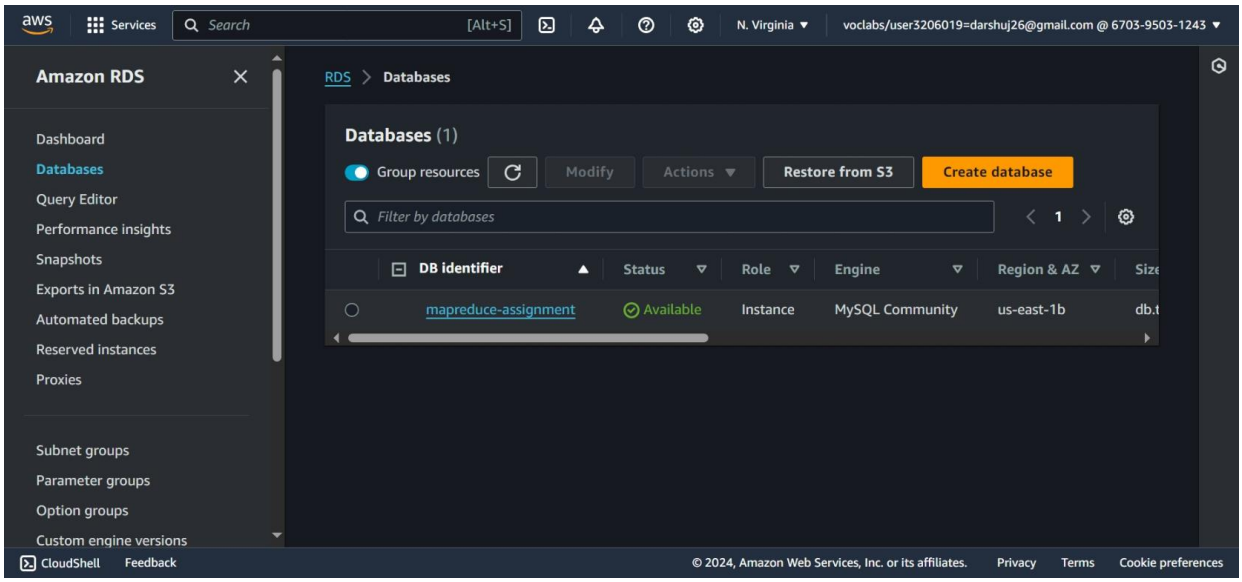
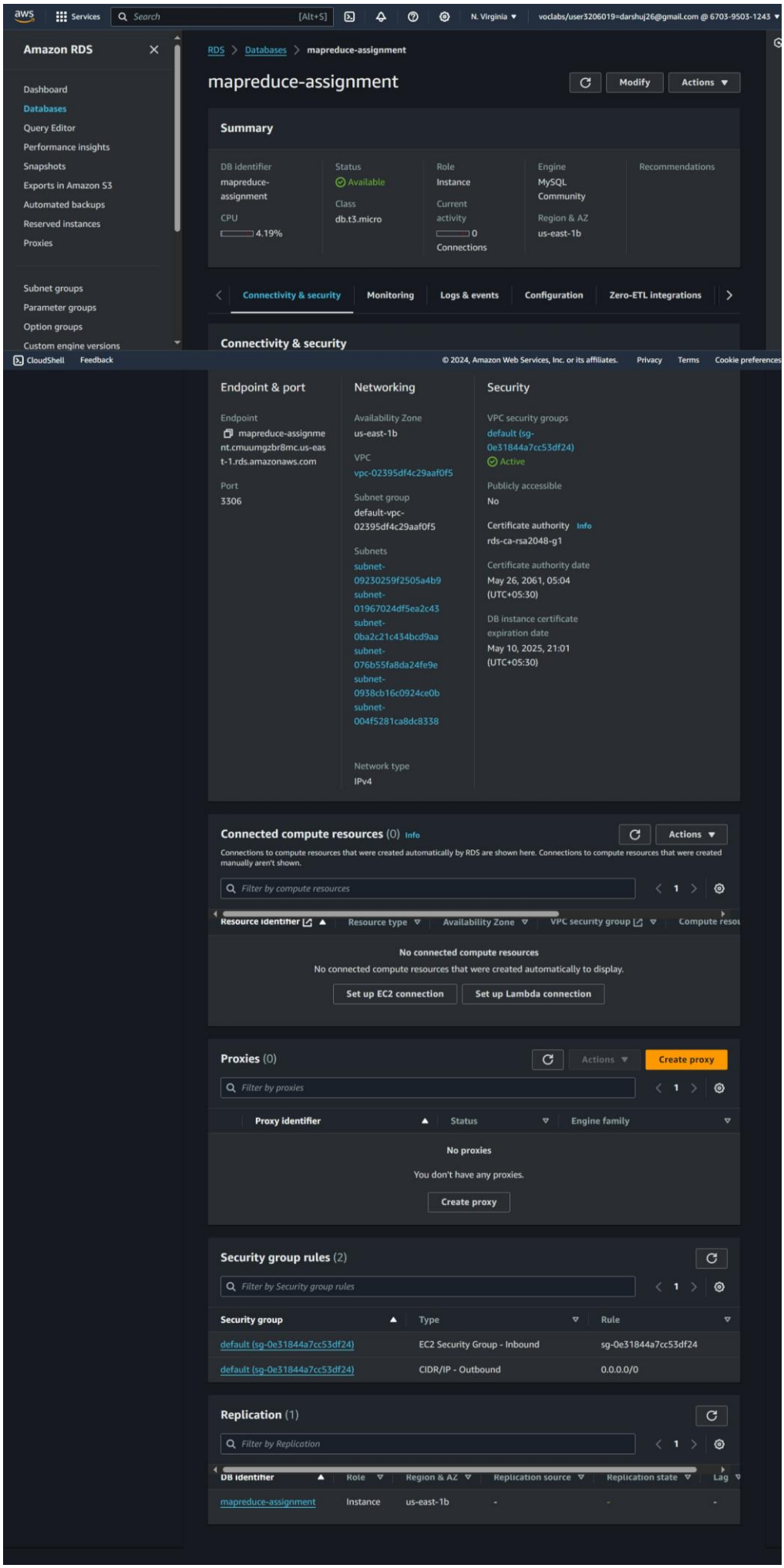
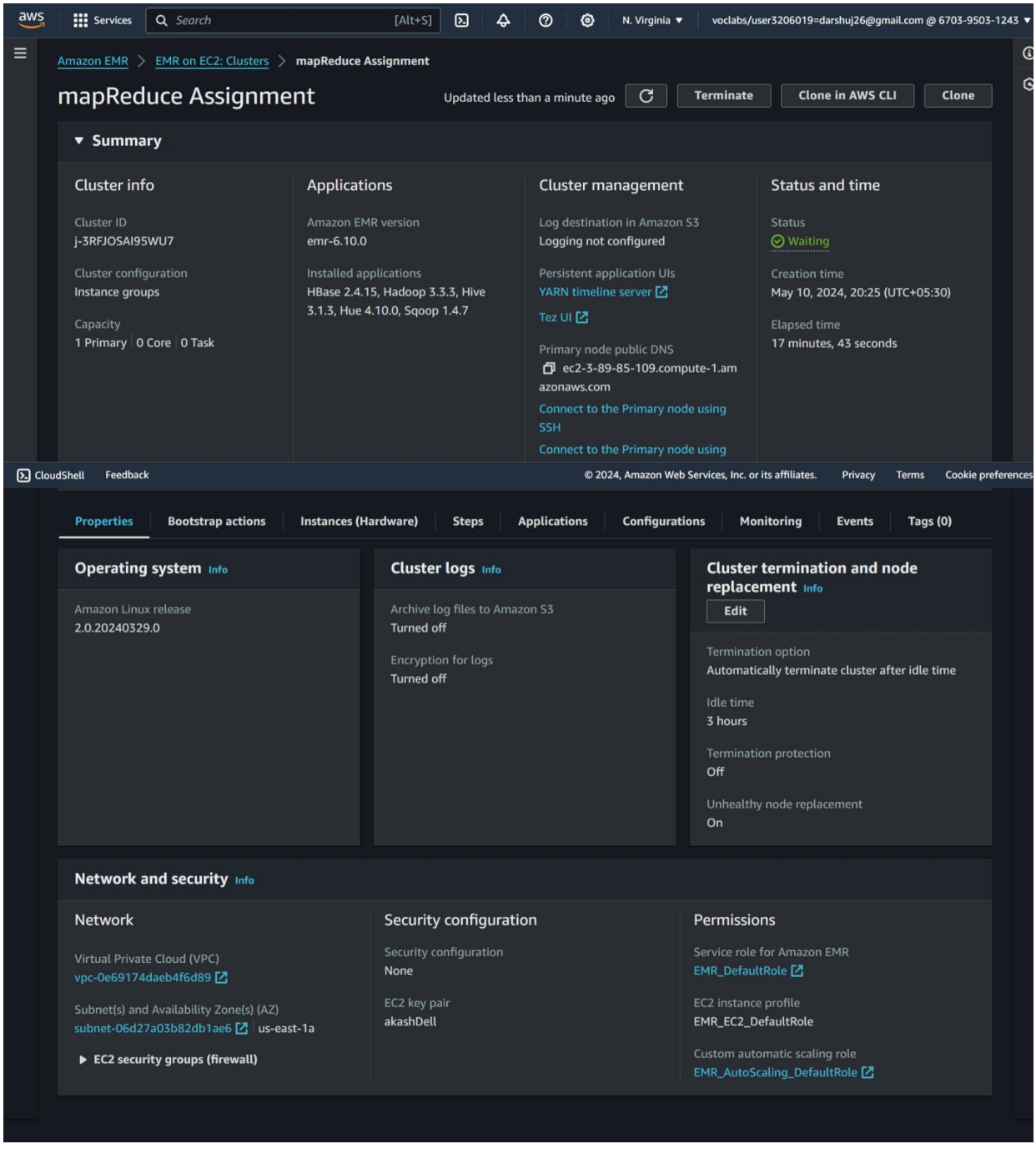
1. **Creation of RDS instance in AWS**

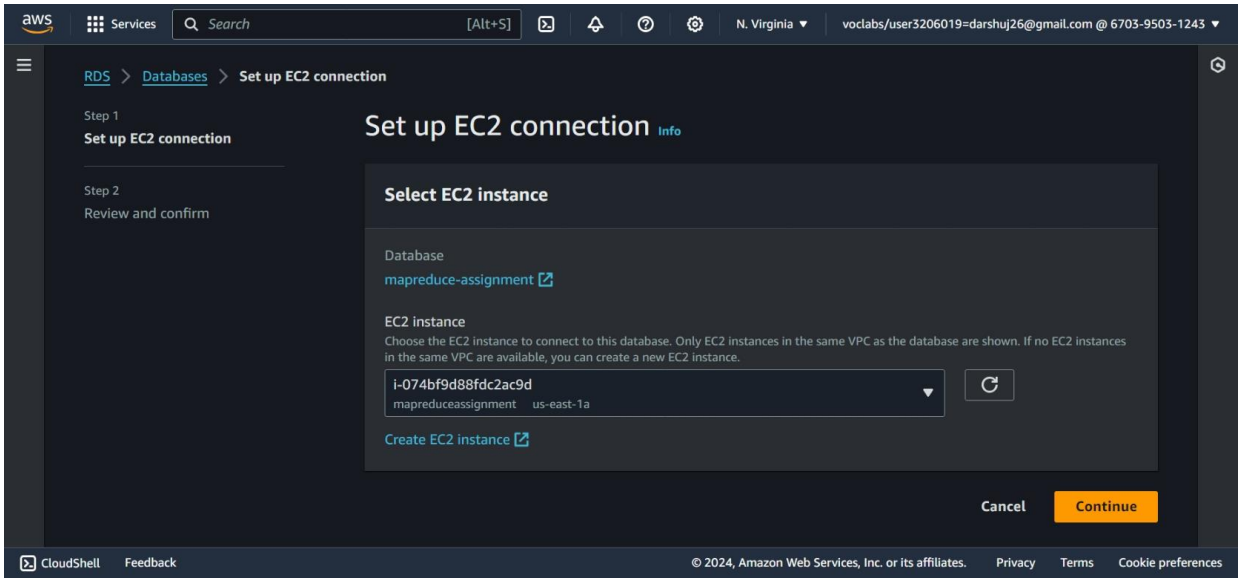



aws
Services
Search
[Alt+S]
N. Virginia
voclabs/user3206019=darshuj26@gmail.com @ 6703-9503-1243
Amazon RDS
Dashboard
Databases
Query Editor
Performance insights
Snapshots
Exports in Amazon S3
Automated backups
Reserved instances
Proxies
Subnet groups
Parameter groups
Option groups
Custom engine versions
RDS > Databases > mapreduce-assignment
