



Panduan Memulai

Amazon Redshift



Amazon Redshift: Panduan Memulai

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

Merek dagang dan tampilan dagang Amazon tidak boleh digunakan sehubungan dengan produk atau layanan apa pun yang bukan milik Amazon, dengan cara apa pun yang dapat menyebabkan kebingungan di antara pelanggan, atau dengan cara apa pun yang menghina atau mendiskreditkan Amazon. Semua merek dagang lain yang tidak dimiliki oleh Amazon merupakan kekayaan masing-masing pemiliknya, yang mungkin atau mungkin tidak berafiliasi, terkait dengan, atau disponsori oleh Amazon.

Table of Contents

Amazon Redshift Tanpa Server	1
Mendaftar untuk AWS	1
Membuat gudang data dengan Amazon Redshift Serverless	1
Memuat data sampel	4
Menjalankan kueri sampel	7
Memuat data dari Amazon S3	9
Cluster yang disediakan Amazon Redshift	18
Mendaftar untuk AWS	20
Tentukan aturan firewall	21
Langkah 1: Buat cluster sampel	21
Langkah 2: Konfigurasi aturan masuk untuk klien SQL	25
Langkah 3: Berikan akses ke klien SQL dan jalankan kueri	25
Memberikan akses ke editor kueri v2	26
Langkah 4: Muat data dari Amazon S3 ke Amazon Redshift	27
Memuat data dari Amazon S3 menggunakan perintah SQL	27
Memuat data dari Amazon S3 menggunakan editor kueri v2	29
Membuat data TICKIT di klaster Anda	29
Langkah 5: Coba contoh kueri menggunakan editor kueri	30
Langkah 6: Setel ulang lingkungan Anda	31
Tugas database umum	32
Menghubungkan ke Amazon Redshift	32
Langkah 1: Buat database	34
Langkah 2: Buat pengguna	34
Langkah 3: Buat skema	35
Langkah 4: Buat tabel	37
Masukkan baris data ke dalam tabel	37
Pilih data dari tabel	38
Langkah 5: Muat data	39
Langkah 6: Kueri tabel sistem	39
Melihat daftar nama tabel	39
Lihat pengguna	41
Lihat pertanyaan terbaru	41
Tentukan ID sesi dari kueri yang sedang berjalan	42
Langkah 7: Batalkan kueri	43

Batalkan kueri menggunakan antrean superuser	45
Menanyakan sumber data luar	46
Query data lake	46
Menanyakan sumber data jarak jauh	47
Mengakses data di klaster lain	47
Melatih model ML dengan data Redshift	48
Gambaran umum Amazon Redshift	49
Sumber daya tambahan	52
Riwayat dokumen	54
.....	lvi

Amazon Redshift Tanpa Server

Jika Anda adalah pengguna pertama kali Amazon Redshift Serverless, kami sarankan Anda membaca bagian berikut untuk membantu Anda mulai menggunakan Amazon Redshift Serverless. Alur dasar Amazon Redshift Serverless adalah membuat sumber daya tanpa server, terhubung ke Amazon Redshift Tanpa Server, memuat data sampel, dan kemudian menjalankan kueri pada data. Dalam panduan ini, Anda dapat memilih untuk memuat data sampel dari Amazon Redshift Tanpa Server atau dari bucket Amazon S3.

- [the section called “Mendaftar untuk AWS”](#)
- [the section called “Membuat gudang data dengan Amazon Redshift Serverless”](#)
- [the section called “Memuat data dari Amazon S3”](#)

Mendaftar untuk AWS

Jika Anda belum memiliki AWS akun, daftar untuk satu. Jika Anda sudah memiliki akun, Anda dapat melewati prasyarat ini dan menggunakan akun yang ada.

1. Buka <https://portal.aws.amazon.com/billing/signup>.
2. Ikuti petunjuk secara online.

Saat Anda mendaftar untuk sebuah AWS akun, pengguna root AWS akun dibuat. Pengguna root memiliki akses ke semua AWS layanan dan sumber daya di akun. Sebagai praktik terbaik keamanan, [tetapkan akses administratif ke pengguna administratif](#), dan gunakan hanya pengguna root untuk melakukan [tugas yang memerlukan akses pengguna root](#).

Membuat gudang data dengan Amazon Redshift Serverless

Saat pertama kali masuk ke konsol Amazon Redshift Tanpa Server, Anda diminta untuk mengakses pengalaman memulai, yang dapat Anda gunakan untuk membuat dan mengelola sumber daya tanpa server. Dalam panduan ini, Anda akan membuat sumber daya tanpa server menggunakan pengaturan default Amazon Redshift Serverless.

Untuk kontrol pengaturan yang lebih terperinci, pilih Sesuaikan pengaturan.

Untuk mengkonfigurasi dengan pengaturan default:

1. [Masuk ke AWS Management Console dan buka konsol Amazon Redshift di https://console.aws.amazon.com/redshiftv2/.](https://console.aws.amazon.com/redshiftv2/)

Pilih Coba Amazon Redshift Tanpa Server.

2. Di bawah Konfigurasi, pilih Gunakan pengaturan default. Amazon Redshift Serverless membuat namespace default dengan workgroup default yang terkait dengan namespace ini. Pilih Simpan konfigurasi.

Tangkapan layar berikut menunjukkan pengaturan default untuk Amazon Redshift Serverless.

Get started with Amazon Redshift Serverless

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

Configuration [Info](#)

Use default settings

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

Customize settings

Customize your settings for your specific needs.

▼ How it works



Using the default settings

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



Customizing the settings

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

Namespace [Info](#)

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

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

Namespace name
default

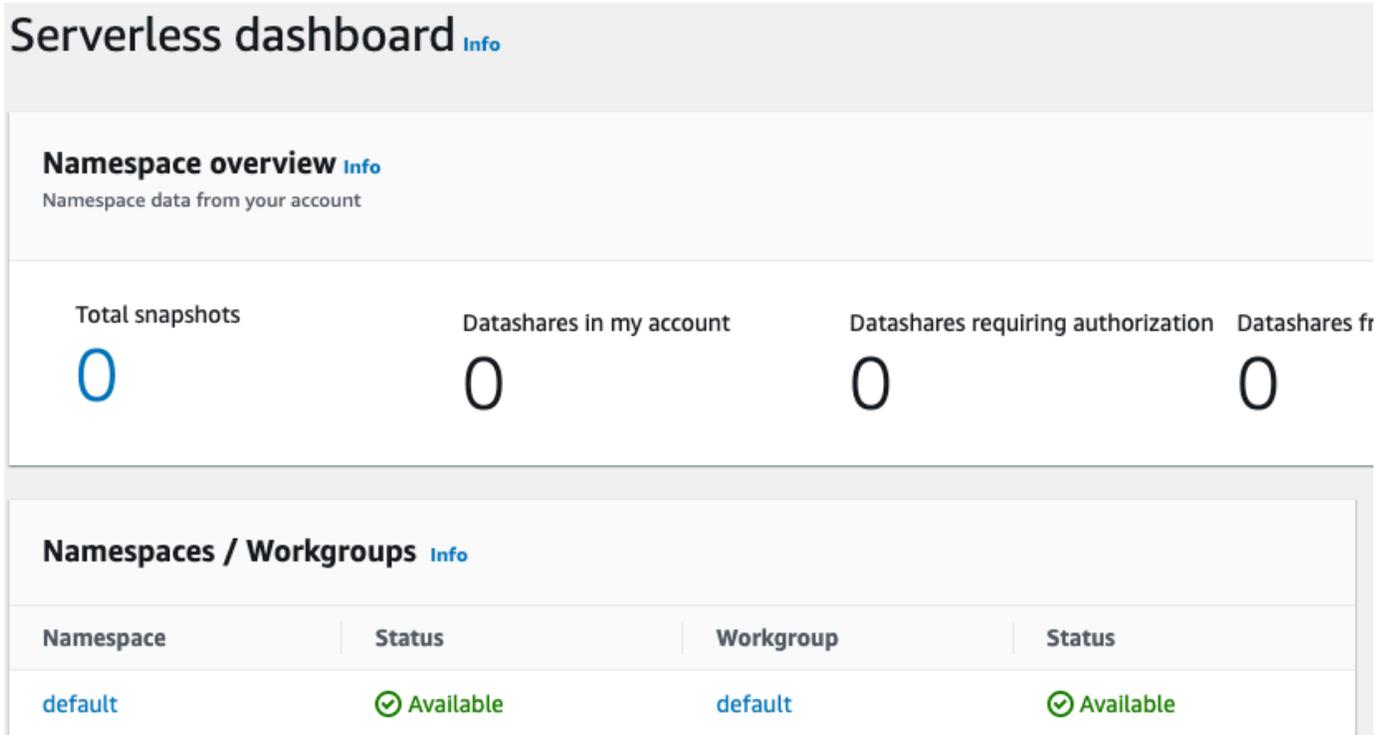
Database name and password

Database name
dev

Admin user credentials
Created based on IAM credentials

Permissions

3. Setelah penyiapan selesai, pilih Lanjutkan untuk pergi ke dasbor Tanpa Server Anda. Anda dapat melihat bahwa workgroup tanpa server dan namespace tersedia.



Serverless dashboard [Info](#)

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

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

Namespaces / Workgroups [Info](#)

Namespace	Status	Workgroup	Status
default	✔ Available	default	✔ Available

Memuat data sampel

Setelah menyiapkan gudang data dengan Amazon Redshift Serverless, Anda dapat menggunakan editor kueri Amazon Redshift v2 untuk memuat data sampel.

1. Untuk meluncurkan editor kueri v2 dari konsol Amazon Redshift Tanpa Server, pilih Kueri data. Saat Anda memanggil editor kueri v2 dari konsol Amazon Redshift Tanpa Server, tab browser baru terbuka dengan editor kueri. Editor kueri v2 terhubung dari mesin klien Anda ke lingkungan Amazon Redshift Tanpa Server.



Amazon Redshift Serverless

Serverless dashboard [Info](#)

[Refresh](#) [Query data](#) [Create workgroup](#)

2. Jika Anda meluncurkan editor kueri v2 untuk pertama kalinya, Anda harus mengonfigurasi AWS KMS enkripsi sebelum dapat melanjutkan. Secara opsional, Anda juga dapat menentukan URI ke bucket S3 untuk pemuatan data nanti. Setelah melakukannya, pilih Konfigurasi akun.

Configure account

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

AWS KMS encryption

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

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

Customize encryption settings (advanced)

S3 bucket - optional

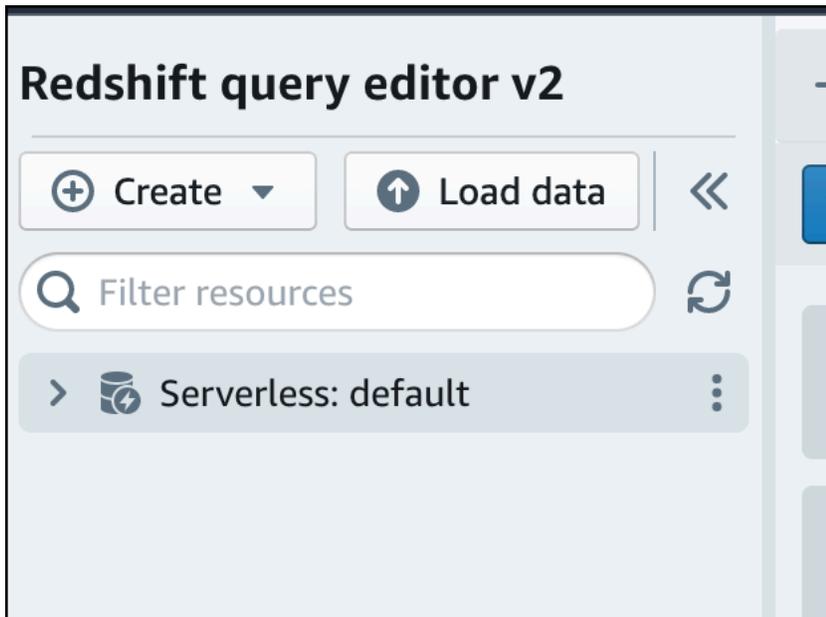
Enter an Amazon S3 path for this account. This path can be used for features such as load data from a local file store. Account users and the S3 bucket must be configured with the proper permissions to secure user's data. [Learn more](#) 

S3 URI

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

Untuk mempelajari lebih lanjut tentang mengonfigurasi editor kueri v2, termasuk izin mana yang diperlukan, lihat [Mengonfigurasi Anda Akun AWS di Panduan](#) Manajemen Amazon Redshift.

3. Untuk menyambung ke workgroup, pilih nama workgroup di panel tampilan pohon.



4. Saat menghubungkan ke workgroup baru untuk pertama kalinya dalam query editor v2, Anda harus memilih jenis otentikasi yang akan digunakan untuk terhubung ke workgroup. Untuk panduan ini, biarkan pengguna Federasi dipilih, dan pilih Buat koneksi.

