

Creation of a Redshift Cluster

Screenshots of Redshift Cluster :

[Alt+S]

N. Virginia
voclabs/user2686176=parthgupta7267@gmail.com @ 8960-8248-7071

Amazon Redshift > Clusters > atm-etl-pipeline-cluster

atm-etl-pipeline-cluster

Actions
Edit
Add partner integration
Query data

General information Info

Cluster identifier

atm-etl-pipeline-cluster

Custom domain name

-

Cluster namespace ARN

arn:aws:redshift:us-east-1:896082487071:namespace:3ab7f27b-f546-47ca-ac33-49588350b907

Cluster configuration

Production

Status

Available

Date created

November 02, 2023, 09:38 (UTC+05:30)

Storage used

-

Multi-AZ

No

Node type

dc2.large

Number of nodes

2

Endpoint

atm-etl-pipeline-cluster.cambwdbgtbpe.us-east-1.redshift.amazonaws.com:5437/dev

JDBC URL

jdbc:redshift://atm-etl-pipeline-cluster.cambwdbgtbpe.us-east-1.redshift.amazonaws.com:5437/dev

ODBC URL

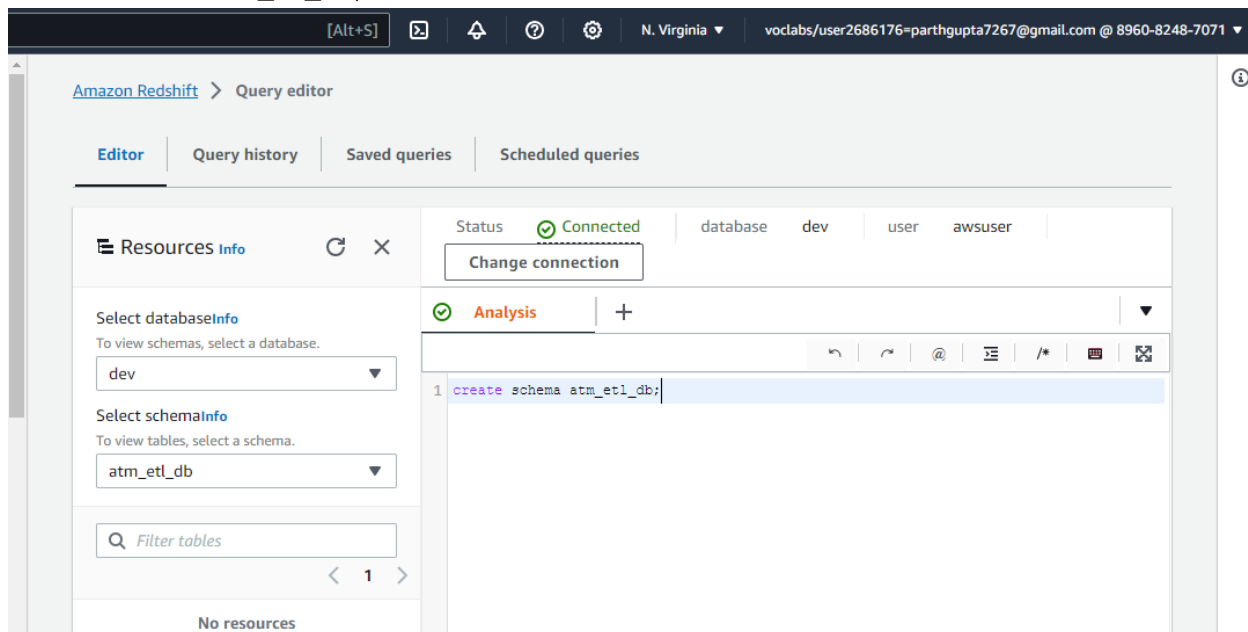
Driver={Amazon Redshift (x64)}; Server=atm-etl-pipeline-cluster.cambwdbgtbpe.us-east-

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:

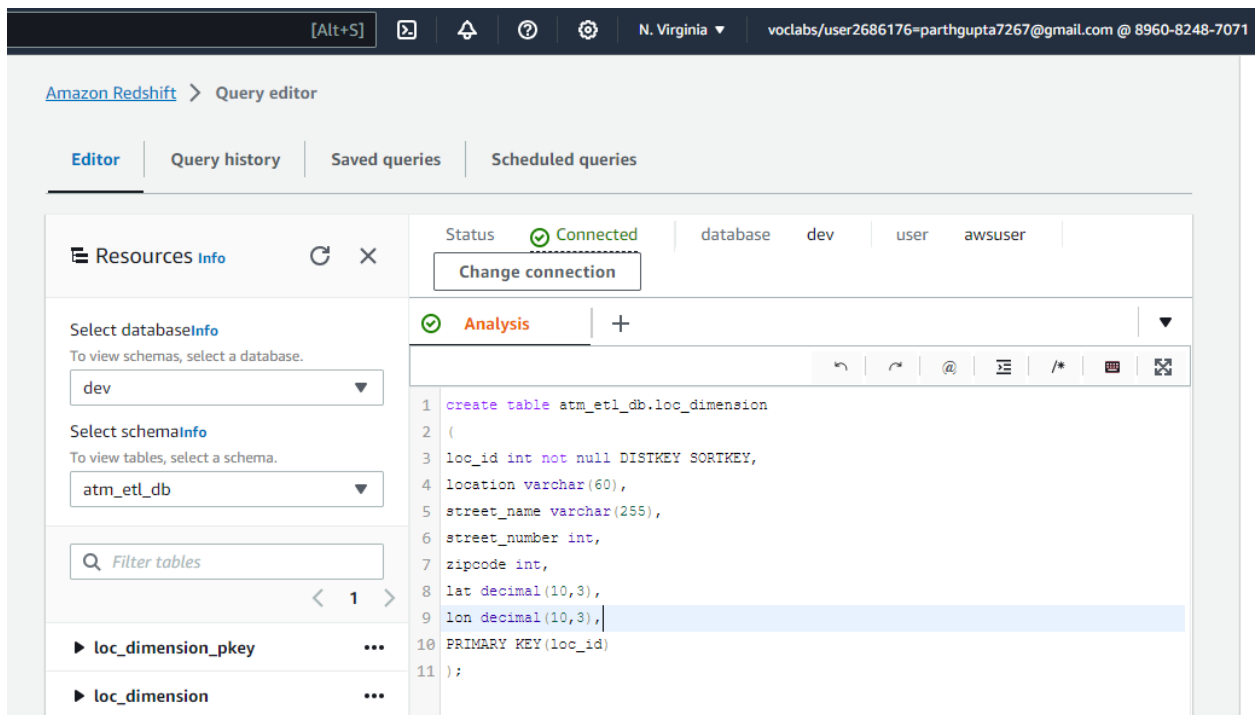
1) Creation of Schema:

```
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)
);
```

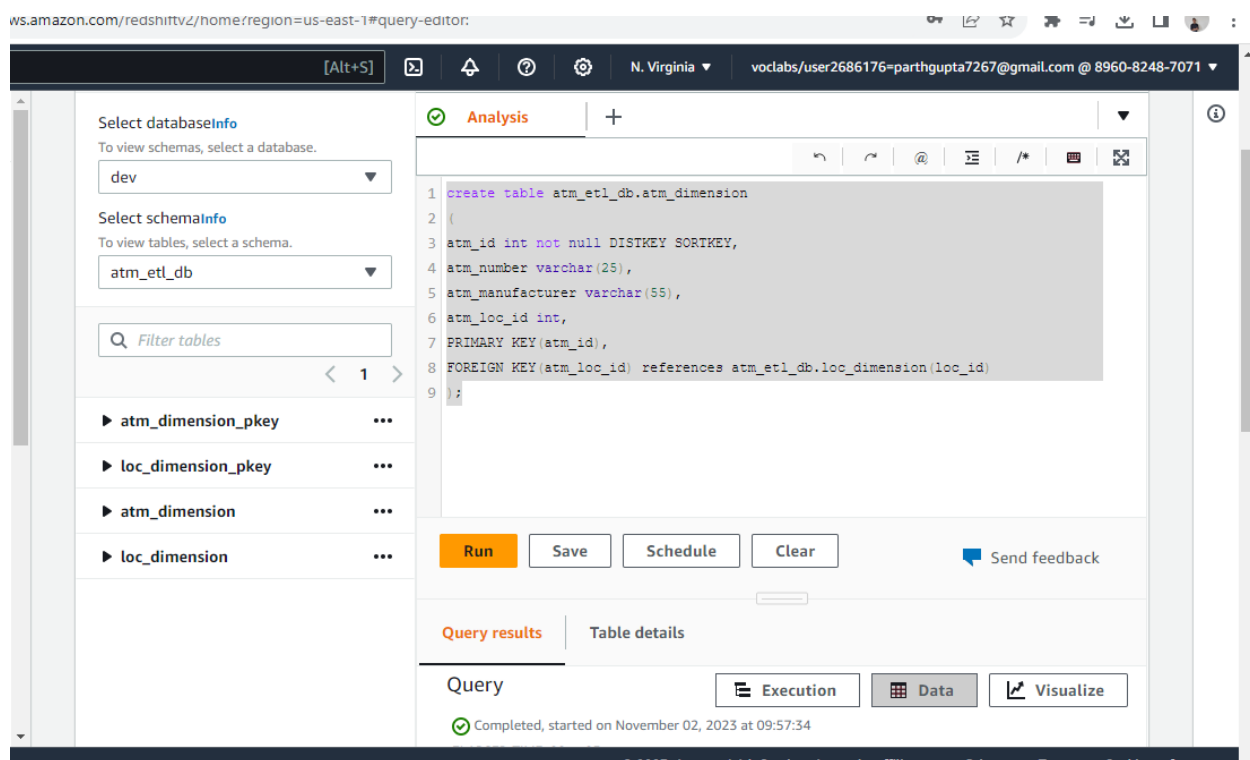


The screenshot shows the Amazon Redshift Query Editor interface. The top bar includes a status bar with [Alt+S], a help icon, a settings icon, a location dropdown (N. Virginia), and a user email (voclabs/user2686176-parthgupta7267@gmail.com @ 8960-8248-7071). The main interface has a navigation bar with 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. The 'Editor' tab is active, showing a 'Resources Info' panel on the left and a 'Query editor' on the right. The 'Resources Info' panel includes 'Select database' (dev), 'Select schema' (atm_etl_db), and a 'Filter tables' search bar. Below this, a list of tables is shown: 'loc_dimension_pkey' and 'loc_dimension'. The 'Query editor' shows a SQL query being executed, with the status 'Connected' and 'Analysis' indicated. The query is:


```
1 create table atm_etl_db.loc_dimension
2 (
3 loc_id int not null DISTKEY SORTKEY,
4 location varchar(60),
5 street_name varchar(255),
6 street_number int,
7 zipcode int,
8 lat decimal(10,3),
9 lon decimal(10,3),
10 PRIMARY KEY(loc_id)
11 );
```

