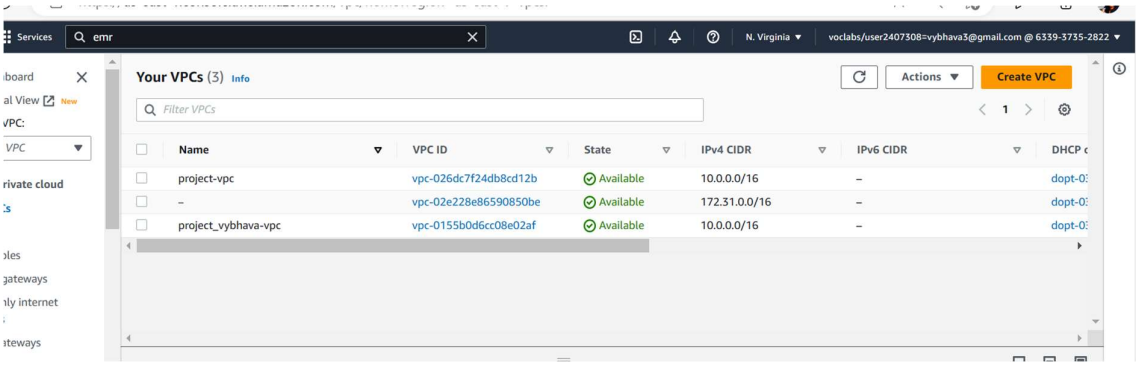
MapReduce - Programming Assignment

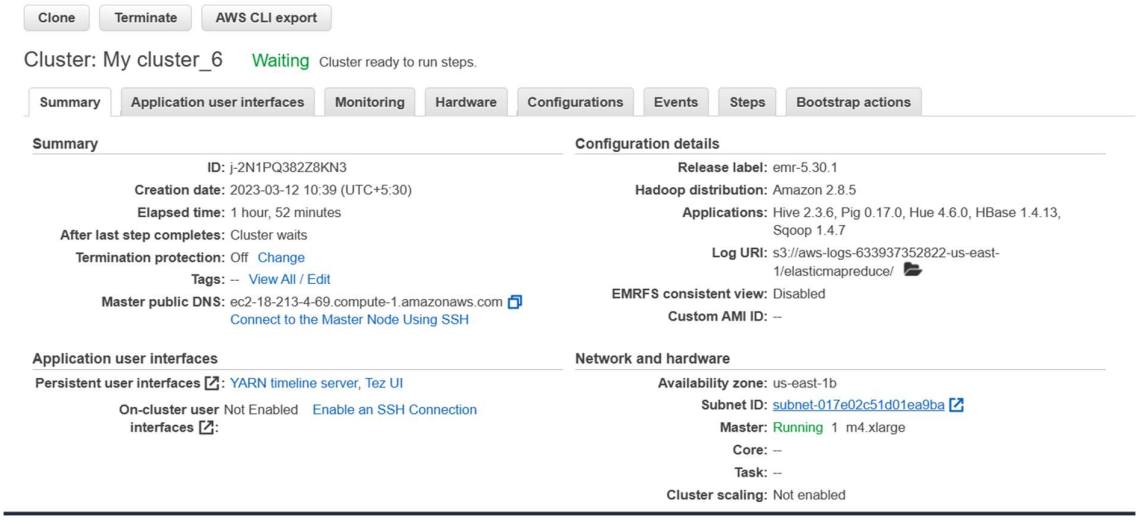
Task 1. Create an RDS instance in your AWS account and upload the data to the RDS instance.

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.

