
Amazon Redshift

Guia de conceitos básicos



Amazon Redshift: Guia de conceitos básicos

Copyright © 2022 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens comerciais da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestigie a Amazon. Todas as outras marcas comerciais que não pertencem à Amazon pertencem a seus respectivos proprietários, que podem ou não ser afiliados, patrocinados pela Amazon ou ter conexão com ela.

Table of Contents

Conceitos básicos do Amazon Redshift	1
Pré-requisitos	1
Cadastre-se no AWS	2
Determinar regras de firewall	2
Conceitos e fluxo de processamento de dados do Amazon Redshift	2
Conceitos do Amazon Redshift	2
Fluxo de processamento de dados típico do Amazon Redshift	3
Conceitos básicos do Amazon Redshift	5
Conceitos básicos do console do Amazon Redshift	5
Conectar-se ao Amazon Redshift	6
Conceitos básicos de clusters e carregamento de dados	7
Usar um conjunto de dados de amostra	7
Trazer seus próprios dados para o Amazon Redshift	10
Conceito básico de tarefas comuns de banco de dados	20
Tarefa 1: Criar um banco de dados	20
Tarefa 2: Criar um usuário	20
Tarefa 3: Criar um esquema	21
Tarefa 4: Criar uma tabela	22
Tarefa 5: Carregar dados de amostra	23
Etapa 6: Consultar as tabelas do sistema	23
Tarefa 7: Cancelar uma consulta	26
Etapa 8: Limpar os recursos	28
Conceitos básicos do Amazon Redshift Serverless	30
Conceitos básicos do console do Amazon Redshift Serverless	30
Conectar-se ao Amazon Redshift Serverless	30
Conceitos básicos do Amazon Redshift Serverless e do compartilhamento de dados	31
Usar um conjunto de dados de amostra	32
Trazer seus próprios dados para o Amazon Redshift Serverless	34
Conceitos básicos da consulta fora do banco de dados	36
Conceitos básicos sobre como consultar data lakes	36
Conceitos básicos sobre consulta de fontes de dados remotas	37
Conceitos básicos sobre acesso a dados em outros clusters	37
Conceitos básicos sobre treinamento de modelos de ML com dados do Redshift	37
Recursos adicionais	39
Histórico do documento	40

Conceitos básicos do Amazon Redshift

Bem-vindo ao Guia de conceitos básicos do Amazon Redshift. O Amazon Redshift é um serviço de data warehouse em escala de petabytes totalmente gerenciado no Nuvem AWS. Um data warehouse do Amazon Redshift é um conjunto de recursos de computação chamados nós, que são organizados em um grupo chamado cluster. Cada cluster executa um mecanismo do Amazon Redshift e contém um ou mais bancos de dados.

Se você for um usuário iniciante do Amazon Redshift, recomendamos que comece lendo as seguintes seções:

- [Visão geral de gerenciamento do Amazon Redshift](#)— Neste tópico, você encontrará uma visão geral do Amazon Redshift.
- [Destaques do serviço e preços](#) — Nesta página de detalhes do produto, você pode encontrar detalhes sobre os destaques do serviço e os preços do Amazon Redshift.
- [Guia de conceitos básicos do Amazon Redshift](#) (este guia): neste guia, você encontrará um tutorial sobre o uso do Amazon Redshift para criar um cluster de exemplo e trabalhar com dados de exemplo.

Neste guia, você encontra tutoriais que orientam sobre:

- [Conceitos básicos do console do Amazon Redshift](#) (p. 5)
- [Conectar-se ao Amazon Redshift](#) (p. 6)
- [Conceitos básicos de clusters do Amazon Redshift e carregamento de dados](#) (p. 7)
- [Conceito básico de tarefas comuns de banco de dados](#) (p. 20)
- [Conceitos básicos sobre como consultar seu data lake](#) (p. 36)
- [Conceitos básicos sobre consulta de dados em fontes de dados remotas](#) (p. 37)
- [Conceitos básicos sobre acesso a dados em outros clusters do Amazon Redshift](#) (p. 37)
- [Conceitos básicos sobre treinamento de modelos de Machine Learning com dados do Amazon Redshift](#) (p. 37)

Se sua organização estiver qualificada, você poderá criar um cluster no programa gratuito de teste do Amazon Redshift. Para fazer isso, escolha [Teste gratuito](#) para criar uma configuração com o tipo de nó dc2.large. Para obter mais informações sobre a escolha de um teste gratuito, consulte [Teste gratuito do Amazon Redshift](#).

Tópicos

- [Pré-requisitos](#) (p. 1)
- [Conceitos e fluxo de processamento de dados do Amazon Redshift](#) (p. 2)

Pré-requisitos

Antes de começar a configurar um cluster do Amazon Redshift, conclua os seguintes pré-requisitos nesta seção:

- [Cadastre-se no AWS](#) (p. 2)
- [Determinar regras de firewall](#) (p. 2)

Cadastre-se no AWS

Caso você ainda não tenha uma Conta da AWS, cadastre-se. Se você já tem uma conta, pode pular esse pré-requisito e usar sua conta existente.

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 inserir um código de verificação no teclado do telefone.

Quando você se cadastra em uma Conta da AWS, é criado um usuário raiz da Conta da AWS. O usuário raiz tem acesso a todos os Serviços da AWS e recursos da conta. Como prática recomendada de segurança, [atribua acesso administrativo a um usuário administrativo](#) e use somente o usuário raiz para executar [tarefas que exijam acesso do usuário raiz](#).

Determinar regras de firewall

Como parte deste tutorial, você especificará uma porta quando ativar o cluster do Amazon Redshift. Você também criará uma regra de entrada em um grupo de segurança para conceder acesso ao cluster por meio da porta.

Se o seu computador cliente estiver atrás de um firewall, verifique se você conhece uma porta aberta que pode ser usada. Essa porta aberta permite que você se conecte ao cluster a partir de uma ferramenta de cliente SQL e execute consultas. Se você não souber isso, trabalhe com alguém que entenda as regras de firewall da rede para determinar uma porta aberta em seu firewall.

Embora o Amazon Redshift use a porta 5439 por padrão, a conexão não funcionará se essa porta não estiver aberta no firewall. Não é possível alterar o número da porta do cluster do Amazon Redshift depois que ela é criada. Portanto, certifique-se de especificar uma porta que funciona em seu ambiente durante o processo de inicialização.

Esse pré-requisito só se aplica quando você traz seus próprios dados para o Amazon Redshift. Para mais informações, consulte [Trazer seus próprios dados para o Amazon Redshift \(p. 10\)](#).

Conceitos e fluxo de processamento de dados do Amazon Redshift

Nas seções a seguir, você encontra os principais conceitos para o Amazon Redshift, bem como uma descrição e um diagrama do fluxo de processamento de dados típico do Amazon Redshift:

- [Conceitos do Amazon Redshift \(p. 2\)](#)
- [Fluxo de processamento de dados típico do Amazon Redshift \(p. 3\)](#)

Conceitos do Amazon Redshift

Veja alguns dos principais conceitos do Amazon Redshift:

- Cluster: em cluster é o principal componente da infraestrutura de um data warehouse do Amazon Redshift.

Um cluster é composto de um ou mais nós de computação. Os nós de computação executam o código compilado.

Se um cluster for provisionado com dois ou mais nós de computação, um nó líder adicional coordenará os nós de computação. O nó líder aborda a comunicação externa com aplicações, como ferramentas de business intelligence e editores de consulta. O aplicativo cliente interage diretamente somente com o nó líder. Os nós de computação são transparentes a aplicativos externos.

- Banco de dados: um cluster contém um ou mais bancos de dados.

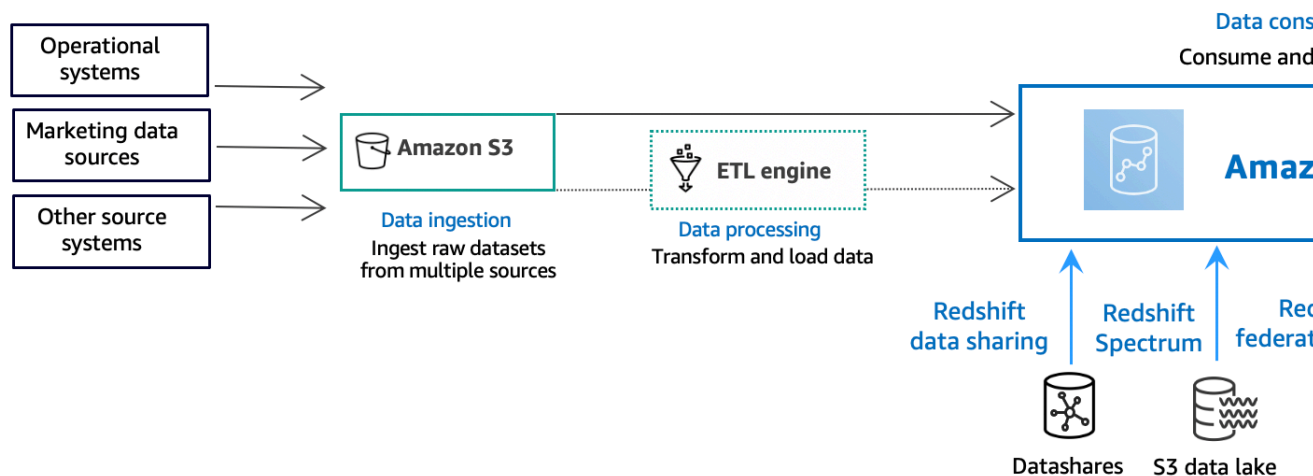
Os dados do usuário são armazenados em um ou mais bancos de dados nos nós de computação. O cliente SQL se comunica com o nó líder, que, por sua vez, coordena as consultas em execução com os nós de computação. Para obter mais informações sobre nós de computação e nós líderes, consulte [Arquitetura do sistema de data warehouse](#). Em um banco de dados, os dados do usuário são organizados em um ou mais esquemas.

O Amazon Redshift é um sistema de gerenciamento de banco de dados relacional (RDBMS) e é compatível com outras aplicações RDBMS. Oferece a mesma funcionalidade de um RDBMS típico, inclusive funções de processamento de transações online (OLTP), como inserir e excluir dados. O Amazon Redshift também é otimizado para análise em lote de alta performance e emissão de relatórios de conjuntos de dados.

A seguir, você encontra uma descrição do fluxo típico de processamento de dados no Amazon Redshift, juntamente com descrições de diferentes partes do fluxo. Para obter mais informações sobre a arquitetura do sistema do Amazon Redshift, consulte [Arquitetura do sistema de data warehouse](#).

Fluxo de processamento de dados típico do Amazon Redshift

O diagrama a seguir ilustra um fluxo de processamento de dados típico no Amazon Redshift.



Um data warehouse do Amazon Redshift é um sistema de gerenciamento e consulta de banco de dados relacional de classe empresarial. O Amazon Redshift oferece suporte a conexões de clientes com muitos tipos de aplicações, incluindo business intelligence (BI), relatórios, dados e ferramentas analíticas. Ao executar consultas analíticas, você recupera, compara e avalia grandes volumes de dados em operações de várias etapas para produzir um resultado final.

Na camada de ingestão de dados, diferentes tipos de origem dos dados carregam continuamente dados estruturados, semiestruturados ou não estruturados para a camada de armazenamento de dados. Essa área de armazenamento de dados serve como uma área de preparação que armazena dados em diferentes estados de disponibilidade para consumo. Um bucket do Amazon Simple Storage Service (Amazon S3) é um exemplo de armazenamento.

Na camada opcional processamento de dados, os dados de origem passam por pré-processamento, validação e transformação usando pipelines extrair, transformar e carregar (ETL) ou extrair, carregar e transformar (ELT). Esses conjuntos de dados brutos são refinados por meio de operações ETL. Um exemplo de um mecanismo ETL é o AWS Glue.

Na camada de consumo de dados, os dados são carregados em seu cluster do Amazon Redshift, onde é possível executar workloads analíticas.

Os dados também podem ser consumidos para workloads analíticas da seguinte forma:

- Use unidades de compartilhamento de dados para compartilhar dados dinâmicos entre clusters do Amazon Redshift para fins de leitura com relativa segurança e facilidade. É possível compartilhar dados em diferentes níveis, como bancos de dados, esquemas, tabelas, visualizações (inclusive visualizações regulares, de vinculação tardia e materializadas) e funções definidas pelo usuário (UDFs) do SQL.

Para obter mais informações sobre compartilhamento de dados, consulte [Conceitos básicos sobre acesso a dados em outros clusters do Amazon Redshift \(p. 37\)](#).

