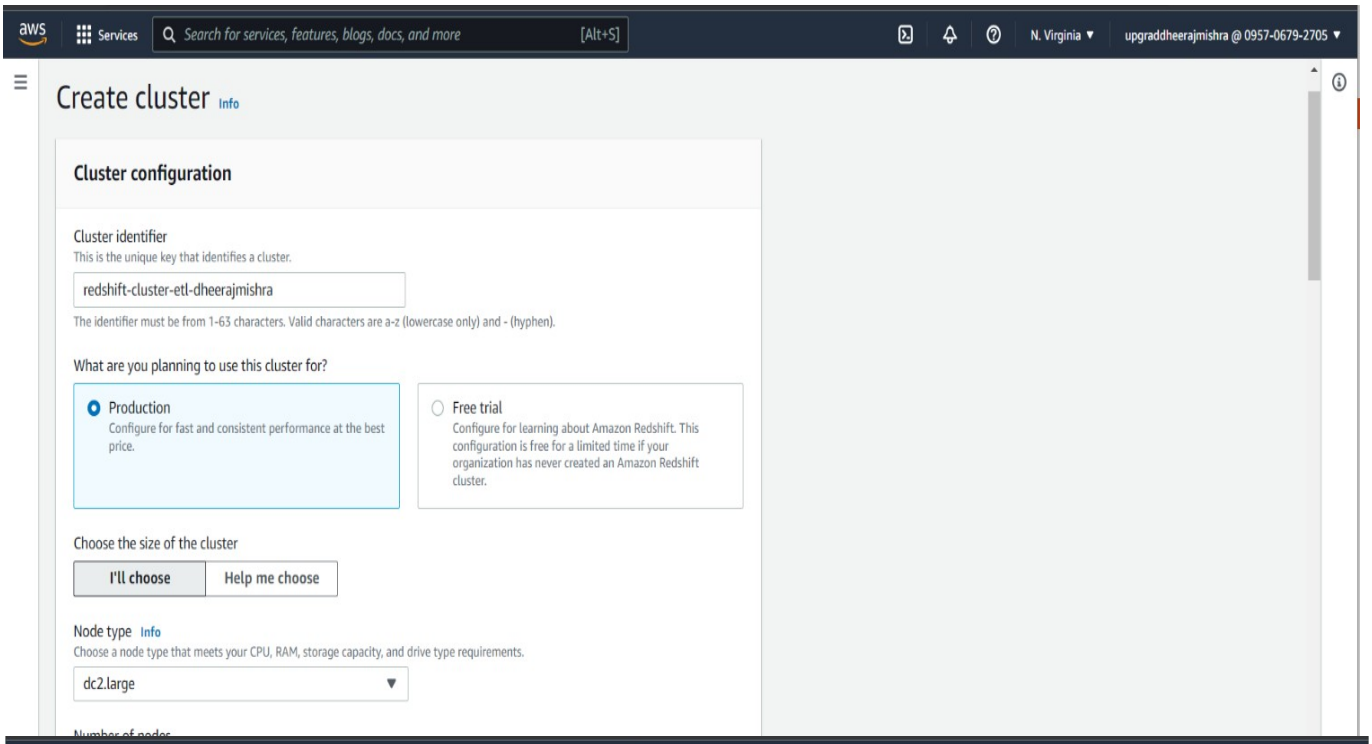
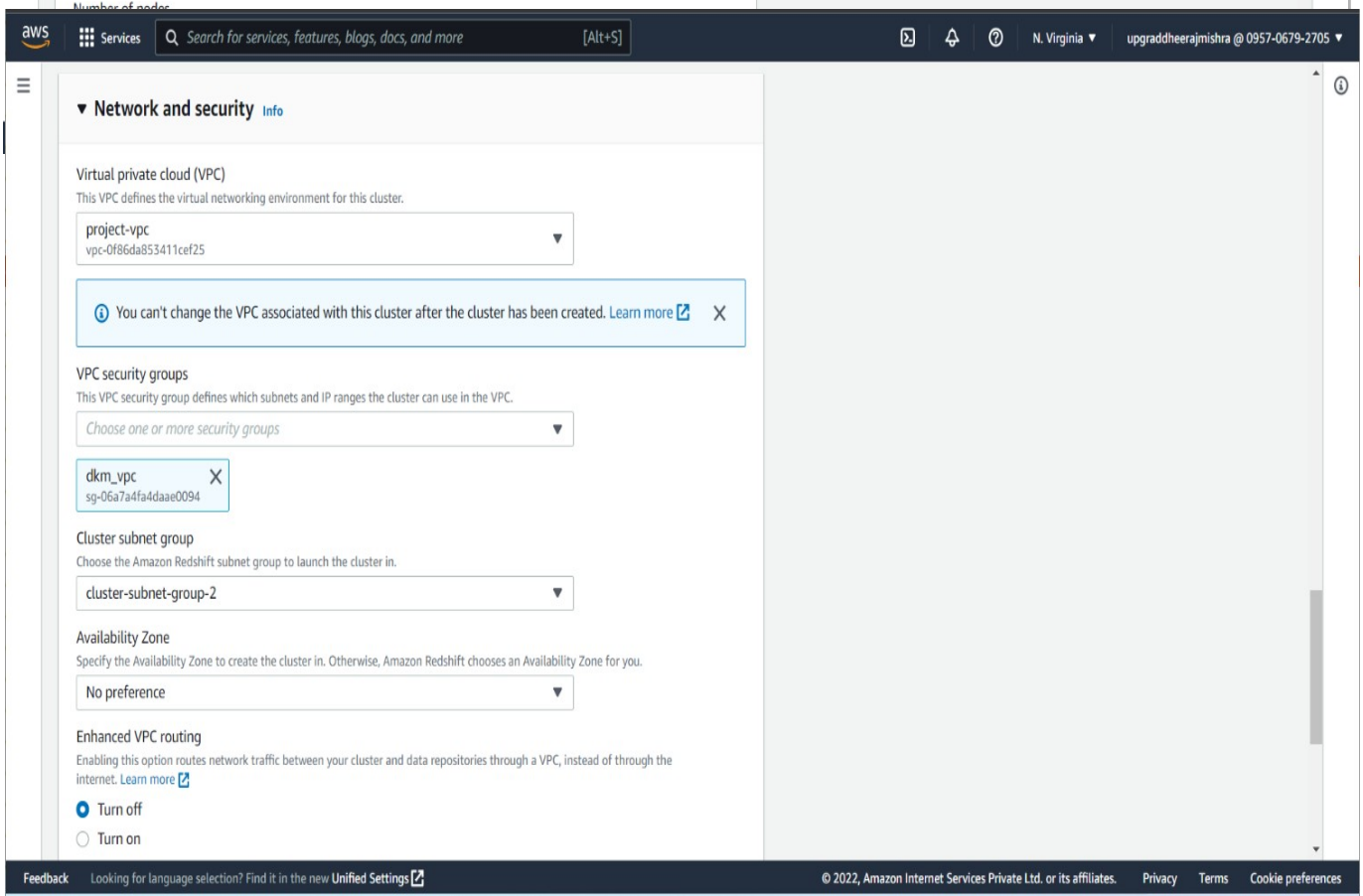
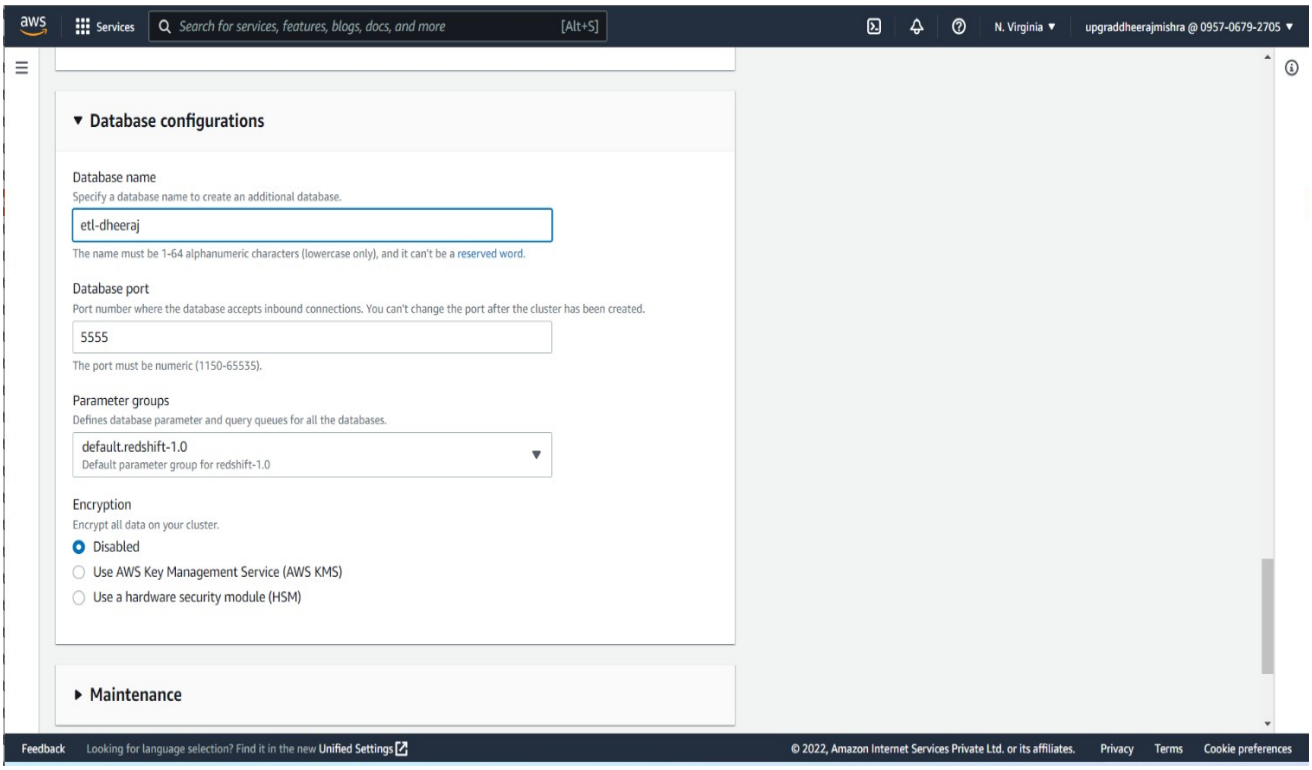
Creation of a Redshift Cluster

