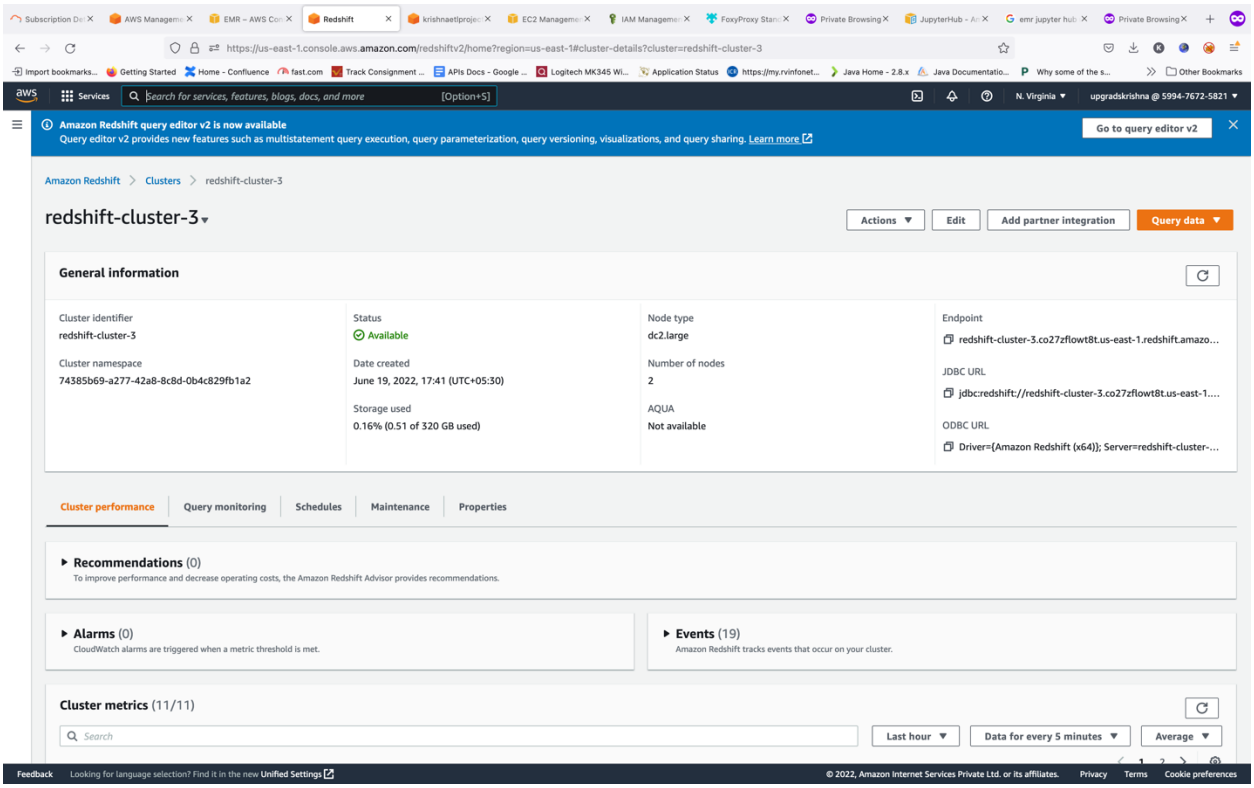
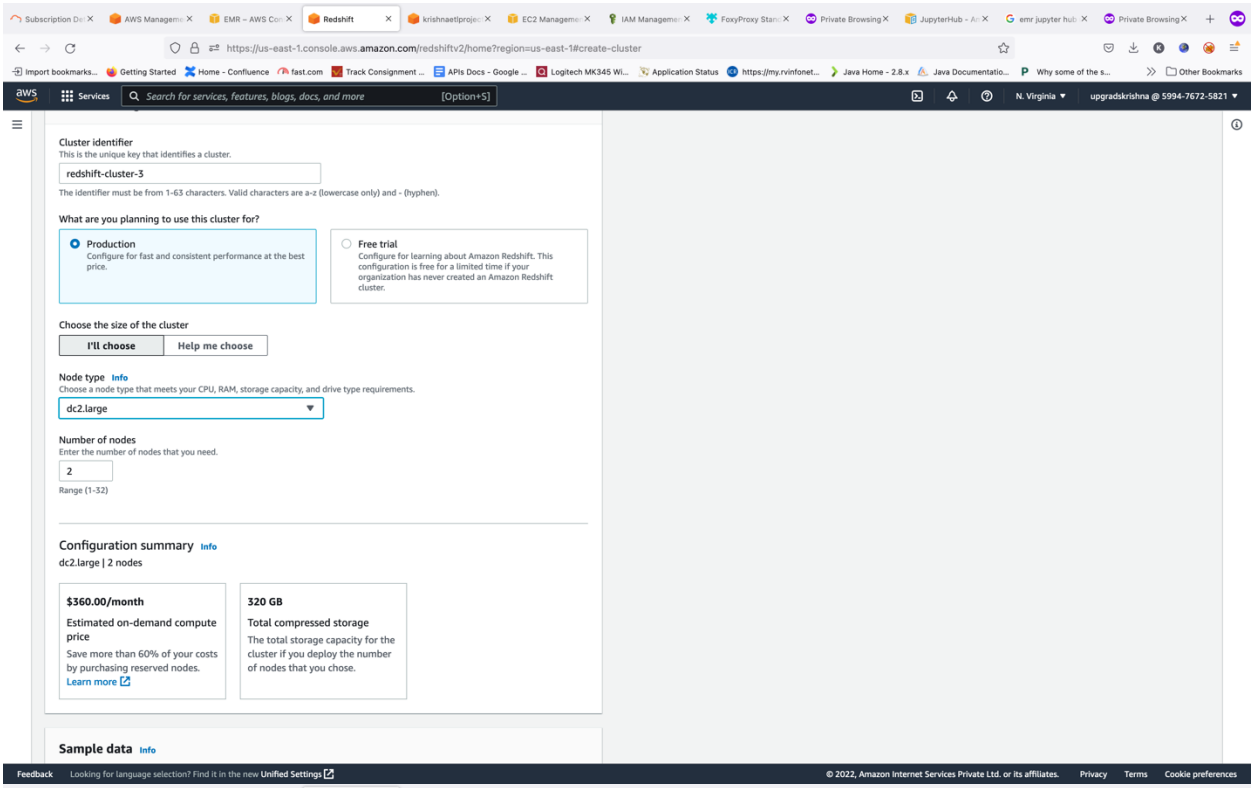
Creation of a Redshift Cluster

