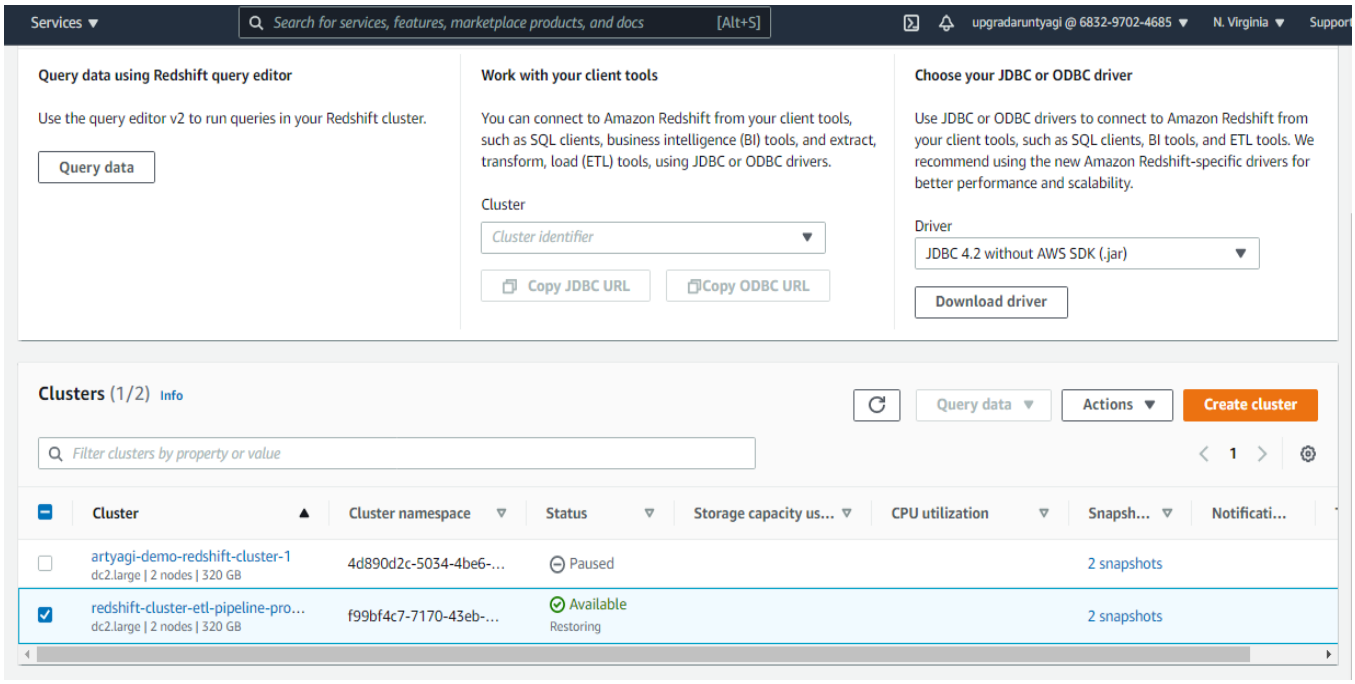
Creation of a RedShift Cluster

~Arun Tyagi

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



Amazon Redshift > Clusters > redshift-cluster-etl-pipeline-project

redshift-cluster-etl-pipeline-project

Actions ▼ | Edit | Add partner integration | Query data ▼

General information

Cluster identifier	Status	Node type	Endpoint
redshift-cluster-etl-pipeline-project	Available	dc2.large	redshift-cluster-etl-pipeline-project.ckqfz1p...
Cluster namespace	**Date created**	**Number of nodes**	**JDBC URL**
f99bf4c7-7170-43eb-b5d7-e218eb7c64cc	November 02, 2021, 03:07(UTC+05:30)	2	jdbc:redshift://redshift-cluster-etl-pipeline-...
	Storage used	**AQUA**	**ODBC URL**
	-	Not available	Driver={Amazon Redshift (x64)}; Server=red...

