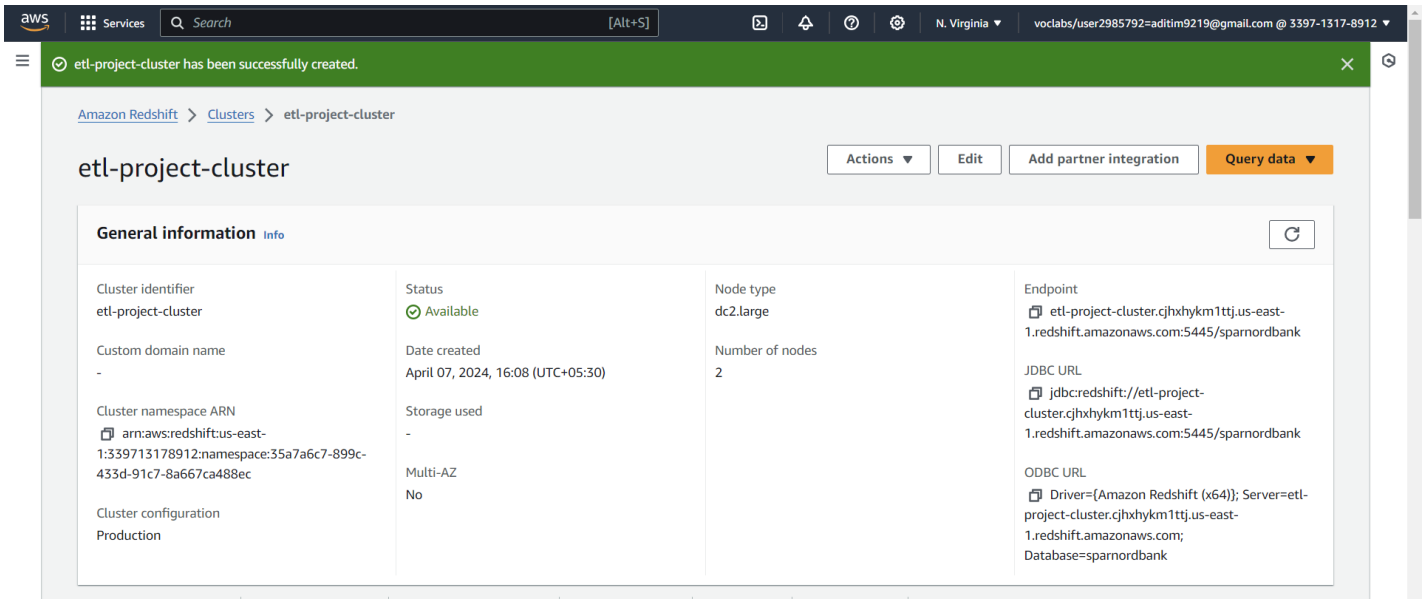
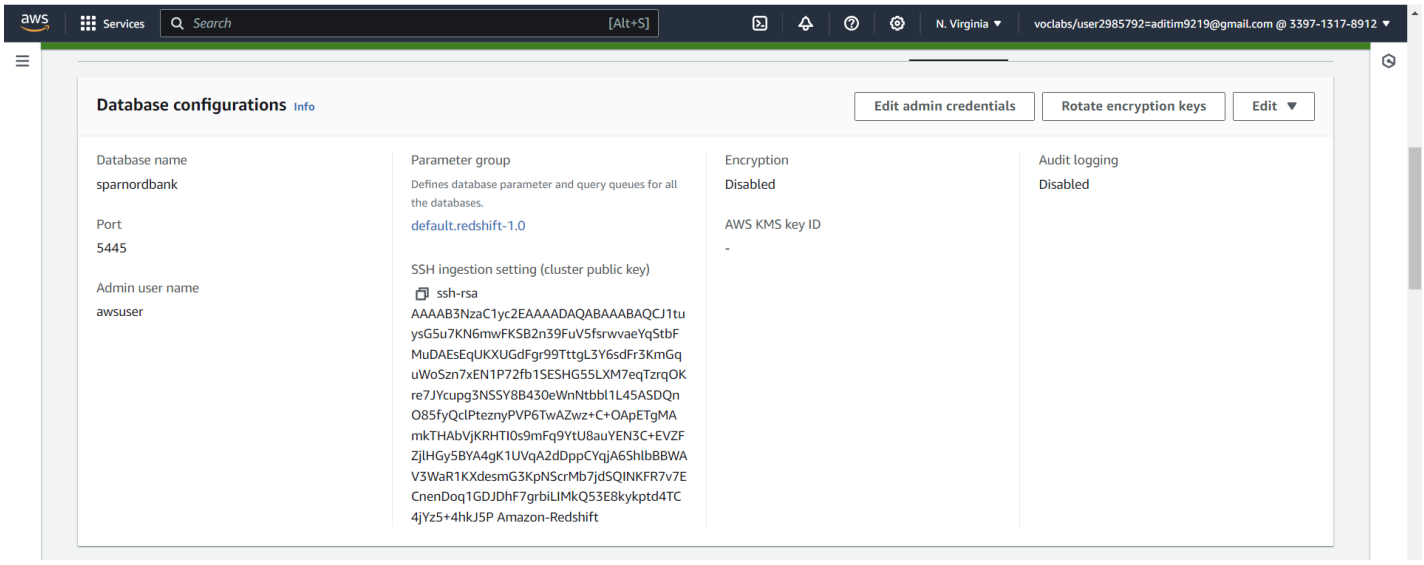
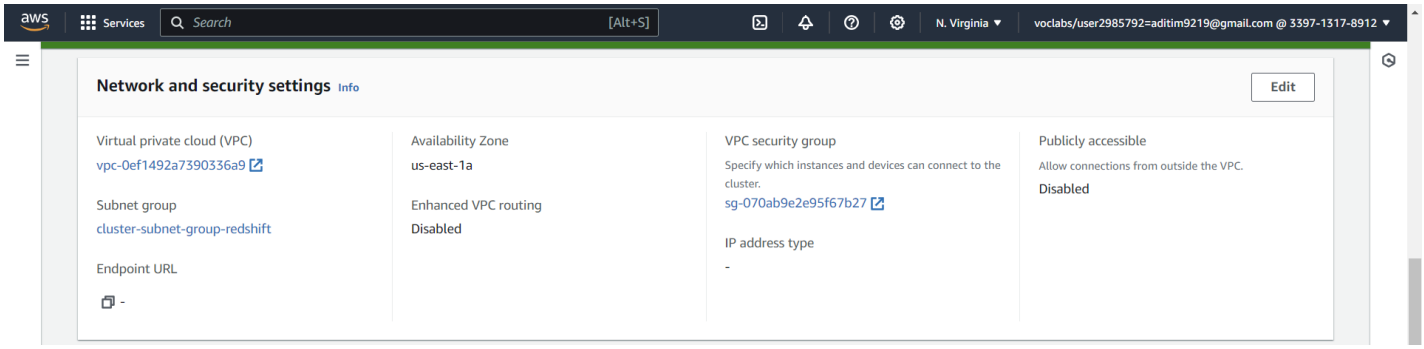
Creation of a Redshift Cluster

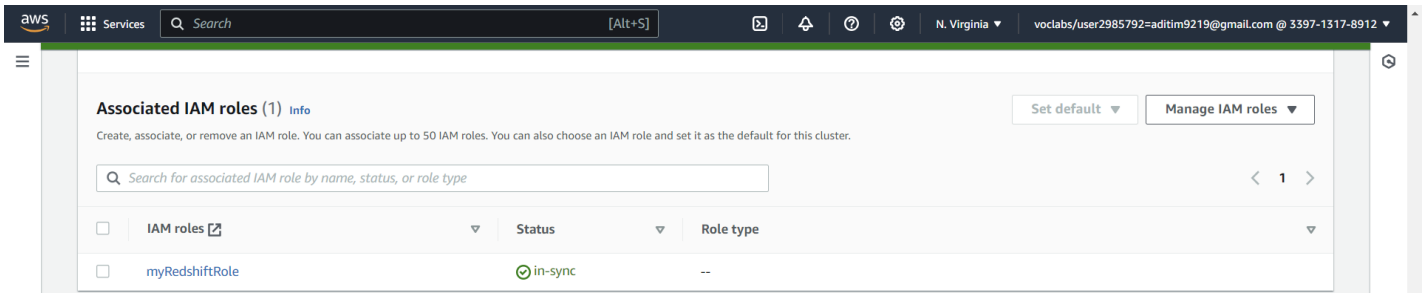
Screenshots of the configuration of the Redshift cluster that you have created:





© Copyright. upGrad Education Pvt. Ltd. All rights reserved



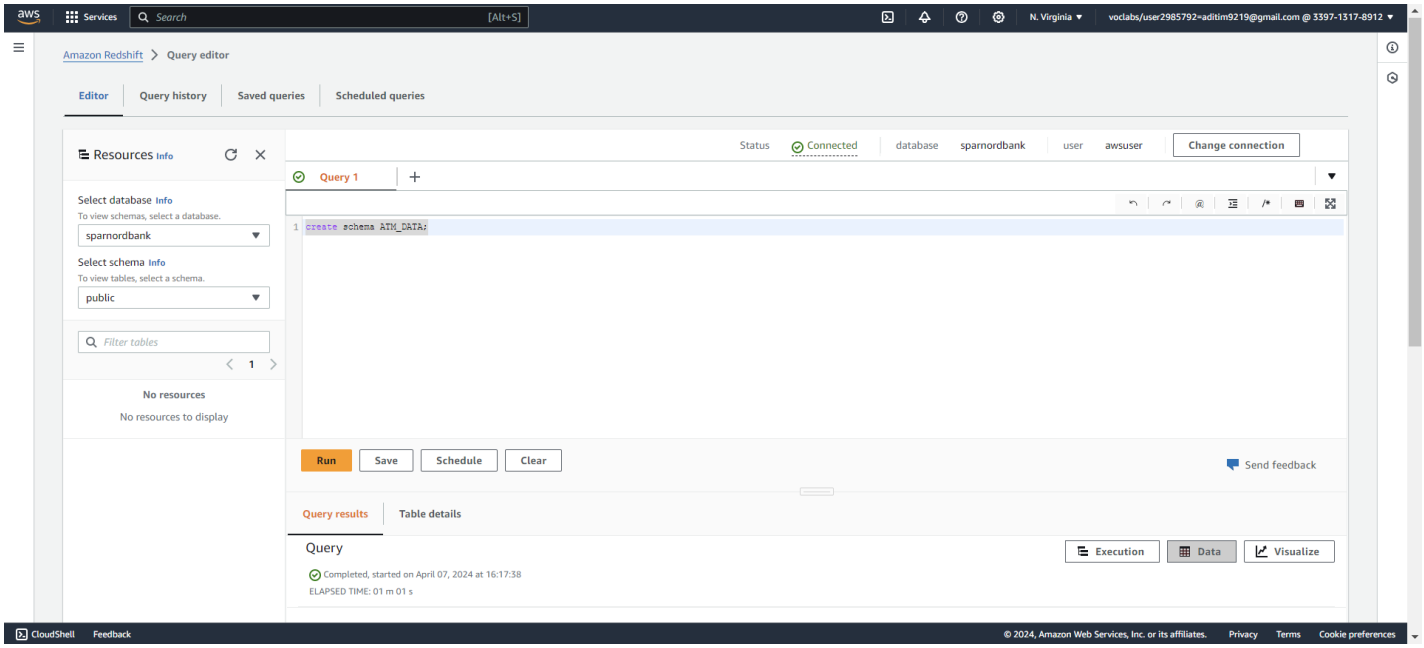


© Copyright. upGrad Education Pvt. Ltd. All rights reserved

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:

Creating Schema:
Query:
create schema ATM_DATA;

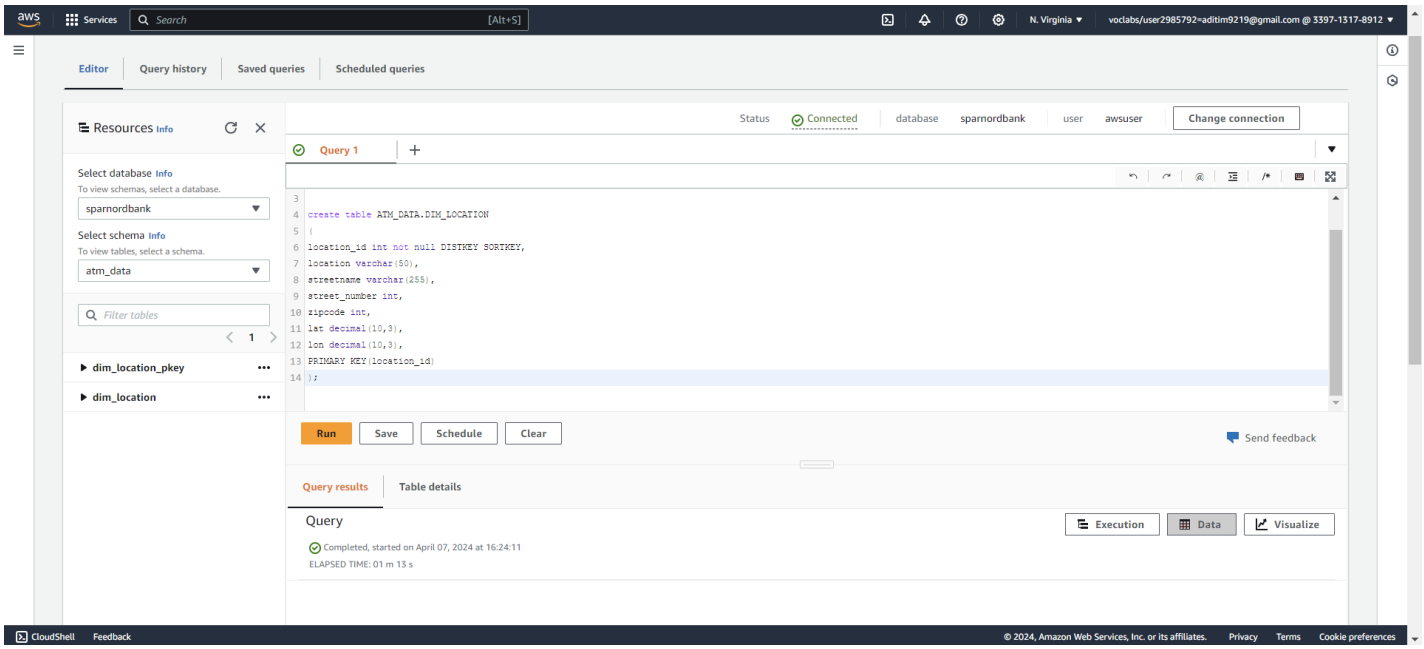


© Copyright. upGrad Education Pvt. Ltd. All rights reserved

Queries for establishing dimension and fact tables, ensuring proper primary and foreign key assignments:

1. **Creating LOCATION dimension table:**
   Query:

```
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)
);
```

