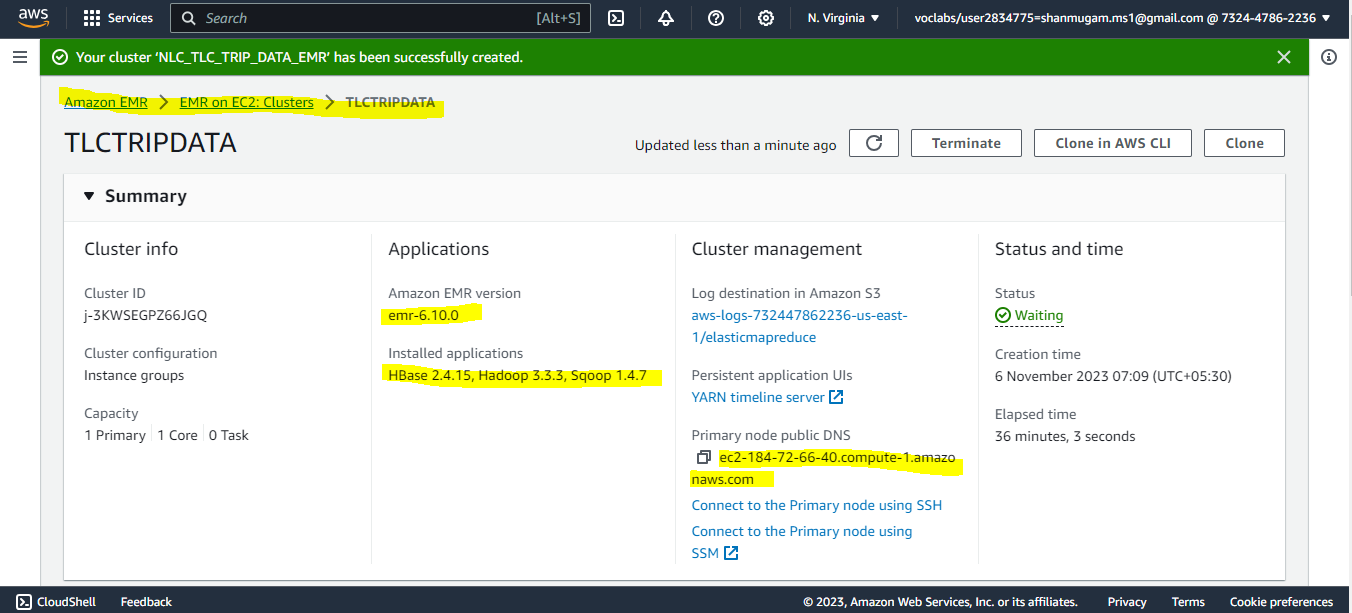
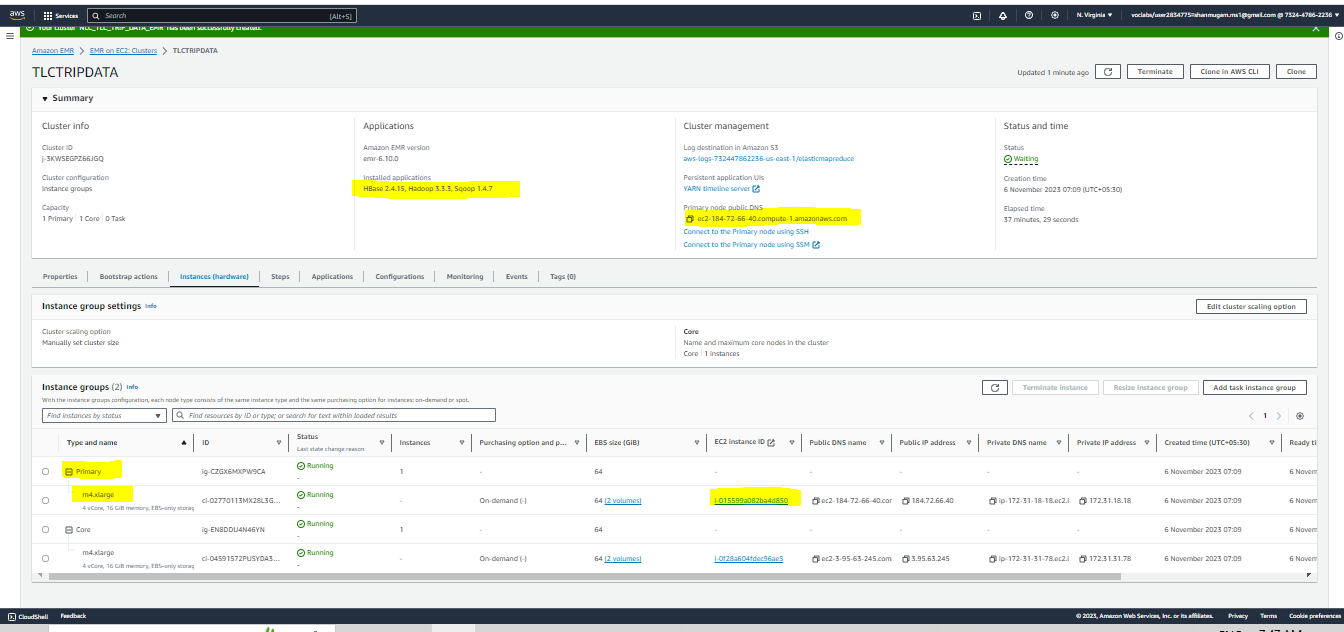
**Task 1:** Create an RDS instance in your AWS account and upload the data to the RDS instance (Note: Instructions on how to work with RDS can be found[**here.)**](https://cdn.upgrad.com/uploads/production/08f82196-b083-426a-ab25-1eb052c43683/Documentation%2B-%2BWorking%2Bwith%2BRDS.pdf)

