

Assignment #2 (Modules 02a & 02b, 20 points)

A20512400

9. (2 points) Execute the following `hdfs` command to list the files or directories that are listed (also indicating which is a file and which a directory):

Take a screen snapshot of names of the files or directories that are listed and include it in your assignment submission.

`hadoop fs -ls /`

```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-237-31-143.compute-1.amazonaws.com — 176x37
/etc/profile.d/lang.sh: line 19: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
pradaap.txt
pradaapss@dhcp54 ~ % ssh -i /Users/pradaapss/Downloads/emr-key-pair.pem hadoop@ec2-54-237-31-143.compute-1.amazonaws.com
Last login: Thu Sep 15 03:49:55 2022

 _ _ | _ _ |
 _ _ | _ _ | /
 _ _ | _ _ | /

Amazon Linux 2 AMI

https://aws.amazon.com/amazon-linux-2/
18 package(s) needed for security, out of 38 available
Run "sudo yum update" to apply all updates.
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory

EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRRRRRRRRR
E:::EEEEEEEEEEEEEEEE M:::M M:::M R:::R
EE:::EEEEEEEEEEEEEEEE M:::M M:::M R:::R
E:::E EEEEE M:::M M:::M R:::R
E:::E M:::M M:::M M:::M R:::R
E:::EEEEEEEEEEEE M:::M M:::M M:::M R:::R
E:::EEEEEEEEEEEE M:::M M:::M M:::M R:::R
E:::E EEEEE M:::M M:::M M:::M R:::R
EE:::EEEEEEEEEEEE M:::M M:::M M:::M R:::R
E:::EEEEEEEEEEEE M:::M M:::M M:::M R:::R
EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRRR RRRRRR

[hadoop@ip-172-31-57-194 ~]$ Hadoop fs -ls /
-bash: Hadoop: command not found
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /
Found 4 items
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-09-15 03:41 /apps
drwxrwxrwt - hdfs hdfsadmingroup 0 2022-09-15 03:43 /tmp
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-09-15 03:41 /user
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-09-15 03:41 /var
[hadoop@ip-172-31-57-194 ~]$
```

10. (2 points) Execute a command (you needed to figure out which one) to list the files and directories under the `hdfs` directory listed below:

Write down the command you executed and also take a screen snapshot of names of the files or directories that are listed and include it in your assignment submission.

`hadoop fs -ls /user`

```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-237-31-143.compute-1....
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /user
Found 6 items
drwxrwxrwx - hadoop hdfsadmingroup 0 2022-09-15 03:41 /user/hadoop
drwxr-xr-x - mapred mapred 0 2022-09-15 03:41 /user/history
drwxrwxrwx - hdfs hdfsadmingroup 0 2022-09-15 03:41 /user/hive
drwxrwxrwx - hue hue 0 2022-09-15 03:41 /user/hue
drwxrwxrwx - oozie oozie 0 2022-09-15 03:43 /user/oozie
drwxrwxrwx - root hdfsadmingroup 0 2022-09-15 03:41 /user/root
```

11. (2 points) Execute a command to create the following HDFS directory:

Record the command you executed and include it in your assignment submission.

`hadoop fs -mkdir /user/csp554`

```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-237-31-143.compute-1....
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -mkdir /user/csp554
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /
Found 4 items
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-09-15 03:41 /apps
drwxrwxrwt - hdfs hdfsadmingroup 0 2022-09-15 03:43 /tmp
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-09-15 03:58 /user
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-09-15 03:41 /var
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /user
Found 7 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2022-09-15 03:58 /user/csp554
drwxrwxrwx - hadoop hdfsadmingroup 0 2022-09-15 03:41 /user/hadoop
drwxr-xr-x - mapred mapred 0 2022-09-15 03:41 /user/history
drwxrwxrwx - hdfs hdfsadmingroup 0 2022-09-15 03:41 /user/hive
drwxrwxrwx - hue hue 0 2022-09-15 03:41 /user/hue
drwxrwxrwx - oozie oozie 0 2022-09-15 03:43 /user/oozie
drwxrwxrwx - root hdfsadmingroup 0 2022-09-15 03:41 /user/root
```

12. (2 points) Execute a command to create the following HDFS directory:

Record the command you executed and include it in your assignment submission.

`hadoop fs -mkdir /user/csp554-2`

```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-237-31-143.compute-1....
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -mkdir /user/csp554-2
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /user/
Found 8 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2022-09-15 04:01 /user/csp554
drwxr-xr-x - hadoop hdfsadmingroup 0 2022-09-15 04:03 /user/csp554-2
drwxrwxrwx - hadoop hdfsadmingroup 0 2022-09-15 04:02 /user/hadoop
drwxr-xr-x - mapred mapred 0 2022-09-15 03:41 /user/history
drwxrwxrwx - hdfs hdfsadmingroup 0 2022-09-15 03:41 /user/hive
drwxrwxrwx - hue hue 0 2022-09-15 03:41 /user/hue
drwxrwxrwx - oozie oozie 0 2022-09-15 03:43 /user/oozie
drwxrwxrwx - root hdfsadmingroup 0 2022-09-15 03:41 /user/root
```

13. (2 points) Execute a command that copies a given local file to the given hdfs directory :

Source local file: /home/hadoop/myname.txt (where the actual name is your name as described above)

Destination HDFS directory: /user/csp554

Record the command you executed and include it in your assignment submission.

