

## 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

### 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).

What are you planning to use this cluster for?

☒ **Production**  
Configure for fast and consistent performance at the best price.

☐ **Free trial**  
Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

Choose the size of the cluster

**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)

### 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](#).

☐ **Auto generate password**  
Amazon Redshift can generate a password for you, or you can specify your own password.

**Admin user password**



☐ **Show 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 `@`.

## ▼ Network and security [Info](#)

### Virtual private cloud (VPC)

This VPC defines the virtual networking environment for this cluster.

redshift-cluster-vpc  
vpc-08e36789ff0cfe085

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

### VPC security groups

This VPC security group defines which subnets and IP ranges the cluster can use in the VPC.

Choose one or more security groups

default  
sg-02457508681555f43

### Cluster subnet group

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

spar-nord-redshift-cluster-subnet-group

### Availability Zone

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

No preference

### 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](#)

☒ Turn off

Amazon Redshift > Clusters > spar-nord-redshift-cluster

## spar-nord-redshift-cluster ▼

Actions ▼

Edit

Add partner integration

Query data ▼

### General information



Cluster identifier  
spar-nord-redshift-cluster

Status  
 Available

Node type  
dc2.large

Endpoint  
 spar-nord-redshift-cluster.cv1kjo31h...

Cluster namespace  
9e23e5a4-8688-40a5-8e66-7b279588235c

Date created  
September 02, 2022, 11:42 (UTC+05:30)

Number of nodes  
2

JDBC URL  
 jdbc:redshift://spar-nord-redshift-clu...

Storage used  
-

AQUA  
Not available

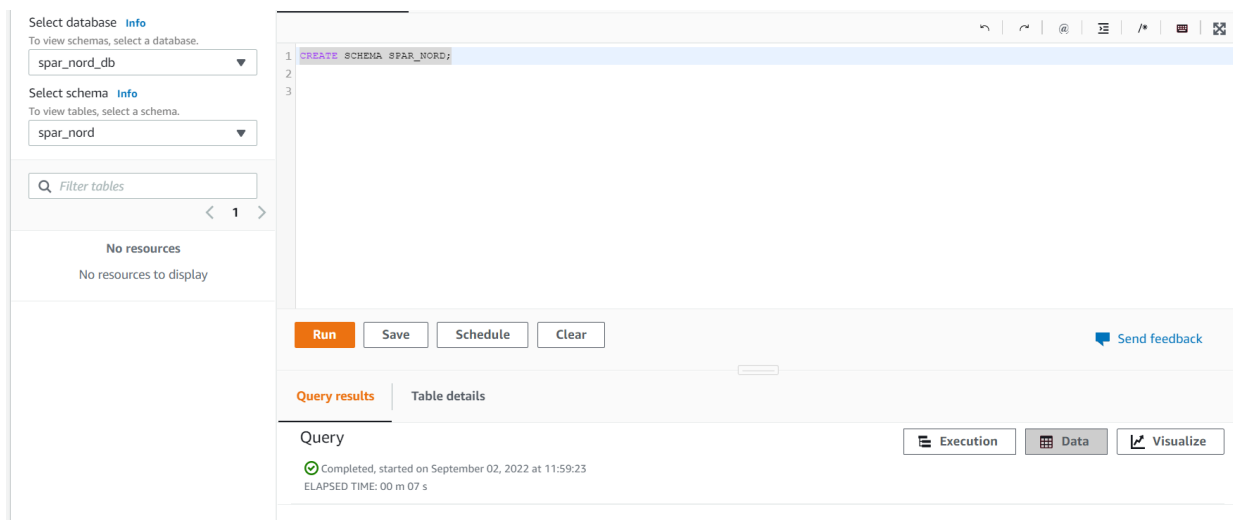
ODBC URL  
 Driver={Amazon Redshift (x64)}; Serv...

## 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 a schema for all the dimensions and fact tables to be created inside it.

