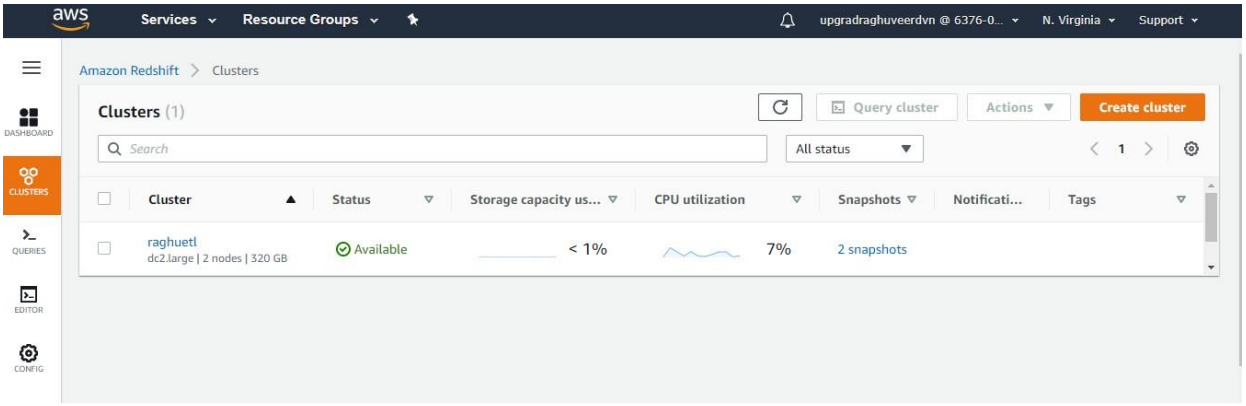
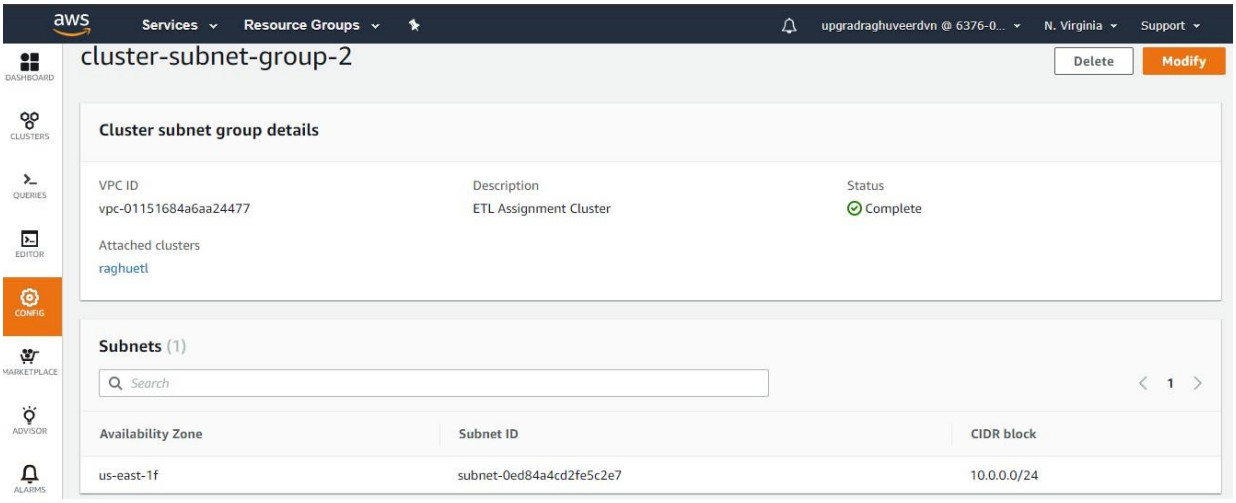
