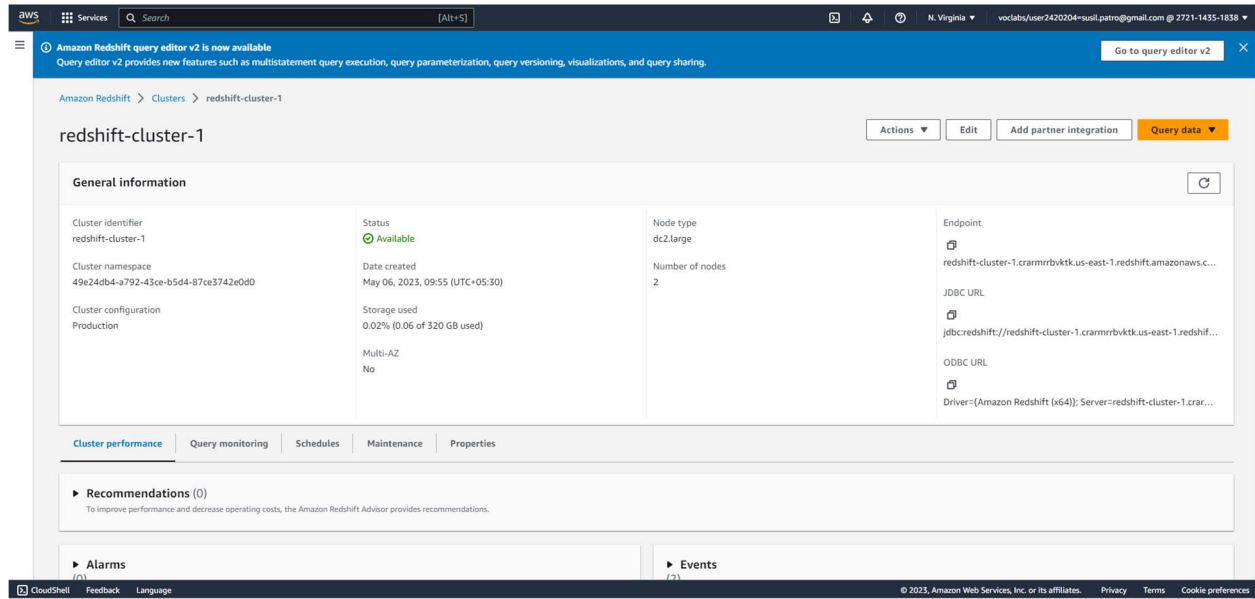


Creation of a Redshift Cluster

Screenshots of the configuration of the Redshift cluster that you have created:

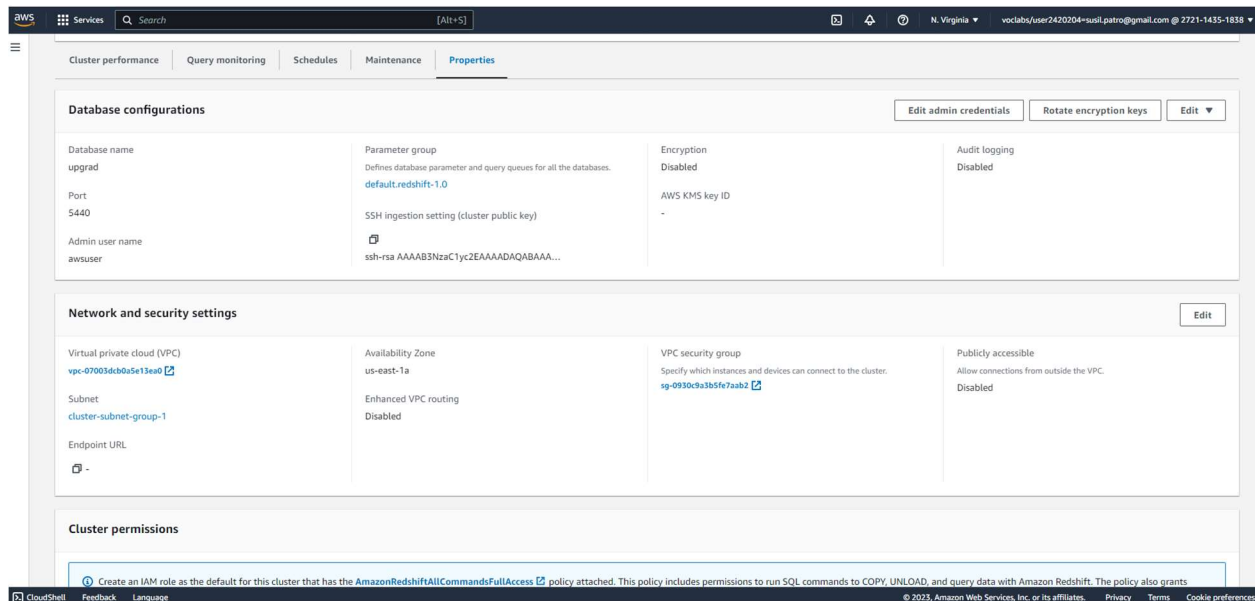
Screenshot of the type of machine used along with number of nodes:



The screenshot shows the Amazon Redshift console interface. At the top, there's a navigation bar with 'AWS' and 'Services' tabs. Below it, a blue banner announces 'Amazon Redshift query editor v2 is now available'. The main content area is titled 'redshift-cluster-1' and includes a 'General information' section. This section contains a table with the following details:

Cluster identifier	Status	Node type	Endpoint
redshift-cluster-1	Available	dc2.large	redshift-cluster-1.crarnrbvktk.us-east-1.redshift.amazonaws.com
Cluster namespace	Date created	Number of nodes	JDBC URL
49e24db4-a792-43ce-b5d4-87ce3742e0d0	May 06, 2023, 09:55 (UTC+05:30)	2	jdbc:redshift://redshift-cluster-1.crarnrbvktk.us-east-1.redshift.amazonaws.com
Cluster configuration	Storage used		ODBC URL
Production	0.02% (0.06 of 320 GB used)		Driver=(Amazon Redshift (x64)); Server=redshift-cluster-1.crar...
	Multi-AZ		
	No		

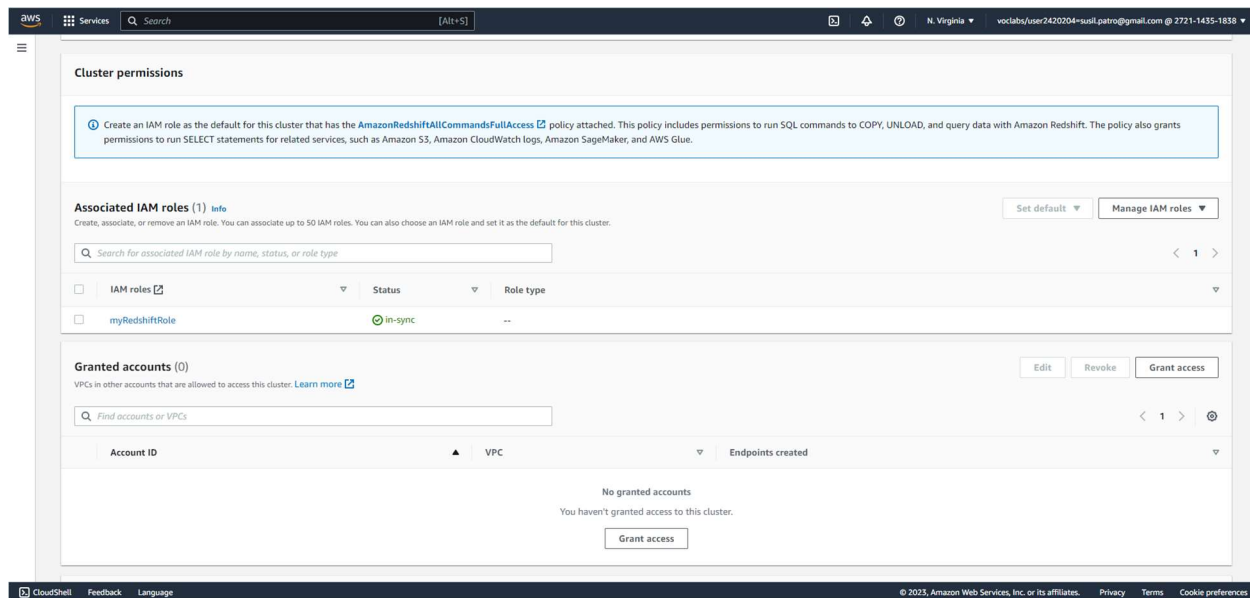
Below the 'General information' section, there are tabs for 'Cluster performance', 'Query monitoring', 'Schedules', 'Maintenance', and 'Properties'. The 'Recommendations (0)' section is also visible, indicating no recommendations are currently present.