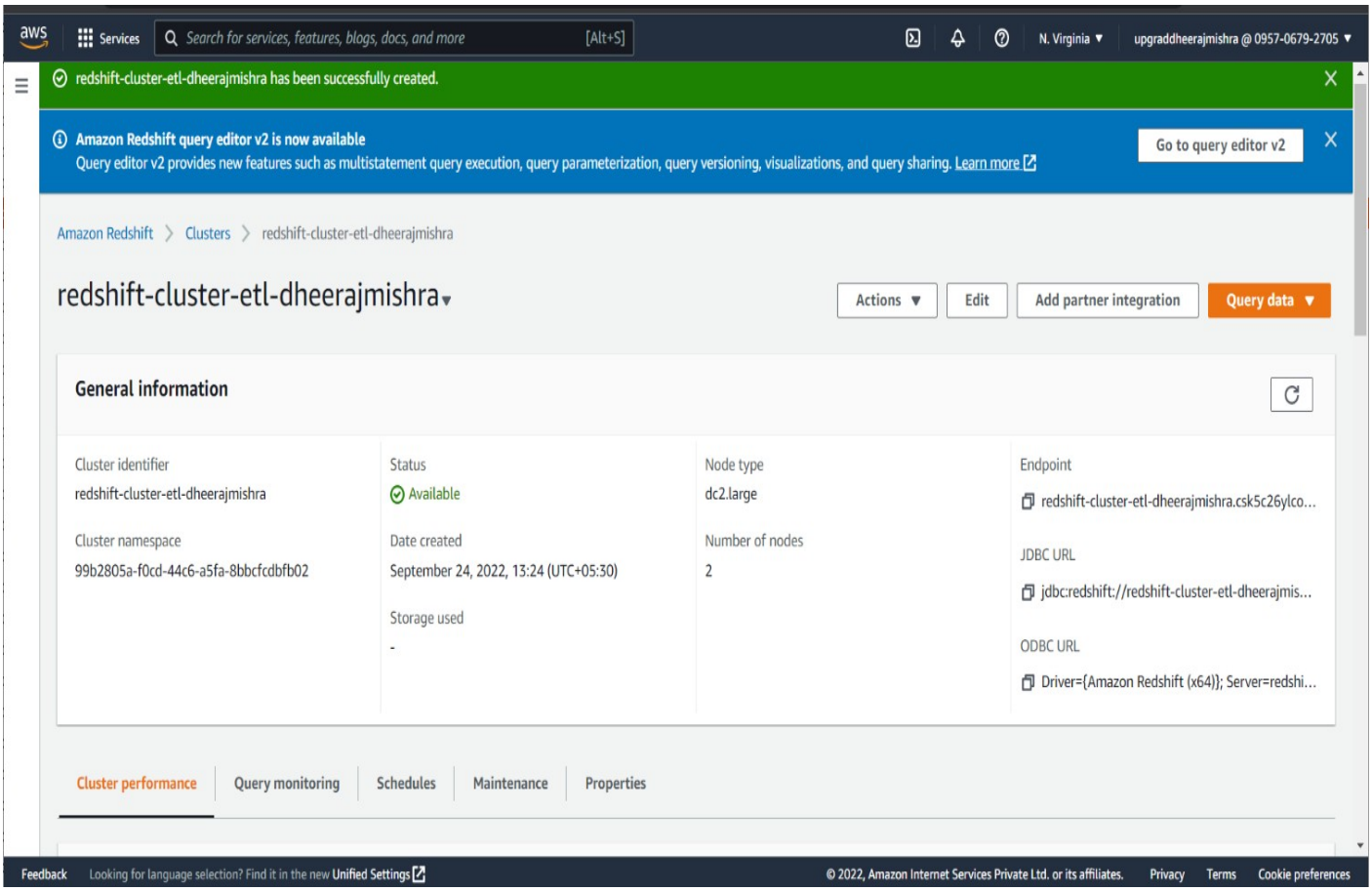
Screenshots of the configuration of the Redshift cluster that I have created:

1.Screenshot of the type of machine used along with number of nodes

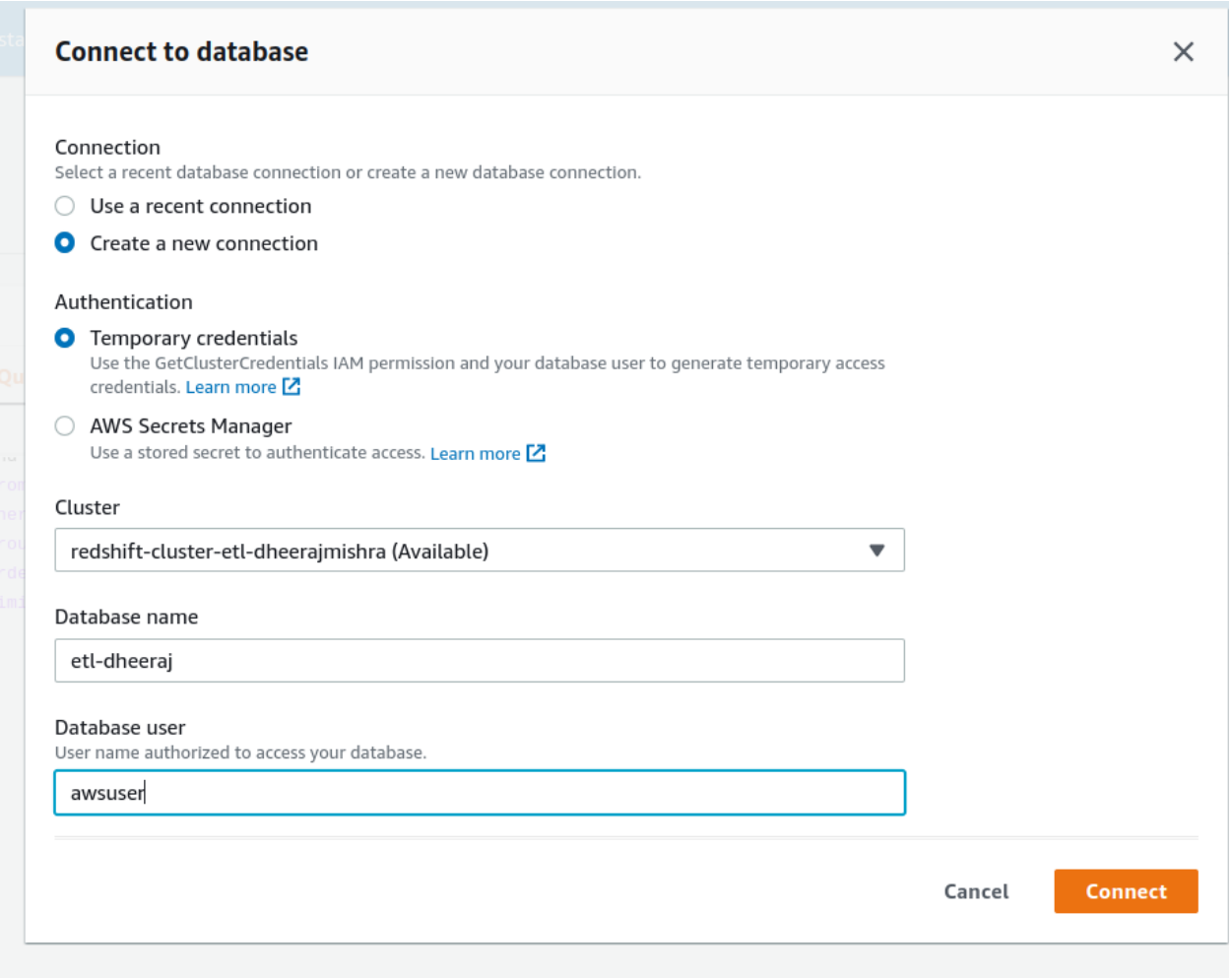




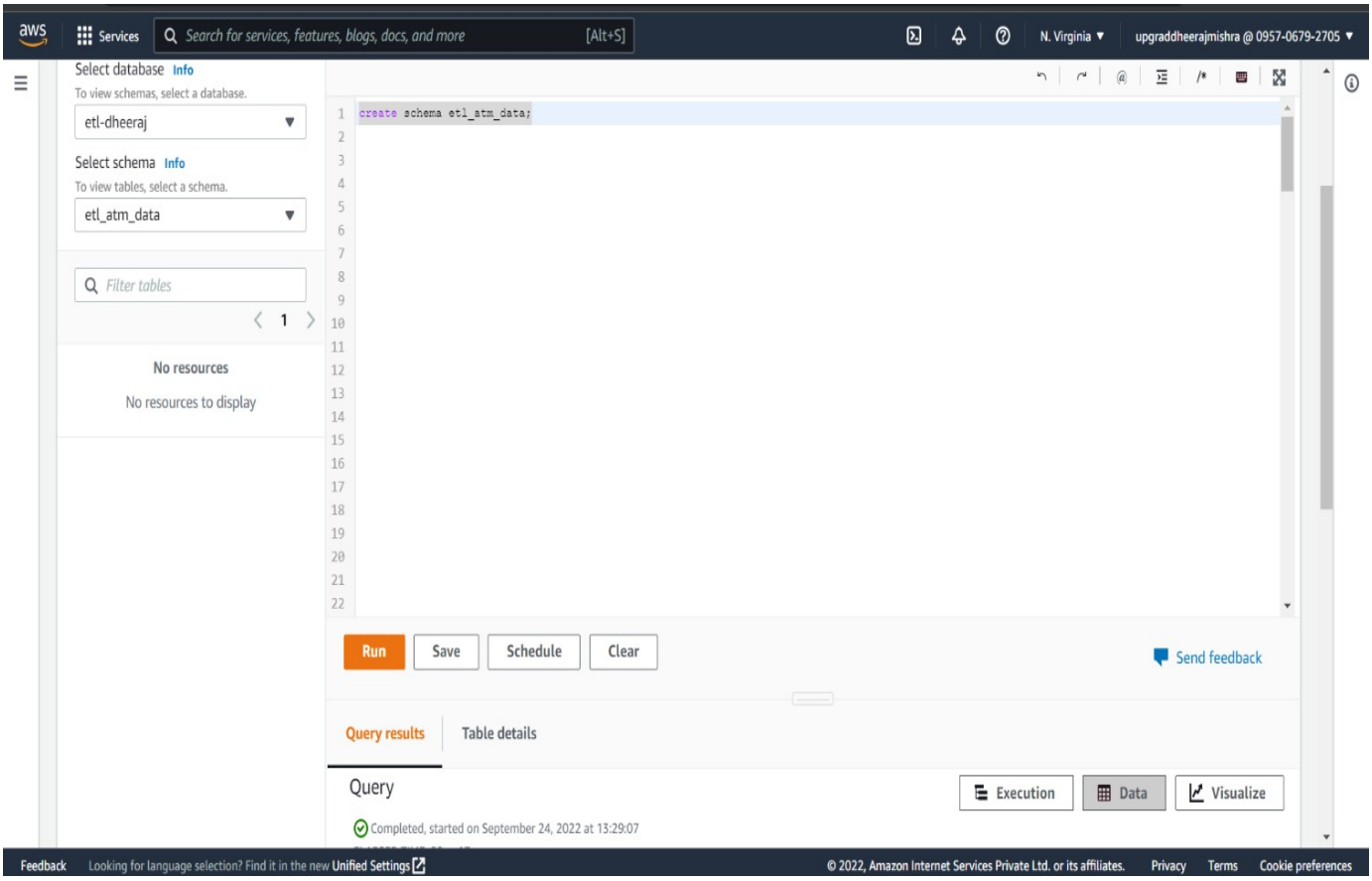




Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables



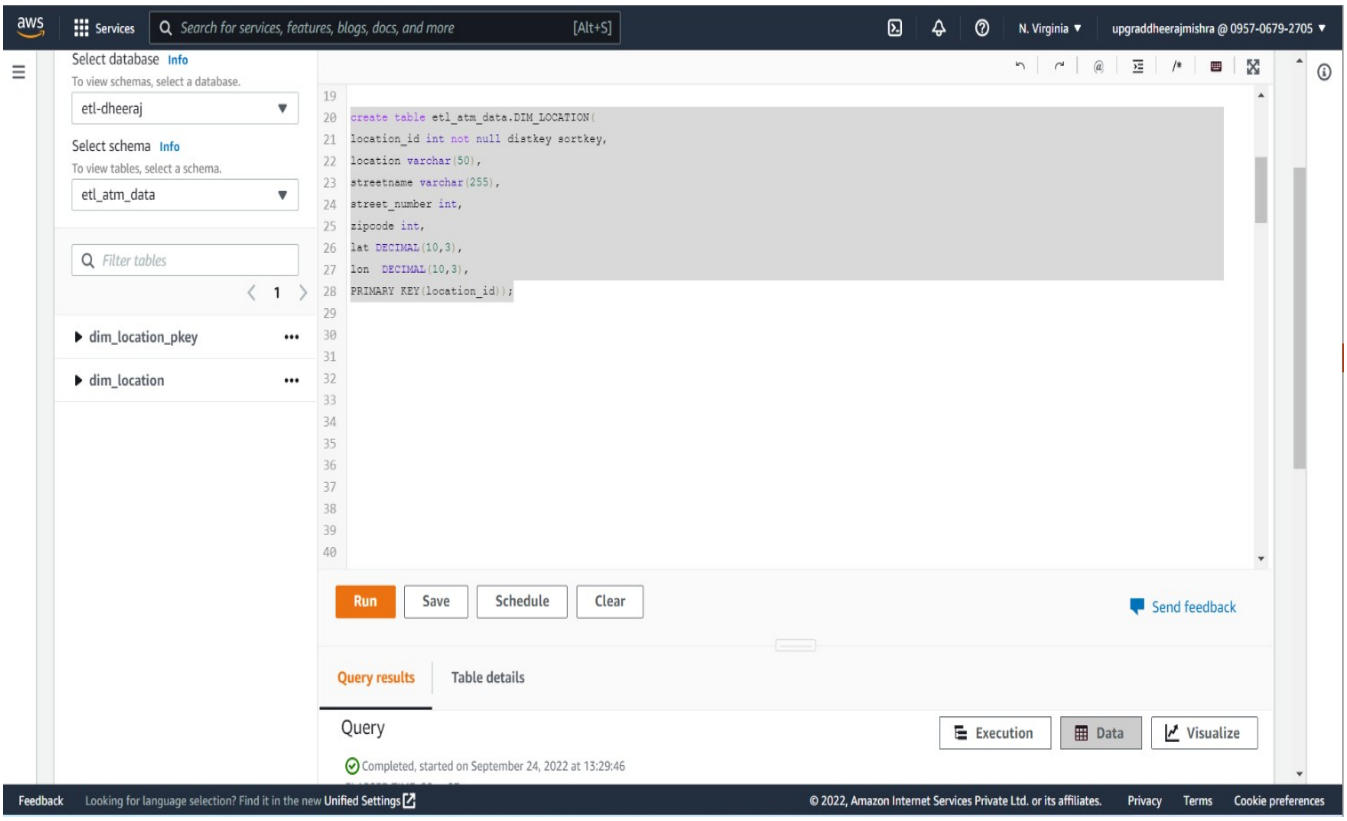
create schema etl_atm_data;



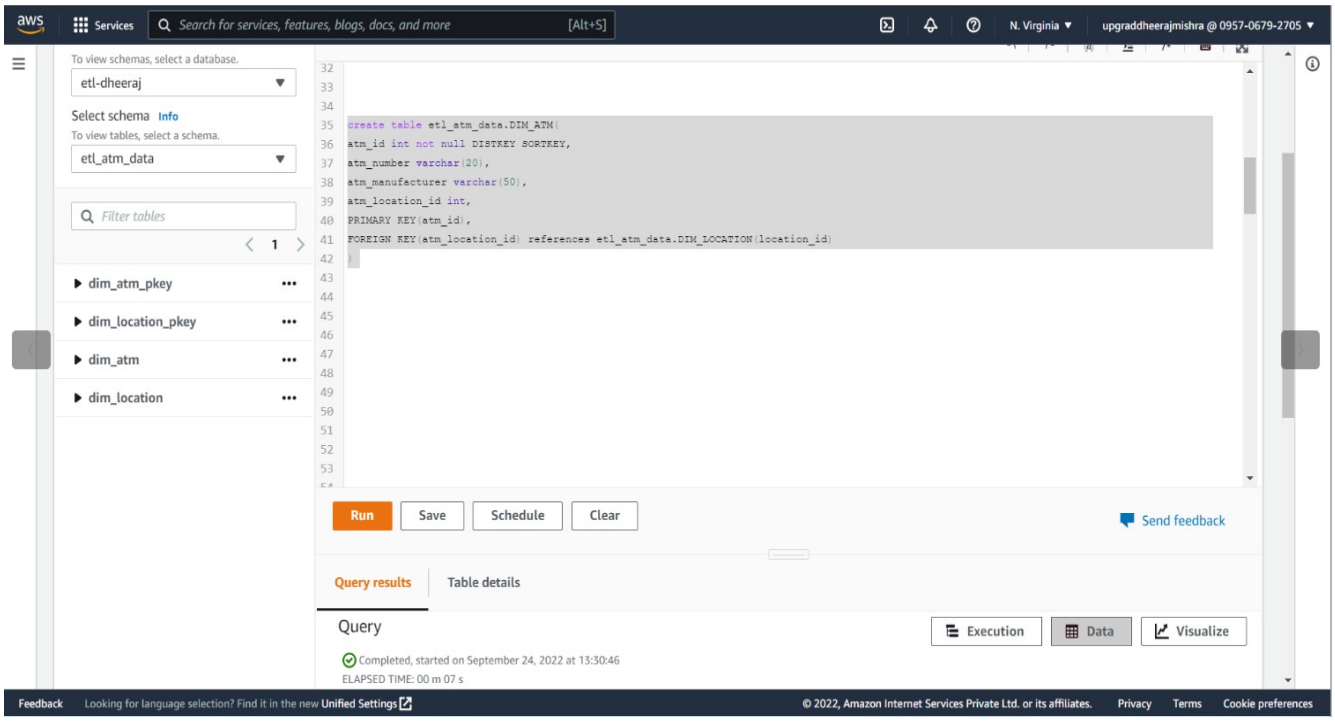
Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

