
BIG DATA

**ANALISIS DE DATOS
REDSHIFT**

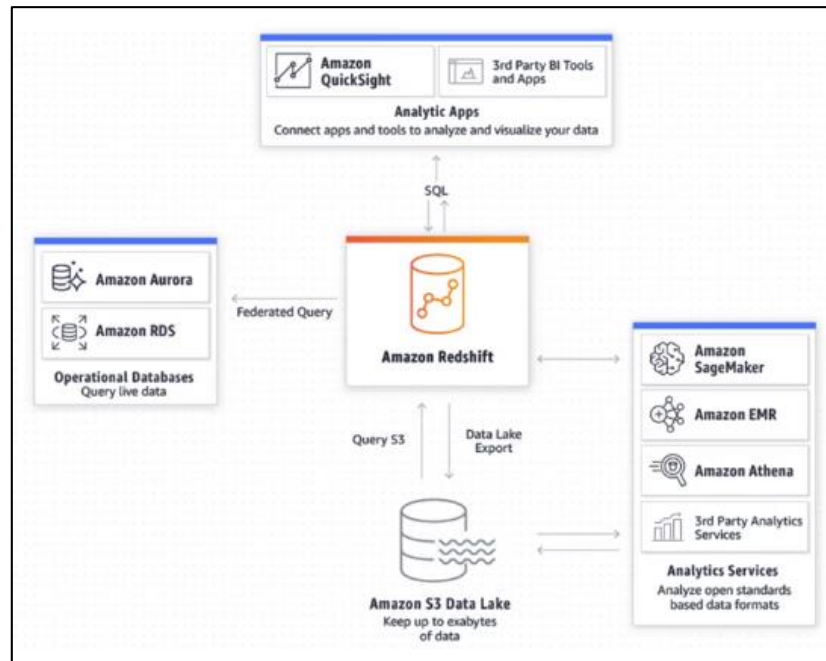
EDUARD LARA

1. INTRODUCCION A REDSHIFT

- ❑ Redshift es un almacén de datos en la nube más rápido y utilizado
- ❑ Ofrece un servicio de Data Warehouse
- ❑ Escala PetaBytes
- ❑ Es rápido y completamente administrado.
- ❑ Está diseñada especialmente para el procesamiento de análisis online OLAP y para el Bussiness Inteligent, BI,
- ❑ No es recomendable el uso para el procesamiento de transaccional online OLTP
- ❑ Es compatible con SQL.
- ❑ Está orientado a columna

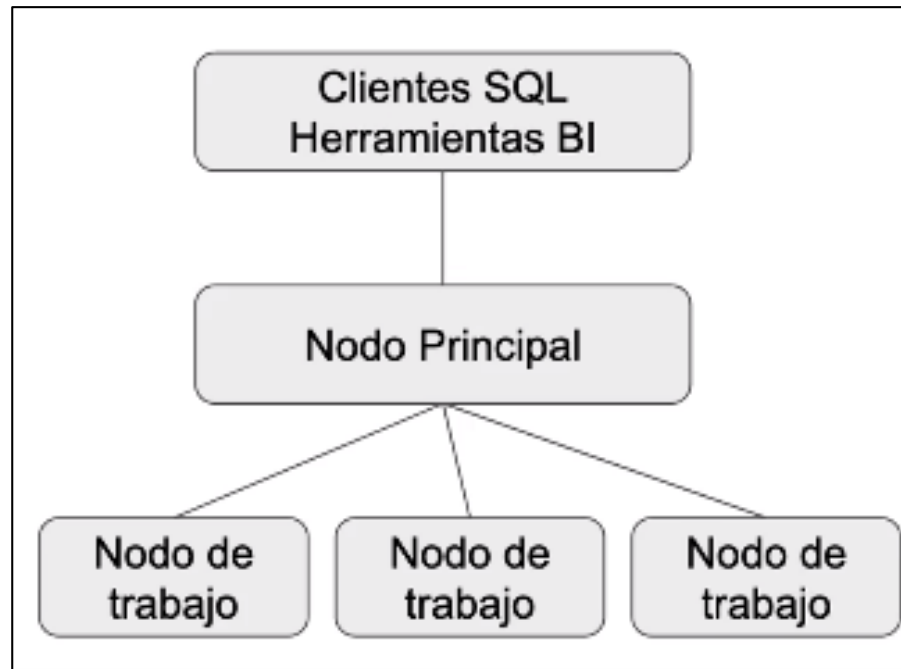
1. INTRODUCCION A REDSHIFT

- ❑ Se pueden aprovisionar múltiples clústers en minutos a bajo precio.
- ❑ Es sencillo de manejar
- ❑ Gestionar grandes cantidades de datos
- ❑ Se integra fácilmente con otros servicios de AWS



2. ARQUITECTURA CLUSTER REDSHIFT

- ❑ Los nodos principales y de trabajo que forman clúster de Redshift están en la misma zona de disponibilidad.
- ❑ RedShift funciona muy bien para grandes volúmenes de datos,
- ❑ Es una base de datos donde el procesamiento se analiza de forma paralela en los distintos nodos de trabajo.



2. ARQUITECTURA CLUSTER REDSHIFT

Nodo principal

- ❑ Facilita la comunicación entre las herramientas BI o clientes SQL con los nodos de trabajo del clúster
- ❑ RedShift ofrece un SQL Endpoint, es decir, una ruta final de acceso a la base de datos de RedShift.
- ❑ Coordina la ejecución de consultas de forma paralela.
- ❑ Cuando un cliente SQL, envía una petición al nodo principal este la distribuye entre los nodos de trabajo que ejecutan dicha consulta y devuelven los resultados al nodo principal que compone la respuesta para enviar al cliente que realizó la consulta SQL.
- ❑ Almacena metadatos de la base de datos de RS.

2. ARQUITECTURA CLUSTER REDSHIFT

Nodo de trabajo

- ❑ Los nodos de trabajo son los que realizan los trabajos SQL y devuelven los resultados al nodo principal.
- ❑ Los nodos de trabajo tienen su propia CPU, memoria y almacenamiento local.
- ❑ Se puede hacer un backup de los nodos de trabajo a Amazon S3.
- ❑ Los nodos de trabajo están divididos en particiones o nodos de partición donde se divide el trabajo a realizar para una ejecución más rápida en paralelo,

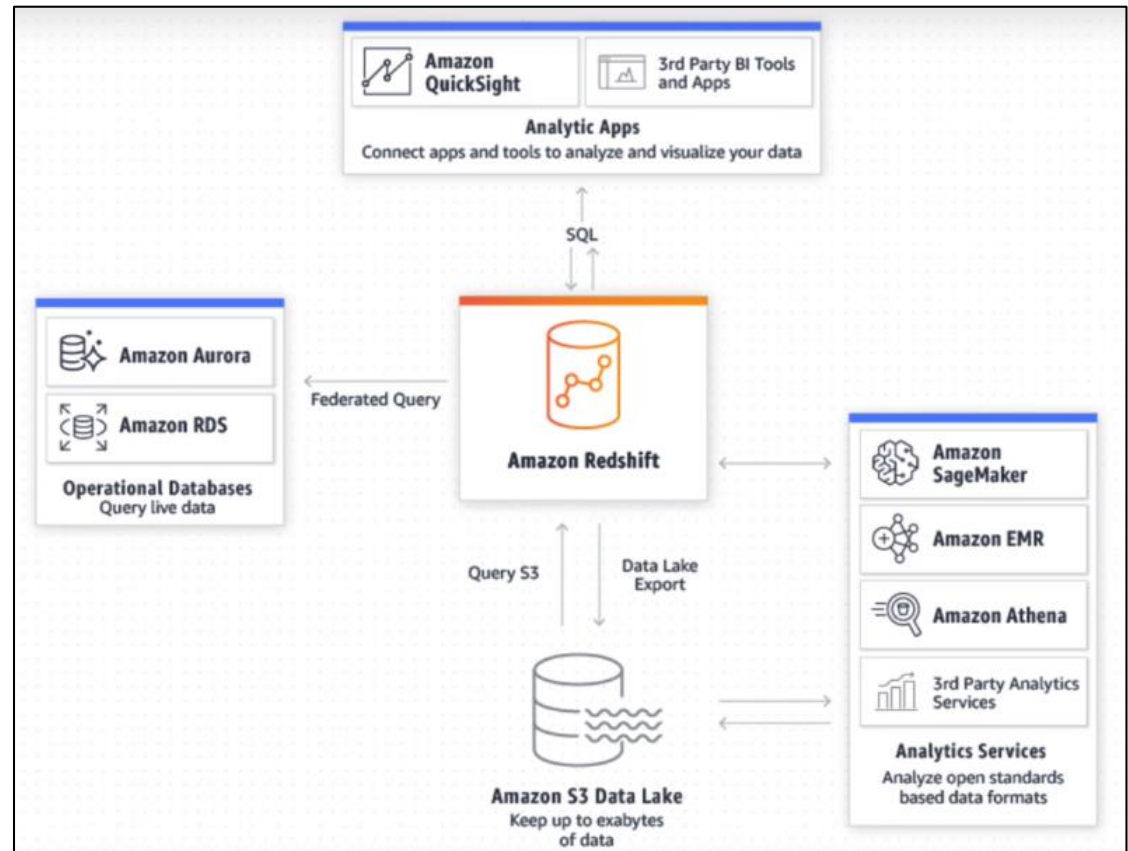
3. INTEGRACIONES REDSHIFT

RedShift se integra fácilmente con otros servicios AWS, como por ejemplo:

- ❑ S3, donde se pueden copiar ficheros a S3 mediante los comandos copy y upload ejecutados de la consola de RedShift
- ❑ DynamoDB, que es una base de datos no SQL
- ❑ EMR
- ❑ EC2
- ❑ Pipeline. Permite mover ficheros entre RedShift y otros servicios AWS mediante Pipeline.
- ❑ Lamba, es decir, un fichero en S3 se puede copiar en RedShift mediante una función Lambda

3. INTEGRACIONES REDSHIFT

- ❑ QuickSight
- ❑ Database migration Service, sistema de migración de bases de datos
- ❑ Kinesis
- ❑ SageMaker
- ❑ Athena



4. BASE DATOS BASADA EN COLUMNAS

- ❑ Una base de datos basada en columnas como RedShift es una base de datos que almacena los datos en columnas en lugar de en filas.
- ❑ Los datos se graban y se leen del disco de forma eficiente para mejorar la rapidez en obtener el resultado de la consulta, mediante la reducción de los datos que son necesarios almacenar en el disco,
- ❑ Tiene algoritmos de compresión que facilitan esta reducción.

4. BASE DATOS BASADA EN COLUMNAS

Diferencias entre una base de datos de columna vs una base de datos de filas.

- ❑ La estructura lógica de los datos es la misma en ambos casos. Los datos son distribuidos en filas y columnas
- ❑ Pero en cuanto a la estructura física es distinta.
- ❑ En la bases de datos por filas los datos se almacenan fila a fila
- ❑ En las base sde datos por columna, los datos se almacenan por columnas completas, bloque a bloque, columna a columna.

4. BASE DATOS BASADA EN COLUMNAS

Beneficios de una base datos de columna

- ☐ Las consultas se hacen sobre unas pocas columnas.
- ☐ Facilita la creación de datos
- ☐ Permite la compresión de los datos almacenados
- ☐ Son más rápidas, ya que las consultas se realizan sobre una menor cantidad de datos y son más baratas, ya que almacena menos información.
- ☐ No es conveniente utilizarlas cuando las cantidades de datos son pequeñas, tampoco con objetos grandes binarios ni con el procesamiento transaccional OLTP.

5. DISEÑO DE UNA TABLA REDSHIFT

Para temas de Big Data es muy importante el diseño de las tablas. Hay distintos estilos de distribución:

- ❑ EVEN → las filas se distribuyen entre los distintos nodos de forma aleatoria. Este es el valor por defecto.
- ❑ KEY → las filas se distribuyen según los valores de una columna. Esto es útil cuando hacemos join de tablas
- ❑ ALL → se distribuye una copia de toda la tabla a cada uno de los nodos. Es útil cuando las tablas no cambian muy a menudo.
- ❑ AUTO → RedShift recibe asigna un estilo de distribución óptimo basado en el tamaño de los datos de la tabla.

5. DISEÑO DE UNA TABLA REDSHIFT

Claves de ordenación de las tablas

- ☐ La clave de ordenación es un tema muy importante para el rendimiento.
- ☐ Los datos se almacenan en el disco ordenados según indica la clave ordenación
- ☐ RedShift utiliza tamaños de bloque de disco de 1 MB
- ☐ Existen mapas de zona, es decir, con valores máximos y mínimos según la clave ordenación para poder realizar búsquedas con mayor rapidez.
- ☐ Existen dos opciones para las claves de ordenación.
 - ☐ Compound
 - ☐ Inteleaved

5. DISEÑO DE UNA TABLA REDSHIFT

Opción Compound para las claves de ordenación

- ❑ En esta opción las columnas son listadas en la definición de la clave de ordenación, es decir, cuando creamos la clave de ordenación, indicamos las columnas.
- ❑ Esta sería la opción por defecto.
- ❑ Es útil cuando hacemos JOINS, ORDER BY, GROUP BY.
- ❑ Con varias columnas en la clave de ordenación, el orden es importante, ya que va ordenando los datos por el orden de las columnas: primero ordena todos los datos por la 1º columna de la clave, luego por la 2º , etc

5. DISEÑO DE UNA TABLA REDSHIFT

Opcion Interleaved para las claves de ordenación

- ☐ En esta opción cada columna de la clave tiene el mismo peso que otras columnas de la clave.
- ☐ Útil para consultas con múltiples filtros en la condición Where
- ☐ El mantenimiento de las tablas es más lento
- ☐ Solo es recomendable utilizarlo en tablas muy grandes de más de 100 millones de filas.

5. DISEÑO DE UNA TABLA REDSHIFT

Selección de columnas para clave de ordenación

Elegiremos aquellas columnas que :

- ☐ Suelen ser de tipo, `TIMESTAMP`, donde podemos hacer ordenaciones por este campo,
- ☐ Columnas utilizadas en condiciones, comparaciones, operaciones de igualdad o `between`
- ☐ Columnas que estén implicadas en operaciones `JOINS` ya que así serán más rápidas.

5. DISEÑO DE UNA TABLA REDSHIFT

Tipos de datos en las tablas RedShift

- ❑ Cada columna de tabla tiene un tipo de dato definido, para saber qué valores podemos guardar en esa columna
- ❑ Los tipos de datos soportados son: INTEGER, DECIMAL, REAL, BOOLEAN, CHAR, VARCHAR, TIMESTAMP, etc
- ❑ Enlace con la lista completa de los tipos de datos

https://docs.aws.amazon.com/redshift/latest/dg/c_Supported_data_types.html

Rango de valores de cada tipo de datos.

- ❑ Un CHAR tiene hasta 4096 bytes
- ❑ VARCHAR tiene un almacén de hasta 65535 bytes

https://docs.aws.amazon.com/redshift/latest/dg/r_Character_types.html

5. DISEÑO DE UNA TABLA REDSHIFT

Comprensión de datos

- ❑ La comprensión de datos reduce la cantidad de datos almacenados en RedShift
- ❑ Cuanto menos datos almacenados menos coste habrá.
- ❑ Cuando las consultas son ejecutadas con menos datos, hay menos E/S y las consultas son más rápidas.
- ❑ ENCODED AUTO, es la opción predeterminada para las tablas en RedShift donde Amazon administra la codificación de la compresión de forma automática.
- ❑ También se puede tener distintos tipos de compresión para cada una de las columnas.
- ❑ Las consultas son más rápidas si tenemos habilitada la compresión en nuestras tablas. Amazon recomienda la compresión automática para sus tablas entre sí

5. DISEÑO DE UNA TABLA REDSHIFT

Restricciones en las columnas de RedShift

- ☐ Reglas que deben cumplir los datos almacenados en cada una de las columnas de la tablas de RedShift
- ☐ Útil para mantener también la integridad de los datos.
- ☐ Está definido a nivel de columna
- ☐ Utilizar las limitaciones en las columnas, permite mejorar el rendimiento de algunas consultas

5. DISEÑO DE UNA TABLA REDSHIFT

Restricciones en las columnas de RedShift

Podemos tener varios tipos:

- ☐ PRIMARY KEY, los valores de esta columna identifican de manera única cada fila de la tabla
- ☐ UNIQUE donde no se permiten valores duplicados para esa columna.
- ☐ NOT NULL. Obliga a insertar algún valor en esa columna, es decir, que no puede estar vacía.
- ☐ REFERENCES. Cuando una columna se relaciona con otra columna, otra tabla,
- ☐ FOREIGN KEY es la clave foránea que tiene relación con la de primaria de otra tabla.

6. GESTION DE LA CARGA DE TRABAJO

- ❑ Consiste en gestionar las consultas cortas y largas para hacer las consultas largas, no atasquen a las cortas
- ❑ Las prioridades de las consultas están basadas en colas.
- ❑ Se crea una serie de colas para gestionar estas consultas y crear una serie de reglas para enrutar las consultas a sus respectivas colas.
- ❑ De esta manera mejoramos el rendimiento y la experiencia del usuario cuando ejecute las consultas.

