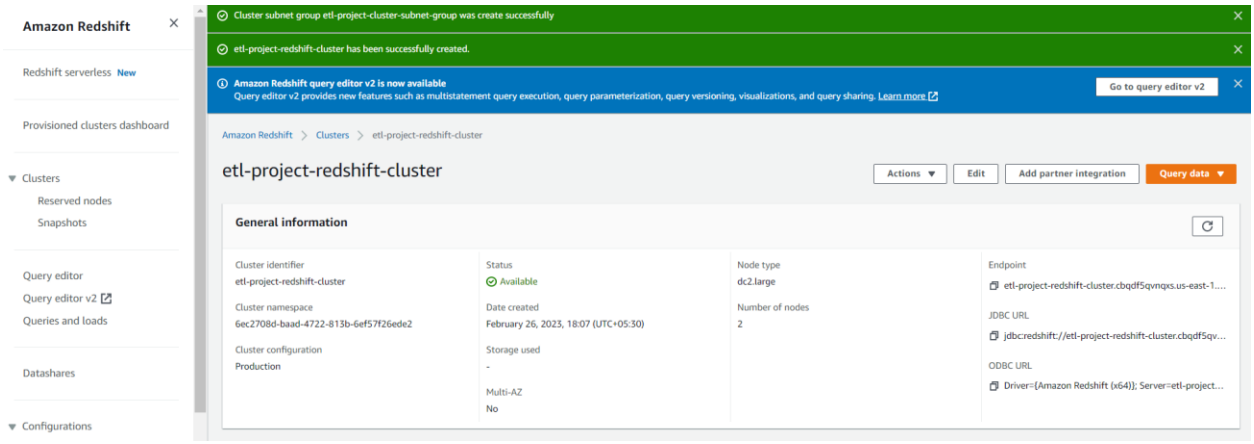
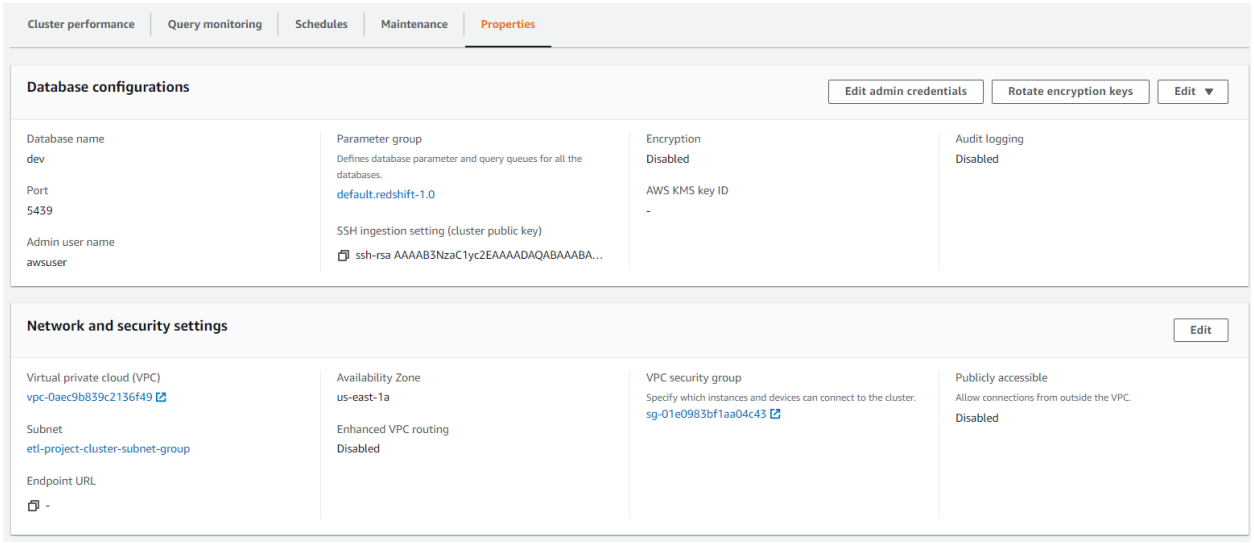
Creation of a Redshift Cluster

