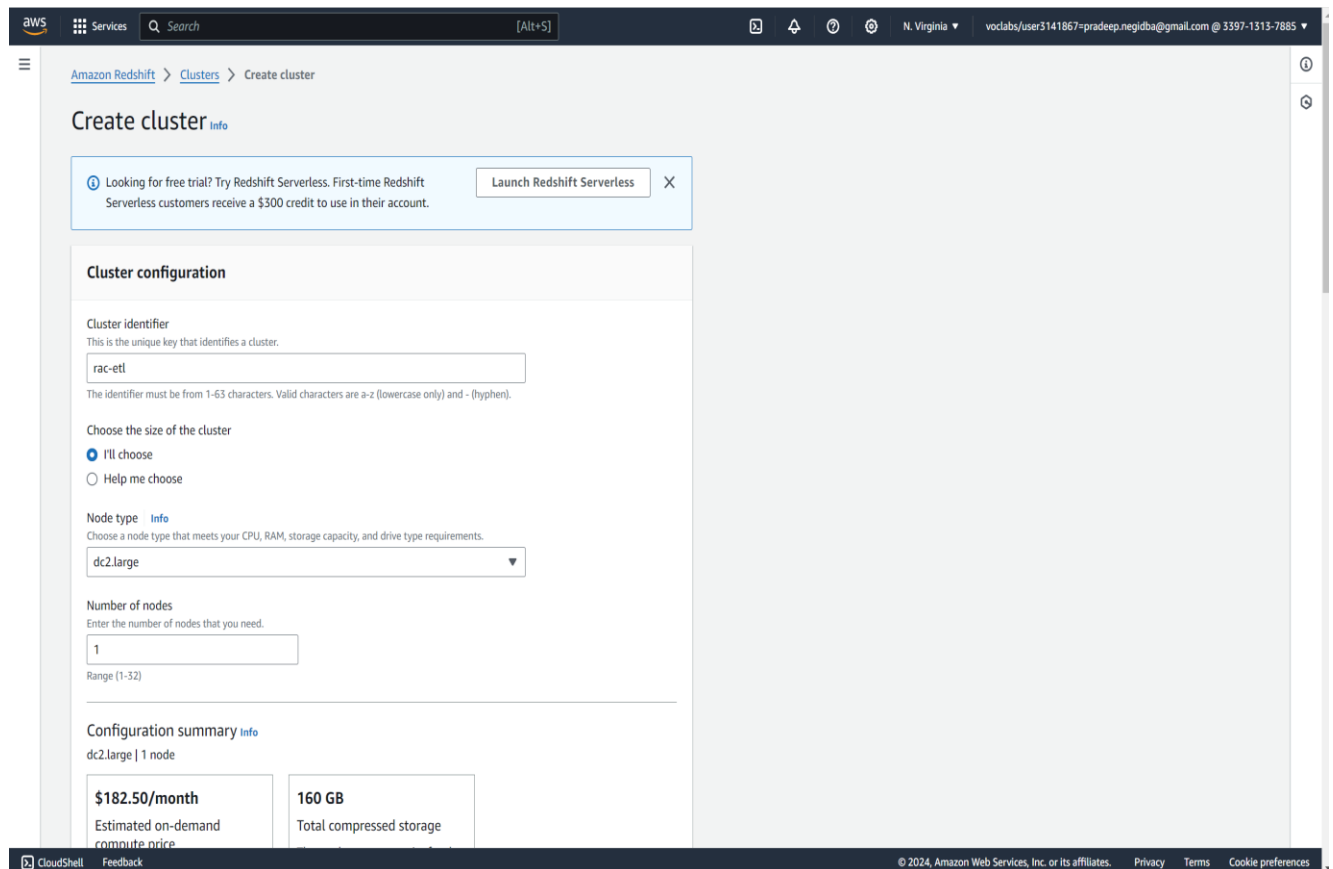


Creation of a Redshift Cluster

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



aws Services Search [Alt+S] N. Virginia voclabs/user3141867-pradeep.negidba@gmail.com @ 3397-1313-7885

Amazon Redshift > Clusters > Create cluster

Create cluster [Info](#)

Looking for free trial? Try Redshift Serverless. First-time Redshift Serverless customers receive a \$300 credit to use in their account. [Launch Redshift Serverless](#)

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.

The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

Choose the size of the cluster
☒ I'll choose
☐ Help me choose

Node type [Info](#)
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.

Number of nodes
Enter the number of nodes that you need.

