

## Creation of a Redshift Cluster

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

#### 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 "@".

## ▼ Database configurations

### Database name

Specify a database name to create an additional database.

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.

The port must be numeric (1150-65535).

### Parameter groups

Defines database parameter and query queues for all the databases.

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 ▼](#) 

 [< 1 >](#) 

<input type="checkbox"/> IAM roles <a href="#"> ⓘ </a>	Status	Role type
<input type="checkbox"/> s3fullredshiftrole	<a href="#"> ⓘ </a> 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 AWS KMS key ID -	Audit logging Disabled
Port 5450	SSH ingestion setting (cluster public key) <a href="#"> ⓘ </a> ssh-rsa AAAAB3NzaC1yc2EAAAADAQAB...		
Admin user name awsuser			

## Network and security settings

 [Edit](#) 

Virtual private cloud (VPC) <a href="#"> vpc-036fbb4c2ae49b03a ⓘ </a>	Availability Zone us-east-1a	VPC security group Specify which instances and devices can connect to the cluster. <a href="#"> sg-0e4702975dc1c3860 ⓘ </a>	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 <a href="#"> ⓘ </a> -			

## ▼ 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 > ⚙

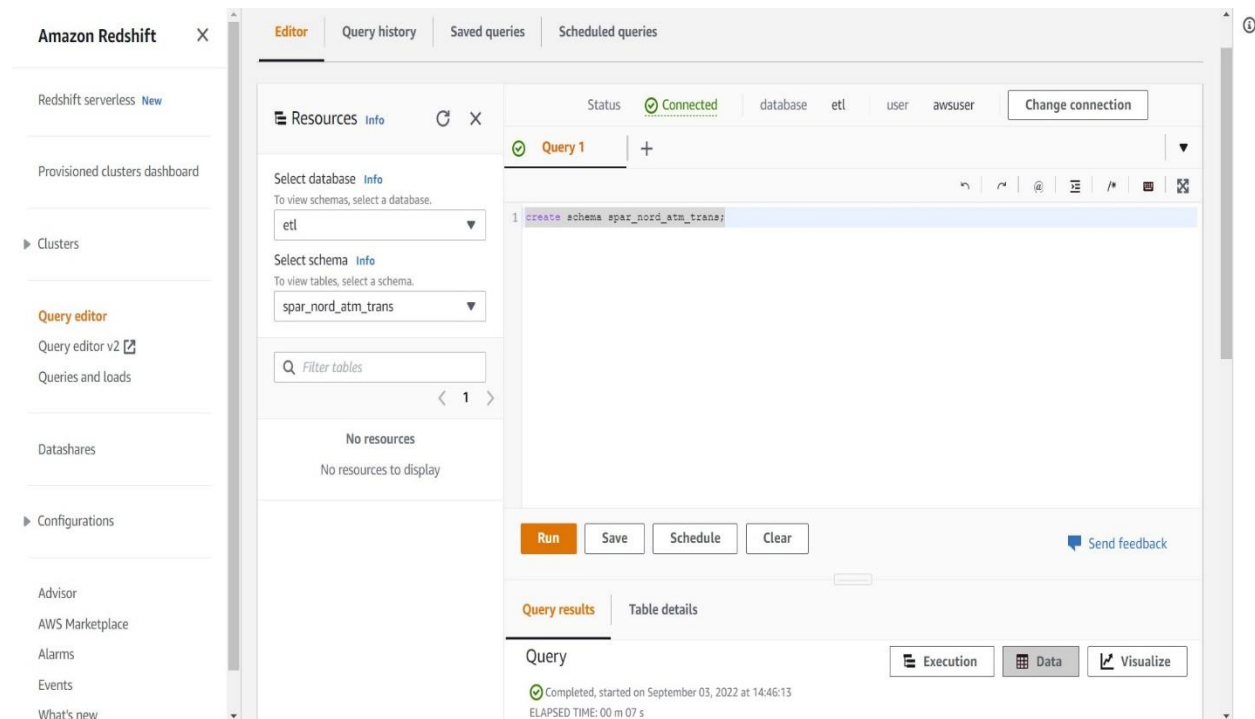
<input type="checkbox"/>	Name ▲	Status ▼	VPC ID ▼	Description ▼	Tags ▼
<input type="checkbox"/>	cluster-subnet-group-1 1 Subnets	Complete	vpc-0bf952dbd58822311	Upgrad demo Redshift cluster	
<input type="checkbox"/>	cluster-subnet-group-2 1 Subnets	Complete	vpc-036fbb4c2ae49b03a	ETL project	