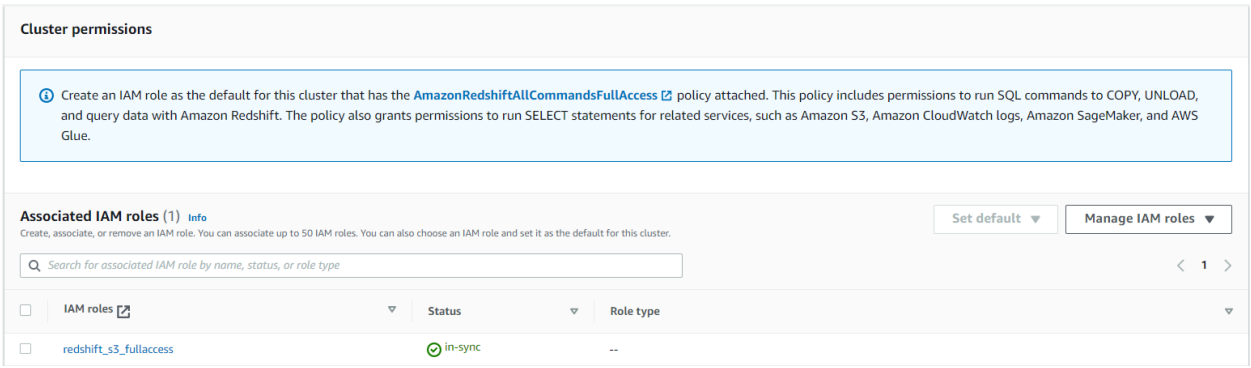
Screenshots of the configuration of the Redshift cluster that we have created:

Screenshot of type of machine used along with number of nodes:



Screenshots of various configurations associated with cluster creation:

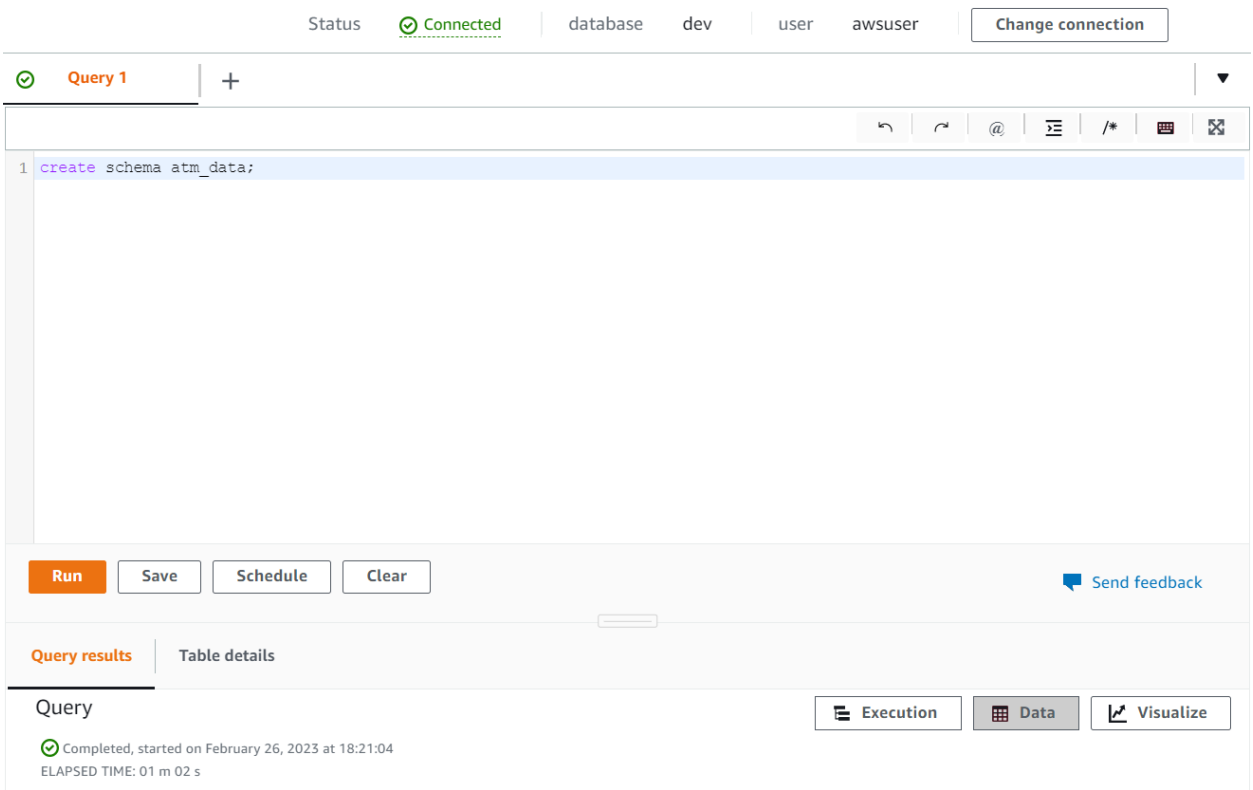




Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

Query to create schema for dimension and fact tables

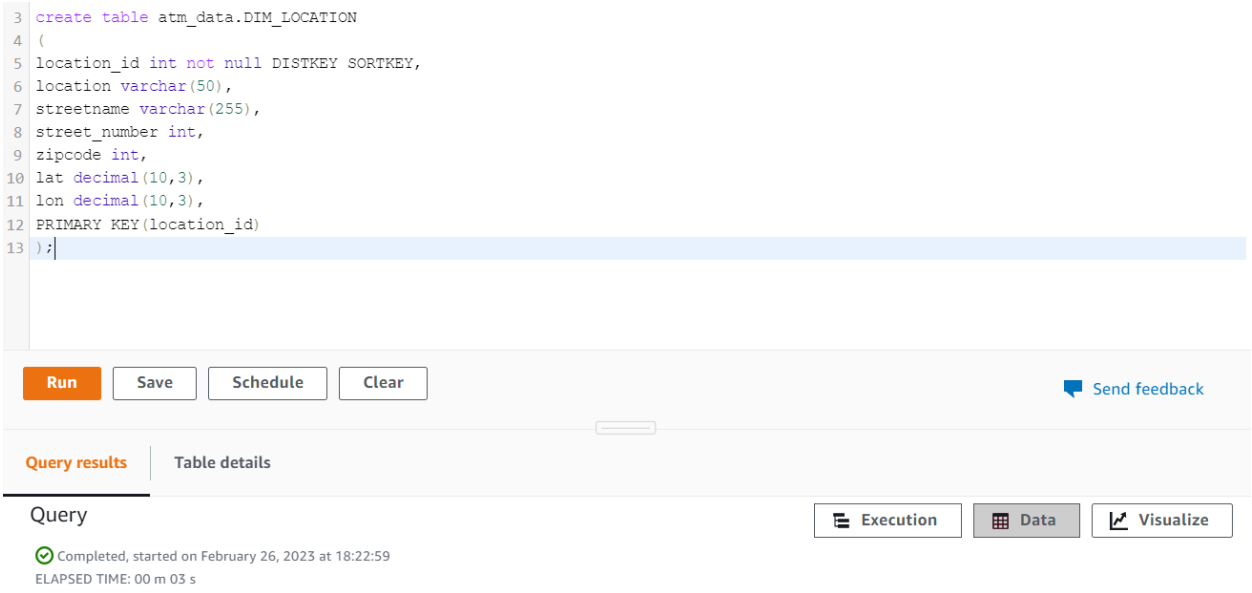
create schema atm_data;