1. CREATE VPC (Amazon virtual private cloud)

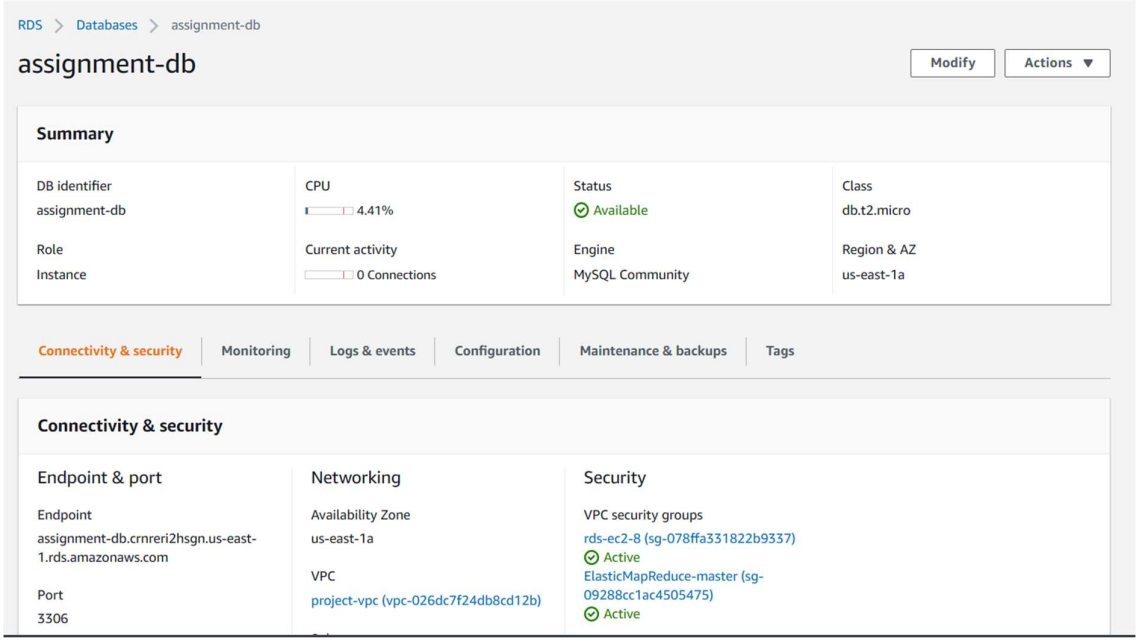


2. CREATE EMR Instance (Creation of amazon elastic map reduce instance)




login as: hadoop
Authenticating with public key "keyppk"
Last login: Sun Mar 12 07:16:49 2023
Amazon Linux 2 AMI
https://aws.amazon.com/amazon-linux-2/
[hadoop@ip-10-0-21-165 ~]$ sudo -i
[root@ip-10-0-21-165 ~]# hbase shell

3. CREATE RDS DATABASE (Creation of RDS database)



4. DOWNLOADING THE DATA /FILES TO EMR CLUSTER:

Code:

a. wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
b. wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv

5. CONNECTING TO RDS FROM EMR CLUSTER

Code:

```
mysql -h assignment-db.crnreri2hsgn.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
```

6. CREATING THE DATA BASE DEMO AND TABLE TLCTRIPDATA IN RDS.

```
create database demo;

use demo;

CREATE TABLE TLCTripData

 (VendorID int,

tpep_pickup_datetime datetime,

tpep_dropoff_datetime datetime,

passenger_count int,

trip_distance float,

RatecodeID int,
```

