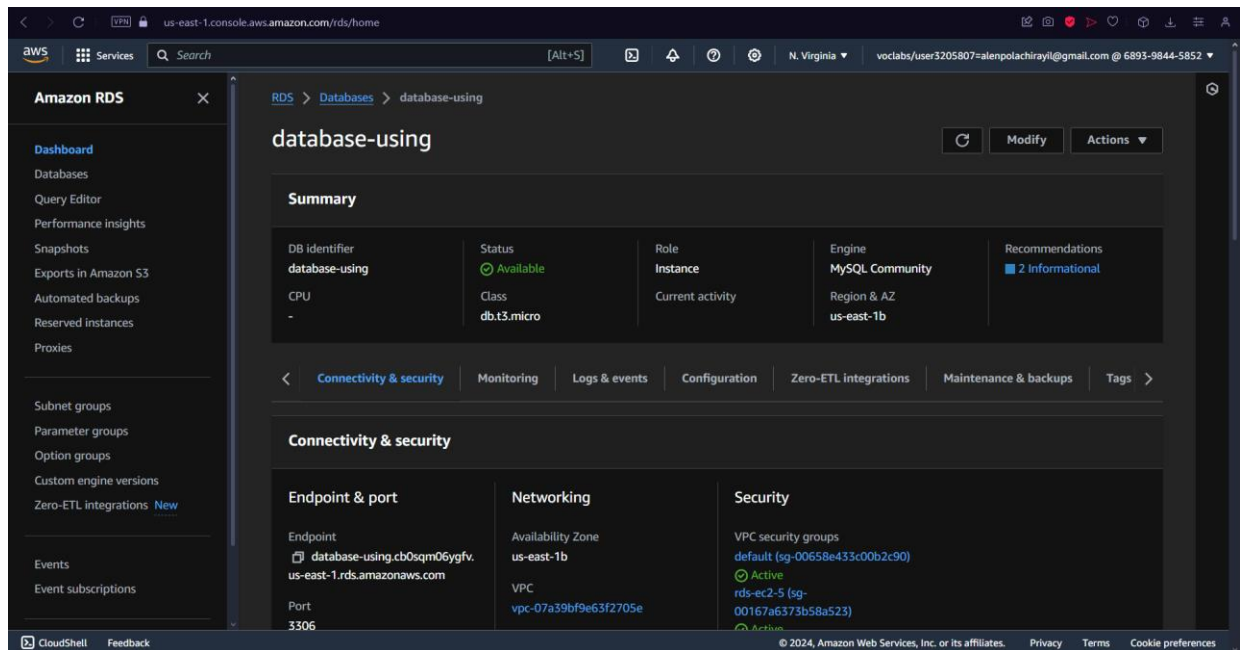
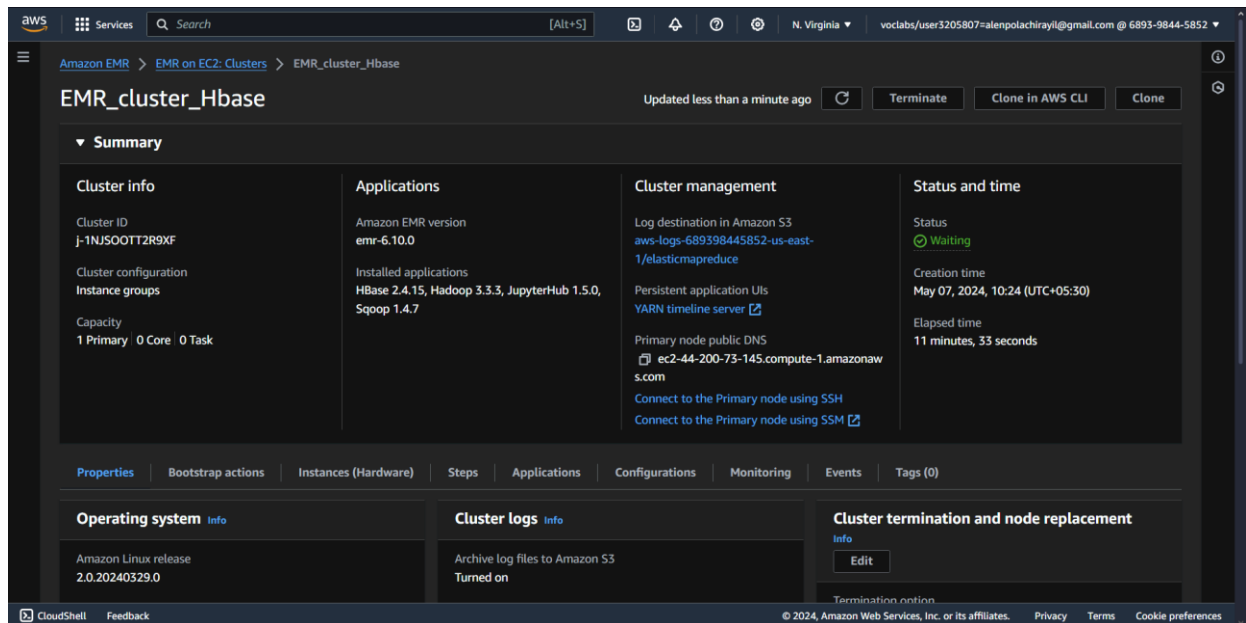