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

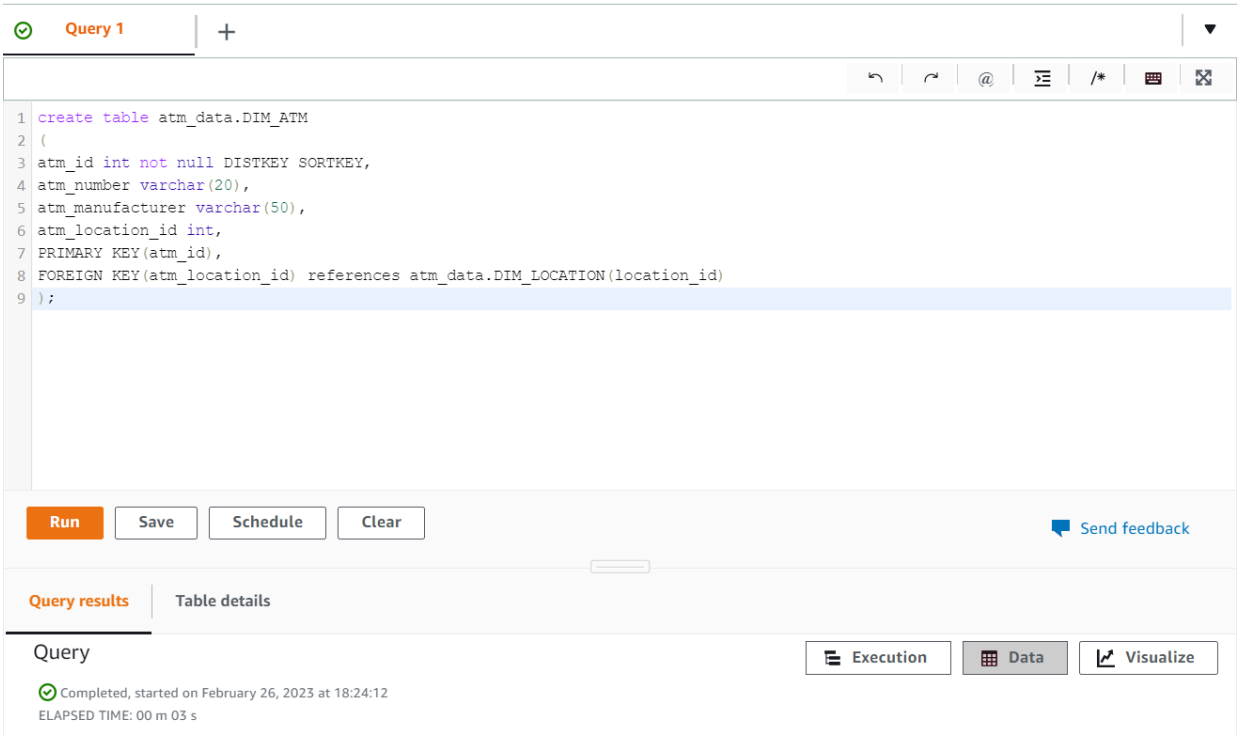
i) Creating Location dimension table

```
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)
);
```



