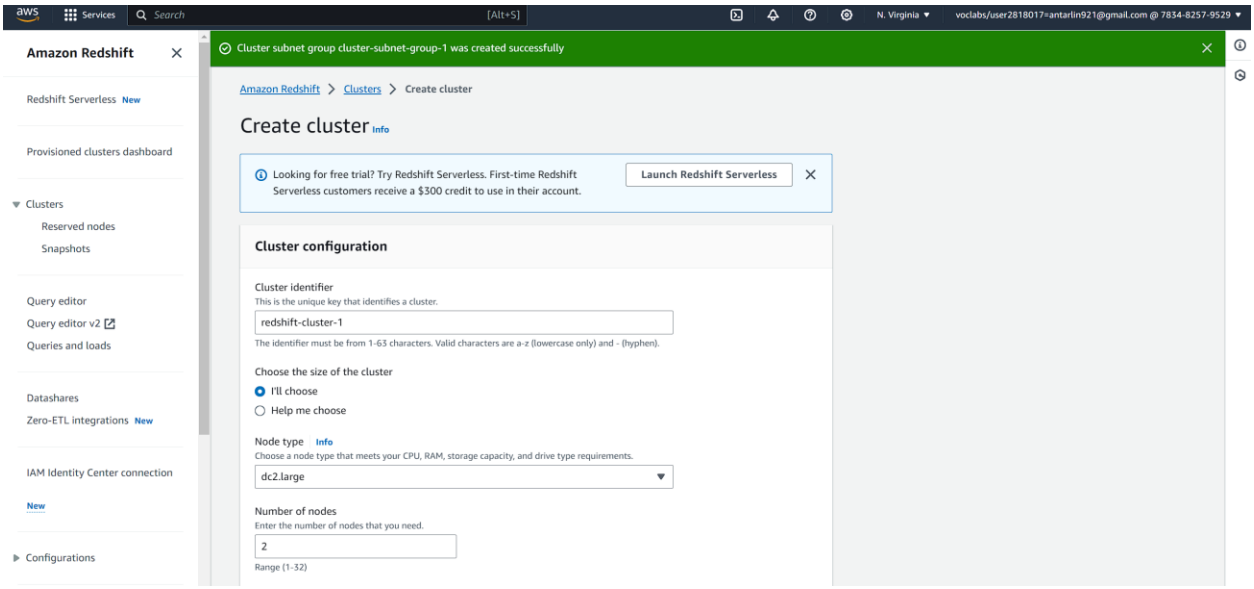
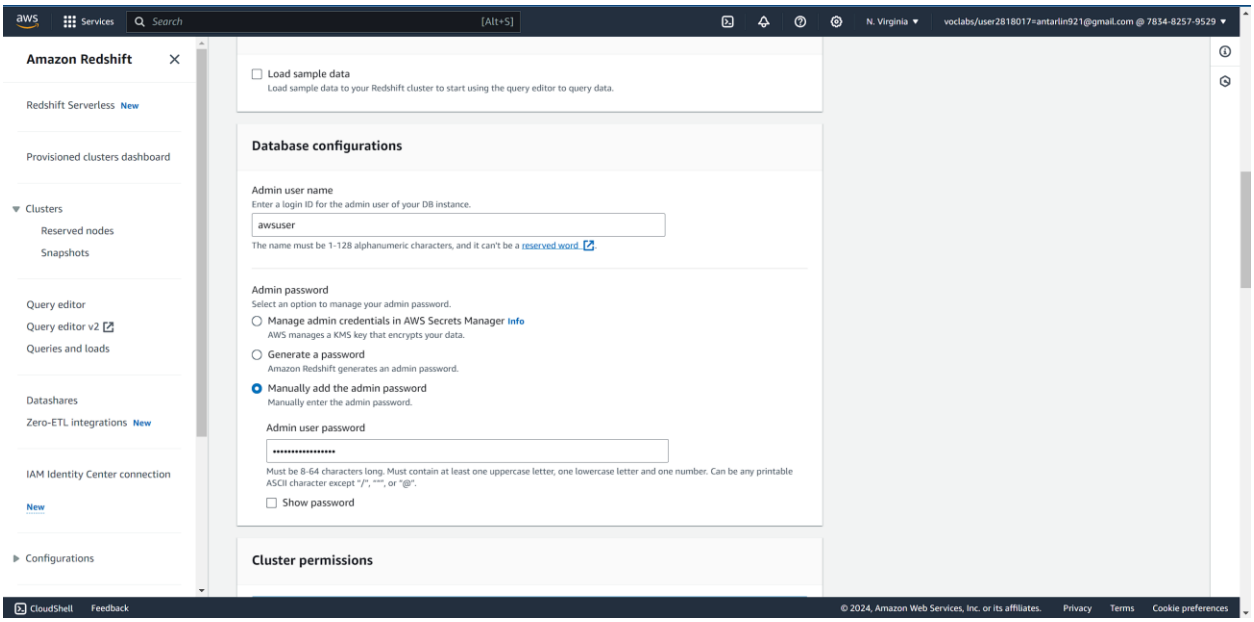
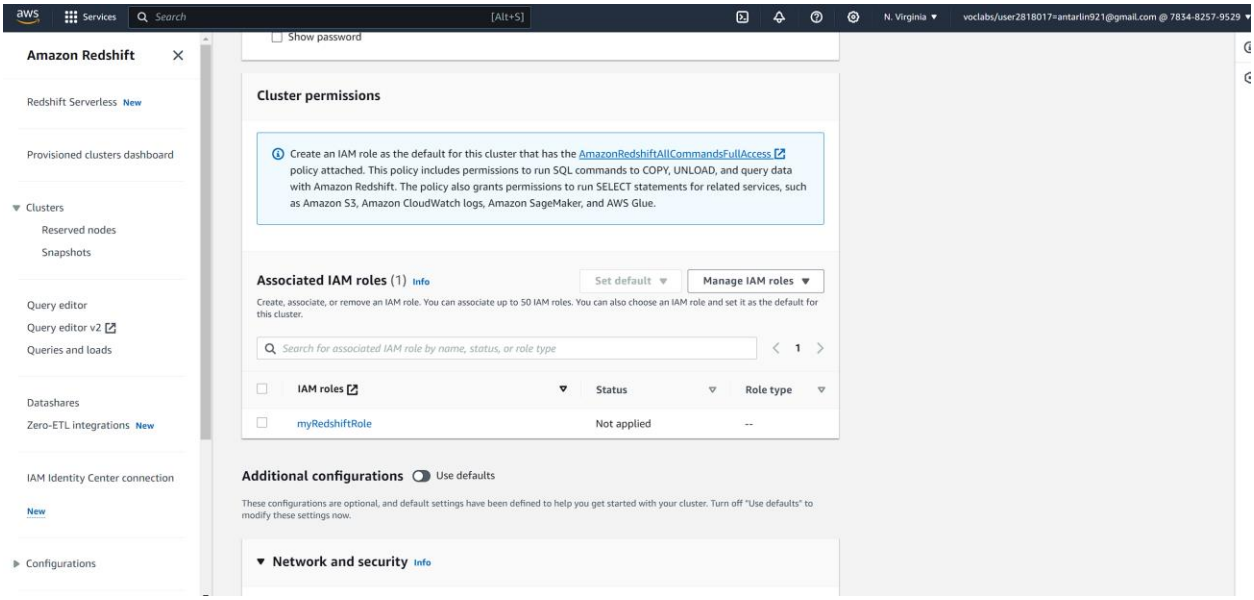
Creation of a Redshift Cluster

