

1) C:\Users\antar>e:

2) E:\MS in Data Science\MS-DataScience\MapReduce - Programming Assignment\Login Files> ssh -i "PEM.pem" ec2-user@ec2-54-158-208-61.compute-1.amazonaws.com

Last login: Tue Oct 31 15:55:29 2023 from 103.166.14.168

3) RDS created in AWS. Giving those details below to login to RDS through EMR

```
[ec2-user@ip-172-31-52-19 ~]$ mysql -h demodb.cr64kmfleoum.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 20

Server version: 8.0.33 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.

4) MySQL [(none)]> show databases;

```
+-----+
| Database          |
+-----+
| information_schema |
| mysql              |
| performance_schema |
| sys                |
+-----+
```

4 rows in set (0.02 sec)

5)MySQL [(none)]> create database demo;

Query OK, 1 row affected (0.01 sec)

6)MySQL [(none)]> use demo;

Database changed

7)MySQL [demo]> CREATE TABLE nyc_taxi_1_2 (

```
->     VendorID INT,  
->     tpep_pickup_datetime DATETIME,  
->     tpep_dropoff_datetime DATETIME,  
->     passenger_count INT,  
->     trip_distance DECIMAL(10, 2),  
->     RatecodeID INT,  
->     store_and_fwd_flag CHAR(1),  
->     PULocationID INT,  
->     DOLocationID INT,  
->     payment_type INT,  
->     fare_amount DECIMAL(10, 2),  
->     extra DECIMAL(5, 2),  
->     mta_tax DECIMAL(5, 2),  
->     tip_amount DECIMAL(10, 2),  
->     tolls_amount DECIMAL(10, 2),  
->     improvement_surcharge DECIMAL(5, 2),  
->     total_amount DECIMAL(10, 2),  
->     congestion_surcharge DECIMAL(5, 2),  
->     airport_fee DECIMAL(10, 2)  
-> );
```

Query OK, 0 rows affected (0.05 sec)

8)MySQL [demo]> show tables;

```

+-----+
| Tables_in_demo |
+-----+
| nyc_taxi_1_2 |
+-----+
1 row in set (0.00 sec)

```

9) MySQL [demo]> Select * from nyc_taxi_1_2;

Empty set (0.00 sec)

MySQL [demo]> Bye

10) Importing Dataset Below through wget

[ec2-user@ip-172-31-62-58 ~]\$ wget

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

[ec2-user@ip-172-31-62-58 ~]\$ wget

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

```

[ec2-user@ip-172-31-54-160 ~]$ https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-03.csv
bash: https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-03.csv: No such file or directory
[ec2-user@ip-172-31-54-160 ~]$ wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-03.csv
--2023-11-01 14:41:52-- https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-03.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 52.216.41.105, 52.216.49.129, 52.2
1, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|52.216.41.105|:443... connected.
HTTP request sent, awaiting response... 200 OK
Content-length: 969809025 (925M) [text/csv]
Saving to: 'yellow_tripdata_2017-03.csv'

0.00%[=====] 969,809,025 12.9MB/s in 5
2023-11-01 14:42:46 (17.2 MB/s) - 'yellow_tripdata_2017-03.csv' saved [969809025/969809025]

[ec2-user@ip-172-31-54-160 ~]$ wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-04.csv
--2023-11-01 14:42:57-- https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-04.csv
Resolving nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)... 52.217.136.185, 52.217.137.33, 52.
129, ...
Connecting to nyc-tlc-upgrad.s3.amazonaws.com (nyc-tlc-upgrad.s3.amazonaws.com)|52.217.136.185|:443... connected.
HTTP request sent, awaiting response... 200 OK
Content-length: 946349441 (903M) [text/csv]
Saving to: 'yellow_tripdata_2017-04.csv'

0.00%[=====] 946,349,441 19.7MB/s in 5
2023-11-01 14:43:48 (17.8 MB/s) - 'yellow_tripdata_2017-04.csv' saved [946349441/946349441]

[ec2-user@ip-172-31-54-160 ~]$

```

```
[ec2-user@ip-172-31-62-58 ~]$ ls
```

```
yellow_tripdata_2017-01.csv
```

```
yellow_tripdata_2017-02.csv
```

```
[ec2-user@ip-172-31-62-58 ~]$ pwd
```

```
/home/ec2-user
```

11) Logging into MariaDB again

```
[ec2-user@ip-172-31-62-58 ~]$ mysql -h demodb.cr64kmfleoum.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 30

Server version: 8.0.33 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 demo;
```

12) Loading Dataset into table

```
MySQL [(demo)]> ]> LOAD DATA LOCAL INFILE '/home/ec2-user/yellow_tripdata_2017-01.csv' INTO  
TABLE nyc_taxi_1_2 FIELDS TERMINATED BY ',';
```

Query OK, 9710821 rows affected, 65535 warnings (2 min 35.81 sec)

Records: 9710821 Deleted: 0 Skipped: 0 Warnings: 19421659