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

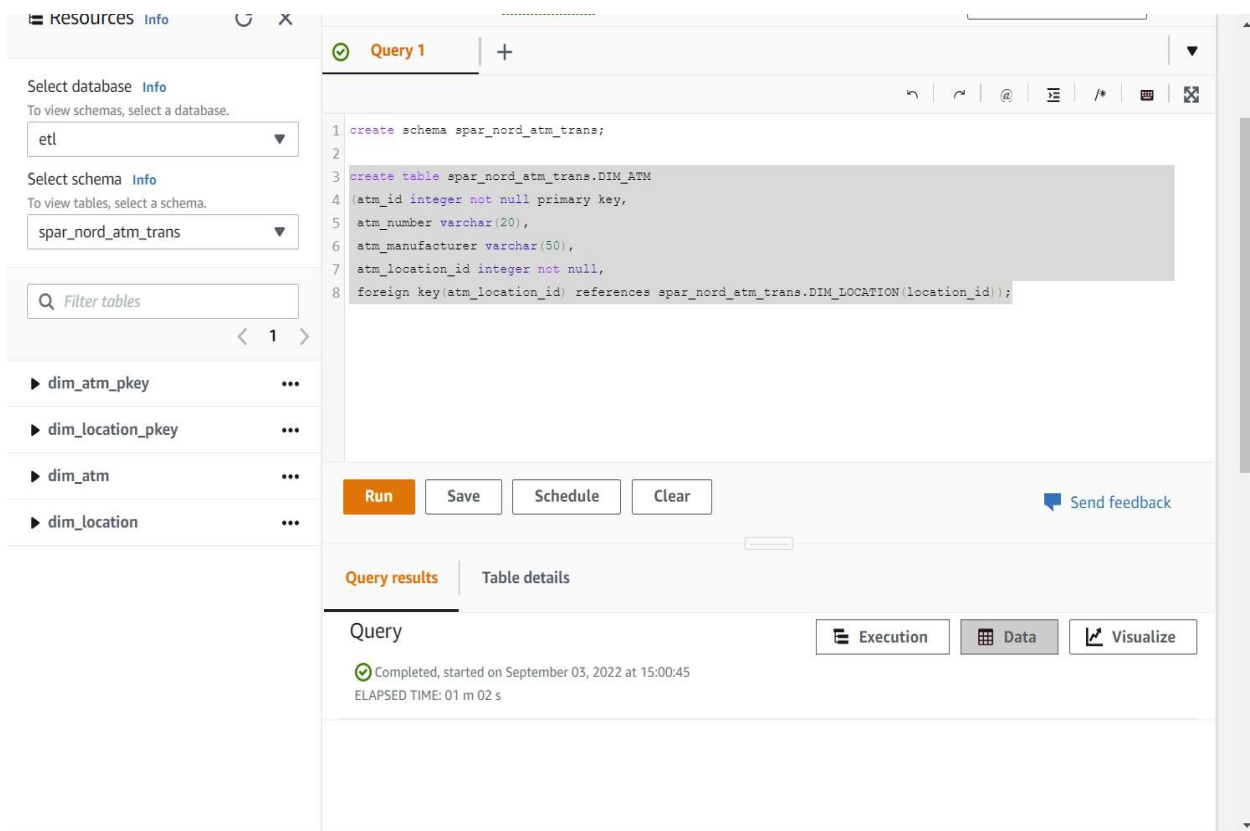
```
create schema spar_nord_atm_trans;
```



The screenshot displays the Amazon Redshift Query Editor interface. On the left, a sidebar contains navigation links for 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor' (highlighted), 'Query editor v2', 'Queries and loads', 'Datashares', 'Configurations', 'Advisor', 'AWS Marketplace', 'Alarms', 'Events', and 'What's new'. The main workspace is divided into several sections. At the top, there are tabs for 'Editor', 'Query history', 'Saved queries', and 'Scheduled queries'. Below these, a 'Resources' panel shows 'Select database' (set to 'etl') and 'Select schema' (set to 'spar\_nord\_atm\_trans'). The central area contains a single SQL query: `1 create schema spar_nord_atm_trans;`. Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. At the bottom, the 'Query results' section shows a status of 'Completed, started on September 03, 2022 at 14:46:13' and 'ELAPSED TIME: 00 m 07 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

## 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));
```



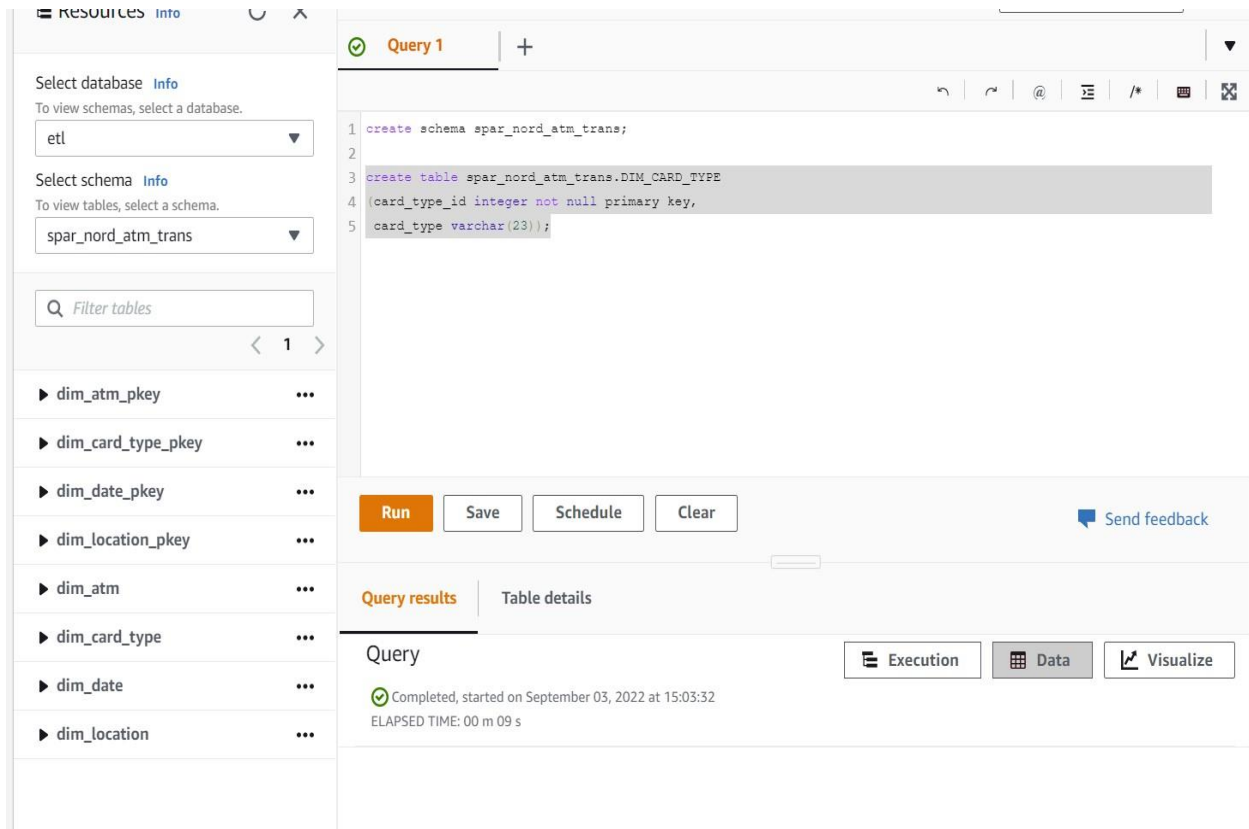
The screenshot shows a database query editor interface. On the left, there is a sidebar with a tree view showing the database structure. The selected database is 'etl' and the selected schema is 'spar\_nord\_atm\_trans'. The tree view lists several tables: 'dim\_atm\_pkey', 'dim\_location\_pkey', 'dim\_atm', and 'dim\_location'. The main area displays a SQL query for creating the 'DIM\_ATM' table. The query is as follows:

```
1 create schema spar_nord_atm_trans;
2
3 create table spar_nord_atm_trans.DIM_ATM
4 (atm_id integer not null primary key,
5  atm_number varchar(20),
6  atm_manufacturer varchar(50),
7  atm_location_id integer not null,
8  foreign key(atm_location_id) references spar_nord_atm_trans.DIM_LOCATION(location_id));
```

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 September 03, 2022, at 15:00:45, with an elapsed time of 01 m 02 s. There are also buttons for 'Execution', 'Data', and 'Visualize'.

## 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));***



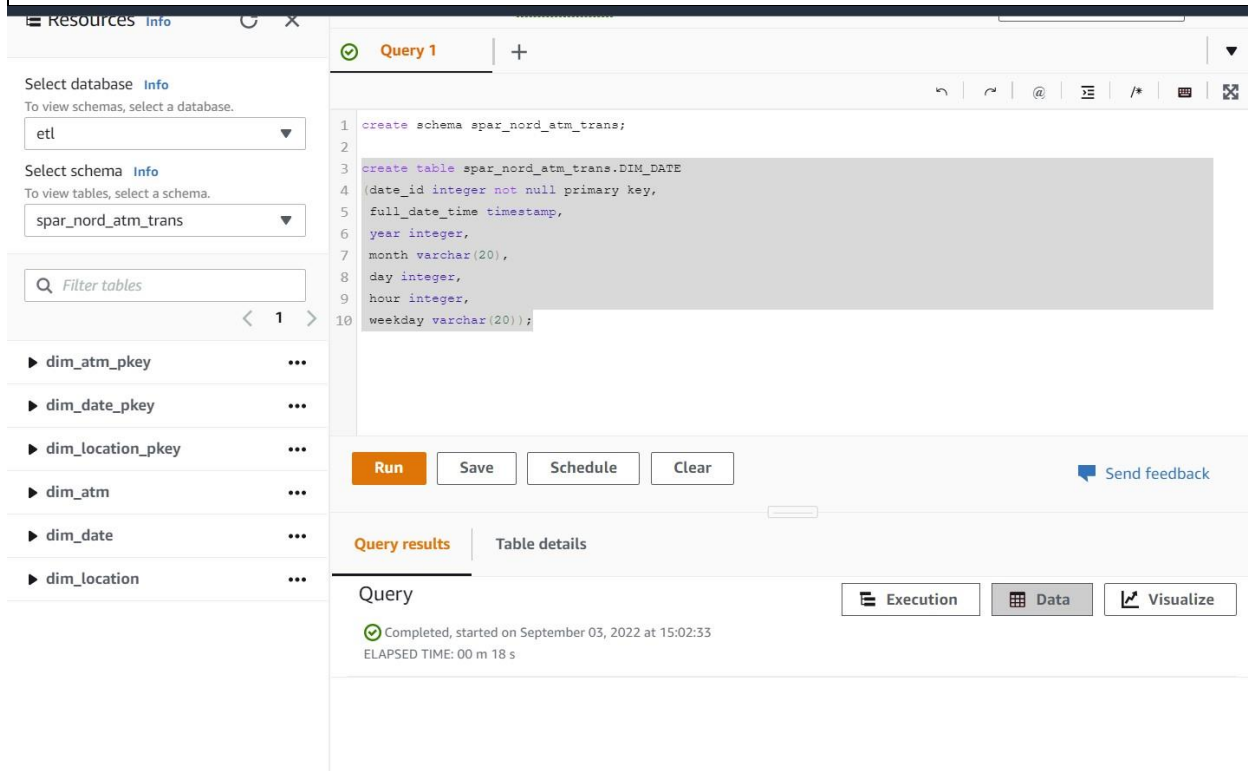
The screenshot shows a database management tool interface. On the left, there is a sidebar with a 'Resources' tab. Under 'Select database', 'etl' is selected. Under 'Select schema', 'spar\_nord\_atm\_trans' is selected. Below this is a search bar labeled 'Filter tables' and a list of tables: dim\_atm\_pkey, dim\_card\_type\_pkey, dim\_date\_pkey, dim\_location\_pkey, dim\_atm, dim\_card\_type, dim\_date, and dim\_location. The main area is titled 'Query 1' and contains the following SQL code:

```
1 create schema spar_nord_atm_trans;
2
3 create table spar_nord_atm_trans.DIM_CARD_TYPE
4 (card_type_id integer not null primary key,
5  card_type varchar(23));
```

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 are tabs for 'Query results' and 'Table details'. The 'Query results' tab is active, showing a status message: 'Completed, started on September 03, 2022 at 15:03:32' and 'ELAPSED TIME: 00 m 09 s'. To the right of the status message are buttons for 'Execution', 'Data', and 'Visualize'.

## 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));
```



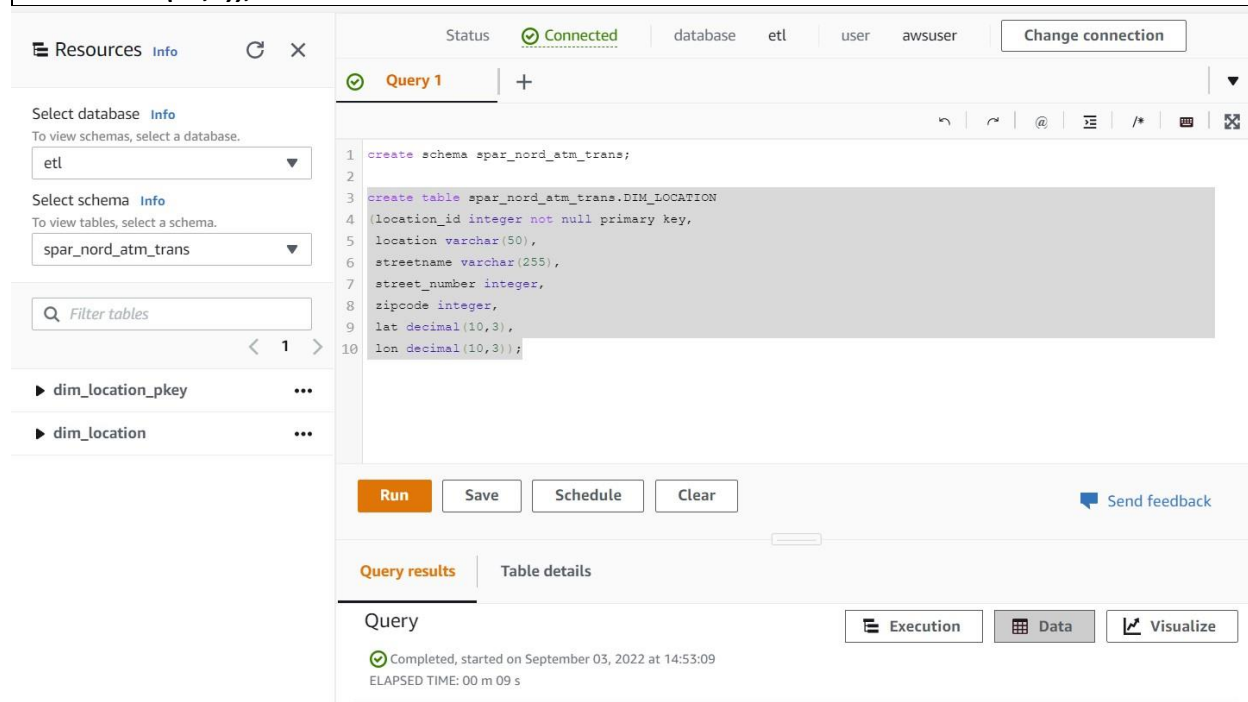
The screenshot shows a SQL IDE interface with the following components:

- Left Panel:**
  - Select database:** A dropdown menu showing 'etl'.
  - Select schema:** A dropdown menu showing 'spar\_nord\_atm\_trans'.
  - Filter tables:** A search bar with the text 'Filter tables'.
  - Table List:** A list of tables including 'dim\_atm\_pkey', 'dim\_date\_pkey', 'dim\_location\_pkey', 'dim\_atm', 'dim\_date', and 'dim\_location'.
- Query Editor:**
  - Query 1:** The SQL query is displayed in a text area, showing the creation of the 'spar\_nord\_atm\_trans.DIM\_DATE' table with columns: 'date\_id' (integer, primary key), 'full\_date\_time' (timestamp), 'year' (integer), 'month' (varchar(20)), 'day' (integer), 'hour' (integer), and 'weekday' (varchar(20)).
  - Buttons:** 'Run', 'Save', 'Schedule', and 'Clear' buttons are located below the query editor.
- Query Results Panel:**
  - Query results:** The 'Query results' tab is selected, showing the execution status: 'Completed, started on September 03, 2022 at 15:02:33' and 'ELAPSED TIME: 00 m 18 s'.
  - Table details:** The 'Table details' tab is also visible.
  - Execution Options:** 'Execution', 'Data', and 'Visualize' buttons are located at the bottom of the results panel.