Range (1-32)

Configuration summary [Info](#)

dc2.large | 1 node

\$182.50/month Estimated on-demand compute price	160 GB Total compressed storage
--	---

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

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

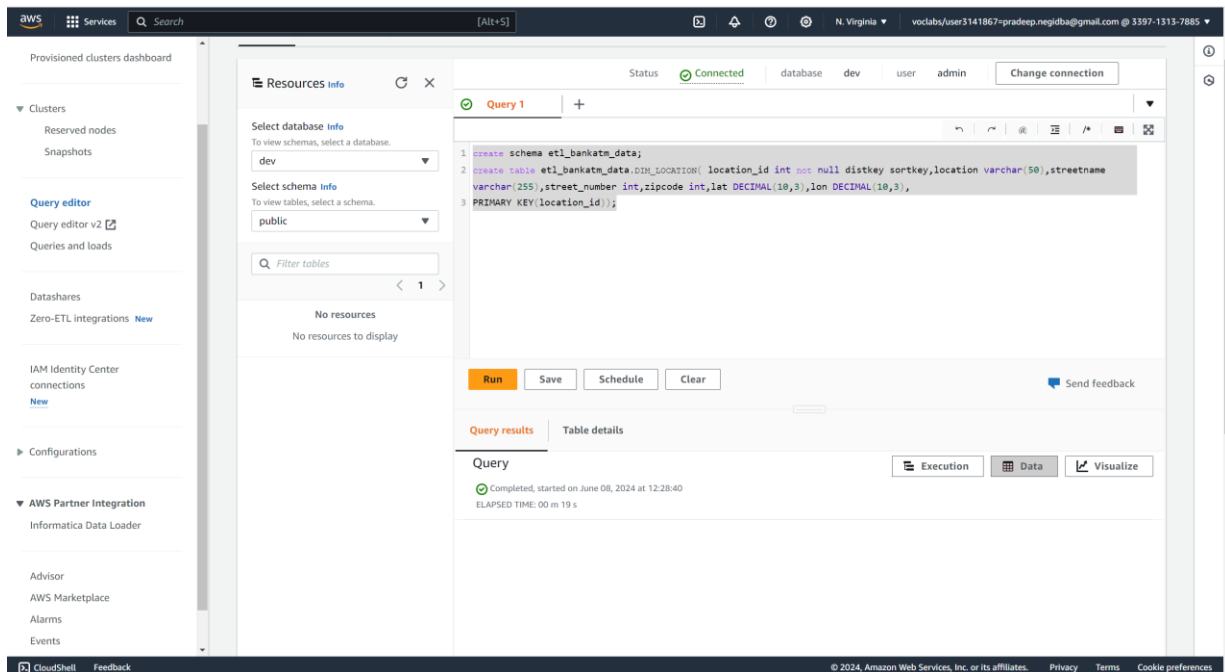
Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

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

create schema etl_bankatm_data;