Connect to default ✕

Authentication [Learn more](#)

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

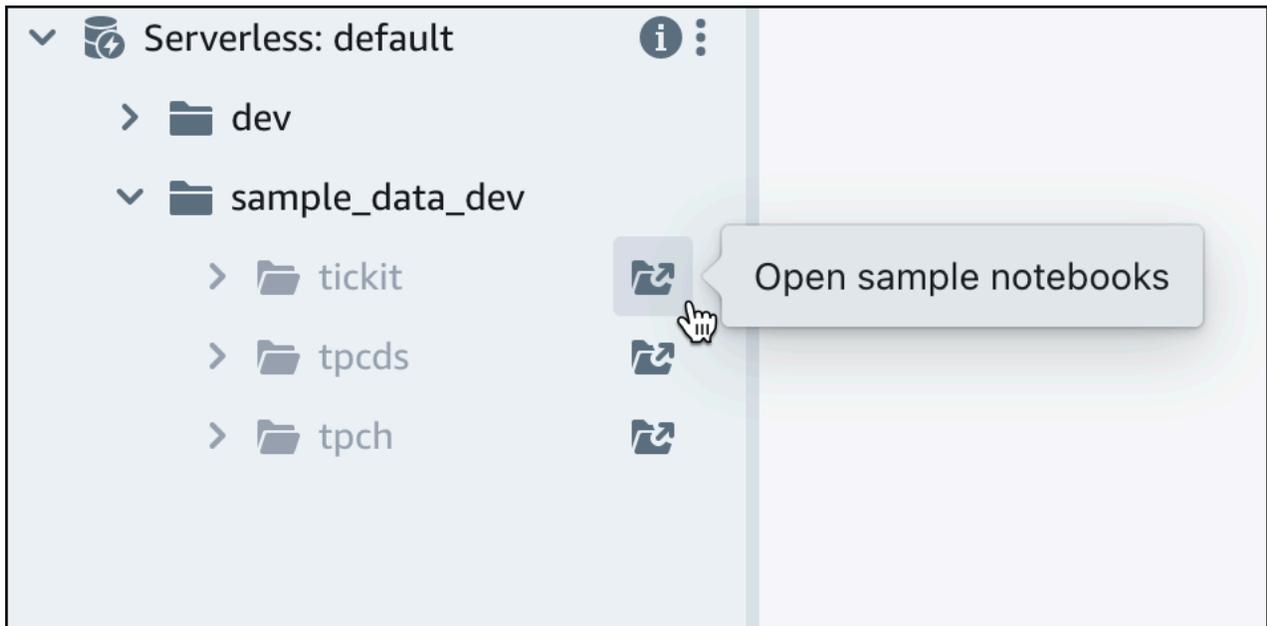
Database user name and password
Provide a database user and password for the database that you are connecting to. The query editor v2 stores your credentials in AWS Secrets Manager on your behalf.

Database

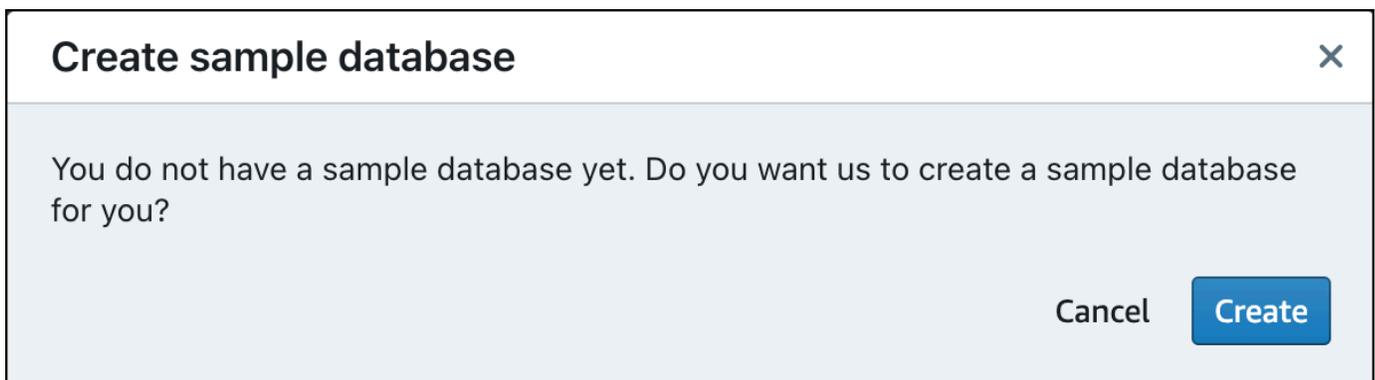
The database name must be 1-64 characters. Valid characters are lowercase alphanumeric characters.

Setelah terhubung, Anda dapat memilih untuk memuat data sampel dari Amazon Redshift Tanpa Server atau dari bucket Amazon S3.

- Di bawah workgroup default Amazon Redshift Serverless, perluas database `sample_data_dev`. Ada tiga skema sampel yang sesuai dengan tiga kumpulan data sampel yang dapat Anda muat ke dalam database Amazon Redshift Tanpa Server. Pilih kumpulan data sampel yang ingin dimuat, dan pilih Buka buku catatan sampel.



- Saat memuat data untuk pertama kalinya, editor kueri v2 akan meminta Anda untuk membuat database sampel. Pilih Buat.



Menjalankan kueri sampel

Setelah menyiapkan Amazon Redshift Tanpa Server, Anda dapat mulai menggunakan kumpulan data sampel di Amazon Redshift Tanpa Server. Amazon Redshift Serverless secara otomatis

memuat kumpulan data sampel, seperti kumpulan data ticket, dan Anda dapat segera melakukan kueri data.

- Setelah Amazon Redshift Serverless selesai memuat data sampel, semua kueri sampel dimuat di editor. Anda dapat memilih Jalankan semua untuk menjalankan semua kueri dari contoh buku catatan.

The screenshot shows the Amazon Redshift console interface. At the top, there's a header "Sales per event" with some navigation icons. Below that, there's a "Run" button and a "Limit 100" toggle. The main area contains a SQL query:

```

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

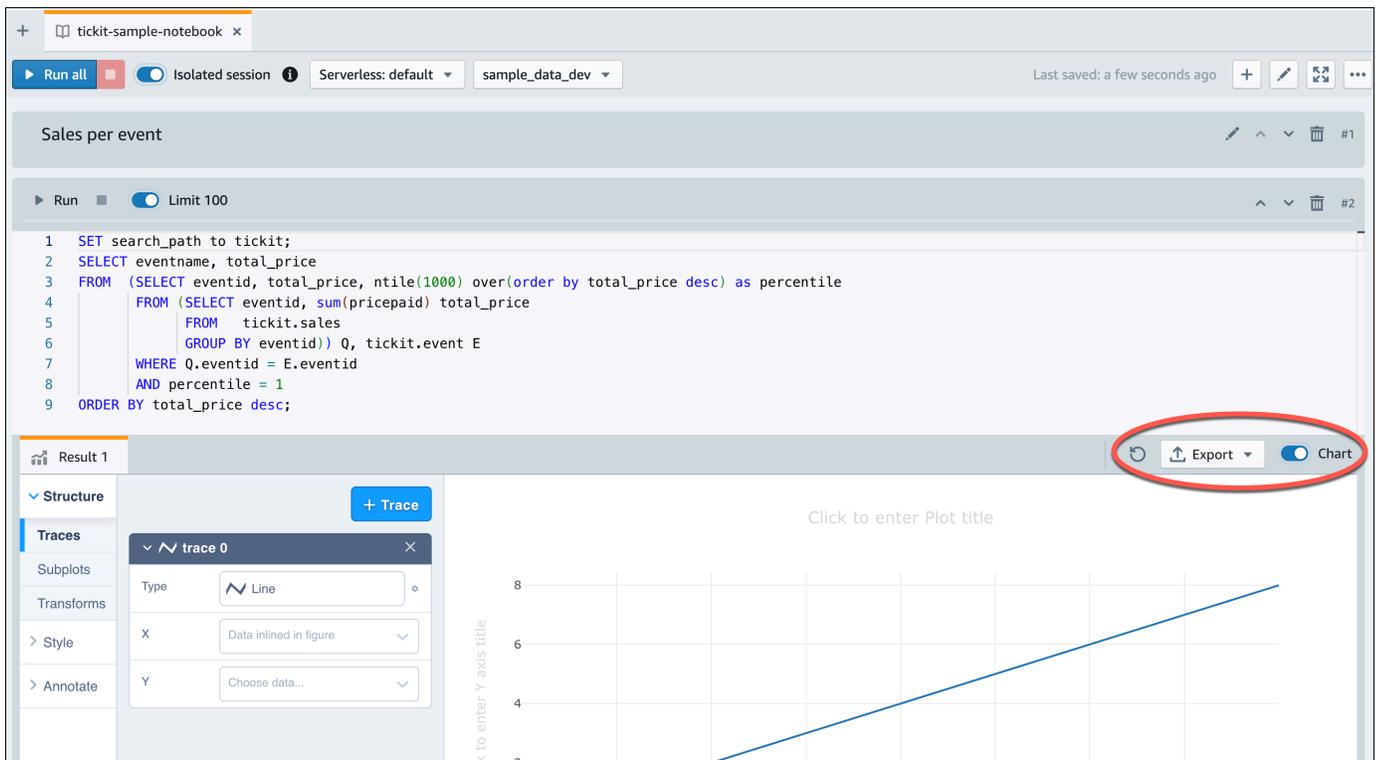
```

Below the query, there are two tabs: "Result 1" and "Result 2 (9)". The "Result 2 (9)" tab is active, showing a table with two columns: "eventname" and "total_price". The table contains 9 rows of data:

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

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

Anda juga dapat mengekspor hasil sebagai file JSON atau CSV atau melihat hasilnya dalam bagan.



The screenshot shows the Amazon Redshift console interface. At the top, there's a tab for 'ticket-sample-notebook'. Below that, there are controls for 'Run all', 'Isolated session', 'Serverless: default', and 'sample_data_dev'. The main area displays a SQL query for 'Sales per event' with a 'Limit 100' option. The query is as follows:

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

Below the query, the 'Result 1' section shows a table with columns for eventname and total_price. To the right of the table, there are buttons for 'Export' and 'Chart', which are circled in red. Below the table, there's a 'Traces' section with a 'trace 0' window showing a line chart. The chart has a Y-axis labeled 'Click to enter Y axis title' and a title 'Click to enter Plot title'. The chart shows a single blue line representing the data.

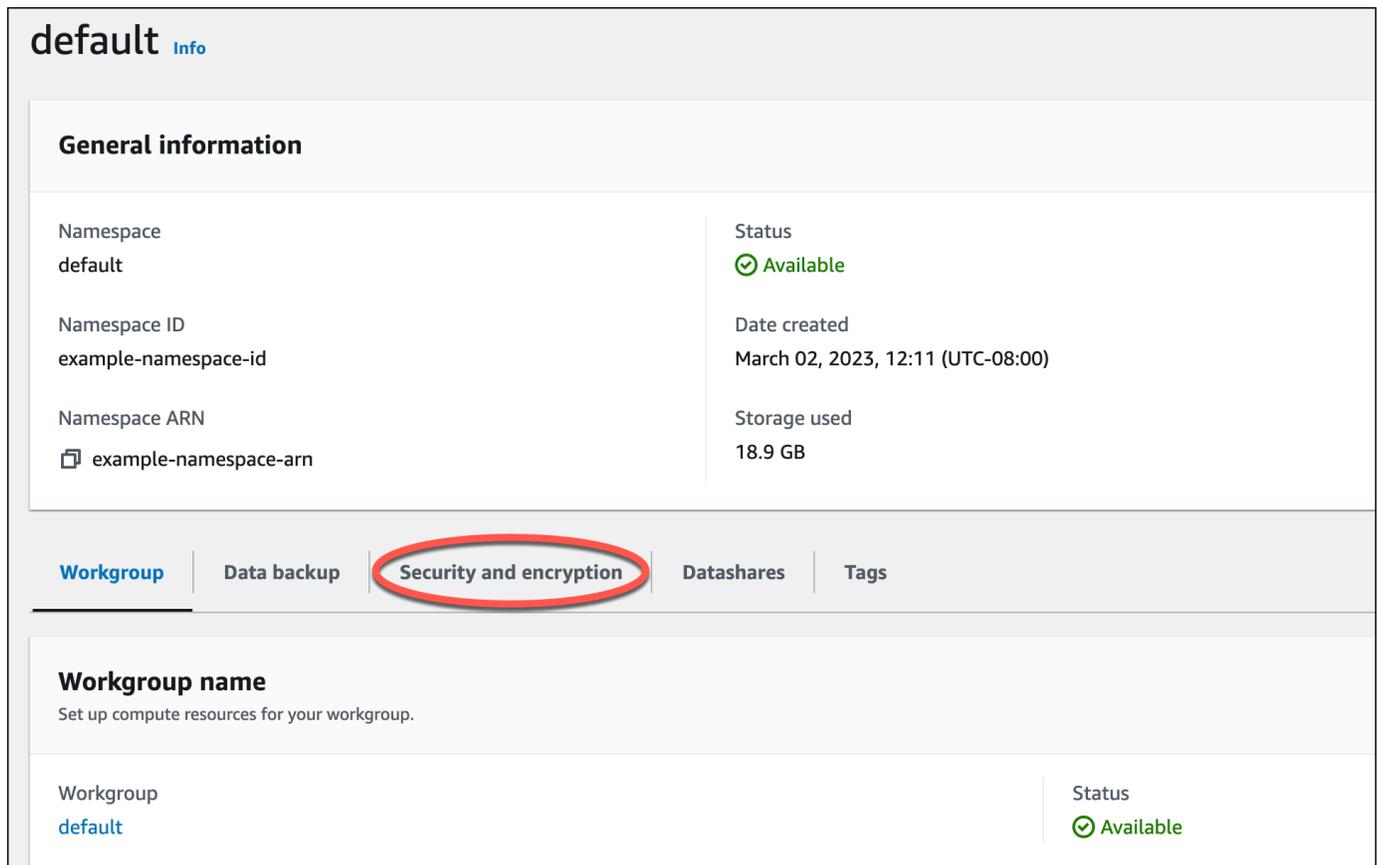
Anda juga dapat memuat data dari bucket Amazon S3. Lihat [the section called “Memuat data dari Amazon S3”](#) untuk mempelajari selengkapnya.

Memuat data dari Amazon S3

Setelah membuat gudang data, Anda dapat memuat data dari Amazon S3.

Pada titik ini, Anda memiliki database bernama dev. Selanjutnya, Anda akan membuat beberapa tabel dalam database, mengunggah data ke tabel, dan mencoba kueri. Demi kenyamanan Anda, data sampel yang Anda muat tersedia di bucket Amazon S3.

1. Sebelum Anda dapat memuat data dari Amazon S3, Anda harus terlebih dahulu membuat peran IAM dengan izin yang diperlukan dan melampirkannya ke namespace tanpa server Anda. Untuk melakukannya, pilih konfigurasi Namespace dari menu navigasi, lalu pilih Keamanan dan enkripsi. Kemudian, pilih Kelola peran IAM.



