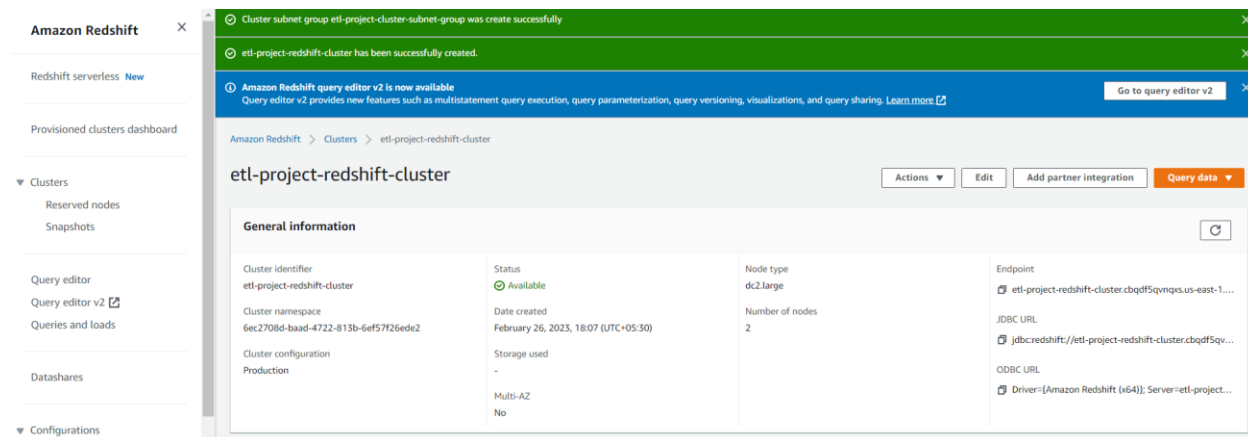


## Creation of a Redshift Cluster

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

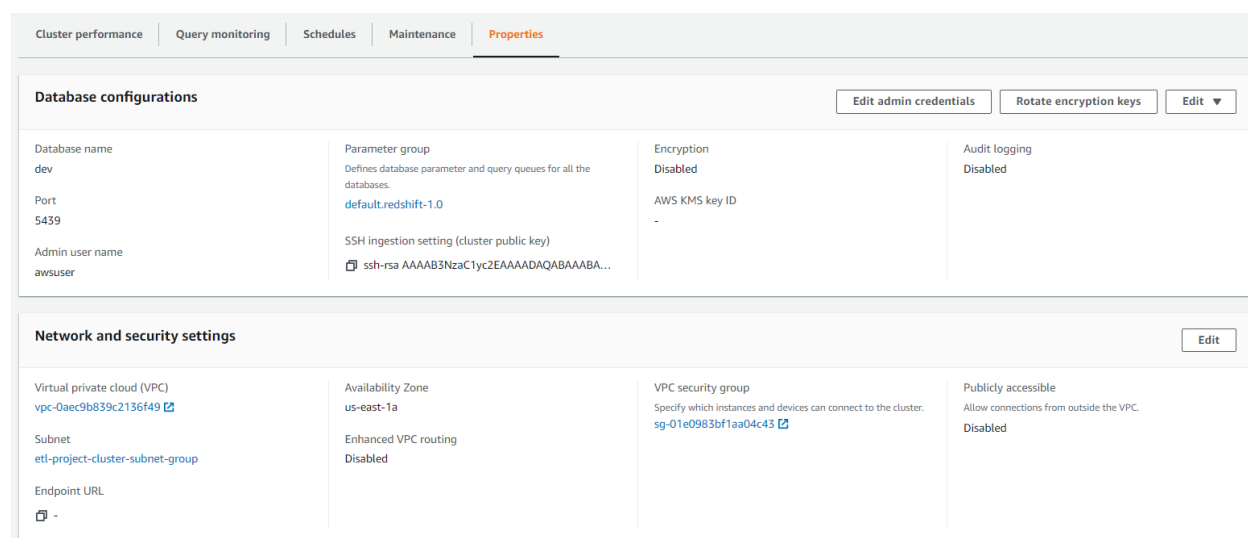
Screenshot of type of machine used along with number of nodes:



The screenshot shows the Amazon Redshift console interface. On the left is a navigation menu with options like 'Redshift serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Queries and loads', 'Dashboards', and 'Configurations'. The main panel displays the details for the cluster 'etl-project-redshift-cluster'. At the top, there are three status messages: 'Cluster subnet group etl-project-cluster-subnet-group was create successfully', 'etl-project-redshift-cluster has been successfully created.', and 'Amazon Redshift query editor v2 is now available'. Below these, the cluster name 'etl-project-redshift-cluster' is shown with buttons for 'Actions', 'Edit', 'Add partner integration', and 'Query data'. A 'General information' section contains a table with the following data:

Cluster identifier	Status	Node type	Endpoint
etl-project-redshift-cluster	Available	dc2.large	etl-project-redshift-cluster.cbqdf5vnmqs.us-east-1...
Cluster namespace	Date created	Number of nodes	JDBC URL
6ec2708d-baad-4722-813b-6ef57f26ede2	February 26, 2023, 18:07 (UTC+05:30)	2	jdbcredshift://etl-project-redshift-cluster.cbqdf5v...
Cluster configuration	Storage used		ODBC URL
Production	-		Driver=(Amazon Redshift (v64)); Server=etl-project...
	Multi-AZ		
	No		

Screenshots of various configurations associated with cluster creation:



The screenshot shows the 'Properties' tab of the Amazon Redshift console for the cluster 'etl-project-redshift-cluster'. It is divided into two main sections: 'Database configurations' and 'Network and security settings'.

**Database configurations:**

Database name	Parameter group	Encryption	Audit logging
dev	default.redshift-1.0	Disabled	Disabled
Port	SSH ingestion setting (cluster public key)	AWS KMS key ID	
5439	ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQ...	-	
Admin user name			
awsuser			

**Network and security settings:**

Virtual private cloud (VPC)	Availability Zone	VPC security group	Publicly accessible
vpc-0aec9b839c2136f49	us-east-1a	sg-01e0983bf1aa04c43	Allow connections from outside the VPC.
Subnet	Enhanced VPC routing		Disabled
etl-project-cluster-subnet-group	Disabled		
Endpoint URL			
-			

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](#)

Set default
Manage IAM roles

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.

< 1 >

<input type="checkbox"/>	IAM roles	Status	Role type
<input type="checkbox"/>	<a href="#">redshift_s3_fullaccess</a>	<span>in-sync</span>	--

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

Query to create schema for dimension and fact tables

create schema atm\_data;

Status
Connected

database
dev
user
awsuser

Change connection

Query 1
+

1

create schema atm\_data;

Run
Save
Schedule
Clear

Send feedback

Query results
Table details

Query

Completed, started on February 26, 2023 at 18:21:04

ELAPSED TIME: 01 m 02 s

Execution
Data
Visualize

## Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

### i) Creating Location dimension table

```
create table 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)
);
```

```
3 create table atm_data.DIM_LOCATION
4 (
5 location_id int not null DISTKEY SORTKEY,
6 location varchar(50),
7 streetname varchar(255),
8 street_number int,
9 zipcode int,
10 lat decimal(10,3),
11 lon decimal(10,3),
12 PRIMARY KEY(location_id)
13 );
```

[Run](#)[Save](#)[Schedule](#)[Clear](#)[Send feedback](#)[Query results](#)[Table details](#)**Query**[Execution](#)[Data](#)[Visualize](#)

✓ Completed, started on February 26, 2023 at 18:22:59  
ELAPSED TIME: 00 m 03 s

### ii) Creating ATM dimension table

```
create table 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 atm_data.DIM_LOCATION(location_id)
);
```

Query 1

+

↶

↷

@

≡

/

\*

SQL

✕

```
1 create table atm_data.DIM_ATM
2 (
3   atm_id int not null DISTKEY SORTKEY,
4   atm_number varchar(20),
5   atm_manufacturer varchar(50),
6   atm_location_id int,
7   PRIMARY KEY(atm_id),
8   FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)
9 );
```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query

Execution

Data

Visualize

Completed, started on February 26, 2023 at 18:24:12

ELAPSED TIME: 00 m 03 s

### iii) Creating Date dimension table

```
create table 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)
);
```

Query 1
+

```

1 create table 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 );

