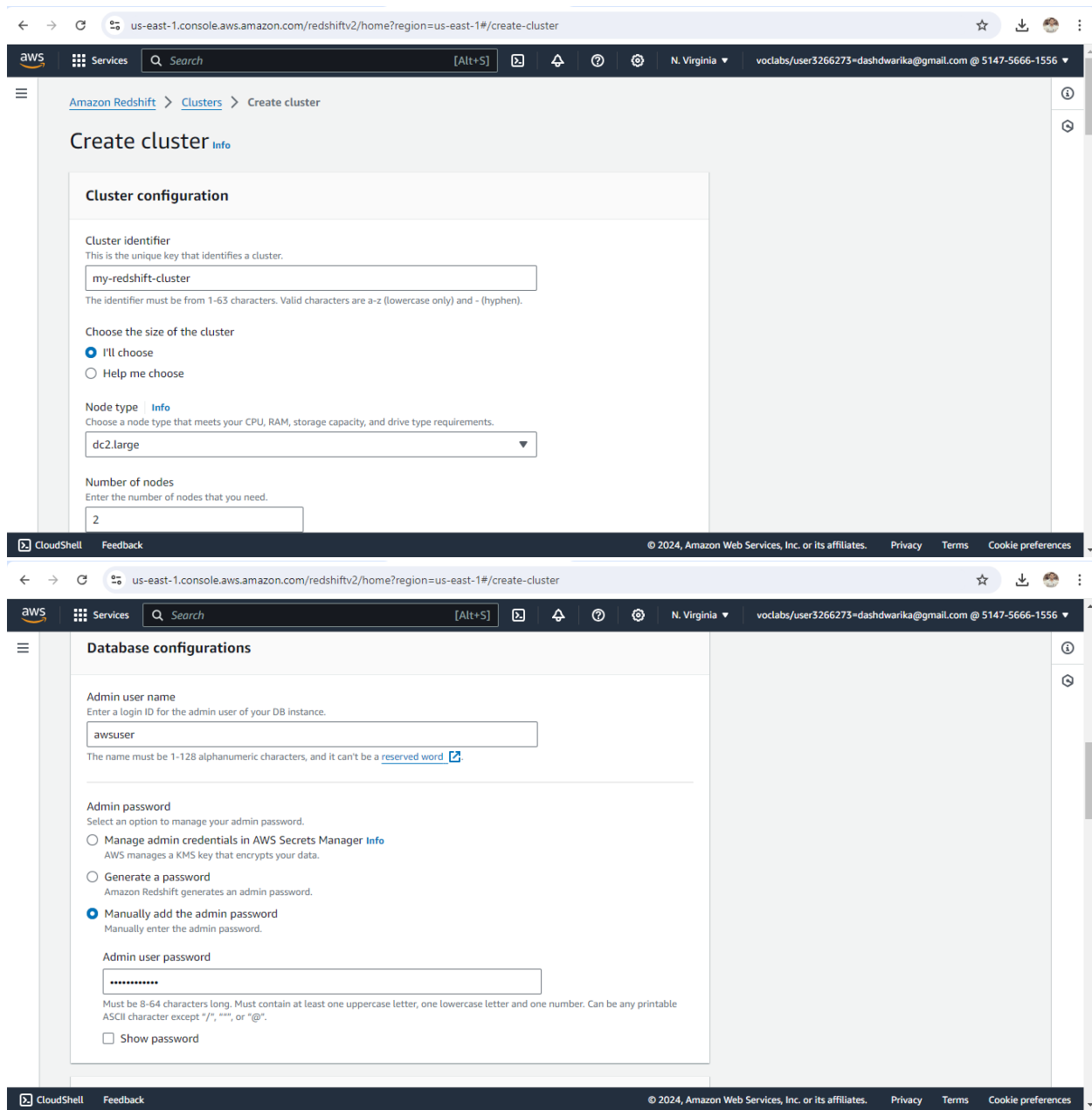


Creation of a Redshift Cluster



The screenshot displays the Amazon Redshift console interface for creating a new cluster. The browser address bar shows the URL: `us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#/create-cluster`. The console header includes the AWS logo, a search bar, and navigation links for Services, CloudShell, and Feedback. The user's profile and account information are visible in the top right corner.

The main content area is titled "Create cluster" and includes an "Info" link. It is divided into two sections: "Cluster configuration" and "Database configurations".

Cluster configuration

- Cluster identifier:** A text input field contains "my-redshift-cluster". Below it, a note states: "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)." A small icon indicates a tooltip is available.
- Choose the size of the cluster:** Two radio buttons are present: "I'll choose" (selected) and "Help me choose".
- Node type:** A dropdown menu shows "dc2.large". Above it, a note says: "Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements." A small icon indicates a tooltip is available.
- Number of nodes:** A text input field contains "2". Above it, a note says: "Enter the number of nodes that you need."

Database configurations

- Admin user name:** A text input field contains "awsuser". Below it, a note states: "Enter a login ID for the admin user of your DB instance. The name must be 1-128 alphanumeric characters, and it can't be a reserved word." A small icon indicates a tooltip is available.
- Admin password:** Three radio buttons are present: "Manage admin credentials in AWS Secrets Manager" (unselected), "Generate a password" (unselected), and "Manually add the admin password" (selected). Below the selected option, a note says: "Manually enter the admin password." A text input field contains "*****". Below the field, a note states: "Must be 8-64 characters long. Must contain at least one uppercase letter, one lowercase letter and one number. Can be any printable ASCII character except '/', '!', ' ', or '@'." A checkbox labeled "Show password" is present.

The footer of the console shows the copyright notice: "© 2024, Amazon Web Services, Inc. or its affiliates." and links for Privacy, Terms, and Cookie preferences.

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#/create-cluster

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3266273=dashdwarika@gmail.com @ 5147-5666-1556

Cluster permissions

Create an IAM role as the default for this cluster that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

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.

