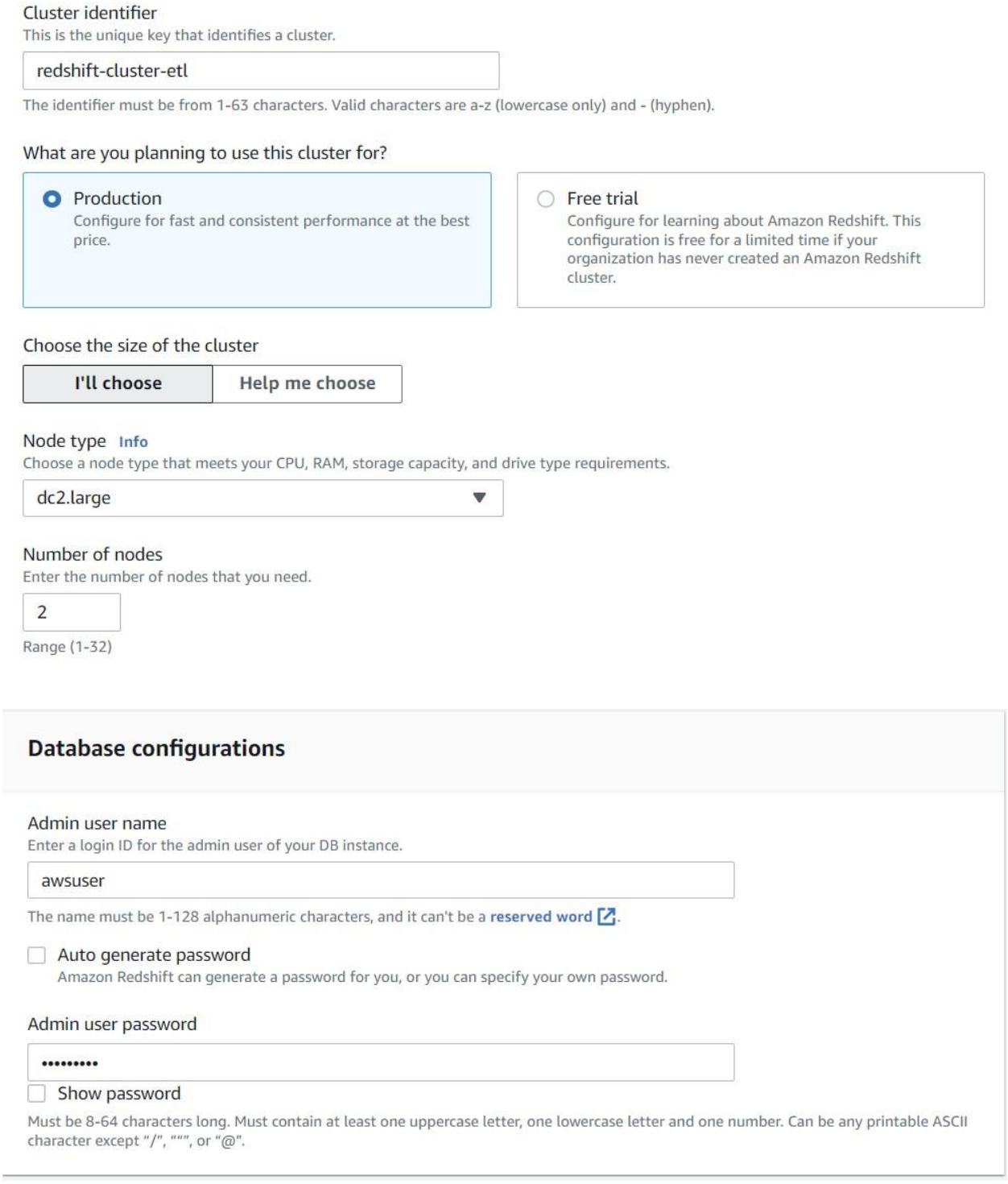
Creation of a Redshift Cluster

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



Database configurations

Database name

Specify a database name to create an additional database.

etl

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. You can't change the port after the cluster has been created.

5450

The port must be numeric (1150-65535).

Parameter groups

Defines database parameter and query queues for all the databases.

default.redshift-1.0
Default parameter group for redshift-1.0

Encryption

Encrypt all data on your cluster.

- (•) Disabled
- ( ) Use AWS Key Management Service (AWS KMS)
- ( ) Use a hardware security module (HSM)

Cluster permissions

Create an IAM role as the default for this cluster that has the AmazonRedshiftAllCommandsFullAccess policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

Associated IAM roles (1) Info

Create, associate, or remove an IAM role. You can associate up to 50 IAM roles. You can also choose an IAM role and set it as the default for this cluster.