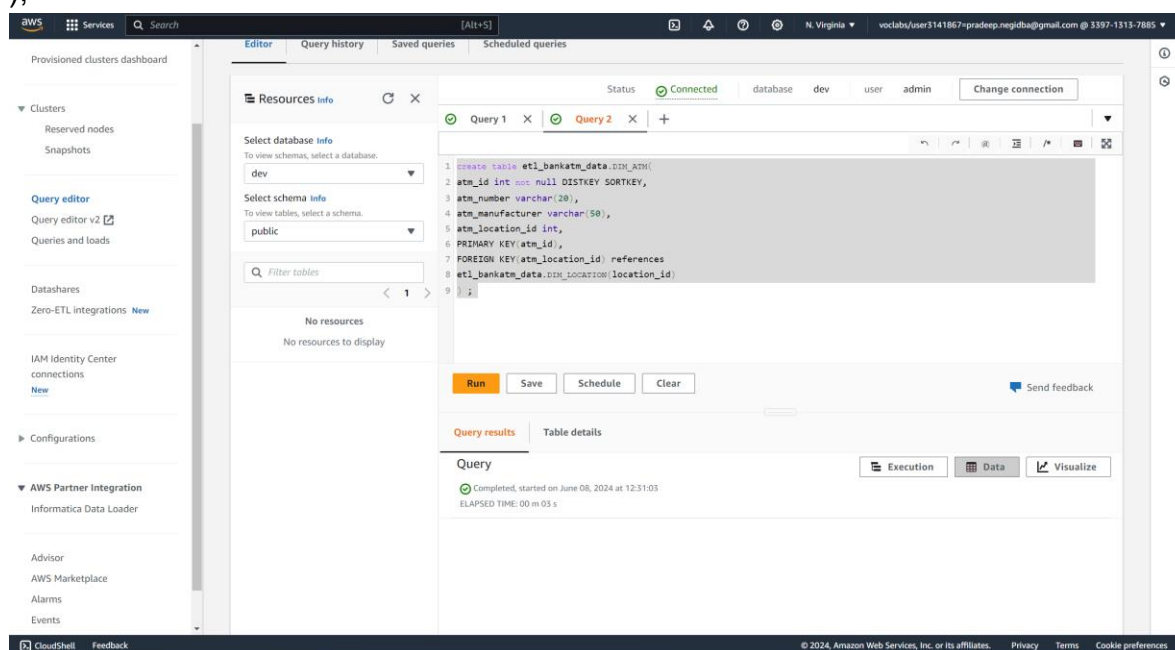
1) Location Dimension Table

Create table etl_bankatm_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)
);



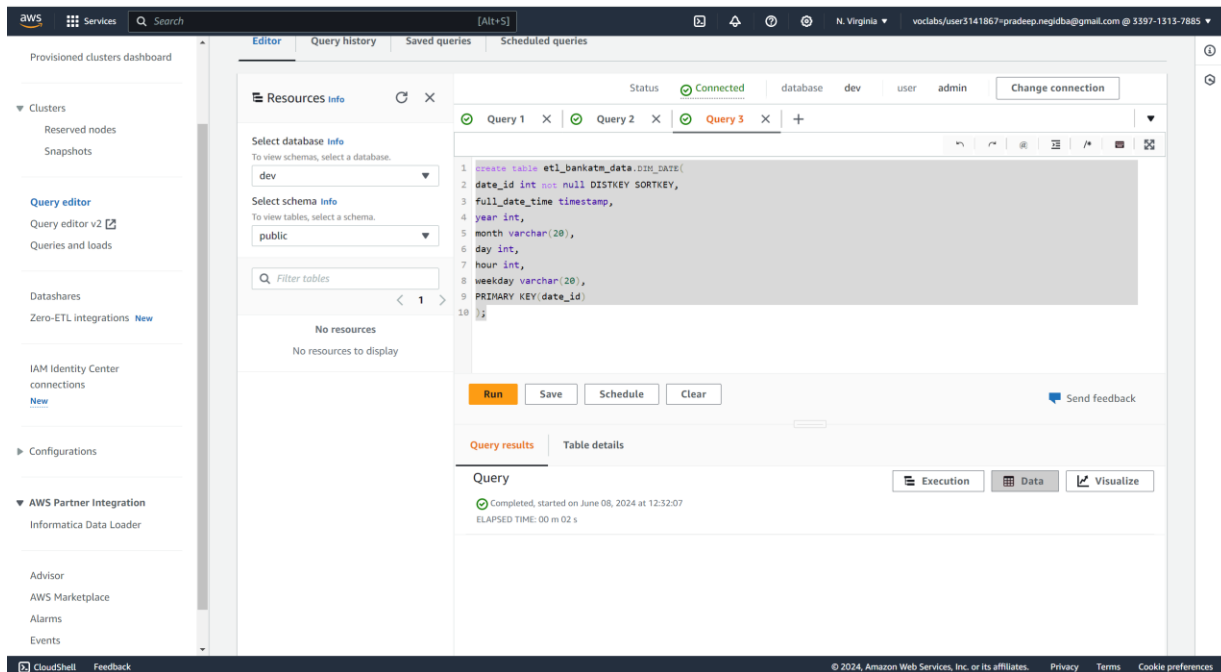
2) ATM Dimension table

Create table etl_bankatm_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
 etl_bankatm_data.DIM_LOCATION(location_id)
);



