

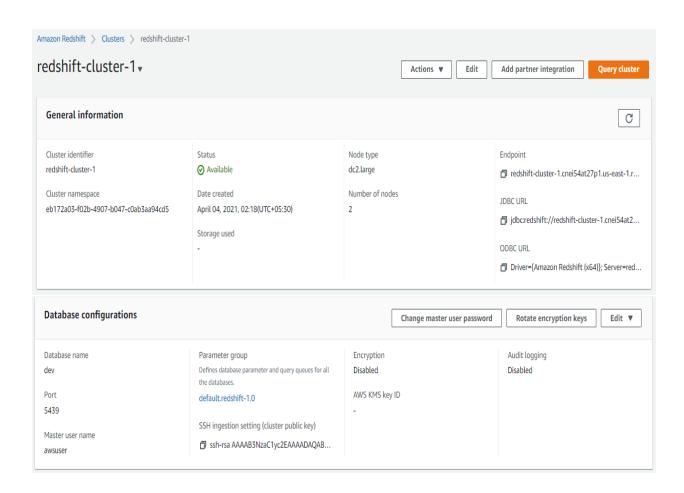


Creation of Redshift Cluster and Loading Data from S3 after creating tables

Creation of a RedShift Cluster

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

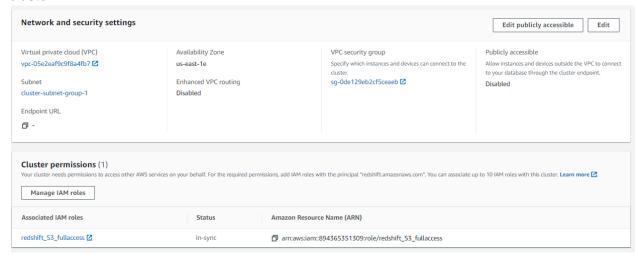
A redshift cluster with the name 'redshift-cluster-1' was created for the purpose of the assignment, Node used was dc2.large with Number of Nodes = 2







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







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 foriegn 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 integerl,
hour integer ,
weekday varchar(20),
primary key(date_id));
```

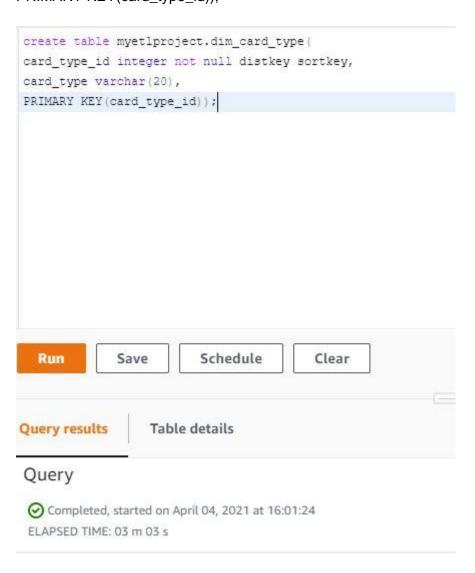
```
1 create table myetlproject.DIM DATE(
2 date_id integer not null,
3 full date time timestamp ,
4 year integer ,
5 month varchar(20) ,
6 day integer ,
7 hour integer ,
8 weekday varchar(20) ,
9 primary key(date_id));
10
11
12
13
14
15
     Run
                Save
                            Schedule
                                            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));



Creating DIM_ATM:

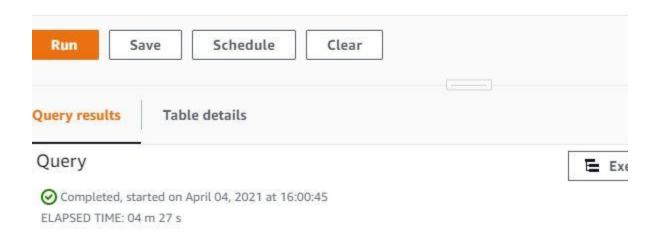
create table myetlproject.dim_atm(atm_id integer not null distkey sortkey, atm_number varchar(20), atm_manufacturer varchar(50), atm_location_id integer,





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));
```



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
 Ocompleted, 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_id integer,
weather_main varchar(50),
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:

```
copy myetlproject.dim_card_type from

's3://myetlproject/DIM_CARD_TYPE/part-00000-a6d78b0d-e375-4620-8497-3c295ab68350-c000.csv'

iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'
delimiter ',' region 'us-east-1';

select count(*) from myetlproject.dim_card_type;

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-b3d5-4d24-83d8-51ee8d616d24c000.csv'
iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'
delimiter ','
emptyasnull
blanksasnull
removequotes
escape region 'us-east-1';



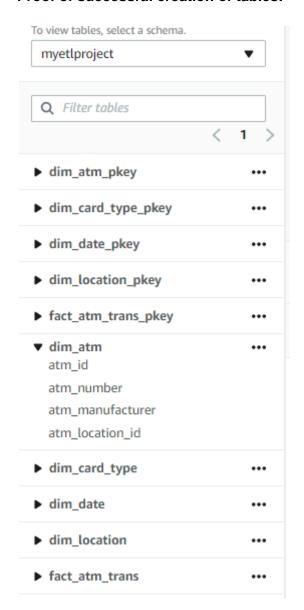


copy myetlproject.fact_atm_trans from
's3://myetlproject/FACT_ATM_TYPE/part-00000-c1c1772e-b3d5-4d24-83d8-51ee8d616d24-c000.csv'
iam_role 'arn:aws:iam::894365351309:role/redshift_S3_fullaccess'
delimiter ','
emptyasnull
blanksasnull
removequotes
escape region 'us-east-1';
select count(*) from myetlproject.fact_atm_trans;
Run Save Schedule Clear
Query results Table details
Query 1549 🖸
○ Completed, started on April 05, 2021 at 06:38:50
ELAPSED TIME: 00 m 02 s





Proof of successful creation of tables:





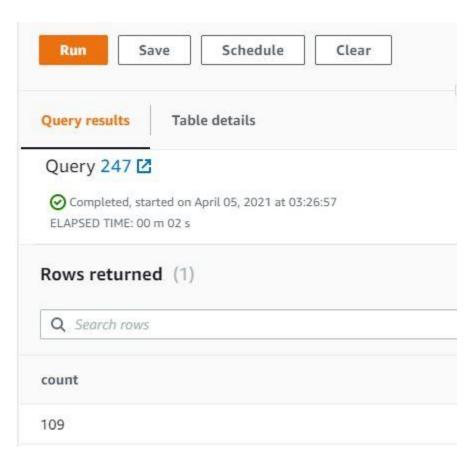


Proofs of data being loaded in the database by using select count(*) from <table_name>

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)
Q Search rows
count
2468571

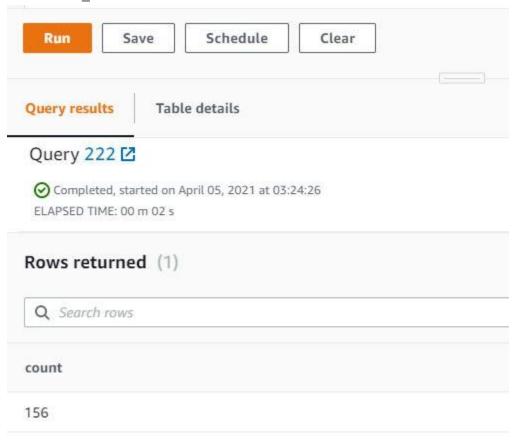
2. DIM_LOCATION







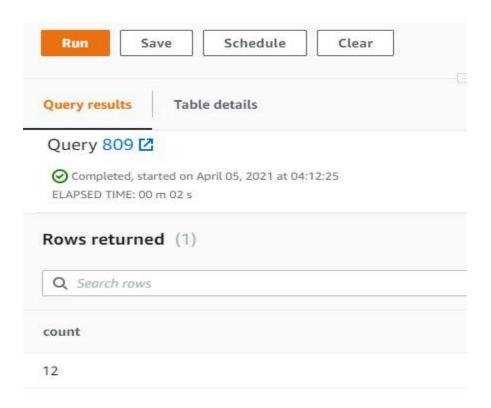
3. DIM_ATM







4. DIM_CARD_TYPE



5. DIM_DATE Run Save Schedule Clear Query results Table details Query 331 [2] Completed, started on April 05, 2021 at 21:47:35 ELAPSED TIME: 00 m 02 s Rows returned (1) Q Search rows count 8685