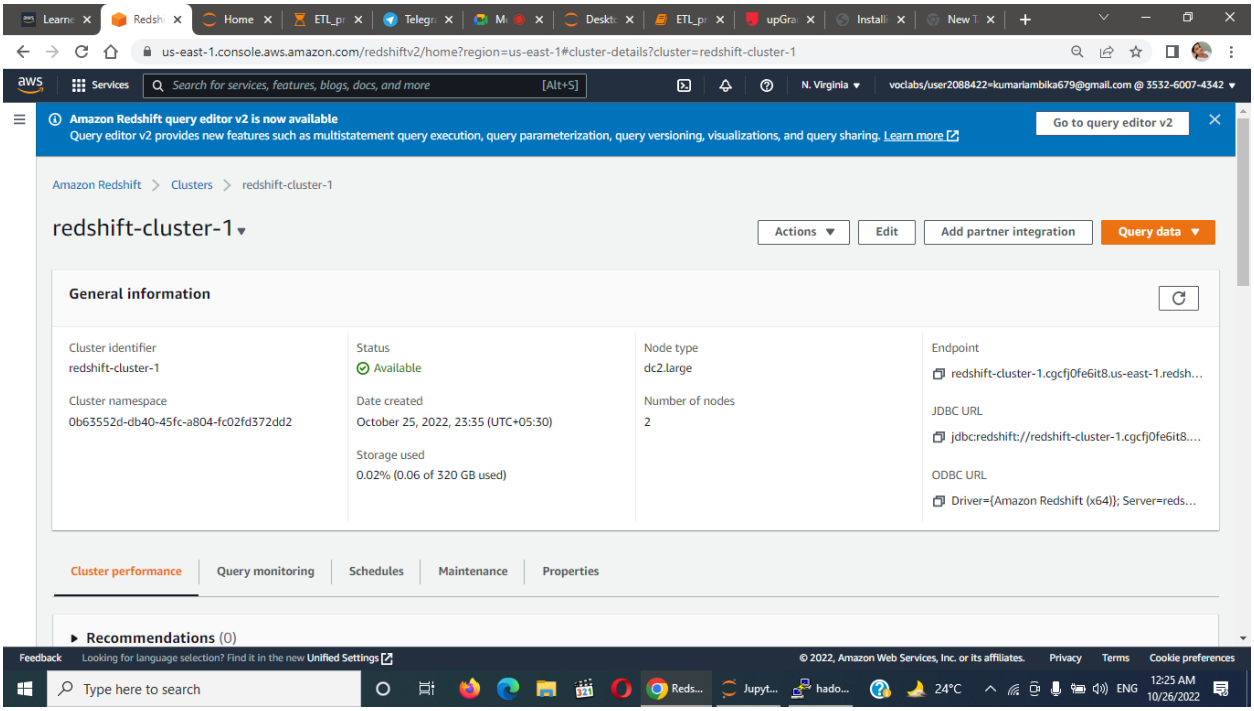
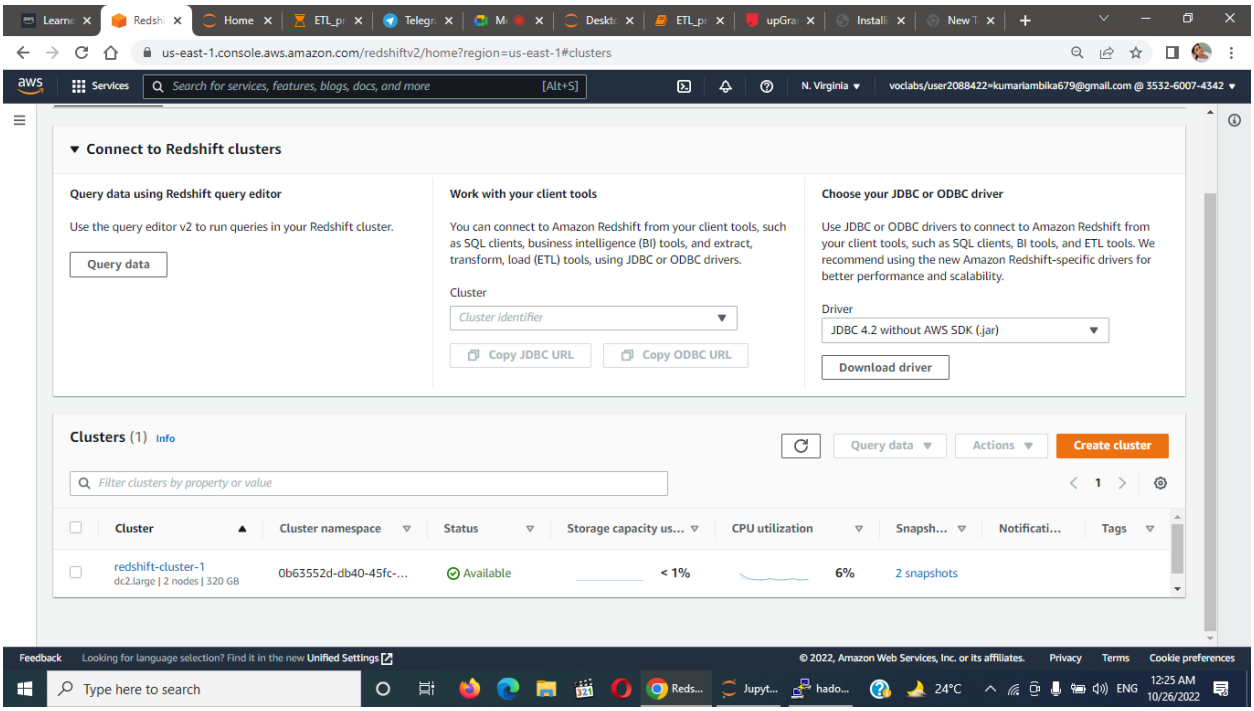
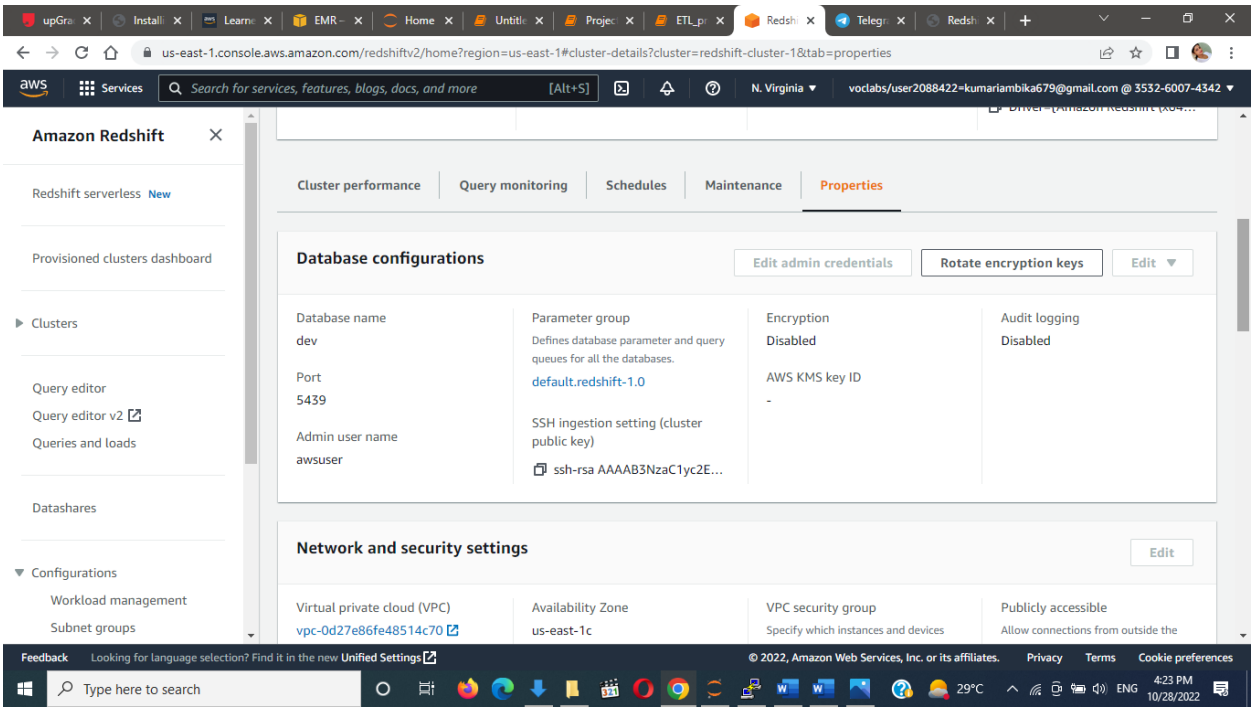
Creation of a Redshift Cluster

