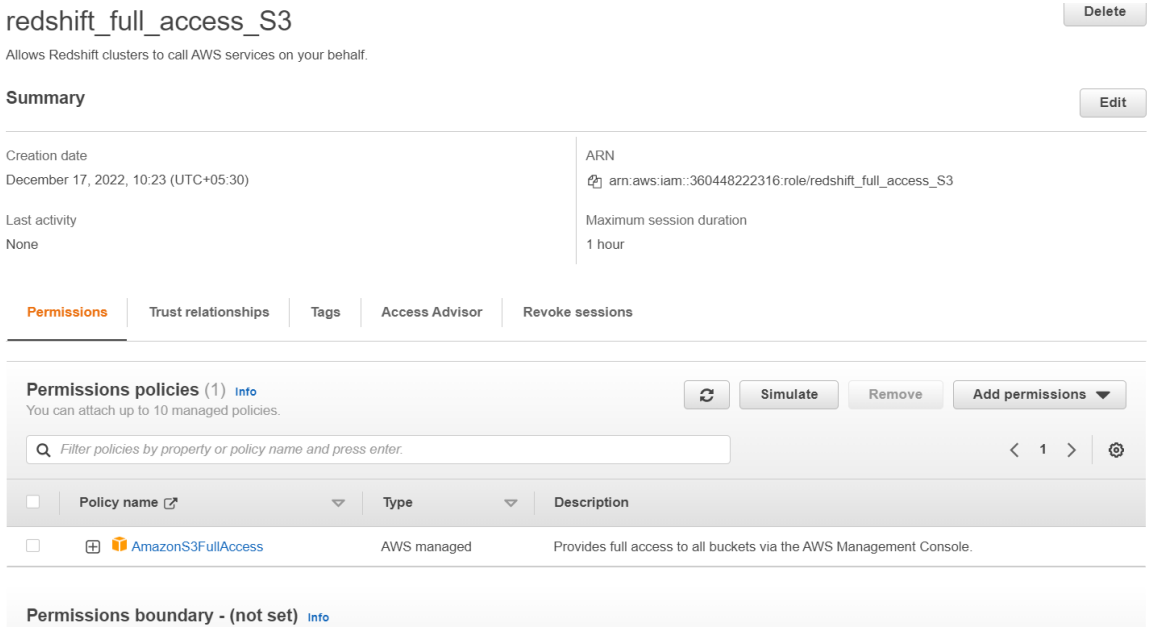
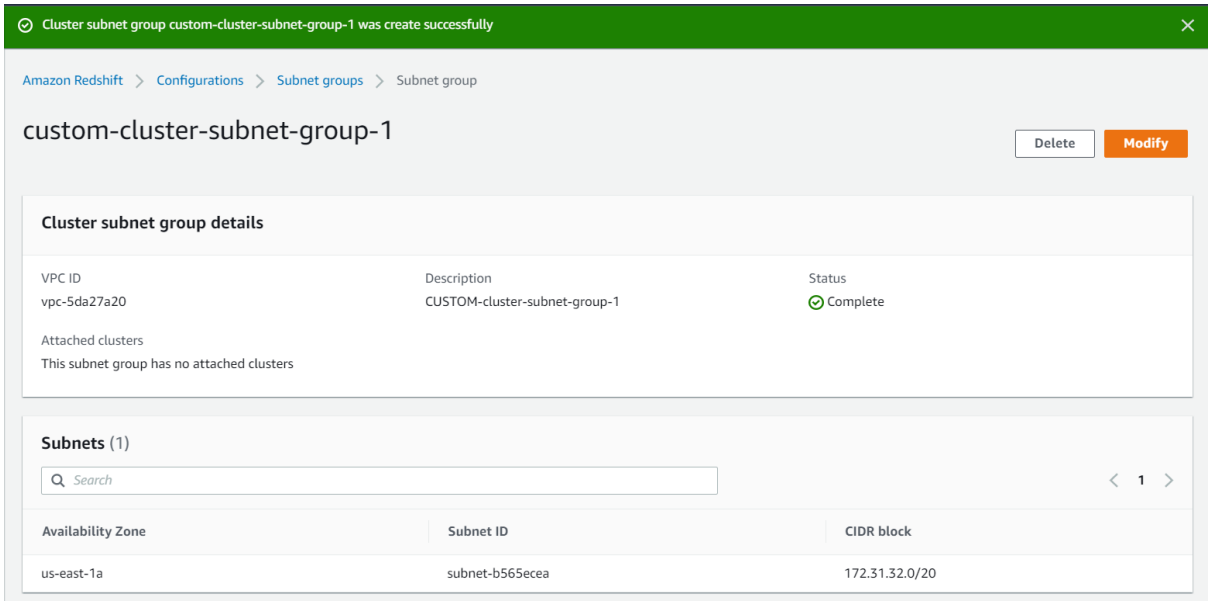
Creation of Redshift Cluster

Creation of IAM Role (for Redshift to get access to S3):



Creation of Cluster Subnet Group



Redshift Cluster Configuration:

Figure 1: Type of node (dc2.large) and number of nodes (2)

