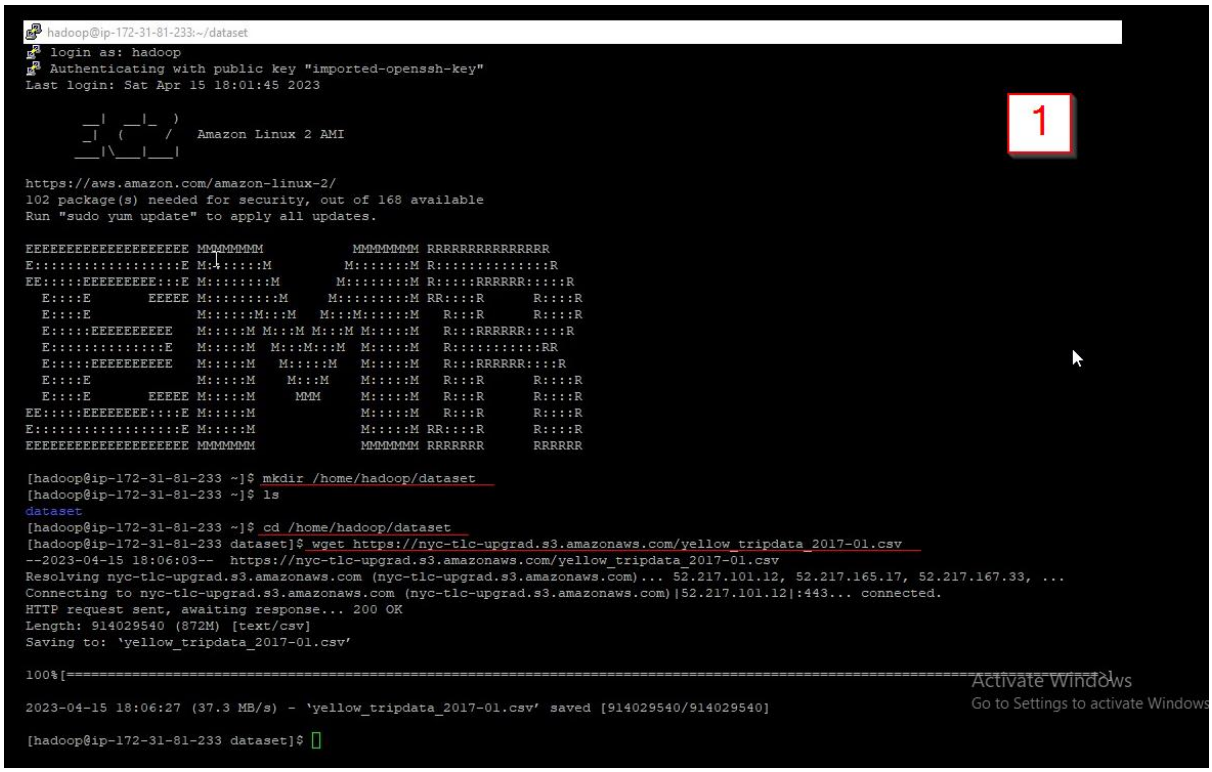
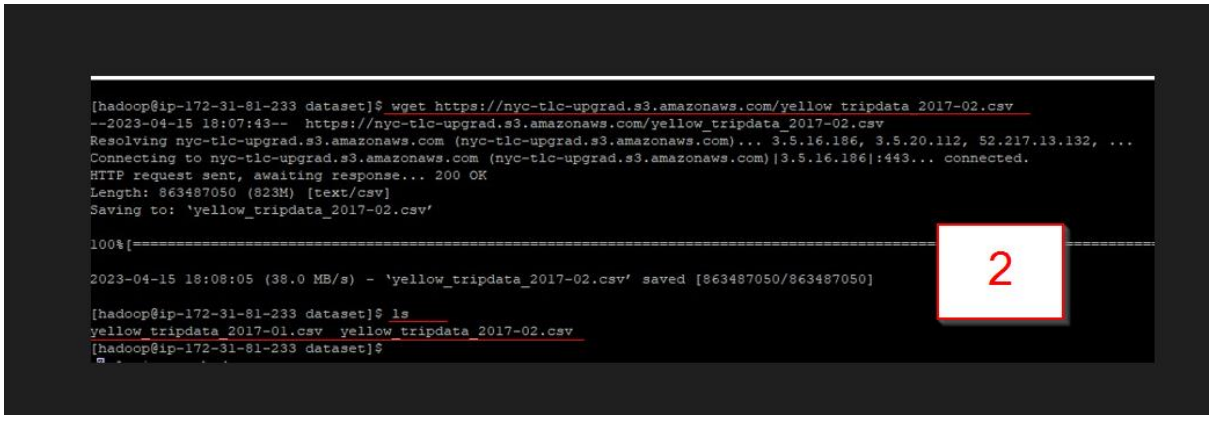
MapReduce - Programming Assignment