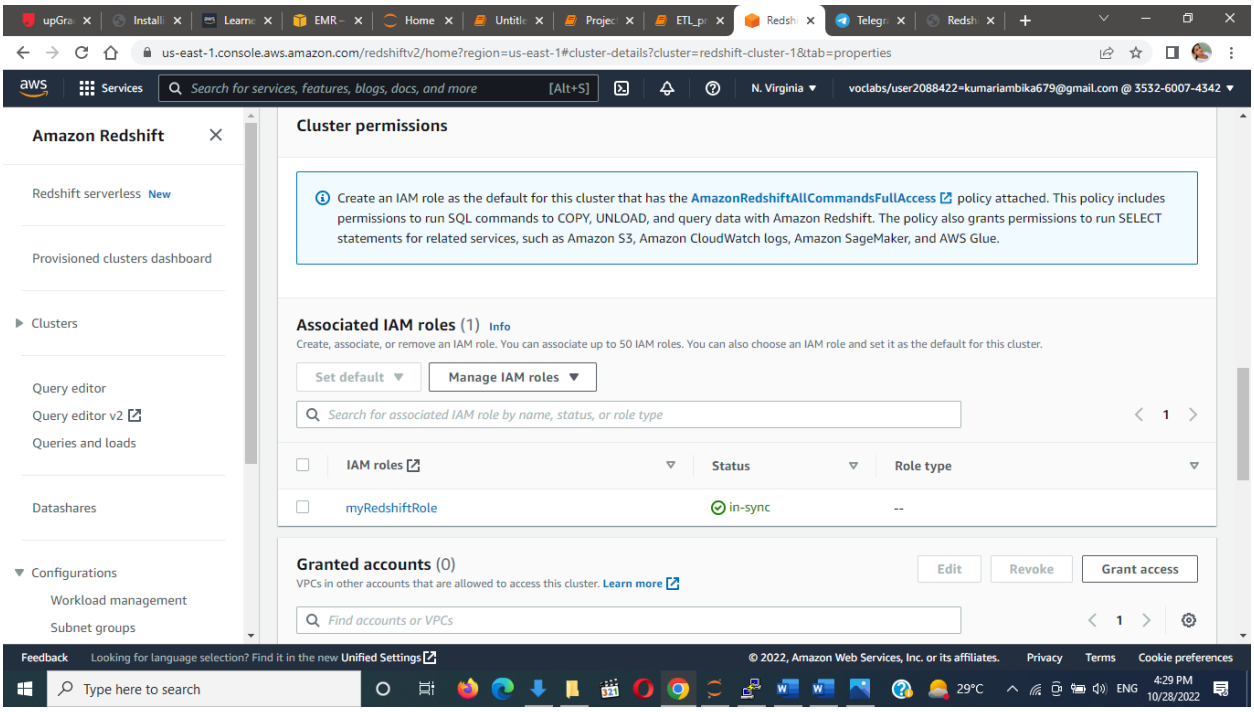
Screenshots of the configuration of the Redshift cluster that you have created:

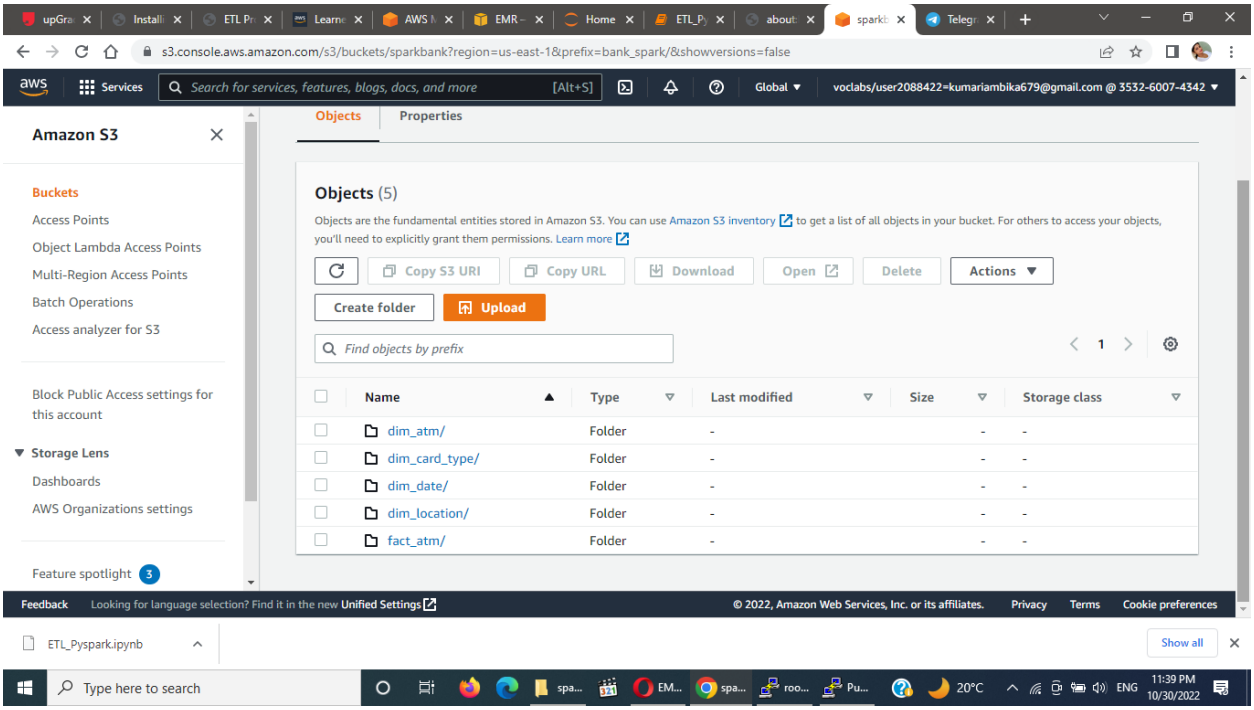
<Screenshot of the type of machine used along with number of nodes>