6. GESTION DE LA CARGA DE TRABAJO

Límites en RedShift

- ☐ Tenemos un límite de 500 conexiones de usuarios concurrentes.
- ☐ Podremos ejecutar hasta 50 consultas simultáneas a la base de datos.
- ☐ Permite un máximo de 8 colas por usuario
- ☐ Existe una cola para el superusuario o el administrador de la base de datos que permite ejecutar alguna acción de emergencia cuando ya se están ejecutando las 50 consultas a la base de datos por otros usuarios.
- ☐ Por defecto, la concurrencia por cola es de 5, aunque se puede modificar.

6. GESTION DE LA CARGA DE TRABAJO

- ❑ Las propiedades estáticas de la gestión de la carga de trabajo que podemos cambiar en la configuración está:
 - ❑ el grupo de usuarios
 - ❑ el grupo de consultas.
- ❑ Las propiedades dinámicas que podemos cambiar en el tiempo de ejecución, sin tener que reiniciar:
 - ❑ La concurrencia,
 - ❑ El porcentaje de memoria utilizado por consulta
 - ❑ El timeout, que sería el tiempo límite para completar una consulta.

7. FORMAS CARGAR DATOS EN REDSHIFT

- ❑ Hay una serie de servicios desde los que podemos importar datos directamente a RedShift mediante el comando *COPY*, que serían S3, EC2, EMR y DynamoDB.
- ❑ Existen otros servicios que necesitan pasar por S3 para importar datos a RedShift, como son Kinesis Firehose y DMS (Database Migration Service)
- ❑ Si llega un fichero a S3, podemos activar la función Lambda para que ejecute el comando *COPY* y copie el fichero a RedShift.

7. FORMAS CARGAR DATOS EN REDSHIFT

Formas de Introducir datos en AWS:

- ☐ AWS CLI, es decir, la terminal de comandos para subir ficheros a S3.
- ☐ Direct Connect, que es un componente que se utiliza para subir ficheros de una forma más rápida.
- ☐ AWS Import/Export
- ☐ Para grandes volúmenes de datos del estilo Terabytes utilizaremos el snowball.

7. FORMAS CARGAR DATOS EN REDSHIFT

Carga mediante comando copy desde S3

- ❑ Desde S3 es la forma más habitual y rápida de subir datos a RedShift mediante el comando copy.
- ❑ Una forma eficiente consiste en dividir el fichero en partes y cargarlos en paralelo desde S3.

```
copy nombre_tabla from 's3://nombre_bucket'  
credentials 'aws_iam_role:rol_con_permisos_operación'
```

- ❑ Los nodos de trabajo se pueden dividir en slices o porciones. Para saber el número de slices que tienen nuestros nodos del cluster RedShift, utilizaremos el comando `Select * from stv_slices`.

7. FORMAS CARGAR DATOS EN REDSHIFT

Carga mediante comando copy desde S3

- ❑ El número de filas de esta consulta indicará el número de slices disponibles. Si tenemos 4 slices, podemos dividir un fichero en 4 u 8 partes para hacer una carga en paralelo.
- ❑ Esto ahorra tiempo y la carga sería más rápida.
- ❑ También podemos utilizar compresión de ficheros para reducir el tamaño de carga.
 - ❑ Los formatos de compresión serían gzip, lzop y bzip2
 - ❑ El tamaño iría desde un mega hasta un giga, que sería el tamaño máximo Después de haber comprimido un fichero.

7. FORMAS CARGAR DATOS EN REDSHIFT

Carga mediante el fichero .manifest

- ❑ La carga de ficheros en RedShift se puede hacer mediante un fichero con extensión .manifest, donde indicaremos los ficheros que queremos subir a S3.
- ❑ Es un fichero de tipo JSON
- ❑ Podemos indicar ficheros de diferentes buckets y también utilizar ficheros con diferentes prefijos.
- ❑ Ejemplo donde indicamos el bucket, el fichero, y si es mandatorio, true o false, si lo vamos a subir o no.

```
{ "entries": [  
  {"URL":"S3://.....1", "MANDATORY":true},  
  {"URL":"S3://.....2", "MANDATORY":true},  
  {"URL":"S3://.....3", "MANDATORY":true},  
]  
}
```

7. FORMAS CARGAR DATOS EN REDSHIFT

- ❑ Comando para copiar o cargar datos via manifiesto:

```
copy nombre_tabla from 's3://..//fichero.manifest'  
credential 'aws_ain_role...' manifest;
```

- ❑ Copiaremos en RedShift los ficheros que indique el fichero .manifest, mediante una serie de entradas

7. FORMAS CARGAR DATOS EN REDSHIFT

Formatos de ficheros

El formato de los ficheros que podemos cargar mediante el comando copy:

- ☐ CSV, aquellos campos separados por comas
- ☐ Delimited, que podemos separar los campos por cualquier separador,
- ☐ FIXED WIDTH aquellos con un ancho fijo, por ejemplo de la 1 a la 10 es un campo de la 10, 10 a la 15 otro campo, etc.
- ☐ formato JSON
- ☐ formato áureo.

7. FORMAS CARGAR DATOS EN REDSHIFT

Tipos de errores en las tablas RedShift

Los tipos errores en las tablas RedShift cuando cargamos los datos podemos tener errores de tipo

- ❑ STL_LOAD ERRORS
- ❑ STL_LOAD ERROR_DETAL

Upsert

- ❑ Es una combinación entre INSERT Y UPDATE, pero en este caso RedShift no lo permite,
- ❑ Como alternativa para utilizar este Upsert sería utilizar una tabla de estado, una tabla intermedia (Staging table) que sirve para simular la operación de Upsert

7. FORMAS CARGAR DATOS EN REDSHIFT

Ficheros encriptados

- ❑ En cuanto al comando copy, podemos subir ficheros encriptados con las siguientes encriptaciones:
- ❑ Encriptación SSE-S3, del lado de servidor con S3 como gestor de claves
- ❑ Encriptación SSE-KMS, del lado del servidor con KMS como gestor de claves
- ❑ Encriptación del lado del cliente utilizando la clave maestra simétrica del cliente
- ❑ El resto de las encriptaciones no están soportadas por RedShift. Sólo permite subir ficheros encriptados de estas tres formas más.

7. FORMAS CARGAR DATOS EN REDSHIFT

Descarga datos comando UPLOAD

- ❑ También podemos descargar datos en S3 mediante el comando upload, es decir, desde RedShift utilizamos el comando Upload para que una select completo sobre una tabla.
- ❑ El resultado de esa shell podemos copiarla en un fichero que esté en un bucket de S3.

`upload ('select ...') to 's3://....' iam_role:aws:iam...`

8. MANTENIMIENTO EN REDSHIFT

Lanzar clúster RedShift

- ❑ Para lanzar un cluster tenemos que hacer una selección de los nodos (DC1, DS2, ...)
 - ❑ El nodo tipo DC1, tiene un alto rendimiento de entrada salida, pero tiene menos almacenamiento por nodo
 - ❑ El nodo tipo DS2 tiene una gran capacidad de almacenamiento.
- ❑ En cuanto al cluster de Redshift, hay que dejar un espacio libre entre 2 y 2,5 veces el tamaño de la tabla más grande para hacer operaciones y que funcione con buen rendimiento.

8. MANTENIMIENTO EN REDSHIFT

Modificar tamaño de un cluster

- ❑ Cuando le damos al botón de Resize del clúster, nos permite cambiar el tipo de nodo, el tipo de clúster y el número de nodos.
- ❑ Cuando Aceptamos, se produce el proceso interno:
 - ❑ hay una notificación de la creación de nuevo Clúster con nuestras características nuevas.
 - ❑ El clastro original se reinicia en modo lectura.
 - ❑ las consultas en el clastro original se pueden hacer también en modo lectura.
 - ❑ Redshift copia los datos del clúster original al nuevo clúster y una vez hecho esto, actualiza el punto final DNS del nuevo Clúster con el valor del punto final DNS del clúster original.
 - ❑ Por último, el clúster original es eliminado y solo quedaría el clúster nuevo

8. MANTENIMIENTO EN REDSHIFT

Comando Vacuum

- ❑ Los bloques de redshift son inmutables, es decir, las actualizaciones se almacenan en nuevos bloques
- ❑ Las filas eliminadas no se borran del disco, simplemente se marcan como eliminadas, por tanto, las viejas filas consumen espacio en disco
- ❑ Para evitar esta pérdida de espacio y degradación de rendimiento, utilizaremos comando vacuum
- ❑ El comando Vacuum permite recuperar el espacio y ordenar las tablas nuevamente.

8. MANTENIMIENTO EN REDSHIFT

Opciones Comando Vacuum

Existen diferentes opciones del comando vacuum:

- ☐ FULL → permite recuperar el espacio y ordenar todas las tablas.
- ☐ FULL + nombre_tabla → Sólo lo haría para esa tabla.
- ☐ SHORT ONLY → Sólo permite hacer la ordenación
- ☐ DELETE ONLY → Solo permite la recuperación de espacio desperdiciado.
- ☐ REINDEX tabla → Reindexa las tablas según las columnas de ordenación interna.

8. MANTENIMIENTO EN REDSHIFT

Comando Vacuum

- ❑ La ejecución del comando vacuum repercute en el rendimiento del sistema, por tanto, sólo debe lanzarse en períodos de poca actividad o ventanas de mantenimiento concretas.
- ❑ El comando vacuum no es recomendable utilizarlo en tablas de más de 700 GB porque es muy lento.
- ❑ En este caso se debe de utilizar el comando Deep Copy, que es utilizar una tabla intermedia de la tabla donde podamos hacer la operación de ordenacion. De esta forma se organizan las tablas y se elimina el espacio que no hace falta.

8. MANTENIMIENTO EN REDSHIFT

Copias de seguridad

- ❑ Se puede hacer de forma automática o manual.
 - ❑ Las copias de seguridad automáticas son continuas o incrementales y se realizan cada 8 horas o 5GB de datos modificados. Se puede cambiar la frecuencia con la que se hace en la configuración de CloudWatch
 - ❑ Las copias de seguridad manuales se pueden hacer en cualquier momento.
- ❑ En la creación de la tabla se puede especificar que nunca se haga backup. Cuando hagamos un backup completo de redshift, esa tabla con no backup, no participaría en esa copia de seguridad

8. MANTENIMIENTO EN REDSHIFT

Copias de seguridad

- ❑ También se pueden hacer copias de seguridad entre regiones o zonas de disponibilidad distintas.
- ❑ Si la tabla está encriptada, podemos hacer una copia de la tabla encriptada.

Restauración de datos

- ❑ Podemos hacer una restauración de datos a nivel de clúster de Redshift, con `restore_from_snapshot`, donde los datos se copiarán en este nuevo cluster.
- ❑ Podemos simplemente restaurar los datos de una única tabla dentro de un snapshot. Nos pedirá los datos del snapshot y los datos de la tabla que queremos restaurar

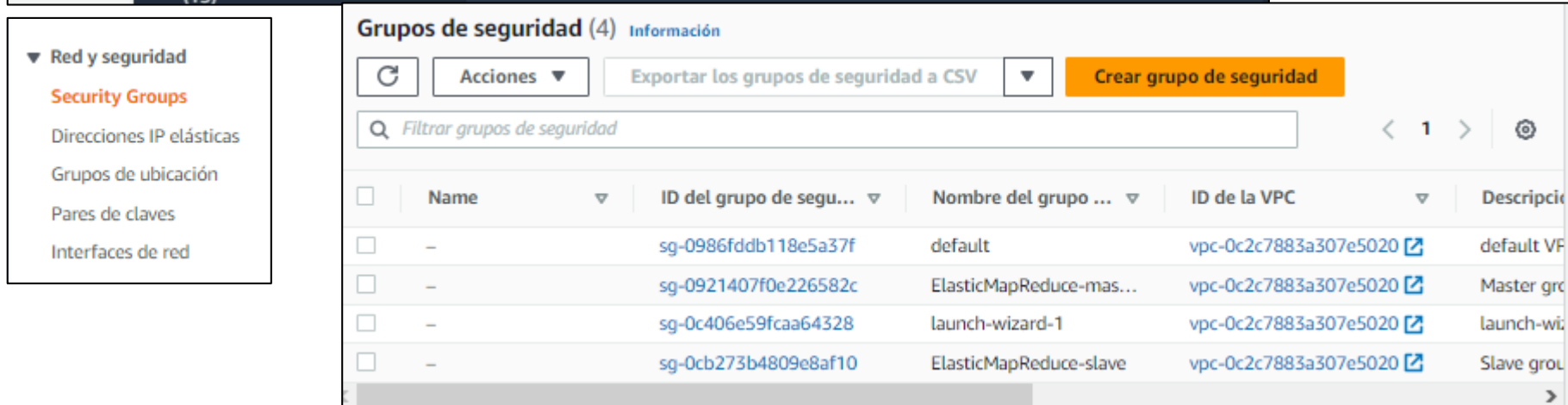
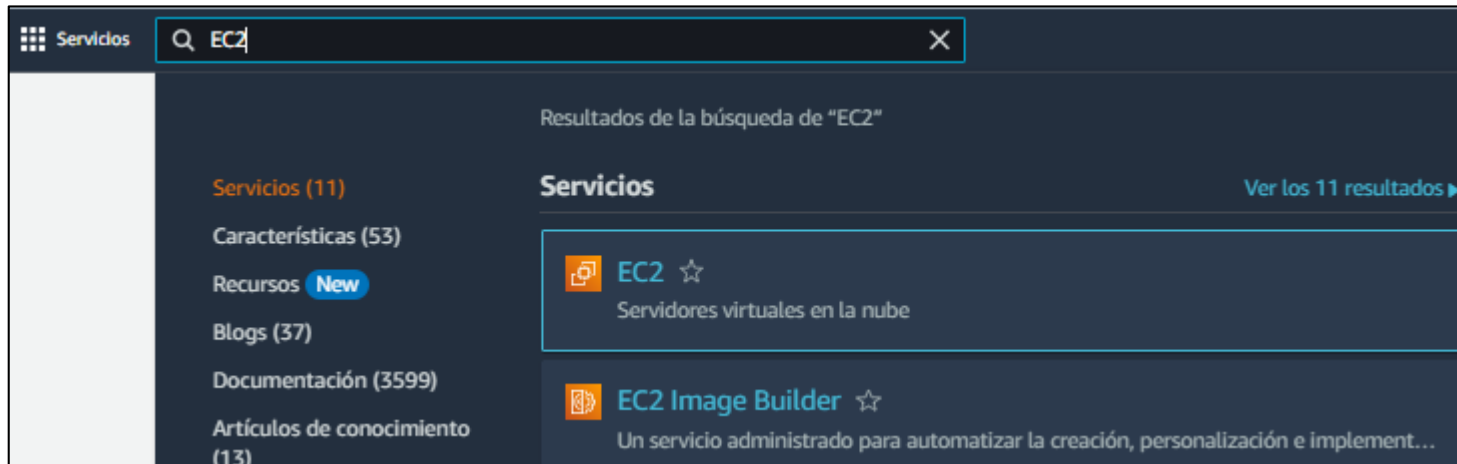
8. MANTENIMIENTO EN REDSHIFT

Monitorizacion

- ❑ Podemos utilizar las métricas de CloudWatch
- ❑ Dentro del cluster RedShift tenemos distintas pestañas como son de rendimiento, de consultas y de carga. Permiten dentro de RedShift monitorizar el rendimiento del cluster, monitorizar las consultas que están en ejecución y también monitorizar la carga de datos.

9. LABORATORIO REDSHIFT

Paso 1. Primero necesitamos crear un grupo de seguridad. Buscamos EC2 y elegimos este servicio. En la parte izquierda, vamos a la parte de Red y seguridad, elegimos la opción de seguridad y grupos



9. LABORATORIO REDSHIFT

Paso 2. Le damos al botón de crear grupos de seguridad. Crearemos un grupo de seguridad de nombre redshift1. En descripción pondremos Grupo de seguridad para redshift.

Elegimos la VPC que tengamos por defecto o podemos crear una nueva.

Crear grupo de seguridad [Información](#)

Un grupo de seguridad actúa como un firewall virtual para que la instancia controle el tráfico de entrada y salida.

Detalles básicos

Nombre del grupo de seguridad [Información](#)

redshift1

El nombre no se puede editar después de su creación.

Descripción [Información](#)

Grupo de seguridad para redshift

VPC [Información](#)

vpc-0c2c7883a307e5020

9. LABORATORIO REDSHIFT

Paso 3. Lo importante es añadir una regla de entrada. Le decimos que el tráfico de entrada que vamos a permitir es Redshift. Si ponemos red nos aparecerá y la seleccionamos. Por defecto nos pondrá que es trafico TCP y el puerto 5439 para el tráfico de entrada de redshift. En Origen elegimos nuestra IP o un rango personalizado de nuestra empresa o ponemos para las pruebas, cualquier IP.

Reglas de entrada

Información

Regla de entrada 1

Eliminar

Tipo

Información

Redshift

Q red

Redshift

Personaliz

Redshift

Protocolo

Información

TCP

Intervalo de puertos

Información

5439

Origen

Información

Q

Descripción: opcional

Información

Agregar regla

Reglas de entrada

Información

Tipo

Información

Redshift

Protocolo

Información

TCP

Intervalo de puertos

Información

5439

Origen

Información

Anywher...

