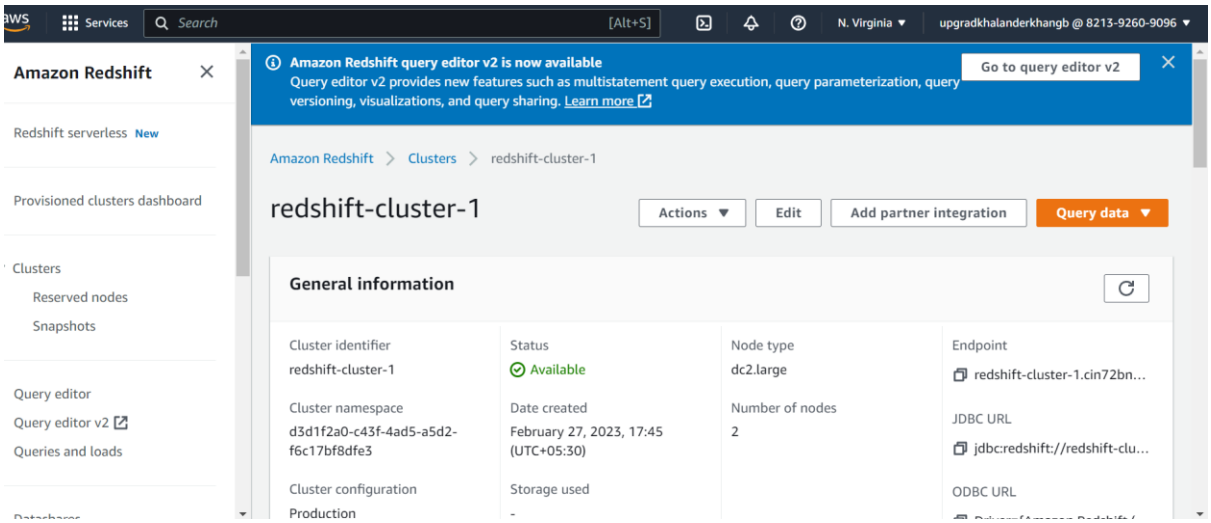
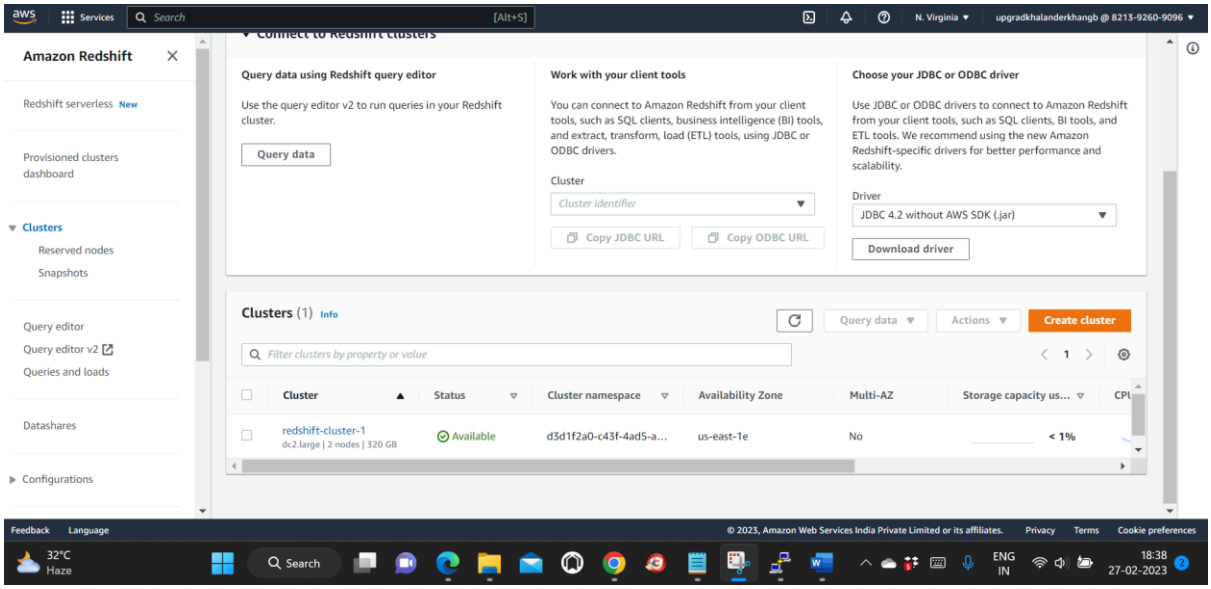
Creation of a Redshift Cluster