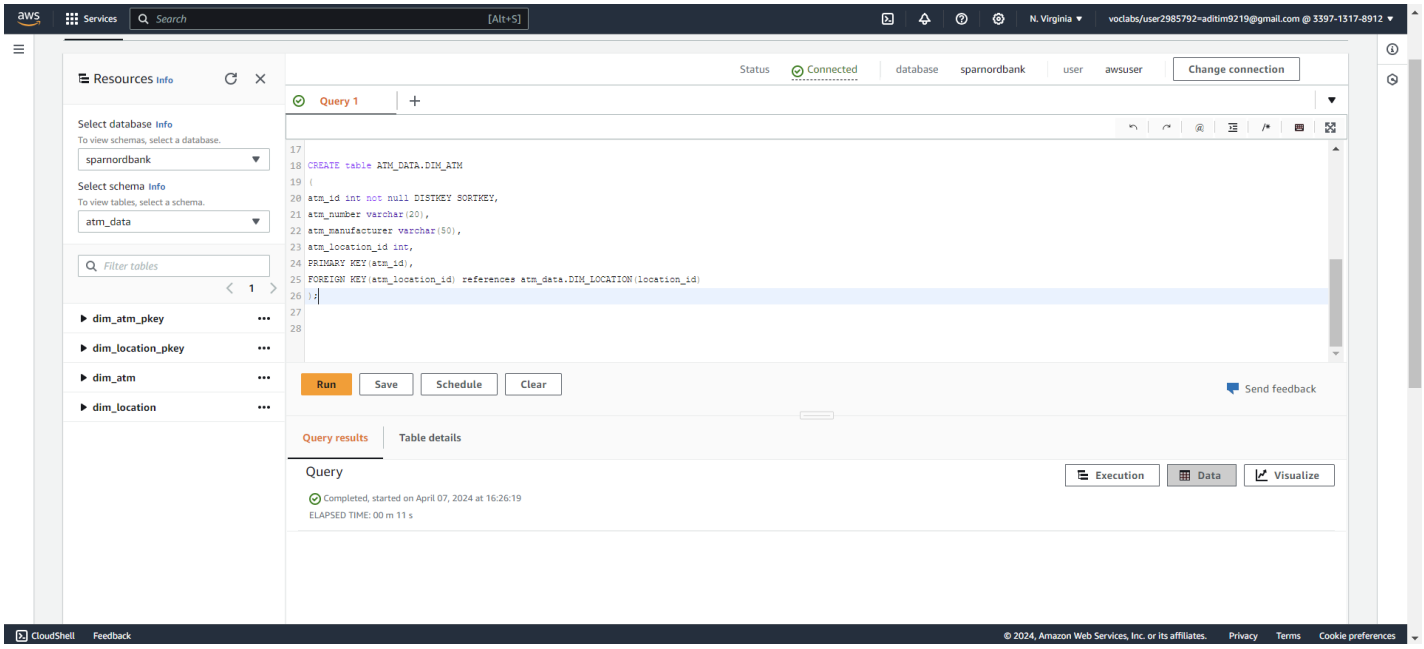


© Copyright. upGrad Education Pvt. Ltd. All rights reserved

2. **Creating ATM dimension table:**

Query:

```
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)
);
```

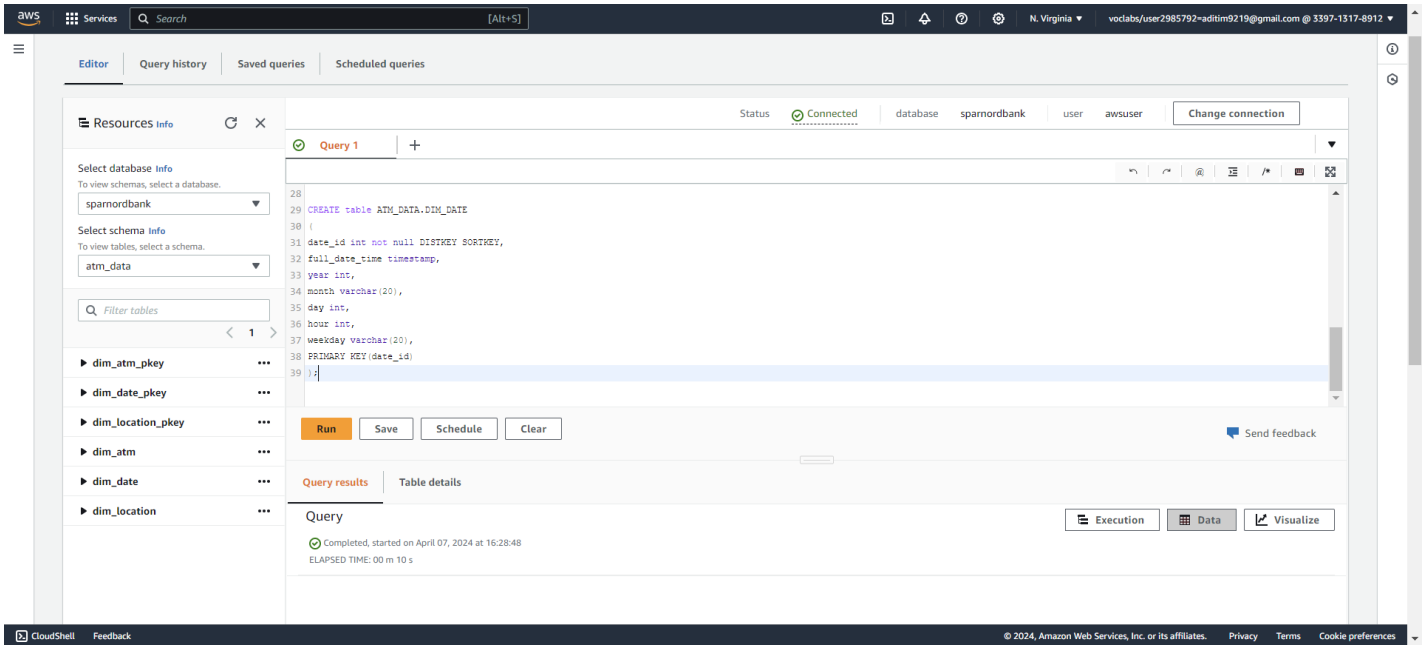


© Copyright. upGrad Education Pvt. Ltd. All rights reserved

3. **Creating DATE dimension table:**

Query:

```
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)
);
```

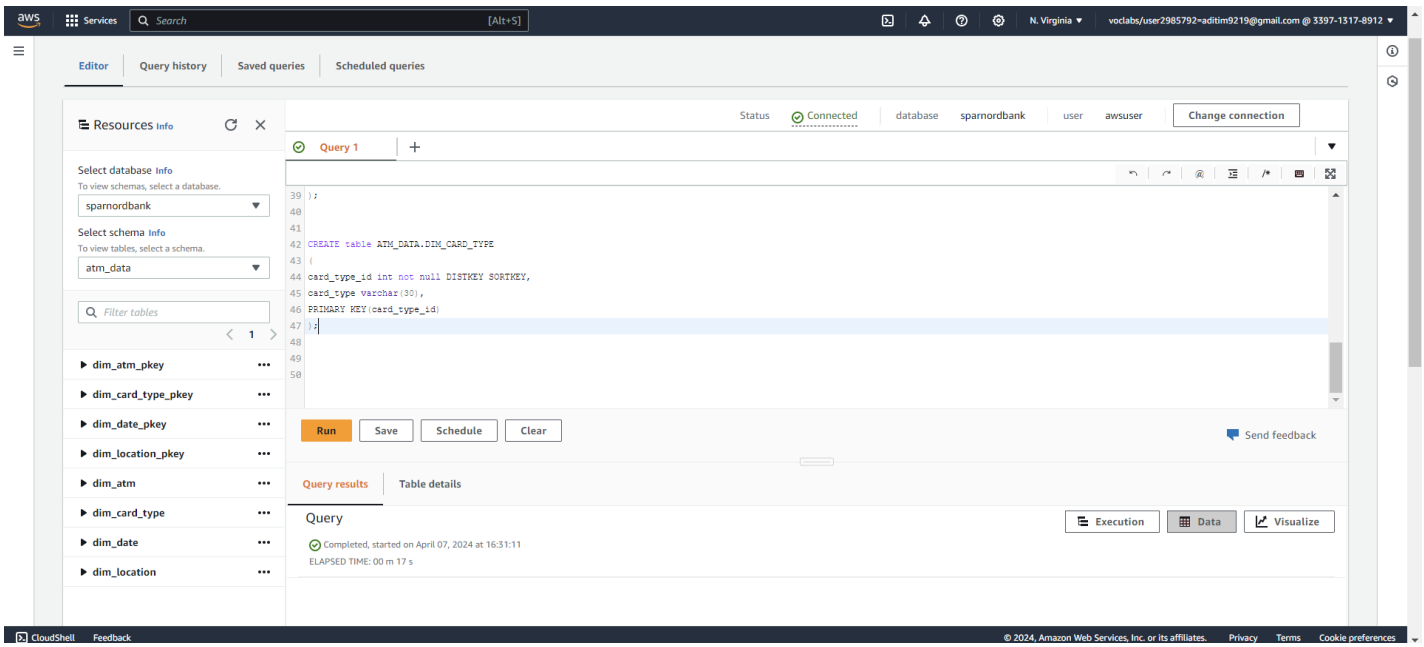


© Copyright. upGrad Education Pvt. Ltd. All rights reserved

4. **Creating CARD TYPE dimension table:**

Query:

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

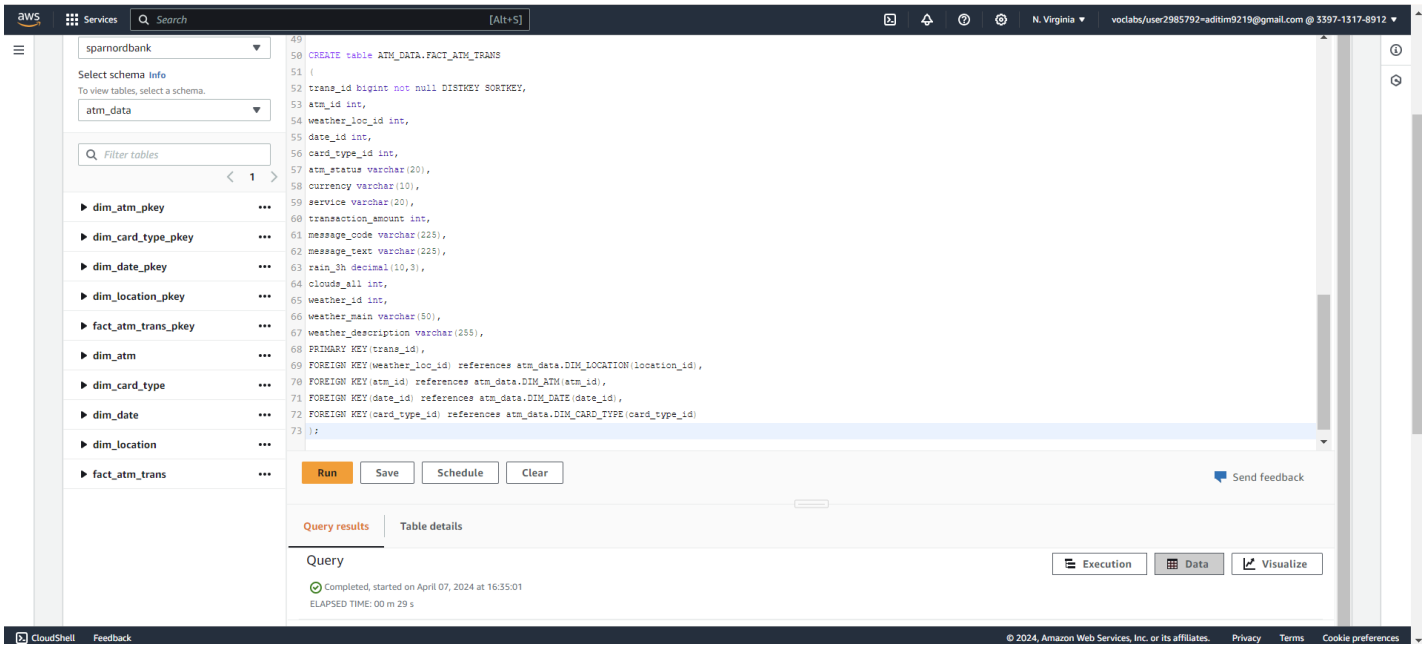


© Copyright. upGrad Education Pvt. Ltd. All rights reserved

5. **Creating ATM TRANSACTIONS FACT table:**

Query:

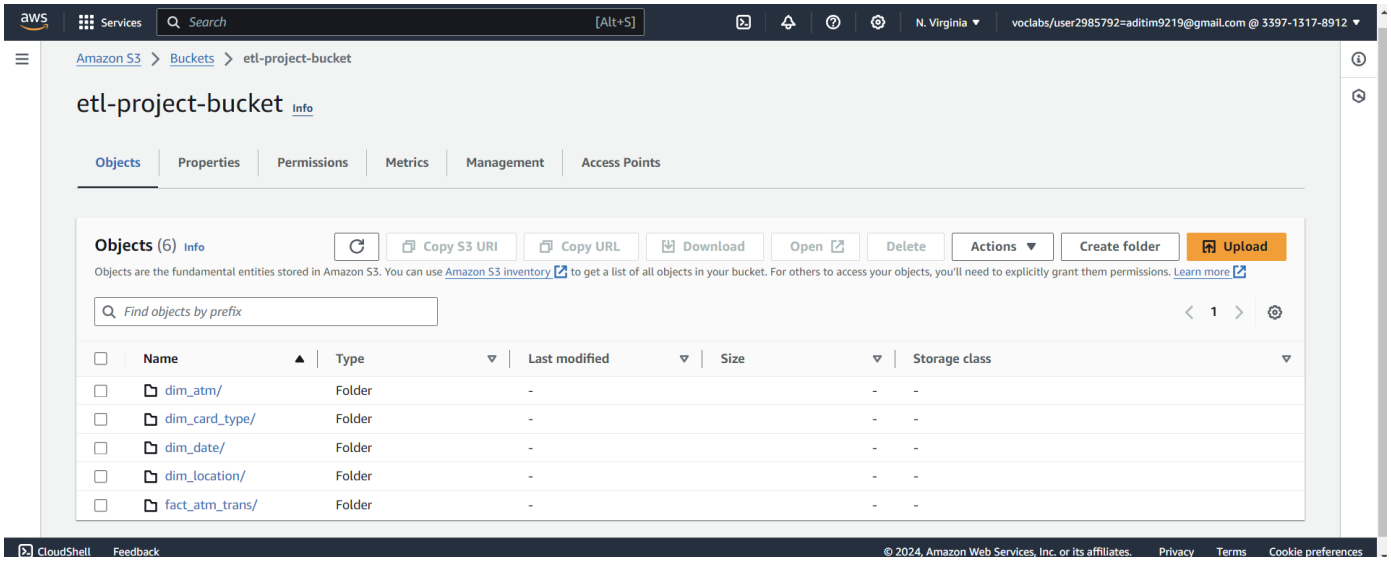
```
CREATE table ATM_DATA.FACT_ATM_TRANS
(
trans_id bigint not null DISTKEY SORTKEY,
atm_id int,
weather_loc_id int,
date_id int,
card_type_id int,
atm_status varchar(20),
currency varchar(10),
service varchar(20),
transaction_amount int,
message_code varchar(225),
message_text varchar(225),
rain_3h 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)
);
```