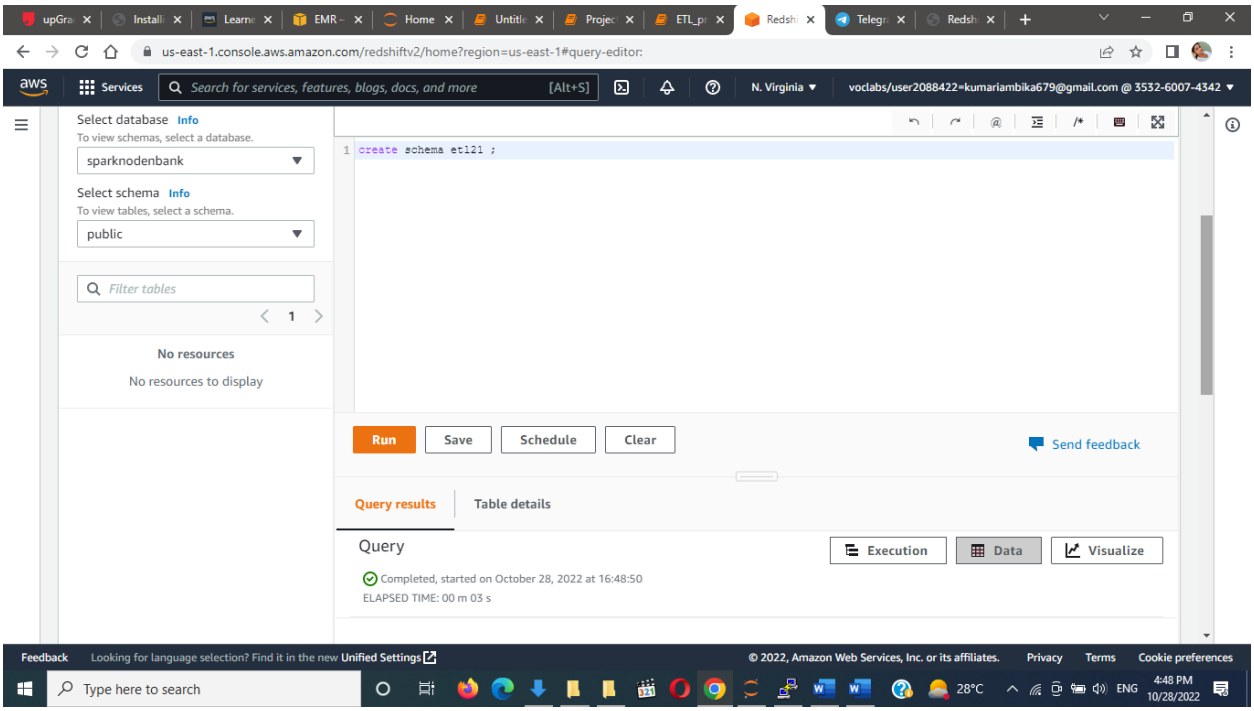






Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

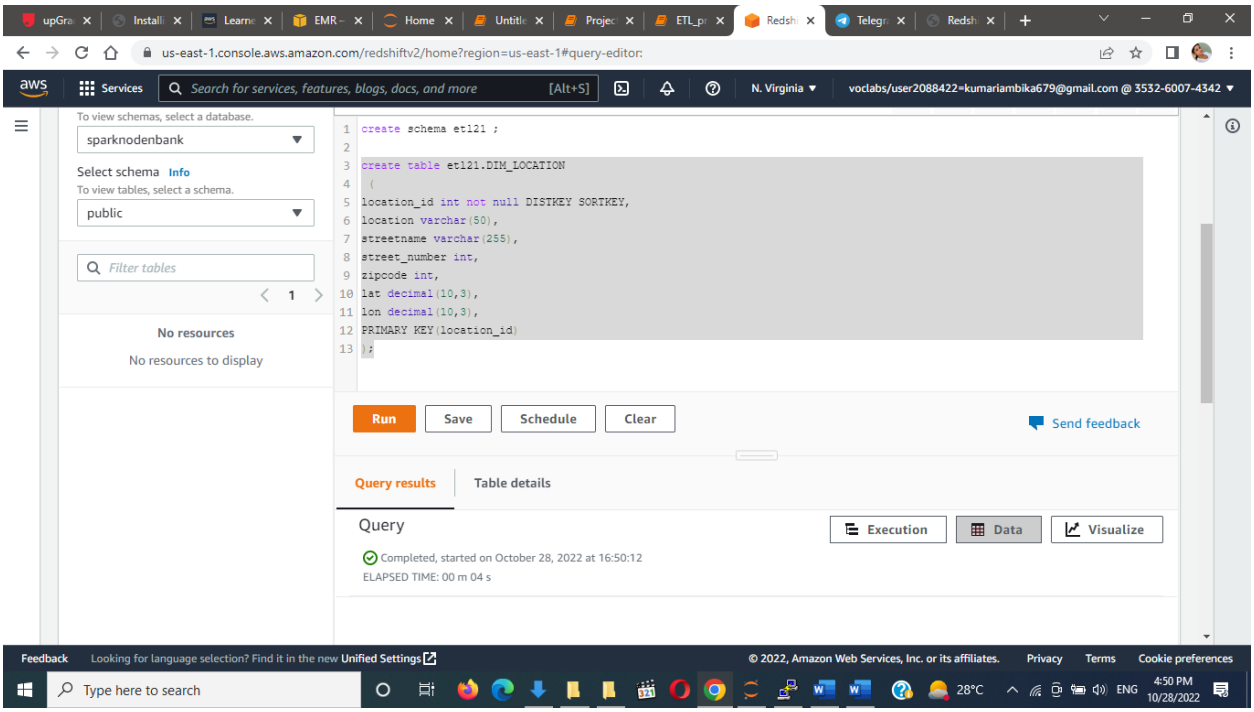
Create schema etl21;



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

• Creating location dimension table

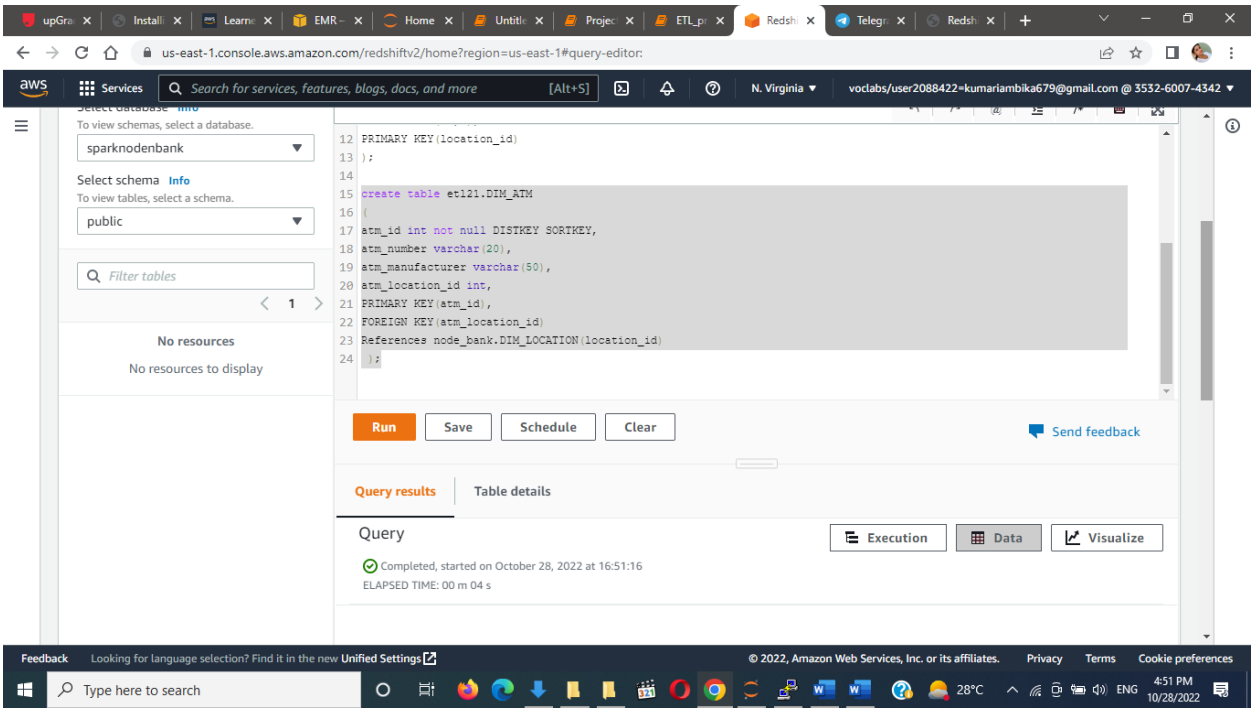
```
create table etl21.DIM_LOCATION
 (
location_id int not null DISTKEY SORTKEY,
 location varchar(50),
streetname varchar(255),
 street_number int,
zipcode int,
lat decimal(10,3),
lon decimal(10,3),
PRIMARY KEY(location_id)
);
```



