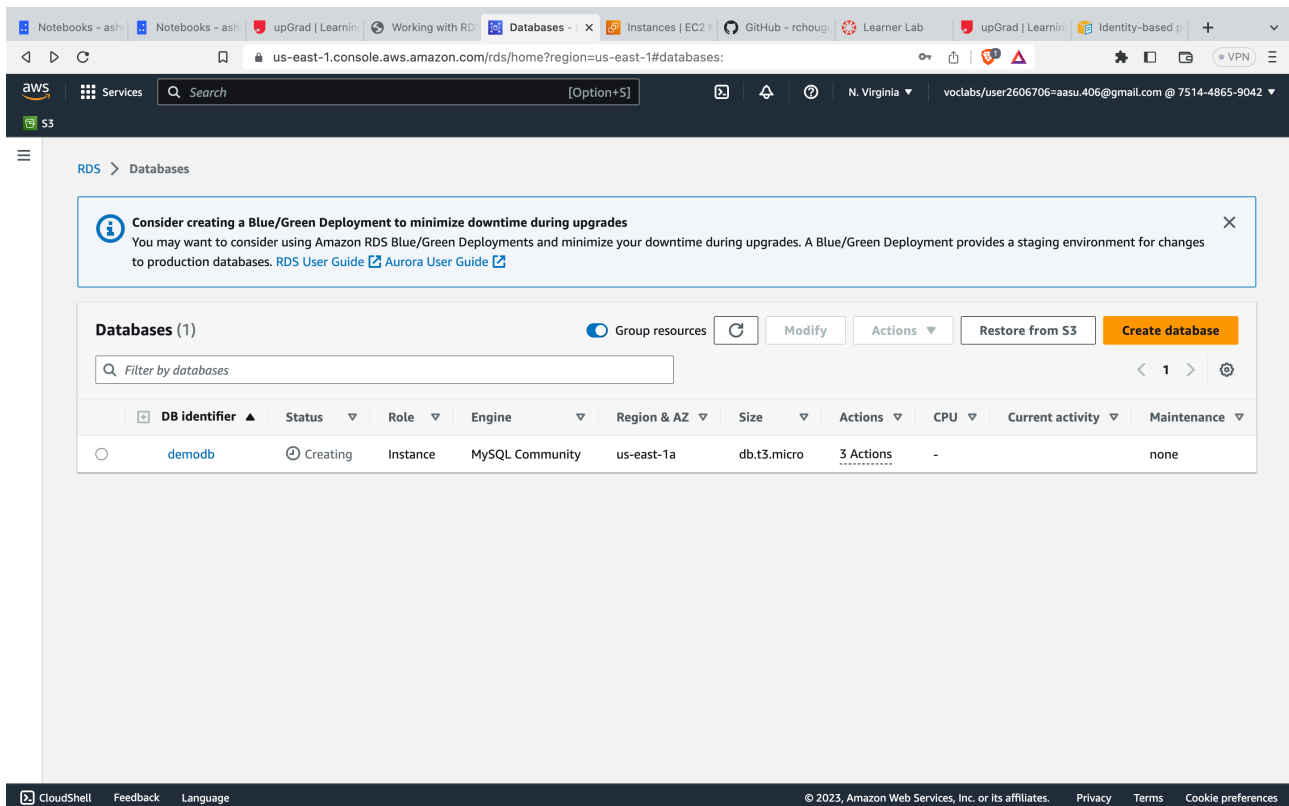


Task 1:

Creating rds cluster in aws, go to aws console search for rds and then click on create cluster select the options according to your requirements and click on create.



Then create emr cluster and open it in your terminal. In your emr cluster give this command to download the civ files.

Wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv

Give these commands to connect to rds.

```
mysql -h demodb.c0tcmelmuto.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
```

And then give password

After entering the db now give this query for creating the table

```
create table data (VendorID INT PRIMARY KEY ,tpep_pickup_datetime DATETIME DEFAULT CURRENT_TIMESTAMP, tpep_dropoff_datetime DATETIME DEFAULT CURRENT_TIMESTAMP, passenger_count INT, trip_distance FLOAT, RatecodeID INT, store_and_fwd_flag VARCHAR (255), PULocationID INT, DOLocationID INT, payment_type INT, fare_amount FLOAT, extra FLOAT, mta_tax FLOAT, tip_amount FLOAT, tolls_amount FLOAT, improvement_surcharge FLOAT, total_amount FLOAT, congestion_surcharge FLOAT, airport_fee FLOAT) ;
```

```

~/Downloads -- ec2-user@ip-172-31-30-45:~ -- zsh ...
...p@ip-172-31-80-28:~ -- ssh -i ~/Downloads/MyEC2.pem hadoop@ec2-3-84-234-77.compute-1.amazonaws.co

[hadoop@ip-172-31-80-28 ~]$ mysql -h demodb.c0tcmelmuto.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 31
Server version: 8.0.32 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)]> create table users
-> (
-> user_id VARCHAR(255),
-> age INT,
-> gender VARCHAR(255),
-> occupation VARCHAR(255),
-> zip_code INT
[
-> ;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near '' at line 7
MySQL [(none)]> create database yellowtrip_data;
Query OK, 1 row affected (0.01 sec)

MySQL [(none)]> use yellowtrip_data;
Database changed
MySQL [yellowtrip_data]> create table data( VendorID INT NOT NULL PRIMARY KEY ,toep_pickup_datetime DATETIME DEFAULT CURRENT_TIMESTAMP,toep_dropoff_datetime DATETIME DEFAULT CURRENT_TIMESTAMP, pa:
ount INT,trip_distance FLOAT,RatecodeID INT,store_and_fwd_flag VARCHAR(255),PULocationID INT,DOLocationID INT,payment_type INT,fare_amount FLOAT,extra FLOAT,mta_tax FLOAT,tip_amount FLOAT,tolls_an
AT,improvement_surcharge FLOAT,total_amount FLOAT,congestion_surcharge FLOAT,airport_fee FLOAT);
Query OK, 0 rows affected (0.05 sec)

MySQL [yellowtrip_data]> SHOW TABLES;
+-----+
| Tables_in_yellowtrip_data |
+-----+
| data |
+-----+
1 row in set (0.00 sec)

MySQL [yellowtrip_data]> █

```

Inserting data csv data using emr in RDS database

```

LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-03.csv'
INTO TABLE data
FIELDS TERMINATED BY ','
LINES TERMINATED BY '\n'
IGNORE 1 LINES;

```

```

~/Downloads -- ec2-user@ip-172-31-80-45:~ -- zsh
p@ip-172-31-80-28:~ -- ssh -i ~/Downloads/MyEC2-pem hadoop@ec2-3-84-234-77.compute-1.amazonaws.com
ERROR 2 (HY000): File '/home/hadoop@ip-172-31-80-28/yellow_tripdata_2017-01.csv' not found (Errcode: 2)
mysql [(none)]> LOAD DATA LOCAL INFILE '/home/hadoop@ip-172-31-80-28/yellow_tripdata_2017-01.csv' INTO TABLE data FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
ERROR 2 (HY000): File '/home/hadoop@ip-172-31-80-28/yellow_tripdata_2017-01.csv' not found (Errcode: 2)
MySQL [yellowtrip_data]> LOAD DATA LOCAL INFILE '/Users/hadoop@ip-172-31-80-28/yellow_tripdata_2017-01.csv' INTO TABLE data FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
ERROR 2 (HY000): File '/Users/hadoop@ip-172-31-80-28/yellow_tripdata_2017-01.csv' not found (Errcode: 2)
MySQL [yellowtrip_data]> exit
Bye
[hadoop@ip-172-31-80-28 ~]$ sudo -i

EEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRR
E::::::::::::::::::::E M:::::M M:::::M R:::::::::::::R
EE::::::::EEEEEEEEEE E M:::::M M:::::M R::::RRRRR::::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 R::RRRRR::::R
E::::::::::::E M::::M M::M M::M R:::::RR
E:::EEEEEEEEEE M::::M M::::M M::::M R::RRRRR::::R
E:::E M::::M M::M M::::M R::R R::::R
E:::E EEEE M::::M MM 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
EEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRR RRRRRR

[root@ip-172-31-80-28 ~]# ls
[root@ip-172-31-80-28 ~]# ll
total 0
[root@ip-172-31-80-28 ~]# cd /
[root@ip-172-31-80-28 /]# ls
bin boot dev device-names.json emr etc home lib lib64 local media mnt mnt1 opt proc root run/sbin/srv/sys/tmp/usr/var
[root@ip-172-31-80-28 /]# cd /home
[root@ip-172-31-80-28 home]# ls
ec2-user hadoop
[root@ip-172-31-80-28 home]# cd /home/hadoop
[root@ip-172-31-80-28 hadoop]# ls
yellow_tripdata_2017-01.csv yellow_tripdata_2017-02.csv
[root@ip-172-31-80-28 hadoop]# exit
logout
[hadoop@ip-172-31-80-28 ~]$ mysql -h demodb.c0tcmelmufso.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
Enter password:
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MySQL connection id is 39
Server version: 8.0.32 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 yellowtrip_data;
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 [yellowtrip_data]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv'
-> INTO TABLE data
-> FIELDS TERMINATED BY ','
-> LINES TERMINATED BY '\n'
-> IGNORE 1 LINES;
Query OK, 2 rows affected, 65535 warnings (58.22 sec)
Records: 9710820 Deleted: 0 Skipped: 9710818 Warnings: 29132458

MySQL [yellowtrip_data]>

```

```

100%[=====>] 910,028,408 42.3MB/s in 21s

2023-07-09 14:53:54 (41.8 MB/s) - 'yellow_tripdata_2017-06.csv' saved [910028408/910028408]

[hadoop@ip-172-31-92-203 ~]$ mysql -h demodb.c0tcmelmuto.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
Enter password:
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MySQL connection id is 1019
Server version: 8.0.32 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 databases;
+-----+
| Database |
+-----+
| demo |
| information_schema |
| mysql |
| performance_schema |
| sys |
| yellow_trip_1 |
| yellowtrip_data |
+-----+
7 rows in set (0.01 sec)

MySQL [(none)]> use yellowtrip_data
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 [yellowtrip_data]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv'
-> INTO TABLE data
-> FIELDS TERMINATED BY ','
-> LINES TERMINATED BY '\n'
-> IGNORE 1 LINES;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'LINES;' at line 5
MySQL [yellowtrip_data]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-03.csv' INTO TABLE data FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'LINES;' at line 1
MySQL [yellowtrip_data]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-03.csv'
-> INTO TABLE data
-> FIELDS TERMINATED BY ','
-> LINES TERMINATED BY '\n'
-> IGNORE 1 LINES;
Query OK, 0 rows affected, 65535 warnings (50.42 sec)
Records: 10295441 Deleted: 0 Skipped: 10295441 Warnings: 30886323

MySQL [yellowtrip_data]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-04.csv' INTO TABLE data FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
Query OK, 0 rows affected, 65535 warnings (49.51 sec)
Records: 10047135 Deleted: 0 Skipped: 10047135 Warnings: 30141405

MySQL [yellowtrip_data]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-05.csv' INTO TABLE data FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
Query OK, 0 rows affected, 65535 warnings (56.98 sec)
Records: 10102127 Deleted: 0 Skipped: 10102127 Warnings: 30306381

MySQL [yellowtrip_data]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-06.csv' INTO TABLE data FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
Query OK, 0 rows affected, 65535 warnings (48.08 sec)
Records: 9656993 Deleted: 0 Skipped: 9656993 Warnings: 28970979

MySQL [yellowtrip_data]> █

```