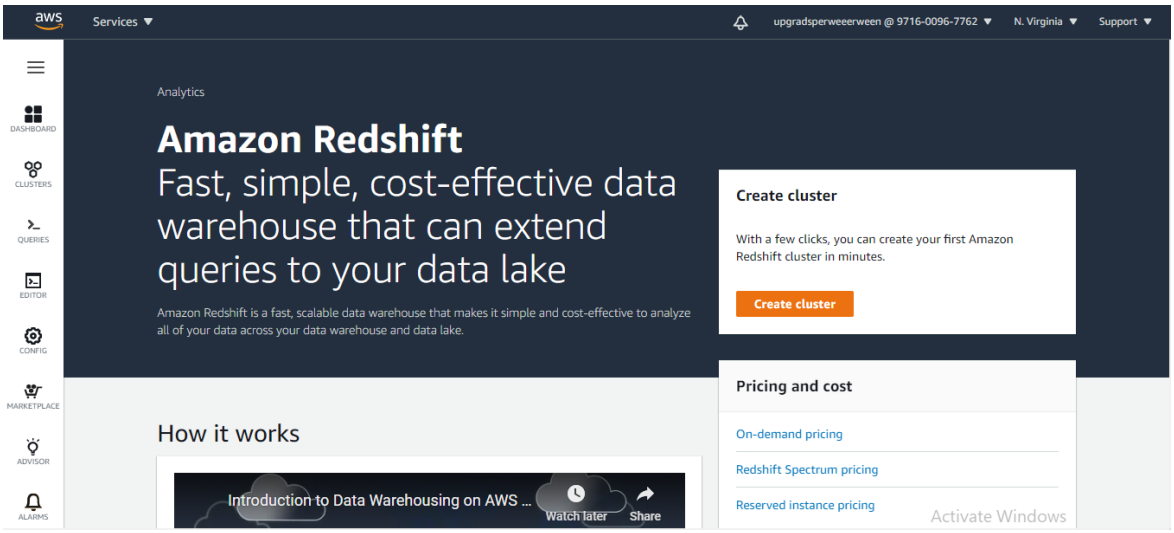
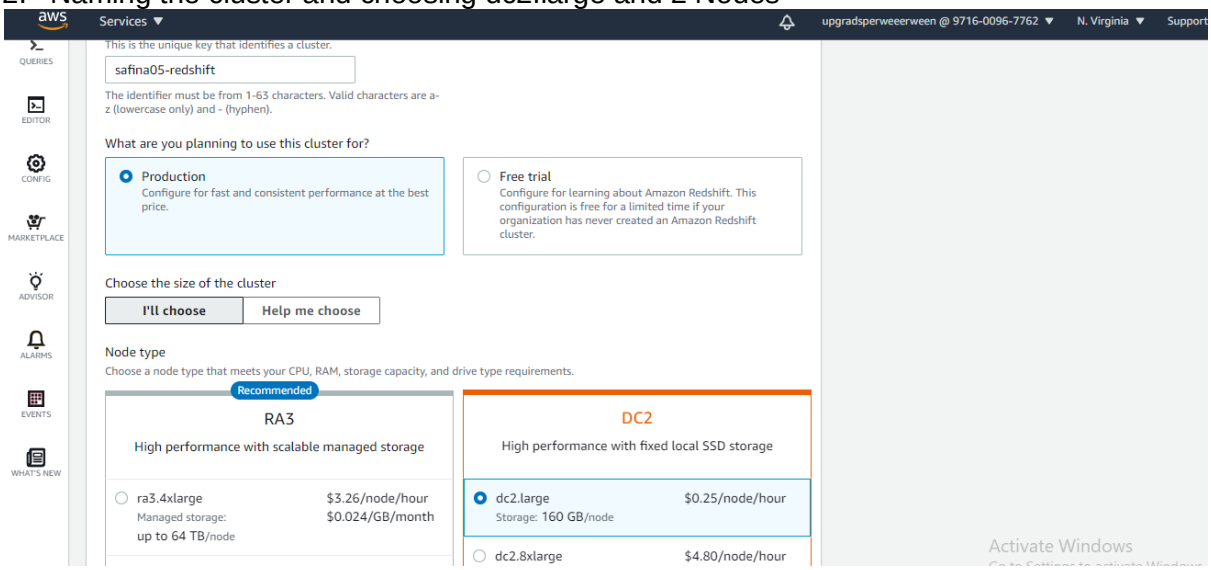
Creation of a RedShift Cluster

