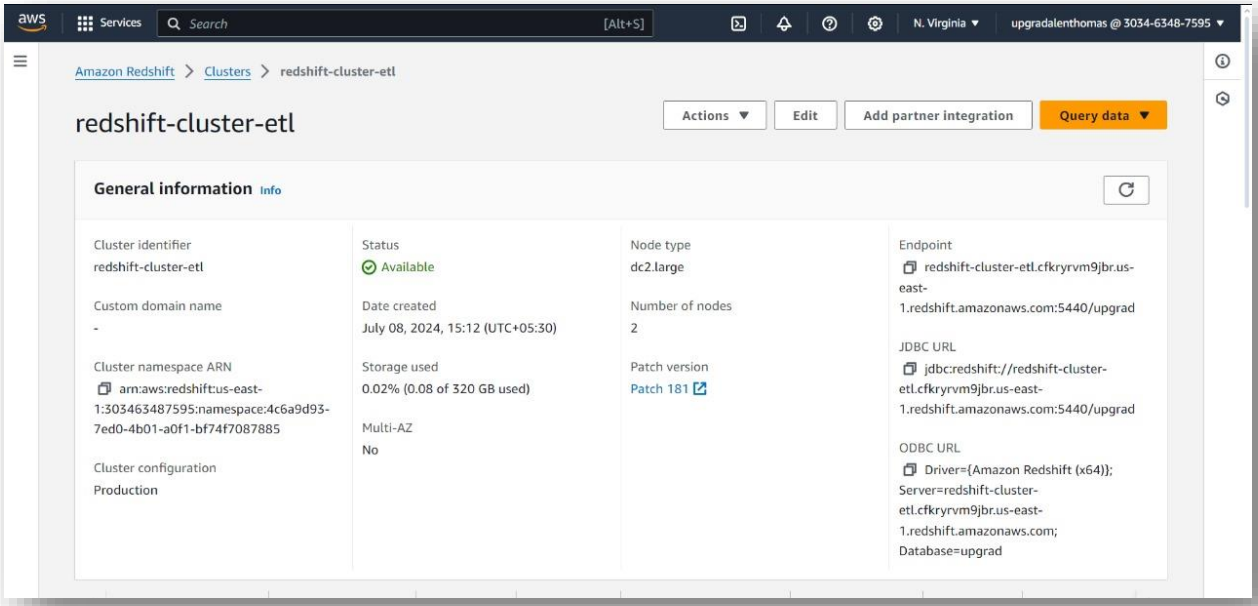
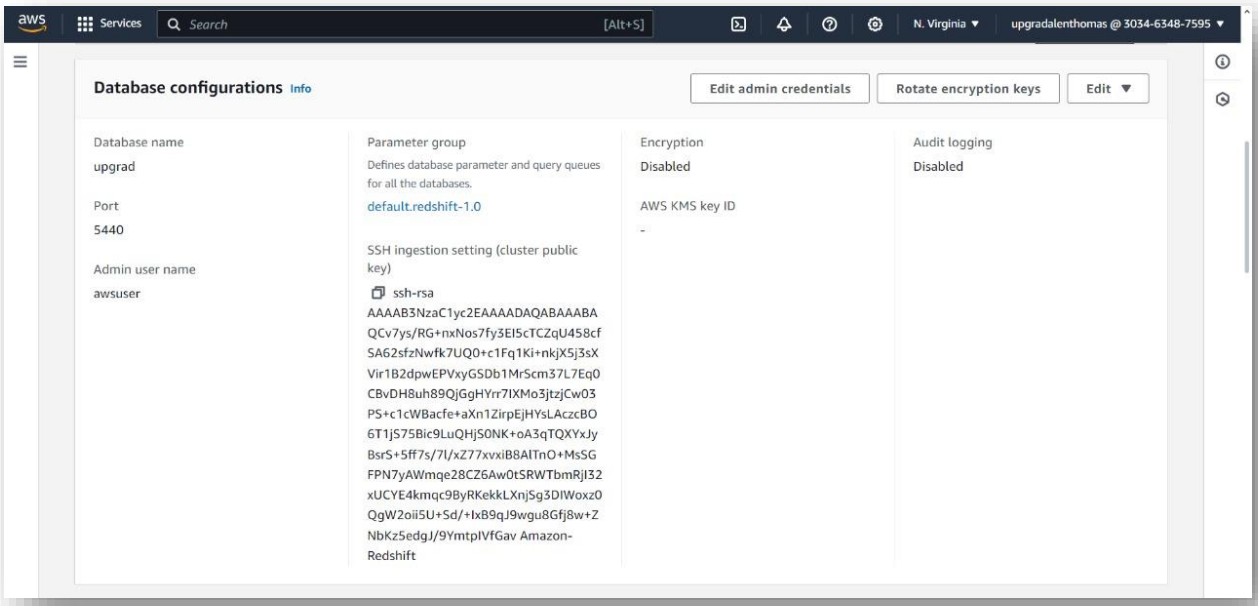
