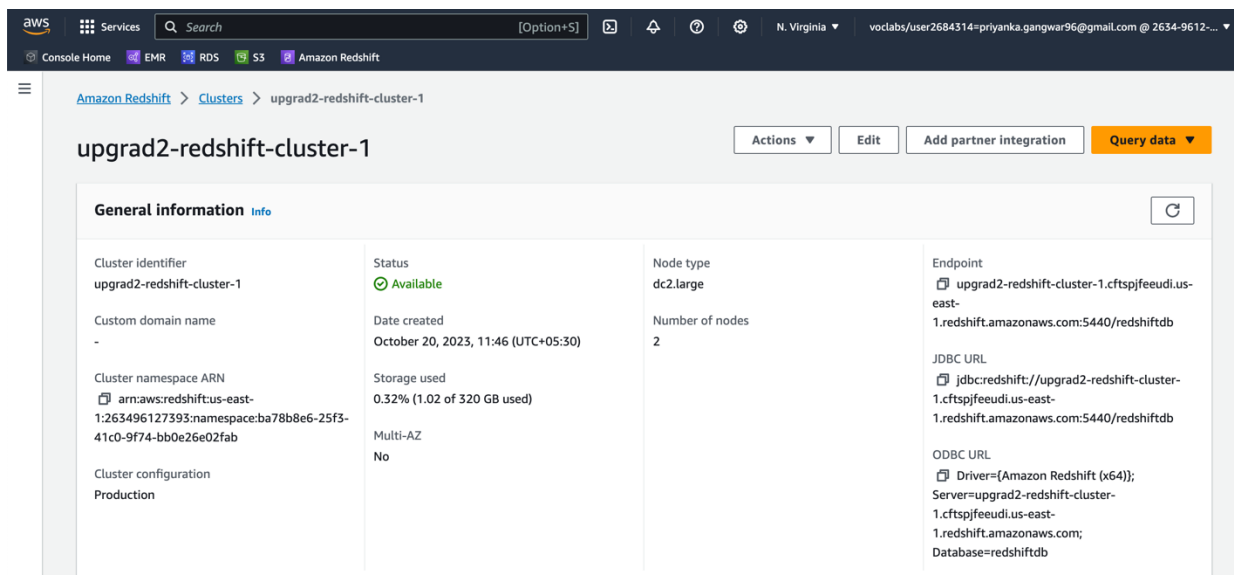


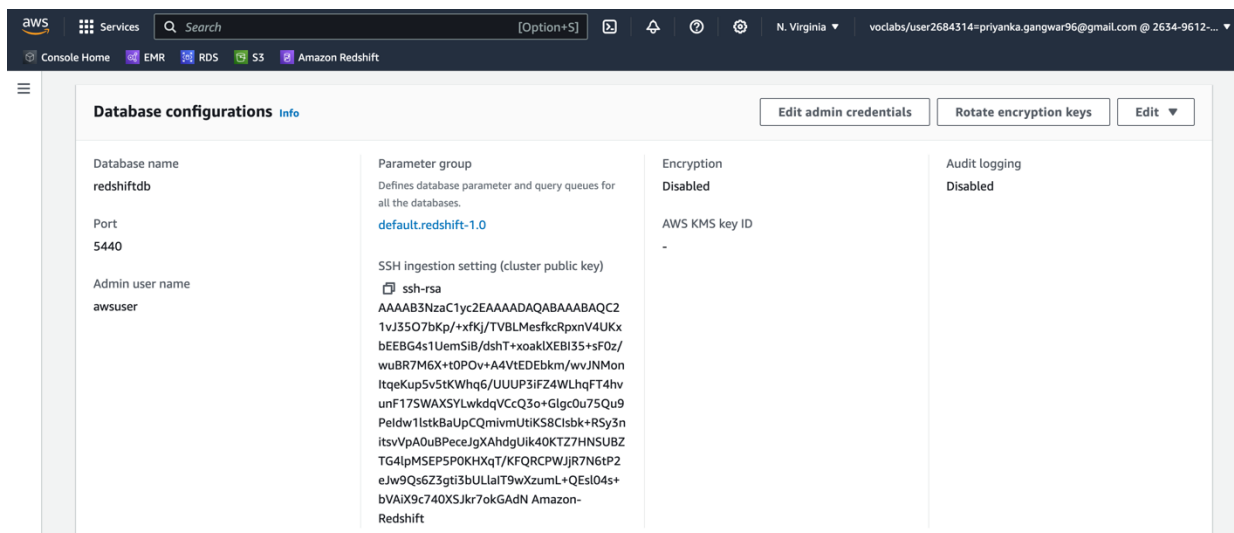
## Creation of a Redshift Cluster

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



The screenshot shows the AWS Management Console for the Redshift service. The cluster name is 'upgrad2-redshift-cluster-1'. The status is 'Available'. The node type is 'dc2.large'. The endpoint is 'upgrad2-redshift-cluster-1.cftspjfeudi.us-east-1.redshift.amazonaws.com:5440/redshiftdb'. The JDBC URL is 'jdbc:redshift://upgrad2-redshift-cluster-1.cftspjfeudi.us-east-1.redshift.amazonaws.com:5440/redshiftdb'. The ODBC URL is 'Driver={Amazon Redshift (x64)}; Server=upgrad2-redshift-cluster-1.cftspjfeudi.us-east-1.redshift.amazonaws.com; Database=redshiftdb'.

General information			
Cluster identifier	Status	Node type	Endpoint
upgrad2-redshift-cluster-1	Available	dc2.large	upgrad2-redshift-cluster-1.cftspjfeudi.us-east-1.redshift.amazonaws.com:5440/redshiftdb