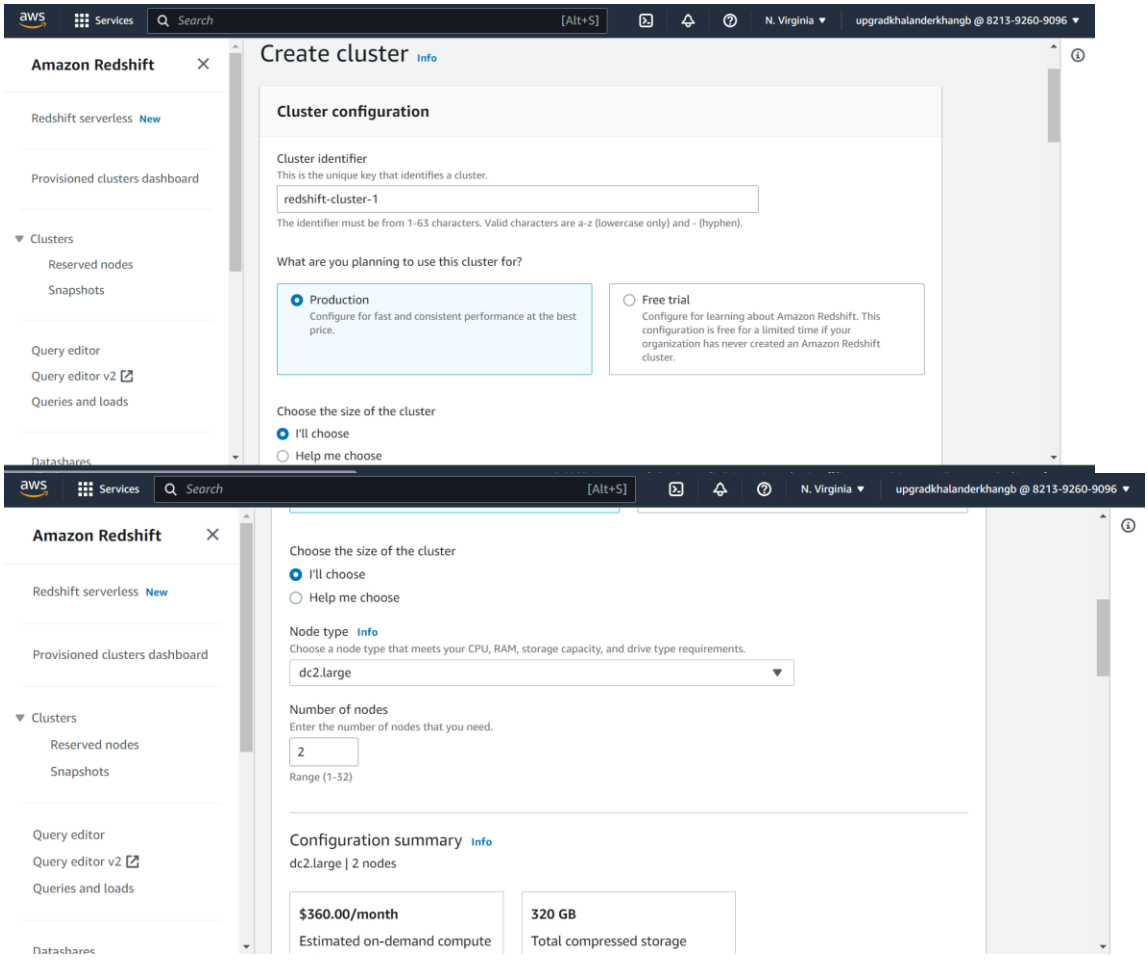
Screenshots of the configuration of the Redshift cluster that you have created:

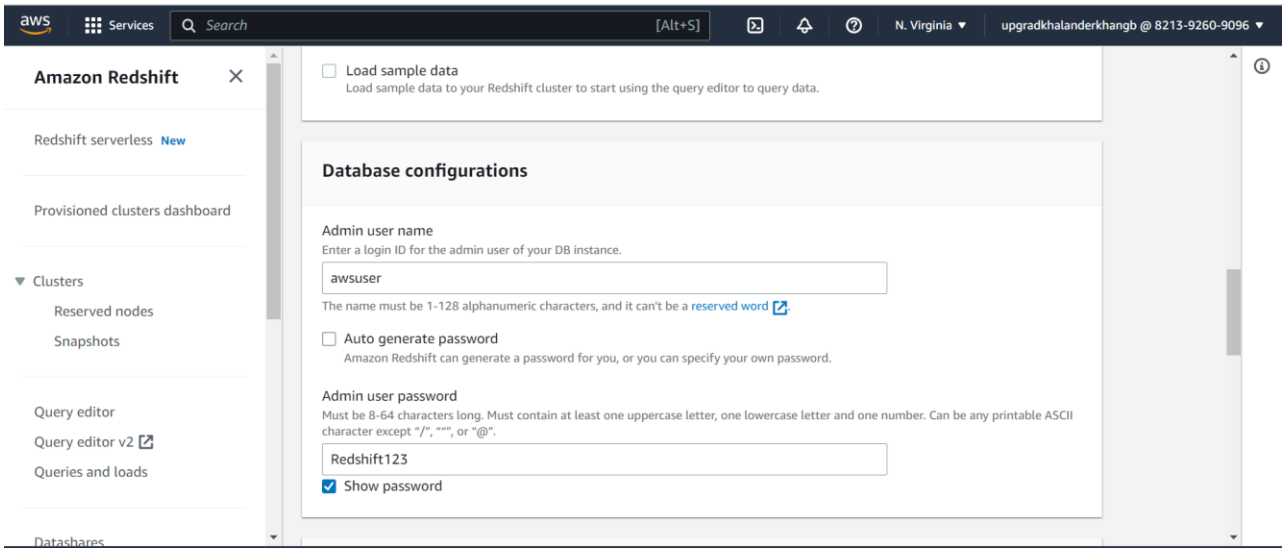
The cluster created is redshift-cluster-1 and there are two nodes used

