Imagen que contiene Texto

Descripción generada automáticamenteTALLER AWS

BIG DATA

Daniel Alejandro Olarte Ávila

Universidad Sergio Arboleda

Universidad Sergio Arboleda Cl. 74 #14-14

Bogotá, Colombia

Correo: [daniel.olarte01@correo.usa.edu.co](mailto:daniel.olarte01@correo.usa.edu.co)

Escuela de Ciencias Exactas e Ingeniería

Profesor: Camilo Rodriguez

26/04/2023

Omitir paso 1-6 si se cuenta con un BD RDS.

1. Instalar mysql en un EC2 con ubuntu, anotar el securitygroup y la subnet. sudo apt update

sudo apt install mysql-server Crear un usuario :

sudo mysql

create user 'admin'@'%' identified by 'xxxxx'; grant all privileges on \*.\* to 'admin'@'%';

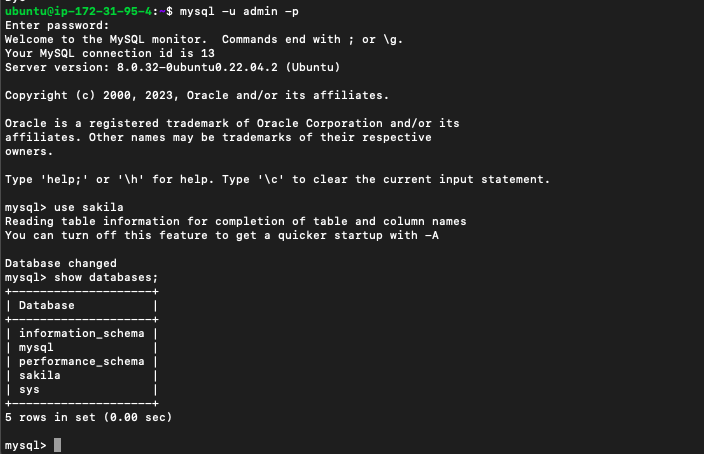
exit

1. Descargar de internet sakila y cargarlo en mysql wget <https://downloads.mysql.com/docs/sakila-db.zip> sudo apt install zip

unzip sakila-db.zip mysql -u admin -p

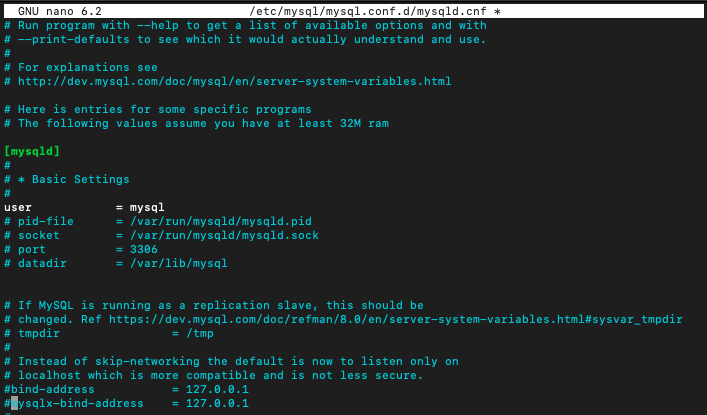
mysql> source sakila-db/sakila-schema.sql mysql> source sakila-db/sakila-data.sql exit

1. Realizar alguna consulta y verificar que todo está bien.



1. Modificar el archivo mysqld.cnf para poder hacer conexión remot sudo nano /etc/mysql/mysql.conf.d/mysqld.cnf

comentar las líneas:(agregar un #) # bind-address=127.0.0.1 #mysqlx-bind-address= 127.0.0.1