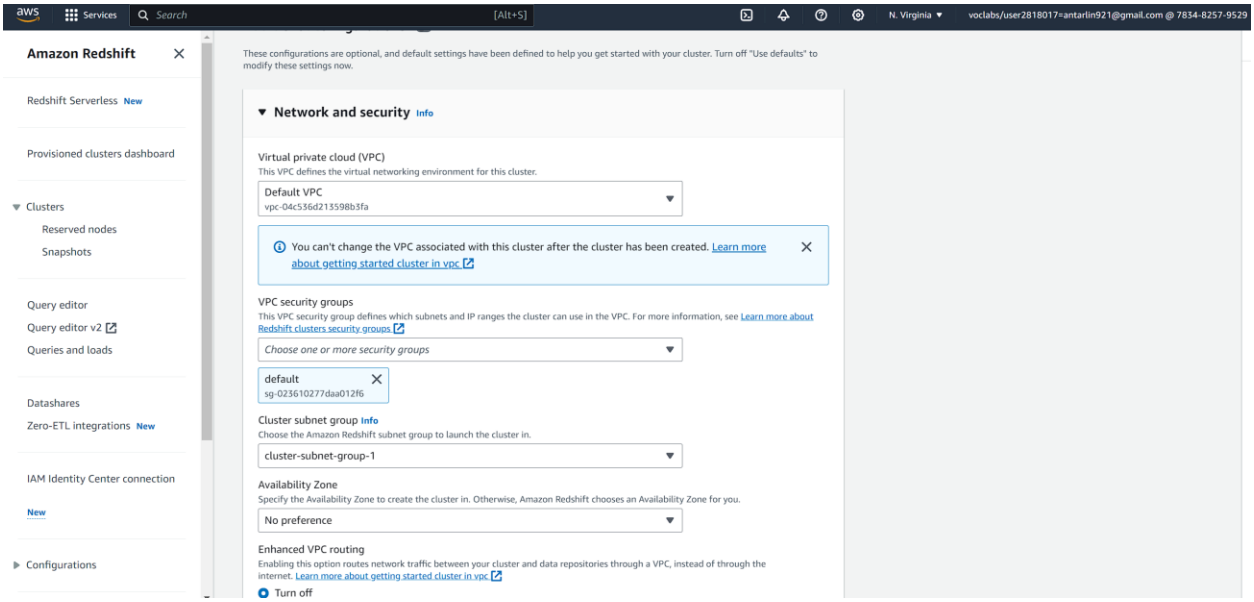
Screenshots of the configuration of the Redshift cluster that you have created:

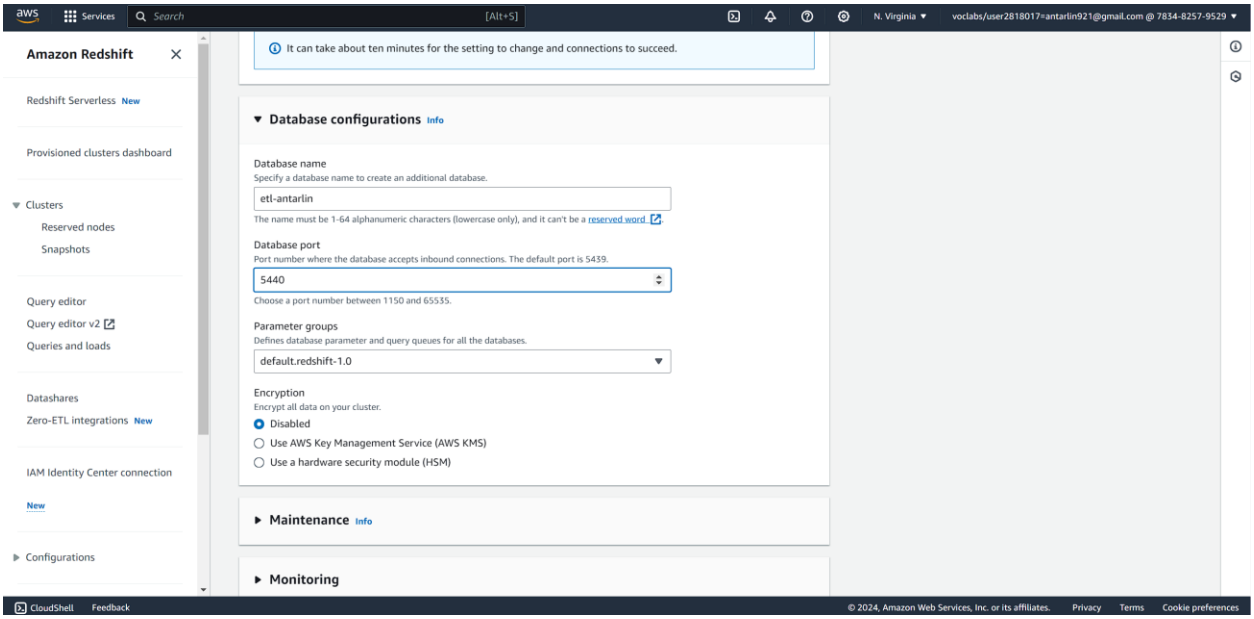
<Screenshot of the type of machine used along with number of nodes>