• Creating atm dimension table

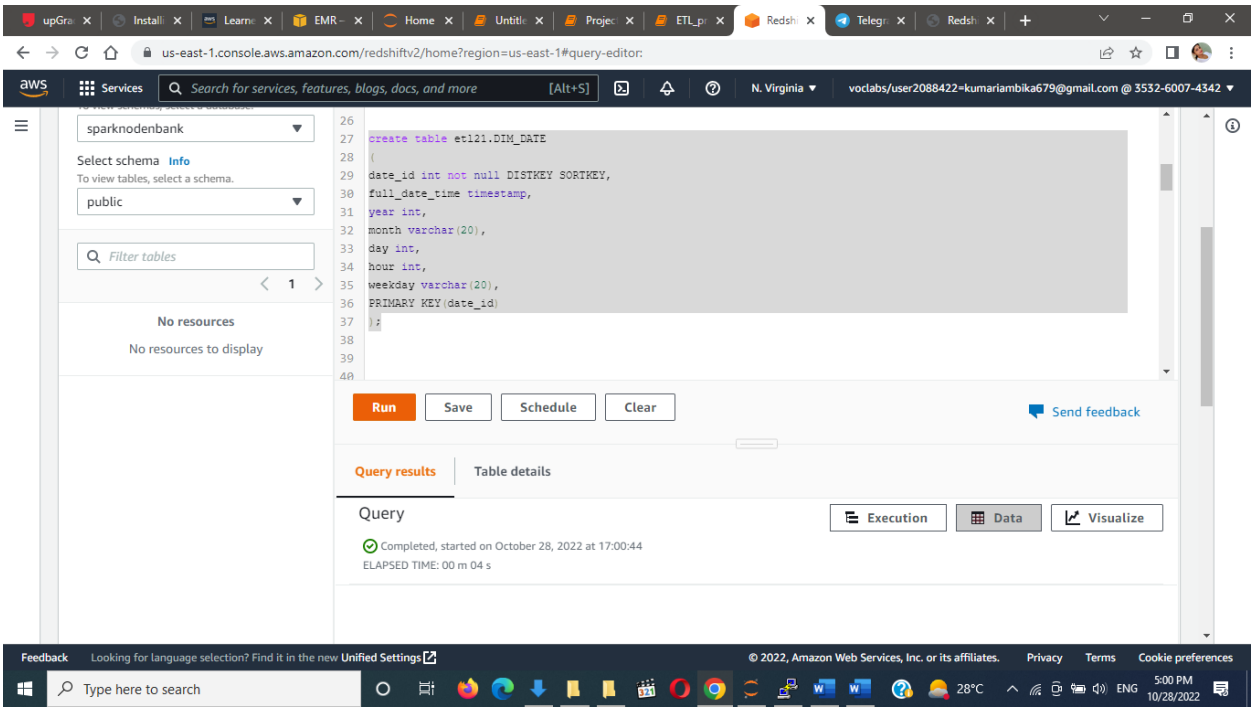
```
create table etl21.DIM_ATM
(  atm_id int not null DISTKEY SORTKEY,
 atm_number varchar(20),
```

atm_manufacturer varchar(50),
atm_location_id int,
PRIMARY KEY(atm_id),
FOREIGN KEY(atm_location_id)
References nodebank.DIM_LOCATION(location_id)
);



• Creating date dimension table

create table etl21.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 etl21.DIM_CARD_TYPE

