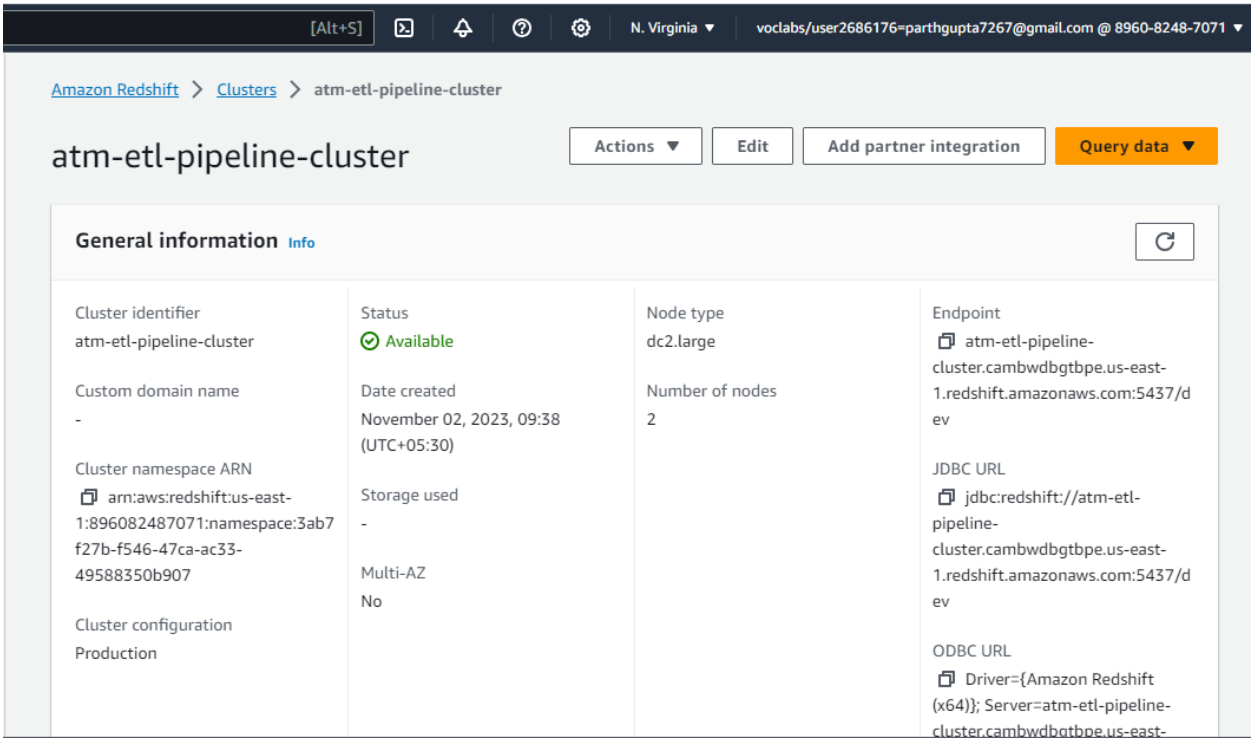
Creation of a Redshift Cluster

Screenshots of Redshift Cluster :