Custom domain name	Date created	Number of nodes	JDBC URL
-	October 20, 2023, 11:46 (UTC+05:30)	2	jdbc:redshift://upgrad2-redshift-cluster-1.cftspjfeudi.us-east-1.redshift.amazonaws.com:5440/redshiftdb
Cluster namespace ARN	Storage used		ODBC URL
arn:aws:redshift:us-east-1:263496127393:namespace:ba78b8e6-25f3-41c0-9f74-bb0e26e02fab	0.32% (1.02 of 320 GB used)		Driver={Amazon Redshift (x64)}; Server=upgrad2-redshift-cluster-1.cftspjfeudi.us-east-1.redshift.amazonaws.com; Database=redshiftdb
Cluster configuration	Multi-AZ		
Production	No		



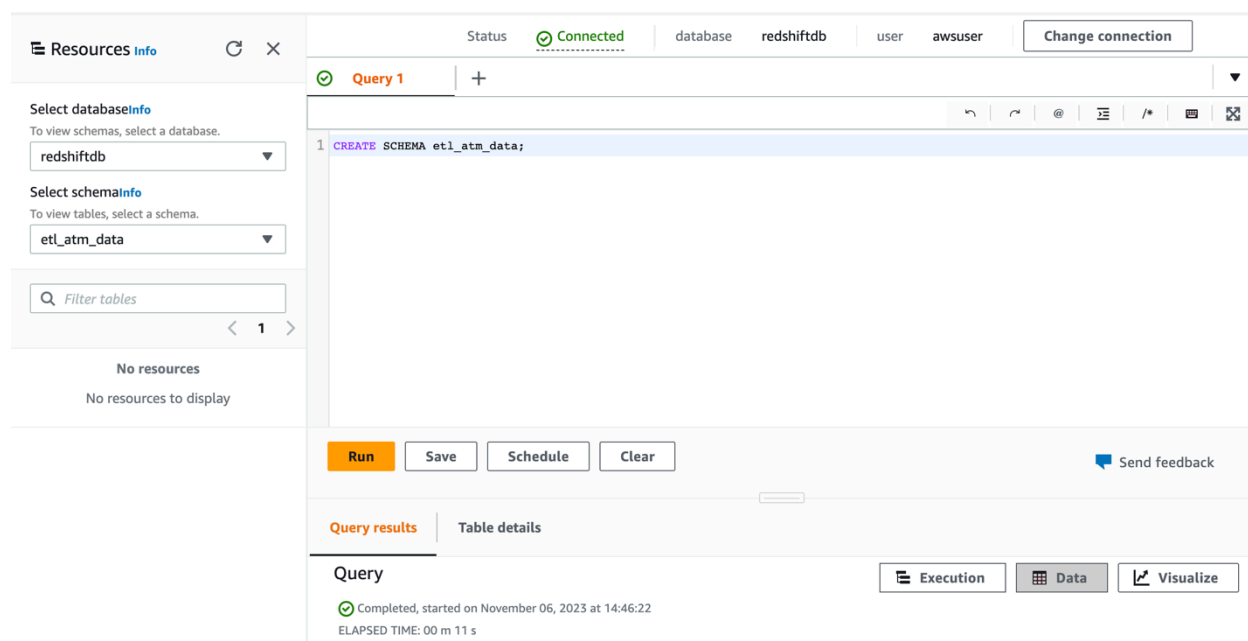
The screenshot shows the AWS Management Console for the Redshift service, specifically the 'Database configurations' tab. The database name is 'redshiftdb'. The port is '5440'. The admin user name is 'awsuser'. The parameter group is 'default.redshift-1.0'. The encryption is 'Disabled'. The audit logging is 'Disabled'.

