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

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	**Status**	**Node type**	**Endpoint**
spar-nord-redshift-cluster	✅ Available	dc2.large	🗐 spar-nord-redshift-cluster.cvljkjo31h…
Cluster namespace	**Date created**	**Number of nodes**	**JDBC URL**
9e23e5a4-8688-40a5-8e66-7b279588235c	September 02, 2022, 11:42 (UTC+05:30)	2	🗐 jdbc:redshift://spar-nord-redshift-clu…
	Storage used	**AQUA**	**ODBC URL**
	-	Not available	🗐 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:

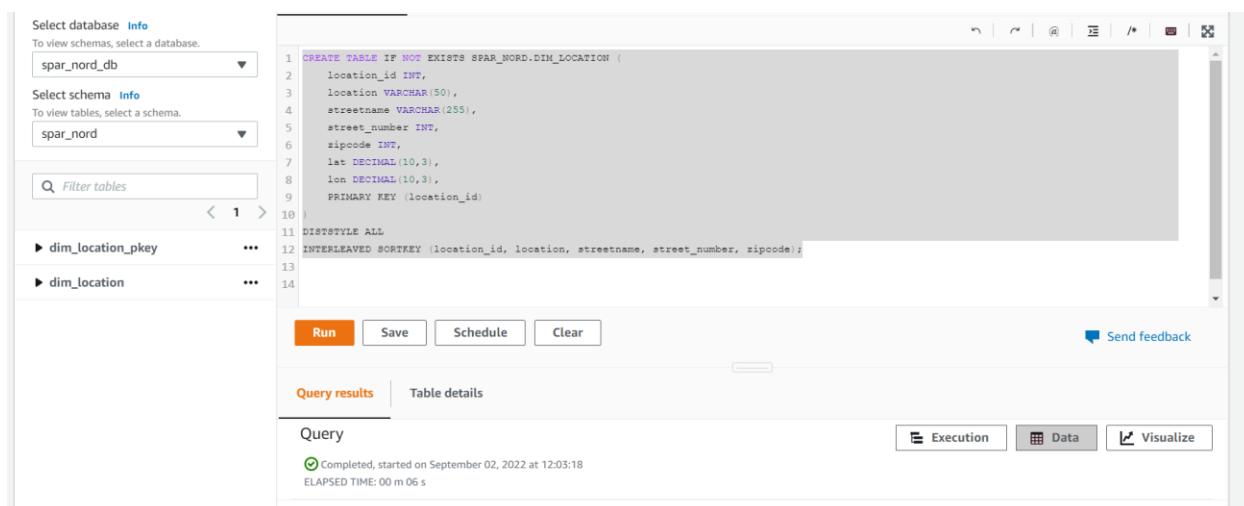**Create a schema for all the dimensions and fact tables to be created inside it.**

    CREATE SCHEMA SPAR_NORD;
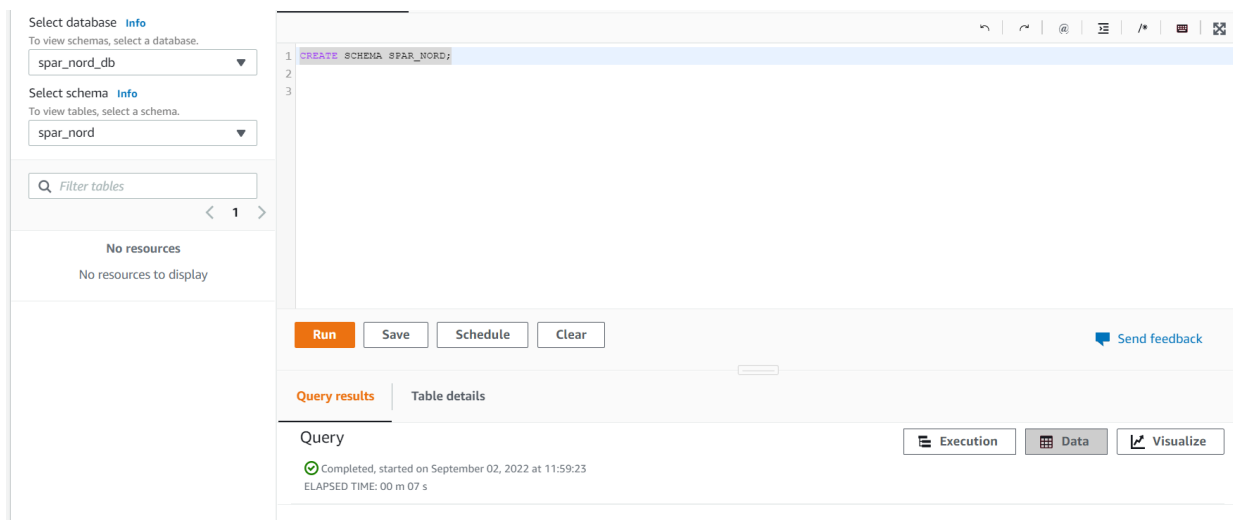
Create all dimensions and facts tables

