

Creation of a Redshift Cluster

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

<Screenshot of the type of machine used along with number of nodes>

redshift-cluster-etl

Actions ▼

Edit

Add partner integration


Query data ▼

General information [Info](#)

Cluster identifier

redshift-cluster-etl


Status

 Available

Node type

dc2.large

Endpoint

 redshift-cluster-etl.ciftgsiaywig.us-east-1.redshift.amazonaws.com:5439/dev

Custom domain name

-


Date created

December 07, 2023, 18:28 (UTC+05:30)


Number of nodes

2

JDBC URL

 jdbc:redshift://redshift-cluster-etl.ciftgsiaywig.us-east-1.redshift.amazonaws.com:5439/dev

Cluster namespace ARN

 arn:aws:redshift:us-east-1:265006242993:namespace:32b36f7d-b6fd-475d-824d-bd6e48a36f97


Storage used

-

Multi-AZ

No

ODBC URL

 Driver={Amazon Redshift (x64)}; Server=redshift-cluster-etl.ciftgsiaywig.us-east-1.redshift.amazonaws.com; Database=dev

Cluster configuration

Production

Types of Machine used dc2.large
Number of nodes=2

Database configurations [Info](#)

Edit admin credentials

Rotate encryption keys

Edit ▼

Database name

dev

Parameter group

Defines database parameter and query queues for all the databases.
[default.redshift-1.0](#)

Encryption

Disabled


Audit logging

Disabled

Port

5439

SSH ingestion setting (cluster public key)

 ssh-rsa
 AAAAB3NzaC1yc2EAAAADAQABAAQCaFc5
 sdtFCg0KVM9w6WWb8vGPwrculH8tMib3b8Nk
 a6Qfbj3+Gy8Cz7zyUC8PWdapGuk1g7cbQW+
 3t4Y1ZzRxn2f8v9YRz+YifhLqaa4ZBusC2vXV
 HZtwjT1Ggdtb0AyKkeFZH3xlyHmHnIrMNNk
 1mRpi7LUTH3h4NFQUQeXvWcOfnsAmsg/AGUB
 6wcHjpTi6itgNKP5i51elgtV0VqU/wf40PQGuX
 Mclgy45yQL8w0UjEXjg/IsQYkvNYHE3hMKMm
 IdAMKjy0pHzK7KLsOkS1GRBI782JtnKqkyz3L
 9v9snrP0NsEnRYboGspktyhvpva3xTG1ykukUS
 1n Amazon-Redshift

AWS KMS key ID

-

Admin user name

awsuser

Network and security settings [Info](#)

Edit

Virtual private cloud (VPC)

[vpc-0fb5c1b69915899bb](#)


Availability Zone

us-east-1e

VPC security group

Specify which instances and devices can connect to the cluster.
[sg-0d293317db1416959](#)


Publicly accessible

Allow connections from outside the VPC.

Subnet group

[default](#)

