

## In-class Lab 1 – HDFS and Linux Commands

1. Create a folder in a directory in your Linux operating system and name it with your first name.

SSH-in-browser



```
Linux biodata-m 5.10.0-0.bpo.15-amd64 #1 SMP Debian 5.10.120-1-bpo10+1 (2022-06-13) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Wed Sep 21 23:07:47 2022 from 35.235.244.33
goodnewsagbadu@biodata-m:~$ pwd
/home/goodnewsagbadu
goodnewsagbadu@biodata-m:~$ mkdir Goodnews
goodnewsagbadu@biodata-m:~$ pwd
/home/goodnewsagbadu
goodnewsagbadu@biodata-m:~$ ls
Goodnews
goodnewsagbadu@biodata-m:~$ ls -lh
total 4.0K
drwxr-xr-x 2 goodnewsagbadu goodnewsagbadu 4.0K Sep 22 13:04 Goodnews
```

2. Create a text file called yourname.txt in the folder you created in Step 1 and place some text in it using any text editor tool.

```
goodnewsagbadu@biodata-m:~$ pwd
/home/goodnewsagbadu
goodnewsagbadu@biodata-m:~$ ls
Goodnews
goodnewsagbadu@biodata-m:~$ ls -lh
total 4.0K
drwxr-xr-x 2 goodnewsagbadu goodnewsagbadu 4.0K Sep 22 13:04 Goodnews
goodnewsagbadu@biodata-m:~$ vi Goodnews.txt
goodnewsagbadu@biodata-m:~$ cat Goodnews.txt
Create a text file called yourname.txt in the folder you created in step 1 and place some text in it using any text editor tool.
goodnewsagbadu@biodata-m:~$
```

3. Create a folder in HDFS under the root directory called BigDataFirstName where “FirstName” is your name.

```
goodnewsagbadu@biodata-m:~$ cd /
goodnewsagbadu@biodata-m:/$ hadoop fs -mkdir /BigDataGoodnews
```

```
goodnewsagbadu@biodata-m:/$ hadoop fs -ls /
Found 4 items
drwxr-xr-x - goodnewsagbadu hadoop 0 2022-09-22 13:25 /BigDataGoodnews
drwxrwxrwt - hdfs hadoop 0 2022-09-21 23:08 /tmp
drwxrwxrwt - hdfs hadoop 0 2022-09-21 23:08 /user
drwx-wx-wx - hive hadoop 0 2022-09-21 23:08 /var
```

4. Copy the text file you created in step 2 from the Linux local file system into the HDFS folder BigData that you created in step 3.

```
goodnewsagbadu@biodata-m:/$ hadoop fs -copyFromLocal /home/goodnewsagbadu/Goodnews.txt /BigDataGoodnews
goodnewsagbadu@biodata-m:/$ hadoop fs -ls /BigDataGoodnews
Found 1 items
-rw-r--r-- 1 goodnewsagbadu hadoop      129 2022-09-22 13:37 /BigDataGoodnews/Goodnews.txt
goodnewsagbadu@biodata-m:/$
```

5. In the class exercises, we have most of the HDFS commands but not all of them. Research a new HDFS command and use it in an example. Provide a link for any external sources you use to answer this question (in other words, cite any sources you use).

```
goodnewsagbadu@biodata-m:/$ hadoop fs -ls /
Found 7 items
drwxr-xr-x - goodnewsagbadu hadoop      0 2022-09-22 13:37 /BigDataGoodnews
-rw-r--r-- 1 goodnewsagbadu hadoop    129 2022-09-22 13:34 /BigDataGoodnews
-rw-r--r-- 1 goodnewsagbadu hadoop      0 2022-09-22 16:20 /new.txt
drwxr-xr-x - goodnewsagbadu hadoop      0 2022-09-22 15:59 /newcmd
drwxrwxrwt - hdfs hadoop      0 2022-09-21 23:08 /tmp
drwxrwxrwt - hdfs hadoop      0 2022-09-21 23:08 /user
drwx-wx-wx - hive hadoop      0 2022-09-21 23:08 /var
goodnewsagbadu@biodata-m:/$ hadoop fs -appendToFile /BigDataGoodnews/Goodnews.txt
appendToFile: missing destination argument
goodnewsagbadu@biodata-m:/$ hadoop fs -appendToFile - /BigDataGoodnews/Goodnews.txt
I am appending this line of text to goodnews.txt
goodnewsagbadu@biodata-m:/$ hadoop fs -cat /BigDataGoodnews/Goodnews.txt
Create a text file called yourname.txt in the folder you created in step 1 and place some text in it using any text editor tool.
I am appending this line of text to goodnews.txt
goodnewsagbadu@biodata-m:/$
```