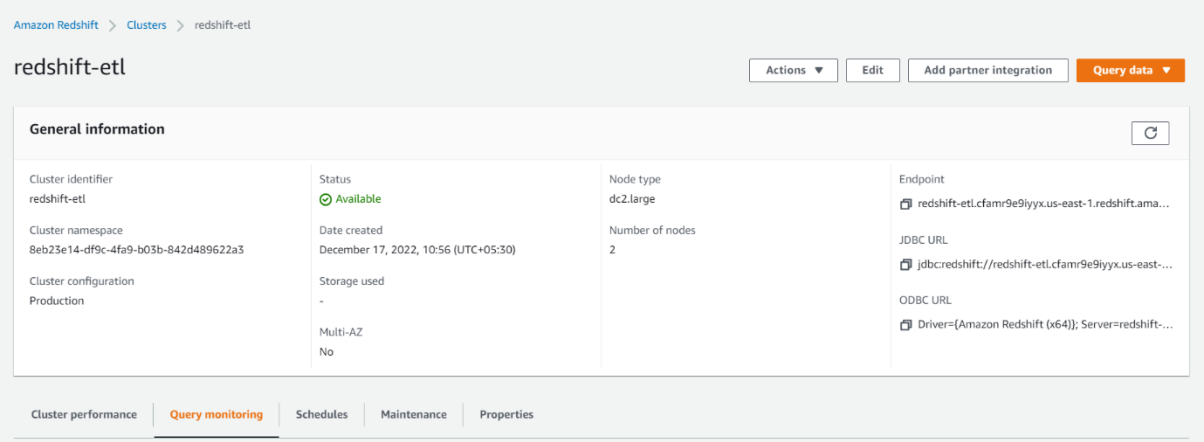


Figure 2: Database configuration, Network and Security Groups

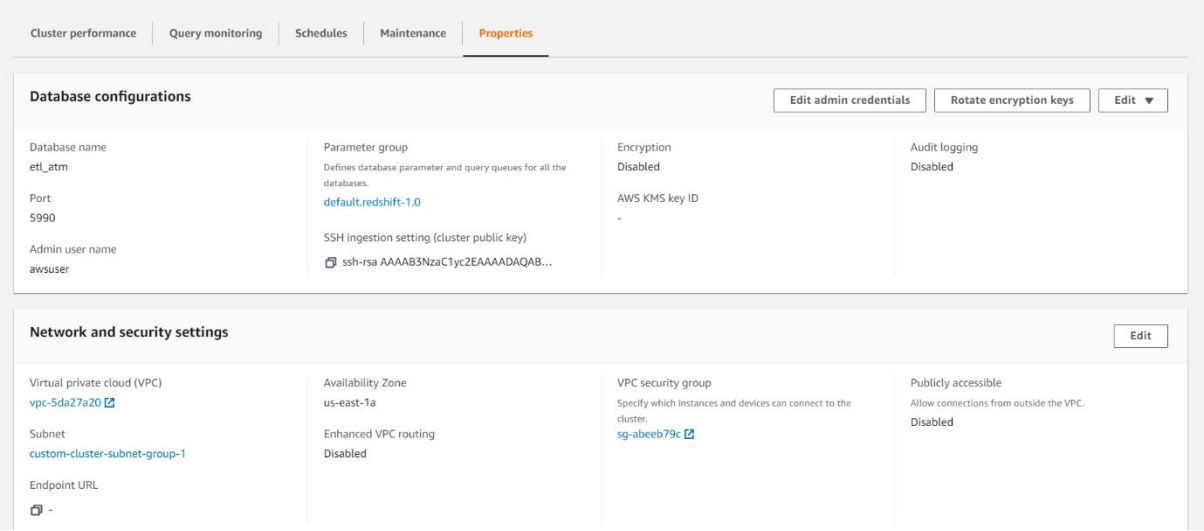


Figure 3: IAM role associated with the cluster

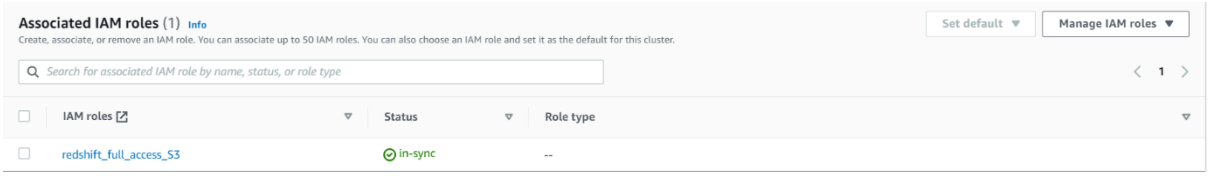


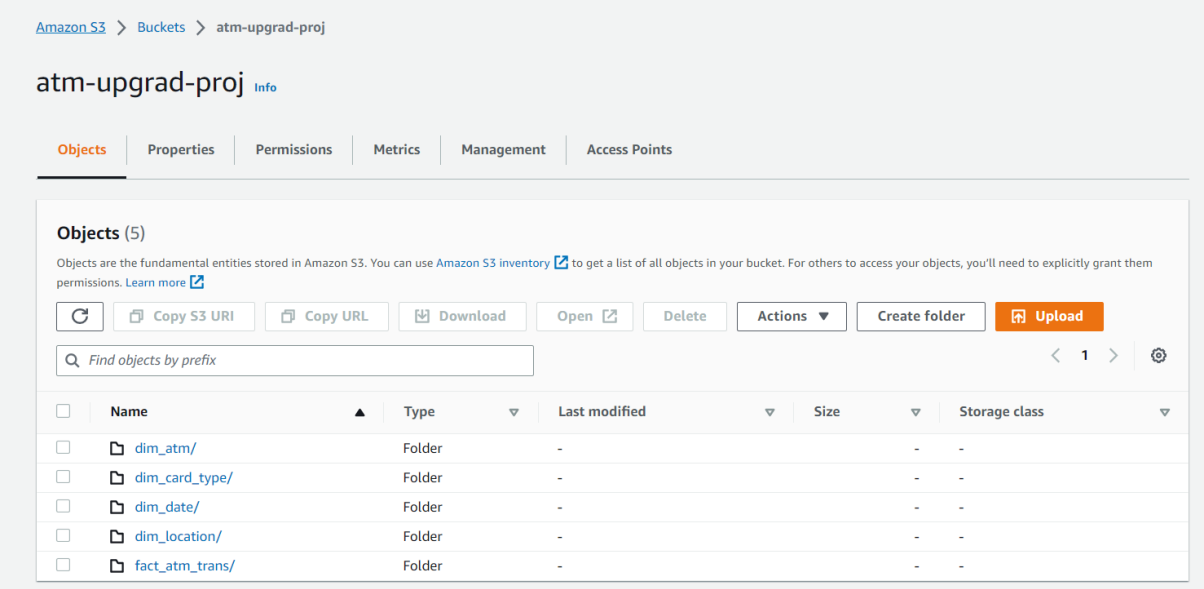
Figure 4: Two Compute nodes and one Leader node

Node IP addresses (3)

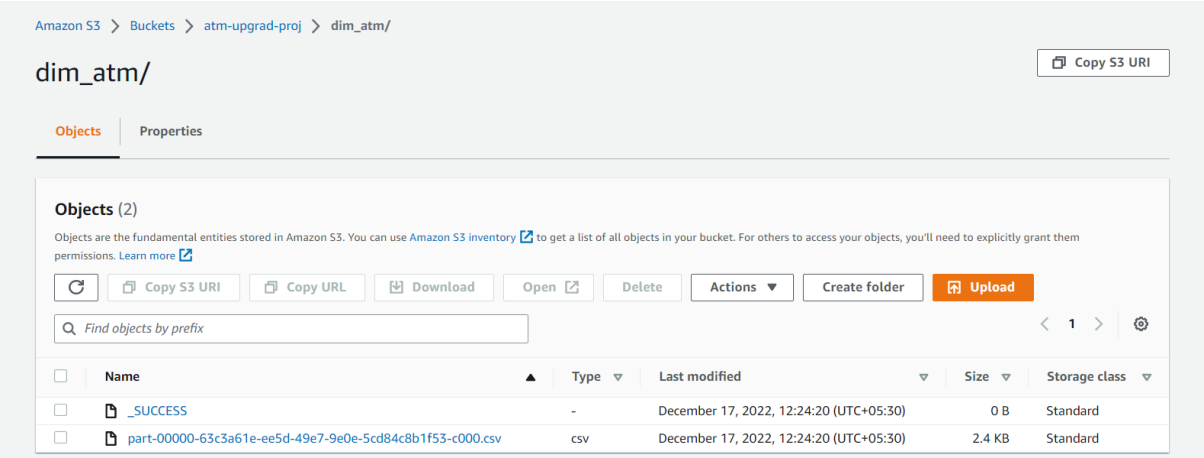
Node role	Public IP address	Private IP address
Leader	18.205.5.137	172.31.41.9
Compute-0	44.213.51.175	172.31.35.243
Compute-1	23.23.103.161	172.31.38.221

View the files created in S3 bucket:

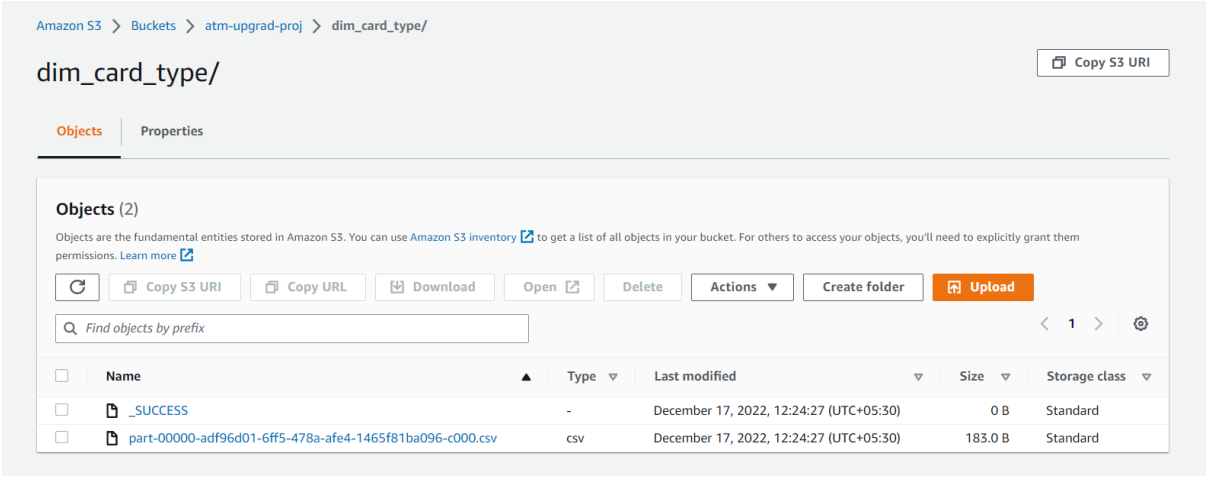
5 folders are created, each for fact/dimension table