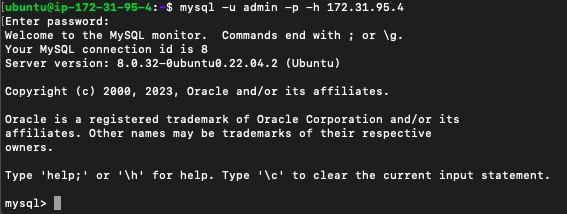


1. reiniciar mysql

sudo service mysql restart

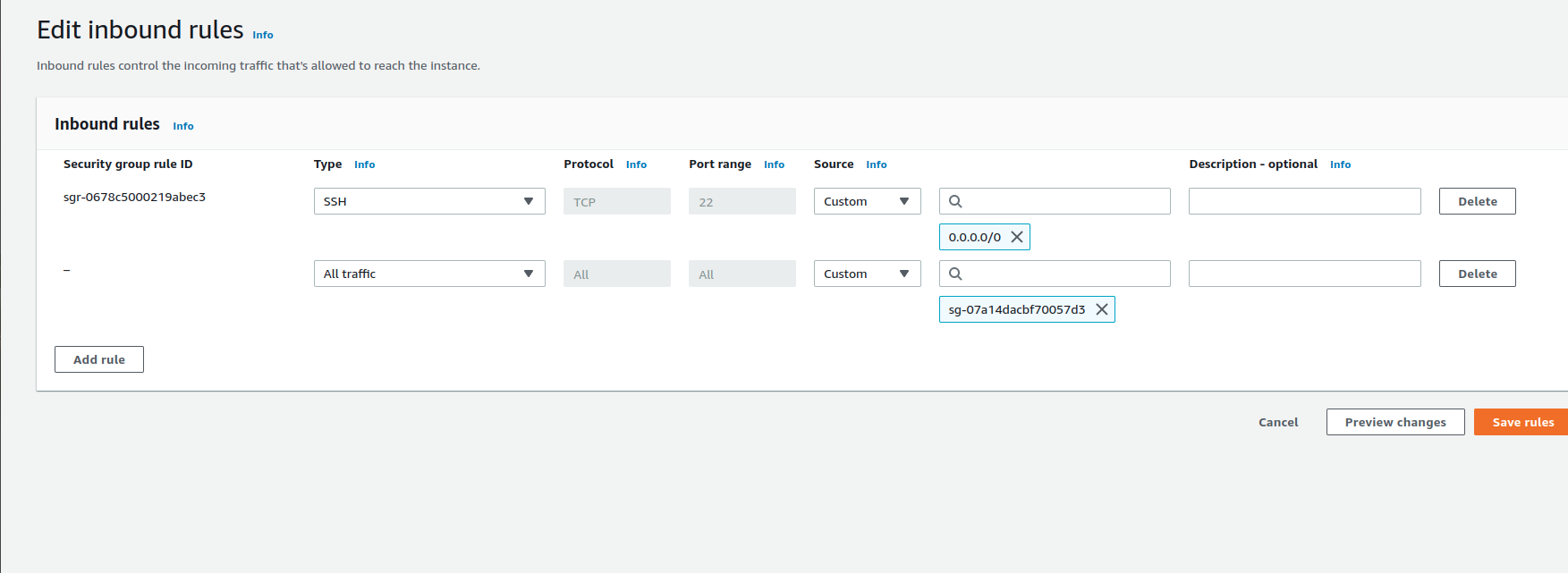
1. intentar hacer conexión usando la ip privada:

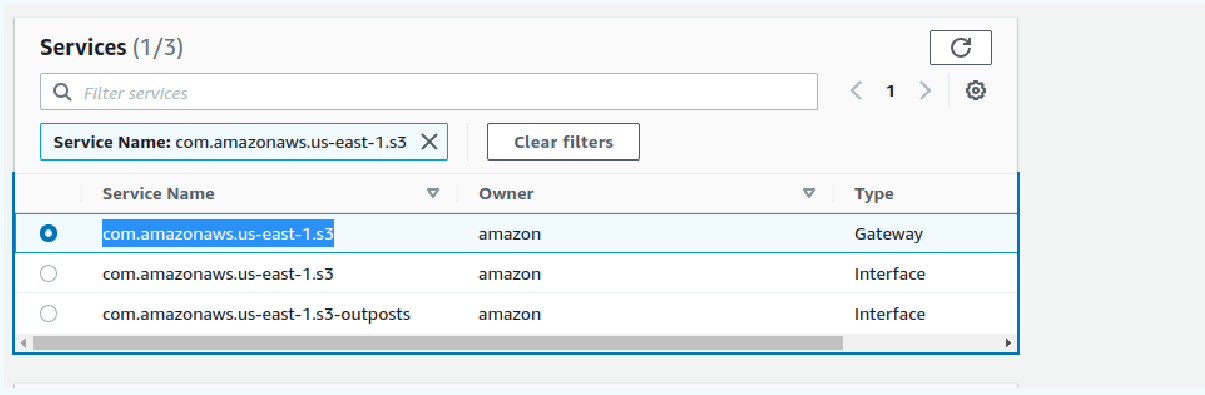
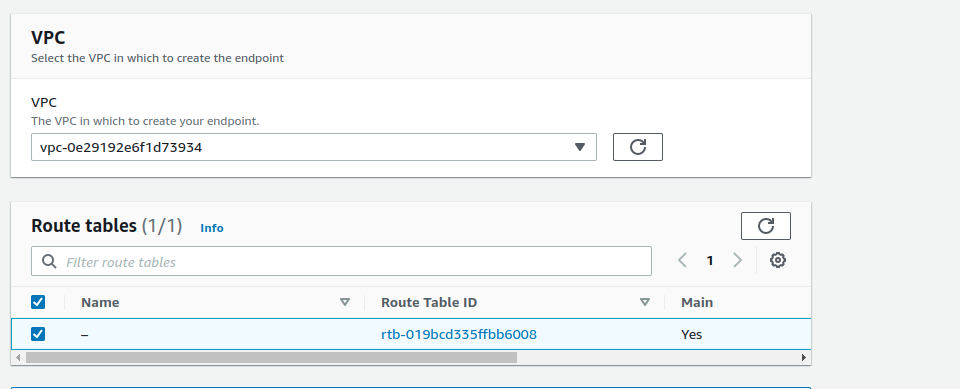
mysql -u admin -p -h 172.x.x.x

