

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-172-31-9-120 ~]$ hadoop fs -ls /  
Found 3 items  
drwxrwxrwt - hdfs hdfsadmingroup 0 2023-09-13 20:24 /tmp  
drwxr-xr-x - hdfs hdfsadmingroup 0 2023-09-13 20:24 /user  
drwxr-xr-x - hdfs hdfsadmingroup 0 2023-09-13 20:24 /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  
Found 5 items  
drwxrwxrwx - hadoop hdfsadmingroup 0 2023-09-13 20:24 /user/hadoop  
drwxrwxrwx - livy livy 0 2023-09-13 20:24 /user/livy  
drwxrwxrwx - root hdfsadmingroup 0 2023-09-13 20:24 /user/root  
drwxrwxrwx - spark spark 0 2023-09-13 20:24 /user/spark  
drwxrwxrwx - zeppelin hdfsadmingroup 0 2023-09-13 20:24 /user/zeppelin
```

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

```
/user/csp554  
[hadoop@ip-172-31-9-120 ~]$ hadoop fs -mkdir /user/csp554  
mkdir: `/user/csp554': File exists  
[hadoop@ip-172-31-9-120 ~]$ hadoop fs -ls /user  
Found 6 items  
drwxr-xr-x - hadoop hdfsadmingroup 0 2023-09-13 21:43 /user/csp554  
drwxrwxrwx - hadoop hdfsadmingroup 0 2023-09-13 20:24 /user/hadoop  
drwxrwxrwx - livy livy 0 2023-09-13 20:24 /user/livy  
drwxrwxrwx - root hdfsadmingroup 0 2023-09-13 20:24 /user/root  
drwxrwxrwx - spark spark 0 2023-09-13 20:24 /user/spark  
drwxrwxrwx - zeppelin hdfsadmingroup 0 2023-09-13 20:24 /user/zeppelin
```

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

```
/user/csp554-2  
Found 7 items  
drwxr-xr-x - hadoop hdfsadmingroup 0 2023-09-13 21:43 /user/csp554  
drwxr-xr-x - hadoop hdfsadmingroup 0 2023-09-13 21:45 /user/csp554-2  
drwxrwxrwx - hadoop hdfsadmingroup 0 2023-09-13 20:24 /user/hadoop  
drwxrwxrwx - livy livy 0 2023-09-13 20:24 /user/livy  
drwxrwxrwx - root hdfsadmingroup 0 2023-09-13 20:24 /user/root  
drwxrwxrwx - spark spark 0 2023-09-13 20:24 /user/spark  
drwxrwxrwx - zeppelin hdfsadmingroup 0 2023-09-13 20:24 /user/zeppelin
```

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-172-31-9-120 ~]$ hadoop fs -copyFromLocal /home/hadoop/kevan.txt /user/csp554
```

```
Found 1 items
-rw-r--r-- 1 hadoop hdfsadmin 21 2023-09-13 21:51 /user/csp554/kevan.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-172-31-9-120 ~]$ hadoop fs -cp /user/csp554/kevan.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-172-31-9-120 ~]$ aws s3 cp s3://myawsbucket-1110/A20522659.txt /home/hadoop/A20522659.txt
download: s3://myawsbucket-1110/A20522659.txt to ./A20522659.txt
[hadoop@ip-172-31-9-120 ~]$ ls /home/hadoop
A20522659.txt  kevan.txt
[hadoop@ip-172-31-9-120 ~]$ |
```

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

```
[hadoop@ip-172-31-9-120 ~]$ hadoop fs -cp s3://myawsbucket-1110/A20522659.txt hdfs:///user/csp554-2
2023-09-13 22:10:17,489 INFO s3n.S3NativeFileSystem: Opening 's3://myawsbucket-1110/A20522659.txt' for reading
[hadoop@ip-172-31-9-120 ~]$ hadoop fs -ls /user/csp554-2
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmin 23 2023-09-13 22:10 /user/csp554-2/A20522659.txt
-rw-r--r-- 1 hadoop hdfsadmin 21 2023-09-13 22:03 /user/csp554-2/kevan.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-172-31-9-120 ~]$ ls
A20522659.txt  kevan.txt
[hadoop@ip-172-31-9-120 ~]$ hadoop fs -cat /user/csp554-2/A20522659.txt
this is the name file
```

Note: By mistake, I changed a text in the .txt file so instead of the id file it is coming name file as an output kindly I apologize for the same.

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

```
[hadoop@ip-172-31-9-120 ~]$ hadoop fs -rm /user/csp554-2/A20522659.txt
Deleted /user/csp554-2/A20522659.txt
[hadoop@ip-172-31-9-120 ~]$ hadoop fs -ls /user/csp554-2
Found 1 items
-rw-r--r-- 1 hadoop hdfsadmin 21 2023-09-13 22:03 /user/csp554-2/kevan.txt
[hadoop@ip-172-31-9-120 ~]$ |
```

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

Clusters (1/1) Info

Filter clusters by status

Find clusters

Filter clusters by creation date-time

View details

Terminate

Clone

Create cluster

	Cluster Id	Cluster name	Status	Creation time (UTC-05:00)	Elapsed time
<input checked="" type="checkbox"/>	j-34N2LMF4FGT9O	My cluster	Terminating User request	September 13, 2023, 15:15	2 hours, 13 minutes

Summary

Source
[s3://myawsbucket-1110](#)

Successfully deleted
1 object, 23.0 B

Failed to delete
0 objects

Failed to delete (0)

Find objects by name

Name	Prefix	Version ID	Type	Last modified	Size	Error
No failed object deletions						