HDFS – CLI

1	**View the `hdfs dfs` Command**
1.1	Enter the following command to view the usage of `hdfs dfs`: **`$ hdfs dfs`**
1.2	Notice that the usage contains options for performing filesystem tasks in HDFS, like copying files from a local folder into HDFS, retrieving a file from HDFS, copying and moving files around, and making and removing directories. In this lab, you will perform these commands, and many others, to help you become comfortable with working with HDFS.
2	**Create a Directory in HDFS**
2.1	Enter the following `-ls` command to view the contents of the user's root directory in HDFS, which is `/user/<user_id>`: **`$ hdfs dfs -ls`** If you do not see any files in `/user/<user_id>`, then you do not have any files yet.
2.2	Enter the following command to create a directory named test in HDFS: $ hdfs dfs -mkdir test
2.3	Verify that the folder was created successfully: $ hdfs dfs -ls

2.4	Create a couple of subdirectories for test:
	$ hdfs dfs -mkdir test/test1
	$ hdfs dfs -mkdir -p test/test2/test3
	Notice how the -p command can be used to create multiple directories.

2.5	Use the -ls command to view the contents
	$ hdfs dfs -ls
	Notice you only see the test directory. To recursively view the contents of a folder, use -ls -R:
	$ hdfs dfs -ls -R
	The output should look like:
	test
	test/test1
	test/test2
	test/test2/test3

3	Delete a Directory
3.1	Delete the test2 folder (and recursively, its sub contents) using the -rm -R command: $ hdfs dfs -rm -R test/test2
3.2	Now run the -ls -R command: $ hdfs dfs -ls -R The directory structure of the output should look like: .Trash .Trash/Current .Trash/Current/user .Trash/Current/user/id .Trash/Current/user/id/test .Trash/Current/user/id/test/test2 .Trash/Current/user/id/test/test2/test3 test test/test1
4	Upload a File to HDFS
4.1	Now let’s put a file into the test folder. Change directories to ~/labs/Lab1: $ cd ~/labs/Lab1/

4.2	Notice this folder contains a file named data.txt: $ tail data.txt
4.3	Run the following -put command to copy data.txt into the test folder in HDFS: $ hdfs dfs -put data.txt test/
4.4	Verify that the file is in HDFS by listing the contents of test: $ hdfs dfs -ls test The output should look like the following: Found 2 items test/data.txt test/test1
5	Copy a File in HDFS
5.1	Now copy the data.txt file in test to another folder in HDFS using the -cp command: $ hdfs dfs -cp test/data.txt test/test1/data2.txt
5.2	Verify that the file is in both places by using the -ls -R command on test. The output should look like the following: $ hdfs dfs -ls -R test test/data.txt test/test1 test/test1/data2.txt
5.3	Now delete the data2.txt file using the -rm command: $ hdfs dfs -rm test/test1/data2.txt
5.4	Verify that the data2.txt file is in the .Trash folder.

6	View the Contents of a File in HDFS
6.1	You can use the -cat command to view text files in HDFS. Enter the following command to view the contents of data.txt: $ hdfs dfs -cat test/data.txt
6.2	You can also use the -tail command to view the end of a file: $ hdfs dfs -tail test/data.txt Notice the output this time is only the last 20 rows of data.txt.
7	Getting a File from HDFS
7.1	See if you can figure out how to use the get command to copy test/data.txt from HDFS into your local folder. Answer: hdfs dfs -get test/data.txt ~/ cd ~/ ls
8	The getmerge Command
8.1	Put the file ~/demos/small_blocks.txt into the test folder in HDFS. You should now have two files in test: data.txt and small_blocks.txt. Answer: hdfs dfs -put ~/demos/small_blocks.txt test/
8.2	Run the following getmerge command: $ hdfs dfs -getmerge test ~/merged.txt
8.3	What did the previous command do? Did you open the file merged.txt to see what happened? Answer: The two files that were in the test folder in HDFS were merged into a single file and stored on the local file system.
9	Specify the Block Size and Replication Factor

9.1	Put ~/labs/Lab1/data.txt into /user/<your_id> in HDFS, giving it a blocksize of 1,048,576 bytes. Hint: The blocksize is defined using the dfs.blocksize property on the command line.
9.2	Run the following fsck command on data.txt: $ hdfs fsck /user/<your_id>/data.txt
9.3	How many blocks are there for this file? Answer: The file should be broken down into two blocks.
	Result: You should now be comfortable with executing the various HDFS commands, including creating directories, putting files into HDFS, copying files out of HDFS, and deleting files and folders.