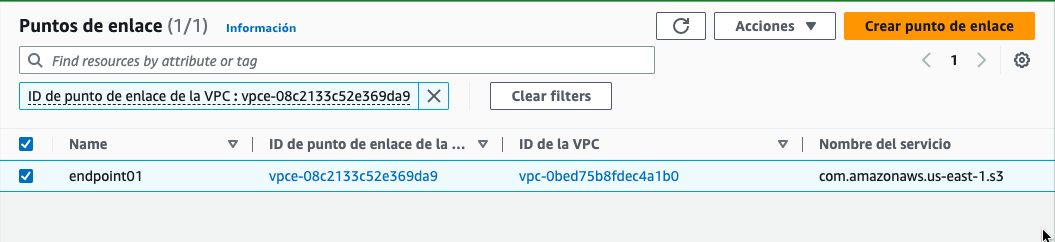


Si puede hacer conexión continuar, de lo contrario revisar los pasos anteriores.

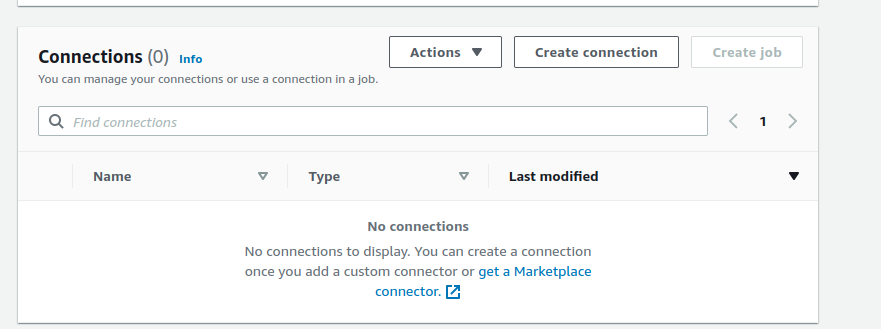
1. abrir todos los puertos de entrada(all trafic) en el security group de la instancia( o de la RDS), permitiendo entradas solo desde direcciones del mismo security group(no, anywhere).