1. Creation of RDS in AWS



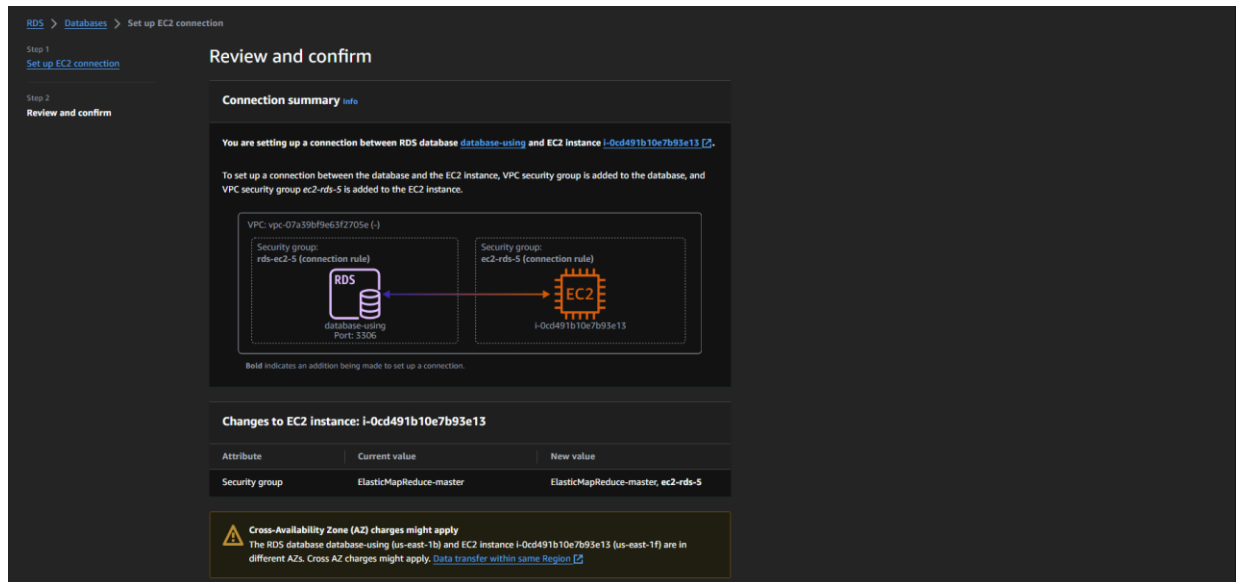
Created an RDS with necessary options selected like Engine is MySQL Community and allowed our IP to access the RDS in the security groups. Name of RDS is database-using

2. Creating EMR cluster with applications HBase, Hadoop, Sqoop

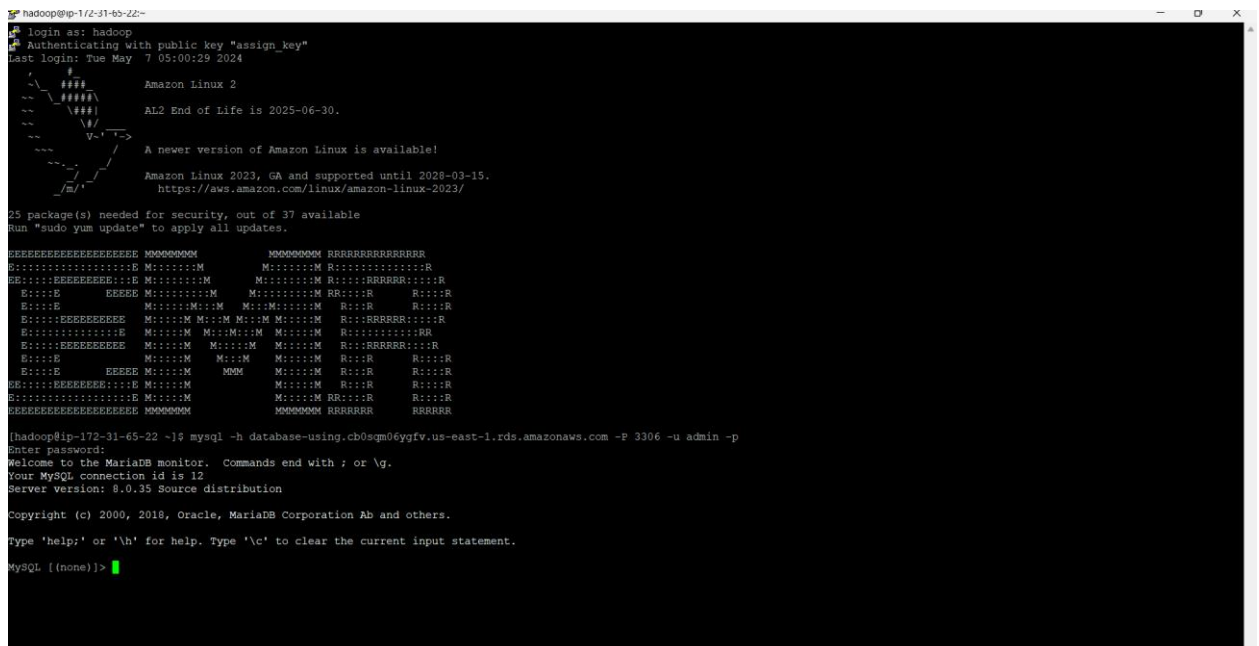


EMR is given access to our IP in security groups to access it. Name of EMR is EMR_cluster_Hbase

3. Setting connection between RDS and EMR instance we choose the action “set up an EC2 instance”



4. Login to RDS using the code “mysql -h database-using.cb0sqm06ygvf.us-east-1.rds.amazonaws.com -P 3306 -u admin -p”



5. Code for Creation of Database

```
'create database taxi;'
```

6. Code for creating table in Database "taxi"

```
create table yellow_taxi ( VendorID VARCHAR(255), tpep_pickup_datetime TIMESTAMP NOT  
NULL DEFAULT '0000-00-00 00:00:00', tpep_dropoff_datetime TIMESTAMP NOT NULL  
DEFAULT '0000-00-00 00:00:00',passenger_count INT,trip_distance DOUBLE,RatecodeID  
VARCHAR(255),store_and_fwd_flag VARCHAR(255),PULocationID  
VARCHAR(255),DOLocationID VARCHAR(255),payment_type VARCHAR(255),fare_amount  
DOUBLE,extra DOUBLE,mta_tax DOUBLE,trip_amount DOUBLE,tolls_amount DOUBLE,  
improvement_surcharge DOUBLE,total_amount DOUBLE,congestion_surcharge  
DOUBLE,airport_fee DOUBLE );
```

- Table name is yellow_taxi

7. Downloading required csv files in local

```
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 to mysql table

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

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

9. `Select count(*) from yellow_taxi;`