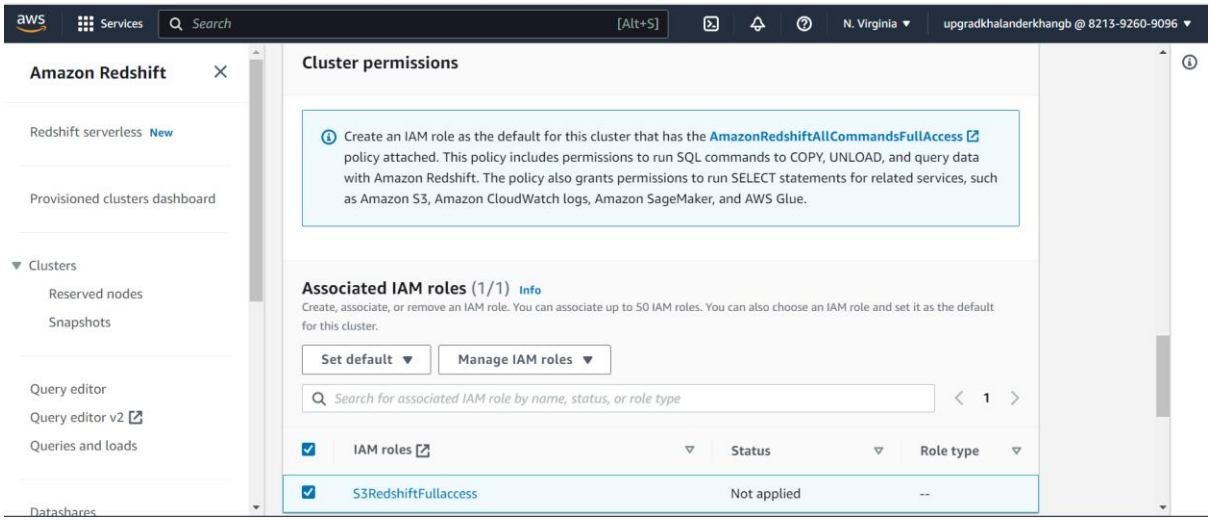


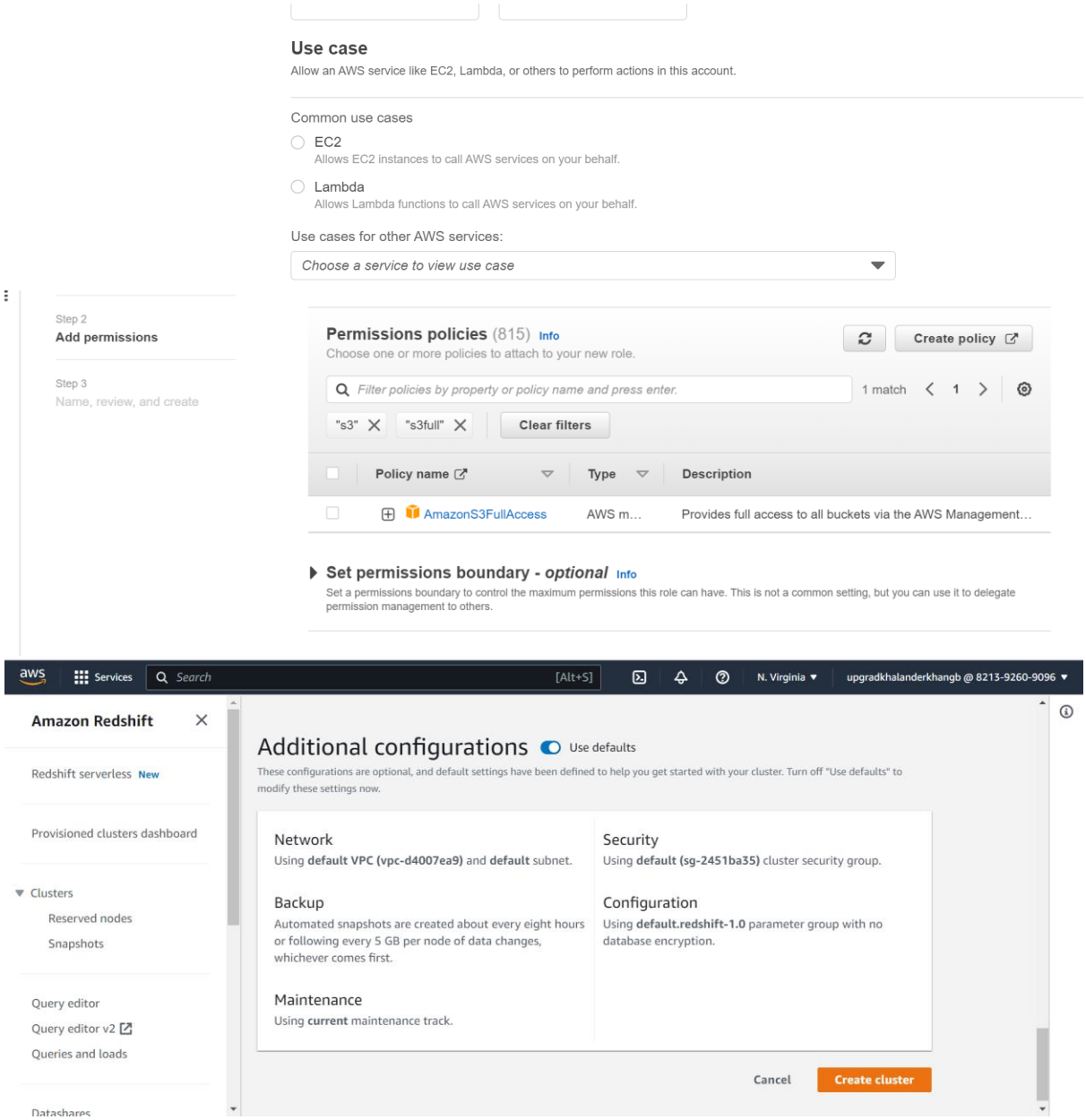


The configurations used are –





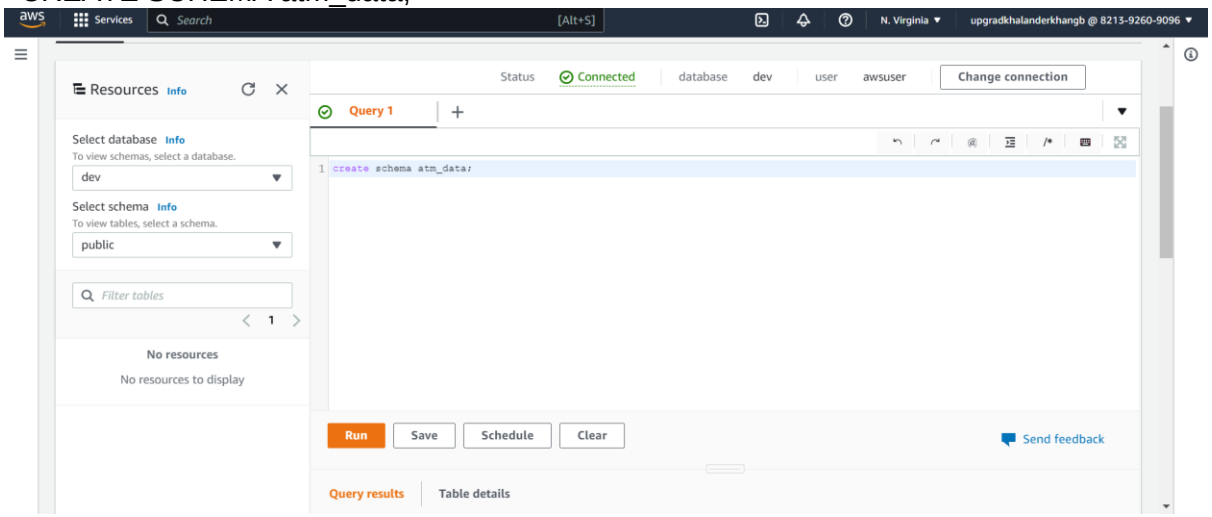




Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Query to create the database's schema:

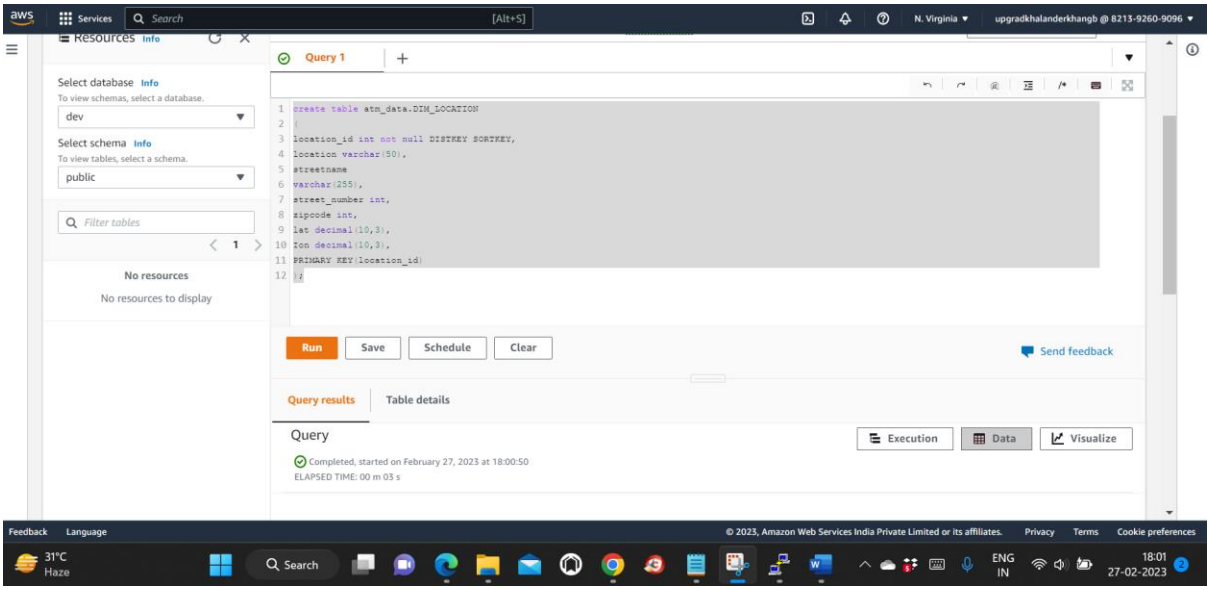
CREATE SCHEMA atm_data;