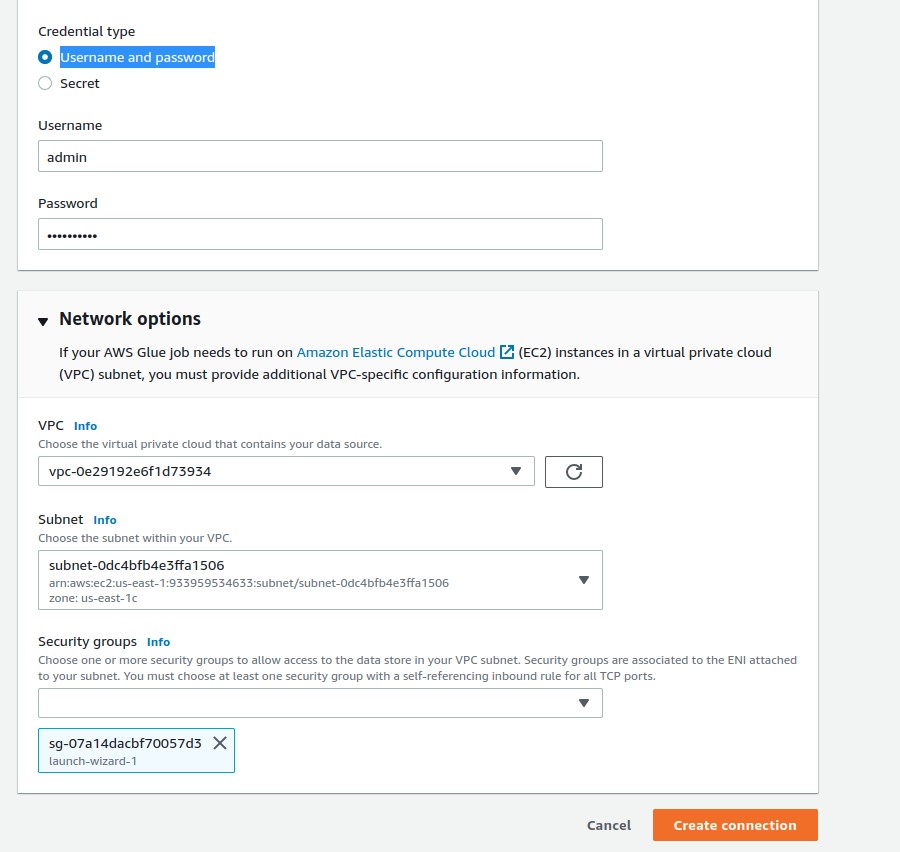
1. - Crear endpoint
   1. En la caja de servicio de Amazon, buscar VPC
   2. En el menú de la izquierda buscar endpoints
   3. Crear endpoint
      * Nombre: endpoint01
      * Services: com.amazonaws.us-east-1.s3 gateway
      * VPC: Default vpc
      * Seleccionar la única route table que aparece
        + Create endpoint



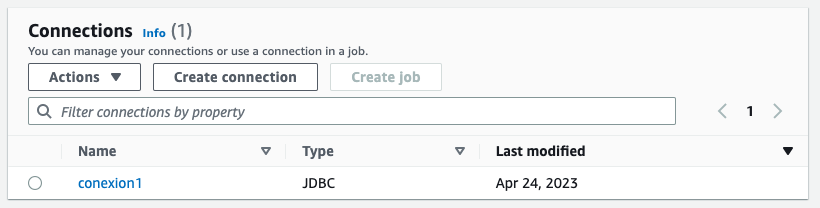
1. - Creación de una Conexión en AWS Glue
   1. ir a AWS GLue
   2. seleccionar en el menú de la izquierda Connections
   3. En connections -> create Connections



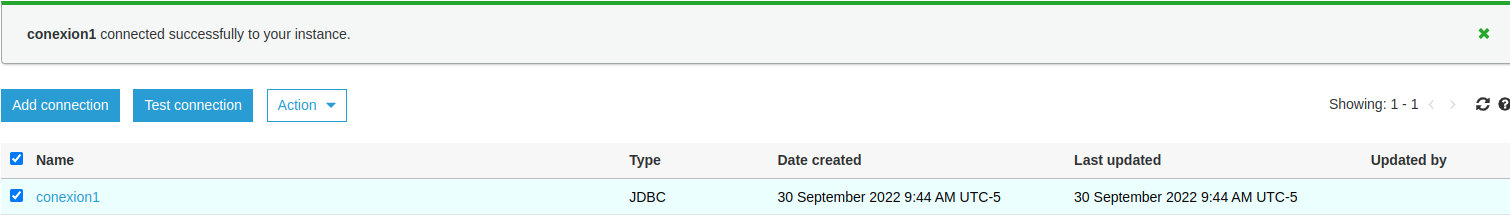
* 1. Name: conexion1
  2. Connection type: jdbc
  3. JDBC URL: jdbc:mysql:// 172.31.93.165:3306/sakila, reemplazar por la ip privada del EC2 donde se encuentra mysql o del dominio de la RDS
  4. Username and password corresponden a los creados en mysql
  5. Networks Options:

Corresponden a la vpc, subnet y security group del EC2 donde se encuentra Mysql

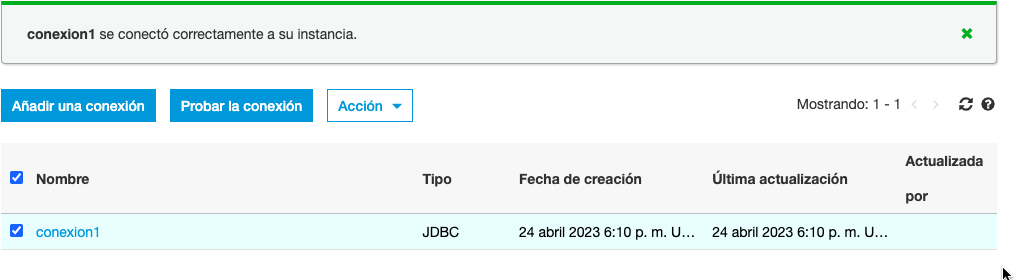
* 1. Create Connection



1. - Probar la conexión
   1. Ir AWS Glue
   2. En el menú de la izquierda ir a Legacy pages - > connections(Legacy)
   3. Seleccionar la conexión y darle clic en Test Connection
   4. Si la conexión es exitosa continuar, si no, revisar los pasos anteriores



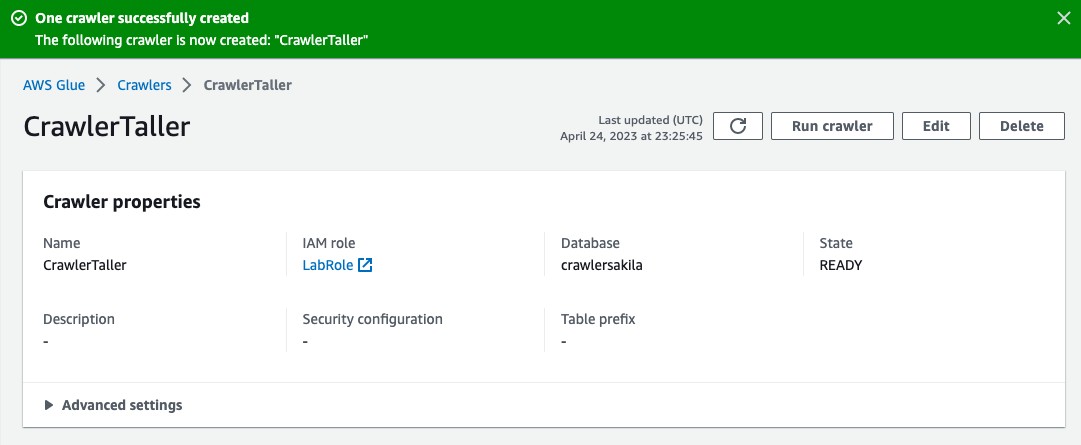
Conexión exitosa



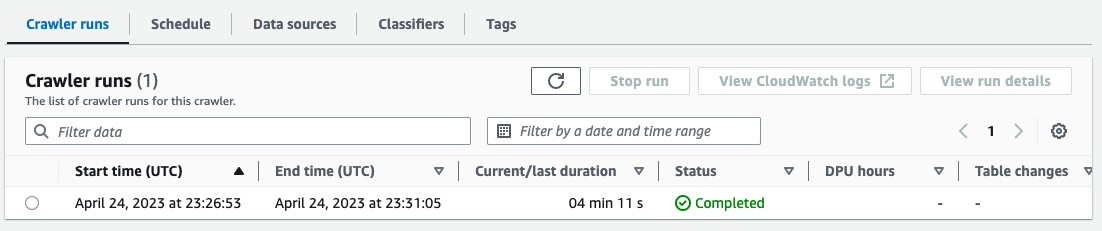
1. Crear un crawler que detecte el esquema de la base de datos y lo registre en el catálogo de Glue

Crear un nuevo crawler y la opción de de Add Data Source:

* + Seleccionar JDBC
  + Seleccionar la conexión creada
  + En include path digitar: sakila/% (quiere decir bd Sakila, todas las tablas)
  + Role: LabRole:
  + En Set output and scheduling: Add Database Sakila
  + Seleccionar BD Sakila
  + Crawler Schedule: On Demand
  + Crear Crwaler



1. Ejecutar el Crawler y verificar que las tablas se crean en el catálogo.



1. Crear y ejecutar un job que copie los datos de la tabla actor a un bucket en s3. Seleccionar la tabla del catálogo de Glue.