The screenshot shows the Amazon Redshift console interface, specifically the 'Properties' tab for the cluster 'redshift-cluster-1'. The 'Database configurations' section is expanded, showing the following details:

Database name	Parameter group	Encryption	Audit logging
upgrad	default.redshift-1.0	Disabled	Disabled
Port	SSH ingestion setting (cluster public key)	AWS KMS key ID	
5440	ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQ...	-	
Admin user name			
awsuser			

Below the 'Database configurations' section, there are tabs for 'Cluster performance', 'Query monitoring', 'Schedules', 'Maintenance', and 'Properties'. The 'Network and security settings' section is also visible, showing details for the Virtual private cloud (VPC), Subnet, and VPC security group.

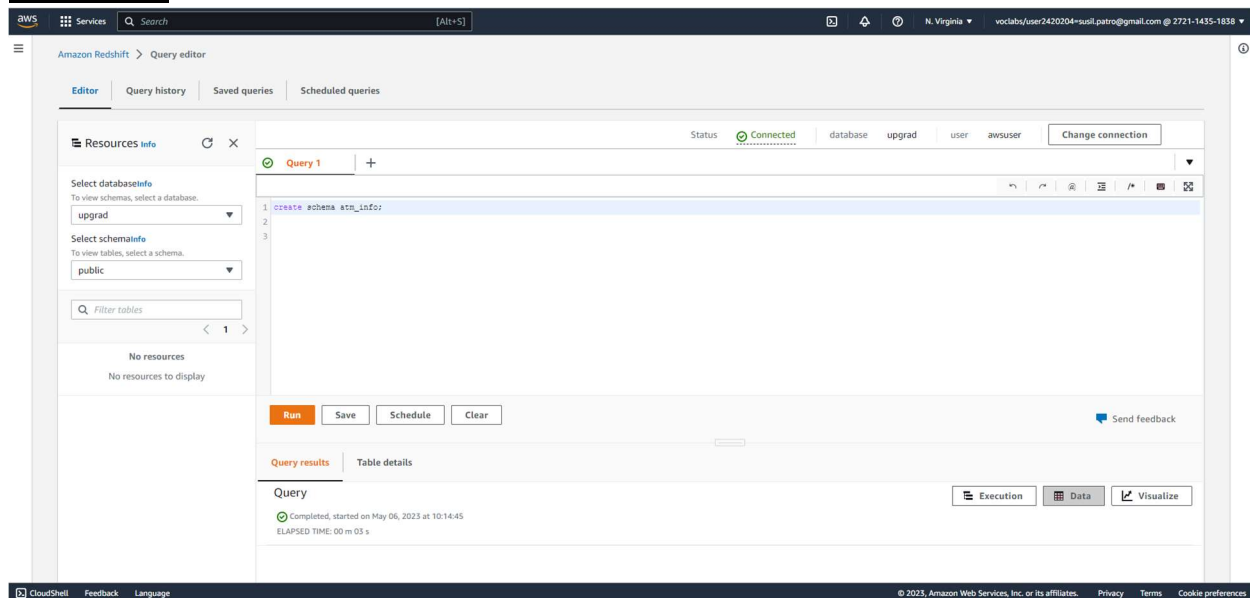


Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Query for creating schema:

create schema atm_info;