Search for associated IAM role by name, status, or role type < 1 >

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	myRedshiftRole	Not applied	--

Additional configurations Use defaults

CloudShell Feedback

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

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#/create-cluster

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3266273=dashdwarika@gmail.com @ 5147-5666-1556

Additional configurations Use defaults

These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

▼ Network and security Info

Virtual private cloud (VPC)
This VPC defines the virtual networking environment for this cluster.

my-vpc-vpc
vpc-0ba7c512e1941c9cb

You can't change the VPC associated with this cluster after the cluster has been created. [Learn more about getting started cluster in vpc](#)

X

VPC security groups
This VPC security group defines which subnets and IP ranges the cluster can use in the VPC. For more information, see [Learn more about Redshift clusters security groups](#).

Choose one or more security groups

my-vpc-security X
sg-0ab3236271d76b388

Cluster subnet group Info
Choose the Amazon Redshift subnet group to launch the cluster in

CloudShell Feedback

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

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#/create-cluster

aws Services Search [Alt+S]

N. Virginia voclabs/user3266273=dashdwarika@gmail.com @ 5147-5666-1556

You can't change the VPC associated with this cluster after the cluster has been created. [Learn more about getting started cluster in vpc](#)

VPC security groups

This VPC security group defines which subnets and IP ranges the cluster can use in the VPC. For more information, see [Learn more about Redshift clusters security groups](#)

Choose one or more security groups

my-vpc-security sg-0ab3236271d76b388

Cluster subnet group Info

Choose the Amazon Redshift subnet group to launch the cluster in.

my-cluster-subnet-group

Create new subnet group

Availability Zone

Specify the Availability Zone to create the cluster in. Otherwise, Amazon Redshift chooses an Availability Zone for you.

us-east-1a

Enhanced VPC routing

Enabling this option routes network traffic between your cluster and data repositories through a VPC, instead of through the internet. [Learn more about getting started cluster in vpc](#)

☒ Turn off

☐ Turn on

Publicly accessible

CloudShell Feedback

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

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#/create-cluster

aws Services Search [Alt+S]

N. Virginia voclabs/user3266273=dashdwarika@gmail.com @ 5147-5666-1556

Database configurations Info

Database name

Specify a database name to create an additional database.

dev

The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a [reserved word](#)

Database port

Port number where the database accepts inbound connections. The default port is 5439.

5439

Choose a port number between 1150 and 65535.

Parameter groups

Defines database parameter and query queues for all the databases.

default.redshift-1.0

AWS KMS

Choose the key to use.

☐ Use AWS Key Management Service (AWS KMS)

Select to encrypt your cluster's data using AWS Key Management Service.

Maintenance Info

CloudShell Feedback

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

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#/create-cluster

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3266273-dashdwarika@gmail.com @ 5147-5666-1556

5439

Choose a port number between 1150 and 65535.

Parameter groups

Defines database parameter and query queues for all the databases.

default.redshift-1.0

AWS KMS

Choose the key to use.

☐ Use AWS Key Management Service (AWS KMS)

Select to encrypt your cluster's data using AWS Key Management Service.

Maintenance Info

Monitoring

Backup

Cancel

Create cluster

CloudShell

Feedback

© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#/clusters

aws

Services

Search

[Alt+S]

N. Virginia

voclabs/user3266273-dashdwarika@gmail.com @ 5147-5666-1556

Create a cluster with preview features. Production use of the cluster is not supported. Use this cluster for testing only.

Use the query editor v2 to run queries in your Redshift cluster.

Query data

You can connect to Amazon Redshift from your client tools, such as SQL clients, business intelligence (BI) tools, and extract, transform, load (ETL) tools, using JDBC or ODBC drivers.

Cluster

my-redshift-cluster

Copy JDBC URL

Copy ODBC URL

Use JDBC or ODBC drivers to connect to Amazon Redshift from your client tools, such as SQL clients, BI tools, and ETL tools. We recommend using the new Amazon Redshift-specific drivers for better performance and scalability.

Driver

JDBC 4.2 without AWS SDK (.jar)

Download driver

Clusters (1) Info

Query data

Actions

Create cluster

Filter clusters by property or value

	Cluster	Status	Cluster namespace	Availability Zone	Multi-AZ	Storage capacity us...
<input type="checkbox"/>	my-redshift-cluster dc2.large 2 nodes 320 GB	Available	5b111d2a-6d1a-43df-...	us-east-1a	No	< 1%

CloudShell

Feedback

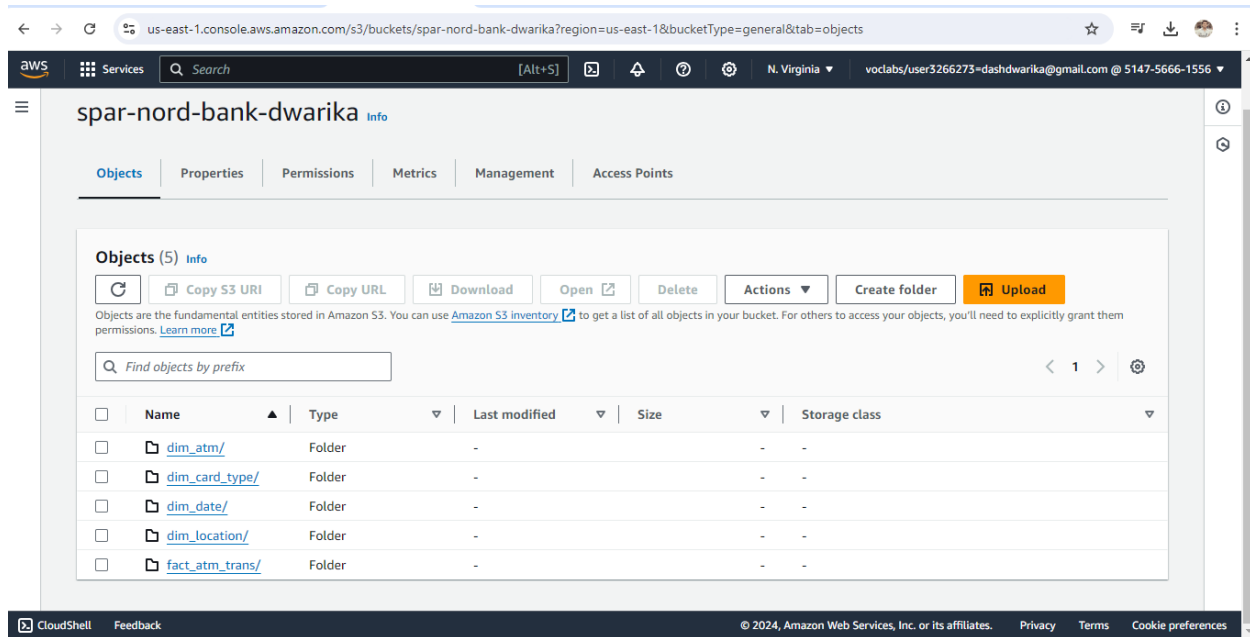
© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

