

## 1. Creation of RDS Instance In AWS

The screenshot displays the AWS Management Console for an Amazon RDS instance. The browser address bar shows the URL: `us-east-1.console.aws.amazon.com/rds/home?region=us-east-1#databaseid=database-1;is-cluster=false`. The console header includes the AWS logo, a search bar, and navigation links for United States (N. Virginia) and the user profile.

The left sidebar shows the navigation menu for Amazon RDS, with 'Databases' selected. The main content area displays the details for 'database-1'.

**Summary**

DB identifier	Status	Role	Engine	Recommendations
database-1	Available	Instance	MySQL Community	

CPU	Class	Current activity	Region & AZ
6.99%	db.t3.micro	0	us-east-1d

Below the summary, there are tabs for 'Connectivity & security', 'Monitoring', 'Logs & events', 'Configuration', 'Zero-ETL integrations', and 'Maintenance'. The 'Connectivity & security' tab is active.

**Connectivity & security**

Endpoint & port	Networking	Security
Endpoint database-1.cm4ah2fwr8jc.us-east-1.rds.amazonaws.com	Availability Zone us-east-1d VPC	VPC security groups default (sg-0154caebf4949f20c) Active

The bottom of the screenshot shows the Windows taskbar with various application icons and the system clock displaying 15:11 on 06-03-2025.

2.

## 3. Creation Of EMR Instance with bundled applications such as Hadoop,hbase ,sqoop

us-east-1.console.aws.amazon.com/emr/home?region=us-east-1#/clusterDetails/j-22AFSQDCAEO4M

aws Search [Alt+S] United States (N. Virginia) voclabs/user3817529=rajatgarg199318@gmail.com @ 4422-4089-1847

Amazon EMR > EMR on EC2: Clusters > My cluster

✓ Your cluster 'My cluster' has been successfully created.

**My cluster** Updated less than a minute ago [Terminate](#) [Clone in AWS CLI](#) [Clone](#)

### ▼ Summary

Cluster info	Applications	Cluster management	Status and time
<b>Cluster ID</b> j-22AFSQDCAEO4M	<b>Amazon EMR version</b> emr-7.4.0	<b>Log destination in Amazon S3</b> <a href="#">aws-logs-442240891847-us-east-1/elasticmapreduce</a>	<b>Status</b> <span>✓ Waiting</span>
<b>Cluster ARN</b> <a href="#">arn:aws:elasticmapreduce:us-east-1:442240891847:cluster/j-22AFSQDCAEO4M</a>	<b>Installed applications</b> Hadoop 3.4.0, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5.2, Sqoop 1.4.7	<b>Persistent application UIs</b> <a href="#">Spark history server</a> <a href="#">YARN timeline server</a> <a href="#">Tez UI</a>	<b>Creation time</b> 6 March 2025 16:23 (UTC+05:30)
<b>Cluster configuration</b> Instance groups		<b>Primary node public DNS</b> <a href="#">ec2-3-235-148-116.compute-1.amazonaws.com</a> <a href="#">Connect to the Primary node using SSH</a> <a href="#">Connect to the Primary node using SSM</a>	<b>Elapsed time</b> 6 minutes, 1 second
<b>Capacity</b> 1 Primary   1 Core   1 Task			

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences 24°C Sunny 16:29 06-03-2025

#### 4. EMR Instance is running through putty

```

hadoop@ip-172-31-67-39:~
Using username "hadoop".
Authenticating with public key "imported-openssh-key"

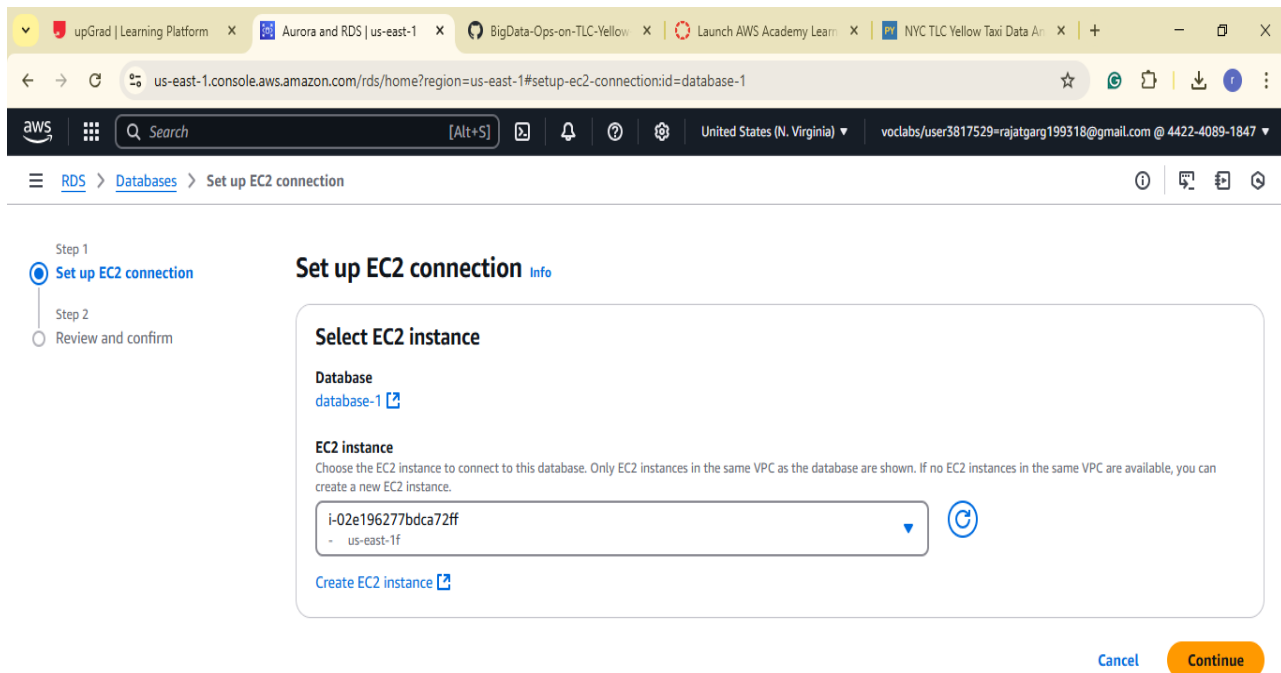
      #_
     _###_   Amazon Linux 2023
    _\#####\
   _\#####|
  _\#####|
 _\#####|
V~'  _-> https://aws.amazon.com/linux/amazon-linux-2023
      /
     /
    /
   /
  /
 /m/'

EEEEEEEEEEEEEEEEEEEE MMMMMMM      MMMMMMM RRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M      M::::::::M R:::::::::R
EE::::::::EEEEEEEE::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::::EEEEEEEE M::::M M::M M::M M::M      R::RRRRR::::R
 E::::::::::::E M::::M M::M::M M::M      R:::::::::RR
 E::::EEEEEEEE M::::M M::M M::M      R::RRRRR::::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

[hadoop@ip-172-31-67-39 ~]$

