NAME : Meith Navlakha
60004190068
SEM-VI B
BDI Practical

EXPERIMENT 2

AIM : Execution OF HDFS Commands

THEORY:

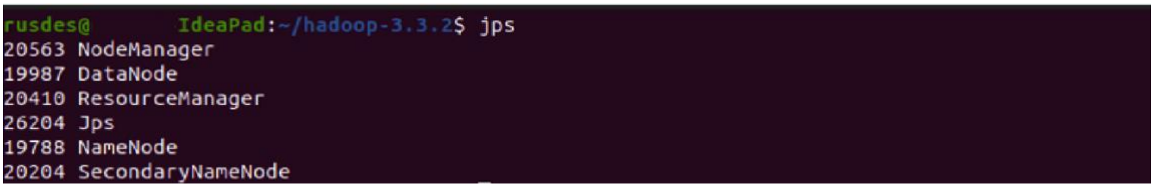
HDFS is the primary or major component of the Hadoop ecosystem which is responsible for storing large data sets of structured or unstructured data across various nodes and thereby maintaining the metadata in the form of log files.

To use the HDFS commands, first you need to start the Hadoop services using the following command:

sbin/start-all.sh

To check the Hadoop services are up and running use the following command:

jps



```
rusdes@        IdeaPad:~/hadoop-3.3.2$ jps
20563 NodeManager
19987 DataNode
20410 ResourceManager
26204 Jps
19788 NameNode
20204 SecondaryNameNode
```

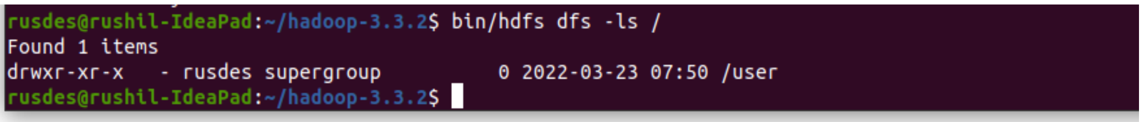
COMMANDS:

1. **ls:** This command is used to list all the files. Use *lsr* for recursive approach. It is useful when we want a hierarchy of a folder.

   Syntax:
   bin/hdfs dfs -ls <path>

   Example:
   bin/hdfs dfs -ls /

   It will print all the directories present in HDFS. bin directory contains executables so, *bin/hdfs* means we want the executables of hdfs particularly *dfs*(Distributed File System) commands.



```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - rusdes supergroup          0 2022-03-23 07:50 /user
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ 
```

2. **mkdir**: To create a directory. In Hadoop *dfs* there is no home directory by default. So let’s first create it.

Syntax:
bin/hdfs dfs -mkdir <folder name>

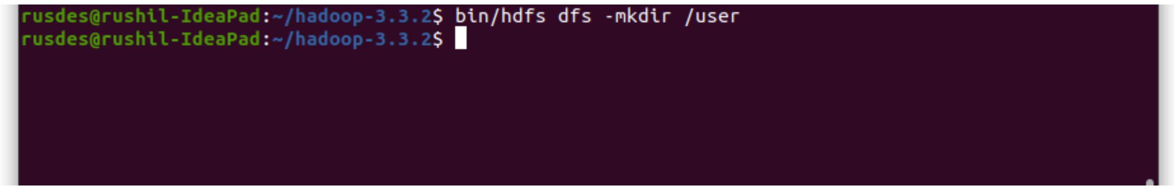
creating home directory:

hdfs/bin -mkdir /user

hdfs/bin -mkdir /user/username -> write the username of your computer

Example:
bin/hdfs dfs -mkdir /meith => '/' means absolute path

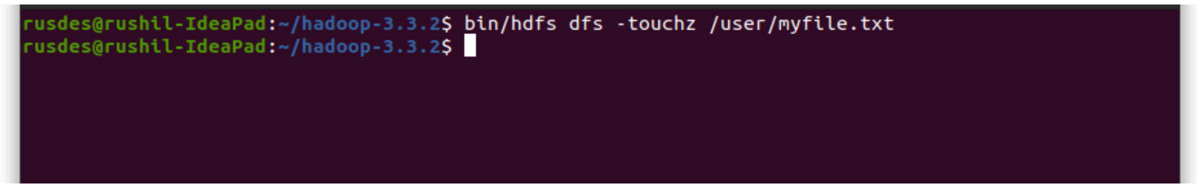
bin/hdfs dfs -mkdir meith2 => Relative path -> the folder will be
created relative to the home directory.



```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -mkdir /user
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ 
```

3. **touchz**: It creates an empty file.
Syntax:
bin/hdfs dfs -touchz <file_path>

Example:
bin/hdfs dfs -touchz /geeks/myfile.txt



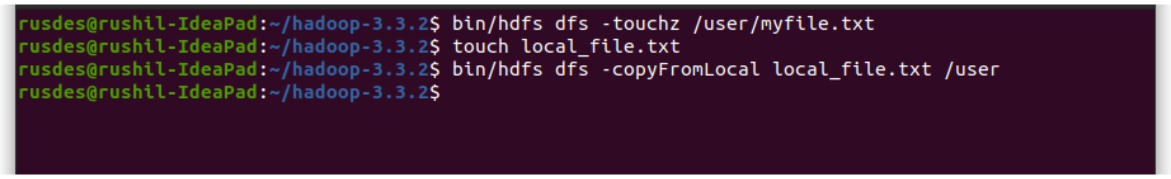
```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -touchz /user/myfile.txt
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ 
```

4. **copyFromLocal (or) put:** To copy files/folders from local file system to hdfs store. This is the most important command. Local filesystem means the files present on the OS.
Syntax:
bin/hdfs dfs -copyFromLocal <local file path> <dest(present on hdfs)>

Example: Let’s suppose we have a file *AI.txt* on Desktop which we want to copy to folder *geeks* present on hdfs.
bin/hdfs dfs -copyFromLocal ../Desktop/AI.txt /geeks

(OR)

bin/hdfs dfs -put ../Desktop/AI.txt /geeks



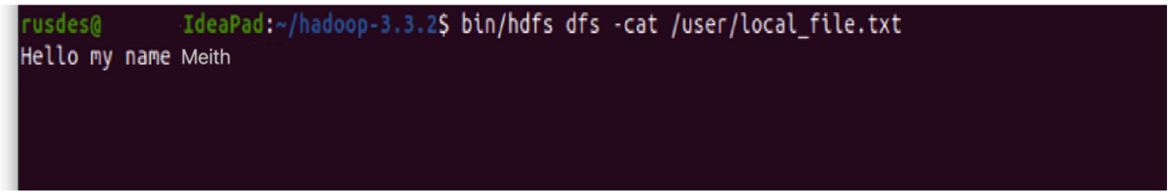
```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -touchz /user/myfile.txt
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ touch local_file.txt
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -copyFromLocal local_file.txt /user
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$
```

5. **cat:** To print file contents.
Syntax:
bin/hdfs dfs -cat <path>

Example:
// print the content of AI.txt present

// inside meith folder.

bin/hdfs dfs -cat /meith/AI.txt ->



```
rusdes@       IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -cat /user/local_file.txt
Hello my name Meith
```

6. **copyToLocal (or) get:** To copy files/folders from hdfs store to local file system.