Set default | Manage IAM roles

Search for associated IAM role by name, status, or role type

1

IAM roles	Status	Role type
s3fullredshiftrole	adding	--

Database configurations

Change admin user password | Rotate encryption keys | Edit

Database name etl	Parameter group Defines database parameter and query queues for all the databases. default.redshift-1.0	Encryption Disabled	Audit logging Disabled
Port 5450	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADAQAB...	AWS KMS key ID -	
Admin user name awsuser			

Network and security settings

Edit

Virtual private cloud (VPC) vpc-036fbb4c2ae49b03a	Availability Zone us-east-1a	VPC security group Specify which instances and devices can connect to the cluster. sg-0e4702975dc1c3860	Publicly accessible Allow instances and devices outside the VPC to connect to the database. Disabled
Subnet cluster-subnet-group-2	Enhanced VPC routing Disabled		
Endpoint URL -			

Network and security Info

Virtual private cloud (VPC)

This VPC defines the virtual networking environment for this cluster.

my_vpc-vpc
vpc-036fbb4c2ae49b03a

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

cloudera
sg-0e4702975dc1c3860

Cluster subnet group

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

cluster-subnet-group-2

Availability Zone

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

us-east-1a

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
- ( ) Turn on

Amazon Redshift > Clusters > redshift-cluster-etl

redshift-cluster-etl

Actions | Edit | Add partner integration | Query data

General information

Cluster identifier redshift-cluster-etl	Status Modifying	Node type dc2.large	Endpoint redshift-cluster-etl.chhyfrge1dap.us-east-1.re...
Cluster namespace 3d1fb684-2867-4df1-8d17-7c7a79270ef7	Date created September 03, 2022, 14:24 (UTC+05:30)	Number of nodes 2	JDBC URL jdbc:redshift://redshift-cluster-etl.chhyfrge1d...
	Storage used 0.02% (0.06 of 320 GB used)	AQUA Not available	ODBC URL Driver={Amazon Redshift (x64)}; Server=redshi...

Cluster subnet group details

Name
You can't modify the name after your subnet group has been created.

cluster-subnet-group-2

The name must be 1-255 characters. Valid characters are A-Z, a-z, 0-9, space, hyphen (-), underscore (_), and period (.).

Description

ETL project

Add subnets

VPC
Choose the VPC that contains the subnets that you want to include in your cluster subnet group.

my_vpc-vpc
vpc-036fbb4c2ae49b03a

Add all the subnets for this VPC

Availability Zone

us-east-1b (2)

Subnet

subnet-024e1438897519909

Add subnet

Subnets in this cluster subnet group (1)

Remove all

Cluster subnet group cluster-subnet-group-2 was create successfully

Amazon Redshift > Configurations > Subnet groups

Cluster subnet groups (2)

Delete | Actions | Create cluster subnet group

Search

1

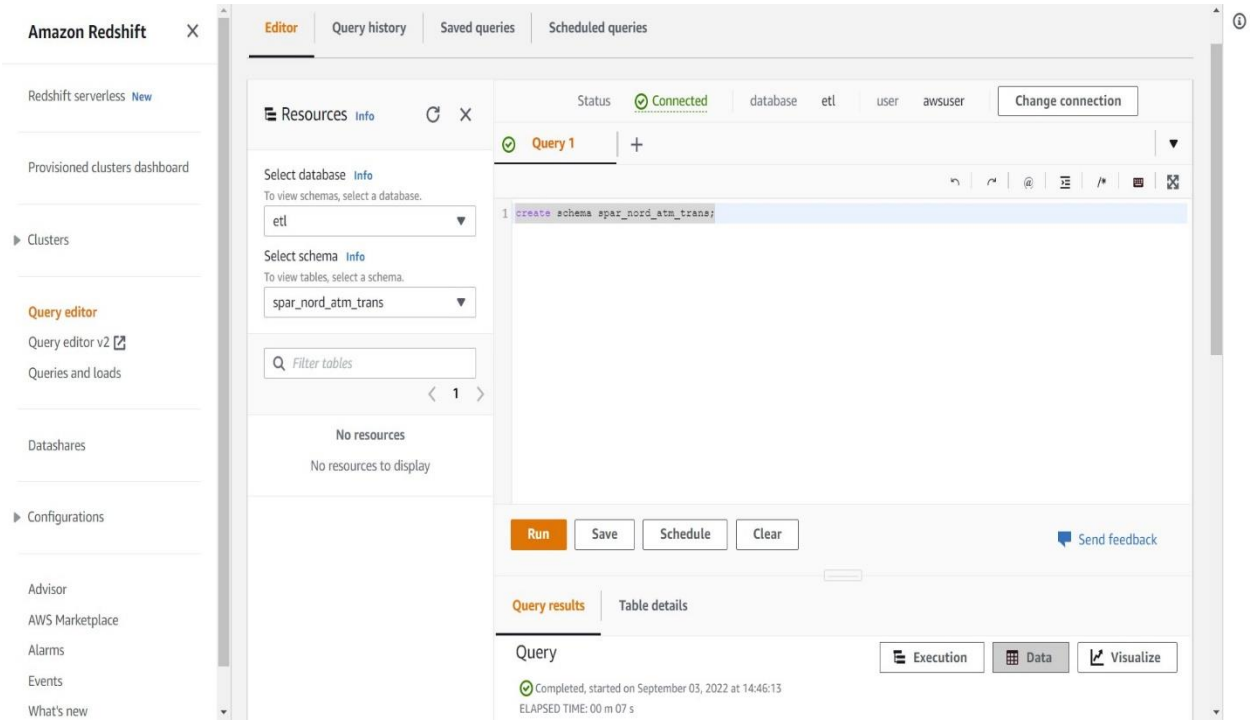
Name	Status	VPC ID	Description	Tags
cluster-subnet-group-1 1 Subnets	Complete	vpc-0bf952dbd58822311	Upgrad demo Redshift cluster	
cluster-subnet-group-2 1 Subnets	Complete	vpc-036fbb4c2ae49b03a	ETL project	

