

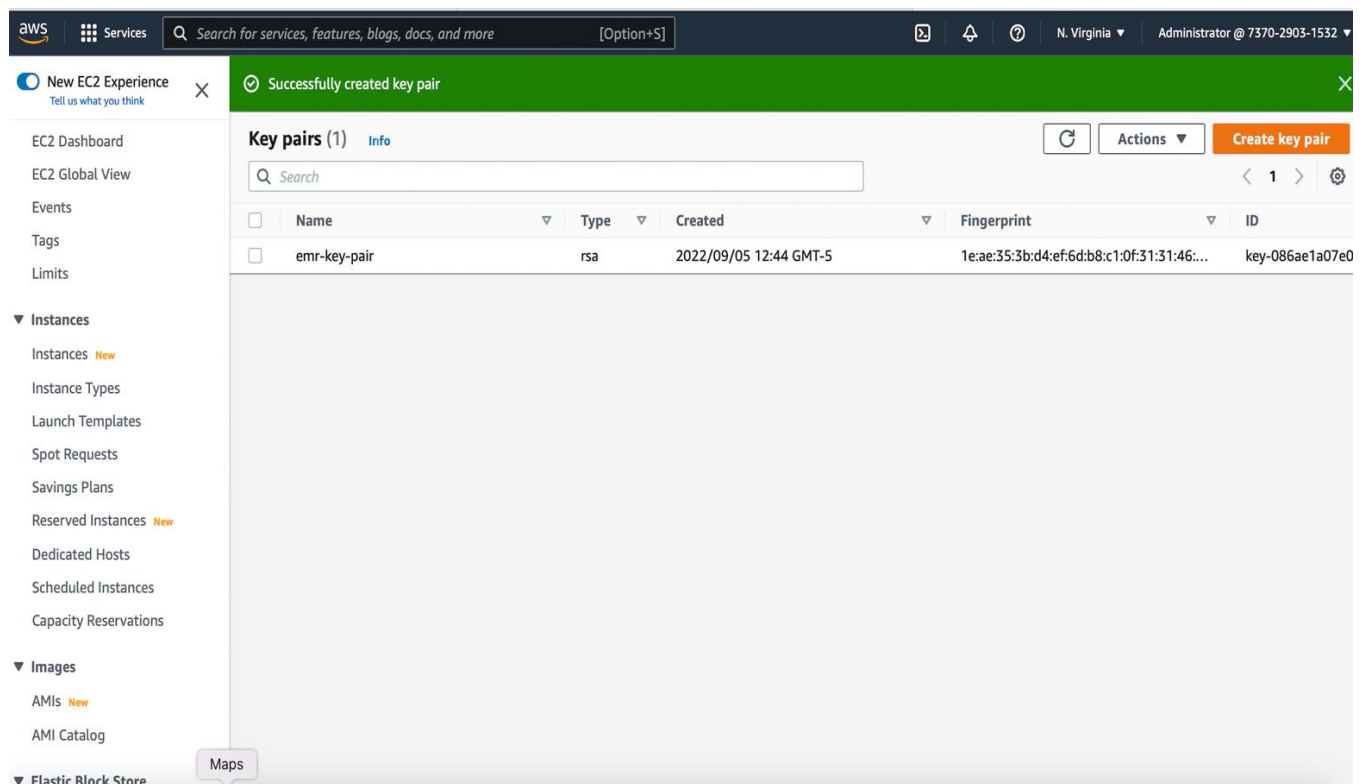
CSP 554 BIG DATA TECHNOLOGIES

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

aws

Services

Search for services, features, blogs, docs, and more

[Option+S]

N. Virginia

Administrator @ 7370-2903-1532

Amazon EMR

EMR Studio

EMR Serverless New

EMR on EC2

Clusters

Notebooks

Git repositories

Security configurations

Block public access

VPC subnets

Events

EMR on EKS

Virtual clusters

Help

What's new

EMR Serverless is now GA.
With EMR Serverless, get the benefits of Amazon EMR such as open source compatibility, latest versions and performance optimized runtime for popular frameworks along with easy provisioning, quick job startup, automatic capacity management, and simple cost controls. [Get Started with EMR Serverless.](#)

Clone Terminate AWS CLI export

Cluster: emr cluster 1 Starting

Summary Application user interfaces Monitoring Hardware Configurations Events Steps Bootstrap actions

Summary

ID: j-1OKJLUHI01KFG

Creation date: 2022-09-05 12:54 (UTC-5)

Elapsed time: 0 seconds

After last step completes: Cluster waits

Termination protection: Off [Change](#)

Tags: -- [View All / Edit](#)

Master public DNS: --

Configuration details

Release label: emr-5.36.0

Hadoop distribution: Amazon 2.10.1

Applications: Hive 2.3.9, Hue 4.10.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2

Log URI: s3://aws-logs-737029031532-us-east-1/elasticmapreduce/

EMRFS consistent view: Disabled

Custom AMI ID: --

Amazon Linux Release: 2.0.20220426.0 [Learn more](#)

