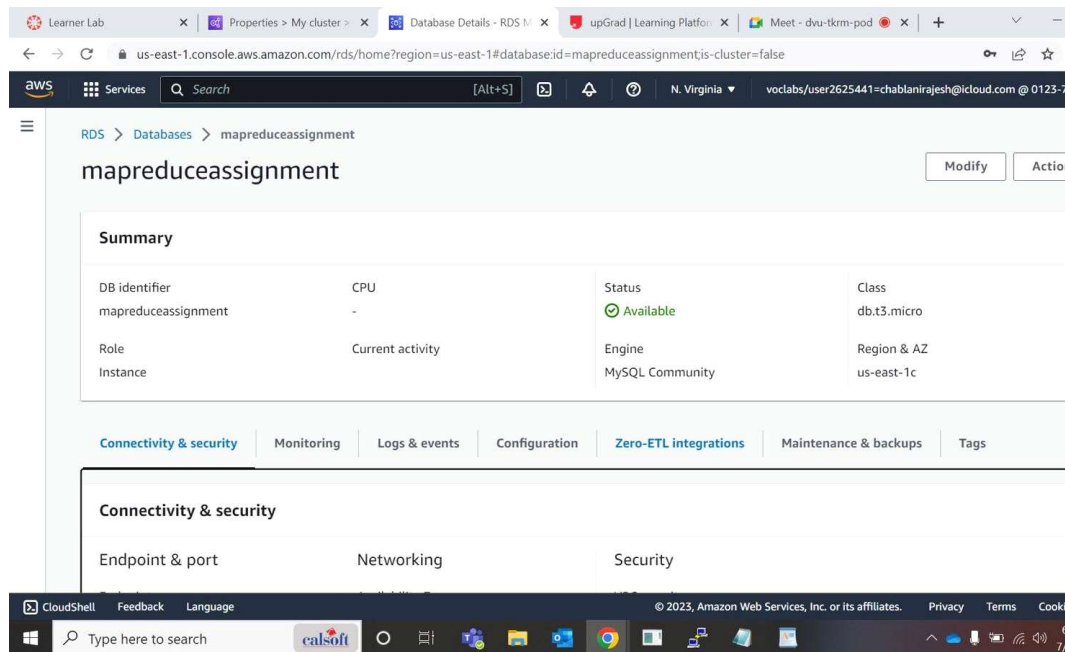
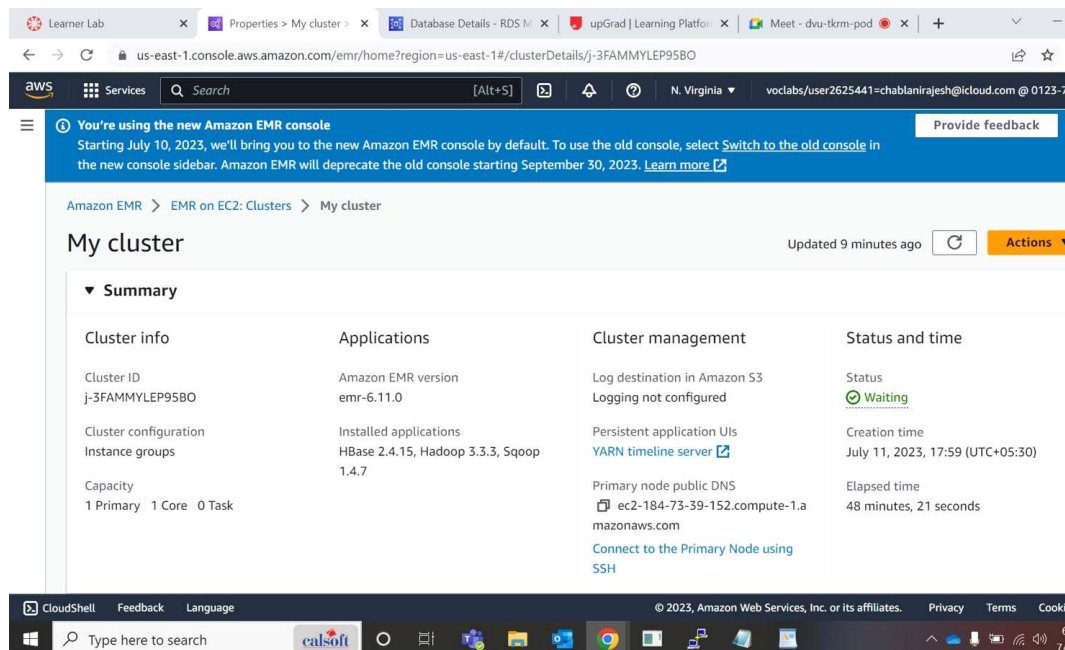


- Creation of RDS instance in AWS



- Creation of EMR instance with bundled applications such as Hadoop, Hbase, Sqoop.



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

- Login to RDS through EMR instance using command: ``mysql -h mapreduceassignment.cnx2n9cjsi3r.us-east-1.rds.amazonaws.com -P 3306 -u admin -p``

- Creation of Database “mapreduceDB” and table “task1”

```
`create database mapreduceDB`
```

```
`CREATE TABLE task1
```

```
(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 VARCHAR(50), Airport_fee VARCHAR(50) );`
```

- 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`
```

- 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 task1 FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
```

```
LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-02.csv' INTO TABLE task1 FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;
```

```
SELECT COUNT(*) FROM mapreduceDB.task1;
```

```
ec2-user@ip-172-31-92-13:~
-> tolls amount FLOAT,
-> improvement_surcharge FLOAT,
-> total amount FLOAT,
-> congestion_surcharge VARCHAR(50),
-> Airport_fee VARCHAR(50)
-> );
Query OK, 0 rows affected (0.05 sec)

MySQL [mapreduceDB]> LOAD DATA LOCAL INFILE '/home/hadoop/yellow_tripdata_2017-01.csv'
-> INTO TABLE task1
-> FIELDS TERMINATED BY ','
-> LINES TERMINATED BY '\n'
-> IGNORE 1 LINES;
ERROR 2 (HY000): File '/home/hadoop/yellow_tripdata_2017-01.csv' not found (Errcode: 2)
MySQL [mapreduceDB]>
MySQL [mapreduceDB]> LOAD DATA LOCAL INFILE '/home/ec2-user/yellow_tripdata_2017-01.csv'
-> INTO TABLE task1
-> FIELDS TERMINATED BY ','
-> LINES TERMINATED BY '\n'
-> IGNORE 1 LINES;
Query OK, 9710820 rows affected (2 min 1.60 sec)
Records: 9710820 Deleted: 0 Skipped: 0 Warnings: 0

MySQL [mapreduceDB]> LOAD DATA LOCAL INFILE '/home/ec2-user/yellow_tripdata_2017-02.csv'
-> INTO TABLE task1
-> FIELDS TERMINATED BY ','
-> LINES TERMINATED BY '\n'
-> IGNORE 1 LINES;
Query OK, 9169775 rows affected (2 min 19.49 sec)
Records: 9169775 Deleted: 0 Skipped: 0 Warnings: 0

MySQL [mapreduceDB]> SELECT COUNT(*) FROM TASK1
-> ;
ERROR 1146 (42S02): Table 'mapreduceDB.TASK1' doesn't exist
MySQL [mapreduceDB]> SELECT COUNT(*) FROM task1
-> ;
+-----+
| COUNT(*) |
+-----+
| 18880595 |
+-----+
1 row in set (57.74 sec)

MySQL [mapreduceDB]>
```