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

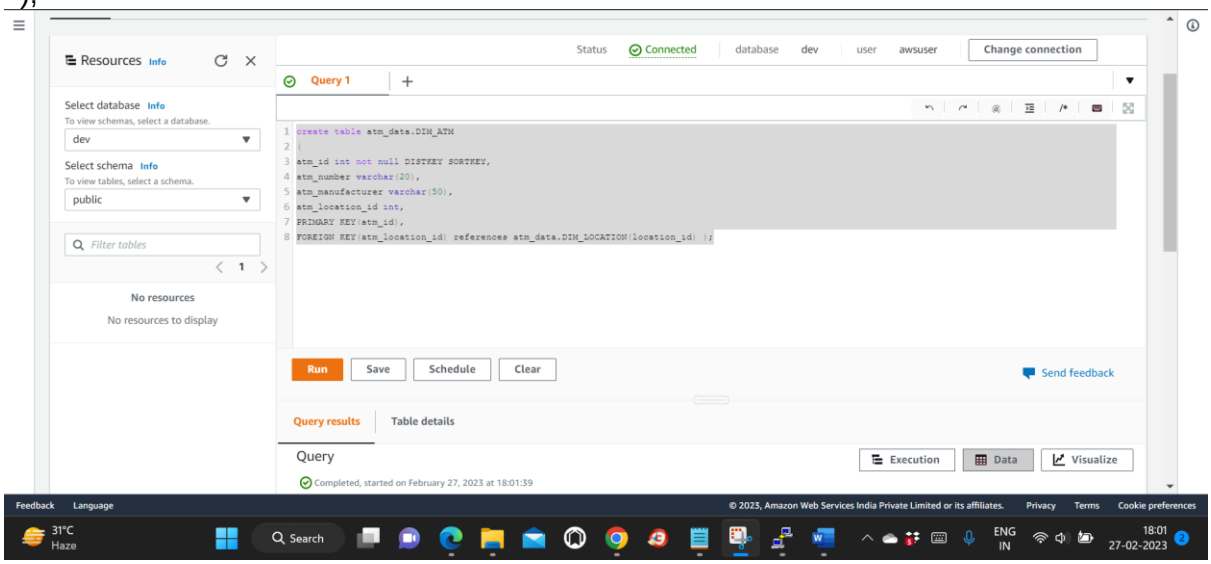
Query to create the dimension table DIM_LOCATION:

```
CREATE TABLE atm_nord.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)
);
```



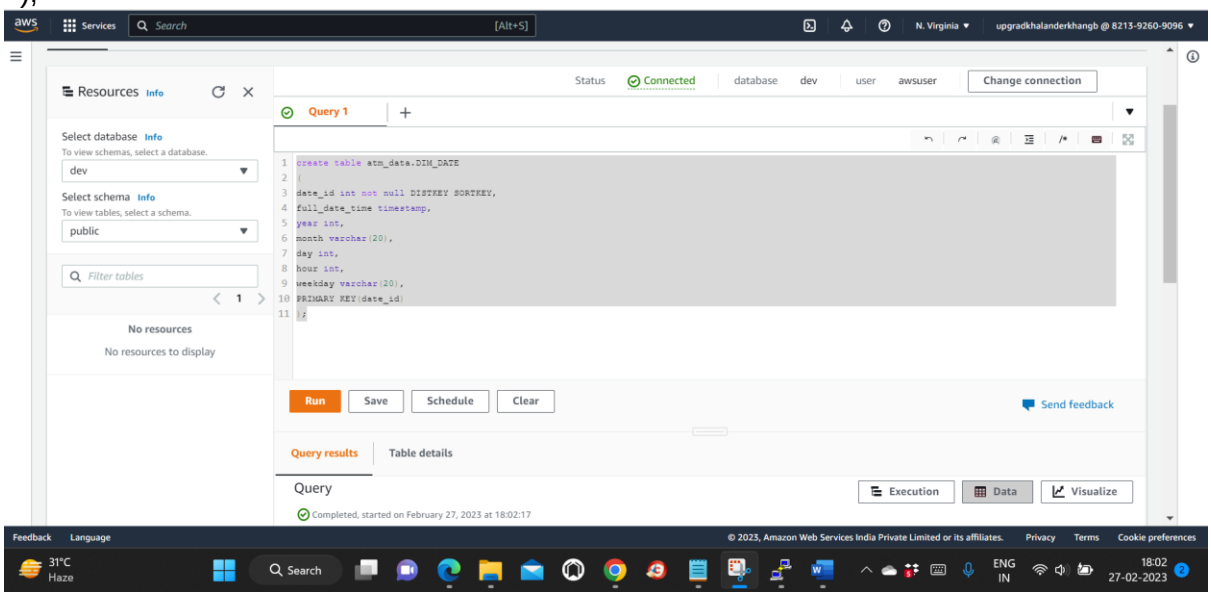
Query to create the dimension table DIM_ATM:

```
CREATE TABLE atm_nord.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_nord.DIM_LOCATION(location_id)
);
```



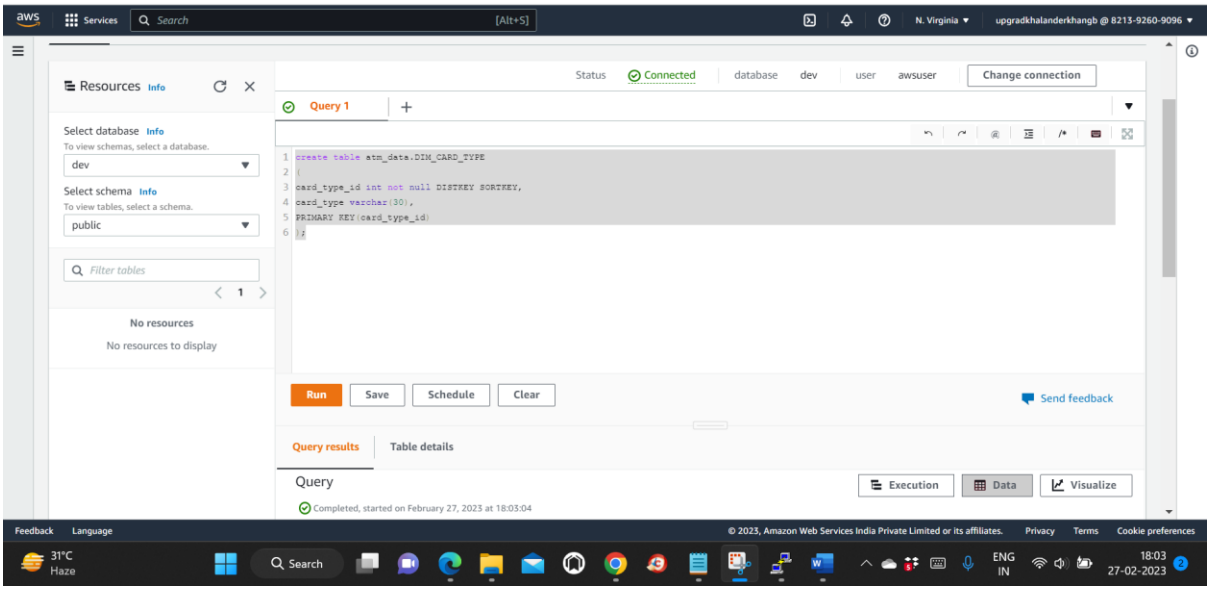
Query to create the dimension table DIM_DATE:

```
CREATE TABLE atm_nord.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)
);
```



