Guia de conceitos básicos

Amazon Redshift

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon Redshift: Guia de conceitos básicos

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens de marcas 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

..... v
Conceitos básicos de data warehouses de tecnologia sem servidor 1
Como se cadastrar na AWS 1
Criar um data warehouse com o Amazon Redshift sem servidor 1
Carregar dados de exemplo 4
Executar consultas de amostra 6
Carregar dados do Amazon S3 7
Conceitos básicos de data warehouses provisionados 14
Como se cadastrar na AWS 16
Determinar regras de firewall 17
Etapa 1: Criar um cluster de exemplo 17
Etapa 2: Configurar regras de entrada para clientes SQL 21
Etapa 3: Conceder acesso a um cliente SQL e executar consultas 21
Conceder acesso ao editore de consultas v2 22
Etapa 4: Carregar dados do Amazon S3 para o Amazon Redshift 23
Carregar dados do Amazon S3 usando comandos SQL 23
Carregar dados do Amazon S3 usando o Editor de Consultas v2 25
Criar dados TICKIT no cluster 26
Etapa 5: Experimentar exemplo de consultas usando o editor de consultas 27
Etapa 6: Redefinir o ambiente 28
Definir e usar um banco de dados no data warehouse 30
Conectar-se ao Amazon Redshift 31
Criar um banco de dados 32
Criar um usuário 33
Criar um esquema 33
Criar uma tabela 35
Inserir linhas de dados em uma tabela 36
Selecionar dados de uma tabela 36
Carregar dados 37
Consultar as visualizações e tabelas do sistema 37
Visualização de uma lista de nomes de tabelas 37
Visualização dos usuários 39
Visualização de consultas recentes 40
Determinar o ID da sessão de uma consulta em execução 40

Cancelar uma consulta 41
Cancelar uma consulta usando a fila de superusuários 43
Consultar dados que não estão em seu banco de dados do Amazon Redshift 45
Consultar data lakes 45
Consultar fontes de dados remotas 46
Acessar dados em outros bancos de dados 46
Treinamento de modelos de ML com dados do Redshift 47
Aprenda conceitos do Amazon Redshift 48
Recursos adicionais de aprendizado 52
Histórico do documento 54

O Amazon Redshift não permitirá mais a criação de UDFs do Python a partir do Patch 198. As UDFs do Python existentes continuarão a funcionar normalmente até 30 de junho de 2026. Para ter mais informações, consulte a publicação de blog .

Conceitos básicos de data warehouse do Amazon Redshift sem servidor

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. O fluxo básico do Amazon Redshift sem servidor é criar recursos de tecnologia sem servidor, conectar-se ao Amazon Redshift sem servidor, carregar dados de amostra e executar consultas aos dados. Neste guia, você pode escolher carregar dados de amostra do Amazon Redshift sem servidor ou de um bucket do Amazon S3. Os dados de amostra são usados em toda a documentação do Amazon Redshift para demonstrar os recursos. Para começar a usar data warehouses provisionados do Amazon Redshift, consulte Conceitos básicos de data warehouses provisionados do Amazon Redshift.

- the section called “Como se cadastrar na AWS”
- the section called “Criar um data warehouse com o Amazon Redshift sem servidor”
- the section called “Carregar dados do Amazon S3”

Como se cadastrar na AWS

Cadastre uma conta da AWS caso você ainda não tenha uma. 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 online.

   Quando você cria uma conta da AWS, um usuário raiz da conta da AWS é criado. O usuário-raiz tem acesso a todos os serviços e recursos da AWS na conta. Como prática recomendada de segurança, atribua acesso administrativo a um usuário administrativo e use o usuário raiz somente para realizar as tarefas que exigem acesso do usuário raiz.

Criar um data warehouse com o Amazon Redshift sem servidor

Na primeira vez que fizer login no console do Amazon Redshift sem servidor, você será solicitado a acessar a experiência de conceitos básicos, que pode ser usada para criar e gerenciar recursos de tecnologia sem servidor. Neste guia, você criará recursos de tecnologia sem servidor usando as configurações padrão do Amazon Redshift sem servidor.

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

> **Note**
>
> O Redshift sem servidor exige uma Amazon VPC com três sub-redes em três zonas de disponibilidade diferentes. O Redshift sem servidor também exige pelo menos três endereços IP disponíveis. Antes de continuar, garanta que a Amazon VPC utilizada para o Redshift sem servidor tenha três sub-redes em três zonas de disponibilidade diferentes e pelo menos três endereços IP disponíveis antes de continuar. Para ter mais informações sobre como criar sub-redes em uma Amazon VPC, consulte Criar uma sub-rede no Manual do usuário da Amazon Virtual Private Cloud. Para ter mais informações sobre endereços IP em uma Amazon VPC, consulte Endereçamento IP para suas VPCs e sub-redes.

Como utilizar as configurações padrão:

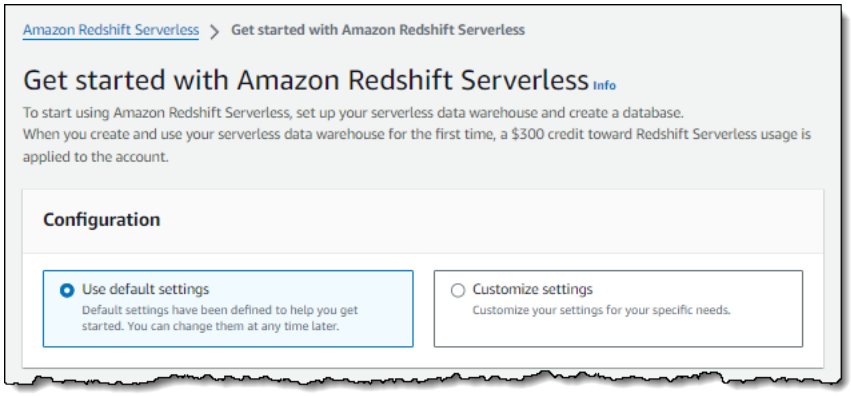
1. Faça login no Console de gerenciamento da AWS e abra o console do Amazon Redshift em https://console.aws.amazon.com/redshiftv2/.

   Selecione Experimente o Amazon Redshift sem servidor.

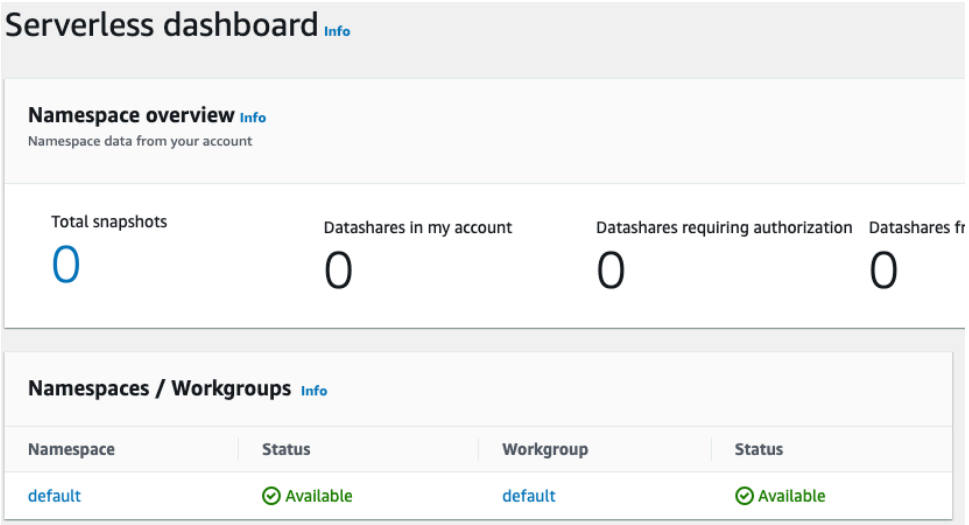
2. Em Configuration (Configuração), selecione Use default settings (Usar configurações padrão). O Amazon Redshift sem servidor cria um namespace padrão com um grupo de trabalho padrão associado a esse namespace. Escolha Save configuration.

   > **Note**
   >
   > Namespace é um conjunto de objetos e usuários do banco de dados. Os namespaces agrupam todos os recursos que você usa no Redshift sem servidor, como esquemas, tabelas, usuários, unidades de compartilhamento de dados e snapshots.
   > Grupo de trabalho é um conjunto de recursos computacionais. Os grupos de trabalho abrigam os recursos computacionais que o Redshift sem servidor usa para executar tarefas computacionais.

   A captura de tela a seguir mostra as configurações padrão para o Amazon Redshift sem servidor.



3. Após a conclusão da configuração, escolha Continue (Continuar) para acessar o Serverless dashboard (Painel do Serverless). Você pode ver que o grupo de trabalho e o namespace de tecnologia sem servidor estão disponíveis.



> **Note**
>
> Se o Redshift sem servidor não criar o grupo de trabalho com êxito, você poderá fazer o seguinte:
>
> - Solucione todos os erros relatados pelo Redshift sem servidor, como ter poucas sub-redes na Amazon VPC.
> - Exclua o namespace selecionando default-namespace no painel do Redshift sem servidor e, depois, escolhendo Ações, Excluir namespace. A exclusão de um namespace leva alguns minutos.
> - Quando você abre o console do Redshift sem servidor novamente, a tela de boas-vindas é exibida.

