Amazon RDS

Data Ingestion Task

The codes and methods we used to import data from two csv files to Amazon RDS using EMR are present in this file.

Files we used:-

- 1. https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
- 2. https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv

Step 1: Login in EMR

Code:

ssh -i demoKey07.pem hadoop@ec2-3-237-71-254.compute-1.amazonaws.com

```
(base) apple@Apples-MacBook-Pro ~ % cd Desktop
[(base] apple@Apples-MacBook-Pro Desktop % ssh -i demoKey07.pem hadoop@ec2-3-237-71-254.compute-1.amazonaws.com
Last login: Thu Dec 15 10:01:42 2022 from 49.36.182.104
      https://aws.amazon.com/amazon-linux-2/
24 package(s) needed for security, out of 46 available
Run "sudo yum update" to apply all updates.
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
EEEEEEEEEEEEEEE MMMMMMM
                                       M::::::M R::::::::R
M::::::M R:::::RRRRRR::::R
                                    M:::::::::::::::::::::::::R
                                                             R::::R
                                                             R::::R
  R:::RRRRRR::::R
                                                 R:::::::::RR
                                                  R:::RRRRRR::::R
                              M:::M
MMM
  E::::E
E::::E
                                                  R:::R
                                                            R::::R
R:::R
                                        M:::::M
                                                             R::::R
                                                  R:::R
                                        M:::::M
                                                             R::::R
                                        M:::::M RR::::R
                                                             R::::R
                                        MMMMMMM RRRRRRR
                                                             RRRRRR
[hadoop@ip-172-31-5-184 ~]$
```

Step 2: Uploading raw data csv files to EMR

Created a dictionary to upload raw csv files **Code:**

mkdir rawDataFiles

```
[[hadoop@ip-172-31-5-184 ~]$ mkdir rawDataFiles
[[hadoop@ip-172-31-5-184 ~]$ ls
rawDataFiles
[hadoop@ip-172-31-5-184 ~]$
```

Upload first file from local to EMR

Code:

scp -i demoKey07.pem yellow_tripdata_2017-01.csv hadoop@ec2-3-237-71-254.compute-1.amazonaws.com:/home/hadoop/rawDataFiles/

Upload second file from local to EMR

Code:

scp -i demoKey07.pem yellow_tripdata_2017-02.csv hadoop@ec2-3-218-250-160.compute-1.amazonaws.com:/home/hadoop/rawDataFiles/

```
(base) apple0Apples-MacBook-Pro Desktop % scp -i demoKey07.pem yellow_tripdata_2017-02.csv hadoop@ec2-3-218-250-160.compute-1.amazonaws.com:/home/hadoop/rawDataFiles/
/etc/profile.d/lang.sh: line 19: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
yellow_tripdata_2017-02.csv

100% 823MB 7.1MB/s 01:56
(base) apple0Apples-MacBook-Pro Desktop %
```

Both the files are now available on EMR Directory

```
[[hadoop@ip-172-31-5-184 rawDataFiles]$ ls
yellow_tripdata_2017-01.csv yellow_tripdata_2017-02.csv
```

Step 2: Login in to RDS using MySQL CLI

Code:

sudo mysql --local-infile=1 -h mapreduceassignment.cjipwhexa9kv.us-east-1.rds.amazonaws.com -u admin -p

```
[hadoop@ip-172-31-5-184 rawDataFiles]$ sudo mysql --local-infile=1 -h mapreduceassignment.cjipwhexa9kv.us-east-1.rds.amazonaws.com -u admin -p [Enter password:
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MySQL connection id is 62
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)]>
```

Step 2: Creating Database and table in RDS

Create new database "MapReduceAssignment" **Code:**

create database MapReduceAssignment;

Use create database

Code:

Use MapReduceAssignment;

```
[MySQL [(none)]> use MapReduceAssignment;

Database changed

MySQL [MapReduceAssignment]>
```

