Faculty: Ms. Sana Shaikh **Hadoop Commands** **Subject: Big Data Analytics**
2020-2021

To start Hadoop
hdfs namenode -format
start-dfs.sh
start-yarn.sh
jps

http://localhost:50070/

stop-all.sh

Basic file commands

Hadoop file commands take the form of:

hadoop fs -cmd <args>

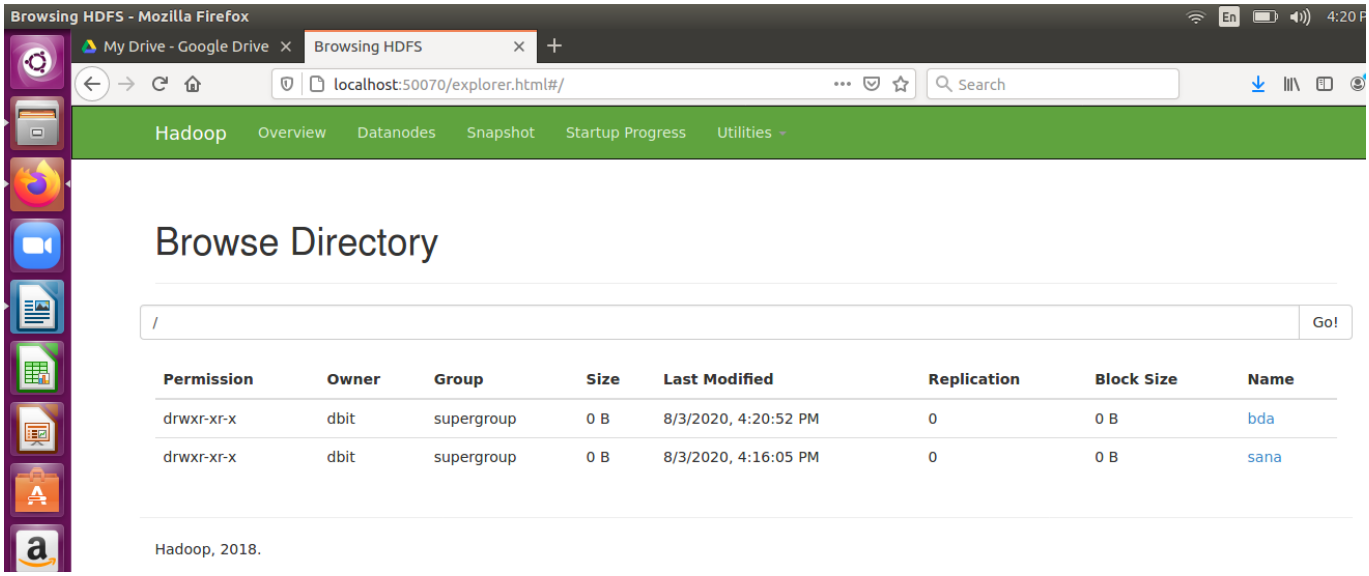
where cmd is the specific file command and <args> is a variable number of arguments.

1. mkdir

hadoop fs -mkdir /sana

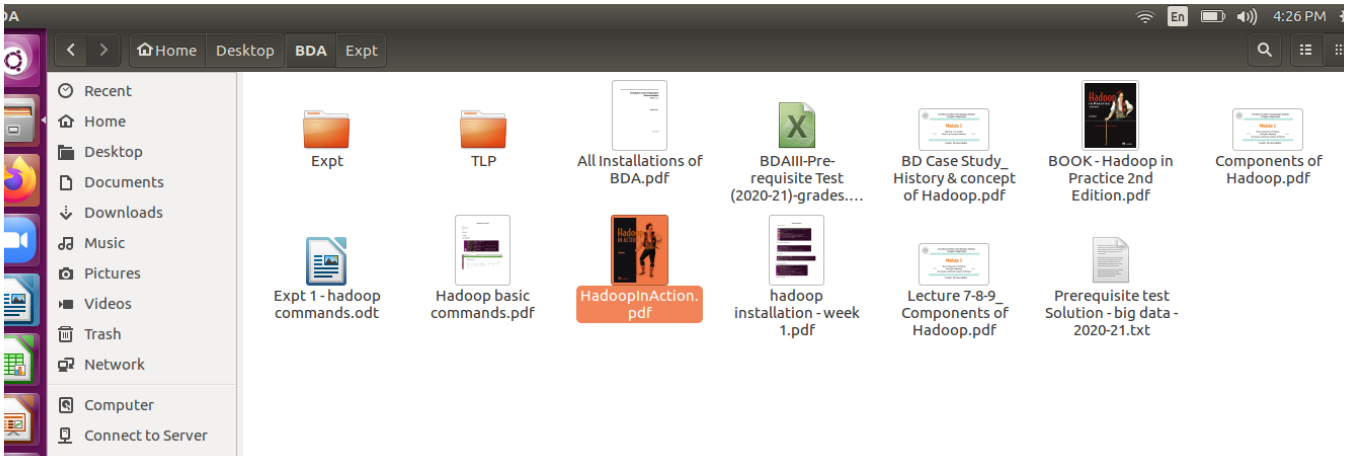
hadoop fs -mkdir /sana .....(repeat the command...message will come...mkdir: `/sana': File exists)