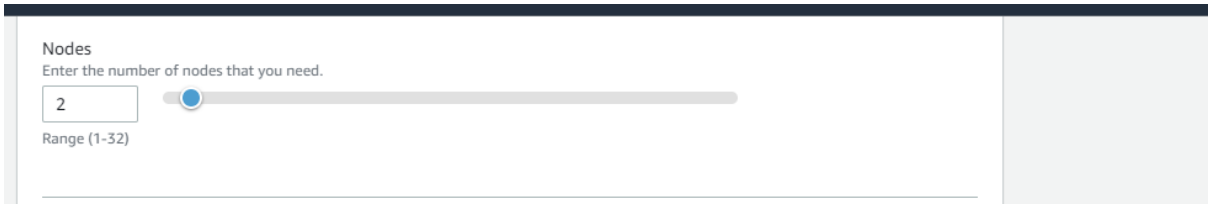
Screenshots of the configuration of the RedShift cluster that we have created:

1. Click on **Create cluster**

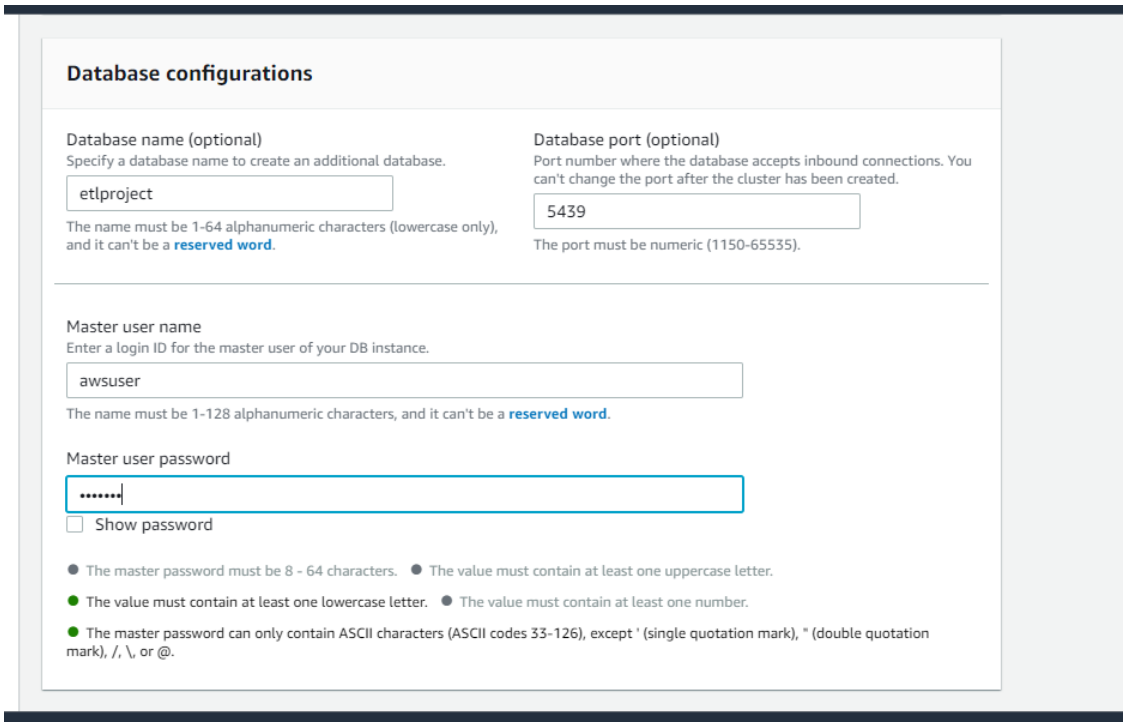


2. Naming the cluster and choosing dc2.large and 2 Nodes

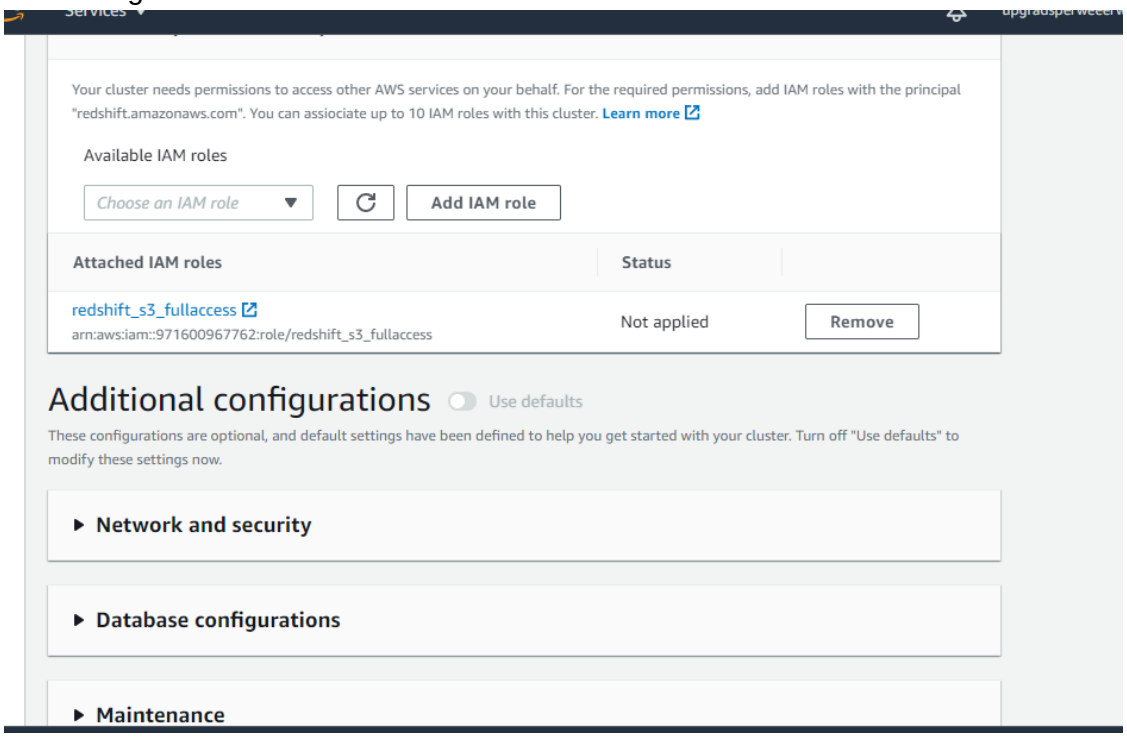




3. Configuring database



4. Adding IAM role



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:

Looking analytical query questions, we see **atm_status** has been frequently filtered on so we make a **sortkey** and the primary keys are **distkeys** and they are frequently joined on

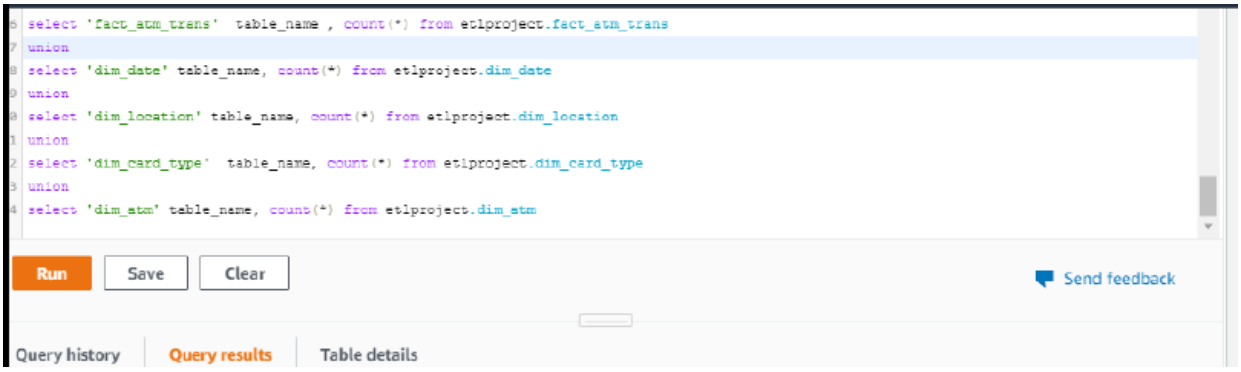
```
copy etlproject.fact_atm_trans from
's3a://safina05/FACT_ATM_TRANS/FACT_ATM_TRANS.csv'
iam_role ‘arn:aws:iam::971600967762:role/redshift_s3_fullaccess’
delimiter ',' region 'us-east-1' REMOVEQUOTES;

copy etlproject.DIM_CARD_TYPE from
's3a://safina05/DIM_CARD_TYPE/DIM_CARD_TYPE.csv'
iam_role ‘arn:aws:iam::971600967762:role/redshift_s3_fullaccess’
delimiter ',' region 'us-east-1';

copy etlproject.dim_location from
's3a://safina05/DIM_LOCATION/DIM_LOCATION.csv'
iam_role ‘arn:aws:iam::971600967762:role/redshift_s3_fullaccess’
delimiter ',' region 'us-east-1';

delete from etlproject.dim_date
copy etlproject.dim_date from
's3a://safina05/DIM_DATE/DIM_DATE.csv'
iam_role ‘arn:aws:iam::971600967762:role/redshift_s3_fullaccess’
delimiter ',' region 'us-east-1';
```

Cross verifying the counts of loaded data



Rows returned (5)

Search rows

table_name	count
dim_date	8685
fact_atm_trans	2468572
dim_atm	156
dim_card_type	12
dim_location	109