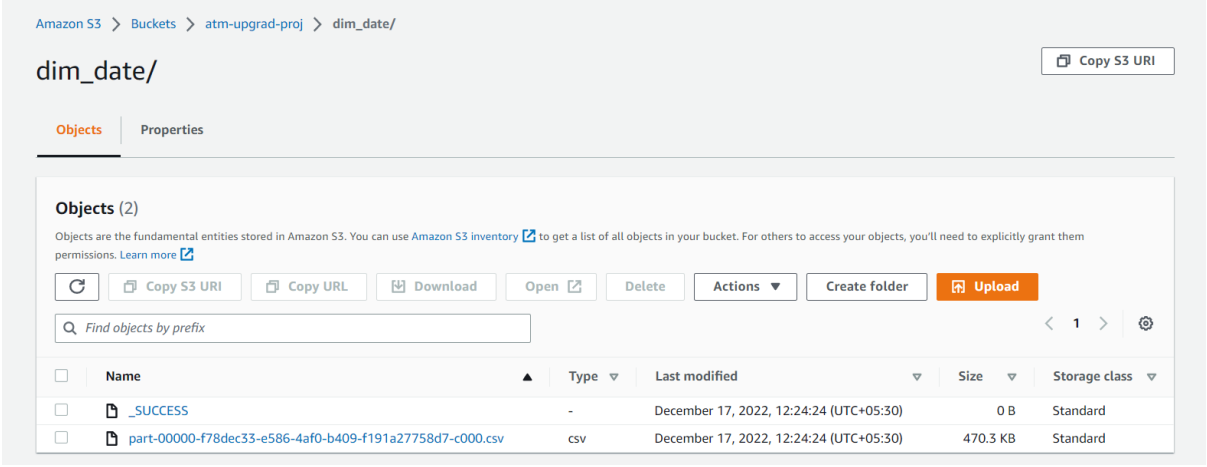
ATM: Dimension table



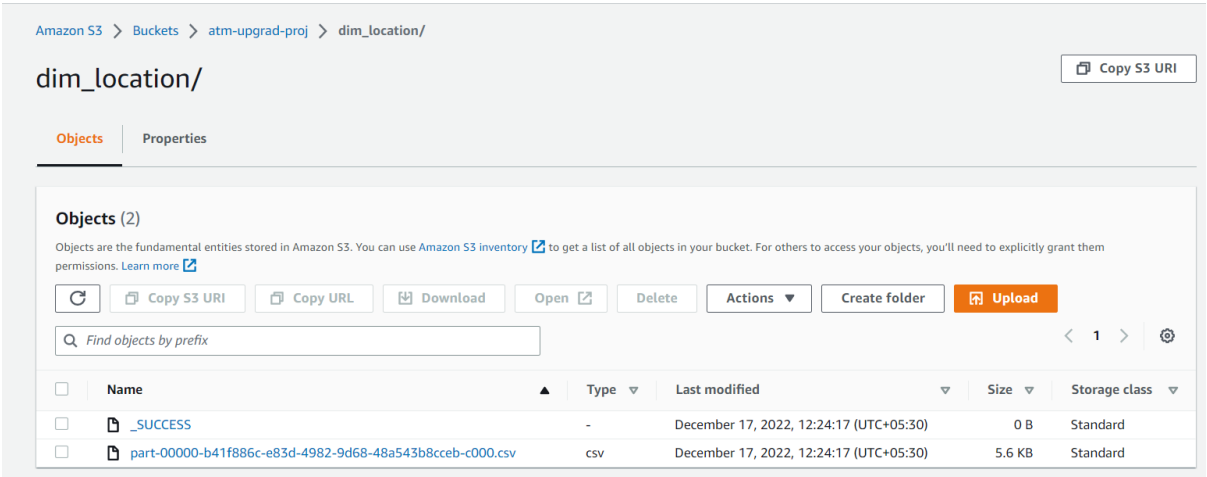
Card Type: Dimension table



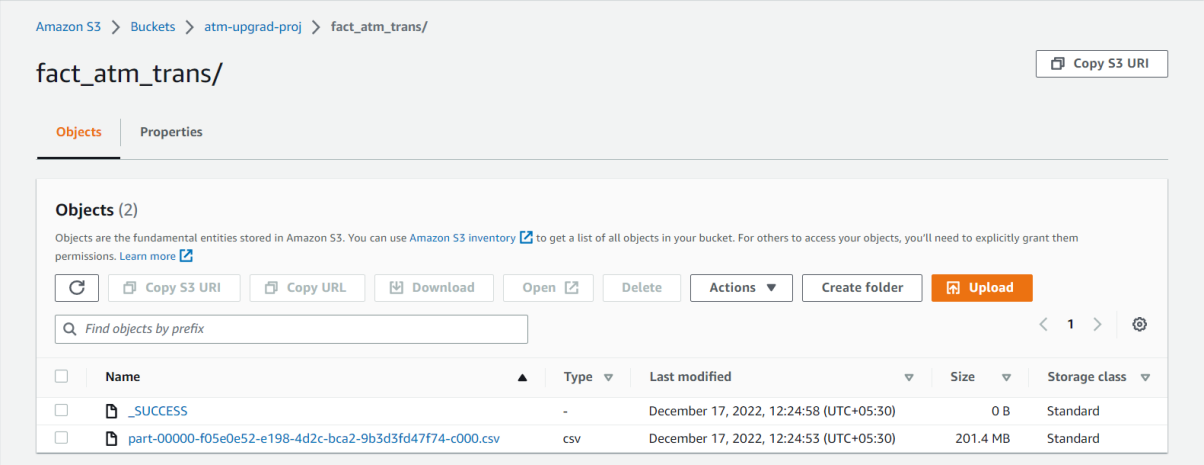
Date: Dimension table



Location: Dimension table



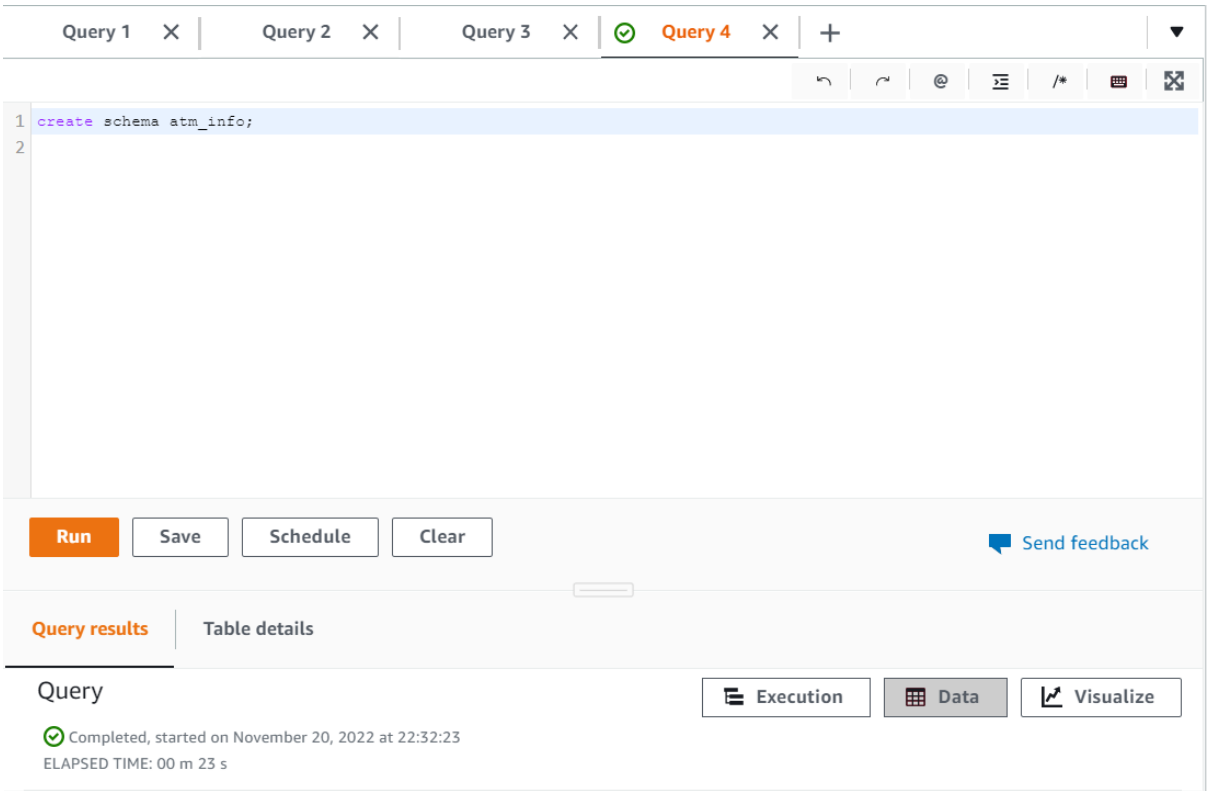
ATM Transactions: Fact table



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

1. Creation of Schema

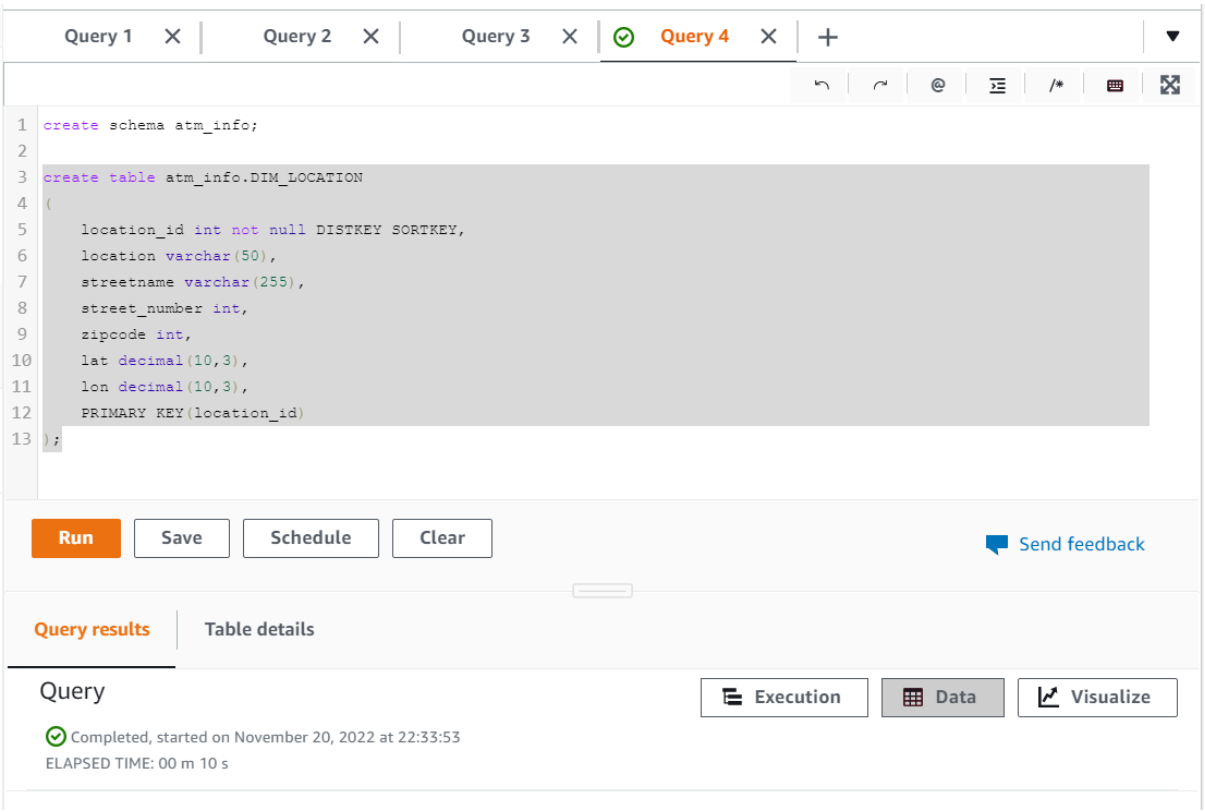
```
create schema atm_info;
```



2. Queries to create the various dimension and fact tables:

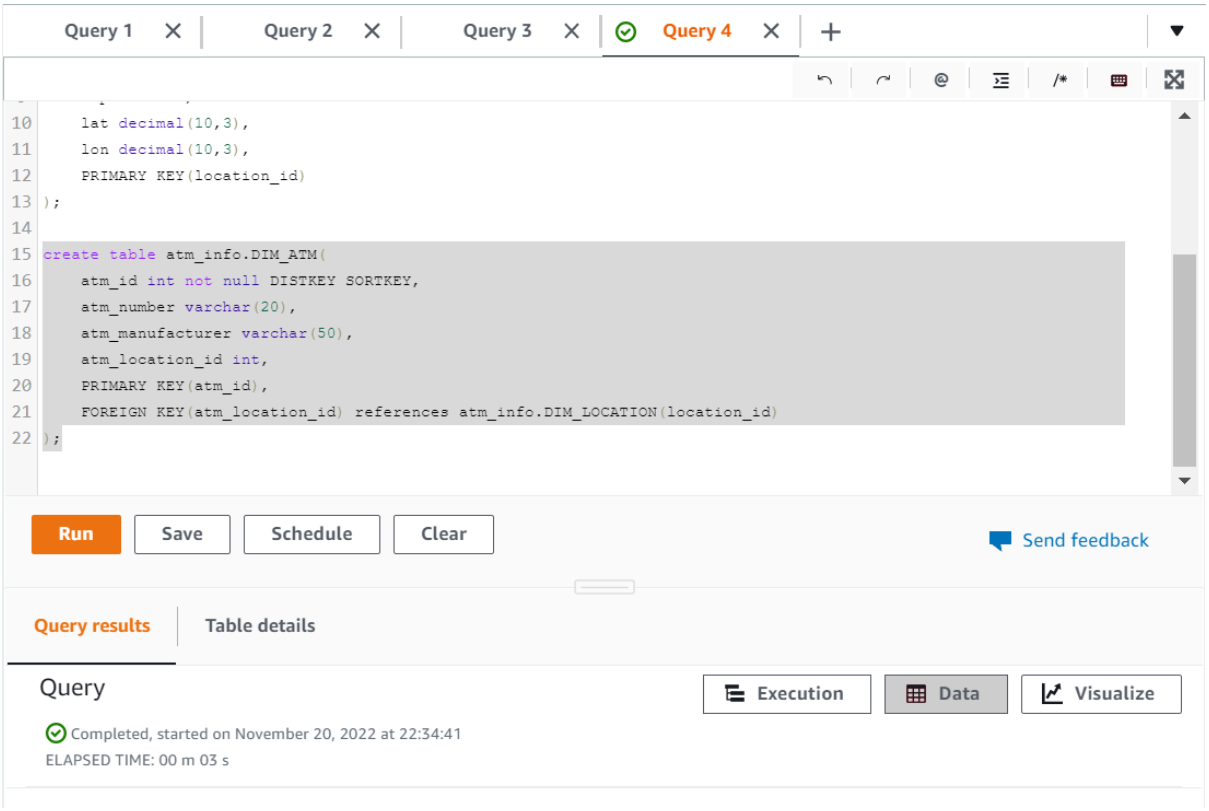
- Creation of DIM_LOCATION table

```
create table atm_info.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)
);
```



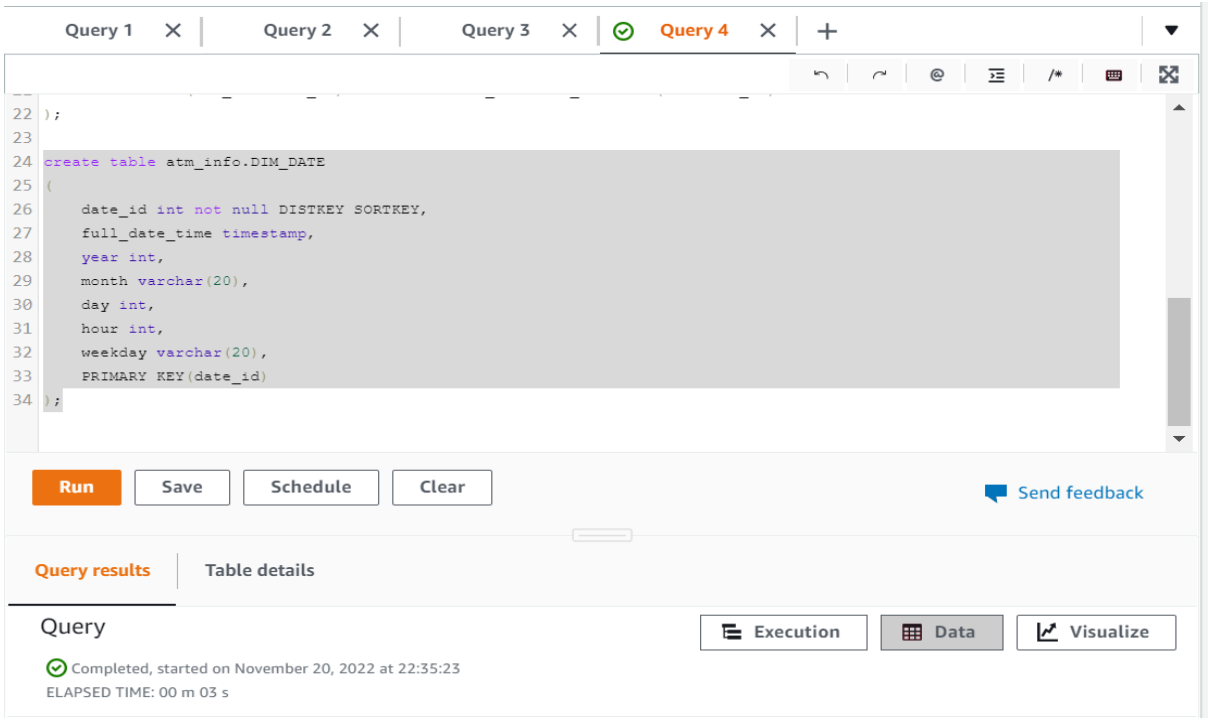
- Creation of DIM_ATM table

```
create table atm_info.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_info.DIM_LOCATION(location_id)
);
```



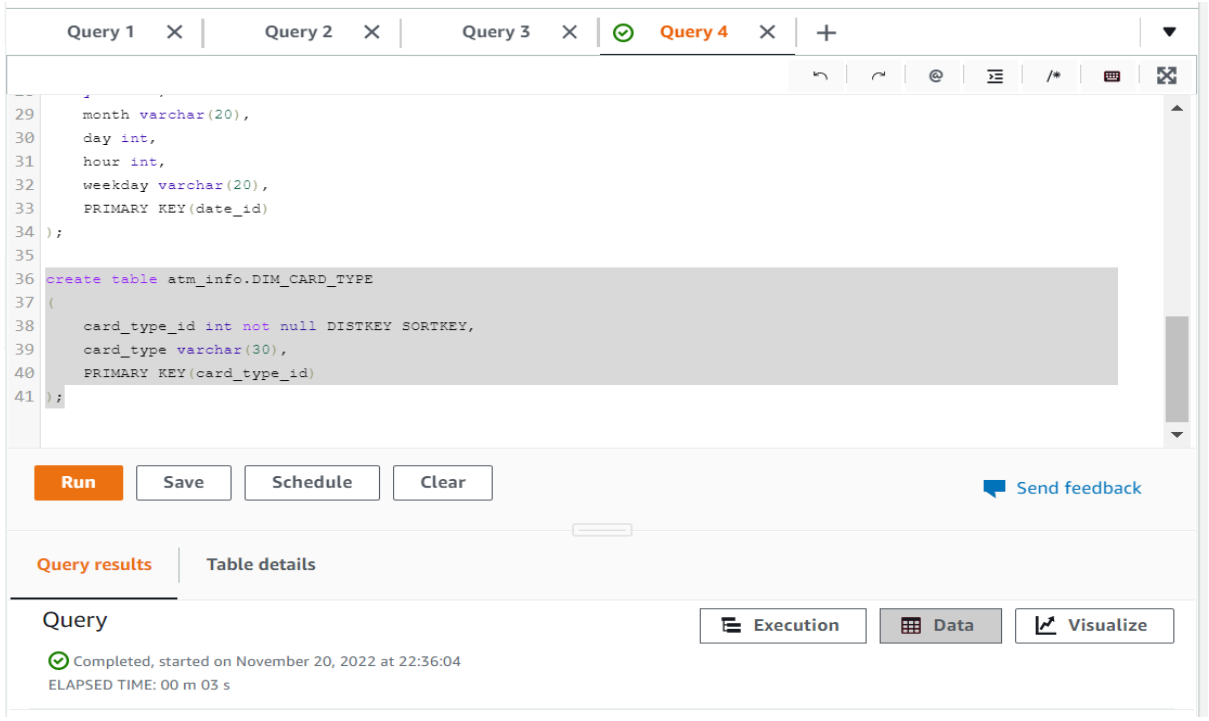
- Creation of DIM_DATE table

```
create table atm_info.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)
);
```