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)
);
```



The screenshot shows the Amazon Redshift Query Editor interface. The left sidebar displays the database structure with the following tables listed:

- atm_dimension_pkey
- loc_dimension_pkey
- atm_dimension
- loc_dimension

The main editor area shows the following SQL query:

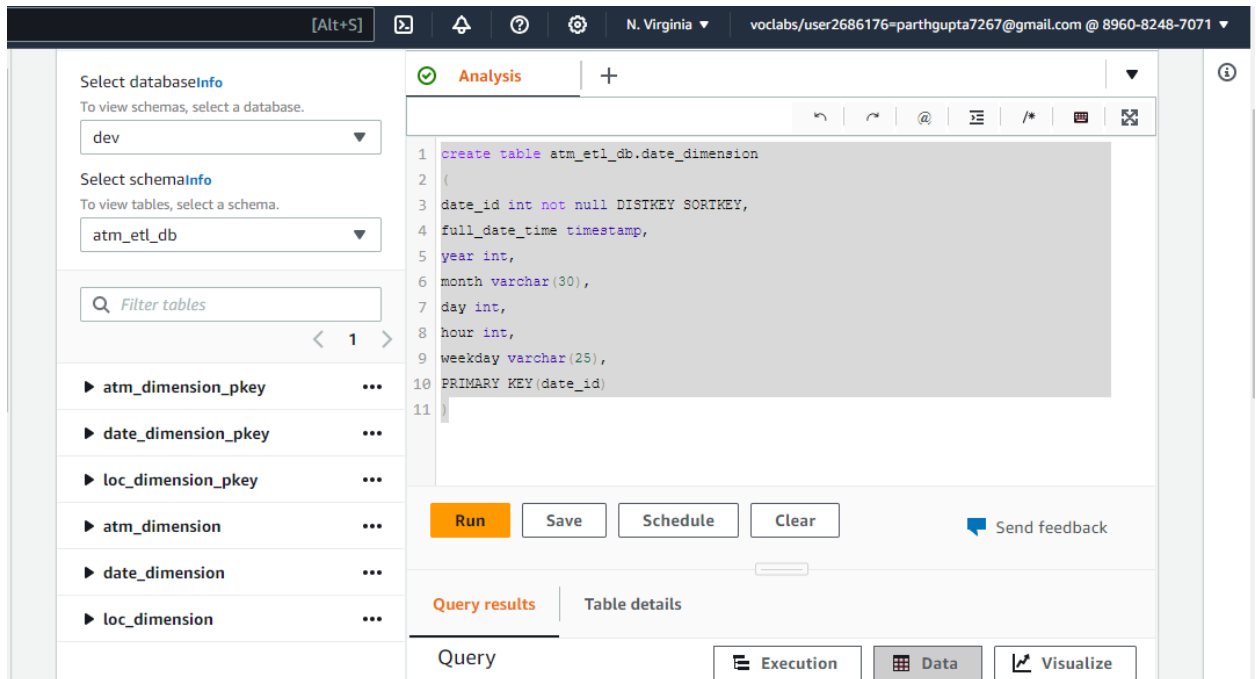
```
1 create table atm_etl_db.atm_dimension
2 (
3 atm_id int not null DISTKEY SORTKEY,
4 atm_number varchar(25),
5 atm_manufacturer varchar(55),
6 atm_loc_id int,
7 PRIMARY KEY(atm_id),
8 FOREIGN KEY(atm_loc_id) references atm_etl_db.loc_dimension(loc_id)
9 );
```

Below the query editor, there are buttons for "Run", "Save", "Schedule", and "Clear". The "Run" button is highlighted in orange. To the right of these buttons is a "Send feedback" link.

At the bottom, the "Query results" tab is active, showing a status message: "Completed, started on November 02, 2023 at 09:57:34". There are also buttons for "Execution", "Data", and "Visualize".

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)
);
```



