

1. Creation of RDS instance in AWS

Connection setup successfully for RDS database `mapreducedb` and EC2 instance `i-0193b21fb52cb9b54` [View details](#)

Introducing Aurora I/O-Optimized

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

Consider creating a Blue/Green Deployment to minimize downtime during upgrades
You may want to consider using Amazon RDS Blue/Green Deployments and minimize your downtime during upgrades. A Blue/Green Deployment provides a staging environment for changes to production databases. [RDS User Guide](#) [Aurora User Guide](#)

Databases (5) ☒ Group resources [Refresh](#) [Modify](#) [Actions](#) [Restore from S3](#) [Create database](#)

<input type="checkbox"/>	DB identifier	Status	Role	Engine	Region & AZ	Size
<input type="radio"/>	demodb	Available	Multi-AZ DB cluster	MySQL Community	ap-south-1	3 instances
<input type="radio"/>	demodb-instance-1	Available	Writer instance	MySQL Community	ap-south-1c	db.m5d.xlarge
<input type="radio"/>	demodb-instance-2	Available	Reader instance	MySQL Community	ap-south-1a	db.m5d.xlarge
<input type="radio"/>	demodb-instance-3	Available	Reader instance	MySQL Community	ap-south-1b	db.m5d.xlarge
<input checked="" type="radio"/>	mapreducedb	Available	Instance	MySQL Community	ap-south-1c	db.t3.micro

CloudShell Feedback

© 2023, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

Connection setup successfully for RDS database **mapreducedb** and EC2 instance **i-0193b21fb62cb9b54**

View details

RDS > Databases > mapreducedb

mapreducedb

Refresh Modify Actions

Summary

mapreducedb

RoleInstance

Current activity

0 Connections

Engine

MySQL Community

Region & AZ

ap-south-1c

Connectivity & security

Monitoring

Logs & events

Configuration

Maintenance & backups

Tags

Connectivity & security

Endpoint & port

Endpoint
mapreducedb.cjd8k4oc9drm.ap-south-1.rds.amazonaws.com

Port
3306

Networking

Availability Zone
ap-south-1c

VPC
vpc-080178f6e1e56830d

Subnet group
default-vpc-080178f6e1e56830d

Subnets
subnet-09449b34cece6ed62
subnet-02425aeaddca7a7817

Network type
IPv4

Security

VPC security groups
rds-ec2-2 (sg-0a1fafc785ce3e275)
Active
default (sg-037d0fa5b4d120713)
Active

Publicly accessible
Yes

CloudShell

Feedback

© 2023, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

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

Connected compute resources (1)

Info

Refresh

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

1

Resource identifier

Resource type

Availability Zone

VPC security group

Compute resource security group

i-0193b21fb62cb9b54

EC2 instance

ap-south-1b

rds-ec2-2

ec2-rds-2

Proxies (0)

Refresh

Actions

Create proxy

Filter by proxies

1

Proxy identifier

Status

Engine family

No proxies

You don't have any proxies.

Create proxy

Security group rules (3)

Refresh

Filter by Security group rules

1

Security group

Type

Rule

default (sg-037d0fa5b4d120713)

EC2 Security Group - Inbound

sg-037d0fa5b4d120713

default (sg-037d0fa5b4d120713)

CIDR/IP - Outbound

0.0.0.0/0

rds-ec2-2 (sg-0a1fafc785ce3e275)

EC2 Security Group - Inbound

sg-0c97b608739f244e3

Replication (1)

Refresh

Filter by Replication

1

DB identifier

Role

Region & AZ

Replication source

Replication state

Lag

mapreducedb

Instance

ap-south-1c

-

-

-

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

The screenshot displays the AWS Management Console interface for an Amazon EMR cluster. At the top, a green notification bar states: "Your cluster 'MapReduce Assignment' has been successfully created." The breadcrumb navigation shows: Amazon EMR > EMR on EC2: Clusters > MapReduce Assignment. The cluster title "MapReduce Assignment" is prominently displayed, along with the update time "Updated less than a minute ago" and action buttons: "Terminate", "Clone in AWS CLI", and "Clone".

