

CSP554—Big Data Technologies

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

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 /`

```
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -ls /
Found 4 items
drwxr-xr-x - hdfs hdfsadmingroup 0 2024-02-01 22:32 /apps
drwxrwxrwt - hdfs hdfsadmingroup 0 2024-02-01 22:35 /tmp
drwxr-xr-x - hdfs hdfsadmingroup 0 2024-02-01 23:44 /user
drwxr-xr-x - hdfs hdfsadmingroup 0 2024-02-01 22:32 /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`

```
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -ls /user
Found 6 items
drwxrwxrwx - hadoop hdfsadmingroup 0 2024-02-01 22:32 /user/hadoop
drwxr-xr-x - mapred mapred 0 2024-02-01 22:32 /user/history
drwxrwxrwx - hdfs hdfsadmingroup 0 2024-02-01 22:32 /user/hive
drwxrwxrwx - hue hue 0 2024-02-01 22:32 /user/hue
drwxrwxrwx - oozie oozie 0 2024-02-01 22:34 /user/oozie
drwxrwxrwx - root hdfsadmingroup 0 2024-02-01 22:32 /user/root
```

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

`/user/csp554`

```
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -mkdir /user/csp554
mkdir: '/user/csp554': File exists
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -ls /user
Found 7 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2024-02-01 23:41 /user/csp554
drwxrwxrwx - hadoop hdfsadmingroup 0 2024-02-01 22:32 /user/hadoop
drwxr-xr-x - mapred mapred 0 2024-02-01 22:32 /user/history
drwxrwxrwx - hdfs hdfsadmingroup 0 2024-02-01 22:32 /user/hive
drwxrwxrwx - hue hue 0 2024-02-01 22:32 /user/hue
drwxrwxrwx - oozie oozie 0 2024-02-01 22:34 /user/oozie
drwxrwxrwx - root hdfsadmingroup 0 2024-02-01 22:32 /user/root
```

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

`/user/csp554-2`

```
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -mkdir /user/csp554-2
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -mkdir /user
mkdir: '/user': File exists
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -ls /user/csp554-2
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -ls /user
Found 8 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2024-02-01 23:41 /user/csp554
drwxr-xr-x - hadoop hdfsadmingroup 0 2024-02-01 23:44 /user/csp554-2
drwxrwxrwx - hadoop hdfsadmingroup 0 2024-02-01 22:32 /user/hadoop
drwxr-xr-x - mapred mapred 0 2024-02-01 22:32 /user/history
drwxrwxrwx - hdfs hdfsadmingroup 0 2024-02-01 22:32 /user/hive
drwxrwxrwx - hue hue 0 2024-02-01 22:32 /user/hue
drwxrwxrwx - oozie oozie 0 2024-02-01 22:34 /user/oozie
drwxrwxrwx - root hdfsadmingroup 0 2024-02-01 22:32 /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

```
[hadoop@ip-172-31-12-51 /]$ hadoop fs -copyFromLocal /home/hadoop/prachi.txt /user/csp554
```

```
Found 1 items
-rw-r--r-- 1 hadoop hdfsadmin group 23 2024-02-02 00:40 /user/csp554-2/prachi.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

```
[hadoop@ip-172-31-12-51 /]$ hadoop fs -cp /user/csp554-2/prachi.txt /user/csp554-2
```

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.

```
[hadoop@ip-172-31-12-51 /]$ aws s3 cp s3://assign2prachia20549927/A20549927.txt /home/hadoop/A20549927.txt
download: s3://assign2prachia20549927/A20549927.txt to home/hadoop/A20549927.txt
[hadoop@ip-172-31-12-51 /]$ ls /home/hadoop
A20549927.txt prachi.txt
```

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

```
[hadoop@ip-172-31-12-51 /]$ hadoop fs -cp s3://assign2prachia20549927/A20549927.txt hdfs:///user/csp554-2
2024-02-02 00:51:10,117 INFO s3n.S3NativeFileSystem: opening 's3://assign2prachia20549927/A20549927.txt' for reading
[hadoop@ip-172-31-12-51 /]$ hadoop fs -ls /user/csp554-2
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmin group 19 2024-02-02 00:51 /user/csp554-2/A20549927.txt
-rw-r--r-- 1 hadoop hdfsadmin group 23 2024-02-02 00:40 /user/csp554-2/prachi.txt
```

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

```
[hadoop@ip-172-31-12-51 ~]$ ls
A20549927.txt prachi.txt
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -cat /user/csp554-2/A20549927.txt
this is the id file[hadoop@icd home^C
```

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

```
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -rm /user/csp554-2/A20549927.txt
Deleted /user/csp554-2/A20549927.txt
[hadoop@ip-172-31-12-51 ~]$ hadoop fs -ls /user/csp554-2
Found 1 items
-rw-r--r-- 1 hadoop hdfsadmin group 23 2024-02-02 00:40 /user/csp554-2/prachi.txt
```

19. This might be very important to your wallet. Bucket deleting and termination section.

My cluster Updated less than a minute ago [Refresh](#) [Terminate](#) [Clone in AWS CLI](#) [Clone](#)

▼ Summary

Cluster info Cluster ID j-1U3ZC70IE3SZ5 Cluster configuration Instance groups Capacity 1 Primary 1 Core 0 Task	Applications Amazon EMR version emr-7.0.0 Installed applications Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2	Cluster management Log destination in Amazon S3 aws-logs-905418339338-us-east-2/elasticmapreduce Persistent application Uls YARN timeline server Tez UI Primary node public DNS ec2-3-14-141-186.us-east-2.compute.amazonaws.com Connect to the Primary node using SSH	Status and time Status Terminated Creation time February 01, 2024, 16:16 (UTC-06:00) Elapsed time 3 hours, 18 minutes End time February 01, 2024, 19:35 (UTC-06:00)
---	--	---	--

Clusters (3) [Info](#) [Refresh](#) [View details](#) [Terminate](#) [Clone](#) [Create cluster](#)

Filter clusters by status [Find clusters](#) Filter clusters by creation date-time < 1 > ⚙️

<input type="checkbox"/>	Cluster ID	Cluster name	Status	Creation time (UTC-06:00)
<input type="checkbox"/>	j-1U3ZC70IE3SZ5	My cluster	Terminated User request	February 01, 2024, 16:16

aws Services Search [Alt+S] Global Prachi Kotadia

Successfully emptied bucket "assign2prachia20549927" View details below. If you want to delete this bucket, use the [delete bucket configuration](#).

Empty bucket: status [Cancel](#) [Exit](#)

The details below are no longer available after you navigate away from this page.

Summary

Source s3://assign2prachia20549927	Successfully deleted ✔️ 1 object, 19.0 B	Failed to delete 0 objects
---------------------------------------	---	-------------------------------