

DATAPIPELINE INSIGHT HUB

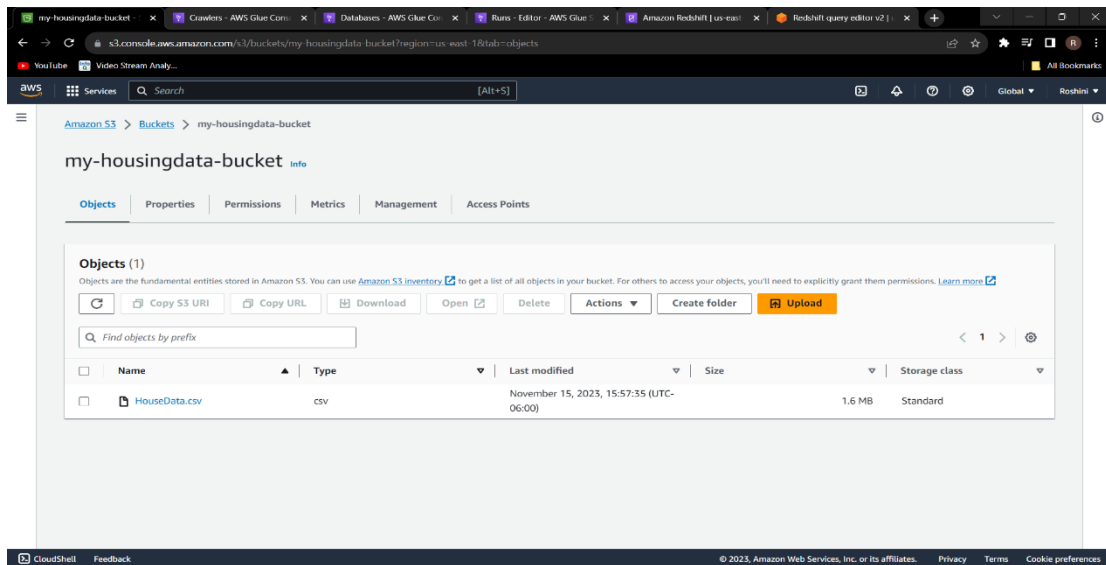
Aim: To create a ETL Pipeline in AWS Cloud to store data and produces insights in Power BI

Services used: Amazon Account, S3 Buckets, AWS Glue, Redshift, and Power BI

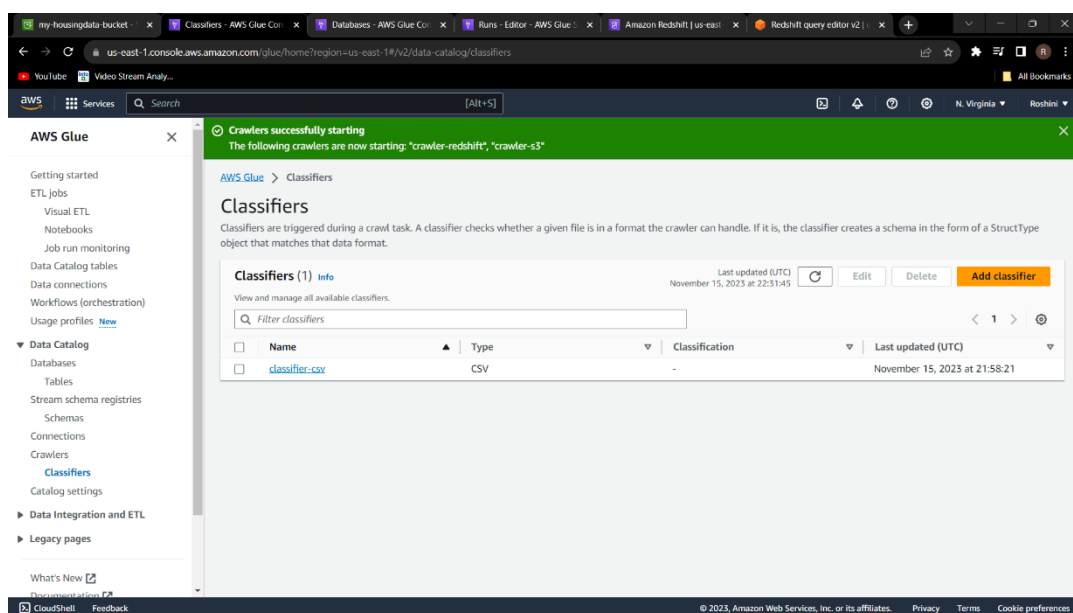
Steps:

The important part of this project is to upload and store the data in AWS Cloud.

