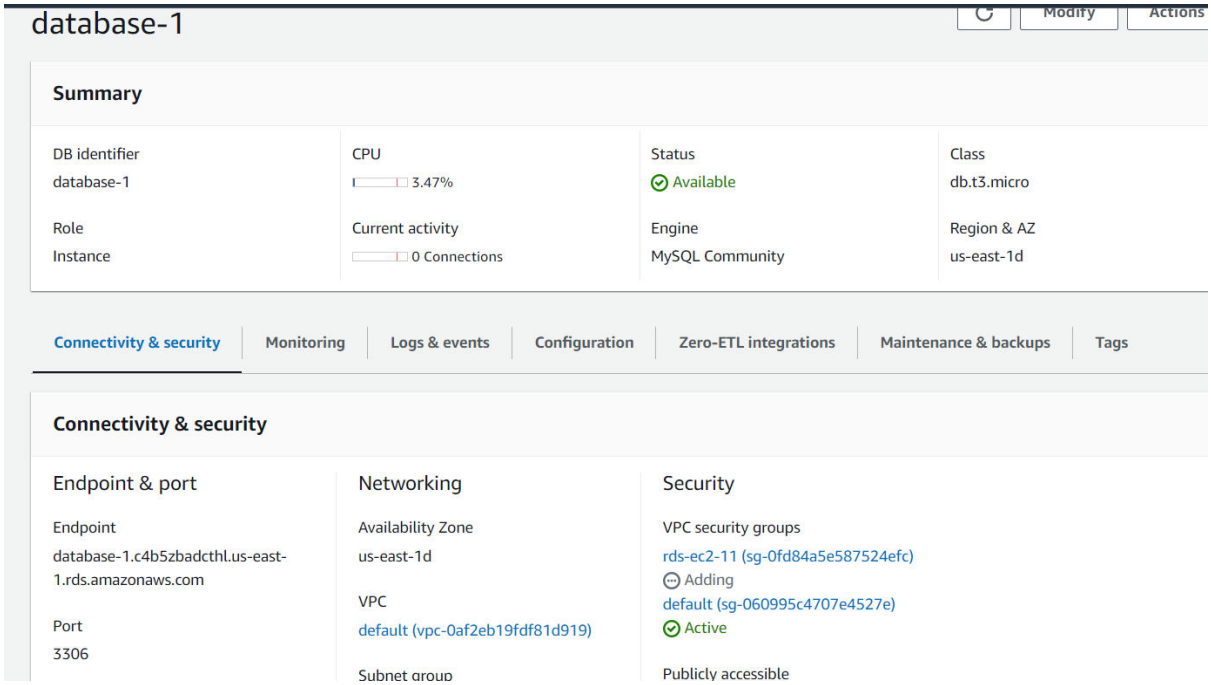
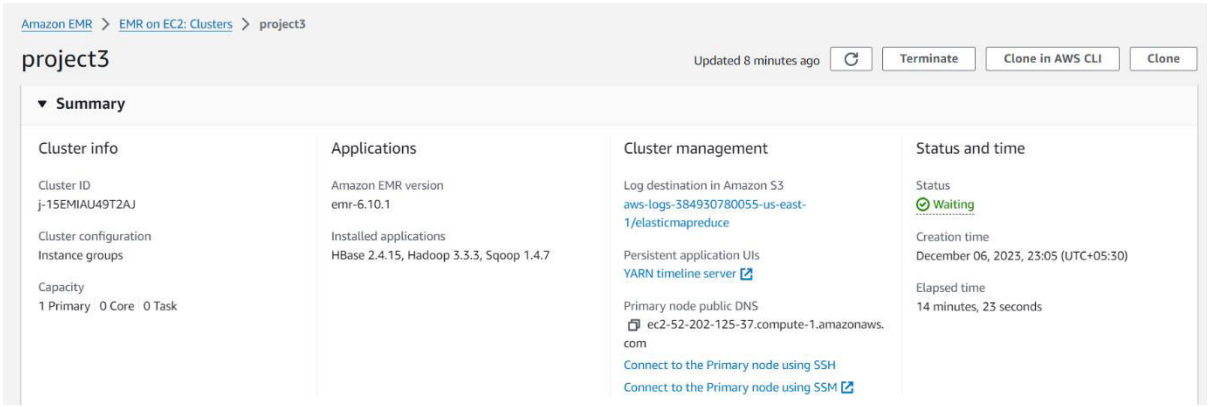