Syntax:
bin/hdfs dfs -copyToLocal <<srcfile(on hdfs)> <local file dest>

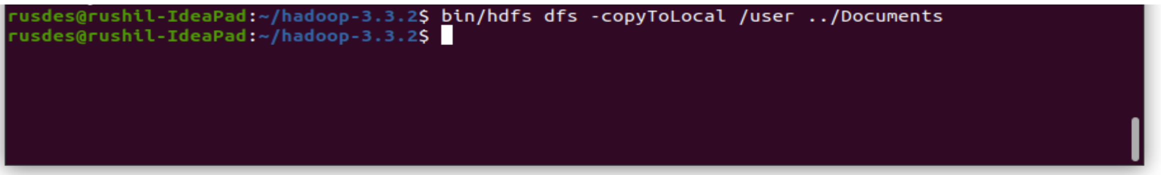
Example:
bin/hdfs dfs -copyToLocal /geeks ../Desktop/hero

(OR)

bin/hdfs dfs -get /meith/myfile.txt ../Desktop/hero

myfile.txt from *meith* folder will be copied to folder *hero* present on *Desktop*.

Note: Observe that we don’t write *bin/hdfs* while checking the things present on local filesystem.

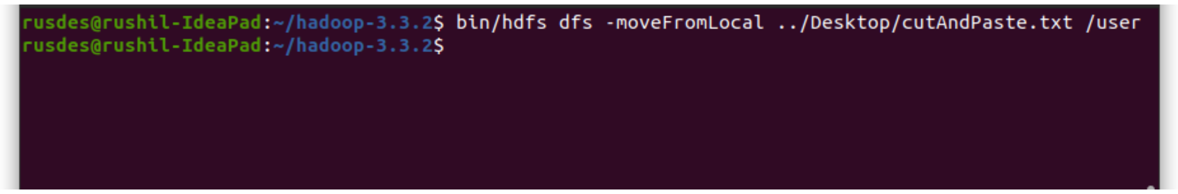


```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -copyToLocal /user ../Documents
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ █
```

7. **moveFromLocal:** This command will move file from local to hdfs.

Syntax:
bin/hdfs dfs -moveFromLocal <local src>   <dest(on hdfs)>

Example:
bin/hdfs dfs -moveFromLocal ../Desktop/cutAndPaste.txt   /geeks

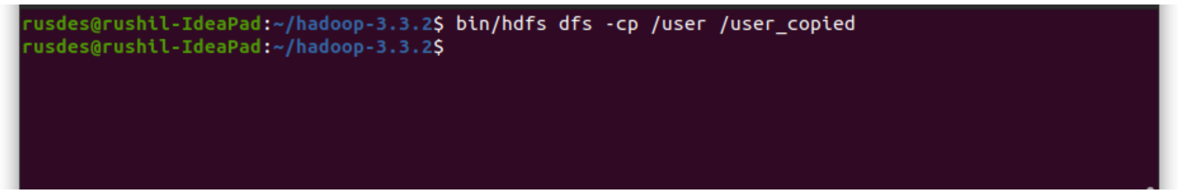


```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -moveFromLocal ../Desktop/cutAndPaste.txt /user
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$
```

8. **cp:** This command is used to copy files within hdfs. Lets copy folder *geeks* to *geeks_copied*.

Syntax:
bin/hdfs dfs -cp <src(on hdfs)> <dest(on hdfs)>

Example:
bin/hdfs -cp /geeks /meith_copied

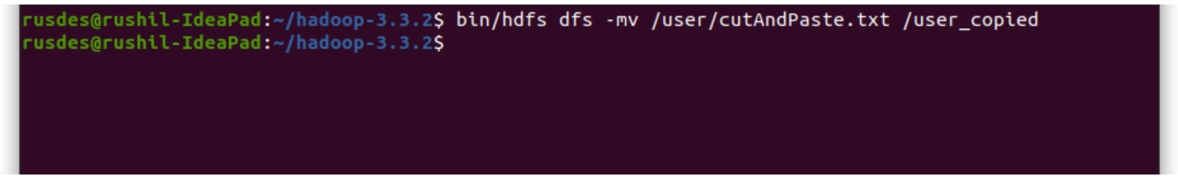


```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -cp /user /user_copied
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$
```

9. **mv:** This command is used to move files within hdfs. Let's cut-paste a file *myfile.txt* from *meith* folder to *geeks_copied*.

