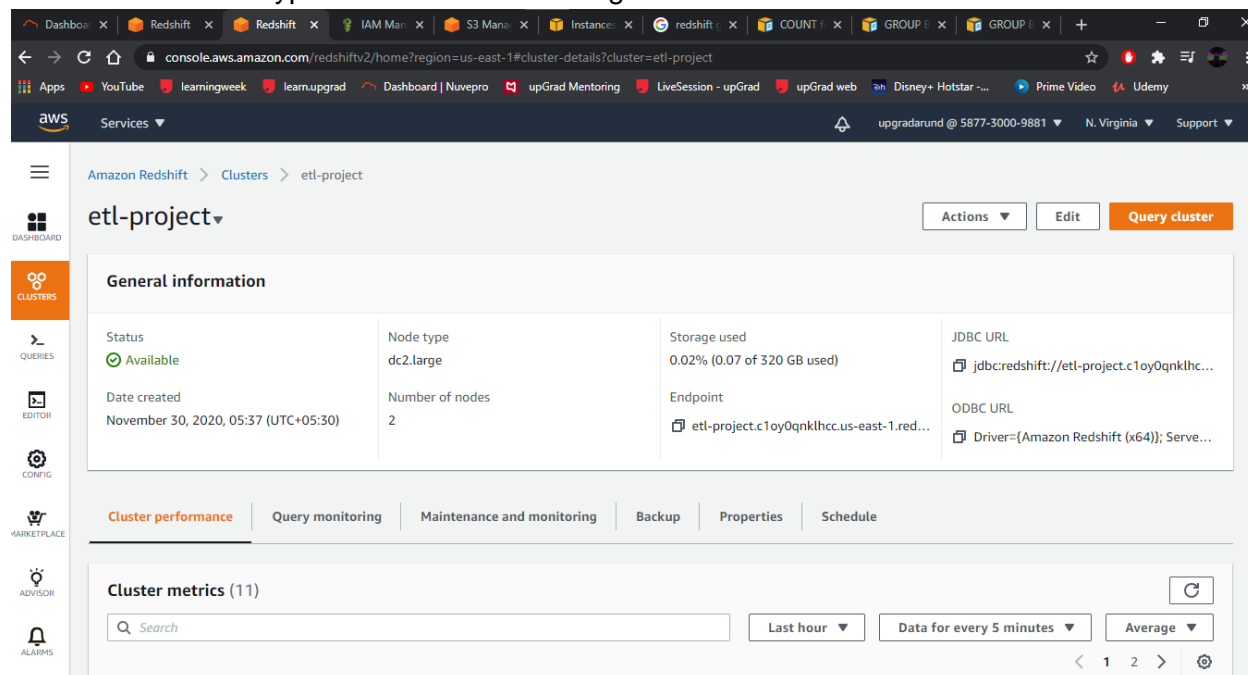


## Creation of a RedShift Cluster

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

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



The screenshot shows the AWS RedShift console for a cluster named 'etl-project'. The 'General information' tab is selected, displaying the following details:

Property	Value
Status	Available
Node type	dc2.large
Storage used	0.02% (0.07 of 320 GB used)
JDBC URL	jdbc:redshift://etl-project.c1oy0qnklhcc.us-east-1.redshift.amazonaws.com:5439/
Date created	November 30, 2020, 05:37 (UTC+05:30)
Number of nodes	2
Endpoint	etl-project.c1oy0qnklhcc.us-east-1.redshift.amazonaws.com
ODBC URL	Driver={Amazon Redshift (x64)}; Server=etl-project.c1oy0qnklhcc.us-east-1.redshift.amazonaws.com

Below the general information, there are tabs for 'Cluster performance', 'Query monitoring', 'Maintenance and monitoring', 'Backup', 'Properties', and 'Schedule'. The 'Cluster metrics' section shows 11 metrics, with a search bar and filters for 'Last hour', 'Data for every 5 minutes', and 'Average'.

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:**

<Queries>

```
create schema redshift_etl_project;
```

```
create table redshift_etl_project.dim_location(
location_id integer not null distkey sortkey,
location varchar(50),
streetname varchar(255),
street_number integer,
```

```
zipcode integer,  
lat DECIMAL(10,3),  
lon DECIMAL(10,3));
```

```
create table redshift_etl_project.dim_atm(  
atm_id integer not null distkey sortkey,  
atm_number varchar(20),  
atm_manufacturer varchar(25),  
location_id integer);
```

```
create table redshift_etl_project.dim_card_type(  
card_type_id integer not null distkey sortkey,  
card_type varchar(20));
```

```
create table redshift_etl_project.dim_date(  
date_id integer not null distkey sortkey,  
full_date_time timestamp,  
year integer,  
month varchar(20),  
day int,  
hour int,  
weekday varchar(20));
```

```
create table redshift_etl_project.fact_atm_trans(  
trans_id integer not null distkey sortkey,  
atm_id integer,  
location_id integer,  
date_id integer,  
card_type_id integer,  
atm_status varchar(20),  
currency varchar(20),  
service varchar(10),  
transaction_amount integer,  
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));
```

## 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>

```
copy redshift_etl_project.dim_location from
's3://devilatom/dim_location/part-00000-30a9ccb8-66df-41ac-b346-890b03b1ef5e-c000.csv'
iam_role 'arn:aws:iam::587730009881:role/upgrad-redshift-s3-access'
delimiter ',' region 'us-east-1' CSV IGNOREHEADER 1;
```

```
copy redshift_etl_project.dim_atm from
's3://devilatom/dim_atm /part-00000-56907704-a52d-43db-bf4e-3f05119405de-c000.csv'
iam_role 'arn:aws:iam::587730009881:role/upgrad-redshift-s3-access'
delimiter ',' region 'us-east-1' CSV IGNOREHEADER 1;
```

```
copy redshift_etl_project.dim_card_type from
's3://devilatom/dim_card/part-00000-3d878a3c-cf0b-47ea-8112-015888e94afa-c000.csv'
iam_role 'arn:aws:iam::587730009881:role/upgrad-redshift-s3-access'
delimiter ',' region 'us-east-1' CSV IGNOREHEADER 1;
```

```
copy redshift_etl_project.dim_date from
's3://devilatom/dim_date/part-00000-bddf5d24-5a26-437f-82f2-e04b71a4f21f-c000.csv'
iam_role 'arn:aws:iam::587730009881:role/upgrad-redshift-s3-access'
delimiter ',' region 'us-east-1' CSV IGNOREHEADER 1;
```

```
copy redshift_etl_project.fact_atm_trans from
's3://devilatom/fact_atm_trans /part-00000-76fb56f0-74d6-4526-a343-b046f3433675-c000.csv'
iam_role 'arn:aws:iam::587730009881:role/upgrad-redshift-s3-access'
delimiter ',' region 'us-east-1' CSV IGNOREHEADER 1;
```