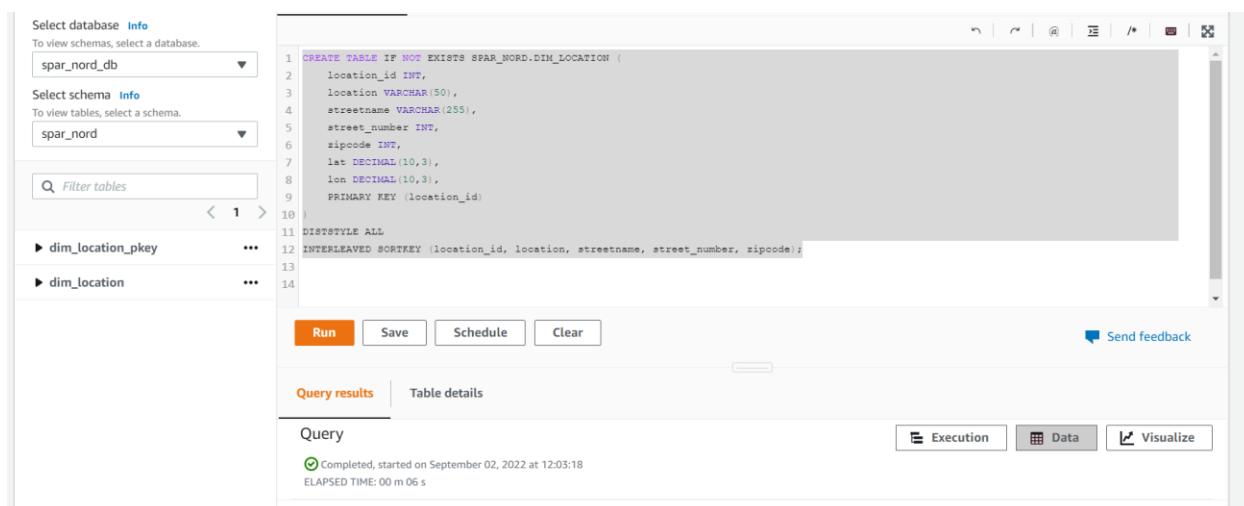
```
CREATE SCHEMA SPAR_NORD;
```



### Create all dimensions and facts tables

#### 1. Location Dimension:

```
CREATE TABLE IF NOT EXISTS SPAR_NORD.DIM_LOCATION (
    location_id INT,
    location VARCHAR(50),
    streetname VARCHAR(255),
    street_number INT,
    zipcode INT,
    lat DECIMAL(10,3),
    lon DECIMAL(10,3),
    PRIMARY KEY (location_id)
)
DISTSTYLE ALL
INTERLEAVED SORTKEY (location_id, location, streetname, street_number, zipcode);
```



Select database [Info](#)  
To view schemas, select a database.  
spar\_nord\_db

Select schema [Info](#)  
To view tables, select a schema.  
spar\_nord

Filter tables

dim\_location\_pkey  
dim\_location

```
1 CREATE TABLE IF NOT EXISTS SPAR_NORD.DIM_LOCATION (
2   location_id INT,
3   location VARCHAR(50),
4   streetname VARCHAR(255),
5   street_number INT,
6   zipcode INT,
7   lat DECIMAL(10,3),
8   lon DECIMAL(10,3),
9   PRIMARY KEY (location_id)
10 )
11 DISTSTYLE ALL
12 INTERLEAVED SORTKEY (location_id, location, streetname, street_number, zipcode);
13
14
```

