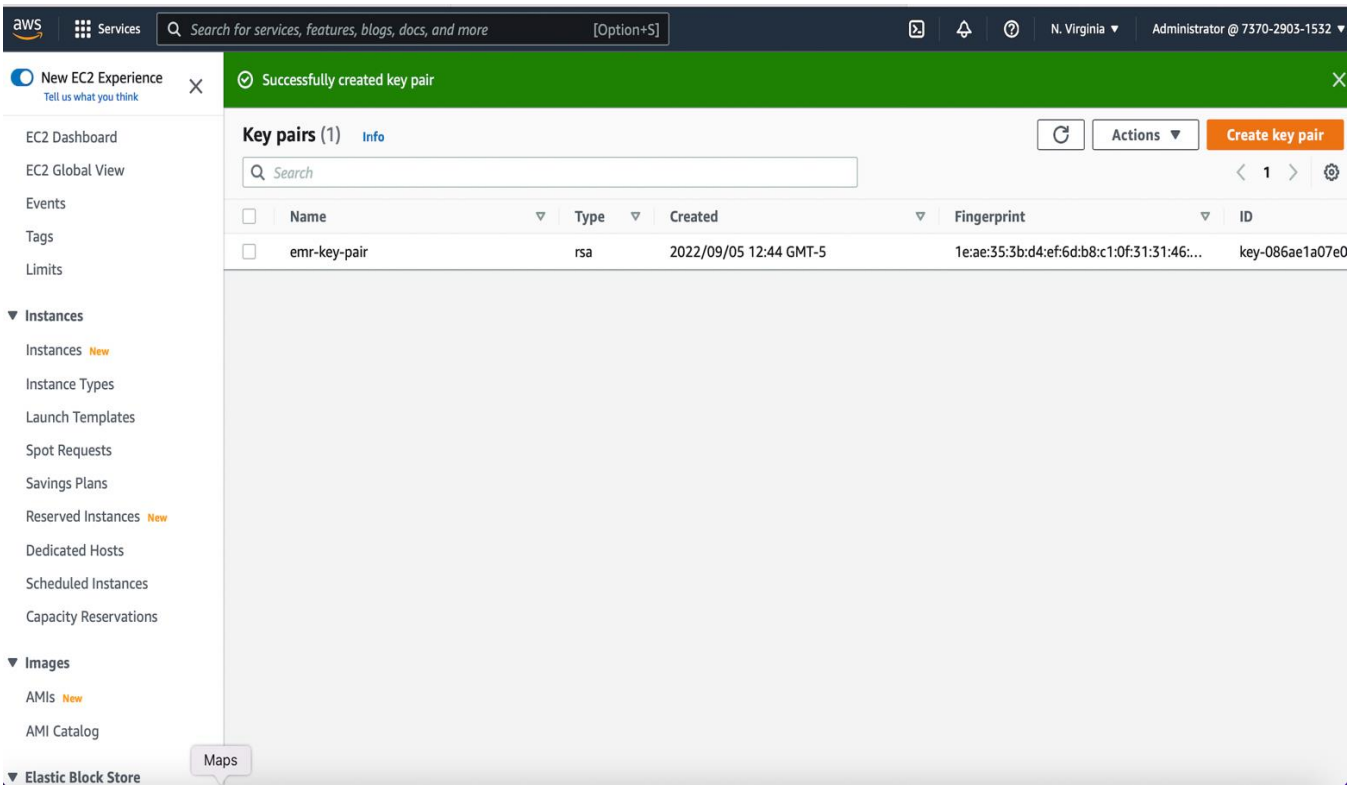
CSP 554 BIG DATA TECHNOLOGIES

MOHAMMED WASIM R D(A20497053)