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:

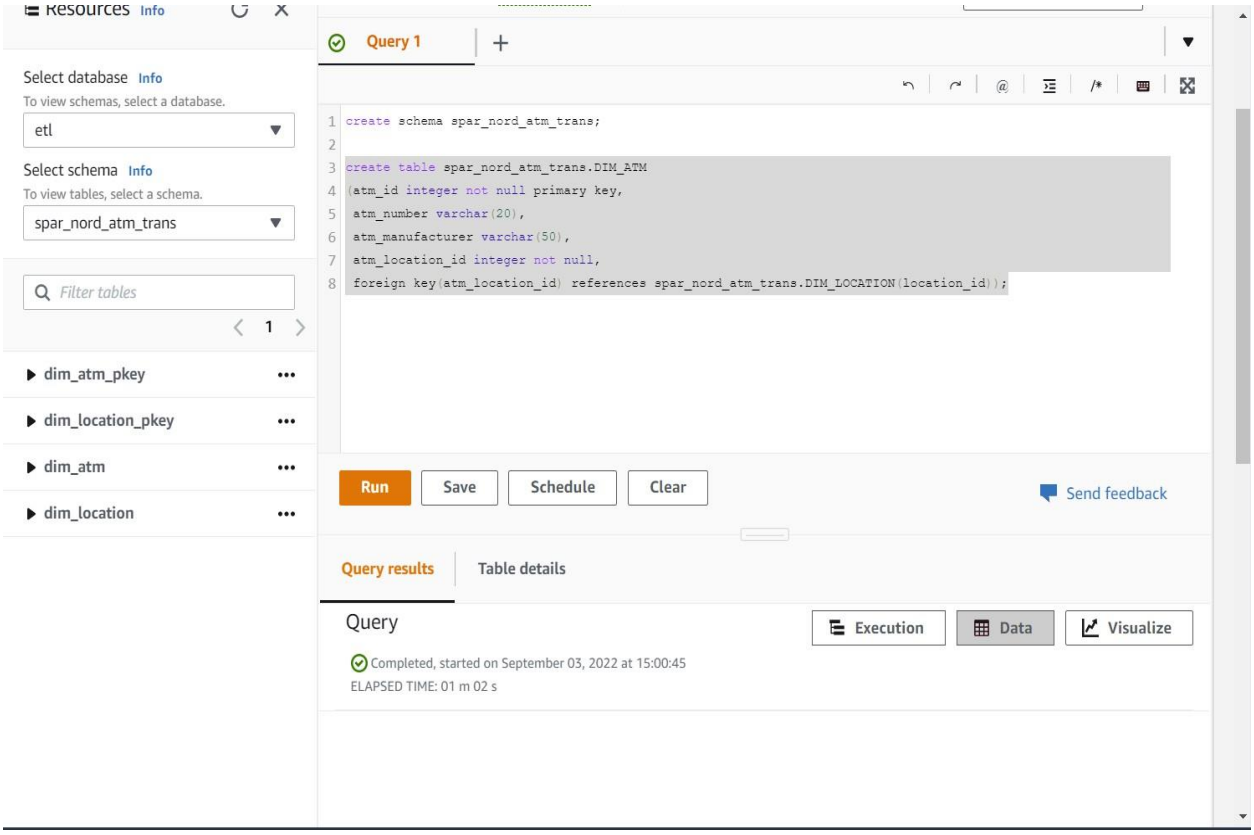
Schema Creation : SPAR_NORD_ATM_TRANS

```
create schema spar_nord_atm_trans;
```