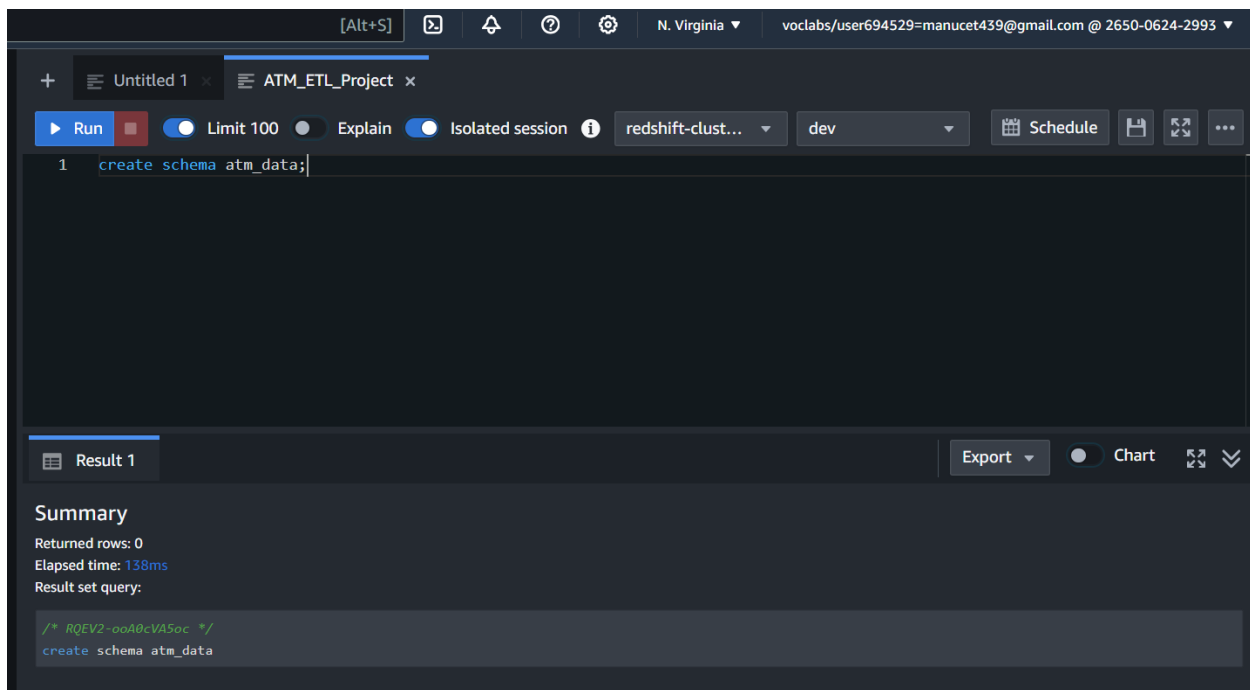
Enhanced VPC routing

Disabled

IP address type

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

create schema atm_data;



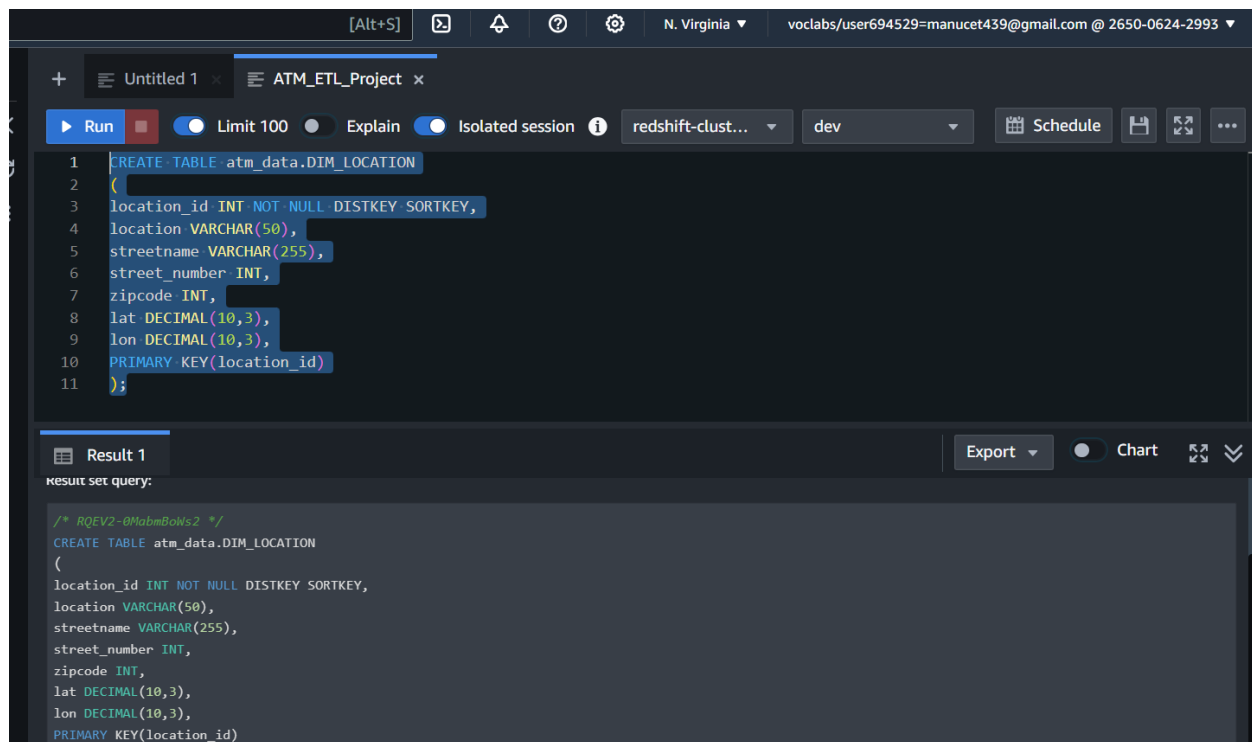
The screenshot shows the AWS Redshift console interface. At the top, there's a navigation bar with the AWS logo, a search bar, and user information. Below this, the console shows a list of clusters. The selected cluster is 'redshift-clust...'. The 'Query Editor' tab is active, displaying a SQL query: 'create schema atm_data;'. The query has been executed successfully, and the 'Result 1' tab shows a summary of the execution: 'Returned rows: 0', 'Elapsed time: 138ms', and 'Result set query: create schema atm_data'. The console also shows various settings like 'Limit 100', 'Explain', 'Isolated session', and 'Schedule'.