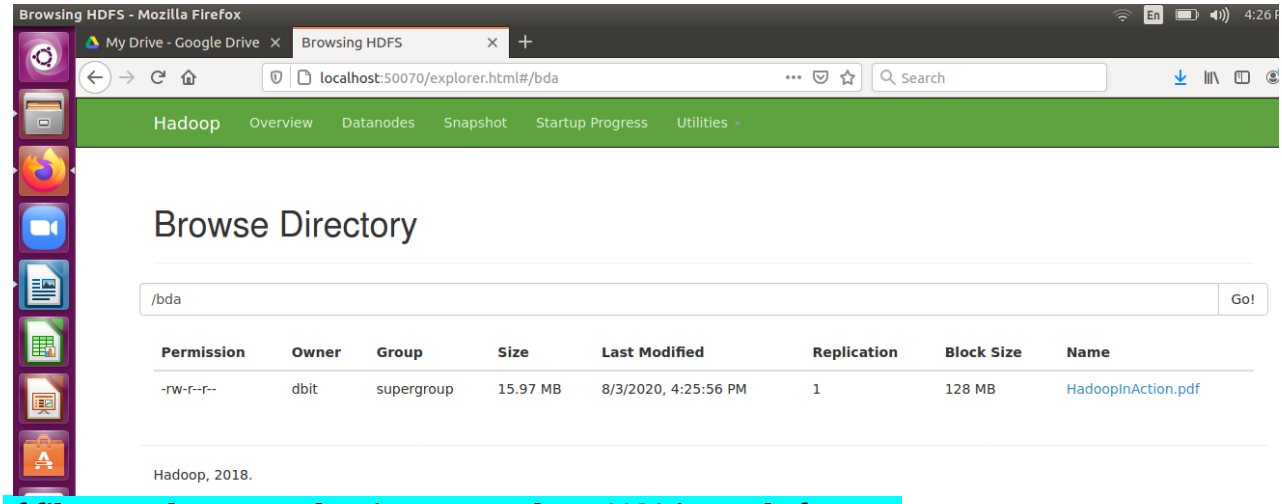
hadoop fs -mkdir /bda



2. put: This command is used to copy files from the local file system to the HDFS filesystem. This command is similar to **–copyFromLocal** command

hadoop fs -put Desktop/BDA/HadoopInAction.pdf /bda/





If file name has space then in command use %20 instead of space:

Example:

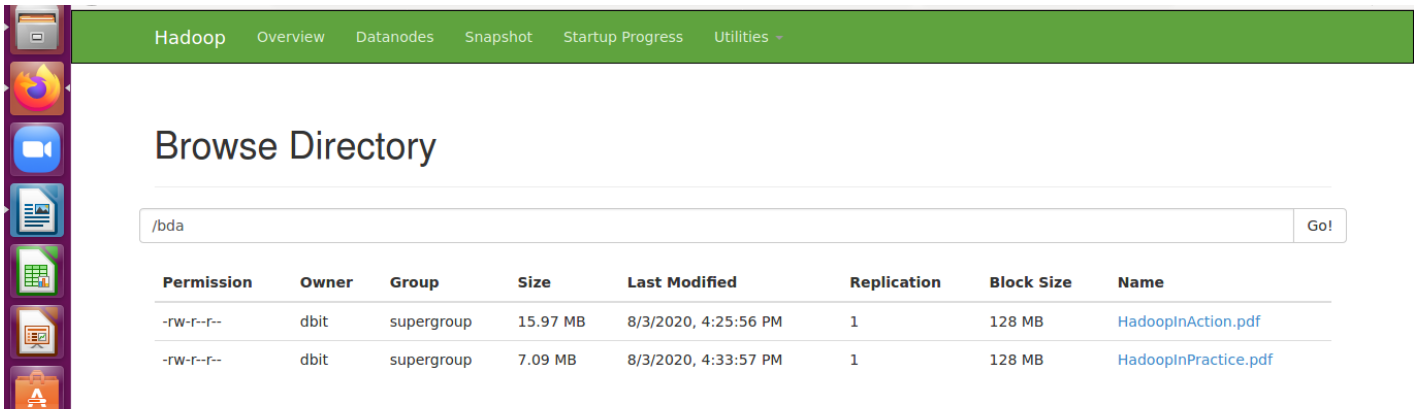
hadoop fs -put /home/dbit/Desktop/BDA/Hadoop%20basic%20commands.pdf /bda/