```

Run
Save
Schedule
Clear
Send feedback

Query results
Table details

Query

Execution
Data
Visualize

Completed, started on February 26, 2023 at 18:25:11  
 ELAPSED TIME: 00 m 03 s

#### iv) Creating Card type dimension table

```

create table atm_data.DIM_CARD_TYPE
(
  card_type_id int not null DISTKEY SORTKEY,
  card_type varchar(30),
  PRIMARY KEY(card_type_id)
);

```

Query 1
+

```

1 create table 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 );

```

Run
Save
Schedule
Clear
Send feedback

Query results
Table details

Query

Execution
Data
Visualize

Completed, started on February 26, 2023 at 18:26:05  
 ELAPSED TIME: 00 m 03 s

## v) Creating ATM transactions fact table

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

Query 1
+

↶
↷
@
≡
/
\*
⌂
⌕

```

1 create table 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,
14        3),
15    clouds_all int,
16    weather_id int,
17    weather_main varchar(50),

```

Run
Save
Schedule
Clear
Send feedback

Query results
Table details

Query
Execution
Data
Visualize

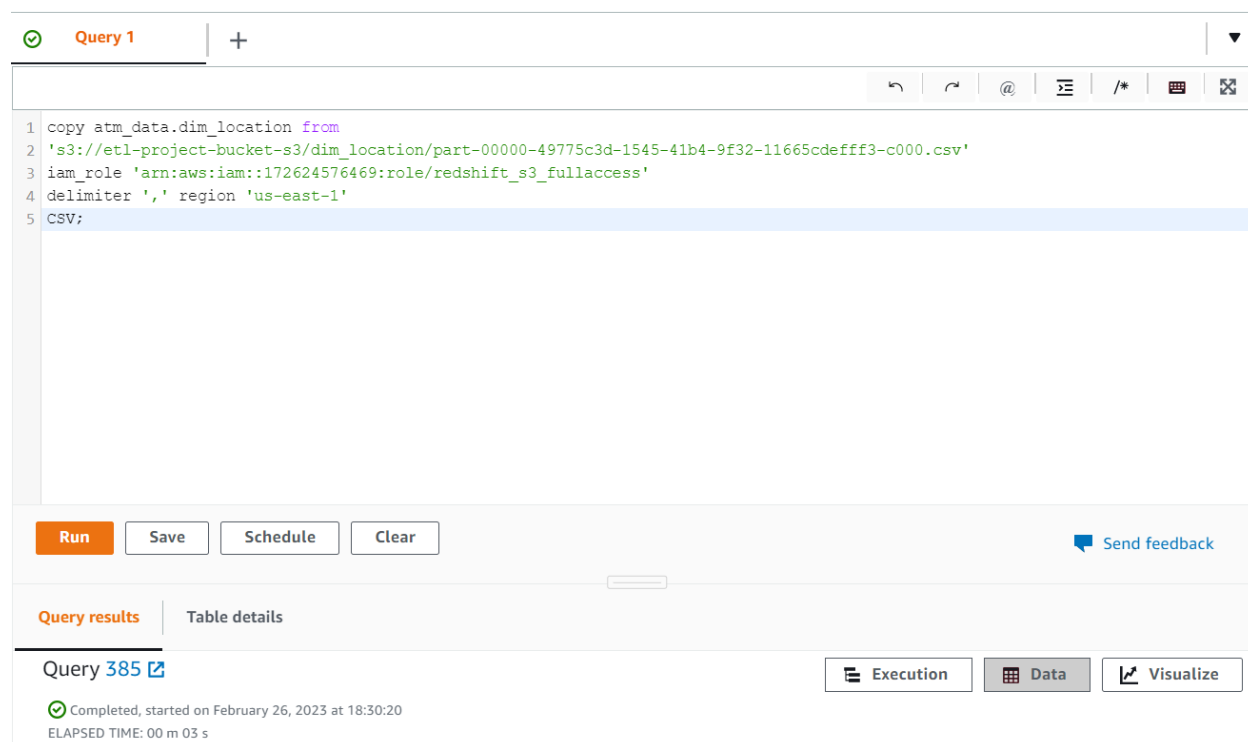
Completed, started on February 26, 2023 at 18:27:38
ELAPSED TIME: 00 m 28 s