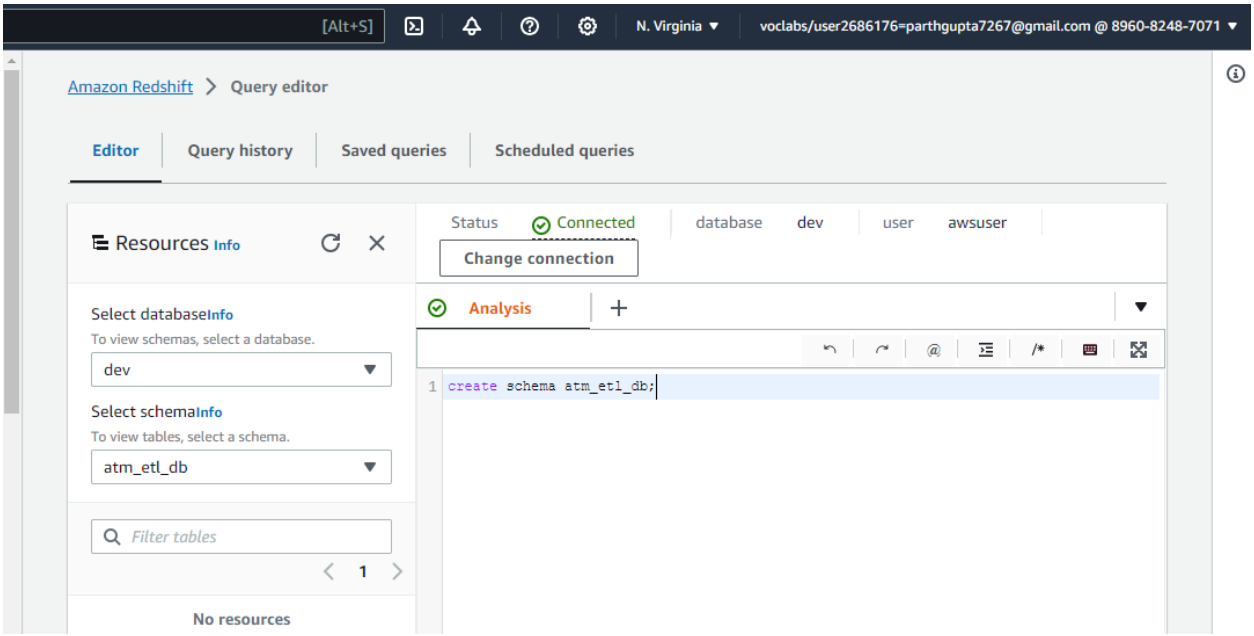


Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Queries to create the various dimension and fact tables:

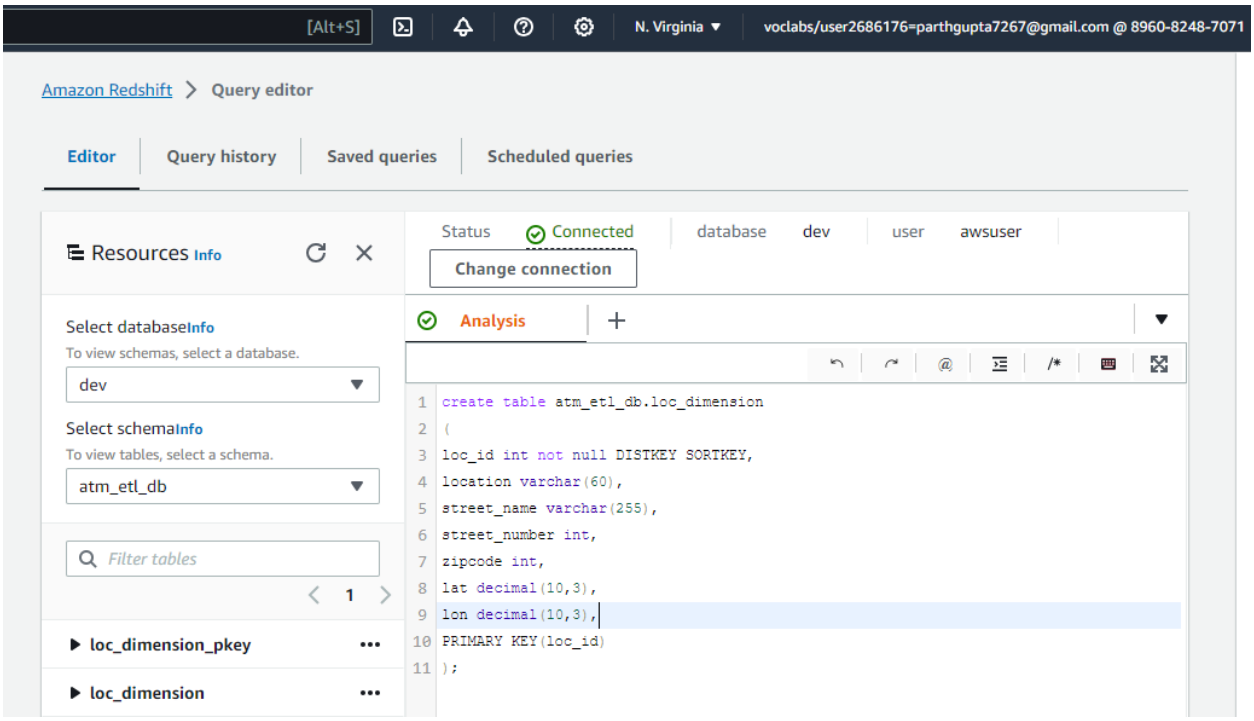
1) **Creation of Schema:**

create schema atm_etl_db;



