

Creation of a Redshift Cluster

Screenshots of the configuration of the Redshift cluster that I have created:
redshift-etl-spar-nord-cluster

[Amazon Redshift](#) > [Clusters](#) > redshift-etl-spar-nord-cluster

redshift-etl-spar-nord-cluster

Actions ▾

Edit

Add partner integration

Query data ▾

General information [Info](#)

Cluster identifier

redshift-etl-spar-nord-cluster

Status

Available

Node type

dc2.large

Endpoint

[redshift-etl-spar-nord-cluster.cplhscgjejm.us-east-1.redshift.amazonaws.com:S439/spar_nord_bank_db](#)

Custom domain name

-

Date created

November 09, 2024, 19:38 (UTC+05:30)

Number of nodes

2

JDBC URL

[jdbc:redshift://redshift-etl-spar-nord-cluster.cplhscgjejm.us-east-1.redshift.amazonaws.com:S439/spar_nord_bank_db](#)

Cluster namespace ARN

[arn:aws:redshift:us-east-1:878643080719:namespace:2b73fad3-c941-4979-8395-49fb6d9c7d72](#)

Multi-AZ

No

Patch version

[Patch 185](#)

ODBC URL

[Driver={Amazon Redshift \(x64\)}; Server=redshift-etl-spar-nord-cluster.cplhscgjejm.us-east-1.redshift.amazonaws.com; Database=spar_nord_bank_db](#)

Cluster configuration

Production

Storage used

-

Cluster performance

Query monitoring

Databases

Datashares

Integrations

Resource Policy

Schedules

Maintenance


Properties


Recommendations (0)

To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations.

Screenshots of configurations while creating cluster

Create cluster [Info](#)

 Looking for free trial? Try Redshift Serverless. First-time Redshift Serverless customers receive a \$300 credit to use in their account.

[Launch Redshift Serverless](#) 

Cluster configuration

Cluster identifier
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).


Choose the size of the cluster
☒ I'll choose
☐ Help me choose

Node type [Info](#)
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.

Number of nodes
Enter the number of nodes that you need.

Range (1-32)

Configuration summary [Info](#)
dc2.large | 2 nodes

\$365.00/month
Estimated on-demand compute price
Save more than 60% of your costs by purchasing reserved nodes. [Learn more about pricing](#) 

320 GB
Total compressed storage
The total storage capacity for the cluster if you deploy the number of nodes that you chose.

Screenshot of database configurations:

Database configurations

Admin user name
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](#).

Admin password
Select an option to manage your admin password.

☐ **Manage admin credentials in AWS Secrets Manager** [Info](#)
AWS manages a KMS key that encrypts your data.

☐ **Generate a password**
Amazon Redshift generates an admin password.

☒ **Manually add the admin password**
Manually enter the admin password.

Admin user password

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 "@"`.

☐ **Show password**

Database encryption
Database encryption helps protect data at rest. Data blocks and system metadata are encrypted for the cluster and its snapshots. [Learn more about service integration](#)

☐ **Enable cluster encryption**
Encrypt your cluster's data, using keys managed by the AWS Key Management Service.

☒ **Disable cluster encryption**
Cluster data won't be encrypted. This isn't reversible after you create the cluster.

Cluster permissions

Associating IAM role (myRedshiftRole) to my cluster and creating spar_nord_bank_db:

Cluster that has the AmazonRedshiftFullAccess IAM policy attached. This policy includes

Associate IAM roles

IAM roles (1/3)
Choose from existing IAM roles. You can associate up to 50 IAM roles with this cluster.