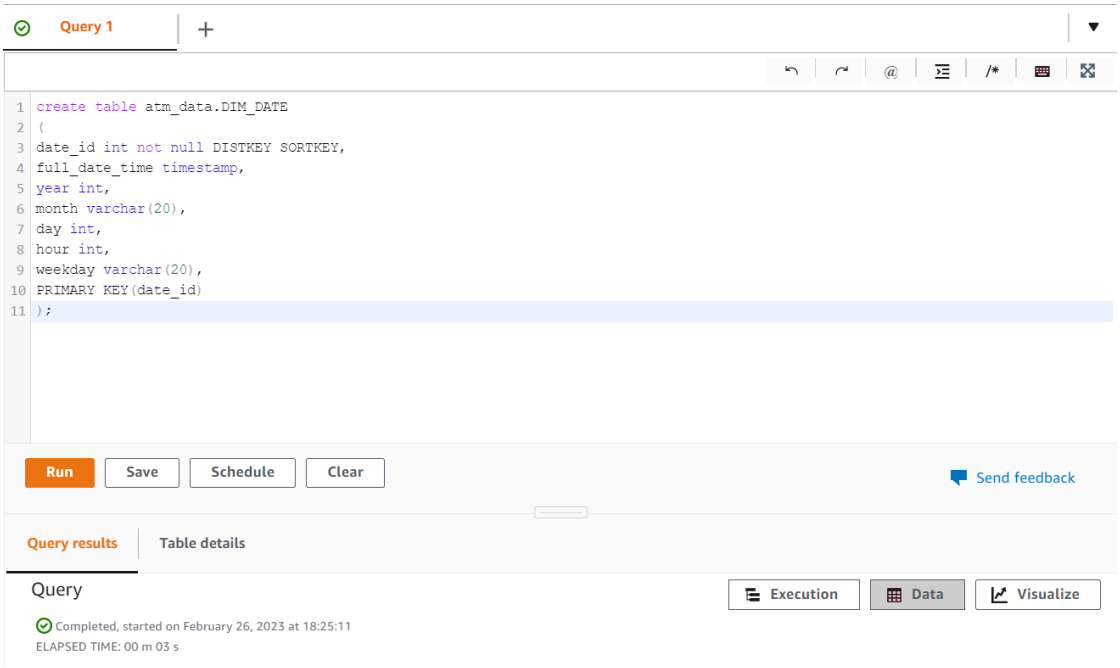
ii) Creating ATM dimension table

```
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)
);
```



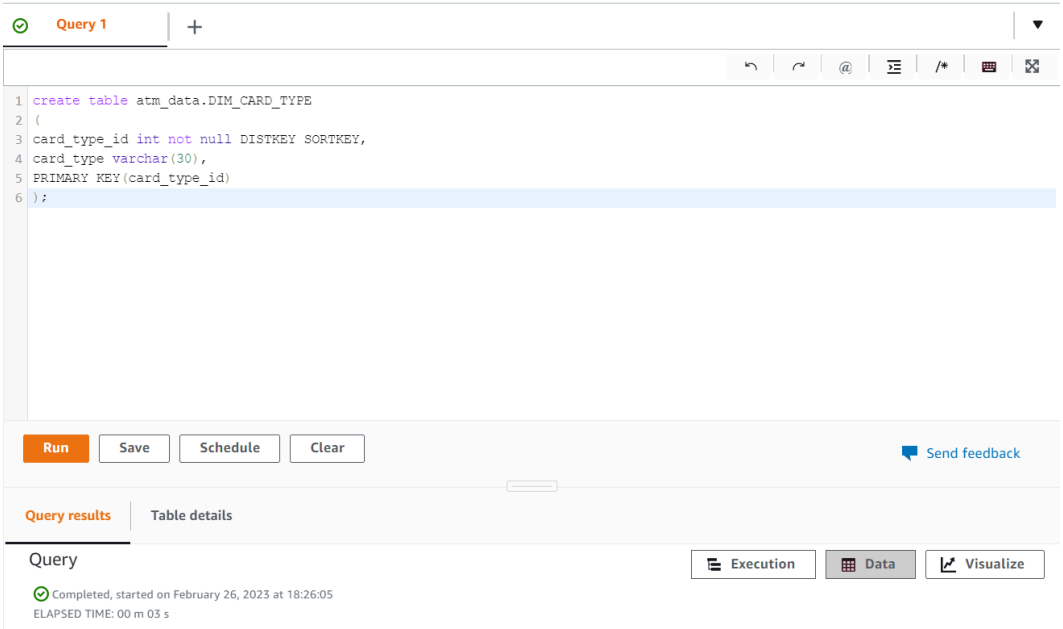
iii) Creating Date dimension table

```
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)
);
```



