

## **MapReduce - Programming Assignment**

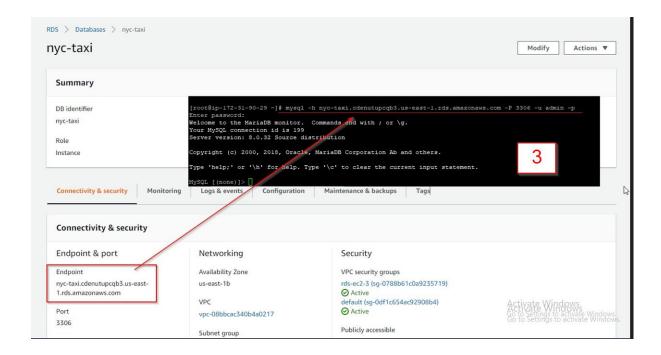
AMR\_RDS\_Task



Dharan R Sanjana Dutta Punith Vadla

## **AWS RDS on EMR Cluster**

```
| Padop@ip-172-31-81-233-/dataset | Padop@ip-172-31-81-233-|5 kd/ /nome/hadop/dataset | Padop@ip-172-31-81-233-|5 kd/ /nom
```



## mysql -h nyc-taxi.cqwodvedrqu4.us-east-1.rds.amazonaws.com -P 3306 -u admin -p

show databases;

create database assignmentdb;

use assignmentdb;

show tables;

```
MySQL [(none)]> show databases;
Database
| information schema |
| performance_schema |
sys
4 rows in set (0.03 sec)
MySQL [(none)]> CREATE DATABASE yellowtaxi;
Query OK, 1 row affected (0.01 sec)
MySQL [(none)]> USE yellowtaxi;
Database changed
MySQL [yellowtaxi]> CREATE TABLE tripdata
    -> VendorID varchar(10),
    -> tpep_pickup_datetime datetime,
   -> tpep_dropoff_datetime datetime,
    -> passenger_count INT,
    -> trip_distance FLOAT,
    -> RatecodeID VARCHAR(10),
   -> store and fwd flag VARCHAR(10),
   -> PULocationID VARCHAR(100),
    -> DOLocationID VARCHAR(100),
   -> payment_type VARCHAR (10),
    -> fare_amount FLOAT,
    -> extra FLOAT,
    -> mta tax FLOAT,
    -> tip amount FLOAT,
                                                    add column with
    -> tolls_amount FLOAT,
                                                   auto increment at
    -> improvement_surcharge FLOAT,
    -> total_amount FLOAT,
                                                   end which will act
    -> Congestion surcharge FLOAT,
                                                      as rowkey in
    -> Airport fee FLOAT,
    -> UniqueID INT NOT NULL AUTO_INCREMENT,
                                                          hbase
    -> PRIMARY KEY (UniqueID)
Query OK, 0 rows affected (0.13 sec)
```

```
create table trip_data
(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(255),
PULocationID INT,
DOLocationID INT,
```

```
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),
congestion_surcharge float(10,2),
airport_fee float(10,2),
UniqueID INT NOT NULL AUTO_INCREMENT,
PRIMARY KEY(ID)
);
```



LOAD DATA LOCAL INFILE '/home/hadoop/dataset/yellow\_tripdata\_2017-01.csv'
INTO TABLE trip\_data
FIELDS TERMINATED BY ','

```
LINES TERMINATED BY '\n'
IGNORE 1 LINES;
```

## show tables;



select \* from tripdata limit 5;

select COUNT(\*) from tripdata;

LOAD DATA LOCAL INFILE '/home/hadoop/dataset/yellow\_tripdata\_2017-02.csv'

INTO TABLE trip\_data

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

IGNORE 1 LINES;

select COUNT(\*) from trip\_data;