Database configurations			
Database name	Parameter group	Encryption	Audit logging
redshiftdb	Defines database parameter and query queues for all the databases.	Disabled	Disabled
Port	SSH ingestion setting (cluster public key)	AWS KMS key ID	
5440	ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQAC2 1vJ35O7bKp/+xrkj/TVBLMesfkcRpxnV4UKx bEEBG4s1UemSIB/dshT+xoakIXEBI35+sF0z/ wuBR7M6X+t0POv+A4VtEDEbkm/vwJNMon ltqeKup5v5tKWq6/UUUP3IFZ4WLhqFT4hv unF175WAXSYLwkdqVCcQ3o+Glgc0u75Qu9 Peldw1lstkBaUpCQmivmUtIKS8CIsbk+RSy3n itsvVpA0uBPecelgXAhdgUik40KTZ7HNSUBZ TG4lpMSEP5POKHqT/KFQRCpWJJR7N6tP2 eJw9Qs6Z3gti3bUllalT9wXzumL+QEsI04s+ bVAIX9c740XSJkr7okGAdN Amazon- Redshift	-	
Admin user name			
awsuser			

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

- Query to create schema in Redshift for the atm data set

<Query> CREATE SCHEMA etl\_atm\_data;



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with 'Resources' and 'Info' tabs. Below 'Resources', there's a 'Select database' dropdown set to 'redshiftdb' and a 'Select schema' dropdown set to 'etl\_atm\_data'. The main area shows a query editor with the text '1 CREATE SCHEMA etl\_atm\_data;'. Below the editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. Below the buttons, there's a 'Query results' tab and a 'Table details' tab. The 'Query results' tab is active, showing a 'Query' section with a green checkmark indicating 'Completed, started on November 06, 2023 at 14:46:22' and 'ELAPSED TIME: 00 m 11 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

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

<Query 1>

```
CREATE TABLE etl_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)
);
```

Resources Info

Select databaseInfo  
To view schemas, select a database.  
redshiftdb

Select schemaInfo  
To view tables, select a schema.  
etl\_atm\_data

Filter tables

dim\_location\_pkey

dim\_location

Status Connected

database redshiftdb

user awsuser

Change connection

Query 1

```

1 CREATE TABLE etl_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 );

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query

Execution

Data

Visualize

Completed, started on November 06, 2023 at 15:00:03

ELAPSED TIME: 00 m 09 s

## <Query 2>

```

CREATE TABLE etl_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 etl_atm_data.dim_location(location_id)
);

```

Resources info

Select database info  
To view schemas, select a database.  
redshiftdb

Select schema info  
To view tables, select a schema.  
etl\_atm\_data

Filter tables

dim\_atm\_pkey

dim\_location\_pkey

dim\_atm

dim\_location

Status Connected
database redshiftdb
user awsuser
Change connection

Query 1

```

1 CREATE TABLE etl_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 etl_atm_data.dim_location(location_id)
9 );

```

Run Save Schedule Clear

Send feedback

Query results
Table details

Query

Completed, started on November 06, 2023 at 15:08:16  
ELAPSED TIME: 00 m 06 s

Execution Data Visualize

### <Query 3>