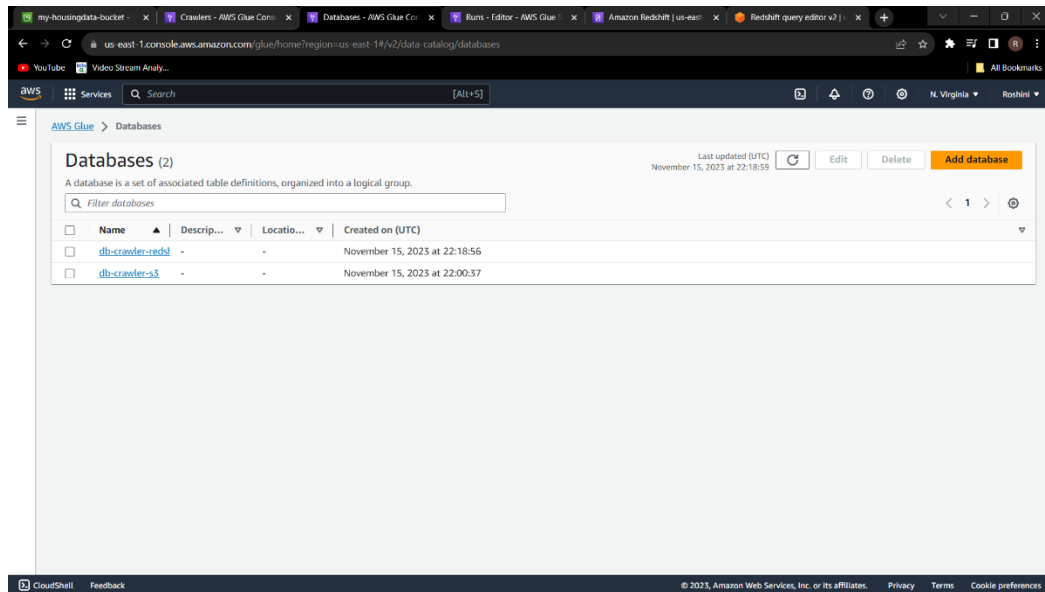
1. I have considered a housing dataset online.
2. Login to your Amazon account and create a S3 Bucket in any region.
3. Upload the file into the bucket created as below.



4. Open AWS Glue from the console and create a classifier.
Classifier helps in reading the data into a data source.



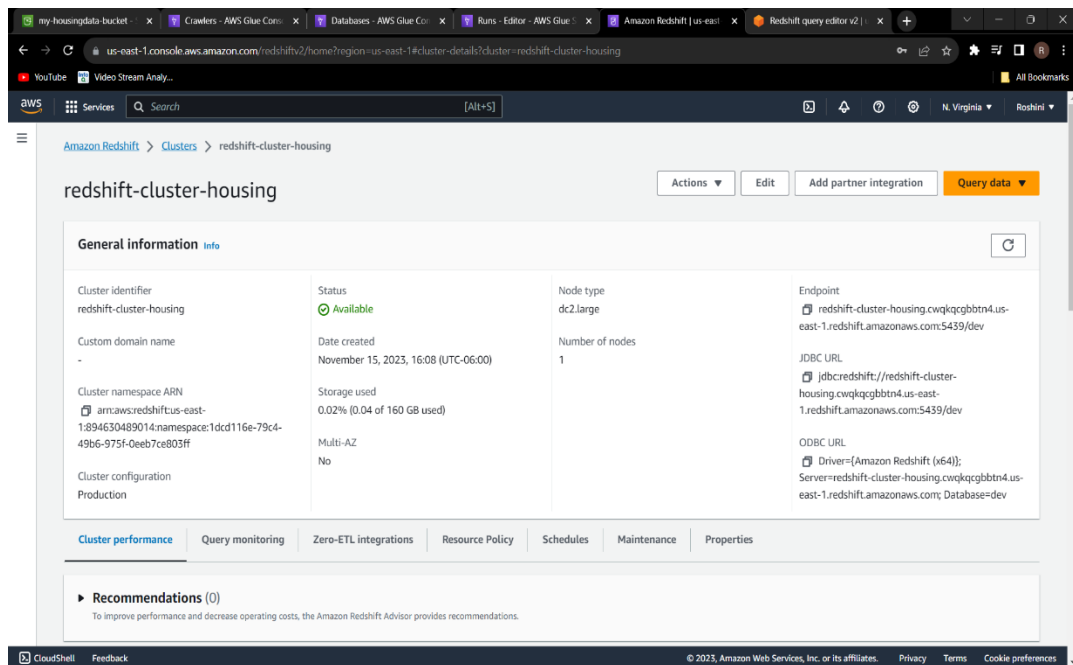
5. Create the databases for both the bucket and the redshift cluster.



