

## Task 1

Create an RDS instance in your AWS account and upload the data to the RDS instance.

### 1. Creating SQL Table to store NY Taxi Trip Details

Creating a table in RDS instance.

- Table name is ***ny\_taxi\_details***.
- rowid is primary key of table and it is auto incremented.

Code to create table is as below:

```
CREATE TABLE ny_taxi_details
( rowid INT AUTO_INCREMENT PRIMARY KEY,
  vendorid INT,
  tpep_pickup_datetime TIMESTAMP,
  tpep_dropoff_datetime TIMESTAMP,
  passenger_count INT,
  trip_distance DECIMAL(10,2),
  ratecodeid INT,
  store_and_fwd_flag CHAR(1),
  pulocationid INT,
  dolocationid INT,
  payment_type INT,
  fare_amount DOUBLE,
  extra DOUBLE,
  mta_tax DOUBLE,
  tip_amount DOUBLE,
  tolls_amount DOUBLE,
  improvement_surcharge DOUBLE,
  total_amount DOUBLE,
  congestion_surcharge DOUBLE,
  airport_fee DOUBLE);
```

```

MySQL [nytlc]> CREATE TABLE ny_taxi_details
-> ( rowid INT AUTO_INCREMENT PRIMARY KEY,
->   vendorid INT,
->   tpep_pickup_datetime TIMESTAMP,
->   tpep_dropoff_datetime TIMESTAMP,
->   passenger_count INT,
->   trip_distance DECIMAL(10,2),
->   ratecodeid INT,
->   store_and_fwd_flag CHAR(1),
->   pulocationid INT,
->   dolocationid INT,
->   payment_type INT,
->   fare_amount DOUBLE,
->   extra DOUBLE,
->   mta_tax DOUBLE,
->   tip_amount DOUBLE,
->   tolls_amount DOUBLE,
->   improvement_surcharge DOUBLE,
->   total_amount DOUBLE,
->   congestion_surcharge DOUBLE,
->   airport_fee DOUBLE);
Query OK, 0 rows affected (0.30 sec)

```

```

MySQL [nytlc]> show tables;
+-----+
| Tables_in_nytlc |
+-----+
| ny_taxi_details |
+-----+
1 row in set (0.00 sec)

```

## 2. Load data from csv files to SQL Table:

Loading data from given csv file (yellow\_tripdata\_2017-01.csv, yellow\_tripdata\_2017-02.csv) to ny\_taxi\_details table using LOAD DATA LOCAL INFILE command.

While loading data fields are terminated by comma (,) and lines are terminated by new line (\n).

### 2.1 Load yellow\_tripdata\_2017-01.csv

```

LOAD DATA LOCAL INFILE '/root/yellow_tripdata_2017-01.csv'
INTO TABLE ny_taxi_details
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,dolocationid,payment_type,fare_amount,e

```

```
xtra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,congestion_surcharge,airport_fee);
```

```
MySQL [nytlc]> LOAD DATA LOCAL INFILE '/root/yellow_tripdata_2017-01.csv'
-> INTO TABLE ny_taxi_details
-> 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,dolocationid,payment_type,fare_amount,extra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,congestion_surcharge,airport_fee);
Query OK, 9710820 rows affected, 65535 warnings (3 min 10.52 sec)
Records: 9710820 Deleted: 0 Skipped: 0 Warnings: 19421640
```

```
MySQL [nytlc]> select COUNT(*) from ny_taxi_details;
+-----+
| COUNT(*) |
+-----+
| 9710820 |
+-----+
1 row in set (30.86 sec)
```

## 2.2 Load yellow\_tripdata\_2017-02.csv

```
LOAD DATA LOCAL INFILE '/root/yellow_tripdata_2017-02.csv'
INTO TABLE ny_taxi_details
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,dolocationid,payment_type,fare_amount,extra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,congestion_surcharge,airport_fee);
```

```
MySQL [nytlc]> LOAD DATA LOCAL INFILE '/root/yellow_tripdata_2017-02.csv'
-> INTO TABLE ny_taxi_details
-> 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,dolocationid,payment_type,fare_amount,extra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,congestion_surcharge,airport_fee);
Query OK, 9169775 rows affected, 65535 warnings (3 min 1.73 sec)
Records: 9169775 Deleted: 0 Skipped: 0 Warnings: 18339550
```

## 2.3 Check Records in table after data load

After importing data into table, we are checking total number of records present in table.

Total **18880595** records are present in table.

### 2.3.1 Select 5 rows from table

`SELECT * FROM ny_taxi_details LIMIT 5;`

```
MySQL [nytlc]> SELECT * FROM ny_taxi_details LIMIT 5;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| rowid | vendorid | tpep_pickup_datetime | tpep_dropoff_datetime | passenger_count | trip_distance | ratecodeid | store_and_fwd_flag | pulocationid | dolocationid | pa |
| yment_type | fare_amount | extra | mta_tax | tip_amount | tolls_amount | improvement_surcharge | total_amount | congestion_surcharge | airport_fee |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | 1 | 2017-01-01 00:32:05 | 2017-01-01 00:37:48 | 1 | 1.20 | 1 | N | 140 | 236 | | | | | | | |
| 2 | 2 | 6.5 | 0.5 | 0.5 | 0 | 0 | 0.3 | 0.70 | 7.8 | 1 | N | 0 | 0 | 237 | 140 |
| 3 | 2 | 5 | 0.5 | 0.5 | 0 | 0 | 0.3 | 0.80 | 6.3 | 1 | N | 0 | 0 | 140 | 237 |
| 4 | 1 | 2017-01-01 00:49:10 | 2017-01-01 00:53:53 | 2 | 5.5 | 1 | 2017-01-01 00:36:42 | 2017-01-01 00:41:09 | 1 | 1 | 1.10 | 1 | N | 0 | 41 | 42 |
| 5 | 2 | 6 | 0.5 | 0.5 | 0 | 0 | 0.3 | 3.00 | 7.3 | 1 | N | 0 | 48 | 263 |
| 6 | 2 | 11 | 0.5 | 0.5 | 0 | 0 | 0.3 | 12.3 | 0 | 0 |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
5 rows in set (0.04 sec)
```

### 2.3.2 Total table records count :

`SELECT COUNT(*) FROM ny_taxi_details;`

```
MySQL [nytlc]> select COUNT(*) from ny_taxi_details;
+-----+
| COUNT(*) |
+-----+
| 18880595 |
+-----+
1 row in set (56.75 sec)
```

Which is equal to the sum of records from both files, ignoring the header row in csv files.

9710820 and 9169775

```
[root@ip-172-31-78-179 ~]# wc -l yellow_tripdata_2017-01.csv
9710821 yellow_tripdata_2017-01.csv
[root@ip-172-31-78-179 ~]# wc -l yellow_tripdata_2017-02.csv
9169776 yellow_tripdata_2017-02.csv
[root@ip-172-31-78-179 ~]#
```