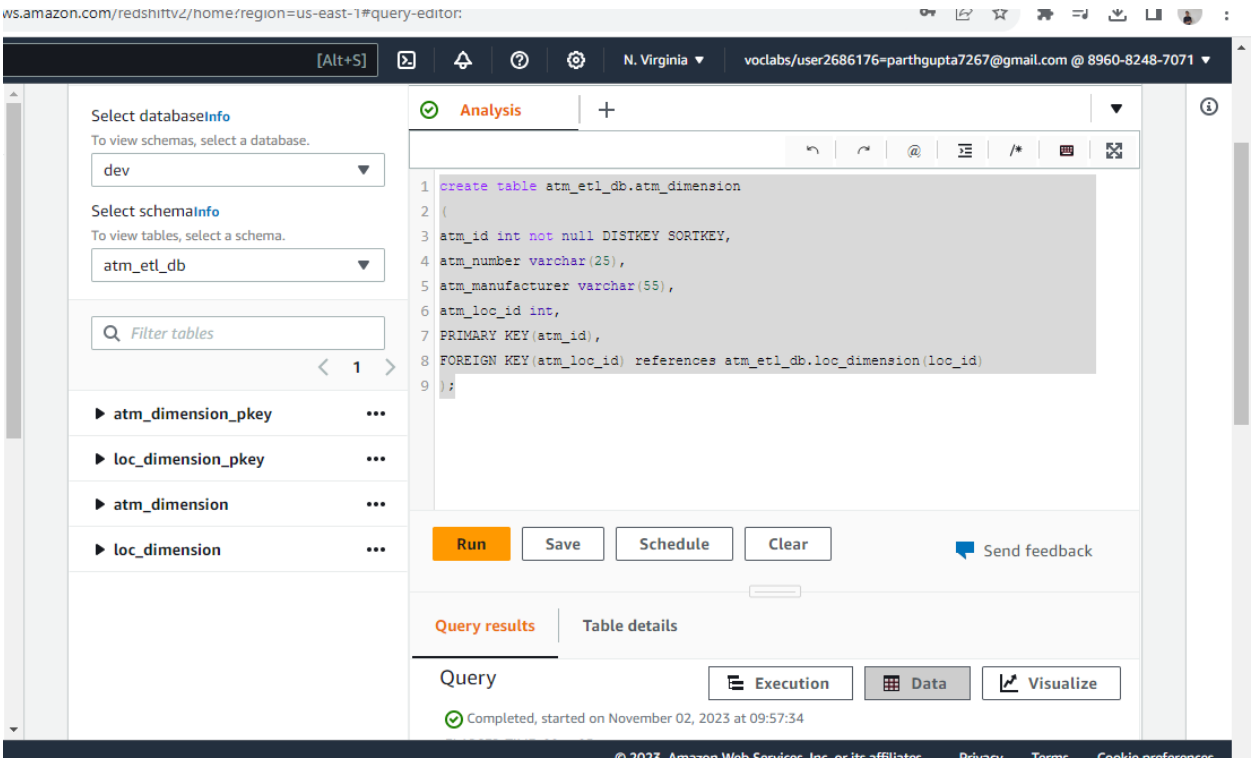
2) Creating location dimension table:

```
create table atm_etl_db.loc_dimension
(
loc_id int not null DISTKEY SORTKEY,
location varchar(60),
street_name varchar(255),
street_number int,
zipcode int,
lat decimal(10,3),
lon decimal(10,3),
PRIMARY KEY(loc_id)
);
```