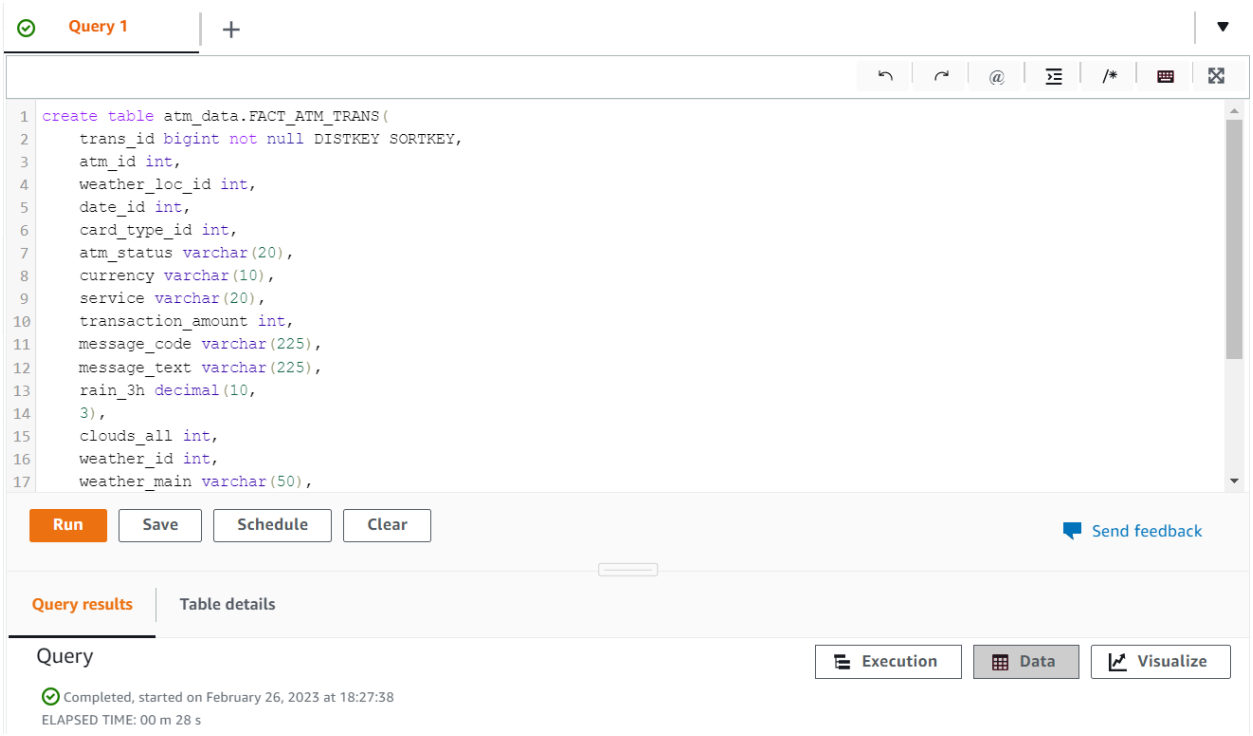
iv) Creating Card type dimension table

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



v) 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)
);
```

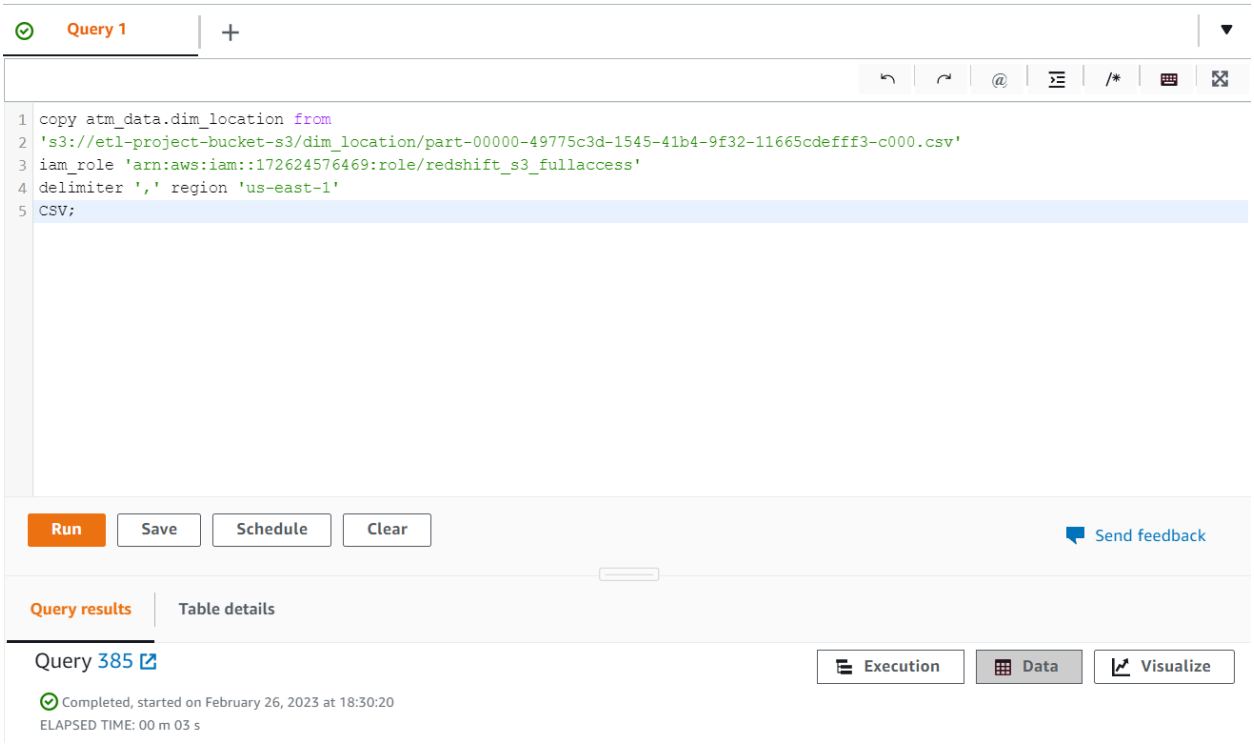


Loading data into a Redshift cluster from Amazon S3 bucket

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

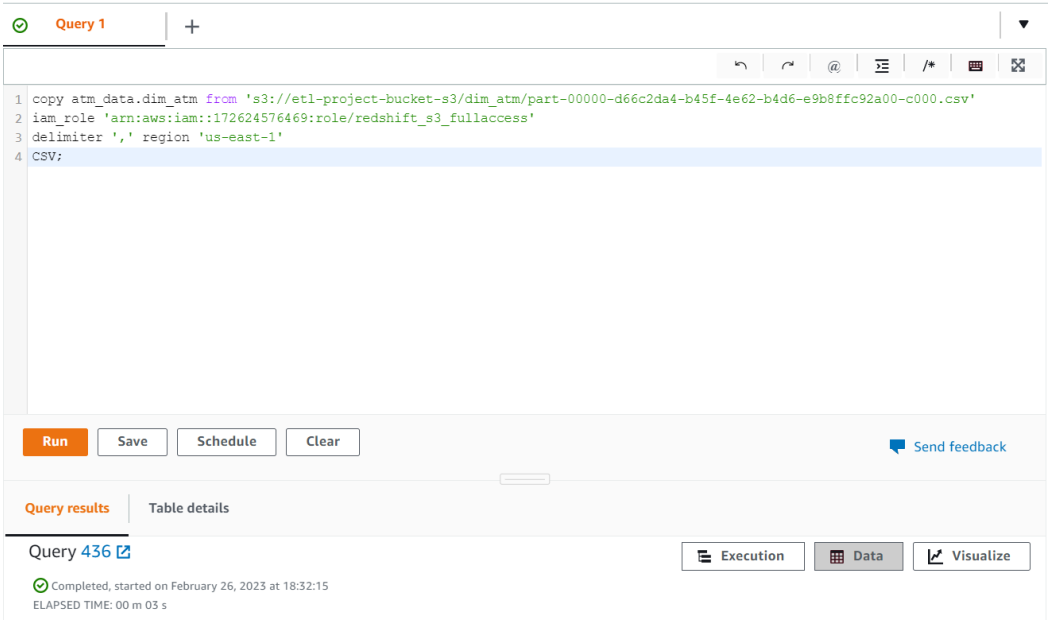
i) Copying data to dim_location table

```
copy atm_data.dim_location from 's3://etl-project-bucket-s3/dim_location/part-00000-49775c3d-1545-41b4-9f32-11665cdefff3-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



