `kinit` pela linha de comando do nó primário para renovar um tíquete expirado.
Relação de confiança entre realms		Especifica uma relação de confiança entre regiões entre um KDC dedicado ao cluster em clusters que usam essa configuração de segurança e um KDC em outro realm do Kerberos. As entidades principais (normalmente usuários) de outro realm são autenticados em clusters que usam essa configuração. É necessário ter configuração adicional no outro realm do Kerberos. Para ter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.
Propriedades de confiança entre realms	Realm	Especifica o nome de realm Kerberos de outro realm na relação de confiança. Por convenção, os nomes de realm do Kerberos são iguais ao nome do domínio, mas em letras maiúsculas.
	Domínio	Especifica o nome de domínio de outro realm na relação de confiança.
	Servidor do administrador	Especifica o nome de domínio totalmente qualificado (FQDN) ou endereço IP do servidor de administrador no outro realm da relação de confiança. O servidor de administração e o servidor de KDC normalmente são executados na mesma máquina com o mesmo FQDN, mas se comunicam por diferentes portas. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com :749`).

Parâmetro		Descrição
	Servidor do KDC	Especifica o nome de domínio totalmente qualificado (FQDN) ou endereço IP do servidor do KDC no outro realm da relação de confiança. O servidor de KDC e o servidor de administração normalmente são executados na mesma máquina com o mesmo FQDN, mas usam diferentes portas. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com:88`).
KDC externo		Especifica que o KDC externo do cluster será usado pelo cluster.
Propriedades do KDC externo	Servidor do administrador	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor do administrador externo. O servidor de administração e o servidor de KDC normalmente são executados na mesma máquina com o mesmo FQDN, mas se comunicam por diferentes portas. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com:749`).

Parâmetro			Descrição
	Servidor do KDC		Especifica o nome de domínio totalmente qualificado (FQDN) do servidor do KDC externo. O servidor de KDC e o servidor de administração normalmente são executados na mesma máquina com o mesmo FQDN, mas usam diferentes portas. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalm ente, você pode especificar a porta (por exemplo, `domain.example.com :88`).
	Integração do Active Directory		Especifica que a autenticação da entidade principal do Kerberos está integrada a um domínio do Microsoft Active Directory.
	Proprieda des de integraçã o do Active Directory	Realm do Active Directory	Especifica o nome do realm do Kerberos do domínio do Active Directory. Por convenção, os nomes de realm do Kerberos geralmente são iguais ao nome do domínio, mas em letras maiúsculas.
		Domínio do Active Directory	Especifica o nome de domínio do Active Directory.
		Servidor do Active Directory	Especifica o nome de domínio totalmente qualificado (FQDN) do controlador de domínio do Microsoft Active Directory.

Especificando as configurações do Kerberos usando o AWS CLI

A tabela de referência a seguir mostra os parâmetros JSON para configurações do Kerberos em uma configuração de segurança. Para exemplos de configuração, consulte Exemplos de configuração.

Parâmetro	Descrição
"AuthenticationConfiguration": {	Obrigatório para o Kerberos. Especifica que uma configuração de autenticação faz parte dessa configuração de segurança.
"KerberosConfiguration": {	Obrigatório para o Kerberos. Especifica as propriedades de configuração do Kerberos.
"Provider": *"ClusterDedicatedKdc",* —ou— "Provider: *"ExternalKdc",*	*ClusterDedicatedKdc* especifica que o Amazon EMR criará um KDC no nó primário de qualquer cluster que usar essa configuração de segurança. Você especifica o nome do realm e a senha de administrador do KDC ao criar o cluster. Você pode referenciar esse KDC por outros clusters, se necessário. Crie esses clusters usando outra configuração de segurança, especifique um KDC externo e use o nome do território e a senha de administrador do KDC que você especificou ao criar o cluster com KDC dedicado ao cluster. *ExternalKdc* especifica que o cluster usa um KDC externo. O Amazon EMR não cria um KDC no nó primário. O cluster que usa essa configuração de segurança deve especificar o nome do realm e a senha de administrador do KDC externo.

Parâmetro	Descrição
`"ClusterDedicatedKdcConfiguration": {`	Obrigatório quando *`ClusterDedicatedKdc`* for especificado.
`"TicketLifetimeInHours": 24,`	Opcional. Especifica o período de validade de um tíquete do Kerberos emitido pelo KDC em clusters que usam essa configuração de segurança. Os ciclos de vida do tíquete são limitados por motivos de segurança. As aplicações e os serviços de cluster renovarão automaticamente os tíquetes quando perderem a validade. Os usuários que se conectam ao cluster via SSH usando credenciais do Kerberos precisam executar `kinit` pela linha de comando do nó primário para renovar um tíquete expirado.

Parâmetro	Descrição
`"CrossRealmTrustConfiguration": {`	Especifica uma relação de confiança entre regiões entre um KDC dedicado ao cluster em clusters que usam essa configuração de segurança e um KDC em outro realm do Kerberos. As entidades principais (normalmente usuários) de outro realm são autenticados em clusters que usam essa configuração. É necessário ter configuração adicional no outro realm do Kerberos. Para ter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.
`"Realm": "KDC2.COM",`	Especifica o nome de realm Kerberos de outro realm na relação de confiança. Por convenção, os nomes de realm do Kerberos são iguais ao nome do domínio, mas em letras maiúsculas.
`"Domain": "kdc2.com",`	Especifica o nome de domínio de outro realm na relação de confiança.

Parâmetro	Descrição
"AdminServer": "*kdc.com:749* ",	Especifica o nome de domínio totalmente qualificado (FQDN) ou endereço IP do servidor de administrador no outro realm da relação de confiança. O servidor de administração e o servidor de KDC normalmente são executados na mesma máquina com o mesmo FQDN, mas se comunicam por diferentes portas. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com` *:749*).
"KdcServer": "*kdc.com:88* "	Especifica o nome de domínio totalmente qualificado (FQDN) ou endereço IP do servidor do KDC no outro realm da relação de confiança . O servidor de KDC e o servidor de administração normalmente são executados na mesma máquina com o mesmo FQDN, mas usam diferentes portas. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com` *:88*).

Parâmetro	Descrição
}	
}	
“ExternalKdcConfiguração”: {	Obrigatório quando *ExternalKdc* for especificado.
"TicketLifetimeInHours": *24*,	Opcional. Especifica o período de validade de um tíquete do Kerberos emitido pelo KDC em clusters que usam essa configuração de segurança. Os ciclos de vida do tíquete são limitados por motivos de segurança . As aplicações e os serviços de cluster renovarão automaticamente os tíquetes quando perderem a validade. Os usuários que se conectam ao cluster via SSH usando credenciais do Kerberos precisam executar `kinit` pela linha de comando do nó primário para renovar um tíquete expirado.
"KdcServerType": "Single",	Especifica que um único servidor do KDC é referenciado. `Single` é o único valor com suporte atualmente.

Parâmetro				Descrição
			“AdminServer“:" *kdc.com:749* “,	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor do administrador externo. O servidor de administração e o servidor de KDC normalmente são executados na mesma máquina com o mesmo FQDN, mas se comunicam por diferentes portas. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com` *:749*).
			“KdcServer“:" *kdc.com:88* “,	Especifica o nome de domínio totalmente qualificado (FQDN) do servidor do KDC externo. O servidor de KDC e o servidor de administr ação normalmente são executados na mesma máquina com o mesmo FQDN, mas usam diferentes portas. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com` *:88*).

Parâmetro	Descrição
"AdIntegrationConfiguration": {	Especifica que a autenticação da entidade principal do Kerberos está integrada a um domínio do Microsoft Active Directory.
"AdRealm": "*AD.DOMAIN.COM* ",	Especifica o nome do realm do Kerberos do domínio do Active Directory. Por convenção, os nomes de realm do Kerberos geralmente são iguais ao nome do domínio, mas em letras maiúsculas.
"AdDomain": "*ad.domain.com* "	Especifica o nome de domínio do Active Directory.
"AdServer": "*ad.domain.com* "	Especifica o nome de domínio totalmente qualificado (FQDN) do controlador de domínio do Microsoft Active Directory.
}	
}	
}	
}	

Configurar perfis do IAM para solicitações do EMRFS para o Amazon S3

Os perfis do IAM para EMRFS permitem que você forneça diferentes permissões para os dados do EMRFS no Amazon S3. Você cria mapeamentos que especificam um perfil do IAM que é usado para permissões quando uma solicitação de acesso contém um identificador especificado. O identificador pode ser um usuário ou um perfil do Hadoop ou um prefixo do Amazon S3.

Para ter mais informações, consulte Configurar perfis do IAM para solicitações do EMRFS para o Amazon S3.

Especificando funções do IAM para o EMRFS usando o AWS CLI

Veja a seguir um exemplo de trecho JSON para especificar perfis do IAM personalizados para o EMRFS em uma configuração de segurança. Ele demonstra mapeamentos de perfil para os três tipos diferentes de identificadores, seguidos por uma referência de parâmetro.

```
{
  "AuthorizationConfiguration": {
    "EmrFsConfiguration": {
      "RoleMappings": [{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_user1",
        "IdentifierType": "User",
        "Identifiers": [ "user1" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_to_MyBuckets",
        "IdentifierType": "Prefix",
        "Identifiers": [ "s3://MyBucket/","s3://MyOtherBucket/" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_AdminGroup",
        "IdentifierType": "Group",
        "Identifiers": [ "AdminGroup" ]
      }]
    }
  }
}
```

Parâmetro	Descrição
`"AuthorizationConfiguration":`	Obrigatório.
`"EmrFsConfiguration":`	Obrigatório. Contém mapeamentos de perfil.
`"RoleMappings":`	Obrigatório. Contém uma ou mais definições de mapeamento de perfil. Os mapeamentos de perfil são avaliados na ordem em que aparecem, de cima para baixo. Se o mapeamento de perfil for avaliado como true para uma chamada do EMRFS para dados

Parâmetro	Descrição
	no Amazon S3, nenhum outro mapeamento de perfil será avaliado, e o EMRFS usará o perfil do IAM especificado para a solicitação. O mapeamento de perfil tem os seguintes parâmetros obrigatórios:
`"Role":`	Especifica o identificador ARN de um perfil do IAM no formato `arn:aws:iam::` *`account-id`* `:role/`*`role-name`* . Essa é o perfil do IAM que o Amazon EMR assume se a solicitação do EMRFS para o Amazon S3 corresponder a qualquer um dos `Identifiers` especificados.
`"IdentifierType":`	Pode ser um dos seguintes: • `"User"` especifica que os identificadores são um ou mais usuários do Hadoop, que podem ser usuários de contas Linux ou entidades principais do Kerberos. Quando a solicitação do EMRFS se origina com os usuários especificados, o perfil do IAM é assumido. • `"Prefix"` especifica que o identificador é um local do Amazon S3. O perfil do IAM é assumido para chamadas para os locais com os prefixos especificados. Por exemplo, o prefixo `s3://mybucket/` corresponde a `s3://mybucket/mydir` e `s3://mybucket/yetanotherdir` . • `"Group"` especifica que os identificadores são um ou mais grupos do Hadoop. O perfil do IAM será assumido se a solicitação for originada de um usuário dos grupos especificados.

Parâmetro	Descrição
`"Identifiers":`	Especifica um ou mais identificadores do tipo de identificador adequado. Separe múltiplos identificadores por vírgulas sem espaços.

Configurar solicitações de serviço de metadados para instâncias do Amazon EC2

Os metadados da instância são dados sobre sua instância que é possível usar para configurar ou gerenciar a instância em execução. É possível acessar metadados de instância em uma instância em execução usando um dos seguintes métodos:

- Serviço de metadados da instância versão 1 (IMDSv1): um método de solicitação/resposta
- Serviço de metadados da instância versão 2 (IMDSv2): um método orientado a sessões

Enquanto o Amazon EC2 oferece suporte ao IMDSv1 e ao IMDSv2, o Amazon EMR oferece suporte ao IMDSv2 no Amazon EMR 5.23.1, 5.27.1, 5.32 ou versões posteriores e 6.2 ou versões posteriores. Nessas versões, os componentes do Amazon EMR usam o IMDSv2 em todas as chamadas do IMDS. Para chamadas do IMDS no código da aplicação, você pode usar IMDSv1 e IMDSv2 ou configurar o IMDS para usar somente IMDSv2 para segurança adicional. Quando você especifica que o IMDSv2 deve ser usado, o IMDSv1 não funciona mais.

Para obter mais informações, consulte Configurar o serviço de metadados da instância no Guia do usuário do Amazon EC2.

> **Note**
>
> Nas versões anteriores do Amazon EMR 5.x ou 6.x, desativar o IMDSv1 causará falha na inicialização do cluster, pois os componentes do Amazon EMR usam o IMDSv1 em todas as chamadas do IMDS. Ao desativar o IMDSv1, verifique se todos os softwares personalizados que utilizam o IMDSv1 estão atualizados para o IMDSv2.

Especificar a configuração do serviço de metadados da instância usando a AWS CLI

Veja a seguir um exemplo de trecho do JSON para especificar o serviço de metadados de instância (IMDS) do Amazon EC2 em uma configuração de segurança. Usar uma configuração de segurança personalizada é opcional.

```
{
  "InstanceMetadataServiceConfiguration" : {
      "MinimumInstanceMetadataServiceVersion": integer,
      "HttpPutResponseHopLimit": integer
   }
}
```

Parâmetro	Descrição
`"InstanceMetadataServiceConfiguration":`	Se você não especificar o IMDS em uma configuração de segurança e usar uma versão do Amazon EMR que exija o IMDSv1, o Amazon EMR usará como padrão o IMDSv1 como a versão mínima do serviço de metadados da instância. Se você quiser usar sua própria configuração, os dois parâmetros a seguir são necessários.
`"MinimumInstanceMetadataServiceVersion":`	Obrigatório. Especifique 1 ou 2. O valor 1 permite o IMDSv1 e o IMDSv2. O valor 2 permite somente IMDSv2.
`"HttpPutResponseHopLimit":`	Obrigatório. O limite de salto de resposta HTTP PUT desejado para solicitações de metadados de instância. Quanto maior o número, mais as solicitações de metadados de instância podem viajar. Padrão: 1. Especifique um número inteiro de 1 a 64.

Especificar a configuração do serviço de metadados da instância usando o console

É possível configurar o uso do IMDS para um cluster ao iniciá-lo no console do Amazon EMR.

Controles de configurações de segurança do IMDS no console do Amazon EMR

Para configurar o uso do IMDS usando o console:

1. Ao criar uma nova configuração de segurança na página Configurações de segurança, selecione Configurar serviço de metadados de instância do EC2 na configuração Serviço de metadados de instância do EC2. Essa configuração é compatível somente com o Amazon EMR 5.23.1, 5.27.1, 5.32 ou posteriores e 6.2 ou posteriores.
2. Na opção Versão mínima do serviço de metadados de instância, selecione:
   - Desativar o IMDSv1 e permitir somente o IMDSv2, se quiser permitir somente o IMDSv2 no cluster. Consulte Transição para o uso do serviço de metadados de instância versão 2 no Guia do usuário do Amazon EC2.
   - Permitir o IMDSv1 e o IMDSv2 no cluster, se quiser permitir o IMDSv1 e o IMDSv2 orientado por sessão no cluster.
3. Para IMDSv2, também é possível configurar o número permitido de saltos de rede para o token de metadados, definindo o limite de salto de resposta HTTP put como um número inteiro de 1 a 64.

Para obter mais informações, consulte Configurar o serviço de metadados da instância no Guia do usuário do Amazon EC2.

Consulte Configurar detalhes da instância e Configurar o serviço de metadados da instância no Guia do usuário do Amazon EC2.

Especificar uma configuração de segurança para um cluster

Ao criar um cluster, você pode especificar configurações de criptografia definindo a configuração de segurança. Você pode usar o AWS Management Console ou AWS CLI o.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Especificar uma configuração de segurança usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Configuração e permissões de segurança, localize o campo Configuração de segurança. Selecione o menu suspenso ou escolha Procurar para selecionar o nome de uma configuração de segurança criada anteriormente. Como alternativa, escolha Criar configuração de segurança para criar uma configuração que você possa usar em seu cluster.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Old console

Especificar uma configuração de segurança usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Na tela Etapa 1: Software e etapas, na lista Versão, escolha emr-4.8.0 ou uma versão mais recente. Escolha as configurações desejadas e selecione Next (Próximo).
4. Na tela Step 2: Hardware (Etapa 2: Hardware), escolha as configurações desejadas e selecione Next (Próximo). Faça o mesmo para a tela Step 3: General Cluster Settings (Etapa 3: Configurações gerais do cluster).
5. Na tela Step 4: Security (Etapa 4: Segurança), em Encryption Options (Opções de criptografia), escolha um valor para Security configuration (Configuração de segurança).
6. Configure outras opções de segurança conforme desejado e escolha Create cluster (Criar cluster).

CLI

Para especificar uma configuração de segurança com o AWS CLI

- Use `aws emr create-cluster` para opcionalmente aplicar uma configuração de segurança usando `--security-configuration` *`MySecConfig`*, em que *`MySecConfig`*

é o nome da configuração de segurança, como mostra o exemplo a seguir. O `--release-label` que você especificar deve ser 4.8.0 ou posterior e o `--instance-type` pode ser qualquer disponível.

```
aws emr create-cluster --instance-type m5.xlarge --release-label emr-5.0.0 --
security-configuration mySecConfig
```

Proteção de dados no Amazon EMR

O modelo de responsabilidade AWS compartilhada se aplica à proteção de dados no Amazon EMR. Conforme descrito neste modelo, AWS é responsável por proteger a infraestrutura global que executa toda a AWS nuvem. Você é responsável por manter o controle sobre seu conteúdo hospedado nessa infraestrutura. Esse conteúdo inclui as tarefas de configuração e gerenciamento de segurança do AWS que você usa. Para obter mais informações sobre a privacidade de dados, consulte as Perguntas frequentes sobre privacidade de dados. Para obter informações sobre proteção de dados na Europa, consulte o modelo de responsabilidade compartilhada da Amazon e a postagem do blog sobre o GDPR no Blog AWS de Segurança.

Para fins de proteção de dados, recomendamos que você proteja as credenciais da AWS conta e configure contas individuais com AWS Identity and Access Management. Dessa maneira, cada usuário receberá apenas as permissões necessárias para cumprir suas obrigações de trabalho. Recomendamos também que você proteja seus dados das seguintes formas:

- Use uma autenticação multifator (MFA) com cada conta.
- Use o TLS para se comunicar com AWS os recursos. O TLS 1.2 é obrigatório.
- Configure a API e o registro de atividades do usuário com AWS CloudTrail.
- Use soluções AWS de criptografia, juntamente com todos os controles de segurança padrão nos AWS serviços.
- Use serviços gerenciados de segurança avançada, como o Amazon Macie, que ajuda a localizar e proteger dados pessoais armazenados no Amazon S3.
- Se você precisar de módulos criptográficos validados pelo FIPS 140-2 ao acessar a AWS por meio de uma interface de linha de comando ou uma API, use um endpoint do FIPS. Para obter mais informações sobre endpoints do FIPS, consulte Federal Information Processing Standard (FIPS) 140-2.

É altamente recomendável que você nunca coloque informações de identificação confidenciais, como números de conta dos seus clientes, em campos de formato livre, como um campo Nome. Isso inclui quando você trabalha com o Amazon EMR ou outros AWS serviços usando o console, a API ou AWS os AWS CLI SDKs. Todos os dados que você insere no Amazon EMR ou em outros serviços podem ser separados para inclusão em logs de diagnóstico. Ao fornecer um URL para um servidor externo, não inclua informações de credenciais no URL para validar a solicitação a esse servidor.

Criptografar dados em repouso e em trânsito

A criptografia de dados ajuda a impedir que usuários não autorizados leiam dados em um cluster e em sistemas de armazenamento físico de dados associados. Isso inclui dados salvos em mídias persistentes, conhecidos como dados em repouso, e dados que podem ser interceptados enquanto viajam pela rede, conhecidos como dados em trânsito.

Começando com o Amazon EMR versão 4.8.0, você pode usar as configurações de segurança do Amazon EMR para definir configurações de criptografia de dados para clusters com mais facilidade. Configurações de segurança oferecem configurações para habilitar a segurança dos dados em trânsito e dos dados em repouso em volumes do Amazon Elastic Block Store (Amazon EBS) e do EMRFS no Amazon S3.

Opcionalmente, a partir do Amazon EMR versão 4.1.0 e posterior, você tem a opção de configurar a criptografia transparente no HDFS, que não é configurada usando configurações de segurança. Para obter mais informações, consulte Transparent encryption in HDFS on Amazon EMR no Guia de lançamento do Amazon EMR.

Tópicos

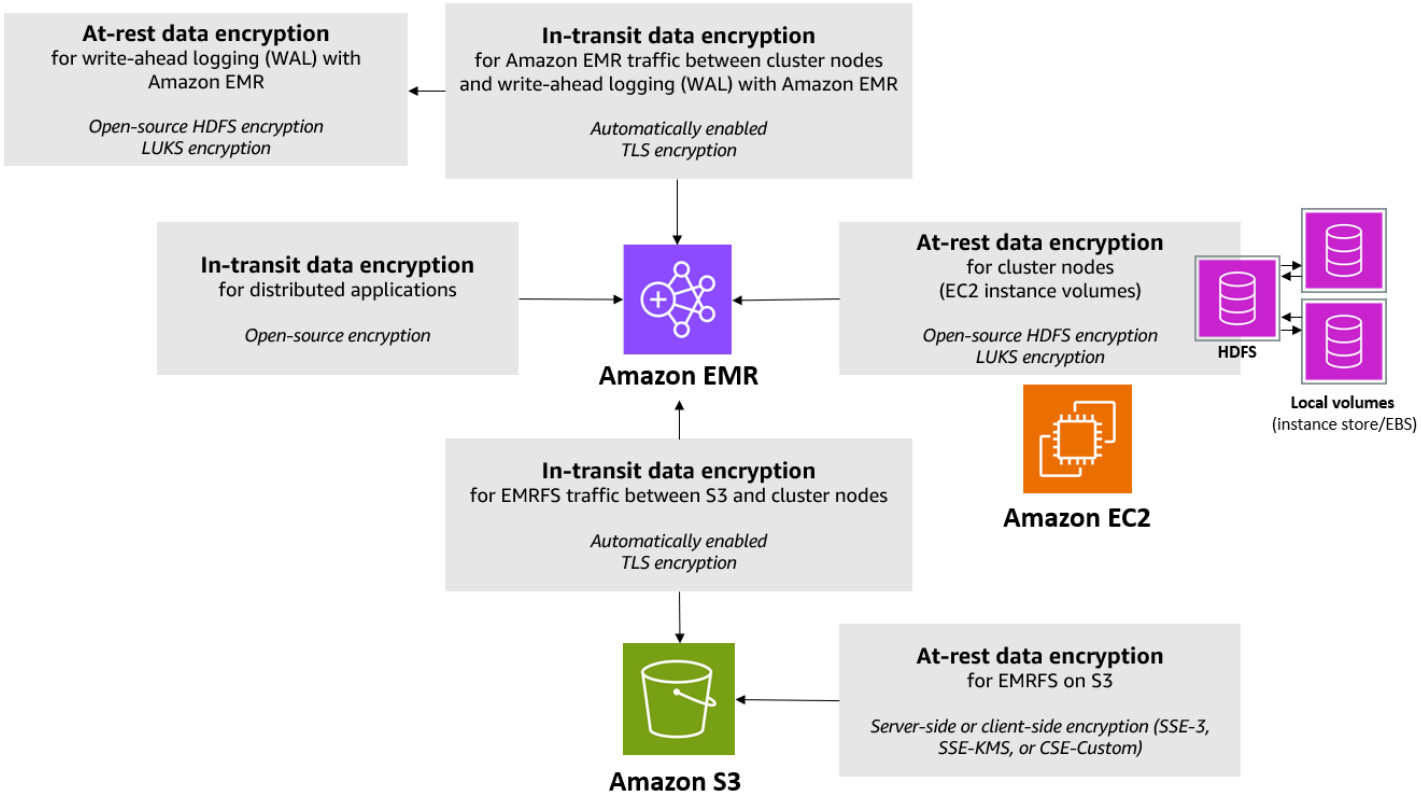
- Opções de criptografia
- Criar chaves e certificados para criptografia de dados

Opções de criptografia

Com as versões 4.8.0 e superiores do Amazon EMR, você pode usar uma configuração de segurança para especificar configurações para criptografar dados em repouso, dados em trânsito ou ambos. Ao habilitar a criptografia de dados em repouso, você tem a opção de criptografar dados do EMRFS no Amazon S3, dados em discos locais ou ambos. Cada configuração de segurança que você cria é armazenada no Amazon EMR, e não na configuração do cluster. Dessa forma, você pode facilmente reutilizar uma configuração para especificar as configurações de criptografia de

dados sempre que criar um cluster. Para ter mais informações, consulte Criar uma configuração de segurança.

O diagrama a seguir mostra as diferentes opções de criptografia de dados disponíveis com as configurações de segurança.



As seguintes opções de criptografia também estão disponíveis e não são configuradas usando uma configuração de segurança:

- Se preferir, com as versões 4.1.0 e posteriores do Amazon EMR, você pode optar por configurar uma criptografia transparente no HDFS. Para obter mais informações, consulte Transparent encryption in HDFS on Amazon EMR no Guia de lançamento do Amazon EMR.
- Se estiver usando uma versão do Amazon EMR que não seja compatível com as configurações de segurança, você poderá configurar a criptografia para dados do EMRFS no Amazon S3 manualmente. Para obter mais informações, consulte Specifying Amazon S3 encryption using EMRFS properties.
- Se você estiver usando uma versão do Amazon EMR anterior à 5.24.0, só haverá suporte para volumes de dispositivo raiz do EBS criptografados ao usar uma AMI personalizada. Para obter mais informações, consulte Creating a custom AMI with an encrypted Amazon EBS root device volume no Guia de gerenciamento do Amazon EMR.

> **Note**
>
> A partir da versão 5.24.0 do Amazon EMR, você pode usar uma opção de configuração de segurança para criptografar o dispositivo raiz e os volumes de armazenamento do EBS ao especificar como seu provedor de chaves. AWS KMS Para ter mais informações, consulte Criptografia de disco local.

A criptografia de dados requer chaves e certificados. Uma configuração de segurança oferece a flexibilidade de escolher entre várias opções, incluindo chaves gerenciadas por AWS Key Management Service, chaves gerenciadas pelo Amazon S3 e chaves e certificados de fornecedores personalizados fornecidos por você. Ao usar AWS KMS como seu provedor de chaves, cobranças se aplicam pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Preços do AWS KMS.

Antes de especificar opções de criptografia, decida quais sistemas de gerenciamento de chaves e certificados você deseja usar, para poder primeiro criar as chaves e os certificados ou os provedores personalizados especificados como parte das configurações de criptografia.

Criptografia em repouso para dados do EMRFS no Amazon S3

A criptografia do Amazon S3 funciona com os objetos do Amazon EMR File System (EMRFS) lidos e gravados no Amazon S3. Você especifica a criptografia do lado do servidor (SSE) ou a criptografia do lado do cliente (CSE) do Amazon S3 como o modo de criptografia padrão ao habilitar a criptografia em repouso. Opcionalmente, você pode especificar diferentes métodos de criptografia para buckets individuais usando Per bucket encryption overrides (Substituições de criptografia por bucket). Independentemente de a criptografia do Amazon S3 estar habilitada, o Transport Layer Security (TLS) criptografa os objetos do EMRFS em trânsito entre os nós do cluster do EMR e o Amazon S3. Para obter mais informações sobre a criptografia do Amazon S3, consulte Proteção de dados usando criptografia no Guia do usuário do Amazon Simple Storage Service.

> **Note**
>
> Quando você usa AWS KMS, cobranças são cobradas pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Preços do AWS KMS.

Criptografia do lado do servidor do Amazon S3

Quando você configura a criptografia do lado do servidor do Amazon S3, o Amazon S3 criptografa os dados no nível do objeto à medida que os grava no disco e os descriptografa quando são acessados. Para ter mais informações sobre o SSE, consulte Proteger os dados usando criptografia do lado do servidor no Guia do usuário do Amazon Simple Storage Service.

Você pode escolher entre dois sistemas de gerenciamento de chaves diferentes ao especificar a SSE no Amazon EMR:

- SSE-S3: o Amazon S3 gerencia as chaves para você.
- SSE-KMS — Você usa um AWS KMS key para configurar políticas adequadas para o Amazon EMR. Para obter mais informações sobre os principais requisitos do Amazon EMR, consulte Uso AWS KMS keys para criptografia.

A SSE com chaves fornecidas pelo cliente (SSE-C) não está disponível para o uso com o Amazon EMR.

Criptografia do lado do cliente do Amazon S3

Com a criptografia do lado do cliente do Amazon S3, a criptografia e a descriptografia do Amazon S3 ocorrem no cliente do EMRFS em seu cluster. Os objetos são criptografados antes de serem carregados no Amazon S3 e descriptografados após serem baixados. O provedor especificado por você fornece a chave de criptografia que o cliente usa. O cliente pode usar chaves fornecidas pelo AWS KMS (CSE-KMS) ou uma classe Java personalizada que fornece a chave raiz do lado do cliente (CSE-C). As especificações de criptografia são ligeiramente diferentes entre a CSE-KMS e a CSE-C, dependendo do provedor especificado e dos metadados do objeto que está sendo descriptografado ou criptografado. Para obter mais informações sobre essas diferenças, consulte Proteger dados usando a criptografia do lado do cliente no Guia do usuário do Amazon Simple Storage Service.

> **Note**
>
> A CSE do Amazon S3 garante somente que os dados do EMRFS trocados com o Amazon S3 sejam criptografados. Não são todos os dados nos volumes de instâncias do cluster que são criptografados. Além disso, como o Hue não usa o EMRFS, os objetos que o navegador de arquivos do S3 para Hue grava no Amazon S3 não são criptografados.

Criptografia em repouso para dados no Amazon EMR WAL

Quando você configura a criptografia do lado do servidor (SSE) para registro antecipado de gravação (WAL), o Amazon EMR criptografa dados em repouso. Você pode escolher entre dois sistemas diferentes de gerenciamento de chaves ao especificar SSE no Amazon EMR:

SSE-EMR-WAL

O Amazon EMR gerencia as chaves para você. Por padrão, o Amazon EMR criptografa os dados que você armazenou no Amazon EMR WAL com. SSE-EMR-WAL

SSE-KMS-WAL

Você usa uma AWS KMS chave para configurar políticas que se aplicam ao Amazon EMR WAL. Para obter mais informações sobre os principais requisitos do Amazon EMR, consulte. Usando AWS KMS keys para criptografia

Você não pode usar sua própria chave com o SSE ao habilitar o WAL com o Amazon EMR. Para obter mais informações, consulte Write-ahead logs (WAL) para o Amazon EMR.

Criptografia de disco local

Os mecanismos a seguir funcionam em conjunto para criptografar discos locais quando você habilita a criptografia de disco local usando uma configuração de segurança do Amazon EMR.

Criptografia HDFS de código aberto

O HDFS troca dados entre instâncias de cluster durante o processamento distribuído. Ele também lê e grava dados em volumes de armazenamento de instâncias e em volumes do EBS anexados às instâncias. As seguintes opções de criptografia Hadoop de código-fonte aberto são ativadas quando você habilita a criptografia do disco local:

- Secure Hadoop RPC: é definida como `Privacy`, que usa a Simple Authentication Security Layer (SASL).
- Data encryption on HDFS block data transfer: é definida como `true` e configurada para usar a criptografia AES de 256 bits.

> **Note**
>
> Você pode ativar a criptografia adicional do Apache Hadoop habilitando a criptografia em trânsito. Para ter mais informações, consulte Criptografia em trânsito. Essas configurações de criptografia não ativam a criptografia transparente do HDFS, que você pode configurar manualmente. Para obter mais informações, consulte Transparent encryption in HDFS on Amazon EMR no Guia de lançamento do Amazon EMR.

Criptografia de armazenamento de instância

Para tipos de instância do EC2 que usam SSDs baseados em NVMe como o volume de armazenamento de instância, a criptografia de NVMe é usada independentemente das configurações de criptografia do Amazon EMR. Para obter mais informações, consulte os volumes SSD NVMe no Guia do usuário do Amazon EC2. Para outros volumes de armazenamento de instância, o Amazon EMR usa LUKS para criptografar o volume de armazenamento de instância quando a criptografia de disco local está habilitada, não importa se os volumes do EBS foram criptografados usando criptografia do EBS ou LUKS.

Criptografia de volume do EBS

Se você criar um cluster em uma região onde a criptografia do Amazon EC2 de volumes do EBS está habilitada por padrão para sua conta, os volumes do EBS serão criptografados mesmo que a criptografia de disco local não esteja habilitada. Para obter mais informações, consulte Criptografia por padrão no Guia do usuário do Amazon EC2. Com a criptografia de disco local habilitada em uma configuração de segurança, as configurações do Amazon EMR têm precedência sobre as configurações do Amazon EC2 para instâncias EC2 encryption-by-default em cluster.

As seguintes opções estão disponíveis para criptografar volumes do EBS usando uma configuração de segurança:

- Criptografia do EBS: a partir do Amazon EMR versão 5.24.0, você tem a opção de habilitar a criptografia do EBS. A opção de criptografia do EBS criptografa o volume do dispositivo raiz do EBS e os volumes de armazenamento anexados. A opção de criptografia do EBS está disponível somente quando você especifica AWS Key Management Service como seu provedor de chaves. Recomendamos usar a criptografia do EBS.
- Criptografia LUKS: se você optar por usar a criptografia LUKS para volumes do Amazon EBS, a criptografia LUKS se aplicará apenas a volumes de armazenamento anexados, e não ao

volume do dispositivo raiz. Para obter mais informações sobre a criptografia LUKS, consulte a Especificação da no disco.

Para seu provedor de chaves, você pode configurar uma AWS KMS key com políticas adequadas para o Amazon EMR ou uma classe Java personalizada que forneça os artefatos de criptografia. Quando você usa AWS KMS, cobranças são cobradas pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Preços do AWS KMS.

> **Note**
>
> Para verificar se a criptografia do EBS está habilitada no cluster, é recomendável usar a chamada de API `DescribeVolumes`. Para obter mais informações, consulte DescribeVolumes. A execução de `lsblk` no cluster só verificará o status da criptografia LUKS, em vez da criptografia do EBS.

Criptografia em trânsito

Vários mecanismos de criptografia estão habilitados com a criptografia em trânsito. Esses são recursos de código aberto, específicos da aplicação e podem variar de acordo com a versão do Amazon EMR. Os atributos de criptografia a seguir específicos da aplicação podem ser habilitados usando configurações de aplicação do Apache. Para obter mais informações, consulte Configure applications.

Hadoop

- O shuffle MapReduce criptografado do Hadoop usa TLS.
- A opção Proteger RPC do Hadoop está definida como “Privacy” e usa SASL (ativada no Amazon EMR quando a criptografia em repouso está habilitada).
- A opção Criptografia de dados na transferência de dados em bloco do HDFS usa a AES 256 (ativada no Amazon EMR quando a criptografia em repouso está habilitada na configuração de segurança).
- Para obter mais informações, consulte Hadoop in secure mode, na documentação do Apache Hadoop.

HBase

- Quando o Kerberos está habilitado, a propriedade `hbase.rpc.protection` é definida como `privacy` para comunicação criptografada.
- Para obter mais informações, consulte Client-side configuration for secure operation na documentação do Apache HBase.
- Para obter mais informações sobre o Kerberos com Amazon EMR, consulte Usar o Kerberos para autenticação com o Amazon EMR.

Hive

- A comunicação do cliente JDBC/ODBC com HiveServer 2 (HS2) é criptografada usando configurações SSL nas versões 6.9.0 e posteriores do Amazon EMR.
- Para obter mais informações, consulte a seção SSL encryption da documentação do Apache Hive.

Spark

- A comunicação de RPC interna entre os componentes do Spark, como o serviço de transferência de blocos e o serviço de shuffle externo, é criptografada usando a criptografia AES-256 nas versões 5.9.0 e posteriores do Amazon EMR. Em versões anteriores, a comunicação RPC interna é criptografada usando SASL com DIGEST-MD5 como a criptografia.
- A comunicação do protocolo HTTP com interfaces de usuário, como o Spark History Server e servidores de arquivos habilitados para HTTPS é criptografada usando a configuração de SSL do Spark. Para obter mais informações, consulte SSL configuration na documentação do Spark.
- Para obter mais informações, consulte a seção Spark security settings da documentação do Apache Spark.

Tez

- A opção Tez Shuffle Handler usa TLS (`tez.runtime.ssl.enable`).

Presto

- A comunicação interna entre nós do Presto usa SSL/TLS (somente no Amazon EMR versão 5.6.0 e posteriores).

Você especifica os artefatos de criptografia usados para a criptografia em trânsito de uma destas duas maneiras: fornecendo um arquivo compactado de certificados, que é carregado no Amazon S3, ou referenciando uma classe Java personalizada, que fornece artefatos de criptografia. Para ter mais informações, consulte Fornecer certificados para criptografia de dados em trânsito com a criptografia do Amazon EMR.

Criar chaves e certificados para criptografia de dados

Antes de especificar as opções de criptografia usando uma configuração de segurança, decida qual provedor você quer usar para as chaves e os artefatos criptográficos. Por exemplo, você pode usar AWS KMS ou um provedor personalizado criado por você. Depois, crie as chaves ou o provedor de chaves conforme descrito nesta seção.

Fornecer chaves para criptografia de dados em repouso com o Amazon EMR

Você pode usar AWS Key Management Service (AWS KMS) ou um provedor de chave personalizado para criptografia de dados em repouso no Amazon EMR. Quando você usa AWS KMS, cobranças são cobradas pelo armazenamento e uso de chaves de criptografia. Para obter mais informações, consulte Preços do AWS KMS.

Este tópico fornece detalhes sobre políticas de chave para uma chave do KMS a ser usada com o Amazon EMR, bem como orientações e exemplos de código para escrever uma classe de provedor de chave personalizada para criptografia do Amazon S3. Para obter mais informações sobre como criar chaves, consulte Creating keys no Guia do desenvolvedor do AWS Key Management Service .

Usando AWS KMS keys para criptografia

A chave de AWS KMS criptografia deve ser criada na mesma região da sua instância de cluster do Amazon EMR e dos buckets do Amazon S3 usados com o EMRFS. Se a chave especificada estiver em uma conta diferente da que foi usada para configurar um cluster, será necessário especificar a chave usando o respectivo ARN.

O perfil do perfil de instância do Amazon EC2 deverá ter permissões para usar a chave do KMS que você especificar. O perfil padrão para o perfil de instância no Amazon EMR é `EMR_EC2_DefaultRole`. Se você usar um perfil diferente para o perfil de instância ou usar perfis do

IAM para solicitações do EMRFS para o Amazon S3, certifique-se de que cada perfil seja adicionado como um usuário de chave, conforme o caso. Isso concede ao perfil permissões para usar a chave do KMS. Para obter mais informações, consulte Using Key Policies no Guia do desenvolvedor do AWS Key Management Service e Configure IAM roles for EMRFS requests to Amazon S3.

Você pode usar o AWS Management Console para adicionar seu perfil de instância ou perfil de instância do EC2 à lista de usuários principais da chave KMS especificada, ou você pode usar o AWS CLI ou um AWS SDK para anexar uma política de chaves apropriada.

O Amazon EMR oferece suporte somente a chaves do KMS simétricas. Não é possível usar uma chave do KMS assimétrica para criptografar dados em repouso em um cluster do Amazon EMR. Para obter ajuda para determinar se uma chave do KMS é simétrica ou assimétrica, consulte Identifying symmetric and asymmetric KMS keys.

O procedimento abaixo descreve como adicionar o perfil de instância do Amazon EMR padrão, `EMR_EC2_DefaultRole`, como um usuário de chave usando o AWS Management Console. Ele pressupõe que você já tenha criado uma chave do KMS. Para criar uma nova chave do KMS, consulte Creating Keys no Guia do desenvolvedor do AWS Key Management Service .

Adicionar o perfil de instância do EC2 para Amazon EMR à lista de usuários de chaves de criptografia

1. Faça login no console AWS Management Console e abra o AWS Key Management Service (AWS KMS) em https://console.aws.amazon.com/kms.
2. Para alterar o Região da AWS, use o seletor de região no canto superior direito da página.
3. Selecione o alias da chave do KMS a ser modificada.
4. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
5. Na caixa de diálogo Add key users (Adicionar usuários da chave) selecione a função apropriada. O nome da função padrão é `EMR_EC2_DefaultRole`.
6. Escolha Adicionar.

Habilitar a criptografia do EBS fornecendo permissões adicionais para chaves do KMS

A partir do Amazon EMR versão 5.24.0, você pode criptografar o dispositivo raiz do EBS e os volumes de armazenamento usando uma opção de configuração de segurança. Para ativar essa opção, você deve especificar AWS KMS como seu provedor de chaves. Além disso, você deve conceder à função `EMR_DefaultRole` de serviço permissões para usar o AWS KMS key que você especificar.

Você pode usar o AWS Management Console para adicionar a função de serviço à lista de usuários principais da chave KMS especificada ou usar o AWS CLI ou um AWS SDK para anexar uma política de chaves apropriada.

O procedimento a seguir descreve como usar o AWS Management Console para adicionar a função de serviço padrão do Amazon EMR `EMR_DefaultRole` como um usuário chave. Ele pressupõe que você já tenha criado uma chave do KMS. Para criar uma nova chave do KMS, consulte Creating keys no Guia do desenvolvedor do AWS Key Management Service .

Para adicionar a função de serviço do Amazon EMR à lista de usuários da chave de criptografia

1. Faça login no console AWS Management Console e abra o AWS Key Management Service (AWS KMS) em https://console.aws.amazon.com/kms.
2. Para alterar o Região da AWS, use o seletor de região no canto superior direito da página.
3. Escolha Chaves gerenciadas pelo cliente na barra lateral esquerda.
4. Selecione o alias da chave do KMS a ser modificada.
5. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
6. Na seção Adicionar usuários-chave, selecione a função apropriada. O nome da função de serviço padrão do Amazon EMR é. `EMR_DefaultRole`
7. Escolha Adicionar.

Criar um provedor de chaves personalizado

Ao usar uma configuração de segurança, você deve especificar um nome de classe de provedor diferente para a criptografia de disco local e para a criptografia do Amazon S3. Os requisitos para o provedor de chaves personalizadas dependem de você usar criptografia de disco local e criptografia Amazon S3, bem como a versão de lançamento do Amazon EMR.

Dependendo do tipo de criptografia que você usa ao criar um provedor de chave personalizado, o aplicativo também deve implementar EncryptionMaterialsProvider interfaces diferentes. Ambas as interfaces estão disponíveis no AWS SDK for Java versão 1.11.0 e posterior.

- Para implementar a criptografia do Amazon S3, use o com.amazonaws.services.s3.model. EncryptionMaterialsProvider interface.
- Para implementar a criptografia de disco local, use com.amazonaws.services.elasticmapreduce.spi.security. EncryptionMaterialsProvider interface.

Você pode usar qualquer estratégia para fornecer materiais de criptografia para a implementação. Por exemplo, você pode optar por fornecer materiais de criptografia estática ou integrá-los a um sistema de gerenciamento de chaves mais complexo.

Se você estiver usando a criptografia do Amazon S3, deverá usar os algoritmos de criptografia NoPaddingAES/GCM/ para materiais de criptografia personalizados.

Se você estiver usando criptografia de disco local, o algoritmo de criptografia a ser usado para materiais de criptografia personalizados varia de acordo com a versão do EMR. Para o Amazon EMR 7.0.0 e versões anteriores, você deve usar AES/GCM/. NoPadding Para o Amazon EMR 7.1.0 e versões posteriores, você deve usar o AES.

A EncryptionMaterialsProvider classe obtém materiais de criptografia por contexto de criptografia. O Amazon EMR popula informações de contexto de criptografia em runtime para ajudar o chamador a determinar os materiais de criptografia corretos a serem retornados.

Example Exemplo: usar um provedor de chaves personalizado para a criptografia do Amazon S3 com o EMRFS

Quando o Amazon EMR busca os materiais de criptografia da EncryptionMaterialsProvider classe para realizar a criptografia, o EMRFS opcionalmente preenche o argumento MaterialsDescription com dois campos: o URI do Amazon S3 para o objeto JobFlowId e o do cluster, que pode ser usado pela classe para retornar materiais de criptografia seletivamente. EncryptionMaterialsProvider

Por exemplo, o provedor pode retornar diferentes chaves para diferentes prefixos de URI do Amazon S3. É a descrição dos materiais de criptografia retornados que acaba sendo armazenada com o objeto do Amazon S3 no lugar do valor de materialsDescription que é gerado pelo EMRFS e transmitido ao provedor. Ao descriptografar um objeto do Amazon S3, a descrição do material de criptografia é passada para a EncryptionMaterialsProvider classe, para que ela possa, novamente, retornar seletivamente a chave correspondente para descriptografar o objeto.

Uma implementação de EncryptionMaterialsProvider referência é fornecida abaixo. Outro provedor personalizado, o EMRFSRSA EncryptionMaterials Provider, está disponível em. GitHub

```
import com.amazonaws.services.s3.model.EncryptionMaterials;
import com.amazonaws.services.s3.model.EncryptionMaterialsProvider;
import com.amazonaws.services.s3.model.KMSEncryptionMaterials;
import org.apache.hadoop.conf.Configurable;
import org.apache.hadoop.conf.Configuration;

import java.util.Map;
```

```
/**
 * Provides KMSEncryptionMaterials according to Configuration
 */
public class MyEncryptionMaterialsProviders implements EncryptionMaterialsProvider,
 Configurable{
  private Configuration conf;
  private String kmsKeyId;
  private EncryptionMaterials encryptionMaterials;

  private void init() {
    this.kmsKeyId = conf.get("my.kms.key.id");
    this.encryptionMaterials = new KMSEncryptionMaterials(kmsKeyId);
  }

  @Override
  public void setConf(Configuration conf) {
    this.conf = conf;
    init();
  }

  @Override
  public Configuration getConf() {
    return this.conf;
  }

  @Override
  public void refresh() {

  }

  @Override
  public EncryptionMaterials getEncryptionMaterials(Map<String, String>
 materialsDescription) {
    return this.encryptionMaterials;
  }

  @Override
  public EncryptionMaterials getEncryptionMaterials() {
    return this.encryptionMaterials;
  }
}
```

Fornecer certificados para criptografia de dados em trânsito com a criptografia do Amazon EMR

Com a versão 4.8.0 ou posteriores do Amazon EMR, há duas opções para especificar artefatos para a criptografia de dados em trânsito usando uma configuração de segurança:

- É possível criar certificados PEM manualmente, incluí-los em um arquivo .zip e referenciar o arquivo .zip no Amazon S3.
- É possível implementar um provedor de certificados personalizado como uma classe Java. Especifique o arquivo JAR da aplicação no Amazon S3 e depois forneça o nome completo da classe desse provedor, conforme declarado na aplicação. A classe deve implementar a ArtifactsProvider interface TLS disponível a partir da AWS SDK for Java versão 1.11.0.

O Amazon EMR baixa os artefatos automaticamente em cada nó do cluster e depois os utiliza para implementar os recursos de criptografia em trânsito de código aberto. Para obter mais informações sobre as opções disponíveis, consulte Criptografia em trânsito.

Usar certificados PEM

Quando você especifica um arquivo de .zip para criptografia em trânsito, a configuração de segurança espera que os arquivos PEM dentro do arquivo .zip tenham exatamente os nomes indicados abaixo:

Certificados de criptografia em trânsito

Nome do arquivo	Obrigatório/opcional	Detalhes
privateKey.pem	Obrigatório	Chave privada
certificateChain.pem	Obrigatório	Cadeia de certificados
trustedCertificates.pem	Opcional	Necessário se o certificado fornecido não estiver assinado pela autoridade de certificação raiz confiável (AC) padrão Java ou por uma AC intermediária que possa estabelecer um vínculo com a AC raiz confiável Java padrão. As ACs

Nome do arquivo	Obrigatório/opcional	Detalhes
		raiz confiáveis Java padrão podem ser encontradas em `jre/lib/security/cacerts` .

Você provavelmente deseja configurar o arquivo PEM de chave particular para ser um certificado curinga que permite o acesso ao domínio da Amazon VPC no qual as suas instâncias de cluster residem. Por exemplo, se o seu cluster reside em us-east-1 (Norte da Virgínia), você pode optar por especificar um nome comum na configuração do certificado que permita o acesso ao cluster, especificando `CN=*.ec2.internal` na definição de requerente do certificado. Se o seu cluster residir em us-west-2 (Oregon), poderá especificar `CN=*.us-west-2.compute.internal`.

Se o arquivo PEM fornecido no artefato de criptografia não tiver um caractere curinga no CN para o domínio, será necessário alterar o valor de `hadoop.ssl.hostname.verifier` para `ALLOW_ALL`. Isso é feito com a classificação `core-site` ao enviar configurações para um cluster ou ao adicionar esse valor ao arquivo `core-site.xml`. Essa alteração é necessária porque o verificador de nome de host padrão não aceitará um nome de host sem curinga, resultando em um erro. Para obter mais informações sobre a configuração do cluster do EMR em uma Amazon VPC, consulte Configurar redes.

O exemplo a seguir demonstra como usar OpenSSL para gerar um certificado X.509 autoassinado com uma chave privada RSA de 1.024 bits. A chave permite o acesso às instâncias de cluster do Amazon EMR do emissor na região `us-west-2` (Oregon) conforme especificado pelo nome de domínio *`*.us-west-2.compute.internal`* como o nome comum.

Outros itens de requerente opcionais como país (C), estado (S) e Localidade (L) são especificados. Como um certificado autoassinado é gerado, o segundo comando no exemplo copia o arquivo `certificateChain.pem` no arquivo `trustedCertificates.pem`. O terceiro comando usa `zip` para criar o arquivo `my-certs.zip` que contém os certificados.

> **Important**
>
> Este exemplo é apenas uma proof-of-concept demonstração. O uso de certificados autoassinados não é recomendado e apresenta um possível risco de segurança. Para

sistemas de produção, use uma autoridade de certificação (AC) confiável para emitir certificados.

```
$ openssl req -x509 -newkey rsa:1024 -keyout privateKey.pem -out certificateChain.pem
 -days 365 -nodes -subj '/C=US/ST=Washington/L=Seattle/O=MyOrg/OU=MyDept/CN=*.us-
west-2.compute.internal'
$ cp certificateChain.pem trustedCertificates.pem
$ zip -r -X my-certs.zip certificateChain.pem privateKey.pem trustedCertificates.pem
```

AWS Identity and Access Management para Amazon EMR

AWS Identity and Access Management (IAM) é uma ferramenta AWS service (Serviço da AWS) que ajuda o administrador a controlar com segurança o acesso aos AWS recursos. Os administradores do IAM controlam quem pode ser autenticado (conectado) e autorizado (ter permissões) para utilizar os recursos do Amazon EMR. O IAM é um AWS service (Serviço da AWS) que você pode usar sem custo adicional.

Tópicos

Público

A forma como você usa AWS Identity and Access Management (IAM) difere, dependendo do trabalho que você faz no Amazon EMR.

Usuário do serviço: se você usar o serviço do Amazon EMR para fazer seu trabalho, o administrador fornecerá as credenciais e as permissões necessárias. À medida que mais recursos do Amazon EMR forem usados para realizar o trabalho, talvez sejam necessárias permissões adicionais.

Entender como o acesso é gerenciado pode ajudar você a solicitar as permissões corretas ao seu administrador. Se você não conseguir acessar um atributo no Amazon EMR, consulte Solução de problemas de identidade e acesso da Amazon EMR.

Administrador do serviço: se você for o responsável pelos recursos do Amazon EMR em sua empresa, provavelmente terá acesso total ao Amazon EMR. Cabe a você determinar quais funcionalidades e recursos do Amazon EMR os usuários do serviço deverão acessar. Assim, você deve enviar solicitações ao administrador do IAM para alterar as permissões dos usuários de seu serviço. Revise as informações nesta página para entender os Introdução ao IAM. Para saber mais sobre como sua empresa pode usar o IAM com o Amazon EMR, consulte Como o Amazon EMR funciona com o IAM.

Administrador do IAM: se você for um administrador do IAM, talvez deseje saber detalhes sobre como escrever políticas para gerenciar o acesso ao Amazon EMR. Para visualizar exemplos de políticas baseadas em identidade do Amazon EMR que podem ser usadas no IAM, consulte Exemplos de políticas baseadas em identidade do Amazon EMR.

Autenticando com identidades

A autenticação é como você faz login AWS usando suas credenciais de identidade. Você deve estar autenticado (conectado AWS) como o Usuário raiz da conta da AWS, como usuário do IAM ou assumindo uma função do IAM.

Você pode entrar AWS como uma identidade federada usando credenciais fornecidas por meio de uma fonte de identidade. AWS IAM Identity Center Usuários (IAM Identity Center), a autenticação de login único da sua empresa e suas credenciais do Google ou do Facebook são exemplos de identidades federadas. Quando você faz login como identidade federada, o administrador já configurou anteriormente a federação de identidades usando perfis do IAM. Ao acessar AWS usando a federação, você está assumindo indiretamente uma função.

Dependendo do tipo de usuário que você é, você pode entrar no AWS Management Console ou no portal de AWS acesso. Para obter mais informações sobre como fazer login em AWS, consulte Como fazer login Conta da AWS no Guia do Início de Sessão da AWS usuário.

Se você acessar AWS programaticamente, AWS fornece um kit de desenvolvimento de software (SDK) e uma interface de linha de comando (CLI) para assinar criptograficamente suas solicitações usando suas credenciais. Se você não usa AWS ferramentas, você mesmo deve assinar as solicitações. Para obter mais informações sobre como usar o método recomendado para assinar

solicitações por conta própria, consulte Assinatura de solicitações de AWS API no Guia do usuário do IAM.

Independente do método de autenticação usado, também pode ser exigido que você forneça informações adicionais de segurança. Por exemplo, AWS recomenda que você use a autenticação multifator (MFA) para aumentar a segurança da sua conta. Para saber mais, consulte Autenticação Multifator no AWS IAM Identity Center Guia do Usuário. Usar a autenticação multifator (MFA) na AWS no Guia do Usuário do IAM.

Conta da AWS usuário root

Ao criar uma Conta da AWS, você começa com uma identidade de login que tem acesso completo a todos Serviços da AWS os recursos da conta. Essa identidade é chamada de usuário Conta da AWS raiz e é acessada fazendo login com o endereço de e-mail e a senha que você usou para criar a conta. É altamente recomendável não usar o usuário raiz para tarefas diárias. Proteja as credenciais do usuário raiz e use-as para executar as tarefas que somente ele pode executar. Para obter a lista completa das tarefas que exigem login como usuário raiz, consulte Tarefas que exigem credenciais de usuário raiz no Guia do usuário do IAM.

Identidade federada

Como prática recomendada, exija que usuários humanos, incluindo usuários que precisam de acesso de administrador, usem a federação com um provedor de identidade para acessar Serviços da AWS usando credenciais temporárias.

Uma identidade federada é um usuário do seu diretório de usuários corporativo, de um provedor de identidade da web AWS Directory Service, do diretório do Identity Center ou de qualquer usuário que acesse usando credenciais fornecidas Serviços da AWS por meio de uma fonte de identidade. Quando as identidades federadas são acessadas Contas da AWS, elas assumem funções, e as funções fornecem credenciais temporárias.

Para o gerenciamento de acesso centralizado, recomendamos usar o . AWS IAM Identity Center Você pode criar usuários e grupos no IAM Identity Center ou pode se conectar e sincronizar com um conjunto de usuários e grupos em sua própria fonte de identidade para uso em todos os seus Contas da AWS aplicativos. Para obter mais informações sobre o Centro de Identidade do IAM, consulte O que é o Centro de Identidade do IAM? no AWS IAM Identity Center Manual do Usuário do.

Usuários e grupos do IAM

Um usuário do IAM é uma identidade dentro da sua Conta da AWS que tem permissões específicas para uma única pessoa ou aplicativo. Sempre que possível, recomendamos depender de credenciais temporárias em vez de criar usuários do IAM com credenciais de longo prazo, como senhas e chaves de acesso. No entanto, se você tiver casos de uso específicos que exijam credenciais de longo prazo com usuários do IAM, recomendamos alternar as chaves de acesso. Para obter mais informações, consulte Alterne Chaves de Acesso Regularmente para Casos de Uso que exijam Credenciais de Longo Prazo no Guia do Usuário do IAM.

Um grupo do IAM é uma identidade que especifica uma coleção de usuários do IAM. Não é possível fazer login como um grupo. É possível usar grupos para especificar permissões para vários usuários de uma vez. Os grupos facilitam o gerenciamento de permissões para grandes conjuntos de usuários. Por exemplo, você pode ter um nome de grupo IAMAdmins e atribuir a esse grupo permissões para administrar recursos do IAM.

Usuários são diferentes de perfis. Um usuário é exclusivamente associado a uma pessoa ou a um aplicativo, mas uma função pode ser assumida por qualquer pessoa que precisar dela. Os usuários têm credenciais permanentes de longo prazo, mas os perfis fornecem credenciais temporárias. Para saber mais, consulte Quando Criar um Usuário do IAM (Ao Invés de uma Função) no Guia do Usuário do IAM.

Perfis do IAM

Uma função do IAM é uma identidade dentro da sua Conta da AWS que tem permissões específicas. Ele é semelhante a um usuário do IAM, mas não está associado a uma pessoa específica. Você pode assumir temporariamente uma função do IAM no AWS Management Console trocando de funções. Você pode assumir uma função chamando uma operação de AWS API AWS CLI ou usando uma URL personalizada. Para obter mais informações sobre métodos para usar perfis, consulte Usando Funções do IAM no Guia do Usuário do IAM.

Funções do IAM com credenciais temporárias são úteis nas seguintes situações:

- Acesso de usuário federado: para atribuir permissões a identidades federadas, você pode criar um perfil e definir permissões para ele. Quando uma identidade federada é autenticada, essa identidade é associada ao perfil e recebe as permissões definidas pelo mesmo. Para obter mais informações sobre perfis para federação, consulte Criando um Perfil para um Provedor de Identidades Terceirizado no Guia do Usuário do IAM. Se você usa o IAM Identity Center, configure um conjunto de permissões. Para controlar o que suas identidades podem acessar após

a autenticação, o IAM Identity Center correlaciona o conjunto de permissões a um perfil no IAM. Para obter informações sobre conjuntos de permissões, consulte Conjuntos de Permissões no AWS IAM Identity Center Manual do Usuário.

- Permissões de usuários temporárias do IAM: um usuário ou perfil do IAM pode assumir um perfil do IAM para obter temporariamente permissões diferentes para uma tarefa específica.
- Acesso entre contas: você pode usar um perfil do IAM para permitir que alguém (uma entidade principal confiável) acesse recursos na sua conta de uma conta diferente. As funções são a forma primária de conceder acesso entre contas. No entanto, com alguns Serviços da AWS, você pode anexar uma política diretamente a um recurso (em vez de usar uma função como proxy). Para aprender a diferença entre funções e políticas baseadas em recurso para acesso entre contas, consulte Como as Funções do IAM Diferem das Políticas Baseadas em Recurso no Guia do Usuário do IAM.
- Acesso entre serviços — Alguns Serviços da AWS usam recursos em outros Serviços da AWS. Por exemplo, quando você faz uma chamada em um serviço, é comum que esse serviço execute aplicativos no Amazon EC2 ou armazene objetos no Amazon S3. Um serviço pode fazer isso usando as permissões de chamada da entidade principal, uma função de serviço ou uma função vinculada ao serviço.
  - Sessões de acesso direto (FAS) — Quando você usa um usuário ou uma função do IAM para realizar ações AWS, você é considerado principal. Ao usar alguns serviços, você pode executar uma ação que inicia outra ação em um serviço diferente. O FAS usa as permissões do diretor chamando um AWS service (Serviço da AWS), combinadas com a solicitação AWS service (Serviço da AWS) para fazer solicitações aos serviços posteriores. As solicitações do FAS são feitas somente quando um serviço recebe uma solicitação que requer interações com outros Serviços da AWS ou com recursos para ser concluída. Nesse caso, você precisa ter permissões para executar ambas as ações. Para obter detalhes da política ao fazer solicitações de FAS, consulte Encaminhar sessões de acesso.
  - Função de Serviço: uma função de serviço é uma função do IAM que um serviço assume para realizar ações em seu nome. Um administrador do IAM pode criar, modificar e excluir um perfil de serviço do IAM. Para obter mais informações, consulte Criando um Perfil para Delegar Permissões a um AWS service (Serviço da AWS) no Guia do Usuário do IAM.
  - Função vinculada ao serviço — Uma função vinculada ao serviço é um tipo de função de serviço vinculada a um. AWS service (Serviço da AWS) O serviço pode assumir o perfil de executar uma ação em seu nome. As funções vinculadas ao serviço aparecem em você Conta da AWS e são de propriedade do serviço. Um administrador do IAM pode visualizar, mas não pode editar as permissões para funções vinculadas a serviço.

- Aplicativos em execução no Amazon EC2 — Você pode usar uma função do IAM para gerenciar credenciais temporárias para aplicativos que estão sendo executados em uma instância do EC2 e fazendo AWS CLI solicitações de API. AWS É preferível fazer isso a armazenar chaves de acesso na instância do EC2. Para atribuir uma AWS função a uma instância do EC2 e disponibilizá-la para todos os seus aplicativos, você cria um perfil de instância anexado à instância. Um perfil de instância contém a perfil e permite que os programas em execução na instância do EC2 obtenham credenciais temporárias. Para mais informações, consulte Usar uma função do IAM para conceder permissões a aplicativos em execução nas instâncias do Amazon EC2 no Guia do usuário do IAM.

Para aprender se deseja usar perfis do IAM, consulte Quando Criar uma Função do IAM (em Vez de um Usuário) no Guia do Usuário do IAM.

Gerenciando acesso usando políticas

Você controla o acesso AWS criando políticas e anexando-as a AWS identidades ou recursos. Uma política é um objeto AWS que, quando associada a uma identidade ou recurso, define suas permissões. AWS avalia essas políticas quando um principal (usuário, usuário raiz ou sessão de função) faz uma solicitação. As permissões nas políticas determinam se a solicitação será permitida ou negada. A maioria das políticas é armazenada AWS como documentos JSON. Para obter mais informações sobre a estrutura e o conteúdo de documentos de políticas JSON, consulte Visão Geral das Políticas JSON no Guia do Usuário do IAM.

Os administradores podem usar políticas AWS JSON para especificar quem tem acesso ao quê. Ou seja, qual entidade principal pode executar ações em quais recursos e em que condições.

Por padrão, usuários e funções não têm permissões. Para conceder aos usuários permissão para executar ações nos recursos de que eles precisam, um administrador do IAM pode criar políticas do IAM. O administrador pode então adicionar as políticas do IAM às funções e os usuários podem assumir as funções.

As políticas do IAM definem permissões para uma ação, independente do método usado para executar a operação. Por exemplo, suponha que você tenha uma política que permite a ação `iam:GetRole`. Um usuário com essa política pode obter informações de função da AWS Management Console AWS CLI, da ou da AWS API.

Políticas baseadas em identidade

As políticas baseadas em identidade são documentos de políticas de permissões JSON que você pode anexar a uma identidade, como usuário do IAM, grupo de usuários ou perfil do IAM. Essas

políticas controlam quais ações os usuários e funções podem realizar, em quais recursos e em quais condições. Para saber como criar uma política baseada em identidade, consulte Criar políticas do IAM no Guia do usuário do IAM.

As políticas baseadas em identidade também podem ser categorizadas como políticas em linha ou políticas gerenciadas. As políticas em linha são incorporadas diretamente a um único usuário, grupo ou função. As políticas gerenciadas são políticas autônomas que você pode associar a vários usuários, grupos e funções em seu Conta da AWS. As políticas AWS gerenciadas incluem políticas gerenciadas e políticas gerenciadas pelo cliente. Para saber como selecionar entre uma política gerenciada ou uma política em linha, consulte Selecionar entre políticas gerenciadas e políticas em linha no Guia do usuário do IAM.

Políticas baseadas em recursos

Políticas baseadas em recursos são documentos de políticas JSON que você anexa a um recurso. São exemplos de políticas baseadas em recursos as políticas de confiança de função do IAM e as políticas do bucket do Amazon S3. Em serviços que suportem políticas baseadas em recursos, os administradores de serviço podem usá-las para controlar o acesso a um recurso específico. Para o recurso ao qual a política está anexada, a política define quais ações uma entidade principal especificada pode executar nesse recurso e em que condições. Você deve especificar uma entidade principal em uma política baseada em recursos. Os diretores podem incluir contas, usuários, funções, usuários federados ou. Serviços da AWS

Políticas baseadas em atributos são políticas em linha que estão localizadas nesse serviço. Você não pode usar políticas AWS gerenciadas do IAM em uma política baseada em recursos.

Listas de controle de acesso (ACLs)

As listas de controle de acesso (ACLs) controlam quais entidades principais (membros, usuários ou funções da conta) têm permissão para acessar um recurso. As ACLs são semelhantes as políticas baseadas em recursos, embora não usem o formato de documento de política JSON.

O Amazon S3 e o Amazon VPC são exemplos de serviços que oferecem suporte a ACLs. AWS WAF Saiba mais sobre ACLs em Configurações da lista de controle de acesso (ACL) no Guia do Desenvolvedor do Amazon Simple Storage Service.

Outros tipos de política

AWS oferece suporte a tipos de políticas adicionais menos comuns. Esses tipos de política podem definir o máximo de permissões concedidas a você pelos tipos de política mais comuns.

- Limites de permissões: um limite de permissões é um recurso avançado no qual você define o máximo de permissões que uma política baseada em identidade pode conceder a uma entidade do IAM (usuário ou perfil do IAM). É possível definir um limite de permissões para uma entidade. As permissões resultantes são a interseção das políticas baseadas em identidade de uma entidade e dos seus limites de permissões. As políticas baseadas em atributo que especificam o usuário ou o perfil no campo `Principal` não são limitadas pelo limite de permissões. Uma negação explícita em qualquer uma dessas políticas substitui a permissão. Para obter mais informações sobre limites de permissões, consulte Limites de Permissões para Entidades do IAM no Guia do Usuário do IAM.
- Políticas de controle de serviço (SCPs) — SCPs são políticas JSON que especificam as permissões máximas para uma organização ou unidade organizacional (OU) em. AWS Organizations AWS Organizations é um serviço para agrupar e gerenciar centralmente vários Contas da AWS que sua empresa possui. Se você habilitar todos os atributos em uma organização, poderá aplicar políticas de controle de serviço (SCPs) a qualquer uma ou a todas as contas. O SCP limita as permissões para entidades nas contas dos membros, incluindo cada uma Usuário raiz da conta da AWS. Para obter mais informações sobre o Organizações e SCPs, consulte Como os SCPs Funcionam no AWS Organizations Manual do Usuário do.
- Políticas de sessão: são políticas avançadas que você transmite como um parâmetro quando cria de forma programática uma sessão temporária para uma função ou um usuário federado. As permissões da sessão resultante são a interseção das políticas baseadas em identidade do usuário ou do perfil e das políticas de sessão. As permissões também podem ser provenientes de uma política baseada em atributo. Uma negação explícita em qualquer uma dessas políticas substitui a permissão. Para obter mais informações, consulte Políticas de sessão no Guia do usuário do IAM.

Vários tipos de política

Quando vários tipos de política são aplicáveis a uma solicitação, é mais complicado compreender as permissões resultantes. Para saber como AWS determinar se uma solicitação deve ser permitida quando vários tipos de políticas estão envolvidos, consulte Lógica de avaliação de políticas no Guia do usuário do IAM.

Como o Amazon EMR funciona com o IAM

Antes de usar o IAM para gerenciar o acesso ao Amazon EMR, entenda que atributos do IAM estão disponíveis para uso com o Amazon EMR.

Atributos do IAM que você pode usar com o Amazon EMR

Atributo do IAM	Suporte ao Amazon EMR
Políticas baseadas em identidade	Sim
Políticas baseadas em atributos	Sim
Ações das políticas	Sim
atributos de políticas	Sim
Chaves de condição de políticas	Sim
ACLs	Não
ABAC (tags em políticas)	Sim
Credenciais temporárias	Sim
Permissões de entidade principal	Sim
Perfis de serviço	Não
Funções vinculadas ao serviço	Sim

Para ter uma visão de alto nível de como o Amazon EMR e AWS outros serviços funcionam com a maioria dos recursos do IAM, AWS consulte os serviços que funcionam com o IAM no Guia do usuário do IAM.

Políticas baseadas em identidade do Amazon EMR

Suporta com políticas baseadas em identidade	Sim

As políticas baseadas em identidade são documentos de políticas de permissões JSON que você pode anexar a uma identidade, como usuário IAM, grupo de usuários ou perfil do IAM. Essas políticas controlam quais ações os usuários e funções podem realizar, em quais recursos e em quais condições. Saiba como criar uma política baseada em identidade consultando Criando Políticas do IAM no Guia do Usuário do IAM.

Com as políticas baseadas em identidade do IAM, é possível especificar ações ou recursos permitidos ou negados, assim como as condições sob as quais as ações são permitidas ou negadas. Você não pode especificar a entidade principal em uma política baseada em identidade porque ela se aplica ao usuário ou função à qual ela está anexada. Para saber mais sobre todos os elementos que podem ser usados em uma política JSON, consulte Referência de elementos da política JSON do IAM no Guia do Usuário do IAM.

Exemplos de políticas baseadas em identidade para o Amazon EMR

Para visualizar exemplos de políticas baseadas em identidade do Amazon EMR, consulte Exemplos de políticas baseadas em identidade do Amazon EMR.

Políticas baseadas em recursos no Amazon EMR

É compatível com políticas baseadas em atributos	Sim

Políticas baseadas em recursos são documentos de políticas JSON que você anexa a um recurso. São exemplos de políticas baseadas em recursos as políticas de confiança de função do IAM e as políticas do bucket do Amazon S3. Em serviços que suportem políticas baseadas em recursos, os administradores de serviço podem usá-las para controlar o acesso a um recurso específico. Para o recurso ao qual a política está anexada, a política define quais ações uma entidade principal especificada pode executar nesse recurso e em que condições. Você deve especificar uma entidade principal em uma política baseada em recursos. Os diretores podem incluir contas, usuários, funções, usuários federados ou. Serviços da AWS

Para permitir o acesso entre contas, você pode especificar uma conta inteira ou as entidades do IAM em outra conta como a entidade principal em uma política baseada em atributo. Adicionar uma entidade principal entre contas à política baseada em atributo é apenas metade da tarefa de estabelecimento da relação de confiança. Quando o principal e o recurso são diferentes Contas da AWS, um administrador do IAM na conta confiável também deve conceder permissão à entidade principal (usuário ou função) para acessar o recurso. Eles concedem permissão ao anexar uma política baseada em identidade para a entidade. No entanto, se uma política baseada em atributo conceder acesso a uma entidade principal na mesma conta, nenhuma política baseada em identidade adicional será necessária. Para obter mais informações, consulte Como os perfis do IAM diferem de políticas baseadas em recursos no Guia do usuário do IAM.

Ações de políticas para o Amazon EMR

Oferece suporte a ações de políticas	Sim

Os administradores podem usar políticas AWS JSON para especificar quem tem acesso ao quê. Ou seja, qual entidade principal pode executar ações em quais recursos, e em que condições.

O elemento `Action` de uma política JSON descreve as ações que você pode usar para permitir ou negar acesso em uma política. As ações de política geralmente têm o mesmo nome da operação de AWS API associada. Existem algumas exceções, como ações somente de permissão, que não têm uma operação de API correspondente. Há também algumas operações que exigem várias ações em uma política. Essas ações adicionais são chamadas de ações dependentes.

Incluem ações em uma política para conceder permissões para executar a operação associada.

Para visualizar uma lista de ações do Amazon EMR, consulte Ações, recursos e chaves de condição para o Amazon EMR na Referência de autorização do serviço.

As ações de política no Amazon EMR usam o seguinte prefixo antes da ação:

```
EMR
```

Para especificar várias ações em uma única instrução, separe-as com vírgulas.

```
"Action": [
      "EMR:action1",
      "EMR:action2"
         ]
```

Para visualizar exemplos de políticas baseadas em identidade do Amazon EMR, consulte Exemplos de políticas baseadas em identidade do Amazon EMR.

Recursos de políticas para o Amazon EMR

Oferece suporte a atributos de políticas	Sim

Os administradores podem usar políticas AWS JSON para especificar quem tem acesso ao quê. Ou seja, qual entidade principal pode executar ações em quais recursos, e em que condições.

O elemento `Resource` de política JSON especifica o objeto ou os objetos aos quais a ação se aplica. As instruções devem incluir um elemento `Resource` ou um elemento `NotResource`. Como prática recomendada, especifique um recurso usando seu nome do recurso da Amazon (ARN). Isso pode ser feito para ações que oferecem suporte a um tipo de atributo específico, conhecido como permissões em nível de atributo.

Para ações não compatíveis com permissões no nível de recurso, como operações de listagem, use um curinga (*) para indicar que a instrução se aplica a todos os recursos.

```
"Resource": "*"
```

Para ver uma lista dos tipos de recursos do Amazon EMR e seus ARNs, consulte Tipos de recursos definidos pelo Amazon EMR na Referência de autorização do serviço. Para saber com quais ações é possível especificar o ARN de cada recurso, consulte Ações, recursos e chaves de condição do Amazon EMR.

Para visualizar exemplos de políticas baseadas em identidade do Amazon EMR, consulte Exemplos de políticas baseadas em identidade do Amazon EMR.

Chaves de condição de política para o Amazon EMR

Suporta chaves de condição de política específicas de serviço	Sim

Os administradores podem usar políticas AWS JSON para especificar quem tem acesso ao quê. Ou seja, qual principal pode executar ações em quais recursos, e em que condições.

O elemento `Condition` (ou bloco `Condition`) permite especificar condições nas quais uma instrução estiver em vigor. O elemento `Condition` é opcional. Você pode criar expressões condicionais que usem operadores de condição, como “igual a” ou “menor que”, para corresponder a condição da política aos valores na solicitação.

Se você especificar vários elementos `Condition` em uma instrução ou várias chaves em um único `Condition` elemento, a AWS os avaliará usando uma operação lógica `AND`. Se você especificar

vários valores para uma única chave de condição, AWS avalia a condição usando uma `OR` operação lógica. Todas as condições devem ser atendidas antes que as permissões da instrução sejam concedidas.

Você também pode usar variáveis de espaço reservado ao especificar condições. Por exemplo, é possível conceder a um usuário do IAM permissão para acessar um atributo somente se ele estiver marcado com seu nome de usuário do IAM. Para obter mais informações, consulte Elementos de Política do IAM: Variáveis e Tags no Guia do Usuário do IAM.

AWS suporta chaves de condição globais e chaves de condição específicas do serviço. Para ver todas as chaves de condição AWS globais, consulte as chaves de contexto de condição AWS global no Guia do usuário do IAM.

Para obter uma lista de chaves de condição do Amazon EMR, das ações e dos recursos que você pode usar, consulte Ações, recursos e chaves de condição do Amazon EMR na Referência de autorização do serviço.

Para visualizar exemplos de políticas baseadas em identidade do Amazon EMR, consulte Exemplos de políticas baseadas em identidade do Amazon EMR.

Listas de controle de acesso (ACLs) no Amazon EMR

Oferece suporte a ACLs	Não

As listas de controle de acesso (ACLs) controlam quais entidades principais (membros, usuários ou funções da conta) têm permissões para acessar um recurso. As ACLs são semelhantes as políticas baseadas em recursos, embora não usem o formato de documento de política JSON.

Controle de acesso baseado em atributos (ABAC) com o Amazon EMR

Oferece suporte a ABAC (tags em políticas)	Sim

O controle de acesso por atributo (ABAC) é uma estratégia de autorização que define permissões com base em atributos. Em AWS, esses atributos são chamados de tags. Você pode anexar tags a entidades do IAM (usuários ou funções) e a vários AWS recursos. A marcação de entidades e atributos é a primeira etapa do ABAC. Em seguida, você cria políticas de ABAC para permitir

operações quando a tag da entidade principal corresponder à tag do recurso que ela estiver tentando acessar.

O ABAC é útil em ambientes que estão crescendo rapidamente e ajuda em situações em que o gerenciamento de políticas se torna um problema.

Para controlar o acesso baseado em tags, forneça informações sobre a tag no elemento de condição de uma política usando as chaves de condição `aws:ResourceTag/key-name`, `aws:RequestTag/key-name` ou `aws:TagKeys`.

Se um serviço oferecer suporte às três chaves de condição para todo tipo de recurso, o valor será Sim para o serviço. Se um serviço oferecer suporte às três chaves de condição somente para alguns tipos de recursos, o valor será Parcial.

Para obter mais informações sobre o ABAC, consulte O que é ABAC? no Guia do Usuário do IAM. Para visualizar um tutorial com etapas para configurar o ABAC, consulte Usar Controle de Acesso Baseado em Atributos (ABAC) no Guia do Usuário do IAM.

Uso de credenciais temporárias com o Amazon EMR

Oferece suporte a credenciais temporárias	Sim

Alguns Serviços da AWS não funcionam quando você faz login usando credenciais temporárias. Para obter informações adicionais, incluindo quais Serviços da AWS funcionam com credenciais temporárias, consulte Serviços da AWS “Trabalhe com o IAM” no Guia do usuário do IAM.

Você está usando credenciais temporárias se fizer login AWS Management Console usando qualquer método, exceto um nome de usuário e senha. Por exemplo, quando você acessa AWS usando o link de login único (SSO) da sua empresa, esse processo cria automaticamente credenciais temporárias. Você também cria automaticamente credenciais temporárias quando faz login no console como usuário e, em seguida, alterna perfis. Para obter mais informações sobre como alternar funções, consulte Alternar para uma Função (Console) no Guia do Usuário do IAM.

Você pode criar manualmente credenciais temporárias usando a AWS API AWS CLI ou. Em seguida, você pode usar essas credenciais temporárias para acessar AWS. AWS recomenda que você gere credenciais temporárias dinamicamente em vez de usar chaves de acesso de longo prazo. Para mais informações, consulte Credenciais de segurança temporárias no IAM.

Permissões de entidades principais entre serviços para o Amazon EMR

Suporte para o recurso Encaminhamento de sessões de acesso (FAS)	Sim

Quando você usa um usuário ou uma função do IAM para realizar ações AWS, você é considerado principal. Ao usar alguns serviços, você pode executar uma ação que inicia outra ação em um serviço diferente. O FAS usa as permissões do diretor chamando um AWS service (Serviço da AWS), combinadas com a solicitação AWS service (Serviço da AWS) para fazer solicitações aos serviços posteriores. As solicitações do FAS são feitas somente quando um serviço recebe uma solicitação que requer interações com outros Serviços da AWS ou com recursos para ser concluída. Nesse caso, você precisa ter permissões para executar ambas as ações. Para obter detalhes da política ao fazer solicitações de FAS, consulte Encaminhar sessões de acesso.

Perfis de serviço para o Amazon EMR

Oferece suporte a perfis de serviço	Não

Perfis vinculados ao serviço para Amazon EMR

Oferece suporte a perfis vinculados ao serviço	Sim

Para obter detalhes sobre como criar ou gerenciar funções vinculadas a serviços, consulte AWS Serviços que Funcionam com o IAM. Encontre um serviço na tabela que inclua um `Yes` na coluna Função Vinculada ao Serviço.. Escolha o link Sim para visualizar a documentação do perfil vinculado ao serviço desse serviço.

Usar etiquetas de caderno e cluster com as políticas de controle de acesso do IAM

A permissão de ações associadas do Amazon EMR com Cadernos do EMR e clusters do EMR pode ser ajustada em detalhes usando o controle de acesso baseado em etiquetas com políticas do IAM baseadas em identidade. Você pode usar chaves de condição em um elemento `Condition` (também denominado bloco `Condition`) para permitir determinadas ações somente quando um bloco de anotações, um cluster ou ambos têm uma chave de tag ou uma combinação de chave-

valor. Também é possível limitar a ação `CreateEditor` (que cria um Caderno do EMR) e a ação `RunJobFlow` (que cria um cluster) para que a solicitação de uma etiqueta tenha que ser enviada quando o recurso for criado.

No Amazon EMR, as chaves de condição que podem ser usadas em um elemento `Condition` se aplicam somente às ações de API do Amazon EMR em que `ClusterID` ou `NotebookID` seja um parâmetro de solicitação obrigatório. Por exemplo, a ação ModifyInstanceGrupos não oferece suporte a chaves de contexto porque `ClusterID` é um parâmetro opcional.

Quando você cria um Caderno do EMR, uma etiqueta padrão é aplicada com uma string de chave de `creatorUserId` definida como o valor do ID de usuário do IAM que criou o caderno. Isso é útil para limitar as ações permitidas para o bloco de anotações apenas ao criador.

As seguintes chaves de condição estão disponíveis no Amazon EMR:

- Use a chave de contexto de condição `elasticmapreduce:ResourceTag/`*`TagKeyString`*, para permitir ou negar ações do usuário em clusters ou blocos de anotações com tags que tenham a *`TagKeyString`* especificada. Se uma ação passar `ClusterID` e `NotebookID`, a condição se aplicará ao cluster e ao bloco de anotações. Isso significa que ambos os recursos devem ter a string de chave de tag ou uma combinação de chave-valor que você especificar. Você pode usar o elemento `Resource` para limitar a declaração para que ela se aplique apenas a clusters ou blocos de anotações, conforme necessário. Para ter mais informações, consulte Exemplos de políticas baseadas em identidade do Amazon EMR.
- Use a chave de contexto de condição `elasticmapreduce:RequestTag/`*`TagKeyString`* para exigir uma tag específica com ações/chamadas de API. Por exemplo, é possível usar essa chave de contexto de condição juntamente com a ação `CreateEditor` para exigir que uma chave com *`TagKeyString`* seja aplicada a um bloco de anotações quando ele é criado.

Exemplos

Para ver uma lista das ações do Amazon EMR, consulte Ações definidas pelo Amazon EMR no Guia do usuário do IAM.

Perfis de runtime para etapas ao Amazon EMR

Uma função de tempo de execução é uma função AWS Identity and Access Management (IAM) que você pode especificar ao enviar um trabalho ou uma consulta para um cluster do Amazon EMR. O trabalho ou consulta que você envia ao seu cluster do Amazon EMR usa a função de tempo de

execução para acessar AWS recursos, como objetos no Amazon S3. Você pode especificar perfis de runtime com o Amazon EMR para trabalhos do Spark e do Hive.

Também é possível especificar perfis de runtime ao se conectar aos clusters do Amazon EMR no Amazon SageMaker e ao anexar um Amazon EMR Studio Workspace a um cluster do EMR. Para obter mais informações, consulte Connect to an Amazon EMR cluster from Studio e Execução de um Workspace do EMR Studio com um perfil de runtime.

Antigamente, os clusters do Amazon EMR executavam trabalhos ou consultas do Amazon EMR com permissões com base na política do IAM anexada ao perfil de instância usado para iniciar o cluster. Assim, as políticas precisavam conter a união de todas as permissões para todos os trabalhos e consultas executados em um cluster do Amazon EMR. Com os perfis de runtime, já é possível gerenciar o controle de acesso para cada trabalho ou consulta individualmente, em vez de compartilhar o perfil de instância do Amazon EMR do cluster.

Nos clusters do Amazon EMR com funções de tempo de execução, você também pode aplicar controle de acesso AWS Lake Formation baseado às tarefas e consultas do Spark, Hive e Presto em seus lagos de dados. Para saber mais sobre como fazer a integração com AWS Lake Formation, consulteIntegre o Amazon EMR com AWS Lake Formation.

> **Note**
>
> Quando você especifica uma função de tempo de execução para uma etapa do Amazon EMR, os trabalhos ou consultas que você envia só podem acessar AWS recursos que as políticas anexadas à função de tempo de execução permitem. Esses trabalhos e consultas não poderão acessar o serviço de metadados de instância nas instâncias do EC2 do cluster nem usar o perfil de instância do EC2 do cluster para acessar recursos da AWS .

Pré-requisitos para iniciar um cluster do Amazon EMR com um perfil de runtime

Tópicos

- Etapa 1: definir configurações de segurança no Amazon EMR
- Etapa 2: configurar um perfil de instância do EC2 para o cluster do Amazon EMR
- Etapa3: configurar uma política de confiança

Etapa 1: definir configurações de segurança no Amazon EMR

Use a estrutura JSON a seguir para criar uma configuração de segurança no AWS Command Line Interface (AWS CLI) e `EnableApplicationScopedIAMRole` defina `true` como. Para obter mais informações sobre configurações de segurança, consulte Usar configurações de segurança para definir a segurança do cluster.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true
        }
    }
}
```

É recomendável habilitar sempre as opções de criptografia em trânsito na configuração de segurança, para que os dados transferidos pela Internet sejam criptografados, em vez de em texto sem formatação. Você pode ignorar essas opções se não quiser se conectar aos clusters do Amazon EMR com funções de tempo de execução do Runtime Studio ou SageMaker do EMR Studio. Para configurar a criptografia de dados, consulte Configure data encryption.

Como alternativa, você pode criar uma configuração de segurança com configurações personalizadas usando o AWS Management Console.

Etapa 2: configurar um perfil de instância do EC2 para o cluster do Amazon EMR

Os clusters do Amazon EMR usam o perfil do perfil de instância do Amazon EC2 para assumir os perfis de runtime. Para usar perfis de runtime com etapas do Amazon EMR, adicione as políticas a seguir ao perfil do IAM que você planeja usar como perfil do perfil de instância. Para adicionar políticas a um perfil do IAM ou editar uma política em linha ou gerenciada já existente, consulte Adicionar e remover permissões de identidade do IAM.

```
{
    "Version":"2012-10-17",
    "Statement":[
        {
            "Sid":"AllowRuntimeRoleUsage",
            "Effect":"Allow",
            "Action":[
                "sts:AssumeRole",
                "sts:TagSession"
```

```
            ],
            "Resource":[
                <runtime-role-ARN>
            ]
        }
    ]
}
```

Etapa3: configurar uma política de confiança

Para cada perfil do IAM que você pretende usar como perfil de runtime, defina a política de confiança a seguir, substituindo `EMR_EC2_DefaultRole` pelo perfil do perfil de instância. Para modificar a política de confiança de um perfil do IAM, consulte Modificar a política de confiança de um perfil.

```
{
    "Sid":"AllowAssumeRole",
    "Effect":"Allow",
    "Principal":{
        "AWS":"arn:aws:iam::<AWS_ACCOUNT_ID>:role/EMR_EC2_DefaultRole"
    },
    "Action":"sts:AssumeRole"
}
```

Iniciar um cluster do Amazon EMR com controle de acesso baseado em perfis

Após definir suas configurações, você pode iniciar um cluster do Amazon EMR com a configuração de segurança de Etapa 1: definir configurações de segurança no Amazon EMR. Para usar perfis de runtime com etapas do Amazon EMR, use o rótulo de versão `emr-6.7.0` ou posteriores e selecione Hive, Spark ou ambos como aplicação de cluster. Para se conectar a partir do SageMaker Studio, use release `emr-6.9.0` ou posterior e selecione Livy, Spark, Hive ou Presto como seu aplicativo de cluster. Para obter instruções sobre como iniciar seu cluster, consulte Especificar uma configuração de segurança para um cluster.

Enviar trabalhos do Spark usando as etapas do Amazon EMR

Veja a seguir um exemplo de como executar o HdfsTest exemplo incluído no Apache Spark. Essa chamada de API só terá êxito se o perfil de runtime fornecido do Amazon EMR puder acessar o `S3_LOCATION`.

```
RUNTIME_ROLE_ARN=<runtime-role-arn>
S3_LOCATION=<s3-path>
```

```
REGION=<aws-region>
CLUSTER_ID=<cluster-id>

aws emr add-steps --cluster-id $CLUSTER_ID \
--steps '[{ "Name": "Spark Example", "ActionOnFailure": "CONTINUE","HadoopJarStep":
 { "Jar":"command-runner.jar","Args" : ["spark-example","HdfsTest",
 "$S3_LOCATION"] } }]' \
--execution-role-arn $RUNTIME_ROLE_ARN \
--region $REGION
```

> **Note**
>
> É recomendável desativar o acesso SSH ao cluster do Amazon EMR e permitir que somente a API `AddJobFlowSteps` do Amazon EMR acesse o cluster.

Enviar trabalhos do Hive usando as etapas do Amazon EMR

O exemplo a seguir usa as etapas do Apache Hive com o Amazon EMR para enviar um trabalho para executar o arquivo `QUERY_FILE.hql`. Essa consulta só será terá êxito se o perfil de runtime fornecido puder acessar o caminho do Amazon S3 do arquivo de consulta.

```
RUNTIME_ROLE_ARN=<runtime-role-arn>
REGION=<aws-region>
CLUSTER_ID=<cluster-id>

aws emr add-steps --cluster-id $CLUSTER_ID \
--steps '[{ "Name": "Run hive query using command-runner.jar - simple
 select","ActionOnFailure":"CONTINUE","HadoopJarStep": { "Jar": "command-
runner.jar","Args" :["hive -
f","s3://DOC_EXAMPLE_BUCKET/QUERY_FILE.hql"] } }]' \
--execution-role-arn $RUNTIME_ROLE_ARN \
--region $REGION
```

Conecte-se aos clusters do Amazon EMR com funções de tempo de execução a partir de um notebook Studio SageMaker

Você pode aplicar funções de tempo de execução do Amazon EMR às consultas que você executa nos clusters do Amazon EMR a partir do Studio. SageMaker Para isso, siga as etapas a seguir.

1. Siga as instruções em Inicie o Amazon SageMaker Studio para criar um SageMaker Studio.

2. Na interface do usuário do SageMaker Studio, inicie um notebook com kernels compatíveis. Por exemplo, inicie uma SparkMagic imagem com um PySpark kernel.
3. Escolha um cluster do Amazon EMR no SageMaker Studio e, em seguida, escolha Connect.
4. Escolha um perfil de runtime e escolha Conectar.

Isso criará uma célula de SageMaker notebook com comandos mágicos para se conectar ao seu cluster do Amazon EMR com a função de tempo de execução escolhida do Amazon EMR. Na célula do caderno, você pode inserir e executar consultas com perfil de runtime e controle de acesso baseado no Lake Formation. Para um exemplo mais detalhado, consulte Aplicar controles refinados de acesso a dados com o AWS Lake Formation Amazon EMR e o Amazon Studio. SageMaker

Controlar o acesso ao perfil de runtime do Amazon EMR

Você pode controlar o acesso ao perfil de runtime usando a chave de condição `elasticmapreduce:ExecutionRoleArn`. A política a seguir permite que uma entidade principal do IAM use um perfil do IAM chamado `Caller`, ou qualquer perfil do IAM que comece com a string `CallerTeamRole`, como o perfil de runtime.

> **Important**
>
> É necessário criar uma condição com base na chave de contexto `elasticmapreduce:ExecutionRoleArn` ao conceder a um chamador acesso para chamar as APIs `AddJobFlowSteps` ou `GetClusterSessionCredentials`, conforme mostra o exemplo a seguir.

```
{
    "Sid":"AddStepsWithSpecificExecRoleArn",
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:AddJobFlowSteps"
    ],
    "Resource":"*",
    "Condition":{
        "StringEquals":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::<AWS_ACCOUNT_ID>:role/Caller"
            ]
```

```
        },
        "StringLike":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::<AWS_ACCOUNT_ID>:role/CallerTeamRole*"
            ]
        }
    }
}
```

Estabelecer confiança entre os perfis de runtime e os clusters do Amazon EMR

O Amazon EMR gera um identificador exclusivo `ExternalId` para cada configuração de segurança com autorização de perfil de runtime ativada. Essa autorização permite que cada usuário tenha um conjunto de perfil de runtime para usar nos clusters que pertencem a eles. Por exemplo, em uma empresa, cada departamento pode usar o próprio ID externo para atualizar a política de confiança em seu próprio conjunto de perfis de runtime.

Você encontra o ID externo com a API `DescribeSecurityConfiguration` do Amazon EMR, conforme mostrado no exemplo a seguir.

```
aws emr describe-security-configuration --name 'iamconfig-with-lf'{"Name": "iamconfig-
with-lf",
    "SecurityConfiguration":
        "{\"AuthorizationConfiguration\":{\"IAMConfiguration\":
{\"EnableApplicationScopedIAMRole\
        ":true,\"ApplicationScopedIAMRoleConfiguration\":{\"PropagateSourceIdentity
\":true,\"Exter
        nalId\":\"FXH5TSACFDWUCDSR3YQE2O7ETPUSM4OBCGLYWODSCUZDNZ4Y\"}},\"Lake
        FormationConfiguration\":{\"AuthorizedSessionTagValue\":\"Amazon EMR\"}}}",
    "CreationDateTime": "2022-06-03T12:52:35.308000-07:00"
}
```

Para obter informações sobre como usar uma ID externa, consulte Como usar uma ID externa ao conceder acesso aos seus AWS recursos a terceiros.

Auditoria

Para monitorar e controlar as ações que os usuários finais realizam com os perfis do IAM, você pode ativar o atributo de identidade de origem. Para saber mais sobre a identidade de origem, consulte Monitorar e controlar ações realizadas com perfis assumidos.

Para rastrear a identidade de origem, defina `ApplicationScopedIAMRoleConfiguration/PropagateSourceIdentity` como `true` em sua configuração de segurança, como mostrado a seguir.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true,
            "ApplicationScopedIAMRoleConfiguration":{
                "PropagateSourceIdentity":true
            }
        }
    }
}
```

Quando você define `PropagateSourceIdentity` como `true`, o Amazon EMR aplica a identidade de origem das credenciais de chamada a um trabalho ou sessão de consulta que você cria com o perfil de runtime. Se não houver nenhuma identidade de origem nas credenciais de chamada, o Amazon EMR não definirá a identidade de origem.

Para usar essa propriedade, forneça permissões `sts:SetSourceIdentity` ao perfil de instância, como mostrado a seguir.

```
{ // PropagateSourceIdentity statement
    "Sid":"PropagateSourceIdentity",
    "Effect":"Allow",
    "Action":"sts:SetSourceIdentity",
    "Resource":[
        <runtime-role-ARN>
    ],
    "Condition":{
        "StringEquals":{
            "sts:SourceIdentity":<source-identity>
        }
    }
}
```

Também é necessário adicionar a instrução `AllowSetSourceIdentity` à política de confiança de seus perfis de runtime.

```
{ // AllowSetSourceIdentity statement
```

```
    "Sid":"AllowSetSourceIdentity",
    "Effect":"Allow",
    "Principal":{
        "AWS":"arn:aws:iam::<AWS_ACCOUNT_ID>:role/EMR_EC2_DefaultRole"
    },
    "Action":[
        "sts:SetSourceIdentity",
        "sts:AssumeRole"
    ],
    "Condition":{
        "StringEquals":{
            "sts:SourceIdentity":<source-identity>
        }
    }
}
```

Considerações adicionais

> **Note**
>
> Com o lançamento do Amazon EMR`emr-6.9.0`, você pode enfrentar falhas intermitentes ao se conectar aos clusters do Amazon EMR a partir do Studio. SageMaker Para resolver esse problema, instale o patch com uma ação de bootstrap ao iniciar o cluster. Para obter detalhes sobre o patch, consulte Amazon EMR release 6.9.0 known issues.

Além disso, considere as informações a seguir ao configurar perfis de runtime para o Amazon EMR.

- O Amazon EMR oferece suporte a perfis de runtime em todas as Regiões da AWS comerciais.
- As etapas do Amazon EMR oferecem suporte a trabalhos do Apache Spark e do Apache Hive com perfis de runtime quando você usa a versão `emr-6.7.0` ou posteriores.
- SageMaker O Studio oferece suporte a consultas Spark, Hive e Presto com funções de tempo de execução quando você usa a versão ou posterior. `emr-6.9.0`
- Os seguintes kernels de notebook SageMaker oferecem suporte a funções de tempo de execução:
  - DataScience — Kernel Python 3
  - DataScience 2.0 — Kernel do Python 3
  - DataScience 3.0 — Kernel do Python 3
  - SparkAnalytics 1.0 — SparkMagic e PySpark grãos

  - SparkAnalytics 2.0 — SparkMagic e PySpark grãos
  - SparkMagic — PySpark núcleo
- O Amazon EMR oferece suporte a etapas que usam `RunJobFlow` somente no momento da criação do cluster. Essa API não é compatível com perfis de runtime.
- O Amazon EMR não oferece suporte a perfis de runtime em clusters configurados para alta disponibilidade.
- Você deve escapar dos argumentos do comando Bash ao executar comandos com o arquivo `command-runner.jar` JAR:

```
aws emr add-steps --cluster-id <cluster-id> --steps '[{"Name":"sample-
step","ActionOnFailure":"CONTINUE","Jar":"command-runner.jar","Properties":"","Args":
["bash","-c","\"aws s3 ls\""],"Type":"CUSTOM_JAR"}]' --execution-role-
arn <IAM_ROLE_ARN>
```

- Os perfis de runtime não oferecem suporte para controlar o acesso a recursos no cluster, como HDFS e HMS.

Configurar perfis de serviço do IAM para permissões do Amazon EMR aos serviços e recursos da AWS

O Amazon EMR e aplicações como o Hadoop e o Spark precisam de permissões para acessar outros recursos da AWS e realizar ações quando são executados. Cada cluster no Amazon EMR deve ter um perfil de serviço e um perfil para o perfil de instância do Amazon EC2. Para obter mais informações, consulte Perfis do IAM e Usar perfis de instância no Guia do usuário do IAM. As políticas do IAM anexadas a esses perfis fornecem permissões para o cluster interoperar com outros serviços da AWS em nome de um usuário.

Um perfil adicional, o perfil do Auto Scaling, será necessário se o cluster usar a ajuste de escala automático no Amazon EMR. A função AWS de serviço para Notebooks EMR é necessária se você usa Notebooks EMR.

O Amazon EMR fornece perfis padrão e políticas gerenciadas padrão para determinar permissões para cada perfil. As políticas gerenciadas são criadas e mantidas por AWS, portanto, são atualizadas automaticamente se os requisitos de serviço mudarem. Consulte AWS managed policies no Guia do usuário do IAM.

Se você estiver criando um cluster ou um caderno pela primeira vez em uma conta, os perfis do Amazon EMR ainda não existirão. Depois de criá-las, você pode visualizar os perfis, as políticas anexadas a eles e as permissões concedidas ou negadas pelas políticas no console do IAM (https://console.aws.amazon.com/iam/). Você pode especificar perfis padrão para o Amazon EMR criar e usar, pode criar seus próprios perfis e especificá-los individualmente ao criar um cluster para personalizar as permissões e pode, ainda, especificar perfis padrão a serem usados ao criar um cluster usando a AWS CLI. Para ter mais informações, consulte Personalizar perfis do IAM.

Modificar políticas baseadas em identidade para permissões a fim de transmitir perfis de serviço ao Amazon EMR

As políticas gerenciadas padrão de permissões completas do Amazon EMR incorporam configurações de segurança `iam:PassRole`, incluindo estas:

- Permissões `iam:PassRole` somente para perfis padrão específicos do Amazon EMR.
- `iam:PassedToService`condições que permitem que você use a política somente com AWS serviços específicos, como `elasticmapreduce.amazonaws.com` `ec2.amazonaws.com` e.

Você pode visualizar a versão JSON das políticas AmazonEMR FullAccess Policy_v2 e AmazonEMR _v2 no console do IAM. ServicePolicy É recomendável criar novos clusters com políticas gerenciadas v2.

Resumo do perfil de serviço

A tabela a seguir lista os perfis de serviço do IAM associadas ao Amazon EMR para referência rápida.

Função	Perfil padrão	Descrição	Política gerenciada padrão
Perfil de serviço para Amazon EMR (perfil do EMR)	`EMR_DefaultRole_V2`	Permite que o Amazon EMR chame outros AWS serviços em seu nome ao provisionar recursos e realizar ações em nível de serviço. Essa	`AmazonEMRServicePolicy_v2` **Important** É necessário ter um perfil

Função	Perfil padrão	Descrição	Política gerenciada padrão
		função é necessária para todos os clusters.	vinculado ao serviço para solicitar instâncias spot. Se esse perfil não existir, o perfil de serviço do Amazon EMR deve ter permissão para criá-lo ou ocorrerá um erro de permissão. Se você pretende solicitar instâncias spot, é necessário atualizar essa política para incluir uma instrução que permita a criação desse perfil vinculado ao serviço. Para obter mais informações, consulte Perfil de serviço para Amazon

Função	Perfil padrão	Descrição	Política gerenciada padrão
			EMR (perfil do EMR) a função vinculada ao serviço para solicitações de instâncias spot no Guia do usuário do Amazon EC2.

Função	Perfil padrão	Descrição	Política gerenciada padrão
Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2)	`EMR_EC2_DefaultRole`	Os processos de aplicativos executados no ecossistema do Hadoop em instâncias de cluster usam essa função quando chamam outros AWS serviços. Para acessar os dados no Amazon S3 usando o EMRFS, você pode especificar diferentes perfis a serem assumidos com base no local dos dados no Amazon S3. Por exemplo, múltiplas equipes podem acessar uma única “conta de armazenamento” de dados do Amazon S3. Para ter mais informações, consulte Configurar perfis do IAM para solicitações do EMRFS para o Amazon S3. Essa função é necessária para todos os clusters.	`AmazonElasticMapReduceforEC2Role` . Para obter mais informações, consulte Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2).

Função	Perfil padrão	Descrição	Política gerenciada padrão
Perfil de serviço para ajuste de escala automático no Amazon EMR (perfil do Auto Scaling)	`EMR_AutoScaling_DefaultRole`	Permite ações adicionais para ambientes de escalabilidade dinâmica. Necessária apenas para clusters que usam a ajuste de escala automático no Amazon EMR. Para ter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.	`AmazonElasticMapReduceforAutoScalingRole` . Para obter mais informações, consulte Perfil de serviço para ajuste de escala automático no Amazon EMR (perfil do Auto Scaling).

Função	Perfil padrão	Descrição	Política gerenciada padrão
Perfil de serviço para Cadernos do EMR	`EMR_Notebooks_DefaultRole`	Fornece as permissões que um notebook EMR precisa para acessar outros AWS recursos e realizar ações. Necessário somente se forem usados cadernos EMR.	`AmazonElasticMapReduceEditorsRole` . Para obter mais informações, consulte Perfil de serviço para Cadernos do EMR. `S3FullAccessPolicy` também é anexado por padrão. Veja a seguir o conteúdo da política. `{ "Version": "2012-10-17", "Statement": [ { "Effect": "Allow", "Action": "s3:*", "Resource": "*" } ] }`

Função	Perfil padrão	Descrição	Política gerenciada padrão
Função vinculada ao serviço	`AWSServiceRoleForEMRCleanup`	O Amazon EMR cria automaticamente um perfil vinculado ao serviço. Se o serviço do Amazon EMR tiver perdido a capacidade de apagar os recursos do Amazon EC2, o Amazon EMR poderá usar esse perfil para fazer isso. Se um cluster usar instâncias spot, a política de permissões anexada ao Perfil de serviço para Amazon EMR (perfil do EMR) deverá permitir a criação de uma função vinculada ao serviço. Para ter mais informações, consulte Usando funções vinculadas a serviços para o Amazon EMR.	`AmazonEMRCleanupPolicy`

Tópicos

- Perfis de serviço do IAM usados pelo Amazon EMR
- Personalizar perfis do IAM
- Configurar perfis do IAM para solicitações do EMRFS para o Amazon S3
- Usa políticas baseadas em recursos para acesso do Amazon EMR ao Catálogo de Dados do AWS Glue

- Usar perfis do IAM com aplicações que chamam diretamente os serviços da AWS
- Permitir que usuários e grupos criem e modifiquem perfis

Perfis de serviço do IAM usados pelo Amazon EMR

O Amazon EMR usa perfis de serviço do IAM para executar ações em seu nome ao provisionar recursos do cluster, executar aplicações, escalar recursos de forma dinâmica e criar e executar Cadernos do EMR. O Amazon EMR usa os seguintes perfis ao interagir com outros serviços da AWS . Cada perfil tem um papel exclusivo no Amazon EMR. Os tópicos desta seção descrevem o papel da função e fornecem as funções padrão e a política de permissões de cada função.

Se você tiver um código de aplicativo em seu cluster que chame AWS serviços diretamente, talvez seja necessário usar o SDK para especificar funções. Para ter mais informações, consulte Usar perfis do IAM com aplicações que chamam diretamente os serviços da AWS.

Tópicos

- Perfil de serviço para Amazon EMR (perfil do EMR)
- Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2)
- Perfil de serviço para ajuste de escala automático no Amazon EMR (perfil do Auto Scaling)
- Perfil de serviço para Cadernos do EMR
- Usando funções vinculadas a serviços para o Amazon EMR

Perfil de serviço para Amazon EMR (perfil do EMR)

O perfil do Amazon EMR define as ações permitidas para o Amazon EMR durante o provisionamento de recursos e a execução de tarefas no nível de serviço que não são executadas no contexto de uma instância do Amazon EC2 em execução em um cluster. Por exemplo, a função de serviço é usada para provisionar instâncias do EC2 quando um cluster é executado.

- O nome de perfil padrão é `EMR_DefaultRole_V2`.
- A política gerenciada com escopo do Amazon EMR padrão anexada a `EMR_DefaultRole_V2` é `AmazonEMRServicePolicy_v2`. Essa política v2 substitui a política gerenciada padrão defasada, `AmazonElasticMapReduceRole`.

`AmazonEMRServicePolicy_v2` depende do acesso de escopo limitado aos recursos que o Amazon EMR provisiona ou usa. Ao usar essa política, é necessário passar a etiqueta de usuário

`for-use-with-amazon-emr-managed-policies = true` ao provisionar o cluster. O Amazon EMR propagará essas etiquetas automaticamente. Além disso, talvez seja necessário adicionar manualmente uma etiqueta de usuário a tipos específicos de recursos, como grupos de segurança do EC2 que não foram criados pelo Amazon EMR. Consulte Etiquetar recursos para usar políticas gerenciadas.

> **Important**
>
> O Amazon EMR usa esse perfil de serviço do Amazon EMR e o perfil `AWSServiceRoleForEMRCleanup` para limpar recursos de cluster em sua conta que você não usa mais, como instâncias do Amazon EC2. Você deve incluir ações nas políticas de perfil para excluir ou encerrar os recursos. Caso contrário, o Amazon EMR não poderá realizar essas ações de limpeza e você poderá ter custos com recursos não utilizados que permanecem no cluster.

O exemplo a seguir mostra o conteúdo de uma política `AmazonEMRServicePolicy_v2` atual. Você também pode ver o conteúdo atual da política gerenciada `AmazonEMRServicePolicy_v2` no console do IAM.

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "CreateInTaggedNetwork",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateNetworkInterface",
    "ec2:RunInstances",
    "ec2:CreateFleet",
    "ec2:CreateLaunchTemplate",
    "ec2:CreateLaunchTemplateVersion"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:subnet/*",
    "arn:aws:ec2:*:*:security-group/*"
   ],
   "Condition": {
    "StringEquals": {
     "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
    }
```

```
    }
   },
   {
    "Sid": "CreateWithEMRTaggedLaunchTemplate",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateFleet",
     "ec2:RunInstances",
     "ec2:CreateLaunchTemplateVersion"
    ],
    "Resource": "arn:aws:ec2:*:*:launch-template/*",
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateEMRTaggedLaunchTemplate",
    "Effect": "Allow",
    "Action": "ec2:CreateLaunchTemplate",
    "Resource": "arn:aws:ec2:*:*:launch-template/*",
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateEMRTaggedInstancesAndVolumes",
    "Effect": "Allow",
    "Action": [
     "ec2:RunInstances",
     "ec2:CreateFleet"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:instance/*",
     "arn:aws:ec2:*:*:volume/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
```

```
  },
  {
   "Sid": "ResourcesToLaunchEC2",
   "Effect": "Allow",
   "Action": [
    "ec2:RunInstances",
    "ec2:CreateFleet",
    "ec2:CreateLaunchTemplate",
    "ec2:CreateLaunchTemplateVersion"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:network-interface/*",
    "arn:aws:ec2:*::image/ami-*",
    "arn:aws:ec2:*:*:key-pair/*",
    "arn:aws:ec2:*:*:capacity-reservation/*",
    "arn:aws:ec2:*:*:placement-group/pg-*",
    "arn:aws:ec2:*:*:fleet/*",
    "arn:aws:ec2:*:*:dedicated-host/*",
    "arn:aws:resource-groups:*:*:group/*"
   ]
  },
  {
   "Sid": "ManageEMRTaggedResources",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateLaunchTemplateVersion",
    "ec2:DeleteLaunchTemplate",
    "ec2:DeleteNetworkInterface",
    "ec2:ModifyInstanceAttribute",
    "ec2:TerminateInstances"
   ],
   "Resource": "*",
   "Condition": {
    "StringEquals": {
     "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "ManageTagsOnEMRTaggedResources",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateTags",
    "ec2:DeleteTags"
```

```
    ],
    "Resource": [
     "arn:aws:ec2:*:*:instance/*",
     "arn:aws:ec2:*:*:volume/*",
     "arn:aws:ec2:*:*:network-interface/*",
     "arn:aws:ec2:*:*:launch-template/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateNetworkInterfaceNeededForPrivateSubnet",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateNetworkInterface"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:network-interface/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "TagOnCreateTaggedEMRResources",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateTags"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:network-interface/*",
     "arn:aws:ec2:*:*:instance/*",
     "arn:aws:ec2:*:*:volume/*",
     "arn:aws:ec2:*:*:launch-template/*"
    ],
    "Condition": {
     "StringEquals": {
      "ec2:CreateAction": [
       "RunInstances",
```

```
      "CreateFleet",
      "CreateLaunchTemplate",
      "CreateNetworkInterface"
     ]
    }
   }
  },
  {
   "Sid": "TagPlacementGroups",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateTags",
    "ec2:DeleteTags"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:placement-group/pg-*"
   ]
  },
  {
   "Sid": "ListActionsForEC2Resources",
   "Effect": "Allow",
   "Action": [
    "ec2:DescribeAccountAttributes",
    "ec2:DescribeCapacityReservations",
    "ec2:DescribeDhcpOptions",
    "ec2:DescribeImages",
    "ec2:DescribeInstances",
    "ec2:DescribeInstanceTypeOfferings",
    "ec2:DescribeLaunchTemplates",
    "ec2:DescribeNetworkAcls",
    "ec2:DescribeNetworkInterfaces",
    "ec2:DescribePlacementGroups",
    "ec2:DescribeRouteTables",
    "ec2:DescribeSecurityGroups",
    "ec2:DescribeSubnets",
    "ec2:DescribeVolumes",
    "ec2:DescribeVolumeStatus",
    "ec2:DescribeVpcAttribute",
    "ec2:DescribeVpcEndpoints",
    "ec2:DescribeVpcs"
   ],
   "Resource": "*"
  },
  {
```

```
    "Sid": "CreateDefaultSecurityGroupWithEMRTags",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateSecurityGroup"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:security-group/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateDefaultSecurityGroupInVPCWithEMRTags",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateSecurityGroup"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:vpc/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "TagOnCreateDefaultSecurityGroupWithEMRTags",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateTags"
    ],
    "Resource": "arn:aws:ec2:*:*:security-group/*",
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true",
      "ec2:CreateAction": "CreateSecurityGroup"
     }
    }
   },
   {
```

```
    "Sid": "ManageSecurityGroups",
    "Effect": "Allow",
    "Action": [
     "ec2:AuthorizeSecurityGroupEgress",
     "ec2:AuthorizeSecurityGroupIngress",
     "ec2:RevokeSecurityGroupEgress",
     "ec2:RevokeSecurityGroupIngress"
    ],
    "Resource": "*",
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateEMRPlacementGroups",
    "Effect": "Allow",
    "Action": [
     "ec2:CreatePlacementGroup"
    ],
    "Resource": "arn:aws:ec2:*:*:placement-group/pg-*"
   },
   {
    "Sid": "DeletePlacementGroups",
    "Effect": "Allow",
    "Action": [
     "ec2:DeletePlacementGroup"
    ],
    "Resource": "*"
   },
   {
    "Sid": "AutoScaling",
    "Effect": "Allow",
    "Action": [
     "application-autoscaling:DeleteScalingPolicy",
     "application-autoscaling:DeregisterScalableTarget",
     "application-autoscaling:DescribeScalableTargets",
     "application-autoscaling:DescribeScalingPolicies",
     "application-autoscaling:PutScalingPolicy",
     "application-autoscaling:RegisterScalableTarget"
    ],
    "Resource": "*"
   },
```

```
  {
   "Sid": "ResourceGroupsForCapacityReservations",
   "Effect": "Allow",
   "Action": [
    "resource-groups:ListGroupResources"
   ],
   "Resource": "*"
  },
  {
   "Sid": "AutoScalingCloudWatch",
   "Effect": "Allow",
   "Action": [
    "cloudwatch:PutMetricAlarm",
    "cloudwatch:DeleteAlarms",
    "cloudwatch:DescribeAlarms"
   ],
   "Resource": "arn:aws:cloudwatch:*:*:alarm:*_EMR_Auto_Scaling"
  },
  {
   "Sid": "PassRoleForAutoScaling",
   "Effect": "Allow",
   "Action": "iam:PassRole",
   "Resource": "arn:aws:iam::*:role/EMR_AutoScaling_DefaultRole",
   "Condition": {
    "StringLike": {
     "iam:PassedToService": "application-autoscaling.amazonaws.com*"
    }
   }
  },
  {
   "Sid": "PassRoleForEC2",
   "Effect": "Allow",
   "Action": "iam:PassRole",
   "Resource": "arn:aws:iam::*:role/EMR_EC2_DefaultRole",
   "Condition": {
    "StringLike": {
     "iam:PassedToService": "ec2.amazonaws.com*"
    }
   }
  }
 ]
}
```

Seu perfil de serviço deve usar a seguinte política de confiança.

> **Important**
>
> A política de confiança a seguir inclui as chaves de condição globais `aws:SourceArn` e `aws:SourceAccount`, que limitam as permissões que você concede ao Amazon EMR para recursos específicos em sua conta. O uso delas pode proteger você contra o problema de "confused deputy".

```
{
    "Version": "2008-10-17",
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "elasticmapreduce.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
                }
            }
        }
    ]
}
```

Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2)

O perfil de serviço para instâncias do EC2 do cluster (também chamada de perfil de instância do EC2 para Amazon EMR) é um tipo especial de perfil de serviço atribuído a cada instância do EC2 no cluster do Amazon EMR quando a instância é iniciada. Os processos de aplicação que são executados no ecossistema do Hadoop assumem esse perfil para que as permissões interajam com outros serviços da AWS .

Para obter mais informações sobre perfis de serviço para instâncias do EC2, consulte Uso de um perfil do IAM para conceder permissões a aplicações em execução em instâncias do Amazon EC2 no Guia do usuário do IAM.

> **Important**
>
> A função de serviço padrão para instâncias EC2 de cluster e sua política gerenciada AWS padrão associada `AmazonElasticMapReduceforEC2Role` estão em vias de descontinuação, sem nenhuma política gerenciada AWS substituta fornecida. Será necessário criar e especificar um perfil de instância para substituir o perfil e a política padrão defasados.

Perfil padrão e política gerenciada

- O nome de perfil padrão é `EMR_EC2_DefaultRole`.
- A política gerenciada `EMR_EC2_DefaultRole` padrão, `AmazonElasticMapReduceforEC2Role`, está chegando ao fim do suporte. Em vez de usar uma política gerenciada padrão para o perfil de instância do EC2, aplique políticas baseadas em recursos aos buckets do S3 e outros recursos que o Amazon EMR precisa, ou use sua própria política gerenciada pelo cliente com um perfil do IAM como perfil de instância. Para ter mais informações, consulte Criar um perfil de serviço para instâncias do EC2 do cluster com permissões de privilégio mínimo.

Veja a seguir o conteúdo da versão 3 de `AmazonElasticMapReduceforEC2Role`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Resource": "*",
            "Action": [
                "cloudwatch:*",
                "dynamodb:*",
                "ec2:Describe*",
                "elasticmapreduce:Describe*",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
```

```
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSteps",
                "kinesis:CreateStream",
                "kinesis:DeleteStream",
                "kinesis:DescribeStream",
                "kinesis:GetRecords",
                "kinesis:GetShardIterator",
                "kinesis:MergeShards",
                "kinesis:PutRecord",
                "kinesis:SplitShard",
                "rds:Describe*",
                "s3:*",
                "sdb:*",
                "sns:*",
                "sqs:*",
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTable",
                "glue:GetTables",
                "glue:GetTableVersions",
                "glue:CreatePartition",
                "glue:BatchCreatePartition",
                "glue:UpdatePartition",
                "glue:DeletePartition",
                "glue:BatchDeletePartition",
                "glue:GetPartition",
                "glue:GetPartitions",
                "glue:BatchGetPartition",
                "glue:CreateUserDefinedFunction",
                "glue:UpdateUserDefinedFunction",
                "glue:DeleteUserDefinedFunction",
                "glue:GetUserDefinedFunction",
                "glue:GetUserDefinedFunctions"
            ]
        }
    ]
```

```
}
```

Seu perfil de serviço deve usar a seguinte política de confiança.

```
{
    "Version": "2008-10-17",
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "ec2.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

Criar um perfil de serviço para instâncias do EC2 do cluster com permissões de privilégio mínimo

Como prática recomendada, é altamente recomendável que você crie uma função de serviço para instâncias do EC2 de cluster e uma política de permissões que tenha as permissões mínimas para outros AWS serviços exigidos pelo seu aplicativo.

A política gerenciada padrão, `AmazonElasticMapReduceforEC2Role`, fornece permissões que facilitam a execução de um cluster inicial. No entanto, `AmazonElasticMapReduceforEC2Role` está em vias de descontinuação e o Amazon EMR não fornecerá uma política padrão AWS gerenciada substituta para a função obsoleta. Para iniciar um cluster inicial, é necessário fornecer uma política gerenciada pelo cliente baseada em recursos ou baseada em ID.

As instruções de política a seguir fornecem exemplos das permissões necessárias para diferentes atributos do Amazon EMR. Recomendamos que você use essas permissões para criar uma política de permissões que restrinja o acesso somente a esses recursos e aos recursos que o cluster exige. Todos os exemplos de declarações de política usam a *`us-west-2`*Região e o ID *`123456789012`*fictício AWS da conta. Faça as substituições apropriadas para o seu cluster.

Para obter mais informações sobre como criar e especificar funções personalizadas, consulte Personalizar perfis do IAM.

> **Note**
>
> Se você criar um perfil personalizado do EMR para o EC2, siga o fluxo de trabalho básico, que criará automaticamente um perfil de instância com o mesmo nome. O Amazon EC2 permite criar perfis e perfis de instância com nomes diferentes, mas o Amazon EMR não oferece suporte a essa configuração, o que resulta em um erro de "perfil de instância inválido" quando você cria o cluster.

Ler e gravar dados no Amazon S3 usando o EMRFS

Quando uma aplicação em execução em um cluster do Amazon EMR faz referência a dados usando o formato `s3://mydata`, o Amazon EMR usa o perfil de instância do EC2 para fazer a solicitação. Normalmente, os clusters leem e gravam dados no Amazon S3 dessa maneira, e o Amazon EMR usa as permissões anexadas ao perfil de serviço para instâncias do cluster do EC2 por padrão. Para ter mais informações, consulte Configurar perfis do IAM para solicitações do EMRFS para o Amazon S3.

Já que os perfis do IAM para o EMRFS se enquadrarão às permissões anexadas ao perfil de serviço para instâncias de cluster do EC2, como prática recomendada, é aconselhável usar perfis do IAM para o EMRFS e limitar as permissões do EMRFS e do Amazon S3 anexadas ao perfil de serviço para instâncias de cluster do EC2.

O exemplo de instrução a seguir demonstra as permissões que o EMRFS exige para fazer solicitações ao Amazon S3.

- *my-data-bucket-in-s3-for-emrfs-reads-and-writes* especifica o bucket no Amazon S3 em que o cluster lê e grava dados e todas as subpastas usando */**. Adicione somente os buckets e as pastas que o aplicativo exige.
- A instrução de política que permite ações dynamodb será necessária somente se a visualização consistente do EMRFS estiver habilitada. *EmrFSMetadata* especifica a pasta padrão para a visualização consistente do EMRFS.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
```

```
            "Action": [
                "s3:AbortMultipartUpload",
                "s3:CreateBucket",
                "s3:DeleteObject",
                "s3:GetBucketVersioning",
                "s3:GetObject",
                "s3:GetObjectTagging",
                "s3:GetObjectVersion",
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads",
                "s3:ListBucketVersions",
                "s3:ListMultipartUploadParts",
                "s3:PutBucketVersioning",
                "s3:PutObject",
                "s3:PutObjectTagging"
            ],
            "Resource": [
                "arn:aws:s3:::my-data-bucket-in-s3-for-emrfs-reads-and-writes",
                "arn:aws:s3:::my-data-bucket-in-s3-for-emrfs-reads-and-writes/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "dynamodb:CreateTable",
                "dynamodb:BatchGetItem",
                "dynamodb:BatchWriteItem",
                "dynamodb:PutItem",
                "dynamodb:DescribeTable",
                "dynamodb:DeleteItem",
                "dynamodb:GetItem",
                "dynamodb:Scan",
                "dynamodb:Query",
                "dynamodb:UpdateItem",
                "dynamodb:DeleteTable",
                "dynamodb:UpdateTable"
            ],
            "Resource": "arn:aws:dynamodb:us-west-2:123456789012:table/EmrFSMetadata"
        },
        {
            "Effect": "Allow",
            "Action": [
                "cloudwatch:PutMetricData",
                "dynamodb:ListTables",
```

```
                "s3:ListBucket"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "sqs:GetQueueUrl",
                "sqs:ReceiveMessage",
                "sqs:DeleteQueue",
                "sqs:SendMessage",
                "sqs:CreateQueue"
            ],
            "Resource": "arn:aws:sqs:us-west-2:123456789012:EMRFS-Inconsistency-*"
        }
    ]
}
```

Arquivar arquivos de log no Amazon S3

A instrução de política a seguir permite que o cluster do Amazon EMR arquive os arquivos de log no local especificada do Amazon S3. No exemplo abaixo, quando o cluster foi criado, *s3://MyLoggingBucket/MyEMRClusterLogs*foi especificado usando a localização da pasta Log S3 no console, usando a `--log-uri` opção do AWS CLI, ou usando o `LogUri` parâmetro no `RunJobFlow` comando. Para ter mais informações, consulte Arquivamento dos arquivos de log no Amazon S3.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:PutObject",
            "Resource": "arn:aws:s3:::MyLoggingBucket/MyEMRClusterLogs/*"
        }
    ]
}
```

Usar as ferramentas de depuração

A instrução de política a seguir permite ações que são necessárias caso você habilite a ferramenta de depuração do Amazon EMR. O arquivamento de arquivos de log no Amazon S3 e as permissões

associadas exibidas no exemplo acima são necessários para a depuração. Para ter mais informações, consulte Habilitar ferramenta de depuração.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "sqs:GetQueueUrl",
                "sqs:SendMessage"
            ],
            "Resource": "arn:aws:sqs:us-west-2:123456789012:AWS-ElasticMapReduce-*"
        }
    ]
}
```

Usando o AWS Glue Data Catalog

A declaração de política a seguir permite ações que são necessárias se você usar o AWS Glue Data Catalog como metastore para aplicativos. Para obter mais informações, consulte Usando o AWS Glue Data Catalog como metastore para o Spark SQL, Usando o AWS Glue Data Catalog como metastore para o Hive e Usando o Presto com o Glue Data AWS Catalog no Guia de lançamento do Amazon EMR.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTable",
                "glue:GetTables",
                "glue:GetTableVersions",
```

```
                "glue:CreatePartition",
                "glue:BatchCreatePartition",
                "glue:UpdatePartition",
                "glue:DeletePartition",
                "glue:BatchDeletePartition",
                "glue:GetPartition",
                "glue:GetPartitions",
                "glue:BatchGetPartition",
                "glue:CreateUserDefinedFunction",
                "glue:UpdateUserDefinedFunction",
                "glue:DeleteUserDefinedFunction",
                "glue:GetUserDefinedFunction",
                "glue:GetUserDefinedFunctions"
            ],
            "Resource": "*",
        }
    ]
}
```

Perfil de serviço para ajuste de escala automático no Amazon EMR (perfil do Auto Scaling)

O perfil do Auto Scaling para o Amazon EMR executa uma função semelhante ao perfil de serviço, mas permite outras ações para ambientes de ajusta de escala dinâmico.

- O nome de perfil padrão é `EMR_AutoScaling_DefaultRole`.
- A política gerenciada padrão anexada a `EMR_AutoScaling_DefaultRole` é `AmazonElasticMapReduceforAutoScalingRole`.

O conteúdo da versão 1 da `AmazonElasticMapReduceforAutoScalingRole` é mostrado a seguir.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "cloudwatch:DescribeAlarms",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
```

```
        }
    ]
}
```

Seu perfil de serviço deve usar a seguinte política de confiança.

> **Important**
>
> A política de confiança a seguir inclui as chaves de condição globais `aws:SourceArn` e `aws:SourceAccount`, que limitam as permissões que você concede ao Amazon EMR para recursos específicos em sua conta. O uso delas pode proteger você contra o problema de "confused deputy".