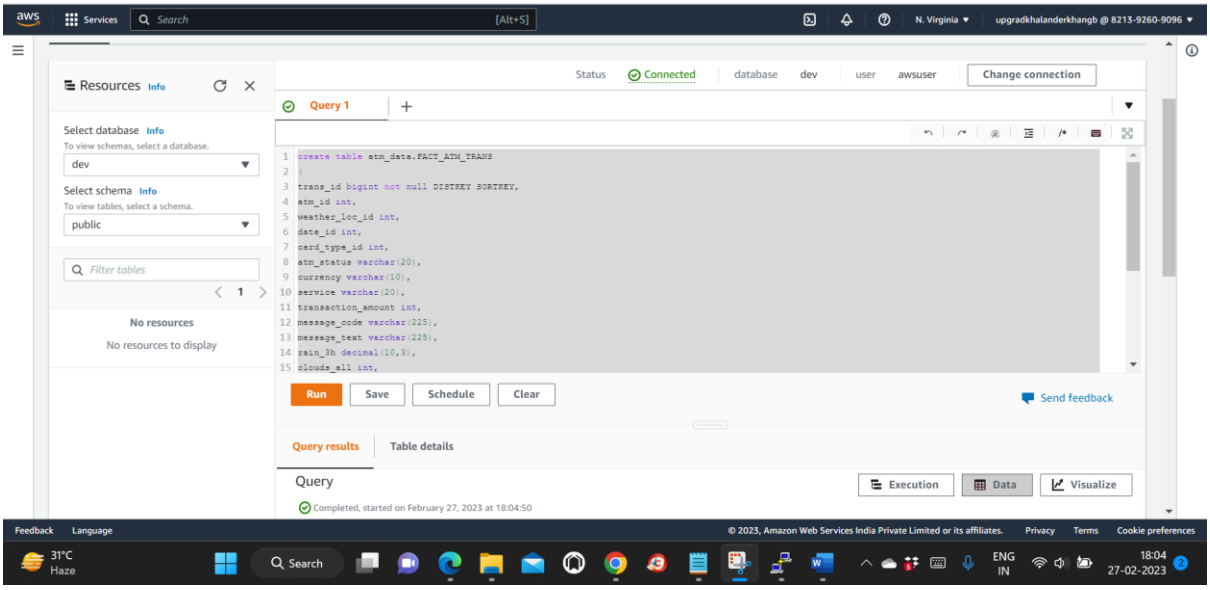
Query to create the dimension table DIM_CARD_TYPE:

```
CREATE TABLE atm_nord.DIM_CARD_TYPE
(
card_type_id INT NOT NULL DISTKEY SORTKEY,
card_type VARCHAR(30),
PRIMARY KEY(card_type_id)
);
```



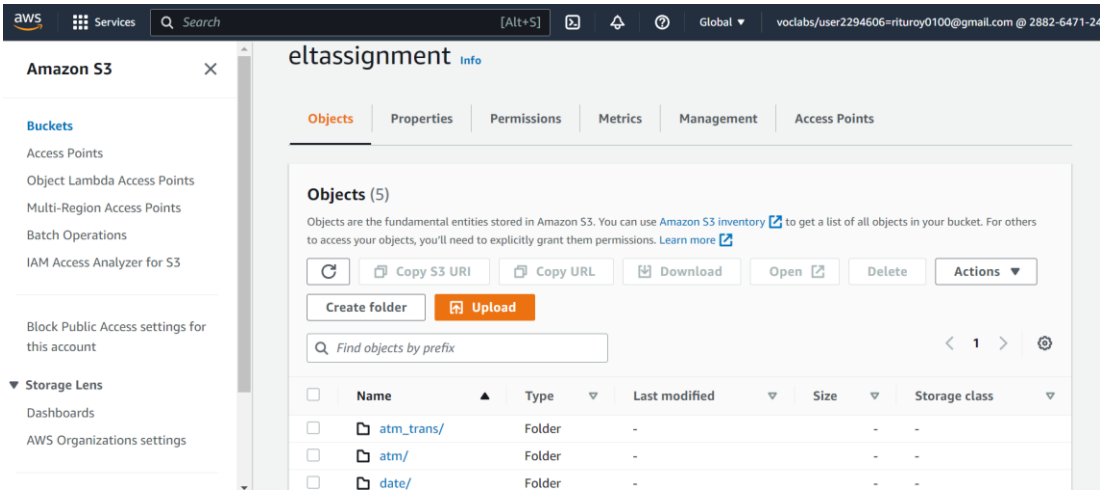
Query to create the fact table FACT_ATM_TRANS:

```
CREATE TABLE atm_nord.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(255),
message_text VARCHAR(255),
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_nord.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references atm_nord.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references atm_nord.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references atm_nord.DIM_CARD_TYPE(card_type_id)
);
```