The screenshot displays the configuration page for a Redshift namespace named 'default'. The page is divided into several sections:

- Header:** 'default' with an 'Info' link.
- General information:** A table of key-value pairs:

Namespace	default	Status	Available
Namespace ID	example-namespace-id	Date created	March 02, 2023, 12:11 (UTC-08:00)
Namespace ARN	example-namespace-arn	Storage used	18.9 GB
- Navigation tabs:** 'Workgroup', 'Data backup', 'Security and encryption' (circled in red), 'Datashares', and 'Tags'.
- Workgroup name:** A section with the title 'Workgroup name' and the instruction 'Set up compute resources for your workgroup.' Below this, a table shows:

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

2. Perluas menu Kelola peran IAM, dan pilih Buat peran IAM.

Manage IAM roles

Permissions

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

Associated IAM roles (1)

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

Set default ▼

Manage IAM roles ▲



Search for associa

Associate IAM roles

Create IAM role

Remove IAM roles

or role type

< 1 >



IAM roles [↗](#)



Status



Role
type



- Pilih level akses bucket S3 yang ingin Anda berikan ke peran ini, dan pilih Buat peran IAM sebagai default.

Create the default IAM role ✕

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

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

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

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

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

4. Pilih Simpan perubahan. Anda sekarang dapat memuat data sampel dari Amazon S3.

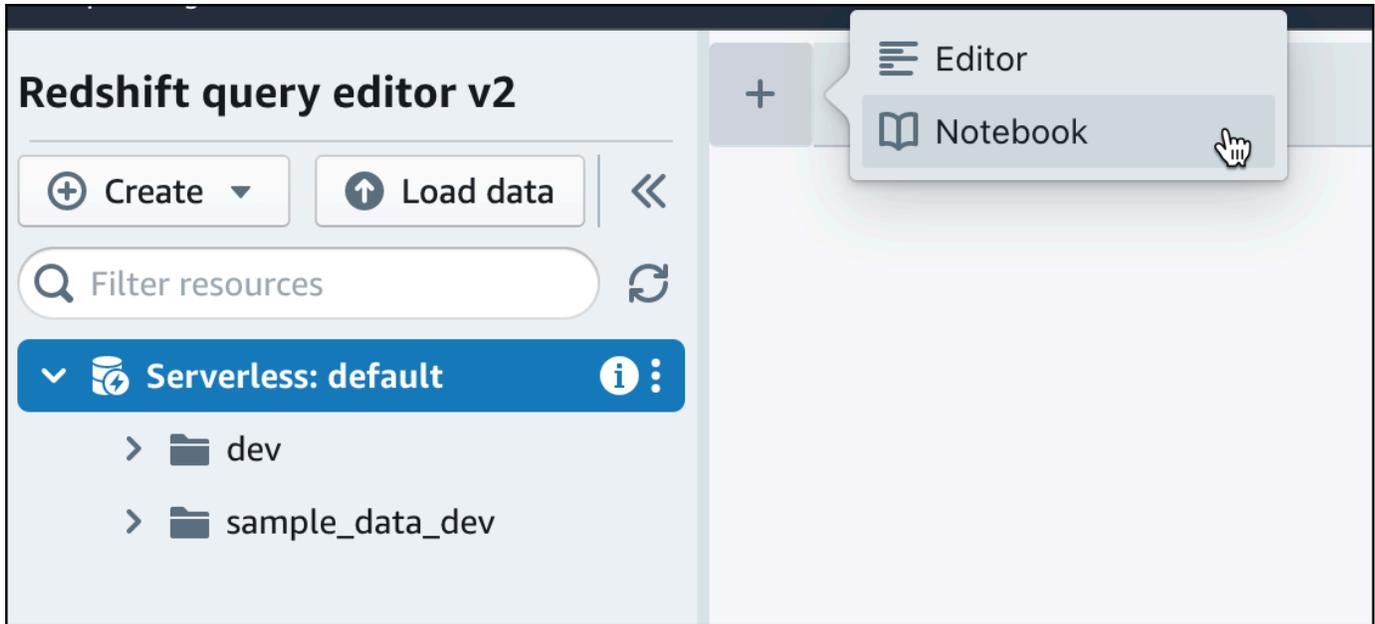
Langkah-langkah berikut menggunakan data dalam bucket Amazon Redshift S3 publik, tetapi Anda dapat mereplikasi langkah yang sama menggunakan bucket S3 dan perintah SQL Anda sendiri.

Muat data sampel dari Amazon S3

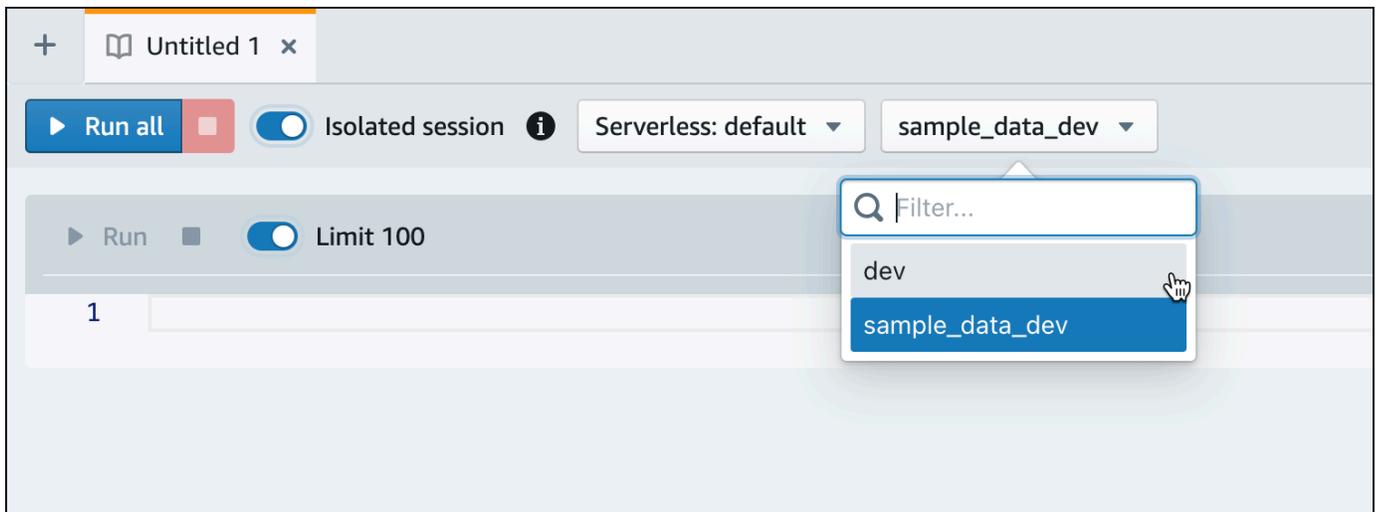
1. Di editor kueri v2, pilih



Tambah, lalu pilih Notebook untuk membuat buku catatan SQL baru.



2. Beralih ke dev database.



3. Buat tabel.

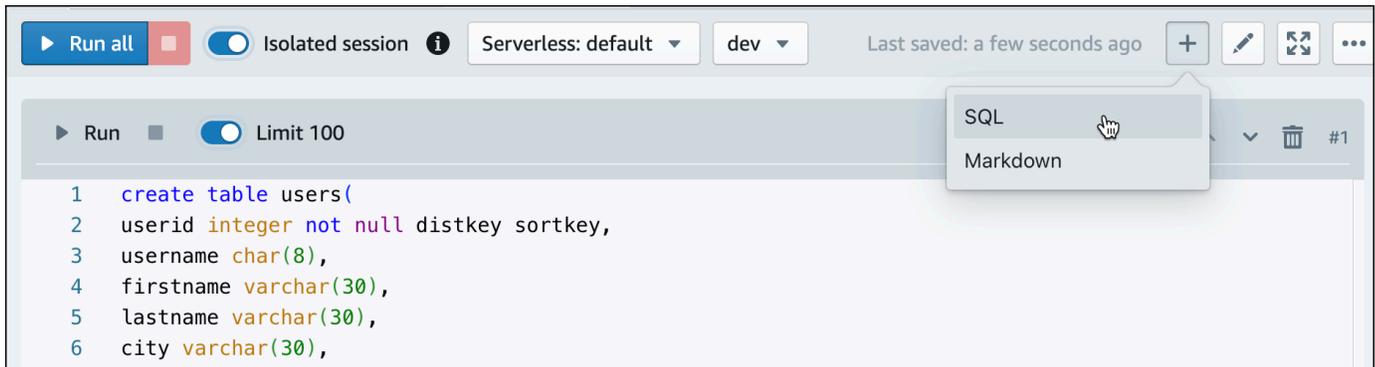
Jika Anda menggunakan query editor v2, copy dan jalankan berikut membuat pernyataan tabel untuk membuat tabel dalam dev database. Untuk informasi selengkapnya tentang sintaks, lihat [MEMBUAT TABEL](#) di Panduan Pengembang Database Amazon Redshift.

```
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. Di editor kueri v2, buat sel SQL baru di buku catatan Anda.



5. Sekarang gunakan perintah COPY di editor kueri v2 untuk memuat kumpulan data besar dari Amazon S3 atau Amazon DynamoDB ke Amazon Redshift. Untuk informasi selengkapnya tentang sintaks COPY, lihat [COPY](#) di Panduan Pengembang Database Amazon Redshift.

Anda dapat menjalankan perintah COPY dengan beberapa data sampel yang tersedia di bucket S3 publik. Jalankan perintah SQL berikut di editor query 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. Setelah memuat data, buat sel SQL lain di buku catatan Anda dan coba beberapa contoh kueri. Untuk informasi selengkapnya tentang bekerja dengan perintah SELECT, lihat [SELECT](#) di

Panduan Pengembang Amazon Redshift. Untuk memahami struktur dan skema data sampel, jelajahi menggunakan editor kueri 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;
```

Sekarang setelah Anda memuat data dan menjalankan beberapa kueri sampel, Anda dapat menjelajahi area lain di Amazon Redshift Tanpa Server. Lihat daftar berikut untuk mempelajari selengkapnya tentang cara menggunakan Amazon Redshift Tanpa Server.

- Anda dapat memuat data dari bucket Amazon S3. Lihat [Memuat data dari Amazon S3](#) untuk informasi selengkapnya.
- Anda dapat menggunakan editor kueri v2 untuk memuat data dari file terpisah karakter lokal yang lebih kecil dari 5 MB. Untuk informasi selengkapnya, lihat [Memuat data dari file lokal](#).
- Anda dapat terhubung ke Amazon Redshift Serverless dengan alat SQL pihak ketiga dengan driver JDBC dan ODBC. Lihat [Menghubungkan ke Amazon Redshift Tanpa Server](#) untuk informasi selengkapnya.
- Anda juga dapat menggunakan Amazon Redshift Data API untuk terhubung ke Amazon Redshift Tanpa Server. Lihat [Menggunakan Amazon Redshift Data API](#) untuk informasi selengkapnya.

- Anda dapat menggunakan data Anda di Amazon Redshift Serverless dengan Redshift ML untuk membuat model pembelajaran mesin dengan perintah CREATE MODEL. Lihat [Tutorial: Membangun model churn pelanggan](#) untuk mempelajari cara membuat model Redshift ML.
- Anda dapat menyalin data dari data lake Amazon S3 tanpa memuat data apa pun ke Amazon Redshift Tanpa Server. Lihat [Menyalin data lake](#) untuk informasi selengkapnya.

Cluster yang disediakan Amazon Redshift

Jika Anda adalah pengguna pertama kali Amazon Redshift, tutorial ini membantu Anda mulai menggunakan cluster Amazon Redshift. Untuk mulai menggunakan Redshift Serverless, lihat.

[Amazon Redshift Tanpa Server](#)

Setelah membuat dan masuk ke konsol yang disediakan Amazon Redshift, Anda dapat membuat dan mengelola objek Amazon Redshift, termasuk cluster, node, dan database. Anda juga dapat menjalankan kueri, melihat kueri, dan melakukan operasi SQL data definition language (DDL) dan data manipulation language (DHTML) lainnya dengan klien SQL.

Important

Cluster yang Anda sediakan untuk latihan ini berjalan di lingkungan hidup. Selama itu berjalan, itu menimbulkan biaya untuk Anda. Akun AWS Untuk informasi harga, lihat halaman [harga Amazon Redshift](#).

Untuk menghindari biaya yang tidak perlu, hapus klaster Anda setelah selesai melakukannya. Langkah terakhir dari tutorial menjelaskan bagaimana melakukannya.

[Masuk ke AWS Management Console dan buka konsol Amazon Redshift di https://console.aws.amazon.com/redshiftv2/.](https://console.aws.amazon.com/redshiftv2/)

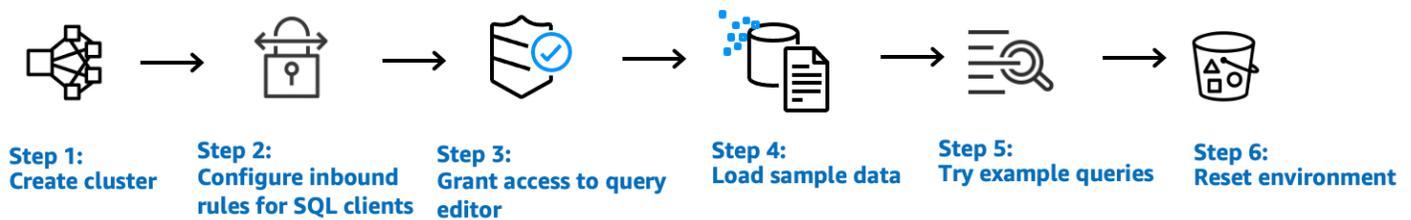
Kami menyarankan Anda memulai dengan membuka dasbor cluster yang disediakan untuk mulai menggunakan konsol Amazon Redshift.

Bergantung pada konfigurasi Anda, item berikut akan muncul di panel navigasi konsol yang disediakan Amazon Redshift:

- Redshift Tanpa Server — Akses dan analisis data tanpa perlu menyiapkan, menyetel, dan mengelola klaster yang disediakan Amazon Redshift.
- Dasbor kluster yang disediakan — Lihat daftar cluster di Anda Wilayah AWS, periksa metrik Cluster, dan ikhtisar kueri untuk wawasan data metrik (seperti pemanfaatan CPU) dan informasi kueri. Menggunakan ini dapat membantu Anda menentukan apakah data kinerja Anda tidak normal selama rentang waktu tertentu.
- Cluster — Lihat daftar cluster Anda dalam hal ini Wilayah AWS, pilih klaster untuk memulai kueri, atau melakukan tindakan terkait cluster. Anda juga dapat membuat cluster baru dari halaman ini.

