# TASK 2

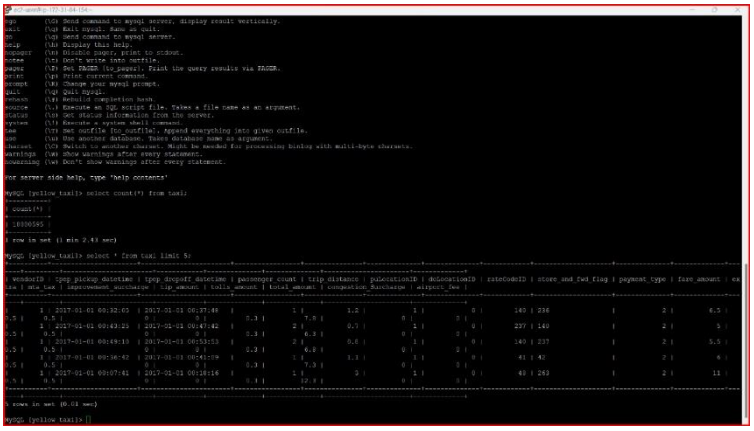
**Task 2. Use Sqoop command to ingest the data from RDS into the HBase Table.**

**The following steps were followed for data ingestion from RDS into HBase table**

1) Load the data to RDS instance

Following screen short for reference from previous step (i.e. Task 1)

```
MySQL [yellow_taxi]> select count(*) from taxi;
+----------+
| count(*) |
+----------+
| 18880595 |
+----------+
1 row in set (1 min 2.43 sec)
```



2) Exit form RDS and load the table data in to the hbase.

Created the hbase table : hbasetaxi, set the column-family as cf. Copied the data from the RDS taxi table

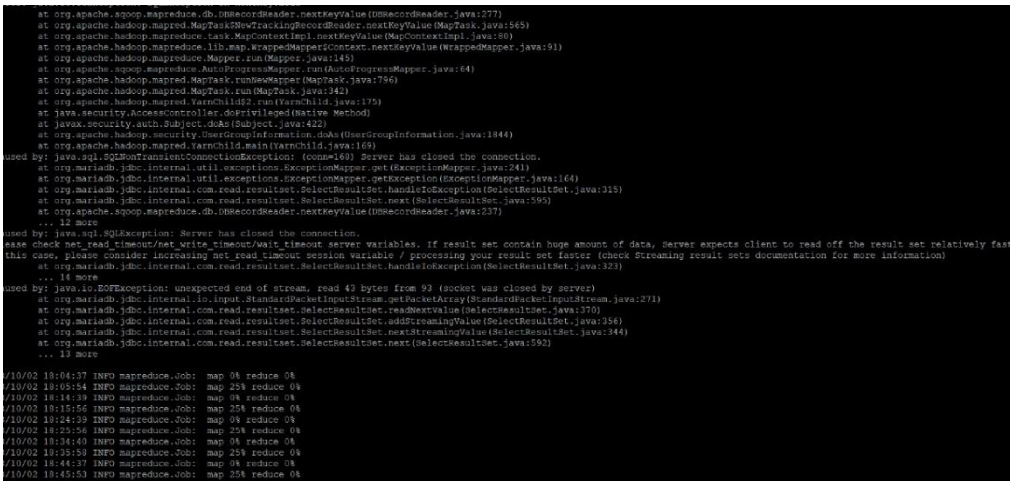
**Sqoop import command:**

```
sqoop import \
  --connect "jdbc:mysql://casestudy.ci2jpset7w3y.us-east-1.rds.amazonaws.com:3306/yellow_taxi" \
  --username root \
  --password 123456789 \
  --table taxi \
  --columns
"vendorID,tpep_pickup_datetime,tpep_dropoff_datetime,passenger_count,trip_distance,puLocationID,doLoc
```

# TASK 2

ationID,rateCodeID,store_and_fwd_flag,payment_type,fare_amount,extra,mta_tax,improvement_surcharge,tip_amount,tolls_amount,total_amount,congestion_Surcharge,airport_fee" \
--hbase-create-table \
--hbase-table hbasetaxi \
--column-family trip_details \
--hbase-row-key "vendorID,tpep_pickup_datetime,tpep_dropoff_datetime" \
--split-by tpep_dropoff_datetime \
-m 8

Screenshots for reference:



# TASK 2

After the mapreduce

Lets check for the hbasetaxi by running

Scan ‘hbasetaxi’

In hbase shell