Q

0.0.0.0/0

Agregar regla

9. LABORATORIO REDSHIFT

Paso 4. La regla de salida la dejamos todo el tráfico, como viene por defecto. Si queremos añadir una nueva etiqueta, lo podemos hacer. En todo caso le damos a crear grupo de seguridad

Reglas de salida

Información

Tipo	Protocolo	Intervalo de puertos	Destino	Descripción: opcional
Información	Información	Información	Información	Información
Todo el tráfico	Todo	Todo	Person... 0.0.0.0/0	
<div>Eliminar</div>				
<div>Agregar regla</div>				

Etiquetas: opcional

Las etiquetas son rótulos que se asignan a un recurso de AWS. Cada etiqueta consta de una clave y un valor opcional. Puede utilizarlas para buscar entre sus recursos y filtrarlos, o para hacer un seguimiento de sus costos en AWS.

No hay etiquetas asociadas a este recurso.

Agregar nueva etiqueta

Puede agregar hasta 50 etiquetas más

Cancelar

Crear grupo de seguridad

9. LABORATORIO REDSHIFT

Paso 5. Crearemos este grupo de seguridad que se llama redshift1 que permite tráfico de RedShift desde todas las IP. Esto lo utilizaremos más adelante para crear el clúster.

✓ El grupo de seguridad (sg-091a5c594642b79a2 | redshift1) se ha creado correctamente
▶ Detalles

EC2 > Grupos de seguridad > sg-091a5c594642b79a2 - redshift1

sg-091a5c594642b79a2 - redshift1 Acciones ▼

Detalles

Nombre del grupo de seguridad redshift1	ID del grupo de seguridad sg-091a5c594642b79a2	Descripción Grupo de seguridad para redshift	ID de la VPC vpc-0c2c7883a307e5020
Propietario 307729659735	Número de reglas de entrada 1 Entrada de permiso	Número de reglas de salida 1 Entrada de permiso	

Reglas de entrada

Reglas de salida

Etiquetas

Ahora puede comprobar la conectividad de red con Reachability Analyzer

Ejecutar Reachability Analyzer

Reglas de entrada (1/1)

Administrar etiquetas

Editar reglas de entrada

Filtrar reglas de grupo de seguridad

< 1 >

9. LABORATORIO REDSHIFT

Paso 6. Vamos a IAM a crear un rol nuevo para poder acceder desde RedShift a algún recurso, por ejemplo, en este caso de S3.

The screenshot displays the AWS IAM console. At the top, a search bar contains 'IAM'. Below it, the search results for 'IAM' are shown, including a link to 'IAM' with a star icon and a description 'Administrar el acceso a los recursos de AWS'. The 'Características principales' section lists 'Grupos', 'Usuarios', 'Roles', 'Políticas', and 'Analizador de acceso'. On the left sidebar, the 'Identity and Access Management (IAM)' panel is visible, with a search bar and a list of links: 'Panel', 'Administración del acceso', 'Grupos de usuarios', 'Usuarios', 'Roles', 'Políticas', 'Proveedores de identidad', and 'Configuración de cuenta'. The main content area shows the 'Panel de IAM' with a 'Recomendaciones de seguridad' section containing two items: 'Agregar MFA para el usuario raíz' and 'Actualice los permisos de acceso para las consolas de facturación, administración de costos y cuentas de AWS'.

Resultados de la búsqueda de "IAM"

Servicios Ver los 14 resultados ▶

IAM ☆
Administrar el acceso a los recursos de AWS

Características principales
[Grupos](#) [Usuarios](#) [Roles](#) [Políticas](#) [Analizador de acceso](#)

Identity and Access Management (IAM)

Buscar en IAM

Panel

▼ **Administración del acceso**

- Grupos de usuarios
- Usuarios
- [Roles](#)
- Políticas
- Proveedores de identidad
- Configuración de cuenta

Panel de IAM

Recomendaciones de seguridad 2

⚠ Agregar MFA para el usuario raíz
Inicie sesión como usuario raíz (o póngase en contacto con el administrador) y registre un dispositivo de autenticación multifactor (MFA) para el usuario raíz a fin de mejorar la seguridad de esta cuenta.

⚠ Actualice los permisos de acceso para las consolas de facturación, administración de costos y cuentas de AWS
Nos encontramos en el proceso de sustituir las siguientes acciones de IAM para las consolas de facturación, administración de costos y cuentas por acciones de IAM detalladas: aws-portal:ViewBilling, aws-portal:ModifyBilling, aws-portal:ViewAccount, aws-portal:ModifyAccount, aws-portal:ViewPaymentMethods, aws-portal:ModifyPaymentMethods, aws-portal:ViewUsage, purchase-orders:ViewPurchaseOrders y purchase-orders:ModifyPurchaseOrders. Para asegurarse de no perder el acceso a las características basadas en la consola de facturación, administración de costos y cuentas de AWS, actualice las políticas de IAM existentes de manera que incluyan las nuevas acciones de IAM antes de julio de 2023. Entre los ejemplos de características afectadas se incluyen el Explorador de costos de AWS, AWS Budgets, la Consola de facturación, entre otros. Para obtener más información, visite [blog](#).

9. LABORATORIO REDSHIFT

Paso 7. Vamos a Roles y daremos al botón Crear Rol

IAM > Roles

Roles (21) [Información](#)

Un rol de IAM es una identidad que podemos crear y que tiene permisos específicos con credenciales que son válidas para períodos cortos. Las entidades en las que confía pueden asumir roles.

12

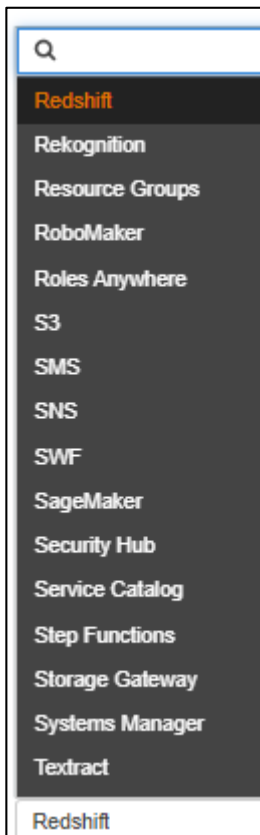
Eliminar

Crear rol

<input type="checkbox"/>	Nombre del rol	Entidades de confianza	Última actividad
<input type="checkbox"/>	AWSServiceRoleForAWSCloud9	Servicio de AWS: cloud9 (Rol vinculado a un servicio)	-
<input type="checkbox"/>	AWSServiceRoleForCloudWatchEvents	Servicio de AWS: events (Rol vinculado a un servicio)	-
<input type="checkbox"/>	AWSServiceRoleForElastiCache	Servicio de AWS: elasticache (Rol vinculado a un servicio)	-
<input type="checkbox"/>	AWSServiceRoleForEMRCleanup	Servicio de AWS: elasticmapreduce (Rol vinculado a un servicio)	Hace 2 días
<input type="checkbox"/>	AWSServiceRoleForOrganizations	Servicio de AWS: organizations (Rol vinculado a un servicio)	-
<input type="checkbox"/>	AWSServiceRoleForSupport	Servicio de AWS: support (Rol vinculado a un servicio)	-
<input type="checkbox"/>	AWSServiceRoleForTrustedAdvisor	Servicio de AWS: trustedadvisor (Rol vinculado a un servicio)	-
<input type="checkbox"/>	c56996a98104513977118f1w307729659735-LambdaSLRRole-DSYACB7CAQ8H	Servicio de AWS: lambda	Hace 17 días

9. LABORATORIO REDSHIFT

Paso 8. Creamos un nuevo rol. El servicio que vamos a utilizar permite que elijamos un caso de uso. En este caso buscamos y elegimos RedShift. Seleccionamos por ejemplo, RedShift Customizable, y hacemos click en Siguiente.



Seleccionar entidad de confianza Información

Tipo de entidad de confianza

☒ **Servicio de AWS**
Permita que servicios de AWS como EC2, Lambda u otros realicen acciones en esta cuenta.

☐ **Cuenta de AWS**
Permitir a las entidades de otras cuentas de AWS que le pertenezcan a usted o a un tercero realizar acciones en esta cuenta.

☐ **Identidad web**
Permite a los usuarios federados por el proveedor de identidad web externo especificado asumir este rol para realizar acciones en esta cuenta.

☐ **Federación SAML 2.0**
Permitir que los usuarios federados con SAML 2.0 a partir de un directorio corporativo realicen acciones en esta cuenta.

☐ **Política de confianza personalizada**
Cree una política de confianza personalizada para permitir que otras personas realicen acciones en esta cuenta.

Caso de uso
Permita que un servicio de AWS, como EC2, Lambda u otros, realicen acciones en esta cuenta.

Casos de uso comunes

☐ **EC2**
Allows EC2 instances to call AWS services on your behalf.

☐ **Lambda**
Allows Lambda functions to call AWS services on your behalf.

Casos de uso para otros servicios de AWS:

Redshift

☒ **Redshift - Customizable**
Allows Redshift clusters to call AWS services on your behalf.

☐ **Redshift**
Allows Redshift clusters to call AWS services on your behalf.

☐ **Redshift - Scheduler**
Allow Redshift Scheduler to call Redshift on your behalf.

Cancelar Siguiente

9. LABORATORIO REDSHIFT

Paso 9. Aquí vamos añadir una política que sea de S3. Buscamos S3 y elegimos AmazonS3FullAccess. Esto lo que hace es permitir que desde RedShift se pueda acceder a S3. Vale, entonces marcamos esta opción, le damos al botón Siguiente

Políticas de permisos (Seleccionado 1/847) Información

Elija una o varias políticas para adjuntarlas al nuevo rol.

9 coincidencias

"S3" X

Quitar filtros

	Nombre de la política	Tipo	Descripción
<input checked="" type="checkbox"/>	AmazonS3FullAccess	Adminis...	Provides full access to all buckets via the AWS Management Console.
<input type="checkbox"/>	AmazonS3ReadOnl...	Adminis...	Provides read only access to all buckets via the AWS Management Console.
<input type="checkbox"/>	AmazonDMSRedsh...	Adminis...	Provides access to manage S3 settings for Redshift endpoints for DMS.
<input type="checkbox"/>	QuickSightAccessF...	Adminis...	Policy used by QuickSight team to access customer data produced by S3 Storage Management Analytics.
<input type="checkbox"/>	AmazonS3Outposts...	Adminis...	Provides full access to Amazon S3 on Outposts via the AWS Management Console.
<input type="checkbox"/>	AmazonS3Outposts...	Adminis...	Provides read only access to Amazon S3 on Outposts via the AWS Management Console.
<input type="checkbox"/>	AmazonS3ObjectLa...	Adminis...	Provides AWS Lambda functions permissions to interact with Amazon S3 Object Lambda. Also grants Lambda permissions to write to Clou...
<input type="checkbox"/>	AWSBackupService...	Adminis...	Policy containing permissions necessary for AWS Backup to restore a S3 backup to a bucket. This includes read/write permissions to all S...
<input type="checkbox"/>	AWSBackupService...	Adminis...	Policy containing permissions necessary for AWS Backup to backup data in any S3 bucket. This includes read access to all S3 objects and...

► Establecer límite de permisos: opcional Información

Establezca un límite de permisos para controlar los permisos máximos que puede tener este rol. Esta configuración no es común, pero puede utilizarla para delegar la administración de permisos a otros usuarios.

Cancelar

Anterior

Siguiente

9. LABORATORIO REDSHIFT

Paso 10. Aquí añadimos un rol de nombre redshift1. Aquí podemos poner un comentario

Paso 1
Seleccionar entidad de confianza

Paso 2
Agregar permisos

Paso 3
Asignar nombre, revisar y crear

Asignar nombre, revisar y crear

Detalles del rol

Nombre del rol

Ingrese un nombre significativo para identificar a este rol.

redshift1

Máximo de 64 caracteres. Utilice caracteres alfanuméricos y '+,=,.,@,-,_'.

Descripción

Agregue una breve explicación para este rol.

Allows Redshift clusters to call AWS services on your behalf.

1000 caracteres como máximo. Utilice caracteres alfanuméricos y '+,=,.,@,-,_'.

Paso 1: seleccionar entidades de confianza

```
1 {  
2   "Version": "2012-10-17",  
3   "Statement": [  
4     {  
5       "Effect": "Allow",  
6       "Action": [  
7         "sts:AssumeRole"  
8       ],  
9       "Principal": {  
10        "Service": [  
11          "redshift.amazonaws.com"  
12        ]  
13      }  
14    ]  
15  }
```

9. LABORATORIO REDSHIFT

Paso 11. Vemos la política elegida `AmazonS3FullAccess` que permite acceder con un control completo a S3. Si quisiéramos por ejemplo cargar un fichero desde un bucket de S3. Crearemos el rol, luego veremos como lo podemos utilizar dentro del clúster de RedShift.

Paso 2: agregar permisos

Editar

Resumen de la política de permisos

Nombre de la política ↗	Tipo ↗	Adjuntado como ↗
AmazonS3FullAccess	Administrada por AWS	Política de permisos

Etiquetas

Agregar etiquetas - optional [Info](#)

Las etiquetas son pares clave-valor que puede agregar a los recursos de AWS para ayudar a identificar, organizar o buscar recursos.

No hay etiquetas asociadas al recurso.

Agregar etiqueta

Puede agregar hasta 50 etiquetas más

CancelarAnteriorCrear rol

9. LABORATORIO REDSHIFT

Paso 12. Nuestro rol ya está creado. Se llama redshift1. Permite desde RedShift acceder a S3 de manera completa.

No se pudo crear el rol redshift1.
User: am:aws:sts::307729659735:assumed-role/voclabs/user1396355=Eduard_Lara is not authorized to perform: iam:CreateRole on resource: am:aws:iam::307729659735:role/redshift1 because no identity-based policy allows the iam:CreateRole action

IAM > Roles > myRedshiftRole

myRedshiftRole

Eliminar

Resumen

Editar

Fecha de creación

April 20, 2023, 18:55 (UTC+02:00)

Última actividad

Ninguno

ARN

am:aws:iam::307729659735:role/myRedshiftRole

Duración máxima de la sesión

1 hora

Permisos

Relaciones de confianza

Etiquetas (1)

Access Advisor

Revocar las sesiones

Políticas de permisos (2) Información

Puede asociar hasta 10 políticas administradas.

Recargar



Simular

Eliminar

Añadir permisos ▼

🔍 Filtre las políticas por propiedad o nombre de política y pulse Intro.

< 1 > ⚙️

<input type="checkbox"/>	Nombre de la política ↗️ ▼	Tipo ▼	Descripción
<input type="checkbox"/>	 AmazonS3ReadOnlyAccess	Administrada por A...	Provides read only access to all buckets via the AWS Management Console.
<input type="checkbox"/>	 RedshiftIAMLabPolicy	Cliente insertado	-

9. LABORATORIO REDSHIFT

Paso 13. Ahora vamos al servicio RedShift y vamos a crear nuestro clúster de la base de datos RedShift. Le damos a crear cluster.

The image shows a screenshot of the AWS Management Console. In the top left, the 'Servicios' (Services) search bar contains the text 'redshift sin servidor'. Below the search bar, a list of search results is displayed, including 'Amazon Redshift' and 'AWS Glue DataBrew'. The 'Amazon Redshift' result is highlighted with a blue border. To the right of the search results, the 'Amazon Redshift' product page is visible. The page features the heading 'Amazon Redshift' and the subheading 'Fast, fully managed, petabyte-scale cloud data warehouse.' Below this, there is a description of Amazon Redshift Serverless and a 'Try Redshift Serverless free trial' button. The page also includes a 'Getting started' section with links to 'Redshift Serverless overview' and 'Evaluation and POC support', and a 'For more granular control' section with a 'Create cluster' button.

aws | Servicios | Q redshift sin servidor X

Identity and Access Management (IAM)

Buscar en IAM

Panel

Administración del acceso

Grupos de usuarios

Resultados de la búsqueda de "reds"

Servicios (27)

Características (63)

Recursos **New**

Blogs (71)

Documentación (4655)

Artículos de conocimiento (14)

Servicios

Ver los 27 resultados ▶

Amazon Redshift ☆

Almacenamiento de datos rápido, sencillo y rentable

AWS Glue DataBrew ☆

Herramienta de

Amazon Redshift

Fast, fully managed, petabyte-scale cloud data warehouse.

Amazon Redshift Serverless combina las mejores capacidades SQL de Amazon Redshift, el rendimiento de precios líder en el sector y la integración perfecta con los lagos de datos con una experiencia sin servidor fácil de utilizar.

Get to powerful insights fast

Get insights from data in seconds without managing data warehouse infrastructure. First-time Redshift Serverless customers receive a \$300 credit to use in their account.

Try Redshift Serverless free trial

Getting started

Redshift Serverless overview

Evaluation and POC support

For more granular control

Create, configure, and manage your cluster to control computing resources.

Create cluster

Funcionamiento

Getting Started with Amazon Redshift Serverless | Amazon

Copiar enl...