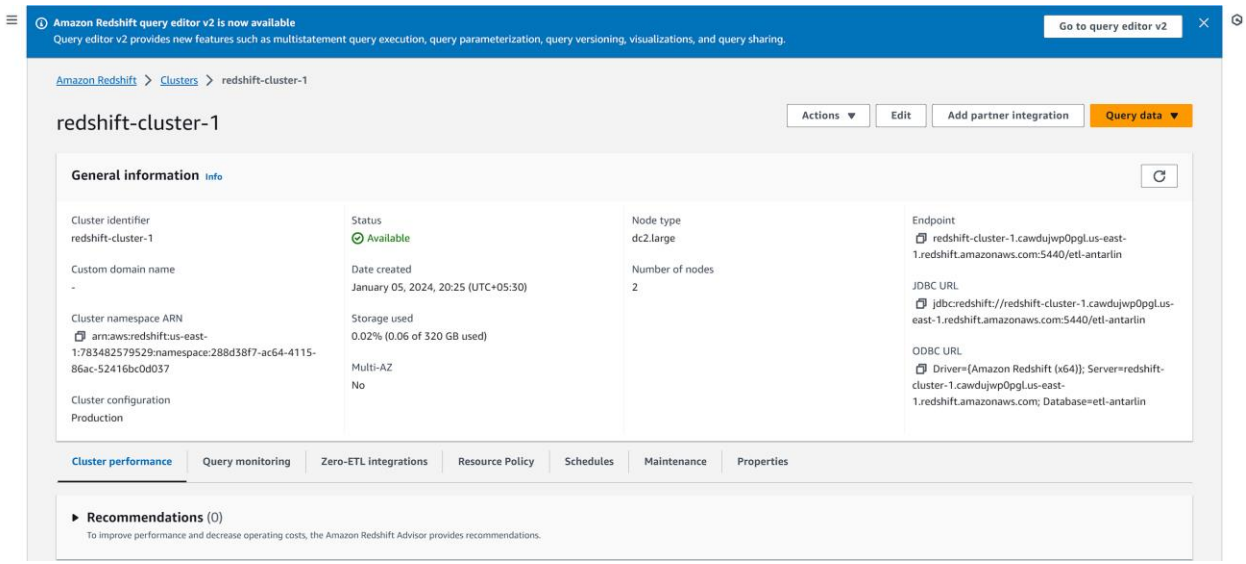






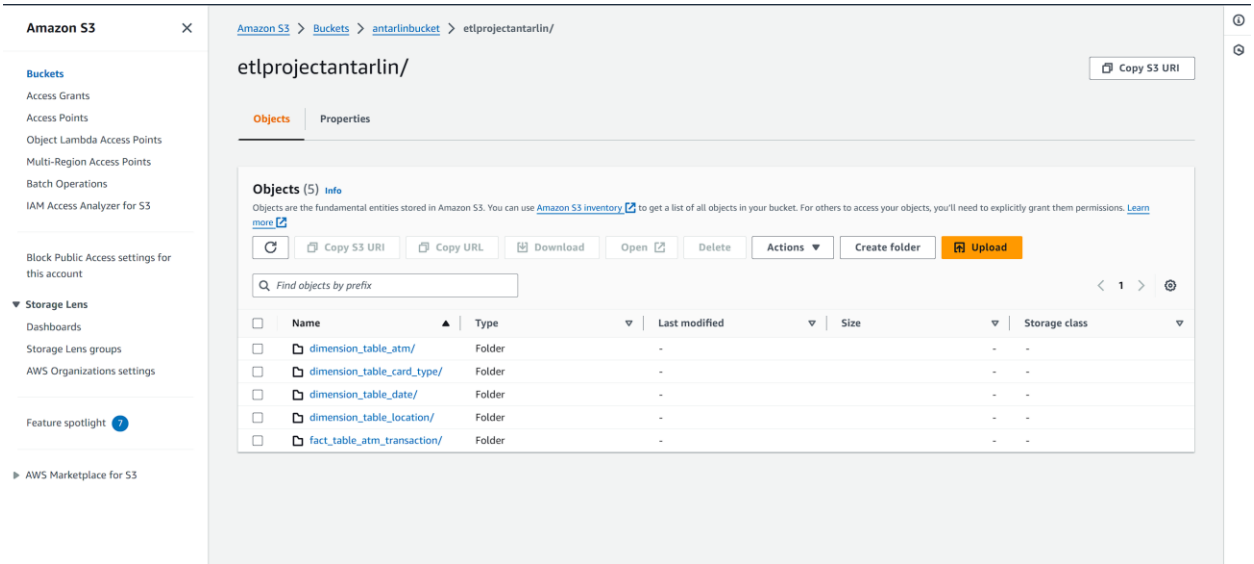




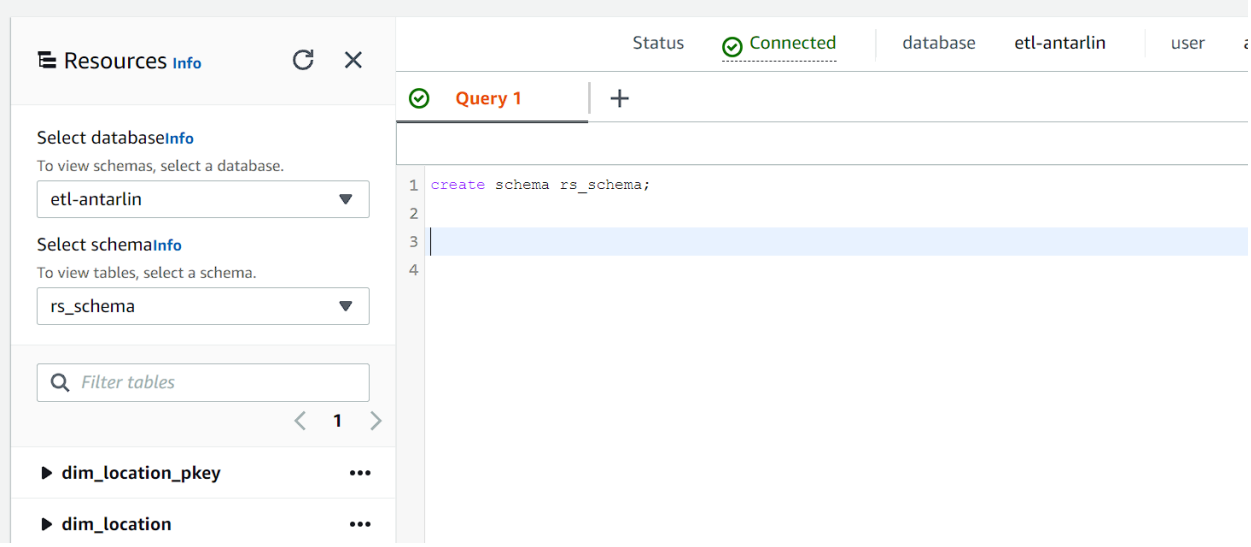


We can see redshift configurations above

