

## 1. Creation of RDS in AWS

The screenshot shows the AWS Management Console for an Amazon RDS instance. The left sidebar contains navigation links for Amazon RDS, including 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, and Events. The main content area displays the details for the instance 'mysqlassignment1'. The 'Summary' section shows the DB identifier, CPU usage (3.73%), Status (Available), Class (db.t2.micro), Role, Current activity (0 Connections), Engine (MySQL Community), and Region & AZ (us-east-1c). Below the summary, there are tabs for Connectivity & security, Monitoring, Logs & events, Configuration, Maintenance & backups, and Tags. The 'Connectivity & security' tab is active, showing the Endpoint & port, Networking, and Security details.

Summary			
DB identifier	CPU	Status	Class
mysqlassignment1	3.73%	Available	db.t2.micro
Role	Current activity	Engine	Region & AZ
Instance	0 Connections	MySQL Community	us-east-1c

Connectivity & security		
Endpoint & port	Networking	Security
Endpoint	Availability Zone	VPC security groups
mysqlassignment1.cuyf7nwft5s.us-east-1.rds.amazonaws.com	us-east-1c	default (sg-072e0aa4ea7502135)
	VPC	Active

The screenshot shows the AWS Management Console for an Amazon EMR instance. The left sidebar contains navigation links for Amazon EMR, including EMR Studio, EMR Serverless, EMR on EC2, Clusters, Notebooks, Git repositories, Security configurations, Block public access, VPC subnets, Events, EMR on EKS, Virtual clusters, Help, and What's new. The main content area displays the details for the instance 'j-29CDKL0L5T8Y7'. The 'Summary' section shows the ID, Creation date, Elapsed time, After last step completes, Termination protection, Tags, and Master public DNS. Below the summary, there are tabs for Configuration details and Application user interfaces. The 'Configuration details' tab is active, showing the Release label, Hadoop distribution, Applications, Log URI, EMRFS consistent view, Custom AMI ID, and Application user interfaces.

Summary	
ID	j-29CDKL0L5T8Y7
Creation date	2023-01-14 08:25 (UTC+5:30)
Elapsed time	1 minute
After last step completes	Cluster waits
Termination protection	Off
Tags	-- View All / Edit
Master public DNS	ec2-54-158-0-133.compute-1.amazonaws.com
	Connect to the Master Node Using SSH

Configuration details	
Release label	emr-5.30.1
Hadoop distribution	Amazon 2.8.5
Applications	Sqoop 1.4.7
Log URI	s3://aws-logs-357485174750-us-east-1/elasticmapreduce/
EMRFS consistent view	Disabled
Custom AMI ID	--
Application user interfaces	
Persistent user interfaces	--

## 2. Login to RDS through EMR instance

```
mysql -h mysqlassignment1.cuyf7nwft5s.us-east-1.rds.amazonaws.com -P 3306 -u srbhrungta -p
```

$$\frac{1}{2} \left( \frac{1}{2} + \frac{1}{2} \right) = \frac{1}{2}$$

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

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

```
[root@ip-172-31-34-81 ~]# mysql -h mysqlclassmgment1.cuyf7nwfvt5.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](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](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.cuyf7nwfvts.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]>
```