Since the dataset is huge, you need to upload the data from only two files (i.e. **yellow\_tripdata\_2017-01.csv & yellow\_tripdata\_2017-02.csv)**from the dataset.

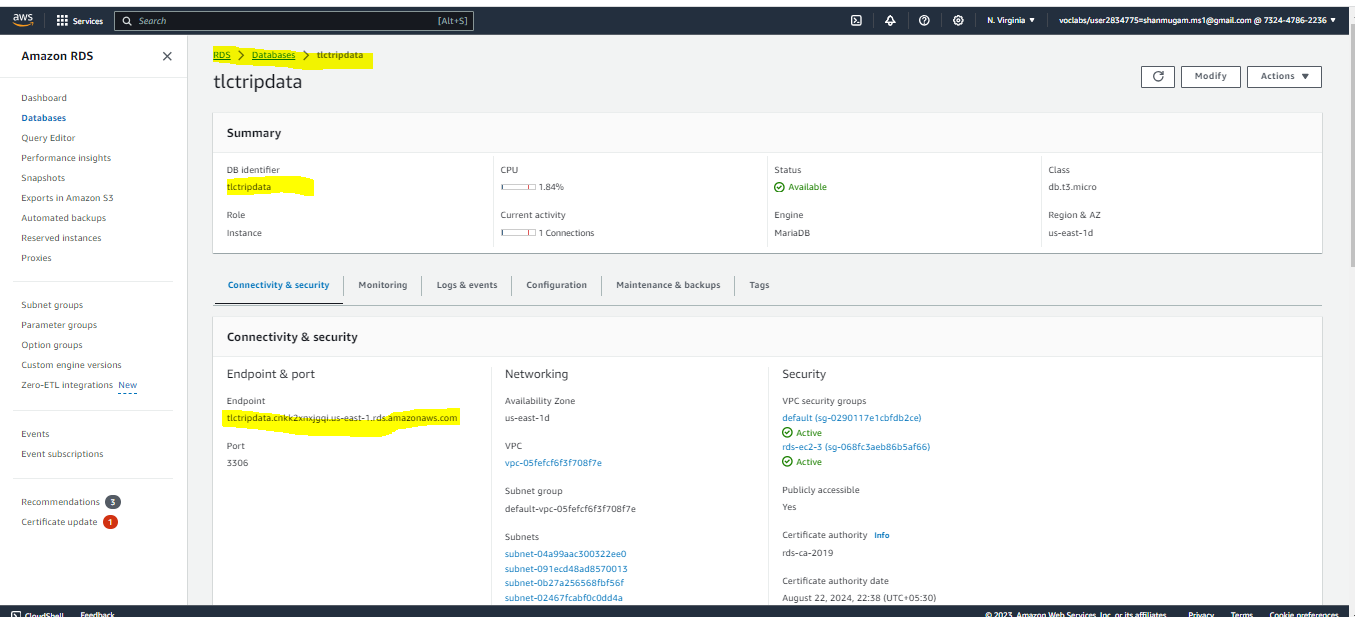
**Note: You will need to create an appropriate schema for the data sets to upload them to RDS (you can find the data dictionary in the previous segments. The steps to work with RDS in given in the Additional Resource).**