1.
create table etl_atm_data.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));



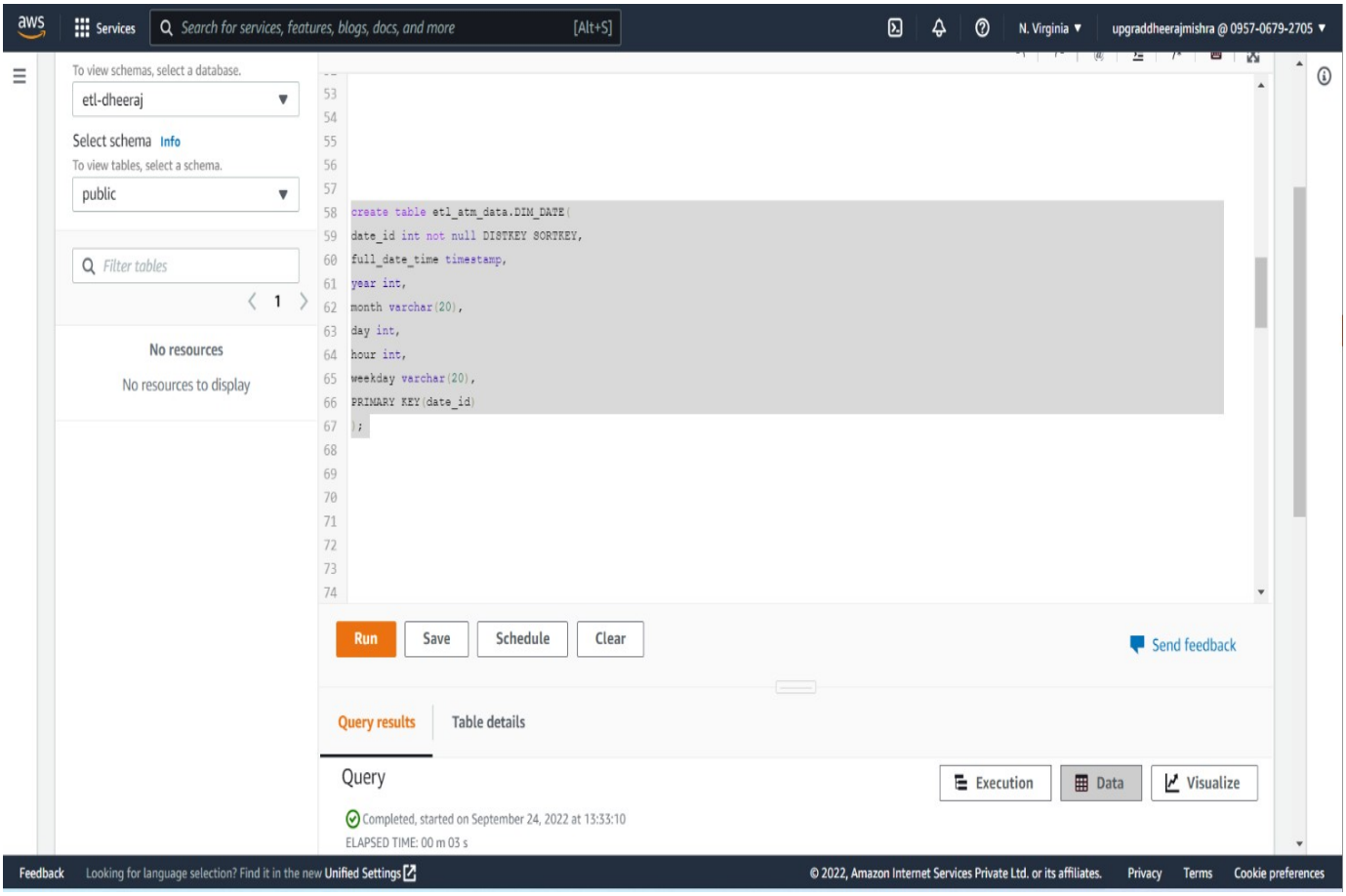
2.
create table etl_atm_data.DIM_ATM(
atm_id int not null DISTKEY SORTKEY,
atm_number varchar(20),
atm_manufacturer varchar(50),
atm_location_id int,
PRIMARY KEY(atm_id),
FOREIGN KEY(atm_location_id) references etl_atm_data.DIM_LOCATION(location_id)
) ;