The screenshot above shows how to append text to already created .txt file using appendToFile. This can be handy when you want to add some information to a file.

I also used CAT to show the contents of Goodnews.txt in the terminal.

6. The class exercises use “hadoop fs” as the start to HDFS commands but we can also use “hdfs dfs”. Research on the web what is the difference between these two commands. Provide a link for any external sources you use to answer this question (in other words, cite any sources you use).

A generic file system called FS refers to any file system, including local, HDFS, and others. This can therefore be utilized while working with many file systems, including Local FS, (S)FTP, S3, and others. While DFS is incredibly unique to HDFS. would be appropriate for HDFS functioning.

7. We went over the most important Linux commands in class but there are many more. Try to find a Linux command that we did not cover in class but you find useful. Describe the command and use it in an example and take a screenshot showing the usage.

```
goodnewsagbadu@biodata-m:~$ ls -lh
total 20K
drwxr-xr-x 2 goodnewsagbadu goodnewsagbadu 4.0K Sep 22 13:04 Goodnews
-rw-r--r-- 1 goodnewsagbadu goodnewsagbadu 129 Sep 22 13:08 Goodnews.txt
-rw-r--r-- 1 goodnewsagbadu goodnewsagbadu 24 Sep 22 17:39 file1.txt
-rw-r--r-- 1 goodnewsagbadu goodnewsagbadu 23 Sep 22 17:40 file2.csv
-rw-r--r-- 1 goodnewsagbadu goodnewsagbadu 20 Sep 22 17:40 file2.txt
goodnewsagbadu@biodata-m:~$ grep goodnews *.txt
file1.txt:goodnews is a good boy
file2.txt:my name is goodnews
goodnewsagbadu@biodata-m:~$
```

The grep tool looks for lines that include a search pattern. The file contents can also be searched using the grep command. Here, I'm looking for the word "goodnews" anywhere in the current directory's text files. This will be handy when you have a large number of files, searching for the keyword can help locate the particular file.

**Reference:**

Team, D. F. (2021, August 25). *Top 10 hadoop HDFS commands with examples and usage*. DataFlair. Retrieved September 22, 2022, from <https://data-flair.training/blogs/top-hadoop-hdfs-commands-tutorial/>

RajRaj 51744 gold badges99 silver badges1818 bronze badges, PradeepKumbharPradeepKumbhar 3, & MrElephantMrElephant 25044 silver badges2323 bronze badges.*How to update a file in HDFS*. Stack Overflow. Retrieved September 22, 2022, from <https://stackoverflow.com/questions/39130092/how-to-update-a-file-in-hdfs>.

Charlie LinCharlie Lin 1. (2020, December 1). *What's the difference between "Hadoop fs" shell commands and "Hdfs dfs" shell commands?* Stack Overflow. Retrieved September 22, 2022, from <https://stackoverflow.com/questions/18142960/whats-the-difference-between-hadoop-fs-shell-commands-and-hdfs-dfs-shell-co>.

McKay, D. (2019, May 8). *37 important linux commands you should know*. How. Retrieved September 22, 2022, from <https://www.howtogeek.com/412055/37-important-linux-commands-you-should-know/>