- Utilize o Amazon Redshift Spectrum para consultar dados em arquivos do Amazon S3 sem ter que carregar os dados nas tabelas do Amazon Redshift. O Amazon Redshift oferece um recurso SQL projetado para processamento analítico online (OLAP) rápido de conjuntos de dados muito grandes que são armazenados em clusters do Amazon Redshift e data lakes do Amazon S3.

Para obter mais informações sobre o Redshift Spectrum, consulte [Conceitos básicos sobre como consultar seu data lake \(p. 36\)](#).

- Faça junção de dados de bancos de dados relacionais, como o Amazon Relational Database Service (Amazon RDS) e Amazon Aurora ou do Amazon S3 a dados de seu banco de dados do Amazon Redshift usando uma consulta federada. Você pode usar o Amazon Redshift para consultar dados operacionais diretamente (sem movê-los), aplicar transformações e inserir dados em suas tabelas do Amazon Redshift.

Para obter mais informações sobre consultas federadas, consulte [Conceitos básicos sobre consulta de dados em fontes de dados remotas \(p. 37\)](#).

- O machine learning (ML) do Amazon Redshift cria modelos, usando dados fornecidos e metadados associados a entradas de dados. Esses modelos capturam padrões nos dados de entrada. Você pode usar esses modelos para gerar previsões para novos dados de entrada. O Amazon Redshift trabalha com o Amazon SageMaker Autopilot para obter automaticamente o melhor modelo e disponibilizar a função de previsão no Amazon Redshift.

Para obter mais informações sobre o Amazon Redshift ML, consulte [Conceitos básicos sobre treinamento de modelos de Machine Learning com dados do Amazon Redshift \(p. 37\)](#).

Conceitos básicos do Amazon Redshift

Se estiver usando o Amazon Redshift pela primeira vez, recomendamos que leia as seguintes seções que ajudarão a começar a usar o Amazon Redshift.

Tópicos

- [Conceitos básicos do console do Amazon Redshift](#) (p. 5)
- [Conectar-se ao Amazon Redshift](#) (p. 6)
- [Conceitos básicos de clusters do Amazon Redshift e carregamento de dados](#) (p. 7)
- [Conceito básico de tarefas comuns de banco de dados](#) (p. 20)

Conceitos básicos do console do Amazon Redshift

Depois de fazer login no console do Amazon Redshift, você poderá criar e gerenciar todos os objetos do Amazon Redshift, inclusive clusters, bancos de dados e nós. Você também pode visualizar consultas, executar consultas e executar outras operações de linguagem de definição de dados (DDL) e linguagem de manipulação de dados (DML).

Se estiver usando o Amazon Redshift pela primeira vez no Amazon Redshift, recomendamos que leia as páginas Painel, Clusters e Editor de consultas v2 para começar a usar o console.

Para começar a usar o console do Amazon Redshift, assista ao seguinte vídeo: [Getting started with Amazon Redshift](#).

A seguir, você encontra uma captura de tela do console do Amazon Redshift e descrições de suas seções.

A seguir, você encontra descrições dos itens do painel de navegação do console do Amazon Redshift:

- Amazon Redshift serverless (Amazon Redshift sem servidor): acesse e analise dados sem a necessidade de configurar, ajustar e gerenciar clusters provisionados do Amazon Redshift.
- Provisioned clusters dashboard (Painel de clusters provisionados): confira Cluster metrics (Métricas de cluster) e Query overview (Visão geral da consulta) para obter insights sobre dados de métricas (como utilização da CPU) e informações de consulta. Usá-los pode ajudar você a determinar se seus dados de performance são anormais em um intervalo de tempo especificado.
- Clusters: veja uma lista de clusters de sua conta da AWS, escolha um cluster para iniciar a consulta ou execute ações relacionadas ao cluster. Também é possível criar um novo cluster a partir desta página.
- Query editor (Editor de consultas): execute consultas em bancos de dados hospedados em seu cluster do Amazon Redshift, salve consultas para reutilização ou agende a execução delas para um momento futuro (somente no editor de consultas).
- Query editor v2 (Editor de consultas v2): use o editor de consultas v2, que é uma aplicação de cliente SQL baseada na Web separada que você usa para criar e executar consultas no data warehouse do Amazon Redshift. Você pode visualizar seus resultados em gráficos e compartilhar suas consultas com outras pessoas de sua equipe.

- Queries and loads (Consultas e cargas): obtenha informações para referência ou solução de problemas, como uma lista de consultas recentes e o texto SQL de cada consulta.
- Datashares (Unidades de compartilhamento de dados): como administrador de uma conta de produtor, autorize contas de consumidor a acessar unidades de compartilhamento de dados ou escolha não autorizar o acesso. Para usar uma unidade de compartilhamento de dados autorizada, um administrador de conta de consumidor pode associar a unidade de compartilhamento de dados a toda uma conta da AWS ou especificar namespaces de cluster em uma conta. Um administrador também pode recusar uma unidade de compartilhamento de dados.
- Configurations (Configurações): conecte-se a clusters do Amazon Redshift por meio de ferramentas de cliente SQL em conexões JDBC (Java Database Connectivity) e ODBC (Open Database Connectivity). Também é possível configurar um endpoint da Virtual Private Cloud (VPC) gerenciado pelo Amazon Redshift. Isso fornece uma conexão privada entre uma VPC com base no serviço da Amazon VPC que contém um cluster e outra VPC que está executando uma ferramenta cliente.
- Advisor (Consultor): obtenha recomendações específicas sobre as alterações que você pode fazer em seu cluster do Amazon Redshift para priorizar suas otimizações.
- AWS Marketplace: obtenha informações sobre outras ferramentas ou serviços da AWS que funcionam com o Amazon Redshift.
- Alarms (Alarmes): crie alarmes em métricas de cluster para exibir dados de performance e rastrear métricas durante um período especificado.
- Events (Eventos): monitore eventos e receba relatórios sobre informações como a data em que o evento ocorreu, uma descrição ou a origem do evento.
- What's new (Novidades): veja os novos recursos e as atualizações de produtos do Amazon Redshift.

Conectar-se ao Amazon Redshift

Para se conectar a clusters do Amazon Redshift, na página Clusters, expanda Connect to Amazon Redshift clusters (Conectar-se a clusters do Amazon Redshift) e siga um destes procedimentos:

- Use o editor de consulta v2 para executar consultas em bancos de dados hospedados por seu cluster do Amazon Redshift. Depois de criar o cluster, é possível executar consultas imediatamente usando o editor de consultas v2.

Para obter mais informações, acesse [Consultar um banco de dados usando o editor de consultas v2 do Amazon Redshift](#).

- Conecte-se ao Amazon Redshift a partir de suas ferramentas de cliente com drivers JDBC ou ODBC copiando a URL do driver JDBC ou ODBC.

Para trabalhar com dados no cluster, você precisa dos drivers do JDBC ou do ODBC para estabelecer a conexão do computador cliente ou da instância. Escreva o código de suas aplicações de modo que usem as operações de API de acesso a dados do JDBC ou do ODBC ou use as ferramentas do cliente SQL que oferecem suporte a JDBC ou ODBC.

Para obter mais informações sobre como encontrar a string de conexão do cluster, consulte [Encontrar a string de conexão do cluster](#).

- Se a ferramenta cliente SQL exigir um driver, você poderá baixar um driver específico do sistema operacional para se conectar ao Amazon Redshift a partir de suas ferramentas de cliente.

Para obter mais informações sobre como instalar o driver apropriado para o cliente SQL, consulte [Configurar uma conexão do driver JDBC versão 2.0](#).

Para obter mais informações sobre como configurar uma conexão ODBC, consulte [Configurar uma conexão ODBC](#).

Conceitos básicos de clusters do Amazon Redshift e carregamento de dados

Nesta seção, você pode encontrar dois tutoriais que orientam você pelo processo de criação de um cluster de amostra do Amazon Redshift. Em um, você usa um conjunto de dados de exemplo e, no outro, traz seu próprio conjunto de dados.

Antes de começar, verifique se você cumpre os pré-requisitos. Para obter mais informações, consulte [Pré-requisitos \(p. 1\)](#).

Tópicos

- [Usar um conjunto de dados de amostra \(p. 7\)](#)
- [Trazer seus próprios dados para o Amazon Redshift \(p. 10\)](#)

Usar um conjunto de dados de amostra

Neste tutorial, você passará pelo processo para criar um cluster do Amazon Redshift usando um conjunto de dados de amostra. O Amazon Redshift carrega automaticamente o conjunto de dados de exemplo quando você está criando um novo cluster. Você pode consultar imediatamente os dados depois que o cluster é criado.

Antes de começar a configurar um cluster do Amazon Redshift, certifique-se de concluir o [Cadastro-se no AWS \(p. 2\)](#) e [Determinar regras de firewall \(p. 2\)](#).

Neste tutorial, você realizará as seguintes etapas:



Tópicos

- [Etapa 1: Criar um cluster de amostra do Amazon Redshift \(p. 7\)](#)
- [Etapa 2: Experimentar exemplo de consultas usando o editor de consultas \(p. 9\)](#)

Important

O exemplo de cluster que você criar será executado em um ambiente dinâmico. A taxa sob demanda é US\$ 0,25 por hora para usar o exemplo de cluster criado neste tutorial até que você o exclua. Para obter mais informações sobre preço, consulte [Preço do Amazon Redshift](#). Se você tiver dúvidas ou tiver dúvidas, entre em contato com a equipe do Amazon Redshift postando em nosso [fórum de discussão](#).

Este tutorial não se destina a ambientes de produção e não discute opções em profundidade. Depois de concluir as etapas neste tutorial, você pode usar [Recursos adicionais \(p. 39\)](#) para localizar informações mais detalhadas. Essas informações podem ajudar a planejar, implantar e manter seus clusters e trabalhar com os dados do seu data warehouse.

Etapa 1: Criar um cluster de amostra do Amazon Redshift

Quando os pré-requisitos forem concluídos, você pode começar a criar seu cluster do Amazon Redshift, com base em um conjunto de dados de amostra.

Para criar um cluster do Amazon Redshift com base em um conjunto de dados de exemplo:

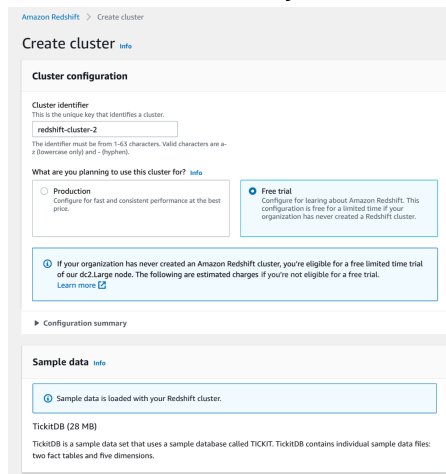
1. Faça login no AWS Management Console e abra o console do Amazon Redshift em <https://console.aws.amazon.com/redshift/>.

2. Para criar um cluster, siga um destes procedimentos:
 - Na página de serviço do Amazon Redshift, escolha Criar cluster. A página Create cluster (Criar cluster) é exibida.
 - No <https://console.aws.amazon.com/redshift/>, escolha DASHBOARD e, depois, escolha Criar cluster.
 - No <https://console.aws.amazon.com/redshift/>, escolha CLUSTERS e, depois, escolha Criar cluster.
3. Na seção Configuração do cluster, especifique um Identificador de cluster. Esse identificador deve ser exclusivo. O identificador deve ter entre 1 e 63 caracteres usando como caracteres válidos a-z (somente minúsculas) e - (hífen).

Para este tutorial, insira **examplecluster**.

4. Se sua organização estiver qualificada, você poderá criar um cluster no programa gratuito de teste do Amazon Redshift. Para fazer isso, escolha Teste gratuito para criar uma configuração com o tipo de nó dc2.large. Para obter mais informações sobre a escolha de um teste gratuito, consulte [Teste gratuito do Amazon Redshift](#).

O console exibe sua seleção, conforme mostrado na captura de tela a seguir.



Se você escolher mais tarde outro tipo de nó, sua organização não será mais elegível para o teste gratuito.

Depois de escolher seu tipo de nó, siga um destes procedimentos:

- Em Sample data (Dados de exemplo), escolha Load sample data (Carregar dados de exemplo) para carregar o conjunto de dados de exemplo em seu cluster do Amazon Redshift. O Amazon Redshift carrega o Tickit do conjunto de dados de exemplo para o banco de dados dev padrão e o esquema public. Você pode começar usando o editor de consulta v2 para consultar dados.
- Para trazer seus próprios dados para o cluster do Amazon Redshift, escolha Produção. Em seguida, em Sample data (Dados de exemplo), escolha Load sample data (Carregar dados de exemplo). Para obter informações sobre como trazer seus próprios dados, consulte [Trazer seus próprios dados para o Amazon Redshift \(p. 10\)](#).