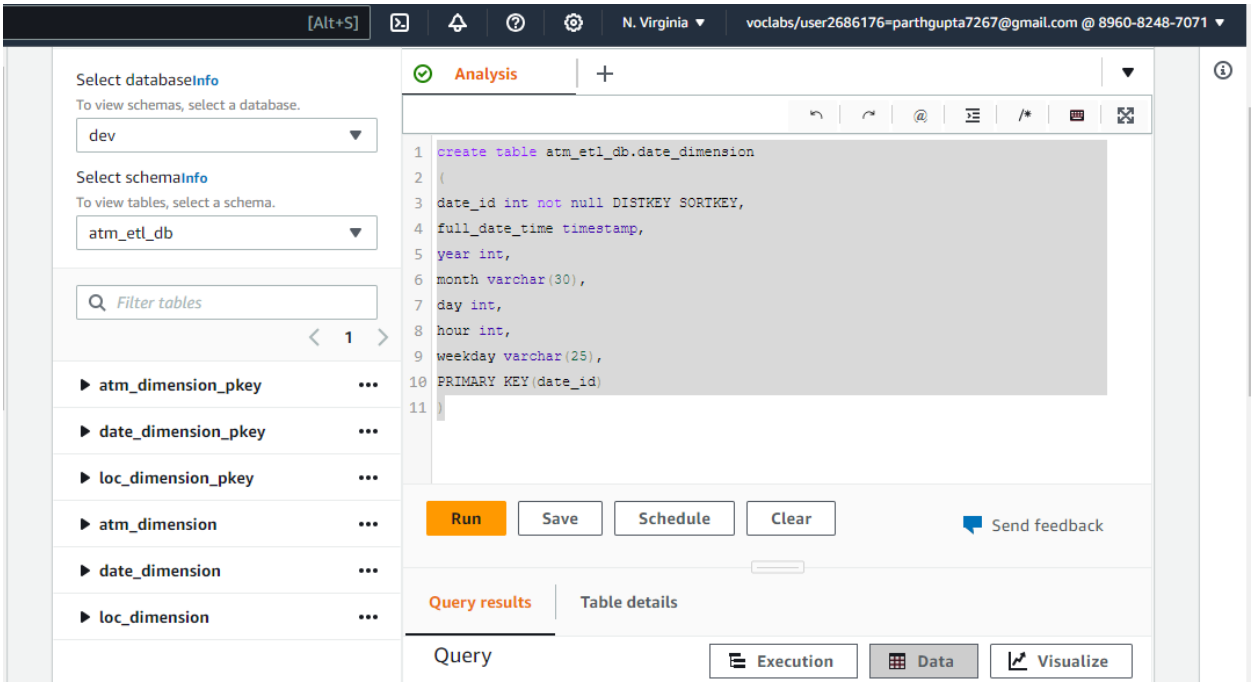
3) Creating ATM dimension table:

```
create table atm_etl_db.atm_dimension
(
atm_id int not null DISTKEY SORTKEY,
atm_number varchar(25),
atm_manufacturer varchar(55),
atm_loc_id int,
PRIMARY KEY(atm_id),
FOREIGN KEY(atm_loc_id) references atm_etl_db.loc_dimension(loc_id)
);
```



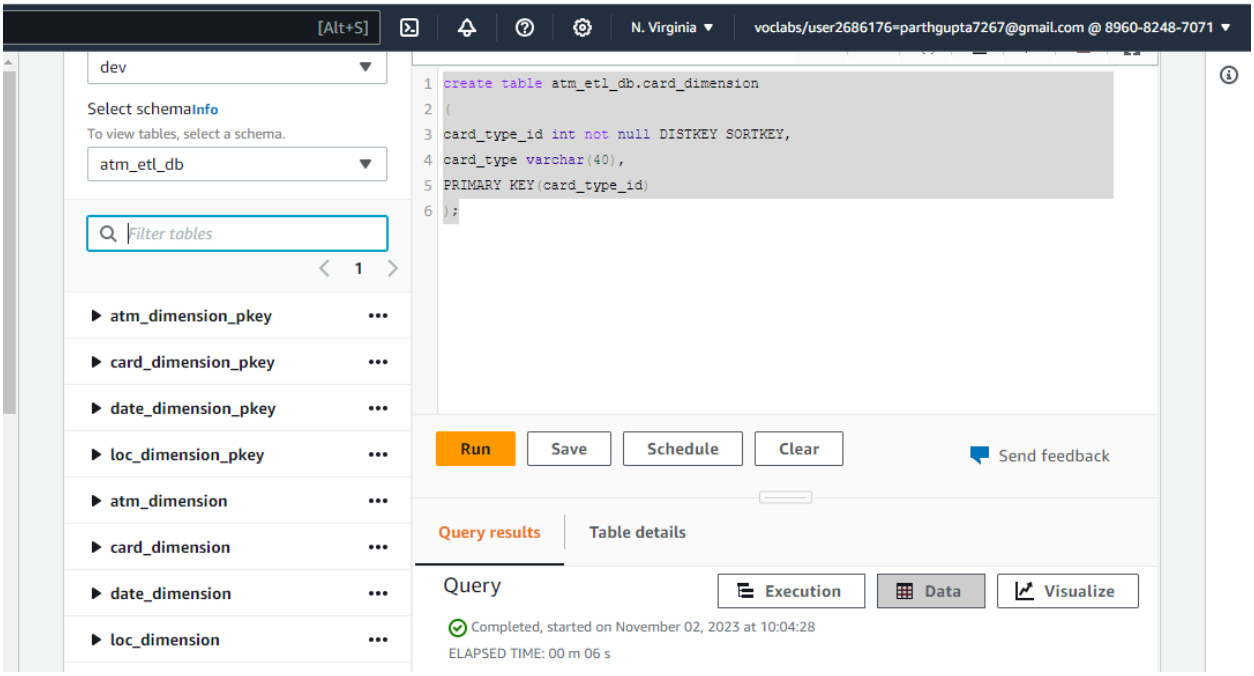
4) Creating date dimension table:

```
create table atm_etl_db.date_dimension
(
date_id int not null DISTKEY SORTKEY,
full_date_time timestamp,
year int,
month varchar(30),
day int,
hour int,
weekday varchar(25),
PRIMARY KEY(date_id)
);
```