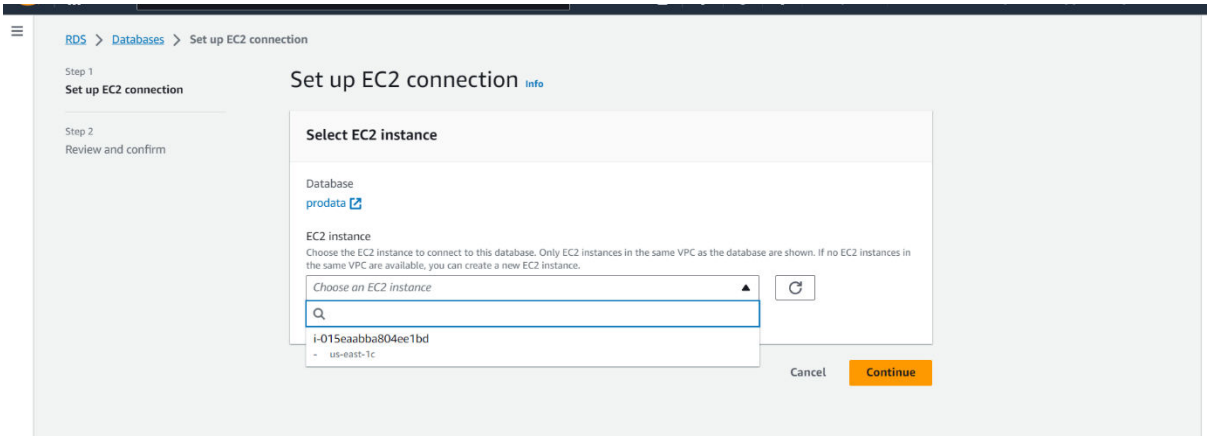
1. Creation of RDS instance called database-1