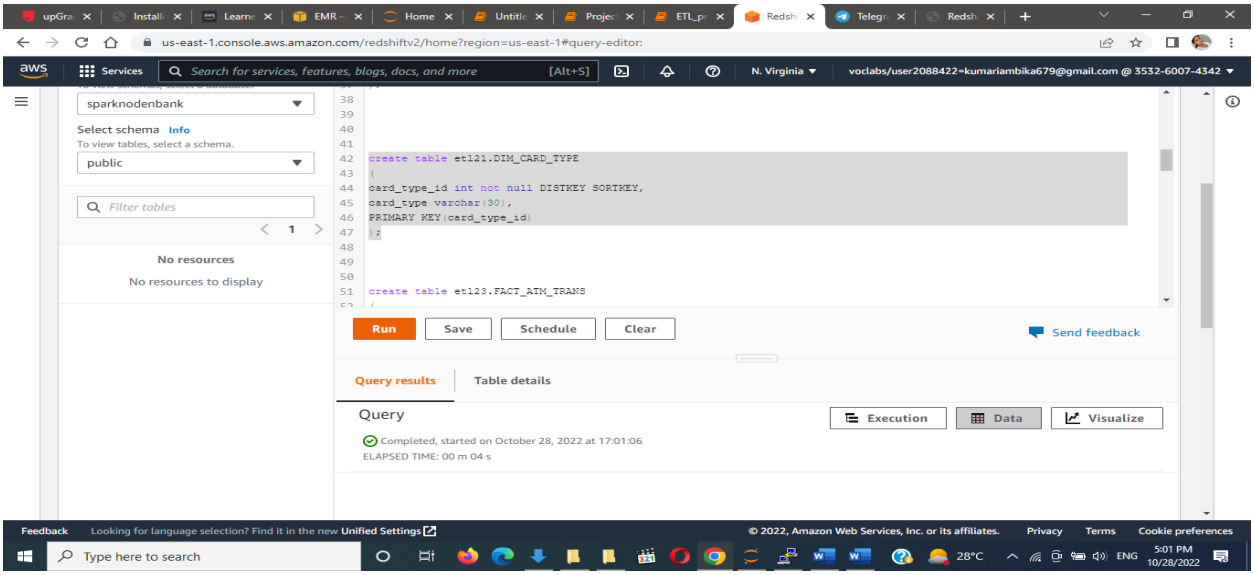
(

card_type_id int not null DISTKEY SORTKEY,

card_type varchar(30),

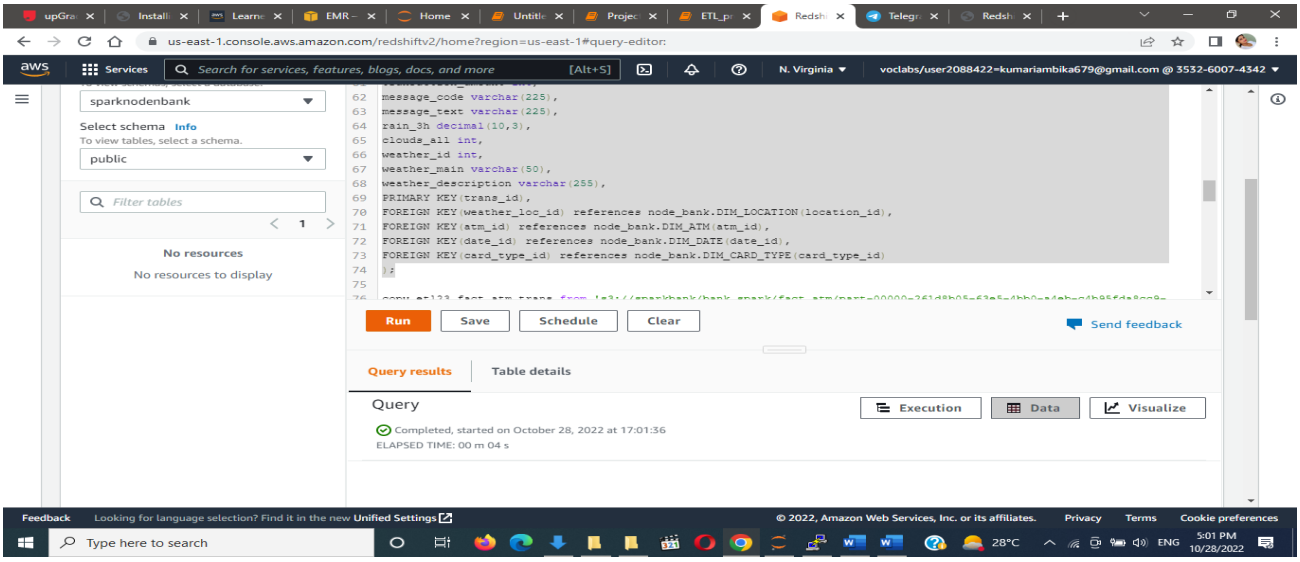
PRIMARY KEY(card_type_id)

);



• **Creating atm transactions fact table**

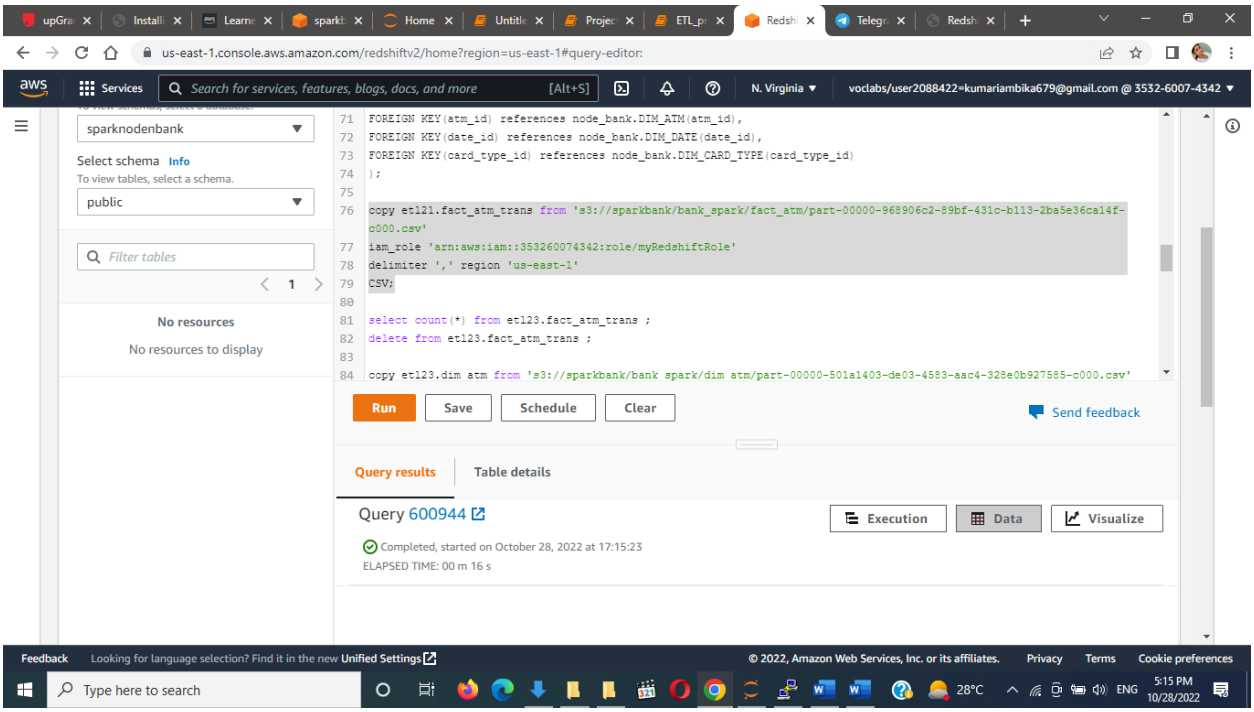
```
create table etl21.FACT_ATM_TRANS
(
trans_id bigint not null DISTKEY SORTKEY,
atm_id int,
weather_loc_id int,
date_id int,
card_type_id int,
atm_status varchar(20),
currency varchar(10),
service varchar(20),
transaction_amount int,
message_code varchar(225),
message_text varchar(225),
rain_3h decimal(10,3),
clouds_all int,
weather_id int,
weather_main varchar(50),
weather_description varchar(255),
PRIMARY KEY(trans_id),
FOREIGN KEY(weather_loc_id) references etl21.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references etl21.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references etl21.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references etl21.DIM_CARD_TYPE(card_type_id)
);
```