```

CREATE TABLE etl_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)
);

```

Resources
Info

Select database
Info
To view schemas, select a database.
redshiftdb

Select schema
Info
To view tables, select a schema.
etl\_atm\_data

Filter tables

dim\_atm\_pkey
dim\_location\_pkey
dim\_atm
dim\_location

Status
Connected
database
redshiftdb
user
awsuser
Change connection

Query 1
+

```

1 CREATE TABLE etl_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 etl_atm_data.dim_location(location_id)
9 );

```

Run
Save
Schedule
Clear
Send feedback

Query results
Table details

Query
Execution
Data
Visualize

Completed, started on November 06, 2023 at 15:08:16
ELAPSED TIME: 00 m 06 s

#### <Query 4>

```

CREATE TABLE etl_atm_data.dim_card_type
(
    card_type_id INT NOT NULL DISTKEY SORTKEY,
    card_type VARCHAR(30),
    PRIMARY KEY(card_type_id)
);

```

Resources
Info

Select database
Info
To view schemas, select a database.
redshiftdb

Select schema
Info
To view tables, select a schema.
etl\_atm\_data

Filter tables

dim\_atm\_pkey
dim\_card\_type\_pkey
dim\_date\_pkey
dim\_location\_pkey
dim\_atm
dim\_card\_type
dim\_date
dim\_location

Status
Connected
database
redshiftdb
user
awsuser
Change connection

Query 1
Query 2
+

```

1 CREATE TABLE etl_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 );

```

Run
Save
Schedule
Clear
Send feedback

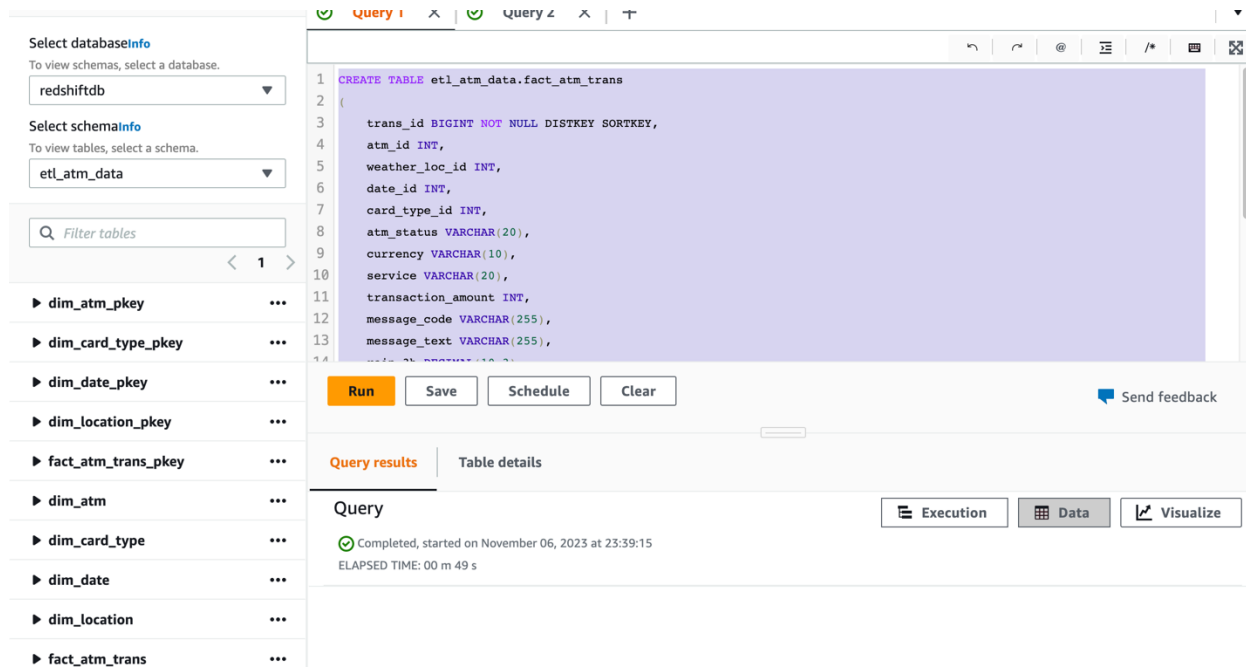
Query results
Table details

Query
Execution
Data
Visualize

Completed, started on November 06, 2023 at 23:31:16
ELAPSED TIME: 00 m 22 s

<Query 5>