© Copyright. upGrad Education Pvt. Ltd. All rights reserved

Loading data into a Redshift cluster from Amazon S3 bucket

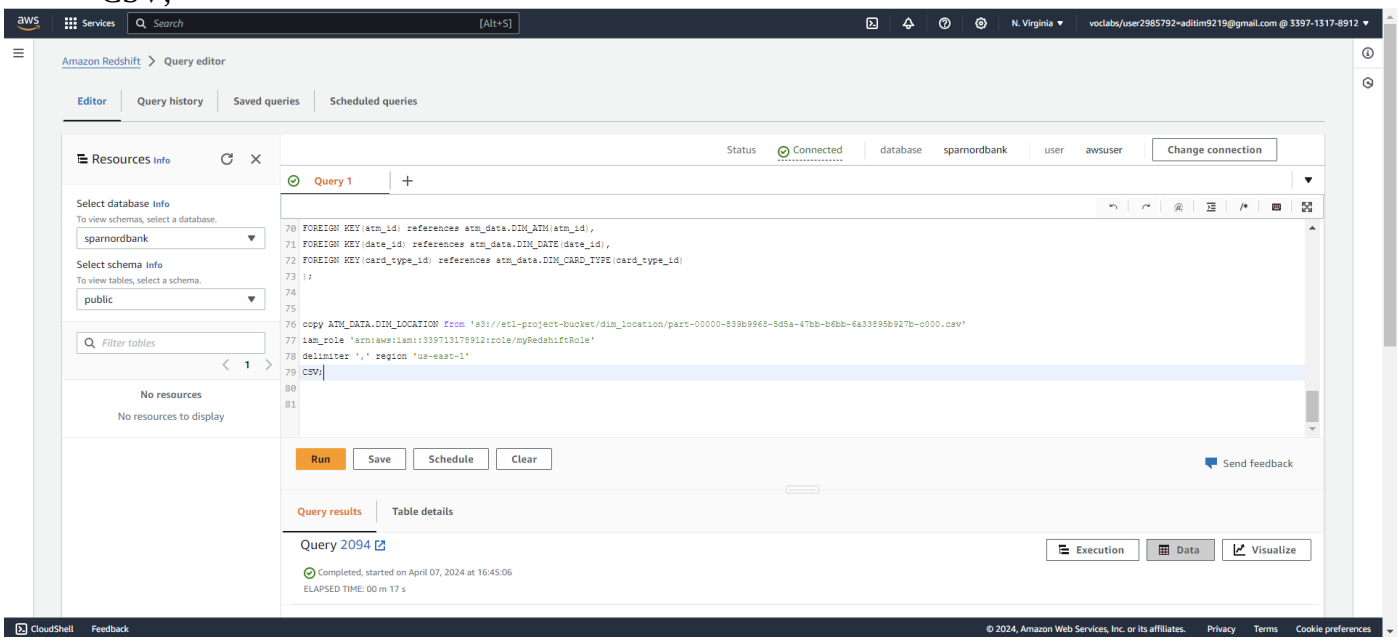
Screenshot of data in S3 bucket:



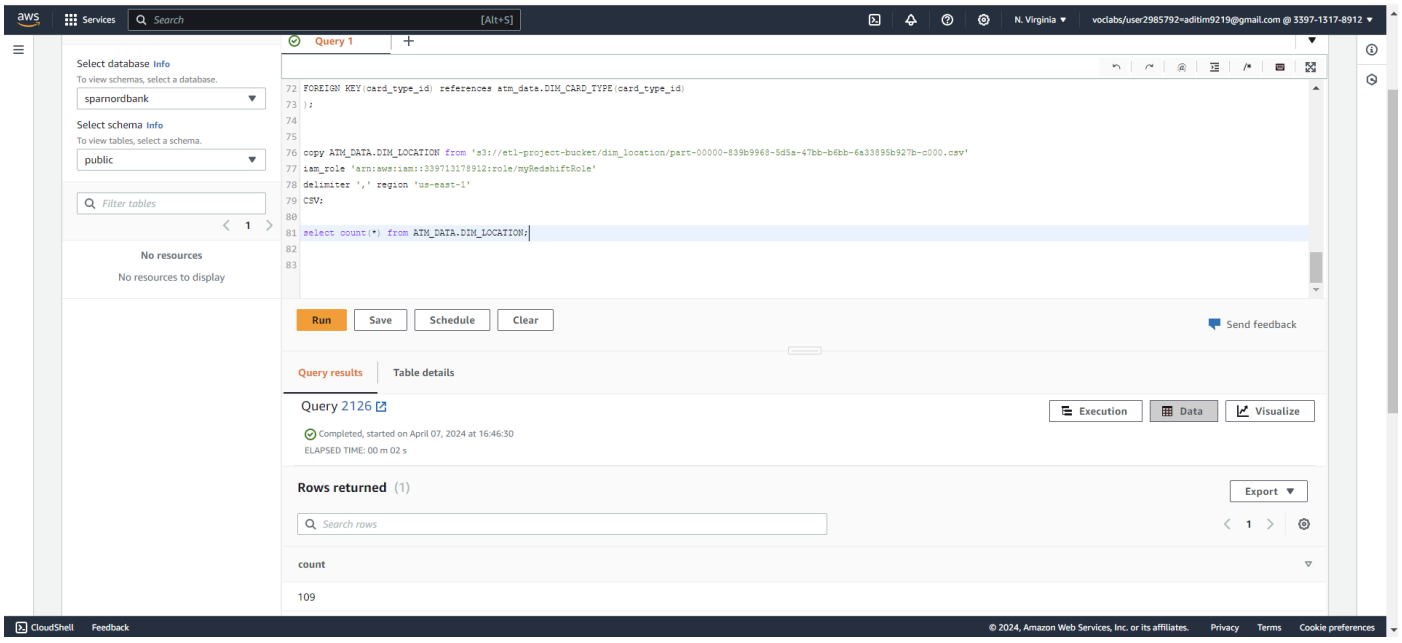
© Copyright. upGrad Education Pvt. Ltd. All rights reserved

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

1. **Copying the data to DIM_LOCATION table:**
   Query:
   copy ATM_DATA.DIM_LOCATION from 's3://etl-project-bucket/dim_location/part-00000-839b9968-5d5a-47bb-b6bb-6a33895b927b-c000.csv'
   iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
   delimiter ',' region 'us-east-1'
   CSV;



Count of rows in DIM_LOCATION table:
Query: select count(*) from ATM_DATA.DIM_LOCATION;



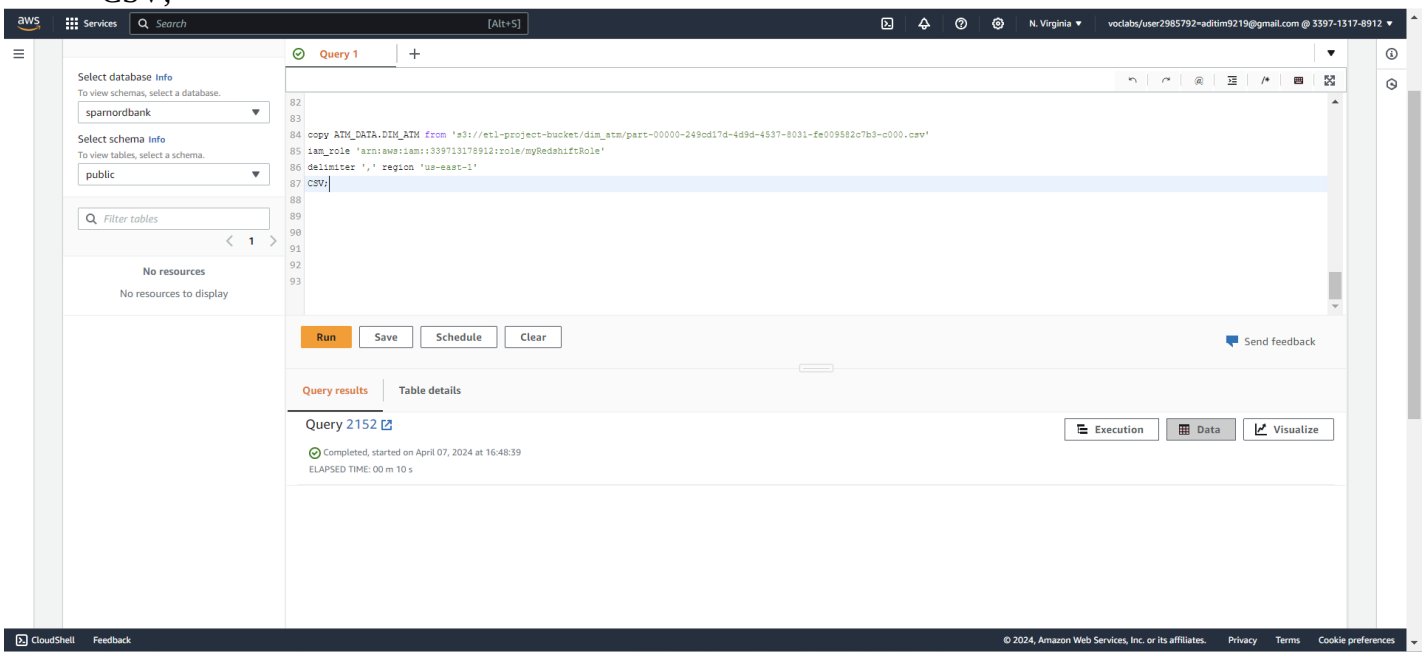
Total number of rows retrieved is **109.**