hadoop fs -put /home/dbit/Desktop/BDA/hadoop%20installation%20-%20week%201.pdf /bda/

hadoop fs -put /home/dbit/Desktop/BDA/map_reduce_tutorial.pdf /bda/

3. (or moveFromLocal)

hadoop fs -moveFromLocal Desktop/BDA/HadoopInPractice.pdf /bda

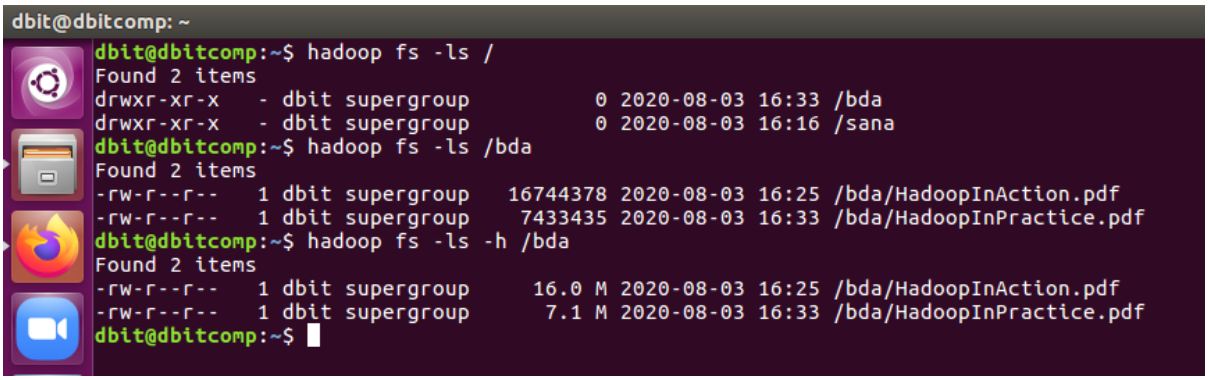


4. ls: it is used for listing the directories present under a specific directory in an HDFS system

options:

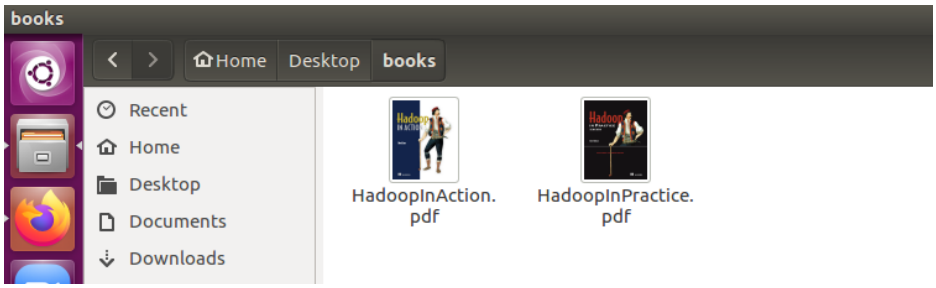
–d	The option is used to list the directories as plain files
–h	The option is used to format the sizes of files into a human-readable manner than just number of bytes
–R	The option is used to recursively list the contents of directories

hadoop fs -ls /
hadoop fs -ls /bda
hadoop fs -ls -h /bda



5. get:
This command is used to copy files from the HDFS file system to the local file system, just the opposite to put command.