```
CREATE TABLE etl_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
etl_atm_data.DIM_LOCATION(location_id),
    FOREIGN KEY(atm_id) REFERENCES etl_atm_data.DIM_ATM(atm_id),
    FOREIGN KEY(date_id) REFERENCES etl_atm_data.DIM_DATE(date_id),
    FOREIGN KEY(card_type_id) REFERENCES
etl_atm_data.DIM_CARD_TYPE(card_type_id)
);
```



The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with 'Select database' (redshiftdb) and 'Select schema' (etl\_atm\_data). Below this is a list of tables: 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 area displays a SQL query in a text editor:

```
1 CREATE TABLE etl_atm_data.fact_atm_trans
2 (
3     trans_id BIGINT NOT NULL DISTKEY SORTKEY,
4     atm_id INT,
5     weather_loc_id INT,
6     date_id INT,
7     card_type_id INT,
8     atm_status VARCHAR(20),
9     currency VARCHAR(10),
10    service VARCHAR(20),
11    transaction_amount INT,
12    message_code VARCHAR(255),
13    message_text VARCHAR(255),
14    ...
15 )
```

Below the query editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. To the right of these buttons is a 'Send feedback' link. Below the query editor, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows a status of 'Completed, started on November 06, 2023 at 23:39:15' and 'ELAPSED TIME: 00 m 49 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

## 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

### <Query 1>

```
COPY etl_atm_data.dim_location FROM
's3://atmdataseforetl/dim_location/part-00000-fb88c980-29d3-4103-a2d7-f6f95fdeed26-
c000.csv'
IAM_ROLE 'arn:aws:iam::263496127393:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;
```

Resources
Info

Select database
Info
To view schemas, select a database.
redshiftdb

Select schema
Info
To view tables, select a schema.
etl\_atm\_data

Filter tables

1

dim\_atm\_pkey
dim\_card\_type\_pkey
dim\_date\_pkey
dim\_location\_pkey
fact\_atm\_trans\_pkey
dim\_atm
dim\_card\_type
dim\_date

Query 1
Query 2

```

1 copy etl_atm_data.dim_location from
2 's3://atmdatasetforetl/dim_location/part-00000-fb88c980-29d3-4103-a2d7-f6f95fdeed26-c000.csv'
3 iam_role 'arn:aws:iam::263496127393:role/myRedshiftRole'
4 delimiter ','
5 region 'us-east-1'
6 csv;
7
8

```

Run
Save
Schedule
Clear
Send feedback

Query results
Table details

Query 807985
Execution
Data
Visualize

Completed, started on November 07, 2023 at 13:28:04
ELAPSED TIME: 00 m 20 s

## <Query 2>

```

COPY etl_atm_data.dim_atm FROM
's3://atmdatasetforetl/dim_atm/part-00000-7518197d-f887-4fd6-8869-2f52abac1728-
c000.csv'
IAM_ROLE 'arn:aws:iam::263496127393:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;

```



Resources info

Select database info

To view schemas, select a database.

redshiftdb

Select schema info

To view tables, select a schema.

etl\_atm\_data

Filter tables

1

dim\_atm\_pkey

dim\_card\_type\_pkey

dim\_date\_pkey

dim\_location\_pkey

fact\_atm\_trans\_pkey

dim\_atm

dim\_card\_type

dim\_date

Status Connected

database redshiftdb

user awsuser

Change connection

Query 1

Query 2

+

```

1 copy etl_atm_data.dim_atm from
2 's3://atmdatasetforetl/dim_atm/part-00000-7518197d-f887-4fd6-8869-2f52abac1728-c000.csv'
3 iam_role 'arn:aws:iam::263496127393:role/myRedshiftRole'
4 delimiter ','
5 region 'us-east-1'
6 csv;
7
8
9

```

Run Save Schedule Clear

Send feedback

Query results

Table details

Query 808038

Execution

Data

Visualize

Completed, started on November 07, 2023 at 13:31:49

ELAPSED TIME: 00 m 13 s

### <Query 3>