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "application-autoscaling.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:application-
autoscaling:<region>:<account-id>:scalable-target/*"
                }
            }
        }
    ]
}
```

Perfil de serviço para Cadernos do EMR

Cada notebook do EMR precisa de permissões para acessar outros AWS recursos e realizar ações. As políticas do IAM anexadas a essa função de serviço fornecem permissões para que o notebook interopere com outros AWS serviços. Ao criar um notebook usando o AWS

Management Console, você especifica uma função AWS de serviço. É possível usar a função padrão, `EMR_Notebooks_DefaultRole`, ou especificar uma função criada por você. Se um bloco de anotações não foi criado anteriormente, é possível optar por criar a função padrão.

- O nome de perfil padrão é `EMR_Notebooks_DefaultRole`.
- As políticas gerenciadas listadas anexadas a `EMR_Notebooks_DefaultRole` são `AmazonElasticMapReduceEditorsRole` e `S3FullAccessPolicy`.

Seu perfil de serviço deve usar a seguinte política de confiança.

> **Important**
>
> A política de confiança a seguir inclui as chaves de condição globais `aws:SourceArn` e `aws:SourceAccount`, que limitam as permissões que você concede ao Amazon EMR para recursos específicos em sua conta. O uso delas pode proteger você contra o problema de "confused deputy".

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "elasticmapreduce.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
                }
            }
        }
    ]
}
```

O conteúdo da versão 1 de `AmazonElasticMapReduceEditorsRole` é o seguinte.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:AuthorizeSecurityGroupEgress",
                "ec2:AuthorizeSecurityGroupIngress",
                "ec2:CreateSecurityGroup",
                "ec2:DescribeSecurityGroups",
                "ec2:RevokeSecurityGroupEgress",
                "ec2:CreateNetworkInterface",
                "ec2:CreateNetworkInterfacePermission",
                "ec2:DeleteNetworkInterface",
                "ec2:DeleteNetworkInterfacePermission",
                "ec2:DescribeNetworkInterfaces",
                "ec2:ModifyNetworkInterfaceAttribute",
                "ec2:DescribeTags",
                "ec2:DescribeInstances",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:ListSteps"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateTags",
            "Resource": "arn:aws:ec2:*:*:network-interface/*",
            "Condition": {
                "ForAllValues:StringEquals": {
                    "aws:TagKeys": [
                        "aws:elasticmapreduce:editor-id",
                        "aws:elasticmapreduce:job-flow-id"
                    ]
                }
            }
        }
    ]
}
```

Veja a seguir o conteúdo de `S3FullAccessPolicy`. `S3FullAccessPolicy` permite que seu perfil de serviço para cadernos EMR executem todas as ações do Amazon S3 em objetos em sua Conta da AWS. Ao criar um perfil de serviço personalizado para Cadernos do EMR, você deve conceder permissões do Amazon S3 ao perfil de serviço.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:*",
            "Resource": "*"
        }
    ]
}
```

Você pode restringir o acesso de leitura e gravação do perfil de serviço ao local do Amazon S3 em que deseja salvar os arquivos do caderno. Use o conjunto mínimo de permissões do Amazon S3 apresentado a seguir.

```
"s3:PutObject",
"s3:GetObject",
"s3:GetEncryptionConfiguration",
"s3:ListBucket",
"s3:DeleteObject"
```

Se o seu bucket do Amazon S3 estiver criptografado, você deverá incluir as seguintes permissões para o AWS Key Management Service.

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
"kms:DescribeKey"
```

Quando você vincula repositórios do Git ao caderno e precisa criar um segredo para o repositório, é necessário adicionar a permissão `secretsmanager:GetSecretValue` à política do IAM anexada ao perfil de serviço dos Cadernos do Amazon EMR. Um exemplo de política é demonstrado a seguir:

```
{
    "Version": "2012-10-17",
```

```
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": "secretsmanager:GetSecretValue",
            "Resource": "*"
        }
    ]
}
```

Permissões de perfil de serviço de Cadernos do EMR

Esta tabela lista as ações que os Cadernos do EMR executam usando o perfil de serviço, junto com as permissões necessárias para cada ação.

Ação	Permissões
Estabelecimento de um canal de rede seguro entre um caderno e um cluster do Amazon EMR e execução das ações de limpeza necessárias.	`"ec2:CreateNetworkInterface", "ec2:CreateNetworkInterfacePermission", "ec2:DeleteNetworkInterface", "ec2:DeleteNetworkInterfacePermission", "ec2:DescribeNetworkInterfaces", "ec2:ModifyNetworkInterfaceAttribute", "ec2:AuthorizeSecurityGroupEgress", "ec2:AuthorizeSecurityGroupIngress", "ec2:CreateSecurityGroup", "ec2:DescribeSecurityGroups", "ec2:RevokeSecurityGroupEgress", "ec2:DescribeTags", "ec2:DescribeInstances", "ec2:DescribeSubnets", "ec2:DescribeVpcs", "elasticmapreduce:ListInstances", "elasticmapreduce:DescribeCluster", "elasticmapreduce:ListSteps"`
Uso das credenciais do Git armazenadas no AWS Secrets Manager para vincular repositórios Git a um caderno.	`"secretsmanager:GetSecretValue"`

Ação	Permissões
Aplique AWS tags à interface de rede e aos grupos de segurança padrão que o EMR Notebooks cria ao configurar o canal de rede seguro. Para obter mais informaçõ es, consulte Etiquetar recursos da AWS.	`"ec2:CreateTags"`

Acesso ou upload de arquivos e metadados de cadernos para o Amazon S3.	`"s3:PutObject", "s3:GetObject", "s3:GetEncryptionConfiguration", "s3:ListBucket", "s3:DeleteObject"`
	
	As permissões a seguir serão necessárias somente se você usar um bucket criptografado do Amazon S3.
	
	`"kms:Decrypt", "kms:GenerateDataKey", "kms:ReEncryptFrom", "kms:ReEncryptTo", "kms:DescribeKey"`

Notebooks EMR: atualizações para políticas gerenciadas AWS

Veja detalhes sobre as atualizações das políticas AWS gerenciadas para Notebooks EMR desde 1º de março de 2021.

Alteração	Descrição	Data
`AmazonElasticMapReduceEditorsRole - Added permissions`	Cadernos do EMR foram adicionadas as permissões `ec2:describeVPCs` e `elastmicmapreduce: ListSteps` para	8 de fevereiro de 2023

Alteração	Descrição	Data
	`AmazonElasticMapReduceEditorsRole` .	
Os Cadernos do EMR começaram a monitorar alterações	O EMR Notebooks começou a monitorar as mudanças em suas políticas gerenciadas AWS .	8 de fevereiro de 2023

Usando funções vinculadas a serviços para o Amazon EMR

O Amazon EMR usa funções vinculadas a serviços AWS Identity and Access Management (IAM). Um perfil vinculado ao serviço é um tipo especial de perfil do IAM vinculado diretamente ao Amazon EMR. As funções vinculadas ao serviço são predefinidas pelo Amazon EMR e incluem todas as permissões que o serviço exige para chamar outros AWS serviços em seu nome.

Tópicos

- Usando funções vinculadas a serviços para limpeza
- Usando funções vinculadas ao serviço para registro antecipado

Para obter informações sobre outros serviços compatíveis com perfis vinculados ao serviço, consulte serviços da AWS que funcionam com o IAM e procure os serviços que apresentam Sim na coluna Perfis vinculados aos serviços. Escolha um Sim com um link para visualizar a documentação do perfil vinculado a esse serviço.

Usando funções vinculadas a serviços para limpeza

O Amazon EMR usa funções vinculadas a serviços AWS Identity and Access Management (IAM). Um perfil vinculado ao serviço é um tipo especial de perfil do IAM vinculado diretamente ao Amazon EMR. As funções vinculadas ao serviço são predefinidas pelo Amazon EMR e incluem todas as permissões que o serviço exige para chamar outros AWS serviços em seu nome.

As funções vinculadas ao serviço funcionam em conjunto com a função de serviço do Amazon EMR e o perfil de instância do Amazon EC2 para o Amazon EMR. Para obter mais informações sobre a função de serviço e o perfil da instância, consulte Configurar perfis de serviço do IAM para permissões do Amazon EMR aos serviços e recursos da AWS.

Uma função vinculada ao serviço facilita a configuração do Amazon EMR porque você não precisa adicionar manualmente as permissões necessárias. O Amazon EMR define as permissões de suas funções vinculadas a serviços e, a menos que seja definido de outra forma, somente o Amazon EMR pode assumir suas funções. As permissões definidas incluem a política de confiança e a política de permissões, que não pode ser anexada a nenhuma outra entidade do IAM.

Você pode excluir essa função vinculada ao serviço para o Amazon EMR somente depois de excluir quaisquer recursos relacionados e encerrar todos os clusters do EMR na conta. Isso protege seus recursos do Amazon EMR para que você não possa remover inadvertidamente a permissão para acessar os recursos.

Usando funções vinculadas a serviços para limpeza

O Amazon EMR usa a função baseada em serviços para conceder permissão AWSServiceRoleForEMRCleanupao Amazon EMR para encerrar e excluir recursos do Amazon EC2 em seu nome se a função vinculada ao serviço do Amazon EMR perder essa capacidade. O Amazon EMR cria a função vinculada ao serviço automaticamente durante a criação do cluster, caso ela ainda não exista.

A função AWSServiceRoleForEMRCleanup vinculada ao serviço confia nos seguintes serviços para assumir a função:

- `elasticmapreduce.amazonaws.com`

A política de permissões AWSServiceRoleForEMRCleanup de função vinculada ao serviço permite que o Amazon EMR conclua as seguintes ações nos recursos especificados:

- Ação: `DescribeInstances` em `ec2`
- Ação: `DescribeSpotInstanceRequests` em `ec2`
- Ação: `ModifyInstanceAttribute` em `ec2`
- Ação: `TerminateInstances` em `ec2`
- Ação: `CancelSpotInstanceRequests` em `ec2`
- Ação: `DeleteNetworkInterface` em `ec2`
- Ação: `DescribeInstanceAttribute` em `ec2`
- Ação: `DescribeVolumeStatus` em `ec2`
- Ação: `DescribeVolumes` em `ec2`
- Ação: `DetachVolume` em `ec2`

- Ação: `DeleteVolume` em `ec2`

Você deve configurar permissões para que uma entidade do IAM (por exemplo, um usuário, grupo ou função) crie, edite ou exclua uma função vinculada a serviço.

Criar um perfil vinculado ao serviço para Amazon EMR

Você não precisa criar a AWSServiceRoleForEMRCleanup função manualmente. Quando você inicia um cluster, seja pela primeira vez ou quando a função AWSServiceRoleForEMRCleanup vinculada ao serviço não está presente, o Amazon EMR cria a função vinculada ao AWSServiceRoleForEMRCleanup serviço para você. Você deve ter permissões para criar uma função vinculada ao serviço. Para obter um exemplo de instrução que acrescenta essa possibilidade à política de permissões de uma entidade do IAM (por exemplo, um usuário, grupo ou perfil), consulte Usando funções vinculadas a serviços para limpeza.

> **Important**
>
> Se você usou o Amazon EMR antes de 24 de outubro de 2017, quando as funções vinculadas ao serviço não eram suportadas, o Amazon EMR criou a função vinculada ao AWSServiceRoleForEMRCleanup serviço em sua conta. Para obter mais informações, consulte Uma nova função apareceu na minha conta do IAM.

Editar um perfil vinculado ao serviço do Amazon EMR

O Amazon EMR não permite que você edite a função vinculada ao AWSServiceRoleForEMRCleanup serviço. Depois de criar uma função vinculada ao serviço, você não pode alterar o nome da função vinculada ao serviço porque várias entidades podem fazer referência à função vinculada ao serviço. No entanto, você pode editar a descrição da função vinculada ao serviço usando o IAM.

Editar a descrição de uma função vinculada ao serviço (console do IAM)

Você pode usar o console do IAM para editar a descrição de uma função vinculada ao serviço.

Para editar a descrição de uma função vinculada ao serviço (console)

1. No painel de navegação do console do IAM, escolha Perfis.
2. Escolha o nome da função a ser modificada.
3. À direita de Descrição do perfil, escolha Editar.

4. Insira uma nova descrição na caixa e escolha Save changes (Salvar alterações).

Editar descrição de um perfil vinculado ao serviço (CLI do IAM)

Você pode usar os comandos do IAM do AWS Command Line Interface para editar a descrição de uma função vinculada ao serviço.

Para alterar a descrição de uma função (CLI)

1. (Opcional) Para visualizar a descrição atual de a uma função, use um dos comandos a seguir:

   ```
   $ aws iam get-role --role-name role-name
   ```

   Use o nome da função, não o nome de recurso da Amazon (ARN), para fazer referência às funções com os comandos da CLI. Por exemplo, se uma função tiver o seguinte nome de recurso da Amazon (ARN): `arn:aws:iam::123456789012:role/myrole`, você fará referência à função como **`myrole`**.
2. Para atualizar a descrição de uma função vinculada a serviço, use um dos comandos a seguir:

   ```
   $ aws iam update-role-description --role-name role-name --description description
   ```

Editar a descrição de uma função vinculada ao serviço (API do IAM)

Você pode usar a API do IAM para editar a descrição de uma função vinculada ao serviço.

Para alterar a descrição de uma função (API)

1. (Opcional) Para visualizar a descrição atual de uma função, use o comando a seguir:

   API do IAM: GetRole
2. Para atualizar a descrição de uma função, use o comando a seguir:

   API IAM: UpdateRoledescrição

Excluir um perfil vinculado ao serviço do Amazon EMR

Se você não precisar mais usar um recurso ou serviço que exija uma função vinculada ao serviço, recomendamos que você exclua essa função vinculada ao serviço. Dessa forma, você não terá uma

entidade não utilizada que não seja monitorada ativamente ou mantida. No entanto, você deve limpar sua função vinculada ao serviço antes de excluí-la.

Limpar um perfil vinculado ao serviço

Antes de usar o IAM para excluir uma função vinculada ao serviço, você deve primeiro confirmar se a função vinculada ao serviço não tem sessões ativas e remover os recursos usados pela função vinculada ao serviço.

Para verificar se a função vinculada ao serviço tem uma sessão ativa no console do IAM

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. No painel de navegação, escolha Perfis. Selecione o nome (não a caixa de seleção) da função AWSServiceRoleForEMRCleanup vinculada ao serviço.
3. Na página Resumo da função vinculada ao serviço selecionada, escolha Access Advisor.
4. Na guia Consultor de acesso, revise a atividade recente para a função vinculada ao serviço.

> **Note**
>
> Se você não tiver certeza se o Amazon EMR está usando AWSServiceRoleForEMRCleanup a função vinculada ao serviço, você pode tentar excluir a função vinculada ao serviço. Se o serviço estiver usando a função vinculada ao serviço, a exclusão falhará e você poderá visualizar as regiões em que a função vinculada ao serviço está sendo usada. Se a função vinculada ao serviço estiver sendo usada, você deverá aguardar o término da sessão antes de excluir a função vinculada ao serviço. Não é possível revogar a sessão de uma função vinculada a um serviço.

Para remover os recursos do Amazon EMR usados pelo AWSServiceRoleForEMRCleanup

- Encerre todos os clusters em sua conta. Para ter mais informações, consulte Terminar um cluster.

Excluir um perfil vinculado ao serviço (console do IAM)

É possível usar o console do IAM para excluir uma função vinculada ao serviço.

Para excluir uma função vinculada ao serviço (console)

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. No painel de navegação, escolha Perfis. Marque a caixa de seleção ao lado AWSServiceRoleForEMRCleanup, não o nome ou a linha em si.
3. Em ações de Função na parte superior da página, escolha a função Excluir.
4. Na caixa de diálogo de confirmação, revise os últimos dados acessados do serviço, que mostram quando cada uma das funções selecionadas acessou um AWS serviço pela última vez. Isso ajuda a confirmar se a função está ativa no momento. Para prosseguir, selecione Yes, Delete.
5. Monitore as notificações do console do IAM para progresso da exclusão da função vinculada ao serviço. Como a exclusão da função vinculada ao serviço do IAM é assíncrona, depois de enviar a função vinculada ao serviço para exclusão, a tarefa de exclusão pode ser bem-sucedida ou falhar. Se a tarefa obtiver êxito, você poderá escolher Visualizar Detalhes ou Visualizar Recursos a partir das notificações para saber por que a exclusão falhou. Se houve falha na exclusão porque há recursos no serviço que estão sendo usados pela função, o motivo da falha incluirá uma lista de recursos.

Excluir um perfil vinculado ao serviço (CLI do IAM)

Você pode usar os comandos do IAM do AWS Command Line Interface para excluir uma função vinculada ao serviço. Como uma função vinculada ao serviço não podem ser excluída se estiver sendo usada ou tiver recursos associados, você deverá enviar uma solicitação de exclusão. Se essas condições não forem atendidas, essa solicitação poderá ser negada.

Para excluir uma função vinculado ao serviço (CLI)

1. Para verificar o status da tarefa de exclusão, você deve capturar o `deletion-task-id` da resposta. Digite o seguinte comando para enviar uma solicitação de exclusão de função vinculada ao serviço:

```
$ aws iam delete-service-linked-role --role-name AWSServiceRoleForEMRCleanup
```

2. Digite o seguinte comando para verificar o estado da tarefa de exclusão:

```
$ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
```

O status da tarefa de exclusão pode ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, ou `FAILED`. Se a exclusão falhar, a chamada informará o motivo de falha para que você possa solucionar o problema.

Excluir uma função vinculada ao serviço (API do IAM)

É possível usar a API do IAM para excluir uma função vinculada ao serviço. Como uma função vinculada ao serviço não podem ser excluída se estiver sendo usada ou tiver recursos associados, você deverá enviar uma solicitação de exclusão. Se essas condições não forem atendidas, essa solicitação poderá ser negada.

Para excluir uma função vinculada ao serviço (API)

1. Para enviar uma solicitação de exclusão para uma função vinculada ao serviço, ligue. DeleteServiceLinkedRole Na solicitação, especifique o nome da AWSServiceRoleForEMRCleanup função.

   Para verificar o status da tarefa de exclusão, você deve capturar o `DeletionTaskId` da resposta.
2. Para verificar o status da exclusão, ligue GetServiceLinkedRoleDeletionStatus. Na solicitação, especifique o `DeletionTaskId`.

   O status da tarefa de exclusão pode ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, ou `FAILED`. Se a exclusão falhar, a chamada informará o motivo de falha para que você possa solucionar o problema.

Regiões suportadas para AWSServiceRoleForEMRCleanup

O Amazon EMR oferece suporte ao uso da função AWSServiceRoleForEMRCleanup vinculada ao serviço nas seguintes regiões.

Nome da região	Identidade da região	Compatível com o Amazon EMR
Leste dos EUA (Norte da Virgínia)	us-east-1	Sim
Leste dos EUA (Ohio)	us-east-2	Sim

Nome da região	Identidade da região	Compatível com o Amazon EMR
Oeste dos EUA (Norte da Califórnia)	us-west-1	Sim
Oeste dos EUA (Oregon)	us-west-2	Sim
Ásia-Pacífico (Mumbai)	ap-south-1	Sim
Asia Pacific (Osaka)	ap-northeast-3	Sim
Ásia-Pacífico (Seul)	ap-northeast-2	Sim
Ásia-Pacífico (Singapura)	ap-southeast-1	Sim
Ásia-Pacífico (Sydney)	ap-southeast-2	Sim
Ásia-Pacífico (Tóquio)	ap-northeast-1	Sim
Canadá (Central)	ca-central-1	Sim
Europa (Frankfurt)	eu-central-1	Sim
Europa (Irlanda)	eu-west-1	Sim
Europa (Londres)	eu-west-2	Sim
Europa (Paris)	eu-west-3	Sim
América do Sul (São Paulo)	sa-east-1	Sim

Usando funções vinculadas ao serviço para registro antecipado

O Amazon EMR usa funções vinculadas a serviços AWS Identity and Access Management (IAM). Um perfil vinculado ao serviço é um tipo especial de perfil do IAM vinculado diretamente ao Amazon EMR. As funções vinculadas ao serviço são predefinidas pelo Amazon EMR e incluem todas as permissões que o serviço exige para chamar outros AWS serviços em seu nome.

As funções vinculadas ao serviço funcionam em conjunto com a função de serviço do Amazon EMR e o perfil de instância do Amazon EC2 para o Amazon EMR. Para obter mais informações

sobre a função de serviço e o perfil da instância, consulte Configurar perfis de serviço do IAM para permissões do Amazon EMR aos serviços e recursos da AWS.

Uma função vinculada ao serviço facilita a configuração do Amazon EMR porque você não precisa adicionar manualmente as permissões necessárias. O Amazon EMR define as permissões de suas funções vinculadas a serviços e, a menos que seja definido de outra forma, somente o Amazon EMR pode assumir suas funções. As permissões definidas incluem a política de confiança e a política de permissões, que não pode ser anexada a nenhuma outra entidade do IAM.

Você pode excluir essa função vinculada ao serviço do Amazon EMR somente depois de excluir os recursos relacionados e encerrar todos os clusters do EMR na conta. Isso protege seus recursos do Amazon EMR para que você não possa remover inadvertidamente a permissão para acessar os recursos.

Permissões de função vinculadas ao serviço para registro antecipado (WAL)

O Amazon EMR usa a função vinculada ao serviço AWSServiceRoleForEMRWALpara recuperar o status de um cluster.

A função AWSServiceRoleForEMRWAL vinculada ao serviço confia nos seguintes serviços para assumir a função:

- `emrwal.amazonaws.com`

A política de `EMRDescribeClusterPolicyForEMRWAL`permissões para a função vinculada ao serviço permite que o Amazon EMR conclua as seguintes ações nos recursos especificados:

- Ação: `DescribeCluster` em `*`

Você deve configurar permissões para permitir que uma entidade do IAM (nesse caso, Amazon EMR WAL) crie, edite ou exclua uma função vinculada ao serviço. Adicione as seguintes declarações conforme necessário à política de permissões do seu perfil de instância:

CreateServiceLinkedRole

Para permitir que uma entidade do IAM crie a função AWSServiceRoleForEMRWAL vinculada ao serviço

Adicione a seguinte declaração à política de permissões da entidade do IAM que precisa criar a função vinculada ao serviço:

```
{
    "Effect": "Allow",
    "Action": [
        "iam:CreateServiceLinkedRole",
        "iam:PutRolePolicy"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/emrwal.amazonaws.com*/
AWSServiceRoleForEMRWAL*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "emrwal.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

UpdateRoleDescription

Para permitir que uma entidade do IAM edite a descrição da função AWSServiceRoleForEMRWAL vinculada ao serviço

Adicione a seguinte declaração à política de permissões da entidade do IAM que precisa editar a descrição de uma função vinculada ao serviço:

```
{
    "Effect": "Allow",
    "Action": [
        "iam:UpdateRoleDescription"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/emrwal.amazonaws.com*/
AWSServiceRoleForEMRWAL*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "emrwal.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

DeleteServiceLinkedRole

Para permitir que uma entidade do IAM exclua a função AWSServiceRoleForEMRWAL vinculada ao serviço

Adicione a seguinte declaração à política de permissões da entidade do IAM que precisa excluir uma função vinculada ao serviço:

```
{
    "Effect": "Allow",
    "Action": [
        "iam:DeleteServiceLinkedRole",
        "iam:GetServiceLinkedRoleDeletionStatus"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/elasticmapreduce.amazonaws.com*/
AWSServiceRoleForEMRCleanup*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "emrwal.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

Criar um perfil vinculado ao serviço para Amazon EMR

Você não precisa criar a AWSServiceRoleForEMRWAL função manualmente. O Amazon EMR cria essa função vinculada ao serviço automaticamente quando você cria um espaço de trabalho do WAL com a CLI do EMRWAL ou de, ou o HBase criará a função vinculada ao serviço quando você configurar um espaço AWS CloudFormation de trabalho para o Amazon EMR WAL e a função vinculada ao serviço ainda não existir. Você deve ter permissões para criar uma função vinculada ao serviço. Por exemplo, declarações que adicionam esse recurso à política de permissões de uma entidade do IAM (como um usuário, grupo ou função), consulte a seção anterior,Permissões de função vinculadas ao serviço para registro antecipado (WAL).

Editar um perfil vinculado ao serviço do Amazon EMR

O Amazon EMR não permite que você edite a função vinculada ao AWSServiceRoleForEMRWAL serviço. Depois de criar uma função vinculada ao serviço, você não pode alterar o nome da função

vinculada ao serviço porque várias entidades podem fazer referência à função vinculada ao serviço. No entanto, você pode editar a descrição da função vinculada ao serviço usando o IAM.

Editar a descrição de uma função vinculada ao serviço (console do IAM)

Você pode usar o console do IAM para editar a descrição de uma função vinculada ao serviço.

Para editar a descrição de uma função vinculada ao serviço (console)

1. No painel de navegação do console do IAM, escolha Perfis.
2. Escolha o nome da função a ser modificada.
3. À direita de Descrição do perfil, escolha Editar.
4. Insira uma nova descrição na caixa e escolha Save changes (Salvar alterações).

Editar descrição de um perfil vinculado ao serviço (CLI do IAM)

Você pode usar os comandos do IAM do AWS Command Line Interface para editar a descrição de uma função vinculada ao serviço.

Para alterar a descrição de uma função (CLI)

1. (Opcional) Para visualizar a descrição atual de a uma função, use um dos comandos a seguir:

   ```
   $ aws iam get-role --role-name role-name
   ```

   Use o nome da função, não o nome de recurso da Amazon (ARN), para fazer referência às funções com os comandos da CLI. Por exemplo, se uma função tiver o seguinte nome de recurso da Amazon (ARN): `arn:aws:iam::123456789012:role/myrole`, você fará referência à função como **`myrole`**.
2. Para atualizar a descrição de uma função vinculada a serviço, use um dos comandos a seguir:

   ```
   $ aws iam update-role-description --role-name role-name --description description
   ```

Editar a descrição de uma função vinculada ao serviço (API do IAM)

Você pode usar a API do IAM para editar a descrição de uma função vinculada ao serviço.

Para alterar a descrição de uma função (API)

1. (Opcional) Para visualizar a descrição atual de uma função, use o comando a seguir:

   API do IAM: GetRole
2. Para atualizar a descrição de uma função, use o comando a seguir:

   API IAM: UpdateRoledescrição

Excluir um perfil vinculado ao serviço do Amazon EMR

Se você não precisar mais usar um recurso ou serviço que exija uma função vinculada ao serviço, recomendamos que você exclua essa função vinculada ao serviço. Dessa forma, você não terá uma entidade não utilizada que não seja monitorada ativamente ou mantida. No entanto, você deve limpar sua função vinculada ao serviço antes de excluí-la.

> **Note**
>
> A operação de registro antecipado de gravação não será afetada se você excluir a AWSServiceRoleForEMRWAL função, mas o Amazon EMR não excluirá automaticamente os registros criados quando o cluster do EMR for encerrado. Portanto, você precisará excluir manualmente os registros WAL do Amazon EMR se excluir a função vinculada ao serviço.

Limpar uma função vinculada ao serviço

Antes de você poder usar o IAM para excluir uma função vinculada ao serviço, você deve primeiro confirmar que a função não tem sessões ativas e remover quaisquer recursos usados pela função.

Para verificar se a função vinculada ao serviço tem uma sessão ativa no console do IAM

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. No painel de navegação, escolha Perfis. Selecione o nome (não a caixa de seleção) da AWSServiceRoleForEMRWAL função.
3. Na página Summary (Resumo) da função selecionada, selecione a guia Access Advisor (Consultor de acesso).
4. Na guia Consultor de acesso, revise a atividade recente para a função vinculada ao serviço.

> **Note**
>
> Se você não tiver certeza se o Amazon EMR está usando AWSServiceRoleForEMRWAL a função, você pode tentar excluir a função vinculada ao serviço. Se o serviço estiver usando a função, a exclusão falhará e você poderá visualizar as regiões em que a função vinculada ao serviço está sendo usada. Se a função vinculada ao serviço estiver sendo usada, você deverá aguardar o término da sessão antes de excluir a função vinculada ao serviço. Não é possível revogar a sessão de uma função vinculada a um serviço.

Para remover os recursos do Amazon EMR usados pelo AWSServiceRoleForEMRWAL

- Encerre todos os clusters em sua conta. Para ter mais informações, consulte Terminar um cluster.

Excluir um perfil vinculado ao serviço (console do IAM)

É possível usar o console do IAM para excluir uma função vinculada ao serviço.

Para excluir uma função vinculada ao serviço (console)

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. No painel de navegação, escolha Perfis. Marque a caixa de seleção ao lado AWSServiceRoleForEMRWAL, não o nome ou a linha em si.
3. Em ações de Função na parte superior da página, escolha a função Excluir.
4. Na caixa de diálogo de confirmação, revise os últimos dados acessados do serviço, que mostram quando cada uma das funções selecionadas acessou um AWS serviço pela última vez. Isso ajuda a confirmar se a função está ativa no momento. Para prosseguir, selecione Yes, Delete.
5. Monitore as notificações do console do IAM para progresso da exclusão da função vinculada ao serviço. Como a exclusão da função vinculada ao serviço do IAM é assíncrona, depois de enviar a função para exclusão, a tarefa pode ou não ser bem-sucedida. Se a tarefa obtiver êxito, você poderá escolher Visualizar Detalhes ou Visualizar Recursos a partir das notificações para saber por que a exclusão falhou. Se houve falha na exclusão porque há recursos no serviço que estão sendo usados pela função, o motivo da falha incluirá uma lista de recursos.

Excluir um perfil vinculado ao serviço (CLI do IAM)

Você pode usar os comandos do IAM do AWS Command Line Interface para excluir uma função vinculada ao serviço. Como uma função vinculada ao serviço não podem ser excluída se estiver sendo usada ou tiver recursos associados, você deverá enviar uma solicitação de exclusão. Se essas condições não forem atendidas, essa solicitação poderá ser negada.

Para excluir uma função vinculado ao serviço (CLI)

1. Para verificar o status da tarefa de exclusão, você deve capturar o `deletion-task-id` da resposta. Digite o seguinte comando para enviar uma solicitação de exclusão de função vinculada ao serviço:

   ```
   $ aws iam delete-service-linked-role --role-name AWSServiceRoleForEMRWAL
   ```

2. Digite o seguinte comando para verificar o estado da tarefa de exclusão:

   ```
   $ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
   ```

   O status da tarefa de exclusão pode ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, ou `FAILED`. Se a exclusão falhar, a chamada informará o motivo de falha para que você possa solucionar o problema.

Excluir uma função vinculada ao serviço (API do IAM)

É possível usar a API do IAM para excluir uma função vinculada ao serviço. Como uma função vinculada ao serviço não podem ser excluída se estiver sendo usada ou tiver recursos associados, você deverá enviar uma solicitação de exclusão. Se essas condições não forem atendidas, essa solicitação poderá ser negada.

Para excluir uma função vinculada ao serviço (API)

1. Para enviar uma solicitação de exclusão para uma função vinculada ao serviço, ligue. DeleteServiceLinkedRole Na solicitação, especifique o nome da AWSServiceRoleForEMRWAL função.

   Para verificar o status da tarefa de exclusão, você deve capturar o `DeletionTaskId` da resposta.

2. Para verificar o status da exclusão, ligue GetServiceLinkedRoleDeletionStatus. Na solicitação, especifique o `DeletionTaskId`.

   O status da tarefa de exclusão pode ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED`, ou `FAILED`. Se a exclusão falhar, a chamada informará o motivo de falha para que você possa solucionar o problema.

Regiões suportadas para AWSServiceRoleForEMRWAL

O Amazon EMR oferece suporte ao uso da função AWSServiceRoleForEMRWAL vinculada ao serviço nas seguintes regiões.

Nome da região	Identidade da região	Compatível com o Amazon EMR
Leste dos EUA (Norte da Virgínia)	us-east-1	Sim
Leste dos EUA (Ohio)	us-east-2	Sim
Oeste dos EUA (Norte da Califórnia)	us-west-1	Sim
Oeste dos EUA (Oregon)	us-west-2	Sim
Ásia-Pacífico (Mumbai)	ap-south-1	Sim
Ásia-Pacífico (Singapura)	ap-southeast-1	Sim
Ásia-Pacífico (Sydney)	ap-southeast-2	Sim
Ásia-Pacífico (Tóquio)	ap-northeast-1	Sim
Europa (Frankfurt)	eu-central-1	Sim
Europa (Irlanda)	eu-west-1	Sim

Personalizar perfis do IAM

Talvez seja útil personalizar as permissões e os perfis de serviço do IAM para limitar os privilégios de acordo com seus requisitos de segurança. Para personalizar as permissões,

recomendamos que você crie novas funções e políticas. Comece com as permissões nas políticas gerenciadas para as funções padrão (por exemplo, `AmazonElasticMapReduceforEC2Role` e `AmazonElasticMapReduceRole`). Em seguida, copie e cole o conteúdo em novas declarações de política, modifique as permissões conforme apropriado e anexe as políticas de permissões modificadas às funções que criar. Você deve ter as permissões apropriadas do IAM para trabalhar com perfis e políticas. Para ter mais informações, consulte Permitir que usuários e grupos criem e modifiquem perfis.

Se você criar um perfil personalizado do EMR para o EC2, siga o fluxo de trabalho básico, que criará automaticamente um perfil de instância com o mesmo nome. O Amazon EC2 permite criar perfis e perfis de instância com nomes diferentes, mas o Amazon EMR não oferece suporte a essa configuração, o que resulta em um erro de “perfil de instância inválido” quando você cria o cluster.

> **Important**
>
> As políticas em linha não são atualizadas automaticamente quando os requisitos do serviço são alterados. Se você criar e anexar políticas em linha, lembre-se de que podem ocorrer atualizações de serviço que causem erros de permissão repentinamente. Para obter mais informações, consulte Políticas gerenciadas e em linha no Guia do usuário do IAM e em Especificar perfis personalizados do IAM ao criar um cluster.

Para obter mais informações sobre como trabalhar com perfis do IAM, consulte os seguintes tópicos no Guia do usuário do IAM:

- Criação de uma função para delegar permissões a um serviço AWS
- Modificar uma função
- Excluir um perfil

Especificar perfis personalizados do IAM ao criar um cluster

Você especifica o perfil de serviço para o Amazon EMR e o perfil para o perfil de instância do Amazon EC2 ao criar um cluster. O usuário que está criando clusters precisa de permissões para recuperar e atribuir perfis a instâncias do EC2 e ao Amazon EMR. Caso contrário, ocorrerá um erro account is not authorized to call EC2. Para ter mais informações, consulte Permitir que usuários e grupos criem e modifiquem perfis.

Usar o console para especificar perfis personalizados

Ao criar um cluster, você pode especificar um perfil de serviço personalizado para o Amazon EMR, um perfil personalizado para o perfil de instância do EC2 e um perfil do Auto Scaling personalizado usando as Opções avançadas. Quando você usa Quick options (Opções avançadas), a função de serviço padrão e a função padrão para o perfil de instância do EC2 são especificadas. Para ter mais informações, consulte Perfis de serviço do IAM usados pelo Amazon EMR.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Especificar perfis do IAM personalizados usando o novo console

Ao criar um cluster usando o novo console, é necessário especificar um perfil de serviço personalizado para o Amazon EMR e um perfil personalizada para o perfil de instância do EC2. Para ter mais informações, consulte Perfis de serviço do IAM usados pelo Amazon EMR.

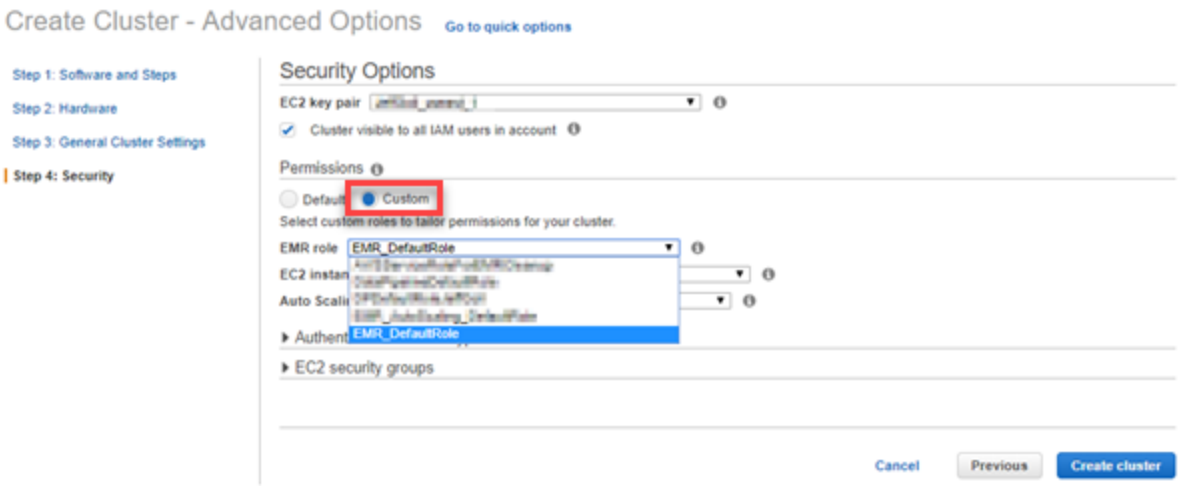
1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Configuração e permissões de segurança, encontre os campos Perfil do IAM para o perfil de instância e Perfil de serviço para o Amazon EMR. Para cada tipo de função, selecione uma função na lista. Apenas as funções em sua conta que têm a política de confiança apropriada para esse tipo de função são listadas.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Old console

Especificar perfis do IAM personalizados usando o console antigo

Ao criar um cluster usando o console antigo, você pode especificar um perfil de serviço personalizado para o Amazon EMR, um perfil personalizado para o perfil de instância do EC2 e um perfil do Auto Scaling personalizado usando as Opções avançadas. Quando você usa Quick options (Opções avançadas), a função de serviço padrão e a função padrão para o perfil de instância do EC2 são especificadas. Para ter mais informações, consulte Perfis de serviço do IAM usados pelo Amazon EMR.

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Escolha as definições adequadas do cluster para seu aplicativo até chegar a Security Options (Opções de segurança). Em Permissões, os perfis Padrão para o Amazon EMR são selecionadas.
4. Escolha Custom (Personalizado).
5. Para cada tipo de função, selecione uma função na lista. Apenas as funções em sua conta que têm a política de confiança apropriada para esse tipo de função são listadas.



6. Escolha outras opções conforme apropriado para seu cluster e, em seguida, escolha Create Cluster (Criar cluster).

Use o AWS CLI para especificar funções personalizadas

É possível especificar um perfil de serviço do Amazon EMR e um perfil de serviço para instâncias de cluster do EC2 usando explicitamente opções com o comando `create-cluster` na AWS CLI. Use a opção `--service-role` para especificar a função de serviço. Use o argumento `InstanceProfile` da opção `--ec2-attributes` para especificar a função para o perfil de instância do EC2.

O perfil do Auto Scaling é especificado usando uma opção separada, `--auto-scaling-role`. Para ter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.

Para especificar funções personalizadas do IAM usando o AWS CLI

- O comando a seguir especifica a função de serviço personalizada, *MyCustomServiceRoleForEMR*, e uma função personalizada para o perfil de instância do EC2, *MyCustomServiceRoleForClusterEC2Instances*, ao iniciar um cluster. Este exemplo usa o perfil padrão do Amazon EMR.

  > **Note**
  >
  > Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.1.0 \
--applications Name=Hive Name=Pig --service-role MyCustomServiceRoleForEMR \
--ec2-attributes InstanceProfile=MyCustomServiceRoleForClusterEC2Instances,\
KeyName=myKey --instance-type m5.xlarge --instance-count 3
```

Você pode usar essas opções para especificar funções padrão explicitamente em vez de usar a opção `--use-default-roles`. A opção `--use-default-roles` especifica a função de serviço e a função do perfil de instância do EC2 definidas no arquivo `config` para a AWS CLI.

O exemplo a seguir demonstra o conteúdo de um `config` arquivo para as funções personalizadas especificadas para AWS CLI o Amazon EMR. Com esse arquivo de configuração, quando a `--use-default-roles` opção é especificada, o cluster é criado usando *MyCustomServiceRoleForEMR*

e *MyCustomServiceRoleForClusterEC2Instances*. Por padrão, o arquivo `config` especifica a `service_role` padrão como `AmazonElasticMapReduceRole` e o `instance_profile` padrão como `EMR_EC2_DefaultRole`.

```
[default]
output = json
region = us-west-1
aws_access_key_id = myAccessKeyID
aws_secret_access_key = mySecretAccessKey
emr =
     service_role = MyCustomServiceRoleForEMR
     instance_profile = MyCustomServiceRoleForClusterEC2Instances
```

Configurar perfis do IAM para solicitações do EMRFS para o Amazon S3

> **Note**
>
> O recurso de mapeamento de perfis do EMRFS descrito nesta página foi aprimorado com a introdução da solução Amazon S3 Access Grants no Amazon EMR 6.15.0. Para uma solução de controle de acesso escalável para os seus dados no Amazon S3, recomendamos usar o S3 Access Grants com o Amazon EMR.

Quando uma aplicação em execução em um cluster faz referência a dados usando o formato `s3://mydata`, o Amazon EMR usa o EMRFS para fazer a solicitação. Para interagir com o Amazon S3, o EMRFS assume as políticas de permissões anexadas ao perfil de instância do Amazon EC2. O mesmo perfil de instância do Amazon EC2 é usado independentemente do usuário ou do grupo que está executando a aplicação ou do local dos dados no Amazon S3.

Se tiver um cluster com vários usuários que precisam de diferentes níveis de acesso aos dados no Amazon S3 por meio do EMRFS, você poderá definir uma configuração de segurança com perfis do IAM para o EMRFS. O EMRFS pode assumir um perfil de serviço diferente para instâncias de cluster do EC2 com base no usuário ou no grupo que faz a solicitação ou com base na localização dos dados no Amazon S3. Cada perfil do IAM para o EMRFS pode ter permissões diferentes para acesso aos dados no Amazon S3. Para obter mais informações sobre o uso de perfis de serviço para instâncias de cluster do EC2. consulte Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2).

Há suporte para uso de perfis personalizados do IAM para o EMRFS nas versões 5.10.0 e posteriores do Amazon EMR. Se usar uma versão anterior ou se tiver requisitos além dos fornecidos pelos perfis do IAM para EMRFS, você poderá criar um provedor de credenciais personalizadas. Para obter mais informações, consulte Authorizing access to EMRFS data in Amazon S3.

Ao usar uma configuração de segurança para especificar perfis do IAM para EMRFS, você configura mapeamentos de perfil. Cada mapeamento de perfil especifica um perfil do IAM que corresponde a identificadores. Esses identificadores determinam a base para o acesso ao Amazon S3 por meio do EMRFS. Os identificadores podem ser usuários, grupos ou prefixos do Amazon S3 que indicam um local de dados. Quando o EMRFS faz uma solicitação ao Amazon S3, se a solicitação corresponder à base para o acesso, o EMRFS fará com que as instâncias do EC2 do cluster assumam o perfil do IAM correspondente para a solicitação. As permissões do IAM anexadas ao perfil se aplicam no lugar das permissões do IAM anexadas ao perfil de serviço para instâncias do EC2 do cluster.

Os usuários e os grupos em um mapeamento de função são usuários e grupos do Hadoop definidos no cluster. Os usuários e os grupos são passados para o EMRFS no contexto do aplicativo que o usa (por exemplo, a personificação de usuário do YARN). O prefixo do Amazon S3 pode ser um especificador do bucket de qualquer profundidade (por exemplo `s3://mybucket` ou `s3://mybucket/myproject/mydata`). Você pode especificar vários identificadores em um único mapeamento de função, mas todos devem ser do mesmo tipo.

> **Important**
>
> Os perfis do IAM para o EMRFS fornecem isolamento no nível da aplicação entre os usuários da aplicação. Isso não fornece isolamento no nível do host entre os usuários no host. Qualquer usuário com acesso ao cluster pode ignorar o isolamento para assumir qualquer uma das funções.

Quando uma aplicação de cluster faz uma solicitação ao Amazon S3 por meio do EMRFS, o EMRFS avalia os mapeamentos do perfil de cima para baixo na ordem em que aparecem na configuração de segurança. Se uma solicitação feita por meio do EMRFS não corresponder a nenhum identificador, o EMRFS retornará ao uso do perfil de serviço para instâncias do EC2 do cluster. Por esse motivo, recomendamos que as políticas anexadas ao perfil limitem as permissões ao Amazon S3. Para ter mais informações, consulte Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2).

Configurar funções do

Antes de definir uma configuração de segurança com funções do IAM para EMRFS, planeje e crie perfis e as políticas de permissão a serem anexadas aos perfis. Para obter mais informações, consulte Como os perfis para as instâncias do Amazon EC2 funcionam? no Guia do usuário do IAM. Ao criar políticas de permissão, recomendamos começar com a política gerenciada anexada ao perfil do Amazon EMR padrão para o EC2 e depois editar essa política de acordo com seus requisitos. O nome de perfil padrão é `EMR_EC2_DefaultRole`, e a política gerenciada padrão a ser editada é `AmazonElasticMapReduceforEC2Role`. Para ter mais informações, consulte Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2).

Atualizar políticas de confiança para permissões para assumir perfil

Cada perfil que o EMRFS usa deve ter uma política de confiança que permite que o perfil do Amazon EMR para o EC2 do cluster o assuma. Da mesma forma, o perfil do Amazon EMR para o EC2 do cluster deve ter uma política de confiança que permita que os perfis do EMRFS o assumam.

O exemplo de política de confiança a seguir está anexado a funções para o EMRFS. A instrução permite que o perfil padrão do Amazon EMR para o EC2 assuma o perfil. Por exemplo, se você tiver duas funções do EMRFS fictícias `EMRFSRole_First` e `EMRFSRole_Second`, esta declaração de política será adicionada às políticas de confiança para cada uma delas.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "AWS":"arn:aws:iam::AWSAcctID:role/EMR_EC2_DefaultRole"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

Além disso, o exemplo a seguir de declaração de política de confiança é adicionado à `EMR_EC2_DefaultRole` para permitir que as duas funções do EMRFS fictícias a assumam.

```
{
   "Version":"2012-10-17",
   "Statement":[
```

```
      {
         "Effect":"Allow",
         "Principal":{
            "AWS": ["arn:aws:iam::AWSAcctID:role/EMRFSRole_First",
 "arn:aws:iam::AWSAcctID:role/EMRFSRole_Second"]
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

Atualizar a política de confiança de um perfil do IAM

Abra o console IAM em https://console.aws.amazon.com/iam/.

1. Selecione Roles (funções), insira o nome da função em Search (Pesquisar) e, em seguida, selecione o Role name (Nome da função).
2. Escolha Trust relationships (Relacionamentos de confiança), Edit trust relationship (Editar relacionamento de confiança).
3. Adicione uma instrução de confiança de acordo com o Documento da política, de acordo com as diretrizes acima e selecione Atualizar política de confiança.

Especificar um perfil como um usuário da chave

Se o perfil permitir acesso a um local no Amazon S3 que é criptografado usando uma AWS KMS key, especifique o perfil como um usuário de chaves. Isso concede permissões ao perfil para usar a chave do KMS. Para obter mais informações, consulte Políticas de chaves no AWS KMS no Guia do desenvolvedor do AWS Key Management Service .

Definir uma configuração de segurança com perfis do IAM para EMRFS

> **Important**
>
> Se nenhuma das perfis do IAM para EMRFS que você especificar for aplicável, o EMRFS retornará para o perfil do Amazon EMR para EC2. Considere a possibilidade de personalizar esse perfil para restringir permissões para o Amazon S3 conforme apropriado para suas aplicação e especificar esse perfil personalizado em vez de `EMR_EC2_DefaultRole` ao criar um cluster. Para obter mais informações, consulte Personalizar perfis do IAM e Especificar perfis personalizados do IAM ao criar um cluster.

Especificar perfis do IAM para solicitações do EMRFS para o Amazon S3 usando o console

1. Crie uma configuração de segurança que especifica os mapeamentos de função:

   a. No console do Amazon EMR, selecione Configurações de segurança, Criar.

   b. Digite um nome em Name (Nome) para a configuração de segurança. Esse nome é usado para especificar a configuração de segurança ao criar um cluster.

   c. Escolha Usa perfis do IAM para solicitações do EMRFS ao Amazon S3.

   d. Selecione um perfil do IAM a ser aplicado e, em Base para acesso, selecione um tipo de identificador (Usuários, Grupos ou Prefixos do S3) na lista e insira os identificadores correspondentes. Se você usar vários identificadores, separe-os com uma vírgula e sem espaço. Para obter mais informações sobre cada tipo de identificador, consulte a JSON configuration reference abaixo.

   e. Escolha Add role (Adicionar função) para configurar mapeamentos de funções adicionais, conforme descrito na etapa anterior.

   f. Configure outras opções de configuração de segurança conforme apropriado e escolha Create (Criar). Para ter mais informações, consulte Criar uma configuração de segurança.

2. Especifique a configuração de segurança criada acima ao criar um cluster. Para ter mais informações, consulte Especificar uma configuração de segurança para um cluster.

Para especificar funções do IAM para solicitações do EMRFS para o Amazon S3 usando o AWS CLI

1. Use o comando `aws emr create-security-configuration`, especificando um nome para a configuração de segurança e os detalhes da configuração de segurança no formato JSON.

   O comando de exemplo mostrado a seguir cria uma configuração de segurança com o nome `EMRFS_Roles_Security_Configuration`. Ele é baseado em uma estrutura JSON no arquivo `MyEmrfsSecConfig.json`, que é salvo no mesmo diretório onde o comando é executado.

   ```
   aws emr create-security-configuration --name EMRFS_Roles_Security_Configuration --
   security-configuration file://MyEmrFsSecConfig.json.
   ```

   Use as diretrizes a seguir para a estrutura do arquivo `MyEmrFsSecConfig.json`. Você pode especificar essa estrutura juntamente com estruturas de outras opções de configuração de segurança. Para ter mais informações, consulte Criar uma configuração de segurança.

Veja a seguir um exemplo de trecho JSON para especificar perfis do IAM personalizados para o EMRFS em uma configuração de segurança. Ele demonstra mapeamentos de perfil para os três tipos diferentes de identificadores, seguidos por uma referência de parâmetro.

```
{
  "AuthorizationConfiguration": {
    "EmrFsConfiguration": {
      "RoleMappings": [{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_user1",
        "IdentifierType": "User",
        "Identifiers": [ "user1" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_to_MyBuckets",
        "IdentifierType": "Prefix",
        "Identifiers": [ "s3://MyBucket/","s3://MyOtherBucket/" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_AdminGroup",
        "IdentifierType": "Group",
        "Identifiers": [ "AdminGroup" ]
      }]
    }
  }
}
```

Parâmetro	Descrição
`"AuthorizationConfiguration":`	Obrigatório.
`"EmrFsConfiguration":`	Obrigatório. Contém mapeamentos de perfil.

Parâmetro	Descrição
`"RoleMappings":`	Obrigatório. Contém uma ou mais definições de mapeamento de perfil. Os mapeamentos de perfil são avaliados na ordem em que aparecem, de cima para baixo. Se o mapeamento de perfil for avaliado como true para uma chamada do EMRFS para dados no Amazon S3, nenhum outro mapeamento de perfil será avaliado, e o EMRFS usará o perfil do IAM especificado para a solicitação. O mapeamento de perfil tem os seguintes parâmetros obrigatórios:
`"Role":`	Especifica o identificador ARN de um perfil do IAM no formato `arn:aws:iam::` *`account-id`* `:role/`*`role-name`*. Essa é o perfil do IAM que o Amazon EMR assume se a solicitação do EMRFS para o Amazon S3 corresponder a qualquer um dos `Identifiers` especificados.

Parâmetro	Descrição
`"IdentifierType":`	Pode ser um dos seguintes: • `"User"` especifica que os identificadores são um ou mais usuários do Hadoop, que podem ser usuários de contas Linux ou entidades principais do Kerberos. Quando a solicitação do EMRFS se origina com os usuários especificados, o perfil do IAM é assumido. • `"Prefix"` especifica que o identificador é um local do Amazon S3. O perfil do IAM é assumido para chamadas para os locais com os prefixos especificados. Por exemplo, o prefixo `s3://mybucket/` corresponde a `s3://mybucket/mydir` e `s3://mybucket/yetanotherdir` . • `"Group"` especifica que os identificadores são um ou mais grupos do Hadoop. O perfil do IAM será assumido se a solicitação for originada de um usuário dos grupos especificados.
`"Identifiers":`	Especifica um ou mais identificadores do tipo de identificador adequado. Separe múltiplos identificadores por vírgulas sem espaços.

2. Use o comando `aws emr create-cluster` para criar um cluster e especifique a configuração de segurança que você criou na etapa anterior.

   O exemplo a seguir cria um cluster com aplicativos Hadoop de núcleo padrão instalados. O cluster usa a configuração de segurança criada acima como `EMRFS_Roles_Security_Configuration` e também usa um perfil do Amazon EMR personalizado para o EC2, `EC2_Role_EMR_Restrict_S3`, que é especificada usando o argumento `InstanceProfile` do parâmetro `--ec2-attributes`.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name MyEmrFsS3RolesCluster \
--release-label emr-7.1.0 --ec2-attributes
 InstanceProfile=EC2_Role_EMR_Restrict_S3,KeyName=MyKey \
--instance-type m5.xlarge --instance-count 3 \
--security-configuration EMRFS_Roles_Security_Configuration
```

Usa políticas baseadas em recursos para acesso do Amazon EMR ao Catálogo de Dados do AWS Glue

Se você usa o AWS Glue em conjunto com o Hive, o Spark ou o Presto no Amazon EMR AWS , o Glue oferece suporte a políticas baseadas em recursos para controlar o acesso aos recursos do catálogo de dados. Esses recursos incluem bancos de dados, tabelas, conexões e funções definidas pelo usuário. Para obter mais informações, consulte Políticas baseadas em recursos no AWS Glue no Guia do desenvolvedor do AWS Glue.

Ao usar políticas baseadas em recursos para limitar o acesso ao AWS Glue a partir do Amazon EMR, o principal que você especifica na política de permissões deve ser o ARN da função associado ao perfil de instância do EC2 que é especificado quando um cluster é criado. Por exemplo, para uma política baseada em recursos anexada a um catálogo, você pode especificar o ARN da função para a função de serviço padrão para instâncias EC2 de cluster, *EMR_EC2_ DefaultRole* como o, usando o formato mostrado no exemplo a `Principal` seguir:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

O *acct-id* pode ser diferente do ID da conta AWS Glue. Isso possibilita o acesso de clusters do EMR em outras contas. Você pode especificar várias entidades principais, cada uma de uma conta diferente.

Usar perfis do IAM com aplicações que chamam diretamente os serviços da AWS

Os aplicativos executados nas instâncias do EC2 de um cluster podem usar o perfil da instância do EC2 para obter credenciais de segurança temporárias ao chamar serviços. AWS

As versões do Hadoop disponíveis com a versão 2.3.0 e posteriores do Amazon EMR já foram atualizadas para usar perfis do IAM. Se seu aplicativo é executado estritamente sobre a arquitetura do Hadoop e não chama diretamente nenhum serviço AWS, ele deve funcionar com funções do IAM sem modificações.

Se seu aplicativo chamar serviços AWS diretamente, você precisará atualizá-lo para aproveitar as funções do IAM. Isso significa que, em vez de obter credenciais de conta de `/etc/hadoop/conf/core-site.xml` nas instâncias do EC2 no cluster, sua aplicação usa um SDK para acessar os recursos usando perfis do IAM ou chama os metadados de instâncias do EC2 para obter as credenciais temporárias.

Para acessar AWS recursos com funções do IAM usando um SDK

- Os tópicos a seguir mostram como usar vários AWS SDKs para acessar credenciais temporárias usando funções do IAM. Cada tópico começa com uma versão de uma aplicação que não usa perfis do IAM e depois percorre o processo de conversão dessa aplicação para usar perfis do IAM.
  - Using IAM roles for Amazon EC2 instances with the SDK for Java no Guia do usuário do AWS SDK for Java
  - Using IAM roles for Amazon EC2 instances with the SDK for .NET no Guia do usuário do AWS SDK for .NET
  - Using IAM roles for Amazon EC2 instances with the SDK for PHP no Guia do usuário do AWS SDK for PHP
  - Using IAM roles for Amazon EC2 instances with the SDK for Ruby no Guia do usuário do AWS SDK for Ruby

Para obter credenciais temporárias de metadados de instâncias do EC2

- Chame a seguinte URL de uma instância do EC2 que está sendo executada com a função IAM especificada, que retorna as credenciais de segurança temporárias associadas (AccessKeyID, SecretAccessKey SessionToken, e expiração). O exemplo a seguir usa o perfil de instância padrão para o Amazon EMR, `EMR_EC2_DefaultRole`.

```
GET http://169.254.169.254/latest/meta-data/iam/security-
credentials/EMR_EC2_DefaultRole
```

Para obter mais informações sobre como escrever aplicativos que usam funções do IAM, consulte Conceder acesso AWS a recursos para aplicativos executados em instâncias do Amazon EC2.

Para obter mais informações sobre credenciais de segurança temporárias, consulte Using temporary security credentials no guia Using Temporary Security Credentials.

Permitir que usuários e grupos criem e modifiquem perfis

As entidades principais do IAM (usuários e grupos) que criam, modificam e especificam os perfis para um cluster, incluindo perfis padrão, devem ter permissão para executar as ações a seguir. Para obter detalhes sobre cada ação, consulte Actions na IAM API Reference.

- `iam:CreateRole`
- `iam:PutRolePolicy`
- `iam:CreateInstanceProfile`
- `iam:AddRoleToInstanceProfile`
- `iam:ListRoles`
- `iam:GetPolicy`
- `iam:GetInstanceProfile`
- `iam:GetPolicyVersion`
- `iam:AttachRolePolicy`
- `iam:PassRole`

A permissão `iam:PassRole` permite a criação do cluster. As permissões restantes permitem a criação das funções padrão.

Para obter informações sobre como atribuir permissões a um usuário, consulte Alteração de permissões de um usuário no Guia do usuário do IAM.

Exemplos de políticas baseadas em identidade do Amazon EMR

Por padrão, usuários e perfis não têm permissão para criar ou modificar recursos do Amazon EMR. Eles também não podem realizar tarefas usando a AWS API AWS Management Console AWS CLI, ou. Um administrador do IAM deve criar políticas do IAM que concedam aos usuários e perfis permissão para executarem operações de API específicas nos recursos especificados de que precisam. O administrador deve anexar essas políticas aos usuários ou grupos que exigem essas permissões.

Para saber como criar uma política baseada em identidade do IAM utilizando esses exemplos de documentos de política JSON, consulte Criar políticas na guia JSON no Guia do usuário do IAM.

Tópicos

Práticas recomendadas de políticas para o Amazon EMR

As políticas baseadas em identidade são muito eficientes. Determinam se alguém pode criar, acessar ou excluir recursos do Amazon EMR em sua conta. Essas ações podem gerar custos para sua AWS conta. Ao criar ou editar políticas baseadas em identidade, siga estas diretrizes e recomendações:

- Comece a usar políticas AWS gerenciadas — Para começar a usar o Amazon EMR rapidamente, use políticas AWS gerenciadas para dar aos seus funcionários as permissões de que precisam. Essas políticas já estão disponíveis em sua conta e são mantidas e atualizadas pela AWS. Para obter mais informações, consulte Comece a usar permissões com políticas AWS gerenciadas no Guia do usuário do IAM Políticas gerenciadas do Amazon EMR e.
- Conceder privilégio mínimo: ao criar políticas personalizadas, conceda apenas as permissões necessárias para executar uma tarefa. Comece com um conjunto mínimo de permissões e conceda permissões adicionais conforme necessário. Fazer isso é mais seguro do que começar

com permissões que são muito lenientes e tentar restringi-las superiormente. Para obter mais informações, consulte Conceder privilégio mínimo no Guia do usuário do IAM.

- Habilitar o MFA para operações confidenciais: para reforçar a segurança, exija que os usuários usem a autenticação multifator (MFA) para acessar recursos ou operações de API sigilosos. Para obter mais informações, consulte Usar autenticação multifator (MFA) AWS no Guia do usuário do IAM.
- Usar condições de política para segurança adicional: na medida do possível, defina as condições sob as quais suas políticas baseadas em identidade permitem o acesso a um recurso. Por exemplo, você pode gravar condições para especificar um intervalo de endereços IP permitidos do qual a solicitação deve partir. Você também pode escrever condições para permitir somente solicitações em uma data especificada ou período ou para exigir o uso de SSL ou MFA. Para obter mais informações, consulte Elementos de política JSON do IAM: condição no Guia do usuário do IAM.

Permitir que usuários visualizem suas próprias permissões

Este exemplo mostra como você pode criar uma política que permite que os usuários visualizem as políticas gerenciadas e em linha anexadas as respectivas identidades de usuário. Essa política inclui permissões para concluir essa ação no console ou programaticamente usando a API AWS CLI ou AWS .

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"ViewOwnUserInfo",
         "Effect":"Allow",
         "Action":[
            "iam:GetUser",
            "iam:GetUserPolicy",
            "iam:ListAttachedUserPolicies",
            "iam:ListGroupsForUser",
            "iam:ListUserPolicies"
         ],
         "Resource":[
            "arn:aws:iam::*:user/${aws:username}"
         ]
      },
      {
```

```
         "Sid":"NavigateInConsole",
         "Effect":"Allow",
         "Action":[
            "iam:GetGroupPolicy",
            "iam:GetPolicy",
            "iam:GetPolicyVersion",
            "iam:ListAttachedGroupPolicies",
            "iam:ListGroupPolicies",
            "iam:ListGroups",
            "iam:ListPolicies",
            "iam:ListPolicyVersions",
            "iam:ListUsers"
         ],
         "Resource":"*"
      }
   ]
}
```

Políticas gerenciadas do Amazon EMR

A maneira mais fácil de conceder acesso total ou acesso somente leitura a ações do Amazon EMR é usar as políticas gerenciadas do IAM para o Amazon EMR. Políticas gerenciadas oferecem o benefício de serem atualizadas automaticamente se os requisitos de permissões forem alterados. Se você usar políticas em linha, podem ocorrer alterações de serviço que provoquem erros de permissão.

O Amazon EMR substituirá as políticas gerenciadas já existentes (políticas v1) em favor de novas políticas gerenciadas (políticas v2). As novas políticas gerenciadas foram reduzidas para se alinharem às melhores práticas. AWS Depois que as políticas gerenciadas v1 existentes forem defasadas, não será possível anexar essas políticas a nenhum novo perfil ou usuário do IAM. Os perfis e os usuários existentes que usam políticas defasadas podem continuar a usá-las. As políticas gerenciadas v2 restringem o acesso usando etiquetas. Elas permitem somente ações específicas do Amazon EMR e exigem recursos de cluster marcados com uma chave específica do EMR. É recomendável analisar cuidadosamente a documentação antes de usar as novas políticas v2.

As políticas v1 serão marcadas como defasadas com um ícone de aviso próximo a elas na lista Políticas no console do IAM. As políticas defasadas terão as seguintes características:

- Continuarão funcionando para todos os usuários, grupos e perfis atualmente conectados. Nada é rompido.

- Não é possível anexar a novos usuários, grupos ou perfis. Se você desvincular uma das políticas de uma entidade atual, não poderá anexá-la novamente.
- Após separar uma política v1 de todas as entidades atuais, a política não estará mais visível e não poderá mais ser usada.

A tabela a seguir resume as alterações entre as políticas atuais (v1) e v2.

Alterações de políticas gerenciadas pelo Amazon EMR

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
Perfil de serviço do EMR padrão e política gerenciada anexada	Nome da função: EMR_ DefaultRole Política V1 (a ser descontinuada): Função AmazonElasticMapReduce(função de serviço do EMR) Nome da política v2 (com escopo reduzido): **AmazonEMRServicePolicy_v2**	Permite que o Amazon EMR chame outros AWS serviços em seu nome ao provisionar recursos e realizar ações em nível de serviço. Essa função é necessária para todos os clusters.	A política adiciona a nova permissão `"ec2:DescribeInstanceTypeOfferings"` . Essa operação de API retorna uma lista de tipos de instância compatíveis com uma lista de determinadas zonas de disponibilidade.
Política gerenciada pelo IAM para acesso total ao Amazon EMR por usuário, função ou grupo vinculado	Nome da política v2 (no escopo): **AmazonEMRServicePolicy_v2**	Concede aos usuários permissões completas para ações do EMR. Inclui iam: PassRole permissões para recursos.	A política adiciona o pré-requisito de que os usuários devem adicionar etiquetas de usuário aos recursos para poderem usar essa política. Consulte Etiquetar recursos

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
			para usar políticas gerenciadas. iam: a PassRole ação requer iam: PassedToService condição definida para o serviço especificado. O acesso ao Amazon EC2, ao Amazon S3 e a outros serviços não é permitido por padrão. Consulte IAM Managed Policy for Full Access (v2 Managed Default Policy).
Política gerenciada do IAM para acesso somente leitura ao EMR por usuário, perfil ou grupo vinculado	Política v1 (a ser defasada) : **AmazonElasticMapReduceReadOnlyAccess** Nome da política v2 (no escopo): **AmazonEMRReadOnlyAccessPolicy_v2**	Concede aos usuários permissões somente leitura para ações do Amazon EMR.	As permissões concedem somente ações específicas de leitura do elasticmapreduce. O acesso ao Amazon S3 é um acesso não concedido por padrão. Consulte IAM Managed Policy for Read-Only Access (v2 Managed Default Policy).

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
Perfil de serviço do EMR padrão e política gerenciada anexada	Nome da função: EMR_ DefaultRole Política V1 (a ser descontinuada): Função AmazonElasticMapReduce(função de serviço do EMR) Nome da política v2 (com escopo reduzido): **AmazonEMRServicePolicy_v2**	Permite que o Amazon EMR chame outros AWS serviços em seu nome ao provisionar recursos e realizar ações em nível de serviço. Essa função é necessária para todos os clusters.	O perfil de serviço v2 e a política padrão v2 substituem o perfil e a política defasados. A política adiciona o pré-requisito de que os usuários devem adicionar etiquetas de usuário aos recursos para poderem usar essa política. Consulte Etiquetar recursos para usar políticas gerenciadas. Consulte Perfil de serviço para Amazon EMR (perfil do EMR).

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2)	Política V1 (a ser descontinuada): DefaultRoleEMR_EC2_ (perfil de instância) Nome da política obsoleta: EC2Role AmazonElastic MapReducefor	Permite que aplicações executadas em um cluster do EMR acessem outros recursos da AWS , como o Amazon S3. Por exemplo, se você executar trabalhos do Apache Spark que processam dados do Amazon S3, a política precisará permitir o acesso a esses recursos.	Tanto o perfil padrão como a política padrão estão prestes a serem defasados . Não há nenhuma função ou política gerenciada AWS padrão de substituição. É necessário fornecer uma política baseada em recursos ou em identidade. Isso significa que, por padrão, as aplicações executados em um cluster do EMR não têm acesso ao Amazon S3 ou a outros recursos, a menos que você os adicione à política manualmente. Consulte Perfil padrão e política gerenciada.

Tipo de política	Nomes de política	Finalidade da política	Alterações na política v2
Outras políticas de perfil de serviço do EC2	Nomes atuais da política: AmazonElasticMapReduceforAutoScalingRole AmazonElasticMapReduceEditorsRole, AmazonEMR CleanupPolicy	Fornece as permissões que o Amazon EMR precisa para acessar outros AWS recursos e realizar ações se estiver usando escalabilidade automática, notebooks ou para limpar recursos do EC2.	Nenhuma alteração na v2.

Objetivo de proteção: PassRole

As políticas gerenciadas padrão de permissões completas do Amazon EMR incorporam configurações de segurança `iam:PassRole`, incluindo estas:

- Permissões `iam:PassRole` somente para perfis padrão específicos do Amazon EMR.
- `iam:PassedToService`condições que permitem que você use a política somente com AWS serviços específicos, como `elasticmapreduce.amazonaws.com` `ec2.amazonaws.com` e.

Você pode visualizar a versão JSON das políticas AmazonEMR FullAccess Policy_v2 e AmazonEMR _v2 no console do IAM. ServicePolicy É recomendável criar novos clusters com políticas gerenciadas v2.

Para criar políticas personalizadas, recomendamos que você comece com as políticas gerenciadas e edite-as de acordo com seus requisitos.

Para obter informações sobre como anexar políticas aos usuários (entidades principais), consulte AWS Management ConsoleWorking with managed policies using the no Guia do usuário do IAM.

Etiquetar recursos para usar políticas gerenciadas

O AmazonEMR ServicePolicy _v2 e o AmazonEMR FullAccess Policy_v2 dependem do acesso limitado aos recursos que o Amazon EMR provisiona ou usa. Obtém-se o escopo reduzido

restringindo o acesso somente aos recursos que têm uma etiqueta de usuário predefinida associada a eles. Ao usar qualquer uma dessas duas políticas, é necessário passar a etiqueta de usuário predefinida `for-use-with-amazon-emr-managed-policies = true` ao provisionar o cluster. O Amazon EMR propagará essa etiqueta automaticamente. Além disso, é necessário adicionar uma etiqueta de usuário aos recursos listados na seção a seguir. Se você usa o console do Amazon EMR para iniciar seu cluster, consulte Considerações sobre o uso do console do Amazon EMR para iniciar clusters com políticas gerenciadas v2.

Para usar políticas gerenciadas, passe a etiqueta de usuário `for-use-with-amazon-emr-managed-policies = true` ao provisionar um cluster com a CLI, o SDK ou outro método.

Quando você passa a etiqueta, o Amazon EMR propaga a etiqueta para a sub-rede privada ENI, a instância do EC2 e os volumes do EBS que ele cria. O Amazon EMR também etiqueta automaticamente os grupos de segurança que ele cria. No entanto, para que o Amazon EMR seja iniciado com determinado grupo de segurança, você deve etiquetá-lo. Para recursos que não são criados pelo Amazon EMR, é necessário adicionar etiquetas a esses recursos. Por exemplo, é necessário etiquetar sub-redes do Amazon EC2, grupos de segurança do EC2 (se não forem criados pelo Amazon EMR) e VPCs (para que o Amazon EMR crie grupos de segurança). Para iniciar clusters com políticas gerenciadas v2 em VPCs, você deve marcar essas VPCs com a etiqueta de usuário predefinida. Consulte Considerações sobre o uso do console do Amazon EMR para iniciar clusters com políticas gerenciadas v2.

Marcação propagada especificada pelo usuário

O Amazon EMR marca os recursos que ele cria usando as etiquetas do Amazon EMR especificadas ao criar um cluster. O Amazon EMR aplica etiquetas aos recursos que cria durante a vida útil do cluster.

O Amazon EMR propaga etiquetas de usuário para os seguintes recursos:

- ENI da sub-rede privada (interfaces de rede elástica de acesso ao serviço)
- Instâncias do EC2
- Volumes do EC2
- Modelos de inicialização do EC2

Grupos de segurança etiquetados automaticamente

O Amazon EMR marca os grupos de segurança do EC2 que ele cria com a etiqueta necessária para políticas gerenciadas v2 para o Amazon EMR, `for-use-with-amazon-emr-managed-`

`policies`, independentemente das etiquetas que você especificar no comando para criar o cluster. Em um grupo de segurança criado antes da introdução das políticas gerenciadas v2, o Amazon EMR não etiqueta o grupo de segurança automaticamente. Para usar políticas gerenciadas v2 com os grupos de segurança padrão que já existem na conta, você precisa etiquetar os grupos de segurança manualmente com `for-use-with-amazon-emr-managed-policies = true`.

Recursos de cluster etiquetados manualmente

É necessário etiquetar alguns recursos do cluster manualmente para que eles possam ser acessados pelos perfis padrão do Amazon EMR.

- Você deve marcar manualmente os grupos de segurança do EC2 e as sub-redes do EC2 com a etiqueta de política gerenciada do Amazon EMR `for-use-with-amazon-emr-managed-policies`.
- Você deve etiquetar manualmente uma VPC, se quiser que o Amazon EMR crie grupos de segurança padrão. O EMR tentará criar um grupo de segurança com a etiqueta específica se o grupo de segurança padrão ainda não existir.

O Amazon EMR marca automaticamente os seguintes recursos:

- Grupos de segurança do EC2 criados pelo EMR

Você deve etiquetar manualmente os seguintes recursos:

- Sub-rede EC2
- Grupos de segurança do EC2

Opcionalmente, você pode etiquetar manualmente os seguintes recursos:

- VPC: somente quando quiser que o Amazon EMR crie grupos de segurança

Considerações sobre o uso do console do Amazon EMR para iniciar clusters com políticas gerenciadas v2

É possível provisionar clusters com políticas gerenciadas v2 usando o console do Amazon EMR. Veja aqui algumas considerações ao usar o console para iniciar clusters do Amazon EMR.

> **Note**
>
> Nós reformulamos o console do Amazon EMR. O recurso de marcação automática ainda não está disponível no novo console, e o novo console também não exibe quais recursos (VPC/sub-redes) precisam ser etiquetados. Consulte Console do Amazon EMR para saber mais sobre as diferenças entre as experiências do console antigo e do novo.

- Não é necessário passar a etiqueta predefinida. O Amazon EMR adiciona a etiqueta automaticamente e a propaga para os componentes adequados.
- Para componentes que precisam ser etiquetados manualmente, o antigo console do Amazon EMR tenta etiquetá-los automaticamente se você tiver as permissões necessárias para etiquetar recursos. Se você não tiver as permissões para etiquetar recursos ou se quiser usar o novo console, peça para o administrador etiquetar esses recursos.
- Não é possível iniciar clusters com políticas gerenciadas v2 sem que todos os pré-requisitos sejam atendidos.
- O console antigo do Amazon EMR mostra quais recursos (VPC/sub-redes) precisam ser etiquetados.

Política gerenciada do IAM para acesso total (política gerenciada v2 padrão)

As políticas gerenciadas padrão do EMR com escopo para v2 concedem privilégios de acesso específicos aos usuários. Eles exigem uma etiqueta de recurso predefinida do Amazon EMR e chaves de condição `iam:PassRole` para os recursos usados pelo Amazon EMR, como a `Subnet` e o `SecurityGroup` que você usa para iniciar o cluster.

Para conceder as ações necessárias para o Amazon EMR, anexe a política gerenciada `AmazonEMRFullAccessPolicy_v2`. Essa política gerenciada padrão atualizada substitui a política gerenciada `AmazonElasticMapReduceFullAccess`.

`AmazonEMRFullAccessPolicy_v2` depende do acesso de escopo limitado aos recursos que o Amazon EMR provisiona ou usa. Ao usar essa política, é necessário passar a etiqueta de usuário `for-use-with-amazon-emr-managed-policies = true` ao provisionar o cluster. O Amazon EMR propagará a etiqueta automaticamente. Além disso, talvez seja necessário adicionar manualmente uma etiqueta de usuário a tipos específicos de recursos, como grupos de segurança do EC2 que não foram criados pelo Amazon EMR. Para ter mais informações, consulte Etiquetar recursos para usar políticas gerenciadas.

A política AmazonEMRFullAccessPolicy_v2 protege os recursos fazendo o seguinte:

- Requer que os recursos sejam marcados com a etiqueta predefinida de políticas gerenciadas do Amazon EMR `for-use-with-amazon-emr-managed-policies` para criação de clusters e acesso ao Amazon EMR.
- Restringe a ação `iam:PassRole` a perfis padrão específicos e acesso `iam:PassedToService` a serviços específicos.
- Não permite mais acesso ao Amazon EC2, ao Amazon S3 e a outros serviços por padrão.

Veja a seguir o conteúdo dessa política.

> **Note**
>
> Você também pode usar o link do console AmazonEMRFullAccessPolicy_v2 para visualizar a política.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "RunJobFlowExplicitlyWithEMRManagedTag",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:RunJobFlow"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
                }
            }
        },
        {
            "Sid": "ElasticMapReduceActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:AddInstanceFleet",
                "elasticmapreduce:AddInstanceGroups",
                "elasticmapreduce:AddJobFlowSteps",
```

```
                "elasticmapreduce:AddTags",
                "elasticmapreduce:CancelSteps",
                "elasticmapreduce:CreateEditor",
                "elasticmapreduce:CreateSecurityConfiguration",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:DeleteSecurityConfiguration",
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:DescribeJobFlows",
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:DescribeReleaseLabel",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
                "elasticmapreduce:GetManagedScalingPolicy",
                "elasticmapreduce:GetAutoTerminationPolicy",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListSupportedInstanceTypes",
                "elasticmapreduce:ModifyCluster",
                "elasticmapreduce:ModifyInstanceFleet",
                "elasticmapreduce:ModifyInstanceGroups",
                "elasticmapreduce:OpenEditorInConsole",
                "elasticmapreduce:PutAutoScalingPolicy",
                "elasticmapreduce:PutBlockPublicAccessConfiguration",
                "elasticmapreduce:PutManagedScalingPolicy",
                "elasticmapreduce:RemoveAutoScalingPolicy",
                "elasticmapreduce:RemoveManagedScalingPolicy",
                "elasticmapreduce:RemoveTags",
                "elasticmapreduce:SetTerminationProtection",
                "elasticmapreduce:StartEditor",
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:TerminateJobFlows",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole"
            ],
            "Resource": "*"
        },
        {
            "Sid": "ViewMetricsInEMRConsole",
```

```
            "Effect": "Allow",
            "Action": [
                "cloudwatch:GetMetricStatistics"
            ],
            "Resource": "*"
        },
        {
            "Sid": "PassRoleForElasticMapReduce",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": [
                "arn:aws:iam::*:role/EMR_DefaultRole",
                "arn:aws:iam::*:role/EMR_DefaultRole_V2"
            ],
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "PassRoleForEC2",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/EMR_EC2_DefaultRole",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "ec2.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "PassRoleForAutoScaling",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/EMR_AutoScaling_DefaultRole",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "application-autoscaling.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "ElasticMapReduceServiceLinkedRole",
```

```
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
            "Resource": "arn:aws:iam::*:role/aws-service-role/
elasticmapreduce.amazonaws.com*/AWSServiceRoleForEMRCleanup*",
            "Condition": {
                "StringEquals": {
                    "iam:AWSServiceName": [
                        "elasticmapreduce.amazonaws.com",
                        "elasticmapreduce.amazonaws.com.cn"
                    ]
                }
            }
        },
        {
            "Sid": "ConsoleUIActions",
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeAccountAttributes",
                "ec2:DescribeAvailabilityZones",
                "ec2:DescribeImages",
                "ec2:DescribeKeyPairs",
                "ec2:DescribeNatGateways",
                "ec2:DescribeRouteTables",
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "ec2:DescribeVpcEndpoints",
                "s3:ListAllMyBuckets",
                "iam:ListRoles"
            ],
            "Resource": "*"
        }
    ]
}
```

Política gerenciada do IAM para acesso total (prestes a ser defasada)

As políticas gerenciadas `AmazonElasticMapReduceFullAccess` e `AmazonEMRFullAccessPolicy_v2` AWS Identity and Access Management (IAM) concedem todas as ações necessárias para o Amazon EMR e outros serviços.

> **Important**
>
> A política gerenciada `AmazonElasticMapReduceFullAccess` está prestes a ser defasada e seu uso com o Amazon EMR não é mais recomendado. Em seu lugar, use `AmazonEMRFullAccessPolicy_v2`. Quando o serviço do IAM eventualmente defasar a política v1, você não poderá vinculá-la a um perfil. No entanto, você pode anexar um perfil já existente a um cluster mesmo que esse perfil use a política defasada.

As políticas gerenciadas padrão de permissões completas do Amazon EMR incorporam configurações de segurança `iam:PassRole`, incluindo estas:

- Permissões `iam:PassRole` somente para perfis padrão específicos do Amazon EMR.
- `iam:PassedToService`condições que permitem que você use a política somente com AWS serviços específicos, como `elasticmapreduce.amazonaws.com` `ec2.amazonaws.com` e.

Você pode visualizar a versão JSON das políticas AmazonEMR FullAccess Policy_v2 e AmazonEMR _v2 no console do IAM. ServicePolicy É recomendável criar novos clusters com políticas gerenciadas v2.

Você pode ver o conteúdo da política v1 obsoleta no endereço. AWS Management Console `AmazonElasticMapReduceFullAccess` A ação `ec2:TerminateInstances` na política concede permissão a um usuário ou perfil para terminar qualquer uma das instâncias do Amazon EC2 associadas à conta do IAM. Inclui instâncias que não fazem parte de um cluster do EMR.

Política gerenciada do IAM para acesso somente leitura (política gerenciada v2 padrão)

Para conceder privilégios de somente leitura ao Amazon EMR, anexe a política gerenciada AmazonEMR _v2. ReadOnly AccessPolicy Essa política gerenciada padrão substitui a política gerenciada `AmazonElasticMapReduceReadOnlyAccess`.
O conteúdo dessa instrução de política é mostrado no trecho a seguir. Em comparação com a política `AmazonElasticMapReduceReadOnlyAccess`, a política `AmazonEMRReadOnlyAccessPolicy_v2` não usa caracteres curinga para o elemento `elasticmapreduce`. Em vez disso, a política v2 padrão define o escopo das ações `elasticmapreduce` que podem ser permitidas.

> **Note**
>
> Você também pode usar o AWS Management Console link AmazonEMRReadOnlyAccessPolicy_v2para ver a política.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ElasticMapReduceActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:DescribeJobFlows",
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:DescribeReleaseLabel",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
                "elasticmapreduce:GetManagedScalingPolicy",
                "elasticmapreduce:GetAutoTerminationPolicy",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListSupportedInstanceTypes",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole"
            ],
            "Resource": "*"
        },
        {
            "Sid": "ViewMetricsInEMRConsole",
            "Effect": "Allow",
            "Action": [
                "cloudwatch:GetMetricStatistics"
            ],
            "Resource": "*"
```

```
        }
    ]
}
```

Política gerenciada do IAM para acesso somente leitura (prestes a ser defasada)

A política gerenciada `AmazonElasticMapReduceReadOnlyAccess` está prestes a ser defasada. Não é possível anexar essa política ao iniciar novos clusters. `AmazonElasticMapReduceReadOnlyAccess` foi substituída pela política gerenciada `AmazonEMRReadOnlyAccessPolicy_v2` padrão do Amazon EMR. O conteúdo dessa instrução de política é mostrado no trecho a seguir. Caracteres curinga para o elemento `elasticmapreduce` especificam que somente as ações que comecem com as strings especificadas serão permitidas. Lembre-se de que, como essa política não nega explicitamente as ações, uma declaração de política diferente ainda pode ser usada para conceder acesso a ações específicas.

> **Note**
>
> Você também pode usar o AWS Management Console para visualizar a política.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:Describe*",
        "elasticmapreduce:List*",
        "elasticmapreduce:ViewEventsFromAllClustersInConsole",
        "s3:GetObject",
        "s3:ListAllMyBuckets",
        "s3:ListBucket",
        "sdb:Select",
        "cloudwatch:GetMetricStatistics"
      ],
      "Resource": "*"
    }
  ]
}
```

AWS política gerenciada: EMR DescribeCluster PolicyFor EMRWAL

Não é possível anexar `EMRDescribeClusterPolicyForEMRWAL` às entidades do IAM. Essa política está vinculada a uma função vinculada ao serviço que permite que o Amazon EMR execute ações em seu nome. Para obter mais informações sobre essa função vinculada ao serviço, consulte. Usando funções vinculadas ao serviço para registro antecipado

Essa política concede permissões somente de leitura que permitem que o serviço WAL do Amazon EMR encontre e retorne o status de um cluster. Para obter mais informações sobre o Amazon EMR WAL, consulte Write-ahead logs (WAL) para o Amazon EMR.

Detalhes de permissão

Esta política inclui as seguintes permissões:

- `emr`— Permite que os diretores descrevam o status do cluster do Amazon EMR. Isso é necessário para que o Amazon EMR possa confirmar quando um cluster foi encerrado e, depois de trinta dias, limpar todos os registros do WAL deixados pelo cluster.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:DescribeCluster"
      ],
      "Resource": "*"
    }
  ]
}
```

AWS políticas gerenciadas para o Amazon EMR

Uma política AWS gerenciada é uma política autônoma criada e administrada por AWS. AWS as políticas gerenciadas são projetadas para fornecer permissões para muitos casos de uso comuns, para que você possa começar a atribuir permissões a usuários, grupos e funções.

Lembre-se de que as políticas AWS gerenciadas podem não conceder permissões de privilégio mínimo para seus casos de uso específicos porque estão disponíveis para uso de todos os AWS

clientes. Recomendamos que você reduza ainda mais as permissões definindo políticas gerenciadas pelo cliente da específicas para seus casos de uso.

Você não pode alterar as permissões definidas nas políticas AWS gerenciadas. Se AWS atualizar as permissões definidas em uma política AWS gerenciada, a atualização afetará todas as identidades principais (usuários, grupos e funções) às quais a política está anexada. AWS é mais provável que atualize uma política AWS gerenciada quando uma nova AWS service (Serviço da AWS) é lançada ou novas operações de API são disponibilizadas para serviços existentes.

Para mais informações, consulte Políticas gerenciadas pela AWS no Manual do usuário do IAM.

Atualizações do Amazon EMR para AWS políticas gerenciadas

Veja detalhes sobre as atualizações das políticas AWS gerenciadas do Amazon EMR desde que esse serviço começou a rastrear essas alterações.

Alteração	Descrição	Data
EMRDescribeCluster PolicyForEMRWAL – Nova política	Foi adicionada uma nova política para que o Amazon EMR possa determinar o status do cluster para limpeza do WAL trinta dias após o término do cluster.	10 de agosto de 2023
AmazonEMRFullAcces sPolicy_v2 e **AmazonEMRReadOnlyA ccessPolicy_v2** : atualizar para uma política já existente	Adicionadas `elasticma preduce:DescribeRe leaseLabel` e `elasticmapreduce:G etAutoTerminationP olicy` .	21 de abril de 2022
AmazonEMRFullAcces sPolicy_v2 : atualização para uma política existente	Adicionou-se `ec2:Descr ibeImages` para Usar uma AMI personalizada.	15 de fevereiro de 2022
Políticas gerenciadas do Amazon EMR	Atualizado para esclarecer o uso de etiquetas de usuário predefinidas.	29 de setembro de 2021

Alteração	Descrição	Data
	Foi adicionada uma seção sobre como usar o AWS console para iniciar clusters com políticas gerenciadas v2.	
AmazonEMRFullAccessPolicy_v2: atualização para uma política existente	Foram alteradas as ações `PassRoleForAutoScaling` e `PassRoleForEC2` para usar o operador de condição `StringLike` para corresponder a `"iam:PassedToService":"application-autoscaling.amazonaws.com*"` e `"iam:PassedToService":"ec2.amazonaws.com*"`, respectivamente.	20 de maio de 2021
AmazonEMRFullAccessPolicy_v2: atualização para uma política existente	A ação inválida `s3:ListBuckets` foi removida e substituída pela ação `s3:ListAllMyBuckets`. Foi atualizada a criação do perfil vinculado ao serviço (SLR) para ser explicitamente reduzida ao único SLR que o Amazon EMR tem com princípios de serviço explícitos. Os SLRs que podem ser criados são exatamente os mesmos de antes da alteração.	23 de março de 2021

Alteração	Descrição	Data
AmazonEMRFullAccessPolicy_v2 – Nova política	O Amazon EMR adicionou novas permissões para definir o escopo do acesso aos recursos e adicionar o pré-requisito de que os usuários devem adicionar uma etiqueta de usuário predefinida aos recursos para poderem usar as políticas gerenciadas do Amazon EMR. A ação `iam:PassRole` requer uma condição `iam:PassedToService` definida para o serviço especificado. O acesso ao Amazon EC2, ao Amazon S3 e a outros serviços não é permitido por padrão.	11 de março de 2021
AmazonEMRServicePolicy_v2 – Nova política	Adiciona o pré-requisito de que os usuários devem adicionar etiquetas de usuário aos recursos para poderem usar essa política.	11 de março de 2021
AmazonEMRReadOnlyAccessPolicy_v2 – Nova política	As permissões concedem somente ações específicas de leitura do elasticmapreduce. O acesso ao Amazon S3 é um acesso não concedido por padrão.	11 de março de 2021

Alteração	Descrição	Data
O Amazon EMR passou a monitorar alterações	O Amazon EMR começou a monitorar as mudanças em suas políticas AWS gerenciadas.	11 de março de 2021

Políticas do IAM para acesso baseado em etiquetas a clusters e Cadernos do EMR

É possível aplicar condições em sua política baseada em identidade para controlar o acesso aos clusters e blocos de anotações do EMR baseados em tags.

Para obter mais informações sobre como adicionar tags a clusters, consulte Marcar clusters do EMR.

Os exemplos a seguir demonstram diferentes cenários e maneiras de usar operadores de condição com chaves de condição do Amazon EMR. Estas instruções de política do IAM são destinadas somente para fins de demonstração e não devem ser usadas em ambientes de produção. Há várias maneiras de combinar declarações de políticas para conceder e negar permissões de acordo com seus requisitos. Para obter mais informações sobre como planejar e testar políticas do IAM, consulte o Guia do usuário do IAM.

> **Important**
>
> Recusar, explicitamente, permissões para ações de uso de tags é uma consideração importante. Isso evita que os usuários façam a marcação de um recurso e, assim, concedam a si mesmos permissões que você não pretendia conceder. Se você não negar as ações de marcação de um recurso, o usuário poderá modificar as etiquetas e contornar a intenção das políticas baseadas em etiquetas.

Exemplo de instruções de políticas baseadas em identidade para clusters

Os exemplos a seguir demonstram a políticas de permissões baseadas em identidade que são usadas para controlar as ações permitidas com clusters do EMR.

> **Important**
>
> A ação `ModifyInstanceGroup` do Amazon EMR não exige que você especifique um ID de cluster. Por isso, negar essa ação com base em etiquetas de cluster requer mais atenção. Para ter mais informações, consulte Negando a ação ModifyInstanceGroup .

Tópicos

- Permitir ações somente em clusters com determinados valores de etiqueta
- Exigir a marcação do cluster quando um cluster é criado
- Permitir ações em clusters com uma etiqueta específica, independentemente do valor da etiqueta

Permitir ações somente em clusters com determinados valores de etiqueta

Os exemplos a seguir demonstram uma política que permite ao usuário executar ações com base na etiqueta de cluster *`department`* com o valor *`dev`* e também permite que o usuário atribua etiquetas a clusters com a mesma etiqueta. O último exemplo de política demonstra como negar privilégios para atribuir tags a clusters do EMR com qualquer coisa, menos a mesma tag.

No exemplo de política a seguir, o operador de condição `StringEquals` tenta corresponder *`dev`* com o valor da tag *`department`*. Se a tag *`department`* ainda não tiver sido adicionada ao cluster ou não contiver o valor *`dev`*, a política não se aplicará e as ações não serão permitidas por essa política. Se nenhuma outra declaração de política permitir as ações, o usuário poderá somente trabalhar com clusters que tenham essa tag com esse valor.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt12345678901234",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:DescribeCluster",
        "elasticmapreduce:ListSteps",
        "elasticmapreduce:TerminateJobFlows",
        "elasticmapreduce:SetTerminationProtection",
        "elasticmapreduce:ListInstances",
        "elasticmapreduce:ListInstanceGroups",
        "elasticmapreduce:ListBootstrapActions",
```

```
        "elasticmapreduce:DescribeStep"
      ],
      "Resource": [
        "*"
      ],
      "Condition": {
        "StringEquals": {
          "elasticmapreduce:ResourceTag/department": "dev"
        }
      }
    }
  ]
}
```

Você também pode especificar vários valores de tag usando um operador de condição. Por exemplo, para permitir todas as ações em clusters em que a tag *department* contenha o valor *dev* ou *test*, você poderia substituir o bloco condicional no exemplo anterior com o seguinte.

```
            "Condition": {
              "StringEquals": {
                "elasticmapreduce:ResourceTag/department":["dev", "test"]
              }
            }
```

Exigir a marcação do cluster quando um cluster é criado

Como no exemplo anterior, o exemplo de política a seguir procura a mesma etiqueta correspondente: o valor *dev* para a etiqueta *department*. Mas neste exemplo, a chave de condição `RequestTag` especifica que a política se aplica durante a criação da etiqueta. Portanto, é necessário criar um cluster com uma etiqueta que corresponda ao valor especificado.

Para criar um cluster com uma etiqueta, também é necessário ter permissão para a ação `elasticmapredue:AddTags`. Para essa instrução, a chave de condição `elasticmapreduce:ResourceTag` garante que o IAM conceda acesso somente aos recursos da etiqueta com o valor *dev* na etiqueta *department*. O elemento `Resource` é usado para limitar essa permissão aos recursos do cluster.

Para os `PassRole` recursos, você deve fornecer o ID ou alias da AWS conta, o nome da função de serviço na `PassRoleForEMR` declaração e o nome do perfil da instância na `PassRoleForEC2`

declaração. Para obter mais informações sobre o formato de ARN do IAM, consulte ARNs do IAM no Guia do usuário do IAM.

Para obter mais informações sobre a correspondência de valores de chave de etiqueta, consulte `aws:RequestTag/tag-key` no Guia do usuário do IAM.

```
{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Sid": "RunJobFlowExplicitlyWithTag",
        "Effect": "Allow",
        "Action": [
          "elasticmapreduce:RunJobFlow"
        ],
        "Resource": "*",
        "Condition": {
          "StringEquals": {
            "aws:RequestTag/department": "dev"
          }
        }
      },
      {
        "Sid": "AddTagsForDevClusters",
        "Effect": "Allow",
        "Action": "elasticmapreduce:AddTags",
        "Resource": "arn:aws:elasticmapreduce:*:*:cluster/*",
        "Condition": {
          "StringEquals": {
            "elasticmapreduce:ResourceTag/department": "dev"
          }
        }
      },
      {
        "Sid": "PassRoleForEMR",
        "Effect": "Allow",
        "Action": "iam:PassRole",
        "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
        "Condition": {
          "StringLike": {
            "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
          }
        }
```

```
      },
      {
        "Sid": "PassRoleForEC2",
        "Effect": "Allow",
        "Action": "iam:PassRole",
        "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
        "Condition": {
          "StringLike": {
            "iam:PassedToService": "ec2.amazonaws.com*"
          }
        }
      }
    ]
}
```

Permitir ações em clusters com uma etiqueta específica, independentemente do valor da etiqueta

Você também pode permitir ações somente em clusters que tenham uma determinada tag, independentemente do valor da tag. Para fazer isso, você pode usar o operador `Null`. Para obter mais informações, consulte Operador de condição para verificar a existência de chaves de condição no Guia do usuário do IAM. Por exemplo, para permitir ações somente em clusters do EMR que tenham a tag *`department`*, independentemente do valor que ela contenha, você poderia substituir o bloco condicional no exemplo anterior pelo seguinte. O operador `Null` procura a presença da tag *`department`* em um cluster do EMR. Se a tag existir, a instrução `Null` será avaliada como falsa, correspondendo à condição especificada nesta declaração de política e as ações apropriadas serão permitidas.

```
"Condition": {
  "Null": {
    "elasticmapreduce:ResourceTag/department":"false"
  }
}
```

A declaração de política a seguir permite que um usuário crie um cluster do EMR somente se o cluster tiver uma tag *`department`*, que possa conter qualquer valor. Para o `PassRole` recurso, você precisa fornecer o ID ou alias da AWS conta e o nome da função de serviço. Para obter mais informações sobre o formato de ARN do IAM, consulte ARNs do IAM no Guia do usuário do IAM.

Para obter mais informações sobre como especificar o operador de condição nulo (“false”), consulte Operador de condição para verificar a existência de chaves de condição no Guia do usuário do IAM.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "CreateClusterTagNullCondition",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:RunJobFlow"
            ],
            "Resource": [
                "*"
            ],
            "Condition": {
                "Null": {
                    "aws:RequestTag/department": "false"
                }
            }
        },
        {
            "Sid": "AddTagsNullCondition",
            "Effect": "Allow",
            "Action": "elasticmapreduce:AddTags",
            "Resource": "arn:aws:elasticmapreduce:*:*:cluster/*",
            "Condition": {
                "Null": {
                    "elasticmapreduce:ResourceTag/department": "false"
                }
            }
        },
        {
            "Sid": "PassRoleForElasticMapReduce",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
                }
            }
        },
```

```
        {
            "Sid": "PassRoleForEC2",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "ec2.amazonaws.com*"
                }
            }
        }
    ]
}
```

Exemplo de instruções de políticas baseadas em identidade para Cadernos do EMR

Os exemplos de instruções de políticas do IAM nesta seção demonstram cenários comuns para usar chaves a fim de limitar as ações permitidas usando Cadernos do EMR. Desde que nenhuma outra política associada à entidade principal (usuário) permita as ações, as chaves de contexto de condição limitam as ações permitidas conforme indicado.

Example : permitir acesso somente aos Cadernos do EMR que o usuário cria com base na marcação

A instrução de política de exemplo a seguir, quando anexada a um perfil ou usuário, permite que o usuário trabalhe apenas com cadernos criados por ele. Esta declaração de política usa a tag padrão aplicada quando um bloco de anotações é criado.

No exemplo, o operador de condição `StringEquals` tenta combinar uma variável que representa o ID do usuário atual (`{aws:userId}`) com o valor de etiqueta `creatorUserID`. Se a tag `creatorUserID` ainda não tiver sido adicionada ao bloco de anotações ou não contiver o valor do ID do usuário atual, a política não se aplicará e as ações não serão permitidas por essa política. Se nenhuma outra declaração de política permitir as ações, o usuário só poderá trabalhar com blocos de anotações que tenham essa tag com esse valor.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:StartEditor",
```

```
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userId}"
                }
            }
        }
    ]
}
```

Example –Exigir marcação de caderno quando um caderno é criado

Neste exemplo, a chave de contexto `RequestTag` é usada. A ação `CreateEditor` será permitida somente se o usuário não alterar nem excluir a tag `creatorUserID` é que é adicionada por padrão. A variável ${aws:userId} especifica o ID de usuário do usuário atualmente ativo, que é o valor padrão da etiqueta.

A declaração de política pode ser usada para ajudar a garantir que os usuários não removam a tag `createUserId` tag nem alterem seu valor.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:RequestTag/creatorUserId": "${aws:userid}"
                }
            }
        }
    ]
}
```

Este exemplo requer que o usuário crie o cluster com uma tag com a string de chave `dept` e um valor definido como um dos seguintes: `datascience`, `analytics`, `operations`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:RequestTag/dept": [
                        "datascience",
                        "analytics",
                        "operations"
                    ]
                }
            }
        }
    ]
}
```

Example –Limitar a criação do caderno para clusters marcados e exigir etiquetas de caderno

Este exemplo permite a criação do bloco de anotações somente se o bloco de anotações for criado com uma tag que tenha a string de chave `owner` definida como um dos valores especificados. Além disso, o bloco de anotações poderá ser criado somente se o cluster tiver uma tag com a string de chave `department` definida como um dos valores especificados.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
```

```
                "StringEquals": {
                    "elasticmapreduce:RequestTag/owner": [
                        "owner1",
                        "owner2",
                        "owner3"
                    ],
                    "elasticmapreduce:ResourceTag/department": [
                        "dep1",
                        "dep3"
                    ]
                }
            }
        }
    ]
}
```

Example –Limitar a capacidade de iniciar um caderno com base em etiquetas

Este exemplo limita a capacidade de iniciar blocos de anotações àqueles que tenham uma tag com a string de chave `owner` definida como um dos valores especificados. Como o elemento `Resource` é usado para especificar apenas o `editor`, a condição não se aplica ao cluster e não precisa ser marcado.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "owner1",
                        "owner2"
                    ]
                }
            }
        }
    ]
```

```
}
```

Este exemplo é semelhante ao exposto acima. No entanto, o limite se aplica apenas a clusters com tags, e não a blocos de anotações.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "dep1",
                        "dep3"
                    ]
                }
            }
        }
    ]
}
```

Este exemplo usa um conjunto diferente de tags de cluster e bloco de anotações. Ele permite que um bloco de anotações seja iniciado somente se:

- O bloco de anotações tiver uma tag com a string de chave `owner` definida como qualquer um dos valores especificados

  —e—
- O cluster tiver uma tag com a string de chave `department` definida como qualquer um dos valores especificados

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
```

```
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "user1",
                        "user2"
                    ]
                }
            }
        },
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "datascience",
                        "analytics"
                    ]
                }
            }
        }
    ]
}
```

Example –Limitar a capacidade de abrir o editor de caderno com base em etiquetas

Este exemplo permite que o editor de blocos de anotações seja aberto somente se:

- O bloco de anotações tiver uma tag com a string de chave `owner` definida como qualquer um dos valores especificados.

  —e—

- O cluster tiver uma tag com a string de chave `department` definida como qualquer um dos valores especificados.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "user1",
                        "user2"
                    ]
                }
            }
        },
        {
            "Action": [
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "datascience",
                        "analytics"
                    ]
                }
            }
        }
    ]
}
```

Negando a ação ModifyInstanceGroup

A ação ModifyInstanceGrupos no Amazon EMR não exige que você forneça um ID de cluster com a ação. Em vez disso, você pode especificar apenas um ID de grupo de instâncias. Por isso, uma política de negação aparentemente simples para essa ação com base no ID do cluster ou em uma etiqueta do cluster pode não ter o efeito pretendido. Considere a seguinte política de exemplo.

```
{
 "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-east-1:123456789012:cluster/
j-12345ABCDEFG67"
        }
    ]

}
```

Se um usuário com essa política anexada realizar uma ação `ModifyInstanceGroup` e especificar somente o ID do grupo de instâncias, a política não se aplicará. Como a ação é permitida em todos os outros recursos, ela tem êxito.

Uma solução para esse problema é anexar uma declaração de política à identidade que usa um NotResourceelemento para negar qualquer `ModifyInstanceGroup` ação emitida sem um ID de cluster. O exemplo de política a seguir adiciona essa instrução de negação para que qualquer solicitação `ModifyInstanceGroups` falhe, a menos que um ID de cluster esteja especificado. Como uma identidade deve especificar um ID de cluster com a ação, as instruções de negação com base no ID do cluster são, portanto, efetivas.

```
{
 "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
```

```
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-east-1:123456789012:cluster/
j-12345ABCDEFG67"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "NotResource": "arn:*:elasticmapreduce:*:*:cluster/*"
        }
    ]

}
```

Um problema semelhante ocorre quando você deseja negar a ação `ModifyInstanceGroups` com base no valor associado a uma etiqueta de cluster. A solução é semelhante. Além de uma instrução de negação que especifica o valor da etiqueta, é possível adicionar uma instrução de política que nega a ação `ModifyInstanceGroup` se a etiqueta especificada não estiver presente, qualquer que seja o valor.

O exemplo a seguir demonstra uma política que, quando anexada a uma identidade, nega à identidade a ação `ModifyInstanceGroups` de qualquer cluster com a etiqueta `department` definida como `dev`. Essa instrução só é efetiva por causa da instrução de negação que usa a condição `StringNotLike` para negar a ação, a menos que a etiqueta `department` esteja presente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        },
```

```
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Condition": {
                "StringEquals": {
                    "aws:ResourceTag/department": "dev"
                }
            },
            "Effect": "Deny",
            "Resource": "*"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Condition": {
                "StringNotLike": {
                    "aws:ResourceTag/department": "?*"
                }
            },
            "Effect": "Deny",
            "Resource": "*"
        }
    ],
}
```

Solução de problemas de identidade e acesso da Amazon EMR

Use as informações a seguir para ajudar a diagnosticar e corrigir problemas comuns que você pode encontrar ao trabalhar com o Amazon EMR e o IAM.

Tópicos

- Não tenho autorização para executar uma ação no Amazon EMR
- Não estou autorizado a realizar iam: PassRole
- Quero permitir que pessoas fora da minha AWS conta acessem meus recursos do Amazon EMR

Não tenho autorização para executar uma ação no Amazon EMR

Se isso AWS Management Console indicar que você não está autorizado a realizar uma ação, entre em contato com o administrador para obter ajuda. O administrador é a pessoa que forneceu o seu nome de usuário e senha.

O erro do exemplo a seguir ocorre quando o usuário `mateojackson` tenta usar o console para visualizar detalhes sobre um recurso do *`my-example-widget`* fictício, mas não tem as permissões fictícias do `EMR:GetWidget`.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 EMR:GetWidget on resource: my-example-widget
```

Neste caso, Mateo pede ao administrador para atualizar suas políticas e permitir o acesso ao recurso *`my-example-widget`* usando a ação `EMR:GetWidget`.

Não estou autorizado a realizar iam: PassRole

Caso receba uma mensagem de erro informando que você não tem autorização para executar a ação `iam:PassRole`, as políticas deverão ser atualizadas para permitir a transmissão de um perfil ao Amazon EMR.

Alguns Serviços da AWS permitem que você passe uma função existente para esse serviço em vez de criar uma nova função de serviço ou uma função vinculada ao serviço. Para fazer isso, é preciso ter permissões para passar o perfil para o serviço.

O erro exemplificado a seguir ocorre quando uma usuária do IAM chamada `marymajor` tenta usar o console para executar uma ação no Amazon EMR. No entanto, a ação exige que o serviço tenha permissões concedidas por um perfil de serviço. Mary não tem permissões para passar o perfil para o serviço.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

Nesse caso, as políticas de Mary devem ser atualizadas para permitir que ela realize a ação `iam:PassRole`.

Se precisar de ajuda, entre em contato com seu AWS administrador. Seu administrador é a pessoa que forneceu suas credenciais de login.

Quero permitir que pessoas fora da minha AWS conta acessem meus recursos do Amazon EMR

Você pode criar uma função que os usuários de outras contas ou pessoas fora da sua organização possam usar para acessar seus recursos. Você pode especificar quem é confiável para assumir o perfil. Para serviços que oferecem suporte a políticas baseadas em recursos ou listas de controle de acesso (ACLs), você pode usar políticas para conceder às pessoas acesso aos seus recursos.

Para saber mais, consulte:

- Para saber se o Amazon EMR é compatível com esses recursos, consulte Como o Amazon EMR funciona com o IAM.
- Para saber como fornecer acesso aos seus recursos em todos os Contas da AWS que você possui, consulte Como fornecer acesso a um usuário do IAM em outro Conta da AWS que você possui no Guia do usuário do IAM.
- Para saber como fornecer acesso aos seus recursos a terceiros Contas da AWS, consulte Como fornecer acesso Contas da AWS a terceiros no Guia do usuário do IAM.
- Saiba como conceder acesso por meio da federação de identidades consultando Concedendo Acesso a Usuários Autenticados Externamente (Federação de Identidades) no Guia do Usuário do IAM.
- Para saber a diferença entre usar perfis e políticas baseadas em recursos para acesso entre contas, consulte Como os perfis do IAM diferem de políticas baseadas em recursos no Guia do usuário do IAM.

Uso do Amazon S3 Access Grants com o Amazon EMR

Visão geral do S3 Access Grants para o Amazon EMR

Com as versões 6.15.0 e superiores do Amazon EMR, o Amazon S3 Access Grants fornece uma solução de controle de acesso escalável que você pode usar para aumentar o acesso aos dados do Amazon S3 por meio do Amazon EMR. Se você tiver uma configuração de permissão complexa ou grande para os dados do S3, poderá usar a funcionalidade Access Grants para escalar as permissões de dados do S3 para usuários, perfis e aplicações no seu cluster.

Use o S3 Access Grants para aumentar o acesso aos dados do Amazon S3 além das permissões concedidas pelo perfil de runtime ou pelos perfis do IAM anexados às identidades com acesso ao seu cluster do EMR. Para obter mais informações, consulte Gerenciar o acesso com o S3 Access Grants no Guia do usuário do Amazon S3.

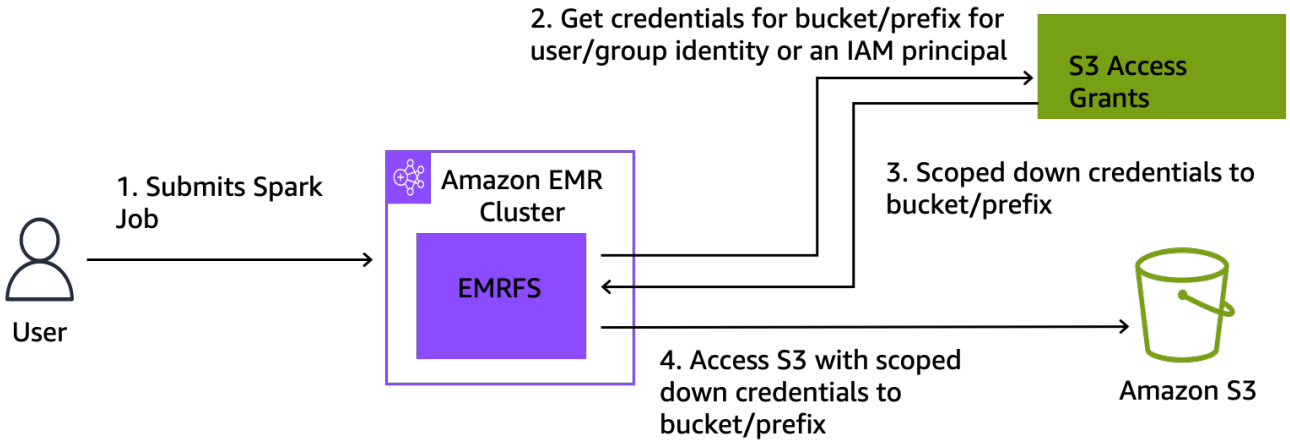
Para ver as etapas de uso do S3 Access Grants com outras implantações do Amazon EMR, consulte a seguinte documentação:

- Using S3 Access Grants with Amazon EMR on EKS
- Using S3 Access Grants with Amazon EMR Serverless

Como o Amazon EMR funciona com o S3 Access Grants

O Amazon EMR 6.15.0 e versões superiores oferecem uma integração nativa com o S3 Access Grants. Você pode habilitar o S3 Access Grants no Amazon EMR e executar trabalhos do Spark. Quando um trabalho do Spark faz uma solicitação de dados do S3, o Amazon S3 fornece credenciais temporárias que têm como escopo o bucket, prefixo ou objeto específico.

Veja a seguir uma visão geral de alto nível sobre como o Amazon EMR obtém acesso aos dados protegidos pela funcionalidade Access Grants do S3.



1. Um usuário envia um trabalho do Spark do Amazon EMR que usa dados armazenados no Amazon S3.
2. O Amazon EMR solicita ao S3 Access Grants que permita o acesso ao bucket, prefixo ou objeto em nome desse usuário.
3. O Amazon S3 retorna credenciais temporárias na forma de um token AWS Security Token Service (STS) para o usuário. O escopo do token é acessar o bucket, prefixo ou objeto do S3.
4. O Amazon EMR usa o token do STS para recuperar dados do S3.
5. O Amazon EMR recebe os dados do S3 e retorna os resultados ao usuário.

Considerações sobre o S3 Access Grants com o Amazon EMR

Observe os comportamentos e as limitações a seguir ao usar o S3 Access Grants com o Amazon EMR.

Suporte a recursos

- O S3 Access Grants é compatível com as versões 6.15.0 e superiores do Amazon EMR.
- O Spark é o único mecanismo de consulta compatível ao usar o S3 Access Grants com o Amazon EMR.
- Delta Lake e Hudi são os únicos formatos de tabela aberta compatíveis ao usar o S3 Access Grants com o Amazon EMR.
- Os seguintes recursos do Amazon EMR não são compatíveis com o S3 Access Grants:
  - Tabelas Apache Iceberg
  - Autenticação nativa LDAP
  - Autenticação nativa do Apache Ranger
  - AWS CLI solicitações para o Amazon S3 que usam funções do IAM
  - Acesso ao S3 por meio do protocolo de código aberto do S3A
- A opção `fallbackToIAM` não é compatível com clusters do EMR que usam a propagação de identidade confiável com o Centro de Identidade do IAM.
- O S3 Access Grants com o AWS Lake Formation só é compatível com clusters do Amazon EMR executados no Amazon EC2.

Considerações comportamentais

- A integração nativa do Apache Ranger com o Amazon EMR possui funcionalidade congruente com o S3 Access Grants como parte do plug-in EMRFS S3 do Apache Ranger. Se você usa o Apache Ranger para controle de acesso refinado (FGAC), recomendamos usar esse plug-in em vez do S3 Access Grants.
- O Amazon EMR fornece um cache de credenciais no EMRFS para garantir que o usuário não precise fazer solicitações repetidas das mesmas credenciais em um trabalho do Spark. Portanto, o Amazon EMR sempre solicita o privilégio de nível padrão quando solicita credenciais. Para obter mais informações, consulte Solicitação de acesso aos dados do S3 no Guia do usuário do Amazon S3.

- No caso de um usuário realizar uma ação que não tenha suporte do S3 Access Grants, o Amazon EMR está configurado para usar o perfil do IAM que foi especificado para a execução do trabalho. Para ter mais informações, consulte Fallback para os perfis do IAM.

Launch an Amazon EMR cluster with S3 Access Grants

Esta seção descreve como iniciar um cluster do EMR que é executado no Amazon EC2 e usa o S3 Access Grants para gerenciar o acesso aos dados no Amazon S3. Para ver as etapas de uso do S3 Access Grants com outras implantações do Amazon EMR, consulte a seguinte documentação:

- Using S3 Access Grants with Amazon EMR on EKS
- Using S3 Access Grants with EMR Serverless

Use as etapas a seguir para iniciar um cluster do EMR que é executado no Amazon EC2 e usa o S3 Access Grants para gerenciar o acesso aos dados no Amazon S3.

1. Configure um perfil de execução de trabalhos para o cluster do EMR. Inclua as permissões do IAM necessárias para executar os trabalhos do Spark, `s3:GetDataAccess` e `s3:GetAccessGrantsInstanceForPrefix`:

```
{
    "Effect": "Allow",
    "Action": [
    "s3:GetDataAccess",
    "s3:GetAccessGrantsInstanceForPrefix"
    ],
    "Resource": [     //LIST ALL INSTANCE ARNS THAT THE ROLE IS ALLOWED TO QUERY
         "arn:aws_partition:s3:Region:account-id1:access-grants/default",
         "arn:aws_partition:s3:Region:account-id2:access-grants/default"
    ]
}
```

> **Note**
>
> Com o Amazon EMR, a funcionalidade S3 Access Grants aumenta as permissões definidas nos perfis do IAM. Se os perfis do IAM especificados para a execução do

trabalho contiverem permissões de acesso direto ao S3, os usuários poderão acessar mais dados do que os definidos por você no S3 Access Grants.

2. Em seguida, use o AWS CLI para criar um cluster com o Amazon EMR 6.15 ou superior e a `emrfs-site` classificação para habilitar o S3 Access Grants, semelhante ao exemplo a seguir:

```
aws emr create-cluster
  --release-label emr-6.15.0 \
  --instance-count 3 \
  --instance-type m5.xlarge \
  --configurations '[{"Classification":"emrfs-site",
 "Properties":{"fs.s3.s3AccessGrants.enabled":"true",
 "fs.s3.s3AccessGrants.fallbackToIAM":"false"}}]'
```

Concessões de acesso ao S3 com AWS Lake Formation

Se você usa o Amazon EMR com a integração do AWS Lake Formation, poderá utilizar o Amazon S3 Access Grants para acesso direto ou tabular aos dados no Amazon S3.

> **Note**
>
> O S3 Access Grants with só AWS Lake Formation é compatível com clusters do Amazon EMR que são executados no Amazon EC2.

Acesso direto

O acesso direto envolve todas as chamadas para acessar dados do S3 que não invocam a API do serviço AWS Glue que a Lake Formation usa como metastore com o Amazon EMR, por exemplo, para chamar: `spark.read`

```
spark.read.csv("s3://...")
```

Quando você usa o S3 Access Grants AWS Lake Formation no Amazon EMR, todos os padrões de acesso direto passam pelo S3 Access Grants para obter credenciais temporárias do S3.

Acesso tabular

O acesso tabular ocorre quando o Lake Formation invoca a API do metastore para acessar sua localização no S3, por exemplo, para consultar dados da tabela:

```
spark.sql("select * from test_tbl")
```

Quando você usa o S3 Access Grants AWS Lake Formation no Amazon EMR, todos os padrões de acesso tabulares passam pelo Lake Formation.

Fallback para os perfis do IAM

Se um usuário tentar realizar uma ação que não tenha o suporte do S3 Access Grants, o Amazon EMR usa como padrão o perfil do IAM que foi especificado para a execução do trabalho quando a configuração `fallbackToIAM` for `true`. Isso permite que os usuários efetuem o fallback do perfil de execução de trabalhos para fornecer credenciais de acesso ao S3 em cenários não cobertos pelo S3 Access Grants.

Com a opção `fallbackToIAM` habilitada, os usuários podem acessar os dados que o Access Grant permite. Se não houver um token do S3 Access Grants para os dados de destino, o Amazon EMR verifica a permissão no perfil de execução de trabalhos.

> **Note**
>
> Recomendamos que você teste suas permissões de acesso com a configuração `fallbackToIAM` habilitada, mesmo que planeje desabilitar a opção para workloads de produção. Com os trabalhos do Spark, há outras maneiras pelas quais os usuários podem acessar todos os conjuntos de permissões com as credenciais do IAM. Quando habilitadas em clusters do EMR, as concessões do S3 dão aos trabalhos do Spark acesso às localizações do S3. Você deve garantir a proteção dessas localizações do S3 contra o acesso fora do EMRFS. Por exemplo, você deve proteger as localizações do S3 contra o acesso de clientes do S3 usados em notebooks ou por aplicações sem o suporte do S3 Access Grants, como Hive ou Presto.

Autenticação em nós de cluster do Amazon EMR

Os clientes SSH podem usar um par de chaves do Amazon EC2 para autenticar-se em instâncias de cluster. Como alternativa, com o Amazon EMR 5.10.0 ou versões posteriores, você pode configurar o Kerberos para autenticar usuários e conexões SSH para o nó primário. E com o Amazon EMR 5.12.0 e versões posteriores, você pode se autenticar com o LDAP.

Tópicos

Usar um par de chaves do EC2 para credenciais SSH

Os nós de cluster do Amazon EMR são executados em instâncias do Amazon EC2. É possível se conectar a nós de cluster da mesma forma que você se conecta a instâncias do Amazon EC2. Você pode usar o Amazon EC2 para criar um par de chaves ou você pode importar um par de chaves. Ao criar um cluster, é possível especificar o par de chaves do Amazon EC2 que será usado nas conexões SSH para todas as instâncias de cluster. Também é possível criar um cluster sem par de chaves. Isso geralmente é feito com clusters transitórios que são iniciados, executam etapas e são encerrados automaticamente.

O cliente SSH que você usa para se conectar ao cluster precisa usar o arquivo de chave privada associado ao par de chaves. Esse é um arquivo .pem para clientes SSH que usam Linux, Unix e macOS. Você deve definir permissões para que apenas o proprietário da chave tenha permissão para acessar o arquivo. É um arquivo .ppk para clientes SSH que usam o Windows, e o arquivo .ppk geralmente é criado a partir do arquivo .pem.

- Para obter mais informações sobre a criação de um par de chaves do Amazon EC2, consulte os pares de chaves do Amazon EC2 no Guia do usuário do Amazon EC2.
- Para obter instruções sobre como usar o PuTTYgen para criar um arquivo.ppk a partir de um arquivo.pem, consulte Convertendo sua chave privada usando o PuTTYgen no Guia do usuário do Amazon EC2.
- Para obter mais informações sobre como definir permissões de arquivos.pem e como se conectar ao nó primário de um cluster do EMR usando métodos diferentes, inclusive do `ssh` Linux ou macOS, do PuTTY do Windows ou AWS CLI de qualquer sistema operacional compatível, consulte. Conectar-se ao nó primário usando SSH

Usar o Kerberos para autenticação com o Amazon EMR

O Amazon EMR 5.10.0 e versões posteriores oferecem suporte ao Kerberos. O Kerberos é um protocolo de autenticação de rede que usa criptografia segredo-chave para fornecer autenticação

forte, de maneira que senhas ou outras credenciais não sejam enviadas pela rede em um formato não criptografado.

No Kerberos, os serviços e os usuários que precisam se autenticar são conhecidos como entidades principais. As entidades principais existem em um realm Kerberos. Dentro do realm, um servidor Kerberos conhecido como Key Distribution Center (KDC) oferece as entidades principais se autenticarem. O KDC faz isso emitindo tíquetes para autenticação. O KDC mantém um banco de dados das entidades principais no realm, as senhas e outras informações administrativas sobre cada entidade principal. Um KDC também pode aceitar credenciais de autenticação de entidades principais em outros realms, o que é conhecido como uma relação de confiança entre realms. Além disso, um cluster do EMR pode usar um KDC externo para autenticar principais.

Um cenário comum para estabelecer uma relação de confiança entre realms ou usar um KDC externo é autenticar usuários de um domínio do Active Directory. Isso permite que os usuários acessem um cluster do EMR com a conta do domínio quando eles usarem o SSH para se conectar a um cluster ou trabalhar com aplicações de big data.

Quando você usa autenticação do Kerberos, o Amazon EMR configura o Kerberos para as aplicações, os componentes e os subsistemas que ele instala no cluster, de maneira que eles sejam autenticados entre si.

> **Important**
>
> O Amazon EMR não oferece suporte AWS Directory Service for Microsoft Active Directory em uma relação de confiança entre regiões ou como um KDC externo.

Antes de configurar o Kerberos usando o Amazon EMR, recomendamos que você se familiarize com os conceitos do Kerberos, os serviços executados em um KDC e as ferramentas para administrar serviços do Kerberos. Para obter mais informações, consulte a documentação do MIT Kerberos, publicada pelo Kerberos Consortium.

Tópicos

- Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory

Aplicações compatíveis

Em um cluster do EMR, as entidades principais do Kerberos são os serviços de aplicações de big data e os subsistemas executados em todos os nós de cluster. O Amazon EMR pode configurar as aplicações e os componentes listados abaixo para usar o Kerberos. Cada aplicativo tem uma entidade principal de usuário Kerberos associada.

O Amazon EMR não oferece suporte a relações de confiança entre realms com AWS Directory Service for Microsoft Active Directory.

O Amazon EMR somente configura os recursos de autenticação do Kerberos de código aberto para as aplicações e os componentes listados abaixo. Todos os outros aplicativos instalados não são Kerberizados, o que pode resultar em uma incapacidade de se comunicar com componentes Kerberizados e causar erros de aplicativo. Os aplicativos e os componentes não Kerberizados não têm autenticação ativada. As aplicações e componentes compatíveis podem variar conforme as diferentes versões do Amazon EMR.

A interface de usuário do Livy é a única interface de usuário da Web hospedada no cluster que é kerberizada.

- Hadoop MapReduce
- Hbase
- HCatalog
- HDFS
- Hive
  - Não habilite o Hive com autenticação LDAP. Isso pode causar problemas na comunicação com o YARN Kerberizado.
- Hue
  - A autenticação de usuário do Hue não é definida automaticamente e pode ser configurada usando-se a API de configuração.
  - O servidor do Hue é Kerberizado. O front-end (UI) do Hue não está configurado para autenticação. A autenticação LDAP pode ser configurada para o Hue UI.
- Livy

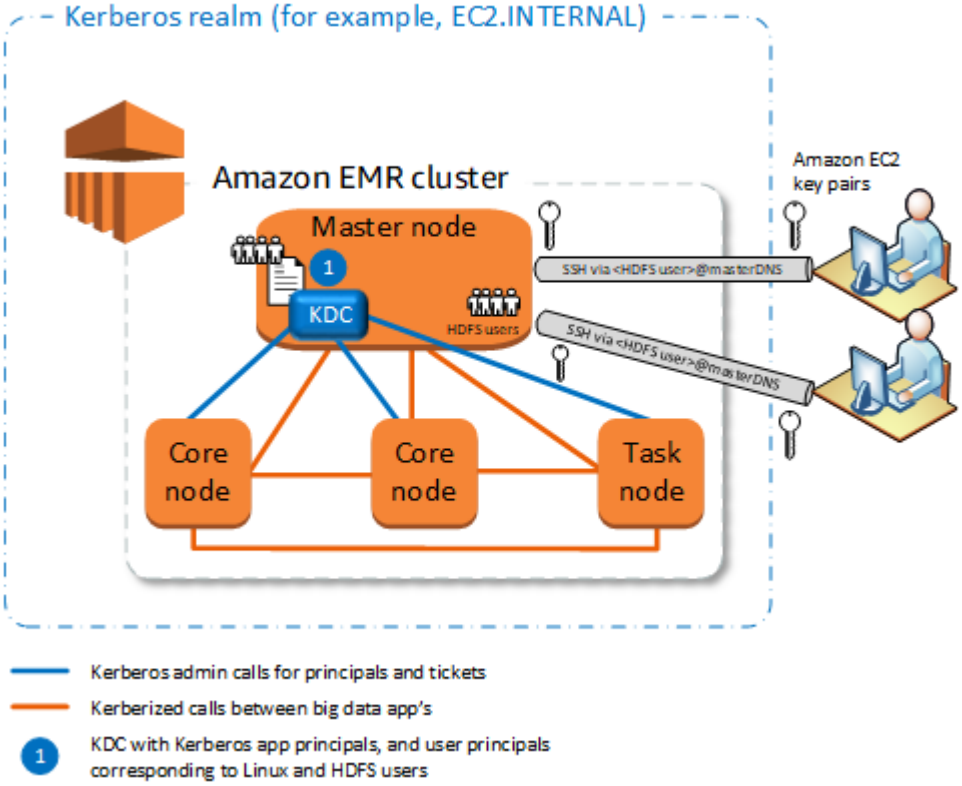
  - A representação do Livy com clusters kerberizados é compatível com as versões 5.22.0 e posteriores do Amazon EMR.
- Oozie
- Phoenix
- Presto
  - O Presto oferece suporte à autenticação Kerberos no Amazon EMR 6.9.0 e versões posteriores.
  - Para usar a autenticação Kerberos com o Presto, é necessário habilitar a criptografia em trânsito.
- Spark
- Tez
- Trino
  - O Trino oferece suporte à autenticação Kerberos no Amazon EMR 6.11.0 e versões posteriores.
  - Para usar a autenticação Kerberos com o Trino, é necessário habilitar a criptografia em trânsito.
- YARN
- Zeppelin
  - O Zeppelin é configurado somente para usar o Kerberos com o intérprete do Spark. Ele não é configurado para outros intérpretes.
  - A representação de usuário não oferece suporte a intérpretes kerberizados do Zeppelin além do Spark.
- Zookeeper
  - O cliente do Zookeeper não é compatível.

Opções de arquitetura do Kerberos

Ao usar o Kerberos com o Amazon EMR, você pode escolher entre as arquiteturas listadas nesta seção. Independentemente da arquitetura que escolhida, você pode configurar o Kerberos usando as mesmas etapas. Você cria uma configuração de segurança, especifica a configuração de segurança do Kerberos e as opções específicas do cluster compatíveis ao criar o cluster, e você cria diretórios do HDFS para usuários do Linux no cluster que correspondam aos usuários principais no KDC. Para obter uma explicação sobre as opções de configuração e configurações de exemplo para cada arquitetura, consulte Configurar o Kerberos no Amazon EMR.

KDC dedicado ao cluster (KDC no nó primário)

Essa configuração está disponível no Amazon EMR 5.10.0 e versões posteriores.