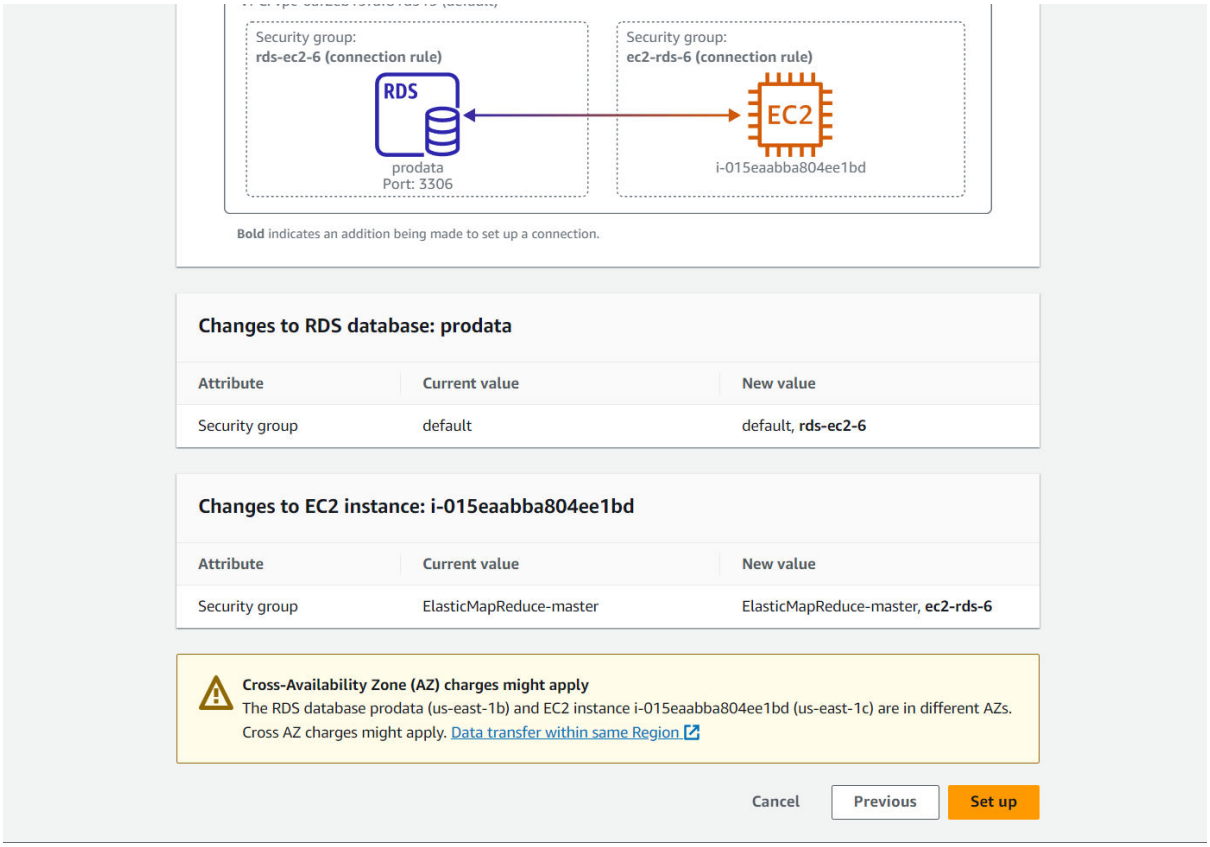


2. Creation of EMR with Hadoop, Hbase and Scoop

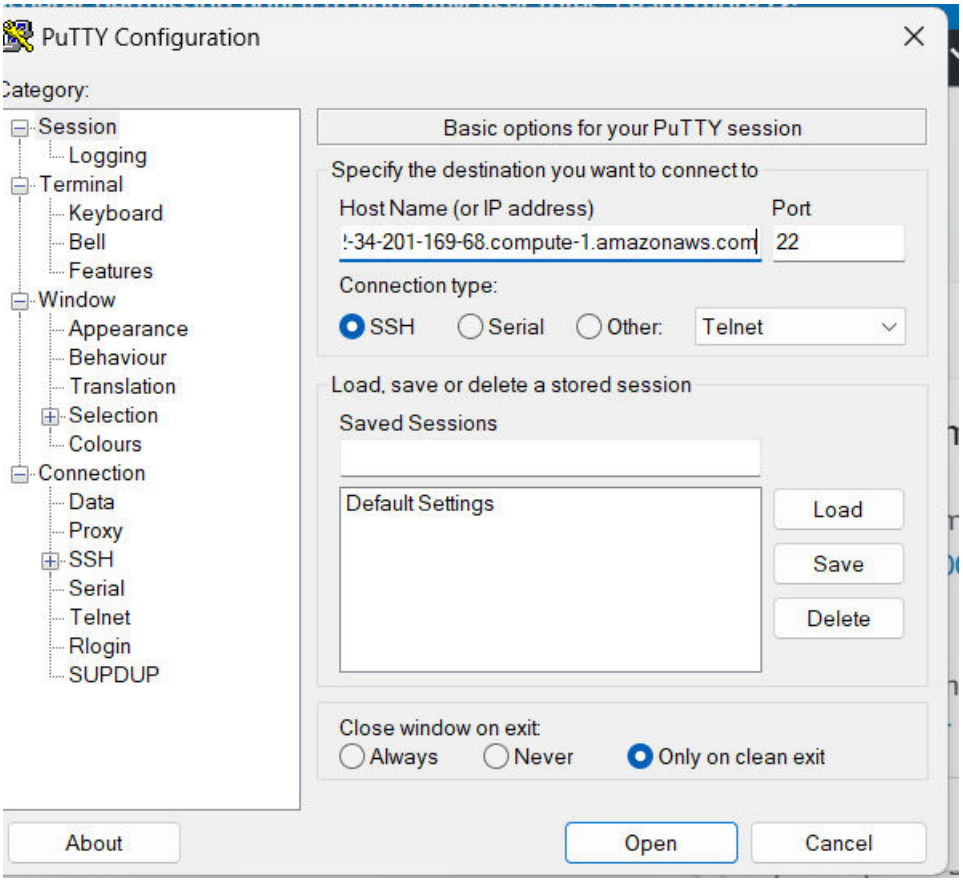


3. Connect RDS with EMR





4. **Connect emr instance through PuTTY**



5. **Login to MYSQL using RDS endpoint**

6. **Creation of Database “taxi_records” and table “trip_log”**

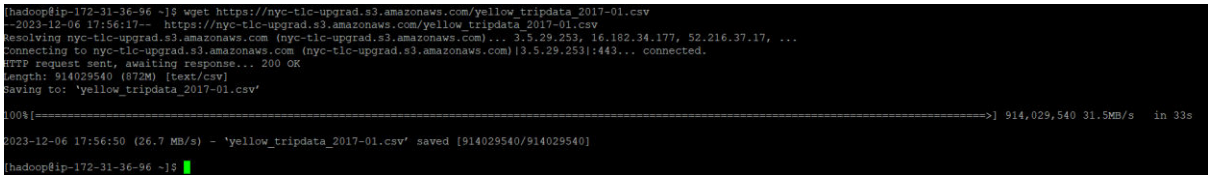
```
CREATE DATABASE taxi_records;

USE 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
);
```