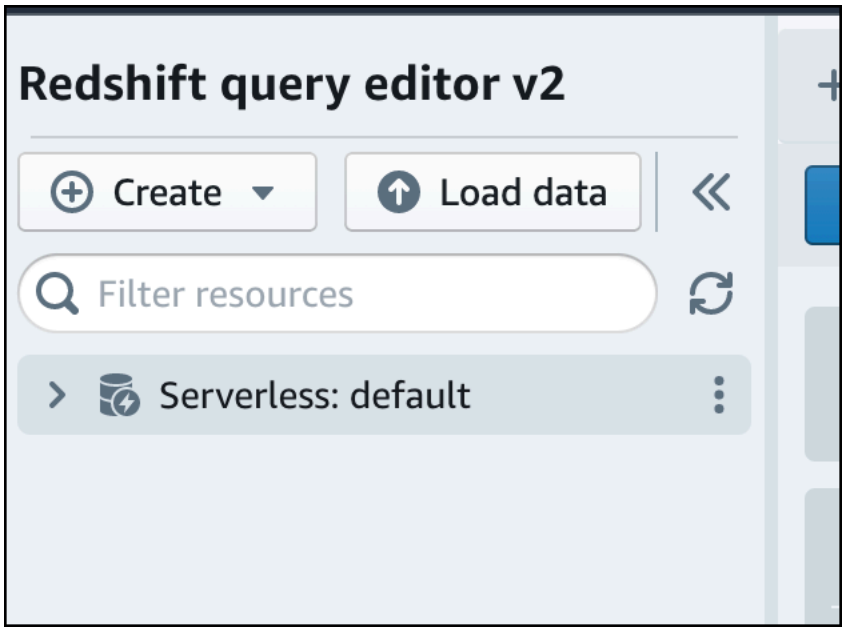
Carregar dados de exemplo

Agora que você configurou seu data warehouse com o Amazon Redshift sem servidor, pode usar o editor de consultas v2 do Amazon Redshift para carregar dados de amostra.

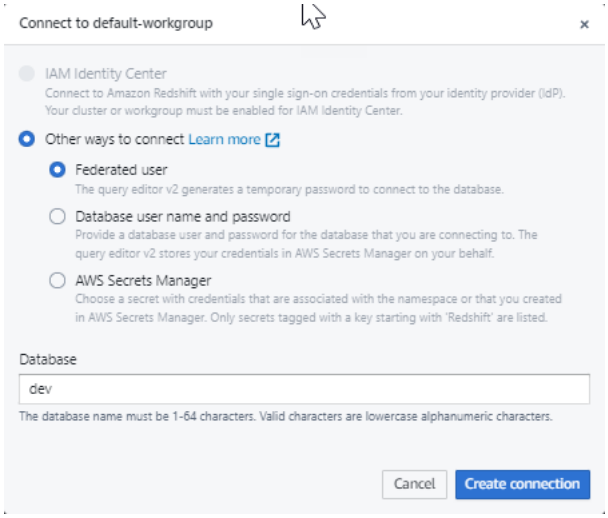
1. Para iniciar o editor de consultas v2 pelo console do Amazon Redshift sem servidor, escolha Consultar dados. Quando você invoca o editor de consultas v2 no console do Amazon Redshift, abre-se uma nova guia no navegador com o editor de consultas. O editor de consultas v2 se conecta da máquina cliente ao ambiente do Amazon Redshift Serverless.

2. Para este guia, você usará sua conta de administrador da AWS e a conta padrão AWS KMS key. Para ter informações sobre como configurar o Editor de Consultas v2, inclusive quais permissões são necessárias, consulte Configurar sua Conta da AWS no Guia de gerenciamento do Amazon Redshift. Para ter informações sobre como configurar o Amazon Redshift para usar uma chave gerenciada pelo cliente ou para alterar a chave do KMS usada pelo Amazon Redshift, consulte Alterar a chave do AWS KMS para um namespace.

3. Para se conectar a um grupo de trabalho, escolha o nome do grupo de trabalho no painel de exibição em árvore.

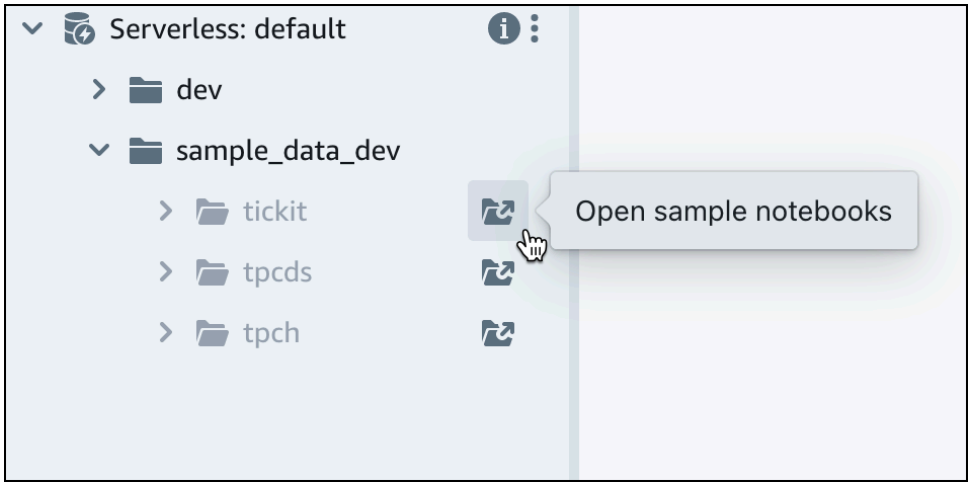


4. Ao se conectar a um novo grupo de trabalho pela primeira vez no editor de consultas v2, selecione o tipo de autenticação que deseja usar para se conectar ao grupo de trabalho. Para este guia, mantenha a seleção de Usuário federado e escolha Criar conexão.



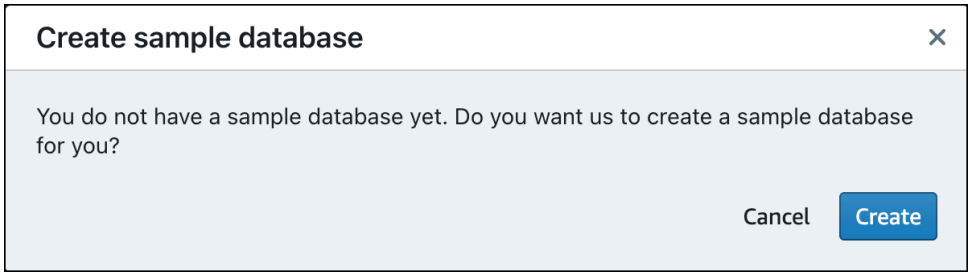
Depois de se conectar, você poderá optar por carregar dados de amostra do Amazon Redshift sem servidor ou de um bucket do Amazon S3.

5. No grupo de trabalho padrão do Amazon Redshift Serverless, expanda o banco de dados sample_data_dev. Há três esquemas de amostra correspondentes a três conjuntos de dados de amostra que você pode carregar no banco de dados do Amazon Redshift sem servidor. Escolha o conjunto de dados de amostra que você deseja carregar e selecione Abrir caderno de exemplo.



> **Note**
>
> Caderno SQL é um contêiner das células SQL e Markdown. É possível usar cadernos para organizar, anotar e compartilhar vários comandos SQL em um único documento.

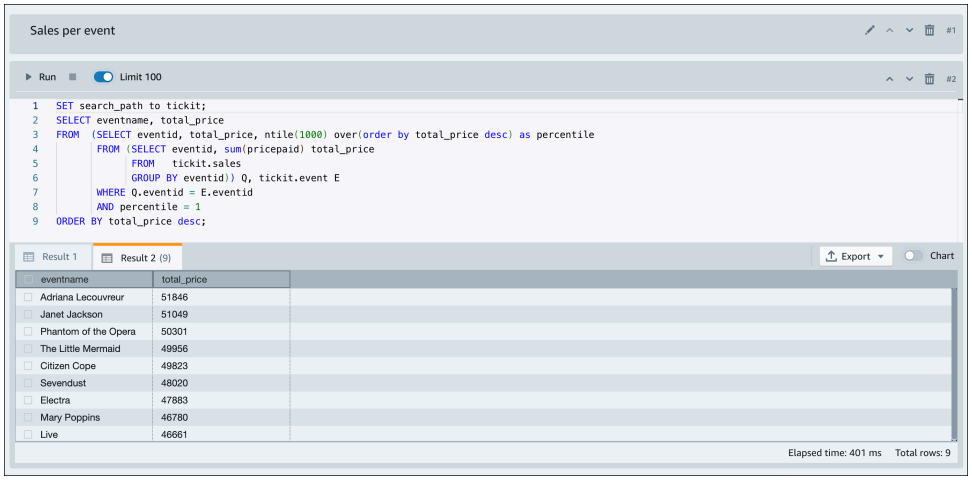
6. Ao carregar dados pela primeira vez, o editor de consultas v2 solicitará que você crie um banco de dados de amostra. Escolha Criar.