Vantagens

- O Amazon EMR tem total propriedade do KDC.
- O KDC no cluster do EMR é independente das implementações centralizadas do KDC, como o Microsoft Active Directory ou o AWS Managed Microsoft AD.
- O impacto no desempenho é mínimo porque o KDC gerencia a autenticação somente para nós locais no cluster.
- Opcionalmente, outros clusters Kerberizados pode fazer referência ao KDC como um KDC externo. Para ter mais informações, consulte KDC externo: nó primário em um cluster diferente.

Considerações e limitações

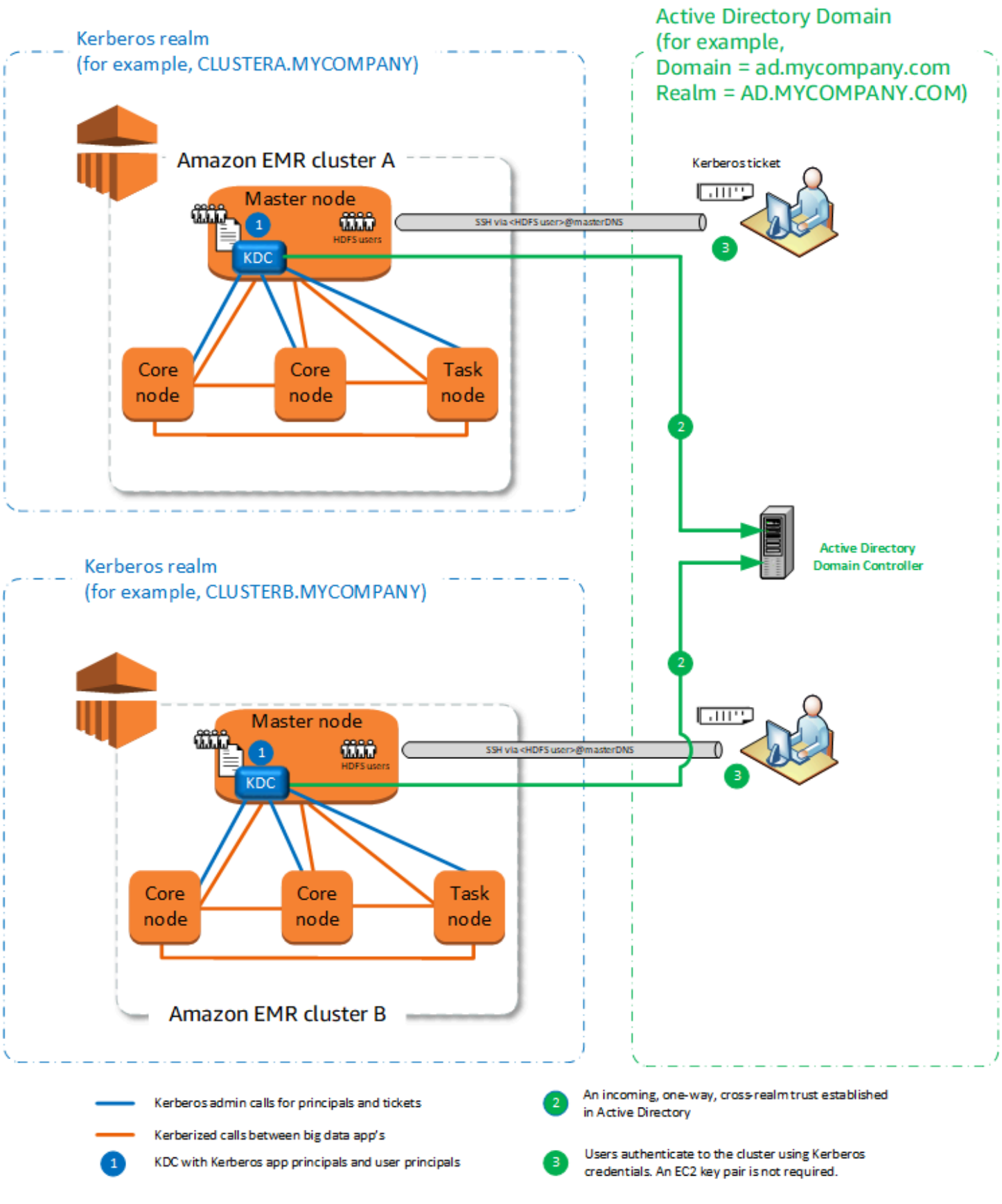
- Os clusters Kerberizados não podem autenticar uns ao outros, portanto, os aplicativos não podem interoperar. Se os aplicativos de cluster precisarem interoperar, você deverá estabelecer uma relação de confiança entre realms entre os clusters, ou configurar um cluster como o KDC externo

para outros clusters. Se uma relação de confiança entre realms for estabelecida, os KDCs deverão ter diferentes realms do Kerberos.

- Você deve criar usuários do Linux na instância do EC2 do nó primário que correspondam aos usuários principais do KDC, juntamente com os diretórios do HDFS para cada usuário.
- Os usuários principais devem usar um arquivo de chave privada do EC2 e credenciais `kinit` para se conectar ao cluster usando SSH.

Relação de confiança entre realms

Nessa configuração, os principais (normalmente usuários) de um realm do Kerberos diferente autenticam componentes do aplicativo em um cluster do EMR Kerberizado, que tem seu próprio KDC. O KDC no nó primário estabelece uma relação de confiança com outro KDC usando uma entidade principal entre realms existente em ambos os KDCs. O nome do principal e a senha coincidem precisamente em cada KDC. Relações de confianças entre realms são mais comuns com implementações do Active Directory, conforme mostrado no diagrama a seguir. Relações de confiança entre realms com um MIT KDC externo ou um KDC em outro cluster do Amazon EMR também são compatíveis.


Active Directory Domain
(for example,
Domain = ad.mycompany.com
Realm = AD.MYCOMPANY.COM)
Kerberos realm
(for example, CLUSTERA.MYCOMPANY)
Amazon EMR cluster A
Master node
1
HDFS users
KDC
SSH via <HDFS user>@masterDNS
Kerberos ticket
3
Core node
Core node
Task node
2
Active Directory Domain Controller
Kerberos realm
(for example, CLUSTERB.MYCOMPANY)
Master node
1
HDFS users
KDC
SSH via <HDFS user>@masterDNS
2
3
Core node
Core node
Task node
Amazon EMR cluster B
Kerberos admin calls for principals and tickets
Kerberized calls between big data app's
1
KDC with Kerberos app principals and user principals
2
An incoming, one-way, cross-realm trust established in Active Directory
3
Users authenticate to the cluster using Kerberos credentials. An EC2 key pair is not required.

Vantagens

- O cluster do EMR no qual o KDC está instalado mantém a total propriedade do KDC.
- Com o Active Directory, o Amazon EMR cria automaticamente usuários do Linux que correspondam aos usuários principais do KDC. Ainda assim é necessário criar diretórios do HDFS para cada usuário. Além disso, os usuário principais no domínio do Active Directory podem acessar clusters Kerberizados usando credenciais `kinit`, sem o arquivo de chave privada do EC2. Isso elimina a necessidade de compartilhar o arquivo de chave privada entre os usuários do cluster.
- Como cada cluster do KDC gerencia a autenticação para os nós no cluster, os efeitos da latência da rede e da sobrecarga de processamento para um grande número de nós nos clusters é minimizado.

Considerações e limitações

- Se estiver estabelecendo uma relação de confiança com um domínio do Active Directory, você deverá fornecer um nome de usuário e senha do Active Directory com permissões para se juntar aos principais do domínio ao criar o cluster.
- As relações de confiança entre realms não podem ser estabelecidas entre realms do Kerberos com o mesmo nome.
- As relações de confiança entre realms deve ser estabelecidas explicitamente. Por exemplo, se o Cluster A e o Cluster B estabelecerem uma relação de confiança entre realms com um KDC, eles não confiarão inerentemente um no outro e seus aplicativos não poderão se autenticar entre si para interoperar.
- Os KDCs deve ser mantidos de forma independente e coordenada para que as credenciais dos usuários principais correspondam exatamente.

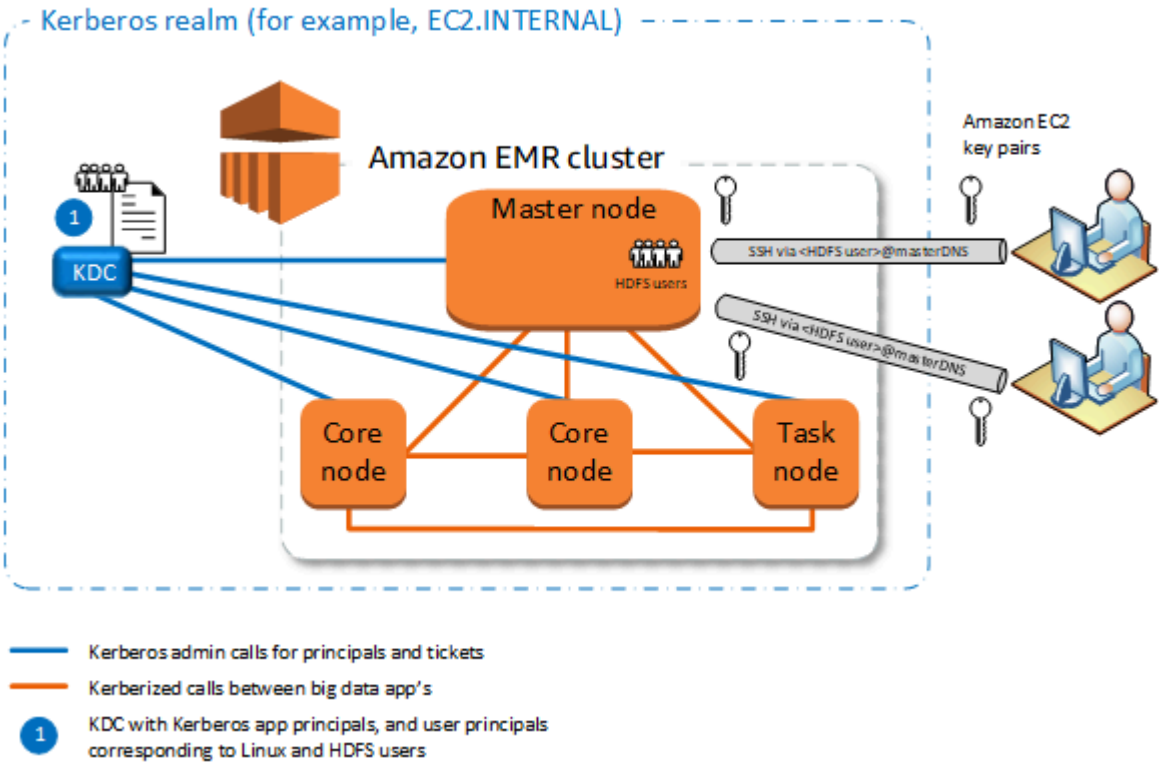
KDC externo

Configurações com um KDC externo são compatíveis com o Amazon EMR 5.20.0 e posteriores.

- KDC externo: MIT KDC
- KDC externo: nó primário em um cluster diferente
- KDC externo: KDC do cluster em um cluster diferente com relação de confiança entre realms do Active Directory

KDC externo: MIT KDC

Essa configuração permite que um ou mais clusters do EMR usem principais definidos e mantidos em um servidor KDC MIT.



Vantagens

- O gerenciamento de principais é consolidado em um único KDC.
- Vários clusters podem usar o mesmo KDC no mesmo realm do Kerberos. Para ter mais informações, consulte Requisitos para usar múltiplos clusters com o mesmo KDC.
- O nó primário em um cluster kerberizado não tem o ônus da performance associada com a manutenção do KDC.

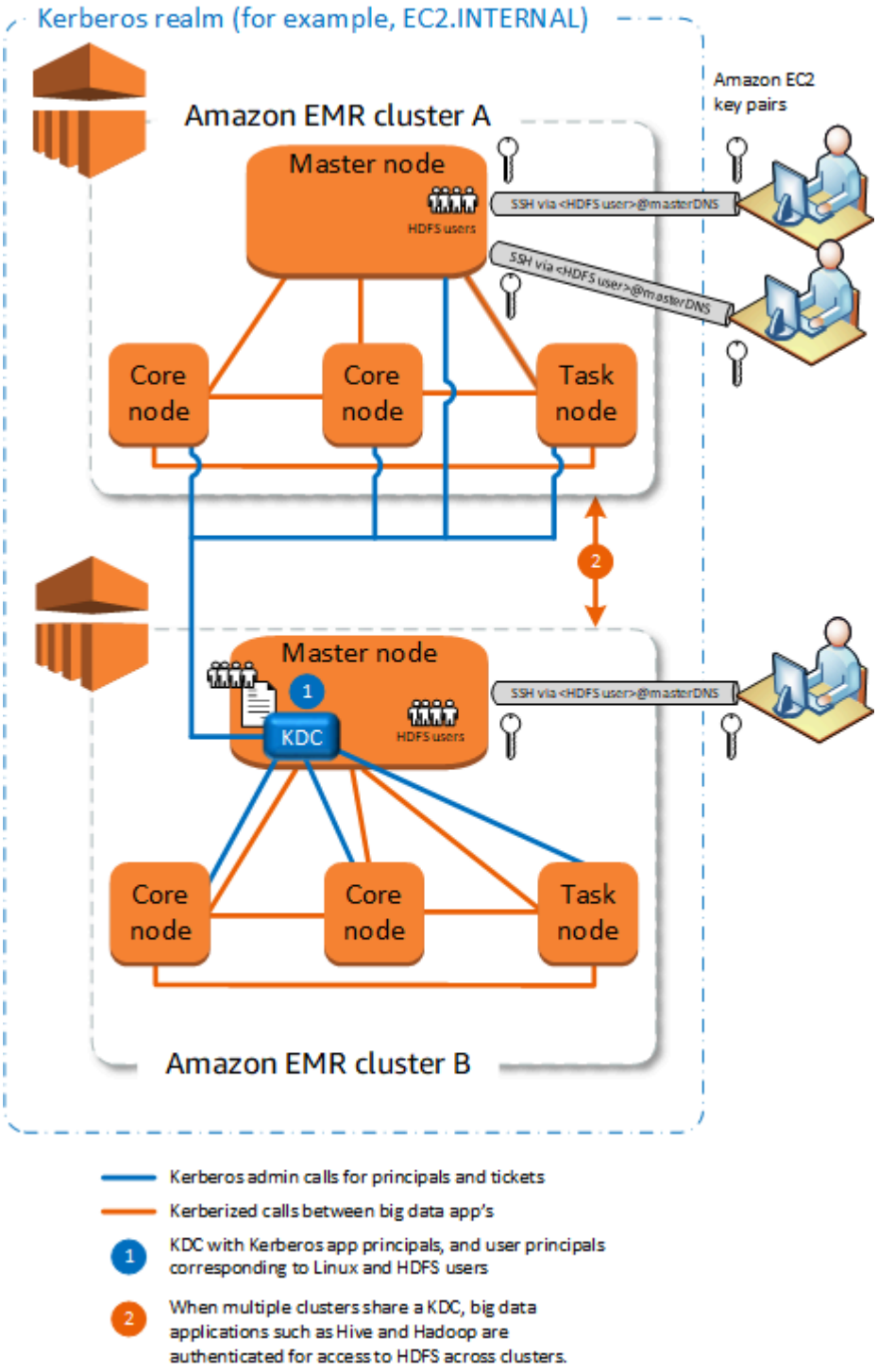
Considerações e limitações

- Você deve criar usuários do Linux na instância do EC2 de cada nó primário do cluster kerberizado que corresponda às entidades principais de usuário do KDC, juntamente com os diretórios do HDFS para cada usuário.
- Os usuários principais devem usar um arquivo de chave privada do EC2 e credenciais `kinit` para se conectar aos clusters Kerberizados usando SSH.

- Cada nó nos clusters do EMR Kerberizados deve ter uma rota de rede para o KDC.
- Cada nó nos clusters Kerberizados coloca uma carga de autenticação no KDC externo, portanto, a configuração do KDC afeta o desempenho do cluster. Ao configurar o hardware do servidor KDC, considere o suporte simultâneo ao número máximo de nós do Amazon EMR.
- O desempenho do cluster depende da latência da rede entre os nós nos clusters Kerberizados e no KDC.
- A solução de problemas pode ser mais difícil devido a interdependências.

KDC externo: nó primário em um cluster diferente

Essa configuração é quase idêntica à implementação do MIT KDC externo acima, porém o KDC está no nó primário de um cluster do EMR. Para obter mais informações, consulte KDC dedicado ao cluster (KDC no nó primário) e Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.



Vantagens

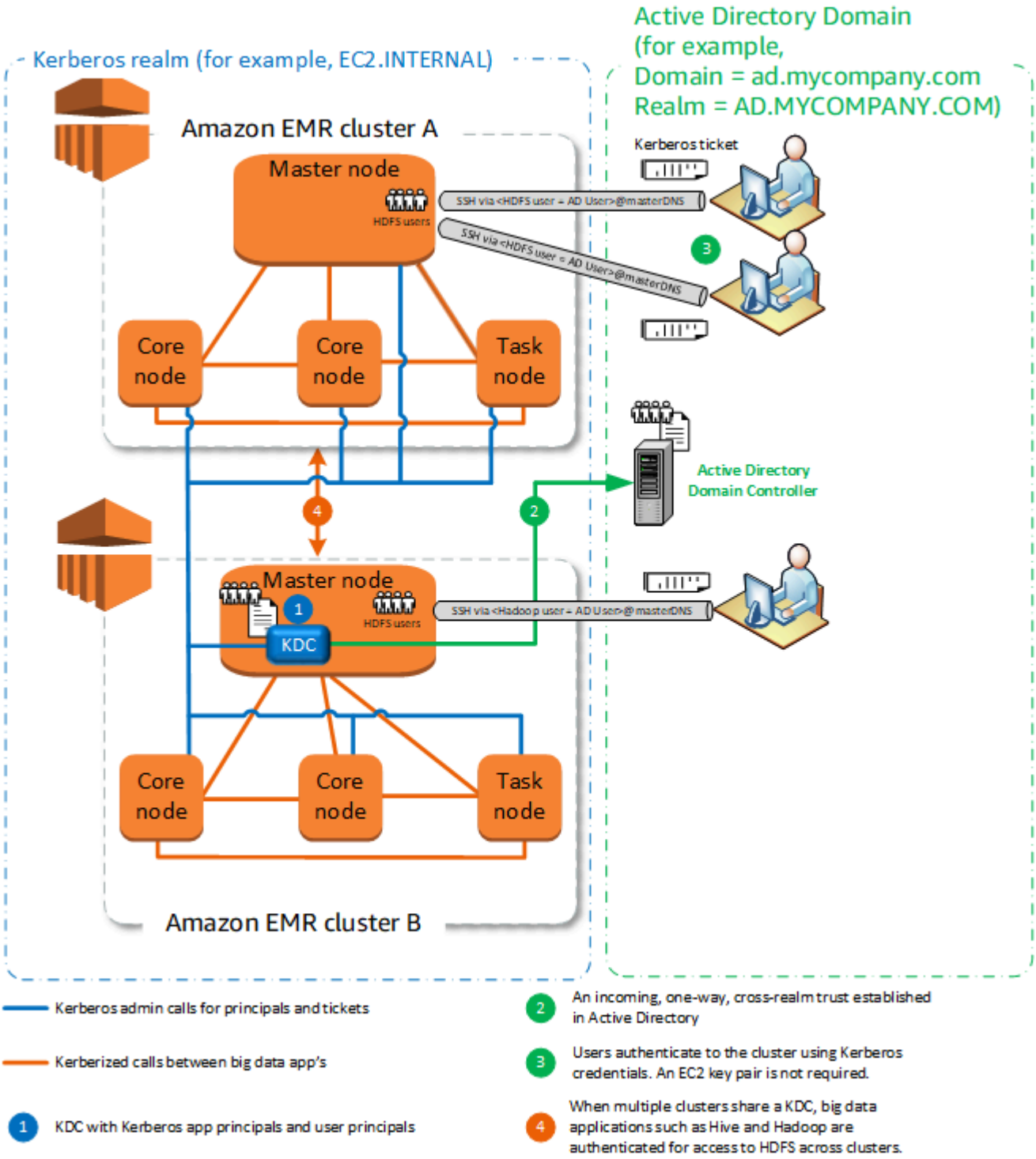
- O gerenciamento de principais é consolidado em um único KDC.
- Vários clusters podem usar o mesmo KDC no mesmo realm do Kerberos. Para ter mais informações, consulte Requisitos para usar múltiplos clusters com o mesmo KDC.

Considerações e limitações

- Você deve criar usuários do Linux na instância do EC2 de cada nó primário do cluster kerberizado que corresponda às entidades principais de usuário do KDC, juntamente com os diretórios do HDFS para cada usuário.
- Os usuários principais devem usar um arquivo de chave privada do EC2 e credenciais `kinit` para se conectar aos clusters Kerberizados usando SSH.
- Cada nó nos clusters do EMR deve ter uma rota de rede para o KDC.
- Cada nó do Amazon EMR nos clusters kerberizados coloca uma carga de autenticação no KDC externo, portanto, a configuração do KDC afeta a performance do cluster. Ao configurar o hardware do servidor KDC, considere o suporte simultâneo ao número máximo de nós do Amazon EMR.
- O desempenho do cluster depende da latência da rede entre os nós nos clusters e no KDC.
- A solução de problemas pode ser mais difícil devido a interdependências.

KDC externo: KDC do cluster em um cluster diferente com relação de confiança entre realms do Active Directory

Nessa configuração, você primeiro cria um cluster com um KDC dedicado ao cluster que tenha uma relação de confiança entre realms unidirecional com o Active Directory. Para ver um tutorial detalhado, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory. Em seguida, inicie clusters adicionais, fazendo referência ao KDC do cluster que tem a confiança como um KDC externo. Para ver um exemplo, consulte KDC externo do cluster com relação de confiança entre realms do Active Directory. Isso permite que cada cluster do Amazon EMR que usa o KDC externo autentique as entidades principais definidas e mantidas em um domínio do Microsoft Active Directory.



Vantagens

- O gerenciamento de principais é consolidado no domínio do Active Directory.

- O Amazon EMR ingressa no realm do Active Directory, o que elimina a necessidade de criar usuários do Linux que correspondam aos usuários do Active Directory. Ainda assim é necessário criar diretórios do HDFS para cada usuário.
- Vários clusters podem usar o mesmo KDC no mesmo realm do Kerberos. Para ter mais informações, consulte Requisitos para usar múltiplos clusters com o mesmo KDC.
- Os usuário principais no domínio do Active Directory podem acessar clusters Kerberizados usando credenciais `kinit`, sem o arquivo de chave privada do EC2. Isso elimina a necessidade de compartilhar o arquivo de chave privada entre os usuários do cluster.
- Somente um nó primário do Amazon EMR tem a carga para manter o KDC, e somente esse cluster deve ser criado com as credenciais do Active Directory para a relação de confiança entre realms entre o KDC e o Active Directory.

Considerações e limitações

- Cada nó nos clusters do EMR deve ter uma rota de rede para o KDC e o controlador de domínio do Active Directory.
- Cada nó do Amazon EMR coloca uma carga de autenticação no KDC externo, portanto, a configuração do KDC afeta a performance do cluster. Ao configurar o hardware do servidor KDC, considere o suporte simultâneo ao número máximo de nós do Amazon EMR.
- O desempenho do cluster depende da latência da rede entre os nós nos clusters e no servidor KDC.
- A solução de problemas pode ser mais difícil devido a interdependências.

Requisitos para usar múltiplos clusters com o mesmo KDC

Vários clusters podem usar o mesmo KDC no mesmo realm do Kerberos. No entanto, se os clusters forem executados simultaneamente, eles poderão falhar se usarem ServicePrincipal nomes Kerberos conflitantes.

Se você tiver múltiplos clusters simultâneos com o mesmo KDC externo, certifique-se de que os clusters usem regiões Kerberos diferentes. Se os clusters precisarem usar o mesmo realm do Kerberos, certifique-se de que os clusters estejam em sub-redes diferentes e que os intervalos de CIDR não se sobreponham.

Configurar o Kerberos no Amazon EMR

Esta seção fornece detalhes da configuração e exemplos para configurar o Kerberos com arquiteturas comuns. Independentemente da arquitetura escolhida, as noções básicas de configuração são as mesmas e a configuração é feita em três etapas. Se usar um KDC externo ou configurar uma relação de confiança entre realms, você deverá garantir que cada nó em um cluster tenha uma rota de rede para o KDC externo, incluindo a configuração aplicável de grupos de segurança para permitir o tráfego de entrada e saída do Kerberos.

Etapa 1: Criar uma configuração de segurança com propriedades do Kerberos

A configuração de segurança especifica detalhes sobre o KDC do Kerberos e permite que a configuração do Kerberos seja reutilizada cada vez que você criar um cluster. Você pode criar uma configuração de segurança usando o console do Amazon EMR AWS CLI, o ou a API do EMR. A configuração de segurança também pode conter outras opções de segurança, como criptografia. Para obter mais informações sobre como criar configurações de segurança e especificar uma configuração de segurança ao criar um cluster, consulte Usar configurações de segurança para definir a segurança do cluster. Para obter informações sobre as propriedades do Kerberos em uma configuração de segurança, consulte Configurações do Kerberos para configurações de segurança.

Etapa 2: Criar um cluster e especificar os atributos do Kerberos específicos do cluster

Ao criar um cluster, você especifica uma configuração de segurança do Kerberos juntamente com e as opções do Kerberos específicas do cluster. Quando o console do Amazon EMR é usado, somente as opções do Kerberos compatíveis com a configuração de segurança especificada estão disponíveis. Ao usar a API AWS CLI ou o Amazon EMR, certifique-se de especificar as opções do Kerberos compatíveis com a configuração de segurança especificada. Por exemplo, se você especificar uma senha principal para uma relação de confiança entre realms ao criar um cluster usando a CLI e a configuração de segurança especificada não for configurada com os parâmetros da relação de confiança entre realms, ocorrerá um erro. Para ter mais informações, consulte Configurações do Kerberos para clusters.

Etapa 3: configurar o nó primário do cluster

Dependendo dos requisitos de sua arquitetura e implantação, configuração adicional no cluster pode ser necessária. Você pode fazer isso depois de criá-lo ou usando etapas ou ações de bootstrap durante o processo de criação.

Para cada usuário autenticado pelo Kerberos que se conecta ao cluster usando SSH, você deve garantir que as contas do Linux criadas correspondam ao usuário do Kerberos. Se as entidades

principais forem fornecidos por um controlador de domínio do Active Directory, como o KDC externo ou por meio de uma relação de confiança entre realms, o Amazon EMR criará contas de usuário do Linux automaticamente. Se o Active Directory não for usado, você deverá criar principais para cada usuário que correspondam ao usuário do Linux. Para ter mais informações, consulte Configurar um cluster para usuários do HDFS autenticados pelo Kerberos e conexões SSH.

Cada usuário também deve ter um diretório de usuário do HDFS que pertença a eles, que você deve criar. Além disso, o SSH deve ser configurado com GSSAPI habilitada para permitir conexões de usuários autenticados pelo Kerberos. A GSSAPI deve ser habilitada no nó primário, e a aplicação SSH cliente deve ser configurada para usar GSSAPI. Para ter mais informações, consulte Configurar um cluster para usuários do HDFS autenticados pelo Kerberos e conexões SSH.

Configuração de segurança e configurações do cluster para Kerberos no Amazon EMR

Ao criar um cluster Kerberizado, você especifica a configuração de segurança com atributos do Kerberos específicos do cluster. Você não pode especificar um conjunto sem o outro, ou ocorrerá um erro.

Este tópico fornece uma visão geral dos parâmetros de configuração disponíveis para o Kerberos quando você cria uma configuração de segurança e um cluster. Além disso, exemplos da CLI para criar configurações de segurança e clusters compatíveis são fornecidos para arquiteturas comuns.

Configurações do Kerberos para configurações de segurança

Você pode criar uma configuração de segurança que especifique os atributos do Kerberos usando o console do Amazon EMR, o AWS CLI ou a API do EMR. A configuração de segurança também pode conter outras opções de segurança, como criptografia. Para ter mais informações, consulte Criar uma configuração de segurança.

Use as referências a seguir para compreender as definições de configuração de segurança disponíveis para a arquitetura do Kerberos que você escolher. As configurações do console do Amazon EMR são exibidas. Para opções da CLI correspondentes, consulte Especificando as configurações do Kerberos usando o AWS CLI ou Exemplos de configuração.

Parâmetro	Descrição
Kerberos	Especifica que o Kerberos está habilitado em clusters que usam essa configuração de segurança. Ao usar essa configuração de segurança, o cluster também

Parâmetro		Descrição
		deverá ter configurações Kerberos especificadas ou ocorrerá um erro.
Provedor	KDC dedicado ao cluster	Especifica que o Amazon EMR criará um KDC no nó primário de qualquer cluster que usar essa configuração de segurança. Você especifica o nome do realm e a senha de administrador do KDC ao criar o cluster. Você pode referenciar esse KDC por outros clusters, se necessário. Crie esses clusters usando outra configuração de segurança, especifique um KDC externo e use o nome do território e a senha de administrador do KDC que você especificar para o KDC dedicado ao cluster.
	KDC externo	Disponível apenas no Amazon EMR 5.20.0 e versões posteriores. Especifica que os clusters que usam essa configuração de segurança autenticarão as entidades principais do Kerberos usando um servidor do KDC fora do cluster. O KDC não é criado no cluster. Ao criar o cluster, especifique o nome do realm e a senha de administrador do KDC para o KDC externo.
Vida útil do tíquete		Opcional. Especifica o período de validade de um tíquete do Kerberos emitido pelo KDC em clusters que usam essa configuração de segurança. Os ciclos de vida do tíquete são limitados por motivos de segurança. As aplicações e os serviços de cluster renovarão automaticamente os tíquetes quando perderem a validade. Os usuários que se conectam ao cluster via SSH usando credenciais do Kerberos precisam executar `kinit` pela linha de comando do nó primário para renovar um tíquete expirado.

Parâmetro		Descrição
Relação de confiança entre realms		Especifica uma relação de confiança entre regiões entre um KDC dedicado ao cluster em clusters que usam essa configuração de segurança e um KDC em outro realm do Kerberos. As entidades principais (normalmente usuários) de outro realm são autenticados em clusters que usam essa configuração. É necessário ter configuração adicional no outro realm do Kerberos. Para ter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.
Propriedades de confiança entre realms	Realm	Especifica o nome de realm Kerberos de outro realm na relação de confiança. Por convenção, os nomes de realm do Kerberos são iguais ao nome do domínio, mas em letras maiúsculas.
	Domínio	Especifica o nome de domínio de outro realm na relação de confiança.
	Servidor do administrador	Especifica o nome de domínio totalmente qualificado (FQDN) ou endereço IP do servidor de administrador no outro realm da relação de confiança. O servidor de administração e o servidor de KDC normalmente são executados na mesma máquina com o mesmo FQDN, mas se comunicam por diferentes portas. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com:749`).

Parâmetro		Descrição
	Servidor do KDC	Especifica o nome de domínio totalmente qualificado (FQDN) ou endereço IP do servidor do KDC no outro realm da relação de confiança. O servidor de KDC e o servidor de administração normalmente são executados na mesma máquina com o mesmo FQDN, mas usam diferentes portas. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com:88`).
KDC externo		Especifica que o KDC externo do cluster será usado pelo cluster.
Propriedades do KDC externo	Servidor do administrador	Especifica o nome de domínio totalmente qualificado (FQDN) ou o endereço IP do servidor do administrador externo. O servidor de administração e o servidor de KDC normalmente são executados na mesma máquina com o mesmo FQDN, mas se comunicam por diferentes portas. Se nenhuma porta especificada, a porta 749 será usada, que é o padrão do Kerberos. Opcionalmente, você pode especificar a porta (por exemplo, `domain.example.com:749`).

Parâmetro			Descrição
	Servidor do KDC		Especifica o nome de domínio totalmente qualificado (FQDN) do servidor do KDC externo. O servidor de KDC e o servidor de administração normalmente são executados na mesma máquina com o mesmo FQDN, mas usam diferentes portas. Se nenhuma porta especificada, a porta 88 será usada, que é o padrão do Kerberos. Opcionalm ente, você pode especificar a porta (por exemplo, `domain.example.com :88`).
	Integração do Active Directory		Especifica que a autenticação da entidade principal do Kerberos está integrada a um domínio do Microsoft Active Directory.
	Proprieda des de integraçã o do Active Directory	Realm do Active Directory	Especifica o nome do realm do Kerberos do domínio do Active Directory. Por convenção, os nomes de realm do Kerberos geralmente são iguais ao nome do domínio, mas em letras maiúsculas.
		Domínio do Active Directory	Especifica o nome de domínio do Active Directory.
		Servidor do Active Directory	Especifica o nome de domínio totalmente qualificado (FQDN) do controlador de domínio do Microsoft Active Directory.

Configurações do Kerberos para clusters

Você pode especificar as configurações do Kerberos ao criar um cluster usando o console do Amazon EMR, o AWS CLI ou a API do EMR.

Use as referências a seguir para compreender as definições de configuração de cluster disponíveis para a arquitetura do Kerberos que você escolher. As configurações do console do Amazon EMR são exibidas. Para opções da CLI correspondentes, consulte Exemplos de configuração.

Parâmetro	Descrição
Realm	O nome do realm do Kerberos para o cluster. A convenção do Kerberos deve ser a mesma do nome de domínio, mas em maiúsculas. Por exemplo, para o domínio `ec2.internal` , usando `EC2.INTERNAL` como o nome do realm.
Senha admin do KDC	A senha usada dentro do cluster para `kadmin` ou `kadmin.local` . Essas são interfaces de linha de comando para o sistema de administração do Kerberos V5, que mantém os principais do Kerberos, as políticas de senha e os keytabs do cluster.
Senha do principal da relação de confiança entre realms (opcional)	Obrigatório quando se estabelece uma relação de confiança entre realms. A senha do principal entre realms, que deve ser idêntica em todos os realms. Use uma senha forte.
Usuário de inclusão no domínio do Active Directory (opcional)	Obrigatório ao usar o Active Directory em uma relação de confiança entre realms. Este é o nome de logon de usuário de uma conta do Active Directory com permissão para integrar computadores ao domínio. O Amazon EMR usa essa identidade para integrar o cluster ao domínio. Para ter mais informações, consulte the section called “Etapa 3: adicionar contas de usuário ao domínio do cluster do EMR”.
Senha de inclusão no domínio do Active Directory (opcional)	A senha para o usuário de inclusão no domínio do Active Directory. Para ter mais informações, consulte the section called “Etapa 3: adicionar

Parâmetro	Descrição
	contas de usuário ao domínio do cluster do EMR".

Exemplos de configuração

Os exemplos a seguir demonstram configurações de segurança e configurações de cluster para cenários comuns. AWS CLI os comandos são mostrados para fins de concisão.

KDC local

Os comandos a seguir criam um cluster com um KDC dedicado ao cluster em execução no nó primário. Configurações adicionais no cluster podem ser necessárias. Para ter mais informações, consulte Configurar um cluster para usuários do HDFS autenticados pelo Kerberos e conexões SSH.

Criar configuração de segurança

```
aws emr create-security-configuration --name LocalKDCSecurityConfig \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ClusterDedicatedKdc",\
"ClusterDedicatedKdcConfiguration": {"TicketLifetimeInHours": 24 }}}}'
```

Criar cluster

```
aws emr create-cluster --release-label emr-7.1.0 \
--instance-count 3 --instance-type m5.xlarge \
--applications Name=Hadoop Name=Hive --ec2-attributes
 InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole \
--security-configuration LocalKDCSecurityConfig \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=MyPassword
```

KDC dedicado ao cluster com relação de confiança entre realms do Active Directory

Os comandos a seguir criam um cluster com um KDC dedicado ao cluster em execução no nó primário com uma relação de confiança entre realms para um domínio do Active Directory. Configuração adicional no cluster e no Active Directory é necessária. Para ter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.

Criar configuração de segurança

```
aws emr create-security-configuration --name LocalKDCWithADTrustSecurityConfig \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ClusterDedicatedKdc", \
"ClusterDedicatedKdcConfiguration": {"TicketLifetimeInHours": 24, \
"CrossRealmTrustConfiguration": {"Realm":"AD.DOMAIN.COM", \
"Domain":"ad.domain.com", "AdminServer":"ad.domain.com", \
"KdcServer":"ad.domain.com"}}}}}'
```

Criar cluster

```
aws emr create-cluster --release-label emr-7.1.0 \
--instance-count 3 --instance-type m5.xlarge --applications Name=Hadoop Name=Hive \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration KDCWithADTrustSecurityConfig \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=MyClusterKDCAdminPassword,\
ADDomainJoinUser=ADUserLogonName,ADDomainJoinPassword=ADUserPassword,\
CrossRealmTrustPrincipalPassword=MatchADTrustPassword
```

KDC externo em um cluster diferente

Os comandos a seguir criam um cluster que referencia um KDC dedicado ao cluster no nó primário de um cluster diferente para autenticar entidades principais. Configurações adicionais no cluster podem ser necessárias. Para ter mais informações, consulte Configurar um cluster para usuários do HDFS autenticados pelo Kerberos e conexões SSH.

Criar configuração de segurança

```
aws emr create-security-configuration --name ExtKDCOnDifferentCluster \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ExternalKdc", \
"ExternalKdcConfiguration": {"KdcServerType": "Single", \
"AdminServer": "MasterDNSOfKDCMaster:749", \
"KdcServer": "MasterDNSOfKDCMaster:88"}}}}'
```

Criar cluster

```
aws emr create-cluster --release-label emr-7.1.0 \
--instance-count 3 --instance-type m5.xlarge \
--applications Name=Hadoop Name=Hive \
```

```
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration ExtKDCOnDifferentCluster \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=KDCOnMasterPassword
```

KDC externo do cluster com relação de confiança entre realms do Active Directory

Os comandos a seguir criam um cluster sem nenhum KDC. O cluster faz referência a um KDC dedicado ao cluster em execução no nó primário de outro cluster para autenticar entidades principais. Esse KDC tem uma relação de confiança entre realms com um controlador de domínio do Active Directory. A configuração adicional no nó primário com o KDC é obrigatória. Para ter mais informações, consulte Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory.

Criar configuração de segurança

```
aws emr create-security-configuration --name ExtKDCWithADIntegration \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ExternalKdc", \
"ExternalKdcConfiguration": {"KdcServerType": "Single", \
"AdminServer": "MasterDNSofClusterKDC:749", \
"KdcServer": "MasterDNSofClusterKDC.com:88", \
"AdIntegrationConfiguration": {"AdRealm":"AD.DOMAIN.COM", \
"AdDomain":"ad.domain.com", \
"AdServer":"ad.domain.com"}}}}}'
```

Criar cluster

```
aws emr create-cluster --release-label emr-7.1.0 \
--instance-count 3 --instance-type m5.xlarge --applications Name=Hadoop Name=Hive \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration ExtKDCWithADIntegration \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=KDCOnMasterPassword,\
ADDomainJoinUser=MyPrivilegedADUserName,ADDomainJoinPassword=PasswordForADDomainJoinUser
```

Configurar um cluster para usuários do HDFS autenticados pelo Kerberos e conexões SSH

O Amazon EMR cria clientes de usuário autenticados pelo Kerberos para aplicações executadas no cluster. Por exemplo, o usuário hadoop, o usuário `spark` e outros. Você também pode adicionar usuários autenticados em processos de cluster usando o Kerberos. Os usuários autenticados podem se conectar ao cluster usando as credenciais do Kerberos e trabalhar com os aplicativos. Para que um usuário faça autenticação no cluster, as seguintes configurações são necessárias:

- Deve haver uma conta Linux que corresponda à entidade principal do Kerberos no KDC no cluster. O Amazon EMR faz isso automaticamente em arquiteturas que se integram ao Active Directory.
- Você deve criar um diretório de usuário do HDFS no nó primário para cada usuário e conceder as permissões ao usuário para o diretório.
- É necessário configurar o serviço SSH para que GSSAPI esteja habilitada no nó primário. Além disso, os usuários devem ter um cliente SSH com GSSAPI habilitada.

Adicionar usuários do Linux e entidades principais do Kerberos ao nó primário

Se não usar o Active Directory, você deverá criar contas do Linux no nó primário do cluster e adicionar entidades principais a esses usuários do Linux para o KDC. Isso inclui uma entidade principal no KDC para o nó primário. Além dos usuários principais, o KDC em execução no nó primário precisa de uma entidade principal para o host local.

Quando sua arquitetura inclui integração com o Active Directory, os usuários do Linux e os principais no KDC local, se aplicável, são criados automaticamente. Você pode ignorar esta etapa. Para obter mais informações, consulte Relação de confiança entre realms e KDC externo: KDC do cluster em um cluster diferente com relação de confiança entre realms do Active Directory.

> **Important**
>
> O KDC, junto com o banco de dados de entidades principais, é perdido quando o nó primário é terminado porque o nó primário usa armazenamento temporário. Se você criar usuários para conexões SSH, é recomendável estabelecer uma relação de confiança entre regiões com um KDC externo configurado para alta disponibilidade. Como alternativa, se você criar usuários para conexões SSH usando contas Linux, automatize o processo de criação da conta usando ações e scripts de bootstrap de modo que possa ser repetido ao criar um novo cluster.

Enviar uma etapa ao cluster depois de criá-lo ou ao criar o cluster é a maneira mais fácil de adicionar usuários e principais do KDC. Como alternativa, você pode se conectar ao nó primário usando um par de chaves do EC2 como o usuário `hadoop` padrão para executar os comandos. Para ter mais informações, consulte Conectar-se ao nó primário usando SSH.

O exemplo a seguir envia um script bash `configureCluster.sh` para um cluster que já existe, fazendo referência ao ID do cluster. O script é salvo no Amazon S3.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,\
Args=["s3://DOC-EXAMPLE-BUCKET/configureCluster.sh"]
```

O exemplo a seguir demonstra o conteúdo do script `configureCluster.sh`. O script também trata da criação de diretórios do usuário do HDFS e habilita GSSAPI para SSH, que são abordados nas seções a seguir.

```
#!/bin/bash
#Add a principal to the KDC for the primary node, using the primary node's returned
 host name
sudo kadmin.local -q "ktadd -k /etc/krb5.keytab host/`hostname -f`"
#Declare an associative array of user names and passwords to add
declare -A arr
arr=([lijuan]=pwd1 [marymajor]=pwd2 [richardroe]=pwd3)
for i in ${!arr[@]}; do
    #Assign plain language variables for clarity
     name=${i}
     password=${arr[${i}]}

     # Create a principal for each user in the primary node and require a new password
 on first logon
     sudo kadmin.local -q "addprinc -pw $password +needchange $name"

     #Add hdfs directory for each user
     hdfs dfs -mkdir /user/$name

     #Change owner of each user's hdfs directory to that user
     hdfs dfs -chown $name:$name /user/$name
done

# Enable GSSAPI authentication for SSH and restart SSH service
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Adicionar diretórios do usuário do HDFS

Para permitir que os usuários façam login no cluster para executar trabalhos do Hadoop, você deve adicionar diretórios do usuário HDFS para contas do Linux e conceder a cada um a propriedade do diretório.

Enviar uma etapa ao cluster depois de criá-lo ou ao criar o cluster é a maneira mais fácil de criar diretórios do HDFS. Como alternativa, você pode se conectar ao nó primário usando um par de chaves do EC2 como o usuário `hadoop` padrão para executar os comandos. Para ter mais informações, consulte Conectar-se ao nó primário usando SSH.

O exemplo a seguir envia um script bash `AddHDFSUsers.sh` para um cluster que já existe, fazendo referência ao ID do cluster. O script é salvo no Amazon S3.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-EXAMPLE-BUCKET/AddHDFSUsers.sh"]
```

O exemplo a seguir demonstra o conteúdo do script `AddHDFSUsers.sh`.

```
#!/bin/bash
# AddHDFSUsers.sh script

# Initialize an array of user names from AD, or Linux users created manually on the
 cluster
ADUSERS=("lijuan" "marymajor" "richardroe" "myusername")

# For each user listed, create an HDFS user directory
# and change ownership to the user

for username in ${ADUSERS[@]}; do
     hdfs dfs -mkdir /user/$username
     hdfs dfs -chown $username:$username /user/$username
done
```

Habilitar GSSAPI para SSH

Para usuários autenticados pelo Kerberos se conectarem ao nó primário usando o SSH, o serviço SSH deve ter a autenticação GSSAPI habilitada. Para habilitar GSSAPI, execute os seguintes

comandos na linha de comando do nó primário ou use uma etapa para executá-lo como um script. Depois de reconfigurar o SSH, você deverá reiniciar o serviço.

```
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Usar o SSH para se conectar a clusters kerberizados

Esta seção demonstra as etapas para que um usuário autenticado pelo Kerberos se conecte ao nó primário de um cluster do EMR.

Cada computador que é usado para uma conexão SSH deve ter aplicativos cliente SSH e cliente Kerberos instalados. Os computadores Linux provavelmente incluem esses aplicativos por padrão. Por exemplo, a OpenSSH está instalada na maioria dos sistemas operacionais Unix, Linux e MacOS X. É possível verificar se existe um cliente SSH digitando ssh na linha de comando. Se o computador não reconhecer o comando, instale um cliente SSH para se conectar ao nó primário. O projeto OpenSSH fornece uma implementação grátis do pacote completo de ferramentas SSH. Para obter mais informações, consulte o site do OpenSSH. Os usuários do Windows podem usar aplicativos, como o PuTTY, como um cliente SSH.

Para obter mais informações sobre conexões SSH, consulte Conectar-se a um cluster.

O SSH usa GSSAPI para autenticar os clientes Kerberos, e você deve habilitar a autenticação GSSAPI para o serviço SSH no nó primário do cluster. Para ter mais informações, consulte Habilitar GSSAPI para SSH. Os clientes SSH também devem usar GSSAPI.

Nos exemplos a seguir, para MasterPublicDNS, use o valor que aparece para ***Master public DNS*** *na guia* ***Resumo*** *do painel de detalhes do cluster – por exemplo, ec2-11-222-33-44.compute-1.amazonaws.com.*

Pré-requisito para krb5.conf (que não é do Active Directory)

Ao usar uma configuração sem integração com o Active Directory, além das aplicações cliente SSH e cliente Kerberos, cada computador cliente deve ter uma cópia do arquivo `/etc/krb5.conf` que corresponde ao arquivo `/etc/krb5.conf` no nó primário do cluster.

Para copiar o arquivo krb5.conf

1. Use o SSH para se conectar ao nó primário usando um par de chaves do EC2 e o usuário `hadoop` padrão; por exemplo, `hadoop@`*`MasterPublicDNS`*. Para obter instruções detalhadas, consulte Conectar-se a um cluster.
2. No nó primário, copie o conteúdo do arquivo `/etc/krb5.conf`. Para ter mais informações, consulte Conectar-se a um cluster.
3. Em cada computador cliente que será usado para se conectar ao cluster, crie um arquivo `/etc/krb5.conf` idêntico com base na cópia feita na etapa anterior.

Usar Kinit e SSH

Cada vez que um usuário se conecta a partir de um computador cliente usando credenciais do Kerberos, o usuário deve primeiro renovar tíquetes Kerberos para seu usuário no computador cliente. Além disso, o cliente SSH deve estar configurado para usar a autenticação GSSAPI.

Para usar o SSH para se conectar a um cluster do EMR Kerberizado

1. Use `kinit` para renovar os tíquetes Kerberos, conforme mostrado no exemplo a seguir

   ```
   kinit user1
   ```

2. Use um cliente `ssh` juntamente com o principal que você criou no KDC dedicado ao cluster ou o nome de usuário do Active Directory. Certifique-se de que a autenticação GSSAPI esteja habilitada, conforme mostrado nos exemplos a seguir.

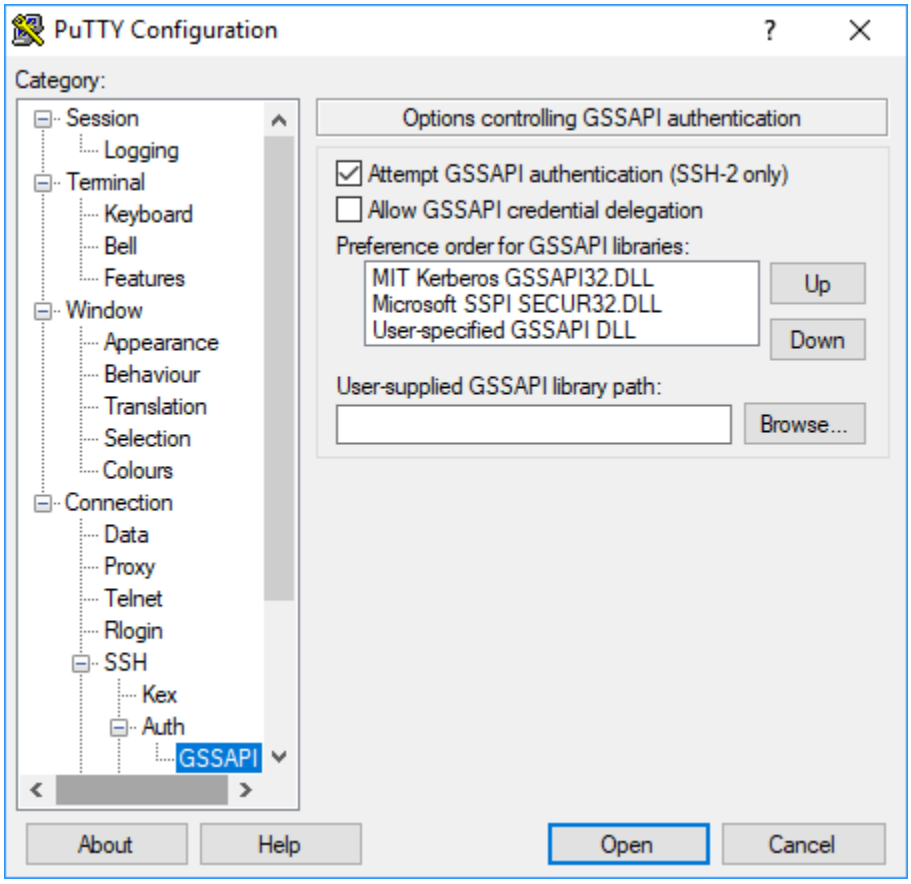
   Exemplo: usuários do Linux

   A opção `-K` especifica a autenticação de GSSAPI.

   ```
   ssh -K user1@MasterPublicDNS
   ```

   Exemplo: usuários do Windows (PuTTY)

   Certifique-se de que a opção de autenticação GSSAPI para a sessão esteja habilitada conforme mostrado:



Tutorial: configurar um KDC dedicado ao cluster

Este tópico orienta você na criação de um cluster com um centro de distribuição de chaves (KDC) dedicado ao cluster, adicionando manualmente contas do Linux a todos os nós do cluster, adicionando entidades principais do Kerberos ao KDC no nó primário e garantindo que os computadores cliente tenham um cliente Kerberos instalado.

Para obter mais informações sobre o suporte do Amazon EMR para Kerberos e KDC, bem como links para a documentação do MIT Kerberos, consulte Usar o Kerberos para autenticação com o Amazon EMR.

Etapa 1: criar o cluster kerberizado

1. Crie uma configuração de segurança que permita o Kerberos. O exemplo a seguir demonstra um `create-security-configuration` comando usando o AWS CLI que especifica a configuração de segurança como uma estrutura JSON embutida. Você também pode fazer referência a um arquivo salvo localmente.

```
aws emr create-security-configuration --name MyKerberosConfig \
--security-configuration '{"AuthenticationConfiguration": {"KerberosConfiguration":
{"Provider": "ClusterDedicatedKdc", "ClusterDedicatedKdcConfiguration":
 {"TicketLifetimeInHours": 24}}}}'
```

2. Crie um cluster que faça referência à configuração de segurança, estabeleça os atributos do Kerberos para o cluster e adicione contas do Linux usando uma ação de bootstrap. O exemplo a seguir demonstra um comando `create-cluster` usando a AWS CLI. O comando faz referência à configuração de segurança criada por você acima, `MyKerberosConfig`. Ele também faz referência a um script simples, `createlinuxusers.sh`, como uma ação de bootstrap, que você cria e carrega no Amazon S3 antes de criar o cluster.

   ```
   aws emr create-cluster --name "MyKerberosCluster" \
   --release-label emr-7.1.0 \
   --instance-type m5.xlarge \
   --instance-count 3 \
   --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2KeyPair \
   --service-role EMR_DefaultRole \
   --security-configuration MyKerberosConfig \
   --applications Name=Hadoop Name=Hive Name=Oozie Name=Hue Name=HCatalog Name=Spark \
   --kerberos-attributes Realm=EC2.INTERNAL,\
   KdcAdminPassword=MyClusterKDCAdminPwd \
   --bootstrap-actions Path=s3://DOC-EXAMPLE-BUCKET/createlinuxusers.sh
   ```

   O código a seguir demonstra o conteúdo do script `createlinuxusers.sh`, que adiciona user1, user2 e user3 a cada nó no cluster. Na próxima etapa, você adicionará esses usuários como principais do KDC.

   ```
   #!/bin/bash
   sudo adduser user1
   sudo adduser user2
   sudo adduser user3
   ```

Etapa 2: adicionar entidades principais ao KDC, criar diretórios de usuário do HDFS e configurar o SSH

O KDC em execução no nó primário precisa de uma entidade principal adicionada para o host local e para cada usuário criado por você no cluster. Você também poderá criar diretórios do HDFS para

cada usuário se eles precisarem se conectar ao cluster e executar trabalhos do Hadoop. Da mesma maneira, configure o SSH para habilitar a autenticação GSSAPI, necessária para o Kerberos. Depois de habilitar GSSAPI, reinicie o serviço SSH.

A maneira mais fácil de realizar essas tarefas é enviar uma etapa para o cluster. O exemplo a seguir envia um `configurekdc.sh` de script bash para o cluster que você criou na etapa anterior, referenciando o ID do cluster. O script é salvo no Amazon S3. Você também pode se conectar ao nó primário usando um par de chaves do EC2 para executar os comandos ou enviar a etapa durante a criação do cluster.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> --steps
 Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,Jar=s3://
myregion.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-EXAMPLE-
BUCKET/configurekdc.sh"]
```

O código a seguir demonstra o conteúdo do script `configurekdc.sh`.

```
#!/bin/bash
#Add a principal to the KDC for the primary node, using the primary node's returned
 host name
sudo kadmin.local -q "ktadd -k /etc/krb5.keytab host/`hostname -f`"
#Declare an associative array of user names and passwords to add
declare -A arr
arr=([user1]=pwd1 [user2]=pwd2 [user3]=pwd3)
for i in ${!arr[@]}; do
    #Assign plain language variables for clarity
     name=${i}
     password=${arr[${i}]}

     # Create principal for sshuser in the primary node and require a new password on
 first logon
     sudo kadmin.local -q "addprinc -pw $password +needchange $name"

     #Add user hdfs directory
     hdfs dfs -mkdir /user/$name

     #Change owner of user's hdfs directory to user
     hdfs dfs -chown $name:$name /user/$name
done

# Enable GSSAPI authentication for SSH and restart SSH service
```

```
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Os usuários que você adicionou agora devem poder se conectar ao cluster usando SSH. Para ter mais informações, consulte Usar o SSH para se conectar a clusters kerberizados.

Tutorial: configurar uma relação de confiança entre realms com um controlador de domínio do Active Directory

Ao configurar uma relação de confiança entre realms, você permite que os principais (normalmente usuários) de um realm do Kerberos diferente se autentiquem em componentes do aplicativo no cluster do EMR. O centro de distribuição de chaves (KDC) dedicado ao cluster estabelece uma relação de confiança com outro KDC usando uma entidade principal entre realms existente em ambos os KDCs. O nome do principal e a senha coincidem precisamente.

Uma relação de confiança entre realms exige que os KDCs possam se alcançar um ao outro pela rede e resolver os nomes de domínio um do outro. As etapas para estabelecer uma relação de confiança entre realms com um controlador de domínio do Microsoft AD em execução como uma instância do EC2 são apresentadas abaixo com uma configuração de rede de exemplo que oferece a conectividade e a resolução de nomes de domínio necessárias. Qualquer configuração de rede que permita o tráfego de rede entre KDCs é aceitável.

Opcionalmente, depois de estabelecer uma relação de confiança entre realms com o Active Directory usando um KDC em um cluster, você poderá criar outro cluster usando uma configuração de segurança diferente para fazer referência ao KDC no primeiro cluster como um KDC externo. Para obter um exemplo de configuração de segurança e a configuração do cluster, consulte KDC externo do cluster com relação de confiança entre realms do Active Directory.

Para obter mais informações sobre o suporte do Amazon EMR para Kerberos e KDC, bem como links para a documentação do MIT Kerberos, consulte Usar o Kerberos para autenticação com o Amazon EMR.

> **Important**
>
> O Amazon EMR não oferece suporte a relações de confiança entre regiões com. AWS Directory Service for Microsoft Active Directory

Etapa 1: configurar a VPC e a sub-rede

As etapas a seguir demonstram como criar uma VPC e uma sub-rede, de maneira que o KDC dedicado ao cluster possa alcançar o controlador de domínio do Active Directory e resolver o nome de domínio. Nessas etapas, a resolução de nomes de domínio é fornecida referenciando-se o controlador de domínio do Active Directory como o servidor de nomes de domínio no conjunto de opções DHCP. Para ter mais informações, consulte Etapa 5: usar uma opção DHCP definida para especificar o controlador de domínio do Active Directory como um servidor DNS da VPC.

O KDC e o controlador de domínio do Active Directory devem poder resolver os nomes de domínio um do outro. Isso permite ao Amazon EMR adicionar computadores ao domínio e configurar automaticamente as contas do Linux correspondentes e os parâmetros SSH em instâncias de cluster.

Se o Amazon EMR não conseguir resolver o nome de domínio, você poderá referenciar a relação de confiança usando o endereço IP do controlador de domínio do Active Directory. No entanto, você deve adicionar manualmente contas do Linux, adicionar entidades principais correspondentes ao KDC dedicado ao cluster e configurar o SSH.

Para configurar a VPC e a sub-rede

1. Crie uma Amazon VPC com uma única sub-rede pública. Para obter mais informações, consulte Step 1: Create the VPC no Amazon VPC Getting Started Guide.

> **Important**
>
> Ao usar um controlador de domínio do Microsoft Active Directory, escolha um bloco CIDR para o cluster do EMR, de maneira que todos os endereços IPv4 tenham menos de nove caracteres (por exemplo, 10.0.0.0/16). Isso ocorre porque os nomes DNS dos computadores de cluster são usados quando os computadores ingressam no diretório do Active Directory. AWS atribui nomes de host DNS com base no endereço IPv4 de forma que endereços IP mais longos possam resultar em nomes DNS com mais de 15 caracteres. O Active Directory tem um limite de 15 caracteres para registrar nomes de computador adicionados e trunca nomes mais longos, o que pode causar erros imprevisíveis.

2. Remova o conjunto de opções DHCP padrão atribuído à VPC. Para obter mais informações, consulte Changing a VPC to use No DHCP options. Posteriormente, você adicionará um novo especificando o controlador de domínio do Active Directory como o servidor DNS.
3. Confirme se o suporte DNS está habilitado para a VPC, ou seja, se os nomes de host e a resolução DNS estão habilitados. Por padrão, as transições estão ativadas. Para obter mais informações, consulte Updating DNS support for your VPC.
4. Confirme se a VPC tem um gateway da Internet anexado, que é o padrão. Para mais informações, consulte Criar e anexar um gateway da Internet.

> **Note**
>
> Um gateway da Internet é usado neste exemplo porque você está estabelecendo um novo controlador de domínio para a VPC. O gateway da Internet talvez não seja necessário para o aplicativo. O único requisito é que o KDC dedicado ao cluster possa acessar o controlador de domínio do Active Directory.

5. Crie uma tabela de rotas personalizada, adicione uma rota com o gateway da Internet como destino e a anexe à sub-rede. Para obter mais informações, consulte Criar uma tabela de rotas personalizada.
6. Ao executar a instância do EC2 do controlador de domínio, ela precisa ter um endereço IPv4 público estático para você se conectar a ela usando RDP. A maneira mais fácil de fazer isso é configurar a sub-rede para atribuir automaticamente endereços IPv4 públicos. Não se trata da configuração padrão quando uma sub-rede é criada. Para obter mais informações, consulte Modificar o atributo de endereçamento IPv4 público para a sub-rede. Você também pode atribuir

o endereço ao iniciar a instância. Para obter mais informações, consulte Assigning a public IPv4 address during instance launch.

7. Quando terminar, anote a VPC e os IDs de sub-rede. Você os usará depois quando iniciar o controlador de domínio do Active Directory e o cluster.

Etapa 2: iniciar e instalar o controlador de domínio do Active Directory

1. Inicie uma instância do EC2 com base na AMI Microsoft Windows Server 2016 Base. Recomendamos um tipo de instância m4.xlarge ou melhor. Para obter mais informações, consulte Lançamento de uma AWS Marketplace instância no Guia do usuário do Amazon EC2.
2. Anote o ID do grupo de segurança associado à instância do EC2. Você precisa dele para o Etapa 6: Iniciar um cluster EMR Kerberizado. Nós usamos *sg-012xrlmdomain345*. Opcionalmente, você pode especificar grupos de segurança diferentes para o cluster do EMR e essa instância que permite o tráfego entre eles. Para obter mais informações, consulte Grupos de segurança do Amazon EC2 para instâncias do Linux no Guia do usuário do Amazon EC2.
3. Conecte-se à instância do EC2 usando o RDP. Para obter mais informações, consulte Conectando-se à sua instância do Windows no Guia do usuário do Amazon EC2.
4. Inicie o Server Manager para instalar e configurar o perfil Active Directory Domain Services no servidor. Promova o servidor para um controlador de domínio e atribua um nome de domínio (o exemplo que usamos aqui é *ad.domain.com*). Anote o nome de domínio porque você vai precisar dele depois ao criar a configuração de segurança do EMR e o cluster. Se estiver começando a configurar o Active Directory, você poderá seguir as instruções em How to setup Active Directory (AD) In Windows Server 2016.

   A instância será reiniciada quando você terminar.

Etapa 3: adicionar contas de usuário ao domínio do cluster do EMR

Use o RDP para o controlador de domínio do Active Directory para criar contas em usuários e computadores do Active Directory para cada usuário do cluster. Para obter mais informações, consulte Create a User Account in Active Directory Users and Computers no site Microsoft Learn. Anote o User logon name (Nome de logon do usuário) de cada usuário. Você precisará dele mais tarde ao configurar o cluster.

Além disso, crie uma conta com privilégios suficientes para integrar computadores ao domínio. Você especifica essa conta ao criar um cluster. O Amazon EMR a usa para integrar instâncias de cluster ao domínio. Você especifica essa conta e a senha em Etapa 6: Iniciar um cluster EMR Kerberizado.

Para delegar privilégios de integração do computador à conta, recomendamos criar um grupo com privilégios de junção e, em seguida, atribuir o usuário ao grupo. Para obter instruções, consulte Delegating directory join privileges no Guia de administração AWS Directory Service .

Etapa 4: configurar uma relação de confiança recebida no controlador de domínio do Active Directory

Os comandos de exemplo abaixo criam uma relação de confiança no Active Directory, que é uma relação de confiança de realm unidirecional, de entrada e não transitiva com o KDC dedicado ao cluster. O exemplo que usamos para no realm do cluster é *EC2.INTERNAL*. Substitua *KDC-FQDN* pelo nome DNS público listado para o nó primário do Amazon EMR que hospeda o KDC. O parâmetro `passwordt` especifica a cross-realm principal password (senha da entidade principal entre realms), determinada por você com o realm do cluster ao criar um cluster. O nome do realm deriva do nome de domínio padrão em `us-east-1` para o cluster. O `Domain` é o domínio do Active Directory no qual você está criando a confiança, que é em minúscula por convenção. O exemplo usa *ad.domain.com*

Abra o prompt de comando do Windows com privilégios de administrador e digite os seguintes comandos para criar a relação de confiança no controlador de domínio do Active Directory:

```
C:\Users\Administrator> ksetup /addkdc EC2.INTERNAL KDC-FQDN
C:\Users\Administrator> netdom trust EC2.INTERNAL /Domain:ad.domain.com /add /realm /
passwordt:MyVeryStrongPassword
C:\Users\Administrator> ksetup /SetEncTypeAttr EC2.INTERNAL AES256-CTS-HMAC-SHA1-96
```

Etapa 5: usar uma opção DHCP definida para especificar o controlador de domínio do Active Directory como um servidor DNS da VPC

Agora que o controlador de domínio do Active Directory está configurado, você deve configurar a VPC para usá-lo como um servidor de nomes de domínio para resolução de nomes em sua VPC. Para isso, anexe um conjunto de opções DHCP. Especifique o Nome do domínio como o nome de domínio do cluster. Por exemplo, `ec2.internal` caso o cluster esteja em us-east-1 ou *region*`.compute.internal` para outras regiões. Para servidores de nomes de domínio, você deve especificar o endereço IP do controlador de domínio do Active Directory (que deve ser acessível a partir do cluster) como a primeira entrada, seguido pelo AmazonProvidedDNS (por exemplo, ***xx.xx.xx.xx***, DNS). AmazonProvided Para obter mais informações, consulte Changing DHCP option sets.

Etapa 6: Iniciar um cluster EMR Kerberizado

1. No Amazon EMR, crie uma configuração de segurança que especifique o controlador de domínio do Active Directory criado por você nas etapas anteriores. Um comando de exemplo é mostrado abaixo. Substitua o domínio, *ad.domain.com*, pelo nome do domínio especificado por você em Etapa 2: iniciar e instalar o controlador de domínio do Active Directory.

```
aws emr create-security-configuration --name MyKerberosConfig \
--security-configuration '{
  "AuthenticationConfiguration": {
    "KerberosConfiguration": {
      "Provider": "ClusterDedicatedKdc",
      "ClusterDedicatedKdcConfiguration": {
        "TicketLifetimeInHours": 24,
        "CrossRealmTrustConfiguration": {
          "Realm": "AD.DOMAIN.COM",
          "Domain": "ad.domain.com",
          "AdminServer": "ad.domain.com",
          "KdcServer": "ad.domain.com"
        }
      }
    }
  }
}'
```

2. Crie o cluster com os seguintes atributos:

   - Use a opção `--security-configuration` para especificar a configuração de segurança que você criou. Usamos *MyKerberosConfig* no exemplo.
   - Use a propriedade `SubnetId` da `--ec2-attributes option` para especificar a sub-rede que você criou em Etapa 1: configurar a VPC e a sub-rede. Nós usamos *step1-subnet* no exemplo.
   - Use `AdditionalMasterSecurityGroups` e `AdditionalSlaveSecurityGroups` da opção `--ec2-attributes` para especificar que o grupo de segurança associado ao controlador de domínio AD do Etapa 2: iniciar e instalar o controlador de domínio do Active Directory está associado ao nó primário do cluster, bem como aos nós centrais e de tarefa. Nós usamos *sg-012xrlmdomain345* no exemplo.

   Use `--kerberos-attributes` para especificar os seguintes atributos Kerberos específicos ao cluster:

- O realm do cluster especificado por você ao configurar o controlador de domínio do Active Directory.
- A senha da entidade principal da relação de confiança entre realms especificada por você como `passwordt` em Etapa 4: configurar uma relação de confiança recebida no controlador de domínio do Active Directory.
- Um `KdcAdminPassword`, que você pode usar para administrar o KDC dedicado ao cluster.
- O nome de logon do usuário e a senha da conta do Active Directory com privilégios de ingresso no computador criados por você em Etapa 3: adicionar contas de usuário ao domínio do cluster do EMR.

O exemplo a seguir inicia um cluster kerberizado.

```
aws emr create-cluster --name "MyKerberosCluster" \
--release-label emr-5.10.0 \
--instance-type m5.xlarge \
--instance-count 3 \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2KeyPair,\
SubnetId=step1-subnet, AdditionalMasterSecurityGroups=sg-012xrlmdomain345,
AdditionalSlaveSecurityGroups=sg-012xrlmdomain345\
--service-role EMR_DefaultRole \
--security-configuration MyKerberosConfig \
--applications Name=Hadoop Name=Hive Name=Oozie Name=Hue Name=HCatalog Name=Spark \
--kerberos-attributes Realm=EC2.INTERNAL,\
KdcAdminPassword=MyClusterKDCAdminPwd,\
ADDomainJoinUser=ADUserLogonName,ADDomainJoinPassword=ADUserPassword,\
CrossRealmTrustPrincipalPassword=MatchADTrustPwd
```

Etapa 7: criar usuários HDFS e definir permissões no cluster para contas do Active Directory

Ao configurar uma relação de confiança com o Active Directory, o Amazon EMR cria usuários do Linux no cluster para cada conta do Active Directory. Por exemplo, o nome de logon de usuário `LiJuan` no Active Directory tem uma conta do Linux de `lijuan`. Os nomes de usuário do Active Directory podem conter letras maiúsculas, mas o Linux não segue o uso de maiúsculas e minúsculas do Active Directory.

Para permitir que os usuários façam login no cluster para executar trabalhos do Hadoop, você deve adicionar diretórios do usuário HDFS para contas do Linux e conceder a cada um a propriedade do diretório. Para isso, recomendamos executar um script salvo no Amazon S3 como uma etapa

de cluster. Você também pode executar os comandos no script abaixo da linha de comando no nó primário. Use o par de chaves do EC2 especificado por você quando criou o cluster para se conectar ao nó primário via SSH como o usuário do Hadoop. Para ter mais informações, consulte Usar um par de chaves do EC2 para credenciais SSH.

Execute o comando a seguir para adicionar uma etapa ao cluster que executa um script, *AddHDFSUsers.sh*.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-EXAMPLE-BUCKET/AddHDFSUsers.sh"]
```

O conteúdo do arquivo *AddHDFSUsers.sh* é o seguinte.

```
#!/bin/bash
# AddHDFSUsers.sh script

# Initialize an array of user names from AD or Linux users and KDC principals created
 manually on the cluster
ADUSERS=("lijuan" "marymajor" "richardroe" "myusername")

# For each user listed, create an HDFS user directory
# and change ownership to the user

for username in ${ADUSERS[@]}; do
     hdfs dfs -mkdir /user/$username
     hdfs dfs -chown $username:$username /user/$username
done
```

Grupos do Active Directory mapeados para grupos do Hadoop

O Amazon EMR usa o Daemon do System Security Services (SSD) para mapear grupos do Active Directory para grupos do Hadoop. Para confirmar mapeamentos de grupos, depois de fazer login no nó primário, conforme descrito em Usar o SSH para se conectar a clusters kerberizados, você poderá usar o comando `hdfs groups` para confirmar que os grupos do Active Directory aos quais sua conta do Active Directory pertence foram mapeados para os grupos do Hadoop para o usuário correspondente do Hadoop no cluster. Você também pode verificar mapeamentos de grupos de outros usuários especificando um ou mais nomes de usuário usando, por exemplo, o comando `hdfs groups` *`lijuan`*. Para obter mais informações, consulte grupos no Guia de comandos HDFS do Apache.

Usar servidores Active Directory ou LDAP para autenticação com o Amazon EMR

Com o Amazon EMR 6.12.0 e versões posteriores, você pode usar o protocolo LDAP sobre SSL (LDAPS) para iniciar um cluster que se integra de forma nativa ao servidor de identidade corporativo. O Lightweight Directory Access Protocol (LDAP) é um protocolo de aplicação aberto e independente de fornecedor que acessa e mantém dados. O LDAP é bastante usado para autenticação de usuários em servidores de identidade corporativa hospedados em aplicações como o Active Directory (AD) e o OpenLDAP. Com essa integração nativa, você pode usar o servidor LDAP para autenticar usuários no Amazon EMR.

Os destaques da integração do LDAP do Amazon EMR incluem:

- O Amazon EMR configura as aplicações compatíveis para se autenticarem com a autenticação LDAP em seu nome.
- O Amazon EMR configura e mantém a segurança das aplicações compatíveis com o protocolo Kerberos. Não é necessário inserir nenhum comando ou script.
- Você recebe controle de acesso refinado (FGAC) por meio da autorização do Apache Ranger para bancos de dados e tabelas do Hive Metastore. Consulte Integrar o Amazon EMR com o Apache Ranger Para mais informações.
- Ao necessitar de credenciais LDAP para acessar um cluster, você recebe controle de acesso refinado (FGAC) sobre quem pode acessar seus clusters do EMR por meio de SSH.

As páginas a seguir fornecem uma visão geral conceitual, os pré-requisitos e as etapas para iniciar um cluster do EMR com a integração LDAP do Amazon EMR.

Tópicos

Visão geral do LDAP com o Amazon EMR

O Lightweight Directory Access Protocol (LDAP) é um protocolo de software que os administradores de rede usam para gerenciar e controlar o acesso aos dados por meio da autenticação de usuários na rede de uma empresa. O protocolo LDAP armazena informações em uma estrutura hierárquica de diretórios em árvore. Para obter mais informações, consulte Basic LDAP Concepts no LDAP.com.

Na rede de uma empresa, muitas aplicações podem usar o protocolo LDAP para autenticar usuários. Com a integração LDAP do Amazon EMR, os clusters do EMR podem usar o mesmo protocolo LDAP de maneira nativa com uma configuração de segurança adicionada.

Há duas implementações principais do protocolo LDAP compatíveis com o Amazon EMR: Active Directory e OpenLDAP. Há outras implementações possíveis, mas a maioria se encaixa nos mesmos protocolos de autenticação do Active Directory ou do OpenLDAP.

Active Directory (AD)

O Active Directory (AD) é um serviço de diretório da Microsoft para redes de domínio Windows. O AD está incluído na maioria dos sistemas operacionais Windows Server e pode se comunicar com clientes pelos protocolos LDAP e LDAPS. Para autenticação, o Amazon EMR tenta usar a associação do usuário com a instância do AD com o nome da entidade principal do usuário (UPN) como nome e senha distintos. O UPN usa o formato padrão `username@domain_name`.

OpenLDAP

O OpenLDAP é uma implementação gratuita e de código aberto do protocolo LDAP. Para autenticação, o Amazon EMR tenta usar a associação do usuário com a instância do OpenLDAP com o nome de domínio totalmente qualificado (FQDN) como nome distinto e senha. O FQDN usa o formato padrão `username_attribute=username,LDAP_user_search_base`. Normalmente, o valor de `username_attribute` é `uid`, e o valor de `LDAP_user_search_base` contém os atributos da árvore que leva ao usuário. Por exemplo, `ou=People,dc=example,dc=com`.

Outras implementações gratuitas e de código aberto do protocolo LDAP normalmente seguem um FQDN semelhante ao OpenLDAP para os nomes distintos dos respectivos usuários.

Componentes LDAP para Amazon EMR

Você pode usar seu servidor LDAP para se autenticar no Amazon EMR e em qualquer aplicação que o usuário utilize diretamente no cluster do EMR por meio dos componentes a seguir.

Agente secreto

O agente secreto é um processo no cluster que autentica todas as solicitações do usuário. O agente secreto cria a associação do usuário para o servidor LDAP em nome das aplicações compatíveis com o cluster do EMR. O agente secreto é executado como o usuário `emrsecretagent` e grava logs no diretório `/emr/secretagent/log`. Esses logs fornecem detalhes sobre o estado da solicitação de autenticação de cada usuário e os erros que possam surgir durante a autenticação do usuário.

System Security Services Daemon (SSSD)

O SSSD é um daemon executado em cada nó de um cluster do EMR habilitado para LDAP. O SSSD cria e gerencia um usuário UNIX para sincronizar sua identidade corporativa remota com cada nó. Aplicações baseadas em YARN, como o Hive e o Spark, exigem que haja um usuário UNIX local em cada nó que executa uma consulta para um usuário.

Suporte de aplicações e considerações com o LDAP para Amazon EMR

Aplicações compatíveis com LDAP para Amazon EMR

> **Important**
>
> As aplicações listadas nesta página são as únicas com suporte do Amazon EMR para LDAP. Para garantir a segurança do cluster, só é possível incluir aplicações compatíveis com LDAP ao criar um cluster do EMR com o LDAP habilitado. Se você tentar instalar outras aplicações sem suporte, o Amazon EMR rejeitará a solicitação de um novo cluster.

O Amazon EMR 6.12 e versões posteriores oferece suporte à integração LDAP com as seguintes aplicações:

- Apache Livy
- Apache Hive até HiveServer 2 (HS2)
- Trino
- Presto
- Hue

Também é possível instalar as seguintes aplicações em um cluster do EMR e configurá-las para atender a suas necessidades de segurança:

- Apache Spark
- Apache Hadoop

Atributos compatíveis com LDAP para Amazon EMR

É possível usar os seguintes recursos do Amazon EMR com a integração do LDAP:

> **Note**
>
> Para manter as credenciais LDAP seguras, é necessário usar criptografia em trânsito para proteger o fluxo de dados dentro e fora do cluster. Para obter mais informações sobre criptografia em trânsito, consulte Criptografar dados em repouso e em trânsito.

- Criptografia em trânsito (obrigatório) e em repouso
- Grupos de instâncias, frotas de instâncias e instâncias spot
- Reconfiguração de aplicações em um cluster em execução
- Criptografia do lado do servidor (SSE) do EMRFS

Atributos não compatíveis

Considere as seguintes limitações ao usar a integração do LDAP com Amazon EMR:

- O Amazon EMR desabilita etapas para clusters com o LDAP habilitado.
- O Amazon EMR não oferece suporte a funções e AWS Lake Formation integrações de tempo de execução para clusters com LDAP habilitado.
- O Amazon EMR não oferece suporte a LDAP com StartTLS.
- O Amazon EMR não oferece suporte ao modo de alta disponibilidade (clusters com múltiplos nós primários) para clusters com LDAP habilitado.
- Não é possível alternar credenciais ou certificados de vinculação para clusters com LDAP habilitado. Se algum desses campos tiver sido alternado, é recomendável iniciar um novo cluster com as credenciais ou certificados de vinculação atualizados.

- Você deve usar bases de pesquisa exatas com o LDAP. A base de pesquisa de usuários e grupos do LDAP não oferece suporte aos filtros de pesquisa do LDAP.

Configurar e iniciar um cluster do EMR com LDAP

Esta seção aborda como configurar o Amazon EMR para uso com autenticação LDAP.

Tópicos

- Adicione AWS Secrets Manager permissões à função de instância do Amazon EMR
- Criar a configuração de segurança do Amazon EMR para integração com LDAP
- Iniciar um cluster do EMR que se autentique com LDAP

Adicione AWS Secrets Manager permissões à função de instância do Amazon EMR

O Amazon EMR usa um perfil de serviço do IAM para realizar ações a seu favor a fim de provisionar e gerenciar clusters. O perfil de serviço para instâncias do EC2 do cluster, também chamada de perfil de instância do EC2 para Amazon EMR, é um tipo especial de perfil de serviço que o Amazon EMR atribui ao iniciar cada instância do EC2 do cluster.

Para definir permissões para que um cluster do EMR interaja com dados do Amazon S3 e outros serviços da AWS , defina um perfil de instância personalizado do Amazon EC2 no lugar de `EMR_EC2_DefaultRole` ao executar o cluster. Para obter mais informações, consulte Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2) e Personalizar perfis do IAM.

Adicione as seguintes instruções ao perfil de instância padrão do EC2 para permitir que o Amazon EMR marque sessões e acesse AWS Secrets Manager aquelas que armazenam certificados LDAP.

```
    {
      "Sid": "AllowAssumeOfRolesAndTagging",
      "Effect": "Allow",
      "Action": ["sts:TagSession", "sts:AssumeRole"],
      "Resource": [
        "arn:aws:iam::111122223333:role/LDAP_DATA_ACCESS_ROLE_NAME",
        "arn:aws:iam::111122223333:role/LDAP_USER_ACCESS_ROLE_NAME"
      ]
    },
    {
        "Sid": "AllowSecretsRetrieval",
        "Effect": "Allow",
```

```
        "Action": "secretsmanager:GetSecretValue",
        "Resource": [
            "arn:aws:secretsmanager:us-east-1:111122223333:secret:LDAP_SECRET_NAME*",
            "arn:aws:secretsmanager:us-
east-1:111122223333:secret:ADMIN_LDAP_SECRET_NAME*"
        ]
    }
```

> **Note**
>
> Suas solicitações de cluster falharão se você esquecer o caractere curinga * no final do nome do segredo ao definir as permissões do Secrets Manager. O curinga representa as versões do segredo.
> Você também deve limitar o escopo da AWS Secrets Manager política somente aos certificados que seu cluster precisa para provisionar instâncias.

Criar a configuração de segurança do Amazon EMR para integração com LDAP

Antes de iniciar um cluster do EMR com integração com LDAP, use as etapas descritas em Criar uma configuração de segurança para criar uma configuração de segurança do Amazon EMR para o cluster. Complete as seguintes configurações no bloco de `LDAPConfiguration` em `AuthenticationConfiguration` ou nos campos correspondentes na seção Configurações de segurança do console do Amazon EMR:

`EnableLDAPAuthentication`

Opção do console: Protocolo de autenticação: LDAP

Para usar a integração com LDAP, defina essa opção como `true` ou selecione-a como protocolo de autenticação ao criar um cluster no console. Por padrão, `EnableLDAPAuthentication` é `true` ao criar uma configuração de segurança no console do Amazon EMR.

`LDAPServerURL`

Opção do console: local do servidor LDAP

A localização do servidor LDAP, incluindo o prefixo: `ldaps://`*`location_of_server`*.

`BindCertificateARN`

Opção do console: certificado SSL LDAP

O AWS Secrets Manager ARN que contém o certificado para assinar o certificado SSL que o servidor LDAP usa. Se seu servidor LDAP for assinado por uma Autoridade Certificadora (CA) pública, você poderá fornecer um AWS Secrets Manager ARN com um arquivo em branco. Para obter mais informações sobre como armazenar seu certificado no Secrets Manager, consulte Armazenar certificados TLS no AWS Secrets Manager.

`BindCredentialsARN`

Opção do console: credenciais de vinculação do servidor LDAP

Um AWS Secrets Manager ARN que contém as credenciais de associação do usuário administrador do LDAP. As credenciais são armazenadas como objeto JSON. Há somente um par de chave-valor nesse segredo; a chave no par é o nome de usuário e o valor é a senha. Por exemplo, `{"uid=admin,cn=People,dc=example,dc=com": "AdminPassword1"}`. Esse é um campo opcional, a menos que você habilite o login SSH para o cluster do EMR. Em muitas configurações, as instâncias do Active Directory exigem credenciais de vinculação para permitir que o SSSD sincronize usuários.

`LDAPAccessFilter`

Opção do console: filtro de acesso LDAP

Especifica o subconjunto de objetos no servidor LDAP que podem ser autenticados. Por exemplo, para conceder acesso a todos os usuários com a classe de objeto `posixAccount` no servidor LDAP, defina o filtro de acesso como `(objectClass=posixAccount)`.

`LDAPUserSearchBase`

Opção do console: base de pesquisa de usuários LDAP

A base de pesquisa à qual seus usuários pertencem no servidor LDAP. Por exemplo, `cn=People,dc=example,dc=com`.

`LDAPGroupSearchBase`

Opção de console: base de pesquisa de grupos LDAP

A base de pesquisa à qual seus grupos pertencem no servidor LDAP. Por exemplo, `cn=Groups,dc=example,dc=com`.

`EnableSSHLogin`

Opção do console: login SSH

Especifica se a autenticação por senha com credenciais LDAP deverá ou não ser permitida. Não é recomendável habilitar essa opção. Os pares de chaves são uma rota mais segura para permitir o acesso aos clusters do EMR. Esse campo é opcional e usa o padrão `false`.

`LDAPServerType`

Opção de console: tipo de servidor LDAP

Especifica o tipo de servidor LDAP ao qual o Amazon EMR se conectará. As opções compatíveis são Active Directory e OpenLDAP. Outros tipos de servidor LDAP podem funcionar, mas o Amazon EMR não é oficialmente compatível com outros tipos de servidor. Para ter mais informações, consulte Componentes LDAP para Amazon EMR.

`ActiveDirectoryConfigurations`

Um sub-bloco necessário para configurações de segurança que utilizam o tipo de servidor Active Directory.

`ADDomain`

Opção do console: domínio do Active Directory

O nome de domínio usado para criar o nome da entidade principal do usuário (UPN) para autenticação do usuário com configurações de segurança que usam o tipo de servidor Active Directory.

Considerações sobre configurações de segurança com LDAP e Amazon EMR

- Para criar uma configuração de segurança com a integração LDAP do Amazon EMR, é necessário usar criptografia em trânsito. Para obter informações sobre criptografia em trânsito, consulte Criptografar dados em repouso e em trânsito.
- Não é possível definir a configuração do Kerberos na mesma configuração de segurança. O Amazon EMR provisiona um KDC que é dedicado automaticamente e gerencia a senha de administrador para o KDC. Os usuários não poderão acessar essa senha de administrador.
- Você não pode definir funções de tempo de execução do IAM e AWS Lake Formation na mesma configuração de segurança.
- `LDAPServerURL` deve ter o protocolo `ldaps://` em seu valor.
- `LDAPAccessFilter` não pode estar vazio.

Usar o LDAP com a integração do Apache Ranger para Amazon EMR

Com a integração LDAP para Amazon EMR, é possível se integrar ainda mais com o Apache Ranger. Ao inserir seus usuários LDAP no Ranger, você pode associar esses usuários a um servidor de políticas Apache Ranger para integração com o Amazon EMR e outras aplicações. Para isso, defina o campo `RangerConfiguration` em `AuthorizationConfiguration` na configuração de segurança que você usa com o cluster do LDAP. Para obter mais informações sobre como definir a configuração de segurança, consulte Criar a configuração de segurança do EMR.

Ao usar o LDAP com o Amazon EMR, não é necessário fornecer uma `KerberosConfiguration` com a integração com o Amazon EMR para Apache Ranger.

Iniciar um cluster do EMR que se autentique com LDAP

Realize as etapas a seguir para iniciar um cluster do EMR com LDAP ou Active Directory.

1. Configure o ambiente:

   - Certifique-se de que os nós em seu cluster do EMR possam se comunicar com o Amazon S3 e. AWS Secrets Manager Para obter mais informações sobre como modificar seu perfil do perfil de instância do EC2 para se comunicar com esses serviços, consulte Adicione AWS Secrets Manager permissões à função de instância do Amazon EMR.
   - Se você planeja executar seu cluster do EMR em uma sub-rede privada, você deve usar endpoints da AWS PrivateLink Amazon VPC ou usar a tradução de endereços de rede (NAT) para configurar a VPC para se comunicar com o S3 e o Secrets Manager. Para obter mais informações, consulte AWS PrivateLink and VPC endpoints e NAT instances no Amazon VPC Getting Started Guide.
   - Verifique se há conectividade de rede entre o cluster do EMR e o servidor LDAP. Seus clusters do EMR devem acessar o servidor LDAP pela rede. Os nós primário, central e de tarefa do cluster se comunicam com o servidor LDAP para sincronizar os dados do usuário. Se o servidor LDAP for executado no Amazon EC2, atualize o grupo de segurança do EC2 para aceitar o tráfego do cluster do EMR. Para ter mais informações, consulte Adicione AWS Secrets Manager permissões à função de instância do Amazon EMR.
2. Criar uma configuração de segurança do Amazon EMR para integração com LDAP. Para ter mais informações, consulte Criar a configuração de segurança do Amazon EMR para integração com LDAP.
3. Agora que você está configurado, use as etapas descritas em Inicialização de um cluster do Amazon EMR para iniciar o cluster com as seguintes configurações:

- Selecione Amazon EMR versão 6.12 ou posterior. É recomendável usar a versão mais recente do Amazon EMR.
- Especifique ou selecione somente aplicações para o cluster compatíveis com LDAP. Para obter uma lista de aplicações compatíveis com LDAP com o Amazon EMR, consulte Suporte de aplicações e considerações com o LDAP para Amazon EMR.
- Aplique a configuração de segurança criada na etapa anterior.

Exemplos usando o LDAP com Amazon EMR

Depois de provisionar um cluster do EMR que usa a integração com LDAP, você pode fornecer suas credenciais LDAP para qualquer aplicação compatível por meio de seu mecanismo de autenticação de nome de usuário e senha incorporado. Esta página mostra alguns exemplos.

Usar a autenticação LDAP com o Apache Hive

Example - Apache Hive

O comando de exemplo a seguir inicia uma sessão do Apache Hive por meio de HiveServer 2 e Beeline:

```
beeline -u "jdbc:hive2://$HOSTNAME:10000/default;ssl=true;sslTrustStore=
$TRUSTSTORE_PATH;trustStorePassword=$TRUSTSTORE_PASS"  -n LDAP_USERNAME -
p LDAP_PASSWORD
```

Usar a autenticação LDAP com o Apache Livy

Example - Apache Livy

O comando de exemplo a seguir inicia uma sessão do Livy por cURL. Substitua *ENCODED-KEYPAIR* com uma string codificada em Base64 por `username:password`.

```
curl -X POST --data '{"proxyUser":"LDAP_USERNAME","kind": "pyspark"}' -H "Content-Type:
 application/json" -H "Authorization: Basic ENCODED-KEYPAIR" DNS_OF_PRIMARY_NODE:8998/
sessions
```

Usar autenticação do LDAP com o Presto

Example - Presto

O comando de exemplo a seguir inicia uma sessão do Presto pela CLI do Presto:

```
presto-cli --user "LDAP_USERNAME" --password --catalog hive
```

Após executar esse comando, digite a senha do LDAP no prompt.

Usar a autenticação LDAP com o Trino

Example - Trino

O comando de exemplo a seguir inicia uma sessão do Trino pela CLI do Trino:

```
trino-cli --user "LDAP_USERNAME" --password --catalog hive
```

Após executar esse comando, digite a senha do LDAP no prompt.

Usar a autenticação LDAP com o Hue

Você pode acessar a interface do usuário do Hue por um túnel SSH criado no cluster ou pode configurar um servidor proxy para transmitir publicamente a conexão com o Hue. Como o Hue não é executado no modo HTTPS por padrão, é recomendável usar uma camada de criptografia adicional para garantir que a comunicação entre os clientes e a interface do usuário do Hue seja criptografada com HTTPS. Isso reduz a chance de expor acidentalmente as credenciais do usuário em texto sem formatação.

Para usar a interface do usuário do Hue, abra a interface do usuário do Hue no navegador e digite a senha do nome de usuário LDAP para fazer login. Se as credenciais estiverem corretas, o Hue fará login e usará sua identidade para autenticar você em todas as aplicações compatíveis.

Usar SSH para autenticação por senha e tíquetes do Kerberos para outras aplicações

> **Important**
>
> Não é recomendável usar a autenticação por senha com um cluster do EMR.

Você pode usar suas credenciais LDAP para fazer SSH em um cluster do EMR. Para isso, defina a configuração `EnableSSHLogin` como `true` na configuração de segurança do Amazon EMR usada para iniciar o cluster. Depois, use o comando a seguir para SSH no cluster depois que ele for iniciado:

```
ssh username@EMR_PRIMARY_DNS_NAME
```

Após executar esse comando, digite a senha do LDAP no prompt.

O Amazon EMR inclui um script no cluster que permite aos usuários gerar um arquivo keytab e um tíquete do Kerberos para usar com aplicações compatíveis que não aceitam credenciais LDAP diretamente. Alguns desses aplicativos incluem `spark-submit` Spark SQL e. PySpark

Execute `ldap-kinit` e siga as instruções. Se a autenticação tiver êxito, o arquivo keytab do Kerberos será exibido no diretório inicial com um tíquete do Kerberos válido. Use o tíquete do Kerberos para executar aplicações como você faria em qualquer ambiente kerberizado.

Integre o Amazon EMR com AWS IAM Identity Center

Com as versões 6.15.0 e superiores do Amazon EMR, você pode usar identidades de para se AWS IAM Identity Center autenticar em um cluster do Amazon EMR. As seções a seguir fornecem uma visão geral conceitual, os pré-requisitos e as etapas necessárias para executar um cluster do EMR com a integração do Centro de Identidade.

Tópicos

Visão geral

A propagação confiável de identidades por meio do IAM Identity Center pode ajudar você a criar ou conectar com segurança suas identidades de força de trabalho e gerenciar centralmente o acesso entre contas e aplicativos. AWS Com esse recurso, um usuário pode entrar no aplicativo que usa propagação de identidade confiável e esse aplicativo pode transmitir a identidade do usuário nas solicitações que ele faz para acessar dados em AWS serviços que também usam propagação de identidade confiável. Como o acesso é gerenciado com base na identidade do usuário, os usuários não precisam usar as credenciais de usuário local do banco de dados nem assumir um perfil do IAM para acessar os dados.

O Identity Center é a abordagem recomendada para autenticação e autorização da força de trabalho em AWS organizações de qualquer tamanho e tipo. Com o Identity Center, você pode criar e gerenciar identidades de usuários ou conectar sua fonte de identidade existente, incluindo Microsoft Active Directory, Okta, Ping Identity JumpCloud, Google Workspace e Microsoft Entra ID (antigo Azure AD). AWS

Para obter mais informações, consulte O que é AWS IAM Identity Center? e propagação confiável de identidade entre aplicativos no Guia do AWS IAM Identity Center usuário.

Atributos e benefícios

A integração do Amazon EMR com o Centro de Identidade do IAM oferece os seguintes benefícios:

- O Amazon EMR fornece credenciais para retransmitir sua identidade do Centro de Identidade a um cluster do EMR.
- O Amazon EMR configura todas as aplicações compatíveis para se autenticarem com as credenciais do cluster.
- O Amazon EMR configura e mantém a segurança das aplicações compatíveis com o protocolo Kerberos, sem que você precise de comandos ou scripts.
- A capacidade de aplicar a autorização no nível de prefixo do Amazon S3 com as identidades do Centro de Identidade em prefixos do S3 gerenciados pelo S3 Access Grants.
- A capacidade de aplicar a autorização em nível de tabela com identidades do Identity Center em tabelas Glue AWS Lake Formation gerenciadas. AWS

Introdução à AWS IAM Identity Center integração com o Amazon EMR

Esta seção ajuda você a configurar o Amazon EMR para integração com o. AWS IAM Identity Center

Tópicos

- Criação de uma instância do Centro de Identidade
- Criação de um perfil do IAM para o Centro de Identidade
- Criação de uma configuração de segurança habilitada para o Centro de Identidade
- Criação e execução de um cluster habilitado para o Centro de Identidade
- Configuração do Lake Formation para um cluster do EMR habilitado para o Centro de Identidade do IAM

- Como trabalhar com o S3 Access Grants em um cluster do EMR habilitado para o Centro de Identidade do IAM

Criação de uma instância do Centro de Identidade

Se ainda não tiver uma, crie uma instância do Centro de Identidade na Região da AWS em que deseja executar o cluster do EMR. Uma instância do Centro de Identidade só pode existir em uma única região para uma Conta da AWS.

Use o AWS CLI comando a seguir para criar uma nova instância chamada*MyInstance*:

```
aws sso-admin create-instance --name MyInstance
```

Criação de um perfil do IAM para o Centro de Identidade

Para integrar o Amazon EMR com AWS IAM Identity Center, crie uma função do IAM que se autentique com o Identity Center a partir do cluster do EMR. O Amazon EMR usa credenciais SigV4 internamente para retransmitir a identidade do Centro de Identidade a serviços downstream, como o AWS Lake Formation. Seu perfil também deve ter as respectivas permissões para invocar os serviços downstream.

Ao criar o perfil, use a seguinte política de permissões:

```
{
  "Statement": [
    {
      "Sid": "IdCPermissions",
      "Effect": "Allow",
      "Action": [
        "sso-oauth:*"
      ],
      "Resource": "*"
    },
    {
      "Sid": "GlueandLakePermissions",
      "Effect": "Allow",
      "Action": [
        "glue:*",
        "lakeformation:GetDataAccess"
      ],
      "Resource": "*"
```

```
    },
    {
      "Sid": "AccessGrantsPermissions",
      "Effect": "Allow",
      "Action": [
        "s3:GetDataAccess",
        "s3:GetAccessGrantsInstanceForPrefix"
      ],
      "Resource": "*"
    }
  ]
}
```

A política de confiança desse perfil permite que o perfil InstanceProfile deixe-o assumir o perfil.

```
{
   "Sid": "AssumeRole",
   "Effect": "Allow",
   "Principal": {
        "AWS": "arn:aws:iam::12345678912:role/EMR_EC2_DefaultRole"
    },
    "Action": [
               "sts:AssumeRole",
               "sts:SetContext"
    ]
 }
```

Criação de uma configuração de segurança habilitada para o Centro de Identidade

Para executar um cluster do EMR com a integração do Centro de Identidade do IAM, use o exemplo de comando a seguir para criar uma configuração de segurança do Amazon EMR que tenha o Centro de Identidade habilitado. Cada configuração é explicada abaixo.

```
aws emr create-security-configuration --name "IdentityCenterConfiguration-with-lf-
accessgrants" --region "us-west-2" --security-configuration '{
 "AuthenticationConfiguration":{
  "IdentityCenterConfiguration":{
   "EnableIdentityCenter":true,
   "IdentityCenterApplicationAssigmentRequired":false,
   "IdentityCenterInstanceARN": "arn:aws:sso:::instance/ssoins-123xxxxxxxxx789",
   "IAMRoleForEMRIdentityCenterApplicationARN": "arn:aws:iam::123456789012:role/tip-
role"
```

```
    }
  },
  "AuthorizationConfiguration": {
   "LakeFormationConfiguration": {
    "EnableLakeFormation": true
   }
  },
  "EncryptionConfiguration": {
   "EnableInTransitEncryption": true,
   "EnableAtRestEncryption": false,
   "InTransitEncryptionConfiguration": {
    "TLSCertificateConfiguration": {
     "CertificateProviderType": "PEM",
     "S3Object": "s3://my-bucket/cert/my-certs.zip"
    }
   }
  }
}'
```

- **`EnableIdentityCenter`**: (obrigatório) habilita a integração do Centro de Identidade.
- **`IdentityCenterApplicationARN`**: (obrigatório) o ARN da instância do Centro de Identidade.
- **`IAMRoleForEMRIdentityCenterApplicationARN`**: (obrigatório) o perfil do IAM que adquire tokens do Centro de Identidade do cluster.
- **`IdentityCenterApplicationAssignmentRequired`** : (booleano) determina se uma atribuição será necessária para usar a aplicação do Centro de Identidade. O valor padrão é `true`.
- **`AuthorizationConfiguration`/`LakeFormationConfiguration`**— Opcionalmente, configure a autorização:
  - **`EnableLakeFormation`**: habilite a autorização do Lake Formation no cluster.

Para habilitar a integração do Centro de Identidade com o Amazon EMR, você deve especificar `EncryptionConfiguration` e `IntransitEncryptionConfiguration`.

Criação e execução de um cluster habilitado para o Centro de Identidade

Agora que configurou o perfil do IAM que se autentica ao Centro de Identidade e criou uma configuração de segurança do Amazon EMR com o Centro de Identidade habilitado, você pode criar e executar seu cluster com reconhecimento de identidade. Para ver as etapas de execução do cluster com a configuração de segurança necessária, consulte Especificar uma configuração de segurança para um cluster.

Opcionalmente, consulte as seguintes seções se deseja usar o cluster habilitado para o Centro de Identidade com outras opções de segurança com suporte do Amazon EMR:

- Como trabalhar com o S3 Access Grants em um cluster do EMR habilitado para o Centro de Identidade do IAM
- Configuração do Lake Formation para um cluster do EMR habilitado para o Centro de Identidade do IAM

Configuração do Lake Formation para um cluster do EMR habilitado para o Centro de Identidade do IAM

Você pode se integrar AWS Lake Formationao seu cluster EMR AWS IAM Identity Center habilitado.

Primeiro, certifique-se de ter uma instância do Centro de Identidade configurada na mesma região do cluster. Para ter mais informações, consulte Criação de uma instância do Centro de Identidade. Você pode encontrar o ARN da instância no console do Centro de Identidade do IAM ao visualizar os detalhes da instância ou usar o seguinte comando para ver os detalhes de todas as instâncias na CLI:

```
aws sso-admin list-instances
```

Em seguida, use o ARN e o ID AWS da sua conta com o comando a seguir para configurar o Lake Formation para ser compatível com o IAM Identity Center:

```
aws lakeformation create-lake-formation-identity-center-configuration --cli-input-json
 file://create-lake-fromation-idc-config.json
json input:
{
    "CatalogId": "account-id/org-account-id",
    "InstanceArn": "identity-center-instance-arn"
}
```

Agora, chame `put-data-lake-settings` e habilite `AllowFullTableExternalDataAccess` com o Lake Formation:

```
aws lakeformation put-data-lake-settings --cli-input-json file://put-data-lake-
settings.json
json input:
```

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier": "admin-ARN"
            }
        ],
        "CreateDatabaseDefaultPermissions": [...],
        "CreateTableDefaultPermissions": [...],
        "AllowExternalDataFiltering": true,
        "AllowFullTableExternalDataAccess": true
    }
}
```

Por fim, conceda permissões completas de tabela ao ARN da identidade do usuário que acessa o cluster do EMR. O ARN contém o ID do usuário do Centro de Identidade. Navegue até o Centro de Identidade no console, selecione Usuários e, em seguida, o usuário para visualizar as configurações de Informações gerais.

Copie o ID do usuário e cole-o no seguinte ARN para *user-id*:

```
arn:aws:identitystore:::user/user-id
```

> **Note**
>
> As consultas no cluster do EMR só funcionam se a identidade do Centro de Identidade do IAM tiver acesso total à tabela protegida do Lake Formation. Se a identidade não tiver acesso total à tabela, a consulta falhará.

Use o seguinte comando para conceder ao usuário acesso total à tabela:

```
aws lakeformation grant-permissions --cli-input-json file://grantpermissions.json
json input:
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:identitystore:::user/user-id"
    },
    "Resource": {
        "Table": {
```

```
            "DatabaseName": "tip_db",
            "Name": "tip_table"
        }
    },
    "Permissions": [
        "ALL"
    ],
    "PermissionsWithGrantOption": [
        "ALL"
    ]
}
```

Como trabalhar com o S3 Access Grants em um cluster do EMR habilitado para o Centro de Identidade do IAM

Você pode integrar o S3 Access Grants ao seu cluster EMR AWS IAM Identity Center habilitado.

Use o S3 Access Grants para autorizar o acesso aos seus conjuntos de dados de clusters que usam o Centro de Identidade. Crie concessões para aumentar as permissões definidas para usuários, grupos e perfis do IAM ou para um diretório corporativo. Para obter mais informações, consulte Using S3 Access Grants with Amazon EMR.

Tópicos

- Como criar uma instância e localização da funcionalidade S3 Access Grants
- Como criar concessões para identidades do Centro de Identidade

Como criar uma instância e localização da funcionalidade S3 Access Grants

Se você ainda não tiver uma, crie uma instância do S3 Access Grants na Região da AWS em que deseja executar seu cluster do EMR.

Use o AWS CLI comando a seguir para criar uma nova instância chamada*MyInstance*:

```
aws s3control-access-grants create-access-grants-instance \
--account-id 12345678912 \
--identity-center-arn "identity-center-instance-arn" \
```

Em seguida, crie uma localização do S3 Access Grants, substituindo os valores vermelhos pelos seus próprios:

```
aws s3control-access-grants create-access-grants-location \
--account-id 12345678912 \
--location-scope s3:// \
--iam-role-arn "access-grant-role-arn" \
--region aa-example-1
```

> **Note**
>
> Defina o parâmetro `iam-role-arn` como o ARN `accessGrantRole`.

Como criar concessões para identidades do Centro de Identidade

Por fim, crie as concessões das identidades que têm acesso ao seu cluster:

```
aws s3control-access-grants create-access-grant \
--account-id 12345678912 \
--access-grants-location-id "default" \
--access-grants-location-configuration S3SubPrefix="s3-bucket-prefix"
--permission READ \
--grantee GranteeType=DIRECTORY_USER,GranteeIdentifier="your-identity-center-user-id"
```

Exemplo de saída:

```
{
"CreatedAt": "2023-09-21T23:47:24.870000+00:00",
"AccessGrantId": "1234-12345-1234-1234567",
"AccessGrantArn": "arn:aws:s3:aa-example-1-1:123456789012:access-grants/default/grant/
xxxx1234-1234-5678-1234-1234567890",
"Grantee": {
"GranteeType": "DIRECTORY_USER",
"GranteeIdentifier": "5678-56789-5678-567890"
},
"AccessGrantsLocationId": "default",
"AccessGrantsLocationConfiguration": {
"S3SubPrefix": "myprefix/*"
},
"Permission": "READ",
"GrantScope": "s3://myprefix/*"
}
```

Considerações e limitações do Amazon EMR com a integração do Centro de Identidade

Considere os seguintes pontos ao usar o Centro de Identidade do IAM com o Amazon EMR:

- A propagação de identidade confiável por meio do Centro de Identidade é compatível com o Amazon EMR 6.15.0 e versões superiores, somente com o Apache Spark.
- Para habilitar clusters do EMR com propagação de identidade confiável, você deve usar o AWS CLI para criar uma configuração de segurança que tenha a propagação de identidade confiável ativada e usar essa configuração de segurança ao iniciar seu cluster. Para ter mais informações, consulte Criação de uma configuração de segurança habilitada para o Centro de Identidade.
- Os clusters do EMR que usam a propagação de identidade confiável só podem invocar serviços que também usam a propagação de identidade confiável.
- Somente o controle de acesso em nível de tabela baseado em AWS Lake Formation está disponível para clusters do EMR que usam propagação de identidade confiável.
- Com clusters do EMR que usam a propagação de identidade confiável, as operações que oferecem suporte ao controle de acesso baseado no Lake Formation com o Apache Spark incluem `SELECT`, `ALTER TABLE` e `DROP TABLE`.
- Com clusters do EMR que usam a propagação de identidade confiável, os controles de acesso baseados no Lake Formation sem compatibilidade com o Apache Spark incluem instruções `INSERT`.
- A propagação de identidade confiável com o Amazon EMR é suportada no seguinte: Regiões da AWS
  - `ap-east-1`: Ásia-Pacífico (Hong Kong)
  - `ap-northeast-1`: Ásia-Pacífico (Tóquio)
  - `ap-northeast-2`: Ásia-Pacífico (Seul)
  - `ap-south-1`: Ásia-Pacífico (Mumbai)
  - `ap-southeast-1`: Ásia-Pacífico (Singapura)
  - `ap-southeast-2`: Ásia-Pacífico (Sydney)
  - `ca-central-1`: Canadá (Central)
  - `eu-central-1`: Europa (Frankfurt)
  - `eu-north-1`: Europa (Estocolmo)
  - `eu-west-1`: Europa (Irlanda)

- `eu-west-2`: Europa (Londres)
- `eu-west-3`: Europa (Paris)
- `me-south-1`: Oriente Médio (Bahrein)
- `sa-east-1`: América do Sul (São Paulo)
- `us-east-1`: Leste dos EUA (Norte da Virgínia)
- `us-east-2`: Leste dos EUA (Ohio)
- `us-west-1`: Oeste dos EUA (Norte da Califórnia)
- `us-west-2`: Oeste dos EUA (Oregon)

Integre o Amazon EMR com AWS Lake Formation

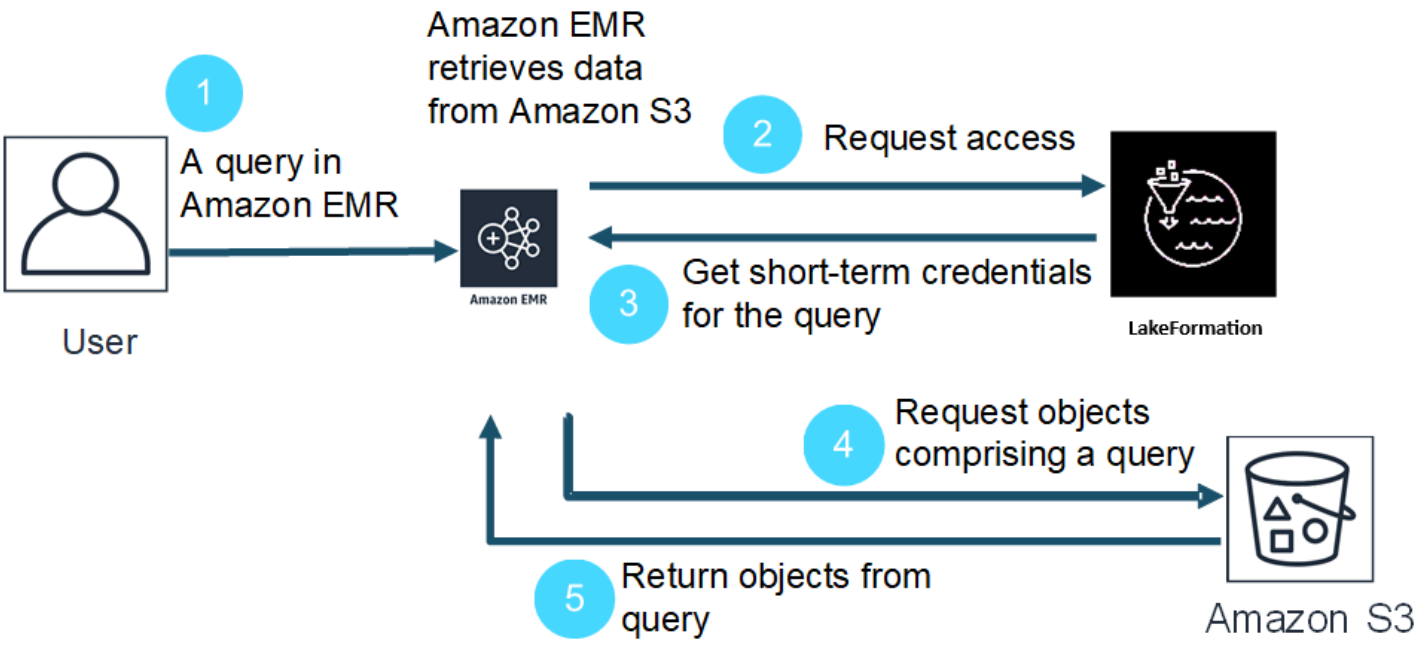
AWS Lake Formation é um serviço gerenciado que ajuda você a descobrir, catalogar, limpar e proteger dados em um data lake do Amazon Simple Storage Service (S3). O Lake Formation fornece acesso refinado em nível de coluna a bancos de dados e tabelas no Glue Data Catalog. AWS Para ter mais informações, consulte O que é o AWS Lake Formation?

Com o Amazon EMR 6.7.0 e versões posteriores, você pode aplicar o controle de acesso baseado no Lake Formation a trabalhos do Spark, Hive e Presto enviados aos clusters do Amazon EMR. Para se integrar ao Lake Formation, é necessário criar um cluster do EMR com um perfil de runtime. O perfil de runtime é um perfil do AWS Identity and Access Management (IAM) que você associa a trabalhos ou consultas do Amazon EMR. Em seguida, o Amazon EMR usa essa função para acessar AWS recursos. Para ter mais informações, consulte Perfis de runtime para etapas ao Amazon EMR.

Como o Amazon EMR funciona com o Lake Formation

Depois de integrar o Amazon EMR com o Lake Formation, você pode executar consultas aos clusters do Amazon EMR com a API ou com o `Step` Studio. SageMaker Em seguida, o Lake Formation fornecerá acesso aos dados por meio de credenciais temporárias para o Amazon EMR. Esse processo chamado de fornecimento de credenciais. Para ter mais informações, consulte O que é o AWS Lake Formation?

Veja a seguir uma visão geral de alto nível sobre como o Amazon EMR obtém acesso aos dados protegidos pelas políticas de segurança do Lake Formation.



1. O usuário envia uma consulta do Amazon EMR para obter dados no Lake Formation.
2. O Amazon EMR solicita credenciais temporárias do Lake Formation para dar acesso aos dados para o usuário.
3. O Lake Formation retorna credenciais temporárias.
4. O Amazon EMR envia a solicitação de consulta para recuperar dados do Amazon S3.
5. O Amazon EMR recebe os dados do Amazon S3, filtra-os e retorna os resultados com base nas permissões de usuário que o usuário definiu no Lake Formation.

Para obter mais informações sobre como adicionar usuários e grupos às políticas do Lake Formation, consulte Granting Data Catalog permissions.

Pré-requisitos

É necessário atender aos seguintes requisitos para integrar o Amazon EMR e o Lake Formation:

- Ative a autorização do perfil de runtime no cluster do Amazon EMR.
- Use o AWS Glue Data Catalog como seu armazenamento de metadados.
- Defina e gerencie permissões no Lake Formation para acessar bancos de dados, tabelas e colunas no AWS Glue Data Catalog. Para ter mais informações, consulte O que é o AWS Lake Formation?

Tópicos

Habilitar o Lake Formation com o Amazon EMR

Com o Amazon EMR 6.15.0 e versões posteriores, ao executar trabalhos do Spark no Amazon EMR em clusters EC2 que acessam dados no AWS Glue Data Catalog, você pode AWS Lake Formation usar para aplicar permissões em nível de tabela, linha, coluna e célula em tabelas baseadas em Hudi, Iceberg ou Delta Lake.

Nesta seção, abordamos como criar uma configuração de segurança e configurar o Lake Formation para trabalhar com o Amazon EMR. Também veremos como iniciar um cluster com a configuração de segurança criada para o Lake Formation.

Etapa 1: configurar um perfil de runtime para o cluster do EMR

Para usar um perfil de runtime para o cluster do EMR, é necessário criar uma configuração de segurança. Com uma configuração de segurança, você pode aplicar opções consistentes de segurança, autorização e autenticação nos clusters.

1. Crie um arquivo chamado `lf-runtime-roles-sec-cfg.json` com a configuração a seguir.

```
{
    "AuthorizationConfiguration": {
        "IAMConfiguration": {
            "EnableApplicationScopedIAMRole": true,
            "ApplicationScopedIAMRoleConfiguration": {
                "PropagateSourceIdentity": true
            }
        },
        "LakeFormationConfiguration": {
            "AuthorizedSessionTagValue": "Amazon EMR"
        }
    },
```

```
    "EncryptionConfiguration": {
            "EnableInTransitEncryption": true,
            "InTransitEncryptionConfiguration": {
            "TLSCertificateConfiguration": {<certificate-configuration>}
        }
    }
}
```

2. Em seguida, para garantir que a etiqueta da sessão possa autorizar o Lake Formation, defina a propriedade `LakeFormationConfiguration/AuthorizedSessionTagValue` como `Amazon EMR`.
3. Use o comando a seguir para criar uma configuração de segurança do Amazon EMR.

```
aws emr create-security-configuration \
--name 'iamconfig-with-iam-lf' \
--security-configuration file://lf-runtime-roles-sec-cfg.json
```

Como alternativa, é possível usar o console do Amazon EMR para criar uma configuração de segurança com configurações personalizadas.

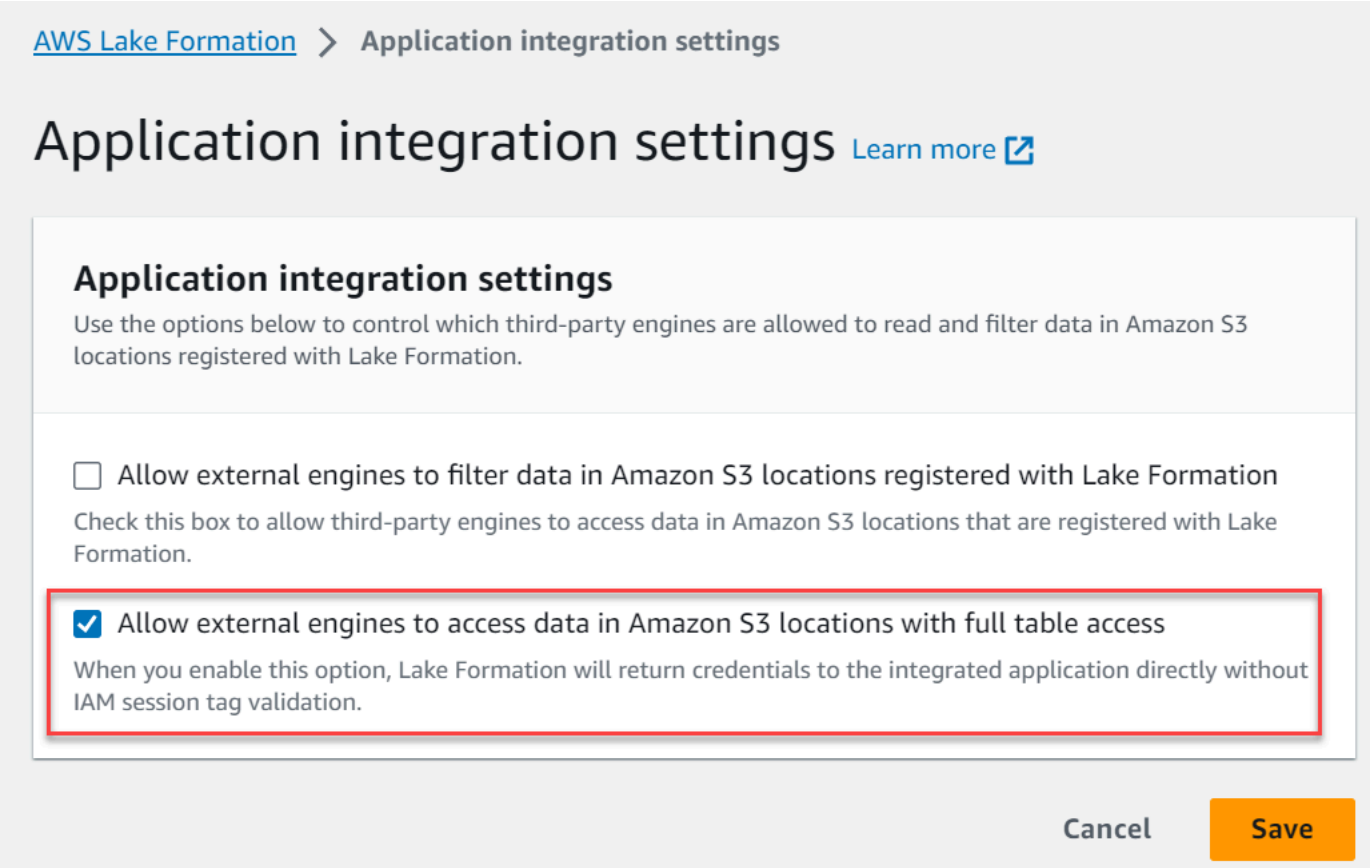
Etapa 2: iniciar um cluster do Amazon EMR

Agora, você já pode iniciar um cluster do EMR com a configuração de segurança criada na etapa anterior. Para obter mais informações sobre configurações de segurança, consulte Usar configurações de segurança para definir a segurança do cluster e Perfis de runtime para etapas ao Amazon EMR.

Etapa 3a: configurar permissões no nível de tabela baseadas no Lake Formation com perfis de runtime do Amazon EMR

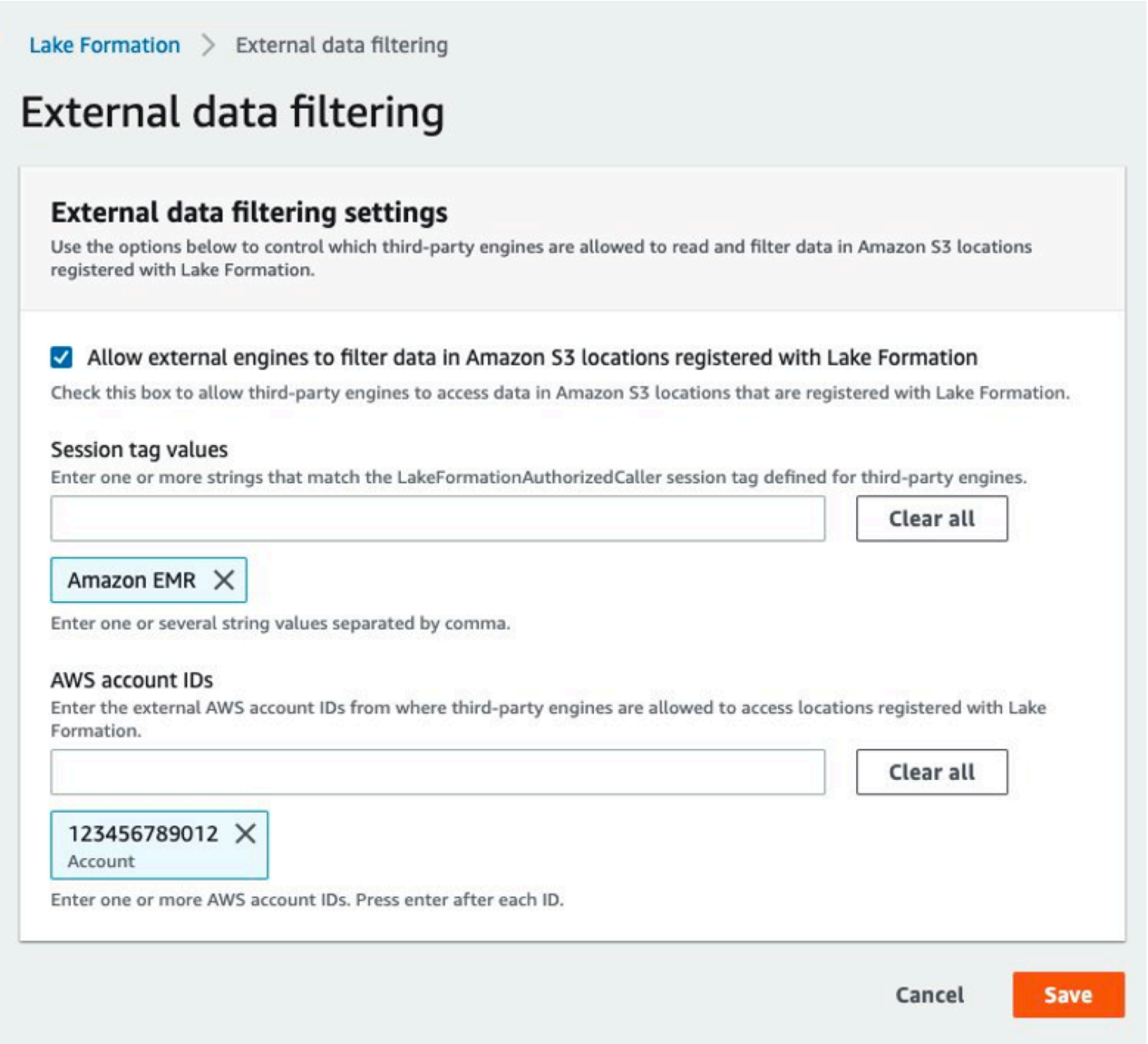
Se você não precisar de um controle de acesso refinado no nível de coluna, linha ou célula, poderá configurar permissões no nível de tabela com o Glue Data Catalog. Para habilitar o acesso em nível de tabela, navegue até o AWS Lake Formation console e selecione a opção Configurações de integração de aplicativos na seção Administração na barra lateral. Em seguida, habilite a seguinte opção e escolha Salvar:

Permitir que mecanismos externos acessem dados em locais do Amazon S3 com acesso total à tabela



Etapa 3b: configurar permissões no nível de coluna, linha ou célula baseadas no Lake Formation com perfis de runtime do Amazon EMR

Para aplicar permissões no nível de tabela e coluna com o Lake Formation, o administrador do data lake no Lake Formation deve definir o `Amazon EMR` como o valor da configuração da tag de sessão, `AuthorizedSessionTagValue`. O Lake Formation usa essa etiqueta de sessão para autorizar os chamadores e fornecer acesso ao data lake. Você pode definir essa etiqueta de sessão na seção Filtragem de dados externos do console do Lake Formation. Substitua *`123456789012`* pelo ID de sua própria Conta da AWS .



Etapa 4: Configurar subsídios do AWS Glue e do Lake Formation para funções de tempo de execução do Amazon EMR

Para continuar com a configuração do controle de acesso baseado em Lake Formation com funções de tempo de execução do Amazon EMR, você deve configurar subsídios do AWS Glue e do Lake Formation para funções de tempo de execução do Amazon EMR. Para permitir que os perfis de runtime do IAM interajam com o Lake Formation, conceda a eles acesso com `lakeformation:GetDataAccess` e `glue:Get*`.

As permissões do Lake Formation controlam o acesso aos recursos do AWS Glue Data Catalog, aos locais do Amazon S3 e aos dados subjacentes nesses locais. As permissões do IAM controlam o acesso às APIs e aos recursos do Lake Formation e do AWS Glue. Embora você possa ter a permissão do Lake Formation para acessar uma tabela no catálogo de dados (SELECT), a operação falhará se você não tiver a permissão do IAM na API `glue:Get*`. Para obter mais detalhes sobre o controle de acesso do Lake Formation, consulte a visão geral do controle de acesso do Lake Formation.

1. Crie o arquivo `emr-runtime-roles-lake-formation-policy.json` com o conteúdo a seguir.

```
{
    "Version": "2012-10-17",
    "Statement": {
        "Sid": "LakeFormationManagedAccess",
        "Effect": "Allow",
        "Action": [
            "lakeformation:GetDataAccess",
            "glue:Get*",
            "glue:Create*",
            "glue:Update*"
        ],
        "Resource": "*"
    }
}
```

2. Crie a política do IAM relacionada ao IAM.

```
aws iam create-policy \
--policy-name emr-runtime-roles-lake-formation-policy \
--policy-document file://emr-runtime-roles-lake-formation-policy.json
```

3. Para atribuir essa política aos perfis de runtime do IAM, siga as etapas descritas em Managing AWS Lake Formation permissions.

Já é possível usar perfis de runtime e o Lake Formation para aplicar permissões em nível de tabela e coluna. Você também pode usar uma identidade de origem para controlar ações e monitorar operações com AWS CloudTrail. Para obter um end-to-end exemplo detalhado, consulte Introdução às funções de tempo de execução para as etapas do Amazon EMR.

Apache Hudi e Lake Formation

As versões 6.15.0 e superiores do Amazon EMR incluem suporte para controle de acesso refinado baseado no Apache Hudi quando você lê e grava dados AWS Lake Formation com o Spark SQL. O Amazon EMR oferece suporte ao controle de acesso no nível de tabela, linha, coluna e célula com o Apache Hudi. Com esse recurso, você pode executar consultas de instantâneos em copy-on-write tabelas para consultar o instantâneo mais recente da tabela em um determinado instante de confirmação ou compactação.

