



Guía de introducción

Amazon Redshift



Amazon Redshift: Guía de introducción

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

Las marcas comerciales y la imagen comercial de Amazon no se pueden utilizar en relación con ningún producto o servicio que no sea de Amazon, de ninguna manera que pueda causar confusión entre los clientes y que menosprecie o desacredite a Amazon. Todas las demás marcas comerciales que no son propiedad de Amazon son propiedad de sus respectivos propietarios, que pueden o no estar afiliados, relacionados o patrocinados por Amazon.

Table of Contents

Amazon Redshift sin servidor	1
Inscripción en AWS	1
Creación de un almacenamiento de datos con Amazon Redshift sin servidor	1
Carga de datos de ejemplo	4
Ejecución de consultas de ejemplo	8
Carga de datos desde Amazon S3	10
Consulta de orígenes de datos externos	18
Consulta de lagos de datos	18
Consulta de orígenes de datos remotos	19
Acceso a los datos en otros clústeres	19
Entrenamiento de modelos de ML con datos de Redshift	20
Clústeres aprovisionados de Amazon Redshift	21
Consola de clústeres aprovisionados de Amazon Redshift	21
Inscripción en AWS	21
Determinación de reglas de firewall	22
Conexión a Amazon Redshift	24
Clústeres y carga de datos de Amazon Redshift	25
Uso de un conjunto de datos de muestra	25
Traer sus propios datos a Amazon Redshift	30
Tareas comunes de base de datos	45
Tarea 1: Crear una base de datos	46
Tarea 2: Crear un usuario	46
Tarea 3: Crear un esquema	47
Tarea 4: Crear una tabla	49
Tarea 5: Cargar datos de muestra	51
Tarea 6: Consultar las tablas de sistema	51
Tarea 7: Cancelar una consulta	55
Tarea 8: Eliminar los recursos	58
Información general conceptual sobre Amazon Redshift	59
Recursos adicionales	63
Historial del documento	65

Amazon Redshift sin servidor

Si es la primera vez que utiliza Amazon Redshift Serverless, recomendamos que lea las siguientes secciones como ayuda para comenzar a utilizar Amazon Redshift Serverless. El flujo básico de Amazon Redshift sin servidor consiste en crear recursos sin servidor, conectarse a Amazon Redshift sin servidor, cargar datos de muestra y, después, ejecutar consultas en los datos. En esta guía, puede elegir cargar los datos de muestra desde Amazon Redshift sin servidor o desde un bucket de Amazon S3.

- [the section called “Inscripción en AWS”](#)
- [the section called “Creación de un almacenamiento de datos con Amazon Redshift sin servidor”](#)
- [the section called “Carga de datos desde Amazon S3”](#)

Inscripción en AWS

Si aún no tiene una cuenta de AWS, regístrese para obtener una. Si ya tiene una cuenta, puede saltarse este requisito previo y utilizar la cuenta existente.

1. Abra <https://portal.aws.amazon.com/billing/signup>.
2. Siga las instrucciones que se le indiquen.

Cuando se registra en una cuenta de AWS, se crea un usuario raíz de la cuenta de AWS. Este usuario tiene acceso a todos los recursos y los servicios de AWS en la cuenta. Como práctica recomendada de seguridad, [asigne acceso administrativo a un usuario administrativo](#) y utilice únicamente el usuario raíz para realizar la ejecución de [tareas que requieren acceso de usuario raíz](#).

Creación de un almacenamiento de datos con Amazon Redshift sin servidor

La primera vez que inicie sesión en la consola de Amazon Redshift sin servidor, se le pedirá que acceda a la experiencia de introducción, que puede utilizar para crear y administrar recursos sin servidor. En esta guía, creará recursos sin servidor mediante la configuración predeterminada de Amazon Redshift sin servidor.

Para obtener un control más detallado de su configuración, elija Customize settings (Personalizar configuración).

Para configurar con los ajustes predeterminados:

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en <https://console.aws.amazon.com/redshiftv2/>.

Elija Pruebe Amazon Redshift Serverless.

2. En Configuration (Configuración), elija Use default settings (Usar configuración predeterminada). Amazon Redshift Serverless crea un espacio de nombres predeterminado con un grupo de trabajo asociado a este espacio de nombres. Seleccione Guardar configuración.

En la siguiente captura de pantalla, se muestra la configuración predeterminada de Amazon Redshift sin servidor.

Get started with Amazon Redshift Serverless

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

Configuration [Info](#)

Use default settings

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

Customize settings

Customize your settings for your specific needs.

▼ How it works



Using the default settings

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



Customizing the settings

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

Namespace [Info](#)

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

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

Namespace name
default

Database name and password

Database name
dev

Admin user credentials
Created based on IAM credentials

Permissions

Default IAM role

<arn:aws:iam::694237154968:role/service-role/AmazonRedshift-CommandsAccessRole-20220809T211900> [↗](#)

- Una vez finalizada la configuración, elija Continue (Continuar) para ir al Serverless dashboard (Panel sin servidor). Puede ver que están disponibles el grupo de trabajo sin servidor y el espacio de nombres.

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

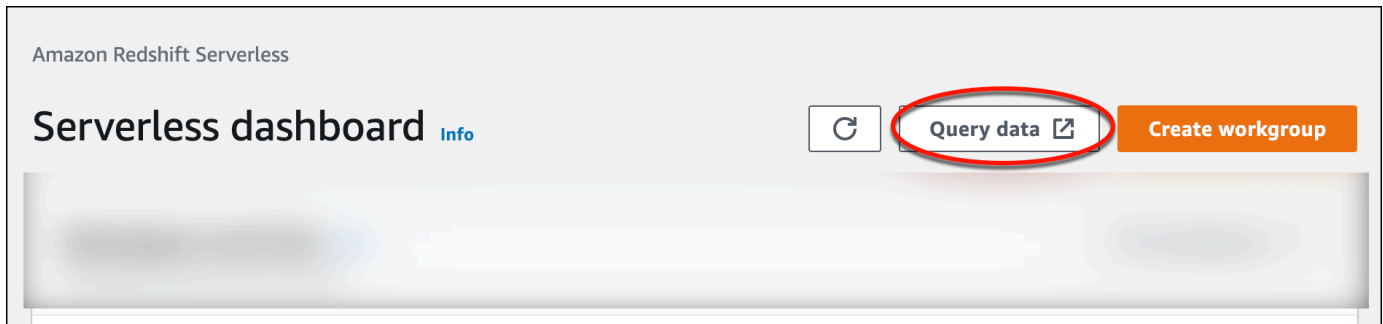
Namespaces / Workgroups [Info](#)

Namespace	Status	Workgroup	Status
default	✔ Available	default	✔ Available

Carga de datos de ejemplo

Ahora que ya ha configurado su almacenamiento de datos con Amazon Redshift sin servidor, puede utilizar el editor de consultas de Amazon Redshift v2 para cargar datos de muestra.

- Para iniciar el editor de consultas v2 desde la consola de Amazon Redshift sin servidor, elija Consultar datos. Cuando se invoca el editor de consultas v2 desde la consola de Amazon Redshift Serverless, se abre una nueva pestaña del navegador con el editor de consultas. El editor de consultas v2 se conecta desde su máquina cliente al entorno de Amazon Redshift Serverless.



2. Si inicia el editor de consultas v2 por primera vez, deberá configurar el cifrado de AWS KMS para poder continuar. Si lo desea, también puede especificar el URI en un bucket de S3 para cargar datos más adelante. Después de hacerlo, elija Configurar cuenta.

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)

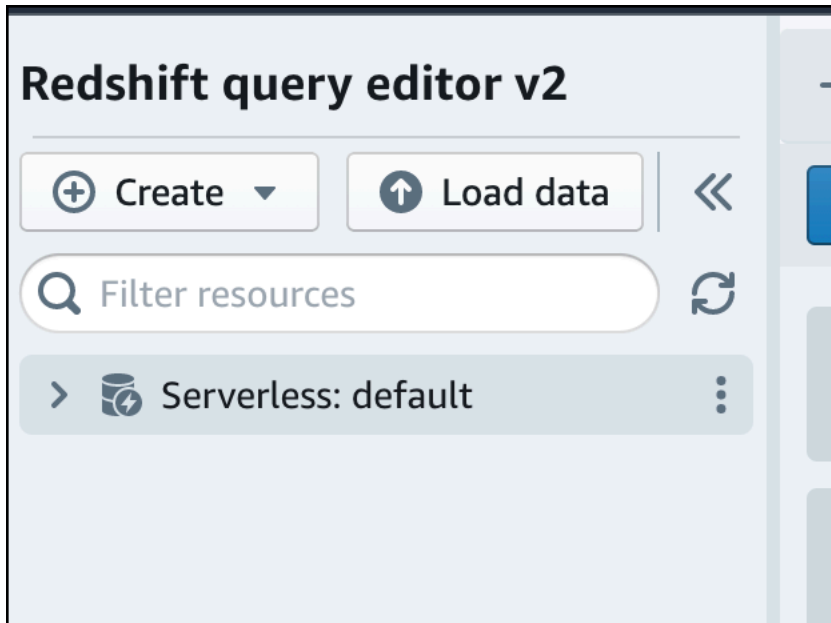
S3 bucket - optional

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

Para obtener más información sobre la configuración del editor de consultas v2, incluidos los permisos necesarios, consulte [Configuración de su Cuenta de AWS](#) en la Guía de administración de Amazon Redshift.

3. Para conectarse a un grupo de trabajo, elija su nombre en el panel de vista de árbol.



4. Cuando se conecte a un nuevo grupo de trabajo por primera vez en el editor de consultas v2, debe seleccionar el tipo de autenticación que utilizará para conectarse al grupo de trabajo. Para esta guía, deje seleccionado Usuario federado y elija Crear conexión.

Connect to default ✕

Authentication [Learn more](#)

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

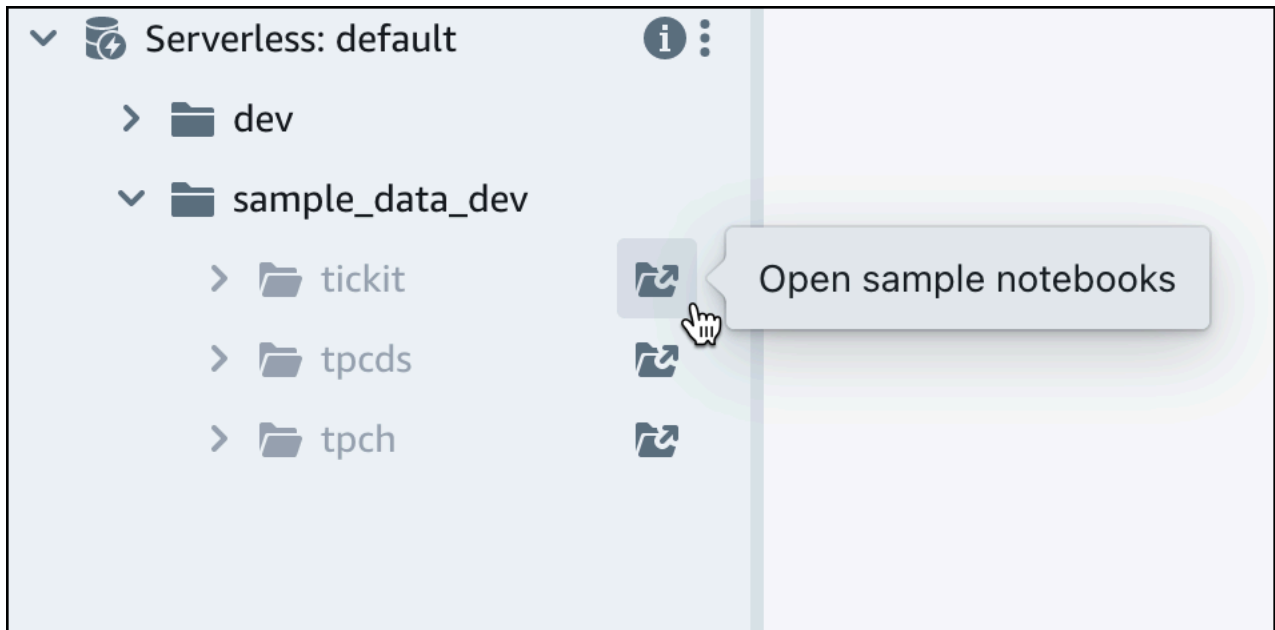
Database user name and password
Provide a database user and password for the database that you are connecting to. The query editor v2 stores your credentials in AWS Secrets Manager on your behalf.

Database

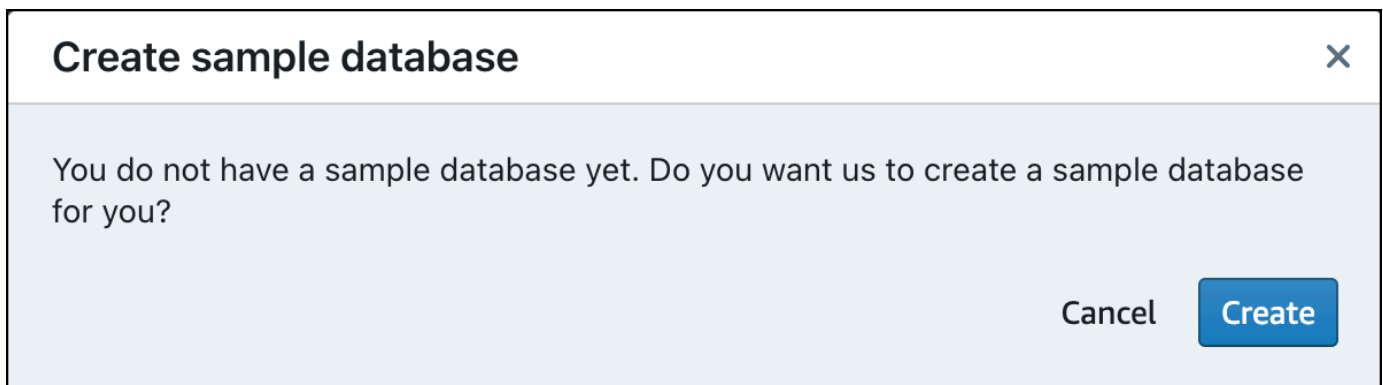
The database name must be 1-64 characters. Valid characters are lowercase alphanumeric characters.