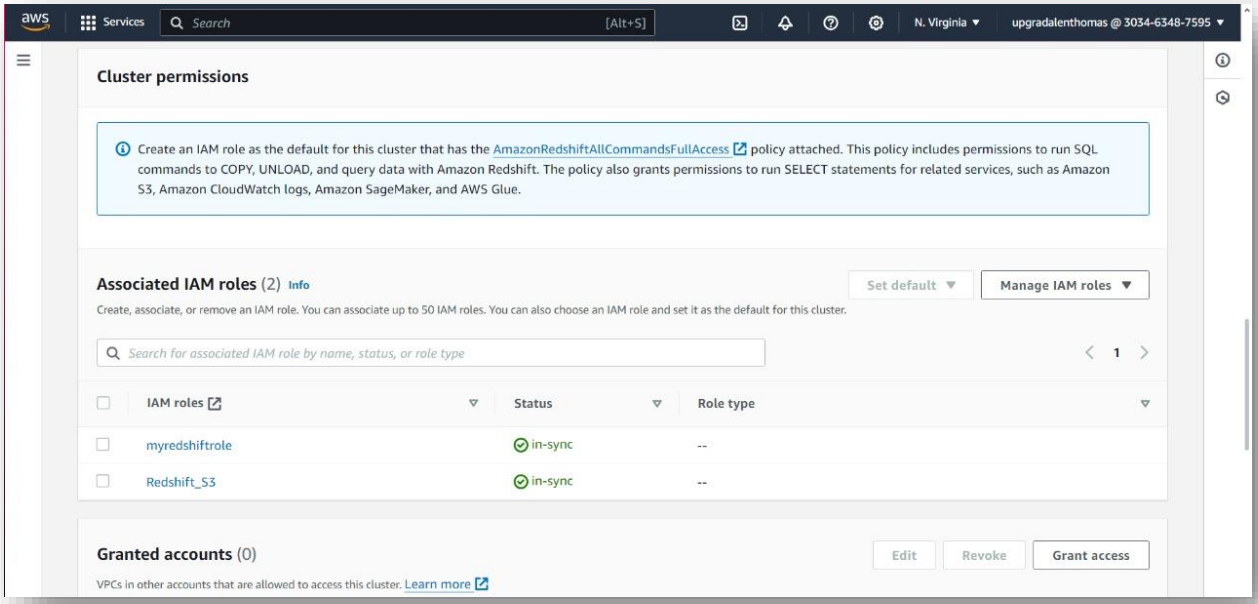
Creating a RedShift Cluster