`hadoop fs -put pradaap.txt /user/csp554`

```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-237-31-143.compute-1....
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -put pradaap.txt /user/csp554
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /user/csp554
Found 1 items
-rw-r--r-- 1 hadoop hdfsadmingroup 25 2022-09-15 04:01 /user/csp554/pradaap.txt
```

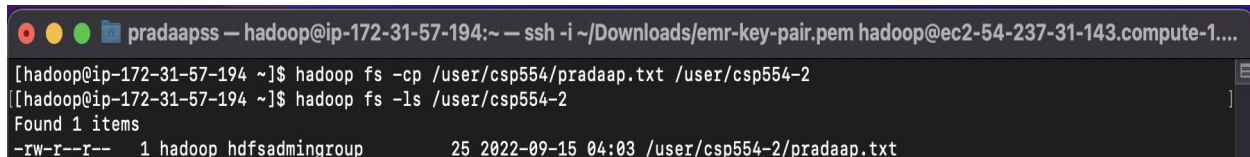
14. (2 points) Copy a file from one hdfs directory to another hdfs directory and write down the command

Source hdfs file: /user/csp554/myname.txt (where the actual name is your name as described above)

Destination HDFS directory: /user/csp554-2

Record the command you executed and include it in your assignment submission.

```
hadoop fs -cp /user/csp554/pradaap.txt /user/csp554-2
```



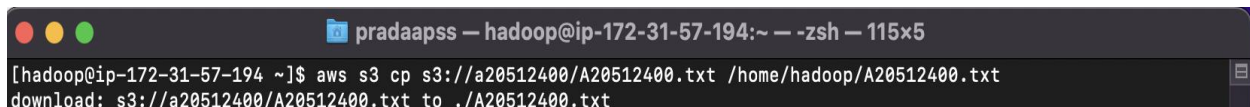
```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-237-31-143.compute-1...
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -cp /user/csp554/pradaap.txt /user/csp554-2
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /user/csp554-2
Found 1 items
-rw-r--r-- 1 hadoop hdfsadmingroup 25 2022-09-15 04:03 /user/csp554-2/pradaap.txt
```

15. (2 points) 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 as above.

Note, Amazon EMR and Hadoop provide a variety of file systems that you can use with EMR. You specify which file system to use with a file system prefix. For example, `s3://myawsbucket/path` references an Amazon S3 bucket using EMRFS (EMR file system). See: <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-file-systems.html>

The way you do this would be as follows to copy an object from an S3 bucket to the Linux file system of the Hadoop master node.

```
aws s3 cp s3://a20512400/A20512400.txt /home/hadoop/A20512400.txt
```

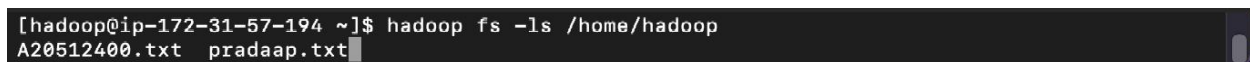


```
pradaapss — hadoop@ip-172-31-57-194:~ — zsh — 115x5
[hadoop@ip-172-31-57-194 ~]$ aws s3 cp s3://a20512400/A20512400.txt /home/hadoop/A20512400.txt
download: s3://a20512400/A20512400.txt to ./A20512400.txt
```

The above is an AWS CLI (command line interpreter) command. For more information about how to use the CLI to manipulate S3 buckets see:

<https://docs.aws.amazon.com/cli/latest/reference/s3/index.html>

After you executed the above command perform an `ls /home/hadoop` and take a screen snapshot of names of the files or directories that are listed and include it in your assignment submission.



```
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /home/hadoop
A20512400.txt pradaap.txt
```

16. (2 points) Copy the same object myid.txt you created in an S3 bucket into HDFS into the directory /users/csp554

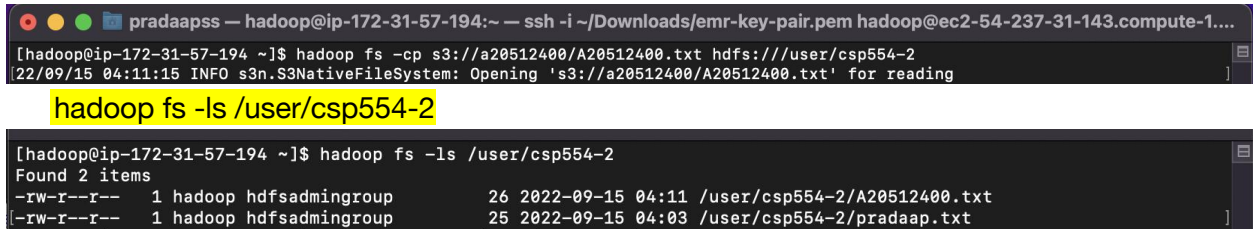
```
hadoop fs -cp s3://mybucket/myid.txt hdfs:///user/csp554-2
```