mapreduce-assignment
Modify
Actions
Summary
DB identifier
mapreduce-assignment
Status
Available
Class
db.t3.micro
CPU
4.19%
Role
Instance
Current activity
0 Connections
Engine
MySQL Community
Region & AZ
us-east-1b
Recommendations
Connectivity & security
Monitoring
Logs & events
Configuration
Zero-ETL integrations
Connectivity & security
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences
Endpoint & port
Endpoint
mapreduce-assignment.cmuumgzbr8mc.us-east-1.rds.amazonaws.com
Port
3306
Networking
Availability Zone
us-east-1b
VPC
vpc-02395df4c29aaf0f5
Subnet group
default-vpc-02395df4c29aaf0f5
Subnets
subnet-09230259f2505a4b9
subnet-01967024df5ea2c43
subnet-0ba2c21c434bcd9aa
subnet-076b55fa8da24fe9e
subnet-0938cb16c0924ce0b
subnet-004f5281ca8dc8338
Network type
IPv4
Security
VPC security groups
default (sg-0e31844a7cc53df24)
Active
Publicly accessible
No
Certificate authority
rds-ca-rsa2048-g1
Certificate authority date
May 26, 2061, 05:04 (UTC+05:30)
DB instance certificate expiration date
May 10, 2025, 21:01 (UTC+05:30)
Connected compute resources (0)
Actions
Connections to compute resources that were created automatically by RDS are shown here. Connections to compute resources that were created manually aren't shown.
Filter by compute resources
Resource identifier
Resource type
Availability Zone
VPC security group
No connected compute resources
No connected compute resources that were created automatically to display.
Set up EC2 connection
Set up Lambda connection
Proxies (0)
Actions
Create proxy
Filter by proxies
Proxy identifier
Status
Engine family
No proxies
You don't have any proxies.