- Editor kueri — Jalankan kueri pada database yang dihosting di klaster Amazon Redshift Anda. Kami merekomendasikan menggunakan Query editor v2 sebagai gantinya.
- Editor kueri v2 - Editor kueri Amazon Redshift v2 adalah aplikasi klien SQL berbasis web terpisah untuk membuat dan menjalankan kueri di gudang data Amazon Redshift Anda. Anda dapat memvisualisasikan hasil Anda dalam bagan dan berkolaborasi dengan membagikan pertanyaan Anda dengan orang lain di tim Anda.
- Kueri dan beban — Dapatkan informasi untuk referensi atau pemecahan masalah, seperti daftar kueri terbaru dan teks SQL untuk setiap kueri.
- Datashares — Sebagai administrator akun produsen, otorisasi akun konsumen untuk mengakses datashares atau memilih untuk tidak mengotorisasi akses. Untuk menggunakan datashare resmi, administrator akun konsumen dapat mengaitkan datashare dengan seluruh Akun AWS atau ruang nama klaster tertentu dalam akun. Administrator juga dapat menolak datashare.
- Integrasi nol-ETL — Kelola integrasi yang membuat data transaksional tersedia di Amazon Redshift setelah ditulis dalam sumber yang didukung.
- Koneksi IAM Identity Center — Konfigurasi koneksi antara Amazon Redshift dan IAM Identity Center.
- Konfigurasi - Hubungkan ke cluster Amazon Redshift dari alat klien SQL melalui koneksi Java Database Connectivity (JDBC) dan Open Database Connectivity (ODBC). Anda juga dapat menyiapkan titik akhir cloud pribadi virtual (VPC) yang dikelola Amazon Redshift. Melakukan hal itu menyediakan koneksi pribadi antara VPC berdasarkan layanan Amazon VPC yang berisi cluster dan VPC lain yang menjalankan alat klien.
- AWS Integrasi Mitra - Buat integrasi dengan AWS Mitra yang didukung.
- Penasihat — Dapatkan rekomendasi spesifik tentang perubahan yang dapat Anda lakukan pada klaster Amazon Redshift untuk memprioritaskan pengoptimalan Anda.
- AWS Marketplace — Dapatkan informasi tentang alat atau AWS layanan lain yang berfungsi dengan Amazon Redshift.
- Alarm — Buat alarm pada metrik klaster untuk melihat data kinerja dan melacak metrik selama periode waktu yang Anda tentukan.
- Acara — Lacak peristiwa dan dapatkan laporan tentang informasi seperti tanggal peristiwa terjadi, deskripsi, atau sumber acara.
- Apa yang baru - Lihat fitur Amazon Redshift baru dan pembaruan produk.

Dalam tutorial ini, Anda melakukan langkah-langkah berikut.



Topik

- [Mendaftar untuk AWS](#)
- [Tentukan aturan firewall](#)
- [Langkah 1: Buat contoh cluster Amazon Redshift](#)
- [Langkah 2: Konfigurasi aturan masuk untuk klien SQL](#)
- [Langkah 3: Berikan akses ke klien SQL dan jalankan kueri](#)
- [Langkah 4: Muat data dari Amazon S3 ke Amazon Redshift](#)
- [Langkah 5: Coba contoh kueri menggunakan editor kueri](#)
- [Langkah 6: Setel ulang lingkungan Anda](#)

Mendaftar untuk AWS

Jika Anda belum memilikinya Akun AWS, daftarkan untuk satu. Jika Anda sudah memiliki akun, Anda dapat melewati prasyarat ini dan menggunakan akun yang ada.

1. Buka <https://portal.aws.amazon.com/billing/signup>.
2. Ikuti petunjuk online.

Bagian dari prosedur pendaftaran melibatkan tindakan menerima panggilan telepon dan memasukkan kode verifikasi di keypad telepon.

Saat Anda mendaftar untuk sebuah Akun AWS, sebuah Pengguna root akun AWS dibuat. Pengguna root memiliki akses ke semua Layanan AWS dan sumber daya di akun. Sebagai praktik keamanan terbaik, tetapkan akses administratif ke pengguna, dan gunakan hanya pengguna root untuk melakukan [tugas yang memerlukan akses pengguna root](#).

Tentukan aturan firewall

Note

Tutorial ini mengasumsikan cluster Anda menggunakan port default 5439 dan Amazon Redshift query editor v2 dapat digunakan untuk menjalankan perintah SQL. Itu tidak masuk ke rincian tentang konfigurasi jaringan atau menyiapkan klien SQL yang mungkin diperlukan di lingkungan Anda.

Di beberapa lingkungan, Anda menentukan port saat meluncurkan klaster Amazon Redshift. Anda juga membuat aturan masuk masuk dalam grup keamanan untuk mengizinkan akses melalui port ke klaster Anda.

Jika komputer klien Anda berada di belakang firewall, pastikan Anda mengetahui port terbuka yang dapat Anda gunakan. Menggunakan port terbuka ini, Anda dapat terhubung ke cluster dari alat klien SQL dan menjalankan kueri. Jika Anda tidak tahu port terbuka, bekerja dengan seseorang yang memahami aturan firewall jaringan Anda untuk menentukan port terbuka di firewall Anda.

Meskipun Amazon Redshift menggunakan port 5439 secara default, koneksi tidak berfungsi jika port itu tidak terbuka di firewall Anda. Anda tidak dapat mengubah nomor port untuk klaster Amazon Redshift setelah dibuat. Jadi, pastikan Anda menentukan port terbuka yang berfungsi di lingkungan Anda selama proses peluncuran.

Langkah 1: Buat contoh cluster Amazon Redshift

Dalam tutorial ini, Anda berjalan melalui proses untuk membuat cluster Amazon Redshift dengan database. Kemudian Anda memuat kumpulan data dari Amazon S3 ke tabel database Anda. Anda dapat menggunakan contoh cluster ini untuk mengevaluasi layanan Amazon Redshift.

Sebelum Anda mulai menyiapkan cluster Amazon Redshift, pastikan Anda menyelesaikan prasyarat yang diperlukan seperti dan. [Mendaftar untuk AWS](#) [Tentukan aturan firewall](#)

Untuk setiap operasi yang mengakses data dari AWS sumber daya lain, klaster Anda memerlukan izin untuk mengakses sumber daya dan data pada sumber daya atas nama Anda. Contohnya adalah menggunakan perintah SQL COPY untuk memuat data dari Amazon Simple Storage Service (Amazon S3). Anda memberikan izin tersebut dengan menggunakan AWS Identity and Access

Management (IAM). Anda dapat melakukan ini melalui peran IAM yang Anda buat dan lampirkan ke cluster Anda. Untuk informasi selengkapnya tentang kredensial dan izin akses, lihat [Kredensial dan izin akses di Panduan Pengembang Database Amazon Redshift](#).

Untuk membuat cluster Amazon Redshift

1. [Masuk ke AWS Management Console dan buka konsol Amazon Redshift di https://console.aws.amazon.com/redshiftv2/.](https://console.aws.amazon.com/redshiftv2/)

 Important

Jika Anda menggunakan kredensial pengguna IAM, pastikan Anda memiliki izin yang diperlukan untuk melakukan operasi klaster. Untuk informasi selengkapnya, lihat [Keamanan di Amazon Redshift di Panduan Manajemen Pergeseran Merah Amazon](#).

2. Di AWS konsol, pilih Wilayah AWS tempat Anda ingin membuat cluster.
3. Pada menu navigasi, pilih Clusters, lalu pilih Create cluster. Halaman Create cluster muncul.
4. Di bagian Konfigurasi cluster, tentukan nilai untuk pengidentifikasi Cluster, tipe Node, dan Node:
 - Cluster identifier: Masukkan **examplecluster** untuk tutorial ini. Pengenal ini harus unik. Pengenal harus dari 1-63 karakter menggunakan karakter yang valid a—z (hanya huruf kecil) dan - (tanda hubung).
 - Pilih salah satu metode berikut untuk mengukur cluster Anda:

 Note

Langkah berikut mengasumsikan Wilayah AWS yang mendukung tipe node RA3. Untuk daftar tipe node RA3 Wilayah AWS yang mendukung, lihat [Ringkasan jenis node RA3](#) di Panduan Manajemen Pergeseran Merah Amazon. Untuk mempelajari lebih lanjut tentang spesifikasi node untuk setiap jenis dan ukuran node, lihat [Detail tipe Node](#).

- Jika Anda tidak tahu seberapa besar ukuran cluster Anda, pilih Bantu saya pilih. Melakukan hal itu membuka kalkulator ukuran yang menanyakan pertanyaan tentang ukuran dan karakteristik kueri data yang Anda rencanakan untuk disimpan di gudang data Anda.

Jika Anda mengetahui ukuran cluster Anda yang diperlukan (yaitu, jenis node dan jumlah node), pilih yang akan saya pilih. Kemudian pilih jenis Node dan jumlah Node untuk mengukur cluster Anda.

Untuk tutorial ini, pilih `ra3.4xlarge` untuk tipe Node dan 2 untuk Jumlah node.

Jika pilihan untuk konfigurasi AZ tersedia, pilih Single-AZ.

- Untuk menggunakan kumpulan data sampel yang disediakan Amazon Redshift, dalam Data sampel, pilih Muat data sampel. Amazon Redshift memuat kumpulan data sampel Tckit ke database dan skema `defaultdev. public`
5. Di bagian konfigurasi Database, tentukan nilai untuk nama pengguna Admin. Untuk kata sandi Admin, pilih dari opsi berikut:
- Buat kata sandi — Gunakan kata sandi yang dihasilkan oleh Amazon Redshift.
 - Tambahkan kata sandi admin secara manual — Gunakan kata sandi Anda sendiri.
 - Kelola kredensi admin di AWS Secrets Manager — Amazon Redshift AWS Secrets Manager menggunakan untuk membuat dan mengelola kata sandi admin Anda. Menggunakan AWS Secrets Manager untuk menghasilkan dan mengelola rahasia kata sandi Anda menimbulkan biaya. Untuk informasi tentang AWS Secrets Manager harga, lihat [AWS Secrets Manager Harga](#).

Untuk tutorial ini, gunakan nilai-nilai ini:

- Nama pengguna admin: `Enterawsuser`.
 - Kata sandi pengguna admin: Masukkan nilai untuk kata sandi.
6. Untuk tutorial ini, buat peran IAM dan tetapkan sebagai default untuk cluster Anda, seperti yang dijelaskan berikut. Hanya ada satu set peran IAM default untuk sebuah cluster.
- a. Di bawah Izin klaster, untuk Kelola peran IAM, pilih Buat peran IAM.
 - b. Tentukan bucket Amazon S3 agar peran IAM dapat diakses dengan salah satu metode berikut:
 - Pilih Tidak ada bucket Amazon S3 tambahan untuk mengizinkan peran IAM yang dibuat hanya mengakses bucket Amazon S3 yang diberi nama. `redshift`
 - Pilih bucket Amazon S3 apa pun untuk memungkinkan peran IAM yang dibuat mengakses semua bucket Amazon S3.

- Pilih Bucket Amazon S3 Spesifik untuk menentukan satu atau beberapa bucket Amazon S3 agar peran IAM yang dibuat dapat diakses. Kemudian pilih satu atau lebih ember Amazon S3 dari tabel.
- c. Pilih Buat peran IAM sebagai default. Amazon Redshift secara otomatis membuat dan menetapkan peran IAM sebagai default untuk kluster Anda.

Karena Anda membuat peran IAM Anda dari konsol, itu memiliki `AmazonRedshiftAllCommandsFullAccess` kebijakan yang dilampirkan. Ini memungkinkan Amazon Redshift untuk menyalin, memuat, menanyakan, dan menganalisis data dari sumber daya Amazon di akun IAM Anda.

Untuk informasi tentang cara mengelola peran IAM default untuk kluster, lihat [Membuat peran IAM sebagai default untuk Amazon Redshift](#) di Panduan Manajemen Pergeseran Merah Amazon.

7. (Opsional) Di bagian Konfigurasi tambahan, matikan Gunakan default untuk mengubah pengaturan Jaringan dan keamanan, konfigurasi Database, Pemeliharaan, Pemantauan, dan Pencadangan.

Dalam beberapa kasus, Anda dapat membuat kluster dengan opsi Muat data sampel dan ingin mengaktifkan perutean Amazon VPC yang disempurnakan. Jika demikian, cluster di virtual private cloud (VPC) Anda memerlukan akses ke endpoint Amazon S3 agar data dapat dimuat.

Untuk membuat cluster dapat diakses publik, Anda dapat melakukan salah satu dari dua hal. Anda dapat mengonfigurasi alamat terjemahan alamat jaringan (NAT) di VPC Anda agar cluster dapat mengakses internet. Atau Anda dapat mengonfigurasi titik akhir VPC Amazon S3 di VPC Anda. Untuk informasi selengkapnya tentang perutean VPC Amazon yang disempurnakan, lihat Perutean [VPC Amazon yang disempurnakan di Panduan Manajemen Pergeseran](#) Merah Amazon.

8. Pilih Buat kluster. Tunggu kluster Anda dibuat dengan `Available` status di halaman Clusters.

Langkah 2: Konfigurasi aturan masuk untuk klien SQL

Note

Kami menyarankan Anda melewati langkah ini dan mengakses kluster Anda menggunakan editor kueri Amazon Redshift v2.

Kemudian dalam tutorial ini, Anda mengakses cluster Anda dari dalam virtual private cloud (VPC) berdasarkan layanan Amazon VPC. Namun, jika Anda menggunakan klien SQL dari luar firewall Anda untuk mengakses cluster, pastikan Anda memberikan akses masuk.

Untuk memeriksa firewall Anda dan memberikan akses masuk ke cluster Anda

1. Periksa aturan firewall Anda jika kluster Anda perlu diakses dari luar firewall. Misalnya, klien Anda mungkin berupa instans Amazon Elastic Compute Cloud (Amazon EC2) atau komputer eksternal.