```
MySQL [demo]> Select passenger_count from nyc_taxi_1_2 limit 10
```

```
-> ;
```

```
+-----+
```

```
| passenger_count |
```

```
+-----+
```

```
|          0 |
```

```
|          1 |
```

```
|          2 |
```

```
|          2 |
```

```
|          1 |
```

```
|          1 |
```

```
|          2 |
```

```
|          2 |
```

```
|          2 |
```

```
|          2 |
```

```
+-----+
```

```
10 rows in set (0.01 sec)
```

```
MySQL [demo]> Select * from nyc_taxi_1_2 limit 10;
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+
```

```
| VendorID | tpep_pickup_datetime | tpep_dropoff_datetime | passenger_count | trip_distance |
RatecodeID | store_and_fwd_flag | PULocationID | DOLocationID | payment_type | fare_amount |
extra | mta_tax | tip_amount | tolls_amount | improvement_surcharge | total_amount |
congestion_surcharge | airport_fee |
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+
```

	0	0000-00-00 00:00:00	0000-00-00 00:00:00			0	
0.00		0	s			0	
0.00		0.00		0.00		0.00	
		0.00		0.00			0.00
	1	2017-01-01 00:32:05	2017-01-01 00:37:48			1	
1.20		1	N		140		2
6.50		0.50		0.00		0.30	
		0.00		0.00			7.80
	1	2017-01-01 00:43:25	2017-01-01 00:47:42			2	
0.70		1	N		237		2
5.00		0.50		0.00		0.30	
		0.00		0.00			6.30
	1	2017-01-01 00:49:10	2017-01-01 00:53:53			2	
0.80		1	N		140		2
5.50		0.50		0.00		0.30	
		0.00		0.00			6.80
	1	2017-01-01 00:36:42	2017-01-01 00:41:09			1	
1.10		1	N		41		2
6.00		0.50		0.00		0.30	
		0.00		0.00			7.30
	1	2017-01-01 00:07:41	2017-01-01 00:18:16			1	
3.00		1	N		48		2
11.00		0.50		0.00		0.30	
12.30				0.00			
				0.00			
	1	2017-01-01 00:20:52	2017-01-01 00:24:59			2	
0.70		1	N		236		2
5.00		0.50		0.00		0.30	
		0.00		0.00			6.30
	1	2017-01-01 00:33:49	2017-01-01 00:42:38			2	
1.60		1	N		236		1
8.00		0.50		1.85		0.30	
		0.00		0.00			11.15
	1	2017-01-01 00:48:22	2017-01-01 00:52:15			2	
0.60		1	N		238		1
5.00		0.50		1.25		0.30	
		0.00		0.00			7.55
	1	2017-01-01 00:57:12	2017-01-01 01:06:28			2	
1.00		1	N		239		1

7.50	0.50	0.50	1.75	0.00	0.30	10.55
		0.00	0.00			

```

+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+
---+-----+-----+

```

10 rows in set (0.00 sec)

MySQL [demo]> LOAD DATA LOCAL INFILE '/home/ec2-user/yellow_tripdata_2017-02.csv' INTO TABLE nyc_taxi_1_2 FIELDS TERMINATED BY ',';

Query OK, 9169776 rows affected, 65535 warnings (2 min 27.03 sec)

Records: 9169776 Deleted: 0 Skipped: 0 Warnings: 18339569

MySQL [demo]> Select * from nyc_taxi_1_2 limit 10;

```

+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+
---+-----+-----+

```

VendorID tpep_pickup_datetime tpep_dropoff_datetime passenger_count trip_distance
RatecodeID store_and_fwd_flag PULocationID DOLocationID payment_type fare_amount
extra mta_tax tip_amount tolls_amount improvement_surcharge total_amount
congestion_surcharge airport_fee

```

+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+
---+-----+-----+

```

	0	0000-00-00 00:00:00		0000-00-00 00:00:00		0
0.00	0	s		0	0	0
0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	0.00			

	1	2017-01-01 00:32:05		2017-01-01 00:37:48		1
1.20	1	N		140	236	2
6.50	0.50	0.50	0.00	0.00	0.30	7.80
		0.00	0.00			

	1	2017-01-01 00:43:25		2017-01-01 00:47:42		2
0.70	1	N		237	140	2
5.00	0.50	0.50	0.00	0.00	0.30	6.30
		0.00	0.00			

	1	2017-01-01 00:49:10	2017-01-01 00:53:53		2			
0.80		1	N		140		237	2
5.50	0.50	0.50	0.00	0.00			0.30	6.80
		0.00	0.00					
	1	2017-01-01 00:36:42	2017-01-01 00:41:09		1			
1.10		1	N		41		42	2
6.00	0.50	0.50	0.00	0.00			0.30	7.30
		0.00	0.00					
	1	2017-01-01 00:07:41	2017-01-01 00:18:16		1			
3.00		1	N		48		263	2
11.00	0.50	0.50	0.00	0.00			0.30	
12.30			0.00	0.00				
	1	2017-01-01 00:20:52	2017-01-01 00:24:59		2			
0.70		1	N		236		262	2
5.00	0.50	0.50	0.00	0.00			0.30	6.30
		0.00	0.00					
	1	2017-01-01 00:33:49	2017-01-01 00:42:38		2			
1.60		1	N		236		238	1
8.00	0.50	0.50	1.85	0.00			0.30	11.15
		0.00	0.00					
	1	2017-01-01 00:48:22	2017-01-01 00:52:15		2			
0.60		1	N		238		239	1
5.00	0.50	0.50	1.25	0.00			0.30	7.55
		0.00	0.00					
	1	2017-01-01 00:57:12	2017-01-01 01:06:28		2			
1.00		1	N		239		48	1
7.50	0.50	0.50	1.75	0.00			0.30	10.55
		0.00	0.00					

```

+-----+-----+-----+-----+-----+-----+-----+-----+
-----+-----+-----+-----+-----+-----+-----+-----+
---+-----+-----+

```

10 rows in set (0.04 sec)

MySQL [demo]> Select count(*) from nyc_taxi_1_2;

```

+-----+
| count(*) |
+-----+

```


| 18880597 |

+-----+

1 row in set (40.76 sec)

```
10 rows in set (0.02 sec)
```

MySQL [demo]> Select * from nyc_taxi_1_2 limit 10;

VendorID	time_pickup_datetime	time_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fwd_flag	PULocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount	congestion_surcharge	airport_fee
0	0000-00-00 00:00:00	0000-00-00 00:00:00	0	0.00	0	s		0		0								
0	00	00	0	0	0													
1	2017-01-01 00:32:05	2017-01-01 00:37:48	1	1.20	1	N		140		236								2
6	50	0	0	0	0													
1	2017-01-01 00:43:25	2017-01-01 00:47:42	2	0.70	1	N		237		140								2
5	00	0	0	0	0													
1	2017-01-01 00:49:10	2017-01-01 00:53:53	2	0.80	1	N		140		237								2
5	00	0	0	0	0													
1	2017-01-01 00:36:42	2017-01-01 00:41:09	1	1.10	1	N		41		42								2
6	00	0	0	0	0													
1	2017-01-01 00:07:41	2017-01-01 00:18:16	1	3.00	1	N		48		263								2
11	00	0	0	0	0													
1	2017-01-01 00:20:52	2017-01-01 00:24:59	2	0.70	1	N		236		262								2
5	00	0	0	0	0													
1	2017-01-01 00:33:49	2017-01-01 00:42:38	2	1.60	1	N		236		238								1
8	00	0	0	0	0													
1	2017-01-01 00:48:22	2017-01-01 00:52:15	2	0.60	1	N		238		239								1
5	00	0	0	0	0													
1	2017-01-01 00:57:12	2017-01-01 01:06:28	2	1.00	1	N		239		48								1
7	50	0	0	0	0													

10 rows in set (0.00 sec)

```
MySQL [demo]> LOAD DATA LOCAL INFILE '/home/ec2-user/yellow_tripdata_2017-02.csv' INTO TABLE nyc_taxi_1_2 FIELDS TERMINATED BY ',';
```

Query OK, 9169776 rows affected, 65535 warnings (2 min 27.03 sec)
Records: 9169776 Deleted: 0 Skipped: 0 Warnings: 18339569

MySQL [demo]> Select * from nyc_taxi_1_2 limit 10;

VendorID	time_pickup_datetime	time_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fwd_flag	PULocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount	congestion_surcharge	airport_fee
0	0000-00-00 00:00:00	0000-00-00 00:00:00	0	0.00	0	s		0		0								
0	00	00	0	0	0													
1	2017-01-01 00:32:05	2017-01-01 00:37:48	1	1.20	1	N		140		236								2
6	50	0	0	0	0													
1	2017-01-01 00:43:25	2017-01-01 00:47:42	2	0.70	1	N		237		140								2
5	00	0	0	0	0													
1	2017-01-01 00:49:10	2017-01-01 00:53:53	2	0.80	1	N		140		237								2
5	00	0	0	0	0													
1	2017-01-01 00:36:42	2017-01-01 00:41:09	1	1.10	1	N		41		42								2
6	00	0	0	0	0													
1	2017-01-01 00:07:41	2017-01-01 00:18:16	1	3.00	1	N		48		263								2
11	00	0	0	0	0													
1	2017-01-01 00:20:52	2017-01-01 00:24:59	2	0.70	1	N		236		262								2
5	00	0	0	0	0													
1	2017-01-01 00:33:49	2017-01-01 00:42:38	2	1.60	1	N		236		238								1
8	00	0	0	0	0													
1	2017-01-01 00:48:22	2017-01-01 00:52:15	2	0.60	1	N		238		239								1
5	00	0	0	0	0													
1	2017-01-01 00:57:12	2017-01-01 01:06:28	2	1.00	1	N		239		48								1
7	50	0	0	0	0													

10 rows in set (0.04 sec)

MySQL [demo]> Select count(*) from nyc_taxi_1_2;

count(*)
18880597

We can clearly see the loaded tables above in RDS database that were loaded from our EMR. Thus this concludes our first task.