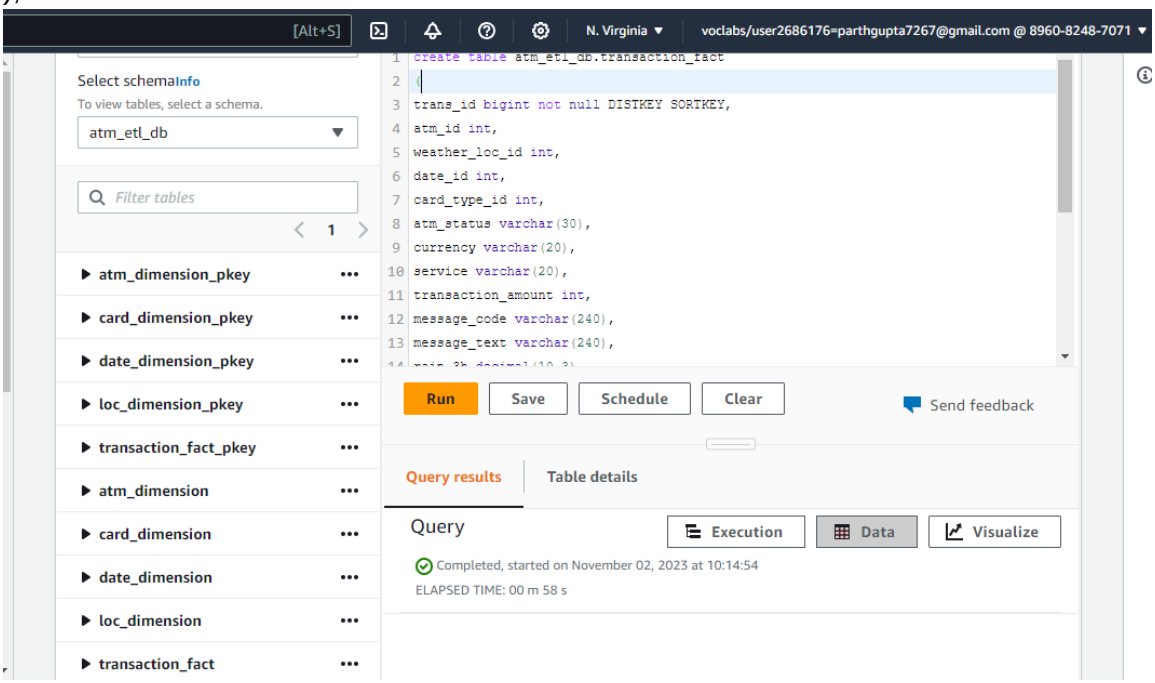
5) Creating card dimension table:

```
create table atm_etl_db.card_dimension
(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(40),
PRIMARY KEY(card_type_id)
);
```



6) Creating Transaction fact table:

```
create table atm_etl_db.transaction_fact
(
trans_id bigint not null DISTKEY SORTKEY,
atm_id int,
weather_loc_id int,
date_id int,
card_type_id int,
atm_status varchar(30),
currency varchar(20),
service varchar(20),
transaction_amount int,
message_code varchar(240),
message_text varchar(240),
rain_3h decimal(10,3),
clouds_all int,
weather_id int,
weather_main varchar(70),
weather_description varchar(255),
PRIMARY KEY(trans_id),
FOREIGN KEY(weather_loc_id) references atm_etl_db.loc_dimension(loc_id),
FOREIGN KEY(atm_id) references atm_etl_db.atm_dimension(atm_id),
FOREIGN KEY(date_id) references atm_etl_db.date_dimension(date_id),
FOREIGN KEY(card_type_id) references atm_etl_db.card_dimension(card_type_id)
);
```