6. Create a Redshift cluster.

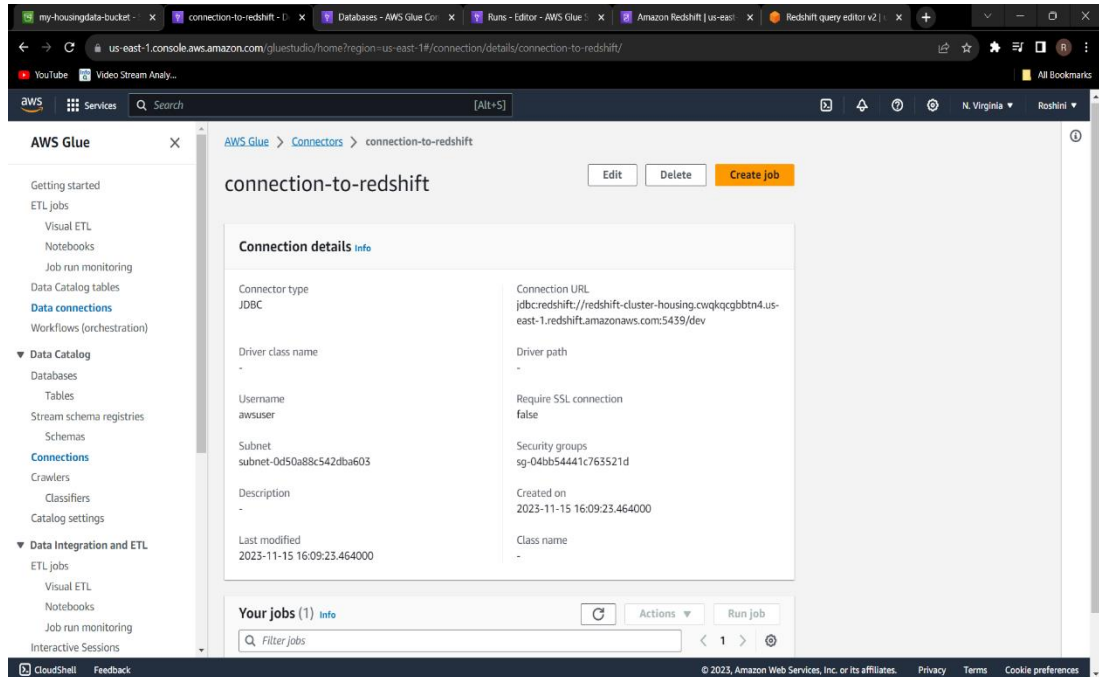
It may take some time to create the cluster.

Create the authentication credentials while creating the cluster to access it using crawlers.