Find iam roles

< 1 > ⚙

	IAM roles
<input type="checkbox"/>	AWSServiceRoleForRedshift
<input type="checkbox"/>	LabRole
<input checked="" type="checkbox"/>	myRedshiftRole

Cancel Associate IAM roles

▼ Database configurations [Info](#)

Database name
Specify a database name to create an additional database.

spar_nord_bank_db

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 ▼

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

Making Sure the files are present in Amazon S3 bucket: sparnordbanketl

Amazon S3 > Buckets > sparnordbanketl

sparnordbanketl [Info](#)

Objects | Properties | Permissions | Metrics | Management | Access Points

Objects (5) [Info](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<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	-	-	-

Query to create Schema for dimension and fact tables:

Create schema spar_nord_atm_data;

Editor | Query history | Saved queries | Scheduled queries

Status: Connected database: spar_nord_bank_db user: awsuser [Change connection](#)

Query 1

```
1 create schema spar_nord_atm_data;
```

[Run](#) [Save](#) [Schedule](#) [Clear](#) [Send feedback](#)

Query results | Table details

Query

Completed, started on November 09, 2024 at 19:46:19
ELAPSED TIME: 00 m 02 s

[Execution](#) [Data](#) [Visualize](#)

Query to create DIM_LOACTION dimension table:

```
create table spar_nord_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));
```

Editor

Query history

Saved queries

Scheduled queries

Resources Info

Select database Info

To view schemas, select a database.

spar_nord_bank_db

Select schema Info

To view tables, select a schema.

public

Filter tables

< 1 >

No resources

No resources to display

Status Connected

database spar_nord_bank_db

user awsuser

Change connection

Query 1

+

```

1 create table spar_nord_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

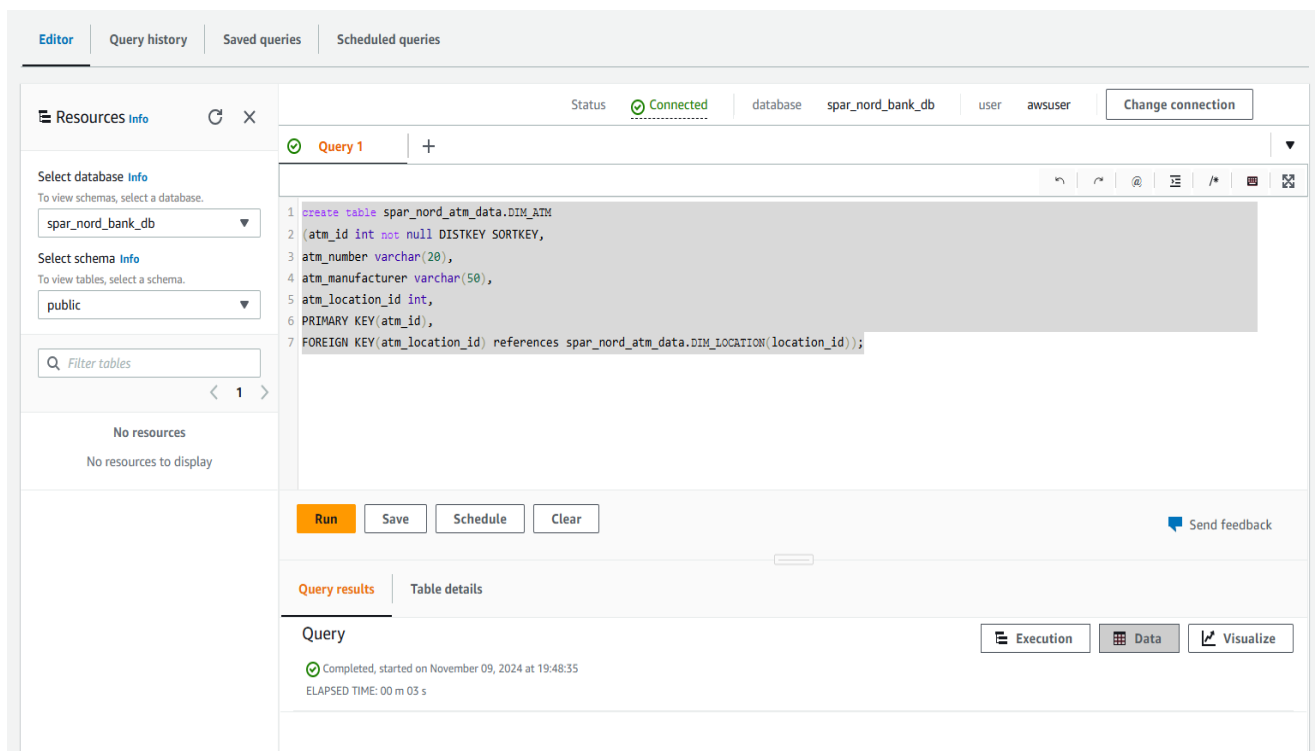
Visualize

Completed, started on November 09, 2024 at 19:47:14

ELAPSED TIME: 00 m 04 s

Query to create DIM_ATM dimension table:

```
create table spar_nord_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_atm_data.DIM_LOCATION(location_id));
```



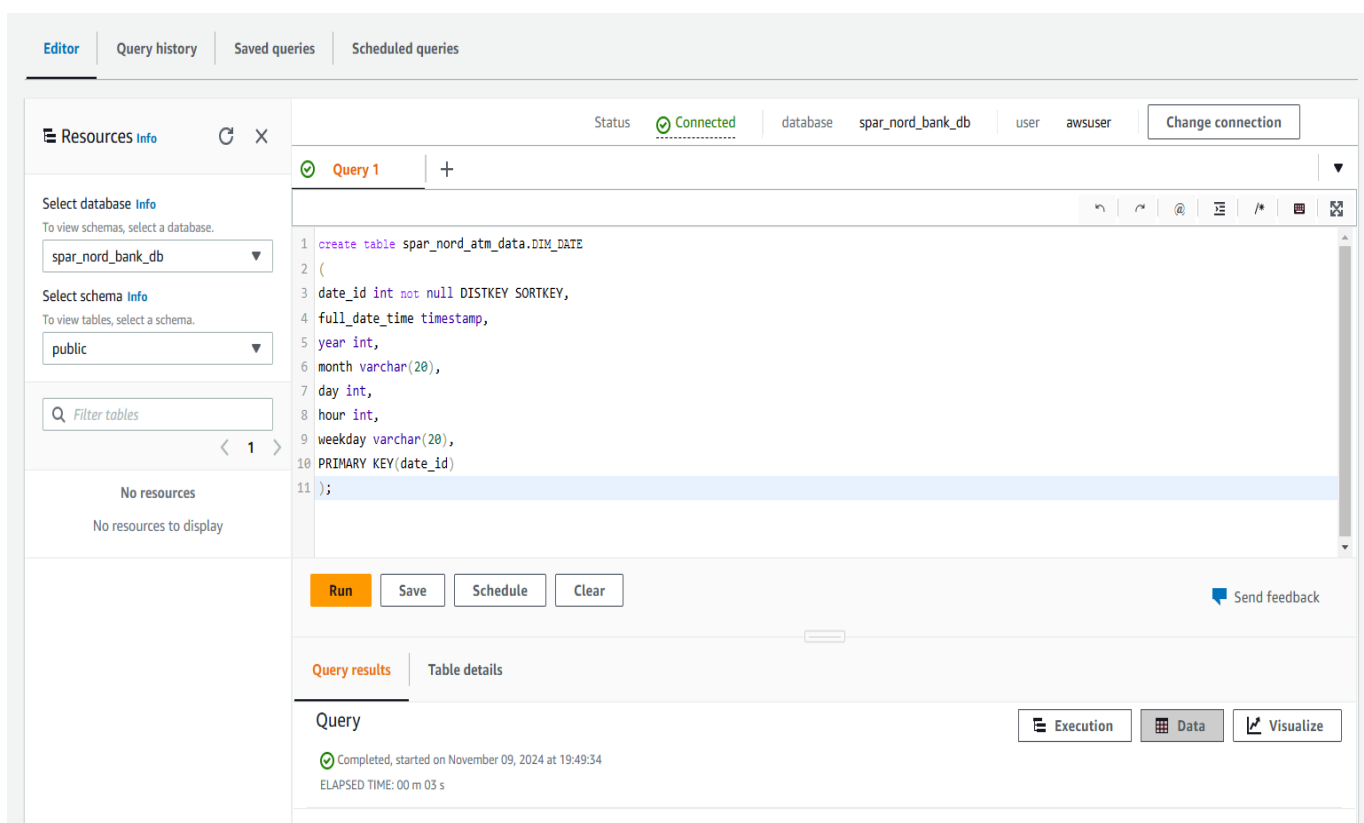
The screenshot shows a database query editor interface. The top navigation bar includes tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. The main interface is divided into several sections:

- Resources Info:** A sidebar on the left with a search bar and filters. It shows 'spar_nord_bank_db' selected as the database and 'public' as the schema. Below this, it states 'No resources' and 'No resources to display'.
- Query Editor:** The central area where the SQL query is entered. The query is:


```
1 create table spar_nord_atm_data.DIM_ATM
2 (atm_id int not null DISTKEY SORTKEY,
3 atm_number varchar(20),
4 atm_manufacturer varchar(50),
5 atm_location_id int,
6 PRIMARY KEY(atm_id),
7 FOREIGN KEY(atm_location_id) references spar_nord_atm_data.DIM_LOCATION(location_id));
```
- Execution Controls:** Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present.
- Query Results:** At the bottom, there is a section for 'Query results' and 'Table details'. The 'Query results' tab is active, showing a status of 'Completed, started on November 09, 2024 at 19:48:35' and 'ELAPSED TIME: 00 m 03 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

Query to create DIM_DATE dimension table:

```
create table spar_nord_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));
```



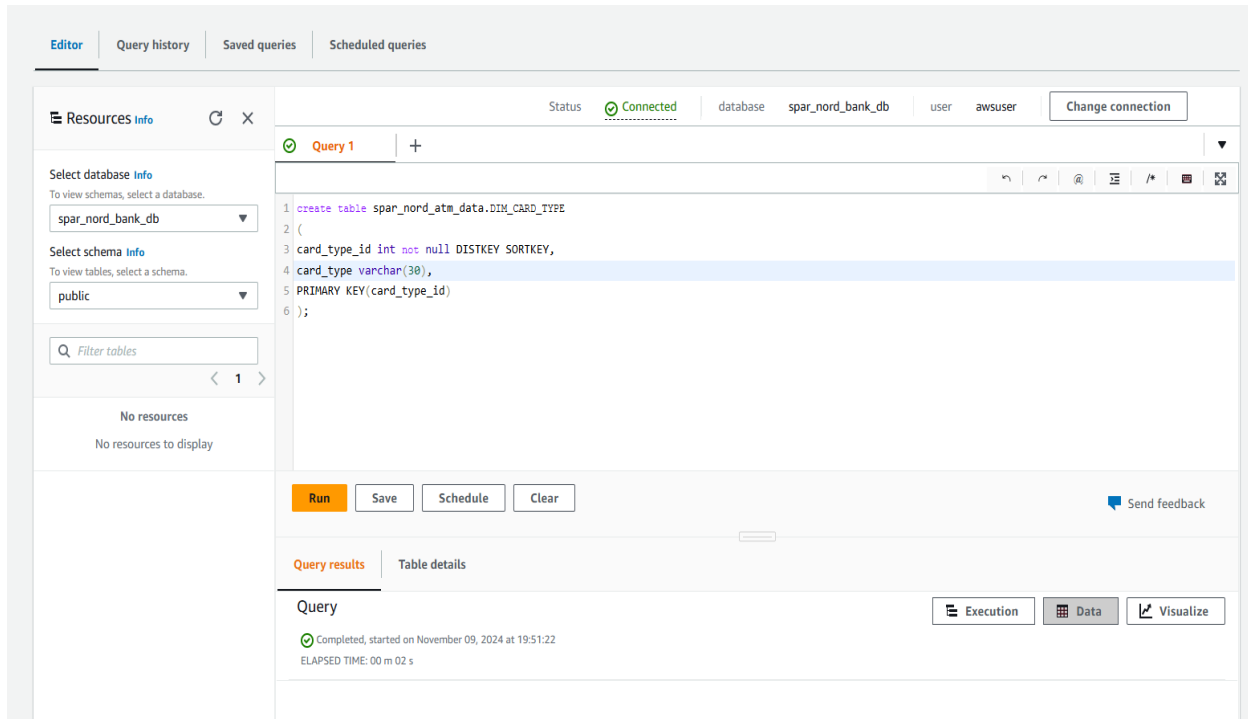
The screenshot shows a SQL query editor interface. At the top, there are tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. The 'Editor' tab is active. On the left side, there is a 'Resources' panel with a search bar and a list of databases and schemas. The 'spar_nord_bank_db' database is selected, and the 'public' schema is chosen. The main editor area displays the SQL query to create the 'spar_nord_atm_data.DIM_DATE' table. The query is as follows:

```
1 create table spar_nord_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 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. The bottom section shows the 'Query results' tab, which indicates that the query was completed successfully on November 09, 2024 at 19:49:34, with an elapsed time of 00 m 03 s. There are also buttons for 'Execution', 'Data', and 'Visualize'.

Query to create DIM_CARD_TYPE dimension table:

```
create table spar_nord_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 upGrad SQL Editor interface. On the left, the 'Resources' panel shows the selected database 'spar_nord_bank_db' and schema 'public'. The main editor area displays the SQL query for creating the 'DIM_CARD_TYPE' table. The query is as follows:

```
1 create table spar_nord_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 );
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. The 'Query results' tab is active, showing the query execution status: 'Completed, started on November 09, 2024 at 19:51:22' and 'ELAPSED TIME: 00 m 02 s'.

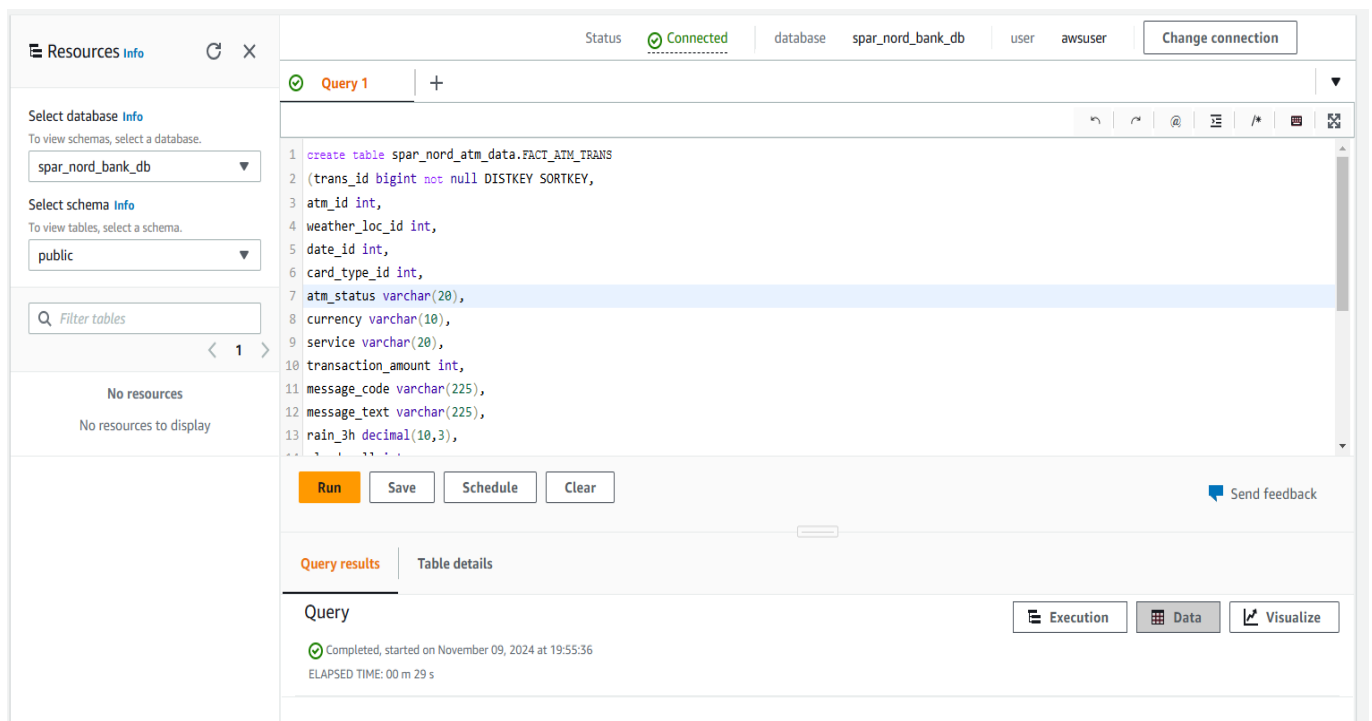
Query to create FACT_ATM_TRANS Fact table:

```
create table spar_nord_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_atm_data.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references spar_nord_atm_data.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references spar_nord_atm_data.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references
spar_nord_atm_data.DIM_CARD_TYPE(card_type_id));