Run Save Schedule Clear

Send feedback

Query results Table details

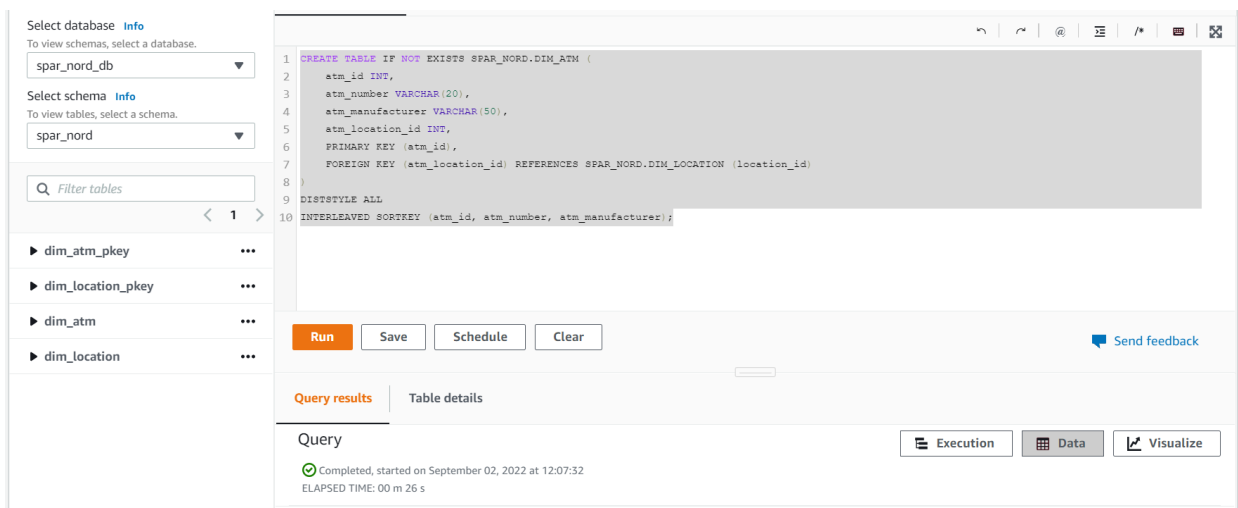
Query

Completed, started on September 02, 2022 at 12:03:18  
ELAPSED TIME: 00 m 06 s

Execution Data Visualize

## 2. ATM Dimension:

```
CREATE TABLE IF NOT EXISTS SPAR_NORD.DIM_ATM (
  atm_id INT,
  atm_number VARCHAR(20),
  atm_manufacturer VARCHAR(50),
  atm_location_id INT,
  PRIMARY KEY (atm_id),
  FOREIGN KEY (atm_location_id) REFERENCES
  SPAR_NORD.DIM_LOCATION (location_id)
)
DISTSTYLE ALL
INTERLEAVED SORTKEY (atm_id, atm_number, atm_manufacturer);
```



Select database [Info](#)  
To view schemas, select a database.  
spar\_nord\_db

Select schema [Info](#)  
To view tables, select a schema.  
spar\_nord

Filter tables

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

```
1 CREATE TABLE IF NOT EXISTS SPAR_NORD.DIM_ATM (
2   atm_id INT,
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.DIM_LOCATION (location_id)
8 )
9 DISTSTYLE ALL
10 INTERLEAVED SORTKEY (atm_id, atm_number, atm_manufacturer);
```