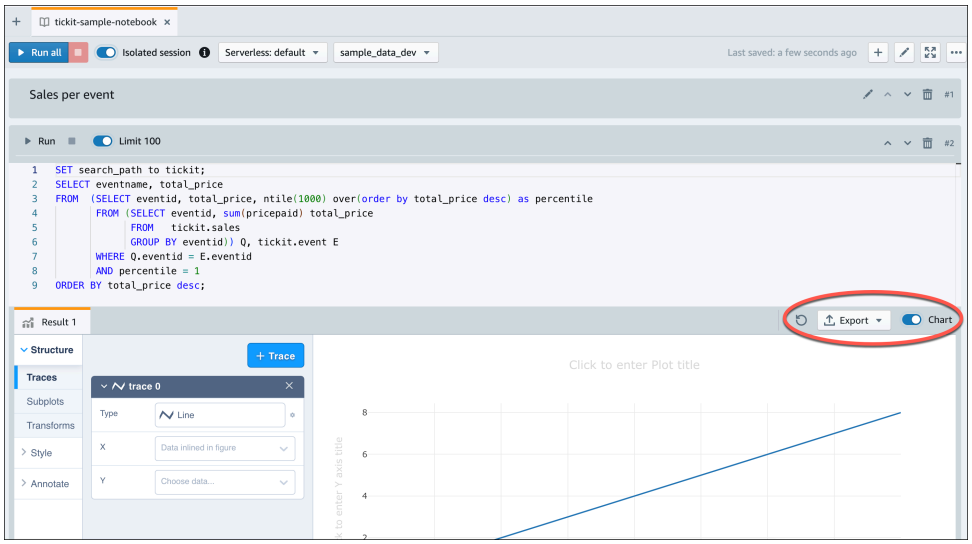
Executar consultas de amostra

Depois de configurar o Amazon Redshift sem servidor, você pode começar a usar um conjunto de dados de amostra no Amazon Redshift sem servidor. O Amazon Redshift sem servidor carrega automaticamente o conjunto de dados de amostra, como o conjunto de dados “tickit”, e você pode consultar os dados imediatamente.

- Quando o Amazon Redshift sem servidor terminar de carregar os dados de amostra, todas as consultas de amostra serão carregadas no editor. Você pode escolher Executar tudo para executar todas as consultas dos cadernos de amostra.



Você também pode exportar os resultados como um arquivo JSON ou CSV, ou exibir os resultados em um gráfico.



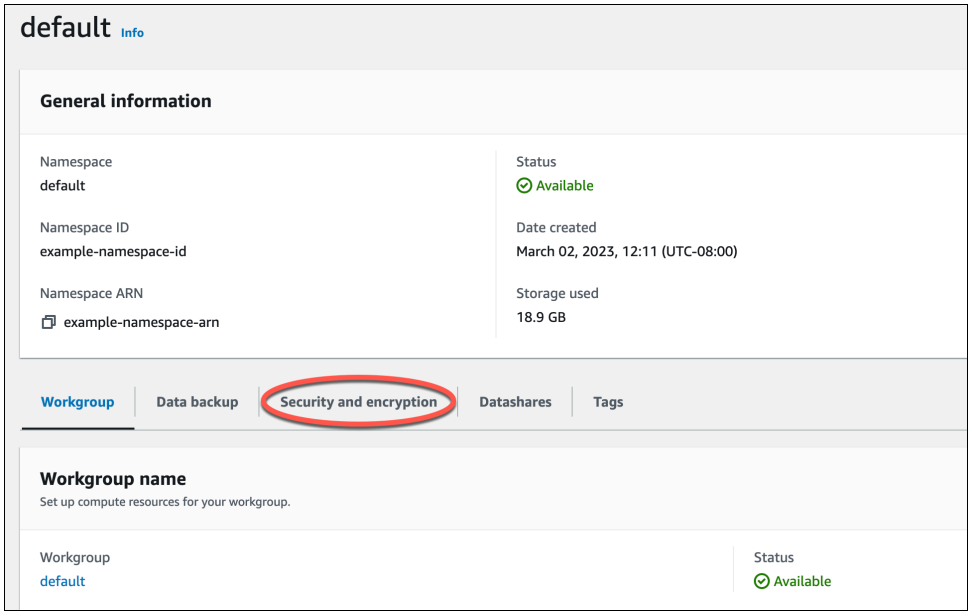
Você também pode carregar dados de um bucket do Amazon S3. Para saber mais, consulte the section called “Carregar dados do Amazon S3”.

Carregar dados do Amazon S3

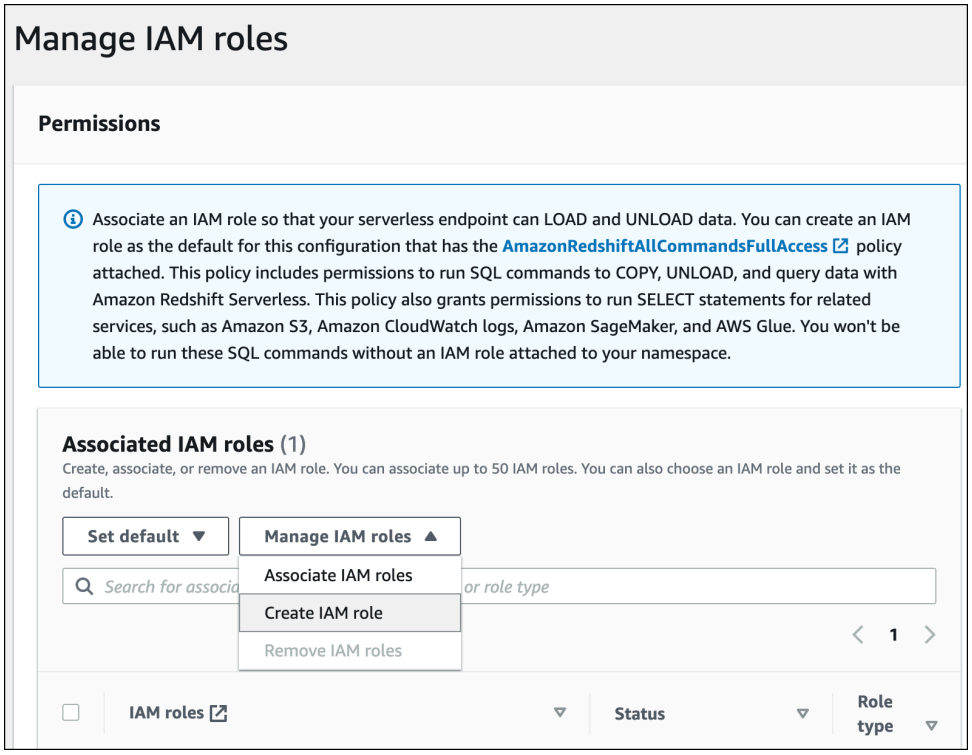
Depois de criar seu data warehouse, você poderá carregar dados do Amazon S3.

Neste ponto, você tem um banco de dados chamado `dev`. Em seguida, 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.

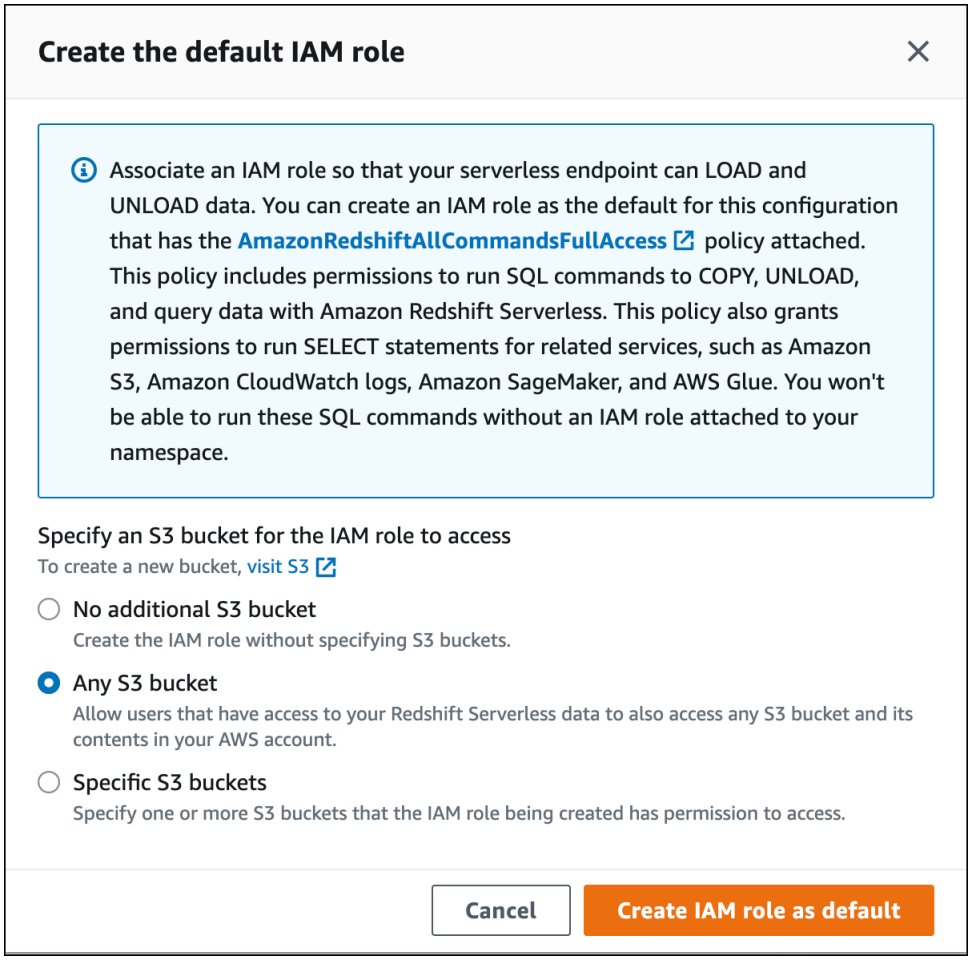
1. Para carregar dados do Amazon S3, primeiro você deve criar um perfil do IAM com as permissões necessárias e anexá-lo ao seu namespace de tecnologia sem servidor. Para fazer isso, retorne ao console do Redshift sem servidor e escolha Configuração do namespace. No menu de navegação, escolha seu namespace e selecione Segurança e criptografia. Escolha Gerenciar funções do IAM.