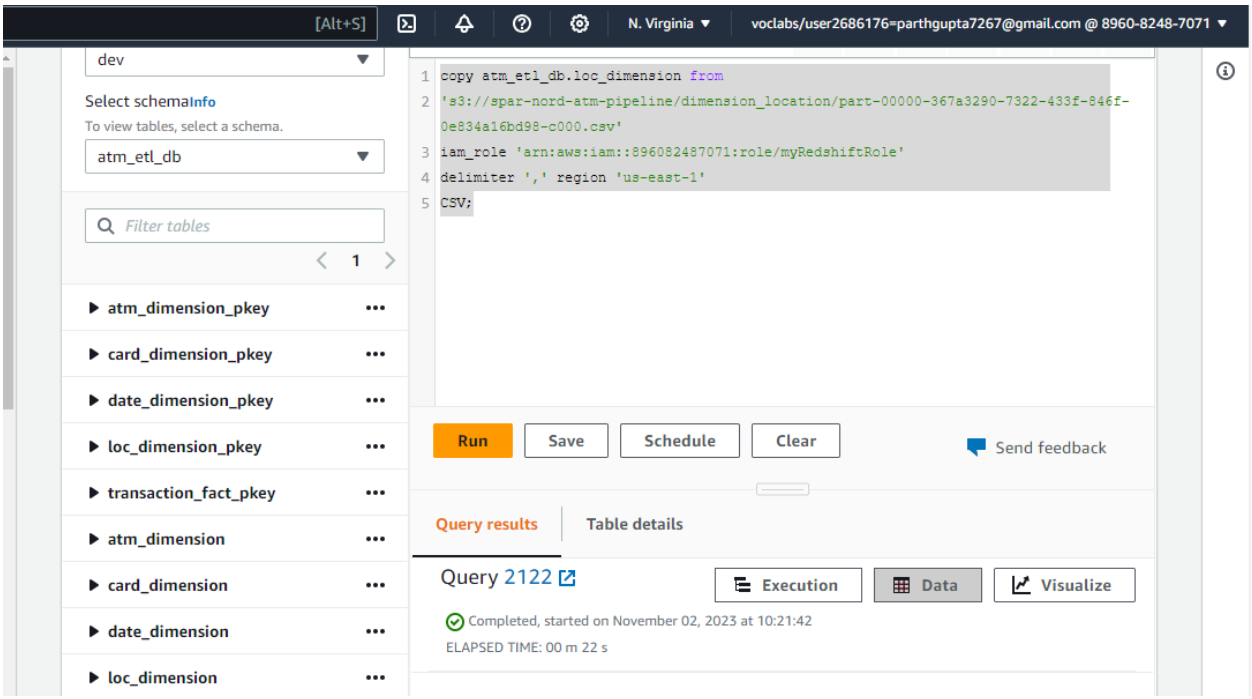


Loading data into a Redshift cluster from Amazon S3 bucket

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

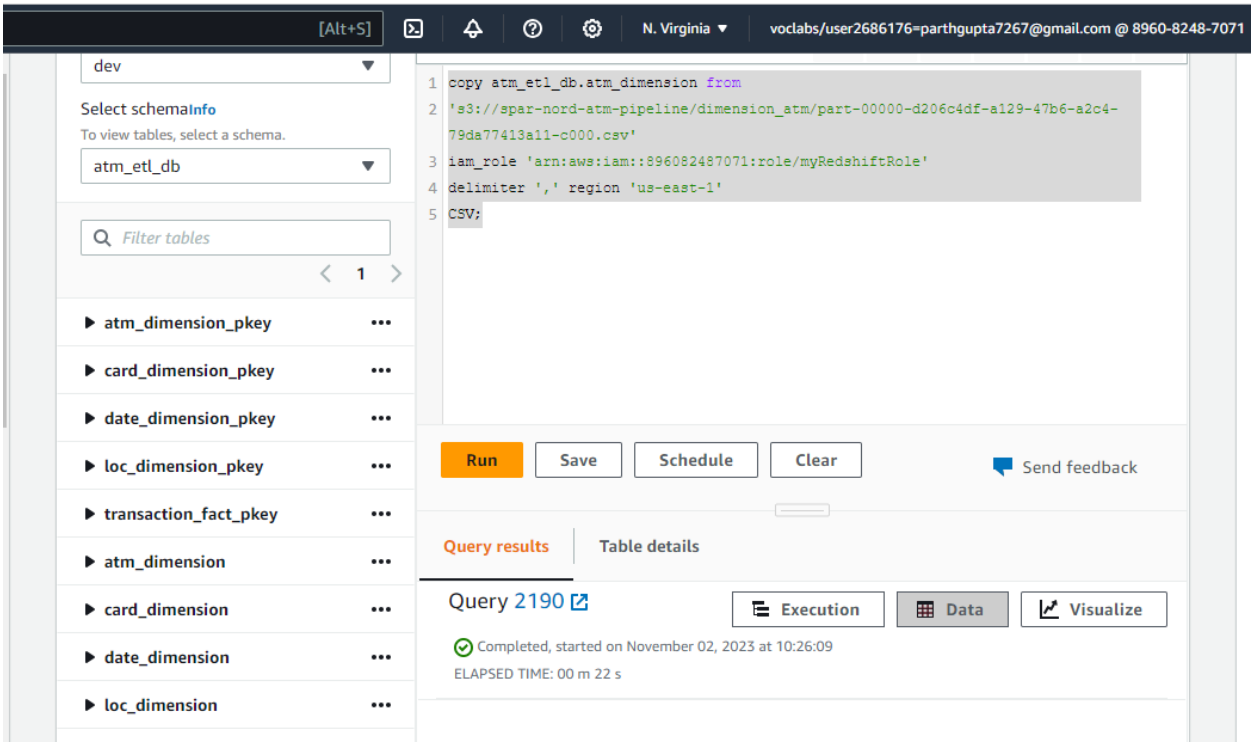
1) **Copying Location Dimension Table data:**

```