Untuk informasi selengkapnya tentang aturan firewall, lihat [Aturan grup keamanan](#) di Panduan Pengguna Amazon EC2.

2. Untuk mengakses dari klien eksternal Amazon EC2, tambahkan aturan masuk ke grup keamanan yang dilampirkan ke kluster Anda yang memungkinkan lalu lintas masuk. Anda menambahkan aturan grup keamanan Amazon EC2 di konsol Amazon EC2. Misalnya, CIDR/IP 192.0.2.0/24 memungkinkan klien dalam rentang alamat IP tersebut untuk terhubung ke cluster Anda. Cari tahu CIDR/IP yang benar untuk lingkungan Anda.

Langkah 3: Berikan akses ke klien SQL dan jalankan kueri

Untuk menanyakan database yang dihosting oleh cluster Amazon Redshift Anda, Anda memiliki beberapa opsi untuk klien SQL. Ini termasuk:

- Hubungkan ke kluster Anda dan jalankan kueri menggunakan editor kueri Amazon Redshift v2.

Jika Anda menggunakan editor kueri v2, Anda tidak perlu mengunduh dan menyiapkan aplikasi klien SQL. Anda meluncurkan editor kueri Amazon Redshift v2 dari konsol Amazon Redshift.

- Connect ke cluster Anda menggunakan RSQL. Untuk informasi selengkapnya, lihat [Menghubungkan dengan Amazon Redshift RSQL di Panduan Manajemen](#) Amazon Redshift.

- Connect ke cluster Anda melalui alat klien SQL, seperti SQL Workbench/J. Untuk informasi selengkapnya, lihat [Connect to your cluster menggunakan SQL Workbench/J di Amazon Redshift Management Guide](#).

Tutorial ini menggunakan editor kueri Amazon Redshift v2 sebagai cara mudah untuk menjalankan kueri pada database yang dihosting oleh cluster Amazon Redshift Anda. Setelah membuat cluster Anda, Anda dapat langsung menjalankan kueri. Untuk detail tentang pertimbangan saat menggunakan editor kueri Amazon Redshift v2, [lihat Pertimbangan saat bekerja dengan editor kueri v2](#) di Panduan Manajemen Amazon Redshift.

Memberikan akses ke editor kueri v2

Pertama kali administrator mengonfigurasi editor kueri v2 untuk Anda Akun AWS, mereka memilih AWS KMS key yang digunakan untuk mengenkripsi sumber daya editor kueri v2. Sumber daya editor kueri Amazon Redshift v2 mencakup kueri, buku catatan, dan bagan yang disimpan. Secara default, kunci yang AWS dimiliki digunakan untuk mengenkripsi sumber daya. Atau, administrator dapat menggunakan kunci yang dikelola pelanggan dengan memilih Amazon Resource Name (ARN) untuk kunci di halaman konfigurasi. Setelah Anda mengonfigurasi akun, pengaturan AWS KMS enkripsi tidak dapat diubah. Untuk informasi selengkapnya, lihat [Mengonfigurasi Anda Akun AWS](#) di Panduan Manajemen Pergeseran Merah Amazon.

Untuk mengakses editor kueri v2, Anda memerlukan izin. Administrator dapat melampirkan salah satu kebijakan AWS terkelola untuk editor kueri Amazon Redshift v2 ke peran IAM atau pengguna untuk memberikan izin. Kebijakan AWS terkelola ini ditulis dengan opsi berbeda yang mengontrol bagaimana sumber daya penandaan memungkinkan berbagi kueri. Anda dapat menggunakan konsol IAM (<https://console.aws.amazon.com/iam/>) untuk melampirkan kebijakan IAM. Untuk informasi selengkapnya tentang kebijakan ini, lihat [Mengakses editor kueri v2](#) di Panduan Manajemen Amazon Redshift.

Anda juga dapat membuat kebijakan sendiri berdasarkan izin yang diizinkan dan ditolak dalam kebijakan terkelola yang disediakan. Jika Anda menggunakan editor kebijakan konsol IAM untuk membuat kebijakan Anda sendiri, pilih SQL Workbench sebagai layanan yang Anda buat kebijakan di editor visual. Query editor v2 menggunakan nama layanan AWS SQL Workbench di editor visual dan IAM Policy Simulator.

Untuk informasi selengkapnya, lihat [Bekerja dengan editor kueri v2](#) di Panduan Manajemen Amazon Redshift.

Langkah 4: Muat data dari Amazon S3 ke Amazon Redshift

Setelah membuat cluster, Anda dapat memuat data dari Amazon S3 ke tabel database Anda. Ada beberapa cara untuk memuat data dari Amazon S3.

- Anda dapat menggunakan klien SQL untuk menjalankan perintah SQL CREATE TABLE untuk membuat tabel di database Anda dan kemudian menggunakan perintah SQL COPY untuk memuat data dari Amazon S3. Editor kueri Amazon Redshift v2 adalah klien SQL.
- Anda dapat menggunakan wizard pemuatan Amazon Redshift query editor v2.

Tutorial ini pertama kali menunjukkan cara menggunakan Amazon Redshift query editor v2 untuk menjalankan perintah SQL untuk CREATE tabel dan COPY data. Luncurkan Query editor v2 dari panel navigasi konsol Amazon Redshift. Dalam query editor v2 buat koneksi ke `examplecluster` cluster dan database bernama `dev` dengan pengguna admin `andaawsuser`. Untuk tutorial ini pilih Kredensi sementara menggunakan nama pengguna database saat Anda membuat koneksi. Untuk detail tentang menggunakan editor kueri Amazon Redshift v2, lihat [Menyambungkan ke database Amazon Redshift](#) di Panduan Manajemen Pergeseran Merah Amazon.

Memuat data dari Amazon S3 menggunakan perintah SQL

Pada panel editor kueri editor kueri v2, konfirmasi bahwa Anda terhubung ke `examplecluster` cluster dan `dev` database. Selanjutnya, buat tabel dalam database dan muat data ke tabel. Untuk tutorial ini, data yang Anda muat tersedia di bucket Amazon S3 yang dapat diakses dari banyak orang. Wilayah AWS

Prosedur berikut membuat tabel dan memuat data dari bucket Amazon S3 publik.

Gunakan Amazon Redshift query editor v2 untuk menyalin dan menjalankan pernyataan create table berikut untuk membuat tabel dalam `public` skema database. `dev` Untuk informasi selengkapnya tentang sintaks, lihat [MEMBUAT TABEL](#) di Panduan Pengembang Database Amazon Redshift.

Untuk membuat dan memuat data menggunakan klien SQL seperti editor kueri v2

1. Jalankan perintah SQL berikut untuk CREATE `sales` tabel.

```
drop table if exists sales;  
create table sales(  
  id integer,  
  name varchar(100),  
  price decimal(10,2),  
  quantity integer,  
  total decimal(10,2),  
  primary key (id)
```

```
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,  
qtsold smallint not null,  
pricepaid decimal(8,2),  
commission decimal(8,2),  
saletime timestamp);
```

2. Jalankan perintah SQL berikut untuk CREATE date tabel.

```
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. Muat sales tabel dari Amazon S3 menggunakan perintah COPY.

Note

Sebaiknya gunakan perintah COPY untuk memuat kumpulan data besar ke Amazon Redshift dari Amazon S3. Untuk informasi selengkapnya tentang sintaks COPY, lihat [COPY](#) di Panduan Pengembang Database Amazon Redshift.

Berikan autentikasi untuk klaster Anda untuk mengakses Amazon S3 atas nama Anda untuk memuat data sampel. Anda memberikan autentikasi dengan mereferensikan peran IAM yang Anda buat dan tetapkan sebagai klaster saat Anda memilih Create IAM role sebagai default saat Anda membuat klaster. default

Muat sales tabel menggunakan perintah SQL berikut. Anda secara opsional dapat mengunduh dan melihat dari Amazon S3 [data sumber untuk sales tabel](#).

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

4. Muat data tabel menggunakan perintah SQL berikut. Anda secara opsional dapat mengunduh dan melihat dari Amazon S3 [data sumber untuk data tabel](#).

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

Memuat data dari Amazon S3 menggunakan editor kueri v2

Editor kueri v2 menyederhanakan pemuatan data saat menggunakan wizard Load data. Perintah COPY yang dihasilkan dan digunakan dalam editor kueri v2 Load data wizard mendukung banyak parameter yang tersedia untuk sintaks perintah COPY untuk memuat data dari Amazon S3. Untuk informasi tentang perintah COPY dan opsinya yang digunakan untuk menyalin beban dari Amazon S3, lihat [COPY dari Amazon Simple Storage Service](#) di Panduan Pengembang Database Amazon Redshift.

Untuk memuat data Anda sendiri dari Amazon S3 ke Amazon Redshift, Amazon Redshift memerlukan peran IAM yang memiliki hak istimewa yang diperlukan untuk memuat data dari bucket Amazon S3 yang ditentukan.

Untuk tutorial ini, pertama buka Amazon Redshift query editor v2 dan sambungkan ke database. Selanjutnya, buat tabel yang akan berisi data yang dimuat. Kemudian muat data Anda sendiri dari Amazon S3 ke Amazon Redshift. Untuk informasi selengkapnya tentang cara bekerja dengan editor kueri v2, lihat [Memuat data ke dalam database](#) di Panduan Manajemen Pergeseran Merah Amazon.

Membuat data TICKIT di klaster Anda

Anda dapat membuat set lengkap tabel TICKIT dan memuat data ke dalam klaster Anda dengan cara berikut:

- Saat membuat cluster di konsol Amazon Redshift, Anda memiliki opsi untuk memuat data TICKIT sampel secara bersamaan. Di konsol Amazon Redshift, pilih Clusters, Create cluster. Di bagian Data sampel, pilih Muat data sampel Amazon Redshift memuat kumpulan data sampelnya ke dev database cluster Amazon Redshift Anda secara otomatis selama pembuatan klaster.
- Dengan editor kueri Amazon Redshift v2, Anda dapat memuat data TICKIT ke dalam database sampel bernama `sample_data_dev`. Arahkan ke database `sample_data_dev`, lalu ke `tickit`, dan dari menu konteks pilih Buka contoh buku catatan. Editor kueri Amazon Redshift v2 membuat database sampel bersama dengan notebook contoh bernama `tickit-sample-notebook`. Anda dapat menjalankan buku catatan ini untuk melakukan kueri data dalam database sampel.

Untuk melihat detail tentang data TICKIT, lihat [Contoh database di Panduan Pengembang Database Amazon Redshift](#).

Langkah 5: Coba contoh kueri menggunakan editor kueri

Untuk menyiapkan dan menggunakan editor kueri Amazon Redshift v2 untuk membuat kueri database, lihat [Bekerja dengan editor kueri v2 di Panduan Manajemen Amazon Redshift](#).

Sekarang, coba beberapa contoh query, seperti yang ditunjukkan berikut. Untuk informasi selengkapnya tentang bekerja dengan perintah SELECT, lihat [SELECT](#) di Panduan Pengembang Database 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;
```

Langkah 6: Setel ulang lingkungan Anda

Pada langkah sebelumnya, Anda telah berhasil membuat kluster Amazon Redshift, memuat data ke dalam tabel, dan data kueri menggunakan klien SQL seperti editor kueri Amazon Redshift v2.

Ketika Anda telah menyelesaikan tutorial ini, kami sarankan Anda mengatur ulang lingkungan Anda ke keadaan sebelumnya dengan menghapus cluster sampel Anda. Anda terus dikenakan biaya untuk layanan Amazon Redshift hingga Anda menghapus kluster.

Namun, Anda mungkin ingin menjaga cluster sampel tetap berjalan jika Anda ingin mencoba tugas di panduan Amazon Redshift lainnya atau tugas yang dijelaskan. [Tugas database umum](#)

Untuk menghapus kluster

1. [Masuk ke AWS Management Console dan buka konsol Amazon Redshift di https://console.aws.amazon.com/redshiftv2/](https://console.aws.amazon.com/redshiftv2/).
2. Pada menu navigasi, pilih Cluster untuk menampilkan daftar cluster Anda.
3. Pilih `examplecluster` cluster. Untuk Tindakan, pilih Hapus. Halaman Delete cluster muncul.
4. Konfirmasikan kluster yang akan dihapus, lalu pilih Hapus kluster.

Pada halaman daftar cluster, status cluster diperbarui saat cluster dihapus.

Setelah Anda menyelesaikan tutorial ini, Anda dapat menemukan informasi lebih lanjut tentang Amazon Redshift dan langkah selanjutnya. [Sumber daya tambahan](#)

Tugas database umum

Baik gudang data Redshift Tanpa Server dan gudang data cluster Amazon Redshift berisi database. Setelah meluncurkan gudang data, Anda dapat mengelola sebagian besar tindakan database menggunakan perintah SQL. Dengan beberapa pengecualian, fungsionalitas dan sintaks SQL sama untuk semua database Amazon Redshift. Untuk detail perintah SQL yang tersedia dengan Amazon Redshift, [lihat perintah SQL](#) di Panduan Pengembang Database Amazon Redshift.

Saat Anda membuat gudang data, di sebagian besar skenario, Amazon Redshift juga membuat database defaultdev. Setelah Anda terhubung ke dev database, Anda dapat membuat database lain.

Langkah-langkah berikut berjalan melalui tugas-tugas database umum saat bekerja dengan database Amazon Redshift. Langkah-langkah dimulai dengan membuat database dan jika Anda melanjutkan ke langkah terakhir Anda dapat menghapus semua sumber daya yang Anda buat dengan menjatuhkan database.

Contoh di bagian ini mengasumsikan hal berikut:

- Anda telah membuat gudang data Amazon Redshift.
- Anda telah membuat koneksi ke gudang data dari alat klien SQL Anda, seperti editor kueri Amazon Redshift v2. Untuk informasi selengkapnya tentang editor kueri v2, lihat [Menanyakan database menggunakan editor kueri Amazon Redshift v2](#) di Panduan Manajemen Amazon Redshift.