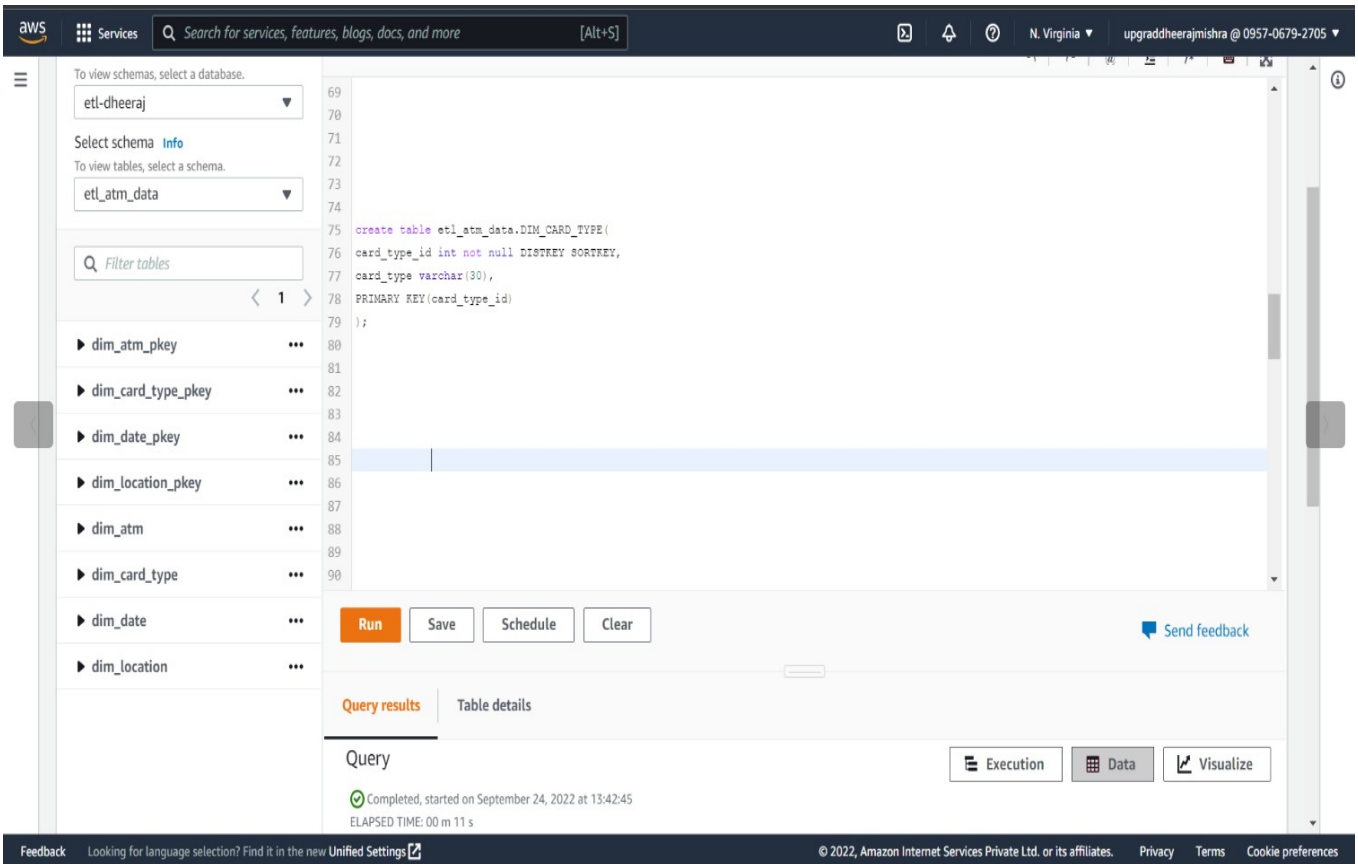
3.

```
create table etl_atm_data.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)
);



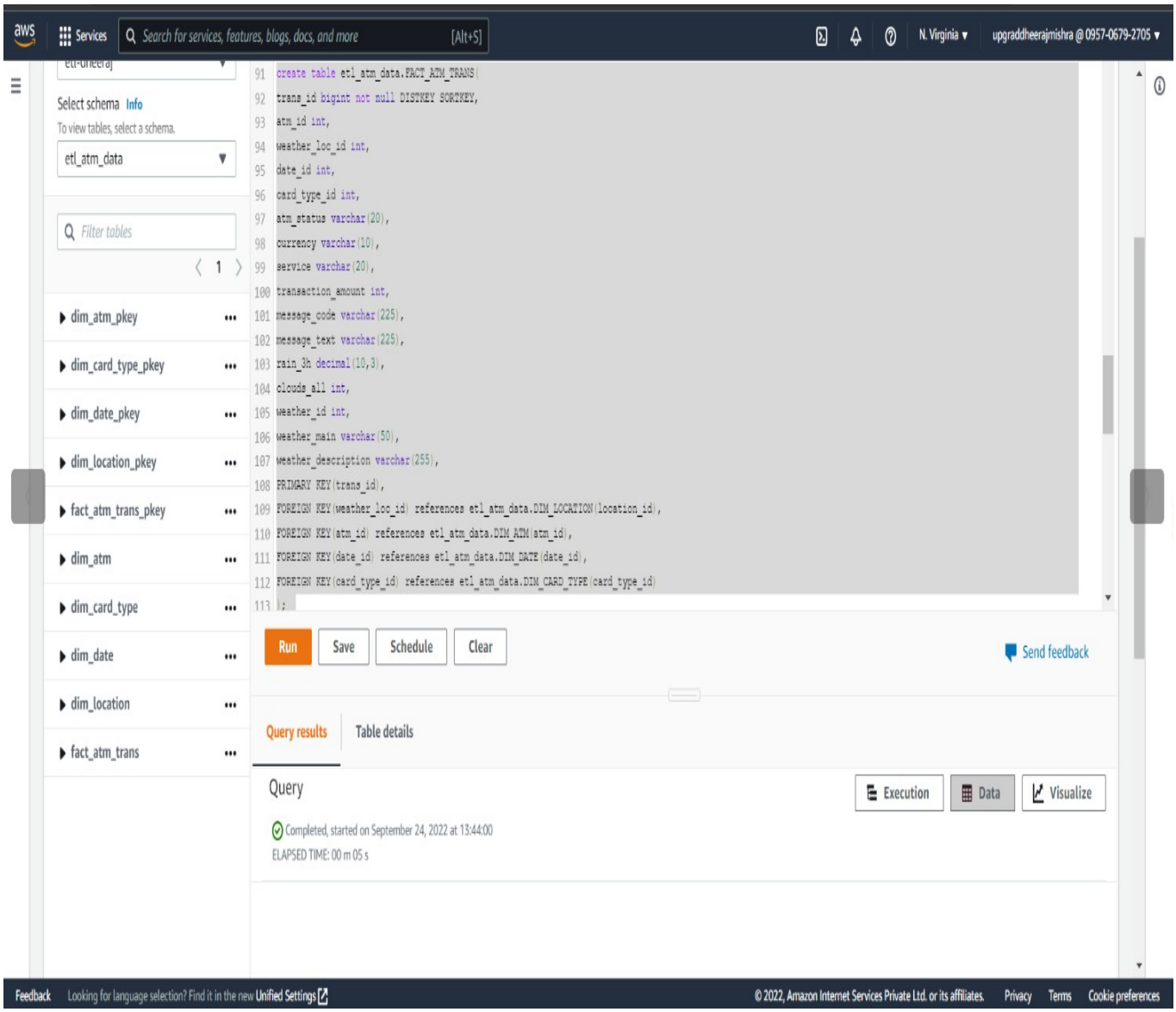
4.
create table etl_atm_data.DIM_CARD_TYPE(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);



5.

