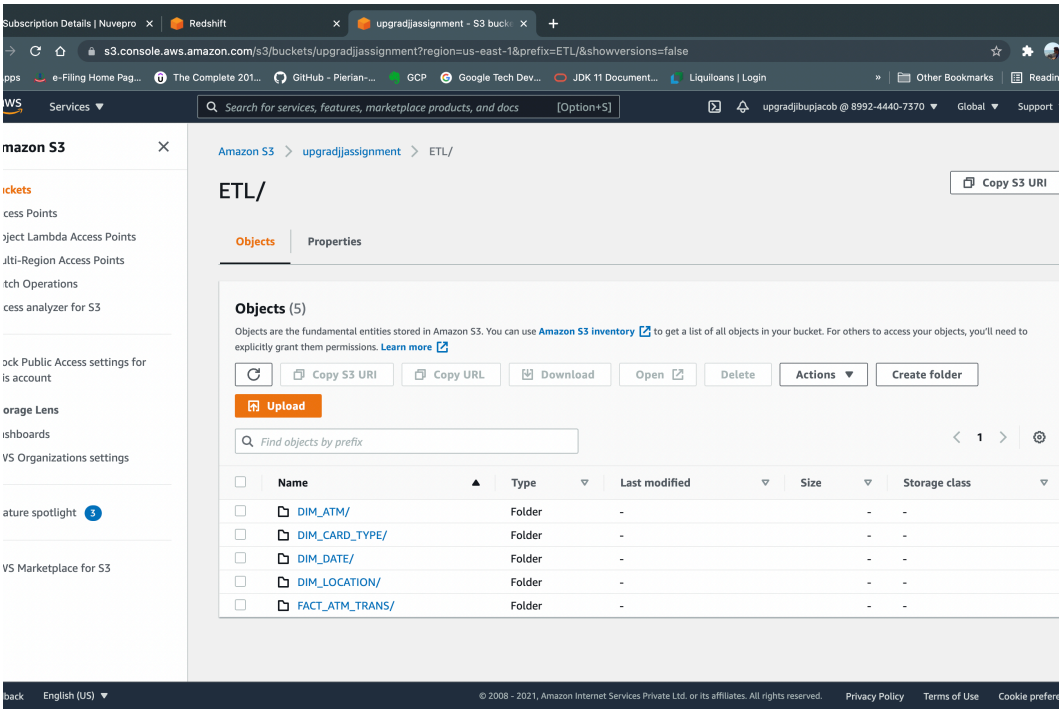
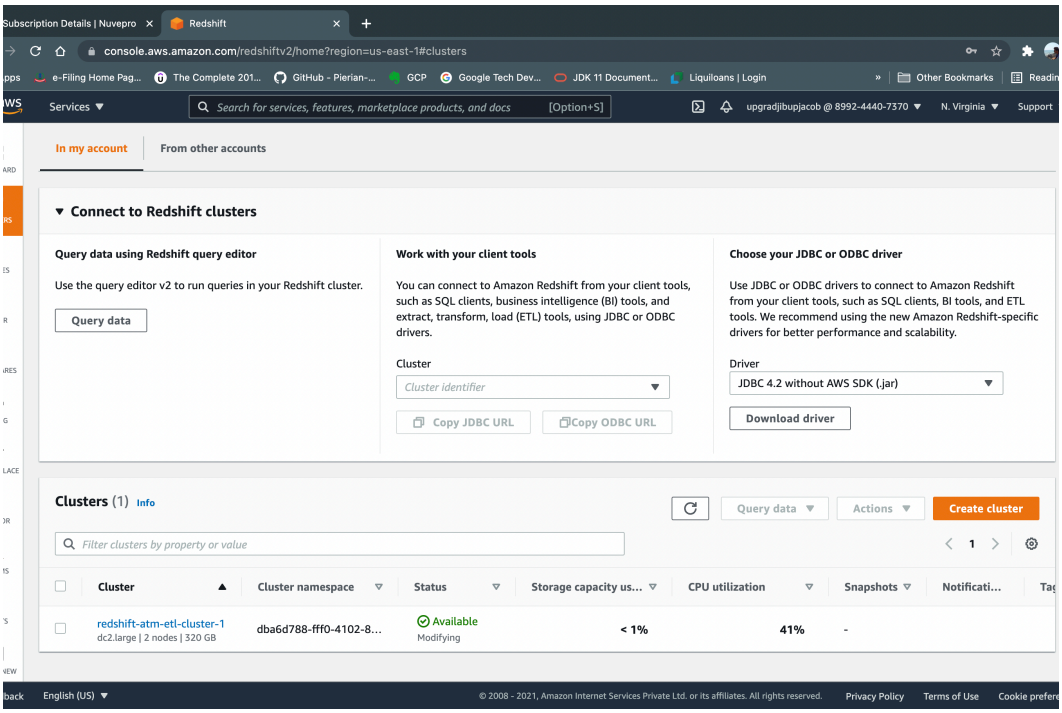
Pre-requisites csv Datasets created via the spark setup to be available in s3