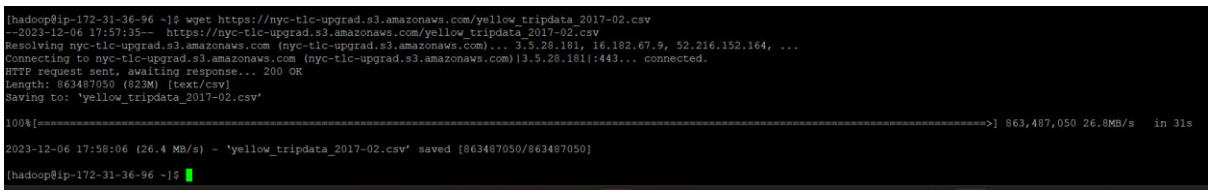
```
MySQL [(none)]> use taxi_records;
Database changed
MySQL [taxi_records]> CREATE TABLE trip_log
    -> (
    -> VendorID INT,
    -> tpep_pickup_datetime VARCHAR(50),
    -> tpep_dropoff_datetime VARCHAR(50),
    -> Passenger_count INT,
    -> Trip_distance FLOAT,
    -> RatecodeID INT,
    -> store_and_fwd_flag 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
    -> );
Query OK, 0 rows affected (0.04 sec)

MySQL [taxi_records]>
```