O Amazon Redshift carrega automaticamente o conjunto de dados de exemplo no cluster do Amazon Redshift de exemplo.

5. Na seção Configuração do banco de dados, especifique valores para Nome do usuário administrador e Senha do usuário administrador. Ou escolha Gerar senha para usar uma senha gerada pelo Amazon Redshift.

Para este tutorial, use estes valores:

- Nome do usuário administrador: digite **awsuser**.
- Senha do usuário administrador: insira um valor para a senha.

6. Selecione Create cluster (Criar cluster).

Este tutorial usa o editor de consultas v2 do Amazon Redshift. É possível usar esse editor para consultar dados imediatamente, depois que o Amazon Redshift terminar de criar o cluster.

Também é possível escolher outras ferramentas do cliente SQL que oferecem suporte aos drivers do JDBC ou do ODBC para trabalhar com dados no cluster. Para obter mais informações, consulte [“Conectar-se a um cluster Amazon Redshift usando ferramentas de cliente SQL”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

Etapa 2: Experimentar exemplo de consultas usando o editor de consultas

Quando o Amazon Redshift está criando seu cluster do Amazon Redshift, ele carrega automaticamente o Tickit do conjunto de dados de amostra. A criação de clusters pode levar alguns minutos para ser concluída. Após a conclusão da criação, o status do cluster torna-se ACTIVE. É possível visualizar as tabelas de Tickit de exemplo do conjunto de dados.

Usar o Query Editor

Você pode visualizar as tabelas Tickit de exemplo no editor de consultas v2 escolhendo o cluster, o banco de dados `dev` e o esquema `public`.

Depois que o cluster do Amazon Redshift for criado, em Conectar-se aos clusters do Amazon Redshift, escolha Consultar dados.

No editor de consultas v2, conecte-se a um banco de dados e escolha o nome do cluster no painel de visualização em árvore. Caso seja solicitado, insira os parâmetros de conexão.

Ao se conectar a um cluster e a seus bancos de dados, você fornece um nome do banco de dados e um nome de usuário. Você também fornece parâmetros necessários para um destes métodos de autenticação:

Nome de usuário e senha do banco de dados

Com este método, forneça também um senha para o banco de dados ao qual você está se conectando.

Credenciais temporárias

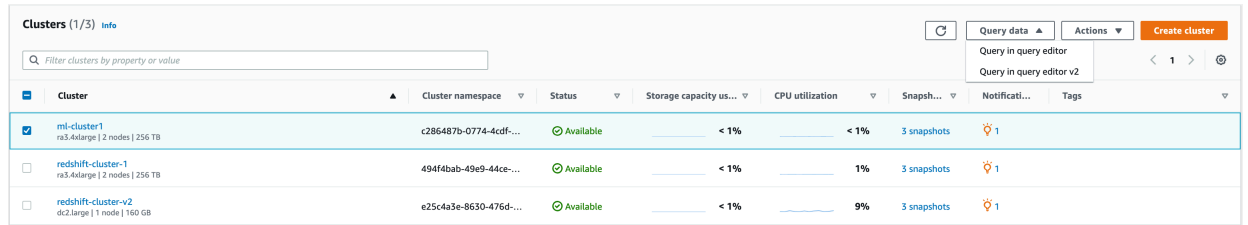
Com esse método, o editor de consultas v2 gera uma senha temporária para se conectar ao banco de dados.

Ao selecionar um cluster com o editor de consultas v2, dependendo do contexto, você pode criar, editar e excluir conexões usando o menu de contexto (clique com o botão direito do mouse).

Por padrão, o Amazon Redshift cria um banco de dados padrão chamado `dev` e um esquema padrão chamado `public`. Para exibir os arquivos de dados individuais da amostra de conjunto de dados, escolha um cluster, acesse o editor de consultas v2 e escolha o banco de dados `dev`, o esquema `public` e `Tables`.

Se preferir, no painel de navegação, escolha Clusters e o cluster em que você deseja consultar dados. Então, em Query data (Consultar dados), escolha Query in query editor (Consultar no editor de consultas)

ou Query in query editor v2 (Consultar no editor de consultas v2) para consultar dados no editor de consultas especificado.



Cluster	Cluster namespace	Status	Storage capacity us...	CPU utilization	Snapsh...	Notificati...	Tags
<input checked="" type="checkbox"/> ml-cluster1 ra3.4xlarge 2 nodes 256 TB	c286487b-0774-4cdf...	Available	< 1%	< 1%	3 snapshots	1	
<input type="checkbox"/> redshift-cluster-1 ra3.4xlarge 2 nodes 256 TB	494f4bab-49e9-44ce...	Available	< 1%	1%	3 snapshots	1	
<input type="checkbox"/> redshift-cluster-v2 dc2.large 1 node 160 GB	e25c4a3e-8630-476d...	Available	< 1%	9%	3 snapshots	1	

Experimentar consultas de exemplo

Experimente alguns exemplos de consultas nos editores de consultas, como mostrado a seguir. Para obter mais informações sobre como trabalhar com o comando SELECT, consulte [SELECT](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

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

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

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

Depois de concluir as etapas neste tutorial, você pode usar [Recursos adicionais \(p. 39\)](#) para localizar informações mais detalhadas. Essas informações podem ajudar a planejar, implantar e manter seus clusters e trabalhar com os dados do seu data warehouse.

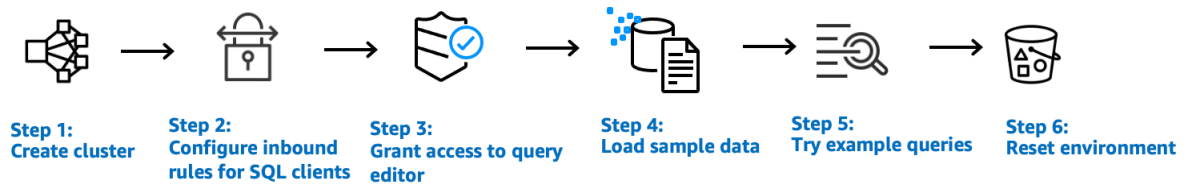
Você também pode experimentar o tutorial [Trazer seus próprios dados para o Amazon Redshift \(p. 10\)](#) para criar um cluster com seu próprio conjunto de dados.

Trazer seus próprios dados para o Amazon Redshift

Neste tutorial, você percorre o processo para criar um cluster do Amazon Redshift trazendo seu próprio conjunto de dados para o Amazon Redshift. Você pode usar esse cluster para avaliar o serviço do Amazon Redshift.

Antes de começar a configurar um cluster do Amazon Redshift, certifique-se de concluir o [Cadastro-se no AWS \(p. 2\)](#) e [Determinar regras de firewall \(p. 2\)](#).

Neste tutorial, você realizará as etapas mostradas a seguir.



Important

O exemplo de cluster que você criar será executado em um ambiente dinâmico. A taxa sob demanda é US\$ 0,25 por hora para usar o exemplo de cluster criado neste tutorial até que você o exclua. Para obter mais informações sobre preços, acesse [a página de preços do Amazon Redshift](#). Se você tiver dúvidas ou tiver dúvidas, entre em contato com a equipe do Amazon Redshift postando em nosso [fórum de discussão](#).

Este tutorial não se destina a ambientes de produção e não discute opções em profundidade. Depois de concluir as etapas neste tutorial, você pode usar [Recursos adicionais](#) (p. 39) para localizar informações mais detalhadas. Essas informações podem ajudar a planejar, implantar e manter seus clusters e trabalhar com os dados do seu data warehouse.

Tópicos

- [Etapa 1: Criar um cluster de amostra do Amazon Redshift](#) (p. 11)
- [Etapa 2: Configurar regras de entrada para clientes SQL](#) (p. 13)
- [Etapa 3: Conceder acesso aos editores de consultas e executar consultas](#) (p. 14)
- [Etapa 4: Carregar dados do Amazon S3 para o Amazon Redshift](#) (p. 15)
- [Etapa 5: Experimentar exemplo de consultas usando o editor de consultas](#) (p. 18)
- [Etapa 6: Redefinir o ambiente](#) (p. 19)

Etapa 1: Criar um cluster de amostra do Amazon Redshift

Para qualquer operação que acesse dados de outro recurso da AWS, seu cluster precisa de permissão para acessar o recurso e os dados no recurso em seu nome. Um exemplo é o uso de um comando COPY para carregar dados do Amazon Simple Storage Service (Amazon S3). Você concede essas permissões usando o AWS Identity and Access Management (IAM). Você pode fazer isso por meio de uma função do IAM associada ao cluster. Ou você pode fornecer a chave de acesso AWS para um usuário do IAM que tenha as permissões necessárias. Para obter mais informações sobre credenciais e permissões de acesso, consulte [Credenciais e permissões de acesso](#).

Para melhor proteger seus dados confidenciais e salvaguardar suas credenciais de acesso da AWS, recomendamos criar uma função do IAM e anexá-la ao seu cluster. Para obter mais informações sobre como fornecer permissões de acesso, consulte [Permissões para acessar outros recursos da AWS](#).

O cluster que você está prestes a criar está ativo e não está sendo executado em uma sandbox. Você será cobrado pelas taxas de uso padrão do Amazon Redshift para o cluster até que o exclua. Se você concluir o tutorial descrito aqui em uma sessão e excluir o cluster quando terminar, o total de cobranças será mínimo.

Para criar um cluster do Amazon Redshift.

1. Faça login no AWS Management Console e abra o console do Amazon Redshift em <https://console.aws.amazon.com/redshift/>.

Important

Se você usar credenciais de usuário do IAM, verifique se o usuário tem as permissões necessárias para realizar as operações de cluster. Para obter mais informações, consulte [“Segurança no Amazon Redshift”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

2. Na parte superior direita, escolha a Região da AWS onde você deseja criar o cluster.
3. No menu de navegação, escolha Clusters e Create cluster (Criar cluster). A página Create cluster (Criar cluster) é exibida.
4. Na seção Configuração do cluster, especifique valores para o Identificador de cluster, o Tipo de nó, e os Nós:
 - Identificador de cluster: digite **examplecluster** para este tutorial. Esse identificador deve ser exclusivo. O identificador deve ter entre 1 e 63 caracteres usando como caracteres válidos a-z (somente minúsculas) e - (hífen).
 - Escolha um dos seguintes métodos para dimensionar o cluster:

Note

A etapa a seguir pressupõe uma região da AWS que oferece suporte a tipos de nó RA3. Para ver uma lista de regiões da AWS compatíveis com tipos de nó RA3, consulte [“Visão geral dos tipos de nó RA3”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

- Se sua região da AWS suporta tipos de nó RA3, escolha Produção ou Teste gratuito para responder à pergunta Para que você está planejando usar esse cluster?

Se sua organização estiver qualificada, você poderá criar um cluster no programa gratuito de teste do Amazon Redshift. Para obter informações sobre como criar clusters usando o programa gratuito de teste, consulte [Usar um conjunto de dados de amostra \(p. 7\)](#). Para fazer isso, selecione Teste gratuito para criar uma configuração com o tipo de nó dc2.large. Para obter mais informações sobre a escolha de um teste gratuito, consulte [Teste gratuito do Amazon Redshift](#).

- Se você não souber o tamanho do cluster, escolha Ajude-me a escolher. Isso inicia uma calculadora de dimensionamento que faz perguntas sobre o tamanho e as características da consulta dos dados que você planeja armazenar no data warehouse.

Se você souber o tamanho necessário do cluster (ou seja, o tipo de nó e o número de nós), escolha Eu escolherei. Em seguida, escolha o Tipo de nó e número de Nós para dimensionar seu cluster para a prova de conceito.

- Para este tutorial, escolha dc2.large em Node type (Tipo de nó) e 2 em Nodes (Nós).
 - Se você escolheu Produção em seu cluster, siga um destes procedimentos:
 - Para usar o conjunto de dados de amostra que o Amazon Redshift fornece, em Dados de amostra, escolha Carregar dados de amostra. O Amazon Redshift carrega o Tckit do conjunto de dados de amostra para o banco de dados dev padrão e o esquema public.
 - Para trazer seus próprios dados para o Amazon Redshift, continue com o restante do tutorial.
5. Na seção Configuração do banco de dados, especifique valores para Nome do banco de dados (opcional), Portabilidade do banco de dados (opcional), Nome do usuário administrador, e Senha do usuário administrador. Ou escolha Gerar senha para usar uma senha gerada pelo Amazon Redshift.

