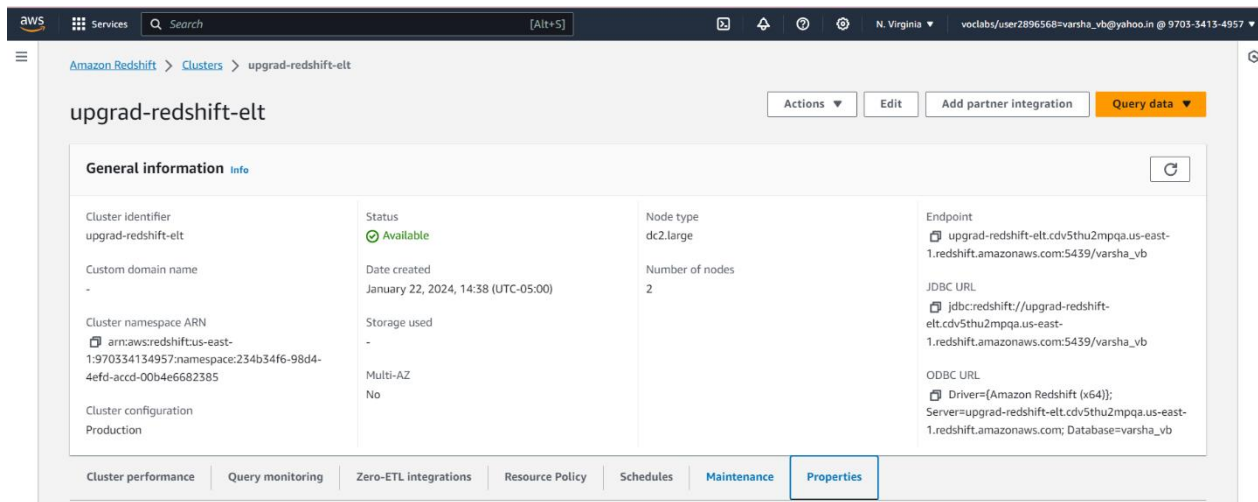


Creation Of a Redshift Cluster

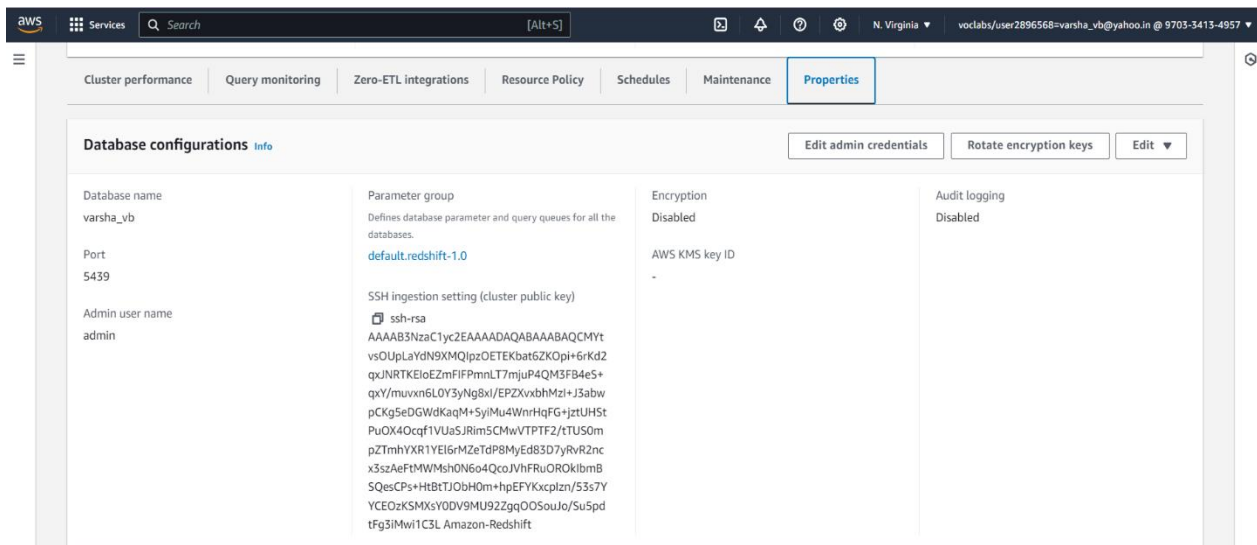
Screenshots of the configuration of the Redshift cluster that you have created:



The screenshot shows the AWS Redshift console interface. The breadcrumb navigation indicates the path: Amazon Redshift > Clusters > upgrad-redshift-elt. The cluster name 'upgrad-redshift-elt' is displayed at the top. Below the cluster name, there are buttons for 'Actions', 'Edit', 'Add partner integration', and 'Query data'. The 'General information' tab is selected, showing the following details:

General information			
Cluster identifier upgrad-redshift-elt	Status Available	Node type dc2.large	Endpoint upgrad-redshift-elt.cdv5thu2mpqa.us-east-1.redshift.amazonaws.com:5439/varsha_vb
Custom domain name -	Date created January 22, 2024, 14:38 (UTC-05:00)	Number of nodes 2	JDBC URL jdbc:redshift://upgrad-redshift-elt.cdv5thu2mpqa.us-east-1.redshift.amazonaws.com:5439/varsha_vb
Cluster namespace ARN arn:aws:redshift:us-east-1:970334134957:namespace:234b34f6-98d4-4efd-acc0-00b4e6682385	Storage used -	Multi-AZ No	ODBC URL Driver=(Amazon Redshift (x64)); Server=upgrad-redshift-elt.cdv5thu2mpqa.us-east-1.redshift.amazonaws.com; Database=varsha_vb
Cluster configuration Production			

At the bottom, there are tabs for 'Cluster performance', 'Query monitoring', 'Zero-ETL integrations', 'Resource Policy', 'Schedules', 'Maintenance', and 'Properties'.



The screenshot shows the AWS Redshift console interface, specifically the 'Properties' tab for the 'upgrad-redshift-elt' cluster. The breadcrumb navigation indicates the path: Amazon Redshift > Clusters > upgrad-redshift-elt. The cluster name 'upgrad-redshift-elt' is displayed at the top. Below the cluster name, there are buttons for 'Edit admin credentials', 'Rotate encryption keys', and 'Edit'. The 'Database configurations' tab is selected, showing the following details:

Database configurations			
Database name varsha_vb	Parameter group default.redshift-1.0	Encryption Disabled	Audit logging Disabled
Port 5439	SSH ingestion setting (cluster public key) ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAQCMYt vsOUplLaYdN9XMQlpzOETEKbat6ZK0pi+6rKd2 qxJNRTKEIoEZmFIFPmnlT7mjup4QM3F84eS+ qxY/muvxn6L0Y3yNg8xl/EPZXvxbhMzl+J3abw pCKg5eDGDWdKaQM+5yIMu4WnrHqFG+ztUHS PuOX4Ocqf1VUaSJrim5CMwVTPTF2/tTUS0m pZTmhYXR1YEl6rMZeTdP8MyEd83D7yRvR2nc x3szAeFTMW/MshON6o4QcoJVfRuOROKlmbB SQesCPs+HtBtTJOH0m+hpEFYKxcpzn/53s7Y YCE0zKSMXsY0DV9MU92ZgqOOSouJo/Su5pd tFg3IMw1C3L Amazon-Redshift	AWS KMS key ID -	

At the bottom, there are tabs for 'Cluster performance', 'Query monitoring', 'Zero-ETL integrations', 'Resource Policy', 'Schedules', 'Maintenance', and 'Properties'.

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 foreign keys:

-- Create schema

```
CREATE SCHEMA atm_data;
```

-- Create table DIM_LOCATION

```
CREATE TABLE atm_data.DIM_LOCATION  
(  
    location_id INT NOT NULL DISTKEY SORTKEY,  
    location VARCHAR(50),  
    streetname VARCHAR(255),  
    street_number INT,  
    zipcode INT,  
    lat DECIMAL(10,3),  
    lon DECIMAL(10,3),  
    PRIMARY KEY (location_id)  
);
```

-- Create table DIM_ATM

```
create table atm_data.DIM_ATM  
(  
    atm_id int not null DISTKEY SORTKEY,  
    atm_number varchar(20),  
    atm_manufacturer varchar(50),  
    atm_location_id int,  
    PRIMARY KEY(atm_id),  
    FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id)  
);
```

-- Create table DIM_DATE

```
create table atm_data.DIM_DATE
(
date_id int not null DISTKEY SORTKEY,
full_date_time timestamp,
year int,
month varchar(20),
day int,
hour int,
weekday varchar(20),
PRIMARY KEY(date_id)
);
```

