

Creation of a RedShift Cluster

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

Cluster configuration

Cluster identifier

This is the unique key that identifies a cluster.

etl-cluster-deepthi

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 configuration is free for a limited time. An organization has never created an Amazon Redshift cluster.

Choose the size of the cluster

I'll choose

Help me choose

Node type

Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.

dc2.large

Nodes

Enter the number of nodes that you need.

2

Range (1-32)

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

Queries to create the various dimension tables:

1. Creating DIM_LOCATION table:

```
create table etl.DIM_LOCATION (  
    location_id INT,  
    location VARCHAR(50),  
    streetname VARCHAR(255),  
    street_number INT,  
    zipcode INT,  
    lat DECIMAL(10,3),  
    lon DECIMAL(10,3)  
);
```

Output:

Query results

Table details


Query

✔ Completed, started on April 12, 2021 at 14:49:11
ELAPSED TIME: 01 m 54 s

2. Creating DIM_ATM table:

```
create table etl.DIM_ATM (  
    atm_id INT,  
    atm_number VARCHAR(20),  
    atm_manufacturer VARCHAR(50),  
    atm_location_id INT  
);
```


Output:

Query results	Table details
Query	
<div>  Completed, started on April 12, 2021 at 14:51:25 </div> <div>ELAPSED TIME: 00 m 32 s</div>	

3. Creating DIM_DATE table:

```
create table etl.DIM_DATE (
  date_id INT,
  full_date_time TIMESTAMP,
  year INT,
  month VARCHAR(20),
  day INT,
  hour INT,
  weekday VARCHAR(20)
);
```

Output:

Query results	Table details
Query	
<div>  Completed, started on April 12, 2021 at 14:53:47 </div> <div>ELAPSED TIME: 00 m 29 s</div>	

4. Creating DIM_CARD_TYPE table:

```
create table etl.DIM_CARD_TYPE(  
    card_type_id INT,  
    card_type VARCHAR(20)  
);
```

Output:

Query results

Table details

Query

✓ Completed, started on April 12, 2021 at 14:55:51
ELAPSED TIME: 00 m 46 s

Query to create fact table:

Creating FACT_ATM_TRANS table:

```
create table etl.FACT_ATM_TABLE (  
    trans_id BIGINT,  
    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(255),  
    message_text VARCHAR(255),  
    rain_3h DECIMAL(10,3),  
    clouds_all INT,  
    weather_id INT,  
    weather_main VARCHAR(50),  
    weather_description VARCHAR(255)  
);
```

Output:

Query results

Table details

Query

✔ Completed, started on April 12, 2021 at 15:00:22
ELAPSED TIME: 00 m 49 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

1. Copying data to DIM_DATE table

```
copy etl.DIM_DATE
from 's3://etlassignment-deepthi/dim_date_df/dim_date_df.csv'
iam_role 'arn:aws:iam::707958701452:role/redshift_s3_fullaccess_etl'
delimiter ',' region 'us-east-1'
CSV;
```


Output:

Query results	Table details
<p>Query</p> <p>  Completed, started on April 12, 2021 at 15:32:03 ELAPSED TIME: 00 m 30 s </p>	

2. Copying data to DIM_ATM table:

```
copy etl.DIM_ATM
from 's3://etlassignment-deepthi/dim_atm_df/dim_atm_df.csv'
iam_role 'arn:aws:iam::707958701452:role/redshift_s3_fullaccess_etl'
delimiter ',' region 'us-east-1'
CSV;
```


Output:

Query results	Table details
<p>Query</p> <p>  Completed, started on April 12, 2021 at 15:33:09 ELAPSED TIME: 00 m 36 s </p>	

3. Copying data to DIM_LOCATION table:

```
copy etl.DIM_LOCATION
from 's3://etlassignment-deepthi/dim_location_df/dim_location_df.csv'
iam_role 'arn:aws:iam::707958701452:role/redshift_s3_fullaccess_etl'
delimiter ',' region 'us-east-1'
CSV;
```


Output:

Query results	Table details
<p>Query</p> <p>  Completed, started on April 12, 2021 at 15:34:19 ELAPSED TIME: 00 m 46 s </p>	

4. Copying data to DIM_CARD_TYPE table:

```
copy etl.DIM_CARD_TYPE
from 's3://etlassignment-deepthi/dim_cardtype_df/dim_cardtype_df.csv'
iam_role 'arn:aws:iam::707958701452:role/redshift_s3_fullaccess_etl'
delimiter ',' region 'us-east-1'
CSV;
```


Output:

Query results	Table details
<p>Query</p> <p>  Completed, started on April 12, 2021 at 15:35:09 ELAPSED TIME: 00 m 36 s </p>	

5. Copying data to FACT_ATM_TRANS:

```
copy etl.FACT_ATM_TABLE
from 's3://etlassignment-deepthi/fact_trans/fact_trans.csv'
iam_role 'arn:aws:iam::707958701452:role/redshift_s3_fullaccess_etl'
delimiter ',' region 'us-east-1'
CSV;
```

Output:

Query results	Table details
<p>Query</p> <p>  Completed, started on April 12, 2021 at 15:36:28 ELAPSED TIME: 00 m 55 s </p>	