Para este tutorial, use estes valores:

- Nome do banco de dados (opcional): digite **dev**.
 - Portabilidade do banco de dados (opcional): digite **5439**.
 - Nome do usuário administrador: digite **awsuser**.
 - Senha do usuário administrador: insira um valor para a senha.
6. Para este tutorial, crie uma função do IAM e a defina como padrão para o cluster, conforme descrito a seguir. Só pode haver um conjunto de funções do IAM padrão por cluster.
 - a. Em Cluster permissions (Permissões de cluster), em Manage IAM roles (Gerenciar funções do IAM), escolha Create IAM role (Criar função do IAM).
 - b. Especifique um bucket do Amazon S3 para que a função do IAM seja acessada por um destes métodos:

- Escolha No additional Amazon S3 bucket (Nenhum bucket adicional do Amazon S3) para permitir que a função do IAM criada acesse somente os buckets do Amazon S3 que são nomeados como `redshift`.
 - Selecione Any Amazon S3 bucket (Qualquer bucket do Amazon S3) para permitir que a função do IAM criada acesse todos os buckets do Amazon S3.
 - Escolha Specific Amazon S3 buckets (Buckets específicos do Amazon S3) para especificar um ou mais buckets do Amazon S3 para que a função do IAM acesse. Em seguida, escolha um ou mais buckets do Amazon S3 na tabela.
- c. Escolha Create IAM role as default (Criar função do IAM como padrão). O Amazon Redshift cria e define automaticamente a função do IAM como padrão para o cluster.

Como você criou sua função do IAM pelo console, ela tem a política `AmazonRedshiftAllCommandsFullAccess` anexada. Isso permite que o Amazon Redshift copie, carregue, consulte e analise dados de recursos da Amazon em sua conta do IAM.

Para obter informações sobre como gerenciar a função do IAM padrão para um cluster, consulte [Criar uma função do IAM como padrão para o Amazon Redshift](#).

7. (Opcional) Na seção Additional configurations (Configurações adicionais), desative Use defaults (Usar padrões) para modificar as configurações Network and security (Rede e segurança), Database configuration (Configuração do banco de dados), Maintenance (Manutenção), Monitoring (Monitoramento) e Backup.

Em alguns casos, você pode criar o cluster com a opção Load sample data (Carregar dados de exemplo) e ativar o roteamento aprimorado da Amazon VPC. Se for o caso, o cluster de sua Virtual Private Cloud (VPC) precisa acessar o endpoint do Amazon S3 para que os dados sejam carregados.

Há dois métodos para tornar o cluster acessível ao público. Configure um endereço de conversão de endereço de rede (NAT) em sua VPC para que o cluster acesse a Internet. Ou configure um endpoint da VPC do Amazon S3 em sua VPC. Para obter mais informações sobre o roteamento aprimorado do Amazon VPC, consulte [“Habilitar o roteamento aprimorado de VPC”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

8. Selecione Create cluster (Criar cluster).

Etapa 2: Configurar regras de entrada para clientes SQL

Mais adiante neste tutorial, você acessará seu cluster de dentro de uma nuvem privada virtual (VPC) com base no serviço da Amazon VPC. No entanto, se você usar um cliente SQL de fora do firewall para acessar o cluster, certifique-se de conceder acesso de entrada.

Você pode pular esta etapa se planeja acessar o cluster com o editor de consulta do Amazon Redshift de dentro do seu VPC.

Para verificar o firewall e conceder acesso de entrada a seu cluster

1. Verifique as regras do firewall quando seu cluster precisar ser acessado de fora de um firewall. Por exemplo, seu cliente pode ser uma instância do Amazon Elastic Compute Cloud (Amazon EC2) ou um computador externo.

Para obter mais informações sobre regras de firewall, consulte [Security group rules](#) (Regras de grupo de segurança) no Guia do usuário do Amazon EC2 para instâncias do Linux.

2. Para acessar de um cliente externo Amazon EC2, adicione uma regra de entrada ao grupo de segurança anexado ao seu cluster que permite o tráfego de entrada. Adicione regras de grupo de segurança do Amazon EC2 no console do Amazon EC2. Por exemplo, um CIDR/IP de `192.0.2.0/24`

permite que os clientes nesse endereço IP se conectem ao seu cluster. Descubra o CIDR/IP correto para o seu ambiente.

Etapa 3: Conceder acesso aos editores de consultas e executar consultas

Para consultar bancos de dados hospedados por seu cluster do Amazon Redshift, você tem duas opções:

1. Conecte-se ao cluster e execute consultas no AWS Management Console com o editor de consultas.

Se usar um dos editores de consultas, você não precisará baixar e configurar uma aplicação de cliente SQL.

2. Conecte-se ao cluster por meio de uma ferramenta de cliente SQL, como o SQL Workbench/J. Para obter mais informações sobre como usar o SQL Workbench/J, consulte [“Conexão com o cluster usando o SQL Workbench/J”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

Usar um dos editores de consultas do Amazon Redshift é a maneira mais fácil de executar consultas em bancos de dados hospedados por seu cluster do Redshift. Depois de criar seu cluster, você pode executar consultas imediatamente usando o console do Amazon Redshift. Para obter detalhes sobre as considerações ao usar o editor de consultas do Amazon Redshift, consulte [“Consultar um banco de dados usando o Query Editor”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

Conceder acesso ao editore de consultas v2

Na primeira vez que um administrador configura o editor de consultas v2 para sua Conta da AWS, ele escolhe a AWS KMS key que será usada para criptografar recursos do editor de consultas v2. Por padrão, uma chave de propriedade da AWS é usada para criptografar recursos. Ou um administrador pode usar uma chave gerenciada pelo cliente escolhendo o nome do recurso da Amazon (ARN) da chave na página de configuração. Depois que você configurar uma conta, as configurações de criptografia do AWS KMS não poderão ser alteradas. Para obter mais informações, consulte [Configurar a Conta da AWS](#).

Para acessar o editor de consultas v2, é necessário ter permissão. Um administrador pode anexar uma das seguintes políticas gerenciadas pela AWS ao usuário ou à função do IAM para conceder permissões. Essas políticas gerenciadas pela AWS são redigidas com diferentes opções que controlam como os recursos de marcação permitem o compartilhamento de consultas. Você pode usar o console do IAM (<https://console.aws.amazon.com/iam/>) para anexar políticas do IAM.

Você também pode criar sua própria política com base nas permissões concedidas e negadas nas políticas gerenciadas fornecidas. Se usar o editor de políticas de console do IAM para criar sua própria política, escolha SQL Workbench como o serviço para o qual você está criando a política no editor visual. O editor de consultas v2 usa o nome do serviço AWS SQL Workbench no editor visual e no IAM Policy Simulator.

Para obter mais informações, consulte [Acessar o editor de consultas v2](#).

Usar o editor de consultas v2

Para usar o editor de consultas v2 para consultar um banco de dados, consulte [Trabalhar com o editor de consultas v2](#).

Conceder acesso ao editor de consultas

Para usar o editor de consulta do Amazon Redshift, você precisa de permissão. Para habilitar o acesso, anexe as políticas `AmazonRedshiftQueryEditor` e `AmazonRedshiftReadOnlyAccess` do IAM ao usuário do IAM que você usa para acessar o cluster.

Se já criou um usuário do IAM para acessar o Amazon Redshift, você pode anexar as políticas `AmazonRedshiftQueryEditor` e `AmazonRedshiftReadOnlyAccess` para esse usuário. Se ainda não criou um usuário do IAM, crie um e anexe as políticas ao usuário do IAM.

Para anexar as políticas do IAM necessárias para o Query Editor

1. Faça login no AWS Management Console e abra o console do IAM em <https://console.aws.amazon.com/iam/>.
2. Selecione Users (Usuários).
3. Escolha o usuário que precisa de acesso ao Query Editor.
4. Escolha Add permissions (Adicionar permissões).
5. Escolha Attach existing policies directly (Anexar políticas existentes diretamente).
6. Para Policy names (Nomes das políticas), selecione AmazonRedshiftQueryEditor e AmazonRedshiftReadOnlyAccess.
7. Escolha Next: Review (Próximo: revisar).
8. Escolha Add permissions (Adicionar permissões).

Usar o Query Editor

Para usar o editor de consultas para consultar um banco de dados, acesse [Consultar um banco de dados usando o editor de consultas](#).

Etapa 4: Carregar dados do Amazon S3 para o Amazon Redshift

Usar um dos editores de consulta do Amazon Redshift é o modo mais fácil de carregar dados em tabelas. Após criar o cluster, é possível carregar dados do Amazon S3 para o cluster usando o console do Amazon Redshift.

Usar o editor de consultas v2 simplifica o carregamento de dados ao usar o assistente Load data (Carregar dados). Para obter mais informações, consulte [Carregar dados do Amazon S3 para o Amazon Redshift usando o editor de consultas v2](#) (p. 15). Também é possível usar o editor de consultas v2 para criar tabelas e carregar seus dados. Para obter mais informações, consulte [Carregar dados de exemplo do Amazon S3 usando o editor de consultas](#) (p. 15).

Carregar dados do Amazon S3 para o Amazon Redshift usando o editor de consultas v2

Para carregar seus dados do Amazon S3 para o Amazon Redshift, o Amazon Redshift requer uma função do IAM que tenha os privilégios necessários para carregar dados do bucket do Amazon S3 especificado.

Primeiro, conecte-se a um banco de dados. Em seguida, crie algumas tabelas no banco de dados. Depois, carregue seus dados do Amazon S3 para o Amazon Redshift. Para obter mais informações sobre como trabalhar com o editor de consultas v2, acesse [“Trabalhar com o editor de consultas v2”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

O comando COPY gerado e usado no assistente Load data (Carregar dados) do editor de consultas v2 oferece suporte a todos os parâmetros disponíveis para a sintaxe do comando COPY para carregar dados do Amazon S3. Para obter informações sobre o comando COPY e suas opções usadas para copiar a carga do Amazon S3, consulte [COPY do Amazon Simple Storage Service](#) no Guia do desenvolvedor do Amazon Redshift.

Carregar dados de exemplo do Amazon S3 usando o editor de consultas

Nesse momento, você tem um banco de dados chamado dev e está conectado a ele. Então, crie algumas tabelas no banco de dados, faça upload de dados nelas e tente fazer uma consulta. Para sua conveniência, os dados de amostra que você carrega estão disponíveis em um bucket do Amazon S3.

Note

Se estiver usando uma ferramenta do cliente SQL, verifique se o cliente SQL está conectado ao cluster.

Depois de concluir esta etapa, você pode fazer o seguinte:

- Experimentar consultas de exemplo em [Etapa 5: Experimentar exemplo de consultas usando o editor de consultas \(p. 18\)](#).
- Redefinir o ambiente no [Etapa 6: Redefinir o ambiente \(p. 19\)](#).
- Encontrar mais informações sobre o Amazon Redshift em [Recursos adicionais \(p. 39\)](#).

Note

Para tentar consultar dados no editor de consulta sem carregar seus próprios dados, escolha Carregar dados de amostra em Dados de amostra. Se você fizer isso, o Amazon Redshift carregará seu conjunto de dados de amostra no cluster do Amazon Redshift automaticamente durante a criação do cluster.

Para carregar dados de amostra do Amazon S3

1. Crie tabelas.

Se você estiver usando o editor de consulta do Amazon Redshift, copie individualmente e execute as instruções de criação de tabelas a seguir para criar tabelas no banco de dados dev. Para obter mais informações sobre a sintaxe, consulte [CREATE TABLE](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

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

```
create table venue(
  venueid smallint not null distkey sortkey,
  venue name varchar(100),
  venuecity varchar(30),
  venuestate char(2),
  venueseats integer);
```

```
create table category(
  catid smallint not null distkey sortkey,
  catgroup varchar(10),
  catname varchar(10),
  catdesc varchar(50));
```

```
create table date(
  dateid smallint not null distkey sortkey,
  caldate date not null,
```

```
day character(3) not null,  
week smallint not null,  
month character(5) not null,  
qtr character(5) not null,  
year smallint not null,  
holiday boolean default('N');
```

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

```
create table listing(  
  listid integer not null distkey,  
  sellerid integer not null,  
  eventid integer not null,  
  dateid smallint not null sortkey,  
  numtickets smallint not null,  
  priceperticket decimal(8,2),  
  totalprice decimal(8,2),  
  listtime timestamp);
```

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

2. Carregue dados de amostra do Amazon S3 usando o comando COPY.

Note

Recomendamos usar o comando COPY para carregar grandes conjuntos de dados no Amazon Redshift do Amazon S3 ou Amazon DynamoDB. Para obter mais informações sobre sintaxe de COPY, consulte [COPY](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

- a. Faça download do arquivo [ticketdb.zip](#) que contém os arquivos de dados de amostra individuais.
- b. Descompacte e carregue os arquivos individuais em uma pasta `ticket` em seu bucket do Amazon S3 em sua Região da AWS.
- c. Edite os comandos COPY neste tutorial para apontar para os arquivos em seu bucket do Amazon S3. Para obter informações sobre como gerenciar arquivos com o Amazon S3, consulte [Criar e configurar um bucket do S3](#) no Guia do usuário do Amazon Simple Storage Service.
- d. Forneça autenticação ao seu cluster para acessar o Amazon S3 em seu nome para carregar os dados de amostra. Nesta etapa, forneça autenticação ao referenciar a função do IAM que criou e anexou ao cluster nas etapas anteriores.