Lets check the S3 folder with the data we imported from HDFS on EMR



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

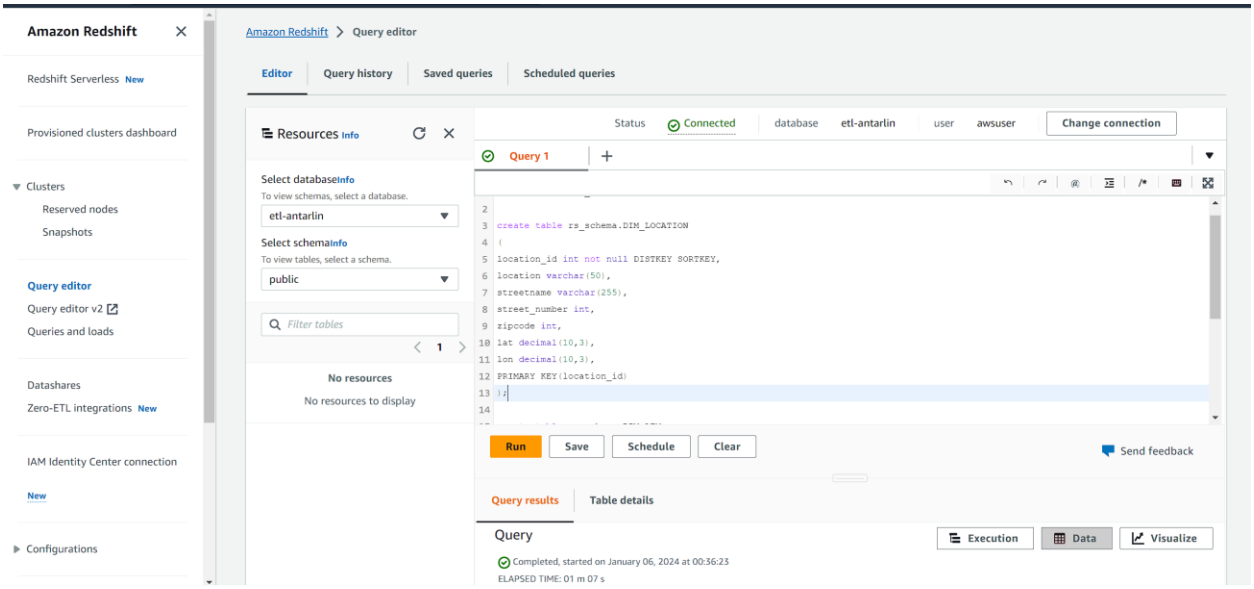


Create schema rs_schema;