copy atm_etl_db.loc_dimension from
's3://spar-nord-atm-pipeline/dimension_location/part-00000-367a3290-7322-433f-846f-0e834a16bd98-c000.csv'
iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



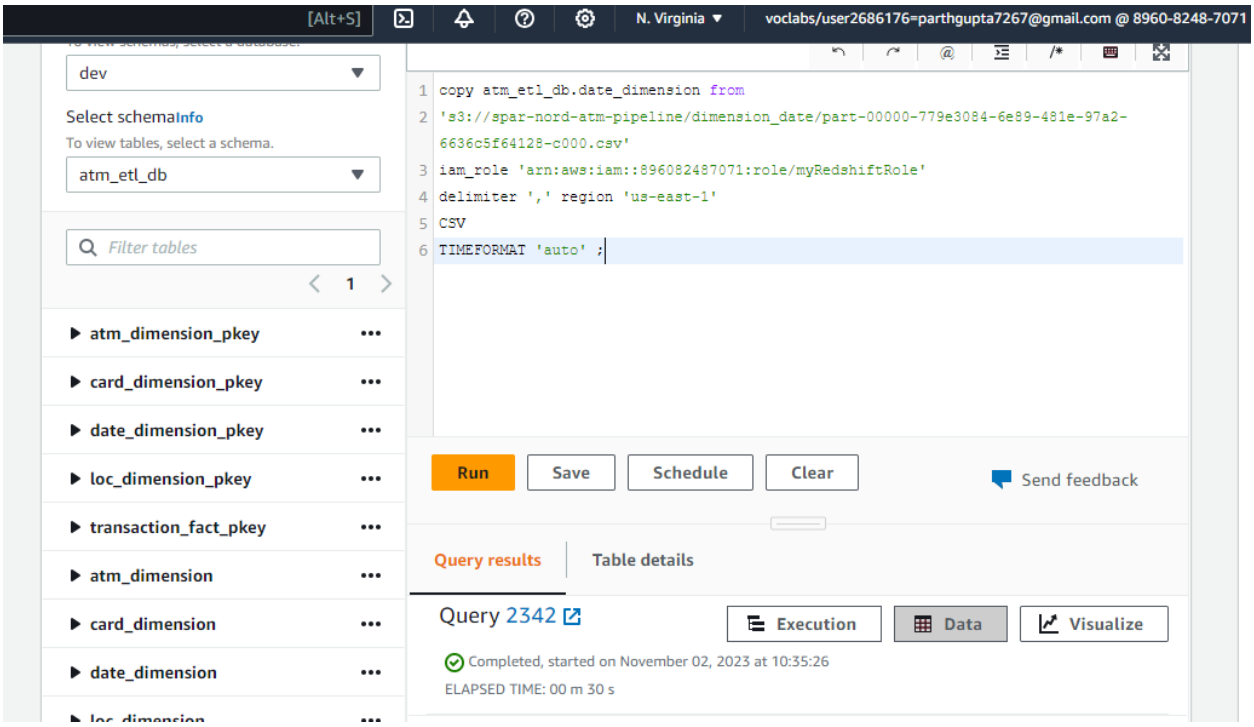
2) **Copying ATM Dimension Table data:**

