

# Creation of Redshift Cluster and Loading Data from S3 after creating tables

## Creation of a RedShift Cluster

**Screenshots of the configuration of the RedShift cluster that you have created:**

A redshift cluster with the name 'redshift-cluster-1' was created for the purpose of the assignment, Node used was dc2.large with Number of Nodes = 2

Amazon Redshift > Clusters > redshift-cluster-1

redshift-cluster-1

Actions

Edit

Add partner integration

Query cluster

General information

Cluster identifier

redshift-cluster-1

Status

Available

Node type

dc2.large

Endpoint

redshift-cluster-1.cnei54at27p1.us-east-1.r...

Cluster namespace

eb172a03-f02b-4907-b047-c0ab3aa94cd5

Date created

April 04, 2021, 02:18(UTC+05:30)

Number of nodes

2

JDBC URL

jdbc:redshift://redshift-cluster-1.cnei54at2...

Storage used

-

ODBC URL

Driver={Amazon Redshift (x64)}; Server=red...

Database configurations

Change master user password

Rotate encryption keys

Edit

Database name

dev

Parameter group

Defines database parameter and query queues for all the databases.

default.redshift-1.0

Encryption

Disabled

Audit logging

Disabled

Port

5439

AWS KMS key ID

-

Master user name

awsuser

SSH ingestion setting (cluster public key)

ssh-rsa AAAAB3NzaC1yc2EAAAADAQAB...

Before creating the cluster the Subnet, VPN and Security Group were chosen to be the ones created as part of the project. Along with that IAM role as below has also been attached to the cluster

**Network and security settings**

Edit publicly accessible

Edit

Virtual private cloud (VPC) <a href="#">vpc-05e2eaf9c9f8a4fb7</a>	Availability Zone us-east-1e	VPC security group Specify which instances and devices can connect to the cluster. <a href="#">sg-0de129eb2cf5ceaeab</a>	Publicly accessible Allow instances and devices outside the VPC to connect to your database through the cluster endpoint. Disabled
Subnet <a href="#">cluster-subnet-group-1</a>	Enhanced VPC routing Disabled		
Endpoint URL -			

**Cluster permissions (1)**  
Your cluster needs permissions to access other AWS services on your behalf. For the required permissions, add IAM roles with the principal "redshift.amazonaws.com". You can associate up to 10 IAM roles with this cluster. [Learn more](#)

Manage IAM roles

Associated IAM roles	Status	Amazon Resource Name (ARN)
<a href="#">redshift_S3_fullaccess</a>	in-sync	arn:aws:iam::894365351309:role/redshift_S3_fullaccess

## 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 schema first:**

```
create schema myetlproject;
```

**Creating DIM\_DATE:**

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

```
1 create table myetlproject.DIM_DATE(  
2 date_id integer not null,  
3 full_date_time timestamp ,  
4 year integer ,  
5 month varchar(20) ,  
6 day integer ,  
7 hour integer ,  
8 weekday varchar(20) ,  
9 primary key(date_id));  
10  
11  
12  
13  
14  
15
```

**Run**

Save

Schedule

Clear

### Creating DIM\_CARD\_TYPE

```
create table myetlproject.dim_card_type(  
card_type_id integer not null distkey sortkey,  
card_type varchar(20),  
PRIMARY KEY(card_type_id));
```

```
create table myetlproject.dim_card_type(  
card_type_id integer not null distkey sortkey,  
card_type varchar(20),  
PRIMARY KEY(card_type_id));
```

**Run**

Save

Schedule

Clear

**Query results**

Table details

#### Query



Completed, started on April 04, 2021 at 16:01:24

ELAPSED TIME: 03 m 03 s

### Creating DIM\_ATM:

```
create table myetlproject.dim_atm(  
atm_id integer not null distkey sortkey,  
atm_number varchar(20),  
atm_manufacturer varchar(50),  
atm_location_id integer,
```

PRIMARY KEY(atm\_id),  
FOREIGN KEY(atm\_location\_id) REFERENCES myetlproject.dim\_location(location\_id));

```
create table myetlproject.dim_atm(  
  atm_id integer not null distkey sortkey,  
  atm_number varchar(20),  
  atm_manufacturer varchar(50),  
  atm_location_id integer,  
  PRIMARY KEY(atm_id),  
  FOREIGN KEY(atm_location_id) REFERENCES myetlproject.dim_location(location_id));
```

Run

Save

Schedule

Clear

Query results

Table details

Query

Completed, started on April 04, 2021 at 16:00:45

ELAPSED TIME: 04 m 27 s

## Create DIM\_LOCATION

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

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

[Run](#)[Save](#)[Schedule](#)[Clear](#)[Query results](#)[Table details](#)