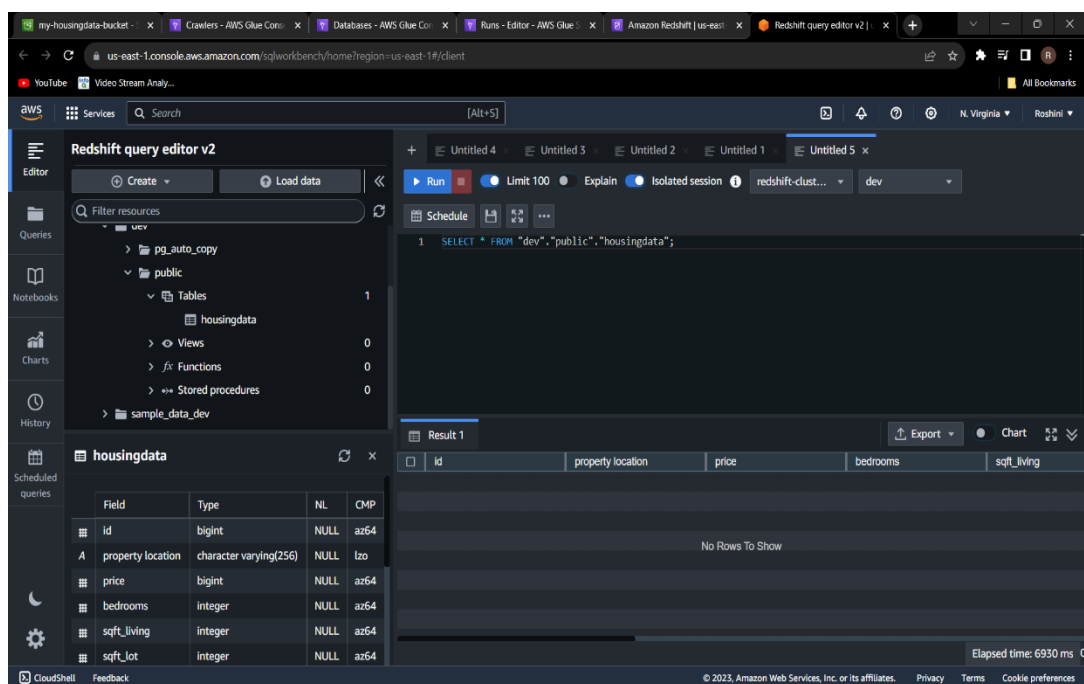


7. Create a crawler for the S3 Bucket.

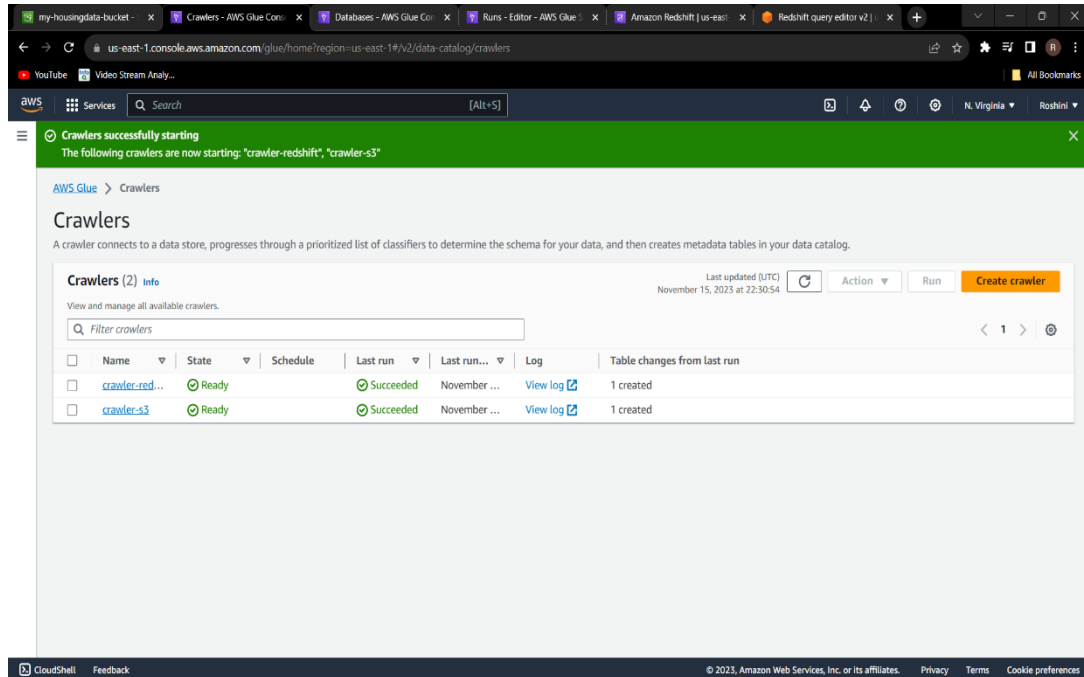
8. Create a connection to the redshift.



9. Create the table in the cluster using the Query Editor available in Redshift cluster console. SQL Query can be used in running the table and adding values to the table.



10. Create the crawlers for the redshift cluster.



11. When both the crawlers are ready, start creating Visual jobs in AWS Glue.
12. Drag and drop the source, Transform used and the destination where data should be transferred.
13. Automatic script will be produced after dragging the tabs onto the dashboard.
14. Start the running the job.
15. When the job is successful, open Power BI.
16. Import data by connecting the Power BI to AWS Redshift Cluster using the plugins.
17. Transform the data in Power BI Desktop and start building the visualization dashboard.