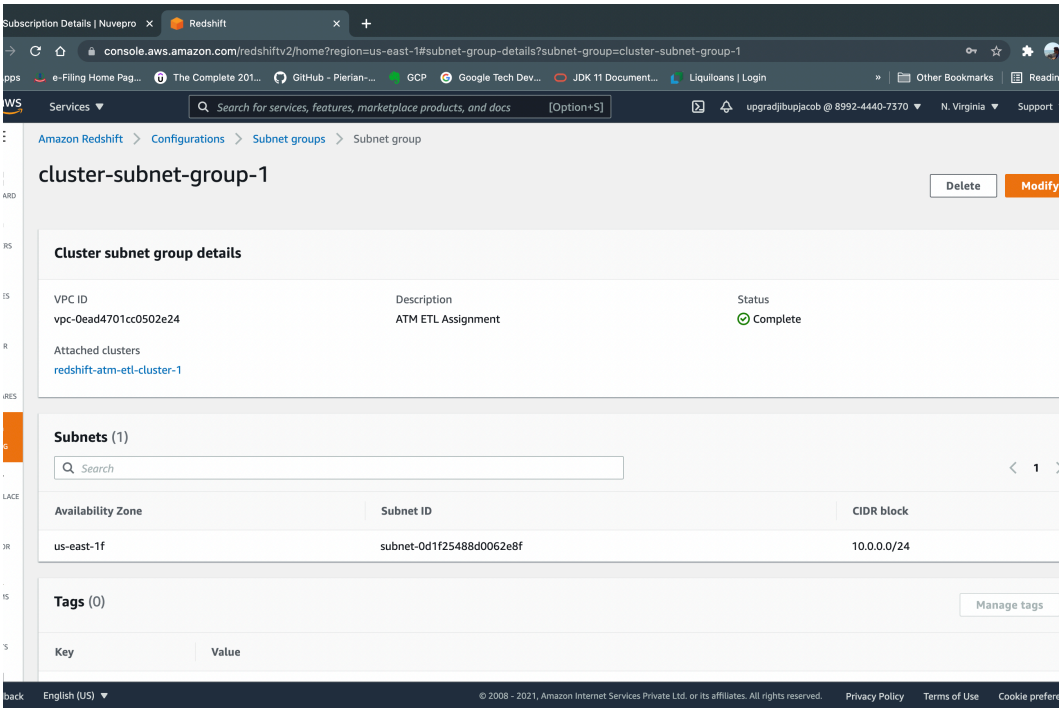


Creation of a Redshift Cluster

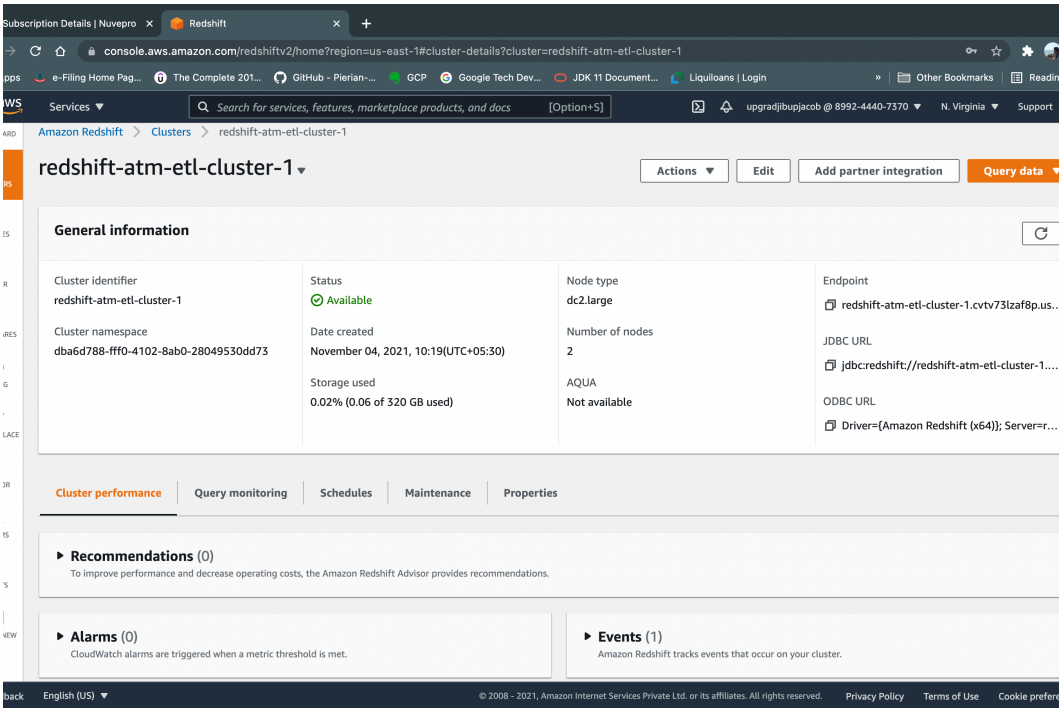
Screenshots of the configuration of the RedShift cluster that you have created:



Cluster Name



Subnet group and subnet



Cluster Details

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

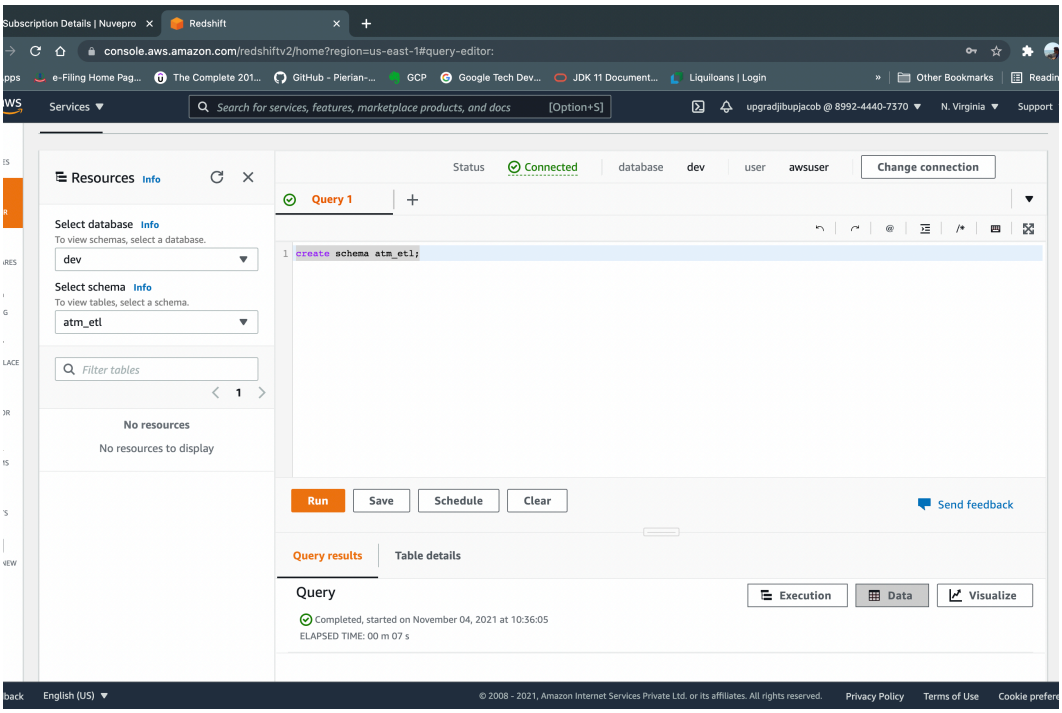
Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