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

<Queries>

Query to create the dimension table DIM_LOCATION:

```
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)
);
```



```

1 CREATE TABLE atm_data.DIM_LOCATION
2 (
3   location_id INT NOT NULL DISTKEY SORTKEY,
4   location VARCHAR(50),
5   streetname VARCHAR(255),
6   street_number INT,
7   zipcode INT,
8   lat DECIMAL(10,3),
9   lon DECIMAL(10,3),
10  PRIMARY KEY(location_id)
11 );
  
```

Result 1

result set query:

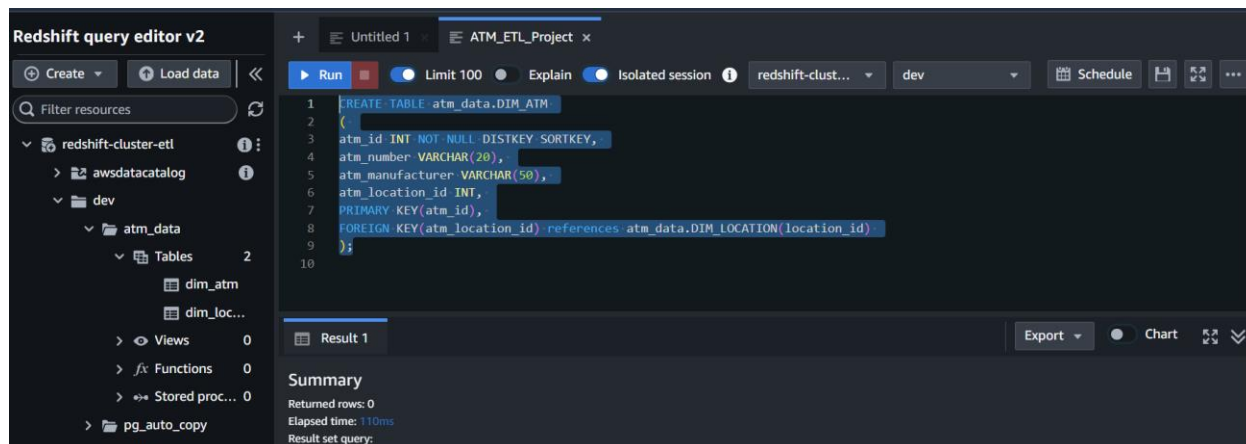
```

/* RQEV2-0MubmBoWs2 */
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)
)
  
```

Query to create the dimension table DIM_ATM:

```

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)
);
  
```



Redshift query editor v2

Filter resources

redshift-cluster-etl

awsdatacatalog

dev

atm_data

Tables

dim_atm

dim_loc...

Views

Functions

Stored proc...

pg_auto_copy

public

sample_data_dev

Untitled 1

ATM_ETL_Project

Run

Limit 100

Explain

Isolated session

redshift-clust...

