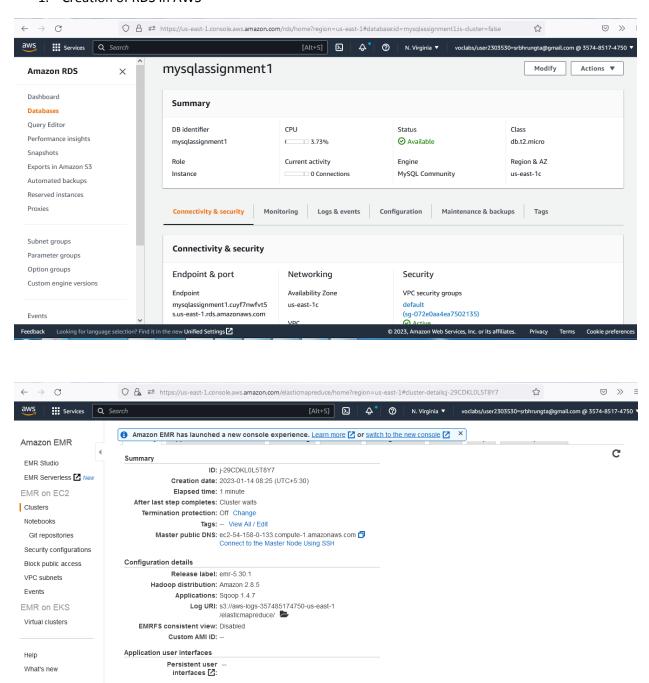1. Creation of RDS in AWS

2. Login to RDS through EMR instance

mysql -h mysqlassignment1.cuyf7nwfvt5s.us-east-1.rds.amazonaws.com -P 3306 -u srbhrungta -p

```
login as: ec2-user
Authenticating with public key "imported-openssh-key"

       __|  __|_  )
       _|  (     /   Amazon Linux 2 AMI
      ___|\___|___|

https://aws.amazon.com/amazon-linux-2/
110 package(s) needed for security, out of 176 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE::::EEEEEEEEE::::E M::::::::M         M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M       M:::::::::M RR::::R      R::::R
  E::::E             M::::::M::::M   M::::M::::::M   R:::R      R::::R
  E::::EEEEEEEEEE    M::::::M M::::M M::::M M::::::M   R:::RRRRRR::::::R
  E:::::::::::::::E  M::::::M  M::::M::::M  M::::::M   R:::::::::::RR
  E::::EEEEEEEEEE    M::::::M   M:::::M   M::::::M   R:::RRRRRR::::R
  E::::E             M::::::M    M:::M    M::::::M   R:::R      R::::R
  E::::E       EEEEE M::::::M     MMM     M::::::M   R:::R      R::::R
EE::::EEEEEEEE::::E M::::::M             M::::::M   R:::R      R::::R
E::::::::::::::::::E M::::::M             M::::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[ec2-user@ip-172-31-34-81 ~]$ sudo -i

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE::::EEEEEEEEE::::E M::::::::M         M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M       M:::::::::M RR::::R      R::::R
  E::::E             M::::::M::::M   M::::M::::::M   R:::R      R::::R
  E::::EEEEEEEEEE    M::::::M M::::M M::::M M::::::M   R:::RRRRRR::::::R
  E:::::::::::::::E  M::::::M  M::::M::::M  M::::::M   R:::::::::::RR
  E::::EEEEEEEEEE    M::::::M   M:::::M   M::::::M   R:::RRRRRR::::R
  E::::E             M::::::M    M:::M    M::::::M   R:::R      R::::R
  E::::E       EEEEE M::::::M     MMM     M::::::M   R:::R      R::::R
EE::::EEEEEEEE::::E M::::::M             M::::::M   R:::R      R::::R
E::::::::::::::::::E M::::::M             M::::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[root@ip-172-31-34-81 ~]#
```

```
  E::::E             M::::::M::::M   M::::M::::::M   R:::R      R::::R
  E::::EEEEEEEEEE    M::::::M M::::M M::::M M::::::M   R:::RRRRRR::::::R
  E:::::::::::::::E  M::::::M  M::::M::::M  M::::::M   R:::::::::::RR
  E::::EEEEEEEEEE    M::::::M   M:::::M   M::::::M   R:::RRRRRR::::R
  E::::E             M::::::M    M:::M    M::::::M   R:::R      R::::R
  E::::E       EEEEE M::::::M     MMM     M::::::M   R:::R      R::::R
EE::::EEEEEEEE::::E M::::::M             M::::::M   R:::R      R::::R
E::::::::::::::::::E M::::::M             M::::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[root@ip-172-31-34-81 ~]# mysql -h mysqlassignment1.cuyf7nwfvt5s.us-east-1.rds.amazonaws.com -P 3306 -u srbhrungta -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 3272
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)]> show tables
    -> ;
ERROR 1046 (3D000): No database selected
MySQL [(none)]> use demo;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A
```