Amazon Redshift Serverless

Getting started with Amazon Redshift Serverless

Demo

9. LABORATORIO REDSHIFT

Paso 14. Le ponemos de nombre al clúster redshit1 por ejemplo. Al elegir el tamaño del cluster podemos pedir que nos ayuden a elegir la configuración del nodo. En función del espacio de almacenamiento que necesitamos, si los datos son de tipo time o no, del número de meses que va a contener nuestro datawarehouse o la frecuencia, etc, nos recomendará utilizar la configuración del nodo dc2.large1, que tendrá unas características dedeterminadas precio/año y precio/nodo/hora

Configuración del clúster

Identificador del clúster
Esta es la clave exclusiva que identifica un clúster.

El identificador debe tener entre 1 y 63 caracteres. Los caracteres válidos son a-z (solo minúsculas) y - (guion).

Elegir el tamaño del clúster

☐ Yo elegiré

☒ Ayúdeme a elegir

¿Esta estimación es para datos comprimidos o sin procesar? [Más información](#)

☐ Mi estimación es para datos comprimidos
Seleccione si la estimación es para datos comprimidos después de cargarlos en Amazon Redshift.

¿Cuál es el espacio de almacenamiento estimado que necesita su almacén de datos?

Los datos cargados en Amazon Redshift son, en promedio, 3 veces más pequeños que los formatos de datos abiertos.

120

Tamaño

¿Cuántos datos se consultan a la vez?

☒ Mis datos se basan en el tiempo
Elija si los datos se agregan en orden temporal al almacén de datos. Por ejemplo, mis datos de ventas se agregan cada mes.

☐ Mis datos no se basan en el tiempo
Elija si los datos no tienen una dimensión temporal. Por ejemplo, enumere las partes del inventario por región geográfica.

¿Cuántos meses de datos contiene su almacén de datos?

Calcule el número de meses de datos que tiene previsto almacenar.

12

1 mes 3 meses 12 meses 36 meses Ilimitado

Meses

¿Cuántos meses de datos consulta frecuentemente en su carga de trabajo?

Calcule el número de meses a los que obtiene acceso la carga de trabajo típica cada vez que se ejecuta.

3

1 semana 1 mes 3 meses 12 meses Ilimitado

Meses

9. LABORATORIO REDSHIFT

Paso 15. Si seleccionamos "Yo elegiré" podemos poner el tipo de nodo que nosotros queramos. Elegiremos dc2.large, que tiene hasta 160 GB por nodo y cuesta 0,25\$/nodo/hora.

Crear clúster Información

Looking for free trial? Try Redshift Serverless. First-time Redshift Serverless customers receive a \$300 credit to use in their account.

Launch Redshift Serverless

Configuración del clúster

Identificador del clúster
Esta es la clave exclusiva que identifica un clúster.

El identificador debe tener entre 1 y 63 caracteres. Los caracteres válidos son a-z (solo minúsculas) y - (guion).

Elegir el tamaño del clúster

☒ Yo elegiré

☐ Ayúdeme a elegir

Tipo de nodo Información
Seleccione un tipo de nodo que cumpla los requisitos de CPU, RAM, capacidad

Número de nodos
Escriba el número de nodos que necesita.

Intervalo (1-32)

Seleccionamos dc2.large, y sólo 1 nodo (podemos hasta 32 nodos)

DC2		
dc2.large	Almacenamiento: 160 GB/nodo \$0,25/nodo/hora 2 vCPU (gen 2)	✓
dc2.8xlarge	Almacenamiento: 2,6 TB/nodo \$4,80/nodo/hora 32 vCPU (gen 2)	

9. LABORATORIO REDSHIFT

Paso 16. Nosotros sólo vamos a utilizar el clúster un rato, lo vamos a crear, cargaremos los datos de pruebas, haremos unas consultas SQL y luego al final lo borraremos para no incurrir en mas costas.

En este caso serían 180 euros/mes si lo dejamos activo todo el mes, pero si lo eliminamos, pues saldrá a lo que ponía antes 0,25/hora/nodo Si tuviéramos una hora arrancado el nodo nos costaría 0.25 dólares.

dc2.large | 1 nodo

<p>\$180,00/mes</p> <p>Precio de cómputo bajo demanda estimado</p> <p>Ahorre más del 60 % de sus costos mediante la compra de nodos reservados. Más información</p>	<p>160 GB</p> <p>Almacenamiento comprimido total</p> <p>La capacidad de almacenamiento total del clúster si implementa el número de nodos elegido.</p>
--	---

9. LABORATORIO REDSHIFT

Paso 17. A la hora de mostrar datos podemos cargar en esta base datos, en este cluster de Redshift unos datos que vienen como ejemplo de 28 megas que podemos utilizar. Le vamos a decir que sí para poder hacer consultas sobre esas tablas de ejemplo

Datos de muestra Información

☒ Cargar datos de muestra
Cargue datos de muestra en su clúster de Redshift para comenzar a utilizar el editor de consultas a fin de consultar datos.

Tickit (28 MB)

Tickit es un conjunto de datos de muestra que utiliza una base de datos de muestra llamada TICKIT. Tickit contiene archivos de datos de muestra individuales: dos tablas de hechos y cinco dimensiones.

9. LABORATORIO REDSHIFT

Paso 18. A nivel de configuración de la base de datos, podemos elegir el usuario con el que vamos a acceder a la base de datos y ponerle un password. Tiene que cumplir estas recomendaciones tenerlas en cuenta a la hora de poner la password. Yo pongo la mía, vosotros podéis seleccionar la vuestra, la apuntáis para luego poder utilizarla con el cluster

Configuraciones de la base de datos

Nombre de usuario del administrador
Ingrese un ID de inicio de sesión para el usuario administrador de la instancia de base de datos.

El nombre debe tener entre 1 y 128 caracteres alfanuméricos y no puede ser una palabra reservada [?](#).

☐ **Generar contraseña de forma automática**
Amazon Redshift puede generar una contraseña en su nombre, o bien puede especificar su propia contraseña.

Contraseña de usuario administrador
Debe tener entre 8 y 64 caracteres. Debe contener al menos una letra mayúscula, una letra minúscula y un número. Puede ser cualquier carácter ASCII imprimible excepto "/", "" o "@".

☒ **Mostrar contraseña**

9. LABORATORIO REDSHIFT

Paso 19. En los permisos del cluster vamos a seleccionar el rol que hemos creado antes redshift1, que permitía a RedShift conectarse a S3 con acceso total para recuperar un fichero que este en un bucket. Asociamos el rol a cluster de Redshift lo cual le permitirá ir a un bucket de S3. Cogera un fichero por ejemplo, para cargarlo en la base de datos o lo que sea.

Asociar roles de IAM

Roles de IAM

Elija entre roles de IAM existentes. Puede asociar hasta 50 roles de IAM a este clúster.

Q Buscar un rol de IAM para asociarlo

< 1

☒ Roles de IAM [?](#)

☐ [LabRole](#)

☒ [myRedshiftRole](#)

Cancelar

Asociar roles de

X

❗ Cree un rol de IAM como predeterminado para este clúster que tenga asociada la política [AmazonRedshiftAllCommandsFullAccess](#). Esta política incluye permisos para ejecutar comandos SQL de COPY, UNLOAD y consulta de datos con Amazon Redshift. La política también concede permisos para ejecutar instrucciones SELECT para servicios relacionados, como Simple Storage Service (Amazon S3), Amazon CloudWatch Logs, Amazon SageMaker y AWS Glue.

Roles de IAM asociados (1) [Información](#)

Cree, asocie o elimine un rol de IAM. Puede asociar hasta 50 roles de IAM. También puede elegir un rol de IAM y establecerlo como predeterminado para este clúster.

Establecer valor predeterminado ▼

Administrar roles de IAM ▼

Q Buscar el rol de IAM asociado por nombre, estado o tipo de rol

< 1 >

<input type="checkbox"/>	Roles de IAM ?	Estado	Tipo de rol
<input type="checkbox"/>	myRedshiftRole	No se aplica	--

9. LABORATORIO REDSHIFT

Paso 20. A nivel de Configuraciones adicionales tendremos que ir al Network Security. Aquí elegimos la VPC que tengamos disponible o podemos crear una. A nivel de Grupos de Seguridad, añadiremos redshift1 que creamos antes. Aquí introducimos una regla de entrada para que deje pasar el tráfico de RedShift. En Grupo de subredes de cluster podemos elegir el que tengamos disponible o sino podemos crear uno. El resto lo dejamos por defecto

▼ Red y seguridad [Información](#)

Virtual Private Cloud (VPC)
Esta VPC define el entorno de red virtual para este clúster.

VPC predeterminada
vpc-0c2c7883a307e5020

ⓘ No puede cambiar la VPC asociada a este clúster después de que se haya creado el clúster. [Más información](#) **×**

Grupos de seguridad de la VPC
Este grupo de seguridad de la VPC define qué subredes e intervalos de IP puede utilizar el clúster en la VPC.

Elegir uno o más grupos de seguridad

sg-0986fddb118e5a37f **×** sg-091a5c594642b79a2 **×**

Grupo de subredes de clúster [Información](#)
Elija el grupo de subredes de Amazon Redshift en el que desea lanzar el clúster.

Zona de disponibilidad
Especifique la zona de disponibilidad en la que desea que se cree el clúster. De lo contrario, Amazon Redshift elige una zona de disponibilidad por usted.

No preference

Enrutamiento de VPC mejorado
Habilitar esta opción fuerza el tráfico de red entre el clúster y los repositorios de datos a través de una VPC, en lugar de

9. LABORATORIO REDSHIFT

Paso 21. En Configuraciones de la base de datos, va a crear la base de datos que se va a llamar DEV (se puede cambiar). 5439 va a ser el puerto y lo demás lo dejamos por defecto.

▼ Configuraciones de la base de datos [Información](#)

Nombre de la base de datos
Especifique un nombre de base de datos para crear una base de datos adicional.

El nombre debe tener entre 1 y 64 caracteres alfanuméricos (solo minúsculas) y no puede ser una [palabra reservada](#).

Puerto de la base de datos
Port number where the database accepts inbound connections. The default port is 5439. You can change the port, if needed.

The port must be numeric and between 1150-65535. We recommend that you choose a port number in the ranges 5431-5455 or 8191-8215 to use disaster recovery, Multi-AZ, and cross-VPC endpoint.

Grupos de parámetros
Define los parámetros de la base de datos y las colas de consultas para todas las bases de datos.

Cifrado
Cifre todos los datos en el clúster.

☒ Deshabilitado

☐ Usar AWS Key Management Service (AWS KMS)

☐ Usar un módulo de seguridad de hardware (HSM)

9. LABORATORIO REDSHIFT

Paso 22. Mantenimiento lo vamos a dejar también por defecto. Monitorización también y backup también.

<div>▼ Mantenimiento</div> <div>Periodo de mantenimiento Elija el inicio del intervalo de tiempo semanal durante el cual desea que se apliquen al clúster las notificaciones pendientes.</div> <div><input checked="" type="checkbox"/> Usar el periodo de mantenimiento predeterminado</div> <div>Pista de mantenimiento La pista de mantenimiento controla qué versión del clúster se aplica durante un periodo de mantenimiento.</div> <div><input checked="" type="radio"/> Actual Utilice la versión del clúster aprobada que es más reciente.</div> <div><input type="radio"/> Anterior Utilice la versión del clúster anterior a la versión actual.</div>	<div>▼ Copia de seguridad</div> <div>Las instantáneas automatizadas son copias de seguridad periódicas del clúster.</div> <div>Programación de instantáneas automatizadas Seleccione la frecuencia con la que Amazon Redshift debe generar instantáneas de este clúster.</div> <div>Programa predeterminado ▼</div> <div>Periodo de retención de instantáneas automatizadas Especifique la cantidad de días de retención de las instantáneas automatizadas.</div> <div>1 ▼</div> <div>El periodo de retención debe ser entre 0 y 35 días. Elija cero (0) para no crear instantáneas automatizadas.</div> <div>Periodo de retención de instantáneas manuales Especifique durante cuánto tiempo desea retener la instantánea.</div> <div>Indefinidamente ▼</div> <div>El periodo de retención debe ser entre 1 y 3653 días.</div> <div>Configurar la instantánea entre regiones Configure Amazon Redshift para que copie automáticamente las instantáneas automatizadas y manuales en otra región de AWS. Se aplicarán costos de transferencia de datos cuando las instantáneas se copien en la región de AWS de destino.</div> <div><input checked="" type="radio"/> Deshabilitado</div> <div><input type="radio"/> Habilitado</div> <div>Reubicación de clústeres Habilite la capacidad de reubicar el clúster en otra zona de disponibilidad. Después de habilitar la reubicación, utilice el punto de enlace de la VPC del clúster para determinar la dirección IP del clúster, en lugar de la dirección IP del nodo principal. Puede encontrar el punto de enlace de la VPC en la sección de Red y seguridad de la página de detalles del clúster. Más información</div> <div><input checked="" type="radio"/> Deshabilitado</div> <div><input type="radio"/> Habilitado</div>
---	---

9. LABORATORIO REDSHIFT

Paso 23. Le damos a Crear clúster y comenzará la creación del clúster redshift1 con los datos de pruebas. Esperaremos unos 5 minutos a que finalice la creación y la carga de datos

The screenshot displays the Amazon Redshift console interface during the cluster creation process. The top section features a blue header with four status items: 1. 'Amazon Redshift is creating redshift1' (indicated by a circle icon). 2. 'Carga de datos de muestra' (Sample data loading) with a sub-note: 'Una vez que se crea el clúster, Amazon Redshift comienza a cargar los datos de muestra.' 3. 'Try new Amazon Redshift features in preview.' with a sub-note: 'Create a cluster with preview features. Production use of the cluster is not supported. Use this cluster for testing only.' and a 'Create preview cluster' button. 4. 'Consulte todos los datos en formatos de archivos abiertos sin cargar datos con la arquitectura de la casa del lago de Amazon Redshift. [Comience](#)' (indicated by an information icon). Below this header, the breadcrumb 'Amazon Redshift > Panel de clústeres aprovisionados' is visible. The main content area shows 'Panel de clústeres aprovisionados' with a link to 'Información'. At the bottom right, there are three buttons: 'Pruebe Amazon Redshift Serverless' (with an external link icon), 'Adquirir nodos reservados', and 'Crear clúster' (highlighted in orange).

Amazon Redshift is creating redshift1

Carga de datos de muestra
Una vez que se crea el clúster, Amazon Redshift comienza a cargar los datos de muestra.

Try new Amazon Redshift features in preview.
Create a cluster with preview features. Production use of the cluster is not supported. Use this cluster for testing only. [Create preview cluster](#)

Consulte todos los datos en formatos de archivos abiertos sin cargar datos con la arquitectura de la casa del lago de Amazon Redshift. [Comience](#)

Amazon Redshift > Panel de clústeres aprovisionados

Panel de clústeres aprovisionados [Información](#)

Pruebe Amazon Redshift Serverless [↗](#) Adquirir nodos reservados [Crear clúster](#)

9. LABORATORIO REDSHIFT

Paso 24. Ha finalizado la creación del clúster de manera exitosa y también ha sido capaz de cargar los datos de pruebas. Nuestro clúster se llama redshift1. Indica el tipo de nodo, 1 nodo con 160GB. El estado está disponible.

✓ **Successfully created cluster redshift1**
To manage clusters or connect to the cluster, go to the [redshift1](#) details page.

[View connection credentials](#) ✕

✓ **Successfully loaded sample data**
To query the sample data, use the Redshift query editor.

[Query data](#) ✕

Amazon Redshift > Clústeres

[En mi cuenta](#) | [De otras cuentas](#)

▼ Conectarse a clústeres de Redshift

Consultar datos con el editor de consultas de Redshift

Utilice el editor de consultas v2 para ejecutar consultas en su clúster de Redshift.

[Datos de consulta](#)

Trabajar con las herramientas de cliente

Puede conectarse a Amazon Redshift desde sus herramientas de cliente, como clientes de SQL, herramientas de inteligencia empresarial (BI) y herramientas de extracción, transformación y carga (ETL), mediante el uso de controladores JDBC u ODBC.

Clúster

redshift1

[Copiar URL de JDBC](#) [Copiar URL de ODBC](#)

Elegir el controlador JDBC u ODBC

Utilice controladores JDBC u ODBC para conectarse a Amazon Redshift desde sus herramientas de cliente, como clientes de SQL, herramientas de inteligencia empresarial y ETL. Recomendamos que utilice los nuevos controladores específicos de Amazon Redshift para mejorar el rendimiento y la escalabilidad.

Controlador

JDBC 4.2 sin AWS SDK (.jar)

[Descargar el controlador](#)

Clústeres (1) [Información](#)

[Filtrar clústeres por propiedad o valor](#)

☐

Clúster

▲

☐

Estado

▼

☐

Espacio de nombre...

▼

☐

Zona de disponibilidad

▼

☐

Multi-AZ

▼

☐

Capacidad de alma...

▼

☐

Utilización de la CPU

▼

☐

Instant...

▼

☐

Notificaci...