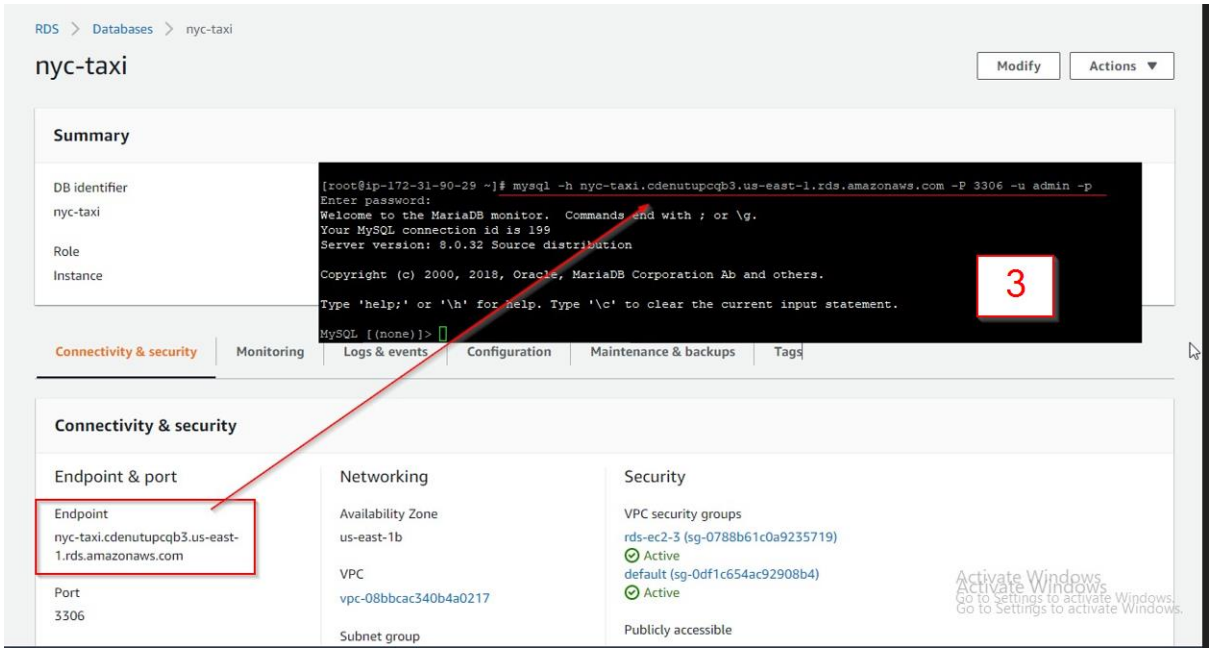
AMR_RDS_Task

BY
Dharan R
Sanjana Dutta
Punith Vadla

AWS RDS on EMR Cluster







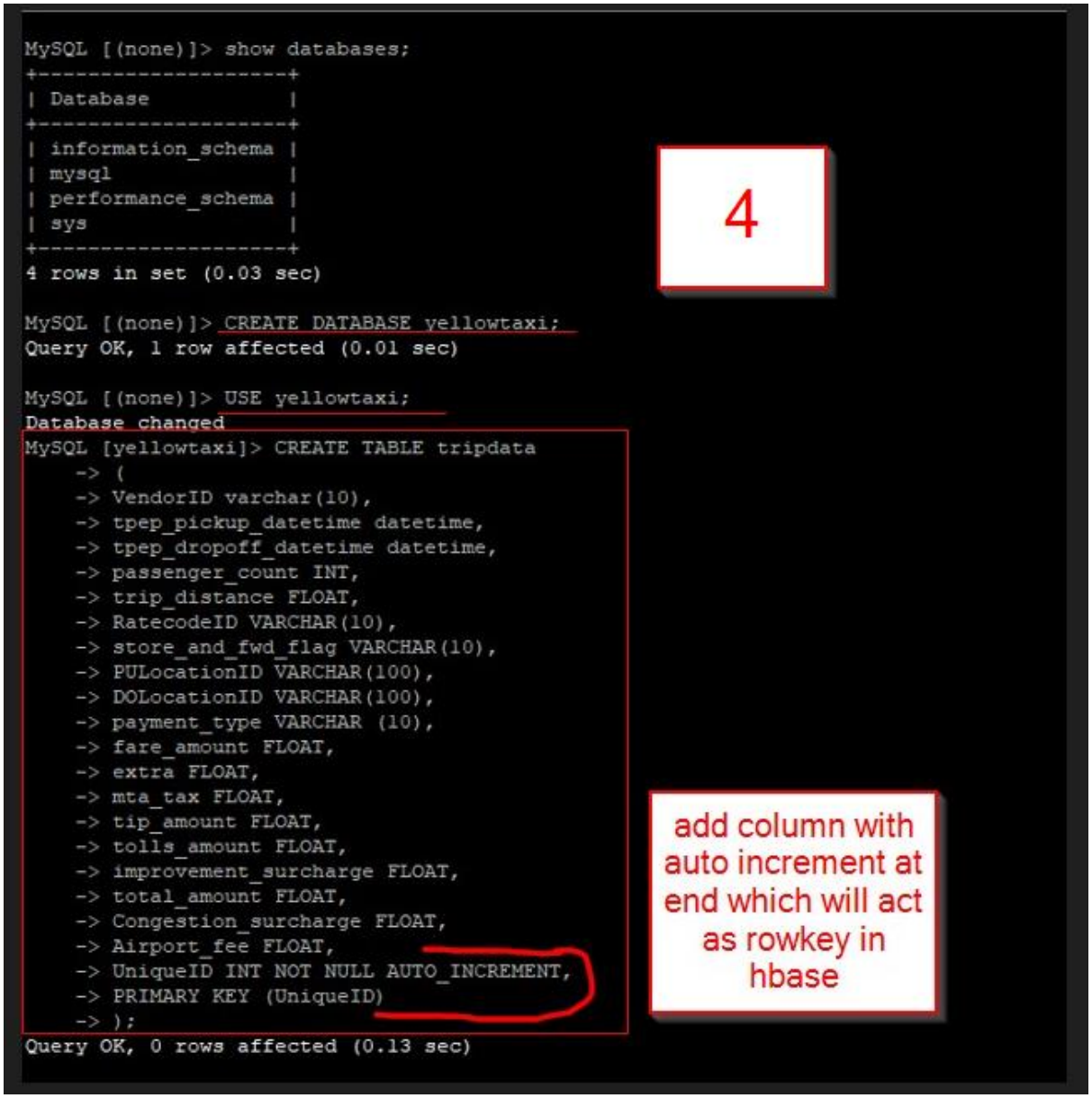
mysql -h nyc-taxi.cqwodvedrqu4.us-east-1.rds.amazonaws.com -P 3306 -u admin -p

show databases;

create database assignmentdb;

use assignmentdb;

show tables;



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),

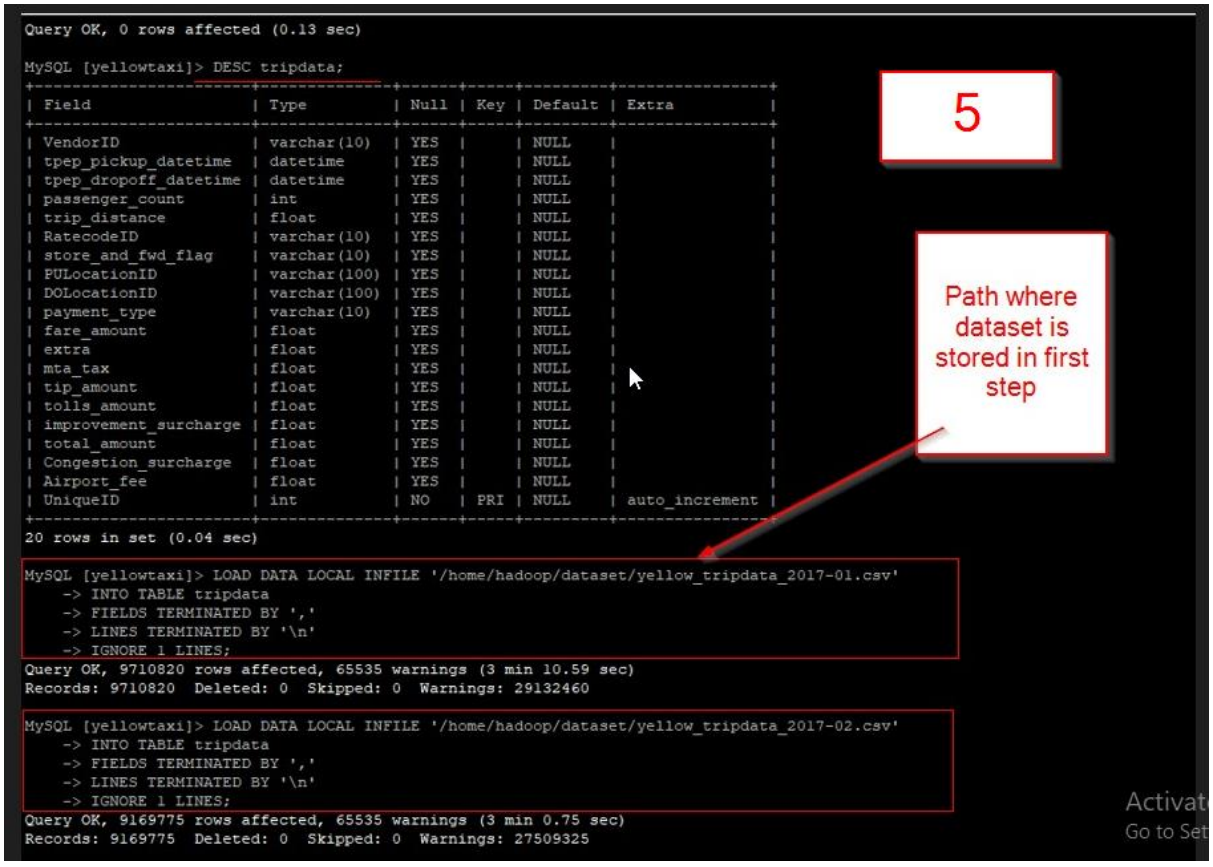
congestion_surcharge float(10,2),

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'

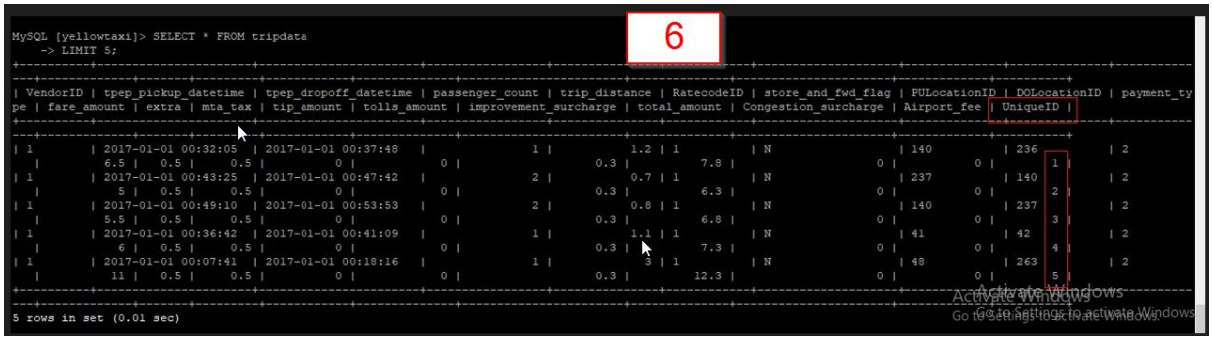
INTO TABLE trip_data

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;