3) Date Dimension table

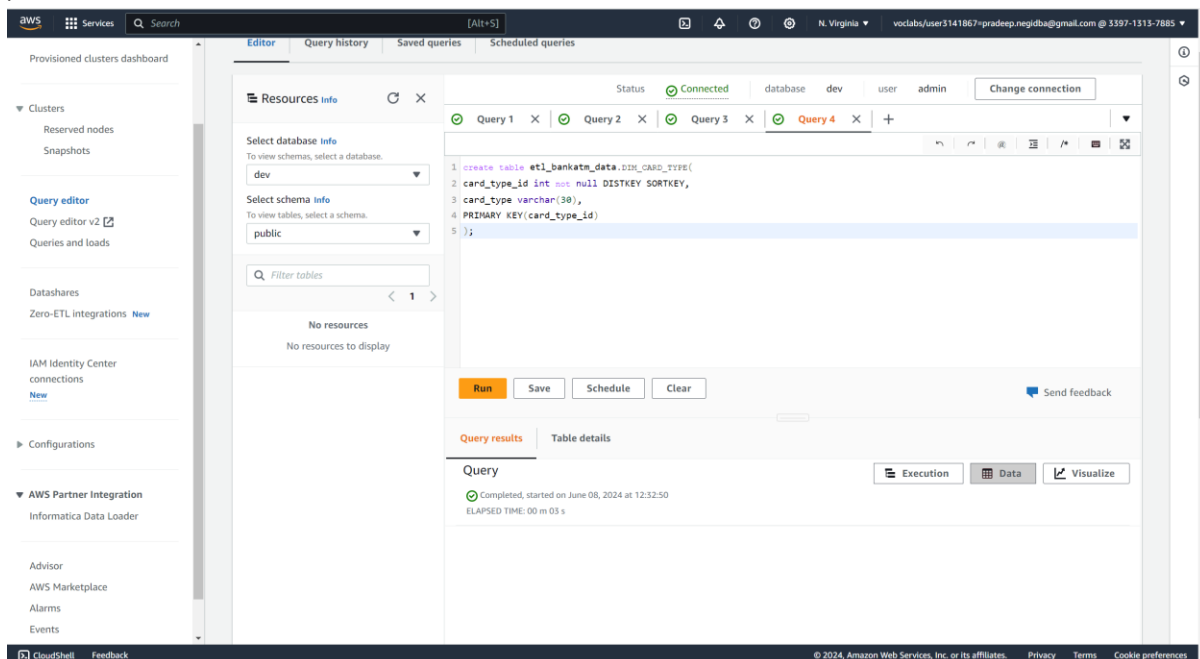
Create table etl_bankatm_data.DIM_DATE(
 date_id int not null DISTKEY SORTKEY,
 full_date_time timestamp,
 year int,
 month varchar(20),
 day int,
 hour int,
 weekday varchar(20),
 PRIMARY KEY(date_id)
);



The screenshot shows the AWS Glue console interface. On the left is a navigation menu with options like Clusters, Query editor, and AWS Partner Integration. The main area is titled 'Editor' and shows 'Query 3' being edited. The query text is: `create table etl_bankatm_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));`. Below the editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' section shows the query is 'Completed, started on June 08, 2024 at 12:32:07' with an 'ELAPSED TIME: 00 m 02 s'.

4) Card Type dimension table

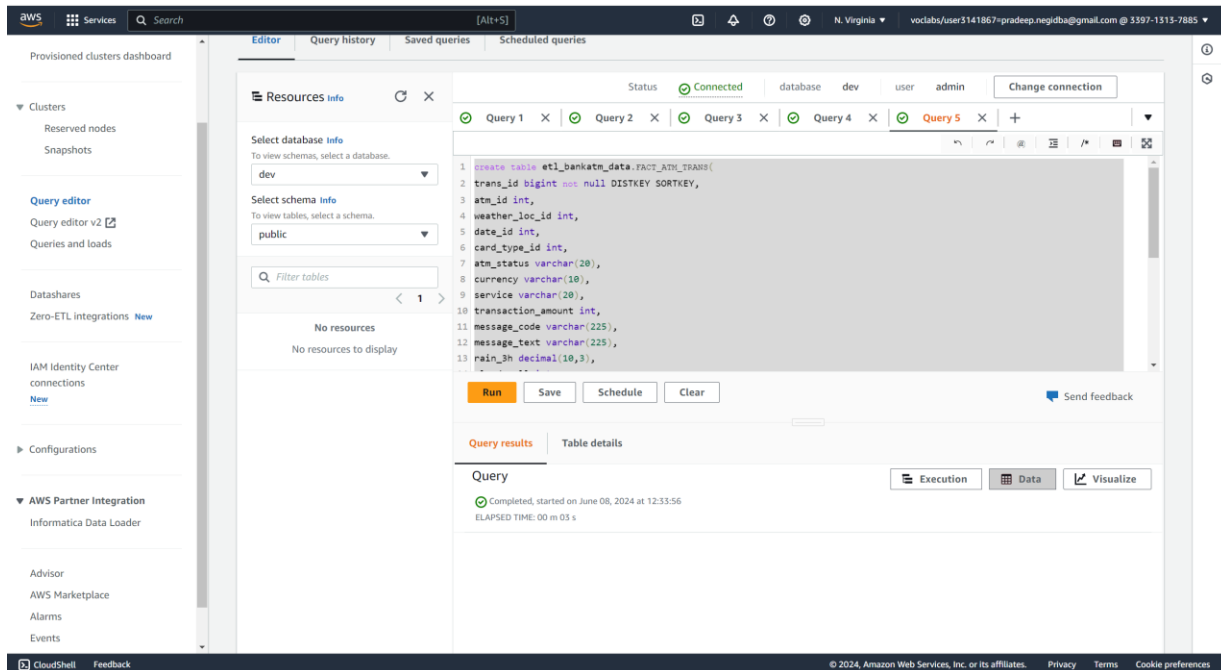
Create table etl_bankatm_data.DIM_CARD_TYPE(
card_type_id int not null DISTKEY SORTKEY,
card_type varchar(30),
PRIMARY KEY(card_type_id)
);