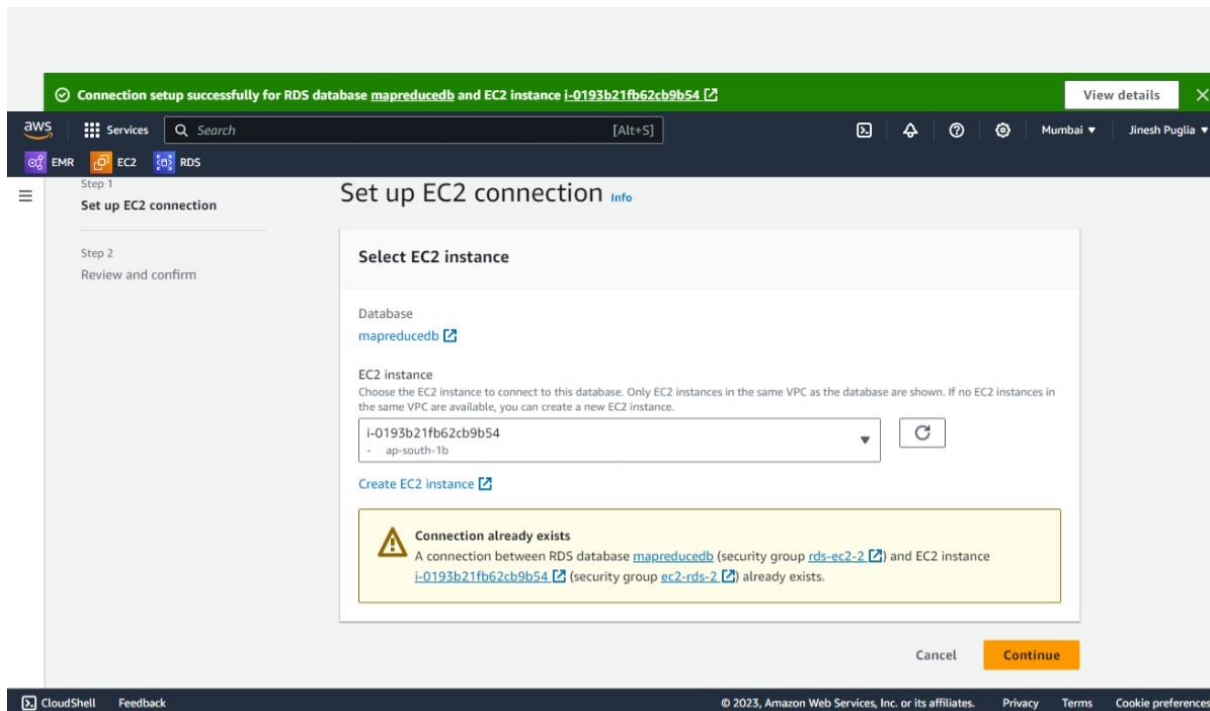
The "Summary" section provides a detailed overview of the cluster:

- Cluster info:** Cluster ID is j-3OUTI0UK9FHHJ. It lists configuration details like instance groups and capacity (1 Primary, 1 Core, 0 Task).
- Applications:** Specifies the Amazon EMR version as emr-5.30.1 and lists installed applications: HBase 1.4.13, Hadoop 2.8.5, Hive 2.3.6, Hue 4.6.0, Spark 2.4.5, and Sqoop 1.4.7.
- Cluster management:** Includes log destination in Amazon S3 (aws-logs-7843587365-ap-south-1/elasticmapreduce), persistent application URLs for Spark History Server, YARN timeline server, and Tez UI, and the primary node public DNS (ec2-65-2-122-142.ap-south-1.compute.amazonaws.com).
- Status and time:** The cluster status is "Waiting". It shows a creation time of November 06, 2023, 10:24 (UTC+05:30) and an elapsed time of 1 hour, 8 minutes.

Below the summary, a horizontal menu allows switching between various tabs: Properties, Bootstrap actions, Instances (Hardware), Steps, Applications, Configurations, Monitoring, Events, and Tags (1). The "Properties" tab is currently selected, showing:

- Cluster logs:** Information about archiving log files to Amazon S3 (turned on) and the Amazon S3 location (s3://aws-logs-7843587365-ap-south-1/elasticmapreduce/).
- Cluster termination:** Details on the termination option (Manually terminate cluster), termination protection (turned off), and idle time.
- Network and security:** A sub-section containing:
 - Network:** Virtual Private Cloud (VPC) vpc-080178f6e1e56830d and Subnet(s) and Availability Zone(s) (AZ) subnet-02425aeada7a7817 in ap-south-1b.
 - Security configuration:** Security configuration set to None and EC2 key pair set to RHEL.
 - Permissions:** Service role for Amazon EMR (AmazonEMR-ServiceRole-20231010T104948), EC2 instance profile (AmazonEMR-InstanceProfile-20231010T104930), and custom automatic scaling role (Not configured).

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``

```
hadoop@ip-172-31-14-82:~  
[hadoop@ip-172-31-14-82 ~]$ mysql -h mapreducedb.cjd8k4oc9drm.ap-south-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.  
  
MySQL [(none)]> show databases  
-> ;  
+-----+  
| Database |  
+-----+  
| information_schema |  
| mysql |  
| performance_schema |  
| sys |  
+-----+  
4 rows in set (0.01 sec)  
  
MySQL [(none)]>
```

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  
);
```

```
hadoop@ip-172-31-14-82:~  
+-----+  
| information_schema |  
| mysql              |  
| performance_schema |  
| sys                |  
+-----+  
4 rows in set (0.01 sec)  
  
MySQL [(none)]> create database taxi_records  
-> ;  
Query OK, 1 row affected (0.01 sec)  
  
MySQL [(none)]> use taxi_records  
Database changed  
MySQL [taxi_records]>  
MySQL [taxi_records]> 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,
```

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;
```

```
hadoop@ip-172-31-14-82:~  
[hadoop@ip-172-31-14-82 ~]$ mysql -h mapreducedb.cjd8k4oc9drm.ap-south-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 27  
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 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
```

```
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]>
```