1. Schema Creation:

```
create schema atm_etl;
```

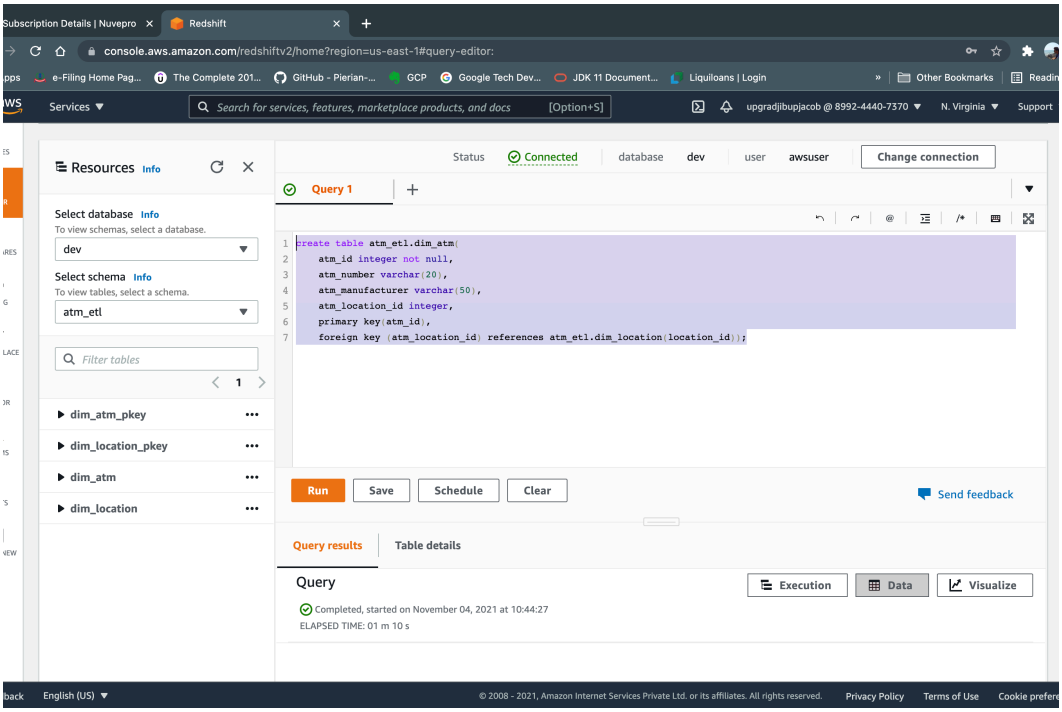
2. DIM_LOCATION dimension table creation

```
create table atm_etl.dim_location(
location_id integer not null,
location varchar(50),
streetname varchar(255),
street_number integer,
zipcode integer,
lat numeric(10,3),
lon numeric(10,3),
primary key(location_id));
```



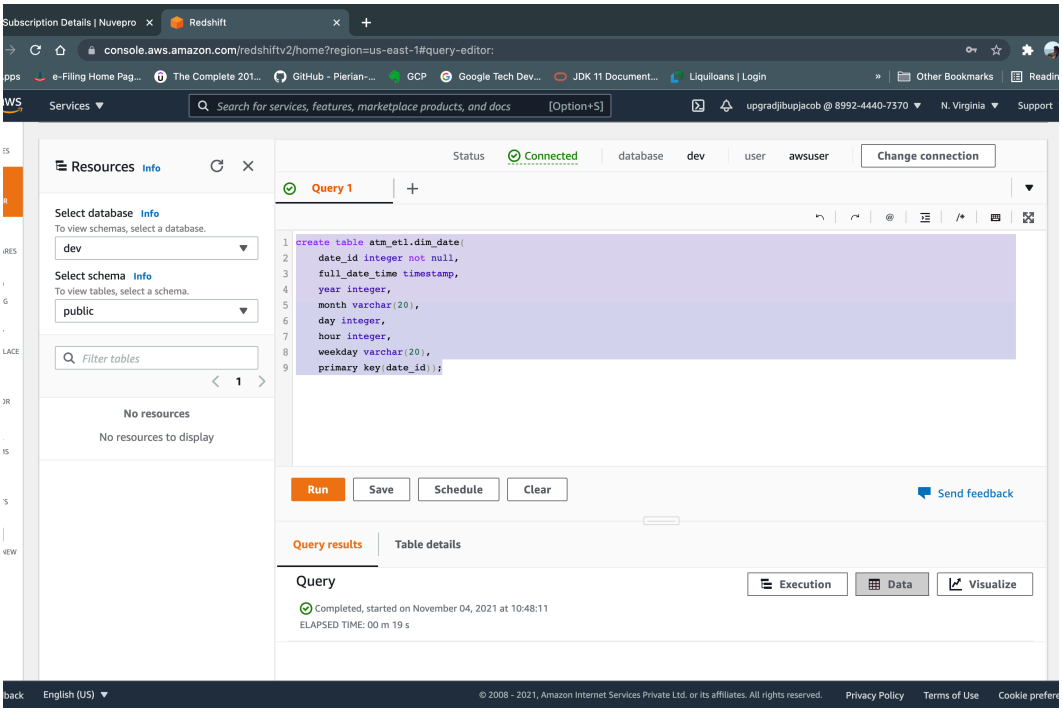
3. DIM_ATM dimension table creation

```
create table atm_etl.dim_atm(
atm_id integer not null,
atm_number varchar(20),
atm_manufacturer varchar(50),
atm_location_id integer,
primary key(atm_id),
foreign key (atm_location_id) references atm_etl.dim_location(location_id));
```