Después de conectarse, puede elegir cargar los datos de muestra desde Amazon Redshift sin servidor o desde un bucket de Amazon S3.

5. En el grupo de trabajo predeterminado de Amazon Redshift Serverless, expanda la base de datos `sample_data_dev`. Hay tres esquemas de muestra que corresponden a tres conjuntos de datos de muestra que se pueden cargar en la base de datos de Amazon Redshift sin servidor. Elija el conjunto de datos de ejemplo que desee cargar y elija Abrir blocs de notas de muestra.



6. Cuando cargue datos por primera vez, el editor de consultas v2 le pedirá que cree una base de datos de muestra. Seleccione Crear.



Ejecución de consultas de ejemplo

Tras configurar Amazon Redshift sin servidor, puede comenzar a utilizar un conjunto de datos de muestra en Amazon Redshift sin servidor. Amazon Redshift sin servidor carga automáticamente el conjunto de datos de muestra, como el conjunto de datos tickit, y puede consultar inmediatamente los datos.

- Cuando Amazon Redshift sin servidor termina de cargar los datos de muestra, todas las consultas de muestra se cargan en el editor. Puede elegir Ejecutar todo para ejecutar todas las consultas desde los cuadernos de muestra.

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

```

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

```

Below the query, the results are shown in 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 results area, it says "Elapsed time: 401 ms Total rows: 9".

También puede exportar los resultados como un archivo JSON o CSV, o ver los resultados en un gráfico.

This screenshot shows the same Amazon Redshift console interface as the previous one, but with the "Export" and "Chart" options highlighted with a red circle. The "Export" button is a dropdown menu, and the "Chart" button is a toggle switch that is currently turned on. Below the query and results, there is a "Structure" panel on the left and a "Traces" panel in the middle. The "Traces" panel shows a "trace 0" with a "Type" of "Line". The "Chart" panel on the right is currently empty, with a placeholder text "Click to enter Plot title".

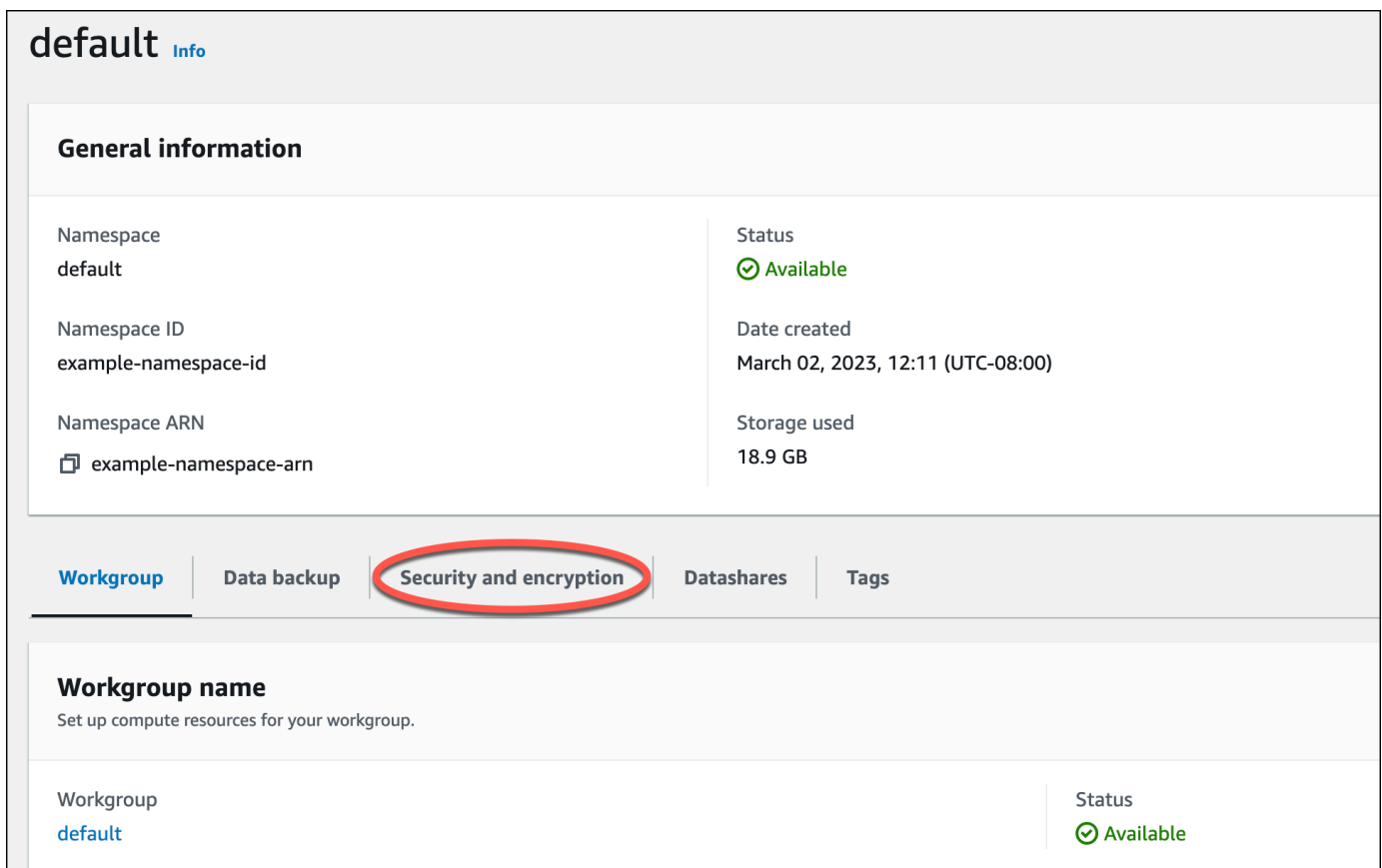
También puede cargar datos desde un bucket de Amazon S3. Consulte [the section called "Carga de datos desde Amazon S3"](#) para obtener más información.

Carga de datos desde Amazon S3

Después de crear el almacenamiento de datos, puede cargar los datos desde Amazon S3.

En este momento, tiene una base de datos denominada dev. Después, creará algunas tablas en la base de datos, cargará datos en ellas y probará una consulta. Para su comodidad, los datos de muestra que se cargan están disponibles en un bucket de Amazon S3.

1. Para poder cargar datos de Amazon S3, primero debe crear un rol de IAM con los permisos necesarios y asociarlo a su espacio de nombres sin servidor. Para ello, elija Configuración del espacio de nombres en el menú de navegación y, a continuación, elija Seguridad y cifrado. A continuación, elija Administrar roles de IAM.



default [Info](#)

General information

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

[Workgroup](#) | [Data backup](#) | **Security and encryption** | [Datashares](#) | [Tags](#)

Workgroup name
Set up compute resources for your workgroup.

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

2. Expanda el menú Administrar roles de IAM y elija Crear rol de 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 associated IAM roles or role type

Associate IAM roles

Create IAM role

Remove IAM roles

< 1 >

IAM roles [🔗](#) ▼ **Status** ▼ **Role type** ▼

3. Elija el nivel de acceso al bucket de S3 que desea conceder a este rol y elija Crear un rol de IAM como predeterminado.

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. Elija Guardar cambios. Ahora puede cargar datos de muestra de Amazon S3.

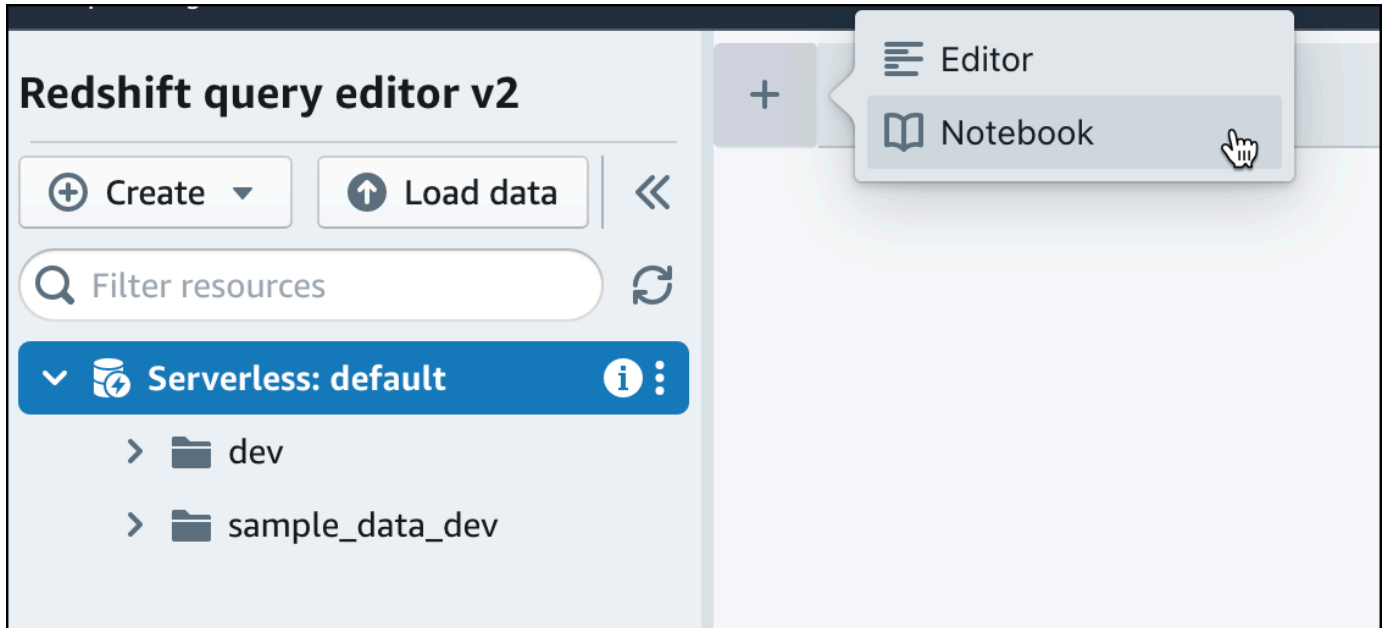
En los siguientes pasos, se utilizan datos en un bucket de S3 público de Amazon Redshift, pero puede replicar los mismos pasos con su propio bucket de S3 y comandos SQL.

Cargar datos de muestra de Amazon S3

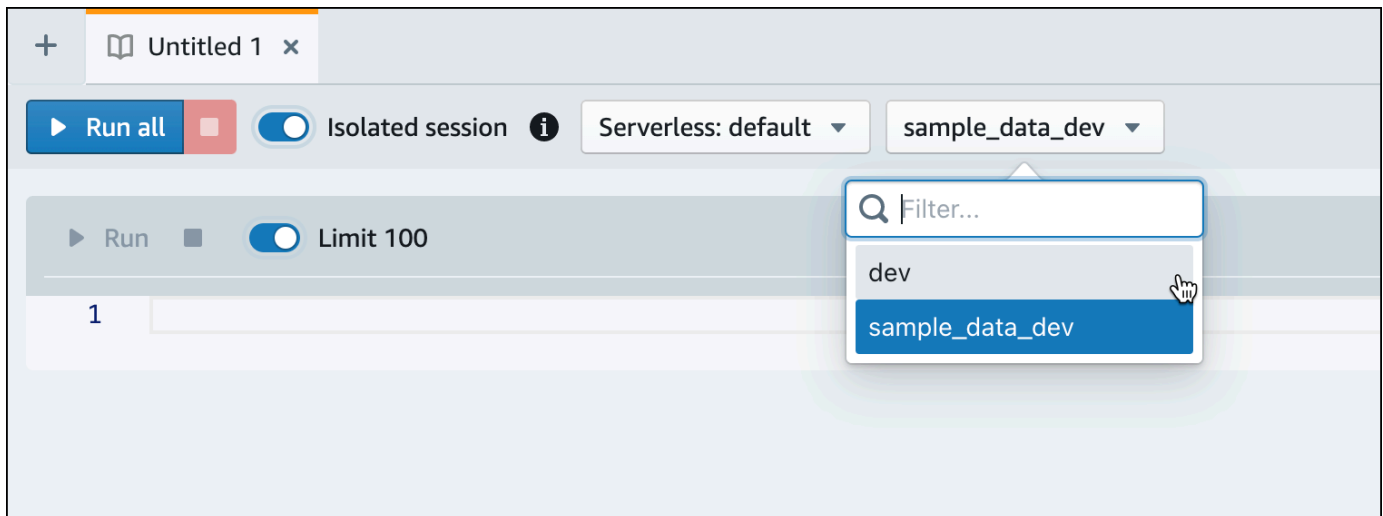
1. En el editor de consultas v2, elija



Agregar y, a continuación, Cuaderno para crear un nuevo cuaderno de SQL.



