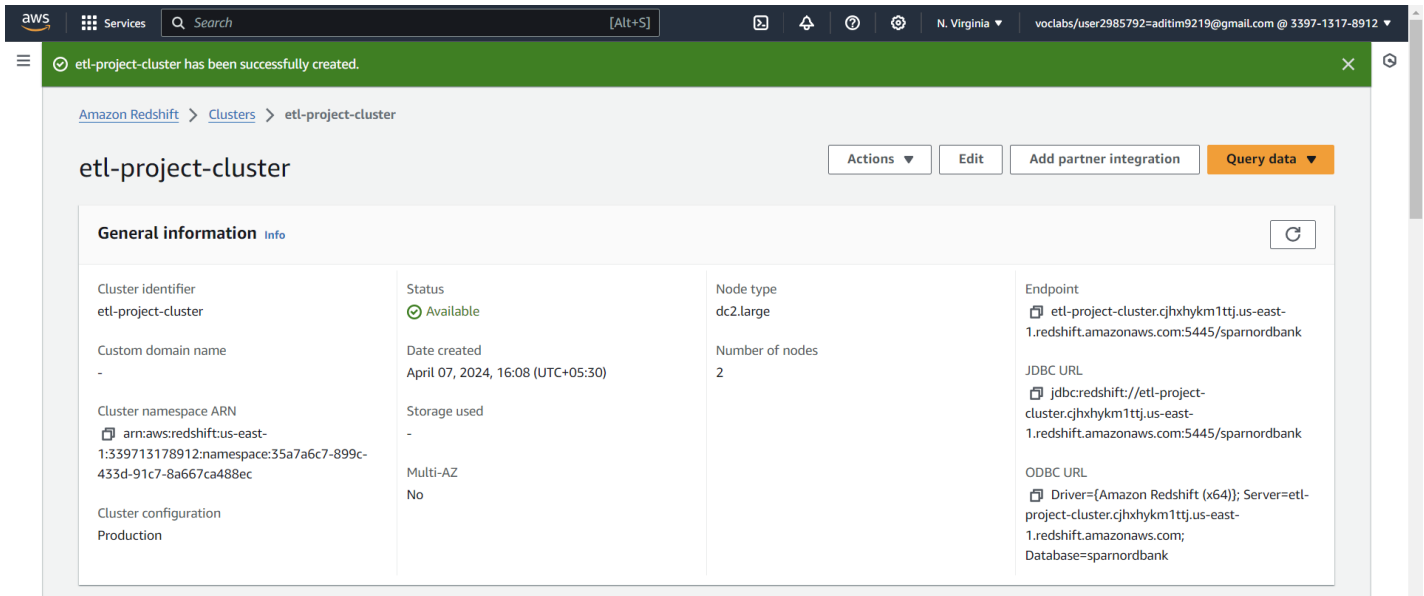


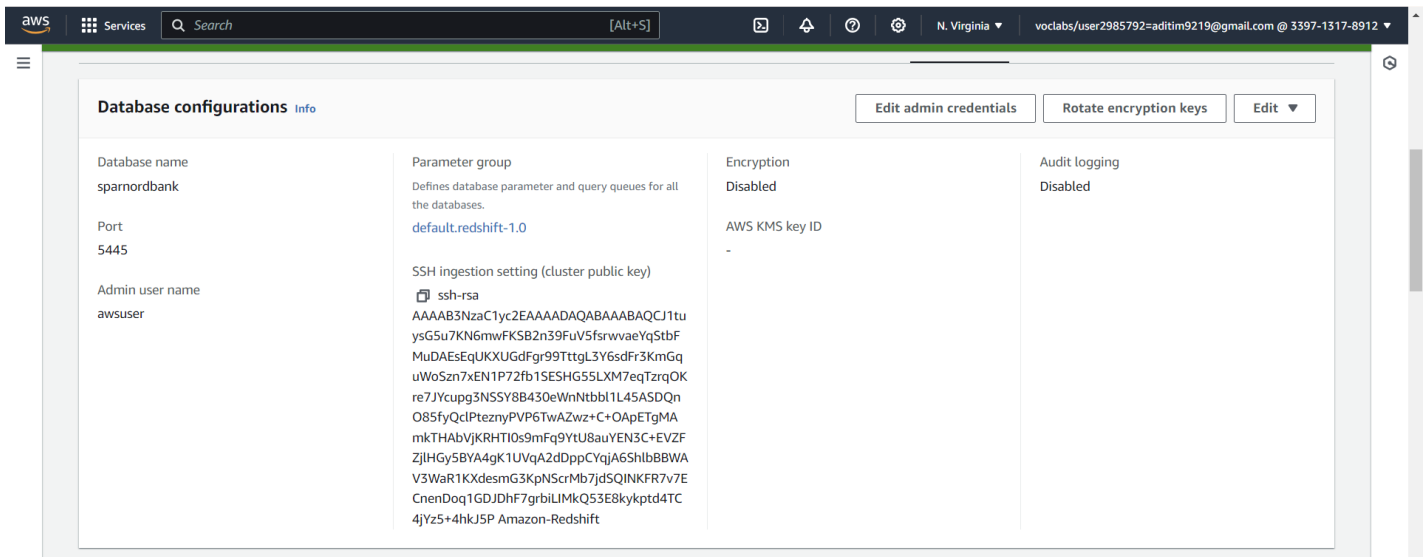
Creation of a Redshift Cluster

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



The screenshot shows the AWS Redshift console for the cluster 'etl-project-cluster'. The cluster is in an 'Available' status. The configuration details are as follows:

General information			
Cluster identifier	etl-project-cluster	Status	Available
Custom domain name	-	Date created	April 07, 2024, 16:08 (UTC+05:30)
Cluster namespace ARN	arn:aws:redshift:us-east-1:339713178912:namespace:35a7a6c7-899c-433d-91c7-8a667ca488ec	Storage used	-
Cluster configuration	Production	Multi-AZ	No
Node type	dc2.large	Number of nodes	2
Endpoint	etl-project-cluster.cjhxhykm1ttj.us-east-1.redshift.amazonaws.com:5445/sparnordbank		
JDBC URL	jdbc:redshift://etl-project-cluster.cjhxhykm1ttj.us-east-1.redshift.amazonaws.com:5445/sparnordbank		
ODBC URL	Driver={Amazon Redshift (x64)}; Server=etl-project-cluster.cjhxhykm1ttj.us-east-1.redshift.amazonaws.com; Database=sparnordbank		



The screenshot shows the 'Database configurations' section for the 'etl-project-cluster'. The configurations are as follows:

Database configurations			
Database name	sparnordbank	Parameter group	default.redshift-1.0
Port	5445	Encryption	Disabled
Admin user name	awsuser	AWS KMS key ID	-
SSH ingestion setting (cluster public key)	ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQCAJ1tuysG5u7KN6mwFKSB2n39FuV5fsrwvaeYqStbFMuDAEsEqUKXUGdFgr99TtL3Y6sdFr3KmGquWoSzn7xEN1P72fb1SESHG55LXM7eqTzrqOKre7JYcupg3NSSY8B430eWnNtbb1L45ASDQnO85fyQcIPtezyPVP6TwAZwz+C+OApETgMAmkTHAbVjKRHTI0s9mFq9YtU8auYEN3C+EVZFZjlHGy5BYA4gK1UVqA2dDppCYqjA6ShibBBWA V3WaR1KXdesmG3KpNScrMb7jdSQINKFR7v7ECnenDoq1GDJDhF7grbiLIMkQ53E8kykptd4TC4jYz5+4hkJP Amazon-Redshift		
Audit logging	Disabled		

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user2985792=aditim9219@gmail.com @ 3397-1317-8912

Network and security settings [Info](#)

Edit

Virtual private cloud (VPC)
vpc-0ef1492a7390336a9 [↗](#)

Subnet group
cluster-subnet-group-redshift

Endpoint URL
 -

Availability Zone
us-east-1a

Enhanced VPC routing
Disabled

VPC security group
Specify which instances and devices can connect to the cluster.
sg-070ab9e2e95f67b27 [↗](#)

IP address type
-

Publicly accessible
Allow connections from outside the VPC.
Disabled

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user2985792=aditim9219@gmail.com @ 3397-1317-8912

Associated IAM roles (1) [Info](#)

Set default [▼](#) Manage IAM roles [▼](#)

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

Q Search for associated IAM role by name, status, or role type

< 1 >

<input type="checkbox"/>	IAM roles ↗	▼	Status ▼	Role type ▼
<input type="checkbox"/>	myRedshiftRole		in-sync	--

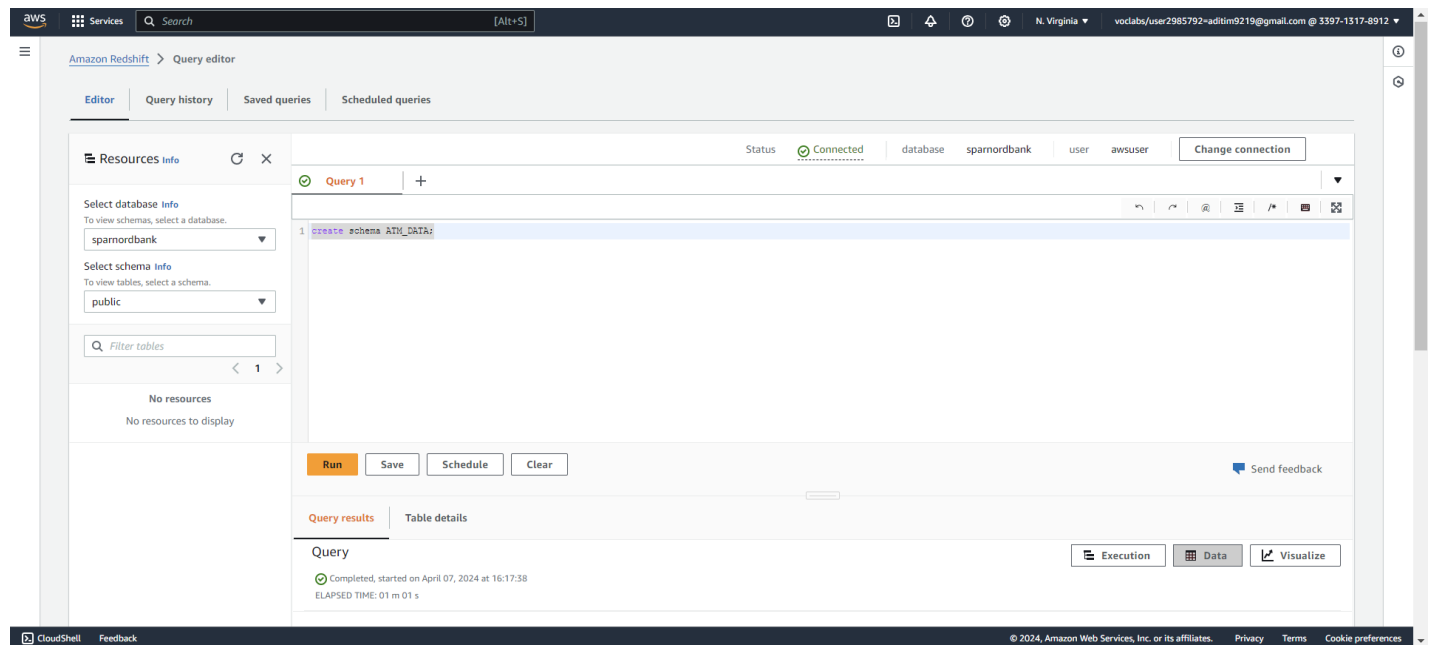
Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

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

Creating Schema:

Query:

create schema ATM_DATA;



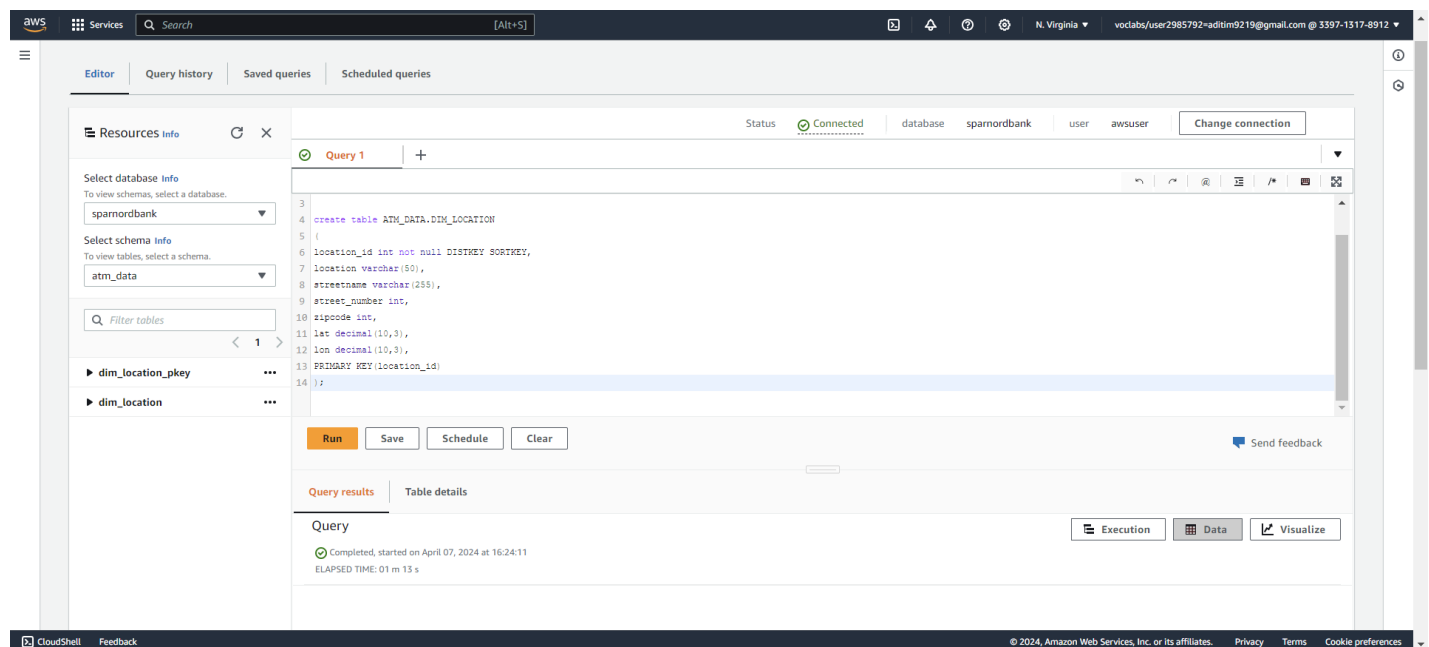
The screenshot displays the Amazon Redshift Query Editor interface. On the left, the 'Resources' panel shows the selected database as 'sparmordbank' and the schema as 'public'. The main editor area contains a single query: `1 create schema ATM_DATA;`. Below the query, the 'Run' button is highlighted in orange. The 'Query results' tab at the bottom shows a status of 'Completed, started on April 07, 2024 at 16:17:38' with an 'ELAPSED TIME' of '01 m 01 s'. The interface also includes a top navigation bar with 'AWS Services' and a search bar, and a bottom status bar with 'CloudShell' and 'Feedback' links.

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)
);
```



The screenshot shows the AWS CloudShell interface. On the left, the 'Resources Info' panel displays the selected database 'sparnordbank' and schema 'atm_data'. The main editor area shows a SQL query to create a table named 'ATM_DATA.DIM_LOCATION'. The query is as follows:

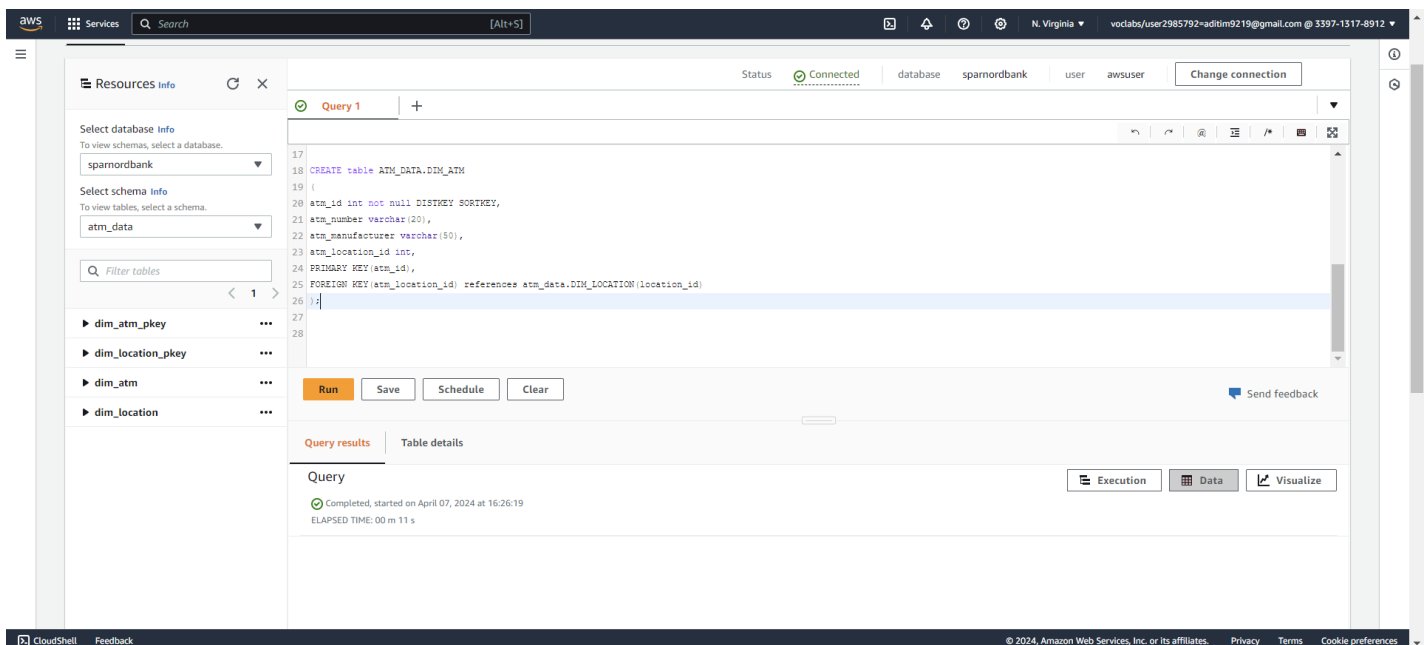
```
3
4 create table ATM_DATA.DIM_LOCATION
5 (
6 location_id int not null DISTKEY SORTKEY,
7 location varchar(50),
8 streetname varchar(255),
9 street_number int,
10 zipcode int,
11 lat decimal(10,3),
12 lon decimal(10,3),
13 PRIMARY KEY(location_id)
14 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button has been clicked, and the 'Query results' tab is active. The results show that the query was completed successfully on April 07, 2024, at 16:24:11, with an elapsed time of 01 m 13 s.

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)
);
```



The screenshot displays the AWS CloudShell interface. On the left, the 'Resources' panel shows the selected database 'sparnordbank' and schema 'atm_data'. The main editor area contains a SQL query to create the 'ATM_DATA.DIM_ATM' table. The query is as follows:

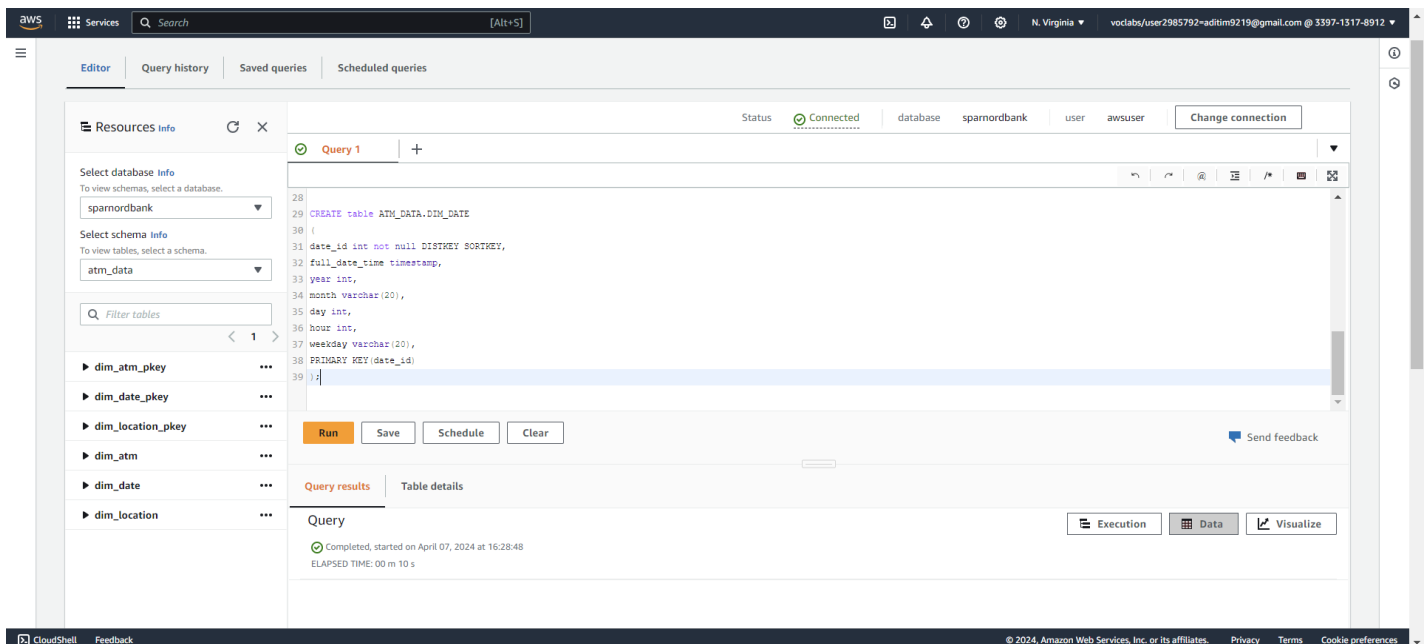
```
17 CREATE table ATM_DATA.DIM_ATM
18 (
19   atm_id int not null DISTKEY SORTKEY,
20   atm_number varchar(20),
21   atm_manufacturer varchar(50),
22   atm_location_id int,
23   PRIMARY KEY(atm_id),
24   FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
25 );
26
27
28
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button has been clicked, and the 'Query results' tab is active. It shows the query execution status: 'Completed, started on April 07, 2024 at 16:26:19' with an 'ELAPSED TIME: 00 m 11 s'. The 'Table details' tab is also visible.

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)
);
```



The screenshot displays the AWS CloudShell interface. On the left, the 'Resources' pane shows the selected database 'sparnordbank' and schema 'atm_data'. The main editor area contains a SQL query to create a table named 'ATM_DATA.DIM_DATE'. The query is as follows:

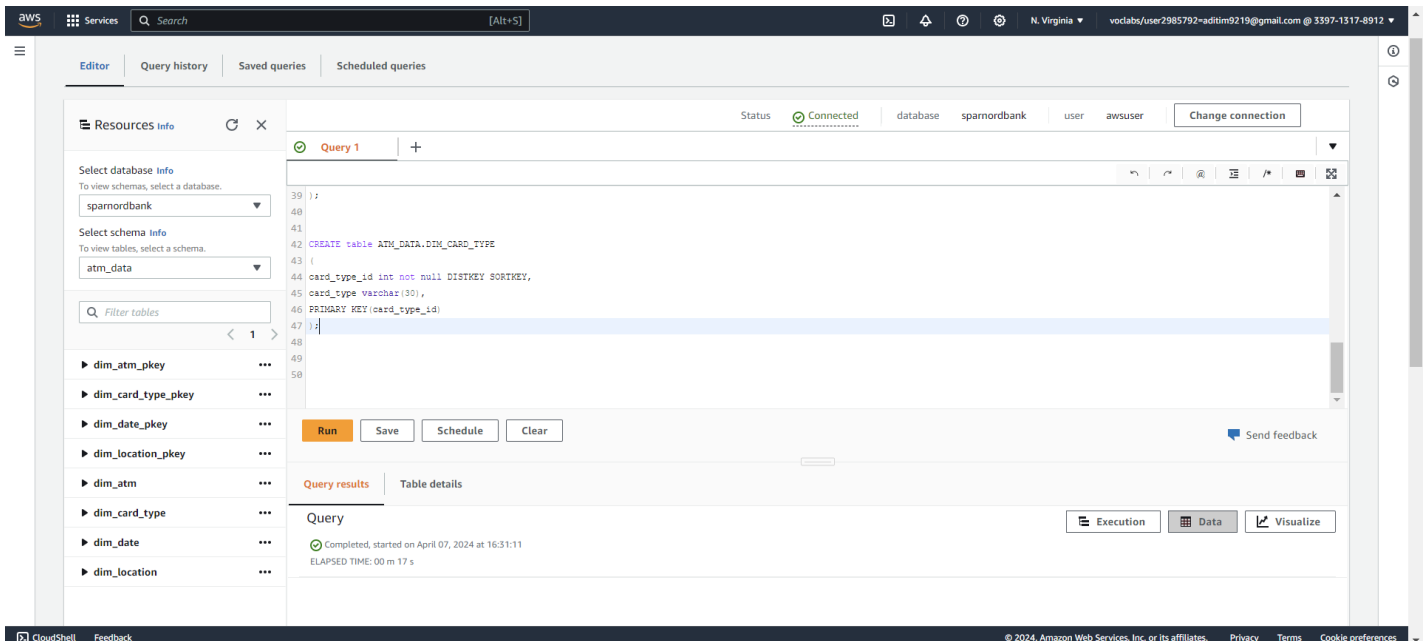
```
28
29 CREATE table ATM_DATA.DIM_DATE
30 (
31 date_id int not null DISTKEY SORTKEY,
32 full_date_time timestamp,
33 year int,
34 month varchar(20),
35 day int,
36 hour int,
37 weekday varchar(20),
38 PRIMARY KEY(date_id)
39 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button has been clicked, and the 'Query results' pane at the bottom shows the status: 'Completed, started on April 07, 2024 at 16:28:48' and 'ELAPSED TIME: 00 m 10 s'.

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)
);
```



The screenshot displays the AWS CloudShell interface. On the left, the 'Resources' pane shows the selected database 'sparnordbank' and schema 'atm_data'. The main editor area contains a SQL query to create a table named 'DIM_CARD_TYPE' in the 'ATM_DATA' schema. The query is as follows:

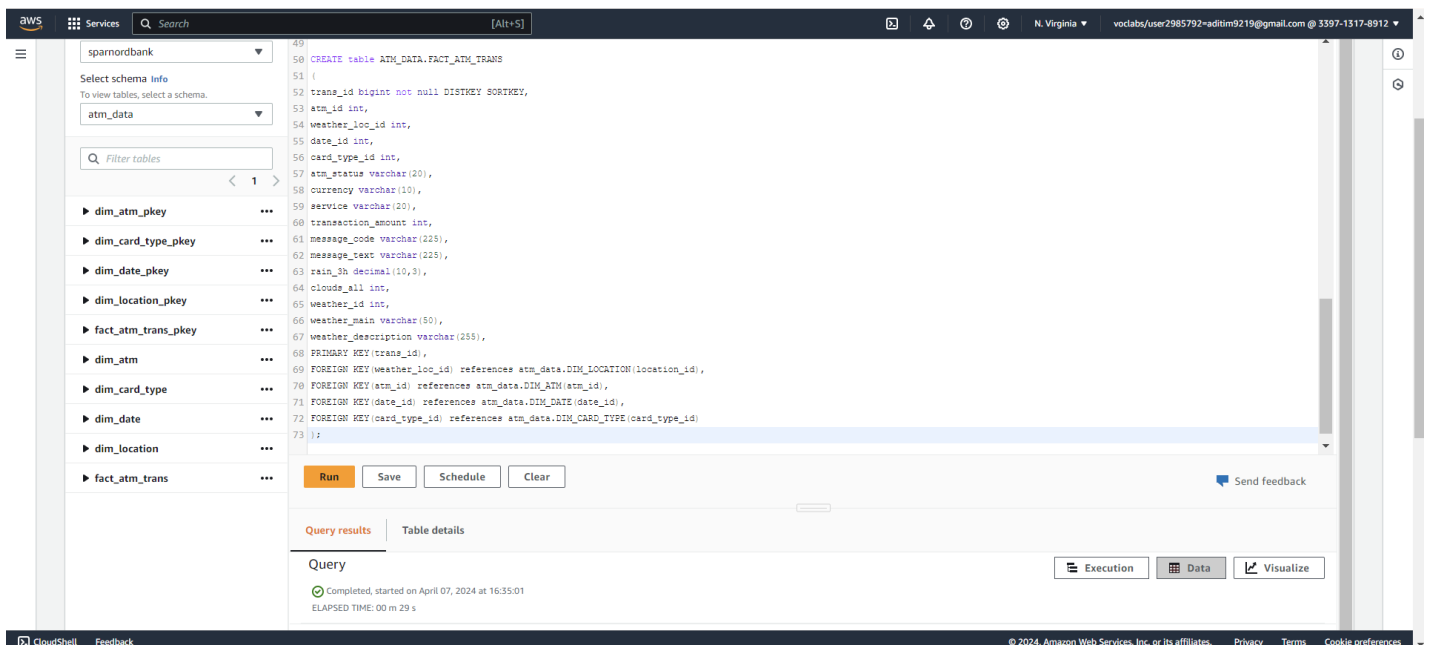
```
CREATE table ATM_DATA.DIM_CARD_TYPE
(
  card_type_id int not null DISTKEY SORTKEY,
  card_type varchar(30),
  PRIMARY KEY(card_type_id)
);
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' pane at the bottom shows the execution status: 'Completed, started on April 07, 2024 at 16:31:11' and 'ELAPSED TIME: 00 m 17 s'.

5. Creating ATM TRANSACTIONS FACT table:

Query:

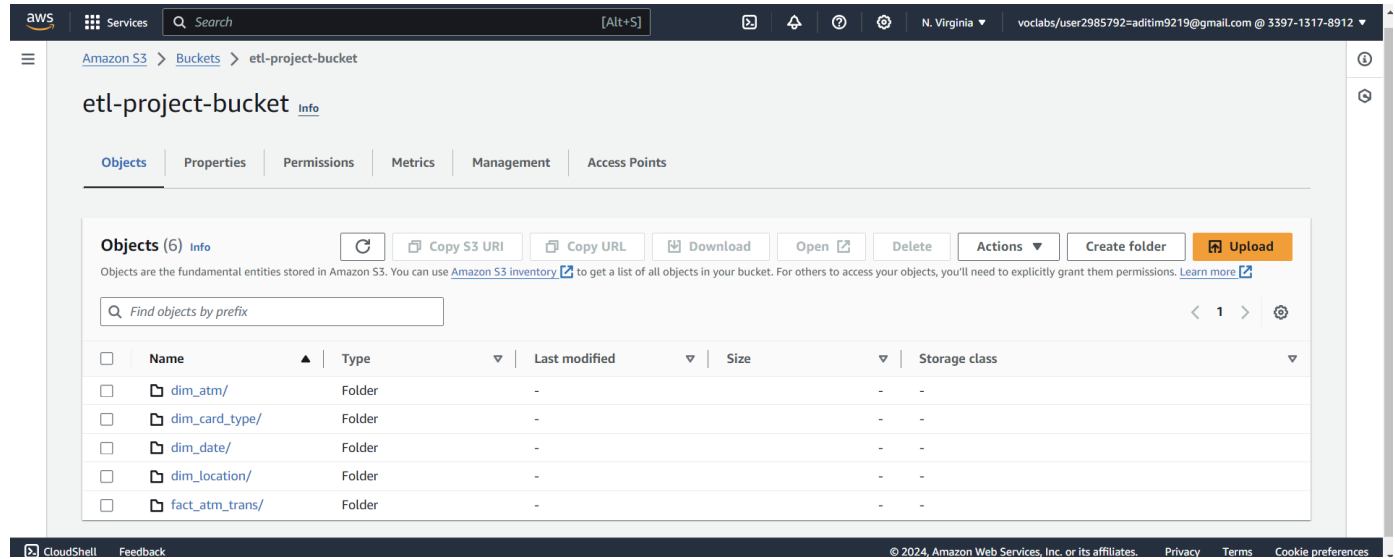
```
CREATE table ATM_DATA.FACT_ATM_TRANS
(
trans_id bigint not null DISTKEY SORTKEY,
atm_id int,
weather_loc_id int,
date_id int,
card_type_id int,
atm_status varchar(20),
currency varchar(10),
service varchar(20),
transaction_amount int,
message_code varchar(225),
message_text varchar(225),
rain_3h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



The screenshot shows the AWS Glue console interface. On the left, the 'Services' menu is open, and the 'atm_data' schema is selected. The 'Filter tables' section shows a list of tables including 'dim_atm_pkey', 'dim_card_type_pkey', 'dim_date_pkey', 'dim_location_pkey', 'fact_atm_trans_pkey', 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The main panel displays the SQL query for creating the 'FACT_ATM_TRANS' table, which is the same query provided in the text above. The query is highlighted in blue. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. At the bottom, the 'Query results' tab is active, showing a message: 'Completed, started on April 07, 2024 at 16:35:01' and 'ELAPSED TIME: 00 m 29 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

Loading data into a Redshift cluster from Amazon S3 bucket

Screenshot of data in S3 bucket:



The screenshot shows the AWS S3 console interface for the bucket 'etl-project-bucket'. The 'Objects' tab is selected, displaying a list of six folders. The table below summarizes the objects shown in the screenshot.

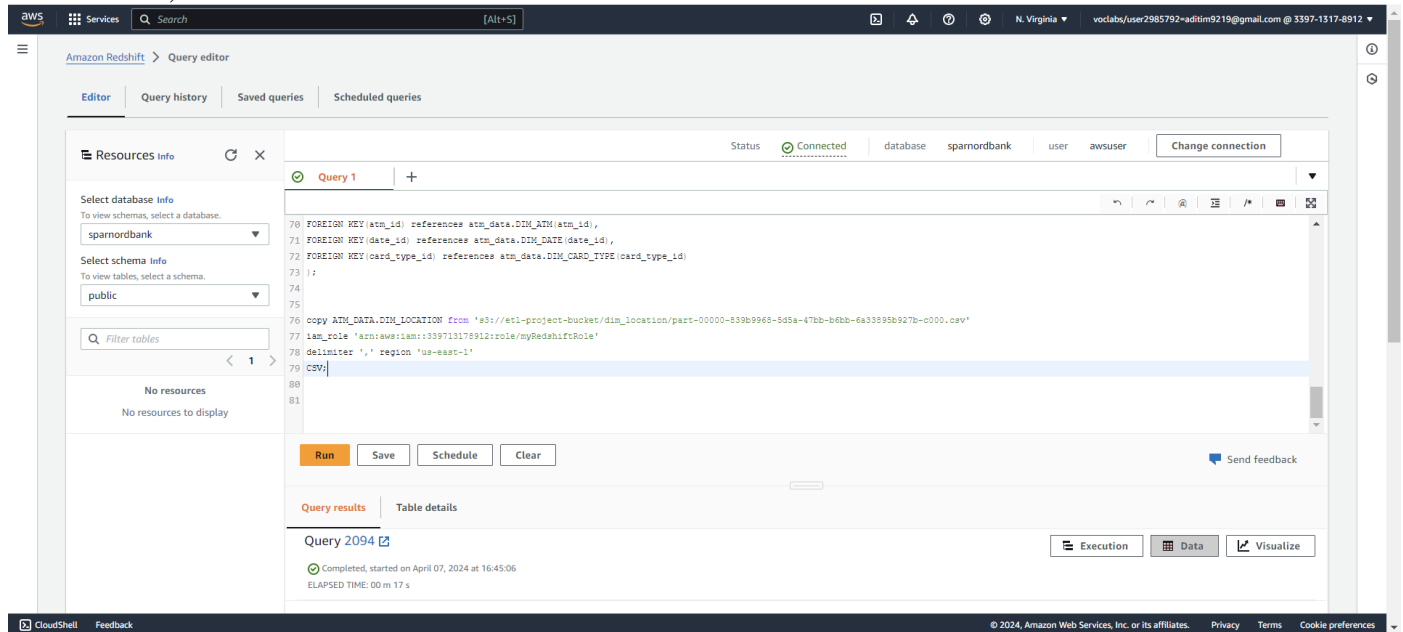
<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	dim_atm/	Folder	-	-	-
<input type="checkbox"/>	dim_card_type/	Folder	-	-	-
<input type="checkbox"/>	dim_date/	Folder	-	-	-
<input type="checkbox"/>	dim_location/	Folder	-	-	-
<input type="checkbox"/>	fact_atm_trans/	Folder	-	-	-

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;



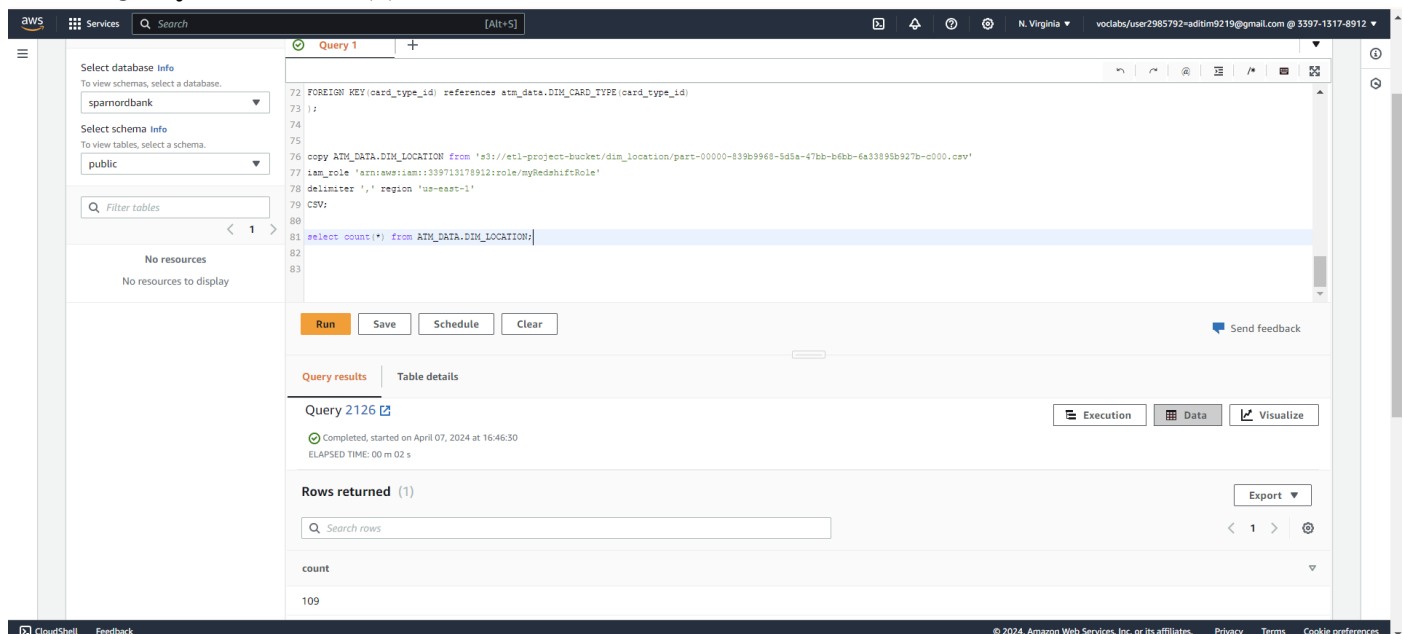
The screenshot shows the Amazon Redshift Query Editor interface. On the left, the 'Resources' panel shows the selected database 'sparnordbank' and schema 'public'. The main editor area displays the following SQL query:

```
70 FOREIGN KEY(atm_id) references atm_data.DIM_ATM(atm_id),
71 FOREIGN KEY(date_id) references atm_data.DIM_DATE(date_id),
72 FOREIGN KEY(card_type_id) references atm_data.DIM_CARD_TYPE(card_type_id)
73 ;;
74
75
76 copy ATM_DATA.DIM_LOCATION from 's3://etl-project-bucket/dim_location/part-00000-839b9968-5d5a-47bb-b6bb-6a33895b927b-c000.csv'
77 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
78 delimiter ',' region 'us-east-1'
79 CSV;
80
81
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab shows the execution status: 'Completed, started on April 07, 2024 at 16:45:06' and 'ELAPSED TIME: 00 m 17 s'.

Count of rows in DIM_LOCATION table:

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



The screenshot shows the Amazon Redshift Query Editor interface. The main editor area displays the following SQL query:

```
72 FOREIGN KEY(card_type_id) references atm_data.DIM_CARD_TYPE(card_type_id)
73 ;;
74
75
76 copy ATM_DATA.DIM_LOCATION from 's3://etl-project-bucket/dim_location/part-00000-839b9968-5d5a-47bb-b6bb-6a33895b927b-c000.csv'
77 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
78 delimiter ',' region 'us-east-1'
79 CSV;
80
81 select count(*) from ATM_DATA.DIM_LOCATION;
82
83
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab shows the execution status: 'Completed, started on April 07, 2024 at 16:46:30' and 'ELAPSED TIME: 00 m 02 s'.

The 'Rows returned (1)' section shows the following result:

count
109

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

2. Copying the data to DIM_ATM table:

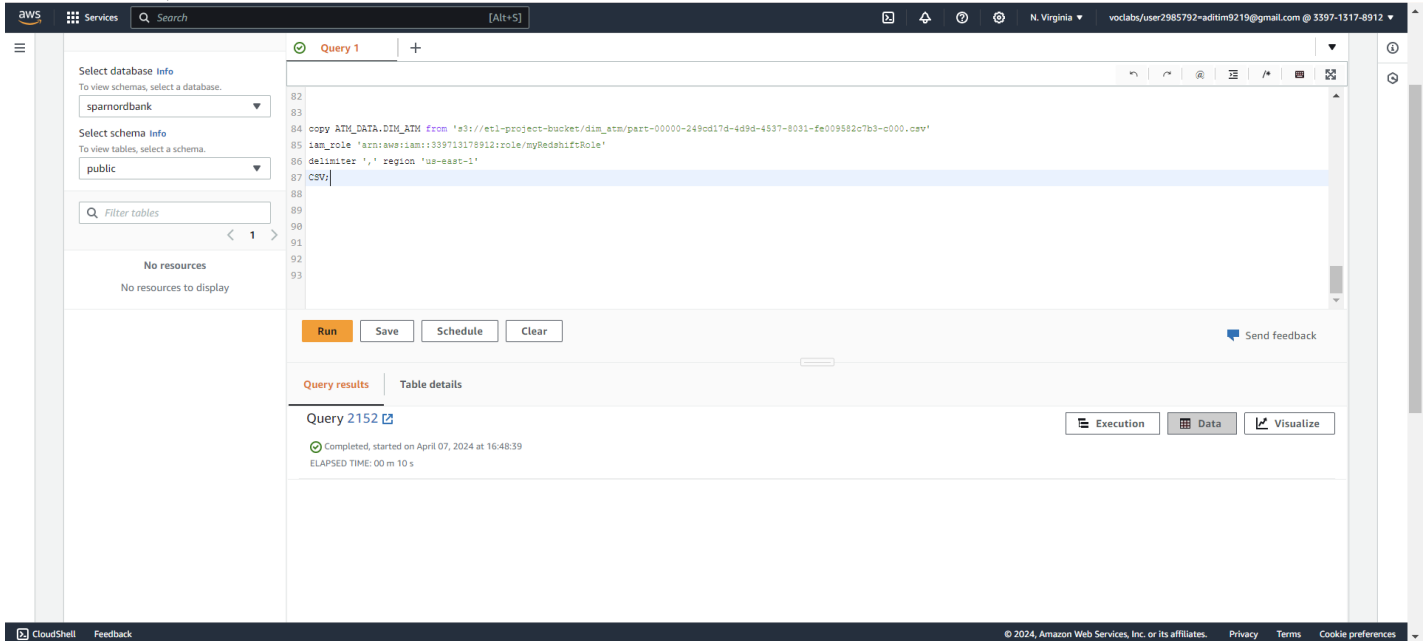
Query:

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;



The screenshot shows the AWS Redshift Query Editor interface. On the left, the 'Select database' dropdown is set to 'sparnordbank' and the 'Select schema' dropdown is set to 'public'. The main query editor displays the following SQL query:

```

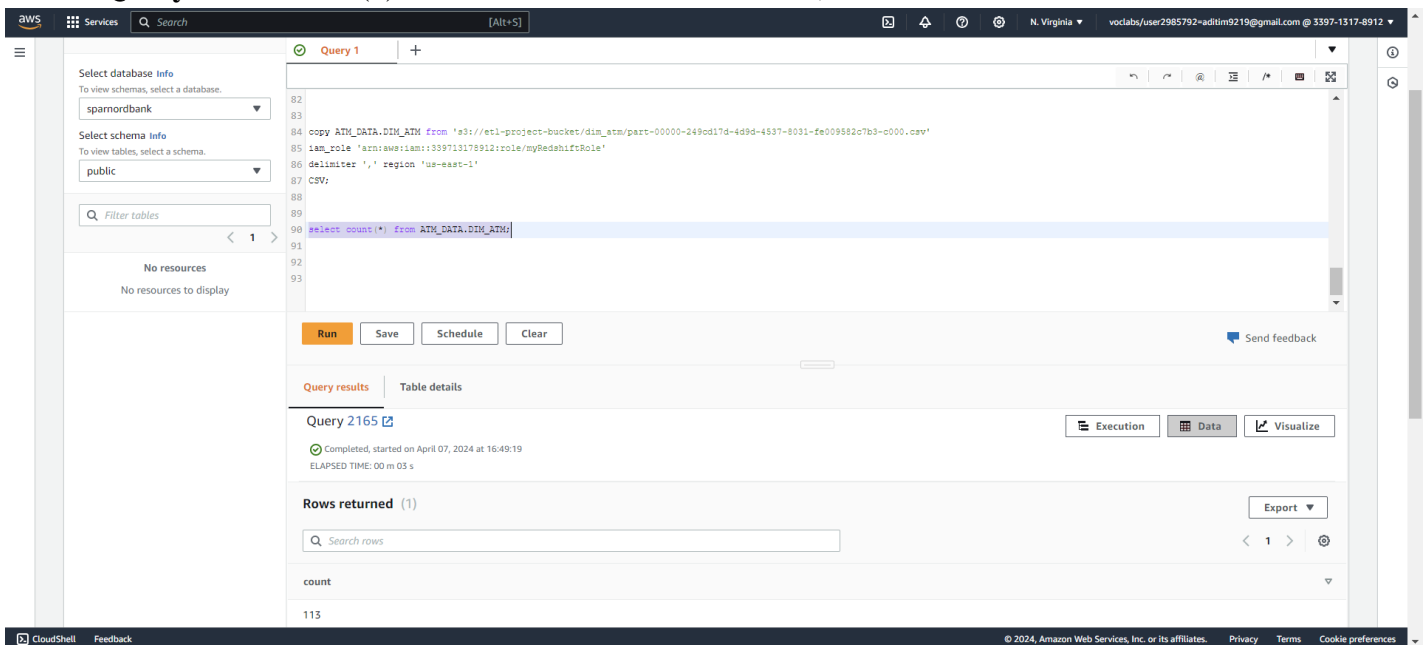
82
83
84 copy ATM_DATA.DIM_ATM from 's3://etl-project-bucket/dim_atm/part-00000-249cd17d-4d9d-4537-8031-fe009582c7b3-c000.csv'
85 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
86 delimiter ',' region 'us-east-1'
87 CSV;
88
89
90
91
92
93

```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab is selected, showing 'Query 2152' with a status of 'Completed, started on April 07, 2024 at 16:48:39' and an 'ELAPSED TIME: 00 m 10 s'.

Count of rows in DIM_ATM table:

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



The screenshot shows the AWS Redshift Query Editor interface. The main query editor displays the following SQL query:

```

82
83
84 copy ATM_DATA.DIM_ATM from 's3://etl-project-bucket/dim_atm/part-00000-249cd17d-4d9d-4537-8031-fe009582c7b3-c000.csv'
85 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
86 delimiter ',' region 'us-east-1'
87 CSV;
88
89
90 select count(*) from ATM_DATA.DIM_ATM;
91
92
93

```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab is selected, showing 'Query 2165' with a status of 'Completed, started on April 07, 2024 at 16:49:19' and an 'ELAPSED TIME: 00 m 03 s'.

Under 'Rows returned (1)', there is a search bar and a table with one row:

count
113

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

3. Copying the data to DIM_DATE table:

Query:

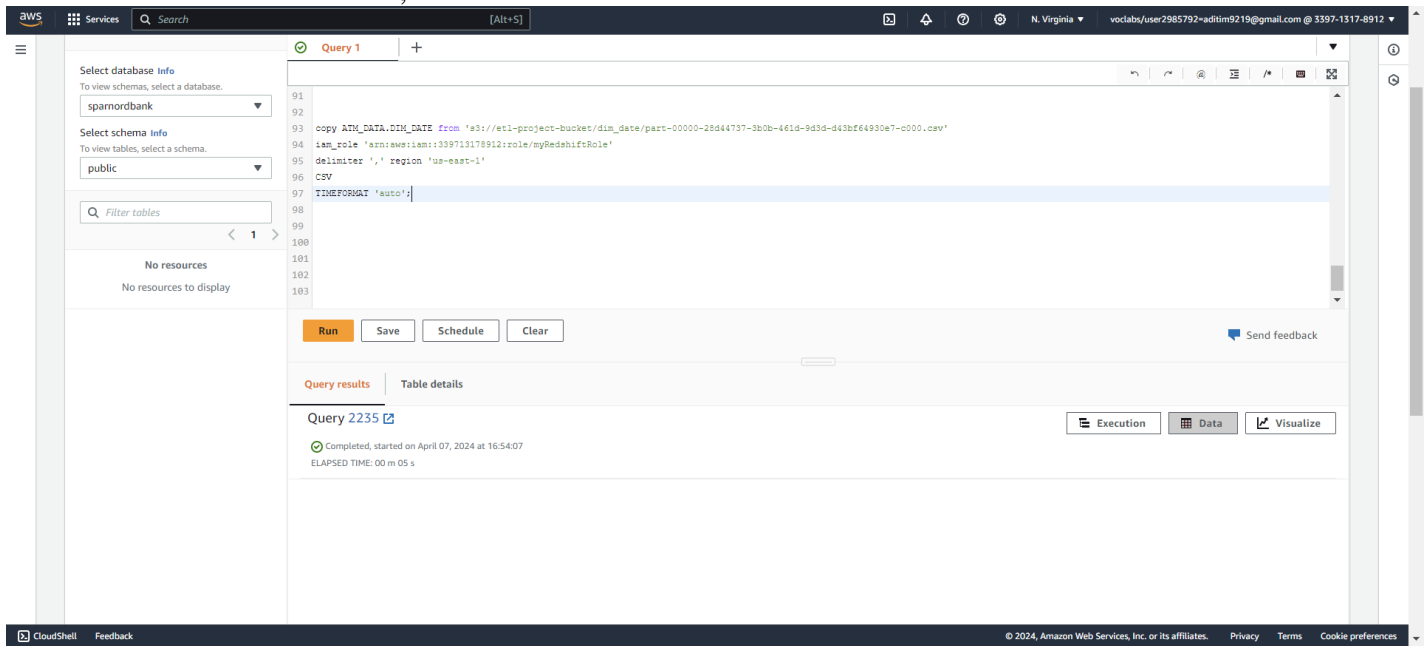
copy ATM_DATA.DIM_DATE from 's3://etl-project-bucket/dim_date/part-00000-28d44737-3b0b-461d-9d3d-d43bf64930e7-c000.csv'

iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'

delimiter ',' region 'us-east-1'

CSV

TIMEFORMAT 'auto';



The screenshot shows the AWS Redshift Query Editor interface. On the left, the 'Select database' dropdown is set to 'sparnordbank' and the 'Select schema' dropdown is set to 'public'. The main query editor displays the following SQL code:

```

91
92
93 copy ATM_DATA.DIM_DATE from 's3://etl-project-bucket/dim_date/part-00000-28d44737-3b0b-461d-9d3d-d43bf64930e7-c000.csv'
94 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
95 delimiter ',' region 'us-east-1'
96 CSV
97 TIMEFORMAT 'auto';
98
99
100
101
102
103

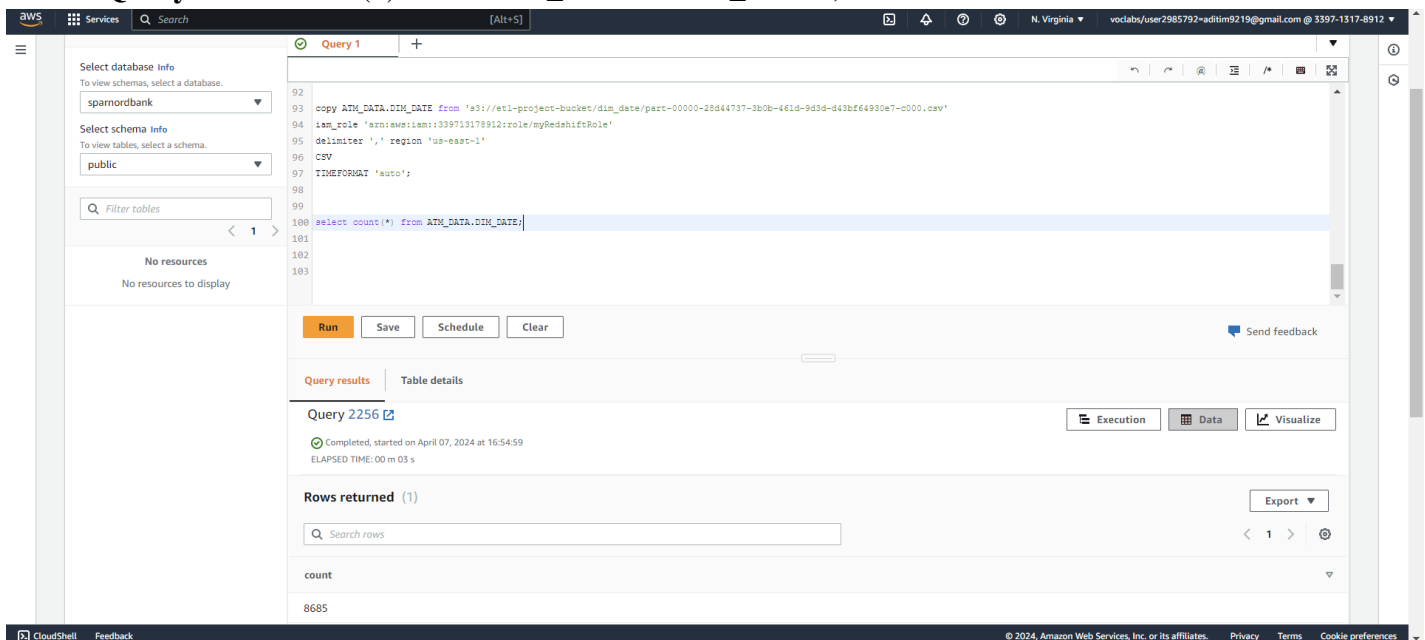
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab is selected, showing the status of the query:

Query 2235
 Completed, started on April 07, 2024 at 16:54:07
 ELAPSED TIME: 00 m 05 s

Count of rows in DIM_DATE table:

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



The screenshot shows the AWS Redshift Query Editor interface. The main query editor displays the following SQL code:

```

92
93 copy ATM_DATA.DIM_DATE from 's3://etl-project-bucket/dim_date/part-00000-28d44737-3b0b-461d-9d3d-d43bf64930e7-c000.csv'
94 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
95 delimiter ',' region 'us-east-1'
96 CSV
97 TIMEFORMAT 'auto';
98
99
100 #select count(*) from ATM_DATA.DIM_DATE;
101
102
103

```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab is selected, showing the status of the query:

Query 2256
 Completed, started on April 07, 2024 at 16:54:59
 ELAPSED TIME: 00 m 03 s

Rows returned (1)

count
8685

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

4. Copying the data to DIM_CARD_TYPE table:

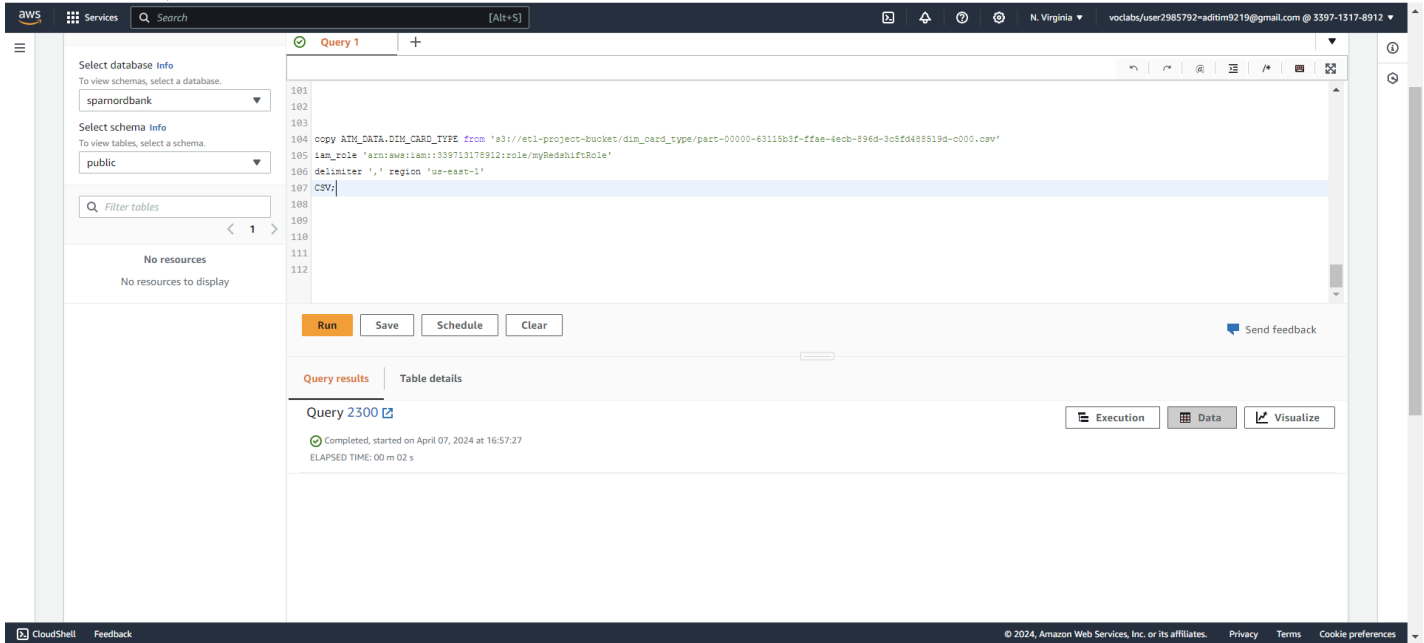
Query:

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;



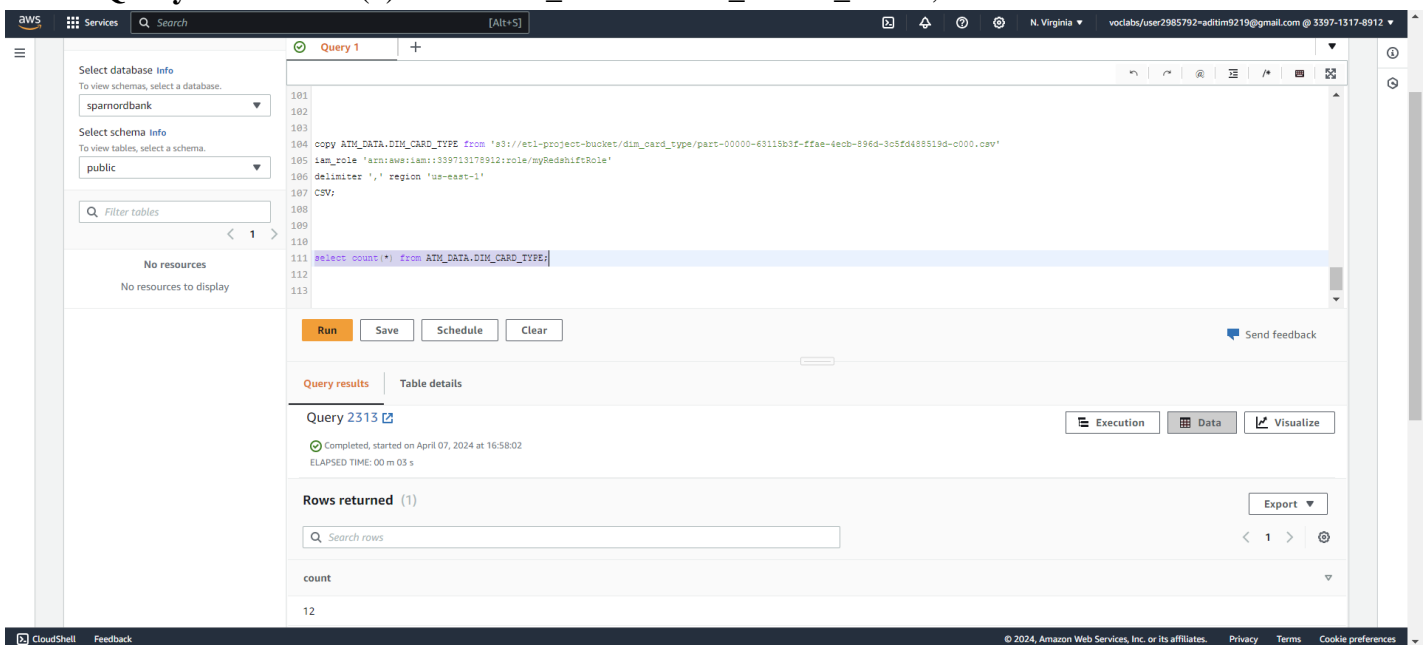
The screenshot shows the AWS Redshift Query Editor interface. On the left, the 'Select database' dropdown is set to 'sparnordbank' and the 'Select schema' dropdown is set to 'public'. The main query editor displays the following SQL command:

```
101
102
103
104 copy ATM_DATA.DIM_CARD_TYPE from 's3://etl-project-bucket/dim_card_type/part-00000-63115b3f-ffae-4ecb-896d-3c5fd488519d-c000.csv'
105 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
106 delimiter ',' region 'us-east-1'
107 CSV;
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab is selected, showing the status: 'Query 2300', 'Completed, started on April 07, 2024 at 16:57:27', and 'ELAPSED TIME: 00 m 02 s'.

Count of rows in DIM_CARD_TYPE table:

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



The screenshot shows the AWS Redshift Query Editor interface. The main query editor displays the following SQL command:

```
101
102
103
104 copy ATM_DATA.DIM_CARD_TYPE from 's3://etl-project-bucket/dim_card_type/part-00000-63115b3f-ffae-4ecb-896d-3c5fd488519d-c000.csv'
105 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
106 delimiter ',' region 'us-east-1'
107 CSV;
111 select count(*) from ATM_DATA.DIM_CARD_TYPE;
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab is selected, showing the status: 'Query 2313', 'Completed, started on April 07, 2024 at 16:58:02', and 'ELAPSED TIME: 00 m 03 s'.

The 'Rows returned' section shows 1 row:

count
12

Total number of rows retrieved is 12.

5. Copying the data to FACT_ATM_TRANS table:

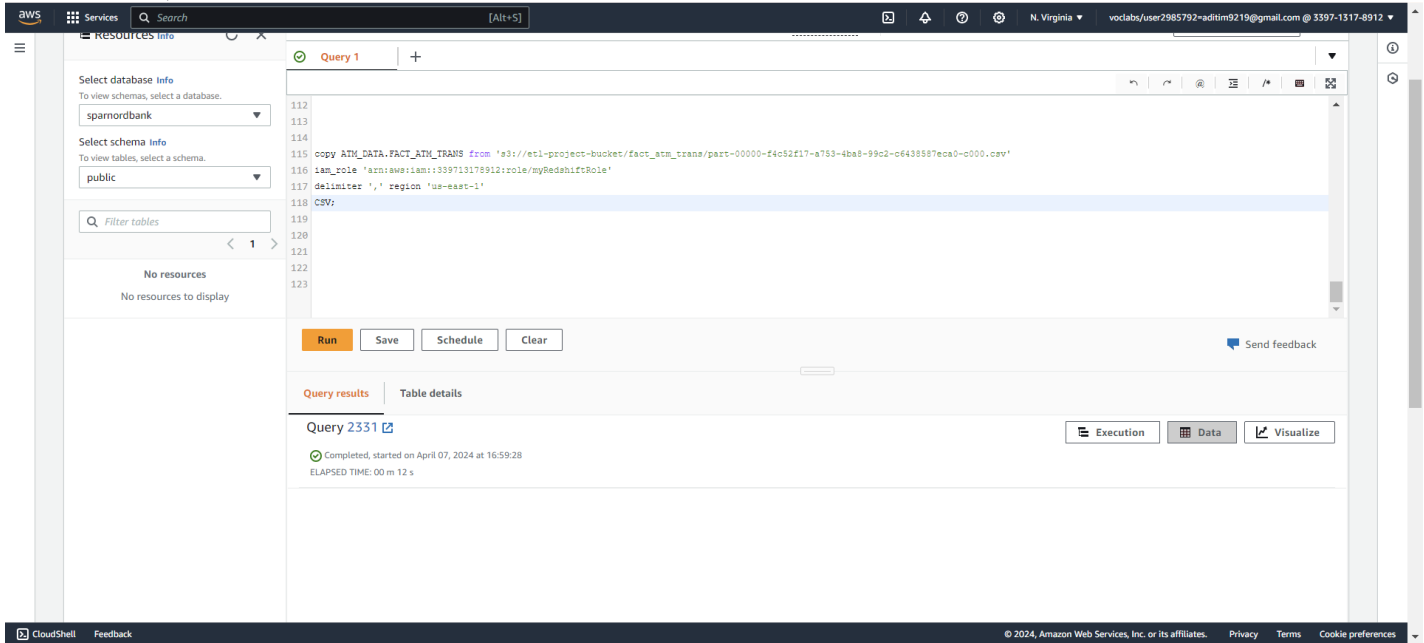
Query:

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;



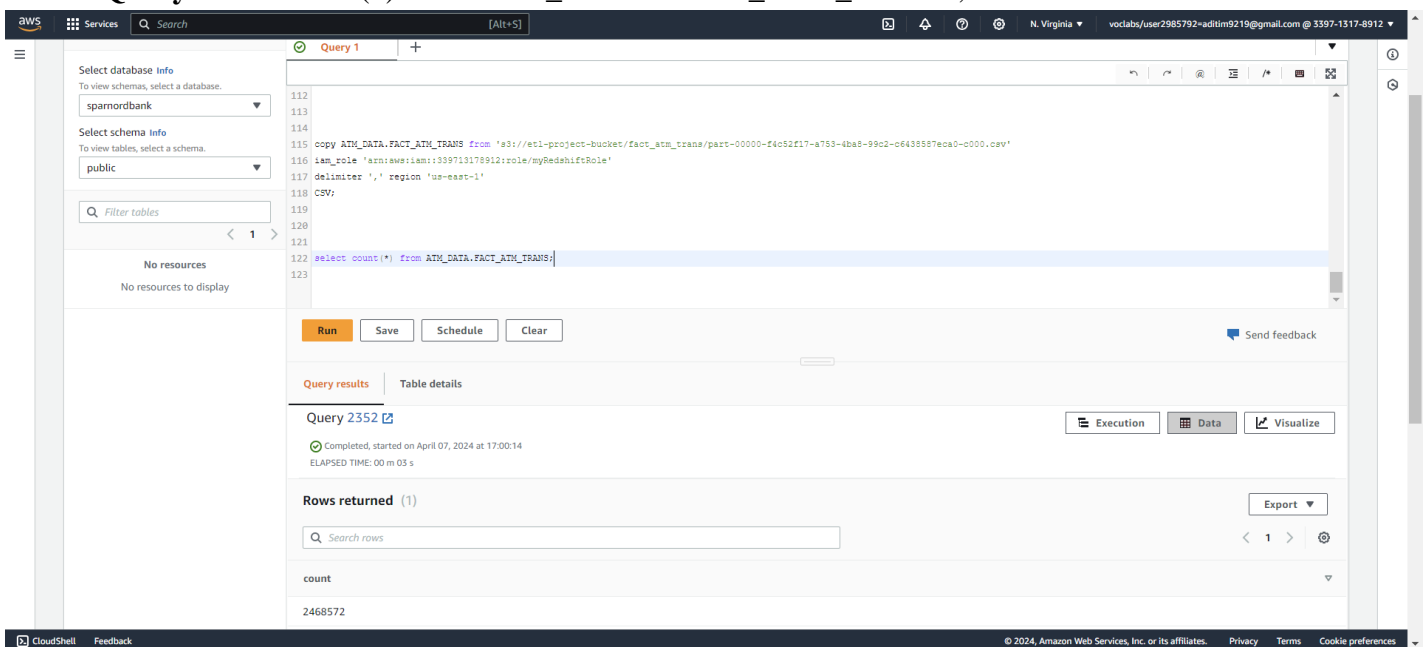
The screenshot shows the AWS Redshift Query Editor interface. On the left, the 'Resources' pane shows the selected database 'sparnordbank' and schema 'public'. The main editor area displays the following SQL query:

```
112 copy ATM_DATA.FACT_ATM_TRANS from 's3://etl-project-bucket/fact_atm_trans/part-00000-f4c52f17-a753-4ba8-99c2-c6438587eca0-c000.csv'
113
114
115
116 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
117 delimiter ',' region 'us-east-1'
118 CSV;
119
120
121
122
123
```

Below the query editor, the 'Query results' tab shows the execution status: 'Completed, started on April 07, 2024 at 16:59:28' and 'ELAPSED TIME: 00 m 12 s'.

Count of rows in FACT_ATM_TRANS table:

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



The screenshot shows the AWS Redshift Query Editor interface. The main editor area displays the following SQL query:

```
112 copy ATM_DATA.FACT_ATM_TRANS from 's3://etl-project-bucket/fact_atm_trans/part-00000-f4c52f17-a753-4ba8-99c2-c6438587eca0-c000.csv'
113
114
115
116 iam_role 'arn:aws:iam::339713178912:role/myRedshiftRole'
117 delimiter ',' region 'us-east-1'
118 CSV;
119
120
121
122 select count(*) from ATM_DATA.FACT_ATM_TRANS;
123
```

Below the query editor, the 'Query results' tab shows the execution status: 'Completed, started on April 07, 2024 at 17:00:14' and 'ELAPSED TIME: 00 m 05 s'.

The 'Rows returned' section shows 1 row:

count
2468572

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