## 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));
```



The screenshot shows a SQL IDE interface with the following components:

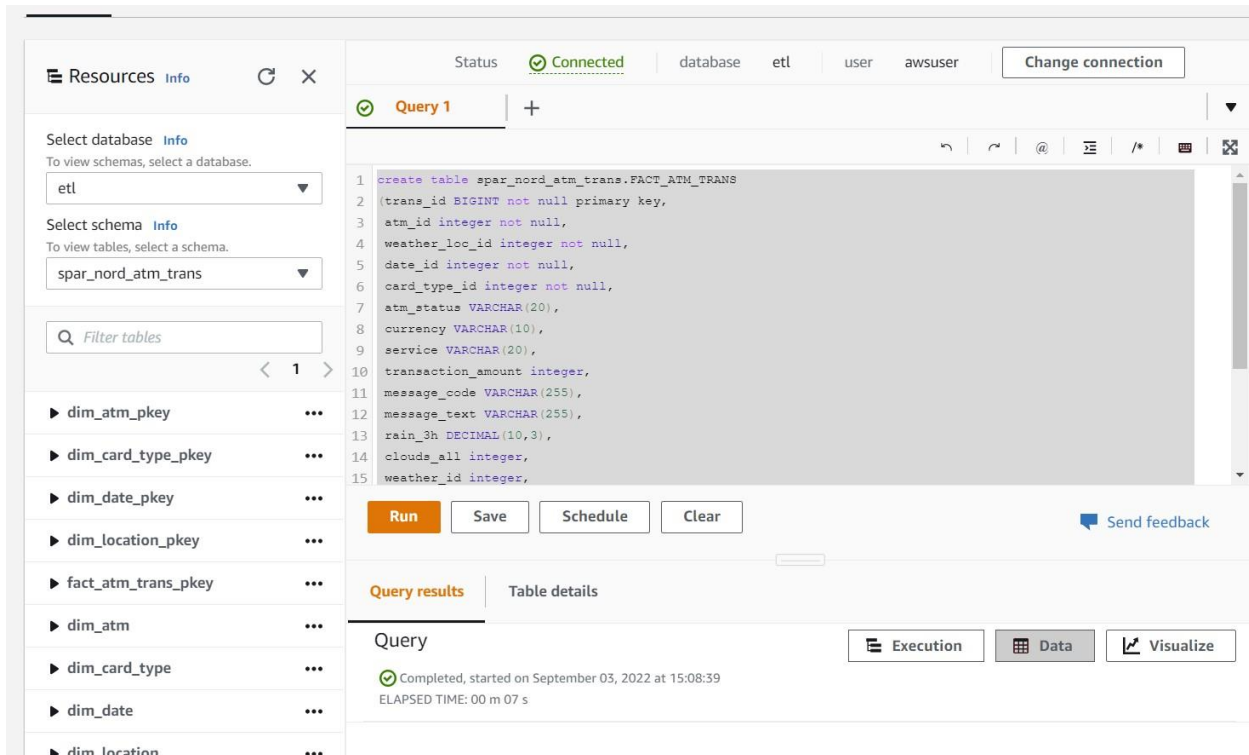
- Resources Panel (Left):**
  - Select database:** etl
  - Select schema:** spar\_nord\_atm\_trans
  - Filter tables:** Search bar with "Filter tables" text.
  - Table list:**
    - dim\_location\_pkey
    - dim\_location
- Query Editor (Center):**
  - Status:** Connected
  - Database:** etl
  - User:** awsuser
  - Change connection:** Button
  - Query 1:**

```
1 create schema spar_nord_atm_trans;
2
3 create table spar_nord_atm_trans.DIM_LOCATION
4 (location_id integer not null primary key,
5  location varchar(50),
6  streetname varchar(255),
7  street_number integer,
8  zipcode integer,
9  lat decimal(10,3),
10 lon decimal(10,3));
```
- Execution Controls (Bottom):**
  - Run:** Button
  - Save:** Button
  - Schedule:** Button
  - Clear:** Button
  - Send feedback:** Link
- Query Results (Bottom):**
  - Query:**
    - Execution:** Completed, started on September 03, 2022 at 14:53:09
    - ELAPSED TIME:** 00 m 09 s
  - Data:** Button
  - Visualize:** Button

## 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));
```



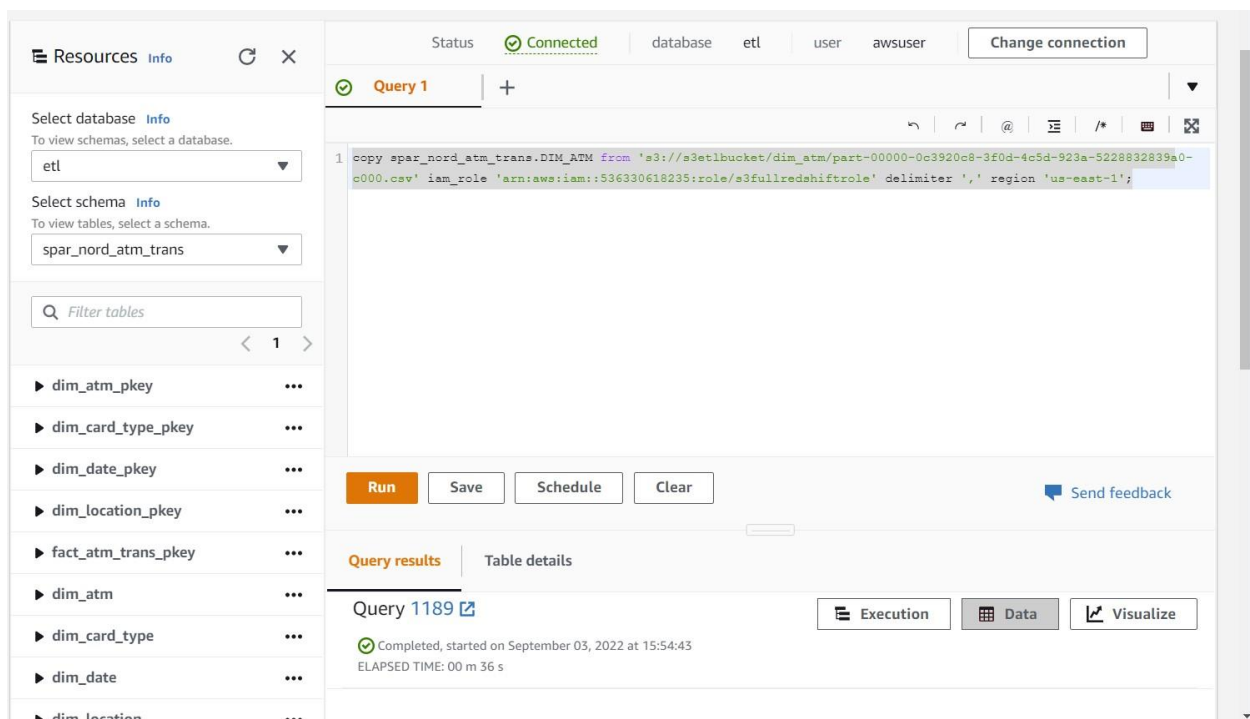
The screenshot displays a database management tool interface. On the left, a sidebar shows a list of resources under the 'etl' database, including various dimension tables like 'dim\_atm\_pkey', 'dim\_card\_type\_pkey', 'dim\_date\_pkey', 'dim\_location\_pkey', 'fact\_atm\_trans\_pkey', 'dim\_atm', 'dim\_card\_type', 'dim\_date', and 'dim\_location'. The main area shows a SQL query editor with a query that creates a table 'spar\_nord\_atm\_trans.FACT\_ATM\_TRANS' with various columns and constraints. Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The bottom section shows the 'Query results' tab, indicating that the query was completed successfully on September 03, 2022, at 15:08:39, with an elapsed time of 00 m 07 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