2. Cambie a la base de datos dev.



3. Cree tablas.

Si utiliza el editor de consultas v2, copie y ejecute las siguientes instrucciones de creación de tabla para crear tablas en la base de datos dev. Para obtener más información acerca de

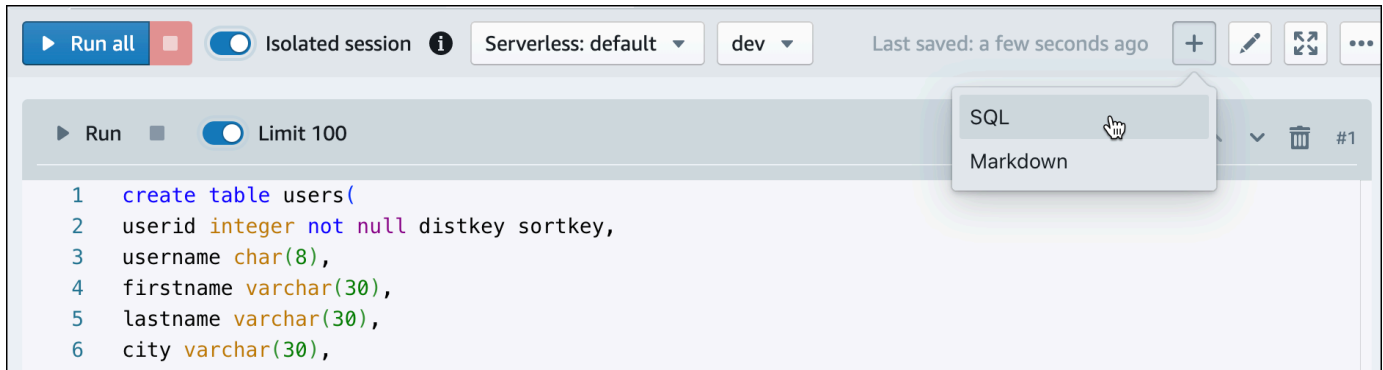
la sintaxis, consulte [CREATE TABLE](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

```
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. En el editor de consultas v2, cree una nueva celda SQL en su cuaderno.



5. Ahora utilice el comando COPY en el editor de consultas v2 para cargar grandes conjuntos de datos de Amazon S3 o Amazon DynamoDB en Amazon Redshift. Para obtener más información acerca de la sintaxis de COPY, consulte [COPY](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Puede ejecutar el comando COPY con algunos datos de ejemplo disponibles en un bucket de S3 público. Ejecute los siguientes comandos SQL en el editor de consultas v2.

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

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

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

- Después de cargar los datos, cree otra celda SQL en el cuaderno y pruebe algunas consultas de ejemplo. Para obtener más información acerca de cómo trabajar con el comando SELECT, consulte [SELECT](#) en la Guía para desarrolladores de Amazon Redshift. Para entender la estructura y los esquemas de los datos de muestra, explore el uso del 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;
```

Ahora que ha cargado los datos y ha ejecutado algunas consultas de muestra, puede explorar otras áreas de Amazon Redshift sin servidor. Consulte la siguiente lista para obtener más información sobre cómo puede utilizar Amazon Redshift sin servidor.

- Puede cargar datos desde un bucket de Amazon S3. Consulte [Carga de datos desde Amazon S3](#) para obtener más información.
- Puede utilizar el editor de consultas v2 para cargar datos de un archivo local separado por caracteres que ocupe menos de 5 MB. Para obtener más información, consulte [Carga de datos desde un archivo local](#).
- Puede conectarse a Amazon Redshift sin servidor con herramientas SQL de terceros con el controlador JDBC y ODBC. Consulte [Conexión a Amazon Redshift sin servidor](#) para obtener más información.
- También puede utilizar la API de datos de Amazon Redshift para conectarse a Amazon Redshift sin servidor. Consulte [Uso de la API de datos de Amazon Redshift](#) para obtener más información.

- Puede utilizar sus datos en Amazon Redshift sin servidor con Redshift ML para crear modelos de machine learning con el comando CREATE MODEL. Consulte [Tutorial: Creación de modelos de deserción de clientes](#) para aprender a crear un modelo de Redshift ML.
- Puede consultar datos de un lago de datos de Amazon S3 sin cargar datos en Amazon Redshift sin servidor. Consulte [Consulta de un lago de datos](#) para obtener más información.

Consulta de orígenes de datos fuera de la base de datos de Amazon Redshift

A continuación, encontrará información sobre cómo comenzar a consultar datos de fuentes remotas, incluidos los clústeres de Amazon Redshift remotos. También encontrará información sobre modelos de formación de machine learning (ML) mediante Amazon Redshift.

Temas

- [Consulta del lago de datos](#)
- [Consulta de datos en orígenes de datos remotos](#)
- [Acceso a los datos de otros clústeres de Amazon Redshift](#)
- [Entrenamiento de modelos de machine learning con datos de Amazon Redshift](#)

Consulta del lago de datos

Puede utilizar Amazon Redshift Spectrum para consultar datos en archivos de Amazon S3 sin tener que cargar los datos en tablas de Amazon Redshift. Amazon Redshift proporciona la capacidad SQL diseñada para un procesamiento de análisis en línea (OLAP) rápido de conjuntos de datos muy grandes que se almacenan tanto en clústeres de Amazon Redshift como en lagos de datos de Amazon S3. Puede consultar datos en muchos formatos, incluidos Parquet, ORC, RCFile, TextFile, SequenceFile, RegexSerde, OpenCSV y AVRO. Puede crear esquemas y tablas externos para definir la estructura de los archivos en Amazon S3. A continuación, utiliza un catálogo de datos externo como AWS Glue o su propio metastore de Apache Hive. Los cambios en cualquier tipo de catálogo de datos están disponibles de inmediato en todos sus clústeres de Amazon Redshift.

Después de registrar sus datos con un catálogo de datos de AWS Glue y habilitarlo con AWS Lake Formation, puede consultarlos mediante Redshift Spectrum.

Redshift Spectrum reside en servidores de Amazon Redshift dedicados que no dependen del clúster. Redshift Spectrum inserta muchas tareas que requieren un uso intensivo de cómputo, como el filtrado y la agrupación de predicados, a la capa de Redshift Spectrum. Redshift Spectrum también escala de forma inteligente para aprovechar el procesamiento masivo en paralelo.

Puede particionar las tablas externas en una o más columnas para optimizar el rendimiento de las consultas a través de la eliminación de particiones. Puede consultar y unir las tablas externas con

las tablas de Amazon Redshift. Puede acceder a tablas externas desde varios clústeres de Amazon Redshift y consultar los datos de Amazon S3 desde cualquier clúster de la misma región de AWS. Cuando actualiza los archivos de datos de Amazon S3, los datos están disponibles de inmediato para consultarlos desde cualquiera de los clústeres de Amazon Redshift.

Para obtener más información acerca de Redshift Spectrum, incluido cómo trabajar con Redshift Spectrum y lagos de datos, consulte [Introducción a Amazon Redshift Spectrum](#) en la Guía para desarrolladores de bases de datos Amazon Redshift.

Consulta de datos en orígenes de datos remotos

Puede combinar datos de una base de datos de Amazon RDS, una base de datos de Amazon Aurora o Amazon S3 con datos en la base de datos de Amazon Redshift mediante una consulta federada. Puede utilizar Amazon Redshift para consultar datos operativos de manera directa (sin moverlos), aplicar transformaciones e insertar datos en las tablas de Redshift. Parte del cómputo de las consultas federadas se distribuye a los orígenes remotos de datos.

Para ejecutar consultas federadas, Amazon Redshift establece primero una conexión con el origen remoto de datos. A continuación, Amazon Redshift recupera metadatos sobre las tablas del origen remoto de datos, emite consultas y, luego, recupera las filas de resultados. Luego, Amazon Redshift distribuye las filas de resultados entre los nodos informáticos de Amazon Redshift para continuar su procesamiento.

Para obtener más información sobre cómo configurar su entorno para consultas federadas, consulte uno de los siguientes temas en la Guía para desarrolladores de bases de datos de Amazon Redshift:

- [Introducción al uso de consultas federadas en PostgreSQL](#)
- [Introducción al uso de consultas federadas en MySQL](#)

Acceso a los datos de otros clústeres de Amazon Redshift

Con el uso compartido de datos de Amazon Redshift, puede compartir datos en directo de forma segura y con mayor facilidad en clústeres o cuentas de AWS para fines de lectura. Puede tener acceso instantáneo, pormenorizado y de alto rendimiento a los datos en los clústeres de Amazon Redshift sin necesidad de copiarlos ni moverlos manualmente. Los usuarios pueden ver la información más actualizada y uniforme a medida que se actualiza en los clústeres de Amazon Redshift. Puede compartir datos en niveles distintos, como bases de datos, esquemas, tablas,

vistas (incluidas las vistas normales, de enlace de tiempo de ejecución y materializadas) y las características definidas por el usuario (UDF) de SQL.

El uso compartido de datos de Amazon Redshift es especialmente útil para estos casos de uso:

- Centralización de las cargas de trabajo esenciales para el negocio: utilice un clúster central de servicio ETL (extracción, transformación y carga) que comparta datos con varios clústeres de análisis o de inteligencia empresarial (BI). Este enfoque proporciona aislamiento de la carga de trabajo de lectura y reintegro para las cargas de trabajo individuales.
- Uso compartido de datos entre entornos: comparta datos entre entornos de desarrollo, prueba y producción. Puede mejorar la agilidad del equipo compartiendo datos con diferentes niveles de detalle.

Para obtener más información acerca del uso compartido de datos, consulte [Introducción al uso compartido de datos](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Entrenamiento de modelos de machine learning con datos de Amazon Redshift

Al usar machine learning de Amazon Redshift (Amazon Redshift ML), puede formar un modelo al proporcionar los datos a Amazon Redshift. A continuación, Amazon Redshift ML crea modelos capaces de detectar patrones en los datos de entrada. Luego, puede utilizar estos modelos para generar predicciones respecto de los nuevos datos de entrada sin incurrir en costos adicionales. A través de Amazon Redshift ML, puede formar modelos de machine learning con instrucciones SQL e invocarlos en consultas SQL para generar predicciones. Puede seguir mejorando la precisión de las predicciones mediante el cambio iterativo de los parámetros y la mejora de los datos de formación.

Amazon Redshift ML permite a los usuarios de SQL crear, formar e implementar modelos de machine learning con más facilidad a través de comandos SQL conocidos. Con Amazon Redshift ML, puede utilizar sus datos en clústeres de Amazon Redshift para entrenar modelos con Piloto automático de Amazon SageMaker y obtener el mejor modelo automáticamente. A continuación, puede localizar los modelos y realizar predicciones desde una base de datos de Amazon Redshift.

Para obtener más información sobre Amazon Redshift ML, consulte, [Introducción a Amazon Redshift ML](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Clústeres aprovisionados de Amazon Redshift

Si es la primera vez que utiliza Amazon Redshift, recomendamos que lea las siguientes secciones como ayuda para comenzar a utilizar Amazon Redshift.

Temas

- [Consola de clústeres aprovisionados de Amazon Redshift](#)
- [Conexión a los clústeres aprovisionados de Amazon Redshift](#)
- [Clústeres y carga de datos de Amazon Redshift](#)
- [Tareas comunes de base de datos](#)

Consola de clústeres aprovisionados de Amazon Redshift

Antes de comenzar a configurar un clúster de Amazon Redshift, asegúrese de completar los siguientes requisitos previos:

- [Inscripción en AWS](#)
- [Determinación de reglas de firewall](#)

Inscripción en AWS

Si aún no tiene una Cuenta de AWS, regístrese para obtener una. Si ya tiene una cuenta, puede saltarse este requisito previo y utilizar la cuenta existente.

1. Abra <https://portal.aws.amazon.com/billing/signup>.
2. Siga las instrucciones que se le indiquen.

Parte del procedimiento de registro consiste en recibir una llamada telefónica e indicar un código de verificación en el teclado del teléfono.

Al registrarse para obtener una Cuenta de AWS, se creará un Usuario raíz de la cuenta de AWS. El usuario raíz tendrá acceso a todos los Servicios de AWS y recursos de esa cuenta. Como práctica recomendada de seguridad, [asigne acceso administrativo a un usuario administrativo](#) y utilice únicamente el usuario raíz para ejecutar [tareas que requieren acceso de usuario raíz](#).

Determinación de reglas de firewall

Como parte de este tutorial, deberá especificar un puerto cuando lance el clúster de Amazon Redshift. También deberá crear una regla de entrada en un grupo de seguridad para permitir el acceso al clúster a través del puerto.

Si el equipo cliente está protegido por un firewall, asegúrese de conocer un puerto abierto que pueda utilizar. Con este puerto abierto, puede conectarse al clúster desde una herramienta cliente SQL y ejecutar consultas. Si no conoce un puerto abierto, deberá trabajar con alguna persona que entienda las reglas del firewall de red para encontrar un puerto abierto en su firewall.

Si bien Amazon Redshift utiliza el puerto 5439 de forma predeterminada, la conexión no funciona si dicho puerto no está abierto en el firewall. No se puede cambiar el número de puerto que corresponde al clúster de Amazon Redshift después de crearlo. Por lo tanto, asegúrese de especificar un puerto abierto que funcione en su entorno durante el proceso de lanzamiento.

Este requisito previo solo se aplica cuando traslada sus propios datos a Amazon Redshift. Para obtener más información, consulte [Traer sus propios datos a Amazon Redshift](#).

Después de iniciar sesión en la consola de Amazon Redshift, puede crear y administrar todos los objetos de Amazon Redshift, incluidos clústeres, bases de datos y nodos. También puede ver consultas, ejecutar consultas y realizar otras operaciones de lenguaje de definición de datos (DDL) y de lenguaje de manipulación de datos (DML).

Si es la primera vez que utiliza Amazon Redshift, recomendamos que empiece por dirigirse a las páginas Dashboard (Panel), Clusters (Clústeres), y query editor v2 (editor de consultas v2) para comenzar a utilizar la consola.

Para comenzar a utilizar la consola de Amazon Redshift, vea el siguiente video: [Getting started with Amazon Redshift](#).

A continuación, encontrará descripciones de los elementos del panel de navegación de la consola de Amazon Redshift:

- Amazon Redshift serverless (Amazon Redshift sin servidor): acceda a los datos y analícelos sin necesidad de configurar, ajustar ni administrar clústeres aprovisionados de Amazon Redshift.
- Provisioned clusters dashboard (Panel de clústeres aprovisionados): compruebe las Cluster metrics (Métricas del clúster) e Query overview (Información general sobre consultas) para obtener información de los datos de las métricas (como uso de la CPU) e información de consulta. El uso

de estas opciones puede ayudarlo a determinar si los datos de rendimiento son anormales en un intervalo temporal especificado.