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

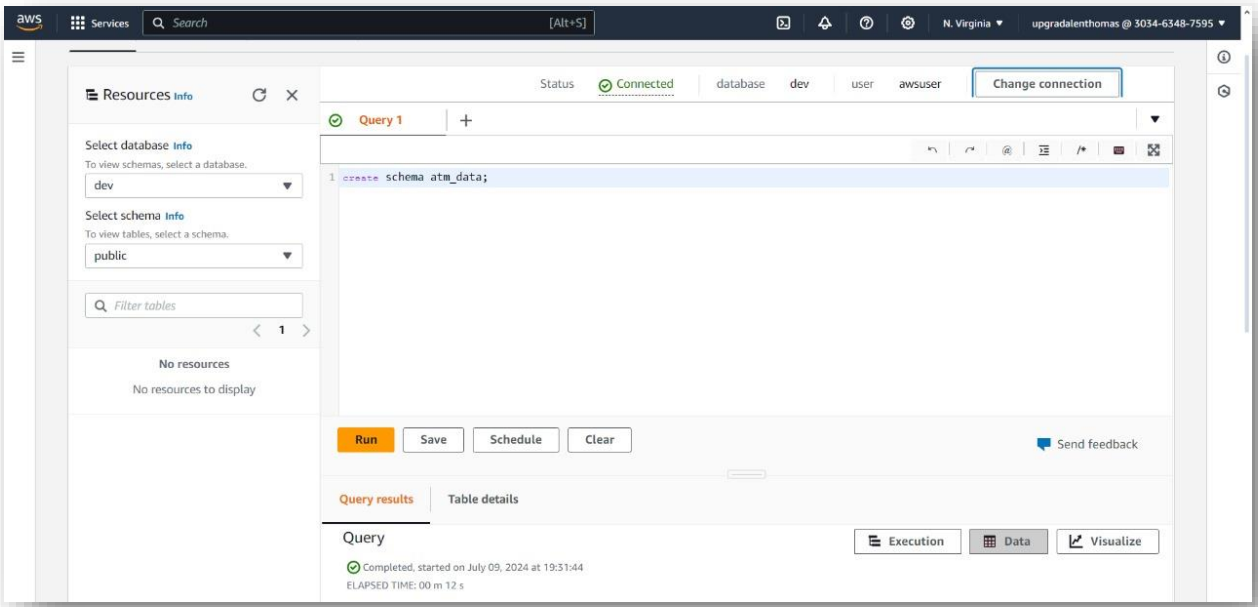






Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

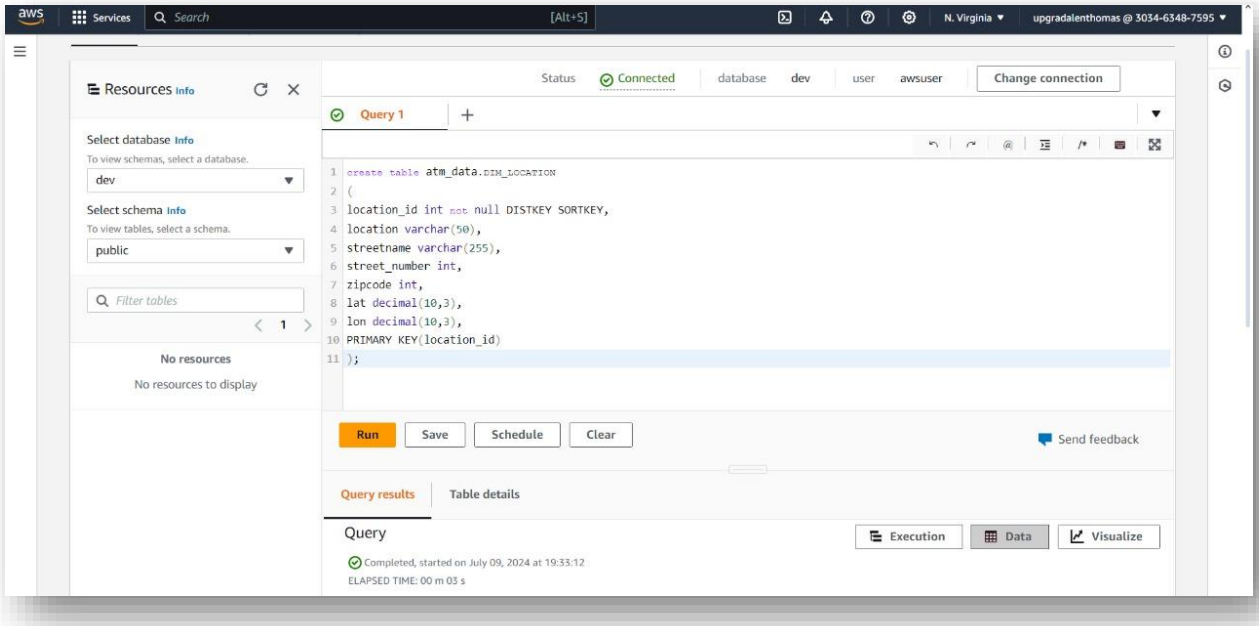
Query for creating schema:



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

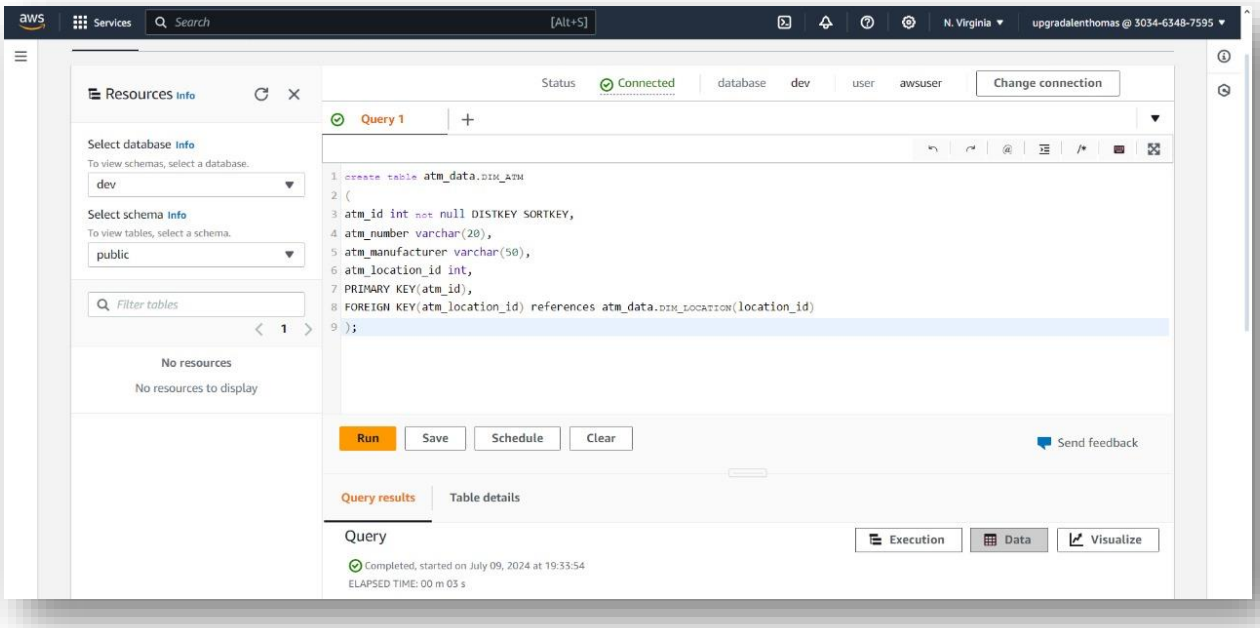
1. **Creating location dimension table**

```
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)
);
```



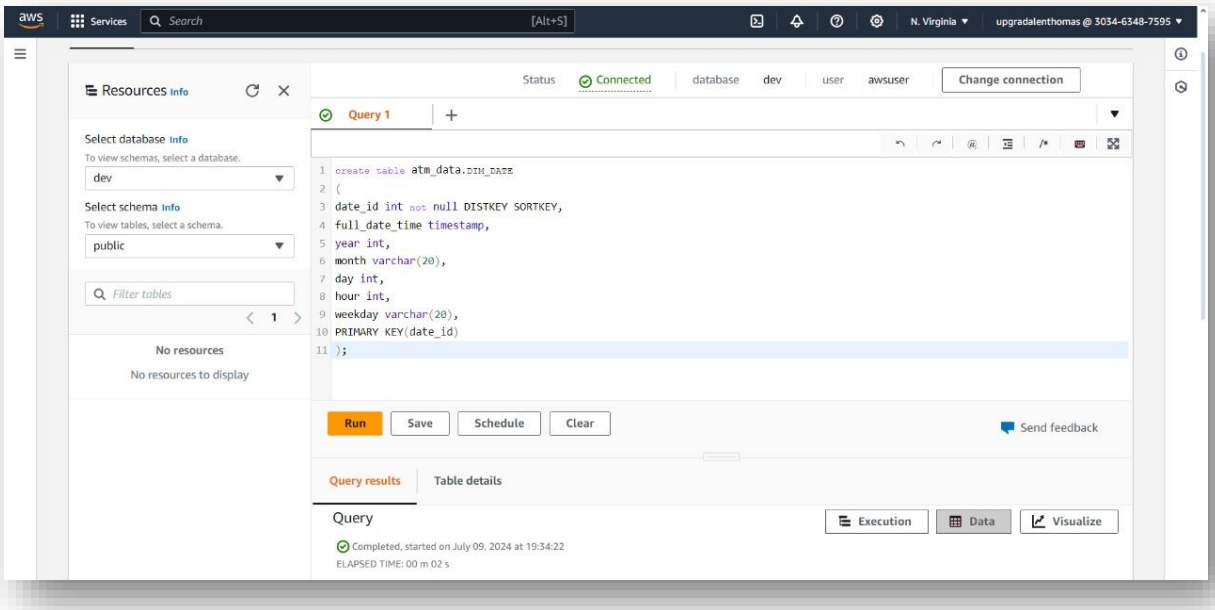
2. **Creating atm dimension table**

```
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) );
```