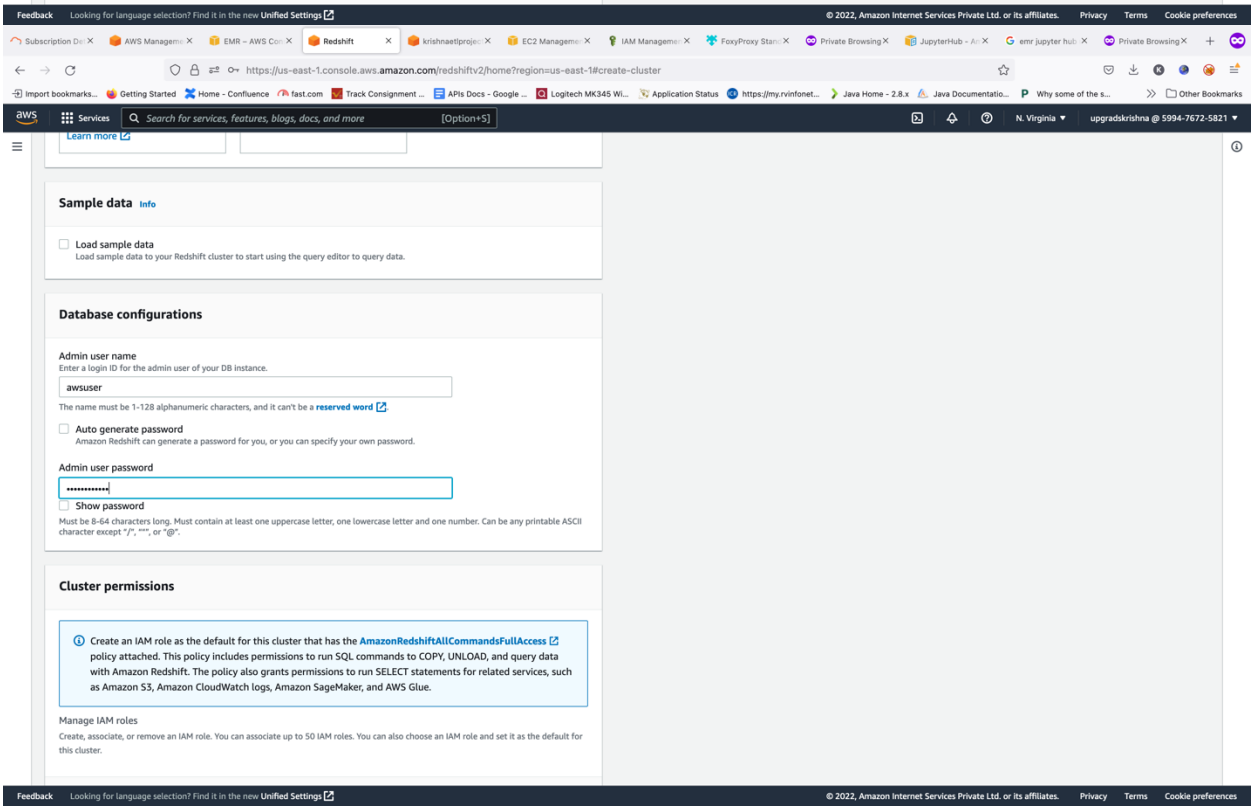
Screenshots of the configuration of the Redshift cluster that you have created:

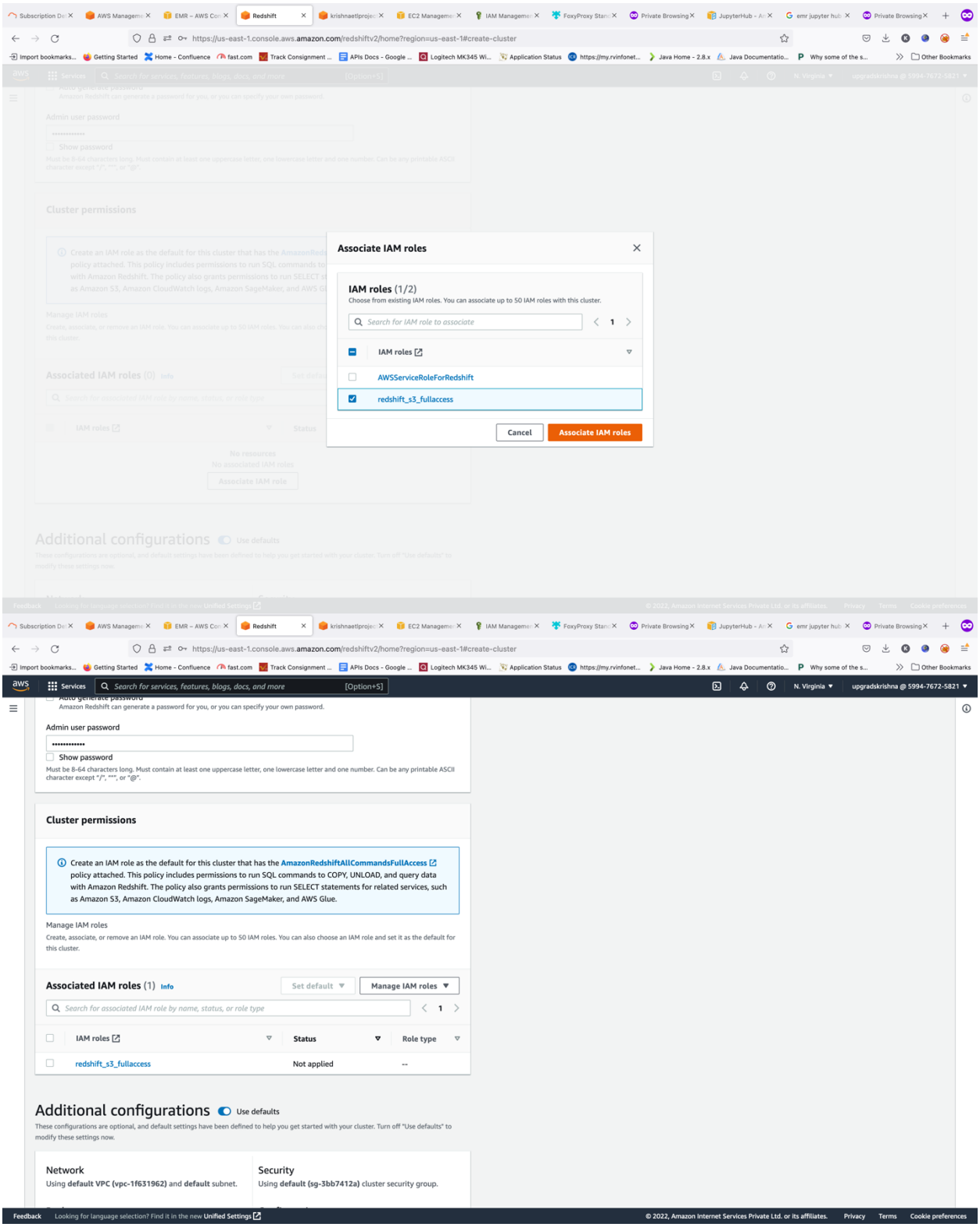
Screenshot of the type of machine used along with number of nodes:

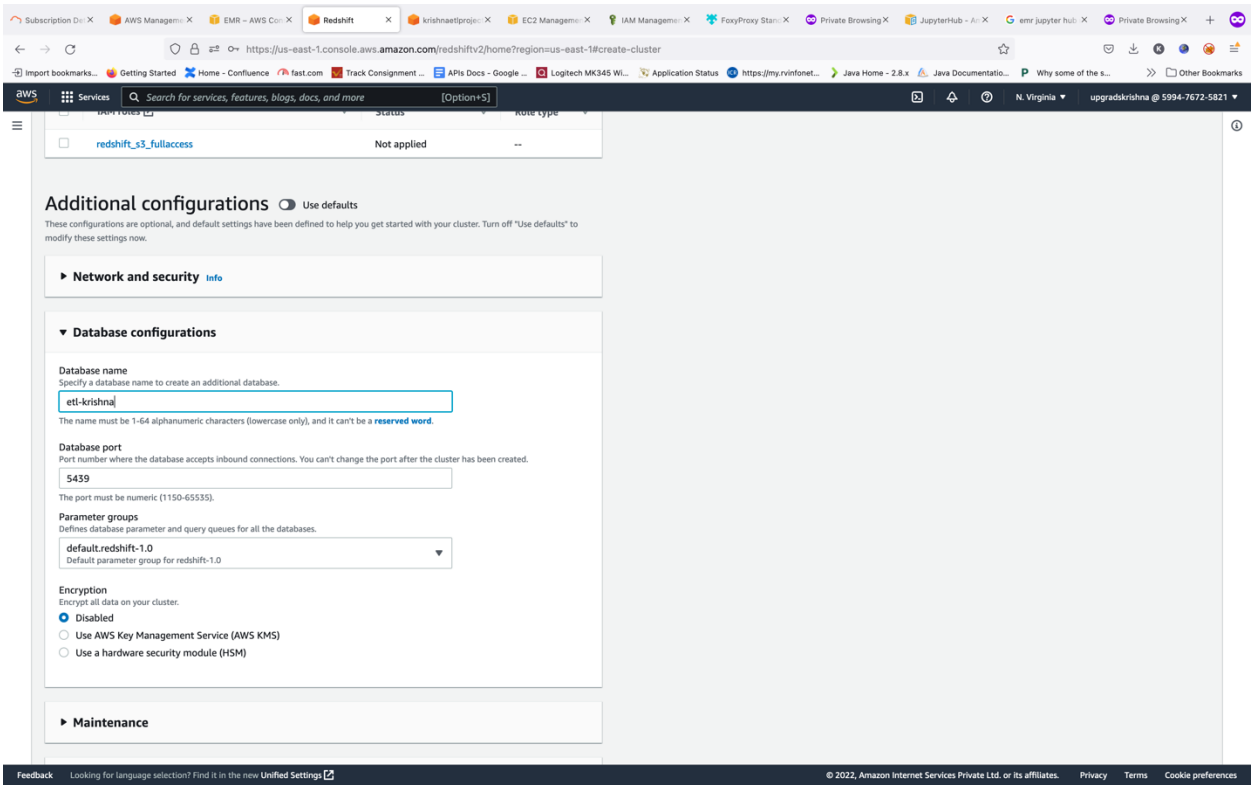


Screenshot of steps taken to create the Redshift cluster:



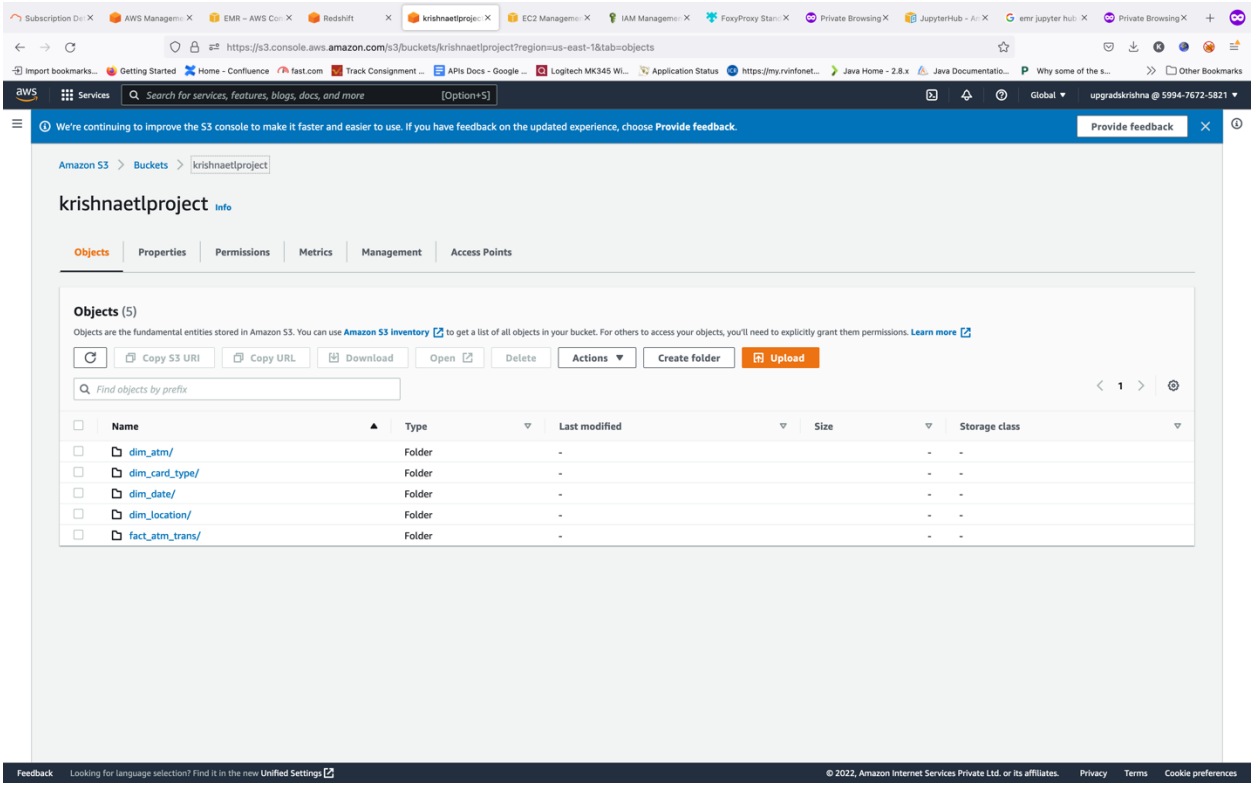






Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Viewing data in s3 bucket:



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

Pre-requisites:

```
create schema atm_data;


drop table atm_data.dim_atm;
drop table atm_data.dim_card_type;
drop table atm_data.dim_date;
drop table atm_data.dim_location;
drop table atm_data.fact_atm_trans;
```

Queries to create the tables:

To create DIM_LOCATION:

```
create table 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)
);
```

To create DIM_ATM:

```
create table 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 atm_data.DIM_LOCATION(location_id)
);
```

To create DIM_DATE:

```
create table 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)
);
```

To create DIM_CARD_TYPE:

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

To create FACT_ATM_TRANS:

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

To copy data to the appropriate tables from s3, I used the following queries:

To copy data from s3 to DIM_LOCATION table:

```
copy atm_data.dim_location from 's3://krishnaetlproject/dim_location/part-00000-
a5d22bc1-c41e-4feb-ae90-c7043bd03103-c000.csv'
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

To copy data from s3 to DIM_ATM table:

```
copy atm_data.dim_atm from 's3://krishnaetlproject/dim_atm/part-00000-a1acba1d-585d-
403c-b6a4-5cfc0704ac4b-c000.csv'
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

To copy data from s3 to DIM_DATE table:

```
copy atm_data.dim_date from 's3://krishnaetlproject/dim_date/part-00000-d6700597-cd65-
4555-b8ef-735e8deb375a-c000.csv'
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
TIMEFORMAT AS 'auto'
CSV;
```

To copy data from s3 to DIM_CARD_TYPE table:

```
copy atm_data.dim_card_type from 's3://krishnaetlproject/dim_card_type/part-00000-
f0097118-fead-46ed-b0fd-cc5606086622-c000.csv'
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

To copy data from s3 to FACT_ATM_TRANS table:

```
copy atm_data.fact_atm_trans from 's3://krishnaetlproject/fact_atm_trans/part-00000-
6341e28c-222c-44b1-a0c1-dfc64080425b-c000.csv'
iam_role 'arn:aws:iam::599476725821:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Screenshot of created tables list: