Name: Deep Pawar(A20545137)
Professor: Joseph Rosen
Institute: Illinois Institute of Technology

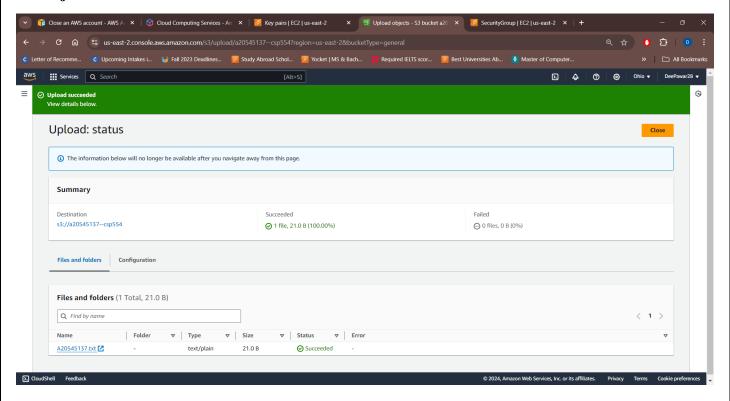# CSP 554: Big Data Technologies

Fall 2024 - Assignment 2

- **Questions and Answers:**

e. Create an S3 bucket of whatever name you choose. For purposes of this discussion I will assume it is called mybucket (but of course your bucket must have a name unique across all of AWS)

f. Using the techniques we learned during assignment #1 upload A12345678.txt to the S3 bucket into an object called A12345678.txt (of course use your id). Note throughout the rest of this document we will still call this object by the name "myid.txt".

Ans:

- **Bucket Name:** a20545137—csp554

- **Object Name:** A20545137.txt

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.

Ans:

- **Command Used:** hadoop fs -ls /

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.

Ans:

- **Command Used:** hadoop fs -ls /user

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

/user/csp554

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

Ans:

- **Command Used:**
  i. **To create HDFS directory:** hadoop fs -mkdir /user/csp554
  ii. **To list the files and directories:** hadoop fs -ls /user

```
hadoop@ip-172-31-24-248:~                                          □   ×
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$ hadoop fs -mkdir /user/csp554
[hadoop@ip-172-31-24-248 ~]$ hadoop fs -ls /user
Found 7 items
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-09-09 18:20 /user/csp554
drwxrwxrwx   - hadoop hdfsadmingroup          0 2024-09-09 17:18 /user/hadoop
drwxr-xr-x   - mapred mapred                  0 2024-09-09 17:18 /user/history
drwxrwxrwx   - hdfs   hdfsadmingroup          0 2024-09-09 17:18 /user/hive
drwxrwxrwx   - hue    hue                     0 2024-09-09 17:18 /user/hue
drwxrwxrwx   - oozie  oozie                   0 2024-09-09 17:20 /user/oozie
drwxrwxrwx   - root   hdfsadmingroup          0 2024-09-09 17:18 /user/root
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
```

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

/user/csp554-2

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

Ans:

- **Command Used:**
  i. **To create HDFS directory:** hadoop fs -mkdir /user/csp554-2
  ii. **To list the files and directories:** hadoop fs -ls /user

```
hadoop@ip-172-31-24-248:~                                          □   ×
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$ hadoop fs -mkdir /user/csp554-2
[hadoop@ip-172-31-24-248 ~]$ hadoop fs -ls /user
Found 8 items
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-09-09 18:20 /user/csp554
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-09-09 18:22 /user/csp554-2
drwxrwxrwx   - hadoop hdfsadmingroup          0 2024-09-09 17:18 /user/hadoop
drwxr-xr-x   - mapred mapred                  0 2024-09-09 17:18 /user/history
drwxrwxrwx   - hdfs   hdfsadmingroup          0 2024-09-09 17:18 /user/hive
drwxrwxrwx   - hue    hue                     0 2024-09-09 17:18 /user/hue
drwxrwxrwx   - oozie  oozie                   0 2024-09-09 17:20 /user/oozie
drwxrwxrwx   - root   hdfsadmingroup          0 2024-09-09 17:18 /user/root
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
[hadoop@ip-172-31-24-248 ~]$
```

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.

Ans:

- **Command Used:**
    - **i.   To copy local file to given hdfs directory:** hadoop fs -copyFromLocal /home/Hadoop/deeppawar.txt /user/csp554
    - **ii. To list the files and directories:** hadoop fs -ls /user/csp554/

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.

Ans:

- **Command Used:**
    - **i.   To copy a file from one hdfs directory to another hdfs directory:** hadoop fs -cp /user/csp554/deeppawar.txt /user/csp554-2
    - **ii. To list the files and directories:** hadoop fs -ls /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.

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.

Ans:

- **Command Used:**
    i. **To copy an object from an S3 bucket to the Linux file system of the Hadoop master node:** aws s3 cp s3://a20545137--csp554/A20545137.txt /home/hadoop/A20545137.txt

    ii. **To list the files and directories:** ls /home/hadoop/

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.

Ans:

- **Command Used:**
  - **i. To copy the same object myid.txt you created in an S3 bucket into HDFS into the directory /users/csp554:** hadoop fs -cp s3://a20545137—csp554/A20545137.txt hdfs:///user/csp554-2/
  - **ii. To list the files and directories:** hadoop fs -ls /user/csp554-2/

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.

Ans:

- **Command Used:**
  - **i. To show the contents of the myid.txt (a20545137.txt) file:** hadoop fs -cat /user/csp554-2/A20545137.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.

Ans:

- **Command Used:**
    - i. **To remove the myid.txt (A20545137.txt) file in the hdfs directory:** hadoop fs -rm /user/csp554-2/A20545137.txt

    - ii. **To list the files and directories:** hadoop fs -ls /user/csp554-2/