▼

<input type="checkbox"/>	redshift1 dc2.large 1 nodo 160 GB	✓ Available	ae993c6f-55bd-4751-...	us-east-1e	No	< 1 %	9 %	2 instantáneas	1
--------------------------	--	-------------	------------------------	------------	----	-------	-----	----------------	---

9. LABORATORIO REDSHIFT

Paso 25. Simplemente pulsando sobre redshift1 podemos ver las características principales. Vale, entonces podemos ver el nombre del clúster, el estado en el que está, cuando se ha creado, el tipo de nodo, cuántos nodos hay en el cluster. el endpoint que se utiliza para grabar la información de redshift

Amazon Redshift > Clústeres > redshift1

redshift1

Acciones ▼ Editar Agregar integración de socios Datos de consulta ▼

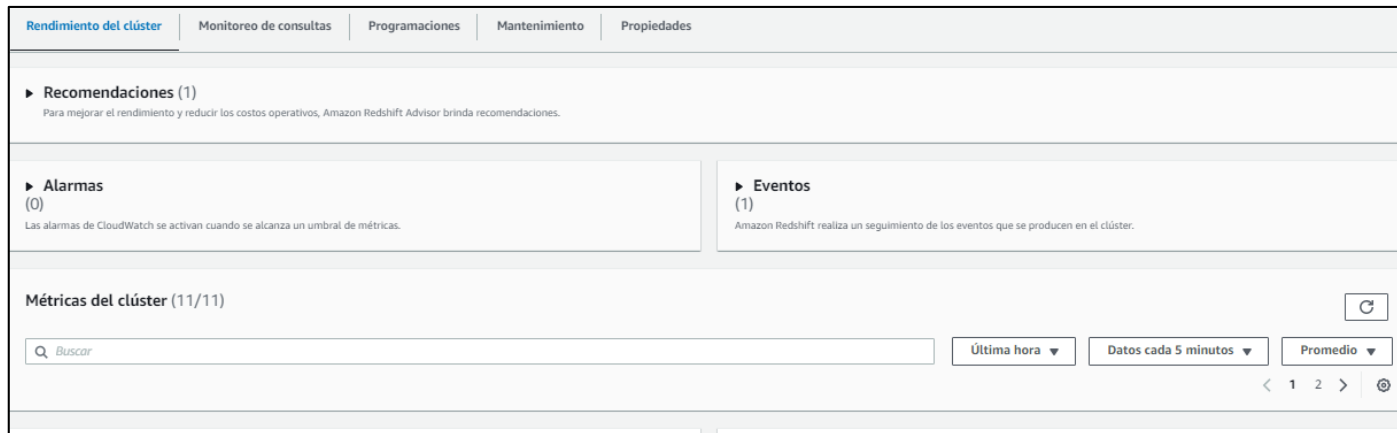
Información general

Identificador del clúster redshift1	Estado 🟢 Available	Tipo de nodo dc2.large	Punto de enlace redshift1.ccx4gbcdfojz.us-east-1.redshift.amazonaws.com...
Espacio de nombres de clúster ae993c6f-55bd-4751-adf1-d63d12836482	Fecha de creación May 08, 2023, 00:17 (UTC+02:00)	Número de nodos 1	URL de JDBC jdbc:redshift://redshift1.ccx4gbcdfojz.us-east-1.redshift.a...
Configuración del clúster Production	Almacenamiento utilizado 0.22 % (0.34 de 160 GB utilizado)		URL de ODBC Driver={Amazon Redshift (x64)}; Server=redshift1.ccx4gb...
	Multi-AZ No		

Rendimiento del clúster | Monitoreo de consultas | Programaciones | Mantenimiento | Propiedades

9. LABORATORIO REDSHIFT

Paso 26. En rendimiento del cluster, podemos ver métricas de cluster, utilización CPU etc. Ahora mismo no hay porque acabamos de crear el cluster. Podemos hacer monitorización de consultas que vayamos realizando, programaciones, el detalle de mantenimiento, donde lo hemos puesto todo por defecto. Y las propiedades nuevamente.



Rendimiento del clúster Monitoreo de consultas Programaciones Mantenimiento Propiedades			
Detalles de mantenimiento Amazon Redshift aplica el mantenimiento al clúster de forma periódica. Más información			
Versión actual del clúster 1.0.49780 🔗		Pista de mantenimiento Current	
Permitir la actualización de la versión Si		Periodo de mantenimiento Cada Wednesday de 09:30 AM a 10:00 AMCEST	
Estado de la versión Actualizado		Último mantenimiento -	
		Posponer el periodo de mantenimiento Deshabilitado	
		Próxima programación En 2 days	

9. LABORATORIO REDSHIFT

Paso 27. En el Panel de clústeres aprovisionados tenemos información del numero total nodos, nodos bajo demanda, nodos reservados, etc

Amazon Redshift > Panel de clústeres aprovisionados

Panel de clústeres aprovisionados [Información](#)

[Pruebe Amazon Redshift Serverless](#) [Adquirir nodos reservados](#) [Crear clúster](#)

Información general sobre los recursos

Datos de recursos para la región de US East (N. Virginia).

Total de nodos	Nodos bajo demanda	Nodos reservados	Nodos reservados disponibles (0 de 0 utilizado)	Instantáneas automatizadas	Instantáneas manuales
1	1	0	0	2	0

Información general acerca del clúster (1)

[Cualquier estado](#)

Clúster	Estado
redshift1	Available

[Ver todos los clústeres](#)

Métricas del clúster

[Cualquier clúster](#) [Última hora](#) [Ver en CloudWatch](#)

Número de consultas	Conexiones de bases de datos	Espacio en disco utilizado	Utilización de la CPU
---------------------	------------------------------	----------------------------	-----------------------

Recursos compartidos de datos

Autorice a otras cuentas de AWS a obtener acceso a los recursos compartidos de datos creados en esta cuenta de AWS. Asocie o rechace recursos compartidos de datos de otras cuentas de AWS.

[Exigir autorización](#)

0

[Exigir asociación](#)

0

9. LABORATORIO REDSHIFT

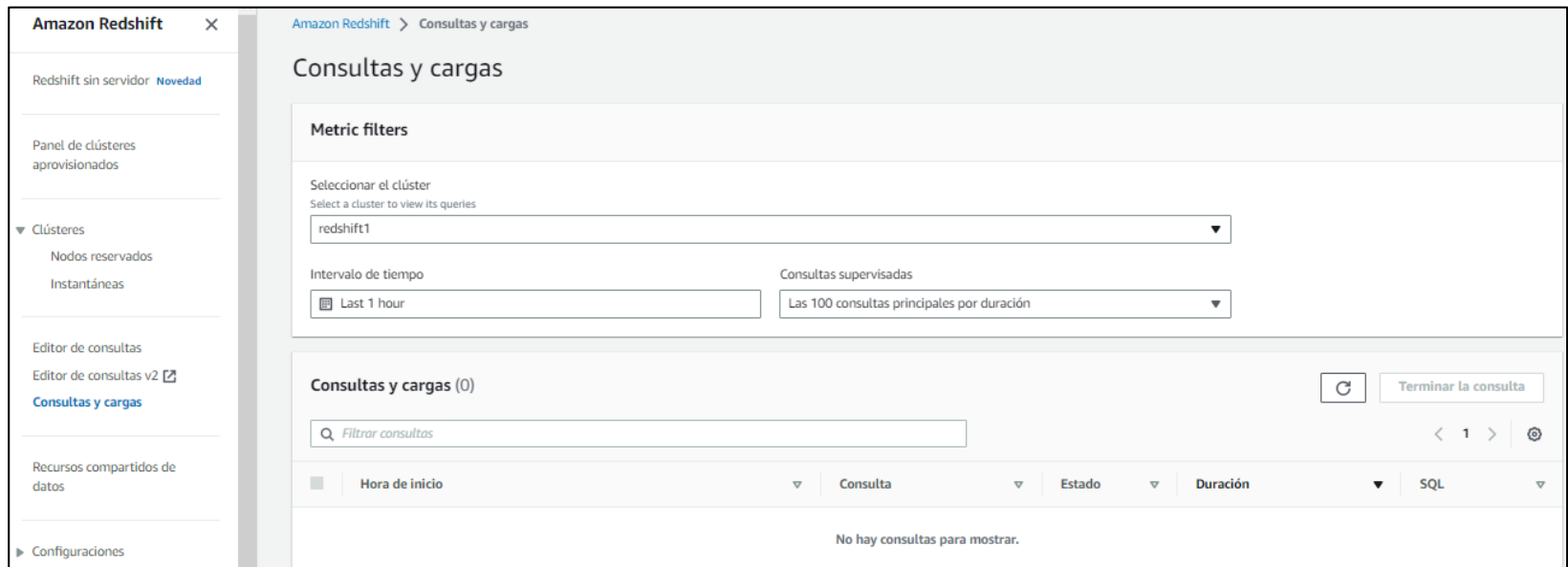
Paso 28. Si vamos a Clústeres tenemos la parte en la que esta la lista de clústeres y podemos ir a Querys and loads, donde podemos hacer búsquedas de consultas que se hayan realizado anteriormente, en este caso de ninguna.



Clústeres (1) Información

Filterar clústeres por propiedad o valor

<input type="checkbox"/>	Clúster	Estado	Espacio de nombre...	Zona de disponibilidad	Multi-AZ	Capacidad de alma...	Utilización de la CPU
<input type="checkbox"/>	redshift1 dc2.large 1 nodo 160 GB	Available	ae993c6f-55bd-4751-...	us-east-1e	No	< 1 %	2 %



Amazon Redshift > Consultas y cargas

Consultas y cargas

Metric filters

Seleccionar el clúster
Select a cluster to view its queries
redshift1

Intervalo de tiempo
Last 1 hour

Consultas supervisadas
Las 100 consultas principales por duración

Consultas y cargas (0)

Filterar consultas

<input type="checkbox"/>	Hora de inicio	Consulta	Estado	Duración	SQL
--------------------------	----------------	----------	--------	----------	-----

No hay consultas para mostrar.

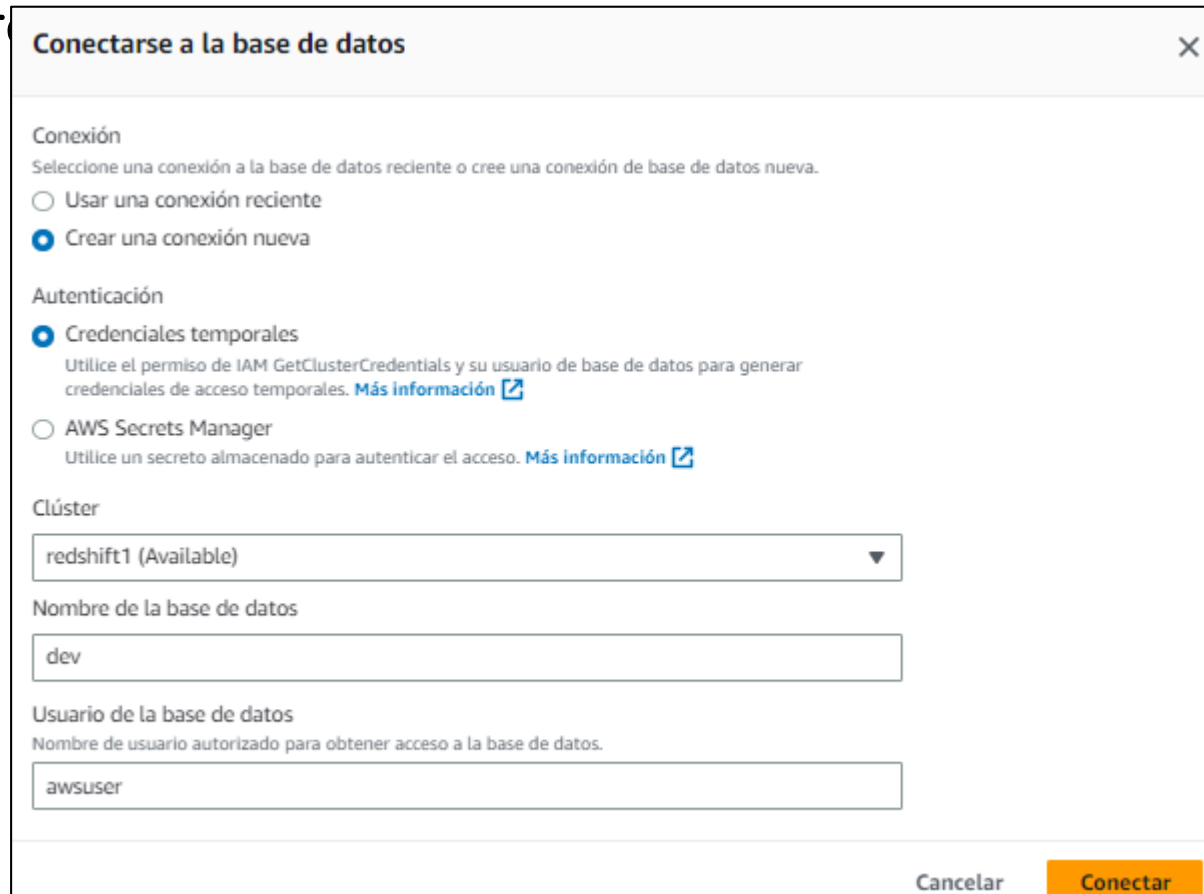
9. LABORATORIO REDSHIFT

Paso 29. Podemos ir a un editor para poder consultar la base de datos de pruebas que hemos cargado. Desde tenemos que crear una nueva conexión a la base de datos. Hacemos click al botón Conectarse a la base de datos

The screenshot shows the Amazon Redshift 'Editor de consultas' (Query Editor) interface. At the top, there's a breadcrumb 'Amazon Redshift > Editor de consultas'. Below it is a navigation bar with tabs: 'Editor' (selected), 'Historial de consultas', 'Consultas guardadas', and 'Consultas programadas'. A light blue banner contains an information icon and the text 'Conéctese a una base de datos para ejecutar consultas y ver los resultados.' Below this, on the left, is a 'Recursos' (Resources) sidebar with 'Información' (Information) section, containing dropdowns for 'Select database' and 'Select schema', and a 'Filter tables' search box. The main area has a toolbar with 'Estado', 'base de datos', and 'usuario' dropdowns, followed by a prominent orange button 'Conectarse a la base de datos'. Below the toolbar, a 'Query 1' tab is active, showing a large text area for writing SQL queries. At the bottom of the sidebar, there are navigation arrows and a page indicator '1'.

9. LABORATORIO REDSHIFT

Paso 30. Indicamos que vamos a crear una nueva conexión, con las credenciales temporales y el clúster se llama redshift1 (ponemos el clúster que toca). El nombre de la base de datos por defecto era dev y el usuario tampoco lo hemos modificado que era awuser. Y le damos al botón Conectar



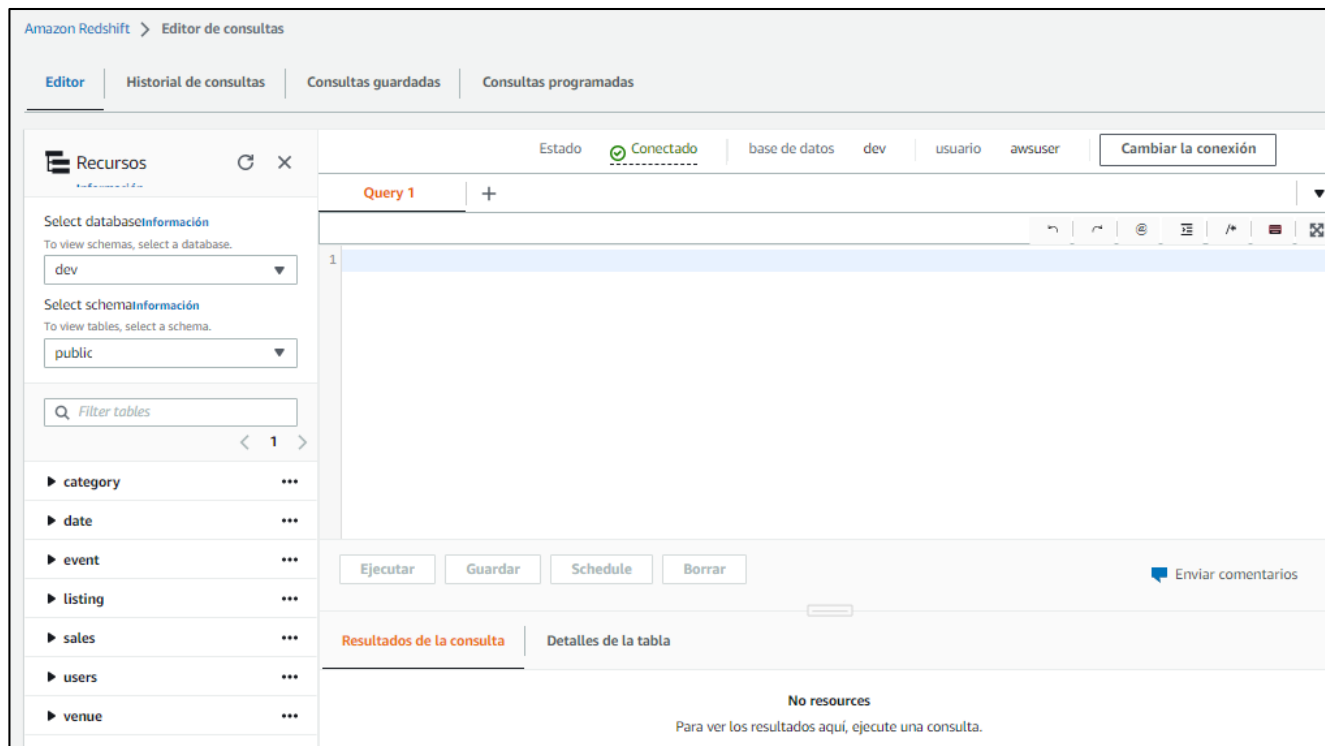
The screenshot shows a modal dialog titled "Conectarse a la base de datos" with a close button (X) in the top right corner. The dialog is divided into several sections:

- Conexión:** "Seleccione una conexión a la base de datos reciente o cree una conexión de base de datos nueva." with two radio buttons: "Usar una conexión reciente" (unselected) and "Crear una conexión nueva" (selected).
- Autenticación:** "Credenciales temporales" (selected) and "AWS Secrets Manager" (unselected). Below "Credenciales temporales" is a description: "Utilice el permiso de IAM GetClusterCredentials y su usuario de base de datos para generar credenciales de acceso temporales. [Más información](#)". Below "AWS Secrets Manager" is a description: "Utilice un secreto almacenado para autenticar el acceso. [Más información](#)".
- Clúster:** A dropdown menu showing "redshift1 (Available)".
- Nombre de la base de datos:** A text input field containing "dev".
- Usuario de la base de datos:** A text input field containing "awsuser".