Os comandos COPY incluem um espaço reservado para o nome do recurso da Amazon (ARN) para a função do IAM, o nome do bucket e uma Região da AWS, conforme mostrado no exemplo a seguir.

```
copy users from 's3://<myBucket>/tickit/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

O comando COPY deve ser semelhante ao exemplo a seguir.

```
copy users from 's3://<myBucket>/tickit/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

Para carregar os dados de exemplo, substitua *<myBucket>* e *<aws-region>* por seus valores nos seguintes comandos COPY. Se você estiver usando o editor de consulta do Amazon Redshift, execute individualmente os comandos a seguir.

```
copy users from 's3://<myBucket>/tickit/allusers_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy venue from 's3://<myBucket>/tickit/venue_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy category from 's3://<myBucket>/tickit/category_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy date from 's3://<myBucket>/tickit/date2008_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy event from 's3://<myBucket>/tickit/allevents_pipe.txt'  
iam_role default  
delimiter '|' timeformat 'YYYY-MM-DD HH:MI:SS' region '<aws-region>;
```

```
copy listing from 's3://<myBucket>/tickit/listings_pipe.txt'  
iam_role default  
delimiter '|' region '<aws-region>;
```

```
copy sales from 's3://<myBucket>/tickit/sales_tab.txt'  
iam_role default  
delimiter '\t' timeformat 'MM/DD/YYYY HH:MI:SS' region '<aws-region>;
```

Etapa 5: Experimentar exemplo de consultas usando o editor de consultas

Agora, tente alguns exemplos de consultas, como mostrado a seguir. Para obter mais informações sobre como trabalhar com o comando SELECT, consulte [SELECT](#) no Guia do desenvolvedor do Amazon Redshift.

```
-- Get definition for the sales table.
```

```
SELECT *
FROM pg_table_def
WHERE tablename = 'sales';

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

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

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

Você criou com sucesso um cluster do Amazon Redshift e consultou dados de seu próprio conjunto de dados usando o editor de consulta do Amazon Redshift.

Etapa 6: Redefinir o ambiente

Ao concluir este tutorial, sugerimos que você reconfigure seu ambiente para o estado anterior, excluindo seu cluster de amostra. Você continuará a ser cobrado pelo serviço do Amazon Redshift até excluir o cluster.

No entanto, você pode querer manter o cluster de amostra em execução se pretende tentar tarefas em outros guias do Amazon Redshift.

Para excluir um cluster

1. Faça login no AWS Management Console e abra o console do Amazon Redshift em <https://console.aws.amazon.com/redshift/>.
2. No menu de navegação, escolha Clusters para exibir sua lista de clusters.
3. Escolha o cluster **examplecluster**. Em Actions, escolha Delete. A página Delete cluster (Excluir cluster) é exibida.
4. Confirme o cluster a ser excluído e escolha Delete cluster (Excluir cluster).

Na página da lista de clusters, o status do cluster é atualizado uma vez que o cluster está excluído.

Depois de concluir este tutorial, você pode encontrar mais informações sobre o Amazon Redshift e as próximas etapas em [Recursos adicionais \(p. 39\)](#).

Conceito básico de tarefas comuns de banco de dados

A seguir, você pode encontrar uma descrição e uma demonstração de tarefas comuns para que você possa começar a usar os bancos de dados do Amazon Redshift.

Depois de se conectar ao banco de dados dev do cluster inicial você pode criar um novo banco de dados. Independentemente de você optar por usar o conjunto de dados de amostra ou trazer seus próprios dados para o Amazon Redshift ao criar um cluster, o Amazon Redshift cria o banco de dados dev.

Para os exemplos nesta seção, suponha o seguinte:

- Você se inscreveu no Amazon Redshift. Para obter mais informações, consulte [Pré-requisitos \(p. 1\)](#).
- Você criou um cluster do Amazon Redshift. Para obter mais informações, consulte [Conceitos básicos de clusters do Amazon Redshift e carregamento de dados \(p. 7\)](#).
- Você estabeleceu uma conexão com o cluster usando a ferramenta de cliente SQL, como o editor de consulta do console do Amazon Redshift. Para obter mais informações, consulte [Etapa 3: Conceder acesso aos editores de consultas e executar consultas \(p. 14\)](#).

Important

O cluster que você implantou para este exercício é executado em um ambiente ativo. Enquanto estiver em execução, ele acumulará cobranças em sua conta da AWS. Para obter informações sobre preço, consulte [a página de preço do Amazon Redshift](#).

Para evitar cobranças desnecessárias, exclua seu cluster quando terminar de usá-lo. O passo final do exercício explica como fazer isso.

Tarefa 1: Criar um banco de dados

Depois de verificar se o cluster está funcionando, você pode criar seu primeiro banco de dados. Este banco de dados é onde você realmente cria tabelas, carrega dados e executa consultas. Um único cluster pode hospedar vários bancos de dados. Por exemplo, você pode ter um banco de dados SALESDB e um banco de dados ORDERSDB no mesmo cluster.

Por exemplo, para criar um banco de dados chamado **SALESDB**, execute o seguinte comando em sua ferramenta de cliente SQL.

```
CREATE DATABASE SALESDB;
```

Para este exercício, aceite os padrões. Para obter informações sobre mais opções de comando, consulte [CREATE DATABASE](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Depois de criar o banco de dados SALESDB, você pode se conectar ao novo banco de dados a partir de seu cliente SQL. Use os mesmos parâmetros de conexão usados para a conexão atual, mas altere o nome do banco de dados para SALESDB.

Tarefa 2: Criar um usuário

Por padrão, apenas o usuário administrador que você criou ao iniciar o cluster tem acesso ao banco de dados inicial no cluster. Para conceder acesso a outros usuários, crie uma ou mais contas. As contas de usuário do banco de dados são globais em todos os bancos de dados em um cluster e não por bancos de dados individuais.

Use o comando `CREATE USER` para criar um novo usuário. Ao criar um novo usuário, especifique o nome do novo usuário e uma senha. Recomendamos que você especifique uma senha para o usuário. Deve ter de 8 a 64 caracteres e incluir pelo menos uma letra maiúscula, uma letra minúscula e um numeral.

Por exemplo, para criar um usuário nomeado **GUEST** com senha **ABCd4321**, execute o comando a seguir.

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

Para se conectar ao banco de dados `SALESDB` como usuário `GUEST`, use a mesma senha quando você criou o usuário, como `ABCd4321`.

Para obter informações sobre mais opções de comando, consulte [CREATE USER](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Tarefa 3: Criar um esquema

Depois de criar um novo banco de dados, você pode criar um novo esquema no banco de dados atual. Um esquema é um namespace que contém objetos nomeados de banco de dados, como tabelas, exibições e funções definidas pelo usuário (UDFs). Um banco de dados pode conter um ou vários esquemas, e cada esquema pertence a apenas um banco de dados. Dois esquemas podem ter objetos diferentes que compartilham o mesmo nome.

Você pode criar vários esquemas no mesmo banco de dados para organizar dados da maneira que você deseja ou agrupar seus dados funcionalmente. Por exemplo, você pode criar um esquema para armazenar todos os dados de preparação e outro esquema para armazenar todas as tabelas de relatórios. Você também pode criar esquemas diferentes para armazenar dados relevantes para diferentes grupos de negócios que estão no mesmo banco de dados. Cada esquema pode armazenar objetos diferentes de banco de dados, como tabelas, exibições e funções definidas pelo usuário (UDFs). Além disso, você pode criar esquemas com a cláusula `AUTHORIZATION`. Esta cláusula fornece propriedade a um usuário especificado ou define uma cota na quantidade máxima de espaço em disco que o esquema especificado pode usar.

O Amazon Redshift cria automaticamente um esquema chamado `public` para cada novo banco de dados. Quando você não especifica o nome do esquema ao criar objetos de banco de dados, os objetos entram no esquema `public`.

Para acessar um objeto em um esquema, qualifique o objeto usando a notação `schema_name.table_name`. O nome qualificado do esquema consiste no nome do esquema e no nome da tabela separados por um ponto. Por exemplo, você pode ter um esquema `sales` que tem uma tabela `price` e um esquema `inventory` que também tem uma tabela `price`. Quando você se refere à tabela `price`, você deve qualificá-la como `sales.price` ou `inventory.price`.

O exemplo a seguir cria um esquema chamado **SALES** para o usuário `GUEST`.

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

Para obter informações sobre mais opções de comando, consulte [CREATE SCHEMA](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Para visualizar a lista de esquemas em seu banco de dados, execute o seguinte comando.

```
select * from pg_namespace;
```

A saída deve ser semelhante à seguinte.

```
nspname          | nspowner | nspacl
```


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

Para obter mais informações sobre como consultar tabelas de catálogo, consulte [Consultar tabelas de catálogo](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Use a instrução GRANT para dar permissões aos usuários para os esquemas.

O exemplo a seguir concede privilégios ao usuário GUEST para selecionar dados de todas as tabelas ou exibições no SALESHEMA usando uma instrução SELECT.

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

O exemplo a seguir concede todos os privilégios disponíveis de uma vez ao usuário GUEST.

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

Tarefa 4: Criar uma tabela

Depois de criar seu novo banco de dados, crie tabelas para armazenar seus dados. Especifique todas as informações da coluna ao criar a tabela.

No exemplo a seguir, o usuário GUEST faz login no SALESDB e cria uma nova tabela.

Por exemplo, para criar uma tabela chamada **DEMO**, execute o comando a seguir.

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

Você também pode criar uma tabela usando a notação `schema_name.object_name` para criar a tabela no esquema SALES.

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

Para visualizar e inspecionar esquemas e suas tabelas, você pode usar o editor de consulta do Amazon Redshift. Ou você pode ver a lista de tabelas em esquemas usando visualizações do sistema. Para obter mais informações, consulte [Etapa 6: Consultar as tabelas do sistema \(p. 23\)](#).

Por padrão, objetos novos de banco de dados, como tabelas, são criados no esquema padrão chamado `public` criado durante a criação do cluster. Você pode usar outro esquema para criar objetos de banco de dados. Para obter mais informações sobre esquemas, consulte [Gerenciando a segurança do banco de dados](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

As colunas `encoding`, `distkey` e `sortkey` são usadas pelo Amazon Redshift para processamento paralelo. Para obter mais informações sobre como projetar tabelas que incorporam esses elementos, consulte [Práticas recomendadas do Amazon Redshift para projetar tabelas](#).

Inserir linhas de dados em uma tabela

Depois de criar uma tabela, insira linhas de dados nessa tabela.

Note

O comando `INSERT` insere linhas em uma tabela. Para carregamentos em massa padrão, use o comando `COPY`. Para obter mais informações, consulte [Usar um comando COPY para carregar dados](#).

Por exemplo, para inserir valores na tabela `DEMO`, execute o comando a seguir.

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

Selecionar dados de uma tabela

Após criar uma tabela e preenchê-la com dados, use a instrução `SELECT` para exibir os dados contidos na tabela. A instrução `SELECT *` retorna todos os nomes de colunas e valores de linha de todos os dados de uma tabela. O uso do `SELECT` é uma boa forma de verificar se os dados adicionados recentemente foram inseridos corretamente na tabela.

Para visualizar os dados inseridos na tabela `DEMO`, execute o comando a seguir.

```
SELECT * from DEMO;
```

O resultado deve ser algo semelhante ao exibido a seguir.