- **Creating location dimension table**

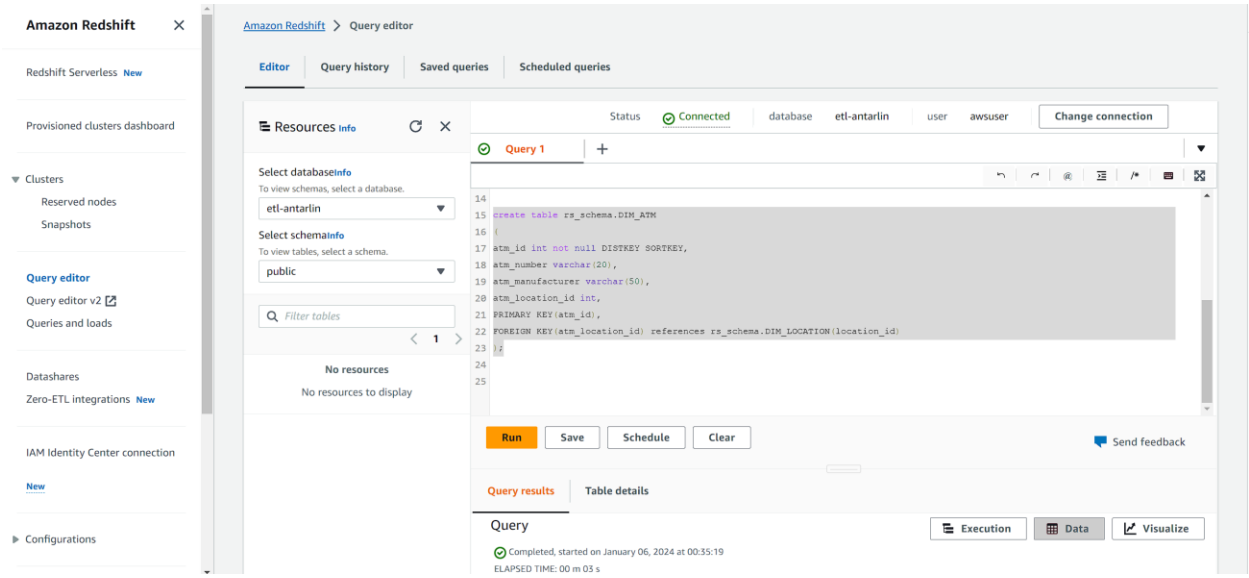
create table rs_schema.DIM_LOCATION
(
location_id int not null DISTKEY SORTKEY,

```
location varchar(50),
streetname varchar(255),
street_number int,
zipcode int,
lat decimal(10,3),
lon decimal(10,3),
PRIMARY KEY(location_id)
);
```



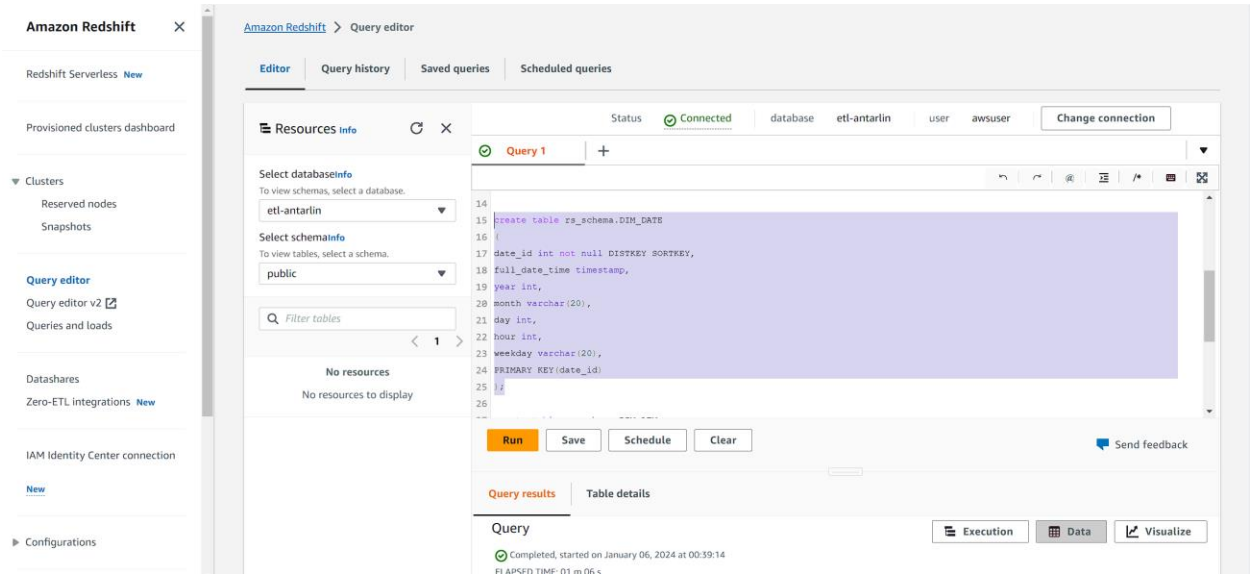
- **Creating atm dimension table**

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



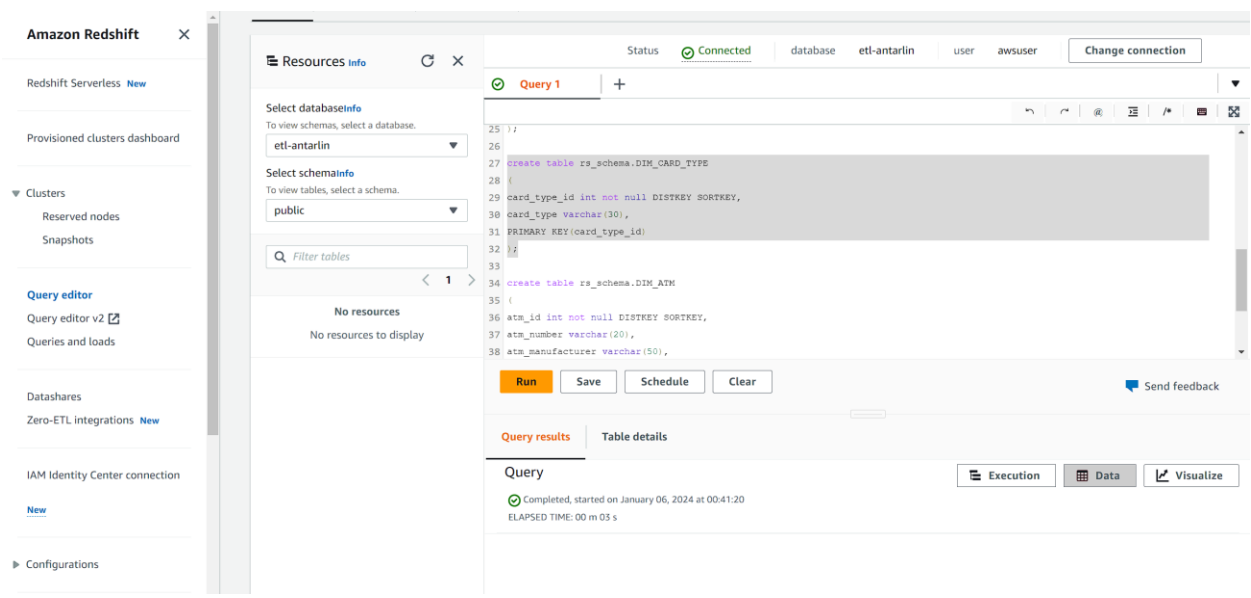
- **Creating date dimension table**

```
create table rs_schema.DIM_DATE
(
date_id int not null DISTKEY SORTKEY,
full_date_time varchar,
year int,
month varchar(20),
day int,
hour int,
weekday varchar(20),
PRIMARY KEY(date_id)
);
```