Atualmente, um cluster Amazon EMR habilitado para Lake Formation deve recuperar a coluna de tempo de confirmação da Hudi para realizar consultas incrementais e consultas de viagem no tempo. Ele não suporta a `timestamp as of` sintaxe e a função do Spark. `Spark.read()` A sintaxe correta é `select * from table where _hoodie_commit_time <= point_in_time`. Para obter mais informações, consulte Consultas de viagem no tempo pontual na tabela Hudi.

A matriz de suporte a seguir lista alguns dos principais recursos do Apache Hudi com o Lake Formation:

	Copiar na gravação	mesclar na leitura
Consultas de snapshots: Spark SQL	✓	✓
Consultas otimizadas para leitura: Spark SQL	✓	✓
Consultas incrementais	✓	✓
Consultas de viagem no tempo	✓	✓
Tabelas de metadados	✓	✓
Comandos **`INSERT`** de DML	✓	✓
Comandos de DDL		
Consultas de fontes de dados do Spark		
Gravações na fonte de dados do Spark		

Consultar tabelas do Hudi

Esta seção mostra como você pode executar as consultas com suporte descritas acima em um cluster habilitado para Lake Formation. A tabela deve ser uma tabela de catálogo registrada.

1. Para iniciar o shell Spark, use os comandos a seguir.

```
spark-sql
--jars /usr/lib/hudi/hudi-spark-bundle.jar \
--conf spark.serializer=org.apache.spark.serializer.KryoSerializer \
--conf
spark.sql.catalog.spark_catalog=org.apache.spark.sql.hudi.catalog.HoodieCatalog \
--conf
spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension,com.amazonaws.em
 \
--conf spark.sql.catalog.spark_catalog.lf.managed=true
```

   Se você quiser que o Lake Formation use o servidor de registros para gerenciar seu catálogo do Spark, `spark.sql.catalog.<managed_catalog_name>.lf.managed` defina como true.

2. Para consultar o instantâneo mais recente das copy-on-write tabelas, use os comandos a seguir.

```
SELECT * FROM my_hudi_cow_table
```

```
spark.read.table("my_hudi_cow_table")
```

3. Para consultar os dados compactados mais recentes das tabelas `MOR`, você pode consultar a tabela otimizada para leitura que tem o sufixo `_ro`:

```
SELECT * FROM my_hudi_mor_table_ro
```

```
spark.read.table("my_hudi_mor_table_ro")
```

> **Note**
>
> A performance das leituras nos clusters do Lake Formation pode ser mais lenta devido às otimizações sem suporte. Esses atributos incluem listagem de arquivos com base nos

metadados do Hudi e salto de dados. É recomendável testar a performance da aplicação para garantir que ela atenda aos seus requisitos.

Apache Iceberg e Lake Formation

As versões 6.15.0 e superiores do Amazon EMR incluem suporte para controle de acesso refinado baseado no Apache Iceberg quando você lê e grava dados AWS Lake Formation com o Spark SQL. O Amazon EMR oferece suporte ao controle de acesso no nível de tabela, linha, coluna e célula com o Apache Iceberg. Com esse recurso, você pode executar consultas de instantâneos em copy-on-write tabelas para consultar o instantâneo mais recente da tabela em um determinado instante de confirmação ou compactação.

Se você quiser usar o formato Iceberg, defina as configurações a seguir. Substitua *DB_LOCATION* pelo caminho do Amazon S3 onde suas tabelas do Iceberg estão localizadas e os espaços reservados para a região e o ID da conta por seus próprios valores.

```
spark-sql \
--conf
 spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions,com.ama

--conf spark.sql.catalog.iceberg_catalog=org.apache.iceberg.spark.SparkCatalog
--conf spark.sql.catalog.iceberg_catalog.warehouse=s3://DB_LOCATION
--conf spark.sql.catalog.iceberg_catalog.catalog-
impl=org.apache.iceberg.aws.glue.GlueCatalog
--conf spark.sql.catalog.iceberg_catalog.io-impl=org.apache.iceberg.aws.s3.S3FileIO
--conf spark.sql.catalog.iceberg_catalog.glue.account-id=ACCOUNT_ID
--conf spark.sql.catalog.iceberg_catalog.glue.id=ACCOUNT_ID
--conf spark.sql.catalog.iceberg_catalog.client.assume-role.region=AWS_REGION
--conf spark.sql.secureCatalog=iceberg_catalog
```

Se você quiser que o Lake Formation use o servidor de registros para gerenciar seu catálogo do Spark, `spark.sql.catalog.<managed_catalog_name>.lf.managed` defina como true.

Você também deve ter cuidado para NÃO aprovar as seguintes configurações de assumir perfis:

```
--conf spark.sql.catalog.my_catalog.client.assume-role.region
--conf spark.sql.catalog.my_catalog.client.assume-role.arn
```

```
--conf spark.sql.catalog.my_catalog.client.assume-
role.tags.LakeFormationAuthorizedCaller
```

A matriz de apoio a seguir lista alguns dos principais recursos do Apache Iceberg com o Lake Formation:

	Copiar na gravação	mesclar na leitura
Consultas de snapshots: Spark SQL	✓	✓
Consultas otimizadas para leitura: Spark SQL	✓	✓
Consultas incrementais	✓	✓
Consultas de viagem no tempo	✓	✓
Tabelas de metadados	✓	✓
Comandos **`INSERT`** de DML	✓	✓
Comandos de DDL		
Consultas de fontes de dados do Spark		
Gravações na fonte de dados do Spark		

Delta Lake e Lake Formation

As versões 6.15.0 e superiores do Amazon EMR incluem suporte para controle de acesso refinado AWS Lake Formation com base no Delta Lake quando você lê e grava dados com o Spark SQL. O Amazon EMR oferece suporte ao controle de acesso no nível de tabela, linha, coluna e célula com o Delta Lake. Com esse recurso, você pode executar consultas de instantâneos em copy-on-write tabelas para consultar o instantâneo mais recente da tabela em um determinado instante de confirmação ou compactação.

Para usar o Delta Lake com o Lake Formation, execute o comando a seguir.

```
spark-sql \
```

```
--conf
 spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserver.co
 \
--conf spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog
 \
--conf spark.sql.catalog.spark_catalog.lf.managed=true
```

Se você quiser que o Lake Formation use o servidor de registros para gerenciar seu catálogo do Spark, `spark.sql.catalog.<managed_catalog_name>.lf.managed` defina como true.

A seguinte matriz de apoio lista alguns dos principais recursos do Delta Lake com o Lake Formation:

	Copiar na gravação	mesclar na leitura
Consultas de snapshots: Spark SQL	✓	✓
Consultas otimizadas para leitura: Spark SQL	✓	✓
Consultas incrementais	Não suportado	Sem suporte
Consultas de viagem no tempo	Não suportado	Sem suporte
Tabelas de metadados	✓	✓
Comandos **`INSERT`** de DML	✓	✓
Comandos de DDL		
Consultas de fontes de dados do Spark		
Gravações na fonte de dados do Spark		

Criação de uma tabela Delta Lake no AWS Glue Data Catalog

O Amazon EMR com Lake Formation não oferece suporte a comandos DDL e à criação de tabelas Delta. Siga estas etapas para criar tabelas no AWS Glue Data Catalog.

1. Use o exemplo a seguir para criar uma tabela Delta. Certifique-se de que sua localização no S3 exista.

```
spark-sql \
--conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \
--conf
 "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"

> CREATE DATABASE if not exists <DATABASE_NAME> LOCATION 's3://<S3_LOCATION>/
transactionaldata/native-delta/<DATABASE_NAME>/';
> CREATE TABLE <TABLE_NAME> (x INT, y STRING, z STRING) USING delta;
> INSERT INTO <TABLE_NAME> VALUES (1, 'a1', 'b1');
```

2. Para ver os detalhes da sua tabela, acesse https://console.aws.amazon.com/glue/.
3. No painel de navegação à esquerda, expanda Catálogo de Dados, escolha Tabelas e escolha a tabela que você criou. Em Esquema, você verá que a tabela Delta que você criou com o Spark armazena todas as colunas em um tipo de dados do `array<string>` AWS Glue.
4. Para definir filtros em nível de coluna e célula no Lake Formation, remova a `col` coluna do seu esquema e, em seguida, adicione as colunas que estão no esquema da tabela. Neste exemplo, adicione as colunas `x`, `y`, `z` e.

Considerações sobre o Amazon EMR com o Lake Formation

Considere o seguinte ao usar o Amazon EMR com. AWS Lake Formation

- O controle de acesso no nível de tabela está disponível em clusters com versões 6.13 e superiores do Amazon EMR.
- O controle de acesso refinado no nível de linha, coluna e célula está disponível em clusters com versões 6.15 e superiores do Amazon EMR.
- Os usuários com acesso a uma tabela podem acessar todas as propriedades da tabela. Se você tiver controle de acesso baseado no Lake Formation em uma tabela, revise a tabela para garantir que as propriedades não contenham dados ou informações sigilosas.
- Os clusters do Amazon EMR com Lake Formation não oferecem suporte ao retorno do Spark para o HDFS quando o Spark coleta estatísticas de tabelas. Isso normalmente ajuda a otimizar a performance da consulta.
- As operações que oferecem suporte a controles de acesso baseados no Lake Formation com tabelas não governadas do Apache Spark incluem `INSERT INTO` e `INSERT OVERWRITE`.

- As operações que oferecem suporte a controles de acesso baseados no Lake Formation com Apache Spark e Apache Hive incluem `SELECT`, `DESCRIBE`, `SHOW DATABASE`, `SHOW TABLE`, `SHOW COLUMN` e `SHOW PARTITION`.
- O Amazon EMR não oferece suporte ao controle de acesso às seguintes operações baseadas no Lake Formation:
  - Grava em tabelas controladas
  - O Amazon EMR não oferece suporte a `CREATE TABLE`. O Amazon EMR 6.10.0 e versões superiores oferecem suporte a `ALTER TABLE`.
  - Instruções DML que não sejam comandos `INSERT`.
- Há diferenças de performance entre a mesma consulta com e sem controle de acesso baseado no Lake Formation.

Integrar o Amazon EMR com o Apache Ranger

Desde a versão 5.32.0 do Amazon EMR, você pode iniciar um cluster que se integre nativamente ao Apache Ranger. O Apache Ranger é uma estrutura de código aberto para habilitar, monitorar e gerenciar uma segurança de dados abrangente em toda a plataforma Hadoop. Para obter mais informações, consulte Apache Ranger. Com a integração nativa, você pode trazer seu próprio Apache Ranger para aplicar um controle de acesso detalhado aos dados no Amazon EMR.

Esta seção fornece uma visão geral conceitual da integração do Amazon EMR com o Apache Ranger. Também inclui os pré-requisitos e as etapas necessárias para iniciar um cluster do Amazon EMR integrado ao Apache Ranger.

Integrar o Amazon EMR de maneira nativa com o Apache Ranger oferece os seguintes benefícios principais:

- Controle de acesso refinado aos bancos de dados e tabelas do Hive Metastore, que permite definir políticas de filtragem de dados no nível de banco de dados, tabela e coluna para aplicações Apache Spark e Apache Hive. A filtragem em nível de linha e o mascaramento de dados são compatíveis com aplicações Hive.
- A capacidade de usar suas políticas atuais do Hive diretamente com o Amazon EMR para aplicações Hive.
- Controle de acesso aos dados do Amazon S3 no nível do prefixo e do objeto, o que permite definir políticas de filtragem de dados para acesso aos dados do S3 usando o sistema de arquivos do EMR.

- A capacidade de usar o CloudWatch Logs para auditoria centralizada.
- O Amazon EMR instala e gerencia os plug-ins do Apache Ranger por você.

Apache Ranger

O Apache Ranger é um framework para habilitar, monitorar e gerenciar uma segurança de dados abrangente em toda a plataforma Hadoop.

O Apache Ranger tem os seguintes atributos:

- Administração de segurança centralizada para gerenciar todas as tarefas relacionadas à segurança em uma IU central ou usando APIs REST.
- Autorização refinada para realizar uma ação ou operação específica usando um componente ou ferramenta do Hadoop, gerenciada por meio de uma ferramenta de administração central.
- Um método de autorização padronizado em todos os componentes do Hadoop.
- Suporte aprimorado para diversos métodos de autorização.
- Auditoria centralizada do acesso do usuário e das ações administrativas (relacionadas à segurança) em todos os componentes do Hadoop.

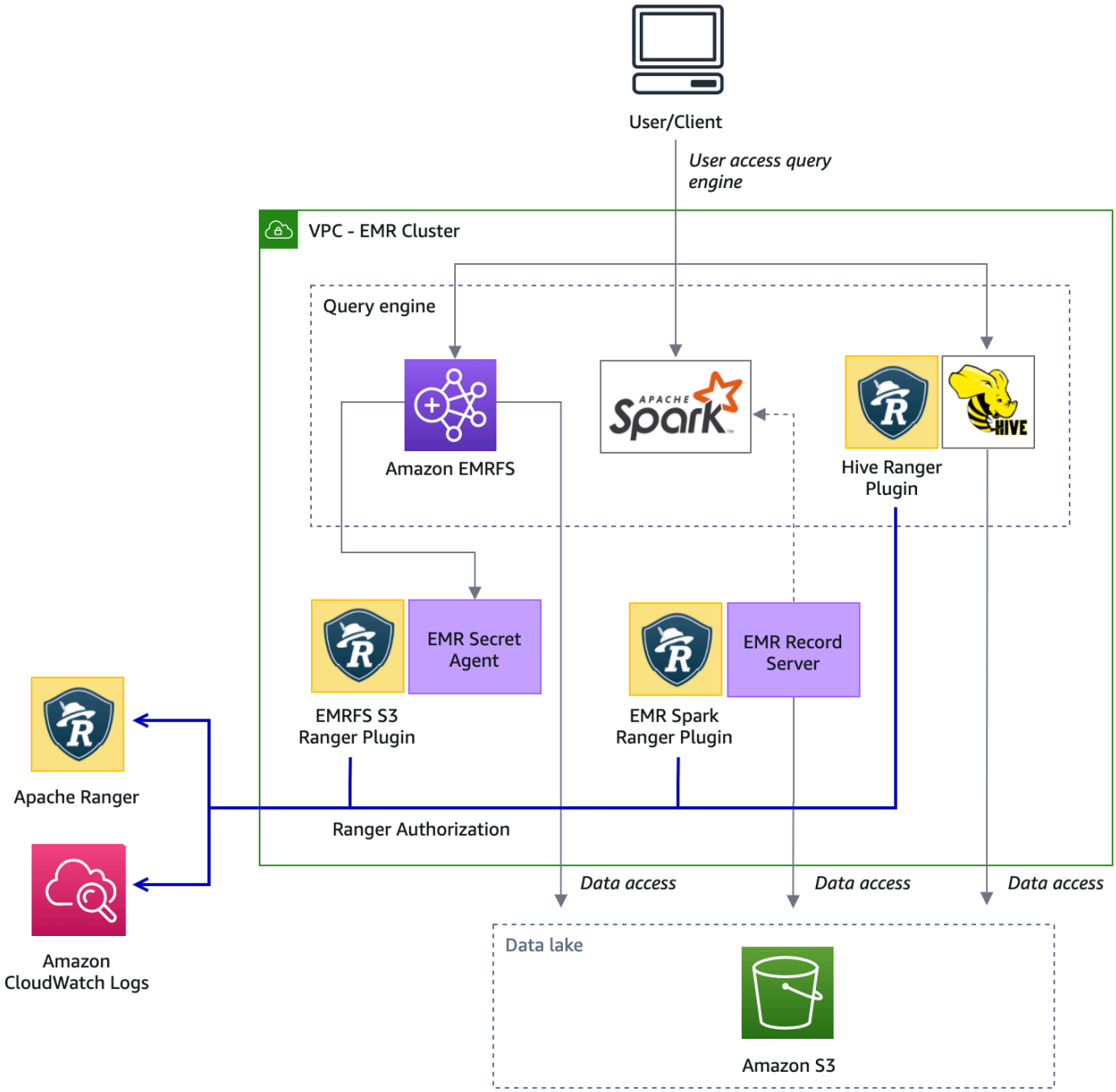
O Apache Ranger usa dois componentes principais para autorização:

- Servidor de administração de políticas Apache Ranger: esse servidor permite definir as políticas de autorização para aplicações Hadoop. Ao fazer a integração com o Amazon EMR, você pode definir e aplicar políticas para que o Apache Spark e o Hive acessem o Hive Metastore e acessem o sistema de arquivos do EMR (EMRFS) de dados do Amazon S3. É possível configurar um novo servidor de administração de políticas Apache Ranger ou usar um já existente para se integrar ao Amazon EMR.
- Plug-in Apache Ranger: esse plug-in valida o acesso de um usuário em relação às políticas de autorização definidas no servidor de administração de políticas do Apache Ranger. O Amazon EMR instala e configura automaticamente o plug-in Apache Ranger para cada aplicação Hadoop selecionada na configuração do Apache Ranger.

Tópicos

- Arquitetura da integração do Amazon EMR com o Apache Ranger
- Componentes do Amazon EMR

Arquitetura da integração do Amazon EMR com o Apache Ranger



Componentes do Amazon EMR

O Amazon EMR possibilita um controle de acesso refinado com o Apache Ranger por meio dos componentes a seguir. Consulte o diagrama de arquitetura para ter uma representação visual desses componentes do Amazon EMR com os plug-ins do Apache Ranger.

Agente secreto: o agente secreto armazena e distribui segredos com segurança para outros componentes ou aplicações do Amazon EMR. Os segredos podem incluir credenciais temporárias de usuário, chaves de criptografia ou tickets Kerberos. O agente secreto é executado em todos os nós do cluster e intercepta chamadas ao serviço de metadados da instância. Para solicitações às credenciais do perfil do perfil de instância, o agente secreto fornece as credenciais dependendo do usuário solicitante e dos recursos solicitados após autorizar a solicitação com o plug-in EMRFS S3 do Ranger. O agente secreto é executado como o usuário *`emrsecretagent`* e grava logs no diretório /emr/secretagent/log. O processo depende de um conjunto específico de regras `iptables` para funcionar. É importante garantir que não `iptables` esteja desabilitado. Se você personalizar a configuração `iptables`, as regras da tabela NAT deverão ser preservadas e deixadas inalteradas.

Servidor de registros do EMR: p servidor de registros recebe solicitações para acessar dados do Spark. Em seguida, ele autoriza as solicitações encaminhando os recursos solicitados ao plug-in Spark Ranger para Amazon EMR. O servidor de registros lê dados do Amazon S3 e retorna dados filtrados que o usuário está autorizado a acessar com base na política do Ranger. O servidor de registros é executado em cada nó do cluster como usuário emr_record_server e grava registros no diretório /var/log/emr-record-server.

Suporte a aplicações e limitações

Aplicações compatíveis

A integração entre o Amazon EMR e o Apache Ranger, na qual o EMR instala os plug-ins do Ranger, atualmente, oferece suporte às seguintes aplicações:

- Apache Spark (disponível no EMR 5.32+ e EMR 6.3+)
- Apache Hive (disponível no EMR 5.32+ e EMR 6.3+)
- Acesso ao S3 por meio do EMRFS (disponível no EMR 5.32+ e EMR 6.3+)

As seguintes aplicações podem ser instaladas em um cluster do EMR e talvez precisem ser configuradas para atender a suas necessidades de segurança:

- Apache Hadoop (disponível no EMR 5.32+ e EMR 6.3+, inclusive YARN e HDFS)
- Apache Livy (disponível no EMR 5.32+ e EMR 6.3+)
- Apache Zeppelin (disponível no EMR 5.32+ e EMR 6.3+)
- Apache Hue (disponível no EMR 5.32+ e EMR 6.3+)

- Ganglia (disponível no EMR 5.32+ e EMR 6.3+)
- HCatalog (disponível no EMR 5.32+ e EMR 6.3+)
- Mahout (disponível no EMR 5.32+ e EMR 6.3+)
- MXNet (disponível no EMR 5.32+ e EMR 6.3+)
- TensorFlow (Disponível com EMR 5.32+ e EMR 6.3+)
- Tez (disponível no EMR 5.32+ e EMR 6.3+)
- Trino (disponível no EMR 6.7+)
- ZooKeeper (Disponível com EMR 5.32+ e EMR 6.3+)

> **Important**
>
> As aplicações listadas acima são as únicas com suporte no momento. Para garantir a segurança do cluster, você tem permissão para criar um cluster do EMR somente com as aplicações da lista acima quando o Apache Ranger está habilitado.
> No momento, não há suporte para outros aplicativos. Para garantir a segurança do cluster, tentar instalar outras aplicações causará a rejeição do cluster.

Atributos compatíveis

Os seguintes atributos do Amazon EMR podem ser usados com o Amazon EMR e o Apache Ranger:

- Criptografia de dados em repouso e em trânsito
- Autenticação Kerberos (obrigatória)
- Grupos de instâncias, frotas de instâncias e instâncias spot
- Reconfiguração de aplicações em um cluster em execução
- Criptografia do lado do servidor (SSE) do EMRFS

> **Note**
>
> As configurações de criptografia do Amazon EMR governam o SSE. Para obter mais informações, consulte Encryption Options.

Limitações de aplicação

Há várias limitações que você deve conhecer ao integrar o Amazon EMR e o Apache Ranger:

- No momento, você não pode usar o console para criar uma configuração de segurança que especifique a opção de integração do AWS Ranger no. AWS GovCloud (US) Region A configuração de segurança do pode ser feita usando a CLI.
- O Kerberos precisa estar instalado no cluster.
- As UIs de aplicativos (interfaces de usuário), como a interface do YARN Resource Manager, a interface do usuário do HDFS e a NameNode interface do usuário do Livy, não são definidas com autenticação por padrão.
- As permissões padrão `umask` do HDFS são configuradas para que os objetos criados sejam definidos como `world wide readable` por padrão.
- O Amazon EMR não oferece suporte ao modo de alta disponibilidade (múltiplo primário) com o Apache Ranger.
- Para ver outras limitações, consulte as limitações de cada aplicação.

> **Note**
>
> As configurações de criptografia do Amazon EMR governam o SSE. Para obter mais informações, consulte Encryption Options.

Limitações de plug-in

Cada plug-in tem limitações específicas. Para ver as limitações do plug-in Apache Hive, consulte as limitações do plug-in Apache Hive. Para ver as limitações do plug-in Apache Spark, consulte as limitações do plug-in Apache Spark. Para ver as limitações do plug-in EMRFS S3, consulte as limitações do plug-in EMRFS S3.

Configurar o Amazon EMR para Apache Ranger

Antes de instalar o Apache Ranger, leia as informações desta seção para garantir que o Amazon EMR esteja configurado corretamente.

Tópicos

- Configurar o servidor do Ranger Admin

- Perfis do IAM para integração nativa com o Apache Ranger
- Criar a configuração de segurança do EMR
- Armazenar certificados TLS no AWS Secrets Manager
- Iniciar um cluster do EMR
- Configurar o Zeppelin para clusters do Amazon EMR habilitados para Apache Ranger
- Problemas conhecidos

Configurar o servidor do Ranger Admin

Para a integração com o Amazon EMR, os plug-ins da aplicação Apache Ranger devem se comunicar com o servidor de administração usando TLS/SSL.

Pré-requisito: habilitar SSL do servidor Ranger Admin

O Apache Ranger no Amazon EMR exige comunicação SSL bidirecional entre os plug-ins e o servidor Ranger Admin. Para garantir que os plug-ins se comuniquem com o servidor Apache Ranger via SSL, habilite o seguinte atributo em ranger-admin-site .xml no servidor Ranger Admin.

```
<property>
    <name>ranger.service.https.attrib.ssl.enabled</name>
    <value>true</value>
</property>
```

Além disso, as configurações a seguir são necessárias.

```
<property>
    <name>ranger.https.attrib.keystore.file</name>
    <value>_<PATH_TO_KEYSTORE>_</value>
</property>

<property>
    <name>ranger.service.https.attrib.keystore.file</name>
    <value>_<PATH_TO_KEYSTORE>_</value>
</property>

<property>
    <name>ranger.service.https.attrib.keystore.pass</name>
    <value>_<KEYSTORE_PASSWORD>_</value>
</property>
```

```
<property>
    <name>ranger.service.https.attrib.keystore.keyalias</name>
    <value><PRIVATE_CERTIFICATE_KEY_ALIAS></value>
</property>

<property>
    <name>ranger.service.https.attrib.clientAuth</name>
    <value>want</value>
</property>

<property>
    <name>ranger.service.https.port</name>
    <value>6182</value>
</property>
```

Certificados TLS

A integração do Apache Ranger com o Amazon EMR exige que o tráfego dos nós do Amazon EMR para o servidor Ranger Admin seja criptografado usando TLS e que os plug-ins do Ranger sejam autenticados no servidor Apache Ranger usando autenticação TLS mútua bidirecional. O serviço Amazon EMR precisa do certificado público do servidor Ranger Admin (especificado no exemplo anterior) e do certificado privado.

Certificados de plug-in do Apache Ranger

Os certificados TLS públicos de plug-in do Apache Ranger devem estar acessíveis ao servidor Apache Ranger Admin para validar quando os plug-ins se conectam. Há três métodos diferentes para isso.

Método 1: configurar um armazenamento confiável no servidor Apache Ranger Admin

Preencha as seguintes configurações em ranger-admin-site .xml para configurar um armazenamento confiável.

```
<property>
    <name>ranger.truststore.file</name>
    <value><LOCATION TO TRUSTSTORE></value>
</property>

<property>
    <name>ranger.truststore.password</name>
    <value><PASSWORD FOR TRUSTSTORE></value>
```

```
</property>
```

Método 2: carregar o certificado no Java cacerts truststore

Se o servidor Ranger Admin não especificar um armazenamento confiável em suas opções da JVM, você poderá colocar os certificados públicos do plug-in no armazenamento cacerts padrão.

Método 3: criar um armazenamento confiável e especificar como parte das opções da JVM

Em `{RANGER_HOME_DIRECTORY}/ews/ranger-admin-services.sh`, modifique `JAVA_OPTS` para incluir "`-Djavax.net.ssl.trustStore=<TRUSTSTORE_LOCATION>`" e "`-Djavax.net.ssl.trustStorePassword=<TRUSTSTORE_PASSWORD>`". Por exemplo, adicione a linha a seguir após o JAVA_OPTS atual.

```
JAVA_OPTS=" ${JAVA_OPTS} -Djavax.net.ssl.trustStore=${RANGER_HOME}/truststore/
truststore.jck -Djavax.net.ssl.trustStorePassword=changeit"
```

> **Note**
>
> Essa especificação pode expor a senha do truststore se algum usuário conseguir fazer login no servidor Apache Ranger Admin e ver os processos em execução, como ao usar o comando `ps`.

Usar certificados autoassinados

Não é recomendável usar certificados autoassinados como certificados. Os certificados autoassinados não podem ser revogados e podem não estar em conformidade com os requisitos internos de segurança.

Instalação da definição de serviço

Uma definição de serviço é usada pelo servidor Ranger Admin para descrever os atributos das políticas de uma aplicação. As políticas são então armazenadas em um repositório de políticas para que os clientes baixem.

Para poder configurar as definições de serviço, as chamadas REST deverão ser feitas para o servidor Ranger Admin. Consulte Apache Ranger PublicAPIsv2 para ver as APIs necessárias na seção a seguir.

Instalar a definição de serviço do Apache Spark

Para instalar a definição de serviço do Apache Spark, consulte Plug-in Apache Spark.

Instalar a definição de serviço do EMRFS

Para instalar a definição de serviço do S3 para Amazon EMR, consulte Plug-in EMRFS S3.

Usar a definição de serviço do Hive

O Apache Hive pode usar a definição de serviço do Ranger já existente que vem com o Apache Ranger 2.0 e versões posteriores. Para ter mais informações, consulte Plug-in Apache Hive.

Regras de tráfego da rede

Quando o Apache Ranger é integrado ao cluster do EMR, o cluster precisa se comunicar com outros servidores e com a AWS.

Todos os nós do Amazon EMR, inclusive os nós centrais e de tarefa, devem ser capazes de se comunicar com os servidores Apache Ranger Admin para baixar as políticas. Se o administrador do Apache Ranger estiver em execução no Amazon EC2, você precisará atualizar o grupo de segurança para poder receber tráfego do cluster do EMR.

Além de se comunicar com o servidor Ranger Admin, todos os nós precisam ser capazes de se comunicar com os seguintes serviços: AWS

- Amazon S3
- AWS KMS (se estiver usando o EMRFS SSE-KMS)
- Amazon CloudWatch
- AWS STS

Se você planeja executar seu cluster do EMR em uma sub-rede privada, configure a VPC para poder se comunicar com esses serviços usando AWS PrivateLink e endpoints da VPC, conforme o Guia do Usuário da Amazon VPC ou usando a instância de conversão de endereços de rede (NAT), seguindo o Guia do usuário da Amazon VPC.

Perfis do IAM para integração nativa com o Apache Ranger

A integração entre o Amazon EMR e o Apache Ranger depende de três perfis principais que você deve criar antes de iniciar o cluster:

- Um perfil de instância do Amazon EC2 personalizado para o Amazon EMR
- Um perfil do IAM para mecanismos do Apache Ranger

- Uma função do IAM para outros AWS serviços

Esta seção fornece uma visão geral desses perfis e das políticas que devem ser incluídas para cada perfil do IAM. Para obter mais informações sobre como criar esses perfis, consulte Configurar o servidor do Ranger Admin.

Perfil de instância do EC2

O Amazon EMR usa um perfil de serviço do IAM para realizar ações a seu favor a fim de provisionar e gerenciar clusters. O perfil de serviço para instâncias do EC2 do cluster, também chamada de perfil de instância do EC2 para Amazon EMR, é um tipo especial de perfil de serviço atribuído ao iniciar cada instância do EC2 do cluster.

Para definir permissões para a interação do cluster EMR com dados do Amazon S3 e com o metastore Hive protegido pelo Apache Ranger e AWS outros serviços, defina um perfil de instância EC2 personalizado para usar em vez do quando você iniciar seu cluster. `EMR_EC2_DefaultRole`

Para obter mais informações, consulte Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2) e Personalizar perfis do IAM.

Você precisa adicionar as seguintes instruções ao perfil de instância EC2 padrão do Amazon EMR para poder marcar sessões e acessar AWS Secrets Manager o que armazena certificados TLS.

```
    {
      "Sid": "AllowAssumeOfRolesAndTagging",
      "Effect": "Allow",
      "Action": ["sts:TagSession", "sts:AssumeRole"],
      "Resource": [
        "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<RANGER_ENGINE-
PLUGIN_DATA_ACCESS_ROLE_NAME>",
        "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<RANGER_USER_ACCESS_ROLE_NAME>"
      ]
    },
    {
        "Sid": "AllowSecretsRetrieval",
        "Effect": "Allow",
        "Action": "secretsmanager:GetSecretValue",
        "Resource": [

 "arn:aws:secretsmanager:<REGION>:<AWS_ACCOUNT_ID>:secret:<PLUGIN_TLS_SECRET_NAME>*",

 "arn:aws:secretsmanager:<REGION>:<AWS_ACCOUNT_ID>:secret:<ADMIN_RANGER_SERVER_TLS_SECRET_NAME>
```

```
        ]
    }
```

> **Note**
>
> Para obter as permissões do Secrets Manager, não esqueça o caractere curinga (“*”) no final do nome do segredo, senão as solicitações falharão. O curinga serve para versões de segredo.

> **Note**
>
> Limite o escopo da AWS Secrets Manager política somente aos certificados necessários para o provisionamento.

Perfil do IAM para Apache Ranger

Esse perfil fornece credenciais para mecanismos de execução confiáveis, como Apache Hive e Amazon EMR Record Server, para acessar os dados do Amazon S3. Use somente esse perfil para acessar dados do Amazon S3, incluindo chaves do KMS, se você estiver usando o SSE-KMS do S3.

Esse perfil deve ser criado com a política mínima indicada no exemplo a seguir.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "CloudwatchLogsPermissions",
      "Action": [
        "logs:CreateLogGroup",
        "logs:CreateLogStream",
        "logs:PutLogEvents"
      ],
      "Effect": "Allow",
      "Resource": [

 "arn:aws:logs:<REGION>:<AWS_ACCOUNT_ID>:<CLOUDWATCH_LOG_GROUP_NAME_IN_SECURITY_CONFIGURATION>:
      ]
    },
```

```
    {
      "Sid": "BucketPermissionsInS3Buckets",
      "Action": [
        "s3:CreateBucket",
        "s3:DeleteBucket",
        "s3:ListAllMyBuckets",
        "s3:ListBucket"
      ],
      "Effect": "Allow",
      "Resource": [
 *"arn:aws:s3:::bucket1",
        "arn:aws:s3:::bucket2"*
        ]
    },
    {
      "Sid": "ObjectPermissionsInS3Objects",
      "Action": [
        "s3:GetObject",
        "s3:DeleteObject",
        "s3:PutObject"
      ],
      "Effect": "Allow",
      "Resource": [
 *"arn:aws:s3:::bucket1/*",
          "arn:aws:s3:::bucket2/*"
*        ]
    }
  ]
}
```

Important

O asterisco “*” no final do recurso de CloudWatch log deve ser incluído para fornecer permissão para gravar nos fluxos de log.

Note

Se você estiver usando a visualização de consistência do EMRFS ou a criptografia S3-SSE, adicione permissões às tabelas do DynamoDB e às chaves do KMS para que os mecanismos de execução possam interagir com esses mecanismos.

O perfil do IAM para Apache Ranger é assumido pelo perfil do perfil de instância do EC2. Use o exemplo a seguir para criar uma política de confiança que permita que o perfil do IAM para o Apache Ranger seja assumido pelo perfil do perfil de instância do EC2.

```
   {
     "Sid": "",
     "Effect": "Allow",
     "Principal": {
       "AWS": "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<EC2 INSTANCE PROFILE ROLE NAME eg.
EMR_EC2_DefaultRole>"
     },
     "Action": ["sts:AssumeRole", "sts:TagSession"]
   }
```

Perfil do IAM para outros serviços da AWS

Essa função fornece aos usuários que não são mecanismos de execução confiáveis credenciais para interagir com os AWS serviços, se necessário. Não use esse perfil do IAM para permitir o acesso aos dados do Amazon S3, a menos que sejam dados que devam ser acessados por todos os usuários.

Esse perfil será assumido pelo perfil do perfil de instância do EC2. Use o exemplo a seguir para criar uma política de confiança que permita que o perfil do IAM para o Apache Ranger seja assumido pelo perfil do perfil de instância do EC2.

```
   {
     "Sid": "",
     "Effect": "Allow",
     "Principal": {
       "AWS": "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<EC2 INSTANCE PROFILE ROLE NAME eg.
EMR_EC2_DefaultRole>"
     },
     "Action": ["sts:AssumeRole", "sts:TagSession"]
   }
```

Validar permissões

Consulte Solução de problemas do Apache Ranger para ver instruções sobre como validar permissões.

Criar a configuração de segurança do EMR

Criar uma configuração de segurança do Amazon EMR para Apache Ranger

Antes de iniciar um cluster do Amazon EMR integrado ao Apache Ranger, crie uma configuração de segurança.

Console

Criar uma configuração de segurança que especifique a opção Integração do AWS Ranger

1. No console do Amazon EMR, selecione Configurações de segurança e depois Criar.
2. Digite um nome em Name (Nome) para a configuração de segurança. Esse nome é usado para especificar a configuração de segurança ao criar um cluster.
3. Em Integração do AWS Ranger, selecione Habilitar controle de acesso granular gerenciado pelo Apache Ranger.
4. Selecione o perfil do IAM para Apache Ranger a ser aplicado. Para ter mais informações, consulte Perfis do IAM para integração nativa com o Apache Ranger.
5. Selecione o Perfil do IAM para outros serviços da AWS a ser aplicado.
6. Configure os plug-ins para se conectar ao servidor Ranger Admin inserindo o ARN do Secret Manager para o servidor Admin e o endereço.
7. Selecione as aplicações para configurar os plug-ins do Ranger. Preencha o ARN do Secret Manager que contém o certificado TLS privado do plug-in.

   Se você não configurar o Apache Spark ou o Apache Hive e eles forem selecionados como uma aplicação para seu cluster, a solicitação falhará.
8. Configure outras opções de configuração de segurança conforme apropriado e escolha Create (Criar). Você deve habilitar a autenticação Kerberos usando o KDC externo ou dedicado ao cluster.

> **Note**
>
> No momento, você não pode usar o console para criar uma configuração de segurança que especifique a opção de integração do AWS Ranger no. AWS GovCloud (US) Region A configuração de segurança do pode ser feita usando a CLI.

CLI

Criar uma configuração de segurança para integração do Apache Ranger

1. *<ACCOUNT ID>*Substitua pelo ID AWS da sua conta.
2. Substitua *<REGION>* pela região em que o recurso está.
3. Especifique um valor para `TicketLifetimeInHours` para determinar o período durante o qual um ticket Kerberos emitido pelo KDC é válido.
4. Especifique o endereço do servidor Ranger Admin para `AdminServerURL`.

```
{
    "AuthenticationConfiguration": {
        "KerberosConfiguration": {
            "Provider": "ClusterDedicatedKdc",
            "ClusterDedicatedKdcConfiguration": {
                "TicketLifetimeInHours": 24
            }
        }
    },
    "AuthorizationConfiguration":{
      "RangerConfiguration":{
         "AdminServerURL":"https://_<RANGER ADMIN SERVER IP>_:6182",
         "RoleForRangerPluginsARN":"arn:aws:iam::_<ACCOUNT ID>_:role/_<RANGER PLUGIN
 DATA ACCESS ROLE NAME>_",
         "RoleForOtherAWSServicesARN":"arn:aws:iam::_<ACCOUNT ID>_:role/_<USER
 ACCESS ROLE NAME>_",
         "AdminServerSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES ADMIN SERVERS PUBLIC TLS CERTIFICATE
 WITHOUT VERSION>_",
         "RangerPluginConfigurations":[
            {
               "App":"Spark",
               "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES SPARK PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
               "PolicyRepositoryName":"<SPARK SERVICE NAME eg. amazon-emr-spark>"
            },
            {
               "App":"Hive",
```

```
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES Hive PLUGIN PRIVATE TLS CERTIFICATE WITHOUT
 VERSION>_",
                "PolicyRepositoryName":"<HIVE SERVICE NAME eg. Hivedev>"
            },
            {
                "App":"EMRFS-S3",
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES EMRFS S3 PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
                "PolicyRepositoryName":"<EMRFS S3 SERVICE NAME eg amazon-emr-emrfs>"
            },
       {
                "App":"Trino",
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES TRINO PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
                "PolicyRepositoryName":"<TRINO SERVICE NAME eg amazon-emr-trino>"
            }
         ],
         "AuditConfiguration":{
            "Destinations":{
               "AmazonCloudWatchLogs":{
                  "CloudWatchLogGroup":"arn:aws:logs:<REGION>:_<ACCOUNT ID>_:log-
group:_<LOG GROUP NAME FOR AUDIT EVENTS>_"
               }
            }
         }
      }
   }
}
```

Esses PolicyRespositoryNames são os nomes dos serviços especificados em seu Apache Ranger Admin.

Crie uma configuração de segurança do Amazon EMR com o comando a seguir. Substitua a configuração de segurança por um nome a sua escolha. Selecione a configuração pelo nome ao criar o cluster.

```
aws emr create-security-configuration \
--security-configuration file://./security-configuration.json \
--name security-configuration
```

Configurar atributos de segurança adicionais

Para integrar o Amazon EMR ao Apache Ranger com segurança, configure os seguintes recursos de segurança do EMR:

- Habilite a autenticação Kerberos usando o KDC externo ou dedicado ao cluster. Para obter instruções, consulte Usar o Kerberos para autenticação com o Amazon EMR.
- (Opcional) Habilite a criptografia em trânsito ou em repouso. Para ter mais informações, consulte Opções de criptografia.

Para ter mais informações, consulte Segurança no Amazon EMR.

Armazenar certificados TLS no AWS Secrets Manager

Os plug-ins Ranger instalados em um cluster do Amazon EMR e o servidor Ranger Admin devem se comunicar por TLS para garantir que os dados da política e outras informações enviadas não possam ser lidos caso sejam interceptados. O EMR também exige que os plug-ins se autentiquem no servidor Ranger Admin fornecendo o próprio certificado TLS e realizando a autenticação TLS bidirecional. Essa configuração exigiu a criação de quatro certificados: dois pares de certificados TLS públicos e de privados. Para obter instruções sobre como instalar o certificado no servidor Ranger Admin, consulte Configurar o servidor do Ranger Admin. Para concluir a configuração, os plug-ins Ranger instalados no cluster do EMR precisam de dois certificados: o certificado TLS público do servidor de administrador e o certificado privado que o plug-in usará para se autenticar no servidor Ranger Admin. Para fornecer esses certificados TLS, eles devem estar no AWS Secrets Manager e fornecidos em uma configuração de segurança do EMR.

> **Note**
>
> É altamente recomendável, mas não obrigatório, criar um par de certificados para cada uma das aplicações para limitar o impacto se um dos certificados do plug-in for comprometido.

> **Note**
>
> É necessário rastrear e alternar os certificados antes da data de vencimento.

Formato do certificado

A importação dos certificados para o AWS Secrets Manager é a mesma, independentemente de ser o certificado de plug-in privado ou o certificado de administrador público do Ranger. Antes de importar os certificados TLS, os certificados deverão estar no formato 509x PEM.

Este é o formato de um exemplo de certificado público:

```
-----BEGIN CERTIFICATE-----
...Certificate Body...
-----END CERTIFICATE-----
```

Este é o formato de um exemplo de certificado privado:

```
-----BEGIN PRIVATE KEY-----
...Private Certificate Body...
-----END PRIVATE KEY-----
-----BEGIN CERTIFICATE-----
...Trust Certificate Body...
-----END CERTIFICATE-----
```

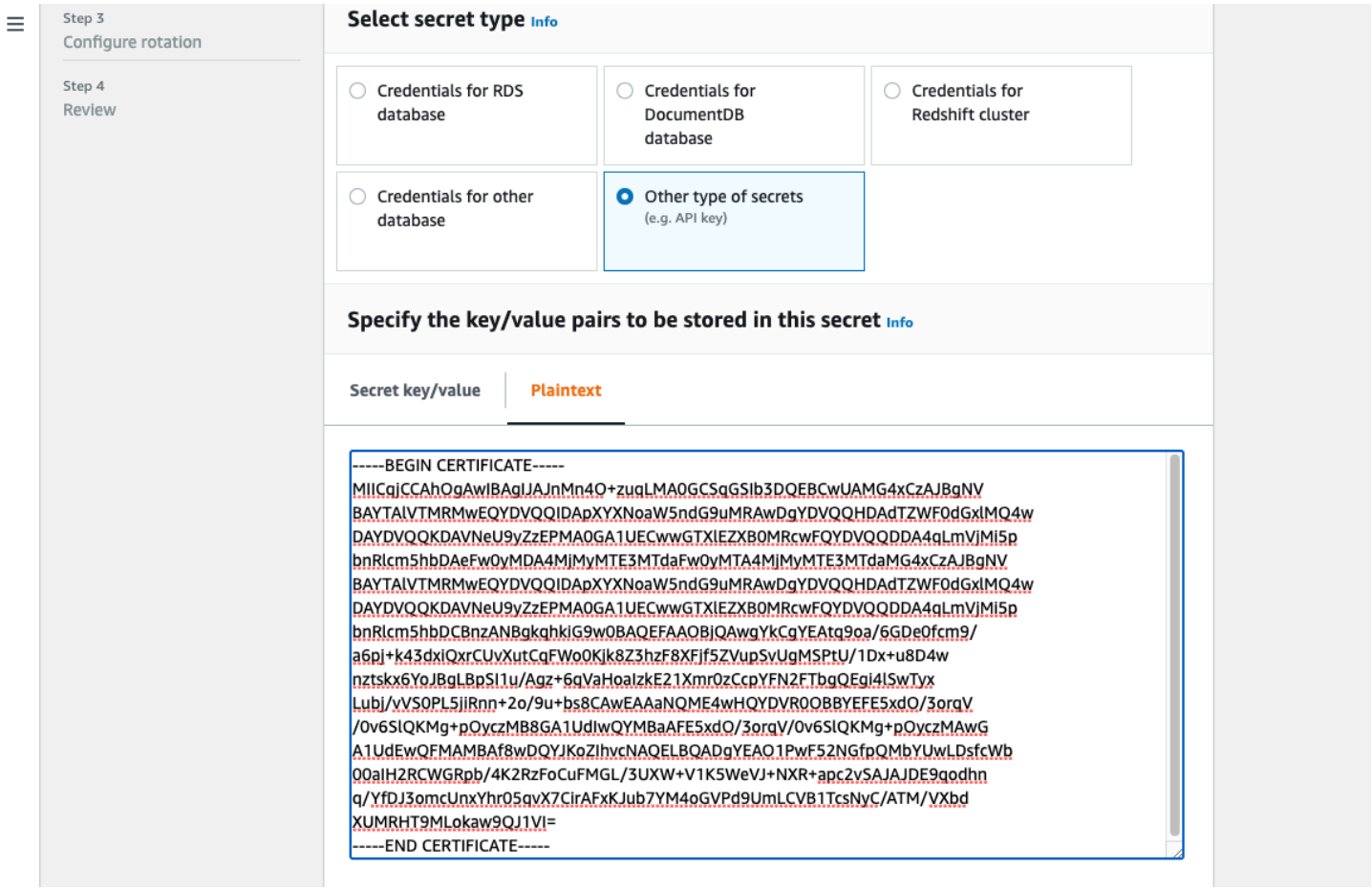
O certificado privado também deverá conter um certificado de confiança.

É possível validar se os certificados estão no formato correto executando o seguinte comando:

```
openssl x509 -in <PEM FILE> -text
```

Importar um certificado para o AWS Secrets Manager

Ao criar seu segredo no Secrets Manager, escolha Outro tipo de segredos em Tipo de segrdo e cole o certificado codificado PEM no campo Texto sem formatação.



Iniciar um cluster do EMR

Antes de iniciar um cluster do Amazon EMR com o Apache Ranger, certifique-se de que todos os componentes atendam aos seguintes requisitos mínimos de versão:

- Amazon EMR 5.32.0 ou versões posteriores ou 6.3.0 ou versões posteriores. É recomendável usar a versão mais recente do Amazon EMR.
- Servidor Apache Ranger Admin 2.x.

Execute as etapas a seguir.

- Instale o Apache Ranger, caso ainda não tenha instalado. Para obter mais informações sobre a instalação, consulte Apache Ranger 0.5.0 installation.
- Verifique se há conectividade de rede entre o cluster do Amazon EMR e o servidor Apache Ranger Admin. Consulte Configurar o servidor do Ranger Admin

- Crie os perfis do IAM necessários. Consulte Perfis do IAM para integração nativa com o Apache Ranger.
- Crie uma configuração de segurança do EMR para a instalação do Apache Ranger. Veja mais informações em Criar a configuração de segurança do EMR.

Configurar o Zeppelin para clusters do Amazon EMR habilitados para Apache Ranger

O tópico aborda como configurar o Apache Zeppelin para um cluster do Amazon EMR habilitado para Apache Ranger para que você possa usar o Zeppelin como um caderno para explorar dados de maneira interativa. O Zeppelin é incluído no Amazon EMR 5.0.0 e versões posteriores. As versões anteriores incluem o Zeppelin como uma aplicação sandbox. Para obter mais informações, consulte Amazon EMR 4.x release versions no Guia de lançamento do Amazon EMR.

Por padrão, o Zeppelin é configurado com um login e uma senha padrão que não são seguros em um ambiente multilocatário.

Para configurar o Zeppelin, siga as etapas a seguir.

1. Modificar o mecanismo de autenticação.

   Modifique o arquivo `shiro.ini` para implementar o mecanismo de autenticação de sua preferência. O Zeppelin oferece suporte a Active Directory, LDAP, PAM e Knox SSO. Consulte Apache Shiro authentication for Apache Zeppelin para obter mais informações.
2. Configurar o Zeppelin para representar o usuário final

   Quando você permite que o Zeppelin represente o usuário final, os trabalhos enviados pelo Zeppelin podem ser executados como esse usuário final. Adicione o seguinte à configuração de `core-site.xml`:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.zeppelin.hosts": "*",
      "hadoop.proxyuser.zeppelin.groups": "*"
    },
    "Configurations": [
    ]
  }
```

```
]
```

Em seguida, adicione a seguinte configuração a `hadoop-kms-site.xml` localizado em `/etc/hadoop/conf`:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.proxyuser.zeppelin.hosts": "*",
      "hadoop.kms.proxyuser.zeppelin.groups": "*"
    },
    "Configurations": [
    ]
  }
]
```

Também é possível adicionar essas configurações ao cluster do Amazon EMR usando o console seguindo as etapas descritas em Reconfigure an instance group in the console.

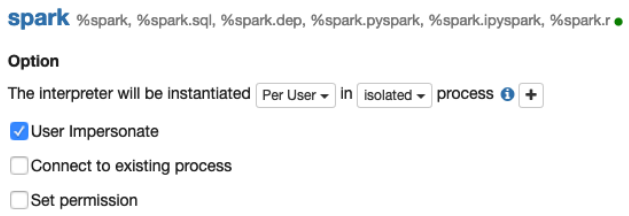
3. Permitir que o Zeppelin se torne o usuário final

   Crie um arquivo `/etc/sudoers.d/90-zeppelin-user` que contenha:

   ```
   zeppelin ALL=(ALL) NOPASSWD:ALL
   ```

4. Modificar as configurações dos intérpretes para executar trabalhos do usuário em seus próprios processos.

   Configure todos os intérpretes para instanciar os intérpretes “Por usuário” em processos “isolados”.

5. Modificar **`zeppelin-env.sh`**

Adicione isto a `zeppelin-env.sh` que o Zeppelin comece a iniciar intérpretes como usuário final:

```
ZEPPELIN_IMPERSONATE_USER=`echo ${ZEPPELIN_IMPERSONATE_USER} | cut -d @ -f1`
export ZEPPELIN_IMPERSONATE_CMD='sudo -H -u ${ZEPPELIN_IMPERSONATE_USER} bash -c'
```

Adicione isto a `zeppelin-env.sh` para alterar as permissões padrão de caderno para somente leitura para o criador:

```
export ZEPPELIN_NOTEBOOK_PUBLIC="false"
```

Por fim, adicione o seguinte `zeppelin-env.sh` para incluir o caminho da RecordServer classe EMR após a primeira `CLASSPATH` declaração:

```
export CLASSPATH="$CLASSPATH:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-connector-common.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-spark-connector.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-client.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-server-
common.jar:/usr/share/aws/emr/record-server/lib/jars/secret-agent-interface.jar"
```

6. Reiniciar o Zeppelin.

   Execute o seguinte comando para reiniciar o Zeppelin:

   ```
   sudo systemctl restart zeppelin
   ```

Problemas conhecidos

Problemas conhecidos

Há um problema conhecido na versão 5.32 do Amazon EMR em que as permissões `hive-site.xml` foram alteradas para que somente usuários privilegiados possam lê-las, pois pode haver credenciais armazenadas nelas. Isso pode impedir que o Hue leia `hive-site.xml` e fazer com que as páginas da Web sejam recarregadas continuamente. Se você tiver esse problema, adicione esta configuração para corrigir o problema:

```
[
  {
```

```
    "Classification": "hue-ini",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "desktop",
        "Properties": {
          "server_group":"hive_site_reader"
         },
        "Configurations":[
        ]
      }
    ]
  }
]
```

Há um problema conhecido de que o plug-in EMRFS S3 para Apache Ranger atualmente não oferece suporte ao atributo Security Zone do Apache Ranger. As restrições de controle de acesso definidas usando o atributo Security Zone não são aplicadas aos clusters do Amazon EMR.

IUs de aplicações

Por padrão, as IUs de aplicações não realizam autenticação. Isso inclui a ResourceManager interface do usuário, a NodeManager interface do usuário, a interface do usuário Livy, entre outras. Além disso, qualquer usuário capaz de acessar as interfaces de usuário poderá visualizar informações sobre os trabalhos de todos os outros usuários.

Se esse comportamento não for desejado, você deve garantir que um grupo de segurança seja usado para restringir o acesso dos usuários às IUs de aplicações.

Permissões padrão do HDFS

Por padrão, os objetos que os usuários criam no HDFS recebem permissões de leitura mundial. Isso poderá tornar os dados legíveis por usuários que não deveriam ter acesso a eles. Para alterar esse comportamento de modo que as permissões de arquivo padrão sejam definidas para leitura e gravação somente pelo criador do trabalho, execute as etapas a seguir.

Ao criar o cluster do EMR, forneça a seguinte configuração:

```
[
  {
    "Classification": "hdfs-site",
```

```
    "Properties": {
      "dfs.namenode.acls.enabled": "true",
      "fs.permissions.umask-mode": "077",
      "dfs.permissions.superusergroup": "hdfsadmingroup"
    }
  }
]
```

Além disso, execute esta ação de bootstrap:

```
--bootstrap-actions Name='HDFS UMask Setup',Path=s3://elasticmapreduce/hdfs/umask/
umask-main.sh
```

Plug-ins Apache Ranger

Os plug-ins Apache Ranger validam o acesso de um usuário em relação às políticas de autorização definidas no servidor de administração de políticas do Apache Ranger.

Tópicos

- Plug-in Apache Hive
- Plug-in Apache Spark
- Plug-in EMRFS S3
- Plug-in Trino

Plug-in Apache Hive

O Apache Hive é um mecanismo de execução bastante usado dentro do ecossistema Hadoop. O Amazon EMR fornece um plug-in Apache Ranger para poder proporcionar controles de acesso refinados para o Hive. O plug-in é compatível com o servidor Apache Ranger Admin de código aberto versão 2.0 e posteriores.

Tópicos

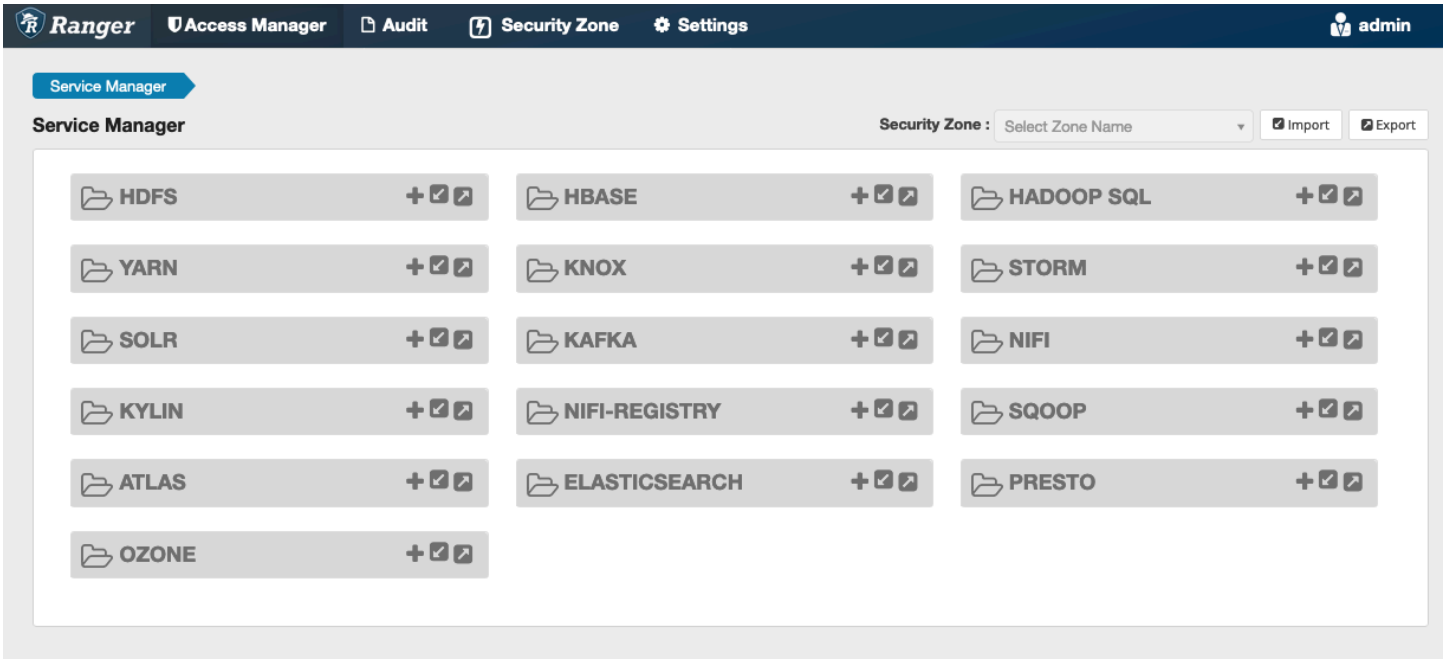
- Atributos compatíveis
- Instalação da configuração de serviço
- Considerações
- Limitações

Atributos compatíveis

O plug-in Apache Ranger para Hive no EMR oferece suporte a todas as funcionalidades do plug-in de código aberto, que inclui controles de acesso em nível de banco de dados, tabela e coluna, filtragem de linhas e mascaramento de dados. Para ver uma tabela dos comandos do Hive e das permissões associadas do Ranger, consulte Hive commands to Ranger permission mapping.

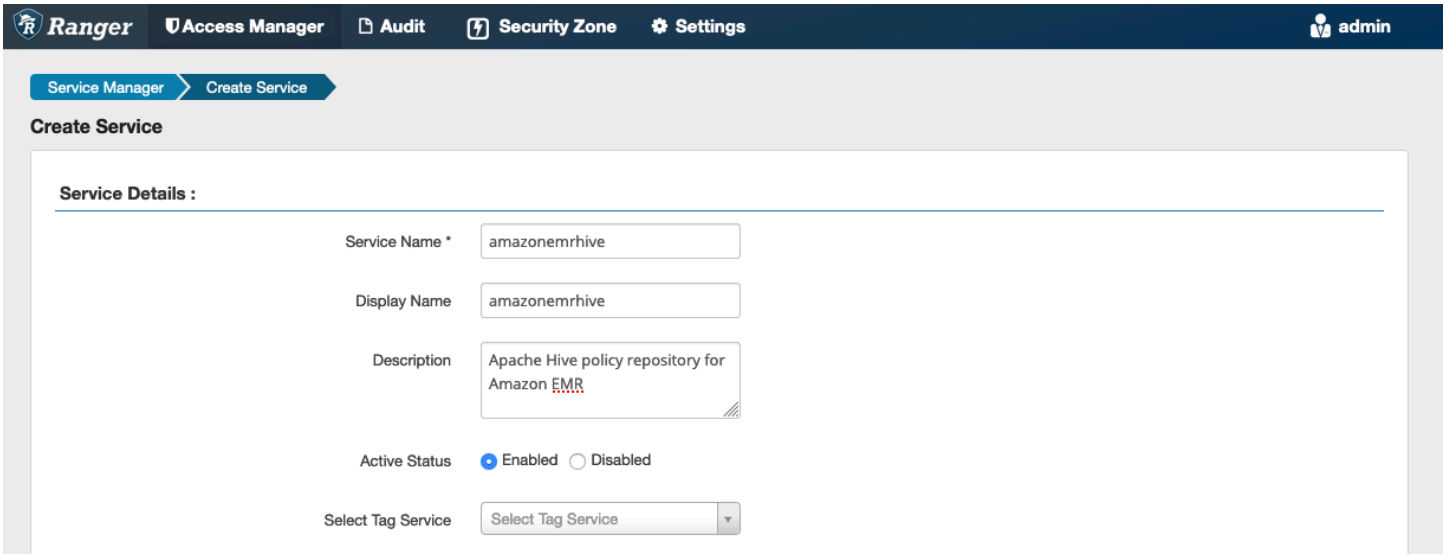
Instalação da configuração de serviço

O plug-in Apache Hive é compatível com a definição de serviço Hive já existente no Apache Hive Hadoop SQL.



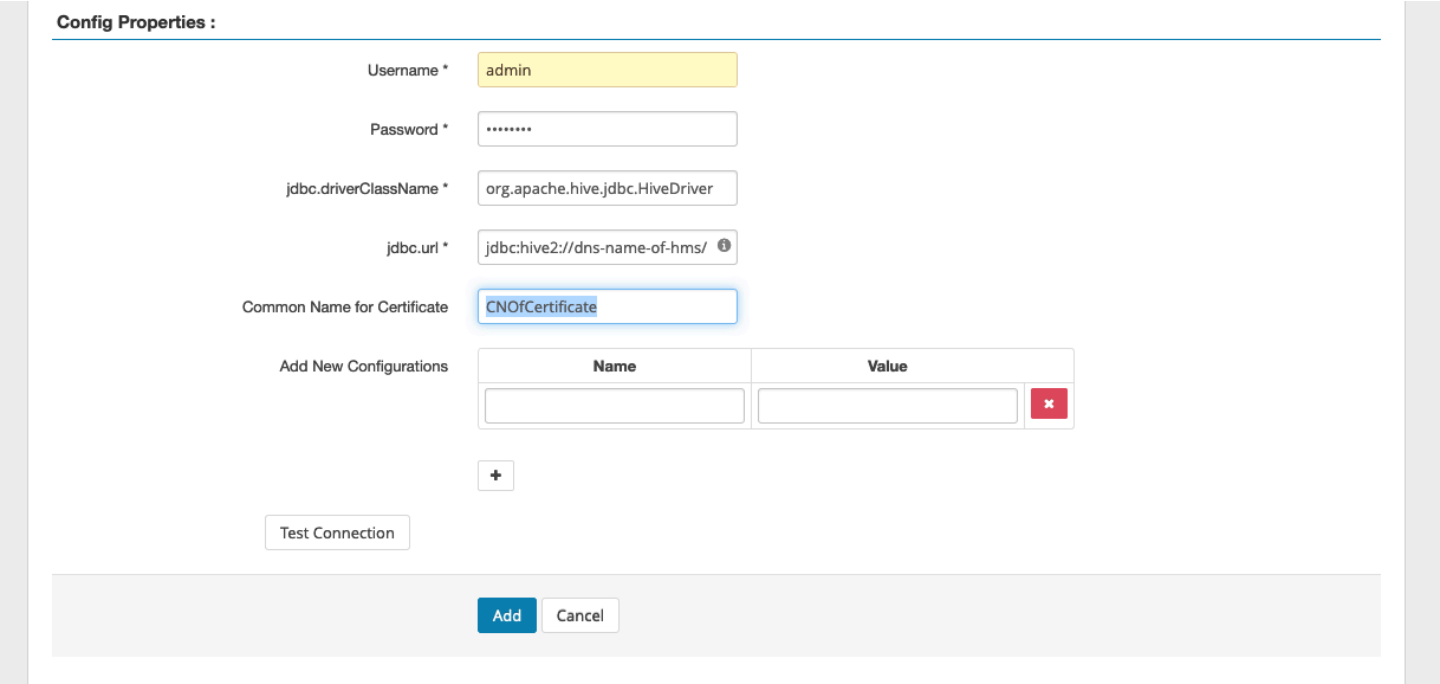
Caso não tenha uma instância de serviço no Hadoop SQL, como mostrado acima, você pode criar uma. Clique em + ao lado do Hadoop SQL.

1. Nome do serviço (se for exibido): insira o nome do serviço. O valor sugerido é **`amazonemrhive`**. Anote esse nome de serviço, pois ele será necessário ao criar uma configuração de segurança do EMR.
2. Nome de exibição: insira o nome a ser exibido para o serviço. O valor sugerido é **`amazonemrhive`**.

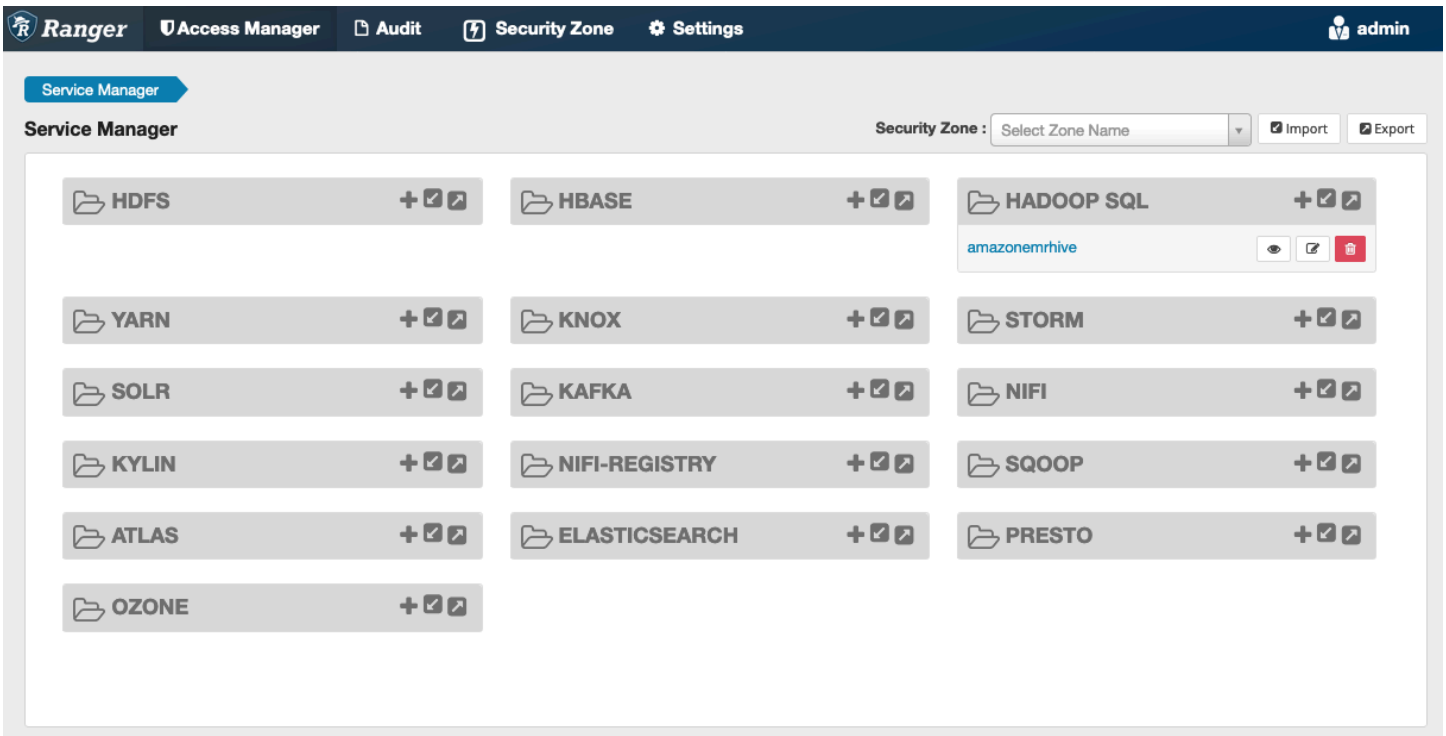


As propriedades de configuração do Apache Hive são usadas para estabelecer uma conexão com seu servidor Apache Ranger Admin com um 2 HiveServer para implementar o preenchimento automático ao criar políticas. As propriedades abaixo não precisam ser precisas se você não tiver um processo persistente HiveServer 2 e puderem ser preenchidas com qualquer informação.

- Nome de usuário: insira um nome de usuário para a conexão JDBC com uma instância de HiveServer 2 instâncias.
- Senha: insira a senha do nome de usuário acima.
- jdbc. driver. ClassName: insira o nome da classe JDBC para conectividade com o Apache Hive. O valor padrão pode ser usado.
- jdbc.url: insira a string de conexão JDBC a ser usada ao se conectar a 2. HiveServer
- Nome comum para certificado: o campo CN dentro do certificado usado para se conectar ao servidor de administração com base em um plug-in cliente. Esse valor deve corresponder ao campo CN do certificado TLS que foi criado para o plug-in.



O botão Testar conexão testa se os valores acima podem ser usados para se conectar com êxito à instância HiveServer 2. Depois que o serviço for criado com êxito, o Service Manager deverá ficará semelhante a isto:



Considerações

Servidor de metadados Hive

O servidor de metadados Hive só pode ser acessado por mecanismos confiáveis, especificamente o Hive e `emr_record_server`, para proteção contra acesso não autorizado. O servidor de metadados Hive também é acessado por todos os nós do cluster. A porta 9083 necessária fornece acesso de todos os nós ao nó principal.

Autenticação

Por padrão, o Apache Hive está configurado para se autenticar usando Kerberos conforme configurado na configuração do EMR Security. HiveServer2 também pode ser configurado para autenticar usuários usando o LDAP. Consulte Implementing LDAP authentication for Hive on a multi-tenant Amazon EMR cluster para obter informações.

Limitações

Estas são as limitações atuais do plug-in Apache Hive no Amazon EMR 5.x:

- Não há suporte para perfis do Hive atualmente. Não há suporte para instruções Grant e Revoke.
- Não há suporte para a CLI do Hive. O JDBC/Beeline é a única forma autorizada de conectar o Hive.
- A configuração `hive.server2.builtin.udf.blacklist` deve ser preenchida com UDFs que você considere inseguras.

Plug-in Apache Spark

O Amazon EMR integrou o EMR RecordServer para fornecer controle de acesso refinado para o SparkSQL. O EMR RecordServer é um processo privilegiado executado em todos os nós em um cluster habilitado para Apache Ranger. Quando um driver ou executor do Spark executa uma instrução SparkSQL, todos os metadados e solicitações de dados passam pelo. RecordServer Para saber mais sobre o EMR RecordServer, consulte a Componentes do Amazon EMR página.

Tópicos

- Atributos compatíveis
- Reimplantar a definição de serviço para usar instruções INSERT, ALTER ou DDL
- Instalação da definição de serviço

- Criar políticas SparkSQL
- Considerações
- Limitações

Atributos compatíveis

Instrução SQL/ação do Ranger	STATUS	Versão do EMR compatível
SELECT	Compatível	A partir da 5.32
SHOW DATABASES	Compatível	A partir da 5.32
SHOW COLUMNS	Compatível	A partir da 5.32
SHOW TABLES	Compatível	A partir da 5.32
SHOW TABLE PROPERTIES	Compatível	A partir da 5.32
DESCRIBE TABLE	Compatível	A partir da 5.32
INSERT OVERWRITE	Compatível	A partir da 5.34 e 6.4
INSERT INTO	Compatível	A partir da 5.34 e 6.4
ALTER TABLE	Compatível	A partir da 6.4
CRIAR TABELA	Compatível	A partir da 5.35 e 6.7
CREATE DATABASE	Compatível	A partir da 5.35 e 6.7
DESCARTAR TABELA	Compatível	A partir da 5.35 e 6.7

Instrução SQL/ação do Ranger	STATUS	Versão do EMR compatível
DROP DATABASE	Compatível	A partir da 5.35 e 6.7
DROP VIEW	Compatível	A partir da 5.35 e 6.7
CREATE VIEW	Sem suporte	

Os seguintes atributos são compatíveis com o uso do SparkSQL:

- Controle de acesso refinado em tabelas dentro do Hive Metastore, e é possível criar políticas em nível de banco de dados, tabela e coluna.
- As políticas do Apache Ranger podem incluir políticas de concessão e políticas de negação para usuários e grupos.
- Os eventos de auditoria são enviados para o CloudWatch Logs.

Reimplantar a definição de serviço para usar instruções INSERT, ALTER ou DDL

> **Note**
>
> A partir do Amazon EMR 6.4, é possível usar o Spark SQL com as instruções: INSERT INTO, INSERT OVERWRITE ou ALTER TABLE. A partir do Amazon EMR 6.7, é possível usar o Spark SQL para criar ou eliminar bancos de dados e tabelas. Se você já tiver uma instalação no servidor Apache Ranger com definições de serviço Apache Spark implantadas, use o código a seguir para reimplantar as definições de serviço.

```
# Get existing Spark service definition id calling Ranger REST API and JSON
 processor
curl --silent -f -u <admin_user_login>:<password_for_ranger_admin_user> \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef/
name/amazon-emr-spark' | jq .id

# Download the latest Service definition
```

```
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/
version-2.0/ranger-servicedef-amazon-emr-spark.json

# Update the service definition using the Ranger REST API
curl -u <admin_user_login>:<password_for_ranger_admin_user> -X PUT -d @ranger-
servicedef-amazon-emr-spark.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/
servicedef/<Spark service definition id from step 1>'
```

Instalação da definição de serviço

A instalação da definição de serviço Apache Spark do EMR exige que o servidor Ranger Admin esteja configurado. Consulte Configurar o servidor do Ranger Admin.

Siga estas etapas para instalar a definição de serviço Apache Spark:

Etapa 1: SSH no servidor Apache Ranger Admin

Por exemplo: .

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

Etapa 2: baixar a definição de serviço e o plug-in do servidor Apache Ranger Admin

Em um diretório temporário, baixe a definição de serviço. Essa definição de serviço é compatível com as versões Ranger 2.x.

```
mkdir /tmp/emr-spark-plugin/
cd /tmp/emr-spark-plugin/

wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-spark-plugin-2.x.jar
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-servicedef-amazon-emr-spark.json
```

Etapa 3: instalar o plug-in Apache Spark para Amazon EMR

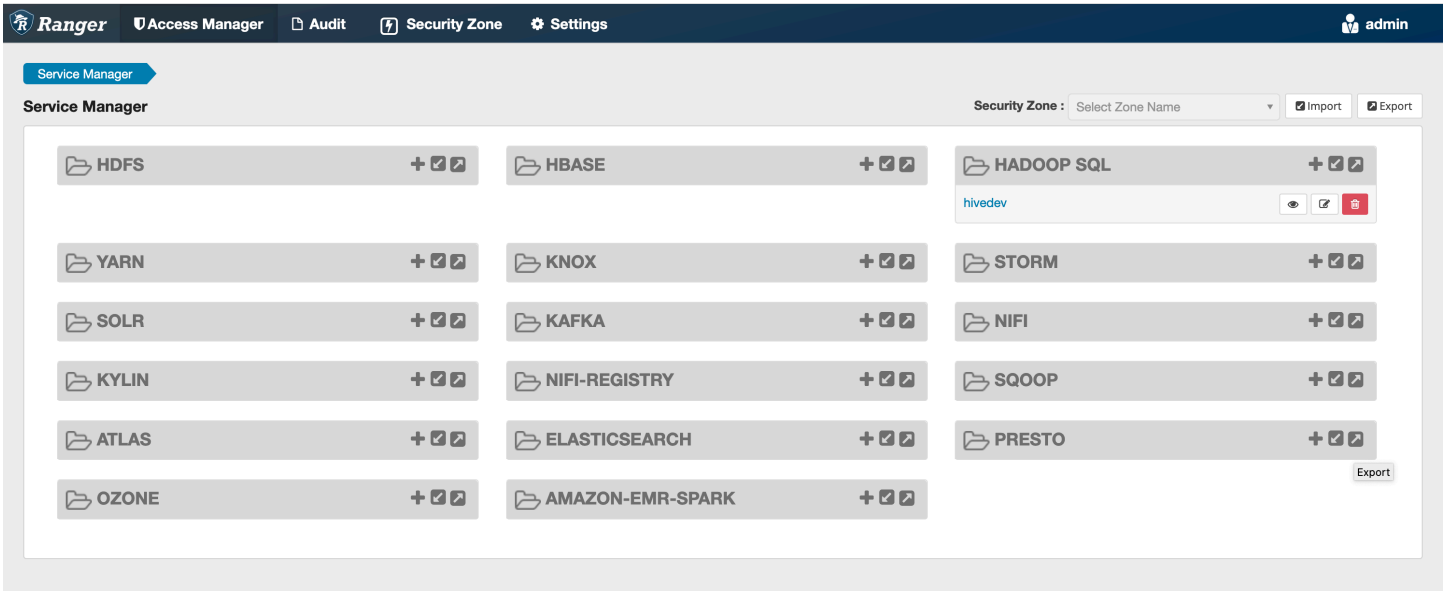
```
export RANGER_HOME=.. # Replace this Ranger Admin's home directory eg /usr/lib/ranger/
ranger-2.0.0-admin
```

```
mkdir $RANGER_HOME/ews/webapp/WEB-INF/classes/ranger-plugins/amazon-emr-spark
mv ranger-spark-plugin-2.x.jar $RANGER_HOME/ews/webapp/WEB-INF/classes/ranger-plugins/
amazon-emr-spark
```

Etapa 4: registrar a definição de serviço Apache Spark para Amazon EMR

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_* -X
 POST -d @ranger-servicedef-amazon-emr-spark.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

Se esse comando for executado com êxito, você verá um novo serviço na interface de usuário do Ranger Admin chamado “AMAZON-EMR-SPARK”, conforme mostrado na imagem a seguir (a versão 2.0 do Ranger é exibida).

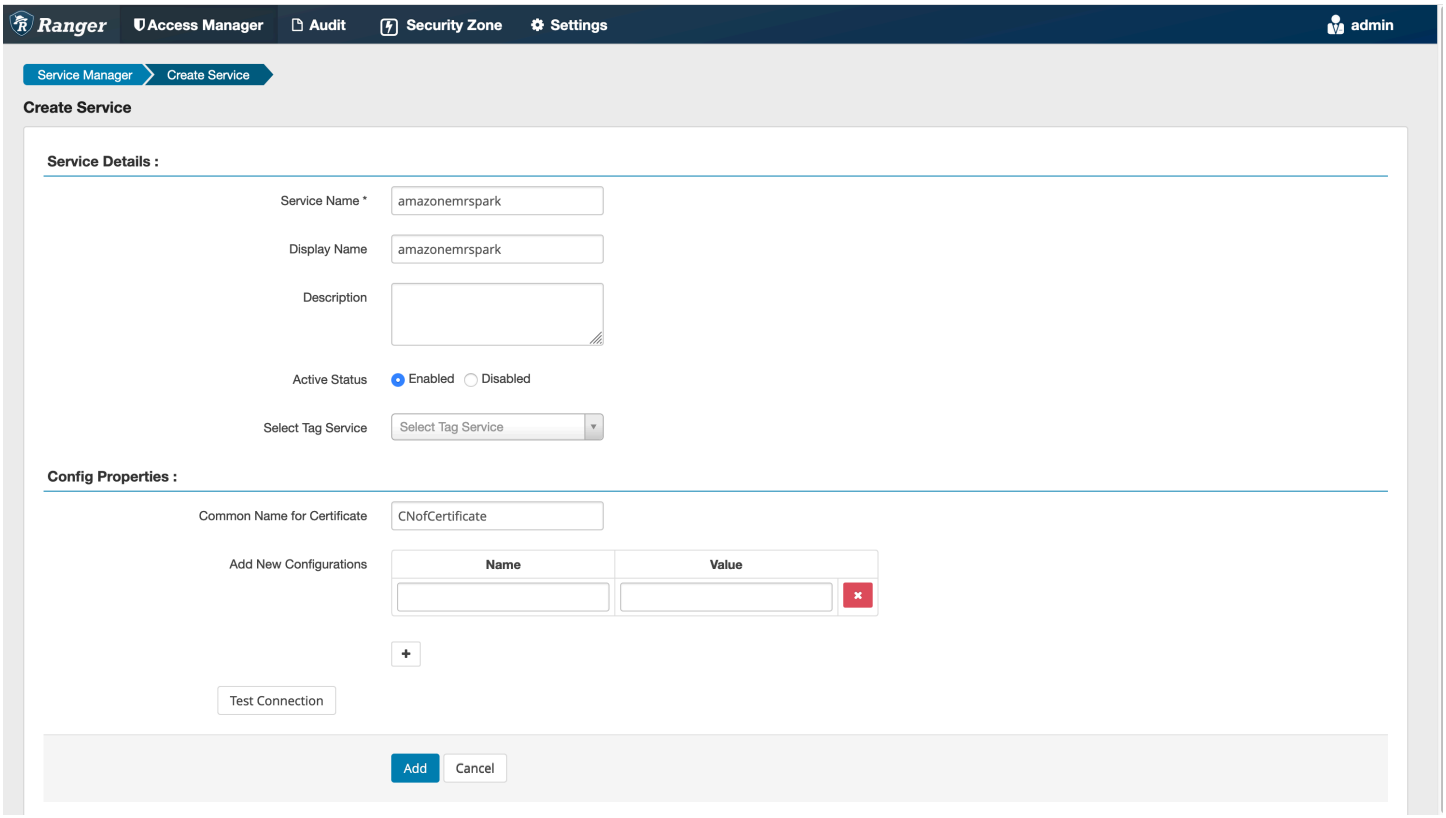


Etapa 5: criar uma instância da aplicação AMAZON-EMR-SPARK

Nome do serviço (se for exibido): o nome do serviço que será usado. O valor sugerido é **`amazonemrspark`**. Anote esse nome de serviço, pois ele será necessário ao criar uma configuração de segurança do EMR.

Nome de exibição: o nome a ser exibido para a instância. O valor sugerido é **`amazonemrspark`**.

Nome comum para certificado: o campo CN dentro do certificado usado para se conectar ao servidor de administração com base em um plug-in cliente. Esse valor deve corresponder ao campo CN do certificado TLS que foi criado para o plug-in.



> **Note**
>
> O certificado TLS para o plug-in deveria ter sido registrado no armazenamento confiável do servidor Ranger Admin. Consulte Certificados TLS para obter mais detalhes.

Criar políticas SparkSQL

Ao criar uma nova política, os campos a serem preenchidos são:

Nome da política: o nome da política.

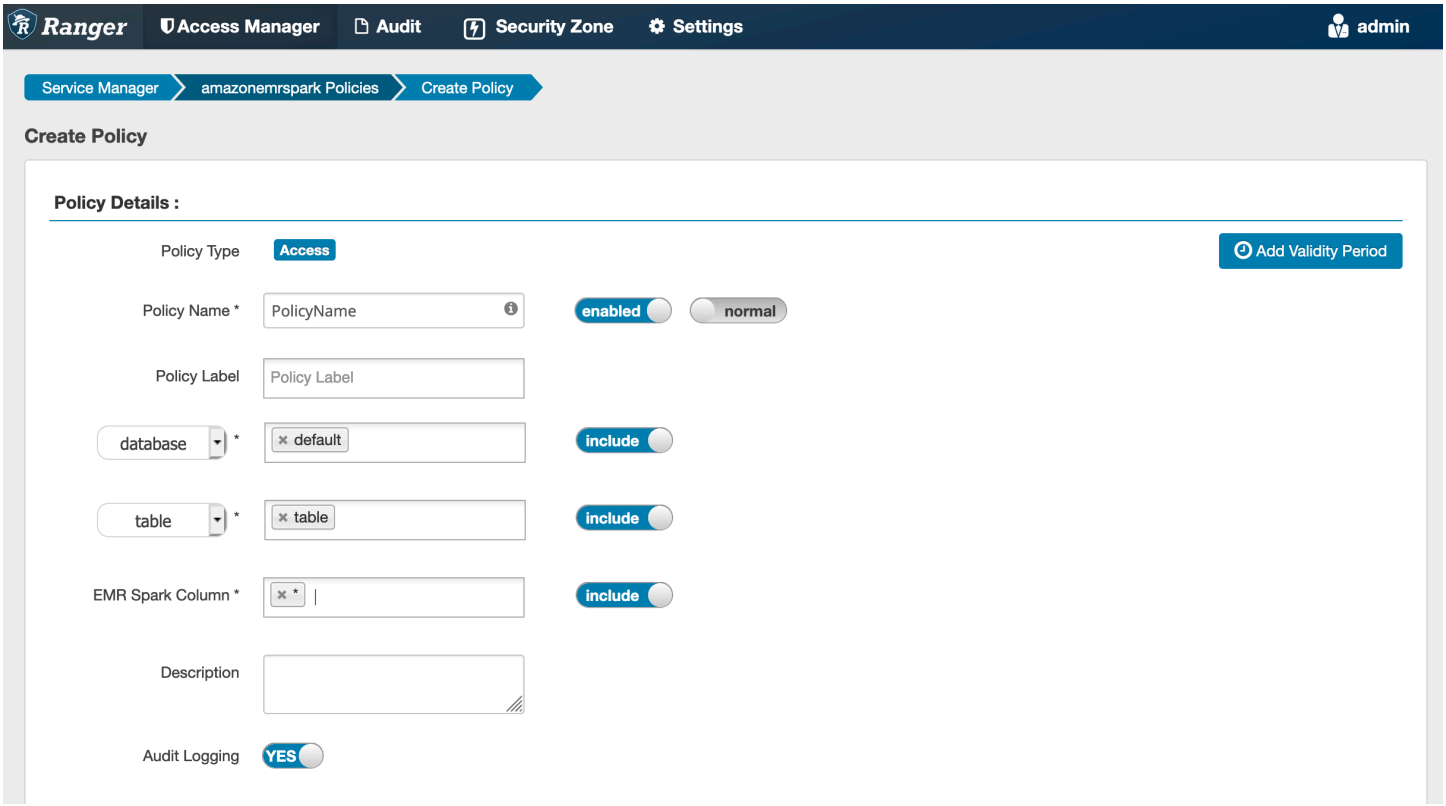
Rótulo de política: um rótulo que você pode colocar na política.

Banco de dados: o banco de dados ao qual a política se aplica. O caractere curinga "*" representa todos os bancos de dados.

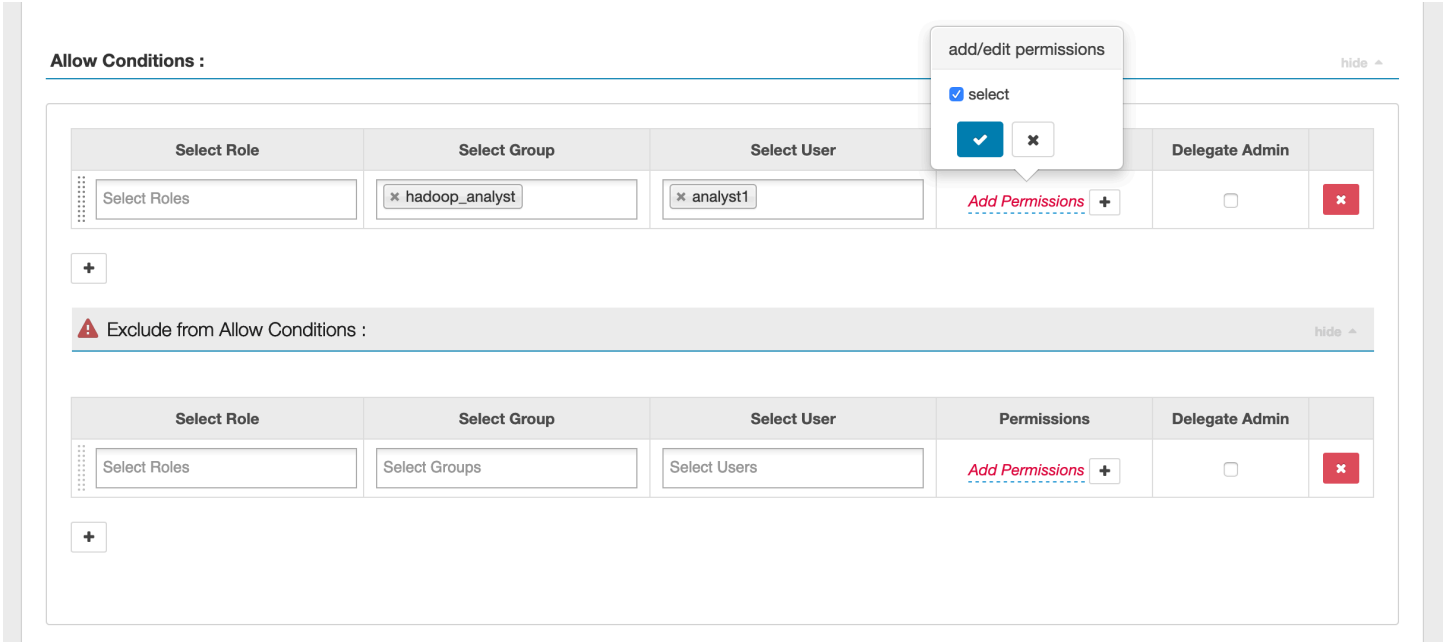
Tabela: as tabelas às quais a política se aplica. O caractere curinga "*" representa todas as tabelas.

Coluna do EMR Spark: as colunas às quais a política se aplica. O caractere curinga "*" representa todas as colunas.

Descrição: uma descrição da política.



Para especificar usuários e grupos, insira os usuários e grupos abaixo para conceder permissões. Também é possível especificar exclusões para as condições de permissão e negação.



Após especificar as condições de permitir e negar, clique em Salvar.

Considerações

Cada nós do cluster do EMR deve ser capaz de se conectar ao nó principal na porta 9083.

Limitações

Estas são as limitações atuais do plug-in Apache Spark:

- O Record Server sempre se conectará ao HMS que está em execução em um cluster do Amazon EMR. Configure o HMS para se conectar ao modo remoto, se necessário. Você não deve colocar valores de configuração no arquivo de configuração Hive-site.xml do Apache Spark.
- As tabelas criadas usando fontes de dados do Spark em CSV ou Avro não podem ser lidas usando o EMR. RecordServer Utilize o Hive para criar e gravar dados e ler usando Record.
- Não há suporte para tabelas Delta Lake e Hudi.
- Os usuários precisam ter acesso ao banco de dados padrão. Esse é um requisito do Apache Spark.
- O servidor Ranger Admin não oferece suporte ao preenchimento automático.
- O plug-in SparkSQL para Amazon EMR não oferece suporte a filtros de linha ou a mascaramento de dados.
- Ao ser usado ALTER TABLE com Spark SQL, o local da partição deve ser o diretório filho do local de uma tabela. Não há suporte para inserção de dados em uma partição em que a localização da partição seja diferente da localização da tabela.

Plug-in EMRFS S3

Para facilitar o fornecimento de controles de acesso contra objetos no S3 em um cluster multilocatário, o plug-in EMRFS S3 fornece controles de acesso aos dados no S3 ao acessá-los pelo EMRFS. Você pode permitir acesso aos recursos do S3 em nível de usuário e grupo.

Para conseguir isso, quando sua aplicação tenta acessar dados no S3, o EMRFS envia uma solicitação de credenciais ao processo do agente secreto, onde a solicitação é autenticada e autorizada em um plug-in Apache Ranger. Se a solicitação for autorizada, o agente secreto assumirá o perfil do IAM para os mecanismos do Apache Ranger com uma política restrita para gerar credenciais que só tenham acesso à política Ranger que permitiu o acesso. As credenciais então são repassadas ao EMRFS para acessar o S3.

Tópicos

Atributos compatíveis

O plug-in EMRFS S3 concede autorização de nível de armazenamento. Políticas podem ser criadas para conceder acesso a usuários e grupos a buckets e prefixos do S3. A autorização é feita somente em relação ao EMRFS.

Instalação da configuração de serviço

Para instalar a definição do serviço EMRFS, você deve configurar o servidor Ranger Admin. Para configurar o servidor, consulteConfigurar o servidor do Ranger Admin.

Siga estas etapas para instalar a definição de serviço do EMRFS.

Etapa 1: SSH no servidor Apache Ranger Admin.

Por exemplo: .

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

Etapa 2: Baixe a definição do serviço EMRFS.

Em um diretório temporário, baixe a definição de serviço do Amazon EMR. Essa definição de serviço é compatível com as versões Ranger 2.x.

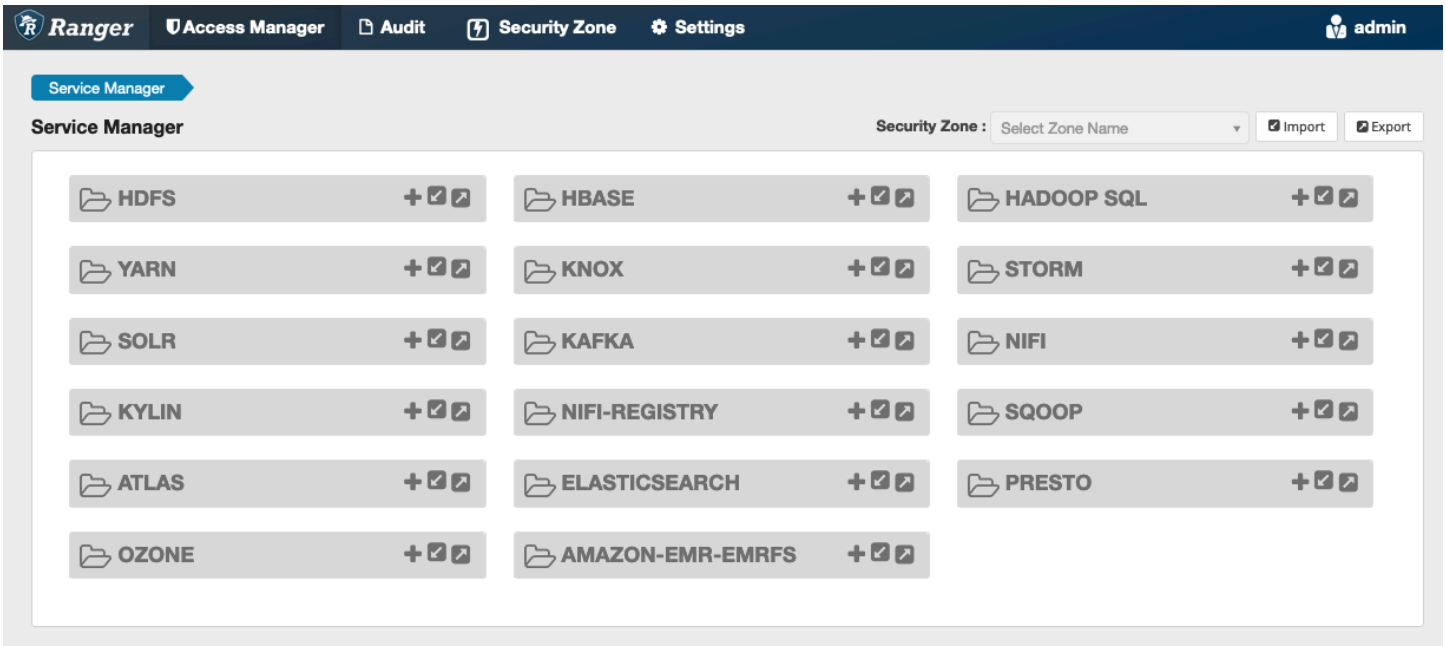
```
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-servicedef-amazon-emr-emrfs.json
```

Etapa 3: Registrar a definição do serviço EMRFS S3.

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_* -X
 POST -d @ranger-servicedef-amazon-emr-emrfs.json \
```

```
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

Se esse comando for executado com êxito, você verá um novo serviço na interface de usuário do Ranger Admin chamado “AMAZON-EMR-S3”, conforme mostrado na imagem a seguir (a versão 2.0 do Ranger é exibida).



Etapa 4: Crie uma instância do aplicativo AMAZON-EMR-EMRFS.

Crie uma instância da definição de serviço.

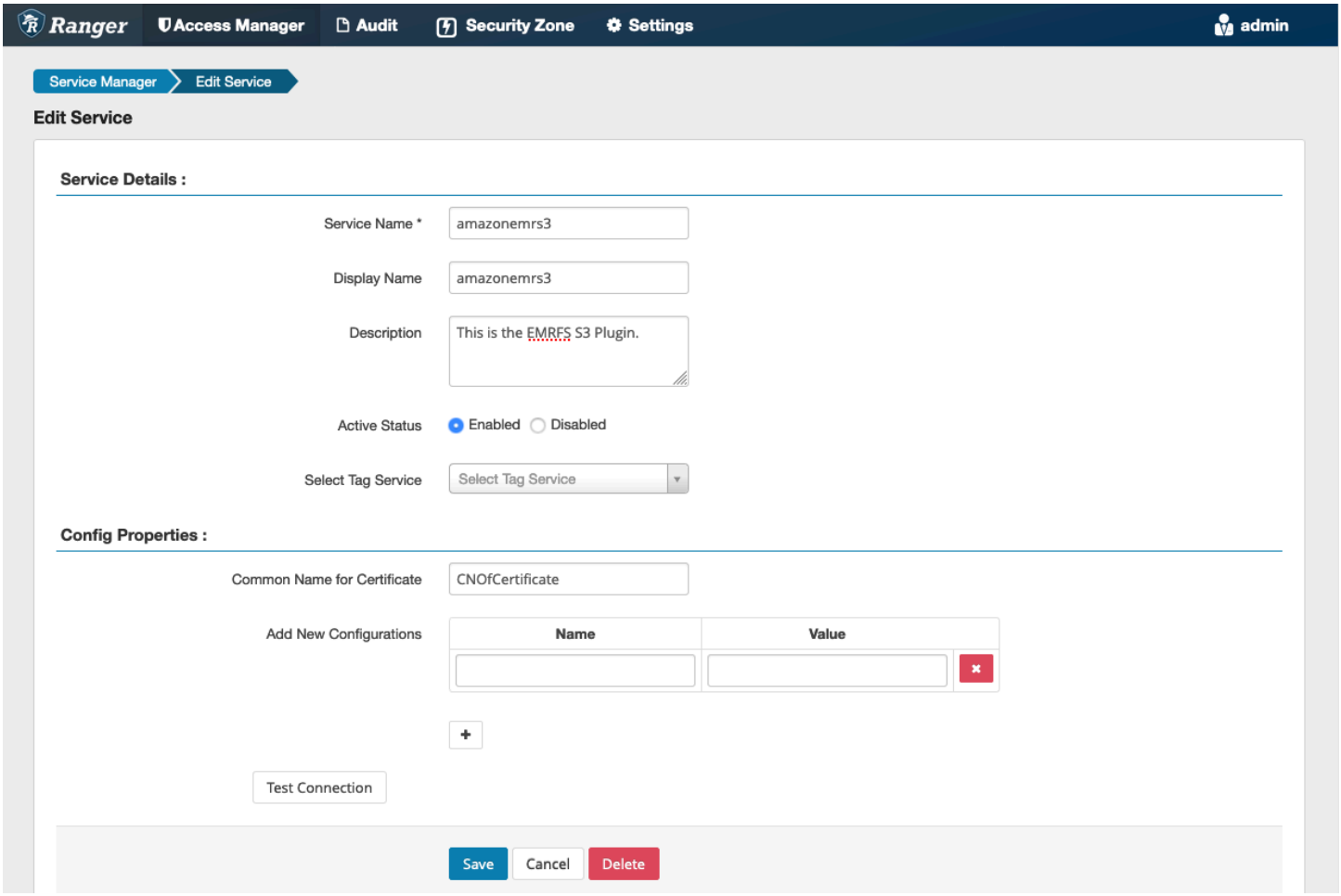
- Clique em + ao lado de AMAZON-EMR-EMRFS.

Preencha os seguintes campos:

Nome do serviço (se for exibido): o valor sugerido é **`amazonemrspark`**. Anote esse nome de serviço, pois ele será necessário ao criar uma configuração de segurança do EMR.

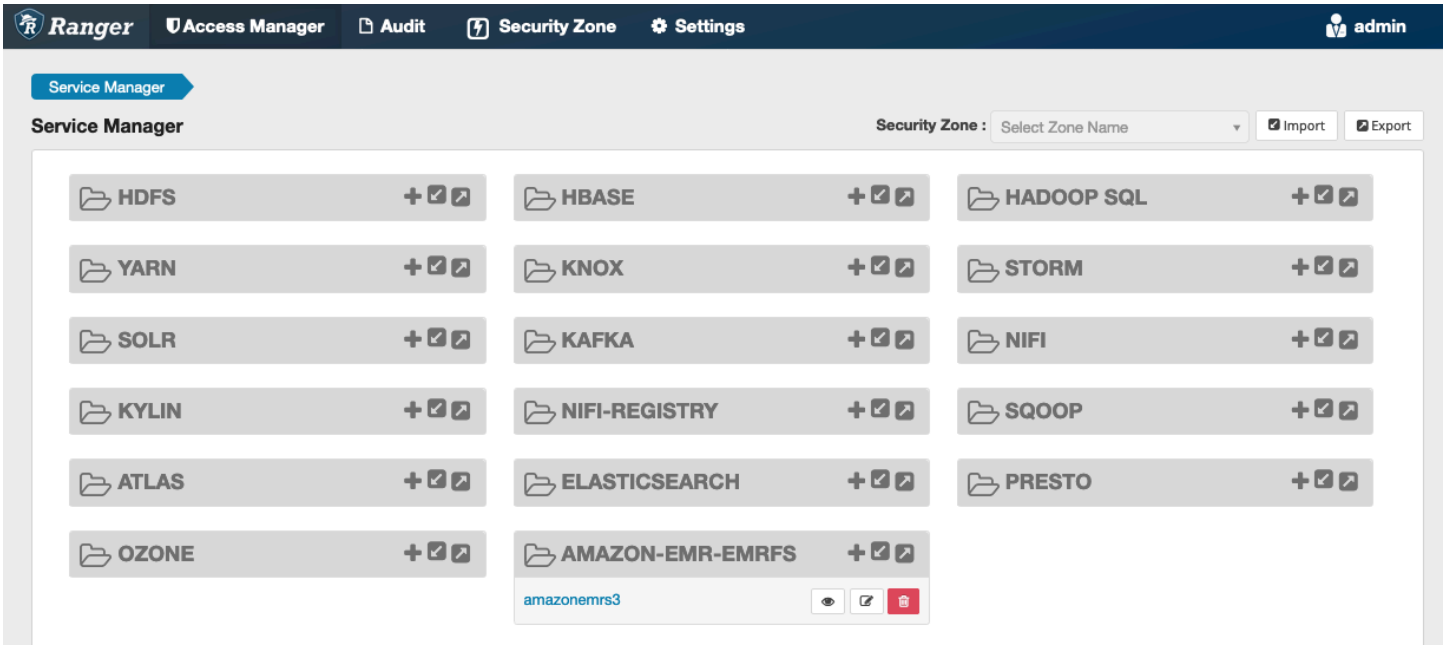
Nome de exibição: o nome exibido para o serviço. O valor sugerido é **`amazonemrspark`**.

Nome comum para certificado: o campo CN dentro do certificado usado para se conectar ao servidor de administração com base em um plug-in cliente. Esse valor deve corresponder ao campo CN no certificado TLS criado para o plug-in.



> **Note**
>
> O certificado TLS para o plug-in deveria ter sido registrado no armazenamento confiável do servidor Ranger Admin. Consulte Certificados TLS para obter mais detalhes.

Quando o serviço é criado, o Service Manager inclui “AMAZON-EMR-EMRFS”, conforme mostra a imagem a seguir.



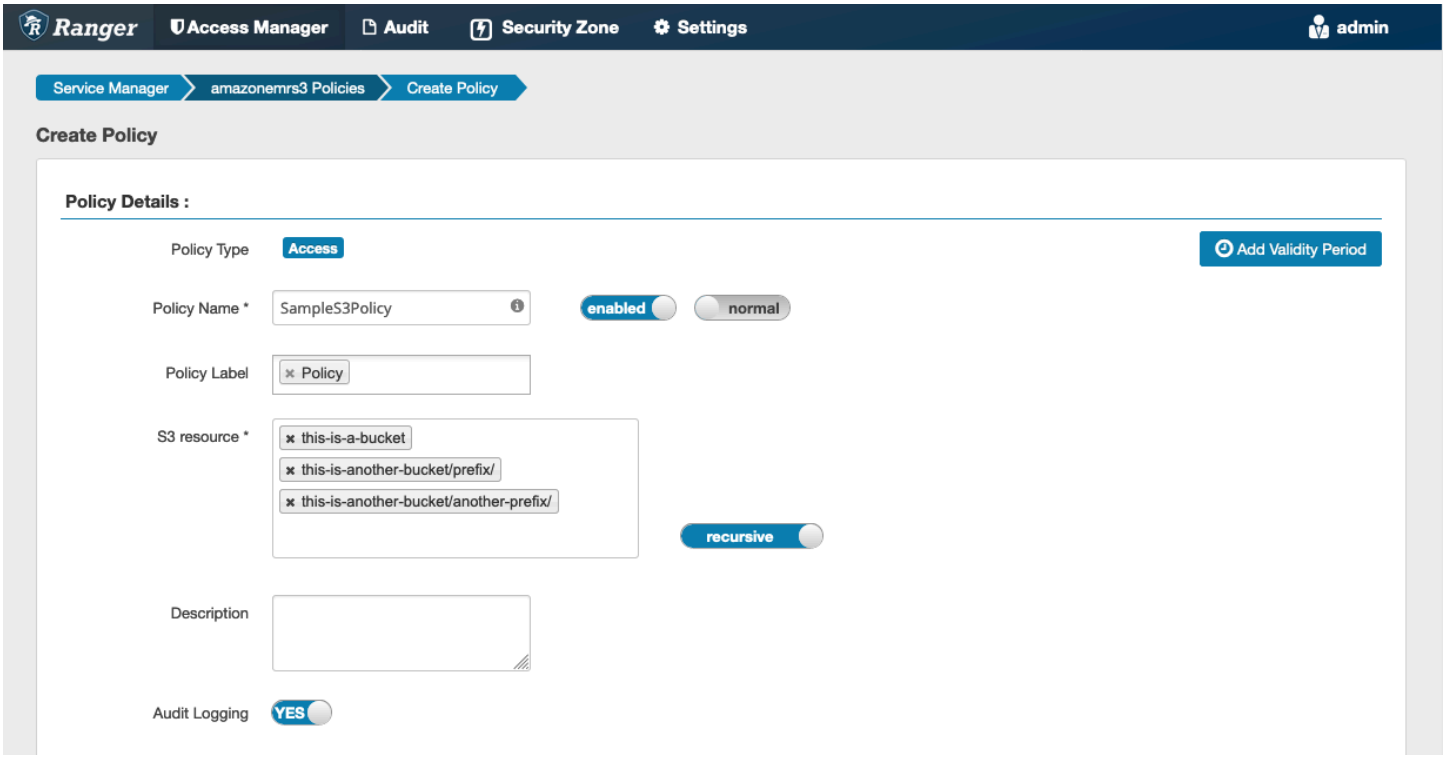
Criar políticas do EMRFS S3

Para criar uma nova política na página Criar política do Service Manager, preencha os campos a seguir.

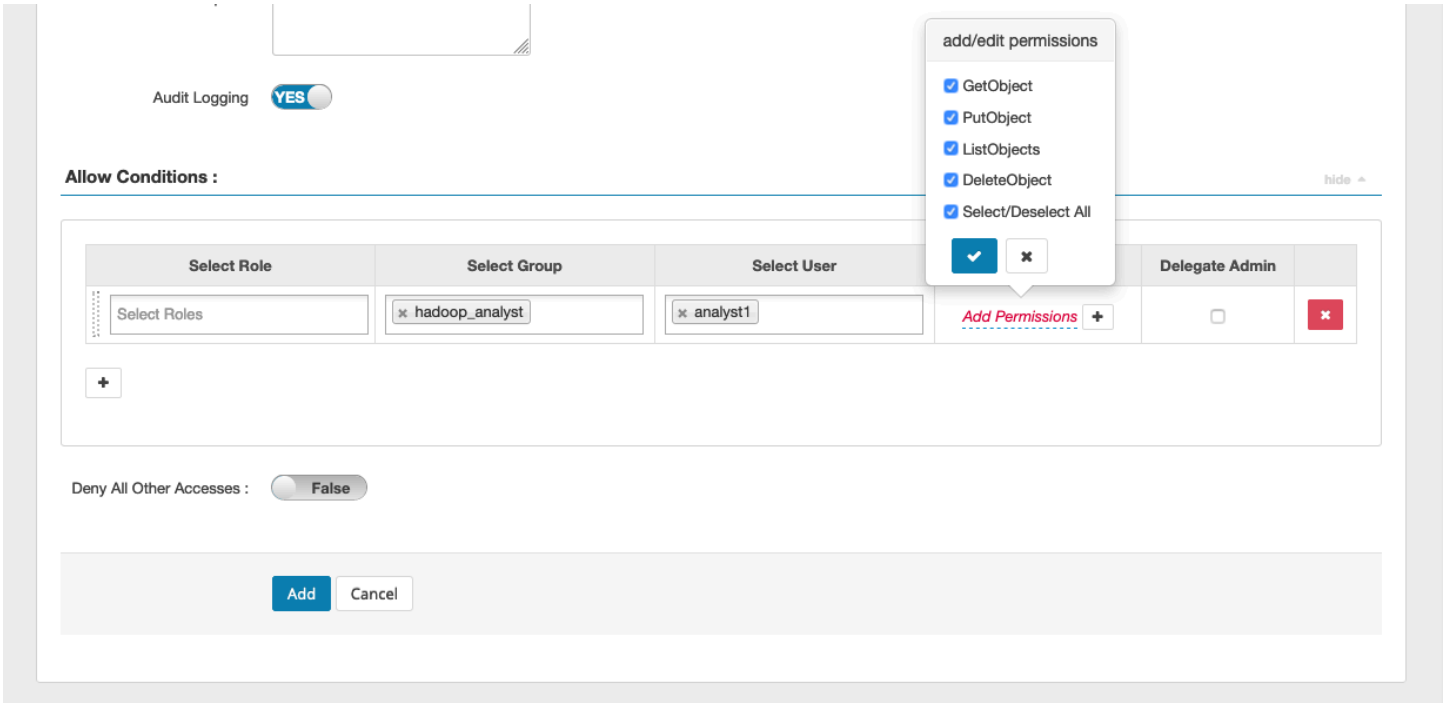
Nome da política: o nome da política.

Rótulo de política: um rótulo que você pode colocar na política.

Recurso do S3: um recurso que começa com o bucket e o prefixo opcional. Consulte Notas sobre o uso das políticas do EMRFS S3 para obter informações sobre práticas recomendadas. Os recursos no servidor Ranger Admin não devem conter **`s3://`**, **`s3a://`** ou **`s3n://`**.



É possível especificar usuários e grupos para conceder permissões. Também é possível especificar exclusões para condições de permissão e negação.



> **Note**
>
> São permitidos no máximo três recursos por política. Adicionar mais de três recursos poderá resultar em um erro quando essa política é usada em um cluster do EMR. Adicionar mais de três políticas exibirá um lembrete sobre o limite da política.

Notas sobre o uso das políticas do EMRFS S3

Ao criar políticas do S3 no Apache Ranger, atente para algumas considerações sobre o uso.

Permissões para múltiplos objetos do S3

É possível usar políticas recursivas e expressões curinga para conceder permissões a vários objetos do S3 com prefixos comuns. As políticas recursivas concedem permissões a todos os objetos com um prefixo comum. As expressões curinga selecionam múltiplos prefixos. Juntos, eles concedem permissões a todos os objetos com múltiplos prefixos comuns, conforme mostrado nos exemplos a seguir.

Example Usar uma política recursiva

Suponha que você queira permissões para listar todos os arquivos parquet em um bucket do S3 organizado da forma a seguir.

```
s3://sales-reports/americas/
    +- year=2000
    |      +- data-q1.parquet
    |      +- data-q2.parquet
    +- year=2019
    |      +- data-q1.json
    |      +- data-q2.json
    |      +- data-q3.json
    |      +- data-q4.json
    |
    +- year=2020
    |      +- data-q1.parquet
    |      +- data-q2.parquet
    |      +- data-q3.parquet
    |      +- data-q4.parquet
    |      +- annual-summary.parquet
    +- year=2021
```

Primeiro, considere os arquivos parquet que tenham o prefixo `s3://sales-reports/americas/year=2000`. Você pode conceder GetObject permissões a todos eles de duas maneiras:

Usar políticas não recursivas: uma opção é usar duas políticas não recursivas separadas, uma para o diretório e outra para os arquivos.

A primeira política concede permissão ao prefixo `s3://sales-reports/americas/year=2020` (não há / final).

```
- S3 resource = "sales-reports/americas/year=2000"
- permission = "GetObject"
- user = "analyst"
```

A segunda política usa a expressão curinga para conceder permissões a todos os arquivos com prefixo `sales-reports/americas/year=2020/` (observe o / final).

```
- S3 resource = "sales-reports/americas/year=2020/*"
- permission = "GetObject"
- user = "analyst"
```

Usar uma política recursiva: uma alternativa mais conveniente é usar uma única política recursiva e conceder permissão recursiva ao prefixo.

```
- S3 resource = "sales-reports/americas/year=2020"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Até agora, apenas os arquivos parquet com o prefixo `s3://sales-reports/americas/year=2000` foram incluídos. Também já é possível incluir os arquivos parquet com outro prefixo, `s3://sales-reports/americas/year=2020`, na mesma política recursiva introduzindo uma expressão curinga da forma a seguir.

```
- S3 resource = "sales-reports/americas/year=20?0"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Políticas PutObject e DeleteObject permissões

Escrever políticas `PutObject` e `DeleteObject` permissões para arquivos no EMRFS precisa de cuidados especiais porque, diferentemente das GetObject permissões, elas precisam de permissões recursivas adicionais concedidas ao prefixo.

Example Políticas PutObject e DeleteObject permissões

Por exemplo, excluir o arquivo `annual-summary.parquet` requer não apenas uma DeleteObject permissão para o arquivo real.

```
- S3 resource = "sales-reports/americas/year=2020/annual-summary.parquet"
- permission = "DeleteObject"
- user = "analyst"
```

Também requer uma política que conceda permissões `GetObject` e `PutObject` recursivas para o prefixo.

Da mesma forma, modificar o arquivo `annual-summary.parquet` requer não apenas uma permissão `PutObject` para o arquivo real.

```
- S3 resource = "sales-reports/americas/year=2020/annual-summary.parquet"
- permission = "PutObject"
- user = "analyst"
```

Também requer uma política que conceda a permissão `GetObject` recursiva para o prefixo.

```
- S3 resource = "sales-reports/americas/year=2020"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Curingas em políticas

Há duas áreas em que é possível especificar caracteres curingas. Ao especificar um recurso do S3, pode-se usar “*” e “?”. O “*” faz correspondência com um caminho do S3 e corresponde a tudo que está depois do prefixo. Por exemplo, a política a seguir.

```
S3 resource = "sales-reports/americas/*"
```

Isso corresponde aos caminhos do S3 a seguir.

```
sales-reports/americas/year=2020/
sales-reports/americas/year=2019/
sales-reports/americas/year=2019/month=12/day=1/afile.parquet
sales-reports/americas/year=2018/month=6/day=1/afile.parquet
sales-reports/americas/year=2017/afile.parquet
```

O curinga “?” corresponde a apenas um caractere. Por exemplo, para a política.

```
S3 resource = "sales-reports/americas/year=201?/"
```

Isso corresponde aos caminhos do S3 a seguir.

```
sales-reports/americas/year=2019/
sales-reports/americas/year=2018/
sales-reports/americas/year=2017/
```

Curingas em usuários

Há dois curingas integrados ao atribuir usuários para fornecer acesso aos usuários. O primeiro é o curinga “{USER}” que concede acesso a todos os usuários. O segundo caractere curinga é “{OWNER}”, que concede acesso direto ao proprietário de um objeto específico. No entanto, atualmente não há suporte para o curinga “{USER}”.

Limitações

Estas são as limitações atuais do plug-in EMRFS S3:

- As políticas do Apache Ranger podem conter no máximo três políticas.
- O acesso ao S3 deve ser feito pelo EMRFS e pode ser usado com aplicações relacionadas ao Hadoop. Não há suporte para:

  - Bibliotecas Boto3

  - AWS SDK e AWK CLI

  - Conector de código aberto S3A
- Não há suporte para políticas de negação do Apache Ranger.
- Atualmente, não há suporte para operações no S3 com chaves com criptografia do CSE-KMS.

- O suporte entre regiões não é compatível.
- Não há suporte para o atributo de zona de segurança do Apache Ranger. As restrições de controle de acesso definidas usando o atributo Security Zone não são aplicadas aos clusters do Amazon EMR.
- O usuário do Hadoop não gera nenhum evento de auditoria, pois o Hadoop sempre acessa o perfil de instância do EC2.
- É recomendável desabilitar a visualização consistente do Amazon EMR. O S3 do tem um alto nível de consistência e, portanto, isso não é mais necessário. Para obter mais informações, consulte Amazon S3 strong consistency.
- O plug-in EMRFS S3 efetua várias chamadas STS. É recomendável fazer testes de carga em uma conta de desenvolvimento e monitorar o volume de chamadas do STS. Também é recomendável que você faça uma solicitação STS para aumentar os limites do AssumeRole serviço.
- O servidor Ranger Admin não oferece suporte ao preenchimento automático.

Plug-in Trino

O Trino (antes chamado PrestoSQL) é um mecanismo de consulta SQL que pode ser usado para executar consultas em fontes de dados como HDFS, armazenamento de objetos, bancos de dados relacionais e bancos de dados NoSQL. Ele elimina a necessidade de migrar dados para um local central e permite a consulta de dados de qualquer lugar. O Amazon EMR fornece um plug-in Apache Ranger para proporcionar controles de acesso refinados para o Trino. O plug-in é compatível com o servidor Apache Ranger Admin de código aberto versão 2.0 e posteriores.

Tópicos

Atributos compatíveis

O plug-in Apache Ranger para Trino no Amazon EMR oferece suporte a todos os recursos do mecanismo de consulta Trino, que é protegido por um controle de acesso refinado. Isso inclui controles de acesso em nível de banco de dados, de tabela e de coluna, filtragem de linhas e

mascaramento de dados. As políticas do Apache Ranger podem incluir políticas de concessão e políticas de negação para usuários e grupos. Os eventos de auditoria também são enviados aos CloudWatch registros.

Instalação da configuração de serviço

A instalação da definição de serviço Trino requer que o servidor Ranger Admin esteja configurado. Para configurar o servidor Ranger Admin, consulte Configurar o servidor do Ranger Admin.

Siga estas etapas para instalar a definição de serviço do Trino.

1. SSH no servidor Apache Ranger Admin.

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

2. Desinstale o plug-in do servidor Presto, se houver. Execute o seguinte comando . Se isso ocorrer com o erro “Serviço não encontrado”, significa que o plug-in do servidor Presto não foi instalado no servidor. Prossiga para a próxima etapa.

```
curl -f -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin
 user_**_>_* -X DELETE -k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/
v2/api/servicedef/name/presto'
```

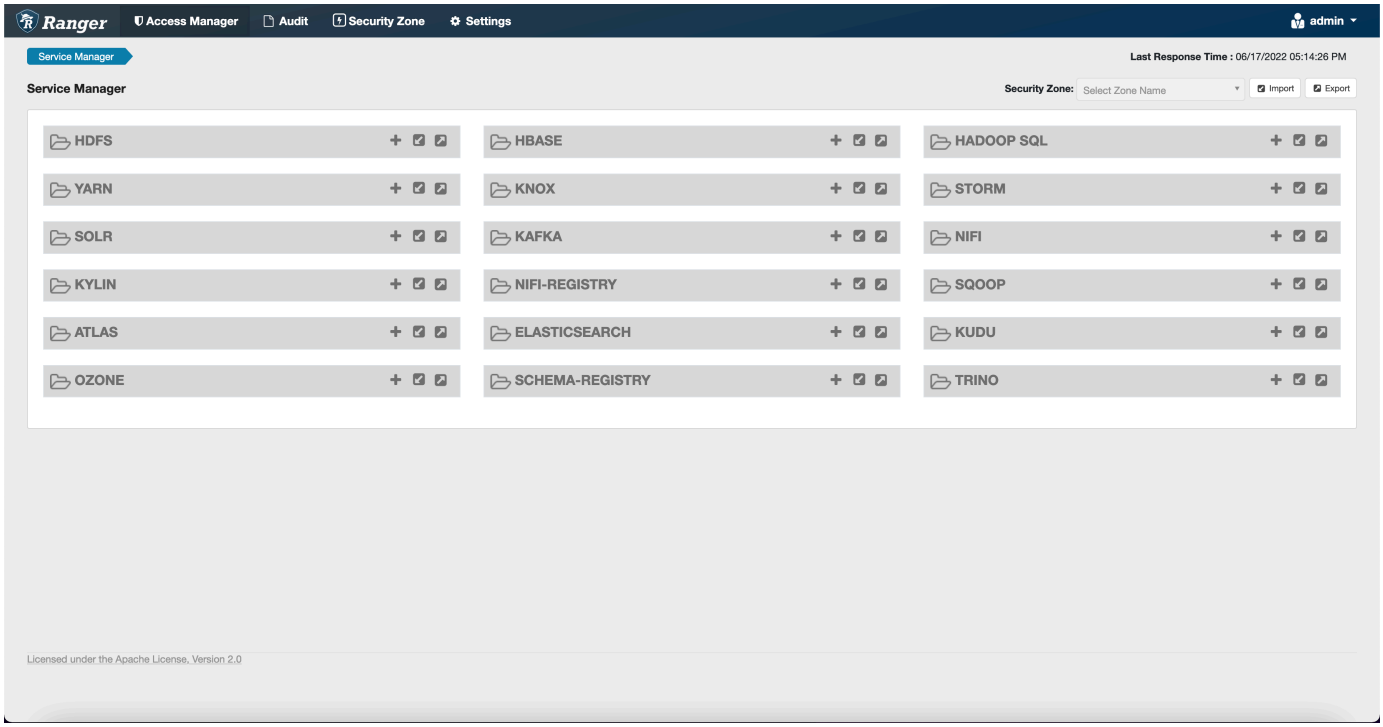
3. Baixe a definição de serviço e o plug-in do servidor Apache Ranger Admin. Em um diretório temporário, baixe a definição de serviço. Essa definição de serviço é compatível com as versões Ranger 2.x.

```
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/
version-2.0/ranger-servicedef-amazon-emr-trino.json
```

4. Registre a definição de serviço do Apache Trino para o Amazon EMR.

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_*
 -X POST -d @ranger-servicedef-amazon-emr-trino.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

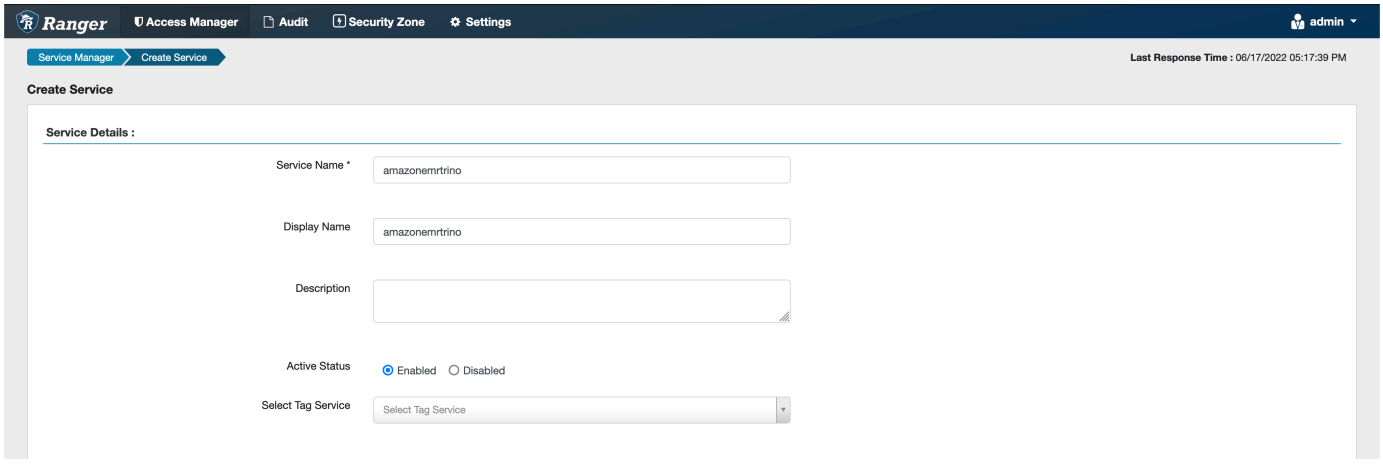
   Se esse comando for executado com êxito, você verá um novo serviço na IU do Ranger Admin chamado `TRINO`, conforme mostrado na imagem.



5. Crie uma instância da aplicação `TRINO`, inserindo as informações a seguir.

   Nome do serviço: o nome do serviço que você usará. O valor sugerido é `amazonemrtrino`. Anote esse nome de serviço, pois ele será necessário ao criar uma configuração de segurança do Amazon EMR.

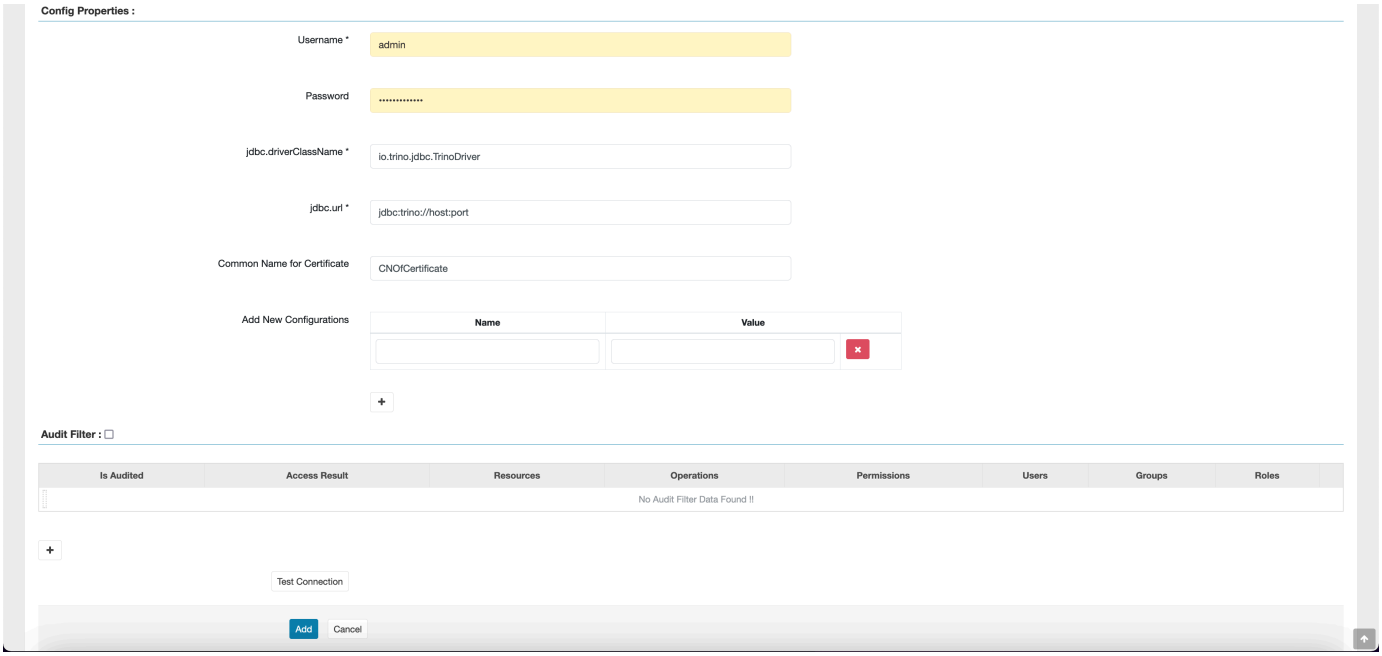
   Nome de exibição: o nome a ser exibido para a instância. O valor sugerido é `amazonemrtrino`.



   jdbc. driver. ClassName: O nome da classe JDBC para conectividade Trino. Você pode usar o valor padrão.

   jdbc.url: a string de conexão JDBC a ser usada ao se conectar ao coordenador Trino.