```
personid |  city
-----+-----
         781 | San Jose
         990 | Palo Alto
(2 rows)
```

Para obter mais informações sobre como usar a instrução `SELECT` para consultar tabelas, consulte [SELECT](#).

Tarefa 5: Carregar dados de amostra

A maioria dos exemplos neste guia usa um banco de dados de amostra chamado TICKIT. Baixe o arquivo [tickitdb.zip](#) que contém os arquivos de dados de amostra individuais.

Você pode carregar os dados de exemplo em seu próprio bucket do Amazon S3.

Para carregar os dados de exemplo do banco de dados, primeiro crie as tabelas. Depois, use o comando `COPY` para carregar as tabelas com os dados de exemplo que estão armazenados em um bucket do Amazon S3. Para ver as etapas de criação de tabelas e carregamento de dados de amostra, consulte [Etapa 4: Carregar dados do Amazon S3 para o Amazon Redshift \(p. 15\)](#).

Etapa 6: Consultar as tabelas do sistema

Além das tabelas que você criou, seu banco de dados contém uma série de tabelas do sistema. Essas tabelas do sistema contêm informações sobre a sua instalação e sobre várias consultas e processos que são executados no sistema. É possível consultar essas tabelas do sistema para coletar informações sobre o banco de dados.

Note

A descrição de cada tabela nesta documentação indica se uma tabela é visível para todos os usuários ou apenas para superusuários. Efetue login como superusuário para consultar tabelas que são visíveis apenas para superusuários.

O Amazon Redshift fornece acesso aos seguintes tipos de tabelas do sistema:

- [Tabelas STL](#)

Essas tabelas de sistema são geradas a partir de arquivos de log do Amazon Redshift para fornecer um histórico do sistema. As tabelas de log têm um prefixo STL.

- [Tabelas STV](#)

Essas tabelas são tabelas virtuais do sistema que contêm snapshots dos dados atuais do sistema. As tabelas de snapshot têm um prefixo STV.

- [Visualizações do sistema](#)

As exibições do sistema contêm um subconjunto de dados encontrados em várias tabelas STL e STV do . As exibições do sistema têm um prefixo SVV ou SVL.

- [Tabelas de catálogo do sistema](#)

As tabelas de catálogo do sistema armazenam os metadados do esquema, como informações sobre as tabelas e as colunas. As tabelas de catálogo do sistema têm o prefixo PG.

Para recuperar informações da tabela do sistema sobre uma consulta, pode ser necessário especificar o ID do processo associado a essa consulta. Para obter mais informações, consulte [Determinar o ID do processo de uma consulta em execução \(p. 26\)](#).

Visualização de uma lista de nomes de tabelas

Para ver uma lista de todas as tabelas em um esquema, você pode consultar a tabela de catálogo do sistema PG_TABLE_DEF. Você pode primeiro examinar a configuração para search_path.

```
SHOW search_path;
```

O resultado deve ser semelhante ao seguinte:

```
search_path
-----
$user, public
(1 row)
```

O exemplo a seguir inclui o esquema SALES para o caminho de pesquisa e mostra todas as tabelas do esquema SALES.

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

SHOW search_path;
      search_path
-----
"$user", public, sales
(1 row)

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

```

schemaname | tablename | column |          type          | encoding | distkey | sortkey
| notnull
-----+-----+-----+-----+-----+-----+-----
sales      | demo     | personid | integer                | az64     | f       | 0
| f
sales      | demo     | city     | character varying(255) | lzo      | f       | 0
| f
(2 rows)
  
```

O exemplo a seguir mostra uma lista de todas as tabelas chamadas DEMO em todos os esquemas do banco de dados atual.

```

select * from pg_table_def where tablename = 'demo';
schemaname | tablename | column |          type          | encoding | distkey | sortkey
| notnull
-----+-----+-----+-----+-----+-----+-----
public     | demo     | personid | integer                | az64     | f       | 0
| f
public     | demo     | city     | character varying(255) | lzo      | f       | 0
| f
sales      | demo     | personid | integer                | az64     | f       | 0
| f
sales      | demo     | city     | character varying(255) | lzo      | f       | 0
| f
(4 rows)
  
```

Para obter mais informações, consulte [PG_TABLE_DEF](#).

Você também pode usar o editor de consultas v2 para visualizar todas as tabelas em um esquema especificado escolhendo primeiro um banco de dados ao qual deseja se conectar.

Visualização dos usuários

Você pode consultar o catálogo PG_USER para ver uma lista de todos os usuários, junto com o ID do usuário (USESYSID) e os privilégios do usuário.

```

SELECT * FROM pg_user;
username   | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
useconfig
-----+-----+-----+-----+-----+-----+-----
rdsdb      | 1        | true        | true     | true      | ***** | infinity |
awsuser    | 100      | true        | true     | false     | ***** |          |
guest      | 104      | true        | false    | false     | ***** |          |
(3 rows)
  
```

O nome de usuário rdsdb é usado internamente pelo Amazon Redshift para realizar tarefas administrativas e de manutenção de rotina. Você pode filtrar sua consulta para mostrar apenas nomes de usuário definidos pelo usuário, adicionando `where usesysid > 1` à sua instrução SELECT.

```

SELECT * FROM pg_user WHERE usesysid > 1;
username   | usesysid | usecreatedb | usesuper | usecatupd | passwd | valuntil |
useconfig
-----+-----+-----+-----+-----+-----+-----
awsuser    | 100      | true        | true     | false     | ***** |          |
guest      | 104      | true        | false    | false     | ***** |          |
(2 rows)
  
```

Visualização de consultas recentes

No exemplo anterior, o ID do usuário (USESYSID) para `adminuser` é 100. Para listar as cinco consultas mais recentes executadas por `adminuser`, você pode consultar a exibição `SVL_QLOG`.

A exibição `SVL_QLOG` é um subconjunto de informações, mais fácil de ler, da tabela `STL_QUERY`. Você pode usar essa visualização para localizar o ID da consulta (`QUERY`) ou o ID do processo (`PID`) para uma consulta executada recentemente. Você também pode usar esta visualização para verificar quanto tempo levou para uma consulta ser concluída. A exibição `SVL_QLOG` inclui os primeiros 60 caracteres da string da consulta (`SUBSTRING`) para ajudá-lo a encontrar uma consulta específica. Use a cláusula `LIMIT` com a instrução `SELECT` para limitar os resultados a cinco linhas.

```
SELECT query, pid, elapsed, substring from svl_qlog
WHERE userid = 100
ORDER BY starttime desc
LIMIT 4;
```

O resultado será semelhante ao seguinte:

query	pid	elapsed	substring
892	21046	55868	SELECT query, pid, elapsed, substring from svl_qlog WHERE us
620	17635	1296265	SELECT query, pid, elapsed, substring from svl_qlog WHERE us
610	17607	82555	SELECT * from DEMO;
596	16762	226372	INSERT INTO DEMO VALUES (100);)

Determinar o ID do processo de uma consulta em execução

Pode ser necessário encontrar o PID de uma consulta que ainda esteja em execução. Por exemplo, você precisará do PID se precisar cancelar uma consulta que está demorando muito para ser executada. Você pode consultar a tabela do sistema `STV_RECENTS` para obter uma lista de IDs de processo das consultas em execução, juntamente com a string da consulta correspondente. Se sua consulta retornar vários PIDs, examine o texto de consulta para determinar o PID de que precisa.

Para determinar o PID de uma consulta em execução, execute a instrução `SELECT` a seguir.

```
SELECT pid, user_name, starttime, query
FROM stv_recents
WHERE status='Running';
```

Tarefa 7: Cancelar uma consulta

Se você executar uma consulta que está demorando muito ou consumindo recursos do cluster excessivos, cancele a consulta. Por exemplo, crie uma lista de vendedores de ingressos que inclua o nome do vendedor e a quantidade de ingressos vendidos. A consulta a seguir seleciona dados da tabela `SALES` e da tabela `USERS` e une as duas tabelas, correspondendo `SELLERID` e `USERID` na cláusula `WHERE`.

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

O resultado é parecido com o mostrado a seguir.

```
sellerid | firstname | lastname | sum
```

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

Note

Essa é uma consulta complexa. Para este tutorial, não se preocupe em saber como a consulta foi criada.

A consulta anterior é executada em segundos e retornará 2.102 linhas.

Suponha que você se esqueça de inserir a cláusula WHERE.

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

O conjunto de resultados inclui todas as linhas da tabela SALES multiplicadas por todas as linhas da tabela USERS (49989*3766). Isso é chamado de junção cartesiana, e não é uma prática recomendada. O resultado é mais de 188 milhões linhas e o tempo de execução é bastante longo.

Para cancelar uma consulta em execução, use o comando CANCEL com o PID da consulta.

Para localizar o ID do processo, inicie uma nova sessão e consulte a tabela STV_RECENTS, conforme mostrado na etapa anterior. O exemplo a seguir mostra como você pode fazer os resultados mais legíveis. Para fazer isso, use a função TRIM para cortar os espaços à direita e mostrar somente os 20 primeiros caracteres da consulta.

```
SELECT pid, trim(user_name), starttime, substring(query,1,20)  
FROM stv_recents  
WHERE status='Running';
```

O resultado é parecido com o mostrado a seguir.

```
pid | btrim | starttime | substring  
-----+-----+-----+-----  
610 | adminuser | 2013-03-28 18:39:49.355918 | select sellerid, fir  
(1 row)
```

Para cancelar a consulta com PID 18764, execute o seguinte comando.

```
CANCEL 610;
```

Note

O comando CANCEL não interrompe uma transação. Para parar ou reverter uma transação, use o comando ABORT ou ROLLBACK. Para cancelar uma consulta associada a uma transação, primeiro cancele a consulta e pare a transação.

Se a consulta que você cancelou estiver associada a uma transação, use o comando ABORT ou ROLLBACK para cancelar a transação e descartar quaisquer alterações feitas nos dados:

```
ABORT;
```

A menos que esteja conectado como um superusuário, você só pode cancelar suas próprias consultas. Os superusuários podem cancelar todas as consultas.

Cancelar uma consulta de outra sessão

Se sua ferramenta de consulta não oferece suporte à execução de consultas simultaneamente, inicie outra sessão para cancelar a consulta. Por exemplo, o editor de consulta que usamos no Guia de conceitos básicos do Amazon Redshift não é compatível com várias consultas simultâneas. Para iniciar outra sessão com o editor de consulta, escolha Arquivo, Nova janela e conecte-se usando os mesmos parâmetros de conexão. Em seguida, você pode encontrar o PID e cancelar a consulta.

Cancelar uma consulta usando a fila de superusuários

Se a sua sessão atual tiver muitas consultas sendo executadas simultaneamente, talvez não seja possível executar o comando CANCEL até que uma outra consulta seja concluída. Nesse caso, execute o comando CANCEL usando uma fila de consulta de gerenciamento de workload diferente.

Usando o gerenciamento de workload, você pode executar consultas em diferentes filas de consulta para que não precise esperar a conclusão de outra consulta. O gerenciador de workload cria uma fila separada, chamada a fila de superusuários, que você pode usar quando estiver solucionando problemas. Para usar a fila Superusuário, faça logon em um superusuário e defina o grupo de consulta como “superusuário” usando o comando SET. Depois de executar os comandos, redefina o grupo de consultas usando o comando RESET.

Para cancelar uma consulta usando a fila Superusuário, execute estes comandos.

```
SET query_group TO 'superuser';  
CANCEL 610;  
RESET query_group;
```

Etapa 8: Limpar os recursos

Se você implantou um cluster para concluir este exercício, quando terminar o exercício, exclua o cluster. A exclusão do cluster interrompe o acúmulo de cobranças para a sua conta da AWS.

Para excluir o cluster, siga as etapas em [“Excluir um cluster”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

Se você deseja manter o cluster, mantenha os dados de amostra para referência. A maioria dos exemplos neste guia usa as tabelas que você criar neste exercício. O tamanho dos dados não terá nenhum efeito significativo no armazenamento disponível.

Se você deseja manter o cluster, mas deseja limpar os dados de amostra, execute o seguinte comando para eliminar o banco de dados SALESDB.

```
DROP DATABASE SALESDB;
```

Se você não criou um banco de dados SALESDB ou se não deseja eliminá-lo, execute os seguintes comandos para eliminar apenas as tabelas.