dev

Schedule

Export

Chart

Summary

Returned rows: 0

Elapsed time: 110ms

Result set query:

```

1 CREATE TABLE atm_data.DIM_ATM
2 (
3   atm_id INT NOT NULL DISTKEY SORTKEY,
4   atm_number VARCHAR(20),
5   atm_manufacturer VARCHAR(50),
6   atm_location_id INT,
7   PRIMARY KEY(atm_id),
8   FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
9 );
10

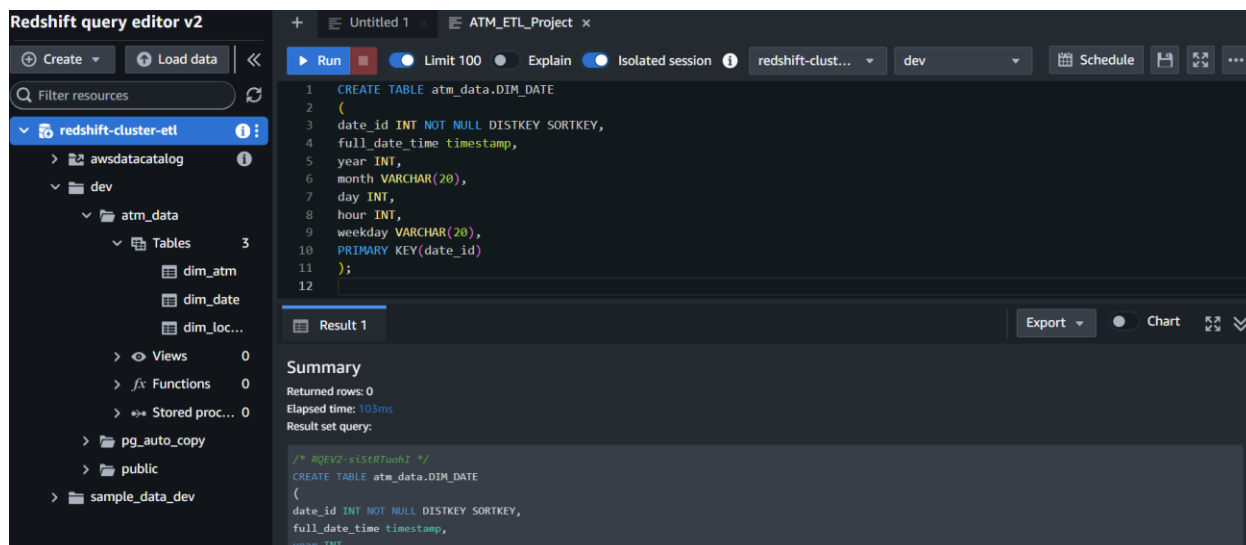
```

Query to create the dimension table DIM_DATE:

```

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)
);

```



Redshift query editor v2

Filter resources

redshift-cluster-etl

awsdatacatalog

dev

atm_data

Tables

dim_atm

dim_date

dim_loc...

Views

Functions

Stored proc...

pg_auto_copy

public

sample_data_dev

Untitled 1

ATM_ETL_Project

Run

Limit 100

Explain

Isolated session

redshift-clust...

dev

Schedule

Export

Chart

Summary

Returned rows: 0

Elapsed time: 103ms

Result set query:

```

1 CREATE TABLE 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,
9   weekday VARCHAR(20),
10  PRIMARY KEY(date_id)
11 );
12