- **Clusters (Clústeres):** vea una lista de clústeres en la cuenta de AWS, elija un clúster para comenzar a realizar consultas o realizar acciones relacionadas con el clúster. También puede crear un clúster nuevo desde esta página.
- **Query editor (Editor de consultas):** ejecute consultas en bases de datos alojadas en el clúster de Amazon Redshift, guarde consultas para reutilizarlas o prográmelas para que se ejecuten en el futuro (solo en el editor de consultas).
- **Editor de consultas v2:** utilice el editor de consultas v2 que es una aplicación de cliente SQL independiente basada en web para crear y ejecutar consultas en el almacenamiento de datos de Amazon Redshift. Puede visualizar los resultados en gráficos y colaborar si comparte sus consultas con otros miembros del equipo.
- **Queries and loads (Consultas y cargas):** obtenga información para referenciar o solucionar problemas, como una lista de consultas recientes y el texto SQL de cada consulta.
- **Datashares (Recursos para compartir datos):** los administradores de una cuenta productora pueden autorizar a las cuentas consumidoras para acceder a los recursos para compartir datos o elegir no autorizar ningún acceso. Para utilizar un recurso para compartir datos autorizado, el administrador de la cuenta consumidora puede asociar el recurso para compartir datos a la cuenta de AWS completa o al espacio de nombres del clúster específico en una cuenta. Un administrador también puede rechazar un recurso para compartir datos.
- **Configurations (Configuraciones):** se puede conectar a clústeres de Amazon Redshift desde herramientas de cliente SQL mediante conexiones Java Database Connectivity (JDBC) y Open Database Connectivity (ODBC). También puede configurar un punto de conexión de Virtual Private Cloud (VPC) administrado por Amazon RedShift. De esta forma, se proporciona una conexión privada entre una VPC basada en el servicio Amazon VPC que contiene un clúster y otra VPC que ejecuta una herramienta cliente.
- **Advisor (Asesor):** obtenga recomendaciones específicas sobre los cambios que puede realizar en su clúster de Amazon Redshift para priorizar sus optimizaciones.
- **AWS Marketplace:** obtenga información sobre otras herramientas o servicios de AWS que funcionan con Amazon Redshift.
- **Alarms (Alarmas):** cree alarmas en métricas de clúster para ver datos de rendimiento y realizar un seguimiento de las métricas durante el lapso de tiempo que especifique.
- **Events (Eventos):** realice un seguimiento de los eventos y obtenga informes sobre información, como la fecha en que se produjo el evento, una descripción o el origen del evento.

- **What's new (Novedades):** vea nuevas características de Amazon Redshift y actualizaciones de productos.

Conexión a los clústeres aprovisionados de Amazon Redshift

Para conectarse a clústeres de Amazon Redshift, desde la página Clusters (Clústeres), amplíe **Connect to Amazon Redshift clusters** (Conectarse a clústeres de Amazon Redshift) y realice alguna de las siguientes operaciones:

- Utilice el editor de consultas v2 para ejecutar consultas en las bases de datos alojadas en el clúster de Amazon Redshift. Después de crear su clúster, puede ejecutar consultas de forma inmediata mediante el editor de consultas v2.

Para obtener más información, vea [Consulta de una base de datos mediante el editor de consultas v2 de Amazon Redshift](#).

- Conéctese a Amazon Redshift desde sus herramientas de cliente mediante controladores JDBC u ODBC a través de la copia de la URL del controlador JDBC u ODBC.

Para trabajar con datos en su clúster, necesita controladores JDBC u ODBC con objeto de establecer una conectividad desde su equipo cliente o instancia. Cifre sus aplicaciones para utilizar operaciones de API de acceso a datos JDBC u ODBC o utilice las herramientas de cliente SQL que sean compatibles con JDBC o bien con ODBC.

Para obtener más información acerca de cómo encontrar la cadena de conexión de clúster, consulte [Obtención de la cadena de conexión a su clúster](#).

- Si la herramienta cliente SQL requiere un controlador, puede descargar un controlador específico del sistema operativo para conectarse a Amazon Redshift desde sus herramientas de cliente.

Para obtener más información acerca de cómo instalar el controlador adecuado para su cliente SQL, consulte [Configuración de una conexión de controlador JDBC versión 2.0](#).

Para obtener más información acerca de cómo configurar una conexión ODBC, consulte [Configuración de una conexión ODBC](#).

Clústeres y carga de datos de Amazon Redshift

En esta sección, encontrará dos tutoriales que le explicarán el proceso de creación de un clúster de Amazon Redshift de muestra. En uno, usará un conjunto de datos de muestra y, en el otro, aportará su propio conjunto de datos.

Temas

- [Uso de un conjunto de datos de muestra](#)
- [Traer sus propios datos a Amazon Redshift](#)

Uso de un conjunto de datos de muestra

En este tutorial, siga el proceso de creación un clúster de Amazon Redshift mediante un conjunto de datos de muestra. Cuando crea un nuevo clúster, Amazon Redshift carga automáticamente el conjunto de datos de muestra. Una vez que el clúster está creado, puede consultar los datos de inmediato.

Antes de comenzar a configurar el clúster de Amazon Redshift, asegúrese de completar [Inscripción en AWS](#) y [Determinación de reglas de firewall](#).

En este tutorial, debe seguir estos pasos:



Step 1:
Create cluster



Step 2:
Try sample queries

Temas

- [Paso 1: Crear un clúster de Amazon Redshift de muestra](#)
- [Paso 2: Probar consultas de ejemplo con los editores de consultas](#)

Important

El clúster de ejemplo que cree se ejecuta en un entorno real. La tarifa bajo demanda es de \$0,25 por hora por el uso del clúster de ejemplo que está diseñado en este tutorial, hasta que lo elimine. Para obtener más información acerca de los precios, consulte [Precios de Amazon](#)

[Redshift](#). Si tiene alguna pregunta o no sabe cómo continuar, puede contactar con el equipo de Amazon Redshift publicando un comentario en nuestro [foro de discusión](#).

Este tutorial no está destinado a los entornos de producción ni analiza las opciones en profundidad. Una vez que haya completado los pasos de este tutorial, puede utilizar [Recursos adicionales](#) para encontrar información más detallada. Esta información puede ayudarle a planificar, implementar y mantener los clústeres, así como a trabajar con los datos del almacén de datos.

Paso 1: Crear un clúster de Amazon Redshift de muestra

Una vez que haya completado los requisitos previos, podrá comenzar a crear su clúster de Amazon Redshift en función de un conjunto de datos de muestra.

Para crear un clúster de Amazon Redshift en función de un conjunto de datos de muestra:

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en <https://console.aws.amazon.com/redshiftv2/>.
2. Para crear un clúster, realice alguna de las siguientes operaciones:
 - En la página de servicio de Amazon Redshift, elija Create cluster (Crear el clúster). Se abrirá la página Create cluster (Crear clúster).
 - En <https://console.aws.amazon.com/redshiftv2/>, elija Panel de clústeres aprovisionados y, a continuación, Crear clúster.
 - En <https://console.aws.amazon.com/redshiftv2/>, elija Clústeres y, a continuación, Crear clúster.
3. En la sección Cluster configuration (Configuración del clúster), especifique un Cluster identifier (Identificador del clúster). Este identificador debe ser único. El identificador debe tener entre 1 y 63 caracteres y utilizar como caracteres válidos letras de la a a la z (solo minúsculas) y el - (guion).

Para este tutorial, escriba **examplecluster**.

4. Si su organización reúne los requisitos necesarios y su clúster se crea en una Región de AWS donde no está disponible Amazon Redshift sin servidor, es posible que pueda crear un clúster en el programa de prueba gratuita de Amazon Redshift. Elige Producción o Prueba gratuita para responder la pregunta ¿Para qué planifica usar este clúster? Si elige Prueba gratuita, cree una configuración con el tipo de nodo dc2.large. Para obtener más información sobre la elección de

una prueba gratuita, consulte [Prueba gratuita de Amazon Redshift](#). Para obtener una lista de las Regiones de AWS donde Amazon Redshift sin servidor está disponible, consulte los puntos de conexión enumerados para la [API de Redshift sin servidor](#) en la Referencia general de Amazon Web Services.

Una vez elegido el tipo de nodo, realice alguna de las siguientes acciones:

- En Sample data (Datos de muestra), elija Load sample data (Cargar datos de muestra) para cargar el conjunto de datos de muestra en el clúster de Amazon Redshift. Amazon Redshift carga el conjunto de datos de muestra Tackit en la base de datos dev y el esquema public predeterminados. Puede comenzar a utilizar el editor de consultas v2 para consultar datos.
- Para llevar sus propios datos al clúster de Amazon Redshift, elija Production (Producción). Luego, en Sample data (Datos de muestra), elija Load sample data (Cargar datos de muestra). Para obtener más información acerca de cómo traer sus propios datos, consulte [Traer sus propios datos a Amazon Redshift](#).

Amazon Redshift carga automáticamente el conjunto de datos de muestra en el clúster de Amazon Redshift de muestra.

5. En la sección Configuración de la base de datos, especifique un valor para Nombre de usuario del administrador. En Contraseña de administrador, elija entre las siguientes opciones:
 - Generar contraseña: use una contraseña generada por Amazon Redshift.
 - Añadir manualmente una contraseña de administrador: use su propia contraseña.
 - Administrar las credenciales de administrador en AWS Secrets Manager: Amazon Redshift usa AWS Secrets Manager para generar y administrar su contraseña de administrador. El uso de AWS Secrets Manager para generar y administrar el secreto de la contraseña conlleva un gasto. Para obtener información sobre precios de AWS Secrets Manager, consulte [Precios de AWS Secrets Manager](#).

Para este tutorial, utilice los valores siguientes:

- Admin user name (Nombre del usuario administrador): ingrese **awsuser**.
- Admin user password (Contraseña del usuario administrador): ingrese un valor para la contraseña.

6. Elija Create cluster.

En este tutorial se utiliza el editor de consultas de Amazon Redshift v2. Puede utilizar este editor para consultar datos inmediatamente, una vez que Amazon Redshift haya terminado de crear el clúster.

También puede elegir otras herramientas de cliente SQL que sean compatibles con los controladores de JDBC u ODBC para trabajar con datos en el clúster. Para obtener más información, consulte [Conexión a un clúster de Amazon Redshift mediante herramientas de cliente SQL](#) en la Guía de administración de Amazon Redshift.

Paso 2: Probar consultas de ejemplo con los editores de consultas

Cuando Amazon Redshift crea el clúster, carga automáticamente el conjunto de datos Tickit de muestra. La creación de clústeres puede tardar algunos minutos en completarse. Una vez finalizada la creación, el estado del clúster se convierte en ACTIVE. Puede ver las tablas de Tickit de muestra desde el conjunto de datos de muestra.

Uso del editor de consultas de

Puede ver las tablas de Tickit de muestra en el editor de consultas v2 si elige el clúster, la base de datos dev y el esquema public.

Después de crear el clúster de Amazon Redshift, en Connect to Amazon Redshift clusters (Conectarse a los clústeres de Amazon Redshift), elija Query data (Consultar datos).

En el editor de consultas v2, conéctese a una base de datos y elija el nombre del clúster en el panel de vista en árbol. Si se le solicita, ingrese los parámetros de conexión.

Cuando se conecta a un clúster y sus bases de datos, debe proporcionar un nombre de Database (Base de datos) y un User name (Nombre de usuario). También debe proporcionar los parámetros necesarios para uno de los siguientes métodos de autenticación:

Nombre de usuario y contraseña de la base de datos

Con este método, también debe proporcionar una Password (Contraseña) para la base de datos a la que se está conectando.

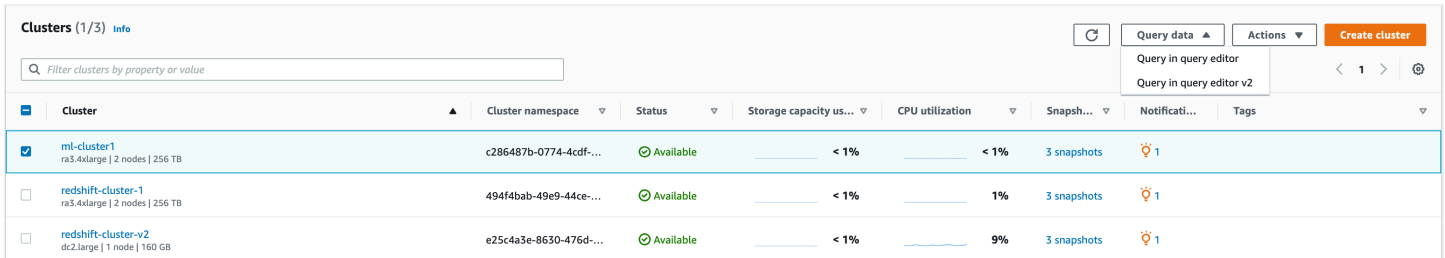
Credenciales temporales

Con este método, el editor de consultas v2 genera una contraseña temporal para conectarse a la base de datos.

Cuando selecciona un clúster con el editor de consultas v2, según el contexto, puede crear, editar y eliminar conexiones mediante el menú contextual (clic derecho).

De manera predeterminada, Amazon Redshift crea una base de datos predeterminada denominada `dev` y un esquema predeterminado denominado `public`. Para ver los archivos de datos individuales del conjunto de datos de muestra, elija un clúster, diríjase al editor de consultas v2 y elija la base de datos `dev`, el esquema `public`, luego `Tables`.

O bien, en el panel de navegación, elija **Clusters (Clústeres)** y el clúster en el que desea consultar datos. A continuación, en **Query data (Consultar datos)**, elija cualquiera de los dos **Query in query editor (Consultar en el editor de consultas)** o **Query in query editor v2 (Consultar en el editor de consultas v2)** para consultar datos en el editor de consultas especificado.