Menghubungkan ke gudang data Amazon Redshift

Untuk terhubung ke cluster Amazon Redshift, dari halaman Clusters konsol Amazon Redshift, perluas cluster Connect to Amazon Redshift dan lakukan salah satu hal berikut:

- Pilih Data kueri untuk menggunakan editor kueri v2 untuk menjalankan kueri pada database yang dihosting oleh klaster Amazon Redshift Anda. Setelah membuat cluster Anda, Anda dapat langsung menjalankan kueri dengan menggunakan editor kueri v2.

Untuk informasi selengkapnya, lihat [Menanyakan database menggunakan editor kueri Amazon Redshift v2](#) di Panduan Manajemen Pergeseran Merah Amazon.

- Di Bekerja dengan alat klien Anda, pilih klaster Anda dan sambungkan ke Amazon Redshift dari alat klien Anda menggunakan driver JDBC atau ODBC dengan menyalin URL driver JDBC atau ODBC. Gunakan URL ini dari komputer klien atau instance Anda. Kode aplikasi Anda untuk

menggunakan operasi API akses data JDBC atau ODBC, atau gunakan alat klien SQL yang mendukung JDBC atau ODBC.

Untuk informasi selengkapnya tentang cara menemukan string koneksi cluster, lihat [Menemukan string koneksi cluster Anda](#).

- Jika alat klien SQL Anda memerlukan driver, Anda dapat Memilih driver JDBC atau ODBC Anda untuk mengunduh driver khusus sistem operasi untuk terhubung ke Amazon Redshift dari alat klien Anda.

Untuk informasi selengkapnya tentang cara menginstal driver yang sesuai untuk klien SQL Anda, lihat [Mengonfigurasi koneksi driver JDBC versi 2.0](#).

Untuk informasi selengkapnya tentang cara mengonfigurasi koneksi ODBC, lihat [Mengonfigurasi koneksi ODBC](#).

Untuk terhubung ke gudang data Redshift Tanpa Server, dari halaman dasbor Tanpa Server konsol Amazon Redshift, lakukan salah satu hal berikut:

- Gunakan editor kueri Amazon Redshift v2 untuk menjalankan kueri pada database yang dihosting oleh gudang data Redshift Tanpa Server Anda. Setelah membuat gudang data Anda, Anda dapat langsung menjalankan kueri dengan menggunakan editor kueri v2.

Untuk informasi selengkapnya, lihat [Menanyakan database menggunakan editor kueri Amazon Redshift v2](#).

- Sambungkan ke Amazon Redshift dari alat klien Anda menggunakan driver JDBC atau ODBC dengan menyalin URL driver JDBC atau ODBC.

Untuk bekerja dengan data di gudang data Anda, Anda memerlukan driver JDBC atau ODBC untuk konektivitas dari komputer klien atau instance Anda. Kode aplikasi Anda untuk menggunakan operasi API akses data JDBC atau ODBC, atau gunakan alat klien SQL yang mendukung JDBC atau ODBC.

Untuk informasi selengkapnya tentang cara menemukan string koneksi, lihat [Menyambung ke Redshift Tanpa Server](#) di Panduan Manajemen Pergeseran Merah Amazon.

Langkah 1: Buat database

Setelah Anda memverifikasi bahwa gudang data Anda aktif dan berjalan, Anda dapat membuat database. Database ini adalah tempat Anda benar-benar membuat tabel, memuat data, dan menjalankan kueri. Gudang data dapat meng-host beberapa database. Misalnya, Anda dapat memiliki database untuk data penjualan bernama SALESDB dan database untuk data pesanan yang disebutkan ORDERSDB di gudang data yang sama.

Untuk membuat database bernama **SALESDB**, jalankan perintah berikut di alat klien SQL Anda.

```
CREATE DATABASE salesdb;
```

Note

Setelah menjalankan perintah, pastikan untuk menyegarkan daftar alat klien SQL Anda dari objek di gudang data Anda untuk melihat yang baru `salesdb`.

Untuk latihan ini, terima default. Untuk informasi tentang opsi perintah lainnya, lihat [MEMBUAT DATABASE](#) di Panduan Pengembang Database Amazon Redshift. Untuk menghapus database dan isinya, lihat [DROP DATABASE](#) di Panduan Pengembang Database Amazon Redshift.

Setelah Anda membuat database SALESDB, Anda dapat terhubung ke database baru dari klien SQL Anda. Gunakan parameter koneksi yang sama seperti yang Anda gunakan untuk koneksi Anda saat ini, tetapi ubah nama database menjadi SALESDB.

Langkah 2: Buat pengguna

Secara default, hanya pengguna admin yang Anda buat saat meluncurkan gudang data yang memiliki akses ke database default di gudang data. Untuk memberikan akses kepada pengguna lain, buat satu atau beberapa akun. Akun pengguna database bersifat global di semua database dalam gudang data, dan bukan per database individu.

Gunakan perintah CREATE USER untuk membuat pengguna baru. Saat Anda membuat pengguna baru, Anda menentukan nama pengguna baru dan kata sandi. Kami menyarankan Anda menentukan kata sandi untuk pengguna. Itu harus memiliki 8-64 karakter, dan harus mencakup setidaknya satu huruf besar, satu huruf kecil, dan satu angka.

Misalnya, untuk membuat pengguna bernama **GUEST** dengan kata sandi **ABCd4321**, jalankan perintah berikut.

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

Untuk terhubung ke SALESDB database sebagai GUEST pengguna, gunakan kata sandi yang sama saat Anda membuat pengguna, seperti ABCd4321.

Untuk informasi tentang opsi perintah lainnya, lihat [MEMBUAT PENGGUNA di Panduan Pengembang Database Amazon Redshift](#).

Langkah 3: Buat skema

Setelah Anda membuat database baru, Anda dapat membuat skema baru dalam database saat ini. Skema adalah namespace yang berisi objek database bernama seperti tabel, tampilan, dan fungsi yang ditentukan pengguna (UDF). Sebuah database dapat berisi satu atau beberapa skema, dan setiap skema hanya milik satu database. Dua skema dapat memiliki objek berbeda yang memiliki nama yang sama.

Anda dapat membuat beberapa skema dalam database yang sama untuk mengatur data dengan cara yang Anda inginkan atau untuk mengelompokkan data Anda secara fungsional. Misalnya, Anda dapat membuat skema untuk menyimpan semua data pementasan dan skema lain untuk menyimpan semua tabel pelaporan. Anda juga dapat membuat skema yang berbeda untuk menyimpan data yang relevan dengan grup bisnis yang berbeda yang berada dalam database yang sama. Setiap skema dapat menyimpan objek database yang berbeda, seperti tabel, tampilan, dan fungsi yang ditentukan pengguna (UDF). Selain itu, Anda dapat membuat skema dengan klausa OTORISASI. Klausul ini memberikan kepemilikan kepada pengguna tertentu atau menetapkan kuota pada jumlah maksimum ruang disk yang dapat digunakan skema tertentu.

Amazon Redshift secara otomatis membuat skema yang dipanggil `public` untuk setiap database baru. Bila Anda tidak menentukan nama skema saat membuat objek database, objek masuk ke `public` skema.

Untuk mengakses objek dalam skema, kualifikasi objek dengan menggunakan notasi `schema_name.table_name`. Nama skema yang memenuhi syarat terdiri dari nama skema dan nama tabel yang dipisahkan oleh titik. Misalnya, Anda mungkin memiliki `sales` skema yang memiliki `price` tabel dan `inventory` skema yang juga memiliki `price` tabel. Ketika Anda merujuk ke `price` tabel, Anda harus memenuhi syarat sebagai `sales.price` atau `inventory.price`.

Contoh berikut membuat skema bernama **SALES** untuk pengguna **GUEST**.

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

Untuk informasi tentang opsi perintah lainnya, lihat [MEMBUAT SKEMA di Panduan Pengembang Database Amazon Redshift](#).

Untuk melihat daftar skema dalam database Anda, jalankan perintah berikut.

```
select * from pg_namespace;
```

Outputnya akan terlihat serupa dengan yang berikut ini:

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}

Untuk informasi selengkapnya tentang cara menanyakan tabel katalog, lihat [Menanyakan tabel katalog di Panduan Pengembang Database Amazon Redshift](#).

Gunakan pernyataan **GRANT** untuk memberikan izin kepada pengguna untuk skema.

Contoh berikut memberikan hak istimewa kepada **GUEST** pengguna untuk memilih data dari semua tabel atau tampilan dalam **SALES** skema menggunakan pernyataan **SELECT**.

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

Contoh berikut memberikan semua hak istimewa yang tersedia sekaligus kepada pengguna. **GUEST**

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

Langkah 4: Buat tabel

Setelah Anda membuat database baru, buat tabel untuk menyimpan data Anda. Tentukan informasi kolom saat Anda membuat tabel.

Misalnya, untuk membuat tabel bernama **DEMO**, jalankan perintah berikut.

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

Secara default, objek database baru, seperti tabel, dibuat dalam skema default bernama `public` dibuat selama pembuatan gudang data. Anda dapat menggunakan skema lain untuk membuat objek database. Untuk informasi selengkapnya tentang skema, lihat [Mengelola keamanan database](#) di Panduan Pengembang Database Amazon Redshift.

Anda juga dapat membuat tabel menggunakan `schema_name.object_name` notasi untuk membuat tabel dalam SALES skema.

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

Untuk melihat dan memeriksa skema dan tabelnya, Anda dapat menggunakan editor kueri Amazon Redshift v2. Atau Anda dapat melihat daftar tabel dalam skema menggunakan tampilan sistem. Untuk informasi selengkapnya, lihat [Langkah 6: Kueri tabel sistem](#).

`sortkey` Kolom `encodingdistkey`, dan digunakan oleh Amazon Redshift untuk pemrosesan paralel. Untuk informasi selengkapnya tentang mendesain tabel yang menggabungkan elemen-elemen ini, lihat [Praktik terbaik Amazon Redshift untuk mendesain](#) tabel.

Masukkan baris data ke dalam tabel

Setelah Anda membuat tabel, masukkan baris data ke dalam tabel itu.

Note

Perintah [INSERT](#) menyisipkan baris ke dalam tabel. Untuk beban massal standar, gunakan perintah [COPY](#). Untuk informasi selengkapnya, lihat [Menggunakan perintah COPY untuk memuat data](#).

Misalnya, untuk memasukkan nilai ke dalam DEMO tabel, jalankan perintah berikut.

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

Untuk menyisipkan data ke dalam tabel yang ada dalam skema tertentu, jalankan perintah berikut.

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

Pilih data dari tabel

Setelah Anda membuat tabel dan mengisinya dengan data, gunakan pernyataan SELECT untuk menampilkan data yang terkandung dalam tabel. Pernyataan SELECT * mengembalikan semua nama kolom dan nilai baris untuk semua data dalam tabel. Menggunakan SELECT adalah cara yang baik untuk memverifikasi bahwa data yang baru ditambahkan dimasukkan dengan benar ke dalam tabel.

Untuk melihat data yang Anda masukkan dalam **DEMO** tabel, jalankan perintah berikut.

```
SELECT * from DEMO;
```

Hasilnya akan terlihat seperti berikut ini.

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

Untuk informasi selengkapnya tentang menggunakan pernyataan SELECT untuk menanyakan tabel, lihat [SELECT](#).

Langkah 5: Muat data

Banyak contoh dalam panduan ini menggunakan kumpulan data sampel TICKIT. Anda dapat mengunduh file [ticketdb.zip](#), yang berisi file data sampel individual. Anda kemudian dapat mengunggah data sampel ke bucket Amazon S3 Anda sendiri.

Untuk memuat data sampel untuk database Anda, buat tabel terlebih dahulu. Kemudian gunakan perintah COPY untuk memuat tabel dengan data sampel yang disimpan dalam bucket Amazon S3. Untuk langkah-langkah untuk membuat tabel dan memuat data sampel, lihat [Langkah 4: Muat data dari Amazon S3 ke Amazon Redshift](#).

Langkah 6: Kueri tabel sistem

Selain tabel yang Anda buat, gudang data Anda berisi sejumlah tabel dan tampilan sistem. Tabel sistem ini berisi informasi tentang instalasi Anda dan tentang berbagai kueri dan proses yang berjalan pada sistem. Anda dapat menanyakan tabel sistem ini untuk mengumpulkan informasi tentang database Anda. Untuk informasi selengkapnya, lihat [Referensi tabel dan tampilan sistem](#) di Panduan Pengembang Database Amazon Redshift. Deskripsi untuk setiap tabel atau tampilan menunjukkan apakah tabel terlihat oleh semua pengguna atau hanya untuk pengguna super. Masuk sebagai pengguna super untuk menanyakan tabel yang hanya terlihat oleh pengguna super.

Melihat daftar nama tabel

Untuk melihat daftar semua tabel dalam skema, Anda dapat menanyakan tabel katalog sistem PG_TABLE_DEF. Anda dapat terlebih dahulu memeriksa pengaturan untuk `search_path`.

```
SHOW search_path;
```

Hasilnya akan terlihat mirip dengan yang berikut ini,

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

Contoh berikut menambahkan SALES skema ke jalur pencarian dan menunjukkan semua tabel dalam SALES skema.

```
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
```

Contoh berikut menunjukkan daftar semua tabel yang disebut DEMO dalam semua skema pada database saat ini.

```
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
```

Untuk informasi selengkapnya, lihat [PG_TABLE_DEF](#).

Anda juga dapat menggunakan editor kueri Amazon Redshift v2 untuk melihat semua tabel dalam skema tertentu dengan terlebih dahulu memilih database yang ingin Anda sambungkan.

Lihat pengguna

Anda dapat menanyakan katalog PG_USER untuk melihat daftar semua pengguna, bersama dengan ID pengguna (USESYSID) dan hak istimewa pengguna.

```
SELECT * FROM pg_user;
```

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

Nama pengguna digunakan `rdsdb` secara internal oleh Amazon Redshift untuk melakukan tugas administrasi dan pemeliharaan rutin. Anda dapat memfilter kueri untuk hanya menampilkan nama pengguna yang ditentukan pengguna dengan menambahkan `where usesysid > 1` ke pernyataan SELECT Anda.

