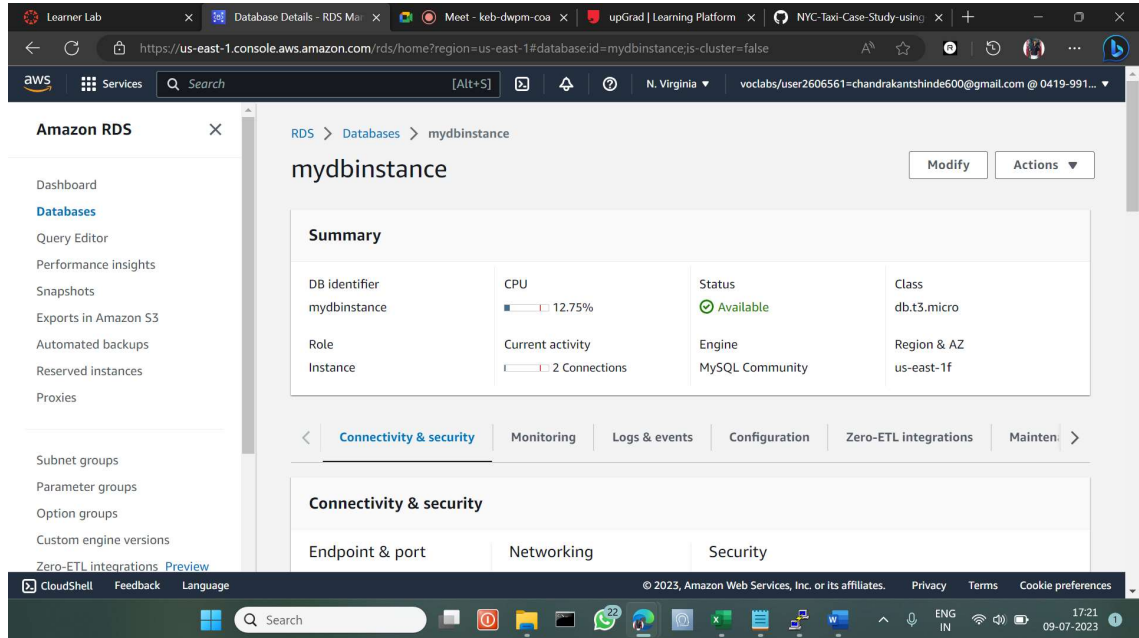
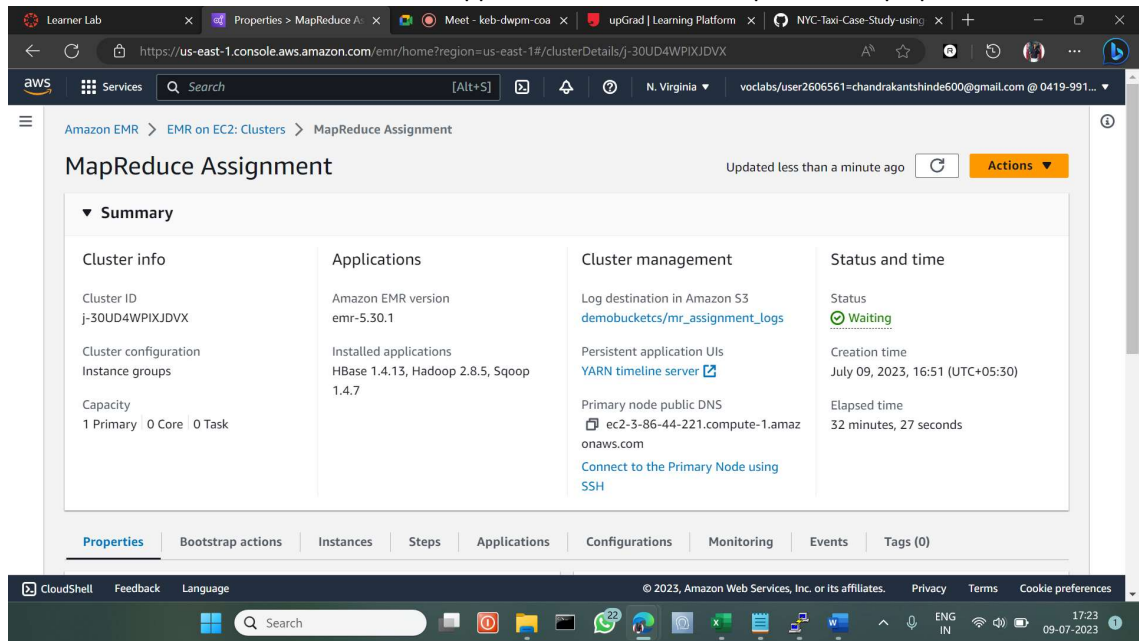