## 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

#### i) Copying data to dim\_location table

```
copy atm_data.dim_location from 's3://etl-project-bucket-s3/dim_location/part-00000-49775c3d-1545-41b4-9f32-11665cdefff3-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```



The screenshot shows the Amazon Redshift console interface. At the top, there's a tab labeled "Query 1" with a plus icon to add more queries. Below the tab is a toolbar with icons for undo, redo, copy, paste, and other editing functions. The main area contains a SQL query:
 

```
1 copy atm_data.dim_location from
2 's3://etl-project-bucket-s3/dim_location/part-00000-49775c3d-1545-41b4-9f32-11665cdefff3-c000.csv'
3 iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
4 delimiter ',' region 'us-east-1'
5 CSV;
```

 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 buttons is a tabbed interface with "Query results" selected and "Table details" as an option. Under "Query results", it shows "Query 385" with a link icon. Below that, it indicates the query is "Completed, started on February 26, 2023 at 18:30:20" with an "ELAPSED TIME: 00 m 03 s". At the bottom right, there are three buttons: "Execution", "Data", and "Visualize".

#### ii) Copying data to dim\_atm table

```
copy atm_data.dim_atm from 's3://etl-project-bucket-s3/dim_atm/part-00000-d66c2da4-b45f-4e62-b4d6-e9b8ffc92a00-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Query 1

+

↶

↷

@

≡

/

\*

📄

✖

```

1 copy atm_data.dim_atm from 's3://etl-project-bucket-s3/dim_atm/part-00000-d66c2da4-b45f-4e62-b4d6-e9b8ffc92a00-c000.csv'
2 iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV;

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 436

Execution

Data

Visualize

Completed, started on February 26, 2023 at 18:32:15

ELAPSED TIME: 00 m 03 s

### iii) Copying data to dim\_date table

```

copy atm_data.dim_date from 's3://etl-project-bucket-s3/dim_date/part-00000-63a9a165-d1c5-4466-b875-068a22bd18dd-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV
TIMEFORMAT 'auto';

```

Query 1

+

↶

↷

@

≡

/

\*

📄

✖

```

1 copy atm_data.dim_date from 's3://etl-project-bucket-s3/dim_date/part-00000-63a9a165-d1c5-4466-b875-068a22bd18dd-c000.csv'
2 iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV
5 TIMEFORMAT 'auto';

```

Run

Save

Schedule

Clear

Send feedback

Query results

Table details

Query 469

Execution

Data

Visualize

Completed, started on February 26, 2023 at 18:33:55

ELAPSED TIME: 00 m 03 s



#### iv) Copying data to dim\_card\_type table

```
copy atm_data.dim_card_type from 's3://etl-project-bucket-s3/dim_card_type/part-00000-4c102809-4c9a-4709-ada4-3c6ea05eecb8-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

Query 1

+

↶

↷

@

≡

/\*

📄

🔍

```
1 copy atm_data.dim_card_type from 's3://etl-project-bucket-s3/dim_card_type/part-00000-4c102809-4c9a-4709-ada4-3c6ea05eecb8-c000.csv'
2 iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
3 delimiter ',' region 'us-east-1'
4 CSV;
```

Run

Save

Schedule

Clear

💬 Send feedback

Query results

Table details

Query 497

Completed, started on February 26, 2023 at 18:35:28  
ELAPSED TIME: 00 m 03 s

📄 Execution

📊 Data

📈 Visualize

### v) Copying data to fact\_atm\_trans table

```
copy atm_data.fact_atm_trans from 's3://etl-project-bucket-s3/fact_atm_trans/part-00000-6e6dd7a3-cb78-48d4-b22d-104e04374072-c000.csv'
iam_role 'arn:aws:iam::172624576469:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1'
CSV;
```

✔ Query 1

+

▼

↶ ↷ @ ⌵ /\* 📄 🗑

1 copy atm\_data.fact\_atm\_trans from 's3://etl-project-bucket-s3/fact\_atm\_trans/part-00000-6e6dd7a3-cb78-48d4-b22d-104e04374072-c000.csv'  
2 iam\_role 'arn:aws:iam::172624576469:role/redshift\_s3\_fullaccess'  
3 delimiter ',' region 'us-east-1'  
4 CSV;

Run Save Schedule Clear

Send feedback

Query results Table details

Query 534

Execution Data Visualize

✔ Completed, started on February 26, 2023 at 18:37:08  
ELAPSED TIME: 00 m 14 s