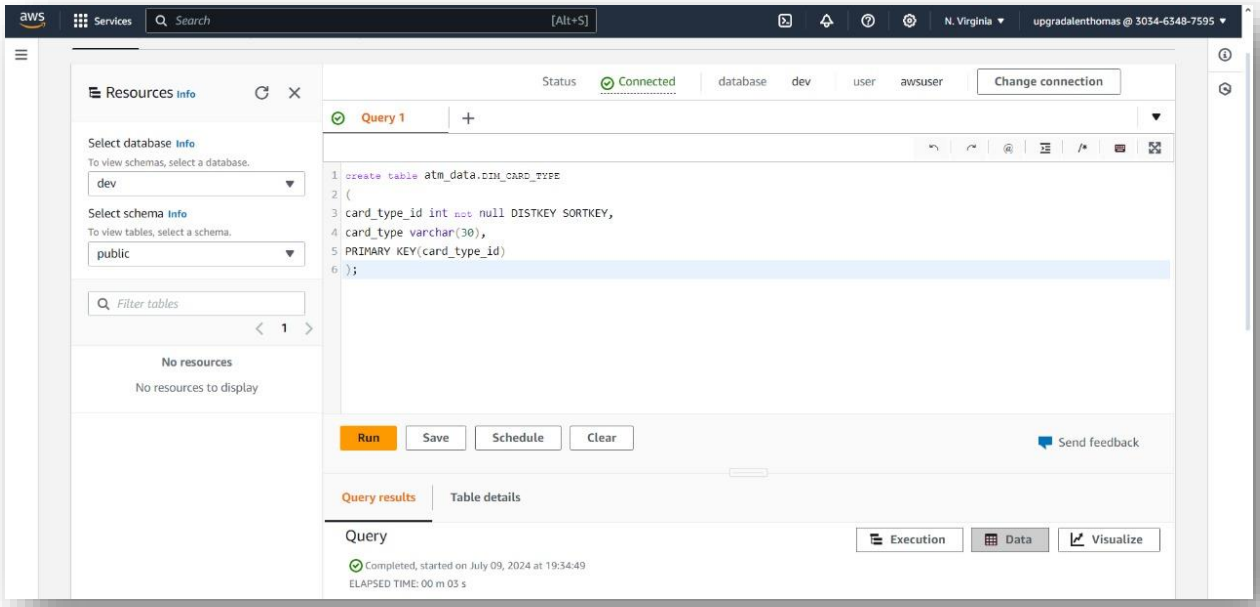
3. **Creating date dimension table**

```
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) );
```



4. **Creating card type dimension table**

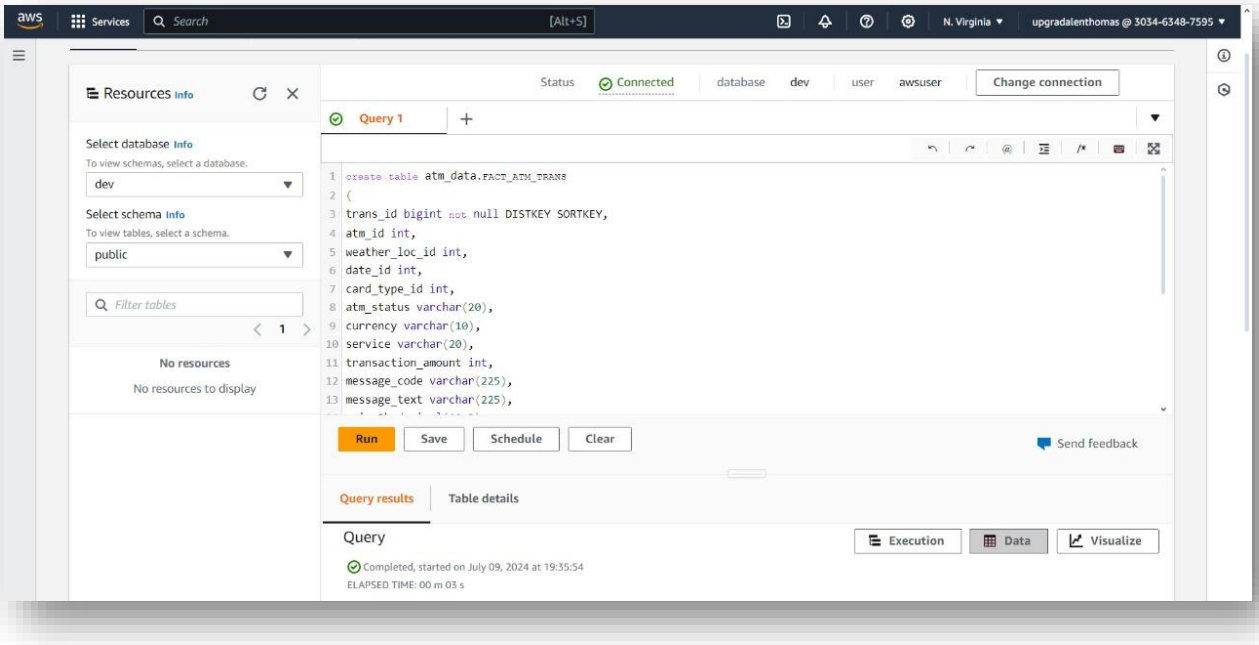
```
create table atm_data.DIM_CARD_TYPE
(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);
```