store_and_fwd_flag char,

PULocationID int,

DOLocationID int,

payment_type int,

fare_amount float,

extra float,

mta_tax float,

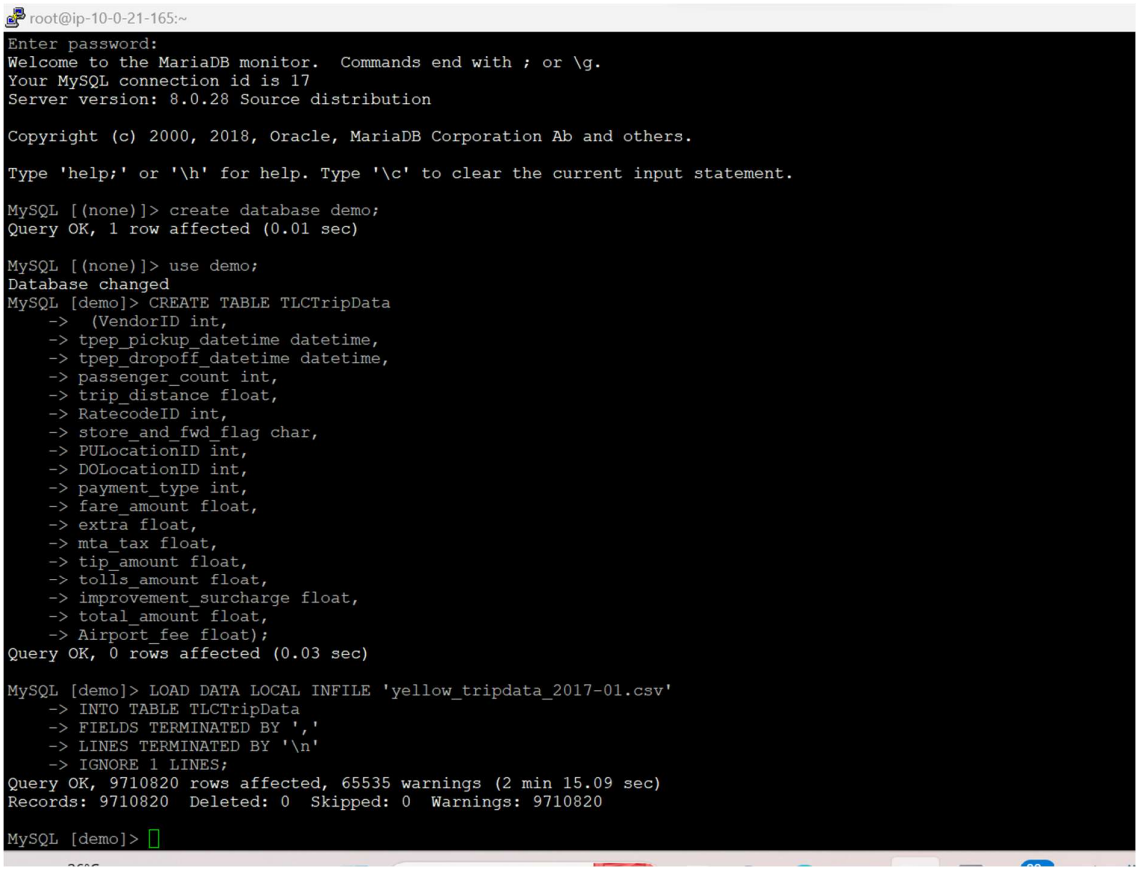
tip_amount float,

tolls_amount float,

improvement_surcharge float,

total_amount float,

Airport_fee float);



7. LOADING THE DATA TO RDS TABLE : TLCTRIPDATA

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

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

```
MySQL [demo]> LOAD DATA LOCAL INFILE 'yellow_tripdata_2017-01.csv'
    -> INTO TABLE TLCTripData
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 LINES;

Query OK, 9710820 rows affected, 65535 warnings (2 min 13.18 sec)
Records: 9710820  Deleted: 0  Skipped: 0  Warnings: 9710820

MySQL [demo]>
MySQL [demo]> LOAD DATA LOCAL INFILE 'yellow_tripdata_2017-02.csv'
    -> INTO TABLE TLCTripData
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 LINES;
Query OK, 9169775 rows affected, 65535 warnings (2 min 50.66 sec)
Records: 9169775  Deleted: 0  Skipped: 0  Warnings: 9169775

MySQL [demo]> select count(*) from TLCTripData;
+----------+
| count(*) |
+----------+
| 18880595 |
+----------+
1 row in set (42.95 sec)

MySQL [demo]>
```

Assignment Submitted by:

Susil Patro, Vybhava P, Vivek Agrawal