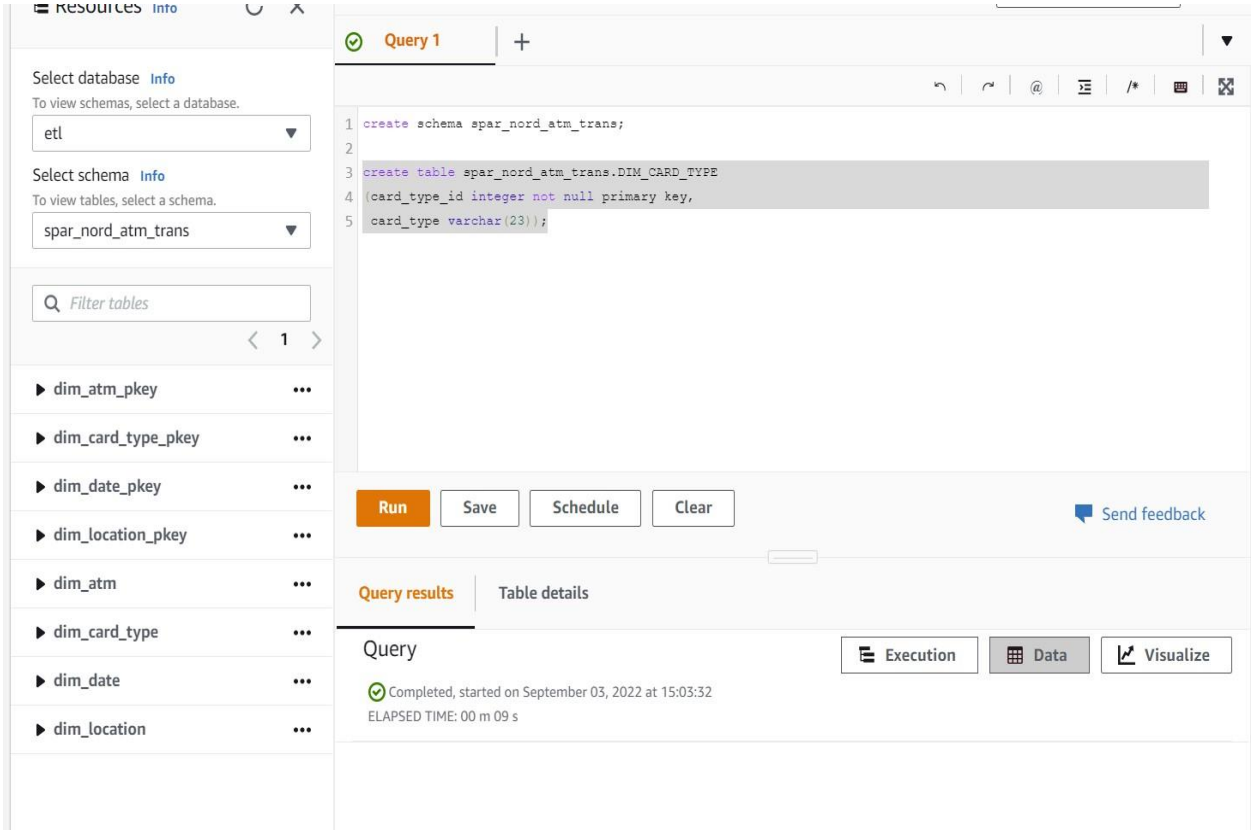
DIMENSION TABLE 1: DIM_ATM

```
create table spar_nord_atm_trans.DIM_ATM
(atm_id integer not null primary key,
 atm_number varchar(20),
 atm_manufacturer varchar(50),
 atm_location_id integer not null,
 foreign key(atm_location_id) references spar_nord_atm_trans.DIM_LOCATION(location_id));
```



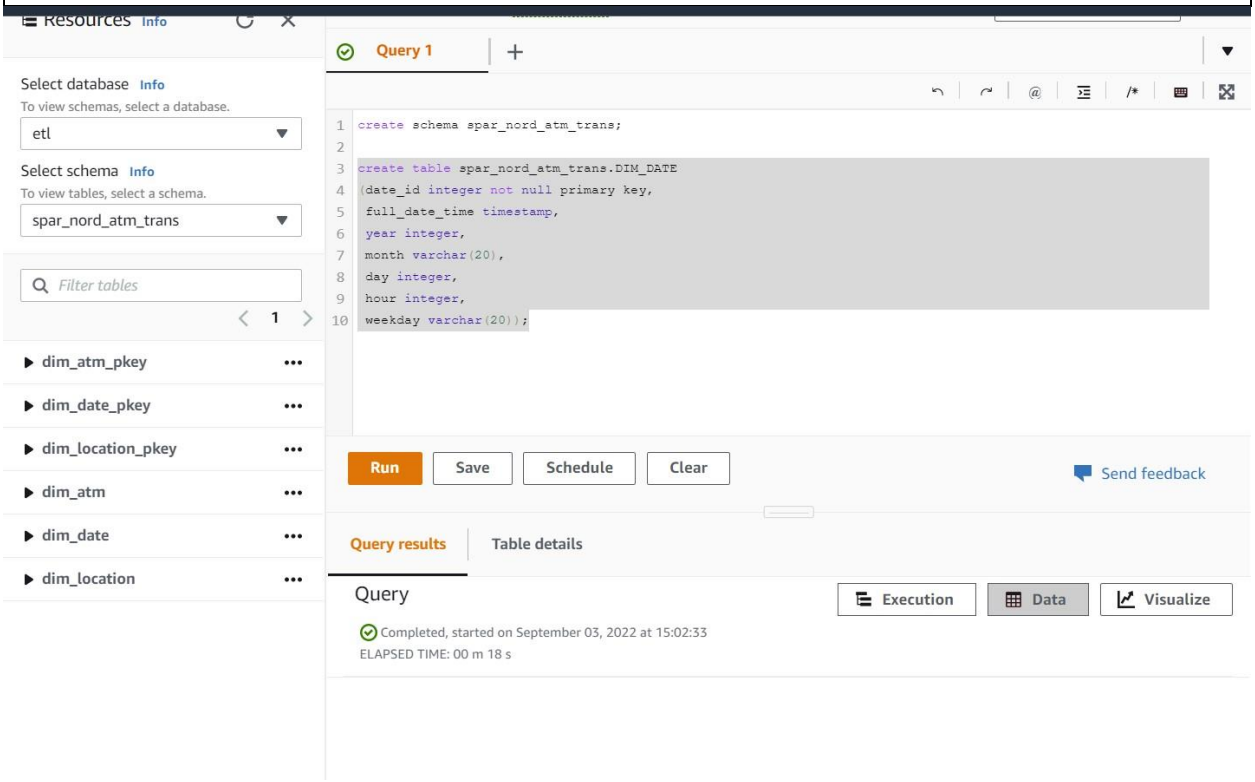
DIMENSION TABLE 2: DIM_CARD_TYPE

```
create table spar_nord_atm_trans.DIM_CARD_TYPE
(card_type_id integer not null primary key,
 card_type varchar(23));
```