5. **Creating atm transactions fact table**

```
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),
```

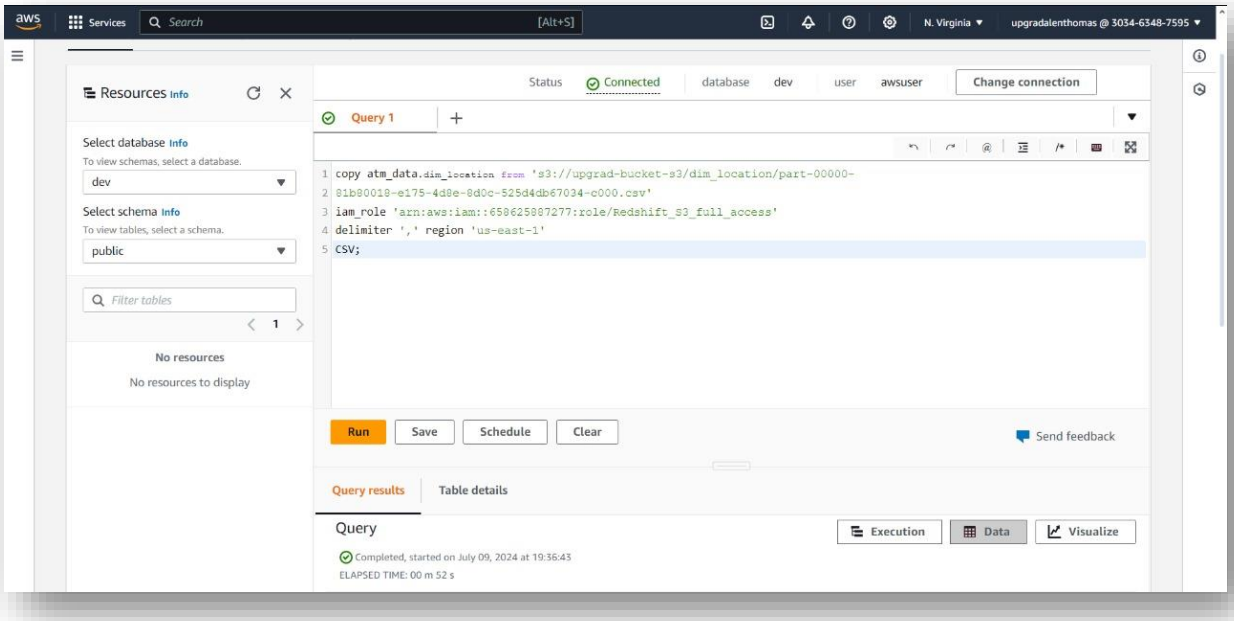
```
PRIMARY KEY(trans_id),
FOREIGN KEY(weather_loc_id) references atm_data.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references atm_data.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references atm_data.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references atm_data.DIM_CARD_TYPE(card_type_id)
);
```