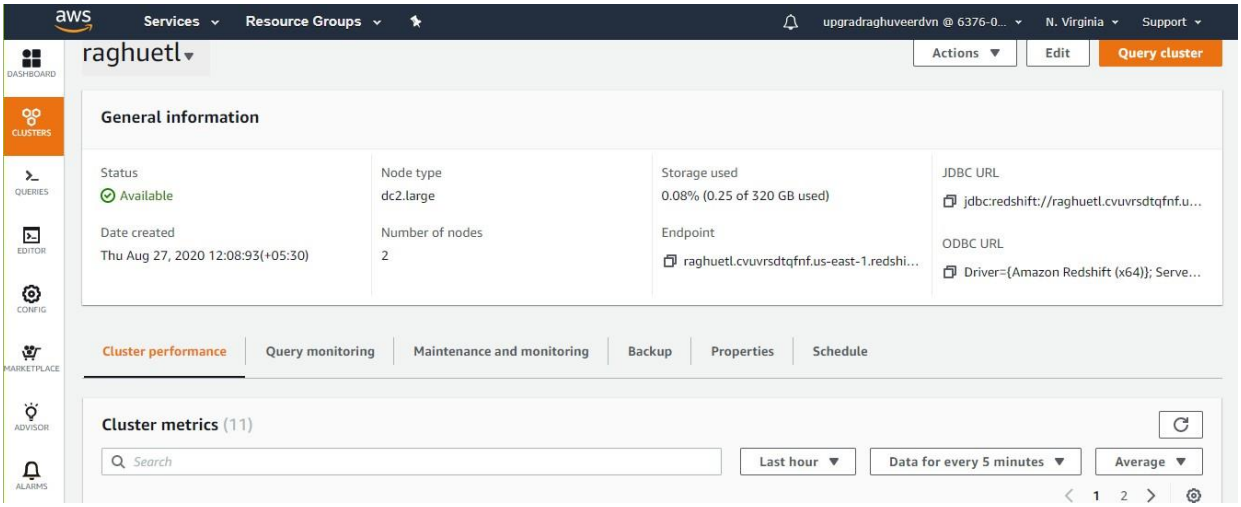
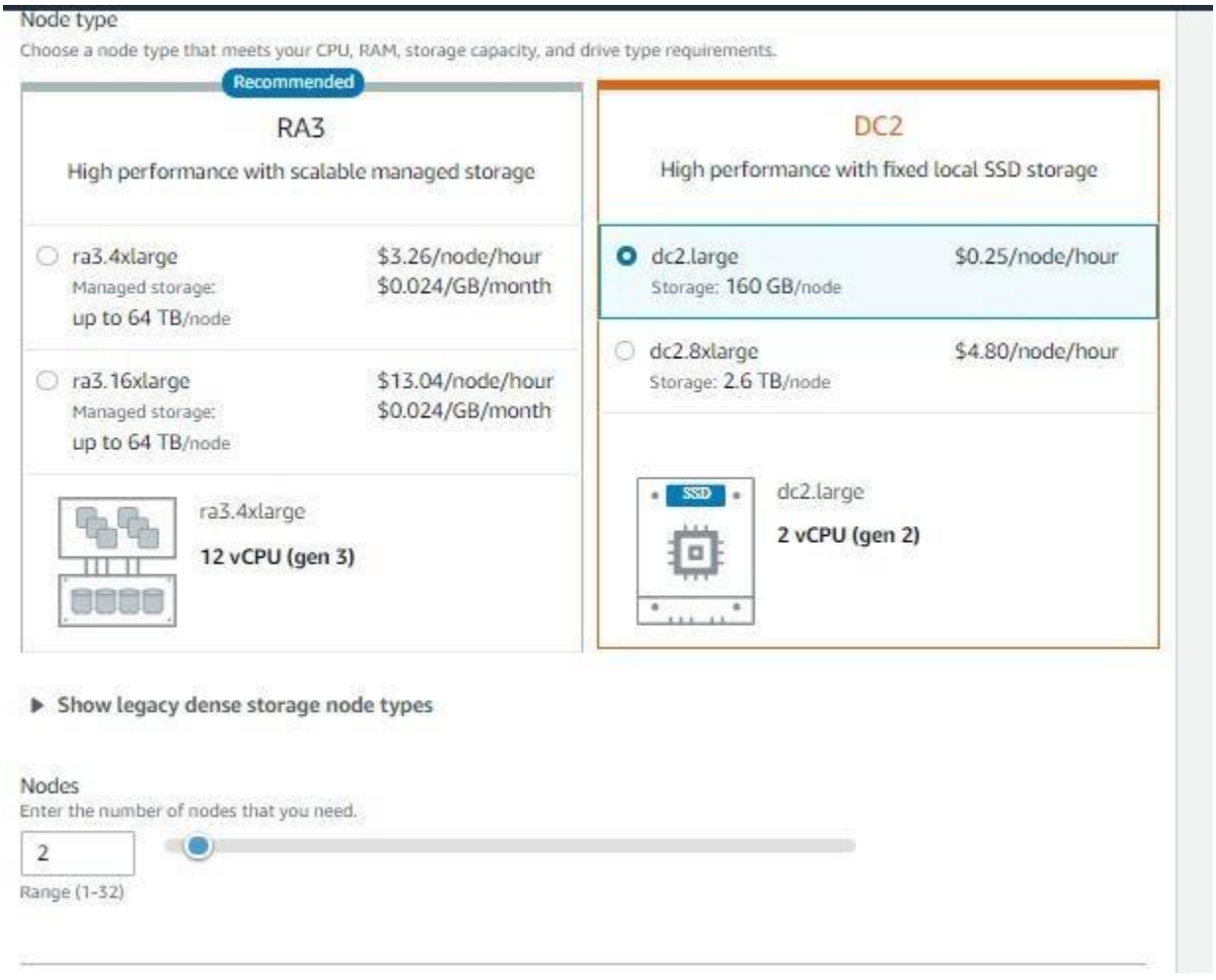
# Creation of a RedShift Cluster