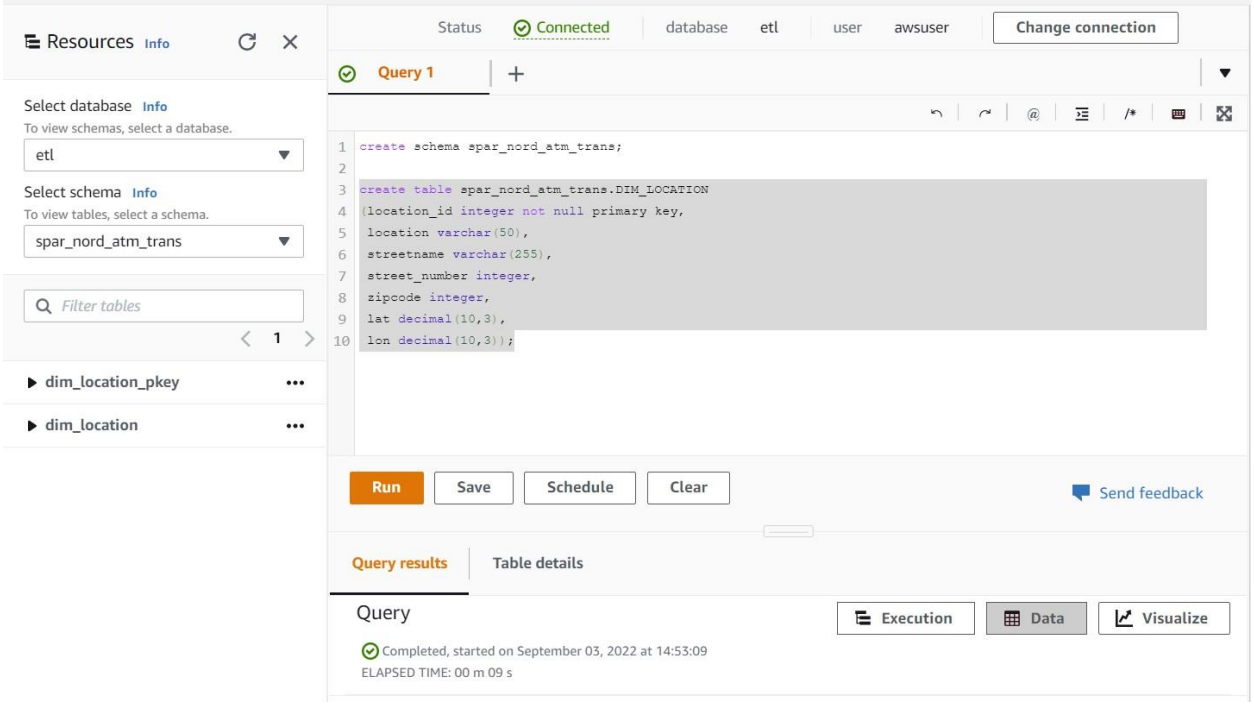
DIMENSION TABLE 3: DIM_DATE

```
create table spar_nord_atm_trans.DIM_DATE
(date_id integer not null primary key,
 full_date_time timestamp,
 year integer,
 month varchar(20),
 day integer,
 hour integer,
 weekday varchar(20));
```