ASSIGNMENT 2

Step 1 : Create an Amazon EC2 Key Pair

Key name : emr-key-pair

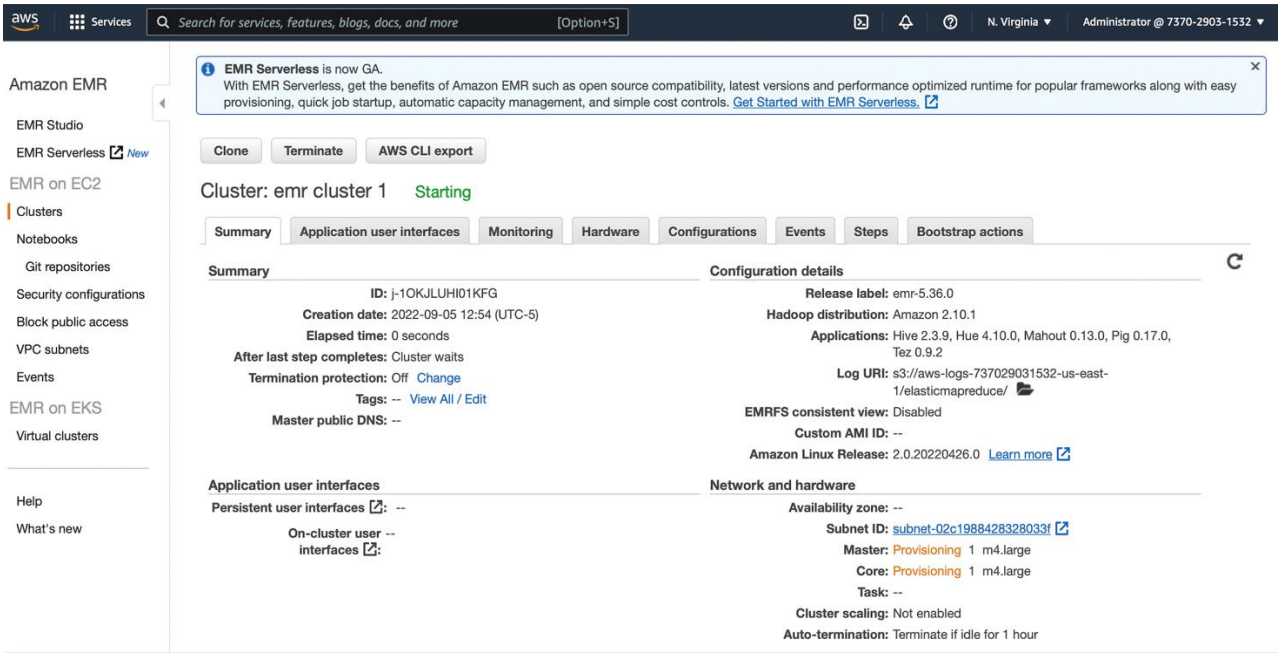


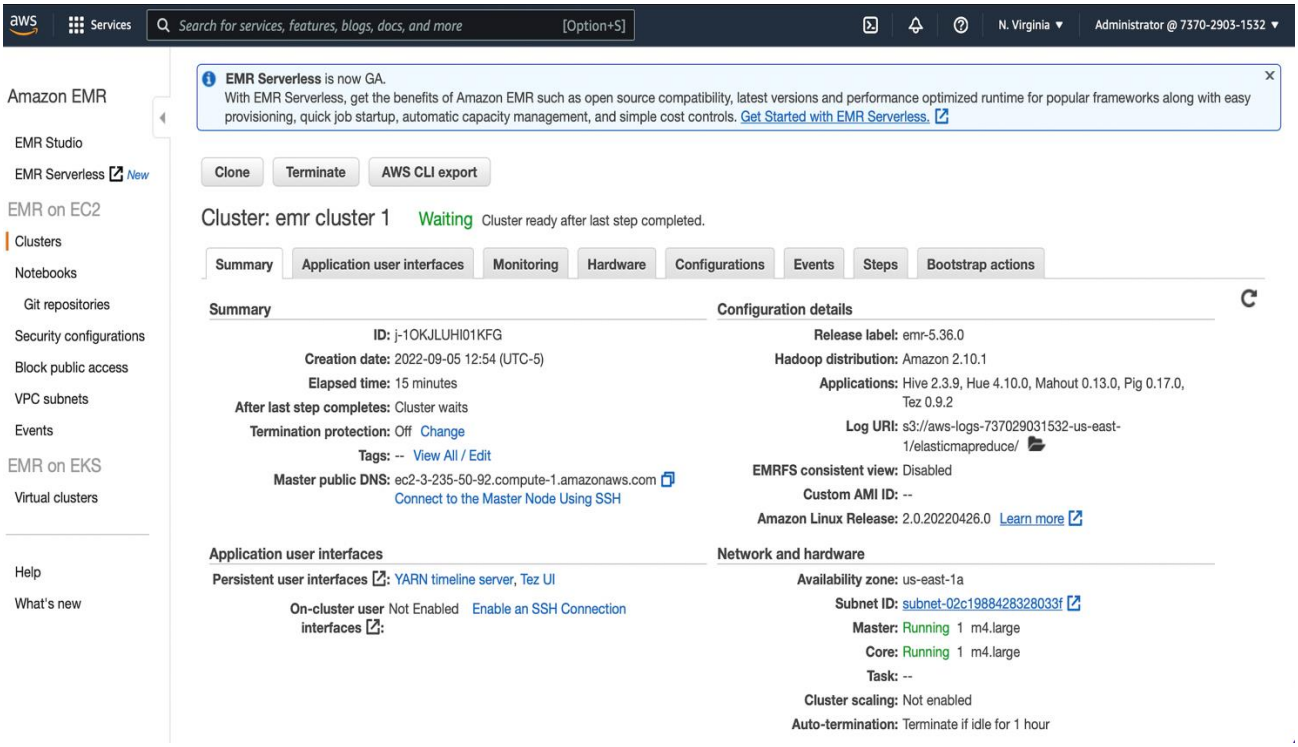
Step 2 : Using the "terminal" program on the MAC execute the following command to set the permissions of your private key file so that only you can read it. Note, use the appropriate path and file name for your situation.