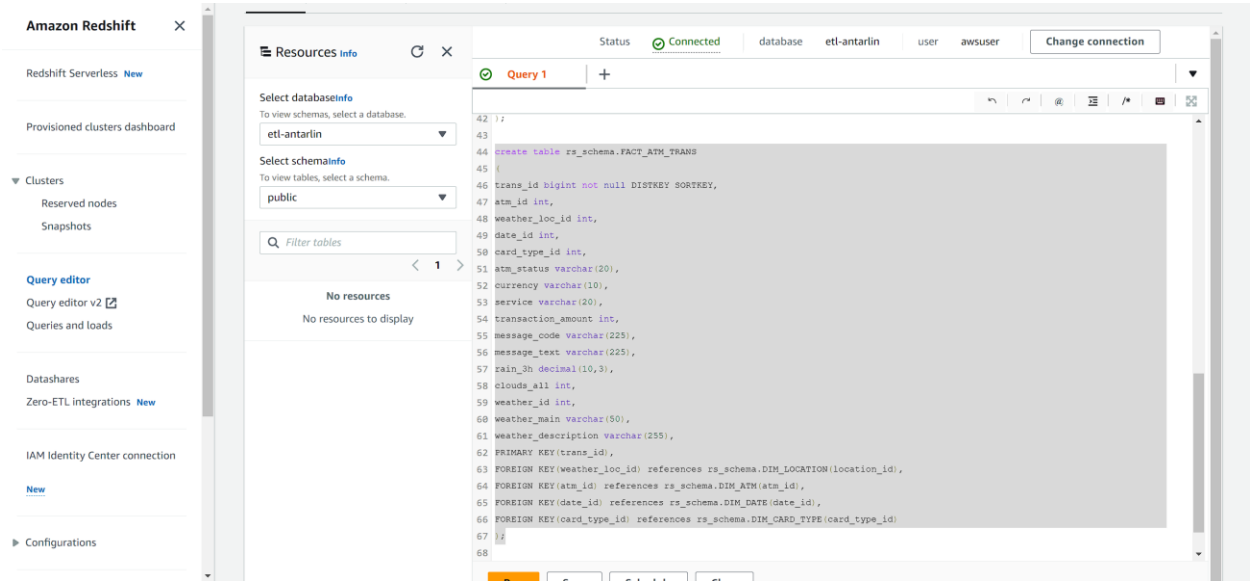
- **Creating card type dimension table**

```
create table rs_schema.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**

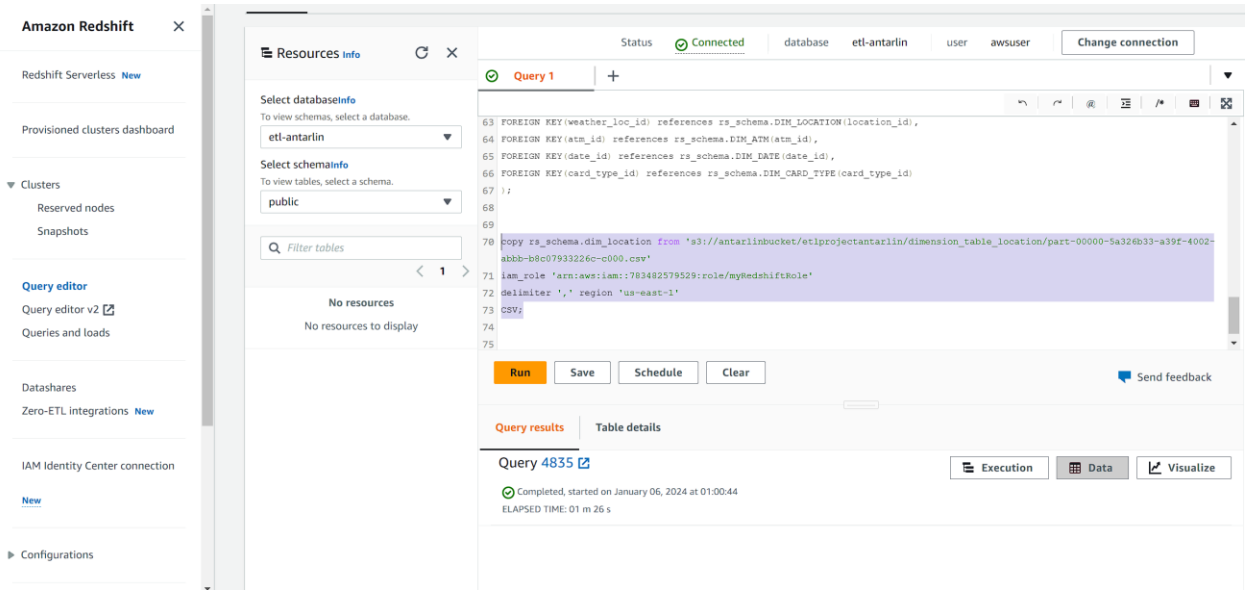
```
create table rs_schema.FACT_ATM_TRANS
(
trans_id bigint not null DISTKEY SORTKEY,
atm_id int,
weather_loc_id int,
date_id int,
card_type_id int,
atm_status varchar(20),
currency varchar(10),
service varchar(20),
transaction_amount int,
message_code varchar(225),
message_text varchar(225),
rain_3h 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 rs_schema.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references rs_schema.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references rs_schema.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references rs_schema.DIM_CARD_TYPE(card_type_id)
);
```



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

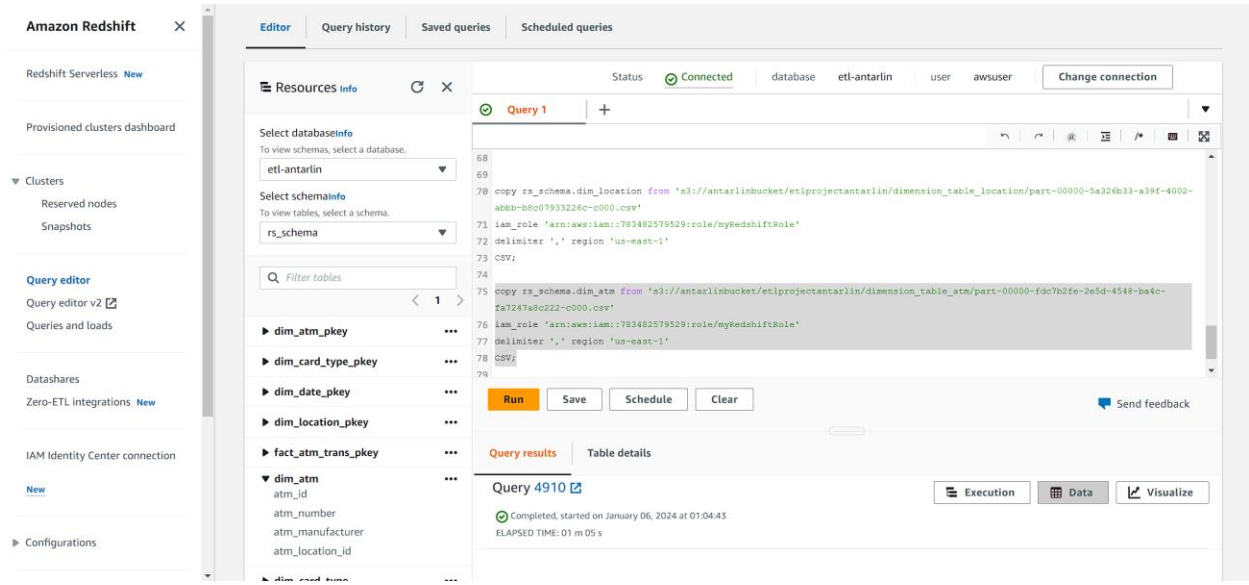
- Copying the data to dim_location table