Loading data into a Redshift cluster from Amazon S3 bucket

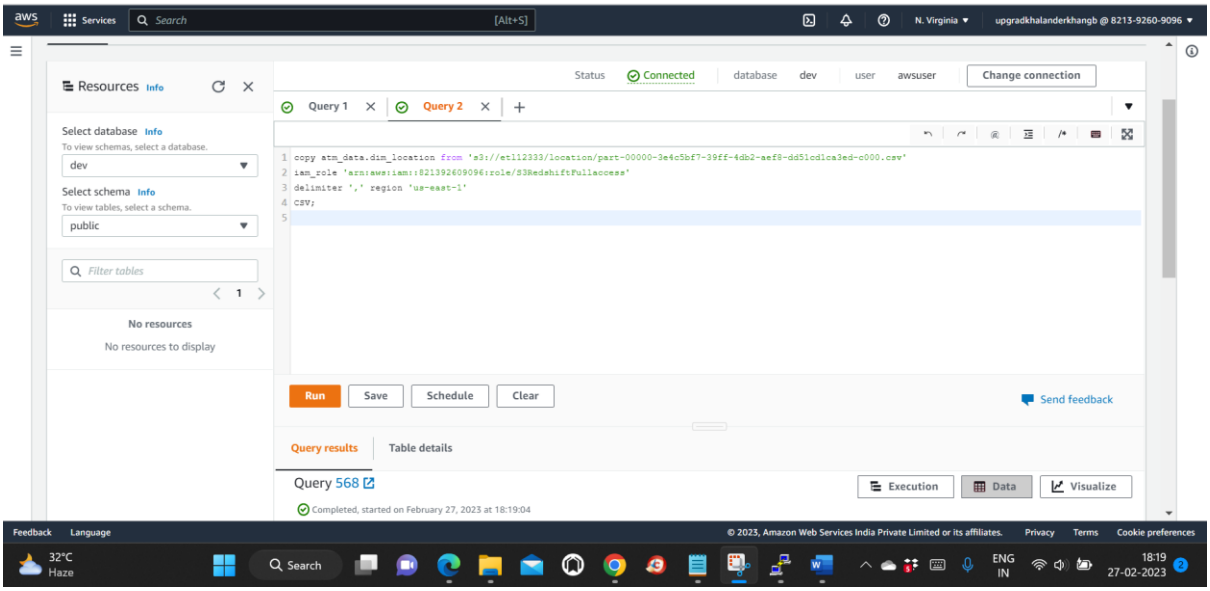
The different folders created by us for writing the data in csv format for dimension and fact table



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

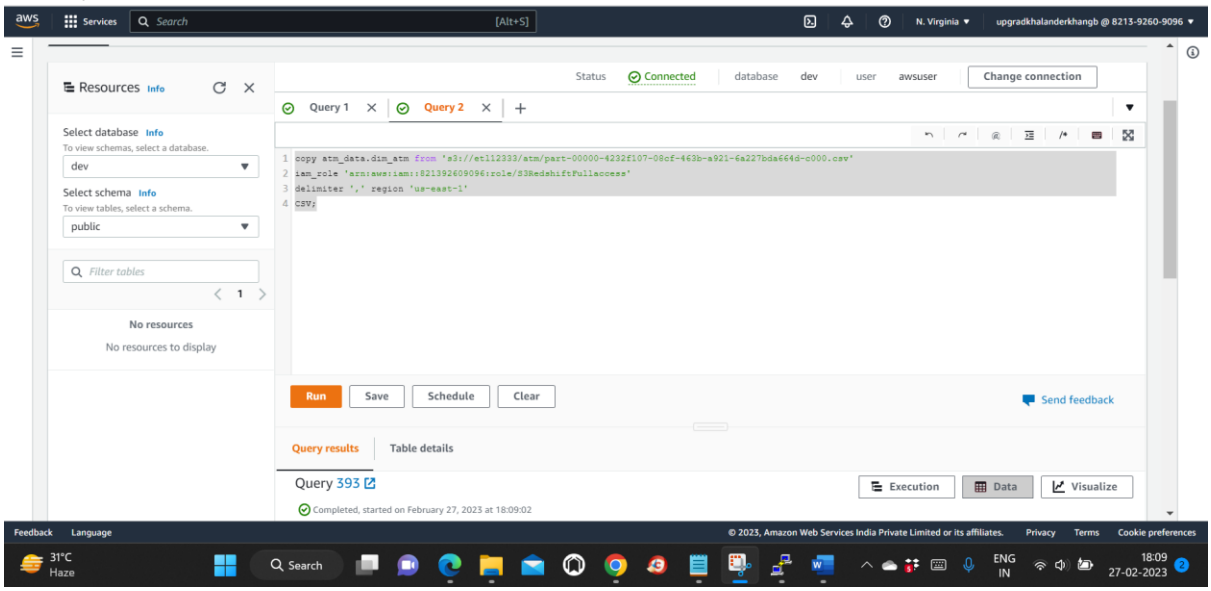
Query to copy the data from S3 bucket – dev to dim_location table

```
copy atm_data.dim_location from 's3://etl12333/location/part-00000-3e4c5bf7-39ff-4db2-aef8-dd51cd1ca3ed-c000.csv'
iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



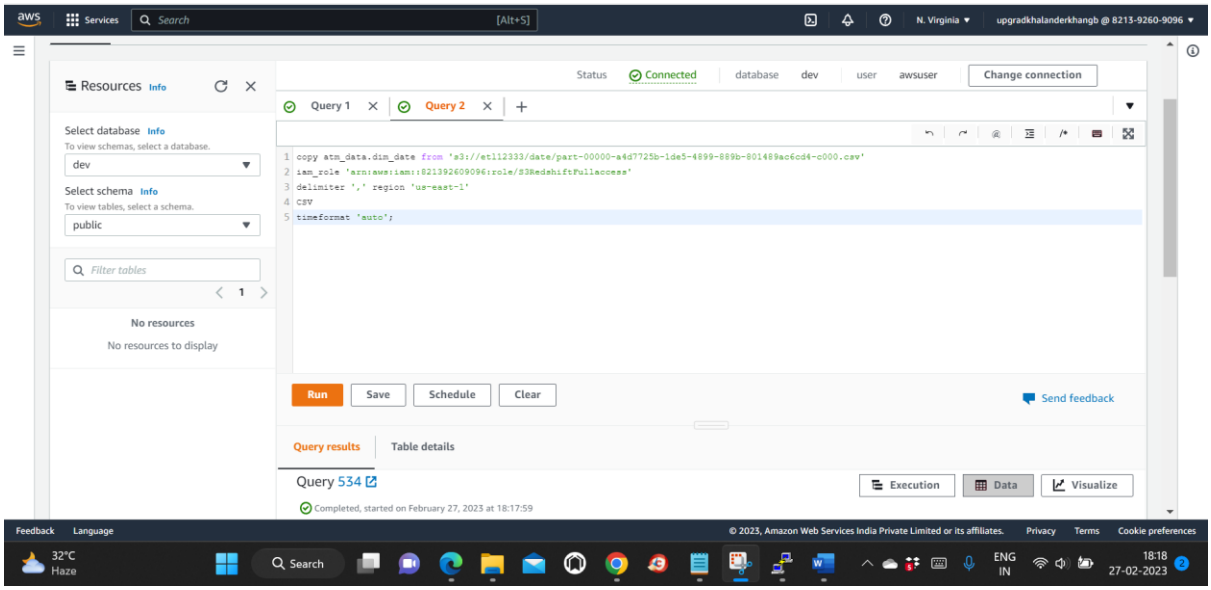
Query to copy the data from S3 bucket – dev to dim_atm table