chmod 400 <path-to-file>/*emr-key-pair*.pem

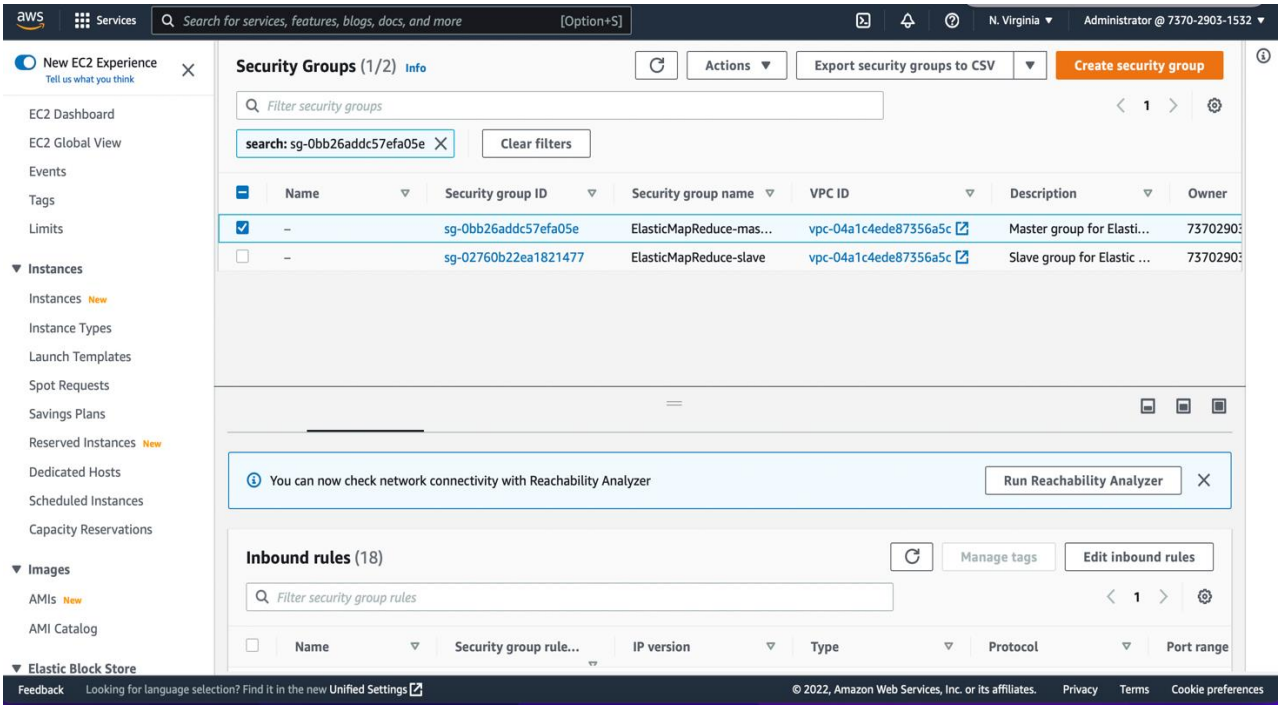
```
(base) mohammedwasimrd@dhcp3 ~ % chmod 400 </Users/mohammedwasimrd/Documents/CSP554>/emr-key-pair.pem
zsh: read-only file system: /emr-key-pair.pem
```