```
copy rs_schema.dim_location from
's3://antarlinbucket/etlprojectantarlin/dimension_table_location/part-00000-5a326b33-
a39f-4002-abbb-b8c07933226c-c000.csv'
iam_role 'arn:aws:iam::783482579529:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



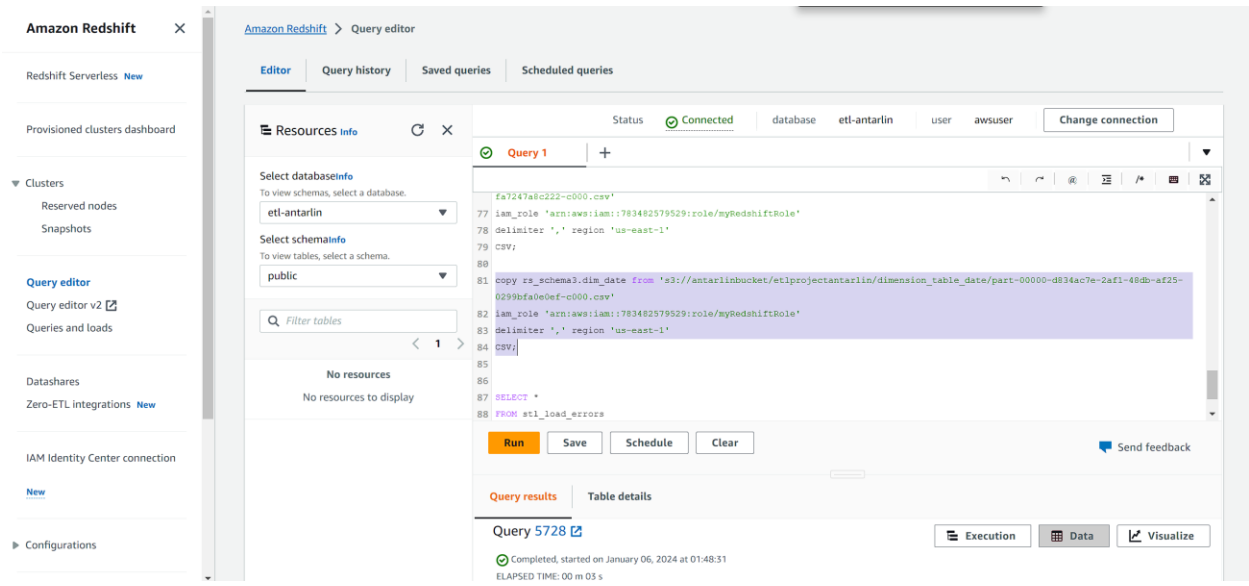
- **Copying the data to dim_atm table**

```
copy rs_schema.dim_atm from 's3://antarlinbucket/etlprojectantarlin/dimension_table_atm/part-
00000-fdc7b2fe-2e5d-4548-ba4c-fa7247a8c222-c000.csv'
iam_role 'arn:aws:iam::783482579529:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



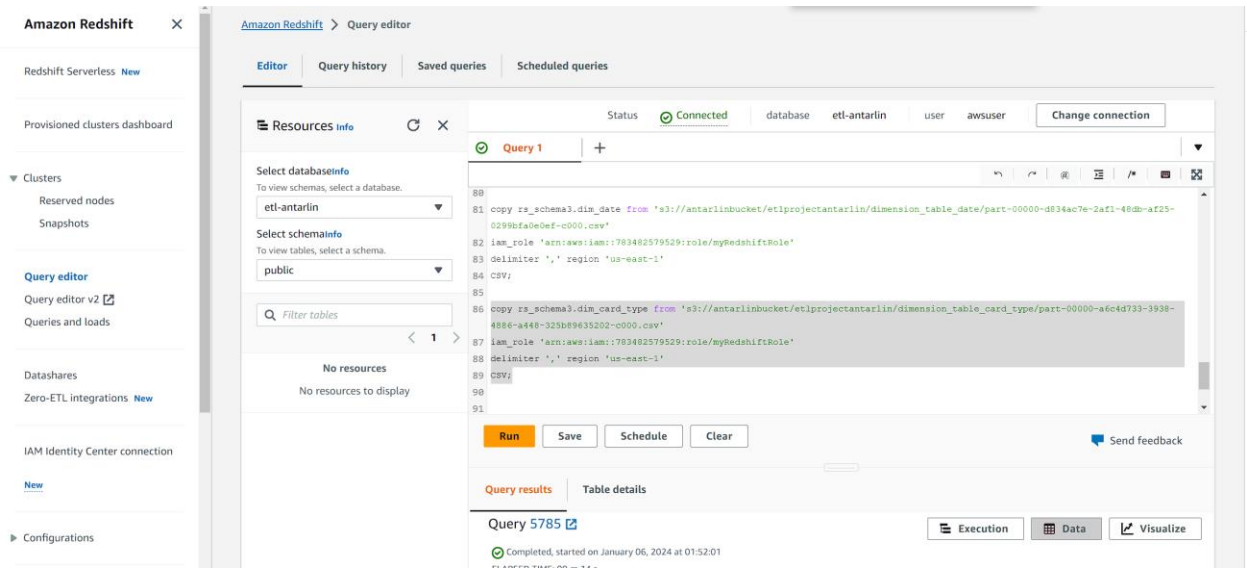
- **Copying the data to dim_date table**

copy rs_schema.dim_date from
's3://antarlinbucket/etlprojectantarlin/dimension_table_date/part-00000-d834ac7e-2af1-48db-af25-0299bfa0e0ef-c000.csv'