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

2. **Copying the data to DIM_ATM table:**

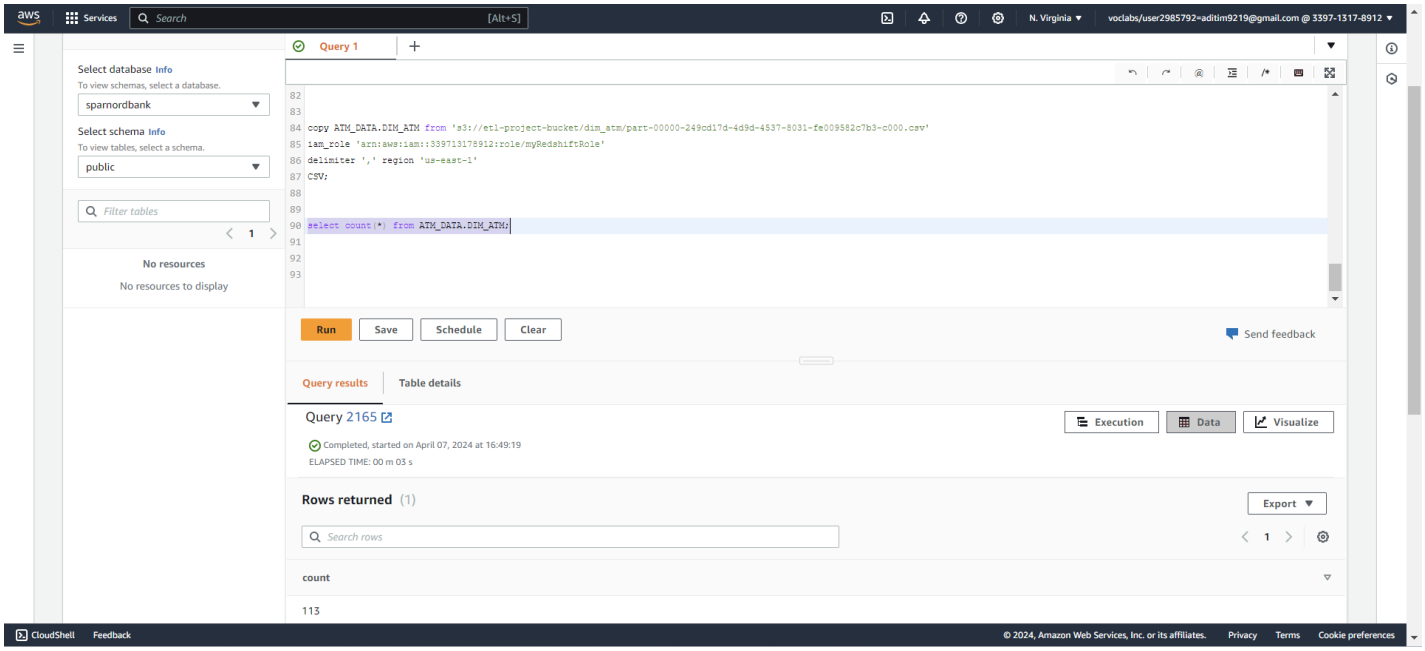
Query:

```
copy ATM_DATA.DIM_ATM from 's3://etl-project-bucket/dim_atm/part-00000-249cd17d-4d9d-4537-8031-fe009582c7b3-c000.csv'
iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



Count of rows in DIM_ATM table:

Query: select count(*) from ATM_DATA. DIM_ATM;



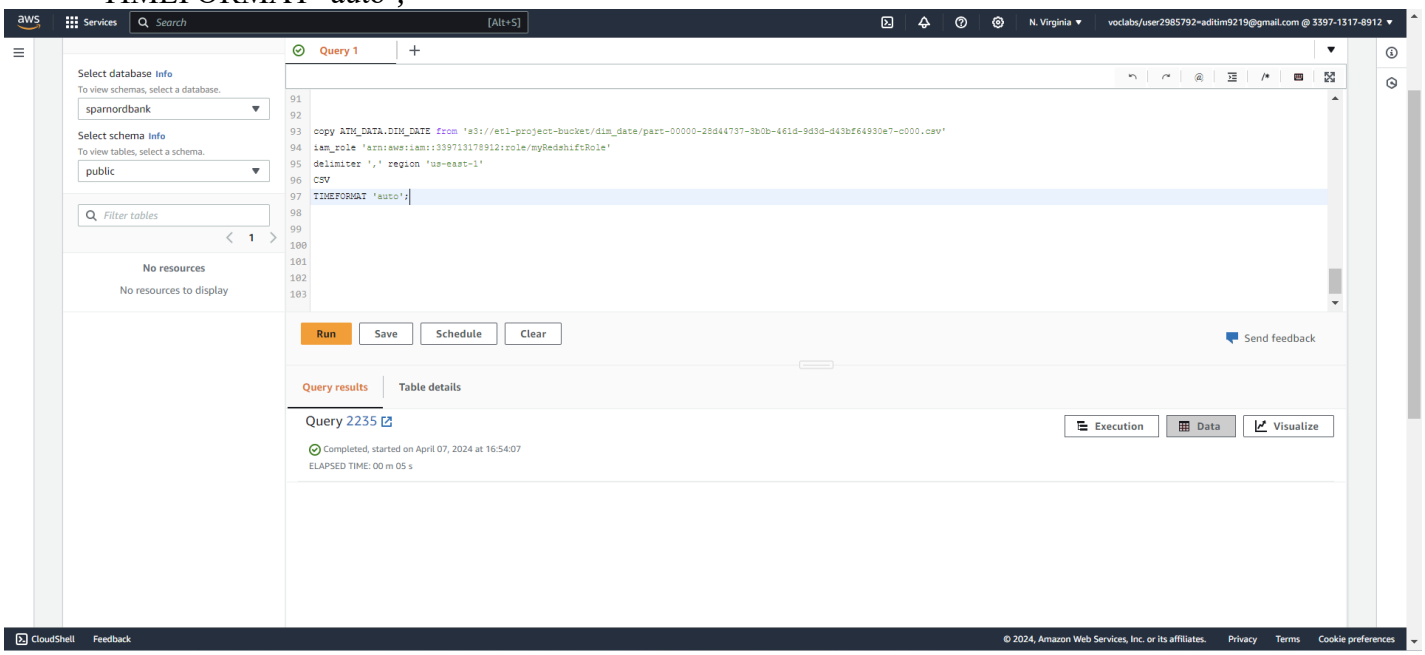
Total number of rows retrieved is **113.**

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

3. **Copying the data to DIM_DATE table:**

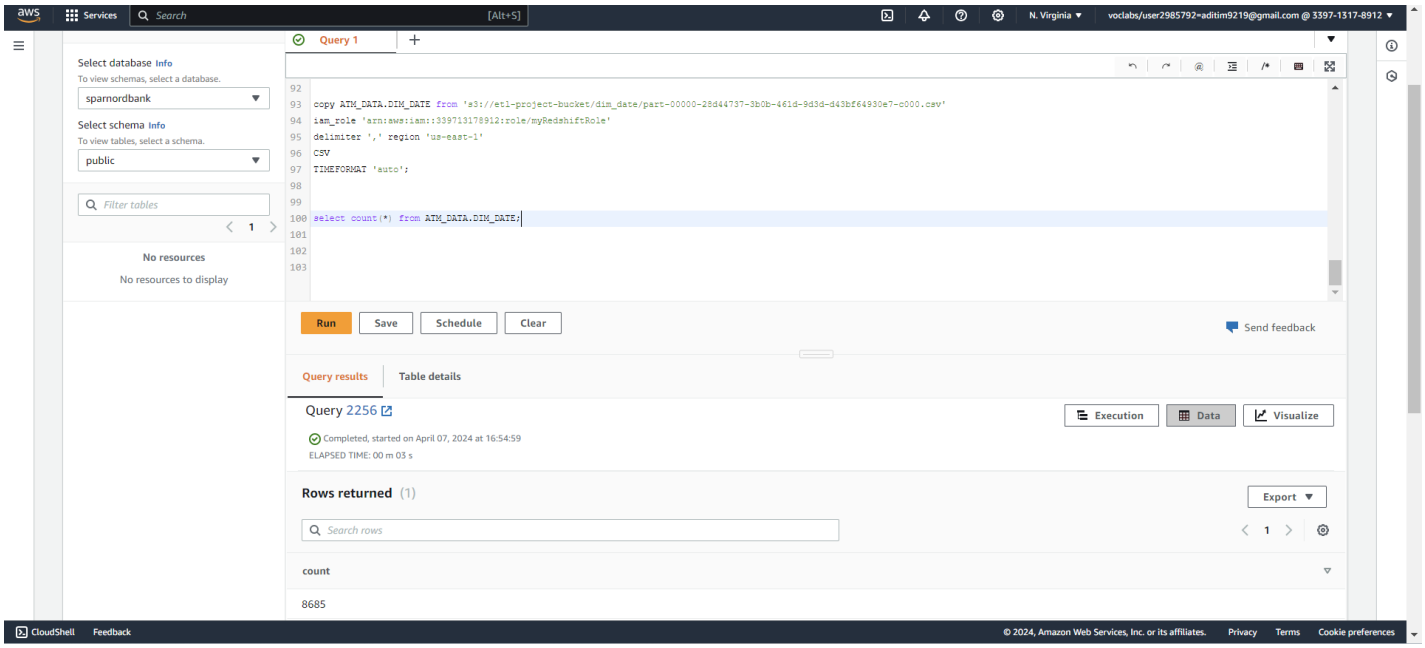
Query:

```
copy ATM_DATA.DIM_DATE from 's3://etl-project-bucket/dim_date/part-00000-28d44737-3b0b-461d-9d3d-d43bf64930e7-c000.csv'
iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';
```



Count of rows in DIM_DATE table:

Query: select count(*) from ATM_DATA. DIM_DATE;



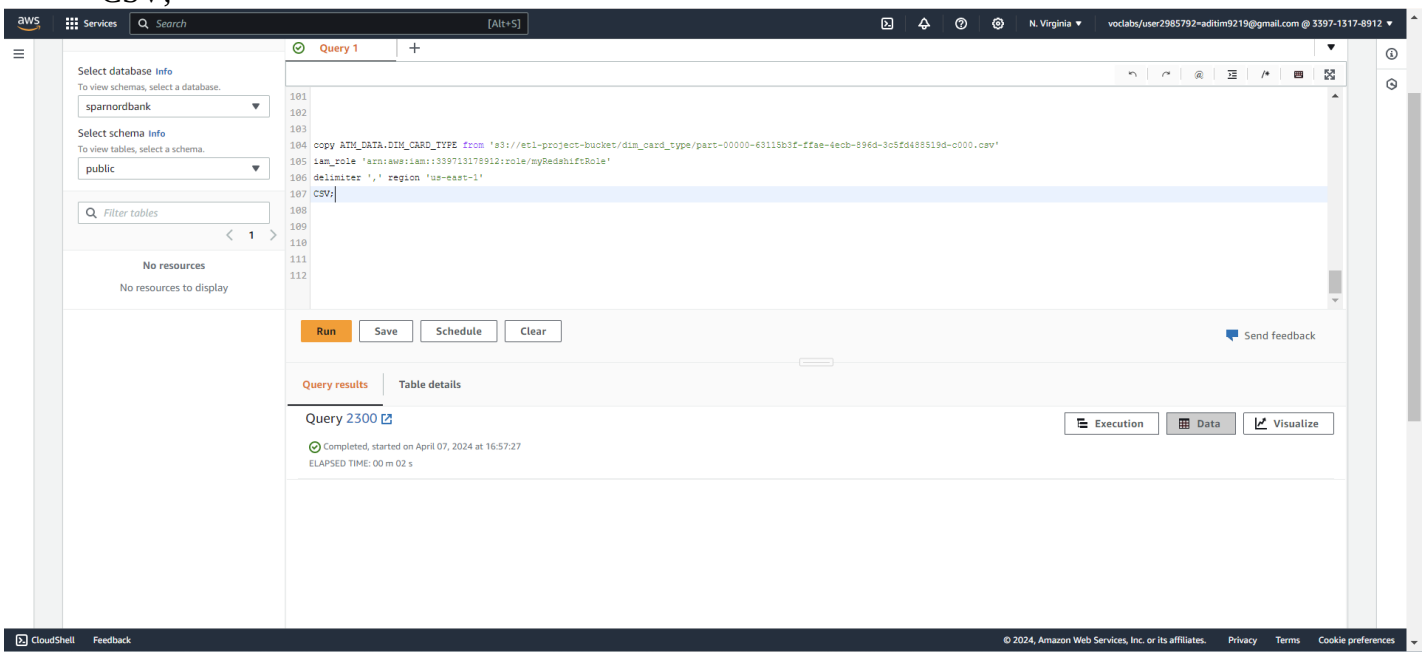
Total number of rows retrieved is **8685.**

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

4. **Copying the data to DIM_CARD_TYPE table:**

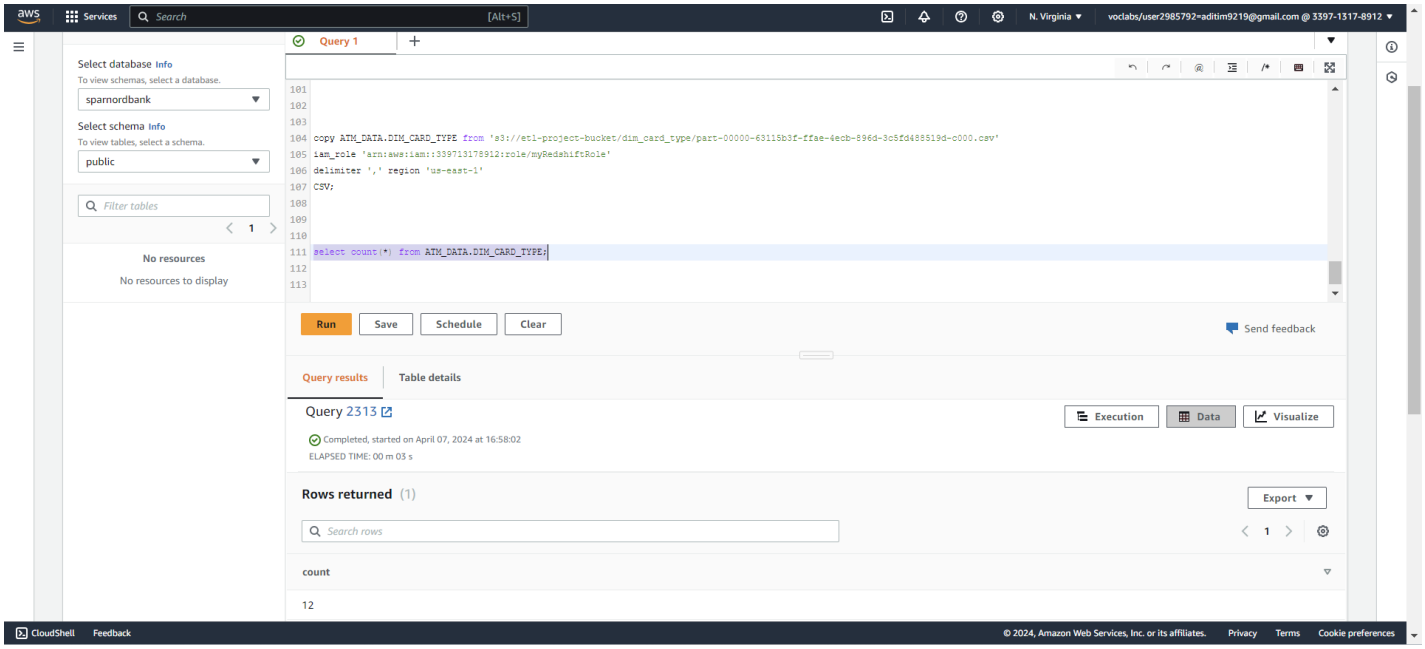
Query:

```
copy ATM_DATA.DIM_CARD_TYPE from 's3://etl-project-bucket/dim_card_type/part-00000-63115b3f-ffae-4ecb-896d-3c5fd488519d-c000.csv'
iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



Count of rows in DIM_CARD_TYPE table:

Query: select count(*) from ATM_DATA. DIM_CARD_TYPE;



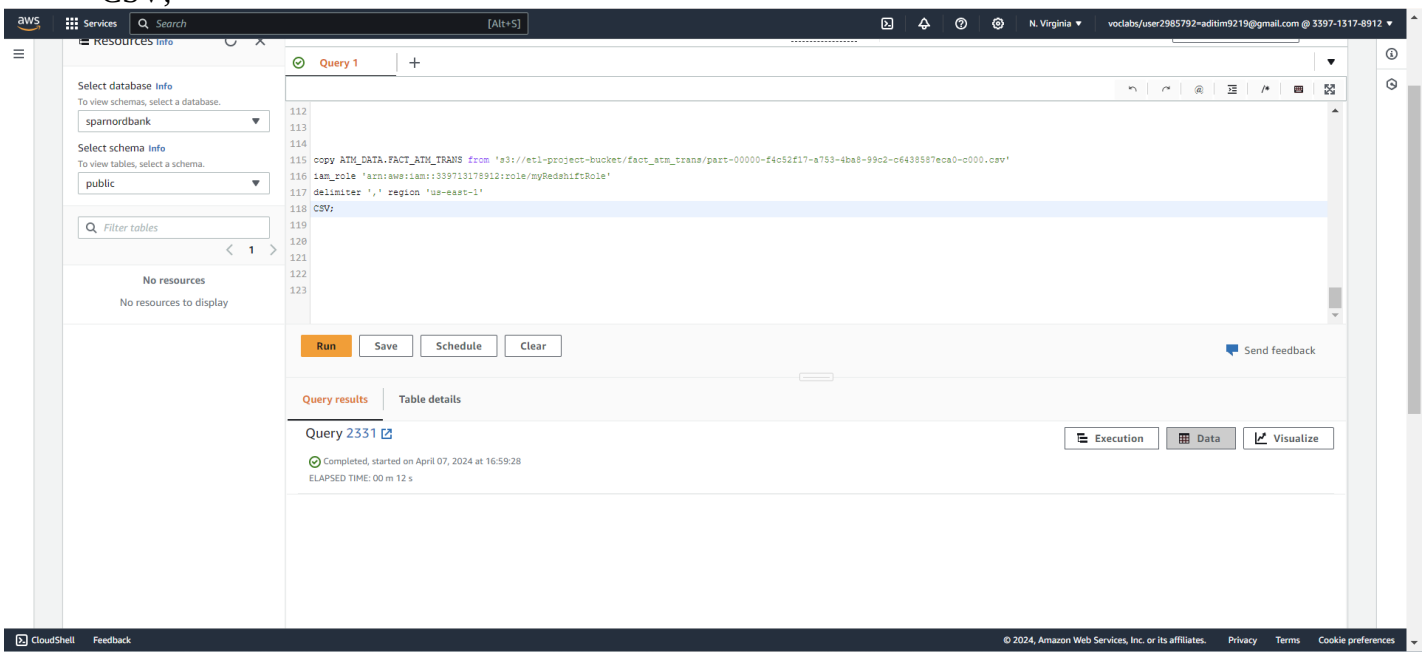
Total number of rows retrieved is **12.**

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

5. **Copying the data to FACT_ATM_TRANS table:**

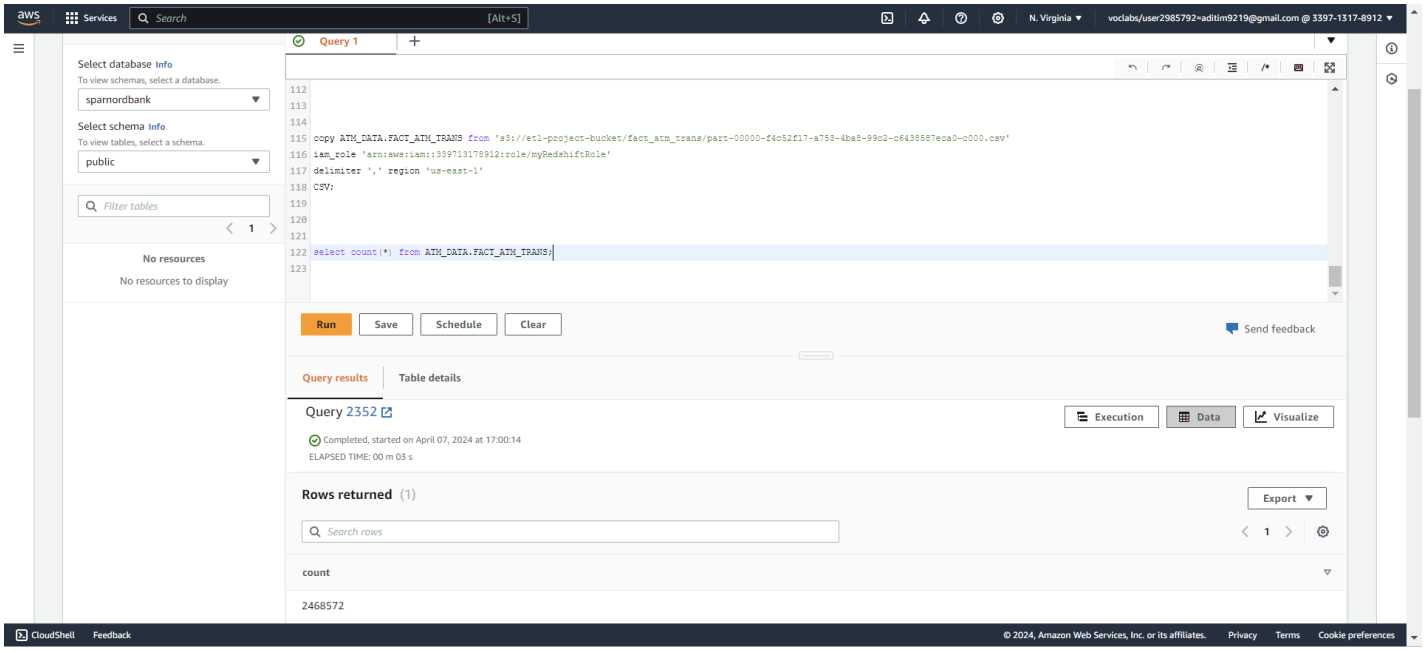
Query:

```
copy ATM_DATA.FACT_ATM_TRANS from 's3://etl-project-bucket/fact_atm_trans/part-00000-f4c52f17-a753-4ba8-99c2-c6438587eca0-c000.csv'
iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



Count of rows in FACT_ATM_TRANS table:

Query: select count(*) from ATM_DATA. FACT_ATM_TRANS;



Total number of rows retrieved is **2468572.**

© Copyright. upGrad Education Pvt. Ltd. All rights reserved