


- Created an RDS instance in AWS RDS.

Summary

DB identifier yellowtaxi	CPU -	Status 🟢 Available	Class db.t3.micro
Role Instance	Current activity  3 Connections	Engine MySQL Community	Region & AZ us-east-1d

- Connected to RDS instance via mysql.

```
[hadoop@ip-172-31-76-22 ~]$ cd mysql-connector-java-8.0.25
[hadoop@ip-172-31-76-22 mysql-connector-java-8.0.25]$ mysql -h yellowtaxi.chlmu6dvhmli.us-east-1.rds.amazonaws.com -P 3306 -u dharani -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 1914
Server version: 8.0.28 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)]> use yellow;
```

- Created table using create table query.

```
create table taxi
(
  VendorID int,
  tpep_pickup_datetime timestamp,
  tpep_dropoff_datetime datetime,
  passenger_count int,
  trip_distance float,
  RatecodeID int,
  store_and_fwd_flag char,
  PULocationID varchar(10),
  DOLocationID varchar(10),
  payment_type int,
  fare_amount decimal(10,3),
  extra decimal(10,3),
  mta_tax decimal(10,3),
  tip_amount decimal(10,3),
  tolls_amount float,
  improvement_surcharge decimal(10,3),
```

total_amount decimal(10,3),
 congestion_surcharges float,
 airport_fee float)

- table with schema based on dataset using mysql.

```
MySQL [yellow]> show tables;
+-----+
| Tables_in_yellow |
+-----+
| taxi              |
+-----+
1 row in set (0.00 sec)

MySQL [yellow]> show columns from taxi;
+-----+-----+-----+-----+-----+-----+
| Field          | Type          | Null | Key | Default | Extra          |
+-----+-----+-----+-----+-----+-----+
| id             | int           | NO   | PRI | NULL    | auto_increment |
| VendorID       | int           | YES  |     | NULL    |                |
| tpep_pickup_datetime | timestamp     | YES  |     | NULL    |                |
| tpep_dropoff_datetime | datetime      | YES  |     | NULL    |                |
| passenger_count | int           | YES  |     | NULL    |                |
| trip_distance  | float         | YES  |     | NULL    |                |
| RatecodeID     | int           | YES  |     | NULL    |                |
| store_and_fwd_flag | char(1)       | YES  |     | NULL    |                |
| PULocationID   | varchar(10)   | YES  |     | NULL    |                |
| DOLocationID   | varchar(10)   | YES  |     | NULL    |                |
| payment_type    | int           | YES  |     | NULL    |                |
| fare_amount     | decimal(10,3) | YES  |     | NULL    |                |
| extra           | decimal(10,3) | YES  |     | NULL    |                |
| mta_tax         | decimal(10,3) | YES  |     | NULL    |                |
| tip_amount      | decimal(10,3) | YES  |     | NULL    |                |
| tolls_amount    | float         | YES  |     | NULL    |                |
| improvement_surcharge | decimal(10,3) | YES  |     | NULL    |                |
| total_amount    | decimal(10,3) | YES  |     | NULL    |                |
| congestion_surcharge | float         | YES  |     | NULL    |                |
| airport_fee     | float         | YES  |     | NULL    |                |
+-----+-----+-----+-----+-----+-----+
20 rows in set (0.03 sec)
```

- Loaded dataset 1 and 2 to RDS using Load data command.

>Load data local infile '/home/hadoop/yellow_tripdata_2017-03.csv' into table taxi
 fields terminated by ',' lines terminated by '\n' ignore 1 lines;

```
MySQL [yellow]> load data local infile '/home/hadoop/yellow_tripdata_2017-02.csv' into table taxi fields terminated by ',' lines terminated by '\n' ignore 1
lines;
Query OK, 9169775 rows affected, 65535 warnings (2 min 2.13 sec)
Records: 9169775 Deleted: 0 Skipped: 0 Warnings: 18339550

MySQL [yellow]> select count(*) from taxi;
+-----+
| count(*) |
+-----+
| 18880595 |
+-----+
```