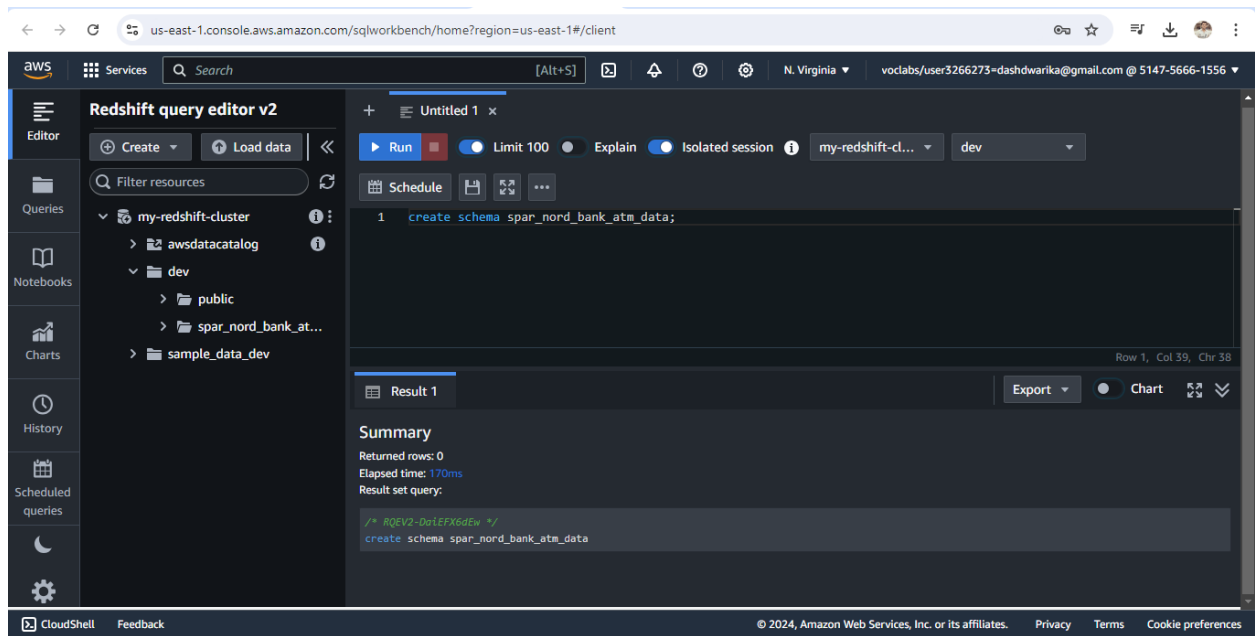
S3 bucket containing files:



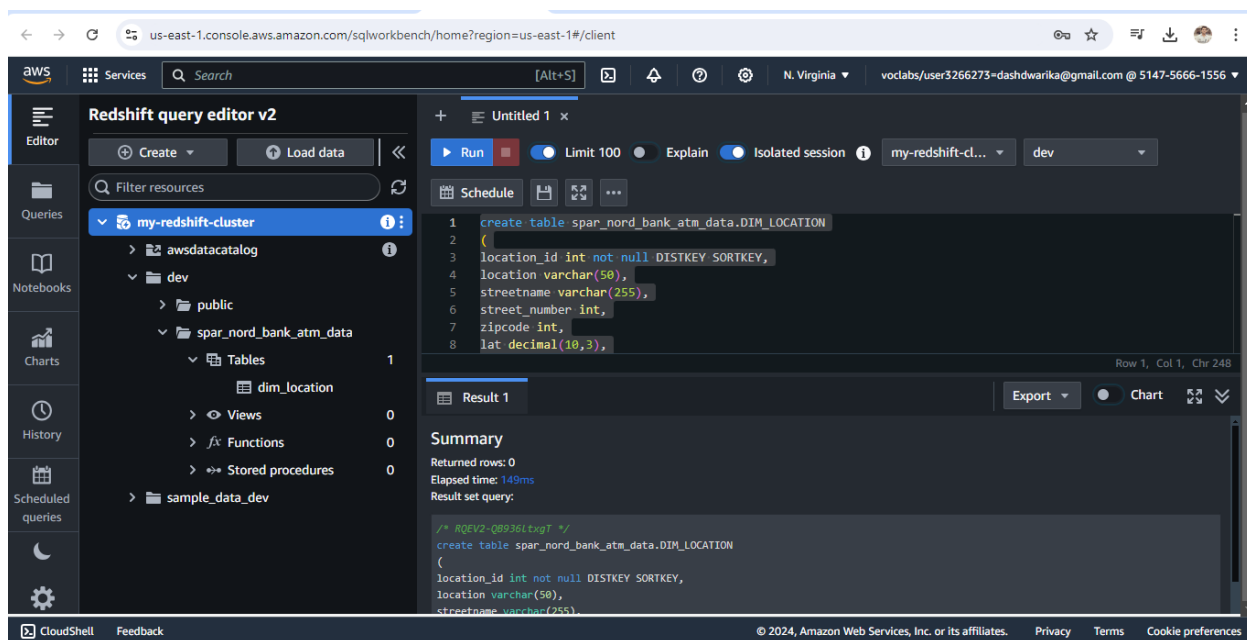
Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables

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

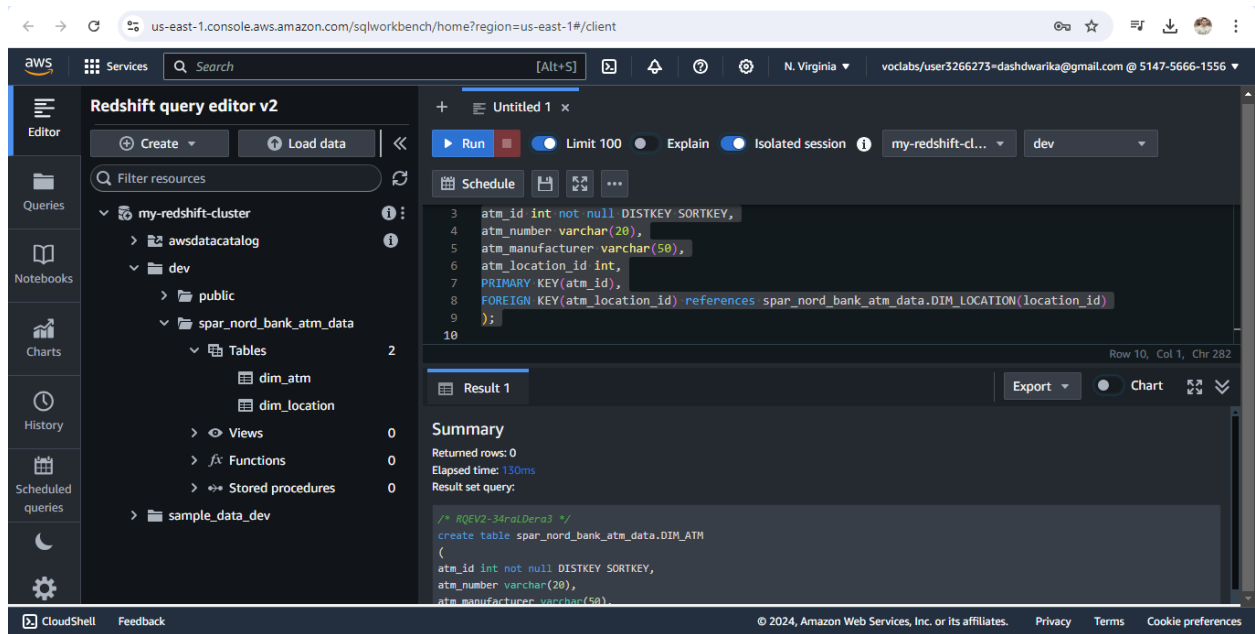
```
create schema spar_nord_bank_atm_data;
```