Step 3: Create EMR Cluster

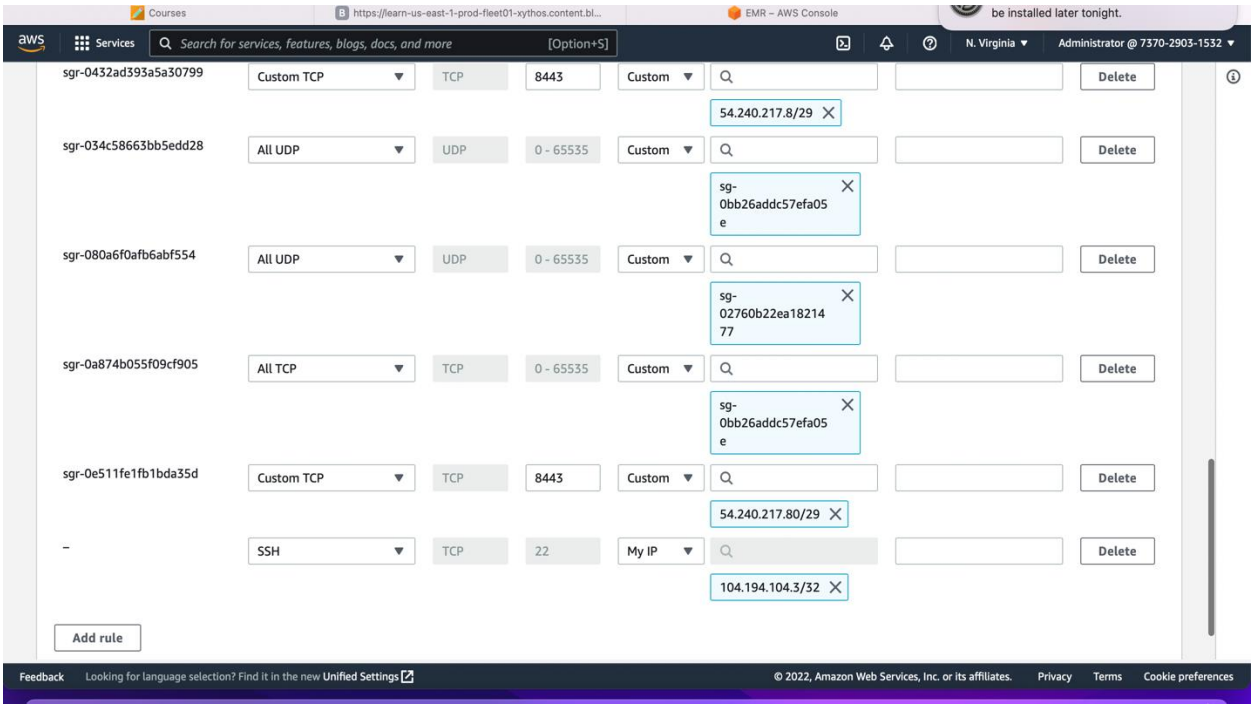




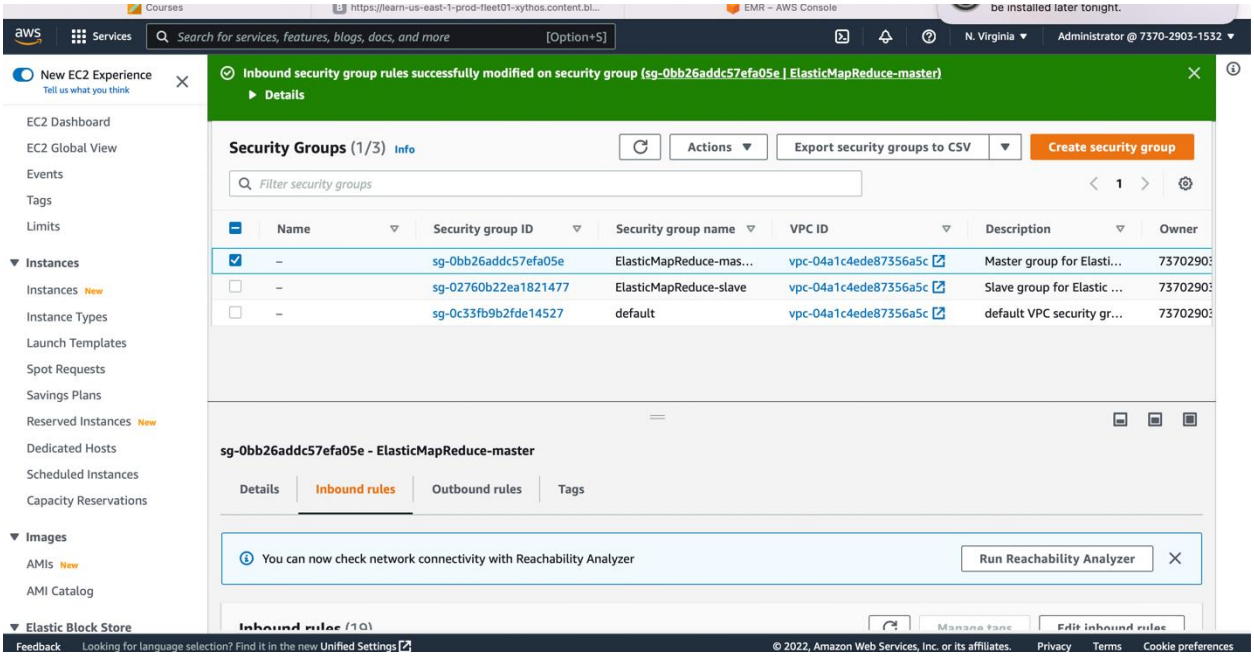
Step 4: Choose ElasticMapReduce-master from the list. Select the ElasticMapReduce-master by clicking on its row.