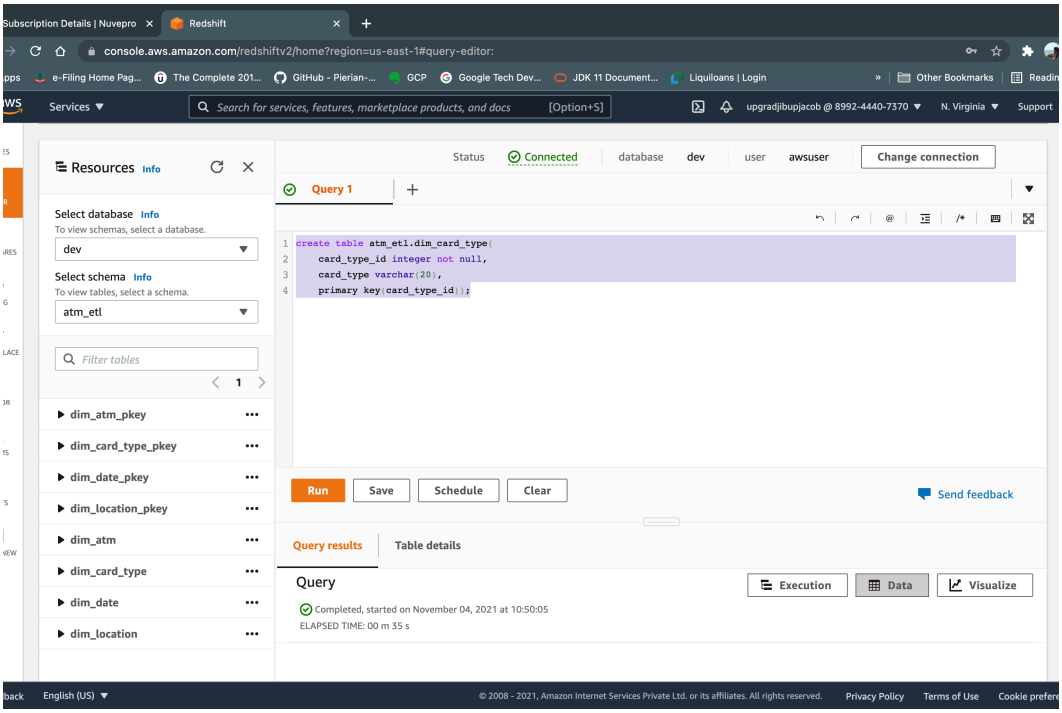
4.DIM_DATE dimension table creation

```
create table atm_etl.dim_date(
date_id integer not null,
full_date_time timestamp,
year integer,
month varchar(20),
day integer,
hour integer,
weekday varchar(20),
primary key(date_id));
```



5. DIM_CARD_TYPE dimension table creation

```
create table atm_etl.dim_card_type(
card_type_id integer not null,
card_type varchar(20),
primary key(card_type_id));
```