```
copy atm_data.dim_atm from 's3://etl12333/atm/part-00000-4232f107-08cf-463b-a921-6a227bda664d-c000.csv'
iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



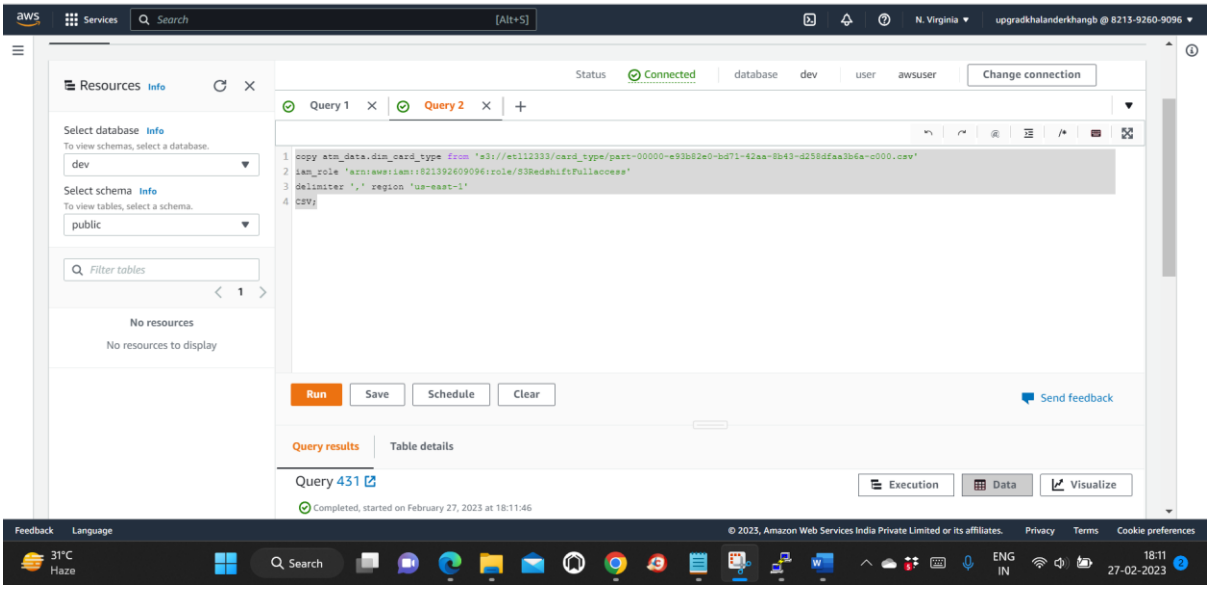
Query to copy the data from S3 bucket – dev to dim_date table

```
copy atm_data.dim_date from
's3://etl12333/date/part-00000-
a4d7725b-1de5-4899-889b-
801489ac6cd4-c000.csv'
iam_role
'arn:aws:iam::821392609096:role/S3
RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV
timeformat 'auto';
```



Query to copy the data from S3 bucket – dev to dim_card_type table

```
copy atm_data.dim_card_type from 's3://etl12333/card_type/part-00000-e93b82e0-bd71-42aa-
8b43-d258dfaa3b6a-c000.csv'
iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



Query to copy the data from S3 bucket – dev to fact_atm_trans table

copy atm_data.fact_atm_trans from 's3://etl12333/fact_table/part-00000-207b7c36-f333-4922-9107-27bffa6b1ac2-c000.csv'
iam_role 'arn:aws:iam::821392609096:role/S3RedshiftFullaccess'
delimiter ',' region 'us-east-1'
CSV;