```

root@hadoopip-172-31-65-22 ~#
┌─┴─┐
└───┘
      Amazon Linux 2023, GA and supported until 2028-03-15.
      https://aws.amazon.com/linux/amazon-linux-2023/

25 package(s) needed for security, out of 37 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEE MMMMMMM          MMMMMMMM RRRRRRRRRRRR
E::::::::::::::::::E M::::::::M       M::::::::M R::::::::::::R
EE::::::::EEEEEEEE::E M::::::::M     M::::::::M R::::RRRRRR::R
E::::E   EEEE M::::::::M    M::::::::M RR::::R   R::::R
E::::E   M::::::::M    M:M:M:M:M R::R   R::::R
E::::EEEEEEEEEE M::::::::M M::M M::M M::M R::RRRRR::R
E::::::::::::::E M::M:M M::M:M M::M R:::RRRR::R
E::::EEEEEEEEEE M::M:M M::M:M M::M R::RRRRR::R
E::::E   M::M:M M::M:M M::M:M R::R   R::::R
E::::E   EEEE M::M:M   MM M::M:M R::R   R::::R
EE::::EEEEEEEE::E M::M:M           M::M:M R::R   R::::R
E::::::::::::::E M::M:M           M::M:M RR::::R   R::::R
EEEEEEEEEEEEEEEEEE MMMMMMM          MMMMMMMM RRRRRR   RRRRR

[hadoop@ip-172-31-65-22 ~]$ mysql -h database-using.cb0sqm06ygvf.us-east-1.rds.a
amazonaws.com -P 3306 -u admin -p
Enter password:
Welcome to the MariaDB monitor. Commands end with ; or \g.
Your MySQL connection id is 14
Server version: 8.0.35 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;
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]> 9.Select count(*) from yellow_taxi;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that
corresponds to your MySQL server version for the right syntax to use near '9.Sel
ect count(*) from yellow_taxi' at line 1
MySQL [taxi]> select count(*) from yellow_taxi;
+-----+
| count(*) |
+-----+
| 18860595 |
+-----+
1 row in set (53.45 sec)

MySQL [taxi]> 

```

10. `Select * from yellow_taxi limit 5;`

```
hadoop@ip-172-31-65-22~$ zshazawes.com -P 3306 --u admin -p  
Enter password:  
Welcome to the MariaDB monitor. Commands end with ; or \g.  
Your MySQL connection id is 14  
Server version: 8.0.35 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;  
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]> select count(*) from yellow_taxi;
```

ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near '9.sel ect count(*) from yellow_taxi' at line 1

```
MySQL [taxi]> Select count(*) from yellow_taxi;
```

	count (*)
	18880595

1 row in set (53.45 sec)

```
MySQL [taxi]> Select * from yellow_taxi limit 5;
```

ID	VendorID	Trip pickup datetime	Trip dropoff datetime	passenger count	Trip distance	RateCodeID	store_and_fwd_flag	PULocationID	DOLocationID	payment_type	Fare amount	extra	mta_tax	Trip amount	tolls amount	improvement surcharge	Total amount	congestion surcharge	Airport fee
1	1	2017-01-01 00:32:05	2017-01-01 00:37:48	1	7.8	1	N	140	236	2	6.5	0.5	0	0	0	0	6.5	0	0
2	1	2017-01-01 02:12:23	2017-01-01 02:27:09	1	6.95	1	N	74	232	2	21	0.5	0	0	0	0	21	0	0
3	1	2017-01-01 00:43:25	2017-01-01 00:47:42	2	0.7	1	N	237	140	2	5	0.5	0	0	0	0	5	0	0
4	2	2017-01-01 02:31:18	2017-01-01 02:44:34	1	6.3	1	N	79	231	2	11	0.5	0	0	0	0	11	0	0
5	1	2017-01-01 00:49:10	2017-01-01 00:53:53	2	0.8	1	N	140	237	2	5.5	0.5	0	0	0	0	5.5	0	0

5 rows in set (0.03 sec)

```
MySQL [taxi]>
```