7. **Downloading required csv files from internet in local using command**

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

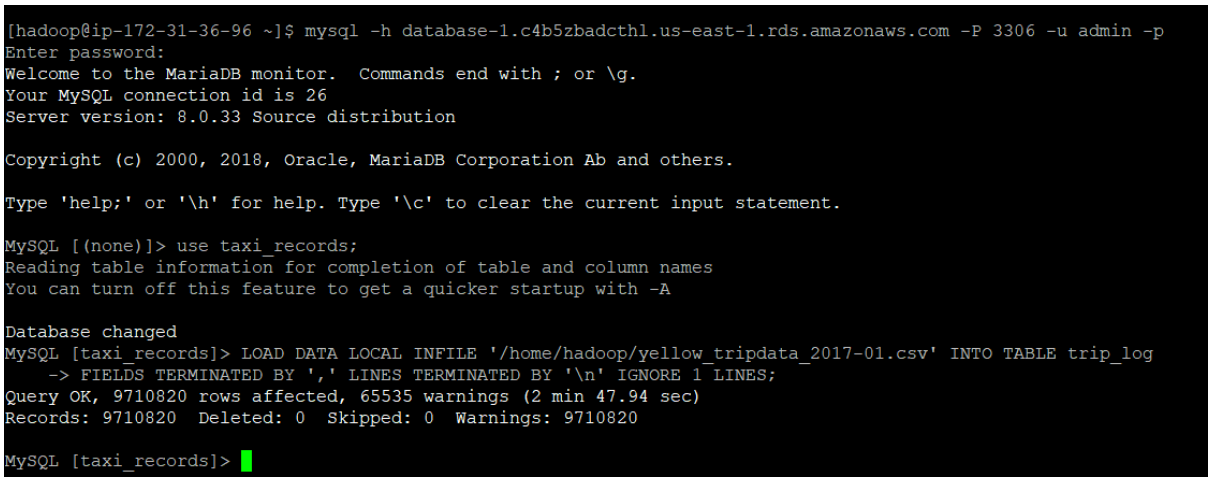


wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv

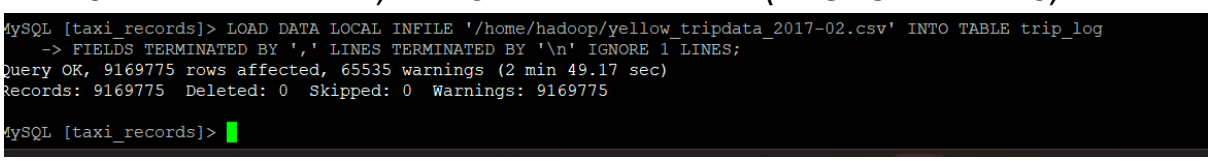


8. **To load data in mysql table we have to login and then run sql command:**

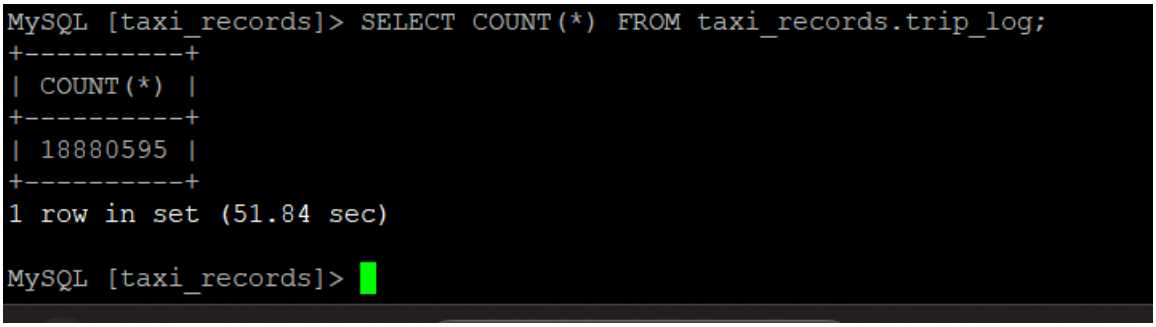
LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv'
INTO TABLE trip_log
FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;



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;


MySQL [taxi_records]> SELECT COUNT(*) FROM taxi_records.trip_log;
+----------+
| COUNT(*) |
+----------+
| 18880595 |
+----------+
1 row in set (51.84 sec)
MySQL [taxi_records]>