Note, the three slashes after the “hdfs:”

After you executed the above command, execute another command (you needed to figure out which one) to list the files and directories under the hdfs directory listed below:

Write down the command you executed and also take a screen snapshot of names of the files or directories that are listed and include it in your assignment submission.

```
hadoop fs -cp s3://a20512400/A20512400.txt hdfs:///user/csp554-2
```



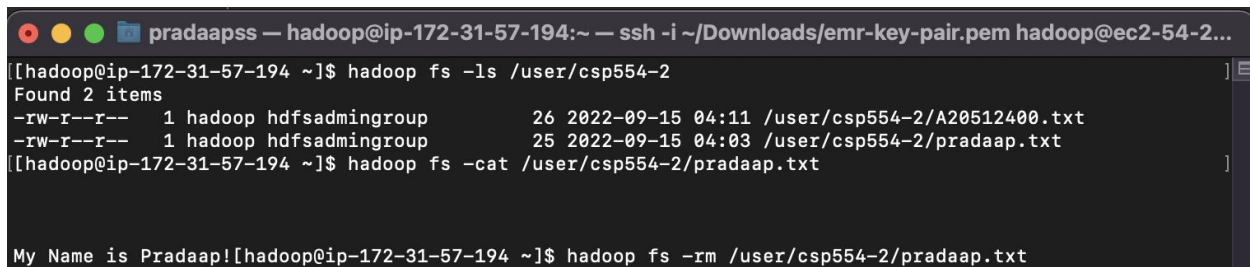
```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-237-31-143.compute-1...  
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -cp s3://a20512400/A20512400.txt hdfs:///user/csp554-2  
22/09/15 04:11:15 INFO s3n.S3NativeFileSystem: Opening 's3://a20512400/A20512400.txt' for reading  
  
hadoop fs -ls /user/csp554-2  
  
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /user/csp554-2  
Found 2 items  
-rw-r--r-- 1 hadoop hdfsadmin group 26 2022-09-15 04:11 /user/csp554-2/A20512400.txt  
-rw-r--r-- 1 hadoop hdfsadmin group 25 2022-09-15 04:03 /user/csp554-2/pradaap.txt
```

17. (2 points) Execute a command to show the contents of the myid.txt file in the hdfs directory /user/csp554-2

Clue: look up about how to use the “cat” command in the file system shell document.

Write down the command you executed and also take a screen snapshot of the listed content of the file and include it in your assignment submission.

```
hadoop fs -ls /user/csp554-2
```



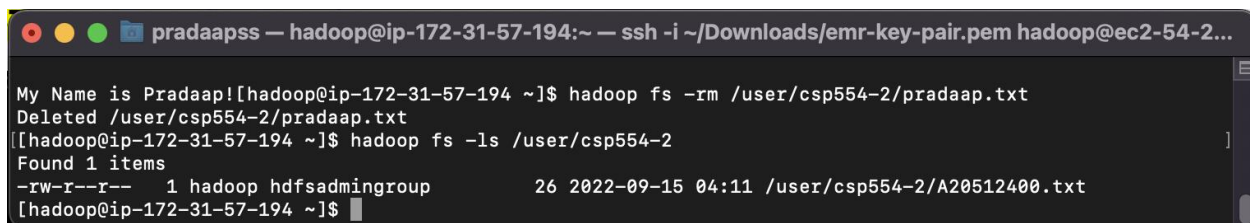
```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-2...  
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /user/csp554-2  
Found 2 items  
-rw-r--r-- 1 hadoop hdfsadmin group 26 2022-09-15 04:11 /user/csp554-2/A20512400.txt  
-rw-r--r-- 1 hadoop hdfsadmin group 25 2022-09-15 04:03 /user/csp554-2/pradaap.txt  
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -cat /user/csp554-2/pradaap.txt  
  
My Name is Pradaap! [hadoop@ip-172-31-57-194 ~]$ hadoop fs -rm /user/csp554-2/pradaap.txt
```

18. (2 points) Execute a command to remove the myid.txt file in the hdfs directory /user/csp554-2

Clue: look up about how to use the “rm” command in the file system shell document.

Write down the command you executed, then list the content of the /user/csp554-2 HDFS directory and take a screen snapshot of the listed content of the directory and include it in your assignment submission.

```
hadoop fs -rm /user/csp554-2/pradaap.txt
```



```
pradaapss — hadoop@ip-172-31-57-194:~ — ssh -i ~/Downloads/emr-key-pair.pem hadoop@ec2-54-2...  
My Name is Pradaap! [hadoop@ip-172-31-57-194 ~]$ hadoop fs -rm /user/csp554-2/pradaap.txt  
Deleted /user/csp554-2/pradaap.txt  
[hadoop@ip-172-31-57-194 ~]$ hadoop fs -ls /user/csp554-2  
Found 1 items  
-rw-r--r-- 1 hadoop hdfsadmin group 26 2022-09-15 04:11 /user/csp554-2/A20512400.txt  
[hadoop@ip-172-31-57-194 ~]$
```