Step 5: Adding Inbound Rules



Step 6: Retrieve the Public DNS Name of the Master Node



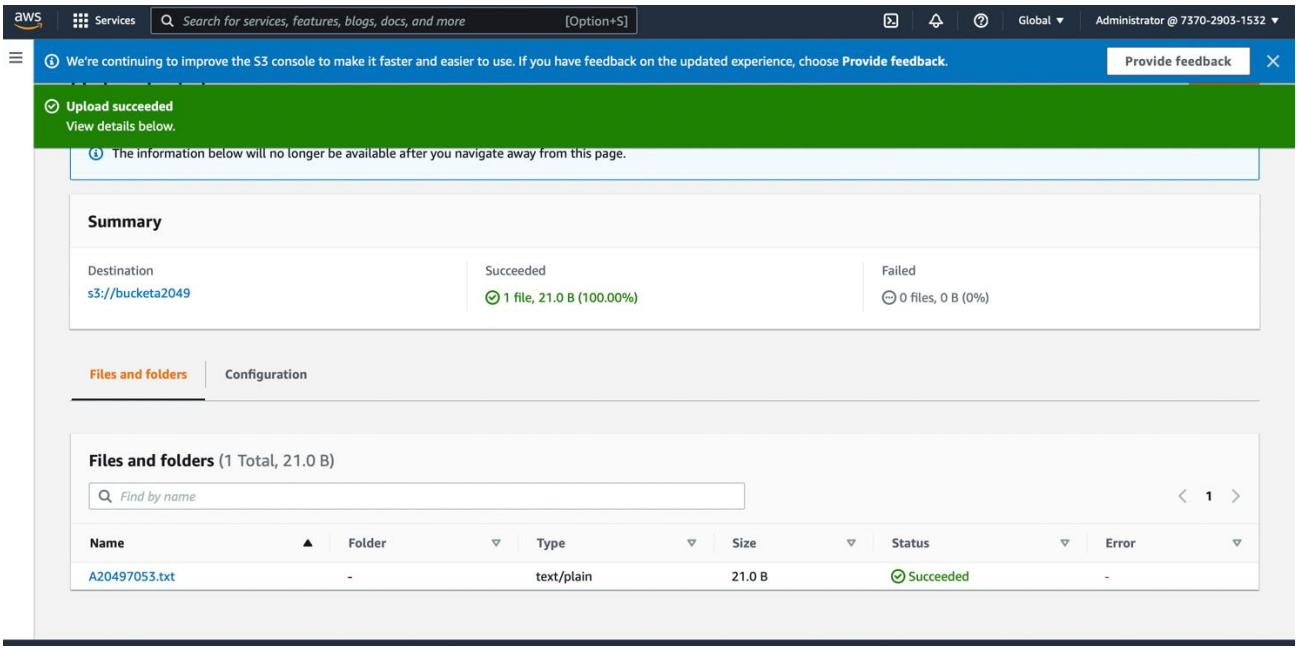
Step 7 : To connect to the Master Node Using SSH and an Amazon EC2 Private Key

```
Last login: Mon Sep  5 16:11:56 on ttys000
(base) mohammedwasimrd@dhcp3 ~ % chmod 400 </Users/mohammedwasimrd/Documents/CSP554>/emr-key-pair.pem
zsh: read-only file system: /emr-key-pair.pem
(base) mohammedwasimrd@dhcp3 ~ % ssh -i /Users/mohammedwasimrd/Documents/CSP554/to/emr-key-pair.pem hadoop@ec2-52-55-13-12.compute-1.amazonaws.com
Warning: Identity file /Users/mohammedwasimrd/Documents/CSP554/to/emr-key-pair.pem not accessible: No such file or directory.
hadoop@ec2-52-55-13-12.compute-1.amazonaws.com: Permission denied (publickey,gssapi-keyex,gssapi-with-mic).
(base) mohammedwasimrd@dhcp3 ~ % chmod 400 /Users/mohammedwasimrd/Documents/CSP554/emr-key-pair.pem
(base) mohammedwasimrd@dhcp3 ~ % ssh -i /Users/mohammedwasimrd/Documents/CSP554/to/emr-key-pair.pem hadoop@ec2-52-55-13-12.compute-1.amazonaws.com
Warning: Identity file /Users/mohammedwasimrd/Documents/CSP554/to/emr-key-pair.pem not accessible: No such file or directory.
hadoop@ec2-52-55-13-12.compute-1.amazonaws.com: Permission denied (publickey,gssapi-keyex,gssapi-with-mic).
(base) mohammedwasimrd@dhcp3 ~ %
(base) mohammedwasimrd@dhcp3 ~ %
(base) mohammedwasimrd@dhcp3 ~ %