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

Prueba de consultas de ejemplo

Pruebe algunas consultas de ejemplo en el editor de consultas, como se muestra a continuación. Para obtener más información acerca de cómo trabajar con el comando `SELECT`, consulte [SELECT](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

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

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

-- Find events in the 99.9 percentile in terms of all time gross sales.
SELECT eventname, total_price
FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as
      percentile
```



```

FROM (SELECT eventid, sum(pricepaid) total_price
      FROM sales
      GROUP BY eventid)) Q, event E
WHERE Q.eventid = E.eventid
      AND percentile = 1
ORDER BY total_price desc;

```

Una vez que haya completado los pasos de este tutorial, puede utilizar [Recursos adicionales](#) para encontrar información más detallada. Esta información puede ayudarle a planificar, implementar y mantener los clústeres, así como a trabajar con los datos del almacén de datos.

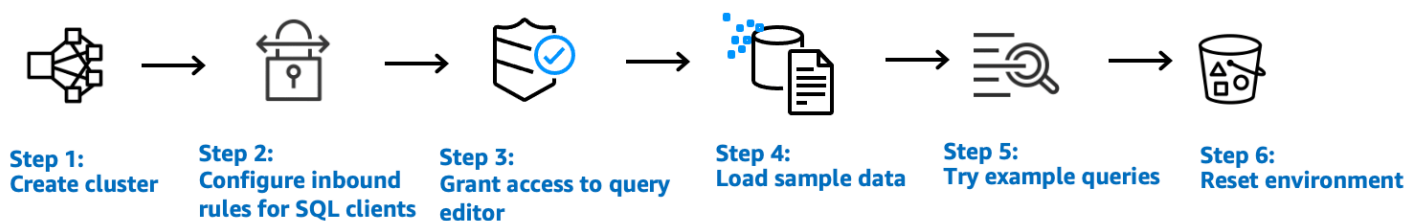
También puede probar el tutorial [Traer sus propios datos a Amazon Redshift](#) para crear un clúster con su propio conjunto de datos.

Traer sus propios datos a Amazon Redshift

En este tutorial, siga el proceso de creación de un clúster de Amazon Redshift mediante su propio conjunto de datos. Puede utilizar este clúster de ejemplo para evaluar el servicio de Amazon Redshift.

Antes de comenzar a configurar el clúster de Amazon Redshift, asegúrese de completar [Inscripción en AWS](#) y [Determinación de reglas de firewall](#).

En este tutorial, debe seguir estos pasos.



⚠ Important

El clúster de ejemplo que cree se ejecuta en un entorno real. La tarifa bajo demanda es de \$0,25 por hora por el uso del clúster de ejemplo que está diseñado en este tutorial, hasta que lo elimine. Para obtener más información acerca de los precios, visite [la página de precios de Amazon Redshift](#). Si tiene alguna pregunta o no sabe cómo continuar, puede contactar con el equipo de Amazon Redshift publicando un comentario en nuestro [foro de discusión](#).

Este tutorial no está destinado a los entornos de producción ni analiza las opciones en profundidad. Una vez que haya completado los pasos de este tutorial, puede utilizar [Recursos adicionales](#) para encontrar información más detallada. Esta información puede ayudarle a planificar, implementar y mantener los clústeres, así como a trabajar con los datos del almacén de datos.

Temas

- [Paso 1: Crear un clúster de Amazon Redshift de muestra](#)
- [Paso 2: Configurar las reglas de entrada para los clientes SQL](#)
- [Paso 3: Otorgar acceso a uno de los editores de consultas y ejecutar consultas](#)
- [Paso 4: Cargar datos desde Amazon S3 en Amazon Redshift](#)
- [Paso 5: Probar consultas de ejemplo mediante el editor de consultas](#)
- [Paso 6: Restablecer su entorno](#)

Paso 1: Crear un clúster de Amazon Redshift de muestra

Para cualquier operación que acceda a datos que estén en otro recurso de AWS, el clúster necesita permiso para acceder en su nombre al recurso y a los datos del recurso. Un ejemplo es el uso de un comando COPY para cargar datos desde Amazon Simple Storage Service (Amazon S3). Estos permisos los concede utilizando AWS Identity and Access Management (IAM). Puede hacerlo a través de un rol de IAM que esté adjunto al clúster. O bien, puede proporcionar la clave de acceso de AWS para un usuario de IAM que tenga los permisos necesarios. Para obtener más información acerca de las credenciales y los permisos de acceso, consulte [Credenciales y permisos de acceso](#).

Para proteger mejor la información confidencial y salvaguardar sus credenciales de acceso de AWS, le recomendamos crear un rol de IAM y adjuntarlo al clúster. Para obtener más información acerca de cómo proporcionar permisos de acceso, consulte [Permisos para acceder a otros recursos de AWS](#).

El clúster que está a punto de crear, se ejecutará en vivo y no en un entorno de pruebas. Debe pagar las tarifas de uso estándar de Amazon Redshift por el clúster hasta que lo elimine. Si completa el tutorial descrito aquí de una sola vez y elimina el clúster al finalizar, los cargos totales son mínimos.

Para crear un clúster de Amazon Redshift

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en <https://console.aws.amazon.com/redshiftv2/>.

⚠ Important

Si utiliza las credenciales de usuario de IAM, asegúrese de que el usuario cuente con los permisos necesarios para realizar las operaciones del clúster. Para obtener más información, consulte [Control de acceso a los usuarios de IAM](#) en la Guía de administración de Amazon Redshift.

2. En la parte superior derecha, elija la Región de AWS en la que desee crear el clúster.
3. En el menú de navegación, elija Clusters (Clústeres) y, a continuación, elija Create cluster (Crear clúster). Se abrirá la página Create cluster (Crear clúster).
4. En la sección Cluster configuration (Configuración del clúster), especifique valores para Cluster identifier (Identificador del clúster), Node type (Tipo de nodo) y Nodes (Nodos):
 - En Cluster identifier (Identificador del clúster), ingrese **examplecluster** para este tutorial. Este identificador debe ser único. El identificador debe tener entre 1 y 63 caracteres y utilizar como caracteres válidos letras de la a a la z (solo minúsculas) y el - (guion).
 - Elija uno de los siguientes métodos para ajustar el tamaño del clúster:

ℹ Note

En el siguiente paso, se considera una región de AWS compatible con tipos de nodo RA3. Para obtener una lista de las regiones de AWS que admiten los tipos de nodo RA3, consulte [Información general sobre los tipos de nodo RA3](#) en la Guía de administración de Amazon Redshift. Para obtener más información sobre las especificaciones de cada tipo y tamaño de nodo, consulte [Detalles acerca de los tipos de nodos](#).

- Si la región de AWS admite tipos de nodo RA3, elija Production (Producción) o Free trial (Prueba gratuita) para responder la pregunta What are you planning to use this cluster for? (¿Para qué planea usar este clúster?).
- Si no sabe cuál sería el tamaño adecuado para el clúster, elija Help me choose (Ayúdeme a elegir). De esta forma, se inicia una calculadora de tamaño que le hace preguntas sobre el tamaño y las características de consulta de los datos que planea almacenar en el almacenamiento de datos.

Si conoce el tamaño requerido para su clúster (es decir, el tipo de nodo y la cantidad de nodos), elija `I'll choose` (Yo elegiré). A continuación, elija el `Node type` (Tipo de nodo) y la cantidad de `Nodes` (nodos) para determinar el tamaño del clúster para la prueba de concepto.

- Para este tutorial, elija `dc2.large` para `Node type` (Tipo de nodo) y 2 para `Nodes` (Nodos).
 - Si eligió `Production` (Producción) para su clúster, realice alguna de las siguientes operaciones:
 - Para usar el conjunto de datos de muestra que proporciona Amazon Redshift, en `Sample data` (Datos de muestra), elija `Load sample data` (Cargar datos de muestra). Amazon Redshift cargará el conjunto de datos de muestra `Tickit` en la base de datos `dev` y el esquema `public` predeterminados.
 - Para traer sus propios datos a Amazon Redshift, continúe con el resto del tutorial.
5. En la sección `Configuración de la base de datos`, especifique un valor para `Nombre de usuario del administrador`. En `Contraseña de administrador`, elija entre las siguientes opciones:
- `Generar contraseña`: use una contraseña generada por Amazon Redshift.
 - `Añadir manualmente una contraseña de administrador`: use su propia contraseña.
 - `Administrar las credenciales de administrador en AWS Secrets Manager`: Amazon Redshift usa `AWS Secrets Manager` para generar y administrar su contraseña de administrador. El uso de `AWS Secrets Manager` para generar y administrar el secreto de la contraseña conlleva un gasto. Para obtener información sobre precios de `AWS Secrets Manager`, consulte [Precios de AWS Secrets Manager](#).

Para este tutorial, utilice los valores siguientes:

- `Admin user name` (Nombre del usuario administrador): ingrese `awsuser`.
 - `Admin user password` (Contraseña del usuario administrador): ingrese un valor para la contraseña.
6. Para este tutorial, cree un rol de IAM y configúrelo como predeterminado para su clúster, como se describe a continuación. Solo se puede configurar un rol de IAM como predeterminado para un clúster.
- a. En `Cluster permissions` (Permisos de clúster), para `Manage IAM roles` (Administrar roles de IAM), elija `Create IAM role` (Crear rol de IAM).

- b. Especifique un bucket de Amazon S3 para que el rol de IAM tenga acceso mediante uno de los siguientes métodos:
- Elija No additional Amazon S3 bucket (Sin bucket adicional de Amazon S3), para permitir que el rol de IAM creado acceda solo a los depósitos de Amazon S3 denominados `redshift`.
 - Elija Any Amazon S3 bucket (Cualquier bucket de Amazon S3), para permitir que el rol de IAM creado acceda a todos los buckets de Amazon S3.
 - Elija Specific Amazon S3 buckets (Buckets específicos de Amazon S3), para especificar uno o más buckets de Amazon S3 para el rol de IAM creado al que puede acceder. A continuación, elija uno o más buckets de Amazon S3 de la tabla.
- c. Elija Create IAM role as default (Crear un rol de IAM como predeterminado). Amazon Redshift crea y configura automáticamente el rol de IAM como predeterminado para su clúster.

Debido a que ha creado su rol de IAM desde la consola, este tiene la política `AmazonRedshiftAllCommandsFullAccess` adjunta. Esto permite a Amazon Redshift copiar, cargar, consultar y analizar datos de los recursos de Amazon en su cuenta de IAM.

Para obtener más información acerca de cómo administrar el rol de IAM predeterminado para un clúster, consulte [Creación de un rol de IAM como predeterminado para Amazon Redshift](#).

7. (Opcional) En la sección Additional configurations (Configuraciones adicionales), desactive Use defaults (Utilizar valores predeterminados) para modificar las opciones de configuración Network and security (Redes y seguridad), Database configuration (Configuración de base de datos), Maintenance (Mantenimiento), Monitoring (Supervisión) y Backup (Copia de seguridad).

En algunos casos, puede crear su clúster con la opción Load sample data (Cargar datos de muestra) y quizá desee activar el enrutamiento mejorado de Amazon VPC. De ser así, el clúster de su nube virtual privada requiere acceso al punto de conexión de Amazon S3 para que se carguen los datos.

Para que el clúster sea accesible públicamente, puede optar por una de estas dos opciones. Puede configurar una dirección de traducción de direcciones de red (NAT) en su VPC para que el clúster acceda a Internet. O bien, puede configurar un punto de conexión de la VPC de Amazon S3 en la VPC. Para obtener más información acerca del enrutamiento mejorado de

Amazon VPC, consulte [Habilitación del enrutamiento mejorado de Amazon VPC](#) en la Guía de administración de Amazon Redshift.

8. Elija Create cluster.

Paso 2: Configurar las reglas de entrada para los clientes SQL

Luego, en este tutorial, puede acceder a su clúster desde una nube virtual privada (VPC) basada en el servicio Amazon VPC. No obstante, si utiliza un cliente SQL desde fuera de su firewall para acceder al clúster, asegúrese de otorgar acceso de entrada.

Puede omitir este paso si su objetivo es acceder al clúster con el editor de consultas de Amazon Redshift desde dentro de su VPC.

Para comprobar el firewall y otorgar acceso entrante a su clúster

1. Compruebe las reglas de su firewall si necesita acceder al clúster desde fuera de un firewall. Por ejemplo, su cliente podría ser una instancia de Amazon Elastic Compute Cloud (Amazon EC2) o un equipo externo.

Para obtener más información, consulte [Reglas del grupo de seguridad](#) en la Guía del usuario de Amazon EC2 para instancias de Linux.

2. Para acceder desde un cliente externo de Amazon EC2, agregue una regla de entrada al grupo de seguridad adjunto a su clúster que permita el tráfico entrante. Las reglas del grupo de seguridad de Amazon EC2 se agregan en la consola de Amazon EC2. Por ejemplo, un CIDR/IP de 192.0.2.0/24 permite a los clientes de ese intervalo de direcciones IP conectarse a su clúster. Descubra cuál es el CIDR/IP correcto para su entorno.

Paso 3: Otorgar acceso a uno de los editores de consultas y ejecutar consultas

Para consultar las bases de datos alojadas en el clúster de Amazon Redshift, tiene dos opciones:

1. Conéctese a su clúster y ejecute consultas en la AWS Management Console con uno de los editores de consultas.

Si utiliza uno de los editores de consultas, no tiene que descargar y configurar una aplicación de cliente SQL.

2. Conéctese al clúster a través de una herramienta de cliente SQL, como SQL Workbench/J. Para obtener más información acerca del uso de SQL Workbench/J, consulte [Conexión al clúster mediante SQL Workbench/J](#) en la Guía de administración de Amazon Redshift.

Utilizar uno de los editores de consultas de Amazon Redshift es la forma más sencilla de ejecutar consultas en bases de datos alojadas por el clúster de Amazon Redshift. Después de crear su clúster, podrá ejecutar consultas de forma inmediata con la consola de Amazon Redshift. Para obtener más información acerca de los aspectos que se deben tener en cuenta al usar el editor de consultas de Amazon Redshift, visite [Consulta de una base de datos mediante el editor de consultas](#) en la Guía de administración de Amazon Redshift.

Otorgamiento de acceso al editor de consultas v2

La primera vez que un administrador configura el editor de consultas v2 para su Cuenta de AWS, este elige la AWS KMS key que se utiliza para cifrar los recursos del editor de consultas v2. De manera predeterminada, se utiliza una clave propia de AWS para cifrar los recursos. O bien, un administrador puede utilizar una clave administrada por el cliente seleccionando el nombre de recurso de Amazon (ARN) para la clave en la página de configuración. Luego de configurar una cuenta, la configuración de cifrado AWS KMS no se puede modificar. Para obtener más información, consulte [Configuración de su Cuenta de AWS](#).

Para obtener acceso al editor de consultas v2, necesita permiso. Un administrador puede adjuntar una de las siguientes políticas administradas por AWS al rol o usuario de IAM para conceder permisos. Estas políticas administradas por AWS se escriben con diferentes opciones que controlan cómo los recursos de etiquetado permiten compartir consultas. Puede utilizar la consola de IAM (<https://console.aws.amazon.com/iam/>) para adjuntar políticas de IAM.

También puede crear su propia política en función de los permisos permitidos y denegados en las políticas administradas proporcionadas. Si utiliza el editor de políticas de la consola de IAM para crear su propia política, elija SQL Workbench como servicio para el que crea la política en el editor visual. El editor de consultas v2 utiliza el nombre del servicio AWS SQL Workbench en el editor visual y el Simulador de políticas de IAM.

Para obtener más información, consulte [Acceso al editor de consultas v2](#).

Uso del editor de consultas v2

Para utilizar el editor de consultas v2 para consultar una base de datos, consulte [Trabajo con el editor de consultas v2](#).

Otorgamiento del acceso al editor de consultas

Para utilizar el editor de consultas de Amazon Redshift, necesita permiso. Para establecer el acceso, asocie las políticas de IAM `AmazonRedshiftQueryEditor` y `AmazonRedshiftReadOnlyAccess` a un rol de IAM y, a continuación, asigne ese rol al usuario que utilice para acceder a su clúster.

Para asociar las políticas de IAM necesarias al editor de consultas

1. Inicie sesión en AWS Management Console y abra la consola IAM en <https://console.aws.amazon.com/iam/>.
2. Elija Roles.
3. Elija el rol que necesita acceso al editor de consultas.
4. En la pestaña Permisos, seleccione Agregar permisos.
5. Seleccione Asociar políticas.
6. Para Nombres de políticas, selecciona `AmazonRedshiftQueryEditor` y `AmazonRedshiftReadOnlyAccess`.
7. Elija Añadir permisos.


Para crear un rol de IAM que permita al clúster de Amazon Redshift comunicarse con otros servicios de AWS en su nombre, siga estos pasos. Los valores utilizados en esta sección son ejemplos; puede elegir valores en función de sus necesidades.

Para crear un rol de IAM que permita a Amazon Redshift el acceso a los servicios de AWS

1. Abra la [consola de IAM](#).
2. Seleccione Roles en el panel de navegación.
3. Elija Crear rol.
4. Elija AWS service (Servicio de AWS) y, a continuación, elija Redshift.
5. En Select your use case (Seleccione su caso de uso), elija Redshift - Customizable (Redshift - Personalizable) y Next: Permissions (Siguiente: permisos). Aparece la página Attach permissions policy (Adjuntar política de permisos).
6. Agregue los permisos `AmazonRedshiftQueryEditor` y `AmazonRedshiftReadOnlyAccess`.

Elija Siguiente: etiquetas.
7. Aparece la página Add tags (Agregar etiquetas). Si lo desea, puede agregar etiquetas. Elija Siguiente: Revisar.

8. En Role Name (Nombre de la función), escriba un nombre para la función; por ejemplo, **RedshiftQueryEditorAccess**. Seleccione Crear rol.
9. El nuevo rol está disponible para todos los usuarios en clústeres que utilizan el rol. Para restringir el acceso solo para usuarios específicos o para clústeres en regiones específicas, edite la relación de confianza del rol. Para obtener más información, consulte [Restricción de acceso a los roles de IAM](#) en la Guía de administración de Amazon Redshift.
10. Asocie el rol al clúster. Puede asociar un rol de IAM a un clúster cuando crea el clúster o agrega el rol a un clúster existente. Para obtener más información, consulte [Asociación de roles de IAM a los clústeres](#) en la Guía de administración de Amazon Redshift.

 Note

Para restringir el acceso a datos específicos, utilice un rol de IAM que otorgue los privilegios mínimos necesarios.

Uso del editor de consultas de

Para utilizar el editor de consultas con el objetivo de consultar una base de datos, vea [Consulta de una base de datos mediante el editor de consultas](#).

Paso 4: Cargar datos desde Amazon S3 en Amazon Redshift

Utilizar uno de los editores de consultas de Amazon Redshift es la forma más sencilla de cargar datos en tablas. Después de crear su clúster, puede cargar datos de Amazon S3 en su clúster mediante la consola de Amazon Redshift.

El uso del editor de consultas v2 simplifica la carga de datos cuando se utiliza el asistente de carga de datos. Para obtener más información, consulte [Carga de sus propios datos desde Amazon S3 en Amazon Redshift mediante el editor de consultas v2](#). Puede utilizar el editor de consultas v2 para crear tablas y cargar los datos. Para obtener más información, consulte [Carga de datos de muestra desde Amazon S3 mediante el editor de consultas](#).

Carga de sus propios datos desde Amazon S3 en Amazon Redshift mediante el editor de consultas v2


Para cargar sus propios datos de Amazon S3 en Amazon Redshift, Amazon Redshift, requiere un rol de IAM que tenga los privilegios necesarios para cargar datos del bucket de Amazon S3 especificado.

Primero, conéctese a una base de datos. A continuación, cree algunas tablas en la base de datos. Luego, cargue sus propios datos desde Amazon S3 en Amazon Redshift. Para obtener más información acerca de cómo trabajar con el editor de consultas v2, consulte [Trabajo con el editor de consultas v2](#) en la Guía de administración de Amazon Redshift.

El comando COPY generado y utilizado en el asistente para cargar datos del editor de consultas v2 admite todos los parámetros disponibles para la sintaxis del comando COPY para cargar datos de Amazon S3. Para obtener información acerca del comando COPY y las opciones que se utilizan para copiar la carga de Amazon S3, consulte [Comando COPY de Amazon Simple Storage Service](#) en la Guía para el desarrollador de base de datos de Amazon Redshift.

Carga de datos de muestra desde Amazon S3 mediante el editor de consultas


En este punto, cuenta con una base de datos denominada dev y está conectado a esta. A continuación, cree algunas tablas en la base de datos, cargue datos en ellas y pruebe una consulta. Para su comodidad, los datos de muestra que se cargan están disponibles en un bucket de Amazon S3.

 Note

Si está utilizando una herramienta de cliente SQL, asegúrese de que el cliente SQL está conectado al clúster.

Una vez completado este paso, podrá realizar lo siguiente:

- probar consultas de ejemplo en [Paso 5: Probar consultas de ejemplo mediante el editor de consultas](#)
- reestablecer su entorno en [Paso 6: Restablecer su entorno](#)
- obtener más información acerca de Amazon Redshift en [Recursos adicionales](#)

 Note

Para probar consultar datos en el editor de consultas sin cargar sus propios datos, elija Load sample data (Cargar datos de muestra) en Sample data (Datos de muestra). Si lo hace, Amazon Redshift cargará automáticamente el conjunto de datos de muestra en el clúster de Amazon Redshift durante la creación del clúster.

Para cargar datos de muestra desde Amazon S3

1. Cree tablas.

Si utiliza el editor de consultas de Amazon Redshift, copie y ejecute de forma individual las siguientes instrucciones de creación de tablas para crear tablas en la base de datos dev. Para obtener más información acerca de la sintaxis, consulte [CREATE TABLE](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

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