iam_role 'arn:aws:iam::783482579529:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



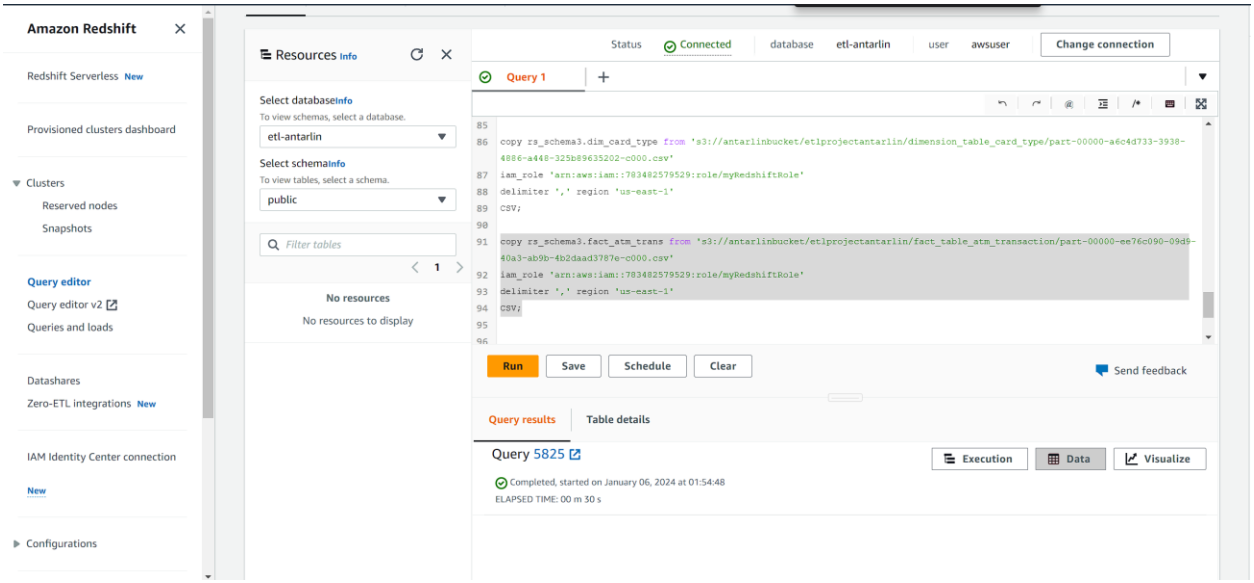
- **Copying the data to dim_card_type table**

copy rs_schema.dim_card_type from
's3://antarlinbucket/etlprojectantarlin/dimension_table_card_type/part-00000-a6c4d733-3938-4886-a448-325b89635202-c000.csv'
iam_role 'arn:aws:iam::783482579529:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



- **Copying the data to fact_atm_trans table**

copy rs_schema.fact_atm_trans from
's3://antarlinbucket/etlprojectantarlin/fact_table_atm_transaction/part-00000-ee76c090-09d9-40a3-ab9b-4b2daad3787e-c000.csv'
iam_role 'arn:aws:iam::783482579529:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;



----------The End----------