2. Expanda o menu Gerenciar perfis do IAM e escolha Criar perfil do IAM.



3. Escolha o nível de acesso ao bucket do S3 que você deseja conceder a esse perfil e selecione Criar perfil do IAM como padrão.



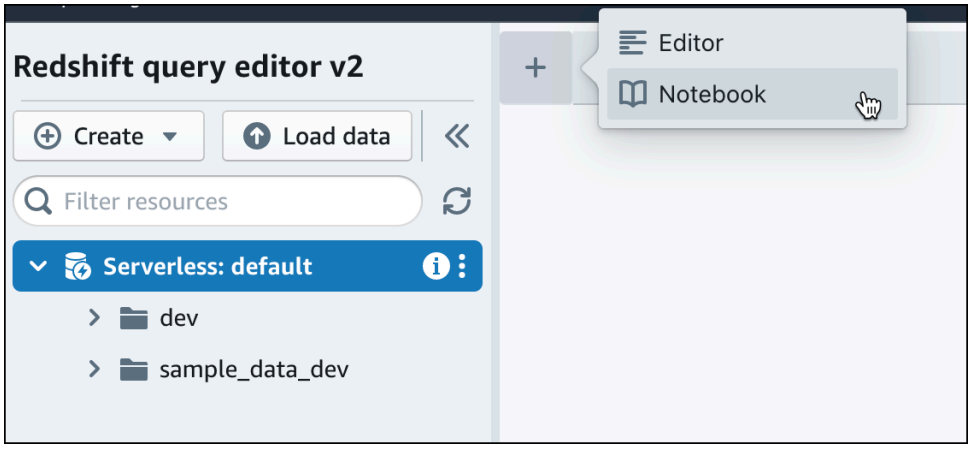
4. Escolha Salvar alterações. Agora é possível carregar dados de amostra do Amazon S3.

As etapas a seguir usam dados em um bucket público do S3 no Amazon Redshift, mas você pode replicar as mesmas etapas usando seu próprio bucket do S3 e comandos SQL.

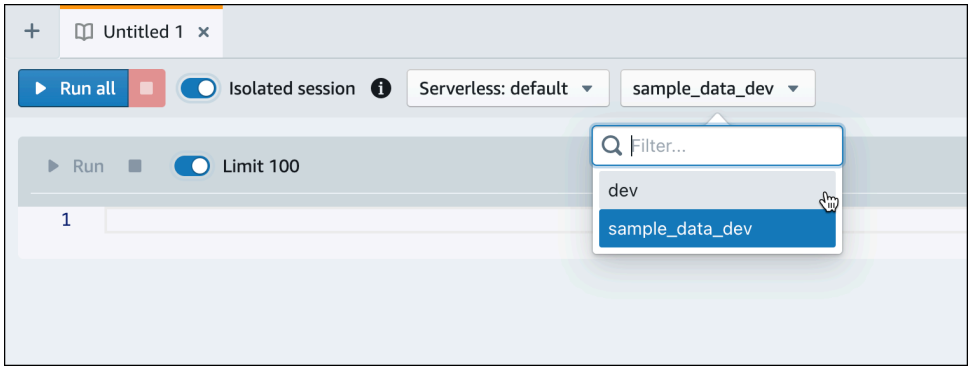
Carregar dados de amostra do Amazon S3

1. No editor de consultas v2, escolha

   Adicionar e selecione Caderno para criar um caderno SQL.



2. Mude para o banco de dados `dev`.



3. Crie tabelas.

   Se você estiver usando o editor de consultas v2, copie 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 event(
eventid integer not null distkey,
venueid smallint not null,
catid smallint not null,
dateid smallint not null sortkey,
eventname varchar(200),
starttime timestamp);

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

4. No editor de consultas v2, crie uma célula SQL em seu caderno.

5. Agora use o comando COPY no editor de consultas v2 para carregar grandes conjuntos de dados do Amazon S3 ou Amazon DynamoDB no Amazon Redshift. Para obter mais informações sobre sintaxe de COPY, consulte COPY no Guia do desenvolvedor de banco de dados do Amazon Redshift.

   Você pode executar o comando COPY com alguns dados de amostra disponíveis em um bucket público do S3. Execute os comandos a seguir no editor de consultas v2.

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

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

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

6. Depois de carregar os dados, crie outra célula SQL em seu caderno e tente realizar algumas consultas. Para obter mais informações sobre como trabalhar com o comando SELECT, consulte SELECT no Guia do desenvolvedor do Amazon Redshift. Para entender a estrutura e os esquemas dos dados de amostra, explore o uso do editor de consultas v2.

```
-- 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;
```

Agora que você carregou dados e executou algumas consultas de amostra, já pode explorar outras áreas do Amazon Redshift sem servidor. Confira a lista a seguir para saber mais sobre como você pode usar o Amazon Redshift sem servidor.

- Você pode carregar dados de um bucket do Amazon S3. Para obter mais informações, consulte Carregar dados do Amazon S3.
- É possível usar o editor de consultas v2 para carregar dados de um arquivo local separado por caracteres de até 5 MB. Para obter mais informações, consulte Carregar dados de um arquivo local.
- Você pode se conectar ao Amazon Redshift sem servidor com ferramentas SQL de terceiros com o driver JDBC e ODBC. Para obter mais informações, consulte Conectar-se ao Amazon Redshift sem servidor.
- Também é possível usar a API de dados do Amazon Redshift para se conectar ao Amazon Redshift Serverless. Para obter mais informações, consulte Usar a API de dados do Amazon Redshift.
- Você pode usar seus dados no Amazon Redshift sem servidor com o Redshift ML para criar modelos de machine learning com o comando CREATE MODEL. Consulte Tutorial: Como criar modelos de rotatividade de clientes para aprender a criar um modelo de ML do Redshift.
- Você pode consultar dados de um data lake do Amazon S3 sem carregar nenhum dado no Amazon Redshift sem servidor. Para obter mais informações, consulte Consultar um data lake.

Conceitos básicos de data warehouses provisionados do Amazon Redshift

Se estiver usando o Amazon Redshift pela primeira vez, recomendamos que leia as seções a seguir que ajudarão a começar a usar clusters provisionados. O fluxo básico do Amazon Redshift é criar recursos provisionados, conectar-se ao Amazon Redshift, carregar dados de amostra e executar consultas nos dados. Neste guia, você pode escolher carregar dados de amostra do Amazon Redshift ou de um bucket do Amazon S3. Os dados de amostra são usados em toda a documentação do Amazon Redshift para demonstrar os recursos.

Este tutorial demonstra como usar clusters provisionados pelo Amazon Redshift, que são objetos de data warehouse da AWS para os quais você gerencia recursos do sistema. Você também pode usar o Amazon Redshift com grupos de trabalho sem servidor, que são objetos de data warehouse que escalam automaticamente em resposta ao uso. Para começar a usar o Redshift sem servidor, consulte Conceitos básicos de data warehouse do Amazon Redshift sem servidor.

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

> **Important**
>
> O cluster provisionado para este exercício é executado em um ambiente ativo. Enquanto estiver em execução, ele vai acumular cobranças na 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. A seção final deste capítulo explica como fazer isso.

Faça login no Console de gerenciamento da AWS e abra o console do Amazon Redshift em https://console.aws.amazon.com/redshiftv2/.

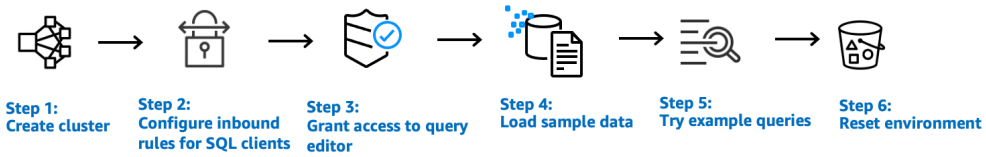
Recomendamos que você comece acessando o Painel de clusters provisionado para começar a usar o console do Amazon Redshift.