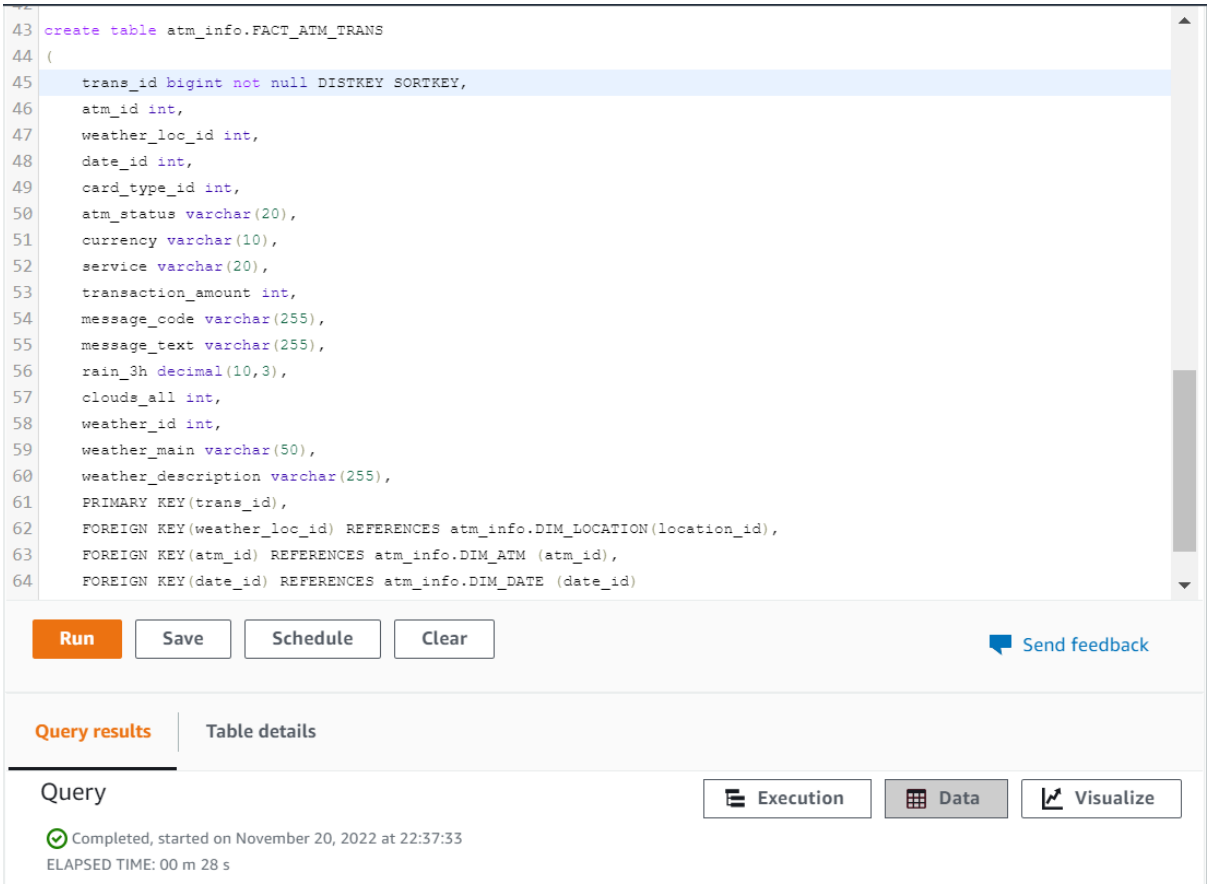
- Creation of DIM_CARD_TYPE table

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



- Creation of FACT_ATM_TRANS table

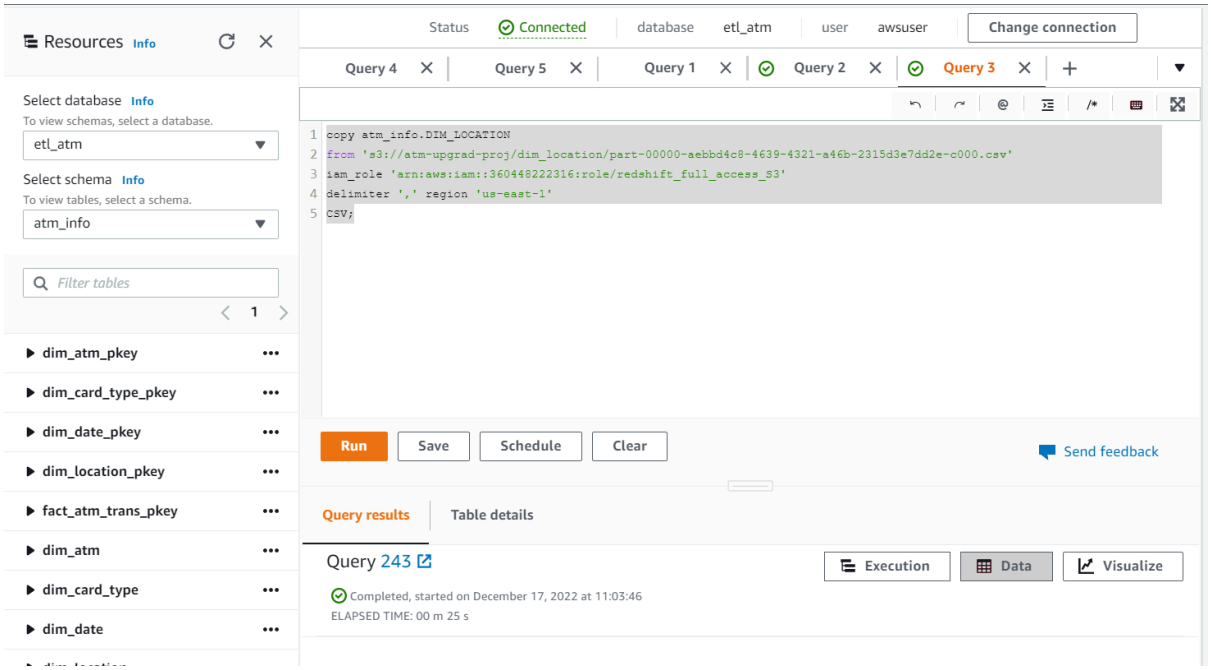
```
create table atm_info.FACT_ATM_TRANS
(
        trans_id bigint not null DISTKEY SORTKEY,
        atm_id int,
        weather_loc_id int,
        date_id int,
        card_type_id int,
        atm_status varchar(20),
        currency varchar(10),
        service varchar(20),
        transaction_amount int,
        message_code varchar(255),
        message_text varchar(255),
        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_info.DIM_LOCATION(location_id),
        FOREIGN KEY(atm_id) REFERENCES atm_info.DIM_ATM (atm_id),
        FOREIGN KEY(date_id) REFERENCES atm_info.DIM_DATE (date_id)
);
```