```
copy atm_etl_db.atm_dimension from
's3://spar-nord-atm-pipeline/dimension_atm/part-00000-d206c4df-a129-47b6-a2c4-
79da77413a11-c000.csv'
iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



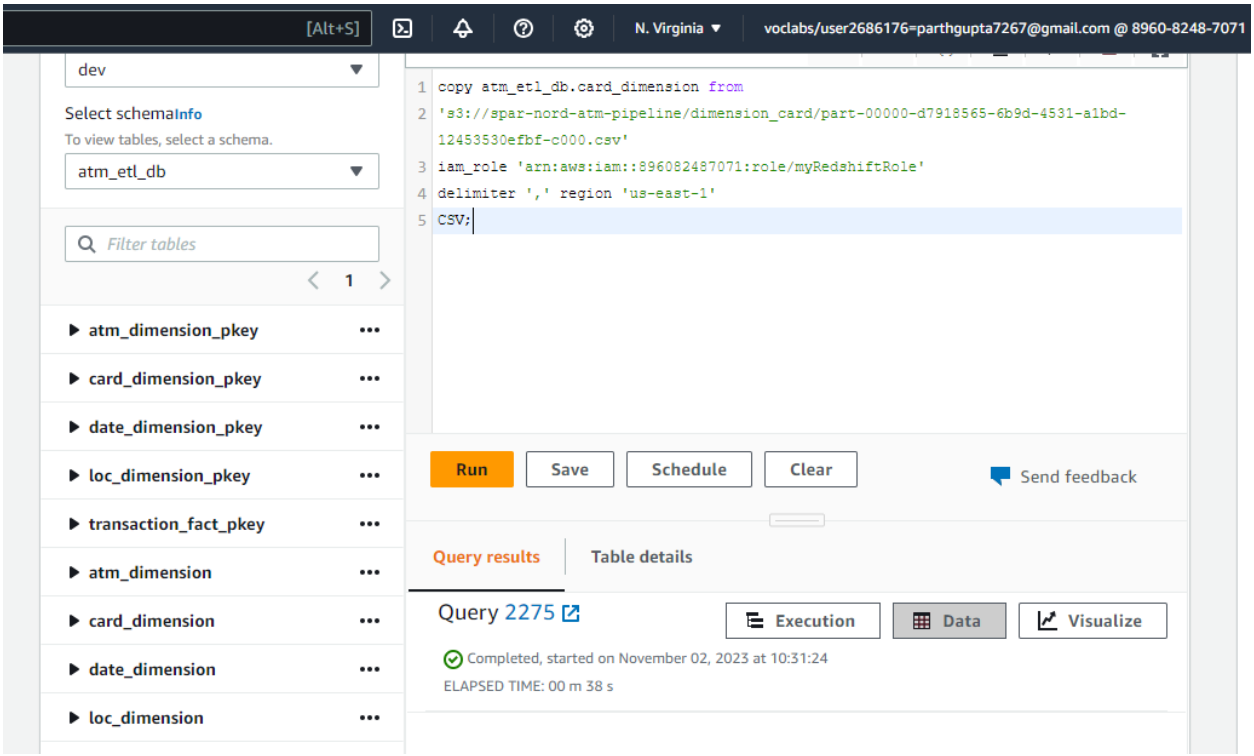
3) **Copying Date Dimension Table data:**

```
copy atm_etl_db.date_dimension from
's3://spar-nord-atm-pipeline/dimension_date/part-00000-779e3084-6e89-481e-97a2-
6636c5f64128-c000.csv'
iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto' ;
```



4) **Copying Card Dimension Table data:**

```
copy atm_etl_db.card_dimension from
's3://spar-nord-atm-pipeline/dimension_card/part-00000-d7918565-6b9d-4531-a1bd-12453530efbf-c000.csv'
iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



5) **Copying Transaction Fact Table data:**

```
copy atm_etl_db.transaction_fact from
's3://spar-nord-atm-pipeline/fact_transaction/part-00000-122bc3a0-af75-41bb-bcb8-92d363b20067-c000.csv'
iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```