Dependendo da configuração, os seguintes itens aparecem no painel de navegação do console provisionado do Amazon Redshift:

- Redshift sem servidor: acesse e analise dados sem a necessidade de configurar, ajustar e gerenciar clusters provisionados do Amazon Redshift.
- Painel de clusters provisionado: veja a lista de clusters na Região da AWS, confira as Métricas de cluster e a Visão geral da consulta para ter 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 a lista de clusters nessa Região da AWS, escolha um cluster para iniciar a consulta ou realize ações relacionadas ao cluster. Também é possível criar um novo cluster a partir desta página.
- Editor de consultas: execute consultas em bancos de dados hospedados no cluster do Amazon Redshift. Em vez disso, recomendamos usar o Editor de Consultas v2.
- Editor de Consultas v2: o Editor de Consultas do Amazon Redshift v2 é uma aplicação de cliente SQL separada, baseada na web, 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 associá-la 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.
- Integrações ETL zero: gerencie integrações que disponibilizam dados transacionais no Amazon Redshift depois de serem gravados em fontes compatíveis.
- Conexões do Centro de Identidade do IAM: configure a conexão entre o Amazon Redshift e o Centro de Identidade do IAM.
- 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.
- Integração de parceiro da AWS: crie integração com um parceiro da AWS compatível.

- 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.

Neste tutorial, você realizará as seguintes etapas:



Tópicos

- Como se cadastrar na AWS
- Determinar regras de firewall
- Etapa 1: Criar um cluster de amostra do Amazon Redshift
- Etapa 2: Configurar regras de entrada para clientes SQL
- Etapa 3: Conceder acesso a um cliente SQL e executar consultas
- Etapa 4: Carregar dados do Amazon S3 para o Amazon Redshift
- Etapa 5: Experimentar exemplo de consultas usando o editor de consultas
- Etapa 6: Redefinir o ambiente

Como se cadastrar na 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 online.

Parte do procedimento de inscrição envolve receber uma chamada telefônica ou uma mensagem de texto e inserir um código de verificação pelo teclado do telefone.

Quando você se inscreve para uma Conta da AWS, um Usuário raiz da conta da AWS é criado. O usuário-raiz tem acesso a todos os Serviços da AWS e recursos na conta. Como prática recomendada de segurança, atribua o acesso administrativo a um usuário e use somente o usuário-raiz para executar tarefas que exigem acesso de usuário-raiz.

Determinar regras de firewall

> **Note**
>
> Este tutorial pressupõe que o cluster use a porta padrão 5439 e que o Editor de Consultas do Amazon Redshift v2 possa ser usado para executar comandos SQL. Ele não entra em detalhes sobre as configurações de rede nem sobre a configuração de um cliente SQL que possa ser necessário no ambiente.

Em alguns ambientes, você deve especificar uma porta ao iniciar o cluster do Amazon Redshift. Você usa essa porta com o URL do endpoint do cluster para acessar o cluster. 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.

Etapa 1: Criar um cluster de amostra do Amazon Redshift

Neste tutorial, você acompanhará o processo para criar um cluster do Amazon Redshift com um banco de dados. Depois, você carregará um conjunto de dados do Amazon S3 em tabelas do banco de dados. Você pode usar esse cluster para avaliar o serviço do Amazon Redshift.

Antes de começar a configurar um cluster do Amazon Redshift, não se esqueça de cumprir todos os pré-requisitos necessários, como Como se cadastrar na AWS e Determinar regras de firewall.

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 SQL COPY para carregar dados do Amazon Simple Storage Service (Amazon S3). Você concede essas permissões usando o AWS Identity and Access Management (IAM). É possível fazer isso por meio de um perfil do IAM criado e associado ao cluster. Para ter mais informações sobre credenciais e permissões de acesso, consulte Credenciais e permissões de acesso no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Para criar um cluster do Amazon Redshift.

1. Faça login no Console de gerenciamento da AWS e abra o console do Amazon Redshift em https://console.aws.amazon.com/redshiftv2/.

   > **Important**
   >
   > Se você usar credenciais de usuário do IAM, verifique se o usuário tem as permissões necessárias para realizar as operações de cluster. Para ter mais informações, consulte Segurança no Amazon Redshift no Guia de gerenciamento do Amazon Redshift.

2. No Console da AWS, 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 comporte tipos de nó RA3. Para conferir a lista de Regiões da AWS compatíveis com tipos de nó RA3, consulte Visão geral dos tipos de nó RA3 no Guia de gerenciamento do Amazon Redshift. Para saber

mais sobre as especificações de cada tipo e tamanho de nó, consulte Detalhes do tipo de nó.

- Se você não souber o tamanho do cluster, escolha Ajude-me a escolher. Isso abre 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 Node type (Tipo de nó) e número de Nodes (Nós) para dimensionar seu cluster para a prova de conceito.

  Para este tutorial, escolha ra3.4xlarge em Tipo de nó e 2 em Número de nós.

  Se uma opção para a Configuração AZ estiver disponível, selecione Single-AZ.
- 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 Tickit do conjunto de dados de amostra para o banco de dados `dev` padrão e o esquema `public`.

5. Na seção Configuração do banco de dados, especifique um valor para Nome do usuário administrador. Em Senha do administrador, escolha uma das seguintes opções:

   - Gere uma senha: use uma senha gerada pelo Amazon Redshift.
   - Adicionar manualmente uma senha de administrador: use a própria senha.
   - Gerenciar credenciais de administrador no AWS Secrets Manager: o Amazon Redshift usa AWS Secrets Manager para gerar e gerenciar a senha de administrador. O uso do AWS Secrets Manager para gerar e gerenciar o segredo da senha incorre em uma taxa. Para obter informações sobre definição de preços do AWS Secrets Manager, consulte Definição de preços do AWS Secrets Manager.

   Para este tutorial, use estes valores:

   - Nome do usuário administrador: digite **`awsuser`**.
   - Senha do usuário administrador: insira **`Changeit1`** 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 ter informações sobre como gerenciar o perfil padrão do IAM para um cluster, consulte Criar uma função do IAM como padrão para o Amazon Redshift no Guia de gerenciamento do Amazon Redshift.

7. (Opcional) Na seção Configurações adicionais, desative Usar padrões para modificar as configurações Rede e segurança, Configuração do banco de dados, Manutenção, 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 ter mais informações

sobre o roteamento aprimorado da Amazon VPC, consulte Enhanced VPC routing no Guia de gerenciamento do Amazon Redshift.

8. Selecione Criar cluster. Aguarde até que o cluster seja criado com o status `Available` na página Clusters.

Etapa 2: Configurar regras de entrada para clientes SQL

> **Note**
>
> Recomendamos que você ignore esta etapa e acesse o cluster usando o Editor de Consultas do Amazon Redshift v2.

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.

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.

   Consulte mais informações sobre regras de firewall em Regras de grupos de segurança no Guia do usuário do Amazon EC2.
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 a um cliente SQL e executar consultas

Para consultar bancos de dados hospedados pelo cluster do Amazon Redshift, há algumas opções para clientes SQL. Isso inclui:

- Conecte-se ao cluster e execute consultas usando o Editor de Consultas do Amazon Redshift v2.

Se você usar o Editor de Consultas v2, não precisará baixar e configurar uma aplicação de cliente SQL. Você pode iniciar o Editor de Consultas do Amazon Redshift v2 por meio do console do Amazon Redshift.

- Conecte-se ao cluster usando RSQL. Para ter mais informações, consulte Conectar-se com o Amazon Redshift RSQL no Guia de gerenciamento do Amazon Redshift.
- Conecte-se ao cluster por meio de uma ferramenta de cliente SQL, como o SQL Workbench/J. Para ter mais informações, consulte Conectar-se com SQL Workbench/J no Guia de gerenciamento do Amazon Redshift.

Este tutorial usa o Editor de Consultas do Amazon Redshift v2 como um método fácil para executar consultas em bancos de dados hospedados pelo cluster do Amazon Redshift. Após a criação do cluster, é possível executar consultas imediatamente. Para ter detalhes sobre as considerações ao usar o Editor de Consultas do Amazon Redshift v2, consulte Considerações ao trabalhar com o editor de consultas v2 no Guia de gerenciamento 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. Os recursos do Editor de Consultas do Amazon Redshift v2 incluem consultas, cadernos e gráficos salvos. Por padrão, uma chave de propriedade da AWS é usada para criptografar recursos. Como alternativa, 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 ter mais informações, consulte Configurar sua Conta da AWS no Guia de gerenciamento do Amazon Redshift.

Para acessar o editor de consultas v2, é necessário ter permissão. Um administrador pode associar uma das políticas gerenciadas pela AWS para o Editor de Consultas do Amazon Redshift v2 ao usuário ou perfil 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. Consulte mais informações sobre essas políticas em Acessar o editor de consultas v2 no Guia de gerenciamento do Amazon Redshift.

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 Trabalhar com o editor de consultas v2 no Guia de gerenciamento do Amazon Redshift.

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

Depois de criar o cluster, você poderá carregar dados do Amazon S3 nas tabelas do banco de dados. Há várias maneiras de carregar dados do Amazon S3.

