This documentation contains the steps to set up an RDS instance
The steps covered in this document are:
1. Connecting to the RDS instance:
    - You will use an EMR cluster to connect to the RDS instance
    - You can also do this from your local machine if you have SQL installed on it or use MySQLWorkbench and connect to the RDS instance.
2. Creating tables on the RDS instance
3. Loading these tables with data present in the file

Creation of the RDS instance

This document assumes that you've already created an RDS instance on AWS. All the commands in this documentation have been executed using the following RDS specifications:
Database Name: demoDB
User: admin
Password: user123

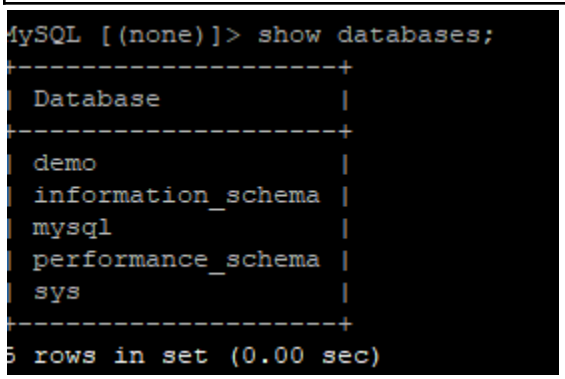
Connecting with the RDS:

```
mysql -h demodb.cqsesz6h9yjg.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
```

After entering this command, you'll need to enter the password.
You may not get connected to the RDS instance, because the RDS instance may not have privileges to connect to this cluster. To enable this you need to edit the security groups by adding a new rule which enables an SQL connection to the EMR ip address. For eg: 172.31.93.189/32
Alternatively, you'll be able to set up a connection between the RDS instance and the cluster/instance. The steps for the same are at the end of this documentation.

Creating a database of the name demo

```
show databases;
create database demo;
use demo;
```



Create a table

```
create table users
(
    user_id VARCHAR(255),
    age INT,
    gender VARCHAR(255),
    occupation VARCHAR(255),
    zip_code INT
);
```

```
show tables;
```

```
MySQL [demo]> show tables;
+----------------+
| Tables_in_demo |
+----------------+
| users          |
+----------------+
1 row in set (0.00 sec)
```

Loading the data into these tables

a. Downloading necessary data on your local file system. We'll be using the MovieLens Dataset from the following link and transfer using WinSCP
https://grouplens.org/datasets/movielens/100k/

Depending on the data source, you can also use the following methods to download the data to your instance

Mac/Linux	scp -i C:\User\Downloads\XXXXX.pem ~/Downloads/crm1.csv hadoop@ec2-34-203-220-27.compute-1.amazonaws.com:/home/hadoop
Windows	wget https://files.grouplens.org/datasets/movielens/ml-100k.zip

And use the unzip command to extract the files

b. Loading this data onto the tables SQL tables using our EMR instance
Connect to your RDS instance

```
mysql -h demodb.cqsesz6h9yjg.us-east-1.rds.amazonaws.com -P 3306 -u admin -p
```

Go to your database and load the tables that you have created

```
LOAD DATA LOCAL INFILE '/home/hadoop/ml-100k/u.user'
INTO TABLE users
FIELDS TERMINATED BY ','
```

```
LINES TERMINATED BY '\n'
IGNORE 1 LINES;
```

```
MySQL [demo]> LOAD DATA LOCAL INFILE '/home/hadoop/ml-100k/u.user'
    -> INTO TABLE users
    -> FIELDS TERMINATED BY ','
    -> LINES TERMINATED BY '\n'
    -> IGNORE 1 LINES;
Query OK, 942 rows affected, 3768 warnings (0.06 sec)
Records: 942  Deleted: 0  Skipped: 0  Warnings: 3768
```

It's always good to try a few commands to make sure that your tables have indeed been loaded

```
select * from users limit 5;
```

```
MySQL [demo]> select * from users limit 5;
+-------------------------+------+--------+------------+----------+
| user_id                 | age  | gender | occupation | zip_code |
+-------------------------+------+--------+------------+----------+
| 2|53|F|other|94043      | NULL | NULL   | NULL       |     NULL |
| 3|23|M|writer|32067     | NULL | NULL   | NULL       |     NULL |
| 4|24|M|technician|43537 | NULL | NULL   | NULL       |     NULL |
| 5|33|F|other|15213      | NULL | NULL   | NULL       |     NULL |
| 6|42|M|executive|98101  | NULL | NULL   | NULL       |     NULL |
+-------------------------+------+--------+------------+----------+
5 rows in set (0.00 sec)
```

```
select COUNT(*) from users;
```

```
MySQL [demo]> select COUNT(*) from users;
+----------+
| COUNT(*) |
+----------+
|      942 |
+----------+
1 row in set (0.03 sec)
```