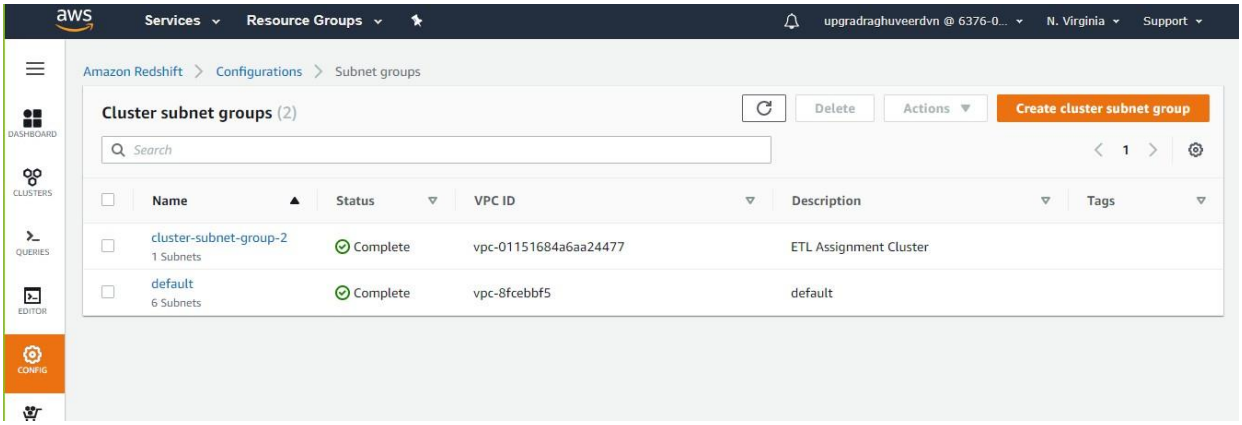
**Screenshots of the configuration of the RedShift cluster that you have created:**