- É possível usar um cliente SQL para executar o comando SQL CREATE TABLE e criar uma tabela no banco de dados e, depois, usar o comando SQL COPY para carregar dados do Amazon S3. O Editor de Consultas do Amazon Redshift v2 é um cliente SQL.
- É possível usar o assistente de carregamento do Editor de Consultas do Amazon Redshift v2.

Este tutorial demonstra como usar o Editor de Consultas V2 do Amazon Redshift para executar comandos SQL para CRIAR tabelas e COPIAR dados. Inicie o Editor de Consultas v2 no painel de navegação do console do Amazon Redshift. No Editor de Consultas v2, crie uma conexão com o cluster `examplecluster` e o banco de dados denominado `dev` com o usuário administrador `awsuser`. Para este tutorial, escolha Credenciais temporárias usando um nome de usuário do banco de dados ao criar a conexão. Para ter detalhes sobre como usar o Editor de Consultas do Amazon Redshift, v2, consulte Conectar-se a um banco de dados do Amazon Redshift no Guia de gerenciamento do Amazon Redshift.

Carregar dados do Amazon S3 usando comandos SQL

No painel de edição de consultas do Editor de Consultas v2, confirme que você já se conectou com o cluster `examplecluster` e o banco de dados `dev`. Depois, crie tabelas no banco de dados e carregue dados nelas. Para este tutorial, os dados carregados estão disponíveis em um bucket do Amazon S3, acessível por meio de várias Regiões da AWS.

O procedimento a seguir cria tabelas e carrega dados de um bucket público do Amazon S3.

Use o Editor de Consultas do Amazon Redshift v2 para copiar e executar a declaração de criação de tabelas e criar uma tabela no esquema `public` do 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.

Como criar e carregar dados usando um cliente SQL, como o Editor de Consultas v2

1. Execute o comando SQL a seguir para CRIAR a tabela `sales`.

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

2. Execute o comando SQL a seguir para CRIAR a tabela `date`.

```
drop table if exists date;
create table date(
 dateid smallint not null distkey sortkey,
 caldate date not null,
 day character(3) not null,
 week smallint not null,
 month character(5) not null,
 qtr character(5) not null,
 year smallint not null,
 holiday boolean default('N'));
```

3. Carregue a tabela `sales` do Amazon S3 usando o comando COPY.

> **Note**
>
> Recomendamos usar o comando COPY para carregar grandes conjuntos de dados no Amazon Redshift por meio do Amazon S3. Para obter mais informações sobre sintaxe

de COPY, consulte COPY no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Forneça autenticação ao seu cluster para acessar o Amazon S3 em seu nome para carregar os dados de amostra. Você pode oferecer autenticação fazendo referência ao perfil do IAM criado e definido como `default` para o cluster ao escolher Criar função do IAM como padrão na criação do cluster.

Carregue a tabela `sales` usando o comando SQL a seguir. Como alternativa, é possível baixar e visualizar os dados de origem do Amazon S3 para a tabela `sales`. .

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

4. Carregue a tabela `date` usando o comando SQL a seguir. Como alternativa, é possível baixar e visualizar os dados de origem do Amazon S3 para a tabela `date`. .

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

Carregar dados do Amazon S3 usando o Editor de Consultas v2

Esta seção descreve de que forma você carrega seus dados em um cluster do Amazon Redshift. O Editor de Consultas v2 simplifica o carregamento de dados ao usar o assistente Carregar dados. O comando COPY gerado e utilizado no assistente Carregar dados do Editor de Consultas v2 é compatível com muitos 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.

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.

Para carregar seus dados do Amazon S3 para o Amazon Redshift, você pode usar o assistente de carregamento de dados do Editor de Consultas V2. Para ter informações sobre como usar o assistente de carregamento de dados, consulte Carregar dados do Amazon S3 no Guia de gerenciamento do Amazon Redshift.

Criar dados TICKIT no cluster

O TICKIT é um banco de dados de exemplo que você pode carregar opcionalmente em seu cluster do Amazon Redshift com o objetivo de aprender a consultar dados no Amazon Redshift. É possível criar o conjunto completo de tabelas TICKIT e carregar dados no cluster das seguintes maneiras:

- Ao criar um cluster no console do Amazon Redshift, você tem a opção de carregar dados TICKIT de exemplo simultaneamente. No console do Amazon Redshift, selecione Clusters e Criar cluster. Na seção Dados de exemplo, selecione Carregar dados de exemplo. O Amazon Redshift carregará automaticamente o conjunto de dados de exemplo para o banco de dados `dev` do cluster do Amazon Redshift durante a criação do cluster.
- Para se conectar a um cluster existente, faça o seguinte:
  - No console do Amazon Redshift, no painel de navegação, selecione Clusters.
  - Escolha seu cluster no painel Clusters.
  - Escolha Consultar dados e Consultar no Editor de Consultas v2.
  - Expanda examplecluster na lista de recursos. Se esta for a primeira vez que você se conecta ao cluster, a opção Conectar a examplecluster é exibida. Escolha Nome de usuário e senha do banco de dados. Deixe o banco de dados como **`dev`**. Especifique **`awsuser`** para o nome de usuário e **`Changeit1`** para a senha.
  - Escolha Criar conexão.
- Com o Editor de Consultas do Amazon Redshift v2, é possível carregar dados TICKIT em um banco de dados de exemplo chamado sample_data_dev. Escolha o banco de dados sample_data_dev na lista de recursos. Ao lado do nó tickit, escolha o ícone Abrir cadernos de exemplo. Confirme que você deseja criar o banco de dados de exemplo.
- O Editor de Consultas do Amazon Redshift v2 cria o banco de dados de exemplo além de um caderno de exemplo chamado tickit-sample-notebook. Você pode escolher Executar tudo para executar esse caderno e consultar dados no banco de dados de exemplo.

Para visualizar detalhes sobre os dados TICKIT, consulte Banco de dados de exemplo no Guia do desenvolvedor de banco de dados do Amazon Redshift.

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

Para configurar e usar o Editor de Consultas do Amazon Redshift v2 e consultar um banco de dados, confira Trabalhar com o editor de consultas v2 no Guia de gerenciamento do Amazon Redshift.

Agora, tente alguns exemplos de consultas, como mostrado a seguir. Para criar consultas no Editor de Consultas V2, escolha o ícone + no canto superior direito do painel de consulta e selecione SQL. Uma nova página de consulta é exibida, na qual você pode copiar e colar as consultas SQL a seguir.

> **Note**
>
> Primeiro, você deve executar a primeira consulta no caderno, que define a configuração `search_path` do servidor para o esquema `tickit` usando o seguinte comando SQL:
>
> ```
> set search_path to tickit;
> ```

Para ter mais informações sobre como trabalhar com o comando SELECT, consulte SELECT no Guia do desenvolvedor de banco de dados 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;
```

Etapa 6: Redefinir o ambiente

Nas etapas anteriores, você criou com êxito um cluster do Amazon Redshift, carregou dados em tabelas e consultou dados usando um cliente SQL, como o Editor de Consultas do Amazon Redshift v2.

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, convém manter o cluster de exemplo em execução se pretende experimentar tarefas em outros guias do Amazon Redshift ou tarefas descritas em Executar comandos para definir e usar um banco de dados no data warehouse.

Para excluir um cluster

1. Faça login no Console de gerenciamento da AWS e abra o console do Amazon Redshift em https://console.aws.amazon.com/redshiftv2/.
2. No menu de navegação, escolha Clusters para exibir sua lista de clusters.
3. Escolha o cluster `examplecluster`. Em Ações, escolha Excluir. A página Delete examplecluster? é exibida.

4. Confirme o cluster a ser excluído, desmarque a configuração Criar snapshot final e insira **`delete`** para confirmar a exclusão. 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 para saber mais sobre o Amazon Redshift.

Executar comandos para definir e usar um banco de dados no data warehouse

Tanto os data warehouses do Redshift sem servidor quanto os data warehouses provisionados do Amazon Redshift contêm bancos de dados. Depois que o data warehouse é iniciado, é possível gerenciar a maioria das ações do banco de dados usando comandos SQL. Com poucas exceções, a funcionalidade e a sintaxe do SQL são as mesmas para todos os bancos de dados do Amazon Redshift. Para ter detalhes sobre os comandos SQL disponíveis com o Amazon Redshift, consulte Comandos SQL no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Quando se cria um data warehouse, o Amazon Redshift também cria o banco de dados `dev` padrão na maioria das situações. Após a conexão com o banco de dados `dev`, é possível criar outro banco de dados.

As seções a seguir descrevem tarefas comuns de banco de dados quando se trabalha com bancos de dados do Amazon Redshift. As tarefas começam com a criação de um banco de dados e, se você prosseguir até a última tarefa, poderá descartar o banco de dados para excluir todos os recursos criados.

Para os exemplos nesta seção, suponha o seguinte:

- Você criou um data warehouse do Amazon Redshift.
- Você estabeleceu uma conexão com o data warehouse por meio de sua ferramenta de cliente SQL, como o Editor de Consultas do Amazon Redshift v2. Para ter mais informações sobre o Editor de Consultas v2, confira Consultar um banco de dados usando o editor de consultas v2 do Amazon Redshift no Guia de gerenciamento do Amazon Redshift.

Tópicos

- Conectar-se a data warehouses de dados do Amazon Redshift
- Criar um banco de dados
- Criar um usuário
- Criar um esquema
- Criar uma tabela
- Carregar dados
- Consultar as visualizações e tabelas do sistema

- Cancelar uma consulta

Conectar-se a data warehouses de dados do Amazon Redshift

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

- Selecione Consultar dados para usar o Editor de Consultas v2 e executar consultas em bancos de dados hospedados pelo cluster do Amazon Redshift. Depois de criar o cluster, é possível executar consultas imediatamente usando o editor de consultas v2.

  Para ter mais informações, confira Consultar um banco de dados usando o editor de consultas v2 do Amazon Redshift no Guia de gerenciamento do Amazon Redshift.
- Em Trabalhar com ferramentas de cliente, selecione o cluster e conecte-se ao Amazon Redshift por meio de suas ferramentas de cliente. Para isso, utilize drivers JDBC ou ODBC copiando o URL do driver JDBC ou ODBC. Use esse URL 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 de cliente SQL exigir um driver, você poderá Escolher um driver JDBC ou ODBC para baixar um driver específico do sistema operacional para se conectar ao Amazon Redshift por meio 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 2.2.

  Para obter mais informações sobre como configurar uma conexão ODBC, consulte Configurar uma conexão ODBC.

Para se conectar ao data warehouse do Redshift sem servidor, na página Painel da tecnologia sem servidor do console do Amazon Redshift, faça o seguinte:

- Use o Editor de Consultas do Amazon Redshift v2 para executar consultas em bancos de dados hospedados pelo data warehouse do Redshift sem servidor. Após a criação do data warehouse, 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 data warehouse, são necessários drivers JDBC ou ODBC para estabelecer 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 ter mais informações sobre como encontrar a string de conexão, consulte Conectar-se ao Amazon Redshift Serverless no Guia de gerenciamento do Amazon Redshift.

Criar um banco de dados

Depois de verificar se o data warehouse está funcionando, você pode criar um banco de dados. Este banco de dados é onde você realmente cria tabelas, carrega dados e executa consultas. Um data warehouse pode hospedar vários bancos de dados. Por exemplo, é possível ter um banco de dados para dados de vendas denominado `SALESDB` e um banco de dados para dados de pedidos chamado `ORDERSDB` no mesmo data warehouse.

Para criar um banco de dados chamado **`SALESDB`**, execute o comando a seguir na ferramenta de cliente SQL.