The screenshot shows the AWS Glue console interface. On the left is a navigation menu. The main area is titled 'Editor' and shows 'Query 4' being edited. The query text is: `create table etl_bankatm_data.DIM_CARD_TYPE(card_type_id int not null DISTKEY SORTKEY, card_type varchar(30), PRIMARY KEY(card_type_id));`. Below the editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' section shows the query is 'Completed, started on June 08, 2024 at 12:32:50' with an 'ELAPSED TIME: 00 m 03 s'.

5) ATM_TRANS fact table

```
create table etl_bankatm_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  
etl_bankatm_data.DIM_LOCATION(location_id),  
  FOREIGN KEY(atm_id) references etl_bankatm_data.DIM_ATM(atm_id),  
  FOREIGN KEY(date_id) references etl_bankatm_data.DIM_DATE(date_id),  
  FOREIGN KEY(card_type_id) references  
etl_bankatm_data.DIM_CARD_TYPE(card_type_id)  
);
```

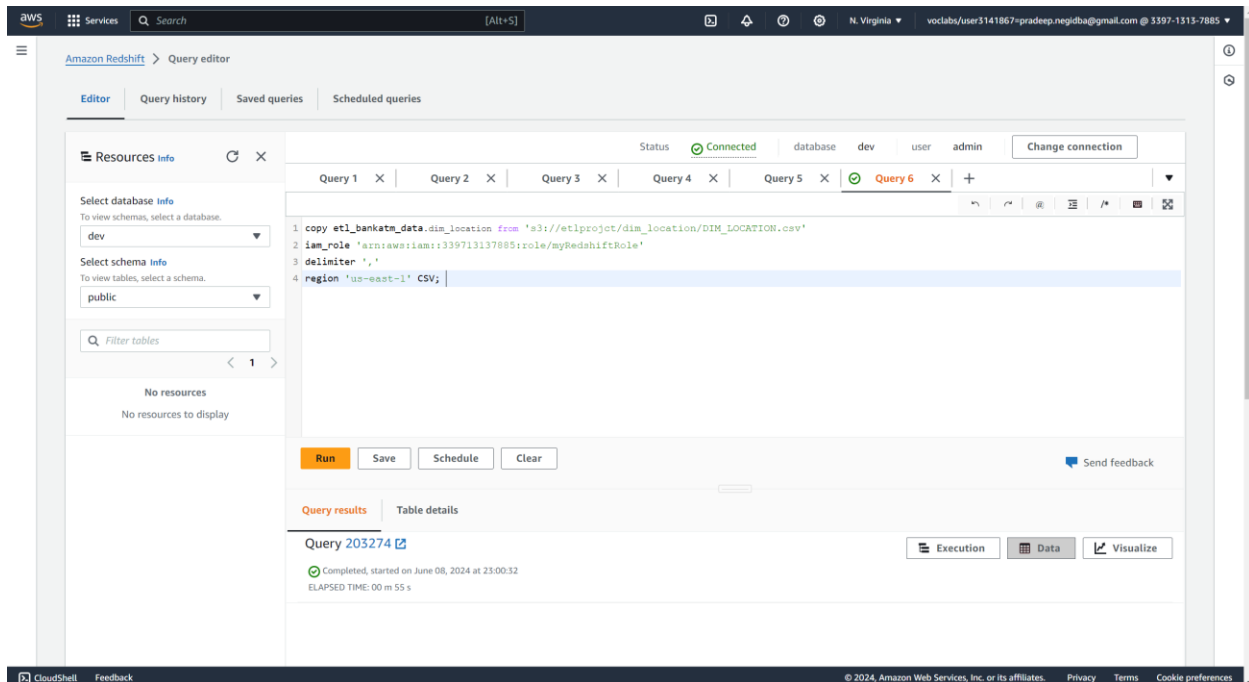


Loading data into a Redshift cluster from Amazon S3 bucket

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

1) Loading Data into dim_location table

copy etl_bankatm_data.dim_location from 's3://etlprojct/dim_location/DIM_LOCATION.csv'
iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV



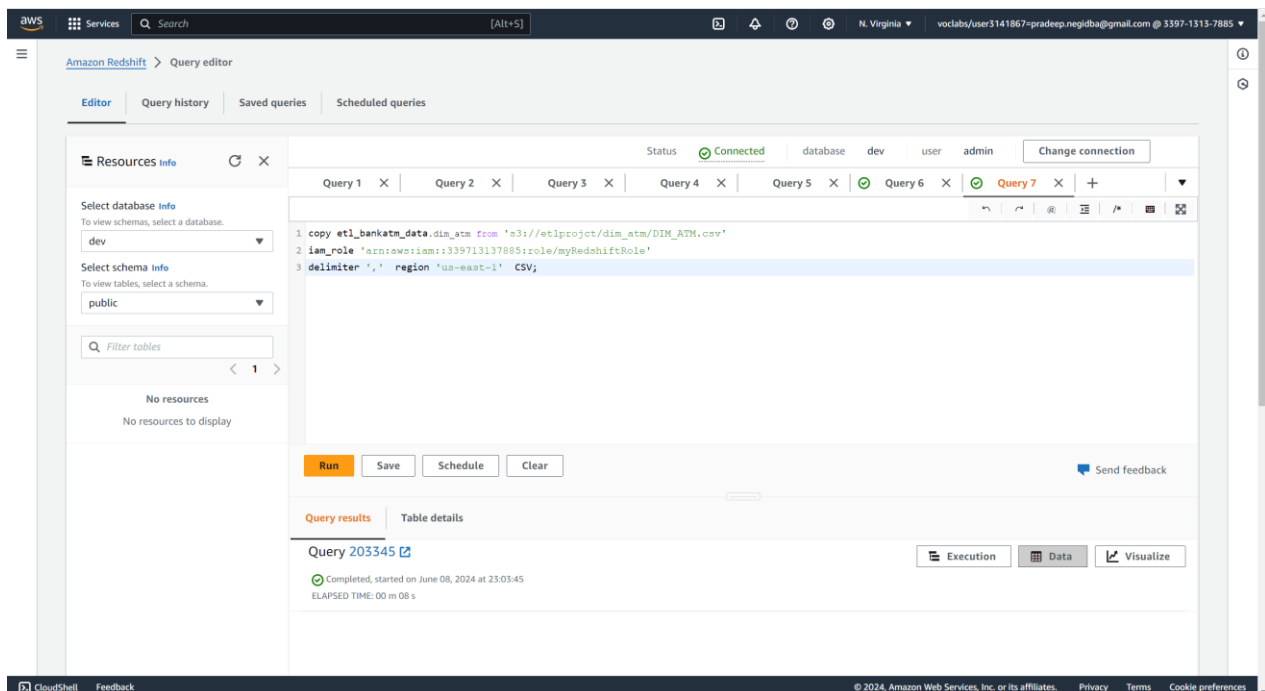
The screenshot shows the Amazon Redshift Query Editor interface. On the left, the 'Resources' panel is open, showing 'dev' as the selected database and 'public' as the selected schema. The main editor area displays the following SQL query for Query 6:

```
1 copy etl_bankatm_data.dim_location from 's3://etlprojct/dim_location/DIM_LOCATION.csv'
2 iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole'
3 delimiter ','
4 region 'us-east-1' CSV;
```

Below the query, the 'Run' button is highlighted. The 'Query results' tab is active, showing the status 'Completed, started on June 08, 2024 at 23:00:32' and 'ELAPSED TIME: 00 m 55 s'.

2) Loading Data into dim_atm table

copy etl_bankatm_data.dim_atm from 's3://etlprojct/dim_atm/DIM_ATM.csv' iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole' delimiter ',' region 'useast-1' CSV;