Nome comum para certificado: o campo CN dentro do certificado usado para se conectar ao servidor de administração com base em um plug-in cliente. Esse valor deve corresponder ao campo CN do certificado TLS que foi criado para o plug-in.



Observe que o certificado TLS para o plug-in deveria ter sido registrado no armazenamento confiável do servidor Ranger Admin. Para obter mais informações, consulte TLS certificates.

Criar políticas do Trino

Ao criar uma nova política, preencha os campos a seguir.

Nome da política: o nome da política.

Rótulo de política: um rótulo que você pode colocar na política.

Catálogo: o catálogo ao qual a política se aplica. O curinga “*” representa todos os catálogos.

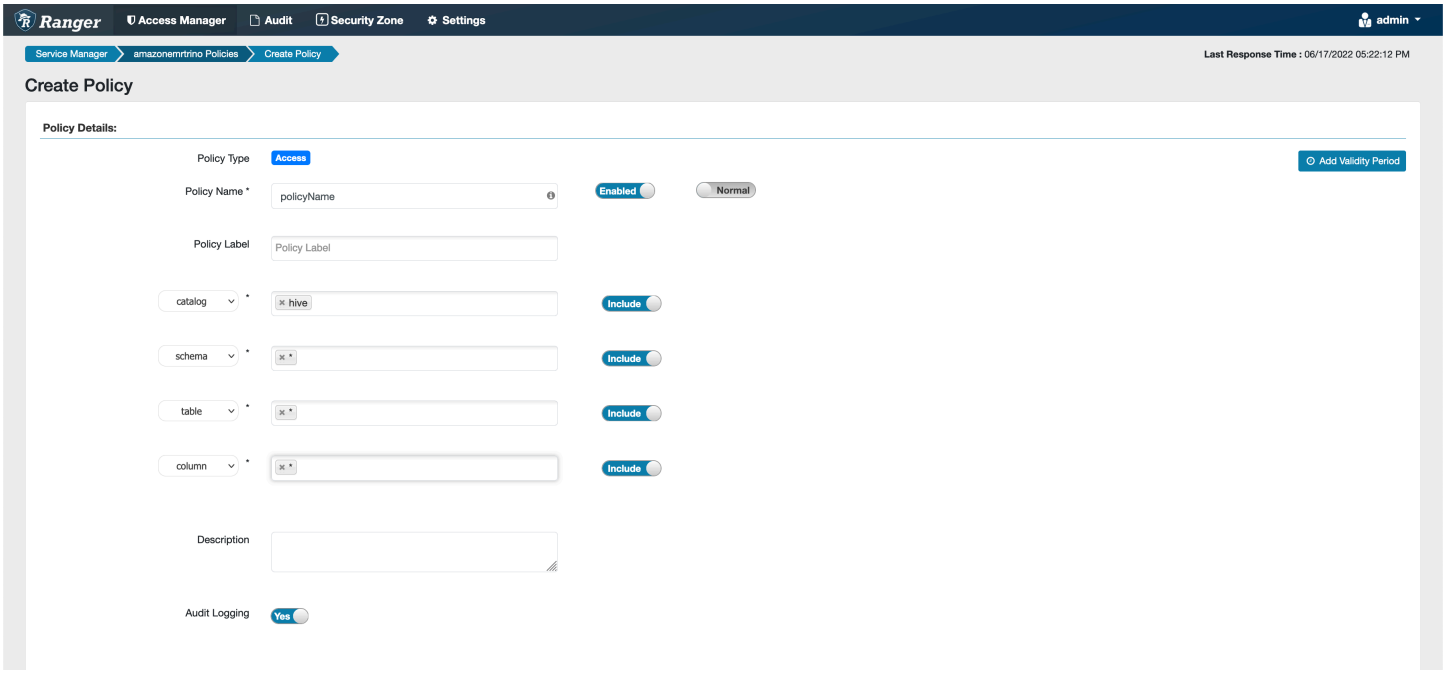
Esquema: os esquemas aos quais a política se aplica. O curinga “*” representa todos os esquemas.

Tabela: as tabelas às quais a política se aplica. O caractere curinga “*” representa todas as tabelas.

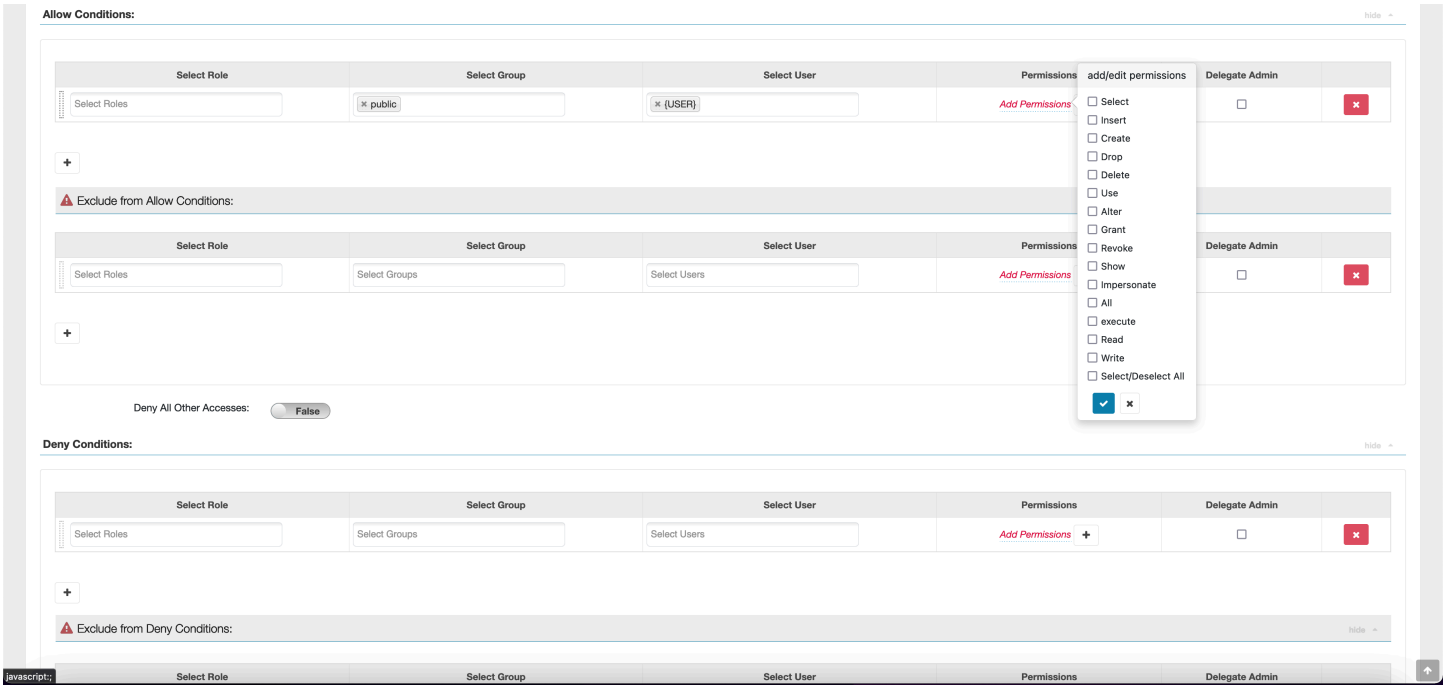
Coluna: as colunas às quais a política se aplica. O caractere curinga “*” representa todas as colunas.

Descrição: uma descrição da política.

Há outros tipos de políticas para o usuário Trino (para acesso à representação do usuário), a propriedade do sistema/sessão Trino (para alterar o sistema do mecanismo ou as propriedades da sessão), funções/procedimentos (para permitir chamadas de função ou procedimento) e o URL (para conceder acesso de leitura/gravação ao mecanismo em locais de dados).



Para conceder permissões a usuários e grupos específicos, insira os usuários e grupos. Também é possível especificar exclusões para condições de permissão e negação.



Após especificar as condições de permitir e negar, escolha Salvar.

Considerações

Ao criar políticas do Trino no Apache Ranger, atente para algumas considerações sobre o uso.

Servidor de metadados Hive

O servidor de metadados Hive só pode ser acessado por mecanismos confiáveis, especificamente o mecanismo Trino, para proteção contra acesso não autorizado. O servidor de metadados Hive também é acessado por todos os nós do cluster. A porta 9083 necessária fornece acesso de todos os nós ao nó principal.

Autenticação

Por padrão, o Trino é configurado para se autenticar usando Kerberos conforme definido na configuração de segurança do Amazon EMR.

A criptografia em trânsito é obrigatória

O plug-in Trino exige que a criptografia em trânsito esteja habilitada na configuração de segurança do Amazon EMR. Para ativar a criptografia, consulte Criptografia em trânsito.

Limitações

Estas são as limitações atuais do plug-in Trino:

- O servidor Ranger Admin não oferece suporte ao preenchimento automático.

Solução de problemas do Apache Ranger

Aqui estão alguns problemas diagnosticados com frequência relacionados ao uso do Apache Ranger.

Recomendações

- Teste usando um único cluster de nó principal: clusters principais de nó único são provisionados mais rapidamente do que um cluster de múltiplos nós, o que pode diminuir o tempo de cada iteração de teste.
- Defina o modo de desenvolvimento no cluster. Ao iniciar o cluster do EMR, defina o parâmetro `--additional-info"` como:

```
'{"clusterType":"development"}'
```

Esse parâmetro só pode ser definido por meio da AWS CLI ou do AWS SDK e não está disponível no console do Amazon EMR. Quando esse sinalizador é definido e o principal falha no provisionamento, o serviço Amazon EMR mantém o cluster ativo por algum tempo antes de desativá-lo. Esse momento é muito útil para testar vários arquivos de log antes que o cluster seja terminado.

Falha no provisionamento do cluster do EMR

Há vários motivos para um cluster do Amazon EMR poder falhar ao iniciar. Veja aqui algumas maneiras de diagnosticar o problema.

Verificar os logs de provisionamento do EMR

O Amazon EMR usa o Puppet para instalar e configurar aplicações em um cluster. A análise dos logs fornecerá detalhes sobre a ocorrência de erros durante a fase de provisionamento de um cluster. Os logs podem ser acessados no cluster ou no S3 se os logs estiverem configurados para serem enviados ao S3.

Os logs são armazenados em `/var/log/provision-node/apps-phase/0/{UUID}/puppet.log` no disco e em `s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/provision-node/apps-phase/0/{UUID}/puppet.log.gz`.

Mensagens de erro comuns

Mensagem de erro	Causa
Puppet (err): Falha na inicialização do Systemd! emr-record-server log journalctl para emr-record-server:	Falha ao iniciar o EMR Record Server. Veja abaixo os logs do EMR Record Server.
Puppet (err): Falha na inicialização do Systemd! emr-record-server registro journalctl para emrsecretagent:	O agente secreto do EMR falhou ao iniciar. Veja abaixo os logs do agente secreto.

Mensagem de erro	Causa
/Stage[main]/Ranger_plugins::Ranger_hive_plugin/Ranger_plugins::Prepare_two_way_tls[configure 2-way TLS in Hive plugin]/Exec[create keystore and truststore for Ranger Hive plugin]/returns (notice): 140408606197664:error:0906D06C:PEM routines:PEM_read_bio:no start line:pem_lib.c:707:Expecting: ANY PRIVATE KEY	O certificado TLS privado do Secret Manager para o certificado do plug-in Apache Ranger não está no formato correto ou não é um certificado privado. Consulte Certificados TLS para ver os formatos de certificado.
/Stage [main] /Ranger_Plugins: :Ranger_S3_Plugin/Ranger_Plugins: :Prepare_TWO_way_TLS [configure TLS bidirecional no plugin Ranger s3] /Exec [crie keystore e truststore para o plugin Ranger 3] /returns (aviso): Ocorreu um erro (exceção) ao chamar a operação: Usuário: arn:aws:sts: XXXXXXX:assumed-role/EMR_EC2_ /i-XXXXXXXXXXXX não está autorizado a executar: secretsmanager: Valor no recurso: arn:aws:secretsmanager:us-east-1:xxxxxxxxxx:secret: amazon-emr-s -XXXXX AccessDenied GetSecretValue DefaultRole GetSecret AdminServer	O perfil do perfil de instância do EC2 não tem as permissões corretas para recuperar os certificados TLS do Secrets Agent.

Verifique SecretAgent os registros

Os logs do Secret Agent estão localizados em `/emr/secretagent/log/` em um nó do EMR ou no diretório `s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/daemons/secretagent/` do S3.

Mensagens de erro comuns

Mensagem de erro	Causa

Mensagem de erro	Causa
Exceção no tópico “main” com.amazonaws.services.securitytoken.model. AWSSecurityTokenServiceException: Usuário: arn:aws:sts: :xxxxxxxxxxxx:assumed-role/emr_ec2_DefaultRole /i-xxxxxxxxxxxxxxxxxxx não está autorizado a executar: sts: no recurso: arn:aws:iam: :xxxxxxxxxxxx:role/* Role* (Serviço:; Código de status: 403; Código de erro:; ID da solicitação: XXXXXXXX-XXXX-XXXX-XXXX-XXXXXX-XXXXXXXXXX; AssumeRole Proxy: nulo) RangerPlugin DataAccess AWSSecurityTokenService AccessDenied	A exceção acima significa que a função de perfil da instância EC2 do EMR não tem permissões para assumir a função. RangerPlugin DataAccess Consulte Perfis do IAM para integração nativa com o Apache Ranger.
ERROR qtp54617902-149: Web App Exception Occurred javax.ws.rs. NotAllowedExceção: método HTTP 405 não permitido	Esses erros podem ser ignorados com segurança.

Verificar logs do Record Server (para SparkSQL)

Os logs do EMR Record Server estão disponíveis em /var/log/emr-record-server/ em um nó do EMR, ou podem ser encontrados no diretório s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/daemons/emr-record-server/ do S3.

Mensagens de erro comuns

Mensagem de erro	Causa
InstanceMetadataServiceResourceFetcher:105 - [] Falha ao recuperar o token com.amazonaws. SdkClientExceção: falha na conexão com o endpoint de serviço	O EMR SecretAgent não apareceu ou está com problemas. Inspecione os SecretAgent registros em busca de erros e o script de

Mensagem de erro	Causa
	marionete para determinar se houve algum erro de provisionamento.

As consultas estão falhando inesperadamente

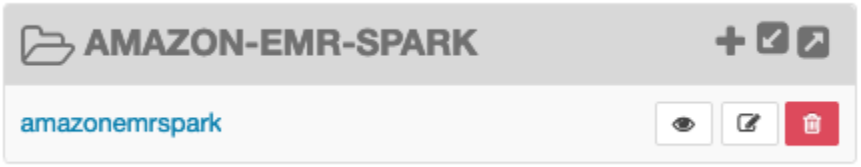
Verifique os registros do plug-in Apache Ranger (registros do Apache Hive, RecordServer EMR, EMR, SecretAgent etc.)

Essa seção é comum em todos os aplicativos que se integram ao plug-in Ranger, como Apache Hive, EMR Record Server e EMR. SecretAgent

Mensagens de erro comuns

Mensagem de erro	Causa
ERROR:272 PolicyRefresher - [] (PolicyRefresherserviceName=policy-repository): falha ao encontrar o serviço. Limpará o cache local de políticas (-1)	Essas mensagens de erro significam que o nome do serviço que você forneceu na configuração de segurança do EMR não corresponde a um repositório de políticas de serviço no servidor Ranger Admin.

Se, no servidor Ranger Admin, o serviço AMAZON-EMR-SPARK for semelhante ao exemplo a seguir, você deverá inserir **`amazonemrspark`** como nome do serviço.



Trabalhando com visualizações do AWS Glue Data Catalog (pré-visualização)

> **Note**
>
> AWS As visualizações do Glue Data Catalog no Amazon EMR estão em versão prévia e estão sujeitas a alterações. O recurso é fornecido como um serviço de Pré-visualização, conforme definido nos Termos de Serviço da AWS.

Você pode criar e gerenciar visualizações comuns únicas no AWS Glue Data Catalog. Visualizações comuns únicas são úteis porque oferecem suporte a vários mecanismos de consulta SQL, para que você possa acessar a mesma visualização em diferentes Serviços da AWS, como Amazon EMR, Amazon Athena e Amazon Redshift.

Ao criar uma exibição no Catálogo de Dados, você pode usar concessões de recursos e controles de acesso baseados em tags AWS Lake Formation para conceder acesso a uma exibição do Catálogo de Dados. Usando esse método de controle de acesso, você não precisa configurar acesso adicional às tabelas referenciadas ao criar a exibição. Esse método de concessão de permissões é chamado de semântica definidora, e essas visualizações são chamadas de visualizações definidoras. Para obter mais informações sobre controle de acesso no Lake Formation, consulte Conceder e revogar permissões nos recursos do Catálogo de Dados. no Guia do AWS Lake Formation desenvolvedor.

As visualizações do Catálogo de Dados são úteis para os seguintes casos de uso:

- Controle de acesso granular — crie uma visualização que restrinja o acesso aos dados com base nas permissões de que o usuário precisa. Por exemplo, você pode usar as exibições do Data Catalog para evitar que funcionários que não trabalham no departamento de RH vejam informações de identificação pessoal (PII).
- Definição completa da visualização — ao aplicar determinados filtros à sua exibição no Catálogo de Dados, você garante que os registros de dados dentro de uma exibição no Catálogo de Dados estejam sempre completos.
- Segurança aprimorada — a definição da consulta usada para criar a exibição deve estar completa. Esse benefício significa que as visualizações no Catálogo de Dados são menos suscetíveis aos comandos SQL de jogadores mal-intencionados.

- Compartilhamento simples de dados — compartilhe dados com outras Contas da AWS pessoas sem mover nenhum dado. Para obter mais informações, consulte Compartilhamento de dados entre contas no Lake Formation.

Criação de uma visualização do Catálogo de Dados

> **Important**
>
> Durante essa versão prévia, o Amazon EMR não valida o Spark-SQL que você usa ao criar a visualização. Para reduzir os riscos, recomendamos que você limite os usuários aos quais você concede permissões de criação de visualizações.

Para criar uma visualização do catálogo de dados, você deve usar uma função do IAM que tenha a `SELECT` permissão total com `Grantable` opções em todas as tabelas que você deseja referenciar ao criar a visualização. Essa função é chamada de função definidora. Para obter uma lista completa das permissões e pré-requisitos necessários para criar uma visualização do Catálogo de Dados, consulte Como trabalhar com exibições no Guia do AWS Lake Formation Desenvolvedor. Você deve usar o AWS CLI para configurar sua função do IAM. Consulte Usar uma função do IAM no AWS CLI para obter mais informações.

Siga estas etapas para criar uma exibição do Catálogo de Dados.

> **Note**
>
> Para acessar uma visualização do catálogo de dados do Apache Spark no Amazon EMR, você deve definir o dialeto como e para. SPARK `DialectVersion` `3.4.1-amzn-2`

1. Primeiro, baixe o modelo de pré-visualização.

```
aws s3 cp s3://emr-data-access-control-us-east-1/beta/glue-views/model/
service-2.json
```

2. Configure o AWS CLI para usar o modelo de pré-visualização.

```
aws configure add-model --service-model file:///<path-to-preview-model>/
service-2.json --service-name glue-views
```

3. Crie a visualização.

```
aws glue-views create-table --cli-input-json '{
  "DatabaseName": "<database>",
  "TableInput": {
    "Name": "<view>",
    "StorageDescriptor": {
      "Columns": [
        {
          "Name": "<col1>",
          "Type": "<data-type>"
        },
        ...
        {
          "Name": "<colN>",
          "Type": "<data-type>"
        }
      ]
    },
    "ViewDefinition": {
      "SubObjects": [
        "arn:aws:glue:<aws-region;>:<aws-account-id>:table/<database>/<referenced-
table1>",
        ...
        "arn:aws:glue:<aws-region>:<aws-account-id>:table/<database>/<referenced-
tableN>",
       ],
       "IsProtected": true,
      "Representations": [
        {
          "Dialect": "SPARK",
          "DialectVersion": "3.4.1-amzn-2",
          "ViewOriginalText": "<Spark-SQL>",
          "ViewExpandedText": "<Spark-SQL>"
        }
      ]
    }
  }
}'
```

Habilitando o acesso a uma visualização do Catálogo de Dados

> **Important**
>
> Recomendamos que você habilite o acesso às visualizações do catálogo de dados somente com clusters do EMR em ambientes de teste e não em ambientes de produção.

Para acessar a visualização do catálogo de dados do Apache Spark no Amazon EMR, primeiro você deve habilitar o suporte para Lake Formation e usar o script abaixo para habilitar o suporte para visualizações com o Spark no Amazon EMR. Para obter mais informações sobre como habilitar o suporte, consulte Habilitar o Lake Formation com o Amazon EMR e Usar ações de bootstrap personalizadas.

```
# Download the script and upload it to Amazon S3
wget https://emr-data-access-control-us-east-1.s3.amazonaws.com/beta/glue-views/ba/
enable-mdv.sh /Users/$USER/enable-mdv.sh
aws s3 cp /Users/$USER/enable-views.sh s3://<bucket>/<prefix>/enable-views.sh

# EMR Security Configuration
cat <<EOT > /Users/$USER/lakeformation-protection.json
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true
        },
        "LakeFormationConfiguration":{
            "AuthorizedSessionTagValue":"Amazon EMR"
        }
    },
     "EncryptionConfiguration": {
            "EnableInTransitEncryption": true,
            "InTransitEncryptionConfiguration": {
            "TLSCertificateConfiguration": {
                 "CertificateProviderType": "PEM",
                 "S3Object": "s3://<BUCKET>/<PREFIX>/certificates.zip"
            }
        }
    }
}
EOT
```

```
SECURITY_CONFIG="RuntimeRolesWithAWSLakeFormation"

aws emr create-security-configuration \
--name $SECURITY_CONFIG \
--security-configuration file:///Users/$USER/lakeformation-protection.json

# EMR Cluster version
RELEASE_LABEL="emr-6.15.0"
```

Em seguida, use o AWS CLI comando a seguir que usa a ação bootstrap para criar um cluster EMR compatível com visualizações do catálogo de dados.

```
aws emr create-cluster \
...
--release-label $RELEASE_LABEL \
--security-configuration $SECURITY_CONFIG \
--bootstrap-actions \
 Name='Enable Views',Path="s3://<bucket>/<prefix>/enable-views.sh"
```

Consulta de uma visualização do Catálogo de Dados

> **Important**
>
> Durante esta versão prévia, recomendamos que você acesse exibições somente de fontes confiáveis. Na versão prévia, o Amazon EMR tem uma quantidade limitada de validações que protegem seu cluster do EMR.

Depois de criar uma visualização do catálogo de dados, agora você pode usar uma função do IAM para consultar a visualização. A função do IAM deve ter a `SELECT` permissão na visualização do catálogo de dados. Você não precisa conceder acesso às tabelas subjacentes mencionadas na exibição. Você deve usar essa função do IAM como uma função de tempo de execução. Você pode acessar a visualização de um cluster do EMR usando uma função de tempo de execução das etapas do Amazon EMR, do EMR Studio e do Studio. SageMaker Para obter mais informações sobre funções de tempo de execução, consulte Funções de tempo de execução para etapas do Amazon EMR.

Depois de configurar tudo, você pode consultar sua visualização. Por exemplo, depois de anexar o cluster do EMR ao seu espaço de trabalho no EMR Studio, você pode executar a consulta a seguir para acessar uma visualização.

```
SELECT * from <database>.<glue-data-catalog-view> LIMIT 10
```

Limitações

Considere as seguintes limitações ao usar as exibições do Catálogo de Dados.

- Você só pode criar visualizações do catálogo de dados com o Amazon EMR 6.15.0.
- Você só pode referenciar até 10 tabelas na definição da exibição.
- Você só pode criar visualizações do Catálogo de `PROTECTED` Dados. `UNPROTECTED`visualizações não são suportadas.
- Você não pode referenciar tabelas em outras Conta da AWS nas visualizações do Catálogo de Dados.
- As funções definidas pelo usuário (UDFs) não são suportadas.
- Você não pode referenciar formatos de tabela aberta, como Apache Hudi ou Apache Iceberg, nas visualizações do Catálogo de Dados.
- Você não pode referenciar outras visualizações nas visualizações do Catálogo de Dados.

Controle do tráfego de rede com grupos de segurança

Grupos de segurança atuam como firewalls virtuais para suas instâncias do EC2 em seu cluster para controlar o tráfego de entrada e saída. Cada grupo de segurança tem um conjunto de regras que controla o tráfego de entrada para as instâncias e um conjunto de regras separado para controlar o tráfego de saída. Para obter mais informações, consulte Grupos de segurança do Amazon EC2 para instâncias do Linux no Guia do usuário do Amazon EC2.

Você usa duas classes de grupos de segurança com o Amazon EMR: grupos de segurança gerenciados pelo Amazon EMR e grupos de segurança adicionais.

Cada cluster tem grupos de segurança gerenciados associados a ele. Você pode usar grupos de segurança gerenciados padrão que o Amazon EMR cria ou especificar grupos de segurança gerenciados personalizados. De qualquer forma, o Amazon EMR adiciona automaticamente regras

aos grupos de segurança gerenciados que um cluster precisa para se comunicar entre instâncias e AWS serviços do cluster.

Grupos de segurança adicionais são opcionais. Você pode especificá-los além dos grupos de segurança gerenciados para personalizar o acesso às instâncias do cluster. Os grupos de segurança adicionais contêm apenas regras definidas por você. O Amazon EMR não os modifica.

As regras que o Amazon EMR cria em grupos de segurança gerenciados permitem que o cluster se comunique entre componentes internos. Para permitir que usuários e aplicativos acessem um cluster de fora do cluster, você pode editar regras em grupos de segurança gerenciados, criar grupos de segurança adicionais com regras adicionais ou ambos.

> **Important**
>
> A edição de regras em grupos de segurança gerenciados pode ter consequências inesperadas. Você pode bloquear acidentalmente o tráfego necessário para os clusters funcionarem corretamente e causar erros porque os nós estão inacessíveis. Tenha cuidado ao planejar e testar configurações de grupo de segurança antes da implementação.

Você pode especificar grupos de segurança somente ao criar um cluster. Eles não podem ser adicionados a um cluster ou instâncias do cluster enquanto um cluster está em execução, mas é possível editar, adicionar e remover regras de grupos de segurança existentes. As regras entram em vigor assim que você as salva.

Grupos de segurança são restritivos por padrão. A menos que seja adicionada uma regra que permita o tráfego, ele é rejeitado. Se houver mais de uma regra que se aplique ao mesmo tráfego e à mesma origem, a regra mais permissiva será aplicada. Por exemplo, se você tiver uma regra que permita SSH do endereço IP 192.0.2.12/32 e outra regra que permita acesso a todo o tráfego TCP do mesmo intervalo 192.0.2.0/24, a regra que permitir todo o tráfego TCP do intervalo que inclua 192.0.2.12 terá precedência. Nesse caso, o cliente em 192.0.2.12 pode ter mais acesso do que o desejado.

> **Important**
>
> Tome cuidado ao editar as regras de grupo de segurança para portas abertas. Adicione regras que só permitam tráfego de clientes confiáveis e autenticados para os protocolos e portas que sejam necessários para executar suas workloads.

Você poderá configurar o bloqueio de acesso público do Amazon EMR em cada região usada para impedir a criação de cluster se uma regra permitir acesso público em qualquer porta que você não adicionar a uma lista de exceções. Para AWS contas criadas após julho de 2019, o bloqueio de acesso público do Amazon EMR está ativado por padrão. Para AWS contas que criaram um cluster antes de julho de 2019, o bloqueio de acesso público do Amazon EMR está desativado por padrão. Para ter mais informações, consulte Usar o bloqueio de acesso público do Amazon EMR.

Tópicos

> **Note**
>
> O Amazon EMR procura usar alternativas inclusivas para termos setoriais potencialmente ofensivos ou não inclusivos, como “mestre” e “escravo”. Fizemos a transição para uma nova terminologia para promover uma experiência mais inclusiva e facilitar a compreensão dos componentes do serviço.
> Agora descrevemos “nós” como instâncias e descrevemos os tipos de instância do Amazon EMR como instâncias primárias, centrais e de tarefa. Durante a transição, ainda é possível encontrar referências antigas a termos desatualizados, como aqueles que dizem respeito aos grupos de segurança do Amazon EMR.

Trabalhar com grupos de segurança gerenciados pelo Amazon EMR

> **Note**
>
> O Amazon EMR procura usar alternativas inclusivas para termos setoriais potencialmente ofensivos ou não inclusivos, como “mestre” e “escravo”. Fizemos a transição para uma nova terminologia para promover uma experiência mais inclusiva e facilitar a compreensão dos componentes do serviço.
> Agora descrevemos “nós” como instâncias e descrevemos os tipos de instância do Amazon EMR como instâncias primárias, centrais e de tarefa. Durante a transição, ainda é possível

encontrar referências antigas a termos desatualizados, como aqueles que dizem respeito aos grupos de segurança do Amazon EMR.

Diferentes grupos de segurança gerenciados estão associados à instância primária e às instâncias centrais e de tarefa em um cluster. Um grupo de segurança gerenciado adicional para acesso de serviço é necessário quando você cria um cluster em uma sub-rede privada. Para obter mais informações sobre a função de grupos de segurança gerenciados com respeito à configuração de sua rede, consulte Opções da Amazon VPC.

Ao especificar grupos de segurança gerenciados para um cluster, você deve usar o mesmo tipo de grupo de segurança, padrão ou personalizado, para todos os grupos de segurança gerenciados. Por exemplo, você não pode especificar um grupo de segurança personalizado para a instância primária e, em seguida, não especificar um grupo de segurança personalizado para instâncias centrais e de tarefa.

Se você usar grupos de segurança gerenciados padrão, não será necessário especificá-los ao criar um cluster. O Amazon EMR usa os padrões automaticamente. Além disso, se os padrões ainda não existirem na VPC do cluster, o Amazon EMR os criará. O Amazon EMR também os criará se você os especificar explicitamente e eles ainda não existirem.

É possível editar regras em grupos de segurança gerenciados depois que os clusters forem criados. Quando você criar um novo cluster, o Amazon EMR verificará as regras nos grupos de segurança gerenciados que você especificar e criará as regras de entrada ausentes necessárias para o novo cluster, além de regras que podem ter sido adicionadas anteriormente. A menos que esteja definido de forma diferente, cada regra para grupos de segurança padrão gerenciados pelo Amazon EMR também é aplicada aos grupos de segurança personalizados gerenciados pelo Amazon EMR que você especificar.

Os grupos de segurança gerenciados padrão são os seguintes:

- ElasticMapReduzir o primário

  Para regras nesse grupo de segurança, consulte Grupo de segurança gerenciado pelo Amazon EMR para a instância primária (sub-redes públicas).
- ElasticMapReduzir o núcleo

  Para regras nesse grupo de segurança, consulte Grupo de segurança gerenciado pelo Amazon EMR para instâncias centrais e de tarefa (sub-redes públicas).

- ElasticMapReduza o nível primário e o privado

  Para regras nesse grupo de segurança, consulte Grupo de segurança gerenciado pelo Amazon EMR para a instância primária (sub-redes privadas).
- ElasticMapReduza o núcleo privado

  Para regras nesse grupo de segurança, consulte Grupo de segurança gerenciado pelo Amazon EMR para instâncias centrais e de tarefa (sub-redes privadas).
- ElasticMapReduzir- ServiceAccess

  Para regras nesse grupo de segurança, consulte Grupo de segurança gerenciado pelo Amazon EMR para acesso de serviço (sub-redes privadas).

Grupo de segurança gerenciado pelo Amazon EMR para a instância primária (sub-redes públicas)

O grupo de segurança gerenciado padrão para a instância primária em sub-redes públicas tem o nome do ElasticMap grupo Reduce-primary. Tem as regras a seguir. Se você especificar um grupo de segurança gerenciado personalizado, o Amazon EMR adicionará todas as mesmas regras ao grupo de segurança personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada				
Todos ICMPs - IPv4	Todos	N/D	O ID do grupo de segurança gerenciado da instância principal. Em outras palavras, o mesmo grupo de segurança em que a regra é exibida.	Essas regras reflexivas permitem o tráfego de entrada de qualquer instância associada ao grupo de segurança especificado. O uso do `ElasticMapReduce-primary` padrão para vários clusters permite que os nós core e de tarefa desses clusters se comunique m entre si por ICMP ou qualquer porta TCP ou UDP. Especifique grupos de segurança gerenciados personalizados para restringir o acesso entre clusters.
Todos os TCP	TCP	Todos		
Todos os UDP	UDP	Todos		

Tipo	Protoc	Interva de portas	Origem	Detalhes
Todos ICMPs - IPV4	Todos	N/D	O ID de grupo de segurança gerenciado especificado para nós core e de tarefa.	Essas regras permitem todo o tráfego ICMP de entrada e o tráfego por qualquer porta TCP ou UDP de quaisquer instâncias core e de tarefa que estão associadas com o grupo de segurança especificado, mesmo se as instâncias estiverem em clusters diferentes.
Todos os TCP	TCP	Todos		
Todos os UDP	UDP	Todos		
Personali zar	TCP	8443	Vários intervalos de endereços IP da Amazon	Essas regras permitem que o gerenciador de clusters se comunique com o nó primário.

Para conceder acesso SSH a fontes confiáveis ao grupo de segurança primário com o console

Para editar seus grupos de segurança, você deve ter permissão para gerenciar os grupos de segurança para a VPC na qual o cluster está localizado. Para obter mais informações, consulte Alteração de permissões de um usuário e o exemplo de política que permite o gerenciamento de grupos de segurança do EC2 no Guia do usuário do IAM.

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Escolha Clusters. Escolha o ID do cluster que você deseja modificar.
3. No painel Rede e segurança, expanda o menu suspenso Grupos de segurança (firewall) do EC2.
4. Em Nó primário, escolha seu grupo de segurança.
5. Escolha Editar regras de entrada.
6. Verifique se há uma regra de entrada que permita acesso público com as configurações a seguir. Se existir, escolha Excluir para removê-la.

    - Tipo

SSH

- Porta

  22

- Origem

  Personalizado 0.0.0.0/0

> **Warning**
>
> Antes de dezembro de 2020, havia uma regra pré-configurada para permitir o tráfego de entrada na Porta 22 de todas as fontes. Esta regra foi criada para simplificar as conexões SSH iniciais com o nó primário. Recomendamos fortemente remover esta regra de entrada e restringir o tráfego para origens confiáveis.

7. Role até o final da lista de regras e escolha Adicionar regra.
8. Em Type (Tipo), selecione SSH.

   Selecionar SSH insere automaticamente TCP para Protocolo e 22 para Intervalo de portas.
9. Para a origem, selecione Meu IP para adicionar automaticamente seu endereço IP como o endereço de origem. Você também pode adicionar um intervalo personalizado de endereços IP de clientes confiáveis ou criar regras adicionais para outros clientes. Diversos ambientes de rede alocam endereços IP dinamicamente, portanto, pode ser necessário atualizar os endereços IP para clientes confiáveis no futuro.
10. Selecione Save (Salvar).
11. Opcionalmente, escolha o outro grupo de segurança em Nós principais e de tarefas no painel Rede e segurança e repita as etapas acima para permitir que o cliente SSH acesse os nós principais e de tarefas.

Grupo de segurança gerenciado pelo Amazon EMR para instâncias centrais e de tarefa (sub-redes públicas)

O grupo de segurança gerenciado padrão para instâncias principais e de tarefas em sub-redes públicas tem o nome do ElasticMap grupo Reduce-core. O grupo de segurança gerenciado padrão

tem as regras a seguir, e o Amazon EMR adicionará as mesmas regras se você especificar um grupo de segurança gerenciado personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada				
Todos ICMPs - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado para instâncias core e de tarefa. Em outras palavras, o mesmo grupo de segurança em que a regra é exibida.	Essas regras reflexivas permitem o tráfego de entrada de qualquer instância associada ao grupo de segurança especificado. O uso do `ElasticMapReduce-core` padrão para vários clusters permite que as instâncias core e de tarefa desses clusters se comunique m entre si por ICMP ou qualquer porta TCP ou UDP. Especifique grupos de segurança gerenciados personalizados para restringir o acesso entre clusters.
Todos os TCP	TCP	Todos		
Todos os UDP	UDP	Todos		
Todos ICMPs - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado da instância principal.	Essas regras permitem todo o tráfego ICMP de entrada e o tráfego por qualquer porta TCP ou UDP de quaisquer instâncias primárias que estão associadas com o grupo de segurança especificado, mesmo se as instâncias estiverem em clusters diferentes.
Todos os TCP	TCP	Todos		
Todos os UDP	UDP	Todos		

Grupo de segurança gerenciado pelo Amazon EMR para a instância primária (sub-redes privadas)

O grupo de segurança gerenciado padrão para a instância primária em sub-redes privadas tem o nome do grupo Reduce-Primary-Private. ElasticMap O grupo de segurança gerenciado padrão tem as regras a seguir, e o Amazon EMR adicionará as mesmas regras se você especificar um grupo de segurança gerenciado personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada				
Todos ICMPs - IPv4	Todos	N/D	O ID do grupo de segurança gerenciado da instância principal. Em outras palavras, o mesmo grupo de segurança em que a regra é exibida.	Essas regras reflexivas permitem o tráfego de entrada de todas as instâncias associadas com o grupo de segurança especificado e acessíveis a partir da sub-rede privada. O uso do `ElasticMapReduce-Primary-Private` padrão para vários clusters permite que os nós core e de tarefa desses clusters se comuniquem entre si por ICMP ou qualquer porta TCP ou UDP. Especifique grupos de segurança gerenciados personalizados para restringir o acesso entre clusters.
Todos os TCP	TCP	Todos		
Todos os UDP	UDP	Todos		
Todos ICMPs - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado para nós core e de tarefa.	Essas regras permitem todo o tráfego ICMP de entrada e o tráfego por qualquer porta TCP ou UDP de quaisquer instâncias core e de tarefa que estão associadas com o grupo de segurança especificado e acessíveis a partir da sub-rede privada, mesmo se as instâncias estiverem em clusters diferentes.
Todos os TCP	TCP	Todos		
Todos os UDP	UDP	Todos		
HTTPS (8443)	TCP	8443	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Essa regra permite que o gerenciador de clusters se comunique com o nó primário.
Regras de saída				

Tipo	Protoc	Interva de portas	Origem	Detalhes
Todo o tráfego	Tudo	Tudo	0.0.0.0/0	Fornece acesso de saída à Internet.
TCP personalizado	TCP	9443	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Se a regra de saída padrão “Todo o tráfego” acima for removida, essa regra será um requisito mínimo para o Amazon EMR 5.30.0 e versões posteriores. **Note** O Amazon EMR não adiciona a regra quando você usa um grupo de segurança gerenciado personalizado.
TCP personalizado	TCP	80 (http) ou 443 (https)	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Se a regra de saída padrão “Todo o tráfego” acima for removida, essa regra será um requisito mínimo para o Amazon EMR 5.30.0 e versões posteriores para se conectar ao Amazon S3 por https. **Note** O Amazon EMR não adiciona a regra quando você usa um grupo de segurança gerenciado personalizado.

Grupo de segurança gerenciado pelo Amazon EMR para instâncias centrais e de tarefa (sub-redes privadas)

O grupo de segurança gerenciado padrão para instâncias principais e de tarefas em sub-redes privadas tem o nome do grupo Reduce-Core-Private. ElasticMap O grupo de segurança gerenciado

padrão tem as regras a seguir, e o Amazon EMR adicionará as mesmas regras se você especificar um grupo de segurança gerenciado personalizado.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada				
Todos ICMPs - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado para instâncias core e de tarefa. Em outras palavras, o mesmo grupo de segurança em que a regra é exibida.	Essas regras reflexivas permitem o tráfego de entrada de qualquer instância associada ao grupo de segurança especificado. O uso do `ElasticMapReduce-core` padrão para vários clusters permite que as instâncias core e de tarefa desses clusters se comunique m entre si por ICMP ou qualquer porta TCP ou UDP. Especifique grupos de segurança gerenciados personalizados para restringir o acesso entre clusters.
Todos os TCP	TCP	Todos		
Todos os UDP	UDP	Todos		
Todos ICMPs - IPV4	Todos	N/D	O ID do grupo de segurança gerenciado da instância principal.	Essas regras permitem todo o tráfego ICMP de entrada e o tráfego por qualquer porta TCP ou UDP de quaisquer instâncias primárias que estão associadas com o grupo de segurança especificado, mesmo se as instâncias estiverem em clusters diferentes.
Todos os TCP	TCP	Todos		
Todos os UDP	UDP	Todos		
HTTPS (8443)	TCP	8443	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Essa regra permite que o gerenciador de clusters se comunique com os nós core e de tarefa.
Regras de saída				

Tipo	Protoc	Interva de portas	Origem	Detalhes
Todo o tráfego	Tudo	Tudo	0.0.0.0/0	Consulte Editar regras de saída abaixo.
TCP personali zado	TCP	80 (http) ou 443 (https)	O ID do grupo de segurança gerenciado para acesso de serviço em uma sub-rede privada.	Se a regra de saída padrão “Todo o tráfego” acima for removida, essa regra será um requisito mínimo para o Amazon EMR 5.30.0 e versões posteriores para se conectar ao Amazon S3 por https. **Note** O Amazon EMR não adiciona a regra quando você usa um grupo de segurança gerenciado personalizado.

Editar regras de saída

Por padrão, o Amazon EMR cria o grupo de segurança com regras de saída que permitem todo o tráfego de saída em todos os protocolos e portas. A opção de permitir todo o tráfego de saída é selecionada porque várias aplicações do Amazon EMR e do cliente que podem ser executadas em clusters do Amazon EMR podem exigir regras de saída diferentes. O Amazon EMR não consegue prever essas configurações específicas ao criar grupos de segurança padrão. Você pode reduzir o escopo da saída em seus grupos de segurança para incluir somente as regras adequadas a seus casos de uso e políticas de segurança. No mínimo, esse grupo de segurança exige as regras de saída a seguir, mas algumas aplicações podem precisar de saída adicional.

Tipo	Protoc	Interva de portas	Destino	Detalhes
Todos os TCP	TCP	Todos	pl-*xxxxxxxx*	Lista gerenciada de prefixos do Amazon S3 `com.amazonaws.` *MyRegion* `.s3`.

Tipo	Protoc	Interva de portas	Destino	Detalhes
Todo o tráfego	Tudo	Todos	sg-*xxxxxxxxx xxxxxxxx*	O ID do grupo de segurança `ElasticMa pReduce-Core-Private` .
Todo o tráfego	Tudo	Todos	sg-*xxxxxxxxx xxxxxxxx*	O ID do grupo de segurança `ElasticMa pReduce-Primary-Private` .
TCP personali zado	TCP	9443	sg-*xxxxxxxxx xxxxxxxx*	O ID do grupo de segurança `ElasticMa pReduce-ServiceAccess` .

Grupo de segurança gerenciado pelo Amazon EMR para acesso de serviço (sub-redes privadas)

O grupo de segurança gerenciado padrão para acesso ao serviço em sub-redes privadas tem o nome do grupo ElasticMap Reduce -. ServiceAccess Ele tem regras de entrada e regras de saída que permitem o tráfego por HTTPS (porta 8443, porta 9443) para os outros grupos de segurança gerenciados em sub-redes privadas. Essas regras permitem que o gerenciador de clusters se comunique com o nó primário e com os nós centrais e de tarefa. As mesmas regras serão necessárias se você estiver usando grupos de segurança personalizados.

Tipo	Protoc	Interva de portas	Origem	Detalhes
Regras de entrada: necessárias para clusters do Amazon EMR com o Amazon EMR versão 5.30.0 e posteriores.				
TCP personali zado	TCP	9443	O ID do grupo de segurança gerenciado para a instância primária.	Essa regra permite a comunicação entre o grupo de segurança da instância principal e o grupo de segurança de acesso ao serviço.
Regras de saída necessárias para todos os clusters do Amazon EMR				

Tipo	Protoc	Interva de portas	Origem	Detalhes
TCP personali zado	TCP	8443	O ID do grupo de segurança gerenciado para a instância primária.	Essas regras permitem que o gerenciador de clusters se comunique com o nó primário e com os nós centrais e de tarefa.
TCP personali zado	TCP	8443	O ID do grupo de segurança gerenciado para instâncias core e de tarefa.	Essas regras permitem que o gerenciador de clusters se comunique com o nó primário e com os nós centrais e de tarefa.

Trabalhar com grupos de segurança adicionais

Independentemente de você usar os grupos de segurança gerenciados padrão ou especificar grupos de segurança gerenciados personalizados, é possível usar grupos de segurança adicionais. Os grupos de segurança adicionais oferecem a você a flexibilidade para adaptar o acesso entre diferentes clusters e de clientes externos, recursos e aplicativos.

Considere os seguintes cenários como um exemplo. Você tem vários clusters que devem se comunicar uns com os outros, mas deseja permitir acesso SSH de entrada à instância primária apenas para um subconjunto específico de clusters. Para fazer isso, você pode usar o mesmo conjunto de grupos de segurança gerenciados para os clusters. Em seguida, você cria grupos de segurança adicionais que permitem acesso SSH de entrada de clientes confiáveis e especifica grupos de segurança adicionais para a instância primária a cada cluster no subconjunto.

Você pode aplicar até 15 grupos de segurança adicionais para a instância primária, 15 para instâncias principais e de tarefas e 15 para acesso ao serviço (em sub-redes privadas). Se necessário, você pode especificar o mesmo grupo de segurança adicional para instâncias primárias, instâncias centrais e de tarefa e acesso de serviço. O número máximo de grupos de segurança e regras em sua conta está sujeito a limites da conta. Para obter mais informações, consulte os limites de grupos de segurança no Manual do usuário da Amazon VPC.

Especificar grupos de segurança gerenciados pelo Amazon EMR e adicionais

Você pode especificar grupos de segurança usando a AWS Management Console AWS CLI, a ou a API do Amazon EMR. Se você não especificar grupos de segurança, o Amazon EMR criará grupos de segurança padrão. A especificação de grupos de segurança adicionais é opcional. Você pode atribuir grupos de segurança adicionais para instâncias primárias, instâncias centrais e de tarefa e acesso de serviço (somente sub-redes privadas).

New console

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

Especificar grupos de segurança usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Redes, selecione a seta ao lado dos Grupos de segurança do EC2 (firewall) para expandir a seção. Em Nó primário e Nós centrais e de tarefa, os grupos de segurança gerenciados padrão do Amazon EMR são selecionados por padrão. Se você usa uma sub-rede privada, também tem a opção de selecionar um grupo de segurança em Acesso ao serviço.
4. Para alterar o grupo de segurança gerenciado do Amazon EMR, use o menu suspenso Escolher grupos de segurança para selecionar outra opção da lista de opções Grupo de segurança gerenciado pelo Amazon EMR. Você tem um grupo de segurança gerenciado do Amazon EMR para o nó primário e os nós centrais e de tarefa.
5. Para adicionar grupos de segurança personalizados, use o mesmo menu suspenso Escolher grupos de segurança para selecionar até quatro grupos de segurança personalizados na lista de opções Grupo de segurança personalizado. Você pode ter até quatro grupos de segurança personalizados para o nó primário e os nós centrais e de tarefa.

6. Escolha qualquer outra opção que se aplique ao cluster.
7. Para iniciar o cluster, escolha Criar cluster.

Old console

Especificar grupos de segurança usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Escolha as opções para o seu cluster até chegar à Step 4: Security (Etapa 4: Segurança).
4. Escolha EC2 Security Groups (Grupos de segurança do EC2) para expandir a seção.

   Em EMR managed security groups (Grupos de segurança gerenciados pelo EMR), os grupos de segurança gerenciados padrão são selecionados por padrão. Se não existir um padrão na VPC para Master (Principal), Core & Task (Core e tarefa) ou Service Access (Acesso de serviço) (somente sub-rede privada), Create (Criar) aparecerá antes do nome do grupo de segurança associado.
5. Se você usar grupos de segurança gerenciados personalizados, selecione-os nas listas EMR managed security groups (Grupos de segurança gerenciados pelo EMR).

   Se você selecionar um grupo de segurança gerenciado personalizado, uma mensagem solicitará que você selecione um grupo de segurança personalizado para as outras instâncias. Você pode usar apenas grupos de segurança gerenciados padrão ou personalizados para um cluster.
6. Como opção, em Additional security groups (Grupos de segurança adicionais), escolha o ícone de lápis, selecione até quatro grupos de segurança na lista e selecione Assign security groups (Atribuir grupos de segurança). Repita para cada Master (Principal), Core & Task (Core e tarefa) e Service Access (Acesso de serviço) conforme desejado.
7. Selecione Create Cluster (Criar cluster).

Especificar grupos de segurança com a AWS CLI

Para especificar grupos de segurança usando o, AWS CLI você usa o `create-cluster` comando com os seguintes parâmetros da `--ec2-attributes` opção:

Parâmetro	Descrição
`EmrManagedPrimarySecurityGroup`	Use esse parâmetro para especificar um grupo de segurança gerenciado personalizado para a instância primária. Se esse parâmetro for especificado, `EmrManagedCoreSecurityGroup` também deve ser especificado. Para clusters em sub-redes privadas, `ServiceAccessSecurityGroup` também deverá ser especificado.
`EmrManagedCoreSecurityGroup`	Use esse parâmetro para especificar um grupo de segurança gerenciado personalizado para instâncias core e de tarefa. Se esse parâmetro for especificado, `EmrManagedPrimarySecurityGroup` também deve ser especificado. Para clusters em sub-redes privadas, `ServiceAccessSecurityGroup` também deverá ser especificado.
`ServiceAccessSecurityGroup`	Use esse parâmetro para especificar um grupo de segurança gerenciado personalizado para acesso de serviço, o que se aplica apenas a clusters em sub-redes privadas. O grupo de segurança que você especificar como `ServiceAccessSecurityGroup` não deve ser usado para nenhuma outra finalidade e também deve ser reservado ao Amazon EMR. Se esse parâmetro for especificado, `EmrManagedPrimarySecurityGroup` também deve ser especificado.
`AdditionalPrimarySecurityGroups`	

Parâmetro	Descrição
	Use esse parâmetro para especificar até quatro grupos de segurança adicionais para a instância primária.
`AdditionalCoreSecurityGroups`	Use esse parâmetro para especificar até quatro grupos de segurança adicionais para instâncias core e de tarefa.

Example : especifique grupos de segurança gerenciados pelo Amazon EMR e grupos de segurança adicionais

O exemplo a seguir especifica grupos de segurança gerenciados pelo Amazon EMR para um cluster em uma sub-rede privada, vários grupos de segurança adicionais para a instância primária e um único grupo de segurança adicional de instâncias centrais e de tarefa.

> **Note**
>
> Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --name "ClusterCustomManagedAndAdditionalSGs" \
--release-label emr-emr-7.1.0 --applications Name=Hue Name=Hive \
Name=Pig --use-default-roles --ec2-attributes \
SubnetIds=subnet-xxxxxxxxxxxx,KeyName=myKey,\
ServiceAccessSecurityGroup=sg-xxxxxxxxxxxx,\
EmrManagedPrimarySecurityGroup=sg-xxxxxxxxxxxx,\
EmrManagedCoreSecurityGroup=sg-xxxxxxxxxxx,\
AdditionalPrimarySecurityGroups=['sg-xxxxxxxxxxx',\
'sg-xxxxxxxxxxx','sg-xxxxxxxxxx'],\
AdditionalCoreSecurityGroups=sg-xxxxxxxxxxx \
--instance-type m5.xlarge
```

Para obter mais informações, consulte create-cluster na AWS CLI Command Reference.

Especificar grupos de segurança do EC2 para Cadernos do EMR

Quando você cria um Caderno do EMR, dois grupos de segurança são usados para controlar o tráfego de rede entre o Caderno do EMR e o cluster do Amazon EMR quando o editor de caderno é usado. Os grupos de segurança padrão têm o mínimo de regras que permitem somente o tráfego de rede entre o serviço de Cadernos do EMR e os clusters aos quais os cadernos estão anexados.

Um Caderno do EMR usa o Apache Livy para se comunicar com o cluster por meio de um proxy pela porta TCP 18888. Ao criar grupos de segurança personalizados com regras personalizadas para seu ambiente, você pode limitar o tráfego de rede para que apenas um subconjunto de cadernos possa executar código no editor de cadernos em determinados clusters. O cluster usa segurança personalizada, além dos grupos de segurança padrão do cluster. Para obter mais informações, consulte Control network traffic with security groups no Guia de gerenciamento do Amazon EMR e no Especificar grupos de segurança do EC2 para Cadernos do EMR.

Grupo de segurança padrão do EC2 para a instância primária

O grupo de segurança padrão do EC2 para a instância primária está associado à instância primária do cluster, além dos grupos de segurança para a instância primária.

Nome do grupo: ElasticMapReduceEditors-Livy

Regras

- Entrada

  Permitir a porta TCP 18888 de todos os recursos no grupo de segurança padrão do EC2 para Cadernos do EMR
- Saída

  Nenhum

Grupo de segurança padrão do EC2 para Cadernos do EMR

O grupo de segurança padrão do EC2 para o Caderno do EMR está associado ao editor de cadernos para qualquer Caderno do EMR ao qual ele esteja atribuído.

Nome do grupo: ElasticMapReduceEditors-Editor

Regras

- Entrada

  Nenhum

- Saída

  Permitir a porta TCP 18888 a todos os recursos no grupo de segurança padrão do EC2 para Cadernos do EMR.

Grupo de segurança personalizado do EC2 para o Cadernos do EMR ao associar cadernos a repositórios do Git

Para vincular um repositório do Git ao caderno, o grupo de segurança do Caderno do EMR deve incluir uma regra de saída para permitir que o caderno encaminhe o tráfego para a Internet. É recomendável criar um grupo de segurança para essa finalidade. A atualização do grupo de segurança padrão ElasticMapReduceEditors-Editor pode fornecer as mesmas regras de saída para outros notebooks anexados a esse grupo de segurança.

Regras

- Entrada

  Nenhum

- Saída

  Permita que o caderno encaminhe o tráfego para a Internet por meio do cluster, como demonstra o exemplo a seguir. Utiliza-se o valor 0.0.0.0/0 para fins de exemplo. É possível modificar essa regra para especificar os endereços IP dos repositórios baseados em Git.

Tipo	Protocolo	Intervalo de portas	Destino
Regra personalizada de TCP	TCP	18888	SG-
HTTPS	TCP	443	0.0.0.0/0

Usar o bloqueio de acesso público do Amazon EMR

O bloqueio de acesso público (BPA) do Amazon EMR impede que você inicie um cluster em uma sub-rede pública se o cluster tiver uma configuração de segurança que permita tráfego de entrada de endereços IP públicos em uma porta.

> **Important**
>
> O bloqueio de acesso público é habilitado por padrão. Para aumentar a proteção da conta, é recomendável mantê-la habilitada.

Noções básicas do bloqueio ao acesso público

É possível usar a configuração em nível de conta de bloqueio de acesso público para gerenciar o acesso à rede pública aos clusters do Amazon EMR de maneira centralizada.

Quando um usuário do seu Conta da AWS executa um cluster, o Amazon EMR verifica as regras de porta no grupo de segurança do cluster e as compara com suas regras de tráfego de entrada. Se o grupo de segurança tiver uma regra de entrada que abra portas para os endereços IP públicos IPv4 0.0.0.0/0 ou IPv6 ::/0, e essas portas não forem especificadas como exceções para a conta, o Amazon EMR não permitirá que o usuário crie o cluster.

Se um usuário modificar as regras do grupo de segurança de um cluster em execução em uma sub-rede pública para ter uma regra de acesso público que viole a configuração do BPA da conta, o Amazon EMR revogará a nova regra se tiver permissão para isso. Se o Amazon EMR não tiver permissão para revogar a regra, ele criará um evento no painel AWS Health que descreva a violação. Para conceder a permissão de revogação da regra ao Amazon EMR, consulte Configurar o Amazon EMR para revogar regras do grupo de segurança.

O bloqueio de acesso público é habilitado por padrão para todos os clusters em cada Região da AWS de sua Conta da AWS. O BPA se aplica a todo o ciclo de vida de um cluster, mas não se aplica aos clusters criados em sub-redes privadas. É possível configurar exceções à regra do BPA; a porta 22 é uma exceção por padrão. Para obter mais informações sobre como configurar exceções, consulte Configurar o bloqueio de acesso público.

Configurar o bloqueio de acesso público

Você pode atualizar os grupos de segurança e a configuração de bloqueio de acesso público de suas contas a qualquer momento.

Você pode ativar e desativar as configurações de bloqueio de acesso público (BPA) com a API AWS Management Console, a AWS Command Line Interface (AWS CLI) e a API do Amazon EMR. As configurações se aplicam à sua conta com base na Região. Para manter a segurança do cluster, é recomendável usar o BPA.

New console

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

Configurar o bloqueio de acesso público usando o novo console

1. Faça login no e AWS Management Console, em seguida, abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Na barra de navegação superior, selecione a região que você deseja configurar, se ainda não estiver selecionada.
3. Em EMR no EC2, no painel de navegação esquerdo, escolha Bloqueio de acesso público.
4. Em Block public access settings (Configurações de bloqueio de acesso público), conclua as etapas a seguir.

Para...	Fazer isso...
Ativar ou desativar o bloqueio de acesso público	Escolha Editar, escolha Ativar ou Desativar, conforme o caso, e escolha Salvar.
Editar portas na lista de exceções	1. Escolha Editar e encontre a seção Exceções do intervalo de portas. 2. Para adicionar portas à lista de exceções, escolha Add a port range (Adicionar um intervalo de portas) e

Para...	Fazer isso...
	insira uma nova porta ou um intervalo de portas. Repita para cada porta ou intervalo de portas a ser adicionado. 3. Para remover uma porta ou um intervalo de portas, escolha Remover ao lado da entrada na lista de intervalos de portas. 4. Selecione Save (Salvar).

Old console

Visualizar a configuração do bloqueio de acesso público usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Na barra de navegação, verifique se a região que você deseja configurar está selecionada.
3. Escolha Block public access (Bloqueio de acesso público).
4. Em Block public access settings (Configurações de bloqueio de acesso público), conclua as etapas a seguir.

Para...	Fazer isso...
Ativar ou desativar o bloqueio de acesso público	Escolha Change (Alterar), selecione On (Ativado) ou Off (Desativado), conforme apropriado, e escolha a marca de seleção para confirmar.
Editar portas na lista de exceções	1. Em Exceptions (Exceções), escolha Edit (Editar). 2.

Para...	Fazer isso...
	Para adicionar portas à lista de exceções, escolha Add a port range (Adicionar um intervalo de portas) e insira uma nova porta ou um intervalo de portas. Repita para cada porta ou intervalo de portas a ser adicionado. 3. Para remover uma porta ou um intervalo de portas, escolha x ao lado da entrada na lista Port range (Intervalos de portas). 4. Escolha Salvar alterações.

AWS CLI

Para configurar, bloquear o acesso público usando o AWS CLI

- Use o comando `aws emr put-block-public-access-configuration` para configurar o bloqueio de acesso público, conforme mostrado nos exemplos a seguir.

Para...	Fazer isso...
Ativar o bloqueio de acesso público	Defina `BlockPublicSecurityGroupRules` como `true`, conforme mostrado no exemplo a seguir. Para que o cluster seja iniciado, nenhum grupo de segurança associado a um cluster pode ter uma regra de entrada que permita acesso público. `aws emr put-block-public-access-configuration --block-public-access-configuration BlockPublicSecurityGroupRules=true`
Desativar o bloqueio de acesso público	Defina `BlockPublicSecurityGroupRules` como `false`, conforme mostrado no exemplo a seguir. Os grupos de segurança associados a um cluster podem ter regras de entrada que permitam o acesso público em qualquer porta. Não recomendamos essa configuração. `aws emr put-block-public-access-configuration --block-public-access-configuration BlockPublicSecurityGroupRules=false`

Para...	Fazer isso...
Ativar o bloqueio de acesso público e especificar portas como exceções	O exemplo a seguir ativa o bloqueio de acesso público e especifica a porta 22 e as portas 100-101 como exceções. Isso permite que os clusters sejam criados se um grupo de segurança associado tiver uma regra de entrada que permita o acesso público na porta 22, na porta 100 ou na porta 101. `aws emr put-block-public-access-configuration --block-public-access-configuration '{ "BlockPublicSecurityGroupRules": true, "PermittedPublicSecurityGroupRuleRanges": [ { "MinRange": 22, "MaxRange": 22 }, { "MinRange": 100, "MaxRange": 101 } ] }'`

Configurar o Amazon EMR para revogar regras do grupo de segurança

O Amazon EMR precisa de permissão para revogar regras do grupo de segurança e cumprir sua configuração de bloqueio de acesso público. Você pode usar uma destas abordagens para dar a permissão necessária ao Amazon EMR:

- (Recomendado) Anexe a política gerenciada `AmazonEMRServicePolicy_v2` ao perfil de serviço. Para ter mais informações, consulte Perfil de serviço para Amazon EMR (perfil do EMR).
- Crie uma nova política em linha que permita a ação `ec2:RevokeSecurityGroupIngress` em grupos de segurança. Para obter mais informações sobre como modificar uma política de permissões de perfil, consulte Modificar uma política de permissões de perfil com o console do IAM, a API da AWS e a AWS CLI no Guia do usuário do IAM.

Resolver violações ao bloqueio de acesso público

Se ocorrer uma violação ao bloqueio de acesso público, você poderá mitigá-la com uma destas ações:

- Se quiser acessar uma interface da Web no cluster, use uma das opções descritas em Visualizar interfaces Web hospedadas em clusters do Amazon EMR para acessar a interface por meio de SSH (porta 22).
- Para permitir o tráfego no cluster com base em endereços IP específicos em vez do endereço IP público, adicione uma regra de grupo de segurança. Para obter mais informações, consulte Adicionar regras a um grupo de segurança, no Guia de conceitos básicos do Amazon EC2.
- (Não recomendado) É possível configurar as exceções de BPA do Amazon EMR para incluir a porta ou o intervalo de portas desejado. Ao especificar uma exceção de BPA, você introduz riscos com uma porta desprotegida. Se você pretende especificar uma exceção, remova a exceção assim que ela não for mais necessária. Para ter mais informações, consulte Configurar o bloqueio de acesso público.

Identificar clusters associados às regras do grupo de segurança

Talvez seja necessário identificar todos os clusters associados a determinada regra de grupo de segurança ou encontrar a regra de grupo de segurança de determinado cluster.

- Se você conhece o grupo de segurança, poderá identificar os clusters associados se encontrar as interfaces de rede do grupo de segurança. Para obter mais informações, consulte How can I find the resources associated with an Amazon EC2 security group? no AWS re:Post. As instâncias do Amazon EC2 que estão conectadas a essas interfaces de rede serão marcadas com o ID do cluster ao qual pertencem.
- Se você quiser encontrar os grupos de segurança de um cluster conhecido, siga as etapas descritas em Visualizar o status e os detalhes do cluster. Você pode encontrar os grupos de segurança do cluster no painel Rede e segurança no console ou no campo `Ec2InstanceAttributes` da AWS CLI.

Validação de conformidade para o Amazon EMR

Auditores terceirizados avaliam a segurança e a conformidade do Amazon EMR como parte de AWS vários programas de conformidade. Isso inclui SOC, PCI, FedRAMP, HIPAA e outros.

Para obter uma lista de AWS serviços no escopo de programas de conformidade específicos, consulte AWS serviços no escopo por programa de conformidade. Para obter informações gerais, consulte Programas de conformidade da AWS.

Você pode baixar relatórios de auditoria de terceiros usando AWS Artifact. Para obter mais informações, consulte Baixar relatórios em AWS Artifact.

Sua responsabilidade de conformidade ao usar o Amazon EMR é determinada pela confidencialidade dos dados, pelos objetivos de conformidade da empresa e pelos regulamentos e leis aplicáveis. Se seu uso do Amazon EMR estiver sujeito à conformidade com padrões como HIPAA, PCI ou FedRAMP, fornece recursos para ajudar a: AWS

- Guias de início rápido sobre segurança e conformidade — Esses guias de implantação discutem considerações arquitetônicas e fornecem etapas para a implantação de ambientes básicos focados em segurança e conformidade em. AWS
- Documento técnico sobre arquitetura para segurança e conformidade com a HIPAA — Este whitepaper descreve como as empresas podem usar para criar aplicativos compatíveis com a HIPAA. AWS
- AWS recursos de conformidade — essa coleção de pastas de trabalho e guias pode ser aplicada ao seu setor e local.
- AWS Config— Esse AWS serviço avalia se suas configurações de recursos estão em conformidade com as práticas internas, as diretrizes e os regulamentos do setor.
- AWS Security Hub— Esse AWS serviço fornece uma visão abrangente do seu estado de segurança interno, AWS que ajuda você a verificar sua conformidade com os padrões e as melhores práticas do setor de segurança.

Resiliência no Amazon EMR

A infraestrutura AWS global é construída em torno de AWS regiões e zonas de disponibilidade. AWS As regiões fornecem várias zonas de disponibilidade fisicamente separadas e isoladas, conectadas a redes de baixa latência, alta taxa de transferência e alta redundância. Com as zonas de disponibilidade, você pode projetar e operar aplicativos e bancos de dados que executam o failover automaticamente entre as zonas de disponibilidade sem interrupção. As Zonas de Disponibilidade são mais altamente disponíveis, tolerantes a falhas e escaláveis que uma ou várias infraestruturas de datacenter tradicionais.

Para obter mais informações sobre AWS regiões e zonas de disponibilidade, consulte infraestrutura AWS global.

Além da infraestrutura AWS global, o Amazon EMR oferece vários recursos para ajudar a suportar suas necessidades de resiliência e backup de dados.

- Integração com o Amazon S3 por meio do EMRFS
- Suporte a vários nós principais

Segurança da infraestrutura no Amazon EMR

Como um serviço gerenciado, o Amazon EMR é protegido pela segurança de rede AWS global. Para obter informações sobre serviços AWS de segurança e como AWS proteger a infraestrutura, consulte AWS Cloud Security. Para projetar seu AWS ambiente usando as melhores práticas de segurança de infraestrutura, consulte Proteção de infraestrutura no Security Pillar AWS Well-Architected Framework.

Você usa chamadas de API AWS publicadas para acessar o Amazon EMR pela rede. Os clientes devem ser compatíveis com:

- Transport Layer Security (TLS). Exigimos TLS 1.2 e recomendamos TLS 1.3.
- Conjuntos de criptografia com Perfect Forward Secrecy (PFS) como DHE (Ephemeral Diffie-Hellman) ou ECDHE (Ephemeral Elliptic Curve Diffie-Hellman). A maioria dos sistemas modernos, como Java 7 e versões posteriores, suporta esses modos.

Além disso, as solicitações devem ser assinadas utilizando um ID da chave de acesso e uma chave de acesso secreta associada a uma entidade principal do IAM. Ou é possível usar o AWS Security Token Service (AWS STS) para gerar credenciais de segurança temporárias para assinar solicitações.

Tópicos

- Conectar-se ao Amazon EMR usando um endpoint da VPC de interface

Conectar-se ao Amazon EMR usando um endpoint da VPC de interface

Você pode se conectar diretamente ao Amazon EMR usando uma interface VPC endpoint (AWS PrivateLink) na sua Virtual Private Cloud (VPC) em vez de se conectar pela Internet. Quando você

usa uma interface VPC endpoint, a comunicação entre sua VPC e o Amazon EMR é conduzida inteiramente dentro da rede. AWS Cada endpoint da VPC é representado por uma ou mais interfaces de rede elástica (ENIs) com endereços IP privados nas sub-redes da VPC.

A interface VPC endpoint conecta sua VPC diretamente ao Amazon EMR sem um gateway de internet, dispositivo NAT, conexão VPN ou conexão. AWS Direct Connect As instâncias em sua VPC não precisam de endereços IP públicos para se comunicarem com a API do Amazon EMR.

Para usar o Amazon EMR por meio da VPC, você deve se conectar de uma instância que esteja dentro da VPC ou conectar sua rede privada à VPC usando a rede privada virtual (VPN) da Amazon ou o AWS Direct Connect. Para obter informações sobre o Amazon VPN, consulte Conexões VPN no Guia do usuário do Amazon Virtual Private Cloud. Para obter informações sobre AWS Direct Connect, consulte Criação de uma conexão no Guia AWS Direct Connect do usuário.

Você pode criar uma interface VPC endpoint para se conectar ao Amazon EMR usando o AWS console ou os comandos (). AWS Command Line Interface AWS CLI Para obter mais informações, consulte Creating an interface endpoint (Criação de um endpoint de interface).

Após criar um endpoint da VPC de interface, se você habilitar nomes de host DNS privados para o endpoint, o endpoint padrão do Amazon EMR será resolvido para seu endpoint da VPC. O endpoint de nome de serviço padrão para o Amazon EMR estará no formato a seguir.

```
elasticmapreduce.Region.amazonaws.com
```

Se você não habilitar nomes de host DNS privados, a Amazon VPC fornecerá um nome de endpoint DNS que poderá ser usado no formato a seguir.

```
VPC_Endpoint_ID.elasticmapreduce.Region.vpce.amazonaws.com
```

Para obter mais informações, consulte Interface VPC endpoints (AWS PrivateLink) no Guia do usuário da Amazon VPC.

O Amazon EMR oferece suporte a chamadas para todas as ações de API dentro da VPC.

Você pode anexar políticas de endpoint da VPC a um endpoint da VPC para controlar o acesso de entidades principais do IAM. Também é possível associar grupos de segurança a um VPC endpoint para controlar o acesso de entrada e saída com base na origem e no destino do tráfego de rede, como um intervalo de endereços IP. Para obter mais informações, consulte Controlling access to services with VPC endpoints.

Criar uma política de endpoint da VPC para o Amazon EMR

É possível criar uma política para endpoints da Amazon VPC para o Amazon EMR para especificar o seguinte:

- O principal que pode ou não executar ações
- As ações que podem ser executadas
- Os recursos nos quais as ações podem ser executadas

Para mais informações, consulte Controlar o acesso a serviços com VPC endpoints no Guia do usuário da Amazon VPC.

Example — Política de VPC endpoint para negar todo o acesso de uma conta especificada AWS

A política de VPC endpoint a seguir nega à AWS conta *123456789012* todo o acesso aos recursos usando o endpoint.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
            "Action": "*",
            "Effect": "Deny",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "123456789012"
                ]
            }
        }
    ]
}
```

Example – Política de endpoint da VPC para permitir o acesso à VPC somente a uma entidade principal do IAM (usuário) especificada

A política de VPC endpoint a seguir permite acesso total somente ao usuário lijuan na conta 123456789012. AWS Todos os outros principais IAM têm acesso negado usando o endpoint.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "arn:aws:iam::123456789012:user/lijuan"
                ]
            }
        }]
}
```

Example – Política de endpoint da VPC para permitir operações somente leitura do EMR

A política de VPC endpoint a seguir permite que somente a AWS conta *123456789012* execute as ações especificadas do Amazon EMR.

As ações especificadas fornecem o equivalente ao acesso somente leitura para o Amazon EMR. Todas as outras ações na VPC serão negadas para a conta especificada. Todas as outras contas terão acesso negado. Para obter uma lista de ações do Amazon EMR, consulte Ações, recursos e chaves de condição do Amazon EMR.

```
{
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:DescribeCluster",
```

```
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "123456789012"
                ]
            }
        }
    ]
}
```

Example – Política de endpoint da VPC negando acesso a um cluster especificado

A política de VPC endpoint a seguir permite acesso total a todas as contas e diretores, mas nega qualquer acesso da AWS conta 123456789012 às ações realizadas no cluster do Amazon EMR com o ID de cluster j-a1b2cd34ef5g. Outras ações do Amazon EMR que não oferecem suporte a permissões de nível de recurso para clusters ainda são permitidas. Para obter uma lista de ações do Amazon EMR e seus tipos de recurso correspondentes, consulte Ações, recursos e chaves de condição do Amazon EMR.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
            "Action": "*",
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-west-2:123456789012:cluster/j-
A1B2CD34EF5G",
            "Principal": {
                "AWS": [
```

```
                    "123456789012"
                ]
            }
        }
    ]
}
```

Gerenciar clusters

Após iniciar seu cluster, você pode monitorá-lo e gerenciá-lo. O Amazon EMR fornece várias ferramentas que você pode usar para se conectar ao cluster e controlá-lo.

Tópicos

Conectar-se a um cluster

Ao executar um cluster do Amazon EMR, muitas vezes, tudo o que você precisa fazer é executar uma aplicação para analisar seus dados e depois coletar a saída de um bucket do Amazon S3. Às vezes, você pode querer interagir com o nó primário enquanto o cluster está em execução. Por exemplo, talvez você queira se conectar ao nó primário para executar consultas interativas, verificar arquivos de log, depurar um problema com o cluster, monitorar a performance usando uma aplicação como o Ganglia, que é executada no nó primário e assim por diante. As seções a seguir descrevem técnicas que você pode usar para conectar-se ao nó primário.

Em um cluster do EMR, o nó primário é uma instância do Amazon EC2 que coordena as instâncias do EC2 em execução como nós centrais e de tarefa. O nó primário expõe um nome DNS público que você pode usar para conectar-se a ele. Por padrão, o Amazon EMR cria regras de grupo de segurança para o nó primário e para os nós centrais e de tarefa, que determinam como você acessa esses nós.

> **Note**
>
> Você pode conectar-se ao nó primário somente enquanto o cluster está em execução. Quando o cluster for encerrado, a instância do EC2 atuando como o nó primário será terminada e não estará mais disponível. Para se conectar ao nó primário, você também deve

se autenticar para o cluster. Você pode usar o Kerberos para autenticação ou especificar uma chave privada do par de chaves do Amazon EC2 ao iniciar o cluster. Para obter mais informações sobre como configurar o Kerberos e se conectar, consulte Usar o Kerberos para autenticação com o Amazon EMR. Quando executar um cluster no console, a chave privada do par de chaves do Amazon EC2 será especificada na seção Segurança e acesso da página Criar cluster.

Por padrão, o grupo de segurança ElasticMapReduce -master não permite acesso SSH de entrada. Talvez seja necessário adicionar uma regra de entrada que permita acesso SSH (porta TCP 22) a partir das origens às quais você deseja ter acesso. Para obter mais informações sobre a modificação das regras do grupo de segurança, consulte Adicionar regras a um grupo de segurança no Guia do usuário do Amazon EC2.

Important

Não modifique as regras restantes no grupo de segurança ElasticMapReduce -master. Modificar essas regras pode interferir com o funcionamento do cluster.

Tópicos

- Antes de se conectar: autorize o tráfego de entrada
- Conectar-se ao nó primário usando SSH

Antes de se conectar: autorize o tráfego de entrada

Antes de se conectar a um cluster do Amazon EMR, é necessário autorizar o tráfego SSH de entrada (porta 22) de clientes confiáveis, como o endereço IP do computador. Para isso, edite as regras do grupo de segurança gerenciado para os nós aos quais deseja se conectar. Por exemplo, as instruções a seguir mostram como adicionar uma regra de entrada para acesso SSH ao grupo de segurança ElasticMapReduce -master padrão.

Para ter mais informações sobre usar grupos de segurança com o Amazon EMR, consulte Controle do tráfego de rede com grupos de segurança.

New console

Conceder o acesso SSH a fontes confiáveis ao grupo de segurança primário com o novo console

Para editar seus grupos de segurança, você deve ter permissão para gerenciar os grupos de segurança para a VPC na qual o cluster está localizado. Para obter mais informações, consulte Alteração de permissões de um usuário e o exemplo de política que permite o gerenciamento de grupos de segurança do EC2 no Guia do usuário do IAM.

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha o cluster que você deseja atualizar. Isso abre a página de detalhes do cluster. A guia Propriedades da página será pré-selecionada.
3. Em Redes na guia Propriedades, selecione a seta ao lado de Grupos de segurança do EC2 (firewall) para expandir esta seção. Em Nó primário, selecione o link do grupo de segurança. Isso abre o console do EC2.
4. Escolha a guia Regras de entrada e escolha Editar regras.
5. Verifique se há uma regra de entrada que permita acesso público com as configurações a seguir. Se existir, escolha Excluir para removê-la.

   - Tipo

     SSH
   - Porta

     22
   - Origem

     Personalizado 0.0.0.0/0

> **Warning**
>
> Antes de dezembro de 2020, o grupo de segurança ElasticMapReduce -master tinha uma regra pré-configurada para permitir tráfego de entrada na Porta 22 de todas as fontes. Esta regra foi criada para simplificar as conexões SSH iniciais com o nó

primário. Recomendamos fortemente remover esta regra de entrada e restringir o tráfego para origens confiáveis.

6. Role até o final da lista de regras e escolha Adicionar regra.
7. Em Type (Tipo), selecione SSH. Essa seleção SSH insere automaticamente TCP para Protocolo e 22 para Intervalo de portas.
8. Para a origem, selecione Meu IP para adicionar automaticamente seu endereço IP como o endereço de origem. Você também pode adicionar um intervalo personalizado de endereços IP de clientes confiáveis ou criar regras adicionais para outros clientes. Diversos ambientes de rede alocam endereços IP dinamicamente, portanto, pode ser necessário atualizar os endereços IP para clientes confiáveis no futuro.
9. Selecione Save (Salvar).
10. Opcionalmente, retorne à Etapa 3, escolha os Nós centrais e de tarefa e repita as Etapas 4 a 8. Isso concede aos nós centrais e de tarefa acesso ao cliente SSH.

Old console

Para conceder acesso SSH a fontes confiáveis ao grupo de segurança primário com o console

Para editar seus grupos de segurança, você deve ter permissão para gerenciar os grupos de segurança para a VPC na qual o cluster está localizado. Para obter mais informações, consulte Alteração de permissões de um usuário e o exemplo de política que permite o gerenciamento de grupos de segurança do EC2 no Guia do usuário do IAM.

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Escolha Clusters. Escolha o ID do cluster que você deseja modificar.
3. No painel Rede e segurança, expanda o menu suspenso Grupos de segurança (firewall) do EC2.
4. Em Nó primário, escolha seu grupo de segurança.
5. Escolha Editar regras de entrada.
6. Verifique se há uma regra de entrada que permita acesso público com as configurações a seguir. Se existir, escolha Excluir para removê-la.

   - Tipo

SSH

- Porta

  22

- Origem

  Personalizado 0.0.0.0/0

> **Warning**
>
> Antes de dezembro de 2020, havia uma regra pré-configurada para permitir o tráfego de entrada na Porta 22 de todas as fontes. Esta regra foi criada para simplificar as conexões SSH iniciais com o nó primário. Recomendamos fortemente remover esta regra de entrada e restringir o tráfego para origens confiáveis.

7. Role até o final da lista de regras e escolha Adicionar regra.
8. Em Type (Tipo), selecione SSH.

   Selecionar SSH insere automaticamente TCP para Protocolo e 22 para Intervalo de portas.
9. Para a origem, selecione Meu IP para adicionar automaticamente seu endereço IP como o endereço de origem. Você também pode adicionar um intervalo personalizado de endereços IP de clientes confiáveis ou criar regras adicionais para outros clientes. Diversos ambientes de rede alocam endereços IP dinamicamente, portanto, pode ser necessário atualizar os endereços IP para clientes confiáveis no futuro.
10. Selecione Save (Salvar).
11. Opcionalmente, escolha o outro grupo de segurança em Nós principais e de tarefas no painel Rede e segurança e repita as etapas acima para permitir que o cliente SSH acesse os nós principais e de tarefas.

Conectar-se ao nó primário usando SSH

O Secure Shell (SSH) é um protocolo de rede que você pode usar para criar uma conexão segura com um computador remoto. Depois de estabelecer uma conexão, o terminal no computador local se comporta como se estivesse em execução no computador remoto. Os comandos que você emitir

localmente serão executados no computador remoto, e a saída do comando do computador remoto será exibida na janela do terminal.

Ao usar SSH com AWS, você está se conectando a uma instância do EC2, que é um servidor virtual executado na nuvem. Ao trabalhar com o Amazon EMR, o uso mais comum do SSH é para conexão com a instância do EC2 que está atuando como o nó primário do cluster.

O uso do SSH para conectar-se ao nó primário permite monitorar o cluster e interagir com ele. Você pode emitir comandos do Linux no nó primário, executar aplicações como o Hive e o Pig interativamente, pesquisar diretórios, ler arquivos de log e assim por diante. Também pode criar um túnel na sua conexão SSH para visualizar as interfaces Web hospedadas no nó primário. Para ter mais informações, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

Para se conectar ao nó primário usando o SSH, você precisa do nome DNS público do nó primário. Além disso, o grupo de segurança associado ao nó primário deve ter uma regra de entrada que permite o tráfego SSH (porta TCP 22) de uma fonte que inclui o cliente onde a conexão SSH se origina. Talvez seja necessário adicionar uma regra para permitir uma conexão SSH do seu cliente. Para obter mais informações sobre a modificação das regras do grupo de segurança, consulte Controle do tráfego de rede com grupos de segurança Adicionar regras a um grupo de segurança no Guia do usuário do Amazon EC2.

Recuperar o nome DNS público do nó primário

Você pode recuperar o nome DNS público primário usando o console do Amazon EMR e a AWS CLI.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Recuperar o nome DNS público do nó primário usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e, em seguida, selecione o cluster em que deseja recuperar o nome DNS público.

3. Observe o valor do DNS público do nó primário na seção Resumo da página de detalhes do cluster.

Old console

Recuperar o nome DNS público do nó primário usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Na página Cluster List (Lista de clusters), selecione o link para o seu cluster.
3. Observe o valor de DNS público principal que é exibido na seção Resumo da página Detalhes do cluster.

> **Note**
>
> Você também pode escolher o link SSH para obter instruções sobre como criar uma conexão SSH com o nó primário.

CLI

Para recuperar o nome DNS público do nó primário com o AWS CLI

1. Para recuperar o identificador do cluster, digite o seguinte comando.

```
aws emr list-clusters
```

A saída lista seus clusters, incluindo os IDs dos clusters. Observe o ID do cluster ao qual você está se conectando.

```
"Status": {
    "Timeline": {
        "ReadyDateTime": 1408040782.374,
        "CreationDateTime": 1408040501.213
    },
    "State": "WAITING",
    "StateChangeReason": {
        "Message": "Waiting after step completed"
```

```
    }
},
"NormalizedInstanceHours": 4,
"Id": "j-2AL4XXXXXX5T9",
"Name": "My cluster"
```

2. Para listar as instâncias de cluster, incluindo o nome DNS público do cluster, digite um dos comandos a seguir. Substitua *j-2AL4XXXXXX5T9* pelo ID do cluster retornado pelo comando anterior.

```
aws emr list-instances --cluster-id j-2AL4XXXXXX5T9
```

Ou:

```
aws emr describe-cluster --cluster-id j-2AL4XXXXXX5T9
```

A saída lista as instâncias de cluster, incluindo nomes DNS e endereços IP. Observe o valor para `PublicDnsName`.

```
"Status": {
    "Timeline": {
        "ReadyDateTime": 1408040779.263,
        "CreationDateTime": 1408040515.535
    },
    "State": "RUNNING",
    "StateChangeReason": {}
},
"Ec2InstanceId": "i-e89b45e7",
"PublicDnsName": "ec2-###-##-##-###.us-west-2.compute.amazonaws.com"

"PrivateDnsName": "ip-###-##-##-###.us-west-2.compute.internal",
"PublicIpAddress": "##.###.###.##",
"Id": "ci-12XXXXXXXXFMH",
"PrivateIpAddress": "###.##.#.###"
```

Para obter mais informações, consulte os comandos do Amazon EMR na AWS CLI.

Conectar-se ao nó primário usando SSH e uma chave privada do Amazon EC2 no Linux, Unix e Mac OS X

Para criar uma conexão SSH autenticada com um arquivo de chave privada, você precisa especificar a chave privada do par de chaves do Amazon EC2 ao iniciar um cluster. Para obter mais informações sobre como acessar seu par de chaves, consulte os pares de chaves do Amazon EC2 no Guia do usuário do Amazon EC2.

Seu computador Linux muito provavelmente inclui um cliente SSH por padrão. Por exemplo, a OpenSSH está instalada na maioria dos sistemas operacionais Unix, Linux e MacOS X. É possível verificar se existe um cliente SSH digitando ssh na linha de comando. Se o computador não reconhecer o comando, instale um cliente SSH para se conectar ao nó primário. O projeto OpenSSH fornece uma implementação grátis do pacote completo de ferramentas SSH. Para obter mais informações, consulte o site do OpenSSH.

As instruções a seguir demonstram como abrir uma conexão SSH com o nó primário do Amazon EMR no Linux, Unix e Mac OS X.

Para configurar as permissões do arquivo de chave privada do par de chaves

Antes de usar sua chave privada do par de chaves do Amazon EC2 para criar uma conexão SSH, você deve definir permissões no arquivo `.pem` para que apenas o proprietário da chave tenha permissão para acessar o arquivo. Isso é necessário para criar uma conexão SSH usando o terminal ou o. AWS CLI

1. Verifique se você permitiu tráfego SSH de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Localize seu arquivo `.pem`. Estas instruções pressupõem que o arquivo se chame `mykeypair.pem` e esteja armazenado no diretório inicial do usuário atual.
3. Digite o seguinte comando para definir as permissões. Substitua *`~/mykeypair.pem`* pelo caminho completo e nome do arquivo de chave privada de par de chaves. Por exemplo, `C:/Users/<username>/.ssh/mykeypair.pem`.

   ```
   chmod 400 ~/mykeypair.pem
   ```

   Se você não definir permissões no arquivo `.pem`, receberá um erro indicando que o arquivo de chave está desprotegido e que a chave será rejeitada. Para conectar, você só precisa definir permissões no arquivo de chave privada do par de chaves ao usá-lo pela primeira vez.

Conectar-se ao nó primário usando o terminal

1. Abra uma janela do terminal. No Mac OS X, escolha Applications > Utilities > Terminal (Aplicativos > Utilitários > Terminal). Em outras distribuições do Linux, o terminal está normalmente localizado em Applications > Accessories > Terminal (Aplicativos > Acessórios > Terminal).
2. Para estabelecer uma conexão com o nó primário, digite o comando a seguir. Substitua *ec2-###-##-##-###.compute-1.amazonaws.com* pelo nome DNS público primário de seu cluster e substitua *~/mykeypair.pem* pela caminho completo e nome de arquivo do arquivo `.pem`. Por exemplo, `C:/Users/<username>/.ssh/mykeypair.pem`.

   ```
   ssh hadoop@ec2-###-##-##-###.compute-1.amazonaws.com -i ~/mykeypair.pem
   ```

   > **Important**
   >
   > Você deve usar o nome de logon `hadoop` ao se conectar ao nó primário do Amazon EMR; caso contrário, poderá ver um erro semelhante a `Server refused our key`.

3. Um aviso afirma que não foi possível verificar a autenticidade do host ao qual você está se conectando. Digite `yes` para continuar.
4. Quando você encerrar de trabalhar no nó primário, digite o seguinte comando para encerrar a conexão SSH.

   ```
   exit
   ```

Caso tenha dificuldades ao usar o SSH para se conectar ao nó primário, consulte Solução de problemas para conectar-se à sua instância.

Conectar-se ao nó primário usando SSH no Windows

Os usuários do Windows podem usar um cliente SSH, como o PuTTY para se conectarem ao nó primário. Antes de se conectar ao nó primário do Amazon EMR, você deve baixar e instalar PuTTY e PuTTYgen. Você pode baixar essas ferramentas na página de download do PuTTY.

O PuTTY não oferece suporte nativamente ao formato de arquivo de chave privada com par de chaves (`.pem`) gerado pelo Amazon EC2. Você usa o PuTTY para converter seu arquivo de chaves

no formato PuTTY necessário (`.ppk`). É necessário converter a chave nesse formato (`.ppk`) antes de tentar se conectar ao nó primário usando o PuTTY.

Para obter mais informações sobre a conversão da sua chave, consulte Como converter sua chave privada usando o PuTTYgen no Guia do usuário do Amazon EC2.

Conectar-se ao nó primário usando PuTTY

1. Verifique se você permitiu tráfego SSH de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Abra o `putty.exe`. Você também pode executar o PuTTY na lista de programas do Windows.
3. Se necessário, na lista Category (Categoria), escolha Session (Sessão).
4. Em Nome do host (ou endereço IP), digite `hadoop@` *`MasterPublicDNS`*. Por exemplo, `hadoop@`*`ec2-###-##-##-###.compute-1.amazonaws.com`*.
5. Na lista Category (Categoria), escolha Connection > SSH (Conexão > SSH), Auth.
6. Para Private key file for authentication (Arquivo de chave privada para autenticação), escolha Browse (Procurar) e selecione o arquivo `.ppk` que você gerou.
7. Escolha Abrir e depois Sim para descartar o alerta de segurança do PuTTY.

> **Important**
>
> Quando fizer login no nó primário, digite `hadoop` se for solicitado a especificar um nome de usuário.

8. Quando terminar de trabalhar no nó primário, você poderá encerrar a conexão SSH fechando o PuTTY.

> **Note**
>
> Para evitar que a conexão SSH atinja o tempo limite, é possível escolher Connection (Conexão) na lista Category (Categoria) e selecionar a opção Enable TCP_keepalives (Habilitar TCP_keepalives). Se você tiver uma sessão SSH ativa no PuTTY, poderá alterar suas configurações abrindo o contexto (botão direito do mouse) para a barra de título do PuTTY e escolhendo Alterar configurações.

Caso tenha dificuldades ao usar o SSH para se conectar ao nó primário, consulte Solução de problemas para conectar-se à sua instância.

Conectar-se ao nó primário usando a AWS CLI

Você pode criar uma conexão SSH com o nó primário usando o AWS CLI no Windows e no Linux, Unix e Mac OS X. Independentemente da plataforma, você precisa do nome DNS público do nó primário e da chave privada do par de chaves do Amazon EC2. Se você estiver usando o AWS CLI no Linux, Unix ou Mac OS X, também deverá definir permissões no arquivo (`.pemou.ppk`) chave privada, conforme mostrado emPara configurar as permissões do arquivo de chave privada do par de chaves.

Para se conectar ao nó primário usando o AWS CLI

1. Verifique se você permitiu tráfego SSH de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Para recuperar o identificador de cluster, digite:

   ```
   aws emr list-clusters
   ```

   A saída lista seus clusters, incluindo os IDs dos clusters. Observe o ID do cluster ao qual você está se conectando.

   ```
   "Status": {
       "Timeline": {
           "ReadyDateTime": 1408040782.374,
           "CreationDateTime": 1408040501.213
       },
       "State": "WAITING",
       "StateChangeReason": {
           "Message": "Waiting after step completed"
       }
   },
   "NormalizedInstanceHours": 4,
   "Id": "j-2AL4XXXXXX5T9",
   "Name": "AWS CLI cluster"
   ```

3. Digite o comando a seguir para abrir uma conexão SSH com o nó primário. No exemplo a seguir, substitua *j-2AL4XXXXXX5T9* pelo ID do cluster e substitua *~/mykeypair.key* pelo

caminho completo e nome do arquivo `.pem` (para Linux, Unix e Mac OS X) ou arquivo `.ppk` (para Windows). Por exemplo, `C:\Users\<username>\.ssh\mykeypair.pem`.

```
aws emr ssh --cluster-id j-2AL4XXXXXX5T9 --key-pair-file ~/mykeypair.key
```

4. Quando terminar de trabalhar no nó primário, feche a AWS CLI janela.

Para obter mais informações, consulte os comandos do Amazon EMR na AWS CLI. Caso tenha dificuldades ao usar o SSH para se conectar ao nó primário, consulte Solução de problemas para conectar-se à sua instância.

Portas de serviços do Amazon EMR

> **Note**
>
> Veja a seguir interfaces e portas de serviço para componentes do Amazon EMR. Esta não é uma lista completa de portas de serviço. Não estão listados serviços não padrão, como portas SSL e outros tipos de protocolos.

> **Important**
>
> Tome cuidado ao editar as regras de grupo de segurança para portas abertas. Adicione regras que só permitam tráfego de clientes confiáveis e autenticados para os protocolos e portas que sejam necessários para executar suas workloads.

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Porta	Chave de configuração
Hadoop	API REST DO KMS PARA HTTP	Sim	9600	hadoop.kms.http.port
HDFS	IU da Web do Namenode	Sim	9870	dfs.namenode.http-address

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Porta	Chave de configuração
	RPC do Nomenode	Sim	8020	dfs.namenode.rpc-address
	DataNode UI da Web	Sim	9864	dfs.datanode.http.address
	HTTP do Datanode para transferência de dados	Sim	986	dfs.datanode.address
	RPC do Datanode para transferência de dados	Sim	9867	dfs.datanode.ipc.address
Hive	HiveServer2 Parcimônia	Sim	10000	hive.server2.thrift.port
	HiveServer2 HTTP	Não	10001	hive.server2.thrift.http.port
	HiveServer2 Interface de usuário da Web	Sim	10002	hive.server2.webui.port
	Hive Metastore	Sim	9083	hive.metastore.port / metastore.thrift.port
	WebHCat	Não	50111	templeton.port
	Serviço de gerenciamento de daemon do LLAP (RPC)	Não	15004	hive.llap.management.rpc.port

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Porta	Chave de configuração
	Porta de shuffle do YARN para shuffle hospedado pelo daemon do LLAP	Não	15551	hive.llap.daemon.yarn.shuffle.port
	O RPC do daemon do LLAP	Não	Dinâmico	hive.llap.daemon.rpc.port
	IU da Web do daemon do LLAP	Não	15002	hive.llap.daemon.web.port
	Serviço de saída do daemon do LLAP	Não	15003	hive.llap.daemon.output.service.port
Oozie		Sim	11000	
Tez	IU Tez	Sim	8080	
YARN	Shuffle	Sim	13562	mapreduce.shuffle.port
	RPC do localizador	Sim	8040	yarn.nodemanager.localizer.address
		Sim	8041	
	Endereço do NM Webapp	Sim	8042	yarn.nodemanager.webapp.address

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Porta	Chave de configuração
	Aplicação Web RM	Sim	8088	yarn.resourcemanager.webapp.address
		Sim	8025	
	Scheduler	Sim	8030	yarn.resourcemanager.scheduler.address
	interface do gerenciador de aplicações	Sim	8032	yarn.resourcemanager.address
	Interface do administrador do RM	Sim	8033	yarn.resourcemanager.admin.address
	JobHistory UI da Web do servidor	Sim	1988	mapreduce.jobhistory.webapp.address
	JobHistory UI da Web para administrador do servidor	Sim	10033	mapreduce.jobhistory.admin.address
	JobHistory Servidor (RPC)	Sim	10020	mapreduce.jobhistory.address

Componente	Service description (Descrição do serviço)	Serviço em execução por padrão	Porta	Chave de configuração
	Application Timeline Server (RPC)	Sim	10200	yarn.timeline-service.address
	IU da Web do HTTP do Application Timeline Server	Sim	8188	yarn.timeline-service.webapp.address
	IU da Web do HTTPS do Application Timeline Server	Não	8190	yarn.timeline-service.webapp.https.address
		Sim	2088	
Zookeeper	Porta de cliente	Sim	2181	
		Sim	37301	
		Sim	8341	

Visualizar interfaces Web hospedadas em clusters do Amazon EMR

Important

É possível configurar um grupo de segurança personalizado para permitir acesso de entrada a essas interfaces da Web. Lembre-se de que qualquer porta na qual você permita o tráfego de entrada representa uma possível vulnerabilidade de segurança. Revise atentamente os grupos de segurança personalizados para minimizar vulnerabilidades. Para ter mais informações, consulte Controle do tráfego de rede com grupos de segurança.

O Hadoop e outras aplicações que você instala no cluster do EMR publicam interfaces de usuário como sites hospedados no nó primário. Por motivos de segurança, ao usar grupos de segurança gerenciados pelo Amazon EMR, esses sites estão disponíveis somente no nó primário do servidor Web local. Por isso, é necessário se conectar ao nó primário para visualizar as interfaces Web. Para ter mais informações, consulte Conectar-se ao nó primário usando SSH. O Hadoop também publica interfaces de usuário como sites hospedados nos nós core e escravos. Esses sites também só estão disponíveis em servidores Web locais nos nós.

A tabela a seguir lista as interfaces da web que você pode visualizar nas instâncias do cluster. Essas interfaces Hadoop estão disponíveis em todos os clusters. Para as interfaces da instância principal, substitua *master-public-dns-name* pelo DNS público principal listado na guia Resumo do cluster no console do Amazon EMR. Para interfaces de instâncias core e de tarefas, substitua *coretask-public-dns-name* pelo Public DNS name (Nome público DNS) listado para a instância. Para encontrar um PNome público DNS, no console do Amazon EMR, escolha seu cluster na lista, escolha a guia Hardware, escolha o ID do grupo de instâncias que contém a instância à qual você deseja se conectar e anote o Nome público DNS listado para a instância.

Nome da interface	URI
Servidor de histórico do Flink (EMR versão 5.33 e posteriores)	http://*master-public-dns-name* :8082/
Ganglia	http://*master-public-dns-name* /ganglia/
Hadoop HDFS (versão NameNode EMR anterior à 6.x)	https://*master-public-dns-name* :50470/
Hadoop HDFS NameNode	http://*master-public-dns-name* :50070/
Hadoop HDFS DataNode	http://*coretask-public-dns-name* :50075/
Hadoop HDFS (NameNode EMR versão 6.x)	https://*master-public-dns-name* :9870/
Hadoop HDFS (versão DataNode EMR anterior à 6.x)	https://*coretask-public-dns-name* :50475/

Nome da interface	URI
Hadoop HDFS (DataNode EMR versão 6.x)	https://*coretask-public-dns-name* :9865/
HBase	http://*master-public-dns-name* :16010/
Hue	http://*master-public-dns-name* :8888/
JupyterHub	https://*master-public-dns-name* :9443/
Livy	http://*master-public-dns-name* :8998/
Faísca HistoryServer	http://*master-public-dns-name* :18080/
Tez	http://*master-public-dns-name* :8080/tez-ui
FIO NodeManager	http://*coretask-public-dns-name* :8042/
FIO ResourceManager	http://*master-public-dns-name* :8088/
Zeppelin	http://*master-public-dns-name* :8890/

Como existem várias interfaces específicas de aplicações disponíveis no nó primário, mas não disponíveis nos nós centrais e de tarefa, as instruções neste documento são específicas para o nó primário do Amazon EMR. O acesso as interfaces Web em todos os nós centrais e de tarefa pode ser feito da mesma maneira como você acessaria as interfaces Web no nó primário.

Existem várias maneiras de acessar as interfaces Web no nó primário. O método mais fácil e rápido é usar o SSH para conectar-se ao nó primário e usar o navegador baseado em texto, o Lynx, para visualizar os sites no cliente SSH. No entanto, o Lynx é um navegador baseado em texto com uma interface de usuário limitada que não pode exibir gráficos. O exemplo a seguir mostra como abrir a ResourceManager interface do Hadoop usando o Lynx (URLs do Lynx também são fornecidos quando você faz login no nó primário usando SSH).

```
lynx http://ip-###-##-##-###.us-west-2.compute.internal:8088/
```

Existem duas opções restantes para acessar interfaces Web no nó primário que fornecem funcionalidade de navegador completa. Escolha uma das seguintes opções:

- Opção 1 (recomendada para usuários mais técnicos): use um cliente SSH para conectar-se ao nó primário, configurar o túnel SSH com o encaminhamento de porta local e usar um navegador da Internet para abrir interfaces Web hospedadas no nó primário. Esse método permite que você configure o acesso à interface Web sem usar um proxy SOCKS.
- Opção 2 (recomendada para novos usuários): use um cliente SSH para se conectar ao nó primário, configure o tunelamento SSH com encaminhamento dinâmico de portas e configure seu navegador da Internet para usar um complemento, como o Firefox ou SwitchyOmega o Chrome, FoxyProxy para gerenciar suas configurações de proxy SOCKS. Esse método permite filtrar URLs automaticamente com base em padrões de texto e limitar as configurações de proxy para domínios que correspondam ao formato do nome DNS do nó primário. Para obter mais informações sobre como configurar FoxyProxy para o Firefox e o Google Chrome, consulteOpção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário.

> **Note**
>
> Se você modificar a porta em que a aplicação é executada por meio da configuração do cluster, o hiperlink para a porta não será atualizado no console do Amazon EMR. Isso ocorre porque o console não tem a funcionalidade de ler a configuração `server.port`.

Com a versão 5.25.0 ou posterior do Amazon EMR, você pode acessar a interface do usuário do servidor de histórico do Spark a partir do console sem configurar um proxy da web por meio de uma conexão SSH. Para obter mais informações, consulte One-click access to persistent Spark history server.

Tópicos

Opção 1: configurar um túnel SSH ao nó primário usando o encaminhamento de portas locais

Para se conectar ao servidor Web local no nó primário, crie um túnel SSH entre seu computador e o nó primário. Isso também é conhecido como encaminhamento de portas. Se não quiser usar um proxy SOCKS, você poderá configurar um túnel SSH para o nó primário usando o encaminhamento de portas locais. Com o encaminhamento de portas locais, você pode especificar portas locais que são utilizadas para encaminhar o tráfego a portas remotas específicas no servidor Web local do nó primário.

A configuração de um túnel SSH usando o encaminhamento de portas locais requer o nome DNS público do nó primário e seu arquivo de chave privada do par de chaves. Para obter informações sobre como localizar o nome DNS público principal, consulte Recuperar o nome DNS público do nó primário usando o console antigo. Para obter mais informações sobre como acessar seu par de chaves, consulte os pares de chaves do Amazon EC2 no Guia do usuário do Amazon EC2. Para obter mais informações sobre os sites que você pode querer visualizar no nó primário, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

Configurar um túnel SSH ao nó primário usando o encaminhamento de portas locais com OpenSSH

Para configurar um túnel SSH usando o encaminhamento de portas locais no terminal

1. Verifique se você permitiu tráfego SSH de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Abra uma janela do terminal. No Mac OS X, escolha Applications > Utilities > Terminal (Aplicativos > Utilitários > Terminal). Em outras distribuições do Linux, o terminal está normalmente localizado em Applications > Accessories > Terminal (Aplicativos > Acessórios > Terminal).
3. Digite o comando a seguir para abrir um túnel SSH em sua máquina local. Este exemplo de comando acessa a interface ResourceManager da web encaminhando o tráfego na porta local 8157 (uma porta local não usada escolhida aleatoriamente) para a porta 8088 no servidor web local do nó principal.

   No comando, substitua *~/mykeypair.pem* pela localização e pelo nome do arquivo `.pem` e substitua *ec2-###-##-##-###.compute-1.amazonaws.com* pelo nome DNS público principal do cluster. Para acessar outra interface da Web, substitua `8088` pelo número de porta aplicável. Por exemplo, substitua `8088` por `8890` na interface do Zeppelin.

```
ssh -i ~/mykeypair.pem -N -L 8157:ec2-###-##-##-
###.compute-1.amazonaws.com:8088 hadoop@ec2-###-##-##-###.compute-1.amazonaws.com
```

`-L` significa o uso do encaminhamento de portas locais, que permite especificar uma porta local usada para encaminhar dados à porta remota identificada no servidor Web local do nó principal.

Após a emissão desse comando, o terminal permanece aberto e não retorna uma resposta.

4. Para abrir a interface ResourceManager da web em seu navegador, digite `http://localhost:8157/` na barra de endereço.
5. Quando terminar de trabalhar com as interfaces Web no nó primário, feche as janelas do terminal.

Opção 2, parte 1: configurar um túnel SSH para o nó primário usando o encaminhamento de portas dinâmicas

Para se conectar ao servidor Web local no nó primário, crie um túnel SSH entre seu computador e o nó primário. Isso também é conhecido como encaminhamento de portas. Se você criar seu túnel SSH usando o encaminhamento de portas dinâmicas, todo o tráfego encaminhado para uma porta local não utilizada especificada será encaminhado ao servidor Web local no nó primário. Isso cria um proxy SOCKS. Em seguida, você pode configurar seu navegador da Internet para usar um complemento, como FoxyProxy ou SwitchyOmega para gerenciar suas configurações de proxy SOCKS.

Usar um complemento de gerenciamento de proxy permite filtrar URLs automaticamente com base em padrões de texto e limitar as configurações de proxy para domínios que correspondam ao formato do nome DNS do nó primário. O complemento do navegador manipula automaticamente a ativação e desativação do proxy quando você alterna entre visualizar sites hospedados no nó primário e aqueles na Internet.

Antes de começar, você precisa do nome DNS público do nó primário e do arquivo de chave privada do par de chaves. Para obter informações sobre como localizar o nome DNS público primário, consulte Recuperar o nome DNS público do nó primário usando o console antigo. Para obter mais informações sobre como acessar seu par de chaves, consulte os pares de chaves do Amazon EC2 no Guia do usuário do Amazon EC2. Para obter mais informações sobre os sites que você pode querer visualizar no nó primário, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

Para configurar um túnel SSH usando o encaminhamento de portas dinâmicas do nó primário com o OpenSSH

Configurar um túnel SSH usando o encaminhamento de portas dinâmicas com OpenSSH

1. Verifique se você permitiu tráfego SSH de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Abra uma janela do terminal. No Mac OS X, escolha Applications > Utilities > Terminal (Aplicativos > Utilitários > Terminal). Em outras distribuições do Linux, o terminal está normalmente localizado em Applications > Accessories > Terminal (Aplicativos > Acessórios > Terminal).
3. Digite o seguinte comando para abrir um túnel SSH na sua máquina local. Substitua *~/mykeypair.pem* pela localização e pelo nome do arquivo, substitua *8157* por um número de porta local não utilizado e substitua `.pem` *ec2-###-##-##-##-###.compute-1.amazonaws.com* pelo nome DNS público primário do seu cluster.

   ```
   ssh -i ~/mykeypair.pem -N -D 8157 hadoop@ec2-###-##-##-###.compute-1.amazonaws.com
   ```

   Após a execução desse comando, o terminal permanece aberto e não retorna uma resposta.

   > **Note**
   >
   > `-D` significa o uso do encaminhamento de portas dinâmicas, que permite especificar uma porta local usada para encaminhar dados a todas as portas remotas identificadas no servidor Web local do nó primário. O encaminhamento de portas dinâmicas cria um proxy SOCKS local que escuta na porta especificada no comando.

4. Depois que o túnel estiver ativo, configure um proxy SOCKS para o seu navegador. Para ter mais informações, consulte Opção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário.
5. Quando terminar de trabalhar com as interfaces Web no nó primário, feche a janela do terminal.

Configure um túnel SSH usando o encaminhamento dinâmico de portas com o AWS CLI

Você pode criar uma conexão SSH com o nó primário usando o AWS CLI no Windows e no Linux, Unix e Mac OS X. Se você estiver usando o AWS CLI no Linux, Unix ou Mac OS X, deverá definir as permissões no `.pem` arquivo conforme mostrado em. Para configurar as permissões do arquivo de

chave privada do par de chaves Se você estiver usando o AWS CLI no Windows, o PuTTY deverá aparecer na variável de ambiente path ou você poderá receber um erro como OpenSSH ou PuTTY não disponível.

Para configurar um túnel SSH usando o encaminhamento dinâmico de portas com o AWS CLI

1. Verifique se você permitiu tráfego SSH de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Crie uma conexão SSH com o nó primário, conforme mostrado em Conectar-se ao nó primário usando a AWS CLI.
3. Para recuperar o identificador de cluster, digite:

   ```
   aws emr list-clusters
   ```

   A saída lista seus clusters, incluindo os IDs dos clusters. Observe o ID do cluster ao qual você está se conectando.

   ```
   "Status": {
       "Timeline": {
           "ReadyDateTime": 1408040782.374,
           "CreationDateTime": 1408040501.213
       },
       "State": "WAITING",
       "StateChangeReason": {
           "Message": "Waiting after step completed"
       }
   },
   "NormalizedInstanceHours": 4,
   "Id": "j-2AL4XXXXXX5T9",
   "Name": "AWS CLI cluster"
   ```

4. Digite o seguinte comando para abrir um túnel SSH para o nó primário usando o encaminhamento de portas dinâmicas. No exemplo a seguir, substitua *j-2AL4XXXXXX5T9* pelo ID do cluster e substitua *~/mykeypair.key* pelo local e nome do seu arquivo `.pem` (para Linux, Unix e Mac OS X) ou arquivo `.ppk` (para Windows).

   ```
   aws emr socks --cluster-id j-2AL4XXXXXX5T9 --key-pair-file ~/mykeypair.key
   ```

> **Note**
>
> O comando socks configura automaticamente o encaminhamento de portas dinâmicas na porta local 8157. Atualmente, essa configuração não pode ser modificada.

5. Depois que o túnel estiver ativo, configure um proxy SOCKS para o seu navegador. Para ter mais informações, consulte Opção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário.
6. Quando terminar de trabalhar com as interfaces da Web no nó primário, feche a AWS CLI janela.

   Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Configurar um túnel SSH para o nó primário usando PuTTY

Os usuários do Windows podem usar um cliente SSH, como o PuTTY, para criar um túnel SSH para o nó primário. Antes de se conectar ao nó primário do Amazon EMR, você deve baixar e instalar PuTTY e PuTTYgen. Você pode baixar essas ferramentas na página de download do PuTTY.

O PuTTY não oferece suporte nativamente ao formato de arquivo de chave privada com par de chaves (`.pem`) gerado pelo Amazon EC2. Você usa o PuTTY para converter seu arquivo de chaves no formato PuTTY necessário (`.ppk`). É necessário converter a chave nesse formato (`.ppk`) antes de tentar se conectar ao nó primário usando o PuTTY.

Para obter mais informações sobre a conversão da sua chave, consulte Como converter sua chave privada usando o PuTTYgen no Guia do usuário do Amazon EC2.

Configurar um túnel SSH usando o encaminhamento de portas dinâmicas usando PuTTY

1. Verifique se você permitiu tráfego SSH de entrada. Para obter instruções, consulte Antes de se conectar: autorize o tráfego de entrada.
2. Clique duas vezes em `putty.exe` para iniciar o PuTTY. Você também pode executar o PuTTY na lista de programas do Windows.

> **Note**
>
> Se você já tiver uma sessão SSH ativa com o nó primário, poderá adicionar um túnel clicando com o botão direito do mouse na barra de título do PuTTY e escolhendo Alterar configurações.

3. Se necessário, na lista Category (Categoria), escolha Session (Sessão).
4. No campo Nome do host, digite **hadoop@** *MasterPublicDNS*. Por exemplo, **hadoop@***ec2-###-##-##-###.compute-1.amazonaws.com*.
5. Na lista Category (Categoria), expanda Connection > SSH (Conexão > SSH) e escolha Auth.
6. Para Private key file for authentication (Arquivo de chave privada para autenticação), escolha Browse (Procurar) e selecione o arquivo `.ppk` que você gerou.

> **Note**
>
> O PuTTY não oferece suporte nativamente ao formato de arquivo de chave privada com par de chaves (`.pem`) gerado pelo Amazon EC2. Você usa o PuTTY para converter seu arquivo de chaves no formato PuTTY necessário (`.ppk`). É necessário converter a chave nesse formato (`.ppk`) antes de tentar se conectar ao nó primário usando o PuTTY.

7. Na lista Category (Categoria), expanda Connection > SSH (Conexão > SSH) e escolha Tunnels (Túneis).
8. No campo Porta de origem, digite `8157` (uma porta local não utilizada) e escolha Adicionar.
9. Deixe o campo Destination (Destino) em branco.
10. Selecione as opções Dynamic (Dinâmico) e Auto.
11. Escolha Open (Abrir).
12. Escolha Yes (Sim) para descartar o alerta de segurança do PuTTY.

> **Important**
>
> Ao fazer login no nó primário, digite `hadoop` se for solicitado um nome de usuário.

13. Depois que o túnel estiver ativo, configure um proxy SOCKS para o seu navegador. Para ter mais informações, consulte Opção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário.
14. Quando terminar de trabalhar com as interfaces Web no nó primário, feche a janela do PuTTY.

Opção 2, parte 2: configurar definições de proxy para visualizar sites hospedados no nó primário

Se você usar um túnel SSH com o encaminhamento de portas dinâmicas, deverá usar um complemento de gerenciamento de proxy SOCKS para controlar as configurações de proxy no seu navegador. Usar uma ferramenta de gerenciamento de proxy SOCKS permite filtrar URLs automaticamente com base em padrões de texto e limitar as configurações de proxy para domínios que correspondam ao formato do nome DNS do nó primário. O complemento do navegador manipula automaticamente a ativação e desativação do proxy quando você alterna entre visualizar sites hospedados no nó primário e aqueles na Internet. Para gerenciar suas configurações de proxy, configure seu navegador para usar um complemento como FoxyProxy ou SwitchyOmega.

Para obter mais informações sobre como criar um túnel SSH, consulte Opção 2, parte 1: configurar um túnel SSH para o nó primário usando o encaminhamento de portas dinâmicas. Para obter mais informações sobre as interfaces Web disponíveis, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

Inclua as seguintes configurações ao definir o complemento de proxy:

- Use localhost como endereço do host.
- Use o mesmo número de porta local que você selecionou para estabelecer o túnel SSH com o nó primário em Opção 2, parte 1: configurar um túnel SSH para o nó primário usando o encaminhamento de portas dinâmicas. Por exemplo, porta *8157*. Essa porta também deve corresponder ao número de porta que você usa no PuTTY ou em qualquer outro emulador de terminal usado para conexão.
- Especifique o protocolo SOCKS v5. Com o SOCKS v5, opcionalmente, você pode configurar a autorização do usuário.
- URL Patterns (Padrões de URL)

  Os seguintes padrões de URL devem ser allow-listed e estão especificados com um tipo de padrão de curinga:

  - Os padrões *ec2*.*compute*.amazonaws.com* e *10*.amazonaws.com* para corresponder ao nome DNS público dos clusters nas regiões dos EUA.

- Os padrões *ec2*.compute* e *10*.compute* correspondem ao nome DNS público dos clusters de todas as outras regiões.
- UM 10. * padrão para fornecer acesso aos arquivos de JobTracker log no Hadoop. Altere esse filtro se ele entrar em conflito com seu plano de acesso de rede.
- Os padrões *.ec2.internal* e *.compute.internal* correspondem aos nomes DNS privados (internos) dos clusters na região `us-east-1` e em todas as outras regiões, respectivamente.

Exemplo: Configurar FoxyProxy para o Firefox

O exemplo a seguir demonstra uma configuração FoxyProxy padrão (versão 7.5.1) para o Mozilla Firefox.

FoxyProxy fornece um conjunto de ferramentas de gerenciamento de proxy. Com ele, você pode usar um servidor proxy para URLs que correspondam aos padrões referentes aos domínios usados pelas instâncias do Amazon EC2 no cluster do Amazon EMR.

Para instalar e configurar FoxyProxy usando o Mozilla Firefox

1. No Firefox, acesse https://addons.mozilla.org/, pesquise por FoxyProxy Padrão e siga as instruções para adicionar FoxyProxy ao Firefox.
2. Usando um editor de texto, crie um arquivo JSON chamado `foxyproxy-settings.json` com base no exemplo de configuração a seguir.

```
{
  "k20d21508277536715": {
    "active": true,
    "address": "localhost",
    "port": 8157,
    "username": "",
    "password": "",
    "type": 3,
    "proxyDNS": true,
    "title": "emr-socks-proxy",
    "color": "#0055E5",
    "index": 9007199254740991,
    "whitePatterns": [
      {
        "title": "*ec2*.*compute*.amazonaws.com*",
        "active": true,
        "pattern": "*ec2*.*compute*.amazonaws.com*",
```

```
      "importedPattern": "*ec2*.*compute*.amazonaws.com*",
      "type": 1,
      "protocols": 1
    },
    {
      "title": "*ec2*.compute*",
      "active": true,
      "pattern": "*ec2*.compute*",
      "importedPattern": "*ec2*.compute*",
      "type": 1,
      "protocols": 1
    },
    {
      "title": "10.*",
      "active": true,
      "pattern": "10.*",
      "importedPattern": "http://10.*",
      "type": 1,
      "protocols": 2
    },
    {
      "title": "*10*.amazonaws.com*",
      "active": true,
      "pattern": "*10*.amazonaws.com*",
      "importedPattern": "*10*.amazonaws.com*",
      "type": 1,
      "protocols": 1
    },
    {
      "title": "*10*.compute*",
      "active": true,
      "pattern": "*10*.compute*",
      "importedPattern": "*10*.compute*",
      "type": 1,
      "protocols": 1
    },
    {
      "title": "*.compute.internal*",
      "active": true,
      "pattern": "*.compute.internal*",
      "importedPattern": "*.compute.internal*",
      "type": 1,
      "protocols": 1
    },
```

```
      {
        "title": "*.ec2.internal* ",
        "active": true,
        "pattern": "*.ec2.internal*",
        "importedPattern": "*.ec2.internal*",
        "type": 1,
        "protocols": 1
      }
    ],
    "blackPatterns": []
  },
  "logging": {
    "size": 100,
    "active": false
  },
  "mode": "patterns",
  "browserVersion": "68.12.0",
  "foxyProxyVersion": "7.5.1",
  "foxyProxyEdition": "standard"
}
```

3. Abra a página Gerenciamento de extensões do Firefox (acesse about:addons e escolha Extensões).
4. Escolha FoxyProxy Padrão e, em seguida, escolha o botão de mais opções (o botão que parece uma elipse).
5. Selecione Opções no menu suspenso.
6. Escolha Importar configurações no menu esquerdo.
7. Na página Configurações de importação, escolha Importar configurações em Importar configurações do FoxyProxy 6.0+, navegue até o local do **`foxyproxy-settings.json`** arquivo que você criou, selecione o arquivo e escolha Abrir.
8. Escolha OK quando solicitado para substituir as configurações atuais e salvar a nova configuração.

Exemplo: Configurar SwitchyOmega para o Chrome

O exemplo a seguir demonstra como configurar a SwitchyOmega extensão para o Google Chrome. SwitchyOmega permite configurar, gerenciar e alternar entre vários proxies.

Para instalar e configurar SwitchyOmega usando o Google Chrome

1. Acesse https://chrome.google.com/webstore/category/extensions, pesquise por Proxy SwitchyOmega e adicione-o ao Chrome.
2. Escolha Novo perfil e insira `emr-socks-proxy` como nome do perfil.
3. Escolha Perfil PAC e Criar. Os arquivos de configuração automática de proxy (PAC) ajudam a definir uma lista de permissões para solicitações do navegador que devem ser encaminhadas a um servidor proxy da Web.
4. No campo Script PAC, substitua o conteúdo pelo script a seguir, que define quais URLs deverão ser encaminhadas pelo servidor proxy da Web. Se você especificou outro número de porta ao configurar o túnel SSH, substitua *8157* pelo número da porta.

```
function FindProxyForURL(url, host) {
    if (shExpMatch(url, "*ec2*.*compute*.amazonaws.com*")) return 'SOCKS5
 localhost:8157';
    if (shExpMatch(url, "*ec2*.compute*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "http://10.*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*10*.compute*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*10*.amazonaws.com*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*.compute.internal*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*ec2.internal*")) return 'SOCKS5 localhost:8157';
    return 'DIRECT';
}
```

5. Em Ações, escolha Aplicar alterações para salvar as configurações de proxy.
6. Na barra de ferramentas do Chrome, escolha SwitchyOmega e selecione o `emr-socks-proxy` perfil.

Acessar uma interface da Web no navegador

Para abrir uma interface Web, insira o nome DNS público do seu nó primário ou central seguido pelo número da porta da interface escolhida na barra de endereço do navegador. O exemplo a seguir mostra a URL que você digitaria para se conectar ao Spark HistoryServer.

```
http://master-public-dns-name:18080/
```

Para obter instruções sobre como recuperar o nome DNS público de um nó, consulte Recuperar o nome DNS público do nó primário. Para obter uma lista completa dos URLs de interface da Web, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

Enviar trabalhos a um cluster

Esta seção descreve os métodos que você pode usar para enviar trabalhos a um cluster do Amazon EMR. Para enviar trabalhos, é possível adicionar etapas ou enviar trabalhos do Hadoop de forma interativa para o nó primário.

Considere estas regras de comportamento de etapas ao enviar etapas para um cluster:

- Um ID de etapa pode conter até 256 caracteres.
- Pode haver até 256 etapas PENDING e RUNNING em um cluster.
- Mesmo com 256 etapas ativas em execução no cluster, é possível enviar trabalhos de forma interativa ao nó primário. Você pode enviar um número ilimitado de etapas durante a vida útil de um cluster de execução prolongada, mas apenas 256 etapas podem estar no estado RUNNING ou PENDING em um determinado momento.
- Com as versões 4.8.0 e posteriores do Amazon EMR, exceto a versão 5.0.0, você pode cancelar etapas pendentes. Para ter mais informações, consulte Cancelar etapas.
- Com o Amazon EMR 5.28.0 e versões posteriores, você pode cancelar as etapas pendentes e em execução. Você também pode optar por executar várias etapas em paralelo para melhorar a utilização de cluster e economizar custos. Para ter mais informações, consulte Considerações sobre a execução de várias etapas em paralelo.

> **Note**
>
> Para obter o melhor desempenho, recomendamos que você armazene ações de bootstrap, scripts e outros arquivos personalizados que você deseja usar com o Amazon EMR em um bucket do Amazon S3 que esteja no Região da AWS mesmo que seu cluster.

Tópicos

Adicionar etapas a um cluster com o Console de Gerenciamento do Amazon EMR

Realize os procedimentos a seguir para adicionar etapas a um cluster usando o AWS Management Console. Para obter informações detalhadas sobre como enviar etapas para aplicações específicas de big data, consulte as seguintes seções do Guia de lançamento do Amazon EMR:

- Enviar uma etapa de JAR personalizado
- Enviar uma etapa de transmissão do Hadoop
- Enviar uma etapa do Spark
- Enviar uma etapa do Pig
- Executar um comando ou script como etapa
- Transmitir valores em etapas para executar scripts do Hive

Adicionar etapas durante a criação do cluster

A partir do AWS Management Console, você pode adicionar etapas ao criar um cluster.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Adicionar etapas ao criar um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Em Etapas, escolha Adicionar etapa. Insira os valores apropriados nos campos da caixa de diálogo Adicionar etapa. Para obter informações sobre como formatar argumentos de etapa, consulte Adicionar argumentos de etapas. As opções diferem dependendo do tipo de etapa. Para adicionar a etapa e sair da caixa de diálogo, selecione Adicionar etapa.

4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Old console

Adicionar etapas ao criar um cluster usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/home. EscolhaCriar cluster: opções avançadas.
2. Na página Step 1: Software and Steps (Etapa 1: Software e etapas), em Steps (optional) (Etapas (opcional)), selecione Run multiple steps in parallel to improve cluster utilization and save cost (Executar várias etapas em paralelo para melhorar a utilização do cluster e economizar custos). O valor padrão para o nível de simultaneidade é 10. Você pode escolher de 2 a 256 etapas que podem ser executadas em paralelo.

   > **Note**
   >
   > A execução de várias etapas em paralelo só é compatível com O Amazon EMR 5.28.0 e versões posteriores.

3. Em After last step completes (Após a conclusão da última etapa), escolha Cluster enters waiting state (Cluster no estado de espera) ou Auto-terminate the cluster (Encerrar automaticamente o cluster).
4. Escolha Step type (Tipo de etapa), e Add step (Adicionar etapa).
5. Digite os valores apropriados nos campos da caixa de diálogo Add Step (Adicionar etapa). Para obter informações sobre como formatar argumentos de etapa, consulte Adicionar argumentos de etapas. As opções diferem dependendo do tipo de etapa. Se você tiver habilitado Executar várias etapas em paralelo para melhorar a utilização do cluster e economizar custos, a única opção disponível para Ação em caso de falha será Continuar. Em seguida, escolha Add (Adicionar).

Adicionar etapas a um cluster em execução

Com o AWS Management Console, você pode adicionar etapas a um cluster com a opção de encerramento automático desativada.

New console

Adicionar etapas a um cluster em execução usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Etapas da página de detalhes do cluster, selecione a guia Adicionar etapa. Para clonar uma etapa já existente, escolha o menu suspenso Ações e selecione Clonar etapa.
4. Insira os valores apropriados nos campos da caixa de diálogo Adicionar etapa. As opções diferem dependendo do tipo de etapa. Para adicionar a etapa e sair da caixa de diálogo, escolha Adicionar etapa.

Old console

Adicionar etapas a um cluster em execução usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/home. Na página Cluster List (Lista de clusters), selecione o link para o seu cluster.
2. Na página Cluster Details (Detalhes do Cluster), escolha a guia Steps (Etapas) .
3. Na guia Steps (Etapas), escolha Add step (Adicionar etapa).
4. Digite os valores apropriados nos campos da caixa de diálogo Add Step (Adicionar etapa) e selecione Add (Adicionar). As opções diferem dependendo do tipo de etapa.

Modificar o nível de simultaneidade da etapa em um cluster em execução

Com o AWS Management Console, você pode modificar o nível de simultaneidade de etapas em um cluster em execução.

> **Note**
>
> Só é possível executar múltiplas etapas em paralelo com o Amazon EMR versão 5.28.0 e posteriores.

New console

Modificar a simultaneidade da etapa em um cluster em execução usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar. O cluster deve estar em execução para alterar o respectivo atributo de simultaneidade.
3. Na guia Etapas da página de detalhes do cluster, encontre a seção Atributos. Selecione Editar para alterar a simultaneidade. Insira um valor entre 1 e 256.

Old console

Modificar a simultaneidade da etapa em um cluster em execução usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/home. Na página Cluster List (Lista de clusters), selecione o link para o seu cluster.
2. Na página Cluster Details (Detalhes do Cluster), escolha a guia Steps (Etapas) .
3. Em Concurrency (Simultaneidade), escolha Change (Alterar). Selecione um novo valor para o nível de simultaneidade da etapa e salve-o.