```



The screenshot displays the upGrad SQL IDE interface. On the left, the 'Resources' panel shows the selected database 'spar_nord_bank_db' and schema 'public'. The main editor area contains a SQL query to create a table 'spar_nord_atm_data.FACT_ATM_TRANS' with various columns and constraints. The query is as follows:

```

1 create table spar_nord_atm_data.FACT_ATM_TRANS
2 (trans_id bigint not null DISTKEY SORTKEY,
3  atm_id int,
4  weather_loc_id int,
5  date_id int,
6  card_type_id int,
7  atm_status varchar(20),
8  currency varchar(10),
9  service varchar(20),
10 transaction_amount int,
11 message_code varchar(225),
12 message_text varchar(225),
13 rain_3h decimal(10,3),

```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. To the right of these buttons is a 'Send feedback' link. Below the query editor, the 'Query results' tab is active, showing a status message: 'Completed, started on November 09, 2024 at 19:55:36' and 'ELAPSED TIME: 00 m 29 s'. At the bottom right, there are 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

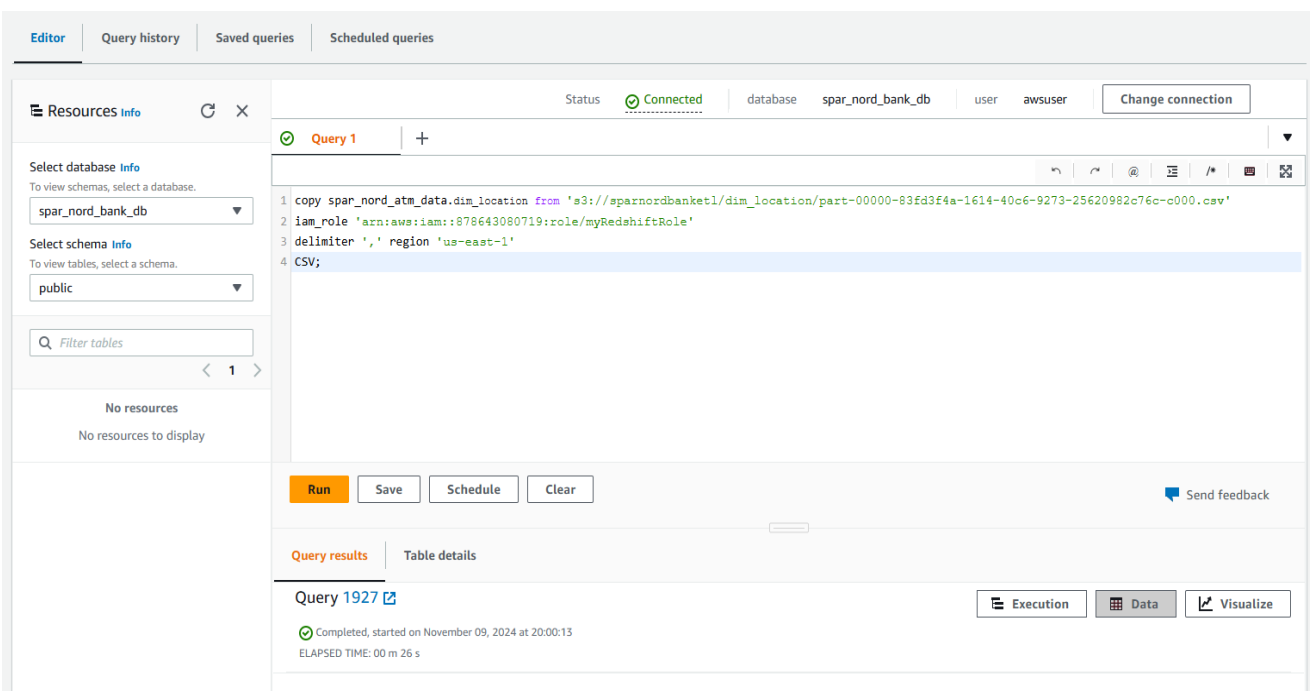
Copying data to dim_location table from S3 bucket:

copy spar_nord_atm_data.dim_location from 's3://sparnordbanketl/dim_location/part-00000-83fd3f4a-1614-40c6-9273-25620982c76c-c000.csv'

iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'

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

CSV;



The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with 'Resources Info' and 'Select database' (set to 'spar_nord_bank_db') and 'Select schema' (set to 'public'). The main area displays a SQL query in the 'Query 1' editor. The query is:
 1 copy spar_nord_atm_data.dim_location from 's3://sparnordbanketl/dim_location/part-00000-83fd3f4a-1614-40c6-9273-25620982c76c-c000.csv'
 2 iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'
 3 delimiter ',' region 'us-east-1'
 4 CSV;
 Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, the 'Query results' tab is active, showing 'Query 1927' with a status of 'Completed, started on November 09, 2024 at 20:00:13' and 'ELAPSED TIME: 00 m 26 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

Copying data to dim_atm table from S3 bucket:

copy spar_nord_atm_data.dim_atm from 's3://sparnordbanketl/dim_atm/part-00000-752abbfe-1f64-425d-81e9-0ae51f56c86e-c000.csv'

iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'

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

CSV;

Editor

Query history

Saved queries

Scheduled queries

Resources Info

Select database Info

To view schemas, select a database.

spar_nord_bank_db

Select schema Info

To view tables, select a schema.

public

Filter tables

< 1 >

No resources

No resources to display

Status Connected

database spar_nord_bank_db

user awsuser

Change connection

Query 1

+

1 copy spar_nord_atm_data.dim_atm from 's3://sparnordbanketl/dim_atm/part-00000-752abbfe-1f64-425d-81e9-0ae51f56c86e-c000.csv'

2 iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'

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

4 CSV;

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 1975

Execution

Data

Visualize

Completed, started on November 09, 2024 at 20:03:18

ELAPSED TIME: 00 m 21 s

Copying data to dim_date table from S3 bucket:

copy spar_nord_atm_data.dim_date from 's3://sparnordbanketl/dim_date/part-00000-7eb2ed5d-94e9-424f-9bc2-fccb623110e9-c000.csv'

iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'

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

CSV;

Resources Info

Select database Info

To view schemas, select a database.

spar_nord_bank_db

Select schema Info

To view tables, select a schema.

public

Filter tables

< 1 >

No resources

No resources to display

Status Connected

database spar_nord_bank_db

user awsuser

Change connection

Query 1

+

1 copy spar_nord_atm_data.dim_date from 's3://sparnordbanketl/dim_date/part-00000-7eb2ed5d-94e9-424f-9bc2-fccb623110e9-c000.csv'

2 iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'