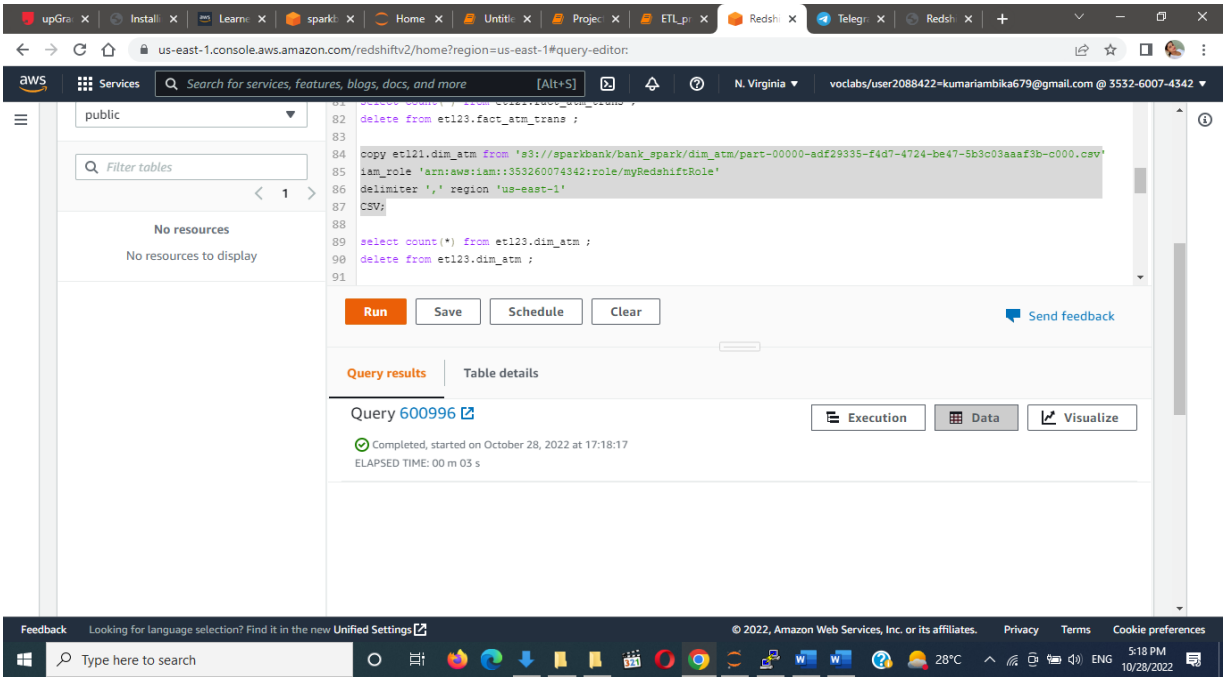
Loading data into a Redshift cluster from Amazon S3 bucket