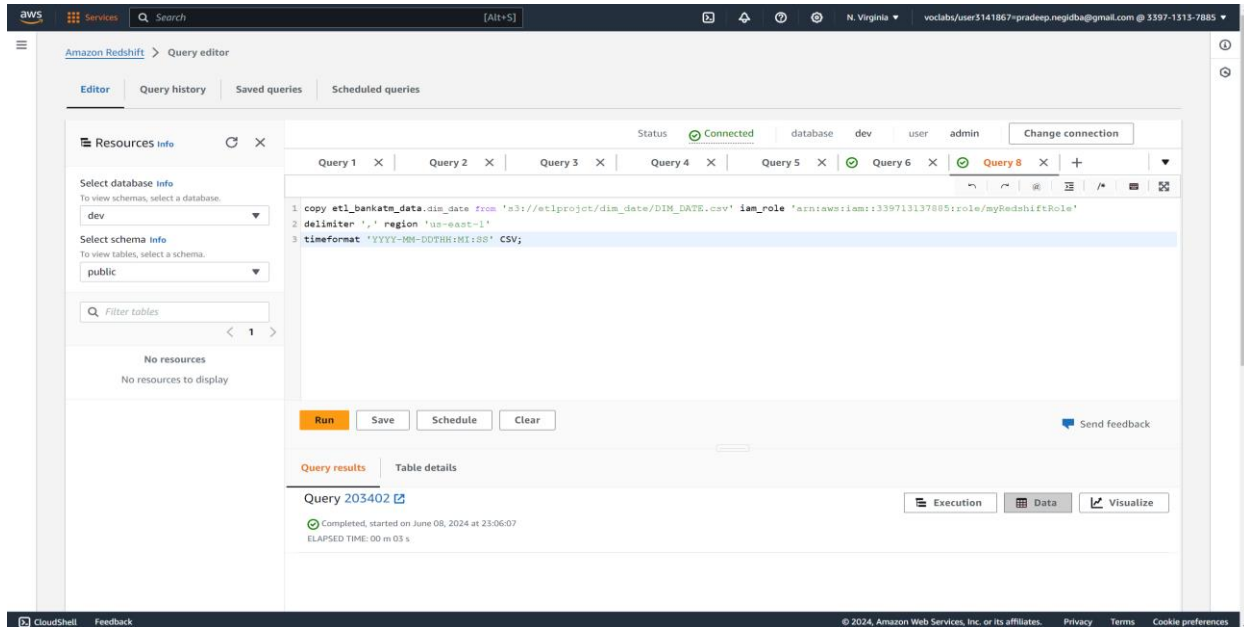
The screenshot shows the Amazon Redshift Query Editor interface. On the left, the 'Resources' panel is open, showing 'dev' as the selected database and 'public' as the selected schema. The main editor area displays the following SQL query for Query 7:

```
1 copy etl_bankatm_data.dim_atm from 's3://etlprojct/dim_atm/DIM_ATM.csv'
2 iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole'
3 delimiter ',' region 'us-east-1' CSV;
```

Below the query, the 'Run' button is highlighted. The 'Query results' tab is active, showing the status 'Completed, started on June 08, 2024 at 23:03:45' and 'ELAPSED TIME: 00 m 08 s'.

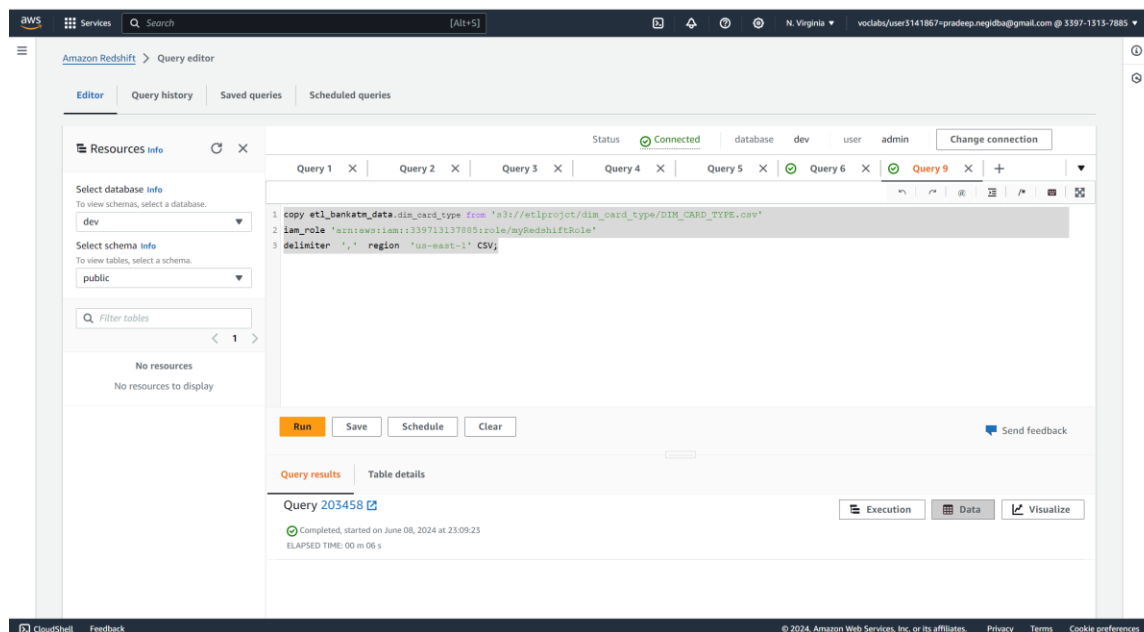
3) Loading data into dim_date table

copy etl_bankatm_data.dim_date from 's3://etlprojct/dim_date/DIM_DATE.csv' iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole' delimiter ',' region 'us-east-1' timeformat 'YYYY-MM-DDTHH:MI:SS' CSV;