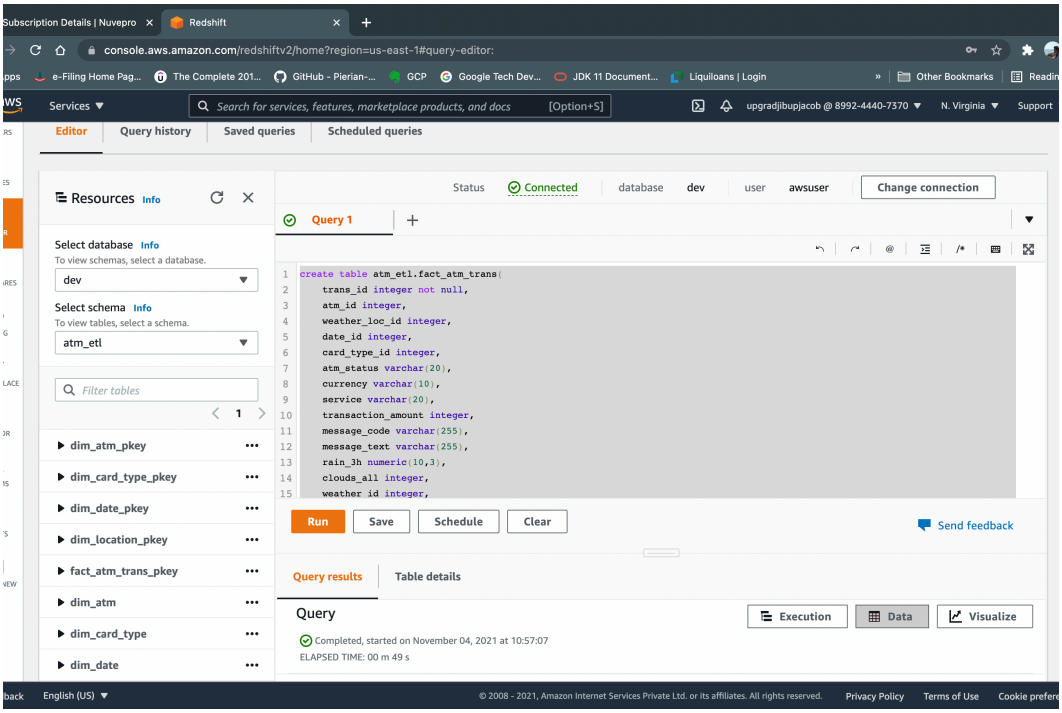


Altered the size of the card type to 23 to accommodate the data

```
alter table atm_etl.dim_card_type alter column card_type type varchar(23);
```

6. FACT_ATM_TRANS fact table creation

```
create table atm_etl.fact_atm_trans(
trans_id integer not null,
atm_id integer,
weather_loc_id integer,
date_id integer,
card_type_id integer,
atm_status varchar(20),
currency varchar(10),
service varchar(20),
transaction_amount integer,
message_code varchar(255),
message_text varchar(255),
rain_3h numeric(10,3),
clouds_all integer,
weather_id integer,
weather_main varchar(50),
weather_description varchar(255),
primary key(trans_id),
foreign key (weather_loc_id) references atm_etl.dim_location(location_id),
foreign key (atm_id) references atm_etl.dim_atm(atm_id),
foreign key (date_id) references atm_etl.dim_date(date_id),
foreign key (card_type_id) references atm_etl.dim_card_type(card_type_id));
```

