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

franciszero@cluster-b275-m:~$ mkdir francis

A picture containing text, font, screenshot

Description automatically generated

1. 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.

A picture containing text, screenshot, font

Description automatically generated

A screenshot of a computer

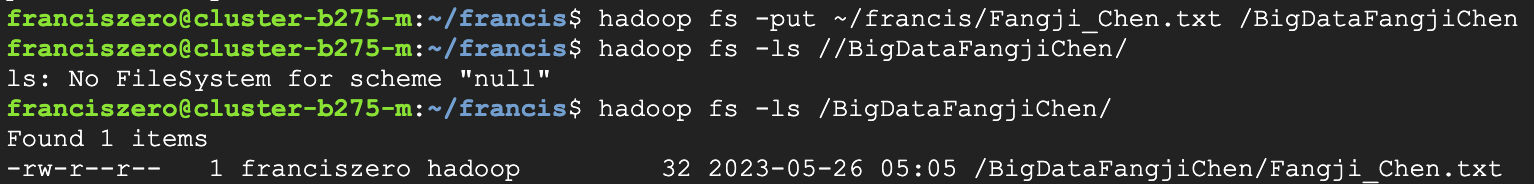
Description automatically generated with low confidence

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

A screenshot of a computer

Description automatically generated with medium confidence

1. 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.



1. 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).

“df”: Displays free space.

“du”: Displays sizes of files and directories contained in the given directory or the length of a file in case its just a file.

“count”: Count the number of directories, files and bytes under the paths that match the specified file pattern.

An excess of small files generated by certain users is putting pressure on the namenode. Using “du”, “dh” and “count” as a combination is an effective strategy for continuously monitoring the health status of system storage, especially for locating folders/users that generate too much small files.

A screenshot of a computer program

Description automatically generated with medium confidence

1. 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).

Based on the explaination in the website: <https://www.quora.com/Is-there-any-difference-between-HDFS-dfs-and-Hadoop-fs-shell-commands>, hadoop fs is more “generic” command that allows you to interact with multiple file systems including Hadoop, whereas hdfs dfs is the command that is specific to HDFS. Now hadoop fs is deprecated.

A diagram of a diagram

Description automatically generated with low confidence

Figure: <https://intellipaat.com/community/42860/hadoop-fs-or-hdfs-dfs-and-whats-the-difference>

1. 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 it useful. Describe the command and use it in an example and take a screenshot showing the usage.

Using xargs to read searching results from Stdin of *find* command is an effective choice. Similar commands like “rm”, “chmod” are always used with find.

*A black background with white text

Description automatically generated with low confidence*