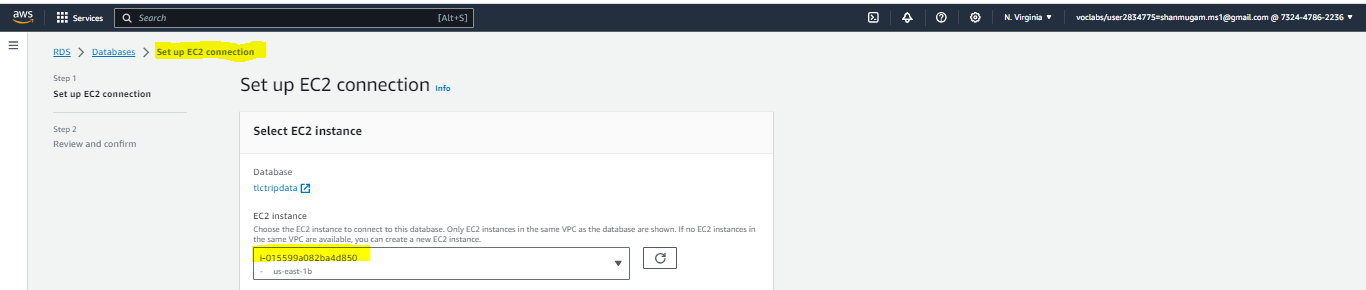
**EMR creation:**

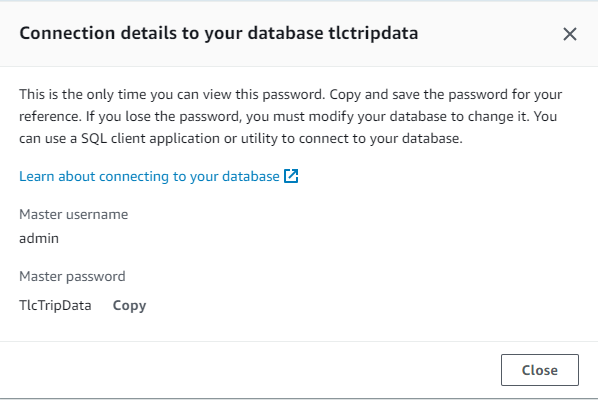


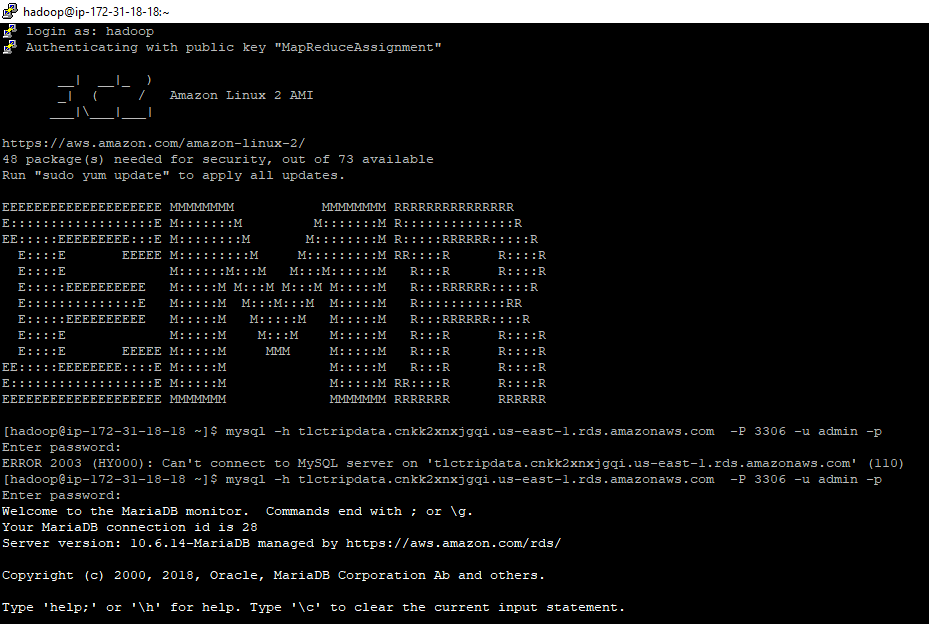


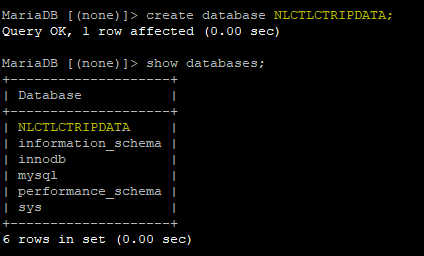
**RDS Instance creation:**











create database NLCTLCTRIPDATA;

use NLCTLCTRIPDATA;

CREATE TABLE yellow\_taxi\_trip\_data( VendorID INT, tpep\_pickup\_datetime VARCHAR(50), tpep\_dropoff\_datetime VARCHAR(50), Passenger\_count INT, Trip\_distance FLOAT, PULocationID INT, DOLocationID INT, RateCodeID INT, Store\_and\_fwd\_flag VARCHAR(2), payment\_type INT, Fare\_amount FLOAT, Extra FLOAT, MTA\_tax FLOAT, improvement\_surcharge FLOAT, Tip\_amount FLOAT, Tolls\_amount FLOAT, total\_amount FLOAT, Congestion\_Surcharge FLOAT, Airport\_fee FLOAT );

wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\_tripdata\_2017-01.csv;

wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\_tripdata\_2017-02.csv;

LOAD DATA LOCAL INFILE '/home/hadoop/yellow\_tripdata\_2017-01.csv' INTO TABLE yellow\_taxi\_trip\_data FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

LOAD DATA LOCAL INFILE '/home/hadoop/yellow\_tripdata\_2017-02.csv' INTO TABLE yellow\_taxi\_trip\_data FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

select count(\*) from yellow\_taxi\_trip\_data;