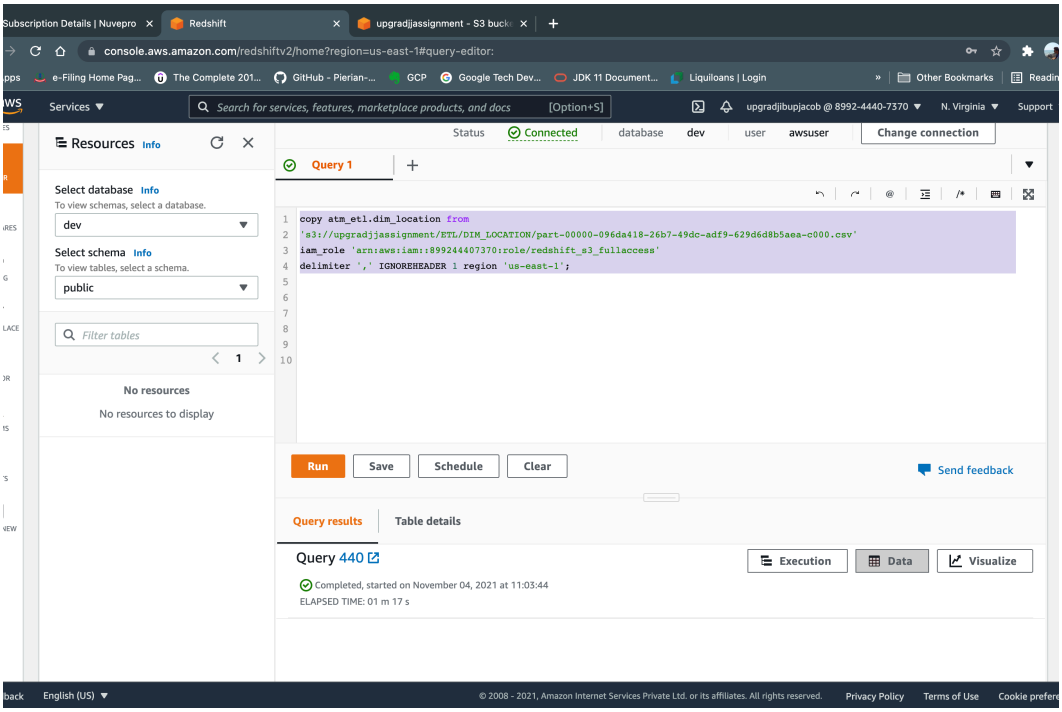


Loading data into a RedShift cluster from Amazon S3 bucket

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

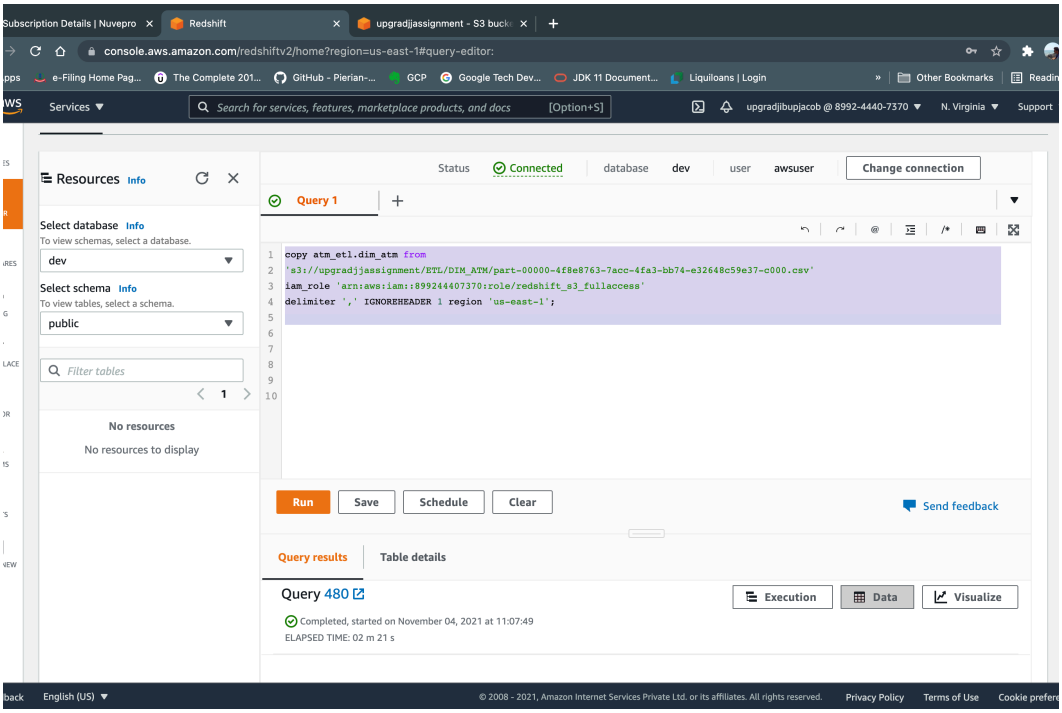
1. Copy DIM_LOCATION data from S3 to Redshift

```
copy atm_etl.dim_location from
's3://upgradjjassignment/ETL/DIM_LOCATION/part-00000-096da418-26b7-49dc-adf9-629d6d8b5aea-c000.csv'
iam_role 'arn:aws:iam::899244407370:role/redshift_s3_fullaccess'
delimiter ',' IGNOREHEADER 1 region 'us-east-1';
```



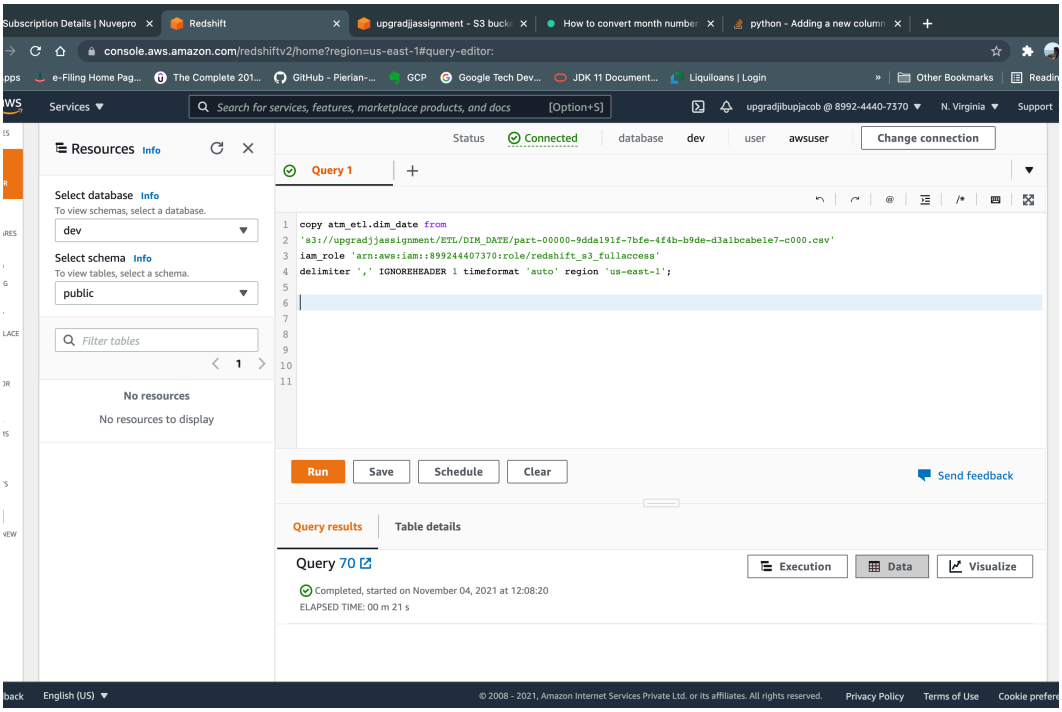
2. Copy DIM_ATM data from S3 to Redshift