3. Loading data from Amazon S3 buckets in the appropriate tables in a Redshift cluster

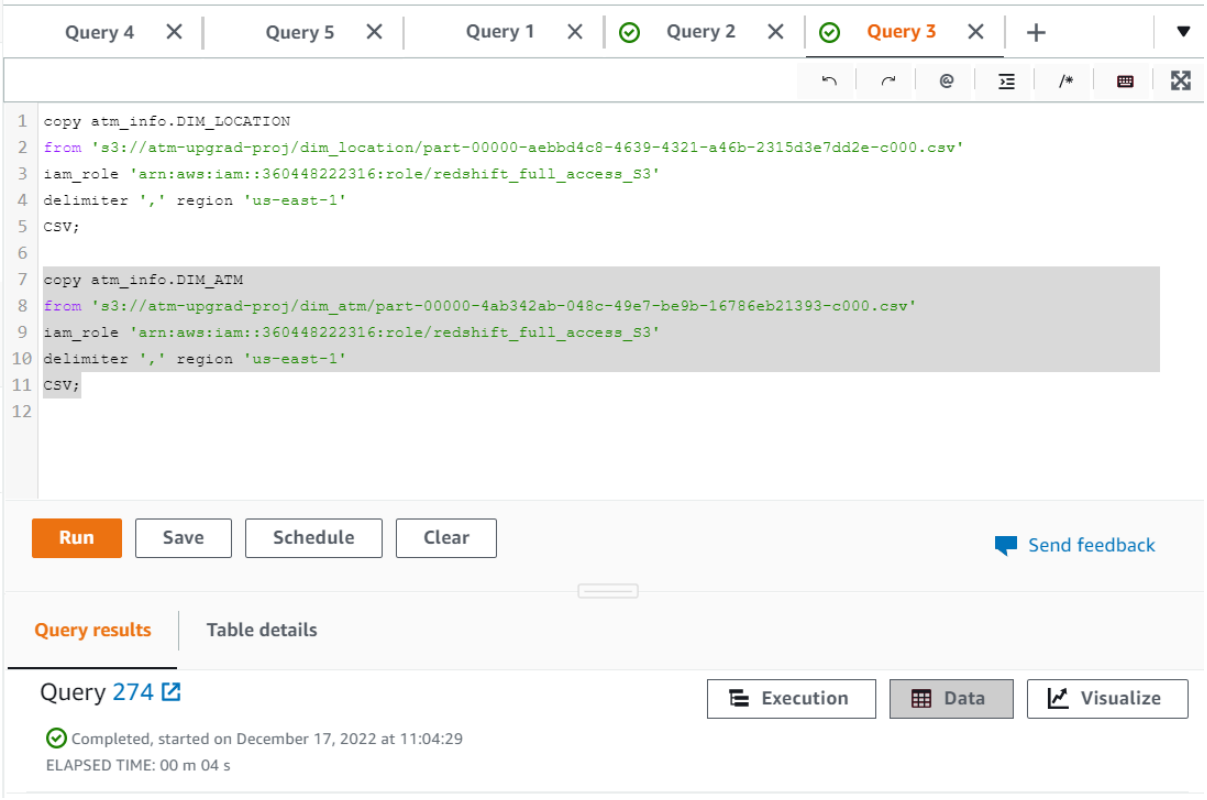
- Load data into DIM_LOCATION table

```
copy atm_info.DIM_LOCATION
from 's3://atm-upgrad-proj/dim_location/part-00000-b41f886c-e83d-4982-9d68-48a543b8cceb-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
```



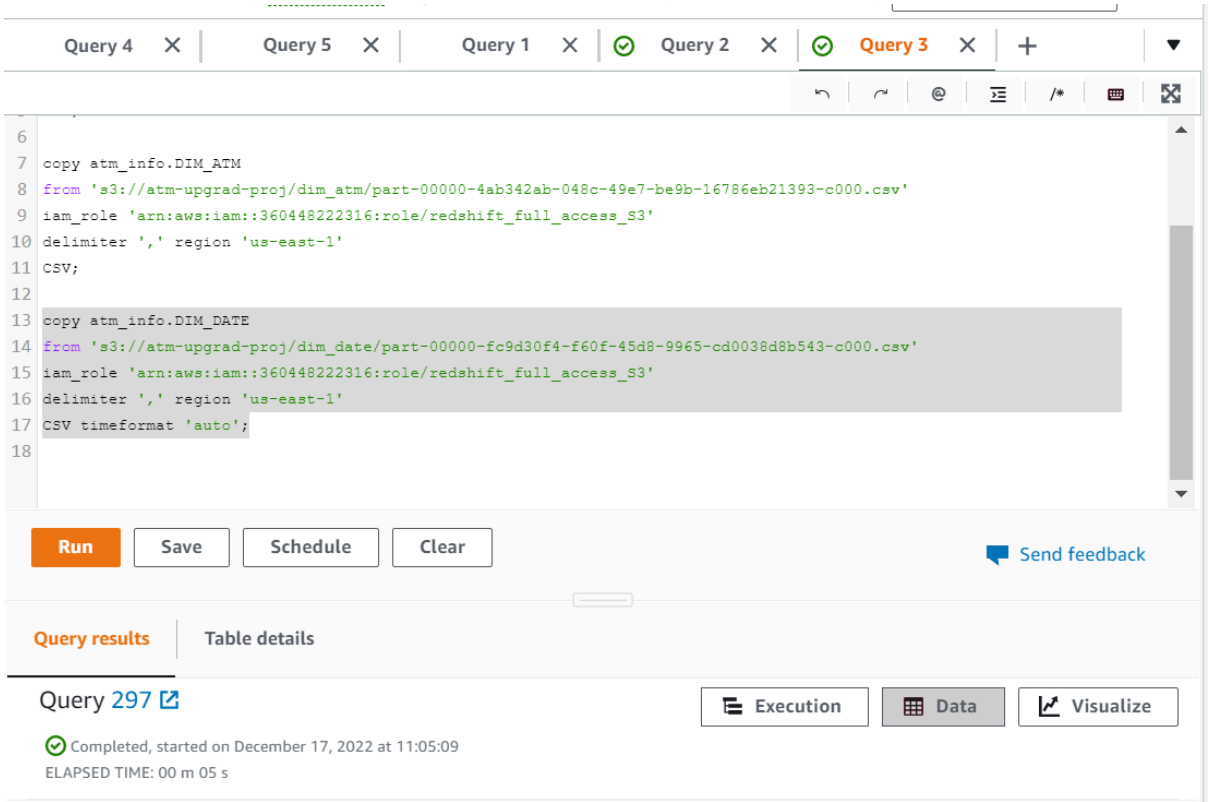
- Load data into DIM_ATM table

```
copy atm_info.DIM_ATM
from 's3://atm-upgrad-proj/dim_atm/part-00000-63c3a61e-ee5d-49e7-9e0e-5cd84c8b1f53-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
```