DIMENSION TABLE 4: DIM_LOCATION

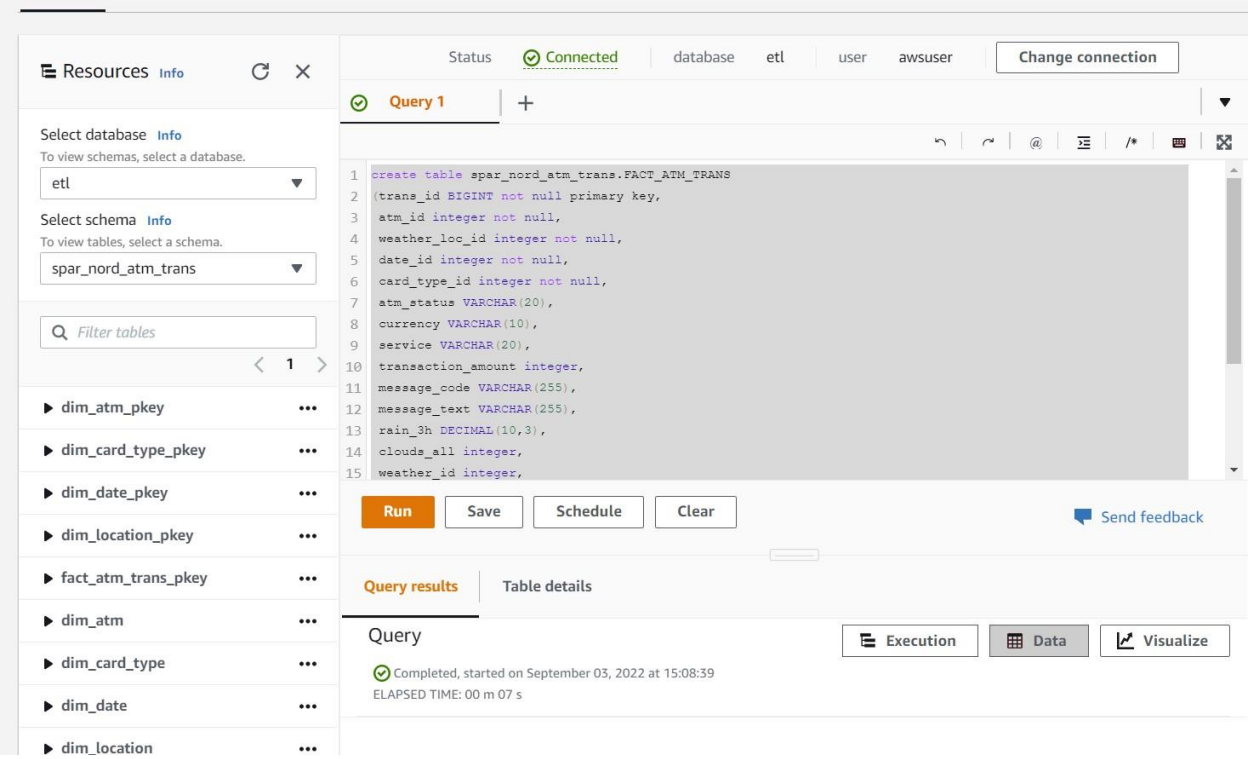
```
create table spar_nord_atm_trans.DIM_LOCATION
(location_id integer not null primary key,
location varchar(50),
streetname varchar(255),
street_number integer,
zipcode integer,
lat decimal(10,3),
lon decimal(10,3));
```



FACT TABLE : FACT_ATM_TRANS

```
create table spar_nord_atm_trans.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_trans.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references spar_nord_atm_trans.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references spar_nord_atm_trans.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references spar_nord_atm_trans.DIM_CARD_TYPE(card_type_id));
```

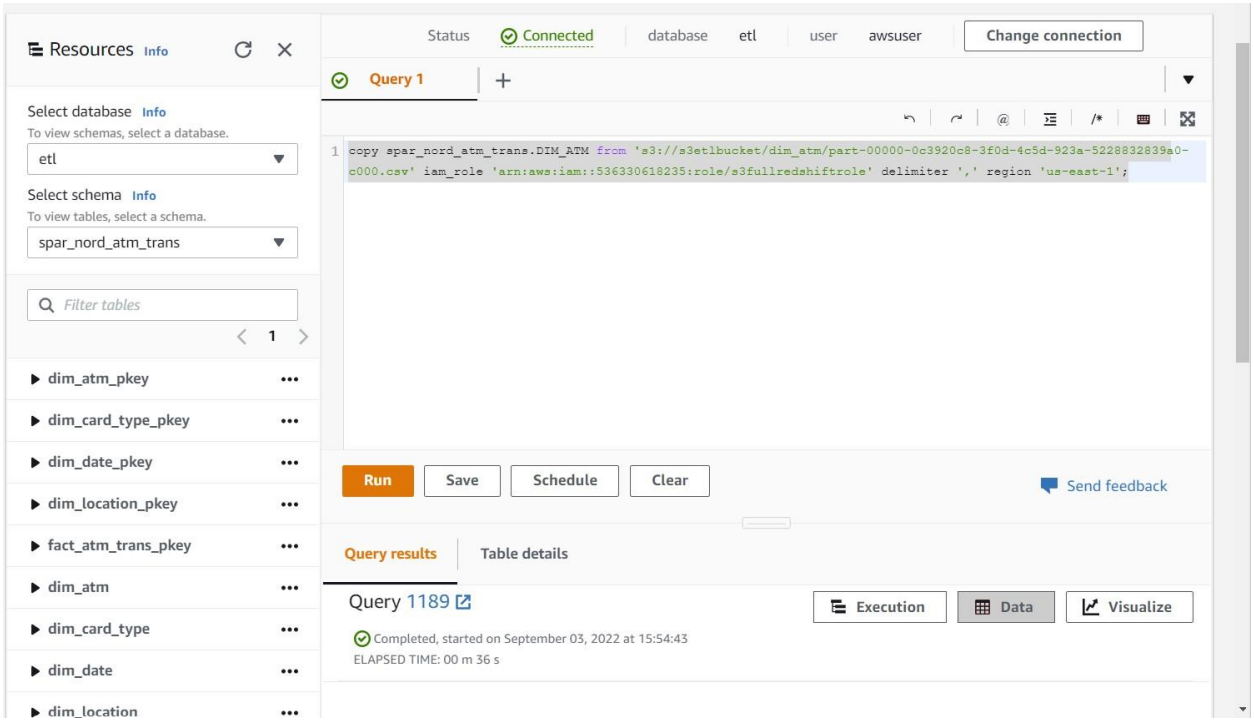


Loading data into a Redshift cluster from Amazon S3 bucket

Queries to copy the data from S3 buckets to the Redshift cluster in the appropriate tables

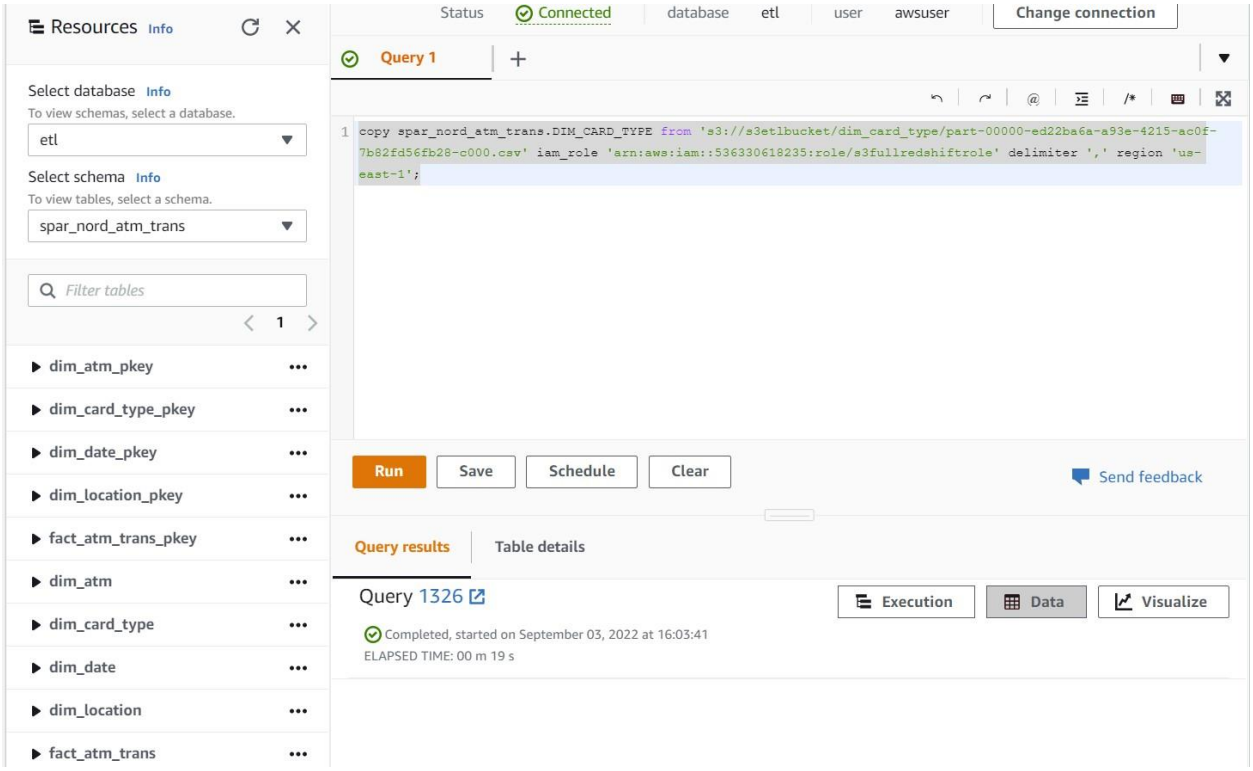
S3 to Redshift DIM_ATM

```
copy spar_nord_atm_trans.DIM_ATM from 's3://s3etlbucket/dim_atm/part-00000-f38f2738-f84e-4581-af21-6eaebeedef11-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1';
```



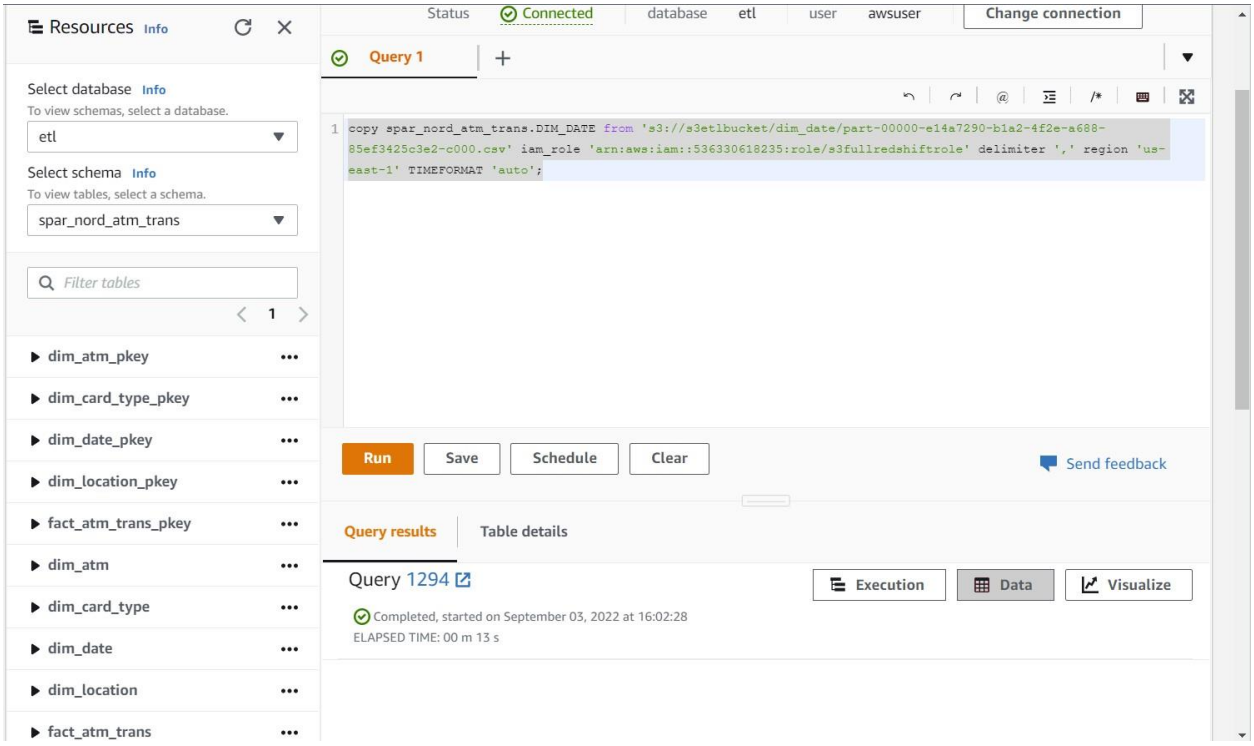
S3 to Redshift DIM_CARD_TYPE

```
copy spar_nord_atm_trans.DIM_CARD_TYPE from 's3://s3etlbucket/dim_card_type/part-00000-e7170186-2eb6-4b2d-a1f0-f23527062a3b-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1';
```



