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.

Here the step by step procedure for this task:

Step 1:download the data from the link provided with command wget for both

yellow_tripdata_2017-01.csv & yellow_tripdata_2017-02.csv.

Step2 : Connect with rds using command mysql –h <link> -P 3306 –u admin –p and enter the

password.

```
[hadoop@ip-172-31-32-103 ~]$ mysql -h database-1.crkohtdqsjvo.us-east-1.rds.amazonaws.com -P 3306 -u admin
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 11
Server version: 8.0.39 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)]> show Databases
    -> \c
MySQL [(none)]> show Databases;
+--------------------+
| Database           |
+--------------------+
| assignment         |
| information_schema |
| mysql              |
| performance_schema |
| sys                |
+--------------------+
```

Step 3: Create Database and also create the schema for the table TaxiTrip Data

Step4 : Now ingest the data through the command as mentioned in the below snip.

```
MySQL [assignment]> CREATE TABLE TaxiTripData (
    ->     VendorID INT,
    ->     tpep_pickup_datetime DATETIME,
    ->     tpep_dropoff_datetime DATETIME,
    ->     passenger_count INT,
    ->     trip_distance FLOAT,
    ->     RatecodeID INT,
    ->     store_and_fwd_flag CHAR(1),
    ->     PULocationID INT,
    ->     DOLocationID INT,
    fare_amount FLOAT,
    extra FLOAT,
    mta_tax FLOAT,
    tip_amount FLOAT,
    tolls_amount FLOAT,
    improvement_surcharge FLOAT,
    total_amount FLOAT,
    congestion_surcharge FLOAT,
    airport_fee FLOAT);
    ->     payment_type INT,
    ->     fare_amount FLOAT,
    ->     extra FLOAT,
    ->     mta_tax FLOAT,
    ->     tip_amount FLOAT,
    ->     tolls_amount FLOAT,
    ->     improvement_surcharge FLOAT,
    ->     total_amount FLOAT,
    ->     congestion_surcharge FLOAT,
    ->     airport_fee FLOAT);
Query OK, 0 rows affected (0.042 sec)

MySQL [assignment]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv'
```

Step 5: Now we will verify the count as the data has been ingested properly or not.

```
MySQL [assignment]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv'
    -> INTO TABLE TaxiTripData
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 LINES
    -> (VendorID, tpep_pickup_datetime, tpep_dropoff_datetime, passenger_count, trip_distance, RatecodeID, store_and_fwd_flag, PULocationID, DOLocation
mta_tax, tip_amount, tolls_amount, improvement_surcharge, total_amount, congestion_surcharge, airport_fee);
Query OK, 9710820 rows affected, 65535 warnings (2 min 29.122 sec)
Records: 9710820  Deleted: 0  Skipped: 0  Warnings: 19421640

MySQL [assignment]> SELECT COUNT(*) FROM TaxiTripData;
+----------+
| COUNT(*) |
+----------+
|  9710820 |
+----------+
1 row in set (19.097 sec)

MySQL [assignment]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-02.csv'
    -> INTO TABLE TaxiTripData
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 LINES
    -> (VendorID, tpep_pickup_datetime, tpep_dropoff_datetime, passenger_count, trip_distance, RatecodeID, store_and_fwd_flag, PULocationID, DOLocation
mta_tax, tip_amount, tolls_amount, improvement_surcharge, total_amount, congestion_surcharge, airport_fee);
Query OK, 9169775 rows affected, 65535 warnings (2 min 17.489 sec)
Records: 9169775  Deleted: 0  Skipped: 0  Warnings: 18339550

MySQL [assignment]> SELECT COUNT(*) FROM TaxiTripData;
```