```

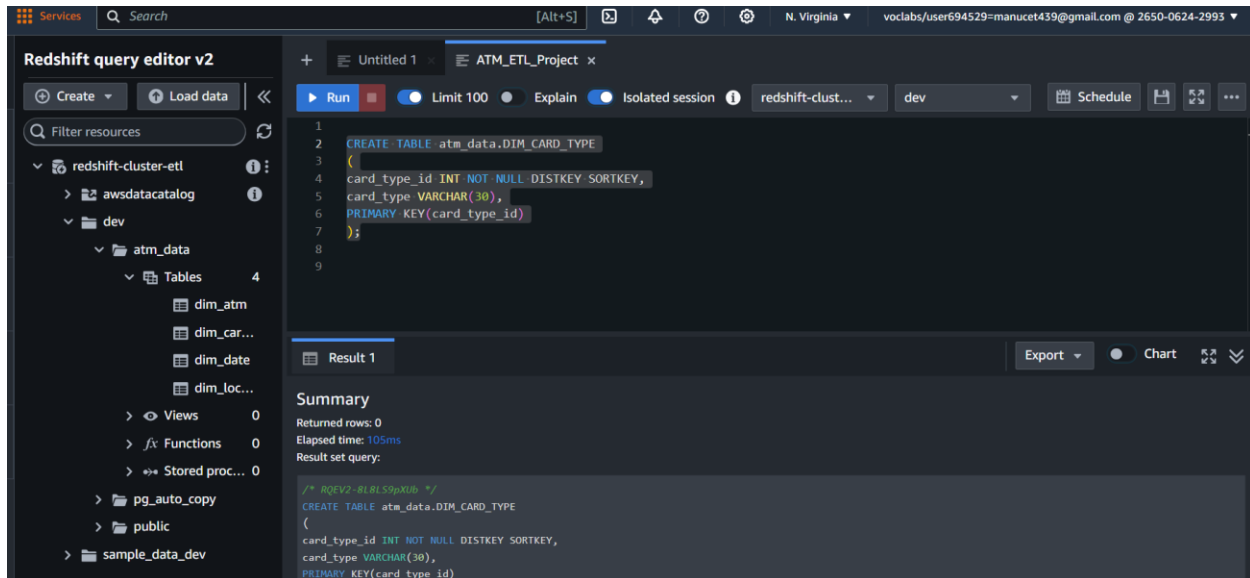
Query to create the dimension table DIM_CARD_TYPE:

```

CREATE TABLE atm_data.DIM_CARD_TYPE
(

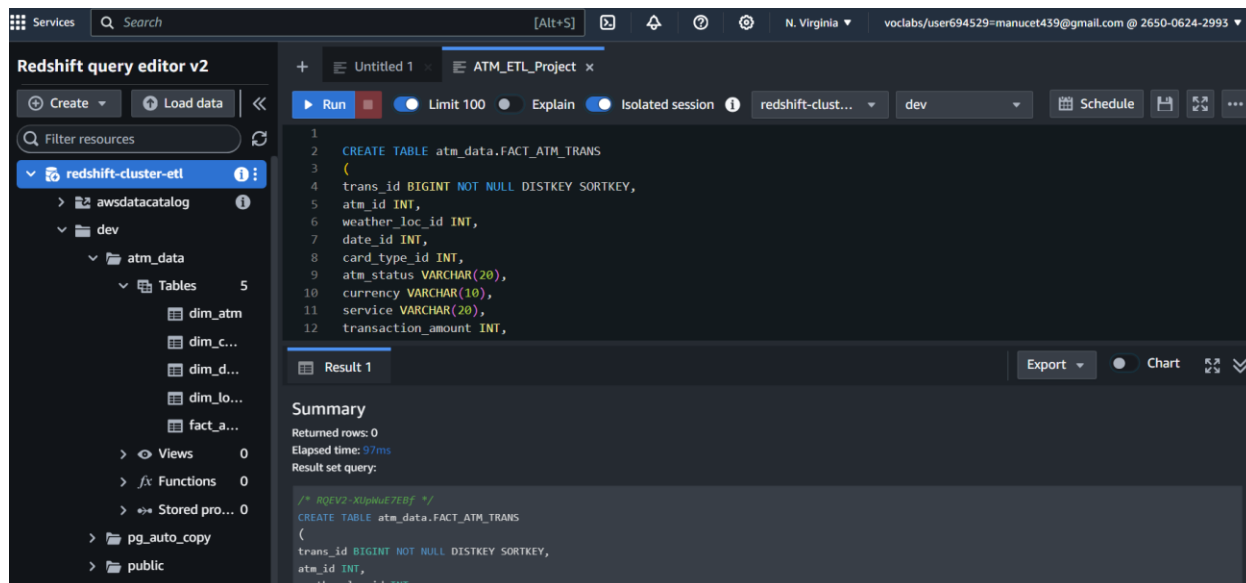
```

```
card_type_id INT NOT NULL DISTKEY SORTKEY,
card_type VARCHAR(30),
PRIMARY KEY(card_type_id)
);
```



Query to create the fact table FACT_ATM_TRANS:

```
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(255),
message_text VARCHAR(255),
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

<Queries>

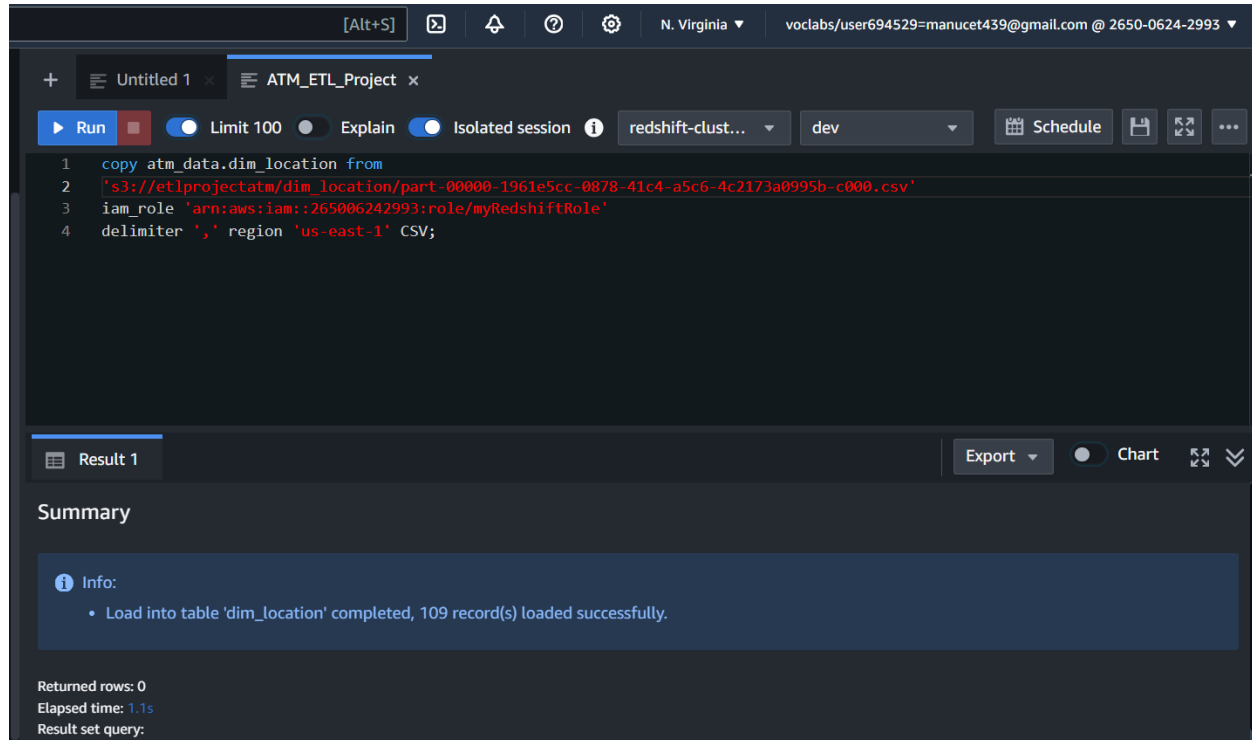
Query to copy the data from S3 bucket – etlprojectatm/dim_location folder to dim_location table

copy atm_data.dim_location from

's3://etlprojectatm/dim_location/part-00000-1961e5cc-0878-41c4-a5c6-4c2173a0995b-c000.csv'

iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'

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



The screenshot shows the Amazon Redshift console interface. At the top, there's a navigation bar with the AWS logo, a search bar, and user information. Below that, the console title is "ATM_ETL_Project". The main area displays a SQL query in a text editor:

```
1 copy atm_data.dim_location from
2 's3://etlprojectatm/dim_location/part-00000-1961e5cc-0878-41c4-a5c6-4c2173a0995b-c000.csv'
3 iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1' CSV;
```

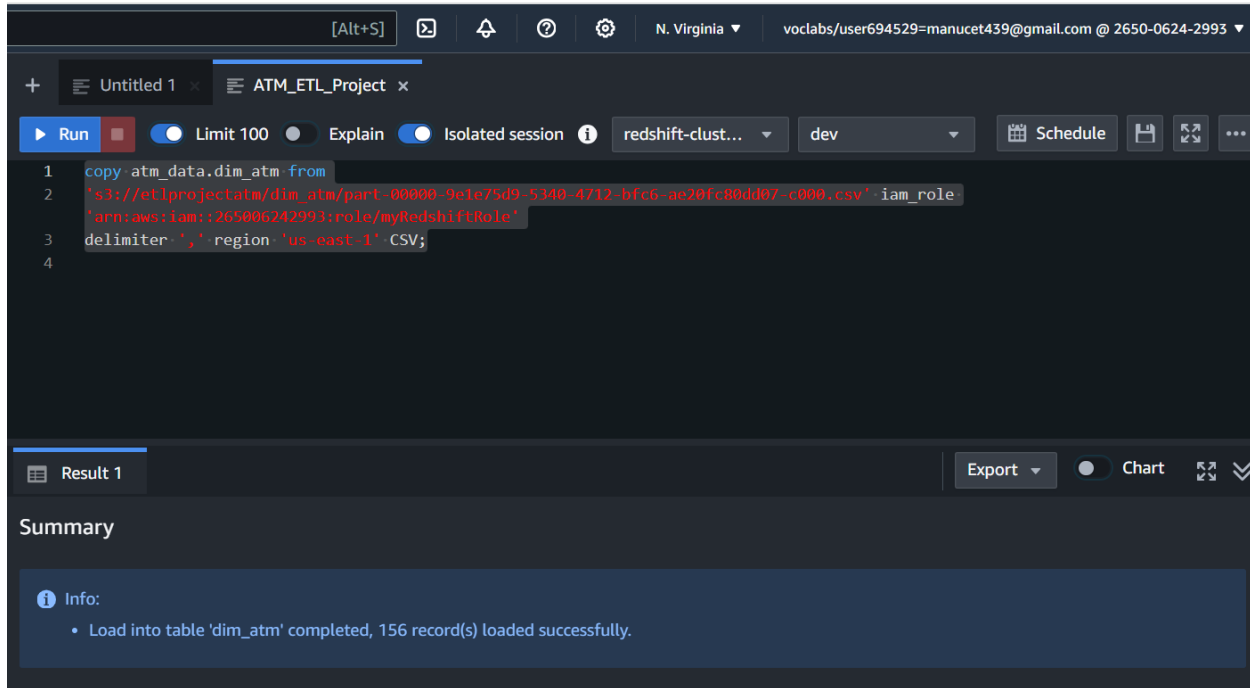
Below the query editor, the "Result 1" tab is active, showing a "Summary" section. It contains an "Info" message:

- Load into table 'dim_location' completed, 109 record(s) loaded successfully.

At the bottom, it shows "Returned rows: 0", "Elapsed time: 1.1s", and "Result set query:".

Query to copy the data from S3 bucket – etlprojectatm/dim_atm folder to dim_atm table