```
CREATE DATABASE salesdb;
```

> **Note**
>
> Depois de executar o comando, não se esqueça de atualizar a lista de objetos da ferramenta de cliente SQL no data warehouse para ver o novo `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. Para excluir um banco de dados e o respectivo conteúdo, consulte DROP DATABASE no Guia do desenvolvedor 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`.

Criar um usuário

Por padrão, apenas o usuário administrador que você criou ao iniciar o data warehouse tem acesso ao banco de dados padrão no data warehouse. 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 data warehouse e não por banco de dados individual.

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.

Criar um esquema

Depois de criar um novo banco de dados, você pode criar um novo esquema no banco de dados atual. Um esquema 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 visualizações no esquema `SALES` usando uma declaração 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;
```

Criar uma tabela

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

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

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

Por padrão, novos objetos de banco de dados, como tabelas, são criados no esquema padrão denominado `public` que foi criado simultaneamente com o data warehouse. 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.

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 as respectivas tabelas, é possível usar o Editor de Consultas do Amazon Redshift v2. Ou você pode ver a lista de tabelas em esquemas usando visualizações do sistema. Para obter mais informações, consulte Consultar as visualizações e tabelas do sistema.

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 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');
```

Para inserir dados em uma tabela que está em um esquema específico, execute o comando a seguir.

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

Selecionar dados de uma tabela

Após criar uma tabela e preenche-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.

Carregar dados

A maioria dos exemplos neste guia usa o conjunto de dados de exemplo TICKIT. Baixe o arquivo tickitdb.zip que contém os arquivos de dados de amostra individuais. Depois, 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.

Consultar as visualizações e tabelas do sistema

Além das tabelas que você criou, o data warehouse contém uma série de tabelas e visualizações do sistema. Essas visualizações e tabelas contêm informações sobre a instalação e sobre várias consultas e processos que são executados no sistema. É possível consultar essas visualizações e tabelas do sistema para coletar informações sobre o banco de dados. Para ter mais informações, consulte Referência de visualizações e tabelas do sistema no Guia do desenvolvedor de banco de dados do Amazon Redshift. A descrição de cada tabela ou visualização 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.

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
```

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


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
```

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

```
set search_path to '$user', 'public', 'sales';
select * from pg_table_def where tablename = 'demo';

 schemaname | tablename |  column  |          type          | encoding | distkey |
 sortkey | notnull
------------+-----------+----------+------------------------+----------+---------
+---------+---------
```

```
 public     | demo      | personid | integer                | az64     | f       |
  0 | f
 public     | demo      | city     | character varying(255) | lzo      | f       |
  0 | f
 sales      | demo      | personid | integer                | az64     | f       |
  0 | f
 sales      | demo      | city     | character varying(255) | lzo      | f       |
  0 | f
```

Para obter mais informações, consulte PG_TABLE_DEF.

Também é possível usar o Editor de Consultas do Amazon Redshift v2 para exibir todas as tabelas em um esquema especificado escolhendo primeiro um banco de dados ao qual você 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;

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

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;

  usename   | usesysid | usecreatedb | usesuper | usecatupd |  passwd  | valuntil |
 useconfig
------------+----------+-------------+----------+-----------+----------+----------
+-----------
```

```
 awsuser    |      100 | true        | true     | false     | ******** |          |
 guest      |      104 | true        | false    | false     | ******** |          |
```

Visualização de consultas recentes

No exemplo anterior, o ID do usuário (user_id) para `adminuser` é 100. Para listar as quatro consultas mais recentes executadas por `adminuser`, consulte a visualização SYS_QUERY_HISTORY.

É possível usar essa visualização para localizar o ID da consulta (QUERY) ou o ID do processo (session_id) para uma consulta executada recentemente. Você também pode usar esta visualização para verificar quanto tempo levou para uma consulta ser concluída. SYS_QUERY_HISTORY inclui os primeiros quatro mil caracteres da string de consulta (query_text) para ajudar você a localizar uma consulta específica. Use a cláusula LIMIT com a declaração SELECT para limitar os resultados.

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

O resultado será semelhante ao seguinte:

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

Determinar o ID da sessão de uma consulta em execução

Para recuperar informações da tabela do sistema sobre uma consulta, pode ser necessário especificar o ID da sessão (ID do processo) associado a essa consulta. Ou pode ser necessário encontrar o ID da sessão de uma consulta que ainda esteja em execução. Por exemplo, você precisará do ID da sessão se houver necessidade de cancelar uma consulta que está demorando muito para ser executada em um cluster provisionado. É possível consultar a tabela do sistema

STV_RECENTS para ter uma lista de IDs de sessão das consultas em execução, além da string da consulta correspondente. Se a consulta exibir várias sessões, examine o texto da consulta para determinar qual ID de sessão é necessário.

Para determinar o ID da sessão de uma consulta em execução, execute a declaração SELECT a seguir.

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

Cancelar uma consulta

Se você executar uma consulta que esteja demorando muito ou consumindo uma quantidade excessiva de recursos, cancele-a. 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 ID da sessão da consulta. Com o Editor de Consultas do Amazon Redshift v2, é possível cancelar uma consulta escolhendo o botão Cancelar enquanto a consulta está em execução.

Para localizar o ID da sessão, 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.

Para determinar o ID da sessão de uma consulta em execução, execute a declaração SELECT a seguir.

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

O resultado é parecido com o mostrado a seguir.

```
 user_id |   session_id  |   start_time               |   query_text
---------+---------------+----------------------------
+-----------------------------------------------------------------
```

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

Para cancelar a consulta com o ID de sessão `1073791534`, execute o comando a seguir.

```
CANCEL 1073791534;
```

> **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.

Se sua ferramenta de consulta não oferece suporte à execução de consultas simultaneamente, inicie outra sessão para cancelar a consulta.

Para ter mais informações sobre cancelamento de uma consulta, confira CANCEL no Guia do desenvolvedor de banco de dados do Amazon Redshift.

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 de superusuários, execute estes comandos:

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

Consultar dados que não estão em seu banco de dados do Amazon Redshift

A seguir, você encontrará informações sobre como começar a consultar dados em fontes remotas, inclusive dados do Amazon S3, gerenciadores de dados remotos, bancos de dados remotos do Amazon Redshift e sobre como treinar modelos de machine learning (ML) com o Amazon Redshift.

Tópicos

- Consultar um data lake
- Consultar dados em gerenciadores de bancos de dados remotos
- Acessar dados em outros bancos de dados do Amazon Redshift
- Treinamento de modelos de machine learning com dados do Amazon Redshift

Consultar um 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. 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. 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.