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

4 CSV;

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 2501

Execution

Data

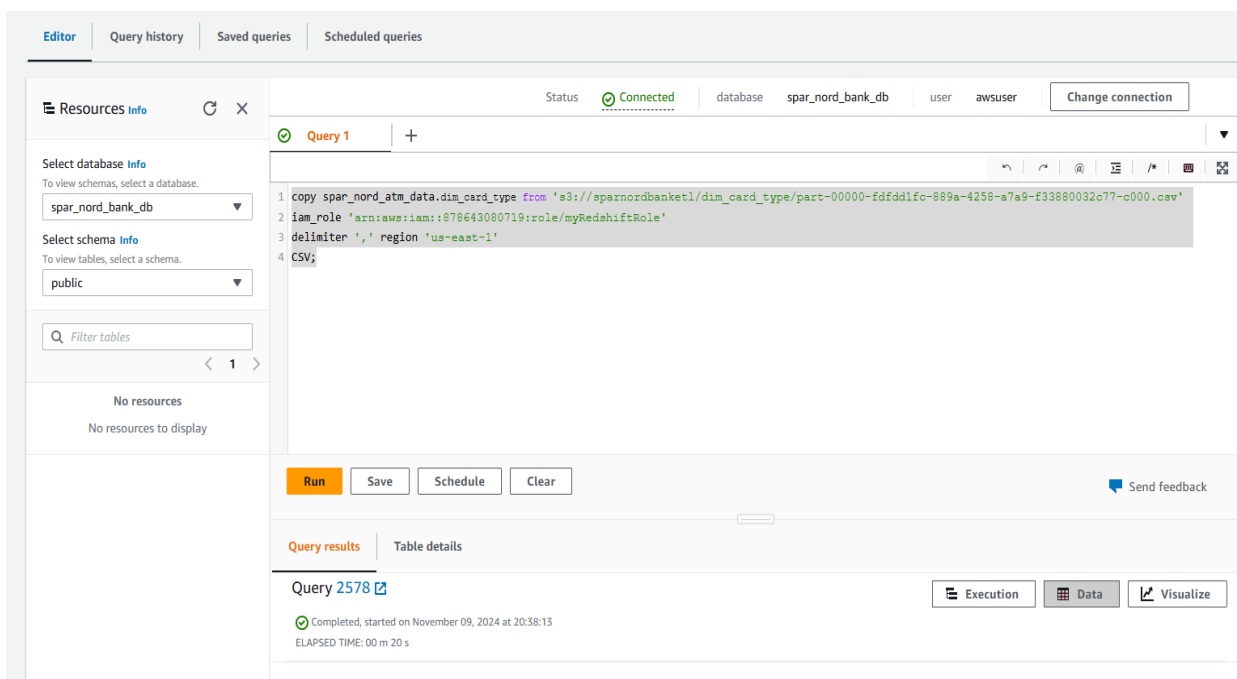
Visualize

Completed, started on November 09, 2024 at 20:33:54

ELAPSED TIME: 00 m 05 s

Copying data to dim_card_type table from S3 bucket:

```
copy spar_nord_atm_data.dim_card_type from 's3://sparnordbanketl/dim_card_type/part-00000-fdfdd1fc-889a-4258-a7a9-f33880032c77-c000.csv'
iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift Query Editor interface. The top navigation bar includes tabs for Editor, Query history, Saved queries, and Scheduled queries. The main workspace displays a SQL query (Query 1) that has been executed successfully. The query is as follows:

```
1 copy spar_nord_atm_data.dim_card_type from 's3://sparnordbanketl/dim_card_type/part-00000-fdfdd1fc-889a-4258-a7a9-f33880032c77-c000.csv'
2 iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

The interface shows the query was completed on November 09, 2024 at 20:38:13, with an elapsed time of 00 m 20 s. The query results are displayed in a table format, and the table details are visible.

Copying data to fact_atm_trans table from S3 bucket:

```
copy spar_nord_atm_data.fact_atm_trans from
's3://sparnordbanketl/fact_atm_trans/part-00000-f3a78bbe-4298-409f-bb9f-b6ed344ffb4f-
c000.csv'
iam_role 'arn:aws:iam::878643080719:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

Editor

Query history

Saved queries

Scheduled queries

Resources Info

Select database Info

Select schema Info

Filter tables

No resources

Status

Connected

database

spar_nord_bank_db

user

awsuser

Change connection

Query 1

+

1

2

3

4

copy spar_nord_atm_data.fact_atm_trans from 's3://spaxnordbanket1/fact_atm_trans/part-00000-f3a78bbe-4298-409f-bb9f-b6ed344ffb4f-c000.csv'

iam_role 'arn:aws:iam:878643080719:role/myRedshiftRole'

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

CSV;

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 2626

Execution

Data

Visualize

Completed, started on November 09, 2024 at 20:40:31

ELAPSED TIME: 00 m 47 s