(base) mohammedwasimrd@dhcp3 ~ % ssh -i /Users/mohammedwasimrd/Documents/CSP554/(base) mohammedwasimrd@dhcp3 ~ % ssh -i /Users/mohammedwasimrd/Documents/CSP554/(base) mohammedwasimrd@dhcp3 ~ % ssh -i /Us
rs/mohammedwasimrd/Documents/CSP554/emr-key-pair.pem hadoop@ec2-52-55-13-12.compute-1.amazonaws.com
Last login: Mon Sep  5 21:21:10 2022

       __|  __|_  )
       _|  (     /   Amazon Linux 2 AMI
      ___|\___|___|

https://aws.amazon.com/amazon-linux-2/
17 package(s) needed for security, out of 37 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEE MMMMMMM           MMMMMMM RRRRRRRRRRRRRR
E::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE:::::EEEEEEEEE:::E M::::::::M       M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M     M:::::::::M RR::::R      R::::R
  E::::E             M::::::M:::M   M:::M::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE   M:::::M M:::M M:::M M:::::M   R:::RRRRRR:::::R
  E::::::::::::::E   M:::::M  M:::M:::M  M:::::M   R::::::::::::RR
  E:::::EEEEEEEEEE   M:::::M   M:::::M   M:::::M   R:::RRRRRR::::R
  E::::E             M:::::M    M:::M    M:::::M   R:::R      R::::R
  E::::E       EEEEE M:::::M     MMM     M:::::M   R:::R      R::::R
EE:::::EEEEEEEE:::E M:::::M             M:::::M   R:::R      R::::R
E::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEE MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-88-199 ~]$
```