```

COPY etl_atm_data.dim_date FROM
's3://atmdatasetforetl/dim_date/part-00000-9f639a89-2041-4ba7-822b-2f87aefb54c3-
c000.csv'
IAM_ROLE 'arn:aws:iam::263496127393:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;

```

Resources info

Select database info

To view schemas, select a database.

redshiftdb

Select schema info

To view tables, select a schema.

etl\_atm\_data

Filter tables

1

dim\_atm\_pkey

dim\_card\_type\_pkey

dim\_date\_pkey

dim\_location\_pkey

fact\_atm\_trans\_pkey

dim\_atm

dim\_card\_type

dim\_date

Status Connected

database redshiftdb

user awsuser

Change connection

Query 1

Query 2

Query 3

+

```

1 COPY etl_atm_data.dim_date FROM 's3://atmdatasetforetl/dim_date/part-00000-4935e392-0ef7-44f7-add7-d967b0a51f18-c000.csv'
2 IAM_ROLE 'arn:aws:iam::263496127393:role/myRedshiftRole'
3 DELIMITER ','
4 REGION 'us-east-1'
5 CSV
6 TIMEFORMAT 'auto';
7
8

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 810182

Execution

Data

Visualize

Completed, started on November 07, 2023 at 16:15:29

ELAPSED TIME: 00 m 25 s

## <Query 4>

```

COPY etl_atm_data.dim_card_type FROM
's3://atmdatasetforetl/dim_card_type/part-00000-abef2a30-ae47-4913-a737-
8109ec1af6a1-c000.csv'
IAM_ROLE 'arn:aws:iam::263496127393:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;

```

Resources Info

Select database info

To view schemas, select a database.

redshiftdb

Select schema info

To view tables, select a schema.

etl\_atm\_data

Filter tables

1

dim\_atm\_pkey

dim\_card\_type\_pkey

dim\_date\_pkey

dim\_location\_pkey

fact\_atm\_trans\_pkey

dim\_atm

dim\_card\_type

dim\_date

Status Connected

database redshiftdb

user awsuser

Change connection

Query 1

Query 2

+

```

1 COPY etl_atm_data.dim_card_type FROM
2 's3://atmdatasetforetl/dim_card_type/part-00000-abef2a30-ae47-4913-a737-8109ec1af6a1-c000.csv'
3 IAM_ROLE 'arn:aws:iam::263496127393:role/myRedshiftRole'
4 DELIMITER ','
5 REGION 'us-east-1'
6 CSV;
7
8
9

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 808106

Execution

Data

Visualize

Completed, started on November 07, 2023 at 13:37:16

ELAPSED TIME: 00 m 41 s

## <Query 5>

```

COPY etl_atm_data.fact_atm_trans FROM
's3://atmdatasetforetl/fact_atm_trans/part-00000-60b4e63b-70ac-489d-b7c7-
5a605bab12db-c000.csv'
IAM_ROLE 'arn:aws:iam::263496127393:role/myRedshiftRole'
DELIMITER ','
REGION 'us-east-1'
CSV;

```

Resources Info

Select database Info

To view schemas, select a database.

redshiftdb

Select schema Info

To view tables, select a schema.

etl\_atm\_data

Filter tables

< 1 >

dim\_atm\_pkey

dim\_card\_type\_pkey

dim\_date\_pkey

dim\_location\_pkey

fact\_atm\_trans\_pkey

dim\_atm

dim\_card\_type

dim\_date

Status Connected

database redshiftdb

user awsuser

Change connection

Query 1

Query 2

+

1 COPY etl\_atm\_data.fact\_atm\_trans FROM

2 's3://atmdatasetforetl/fact\_atm\_trans/part-00000-60b4e63b-70ac-489d-b7c7-5a605bab12db-c000.csv'

3 IAM\_ROLE 'arn:aws:iam::263496127393:role/myRedshiftRole'

4 DELIMITER ','

5 REGION 'us-east-1'

6 CSV;

7

8

9

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 808146

Execution

Data

Visualize

Completed, started on November 07, 2023 at 13:39:18

ELAPSED TIME: 00 m 40 s