Database configurations

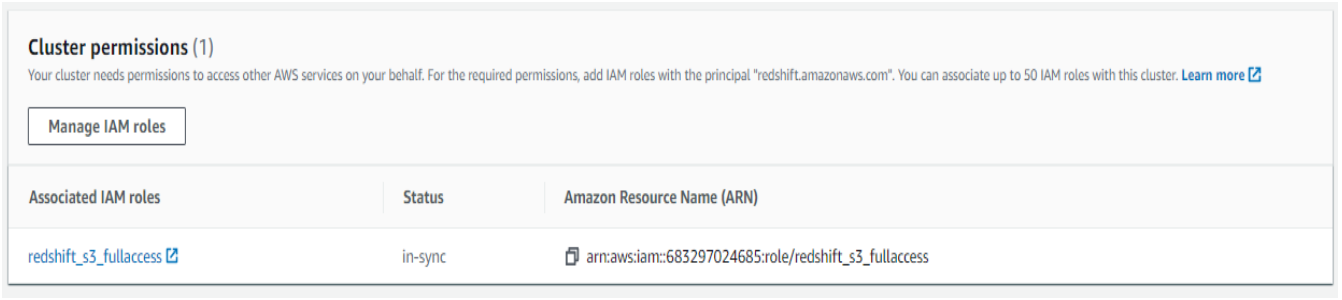
Change admin user password | Rotate encryption keys | Edit ▼

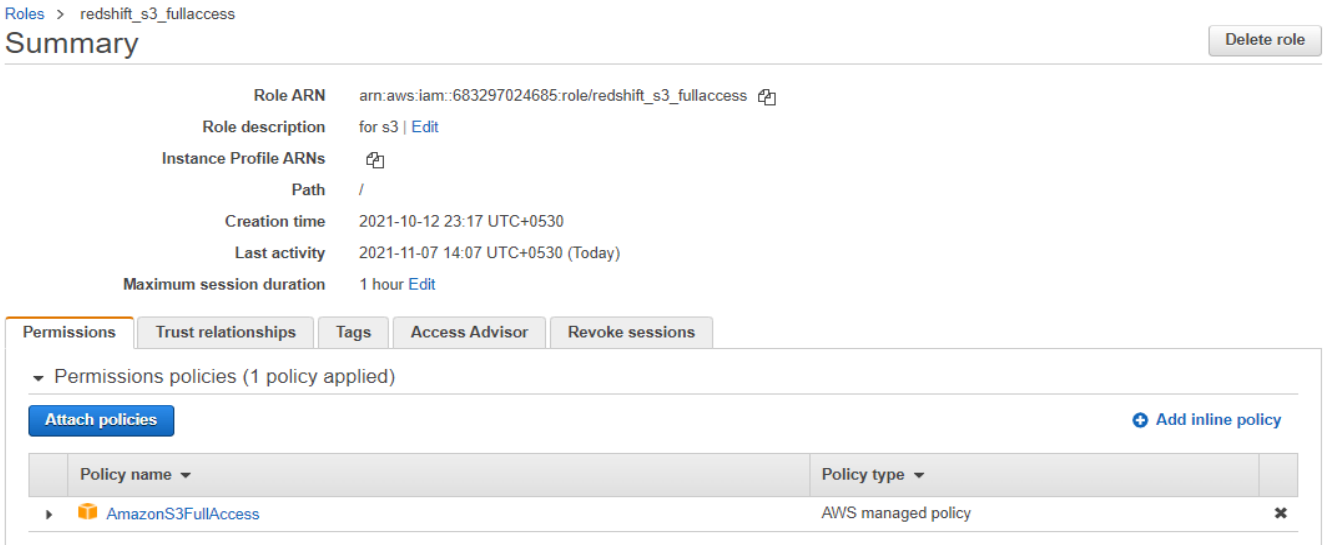
Database name	Parameter group	Encryption	Audit logging
artyagi-etl-project-db	Defines database parameter and query queues for all the databases.	Disabled	Disabled
Port	default.redshift-1.0	**AWS KMS key ID**	
5440	**SSH ingestion setting (cluster public key)**	-	
Admin user name	ssh-rsa AAAAB3NzaC1yc2EAAAADAQAB...		
awsuser			

Network and security settings

Edit

Virtual private cloud (VPC)	Availability Zone	VPC security group	Publicly accessible
vpc-0b13e60f8c2f5bebc	us-east-1e	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-03579dafd2f0a6a45	Disabled
cluster-subnet-group-etl-project	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 50 IAM roles with this cluster. Learn more
Manage IAM roles
Associated IAM roles
Status
Amazon Resource Name (ARN)
redshift_s3_fullaccess
in-sync
arn:aws:iam::683297024685:role/redshift_s3_fullaccess