Step 8 : Connect to local window with other terminal

```
Last login: Mon Sep  5 16:13:56 on ttys000
(base) mohammedwasimrd@dhcp3 ~ % scp -i /Users/mohammedwasimrd/Documents/CSP554/emr-key-pair.pem /Users/mohammedwasimrd/Documents/CSP554/mohammedwasim.txt hadoop@ec2-52-55-13-12.compute-1.amazonaws.com:/h
ome/hadoop
mohammedwasim.txt                             100%   23     0.9KB/s   00:00
(base) mohammedwasimrd@dhcp3 ~ % scp -i /Users/mohammedwasimrd/Documents/CSP554/emr-key-pair.pem /Users/mohammedwasimrd/Documents/CSP554/mohammedwasim.txt hadoop@ec2-54-208-219-119.compute-1.amazonaws.com
:/home/hadoop
mohammedwasim.txt                             100%   23     0.8KB/s   00:00
(base) mohammedwasimrd@dhcp3 ~ %
```

Step 9: Create s3 bucket and addA20497053.txt



EXECUTING COMMANDS :

1) hadoop fs –ls /

```
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /
Found 4 items
drwxr-xr-x   - hdfs hdfsadmingroup          0 2022-09-05 22:04 /apps
drwxrwxrwt   - hdfs hdfsadmingroup          0 2022-09-05 22:06 /tmp
drwxr-xr-x   - hdfs hdfsadmingroup          0 2022-09-05 22:04 /user
drwxr-xr-x   - hdfs hdfsadmingroup          0 2022-09-05 22:04 /var
[hadoop@ip-172-31-84-171 ~]$
```

2) hadoop fs –ls /user

```
[[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user
Found 6 items
drwxrwxrwx   - hadoop hdfsadmingroup          0 2022-09-05 22:04 /user/hadoop
drwxr-xr-x   - mapred mapred                  0 2022-09-05 22:04 /user/history
drwxrwxrwx   - hdfs   hdfsadmingroup          0 2022-09-05 22:04 /user/hive
drwxrwxrwx   - hue    hue                     0 2022-09-05 22:04 /user/hue
drwxrwxrwx   - oozie  oozie                   0 2022-09-05 22:06 /user/oozie
drwxrwxrwx   - root   hdfsadmingroup          0 2022-09-05 22:04 /user/root
[hadoop@ip-172-31-84-171 ~]$
```

3) Execute a command to create the CSP554 HDFS directory:

hadoop fs –mkdir/user/csp/csp554

```
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -mkdir /user/csp554
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user/csp554
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user/csp554
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user
Found 7 items
drwxr-xr-x   - hadoop hdfsadmingroup          0 2022-09-05 22:39 /user/csp554
drwxrwxrwx   - hadoop hdfsadmingroup          0 2022-09-05 22:04 /user/hadoop
drwxr-xr-x   - mapred mapred                  0 2022-09-05 22:04 /user/history
drwxrwxrwx   - hdfs   hdfsadmingroup          0 2022-09-05 22:04 /user/hive
drwxrwxrwx   - hue    hue                     0 2022-09-05 22:04 /user/hue
drwxrwxrwx   - oozie  oozie                   0 2022-09-05 22:06 /user/oozie
drwxrwxrwx   - root   hdfsadmingroup          0 2022-09-05 22:04 /user/root
[hadoop@ip-172-31-84-171 ~]$
```