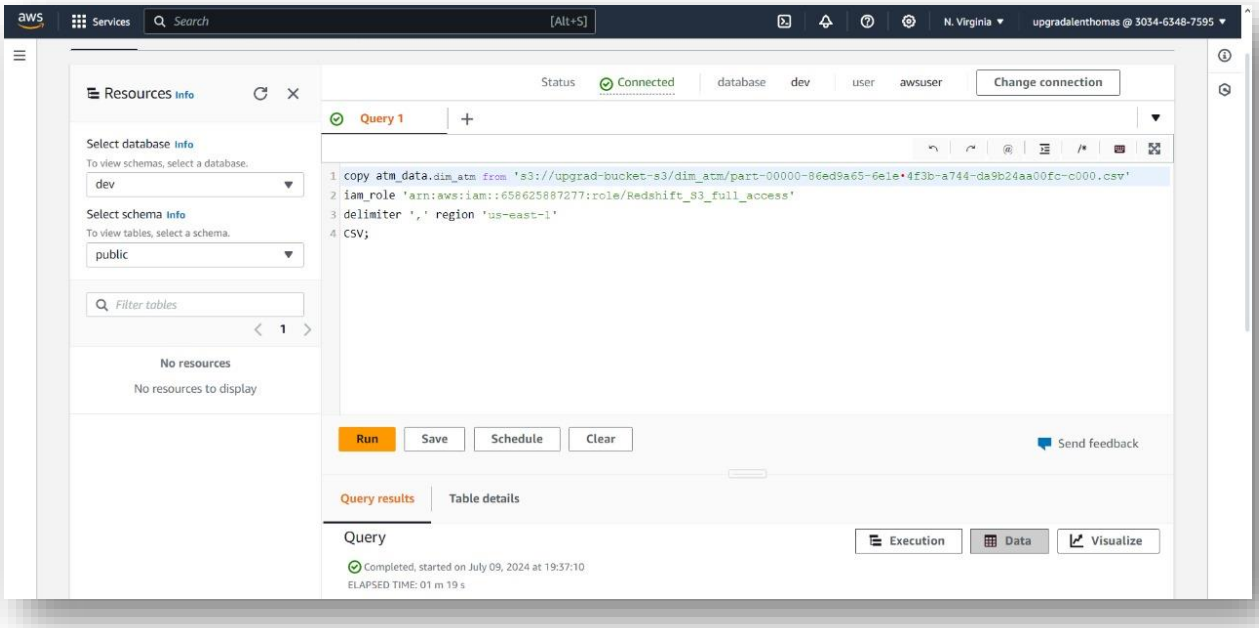
Loading data into a Redshift cluster from Amazon S3 bucket

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

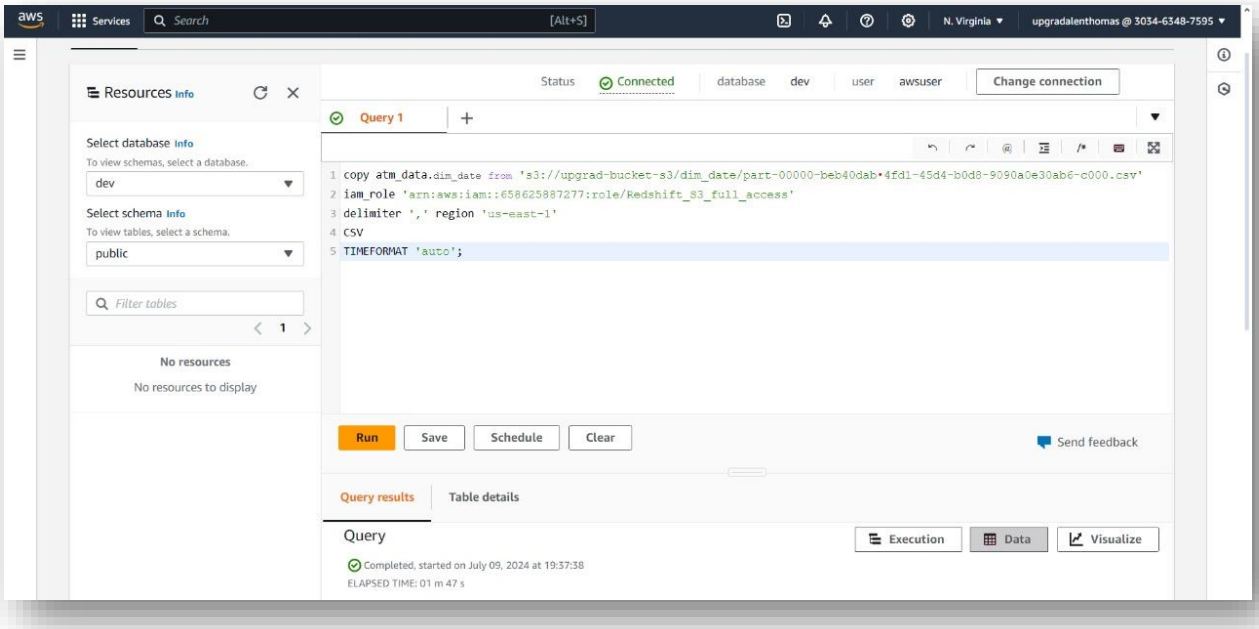
1. **Copying the data to dim_location table**
   copy atm_data.dim_location from 's3://upgrad-bucket-s3/dim_location/part-00000-81b80018-e175-4d8e-8d0c-525d4db67034-c000.csv' iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access' delimiter ',' region 'us-east-1' CSV;



