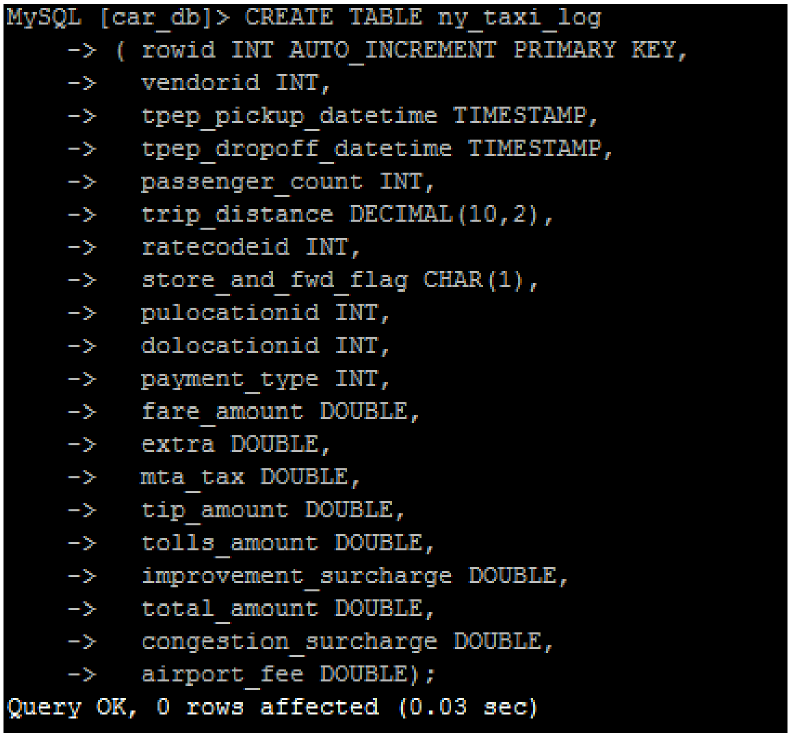
Task 1

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

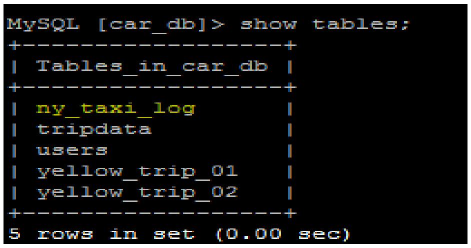
1. Creating SQL Table to store NY Taxi Trip Details

Creating a table in RDS instance.
Table name is ***ny_taxi_log***.
rowid is primary key of table and it is auto incremented.
To create a table code is present below:

```
CREATE TABLE ny_taxi_log
( 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 [car_db]> CREATE TABLE ny_taxi_log
    -> ( 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 (0.03 sec)
```

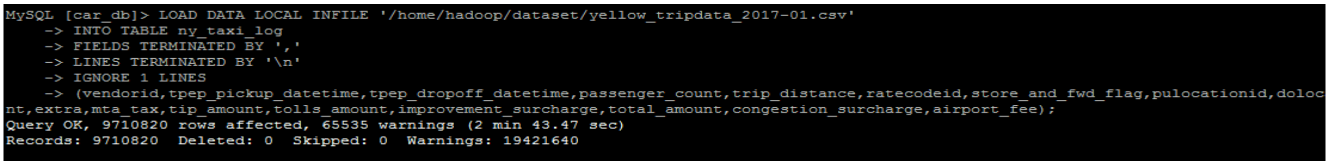


```
MySQL [car_db]> show tables;
+-------------------+
| Tables_in_car_db |
+-------------------+
| ny_taxi_log       |
| tripdata          |
| users             |
| yellow_trip_01    |
| yellow_trip_02    |
+-------------------+
5 rows in set (0.00 sec)
```

