First create a warehouse, database and role in snowflake

Using role account admin – which is kind of super user, create warehouse

And also create database and role

Grant usage on warehouse to the role created

Grant all access on database to the role

And then using the role, create a schema in the database

The goal is to transfer the data from Snowflake to Amazon S3 using the spark and Glue job

So basically 2 connectors are required

1. Connecting Spark with the snowflake
2. JDBC driver to connect snowflake with the Glue

Created a database called mydb in the snowflake

Also create table called EMPDATA inside the public schema of the mydb.

create or replace TABLE MYDB.PUBLIC.EMPDATA ( ID VARCHAR(16777216), FIRST\_NAME VARCHAR(16777216), LAST\_NAME VARCHAR(16777216), EMAIL VARCHAR(16777216), GENDER VARCHAR(16777216), IP\_ADDRESS VARCHAR(16777216) );

Now need to insert like at least 1000 rows of data into the snowflake.

Downloaded the randomly generated data into a csv file. And then uploaded to the snowflake

<https://jcuriwa-ot80342.snowflakecomputing.com>

Next……download the connectors…and what all connectors are needed

Refer to the Glue scripts In the following paths.

* C:\Users\srias\Documents\Data Engineer\SnowFlake\_learning\SnowFlake to AWS S3 Data Transfer

Drivers used for this tutorial are :

* snowflake-jdbc-3.13.22.jar
* spark-snowflake\_2.12-2.13.0-spark\_3.3.jar

Issues faced while running the glue job are:

1. Initially I used another driver which is not compatible with the glue selection:

spark-snowflake\_2.11-2.9.3-spark\_2.4.jar

This made the job fail for multiple times, then I upgraded the driver to

* spark-snowflake\_2.12-2.13.0-spark\_3.3.jar

which is supported by - **Glue 4.0 –**Supports Spark 3.3.0 and Python 3.10.

After the Job is succeeded, check for the Output logs

<https://repo1.maven.org/maven2/net/snowflake/spark-snowflake_2.12/2.13.0-spark_3.3/>

<https://repo1.maven.org/maven2/net/snowflake/snowflake-jdbc/3.13.22/>

A screenshot of a computer

Description automatically generated