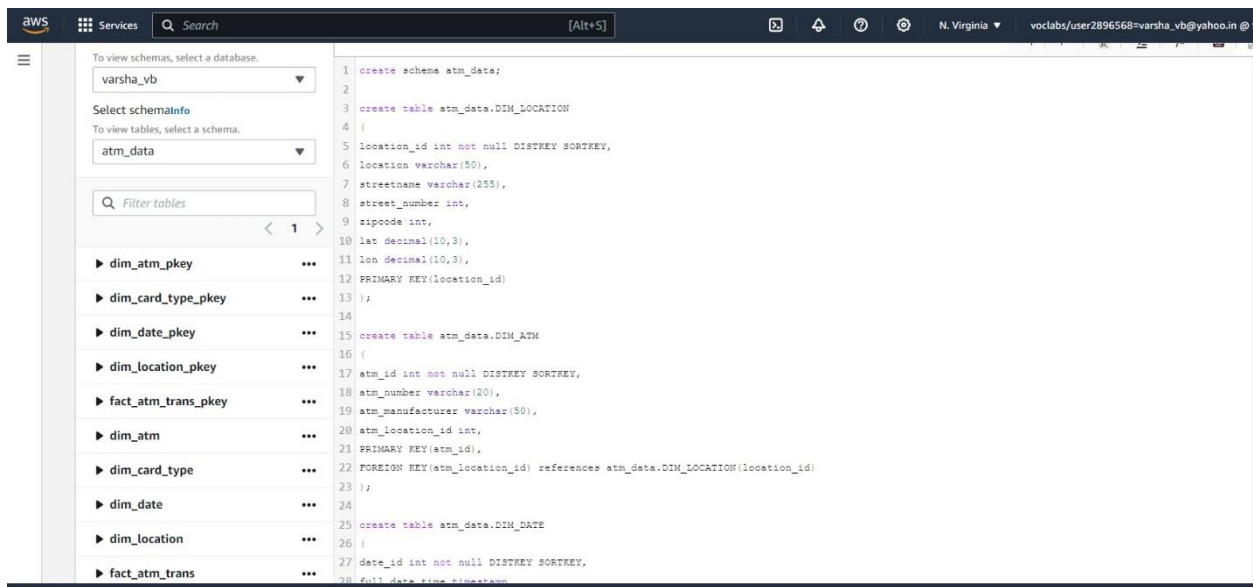
-- Create table DIM_CARD_TYPE

```
CREATE TABLE atm_data.DIM_CARD_TYPE
(
card_type_id INT NOT NULL DISTKEY SORTKEY,
card_type VARCHAR(30),
PRIMARY KEY (card_type_id)
);
```

-- Create table FACT_ATM_TRANS

```
create table atm_data.FACT_ATM_TRANS
(
trans_id bigint not null DISTKEY SORTKEY,
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(225),
message_text varchar(225),
rain_3h decimal(10,3),
clouds_all int,
weather_id int,
weather_main varchar(50),
weather_description varchar(255),
PRIMARY KEY(trans_id),
FOREIGN KEY(weather_loc_id) references atm_data.DIM_LOCATION(location_id),
FOREIGN KEY(atm_id) references atm_data.DIM_ATM(atm_id),
FOREIGN KEY(date_id) references atm_data.DIM_DATE(date_id),
FOREIGN KEY(card_type_id) references atm_data.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

```
copy atm_data.dim_location from 's3://s3bucket4upgradredshift/ETL-Upgrad/DIM_LOCATION/'
iam_role 'arn:aws:iam::970334134957:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

```
copy atm_data.dim_atm from 's3://s3bucket4upgradredshift/ETL-Upgrad/DIM_ATM/'
iam_role 'arn:aws:iam::970334134957:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

```
copy atm_data.dim_date from 's3://s3bucket4upgradredshift/ETL-Upgrad/DIM_DATE/'
iam_role 'arn:aws:iam::970334134957:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

```
copy atm_data.dim_card_type from 's3://s3bucket4upgradredshift/ETL-Upgrad/DIM_CARD/'
iam_role 'arn:aws:iam::970334134957:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

```
copy atm_data.FACT_ATM_TRANS from 's3://s3bucket4upgradredshift/ETL-
Upgrad/FACT_ATM_TRANS/'
iam_role 'arn:aws:iam::970334134957:role/myRedshiftRole'
delimiter ',' region 'us-east-1'
CSV;
```