hadoop fs -get /bda Desktop/books

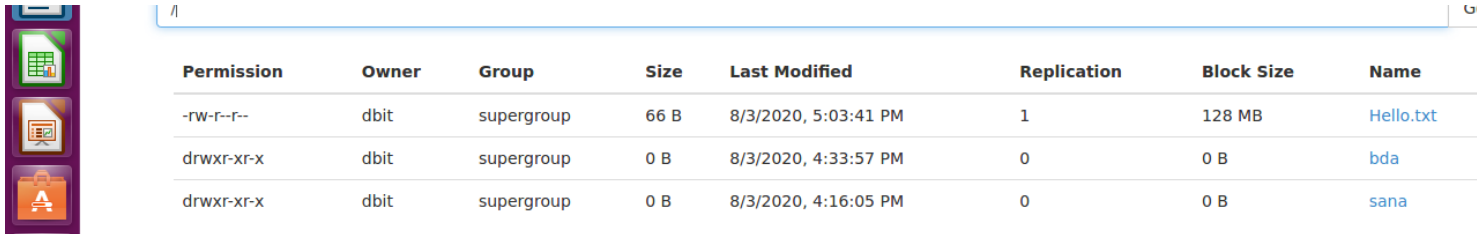


6. cat:

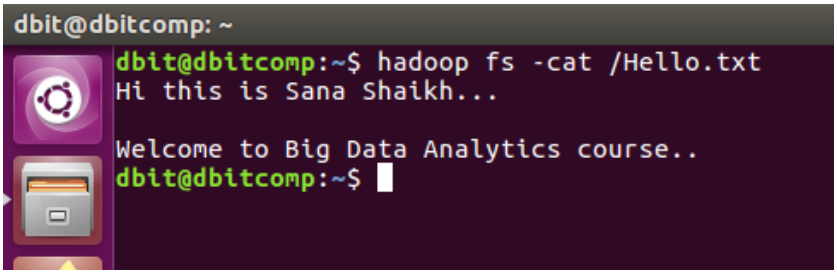
This command is used for displaying the contents of a file on the console.

```
hadoop fs -cat /hello.txt
```

```
hadoop fs -cat /bda/HadoopInAction.pdf
```

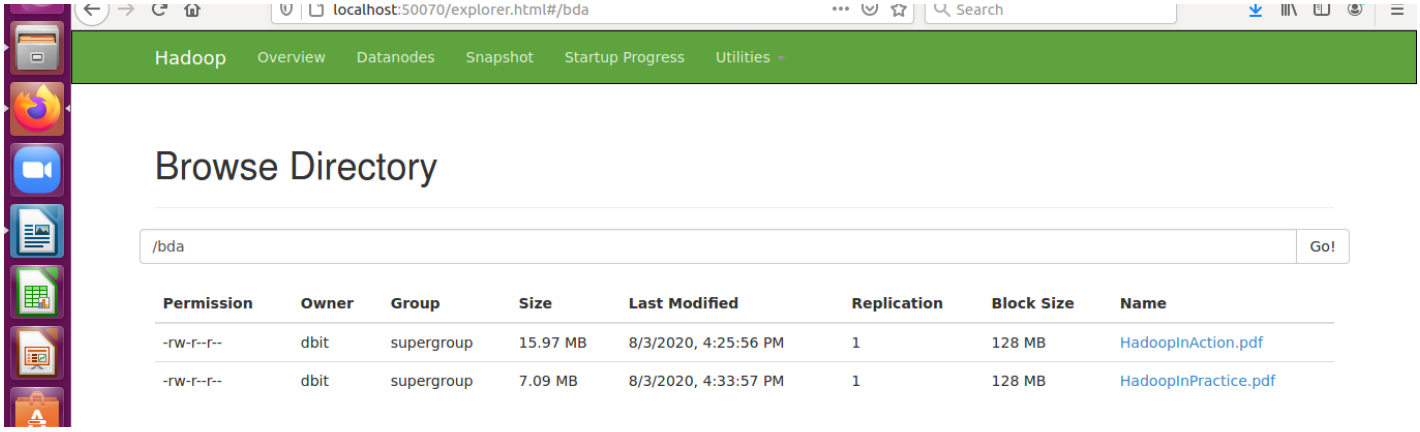


Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	dbit	supergroup	66 B	8/3/2020, 5:03:41 PM	1	128 MB	Hello.txt
drwxr-xr-x	dbit	supergroup	0 B	8/3/2020, 4:33:57 PM	0	0 B	bda
drwxr-xr-x	dbit	supergroup	0 B	8/3/2020, 4:16:05 PM	0	0 B	sana