```
create table etl_atm_data.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 etl_atm_data.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references etl_atm_data.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references etl_atm_data.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references etl_atm_data.DIM_CARD_TYPE(card_type_id)
);
```

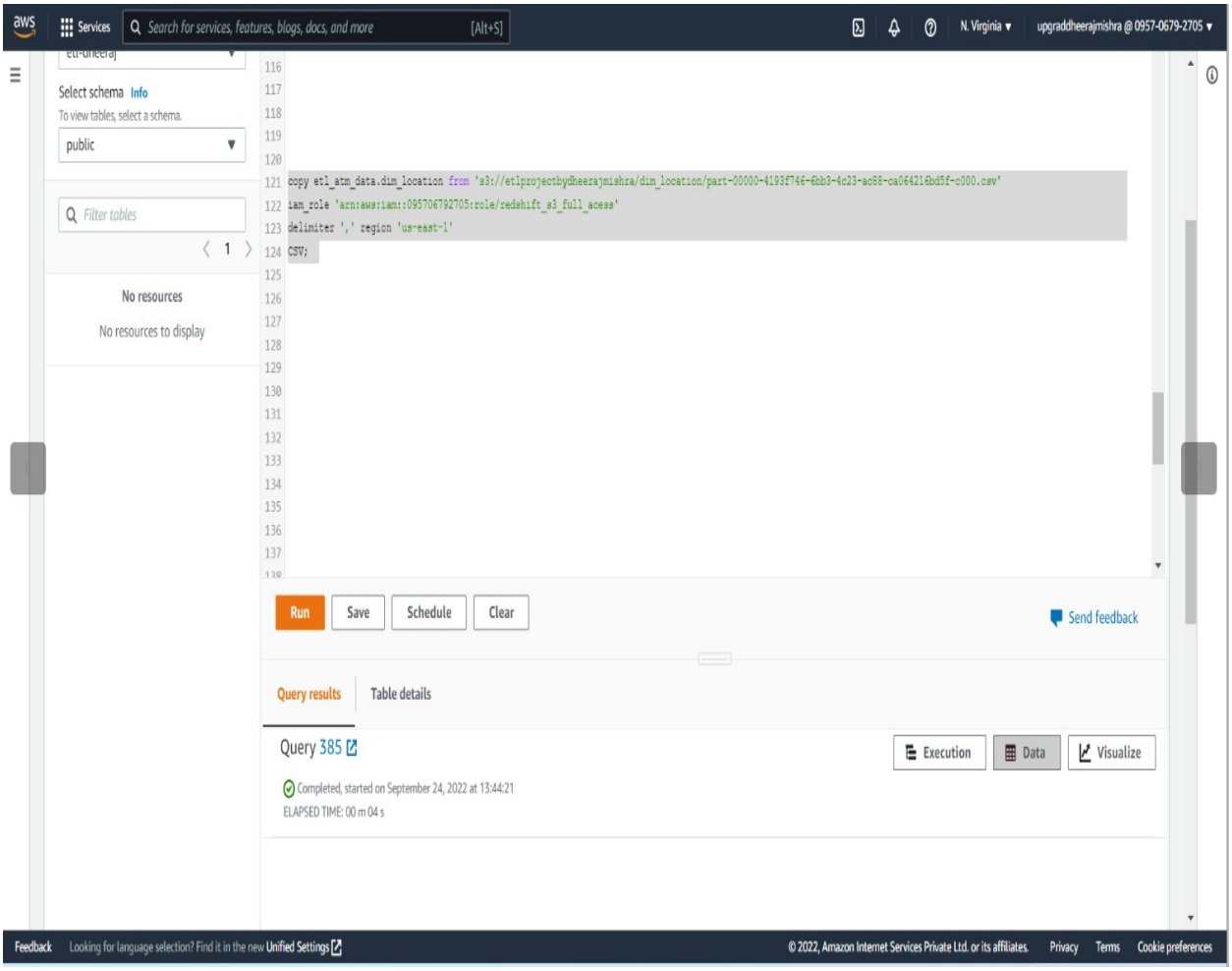


Loading data into a Redshift cluster from Amazon S3 bucket

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

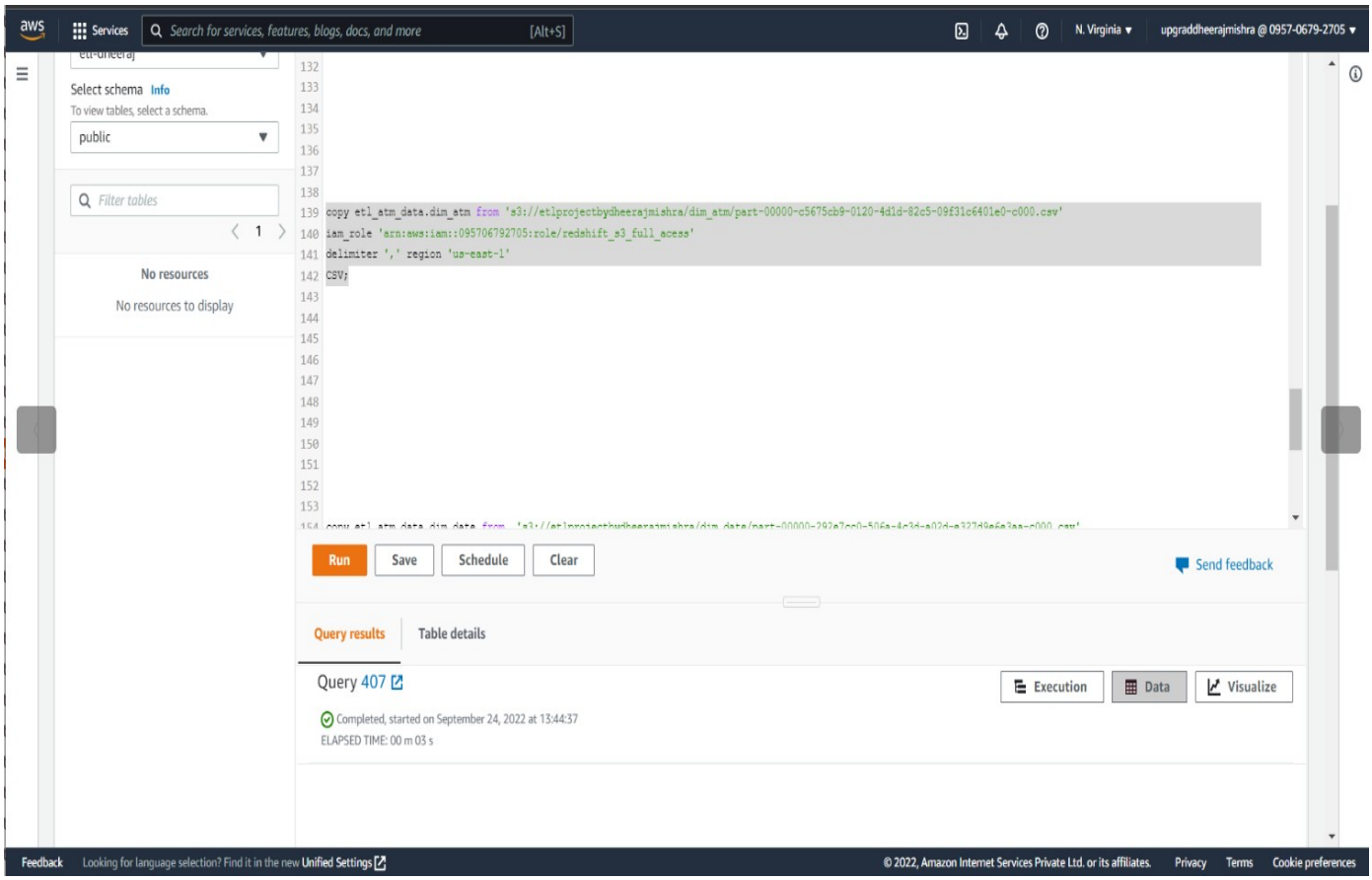
1.