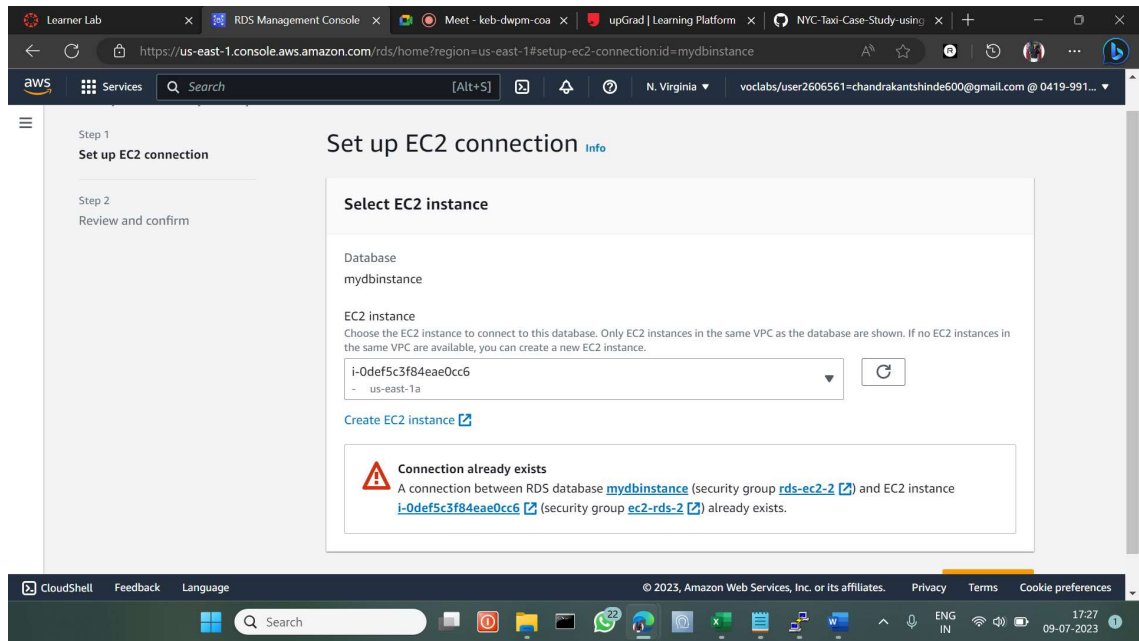
## 1. Creation of RDS instance in AWS



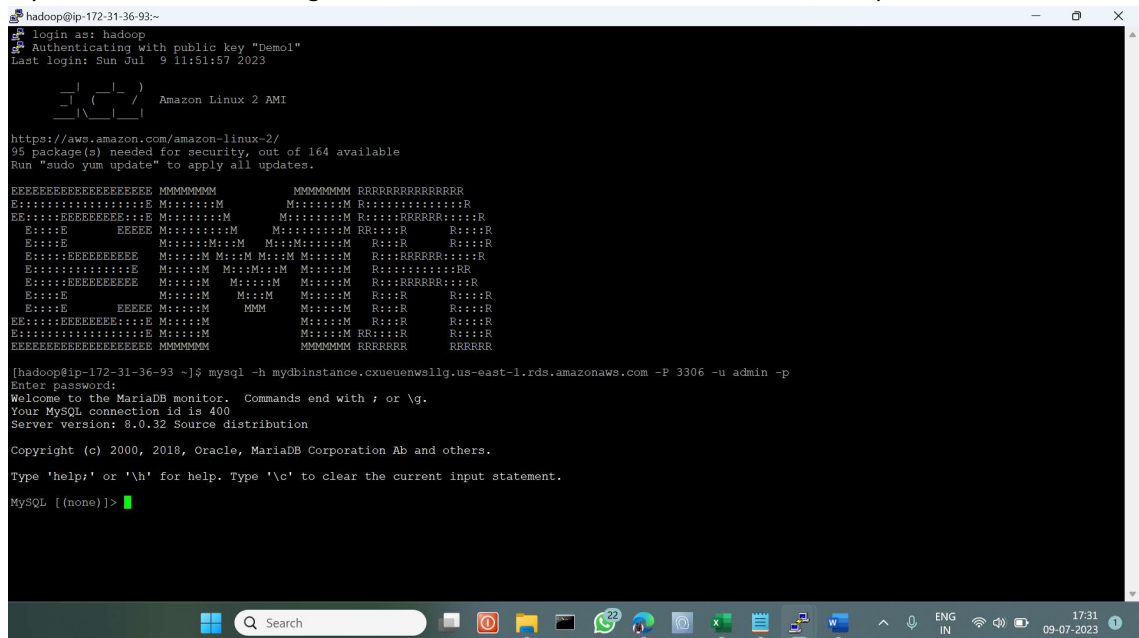
## 2. Creation of EMR instance with bundled applications such as Hadoop, Hbase, Sqoop.



3. To connect RDS with EMR instance, we have to click on "Action" button on RDS instance menu and then "Set up an EC2 Instance".



4. Login to RDS through EMR instance using command: ``mysql -h mydbinstance.cxueuenwslg.us-east-1.rds.amazonaws.com -P 3306 -u admin -p``



5. Creation of Database "taxi\_records" and table "trip\_log"  
``create database taxi_records``  
``CREATE TABLE trip_log`  
(VendorID INT, tpep\_pickup\_datetime VARCHAR(50), tpep\_dropoff\_datetime VARCHAR(50), Passenger\_count INT, Trip\_distance FLOAT, RatecodeID INT, store\_and\_fwd\_flag VARCHAR(2), PULocationID INT, DOLocationID INT, payment\_type INT, fare\_amount FLOAT, extra FLOAT, mta\_tax FLOAT, tip\_amount FLOAT, tolls\_amount FLOAT, improvement\_surcharge FLOAT, total\_amount FLOAT, Airport\_fee FLOAT );'

6. Downloading required csv files from internet in local using command  

```
` 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`
```
7. To load data in mysql table we have to login and then run sql command:  

```
LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv' INTO TABLE trip_log  
FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
```

```
LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-02.csv' INTO TABLE trip_log
FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
```

```
SELECT COUNT(*) FROM taxi_records.trip_log;
```

[illegible]