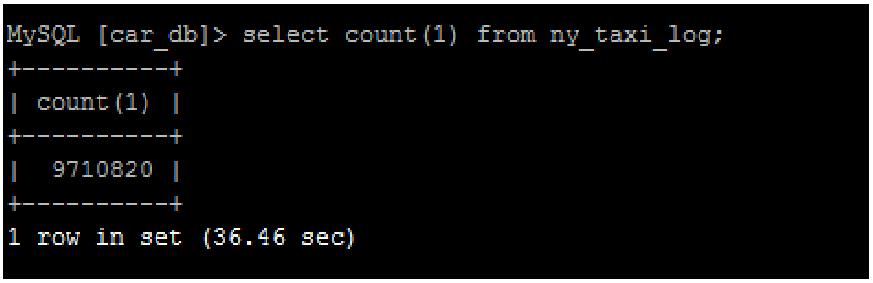
2. Load data from csv files to SQL Table:

2.1 Load yellow_tripdata_2017-01.csv

```
LOAD DATA LOCAL INFILE '/home/hadoop/dataset/yellow_tripdata_2017-01.csv'
INTO TABLE ny_taxi_log
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);
```



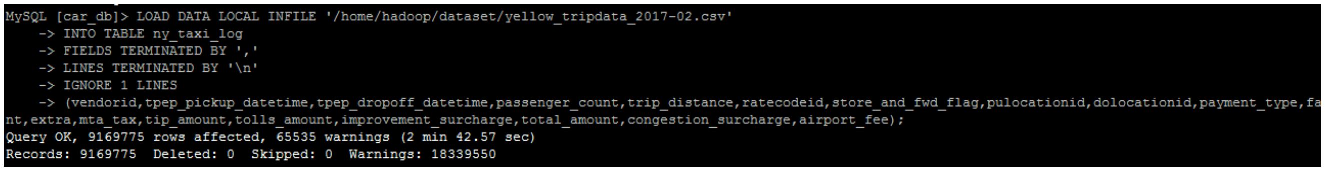
```
MySQL [car_db]> LOAD DATA LOCAL INFILE '/home/hadoop/dataset/yellow_tripdata_2017-01.csv'
    -> INTO TABLE ny_taxi_log
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 LINES
    -> (vendorid,tpep_pickup_datetime,tpep_dropoff_datetime,passenger_count,trip_distance,ratecodeid,store_and_fwd_flag,pulocationid,doloc
nt,extra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,congestion_surcharge,airport_fee);
Query OK, 9710820 rows affected, 65535 warnings (2 min 43.47 sec)
Records: 9710820  Deleted: 0  Skipped: 0  Warnings: 19421640
```



```
MySQL [car_db]> select count(1) from ny_taxi_log;
+----------+
| count(1) |
+----------+
|  9710820 |
+----------+
1 row in set (36.46 sec)
```

2.2 Load yellow_tripdata_2017-02.csv

```
LOAD DATA LOCAL INFILE '/home/hadoop/dataset/yellow_tripdata_2017-02.csv'
INTO TABLE ny_taxi_log
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,congestio
n_surcharge,airport_fee);
```



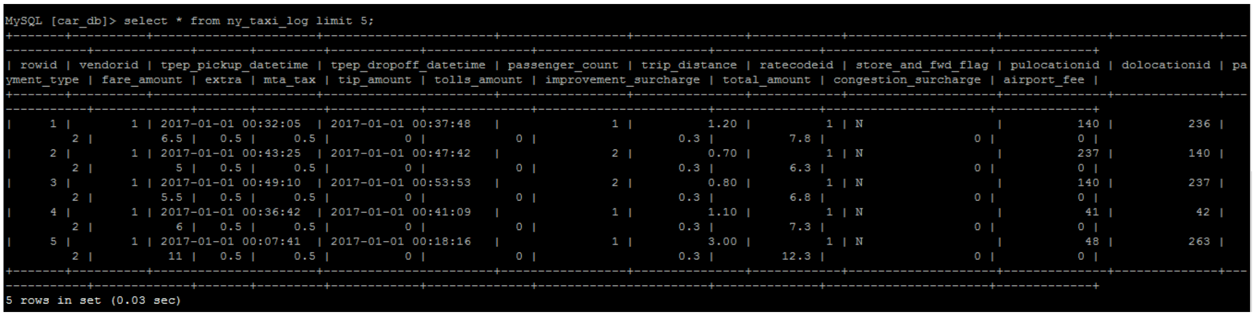
```
MySQL [car_db]> LOAD DATA LOCAL INFILE '/home/hadoop/dataset/yellow_tripdata_2017-02.csv'
    -> INTO TABLE ny_taxi_log
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 LINES
    -> (vendorid,tpep_pickup_datetime,tpep_dropoff_datetime,passenger_count,trip_distance,ratecodeid,store_and_fwd_flag,pulocationid,dolocationid,payment_type,fa
nt,extra,mta_tax,tip_amount,tolls_amount,improvement_surcharge,total_amount,congestion_surcharge,airport_fee);
Query OK, 9169775 rows affected, 65535 warnings (2 min 42.57 sec)
Records: 9169775  Deleted: 0  Skipped: 0  Warnings: 18339550
```