```
copy etl_atm_data.dim_location from 's3://etlprojectbydheerajmishra/dim_location/part-
00000-dab6b198-a75d-4dbd-8629-671d79c580ef-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_acess'
delimiter ',' region 'us-east-1'
CSV;
```



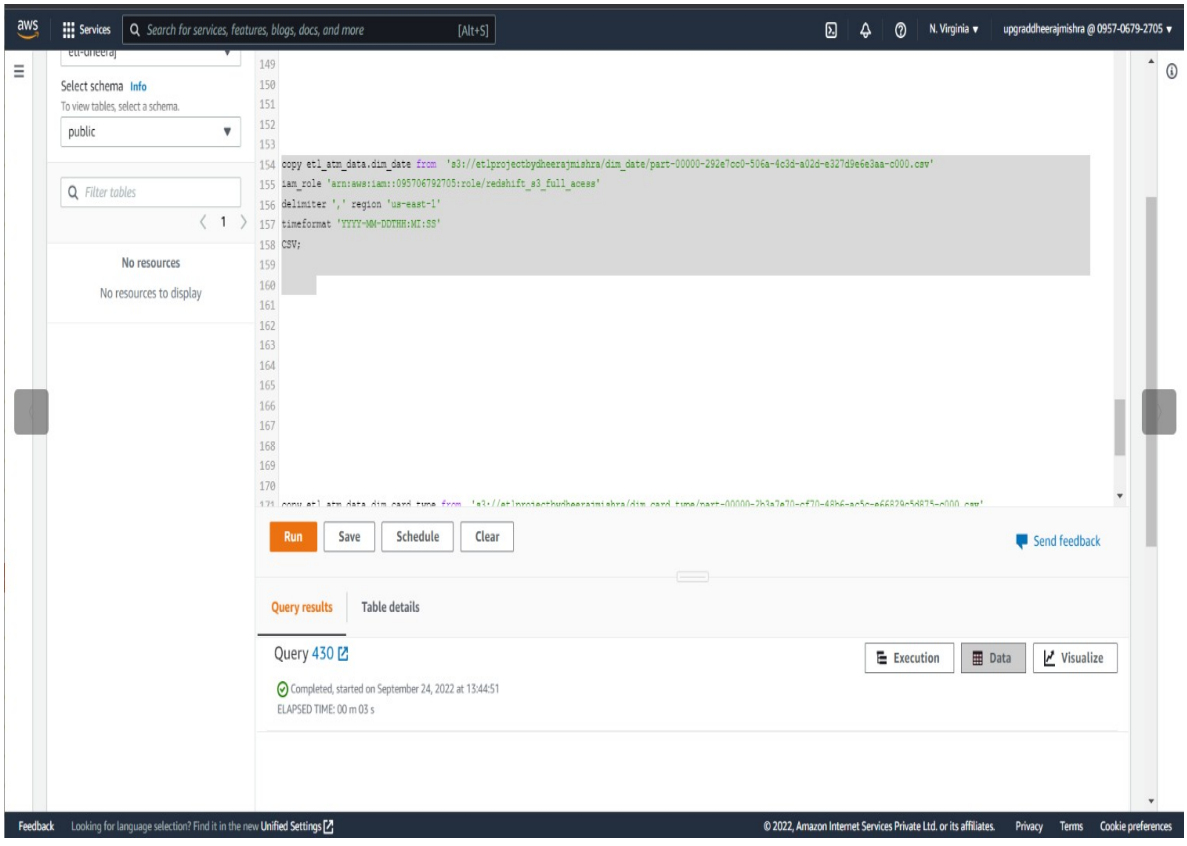
2.

```
copy etl_atm_data.dim_atm from 's3://etlprojectbydheerajmishra/dim_atm/part-00000-
c9dc5033-485a-462c-a434-27aa8851297b-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_acess'
delimiter ',' region 'us-east-1'
CSV;
```



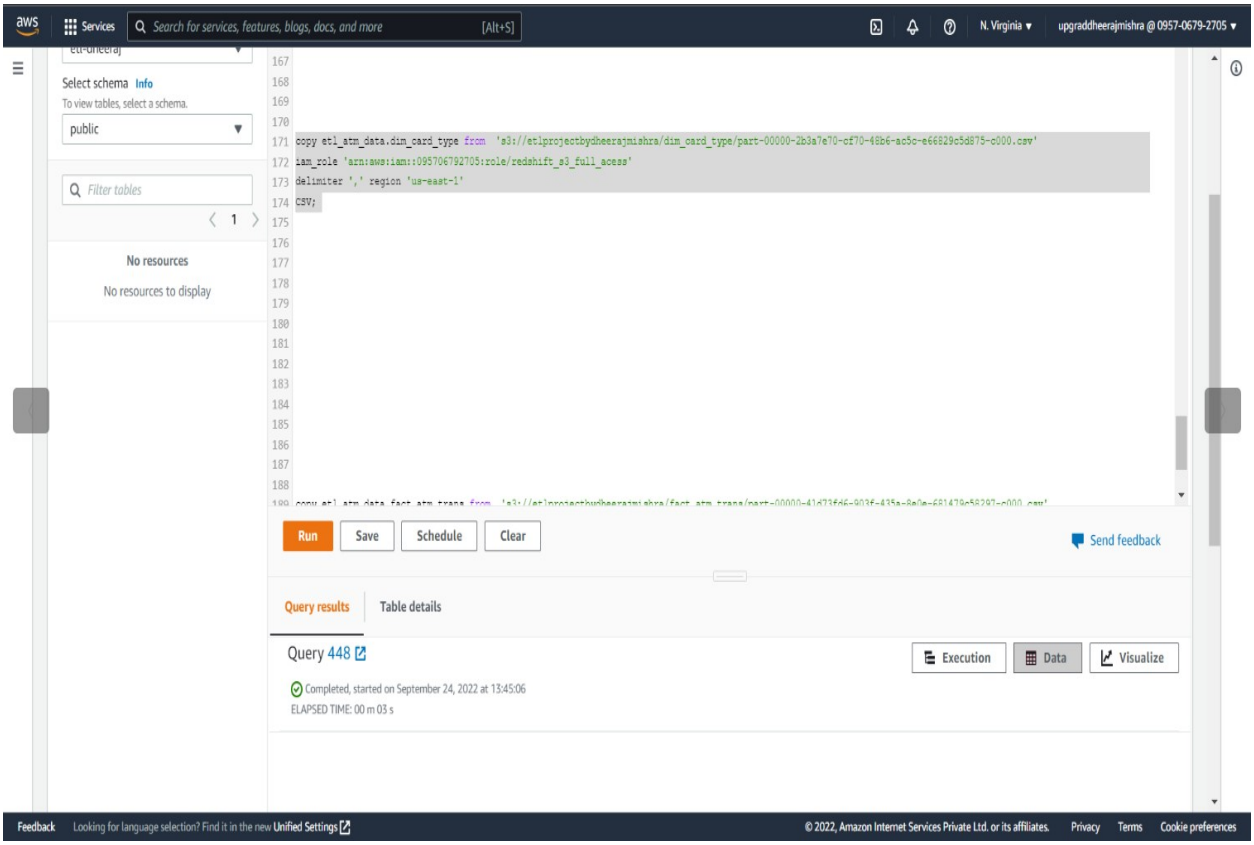
3.

```
copy etl_atm_data.dim_date from  's3://etlprojectbydheerajmishra/dim_date/part-00000-02bf4ea0-0dad-44ce-b28a-1d76ef2b8aa6-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_acess'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;
```



4.

```
copy etl_atm_data.dim_card_type from
's3://etlprojectbydheerajmishra/dim_card_type/part-00000-438a36c1-9e05-4faa-9839-
6625e2e0be3a-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_acess'
delimiter ',' region 'us-east-1'
CSV;
```



5.

```
copy etl_atm_data.fact_atm_trans from
's3://etlprojectbydheerajmishra/fact_atm_trans/part-00000-5cbe2097-b629-4447-8fee-0f427ec95a5f-c000.csv'
iam_role 'arn:aws:iam::095706792705:role/redshift_s3_full_acess'
delimiter ',' region 'us-east-1'
CSV;
```