Screenshot:

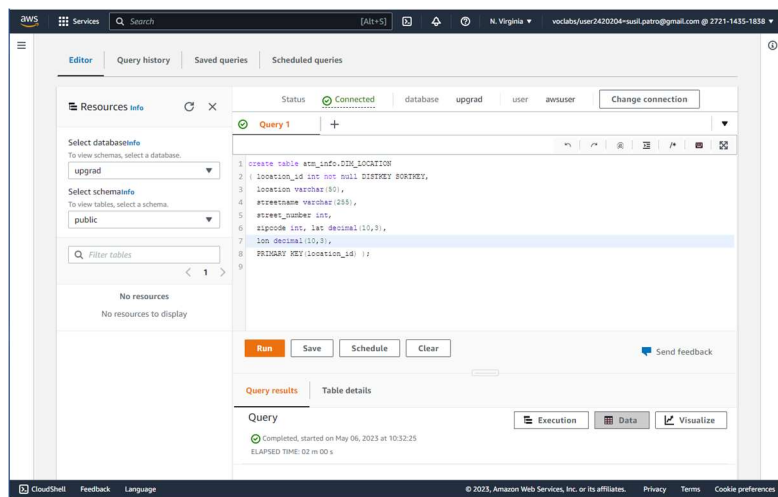


Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

Queries:

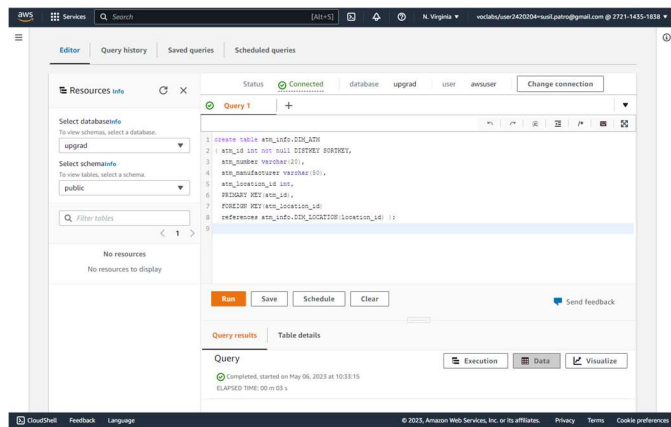
Creating location dimension table:

```
create table atm_info.DIM_LOCATION
( location_id int not null DISTKEY SORTKEY,
  location varchar(50),
  streetname varchar(255),
  street_number int,
  zipcode int, lat decimal(10,3),
  lon decimal(10,3),
  PRIMARY KEY(location_id) );
```



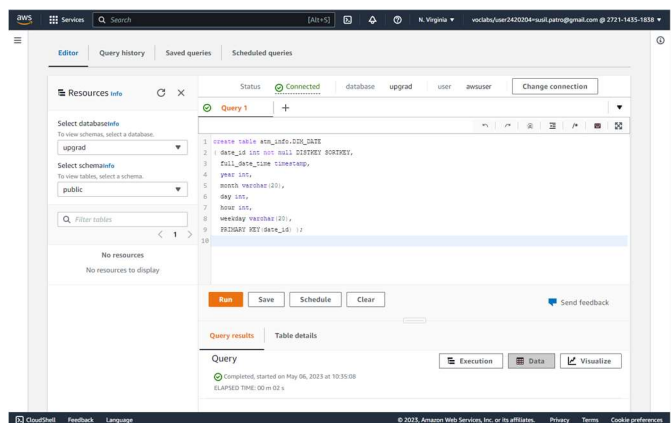
Creating atm dimension table:

```
create table atm_info.DIM_ATM
( atm_id int not null DISTKEY SORTKEY,
  atm_number varchar(20),
  atm_manufacturer varchar(50),
  atm_location_id int,
  PRIMARY KEY(atm_id),
  FOREIGN KEY(atm_location_id)
  references atm_info.DIM_LOCATION(location_id) );
```