Consultar dados em gerenciadores de bancos de dados remotos

É possível juntar dados de um banco de dados do Amazon RDS e de um banco de dados do Amazon Aurora 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 ao PostgreSQL
- Conceitos básicos do uso de consultas federadas no MySQL

Acessar dados em outros bancos de dados 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 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. 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.

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. Você pode melhorar a agilidade da equipe compartilhando dados em diferentes níveis de granularidade.

Consulte mais informações sobre o compartilhamento de dados em Gerenciamento de tarefas de compartilhamento de dados no Guia do desenvolvedor de banco de dados do Amazon Redshift.

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. Ao utilizar o Amazon Redshift ML, você pode usar os dados em clusters do Amazon Redshift para treinar modelos com o Amazon SageMaker AI Autopilot e obter automaticamente o melhor modelo. Depois, você pode localizar os modelos e fazer previsões de dentro de 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.

Aprenda conceitos do Amazon Redshift

O Amazon Redshift sem servidor permite acessar e analisar dados sem todas as configurações de um data warehouse provisionado. Os recursos são provisionados automaticamente e a capacidade do data warehouse escala de maneira inteligente para oferecer performance rápida até mesmo às workloads mais exigentes e imprevisíveis. O tempo em que o data warehouse fica ocioso não é cobrado, portanto você paga apenas pelo que usa. Você pode carregar dados e começar a consultar imediatamente no editor de consultas v2 do Amazon Redshift ou na sua ferramenta de business intelligence (BI) favorita. Aproveite a melhor relação preço/performance e recursos de SQL familiares em um ambiente fácil de usar e que não exige administração.

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

- Visão geral dos recursos do Amazon Redshift sem servidor: neste tópico, você encontrará uma visão geral do Amazon Redshift sem servidor e de seus principais recursos.
- Destaques do serviço e preços: nesta página de detalhes do produto, você encontrará detalhes sobre os destaques e os preços do Amazon Redshift sem servidor.
- Conceitos básicos de data warehouse do Amazon Redshift sem servidor: neste tópico, você pode aprender mais sobre como criar um data warehouse do Amazon Redshift sem servidor e começar a consultar dados usando o Editor de Consultas v2.

Se você preferir gerenciar seus recursos do Amazon Redshift manualmente, poderá criar clusters provisionados para suas necessidades de consulta de dados. Para obter mais informações, consulte Clusters do Amazon Redshift.

Se sua organização for elegível e seu cluster estiver sendo criado em uma Região da AWS em que o Amazon Redshift sem servidor não está indisponível, você poderá criar um cluster no programa de teste gratuito do Amazon Redshift. Escolha Produção ou Teste gratuito para responder à pergunta Para que você está planejando usar esse cluster? Ao escolher Teste gratuito, você crie 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. Para obter uma lista de Regiões da AWS nas quais o Amazon Redshift sem servidor está disponível, consulte os endpoints listados para a API do Redshift sem servidor na Referência geral da Amazon Web Services.

A seguir estão alguns dos principais conceitos do Amazon Redshift sem servidor.

- Namespace: um conjunto de objetos e usuários do banco de dados. Os namespaces agrupam todos os recursos que você usa no Amazon Redshift sem servidor, como esquemas, tabelas, usuários, unidades de compartilhamento de dados e snapshots.
- Grupo de trabalho: um conjunto de recursos de computação. Os grupos de trabalho abrigam os recursos computacionais que o Amazon Redshift sem servidor usa para executar tarefas computacionais. Alguns exemplos desses recursos incluem unidades de processamento do Redshift (RPUs), grupos de segurança e limites de uso. Os grupos de trabalho têm configurações de rede e segurança que você pode definir usando o console do Amazon Redshift sem servidor, a AWS Command Line Interface ou as APIs do Amazon Redshift sem servidor.

Para obter mais informações sobre como configurar recursos de namespaces e grupos de trabalho, consulte Trabalhar com namespaces e Trabalhar com grupos de trabalho.

Veja a seguir alguns dos principais conceitos de clusters provisionados 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.

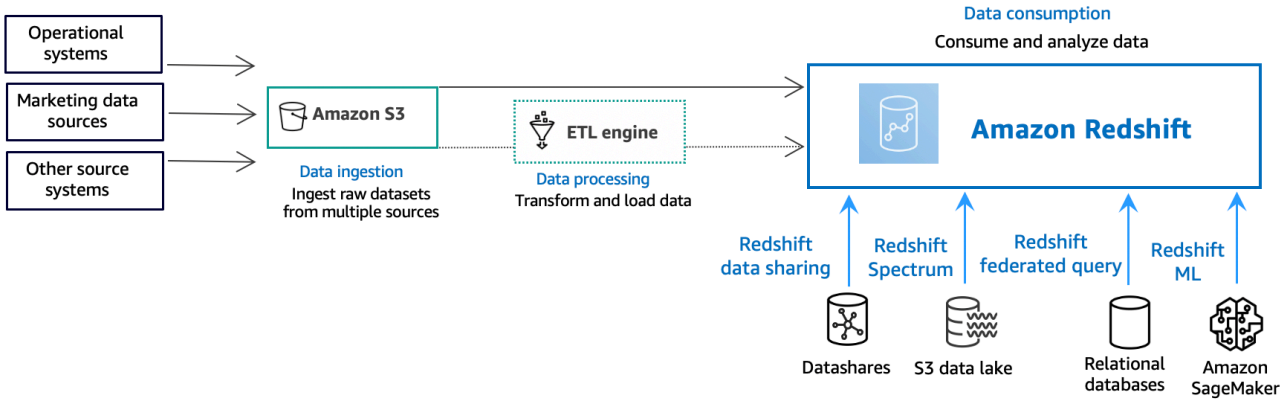
  Os dados do usuário são armazenados em um ou mais bancos de dados nos nós de computação. O cliente SQL se comunica com o nó líder, que, por sua vez, coordena as consultas em execução com os nós de computação. Para obter mais informações sobre nós de computação e nós líderes, consulte Arquitetura do sistema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

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.

Para ver alguns exemplos de workloads analíticas, consulte Consultar fontes de dados externas.

Recursos adicionais para saber mais sobre o Amazon Redshift

Para obter mais informações sobre o Amazon Redshift sem servidor, recomendamos que continue aprendendo sobre os conceitos apresentados neste guia usando os seguintes recursos do Amazon Redshift:

- Vídeos de recursos: esses vídeos ajudam você a aprender sobre os recursos do Amazon Redshift.
  - Para saber como funciona o Amazon Redshift sem servidor, assista ao vídeo a seguir. Amazon Redshift sem servidor explicado em 90 segundos.
  - Para saber como configurar um data warehouse sem servidor e começar a consultar dados, assista ao vídeo a seguir. Conceitos básicos do Amazon Redshift sem servidor.
- Guia de gerenciamento de clusters do Amazon Redshift: esse guia baseia-se neste Guia de conceitos básicos do Amazon Redshift. Fornece informações detalhadas sobre os conceitos e as tarefas de criação, gerenciamento e monitoramento de clusters provisionados do Amazon Redshift sem servidor e do Amazon Redshift.
- Guia do desenvolvedor de banco de dados do Amazon Redshift: este guia baseia-se nestes 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.
  - Referência de visualizações e tabelas do sistema: esse tópico descreve visualizações e tabelas do sistema para o Amazon Redshift.
- Tutoriais para o Amazon Redshift: este tópico mostra tutoriais sobre os recursos do 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.
  - Conceitos básicos do compartilhamento de dados: esta seção descreve como compartilhar e acessar dados em outros clusters do Amazon Redshift.
  - Usando funções SQL espaciais com o Amazon Redshift: este tutorial demonstra como usar algumas das funções SQL espaciais com o Amazon Redshift.

  - 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.
  - Conceitos básicos do Amazon Redshift ML: esta seção descreve como os usuários podem criar, treinar e implantar modelos de machine learning usando comandos SQL familiares.
- Novidades: esta página da Web lista os novos recursos e atualizações de produtos do Amazon Redshift.

Histórico do documento

> **Note**
>
> Para ver uma descrição dos novos recursos no Amazon Redshift, consulte What's new.

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

Alteração	Descrição	Data de lançamento
Atualização da documentação	O guia foi atualizado para mostrar as alterações em políticas gerenciadas do Editor de Consultas V2 e as permissões aprimoradas de acesso ao namespace e grupo de trabalho sem servidor.	21 de fevereiro de 2024
Atualização da documentação	Capturas de tela e procedimentos atualizados para mostrar as melhorias mais recentes na interface do console e os aprimoramentos do Editor de Consultas V2.	11 de março de 2023
Novo recurso	O guia foi atualizou para incluir procedimentos e fluxos de trabalho básicos do Amazon Redshift sem servidor. Foi adicionada uma seção abrangente sobre criação e gerenciamento de data warehouses sem servidor.	12 de julho de 2022
Atualização da documentação	O guia foi atualizado para apresentar o Editor de Consultas v2 como a principal interface de consulta, substituindo as referências ao editor de consultas antigo.	Fevereiro de 2022
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	30 de junho de 2021

Alteração	Descrição	Data de lançamento
	remotas, compartilhar dados e treinar modelos de Machine Learning com dados do Amazon Redshift.	
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
Novo 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

Alteração	Descrição	Data de lançamento
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