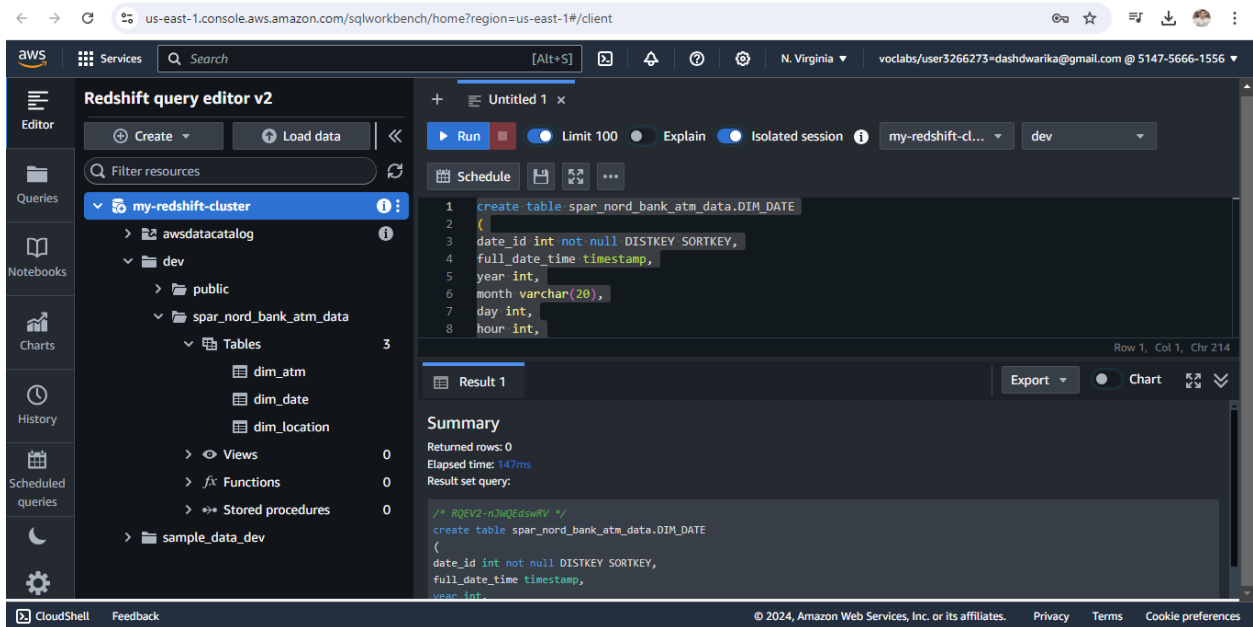
```
create table spar_nord_bank_atm_data.DIM_LOCATION
(
location_id int not null DISTKEY SORTKEY,
location varchar(50),
streetname varchar(255),
street_number int,
zipcode int,
lat decimal(10,3),
lon decimal(10,3),
PRIMARY KEY(location_id)
);
```



```
create table spar_nord_bank_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
spar_nord_bank_atm_data.DIM_LOCATION(location_id)
);
```



```
create table spar_nord_bank_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 shows the AWS Redshift Query Editor v2 interface. The left sidebar displays the resource tree for 'my-redshift-cluster', including 'awsdatacatalog', 'dev', 'public', 'spar_nord_bank_atm_data', 'Tables', 'Views', 'Functions', and 'Stored procedures'. The main editor area shows a SQL query to create a table named 'spar_nord_bank_atm_data.DIM_DATE'. The query is as follows:

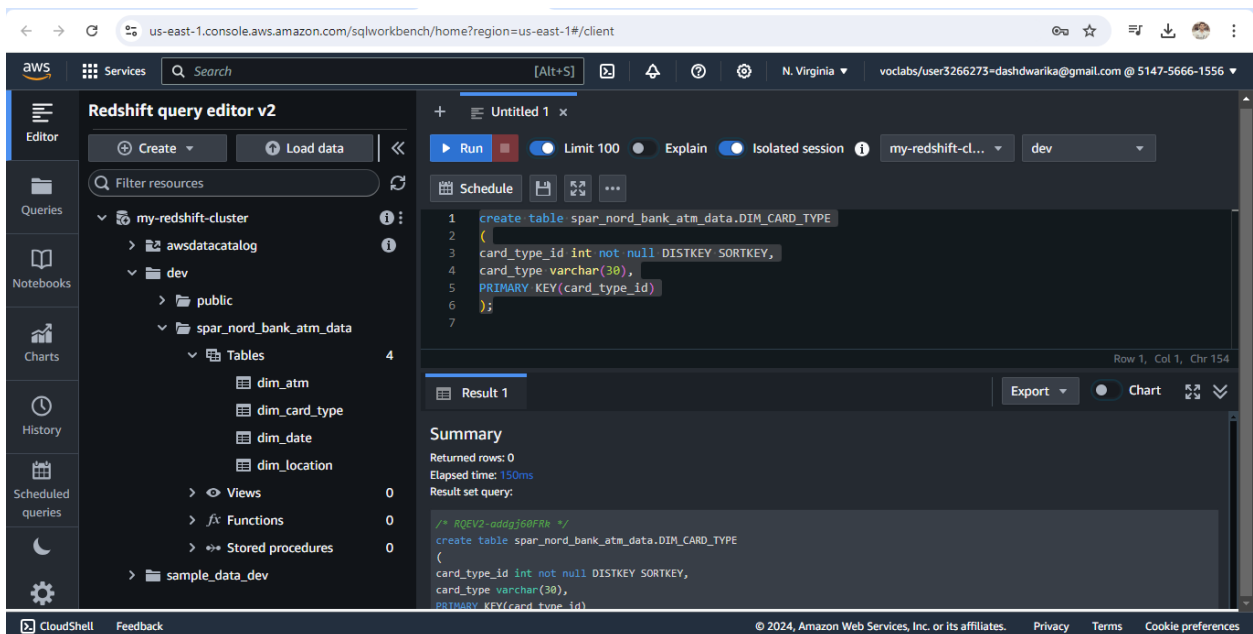
```
1 create table spar_nord_bank_atm_data.DIM_DATE
2 (
3   date_id int not null DISTKEY SORTKEY,
4   full_date_time timestamp,
5   year int,
6   month varchar(20),
7   day int,
8   hour int,
```

The 'Result 1' section shows a summary of the query execution:

- Returned rows: 0
- Elapsed time: 147ms
- Result set query:

The bottom of the interface shows the AWS CloudShell and Feedback buttons, along with the copyright notice for Amazon Web Services, Inc. or its affiliates.

```
create table spar_nord_bank_atm_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 Redshift Query Editor v2 interface. The left sidebar displays the resource tree for 'my-redshift-cluster', including 'awsdatacatalog', 'dev', 'public', 'spar_nord_bank_atm_data', 'Tables', 'Views', 'Functions', and 'Stored procedures'. The main editor area shows a SQL query to create a table named 'spar_nord_bank_atm_data.DIM_CARD_TYPE'. The query is as follows:

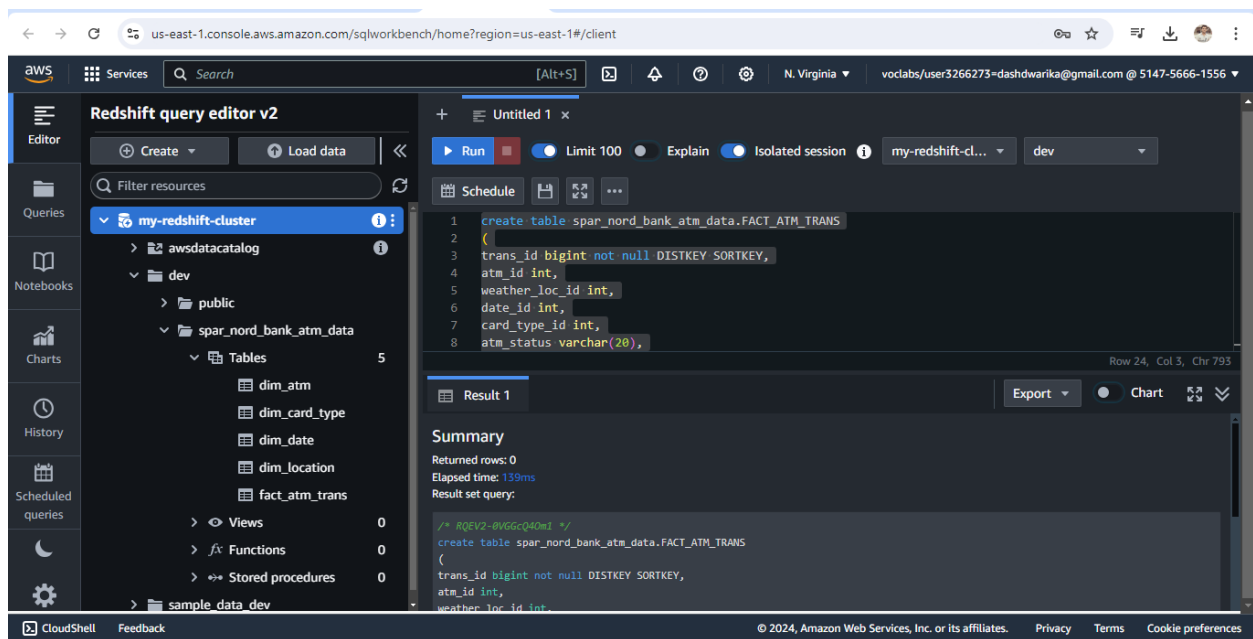
```
1 create table spar_nord_bank_atm_data.DIM_CARD_TYPE
2 (
3   card_type_id int not null DISTKEY SORTKEY,
4   card_type varchar(30),
5   PRIMARY KEY(card_type_id)
6 );
7
```

