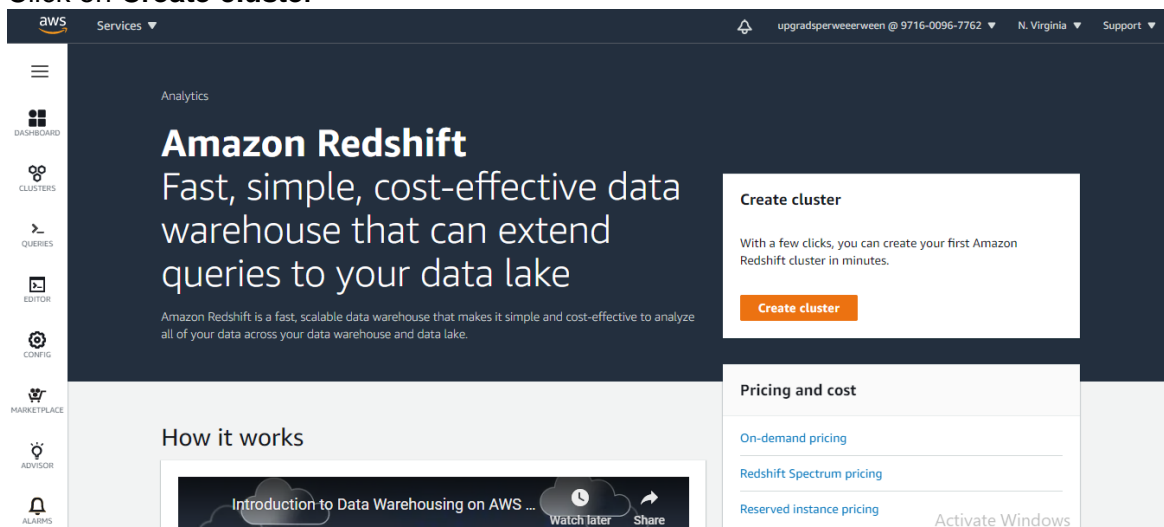


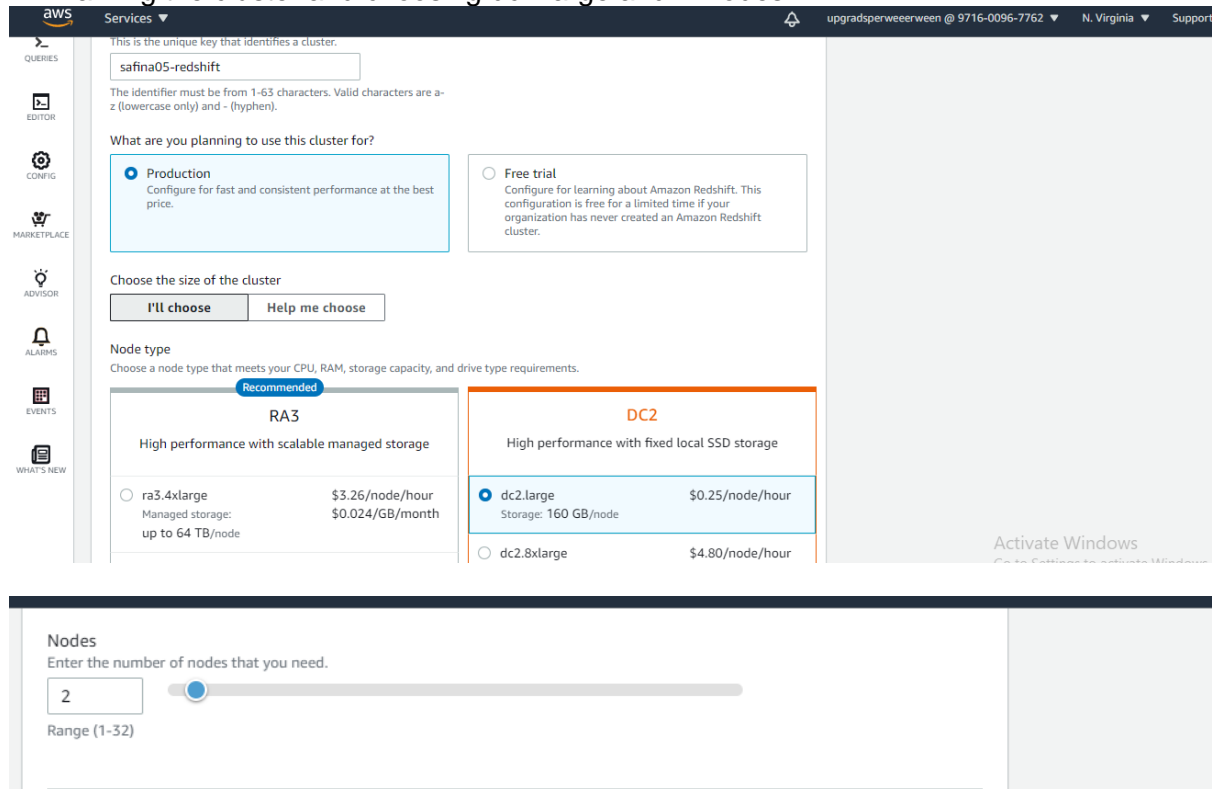
Creation of a RedShift Cluster

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

Database configurations

Database name (optional)
Specify a database name to create an additional database.

The name must be 1-64 alphanumeric characters (lowercase only), and it can't be a **reserved word**.

Database port (optional)
Port number where the database accepts inbound connections. You can't change the port after the cluster has been created.

The port must be numeric (1150-65535).

Master user name
Enter a login ID for the master user of your DB instance.

The name must be 1-128 alphanumeric characters, and it can't be a **reserved word**.

Master user password

☐ Show password

- The master password must be 8 - 64 characters.
- The value must contain at least one uppercase letter.
- The value must contain at least one lowercase letter.
- The value must contain at least one number.
- The master password can only contain ASCII characters (ASCII codes 33-126), except ' (single quotation mark), " (double quotation mark), /, \, or @.