S3 to Redshift DIM\_ATM

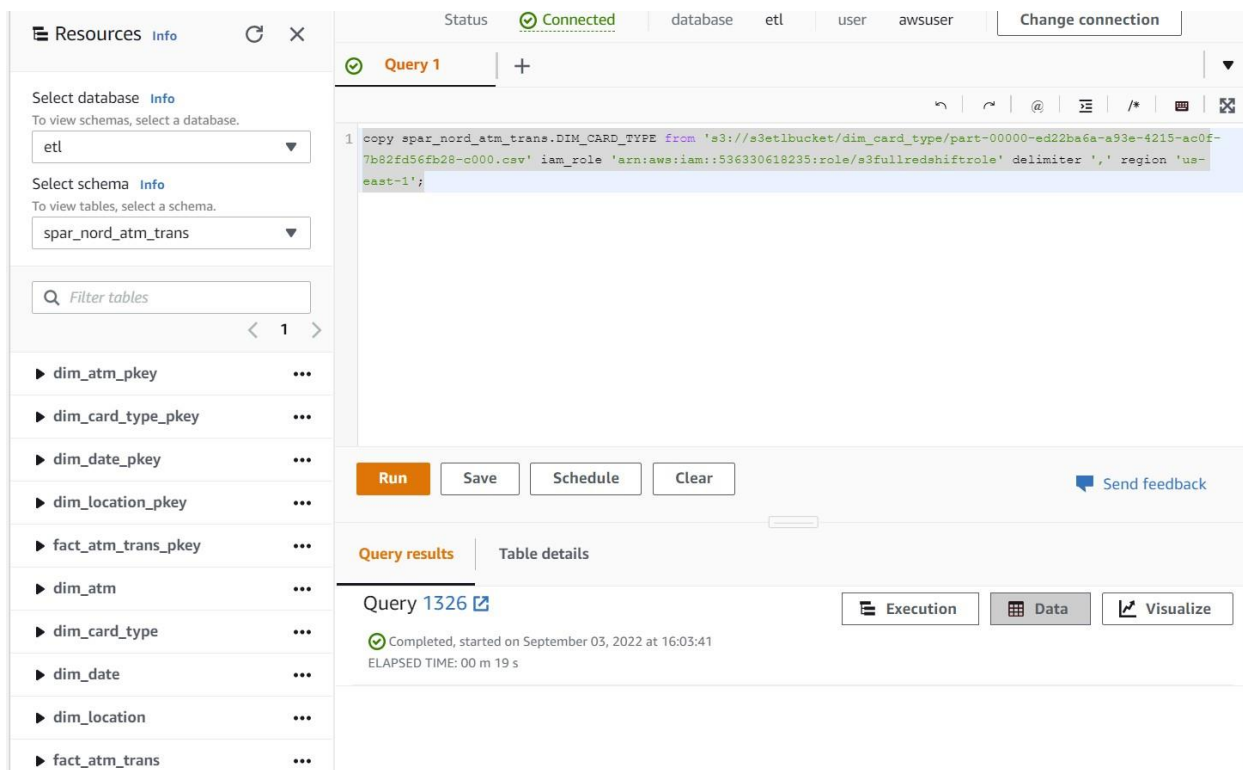
```
copy spar_nord_atm_trans.DIM_ATM from 's3://s3etlbucket/dim_atm/part-00000-f38f2738-f84e-4581-af21-6eaebeedef11-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1';
```



The screenshot displays the AWS Redshift console interface. On the left, the 'Resources' panel shows the selected database 'etl' and schema 'spar\_nord\_atm\_trans'. The main area shows a SQL query for copying data from an S3 bucket to a Redshift table. The query is: `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';`. Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing 'Query 1189' with a status of 'Completed' and an elapsed time of '00 m 36 s'. The 'Table details' tab is also visible.

