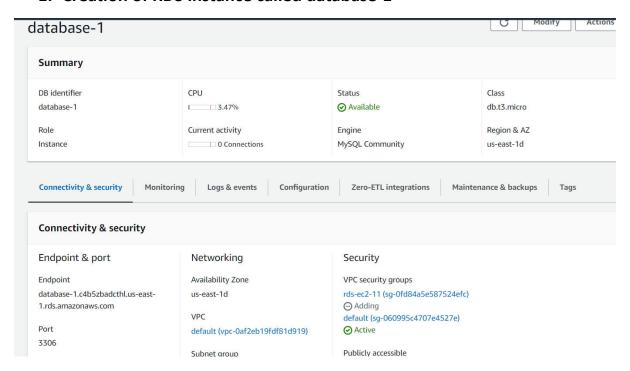
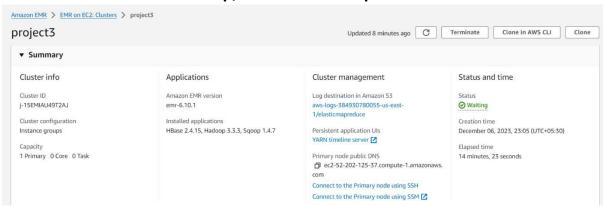
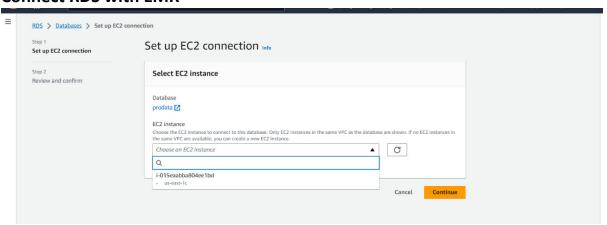
1. Creation of RDS instance called database-1

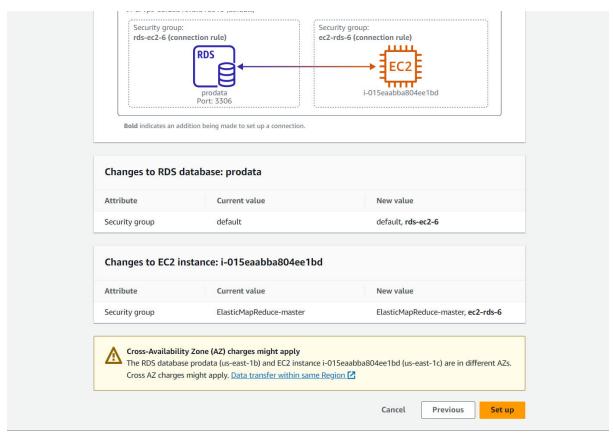


2. Creation of EMR with Hadoop, Hbase and Scoop

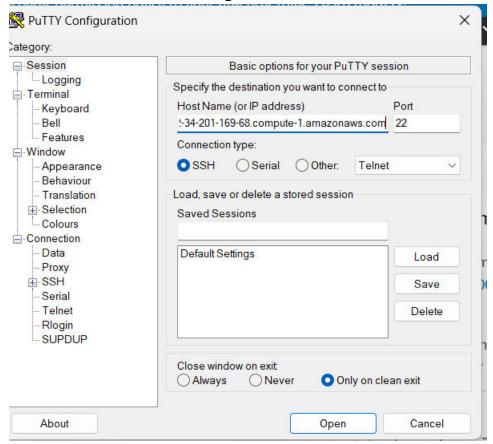


3. Connect RDS with EMR





4. Connect emr instance through PuTTY



5. Login to MYSQL using RDS endpoint

```
hadoop@ip-172-31-36-96:~
     login as: hadoop
     Authenticating with public key "imported-openssh-key"
   Last login: Wed Dec 6 17:47:57 2023
          ####
                     Amazon Linux 2
                      AL2 End of Life is 2025-06-30.
           \###|
                      A newer version of Amazon Linux is available!
                      Amazon Linux 2023, GA and supported until 2028-03-15.
                       https://aws.amazon.com/linux/amazon-linux-2023/
   26 package(s) needed for security, out of 35 available
   Run "sudo yum update" to apply all updates.
   EEEEEEEEEEEEEEEEE MMMMMMMM
                                      M::::::: M R:::::::::::::R
   EE:::::EEEEEEEEE:::E M:::::::M
                                    M:::::::M R:::::RRRRRR:::::R
                                   M:::::::M RR::::R
    E::::E
              EEEEE M:::::::M
    E::::E
                     R::::R
    E:::::EEEEEEEEE M::::M M:::M M::::M R:::RRRRRR:::::R
    E:::::::::E M:::::M M::::M M:::::M R::::::RR
    E:::::EEEEEEEEE M:::::M M:::::M M:::::M
                                               R:::RRRRRR::::R
    E::::E
                     M:::::M
                                       M:::::M
                                                R:::R
                                                         R::::R
     E::::E
                EEEEE M:::::M
                                       M:::::M
                                                R:::R
                                                          R::::R
   EE:::::EEEEEEEE::::E M:::::M
                                       M:::::M
                                                R:::R
   M:::::M RR::::R
                                                          R::::R
   EEEEEEEEEEEEEEEEE MMMMMMM
                                       MMMMMMM RRRRRRR
                                                          RRRRRR
   [hadoop@ip-172-31-36-96 ~]$ mysql -h database-1.c4b5zbadcthl.us-east-1.rds.amazo
   naws.com -P 3306 -u admin -p
   Enter password:
   Welcome to the MariaDB monitor. Commands end with ; or \g.
   Your MySQL connection id is 23
   Server version: 8.0.33 Source distribution
   Copyright (c) 2000, 2018, Oracle, MariaDB Corporation Ab and others.
   Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
   MySQL [(none)]>
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

```
| Thadoop@ip-172-31-36-96 -]$ wget https://nyc-tlc-upgrad.s3.amazonavs.com/yellow_tripdata_2017-01.csv -_2023-12-06_17:56:17-- https://nyc-tlc-upgrad.s3.amazonavs.com/yellow_tripdata_2017-01.csv -_2023-12-06_17:56:17-- https://nyc-tlc-upgrad.s3.amazonavs.com/yellow_tripdata_2017-01.csv -_2023-12-06_17:56:17-- https://nyc-tlc-upgrad.s3.amazonavs.com/yellow_tripdata_2017-01.csv -_2020-12-06_17:56:17-- https://nyc-tlc-upgrad.s3.amazonavs.com/yellow_tripdata_2017-01.csv -_2020-12-06_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56:50_17:56
```

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;

```
4ySQL [taxi_records]> 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;
Query OK, 9169775 rows affected, 65535 warnings (2 min 49.17 sec)
Records: 9169775 Deleted: 0 Skipped: 0 Warnings: 9169775

4ySQL [taxi_records]>
```

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]>
```