```

- 5.
6. To connect RDS with EMR instance, we have to click on “Action” button on RDS instance menu and then “Set up an EC2 Instance”.



7.



## 8. Login to RDS through EMR instance using command: `mysql -h`

```
hadoop@ip-172-31-36-93:~$ ssh -i /home/hadoop/.ssh/rsa.pem ec2-user@rds-1.amazonaws.com
Last login: Sun Jul 9 11:51:57 2023

Amazon Linux 2 AMI

https://www.amazon.com/amazon-linux-2/
95 package(s) needed for security, out of 164 available
Run "sudo yum update" to apply all updates.

[ec2-user@ip-172-31-36-93 ~]$ mysql -h rds-1.amazonaws.com -P 3306 -u admin -p
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 480
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.

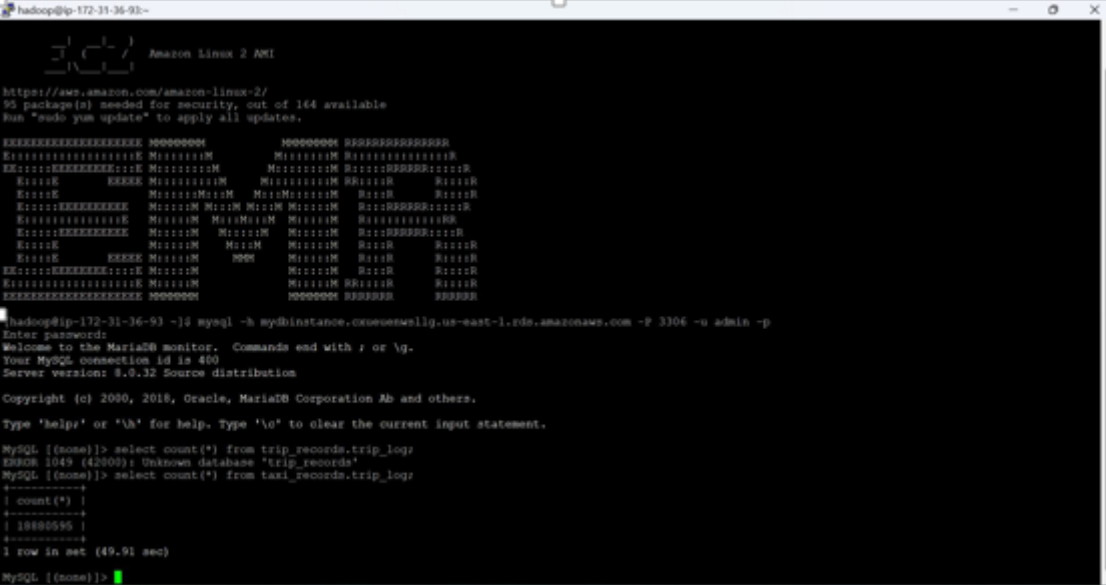
MySQL [(none)]>
```

9. Creation of Database "taxi\_records" and table "trip\_log"
  - `create database taxi\_records`
  - `CREATE TABLE trip\_log`
  - (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 );'**

1. Downloading required csv files from internet in local using command  
` wget [https://nyc-tlc-upgrad.s3.amazonaws.com/yellow\\_tripdata\\_2017-01.csv](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](https://nyc-tlc-upgrad.s3.amazonaws.com/yellow_tripdata_2017-02.csv)`
2. 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;**



```
hadoop@ip-172-31-36-93:~$  
Amazon Linux 2 AMI  
https://aws.amazon.com/amazon-linux-2/  
93 package(s) needed for security, out of 164 available  
Run "sudo yum update" to apply all updates.  
hadoop@ip-172-31-36-93:~$ mysql -h mydbinstance.c9ueuuewllg.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 400  
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)]> select count(*) from trip_records.trip_log;  
ERROR 1049 (42009): Unknown database 'trip_records'  
MySQL [(none)]> select count(*) from taxi_records.trip_log;  
+-----+  
| count(*) |  
+-----+  
| 18880595 |  
+-----+  
1 row in set (49.91 sec)  
MySQL [(none)]>
```

**SELECT COUNT(\*) FROM taxi\_records.trip\_log;**