### Query



Completed, started on April 04, 2021 at 15:59:58

ELAPSED TIME: 06 m 15 s

### Create FACT\_ATM\_TRANS:

```
create table myetlproject.fact_atm_trans(  
  trans_id bigint not null distkey sortkey,  
  atm_id integer,  
  weather_loc_id integer,  
  date_id integer,  
  card_type_id integer,  
  atm_status varchar(20),  
  clouds_all integer,  
  weather_id integer,  
  weather_main varchar(50),  
  weather_description varchar(255),  
  transaction_amount integer,
```

```
message_code varchar(255),
message_text varchar(255),
currency varchar(10),
service varchar(20),
PRIMARY KEY(trans_id),
FOREIGN KEY(weather_loc_id) REFERENCES myetlproject.dim_location(location_id),
FOREIGN KEY(atm_id) REFERENCES myetlproject.dim_atm(atm_id),
FOREIGN KEY(date_id) REFERENCES myetlproject.dim_date(date_id),
FOREIGN KEY(card_type_id) REFERENCES myetlproject.dim_card_type(card_type_id));
```

```
create table myetlproject.fact_atm_trans(
trans_id bigint not null distkey sortkey,
atm_id integer,
weather_loc_id integer,
date_id integer,
card_type_id integer,
atm_status varchar(20),
clouds_all integer,
weather_id integer,
weather_main varchar(50),
weather_description varchar(255),
transaction_amount integer,
message_code varchar(255),
message_text varchar(255),
currency varchar(10),
service varchar(20),
PRIMARY KEY(trans_id),
FOREIGN KEY(weather_loc_id) REFERENCES myetlproject.dim_location(location_id),
FOREIGN KEY(atm_id) REFERENCES myetlproject.dim_atm(atm_id),
FOREIGN KEY(date_id) REFERENCES myetlproject.dim_date(date_id),
FOREIGN KEY(card_type_id) REFERENCES myetlproject.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

Loading into DIM\_DATE:

```
copy myetlproject.dim_date from
's3://myetlproject/DIM_DATE/part-00000-47d5705b-a5cf-4d46-9006-ad7db2d01995-c000.csv'
iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'
delimiter ',' timeformat 'YYYY-MM-DDTHH:MI:SS' region 'us-east-1';

select count(*) from myetlproject.dim_date;
```

Run

Save

Schedule

Clear



**Loading into DIM\_CARD\_TYPE:**

```
1 copy myetlproject.dim_card_type from
2 's3://myetlproject/DIM_CARD_TYPE/part-00000-a6d78b0d-e375-4620-8497-3c295ab68350-c000.csv'
3 iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'
4 delimiter ',' region 'us-east-1';
5
6
7
8 select count(*) from myetlproject.dim_card_type;
9
10
11
12
13
14
15
```

**Run** **Save** **Schedule** **Clear**

**Loading into DIM\_LOCATION:**

```
copy myetlproject.dim_location from
's3://myetlproject/DIM_LOCATION/part-00000-5e89fc61-e5f4-4067-a135-3b3f265bc219-c000.csv'
iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'
delimiter ',' region 'us-east-1';

select count(*) from myetlproject.dim_location;
```

**Run** **Save** **Schedule** **Clear**

**Loading into DIM\_ATM:**

```
copy myetlproject.dim_atm from  
's3://myetlproject/DIM_ATM/part-00000-15edf02f-522f-4acd-b6e3-d6ddb719eb40-c000.csv'  
iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'  
delimiter ',' region 'us-east-1';
```

```
select count(*) from myetlproject.dim_atm;
```

**Run**

Save

Schedule

Clear

**Loading into FACT\_ATM\_TRANS:**

```
copy myetlproject.fact_atm_trans from  
's3://myetlproject/FACT_ATM_TYPE/part-00000-c1c1772e-b3d5-4d24-83d8-51ee8d616d24-  
c000.csv'  
iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'  
delimiter ','  
emptyasnull  
blanksasnull  
removequotes  
escape region 'us-east-1';
```

```
copy myetlproject.fact_atm_trans from  
's3://myetlproject/FACT_ATM_TYPE/part-00000-c1c1772e-b3d5-4d24-83d8-51ee8d616d24-c000.csv'  
iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'  
delimiter ','  
emptyasnull  
blanksasnull  
removequotes  
escape region 'us-east-1';
```

```
select count(*) from myetlproject.fact_atm_trans;
```

[Run](#)[Save](#)[Schedule](#)[Clear](#)[Query results](#)[Table details](#)

Query 1549 [🔗](#)

[Execution](#)

✅ Completed, started on April 05, 2021 at 06:38:50

ELAPSED TIME: 00 m 02 s

## Proof of successful creation of tables:

To view tables, select a schema.

myetlproject ▼

Q Filter tables

< 1 >

▶ dim_atm_pkey	...
▶ dim_card_type_pkey	...
▶ dim_date_pkey	...
▶ dim_location_pkey	...
▶ fact_atm_trans_pkey	...
▼ dim_atm	...
atm_id	
atm_number	
atm_manufacturer	
atm_location_id	
▶ dim_card_type	...
▶ dim_date	...
▶ dim_location	...
▶ fact_atm_trans	...


Proofs of data being loaded in the database by using select count(\*) from <table\_name>

### 1. Fact\_Atm\_trans

Query results

Table details

Query [295](#)

 Completed, started on April 05, 2021 at 21:45:12  
 ELAPSED TIME: 00 m 02 s

Rows returned (1)

count
2468571

### 2. DIM\_LOCATION

Run

Save


Schedule

Clear

Query results

Table details

Query [247](#)

 Completed, started on April 05, 2021 at 03:26:57  
 ELAPSED TIME: 00 m 02 s

Rows returned (1)

count
109

### 3. DIM\_ATM

Run

Save

Schedule

Clear

Query results

Table details

Query [222](#)

✓

Completed, started on April 05, 2021 at 03:24:26  
ELAPSED TIME: 00 m 02 s

Rows returned (1)

Q

Search rows

count
156

#### 4. DIM\_CARD\_TYPE

Run
Save
Schedule
Clear

Query results
Table details

Query [809](#)

✓ Completed, started on April 05, 2021 at 04:12:25  
 ELAPSED TIME: 00 m 02 s

**Rows returned (1)**

Search rows

count
12

#### 5. DIM\_DATE

Run
Save
Schedule
Clear

Query results
Table details

Query [331](#)

✓ Completed, started on April 05, 2021 at 21:47:35  
 ELAPSED TIME: 00 m 02 s

**Rows returned (1)**

Search rows

count
8685