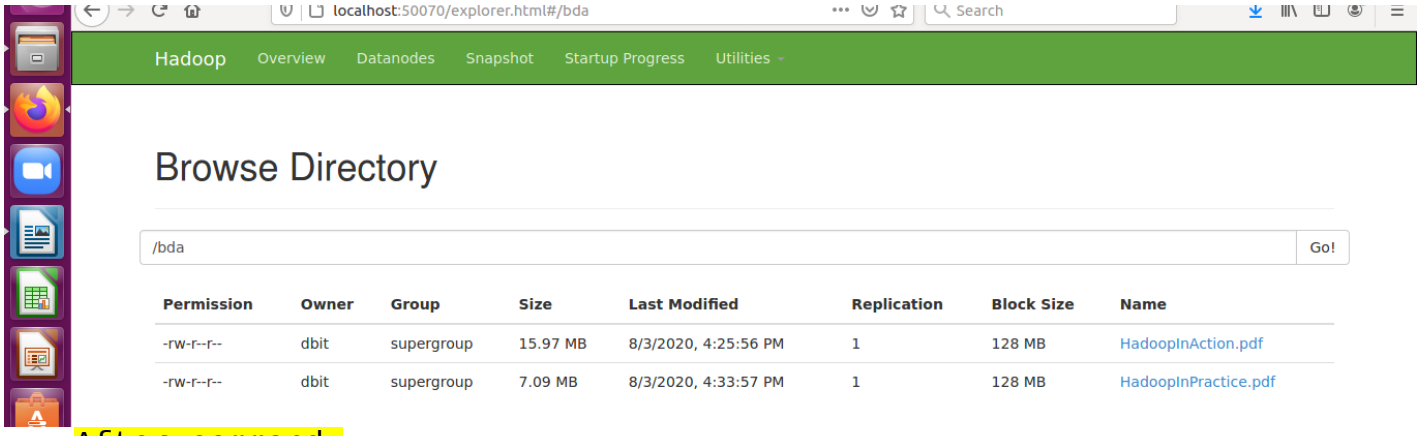


7. cp: This command is used for copying files from one directory to another directory within the HDFS file system. `Sample: hadoop fs -cp /sana /sana1`

`Before command:`

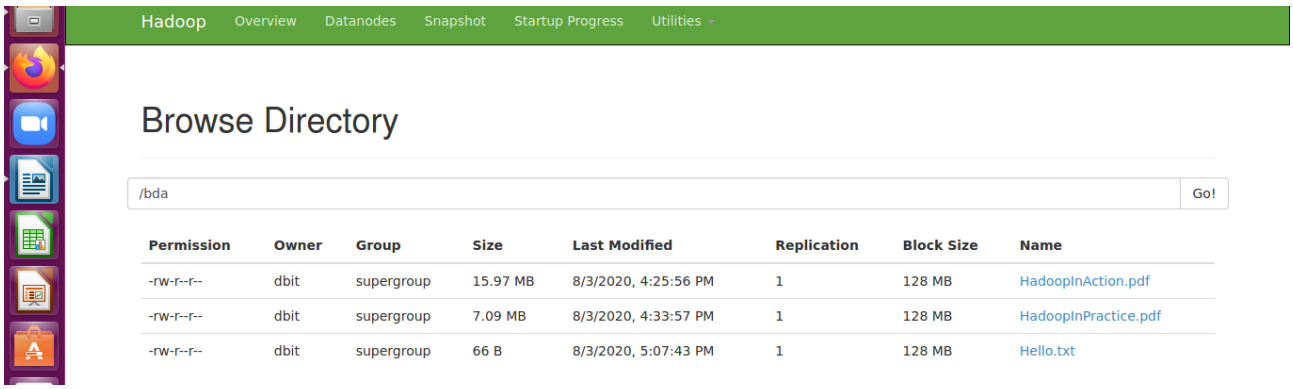


Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	dbit	supergroup	15.97 MB	8/3/2020, 4:25:56 PM	1	128 MB	HadoopInAction.pdf
-rw-r--r--	dbit	supergroup	7.09 MB	8/3/2020, 4:33:57 PM	1	128 MB	HadoopInPractice.pdf