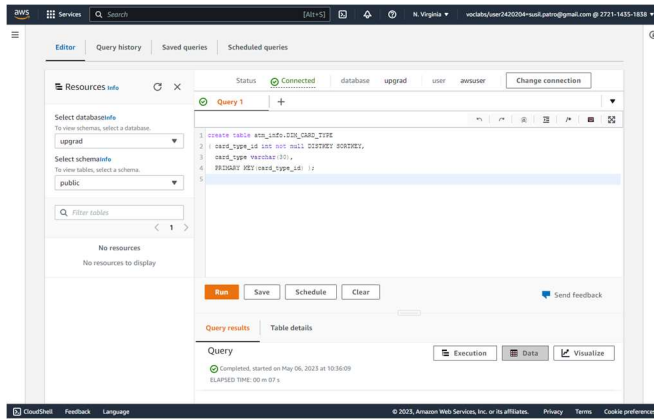
Creating date dimension table:

```
create table atm_info.DIM_DATE
( date_id int not null DISTKEY SORTKEY,
  full_date_time timestamp,
  year int,
  month varchar(20),
  day int,
  hour int,
  weekday varchar(20),
  PRIMARY KEY(date_id) );
```



Creating card type dimension table:

create table atm_info.DIM_CARD_TYPE
 (card_type_id int not null DISTKEY SORTKEY,
 card_type varchar(30),
 PRIMARY KEY(card_type_id));



Creating atm transactions fact table:

create table atm_info.FACT_ATM_TRANS

(trans_id bigint not null DISTKEY SORTKEY,

atm_id int, weather_loc_id int,

date_id int,

card_type_id int,

atm_status varchar(20),

currency varchar(10),

service varchar(20),

transaction_amount int,

message_code varchar(225),

message_text varchar(225),

rain_3h decimal(10,3),

clouds_all int,

weather_id int,

weather_main varchar(50),

weather_description varchar(255),

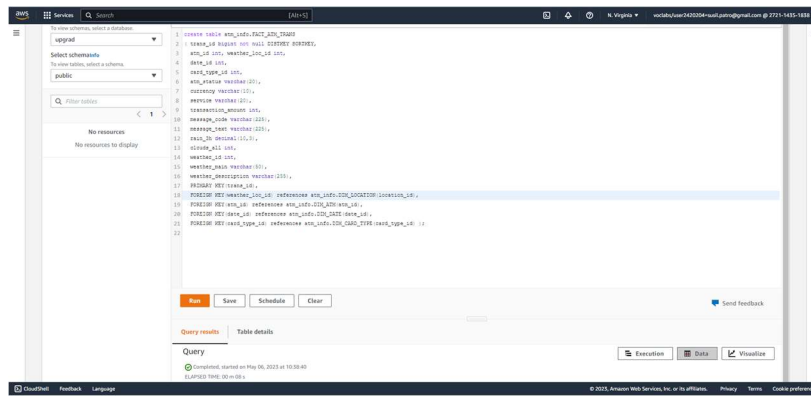
PRIMARY KEY(trans_id),

FOREIGN KEY(weather_loc_id) references atm_info.DIM_LOCATION(location_id),

FOREIGN KEY(atm_id) references atm_info.DIM_ATM(atm_id),

FOREIGN KEY(date_id) references atm_info.DIM_DATE(date_id),

FOREIGN KEY(card_type_id) references atm_info.DIM_CARD_TYPE(card_type_id));

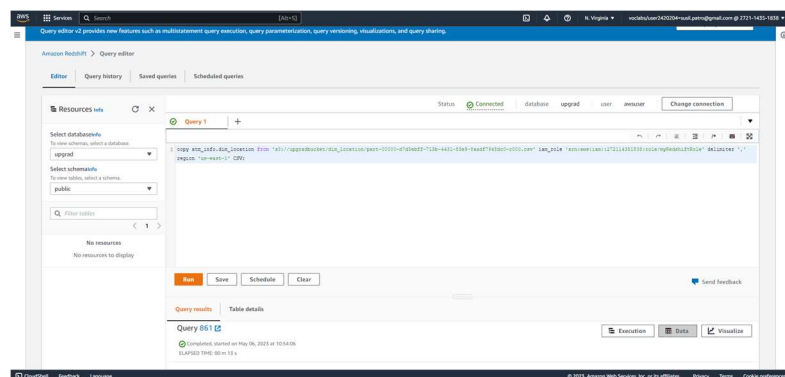


Loading data into a Redshift cluster from Amazon S3 bucket

Queries to copy the data from S3 buckets to the Redshift cluster in the appropriate tables