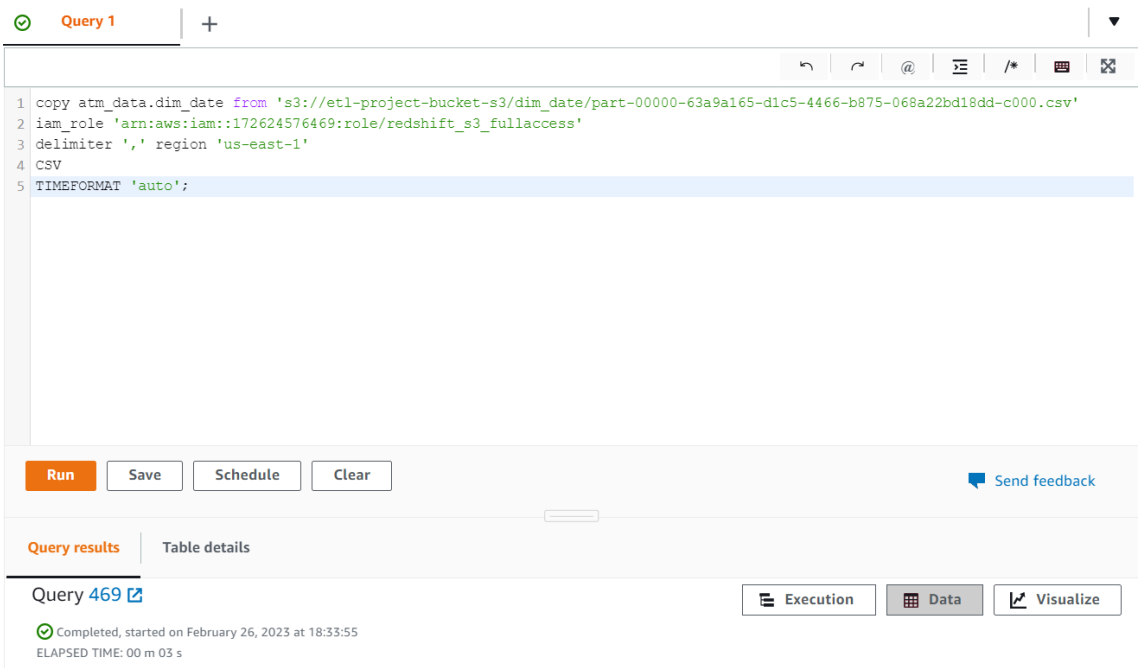
ii) Copying data to dim_atm table

```
copy atm_data.dim_atm from 's3://etl-project-bucket-s3/dim_atm/part-00000-d66c2da4-b45f-4e62-b4d6-e9b8ffc92a00-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



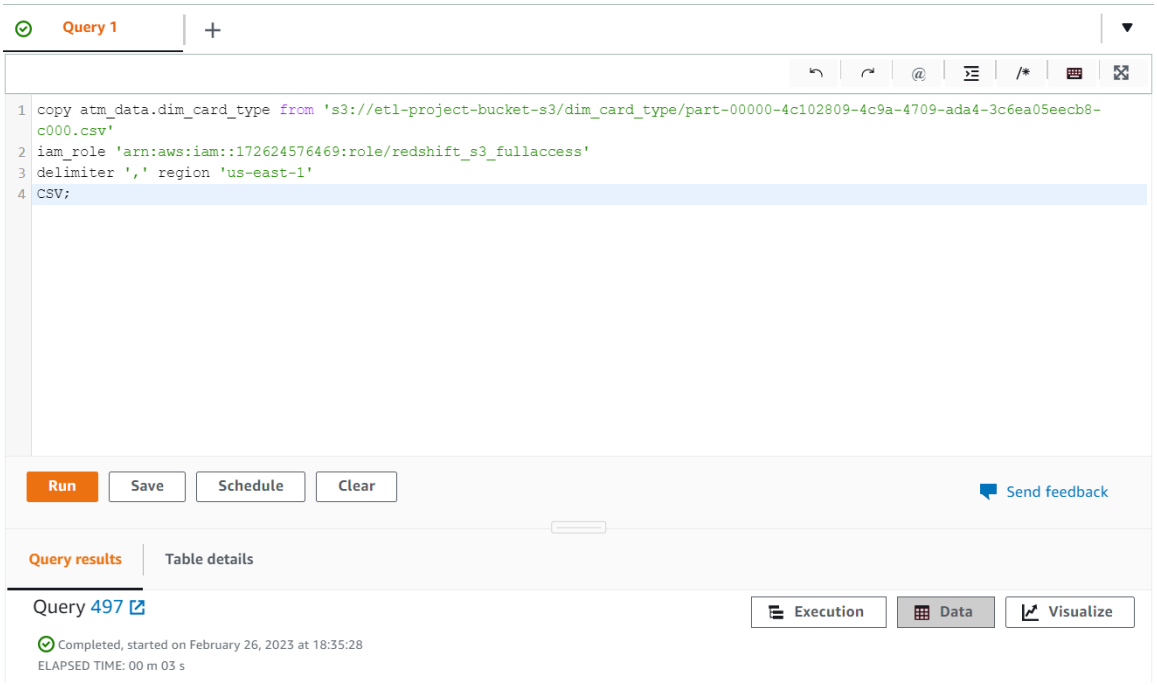
iii) Copying data to dim_date table