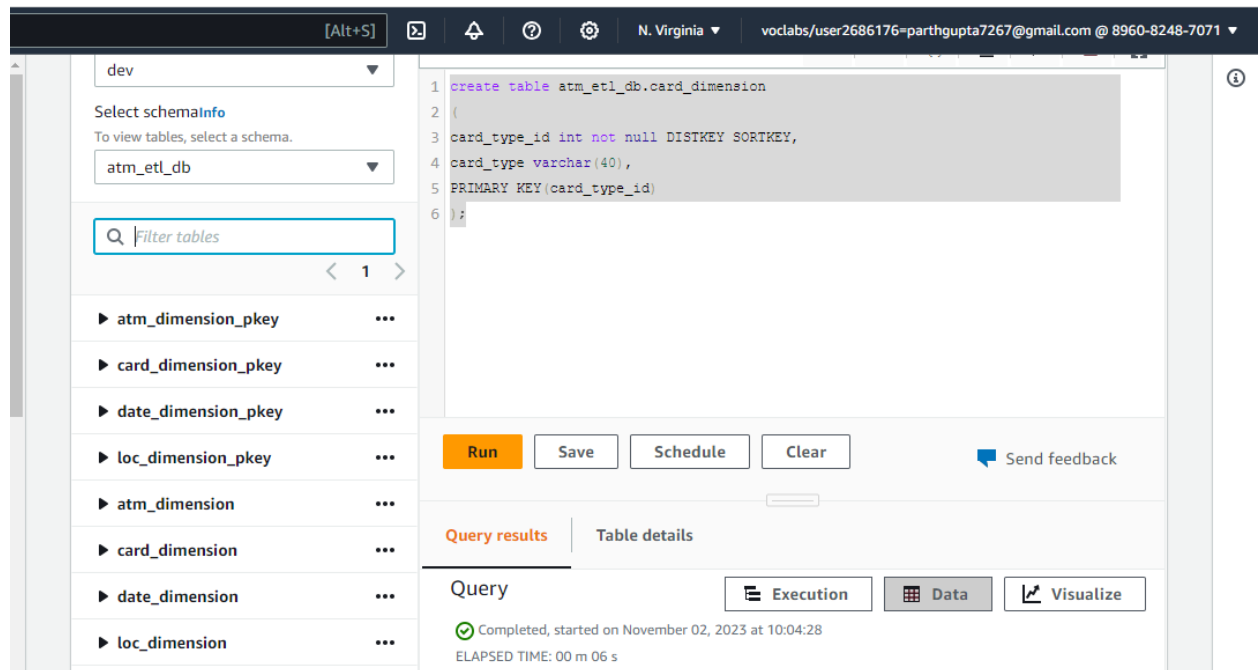
The screenshot shows a SQL query editor interface. On the left, there's a sidebar with a tree view showing the database structure. The main area displays a SQL query to create a table named 'date_dimension' in the 'atm_etl_db' database. The query is as follows:

```
1 create table atm_etl_db.date_dimension
2 (
3 date_id int not null DISTKEY SORTKEY,
4 full_date_time timestamp,
5 year int,
6 month varchar(30),
7 day int,
8 hour int,
9 weekday varchar(25),
10 PRIMARY KEY(date_id)
11 );
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. To the right of these buttons is a 'Send feedback' link. At the bottom, there are tabs for 'Query results' and 'Table details', and buttons for 'Execution', 'Data', and 'Visualize'.

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)
);
```



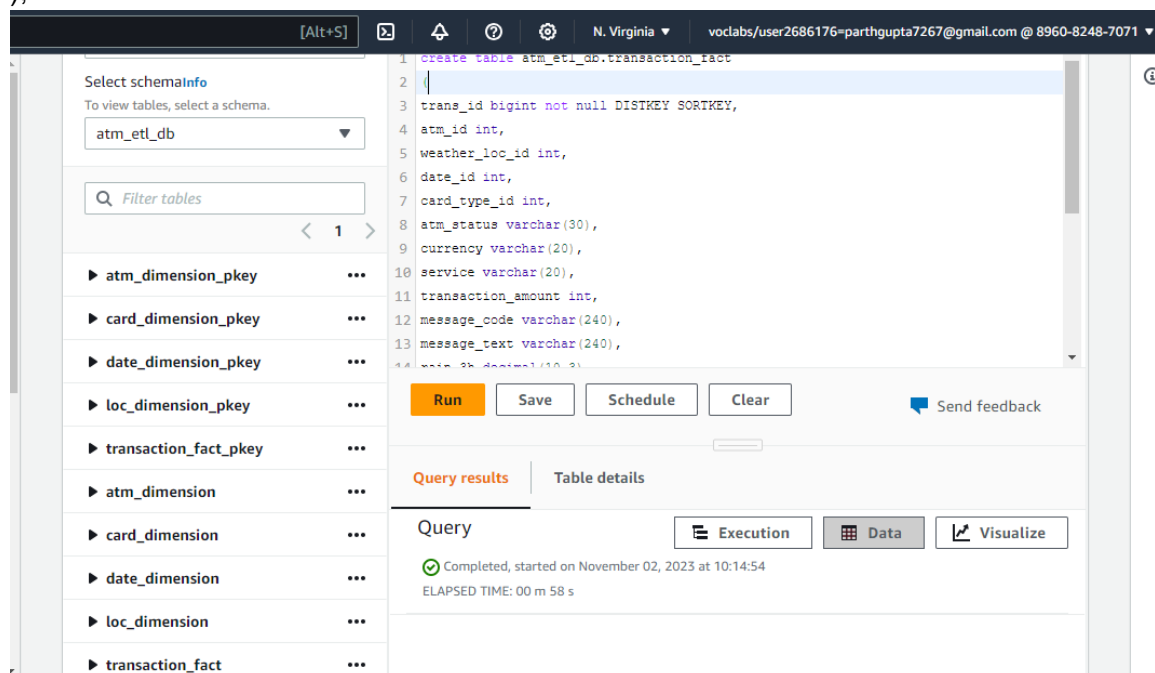
The screenshot shows a SQL query editor interface. On the left, there is a sidebar with a schema browser. The 'dev' schema is selected, and the 'atm_etl_db' schema is chosen. Below this, a list of tables is displayed, including 'atm_dimension_pkey', 'card_dimension_pkey', 'date_dimension_pkey', 'loc_dimension_pkey', 'atm_dimension', 'card_dimension', 'date_dimension', and 'loc_dimension'. The main area shows a SQL query being executed:

```
1 create table atm_etl_db.card_dimension
2 (
3 card_type_id int not null DISTKEY SORTKEY,
4 card_type varchar(40),
5 PRIMARY KEY(card_type_id)
6 );
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is active, showing a message: 'Query Completed, started on November 02, 2023 at 10:04:28' and 'ELAPSED TIME: 00 m 06 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

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)
);
```



The screenshot shows a SQL query editor interface. On the left, there is a sidebar with a search bar and a list of tables under the 'atm_etl_db' schema. The tables listed are: atm_dimension_pkey, card_dimension_pkey, date_dimension_pkey, loc_dimension_pkey, transaction_fact_pkey, atm_dimension, card_dimension, date_dimension, loc_dimension, and transaction_fact. The main area displays the SQL query for creating the 'transaction_fact' table, which matches the code provided in the previous block. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, there is a section for 'Query results' and 'Table details', with tabs for 'Query', 'Execution', 'Data', and 'Visualize'. The 'Query' tab is currently selected, showing a status message: 'Completed, started on November 02, 2023 at 10:14:54' and 'ELAPSED TIME: 00 m 58 s'.

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:

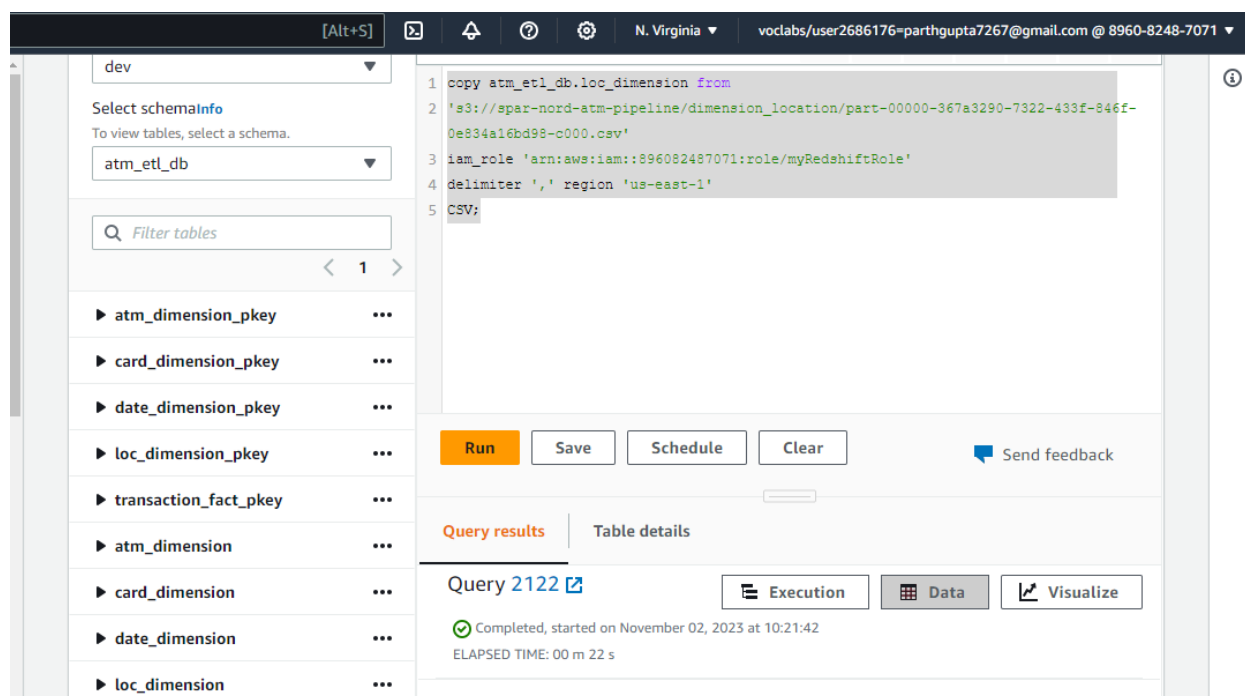
copy atm_etl_db.loc_dimension from

'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;



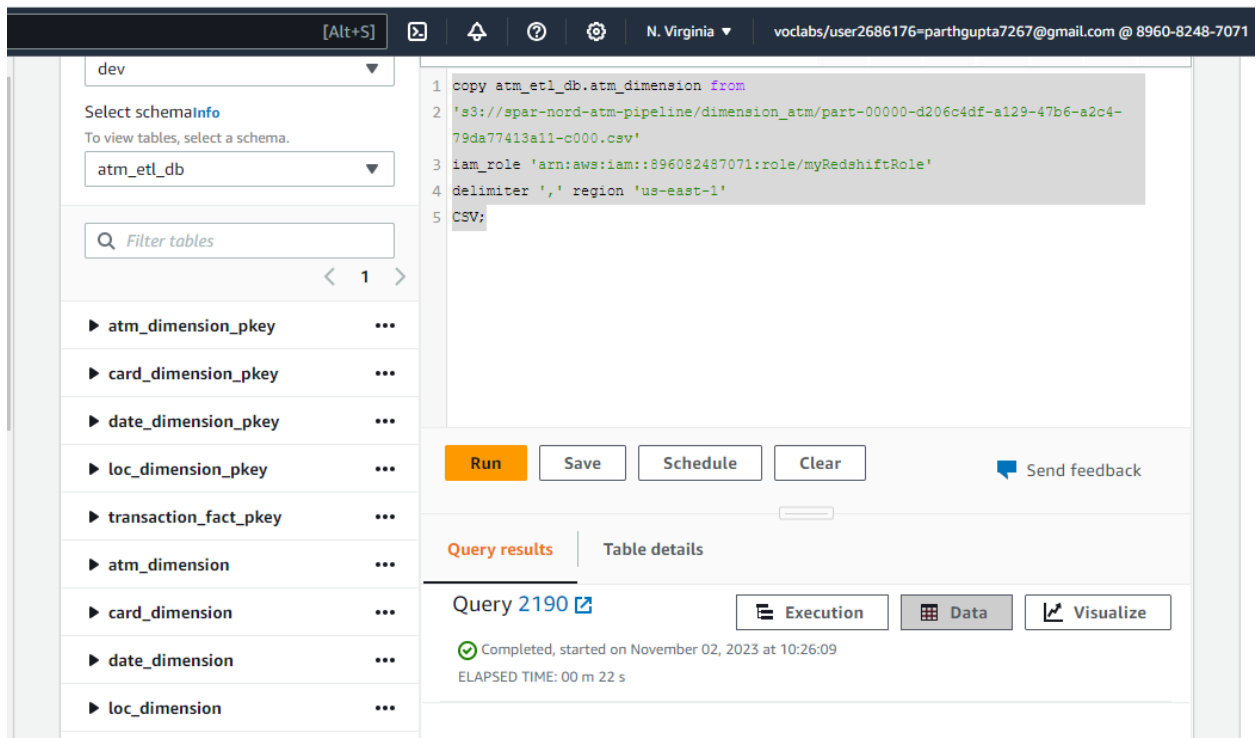
The screenshot shows the AWS Redshift console interface. On the left, the 'dev' schema is selected, and a list of tables is displayed, including 'atm_dimension_pkey', 'card_dimension_pkey', 'date_dimension_pkey', 'loc_dimension_pkey', 'transaction_fact_pkey', 'atm_dimension', 'card_dimension', 'date_dimension', and 'loc_dimension'. The main area shows a SQL query being executed:

```
1 copy atm_etl_db.loc_dimension from
2 's3://spar-nord-atm-pipeline/dimension_location/part-00000-367a3290-7322-433f-846f-
3 0e834a16bd98-c000.csv'
4 iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
5 delimiter ',' region 'us-east-1'
6 CSV;
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of the buttons is a 'Send feedback' link. Below the buttons, the 'Query results' tab is selected, showing the query ID '2122'. The status indicates the query is 'Completed, started on November 02, 2023 at 10:21:42' with an 'ELAPSED TIME: 00 m 22 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

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;
```



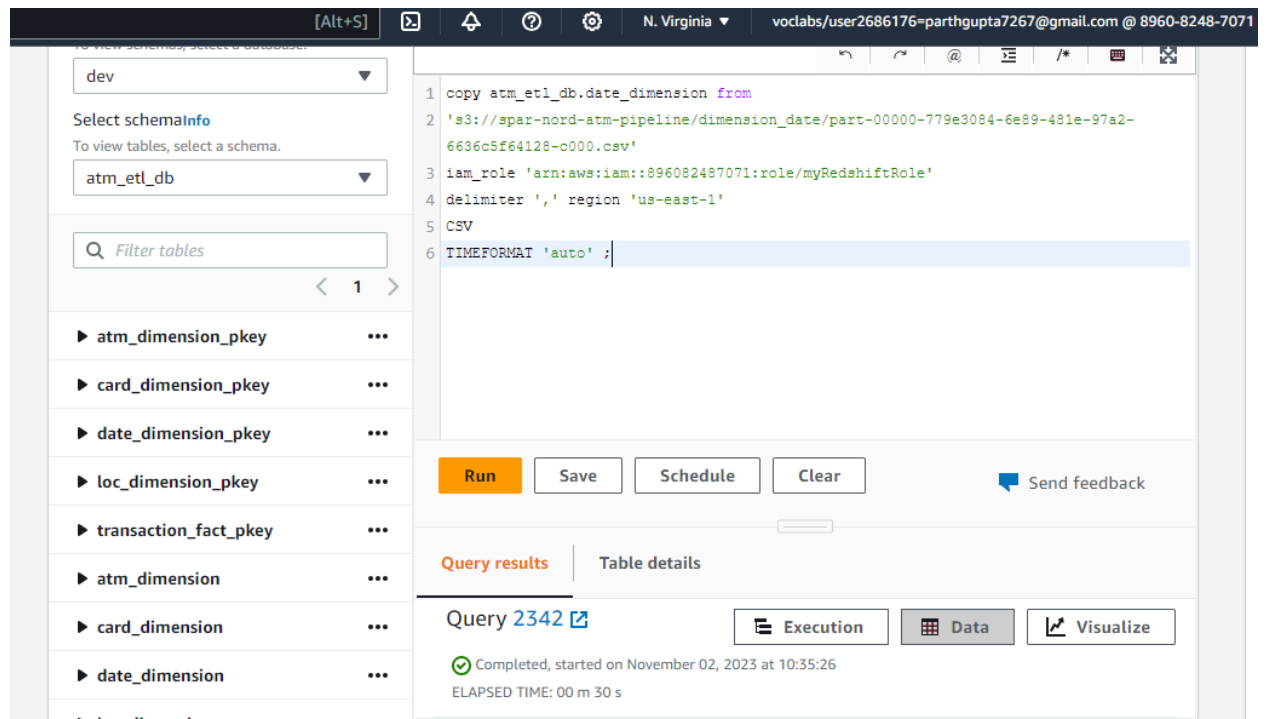
The screenshot displays the AWS Redshift console interface. On the left, a sidebar shows the database schema 'dev' with a list of tables including 'atm_dimension_pkey', 'card_dimension_pkey', 'date_dimension_pkey', 'loc_dimension_pkey', 'transaction_fact_pkey', 'atm_dimension', 'card_dimension', 'date_dimension', and 'loc_dimension'. The main area contains a SQL query editor with the following text:

```
1 copy atm_etl_db.atm_dimension from
2 's3://spar-nord-atm-pipeline/dimension_atm/part-00000-d206c4df-a129-47b6-a2c4-
3 79da77413a11-c000.csv'
4 iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
5 delimiter ',' region 'us-east-1'
6 CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is active, showing the query ID 'Query 2190' and its execution status: 'Completed, started on November 02, 2023 at 10:26:09' with an 'ELAPSED TIME: 00 m 22 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

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' ;
```



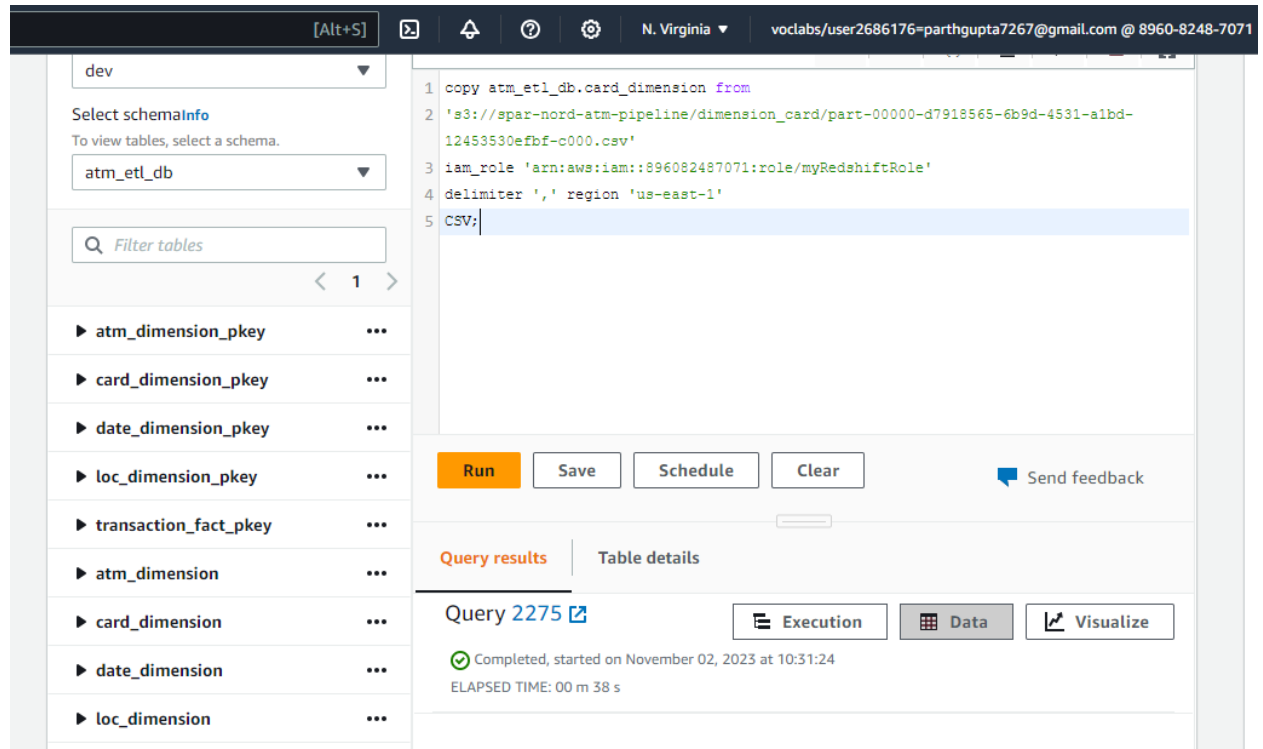
The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with a dropdown menu set to 'dev' and a list of tables including 'atm_dimension_pkey', 'card_dimension_pkey', 'date_dimension_pkey', 'loc_dimension_pkey', 'transaction_fact_pkey', 'atm_dimension', 'card_dimension', 'date_dimension', and 'loc_dimension'. The main area displays a SQL query being executed:

```
1 copy atm_etl_db.date_dimension from
2 's3://spar-nord-atm-pipeline/dimension_date/part-00000-779e3084-6e89-481e-97a2-
3 6636c5f64128-c000.csv'
4 iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
5 delimiter ',' region 'us-east-1'
6 CSV
7 TIMEFORMAT 'auto' ;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is active, showing 'Query 2342' with a status of 'Completed, started on November 02, 2023 at 10:35:26' and an 'ELAPSED TIME: 00 m 30 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

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;
```



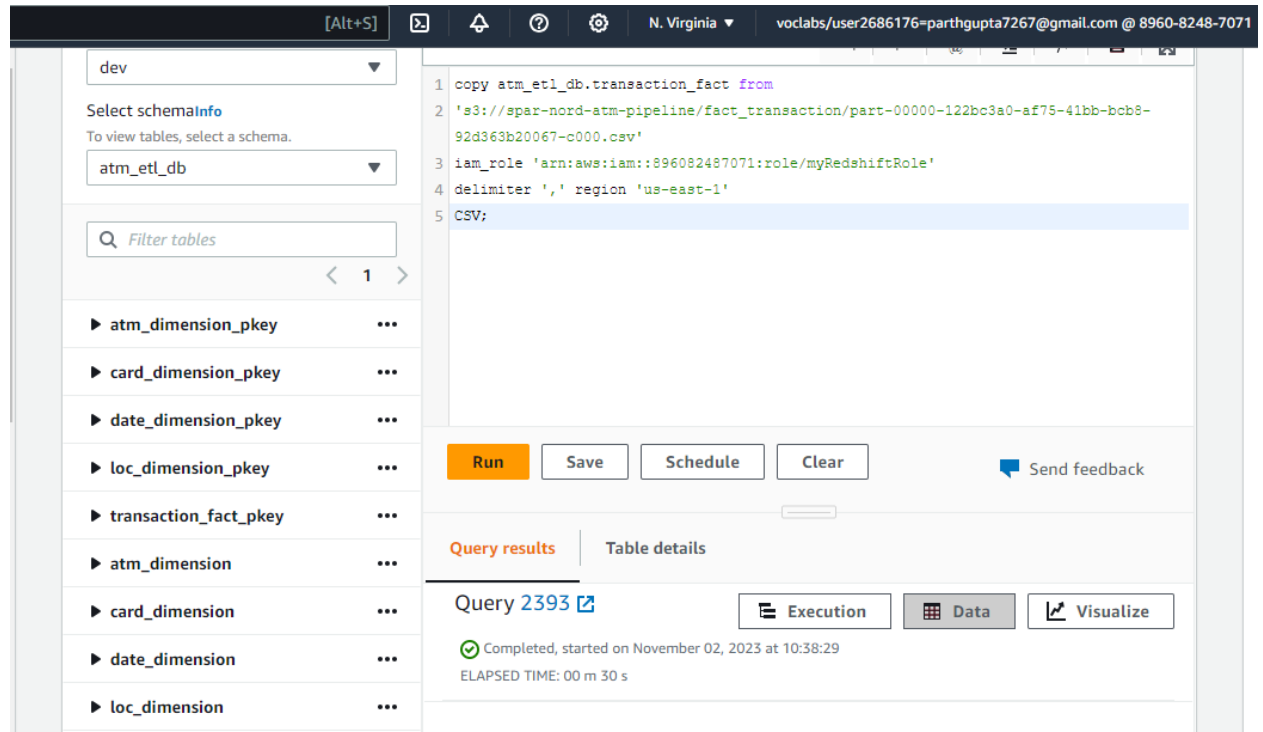
The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with a schema dropdown set to 'dev' and a list of tables including 'atm_dimension_pkey', 'card_dimension_pkey', 'date_dimension_pkey', 'loc_dimension_pkey', 'transaction_fact_pkey', 'atm_dimension', 'card_dimension', 'date_dimension', and 'loc_dimension'. The main area contains a SQL query editor with the following text:

```
1 copy atm_etl_db.card_dimension from
2 's3://spar-nord-atm-pipeline/dimension_card/part-00000-d7918565-6b9d-4531-a1bd-
3 12453530efbf-c000.csv'
4 iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
5 delimiter ',' region 'us-east-1'
6 CSV;
```

Below the query editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is active, showing 'Query 2275' with a status of 'Completed, started on November 02, 2023 at 10:31:24' and an 'ELAPSED TIME: 00 m 38 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

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;
```



The screenshot shows the AWS Redshift console interface. On the left, the 'dev' database is selected, and the 'atm_etl_db' schema is chosen. A list of tables is displayed, including 'atm_dimension_pkey', 'card_dimension_pkey', 'date_dimension_pkey', 'loc_dimension_pkey', 'transaction_fact_pkey', 'atm_dimension', 'card_dimension', 'date_dimension', and 'loc_dimension'. The main area shows a SQL query being executed:

```
1 copy atm_etl_db.transaction_fact from
2 's3://spar-nord-atm-pipeline/fact_transaction/part-00000-122bc3a0-af75-41bb-bcb8-
3 92d363b20067-c000.csv'
4 iam_role 'arn:aws:iam::896082487071:role/myRedshiftRole'
5 delimiter ',' region 'us-east-1'
6 CSV;
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The 'Query results' tab is active, showing the query ID 'Query 2393' and its execution status: 'Completed, started on November 02, 2023 at 10:38:29' with an 'ELAPSED TIME: 00 m 30 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.