At the bottom right, there are two buttons: "Cancelar" (grey) and "Conectar" (orange).

9. LABORATORIO REDSHIFT

Paso 31. Lo que hace es conectarnos, hacer una conexión con la base de datos para poder hacer consultas. Aquí ha seleccionado la base de datos dev (solo hay esta), el esquema public y aquí vemos todas las tablas que ha cargado esta base de datos de pruebas.



9. LABORATORIO REDSHIFT

Paso 32. si hacemos click en la tabla category, nos carga todos los campos que tiene. Incluso podemos ver información de la tabla, haciendo clic sobre los 3 puntos seleccionamos Vista previa de los datos. En la parte de abajo podemos ver los 10 primeros datos de la base de datos.

▼ category	...
catid	Insertar en SQL
catgroup	Mostrar el esquema
catname	Vista previa de los datos
catdesc	Agregar fila
► date	Duplicar la tabla
► event	Soltar la tabla
► listing	
► sales	...
► users	...

category				
public.category				
			Mostrar el esquema	Vista previa de los datos
Buscar en las filas				
catid	catgroup	catname	catdesc	
1	Sports	MLB	Major League Baseball	
3	Sports	NFL	National Football League	
6	Shows	Musicals	Musical theatre	
8	Shows	Opera	All opera and light opera	
11	Concerts	Classical	All symphony, concerto, and choir concerts	
2	Sports	NHL	National Hockey League	
4	Sports	NBA	National Basketball Association	
5	Sports	MLS	Major League Soccer	
7	Shows	Plays	All non-musical theatre	
9	Concerts	Pop	All rock and pop music concerts	

9. LABORATORIO REDSHIFT

Paso 33. Lo mismo con la tabla de ventas. Podemos hacer una consulta de los datos van. Y que precede a toda la información.

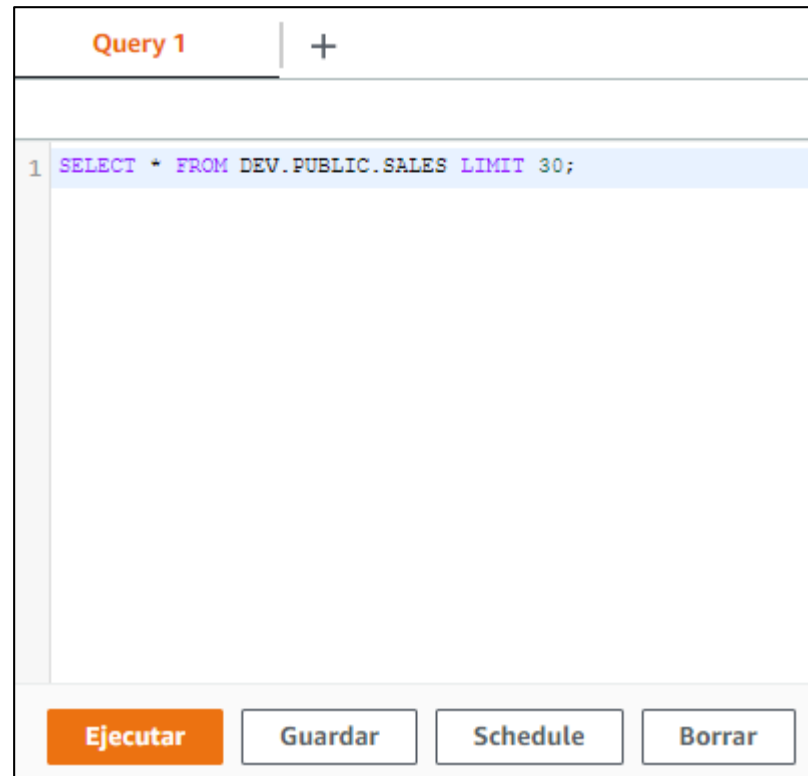
<div><div>▶ category</div><div>...</div></div>		<div>las filas</div> <div>< 1 ></div> <div></div>																
<div><div>▶ date</div><div>...</div></div>																		
<div><div>▶ event</div><div>...</div></div>		listid	▼	sellerid	▼	buyerid	▼	eventid	▼	dateid	▼	qtysold	▼	pricepaid	▼	commissi		
<div><div>▶ listing</div><div>...</div></div>		36572		30047		660		2903		1827		2		234.00		35.10		
<div><div>▼ sales</div><div>...</div><div>salesid</div><div>listid</div><div>sellerid</div><div>buyerid</div><div>eventid</div><div>dateid</div><div>qtysold</div><div>pricepaid</div><div>commission</div><div>saletime</div></div>		100813		45818		698		8649		1827		4		836.00		125.40		
		127048		37631		116		1749		1827		1		337.00		50.55		
		173969		48680		816		8762		1827		2		688.00		103.20		
		206999		3003		157		6605		1827		1		1730.00		259.50		
		1176		37614		301		5414		1828		1		218.00		32.70		
		2067		27144		256		6977		1828		2		494.00		74.10		
		8942		23600		1078		2557		1828		5		525.00		78.75		
		40331		45102		13005		1091		1756		1828		2		58.00		8.70
<div><div>▶ users</div><div>...</div></div>																		
<div><div>▶ venue</div><div>...</div></div>																		
		40741		45570		18692		226		4650		1828		3		90.00		13.50

9. LABORATORIO REDSHIFT

Paso 34. También podemos hacer consultas directamente aquí:

```
SELECT * FROM DEV.PUBLIC.SALES LIMIT 30;  
DATABASE.SCHEMA.TABLE
```

Una vez hecha la Select le damos Ejecutar



9. LABORATORIO REDSHIFT

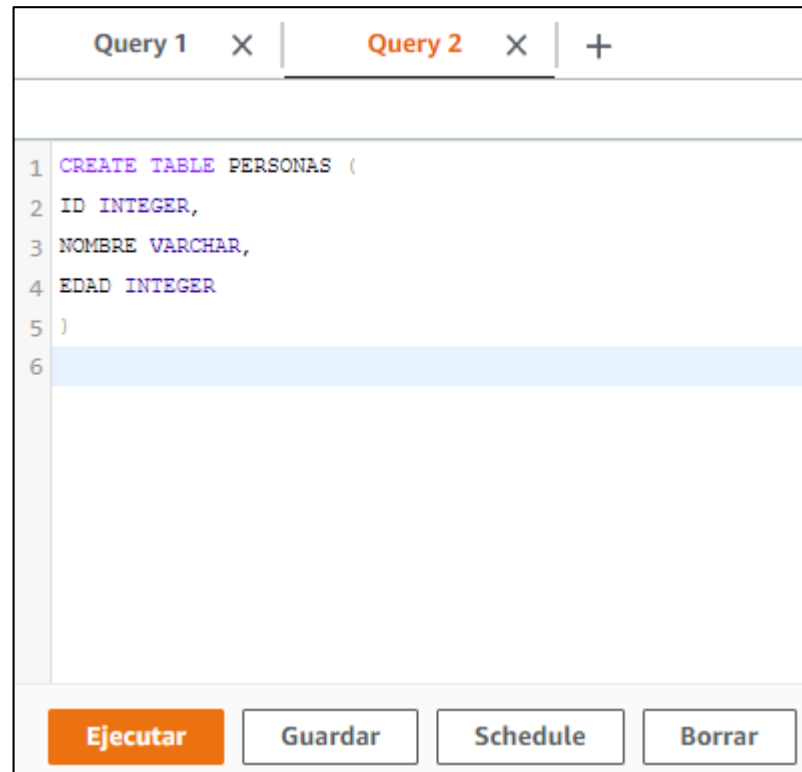
Paso 35. Debajo aparece el resultado de la query que hemos hecho, ha sido completada correctamente y el tiempo de ejecución. Devuelve 30 filas, y se pueden paginar porque hay 3 páginas con 10 resultados por página. Podemos ir pasando de una página a otra, viendo los resultados

Resultados de la consulta		Detalles de la tabla						
Consulta 12118 🔗		<div>Ejecución Datos Visualizar</div>						
✅ Completed, started on Mayo 08, 2023 at 19:56:23		TIEMPO TRANSCURRIDO: 00 min 02 s						
Rows returned (30)		Export ▼						
<input type="text" value="Buscar en las filas"/>		< 1 2 3 > ⚙️						
salesid ▼	listid ▼	sellerid ▼	buyerid ▼	eventid ▼	dateid ▼	qtysold ▼	pricepaid ▼	
33095	36572	30047	660	2903	1827	2	234.00	
88268	100813	45818	698	8649	1827	4	836.00	
110917	127048	37631	116	1749	1827	1	337.00	

9. LABORATORIO REDSHIFT

Paso 36. Vamos a crear una tabla nueva, Abrimos una nueva pestaña query y desde ponemos el comando para crear una tabla.

```
CREATE TABLE PERSONAS ( ID INTEGER, NOMBRE VARCHAR,  
EDAD INTEGER )
```




9. LABORATORIO REDSHIFT

Paso 37. Entonces simplemente este SQL la podemos seleccionar y ejecutar o Si sólo esta este sql en este query , pues podemos lanzarla directamente. Le damos a Run y aquí aparecería el resultado está ejecutando en verde decir se ha completado satisfactoriamente y el tiempo que le ha llevado. Entonces si ahira automáticamente vemos que existe ya la tabla personas. La acabamos de crear con sus tres campos y de nombre y edad y de nombre.

Resultados de la consulta | Detalles de la tabla

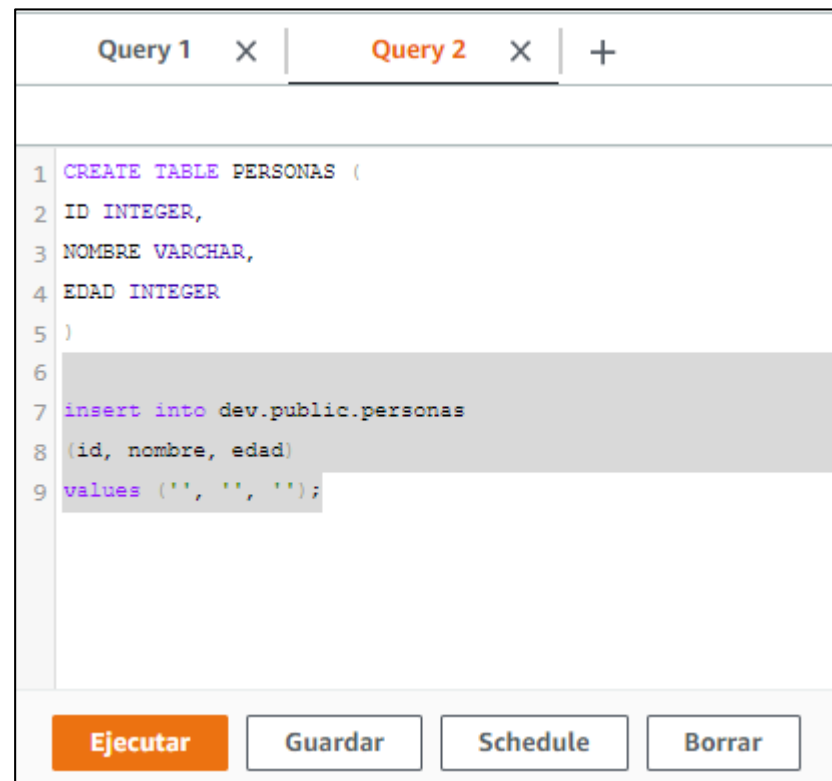
Consulta

 Completed, started on Mayo 08, 2023 at 21:11:33
TIEMPO TRANSCURRIDO: 00 min 03 s

▼ personas	...
id	
nombre	
edad	

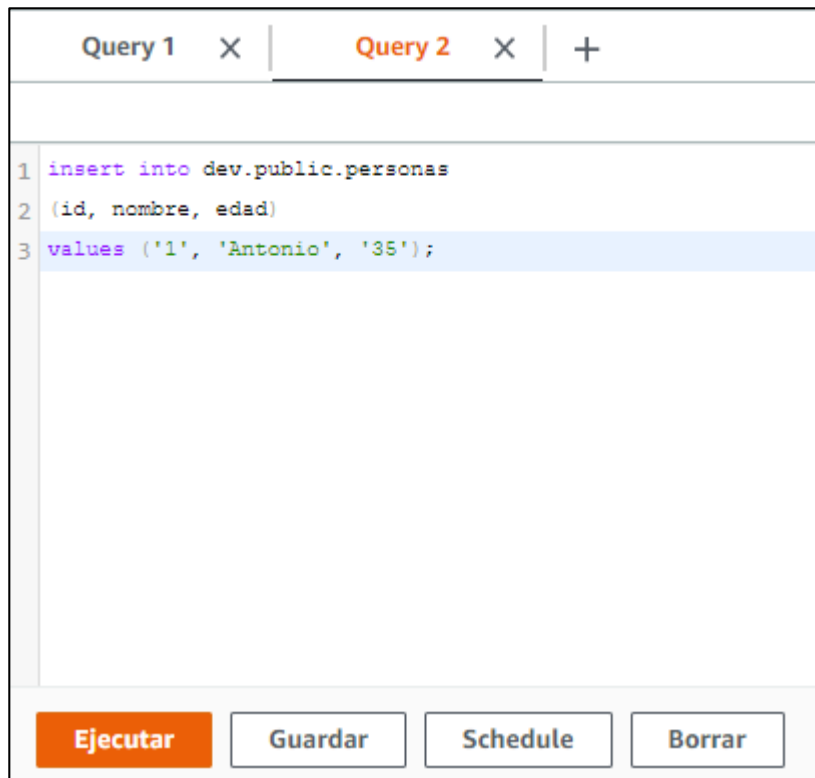
9. LABORATORIO REDSHIFT

Paso 38. Si queremos podemos insertar datos en la tabla. Desde los 3 puntos podemos darle a añadir fila. Esto nos crea la SQL que necesitamos para insertar. Inserta dentro de la base de datos dev, el esquema public y la tabla personas. Entre paréntesis, los tres campos que figuran dentro de la tabla, id, nombre y edad y los valores en este caso están sin rellenar.



9. LABORATORIO REDSHIFT

Paso 39. Ponemos por ejemplo: id=1, nombre=Antonio y edad=35. Le vamos a ejecutar. Vemos que se ha completado satisfactoriamente la inserción en la tabla.



The screenshot shows the Amazon Redshift console interface. At the top, there are tabs for 'Query 1' and 'Query 2'. 'Query 2' is selected and highlighted in orange. Below the tabs, the SQL query is displayed in a text area:

```
1 insert into dev.public.personas
2 (id, nombre, edad)
3 values ('1', 'Antonio', '35');
```

At the bottom of the console, there are four buttons: 'Ejecutar' (highlighted in orange), 'Guardar', 'Schedule', and 'Borrar'.