```
copy atm_data.dim_date from 's3://etl-project-bucket-s3/dim_date/part-00000-63a9a165-d1c5-4466-b875-068a22bd18dd-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';
```



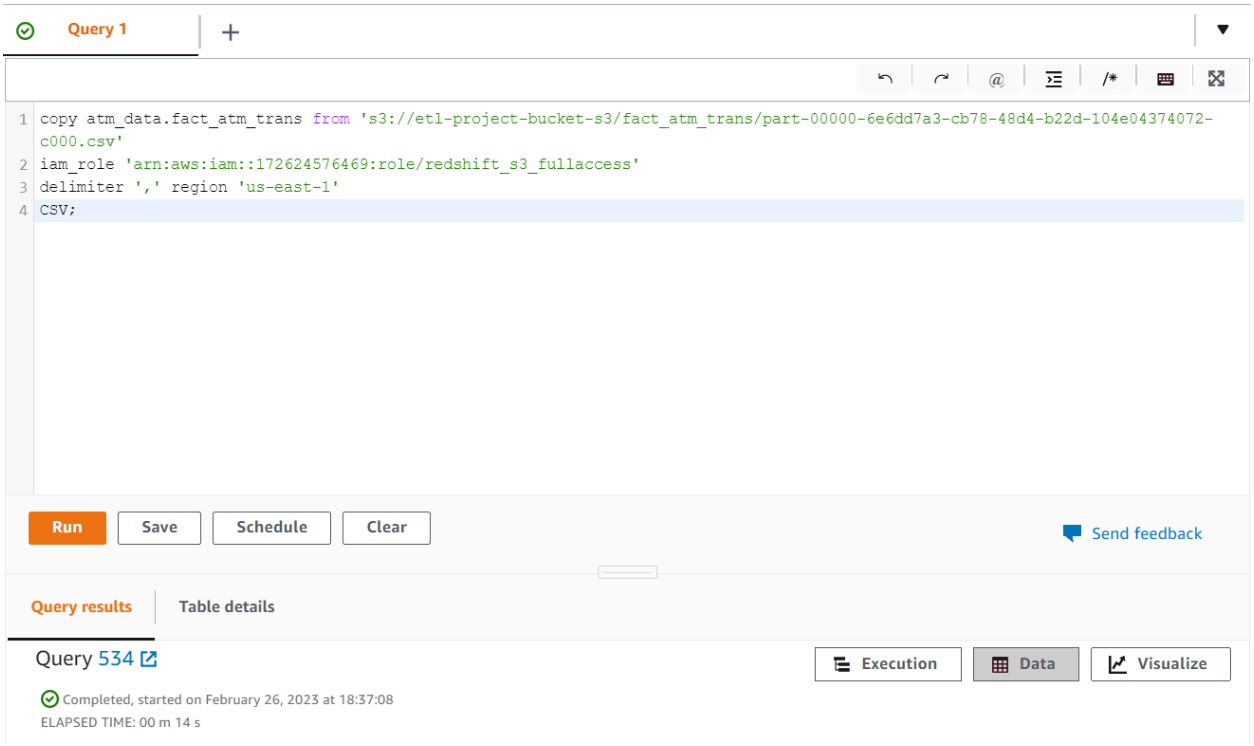
iv) Copying data to dim_card_type table

```
copy atm_data.dim_card_type from 's3://etl-project-bucket-s3/dim_card_type/part-00000-4c102809-4c9a-4709-ada4-3c6ea05eecb8-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



v) Copying data to fact_atm_trans table

```
copy atm_data.fact_atm_trans from 's3://etl-project-bucket-s3/fact_atm_trans/part-00000-6e6dd7a3-cb78-48d4-b22d-104e04374072-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```


Query 1
copy atm_data.fact_atm_trans from 's3://etl-project-bucket-s3/fact_atm_trans/part-00000-6e6dd7a3-cb78-48d4-b22d-104e04374072-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
Run
Save
Schedule
Clear
Send feedback
Query results
Table details
Query 534
Execution
Data
Visualize
Completed, started on February 26, 2023 at 18:37:08
ELAPSED TIME: 00 m 14 s