Application user interfaces

Persistent user interfaces: --

On-cluster user interfaces: --

Network and hardware

Availability zone: --

Subnet ID: [subnet-02c1988428328033f](#)

Master: Provisioning 1 m4.large

Core: Provisioning 1 m4.large

Task: --

Cluster scaling: Not enabled

Auto-termination: Terminate if idle for 1 hour

aws

Services

Search for services, features, blogs, docs, and more

[Option+S]

N. Virginia

Administrator @ 7370-2903-1532

Amazon EMR

EMR Studio

EMR Serverless New

EMR on EC2

Clusters

Notebooks

Git repositories

Security configurations

Block public access

VPC subnets

Events

EMR on EKS

Virtual clusters

Help

What's new

EMR Serverless is now GA.
With EMR Serverless, get the benefits of Amazon EMR such as open source compatibility, latest versions and performance optimized runtime for popular frameworks along with easy provisioning, quick job startup, automatic capacity management, and simple cost controls. [Get Started with EMR Serverless.](#)

Clone Terminate AWS CLI export

Cluster: emr cluster 1 Waiting Cluster ready after last step completed.

Summary Application user interfaces Monitoring Hardware Configurations Events Steps Bootstrap actions

Summary

ID: j-1OKJLUHI01KFG

Creation date: 2022-09-05 12:54 (UTC-5)

Elapsed time: 15 minutes

After last step completes: Cluster waits

Termination protection: Off [Change](#)

Tags: -- [View All / Edit](#)

Master public DNS: ec2-3-235-50-92.compute-1.amazonaws.com [Connect to the Master Node Using SSH](#)

Configuration details

Release label: emr-5.36.0

Hadoop distribution: Amazon 2.10.1

Applications: Hive 2.3.9, Hue 4.10.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2

Log URI: s3://aws-logs-737029031532-us-east-1/elasticmapreduce/

EMRFS consistent view: Disabled

Custom AMI ID: --

Amazon Linux Release: 2.0.20220426.0 [Learn more](#)

Application user interfaces

Persistent user interfaces: [YARN timeline server](#), [Tez UI](#)

On-cluster user interfaces: Not Enabled [Enable an SSH Connection](#)

Network and hardware

Availability zone: us-east-1a

Subnet ID: [subnet-02c1988428328033f](#)

Master: Running 1 m4.large

Core: Running 1 m4.large

Task: --

Cluster scaling: Not enabled

Auto-termination: Terminate if idle for 1 hour

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

The screenshot shows the AWS Management Console interface. On the left is a navigation sidebar with options like EC2 Dashboard, EC2 Global View, Events, Tags, Limits, Instances, Images, and Elastic Block Store. The main content area is titled 'Security Groups (1/2) Info'. It features a search bar with the text 'search: sg-0bb26addc57efa05e' and a 'Clear filters' button. Below the search bar is a table listing security groups. The first row is selected, showing details for 'sg-0bb26addc57efa05e', which is the 'ElasticMapReduce-master' group. The second row is 'sg-02760b22ea1821477', the 'ElasticMapReduce-slave' group. Below the table, there's a section for 'Inbound rules (18)' with a search bar and a table of rules. At the bottom, there's a footer with 'Feedback', 'Looking for language selection? Find it in the new Unified Settings', and copyright information for Amazon Web Services.

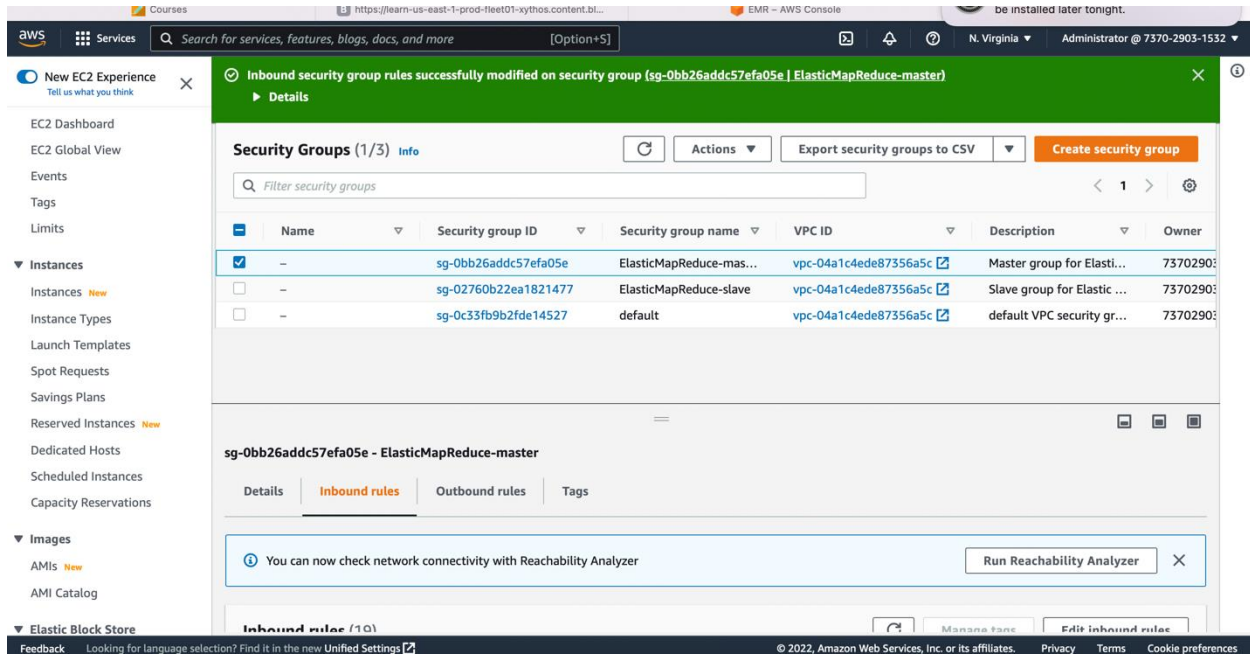
Name	Security group ID	Security group name	VPC ID	Description	Owner
-	sg-0bb26addc57efa05e	ElasticMapReduce-mas...	vpc-04a1c4ede87356a5c	Master group for Elastic...	7370290:
-	sg-02760b22ea1821477	ElasticMapReduce-slave	vpc-04a1c4ede87356a5c	Slave group for Elastic ...	7370290:

Step 5: Adding Inbound Rules

The screenshot shows the 'Add rule' dialog in the AWS Management Console. The dialog is open, showing a list of rules to be added. The rules are listed in a table with columns for 'Rule ID', 'Protocol', 'Port Range', 'Source', and 'Action'. The rules are: 'sgr-0432ad393a5a30799' (Custom TCP, 8443, Custom, 54.240.217.8/29), 'sgr-034c58663bb5edd28' (All UDP, 0 - 65535, Custom, 54.240.217.8/29), 'sgr-080a6f0afb6abf554' (All UDP, 0 - 65535, Custom, 54.240.217.8/29), 'sgr-0a874b055f09cf905' (All TCP, 0 - 65535, Custom, 54.240.217.8/29), 'sgr-0e511fe1fb1bda35d' (Custom TCP, 8443, Custom, 54.240.217.8/29), and 'SSH' (SSH, TCP, 22, My IP, 104.194.104.3/32). The 'Add rule' button is at the bottom left.

Rule ID	Protocol	Port Range	Source	Action
sgr-0432ad393a5a30799	Custom TCP	8443	Custom	54.240.217.8/29
sgr-034c58663bb5edd28	All UDP	0 - 65535	Custom	54.240.217.8/29
sgr-080a6f0afb6abf554	All UDP	0 - 65535	Custom	54.240.217.8/29
sgr-0a874b055f09cf905	All TCP	0 - 65535	Custom	54.240.217.8/29
sgr-0e511fe1fb1bda35d	Custom TCP	8443	Custom	54.240.217.8/29
-	SSH	TCP	22	My IP

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.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::E M:::::::::M M:::::::::M R::::::::::::R
EE::::::::::::::::::E M:::::::::M M:::::::::M R::::RRRRR::::R
E::::E EEEEE M:::::::::M M:::::::::M RR::::R R::::R
E::::E M:::::::::M M:::::::::M R::::R R::::R
E::::EEEEEEEEEEEE M:::::::::M M:::::::::M R::::RRRRR::::R
E::::::::::::::::::E M:::::::::M M:::::::::M R::::::::::::R
E::::EEEEEEEEEEEE M:::::::::M M:::::::::M R::::RRRRR::::R
E::::E M:::::::::M M:::::::::M M:::::::::M R::::R R::::R
E::::E EEEEE M:::::::::M M M M:::::::::M R::::R R::::R
EE::::::::::::::::::E M:::::::::M M:::::::::M R::::R R::::R
E::::::::::::::::::E M:::::::::M M:::::::::M RR::::R R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRR 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:/home/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

The screenshot shows the AWS S3 console interface. At the top, there's a navigation bar with the AWS logo, 'Services' link, a search bar, and user information. Below the navigation bar, a green banner indicates 'Upload succeeded' with a 'View details below' link. The main content area shows a 'Summary' section with a table of upload results. The table has three columns: Destination, Succeeded, and Failed. The Destination is 's3://bucketa2049'. The Succeeded column shows '1 file, 21.0 B (100.00%)' with a green checkmark. The Failed column shows '0 files, 0 B (0%)' with a red X. Below the summary, there are tabs for 'Files and folders' (selected) and 'Configuration'. The 'Files and folders' tab shows a list of files and folders. The list has a search bar and a table with columns: Name, Folder, Type, Size, Status, and Error. The table contains one entry: 'A20497053.txt' in the 'Name' column, '-' in the 'Folder' column, 'text/plain' in the 'Type' column, '21.0 B' in the 'Size' column, 'Succeeded' in the 'Status' column, and '-' in the 'Error' column.

Destination	Succeeded	Failed
s3://bucketa2049	1 file, 21.0 B (100.00%)	0 files, 0 B (0%)

Name	Folder	Type	Size	Status	Error
A20497053.txt	-	text/plain	21.0 B	Succeeded	-

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 ~]$
```