**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,

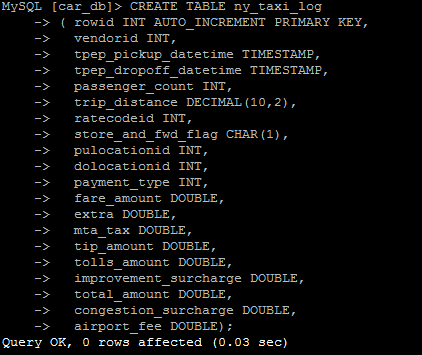
tolls\_amount DOUBLE,

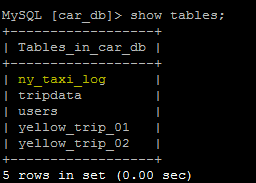
improvement\_surcharge DOUBLE,

total\_amount DOUBLE,

congestion\_surcharge DOUBLE,

airport\_fee DOUBLE);





1. **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'

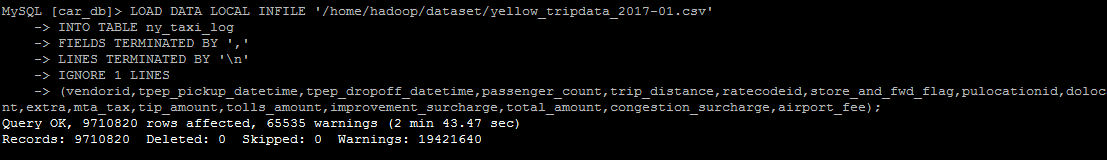
INTO TABLE ny\_taxi\_log

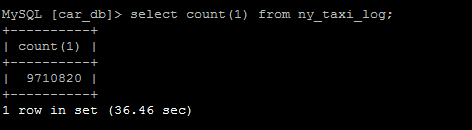
FIELDS TERMINATED BY ','

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);





* 1. **Load yellow\_tripdata\_2017-02.csv**

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

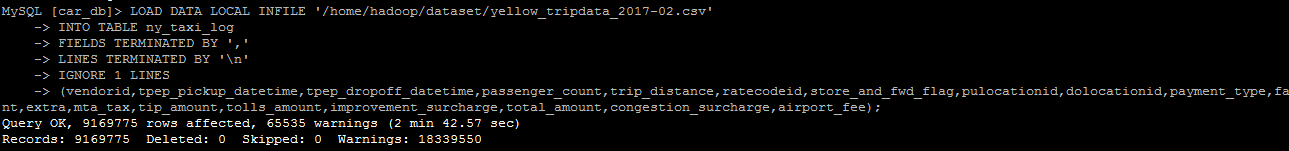
INTO TABLE ny\_taxi\_log

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,extra,mta\_tax,tip\_amount,tolls\_amount,improvement\_surcharge,total\_amount,congestion\_surcharge,airport\_fee);

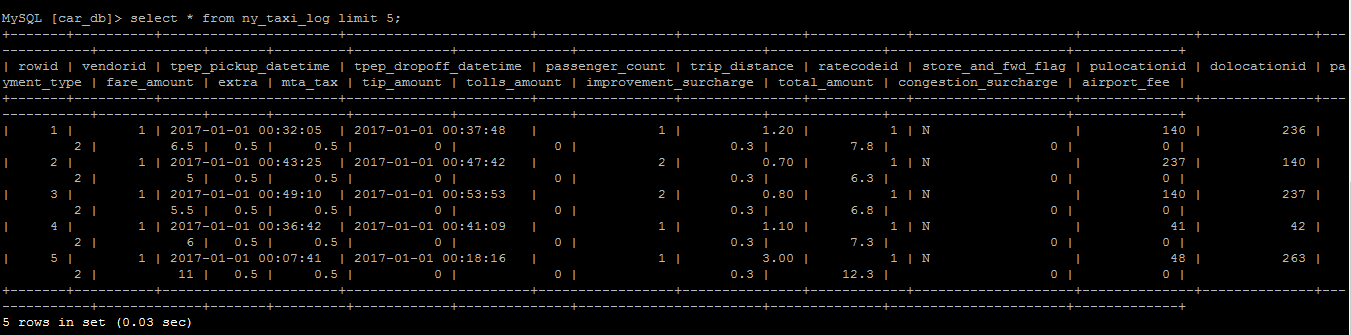


* 1. **Check Records in table after data load**

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

**2.3.1 Select 5 rows from table**

SELECT \* FROM ny\_taxi\_log LIMIT 5;



**2.3.2 Total table records count :**

SELECT COUNT(1) FROM ny\_taxi\_log;