The 'Result 1' section shows a summary of the query execution:

- Returned rows: 0
- Elapsed time: 150ms
- Result set query:

The bottom of the interface shows the AWS CloudShell and Feedback buttons, along with the copyright notice for Amazon Web Services, Inc. or its affiliates.

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

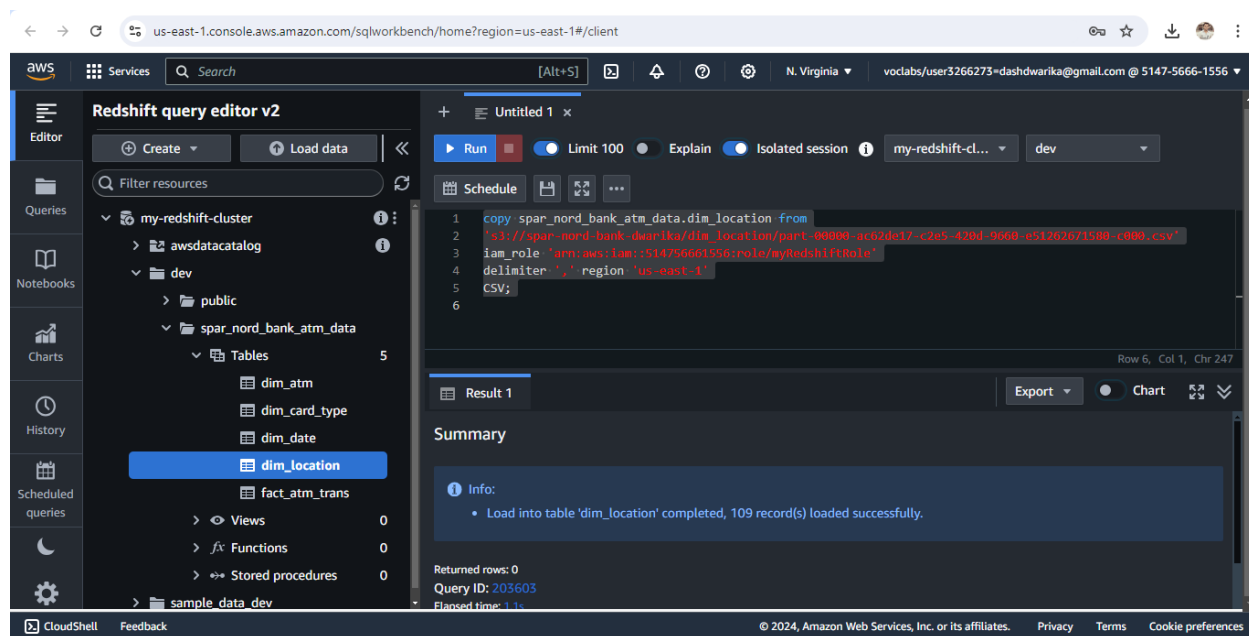


The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar displays the 'my-redshift-cluster' resource tree, including tables like 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The main editor area contains a SQL query to create a table 'FACT_ATM_TRANS' with various attributes and foreign key references. The query is executed, and the results are displayed in the 'Result 1' tab, showing a summary of the execution: 'Returned rows: 0', 'Elapsed time: 139ms', and 'Result set query: create table spar_nord_bank_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))'.

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

```
copy spar_nord_bank_atm_data.dim_location from
's3://spar-nord-bank-dwarika/dim_location/part-00000-ac62de17-c2e5-420d-9660-
e51262671580-c000.csv'
iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar displays the resource tree with the following structure:

- my-redshift-cluster
 - awsdatacatalog
 - dev
 - public
 - spar_nord_bank_atm_data
 - Tables (5)
 - dim_atm
 - dim_card_type
 - dim_date
 - dim_location** (selected)
 - fact_atm_trans
 - Views (0)
 - Functions (0)
 - Stored procedures (0)
 - sample_data_dev

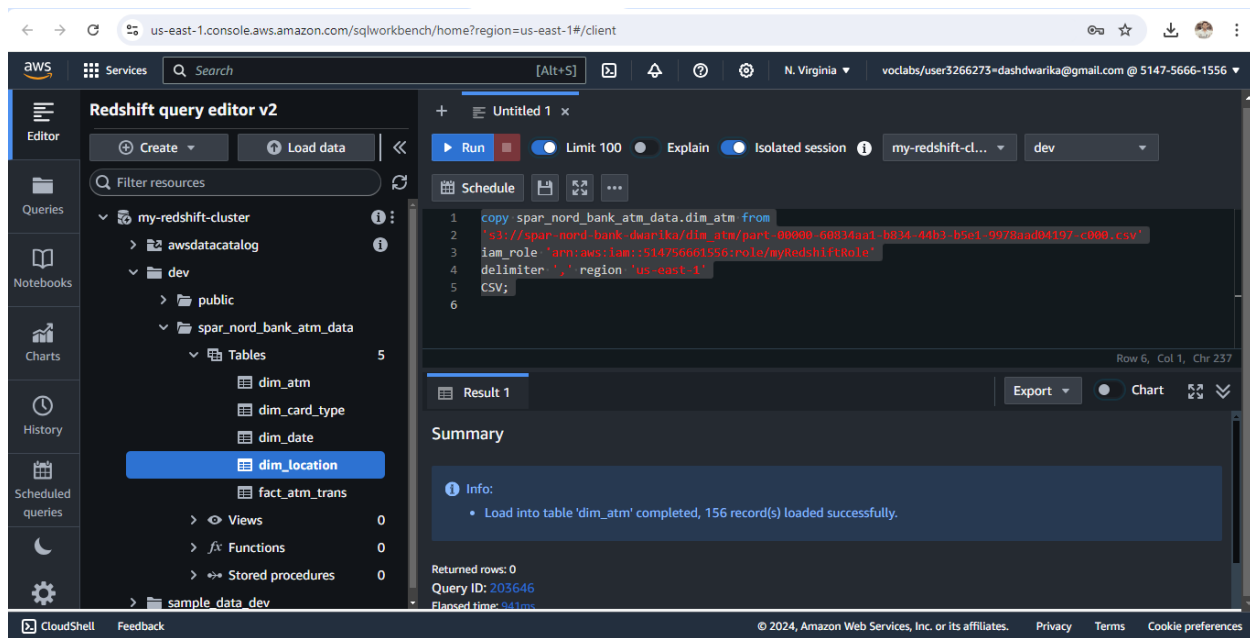
The main editor area shows the following SQL query:

```
1 copy spar_nord_bank_atm_data.dim_location from
2 's3://spar-nord-bank-dwarika/dim_location/part-00000-ac62de17-c2e5-420d-9660-e51262671580-c000.csv'
3 iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1'
5 CSV;
6
```

The query was executed successfully, and the summary shows:

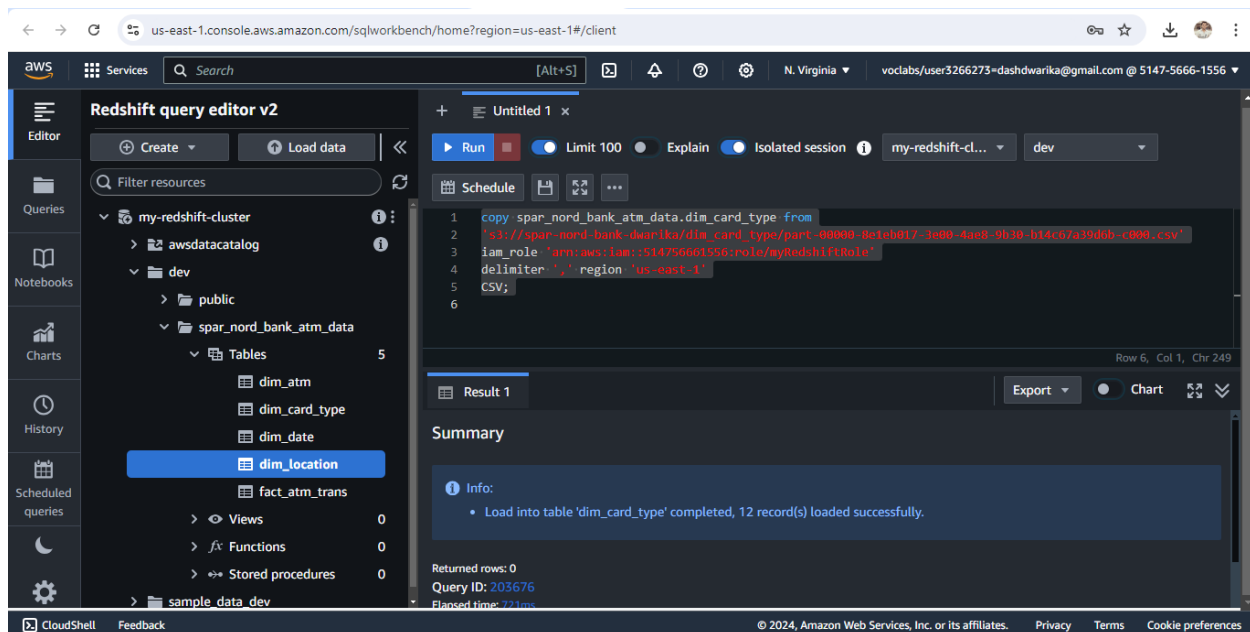
- Info: Load into table 'dim_location' completed, 109 record(s) loaded successfully.
- Returned rows: 0
- Query ID: 203603
- Elapsed time: 1.1s

```
copy spar_nord_bank_atm_data.dim_atm from
's3://spar-nord-bank-dwarika/dim_atm/part-00000-60834aa1-b834-44b3-b5e1-
9978aad04197-c000.csv'
iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift Query Editor v2 interface. The left sidebar displays the resource tree with 'my-redshift-cluster' expanded, showing 'awsdatacatalog' and 'dev' schemas. The 'dev' schema is expanded, showing 'public' and 'spar_nord_bank_atm_data' schemas. The 'spar_nord_bank_atm_data' schema is expanded, showing 'Tables' and 'Views'. The 'Tables' section is expanded, showing 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The 'dim_location' table is selected. The main editor shows a SQL query: `copy spar_nord_bank_atm_data.dim_atm from 's3://spar-nord-bank-dwarika/dim_atm/part-00000-68834aa1-b834-44b3-b5e1-9978aad04197-c000.csv' iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;`. The 'Run' button is clicked, and the 'Result 1' tab shows a summary: 'Load into table 'dim_atm' completed, 156 record(s) loaded successfully.' The 'Summary' section also shows 'Returned rows: 0' and 'Query ID: 203646'.