4. Adding IAM role

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 10 IAM roles with this cluster. [Learn more](#)

Available IAM roles

Attached IAM roles	Status	
redshift_s3_fullaccess arn:aws:iam::971600967762:role/redshift_s3_fullaccess	Not applied	<input type="button" value="Remove"/>

Additional configurations ☒ Use defaults

These configurations are optional, and default settings have been defined to help you get started with your cluster. Turn off "Use defaults" to modify these settings now.

▶ Network and security

▶ Database configurations

▶ Maintenance

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:

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

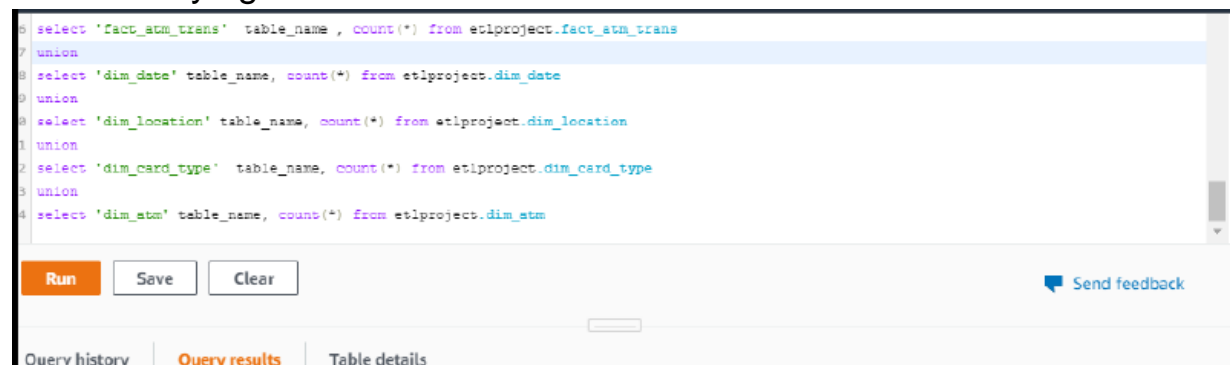
```
copy etlproject.fact_atm_trans from
's3a://safina05/FACT_ATM_TRANS/FACT_ATM_TRANS.csv'
iam_role 'arn:aws:iam::971600967762:role/redshift_s3_fullaccess'
delimiter ',' region 'us-east-1' REMOVEQUOTES;
```

```
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)

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