Create proxy
Security group rules (2)
Filter by Security group rules
Security group
Type
Rule
default (sg-0e31844a7cc53df24)
EC2 Security Group - Inbound
sg-0e31844a7cc53df24
default (sg-0e31844a7cc53df24)
CIDR/IP - Outbound
0.0.0.0/0
Replication (1)
Filter by Replication
DB identifier
Role
Region & AZ
Replication source
Replication state
Lag
mapreduce-assignment
Instance
us-east-1b

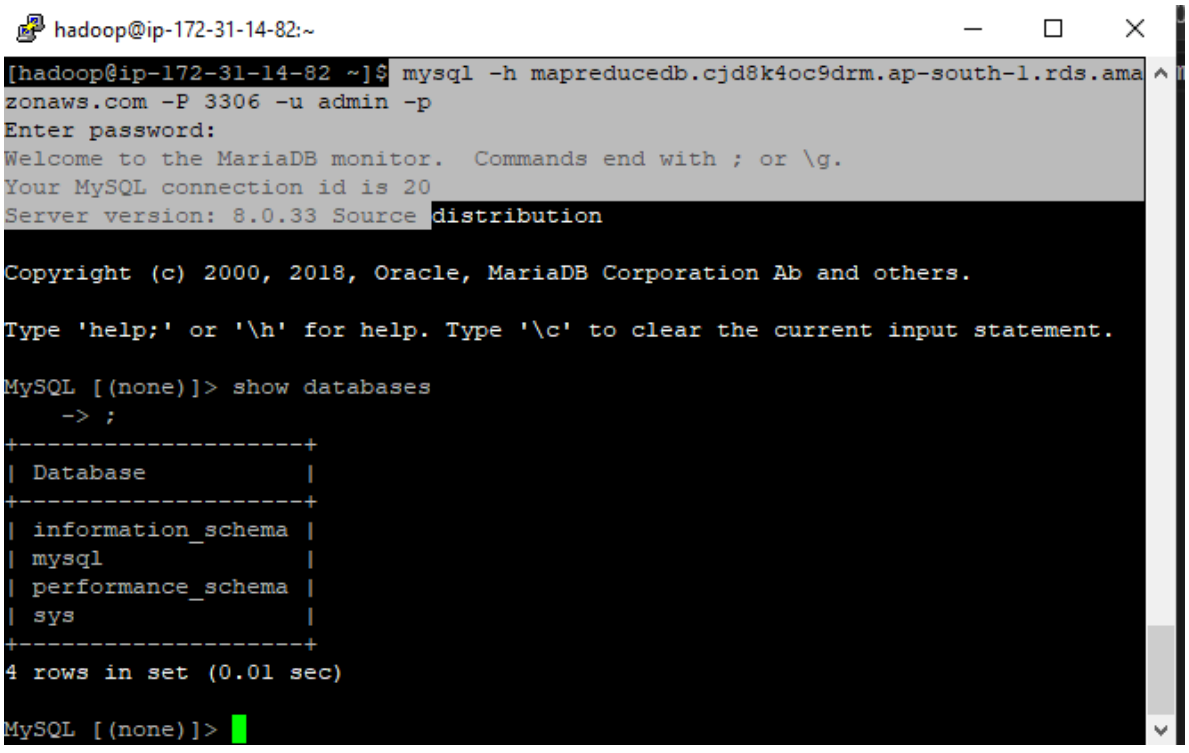
2. **Creation of EMR instance with bundled applications such as Hadoop, Hbase, Hue, Hive, Sqoop.**



3. **To connect RDS with EMR instance, we have to click on “Action” button on RDS instance menu and then “Set up an EC2 Instance”.**



4. **Login to RDS through EMR instance using command: `mysql -h demodb.cluster-cjd8k4oc9drm.ap-south-1.rds.amazonaws.com -P 3306 -u admin -p`**

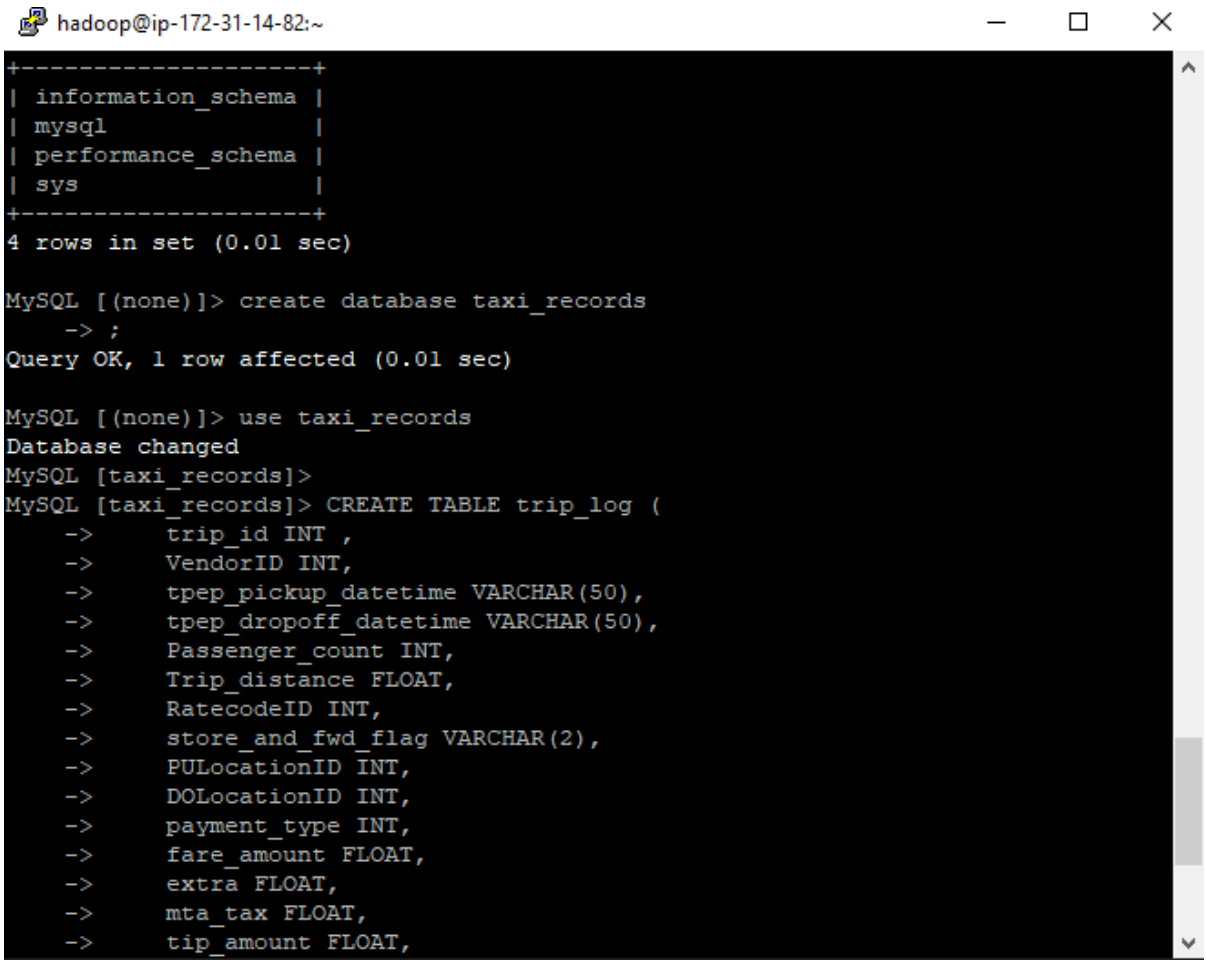


5. **Create Database “taxi_records”**

`create database taxi_records'

6. Create table "trip_log"

```
CREATE TABLE trip_log (
   trip_id INT,
   VendorID INT,
   tpep_pickup_datetime VARCHAR(50),
   tpep_dropoff_datetime VARCHAR(50),
   Passenger_count INT,
   Trip_distance FLOAT,
   RatecodeID INT,
   store_and_fwd_flag VARCHAR(2),
   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,
   Airport_fee FLOAT
);
```



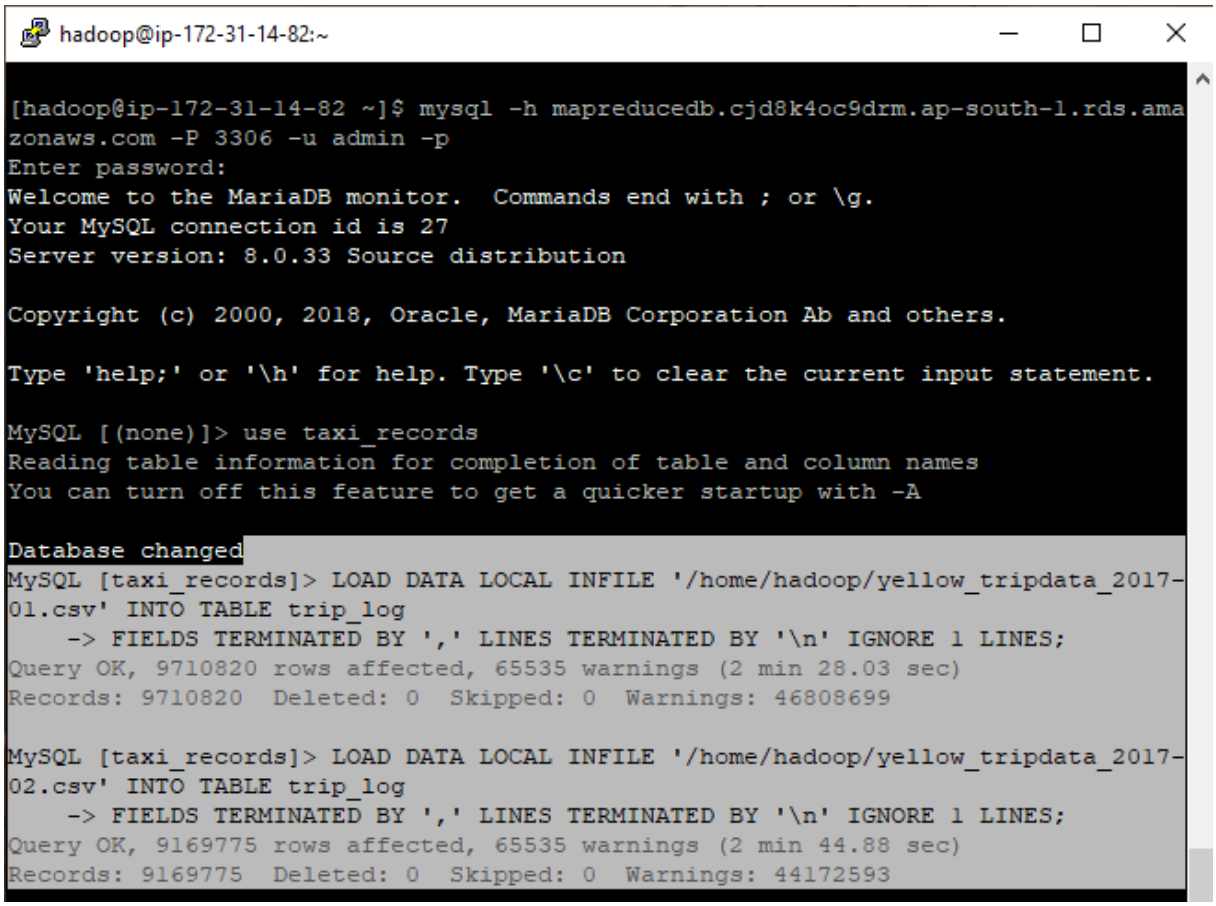