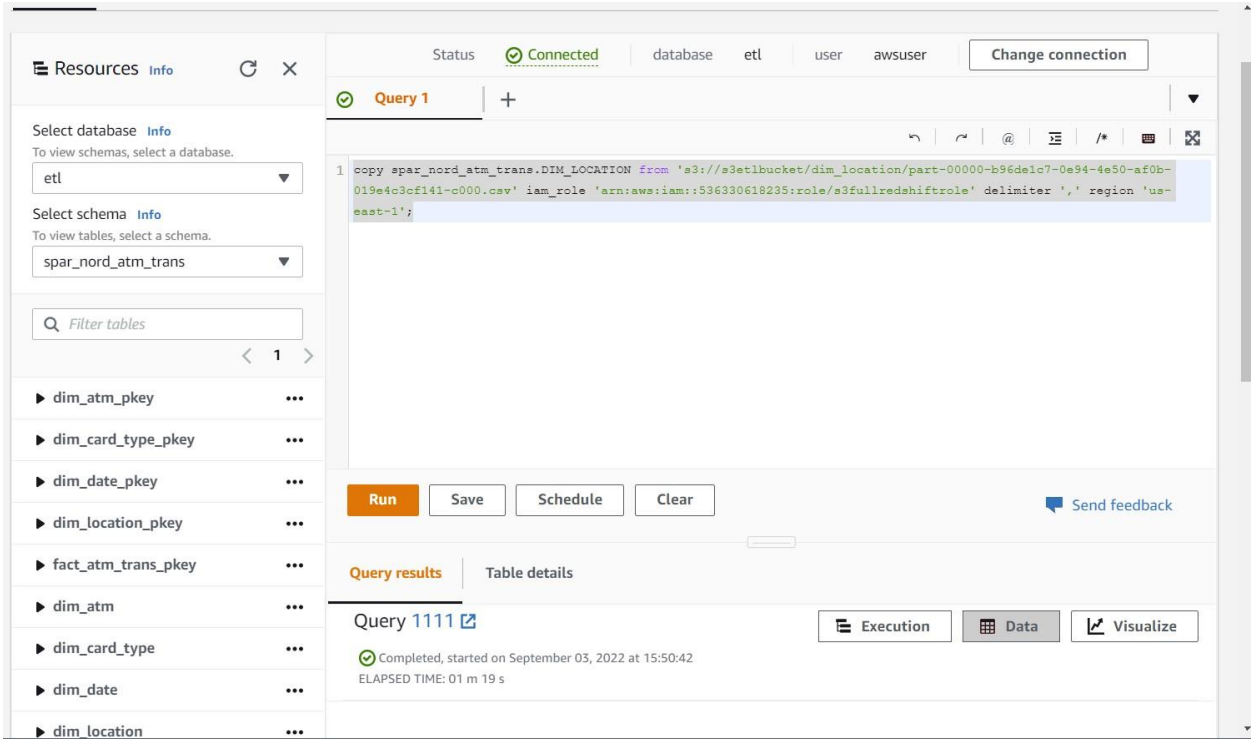
S3 to Redshift DIM_DATE

```
copy spar_nord_atm_trans.DIM_DATE from 's3://s3etlbucket/dim_date/part-00000-ace51bd0-f726-40e3-9a8f-3c9bfe6066de-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1' TIMEFORMAT 'auto';
```



S3 to Redshift DIM_LOCATION

```
copy spar_nord_atm_trans.DIM_LOCATION from 's3://s3etlbucket/dim_location/part-00000-4083086e-aae6-44e1-833d-88b52c5f5f0d-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1';
```



S3 to Redshift FACT_ATM_TRANS

```
copy spar_nord_atm_trans.FACT_ATM_TRANS from 's3://s3etlbucket/fact_atm_trans/part-00000-6d0fcdcf-adb2-4e08-bfd6-e24254fc7d11-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1' CSV;
```