

CSP554—Big Data Technologies

Assignment #2

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):

```
hadoop fs -ls /
```

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

There are 4 directories apps, tmp, user, var

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -ls /  
Found 4 items  
drwxr-xr-x - hdfs hadoop      0 2019-09-03 23:04 /apps  
drwxrwxrwt - hdfs hadoop      0 2019-09-03 23:06 /tmp  
drwxr-xr-x - hdfs hadoop      0 2019-09-03 23:04 /user  
drwxr-xr-x - hdfs hadoop      0 2019-09-03 23:04 /var
```

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:

/user

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

There are 6 directories Hadoop, history, hive, hue, oozie, root

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -ls /user  
Found 6 items  
drwxrwxrwx - hadoop hadoop      0 2019-09-03 23:04 /user/hadoop  
drwxr-xr-x - mapred mapred      0 2019-09-03 23:04 /user/history  
drwxrwxrwx - hdfs hadoop      0 2019-09-03 23:04 /user/hive  
drwxrwxrwx - hue hue          0 2019-09-03 23:04 /user/hue  
drwxrwxrwx - oozie oozie      0 2019-09-03 23:04 /user/oozie  
drwxrwxrwx - root hadoop      0 2019-09-03 23:04 /user/root
```

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

/user/csp554

Ans: Command used is **hadoop fs -mkdir /user/csp554**

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -mkdir /user/csp554
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -ls /user
Found 7 items
drwxr-xr-x - hadoop hadoop 0 2019-09-04 00:48 /user/csp554
drwxrwxrwx - hadoop hadoop 0 2019-09-03 23:04 /user/hadoop
drwxr-xr-x - mapred mapred 0 2019-09-03 23:04 /user/history
drwxrwxrwx - hdfs hadoop 0 2019-09-03 23:04 /user/hive
drwxrwxrwx - hue hue 0 2019-09-03 23:04 /user/hue
drwxrwxrwx - oozie oozie 0 2019-09-03 23:04 /user/oozie
drwxrwxrwx - root hadoop 0 2019-09-03 23:04 /user/root
[hadoop@ip-172-31-5-53 ~]$ Connection reset by 3.15.220.114 port 22
```

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

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

/user/csp554-2

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

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

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -mkdir /user/csp554-2
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -ls /user
Found 8 items
drwxr-xr-x - hadoop hadoop 0 2019-09-04 00:48 /user/csp554
drwxr-xr-x - hadoop hadoop 0 2019-09-04 02:29 /user/csp554-2
drwxrwxrwx - hadoop hadoop 0 2019-09-03 23:04 /user/hadoop
drwxr-xr-x - mapred mapred 0 2019-09-03 23:04 /user/history
drwxrwxrwx - hdfs hadoop 0 2019-09-03 23:04 /user/hive
drwxrwxrwx - hue hue 0 2019-09-03 23:04 /user/hue
drwxrwxrwx - oozie oozie 0 2019-09-03 23:04 /user/oozie
drwxrwxrwx - root hadoop 0 2019-09-03 23:04 /user/root
[hadoop@ip-172-31-5-53 ~]$ |
```

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 /home/hadoop/juhideshpande.txt /user/csp554`

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -put /home/hadoop/juhideshpande.txt /user/csp554
[hadoop@ip-172-31-5-53 ~]$ ls /user
ls: cannot access /user: No such file or directory
[hadoop@ip-172-31-5-53 ~]$ hdfs dfs -ls /user/csp554
Found 1 items
-rw-r--r-- 1 hadoop hadoop          20 2019-09-05 00:36 /user/csp554/juhideshpande.txt
[hadoop@ip-172-31-5-53 ~]$
```

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 -put /home/ hadoop /juhideshpande.txt /user/csp554-2`

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -put /home/hadoop/juhideshpande.txt /user/csp554-2
[hadoop@ip-172-31-5-53 ~]$ hdfs dfs -ls /user/csp554-2
Found 1 items
-rw-r--r-- 1 hadoop hadoop          20 2019-09-05 00:38 /user/csp554-2/juhideshpande.txt
[hadoop@ip-172-31-5-53 ~]$
```

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://mybucket/myid.txt /home/hadoop/myid.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.

There are 2 files listed:

juhideshpande.txt

A20430435.txt

```
[hadoop@ip-172-31-5-53 ~]$ aws s3 cp s3://juhideshpande/A20430435.txt /home/hadoop/A20430435.txt
download: s3://juhideshpande/A20430435.txt to ./A20430435.txt
[hadoop@ip-172-31-5-53 ~]$ ls /home/hadoop
A20430435.txt juhideshpande.txt
[hadoop@ip-172-31-5-53 ~]$
```

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:

/user/csp554-2

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.

The command used for listing is **hdfs dfs -ls /user/csp554-2**

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -cp s3://juhideshpande/A20430435.txt hdfs:///user/csp554-2
19/09/05 01:04:00 INFO executor.GlobalS3Executor: Bucket juhideshpande is in the us-west-2 region. Please configure the proper region to avoid multiple unnecessary redirects
19/09/05 01:04:01 INFO s3n.S3NativeFileSystem: Opening 's3://juhideshpande/A20430435.txt' for reading
[hadoop@ip-172-31-5-53 ~]$ hdfs dfs -ls /user/csp554-2
Found 2 items
-rw-r--r-- 1 hadoop hadoop      21 2019-09-05 01:04 /user/csp554-2/A20430435.txt
-rw-r--r-- 1 hadoop hadoop      20 2019-09-05 00:38 /user/csp554-2/juhideshpande.txt
[hadoop@ip-172-31-5-53 ~]$
```

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 -cat /user/csp554-2/A20430435.txt
```

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -cat /user/csp554-2/A20430435.txt
this is the id file
[hadoop@ip-172-31-5-53 ~]$
```

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/A20430435.txt
There is 1 file juhideshpande.txt present

```
[hadoop@ip-172-31-5-53 ~]$ hadoop fs -rm /user/csp554-2/A20430435.txt
Deleted /user/csp554-2/A20430435.txt
[hadoop@ip-172-31-5-53 ~]$ hdfs dfs -ls /user/csp554-2
Found 1 items
-rw-r--r-- 1 hadoop hadoop          20 2019-09-05 00:38 /user/csp554-2/juhideshpande.txt
[hadoop@ip-172-31-5-53 ~]$
```

19. This might be very important to your wallet 😊.

Follow Step 7 in the AWS EMR Instructions document to terminate your cluster and delete any buckets you have created. If you forget you may end up paying (a lot) more than you need to.

The screenshot shows the AWS Management Console interface for an Amazon EMR cluster. The cluster is named 'my-first-emr-cluster' and is in a 'Terminated' state. The console displays various tabs for cluster management, including Summary, Application history, Monitoring, Hardware, Configurations, Events, Steps, and Bootstrap actions. The Summary tab is selected, showing details such as the cluster ID (j-3GTJAUSBY7AU), creation date (2019-09-03 17:57 UTC-5), end date (2019-09-04 20:17 UTC-5), and elapsed time (1 day, 2 hours). The configuration details section shows the release label (emr-5.26.0), Hadoop distribution (Amazon 2.8.5), and applications (Ganglia 3.7.2, Hive 2.3.5, Hue 4.4.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2). The network and hardware section shows the availability zone (us-east-2a) and subnet ID (subnet-d09b89b8). The security and access section shows the key name (emr-key-pair) and EC2 instance profile (EMR_EC2_DefaultRole).

Also deleted the bucket successfully