```
create table venue(  
  venueid smallint not null distkey sortkey,  
  venue name varchar(100),  
  venue city varchar(30),  
  venue state char(2),  
  venue seats integer);
```

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

```
catdesc varchar(50));
```

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

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

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

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

```
saletime timestamp);
```

2. Cargue datos de muestra desde Amazon S3 mediante el comando COPY.

Note

Le recomendamos utilizar el comando COPY para cargar grandes conjuntos de datos en Amazon Redshift desde Amazon S3 o Amazon DynamoDB. Para obtener más información acerca de la sintaxis de COPY, consulte [COPY](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

- Descargue el archivo [ticketdb.zip](#) que contiene archivos de datos de muestra individuales.
- Descomprima y cargue los distintos archivos en una carpeta `ticket` de su bucket de Amazon S3 en la Región de AWS que utilice.
- Edite los comandos COPY de este tutorial para que se dirijan a los archivos de su bucket de Amazon S3. Para obtener más información acerca de cómo administrar archivos con Amazon S3, consulte [Creación y configuración de un bucket de S3](#) en la Guía del usuario de la consola de Amazon Simple Storage Service.
- Proporcione autenticación para que el clúster acceda a Amazon S3 en su nombre para cargar los datos de muestra. Para este paso, deberá proporcionar autenticación mediante la referencia al rol de IAM que creó y configuró como predeterminado para el clúster en los pasos anteriores.

Los comandos COPY incluyen un marcador de posición para el Nombre de recurso de Amazon (ARN) del rol de IAM, el nombre del bucket y una Región de AWS, como se muestra en el siguiente ejemplo.

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

El comando COPY debería ser similar al siguiente ejemplo.

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

Para cargar los datos de muestra, reemplace `<myBucket>` y `<aws-region>` en los siguientes comandos COPY por sus valores. Si utiliza el editor de consultas de Amazon Redshift, ejecute individualmente los comandos siguientes.

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

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

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

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

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

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

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

Paso 5: Probar consultas de ejemplo mediante el editor de consultas

Ahora, pruebe algunas consultas de ejemplo, como se muestra a continuación. Para obtener más información acerca de cómo trabajar con el comando SELECT, consulte [SELECT](#) en la Guía para desarrolladores de 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;
```

Ha creado correctamente un clúster de Amazon Redshift y ha consultado los datos de su propio conjunto de datos usando el editor de consultas de Amazon Redshift.

Paso 6: Restablecer su entorno

Cuando haya completado este tutorial, le sugerimos restablecer el entorno a su estado anterior eliminando el clúster de muestra. Se le seguirá cobrando por el servicio Amazon Redshift hasta que elimine el clúster.

No obstante, es posible que desee continuar ejecutando el clúster de muestra si planea probar tareas de otras guías de Amazon Redshift.

Para eliminar un clúster

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon Redshift en <https://console.aws.amazon.com/redshiftv2/>.
2. En el menú de navegación, elija Clusters (Clústeres) para mostrar la lista de clústeres.
3. Seleccione el clúster de **examplecluster**. En Actions (Acciones), seleccione Delete (Eliminar). Se abrirá la página Delete cluster (Eliminar clúster).
4. Confirme el clúster que desea eliminar y luego seleccione Delete cluster (Eliminar clúster).

En la página de lista de clúster, se actualiza el estado del clúster a medida que se elimina el clúster.

Luego de completar este tutorial, podrá encontrar más información acerca de Amazon Redshift y los pasos siguientes en [Recursos adicionales](#).

Tareas comunes de base de datos

A continuación, puede encontrar una descripción y una explicación de tareas comunes para comenzar a utilizar bases de datos de Amazon Redshift.

Después de conectarse a la base de datos dev del clúster inicial, puede crear una nueva base de datos. Independientemente de si elige utilizar el conjunto de datos de ejemplo o llevar sus propios datos a Amazon Redshift mientras crea un clúster, Amazon Redshift crea la base de datos dev.

Los ejemplos de esta sección suponen lo siguiente:

- Ha creado un clúster de Amazon Redshift. Para obtener más información, consulte [Clústeres y carga de datos de Amazon Redshift](#).
- Ha establecido una conexión con el clúster desde su herramienta de cliente SQL, como el editor de consultas de la consola de Amazon Redshift. Para obtener más información, consulte [Paso 3: Otorgar acceso a uno de los editores de consultas y ejecutar consultas](#).

Important

El clúster que implementó para este ejercicio se ejecuta en un entorno real. Su clúster acumula cargos en su cuenta de AWS mientras esté en ejecución. Para obtener información acerca de los precios, consulte la [página de precios de Amazon Redshift](#).

Para evitar cargos innecesarios, elimine su clúster cuando termine de usarlo. En el último paso del ejercicio, se explica cómo hacerlo.

Tarea 1: Crear una base de datos

Después de comprobar que su clúster está activo y en ejecución, puede crear su primera base de datos. Esta base de datos es donde crea tablas, carga datos y ejecuta consultas. Un mismo clúster puede alojar distintas bases de datos. Por ejemplo, puede tener una base de datos SALESDB y una base de datos ORDERSDB en el mismo clúster.

Por ejemplo, para crear una base de datos denominada **SALESDB**, ejecute el siguiente comando en la herramienta de cliente SQL.

```
CREATE DATABASE SALESDB;
```

Para este ejercicio, acepte los valores predeterminados. Para obtener más información acerca de opciones de comandos, consulte [CREAR BASE DE DATOS](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Después de haber creado la base de datos SALESDB, puede conectarse a la base de datos nueva desde su cliente SQL. Utilice los mismos parámetros de conexión que utilizó en su conexión actual, pero cambie el nombre de la base de datos por SALESDB.

Tarea 2: Crear un usuario

De manera predeterminada, solo el usuario administrador que haya creado cuando lanzó el clúster tiene acceso a la base de datos inicial del clúster. Para otorgarles acceso a otros usuarios, cree una o más cuentas. Las cuentas de usuario de base de datos se aplican globalmente a todas las bases de datos de un clúster y no pertenecen a bases de datos individuales.

Utilice el comando CREATE USER para crear un nuevo usuario. Al crear un nuevo usuario, debe especificar el nombre del usuario nuevo y una contraseña. Le recomendamos que especifique una

contraseña para el usuario. Debe tener entre 8 y 64 caracteres y debe incluir al menos una letra en mayúsculas, una letra en minúsculas y un número.

Por ejemplo, para crear un usuario llamado **GUEST** con la contraseña **ABCd4321**, ejecute el siguiente comando.

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

Para conectarse a la base de datos SALESDB como el usuario GUEST, utilice la misma contraseña que usó al crear el usuario, como ABCd4321.

Para obtener información acerca de otras opciones de comandos, consulte [CREATE USER](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Tarea 3: Crear un esquema

Después de crear una nueva base de datos, puede crear un nuevo esquema en la base de datos actual. Un esquema es un espacio de nombres que contiene objetos de base de datos nombrados como tablas, vistas y funciones definidas por el usuario (UDF). Una base de datos puede contener uno o más esquemas, y cada esquema pertenece solo a una base de datos. Dos esquemas pueden tener objetos diferentes que comparten el mismo nombre.

Puede crear varios esquemas en la misma base de datos para organizar los datos de la forma que desee o para agrupar los datos funcionalmente. Por ejemplo, puede crear un esquema para almacenar todos los datos transitorios y otro esquema para almacenar todas las tablas de informes. También puede crear esquemas diferentes para almacenar datos relevantes para diferentes grupos empresariales que se encuentran en la misma base de datos. Cada esquema puede almacenar diferentes objetos de base de datos, como tablas, vistas y funciones definidas por el usuario (UDF). Además, puede crear esquemas con la cláusula `AUTHORIZATION`. Esta cláusula otorga la propiedad a un usuario especificado o establece una cuota en la cantidad máxima de espacio en disco que puede utilizar el esquema especificado.