Adicionar argumentos de etapas

Ao usar o AWS Management Console para adicionar uma etapa ao seu cluster, você pode especificar argumentos para essa etapa no campo Argumentos. É necessário separar argumentos com espaço em branco e cercar com aspas argumentos de sequência de caracteres formados por caracteres e espaços em branco.

Example : Argumentos corretos

Os argumentos de exemplo a seguir estão formatados corretamente para o AWS Management Console, com aspas ao redor do argumento final da string.

```
bash -c "aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh ."
```

Também é possível colocar cada argumento em uma linha separada para facilitar a leitura, como mostra o exemplo a seguir.

```
bash
-c
"aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh ."
```

Example : Argumentos incorretos

Os argumentos de exemplo a seguir estão formatados incorretamente para o AWS Management Console. Observa-se que o argumento final da string ,`aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh .`, contém espaços em branco e não está entre aspas.

```
bash -c aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh .
```

Adicionando etapas a um cluster com o AWS CLI

Os procedimentos a seguir demonstram como adicionar etapas a um cluster recém-criado e a um cluster em execução com a AWS CLI. Ambos os exemplos usam o subcomando `--steps` para adicionar etapas ao cluster.

Para adicionar etapas durante a criação do cluster

- Digite o seguinte comando para criar um cluster e adicionar uma etapa do Apache Pig. Substitua *myKey* pelo nome do par de chaves do Amazon EC2.

```
aws emr create-cluster --name "Test cluster" \
--applications Name=Spark \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=PRIMARY,InstanceCount=1,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=2,InstanceType=m5.xlarge \
--steps '[{"Args":["spark-submit","--deploy-mode","cluster","--
class","org.apache.spark.examples.SparkPi","/usr/lib/spark/examples/jars/spark-
examples.jar","5"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"command-
runner.jar","Properties":"","Name":"Spark application"}]'
```

> **Note**
>
> A lista de argumentos muda dependendo do tipo de etapa.

Por padrão, o nível de simultaneidade da etapa é 1. É possível definir o nível de simultaneidade da etapa usando o parâmetro `StepConcurrencyLevel` ao criar um cluster.

A saída é um identificador de cluster semelhante ao seguinte:

```
{
    "ClusterId": "j-2AXXXXXXGAPLF"
}
```

Para adicionar uma etapa a um cluster em execução

- Digite o seguinte comando para adicionar uma etapa a um cluster em execução. Substitua *j-2AXXXXXXGAPLF* por seu próprio ID do cluster.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps '[{"Args":["spark-submit","--deploy-mode","cluster","--
class","org.apache.spark.examples.SparkPi","/usr/lib/spark/examples/jars/spark-
examples.jar","5"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"command-
runner.jar","Properties":"","Name":"Spark application"}]'
```

A saída é um identificador de etapa semelhante ao seguinte:

```
{
    "StepIds": [
 "s-Y9XXXXXXAPMD"
    ]
}
```

Para modificar o StepConcurrencyLevel em um cluster em execução

1. Em um cluster em execução, é possível modificar `StepConcurrencyLevel` com a API `ModifyCluster`. Por exemplo, digite o seguinte comando para aumentar o `StepConcurrencyLevel` para `10`. Substitua *j-2AXXXXXXGAPLF* pelo nome do ID do cluster.

```
aws emr modify-cluster --cluster-id j-2AXXXXXXGAPLF --step-concurrency-level 10
```

2. A saída é semelhante à seguinte.

```
{
"StepConcurrencyLevel": 10
}
```

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte a Referência de AWS CLI comandos.

Considerações sobre a execução de várias etapas em paralelo

- As etapas executadas em paralelo podem ser concluídas em qualquer ordem, mas as etapas pendentes na fila passam para o estado de execução na ordem em que são enviadas.
- Ao selecionar um nível de simultaneidade da etapa para o cluster, você deve considerar se o tipo de instância de nó primário atende ou não aos requisitos de memória das workloads do usuário. O processo executor da etapa principal é executado no nó primário de cada etapa. Executar várias etapas em paralelo requer mais memória e utilização da CPU do nó primário do que executar uma etapa de cada vez.
- Para alcançar a programação complexa e o gerenciamento de recursos de etapas simultâneas, você pode usar recursos de programação do YARN, como `FairScheduler` ou `CapacityScheduler`. Por exemplo, você pode usar o `FairScheduler` com um conjunto `queueMaxAppsDefault` para impedir que mais de um determinado número de trabalhos seja executado por vez.
- O nível de simultaneidade da etapa está sujeito às configurações dos gerenciadores de recursos. Por exemplo, se o YARN estiver configurado com apenas um paralelismo de 5, você só poderá ter cinco aplicativos do YARN sendo executados em paralelo, ainda que o `StepConcurrencyLevel` esteja definido como `10`. Para obter mais informações sobre a configuração de gerenciadores de recursos, consulte Configure applications no Guia de lançamento do Amazon EMR.
- Não é possível adicionar uma etapa com `ActionOnFailure` que não seja CONTINUE enquanto o nível de simultaneidade de etapas do cluster for maior que 1.
- Se o nível de simultaneidade de etapas do cluster for maior que 1, o atributo `ActionOnFailure` da etapa não será ativado.
- Se o nível de simultaneidade de etapas do cluster for 1, mas houver várias etapas em execução, `TERMINATE_CLUSTER ActionOnFailure` poderá ser ativada, mas não `CANCEL_AND_WAIT ActionOnFailure` será. Esse caso extremo ocorre quando o nível de simultaneidade da etapa do cluster é maior que 1, mas diminui durante a execução de várias etapas.

- Você pode usar a escalabilidade automática no EMR para aumentar e diminuir a escala com base nos recursos do YARN de modo a evitar contenção de recursos. Para obter mais informações, consulte Using automatic scaling with a custom policy for instance groups no Guia de gerenciamento do Amazon EMR.
- Quando você diminui o nível de simultaneidade da etapa, o EMR permite que as etapas em execução sejam concluídas antes de reduzir o número de etapas. Se os recursos estiverem esgotados porque o cluster está executando muitas etapas simultâneas, recomendamos cancelar as etapas em execução manualmente para liberar recursos.

Visualizar etapas

Você pode ver até 10.000 etapas que o Amazon EMR concluiu nos últimos sete dias. Você também pode visualizar 1.000 etapas que o Amazon EMR concluiu a qualquer momento. Esse total inclui tanto etapas do sistema quanto etapas enviadas pelo usuário.

Se você enviar novas etapas quando o cluster atingir o limite de registros de 1.000 etapas, o Amazon EMR excluirá as etapas inativas enviadas pelo usuário cujos status foram CONCLUÍDOS, CANCELADOS ou FALHADOS por mais de sete dias. Se você enviar etapas além do limite de registro de 10.000 etapas, o Amazon EMR excluirá os registros de etapas inativos enviados pelo usuário, independentemente de sua duração inativa. O Amazon EMR não remove esses registros dos arquivos de log. O Amazon EMR os remove do AWS console e eles não são retornados quando você usa a API AWS CLI ou para recuperar informações do cluster. Registros de etapas do sistema nunca são removidos.

As informações de etapas que você pode visualizar dependem do mecanismo usado para recuperar informações do cluster. A tabela a seguir indica as informações de etapa retornadas por cada uma das opções disponíveis.

Opção	DescribeJobFlow ou --describe --jobflow	ListSteps ou lista-etapas
SDK	256 etapas	Até 10.000 etapas
CLI do Amazon EMR	256 etapas	N/D
AWS CLI	N/D	Até 10.000 etapas

Opção	DescribeJobFlow ou --describe --jobflow	ListSteps ou lista-etapas
API	256 etapas	Até 10.000 etapas

Cancelar etapas

Você pode cancelar etapas pendentes e em execução da AWS Management Console AWS CLI, da ou da API do Amazon EMR.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Cancelar etapas usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e, em seguida, selecione o cluster que você deseja atualizar.
3. Na guia Etapas da página de detalhes do cluster, marque a caixa de seleção ao lado da etapa que você deseja cancelar. Escolha o menu suspenso Ações e selecione Cancelar etapas.
4. Na caixa de diálogo Cancelar a etapa, escolha entre cancelar a etapa e esperar a saída ou cancelar a etapa e forçar a saída. Depois, selecione Confirm (Confirmar).
5. O status das etapas na tabela Etapas é alterado para `CANCELLED`.

Old console

Cancelar etapas usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Na página Cluster Details (Detalhes do cluster), expanda a seção Steps (Etapas).
3. Para cada etapa a cancelar, selecione a etapa na lista de Steps (Etapas). Em seguida, escolha Cancel step (Cancelar etapa).
4. Na caixa de diálogo Cancel step (Cancelar etapa), mantenha a opção padrão Cancel the step and wait for it to exit (Cancelar a etapa e aguardar que ela saia). Se quiser encerrar a etapa imediatamente sem aguardar a conclusão de nenhum processo, escolha Cancel the step and force it to exit (Cancelar a etapa e forçá-la a sair).
5. Escolha Cancel step (Cancelar etapa).

CLI

Para cancelar usando o AWS CLI

- Use o comando `aws emr cancel-steps`, especificando o cluster e as etapas a serem canceladas. O exemplo a seguir demonstra um comando da AWS CLI para cancelar duas etapas.

```
aws emr cancel-steps --cluster-id j-2QUAXXXXXXXXX \
--step-ids s-3M8DXXXXXXXXX s-3M8DXXXXXXXXX \
--step-cancellation-option SEND_INTERRUPT
```

Com o Amazon EMR versão 5.28.0, você pode escolher uma das duas opções de cancelamento a seguir para o parâmetro `StepCancellationOption` ao cancelar etapas.

- `SEND_INTERRUPT`: essa é a opção padrão. Quando uma solicitação de cancelamento de etapa é recebida, o EMR envia um sinal `SIGTERM` para a etapa. Adicione um processador de sinal `SIGTERM` à lógica de etapas para capturar esse sinal e terminar os processos da etapa descendente ou aguarde eles serem concluídos.

- `TERMINATE_PROCESS`: quando essa opção é selecionada, o EMR envia um sinal `SIGKILL` para a etapa e para todos os seus processos descendentes, o que os termina imediatamente.

Considerações sobre o cancelamento de etapas

- Cancelar uma etapa em execução ou pendente removerá a etapa da contagem de etapas ativas.
- Cancelar uma etapa em execução não permite que uma etapa pendente comece a ser executada, supondo que não haja alteração em `stepConcurrencyLevel`.
- O cancelamento de uma etapa em execução não aciona a etapa `ActionOnFailure`.
- Para o EMR 5.32.0 e versões posteriores, `SEND_INTERRUPT StepCancellationOption` envia um sinal `SIGTERM` para o processo filho da etapa. Observe esse sinal e faça uma limpeza e desligue-o normalmente. `TERMINATE_PROCESS StepCancellationOption` envia um sinal `SIGKILL` para o processo filho da etapa e para todos os seus processos descendentes; mas os processos assíncronos não são afetados.

Visualizar e monitorar um cluster

O Amazon EMR fornece várias ferramentas que você pode usar para coletar informações sobre o cluster. Você pode acessar informações sobre o cluster a partir do console, da CLI ou de forma programática. As interfaces Web padrão do Hadoop e os arquivos de log estão disponíveis no nó primário. Você também pode usar serviços de monitoramento, como o CloudWatch Ganglia, para monitorar o desempenho do seu cluster.

O histórico do aplicativo também está disponível no console usando as interfaces do usuário de aplicativos “persistentes” para o servidor de histórico do Spark a partir da versão Amazon EMR 5.25.0. Com o Amazon EMR 6.x, o servidor persistente de linha de tempo do YARN e as interfaces do usuário Tez também estão disponíveis. Esses serviços são hospedados fora do cluster, portanto, você pode acessar o histórico de aplicativos por 30 dias após o encerramento do cluster, sem a necessidade de uma conexão SSH ou proxy da Web. Consulte Visualizar o histórico da aplicação.

Tópicos

- Visualizar instâncias de cluster no Amazon EC2
- CloudWatch eventos e métricas
- Visualizar métricas para aplicações de cluster com o Ganglia
- Registro de chamadas de API do Amazon EMR em AWS CloudTrail

Visualizar o status e os detalhes do cluster

Depois de criar um cluster, você pode monitorar seu status e obter informações detalhadas sobre sua execução e erros que podem ter ocorrido, mesmo depois de ele ter sido terminado. O Amazon EMR salva metadados sobre clusters terminados para sua referência por dois meses, e os metadados são excluídos após esse período. Você não pode excluir clusters do histórico de clusters, mas, usando o AWS Management Console, você pode usar o Filter (Filtro) e, usando a AWS CLI, você pode usar opções com o comando `list-clusters` para focalizar nos clusters que interessam a você.

Você pode acessar o histórico do aplicativo armazenado no cluster por uma semana a partir de sua gravação, independentemente de se o cluster está em execução ou encerrado. Além disso, as interfaces do usuário de aplicativos persistentes armazenam o histórico de aplicativos fora do cluster por 30 dias após o encerramento de um cluster. Consulte Visualizar o histórico da aplicação.

Para obter mais informações sobre estados de cluster, como Waiting e Running, consulte Noções básicas sobre o ciclo de vida do cluster.

Visualizar os detalhes do cluster usando o AWS Management Console

A lista de clusters em https://console.aws.amazon.com/emr lista todos os clusters em sua conta e AWS região, incluindo clusters encerrados. A lista mostra o seguinte para cada cluster: o Nome e o ID, o Status e os Detalhes do status, a Hora da criação, o Tempo decorrido em que o cluster esteve em execução e as Horas da instância normalizadas que foram acumuladas para todas as instâncias do EC2 no cluster. Essa lista é o ponto de partida para monitorar o status dos clusters. Ela foi criada para que você possa analisar detalhadamente cada cluster para análise e solução de problemas.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Visualizar as informações do cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja visualizar.
3. Use o painel Resumo para visualizar as informações básicas sobre a configuração do cluster, como o status do cluster, as aplicações de código aberto que o Amazon EMR instalou no cluster e a versão do Amazon EMR usada para criar o cluster. Use cada guia abaixo do Resumo para visualizar informações, conforme descrito na tabela a seguir.

Old console

Visualizar as informações do cluster usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Para visualizar um breve resumo das informações do cluster, selecione a seta para baixo ao lado do link do cluster em Nome. A linha do cluster é expandida para fornecer mais informações sobre o cluster, o hardware, as etapas e as ações de bootstrap. Use os links desta seção para analisar informações específicas. Por exemplo, clique em um link em Steps (Etapas) para acessar os arquivos de log das etapas, consultar o JAR associado à etapa, analisar os trabalhos e as tarefas das etapas e acessar os arquivos de log.
3. Para visualizar informações detalhadas sobre o cluster, escolha o link do cluster em Nome para abrir a página de detalhes do cluster. As informações a seguir estão disponíveis na página de detalhes do cluster do console antigo:

Guia (console antigo)	Descrição (console antigo)
Properties	Use essa guia para visualizar o sistema operacional do cluster, as configurações de término e segurança do cluster, as informaçõ

Guia (console antigo)	Descrição (console antigo)
	es sobre a VPC e a sub-rede e onde os logs são armazenados no Amazon S3.
Bootstrap actions (Ações de bootstrap)	Use esta guia para visualizar o status de todas as ações de bootstrap que o cluster executa quando é iniciado. As ações de bootstrap são usadas para instalações de software personalizadas e configurações avançadas. Para ter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais.
Monitoramento	Use essa guia para visualizar as principais métricas de operação do cluster. Você pode visualizar os dados em nível de cluster, os dados em nível de nó e informações sobre E/S e armazenamento físico de dados.
Instâncias	Use essa guia para visualizar informaçõ es sobre nós no cluster, incluindo IDs de instâncias do EC2, nomes DNS, volumes do EBS, endereços IP e muito mais.
Etapas	Use esta guia para ver o status e acessar os arquivos de log das etapas que você enviou. Para mais informações sobre as etapas, consulte Enviar trabalhos a um cluster.

Guia (console antigo)	Descrição (console antigo)
Aplicativos	Use esta guia para exibir detalhes do servidor de linha de tempo do YARN persistente fora do cluster e do aplicativo de interface do usuário do Tez. Também é possível visualizar informações sobre as aplicações instaladas, as configurações do cluster e os grupos de instâncias. As interfaces do usuário do aplicativo no cluster estão disponíveis enquanto o cluster está em execução.
Eventos	Use esta guia para visualizar o log de eventos de seu cluster. Para ter mais informações, consulte Monitorando eventos do Amazon EMR com CloudWatch.
Tags	Use essa guia para visualizar todas as etiquetas aplicadas ao cluster.

Visualize os detalhes do cluster usando o AWS CLI

Os exemplos a seguir demonstram como recuperar detalhes de cluster usando a AWS CLI. Para obter mais informações sobre os comandos disponíveis, consulte AWS CLI Command Reference for Amazon EMR. Você pode usar o comando describe-cluster para visualizar detalhes em nível de cluster, incluindo status, configuração de hardware e software, configurações da VPC, ações de bootstrap, grupos de instâncias, etc. Para obter mais informações sobre estados de cluster, consulte Noções básicas sobre o ciclo de vida do cluster. O exemplo a seguir demonstra o uso do comando `describe-cluster`, seguido por exemplos do comando list-clusters.

Example Visualizar o status do cluster

Para usar o comando `describe-cluster`, você precisa do ID do cluster. Este exemplo demonstra o uso para obter uma lista dos clusters criados em um determinado intervalo de datas e, em seguida, o uso de um dos IDs de cluster retornados para listar mais informações sobre o status de um cluster individual.

O comando a seguir descreve o cluster *j-1K48XXXXXXHCB* que você substitui pelo ID do seu cluster.

```
aws emr describe-cluster --cluster-id j-1K48XXXXXXHCB
```

A saída do comando é semelhante à seguinte.

```
{
    "Cluster": {
        "Status": {
            "Timeline": {
                "ReadyDateTime": 1438281058.061,
                "CreationDateTime": 1438280702.498
            },
            "State": "WAITING",
            "StateChangeReason": {
                "Message": "Waiting for steps to run"
            }
        },
        "Ec2InstanceAttributes": {
            "EmrManagedMasterSecurityGroup": "sg-cXXXXX0",
            "IamInstanceProfile": "EMR_EC2_DefaultRole",
            "Ec2KeyName": "myKey",
            "Ec2AvailabilityZone": "us-east-1c",
            "EmrManagedSlaveSecurityGroup": "sg-example"
        },
        "Name": "Development Cluster",
        "ServiceRole": "EMR_DefaultRole",
        "Tags": [],
        "TerminationProtected": false,
        "ReleaseLabel": "emr-4.0.0",
        "NormalizedInstanceHours": 16,
        "InstanceGroups": [
            {
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "ReadyDateTime": 1438281058.101,
                        "CreationDateTime": 1438280702.499
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
```

```
                }
            },
            "Name": "CORE",
            "InstanceGroupType": "CORE",
            "Id": "ig-2EEXAMPLEXXP",
            "Configurations": [],
            "InstanceType": "m5.xlarge",
            "Market": "ON_DEMAND",
            "RunningInstanceCount": 1
        },
        {
            "RequestedInstanceCount": 1,
            "Status": {
                "Timeline": {
                    "ReadyDateTime": 1438281023.879,
                    "CreationDateTime": 1438280702.499
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "Name": "MASTER",
            "InstanceGroupType": "MASTER",
            "Id": "ig-2A1234567XP",
            "Configurations": [],
            "InstanceType": "m5.xlarge",
            "Market": "ON_DEMAND",
            "RunningInstanceCount": 1
        }
    ],
    "Applications": [
        {
            "Version": "1.0.0",
            "Name": "Hive"
        },
        {
            "Version": "2.6.0",
            "Name": "Hadoop"
        },
        {
            "Version": "0.14.0",
            "Name": "Pig"
        },
```

```
            {
                "Version": "1.4.1",
                "Name": "Spark"
            }
        ],
        "BootstrapActions": [],
        "MasterPublicDnsName": "ec2-X-X-X-X.compute-1.amazonaws.com",
        "AutoTerminate": false,
        "Id": "j-jobFlowID",
        "Configurations": [
            {
                "Properties": {
                    "hadoop.security.groups.cache.secs": "250"
                },
                "Classification": "core-site"
            },
            {
                "Properties": {
                    "mapreduce.tasktracker.reduce.tasks.maximum": "5",
                    "mapred.tasktracker.map.tasks.maximum": "2",
                    "mapreduce.map.sort.spill.percent": "90"
                },
                "Classification": "mapred-site"
            },
            {
                "Properties": {
                    "hive.join.emit.interval": "1000",
                    "hive.merge.mapfiles": "true"
                },
                "Classification": "hive-site"
            }
        ]
    }
}
```

Example Listar clusters por data de criação

Para recuperar clusters criados em um intervalo de dados específicos, use o comando `list-clusters` com os parâmetros `--created-after` e `--created-before`.

O comando a seguir lista todos os clusters criados entre 9 e 12 de outubro de 2019.

```
aws emr list-clusters --created-after 2019-10-09T00:12:00 --created-
before 2019-10-12T00:12:00
```

Example Listar clusters por estado

Para listar clusters por estado, use o comando `list-clusters` com o parâmetro `--cluster-states`. Os estados de cluster válidos incluem: STARTING, BOOTSTRAPPING, RUNNING, WAITING, TERMINATING, TERMINATED e TERMINATED_WITH_ERRORS.

```
aws emr list-clusters --cluster-states TERMINATED
```

Você também pode usar os seguintes parâmetros de atalho para listar todos os clusters nos estados especificados:

- `--active` – filtra clusters nos estados STARTING, BOOTSTRAPPING, RUNNING, WAITING ou TERMINATING.
- `--terminated` filtra clusters no estado TERMINATED.
- O parâmetro `--failed` filtra clusters no estado TERMINATED_WITH_ERRORS.

As seguintes comandos retornam o mesmo resultado.

```
aws emr list-clusters --cluster-states TERMINATED
```

```
aws emr list-clusters --terminated
```

Para obter mais informações sobre estados de cluster, consulte Noções básicas sobre o ciclo de vida do cluster.

Etapa aprimorada de depuração

Se houver falha em uma etapa do Amazon EMR e você enviou seu trabalho usando a operação de etapa de API com uma AMI de versão 5.x ou posterior, o Amazon EMR poderá identificar e retornar a causa raiz da falha na etapa em alguns casos, juntamente com o nome do arquivo de log relevante e uma parte do rastreamento da pilha de aplicações pela API. Por exemplo, as seguintes falhas podem ser identificadas:

- Um erro do Hadoop comum, como o diretório de saída já existe, o diretório de entrada não existe ou um aplicativo ficou sem memória.

- Erros de Java, como um aplicativo que foi compilado com uma versão incompatível do Java ou executado com uma classe principal não encontrada.
- Um problema ao acessar objetos armazenados no Amazon S3.

Essas informações estão disponíveis usando as operações DescribeStepe ListStepsda API. O FailureDetailscampo do StepSummaryretornado por essas operações. Para acessar as FailureDetails informações, use a AWS CLI, o console ou AWS o SDK.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

O novo console do Amazon EMR não oferece depuração por etapas. No entanto, é possível visualizar os detalhes do encerramento do cluster realizando as etapas a seguir.

Visualizar os detalhes usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e, em seguida, selecione o cluster que você deseja visualizar.
3. Observe o valor do Status na seção Resumo da página de detalhes do cluster. Se o status for Terminado com erros, passe o mouse sobre o texto para visualizar os detalhes da falha do cluster.

Old console

Visualizar os detalhes da falha usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.

2. Escolha Cluster List (Lista de clusters) e selecione um cluster.
3. Selecione o ícone de seta ao lado de cada etapa para exibir mais detalhes. Se houve falha na etapa e o Amazon EMR puder identificar a causa raiz, você verá os detalhes da falha.

CLI

Para ver os detalhes da falha com o AWS CLI

- Para obter detalhes da falha de uma etapa com o AWS CLI, use o `describe-step` comando.

```
aws emr describe-step --cluster-id j-1K48XXXXXHCB --step-id s-3QM0XXXXXM1W
```

A saída será semelhante à seguinte:

```
{
  "Step": {
    "Status": {
      "FailureDetails": {
        "LogFile": "s3://myBucket/logs/j-1K48XXXXXHCB/steps/s-3QM0XXXXXM1W/
stderr.gz",
        "Message": "org.apache.hadoop.mapred.FileAlreadyExistsException: Output
 directory s3://myBucket/logs/beta already exists",
        "Reason": "Output directory already exists."
      },
      "Timeline": {
        "EndDateTime": 1469034209.143,
        "CreationDateTime": 1469033847.105,
        "StartDateTime": 1469034202.881
      },
      "State": "FAILED",
      "StateChangeReason": {}
    },
    "Config": {
      "Args": [
        "wordcount",
        "s3://myBucket/input/input.txt",
        "s3://myBucket/logs/beta"
      ],
      "Jar": "s3://myBucket/jars/hadoop-mapreduce-examples-2.7.2-amzn-1.jar",
      "Properties": {}
```

```
    },
    "Id": "s-3QM0XXXXXM1W",
    "ActionOnFailure": "CONTINUE",
    "Name": "ExampleJob"
  }
}
```

Visualizar o histórico da aplicação

É possível visualizar os detalhes da aplicação de serviço do Spark History Server e do cronograma do YARN na página de detalhes do cluster no console. O uso do histórico de aplicação do Amazon EMR facilita a solução de problemas e a análise de trabalhos ativos e do histórico de trabalhos.

> **Note**
>
> Para aumentar a segurança das aplicações fora do console que podem ser usadas com o Amazon EMR, os domínios de hospedagem das aplicações são registrados na Public Suffix List (PSL). Exemplos desses domínios de hospedagem incluem os seguintes: `emrstudio-prod.us-east-1.amazonaws.com`, `emrnotebooks-prod.us-east-1.amazonaws.com`, `emrappui-prod.us-east-1.amazonaws.com`. Para maior segurança, se precisar definir cookies confidenciais no nome de domínio padrão, recomendamos que você use cookies com um prefixo `__Host-`. Isso ajuda a defender seu domínio contra tentativas de falsificação de solicitação entre sites (CSRF). Para obter mais informações, consulte a página Set-Cookie em Mozilla Developer Network.

A seção Interfaces de usuário da aplicação da guia Aplicações fornece várias opções de visualização, conforme o status do cluster e das aplicações instaladas no cluster.

- Acesso fora do cluster a interfaces de usuário de aplicações persistentes: a partir do Amazon EMR versão 5.25.0, os links de interface de usuário de aplicações persistentes estão disponíveis para a interface de usuário do Spark e o Spark History Service. Com o Amazon EMR 5.30.1 e versões posteriores, a IU do Tez e o servidor de linha do tempo do YARN também possuem interfaces do usuário de aplicações persistentes. O servidor de linha do tempo do YARN e a interface do usuário do Tez são aplicativos de código aberto que fornecem métricas para clusters ativos e encerrados. A interface do usuário do Spark fornece detalhes sobre as tarefas e os estágios do programador, tamanhos de RDD e uso de memória, informações sobre o ambiente e informações sobre os executores em execução. As interfaces do usuário de aplicativos persistentes são executadas

fora do cluster, portanto, as informações de cluster e os logs ficam disponíveis por 30 dias após o encerramento de um aplicativo. Ao contrário das interfaces do usuário de aplicativos no cluster, as interfaces do usuário de aplicativos persistentes não exigem que você configure um proxy da Web por meio de uma conexão SSH.

- Interfaces de usuário de aplicações no cluster: há uma variedade de interfaces de usuário de histórico de aplicações que podem ser executadas em um cluster. As interfaces do usuário no cluster são hospedadas no nó principal e exigem que você configure uma conexão SSH com o servidor Web. As interfaces do usuário de aplicativos no cluster mantêm o histórico de aplicativos por uma semana após o encerramento do aplicativo. Para obter mais informações e instruções sobre como configurar um túnel SSH, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

  Com exceção do servidor de histórico do Spark, do servidor de linha de tempo do YARN e das aplicações Hive, o histórico de aplicações no cluster só pode ser visualizado enquanto o cluster estiver em execução.

Visualizar interfaces do usuário de aplicações persistentes

O Amazon EMR versão 5.25.0, você pode se conectar aos detalhes da aplicação servidor de histórico do Spark persistente hospedado fora do cluster usando a página Resumo do cluster ou a guia Interfaces de usuário da aplicação no console. A IU do Tez e as interfaces de aplicações persistentes do servidor de linha de tempo do YARN estão disponíveis a partir do Amazon EMR versão 5.30.1. O acesso com um clique, por meio de um link, ao histórico de aplicativos persistente fornece os seguintes benefícios:

- Você pode analisar e solucionar problemas de trabalhos ativos e histórico de trabalhos sem configurar um proxy da Web por meio de uma conexão SSH.
- Você pode acessar o histórico de aplicativos e os arquivos de log relevantes para clusters ativos e encerrados. Os logs ficam disponíveis por 30 dias após o aplicativo ser encerrado.

Navegue até os detalhes do seu cluster no console e selecione a guia Aplicações. Selecione a interface do usuário da aplicação que você deseja após a inicialização do cluster. A interface do usuário da aplicação abre em uma nova guia do navegador. Para obter mais informações, consulte Monitoring and instrumentation.

Você pode exibir logs de contêiner do YARN por meio dos links no servidor de histórico do Spark, no servidor de linha de tempo do YARN e na interface do usuário do Tez.

> **Note**
>
> Para acessar os logs de contêiner do YARN pelo servidor de histórico do Spark, servidor de linha de tempo do YARN e interface do usuário do Tez, é necessário habilitar o registro em log no Amazon S3 para o cluster. Se você não habilitar o registro em log, os links para os logs de contêiner do YARN não funcionarão.

Coleta de logs

Para habilitar o acesso com um clique às interfaces do usuário de aplicações persistentes, o Amazon EMR coleta dois tipos de logs:

- Os logs de eventos do aplicativo são coletados em um bucket do sistema EMR. Os logs de eventos são criptografados em repouso usando a criptografia no lado do servidor com as chaves gerenciadas pelo Amazon S3 (SSE-S3). Se você usar uma sub-rede privada para o cluster, inclua "`arn:aws:s3:::prod.MyRegion.appinfo.src/*`" na lista de recursos da política do Amazon S3 da sub-rede privada. Para obter mais informações, consulte Minimum Amazon S3 policy for private subnet.
- Os logs de contêiner do YARN são coletados em um bucket do Amazon S3 de sua propriedade. Você deve habilitar o registro em log para que seu cluster acesse logs de contêiner do YARN. Para obter mais informações, consulte Configurar registro em log e depuração do cluster.

Se você precisar desabilitar esse recurso por motivos de privacidade, será possível interromper o daemon usando um script de bootstrap ao criar um cluster, como demonstra o exemplo a seguir.

```
aws emr create-cluster --name "Stop Application UI Support" --release-label emr-7.1.0 \
--applications Name=Hadoop Name=Spark --ec2-attributes KeyName=<myEMRKeyPairName> \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m3.xlarge
 InstanceGroupType=CORE,InstanceCount=1,InstanceType=m3.xlarge
 InstanceGroupType=TASK,InstanceCount=1,InstanceType=m3.xlarge \
--use-default-roles --bootstrap-actions Path=s3://region.elasticmapreduce/bootstrap-
actions/run-if,Args=["instance.isMaster=true","echo Stop Application UI | sudo tee /
etc/apppusher/run-apppusher; sudo systemctl stop apppusher || exit 0"]
```

Após executar esse script de bootstrap, o Amazon EMR não coletará nenhum log de eventos do servidor de histórico do Spark ou servidor de linha de tempo do YARN no bucket do sistema EMR.

Nenhuma informação do histórico do aplicativo estará disponível na guia Interfaces do usuário do aplicativo e você perderá acesso a todas as interfaces do usuário do aplicativo do console.

Arquivos grandes de log de eventos do Spark

Em alguns casos, trabalhos de execução prolongada do Spark, como transmissão do Spark, e trabalhos grandes, como consultas SQL do Spark, podem gerar grandes logs de eventos. Com grandes logs de eventos, é possível usar rapidamente espaço em disco nas instâncias de computação e encontrar erros `OutOfMemory` ao carregar interfaces de usuário persistentes. Para evitar esses problemas, recomenda-se ativar o atributo de rolagem e compactação do log de eventos do Spark. Esse atributo está disponível no Amazon EMR versões emr-6.1.0 e posteriores. Para obter mais detalhes sobre rolagem e compactação, consulte Applying compaction on rolling event log files na documentação do Spark.

Para ativar o atributo de rolagem e compactação do log de eventos do Spark, ative as configurações do Spark a seguir.

- `spark.eventLog.rolling.enabled`: ativa a rolagem do log de eventos com base no tamanho. Essa configuração é desativada por padrão.
- `spark.eventLog.rolling.maxFileSize`: quando a rolagem é ativada, especifica o tamanho máximo do arquivo de log de eventos antes da rolagem. O padrão é 128 MB.
- `spark.history.fs.eventLog.rolling.maxFilesToRetain`: especifica o número máximo de arquivos de log de eventos não compactados a serem retidos. Por padrão, todos os arquivos de log de eventos são mantidos. Defina com um número menor para compactar logs de eventos mais antigos. O valor mais baixo é 1.

A compactação tenta excluir eventos com arquivos de log de eventos desatualizados, como os apresentados a seguir. Se ele descartar eventos, eles não serão mais exibidos na interface do Spark History Server.

- Eventos para trabalhos concluídos e eventos relacionados de preparação ou de tarefa.
- Eventos para executores terminados.
- Eventos para consultas SQL concluídas e eventos relacionados de trabalho, preparação e tarefas.

Para iniciar um cluster com rolagem e compactação habilitadas

1. Crie um arquivo `spark-configuration.json` com a configuração a seguir.

```
[
   {
     "Classification": "spark-defaults",
        "Properties": {
           "spark.eventLog.rolling.enabled": true,
           "spark.history.fs.eventLog.rolling.maxFilesToRetain": 1
        }
   }
]
```

2. Crie o cluster com a configuração de compactação contínua do Spark da forma exibida a seguir.

```
aws emr create-cluster \
--release-label emr-6.6.0 \
--instance-type m4.large \
--instance-count 2 \
--use-default-roles \
--configurations file://spark-configuration.json
```

Considerações e limitações

No momento, o acesso com um clique às interfaces do usuário de aplicações persistentes tem as limitações a seguir.

- Haverá um atraso de pelo menos dois minutos quando os detalhes da aplicação forem exibidos na interface do Spark History Server.
- Esse recurso só funciona quando o diretório de log de eventos do aplicativo está no HDFS. Por padrão, o Amazon EMR armazena logs de eventos em um diretório do HDFS. Se você alterar o diretório padrão para um sistema de arquivos diferente, como, por exemplo o Amazon S3, esse atributo não funcionará.
- Esse recurso não está disponível no momento para clusters do EMR com vários nós principais ou para clusters do EMR integrados ao AWS Lake Formation.
- Para habilitar o acesso com um clique às interfaces do usuário de aplicações persistentes, é preciso ter permissão para a ação `DescribeCluster` do Amazon EMR. Se você negar permissão para um principal do IAM a essa ação, levará aproximadamente cinco minutos para que a alteração na permissão seja propagada.

- Se você reconfigurar os aplicativos em um cluster em execução, o histórico do aplicativo não estará disponível na interface do usuário do aplicativo.
- Para cada um Conta da AWS, o limite padrão para UIs de aplicativos ativos é 200.
- A seguir Regiões da AWS, você pode acessar as UIs do aplicativo a partir do console com o Amazon EMR 6.14.0 e superior:
  - Ásia-Pacífico (Jacarta) (ap-southeast-3)
  - Europa (Espanha) (eu-south-2)
  - Ásia-Pacífico (Melbourne) (ap-southeast-4)
  - Israel (Tel Aviv) (il-central-1)
  - Oriente Médio (EAU) (me-central-1)
- A seguir Regiões da AWS, você pode acessar as UIs do aplicativo a partir do console com o Amazon EMR 5.25.0 e superior:
  - Leste dos EUA (Norte da Virgínia) (us-east-1)
  - Oeste dos EUA (Oregon) (us-west-2)
  - Ásia-Pacífico (Mumbai) (ap-south-1)
  - Ásia-Pacífico (Seul) (ap-northeast-2)
  - Ásia-Pacífico (Singapura) (ap-southeast-1)
  - Ásia-Pacífico (Sydney) (ap-southeast-2)
  - Ásia Pacific (Tóquio) (ap-northeast-1)
  - Canadá (Central) (ca-central-1)
  - América do Sul (São Paulo) (sa-east-1)
  - Europa (Frankfurt) (eu-central-1)
  - Europa (Irlanda) (eu-west-1)
  - Europa (Londres) (eu-west-2)
  - Europa (Paris) (eu-west-3)
  - UE (Estocolmo) (eu-north-1)
  - China (Pequim) (cn-north-1)
  - China (Ningxia) (cn-northwest-1)

Visualizar um histórico de aplicações de alto nível

> **Note**
>
> Recomenda-se usar a interface da aplicação persistente para melhorar a experiência do usuário e reter o histórico da aplicação por até 30 dias. O histórico de aplicações de alto nível descrito nesta página não está disponível no novo console do Amazon EMR (https://console.aws.amazon.com/emr). Para ter mais informações, consulte Visualizar interfaces do usuário de aplicações persistentes.

Com as versões do Amazon EMR 5.8.0 a 5.36.0 e 6.x a 6.8.0, é possível visualizar um histórico de aplicações de alto nível na guia Interfaces de usuário da aplicação do console antigo do Amazon EMR. A Interface de usuário da aplicação do Amazon EMR mantém o resumo do histórico da aplicação por sete dias após a aplicação ser concluída.

Considerações e limitações

Considere as limitações a seguir ao usar a guia Interfaces de usuário da aplicação no console antigo do Amazon EMR.

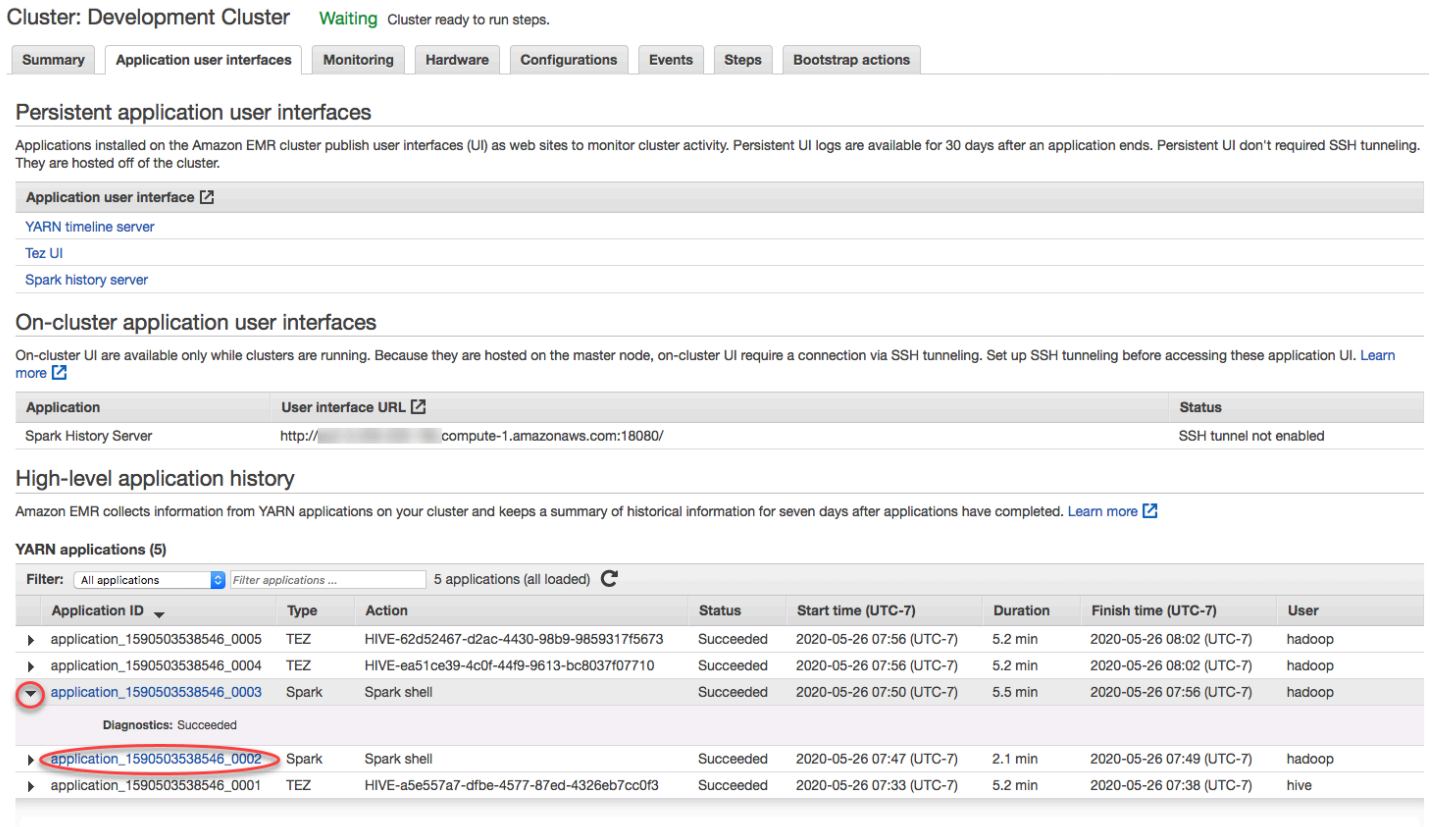
- Só é possível acessar o atributo de histórico de aplicações de alto nível usando as versões 5.8.0 a 5.36.0 e 6.x a 6.8.0 do Amazon EMR. A partir de 23 de janeiro de 2023, o Amazon EMR descontinuará o histórico de aplicações de alto nível em todas as versões. Se você usa o Amazon EMR 5.25.0 ou versões posteriores, recomenda-se usar a interface de usuário da aplicação persistente.
- O atributo de histórico de aplicações de alto nível não é compatível com aplicações Spark Streaming.
- No momento, o acesso com um clique às interfaces de usuário de aplicações persistentes não está disponível para clusters do Amazon EMR com múltiplos nós principais ou para clusters do Amazon EMR integrados ao AWS Lake Formation.

Exemplo: visualizar um histórico de aplicações de alto nível

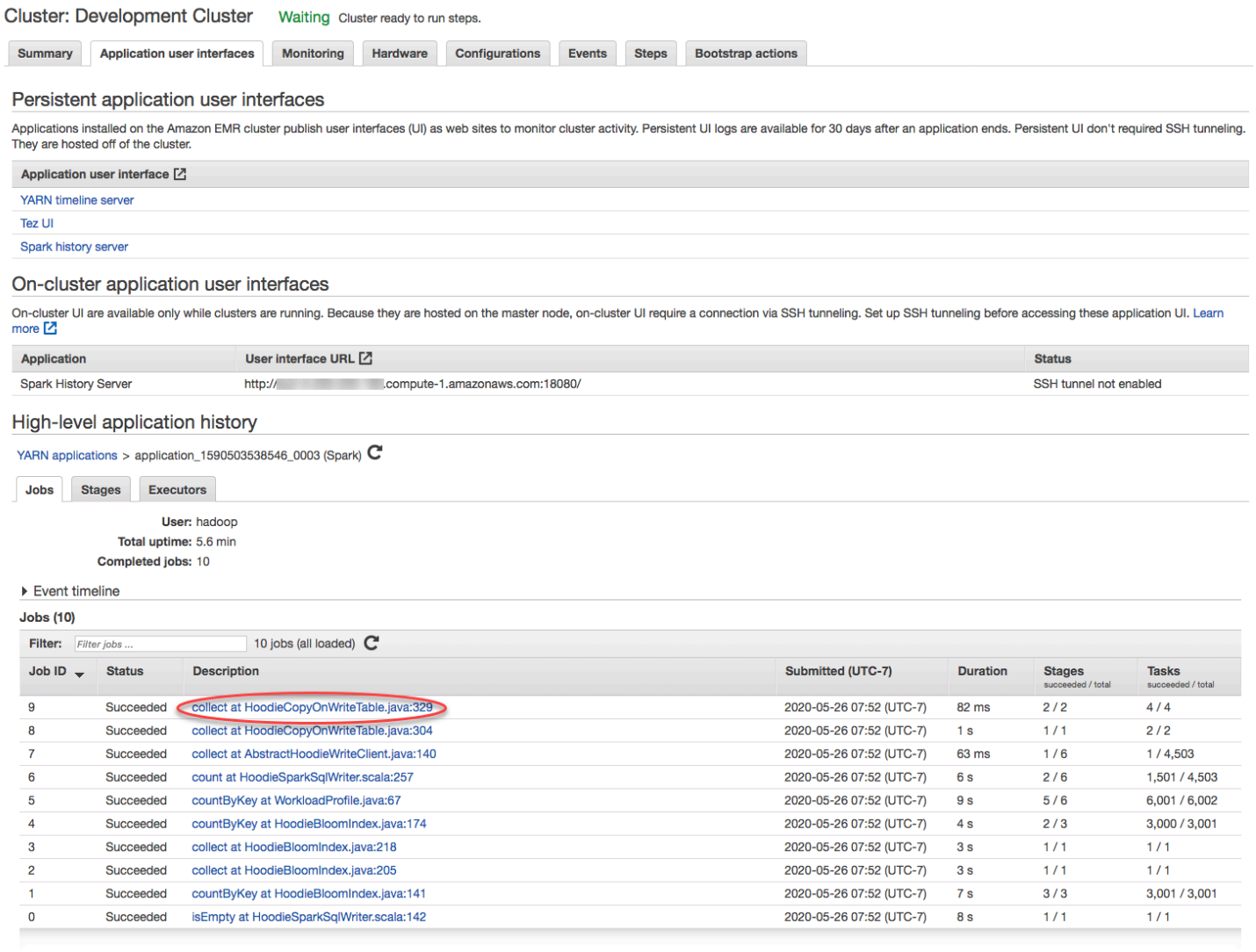
A sequência a seguir demonstra uma busca detalhada de uma aplicação YARN ou Spark nos detalhes do trabalho usando a guia Interfaces do usuário da aplicação na página de detalhes do cluster do console antigo.

Para visualizar detalhes do cluster, selecione o Nome de um cluster na lista Clusters. Para visualizar informações sobre os logs de contêiner do YARN, é necessário habilitar o registro em log do cluster. Para obter mais informações, consulte Configurar registro em log e depuração do cluster. Para o histórico da aplicação Spark, as informações fornecidas na tabela de resumo são apenas um subconjunto das informações disponíveis pela IU do servidor de histórico do Spark.

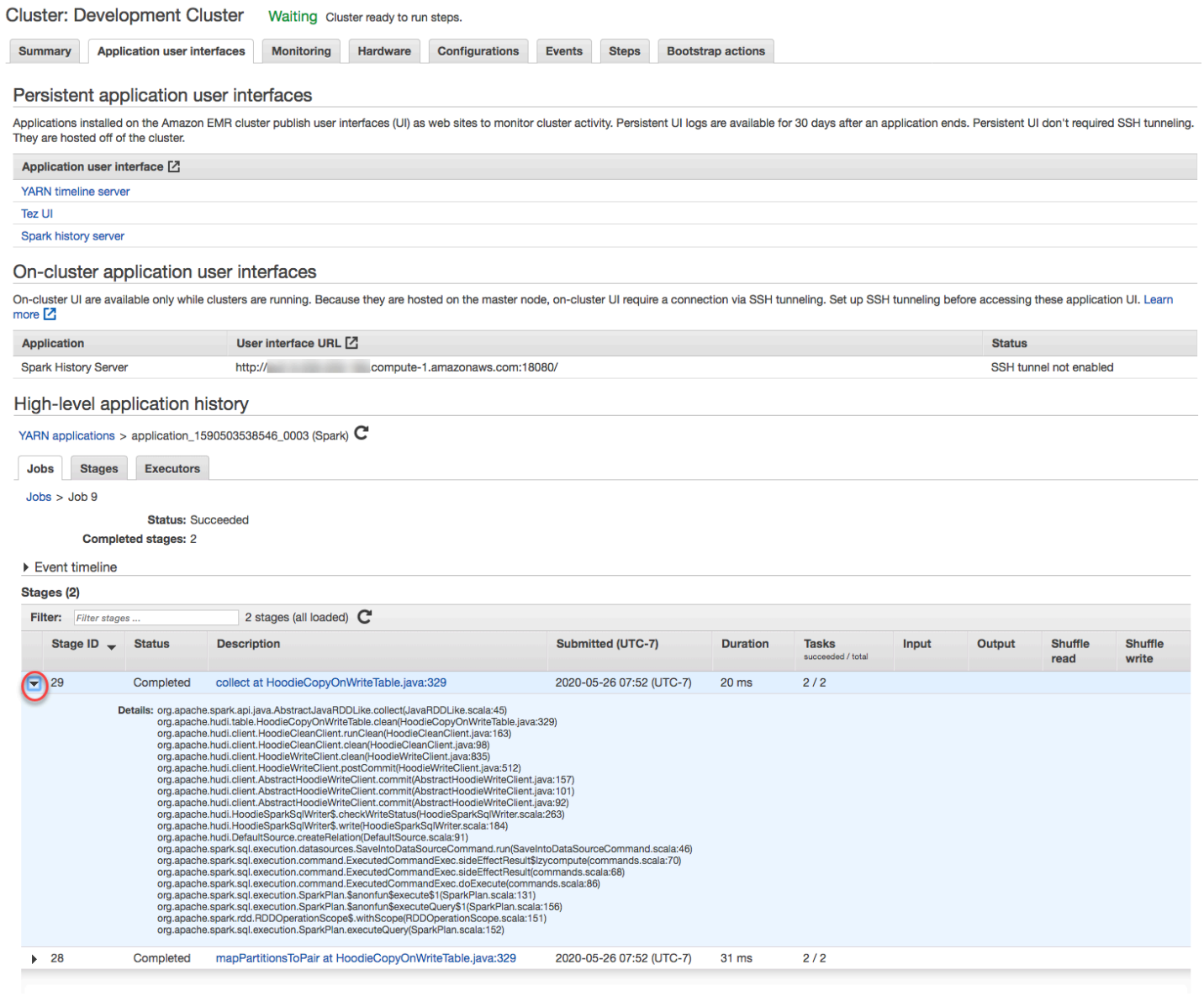
Na guia Interfaces de usuário da aplicação, em Histórico da aplicação de alto nível, você pode expandir uma linha para exibir o resumo do diagnóstico de uma aplicação Spark ou selecionar um link de ID da aplicação para visualizar detalhes sobre outra aplicação.



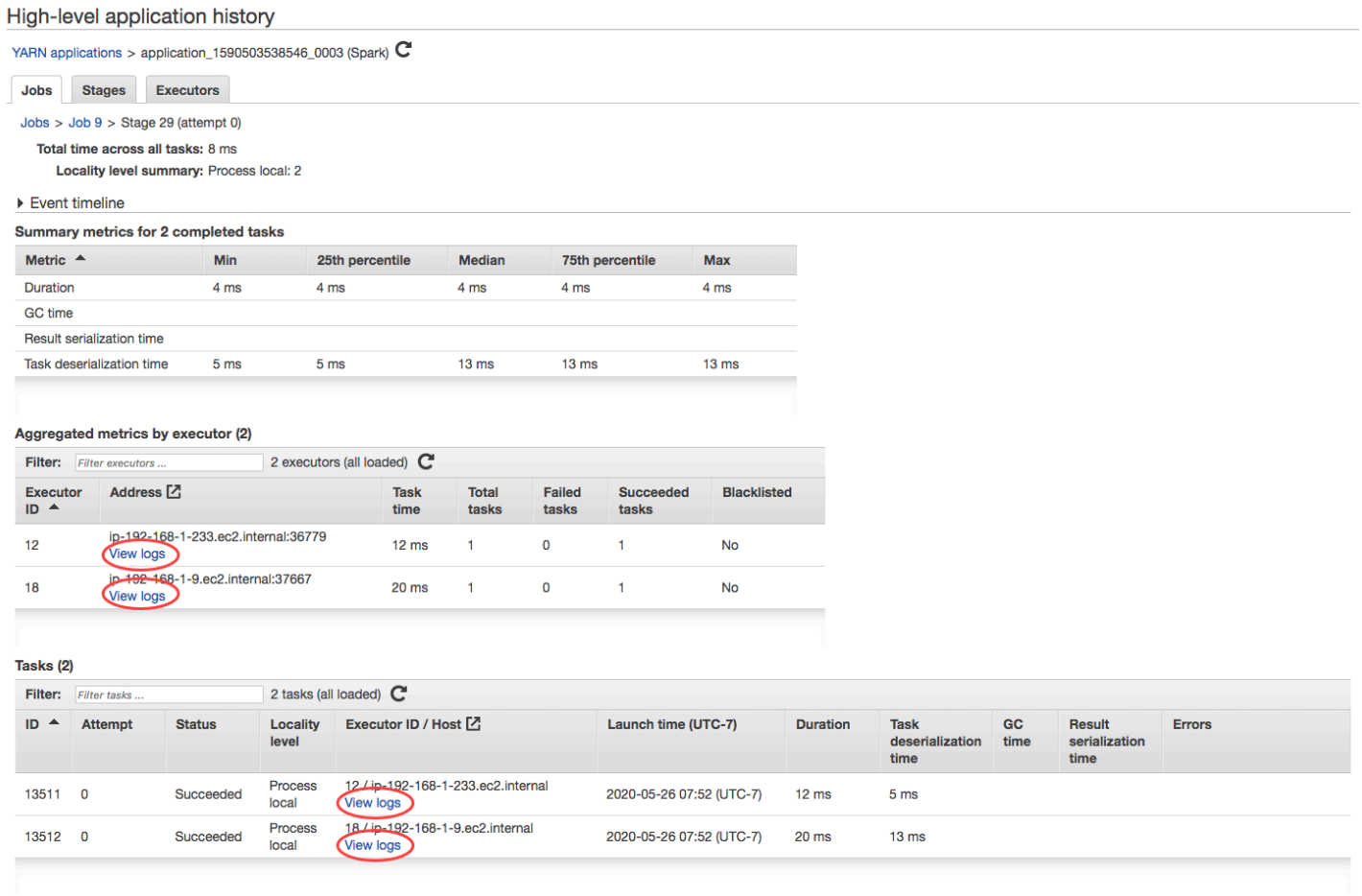
Quando você seleciona um link de ID da aplicação, a interface do usuário passa a exibir os detalhes da aplicação YARN daquela aplicação. Na guia Trabalhos dos detalhes da aplicação YARN, é possível escolher o link Descrição de um trabalho para exibir os detalhes desse trabalho.



Na página de detalhes do trabalho, você pode expandir as informações sobre a preparação do trabalho individual e selecionar o link Descrição para ver os detalhes da preparação.



Na página de detalhes da preparação, você pode visualizar as principais métricas para tarefas e executores do preparação. Também é possível visualizar os logs de tarefa e do executor usando os links Visualizar logs.



Exibir arquivos de log do

Tanto o Amazon EMR como o Hadoop produzem arquivos de log que informam o status no cluster. Por padrão, esses são gravados no nó primário, no diretório `/mnt/var/log/`. Dependendo de como você configurou seu cluster quando o executou, esses logs também podem ser arquivados no Amazon S3 e podem ser visualizados na ferramenta de depuração gráfica.

Há muitos tipos de logs gravados no nó primário. O Amazon EMR grava logs de etapa, ação de bootstrap e estado de instâncias. O Apache Hadoop grava logs para informar o processamento de trabalhos, tarefas e tentativas de tarefas. O Hadoop também registra logs de seus daemons. Para obter mais informações sobre os registros escritos pelo Hadoop, acesse http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/ ClusterSetup .html.

Visualizar arquivos de log no nó primário

A tabela a seguir lista alguns dos arquivos de log que você encontrará no nó primário.

Local	Descrição
/emr/instance-controller/log/bootstrap-actions	Logs gravados durante o processamento das ações de bootstrap.
/mnt/var/log/hadoop-state-pusher	Logs gravados pelo processo de agente de envio de estado do Hadoop.
/emr/instance-controller/log	Logs do controlador de instâncias.
/emr/instance-state	Logs de estado de instância. Contém informaçõ es sobre a CPU, o estado da memória e os threads de coletor de lixo do nó.
/emr/service-nanny	Logs gravados pelo processo nanny de serviço.
/mnt/var/log/*application*	Logs específicos de um aplicativo, como o Hadoop, o Spark ou o Hive.
/mnt/var/log/hadoop/steps/*N*	Logs de etapa que contêm informações sobre o processamento da etapa. O valor de *N* indica o stepId atribuído pelo Amazon EMR. Por exemplo, um cluster tem duas etapas: `s-1234ABCDEFGH` e `s-5678IJKLMNOP`. A primeiro etapa está localizado em `/mnt/var/log/hadoop/steps/s-1234ABCDEFGH/` e segundo etapa, em `/mnt/var/log/hadoop/steps/s-5678IJKLMNOP/`. Os logs de etapas escritos pelo Amazon EMR estão apresentados a seguir. • controller: informações sobre o processam ento da etapa. Se a etapa falhar durante o carregamento, você encontrará o rastreame nto da pilha nesse log.

Local	Descrição
	• syslog: descreve a execução dos trabalhos do Hadoop na etapa. • stderr: o canal de erro padrão do Hadoop enquanto ele processa a etapa. • stdout: o canal de saída padrão do Hadoop enquanto ele processa a etapa.

Para visualizar arquivos de log no nó primário usando a AWS CLI.

1. Use o SSH para conectar-se ao nó primário, conforme descrito em Conectar-se ao nó primário usando SSH.
2. Navegue até o diretório que contém as informações do arquivo de log que você deseja visualizar. A tabela anterior fornece uma lista dos tipos de arquivos de log que estão disponíveis e onde você os encontrará. O exemplo a seguir mostra o comando para navegar até o log de etapas com um ID `s-1234ABCDEFGH`.

   ```
   cd /mnt/var/log/hadoop/steps/s-1234ABCDEFGH/
   ```

3. Use um visualizador de arquivos de sua preferência para visualizar o arquivo de log. O exemplo a seguir usa o comando `less` do Linux para visualizar o arquivo de log `controller`.

   ```
   less controller
   ```

Visualizar arquivos de log arquivados no Amazon S3

Por padrão, os clusters do Amazon EMR iniciados com o uso do console arquivam automaticamente os arquivos de log no Amazon S3. Você pode especificar seu próprio caminho de log ou pode permitir que o console gere automaticamente um caminho de log para você. Para clusters iniciados com o uso da CLI ou da API, você deve configurar o arquivamento de log do Amazon S3 manualmente.

Quando o Amazon EMR está configurado para arquivar arquivos de log no Amazon S3, ele armazena os arquivos no local do S3 que você especificou, na pasta /*cluster-id*/, em que *cluster-id* é o ID do cluster.

A tabela a seguir lista alguns dos arquivos de log que você encontrará no Amazon S3.

Local	Descrição
/*cluster-id* /node/	Logs de nós, incluindo logs de ações de bootstrap, estado da instância e aplicativo para o nó. Os logs para cada nó são armazenados em uma pasta rotulada com o identificador da instância do EC2 desse nó.
/*cluster-id* /node/*instance-id* /*application*	Os logs criados por cada aplicativo ou daemon associado a um aplicativo. Por exemplo, o log do servidor Hive está localizado em *cluster-id* /node/*instance-id* /hive/hive-server.log .
/*cluster-id* /steps/*step-id*/	Logs de etapa que contêm informações sobre o processamento da etapa. O valor de *step-id* indica o ID de etapa atribuído pelo Amazon EMR. Por exemplo, um cluster tem duas etapas: s-1234ABCDEFGH e s-5678IJKLMNOP . A primeiro etapa está localizado em /mnt/var/log/hadoop/steps/s-1234ABCDEFGH/ e segundo etapa, em /mnt/var/log/hadoop/steps/s-5678IJKLMNOP/ . Os logs de etapas escritos pelo Amazon EMR estão apresentados a seguir. • controller: informações sobre o processamento da etapa. Se a etapa falhar durante o carregamento, você encontrará o rastreamento da pilha nesse log. • syslog: descreve a execução dos trabalhos do Hadoop na etapa.

Local	Descrição
	• stderr: o canal de erro padrão do Hadoop enquanto ele processa a etapa. • stdout: o canal de saída padrão do Hadoop enquanto ele processa a etapa.
/*cluster-id* /containers	Logs de contêiner de aplicativo. Os logs para cada aplicativo YARN são armazenados nesses locais.
/*cluster-id* /hadoop-mapreduce/	Os registros que contêm informações sobre detalhes de configuração e histórico de MapReduce trabalhos.

Visualizar os arquivos de log arquivados no Amazon S3 usando o console do Amazon S3

1. Faça login no AWS Management Console e abra o console do Amazon S3 em https://console.aws.amazon.com/s3/.
2. Abra o bucket do S3 especificado quando você configurou o cluster para arquivar arquivos de log no Amazon S3.
3. Navegue até o arquivo de log que contém as informações a serem exibidas. A tabela anterior fornece uma lista dos tipos de arquivos de log que estão disponíveis e onde você os encontrará.
4. Baixe o objeto do arquivo de log para visualizá-lo. Para obter instruções, consulte Fazer download de um objeto.

Visualizar arquivos de log na ferramenta de depuração

O Amazon EMR não habilita automaticamente a ferramenta de depuração. Você deve configurá-la ao executar o cluster. O novo console do Amazon EMR não oferece a ferramenta de depuração.

Visualizar os logs do cluster usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.

2. Na página Lista de clusters, escolha o ícone de detalhes ao lado do cluster que você deseja visualizar.

   Isso abrirá a página Detalhes do cluster. Na seção Etapas, os links à direita de cada etapa exibem os vários tipos de logs disponíveis para a etapa. Esses logs são gerados pelo Amazon EMR.
3. Para visualizar uma lista de trabalhos do Hadoop associados a determinada etapa, escolha o link Visualizar trabalhos à direita da etapa.
4. Para visualizar uma lista de tarefas do Hadoop associadas a determinado trabalho, escolha o link Visualizar tarefas à direita do trabalho.
5. Para visualizar uma lista das tentativas que determinada tarefa executou ao tentar concluir, escolha o link Visualizar tentativas à direita da tarefa.
6. Para visualizar os logs gerados por uma tentativa da tarefa, escolha os links stderr, stdout e syslog à direita da tentativa da tarefa.

A ferramenta de depuração exibe links para os arquivos de log depois que o Amazon EMR os carrega para seu bucket no Amazon S3. Como os arquivos de log são carregados para o Amazon S3 a cada cinco minutos, pode levar alguns minutos para o carregamento desses arquivos terminar após a conclusão da etapa.

O Amazon EMR atualiza periodicamente o status de trabalhos, tarefas e tentativas de tarefa do Hadoop na ferramenta de depuração. Você pode clicar em Atualizar lista nos painéis de depuração para obter o status máximo up-to-date desses itens.

Visualizar instâncias de cluster no Amazon EC2

Para ajudá-lo a gerenciar seus recursos, o Amazon EC2 permite atribuir metadados a recursos no formato de tags. Cada tag do Amazon EC2 consiste em uma chave e um valor. Tags permitem categorizar seus recursos do Amazon EC2 de diferentes maneiras: por exemplo, por finalidade, proprietário ou ambiente.

Você pode pesquisar e filtrar recursos com base nessas tags. As tags que você atribui aos recursos por meio de sua AWS conta estão disponíveis somente para você. Outras contas que compartilham o recurso não podem visualizar suas etiquetas.

O Amazon EMR marca automaticamente cada instância do EC2 que ele executa com pares de chave-valor. As chaves identificam o cluster e o grupo de instâncias ao qual a instância pertence.

Isso facilita a filtragem de instâncias do EC2 para exibir, por exemplo, somente aquelas que pertencem a um determinado cluster ou para exibir todas as instâncias atualmente em execução no grupo de instâncias para a tarefa. Isso é especialmente útil quando você executa vários clusters simultaneamente ou gerencia um grande número de instâncias do EC2.

Estes são os pares de chave-valor predefinidos que o Amazon EMR atribui:

Chave	Valor	Definição de valor
aws:elasticmapreduce:job-flow-id	*job-flow-identifier*	O ID do cluster para o qual a instância está provisionada. Ele é exibido no formato `j-XXXXXXXXXXXXX` e pode conter até 256 caracteres.
aws:elasticmapreduce:instance-group-role	*group-role*	O tipo de grupo de instâncias, inserido como um destes valores: `master`, `core` ou `task`.

Você pode visualizar e filtrar com base nas etiquetas adicionadas pelo Amazon EMR. Para obter mais informações, consulte Uso de tags no Guia do usuário do Amazon EC2. Como as etiquetas definidas pelo Amazon EMR são etiquetas do sistema e não podem ser editadas ou excluídas, as seções sobre como exibir e filtrar etiquetas são as mais relevantes.

> **Note**
>
> O Amazon EMR adiciona etiquetas à instância do EC2 quando seu status é atualizado para Running. Se a latência ocorrer entre o momento em que a instância do EC2 for provisionada e o momento em que seu status for definido como Running, as etiquetas definidas pelo Amazon EMR serão exibidas quando a instância for iniciada. Se você não vir as tags, aguarde alguns minutos e atualize a exibição.

CloudWatch eventos e métricas

Use eventos e métricas para rastrear a atividade e a integridade de um cluster do Amazon EMR. Eventos são úteis para monitorar uma ocorrência específica em um cluster. Por exemplo, quando um cluster muda do estado iniciando para em execução. Métricas são úteis para monitorar um valor

específico, como, por exemplo, a porcentagem de espaço em disco disponível que o HDFS está usando em um cluster.

Para obter mais informações sobre CloudWatch eventos, consulte o Guia do usuário do Amazon CloudWatch Events. Para obter mais informações sobre CloudWatch métricas, consulte Uso de CloudWatch métricas da Amazon e Criação de CloudWatch alarmes da Amazon no Guia do CloudWatch usuário da Amazon.

Tópicos

Monitorando métricas do Amazon EMR com CloudWatch

As métricas são atualizadas a cada cinco minutos e coletadas e enviadas automaticamente CloudWatch para cada cluster do Amazon EMR. Esse intervalo não é configurável. Não há cobrança pelas métricas do Amazon EMR relatadas em. CloudWatch Essas métricas de ponto de dados de cinco minutos são arquivadas por 63 dias, e os dados são descartados após esse período.

Como usar métricas do Amazon EMR?

A tabela a seguir mostra os usos comuns das métricas informadas pelo Amazon EMR. Essas são sugestões para você começar, e não uma lista abrangente. Para obter uma lista completa das métricas relatadas pelo Amazon EMR, consulte Métricas relatadas pelo Amazon EMR em CloudWatch.

Como eu faço para...	Métricas relevantes
Controlar o progresso do meu cluster	Examine as métricas `RunningMapTasks` , `RemainingMapTasks` , `RunningReduceTasks` e `RemainingReduceTasks` .
Detectar clusters ociosos	A métrica `IsIdle` controla se um cluster está ativo, mas não executando tarefas no momento. Você pode definir um alarme a ser disparado quando o cluster permanecer

Como eu faço para...	Métricas relevantes
	ocioso por um determinado tempo, como trinta minutos.
Detectar quando um nó fica sem armazenamento	A métrica `MRUnhealthyNodes` rastreia quando um ou mais nós centrais ou de tarefa ficam sem armazenamento em disco local e fazem a transição para o estado `UNHEALTHY` do YARN. Por exemplo, os nós centrais ou de tarefa estão com pouco espaço em disco e não poderão executar tarefas.
Detectar quando um cluster fica sem armazenamento	A métrica `HDFSUtilization` monitora a capacidade HDFS combinada do cluster e pode exigir o redimensionamento do cluster para adicionar mais nós centrais. Por exemplo, a utilização do HDFS é alta, o que pode afetar os trabalhos e a integridade do cluster.
Detectar quando um cluster está em execução com capacidade reduzida	A métrica `MRLostNodes` rastreia quando um ou mais nós centrais ou de tarefa não conseguem se comunicar com o nó principal . Por exemplo, o nó principal não consegue acessar o nó central ou de tarefa.

Para obter mais informações, consulte O cluster é terminado com NO_SLAVE_LEFT e nós centrais FAILED_BY_MASTER e AWSSupport-AnalyzeEMRLogs.

CloudWatch Métricas de acesso para o Amazon EMR

Você pode visualizar as métricas às quais o Amazon EMR reporta CloudWatch usando o console do Amazon EMR ou o console. CloudWatch Você também pode recuperar métricas usando o `mon-get-stats` comando CloudWatch CLI ou CloudWatch `GetMetricStatistics` a API. Para obter mais informações sobre a visualização ou recuperação de métricas para CloudWatch uso do Amazon EMR, consulte o Guia do usuário da CloudWatch Amazon.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Visualizar métricas usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha o cluster cujas métricas você deseja visualizar. Isso abrirá a página de detalhes do cluster.
3. Selecione a guia Monitoramento da página de detalhes do cluster. Escolha qualquer uma das opções Status do cluster, Status do nó ou Entradas e saídas para carregar os relatórios sobre o progresso e a integridade do cluster.
4. Após escolher uma métrica para visualizar, você poderá aumentar cada grafo. Para filtrar o período de tempo do grafo, selecione uma opção pré-preenchida ou escolha Personalizado.

Old console

Visualizar as métricas usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/.
2. Para visualizar as métricas de um cluster, selecione esse cluster para exibir o painel Summary (Resumo).
3. Escolha Monitoring (Monitoramento) para visualizar informações sobre esse cluster. Escolha qualquer uma das guias Status do cluster, Map/Reduce, Status do nó, ou ES para carregar os relatórios sobre o progresso e a integridade do cluster.
4. Depois que você escolher uma métrica para visualizar, você poderá selecionar um tamanho de gráfico. Edite os campos Start (Iniciar) e End (Finalizar) para filtrar as métricas para um período específico.

Métricas relatadas pelo Amazon EMR em CloudWatch

As tabelas a seguir listam as métricas que o Amazon EMR reporta no console e para as quais envia. CloudWatch

Métricas do Amazon EMR

O Amazon EMR envia dados de várias métricas para. CloudWatch Todos os clusters do Amazon EMR enviam métricas automaticamente em intervalos de cinco minutos. As métricas são arquivadas por duas semanas. Depois desse período, os dados serão descartados.

O namespace `AWS/ElasticMapReduce` inclui as métricas a seguir.

> **Note**
>
> O Amazon EMR extrai métricas de um cluster. Se um cluster torna-se inacessível, nenhuma métrica é relatada até que o cluster fique disponível novamente.

As métricas a seguir estão disponíveis para clusters que executam o Hadoop versões 2.x.

Métrica	Descrição
Status do cluster	
IsIdle	Indica que um cluster não está mais executando nenhum trabalho, mas ainda está ativo e acumulando cobranças. É definido como 1 se nenhuma tarefa ou nenhum trabalho estiver em execução, caso contrário, é definido como 0. Esse valor é verificado em intervalos de 5 minutos, sendo que um valor de 1 indica somente que o cluster estava ocioso no momento da verificação, e não que ele ficou ocioso durante todo o período de 5 minutos. Para evitar falsos positivos, você deve gerar um alerta quando esse valor for 1 em mais de uma verificação consecutiva de 5 minutos. Por exemplo, você pode gerar um alerta para esse valor se ele for 1 por 30 minutos ou mais. Caso de uso: monitorar a performance do cluster

Métrica	Descrição
	Unidade: booliano
ContainerAllocated	O número de contêineres de recursos alocados peloResourceManager. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ContainerReserved	O número de contêineres reservados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ContainerPending	O número de contêineres na fila que ainda não foram alocados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ContainerPendingProporção	A proporção de contêineres pendentes em relação aos contêineres alocados (ContainerPendingRatio = ContainerPending / ContainerAllocated). Se ContainerAllocated = 0, então ContainerPendingRatio =ContainerPending. O valor de ContainerPendingRatio representa um número, não uma porcentagem. Esse valor é útil para escalonar recursos de cluster com base no comportamento de alocação do contêiner. Unidades: Contagem
AppsCompleted	O número de aplicativos enviados para o YARN que foram concluídos. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
AppsFailed	O número de aplicativos enviados para o YARN que apresentaram falha ao concluir. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem
AppsKilled	O número de aplicativos enviados para o YARN que foram interrompidos. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem
AppsPending	O número de aplicativos enviados para o YARN em estado pendente. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
AppsRunning	O número de aplicativos enviados para o YARN que estão em execução. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
AppsSubmitted	O número de aplicativos enviados para o YARN. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
Status do nó	

Métrica	Descrição
CoreNodesCorrendo	O número de nós core em funcionamento. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
CoreNodesPendente	O número de nós core aguardando atribuição. Todos os nós core solicitados podem não estar disponíveis imediatam ente; essa métrica reporta as solicitações pendentes. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
LiveDataNodos	A porcentagem de nós de dados que estão recebendo trabalho do Hadoop. Caso de uso: monitorar a integridade do cluster Unidade: percentual
SR. TotalNodes	O número de nós atualmente disponíveis para MapReduce trabalhos. Equivalente ao `mapred.resourceman ager.TotalNodes` da métrica YARN. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
SR. ActiveNodes	O número de nós que estão executando MapReduce tarefas ou trabalhos no momento. Equivalente ao `mapred.resourcemanager.NoOfActiveNodes` da métrica YARN. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
SR. LostNodes	O número de nós alocados MapReduce que foram marcados no estado LOST. Equivalente ao `mapred.resourcemanager.NoOfLostNodes` da métrica YARN. Caso de uso: monitorar a integridade do cluster, monitorar o progresso do cluster Unidades: Contagem
SR. UnhealthyNodes	O número de nós disponíveis para MapReduce trabalhos marcados em um estado UNHEALTHY. Equivalente ao `mapred.resourcemanager.NoOfUnhealthyNodes` da métrica YARN. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
SR. DecommissionedNodes	O número de nós alocados para MapReduce aplicativos que foram marcados no estado DESCOMISSIONADO. Equivalente ao `mapred.resourcemanager.NoOfDecommissionedNodes` da métrica YARN. Caso de uso: monitorar a integridade do cluster, monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
SR. RebootedNodes	O número de nós disponíveis MapReduce que foram reinicializados e marcados no estado REINICIALIZADO. Equivalente ao `mapred.resourcemanager.NoOf RebootedNodes` da métrica YARN. Caso de uso: monitorar a integridade do cluster, monitorar o progresso do cluster Unidades: Contagem
MultiMasterInstanceGroupNodesRunning	O número de nós principais em execução. Caso de uso: monitorar falhas do nó principal e substituição Unidades: Contagem
MultiMasterInstanceGroupNodesRunningPorcentagem	A porcentagem de nós principais em execução sobre a contagem solicitada de instâncias de nós principais. Caso de uso: monitorar falhas do nó principal e substituição Unidade: percentual
MultiMasterInstanceGroupNodesRequested	O número de nós principais solicitados. Caso de uso: monitorar falhas do nó principal e substituição Unidades: Contagem
IO	
S3 BytesWritten	O número de bytes gravados no Amazon S3. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho no Amazon EMR. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
S3 BytesRead	O número de bytes lidos no Amazon S3. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho no Amazon EMR. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
HDFSUtilization	O percentual de armazenamento do HDFS em uso no momento. Caso de uso: analisar a performance do cluster Unidade: percentual
HDFS BytesRead	O número de bytes lidos do HDFS. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho no Amazon EMR. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
HDFS BytesWritten	O número de bytes gravados no HDFS. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho no Amazon EMR. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
MissingBlocks	O número de blocos em que o HDFS não tem réplicas. Esses podem ser blocos danificados. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
CorruptBlocks	O número de blocos que o HDFS reporta como corrompid os. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
TotalLoad	O número total de transferências simultâneas de dados. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
MemoryTotalMB	A quantidade total de memória no cluster. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MemoryReservedMB	A quantidade de memória reservada. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MemoryAvailableMB	A quantidade de memória disponível para ser alocada. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
Porcentagem de YARN MemoryAvailable	A porcentagem de memória restante disponível para o YARN (YARN MemoryAvailablePercentage = MemoryAvailable MB/MemoryTotalMB). Esse valor é útil para escalonar recursos de cluster com base no uso da memória YARN. Unidade: percentual
MemoryAllocatedMB	A quantidade de memória alocada para o cluster. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
PendingDeletionBlocos	O número de blocos marcados para exclusão. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem
UnderReplicatedBlocos	O número de blocos que precisam ser replicados uma ou mais vezes. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem
DfsPendingReplicationBlocks	O status da replicação de bloco: blocos sendo replicados, idade das solicitações de replicação e solicitações de replicação sem sucesso. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
CapacityRemainingGB	A quantidade de capacidade de disco do HDFS restante. Caso de uso: monitorar o progresso do cluster, monitorar a integridade do cluster Unidades: Contagem

Veja a seguir as métricas do Hadoop 1:

Métrica	Descrição
Status do cluster	
IsIdle	Indica que um cluster não está mais executando nenhum trabalho, mas ainda está ativo e acumulando cobranças. É definido como 1 se nenhuma tarefa ou nenhum trabalho estiver em execução, caso contrário, é definido como 0. Esse valor é verificado em intervalos de 5 minutos, sendo que um valor de 1 indica somente que o cluster estava ocioso no momento da verificação, e não que ele ficou ocioso durante todo o período de 5 minutos. Para evitar falsos positivos, você deve gerar um alerta quando esse valor for 1 em mais de uma verificação consecutiva de 5 minutos. Por exemplo, você pode gerar um alerta para esse valor se ele for 1 por 30 minutos ou mais. Caso de uso: monitorar a performance do cluster Unidade: booliano
JobsRunning	O número de trabalhos no cluster que estão em execução no momento. Caso de uso: monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
JobsFailed	O número de trabalhos no cluster que apresentaram falha. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
Map/Reduce	
MapTasksCorrendo	O número de tarefas de mapeamento em execução para cada trabalho. Se você tiver um programador instalado e vários trabalhos em execução, vários gráficos são gerados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MapTasksRestante	O número de tarefas de mapeamento restantes para cada trabalho. Se você tiver um programador instalado e vários trabalhos em execução, vários gráficos são gerados. Uma tarefa de mapeamento restante não está em nenhum dos seguintes estados: Running, Killed ou Completed. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
MapSlotsAberto	A capacidade não utilizada da tarefa de mapeament o. É calculado como o número máximo de tarefas de mapeamento para um determinado cluster, menos o número total de tarefas de mapeamento em execução no momento nesse cluster. Caso de uso: analisar a performance do cluster Unidades: Contagem

Métrica	Descrição
RemainingMapTasksPerSlot	A razão entre o total de tarefas de mapeamento restantes e o total de slots de mapeamento disponíveis no cluster. Caso de uso: analisar a performance do cluster Unidade: razão
ReduceTasksCorrendo	O número de tarefas de redução em execução para cada trabalho. Se você tiver um programador instalado e vários trabalhos em execução, vários gráficos são gerados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ReduceTasksRestante	O número de tarefas de redução restantes para cada trabalho. Se você tiver um programador instalado e vários trabalhos em execução, vários gráficos são gerados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ReduceSlotsAberto	Capacidade não utilizada das tarefas de redução. É calculado como a capacidade máxima da tarefa de redução para um determinado cluster, menos o número total de tarefas de redução em execução no momento nesse cluster. Caso de uso: analisar a performance do cluster Unidades: Contagem
Status do nó	

Métrica	Descrição
CoreNodesCorrendo	O número de nós core em funcionamento. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
CoreNodesPendente	O número de nós core aguardando atribuição. Todos os nós core solicitados podem não estar disponíveis imediatam ente; essa métrica reporta as solicitações pendentes. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
LiveDataNodos	A porcentagem de nós de dados que estão recebendo trabalho do Hadoop. Caso de uso: monitorar a integridade do cluster Unidade: percentual
TaskNodesCorrendo	O número de nós da tarefa trabalhando. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
TaskNodesPendente	O número de nós de tarefa aguardando atribuição. Todos os nós de tarefa solicitados podem não estar disponíveis imediatamente; essa métrica reporta as solicitações pendentes. Os pontos de dados para essa métrica são relatados somente quando existe um grupo de instâncias correspondente. Caso de uso: monitorar a integridade do cluster Unidades: Contagem
LiveTaskRastreadores	O percentual dos rastreadores de tarefas que estão funcionando. Caso de uso: monitorar a integridade do cluster Unidade: percentual
IO	
S3 BytesWritten	O número de bytes gravados no Amazon S3. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho no Amazon EMR. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
S3 BytesRead	O número de bytes lidos no Amazon S3. Essa métrica agrega somente MapReduce trabalhos e não se aplica a outras cargas de trabalho no Amazon EMR. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
HDFSUtilization	O percentual de armazenamento do HDFS em uso no momento. Caso de uso: analisar a performance do cluster Unidade: percentual
HDFS BytesRead	O número de bytes lidos do HDFS. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
HDFS BytesWritten	O número de bytes gravados no HDFS. Caso de uso: analisar a performance do cluster, monitorar o progresso do cluster Unidades: Contagem
MissingBlocks	O número de blocos em que o HDFS não tem réplicas. Esses podem ser blocos danificados. Caso de uso: monitorar a integridade do cluster Unidades: Contagem

Métrica	Descrição
TotalLoad	O número total atual de leitores e escritores relatados por todos DataNodes em um cluster. Caso de uso: diagnostique até que ponto a alta taxa de E/S pode continuar contribuindo para o desempenho insatisfatório da execução do trabalho. Os nós de trabalho que executam o DataNode daemon também devem realizar tarefas de mapeamento e redução. TotalLoad Valores persistentemente altos ao longo do tempo podem indicar que a alta E/S pode ser um fator que contribui para o baixo desempenho. Os picos ocasionais nesse valor são típicos e geralmente não são indícios de problema. Unidades: Contagem

Métricas de capacidade de cluster

As métricas a seguir indicam as capacidades atuais ou de destino de um cluster. Essas métricas só estão disponíveis quando o ajuste de escala gerenciado ou o término automático estão habilitados.

Para clusters compostos por frotas de instâncias, as métricas de capacidade de cluster são medidas em `Units`. Para clusters compostos por grupos de instâncias, as métricas de capacidade de cluster são medidas em `Nodes` ou `VCPU` com base no tipo de unidade usado na política de escalabilidade gerenciada. Para obter mais informações, consulte Using EMR-managed scaling no Guia de gerenciamento do Amazon EMR.

Métrica	Descrição
• `TotalUnitsRequested` • `TotalNodesRequested` • `TotalVCPURequested`	O número total desejado de unidades/nós/vCPUs em um cluster, conforme determinado pela escalabilidade gerenciada. Unidades: Contagem

Métrica	Descrição
• `TotalUnitsRunning` • `TotalNodesRunning` • `TotalVCPURunning`	O número total atual de unidades/nós/vCPUs disponíveis em um cluster em execução. Quando um redimensionamento de cluster for solicitado, essa métrica será atualizada depois que as novas instâncias forem adicionadas ou removidas do cluster. Unidades: Contagem
• `CoreUnitsRequested` • `CoreNodesRequested` • `CoreVCPURequested`	O número desejado de unidades/nós/vCPUs CORE em um cluster, conforme determinado pela escalabilidade gerenciada. Unidades: Contagem
• `CoreUnitsRunning` • `CoreNodesRunning` • `CoreVCPURunning`	O número atual de unidades/nós/vCPUs CORE em execução em um cluster. Unidades: Contagem
• `TaskUnitsRequested` • `TaskNodesRequested` • `TaskVCPURequested`	O número desejado de unidades/nós/vCPUs TASK em um cluster, conforme determinado pela escalabilidade gerenciada. Unidades: Contagem

Métrica	Descrição
• `TaskUnitsRunning` • `TaskNodesRunning` • `TaskVCPURunning`	O número atual de unidades/nós/vCPUs TASK em execução em um cluster. Unidades: Contagem

O Amazon EMR emite as métricas a seguir em uma granularidade de um minuto quando você habilita o término automático usando uma política de término automático. Algumas métricas estão disponíveis somente para o Amazon EMR 6.4.0 e versões posteriores. Para saber mais sobre término automático, consulte Usar uma política de término automático.

Métrica	Descrição
`TotalNotebookKernels`	O total de kernels de cadernos em execução e ociosos no cluster. Essa métrica está disponível somente para o Amazon EMR 6.4.0 e versões posteriores.
`AutoTerminationIsClusterIdle`	Indica se o cluster está em uso. O valor 0 indica que o cluster está sendo usado ativamente por um destes componentes: • Uma aplicação YARN • HDFS • Um caderno • Uma interface de usuário no cluster, como Spark History Server

Métrica	Descrição
	O valor 1 indica que o cluster está ocioso. O Amazon EMR verifica a ociosidade contínua do cluster (`AutoTerminationIsClusterIdle = 1`). Quando o tempo ocioso de um cluster é igual ao valor `IdleTimeout` na política de término automático, o Amazon EMR termina o cluster.

Dimensões para métricas do Amazon EMR

Os dados do Amazon EMR podem ser filtrados usando qualquer uma das dimensões da tabela a seguir.

Dimensão	Descrição
JobFlowIdentificação	Igual ao ID do cluster, que é o identificador exclusivo de um cluster no formato `j-XXXXXXXXXXXX`. Encontre esse valor clicando no cluster do console do Amazon EMR.

Monitorando eventos do Amazon EMR com CloudWatch

O Amazon EMR controla eventos e mantém as informações sobre eles por até sete dias no console do Amazon EMR. O Amazon EMR registra eventos quando há uma alteração no estado de clusters, grupos de instâncias, frotas de instâncias, políticas de ajuste de escala automático ou etapas. Os eventos capturam a data e a hora em que o evento ocorreu, detalhes sobre os elementos afetados e outros pontos de dados essenciais.

A tabela apresentada a seguir lista os eventos do Amazon EMR em conjunto com o estado ou a alteração de estado que o evento indica, a gravidade do evento, o tipo de evento, o código do evento e as mensagens do evento. O Amazon EMR representa eventos como objetos JSON e os envia automaticamente para um fluxo de eventos. O objeto JSON é importante quando você configura regras para processamento de CloudWatch eventos usando Eventos porque as regras buscam corresponder aos padrões no objeto JSON. Para obter mais informações, consulte Eventos e padrões de eventos e eventos do Amazon EMR no Guia do usuário do Amazon CloudWatch Events.

> **Note**
>
> Para garantir que forneceremos as informações mais pertinentes, refinamos continuamente nossas mensagens de erro. Por isso, não é recomendável analisar o texto das mensagens para iniciar as próximas ações do fluxo de trabalho.

Eventos de início de cluster

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	Provisionamento da frota de instâncias do Amazon EMR	Provisionamento do EC2: capacidade de instância insuficiente	Não foi possível criar o cluster `ClusterId (ClusterName)` do Amazon EMR para a frota de instâncias `InstanceFleetID`. O Amazon EC2 tem capacidade spot insuficiente para o tipo de instância `[Instancetype1, Instancetype2]` e capacidade sob demanda insuficiente para o tipo de instância

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				`[Instance type3, Instancetype4]` na zona de disponibilidade `[AvailabilityZone1, AvaliabilityZone2]`. Confira aqui a documentação para obter mais informações sobre como responder a esse evento.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`CREATING`	`WARN`	Provisionamento do grupo de instâncias do Amazon EMR	Provisionamento do EC2: capacidade de instância insuficiente	Não foi possível criar o cluster `ClusterId (ClusterName)` do Amazon EMR para a grupo de instâncias `InstancegroupID`. O Amazon EC2 tem capacidade `[Spot or On-Demand]` insuficiente para o tipo de instância `Instancetype` na zona de disponibilidade `AvailabilityZone`. Confira aqui a documentação para obter mais informações sobre como responder a esse evento.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`STARTING`	`INFO`	Alteração de estado de clusters do EMR	none	O cluster `ClusterId (ClusterName)` do Amazon EMR foi solicitado às `Time` e está sendo criado.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`STARTING`	`INFO`	Alteração de estado de clusters do EMR	none	**Note** Aplica-se apenas a clusters com a configuração de frotas de instâncias e várias zonas de disponibilidade selecionadas no Amazon EC2. O cluster `ClusterId (ClusterName)` do Amazon EMR está sendo criado na zona (`AvailabilityZoneID`), que foi escolhida entre as opções

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				de zona de disponibilidade.
`STARTING`	`INFO`	Alteração de estado de clusters do EMR	none	O cluster `ClusterId (ClusterName)` do Amazon EMR começou a executar etapas às `Time`.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`WAITING`	`INFO`	Alteração de estado de clusters do EMR	none	O cluster `ClusterId (ClusterName)` do Amazon EMR foi criado às `Time` e está pronto para uso. - ou - O cluster `ClusterId (ClusterName)` do Amazon EMR concluiu a execução de todas as etapas pendentes às `Time`. **Note** Um cluster no estado `WAITING` pode ainda estar processan

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				do trabalhos .

> **Note**
>
> Os eventos com código de evento `EC2 provisioning - Insufficient Instance Capacity` são emitidos periodicamente quando o cluster do EMR encontra um erro de capacidade insuficiente do Amazon EC2 para a frota de instâncias ou grupo de instâncias durante a criação ou operação de redimensionamento do cluster. Para obter informações sobre como responder a esses eventos, consulte Responder eventos de capacidade de instância insuficiente do cluster do Amazon EMR.

Eventos de término de clusters

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`TERMINATED`	A gravidade depende do motivo da mudança de estado, conforme mostrado a seguir: • **`CRITICAL`** se o cluster	Alteração de estado de clusters do EMR	none	O cluster `ClusterId (ClusterName)` do Amazon EMR foi terminado às `Time` pelo motivo `StateChangeReason: Code` .

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
	terminou com qualquer um dos seguintes motivos de mudança de estado: `INTERNAL_ERROR`, `VALIDATION_ERROR`, `INSTANCE_FAILURE`, `BOOTSTRAP_FAILURE` ou `STEP_FAILURE`. • **`INFO`** se o cluster terminou com qualquer um dos seguintes motivos de mudança de estado: `USER_REQUEST` ou `ALL_STEPS_COMPLETED`.			

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`TERMINATED_WITH_ERRORS`	`CRITICAL`	Alteração de estado de clusters do EMR	none	O cluster `ClusterId (ClusterName)` do Amazon EMR foi terminado com erros às `Time` pelo motivo `StateChangeReason: Code` .

Eventos de alteração de estado da frota de instâncias

> **Note**
>
> A configuração de frotas de instância só está disponível em versões do Amazon EMR 4.8.0 e posteriores, exceto versões 5.0.0 e 5.0.3.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
De `PROVISIONING` até `WAITING`	`INFO`		none	O provisionamento da frota de instâncias `InstanceFleetID` no cluster do Amazon EMR foi concluído

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				`ClusterId (ClusterName)` . O provisionamento começou às `Time` e levou `Num` minutos. Agora, a frota de instâncias tem capacidade sob demanda de `Num` e capacidade spot de `Num`. A capacidade sob demanda de destino era `Num`, e a capacidade spot de destino era `Num`.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
De `WAITING` até `RESIZING`	`INFO`		none	Um redimensionamento da frota de instâncias `InstanceFleetID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi iniciado às `Time`. A frota de instâncias está sendo redimensionada de uma capacidade sob demanda de `Num` para um destino de `Num` e de uma capacidade spot de `Num` para um destino de `Num`.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
De `RESIZING` até `WAITING`	`INFO`		none	A operação de redimensionamento da frota de instâncias `InstanceFleetID` no cluster `ClusterId` (`ClusterName`) do Amazon EMR foi concluída. O redimensionamento começou às `Time` e durou `Num` minutos. Agora, a frota de instâncias tem capacidade sob demanda de `Num` e capacidade spot de `Num`. A capacidade sob demanda de destino era `Num`, e a capacidade spot de destino era `Num`.

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
De `RESIZING` até `WAITING`	`INFO`		none	A operação de redimensionamento da frota de instâncias `InstanceFleetID` no cluster `ClusterId` (`ClusterName`) do Amazon EMR atingiu o tempo limite e foi interrompida. O redimensionamento começou às `Time` e foi interrompido após `Num` minutos. Agora, a frota de instâncias tem capacidade sob demanda de `Num` e capacidade spot de `Num`. A capacidade sob demanda de destino era `Num`, e a capacidade

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				spot de destino era Num.
SUSPENDED	ERROR		none	A frota de instâncias InstanceFleetID no cluster ClusterId (ClusterName) do Amazon EMR foi presa às Time pelo seguinte motivo: ReasonDesc .
RESIZING	WARNING		none	A operação de redimensionamento da frota de instâncias InstanceFleetID no cluster ClusterId (ClusterName) do Amazon EMR está paralisada pelo seguinte motivo: ReasonDesc .

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
`WAITING` ou `Running`	`INFO`		none	Não foi possível concluir a operação de redimensionamento da frota de instâncias `InstanceFleetID` no cluster `ClusterId (ClusterName)` do Amazon EMR enquanto o Amazon EMR adicionava capacidade spot à zona de disponibilidade `AvailabilityZone`. Cancelamos sua solicitação para provisionar uma capacidade spot maior. Para ver as ações recomendadas, verifique Práticas recomendadas para flexibilidade de instâncias

Estado ou alteração de estado	Gravidade	Tipo de evento	Código do evento	Message
				e de zona de disponibilidade e tente novamente .
`WAITING` ou `Running`	`INFO`		none	Uma operação de redimensionamento da frota de instâncias `InstanceFleetID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi iniciada por `Entity` às `Time`.

Eventos de redimensionamento da frota de instâncias

Tipo de evento	Gravidade	Código do evento	Message
Redimensionamento da frota de instâncias do Amazon EMR	`ERROR`	Tempo limite de provisionamento spot	Não foi possível concluir a operação de redimensionamento da frota de instâncias `InstanceFleetID` no cluster `ClusterId`

Tipo de evento	Gravidade	Código do evento	Message
			`(ClusterName)` do Amazon EMR durante a aquisição da capacidade spot na AZ `AvailabilityZone` . Já cancelamos a solicitação e paramos de tentar provisionar qualquer capacidade spot adicional, e a frota de instâncias provisionou a capacidade spot de `num`. A capacidade spot de destino era `num`. Para obter mais informações e ações recomendadas, consulte a página de documentação aqui e tente novamente.

Tipo de evento	Gravidade	Código do evento	Message
Redimensionamento da frota de instâncias do Amazon EMR	ERROR	Tempo limite de provisionamento sob demanda	Não foi possível concluir a operação de redimensionamento da frota de instâncias `InstanceFleetID` no cluster `ClusterId (ClusterName)` do Amazon EMR durante a aquisição da capacidade sob demanda na AZ `AvailabilityZone`. Já cancelamos a solicitação e paramos de tentar provisionar qualquer capacidade sob demanda adicional, e a frota de instâncias provisionou a capacidade sob demanda de `num`. A capacidade sob demanda de destino era `num`. Para obter mais informações e ações recomendadas, consulte a página de documentação aqui e tente novamente.

Tipo de evento	Gravidade	Código do evento	Message
Redimensionamento da frota de instâncias do Amazon EMR	`WARNING`	Provisionamento do EC2: capacidade de instância insuficiente	Não foi possível concluir a operação de redimensionamento da frota de instâncias `InstanceFleetID` no cluster `ClusterId (ClusterName)` do EMR, pois o Amazon EC2 tem capacidade spot insuficiente para tipos de instância `[Instancetype1, Instancetype2]` e capacidade sob demanda insuficiente para tipos de instância `[Instancetype3, Instancetype4]` na zona de disponibilidade `[AvailabilityZone1]` . Até agora, a frota de instâncias provisionou a capacidade sob demanda de `num`, e a capacidade sob demanda de destino era `num`. A capacidade spot provisionada é `num`, e a capacidade spot de destino

Tipo de evento	Gravidade	Código do evento	Message
			era num. Confira aqui a documentação para obter mais informações sobre como responder a esse evento.

Tipo de evento	Gravidade	Código do evento	Message
Redimensionamento da frota de instâncias do Amazon EMR	`WARNING`	Tempo limite de provisionamento spot: redimensionamento contínuo	Ainda estamos provisionando a capacidade spot para a operação de redimensionamento da frota de instâncias que foi iniciada às `time` para o ID da frota de instâncias `InstanceFleetID` no cluster `ClusterId (ClusterName)` do Amazon EMR para `[Instancetype1, Instancetype2]` ou na AZ `AvailabilityZone`. Para a operação de redimensionamento anterior iniciada às `time`, o período de tempo limite expirou, então o Amazon EMR parou de provisionar a capacidade spot após adicionar `num` das `num` instâncias solicitadas à frota de instâncias. Para obter mais informações e ações recomendadas,

Tipo de evento	Gravidade	Código do evento	Message
			confira a página de documentação aqui.

Tipo de evento	Gravidade	Código do evento	Message
Redimensionamento da frota de instâncias do Amazon EMR	`WARNING`	Tempo limite de provisionamento sob demanda: redimensionamento contínuo	Ainda estamos provisionando a capacidade sob demanda para a operação de redimensionamento da frota de instâncias que foi iniciada às `time` para o ID da frota de instâncias `InstanceFleetID` no cluster `ClusterId (ClusterName)` do Amazon EMR para `[Instancetype1, Instancetype2]` ou na AZ `AvailabilityZone` . Para a operação de redimensionamento anterior iniciada às `time`, o período de tempo limite expirou, então o Amazon EMR parou de provisionar a capacidade sob demanda após adicionar `num` das `num` instâncias solicitadas à frota de instâncias. Para obter mais informações e

Tipo de evento	Gravidade	Código do evento	Message
			ações recomendadas, confira a página de documentação aqui.

> **Note**
>
> Os eventos de tempo limite de provisionamento são emitidos quando o Amazon EMR interrompe o provisionamento de capacidade spot ou sob demanda da frota após o tempo limite expirar. Para obter informações sobre como responder a esses eventos, consulte Responder a eventos de tempo limite de redimensionamento da frota de instâncias de cluster do Amazon EMR.

Eventos de instâncias de grupos

Tipo de evento	Gravidade	Código do evento	Message
De `RESIZING` até `Running`	`INFO`	none	A operação de redimensionamento do grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi concluída. Agora, ele tem uma contagem de instâncias de `Num`. O redimensionamento começou às `Time` e levou `Num` minutos para ser concluído.
De `RUNNING` até `RESIZING`	`INFO`	none	Um redimensionamento do

Tipo de evento	Gravidade	Código do evento	Message
			grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi iniciado às `Time`. Ele está sendo redimensionado de uma contagem de instâncias de `Num` a `Num`.
`SUSPENDED`	`ERROR`	none	O grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi preso às `Time` pelo seguinte motivo: `ReasonDesc` .
`RESIZING`	`WARNING`	none	A operação de redimensionamento do grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR está paralisada pelo seguinte motivo: `ReasonDesc` .

Tipo de evento	Gravidade	Código do evento	Message
Redimensionamento do grupo de instâncias do Amazon EMR	`WARNING`	Provisionamento do EC2: capacidade de instância insuficiente	Não foi possível concluir a operação de redimensionamento iniciada às `time` para o grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do EMR, pois o Amazon EC2 não tem capacidade `Spot/On Demand` suficiente para o tipo de instância `[Instancetype]` na zona de disponibilidade `[AvailabilityZone1]` . Até agora, o grupo de instâncias tem uma contagem de instâncias em execução de `num`, e a contagem de instâncias solicitadas era `num`. Confira aqui a documentação para obter mais informações sobre como responder a esse evento.

Tipo de evento	Gravidade	Código do evento	Message
De `RUNNING` até `RESIZING`	`INFO`	none	Um redimensionamento do grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi iniciado por `Entity` às `Time`.

> **Note**
>
> Com as versões 5.21.0 e posteriores do Amazon EMR, você pode substituir as configurações de cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o console do Amazon EMR, o AWS Command Line Interface (AWS CLI) ou o AWS SDK. Para obter mais informações, consulte Supplying a Configuration for an Instance Group in a Running Cluster.

A tabela a seguir lista eventos do Amazon EMR para a operação de reconfiguração, juntamente com o estado ou a alteração de estado que cada um indica, a gravidade do evento e as mensagens do evento.

Estado ou alteração de estado	Gravidade	Message
`RUNNING`	`INFO`	Uma reconfiguração do grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi iniciada pelo usuário às `Time`. A versão da configuração solicitada é `Num`.

Estado ou alteração de estado	Gravidade	Message
De RECONFIGURING até Running	INFO	A operação de reconfiguração do grupo de instâncias InstanceGroupID no cluster ClusterId (ClusterName) do Amazon EMR foi concluída. A reconfiguração começou às Time e levou Num minutos para ser concluída. A versão de configuração atual é Num.
De RUNNING até RECONFIGURING em	INFO	Uma reconfiguração para o grupo de instâncias InstanceGroupID no cluster ClusterId (ClusterName) do Amazon EMR foi iniciada às Time. Ela é a configuração do número da versão Num ao número da versão Num.
RESIZING	INFO	A operação de reconfiguração para a versão de configuração Num do grupo de instâncias InstanceGroupID no cluster ClusterId (ClusterName) do Amazon EMR é temporariamente bloqueada às Time porque o grupo de instâncias está em State.

Estado ou alteração de estado	Gravidade	Message
RECONFIGURING	INFO	A operação de redimensionamento para a contagem de instâncias Num do grupo de instâncias InstanceGroupID no cluster ClusterId (ClusterName) do Amazon EMR é temporariamente bloqueada às Time porque o grupo de instâncias está em State.
RECONFIGURING	WARNING	A operação de reconfiguração do grupo de instâncias InstanceGroupID no cluster ClusterId (ClusterName) do Amazon EMR falhou às Time e levou Num minutos para falhar. A versão de configuração com falha é Num.
RECONFIGURING	INFO	As configurações estão sendo revertidas com êxito para o número da versão anterior Num do grupo de instâncias InstanceGroupID no cluster ClusterId (ClusterName) do Amazon EMR às Time. A nova versão de configuração é Num.

Estado ou alteração de estado	Gravidade	Message
De RECONFIGURING até Running	INFO	As configurações foram revertidas com êxito para a versão anterior Num do grupo de instância s InstanceGroupID no cluster ClusterId (ClusterName) do Amazon EMR às Time. A nova versão de configuração é Num.
De RECONFIGURING até SUSPENDED	CRITICAL	Falha ao reverter para a versão com êxito anterior Num do grupo de instância s InstanceGroupID no cluster ClusterId (ClusterName) do Amazon EMR às Time.

Eventos de política do Auto Scaling

Estado ou alteração de estado	Gravidade	Message
PENDING	INFO	Uma política do Auto Scaling foi adicionada para o grupo de instâncias InstanceG roupID no cluster ClusterId (Cluster Name) do Amazon EMR às Time. A política tem um anexo pendente. - ou -

Estado ou alteração de estado	Gravidade	Message
		A política do Auto Scaling para o grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi atualizada às `Time`. A política tem um anexo pendente.
`ATTACHED`	`INFO`	A política do Auto Scaling para o grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi anexada às `Time`.
`DETACHED`	`INFO`	A política do Auto Scaling para o grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR foi desvinculada às `Time`.

Estado ou alteração de estado	Gravidade	Message
FAILED	ERROR	Não foi possível anexar a política do Auto Scaling para o grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR, que falhou às `Time`. - ou - Não foi possível desvincular a política do Auto Scaling para o grupo de instâncias `InstanceGroupID` no cluster `ClusterId (ClusterName)` do Amazon EMR, que falhou às `Time`.

Eventos de etapa

Estado ou alteração de estado	Gravidade	Message
PENDING	INFO	A etapa `StepID (StepName)` foi adicionada ao cluster `ClusterId (ClusterName)` do Amazon EMR às `Time` e está com execução pendente.
CANCEL_PENDING	WARN	A etapa `StepID (StepName)` no cluster `ClusterId (ClusterName)` do Amazon EMR foi cancelada às

Estado ou alteração de estado	Gravidade	Message
		Time e está com cancelamento pendente.
RUNNING	INFO	A etapa StepID (StepName) no cluster ClusterId (ClusterName) do Amazon EMR começou a ser executada às Time.
COMPLETED	INFO	A etapa StepID (StepName) no cluster ClusterId (ClusterName) do Amazon EMR concluiu a execução em Time. A etapa começou a ser executada às Time e levou Num minutos para ser concluída.
CANCELLED	WARN	A solicitação de cancelamento teve êxito na etapa do cluster StepID (StepName) no cluster ClusterId (ClusterName) do Amazon EMR às Time, e a etapa já foi cancelada.
FAILED	ERROR	A etapa StepID (StepName) no cluster ClusterId (ClusterName) do Amazon EMR falhou às Time.

Eventos de substituição de nós não íntegros

Tipo de evento	Gravidade	Código do evento	Message
Substituição de nós não íntegra do Amazon EMR	INFO	Detectado um nó central não íntegro	O Amazon EMR identificou que a instância principal [instanceID (Instance Name)] InstanceGroup/Fleet no cluster do Amazon EMR é. clusterID (ClusterName) UNHEALTHY O Amazon EMR tentará recuperar ou substituir a instância sem problemas. UNHEALTHY
Substituição de nós não íntegra do Amazon EMR	INFO	Nó central não íntegro - substituição desativada	O Amazon EMR identificou que a instância principal [instanceID (Instance Name)] InstanceGroup/Fleet no cluster do

Tipo de evento	Gravidade	Código do evento	Message	
			Amazon EMR é. `{clusterID}` `(ClusterName)` `UNHEALTHY` Ative a substituição normal de nós principais não íntegros em seu cluster para permitir que o Amazon EMR substitua as `UNHEALTHY` instâncias sem problemas, caso elas não possam ser recuperadas.	

Tipo de evento	Gravidade	Código do evento	Message
Substituição de nós não íntegra do Amazon EMR	WARN	O nó central não íntegro não foi substituído	O Amazon EMR não pode substituir sua instância *UNHEALTHY* principal *[instanceID (Instance Name)]* *InstanceGroup/Fleet* no cluster *clusterID (ClusterName)* do Amazon EMR por esse motivo. **Note** O motivo pelo qual o Amazon EMR não pode substituir seu nó principal varia de acordo com seu cenário. Por

Tipo de evento	Gravidade	Código do evento	Message	
			exemplo, uma razão pela qual o Amazon EMR não pode excluir um nó é porque um cluster não teria nenhum nó principal restante.	

Tipo de evento	Gravidade	Código do evento	Message	
Substituição de nós não íntegra do Amazon EMR	INFO	Nó central não íntegro recuperado	O Amazon EMR recuperou suas instâncias UNHEALTHY principais [instanceID (InstanceName)] InstanceGroup/Fleet no cluster do Amazon EMR clusterID (ClusterName)	

Para obter mais informações sobre a substituição de nós não íntegros, consulte Substituindo nós não íntegros.

Visualizar eventos usando o console do Amazon EMR

Para cada cluster, você pode visualizar uma lista simples de eventos no painel de detalhes, que lista os eventos em ordem decrescente de ocorrência. Você também pode visualizar todos os eventos para todos os clusters de uma região em ordem decrescente de ocorrência.

Se não quiser que um usuário veja todos os eventos de cluster para uma região, adicione uma instrução que negue permissão (`"Effect": "Deny"`) para a ação `elasticmapreduce:ViewEventsFromAllClustersInConsole` a uma política anexada a esse usuário.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Visualizar eventos de todos os clusters em uma região usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Eventos.

Visualizar os eventos de um determinado cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha um cluster.
3. Para visualizar todos os seus eventos, selecione a guia Eventos na página de detalhes do cluster.

Old console

Visualizar eventos de todos os clusters em uma região usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha Events (Eventos).

Visualizar os eventos de um determinado cluster usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce/.
2. Escolha Cluster List (Lista de clusters), selecione um cluster e escolha View details (Exibir detalhes).

3. Escolha Events (Eventos) no painel de detalhes do cluster.

Respondendo a eventos CloudWatch

Esta seção descreve várias maneiras de responder a eventos acionáveis que o Amazon EMR emite CloudWatch como mensagens de eventos.

Tópicos

- Criação de regras para eventos do Amazon EMR com CloudWatch
- Configurando alarmes nas métricas CloudWatch
- Responder eventos de capacidade de instância insuficiente do cluster do Amazon EMR
- Responder a eventos de tempo limite de redimensionamento da frota de instâncias de cluster do Amazon EMR

Criação de regras para eventos do Amazon EMR com CloudWatch

O Amazon EMR envia automaticamente eventos para um CloudWatch stream de eventos. Você pode criar regras que correspondem eventos de acordo com um padrão especificado e rotear esses eventos para destinos a fim de realizar ações, como enviar uma notificação por e-mail. Os padrões são correspondidos ao objeto JSON do evento. Para obter mais informações sobre os detalhes dos eventos do Amazon EMR, consulte Eventos do Amazon EMR no Guia do usuário do Amazon CloudWatch Events.

Para obter informações sobre como configurar regras de CloudWatch eventos, consulte Criação de uma CloudWatch regra que é acionada em um evento.

Configurando alarmes nas métricas CloudWatch

O Amazon EMR envia métricas para a Amazon. CloudWatch Em resposta, você pode usar CloudWatch para definir alarmes em suas métricas do Amazon EMR. Por exemplo, você pode configurar um alarme CloudWatch para enviar um e-mail sempre que a utilização do HDFS ultrapassar 80%. Para obter instruções detalhadas, consulte Criar ou editar um CloudWatch alarme no Guia do CloudWatch usuário da Amazon.

Responder eventos de capacidade de instância insuficiente do cluster do Amazon EMR

Visão geral

Os clusters do Amazon EMR retornam o código de evento `EC2 provisioning - Insufficient Instance Capacity` quando a zona de disponibilidade selecionada não tem capacidade suficiente para solucionar a solicitação de inicialização ou redimensionamento do cluster. O evento é emitido periodicamente com grupos de instâncias e frotas de instâncias se o Amazon EMR encontrar repetidamente exceções de capacidade insuficientes e não puder solucionar solicitação de provisionamento para iniciar ou redimensionar o cluster.

Esta página descreve como você pode responder melhor a esse tipo de evento quando ele ocorre no cluster do EMR.

Solução recomendada a um evento de capacidade insuficiente

Recomendamos responder a um evento de capacidade insuficiente de uma das seguintes maneiras:

- Aguarde a recuperação da capacidade. Como a capacidade muda com frequência, uma exceção de capacidade insuficiente pode se recuperar sozinha. Os clusters começarão ou terminarão de ser redimensionados assim que a capacidade do Amazon EC2 estiver disponível.
- Como alternativa, você pode encerrar o cluster, modificar as configurações de tipo de instância e criar um novo cluster com a solicitação de configuração de cluster atualizada. Para ter mais informações, consulte Práticas recomendadas para flexibilidade de instâncias e de zona de disponibilidade.

Também é possível configurar regras ou respostas automatizadas para um evento de capacidade insuficiente, conforme descrito na próxima seção.

Recuperação automatizada de um evento de capacidade insuficiente

É possível criar automação em resposta aos eventos do Amazon EMR, como aqueles com código de evento `EC2 provisioning - Insufficient Instance Capacity`. Por exemplo, a AWS Lambda função a seguir encerra um cluster do EMR com um grupo de instâncias que usa instâncias sob demanda e, em seguida, cria um novo cluster do EMR com um grupo de instâncias que contém tipos de instância diferentes da solicitação original.

Estas condições acionam a ocorrência do processo automatizado:

- O evento de capacidade insuficiente foi emitido para nós primários ou centrais durante mais de 20 minutos.
- O cluster não está no estado READY ou WAITING. Para obter mais informações sobre estados de cluster do EMR, consulte Noções básicas sobre o ciclo de vida do cluster.

> **Note**
>
> Ao criar um processo automatizado para uma exceção de capacidade insuficiente, considere que o evento de capacidade insuficiente é recuperável. A capacidade muda com frequência, e os clusters retomarão o redimensionamento ou iniciarão a operação assim que a capacidade do Amazon EC2 estiver disponível.

Example função para responder ao evento de capacidade insuficiente

```
// Lambda code with Python 3.10 and handler is lambda_function.lambda_handler
// Note: related IAM role requires permission to use Amazon EMR

import json
import boto3
import datetime
from datetime import timezone

INSUFFICIENT_CAPACITY_EXCEPTION_DETAIL_TYPE = "EMR Instance Group Provisioning"
INSUFFICIENT_CAPACITY_EXCEPTION_EVENT_CODE = (
    "EC2 provisioning - Insufficient Instance Capacity"
)
ALLOWED_INSTANCE_TYPES_TO_USE = [
    "m5.xlarge",
    "c5.xlarge",
    "m5.4xlarge",
    "m5.2xlarge",
    "t3.xlarge",
]
CLUSTER_START_ACCEPTABLE_STATES = ["WAITING", "RUNNING"]
CLUSTER_START_SLA = 20

CLIENT = boto3.client("emr", region_name="us-east-1")
```

```
# checks if the incoming event is 'EMR Instance Fleet Provisioning' with eventCode 'EC2
 provisioning - Insufficient Instance Capacity'
def is_insufficient_capacity_event(event):
    if not event["detail"]:
        return False
    else:
        return (
            event["detail-type"] == INSUFFICIENT_CAPACITY_EXCEPTION_DETAIL_TYPE
            and event["detail"]["eventCode"]
            == INSUFFICIENT_CAPACITY_EXCEPTION_EVENT_CODE
        )


# checks if the cluster is eligible for termination
def is_cluster_eligible_for_termination(event, describeClusterResponse):
    # instanceGroupType could be CORE, MASTER OR TASK
    instanceGroupType = event["detail"]["instanceGroupType"]
    clusterCreationTime = describeClusterResponse["Cluster"]["Status"]["Timeline"][
        "CreationDateTime"
    ]
    clusterState = describeClusterResponse["Cluster"]["Status"]["State"]

    now = datetime.datetime.now()
    now = now.replace(tzinfo=timezone.utc)
    isClusterStartSlaBreached = clusterCreationTime < now - datetime.timedelta(
        minutes=CLUSTER_START_SLA
    )

    # Check if instance group receiving Insufficient capacity exception is CORE or
 PRIMARY (MASTER),
    # and it's been more than 20 minutes since cluster was created but the cluster
 state and the cluster state is not updated to RUNNING or WAITING
    if (
        (instanceGroupType == "CORE" or instanceGroupType == "MASTER")
        and isClusterStartSlaBreached
        and clusterState not in CLUSTER_START_ACCEPTABLE_STATES
    ):
        return True
    else:
        return False


# Choose item from the list except the exempt value
def choice_excluding(exempt):
```

```
    for i in ALLOWED_INSTANCE_TYPES_TO_USE:
        if i != exempt:
            return i


# Create a new cluster by choosing different InstanceType.
def create_cluster(event):
    # instanceGroupType cloud be CORE, MASTER OR TASK
    instanceGroupType = event["detail"]["instanceGroupType"]

    # Following two lines assumes that the customer that created the cluster already
 knows which instance types they use in original request
    instanceTypesFromOriginalRequestMaster = "m5.xlarge"
    instanceTypesFromOriginalRequestCore = "m5.xlarge"

    # Select new instance types to include in the new createCluster request
    instanceTypeForMaster = (
        instanceTypesFromOriginalRequestMaster
        if instanceGroupType != "MASTER"
        else choice_excluding(instanceTypesFromOriginalRequestMaster)
    )
    instanceTypeForCore = (
        instanceTypesFromOriginalRequestCore
        if instanceGroupType != "CORE"
        else choice_excluding(instanceTypesFromOriginalRequestCore)
    )

    print("Starting to create cluster...")
    instances = {
        "InstanceGroups": [
            {
                "InstanceRole": "MASTER",
                "InstanceCount": 1,
                "InstanceType": instanceTypeForMaster,
                "Market": "ON_DEMAND",
                "Name": "Master",
            },
            {
                "InstanceRole": "CORE",
                "InstanceCount": 1,
                "InstanceType": instanceTypeForCore,
                "Market": "ON_DEMAND",
                "Name": "Core",
            },
```

```
        ]
    }
    response = CLIENT.run_job_flow(
        Name="Test Cluster",
        Instances=instances,
        VisibleToAllUsers=True,
        JobFlowRole="EMR_EC2_DefaultRole",
        ServiceRole="EMR_DefaultRole",
        ReleaseLabel="emr-6.10.0",
    )

    return response["JobFlowId"]


# Terminated the cluster using clusterId received in an event
def terminate_cluster(event):
    print("Trying to terminate cluster, clusterId: " + event["detail"]["clusterId"])
    response = CLIENT.terminate_job_flows(JobFlowIds=[event["detail"]["clusterId"]])
    print(f"Terminate cluster response: {response}")


def describe_cluster(event):
    response = CLIENT.describe_cluster(ClusterId=event["detail"]["clusterId"])
    return response


def lambda_handler(event, context):
    if is_insufficient_capacity_event(event):
        print(
            "Received insufficient capacity event for instanceGroup, clusterId: "
            + event["detail"]["clusterId"]
        )

        describeClusterResponse = describe_cluster(event)

        shouldTerminateCluster = is_cluster_eligible_for_termination(
            event, describeClusterResponse
        )
        if shouldTerminateCluster:
            terminate_cluster(event)

            clusterId = create_cluster(event)
            print("Created a new cluster, clusterId: " + clusterId)
        else:
```

```
            print(
                "Cluster is not eligible for termination, clusterId: "
                + event["detail"]["clusterId"]
            )

    else:
        print("Received event is not insufficient capacity event, skipping")
```

Responder a eventos de tempo limite de redimensionamento da frota de instâncias de cluster do Amazon EMR

Visão geral

Os clusters do Amazon EMR emitem eventos enquanto executam a operação de redimensionamento para clusters de frotas de instâncias. Os eventos de tempo limite de provisionamento são emitidos quando o Amazon EMR interrompe o provisionamento de capacidade spot ou sob demanda da frota após o tempo limite expirar. O usuário pode configurar a duração do tempo limite como parte das especificações de redimensionamento das frotas de instâncias. Em cenários de redimensionamento consecutivo para a mesma frota de instâncias, o Amazon EMR emite os eventos `Spot provisioning timeout - continuing resize` ou `On-Demand provisioning timeout - continuing resize` quando o tempo limite da operação de redimensionamento atual expira. Em seguida, começa a provisionar capacidade para a próxima operação de redimensionamento da frota.

Responder a eventos de tempo limite de redimensionamento da frota de instâncias

Recomendamos responder a um evento de tempo limite de aprovisionamento de uma das seguintes maneiras:

- Revisite as especificações de redimensionamento e repita a operação de redimensionamento. Como a capacidade muda com frequência, os clusters serão redimensionados com êxito assim que a capacidade do Amazon EC2 estiver disponível. Recomenda-se que os clientes configurem valores mais baixos para a duração do tempo limite dos trabalhos que exigem SLAs mais rigorosos.
- Como alternativa, você pode:
  - Iniciar um novo cluster com tipos de instância diversificados com base nas práticas recomendadas para instâncias e na flexibilidade da zona de disponibilidade ou
  - Iniciar um cluster com capacidade sob demanda

- Para o evento de tempo limite de provisionamento e redimensionamento contínuo, você também pode aguardar o processamento das operações de redimensionamento. O Amazon EMR continuará processando sequencialmente as operações de redimensionamento acionadas para a frota, atendendo às especificações de redimensionamento configuradas.

Também é possível configurar regras ou respostas automatizadas para este evento, conforme descrito na próxima seção.

Recuperação automatizada de um evento de tempo limite de provisionamento

É possível criar automação em resposta aos eventos do Amazon EMR com código de evento `Spot Provisioning timeout`. Por exemplo, a função do AWS Lambda a seguir desativa um cluster do EMR com uma frota de instâncias que usa instâncias spot para nós de tarefa e cria um novo cluster do EMR com uma frota de instâncias que contém tipos de instância mais diversificados do que a solicitação original. Neste exemplo, o evento `Spot Provisioning timeout` emitido para os nós de tarefa acionará a execução da função do Lambda.

Example Exemplo de função para responder ao evento **`Spot Provisioning timeout`**

```
// Lambda code with Python 3.10 and handler is lambda_function.lambda_handler
// Note: related IAM role requires permission to use Amazon EMR

import json
import boto3
import datetime
from datetime import timezone

SPOT_PROVISIONING_TIMEOUT_EXCEPTION_DETAIL_TYPE = "EMR Instance Fleet Resize"
SPOT_PROVISIONING_TIMEOUT_EXCEPTION_EVENT_CODE = (
    "Spot Provisioning timeout"
)

CLIENT = boto3.client("emr", region_name="us-east-1")

# checks if the incoming event is 'EMR Instance Fleet Resize' with eventCode 'Spot
 provisioning timeout'
def is_spot_provisioning_timeout_event(event):
    if not event["detail"]:
        return False
    else:
        return (
```

```
            event["detail-type"] == SPOT_PROVISIONING_TIMEOUT_EXCEPTION_DETAIL_TYPE
            and event["detail"]["eventCode"]
            == SPOT_PROVISIONING_TIMEOUT_EXCEPTION_EVENT_CODE
        )


# checks if the cluster is eligible for termination
def is_cluster_eligible_for_termination(event, describeClusterResponse):
    # instanceFleetType could be CORE, MASTER OR TASK
    instanceFleetType = event["detail"]["instanceFleetType"]

    # Check if instance fleet receiving Spot provisioning timeout event is TASK
    if (instanceFleetType == "TASK"):
        return True
    else:
        return False


# create a new cluster by choosing different InstanceType.
def create_cluster(event):
    # instanceFleetType cloud be CORE, MASTER OR TASK
    instanceFleetType = event["detail"]["instanceFleetType"]

    # the following two lines assumes that the customer that created the cluster
 already knows which instance types they use in original request
    instanceTypesFromOriginalRequestMaster = "m5.xlarge"
    instanceTypesFromOriginalRequestCore = "m5.xlarge"

    # select new instance types to include in the new createCluster request
    instanceTypesForTask = [
        "m5.xlarge",
        "m5.2xlarge",
        "m5.4xlarge",
        "m5.8xlarge",
        "m5.12xlarge"
    ]

    print("Starting to create cluster...")
    instances = {
        "InstanceFleets": [
            {
                "InstanceFleetType":"MASTER",
                "TargetOnDemandCapacity":1,
                "TargetSpotCapacity":0,
```

```
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesFromOriginalRequestMaster,
                        "WeightedCapacity":1,
                    }
                ]
            },
            {
                "InstanceFleetType":"CORE",
                "TargetOnDemandCapacity":1,
                "TargetSpotCapacity":0,
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesFromOriginalRequestCore,
                        "WeightedCapacity":1,
                    }
                ]
            },
            {
                "InstanceFleetType":"TASK",
                "TargetOnDemandCapacity":0,
                "TargetSpotCapacity":100,
                "LaunchSpecifications":{},
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesForTask[0],
                        "WeightedCapacity":1,
                    },
                    {
                        'InstanceType': instanceTypesForTask[1],
                        "WeightedCapacity":2,
                    },
                    {
                        'InstanceType': instanceTypesForTask[2],
                        "WeightedCapacity":4,
                    },
                    {
                        'InstanceType': instanceTypesForTask[3],
                        "WeightedCapacity":8,
                    },
                    {
                        'InstanceType': instanceTypesForTask[4],
                        "WeightedCapacity":12,
                    }
```

```
                ],
                "ResizeSpecifications": {
                    "SpotResizeSpecification": {
                        "TimeoutDurationMinutes": 30
                    }
                }
            }
        ]
    }
    response = CLIENT.run_job_flow(
        Name="Test Cluster",
        Instances=instances,
        VisibleToAllUsers=True,
        JobFlowRole="EMR_EC2_DefaultRole",
        ServiceRole="EMR_DefaultRole",
        ReleaseLabel="emr-6.10.0",
    )

    return response["JobFlowId"]


# terminated the cluster using clusterId received in an event
def terminate_cluster(event):
    print("Trying to terminate cluster, clusterId: " + event["detail"]["clusterId"])
    response = CLIENT.terminate_job_flows(JobFlowIds=[event["detail"]["clusterId"]])
    print(f"Terminate cluster response: {response}")


def describe_cluster(event):
    response = CLIENT.describe_cluster(ClusterId=event["detail"]["clusterId"])
    return response


def lambda_handler(event, context):
    if is_spot_provisioning_timeout_event(event):
        print(
            "Received spot provisioning timeout event for instanceFleet, clusterId: "
            + event["detail"]["clusterId"]
        )

        describeClusterResponse = describe_cluster(event)

        shouldTerminateCluster = is_cluster_eligible_for_termination(
            event, describeClusterResponse
```

```
        )
        if shouldTerminateCluster:
            terminate_cluster(event)

            clusterId = create_cluster(event)
            print("Created a new cluster, clusterId: " + clusterId)
        else:
            print(
                "Cluster is not eligible for termination, clusterId: "
                + event["detail"]["clusterId"]
            )

    else:
        print("Received event is not spot provisioning timeout event, skipping")
```

Visualizar métricas para aplicações de cluster com o Ganglia

O Ganglia está disponível com as versões do Amazon EMR entre 4.2 e 6.15. O Ganglia é um projeto de código aberto que é um sistema distribuído e escalável projetado para monitorar clusters e grades e, ao mesmo tempo, minimizar o impacto no desempenho. Quando você habilita o Ganglia no seu cluster, pode gerar relatórios e visualizar o desempenho do cluster como um todo, bem como inspecionar o desempenho de instâncias de nós individuais. O Ganglia também é configurado para analisar e visualizar as métricas do Hadoop e do Spark. Para obter mais informações, consulte Ganglia no Guia de lançamento do Amazon EMR.

Registro de chamadas de API do Amazon EMR em AWS CloudTrail

O Amazon EMR é integrado com AWS CloudTrail, um serviço que fornece um registro das ações realizadas por um usuário, função ou AWS serviço no Amazon EMR. CloudTrail captura todas as chamadas de API para o Amazon EMR como eventos. As chamadas capturadas incluem as chamadas do console do Amazon EMR e as chamadas de código para as operações da API do Amazon EMR. Se você criar uma trilha, poderá habilitar a entrega contínua de CloudTrail eventos para um bucket do Amazon S3, incluindo eventos para o Amazon EMR. Se você não configurar uma trilha, ainda poderá ver os eventos mais recentes no CloudTrail console no Histórico de eventos. Usando as informações coletadas por CloudTrail, você pode determinar a solicitação que foi feita ao Amazon EMR, o endereço IP a partir do qual a solicitação foi feita, quem fez a solicitação, quando ela foi feita e detalhes adicionais.

Para saber mais sobre isso CloudTrail, consulte o Guia AWS CloudTrail do usuário.

Informações do Amazon EMR em CloudTrail

CloudTrail é ativado em sua AWS conta quando você cria a conta. Quando a atividade ocorre no Amazon EMR, essa atividade é registrada em um CloudTrail evento junto com outros eventos de AWS serviço no histórico de eventos. Você pode visualizar, pesquisar e baixar eventos recentes em sua AWS conta. Para obter mais informações, consulte Visualização de eventos com histórico de CloudTrail eventos.

Para um registro contínuo de eventos em sua AWS conta, incluindo eventos para o Amazon EMR, crie uma trilha. Uma trilha permite CloudTrail entregar arquivos de log para um bucket do Amazon S3. Por padrão, quando você cria uma trilha no console, a trilha se aplica a todas as AWS regiões. A trilha registra eventos de todas as regiões na AWS partição e entrega os arquivos de log ao bucket do Amazon S3 que você especificar. Além disso, você pode configurar outros AWS serviços para analisar e agir com base nos dados de eventos coletados nos CloudTrail registros. Para mais informações, consulte:

- Visão geral da criação de uma trilha
- CloudTrail serviços e integrações suportados
- Configurando notificações do Amazon SNS para CloudTrail
- Recebendo arquivos de CloudTrail log de várias regiões e Recebendo arquivos de CloudTrail log de várias contas

Todas as ações do Amazon EMR são registradas CloudTrail e documentadas na Referência da API do Amazon EMR. Por exemplo, chamadas para o `RunJobFlow` `ListCluster` e `DescribeCluster` as ações geram entradas nos arquivos de CloudTrail log.

Cada entrada de log ou evento contém informações sobre quem gerou a solicitação. As informações de identidade ajudam a determinar:

- Se a solicitação foi feita com credenciais de usuário root ou AWS Identity and Access Management (IAM).
- Se a solicitação foi feita com credenciais de segurança temporárias de um perfil ou de um usuário federado.
- Se a solicitação foi feita por outro AWS serviço.

No caso de um processo, em vez de um usuário criar um cluster, você pode usar o identificador `principalId` para determinar o usuário associado à criação do cluster. Para obter mais informações, consulte o elemento CloudTrail userIdentity.

Exemplo: entradas de arquivo de log do Amazon EMR

Uma trilha é uma configuração que permite a entrega de eventos como arquivos de log para um bucket do Amazon S3 que você especificar. CloudTrail os arquivos de log contêm uma ou mais entradas de log. Um evento representa uma única solicitação de qualquer fonte e inclui informações sobre a ação solicitada, a data e a hora da ação, os parâmetros da solicitação e assim por diante. CloudTrail os arquivos de log não são um rastreamento de pilha ordenado das chamadas públicas de API, portanto, eles não aparecem em nenhuma ordem específica.

O exemplo a seguir mostra uma entrada de CloudTrail registro que demonstra a ação RunJobFlow.