```
copy atm_data.dim_atm from
's3://etlprojectatm/dim_atm/part-00000-9e1e75d9-5340-4712-bfc6-ae20fc80dd07-c000.csv'
iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'
delimiter ',' region 'us-east-1' CSV;
```



The screenshot shows a Redshift console interface with a SQL query editor. The query is as follows:

```
1 copy atm_data.dim_atm from
2 's3://etlprojectatm/dim_atm/part-00000-9e1e75d9-5348-4712-bfc6-a020fc80dd07-c000.csv' iam_role
3 'arn:aws:iam::265006242993:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1' CSV;
```

Below the query editor, the 'Result 1' tab is selected, showing a 'Summary' section with an 'Info' message:

- Load into table 'dim_atm' completed, 156 record(s) loaded successfully.

Query to copy the data from S3 bucket – etlprojectatm/dim_date to dim_date table

copy atm_data.dim_date from

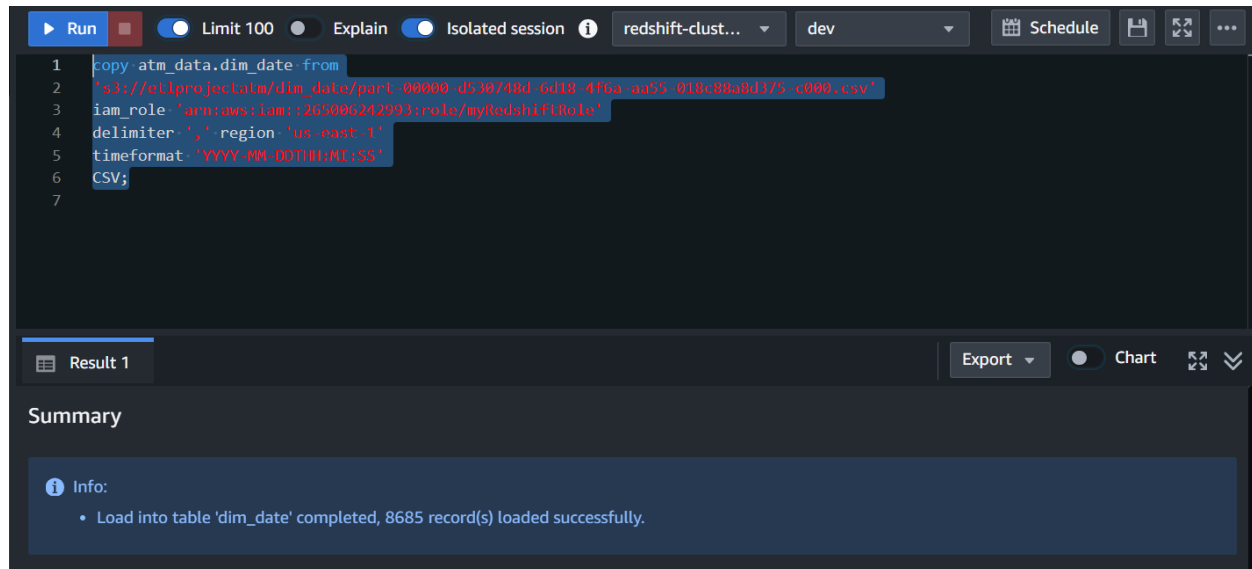
's3://etlprojectatm/dim_date/part-00000-d530748d-6d18-4f6a-aa55-018c88a8d375-c000.csv'

iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'

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

timeformat 'YYYY-MM-DDTHH:MI:SS'

CSV;