```
DROP TABLE DEMO;
```

```
DROP TABLE users;  
DROP TABLE venue;  
DROP TABLE category;  
DROP TABLE date;  
DROP TABLE event;  
DROP TABLE listing;  
DROP TABLE sales;
```


Conceitos básicos do Amazon Redshift Serverless

Se estiver usando o Amazon Redshift Serverless pela primeira vez, recomendamos que leia as seguintes seções que ajudarão a começar a usar o Amazon Redshift Serverless.

Tópicos

- [Conceitos básicos do console do Amazon Redshift Serverless \(p. 30\)](#)
- [Conectar-se ao Amazon Redshift Serverless \(p. 30\)](#)
- [Conceitos básicos do Amazon Redshift Serverless e do compartilhamento de dados \(p. 31\)](#)

Conceitos básicos do console do Amazon Redshift Serverless

No console do Amazon Redshift Serverless, é possível gerenciar todos os seus recursos com tecnologia sem servidor, como grupos de trabalho, namespaces, backups de dados e alarmes. Na primeira vez que fizer login no console do Amazon Redshift Serverless, você será solicitado a acessar a experiência de conceitos básicos, que pode ser usada para configurar o Amazon Redshift Serverless.

Conectar-se ao Amazon Redshift Serverless

Para se conectar ao Amazon Redshift Serverless, use um dos seguintes métodos:

- Com o cliente SQL de sua preferência, use o driver JDBC versão 2 fornecido pelo Amazon Redshift. Para obter mais informações, consulte [Conectar-se ao Amazon Redshift Serverless por meio de drivers JDBC](#).
- Para usar a API de dados do Amazon Redshift para se conectar ao Amazon Redshift Serverless, omita o parâmetro `cluster-identifier` e inclua `workgroup-name` em chamadas AWS CLI para encaminhar as chamadas para o Amazon Redshift Serverless. Para obter mais informações sobre a API de dados, consulte [Usar a API de dados do Amazon Redshift](#).
- Para configurar uma conexão Secure Sockets Layer (SSL) para criptografar consultas e dados, utilize a mesma configuração usada para configurar uma conexão com um cluster do Redshift provisionado. Para obter mais informações, consulte [Configurar as opções de segurança para conexões](#).
- Também é possível conectar o Amazon Redshift Serverless por um endpoint da VPC gerenciado pelo Amazon Redshift. Para obter mais informações, consulte [Conexão com o Amazon Redshift Serverless por um endpoint da VPC gerenciado pelo Amazon Redshift](#).
- É possível criar um Amazon Redshift Serverless acessível ao público para consultá-lo de um cliente SQL na Internet pública. Para obter mais informações, consulte, [Criar uma instância do Amazon Redshift Serverless acessível ao público e conectar-se a ela](#).
- Quando você invoca o editor de consultas v2 no console do Amazon Redshift Serverless, uma nova guia do navegador é aberta com o editor de consultas. O editor de consultas v2 se conecta da máquina

cliente ao ambiente do Amazon Redshift Serverless. Para mais informações, consulte [Etapa 2: Consultar dados de exemplo no editor de consultas v2 do Amazon Redshift](#) (p. 33).

Conceitos básicos do Amazon Redshift Serverless e do compartilhamento de dados

Para configurar e usar o Amazon Redshift Serverless, configure o data warehouse sem servidor e crie um banco de dados. Para fazer isso, você precisa de permissões do IAM, conforme descrito em “[Usar políticas baseadas em identidade \(políticas do IAM\) para o Amazon Redshift](#)” no Guia de gerenciamento de clusters do Amazon Redshift. Além disso, você deve anexar uma política semelhante à política a seguir ao perfil do IAM ou ao usuário do IAM.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "redshift-serverless:*",
      "Resource": "*"
    }
  ]
}
```

Para começar, abra o AWS Management Console, escolha o console do Amazon Redshift e selecione Try Amazon Redshift Serverless (Experimentar o Amazon Redshift Serverless).

Se tiver as permissões do AWS Identity and Access Management (IAM), na primeira vez em que acessar o console do Amazon Redshift Serverless você visualizará a página Get started with Amazon Redshift Serverless (Conceitos básicos do Amazon Redshift Serverless).

Aqui é onde você pode optar por Use default settings (Usar configurações padrão) ou Customize settings (Personalizar configurações) para criar o namespace, o banco de dados e o grupo de trabalho.

O Amazon Redshift Serverless inicializa os recursos para sua Conta da AWS na Região da AWS atual. O processo de inicialização pode levar alguns minutos para configurar o ambiente. O editor de consultas v2 do Amazon Redshift é aberto em uma nova guia para você começar a usar o Amazon Redshift Serverless.

Permissões necessárias para usar o Amazon Redshift Serverless

Ao usar o Amazon Redshift Serverless, o perfil do IAM que você associa ao Amazon Redshift Serverless precisa de uma relação de confiança com `redshift.amazonaws.com` e `redshift-serverless.amazonaws.com` para permitir que o Amazon Redshift assuma permissões em seu nome. O exemplo a seguir mostra o documento de política no formato JSON para configurar uma relação de confiança com o Amazon Redshift Serverless.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": [
          "redshift-serverless.amazonaws.com",
          "redshift.amazonaws.com"
        ]
      }
    }
  ],
}
```