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

username	usesysid	usecreatedb	usesuper	usecatupd	passwd	valuntil	useconfig
awsuser	100	true	true	false	*****		
guest	104	true	false	false	*****		

Lihat pertanyaan terbaru

Pada contoh sebelumnya, ID pengguna (`user_id`) untuk `adminuser` adalah 100. Untuk mencantumkan empat kueri terbaru yang dijalankan oleh `adminuser`, Anda dapat menanyakan tampilan `SYS_QUERY_HISTORY`.

Anda dapat menggunakan tampilan ini untuk menemukan ID kueri (`query_id`) atau ID proses (`session_id`) untuk kueri yang baru saja dijalankan. Anda juga dapat menggunakan tampilan ini untuk memeriksa berapa lama waktu yang dibutuhkan kueri untuk diselesaikan. `SYS_QUERY_HISTORY` menyertakan 4.000 karakter pertama dari string kueri (`query_text`) untuk membantu Anda

menemukan kueri tertentu. Gunakan klausa LIMIT dengan pernyataan SELECT Anda untuk membatasi hasil.

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

Hasilnya terlihat seperti berikut ini.

```
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);
```

Tentukan ID sesi dari kueri yang sedang berjalan

Untuk mengambil informasi tabel sistem tentang kueri, Anda mungkin perlu menentukan ID sesi (ID proses) yang terkait dengan kueri tersebut. Atau, Anda mungkin perlu menemukan ID sesi untuk kueri yang masih berjalan. Misalnya, Anda memerlukan ID sesi jika Anda perlu membatalkan kueri yang terlalu lama untuk dijalankan di klaster yang disediakan. Anda dapat menanyakan tabel sistem STV_RECENTS untuk mendapatkan daftar ID sesi untuk menjalankan kueri, bersama dengan string kueri yang sesuai. Jika kueri Anda mengembalikan beberapa sesi, Anda dapat melihat teks kueri untuk menentukan ID sesi mana yang Anda butuhkan.

Untuk menentukan ID sesi dari kueri yang sedang berjalan, jalankan pernyataan SELECT berikut.

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

Langkah 7: Batalkan kueri

Jika Anda menjalankan kueri yang memakan waktu terlalu lama atau menghabiskan sumber daya yang berlebihan, batalkan kueri. Misalnya, buat daftar penjual tiket yang mencakup nama penjual dan jumlah tiket yang terjual. Query berikut memilih data dari SALES tabel dan tabel dan USERS menggabungkan dua tabel dengan mencocokkan SELLERID dan USERID dalam klausa WHERE.

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

Hasilnya terlihat seperti berikut ini.

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

Ini adalah kueri yang kompleks. Untuk tutorial ini, Anda tidak perlu khawatir tentang bagaimana query ini dibangun.

Query sebelumnya berjalan dalam hitungan detik dan mengembalikan 2.102 baris.

Misalkan Anda lupa memasukkan klausa WHERE.

```
SELECT sellerid, firstname, lastname, sum(qtysold)
FROM sales, users
GROUP BY sellerid, firstname, lastname
```

```
ORDER BY 4 desc;
```

Set hasil mencakup semua baris dalam tabel dikalikan dengan semua baris dalam SALES tabel (USERS49989*3766). Ini disebut bergabung Cartesian, dan itu tidak direkomendasikan. Hasilnya lebih dari 188 juta baris dan membutuhkan waktu lama untuk dijalankan.

Untuk membatalkan kueri yang sedang berjalan, gunakan perintah CANCEL dengan ID sesi kueri. Dengan editor kueri Amazon Redshift v2 Anda dapat membatalkan kueri dengan memilih tombol batal saat kueri sedang berjalan.

Untuk menemukan ID sesi, mulai sesi baru dan kueri tabel STV_RECENTS, seperti yang ditunjukkan pada langkah sebelumnya. Contoh berikut menunjukkan bagaimana Anda dapat membuat hasil lebih mudah dibaca. Untuk melakukan ini, gunakan fungsi TRIM untuk memangkas spasi tambahan dan hanya menampilkan 20 karakter pertama dari string kueri.

Untuk menentukan ID sesi dari kueri yang sedang berjalan, jalankan pernyataan SELECT berikut.

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

Hasilnya terlihat seperti berikut ini.

```
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 ...
```

Untuk membatalkan kueri dengan ID sesi1073791534, jalankan perintah berikut.

```
CANCEL 1073791534;
```

Note

Perintah CANCEL tidak menghentikan transaksi. Untuk menghentikan atau memutar kembali transaksi, gunakan perintah ABORT atau ROLLBACK. Untuk membatalkan kueri yang terkait dengan transaksi, pertama-tama batalkan kueri lalu hentikan transaksi.

Jika kueri yang Anda batalkan dikaitkan dengan transaksi, gunakan perintah ABORT atau ROLLBACK untuk membatalkan transaksi dan membuang setiap perubahan yang dilakukan pada data:

```
ABORT;
```

Kecuali Anda masuk sebagai pengguna super, Anda hanya dapat membatalkan pertanyaan Anda sendiri. Superuser dapat membatalkan semua kueri.

Jika alat kueri Anda tidak mendukung kueri yang berjalan secara bersamaan, mulailah sesi lain untuk membatalkan kueri.

Untuk informasi selengkapnya tentang membatalkan kueri, lihat [CANCEL](#) di Panduan Pengembang Database Amazon Redshift.

Batalkan kueri menggunakan antrean superuser

Jika sesi Anda saat ini memiliki terlalu banyak kueri yang berjalan secara bersamaan, Anda mungkin tidak dapat menjalankan perintah CANCEL sampai kueri lain selesai. Dalam hal ini, jalankan perintah CANCEL menggunakan antrean kueri manajemen beban kerja yang berbeda.

Dengan menggunakan manajemen beban kerja, Anda dapat menjalankan kueri dalam antrian kueri yang berbeda sehingga Anda tidak perlu menunggu kueri lain selesai. Manajer beban kerja membuat antrian terpisah, yang disebut antrian Superuser, yang dapat Anda gunakan untuk pemecahan masalah. Untuk menggunakan antrian Superuser, masuk ke superuser dan atur grup kueri ke 'superuser' menggunakan perintah SET. Setelah menjalankan perintah Anda, setel ulang grup kueri menggunakan perintah RESET.

Untuk membatalkan kueri menggunakan antrean superuser, jalankan perintah ini.

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

Mengkueri sumber data di luar database Amazon Redshift Anda

Berikut ini, Anda dapat menemukan informasi tentang cara memulai kueri data pada sumber jarak jauh, termasuk klaster Amazon Redshift dari jarak jauh. Anda juga dapat menemukan informasi tentang model training machine learning (ML/machine learning) menggunakan Amazon Redshift.

Topik

- [Menanyakan data lake Anda](#)
- [Menanyakan data pada sumber data jarak jauh](#)
- [Mengakses data di klaster Amazon Redshift lainnya](#)
- [Melatih model machine learning dengan data Amazon Redshift](#)

Menanyakan data lake Anda

Anda dapat menggunakan Amazon Redshift Spectrum untuk melakukan kueri data dalam file Amazon S3 tanpa harus memuat data ke tabel Amazon Redshift. Amazon Redshift menyediakan kemampuan SQL yang dirancang untuk pemrosesan analitik online cepat (OLAP) dari kumpulan data yang sangat besar yang disimpan di klaster Amazon Redshift dan data lake Amazon S3. Anda dapat query data dalam berbagai format, termasuk Parquet, ORC, RCFile, TextFile, SequenceFile, RegexSerde, OpenCSV, dan AVRO. Untuk menentukan struktur file di Amazon S3, Anda membuat skema dan tabel eksternal. Kemudian, Anda menggunakan katalog data eksternal seperti AWS Glue atau metastore Apache Hive Anda sendiri. Perubahan pada salah satu jenis katalog data segera tersedia untuk kluster Amazon Redshift Anda.

Setelah data Anda terdaftar dengan Katalog AWS Glue Data dan diaktifkan dengan AWS Lake Formation, Anda dapat query dengan menggunakan Redshift Spectrum.

Redshift Spectrum berada di server Amazon Redshift khusus yang tidak bergantung pada klaster Anda. Redshift Spectrum mendorong banyak tugas intensif komputasi, seperti penyaringan predikat dan agregasi, ke layer Redshift Spectrum. Redshift Spectrum juga menskalakan secara cerdas untuk memanfaatkan pemrosesan parallel secara masif.

Anda dapat partisi tabel eksternal pada satu atau lebih kolom untuk mengoptimalkan kinerja query melalui penghapusan partisi. Anda dapat melakukan kueri dan bergabung dengan tabel eksternal

dengan tabel Amazon Redshift. Anda dapat mengakses tabel eksternal dari beberapa kluster Amazon Redshift dan mengkueri data Amazon S3 dari kluster mana pun di AWS Wilayah yang sama. Saat Anda memperbarui file data Amazon S3, data segera tersedia untuk kueri dari kluster Amazon Redshift Anda.

Untuk informasi selengkapnya tentang Redshift Spectrum, termasuk cara bekerja dengan Redshift Spectrum dan data lake, lihat [Memulai Amazon Redshift Spectrum](#) di Panduan Pengembang Database Amazon Redshift.

Menanyakan data pada sumber data jarak jauh

Anda dapat menggabungkan data dari database Amazon RDS, database Amazon Aurora, atau Amazon S3 dengan data dalam database Amazon Redshift Anda menggunakan kueri federasi. Anda dapat menggunakan Amazon Redshift untuk melakukan kueri data operasional secara langsung (tanpa memindahkannya), menerapkan transformasi, dan memasukkan data ke tabel Redshift Anda. Beberapa perhitungan untuk kueri federasi didistribusikan ke sumber data jarak jauh.

Untuk menjalankan kueri federasi, Amazon Redshift terlebih dahulu membuat sambungan ke sumber data jarak jauh. Amazon Redshift kemudian mengambil metadata tentang tabel di sumber data jarak jauh, mengeluarkan kueri, dan kemudian mengambil baris hasil. Amazon Redshift kemudian mendistribusikan baris hasil ke node komputasi Amazon Redshift untuk diproses lebih lanjut.

Untuk informasi tentang pengaturan lingkungan Anda untuk kueri federasi, lihat salah satu topik berikut dalam Panduan Pengembang Database Amazon Redshift:

- [Memulai dengan menggunakan kueri federasi ke PostgreSQL](#)
- [Memulai dengan menggunakan kueri federasi ke MySQL](#)

Mengakses data di kluster Amazon Redshift lainnya

Menggunakan berbagi data Amazon Redshift, Anda dapat berbagi data langsung dengan keamanan tinggi dan kemudahan yang lebih besar di seluruh kluster atau AWS akun Amazon Redshift untuk tujuan baca. Anda dapat memiliki akses instan, terperinci, dan berkinerja tinggi ke data di seluruh kluster Amazon Redshift tanpa menyalin atau memindahkannya secara manual. Pengguna Anda dapat melihat informasi yang paling up-to-date dan konsisten saat diperbarui di kluster Amazon Redshift. Anda dapat berbagi data pada tingkat yang berbeda, seperti database, skema, tabel, tampilan (termasuk tampilan reguler, pengikatan akhir, dan terwujud), dan fungsi yang ditentukan pengguna SQL (UDF).

Berbagi data Amazon Redshift sangat berguna untuk kasus penggunaan ini:

- Memusatkan beban kerja penting bisnis - Gunakan kluster sentral, transformasi, dan muat (ETL) yang berbagi data dengan beberapa kluster intelijen bisnis (BI) atau kluster analitik. Pendekatan ini menyediakan isolasi beban kerja baca dan tolak bayar untuk beban kerja individu.
- Berbagi data antara lingkungan — Berbagi data antara pengembangan, pengujian, dan lingkungan produksi. Anda dapat meningkatkan kelincahan tim dengan berbagi data pada berbagai tingkat rincian rincian rincian rincian yang berbeda.

Untuk informasi selengkapnya tentang berbagi data, lihat [Memulai berbagi data](#) di Panduan Pengembang Database Amazon Redshift.

Melatih model machine learning dengan data Amazon Redshift

Dengan menggunakan pembelajaran mesin Amazon Redshift (Amazon Redshift ML), Anda dapat melatih model dengan menyediakan data ke Amazon Redshift. Kemudian Amazon Redshift ML menciptakan model yang menangkap pola dalam data input. Anda kemudian dapat menggunakan model ini untuk menghasilkan prediksi untuk data input baru tanpa menimbulkan biaya tambahan. Dengan menggunakan Amazon Redshift, Anda dapat melatih model machine learning menggunakan pernyataan SQL dan memanggilnya dalam kueri SQL untuk prediksi. Anda dapat terus meningkatkan akurasi prediksi dengan mengubah parameter secara berulang dan meningkatkan data pelatihan Anda.

Amazon Redshift ML memudahkan pengguna SQL untuk membuat, melatih, dan menerapkan model machine learning menggunakan perintah SQL yang sudah dikenal. Dengan menggunakan Amazon Redshift, Anda dapat menggunakan data Anda di kluster Amazon Redshift untuk melatih model dengan Amazon SageMaker Autopilot dan secara otomatis mendapatkan model terbaik. Anda kemudian dapat melokalisasi model dan membuat prediksi dari dalam database Amazon Redshift.

Untuk informasi selengkapnya tentang Amazon Redshift ML, lihat [Memulai Amazon Redshift ML dalam](#) Panduan Pengembang Database Amazon Redshift.

Gambaran umum Amazon Redshift

Amazon Redshift Tanpa Server memungkinkan Anda mengakses dan menganalisis data tanpa semua konfigurasi gudang data yang disediakan. Sumber daya disediakan secara otomatis dan kapasitas gudang data diskalakan secara cerdas untuk memberikan kinerja yang cepat bahkan untuk beban kerja yang paling menuntut dan tidak dapat diprediksi. Anda tidak dikenakan biaya saat gudang data tidak aktif, jadi Anda hanya membayar atas apa yang Anda gunakan. Anda dapat memuat data dan langsung mulai melakukan kueri di editor kueri Amazon Redshift v2 atau di alat kecerdasan bisnis (BI) favorit Anda. Nikmati kinerja harga terbaik dan fitur SQL yang easy-to-use sudah dikenal di lingkungan administrasi nol.