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);
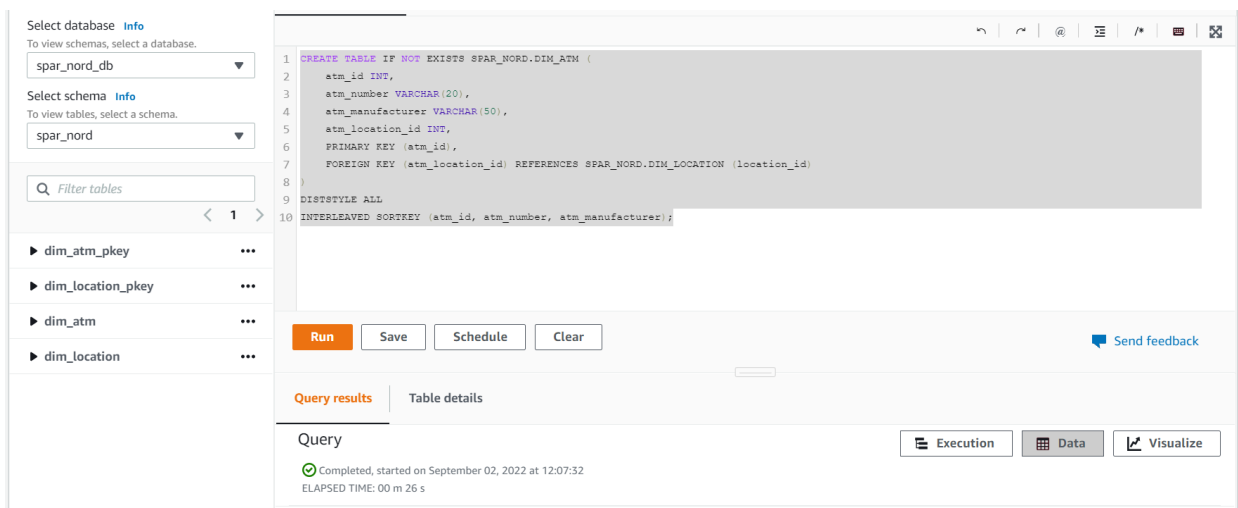
2. ATM Dimension:

```
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);
```
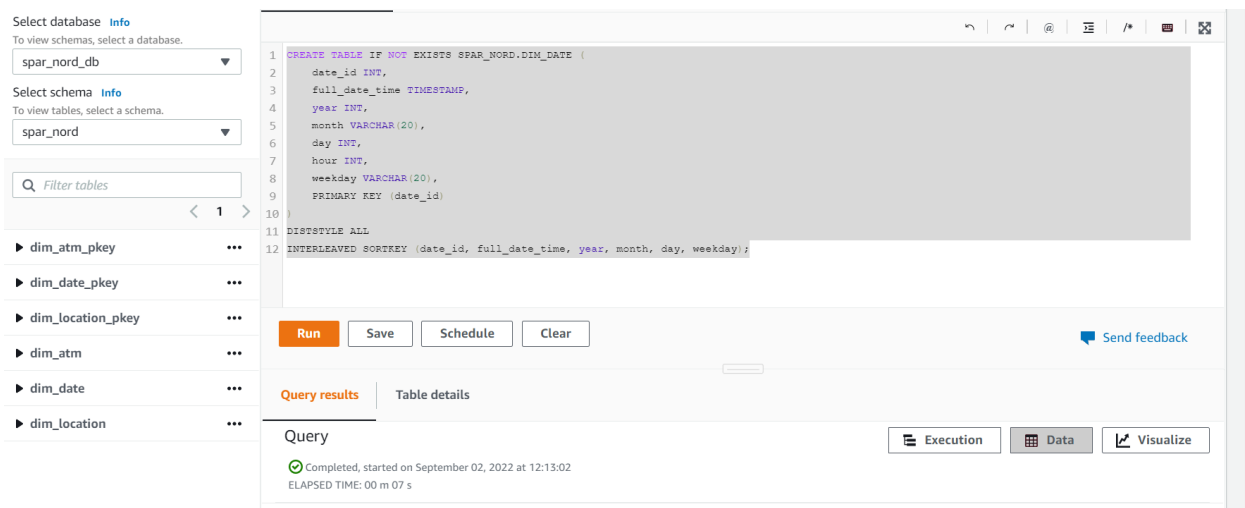
3. **Date Dimension:**

```
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);
```


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);
```
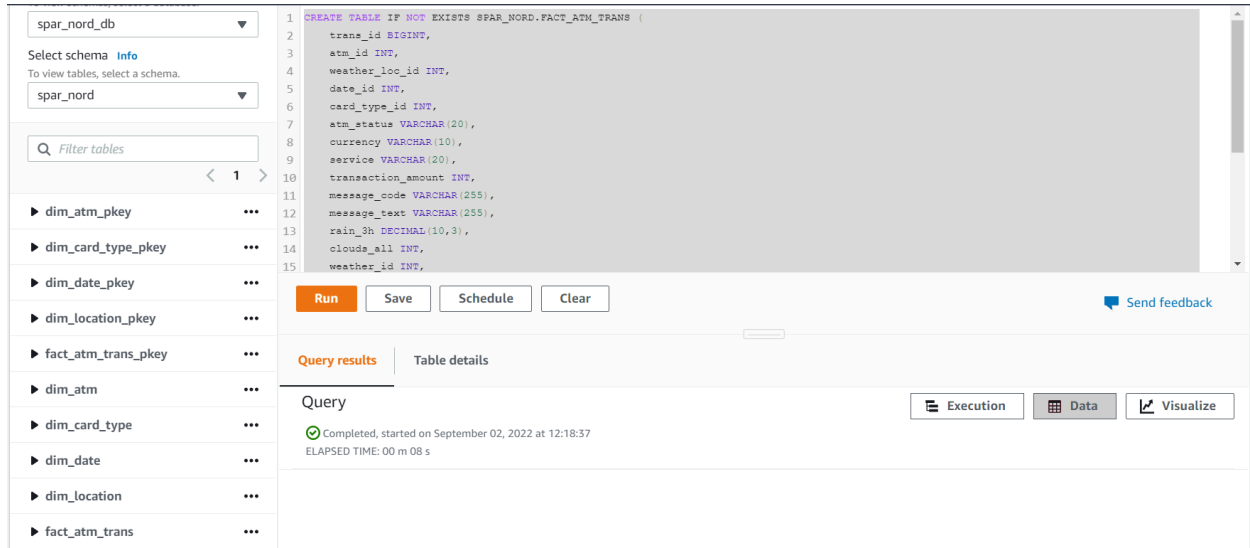
5. **ATM Transactions fact:**

```
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);
```

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
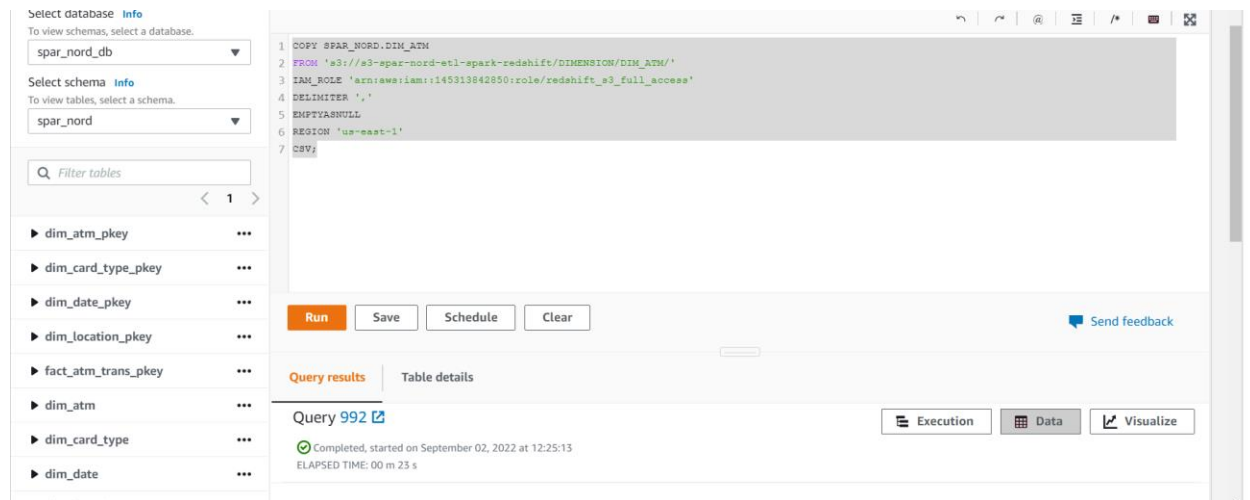
1. **Location Dimension:**

```
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;
```

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;
```


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;
```

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;
```
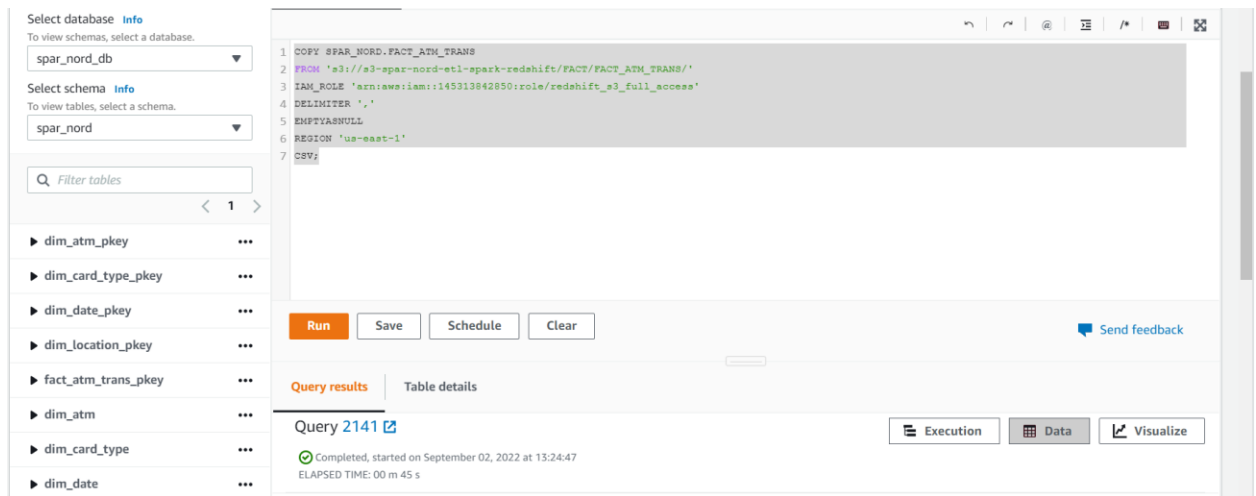
5. ATM Transaction Fact:

```
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;
```