Create table command

Code:

```
CREATE TABLE trip record(
 VendorID
                 INTEGER
 ,tpep_pickup_datetime DATETIME
 ,tpep_dropoff_datetime DATETIME
 ,passenger_count INTEGER
 ,trip_distance
                 NUMERIC(5,2)
 .RatecodeID
                 INTEGER
 ,store_and_fwd_flag
                        VARCHAR(1)
,PULocationID
                 INTEGER
 .DOLocationID
                 INTEGER
 ,payment_type
                 INTEGER
 ,fare amount
                 NUMERIC(4,1)
                 NUMERIC(3,1)
 ,extra
,mta_tax
                 NUMERIC(3,1)
                 NUMERIC(5,2)
tip_amount,
 ,tolls_amount
                 NUMERIC(4,2)
 ,improvement_surcharge NUMERIC(3,1)
```

```
,total_amount NUMERIC(5,2)
,congestion_surcharge NUMERIC(4,1)
,airport fee NUMERIC(4,1));
```

```
MySQL [MapReduceAssignment]> CREATE TABLE trip_record(
    ->
          VendorID
                                  INTEGER
         ,tpep_pickup_datetime
                                  DATETIME
    ->
    ->
         ,tpep_dropoff_datetime DATETIME
         ,passenger_count
                                  INTEGER
    ->
         ,trip_distance
                                  NUMERIC(5,2)
    ->
         ,RatecodeID
                                  INTEGER
    ->
         ,store_and_fwd_flag
                                  VARCHAR(1)
    ->
         , PULocationID
    ->
                                  INTEGER
         ,DOLocationID
    ->
                                  INTEGER
    ->
         ,payment_type
                                  INTEGER
                                  NUMERIC(4,1)
    ->
         ,fare_amount
    ->
                                  NUMERIC(3,1)
         ,extra
                                  NUMERIC(3,1)
    ->
         ,mta_tax
         ,tip_amount
                                  NUMERIC(5,2)
    ->
         ,tolls_amount
                                  NUMERIC(4,2)
    ->
         ,improvement_surcharge NUMERIC(3,1)
    ->
                                  NUMERIC(5,2)
         ,total_amount
    ->
         ,congestion_surcharge
                                  NUMERIC(4,1)
    ->
         ,airport_fee
                                 NUMERIC(4,1)
    ->
    -> );
Query OK, 0 rows affected (0.04 sec)
```

Importing data from first file

Code:

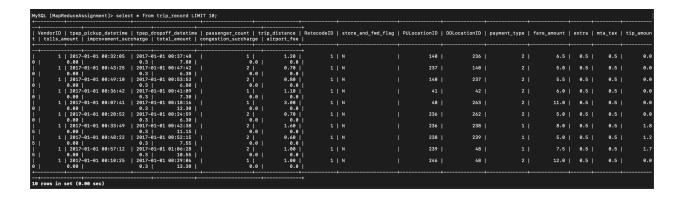
LOAD DATA LOCAL INFILE '/home/hadoop/rawDataFiles/yellow_tripdata_2017-01.csv' INTO TABLE trip_record FIELDS TERMINATED BY ',' ENCLOSED BY ''" LINES TERMINATED BY '\n' IGNORE 1 LINES;

WySQL [MapReduceAssignment]> LOAD DATA LOCAL INFILE '/home/hadoop/rawDataFiles/yellow_tripdata_2017-01.csv' INTO TABLE trip_record FIELDS TERMINATED BY ',' ENCLOSED BY '" LINES TERMINATED BY '\n' IGNORE 1 LINES Query OK, 9718820 rows affected, 66535 warnings (1 min 48.88 sec)
Records: 9718820 Boleted: 0 Skipped: 0 Marnings: 19424237

Checking whether data is loaded or not

Code:

select * from trip record LIMIT 10;



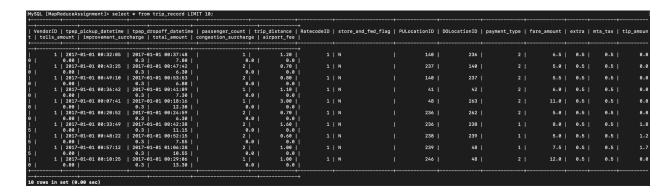
Importing data from second file

Code:

LOAD DATA LOCAL INFILE '/home/hadoop/rawDataFiles/yellow_tripdata_2017-02.csv' INTO TABLE trip_record FIELDS TERMINATED BY ',' ENCLOSED BY ''" LINES TERMINATED BY '\n' IGNORE 1 LINES;

```
| MySQL [MapReduceAssignment]> LOAD DATA LOCAL INFILE '/home/hadoop/rawDateFiles/yellow_tripdate_2017-02.csv' INTO TABLE trip_record FIELDS TERMINATED BY ',' ENCLOSED BY ''' LINES TERMINATED BY '\n' IGNORE 1 LINES. Query 0K, 9169775 Cross affected, 68535 warnings (1 min 45.87 sec)
Records: 9169775 Deleted: 0 Skipped: 0 Marnings: 18342814
```

Checking whether data is loaded or not



Total row in RDS:

```
[MySQL [MapReduceAssignment]> SELECT COUNT(*) FROM trip_record;
+-----+
| COUNT(*) |
+-----+
| 18880595 |
+-----+
1 row in set (36.53 sec)
```

Submitted By:

Member 1: Nikhil Dhiman (nikhildhiman3644@gmail.com)

Member 2: Rohan kulkarni (rohan.kulkarni951@gmail.com)

Member 3: Mahesh Kumar Patra (maheshkumarpatra5@gmail.com)