Run Save Schedule Clear

Send feedback

Query results Table details

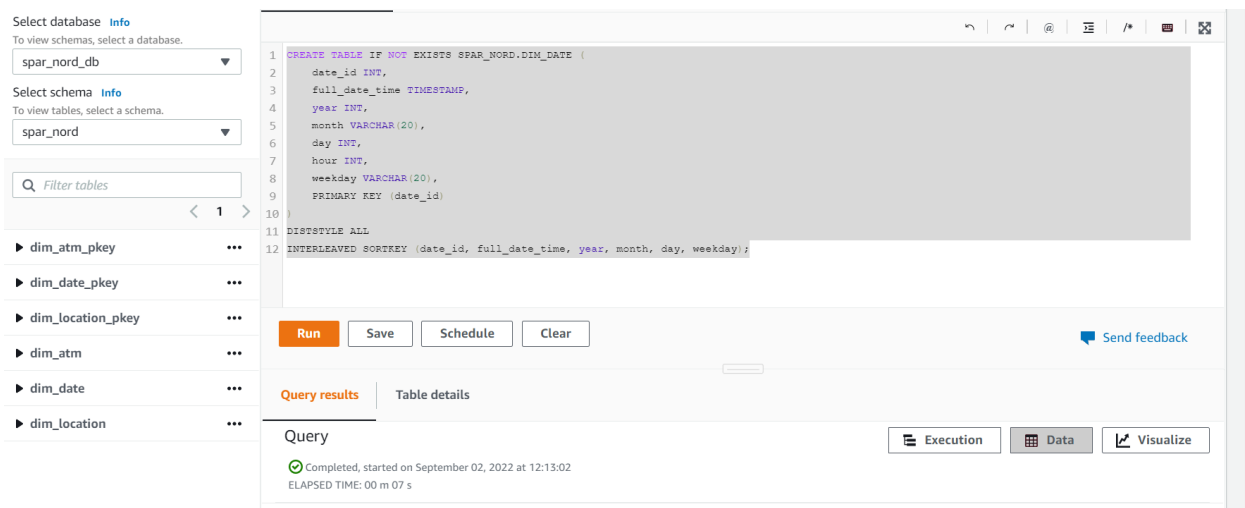
Query

Completed, started on September 02, 2022 at 12:07:32  
ELAPSED TIME: 00 m 26 s

Execution Data Visualize

### 3. Date Dimension:

```
CREATE TABLE IF NOT EXISTS SPAR_NORD.DIM_DATE (
    date_id INT,
    full_date_time TIMESTAMP,
    year INT,
    month VARCHAR(20),
    day INT,
    hour INT,
    weekday VARCHAR(20),
    PRIMARY KEY (date_id)
)
DISTSTYLE ALL
INTERLEAVED SORTKEY (date_id, full_date_time, year, month, day, weekday);
```



The screenshot shows a SQL IDE interface. On the left, there's a sidebar with a tree view showing the database structure: 'spar\_nord\_db' is selected, and under it, 'spar\_nord' schema is chosen. Below this, a list of tables is shown: 'dim\_atm\_pkey', 'dim\_date\_pkey', 'dim\_location\_pkey', 'dim\_atm', 'dim\_date', and 'dim\_location'. The main area displays the SQL query for creating the 'DIM\_DATE' table. The query is:
 

```
1 CREATE TABLE IF NOT EXISTS SPAR_NORD.DIM_DATE (
2     date_id INT,
3     full_date_time TIMESTAMP,
4     year INT,
5     month VARCHAR(20),
6     day INT,
7     hour INT,
8     weekday VARCHAR(20),
9     PRIMARY KEY (date_id)
10 )
11 DISTSTYLE ALL
12 INTERLEAVED SORTKEY (date_id, full_date_time, year, month, day, weekday);
```

 Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted. Below these buttons, there's a section for 'Query results' and 'Table details'. The 'Query results' section shows a status message: 'Completed, started on September 02, 2022 at 12:13:02' and 'ELAPSED TIME: 00 m 07 s'. At the bottom right, there are tabs for 'Execution', 'Data', and 'Visualize'.

### 4. Card Type dimension:

```
CREATE TABLE IF NOT EXISTS SPAR_NORD.DIM_CARD_TYPE (
    card_type_id INT,
    card_type VARCHAR(30),
    PRIMARY KEY (card_type_id)
)
DISTSTYLE ALL
INTERLEAVED SORTKEY (card_type_id, card_type);
```

Select database [Info](#)

To view schemas, select a database.

spar\_nord\_db

Select schema [Info](#)

To view tables, select a schema.

spar\_nord

Filter tables

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

```

1 CREATE TABLE IF NOT EXISTS SPAR_NORD.DIM_CARD_TYPE (
2   card_type_id INT,
3   card_type VARCHAR(30),
4   PRIMARY KEY (card_type_id)
5 )
6 DISTSTYLE ALL
7 INTERLEAVED SORTKEY (card_type_id, card_type);

```

Run Save Schedule Clear

[Send feedback](#)

Query results Table details

Query

Completed, started on September 02, 2022 at 12:14:04  
ELAPSED TIME: 00 m 06 s

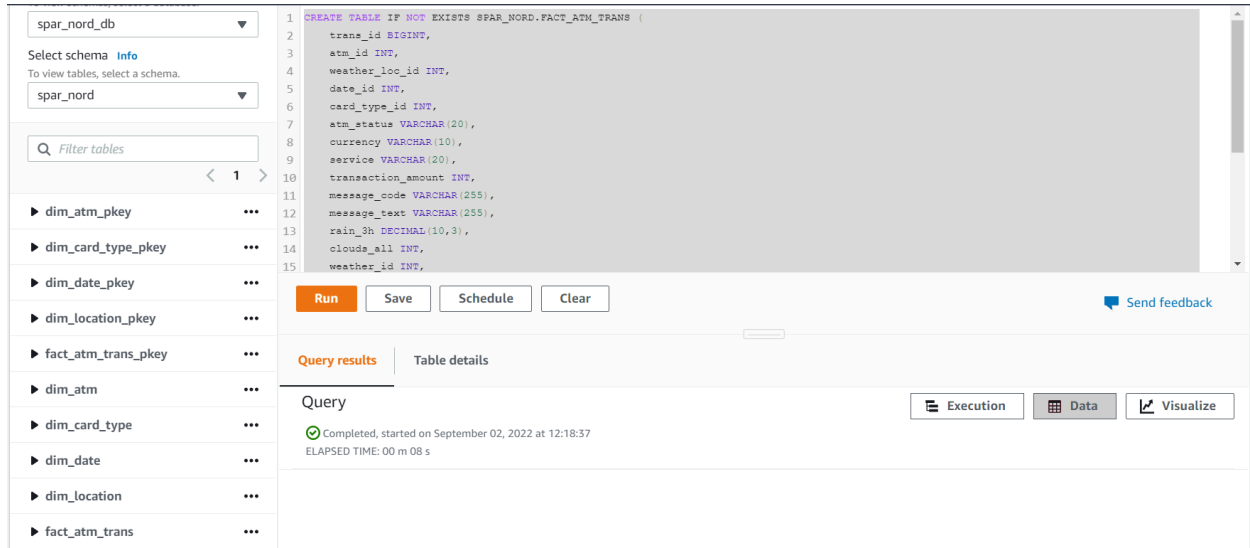
Execution Data Visualize

## 5. ATM Transactions fact:

```

CREATE TABLE IF NOT EXISTS SPAR_NORD.FACT_ATM_TRANS (
  trans_id BIGINT,
  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 SPAR_NORD.DIM_LOCATION
(location_id),
  FOREIGN KEY (atm_id) REFERENCES SPAR_NORD.DIM_ATM (atm_id),
  FOREIGN KEY (date_id) REFERENCES SPAR_NORD.DIM_DATE (date_id),
  FOREIGN KEY (card_type_id) REFERENCES SPAR_NORD.DIM_CARD_TYPE
(card_type_id)
)
DISTSTYLE KEY DISTKEY (trans_id)
INTERLEAVED SORTKEY (trans_id, atm_status, currency, service, message_code,
weather_main);

```



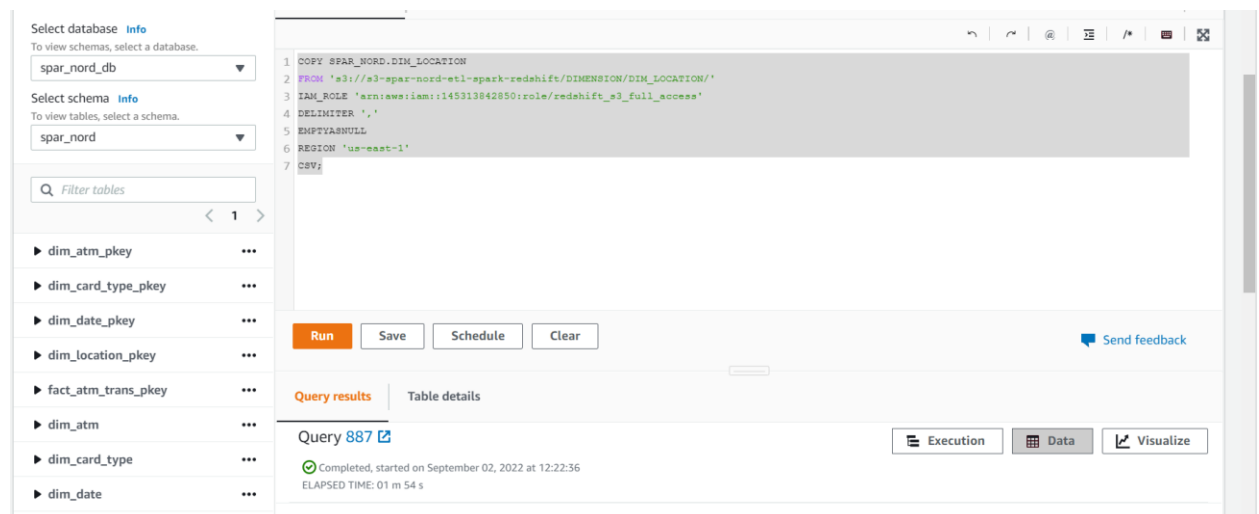
The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with a dropdown menu for 'spar\_nord\_db' and a list of tables including 'dim\_atm\_pkey', 'dim\_card\_type\_pkey', 'dim\_date\_pkey', 'dim\_location\_pkey', 'fact\_atm\_trans\_pkey', 'dim\_atm', 'dim\_card\_type', 'dim\_date', 'dim\_location', and 'fact\_atm\_trans'. The main area displays a SQL query to create a table 'FACT\_ATM\_TRANS' with various columns like 'trans\_id', 'atm\_id', 'weather\_loc\_id', 'date\_id', 'card\_type\_id', 'atm\_status', 'currency', 'service', 'transaction\_amount', 'message\_code', 'message\_text', 'rain\_3h', 'clouds\_all', and 'weather\_id'. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is selected, showing a status of 'Completed' and an elapsed time of '00 m 08 s'.

## 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

### 1. Location Dimension:

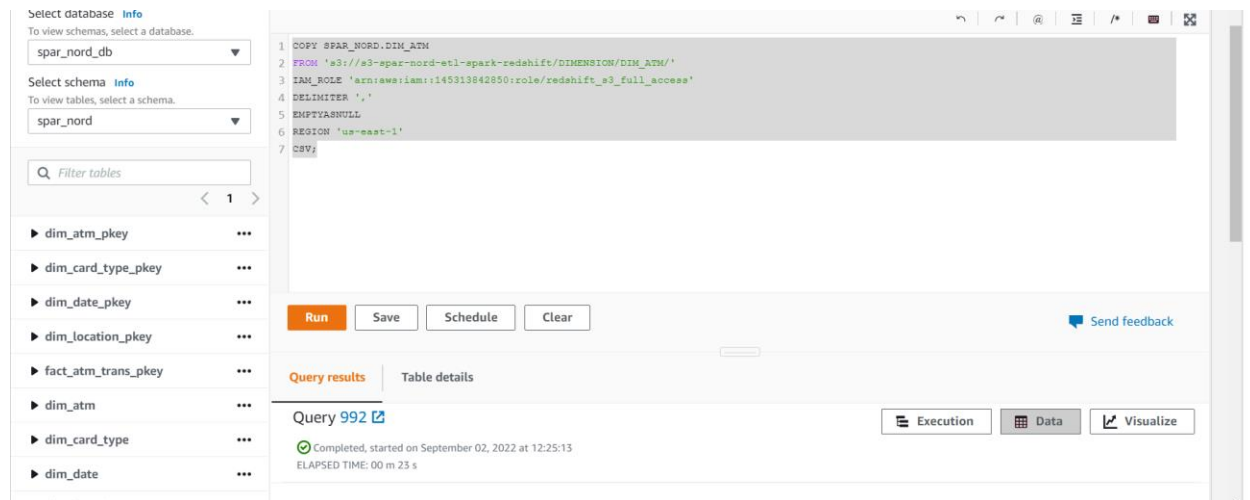
```
COPY SPAR_NORD.DIM_LOCATION
FROM 's3://s3-spar-nord-etl-spark-redshift/DIMENSION/DIM_LOCATION/'
IAM_ROLE 'arn:aws:iam::145313842850:role/redshift_s3_full_access'
DELIMITER ','
EMPTYASNULL
REGION 'us-east-1'
CSV;
```