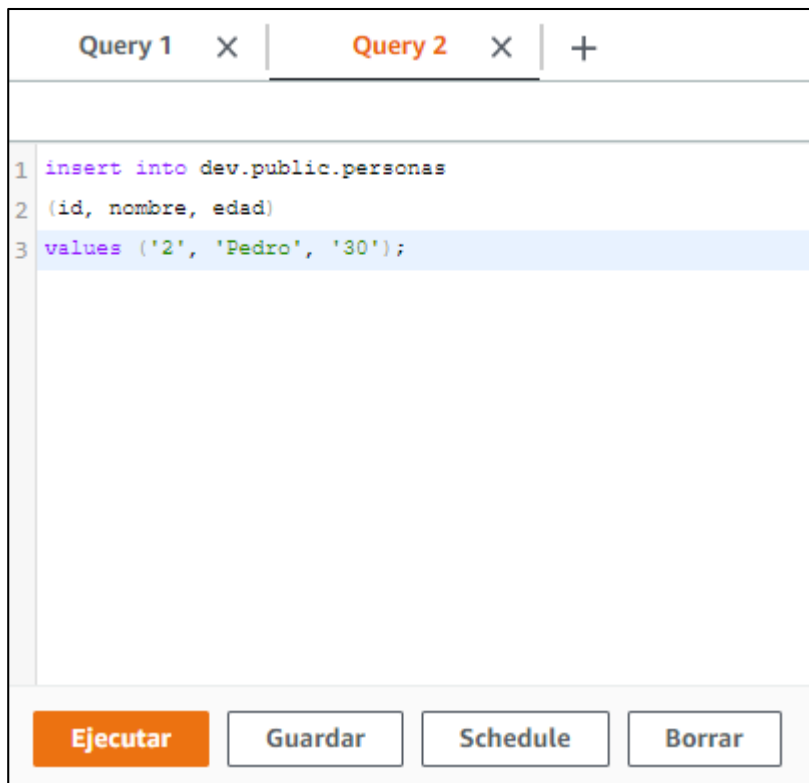


The screenshot shows the 'Resultados de la consulta' (Query Results) tab in the Amazon Redshift console. The query ID is 'Consulta 13287'. The status is 'Completed, started on Mayo 08, 2023 at 21:40:53'. The execution time is 'TIEMPO TRANSCURRIDO: 00 min 03 s'.

Resultados de la consulta	Detalles de la tabla
Consulta 13287 🔗	
✅ Completed, started on Mayo 08, 2023 at 21:40:53	
TIEMPO TRANSCURRIDO: 00 min 03 s	

9. LABORATORIO REDSHIFT

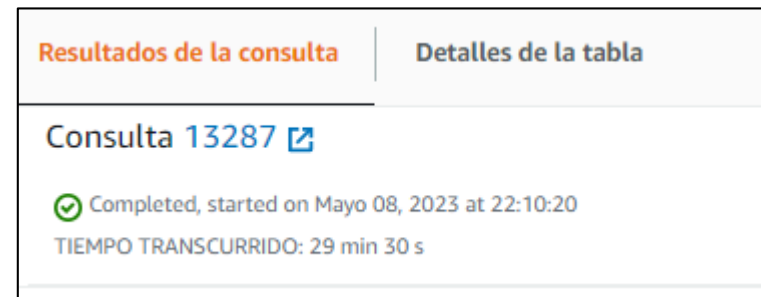
Paso 40. Ponemos añadir más elementos para después visualizarlos: id=2 , nombre=Pedro, edad=30 y le damos a ejecutar. Hemos insertado el segundo elemento de personas



The screenshot shows the Amazon Redshift console interface. At the top, there are tabs for 'Query 1' and 'Query 2', with 'Query 2' being the active tab. Below the tabs, the SQL query is displayed in a text area:

```
1 insert into dev.public.personas
2 (id, nombre, edad)
3 values ('2', 'Pedro', '30');
```

At the bottom of the console, there are four buttons: 'Ejecutar' (Execute), 'Guardar' (Save), 'Schedule', and 'Borrar' (Delete). The 'Ejecutar' button is highlighted in orange.



The screenshot shows the 'Resultados de la consulta' (Query Results) tab in the Amazon Redshift console. It displays the following information:

- Consulta 13287** with a link icon.
- A green checkmark icon followed by the text: 'Completed, started on Mayo 08, 2023 at 22:10:20'.
- The text: 'TIEMPO TRANSCURRIDO: 29 min 30 s'.

9. LABORATORIO REDSHIFT

Paso 41. Por último, la tercera persona: id=3, nombre=María, y edad=28. Vale, así que le vamos a ejecutar y tendríamos los tres elementos insertados y funciona correctamente.



The screenshot shows the Amazon Redshift Query Editor interface. At the top, there are tabs for 'Query 1' and 'Query 2', with a plus sign to add more. The 'Query 2' tab is active. Below the tabs, the SQL editor contains the following code:

```
1 insert into dev.public.personas
2 (id, nombre, edad)
3 values ('3', 'Maria', '28');
```

At the bottom of the editor, there are four buttons: 'Ejecutar' (Execute), 'Guardar' (Save), 'Schedule', and 'Borrar' (Delete).



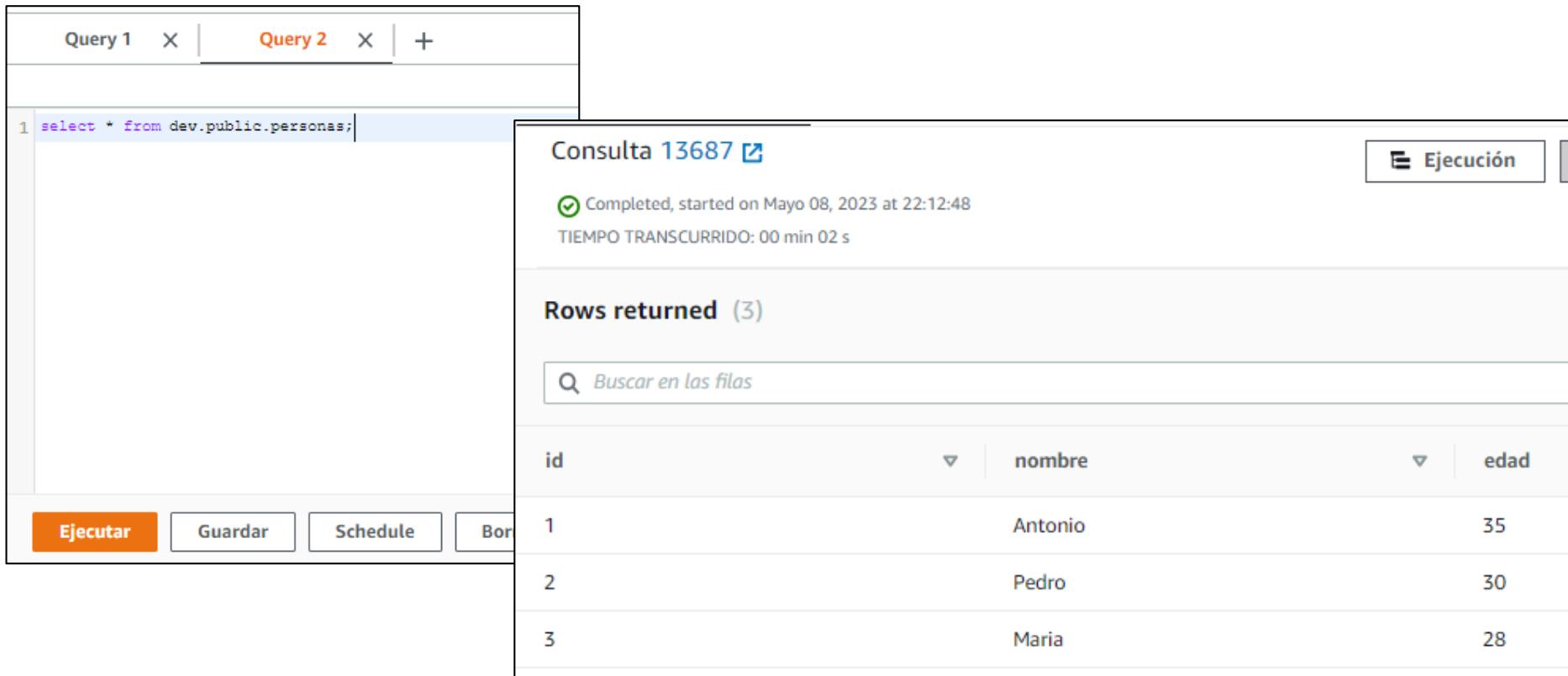
The screenshot shows the Amazon Redshift Query Results page. It has two tabs: 'Resultados de la consulta' (Query Results) and 'Detalles de la tabla' (Table Details). The 'Resultados de la consulta' tab is active. Below the tabs, the query ID is 'Consulta 13640' with a link icon. Below that, there is a green checkmark icon and the text 'Completed, started on Mayo 08, 2023 at 22:12:07'. At the bottom, it says 'TIEMPO TRANSCURRIDO: 00 min 03 s'.

9. LABORATORIO REDSHIFT

Paso 42. Para ver la información, ejecutamos una simple select:

```
select * from dev.public.personas
```

Como sólo hay solo tres filas, no hace falta poner un limit. La ejecutamos y nos salen los tres registros que hemos insertado



Query 1 X Query 2 X +

```
1 select * from dev.public.personas;
```

Ejecutar Guardar Schedule Bor

Consulta 13687 [🔗](#) Ejecución

✅ Completed, started on Mayo 08, 2023 at 22:12:48
TIEMPO TRANSCURRIDO: 00 min 02 s









Rows returned (3)

🔍 *Buscar en las filas*

id	nombre	edad
1	Antonio	35
2	Pedro	30
3	Maria	28

9. LABORATORIO REDSHIFT

Paso 43. Vamos a Consultas y cargas. Podemos ver todas las queries y ejecuciones que hemos hecho: cuando se han lanzado, el tipo de query, su estado, el tiempo de duración de cada una de las queries, etc

Consultas y cargas (8)							Terminar la consulta
<input type="text" value="Filtrar consultas"/>						< 1 > 	
<input type="checkbox"/>	Hora de inicio ▼	Consulta ▼	Estado ▼	Duración ▼	SQL ▼		
<input type="checkbox"/>	May 8th, 2023 09:33:13 PM 2 days ago	13155	✔ Completado	216 ms	 insert into dev.public.personas (id, nombre, edad) values ('1', 'Antonio', '35');		
<input type="checkbox"/>	May 8th, 2023 05:52:22 PM 2 days ago	10822	✔ Completado	159 ms	 select * from "dev"."public"."category" limit 10;		
<input type="checkbox"/>	May 8th, 2023 10:11:41 PM a day ago	13620	✔ Completado	113 ms	 select * from dev.public.personas;		
<input type="checkbox"/>	May 8th, 2023 07:51:43 PM 2 days ago	12067	✔ Completado	108 ms	 select * from "dev"."public"."sales" limit 10;		
<input type="checkbox"/>	May 8th, 2023 09:40:51 PM 2 days ago	13287	✔ Completado	15 ms	 insert into dev.public.personas (id, nombre, edad) values ('2', 'Pedro', '30');		
<input type="checkbox"/>	May 8th, 2023 10:12:05 PM a day ago	13640	✔ Completado	15 ms	 insert into dev.public.personas (id, nombre, edad) values ('3', 'Maria', '28');		

9. LABORATORIO REDSHIFT

Paso 44. También podemos ir a Recursos compartidos de datos


aprovechados

▼ Clústeres

Nodos reservados

Instantáneas

Editor de consultas

Editor de consultas v2 

Consultas y cargas

Recursos compartidos de datos


▼ Configuraciones

Administración de cargas de trabajo

Recursos compartidos de datos


▼ Funcionamiento

Control datashare access from other accounts




For datashares created in your account, choose AWS accounts to authorize access to or to remove authorization from the datashares. Authorized AWS accounts are notified to accept the datashares.

Associate datashares with your AWS account or namespace




For datashares from other accounts, associate the datashares with your AWS account or specific namespaces in your account. Consumers can work with datashares from the datashare page after you associate the datashares.

Create a product on AWS Data Exchange



To license your Amazon Redshift data on AWS Data Exchange, publish a data product and make it available in the AWS Marketplace catalog. Subscribers can access the datashare and query specific data.

Share via AWS Data Catalog



To centrally enforce permissions on the datashare and its objects, authorize a Lake Formation account to access the datashare. Lake Formation administrators can define security policies to restrict user access to certain objects in the datashare.


9. LABORATORIO REDSHIFT

Paso 45. En Configuraciones, si queremos crear un punto de enlace (endpoint) de VPC para RedShit.


Configuraciones

Configure conexiones entre Amazon Redshift y otras herramientas.

Puntos de enlace de VPC administrados por Redshift (0)
Puntos de enlace para obtener acceso a clústeres de que están en otra VPC o subred.

 [Editar](#) [Eliminar](#) [Crear punto de enlace](#)

< 1 >



Nombre del punto de enlace ▲	ID de la cuenta ▼	Identificador del clúster ▼	Estado ▼
No hay puntos de enlace disponibles para mostrar			

9. LABORATORIO REDSHIFT

Paso 46. También podemos conectarnos desde herramientas clientes a nuestra base de datos de RedShift. Podemos elegir el tipo de conexión que vamos a hacer, si va a ser una conexión JDBC o ODBC contra nuestra base RedShift, si es JDBC, podemos elegir el driver que vamos a utilizar y descargarlo para poder utilizarlo en nuestro ordenador para que nuestra cliente se conecte a RedShift.

Conectarse a las herramientas del cliente [Más información](#)

Amazon Redshift está diseñado para funcionar con la herramienta de inteligencia empresarial que prefiera. Si no dispone de las [herramientas de nuestros socios](#) o [SQL Workbench/J](#) de forma gratuita.

1

Elegir el tipo de conexión

Elija el tipo de conexión que utiliza la herramienta del cliente para la conexión del clúster.

☒ JDBC ☐ ODBC

Conectividad de bases de datos Java Conectividad de bases de datos abiertas

2

Descargar los controladores

Para conectarse a la herramienta del cliente, puede utilizar un controlador JDBC u ODBC de Amazon Redshift personalizado.

JDBC 4.2 sin AWS SDK (.jar)

▼

Descargar el controlador

Versión v1.2.45.1069 de los controladores JDBC

3


URL de conexión del clúster

Elija el clúster al que desea conectarse.


Elegir un clúster

▼

Copiar la URL de JDBC



redshift-jdbc42-2.1.0.12.jar
Descarga terminada



87

9. LABORATORIO REDSHIFT


Paso 47. En la URL de conexión al cluster, elegimos nuestro clúster redshift1 y obtenemos la ruta que tendríamos para conectarnos a él desde otra herramienta cliente. Abajo hay mas información:

3

URL de conexión del clúster

Elija el clúster al que desea conectarse.

redshift1 ▼

 Copiar la URL de JDBC

jdbc:redshift://redshift1.ccx4gbcdfjz.us-east-1.redshift.amazonaws.com:5439/dev

Puede [agregar opciones de configuración](#) a la URL de conexión, como el nombre de usuario y la contraseña.

Administración de cargas de trabajo

Administre la carga de trabajo de consulta mediante un conjunto ordenado de colas de consultas que definan de qué manera se asignan los recursos y cómo se guían las consultas para su procesamiento. Aplique parámetros y ajustes de carga de trabajo similares a varios clústeres con los grupos de parámetros.

[Administrar los grupos de parámetros >](#)

Conexiones de HSM

Cifre la base de datos mediante un módulo de seguridad de hardware (HSM). Configure los certificados de cliente y servidor para establecer un enlace de red de confianza entre Amazon Redshift y su HSM.

Grupos de subredes

Aprovisione el clúster en una de las subredes que proporcione al grupo de subredes. Un grupo de subredes de clúster le permite especificar un conjunto de subredes en su VPC.

172.16.0.0

172.16.1.0

172.16.2.0

Administrar las etiquetas

Organice todos los recursos de Amazon Redshift etiquetándolos. Aplique claves y valores de etiqueta a varios recursos a la vez.

[Administrar las etiquetas >](#)

9. LABORATORIO REDSHIFT

Paso 48. En AWS Marketplace, podemos tener otros productos como Tableau Server que podemos, que podemos ver y utilizar


Productos de socios de Amazon Redshift

Mejore el almacenamiento de datos de Amazon Redshift con las herramientas y los expertos líderes del sector para cargar, transformar y visualizar datos a través de AWS Marketplace.

AWS Marketplace

Buscar productos de Amazon Redshift

Inteligencia empresarial [Ver todos los productos](#)



Alteryx Server

Por Alteryx | Ver 2022.3.1.486

Alteryx on AWS is a cost-effective and flexible way to manage and deploy various configurations of Alteryx Server.

[Ver producto](#)




Tableau Server (BYOL)

Por Tableau

Quickly deploy or evaluate a single-node instance of Tableau Server on AWS to enable data-driven decision making throughout your organization.

[Ver producto](#)

9. LABORATORIO REDSHIFT

Paso 49. En Advisor, de recomendaciones no tenemos ninguna,

Amazon Redshift > Recomendaciones

💡 Recomendaciones de Advisor (0)

📄 Exportar

Agrupar por clústeres

Agrupar por recomendación

Ordenar por cluster ▼

🔍 Buscar

< 1 >



No hay recomendaciones para mostrar

De alarmas tampoco tenemos ninguna. Podemos crear alarmas,

Alarmas (0)



Acciones ▼

Crear una alarma

🔍 Buscar

Alarma ▼

Cualquier clúster ▼

< 1 >



Nombre de la alarma ▼

Descripción ▼

Clúster ▼



Estado



No hay alarmas.

Comience a monitorizar el clúster con alarmas de CloudWatch

Crear una alarma

9. LABORATORIO REDSHIFT

Paso 50. Los eventos que lo que ha ido pasando con todo lo que hemos ido haciendo

Eventos

Última semana ▼

2

Total de eventos

0

Próximos eventos

0

Errores

EVENTOS POR CATEGORÍA

1

Eventos de configuración

0

Eventos de monitoreo

1

Eventos de administración

0

Eventos de seguridad

EVENTOS POR ORIGEN

1

Eventos de clúster

0

Eventos de instantáneas

0

Acción programada

1

Eventos de grupo de parámetros

0

Eventos de grupo de seguridad

Q

Buscar

Todas las cate...