## 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';
```



The screenshot shows the AWS Redshift console interface. On the left, the 'Resources' panel displays a list of databases and schemas. The 'etl' database is selected, and the 'spar\_nord\_atm\_trans' schema is chosen. Below this, a list of tables is shown, 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 panel shows a query editor with the following SQL query:

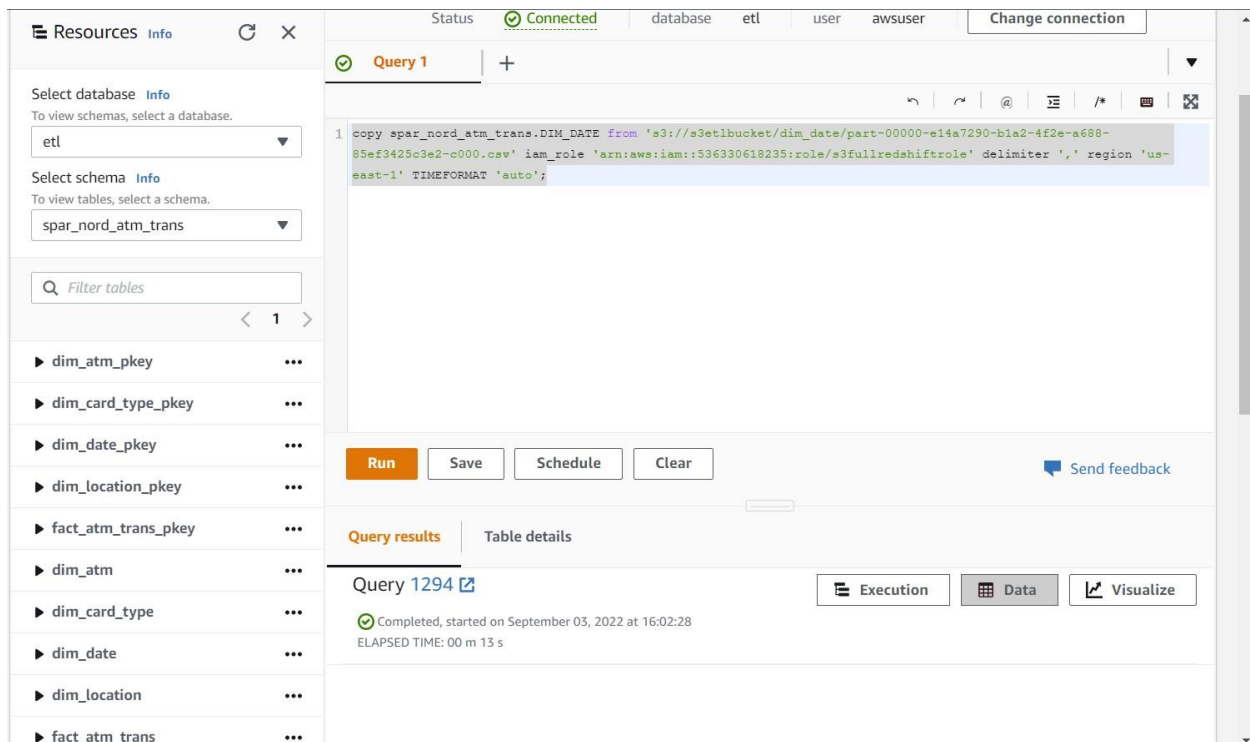
```
1 copy spar_nord_atm_trans.DIM_CARD_TYPE from 's3://s3etlbucket/dim_card_type/part-00000-ed22ba6a-a93e-4215-ac0f-7b82fd56fb28-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1';
```

Below the query editor, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Run' button is highlighted in orange. To the right of these buttons is a 'Send feedback' link.

The bottom section of the console shows the 'Query results' tab. It indicates that the query was completed successfully on September 03, 2022, at 16:03:41, with an elapsed time of 00 m 19 s. There are also tabs for 'Execution', 'Data', and 'Visualize'.

## 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';
```



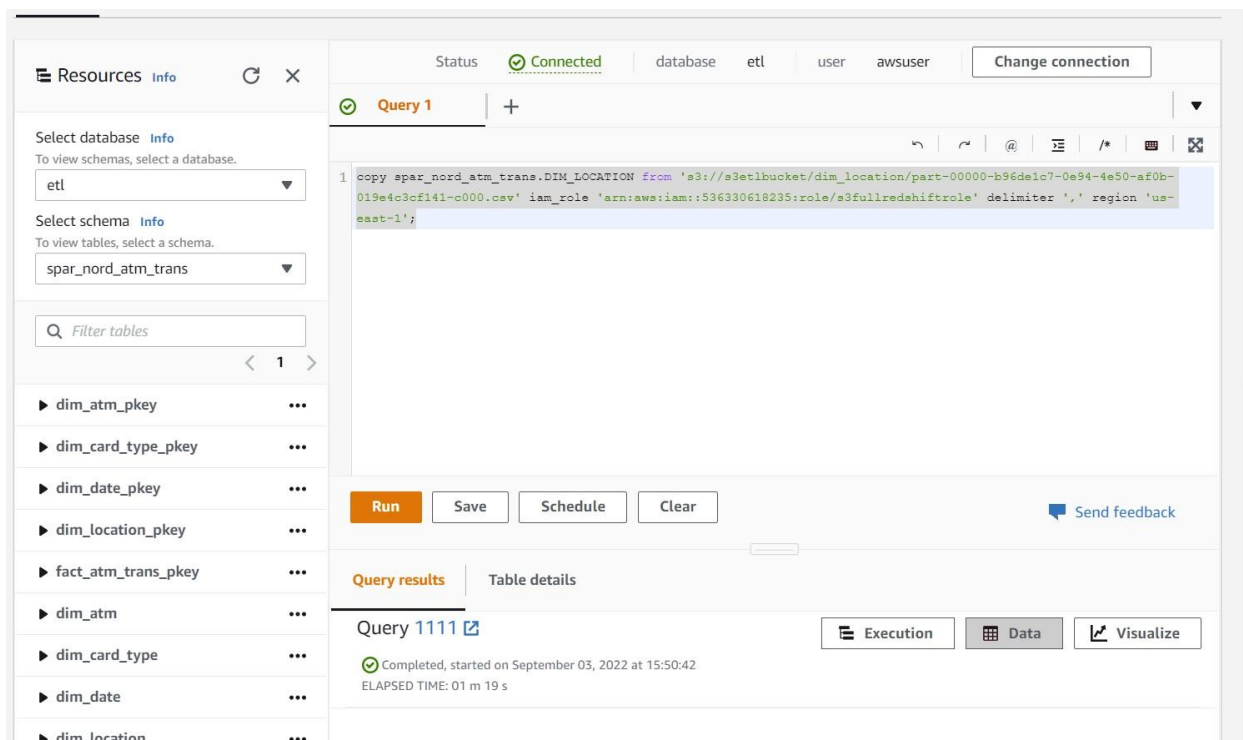
The screenshot shows the AWS Redshift console interface. On the left, the 'Resources' panel is expanded, showing the 'etl' database and 'spar\_nord\_atm\_trans' schema. A list of tables is visible, including 'dim\_atm\_pkey', 'dim\_card\_type\_pkey', 'dim\_date\_pkey', 'dim\_location\_pkey', 'fact\_atm\_trans\_pkey', 'dim\_atm', 'dim\_card\_type', 'dim\_date', 'dim\_location', and 'fact\_atm\_trans'. The main area displays 'Query 1' with the following SQL statement:

```
1 copy spar_nord_atm_trans.DIM_DATE from 's3://s3etlbucket/dim_date/part-00000-e14e7290-b1a2-4f2e-a688-85ef3425c3e2-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1' TIMEFORMAT 'auto';
```