```
{
 "Records": [
 {
         "eventVersion":"1.01",
         "userIdentity":{
            "type":"IAMUser",
            "principalId":"EX_PRINCIPAL_ID",
            "arn":"arn:aws:iam::123456789012:user/temporary-user-xx-7M",
            "accountId":"123456789012",
            "userName":"temporary-user-xx-7M"
         },
         "eventTime":"2018-03-31T17:59:21Z",
         "eventSource":"elasticmapreduce.amazonaws.com",
         "eventName":"RunJobFlow",
         "awsRegion":"us-west-2",
         "sourceIPAddress":"192.0.2.1",
         "userAgent":"aws-sdk-java/unknown-version Linux/xx Java_HotSpot(TM)_64-
Bit_Server_VM/xx",
         "requestParameters":{
            "tags":[
               {
                  "value":"prod",
                  "key":"domain"
               },
               {
                  "value":"us-west-2",
```

```
                  "key":"realm"
               },
               {
                  "value":"VERIFICATION",
                  "key":"executionType"
               }
            ],
            "instances":{
               "slaveInstanceType":"m5.xlarge",
               "ec2KeyName":"emr-integtest",
               "instanceCount":1,
               "masterInstanceType":"m5.xlarge",
               "keepJobFlowAliveWhenNoSteps":true,
               "terminationProtected":false
            },
            "visibleToAllUsers":false,
            "name":"MyCluster",
            "ReleaseLabel":"emr-5.16.0"
         },
         "responseElements":{
            "jobFlowId":"j-2WDJCGEG4E6AJ"
         },
         "requestID":"2f482daf-b8fe-11e3-89e7-75a3d0e071c5",
         "eventID":"b348a38d-f744-4097-8b2a-e68c9b424698"
      },
  ...additional entries
   ]
}
```


Usar ajuste de escala de clusters

Você pode ajustar o número de instâncias do Amazon EC2 disponíveis para um cluster do Amazon EMR automaticamente ou manualmente, em resposta a workloads com demandas variáveis.
Há duas opções para usar a escalabilidade automática. É possível habilitar o Ajuste de Escala Gerenciado do Amazon EMR ou criar uma política personalizada de ajuste de escala automático. A tabela a seguir descreve as diferenças entre as duas opções.

	Ajuste de Escala Gerenciado do Amazon EMR	Escalabilidade automática personalizada
Políticas e regras de escalabilidade	Nenhuma política necessária. O Amazon EMR gerencia a ação de ajuste de escala automático avaliando continuamente as métricas de cluster e tomando decisões otimizadas de ajuste de escala.	É necessário definir e gerenciar as políticas e as regras de ajuste de escala automático, como as condições específicas que acionam ações de ajuste de escala, períodos de avaliação, períodos de esfriamento etc.
Versões do Amazon EMR compatíveis	Amazon EMR versão 5.30.0 e posteriores (exceto Amazon EMR versão 6.0.0)	Amazon EMR 4.0.0 e versões posteriores
Composição de cluster compatível	Grupos de instâncias ou frotas de instâncias	Somente grupos de instâncias
Configuração de limites de escalabilidade	Os limites de escalabilidade são configurados para todo o cluster.	Os limites de escalabilidade só podem ser configurados para cada grupo de instâncias.
Frequência da avaliação de métricas	A cada 5 a 10 segundos A avaliação mais frequente de métricas permite que o Amazon EMR tome decisões mais precisas relacionadas à ao ajuste de escala.	É possível definir os períodos de avaliação apenas em incrementos de cinco minutos.
Aplicações compatíveis	Somente aplicativos do YARN são compatíveis, como Spark, Hadoop, Hive e Flink. O Ajuste de Escala Gerenciado do Amazon EMR não oferece suporte a aplicações que não	Você pode escolher quais aplicativos são compatíveis ao definir as regras de escalabilidade automática.

	Ajuste de Escala Gerenciado do Amazon EMR	Escalabilidade automática personalizada
	sejam baseadas no YARN, como o Presto ou o HBase.	

Considerações

- Um cluster do Amazon EMR sempre consiste em um ou três nós primários. Depois de configurar o cluster inicialmente, você só pode escalar os nós centrais e de tarefas. Você não pode escalar o número de nós primários para o cluster.
- Para grupos de instâncias, as operações de reconfiguração e redimensionamento ocorrem consecutivamente e não simultaneamente. Se você iniciar uma reconfiguração enquanto um grupo de instâncias estiver sendo redimensionado, a reconfiguração será iniciada quando o grupo de instâncias concluir o redimensionamento em andamento. Por outro lado, se você iniciar uma operação de redimensionamento enquanto uma instância agrupa sua reconfiguração.

Usar o ajuste de escala gerenciado no Amazon EMR

> **Important**
>
> É altamente recomendável que você use a versão mais recente do Amazon EMR (Amazon EMR 7.1.0) para escalabilidade gerenciada. Em versões anteriores, você pode enfrentar falhas de aplicações intermitentes ou atrasos no ajuste de escala. O Amazon EMR resolveu esse problema nas versões 5.x: 5.30.2, 5.31.1, 5.32.1, 5.33.1 e posteriores, e nas versões 6.x: 6.1.1, 6.2.1, 6.3.1 e posteriores. Para ter mais informações sobre Regiões e disponibilidade de versões, consulte Disponibilidade gerenciada de ajuste de escala.

Visão geral

Com o Amazon EMR 5.30.0 e versões posteriores (exceto para o Amazon EMR 6.0.0), você pode habilitar o Ajuste de Escala Gerenciado do Amazon EMR. Com o ajuste de escala gerenciado, é possível aumentar ou diminuir automaticamente o número de instâncias ou unidades no cluster com base na workload. O Amazon EMR avalia continuamente as métricas do cluster para tomar decisões

de ajuste de escala que otimizam os clusters em termos de custo e velocidade. O ajuste de escala gerenciado está disponível para clusters compostos por grupos de instâncias ou frotas de instâncias.

Disponibilidade gerenciada de ajuste de escala

- A seguir Regiões da AWS, a escalabilidade gerenciada do Amazon EMR está disponível com o Amazon EMR 6.14.0 e superior:
  - Ásia-Pacífico (Hyderabad) (ap-south-2)
  - Ásia-Pacífico (Jacarta) (ap-southeast-3)
  - Europa (Espanha) (eu-south-2)
- A seguir Regiões da AWS, a escalabilidade gerenciada do Amazon EMR está disponível com o Amazon EMR 5.30.0 e 6.1.0 ou superior:
  - Leste dos EUA (Norte da Virgínia) (us-east-1)
  - Leste dos EUA (Ohio) (us-east-2)
  - Oeste dos EUA (Oregon) (us-west-2)
  - Oeste dos EUA (Norte da Califórnia) (us-west-1)
  - África (Cidade do Cabo) (af-south-1)
  - Ásia-Pacífico (Hong Kong) (ap-east-1)
  - Ásia-Pacífico (Mumbai) (ap-south-1)
  - Ásia-Pacífico (Seul) (ap-northeast-2)
  - Ásia-Pacífico (Singapura) (ap-southeast-1)
  - Ásia-Pacífico (Sydney) (ap-southeast-2)
  - Ásia Pacific (Tóquio) (ap-northeast-1)
  - Canadá (Central) (ca-central-1)
  - América do Sul (São Paulo) (sa-east-1)
  - Europa (Frankfurt) (eu-central-1)
  - Europa (Irlanda) (eu-west-1)
  - Europa (Londres) (eu-west-2)
  - UE (Milão) (eu-south-1)
  - Europa (Paris) (eu-west-3)
  - UE (Estocolmo) (eu-north-1)
  - China (Pequim) (cn-north-1)

  - China (Ningxia) (cn-northwest-1)
  - AWS GovCloud (Leste dos EUA) (us-gov-east-1)
  - AWS GovCloud (Oeste dos EUA) (us-gov-west-1)
- O Ajuste de Escala Gerenciado do Amazon EMR só funciona com aplicações YARN, como Spark, Hadoop, Hive e Flink. Não oferece suporte a aplicações que não sejam baseadas em YARN, como o Presto e HBase.

Parâmetros do ajuste de escala gerenciado

É necessário configurar os parâmetros a seguir para ajuste de escala gerenciado. O limite só se aplica aos nós core e de tarefa. Não é possível escalar o nó primário após a configuração inicial.

- Mínimo (`MinimumCapacityUnits`): o limite inferior da capacidade permitida do EC2 em um cluster. É medido por meio de núcleos ou instâncias da unidade central de processamento virtual (vCPU) para grupos de instâncias. É medido por meio de unidades para frotas de instâncias.
- Máximo (`MaximumCapacityUnits`): o limite superior da capacidade permitida do EC2 em um cluster. É medido por meio de núcleos ou instâncias da unidade central de processamento virtual (vCPU) para grupos de instâncias. É medido por meio de unidades para frotas de instâncias.
- Limite sob demanda (`MaximumOnDemandCapacityUnits`) (opcional): o limite superior da capacidade permitida do EC2 para o tipo de mercado sob demanda em um cluster. Se este parâmetro não for especificado, o valor `MaximumCapacityUnits` será usado como padrão.
  - Esse parâmetro é usado para dividir a alocação de capacidade entre instâncias sob demanda e spot. Por exemplo, se você definir o parâmetro mínimo como duas instâncias, o parâmetro máximo como cem instâncias e o limite sob demanda como dez instâncias, o Ajuste de Escala Gerenciado do Amazon EMR escalará até dez instâncias sob demanda e alocará a capacidade restante para instâncias spot. Para ter mais informações, consulte Cenários de alocação de nós.
- Máximo de nós centrais (`MaximumCoreCapacityUnits`) (opcional): o limite superior da capacidade permitida do EC2 para o tipo de nó central em um cluster. Se este parâmetro não for especificado, o valor `MaximumCapacityUnits` será usado como padrão.
  - Esse parâmetro é usado para dividir a alocação de capacidade entre nós de centrais e de tarefa. Por exemplo, se você definir o parâmetro mínimo como duas instâncias, o máximo como cem instâncias e o nó central máximo como 17 instâncias, o Ajuste de Escala Gerenciado do Amazon EMR escalará até 17 nós principais e alocará as 83 instâncias restantes aos nós de tarefa. Para ter mais informações, consulte Cenários de alocação de nós.

Para obter mais informações sobre parâmetros de ajuste de escala gerenciado, consulte `ComputeLimits`.

Considerações sobre Ajuste de Escala Gerenciado do Amazon EMR

- A escalabilidade gerenciada é suportada em versões limitadas Regiões da AWS e do Amazon EMR. Para ter mais informações, consulte Disponibilidade gerenciada de ajuste de escala.
- Você deve configurar os parâmetros necessários para o Ajuste de Escala Gerenciado do Amazon EMR. Para ter mais informações, consulte Parâmetros do ajuste de escala gerenciado.
- Para usar o ajuste de escala gerenciado, o processo coletor de métricas deve ser capaz de se conectar ao endpoint público da API para o ajuste de escala gerenciado no API Gateway. Se você usar um nome DNS privado com Amazon Virtual Private Cloud, o escalonamento gerenciado não funcionará corretamente. Para garantir que o ajuste de escala gerenciado funcione, é recomendável executar uma das seguintes ações:
  - Remova o endpoint da VPC de interface do API Gateway da Amazon VPC.
  - Siga as instruções em Por que ocorre um erro HTTP 403 Proibido ao conectar APIs do API Gateway de uma VPC? para desabilitar a configuração de nome DNS privado.
  - Em vez disso, inicie o cluster em sua sub-rede privada. Para obter mais informações, consulte o tópico em Sub-redes privadas.
- Se os trabalhos do YARN ficarem intermitentemente lentos durante a redução da escala verticalmente e os logs do YARN Resource Manager mostrarem que a maioria dos nós foram listados como negados durante o período, você poderá ajustar o limite do tempo limite de desativação.

  Reduza `spark.blacklist.decommissioning.timeout` de uma hora para um minuto para disponibilizar o nó para que outros contêineres pendentes continuem o processamento de tarefa.

  Defina também `YARN.resourcemanager.nodemanager-graceful-decommission-timeout-secs` como um valor maior para garantir que o Amazon EMR não force o término do nó enquanto a “Tarefa do Spark” mais longa ainda estiver em execução no nó. O padrão atual é 60 minutos, o que significa que o YARN força o término do contêiner após 60 minutos, quando o nó entra no estado de desativação.

  O exemplo a seguir da linha de log do YARN Resource Manager mostra os nós adicionados ao estado de desativação:

```
2021-10-20 15:55:26,994 INFO
 org.apache.hadoop.YARN.server.resourcemanager.DefaultAMSProcessor
 (IPC Server handler 37 on default port 8030): blacklist are updated in
 Scheduler.blacklistAdditions: [ip-10-10-27-207.us-west-2.compute.internal,
 ip-10-10-29-216.us-west-2.compute.internal, ip-10-10-31-13.us-
west-2.compute.internal, ... , ip-10-10-30-77.us-west-2.compute.internal],
 blacklistRemovals: []
```

Veja mais detalhes sobre como o Amazon EMR se integra à lista de negação do YARN durante o desativação de nós, casos em que nós no Amazon EMR podem ser listados como negados e como configurar o comportamento de desativação de nós do Spark.

- A utilização excessiva dos volumes do EBS pode causar problemas de ajuste de escala gerenciado. É recomendável manter o volume do EBS abaixo de 90% de utilização. Para ter mais informações, consulte Armazenamento de instâncias.
- CloudWatch As métricas da Amazon são essenciais para a operação da escalabilidade gerenciada do Amazon EMR. Recomendamos que você monitore de perto CloudWatch as métricas da Amazon para garantir que os dados não estejam ausentes. Para obter mais informações sobre como você pode configurar CloudWatch alarmes para detectar métricas ausentes, consulte Usando CloudWatch alarmes da Amazon.
- As operações de ajuste de escala gerenciado nos clusters das versões 5.30.0 e 5.30.1 sem o Presto instalado podem causar falhas na aplicação ou fazer com que um grupo de instâncias ou uma frota de instâncias uniforme permaneça no estado `ARRESTED`, sobretudo quando uma operação de redução da escala verticalmente logo é seguida por uma operação de aumento da escala verticalmente.

  Como solução alternativa, escolha o Presto como uma aplicação a ser instalada ao criar um cluster com as versões 5.30.0 e 5.30.1 do Amazon EMR, mesmo que o trabalho não exija o Presto.
- Ao definir o nó central máximo e o limite sob demanda para o Ajuste de Escala Gerenciado do Amazon EMR, leve em consideração as diferenças entre grupos de instâncias e frotas de instâncias. Cada grupo de instâncias consiste no mesmo tipo de instância e na mesma opção de compra para instâncias: sob demanda ou spot. Para cada frota de instâncias, você pode especificar até cinco tipos de instâncias, que podem ser configurados como instâncias sob demanda e spot. Para obter mais informações, consulte Create a cluster with instance fleets or uniform instance groups, Instance fleet options e Cenários de alocação de nós.
- Com o Amazon EMR 5.30.0 e versões posteriores, ao remover a regra de saída Permitir tudo padrão para 0.0.0.0/ para o grupo de segurança principal, você deverá adicionar uma regra

que permita a conectividade TCP de saída ao grupo de segurança para acesso ao serviço na porta 9443. O grupo de segurança para acesso ao serviço também deve permitir tráfego TCP de entrada na porta 9443 do grupo de segurança principal. Para obter mais informações sobre como configurar grupos de segurança, consulte Amazon EMR-managed security group for the primary instance (private subnets).

- O ajuste de escala gerenciado não é compatível com o atributo de rótulos de nós do YARN. Evite usar rótulos de nós em clusters com ajuste de escala gerenciado. Por exemplo, não permita que os executores sejam executados somente em nós de tarefa. Ao usar rótulos de nós em clusters do Amazon EMR, você pode descobrir que seu cluster não está aumentando a escala verticalmente, o que pode deixar sua aplicação mais lenta.
- Você pode usar AWS CloudFormation para configurar a escalabilidade gerenciada do Amazon EMR. Para obter mais informações, consulte AWS::EMR::Clustero Guia AWS CloudFormation do usuário.

Histórico de recursos

Esta tabela lista as atualizações na funcionalidade de ajuste de escala gerenciado do Amazon EMR.

Data de lançamento	Recurso	Versões do Amazon EMR
31 de março de 2024	O escalonamento gerenciado está disponível na região `ap-south-2` Ásia-Pacífico (Hyderabad).	6.14.0 e posterior
13 de fevereiro de 2024	O escalonamento gerenciado está disponível na região `eu-south-2` Europa (Espanha).	6.14.0 e posterior
10 de outubro de 2023	Ajuste de Escala Gerenciado disponível na região `ap-southeast-3` Ásia-Pacífico (Jacarta).	6.14.0 e posterior
28 de julho de 2023	O ajuste de escala gerenciado foi aprimorado para alternar para um grupo de instância	5.34.0 e posteriores, 6.4.0 e posteriores

Data de lançamento	Recurso	Versões do Amazon EMR
	s de tarefa diferente ao aumentar a escala verticalm ente quando o Amazon EMR sofre um atraso ao aumentar a escala verticalmente com o grupo de instâncias atual.	
16 de junho de 2023	O ajuste de escala gerenciado foi aprimorado para reconhece r os nós que executam a aplicação principal, de forma que esses nós não sejam reduzidos. Para ter mais informações, consulte Noções básicas da estratégia e dos cenários de alocação de nós.	5.34.0 e posteriores, 6.4.0 e posteriores

Data de lançamento	Recurso	Versões do Amazon EMR
21 de março de 2022	Foi adicionado o reconhecimento de dados de shuffle do Spark usado ao reduzir a escala verticalmente de clusters. Para clusters do Amazon EMR com o Apache Spark e o atributo de ajuste de escala gerenciado habilitado, o Amazon EMR monitora continuamente os executores do Spark e os locais intermediários de dados de shuffle. Com essas informações, o Amazon EMR reduz a escala verticalmente apenas das instâncias subutilizadas que não contêm dados de shuffle usados ativamente. Isso evita o recálculo de dados de shuffle perdidos, ajudando a reduzir custos e melhorar a performance do trabalho. Para obter mais informações, consulte o Spark Programming Guide.	5.34.0 e posteriores, 6.4.0 e posteriores

Configurar o ajuste de escala gerenciado para o Amazon EMR

As seções a seguir explicam como iniciar um cluster do EMR que usa escalabilidade gerenciada com o AWS Management Console AWS SDK for Java, o ou o. AWS Command Line Interface

Tópicos

- Use o AWS Management Console para configurar o escalonamento gerenciado
- Use o AWS CLI para configurar o escalonamento gerenciado

- Use AWS SDK for Java para configurar o escalonamento gerenciado

Use o AWS Management Console para configurar o escalonamento gerenciado

Você pode usar o console do Amazon EMR para configurar o ajuste de escala gerenciado ao criar um cluster ou para alterar uma política de ajuste de escala gerenciado para um cluster em execução.

New console

Configurar o ajuste de escala gerenciado ao criar um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Escolha uma versão emr-5.30.0 ou posterior do Amazon EMR, exceto a versão emr-6.0.0.
4. Em Opção de ajuste de escala e provisionamento de clusters, escolha Usar ajuste de escala gerenciado pelo EMR. Especifique o número mínimo e máximo de instâncias, o máximo de instâncias do nó central e o máximo de instâncias sob demanda.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

Configurar o ajuste de escala gerenciado em um cluster já existente usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar.
3. Na guia Instâncias da página de detalhes do cluster, encontre a seção Configurações do grupo de instâncias. Na seção Editar ajuste de escala do cluster, especifique novos valores para os números Mínimo e Máximo de instâncias e o limite Sob demanda.

Old console

Ao criar um cluster no console antigo, é possível configurar o ajuste de escala gerenciado usando a opção rápida ou as opções avançadas de configuração de cluster. Também é possível criar ou

alterar uma política de ajuste de escala gerenciado de um cluster em execução modificando as configurações de Ajuste de escala gerenciado na página Resumo ou Hardware.

Usar opções rápidas para configurar a ajuste de escala gerenciado ao criar um cluster usando o console antigo

1. Abra o console do Amazon EMR, escolha Criar cluster e abra Criar cluster: opções rápidas.
2. Na seção Configuração de hardware ao lado da Opção de ajuste de escala e provisionamento de clusters, marque a caixa de seleção para habilitar escalar nós do cluster com base na workload.
3. Em Unidades centrais e de tarefa, especifique o número Mínimo e o Máximo de instâncias centrais e de tarefa.

Usar a opção avançada para configurar o ajuste de escala gerenciado ao criar um cluster usando o console antigo

1. No console do Amazon EMR, selecione Criar cluster, selecione Ir para opções avançadas, escolha as opções em Etapa 1: software e etapas e vá para Etapa 2: configuração do hardware.
2. Na seção Composição do cluster, selecione Frotas de instâncias ou Grupos de instâncias uniformes.
3. Em Opção de ajuste de escala e provisionamento de clusters, selecione Habilitar ajuste de escala de clusters. Selecione Usar ajuste de escala gerenciado do EMR. Em Unidades centrais e de tarefa, especifique o número mínimo e máximo de instâncias ou unidades da frota de instâncias, o limite sob demanda e a contagem máxima de nós centrais.

   Para clusters compostos por grupos de instâncias, também é possível escolher Criar uma política de escalabilidade automática personalizada se quiser definir políticas de escalabilidade automática personalizadas para cada grupo de instâncias. Para ter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.

Modificar o ajuste de escala gerenciado em um cluster já existente usando o console antigo

1. Abra o console do Amazon EMR, selecione o cluster na lista de clusters e escolha a guia Hardware.

2. Na seção Opção de ajuste de escala e provisionamento de clusters, selecione Editar para Ajuste de Escala Gerenciado do Amazon EMR.
3. Na seção Opção de ajuste de escala e provisionamento de clusters, especifique novos valores para os números Mínimo e Máximo de instâncias e o Limite sob demanda.

Use o AWS CLI para configurar o escalonamento gerenciado

Você pode usar AWS CLI comandos do Amazon EMR para configurar a escalabilidade gerenciada ao criar um cluster. Você pode usar uma sintaxe abreviada, especificando a configuração do JSON nas linhas dos comandos relevantes, ou pode fazer referência a um arquivo que contém a configuração do JSON. Também é possível aplicar uma política de escalabilidade gerenciada a um cluster existente e remover uma política de escalabilidade gerenciada que foi aplicada anteriormente. Além disso, você pode recuperar os detalhes da configuração de uma política de escalabilidade de um cluster em execução.

Habilitar a escalabilidade gerenciada durante a execução do cluster

É possível habilitar a escalabilidade gerenciada durante a execução do cluster, conforme demonstra o exemplo a seguir.

```
aws emr create-cluster \
 --service-role EMR_DefaultRole \
 --release-label emr-7.1.0 \
 --name EMR_Managed_Scaling_Enabled_Cluster \
 --applications Name=Spark Name=Hbase \
 --ec2-attributes KeyName=keyName,InstanceProfile=EMR_EC2_DefaultRole \
 --instance-groups InstanceType=m4.xlarge,InstanceGroupType=MASTER,InstanceCount=1
 InstanceType=m4.xlarge,InstanceGroupType=CORE,InstanceCount=2 \
 --region us-east-1 \
 --managed-scaling-policy
 ComputeLimits='{MinimumCapacityUnits=2,MaximumCapacityUnits=4,UnitType=Instances}'
```

Você também pode especificar uma configuração de política gerenciada usando a managed-scaling-policy opção -- ao usar`create-cluster`.

Aplicar uma política de escalabilidade gerenciada a um cluster existente

É possível aplicar uma política de escalabilidade gerenciada a um cluster existente, conforme demonstra o exemplo a seguir.

```
aws emr put-managed-scaling-policy
--cluster-id j-123456
--managed-scaling-policy ComputeLimits='{MinimumCapacityUnits=1,
MaximumCapacityUnits=10,  MaximumOnDemandCapacityUnits=10, UnitType=Instances}'
```

Também é possível aplicar uma política de escalabilidade gerenciada a um cluster existente usando o comando `aws emr put-managed-scaling-policy`. O exemplo a seguir usa uma referência a um arquivo JSON `managedscaleconfig.json`, que especifica a configuração da política de escalabilidade gerenciada.

```
aws emr put-managed-scaling-policy --cluster-id j-123456 --managed-scaling-policy
 file://./managedscaleconfig.json
```

O exemplo a seguir mostra o conteúdo do arquivo `managedscaleconfig.json`, que define a política de escalabilidade gerenciada.

```
{
    "ComputeLimits": {
        "UnitType": "Instances",
        "MinimumCapacityUnits": 1,
        "MaximumCapacityUnits": 10,
        "MaximumOnDemandCapacityUnits": 10
    }
}
```

Recuperar uma configuração de política de escalabilidade gerenciada

O comando `GetManagedScalingPolicy` recupera a configuração da política. Por exemplo, o comando a seguir recupera a configuração de um cluster com o ID `j-123456`.

```
aws emr get-managed-scaling-policy --cluster-id j-123456
```

Este comando gera o seguinte exemplo de saída.

```
{
   "ManagedScalingPolicy": {
      "ComputeLimits": {
         "MinimumCapacityUnits": 1,
         "MaximumOnDemandCapacityUnits": 10,
         "MaximumCapacityUnits": 10,
         "UnitType": "Instances"
```

```
      }
   }
}
```

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Remover a política de escalabilidade gerenciada

O comando `RemoveManagedScalingPolicy` remove a configuração da política. Por exemplo, o comando a seguir remove a configuração de um cluster com o ID `j-123456`.

```
aws emr remove-managed-scaling-policy --cluster-id j-123456
```

Use AWS SDK for Java para configurar o escalonamento gerenciado

O trecho do programa a seguir mostra como configurar a escalabilidade gerenciada usando o AWS SDK for Java:

```
package com.amazonaws.emr.sample;

import java.util.ArrayList;
import java.util.List;

import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.Application;
import com.amazonaws.services.elasticmapreduce.model.ComputeLimits;
import com.amazonaws.services.elasticmapreduce.model.ComputeLimitsUnitType;
import com.amazonaws.services.elasticmapreduce.model.InstanceGroupConfig;
import com.amazonaws.services.elasticmapreduce.model.JobFlowInstancesConfig;
import com.amazonaws.services.elasticmapreduce.model.ManagedScalingPolicy;
import com.amazonaws.services.elasticmapreduce.model.RunJobFlowRequest;
import com.amazonaws.services.elasticmapreduce.model.RunJobFlowResult;

public class CreateClusterWithManagedScalingWithIG {

 public static void main(String[] args) {
```

```
 AWSCredentials credentialsFromProfile = getCreadentials("AWS-Profile-Name-Here");

 /**
  * Create an Amazon EMR client with the credentials and region specified in order to
create the cluster
  */
 AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
  .withCredentials(new AWSStaticCredentialsProvider(credentialsFromProfile))
  .withRegion(Regions.US_EAST_1)
  .build();

 /**
  * Create Instance Groups - Primary, Core, Task
  */
 InstanceGroupConfig instanceGroupConfigMaster = new InstanceGroupConfig()
   .withInstanceCount(1)
   .withInstanceRole("MASTER")
   .withInstanceType("m4.large")
   .withMarket("ON_DEMAND");

 InstanceGroupConfig instanceGroupConfigCore = new InstanceGroupConfig()
  .withInstanceCount(4)
  .withInstanceRole("CORE")
  .withInstanceType("m4.large")
  .withMarket("ON_DEMAND");

 InstanceGroupConfig instanceGroupConfigTask = new InstanceGroupConfig()
  .withInstanceCount(5)
  .withInstanceRole("TASK")
  .withInstanceType("m4.large")
  .withMarket("ON_DEMAND");

 List<InstanceGroupConfig> igConfigs = new ArrayList<>();
 igConfigs.add(instanceGroupConfigMaster);
 igConfigs.add(instanceGroupConfigCore);
 igConfigs.add(instanceGroupConfigTask);

       /**
        *  specify applications to be installed and configured when Amazon EMR creates
the cluster
        */
 Application hive = new Application().withName("Hive");
 Application spark = new Application().withName("Spark");
 Application ganglia = new Application().withName("Ganglia");
```

```
 Application zeppelin = new Application().withName("Zeppelin");

 /**
  * Managed Scaling Configuration -
        * Using UnitType=Instances for clusters composed of instance groups
  *
        * Other options are:
        * UnitType = VCPU ( for clusters composed of instance groups)
        * UnitType = InstanceFleetUnits ( for clusters composed of instance fleets)
        **/
 ComputeLimits computeLimits = new ComputeLimits()
   .withMinimumCapacityUnits(1)
   .withMaximumCapacityUnits(20)
   .withUnitType(ComputeLimitsUnitType.Instances);

 ManagedScalingPolicy managedScalingPolicy = new ManagedScalingPolicy();
 managedScalingPolicy.setComputeLimits(computeLimits);

 // create the cluster with a managed scaling policy
 RunJobFlowRequest request = new RunJobFlowRequest()
         .withName("EMR_Managed_Scaling_TestCluster")
         .withReleaseLabel("emr-7.1.0")          // Specifies the version label for
the Amazon EMR release; we recommend the latest release
         .withApplications(hive,spark,ganglia,zeppelin)
         .withLogUri("s3://path/to/my/emr/logs")  // A URI in S3 for log files is
required when debugging is enabled.
         .withServiceRole("EMR_DefaultRole")      // If you use a custom IAM service
role, replace the default role with the custom role.
         .withJobFlowRole("EMR_EC2_DefaultRole")  // If you use a custom Amazon EMR
role for EC2 instance profile, replace the default role with the custom Amazon EMR
role.
         .withInstances(new JobFlowInstancesConfig().withInstanceGroups(igConfigs)
             .withEc2SubnetId("subnet-123456789012345")
             .withEc2KeyName("my-ec2-key-name")
             .withKeepJobFlowAliveWhenNoSteps(true))
         .withManagedScalingPolicy(managedScalingPolicy);
    RunJobFlowResult result = emr.runJobFlow(request);

    System.out.println("The cluster ID is " + result.toString());
}

public static AWSCredentials getCredentials(String profileName) {
 // specifies any named profile in .aws/credentials as the credentials provider
 try {
```

```
   return new ProfileCredentialsProvider("AWS-Profile-Name-Here")
     .getCredentials();
        } catch (Exception e) {
            throw new AmazonClientException(
                    "Cannot load credentials from .aws/credentials file. " +
                    "Make sure that the credentials file exists and that the profile
 name is defined within it.",
                    e);
        }
 }

 public CreateClusterWithManagedScalingWithIG() { }
}
```

Noções básicas da estratégia e dos cenários de alocação de nós

Esta seção fornece uma visão geral da estratégia de alocação de nós e dos cenários comuns de ajuste de escala que você pode usar com o Ajuste de Escala Gerenciado do Amazon EMR.

Estratégia de alocação de nós

O Ajuste de Escala Gerenciado do Amazon EMR aloca nós centrais e de tarefa com base nas seguintes estratégias de aumento e redução da escala verticalmente:

Estratégia de aumento da escala verticalmente

- O Ajuste de Escala Gerenciado do Amazon EMR primeiro adiciona capacidade aos nós centrais e depois aos nós de tarefa até atingir a capacidade máxima permitida ou até alcançar a capacidade desejada de aumento da escala verticalmente.
- Quando o Amazon EMR sofre um atraso ao aumentar a escala verticalmente com o grupo de instâncias atual, os clusters que usam o ajuste de escala gerenciado alternam automaticamente para outro grupo de instâncias de tarefa.
- Se o parâmetro `MaximumCoreCapacityUnits` estiver definido, o Amazon EMR escalará os nós centrais até que as unidades centrais atinjam o limite máximo permitido. Toda a capacidade restante é adicionada aos nós de tarefa.
- Se o parâmetro `MaximumOnDemandCapacityUnits` estiver definido, o Amazon EMR escalará o cluster usando as instâncias sob demanda até que as unidades sob demanda atinjam o limite máximo permitido. Toda a capacidade restante é adicionada usando instâncias spot.
- Se os parâmetros `MaximumCoreCapacityUnits` e `MaximumOnDemandCapacityUnits` estiverem definidos, o Amazon EMR considerará os dois limites durante o ajuste de escala.

Por exemplo, se `MaximumCoreCapacityUnits` for menor que `MaximumOnDemandCapacityUnits`, o Amazon EMR primeiro escala os nós centrais até atingir o limite de capacidade do núcleo. Para a capacidade restante, o Amazon EMR primeiro usa instâncias sob demanda para escalar nós de tarefa até atingir o limite sob demanda e usa instâncias spot para nós de tarefa.

Estratégia de redução da escala verticalmente

- O Amazon EMR 5.34.0 e versões posteriores e o Amazon EMR 6.4.0 e versões posteriores oferecem suporte ao ajuste de escala gerenciado que reconhece os dados de shuffle do Spark (dados que o Spark redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações de shuffle, consulte o Guia de programação do Spark. O ajuste de escala gerenciado reduz somente as instâncias que são subutilizadas e que não contêm dados de shuffle usados ativamente. Esse ajuste de escala inteligente evita a perda não intencional de dados de shuffle, evitando a necessidade de novas tentativas de trabalho e recálculo de dados intermediários.
- O Ajuste de Escala Gerenciado do Amazon EMR primeiro remove os nós de tarefa e depois remove os nós centrais até alcançar a capacidade desejada de redução da escala verticalmente. O cluster jamais escala abaixo das restrições mínimas na política de ajuste de escala gerenciado.
- Em cada tipo de nó (nós centrais ou nós de tarefa), o Ajuste de Escala Gerenciado do Amazon EMR remove primeiro as instâncias spot e depois remove as instâncias sob demanda.
- Para clusters que são iniciados com o Amazon EMR 5.x versões 5.34.0 e posteriores e 6.x versões 6.4.0 e posteriores, o Ajuste de Escala Gerenciado do Amazon EMR não reduz a escala verticalmente dos nós com `ApplicationMaster` para o Apache Spark em execução neles. Isso minimiza falhas e novas tentativas de trabalho, o que ajuda a melhorar a performance do trabalho e reduzir custos. Para confirmar quais nós do cluster estão executando `ApplicationMaster`, acesse o Spark History Server e filtre o driver na guia Executores do ID da aplicação Spark.

Se o cluster não tiver nenhuma carga, o Amazon EMR cancelará a adição de novas instâncias de uma avaliação anterior e executará operações de redução da escala verticalmente. Se o cluster tiver uma carga pesada, o Amazon EMR cancelará a remoção de instâncias e executará operações de aumento da escala verticalmente.

Considerações sobre alocação de nós

É recomendável usar a opção de compra sob demanda para os nós centrais para evitar a perda de dados do HDFS em caso de recuperação spot. Você pode usar a opção de compra spot para nós de tarefa para reduzir custos e obter uma execução mais rápida do trabalho quando mais instâncias spot são adicionadas aos nós de tarefa.

Cenários de alocação de nós

É possível criar vários cenários de ajuste de escala com base em suas necessidades configurando os parâmetros máximo, mínimo, limite sob demanda e nó central máximo em combinações diferentes.

Cenário 1: escalar somente os nós centrais

Para escalar somente os nós centrais, os parâmetros do ajuste de escala gerenciado devem atender aos seguintes requisitos:

- O limite sob demanda é igual ao limite máximo.
- O nó central máximo é igual ao limite máximo.

Quando o limite sob demanda e os parâmetros máximos do nó central não estão especificados, ambos os parâmetros assumem o limite máximo como padrão.

Os exemplos a seguir demonstram o cenário de ajuste de escala somente para os nós centrais.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma sob demanda De tarefa: uma sob demanda e uma spot	`UnitType`: instâncias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 20	Escale de 1 a 20 instâncias ou unidades da frota de instâncias nos nós centrais usando o tipo sob demanda.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Frotas de instâncias Central: uma sob demanda De tarefa: uma sob demanda e uma spot	UnitType: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 20	Sem ajuste de escala nos nós de tarefa.

Cenário 2: escalar somente nós de tarefa

Para escalar somente os nós de tarefa, os parâmetros do ajuste de escala gerenciado devem atender ao seguinte requisito:

- O nó central máximo deve ser igual ao limite mínimo.

Os exemplos a seguir demonstram o cenário de ajuste de escala somente para os nós de tarefa.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: duas sob demanda De tarefa: uma spot	`UnitType`: instâncias `MinimumCapacityUnits` : 2 `MaximumCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 2	Mantenha os nós centrais estáveis em 2 e escale somente os nós de tarefa de 0 a 18 instâncias ou unidades da frota de instância

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Frotas de instâncias Central: duas sob demanda De tarefa: uma spot	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 2 `MaximumCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 2	s. A capacidade entre os limites mínimo e máximo é adicionada somente aos nós de tarefa.

Cenário 3: somente instâncias sob demanda no cluster

Para ter somente instâncias sob demanda, o cluster e os parâmetros de ajuste de escala gerenciado devem atender aos seguintes requisitos:

- O limite sob demanda é igual ao limite máximo.

  Quando o limite sob demanda não é especificado, o valor do parâmetro assume o limite máximo como padrão. O valor padrão indica que o Amazon EMR escalará somente instâncias sob demanda.

Se o nó central máximo for menor que o limite máximo, o parâmetro do nó central máximo poderá ser usado para dividir a alocação de capacidade entre os nós centrais e os nós de tarefa.

Para habilitar esse cenário em um cluster composto por grupos de instâncias, todos os grupos de nós do cluster devem usar o tipo de mercado sob demanda durante a configuração inicial.

Os exemplos a seguir demonstram o cenário de ter instâncias sob demanda em todo o cluster.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma sob demanda	`UnitType`: instâncias `MinimumCapacityUnits` : 1	Escale de 1 a 12 instâncias

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
De tarefa: uma sob demanda	`MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 12	ou unidades da frota de instâncias nos nós centrais usando o tipo sob demanda. Escale a capacidade restante usando sob demanda em nós de tarefa. Sem ajuste de escala usando instâncias spot.
Frotas de instâncias Central: uma sob demanda De tarefa: uma sob demanda	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 12	

Cenário 4: somente instâncias spot no cluster

Para ter somente instâncias spot, os parâmetros de ajuste de escala gerenciado devem atender aos seguintes requisitos:

- O limite sob demanda está definido como 0.

Se o nó central máximo for menor que o limite máximo, o parâmetro do nó central máximo poderá ser usado para dividir a alocação de capacidade entre os nós centrais e os nós de tarefa.

Para habilitar esse cenário em um cluster composto por grupos de instâncias, o grupo de instâncias central deve usar a opção de compra spot durante a configuração inicial. Se não houver nenhuma instância spot no grupo de instâncias de tarefa, o Ajuste de Escala Gerenciado do Amazon EMR criará um grupo de tarefas usando instâncias spot quando necessário.

Os exemplos a seguir demonstram o cenário de ter instâncias spot em todo o cluster.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma spot De tarefa: uma spot	`UnitType`: instâncias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 0	Escale de 1 a 20 instâncias ou unidades da frota de instâncias nos nós centrais usando spot. Sem ajuste de escala usando o tipo sob demanda.
Frotas de instâncias Central: uma spot De tarefa: uma spot	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 0	

Cenário 5: escalar instâncias sob demanda nos nós centrais e instâncias spot nos nós de tarefa

Para escalar instâncias sob demanda em nós centrais e instâncias spot em nós de tarefa, os parâmetros de ajuste de escala gerenciado devem atender aos seguintes requisitos:

- O limite sob demanda deve ser igual ao nó central máximo.
- Tanto o limite sob demanda como o nó central máximo devem ser menores que o limite máximo.

Para habilitar esse cenário em um cluster composto por grupos de instâncias, o grupo de nós centrais deve usar a opção de compra sob demanda.

Os exemplos a seguir demonstram o cenário de ajuste de escala de instâncias sob demanda nos nós centrais e instâncias spot nos nós de tarefa.

Estado inicial do cluster	Parâmetros de escalabilidade	Comportamento do ajuste de escala
Grupos de instâncias Central: uma sob demanda De tarefa: uma sob demanda e uma spot	`UnitType`: instâncias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 7 `MaximumCoreCapacityUnits` : 7	Escale até 6 unidades sob demanda no nó central, pois já existe 1 unidade sob demanda no nó de tarefa, e o limite máximo para sob demanda é 7. Depois escale até 13 unidades spot nos nós de tarefa.
Frotas de instâncias Central: uma sob demanda De tarefa: uma sob demanda e uma spot	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 7 `MaximumCoreCapacityUnits` : 7	

Noções básicas sobre métricas de ajuste de escala gerenciado

O Amazon EMR publica métricas de alta resolução com dados em uma granularidade de um minuto quando o ajuste de escala gerenciado está habilitado em um cluster. Você pode visualizar eventos em cada iniciação e conclusão de redimensionamento controlados pela escalabilidade gerenciada com o console do Amazon EMR ou o console da Amazon. CloudWatch CloudWatch as métricas são essenciais para a operação da escalabilidade gerenciada do Amazon EMR. Recomendamos que você monitore de perto CloudWatch as métricas para garantir que os dados não estejam ausentes. Para obter mais informações sobre como você pode configurar CloudWatch alarmes para detectar métricas ausentes, consulte Usando CloudWatch alarmes da Amazon. Para obter mais informações sobre o uso de CloudWatch eventos com o Amazon EMR, consulte Monitorar CloudWatch eventos.

As métricas a seguir indicam as capacidades atuais ou de destino de um cluster. Essas métricas só estão disponíveis quando a escalabilidade gerenciada está habilitada. Para clusters compostos por frotas de instâncias, as métricas de capacidade de cluster são medidas em `Units`. Para clusters compostos por grupos de instâncias, as métricas de capacidade de cluster são medidas em `Nodes` ou `vCPU` com base no tipo de unidade usado na política de escalabilidade gerenciada.

Métrica	Descrição
• `TotalUnitsRequested` • `TotalNodesRequested` • `TotalVCPURequested`	O número total desejado de unidades/nós/vCPUs em um cluster, conforme determinado pela escalabilidade gerenciada. Unidades: Contagem
• `TotalUnitsRunning` • `TotalNodesRunning` • `TotalVCPURunning`	O número total atual de unidades/nós/vCPUs disponíveis em um cluster em execução. Quando um redimensionamento de cluster for solicitado, essa métrica será atualizada depois que as novas instâncias forem adicionadas ou removidas do cluster. Unidades: Contagem
• `CoreUnitsRequested` • `CoreNodesRequested` • `CoreVCPURequested`	O número desejado de unidades/nós/vCPUs CORE em um cluster, conforme determinado pela escalabilidade gerenciada. Unidades: Contagem
• `CoreUnitsRunning` • `CoreNodesRunning` • `CoreVCPURunning`	O número atual de unidades/nós/vCPUs CORE em execução em um cluster. Unidades: Contagem

Métrica	Descrição
• `TaskUnitsRequested` • `TaskNodesRequested` • `TaskVCPURequested`	O número desejado de unidades/nós/vCPUs TASK em um cluster, conforme determinado pela escalabilidade gerenciada. Unidades: Contagem
• `TaskUnitsRunning` • `TaskNodesRunning` • `TaskVCPURunning`	O número atual de unidades/nós/vCPUs TASK em execução em um cluster. Unidades: Contagem

As métricas a seguir indicam o status de uso do cluster e dos aplicativos. Essas métricas estão disponíveis para todos os recursos do Amazon EMR mas são publicadas em uma resolução mais alta com dados em uma granularidade de um minuto quando o ajuste de gerenciado é habilitado para um cluster. É possível correlacionar as métricas a seguir com as métricas de capacidade do cluster na tabela anterior para entender as decisões de escalabilidade gerenciada.

Métrica	Descrição
`AppsCompleted`	O número de aplicativos enviados para o YARN que foram concluídos. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
`AppsPending`	O número de aplicativos enviados para o YARN em estado pendente. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
AppsRunning	O número de aplicativos enviados para o YARN que estão em execução. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ContainerAllocated	O número de contêineres de recursos alocados peloResourceManager. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ContainerPending	O número de contêineres na fila que ainda não foram alocados. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
ContainerPendingRatio	A proporção de contêineres pendentes em relação aos contêineres alocados (ContainerPendingRatio = ContainerPending / ContainerAllocated). Se ContainerAllocated = 0, então ContainerPendingRatio =ContainerPending. O valor de ContainerPendingRatio representa um número, não uma porcentagem. Esse valor é útil para escalonar recursos de cluster com base no comportamento de alocação do contêiner. Unidades: Contagem

Métrica	Descrição
`HDFSUtilization`	O percentual de armazenamento do HDFS em uso no momento. Caso de uso: analisar a performance do cluster Unidade: percentual
`IsIdle`	Indica que um cluster não está mais executando nenhum trabalho, mas ainda está ativo e acumulando cobranças. É definido como 1 se nenhuma tarefa ou nenhum trabalho estiver em execução, caso contrário, é definido como 0. Esse valor é verificado em intervalos de 5 minutos, sendo que um valor de 1 indica somente que o cluster estava ocioso no momento da verificação, e não que ele ficou ocioso durante todo o período de 5 minutos. Para evitar falsos positivos, é necessário gerar um alarme quando esse valor for 1 em mais de uma verificação consecutiva de cinco minutos. Por exemplo, você pode gerar um alerta para esse valor se ele for 1 por 30 minutos ou mais. Caso de uso: monitorar a performance do cluster Unidade: booliano
`MemoryAvailableMB`	A quantidade de memória disponível para ser alocada. Caso de uso: monitorar o progresso do cluster Unidades: Contagem

Métrica	Descrição
`MRActiveNodes`	O número de nós que estão executando MapReduce tarefas ou trabalhos no momento. Equivalente ao `mapred.resourcemanager.NoOfActiveNodes` da métrica YARN. Caso de uso: monitorar o progresso do cluster Unidades: Contagem
`YARNMemoryAvailablePercentage`	A porcentagem de memória restante disponível para o YARN (YARN MemoryAvailablePercentage = MemoryAvailable MB/MemoryTotalMB). Esse valor é útil para escalonar recursos de cluster com base no uso da memória YARN. Unidade: percentual

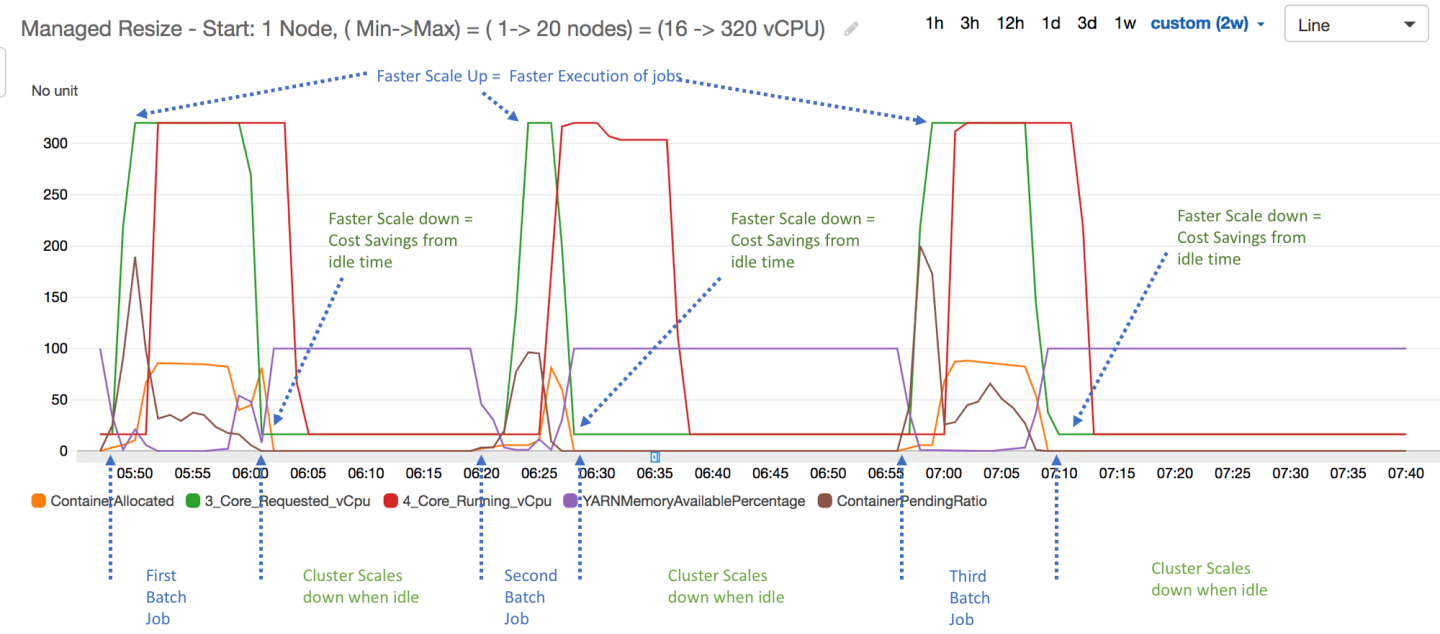
Criar grafos de métricas de ajuste de escala gerenciado

É possível criar grafos de métricas para visualizar os padrões de workload do cluster e as decisões de ajuste de escala correspondentes tomadas pelo Ajuste de Escala Gerenciado do Amazon EMR, conforme demonstrado nas etapas a seguir.

Para representar graficamente as métricas de escalabilidade gerenciadas no console CloudWatch

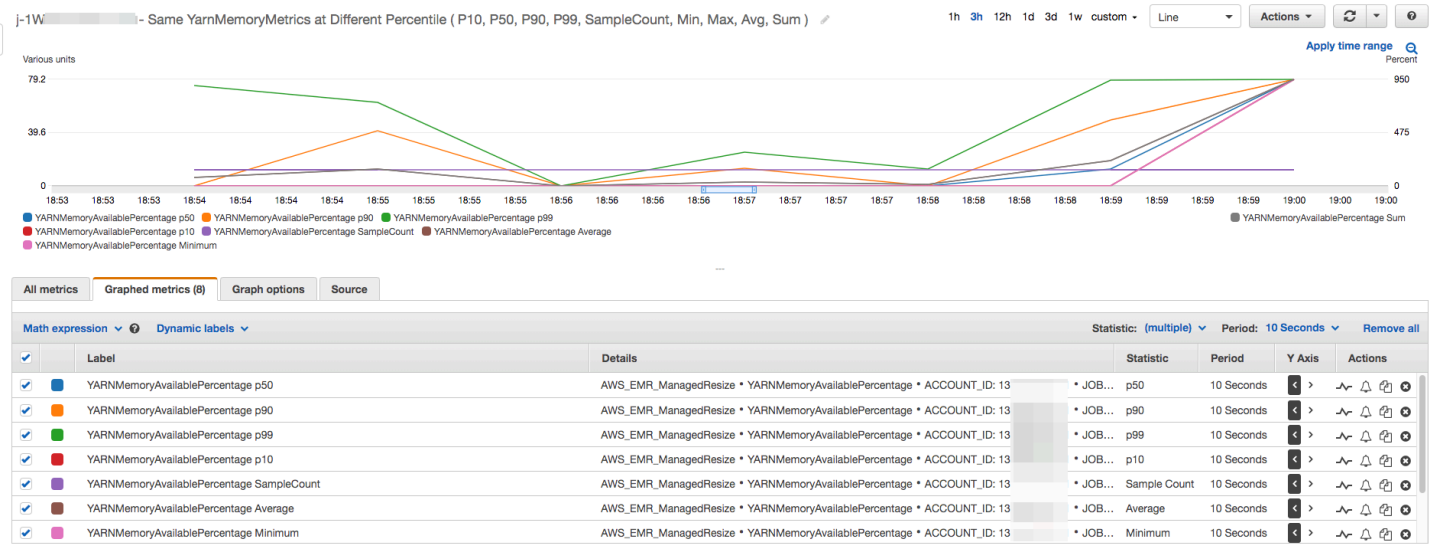
1. Abra o console de CloudWatch.
2. No painel de navegação, escolha o Amazon EMR. Você pode pesquisar com base no identificador do cluster para monitoramento.
3. Role para baixo até a métrica para exibição em gráfico. Abra uma métrica para exibir o gráfico.
4. Para criar um gráfico de uma ou mais métricas, marque a caixa de seleção ao lado de cada métrica.

O exemplo a seguir ilustra a ação de Ajuste de Escala Gerenciado do Amazon EMR de um cluster. O gráfico mostra três períodos de redução automática, que economizam custos quando há uma workload menos ativa.



Todas as métricas de capacidade e uso do cluster são publicadas em intervalos de um minuto. As informações estatísticas adicionais também estão associadas a cada dado de um minuto, o que permite representar várias funções como `Percentiles`, `Min`, `Max`, `Sum`, `Average` e `SampleCount`.

Por exemplo, o gráfico a seguir representa graficamente a mesma métrica `YARNMemoryAvailablePercentage` em percentis diferentes, P10, P50, P90 e P99, juntamente com `Sum`, `Average`, `Min` e `SampleCount`.



Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias

A escalabilidade automática com uma política personalizada nas versões 4.0 e superiores do Amazon EMR permite que você escale e escale programaticamente os nós principais e os nós de tarefas com base em CloudWatch uma métrica e em outros parâmetros que você especifica em uma política de escalabilidade. A escalabilidade automática com uma política personalizada está disponível com a configuração de grupos de instâncias e não está disponível ao usar frotas de instâncias. Para obter mais informações sobre os grupos de instâncias e frotas de instâncias, consulte Criar um cluster com frotas de instâncias ou grupos de instâncias uniformes.

> **Note**
>
> Para usar o ajuste de escala automático com um atributo de política personalizada no Amazon EMR, é necessário definir `true` para o parâmetro `VisibleToAllUsers` ao criar um cluster. Para obter mais informações, consulte SetVisibleToAllUsuários.

A política de escalabilidade é parte da configuração de um grupo de instâncias. Você pode especificar uma política durante a configuração inicial de um grupo de instâncias ou pode modificar um grupo de instâncias de um cluster existente, mesmo que esse grupo de instâncias esteja ativo. Cada grupo de instâncias em um cluster, com exceção do grupo de instâncias primário, pode ter sua própria política de ajuste de escala, que consiste em regras de aumento ou redução da escala na horizontal. As regras de expansão e redução podem ser configuradas de forma independente, com parâmetros diferentes para cada regra.

Você pode configurar políticas de escalabilidade com a AWS Management Console AWS CLI, a ou a API do Amazon EMR. Ao usar a API AWS CLI ou o Amazon EMR, você especifica a política de escalabilidade no formato JSON. Além disso, com a API do Amazon EMR AWS CLI ou com a API do Amazon EMR, você pode especificar métricas personalizadas CloudWatch . As métricas personalizadas não estão disponíveis para seleção ao usar o AWS Management Console. Quando você cria inicialmente uma política de ajuste de escala usando o console, uma política padrão adequada para muitas aplicações é pré-configurada para ajudar você a começar. Você pode excluir ou modificar as regras padrão.

Embora o escalonamento automático permita ajustar a on-the-fly capacidade do cluster do EMR, você ainda deve considerar os requisitos básicos de carga de trabalho e planejar suas configurações

de nós e grupos de instâncias. Para obter mais informações, consulte Cluster configuration guidelines.

> **Note**
>
> Para a maioria das cargas de trabalho, a configuração de ambas as regras de expansão e redução é desejável para otimizar a utilização de recursos. Definir uma regra sem a outra significa que você precisaria manualmente redimensionar o número de instâncias após uma ação de escalabilidade. Em outras palavras, isso definiria uma política "unidirecional" automática de expansão ou redução com uma reinicialização manual.

Criar o perfil do IAM para ajuste de escala automático

O ajuste de escala automático no Amazon EMR requer um perfil do IAM com permissões para adicionar e terminar instâncias quando as ações de ajuste de escala são iniciadas. Uma função padrão `EMR_AutoScaling_DefaultRole`, configurada com as políticas de função e de confiança adequadas, está disponível para esse objetivo. Quando você cria um cluster com uma política de escalabilidade pela primeira vez com o AWS Management Console, o Amazon EMR cria a função padrão e anexa a política gerenciada padrão para permissões,. `AmazonElasticMapReduceforAutoScalingRole`

Ao criar um cluster com uma política de escalabilidade automática com o AWS CLI, você deve primeiro garantir que a função padrão do IAM exista ou que você tenha uma função personalizada do IAM com uma política anexada que forneça as permissões apropriadas. Para criar a função padrão, você pode executar o comando `create-default-roles` antes de criar um cluster. Em seguida, você pode especificar a opção `--auto-scaling-role EMR_AutoScaling_DefaultRole` ao criar um cluster. Como alternativa, você pode criar uma função personalizada de escalabilidade automática e, em seguida, especificá-la ao criar um cluster, por exemplo, `--auto-scaling-role` *`MyEMRAutoScalingRole`*. Se você criar um perfil personalizado de ajuste de escala automático para o Amazon EMR, recomendamos basear as políticas de permissão para o perfil personalizado com base na política gerenciada. Para ter mais informações, consulte Configurar perfis de serviço do IAM para permissões do Amazon EMR aos serviços e recursos da AWS.

Noções básicas sobre as regras de ajuste de escala automático

Quando uma regra de aumento da escala da horizontal aciona uma ação de ajuste de escala para um grupo de instâncias, as instâncias do Amazon EC2 são adicionadas ao grupo de instâncias, de

acordo com as suas regras. Novos nós podem ser usados por aplicações como o Apache Spark, o Apache Hive e o Presto assim que a instância do Amazon EC2 entra no estado `InService`. Você também pode configurar uma regra de redução que encerra as instâncias e remove os nós. Para obter mais informações sobre o ciclo de vida das instâncias do Amazon EC2 que podem escalar automaticamente, consulte Auto Scaling lifecycle no Guia do usuário do Amazon EC2 Auto Scaling.

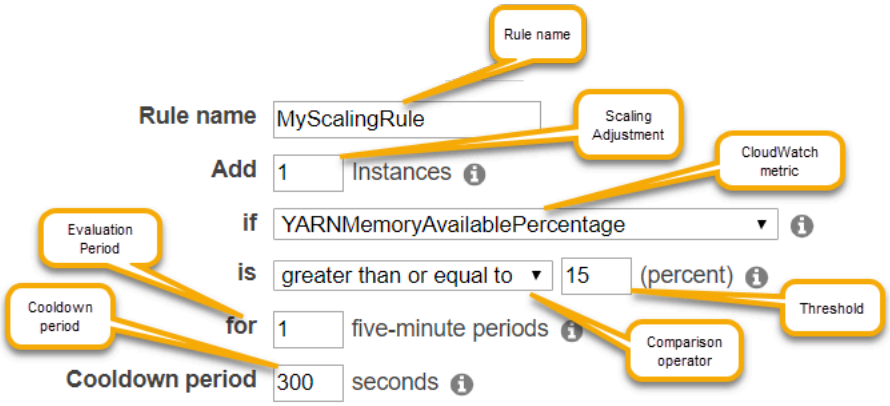
Você pode configurar como um cluster terminará as instâncias do Amazon EC2. Você pode optar por terminar no limite instância-hora do Amazon EC2 para o faturamento ou após a conclusão da tarefa. Esta configuração se aplica tanto ao Auto Scaling quanto ao redimensionamento manual de operações. Para obter mais informações sobre essa configuração, consulte Redução da escala verticalmente do cluster.

Os parâmetros a seguir se referem às regras das políticas e determinam o comportamento do Auto Scaling.

> **Note**
>
> Os parâmetros listados aqui são baseados nos do AWS Management Console Amazon EMR. Quando você usa a API AWS CLI ou o Amazon EMR, opções adicionais de configuração avançada estão disponíveis. Para obter mais informações sobre opções avançadas, consulte a SimpleScalingPolicyConfigurationReferência da API do Amazon EMR.

- Números máximo e mínimo de instâncias. A restrição Máximo de instâncias especifica o número máximo de instâncias do Amazon EC2 que podem estar no grupo de instâncias e se aplica a todas as regras de aumento da escala na horizontal. Da mesma forma, a restrição Mínimo de instâncias especifica o número mínimo de instâncias do Amazon EC2 e se aplica a todas as regras de redução da escala na horizontal.
- O Rule name (Nome da regra), que deve ser único dentro da política.
- O scaling adjustment (ajuste de escalabilidade), que determina o número de instâncias do EC2 a serem adicionadas (para regras de expansão) ou encerradas (para regras de redução) durante a ação de escalabilidade acionada pela regra.
- A CloudWatch métrica, que é observada em busca de uma condição de alarme.
- Um operador de comparação, usado para comparar a CloudWatch métrica com o valor limite e determinar uma condição de gatilho.
- Um período de avaliação, em incrementos de cinco minutos, durante o qual a CloudWatch métrica deve estar em uma condição de gatilho antes que a atividade de escalabilidade seja acionada.

- Um Cooldown period (Desaquecimento), que determina a quantidade de tempo que deve se passar entre uma ação de escalabilidade iniciada por uma regra e o início da próxima ação de escalabilidade, independentemente da regra que o aciona. Quando um grupo de instâncias conclui uma atividade de escalabilidade e atinge seu estado de pós-escala, o período de espera oferece uma oportunidade para que as CloudWatch métricas que podem acionar as atividades de escalabilidade subsequentes se estabilizem. Para obter mais informações, consulte Auto Scaling cooldowns no Guia do usuário do Amazon EC2 Auto Scaling.



Considerações e limitações

- CloudWatch As métricas da Amazon são essenciais para a operação da escalabilidade automática do Amazon EMR. Recomendamos que você monitore de perto CloudWatch as métricas da Amazon para garantir que os dados não estejam ausentes. Para obter mais informações sobre como você pode configurar CloudWatch os alarmes da Amazon para detectar métricas ausentes, consulte Usando CloudWatch alarmes da Amazon.
- A utilização excessiva dos volumes do EBS pode causar problemas de ajuste de escala gerenciado. É recomendável monitorar atentamente o uso do volume do EBS para garantir que o volume do EBS esteja abaixo de 90% de utilização. Consulte Instance storage para obter informações sobre como especificar volumes do EBS adicionais.
- A escalabilidade automática com uma política personalizada nas versões 5.18 a 5.28 do Amazon EMR pode apresentar falhas de escalabilidade causadas pela falta intermitente de dados nas métricas da Amazon. CloudWatch É recomendável usar as versões mais recentes do Amazon EMR para melhorar o ajuste de escala automático. Você também pode entrar em contato com o AWS Support para obter um patch, caso precise usar uma versão do Amazon EMR entre 5.18 e 5.28.

Usando o AWS Management Console para configurar o escalonamento automático

Ao criar um cluster, você configura uma política de ajuste de escala para os grupos de instâncias usando as opções de configuração avançadas do cluster. Você também pode criar ou modificar uma política de escalabilidade para um grupo de instâncias em serviço modificando os grupos de instâncias nas configurações de Hardware de um cluster existente.

> **Note**
>
> O novo console do Amazon EMR (https://console.aws.amazon.com/emr) usa ajuste de escala gerenciado em vez de ajuste de escala automático. Para usar o ajuste de escala automático, verifique se você fez login no console antigo em https://console.aws.amazon.com/elasticmapreduce.

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Se você estiver criando um cluster, no console do Amazon EMR, selecione Criar cluster, em seguida selecione Ir para opções avançadas, escolha as opções em Etapa1: software e etapas e acesse Etapa 2: configuração de hardware.

   - ou -

   Se você estiver modificando um grupo de instâncias em um cluster em execução, selecione o seu cluster na lista de clusters e, em seguida, expanda a seção Hardware.
3. Na seção Opção de ajuste de escala e provisionamento de clusters, selecione Habilitar ajuste de escala de clusters. Selecione Criar uma política personalizada de escalabilidade automática.

   Na tabela de Políticas personalizadas de escalabilidade automática, clique no ícone de lápis que aparece na linha do grupo de instâncias que você deseja configurar. A tela Regras do Auto Scaling é exibida.
4. Digite o número de Maximum instances (Máximo de instâncias) que você deseja que o grupo de instâncias tenha quando houver uma expansão e digite o número de Minimum instances (Mínimo de instâncias) que deseja que o grupo de instâncias tenha quando houver uma redução.
5. Clique no lápis para editar os parâmetros das regras, clique em X para remover uma regra da política e clique em Add rule (Adicionar regra) para acrescentar regras adicionais.

6. Escolha os parâmetros para as regras como descrevemos anteriormente neste tópico. Para obter descrições das CloudWatch métricas disponíveis para o Amazon EMR, consulte as métricas e dimensões do Amazon EMR no Guia do usuário da Amazon. CloudWatch

Usando o AWS CLI para configurar o escalonamento automático

Você pode usar AWS CLI comandos para o Amazon EMR para configurar a escalabilidade automática ao criar um cluster e ao criar um grupo de instâncias. Você pode usar uma sintaxe abreviada, especificando a configuração do JSON nas linhas dos comandos relevantes, ou pode fazer referência a um arquivo que contém a configuração do JSON. Você também pode aplicar uma política de Auto Scaling para um grupo de instâncias existente e remover uma política de Auto Scaling que foi aplicada anteriormente. Além disso, você pode recuperar os detalhes da configuração de uma política de escalabilidade de um cluster em execução.

> **Important**
>
> Ao criar um cluster que tem uma política de ajuste de escala automático, é necessário usar o comando `--auto-scaling-role` *`MyAutoScalingRole`* para especificar o perfil do IAM para o ajuste de escala automático. A função padrão é *`EMR_AutoScaling_DefaultRole`* e pode ser criada com o comando `create-default-roles`. Esta função só pode ser adicionada quando o cluster é criado e não em um cluster existente.

Para obter uma descrição detalhada dos parâmetros disponíveis ao configurar uma política de escalabilidade automática, consulte a Referência da API do PutAutoScalingPolicyAmazon EMR.

Criar um cluster com uma política do Auto Scaling aplicada a um grupo de instâncias

Você pode especificar uma configuração de escalabilidade automática dentro da opção `--instance-groups` do comando `aws emr create-cluster`. O exemplo a seguir ilustra um comando create-cluster em que uma política de Auto Scaling para o grupo de instâncias core é fornecida na linha. O comando cria uma configuração de escalabilidade equivalente à política de escalabilidade horizontal padrão que aparece quando você cria uma política de escalabilidade automática com o for Amazon AWS Management Console EMR. Para não estender a explicação, não mostramos uma política de redução. Não é recomendável criar uma regra de expansão sem uma regra de redução.

```
aws emr create-cluster --release-label emr-5.2.0 --service-role
 EMR_DefaultRole --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole
 --auto-scaling-role EMR_AutoScaling_DefaultRole  --instance-groups
 Name=MyMasterIG,InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1
 'Name=MyCoreIG,InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=2,AutoScalingPolicy
scale-out,Description=Replicates the default scale-out rule in the
 console.,Action={SimpleScalingPolicyConfiguration={AdjustmentType=CHANGE_IN_CAPACITY,ScalingAd
ElasticMapReduce,Period=300,Statistic=AVERAGE,Threshold=15,Unit=PERCENT,Dimensions=[{Key=JobFlo
```

O comando a seguir ilustra como usar a linha de comando para fornecer a definição da política do Auto Scaling como parte de um arquivo de configuração de grupo de instâncias chamado *instancegroupconfig.json*.

```
aws emr create-cluster --release-label emr-5.2.0 --service-role EMR_DefaultRole --ec2-
attributes InstanceProfile=EMR_EC2_DefaultRole --instance-groups file://your/path/to/
instancegroupconfig.json --auto-scaling-role EMR_AutoScaling_DefaultRole
```

O conteúdo do arquivo de configuração é o seguinte:

```
[
{
  "InstanceCount": 1,
  "Name": "MyMasterIG",
  "InstanceGroupType": "MASTER",
  "InstanceType": "m5.xlarge"
},
{
  "InstanceCount": 2,
  "Name": "MyCoreIG",
  "InstanceGroupType": "CORE",
  "InstanceType": "m5.xlarge",
  "AutoScalingPolicy":
    {
     "Constraints":
      {
       "MinCapacity": 2,
       "MaxCapacity": 10
      },
     "Rules":
     [
```

```
      {
       "Name": "Default-scale-out",
       "Description": "Replicates the default scale-out rule in the console for YARN
 memory.",
       "Action":{
        "SimpleScalingPolicyConfiguration":{
          "AdjustmentType": "CHANGE_IN_CAPACITY",
          "ScalingAdjustment": 1,
          "CoolDown": 300
        }
       },
       "Trigger":{
        "CloudWatchAlarmDefinition":{
          "ComparisonOperator": "LESS_THAN",
          "EvaluationPeriods": 1,
          "MetricName": "YARNMemoryAvailablePercentage",
          "Namespace": "AWS/ElasticMapReduce",
          "Period": 300,
          "Threshold": 15,
          "Statistic": "AVERAGE",
          "Unit": "PERCENT",
          "Dimensions":[
             {
               "Key" : "JobFlowId",
               "Value" : "${emr.clusterId}"
             }
          ]
        }
       }
      }
     ]
   }
}
]
```

Adicionar um grupo de instâncias com uma política do Auto Scaling a um cluster

Você pode especificar uma configuração de política de ajuste de escala usando a opção `--instance-groups` com o comando `add-instance-groups` da mesma maneira com que usa o `create-cluster`. O exemplo a seguir usa uma referência a um arquivo JSON *`instancegroupconfig.json`*, com a configuração do grupo de instâncias.

```
aws emr add-instance-groups --cluster-id j-1EKZ3TYEVF1S2 --instance-groups file://your/
path/to/instancegroupconfig.json
```

Aplicar uma política de ajuste de escala automático a um grupo de instâncias atual ou modificar uma política aplicada

Use o comando `aws emr put-auto-scaling-policy` para aplicar uma política de Auto Scaling a um grupo de instâncias existente. O grupo de instâncias deve fazer parte de um cluster que usa o perfil do IAM de ajuste de escala automático. O exemplo a seguir usa uma referência a um arquivo JSON *autoscaleconfig.json*, que especifica a configuração da política de Auto Scaling.

```
aws emr put-auto-scaling-policy --cluster-id j-1EKZ3TYEVF1S2 --instance-group-
id ig-3PLUZBA6WLS07 --auto-scaling-policy file://your/path/to/autoscaleconfig.json
```

O conteúdo do arquivo `autoscaleconfig.json`, que define a mesma regra de expansão apresentada no exemplo anterior, é mostrado a seguir.

```
{
          "Constraints": {
                  "MaxCapacity": 10,
                  "MinCapacity": 2
          },
          "Rules": [{
                  "Action": {
                          "SimpleScalingPolicyConfiguration": {
                                  "AdjustmentType": "CHANGE_IN_CAPACITY",
                                  "CoolDown": 300,
                                  "ScalingAdjustment": 1
                          }
                  },
                  "Description": "Replicates the default scale-out rule in the console
 for YARN memory",
                  "Name": "Default-scale-out",
                  "Trigger": {
                          "CloudWatchAlarmDefinition": {
                                  "ComparisonOperator": "LESS_THAN",
                                  "Dimensions": [{
                                          "Key": "JobFlowID",
                                          "Value": "${emr.clusterID}"
                                  }],
                                  "EvaluationPeriods": 1,
```

```
                                "MetricName": "YARNMemoryAvailablePercentage",
                                "Namespace": "AWS/ElasticMapReduce",
                                "Period": 300,
                                "Statistic": "AVERAGE",
                                "Threshold": 15,
                                "Unit": "PERCENT"
                            }
                    }
            }]
    }
```

Remover uma política de ajuste de escala automático de um grupo de instâncias

```
aws emr remove-auto-scaling-policy --cluster-id j-1EKZ3TYEVF1S2 --instance-group-
id ig-3PLUZBA6WLS07
```

Recuperar configuração de uma política de ajuste de escala automático

O `describe-cluster` comando recupera a configuração da política no InstanceGroup bloco. Por exemplo, o comando a seguir recupera a configuração de um cluster com o ID `j-1CWOHP4PI30VJ`.

```
aws emr describe-cluster --cluster-id j-1CWOHP4PI30VJ
```

Este comando gera o seguinte exemplo de saída.

```
{
    "Cluster": {
        "Configurations": [],
        "Id": "j-1CWOHP4PI30VJ",
        "NormalizedInstanceHours": 48,
        "Name": "Auto Scaling Cluster",
        "ReleaseLabel": "emr-5.2.0",
        "ServiceRole": "EMR_DefaultRole",
        "AutoTerminate": false,
        "TerminationProtected": true,
        "MasterPublicDnsName": "ec2-54-167-31-38.compute-1.amazonaws.com",
        "LogUri": "s3n://aws-logs-232939870606-us-east-1/elasticmapreduce/",
        "Ec2InstanceAttributes": {
            "Ec2KeyName": "performance",
```

```
            "AdditionalMasterSecurityGroups": [],
            "AdditionalSlaveSecurityGroups": [],
            "EmrManagedSlaveSecurityGroup": "sg-09fc9362",
            "Ec2AvailabilityZone": "us-east-1d",
            "EmrManagedMasterSecurityGroup": "sg-0bfc9360",
            "IamInstanceProfile": "EMR_EC2_DefaultRole"
        },
        "Applications": [
            {
                "Name": "Hadoop",
                "Version": "2.7.3"
            }
        ],
        "InstanceGroups": [
            {
                "AutoScalingPolicy": {
                    "Status": {
                        "State": "ATTACHED",
                        "StateChangeReason": {
                            "Message": ""
                        }
                    },
                    "Constraints": {
                        "MaxCapacity": 10,
                        "MinCapacity": 2
                    },
                    "Rules": [
                        {
                            "Name": "Default-scale-out",
                            "Trigger": {
                                "CloudWatchAlarmDefinition": {
                                    "MetricName": "YARNMemoryAvailablePercentage",
                                    "Unit": "PERCENT",
                                    "Namespace": "AWS/ElasticMapReduce",
                                    "Threshold": 15,
                                    "Dimensions": [
                                        {
                                            "Key": "JobFlowId",
                                            "Value": "j-1CWOHP4PI30VJ"
                                        }
                                    ],
                                    "EvaluationPeriods": 1,
                                    "Period": 300,
                                    "ComparisonOperator": "LESS_THAN",
```

```
                            "Statistic": "AVERAGE"
                        }
                    },
                    "Description": "",
                    "Action": {
                        "SimpleScalingPolicyConfiguration": {
                            "CoolDown": 300,
                            "AdjustmentType": "CHANGE_IN_CAPACITY",
                            "ScalingAdjustment": 1
                        }
                    }
                },
                {
                    "Name": "Default-scale-in",
                    "Trigger": {
                        "CloudWatchAlarmDefinition": {
                            "MetricName": "YARNMemoryAvailablePercentage",
                            "Unit": "PERCENT",
                            "Namespace": "AWS/ElasticMapReduce",
                            "Threshold": 75,
                            "Dimensions": [
                                {
                                    "Key": "JobFlowId",
                                    "Value": "j-1CWOHP4PI30VJ"
                                }
                            ],
                            "EvaluationPeriods": 1,
                            "Period": 300,
                            "ComparisonOperator": "GREATER_THAN",
                            "Statistic": "AVERAGE"
                        }
                    },
                    "Description": "",
                    "Action": {
                        "SimpleScalingPolicyConfiguration": {
                            "CoolDown": 300,
                            "AdjustmentType": "CHANGE_IN_CAPACITY",
                            "ScalingAdjustment": -1
                        }
                    }
                }
            ]
        },
        "Configurations": [],
```

```
                "InstanceType": "m5.xlarge",
                "Market": "ON_DEMAND",
                "Name": "Core - 2",
                "ShrinkPolicy": {},
                "Status": {
                    "Timeline": {
                        "CreationDateTime": 1479413437.342,
                        "ReadyDateTime": 1479413864.615
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "RunningInstanceCount": 2,
                "Id": "ig-3M16XBE8C3PH1",
                "InstanceGroupType": "CORE",
                "RequestedInstanceCount": 2,
                "EbsBlockDevices": []
            },
            {
                "Configurations": [],
                "Id": "ig-OP62I28NSE8M",
                "InstanceGroupType": "MASTER",
                "InstanceType": "m5.xlarge",
                "Market": "ON_DEMAND",
                "Name": "Master - 1",
                "ShrinkPolicy": {},
                "EbsBlockDevices": [],
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "CreationDateTime": 1479413437.342,
                        "ReadyDateTime": 1479413752.088
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "RunningInstanceCount": 1
            }
        ],
        "AutoScalingRole": "EMR_AutoScaling_DefaultRole",
```

```
        "Tags": [],
        "BootstrapActions": [],
        "Status": {
            "Timeline": {
                "CreationDateTime": 1479413437.339,
                "ReadyDateTime": 1479413863.666
            },
            "State": "WAITING",
            "StateChangeReason": {
                "Message": "Cluster ready after last step completed."
            }
        }
    }
}
```

Redimensionar manualmente um cluster em execução

Você pode adicionar e remover instâncias de grupos de instâncias principais e de tarefas e frotas de instâncias em um cluster em execução com a AWS Management Console, AWS CLI, ou a API do Amazon EMR. Se um cluster usa grupos de instâncias, você altera explicitamente a contagem de instâncias. Se o cluster usa frotas de instâncias, você pode alterar as unidades de destino para instâncias sob demanda e instâncias spot. A frota de instâncias, em seguida, adiciona e remove instâncias para corresponder ao novo destino. Para ter mais informações, consulte Opções de frotas de instâncias. As aplicações podem usar essas instâncias do Amazon EC2 recém-disponibilizadas para hospedar os nós assim que as instâncias se tornarem disponíveis. Quando as instâncias são removidas, o Amazon EMR desativa as tarefas de uma forma que não interrompe os trabalhos e proteções contra a perda de dados. Para ter mais informações, consulte Terminar na conclusão de tarefas.

Redimensionar um cluster usando o console

Você pode usar o console do Amazon EMR para redimensionar um cluster em execução.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Alterar a contagem de instâncias para um cluster existente usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2 no painel de navegação esquerdo, escolha Clusters e selecione o cluster que você deseja atualizar. O cluster deve estar em execução, e não é possível redimensionar um cluster provisionado ou terminado.
3. Na guia Instâncias da página de detalhes do cluster, visualize o painel Grupos de instâncias.
4. Para redimensionar um grupo de instâncias já existente, selecione o botão de opção ao lado do grupo de instâncias central ou de tarefa que você deseja redimensionar e escolha Redimensionar grupo de instâncias. Especifique o novo número de instâncias do grupo de instâncias e selecione Redimensionar.

   > **Note**
   >
   > Se você optar por reduzir o tamanho de um grupo de instâncias que estão em execução, o Amazon EMR selecionará de forma inteligente as instâncias a serem removidas do grupo para perda mínima de dados. Para um controle mais granular da ação de redimensionamento, você pode selecionar o ID do grupo de instâncias, escolher as instâncias que deseja remover e usar a opção Terminar. Para obter mais informações sobre o comportamento inteligente de redução da escala verticalmente, consulte Redução da escala verticalmente do cluster.

5. Para cancelar a ação de redimensionamento, selecione o botão de opção para um grupo de instâncias com o status Resizing e escolha Interromper redimensionamento na lista de ações.
6. Para adicionar um ou mais grupos de instâncias de tarefa ao cluster em resposta ao aumento da workload, escolha Adicionar grupo de instâncias de tarefa na lista de ações. Escolha o tipo de instância do Amazon EC2, insira o número de instâncias para o grupo de tarefa e selecione Adicionar grupo de instâncias de tarefa para retornar ao painel Grupos de instâncias do cluster.

Old console

Alterar a contagem de instâncias para um cluster existente usando o console antigo

1. Na página Cluster List (Lista de clusters), escolha um cluster para redimensionar.
2. Na página Cluster Details (Detalhes do cluster), escolha Hardware.
3. Se o seu cluster usa grupos de instâncias, escolha Resize (Redimensionar) na coluna Instance count (Contagem de instâncias) para o grupo de instâncias que você deseja redimensionar, digite uma nova contagem de instâncias e, em seguida, clique na marca de seleção verde.

   OU

   Se o cluster usa frotas de instâncias, escolha Redimensionar na coluna Capacidade provisionada, digite novos valores para Unidades sob demanda e Unidades spot e, em seguida, escolha Redimensionar.

Quando você altera o número de nós, o Status do grupo de instâncias é atualizado. Quando a alteração solicitada estiver concluída, o Status muda para Running (Em execução).

Redimensionar um cluster com o AWS CLI

Você pode usar o AWS CLI para redimensionar um cluster em execução. Você pode aumentar ou diminuir o número de nós de tarefa, e pode aumentar o número de nós core de um cluster em execução. Também é possível encerrar uma instância no grupo de instâncias principal com a AWS CLI ou a API. Isso deve ser feito com cuidado. Desativar uma instância no grupo de instâncias centrais expõe você ao risco de perda de dados, e a instância não é substituída automaticamente.

Além de redimensionar os grupos centrais e de tarefa, você também pode adicionar um ou mais grupos de instâncias de tarefa a um cluster em execução usando a AWS CLI.

Para redimensionar um cluster alterando a contagem de instâncias com AWS CLI

Você pode adicionar instâncias ao grupo principal ou ao grupo de tarefas e remover instâncias do grupo de tarefas com o AWS CLI `modify-instance-groups` subcomando com o `InstanceCount` parâmetro. Para adicionar instâncias aos grupos core ou de tarefas, aumente o `InstanceCount`. Para reduzir o número de instâncias no grupo de tarefas, diminua o `InstanceCount`. Alterar o número de instâncias do grupo de tarefas para 0 remove todas as instâncias, mas não o grupo de instâncias.

- Para aumentar o número de instâncias no grupo de instâncias de tarefas de 3 para 4, digite o seguinte comando e substitua *ig-31JXXXXXXBT0* pelo ID do grupo de instâncias.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-31JXXXXXXBT0,InstanceCount=4
```

Para recuperar o `InstanceGroupId`, use o subcomando `describe-cluster`. A saída é um objeto JSON chamado `Cluster` que contém o ID de cada grupo de instâncias. Para usar esse comando, você precisa do ID do cluster (que pode ser recuperado usando o comando `aws emr list-clusters` ou pelo console). Para recuperar o ID do grupo de instâncias, digite o seguinte comando e substitua *j-2AXXXXXXGAPLF* pelo ID do cluster.

```
aws emr describe-cluster --cluster-id j-2AXXXXXXGAPLF
```

Com o AWS CLI, você também pode encerrar uma instância no grupo de instâncias principal com o `--modify-instance-groups` subcomando.

> **Warning**
>
> A especificação de `EC2InstanceIdsToTerminate` deve ser feita com cuidado. As instâncias são encerradas imediatamente, independentemente do status dos aplicativos em execução nelas, e as instâncias não são substituídas automaticamente. Isso é verdadeiro, independentemente da configuração de Scale down behavior (Comportamento da escalabilidade vertical) do cluster. O encerramento de uma instância dessa forma tem o risco de perda de dados e de comportamento imprevisível do cluster.

Para encerrar uma instância específica, você precisa do ID do grupo de instâncias (fornecido pelo subcomando `aws emr describe-cluster --cluster-id`) e do ID da instância (fornecido pelo subcomando `aws emr list-instances --cluster-id`) e, em seguida, digite o seguinte comando, substituindo *ig-6RXXXXXX07SA* pelo ID do grupo de instâncias e *i-f9XXXXf2* pelo ID da instância.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-6RXXXXXX07SA,EC2InstanceIdsToTerminate=i-f9XXXXf2
```

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Para redimensionar um cluster adicionando grupos de instâncias de tarefas com o AWS CLI

Com o AWS CLI, você pode adicionar de 1 a 48 grupos de instâncias de tarefas a um cluster com o `--add-instance-groups` subcomando. Os grupos de instâncias de tarefa só podem ser adicionados a um cluster contendo um grupo de instâncias primárias e um grupo de instâncias centrais. Ao usar o AWS CLI, você pode adicionar até cinco grupos de instâncias de tarefas sempre que usar o `--add-instance-groups` subcomando.

1. Para adicionar um único grupo de instâncias de tarefas a um cluster, digite o seguinte comando e substitua *j-JXBXXXXXX37R* pelo ID do cluster.

```
aws emr add-instance-groups --cluster-id j-JXBXXXXXX37R --instance-groups
 InstanceCount=6,InstanceGroupType=task,InstanceType=m5.xlarge
```

2. Para adicionar vários grupos de instâncias de tarefas a um cluster, digite o seguinte comando e substitua *j-JXBXXXXXX37R* pelo ID do cluster. Você pode adicionar até cinco grupos de instâncias de tarefas em um único comando.

```
aws emr add-instance-groups --cluster-id j-JXBXXXXXX37R --instance-
groups InstanceCount=6,InstanceGroupType=task,InstanceType=m5.xlarge
 InstanceCount=10,InstanceGroupType=task,InstanceType=m5.xlarge
```

   Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Interromper um redimensionamento

Usando o Amazon EMR versão 4.1.0 ou posteriores, você pode iniciar um redimensionamento no meio de uma operação de redimensionamento existente. Além disso, você pode interromper uma solicitação de redimensionamento enviada anteriormente ou enviar uma nova solicitação para substituir uma solicitação anterior, antes mesmo que ela seja concluída. Você também pode interromper um redimensionamento pelo console ou com a chamada de API `ModifyInstanceGroups`, usando a contagem atual como a contagem de destino do cluster.

A imagem a seguir mostra um grupo de instâncias de tarefas que está sendo redimensionado mas pode ser interrompido pela opção de Stop (Interromper).

Para interromper um redimensionamento com o AWS CLI

Você pode usar o AWS CLI para interromper o redimensionamento com o `modify-instance-groups` subcomando. Suponha que você tem seis instâncias em um grupo de instâncias e deseja aumentar este número para 10. E mais tarde você decide cancelar essa solicitação:

- A solicitação inicial:

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-myInstanceGroupId,InstanceCount=10
```

  A segunda solicitação para interromper a primeira solicitação:

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-myInstanceGroupId,InstanceCount=6
```

> **Note**
>
> Como esse processo é assíncrono, é possível que você veja o número de instâncias ser alterado em relação às solicitações anteriores da API antes que as solicitações subsequentes sejam acatadas. Em caso de redução, se você tiver um trabalho em execução nos nós, é possível que o grupo de instâncias não seja reduzido até que os nós tenham concluído seu trabalho.

Estado suspenso

Um grupo de instâncias entra em estado suspenso se encontrar muitos erros durante a tentativa de iniciar os novos nós do cluster. Por exemplo, se os novos nós falham ao executar ações de bootstrap, o grupo de instâncias entra no estado SUSPENDED, em vez de continuar a fornecer novos nós. Depois de resolver o problema básico, redefina o número desejado de nós no grupo de instâncias do cluster e, em seguida, o grupo de instâncias reiniciará a alocação de nós. A

modificação de um grupo de instâncias instrui o Amazon EMR a tentar fornecer nós novamente. Os nós em execução não são reiniciados ou encerrados.

No AWS CLI, o `list-instances` subcomando retorna todas as instâncias e seus estados, assim como o `describe-cluster` subcomando. Se o Amazon EMR detecta uma falha com um grupo de instâncias, ele altera o estado do grupo para SUSPENDED.

Para redefinir um cluster em um estado SUSPENSO com o AWS CLI

Digite o subcomando `describe-cluster` com o parâmetro `--cluster-id` para visualizar o estado das instâncias no cluster.

- Para exibir informações sobre todas as instâncias e grupos de instâncias em um cluster, digite o seguinte comando e substitua *j-3KVXXXXXXY7UG* pelo ID do cluster.

  ```
  aws emr describe-cluster --cluster-id j-3KVXXXXXXY7UG
  ```

  A saída exibe informações sobre os grupos de instâncias e o estado das instâncias:

  ```
  {
      "Cluster": {
          "Status": {
              "Timeline": {
                  "ReadyDateTime": 1413187781.245,
                  "CreationDateTime": 1413187405.356
              },
              "State": "WAITING",
              "StateChangeReason": {
                  "Message": "Waiting after step completed"
              }
          },
          "Ec2InstanceAttributes": {
              "Ec2AvailabilityZone": "us-west-2b"
          },
          "Name": "Development Cluster",
          "Tags": [],
          "TerminationProtected": false,
          "RunningAmiVersion": "3.2.1",
          "NormalizedInstanceHours": 16,
          "InstanceGroups": [
              {
                  "RequestedInstanceCount": 1,
  ```

```
                "Status": {
                    "Timeline": {
                        "ReadyDateTime": 1413187775.749,
                        "CreationDateTime": 1413187405.357
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "Name": "MASTER",
                "InstanceGroupType": "MASTER",
                "InstanceType": "m5.xlarge",
                "Id": "ig-3ETXXXXXXFYV8",
                "Market": "ON_DEMAND",
                "RunningInstanceCount": 1
            },
            {
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "ReadyDateTime": 1413187781.301,
                        "CreationDateTime": 1413187405.357
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "Name": "CORE",
                "InstanceGroupType": "CORE",
                "InstanceType": "m5.xlarge",
                "Id": "ig-3SUXXXXXXQ9ZM",
                "Market": "ON_DEMAND",
                "RunningInstanceCount": 1
            }
...
}
```

Para exibir as informações sobre um grupo de instâncias específico, digite o subcomando `list-instances` com os parâmetros `--cluster-id` e `--instance-group-types`. Você pode visualizar as informações para grupos primários, centrais ou de tarefa.

```
aws emr list-instances --cluster-id j-3KVXXXXXXY7UG --instance-group-types "CORE"
```

Use o subcomando `modify-instance-groups` com o parâmetro `--instance-groups` para redefinir um cluster no estado `SUSPENDED`. O ID do grupo de instâncias é obtido pelo subcomando `describe-cluster`.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-3SUXXXXXXQ9ZM,InstanceCount=3
```

Considerações ao reduzir o tamanho do cluster

Se você optar por reduzir o tamanho de um cluster em execução, leve em consideração o seguinte comportamento e as práticas recomendadas do Amazon EMR:

- Para reduzir o impacto nas tarefas que estão em andamento, o Amazon EMR seleciona de forma inteligente as instâncias a serem removidas. Para obter mais informações sobre o comportamento de redução da escala verticalmente do cluster, consulte Terminar na conclusão de tarefas no Guia de gerenciamento do Amazon EMR.
- Quando você reduz a escala verticalmente de um cluster, o Amazon EMR copia os dados das instâncias que ele remove para as instâncias que permanecem. Verifique se há capacidade de armazenamento suficiente para esses dados nas instâncias que permanecem no grupo.
- O Amazon EMR tenta desativar o HDFS em instâncias do grupo. Antes de reduzir o tamanho de um cluster, é recomendável minimizar a E/S de gravação do HDFS.
- Para obter o controle mais granular ao reduzir o tamanho de um cluster, é possível visualizar o cluster no console e navegar até a guia Instâncias. Selecione o ID do grupo de instâncias que você deseja redimensionar. Em seguida, use a opção Terminar para as instâncias específicas que você deseja remover.

Configurar os tempos limite para a capacidade de provisionamento

Ao usar frotas de instâncias, é possível configurar os tempos limite de provisionamento. Um tempo limite de provisionamento instrui o Amazon EMR a interromper o provisionamento da capacidade da instância se o cluster exceder um limite de tempo especificado durante a inicialização do cluster ou as operações de ajuste de escala do cluster. Os tópicos a seguir abordam como configurar um tempo

limite de provisionamento para a inicialização do cluster e para operações de aumento da escala verticalmente do cluster.

Tópicos

Configurar tempos limite de provisionamento para inicialização de clusters no Amazon EMR

Você pode definir um período de tempo limite para provisionar instâncias spot para cada frota do cluster. Se o Amazon EMR não puder provisionar a capacidade spot, você escolhe entre terminar o cluster ou provisionar a capacidade sob demanda. Se o período de tempo limite acabar durante o processo de redimensionamento do cluster, o Amazon EMR cancelará solicitações spot não provisionadas. As instâncias spot que não foram provisionadas não são transferidas para a capacidade sob demanda.

> **Note**
>
> Não é possível personalizar um período de tempo limite de provisionamento no console antigo. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

Execute as etapas a seguir para personalizar um período de tempo limite de provisionamento para inicialização do cluster usando o console do Amazon EMR.

New console

Configurar o tempo limite de provisionamento ao criar um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Na página Criar cluster, navegue até Configuração do cluster e selecione Frotas de instâncias.

4. Em Opção de ajuste de escala e provisionamento de clusters, especifique o tamanho do spot para suas frotas centrais e de tarefa.
5. Em Configuração de tempo limite spot, selecione Terminar cluster após o tempo limite spot ou Alternar para sob demanda após tempo limite spot. Em seguida, especifique o período de tempo limite para provisionamento de instâncias spot. O valor padrão é uma hora.
6. Escolha qualquer outra opção que se aplique ao cluster.
7. Para iniciar o cluster com o tempo limite configurado, escolha Criar cluster.

AWS CLI

Especificar um tempo limite de provisionamento com o comando **`create-cluster`**

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"LaunchSp
{"OnDemandSpecification":{"AllocationStrategy":"lowest-
price"}},"InstanceTypeConfigs":[{"WeightedCapacity":1,"EbsConfiguration":
{"EbsBlockDeviceConfigs":[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemand
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecifi
{"SpotSpecification":
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification"
{"AllocationStrategy":"lowest-price"}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemand
 - 2"}]'
```

Personalizar um período de tempo limite de provisionamento para redimensionamento do cluster no Amazon EMR

É possível definir um período de tempo limite para provisionar instâncias spot para cada frota do cluster. Se o Amazon EMR não conseguir provisionar a capacidade spot, ele cancelará a solicitação de redimensionamento e interromperá as tentativas de provisionar capacidade spot adicional.

Ao criar um cluster, é possível configurar o tempo limite. Em um cluster em execução, é possível adicionar ou atualizar um tempo limite.

Quando o período de tempo limite expira, o Amazon EMR envia automaticamente os eventos para um stream do Amazon Events. CloudWatch Com CloudWatch, você pode criar regras que correspondam aos eventos de acordo com um padrão especificado e, em seguida, rotear os eventos aos alvos para que sejam executadas ações. Por exemplo, é possível configurar uma regra para enviar uma notificação por e-mail. Para obter mais informações sobre como criar regras, consulte Criação de regras para eventos do Amazon EMR com CloudWatch. Para obter mais informações sobre diferentes detalhes de evento, consulte Eventos de alteração de estado da frota de instâncias.

Exemplos de tempos limite de provisionamento para redimensionamento de clusters

Especifique um tempo limite de provisionamento para redimensionar usando a AWS CLI

O exemplo a seguir usa o comando `create-cluster` para adicionar um tempo limite de provisionamento para redimensionamento.

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"InstanceTyp
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecificat
{"SpotSpecification":
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification":
{"AllocationStrategy":"lowest-price"}},"ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":20},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":25}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 2"}]'
```

O exemplo a seguir usa o comando `modify-instance-fleet` para adicionar um tempo limite de provisionamento para redimensionamento.

```
aws emr modify-instance-fleet \
--cluster-id j-XXXXXXXXXXXX \
--instance-fleet '{"InstanceFleetId":"if-XXXXXXXXXXX","ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":30},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":60}}}' \
--region us-east-1
```

O exemplo a seguir usa `add-instance-fleet-command` para adicionar um tempo limite de provisionamento para redimensionamento.

```
aws emr add-instance-fleet \
--cluster-id j-XXXXXXXXXXXX \
--instance-fleet
 '{"InstanceFleetType":"TASK","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"InstanceTypeCo
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
{"SpotResizeSpecification":{"TimeoutDurationMinutes":30},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":35}}}' \
--region us-east-1
```

Especifique um tempo limite de provisionamento para redimensionar e iniciar com o AWS CLI

O exemplo a seguir usa o comando `create-cluster` para adicionar um tempo limite de provisionamento para redimensionamento e inicialização.

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"LaunchSpeci
{"OnDemandSpecification":{"AllocationStrategy":"lowest-price"}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecificat
{"SpotSpecification":
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification":
{"AllocationStrategy":"lowest-price"}},"ResizeSpecifications":
```

```
{"SpotResizeSpecification":{"TimeoutDurationMinutes":20},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":25}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 2"}]'
```

Considerações para redimensionamento dos tempos limite de provisionamento

Ao configurar os tempos limite de provisionamento de clusters para suas frotas de instâncias, leve em consideração os comportamentos a seguir.

- É possível configurar os tempos limite de provisionamento para instâncias spot e sob demanda. O tempo limite mínimo de provisionamento é de cinco minutos. O tempo limite máximo de provisionamento é de sete dias.
- Só é possível configurar tempos limite de provisionamento para um cluster do EMR que usa frotas de instâncias. É necessário configurar cada frota central e de tarefa separadamente.
- Ao criar um cluster, você pode configurar os tempos limite de provisionamento. É possível adicionar um tempo limite ou atualizar um tempo limite atual para um cluster em execução.
- Se você enviar múltiplas operações de redimensionamento, o Amazon EMR rastreará os tempos limite de provisionamento para cada operação de redimensionamento. Por exemplo, defina o tempo limite de provisionamento de um cluster para *60* minutos. Em seguida, envie uma operação de redimensionamento *R1* no tempo *T1*. Envie uma segunda operação de redimensionamento *R2* no tempo *T2*. O tempo limite de provisionamento para R1 expira em *T1 + 60 minutos*. O tempo limite de provisionamento para R2 expira em *T2 + 60 minutos*.
- Se você enviar uma nova operação de redimensionamento para aumentar a escala verticalmente antes que o tempo limite expire, o Amazon EMR continuará sua tentativa de provisionar capacidade para o cluster do EMR.

Redução da escala verticalmente do cluster

> **Note**
>
> Não há mais suporte para as opções de comportamento de redução da escala verticalmente desde o Amazon EMR versão 5.10.0. Por causa da introdução do faturamento por segundo no Amazon EC2, o comportamento padrão da redução da escala verticalmente para clusters do Amazon EMR agora é terminar na conclusão da tarefa.

Com o Amazon EMR versões 5.1.0 a 5.9.1, há duas opções para o comportamento da redução da escala verticalmente: terminar no limite de instâncias por hora de acordo com o faturamento do Amazon EC2 ou terminar quando a tarefa for concluída. A partir do Amazon EMR versão 5.10.0, a configuração de término ao atingir o limite de hora de instância se torna defasada por causa da introdução do faturamento por segundo no Amazon EC2. Não recomendamos especificar o encerramento no limite de tempo de execução da instância em que a opção está disponível.

> **Warning**
>
> Se você usar o AWS CLI para emitir um `modify-instance-groups` com`EC2InstanceIdsToTerminate`, essas instâncias serão encerradas imediatamente, sem considerar essas configurações e independentemente do status dos aplicativos em execução nelas. O encerramento de uma instância dessa forma tem o risco de perda de dados e de comportamento imprevisível do cluster.

Quando o término na conclusão da tarefa é especificado, o Amazon EMR colocas as tarefas em listas de negação e as remove dos nós antes de terminar as instâncias do Amazon EC2. Em ambos os comportamentos especificados, o Amazon EMR não termina as instâncias do Amazon EC2 em grupos de instâncias centrais se houver a possibilidade de danos ao HDFS.

Terminar na conclusão de tarefas

O Amazon EMR permite que você reduzir a escala verticalmente do cluster sem afetar a sua workload. O Amazon EMR desativa normalmente o YARN, o HDFS e os outros daemons nos nós centrais e nós de tarefa durante uma operação de redimensionamento para redução sem perder dados nem interromper trabalhos. O Amazon EMR reduzirá o tamanho de grupos de instâncias somente se o trabalho atribuído a eles tiver sido concluído e eles estiverem ociosos. Para o YARN NodeManager Graceful Decommission, você pode ajustar manualmente o tempo que um nó espera pelo descomissionamento.

Este tempo é definido usando uma propriedade na a classificação de configuração `YARN-site`. Usando o Amazon EMR versão 5.12.0 e posterior, especifique a propriedade `YARN.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. Ao usar versões anteriores do Amazon EMR, especifique a propriedade `YARN.resourcemanager.decommissioning.timeout`.

Se ainda houver contêineres ou aplicativos do YARN em execução quando o tempo limite de desativação for ultrapassado, o nó será obrigatoriamente desativado e o YARN reprogramará os

contêineres afetados em outros nós. O valor padrão é de 3.600 segundos (uma hora). Você pode definir esse tempo limite com um valor arbitrariamente alto para forçar a redução amigável a esperar mais tempo. Para obter mais informações, consulte Graceful Decommission of YARN nodes na documentação do Apache Hadoop.

Grupos de nós de tarefa

O Amazon EMR seleciona, de forma inteligente, as instâncias que não têm tarefas em execução relacionadas a etapas ou aplicações e as remove primeiro do cluster. Se todas as instâncias do cluster estiverem em uso, o Amazon EMR aguardará a conclusão das tarefas da instância antes de removê-la do cluster. O tempo de espera padrão é 1 hora. Esse valor pode ser alterado com a configuração `YARN.resourcemanager.decommissioning.timeout`. O Amazon EMR usa a nova configuração dinamicamente. Você pode definir isso como um número arbitrariamente grande para garantir que o Amazon EMR não termine nenhuma tarefa ao reduzir o tamanho do cluster.

Grupos de nós centrais

Nos nós principais, os DataNode daemons YARN NodeManager e HDFS devem ser desativados para que o grupo de instâncias seja reduzido. Para o YARN, a redução amigável garante que um nó marcado para desativação seja passado para o estado `DECOMMISSIONED` somente se não houver contêineres ou aplicações pendentes ou não concluídas. A desativação termina imediatamente se não há contêineres em execução no nó no início da desativação.

Para o HDFS, a redução amigável certifica-se de que a capacidade de destino do HDFS é grande o suficiente para comportar todos os blocos existentes. Se a capacidade alvo não for grande o suficiente, apenas uma parte das instâncias core serão desativadas, de forma que os nós restantes possam lidar com os dados atuais residentes no HDFS. Você deve garantir o aumento da capacidade no HDFS para permitir uma desativação mais extensa. Tente também minimizar a E/S de gravação antes de tentar reduzir os grupos de instâncias. O excesso de E/S de gravação poderá atrasar a conclusão da operação de redimensionamento.

Outro limite é o fator de replicação padrão `dfs.replication` no `/etc/hadoop/conf/hdfs-site`. Ao criar um cluster, o Amazon EMR configura o valor com base no número de instâncias no cluster: 1 para clusters com uma a três instâncias, 2 para clusters com quatro a nove instâncias e 3 para clusters com mais de dez instâncias.

> **Warning**
>
> 1. Definir `dfs.replication` como 1 em clusters com menos de quatro nós poderá causar perda de dados do HDFS se um único nó ficar inativo. É recomendável usar um cluster com pelo menos quatro nós centrais para workloads de produção.
> 2. O Amazon EMR não permitirá que os clusters escalem os nós principais abaixo de `dfs.replication`. Por exemplo, se `dfs.replication = 2`, o número mínimo de nós central será 2.
> 3. Ao usar o Ajuste de Escala Gerenciado, o Auto Scaling ou optar por redimensionar manualmente o cluster, é recomendável definir `dfs.replication` como 2 ou mais.

A redução amigável não permite reduzir os nós centrais abaixo do fator de replicação do HDFS. Isso permite que o HDFS feche arquivos devido à insuficiência de réplicas. Para contornar esse limite, diminua o fator de replicação e reinicie o daemon. NameNode

Configurar o comportamento de redução da escala verticalmente do Amazon EMR

> **Note**
>
> A opção de comportamento de redução da escala verticalmente de terminar na hora da instância não é mais compatível com o Amazon EMR 5.10.0 e versões posteriores. As opções de comportamento de redução da escala verticalmente a seguir são exibidas no console do Amazon EMR somente nas versões 5.1.0 a 5.9.1.

Você pode usar a AWS Management Console AWS CLI, a ou a API do Amazon EMR para configurar o comportamento de redução ao criar um cluster.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. Consulte Console do Amazon EMR para conhecer as diferenças entre as experiências do console antigo e novo.

New console

Configurar o comportamento de redução da escala verticalmente usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters e depois Criar cluster.
3. Na seção Opções de ajuste de escala e provisionamento de clusters, encontre Término de cluster e escolha terminar o cluster manualmente ou peça para o Amazon EMR terminar o cluster após certo período de tempo ocioso. Se preferir, ative a proteção contra término contra bugs ou erros.
4. Escolha qualquer outra opção que se aplique ao cluster.
5. Para iniciar o cluster, escolha Criar cluster.

Old console

Configurar o comportamento de redução da escala verticalmente usando o console antigo

1. Abra o console do Amazon EMR em https://console.aws.amazon.com/elasticmapreduce.
2. Selecione Criar cluster. Acesse Opções avançadas e escolha suas configurações na Etapa 1: software e etapas e Etapa 2: hardware.
3. Na Etapa 3: configurações gerais do cluster, selecione o comportamento de redução da escala verticalmente de sua preferência. Conclua as configurações restantes e crie o cluster.

AWS CLI

Para configurar o comportamento de redução de escala com o AWS CLI

- Use a opção `--scale-down-behavior` para especificar `TERMINATE_AT_INSTANCE_HOUR` ou `TERMINATE_AT_TASK_COMPLETION`.

Terminar um cluster

Esta seção descreve os métodos de encerramento de um cluster. Para obter informações sobre como habilitar a proteção contra encerramento e encerrar clusters automaticamente, consulte Controle de término do cluster. Você pode encerrar clusters nos estados `STARTING`,

`RUNNING`ou `WAITING`. Um cluster no estado `WAITING` deve ser encerrado ou ele será executado indefinidamente, gerando encargos para sua conta. Você pode encerrar um cluster que não sai do estado `STARTING` ou que não consegue concluir uma etapa.

Se quiser encerrar um cluster que possui proteção de encerramento, deve primeiro desativar essa proteção antes de encerrar o cluster. Os clusters podem ser encerrados usando o console AWS CLI, o ou programaticamente usando a API. `TerminateJobFlows`

Dependendo da configuração do cluster, pode levar entre 5 a 20 minutos para concluir o encerramento e liberar recursos alocados tais como instâncias do EC2.

> **Note**
>
> Você não pode reiniciar um cluster terminado, mas pode clonar um cluster terminado para reutilizar a configuração dele em um novo cluster. Para ter mais informações, consulte Clonar um cluster usando o console.

> **Important**
>
> O Amazon EMR usa o perfil de serviço do Amazon EMR e a função `AWSServiceRoleForEMRCleanup` para limpar recursos de cluster em sua conta que você não usa mais, como instâncias do Amazon EC2. Você deve incluir ações nas políticas de função para excluir ou encerrar os recursos. Caso contrário, o Amazon EMR não poderá realizar essas ações de limpeza, e você poderá ter custos com recursos não utilizados que permanecem no cluster.

Encerrrar um cluster com o console

Você pode terminar um ou mais clusters usando o console do Amazon EMR. As etapas para encerrar um cluster no console variam de acordo com o estado da proteção de encerramento, ou seja, se a proteção está ativada ou não. Para encerrar um cluster protegido, você deve primeiro desativar a proteção de encerramento.

New console

Terminar um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Escolha Clusters e, em seguida, selecione o cluster que você deseja encerrar.
3. No menu suspenso Ações, escolha Terminar cluster para abrir o prompt Terminar cluster.
4. No prompt, escolha Terminar. Dependendo da configuração do cluster, o encerramento pode demorar de cinco a dez minutos. Para obter mais informações sobre os clusters do Amazon EMR, consulte Terminar um cluster.

Old console

Terminar um cluster com a proteção contra término desativada usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione o cluster para encerrar. Você pode selecionar vários clusters e encerrá-los ao mesmo tempo.
3. Escolha Encerrar.
4. Quando solicitado, escolha Terminate (Encerrar).

   O Amazon EMR termina as instâncias no cluster e interrompe a gravação dos dados de log.

Terminar um cluster com proteção contra término ativada usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Na página Cluster List (Lista de clusters), selecione o cluster para encerrar. Você pode selecionar vários clusters e encerrá-los ao mesmo tempo.
3. Escolha Encerrar.

4. Quando solicitado, escolha Change (Alterar) para desativar a proteção de encerramento. Se você selecionou vários clusters, escolha Turn off all (Desativar tudo) para desativar a proteção de encerramento para todos os clusters de uma só vez.
5. Na caixa de diálogo Terminate clusters (Encerrar clusters), em Termination Protection (Proteção contra o encerramento), escolha Off (Desativado) e, em seguida, clique na marca de verificação para confirmar.
6. Clique em Terminate (Encerrar).

   O Amazon EMR termina as instâncias no cluster e interrompe a gravação dos dados de log.

Encerrar um cluster com a AWS CLI

Para encerrar um cluster desprotegido usando o AWS CLI

Para encerrar um cluster desprotegido usando o AWS CLI, use o `terminate-clusters` subcomando com o parâmetro --cluster-ids.

- Digite o comando a seguir para encerrar um cluster único e substitua *j-3KVXXXXXXX7UG* pelo ID do seu cluster.

  ```
  aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG
  ```

  Para encerrar vários clusters, digite o comando a seguir e substitua *j-3KVXXXXXXX7UG* e *j-WJ2XXXXXX8EU* pelos IDs do seu cluster.

  ```
  aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG j-WJ2XXXXXX8EU
  ```

  Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Para encerrar um cluster protegido usando o AWS CLI

Para encerrar um cluster protegido usando o AWS CLI, primeiro desative a proteção de encerramento usando o `modify-cluster-attributes` subcomando com o `--no-termination-protected` parâmetro. Em seguida, use o subcomando `terminate-clusters` com o parâmetro `--cluster-ids` para encerrá-lo.

1. Digite o comando a seguir para desativar a proteção de encerramento e substitua *j-3KVTXXXXXX7UG* pelo ID do seu cluster.

```
aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-protected
```

2. Para encerrar o cluster, digite o comando a seguir e substitua *j-3KVXXXXXXX7UG* pelo ID do seu cluster.

```
aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG
```

   Para encerrar vários clusters, digite o comando a seguir e substitua *j-3KVXXXXXXX7UG* e *j-WJ2XXXXXX8EU* pelos IDs do seu cluster.

```
aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG j-WJ2XXXXXX8EU
```

   Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte. https://docs.aws.amazon.com/cli/latest/reference/emr

Encerrrar um cluster com a API

A operação `TerminateJobFlows` interrompe o processamento da etapa, carrega dados de log do Amazon EC2 para o Amazon S3 (se configurado) e termina o cluster do Hadoop. Um cluster também é encerrado automaticamente se você definir `KeepJobAliveWhenNoSteps` como `False` em uma solicitação `RunJobFlows`.

Você pode usar esta ação para encerrar um único cluster ou uma lista de clusters usando os IDs de clusters.

Para obter mais informações sobre os parâmetros de entrada exclusivos de`TerminateJobFlows`, consulte _TerminateJobFlows. Para obter mais informações sobre os parâmetros genéricos na solicitação, consulte Common request parameters.

Clonar um cluster usando o console

Você pode usar o console do Amazon EMR para clonar um cluster, que faz uma cópia da configuração do cluster original para ser usada como base em um novo cluster.

> **Note**
>
> Reformulamos o console do Amazon EMR para torná-lo mais fácil de usar. É possível clonar clusters que usam a escalabilidade automática no novo console, mas você poderá criar novos clusters somente se desejar escalá-los manualmente ou usar o ajuste de escala gerenciado. Consulte Console do Amazon EMR para saber mais sobre as diferenças entre as experiências do console antigo e do novo.

New console

Clonar um cluster usando o novo console

1. Faça login no AWS Management Console e abra o console do Amazon EMR em https://console.aws.amazon.com/emr.
2. Em EMR no EC2, no painel de navegação esquerdo, escolha Clusters.
3. Clonar um cluster da lista de clusters
   a. Use as opções de pesquisa e de filtro para encontrar o cluster que você deseja clonar na visualização de lista.
   b. Marque a caixa de seleção à esquerda da linha do cluster que deseja clonar.
   c. A opção Clonar já estará disponível na parte superior da visualização da lista. Selecione Clonar para iniciar o processo de clonagem. Se o cluster tiver etapas configuradas, escolha Incluir etapas e Continuar para clonar as etapas junto com as outras configurações do cluster.
   d. Revise as configurações do novo cluster que foram copiadas do cluster clonado. Ajuste as configurações, se necessário. Quando a configuração do novo cluster estiver satisfatória, selecione Criar cluster para iniciar o novo cluster.
4. Clonar um cluster na página de detalhes do cluster
   a. Para navegar até a página de detalhes do cluster que você deseja clonar, selecione o ID do cluster na visualização da lista de clusters.
   b. Na parte superior da página de detalhes do cluster, selecione Clonar cluster no menu Ações para iniciar o processo de clonagem. Se o cluster tiver etapas configuradas, escolha Incluir etapas e Continuar para clonar as etapas junto com as outras configurações do cluster.

c. Revise as configurações do novo cluster que foram copiadas do cluster clonado. Ajuste as configurações, se necessário. Quando a configuração do novo cluster estiver satisfatória, selecione Criar cluster para iniciar o novo cluster.

Old console

Clonar um cluster usando o console antigo

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte Usar o console antigo.
2. Selecione Criar cluster.
3. Na página Cluster List (Lista de clusters), clique em um cluster a ser clonado.
4. Na parte superior da página Cluster Details (Detalhes do cluster), clique em Clone (Clonar).

   Na caixa de diálogo, selecione Yes (Sim) para incluir as etapas do cluster original no cluster clonado. Selecione No (Não), para clonar a configuração do cluster original sem incluir nenhuma das etapas.

   > **Note**
   >
   > Para clusters criados usando a AMI 3.1.1 e posterior (Hadoop 2.x) ou a AMI 2.4.8 e posterior (Hadoop 1.x), se você clonar um cluster e incluir as etapas, todas as etapas do sistema (como configurar o Hive) serão clonadas juntamente com as etapas enviadas pelo usuário, até um total de 1.000. Todas as etapas mais antigas que já não aparecem no histórico de etapas do console não podem ser clonadas. Para AMIs anteriores, somente 256 etapas podem ser clonadas (incluindo as etapas do sistema). Para ter mais informações, consulte Enviar trabalhos a um cluster.

5. A página Create Cluster (Criar cluster) é exibida com uma cópia da configuração do cluster original. Examine a configuração, faça as alterações necessárias e clique em Create Cluster (Criar cluster).

Automatizar clusters recorrentes usando o AWS Data Pipeline

AWS Data Pipeline é um serviço que automatiza a movimentação e a transformação dos dados. Você pode usá-lo para programar a movimentação de dados de entrada para o Amazon S3 e para programar a inicialização de clusters para processar dados. Por exemplo, considere o caso em que você tenha um servidor web gravando logs de tráfego. Se você quiser executar um cluster semanal para analisar os dados de tráfego, você pode usá-lo AWS Data Pipeline para programar esses clusters. AWS Data Pipeline é um fluxo de trabalho orientado por dados, de modo que uma tarefa (iniciar o cluster) pode depender de outra tarefa (mover os dados de entrada para o Amazon S3). Ele também tem uma funcionalidade de novas tentativas robusta.

Para obter mais informações sobre AWS Data Pipeline, consulte o Guia do AWS Data Pipeline desenvolvedor, especialmente os tutoriais sobre o Amazon EMR:

- Tutorial: iniciar um fluxo de trabalho do Amazon EMR
- Introdução: processe registros da web com o AWS Data Pipeline Amazon EMR e o Hive
- Tutorial: importação e exportação do Amazon DynamoDB usando AWS Data Pipeline

Solução de problemas de clusters

Um cluster EMR é executado em um ecossistema complexo que inclui software de código aberto, código de aplicativo personalizado e. Serviços da AWS Quando ocorre um problema com qualquer uma dessas partes, o cluster pode falhar ou levar mais tempo do que o esperado para conclusão. Os tópicos a seguir podem ajudar a identificar problemas com o cluster e como corrigi-los.

Tópicos

Ao desenvolver uma nova aplicação Hadoop, é recomendável habilitar a depuração e processar um subconjunto pequeno, mas representativo, de seus dados para testar a aplicação. Talvez você também queira executar o aplicativo step-by-step para testar cada etapa separadamente. Para obter mais informações, consulte Configurar registro em log e depuração do cluster e Etapa 5: testar o cluster passo a passo.

Que ferramentas estão disponíveis para a solução de problemas?

Para identificar e corrigir erros de cluster, use as ferramentas descritas nesta página. Talvez seja necessário inicializar algumas ferramentas ao iniciar o cluster. Outras ferramentas estão disponíveis para todos os clusters por padrão.

Tópicos

Visualizar detalhes do cluster do EMR

Você pode usar a API AWS Management Console AWS CLI, ou EMR para recuperar informações detalhadas sobre um cluster do EMR e a execução de trabalhos. Para obter mais informações sobre como usar o AWS Management Console e AWS CLI, consulteVisualizar o status e os detalhes do cluster.

Painel de detalhes do console do Amazon EMR

Na lista Clusters no console do Amazon EMR, você pode ver informações de alto nível sobre o status de cada cluster em sua conta e Região da AWS. A lista exibe todos os clusters ativos e terminados que você iniciou nos últimos dois meses. Na lista Clusters, você pode selecionar um Name (Nome) de cluster para visualizar detalhes do cluster. Essas informações são organizadas em diferentes categorias para facilitar a navegação.

As interfaces do usuário da aplicação disponíveis na página de detalhes do cluster podem ser para solucionar problemas de cluster. Ele fornece o status de aplicações do YARN e, para algumas, como aplicações Spark, você pode se aprofundar em diferentes métricas e facetas, como trabalhos, preparação e executores. Para ter mais informações, consulte Visualizar o histórico da aplicação. Esse atributo está disponível somente no Amazon EMR 5.8.0 e versões posteriores.

Interface de linha de comando do Amazon EMR

Você pode localizar detalhes sobre um cluster usando o `--describe` argumento AWS CLI with the.

API do Amazon EMR

Você pode localizar detalhes sobre um cluster na API usando a ação `DescribeJobFlows`.

Visualizar detalhes do erro do cluster do EMR

Quando um cluster do EMR é terminado com um erro, as APIs `DescribeCluster` e `ListClusters` retornam um código de erro e uma mensagem de erro. Para erros de cluster selecionados, a matriz de dados `ErrorDetail` pode ajudar a solucionar a falha.

Para obter uma lista de códigos de erro que incluam dados `ErrorDetail`, consulte Códigos de erro com ErrorDetail informações.

> **Note**
>
> Refinamos continuamente nossas mensagens de erro para você receber as informações mais recentes e pertinentes. Não é recomendável analisar o texto de `ErrorMessage` porque ele está sujeito a alterações.

Executar scripts e configurar processos do Amazon EMR

Como parte do processo de solução de problemas, talvez seja útil executar scripts personalizados no cluster ou visualizar e configurar processos de cluster.

Visualizar e reiniciar processos da aplicação

Pode ser útil visualizar os processos em execução no cluster para diagnosticar possíveis problemas. Você pode interromper e reiniciar os processos do cluster conectando-se ao nó principal do cluster. Para ter mais informações, consulte Visualizar e reiniciar processos do Amazon EMR e de aplicações (daemons).

Executar comandos e scripts sem uma conexão SSH

Para executar um comando ou script no cluster como uma etapa, você pode usar as ferramentas `command-runner.jar` ou `script-runner.jar` sem estabelecer uma conexão SSH com o nó principal. Para obter mais informações, consulte Run commands and scripts on an Amazon EMR cluster.

Exibir arquivos de log do

Tanto o Amazon EMR como o Hadoop geram arquivos de log conforme o cluster é executado. Você pode acessar esses arquivos de log de várias ferramentas diferentes, dependendo da configuração especificada ao iniciar o cluster. Para ter mais informações, consulte Configurar registro em log e depuração do cluster.

Arquivos de log no nó principal

Cada cluster publica arquivos de logs no diretório /mnt/var/log/ do nó principal. Esses arquivos de log estão disponíveis apenas enquanto o cluster está em execução.

Arquivos de log arquivados no Amazon S3

Se você executar o cluster e especificar um caminho de log do Amazon S3, o cluster copiará os arquivos de log armazenados em /mnt/var/log/ no nó principal para o Amazon S3 em intervalos de cinco minutos. Isso garante que você terá acesso aos arquivos de log, mesmo depois que o cluster for encerrado. Como os arquivos são arquivados em intervalos de 5 minutos, os últimos minutos de um cluster repentinamente encerrado podem não estar disponíveis.

Monitorar a performance do cluster do EMR

O Amazon EMR fornece várias ferramentas para monitorar a performance do cluster.

Interfaces Web do Hadoop

Cada cluster publica um conjunto de interfaces Web no nó principal que contêm informações sobre o cluster. Você pode acessar essas páginas da Web usando um túnel SSH para conectá-las ao nó principal. Para ter mais informações, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

CloudWatch métricas

Cada cluster reporta métricas para CloudWatch. CloudWatch é um serviço da web que rastreia métricas e que você pode usar para definir alarmes sobre essas métricas. Para ter mais informações, consulte Monitorando métricas do Amazon EMR com CloudWatch.

Visualizar e reiniciar processos do Amazon EMR e de aplicações (daemons)

Ao solucionar problemas em um cluster, você pode relacionar os processos em execução. Também pode ser útil interromper ou reiniciar processos. Por exemplo, você pode reiniciar os processos após alterar uma configuração ou observar um problema com um determinado processo após a análise de arquivos de log e mensagens de erro.

Há dois tipos de processos que são executados em um cluster: processos do Amazon EMR (por exemplo, controlador de instância e Log Pusher) e processos associados aos aplicativos instalados no cluster (por exemplo, e). hadoop-hdfs-namenode hadoop-yarn-resourcemanager

Para trabalhar com os processos diretamente em um cluster, primeiro é necessário conectar-se ao nó principal. Para ter mais informações, consulte Conectar-se a um cluster.

Visualizar processos em execução

O método que você usa para visualizar os processos que estão em execução em um cluster difere de acordo com a versão do Amazon EMR utilizada.

EMR 5.30.0 and 6.0.0 and later

Example : Listar todos os processos em execução

O exemplo a seguir usa `systemctl` e especifica `--type` para visualizar todos os processos.

```
systemctl --type=service
```

Example : Listar processos específicos

O exemplo a seguir lista todos os processos com nomes que contenham `hadoop`.

```
systemctl --type=service | grep -i hadoop
```

Resultado do exemplo:

```
 hadoop-hdfs-namenode.service           loaded active running Hadoop namenode
 hadoop-httpfs.service                  loaded active running Hadoop httpfs
 hadoop-kms.service                     loaded active running Hadoop kms
 hadoop-mapreduce-historyserver.service loaded active running Hadoop historyserver
 hadoop-state-pusher.service            loaded active running Daemon process that
 processes and serves EMR metrics data.
 hadoop-yarn-proxyserver.service        loaded active running Hadoop proxyserver
 hadoop-yarn-resourcemanager.service    loaded active running Hadoop resourcemanager
 hadoop-yarn-timelineserver.service     loaded active running Hadoop timelineserver
```

Example : Ver um relatório de status detalhado de um processo específico

O exemplo a seguir exibe um relatório de status detalhado do serviço `hadoop-hdfs-namenode`.

```
sudo systemctl status hadoop-hdfs-namenode
```

Resultado do exemplo:

```
hadoop-hdfs-namenode.service - Hadoop namenode
```

```
   Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
   Active: active (running) since Wed 2021-08-18 21:01:46 UTC; 26min ago
 Main PID: 9733 (java)
    Tasks: 0
   Memory: 1.1M
   CGroup: /system.slice/hadoop-hdfs-namenode.service
           # 9733 /etc/alternatives/jre/bin/java -Dproc_namenode -Xmx1843m -server -
XX:OnOutOfMemoryError=kill -9 %p ...

Aug 18 21:01:37 ip-172-31-20-123 systemd[1]: Starting Hadoop namenode...
Aug 18 21:01:37 ip-172-31-20-123 su[9715]: (to hdfs) root on none
Aug 18 21:01:37 ip-172-31-20-123 hadoop-hdfs-namenode[9683]: starting namenode,
 logging to /var/log/hadoop-hdfs/ha...out
Aug 18 21:01:46 ip-172-31-20-123 hadoop-hdfs-namenode[9683]: Started Hadoop
 namenode:[  OK  ]
Aug 18 21:01:46 ip-172-31-20-123 systemd[1]: Started Hadoop namenode.
Hint: Some lines were ellipsized, use -l to show in full.
```

EMR 4.x - 5.29.0

Example : Listar todos os processos em execução

O exemplo a seguir lista todos os processos que estão em execução.