Todos los orig...

< 1 >

⚙


Fecha	Hora	Evento	Origen
2 days ago 8 may 2023	0:17	Amazon Redshift cluster 'redshift1' has been created at 2023-05-07 22:17 UTC and is ready for use.	redshift1 Clúster
2 days ago 8 may 2023	0:14	Cluster parameter group 'default.redshift-1.0' was created at 2023-05-07 22:14 UTC.	default.redshift-1.0 Grupo de parámetros

9. LABORATORIO REDSHIFT

Paso 51. En Novedades tenemos información general

Novedades

En esta regiónEn todas las regiones



Novedad


Amazon Redshift announces preview of cross-account data sharing

2 years ago
Apr 30, 2021

Amazon Redshift data sharing allows you to share live, transactionally consistent data across different Redshift clusters without the complexity and delays associated with data copies and data movement. Ability to share data across clusters that are in the same AWS account is already generally available. Now you can preview cross-account data sharing to share data across Redshift clusters in different AWS accounts, both within and outside your organization. Cross-account data sharing is available on all Redshift RA3 node types. There is no additional cost to use cross-account sharing on your Amazon Redshift clusters.

[Más información](#)

Disponible en **Northern Virginia (us-east-1)** y **15 otras regiones**



Novedad

Amazon Redshift announces support for hierarchical data queries with Recursive CTE

2 years ago
Apr 29, 2021

Amazon Redshift, a fully-managed cloud data warehouse, now adds support for Recursive Common Table Expression (CTE). Starting today, you can use a Recursive CTE on your Redshift cluster to query hierarchical data such as organizational charts where employees reports to other employees (managers), or multi-level product orders where a product consists of many components, which in turn consist of other components.

[Más información](#)

Disponible en: *todas las regiones de AWS*

9. LABORATORIO REDSHIFT

Paso 52. A nivel de clúster una vez que ya lo hemos utilizado, podemos ir al Rendimiento del clúster. Aquí podemos ver las gráficas de uso de CPU que hemos ido teniendo en función de las consultas realizadas

También el porcentaje de disco usado. En este caso hemos creado la base de datos de pruebas con datos de 28 kbytes, con lo cual tenemos un porcentaje utilizado de solamente 0,2% de 160 GB que pusimos en el nodo

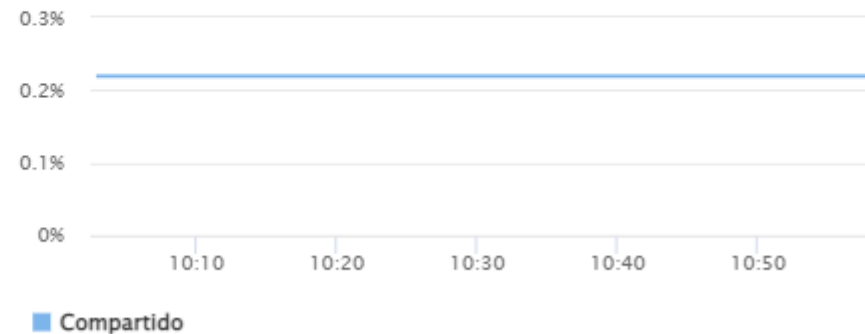
Utilización de la CPU

El porcentaje de utilización de la CPU.



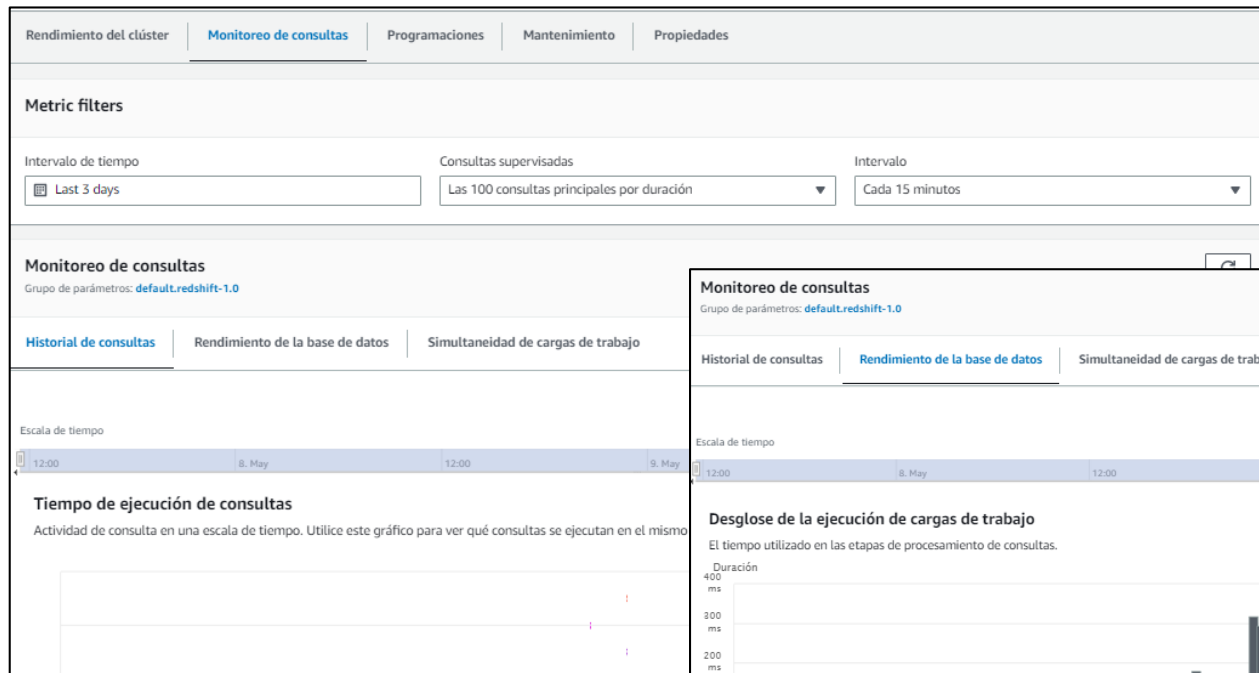
Porcentaje del espacio en disco utilizado

El porcentaje de espacio en disco utilizado.



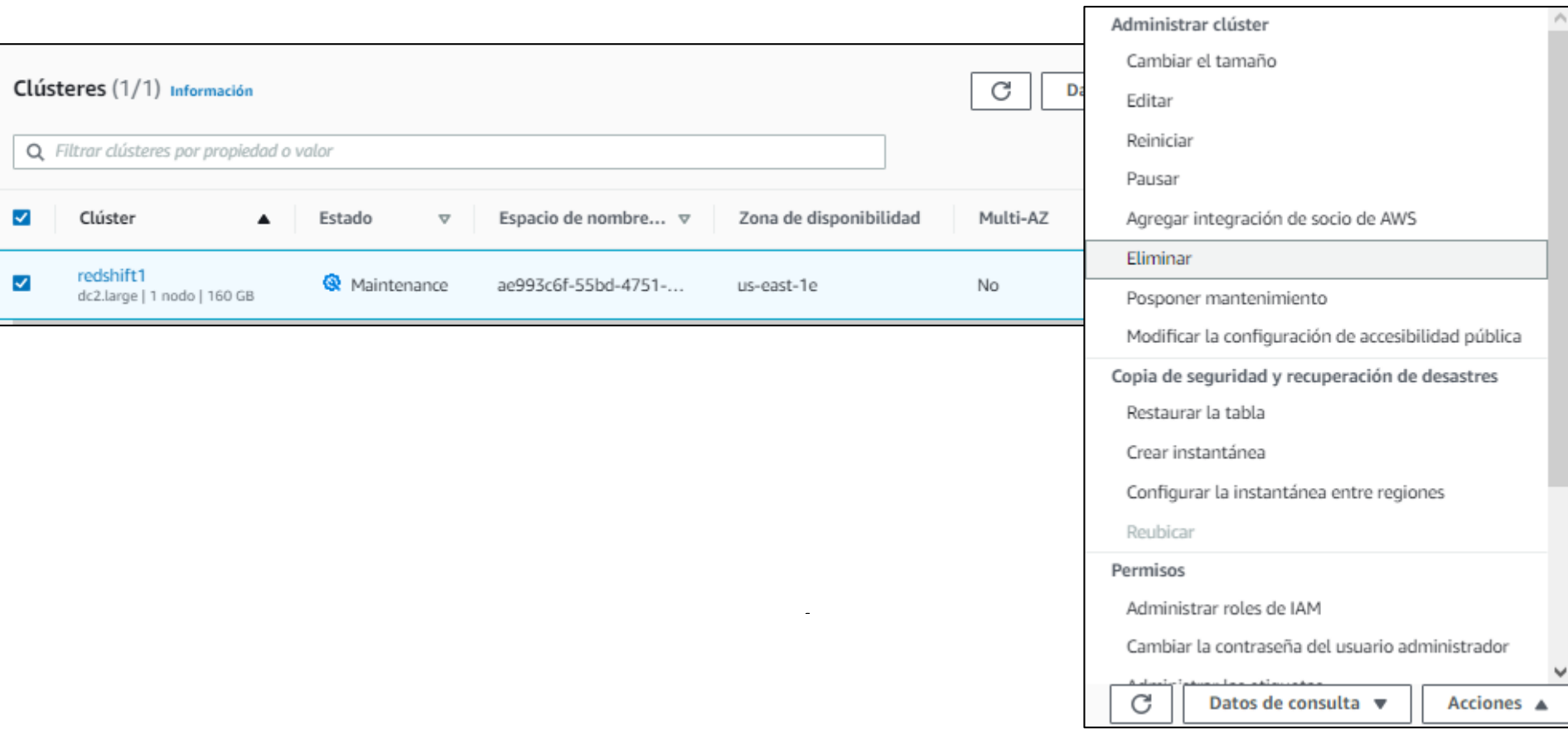
9. LABORATORIO REDSHIFT

Paso 53. En Monitoreo de consultas, Historial de Consultas podemos ver si hay alguna query en ejecución, podemos ver en que estado está. En Rendimiento de la base de datos, trabajo, concurrentes,



9. LABORATORIO REDSHIFT

Paso 54. Vamos a eliminar el cluster RedShift que hemos creado anteriormente para que no nos facturen demás. Vamos a Clústeres y marcamos el cluster redshift que tenemos. Desde aquí podemos ver las acciones que podemos realizar sobre el cluster



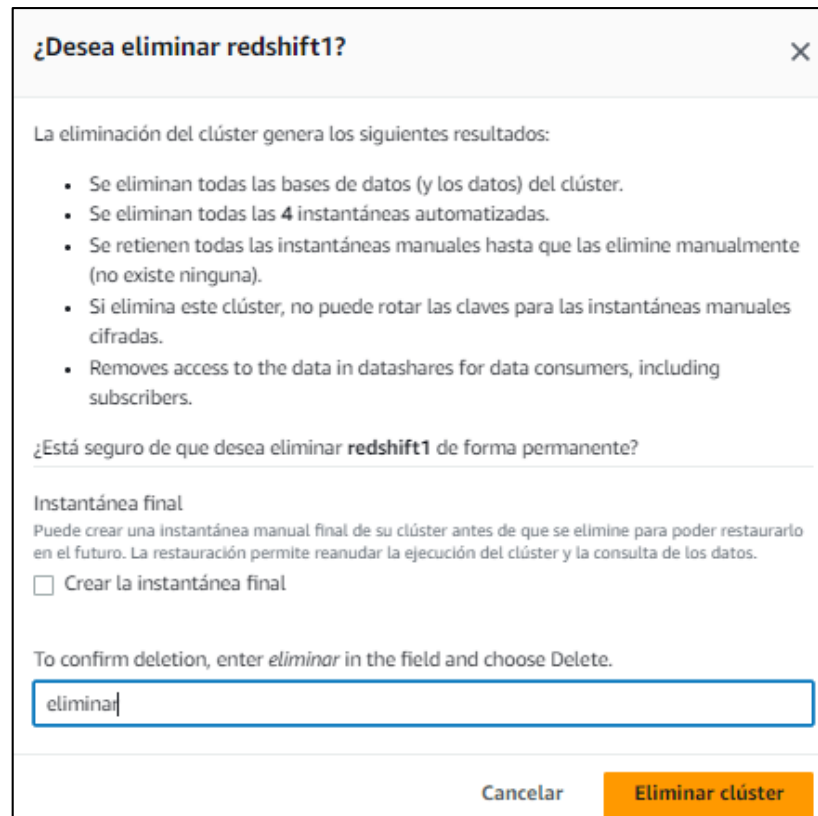
The screenshot displays the AWS Redshift console interface. At the top, there's a header with 'Clústeres (1/1)' and an 'Información' link. Below this is a search bar with the placeholder text 'Filtrar clústeres por propiedad o valor'. The main content area is a table listing Redshift clusters. The table has columns for 'Clúster', 'Estado', 'Espacio de nombre...', 'Zona de disponibilidad', and 'Multi-AZ'. A single cluster, 'redshift1', is listed with a status of 'Maintenance', a name 'dc2.large | 1 nodo | 160 GB', and a zone of 'us-east-1e'. A context menu is open over the 'redshift1' cluster, showing a list of actions: 'Administrar clúster', 'Cambiar el tamaño', 'Editar', 'Reiniciar', 'Pausar', 'Agregar integración de socio de AWS', 'Eliminar' (highlighted), 'Posponer mantenimiento', 'Modificar la configuración de accesibilidad pública', 'Copia de seguridad y recuperación de desastres', 'Restaurar la tabla', 'Crear instantánea', 'Configurar la instantánea entre regiones', 'Reubicar', 'Permisos', 'Administrar roles de IAM', and 'Cambiar la contraseña del usuario administrador'. At the bottom of the menu, there are buttons for 'Datos de consulta' and 'Acciones'.

Clúster	Estado	Espacio de nombre...	Zona de disponibilidad	Multi-AZ
redshift1 dc2.large 1 nodo 160 GB	Maintenance	ae993c6f-55bd-4751-...	us-east-1e	No

- Administrar clúster
 - Cambiar el tamaño
 - Editar
 - Reiniciar
 - Pausar
 - Agregar integración de socio de AWS
- Eliminar**
- Posponer mantenimiento
- Modificar la configuración de accesibilidad pública
- Copia de seguridad y recuperación de desastres
 - Restaurar la tabla
 - Crear instantánea
 - Configurar la instantánea entre regiones
 - Reubicar
- Permisos
 - Administrar roles de IAM
 - Cambiar la contraseña del usuario administrador

9. LABORATORIO REDSHIFT

Paso 55. Seleccionamos la opción Eliminar para borrar nuestro clúster. Nos pregunta si queremos hacer un snapshot una copia de seguridad de los datos que hay actualmente en la base de datos de redshift. En este caso, como no lo vamos a utilizar más, le vamos a decir que no cree ningún snapshot final con los datos y le damos a Eliminar clúster.



¿Desea eliminar redshift1?

La eliminación del clúster genera los siguientes resultados:

- Se eliminan todas las bases de datos (y los datos) del clúster.
- Se eliminan todas las 4 instantáneas automatizadas.
- Se retienen todas las instantáneas manuales hasta que las elimine manualmente (no existe ninguna).
- Si elimina este clúster, no puede rotar las claves para las instantáneas manuales cifradas.
- Removes access to the data in datashares for data consumers, including subscribers.

¿Está seguro de que desea eliminar **redshift1** de forma permanente?

Instantánea final
Puede crear una instantánea manual final de su clúster antes de que se elimine para poder restaurarlo en el futuro. La restauración permite reanudar la ejecución del clúster y la consulta de los datos.

☐ Crear la instantánea final

To confirm deletion, enter *eliminar* in the field and choose Delete.

Cancelar Eliminar clúster

9. LABORATORIO REDSHIFT

Paso 56. Una vez que le damos a Eliminar clúster, se modifica el estado pasando Eliminando. Esto puede durar unos 5 minutos. Después se habrá borrado completamente y ya nos cobrarán más. Este paso es importante porque si dejamos arrancado el clúster tendrá un coste alto aproximadamente 180\$/mes, por lo cual es muy importante verificar que está correctamente eliminado

Clústeres (1/1)

Información

↺

Datos de consulta

↺

Acciones

Crear clúster

🔍

Filtrar clústeres por propiedad o valor

◀

1

▶

⚙️

✓	Clúster ▲	Estado ▼	Espacio de nombre... ▼	Zona de disponibilidad	Multi-AZ	Capacidad de alma... ▼	Utilización de la CPU ▼
✓	redshift1 dc2.large 1 nodo 160 GB	⏸ Modifying Eliminando	ae993c6f-55bd-4751-...	us-east-1e	No	<div></div> < 1 %	<div></div> 3 %

Clústeres (0) [Información](#)

Datos de consulta ▼

Acciones ▼

Crear clúster

Filtrar clústeres por propiedad o valor

<

1

>

Clúster ▲

Estado ▼

Espacio de nombre... ▼

Zona de disponibilidad

Multi-AZ

Capacidad de alma... ▼

Utilización de la CPU ▼

Instant... ▼

Notificaci...

No hay clústeres

Crear un clúster de Amazon Redshift

Crear clúster