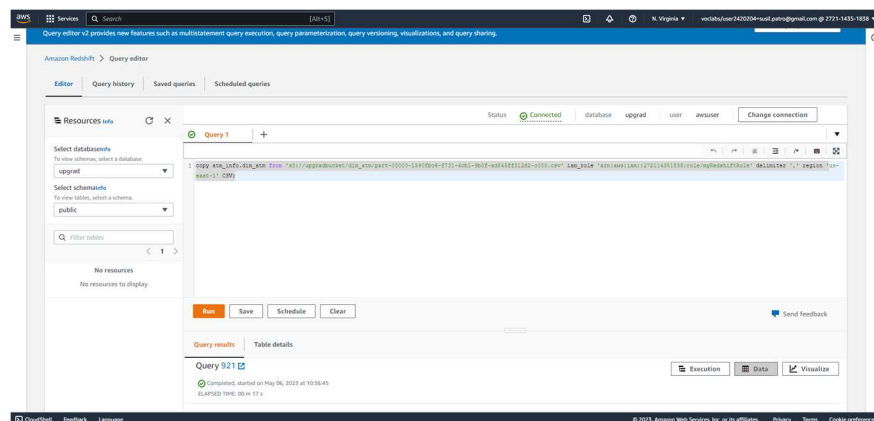
Copying the data to dim_location table:

copy atm_info.dim_location from 's3://upgradbucket/dim_location/part-00000-d7d5ebff-715b-4431-85e9-8aadf7948dc0-c000.csv' iam_role 'arn:aws:iam::272114351838:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;



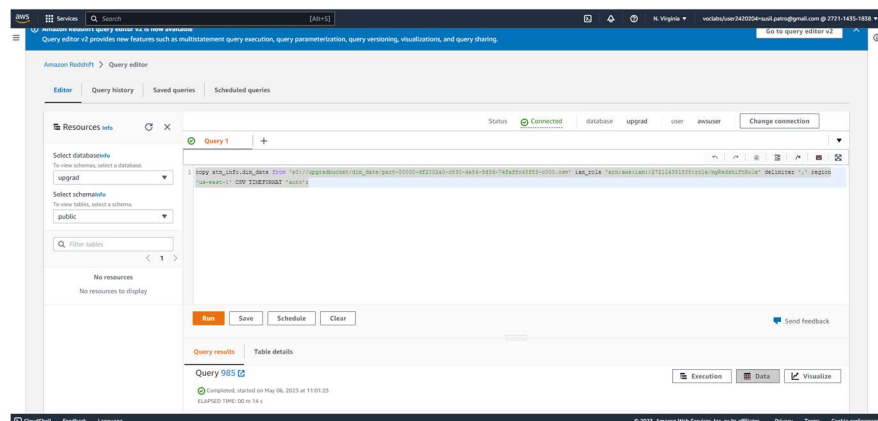
Copying the data to dim_atm table:

copy atm_info.dim_atm from 's3://upgradbucket/dim_atm/part-00000-1590fbc6-8731-4cb1-9b0f-ad845ff312d2-c000.csv' iam_role 'arn:aws:iam::272114351838:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;