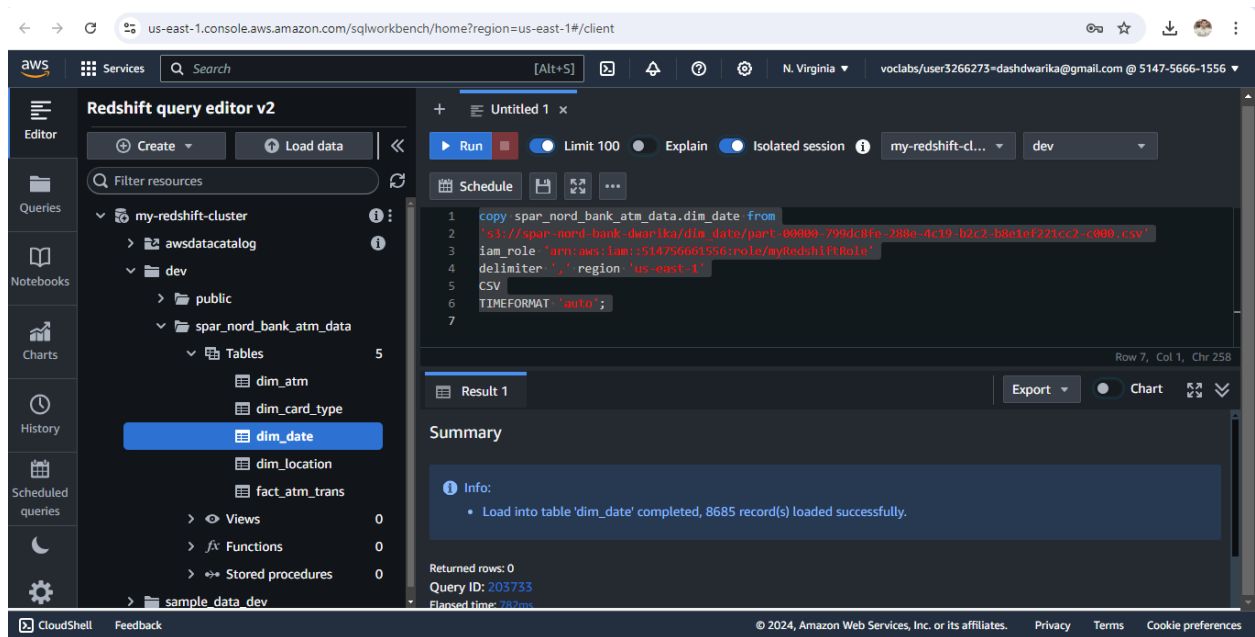
```
copy spar_nord_bank_atm_data.dim_card_type from
's3://spar-nord-bank-dwarika/dim_card_type/part-00000-8e1eb017-3e00-4ae8-9b30-
b14c67a39d6b-c000.csv'
iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift Query Editor v2 interface. The left sidebar displays the resource tree with 'my-redshift-cluster' expanded, showing 'awsdatacatalog' and 'dev' schemas. The 'dev' schema is expanded, showing 'public' and 'spar_nord_bank_atm_data' schemas. The 'spar_nord_bank_atm_data' schema is expanded, showing 'Tables' and 'Views'. The 'Tables' section is expanded, showing 'dim_atm', 'dim_card_type', 'dim_date', 'dim_location', and 'fact_atm_trans'. The 'dim_location' table is selected. The main editor shows a SQL query: `copy spar_nord_bank_atm_data.dim_card_type from 's3://spar-nord-bank-dwarika/dim_card_type/part-00000-8e1eb017-3e00-4ae8-9b30-b14c67a39d6b-c000.csv' iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole' delimiter ',' region 'us-east-1' CSV;`. The 'Run' button is clicked, and the 'Result 1' tab shows a summary: 'Load into table 'dim_card_type' completed, 12 record(s) loaded successfully.' The 'Summary' section also shows 'Returned rows: 0' and 'Query ID: 203676'.

```
copy spar_nord_bank_atm_data.dim_date from
```

```
's3://spar-nord-bank-dwarika/dim_date/part-00000-799dc8fe-288e-4c19-b2c2-b8e1ef221cc2-c000.csv'
iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';
```



The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar displays the resource tree with the 'dim_date' table selected. The main editor area contains the following SQL query:

```
copy spar_nord_bank_atm_data.dim_date from
's3://spar-nord-bank-dwarika/dim_date/part-00000-799dc8fe-288e-4c19-b2c2-b8e1ef221cc2-c000.csv'
iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';
```

The query has been executed successfully, as indicated by the 'Summary' section which states: 'Load into table 'dim_date' completed, 8685 record(s) loaded successfully.' The interface also shows the 'Result 1' tab and the 'Export' button.

```
copy spar_nord_bank_atm_data.fact_atm_trans from
's3://spar-nord-bank-dwarika/fact_atm_trans/part-00000-fd716829-86aa-4836-b5c6-49443a3cd2f7-c000.csv'
iam_role 'arn:aws:iam::514756661556:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

us-east-1.console.aws.amazon.com/sqlworkbench/home?region=us-east-1#/client

aws Services Search [Alt+S] N. Virginia voclabs/user3266273-dashdwarka@gmail.com @ 5147-5666-1556

Redshift query editor v2

Create Load data

Filter resources

- dim_card_type
- dim_date
- dim_location
- fact_atm_trans**

Views 0

fact_atm_trans			
atm_id	integer	NULL	
weather_loc_id	integer	NULL	
date_id	integer	NULL	
card_type_id	integer	NULL	
atm_status	character varying(20)	NULL	
currency	character varying(10)	NULL	

```
1 copy_spar_nord_bank_atm_data.fact_atm_trans from
2 's3://spar-nord-bank-dashdwarka/fact_atm_trans/part-000000-td710029-80aa-4036-b5c6-40441a3c0217-c900.csv'
3 iam_role 'arn:aws:iam::514756661556:role/myredshiftRole'
4 delimiter ',' region 'us-east-1'
5 CSV;
6
```

Run Limit 100 Explain Isolated session my-redshift-cl... dev

Schedule

Result 1

Export Chart

Summary

Info:

- Load into table 'fact_atm_trans' completed, 2468572 record(s) loaded successfully.

Returned rows: 0
Query ID: 204597
Elapsed time: 6.9s

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