```
    "Action": "sts:AssumeRole"  
  }  
]  
}
```

Para obter mais informações sobre entidades confiáveis, consulte [Criar uma função para delegar permissões a um serviço da AWS](#) no Manual do usuário do IAM.

Na primeira vez que fizer login no console do Amazon Redshift Serverless, você será solicitado a acessar a experiência de conceitos básicos, que pode ser usada para configurar o Amazon Redshift Serverless. É possível usar configurações padrão ou personalizar configurações, como credenciais, segurança e log de auditoria. Parte da experiência de conceitos básicos é criar um grupo de trabalho e um namespace, que são coleções de recursos de computação e objetos de banco de dados, respectivamente. Para obter mais informações sobre grupos de trabalho e namespaces do Amazon Redshift Serverless, consulte [Visão geral de grupos de trabalho e namespaces do Amazon Redshift Serverless](#).

Tópicos

- [Usar um conjunto de dados de amostra \(p. 32\)](#)
- [Trazer seus próprios dados para o Amazon Redshift Serverless \(p. 34\)](#)

Usar um conjunto de dados de amostra

Neste tutorial, você passará pelo processo para criar um Amazon Redshift Serverless usando um conjunto de dados de exemplo. O Amazon Redshift Serverless carrega automaticamente o conjunto de dados de exemplo, como o conjunto de dados tickit, quando você está criando um Amazon Redshift Serverless. Você pode consultar imediatamente os dados.

Tópicos

- [Etapa 1: Configurar o Amazon Redshift Serverless pela primeira vez \(p. 32\)](#)
- [Etapa 2: Consultar dados de exemplo no editor de consultas v2 do Amazon Redshift \(p. 33\)](#)

Etapa 1: Configurar o Amazon Redshift Serverless pela primeira vez

A primeira vez que seleciona o Serverless dashboard (Painel do Serverless), você executa as etapas para configurar o Amazon Redshift Serverless. Em Get started with the serverless experience (Conceitos básicos da experiência com a tecnologia sem servidor), você pode selecionar Use default settings (Usar configurações padrão). Depois, o Amazon Redshift Serverless cria um namespace e um grupo de trabalho padrão. Essa configuração usa as configurações padrão e torna-se ativa quando você associa o grupo de trabalho padrão ao namespace padrão. Nesta seção, são usados as configurações, o grupo de trabalho e o namespace padrão.

Para ter um controle mais detalhado da configuração, selecione Customize settings (Personalizar configurações).

Como utilizar as configurações padrão:

1. Em Configuration (Configuração), selecione Use default settings (Usar configurações padrão). O Amazon Redshift Serverless cria um namespace padrão com um grupo de trabalho padrão associado a esse namespace.

As configurações a seguir em Namespace são definidas como valores padrão.

- Nome de usuário e senha do banco de dados

- Credenciais de usuário administrador
- Default IAM role (Perfil do IAM padrão)
- IAM roles (Perfis do IAM): se você não criou nenhum perfil do IAM, escolha Associate IAM role (Associar perfil do IAM) para criar e associar um ou mais perfis do IAM ao namespace. O perfil do IAM que você associa ao namespace deve incluir uma relação de confiança com `redshift-serverless.amazonaws.com` e `redshift.amazonaws.com`.

O Workgroup (Grupo de trabalho) é definido como default (padrão). Além disso, as configurações de rede são definidas como a default VPC (VPC padrão), a sub-rede default (padrão) e o grupo de segurança do cluster default (padrão).

2. Escolha Salvar configurações.
3. Após a conclusão da configuração, escolha Continue (Continuar) para acessar o Serverless dashboard (Painel do Serverless).

Como usar configurações personalizadas:

1. Em Configuration (Configuração), selecione Customize settings (Personalizar configurações). Configure as definições a seguir.
 - Database name (Nome do banco de dados): o nome do banco de dados inicial (padrão) a ser criado no ambiente do Amazon Redshift Serverless. Esse banco de dados é de propriedade de sua conta e foi criado na Região da AWS atual. O nome é dev e não é possível alterá-lo.
 - Admin user credentials (Credenciais de usuário administrador): o nome de usuário e a senha do administrador do banco de dados inicial. Esse usuário tem permissões de propriedade para o banco de dados.
 - Virtual private cloud (VPC): o nome da VPC em que o banco de dados é criado.
 - VPC security groups (grupos de segurança): esses grupos de segurança definem quais sub-redes e intervalos de IP podem ser usados na VPC.
 - Subnet (Sub-rede): as sub-redes na VPC que estão associadas ao banco de dados especificado.
 - A chave do KMS de propriedade da AWS é usada por padrão para criptografar os dados. Em vez de usar a chave KMS de propriedade da AWS, você pode Customize encryption settings (Personalizar configurações de criptografia) para selecionar uma KMS key (Chave do KMS) que você gerencia. A AWS KMS key é usada para criptografar recursos no Amazon Redshift Serverless.
 - Audit logging (Log de auditoria): os tipos de log de auditoria que você deseja exportar.
 - Permissions (Permissões): o perfil do IAM que você associa ao Amazon Redshift Serverless deve incluir uma relação de confiança com `redshift-serverless.amazonaws.com` e `redshift.amazonaws.com`.

O Workgroup (Grupo de trabalho) é definido como default (padrão). Além disso, as configurações de rede são definidas como a default VPC (VPC padrão), a sub-rede default (padrão) e o grupo de segurança do cluster default (padrão).

2. Escolha Salvar configurações.
3. Após a conclusão da configuração, escolha Continue (Continuar) para acessar o Serverless dashboard (Painel do Serverless).

Etapa 2: Consultar dados de exemplo no editor de consultas v2 do Amazon Redshift

É possível gerenciar e consultar dados no editor de consultas v2. O editor de consultas v2 é uma ferramenta completa de cliente SQL baseada na Web para se conectar aos dados do Amazon Redshift

Serverless. Para configurar o uso do editor de consultas v2 do Amazon Redshift, inclusive quais permissões são necessárias, consulte [“Configurar sua Conta da AWS”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

Procure o botão Query data (Consultar dados) para consultar dados no Amazon Redshift Serverless com o editor de consultas v2. Quando você invoca o editor de consultas v2 no console do Amazon Redshift, abre-se uma nova guia no navegador com o editor de consultas. O editor de consultas v2 se conecta da máquina cliente ao ambiente do Amazon Redshift Serverless.

1. Na lista de recursos, selecione o grupo de trabalho Serverless default (Padrão sem servidor). O editor de consultas v2 se conecta automaticamente ao Amazon Redshift Serverless usando credenciais temporárias.
2. No grupo de trabalho padrão do Amazon Redshift Serverless, expanda o banco de dados `sample_data_dev`. Há três esquemas de exemplo correspondentes a três conjuntos de dados de exemplo que você pode carregar no banco de dados do Amazon Redshift.
3. Neste tutorial, selecione o esquema `tickit` para iniciar a criação do banco de dados `sample_data_dev`, do esquema `tickit` e das tabelas de exemplo no esquema `tickit`. O Amazon Redshift Serverless também carrega os dados de vendas do `tickit` nas tabelas de exemplo.
4. Você pode executar as consultas de exemplo com base nos dados de exemplo carregados.

Para obter mais informações sobre o editor de consultas v2, acesse [“Consultar um banco de dados usando o editor de consultas v2 do Amazon Redshift”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

Trazer seus próprios dados para o Amazon Redshift Serverless

Neste tutorial, você passará pelo processo de criação de um Amazon Redshift Serverless e carregará seus próprios dados.

Tópicos

- [Etapa 1: Configurar o Amazon Redshift Serverless pela primeira vez \(p. 34\)](#)
- [Etapa 2: Consultar seus próprios dados no editor de consultas v2 do Amazon Redshift \(p. 34\)](#)

Etapa 1: Configurar o Amazon Redshift Serverless pela primeira vez

Para configurar o Amazon Redshift Serverless pela primeira vez, siga as etapas em [Etapa 1: Configurar o Amazon Redshift Serverless pela primeira vez \(p. 32\)](#).

Etapa 2: Consultar seus próprios dados no editor de consultas v2 do Amazon Redshift

É possível consultar dados de diferentes fontes de dados no Amazon Redshift Serverless. Veja a seguir uma lista de fontes de dados nas quais você pode consultar dados:

- Para consultar dados no Amazon S3, você pode carregar os dados em uma tabela Amazon Redshift existente do Amazon S3 usando o editor de consultas v2. Para obter mais informações, consulte [“Carregar dados do Amazon S3”](#) no Guia de gerenciamento de clusters do Amazon Redshift.
- Para consultar dados em catálogos do AWS Glue que representam bancos de dados em seus data lakes baseados no Amazon S3, você pode criar um esquema externo para consultar seu data lake sem carregar dados no Amazon Redshift Serverless. Para obter mais informações, confira [“Consultar um data lake”](#) no Guia de gerenciamento de clusters do Amazon Redshift.

- Para consultar dados em seus armazenamentos de dados usando consultas federadas, você pode criar um esquema externo usando um banco de dados PostgreSQL ou MySQL do Amazon RDS.

Para obter mais informações sobre como usar consultas federadas no PostgreSQL, consulte [“Conceitos básicos do uso de consultas federadas no PostgreSQL”](#).

Para obter mais informações sobre como começar a usar consultas federadas no MySQL, consulte [“Conceitos básicos do uso de consultas federadas no MySQL”](#).

Para obter mais informações sobre como criar um esquema externo no editor de consultas v2, consulte [Criar o esquema externo](#).

Conceitos básicos da consulta de fontes de dados fora do banco de dados do Amazon Redshift

A seguir, você encontrará informações sobre como começar a consultar dados em fontes remotas, inclusive clusters remotos do Amazon Redshift. Também encontrará informações sobre como treinar modelos de machine learning (ML) usando o Amazon Redshift.

Tópicos

- [Conceitos básicos sobre como consultar seu data lake \(p. 36\)](#)
- [Conceitos básicos sobre consulta de dados em fontes de dados remotas \(p. 37\)](#)
- [Conceitos básicos sobre acesso a dados em outros clusters do Amazon Redshift \(p. 37\)](#)
- [Conceitos básicos sobre treinamento de modelos de Machine Learning com dados do Amazon Redshift \(p. 37\)](#)

Conceitos básicos sobre como consultar seu data lake

É possível usar o Amazon Redshift Spectrum para consultar dados em arquivos do Amazon S3 sem ter que carregar os dados nas tabelas do Amazon Redshift. Você pode consultar dados em vários formatos, incluindo Parquet, ORC, RCFile, TextFile, SequenceFile, RegexSerde, OpenCSV, e AVRO. Para definir a estrutura dos arquivos no Amazon S3, crie esquemas e tabelas externos. Em seguida, você usa um catálogo de dados externo, como AWS Glue ou seu próprio metastore do Apache Hive. As alterações feitas em qualquer tipo de catálogo de dados são disponibilizadas instantaneamente para qualquer um de seus clusters do Amazon Redshift.

Depois que seus dados são registrados com um catálogo de dados do AWS Glue e habilitado com AWS Lake Formation, você pode consultá-la usando o Redshift Spectrum.

O Redshift Spectrum reside em servidores dedicados do Amazon Redshift que são independentes do seu cluster. O Redshift Spectrum envia várias tarefas de computação intensiva para a camada do Redshift Spectrum, como a filtragem e a agregação de predicados. O Redshift Spectrum também é escalado de forma inteligente para aproveitar o processamento massivamente paralelo.

Você pode dividir as tabelas externas em partições de uma ou mais colunas para otimizar a performance da consulta por meio da eliminação de partições. Você pode consultar e unir tabelas externas com tabelas do Amazon Redshift. Você pode acessar tabelas externas de vários clusters do Amazon Redshift e consultar os dados do Amazon S3 de qualquer cluster na mesma região da AWS. Quando você atualiza os arquivos de dados do Amazon S3, os dados são disponibilizados instantaneamente para consulta a partir de qualquer um dos clusters do Amazon Redshift.

Para obter mais informações sobre o Redshift Spectrum, incluindo como trabalhar com o Redshift Spectrum e data lakes, consulte [Conceitos básicos do Amazon Redshift Spectrum](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Conceitos básicos sobre consulta de dados em fontes de dados remotas

É possível juntar dados de um banco de dados do Amazon RDS, de um banco de dados do Amazon Aurora ou do Amazon S3 a dados em seu banco de dados do Amazon Redshift usando uma consulta federada. Você pode usar o Amazon Redshift para consultar dados operacionais diretamente (sem movê-lo), aplicar transformações e inserir dados em suas tabelas do Redshift. Alguns dos cálculos para consultas federadas são distribuídos para as fontes de dados remotas.

Para executar consultas federadas, o Amazon Redshift primeiro faz uma conexão com a origem dos dados remota. Em seguida, o Amazon Redshift recupera metadados sobre as tabelas na origem dos dados remota, emite consultas e recupera as linhas de resultado. O Amazon Redshift então distribui as linhas de resultados para os nós de computação do Amazon Redshift para processamento adicional.

Para obter informações sobre como configurar seu ambiente para consultas federadas, consulte um dos seguintes tópicos no Guia do desenvolvedor do banco de dados do Amazon Redshift:

- [Conceitos básicos do uso de consultas federadas no PostgreSQL](#)
- [Conceitos básicos do uso de consultas federadas no MySQL](#)

Conceitos básicos sobre acesso a dados em outros clusters do Amazon Redshift

Com o compartilhamento de dados do Amazon Redshift, é possível compartilhar dados dinâmicos com alta segurança e mais facilidade entre clusters do Amazon Redshift ou contas da AWS para fins de leitura. Você pode ter acesso instantâneo, detalhado e de alta performance aos dados em clusters do Amazon Redshift sem precisar copiá-los ou movê-los manualmente. Os usuários podem ver as informações mais atualizadas e consistentes conforme elas são atualizadas nos clusters do Amazon Redshift.

O compartilhamento de dados do Amazon Redshift é especialmente útil para esses casos de uso:

- Centralização de workloads essenciais aos negócios: use um cluster central de extração, transformação e carga (ETL) que compartilhe dados com vários clusters de Business Intelligence (BI) ou analíticos. Essa abordagem oferece isolamento de workload de leitura e estorno para workloads individuais.
- Compartilhamento de dados entre ambientes: compartilhe dados entre os ambientes de desenvolvimento, teste e produção. É possível melhorar a agilidade da equipe compartilhando dados em diferentes níveis de detalhes.

Para obter mais informações sobre o compartilhamento de dados, consulte [Conceitos básicos sobre compartilhamento de dados](#) no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Conceitos básicos sobre treinamento de modelos de Machine Learning com dados do Amazon Redshift

Com o Amazon Redshift Machine Learning (Amazon Redshift ML), é possível treinar um modelo ao fornecer os dados para o Amazon Redshift. Em seguida, o Amazon Redshift ML cria modelos que

capturam padrões nos dados de entrada. Você pode usar esses modelos para gerar previsões para novos dados de entrada sem incorrer em custos adicionais. Usando o Amazon Redshift ML, é possível treinar modelos de machine learning usando instruções SQL e chamá-los em consultas SQL para previsão. Você pode continuar a melhorar a precisão das previsões ao alterar os parâmetros de forma iterativa e melhorar seus dados de treinamento.

O Amazon Redshift ML facilita a criação, o treinamento e a implantação de modelos de machine learning para os usuários de SQL usando comandos SQL familiares. Com o Amazon Redshift ML, você pode usar dados em clusters do Amazon Redshift para treinar modelos com o Amazon SageMaker. É possível localizar os modelos, e as previsões podem ser feitas em um banco de dados do Amazon Redshift.

Para obter mais informações sobre o Amazon Redshift ML, consulte [Conceitos básicos do Amazon Redshift ML](#) no Guia do desenvolvedor do banco de dados do Amazon Redshift.

Recursos adicionais

Quando você concluir esses tutoriais, recomendamos que continue aprendendo sobre os conceitos introduzidos neste guia usando os seguintes recursos do Amazon Redshift:

- [Guia de gerenciamento de clusters do Amazon Redshift](#): esse guia baseia-se neste Guia de conceitos básicos do Amazon Redshift. Ele fornece informações detalhadas sobre os conceitos e as tarefas de criação, gerenciamento e monitoramento de clusters.
- [Guia do desenvolvedor de banco de dados do Amazon Redshift](#): este guia baseia-se neste Guia de conceitos básicos do Amazon Redshift. Ele fornece informações detalhadas aos desenvolvedores de bancos de dados sobre como projetar, criar, consultar e manter os bancos de dados que compõem seu data warehouse.
- [Referência SQL](#): este tópico descreve comandos SQL e referências de função para o Amazon Redshift.
- [Tabelas e visualizações de sistema](#): este tópico descreve tabelas de sistema e visualizações para o Amazon Redshift.
- Tutoriais para o Amazon Redshift: este tópico mostra tutoriais para saber mais sobre os recursos do Amazon Redshift.
 - [Usando funções SQL especiais com o Amazon Redshift](#): este tutorial demonstra como usar algumas das funções SQL especiais com o Amazon Redshift.
 - [Carregamento de dados do Amazon S3](#): este tutorial descreve como carregar dados nas tabelas de banco de dados do Amazon Redshift a partir de arquivos de dados em um bucket do Amazon S3.
 - [Consultar dados aninhados com o Amazon Redshift Spectrum](#): este tutorial descreve como usar o Redshift Spectrum para consultar dados aninhados em formatos de arquivo Parquet, ORC, JSON e Ion usando tabelas externas.
 - [Configuração de filas de gerenciamento de workload manual \(WLM\)](#): este tutorial descreve como configurar o WLM no Amazon Redshift.
- Vídeos de recursos: esses vídeos ajudam você a aprender sobre os recursos do Amazon Redshift.
 - Para saber como começar a usar o Amazon Redshift, assista ao seguinte vídeo: [Conceitos básicos do Amazon Redshift](#).
 - Para saber como funciona o compartilhamento de dados do Amazon Redshift, assista ao seguinte vídeo: [Fluxo de trabalho do compartilhamento de dados do Amazon Redshift](#).
 - Para saber mais sobre como o Amazon Redshift Machine Learning (ML) funciona, assista ao seguinte vídeo: [Amazon Redshift ML](#).
 - Para saber como monitorar, isolar e otimizar suas consultas usando os recursos de monitoramento de consultas no console do Amazon Redshift, assista ao seguinte vídeo: [Monitoramento de consultas no Amazon Redshift](#).
- [Novidades](#): esta página da Web lista os novos recursos e atualizações de produtos do Amazon Redshift.

Histórico do documento

A tabela a seguir descreve as alterações importantes do Guia de conceitos básicos do Amazon Redshift.

Última atualização da documentação: 30 de junho de 2021

Alteração	Descrição	Data de lançamento
Atualização da documentação	Guia atualizado para incluir novas seções sobre como começar com tarefas comuns de banco de dados, consultar seu data lake, consultar dados em fontes remotas, compartilhar dados e treinar modelos de Machine Learning com dados do Amazon Redshift.	30 de junho de 2021
Novo recurso	Guia atualizado para descrever o novo procedimento de carregamento para amostra.	04 de junho de 2021
Atualização da documentação	Guia atualizado para remover o console original do Amazon Redshift e melhorar o fluxo de etapas.	14 de agosto de 2020
New console	Guia atualizado para descrever o novo console do Amazon Redshift.	11 de novembro de 2019
Novo recurso	Atualização do guia para descrever o procedimento de cluster para início rápido.	10 de agosto de 2018
Novo recurso	Guia atualizado para iniciar clusters do painel do Amazon Redshift.	28 de julho de 2015
Novo recurso	Atualização do guia para usar os nomes do novo tipo de nó.	9 de junho de 2015
Atualização da documentação	Atualização das capturas de tela e do procedimento para configurar grupos de segurança de VPC.	30 de abril de 2015
Atualização da documentação	Atualização das capturas de tela e dos procedimentos de correspondência do console atual.	12 de novembro de 2014
Atualização da documentação	Os dados de carregamento foram movidos das informações do Amazon S3 para a sua própria seção, e a seção de próximas etapas foi movida para a etapa final com o intuito de melhorar a capacidade de descoberta.	13 de maio de 2014
Atualização da documentação	A página de boas-vindas foi removida, e o conteúdo foi incorporado à página principal de Conceitos básicos.	14 de março de 2014
Atualização da documentação	Esta é uma nova versão do Guia de conceitos básicos do Amazon Redshift que aborda o feedback do cliente e as atualizações de serviço.	14 de março de 2014
Novo guia	Esta é a primeira versão do Guia de conceitos básicos do Amazon Redshift.	14 de fevereiro de 2013