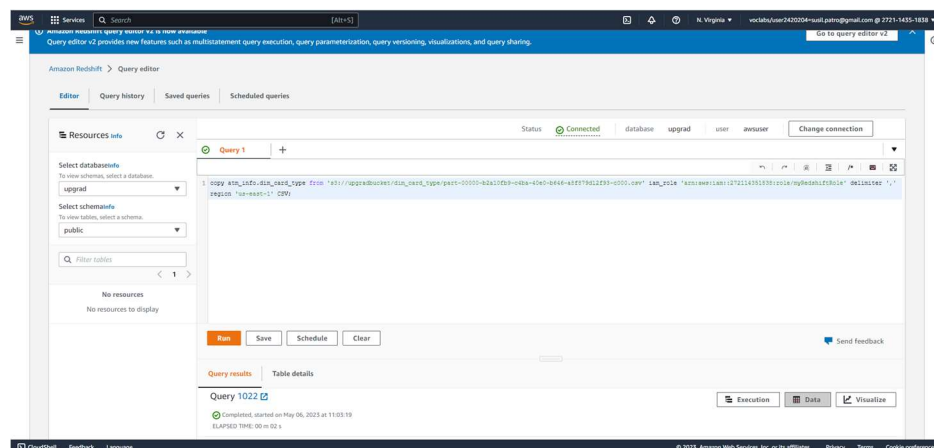
Copying the data to dim_date table:

copy atm_info.dim_date from 's3://upgradbucket/dim_date/part-00000-df2002a0-c830-4a84-8d3d-74faf8c48f83-c000.csv' iam_role 'arn:aws:iam::272114351838:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV TIMEFORMAT 'auto';



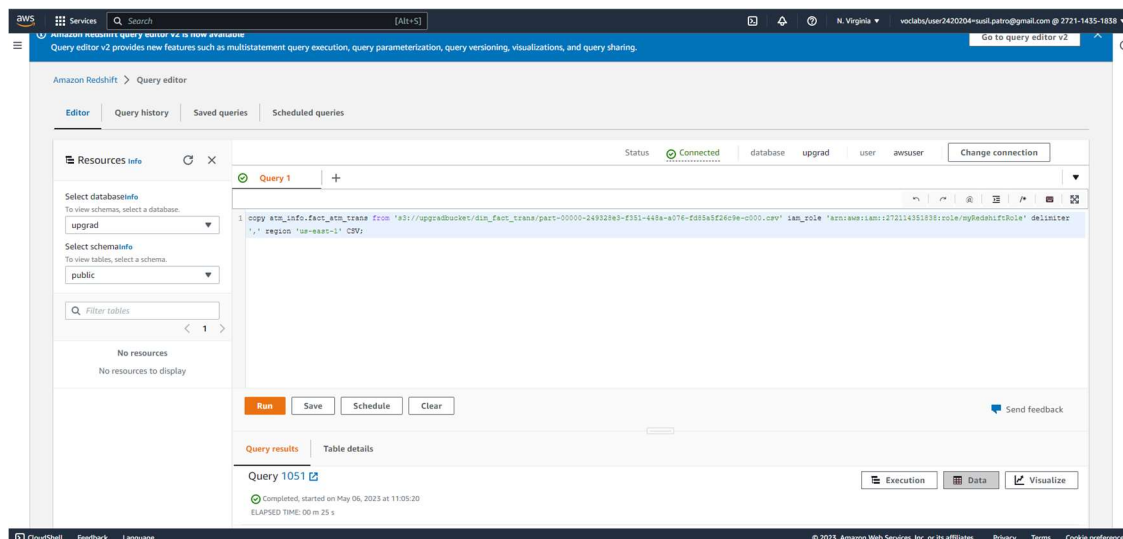
Copying the data to dim_card_type table:

copy atm_info.dim_card_type from 's3://upgradbucket/dim_card_type/part-00000-b2a10fb9-c4ba-40e0-b646-a8f879d12f93-c000.csv' iam_role 'arn:aws:iam::272114351838:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;



Copying the data to fact atm_trans table:

copy atm_info.fact_atm_trans from 's3://upgradbucket/dim_fact_trans/part-00000-249328e3-f351-448a-a076-fd85a5f26c9e-c000.csv' iam_role
'arn:aws:iam::272114351838:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;



Once all data copied onto dimension and fact tables for atm_info schema, queries for analysis will be done. The queries are illustrated with evidence in separate document “RedshiftQueries.pdf”

Case Study submitted by:
Susil Patro, Vivek Agrawal & Harshal Sapkade

