1. **Create an S3 bucket with a unique name and upload the CSV file to the S3 bucket (ensure that the file is in UTF-8 format only)**

Graphical user interface, application, Teams

Description automatically generated Graphical user interface, application, Teams

Description automatically generated

1. Create a folder for placing datasets in it:

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface

Description automatically generated

1. **Create a crawler to crawl the CSV data and generate a metadata catalog.**

Graphical user interface, application

Description automatically generated

* 1. Select the folder that file is in there(not the file itself):

Graphical user interface, application, Teams

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

* 1. Create an IAM

Graphical user interface

Description automatically generated

* 1. After Role is created, run it, and table will be created

Graphical user interface, application

Description automatically generated

1. **Create a Glue job to transform the data into the Parquet format as CSV is not optimal for data warehouse queries**

Create a Job to Convert CSV to Parquet file:

1. Create IAM ROLE FOR GLUE  
     
   A computer screen capture

   Description automatically generated with medium confidence
2. Create a Glue Job  
     
   A screenshot of a computer

   Description automatically generated
3. Create a name, add IAM role in job detail  
     
   Graphical user interface, application

   Description automatically generated
4. Create a folder output for output target:  
     
   Graphical user interface, application

   Description automatically generated
5. Create source, target and transform action to change datatype.

A screenshot of a computer

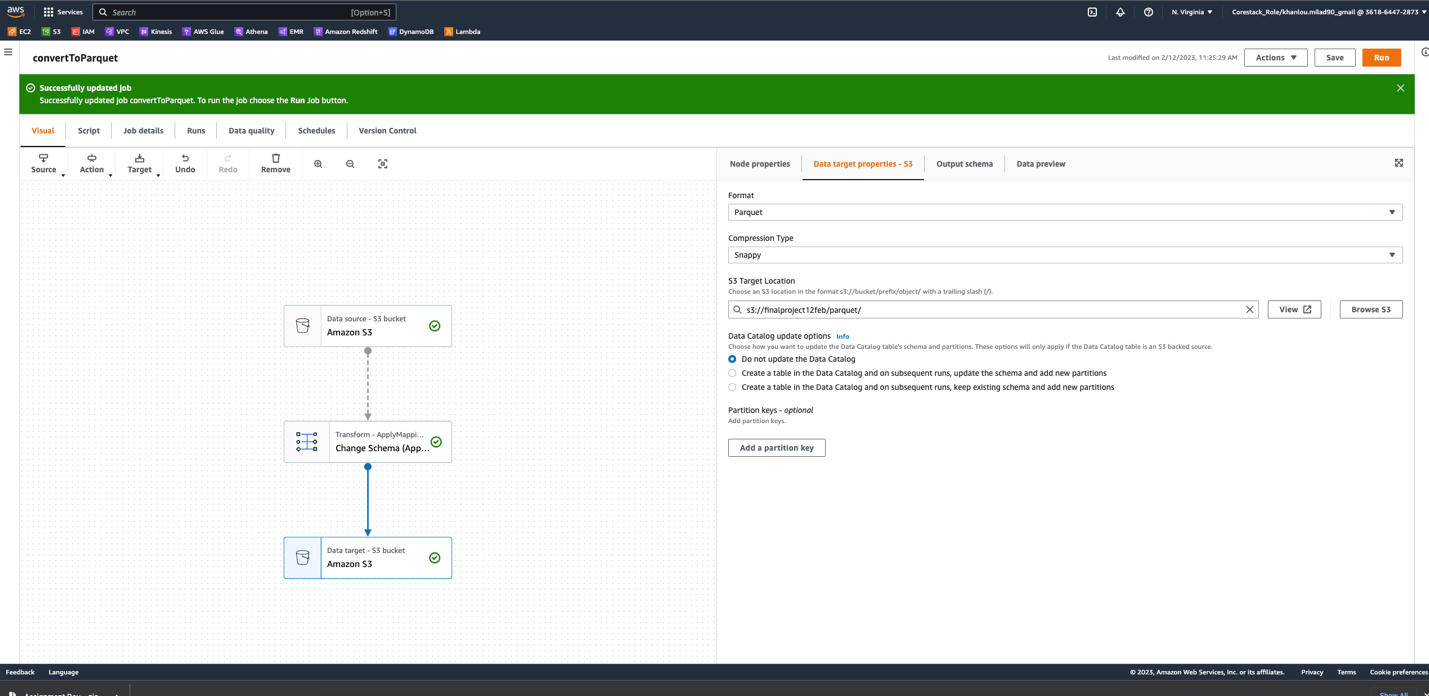
Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Done and save and RUN  
  


JOB CREATED PARQUET FILE IN S3 BUCKET:

Graphical user interface, text, application

Description automatically generated

1. **Add another crawler to crawl the Parquet data files to generate the metadata catalog of the Parquet file in order to query it with Athena**
2. Create a Database:  
     
   Graphical user interface, text, application, email

   Description automatically generated
3. Create a crawler for parquet file  
     
   Graphical user interface, text, application, email

   Description automatically generated  
     
     
     
   Graphical user interface, application

   Description automatically generated  
     
     
   Graphical user interface, text, application, email

   Description automatically generated  
   Create an IAM role:  
     
   Graphical user interface

   Description automatically generated  
     
   Graphical user interface, text, application, email

   Description automatically generated
5. SAVE AND RUN THE CRAWLER:  
     
   A screenshot of a computer

   Description automatically generated  
     
   Graphical user interface

   Description automatically generated
6. TABLE CREATED in separate database  
     
   Graphical user interface

   Description automatically generated
7. **Query the data to identify the best-selling item and countries where customers have bought the most-sold item using Athena**

In Athena choose the database and query:

Graphical user interface, application

Description automatically generated