Below the query, 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 area, the 'Query results' tab is selected, showing 'Query 1294'. The status indicates 'Completed, started on September 03, 2022 at 16:02:28' with an 'ELAPSED TIME: 00 m 13 s'. There are also buttons for 'Execution', 'Data', and 'Visualize'.

## 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';
```



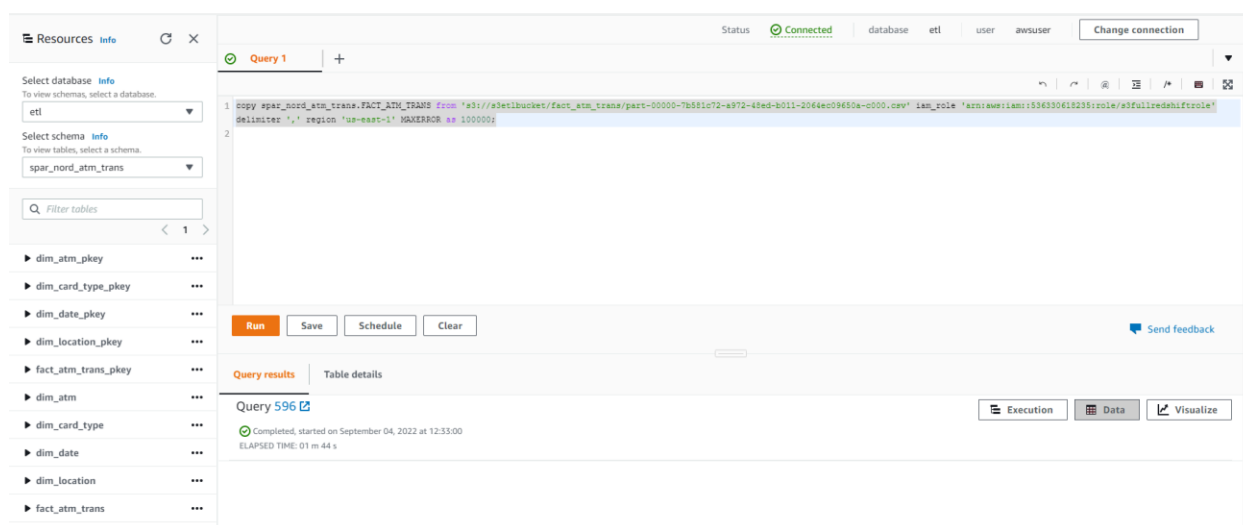
The screenshot displays the AWS Redshift console interface for executing a query. On the left, the 'Resources' panel shows the database 'etl' and schema 'spar\_nord\_atm\_trans'. The main query editor contains the following SQL statement:

```
1 copy spar_nord_atm_trans.DIM_LOCATION from 's3://s3etlbucket/dim_location/part-00000-b96de1c7-0e94-4e50-af0b-019e4c3cf141-c000.csv' iam_role 'arn:aws:iam::536330618235:role/s3fullredshiftrole' delimiter ',' region 'us-east-1';
```

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 buttons, the 'Query results' tab is active, showing the query ID 'Query 1111' and its execution status: 'Completed, started on September 03, 2022 at 15:50:42' with an 'ELAPSED TIME: 01 m 19 s'. At the bottom right, there are tabs for 'Execution', 'Data', and 'Visualize'.

## 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;***



The screenshot displays the AWS Redshift console interface. On the left, the 'Resources' panel shows the selected database 'etl' and schema 'spar\_nord\_atm\_trans'. The main area shows a SQL query for 'Query 1':

```
1 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'
2 delimiter ',' region 'us-east-1' MAXERROR 1000000
```

Below the query, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear'. The 'Query results' tab is active, showing the query ID 'Query 596' and its status: 'Completed, started on September 04, 2022 at 12:33:00' with an 'ELAPSED TIME: 01 m 44 s'. At the bottom right, there are tabs for 'Execution', 'Data', and 'Visualize'.