4) Execute a command to create the CSP554-2 HDFS directory:

hadoop fs –mkdir/user/csp/csp554-2

```
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -mkdir /user/csp554-2
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user
Found 8 items
drwxr-xr-x   - hadoop hdfsadmingroup          0 2022-09-05 22:39 /user/csp554
drwxr-xr-x   - hadoop hdfsadmingroup          0 2022-09-05 22:41 /user/csp554-2
drwxrwxrwx   - hadoop hdfsadmingroup          0 2022-09-05 22:04 /user/hadoop
drwxr-xr-x   - mapred mapred                  0 2022-09-05 22:04 /user/history
drwxrwxrwx   - hdfs   hdfsadmingroup          0 2022-09-05 22:04 /user/hive
drwxrwxrwx   - hue    hue                     0 2022-09-05 22:04 /user/hue
drwxrwxrwx   - oozie  oozie                   0 2022-09-05 22:06 /user/oozie
drwxrwxrwx   - root   hdfsadmingroup          0 2022-09-05 22:04 /user/root
[hadoop@ip-172-31-84-171 ~]$
```

5) Execute a command that copies a given local file to the given hdfs directory

Hadoop fs -put mohammedwasim.txt/user/csp554

```
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -put mohammedwasim.txt /user/csp554
[[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user/csp554
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup         23 2022-09-05 22:44 /user/csp554/mohammedwasim.txt
[hadoop@ip-172-31-84-171 ~]$
```

6) Copy a file from one hdfs directory to another hdfs directory and write down the command.

Hadoop fs -cp/user/csp554/mohammedwasim.txt/user/csp554-2

```
[[hadoop@ip-172-31-84-171 ~]$ hadoop fs -cp /user/csp554/mohammedwasim.txt /user/csp554-2
[[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user/csp554-2
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup         23 2022-09-05 22:49 /user/csp554-2/mohammedwasim.txt
[hadoop@ip-172-31-84-171 ~]$
```

7) Copy the object myid.txt you uploaded to an S3 bucket into the Hadoop master node Linux file system. The actual object includes your student id .

Aws s3 cp s3://bucketa2049/A20497053.txt /home/Hadoop/A20497053.txt

```
[hadoop@ip-172-31-84-171 ~]$ aws s3 cp s3://bucketa2049/A20497053.txt   /home/hadoop/A20497053.txt
download: s3://bucketa2049/A20497053.txt to ./A20497053.txt
[hadoop@ip-172-31-84-171 ~]$ ls
A20497053.txt  mohammedwasim.txt
[hadoop@ip-172-31-84-171 ~]$
```

8) Copy the same object myid.txt you created in an S3 bucket into HDFS into the directory /users/csp554

Hadoop fs -cp s3://bucketa2049/A20497053.txt /home/Hadoop/A20497053.txt hdfs:///user/csp554-2

```
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -cp s3://bucketa2049/A20497053.txt    hdfs:///user/csp554-2
22/09/05 22:56:35 INFO s3n.S3NativeFileSystem: Opening 's3://bucketa2049/A20497053.txt' for reading
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user/csp554-2
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup         21 2022-09-05 22:56 /user/csp554-2/A20497053.txt
-rw-r--r--   1 hadoop hdfsadmingroup         23 2022-09-05 22:49 /user/csp554-2/mohammedwasim.txt
[hadoop@ip-172-31-84-171 ~]$
```

9) Execute a command to show the contents of the myid.txt file in the hdfs directory /user/csp554-2

Hadoop fs -cat/user/csp554-2/A20497053.txt

```
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -ls /user/csp554-2
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup         21 2022-09-05 22:56 /user/csp554-2/A20497053.txt
-rw-r--r--   1 hadoop hdfsadmingroup         23 2022-09-05 22:49 /user/csp554-2/mohammedwasim.txt
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -cat /user/csp554-2/A20497053.txt
this is the id file

[hadoop@ip-172-31-84-171 ~]$
```

10) Execute a command to remove the myid.txt file in the hdfs directory /user/csp554-2

Hadoop fs -rm/user/csp554-2/A20497053.txt

```
[hadoop@ip-172-31-84-171 ~]$ hadoop fs -rm /user/csp554-2/A20497053.txt
Deleted /user/csp554-2/A20497053.txt
[hadoop@ip-172-31-84-171 ~]$
```