Amazon Redshift crea automáticamente un esquema llamado `public` para cada nueva base de datos. Cuando no especifica el nombre del esquema al crear objetos de base de datos, los objetos entran en el esquema `public`.

Para acceder a un objeto de un esquema, califique el objeto mediante la notación `schema_name.table_name`. El nombre calificado del esquema consiste en el nombre del esquema y el nombre de la tabla separados por un punto. Por ejemplo, es posible que tenga un esquema

de sales que tiene una tabla de price y un esquema de inventory que también tiene una tabla de price. Cuando se refiere a la tabla price, debe calificarla como sales.price o inventory.price.

En el siguiente ejemplo, se crea un esquema denominado **SALES** para el usuario GUEST.

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

Para obtener más información acerca de otras opciones de comandos, consulte [CREATE SCHEMA](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Para ver la lista de esquemas de la base de datos, ejecute el siguiente comando.

```
select * from pg_namespace;
```

El resultado de debería parecerse al siguiente.

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 obtener más información acerca de cómo consultar tablas de catálogo, consulte [Consulta de las tablas de catálogos](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Utilice la instrucción GRANT para conceder permisos a usuarios para los esquemas.

En el siguiente ejemplo, se concede privilegio al usuario GUEST para seleccionar datos de todas las tablas o vistas en SALESCHEMA con la instrucción SELECT.

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

En el siguiente ejemplo, se conceden a la vez todos los privilegios disponibles al usuario GUEST.

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

Tarea 4: Crear una tabla

Después de crear su base de datos nueva, cree tablas para almacenar sus datos. Al crear la tabla, especifique toda la información de las columnas.

Por ejemplo, para crear una tabla llamada **DEMO**, ejecute el siguiente comando.

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

También puede crear una tabla con la notación `schema_name.object_name` para crear la tabla en el esquema de SALES.

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

Para ver y revisar esquemas y sus tablas, puede utilizar el editor de consultas de Amazon Redshift. También puede ver la lista de tablas en esquemas con las vistas del sistema. Para obtener más información, consulte [Tarea 6: Consultar las tablas de sistema](#).

De manera predeterminada, los objetos nuevos de la base de datos, como las tablas, se crean en el esquema predeterminado llamado `public`, que se creó con el clúster. Puede utilizar otro esquema para crear objetos de base de datos. Para obtener más información acerca de esquemas, consulte [Administración de la seguridad de las bases de datos](#) en la Guía para desarrolladores de bases de datos de Amazon Redshift.

Amazon Redshift utiliza las columnas `encoding`, `distkey` y `sortkey` para el procesamiento en paralelo. Para obtener más información acerca del diseño de tablas que incorporan estos elementos, consulte [Prácticas recomendadas de Amazon Redshift para diseñar tablas](#).

Inserción de filas de datos en una tabla

Después de crear una tabla, inserte filas de datos en esa tabla.

Note

El comando [INSERT](#) inserta filas en una tabla. Para cargas masivas estándar, utilice el comando [COPY](#). Para obtener más información, consulte [Uso del comando COPY para cargar datos](#).

Por ejemplo, para insertar valores en la tabla DEMO, ejecute el siguiente comando.

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

Para insertar datos en una tabla que está en un esquema, ejecute el siguiente comando.

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

Selección de datos de una tabla

Después de crear una tabla y rellenarla con datos, utilice una instrucción SELECT para mostrar los datos que tiene la tabla. La instrucción SELECT* devuelve todos los nombres de columnas y los valores de filas de todos los datos de una tabla. El uso de SELECT es una buena forma de verificar que los datos que se hayan agregado recientemente se insertaron en la tabla de forma correcta.

Para ver los datos ingresados en la tabla **DEMO**, ejecute el siguiente comando.

```
SELECT * from DEMO;
```

El resultado debe ser similar a lo siguiente.

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

Para obtener más información acerca del uso de la instrucción SELECT para consultar tablas, consulte [SELECT](#).

Tarea 5: Cargar datos de muestra

En la mayoría de los ejemplos de esta guía, se usa un conjunto de datos de muestra TICKIT. Puede descargar el archivo [tickitdb.zip](#) que contiene archivos de datos de muestra individuales.

Puede cargar los datos de muestra en su propio bucket de Amazon S3.

Para cargar los datos de muestra de la base de datos, cree primero las tablas. A continuación, utilice el comando COPY para cargar las tablas con datos de muestra almacenados en un bucket de Amazon S3. Si desea ver los pasos necesarios para crear tablas y cargar datos de muestra, consulte [Paso 4: Cargar datos desde Amazon S3 en Amazon Redshift](#).

Tarea 6: Consultar las tablas de sistema

Además de las tablas que crea, su base de datos tiene una serie de tablas de sistema. Estas tablas de sistema tienen información relacionada con la instalación y con las diferentes consultas y procesos que se están ejecutando en el sistema. Puede consultar estas tablas de sistema para recopilar información relacionada con su base de datos.

Note

La descripción de cada tabla en esta documentación indica si una tabla es visible para todos los usuarios o si es visible solo para los superusuarios. Inicie sesión como superusuario para consultar las tablas que son visibles solo para los superusuarios.

Amazon Redshift proporciona acceso a los siguientes tipos de tablas de sistema:

- [Tablas de STL](#)

Estas tablas de sistema se generan a partir de archivos de registro de Amazon Redshift para proporcionar un historial del sistema. Las tablas de registro tienen un prefijo STL.

- [Tablas de STV](#)

Estas tablas son tablas de sistema virtuales que tienen instantáneas de los datos actuales del sistema. Las tablas de instantáneas tienen un prefijo STV.

- [Vistas de sistema](#)

Las vistas de sistema presentan un subconjunto de datos que se encuentra en distintas tablas de sistema STL y STV. Las vistas de sistemas tienen un prefijo SVV o SVL.

- [Tablas de catálogos de sistema](#)

Las tablas de catálogos de sistemas almacenan metadatos de esquema, como información acerca de tablas y columnas. Las tablas de catálogos de sistemas tienen un prefijo PG.

Es posible que tenga que especificar el ID del proceso asociado a la consulta para recuperar información de las tablas de sistema sobre una consulta. Para obtener más información, consulte [Determinación del ID de proceso de una consulta en ejecución](#).

Vista de una lista de los nombres de las tablas

Para ver una lista de todas las tablas de un esquema, puede consultar la tabla de catálogo del sistema PG_TABLE_DEF. En primer lugar, puede examinar la configuración de search_path.

```
SHOW search_path;
```

El resultado debería ser similar al siguiente.

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

En el siguiente ejemplo, se agrega el esquema SALES en la ruta de búsqueda y muestra todas las tablas en el esquema SALES.

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

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

select * from pg_table_def where schemaname = 'sales';
schemaname | tablename | column |          type          | encoding | distkey |
sortkey | notnull
-----+-----+-----+-----+-----+-----+-----
sales      | demo      | personid | integer                | az64     | f       |
0 | f
```

```

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

```

En el siguiente ejemplo, se muestra una lista de todas las tablas llamadas DEMO en todos los esquemas de la base de datos actual.

```

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

```

Para obtener más información, consulte [PG_TABLE_DEF](#).

También puede utilizar el editor de consultas v2 para ver todas las tablas de un esquema especificado si elige primero una base de datos a la que desea conectarse.

Ver usuarios

Puede consultar el catálogo PG_USER para ver una lista de todos los usuarios, junto con el ID de usuario (USESYSID) y los privilegios de usuario.

```

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

```


Amazon Redshift utiliza internamente el nombre de usuario `rdsdb` para realizar tareas administrativas y de mantenimiento de rutina. Puede filtrar su consulta para que solo se vean los nombres de usuario definidos por el usuario si agrega `where usesysid > 1` a la instrucción `SELECT`.

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

Vista de consultas recientes

En el ejemplo anterior, el ID de usuario (`USESYSID`) para `adminuser` es 100. Para encontrar las cinco consultas más recientes que ejecutó `adminuser`, puede consultar la vista `SVL_QLOG`.

La vista `SVL_QLOG` es un subconjunto de información más simple extraído de la tabla `STL_QUERY`. Puede utilizar esta vista para encontrar el ID de consulta (`QUERY`) o el ID de proceso (`PID`) para una consulta ejecutada recientemente. También puede utilizar esta vista para verificar cuánto demora en completarse una consulta. `SVL_QLOG` incluye los primeros 60 caracteres de la cadena de la consulta (`SUBSTRING`) para ayudarlo a localizar una consulta específica. Utilice la cláusula `LIMIT` con la instrucción `SELECT` para limitar los resultados a cinco filas.

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

El resultado es similar al siguiente:

```
query| pid | elapsed | substring
-----+-----+-----+-----
+-----+
892 | 21046 | 55868 | SELECT query, pid, elapsed, substring from svl_qlog WHERE
us
620 | 17635 | 1296265 | SELECT query, pid, elapsed, substring from svl_qlog WHERE us
610 | 17607 | 82555 | SELECT * from DEMO;
```

```
596 | 16762 | 226372 | INSERT INTO DEMO VALUES (100);)
```

Determinación del ID de proceso de una consulta en ejecución

Es posible que necesite encontrar el PID para una consulta que sigue en ejecución. Por ejemplo, necesita el PID si debe cancelar una consulta que está tardando mucho en ejecutarse. Puede consultar la tabla de sistema STV_RECENTS para obtener una lista de los ID de proceso de las consultas en ejecución, junto con la cadena de consulta correspondiente. Si su consulta devuelve PID múltiples, puede analizar el texto de la consulta para determinar cuál es el que necesita.

Para determinar el PID de una consulta en ejecución, ejecute la siguiente instrucción SELECT.

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

Tarea 7: Cancelar una consulta

Si ejecuta una consulta que demora mucho tiempo o que consume demasiados recursos del clúster, cancele la consulta. Por ejemplo, cree una lista de vendedores de tickets que incluya el nombre del vendedor y la cantidad de tickets que vendió. La siguiente consulta selecciona los datos de la tabla SALES y de la tabla USERS y une las dos tablas haciendo coincidir los parámetros SELLERID y USERID en la 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;
```

El resultado es similar al siguiente:

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

Se trata de una consulta compleja. Para este tutorial, no necesita preocuparse por cómo se construye esta consulta.

La consulta anterior se ejecuta en segundos y devuelve 2102 filas.

Suponga que se olvida de incorporar la cláusula WHERE.

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

El conjunto de resultados incluye todas las filas de la tabla SALES multiplicadas por todas las filas de la tabla USERS (49 989 x 3766). Esta es una unión cartesiana, la que no se recomienda. El resultado es de más de 188 millones de filas y toma mucho tiempo de ejecución.

Para cancelar una consulta en ejecución, utilice el comando CANCEL con el PID de la consulta.

Para encontrar el ID de proceso, comience una nueva sesión y consulte la tabla STV_RECENTS, tal como se muestra en el paso anterior. En el siguiente ejemplo, se muestra cómo puede hacer que los resultados sean más fáciles de leer. Para ello, utilice la función TRIM para eliminar espacios finales y mostrar solo los primeros 20 caracteres de la cadena de consulta.

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

El resultado es similar al siguiente:

pid	btrim	starttime	substring
610	adminuser	2013-03-28 18:39:49.355918	select sellerid, fir

```
(1 row)
```

Para cancelar la consulta con PID 610, ejecute el siguiente comando.

```
CANCEL 610;
```

Note

El comando CANCEL no detiene una transacción. Para detener o revertir una transacción, debe utilizar el comando ABORT o ROLLBACK. Para cancelar una consulta asociada a una transacción, primero cancele la consulta y, luego, detenga la transacción.

Si la consulta que canceló está asociada con una transacción, utilice el comando ABORT o ROLLBACK para cancelar la transacción y descartar los cambios realizados en los datos:

```
ABORT;
```

A menos que haya iniciado una sesión de superusuario, solo puede cancelar sus propias consultas. Un superusuario puede cancelar todas las consultas.

Cancelación de una consulta desde otra sesión

Si su herramienta de consulta no admite la ejecución de consultas de manera simultánea, inicie otra sesión para cancelar la consulta. Por ejemplo, el editor de consultas que utilizamos en la Guía de introducción a Amazon Redshift no admite varias consultas simultáneas. Para iniciar otra sesión con el editor de consultas, elija File (Archivo), New Window (Nueva ventana) y conéctese usando los mismos parámetros de conexión. Podrá encontrar el PID y cancelar la consulta.

Cancelación de una consulta mediante la cola de superusuario

Si su sesión actual tiene demasiadas consultas ejecutándose de forma simultánea, es posible que no pueda ejecutar el comando CANCEL hasta que termine otra consulta. En ese caso, ejecute el comando CANCEL con una cola de consulta de administración de cargas de trabajo diferente.

Al usar la administración de cargas de trabajo, puede ejecutar consultas en diferentes colas de consulta para no tener que esperar que se complete otra consulta. El administrador de cargas de trabajo crea una cola independiente, denominada cola de superusuario, que puede utilizar para

solucionar problemas. Para utilizar la cola de superusuario, inicie una sesión de superusuario y configure el grupo de consultas como “superusuario” con el comando SET. Después de ejecutar los comandos, restablezca el grupo de consultas con el comando RESET.

Para cancelar una consulta mediante la cola de superusuario, ejecute estos comandos.

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

Tarea 8: Eliminar los recursos

Si implementó un clúster para completar este ejercicio, cuando termine el ejercicio, elimine el clúster. Al eliminar el clúster, se detiene la acumulación de cargos a su cuenta de AWS.

Para eliminar el clúster, siga los pasos detallados en [Eliminación de un clúster](#) en la Guía de administración de Amazon Redshift.

Si desea conservar el clúster, conserve los datos de muestra para referencia. En la mayoría de los ejemplos de esta guía, se usan las tablas que crea en este ejercicio. El tamaño de los datos no tendrá ningún efecto importante en su almacenamiento disponible.

Si desea conservar el clúster y eliminar los datos de muestra, ejecute el comando a continuación para eliminar la base de datos SALESDB.

```
DROP DATABASE SALESDB;
```

Si no creó una base de datos SALESDB o si no desea eliminarla, ejecute los comandos siguientes para eliminar solo las tablas.