copy atm_etl.dim_atm from
's3://upgradjjassignment/ETL/DIM_ATM/part-00000-4f8e8763-7acc-4fa3-bb74-e32648c59e37-c000.csv'
iam_role 'arn:aws:iam::899244407370:role/redshift_s3_fullaccess'
delimiter ',' IGNOREHEADER 1 region ‘us-east-1';



3. Copy DIM_DATE data from S3 to Redshift

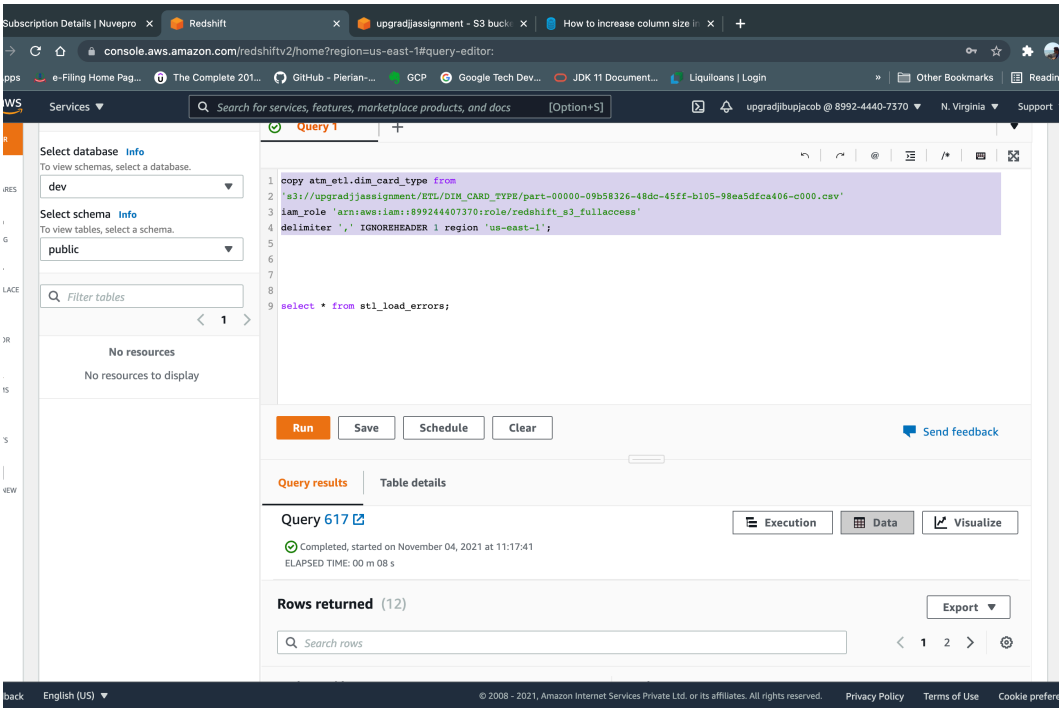
copy atm_etl.dim_date from
's3://upgradjjassignment/ETL/DIM_DATE/part-00000-9dda191f-7bfe-4f4b-b9de-d3a1bcabe1e7-c000.csv'
iam_role 'arn:aws:iam::899244407370:role/redshift_s3_fullaccess'
delimiter ',' IGNOREHEADER 1 timeformat 'auto' region 'us-east-1';



Caption

4. Copy DIM_CARD_TYPE data from S3 to Redshift

copy atm_etl.dim_card_type from
's3://upgradjjassignment/ETL/DIM_CARD_TYPE/part-00000-09b58326-48dc-45ff-b105-98ea5dfca406-c000.csv'
iam_role 'arn:aws:iam::899244407370:role/redshift_s3_fullaccess'
delimiter ',' IGNOREHEADER 1 region 'us-east-1';



5. Copy FACT_ATM_TRANS data from S3 to Redshift

copy atm_etl.fact_atm_trans from
's3://upgradjjassignment/ETL/FACT_ATM_TRANS/part-00000-17259abc-6328-48b2-a7f8-69f53ada18eb-c000.csv'
iam_role 'arn:aws:iam::899244407370:role/redshift_s3_fullaccess'
delimiter ',' IGNOREHEADER 1 region 'us-east-1' TRUNCATECOLUMNS CSV;