```

1 copy atm_data.dim_date from
2 's3://etlprojectatm/dim_date/part-00000-d530748d-6d18-4f6a-aa55-018c88a8d375-c000.csv'
3 iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1'
5 timeformat 'YYYY-MM-DDTHH:MI:SS'
6 CSV;
7

```

Result 1 Export Chart

Summary

Info:

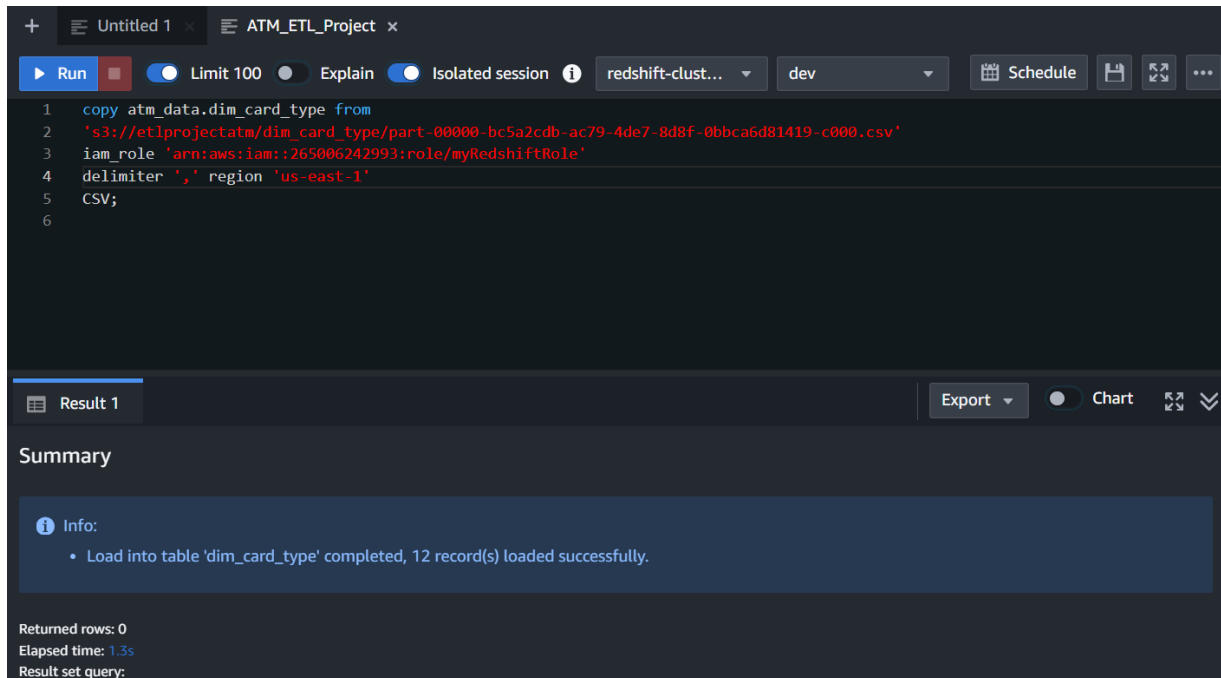
- Load into table 'dim_date' completed, 8685 record(s) loaded successfully.

Query to copy the data from S3 bucket – etlprojectatm/ dim_card_type to dim_card_type table

```

copy atm_data.dim_card_type from
's3://etlprojectatm/dim_card_type/part-00000-bc5a2cdb-ac79-4de7-8d8f-0bbca6d81419-
c000.csv'
iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;

```



```

1 copy atm_data.dim_card_type from
2 's3://etlprojectatm/dim_card_type/part-00000-bc5a2cdb-ac79-4de7-8d8f-0bbca6d81419-c000.csv'
3 iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1'
5 CSV;
6

```

Result 1 | Export | Chart

Summary

Info:

- Load into table 'dim_card_type' completed, 12 record(s) loaded successfully.

Returned rows: 0
Elapsed time: 1.3s
Result set query:

Query to copy the data from S3 bucket – etlprojectatm/ fact_atm_trans to fact_atm_trans table

```

copy atm_data.fact_atm_trans from
's3://etlprojectatm/fact_atm_trans/part-00000-71e4b685-05e2-456e-8e23-a5a68a1524ee-
c000.csv'
iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;

```

+

Untitled 1 x

ATM_ETL_Project x

▶ Run

Limit 100

Explain

Isolated session ⓘ

redshift-clust...

dev

Schedule

...

```
1 copy atm_data.fact_atm_trans from
2 's3://etlprojectatm/fact_atm_trans/part-00000-71e4b685-05e2-456e-8e23-a5a68a1524ee-c000.csv'
3 iam_role 'arn:aws:iam::265006242993:role/myRedshiftRole'
4 delimiter ',' region 'us-east-1'
5 CSV;
6
```

Result 1

Export

Chart

Summary

Info:

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