Creation of a RedShift Cluster

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

<Screenshot of the type of machine used along with number of nodes>

General information

Cluster identifier	Status	Node type	Endpoint
redshift-cluster-1-etl	Available	dc2.large	redshift-cluster-1-etl.cer7xoj4ajun.us-...
Cluster namespace	**Date created**	**Number of nodes**	**JDBC URL**
69b6c2a4-5878-432e-b0ef-0dde28b3e401	June 10, 2021, 05:02(UTC+05:30)	2	jdbc:redshift://redshift-cluster-1-etl.ce...
	Storage used	**AQUA**	**ODBC URL**
	0.17% (0.55 of 320 GB used)	Not available	Driver={Amazon Redshift (x64)}; Serve...

Database configurations

Change admin user password | Rotate encryption keys | Edit

Database name	Parameter group	Encryption	Audit logging
redshiftdb	Defines database parameter and query queues for all the databases.	Disabled	Disabled
Port	default.redshift-1.0	**AWS KMS key ID**	
5439	**SSH ingestion setting (cluster public key)**	-	
Admin user name	ssh-rsa AAAAB3NzaC1yc2EAAAADA...		
awsuser			

Network and security settings

Edit publicly accessible | Edit

Virtual private cloud (VPC)	Availability Zone	VPC security group	Publicly accessible
vpc-0d044f62ddd235eed	us-east-1a	Specify which instances and devices can connect to the cluster.	Allow instances and devices outside the VPC to connect to your database through the cluster endpoint.
Subnet	**Enhanced VPC routing**	sg-0c0797c0c1a04ee4c	Disabled
cluster-subnet-group-1	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 assc cluster. Learn more

Manage IAM roles

Associated IAM roles	Status	Amazon Resource Name (ARN)
redshift_etl_proj	in-sync	arn:aws:iam::000902302705:role/redshift_etl_proj

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:

<Queries>

```
create group spark_redshift_etl_group with user awsuser;

create schema if not exists etl_proj_schema;

--////location
--create table if not exists etl_proj_schema.dim_location
(
location_id varchar(50) not null distkey sortkey primary key,
location VARCHAR(50),
streetname VARCHAR(255),
street_number INTEGER,
zipcode INTEGER,
lat DECIMAL(10,3),
lon DECIMAL(10,3)
);

--select * from etl_proj_schema.dim_location

--/////////-- card type
create table if not exists etl_proj_schema.dim_card_type
(
card_type_id VARCHAR(50) not null distkey sortkey primary key,
card_type VARCHAR(40)
);

---select * from etl_proj_schema.dim_card_type


--///// dim_date
create table if not exists etl_proj_schema.dim_date
(
```

```
date_id VARCHAR(50) not null distkey sortkey primary key,
full_date_time TIMESTAMP,
year INTEGER,
month VARCHAR(20),
day INTEGER,
hour INTEGER,
weekday VARCHAR(20)
);

--//////// atm
create table if not exists etl_proj_schema.dim_atm
(
atm_id VARCHAR(50) not null distkey sortkey primary key,
atm_number VARCHAR(20),
atm_manufacturer VARCHAR(50),
atm_location_id VARCHAR(50) references etl_proj_schema.dim_location(location_id)

);


--//// fact
create table if not exists etl_proj_schema.fact_atm_trans
(
trans_id VARCHAR(50) not null distkey sortkey primary key,
atm_id VARCHAR(50) references etl_proj_schema.dim_atm(atm_id),
weather_loc_id VARCHAR(50) references etl_proj_schema.dim_location(location_id),
date_id VARCHAR(50) references etl_proj_schema.dim_date(date_id),
card_type_id VARCHAR(50) references etl_proj_schema.dim_card_type(card_type_id),
atm_status VARCHAR(20),
currency VARCHAR(10),
service VARCHAR(20),
transaction_amount INTEGER,
message_code VARCHAR(255),
message_text VARCHAR(255),
rain_3h DECIMAL(10,3),
clouds_all INTEGER,
weather_id INTEGER,
weather_main VARCHAR(50),
weather_description VARCHAR(255)
);
```

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

<Queries>

```
--///////// location
copy etl_proj_schema.dim_location from
's3://etlproj/Dim_Location/Dim_Location.csv'
iam_role 'arn:aws:iam::000902302705:role/redshift_etl_proj'
csv region 'us-east-1';


--//// atm
copy etl_proj_schema.dim_atm from
's3://etlproj/Dim_ATM/Dim_ATM.csv'
iam_role 'arn:aws:iam::000902302705:role/redshift_etl_proj'
csv region 'us-east-1';


---///// card type
copy etl_proj_schema.dim_card_type from
's3://etlproj/Dim_Card_Type/Dim_Card_Type.csv'
iam_role 'arn:aws:iam::000902302705:role/redshift_etl_proj'
acceptinvchars csv region 'us-east-1';


-----/// date
copy etl_proj_schema.dim_date from
's3://etlproj/Dim_Date/Dim_Date.parquet'
iam_role 'arn:aws:iam::000902302705:role/redshift_etl_proj'
format as parquet;


--- fact table
copy etl_proj_schema.fact_atm_trans from
's3://etlproj/Fact_Transaction/Fact_Transaction.csv'
iam_role 'arn:aws:iam::000902302705:role/redshift_etl_proj'
csv region 'us-east-1';
```