Jika Anda baru pertama kali menggunakan Amazon Redshift, sebaiknya mulai dengan membaca bagian-bagian berikut:

- [Gambaran umum Amazon Redshift Serverless](#) — Dalam topik ini, Anda dapat menemukan gambaran umum Amazon Redshift Serverless dan kemampuan utamanya.
- [Sorotan dan harga layanan](#) - Pada halaman detail produk ini, Anda dapat menemukan detail tentang sorotan dan harga Amazon Redshift Tanpa Server.
- [Memulai Amazon Redshift Tanpa Server](#) — Dalam topik ini, Anda dapat mempelajari lebih lanjut tentang cara membuat gudang data Amazon Redshift Tanpa Server, dan mulai melakukan kueri data menggunakan editor kueri v2.

Jika Anda lebih suka mengelola sumber daya Amazon Redshift secara manual, Anda dapat membuat klaster yang disediakan untuk kebutuhan kueri data Anda. Untuk informasi selengkapnya, lihat [klaster Amazon Redshift](#).

Jika organisasi Anda memenuhi syarat dan klaster Anda sedang dibuat di Wilayah AWS tempat Amazon Redshift Tanpa Server tidak tersedia, Anda mungkin dapat membuat klaster di bawah program uji coba gratis Amazon Redshift. Pilih Production atau Free trial untuk menjawab pertanyaan Apa yang Anda rencanakan untuk menggunakan klaster ini? Bila Anda memilih Uji coba gratis, Anda membuat konfigurasi dengan tipe node dc2.large. Untuk informasi selengkapnya tentang memilih uji coba gratis, lihat uji [coba gratis Amazon Redshift](#). Untuk daftar Wilayah AWS tempat Amazon Redshift Tanpa Server tersedia, lihat titik akhir yang tercantum untuk [Redshift Serverless API](#) di Referensi Umum Amazon Web Services.

Berikut ini adalah beberapa konsep kunci Amazon Redshift Tanpa Server.

- **Namespace** - Kumpulan objek database dan pengguna. Ruang nama mengelompokkan semua sumber daya yang Anda gunakan di Amazon Redshift Tanpa Server, seperti skema, tabel, pengguna, datashares, dan snapshot.
- **Workgroup** - Kumpulan sumber daya komputasi. Workgroup menampung sumber daya komputasi yang digunakan Amazon Redshift Tanpa Server untuk menjalankan tugas komputasi. Beberapa contoh sumber daya tersebut termasuk Redshift Processing Units (RPU), grup keamanan, batas penggunaan. Workgroup memiliki pengaturan jaringan dan keamanan yang dapat Anda konfigurasi menggunakan Amazon Redshift Serverless console, AWS Command Line Interface, atau Amazon Redshift Serverless API.

Untuk informasi selengkapnya tentang mengonfigurasi sumber daya namespace dan workgroup, lihat [Bekerja dengan ruang nama](#) dan [Bekerja dengan workgroup](#).

Berikut ini adalah beberapa konsep kluster yang disediakan Amazon Redshift utama:

- **Kluster** — Komponen infrastruktur inti dari gudang data Amazon Redshift adalah kluster.

Kluster terdiri dari satu atau lebih node komputasi. Node komputasi menjalankan kode yang dikompilasi.

Jika cluster disediakan dengan dua atau lebih node komputasi, node pemimpin tambahan mengoordinasikan node komputasi. Node pemimpin menangani komunikasi eksternal dengan aplikasi, seperti alat intelijen bisnis dan editor kueri. Aplikasi klien Anda berinteraksi langsung hanya dengan node pemimpin. Node komputasi transparan untuk aplikasi eksternal.

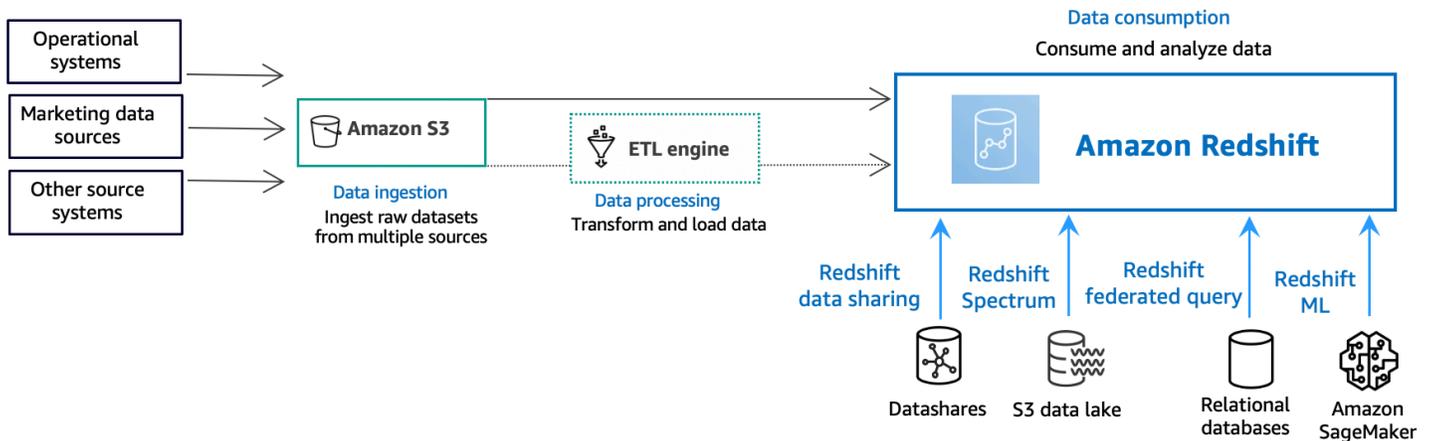
- **Database** - Sebuah cluster berisi satu atau lebih database.

Data pengguna disimpan dalam satu atau lebih database pada node komputasi. Klien SQL Anda berkomunikasi dengan node pemimpin, yang pada gilirannya mengoordinasikan menjalankan kueri dengan node komputasi. Untuk detail tentang node komputasi dan node pemimpin, lihat [Arsitektur sistem gudang data](#). Dalam database, data pengguna diatur ke dalam satu atau lebih skema.

Amazon Redshift adalah sistem manajemen basis data relasional (RDBMS) dan kompatibel dengan aplikasi RDBMS lainnya. Ini menyediakan fungsi yang sama seperti RDBMS khas, termasuk pemrosesan transaksi online (OLTP) fungsi seperti memasukkan dan menghapus data. Amazon Redshift juga dioptimalkan untuk analisis batch kinerja tinggi dan pelaporan set data.

Setelah itu, Anda dapat menemukan deskripsi alur pemrosesan data tipikal di Amazon Redshift, bersama dengan deskripsi bagian alur yang berbeda. Untuk informasi lebih lanjut tentang arsitektur sistem Amazon Redshift, lihat [Arsitektur sistem gudang data](#).

Diagram berikut menggambarkan alur pemrosesan data yang khas di Amazon Redshift.



Gudang data Amazon Redshift adalah kueri database relasional kelas perusahaan dan sistem manajemen. Amazon Redshift mendukung koneksi klien dengan berbagai jenis aplikasi, termasuk intelijen bisnis (BI), pelaporan, data, dan alat analisis. Ketika Anda menjalankan kueri analitik, Anda mengambil, membandingkan, dan mengevaluasi sejumlah besar data dalam operasi beberapa tahap untuk menghasilkan hasil akhir.

Pada lapisan konsumsi data, berbagai jenis sumber data terus mengunggah data terstruktur, semistruktur, atau tidak terstruktur ke lapisan penyimpanan data. Area penyimpanan data ini berfungsi sebagai area pementasan yang menyimpan data di berbagai negara kesiapan konsumsi. Contoh penyimpanan mungkin bucket Amazon Simple Storage Service (Amazon S3).

Pada lapisan pengolahan data opsional, data sumber berjalan melalui preprocessing, validasi, dan transformasi menggunakan ekstrak, transformasi, beban (ETL) atau ekstrak, beban, transformasi (ELT) pipa. Dataset mentah ini kemudian disempurnakan dengan menggunakan operasi ETL. Contoh mesin ETL adalah AWS Glue.

Pada lapisan konsumsi data, data dimuat ke kluster Amazon Redshift, tempat Anda dapat menjalankan beban kerja analitis.

Untuk beberapa contoh beban kerja analitis, lihat [Kueri di luar sumber data](#).

Sumber daya tambahan

Untuk informasi selengkapnya tentang Amazon Redshift Tanpa Server, kami menyarankan Anda untuk terus mempelajari lebih lanjut tentang konsep yang diperkenalkan dalam panduan ini dengan menggunakan sumber daya Amazon Redshift berikut:

- **Video fitur:** Video ini membantu Anda mempelajari fitur Amazon Redshift.
 - Untuk memahami Amazon Redshift Tanpa Server pada tingkat tinggi, tonton video berikut. [Amazon Redshift Tanpa Server Dijelaskan dalam 90 Detik.](#)
 - Untuk mempelajari cara menyiapkan gudang data tanpa server dan mulai melakukan kueri data, tonton video berikut. [Memulai dengan Amazon Redshift Tanpa Server.](#)
- **[Panduan Manajemen Amazon Redshift:](#)** Panduan ini dibuat berdasarkan Panduan Memulai Amazon Redshift ini. Ini memberikan informasi mendalam tentang konsep dan tugas untuk membuat, mengelola, dan memantau kluster yang disediakan Amazon Redshift Tanpa Server dan Amazon Redshift.
- **[Panduan Pengembang Database Amazon Redshift:](#)** Panduan ini juga dibuat berdasarkan Panduan Memulai Amazon Redshift ini. Ini memberikan informasi mendalam untuk pengembang database tentang merancang, membangun, query, dan memelihara database yang membentuk gudang data Anda.
 - **[Referensi SQL:](#)** Topik ini menjelaskan perintah SQL dan referensi fungsi untuk Amazon Redshift.
 - **[Tabel dan tampilan sistem:](#)** Topik ini menjelaskan tabel dan tampilan sistem untuk Amazon Redshift.
- **Tutorial untuk Amazon Redshift:** Topik ini menunjukkan tutorial tentang fitur Amazon Redshift.
 - **[Memuat data dari Amazon S3:](#)** Tutorial ini menjelaskan cara memuat data ke dalam tabel database Amazon Redshift Anda dari file data dalam bucket Amazon S3.
 - **[Memulai berbagi data:](#)** Bagian ini menjelaskan cara berbagi dan mengakses data di kluster Amazon Redshift lainnya.
 - **[Menggunakan fungsi SQL spasial dengan Amazon Redshift:](#)** Tutorial ini menunjukkan cara menggunakan beberapa fungsi SQL spasial dengan Amazon Redshift.
 - **[Mengkueri data bertingkat dengan Amazon Redshift Spectrum:](#)** Tutorial ini menjelaskan cara menggunakan Redshift Spectrum untuk mengkueri data bersarang dalam format file Parquet, ORC, JSON, dan Ion menggunakan tabel eksternal.
 - **[Mengonfigurasi antrean manajemen beban kerja manual \(WLM\):](#)** Tutorial ini menjelaskan cara mengkonfigurasi manajemen beban kerja manual (WLM) di Amazon Redshift.

- [Memulai dengan Amazon Redshift ML:Bagian](#) ini menjelaskan cara pengguna dapat membuat, melatih, dan menerapkan model machine learning menggunakan perintah SQL yang sudah dikenal.
- [Apa yang baru](#): Halaman web ini mencantumkan fitur baru dan pembaruan produk Amazon Redshift.

Riwayat dokumen

Note

Untuk deskripsi fitur baru di Amazon Redshift, lihat [Apa yang baru.](#)

Tabel berikut menjelaskan perubahan dokumentasi penting pada Panduan Memulai Amazon Redshift.

Perubahan	Deskripsi	Tanggal rilis
Pembaruan dokumentasi	Memperbarui panduan untuk menyertakan bagian baru tentang memulai tugas database umum, menanyakan data lake Anda, menanyakan data pada sumber jarak jauh, berbagi data, dan melatih model pembelajaran mesin dengan data Amazon Redshift.	30 Juni 2021
Fitur baru	Memperbarui panduan untuk menjelaskan prosedur pemuatan sampel baru.	4 Juni 2021
Pembaruan dokumentasi	Memperbarui panduan untuk menghapus konsol Amazon Redshift asli dan meningkatkan aliran langkah.	14 Agustus 2020
Konsol baru	Memperbarui panduan untuk menggambarkan konsol Amazon Redshift baru.	11 November 2019
Fitur baru	Memperbarui panduan untuk menjelaskan prosedur cluster peluncuran cepat.	Agustus 10, 2018
Fitur baru	Memperbarui panduan untuk meluncurkan cluster dari dasbor Amazon Redshift.	28 Juli 2015
Fitur baru	Memperbarui panduan untuk menggunakan nama tipe node baru.	Juni 9, 2015

Perubahan	Deskripsi	Tanggal rilis
Pembaruan dokumentasi	Tangkapan layar dan prosedur yang diperbarui untuk mengonfigurasi grup keamanan VPC.	30 April 2015
Pembaruan dokumentasi	Tangkapan layar dan prosedur yang diperbarui agar sesuai dengan konsol saat ini.	12 November 2014
Pembaruan dokumentasi	Memindahkan pemuatan data dari informasi Amazon S3 ke bagiannya sendiri dan memindahkan bagian langkah selanjutnya ke langkah terakhir untuk penemuan yang lebih baik.	13 Mei 2014
Pembaruan dokumentasi	Menghapus halaman Selamat Datang dan memasukkan konten ke halaman Memulai utama.	Maret 14, 2014
Pembaruan dokumentasi	Ini adalah rilis baru dari Panduan Memulai Amazon Redshift yang membahas umpan balik pelanggan dan pembaruan layanan.	Maret 14, 2014
Panduan baru	Ini adalah rilis pertama dari Panduan Memulai Amazon Redshift.	Februari 14, 2013

Terjemahan disediakan oleh mesin penerjemah. Jika konten terjemahan yang diberikan bertentangan dengan versi bahasa Inggris aslinya, utamakan versi bahasa Inggris.