4) Loading data into dim_card_type table

copy etl_bankatm_data.dim_card_type from 's3://etlprojct/dim_card_type/DIM_CARD_TYPE.csv' iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;



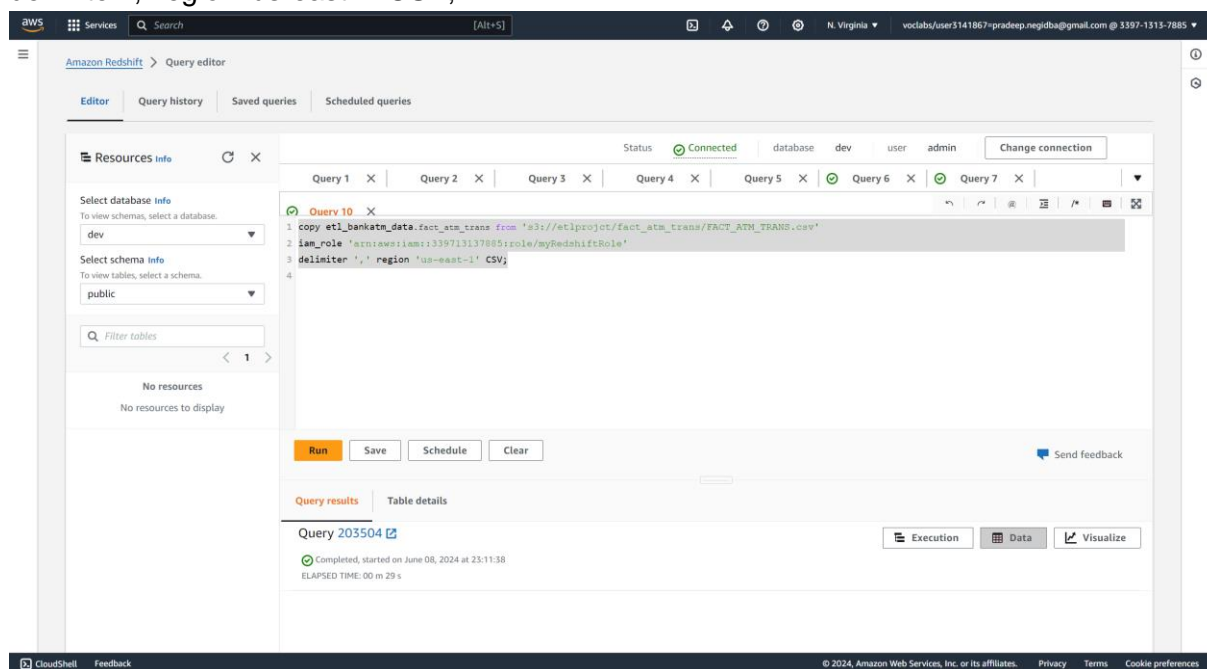
The screenshot shows the Amazon Redshift Query Editor interface. The left sidebar contains the 'Resources' panel with 'dev' database and 'public' schema selected. The main editor area shows a SQL query for Query 9:

```
1 copy etl_bankatm_data.dim_card_type from 's3://etlprojct/dim_card_type/DIM_CARD_TYPE.csv'
2 iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole'
3 delimiter ',' region 'us-east-1' CSV;
```

Below the query editor, the 'Query results' section shows that Query 203458 has completed successfully on June 08, 2024 at 23:09:23, with an elapsed time of 00 m 06 s.

5) Loading Data into fact_atm_trans table

copy etl_bankatm_data.fact_atm_trans from
's3://etlprojct/fact_atm_trans/FACT_ATM_TRANS.csv'
iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole'
delimiter ',' region 'us-east-1' CSV;



The screenshot shows the Amazon Redshift Query Editor interface. The left sidebar contains the 'Resources' panel with 'dev' database and 'public' schema selected. The main editor area shows a SQL query for Query 10:

```
1 copy etl_bankatm_data.fact_atm_trans from 's3://etlprojct/fact_atm_trans/FACT_ATM_TRANS.csv'
2 iam_role 'arn:aws:iam::339713137885:role/myRedshiftRole'
3 delimiter ',' region 'us-east-1' CSV;
```

Below the query editor, the 'Query results' section shows that Query 203504 has completed successfully on June 08, 2024 at 23:11:38, with an elapsed time of 00 m 29 s.