2.3 Check Records in table after data load

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

2.3.1 Select 5 rows from table

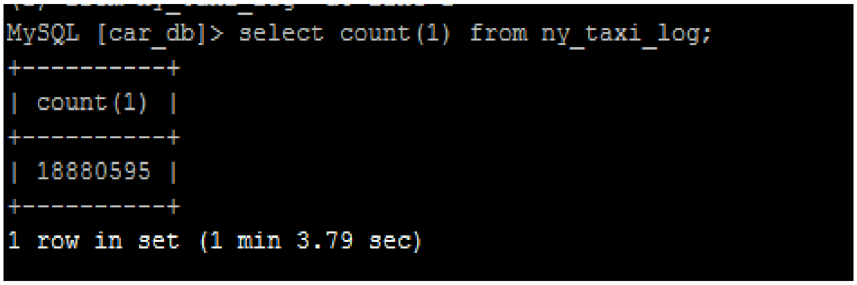
SELECT * FROM ny_taxi_log LIMIT 5;



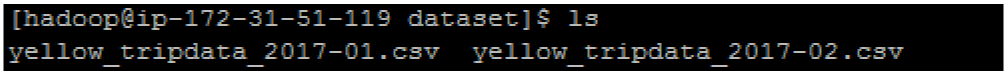
```
MySQL [car_db]> select * from ny_taxi_log limit 5;
+-------+----------+----------------------+-----------------------+-----------------+---------------+------------+--------------------+--------------+--------------+---
-----------+-------------+-------+---------+------------+--------------+-----------------------+--------------+----------------------+-------------+
| 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 |         6.5 |   0.5 |     0.5 |          0 |            0 |                   0.3 |          7.8 |                    0 |           0 |
|     2 |        1 | 2017-01-01 00:43:25  | 2017-01-01 00:47:42   |               2 |          0.70 |          1 | N                  |          237 |          140 |
         2 |           5 |   0.5 |     0.5 |          0 |            0 |                   0.3 |          6.3 |                    0 |           0 |
|     3 |        1 | 2017-01-01 00:49:10  | 2017-01-01 00:53:53   |               2 |          0.80 |          1 | N                  |          140 |          237 |
         2 |         5.5 |   0.5 |     0.5 |          0 |            0 |                   0.3 |          6.8 |                    0 |           0 |
|     4 |        1 | 2017-01-01 00:36:42  | 2017-01-01 00:41:09   |               1 |          1.10 |          1 | N                  |           41 |           42 |
         2 |           6 |   0.5 |     0.5 |          0 |            0 |                   0.3 |          7.3 |                    0 |           0 |
|     5 |        1 | 2017-01-01 00:07:41  | 2017-01-01 00:18:16   |               1 |          3.00 |          1 | N                  |           48 |          263 |
         2 |          11 |   0.5 |     0.5 |          0 |            0 |                   0.3 |         12.3 |                    0 |           0 |
+-------+----------+----------------------+-----------------------+-----------------+---------------+------------+--------------------+--------------+--------------+---
-----------+-------------+-------+---------+------------+--------------+-----------------------+--------------+----------------------+-------------+
5 rows in set (0.03 sec)
```

2.3.2 Total table records count :

SELECT COUNT(1) FROM ny_taxi_log;



```
MySQL [car_db]> select count(1) from ny_taxi_log;
+----------+
| count(1) |
+----------+
| 18880595 |
+----------+
1 row in set (1 min 3.79 sec)
```



```
[hadoop@ip-172-31-51-119 dataset]$ ls
yellow_tripdata_2017-01.csv  yellow_tripdata_2017-02.csv
```