3. Creation of database and table (nyc_yellow_taxi): -
   create table if not exists nyc_yellow_taxi(VendorID int, tpep_pickup_datetime timestamp,

tpep_dropoff_datetime timestamp, Passenger_count int, Trip_distance DECIMAL(9,6),
RateCodeID int, Store_and_fwd_flag CHAR, PULocationID int, DOLocationID int, Payment_type
int, Fare_amount DECIMAL(9,6), Extra DECIMAL(9,6), MTA_tax DECIMAL(9,6), Tip_amount
DECIMAL(9,6), Tolls_amount DECIMAL(9,6), Improvement_surcharge DECIMAL(9,6),
Total_amount DECIMAL(9,6));

```
MySQL [demo]> show tables
    -> ;
Empty set (0.00 sec)

MySQL [demo]> create table if not exists nyc_yellow_taxi(VendorID int, tpep_pickup_datetime  timestamp,
    -> tpep_dropoff_datetime timestamp, Passenger_count int, Trip_distance DECIMAL(9,6), RateCodeID int, Store_and_fwd_flag CHAR,
    -> PULocationID int, DOLocationID int, Payment_type int, Fare_amount DECIMAL(9,6), Extra DECIMAL(9,6), MTA_tax DECIMAL(9,6),
    -> Tip_amount DECIMAL(9,6), Tolls_amount DECIMAL(9,6), Improvement_surcharge DECIMAL(9,6), Total_amount DECIMAL(9,6))
    -> ;
Query OK, 0 rows affected (0.10 sec)

MySQL [demo]> show tables;
+----------------+
| Tables_in_demo |
+----------------+
| nyc_yellow_taxi |
+----------------+
1 row in set (0.00 sec)

MySQL [demo]> select count(*) from nyc_yellow_taxi;
+----------+
| count(*) |
+----------+
|        0 |
+----------+
1 row in set (0.01 sec)

MySQL [demo]>
```

4. Downloading of CSV file given in the assignment
   wget -P /root/ https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
   wget -P /root/ https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv

```
MySQL [demo]> show tables;
+----------------+
| Tables_in_demo |
+----------------+
| nyc_yellow_taxi |
+----------------+
1 row in set (0.00 sec)

MySQL [demo]> select count(*) from nyc_yellow_taxi;
+----------+
| count(*) |
+----------+
|        0 |
+----------+
1 row in set (0.01 sec)

MySQL [demo]> exit
Bye
[root@ip-172-31-34-81 ~]# wget -P /root/ https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
--2023-01-14 03:11:02--  https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 54.231.166.81, 54.231.170.201, 3.5.3.216, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|54.231.166.81|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 914029540 (872M) [text/csv]
Saving to: '/root/yellow_tripdata_2017-01.csv'

100%[===============================================================================================================>] 914,029,540 34.8MB/s   in 27s

2023-01-14 03:11:30 (31.8 MB/s) - '/root/yellow_tripdata_2017-01.csv' saved [914029540/914029540]

[root@ip-172-31-34-81 ~]# wget -P /root/ https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv
--2023-01-14 03:11:51--  https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 52.217.201.113, 54.231.133.225, 54.231.228.73, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|52.217.201.113|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 863487050 (823M) [text/csv]
Saving to: '/root/yellow_tripdata_2017-02.csv'

100%[===============================================================================================================>] 863,487,050 29.3MB/s   in 27s

2023-01-14 03:12:18 (30.3 MB/s) - '/root/yellow_tripdata_2017-02.csv' saved [863487050/863487050]

[root@ip-172-31-34-81 ~]#
```

5. Load in CSV file data in sql table(nyc_yellow_taxi) and checking the count of data.

LOAD DATA LOCAL INFILE 'yellow_tripdata_2017-01.csv' INTO TABLE nyc_yellow_taxi FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

LOAD DATA LOCAL INFILE 'yellow_tripdata_2017-02.csv' INTO TABLE nyc_yellow_taxi FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

Select count(*) from nyc_yellow_taxi;

```
2023-01-14 03:12:18 (30.3 MB/s) - '/root/yellow_tripdata_2017-02.csv' saved [863487050/863487050]

[root@ip-172-31-34-81 ~]# mysql -h mysqlassignment1.cuyf7nwfvt5s.us-east-1.rds.amazonaws.com -P 3306 -u srbhrungta -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 3275
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)]> LOAD DATA LOCAL INFILE 'yellow_tripdata_2017-01.csv' INTO TABLE nyc_yellow_taxi FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
ERROR 1046 (3D000): No database selected
MySQL [(none)]> use demo;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
MySQL [demo]> LOAD DATA LOCAL INFILE 'yellow_tripdata_2017-01.csv' INTO TABLE nyc_yellow_taxi FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
Query OK, 9710820 rows affected, 65535 warnings (2 min 44.19 sec)
Records: 9710820  Deleted: 0  Skipped: 0  Warnings: 9710836

MySQL [demo]> select count(*) from nyc_yellow_taxi;
+----------+
| count(*) |
+----------+
|  9710820 |
+----------+
1 row in set (20.30 sec)

MySQL [demo]> LOAD DATA LOCAL INFILE 'yellow_tripdata_2017-02.csv' INTO TABLE nyc_yellow_taxi FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
Query OK, 9169775 rows affected, 65535 warnings (2 min 28.12 sec)
Records: 9169775  Deleted: 0  Skipped: 0  Warnings: 9169801

MySQL [demo]> select count(*) from nyc_yellow_taxi;
+----------+
| count(*) |
+----------+
| 18880595 |
+----------+
1 row in set (43.02 sec)

MySQL [demo]>
```