7. **Downloading required csv files from internet in local using command**

```
wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-01.csv
wget https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv
```

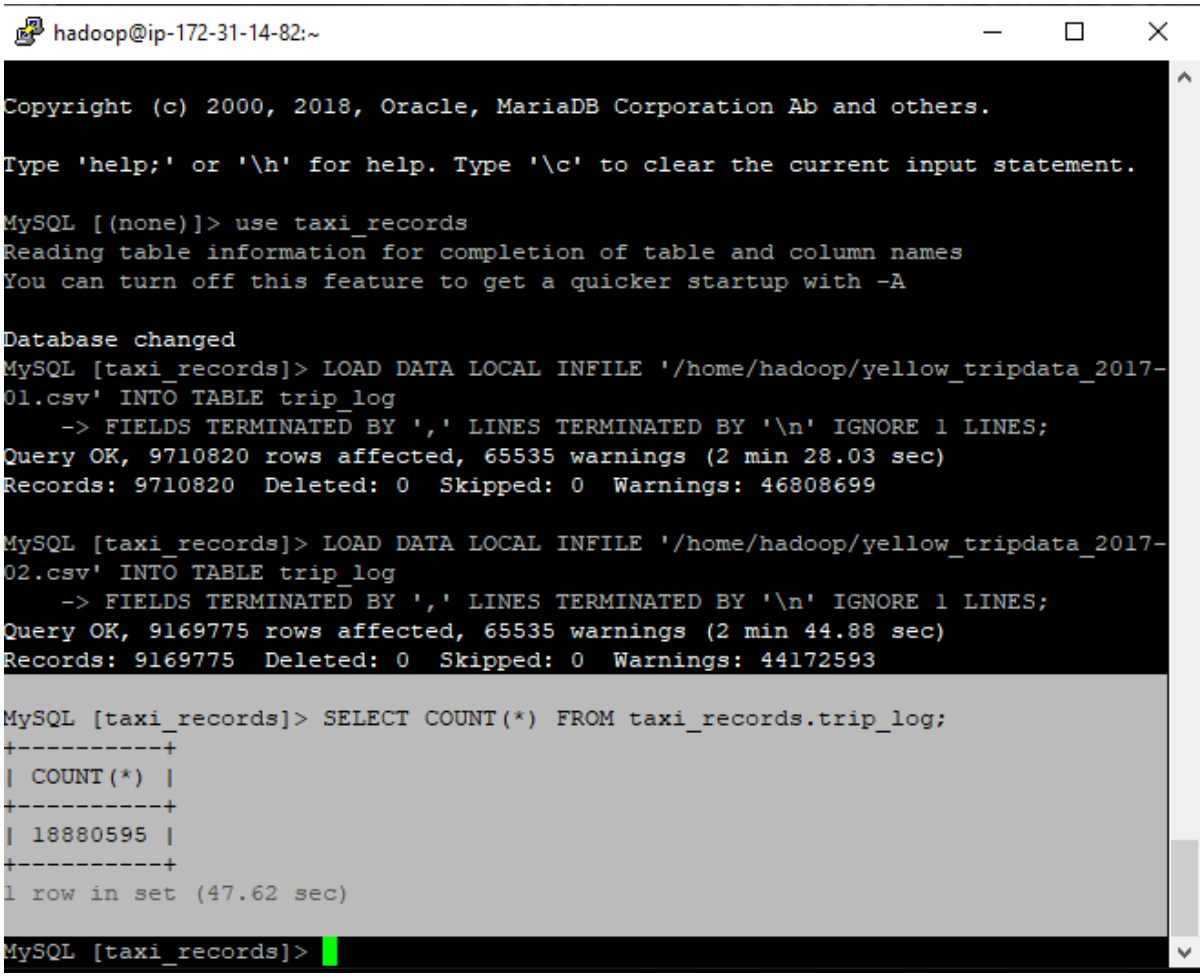
8. **To load data in mysql table we have to login and then run sql command:**

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

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




```
SELECT COUNT(*) FROM taxi_records.trip_log;
```


hadoop@ip-172-31-14-82:~
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 taxi_records
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 [taxi_records]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv' INTO TABLE trip_log
-> FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
Query OK, 9710820 rows affected, 65535 warnings (2 min 28.03 sec)
Records: 9710820 Deleted: 0 Skipped: 0 Warnings: 46808699
MySQL [taxi_records]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-02.csv' INTO TABLE trip_log
-> FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
Query OK, 9169775 rows affected, 65535 warnings (2 min 44.88 sec)
Records: 9169775 Deleted: 0 Skipped: 0 Warnings: 44172593
MySQL [taxi_records]> SELECT COUNT(*) FROM taxi_records.trip_log;
+----------+
| COUNT(*) |
+----------+
| 18880595 |
+----------+
1 row in set (47.62 sec)
MySQL [taxi_records]>