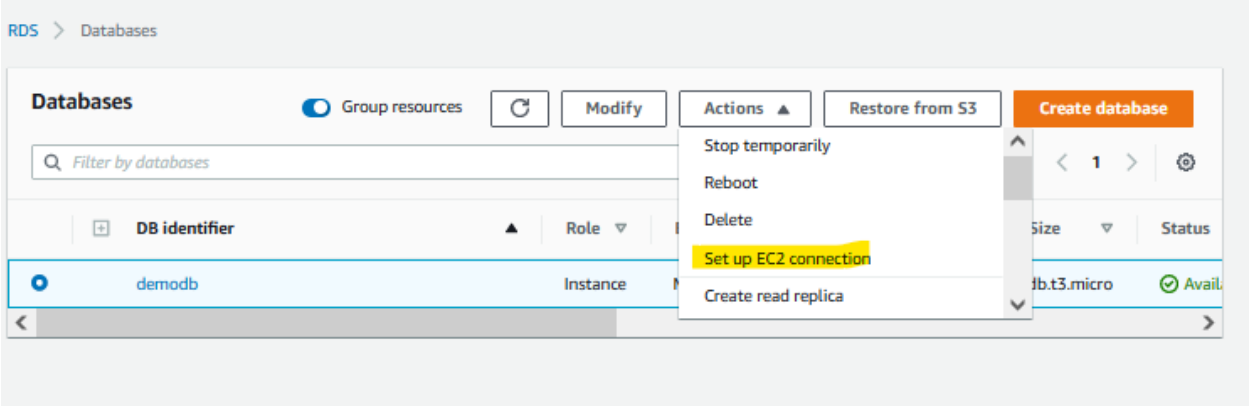
Validate this count with the original u.users file in your instance.

```
wc device1.csv
```

With the RDS now loaded with the data, you can now use Sqoop commands to ingest the data from RDS.

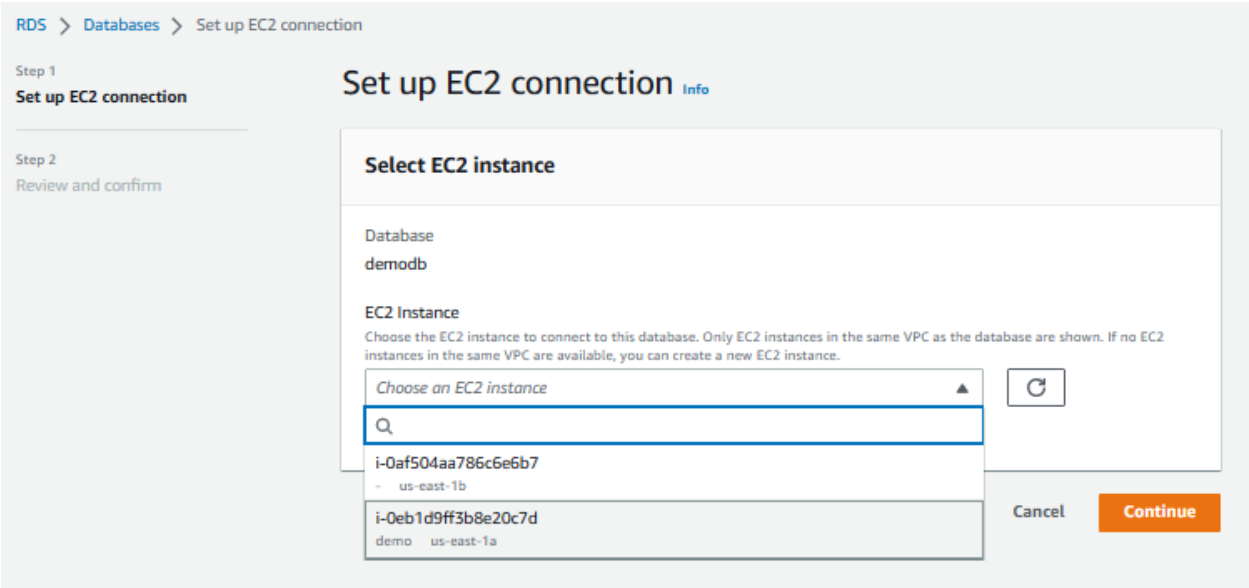
Steps to set up a connection between RDS instance and EC2

Go to your RDS instance and click on **Actions** and “Set up EC2 connection’



Click the running EC2 instance from the dropdown.

If you’re using an EMR cluster, use the instance corresponding to the master node.



Click on ‘Confirm and set up’ button after reviewing the parameters

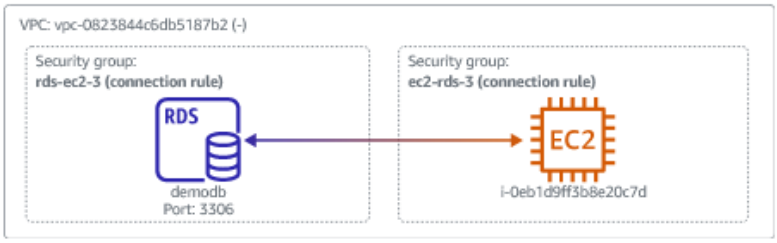
Set up EC2 connection

Step 2
Review and confirm

Connection summary Info

You are setting up a connection bewteen RDS database demodb and EC2 instance i-0eb1d9ff3b8e20c7d.

To set up a connection between the database and the EC2 instance, VPC security group is added to the database, and VPC security group *ec2-rds-3* is added to the EC2 instance.



Bold indicates an addition being made to set up a connection.

Changes to EC2 instance: i-0eb1d9ff3b8e20c7d

Attribute	Current value	New value
Security group	launch-wizard-18	launch-wizard-18, ec2-rds-3

Cross Availability Zone (AZ) charges might apply
The RDS database demodb (us-east-1b) and EC2 instance i-0eb1d9ff3b8e20c7d (us-east-1a) are in different AZs. Cross AZ charges might apply. Data transfer within same region

Cancel | Previous | Confirm and set up