The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with a dropdown menu for 'spar\_nord\_db' and 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 area displays a SQL query to copy data from an S3 bucket to the 'DIM\_LOCATION' table. The query includes the 'FROM' clause with the S3 path, the 'IAM\_ROLE' clause, the 'DELIMITER' clause, the 'EMPTYASNULL' clause, the 'REGION' clause, and the 'CSV' clause. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is selected, showing a status of 'Completed' and an elapsed time of '01 m 54 s'.

## 2. ATM Transaction:

```
COPY SPAR_NORD.DIM_ATM
FROM 's3://s3-spar-nord-etl-spark-redshift/DIMENSION/DIM_ATM/'
IAM_ROLE 'arn:aws:iam::145313842850:role/redshift_s3_full_access'
DELIMITER ','
EMPTYASNULL
REGION 'us-east-1'
CSV;
```



The screenshot shows a SQL query editor interface. On the left, there is a sidebar with a 'Select database' dropdown set to 'spar\_nord\_db' and a 'Select schema' dropdown set to 'spar\_nord'. Below these are search filters and 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', and 'dim\_date'. The main area displays a SQL query with line numbers 1 through 7, matching the code provided in the previous block. Below the query editor are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. A 'Send feedback' link is also present. At the bottom, there is a 'Query results' section showing 'Query 992' with a status of 'Completed, started on September 02, 2022 at 12:25:13' and an 'ELAPSED TIME: 00 m 23 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

## 3. Date Dimension:

```
COPY SPAR_NORD.DIM_DATE
FROM 's3://s3-spar-nord-etl-spark-redshift/DIMENSION/DIM_DATE/'
IAM_ROLE 'arn:aws:iam::145313842850:role/redshift_s3_full_access'
DELIMITER ','
EMPTYASNULL
REGION 'us-east-1'
TIMEFORMAT AS 'auto'
CSV;
```



Select database [Info](#)

To view schemas, select a database.

spar\_nord\_db

Select schema [Info](#)

To view tables, select a schema.

spar\_nord

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
- dim\_location

```

1 COPY SPAR_NORD.DIM_DATE
2 FROM 's3://s3-spar-nord-etl-spark-redshift/DIMENSION/DIM_DATE/'
3 IAM_ROLE 'arn:aws:iam::145313842850:role/redshift_s3_full_access'
4 DELIMITER ','
5 EMPTYASNULL
6 REGION 'us-east-1'
7 TIMEFORMAT AS 'auto'
8 CSV;

```

Run Save Schedule Clear

[Send feedback](#)

Query results Table details

Query 1403 [🔗](#)

Execution Data Visualize

Completed, started on September 02, 2022 at 12:53:56  
ELAPSED TIME: 01 m 22 s

#### 4. Card Type Dimension:

```

COPY SPAR_NORD.DIM_CARD_TYPE
FROM 's3://s3-spar-nord-etl-spark-redshift/DIMENSION/DIM_CARD_TYPE/'
IAM_ROLE 'arn:aws:iam::145313842850:role/redshift_s3_full_access'
DELIMITER ','
EMPTYASNULL
REGION 'us-east-1'
CSV;

```

Select database [Info](#)

To view schemas, select a database.

spar\_nord\_db

Select schema [Info](#)

To view tables, select a schema.