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

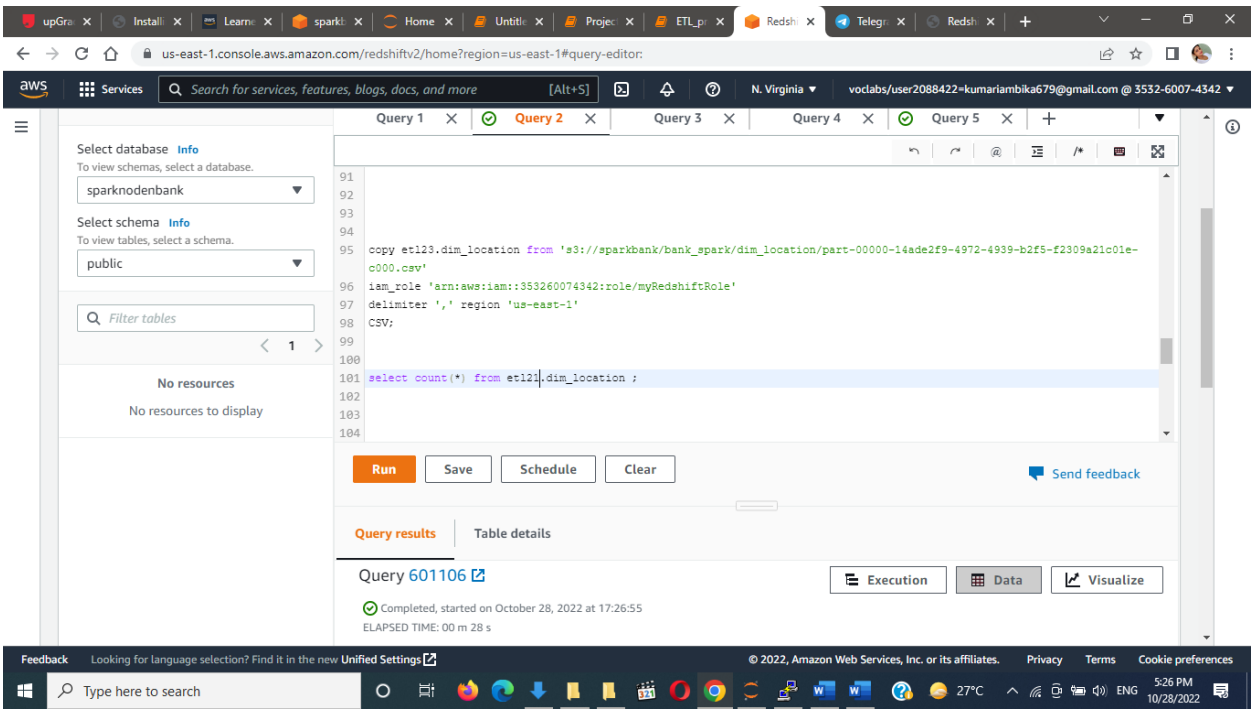
copy etl21.fact_atm_trans from 's3://sparkbank/bank_spark/fact_atm/part-00000-968906c2-89bf-431c-b113-2ba5e36ca14f-c000.csv'
iam_role 'arn:aws:iam::353260074342:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



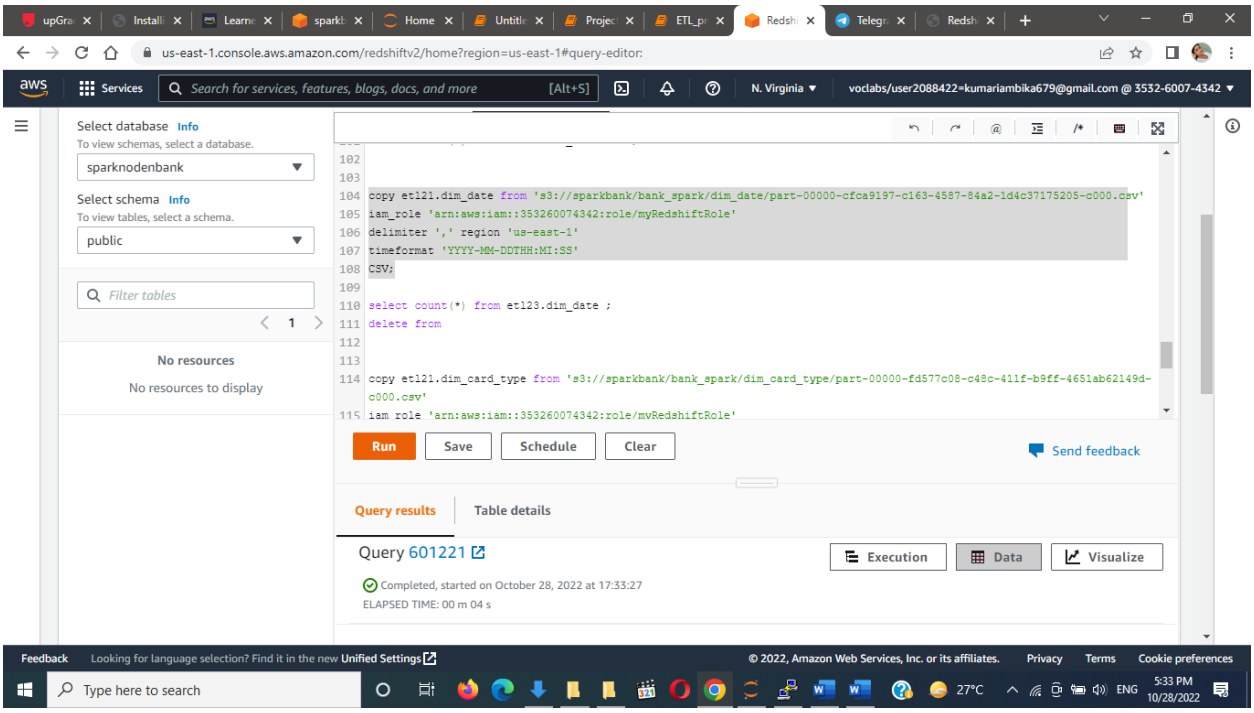
copy etl21.dim_atm from 's3://sparkbank/bank_spark/dim_atm/part-00000-adf29335-f4d7-4724-be47-5b3c03aaaf3b-c000.csv'
iam_role 'arn:aws:iam::353260074342:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



copy etl23.dim_location from 's3://sparkbank/bank_spark/dim_location/part-00000-14ade2f9-4972-4939-b2f5-f2309a21c01e-c000.csv'
iam_role 'arn:aws:iam::353260074342:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



copy etl21.dim_date from 's3://sparkbank/bank_spark/dim_date/part-00000-cfca9197-c163-4587-84a2-1d4c37175205-c000.csv'
iam_role 'arn:aws:iam::353260074342:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
timeformat 'YYYY-MM-DDTHH:MI:SS'
CSV;



copy etl21.dim_card_type from 's3://sparkbank/bank_spark/dim_card_type/part-00000-eefc413e-0958-43e6-b9a0-bf8a594d52a2-c000.csv'
iam_role 'arn:aws:iam::353260074342:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;