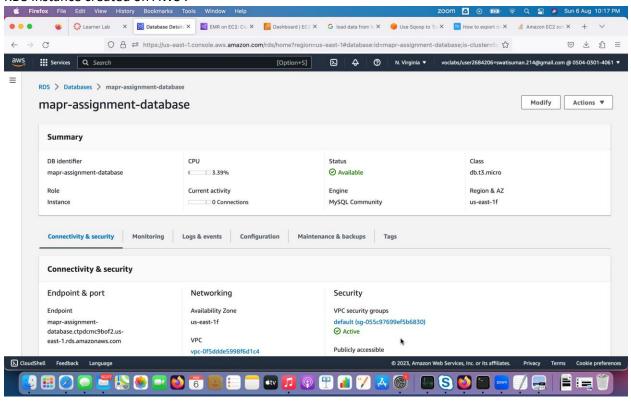
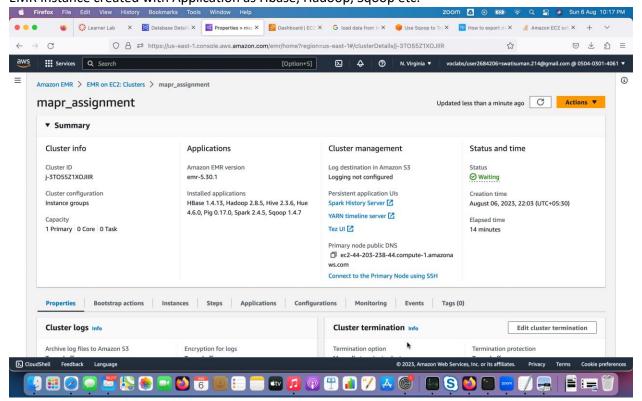
1. RDS Instance created on AWS:



2. EMR Instance created with Application as Hbase, Hadoop, Sqoop etc.

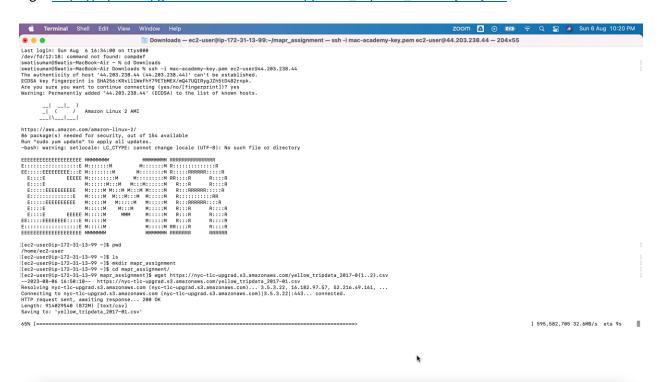


3. Open EMR Instance on Terminal, Create a new Directory and Download NYC data file in new directory using below command:

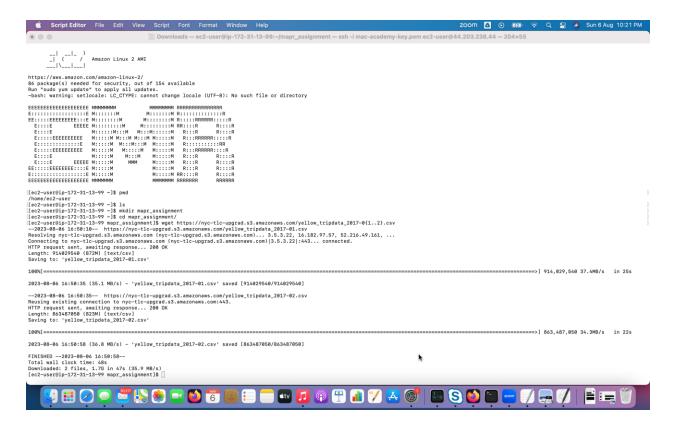
// to make new directory
mkdir mapr_assignment
// move to new directory
cd mapr_assignment

//download NYC data files

wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-0{1..2}.csv







Note: Above command will download first 2 files sequentially.

4. Connect to RDS on EMR instance through terminal using below command:

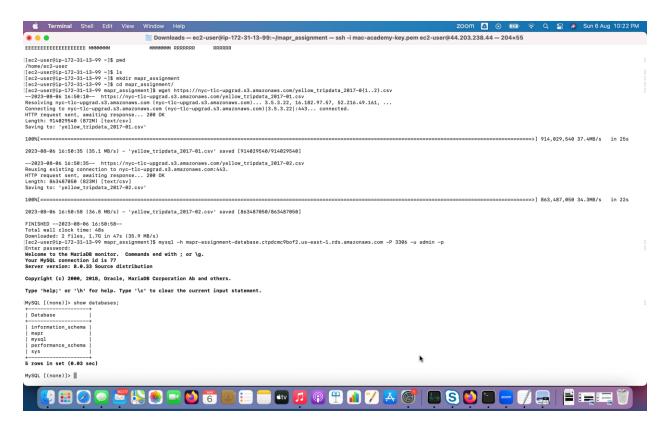
```
//Connect RDS on EMR
mysql -h mapr-assignment-database.ctpdcmc9bof2.us-east-1.rds.amazonaws.com -P 3306 -u
admin -p

//to list database
show databases;

//to create database
create database <database_name>

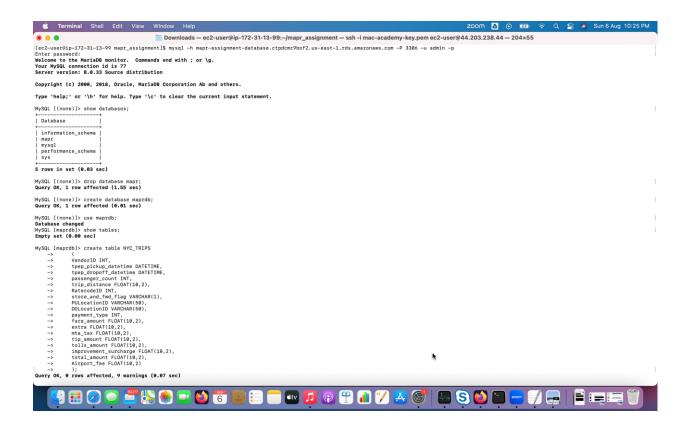
//to go to db
use <database_name>;

//to list tables
show tables;
```



Create New table with name NYC_TRIPS create table NYC_TRIPS

```
VendorID INT,
tpep_pickup_datetime DATETIME,
tpep_dropoff_datetime DATETIME,
passenger count INT,
trip distance FLOAT(10,2),
RatecodeID INT,
store and fwd flag VARCHAR(1),
PULocationID VARCHAR(50),
DOLocationID VARCHAR(50),
payment_type INT,
fare_amount FLOAT(10,2),
extra FLOAT(10,2),
mta_tax FLOAT(10,2),
tip amount FLOAT(10,2),
tolls amount FLOAT(10,2),
improvement surcharge FLOAT(10,2),
total amount FLOAT(10,2),
Airport_fee FLOAT(10,2)
);
```



6. Load Data from File 1 and File 2 to RDS Database

LOAD DATA LOCAL INFILE '/home/ec2-user/mapr_assignment/yellow_tripdata_2017-01.csv' INTO TABLE NYC_TRIPS FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

LOAD DATA LOCAL INFILE '/home/ec2-user/mapr_assignment/yellow_tripdata_2017-02.csv' INTO TABLE NYC_TRIPS FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

//check data in database

Select * from NYC_TRIPS LIMIT 10;

Select count(*) from NYC_TRIPS;