Roles > redshift_s3_fullaccess
Summary
Delete role
Role ARN
arn:aws:iam::683297024685:role/redshift_s3_fullaccess
Role description
for s3 | Edit
Instance Profile ARNs
Path
/
Creation time
2021-10-12 23:17 UTC+0530
Last activity
2021-11-07 14:07 UTC+0530 (Today)
Maximum session duration
1 hour Edit
Permissions
Trust relationships
Tags
Access Advisor
Revoke sessions
Permissions policies (1 policy applied)
Attach policies
Add inline policy
Policy name
Policy type
AmazonS3FullAccess
AWS managed policy

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

1. create schema etl;

2. create table etl.loc(

   location_id integer not null,

   atm_location varchar(50),

   atm_streetname varchar(255),

   atm_street_number integer,

   atm_zipcode integer,

   atm_lat NUMERIC(10,3),

   atm_lon NUMERIC(10,3),

   primary key(location_id));

3. create table etl.atm(

   atm_prim_id integer not null,

   atm_id varchar(20),

   atm_manufacturer varchar(50),

   location_id integer,

   primary key(atm_prim_id),

foreign key(location_id) references etl.loc(location_id));

4. create table etl.date(
   year integer,
   month varchar(20),
   day integer,
   hour integer,
   weekday varchar(20),
   full_date_time timestamp,
   date_id integer,
   primary key(date_id));

5. create table etl.card(
   card_type varchar(23),
   card_type_id integer,
   primary key(card_type_id));

6. create table etl.FACT_ATM_TRANS(

   trans_id BIGINT not null,

   atm_prim_id integer,

   location_id integer,

   date_id integer,

   card_type_id integer,

   atm_status varchar(20),

   currency varchar(10),

   service varchar(20),

   transaction_amount integer,

   message_code varchar(255),

   message_text varchar(255),

   rain_3h NUMERIC(10,3),

   clouds_all integer,

   weather_id integer,

   weather_main varchar(50),

   weather_description varchar(255),

   primary key(trans_id),

   foreign key(atm_prim_id) references etl.atm(atm_prim_id),

   foreign key(location_id) references etl.loc(location_id),

   foreign key(date_id) references etl.date(date_id),

```
foreign key(card_type_id) references etl.card(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:

1. copy etl.atm from

   's3://artyagi-s3-bucket-etl-project/DIM_ATM/part-00000-93319583-0153-4457-ae3c-21f0588999d9-c000.csv'

   iam_role 'arn:aws:iam::683297024685:role/redshift_s3_fullaccess'

   delimiter ',' IGNOREHEADER 1

   region 'us-east-1';

2. copy etl.loc from

   's3://artyagi-s3-bucket-etl-project/DIM_LOC/part-00000-62c6190e-3c9d-48ea-b3f3-1f239e48ec0d-c000.csv'

   iam_role 'arn:aws:iam::683297024685:role/redshift_s3_fullaccess'

   delimiter ',' IGNOREHEADER 1

   region 'us-east-1';

3. copy etl.date from

   's3://artyagi-s3-bucket-etl-project/DIM_DATE/part-00000-af40679b-5327-43cd-a608-bbb5d710365d-c000.csv'

iam_role 'arn:aws:iam::683297024685:role/redshift_s3_fullaccess'

delimiter ',' IGNOREHEADER 1

timeformat 'auto'

region 'us-east-1';

4. copy etl.card from

   's3://artyagi-s3-bucket-etl-project/DIM_CARD/part-00000-9b9e9793-3386-49e1-b127-ec5472249b77-c000.csv'

   iam_role 'arn:aws:iam::683297024685:role/redshift_s3_fullaccess'

   delimiter ',' IGNOREHEADER 1

   region 'us-east-1';

5. copy etl.FACT_ATM_TRANS from

   's3://artyagi-s3-bucket-etl-project/FACT_ATM_TRANS/part-00000-258816ca-68f3-4f7b-9526-648dee21e22d-c000.csv'

   iam_role 'arn:aws:iam::683297024685:role/redshift_s3_fullaccess'

   delimiter ',' IGNOREHEADER 1

   region 'us-east-1'

   TRUNCATECOLUMNS

   CSV;