spar\_nord

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

```

1 COPY SPAR_NORD.DIM_CARD_TYPE
2 FROM 's3://s3-spar-nord-etl-spark-redshift/DIMENSION/DIM_CARD_TYPE/'
3 IAM_ROLE 'arn:aws:iam::145313842850:role/redshift_s3_full_access'
4 DELIMITER ','
5 EMPTYASNULL
6 REGION 'us-east-1'
7 CSV;

```

Run Save Schedule Clear

[Send feedback](#)

Query results Table details

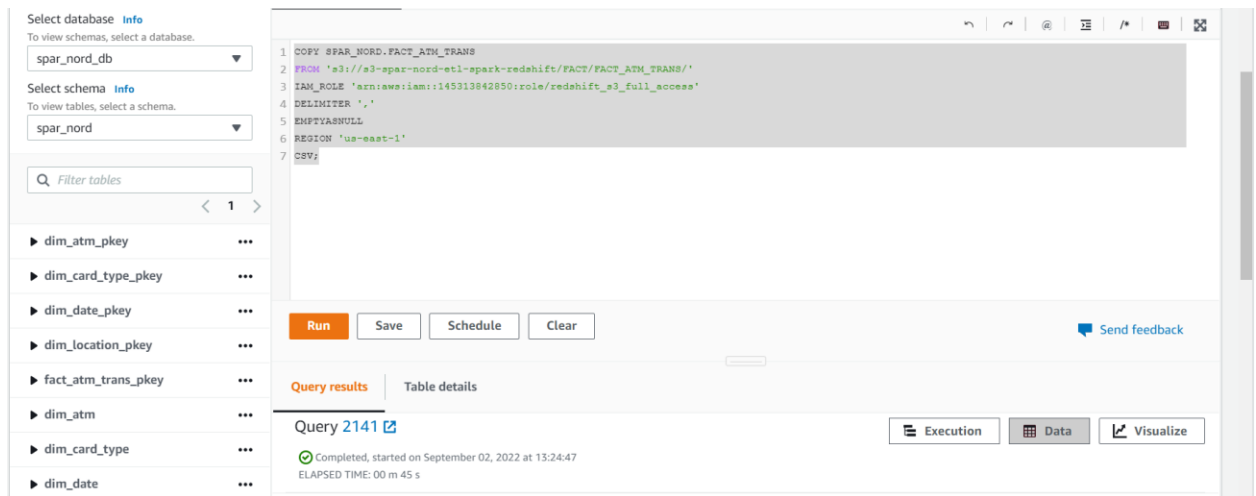
Query 2091 [🔗](#)

Execution Data Visualize

Completed, started on September 02, 2022 at 13:22:42  
ELAPSED TIME: 00 m 22 s

## 5. ATM Transaction Fact:

```
COPY SPAR_NORD.FACT_ATM_TRANS
FROM 's3://s3-spar-nord-etl-spark-redshift/FACT/FACT_ATM_TRANS/'
IAM_ROLE 'arn:aws:iam::145313842850:role/redshift_s3_full_access'
DELIMITER ','
EMPTYASNULL
REGION 'us-east-1'
CSV;
```



Select database [Info](#)  
To view schemas, select a database.  
spar\_nord\_db

Select schema [Info](#)  
To view tables, select a schema.  
spar\_nord

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 ...

```
1 COPY SPAR_NORD.FACT_ATM_TRANS
2 FROM 's3://s3-spar-nord-etl-spark-redshift/FACT/FACT_ATM_TRANS/'
3 IAM_ROLE 'arn:aws:iam::145313842850:role/redshift_s3_full_access'
4 DELIMITER ','
5 EMPTYASNULL
6 REGION 'us-east-1'
7 CSV;
```

Run Save Schedule Clear

[Send feedback](#)

Query results Table details

Query 2141 [🔗](#)

Execution Data Visualize

Completed, started on September 02, 2022 at 13:24:47  
ELAPSED TIME: 00 m 45 s