```
initctl list
```

EMR 2.x - 3.x

Example : Listar todos os processos em execução

O exemplo a seguir lista todos os processos que estão em execução.

```
ls /etc/init.d/
```

Interromper e reiniciar processos

Depois de determinar quais processos estão em execução, você pode interrompê-los e reiniciá-los, se necessário.

EMR 5.30.0 and 6.0.0 and later

Example : Interromper um processo

O exemplo a seguir interrompe o processo `hadoop-hdfs-namenode`.

```
sudo systemctl stop hadoop-hdfs-namenode
```

Consulte o `status` para verificar se o processo foi interrompido.

```
sudo systemctl status hadoop-hdfs-namenode
```

Resultado do exemplo:

```
hadoop-hdfs-namenode.service - Hadoop namenode
  Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
  Active: failed (Result: exit-code) since Wed 2021-08-18 21:37:50 UTC; 8s ago
Main PID: 9733 (code=exited, status=143)
```

Example : Iniciar um processo

O exemplo a seguir inicia o processo `hadoop-hdfs-namenode`.

```
sudo systemctl start hadoop-hdfs-namenode
```

Consulte o status para verificar se o processo está em execução.

```
sudo systemctl status hadoop-hdfs-namenode
```

Resultado do exemplo:

```
hadoop-hdfs-namenode.service - Hadoop namenode
   Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
   Active: active (running) since Wed 2021-08-18 21:38:24 UTC; 2s ago
  Process: 13748 ExecStart=/etc/init.d/hadoop-hdfs-namenode start (code=exited,
 status=0/SUCCESS)
 Main PID: 13800 (java)
    Tasks: 0
   Memory: 1.1M
```

```
  CGroup: /system.slice/hadoop-hdfs-namenode.service
          # 13800 /etc/alternatives/jre/bin/java -Dproc_namenode -Xmx1843m -server
 -XX:OnOutOfMemoryError=kill -9 %p...
```

EMR 4.x - 5.29.0

Example : Interromper um processo em execução

O exemplo a seguir interrompe o serviço `hadoop-hdfs-namenode`.

```
sudo stop hadoop-hdfs-namenode
```

Example : Reiniciar um processo interrompido

O exemplo a seguir reinicia o serviço `hadoop-hdfs-namenode`. Você deve usar o comando `start` em vez de `restart`.

```
sudo start hadoop-hdfs-namenode
```

Example : Verificar o status do processo

O exemplo a seguir busca o status de `hadoop-hdfs-namenode`. Você pode usar o comando `status` para verificar se o processo foi interrompido ou iniciado.

```
sudo status hadoop-hdfs-namenode
```

EMR 2.x - 3.x

Example : Interromper um processo da aplicação

O exemplo a seguir interrompe o serviço `hadoop-hdfs-namenode`, que está associado à versão do Amazon EMR instalada no cluster.

```
sudo /etc/init.d/hadoop-hdfs-namenode stop
```

Example : Reiniciar um processo da aplicação

O exemplo de comando a seguir reinicia o processo `hadoop-hdfs-namenode`:

```
sudo /etc/init.d/hadoop-hdfs-namenode start
```

Example : Interromper um processo do Amazon EMR

O exemplo a seguir interrompe um processo, como instance-controller, que não esteja associado à versão do Amazon EMR no cluster.

```
sudo /sbin/stop instance-controller
```

Example : Reiniciar um processo do Amazon EMR

O exemplo a seguir reinicia um processo, como instance-controller, que não esteja associado à versão do Amazon EMR no cluster.

```
sudo /sbin/start instance-controller
```

> **Note**
>
> Os comandos `/sbin/start, stop` e `restart` são symlinks para `/sbin/intictl`. Para obter mais informações sobre `initctl`, consulte a página do manual do initctl digitando `man initctl` no prompt de comando.

Erros comuns do Amazon EMR

Às vezes, os clusters falham ou demoram para processar os dados. As seções a seguir listam alguns problemas comuns de cluster com sugestões para corrigi-los.

Tópicos

- AWS GovCloud Erros (Oeste dos EUA)
- Encontrar um cluster ausente

Códigos de erro com ErrorDetail informações

Quando um cluster do EMR é terminado com um erro, as APIs `DescribeCluster` e `ListClusters` retornam um código de erro e uma mensagem de erro. Para alguns erros de cluster, a matriz de dados `ErrorDetail` pode ajudar a solucionar a falha.

Os erros que incluem uma matriz `ErrorDetail` fornecem os seguintes detalhes:

`ErrorCode`

Um código de erro exclusivo que pode ser usado para acesso programático.

`ErrorData`

Uma lista de identificadores em pares de chave-valor que podem ser usados para programação ou pesquisa manual. Para obter descrições dos valores de `ErrorData` que um código de erro inclui, consulte a página de solução de problemas do código de erro.

`ErrorMessage`

Descrição do erro com um link para mais informações na documentação do Amazon EMR.

> **Note**
>
> Não é recomendável analisar o texto de `ErrorMessage` porque está sujeito a alterações.

Códigos de erro por categoria

- Códigos de erro de falha de bootstrap
- Códigos de erro internos
- Códigos de erro de falha de validação

Códigos de erro de falha de bootstrap

As seções a seguir fornecem informações sobre solução de problemas para códigos de erro de falha de bootstrap.

Tópicos

BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE

Visão geral

Quando um cluster é terminado com um erro `BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE`, uma ação de bootstrap falhou na instância primária. Para obter mais informações sobre ações de bootstrap, consulte Criar ações de bootstrap para instalar softwares adicionais.

Resolução

Para resolver esse erro, revise os detalhes retornados no erro da API, modifique o script de ação de bootstrap e crie um novo cluster com a ação de bootstrap atualizada.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária em que a ação de bootstrap falhou.

`bootstrap-action`

O número ordinal da ação de bootstrap com falha. Um script com um valor `bootstrap-action` de 1 é a primeira ação de bootstrap a ser executada na instância.

`return-code`

O código de retorno para a ação de bootstrap com falha.

`amazon-s3-path`

O local da ação de bootstrap com falha no Amazon S3.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para identificar e corrigir a causa raiz do erro de ação de bootstrap. Em seguida, inicie um novo cluster.

1. Analise os arquivos de log de ações de bootstrap no Amazon S3 para identificar a causa raiz da falha. Para saber mais sobre como visualizar os logs do Amazon EMR, consulte Exibir arquivos de log do .
2. Se você ativou os logs do cluster ao criar a instância, consulte o log `stdout` para obter mais informações. Você encontra o log `stdout` da ação de bootstrap neste local do Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

   Para obter mais informações sobre logs de clusters, consulte Configurar registro em log e depuração do cluster.
3. Para determinar a falha na ação de bootstrap, revise as exceções nos logs `stdout` e o valor `return-code` em `ErrorData`.
4. Use suas descobertas da etapa anterior para revisar a ação de bootstrap para que ela evite exceções ou consiga lidar com exceções normalmente quando elas ocorrerem.
5. Inicie um novo cluster com a ação de bootstrap atualizada.

BOOTSTRAP_FAILURE_BA_DOWNLOAD_FAILED_PRIMARY

Visão geral

Um cluster é encerrado com o erro `BOOTSTRAP_FAILURE_BA_DOWNLOAD_FAILED_PRIMARY` quando a instância primária não consegue baixar um script de ação de bootstrap no local do Amazon S3 especificado. As possíveis causas incluem:

- O arquivo de script de ação de bootstrap não está no local especificado do Amazon S3.
- O perfil de serviço para instâncias do Amazon EC2 no cluster (também chamada de perfil de instância do EC2 para o Amazon EMR) não tem permissões para acessar o bucket do Amazon S3

onde o script de ação de bootstrap reside. Para obter mais informações sobre perfis de serviço, consulte Perfil de serviço para instâncias do EC2 do cluster (perfil de instância do EC2).

Para obter mais informações sobre ações de bootstrap, consulte Criar ações de bootstrap para instalar softwares adicionais.

Resolução

Para resolver esse erro, certifique-se de que a instância primária tem o devido acesso ao script de ação de bootstrap.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária em que a ação de bootstrap falhou.

`bootstrap-action`

O número ordinal da ação de bootstrap com falha. Um script com um valor `bootstrap-action` de 1 é a primeira ação de bootstrap a ser executada na instância.

`amazon-s3-path`

O local da ação de bootstrap com falha no Amazon S3.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para identificar e corrigir a causa raiz do erro de ação de bootstrap. Em seguida, inicie um novo cluster.

Etapas de solução de problemas

1. Use o valor `amazon-s3-path` da matriz `ErrorData` para encontrar o script de ação de bootstrap relevante no Amazon S3.

2. Se você ativou os logs do cluster ao criar a instância, consulte o log `stdout` para obter mais informações. Você encontra o log `stdout` da ação de bootstrap neste local do Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

   Para obter mais informações sobre logs de clusters, consulte Configurar registro em log e depuração do cluster.
3. Para determinar a falha na ação de bootstrap, revise as exceções nos logs `stdout` e o valor `return-code` em `ErrorData`.
4. Use suas descobertas da etapa anterior para revisar a ação de bootstrap para que ela evite exceções ou consiga lidar com exceções normalmente quando elas ocorrerem.
5. Inicie um novo cluster com a ação de bootstrap atualizada.

BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY

Visão geral

O erro `BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY` indica que a instância primária não consegue encontrar o script de ação de bootstrap que a instância acabou de baixar no bucket do Amazon S3 especificado.

Resolução

Para resolver esse erro, confirme se a instância primária tem o devido acesso ao script de ação de bootstrap.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária em que a ação de bootstrap falhou.

`bootstrap-action`

O número ordinal da ação de bootstrap com falha. Um script com um valor `bootstrap-action` de 1 é a primeira ação de bootstrap a ser executada na instância.

`amazon-s3-path`

O local da ação de bootstrap com falha no Amazon S3.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para identificar e corrigir a causa raiz do erro de ação de bootstrap. Em seguida, inicie um novo cluster.

1. Para encontrar o script de ação de bootstrap relevante no Amazon S3, use o valor `amazon-s3-path` da matriz `ErrorData`.
2. Analise os arquivos de log de ações de bootstrap no Amazon S3 para identificar a causa raiz da falha. Para saber mais sobre como visualizar os logs do Amazon EMR, consulte Exibir arquivos de log do .

   > **Note**
   >
   > Se você não ativou os logs do cluster, será necessário criar um novo cluster com as mesmas configurações e ações de bootstrap. Para verificar se os logs do cluster estão ativados, consulte Configurar registro em log e depuração do cluster.

3. Analise o logs `stdout` de suas ações de bootstrap e confirme se não há processos personalizados que excluam arquivos na pasta `/emr/instance-controller/lib/bootstrap-actions` em suas instâncias primárias. Você encontra o log `stdout` da ação de bootstrap neste local do Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

4. Inicie um novo cluster com a ação de bootstrap atualizada.

Códigos de erro internos

As seções a seguir fornecem informações sobre solução de problemas para códigos de erro interno.

Tópicos

- INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ
- INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY
- INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY

INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ

Visão geral

Um cluster é terminado com um erro `INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ` quando a zona de disponibilidade selecionada não tem capacidade suficiente para atender à solicitação de tipo de instância do Amazon EC2. A sub-rede que você selecionou para um cluster determina a zona de disponibilidade. Para obter mais informações sobre sub-redes para o Amazon EMR, consulte Configurar redes.

Resolução

Para resolver esse erro, modifique as configurações de tipo de instância e crie um novo cluster com a solicitação atualizada.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`instance-type`

O tipo de instância que está fora da capacidade.

`availability-zone`

A zona de disponibilidade para a qual sua sub-rede é resolvida.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Execute estas etapas para identificar e corrigir a causa raiz do erro de configuração do cluster:

- Analise as práticas recomendadas para conexão de cluster. Consulte Práticas recomendadas para configuração de clusters no Guia de gerenciamento do Amazon EMR.

- Solucionar os problemas de inicialização e revisar a configuração. Consulte Solucionar problemas de lançamento de instâncias no Guia do usuário do Amazon EC2.
- Inicie um novo cluster com a configuração de cluster atualizada.

INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY

Visão geral

Um cluster é terminado com um erro `INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY` quando o Amazon EMR não consegue atender à solicitação de instância spot para o nó primário porque as instâncias não estão disponíveis até o preço spot máximo. Para obter mais informações, consulte Instâncias spot no Guia do usuário do Amazon EC2.

Resolução

Para resolver esse erro, especifique os tipos de instância do cluster que estejam dentro da meta de preço ou aumente o limite de preço para o mesmo tipo de instância.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária do cluster que falhou.

`instance-type`

O tipo de instância que está fora da capacidade.

`availability-zone`

A zona de disponibilidade em que a sub-rede reside.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para solucionar problemas da estratégia de configuração de cluster e iniciar um novo cluster:

1. Analise as práticas recomendadas para instâncias spot do Amazon EC2 e analise a estratégia de configuração de cluster. Para obter mais informações, consulte as melhores práticas para o EC2 Spot no Guia do usuário do Amazon EC2 e. Práticas recomendadas para configuração de clusters
2. Modifique as configurações de tipo de instância ou zona de disponibilidade e crie um novo cluster com a solicitação atualizada.
3. Se o problema persistir, use a capacidade sob demanda para a instância primária.

INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY

Visão geral

Um cluster é terminado com um erro `INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY` quando não há capacidade suficiente para atender a uma solicitação de instância spot para o nó primário. Para obter mais informações, consulte Instâncias spot no Guia do usuário do Amazon EC2.

Resolução

Para resolver esse erro, especifique os tipos de instância do cluster que estejam dentro da meta de preço ou aumente o limite de preço para o mesmo tipo de instância.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`primary-instance-id`

O ID da instância primária do cluster que falhou.

`instance-type`

O tipo de instância que está fora da capacidade.

`availability-zone`

A zona de disponibilidade para a qual sua sub-rede é resolvida.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Execute as etapas a seguir para solucionar problemas da estratégia de configuração de cluster e iniciar um novo cluster:

1. Analise as práticas recomendadas para instâncias spot do Amazon EC2 e analise a estratégia de configuração de cluster. Para obter mais informações, consulte as melhores práticas para o EC2 Spot no Guia do usuário do Amazon EC2 e. Práticas recomendadas para configuração de clusters
2. Modifique as configurações de tipo de instância e crie um novo cluster com a solicitação atualizada.
3. Se o problema persistir, use a capacidade sob demanda para a instância primária.

Códigos de erro de falha de validação

As seções a seguir fornecem informações sobre solução de problemas para códigos de erro de falha de validação.

Tópicos

- VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC
- VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC
- VALIDATION_ERROR_INVALID_SSH_KEY_NAME
- VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED

VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC

Visão geral

Quando o cluster e as sub-redes referenciadas para o cluster pertencem a diferentes nuvens privadas virtuais (VPCs), o cluster é terminado com um erro `VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC`. Você pode iniciar clusters usando o Amazon EMR com a configuração de frotas de instâncias em sub-redes em uma VPC. Para obter mais informações sobre frotas de instâncias, consulte Configurar frotas de instâncias no Guia de gerenciamento do Amazon EMR.

Resolução

Para resolver esse erro, use sub-redes que pertençam à mesma VPC do cluster.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`vpc`

Para cada par sub-rede:VPC, o ID da VPC à qual a sub-rede pertence.

`subnet`

Para cada par sub-rede:VPC, o ID da sub-rede.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Realize as etapas a seguir para identificar e corrigir o erro:

1. Analise os IDs de sub-rede listados na matriz `ErrorData` e confirme se eles pertencem à VPC na qual você deseja iniciar o cluster do EMR.
2. Modifique as configurações de sub-rede. Use um dos métodos a seguir para encontrar todas as sub-redes públicas e privadas em uma VPC.
   - Acesse o console da Amazon VPC. Escolha Sub-redes e liste todas as sub-redes que residem dentro do Região da AWS seu cluster. Para encontrar somente sub-redes públicas ou privadas, aplique o filtro de Atribuir automaticamente endereços IPv4 públicos. Para encontrar e selecionar sub-redes na VPC que o cluster usa, use a opção Filtrar por VPC. Para obter mais informações sobre como criar sub-redes, consulte Criar uma sub-rede no Guia do usuário da Amazon Virtual Private Cloud.
   - Use o AWS CLI para encontrar todas as sub-redes públicas e privadas disponíveis na VPC que seu cluster usa. Para obter mais informações, consulte a API describe-subnets. Para criar novas sub-redes em uma VPC, consulte a API create-subnet.
3. Inicie um novo cluster com sub-redes da mesma VPC do cluster.

VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC

Visão geral

Quando o cluster e os grupos de segurança que você atribui ao cluster pertencem a diferentes nuvens privadas virtuais (VPCs), o cluster é terminado com um erro `VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC`. Para obter mais informações sobre grupos de segurança, consulte Especificar grupos de segurança gerenciados pelo Amazon EMR e adicionais e Controle do tráfego de rede com grupos de segurança.

Resolução

Para resolver esse erro, use grupos de segurança que pertençam à mesma VPC do cluster.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`vpc`

Para cada par security-group:VPC, o ID da VPC à qual o grupo de segurança pertence.

`security-group`

Para cada par security-group:VPC, o ID do grupo de segurança.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Realize as etapas a seguir para identificar e corrigir o erro:

1. Analise os IDs de grupos de segurança listados na matriz `ErrorData` e confirme se eles pertencem à VPC na qual você deseja iniciar o cluster do EMR.
2. Acesse o console da Amazon VPC. Escolha Grupos de segurança para listar todos os grupos de segurança da selecionada. Encontre os grupos de segurança da mesma VPC que o cluster e modifique a configuração do grupo de segurança.
3. Inicie um novo cluster com grupos de segurança da mesma VPC do cluster.

VALIDATION_ERROR_INVALID_SSH_KEY_NAME

Visão geral

Um cluster é terminado com um erro `VALIDATION_ERROR_INVALID_SSH_KEY_NAME` quando você usa um par de chaves do Amazon EC2 que não é válido para SSH na instância primária. O nome do par de chaves pode estar incorreto ou o par de chaves pode não existir na solicitação Região da AWS. Para obter mais informações sobre pares de chaves, consulte os pares de chaves do Amazon EC2 e as instâncias Linux no Guia do usuário do Amazon EC2.

Resolução

Para resolver esse erro, crie um novo cluster com um nome de par de chaves SSH válido.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`ssh-key`

O nome do par de chaves SSH fornecido ao criar o cluster.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Realize as etapas a seguir para identificar e corrigir o erro:

1. Verifique o arquivo *`keypair`*.pem e confirme se ele corresponde ao nome da chave SSH que você vê no console do Amazon EMR.
2. Navegue até o console do Amazon EC2. Verifique se o nome da chave SSH que você usou está disponível no nome Região da AWS que seu cluster usa. Você pode encontrar seu Região da AWS próximo ID de conta na parte superior do AWS Management Console.
3. Inicie um novo cluster com um nome de chave SSH válido.

VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED

Visão geral

Um cluster é terminado com um erro `VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED` quando as Região da AWS e as zonas de disponibilidade do cluster não oferecem suporte ao tipo de instância especificado para um ou mais grupos de instâncias. Talvez o Amazon EMR ofereça suporte a um tipo de instância em uma zona de disponibilidade dentro de uma região, mas não em outra. A sub-rede selecionada para um cluster determina a zona de disponibilidade na região. Para obter uma lista de tipos de instância e regiões com suporte do Amazon EMR, consulte Tipos de instâncias compatíveis.

Resolução

Para resolver esse erro, especifique os tipos de instância para seu cluster compatíveis com o Amazon EMR na região e na zona de disponibilidade em que o cluster é solicitado.

Para solucionar o problema do cluster do EMR com falha, consulte as informações de `ErrorDetail` retornadas das APIs `DescribeCluster` e `ListClusters`. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações. A matriz `ErrorData` em `ErrorDetail` retorna as seguintes informações para o código de erro:

`instance-types`

A lista de tipos de instância com suporte.

`availability-zones`

A lista de zonas de disponibilidade para a qual sua sub-rede é resolvida.

`public-doc`

O URL público da documentação do código de erro.

Etapas a serem executadas

Realize as etapas a seguir para identificar e corrigir o erro:

1. Use o AWS CLI para recuperar os tipos de instância disponíveis em uma zona de disponibilidade. Para fazer isso, você pode usar o `ec2 describe-instance-type-offerings` comando para filtrar os tipos de instância disponíveis por local (Região da AWS ou

zona de disponibilidade). Por exemplo, o comando a seguir retorna os tipos de instância que são oferecidos na AZ especificada, *us-east-2a*.

```
aws ec2 describe-instance-type-offerings --location-type "availability-
zone" --filters Name=location,Values=us-east-2a --region us-east-2 --query
 "InstanceTypeOfferings[*].[InstanceType]" --output text | sort
```

Para saber mais sobre como descobrir tipos de instância disponíveis, consulte Localizar um tipo de instância do Amazon EC2.

2. Após determinar os tipos de instância que estão disponíveis na mesma região e zona de disponibilidade do cluster, escolha uma das seguintes resoluções para continuar:

   a. Crie um novo cluster e escolha uma sub-rede para o cluster que esteja em uma zona de disponibilidade onde o tipo de instância que você selecionou esteja disponível e tenha suporte do Amazon EMR.

   b. Crie um novo cluster na mesma região e sub-rede do Amazon EC2 do cluster que falhou, mas com um tipo de instância compatível com o Amazon EMR naquele local.

Para obter uma lista de tipos de instância e regiões com suporte do Amazon EMR, consulte Tipos de instâncias compatíveis. Para comparar os recursos dos tipos de instância do EC2, consulte Tipos de instância do Amazon EC2.

Erros de recursos

Os seguintes erros são geralmente causados pela restrição de recursos no cluster.

Tópicos

O cluster é terminado com NO_SLAVE_LEFT e nós centrais FAILED_BY_MASTER

Geralmente, isso acontece porque a proteção contra encerramento está desabilitada, e todos os nós core excedem a capacidade de armazenamento em disco, conforme especificado por um limite de utilização máxima na classificação de configuração `yarn-site`, que corresponde ao arquivo `yarn-site.xml`. Esse valor é 90%, por padrão. Quando a utilização do disco de um nó principal excede o limite de utilização, o serviço de NodeManager integridade do YARN relata o nó como. `UNHEALTHY` Enquanto ele estiver nesse estado, o Amazon EMR coloca o nó na lista de negação e não aloca contêineres YARN a ele. Se o nó permanecer não íntegro por 45 minutos, o Amazon EMR marcará a instância associada do Amazon EC2 para término como `FAILED_BY_MASTER`. Quando todas as instâncias do Amazon EC2 associadas a nós centrais são marcadas para término, o cluster é terminado com o status `NO_SLAVE_LEFT` porque não há recursos para executar trabalhos.

Ultrapassar a utilização de disco em um nó core pode causar uma reação em cadeia. Se um único nó exceder o limite de utilização de disco por causa do HDFS, outros nós também poderão estar perto do limite. O primeiro nó excede o limite de utilização de disco, então o Amazon EMR o coloca na lista de negação. Isso aumenta a carga de utilização do disco para os nós restantes, pois eles começam a replicar dados do HDFS entre si que foram perdidos no nó que está na lista de negação. Cada nó subsequentemente entra no status `UNHEALTHY` da mesma maneira, e o cluster por fim é encerrado.

Práticas recomendadas e orientações

Configurar o hardware do cluster com armazenamento adequado

Ao criar um cluster, certifique-se de que haja nós core suficientes e que cada um tenha um armazenamento de instâncias adequado e volumes de armazenamento do EBS para HDFS. Para ter mais informações, consulte Calcular a capacidade necessária do HDFS de um cluster. Você também pode adicionar instâncias core aos grupos de instâncias existentes manualmente ou usando a escalabilidade automática. As novas instâncias têm a mesma configuração de armazenamento que outras instâncias no grupo de instâncias. Para ter mais informações, consulte Usar ajuste de escala de clusters.

Habilitar a proteção contra encerramento

Habilitar a proteção contra encerramento. Dessa forma, se um nó central estiver na lista de negação, você poderá se conectar à instância associada do Amazon EC2 usando SSH para solucionar problemas e recuperar dados. Se você habilitar a proteção contra término, lembre-se de que o Amazon EMR não substitui a instância do Amazon EC2 por uma nova instância. Para ter mais informações, consulte Usar a proteção contra término.

Crie um alarme para a UnhealthyNodes CloudWatch métrica MR

Essa métrica informa o número de nós com o status UNHEALTHY. É equivalente à métrica do YARN `mapred.resourcemanager.NoOfUnhealthyNodes`. Você pode configurar uma notificação desse alarme para avisá-lo de nós não íntegros antes que o limite de 45 minutos seja atingido. Para ter mais informações, consulte Monitorando métricas do Amazon EMR com CloudWatch.

Ajustar as configurações com yarn-site

As configurações a seguir podem ser ajustadas de acordo com os requisitos do aplicativo. Por exemplo, talvez você queira aumentar o limite de utilização de disco onde um nó informa UNHEALTHY ao aumentar o valor de `yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage`.

Você pode definir esses valores ao criar um cluster usando a classificação de configuração `yarn-site`. Para obter mais informações, consulte Configuring applications no Guia de lançamento do Amazon EMR. Você também pode se conectar às instâncias do Amazon EC2 associadas a nós centrais usando SSH e, em seguida, adicionar os valores `/etc/hadoop/conf.empty/yarn-site.xml` usando um editor de texto. Depois de fazer a alteração, você deve reiniciar hadoop-yarn-nodemanager conforme mostrado abaixo.

> **Important**
>
> Quando você reinicia o NodeManager serviço, os contêineres ativos do YARN são eliminados, a menos que `yarn.nodemanager.recovery.enabled` esteja configurado para `true` usar a classificação de `yarn-site` configuração ao criar o cluster. Você também deve especificar o diretório no qual armazenar um estado de contêiner usando a propriedade `yarn.nodemanager.recovery.dir`.

```
sudo /sbin/stop hadoop-yarn-nodemanager
sudo /sbin/start hadoop-yarn-nodemanager
```

Para obter mais informações sobre as propriedades `yarn-site` atuais e valores padrão, consulte Configurações padrão do YARN na documentação do Apache Hadoop.

Propriedade	Valor padrão	Descrição
yarn.nodemanager.disk-health-checker.interval-ms	120000	A frequência (em segundos) em que o verificador de integridade do disco é executado.
yarn.nodemanager.disk-health-checker.min-healthy-disks	0.25	A fração mínima do número de discos que devem estar íntegros NodeManager para lançar novos contêineres. Isso corresponde a yarn.nodemanager.local-dirs (por padrão, `/mnt/yarn` no Amazon EMR) e yarn.nodemanager.log-dirs (por padrão `/var/log/hadoop-yarn/containers`, que apresenta um link simbólico para `mnt/var/log/hadoop-yarn/containers` no Amazon EMR).
`yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage`	90.0	A porcentagem máxima de utilização de espaço em disco permitido depois que um disco é marcado como inválido. Os valores variam de 0,0 a 100,0. Se o valor for maior ou igual a 100, NodeManager verificará se há um disco cheio. Isso se aplica a `yarn-nodemanager.local-dirs` e a `yarn.nodemanager.log-dirs`.

Propriedade	Valor padrão	Descrição
`yarn.nodemanager.disk-health-checker.min-free-space-per-disk-mb`	0	O espaço mínimo que deve estar disponível em um disco para que ele seja usado. Isso se aplica a `yarn-nodemanager.local-dirs` e a `yarn.nodemanager.log-dirs` .

Não é possível replicar os blocos, só foi possível replicar para zero nós.

O erro: “Não é possível replicar os blocos, só foi possível replicar para zero nós”. normalmente ocorre quando o cluster não tem armazenamento HDFS suficiente. Esse erro ocorre quando você gera no seu cluster uma quantidade de dados maior do que o HDFS pode armazenar. Você verá esse erro somente enquanto o cluster estiver em execução, porque quando o trabalho é terminado, ele libera o espaço que o HDFS estava usando.

A quantidade de espaço disponível no HDFS para um cluster depende do número e do tipo de instâncias do Amazon EC2 que são usadas como nós centrais. Nós de tarefa não são usados para armazenamento HDFS. Todo o espaço em disco em cada instância do Amazon EC2, incluindo os volumes de armazenamento do EBS anexados, está disponível para o HDFS. Para obter mais informações sobre a quantidade de armazenamento local para cada tipo de instância do EC2, consulte Tipos e famílias de instâncias no Guia do usuário do Amazon EC2.

O outro fator que pode afetar a quantidade de espaço disponível no HDFS é o fator de replicação, que é o número de cópias de cada bloco de dados que são armazenadas no HDFS por redundância. O fator de replicação aumenta de acordo com o número de nós no cluster: são 3 cópias de cada bloco de dados para um cluster com 10 ou mais nós, 2 cópias de cada bloco para um cluster com 4 a 9 nós e 1 cópia (sem redundância) para clusters com 3 ou menos nós. O total de espaço disponível no HDFS é dividido pelo fator de replicação. Em alguns casos, como por exemplo, com o aumento do número de nós de 9 para 10, o aumento no fator de replicação pode realmente fazer com que a quantidade de espaço disponível no HDFS diminua.

Por exemplo, um cluster com 10 nós core do tipo m1.large teria 2833 GB de espaço disponível para o HDFS ((10 nós X 850 GB por nó)/fator de replicação de 3).

Se o seu cluster exceder a quantidade de espaço disponível no HDFS, você pode adicionar mais nós core ao cluster ou usar a compactação de dados para criar mais espaço no HDFS. Se o cluster pode ser interrompido e reiniciado, você pode considerar o uso dos nós centrais de um tipo de instância maior do Amazon EC2. Você também deve considerar um ajuste no fator de replicação. Observe, no entanto, que a redução do fator de replicação diminui a redundância dos dados do HDFS e, consequentemente, a capacidade do seu cluster para recuperar blocos perdidos ou corrompidos do HDFS.

EC2 QUOTA EXCEEDED

Se uma mensagem `EC2 QUOTA EXCEEDED` for exibida, pode haver várias causas. Dependendo das diferenças na configuração, pode demorar entre 5 a 20 minutos para que clusters anteriores sejam encerrados totalmente e liberem os recursos alocados. Se você está recebendo um erro `EC2 QUOTA EXCEEDED` ao tentar iniciar um cluster, pode ser que os recursos de um cluster recém-encerrado ainda não tenham sido liberados. Essa mensagem também pode ser causada pelo redimensionamento de um grupo ou frota de instâncias para um tamanho de destino maior do que a cota de instâncias atual da conta. Isso pode acontecer manualmente ou automaticamente por meio de escalabilidade automática.

Considere as opções a seguir para resolver o problema:

- Siga as instruções descritas em AWS service quotas no Referência geral da Amazon Web Services para solicitar um aumento do limite de serviço. Para algumas APIs, configurar um CloudWatch evento pode ser uma opção melhor do que aumentar os limites. Para obter mais detalhes, consulte Quando configurar eventos do EMR em CloudWatch.
- Se um ou mais clusters em execução não estiverem na capacidade, redimensione os grupos de instâncias ou reduza as capacidades de destino nas frotas de instâncias para os clusters em execução.
- Crie clusters com um número menor de instâncias do EC2 ou reduza a capacidade de destino.

Muitas falhas de busca

A presença de mensagens de erro "Too many fetch-failures (Excesso de falhas de busca)" ou "Error reading task output (Erro ao ler a saída da tarefa)" nas etapas ou em logs de tentativas de tarefas indica que a tarefa em execução está dependendo da saída de uma outra tarefa. Isso geralmente ocorre quando uma tarefa é colocada na fila de execução e necessita da saída de uma ou mais tarefas de mapeamento, e essa saída ainda não está disponível.

Há vários motivos pelos quais a saída pode não estar disponível:

- A tarefa de pré-requisito ainda está em processamento. Essa geralmente é uma tarefa de mapeamento.
- Os dados podem estar indisponíveis devido à conectividade de rede ruim, se os dados estiverem localizados em uma instância diferente.
- Se o HDFS estiver sendo usado para recuperar a saída, pode haver um problema com o HDFS.

A causa mais comum deste erro é que a tarefa anterior ainda está em processamento. Isso é mais provável se os erros estão ocorrendo quando as tarefas de redução estão sendo executadas pela primeira vez. Você pode verificar se é esse o caso examinando o log do syslog para a etapa do cluster que está gerando o erro. Se o syslog mostra que ambas as tarefas de mapeamento e redução estão em andamento, isso indica que a fase de redução foi iniciada e, ao mesmo tempo, há tarefas de mapeamento que ainda não foram concluídas.

Um item a ser pesquisado nos logs é a porcentagem de andamento do mapeamento que vai até 100% e, em seguida, cai para um valor mais baixo. Quando a porcentagem está em 100%, isso não significa que todas as tarefas de mapeamento foram concluídas. Isto significa simplesmente que o Hadoop está executando todas as tarefas de mapeamento. Se esse valor voltar a ficar abaixo de 100%, isso significa que uma tarefa de mapeamento falhou e, dependendo da configuração, o Hadoop pode tentar reprogramar a tarefa. Se a porcentagem do mapa permanecer em 100% nos registros, observe as CloudWatch métricas, especificamente`RunningMapTasks`, para verificar se a tarefa do mapa ainda está sendo processada. Você também pode encontrar essas informações usando a interface da web do Hadoop no nó principal.

Se você está vendo esse problema, pode tentar várias ações:

- Inclua instruções na fase de redução para esperar mais antes de iniciar. Você pode fazer isso alterando a definição da configuração do Hadoop mapred.reduce.slowstart.completed.maps para um tempo maior. Para ter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais.
- Iguale a contagem de reducers com a capacidade total de reducers do cluster. Você pode fazer isso ajustando a definição de configuração do Hadoop mapred.reduce.tasks de acordo com o trabalho.
- Use um código de classe de combiner para minimizar o número de saídas que precisam ser obtidas.

- Verifique se não há problemas com o serviço do Amazon EC2 que estejam afetando a performance de rede do cluster. Você pode fazer isso usando o Painel de status dos serviços.
- Analise os recursos de CPU e memória das instâncias no seu cluster para assegurar-se de que o processamento dos dados não está degradando os recursos dos seus nós. Para ter mais informações, consulte Configurar o hardware e as redes do cluster.
- Verifique a versão da Imagem de máquina da Amazon (AMI) usada no cluster do Amazon EMR. Se a versão estiver entre a 2.3.0 e a 2.4.4, ambas incluídas, atualize para uma versão mais recente. As versões da AMI desse intervalo especificado usam uma versão do Jetty que pode falhar ao produzir uma saída da fase de mapeamento. O erro de busca ocorre quando os reducers não conseguem obter uma saída da fase de mapeamento.

  O Jetty é um servidor de HTTP de código aberto usado para estabelecer a comunicação entre máquinas em um cluster do Hadoop.

O arquivo pode ser replicado somente para 0 nós em vez de 1

Quando um arquivo é gravado no HDFS, ele é replicado para vários nós core. Quando você vê esse erro, isso significa que o NameNode daemon não tem nenhuma DataNode instância disponível para gravar dados no HDFS. Em outras palavras, a replicação de blocos não está sendo realizada. Esse erro pode ser causado por vários problemas:

- O sistema de arquivos do HDFS pode estar com o espaço esgotado. Esta é a causa mais provável.
- DataNode as instâncias podem não estar disponíveis quando o trabalho foi executado.
- DataNode as instâncias podem ter sido bloqueadas de se comunicar com o nó principal.
- As instâncias no grupo de instâncias core podem não estar disponíveis.
- Podem estar faltando permissões. Por exemplo, o JobTracker daemon pode não ter permissões para criar informações do rastreador de tarefas.
- A configuração do espaço reservado para uma DataNode instância pode ser insuficiente. Verifique se esse é o caso, examinando a definição da configuração de dfs.datanode.du.reserved.

Para verificar se esse problema é causado pela falta de espaço em disco do HDFS, veja a `HDFSUtilization` métrica em CloudWatch. Se o valor for muito alto, você pode adicionar mais nós core ao cluster. Se você tem um cluster que acha que pode ficar sem espaço em disco no HDFS, você pode configurar um alarme CloudWatch para alertá-lo quando o valor de `HDFSUtilization`

subir acima de um determinado nível. Para obter mais informações, consulte Redimensionar manualmente um cluster em execução e Monitorando métricas do Amazon EMR com CloudWatch.

Se o HDFS ficar sem espaço não fosse o problema, verifique os registros, os DataNode NameNode registros e a conectividade de rede em busca de outros problemas que poderiam ter impedido o HDFS de replicar dados. Para ter mais informações, consulte Exibir arquivos de log do .

Negar deny-listed

O NodeManager daemon é responsável por lançar e gerenciar contêineres nos nós principais e de tarefas. Os contêineres são alocados ao NodeManager daemon pelo ResourceManager daemon executado no nó principal. Ele ResourceManager monitora o NodeManager nó por meio de um batimento cardíaco.

Há algumas situações em que o ResourceManager daemon deny lista uma NodeManager, removendo-a do pool de nós disponíveis para processar tarefas:

- Se o não NodeManager tiver enviado uma pulsação ao ResourceManager daemon nos últimos 10 minutos (600.000 milissegundos). Esse intervalo de tempo pode ser configurado usando a definição da configuração `yarn.nm.liveness-monitor.expiry-interval-ms`. Para obter mais informações sobre a alteração das definições de configuração do Yarn, consulte Configuring applications no Guia de lançamento do Amazon EMR.
- NodeManager verifica a integridade dos discos determinada por `yarn.nodemanager.local-dirs` e. `yarn.nodemanager.log-dirs` As verificações incluem permissões e espaço livre em disco (< 90%). Se um disco falhar na verificação, ele para de NodeManager usar esse disco específico, mas ainda informa o status do nó como íntegro. Se vários discos falharem na verificação, o nó será reportado como não íntegro ResourceManager e os novos contêineres não serão atribuídos ao nó.

O mestre do aplicativo também pode negar a lista de um NodeManager nó se ele tiver mais de três tarefas com falha. Você pode aumentar esse valor usando o parâmetro de configuração `mapreduce.job.maxtaskfailures.per.tracker`. Outras definições de configuração que você pode alterar controlam o número de tentativas para uma tarefa antes de marcá-la como falha: `mapreduce.map.max.attempts` para tarefas de mapeamento e `mapreduce.reduce.maxattempts` para tarefas de redução. Para obter mais informações sobre a alteração das definições de configuração, consulte Configuring applications no Guia de lançamento do Amazon EMR.

Erros de controle de utilização

Os erros “Throttled from *Amazon EC2* while launching cluster” e “Failed to provision instances due to throttling from *Amazon EC2*” ocorrem quando o Amazon EMR não consegue concluir uma solicitação porque outro serviço limitou a atividade. O Amazon EC2 é a origem mais comum de erros de controle de utilização, mas outros serviços podem ocasionar esses erros. Os limites de serviço da AWS se aplicam por região para melhorar a performance, e um erro de controle de utilização indica que você excedeu o limite de serviço da conta naquela região.

Possíveis causas

A origem mais comum de erros de controle de utilização do Amazon EC2 é um grande número de instâncias do cluster sendo iniciadas, de modo que o limite de serviço para instâncias do EC2 é excedido. As instâncias do cluster podem ser executadas pelos seguintes motivos:

- Novos clusters são criados.
- Os clusters são redimensionados manualmente. Para ter mais informações, consulte Redimensionar manualmente um cluster em execução.
- Os grupos de instâncias em um cluster adicionam instâncias (expandem) como resultado de uma regra de escalabilidade automática. Para ter mais informações, consulte Noções básicas sobre as regras de ajuste de escala automático.
- As frotas de instâncias em um cluster adicionam instâncias para atender a uma maior capacidade de destino. Para ter mais informações, consulte Configurar frotas de instâncias.

Também é possível que a frequência ou tipo de solicitação de API sendo feita ao Amazon EC2 cause erros de controle de utilização. Para obter mais informações sobre como o Amazon EC2 limita solicitações de API, consulte Query API request rate na Amazon EC2 API Reference.

Soluções

Considere as seguintes soluções:

- Siga as instruções descritas em AWS service quotas no Referência geral da Amazon Web Services para solicitar um aumento do limite de serviço. Para algumas APIs, configurar um CloudWatch evento pode ser uma opção melhor do que aumentar os limites. Para obter mais detalhes, consulte Quando configurar eventos do EMR em CloudWatch.
- Se você tiver clusters são executados no mesmo agendamento (por exemplo, no começo da hora) considere intercalar os horários de início.

- Se tiver clusters que são dimensionados para picos de demanda, e você periodicamente tiver capacidade de instância, considere especificar a escalabilidade automática para adicionar e remover instâncias sob demanda. Dessa forma, as instâncias serão usadas de forma mais eficiente e, dependendo do perfil de demanda, menos instâncias poderão ser solicitadas em um determinado momento em uma conta. Para ter mais informações, consulte Usar o ajuste de escala automático com uma política personalizada para grupos de instâncias.

Tipo de instância sem suporte

Se você criar um cluster e ele falhar com a mensagem de erro “O tipo de instância solicitado não *InstanceType* é suportado na zona de disponibilidade solicitada”, significa que você criou o cluster e especificou um tipo de instância para um ou mais grupos de instâncias que não é suportado pelo Amazon EMR na região e na zona de disponibilidade em que o cluster foi criado. O Amazon EMR pode oferecer suporte a um tipo de instância em uma zona de disponibilidade de uma região e não em outra. A sub-rede selecionada para um cluster determina a Zona de disponibilidade na região.

Solução

Determine os tipos de instância disponíveis em uma zona de disponibilidade usando o AWS CLI

- Use o comando `ec2 run-instances` com a opção `--dry-run`. No exemplo abaixo, substitua *m5.xlarge* pelo tipo de instância que você deseja usar, *ami-035be7bafff33b6b6* pela AMI associada a esse tipo de instância e *subnet-12ab3c45* por uma sub-rede na zona de disponibilidade você deseja consultar.

  ```
  aws ec2 run-instances --instance-type m5.xlarge --dry-run --image-
  id ami-035be7bafff33b6b6 --subnet-id subnet-12ab3c45
  ```

  Para obter instruções sobre como encontrar um ID de AMI, consulte Encontre uma AMI do Linux. Para encontrar um ID de sub-rede, você pode usar o comando describe-subnets.

Para saber mais sobre como descobrir tipos de instância disponíveis, consulte Localizar um tipo de instância do Amazon EC2.

Depois de determinar os tipos de instâncias disponíveis, você pode fazer o seguinte:

- Crie o cluster na mesma região e sub-rede do EC2 e selecione um tipo de instância diferente com recursos semelhantes que a escolha inicial. Para obter uma lista dos tipos de instâncias

compatíveis, consulte Tipos de instâncias compatíveis. Para comparar recursos de tipos de instância do EC2, consulte Tipos de instância do Amazon EC2.

- Selecione uma sub-rede para o cluster em uma zona de disponibilidade onde o tipo de instância esteja disponível e tenha suporte do Amazon EMR.

O EC2 está fora da capacidade

Um erro “O EC2 está sem capacidade para *`InstanceType`*” ocorre quando você tenta criar um cluster ou adicionar instâncias a um cluster em uma zona de disponibilidade que não tem mais do tipo de instância EC2 especificado. A sub-rede que você selecionou para um cluster determina a zona de disponibilidade.

Para criar um cluster, siga um destes procedimentos:

- Especificar outro tipo de instância com recursos semelhantes
- Criar o cluster em outra região
- Selecione uma sub-rede em uma zona de disponibilidade em que o tipo de instância desejado possa estar disponível.

Para adicionar instância a um cluster em execução, realize uma destas ações:

- Modifique as configurações do grupo de instâncias ou as configurações da frota de instâncias para adicionar os tipos de instância disponíveis com recursos semelhantes. Para obter uma lista dos tipos de instâncias compatíveis, consulte Tipos de instâncias compatíveis. Para comparar recursos de tipos de instância do EC2, consulte Tipos de instância do Amazon EC2.
- Termine o cluster e o recrie em uma região e zona de disponibilidade em que o tipo de instância está disponível.

Erros de entrada e saída

Os erros a seguir são comuns em operações de entrada e saída do cluster.

Tópicos

- O caminho para o Amazon Simple Storage Service (Amazon S3) com pelo menos três barras?
- Você está tentando, recursivamente, desviar diretórios de entrada?
- Seu diretório de saída já existe?

O caminho para o Amazon Simple Storage Service (Amazon S3) com pelo menos três barras?

Quando especificar um bucket do Amazon S3, inclua uma barra de término no final do URL. Por exemplo, em vez de referenciar um bucket como “s3n://DOC-EXAMPLE-BUCKET1”, use “s3n://DOC-EXAMPLE-BUCKET1/”, senão o cluster apresentará falha no Hadoop na maioria dos casos.

Você está tentando, recursivamente, desviar diretórios de entrada?

O Hadoop não pesquisa recursivamente diretórios de entrada para arquivos. Se você tiver uma estrutura de diretório como /corpus/01/01.txt, /corpus/01/02.txt, /corpus/02/01.txt, etc. e especificar / corpus/ como o parâmetro de entrada para seu cluster, o Hadoop não localizará nenhum arquivo de entrada porque o diretório /corpus/está vazio, e o Hadoop não verificará o conteúdo dos subdiretórios. Da mesma forma, o Hadoop não verifica recursivamente os subdiretórios de buckets do Amazon S3.

Os arquivos de entrada devem estar diretamente no diretório de entrada ou no bucket do Amazon S3 que você especificar, e não nos subdiretórios.

Seu diretório de saída já existe?

Se você especificar um caminho de saída que já existe, seu cluster apresentará falha no Hadoop na maioria dos casos. Isso significa que, se você executar um cluster uma vez e, em seguida, executá-lo novamente com, exatamente, os mesmos parâmetros ele, provavelmente, funcionará na primeira vez e depois nunca mais. Após a primeira execução, o caminho de saída passa a existir e isso faz com que haja falha em todas as execuções sucessivas.

Você está tentando especificar um recurso usando um URL HTTP?

O Hadoop não aceita locais de recursos especificados usando o prefixo http://. Não é possível fazer referência a um recurso usando um URL HTTP. Por exemplo, passar em http://mysite/myjar.jar como o parâmetro JAR faz com que haja falha no cluster.

Você está referenciando um bucket do Amazon S3 usando um nome de formato inválido?

Se você tentar usar um nome de bucket, como “DOC-EXAMPLE-BUCKET1.1” com o Amazon EMR, haverá falha no cluster porque o Amazon EMR exige que os nomes de bucket sejam nomes de host RFC 2396 válidos. O nome não pode terminar com um número. Além disso, devido aos requisitos do Hadoop, os nomes de bucket do Amazon S3 usados com o Amazon EMR devem conter somente letras minúsculas, números, pontos (.) e hífens (-). Para obter informações sobre como formatar nomes de buckets do Amazon S3, consulte Restrições e limitações do bucket no guia do usuário do Amazon Simple Storage Service.

Você está tendo problemas para carregar dados para carregar ou descarregar dados do Amazon S3?

O Amazon S3 é a fonte de entrada e saída mais conhecida do Amazon EMR. Um erro comum é tratar o Amazon S3 como um sistema de arquivos típico. Há diferenças entre o Amazon S3 e um sistema de arquivos que você precisa levar em conta ao executar seu cluster.

- Se ocorrer um erro interno no Amazon S3, sua aplicação deverá lidar com isso normalmente e repetir a operação.
- Se as chamadas para o Amazon S3 levam muito tempo para retornar, talvez seja necessário reduzir a frequência com que a aplicação chama o Amazon S3.
- Listar todos os objetos em um bucket do Amazon S3 é uma chamada de alto custo. O aplicativo deve minimizar o número de vezes que faz isso.

Há várias maneiras de melhorar como seu cluster interage com o Amazon S3.

- Inicie o cluster usando a versão mais recente do Amazon EMR.
- Use o S3 DistCp para mover objetos para dentro e para fora do Amazon S3. O S3 DistCp implementa tratamento de erros, novas tentativas e recuos para atender aos requisitos do Amazon S3. Para obter mais informações, consulte Cópia distribuída usando o S3 DistCp.
- Projete seu aplicativo com consistência eventual em mente. Use o HDFS para armazenamento de dados intermediários enquanto o cluster está em execução e o Amazon S3 somente para a entrada de dados iniciais e a saída dos resultados finais.
- Se os seus clusters confirmarem 200 ou mais transações por segundo para o Amazon S3, entre em contato com o suporte para preparar seu bucket para mais transações por segundo e

considere usar estratégias de partição de chave, descritas em Amazon S3 performance tips and tricks.

- Defina a configuração io.file.buffer.size do Hadoop como 65536. Isso faz com que o Hadoop gaste menos tempo procurando entre objetos do Amazon S3.
- Considere desabilitar o atributo de execução especulativa do Hadoop, se o cluster estiver enfrentando problemas de simultaneidade do Amazon S3. Isso também é útil quando você estiver solucionando problemas de um cluster lento. Você pode fazer isso definindo as propriedades `mapreduce.reduce.speculative` e `mapreduce.map.speculative` como `false`. Ao executar um cluster, você pode definir esses valores usando a classificação de configuração `mapred-env`. Para obter mais informações, consulte Configurar aplicações no Guia de versão do Amazon EMR.
- Se você estiver executando um cluster do Hive, consulte Você está tendo problemas para carregar ou descarregar dados do Amazon S3 no Hive?.

Para obter informações adicionais, consulte Práticas recomendadas com relação a erros do Amazon S3 no Guia do usuário do Amazon Simple Storage Service.

Erros de permissão

Os seguintes erros são comuns quando se utiliza permissões ou credenciais.

Tópicos

- Você está inserindo as credenciais corretas para o SSH?
- Se você está usando o IAM, possui o conjunto apropriado de políticas do Amazon EC2?

Você está inserindo as credenciais corretas para o SSH?

Se você não consegue usar o SSH para se conectar ao nó principal, é muito provável que haja um problema com suas credenciais de segurança.

Primeiro, verifique se o arquivo .pem contendo a sua chave SSH tem as permissões adequadas. Você pode usar o chmod para alterar as permissões de seu arquivo .pem, como mostrado no exemplo a seguir, onde você deve substituir mykey.pem pelo nome do seu próprio arquivo .pem.

```
chmod og-rwx mykey.pem
```

A segunda possibilidade é você não estar usando o par de chaves especificado quando o cluster foi criado. Isso é fácil de acontecer se você tiver criado vários pares de chaves. Verifique os detalhes do cluster no console do Amazon EMR (ou use a opção `--describe` na CLI) para obter o nome do par de chaves que foi especificado quando o cluster foi criado.

Após verificar se está usando o par de chaves correto e se as permissões estão definidas corretamente no arquivo .pem, você pode usar o comando a seguir para se conectar ao nó principal usando o SSH, onde você deve substituir mykey.pem pelo nome do arquivo .pem e `hadoop@ec2-01-001-001-1.compute-1.amazonaws.com` pelo nome DNS público do nó principal (disponível por meio da opção `--describe` na CLI ou do console do Amazon EMR).

> **Important**
>
> Você deve usar o nome de login `hadoop` quando se conectar a um nó do cluster do Amazon EMR, caso contrário, poderá ocorrer um erro semelhante a `Server refused our key`.

```
ssh -i mykey.pem hadoop@ec2-01-001-001-1.compute-1.amazonaws.com
```

Para ter mais informações, consulte Conectar-se ao nó primário usando SSH.

Se você está usando o IAM, possui o conjunto apropriado de políticas do Amazon EC2?

Como o Amazon EMR usa instâncias do EC2 como nós, os usuários do Amazon EMR também precisam ter determinadas políticas definidas para o Amazon EC2 a fim de que o Amazon EMR possa gerenciar essas instâncias em nome do usuário. Se você não tiver as permissões necessárias, o Amazon EMR gerará o erro: “account is not authorized to call EC2”.

Para obter mais informações sobre as políticas do Amazon EC2 que sua conta do IAM precisa ter definidas para executar o Amazon EMR, consulte Como o Amazon EMR funciona com o IAM.

Erros de cluster do Hive

Geralmente, você pode encontrar a causa de um erro do Hive no arquivo `syslog`, que você vincula a partir do painel Steps (Etapas). Se você não conseguir determinar o problema lá, verifique a

mensagem de erro de tentativa de tarefa do Hadoop. Vincule-se a ela no painel Task Attempts (Tentativas da tarefa).

Os erros a seguir são comuns em clusters do Hive.

Tópicos

Você está usando a versão mais recente do Hive?

A versão mais recente do Hive tem todos os patches e correções de erros atuais e pode resolver o problema.

Você encontrou um erro de sintaxe no script do Hive?

Se houver falha em uma etapa, examine o arquivo `stdout` de logs para a etapa que executou o script do Hive. Se o erro não estiver lá, examine o arquivo `syslog` dos logs das tentativas de tarefa que tiveram falha. Para ter mais informações, consulte Exibir arquivos de log do .

Houve falha em um trabalho quando executado interativamente?

Se você estiver executando o Hive interativamente no nó principal e houver falha no cluster, veja as entradas do `syslog` no log de tentativas de tarefa para a tentativa de tarefas com falha. Para ter mais informações, consulte Exibir arquivos de log do .

Você está tendo problemas para carregar ou descarregar dados do Amazon S3 no Hive?

Se você estiver com problemas para acessar dados no Amazon S3, verifique primeiro as possíveis causas listadas em Você está tendo problemas para carregar dados para carregar ou descarregar dados do Amazon S3?. Se nenhum desses problemas for a causa, considere as opções a seguir específicas para o Hive.

- Verifique se você está usando a versão mais recente do Hive, que tem todos os patches e correções de erros atuais e pode resolver o problema. Para obter mais informações, consulte Apache Hive.

- Usar `INSERT OVERWRITE` exige a listagem do conteúdo do bucket ou pasta do Amazon S3. Isso é uma operação cara. Se possível, remova manualmente o caminho, em vez de fazer com que o Hive liste e exclua os objetos existentes.
- Se você usar versões anteriores à 5.0 do Amazon EMR, poderá usar o seguinte comando no HiveQL para pré-armazenar em cache os resultados de uma operação de lista do Amazon S3 localmente no cluster:

```
set hive.optimize.s3.query=true;
```

- Use partições estáticas sempre que possível.
- Em algumas versões do Hive e do Amazon EMR, é possível que haja falha ao usar ALTER TABLES porque a tabela é armazenada em um local diferente do que o esperado pelo Hive. A solução é adicionar ou atualizar o seguinte no `/home/hadoop/conf/core-site.xml`:

```
<property>
    <name>fs.s3n.endpoint</name>
    <value>s3.amazonaws.com</value>
</property>
```

Erros de VPC

Os erros a seguir são comuns na configuração da VPC no Amazon EMR.

Tópicos

Configuração de sub-rede inválida

Na página Cluster Details (Detalhes do cluster), no campo Status, será exibida uma mensagem de erro semelhante ao seguinte:

`The subnet configuration was invalid: Cannot find route to InternetGateway in main RouteTable` *`rtb-id`* `for vpc` *`vpc-id`*`.`

Para resolver esse problema, você deve criar um Gateway da Internet e anexá-lo à sua VPC. Para obter mais informações, consulte Adicionar um gateway da Internet à VPC.

Como alternativa, verifique se você configurou a VPC com as opções Enable DNS resolution (Habilitar resolução DNS) e Enable DNS hostname support (Habilitar suporte de nome de host DNS) ativadas. Para obter mais informações, consulte Como usar o DNS com sua VPC.

Conjunto de opções DHCP ausente

Você verá uma falha de etapa no syslog (log do sistema) do cluster com uma mensagem de erro semelhante ao seguinte:

```
ERROR org.apache.hadoop.security.UserGroupInformation
(main): PriviledgedActionException as:hadoop (auth:SIMPLE)
cause:java.io.IOException:
org.apache.hadoop.yarn.exceptions.ApplicationNotFoundException: Application
with id 'application_id' doesn't exist in RM.
```

ou

```
ERROR org.apache.hadoop.streaming.StreamJob (main): Error Launching job :
org.apache.hadoop.yarn.exceptions.ApplicationNotFoundException: Application
with id 'application_id' doesn't exist in RM.
```

Para resolver esse problema, você deve configurar uma VPC que inclui um conjunto de opções DHCP cujos parâmetros estejam definidos como os seguintes valores:

> **Note**
>
> Se você usar a região AWS GovCloud (Oeste dos EUA), defina nome de domínio como **`us-gov-west-1.compute.internal`** em vez do valor usado no exemplo a seguir.

- domain-name = **`ec2.internal`**

  Use **`ec2.internal`**, se a região for Leste dos EUA (Norte da Virgínia). Para outras regiões, use *`region-name`***`.compute.internal`**. Por exemplo, em us-west-2, use domain-name=**`us-west-2.compute.internal`**.

- domain-name-servers = **`AmazonProvidedDNS`**

Para obter mais informações, consulte Conjuntos de opções de DHCP.

Erros de permissão

Uma falha no log `stderr` para uma etapa indica que um recurso do Amazon S3 não tem as permissões apropriadas. Este é um erro 403 e a mensagem de erro é semelhante a algo como:

```
Exception in thread "main" com.amazonaws.services.s3.model.AmazonS3Exception: Access
 Denied (Service: Amazon S3; Status Code: 403; Error Code: AccessDenied; Request
 ID: REQUEST_ID
```

Se ActionOnFailure for definido como`TERMINATE_JOB_FLOW`, isso resultará no encerramento do cluster com o estado,`SHUTDOWN_COMPLETED_WITH_ERRORS`.

Algumas maneiras de solucionar esse problema incluem:

- Se você estiver usando uma política de bucket do Amazon S3 dentro de uma VPC, certifique-se de dar acesso a todos os buckets, criando um endpoint da VPC e selecionando Permitir todos na opção Política ao criar o endpoint.
- Certifique-se de que as políticas associadas a recursos do S3 incluam a VPC na qual você inicia o cluster.
- Tente executar o seguinte comando a partir de seu cluster, para verificar se você pode acessar o bucket

  ```
  hadoop fs -copyToLocal s3://path-to-bucket /tmp/
  ```

- Você pode obter mais informações de depuração específicas, ao configurar o parâmetro `log4j.logger.org.apache.http.wire` como `DEBUG` no arquivo `/home/hadoop/conf/log4j.properties` no cluster. Você pode verificar o arquivo de log `stderr` depois de tentar acessar o bucket a partir do cluster. O arquivo de log fornecerá informações mais detalhadas:

  ```
  Access denied for getting the prefix for bucket - us-west-2.elasticmapreduce with
   path samples/wordcount/input/
  15/03/25 23:46:20 DEBUG http.wire: >> "GET /?prefix=samples%2Fwordcount%2Finput
  %2F&delimiter=%2F&max-keys=1 HTTP/1.1[\r][\n]"
  15/03/25 23:46:20 DEBUG http.wire: >> "Host: us-
  west-2.elasticmapreduce.s3.amazonaws.com[\r][\n]"
  ```

Erros que resultam em **START_FAILED**

Antes da AMI 3.7.0, para VPCs em que um nome de host é especificado, o Amazon EMR, mapeia os nomes de hosts internos da sub-rede com endereços de domínio personalizados da seguinte forma: `ip-`*`X.X.X.X.customdomain.com`*`.tld`. Por exemplo, se o nome do host fosse `ip-10.0.0.10` e a VPC tivesse a opção nome de domínio definida como customdomain.com, o nome de host resultante mapeado pelo Amazon EMR seria `ip-10.0.1.0.customdomain.com`. Uma entrada é incluída em `/etc/hosts` para resolver o nome do host como 10.0.0.10. Esse comportamento é alterado com a AMI 3.7.0 e agora o Amazon EMR honra totalmente a configuração de DHCP da VPC. Anteriormente, os clientes também podiam usar uma ação de bootstrap para especificar um mapeamento de nome de host.

Se quiser preservar esse comportamento, você deve fornecer o DNS e encaminhar a configuração de resolução que você precisa para o domínio personalizado.

Cluster **Terminated with errors** e NameNode falha ao iniciar

Ao iniciar um cluster do EMR em uma VPC que faz o uso de um nome de domínio DNS personalizado, pode haver falha no cluster com a seguinte mensagem de erro no console:

```
Terminated with errors  On the master instance(instance-id), bootstrap action 1
 returned a  non-zero return code
```

A falha é resultado da NameNode impossibilidade de inicialização. Isso resultará no seguinte erro encontrado nos NameNode registros, cujo URI do Amazon S3 tem o formato: `s3://`*`mybucket`*`/`*`logs`*`/`*`cluster-id`*`/daemons/`*`master instance-id`*`/hadoop-hadoop-namenode-`*`master node hostname`*`.log.gz`

```
2015-07-23 20:17:06,266 WARN
      org.apache.hadoop.hdfs.server.namenode.FSNamesystem (main): Encountered
 exception
      loading fsimage  java.io.IOException: NameNode is not formatted.
      at

 org.apache.hadoop.hdfs.server.namenode.FSImage.recoverTransitionRead(FSImage.java:212)
           at

 org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFSImage(FSNamesystem.java:1020)
           at

 org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFromDisk(FSNamesystem.java:739)
```

```
          at
     org.apache.hadoop.hdfs.server.namenode.NameNode.loadNamesystem(NameNode.java:537)
          at
     org.apache.hadoop.hdfs.server.namenode.NameNode.initialize(NameNode.java:596)

     at  org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:765)
          at
     org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:749)
     at

org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1441)
          at
     org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1507)
```

Isso é causado por um possível problema em que uma instância do EC2 pode ter vários conjuntos de nomes de domínio totalmente qualificados ao iniciar clusters do EMR em uma VPC que faz uso de um servidor DNS fornecido pela AWS e um servidor DNS personalizado fornecido pelo usuário. Se o servidor DNS fornecido pelo usuário não fornecer nenhum PTR (registro do apontador) para qualquer um dos registros A usados para designar nós em um cluster do EMR, haverá falha de inicialização nos clusters quando configurado desta maneira. A solução é adicionar um registro PTR para cada registro A criado quando uma instância do EC2 é executada em qualquer uma das sub-redes na VPC.

Erros em clusters de transmissão

Em geral, você pode encontrar a causa de um erro de streaming em um arquivo `syslog`. Estabeleça um link para ela no painel Steps (Etapas).

Os seguintes erros são comuns em clusters de streaming.

Tópicos

- Os dados estão sendo enviados ao mapeador no formato errado?
- Seu script está perdendo a validade?
- Você está transmitindo argumentos de streaming inválidos?
- Seu script foi encerrado com um erro?

Os dados estão sendo enviados ao mapeador no formato errado?

Para verificar se esse é o caso, procure uma mensagem de erro no arquivo `syslog` de uma tentativa de tarefa com falha nos logs de tentativas de tarefas. Para ter mais informações, consulte Exibir arquivos de log do .

Seu script está perdendo a validade?

O tempo limite padrão para um script de mapeador ou reducer é de 600 segundos. Se o script demorar mais do que isso, a tentativa de tarefa falhará. Você pode verificar se esse é o caso consultando o arquivo `syslog` de uma tentativa de tarefa com falha nos logs de tentativas de tarefas. Para ter mais informações, consulte Exibir arquivos de log do .

Você pode alterar o limite de tempo definindo um novo valor para a definição de configuração `mapred.task.timeout`. Essa configuração especifica o número de milissegundos após os quais o Amazon EMR encerrará uma tarefa que não lei entradas, gravou saídas ou atualizou sua string de status. Você pode atualizar esse valor transmitindo um argumento de streaming adicional `-jobconf mapred.task.timeout=800000`.

Você está transmitindo argumentos de streaming inválidos?

O streaming do Hadoop oferece suporte apenas aos seguintes argumentos. Se você transmitir argumentos diferentes dos listados abaixo, o cluster falhará.

```
-blockAutoGenerateCacheFiles
-cacheArchive
-cacheFile
-cmdenv
-combiner
-debug
-input
-inputformat
-inputreader
-jobconf
-mapper
-numReduceTasks
-output
-outputformat
-partitioner
-reducer
```

```
-verbose
```

Além disso, o streaming do Hadoop só reconhece argumentos transmitidos usando a sintaxe Java; ou seja, precedidos por um único hífen. Se você transmitir argumentos precedidos de um hífen duplo, o cluster falhará.

Seu script foi encerrado com um erro?

Se a saída do seu script de mapeador ou reducer for gerada com um erro, você poderá localizar esse erro no arquivo `stderr` dos logs de tentativas da tarefa com falha. Para ter mais informações, consulte Exibir arquivos de log do .

Erros de cluster com JAR personalizado

Os seguintes erros são comuns em clusters com JAR personalizado.

Tópicos

- Seu JAR está lançando uma exceção antes de criar um trabalho?
- Seu JAR está lançando um erro dentro em uma tarefa map?

Seu JAR está lançando uma exceção antes de criar um trabalho?

Se o programa principal do seu JAR personalizado lançar uma exceção ao criar o trabalho do Hadoop, o melhor lugar para procurar é no arquivo `syslog` dos logs de etapas. Para ter mais informações, consulte Exibir arquivos de log do .

Seu JAR está lançando um erro dentro em uma tarefa map?

Se o seu JAR personalizado e o mapeador lançarem uma exceção ao processarem dados de entrada, o melhor lugar para procurar é no arquivo `syslog` dos logs de tentativas de tarefas. Para ter mais informações, consulte Exibir arquivos de log do .

AWS GovCloud Erros (Oeste dos EUA)

A região AWS GovCloud (Oeste dos EUA) difere de outras regiões em sua segurança, configuração e configurações padrão. Como resultado, use a lista de verificação a seguir para solucionar erros do Amazon EMR que são específicos da região AWS GovCloud (Oeste dos EUA) antes de usar recomendações mais gerais de solução de problemas.

- Verifique se os perfis do IAM estão configuradas corretamente. Para ter mais informações, consulte Configurar perfis de serviço do IAM para permissões do Amazon EMR aos serviços e recursos da AWS.
- Certifique-se de que a sua configuração de VPC configurou corretamente os parâmetros de resolução de DNS/suporte de nome de host, do Internet Gateway e do conjunto de opções de DHCP. Para ter mais informações, consulte Erros de VPC.

Se essas etapas não resolverem o problema, continue com as etapas para solucionar os erros comuns do Amazon EMR. Para ter mais informações, consulte Erros comuns do Amazon EMR.

Encontrar um cluster ausente

Se o cluster estiver ausente da lista do console ou da API `ListClusters`, verifique o seguinte:

- Confirme se a idade do cluster, a partir do momento da conclusão, é inferior a dois meses. O Amazon EMR preserva informações de metadados de clusters concluídos gratuitamente, por dois meses. Não é possível excluir clusters concluídos do console. Em vez disso, o Amazon EMR limpa os clusters concluídos automaticamente após dois meses.
- Confirme que você tem permissões de perfil para visualizar o cluster.
- Confirme se você está visualizando o mesmo Região da AWS local em que o cluster reside.

Solucionar problemas em um cluster com falha

Esta seção orienta você durante o processo de solução de problemas de um cluster que apresentou falha. Isso significa que o cluster foi encerrado com um código de erro.

> **Note**
>
> Quando um cluster do EMR é terminado com um erro, as APIs `DescribeCluster` e `ListClusters` retornam um código de erro e uma mensagem de erro. Para alguns erros de cluster, a matriz de dados `ErrorDetail` também ajuda a solucionar a falha. Para ter mais informações, consulte Códigos de erro com ErrorDetail informações.

Se o cluster é executado, mas leva muito tempo para retornar resultados, consulte Solucionar problemas com um cluster lento.

Tópicos

Etapa 1: coletar dados sobre o problema

A primeira etapa para solucionar problemas de um cluster é coletar informações sobre o que deu errado e o status e a configuração atuais do cluster. Essas informações serão usadas nas etapas a seguir para confirmar ou descartar as possíveis causas do problema.

Definir o problema

Começamos fazendo uma definição clara do problema. Algumas perguntas para se fazer:

- O que eu esperava que acontecesse? O que aconteceu em vez disso?
- Quando o problema ocorreu pela primeira vez? Com que frequência ele ocorreu desde então?
- Alguma coisa mudou na forma como eu configuro ou executo o cluster?

Detalhes do cluster

Os detalhes do cluster a seguir são úteis para ajudar a monitorar problemas. Para obter mais informações sobre como reunir essas informações, consulte Visualizar o status e os detalhes do cluster.

- Identificador do cluster. (Também chamado de identificador de fluxo de trabalho.)
- Região da AWS e na Zona de Disponibilidade em que o cluster foi lançado.
- Estado do cluster, inclusive detalhes da última alteração de estado.
- Tipo e número de instâncias do EC2 especificados para os nós principal, central e de tarefa.

Etapa 2: verificar o ambiente

O Amazon EMR opera como parte de um ecossistema de serviços da Web e software de código aberto. As coisas que afetam essas dependências podem afetar a performance do Amazon EMR.

Tópicos

Verificar a existência de interrupções de serviço

O Amazon EMR usa diversos Amazon Web Services internamente. Ele executa servidores virtuais no Amazon EC2, armazena dados e scripts no Amazon S3 e reporta métricas para. CloudWatch Os eventos que interrompem esses serviços são raros, mas, quando ocorrem, podem causar problemas no Amazon EMR.

Antes de avançar, verifique o Painel de status dos serviços. Verifique a região onde você iniciou o cluster para saber se há um eventos de interrupção em qualquer um desses serviços.

Verificar os limites de uso

Se você estiver iniciando um cluster grande, tiver lançado vários clusters simultaneamente ou se for um usuário compartilhando um Conta da AWS com outros usuários, o cluster pode ter falhado porque você excedeu um limite de AWS serviço.

O Amazon EC2 limita o número de instâncias de servidores virtuais em execução em uma única AWS região a 20 instâncias sob demanda ou reservadas. Se você iniciar um cluster com mais de 20 nós ou executar um cluster que faça com que o número total de instâncias do EC2 ativas em você Conta da AWS exceda 20, o cluster não poderá executar todas as instâncias do EC2 necessárias e poderá falhar. Quando isso acontece, o Amazon EMR retorna um erro `EC2 QUOTA EXCEEDED`. Você pode solicitar o AWS aumento do número de instâncias do EC2 que você pode executar em sua conta enviando uma solicitação para aumentar o aplicativo Amazon EC2 Instance Limit.

Outra coisa que pode fazer você exceder os limites de uso é o atraso entre quando um cluster é encerrado e quando ele libera todos os recursos. Dependendo da configuração, pode demorar de 5 a 20 minutos para um cluster ser encerrado totalmente e liberar os recursos alocados. Se você estiver

recebendo um erro `EC2 QUOTA EXCEEDED` ao tentar iniciar um cluster, isso poderá acontecer porque os recursos de um cluster recém-encerrado talvez ainda não tenham sido liberados. Nesse caso, é possível solicitar que sua cota do Amazon EC2 seja aumentada ou esperar 20 minutos e reiniciar o cluster.

O Amazon S3 limita a cem o número de buckets criados em uma conta. Se o cluster criar um bucket novo que exceda esse limite, haverá falha na criação do bucket e poderá fazer com que haja uma falha no cluster.

Verificar a versão

Compare o rótulo da versão usada para iniciar o cluster com a versão do Amazon EMR mais recente. Cada versão do Amazon EMR inclui melhorias, como novas aplicações, recursos, patches e correções de bug. O problema que está afetando o cluster já pode ter sido corrigido na versão mais recente. Se possível, execute o cluster novamente usando a versão da mais recente.

Verificar a configuração da sub-rede da Amazon VPC

Se o cluster foi iniciado em uma sub-rede da Amazon VPC, a sub-rede precisa ser configurada conforme descrito em Configurar redes. Além disso, verifique se a sub-rede na qual o cluster é iniciado tem endereços IP elásticos livres suficientes para atribuir um a cada nó do cluster.

Etapa 3: conferir a última alteração de estado

A última alteração de estado fornece informações sobre o que ocorreu na última vez em que o estado do cluster foi alterado. Isso, geralmente, tem informações que podem determinar o que deu errado, conforme o estado de um cluster muda para `FAILED`. Por exemplo, se você iniciar um cluster de transmissão e especificar um local de saída que já exista no Amazon S3, haverá falha no cluster com uma última alteração de estado de “Streaming output directory already exists”.

Você pode localizar o último valor de alteração de estado a partir do console, visualizando o painel de detalhes do cluster; a partir da CLI, usando os argumentos `list-steps` ou `describe-cluster` ou a partir da API, usando as ações `DescribeCluster` e `ListSteps`. Para ter mais informações, consulte Visualizar o status e os detalhes do cluster.

Etapa 4: examinar os arquivos de log

A próxima etapa é examinar os arquivos de log para localizar um código de erro ou outra indicação do problema que o cluster enfrentou. Para obter informações sobre os arquivos de log disponíveis, onde encontrá-los e como visualizá-los, consulte Exibir arquivos de log do .

Pode ser necessário realizar algum trabalho investigativo para determinar o que aconteceu. O Hadoop executa o trabalho em tentativas de tarefa em múltiplos nós do cluster. O Amazon EMR pode iniciar tentativas de tarefa especulativas, terminado as outras tentativas de tarefa que não foram concluídas primeiro. Isso gera uma atividade considerável que é registrada nos arquivos de log controller, stderr e syslog quando isso acontece. Além disso, várias tentativas de tarefa são executadas simultaneamente, mas um arquivo de log só pode exibir os resultados de forma linear.

Comece verificando os logs de ações de bootstrap em busca de erros ou alterações inesperadas na configuração durante a inicialização do cluster. A partir daí, consulte os logs de etapas para identificar trabalhos do Hadoop iniciados como parte de uma etapa com erros. Examine os logs de trabalhos do Hadoop para identificar as tentativas de tarefa com falha. O logs de tentativas de tarefa conterá detalhes sobre o que causou a falha de uma tentativa de tarefa.

As seções a seguir descrevem como usar os diversos arquivos de log para identificar erros no cluster.

Verificar os logs de ação de bootstrap

As ações de bootstrap executam scripts no cluster quando ele é iniciado. Geralmente são usados para instalar outros softwares no cluster ou para alterar as configurações com base nos valores padrão. Verificar esses logs pode fornecer insights sobre os erros que ocorreram durante a configuração do cluster, bem como das alterações nas configurações que podem afetar a performance.

Verificar os logs de etapa

Há quatro tipos de logs de etapas.

- controller: contém arquivos gerados pelo Amazon EMR (Amazon EMR) que surgem de erros encontrados ao tentar executar a etapa. Se a etapa falhar durante o carregamento, você encontrará o rastreamento da pilha nesse log. Os erros ao carregar ou acessar a aplicação muitas vezes são descritos aqui, assim como os erros ausentes do arquivo do mapeador.
- stderr: contém mensagens de erro que ocorreram durante o processamento da etapa. Os erros de carregamento de aplicações muitas vezes são descritos aqui. Às vezes, esse log contém um rastreamento de pilha.
- stdout: contém o status gerado pelos executáveis do mapeador e do redutor. Os erros de carregamento de aplicações muitas vezes são descritos aqui. Às vezes, o log contém mensagens de erro da aplicação.

- syslog: contém registros de softwares que não são da Amazon, como Apache e Hadoop. Os erros de transmissão muitas vezes são descritos aqui.

Verifique se há erros óbvios em stderr. Se stderr exibe uma pequena lista de erros, a etapa foi interrompida rapidamente com um erro gerado. Isso geralmente é causado por um erro nas aplicações mapeadoras e redutoras que estão sendo executadas no cluster.

Verifique se há em avisos de erros ou falhas nas últimas linhas do controller e do syslog. Siga todos os avisos sobre tarefas que falharam, sobretudo se estiver escrito “Job Failed”.

Verificar os logs de tentativa de tarefas

Se a análise anterior dos logs de etapas retornou uma ou mais tarefas com falha, investigue os logs das tentativas de tarefa correspondentes para obter informações mais detalhadas sobre o erro.

Etapa 5: testar o cluster passo a passo

Uma técnica útil quando você está tentando rastrear a origem de um erro é reiniciar o cluster e enviar as etapas a ele uma por uma. Isso permite que você verifique os resultados de cada etapa antes de processar a seguinte e dá a você a oportunidade de corrigir e executar novamente uma etapa que tenha apresentado falha. Isso também permite que você carregue seus dados de entrada somente uma vez.

Para testar um cluster passo a passo

1. Execute um novo cluster, com as proteções de encerramento e keep alive ativadas. A proteção keep alive mantém o cluster em execução após ter processado todas as suas etapas pendentes. A proteção de encerramento impede que um cluster seja encerrado no caso de um erro. Para obter mais informações, consulte Configurar um cluster para continuar ou terminar após a execução da etapa e Usar a proteção contra término.
2. Envie uma etapa para o cluster. Para ter mais informações, consulte Enviar trabalhos a um cluster.
3. Quando a etapa for concluída, verifique se há erros de processamento nos arquivos de log da etapa. Para ter mais informações, consulte Etapa 4: examinar os arquivos de log. A maneira mais rápida de localizar esses arquivos de log é estabelecer uma conexão com o nó principal e exibir os arquivos de log. Os arquivos de log da etapa não serão exibidos até que a etapa seja executada por algum tempo, seja concluída ou apresente uma falha.

4. Se a etapa for concluída com êxito, execute a próxima etapa. Se houver erros, investigue o erro nos arquivos de log. Se houve um erro em seu código, faça a correção e execute novamente a etapa. Continue até que todas as etapas sejam executadas sem erros.
5. Quando você terminar a depuração do cluster e quiser encerrá-lo, deverá fazê-lo manualmente. Isso é necessário porque o cluster foi iniciado com a proteção de encerramento ativada. Para ter mais informações, consulte Usar a proteção contra término.

Solucionar problemas com um cluster lento

Esta seção orienta você durante o processo de solução de problemas com um cluster que ainda está em execução, mas está demorando muito para retornar resultados. Para obter mais informações sobre o que fazer se o cluster tiver sido encerrado com um código de erro, consulte Solucionar problemas em um cluster com falha

O Amazon EMR permite que você especifique o número e o tipo de instâncias no cluster. Essas especificações são os principais meios de afetar a velocidade com a qual o processamento dos seus dados é concluída. Uma ação que você pode considerar é reexecutar o cluster, dessa vez especificando instâncias do EC2 com recursos maiores ou especificando um número maior de instâncias no cluster. Para ter mais informações, consulte Configurar o hardware e as redes do cluster.

Os tópicos a seguir você orientam você durante o processo de identificar as causas alternativas de um cluster lento.

Tópicos

- Etapa 1: coletar dados sobre o problema
- Etapa 2: verificar o ambiente
- Etapa 3: examinar os arquivos de log
- Etapa 4: verificar a integridade do cluster e das instâncias
- Etapa 5: verificar se há grupos suspensos
- Etapa 6: revisar as configurações
- Etapa 7: examinar dados de entrada

Etapa 1: coletar dados sobre o problema

A primeira etapa para solucionar problemas de um cluster é coletar informações sobre o que deu errado e o status e a configuração atuais do cluster. Essas informações serão usadas nas etapas a seguir para confirmar ou descartar as possíveis causas do problema.

Definir o problema

Começamos fazendo uma definição clara do problema. Algumas perguntas para se fazer:

- O que eu esperava que acontecesse? O que aconteceu em vez disso?
- Quando o problema ocorreu pela primeira vez? Com que frequência ele ocorreu desde então?
- Alguma coisa mudou na forma como eu configuro ou executo o cluster?

Detalhes do cluster

Os detalhes do cluster a seguir são úteis para ajudar a monitorar problemas. Para obter mais informações sobre como reunir essas informações, consulte Visualizar o status e os detalhes do cluster.

- Identificador do cluster. (Também chamado de identificador de fluxo de trabalho.)
- Região da AWS e na Zona de Disponibilidade em que o cluster foi lançado.
- Estado do cluster, inclusive detalhes da última alteração de estado.
- Tipo e número de instâncias do EC2 especificados para os nós principal, central e de tarefa.

Etapa 2: verificar o ambiente

Tópicos

- Verificar a existência de interrupções de serviço
- Verificar os limites de uso
- Verificar a configuração da sub-rede da Amazon VPC
- Reiniciar o cluster

Verificar a existência de interrupções de serviço

O Amazon EMR usa diversos Amazon Web Services internamente. Ele executa servidores virtuais no Amazon EC2, armazena dados e scripts no Amazon S3 e reporta métricas para. CloudWatch Os eventos que interrompem esses serviços são raros, mas, quando ocorrem, podem causar problemas no Amazon EMR.

Antes de avançar, verifique o Painel de status dos serviços. Verifique a região onde você iniciou o cluster para saber se há um eventos de interrupção em qualquer um desses serviços.

Verificar os limites de uso

Se você estiver iniciando um cluster grande, tiver lançado vários clusters simultaneamente ou se for um usuário compartilhando um Conta da AWS com outros usuários, o cluster pode ter falhado porque você excedeu um limite de AWS serviço.

O Amazon EC2 limita o número de instâncias de servidores virtuais em execução em uma única AWS região a 20 instâncias sob demanda ou reservadas. Se você iniciar um cluster com mais de 20 nós ou executar um cluster que faça com que o número total de instâncias do EC2 ativas em você Conta da AWS exceda 20, o cluster não poderá executar todas as instâncias do EC2 necessárias e poderá falhar. Quando isso acontece, o Amazon EMR retorna um erro `EC2 QUOTA EXCEEDED`. Você pode solicitar o AWS aumento do número de instâncias do EC2 que você pode executar em sua conta enviando uma solicitação para aumentar o aplicativo Amazon EC2 Instance Limit.

Outra coisa que pode fazer você exceder os limites de uso é o atraso entre quando um cluster é encerrado e quando ele libera todos os recursos. Dependendo da configuração, pode demorar de 5 a 20 minutos para um cluster ser encerrado totalmente e liberar os recursos alocados. Se você estiver recebendo um erro `EC2 QUOTA EXCEEDED` ao tentar iniciar um cluster, isso poderá acontecer porque os recursos de um cluster recém-encerrado talvez ainda não tenham sido liberados. Nesse caso, é possível solicitar que sua cota do Amazon EC2 seja aumentada ou esperar 20 minutos e reiniciar o cluster.

O Amazon S3 limita a cem o número de buckets criados em uma conta. Se o cluster criar um bucket novo que exceda esse limite, haverá falha na criação do bucket e poderá fazer com que haja uma falha no cluster.

Verificar a configuração da sub-rede da Amazon VPC

Se o cluster foi iniciado em uma sub-rede da Amazon VPC, a sub-rede precisa ser configurada conforme descrito em Configurar redes. Além disso, verifique se a sub-rede na qual o cluster é iniciado tem endereços IP elásticos livres suficientes para atribuir um a cada nó do cluster.

Reiniciar o cluster

A lentidão no processamento pode ser causada por uma condição transitória. Considere encerrar e reiniciar o cluster para ver se o desempenho melhora.

Etapa 3: examinar os arquivos de log

A próxima etapa é examinar os arquivos de log para localizar um código de erro ou outra indicação do problema que o cluster enfrentou. Para obter informações sobre os arquivos de log disponíveis, onde encontrá-los e como visualizá-los, consulte Exibir arquivos de log do .

Pode ser necessário realizar algum trabalho investigativo para determinar o que aconteceu. O Hadoop executa o trabalho em tentativas de tarefa em múltiplos nós do cluster. O Amazon EMR pode iniciar tentativas de tarefa especulativas, terminado as outras tentativas de tarefa que não foram concluídas primeiro. Isso gera uma atividade considerável que é registrada nos arquivos de log controller, stderr e syslog quando isso acontece. Além disso, várias tentativas de tarefa são executadas simultaneamente, mas um arquivo de log só pode exibir os resultados de forma linear.

Comece verificando os logs de ações de bootstrap em busca de erros ou alterações inesperadas na configuração durante a inicialização do cluster. A partir daí, consulte os logs de etapas para identificar trabalhos do Hadoop iniciados como parte de uma etapa com erros. Examine os logs de trabalhos do Hadoop para identificar as tentativas de tarefa com falha. O logs de tentativas de tarefa conterá detalhes sobre o que causou a falha de uma tentativa de tarefa.

As seções a seguir descrevem como usar os diversos arquivos de log para identificar erros no cluster.

Verificar os logs de ação de bootstrap

As ações de bootstrap executam scripts no cluster quando ele é iniciado. Geralmente são usados para instalar outros softwares no cluster ou para alterar as configurações com base nos valores padrão. Verificar esses logs pode fornecer insights sobre os erros que ocorreram durante a configuração do cluster, bem como das alterações nas configurações que podem afetar a performance.

Verificar os logs de etapa

Há quatro tipos de logs de etapas.

- controller: contém arquivos gerados pelo Amazon EMR (Amazon EMR) que surgem de erros encontrados ao tentar executar a etapa. Se a etapa falhar durante o carregamento, você encontrará o rastreamento da pilha nesse log. Os erros ao carregar ou acessar a aplicação muitas vezes são descritos aqui, assim como os erros ausentes do arquivo do mapeador.
- stderr: contém mensagens de erro que ocorreram durante o processamento da etapa. Os erros de carregamento de aplicações muitas vezes são descritos aqui. Às vezes, esse log contém um rastreamento de pilha.
- stdout: contém o status gerado pelos executáveis do mapeador e do redutor. Os erros de carregamento de aplicações muitas vezes são descritos aqui. Às vezes, o log contém mensagens de erro da aplicação.
- syslog: contém registros de softwares que não são da Amazon, como Apache e Hadoop. Os erros de transmissão muitas vezes são descritos aqui.

Verifique se há erros óbvios em stderr. Se stderr exibe uma pequena lista de erros, a etapa foi interrompida rapidamente com um erro gerado. Isso geralmente é causado por um erro nas aplicações mapeadoras e redutoras que estão sendo executadas no cluster.

Verifique se há em avisos de erros ou falhas nas últimas linhas do controller e do syslog. Siga todos os avisos sobre tarefas que falharam, sobretudo se estiver escrito “Job Failed”.

Verificar os logs de tentativa de tarefas

Se a análise anterior dos logs de etapas retornou uma ou mais tarefas com falha, investigue os logs das tentativas de tarefa correspondentes para obter informações mais detalhadas sobre o erro.

Verificar os logs de daemons do Hadoop

Em casos raros, o Hadoop em si poderá falhar. Para ver se esse é o caso, é necessário examinar os logs do Hadoop. Eles estão localizados em cada nó do `/var/log/hadoop/`.

Você pode usar os JobTracker registros para mapear uma tentativa de tarefa malsucedida para o nó em que ela foi executada. Depois de conhecer o nó associado à tentativa de tarefa, verifique a integridade da instância do EC2 que hospeda esse nó para ver se houve algum problema, como falta de CPU ou de memória.

Etapa 4: verificar a integridade do cluster e das instâncias

Um cluster do Amazon EMR é formado por nós em execução em instâncias do Amazon EC2. Se essas instâncias tornarem-se limitadas por recursos (por exemplo, se ficarem sem memória ou CPU), passarem por problemas de conectividade de rede ou forem encerradas, a velocidade de processamento do cluster será prejudicada.

Existem até três tipos de nós em um cluster:

- nó principal: gerencia o cluster. Se ele sofrer um problema de desempenho, todo o cluster será afetado.
- nós core: processam tarefas map/reduce e mantêm o Sistema de Arquivos Distribuído do Hadoop (HDFS). Se um dos nós passar por um problema de desempenho, isso poderá retardar as operações do HDFS, bem como o processamento de map/reduce. Você pode adicionar outros nós core a um cluster para melhorar o desempenho, mas não pode remover nós core. Para ter mais informações, consulte Redimensionar manualmente um cluster em execução.
- nós de tarefa: processam tarefas map/reduce. Estes são recursos puramente de computação e não armazenam dados. Você pode adicionar nós de tarefas a um cluster para acelerar o desempenho ou pode remover nós de tarefas que não são necessários. Para ter mais informações, consulte Redimensionar manualmente um cluster em execução.

Ao examinar a integridade de um cluster, você deve considerar o desempenho do cluster como um todo, bem como o desempenho de instâncias individuais. Existem várias ferramentas que pode ser usadas:

Verifique a integridade do cluster com CloudWatch

Cada cluster do Amazon EMR reporta métricas para. CloudWatch Essas métricas fornecem informações de desempenho resumidas sobre o cluster, como a carga total, a utilização do HDFS, as tarefas em execução, as tarefas restantes, os blocos corrompidos e muito mais. A análise das CloudWatch métricas fornece uma visão geral do que está acontecendo com seu cluster e pode fornecer informações sobre o que está causando a lentidão no processamento. Além de usar CloudWatch para analisar um problema de desempenho existente, você pode definir alarmes que CloudWatch causem alertas caso ocorra um problema de desempenho futuro. Para ter mais informações, consulte Monitorando métricas do Amazon EMR com CloudWatch.

Verificar a integridade do status do trabalho e do HDFS

Use as Interfaces do usuário do aplicativo na página de detalhes do cluster para visualizar os detalhes do aplicativo YARN. Para determinados aplicativos, você pode analisar diretamente os logs de acesso em mais detalhes. Isso é útil principalmente para aplicativos Spark. Para ter mais informações, consulte Visualizar o histórico da aplicação.

O Hadoop fornece uma série de interfaces Web que você pode usar para visualizar informações. Para obter mais informações sobre como acessar essas interfaces Web, consulte Visualizar interfaces Web hospedadas em clusters do Amazon EMR.

- JobTracker — fornece informações sobre o progresso do trabalho que está sendo processado pelo cluster. Você pode usar essa interface para identificar quando um trabalho ficou preso.
- HDFS NameNode — fornece informações sobre a porcentagem de utilização do HDFS e o espaço disponível em cada nó. Você pode usar essa interface para identificar quando o HDFS está se tornando limitado por recursos e requer capacidade adicional.
- TaskTracker — fornece informações sobre as tarefas do trabalho que está sendo processado pelo cluster. Você pode usar essa interface para identificar quando uma tarefa ficou presa.

Verificar a integridade da instância com o Amazon EC2

Outra maneira de procurar informações sobre o status das instâncias no cluster é usar o console do Amazon EC2. Como cada nó do cluster é executado em uma instância do EC2, você pode usar as ferramentas fornecidas pelo Amazon EC2 para verificar seu status. Para ter mais informações, consulte Visualizar instâncias de cluster no Amazon EC2.

Etapa 5: verificar se há grupos suspensos

Um grupo de instâncias fica suspenso quando encontra muitos erros ao tentar executar nós. Por exemplo, se novos nós falharem repetidamente durante a execução de ações de bootstrap, depois de algum tempo, o grupo de instâncias entrará no estado `SUSPENDED` em vez de tentar provisionar continuamente novos nós.

Um nó poderá falhar se:

- O Hadoop ou o cluster estiver de alguma forma com problemas e não aceitar um novo nó no cluster
- Uma ação de bootstrap falhar no novo nó

- O nó não estava funcionando corretamente e não conseguir fazer check-in no Hadoop

Se um grupo de instâncias estiver no estado `SUSPENDED`, e o cluster estiver em um estado `WAITING`, você poderá adicionar uma etapa de cluster para redefinir o número desejado de nós core e de tarefa. Adicionar a etapa retoma o processamento do cluster e coloca o grupo de instâncias em um estado `RUNNING`.

Para obter mais informações sobre como redefinir um cluster em um estado suspenso, consulte Estado suspenso.

Etapa 6: revisar as configurações

Definições de configuração especificam detalhes sobre como um cluster é executado, como quantas vezes uma tarefa deve ser repetida e quanta memória está disponível para classificação. Quando você executa um cluster usando o Amazon EMR, existem configurações específicas do Amazon EMR, além das definições de configuração padrão do Hadoop. As definições de configuração são armazenadas no nó principal do cluster. Você pode verificar as definições de configuração para garantir que o cluster tenha os recursos necessários para um funcionamento eficiente.

O Amazon EMR define definições de configuração do Hadoop padrão que ele utiliza para iniciar um cluster. Os valores se baseiam na AMI e no tipo de instância que você especifica para o cluster. Você pode modificar os valores padrão das definições de configuração usando uma ação de bootstrap ou especificando novos valores em parâmetros de execução de trabalho. Para ter mais informações, consulte Criar ações de bootstrap para instalar softwares adicionais. Para determinar se uma ação de bootstrap alterou as definições de configuração, verifique os logs dessa ação.

O Amazon EMR registra em log as configurações do Hadoop usadas para executar cada trabalho. Os dados de log são armazenados em um arquivo `job_`*`job-id`*`_conf.xml` no diretório `/mnt/var/log/hadoop/history/` do nó principal, em que *`job-id`* é substituído pelo identificador do trabalho. Se você tiver habilitado o arquivamento em log, esses dados serão copiados para o Amazon S3 na pasta `logs/`*`date`*`/`*`jobflow-id`*`/jobs`, em que *`date`* é a data em que o trabalho foi executado, e *`jobflow-id`* é o identificador do cluster.

As seguintes definições de configuração de trabalhos do Hadoop são especialmente úteis para investigar problemas de desempenho. Para obter mais informações sobre as definições de configuração do Hadoop e como elas afetam o comportamento do Hadoop, acesse http://hadoop.apache.org/docs/.

⚠ Warning

1. Definir `dfs.replication` como 1 em clusters com menos de quatro nós poderá causar perda de dados do HDFS se um único nó ficar inativo. É recomendável usar um cluster com pelo menos quatro nós centrais para workloads de produção.
2. O Amazon EMR não permitirá que os clusters escalem os nós principais abaixo de `dfs.replication`. Por exemplo, se `dfs.replication = 2`, o número mínimo de nós central será 2.
3. Ao usar o Ajuste de Escala Gerenciado, o Auto Scaling ou optar por redimensionar manualmente o cluster, é recomendável definir `dfs.replication` como 2 ou mais.

Definição da configuração	Descrição
dfs.replication	O número de nós HDFS para os quais um único bloco (como o bloco de disco rígido) é copiado a fim de produzir um ambiente semelhante ao RAID. Determina o número de nós HDFS que contêm uma cópia do bloco.
io.sort.mb	Total de memória disponível para classificação. Esse valor deve ser 10x io.sort.factor. Essa configuração também pode ser usada para calcular o total de memória usado pelo nó de tarefas, contando io.sort.mb multiplic ado por mapred.tasktracker.ap.tasks.maximum.
io.sort.spill.percent	Usado durante a classificação, no momento em que o disco começa a ser usado porque a memória de classific ação alocada está ficando cheia.
mapred.child.java.opts	Suspenso. Em vez disso, use mapred.map.child.java.opts e mapred.reduce.child.java.opts. As opções Java são TaskTracker usadas ao iniciar uma JVM para que uma tarefa seja executada nela. Um parâmetro comum é “-Xmx” para configurar o tamanho máximo da memória.
mapred.map.child.java.opts	As opções Java são TaskTracker usadas ao iniciar uma JVM para que uma tarefa de mapeamento seja

Definição da configuração	Descrição
	executada nela. Um parâmetro comum é “-Xmx” para configurar o tamanho máximo do heap de memória.
mapred.map.tasks.speculative.execution	Determina se tentativas de tarefas map da mesma tarefa podem ser executadas em paralelo.
mapred.reduce.tasks.speculative.execution	Determina se tentativas de tarefas reduce da mesma tarefa podem ser executadas em paralelo.
mapred.map.max.attempts	O número máximo de vezes que uma tarefa map pode ser tentada. Se tudo falhar, a tarefa map será marcada como falha.
mapred.reduce.child.java.opts	As opções Java são TaskTracker usadas ao iniciar uma JVM para que uma tarefa reduzida seja executada nela. Um parâmetro comum é “-Xmx” para configurar o tamanho máximo do heap de memória.
mapred.reduce.max.attempts	O número máximo de vezes que uma tarefa reduce pode ser tentada. Se tudo falhar, a tarefa map será marcada como falha.
mapred.reduce.slowstart.completed.maps	A quantidade de tarefas map que devem ser concluídas antes que tarefas reduce sejam tentadas. Uma espera insuficiente pode causar erros “Too many fetch-failure” em tentativas.
mapred.reuse.jvm.num.tasks	Uma tarefa é executada em uma única JVM. Especifica quantas tarefas podem reutilizar a mesma JVM.
mapred.tasktracker.map.tasks.maximum	A quantidade máxima de tarefas que podem ser executadas em paralelo por nó de tarefa durante o mapeamento.
mapred.tasktracker.reduce.tasks.maximum	A quantidade máxima de tarefas que podem ser executadas em paralelo por nó de tarefa durante a redução.

Se as suas tarefas de cluster consumirem muita memória, você poderá melhorar o desempenho usando menos tarefas por nó core e reduzindo seu tamanho do heap do rastreador de trabalhos.

Etapa 7: examinar dados de entrada

Observe seus dados de entrada. Eles estão distribuídos uniformemente entre seus valores de chave? Se os seus dados estiverem fortemente desviados para um ou alguns valores de chave, a carga de processamento pode estar mapeada para um pequeno número de nós, enquanto outros nós estão ociosos. Essa distribuição desequilibrada de trabalho pode resultar em tempos de processamento mais lentos.

Um exemplo de um conjunto de dados desequilibrado seria executar um cluster para colocar palavras em ordem alfabética, mas ter um conjunto de dados contendo apenas palavras que começam com a letra "a". Quando o trabalho fosse mapeado, o nó processando valores que começam com "a" seria sobrecarregado, enquanto os nós processando palavras que começam com outras letras ficariam ociosos.

Solucionar problemas de um cluster do Lake Formation

Esta seção orienta você no processo de solução de problemas comuns ao usar o Amazon EMR com o AWS Lake Formation.

O acesso ao data lake não é permitido

É necessário optar explicitamente pela filtragem de dados nos clusters do Amazon EMR para poder analisar e processar dados no data lake. Quando o acesso aos dados falhar, você verá uma mensagem genérica `Access is not allowed` na saída das entradas do caderno.

Para aceitar e permitir a filtragem de dados no Amazon EMR, consulte as instruções em Allow data filtering on Amazon EMR no Guia do desenvolvedor do AWS Lake Formation .

Expiração da sessão

O tempo limite da sessão para Cadernos do EMR e Zeppelin é controlado pelo perfil do IAM para a configuração `Maximum CLI/API session duration` do Lake Formation. O valor padrão para essa configuração é uma hora. Quando ocorrer um tempo limite de sessão, você verá a seguinte mensagem na saída das entradas do bloco de anotações ao tentar executar comandos do Spark SQL.

```
Error 401    HTTP ERROR: 401 Problem accessing /sessions/2/statements.
Reason:  JWT token included in request failed validation.
Powered by Jetty:// 9.3.24.v20180605
 org.springframework.web.client.HttpClientErrorException: 401 JWT token included in
 request failed validation…
```

Para validar sua sessão, atualize a página. Será solicitado que você faça a autenticação novamente usando seu IdP e seja redirecionado de volta para o bloco de anotações. Você pode continuar a executar consultas após a nova autenticação.

Não há permissões para o usuário na tabela solicitada

Ao tentar acessar uma tabela à qual você não tem acesso, você verá a seguinte exceção na saída das entradas do bloco de anotações ao tentar executar comandos do Spark SQL.

```
org.apache.spark.sql.AnalysisException:
 org.apache.hadoop.hive.ql.metadata.HiveException: Unable to fetch table table.
Resource does not exist or requester is not authorized to access requested
 permissions.
(Service: AWSGlue; Status Code: 400; Error Code: AccessDeniedException; Request ID: …
```

Para acessar a tabela, você deve conceder acesso ao usuário atualizando as permissões associadas a essa tabela no Lake Formation.

Consultar dados de várias contas compartilhados com o Lake Formation

Quando você usa o Amazon EMR para acessar dados de outra conta compartilhados com você, algumas bibliotecas do Spark tentarão chamar a operação de API `Glue:GetUserDefinedFunctions`. Como as versões 1 e 2 das permissões AWS RAM gerenciadas não oferecem suporte a essa ação, você recebe a seguinte mensagem de erro:

```
"ERROR: User: arn:aws:sts::012345678901:assumed-role/my-
spark-role/i-06ab8c2b59299508a is not authorized to perform:
glue:GetUserDefinedFunctions on resource: arn:exampleCatalogResource
because no resource-based policy allows the glue:GetUserDefinedFunctions
action"
```

Para resolver esse erro, o administrador do data lake que criou o compartilhamento de recursos deve atualizar as permissões AWS RAM gerenciadas anexadas ao compartilhamento de recursos. A

versão 3 das permissões gerenciadas pelo AWS RAM permite que as entidades principais executem a ação `glue:GetUserDefinedFunctions`.

Se você criar um novo compartilhamento de recursos, o Lake Formation aplicará a versão mais recente da permissão AWS RAM gerenciada por padrão, e nenhuma ação será exigida por você. Para habilitar o acesso a dados entre contas para compartilhamentos de recursos existentes, você precisa atualizar as permissões AWS RAM gerenciadas para a versão 3.

Você pode ver as AWS RAM permissões atribuídas aos recursos compartilhados com você em AWS RAM. As permissões incluídas na versão 3 são estas:

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase

AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Para atualizar a versão de permissões AWS RAM gerenciadas dos compartilhamentos de recursos existentes

Você (administrador do data lake) pode atualizar as permissões AWS RAM gerenciadas para uma versão mais recente seguindo as instruções no Guia do AWS RAM usuário ou revogar todas as permissões existentes para o tipo de recurso e concedê-las novamente. Se você revogar as permissões, AWS RAM excluirá o compartilhamento AWS RAM de recursos associado ao tipo de recurso. Quando você concede permissões novamente, AWS RAM cria novos compartilhamentos de recursos anexando a versão mais recente das permissões AWS RAM gerenciadas.

Inserir, criar e alterar tabelas

Não há suporte para a inserção, a criação ou a alteração de tabelas em bancos de dados protegidos por políticas do Lake Formation. Ao executar essas operações, você verá a seguinte exceção na saída das entradas do bloco de anotações ao tentar executar comandos do Spark SQL:

```
java.io.IOException:
 com.amazon.ws.emr.hadoop.fs.shaded.com.amazonaws.services.s3.model.AmazonS3Exception:
           Access Denied (Service: Amazon S3; Status Code: 403; Error Code:
 AccessDenied; Request ID: …
```

Para obter mais informações, consulte Limitações da integração do Amazon EMR com. AWS Lake Formation

Escrita de aplicações que iniciam e gerenciam clusters

Tópicos

Você pode acessar a funcionalidade fornecida pela API do Amazon EMR chamando funções de wrapper em um dos SDKs. AWS Os AWS SDKs fornecem funções específicas de linguagem que envolvem a API do serviço web e simplificam a conexão com o serviço web, lidando com muitos dos detalhes da conexão para você. Para obter mais informações sobre como chamar o Amazon EMR usando um dos SDKs, consulte Uso de SDKs para chamar APIs do Amazon EMR.

> **Important**
>
> A taxa máxima de solicitações para o Amazon EMR é de uma solicitação a cada dez segundos.

E Exemplo de código-fonte Java do nd-to-end Amazon EMR

Os desenvolvedores podem chamar a API do Amazon EMR usando código Java personalizado para fazer as mesmas coisas que fariam com o console ou com a CLI do Amazon EMR. Esta seção fornece as end-to-end etapas necessárias para instalar AWS Toolkit for Eclipse e executar uma amostra de código-fonte Java totalmente funcional que adiciona etapas a um cluster do Amazon EMR.

> **Note**
>
> Este exemplo se concentra em Java, mas o Amazon EMR também oferece suporte a diversas linguagens de programação com uma coleção de SDKs do Amazon EMR. Para ter mais informações, consulte Uso de SDKs para chamar APIs do Amazon EMR.

Este exemplo de código-fonte Java demonstra como executar as seguintes tarefas usando a API do Amazon EMR:

- Recupere AWS as credenciais e envie-as ao Amazon EMR para fazer chamadas de API
- Configurar uma nova etapa personalizada e uma nova etapa predefinida
- Adicionar novas etapas a um cluster existente do Amazon EMR
- Recuperar os IDs das etapas de um cluster em execução

> **Note**
>
> Este exemplo demonstra como adicionar etapas a um cluster existente e, portanto, requer um cluster ativo na sua conta.

Antes de começar, instale a versão do Eclipse IDE for Java EE Developers (Eclipse IDE para desenvolvedores de Java EE) que corresponda a sua plataforma do computador. Para obter mais informações, acesse a página de downloads do Eclipse.

Em seguida, instale o plug-in de desenvolvimento de banco de dados para o Eclipse.

Instalar o plug-in de desenvolvimento de banco de dados para o Eclipse

1. Abra o Eclipse IDE.
2. Escolha Help (Ajuda) e Install New Software (Instalar novo software).
3. No campo Work with: (Trabalhar com:), digite **`http://download.eclipse.org/releases/kepler`** ou o caminho que corresponda ao número da versão do seu Eclipse IDE.
4. Na lista de itens, escolha Database Development (Desenvolvimento de banco de dados) e Finish (Concluir).
5. Reinicie o Eclipse quando solicitado.

Em seguida, instale o kit de ferramentas para Eclipse a fim de disponibilizar os modelos de projeto de código-fonte úteis e configurados previamente.

Instalar o kit de ferramentas para Eclipse

1. Abra o Eclipse IDE.

2. Escolha Help (Ajuda) e Install New Software (Instalar novo software).
3. No campo Work with: (Trabalhar com:), digite **`https://aws.amazon.com/eclipse`**.
4. Na lista de itens, escolha AWS Toolkit for Eclipse e Finish.
5. Reinicie o Eclipse quando solicitado.

Em seguida, crie um novo projeto AWS Java e execute o código-fonte Java de amostra.

Para criar um novo projeto AWS Java

1. Abra o Eclipse IDE.
2. Escolha File (Arquivo), New (Novo) e Other (Outros).
3. Na caixa de diálogo Select a wizard, escolha AWS Java Project e Next.
4. Na caixa de diálogo Novo projeto AWS Java, no **`Project name:`** campo, insira o nome do seu novo projeto, por exemplo**`EMR-sample-code`**.
5. Escolha Configurar AWS contas..., insira suas chaves de acesso públicas e privadas e escolha Concluir. Para obter mais informações sobre a criação de chaves de acesso, consulte How do I get security credentials? na Referência geral da Amazon Web Services.

   > **Note**
   >
   > Você não deve incorporar chaves de acesso diretamente no código. O SDK do Amazon EMR permite colocar as chaves de acesso em locais conhecidos para que você não precise mantê-las em código.

6. No novo projeto Java, clique com o botão direito do mouse na pasta src e, em seguida, escolha New (Novo) e Class (Classe).
7. Na caixa de diálogo Java Class (Classe Java), no campo Name (Nome), insira um nome para sua nova classe, por exemplo, **`main`**.
8. Na seção Which method stubs would you like to create? (Quais stubs de método você gostaria de criar?) escolha public static void main (String [] args) e Finish (Concluir).
9. Insira o código-fonte em Java dentro de sua nova classe e adicione as declarações adequadas de import (importação)para as classes e métodos no exemplo. Para sua conveniência, a listagem do código-fonte completo é mostrada abaixo.

> **Note**
>
> No código de exemplo a seguir, substitua o exemplo de ID de cluster (JobFlowId) *j-xxxxxxxxxxxx*,, por um ID de cluster válido em sua conta encontrado no AWS Management Console ou usando o seguinte AWS CLI comando:
>
> ```
> aws emr list-clusters --active | grep "Id"
> ```
>
> Além disso, substitua o caminho de exemplo do Amazon S3, *s3://path/to/my/jarfolder*, pelo caminho válido para o seu JAR. Por fim, substitua o nome da classe do exemplo, *com.my.Main1*, pelo nome correto da classe em seu JAR, se aplicável.

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;

public class Main {

 public static void main(String[] args) {
  AWSCredentials credentials_profile = null;
  try {
   credentials_profile = new
 ProfileCredentialsProvider("default").getCredentials();
  } catch (Exception e) {
   throw new AmazonClientException(
     "Cannot load credentials from .aws/credentials file. " +
       "Make sure that the credentials file exists and the profile name is
 specified within it.",
     e);
  }

  AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
    .withCredentials(new AWSStaticCredentialsProvider(credentials_profile))
    .withRegion(Regions.US_WEST_1)
```

```
    .build();

  // Run a bash script using a predefined step in the StepFactory helper class
  StepFactory stepFactory = new StepFactory();
  StepConfig runBashScript = new StepConfig()
    .withName("Run a bash script")
    .withHadoopJarStep(stepFactory.newScriptRunnerStep("s3://jeffgoll/emr-scripts/
create_users.sh"))
    .withActionOnFailure("CONTINUE");

  // Run a custom jar file as a step
  HadoopJarStepConfig hadoopConfig1 = new HadoopJarStepConfig()
    .withJar("s3://path/to/my/jarfolder") // replace with the location of the jar
 to run as a step
    .withMainClass("com.my.Main1") // optional main class, this can be omitted if
 jar above has a manifest
    .withArgs("--verbose"); // optional list of arguments to pass to the jar
  StepConfig myCustomJarStep = new StepConfig("RunHadoopJar", hadoopConfig1);

  AddJobFlowStepsResult result = emr.addJobFlowSteps(new AddJobFlowStepsRequest()
    .withJobFlowId("j-xxxxxxxxxxxx") // replace with cluster id to run the steps
    .withSteps(runBashScript, myCustomJarStep));

  System.out.println(result.getStepIds());

 }
}
```

10. Escolha Run (Executar), Run As (Executar como) e Java Application (Aplicativo Java).
11. Se o exemplo for executado corretamente, uma lista de IDs para as novas etapas aparece na janela do console do Eclipse IDE. A saída correta é semelhante à seguinte:

```
[s-39BLQZRJB2E5E, s-1L6A4ZU2SAURC]
```

Conceitos comuns para chamadas de API

Tópicos

- Endpoints para o Amazon EMR
- Especificação dos parâmetros de cluster no Amazon EMR
- Zonas de disponibilidade no Amazon EMR

- Como usar arquivos e bibliotecas adicionais em clusters do Amazon EMR

Ao escrever uma aplicação que chama a API do Amazon EMR, há vários conceitos que se aplicam ao chamar uma das funções wrapper de um SDK.

Endpoints para o Amazon EMR

Um endpoint é um URL que é o ponto de entrada para um serviço da Web. Toda solicitação de serviço da web deve conter um endpoint. O endpoint especifica a AWS região em que os clusters são criados, descritos ou encerrados. Ele tem o formato `elasticmapreduce.`*`regionname`*`.amazonaws.com`. Se você especificar o endpoint geral (`elasticmapreduce.amazonaws.com`), o Amazon EMR direcionará sua solicitação para um endpoint na região padrão. Para contas criadas a partir de 8 de março de 2013, a região padrão é us-west-2; para contas mais antigas, a região padrão é us-east-1.

Para obter mais informações sobre os endpoints do Amazon EMR, consulte Regions and endpoints na Referência geral da Amazon Web Services.

Especificação dos parâmetros de cluster no Amazon EMR

Os parâmetros `Instances` permitem que você configure os tipos e o número de instâncias do EC2 para criar os nós que vão processar os dados. O Hadoop distribui o processamento dos dados entre os vários nós do cluster. O nó principal é responsável por acompanhar a integridade dos nós core e de tarefas e por sondar os nós para obter o status dos resultados de trabalhos. Os nós core e de tarefa realizam o processamento real dos dados. Se você tem um cluster com um único nó, este nó serve como nó principal e também como nó core.

O parâmetro `KeepJobAlive` em uma solicitação `RunJobFlow` determina se um cluster deve ser encerrado quando não tem mais etapas para executar. Defina este valor como `False` quando você sabe que o cluster está sendo executado como esperado. Quando você estiver tentando resolver problemas no fluxo de trabalho e adicionando etapas enquanto a execução do cluster é suspensa, defina este valor como `True`. Isso reduz a quantidade de tempo e as despesas de upload dos resultados para o Amazon Simple Storage Service (Amazon S3), apenas para repetir o processo após a modificação de uma etapa para reiniciar o cluster.

Em caso `KeepJobAlive` `true` afirmativo, depois de fazer com que o cluster conclua seu trabalho, você deve enviar uma `TerminateJobFlows` solicitação ou o cluster continuará em execução e gerará AWS cobranças.

Para obter mais informações sobre parâmetros exclusivos de`RunJobFlow`, consulte RunJobFlow. Para obter mais informações sobre os parâmetros genéricos na solicitação, consulte Common request parameters.

Zonas de disponibilidade no Amazon EMR

O Amazon EMR usa instâncias do EC2 como nós para o processamento de clusters. Essas instâncias do EC2 têm locais que são compostos por regiões e zonas de disponibilidade. As regiões são dispersas e localizadas em diferentes áreas geográficas. As zonas de disponibilidade são locais distintos dentro de uma região, que são isolados de falhas que ocorrem em outras zonas de disponibilidade. Cada zona de disponibilidade fornece conectividade de rede de baixa latência e custo reduzido para outras zonas de disponibilidade na mesma região. Para obter uma lista de regiões e endpoints para o Amazon EMR, consulte Regions and endpoints na Referência geral da Amazon Web Services.

O parâmetro `AvailabilityZone` especifica a localização geral do cluster. Esse parâmetro é opcional e, em geral, não recomendamos o seu uso. Quando `AvailabilityZone` não é especificado, o Amazon EMR escolhe automaticamente o melhor valor de `AvailabilityZone` para o cluster. Esse parâmetro pode ser útil se você desejar compartilhar os locais de suas instâncias com outras instâncias existentes em execução e seu cluster precisar ler ou gravar dados dessas instâncias. Para obter mais informações, consulte o Guia do usuário do Amazon EC2.

Como usar arquivos e bibliotecas adicionais em clusters do Amazon EMR

Em algumas ocasiões você pode querer usar arquivos adicionais ou bibliotecas personalizadas com seus aplicativos de mapeador ou reducer. Por exemplo, você pode querer usar uma biblioteca que converte um arquivo PDF em texto simples.

Para armazenar um arquivo em cache a ser usado pelo mapeador ou reducer quando usarem o streaming do Hadoop

- No campo `args` do JAR, adicione o seguinte argumento:

  ```
  -cacheFile s3://bucket/path_to_executable#local_path
  ```

  O arquivo `local_path` está no diretório de trabalho do mapeador, que pode referenciar o arquivo.

Uso de SDKs para chamar APIs do Amazon EMR

Tópicos

Os AWS SDKs fornecem funções que envolvem a API e cuidam de muitos detalhes da conexão, como calcular assinaturas, lidar com novas tentativas de solicitação e tratamento de erros. Os SDKs também contêm exemplos de código, tutoriais e outros recursos para ajudar você a começar a criar aplicativos que chamam. AWS Chamar as funções do wrapper em um SDK pode simplificar muito o processo de criação de um AWS aplicativo.

Para obter mais informações sobre como baixar e usar os AWS SDKs, consulte SDKs em Tools for Amazon Web Services.

Usando o AWS SDK for Java para criar um cluster do Amazon EMR

AWS SDK for Java Ele fornece três pacotes com a funcionalidade do Amazon EMR:

- com.amazonaws.services.elasticmapreduce
- com.amazonaws.services.elasticmapreduce.model
- com.amazonaws.services.elasticmapreduce.util

Para obter mais informações sobre esses pacotes, consulte Referência da API do AWS SDK for Java.

O exemplo a seguir ilustra como os SDKs podem simplificar a programação com o Amazon EMR. O exemplo de código apresentado abaixo usa o objeto `StepFactory`, uma classe auxiliar para criar tipos de etapas comuns do Amazon EMR, para criar um cluster do Hive interativo com a depuração habilitada.

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;
```

```
public class Main {

 public static void main(String[] args) {
  AWSCredentialsProvider profile = null;
  try {
   credentials_profile = new ProfileCredentialsProvider("default"); // specifies any
 named profile in
                                    // .aws/credentials as the credentials provider
  } catch (Exception e) {
   throw new AmazonClientException(
     "Cannot load credentials from .aws/credentials file. " +
       "Make sure that the credentials file exists and that the profile name is defined
 within it.",
     e);
  }

  // create an EMR client using the credentials and region specified in order to
  // create the cluster
  AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
    .withCredentials(credentials_profile)
    .withRegion(Regions.US_WEST_1)
    .build();

  // create a step to enable debugging in the AWS Management Console
  StepFactory stepFactory = new StepFactory();
  StepConfig enabledebugging = new StepConfig()
    .withName("Enable debugging")
    .withActionOnFailure("TERMINATE_JOB_FLOW")
    .withHadoopJarStep(stepFactory.newEnableDebuggingStep());

  // specify applications to be installed and configured when EMR creates the
  // cluster
  Application hive = new Application().withName("Hive");
  Application spark = new Application().withName("Spark");
  Application ganglia = new Application().withName("Ganglia");
  Application zeppelin = new Application().withName("Zeppelin");

  // create the cluster
  RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("MyClusterCreatedFromJava")
    .withReleaseLabel("emr-5.20.0") // specifies the EMR release version label, we
 recommend the latest release
    .withSteps(enabledebugging)
```

```
    .withApplications(hive, spark, ganglia, zeppelin)
    .withLogUri("s3://path/to/my/emr/logs") // a URI in S3 for log files is required
 when debugging is enabled
    .withServiceRole("EMR_DefaultRole") // replace the default with a custom IAM
 service role if one is used
    .withJobFlowRole("EMR_EC2_DefaultRole") // replace the default with a custom EMR
 role for the EC2 instance
                        // profile if one is used
    .withInstances(new JobFlowInstancesConfig()
      .withEc2SubnetId("subnet-12ab34c56")
      .withEc2KeyName("myEc2Key")
      .withInstanceCount(3)
      .withKeepJobFlowAliveWhenNoSteps(true)
      .withMasterInstanceType("m4.large")
      .withSlaveInstanceType("m4.large"));

  RunJobFlowResult result = emr.runJobFlow(request);
  System.out.println("The cluster ID is " + result.toString());

 }

}
```

No mínimo, você deve passar uma função de serviço e uma função de fluxo de trabalho correspondentes ao EMR_ DefaultRole e ao EMR_EC2_, respectivamente. DefaultRole Você pode fazer isso invocando esse AWS CLI comando para a mesma conta. Primeiro, verifique se as funções já existem:

```
aws iam list-roles | grep EMR
```

Tanto o perfil da instância (EMR_EC2_DefaultRole) quanto a função de serviço (EMR_DefaultRole) serão exibidos se existirem:

```
"RoleName": "EMR_DefaultRole",
            "Arn": "arn:aws:iam::AccountID:role/EMR_DefaultRole"
            "RoleName": "EMR_EC2_DefaultRole",
            "Arn": "arn:aws:iam::AccountID:role/EMR_EC2_DefaultRole"
```

Se os perfis padrão não existirem, você poderá usar o seguinte comando para criá-los:

```
aws emr create-default-roles
```

Gerenciamento de cotas de serviço do Amazon EMR

Tópicos

Os tópicos desta seção descrevem as cotas de serviço do EMR (anteriormente chamadas de limites de serviço), como gerenciá-las no e quando é vantajoso usar CloudWatch eventos em vez de cotas de serviço para monitorar clusters e acionar ações. AWS Management Console

O que são as cotas de serviço do Amazon EMR

Sua AWS conta tem cotas de serviço padrão, também conhecidas como limites, para cada AWS serviço. O serviço EMR tem dois tipos de limites:

- Limites de recursos: você pode usar o EMR para criar recursos do EC2. Contudo, estes recursos do EC2 estão sujeitos a cotas de serviço. As limitações de recursos nesta categoria são:
  - O número máximo de clusters ativos que podem ser executados ao mesmo tempo.
  - O número máximo de instâncias ativas por grupo de instâncias.
- Limites de APIs: ao usar APIs do EMR, os dois tipos de limitações são:
  - Limite de intermitência: este é o número máximo de chamadas de API que você pode fazer de uma só vez. Por exemplo, o número máximo de solicitações de AddInstanceFleet API que você pode fazer por segundo é definido como 5 chamadas/segundo como padrão. Isso significa que o limite de intermitência da AddInstanceFleet API é de 5 chamadas/segundo ou que, a qualquer momento, você pode fazer no máximo 5 AddInstanceFleet chamadas de API. Entretanto, depois de usar o limite de intermitência, as chamadas subsequentes serão limitadas pelo limite de taxa.
  - Limite de taxa: esta é a taxa de reabastecimento da capacidade de expansão da API. Por exemplo, a taxa de reabastecimento de AddInstanceFleet chamadas é definida como 0,5 chamadas/segundo como padrão. Isso significa que, depois de atingir o limite de intermitência, você terá que esperar, no mínimo, dois segundos (0,5 chamadas por segundo X 2 segundos = 1 chamada) para chamar a API. Se você fizer uma chamada antes disso, sofrerá o controle de utilização pelo serviço Web do EMR. A qualquer momento, você pode fazer somente a quantidade de chamadas correspondente à capacidade de expansão sem sofrer o controle de

utilização. A cada segundo adicional que você espera, a capacidade de expansão aumenta em 0,5 chamadas até atingir o limite máximo de cinco, que corresponde ao limite de intermitência.

Como gerenciar cotas de serviço do Amazon EMR

O Service Quotas é um AWS recurso que você pode usar para visualizar e gerenciar suas cotas ou limites de serviços do Amazon EMR a partir de um local central usando a API ou a AWS Management Console CLI. Para saber mais sobre como visualizar quotas e solicitar aumentos, consulte AWS service quotas na Referência geral da Amazon Web Services.

Para algumas APIs, configurar um CloudWatch evento pode ser uma opção melhor do que aumentar as cotas de serviço. Você também pode economizar tempo usando CloudWatch para definir alarmes e acionar solicitações de aumento de forma proativa, antes de atingir a cota de serviço. Para obter mais detalhes, consulte Quando configurar eventos do EMR em CloudWatch.

Quando configurar eventos do EMR em CloudWatch

Para algumas APIs de pesquisa, como DescribeCluster,, e DescribeStep ListClusters, configurar um CouldWatch evento pode reduzir o tempo de resposta às mudanças e liberar suas cotas de serviço. Por exemplo, se você tiver uma função do Lambda configurada para ser executada quando o estado de um cluster for alterado, como quando uma etapa for concluída ou um cluster for encerrado, você poderá usar esse acionador para iniciar a próxima ação em seu fluxo de trabalho em vez de aguardar pela próxima sondagem. Caso contrário, se você tiver instâncias dedicadas do Amazon EC2 ou funções do Lambda sondando constantemente a API do EMR em busca de alterações, você não somente desperdiçará recursos de computação, mas também poderá atingir sua cota de serviço.

A seguir são apresentados alguns casos nos quais você pode se beneficiar ao migrar para uma arquitetura orientada a eventos.

Caso 1: Sondagem do EMR DescribeCluster usando chamadas de API para conclusão da etapa

Example Pesquisando o EMR DescribeCluster usando chamadas de API para conclusão da etapa

Um padrão comum é enviar uma etapa para um cluster em execução e consultar o Amazon EMR para obter o status da etapa, normalmente usando DescribeCluster as DescribeStep APIs ou. Essa tarefa também pode ser realizada com atraso mínimo ao se conectar ao evento de alteração de etapa ou de status do Amazon EMR.

Este evento inclui as informações apresentadas a seguir em sua carga útil.

```
{
  "version": "0",
  "id": "999cccaa-eaaa-0000-1111-123456789012",
  "detail-type": "EMR Step Status Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T20:53:09Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "ERROR",
    "actionOnFailure": "CONTINUE",
    "stepId": "s-ZYXWVUTSRQPON",
    "name": "CustomJAR",
    "clusterId": "j-123456789ABCD",
    "state": "FAILED",
    "message": "Step s-ZYXWVUTSRQPON (CustomJAR) in Amazon EMR cluster j-123456789ABCD
 (Development Cluster) failed at 2016-12-16 20:53 UTC."
  }
}
```

No mapa detalhado, uma função do Lambda pode analisar “state”, “stepId” ou “clusterId” para localizar informações pertinentes.

Caso 2: sondagem do EMR para clusters disponíveis para a execução de fluxos de trabalho

Example Sondagem do EMR para clusters disponíveis para a execução de fluxos de trabalho

Um padrão para clientes que executam vários clusters é executar fluxos de trabalho em clusters assim que estiverem disponíveis. Se houver muitos clusters em execução e um fluxo de trabalho precisar ser executado em um cluster que está aguardando, um padrão pode ser pesquisar o EMR DescribeCluster usando ListClusters chamadas de API para os clusters disponíveis. Outra maneira de reduzir o atraso em saber quando um cluster está pronto para uma etapa seria processar o evento de alteração de estado do cluster do Amazon EMR.

Este evento inclui as informações apresentadas a seguir em sua carga útil.

```
{
```

```
  "version": "0",
  "id": "999cccaa-eaaa-0000-1111-123456789012",
  "detail-type": "EMR Cluster State Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T20:43:05Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "INFO",
    "stateChangeReason": "{\"code\":\"\"}",
    "name": "Development Cluster",
    "clusterId": "j-123456789ABCD",
    "state": "WAITING",
    "message": "Amazon EMR cluster j-123456789ABCD ..."
  }
}
```

Para este evento, uma função do Lambda pode ser configurada para enviar imediatamente um fluxo de trabalho em espera para um cluster assim que seu status for alterado para WAITING.

Caso 3: sondagem do EMR para o encerramento de um cluster

Example Sondagem do EMR para o encerramento de um cluster

Um padrão comum para clientes que executam vários clusters do EMR é sondar o Amazon EMR em busca de clusters encerrados para que o trabalho não seja mais enviado a eles. Você pode implementar esse padrão com as chamadas de ListClusters API DescribeCluster e usando o evento Amazon EMR Cluster State Change em.

Após o encerramento do cluster, o evento emitido é semelhante ao exemplo apresentado a seguir.

```
{
  "version": "0",
  "id": "1234abb0-f87e-1234-b7b6-000000123456",
  "detail-type": "EMR Cluster State Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T21:00:23Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "INFO",
```

```
    "stateChangeReason": "{\"code\":\"USER_REQUEST\",\"message\":\"Terminated by user
 request\"}",
    "name": "Development Cluster",
    "clusterId": "j-123456789ABCD",
    "state": "TERMINATED",
    "message": "Amazon EMR Cluster jj-123456789ABCD (Development Cluster) has
 terminated at 2016-12-16 21:00 UTC with a reason of USER_REQUEST."
  }
}
```

A seção “Detalhes” da carga útil inclui o clusterId e o estado que podem ser utilizados.

Glossário do AWS

Para obter a terminologia mais recente da AWS, consulte o glossário da AWS na Referência do Glossário da AWS.

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.