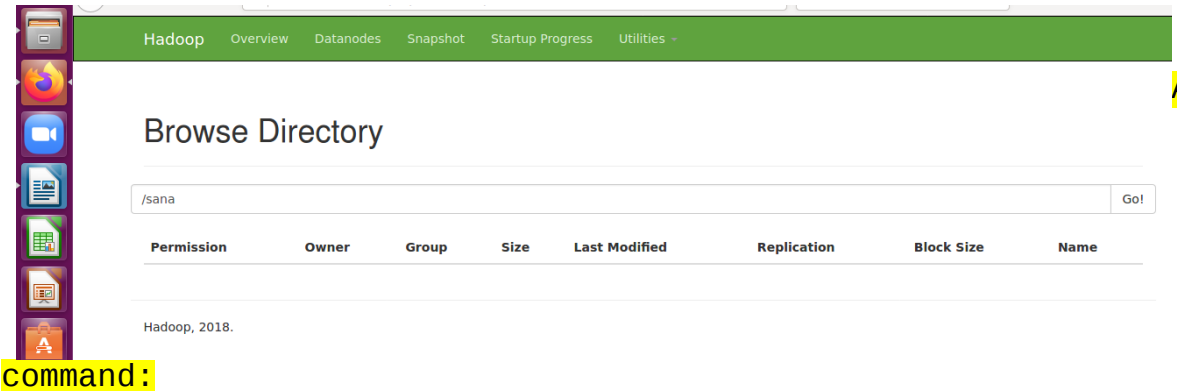
After command:

```
hadoop fs -cp /Hello.txt /bda/
```



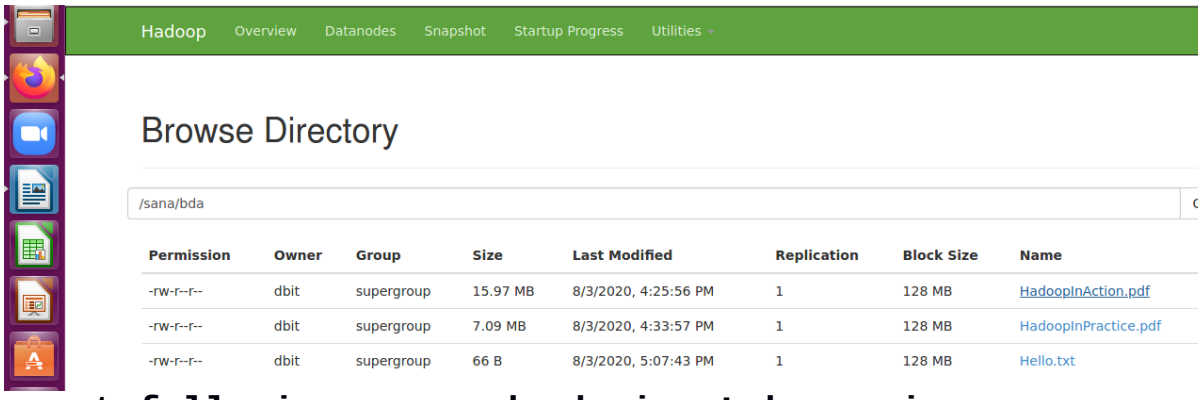
8. mv: This command is used for moving a file from one directory to another directory within the HDFS file system.

Before command:



After command:

```
hadoop fs -mv /bda /sana
```



Try out following commands during Lab session:

rm - it is used for removing a file from the HDFS file system.

rm -r: This command deletes a file from HDFS *recursively*. It is very useful command when you want to delete a *non-empty directory*.

tail - This command is used to show the last 1KB of the file

stat - This command is used to print the statistics about the file/directory in the specified format

appendToFile - This command appends the contents of all the given local files to the provided destination file on the HDFS filesystem.
hadoop fs -appendToFile Desktop/hi1.txt Desktop/hi2.txt Desktop/hi3.txt /hi.txt

setrep - Change replication factor from 1 to 2
hadoop fs -setrep -R 4 /sana (applied on all files)
hadoop fs -setrep -R -w 6 /sana/hello.txt (applied on specific file)

text - takes the source file and outputs the file in the **text** format. It detects the encoding of the file and decodes it to plain **text**.

`df - it` is used to show the capacity, free and used space available on the HDFS filesystem.

`du- it` is used to show the amount of space in bytes that have been used by the files that match the specified file pattern

du -s: This command will give the total size of directory/file.

```
hadoop fs -du -s /
```

`Count-` This command is used to count the number of directories, files, and bytes under the path that matches the provided file pattern.

hadoop -fs count /

touchz: It creates an empty file.