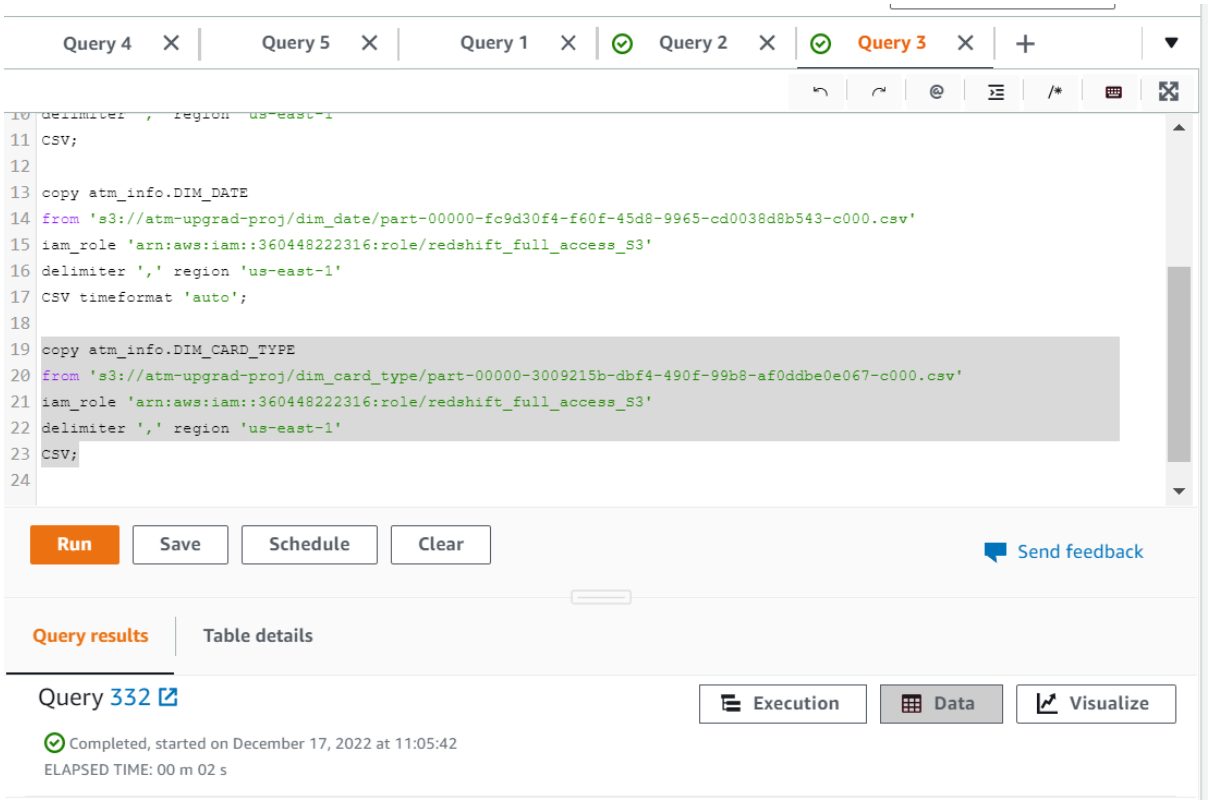
- Load data into DIM_DATE table

```
copy atm_info.DIM_DATE
from 's3://atm-upgrad-proj/dim_date/part-00000-f78dec33-e586-4af0-b409-f191a27758d7-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV timeformat 'auto';
```



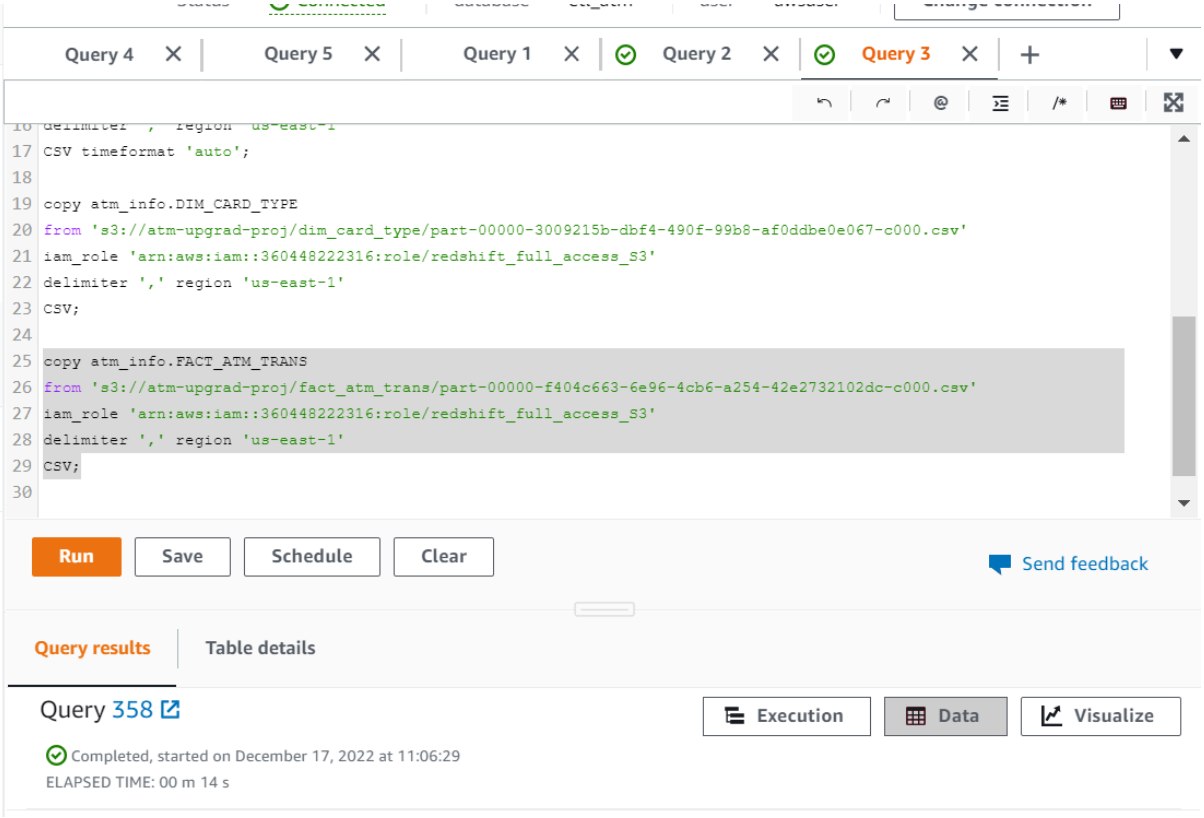
- Load data into DIM_CARD_TYPE table

```
copy atm_info.DIM_CARD_TYPE
from 's3://atm-upgrad-proj/dim_card_type/part-00000-adf96d01-6ff5-478a-afe4-1465f81ba096-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
```



- Load data into FACT_ATM_TRANS table

```
copy atm_info.FACT_ATM_TRANS
from 's3://atm-upgrad-proj/fact_atm_trans/part-00000-f05e0e52-e198-4d2c-bca2-9b3d3fd47f74-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
```


Query 4
Query 5
Query 1
Query 2
Query 3
delimiter ',' region 'us-east-1'
CSV timeformat 'auto';
copy atm_info.DIM_CARD_TYPE
from 's3://atm-upgrad-proj/dim_card_type/part-00000-3009215b-dbf4-490f-99b8-af0ddbe0e067-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
copy atm_info.FACT_ATM_TRANS
from 's3://atm-upgrad-proj/fact_atm_trans/part-00000-f404c663-6e96-4cb6-a254-42e2732102dc-c000.csv'
iam_role 'arn:aws:iam::360448222316:role/redshift_full_access_S3'
delimiter ',' region 'us-east-1'
CSV;
Run
Save
Schedule
Clear
Send feedback
Query results
Table details
Query 358
Execution
Data
Visualize
Completed, started on December 17, 2022 at 11:06:29
ELAPSED TIME: 00 m 14 s