

Actividad:	RedShift							
Ciclo:	IABD	Modulo:	SBD	Curso:		Agrupación:	1	
Alumno/a:							Grupo:	
CIPFP www.fpmislata.com								

Tarea 1. Iniciar un clúster de Amazon Redshift

En esta tarea, iniciarás un clúster de Amazon Redshift que es un Data Warehouse completamente gestionado formado por un conjunto de nodos de computación.

Cada clúster pone en marcha un motor de Amazon Redshift y contiene una o más bases de datos.

Al iniciar un clúster, una de las opciones que tendrás que especificar:

- El node type (tipo de nodo). El tipo de nodo determina la CPU, la RAM, la capacidad de almacenamiento y el tipo de unidad de almacenamiento de cada nodo.
- Los tipos de nodo están disponibles en varios tamaños.
- El tamaño del nodo y el número de nodos determina el almacenamiento total de un clúster.

Pasos:

1. En la consola de administración de AWS, en el menú Servicios, haz clic en Amazon Redshift.
2. En el panel de navegación de la izquierda, haz clic en Clústeres.
3. Haz clic en Crear clúster para abrir el asistente de creación de clústeres de Redshift.
4. En la sección Configuración del clúster, configura lo siguiente:
 - ☐ Identificador de clúster: lab
 - Tipo de nodo: dc2.large
 - Number of nodes (Cantidad de nodos): 2
5. En la sección Configuraciones de la base de datos, configura lo siguiente:
 - Nombre de usuario del administrador: master
 - Contraseña de usuario administrador: Redshift123
 - En Roles de IAM asociados, haz clic en Rol de IAM asociado y selecciona LabRole (de AWS Academy).
6. Haz clic en Roles de IAM asociados.
7. El rol concede permiso a Amazon Redshift para leer datos de Amazon S3.
8. Clic en Crear clúster.
9. El clúster tardará unos minutos en iniciarse.
10. Haz clic en el nombre de tu clúster (lab).
 - ☐ La configuración del clúster aparecerá en pantalla.

Tarea 2. Utilizar el editor de consultas de Redshift para contactar con tu clúster de Redshift

Actividad:	RedShift							
Ciclo:	IABD	Modulo:	SBD	Curso:		Agrupación:	1	
Alumno/a:							Grupo:	
CIPFP www.fpmislata.com								

1. En el panel de navegación de la izquierda, haz clic en Query editor (Editor de consultas) y selecciona Conectar con la base de datos. A continuación, configura lo siguiente:
 - ☐ Clúster: lab
 - ☐ Nombre de la base de datos: dev
 - ☐ Usuario de la base de datos: master
 - ☐ Contraseña: Redshift123
2. Haz clic en Conectar

Tarea 3: Crear una base de datos y una tabla

1. Pulsa en el cluster en la parte izquierda y en botón CREATE elegie Database y nombre labdb.
2. Copia este comando SQL, pégalo en la ventana Query 1 (Consulta 1) elige en los desplegables la base de datos labdb y el usuario master. Ejecuta la consulta.

```
CREATE TABLE users (
  userid INTEGER NOT NULL,
  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,
```

Actividad:	RedShift							
Ciclo:	IABD	Modulo:	SBD	Curso:		Agrupación:	1	
Alumno/a:							Grupo:	
CIPFP www.fpmislata.com								

```

likerock BOOLEAN,

likevegas BOOLEAN,

likebroadway BOOLEAN,

likemusicals BOOLEAN

);

```

3. Este comando creará una tabla denominada usuarios. Contiene el nombre, la dirección y detalles sobre el tipo de música que le gusta al usuario.

Tarea 4: Cargar datos de muestra desde Amazon S3

Amazon Redshift puede importar datos desde Amazon S3. Admite varios formatos de archivo, campos de longitud fija, valores separados por comas (CSV) y delimitadores personalizados. Los datos de este laboratorio están separados por plecas (|).

1. Elimina la consulta existente y pega este comando SQL en la ventana Query 1 (Consulta 1).

```

COPY users FROM 's3://midemobucketxxxxxx/allusers_pipe.txt'

CREDENTIALS 'aws_iam_role=YOUR-ROLE'

DELIMITER '|';

```

2. Crea un bucket en la región de AWS Acadmy Nort Virginia con el nombre midemobucketxxxxxx siendo xxxxx cualquier combinación de números y letras.
3. Copia el valor de Rol de LabRole desde IAM en la cuenta de AWS Academ. Empieza así: arn:aws:iam::
4. Pégallo en la ventana de la consulta, reemplazando el texto YOUR-ROLE (TU ROL).
5. Esta segunda línea ahora debería tener este aspecto: CREDENCIALES 'aws_iam_role=arn:aws:iam::
6. Haz clic en Ejecutar. El comando tardará unos 10 segundos en cargar 49 990 filas de datos.

Actividad:	RedShift							
Ciclo:	IABD	Modulo:	SBD	Curso:		Agrupación:	1	
Alumno/a:							Grupo:	
CIPFP www.fpmislata.com								

Tarea 5: Consultar datos

Ahora que tienes datos en tu base de datos de Redshift, puedes consultarlos mediante determinados enunciados y consultas SQL. Si conoces SQL, no dudes en probar otros comandos para consultar los datos.

1. Pon en marcha esta consulta para contar el número de filas de la tabla usuarios:

```
SELECT COUNT(*) FROM users;
```

El resultado muestra que la tabla tiene casi 50 000 filas.

2. Pon en marcha esta consulta:

```
SELECT userid, firstname, lastname, city, state
```

```
FROM users
```

```
WHERE likesports AND NOT likeopera AND state = 'OH'
```

```
ORDER BY firstname;
```

Esta consulta muestra los usuarios residentes en Ohio (OH) a los que les gusta el deporte, pero no la ópera. La lista está ordenada según el nombre de los usuarios.

3. Pon en marcha esta consulta:

```
SELECT city, COUNT(*) AS count
```

```
FROM users
```

```
WHERE likejazz
```

```
GROUP BY city
```

```
ORDER BY count DESC
```

```
LIMIT 10;
```

Nota:

Esta práctica esta sacada del Github de Santos Pardos

- <https://github.com/santos-pardos/Hands-On-Lab-in-AWS>