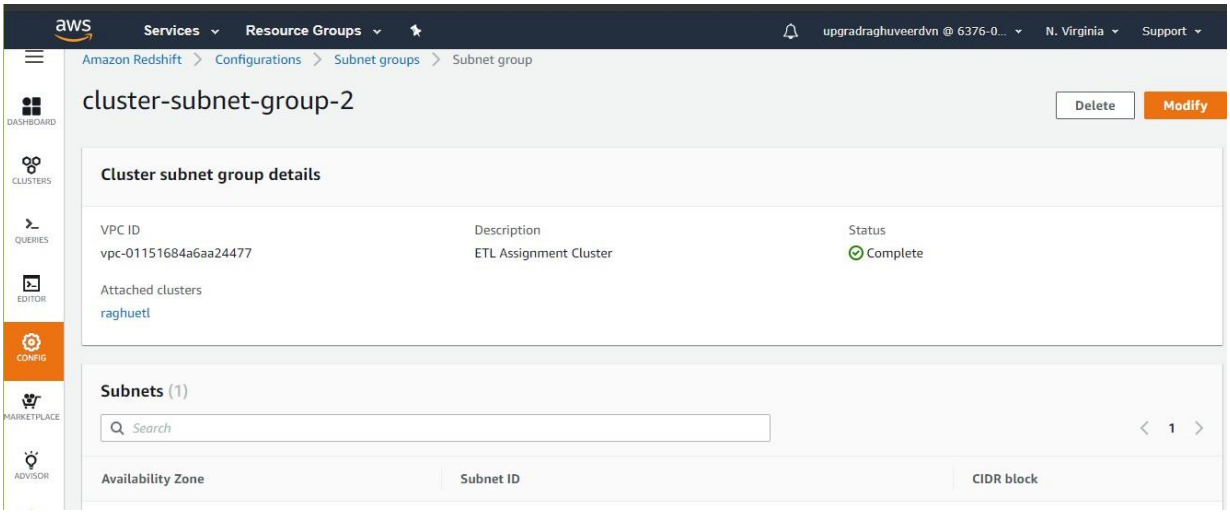



aws
Services
Resource Groups
upgradraghuveerdvn @ 6376-0...
N. Virginia
Support
DASHBOARD
CLUSTERS
QUERIES
EDITOR
CONFIG
MARKETPLACE
ADVISOR
ALARMS
raghuetl
Actions
Edit
Query cluster
General information
Status
Available
Node type
dc2.large
Storage used
0.08% (0.25 of 320 GB used)
JDBC URL
jdbc:redshift://raghuetl.cvuvrsdtqfnf.u...
Date created
Thu Aug 27, 2020 12:08:93(+05:30)
Number of nodes
2
Endpoint
raghuetl.cvuvrsdtqfnf.us-east-1.redshi...
ODBC URL
Driver={Amazon Redshift (x64)}; Serve...
Cluster performance
Query monitoring
Maintenance and monitoring
Backup
Properties
Schedule
Cluster metrics (11)
Search
Last hour
Data for every 5 minutes
Average
1
2


Node type
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.
Recommended
RA3
High performance with scalable managed storage
ra3.4xlarge
Managed storage:
up to 64 TB/node
$3.26/node/hour
$0.024/GB/month
ra3.16xlarge
Managed storage:
up to 64 TB/node
$13.04/node/hour
$0.024/GB/month
ra3.4xlarge
12 vCPU (gen 3)
DC2
High performance with fixed local SSD storage
dc2.large
Storage: 160 GB/node
$0.25/node/hour
dc2.8xlarge
Storage: 2.6 TB/node
$4.80/node/hour
SSD
dc2.large
2 vCPU (gen 2)
Show legacy dense storage node types
Nodes
Enter the number of nodes that you need.
2
Range (1-32)


aws
Services
Resource Groups
upgradraghuveerdvn @ 6376-0...
N. Virginia
Support
Amazon Redshift
Configurations
Subnet groups
Cluster subnet groups (2)
Delete
Actions
Create cluster subnet group
Search
1
Name
Status
VPC ID
Description
Tags
cluster-subnet-group-2
1 Subnets
Complete
vpc-01151684a6aa24477
ETL Assignment Cluster
default
6 Subnets
Complete
vpc-8fcebbf5
default
DASHBOARD
CLUSTERS
QUERIES
EDITOR
CONFIG


aws
Services
Resource Groups
upgradraghuveerdvn @ 6376-0...
N. Virginia
Support
Amazon Redshift
Configurations
Subnet groups
Subnet group
cluster-subnet-group-2
Delete
Modify
Cluster subnet group details
VPC ID
vpc-01151684a6aa24477
Description
ETL Assignment Cluster
Status
Complete
Attached clusters
raghuetl
Subnets (1)
Search
1
Availability Zone
Subnet ID
CIDR block
DASHBOARD
CLUSTERS
QUERIES
EDITOR
CONFIG
MARKETPLACE
ADVISOR

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 etl;

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));


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));


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));
```

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


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**

```
copy etl.atm from
's3://redshiftdvn/ETL/DIM_ATM/part-00000-b81afa6c-3dbf-4a2d-9d33-0c7eaa9de097-c000'
iam_role 'arn:aws:iam::637602732672:role/upgrad-redshift-s3-access'
delimiter ',' IGNOREHEADER 1
region 'us-east-1';

copy etl.loc from
's3://redshiftdvn/ETL/DIM_LOC/part-00000-25b4b17c-2492-41e9-8773-33a2751cbf24-c000.csv'
iam_role 'arn:aws:iam::637602732672:role/upgrad-redshift-s3-access'
delimiter ',' IGNOREHEADER 1
region 'us-east-1';

copy etl.date from
's3://redshiftdvn/ETL/DIM_DATE/part-00000-59ccb227-a4b1-49f2-8689-5c39c333405f-
c000.csv'
iam_role 'arn:aws:iam::637602732672:role/upgrad-redshift-s3-access'
delimiter ',' IGNOREHEADER 1
timeformat 'auto'
region 'us-east-1';

copy etl.card from
's3://redshiftdvn/ETL/DIM_CARD/part-00000-f4f7cd66-edce-4415-99ae-9f3368e47dcc-
c000.csv'
iam_role 'arn:aws:iam::637602732672:role/upgrad-redshift-s3-access'
delimiter ',' IGNOREHEADER 1
region 'us-east-1';

copy etl.FACT_ATM_TRANS from
's3://redshiftdvn/ETL/FACT_ATM_TRANS/part-00000-5bc708bb-022d-4d78-85dc-
e49447a9402a-c000.csv'
iam_role 'arn:aws:iam::637602732672:role/upgrad-redshift-s3-access'
delimiter ',' IGNOREHEADER 1
region 'us-east-1'
TRUNCATECOLUMNS
CSV;
```