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

Información general conceptual sobre Amazon Redshift

Amazon Redshift sin servidor le permite acceder a los datos y analizarlos sin todas las configuraciones de un almacenamiento de datos aprovisionado. Los recursos se aprovisionan automáticamente y la capacidad del almacenamiento de datos se escala de forma inteligente para ofrecer un rendimiento rápido incluso para las cargas de trabajo más exigentes e impredecibles. No incurrirá en gastos cuando el almacenamiento de datos esté inactivo, por lo que solo pagará por lo que utilice. Puede cargar datos y comenzar a realizar consultas de inmediato en el editor de consultas de Amazon Redshift v2 o en su herramienta de inteligencia empresarial (BI) favorita. Disfrute de la mejor relación precio-rendimiento y de las conocidas características de SQL en un entorno sin administración y fácil de utilizar.

Si es la primera vez que utiliza Amazon Redshift, le recomendamos que comience leyendo las siguientes secciones:

- [Información general sobre las características de Amazon Redshift sin servidor](#): en este tema encontrará información general sobre Amazon Redshift sin servidor y sus capacidades clave.
- [Aspectos destacados y precios del servicio](#): en esta página de detalles del producto, puede encontrar información detallada sobre los aspectos destacados y los precios de Amazon Redshift sin servidor.
- [Introducción a Amazon Redshift sin servidor](#): en este tema, puede obtener más información sobre cómo crear un almacenamiento de datos de Amazon Redshift sin servidor y comenzar a consultar datos mediante el editor de consultas v2.

Si prefiere administrar sus recursos de Amazon Redshift manualmente, puede crear clústeres aprovisionados para sus necesidades de consulta de datos. Para obtener más información, consulte [Clústeres de Amazon Redshift](#).

Si su organización reúne los requisitos necesarios y su clúster se crea en una Región de AWS donde no está disponible Amazon Redshift sin servidor, es posible que pueda crear un clúster en el programa de prueba gratuita de Amazon Redshift. Elige Producción o Prueba gratuita para responder la pregunta ¿Para qué planifica usar este clúster? Si elige Prueba gratuita, cree una configuración con el tipo de nodo dc2.large. Para obtener más información sobre la elección de una prueba gratuita, consulte [Prueba gratuita de Amazon Redshift](#). Para obtener una lista de las Regiones de AWS donde Amazon Redshift sin servidor está disponible, consulte los puntos de

conexión enumerados para la [API de Redshift sin servidor](#) en la Referencia general de Amazon Web Services.

A continuación, se presentan algunos conceptos clave de Amazon Redshift sin servidor.

- **Espacio de nombres:** una colección de objetos y usuarios de base de datos. Los espacios de nombres agrupan todos los recursos que se utilizan en Amazon Redshift sin servidor, como esquemas, tablas, usuarios, recursos compartidos de datos e instantáneas.
- **Grupo de trabajo:** una colección de recursos de computación. Los grupos de trabajo alojan recursos de computación que Amazon Redshift sin servidor utiliza para ejecutar tareas de computación. Algunos ejemplos de estos recursos son las unidades de procesamiento de Redshift (RPU), los grupos de seguridad y los límites de uso. Los grupos de trabajo disponen de una configuración de red y de seguridad que puede establecer mediante la consola de Amazon Redshift sin servidor, la AWS Command Line Interface o las API de Amazon Redshift sin servidor.

Para obtener más información sobre la configuración de los recursos de espacios de nombres y de grupos de trabajo, consulte [Uso de espacios de nombres](#) y [Uso de grupos de trabajo](#).

A continuación, se presentan algunos conceptos clave sobre los clústeres aprovisionados de Amazon Redshift:

- **Clúster:** el principal componente de la infraestructura de un almacenamiento de datos de Amazon Redshift es el clúster.

Un clúster se compone de uno o varios nodos de computación. Los nodos informáticos ejecutan el código compilado.

Si un clúster se aprovisiona con dos o más nodos informáticos, un nodo principal adicional coordina los nodos informáticos. El nodo principal gestiona la comunicación externa con aplicaciones, como herramientas de inteligencia empresarial y editores de consultas. La aplicación cliente interactúa de forma directa solo con el nodo principal. Los nodos de computación son transparentes para las aplicaciones externas.

- **Base de datos:** un clúster contiene una o varias bases de datos.

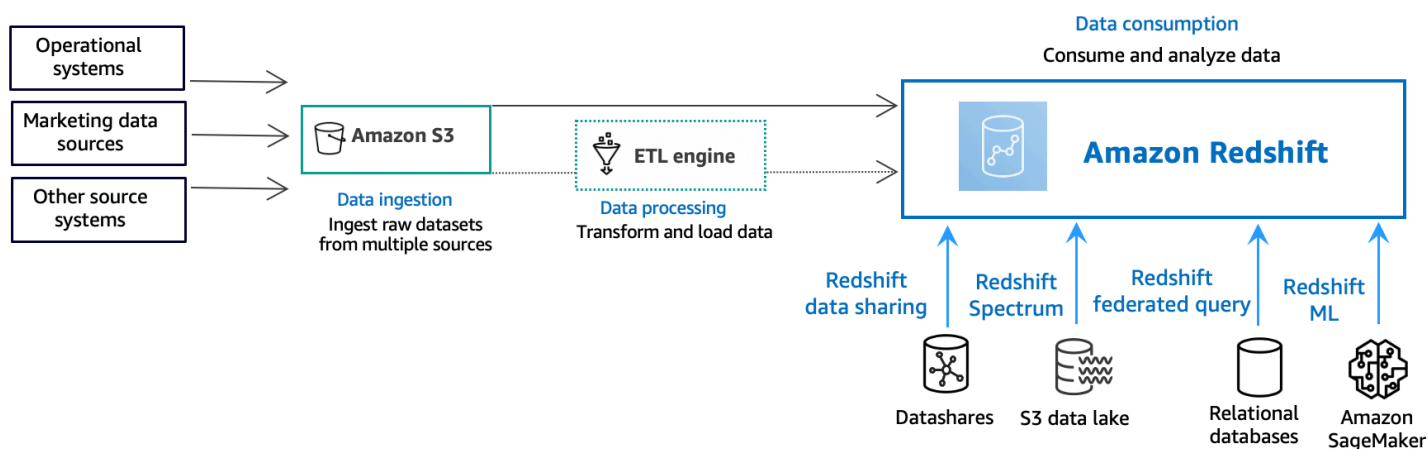
Los datos de usuario se almacenan en una o más bases de datos de los nodos informáticos. El cliente SQL se comunica con el nodo principal y este coordina la ejecución de consultas con los nodos de informática. Para obtener más información sobre los nodos principales y los nodos de

informática, consulte [Arquitectura del sistema de almacenamiento de datos](#). Dentro de una base de datos, los datos de usuario se organizan en uno o más esquemas.

Amazon Redshift es un sistema de administración de base de datos relacional (RDBMS) y es compatible con otras aplicaciones de RDBMS. Aunque proporciona la misma funcionalidad que una RDBMS típica, incluidas las funciones de procesamiento de transacciones en línea (OLTP) como insertar y eliminar datos. Amazon Redshift también está optimizado para análisis por lotes de alto rendimiento y generación de informes de conjuntos de datos.

A continuación, encontrará una descripción del flujo de procesamiento de datos típico en Amazon Redshift, junto con descripciones de distintas partes del flujo. Para obtener más información sobre la arquitectura del sistema de Amazon Redshift, consulte [Arquitectura del sistema de almacenamiento de datos](#).

En el siguiente diagrama, se ilustra un flujo de procesamiento de datos típico en Amazon Redshift.



Un almacenamiento de datos de Amazon Redshift es un sistema de administración y consulta de bases de datos relacionales de clase empresarial. Amazon Redshift admite las conexiones de clientes con muchos tipos de aplicaciones, incluidas las herramientas de análisis, datos, generación de informes e inteligencia empresarial (BI). Cuando ejecuta consultas de análisis, está recuperando, comparando y evaluando grandes cantidades de datos en operaciones de varias etapas para producir un resultado final.

En la capa de captura de datos, distintos tipos de orígenes de datos cargan continuamente datos estructurados, semiestructurados o no estructurados en la capa de almacenamiento de datos. Esta área de almacenamiento de datos sirve como área de almacenamiento provisional que almacena

datos en diferentes estados de preparación para el consumo. Un ejemplo de almacenamiento podría ser un bucket de Amazon Simple Storage Service (Amazon S3).

En la capa opcional de procesamiento de datos, los datos de origen pasan por preprocesamiento, validación y transformación mediante canalizaciones de extracción, transformación, carga (ETL) o extracción, carga, transformación (ELT). Estos conjuntos de datos sin procesar luego se perfeccionan mediante operaciones de ETL. Un ejemplo de motor de ETL es AWS Glue.

En la capa de consumo de datos, los datos se cargan en el clúster de Amazon Redshift, donde puede ejecutar cargas de trabajo de análisis.

Para ver algunos ejemplos de cargas de trabajo analíticas, consulte [Consulta de orígenes de datos externos](#).

Recursos adicionales

Para obtener más información sobre Amazon Redshift sin servidor, le recomendamos que continúe profundizando en los conceptos presentados en esta guía mediante los siguientes recursos de Amazon Redshift:

- Videos de características: estos videos lo ayudarán a obtener información sobre las características de Amazon Redshift.
 - Para conocer Amazon Redshift sin servidor de forma general, vea el siguiente video. [Explicación de Amazon Redshift sin servidor en 90 segundos](#).
 - Para aprender a configurar un almacenamiento de datos sin servidor y empezar a consultar datos, vea el siguiente video. [Introducción a Amazon Redshift sin servidor](#).
- [Guía de administración de Amazon Redshift](#): esta guía se basa en esta Guía de introducción a Amazon Redshift. Proporciona información detallada sobre los conceptos y las tareas de creación, administración y monitoreo de clústeres aprovisionados por Amazon Redshift sin servidor y Amazon Redshift.
- [Guía para desarrolladores de bases de datos de Amazon Redshift](#): esta guía también se basa en esta Guía de introducción a Amazon Redshift. Proporciona a los desarrolladores de bases de datos información detallada acerca del diseño, la creación, la consulta y el mantenimiento de las bases de datos que componen el almacenamiento de datos.
 - [Referencia de SQL](#): en este tema, se describen las referencias de las funciones y los comandos de SQL para Amazon Redshift.
 - [Tablas y vistas de sistema](#): en este tema, se describen las tablas y las vistas de sistema de Amazon Redshift.
- Tutoriales sobre Amazon Redshift: en este tema se muestran tutoriales sobre las características de Amazon Redshift.
 - [Carga de datos desde Amazon S3](#): en este tutorial, se describe cómo cargar datos en las tablas de bases de datos de Amazon Redshift desde los archivos de datos de un bucket de Amazon S3.
 - [Introducción al uso compartido de datos](#): en esta sección se describe cómo compartir datos y acceder a ellos en otros clústeres de Amazon Redshift.
 - [Uso de funciones espaciales de SQL con Amazon Redshift](#): en este tutorial, se muestra cómo utilizar algunas de las funciones espaciales de SQL con Amazon Redshift.

- [Consulta de datos anidados con Amazon Redshift Spectrum](#): en este tutorial, se describe cómo utilizar Redshift Spectrum para consultar datos anidados en los formatos de archivo Parquet, ORC, JSON e Ion mediante el uso de tablas externas.
- [Configuración de colas de administración de carga de trabajo \(WLM\) manual](#): en este tutorial, se describe cómo configurar la administración de carga de trabajo (WLM) manual en Amazon Redshift.
- [Introducción a Amazon Redshift ML](#): en esta sección se describe cómo los usuarios pueden crear, entrenar e implementar modelos de machine learning mediante comandos SQL conocidos.
- [Novedades](#): en esta página web, se enumeran las nuevas características de Amazon Redshift y las actualizaciones de productos.

Historial del documento

Note

Para obtener una descripción de las nuevas características de Amazon Redshift, consulte [Novedades](#).

La siguiente tabla describe los cambios importantes en la documentación en la Guía de introducción a Amazon Redshift.

Cambio	Descripción	Fecha de lanzamiento de la nueva versión
Actualización de la documentación	Actualizamos la guía para incluir nuevas secciones sobre cómo empezar a trabajar con tareas comunes de base de datos, consultar el lago de datos, consultar datos de fuentes remotas, compartir datos y formar modelos de machine learning con datos de Amazon Redshift.	30 de junio de 2021
Nueva característica	Actualizamos la guía para describir el nuevo procedimiento de carga de muestra.	4 de junio de 2021
Actualización de la documentación	Actualizamos la guía para eliminar la consola original de Amazon Redshift y mejorar el flujo de pasos.	14 de agosto de 2020
New console	Actualizamos la guía para describir la nueva consola de Amazon Redshift.	11 de noviembre de 2019
Nueva característica	Actualizamos la guía para describir el procedimiento de lanzamiento rápido del clúster.	10 de agosto de 2018
Nueva característica	Actualizamos la guía para lanzar clústeres desde el panel de Amazon Redshift.	28 de julio de 2015

Cambio	Descripción	Fecha de lanzamiento de la nueva versión
Nueva característica	Actualizamos la guía para usar nuevos nombres para los tipos de nodo.	9 de junio de 2015
Actualización de la documentación	Actualizamos las capturas de pantalla y los procedimientos para configurar grupos de seguridad de Virtual Private Cloud (VPC, Nube privada virtual).	30 de abril de 2015
Actualización de la documentación	Actualizamos las capturas de pantalla y los procedimientos para que coincidan con la consola actual.	12 de noviembre de 2014
Actualización de la documentación	Trasladamos la carga de datos de la información de Amazon S3 a su propia sección y trasladamos la sección de pasos siguientes al paso final para mejorar la visibilidad.	13 de mayo de 2014
Actualización de la documentación	Eliminamos la página de bienvenida e incorporamos el contenido en la página principal Introducción.	14 de marzo de 2014
Actualización de la documentación	Esta es una nueva publicación de la Guía de introducción a Amazon Redshift que aborda los comentarios de los clientes y las actualizaciones del servicio.	14 de marzo de 2014
Nueva guía	Esta es la primera publicación de la Guía de introducción a Amazon Redshift.	14 de febrero de 2013