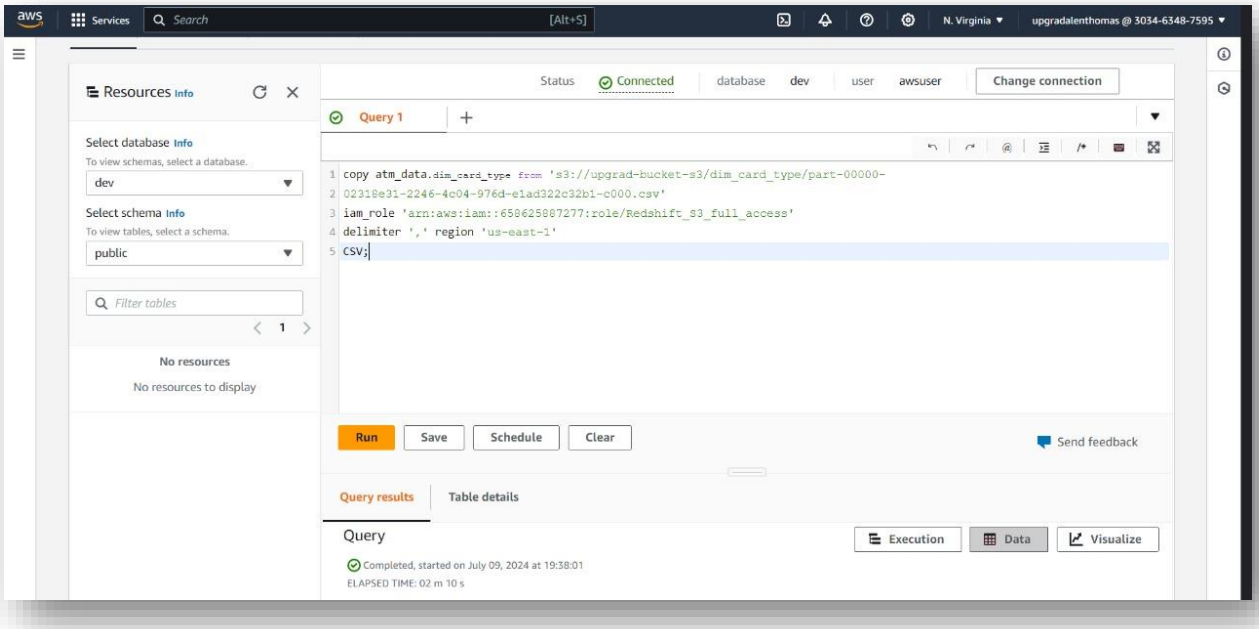
2. **Copying the data to dim_atm table**
   copy atm_data.dim_atm from 's3://upgrad-bucket-s3/dim_atm/part-00000-86ed9a65-6e1e4f3b-a744-da9b24aa00fc-c000.csv' iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access' delimiter ',' region 'us-east-1' CSV;



3. **Copying the data to dim_date table**
   copy atm_data.dim_date from 's3://upgrad-bucket-s3/dim_date/part-00000-beb40dab4fd1-45d4-b0d8-9090a0e30ab6-c000.csv' iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access' delimiter ',' region 'us-east-1' CSV TIMEFORMAT 'auto';



4. **Copying the data to dim_card_type table**
   copy atm_data.dim_card_type from 's3://upgrad-bucket-s3/dim_card_type/part-00000-02318e31-2246-4c04-976d-e1ad322c32b1-c000.csv' iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access' delimiter ',' region 'us-east-1' CSV;



5. **Copying the data to fact_atm_trans table**

copy atm_data.fact_atm_trans from 's3://upgrad-bucket-s3/fact_atm_trans/part-00000-fb72e768-f589-4caa-bd26-1bc6fcbbdb8a-c000.csv' iam_role 'arn:aws:iam::658625887277:role/Redshift_S3_full_access' delimiter ',' region 'us-east-1' CSV;