SP554—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-172-31-49-62 ~]$ hadoop fs -ls
SLF4J: Class path contains multiple SLF4J bindings
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Found 4 items
drwxr-xr-x - hdfs hdfsadmingroup
                                                   0 2022-01-20 17:32 /apps
 rwxrwxrwt - hdfs hdfsadmingroup
                                                   0 2022-01-20 17:34 /tmp
             - hdfs hdfsadmingroup
                                                   0 2022-01-20 17:32 /user
 rwxr-xr-x
drwxr-xr-x - hdfs hdfsadmingroup
                                                   0 2022-01-20 17:32 /var
[hadoop@ip-172-31-49-62 ~]$
```

The above command display four directory in the first column first letter is d that indicate directory there is no file .

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 fs -ls /user

```
@ip-172-31-49-62 ~]$ hadd
SLF4J: Class path contains multiple SLF4J bindings
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an_explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
 Found 6 items

    hadoop hdfsadmingroup

                                                            0 2022-01-20 17:32 /user/hadoop
drwxrwxrwx
                                                            0 2022-01-20 17:32 /user/history
drwxr-xr-x
                 - mapred mapred
 drwxrwxrwx
                 - hdfs hdfsadmingroup
                                                            0 2022-01-20 17:32 /user/hive
                                                           0 2022-01-20 17:32 /user/hue
drwxrwxrwx
                - hue
                            hue
                - oozie oozie
                                                           0 2022-01-20 17:34 /user/oozie
drwxrwxrwx
                            hdfsadmingroup
 rwxrwxrwx
                 - root
                                                            0 2022-01-20 17:32 /user/root
[hadoop@ip-172-31-49-62 ~]$
```

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

hadoop fs -mkdir /user/csp554

```
[hadoop@ip-172-31-49-62 ~]$ hadoop fs -mkdir /user/csp554

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[hadoop@ip-172-31-49-62 ~]$
```

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

/user/csp554-2

hadoop fs -mkdir /user/csp554-2

```
[hadoop@ip-172-31-49-62 ~]$ hadoop fs -mkdir /user/csp554-2
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j1
2-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf
4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for a
n explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactor
y]
[hadoop@ip-172-31-49-62 ~]$
```

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 fs -copyFromLocal Shraddha.txt /user/csp554

```
[hadoop@ip-172-31-49-62 ~]$ hadoop fs -copyFromLocal Shraddha.txt /user/csp554

SLF4]: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

[hadoop@ip-172-31-49-62 ~]$
```

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 fs -cp /user/csp554/Shraddha.txt /user/csp554-2/

```
[hadoop@ip-172-31-49-62 ~]$ hadoop fs -cp /user/csp554/Shraddha.txt /user/csp554-2/
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
[hadoop@ip-172-31-49-62 ~]$
```

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.

```
aws s3 cp s3://a20499171-csp554/A20499171.txt /home/hadoop/
```

After you executed the above command perform an "Is /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-172-31-49-62 ~]$ aws s3 cp s3://a20499171-csp554/A20499171.txt /home/hadoop/download: s3://a20499171-csp554/A20499171.txt to ./A20499171.txt [hadoop@ip-172-31-49-62 ~]$ ls A20499171.txt shraddha.txt [hadoop@ip-172-31-49-62 ~]$
```

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:"

```
hadoop fs -cp s3://a20499171-csp554/A20499171.txt hdfs:///user/csp554-2
```

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
```

hadoop fs -1s /user/csp554-2

```
[hadoop@ip-172-31-49-62 ~]$ hadoop fs -ls /user/csp554-2
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.cla
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmingroup 21 2022-01-20 18:37 /user/csp554-2/A20499171.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 23 2022-01-20 18:22 /user/csp554-2/Shraddha.txt
[hadoop@ip-172-31-49-62 ~]$
```

17. (2 points) 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/A20499171.txt

```
[hadoop@ip-172-31-49-62 ~]$ hadoop fs -cat /user/csp554-2/A20499171.txt

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.cla

ss]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

this is the id file
```

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

hadoop fs -rm /user/csp554-2/A20499171.txt

hadoop fs -ls /user/csp554-2/

```
[hadoop@ip-172-31-49-62 ~]$ hadoop fs -rm /user/csp554-2/A20499171.txt

SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/usr/lib/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/share/aws/emr/emrfs/lib/slf4j-log4j12-1.7.12.jar!/org/slf4j/impl/StaticLoggerBinder.cla

ss]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]

Deleted /user/csp554-2/A20499171.txt
```

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.