Syntax:
bin/hdfs dfs -mv <src(on hdfs)> <src(on hdfs)>

Example:
bin/hdfs -mv /user/cutAndPaste.txt   /user_copied

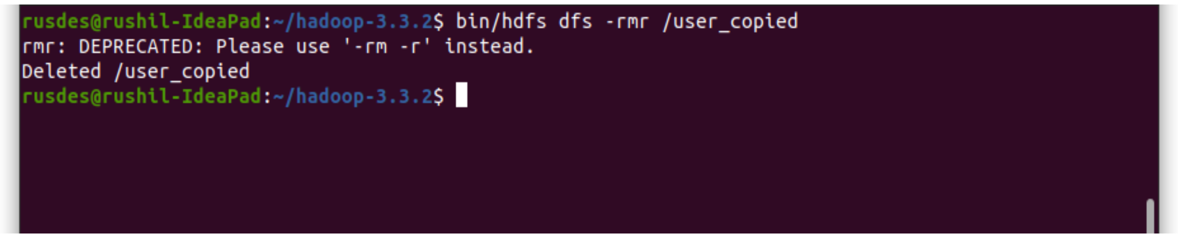


```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -mv /user/cutAndPaste.txt /user_copied
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$
```

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

Syntax:
bin/hdfs dfs -rmr <filename/directoryName>

Example:
bin/hdfs dfs -rmr /geeks_copied -> It will delete all the content inside the directory then the directory itself.

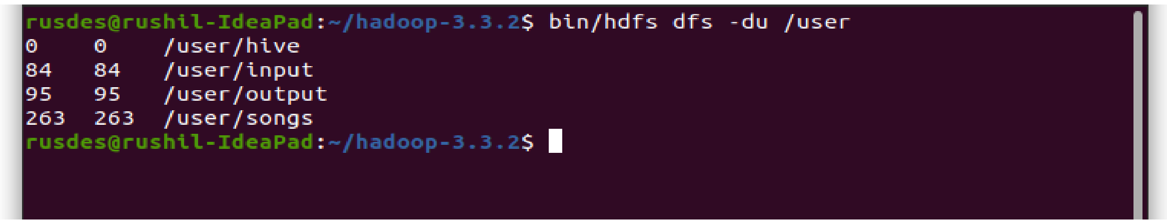


```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -rmr /user_copied
rmr: DEPRECATED: Please use '-rm -r' instead.
Deleted /user_copied
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ 
```

11. **du:** It will give the size of each file in directory.

Syntax:
bin/hdfs dfs -du <dirName>

Example:
bin/hdfs dfs -du /user

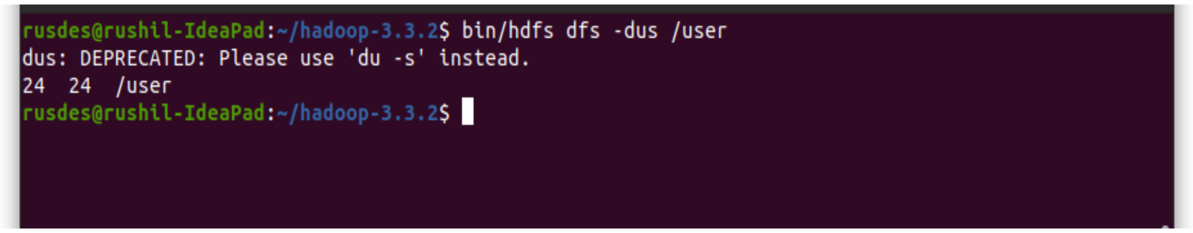


```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -du /user
0     0     /user/hive
84    84    /user/input
95    95    /user/output
263   263   /user/songs
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ 
```

12. **dus:**: This command will give the total size of directory/file.

Syntax:
bin/hdfs dfs -dus <dirName>

Example:
bin/hdfs dfs -dus /user

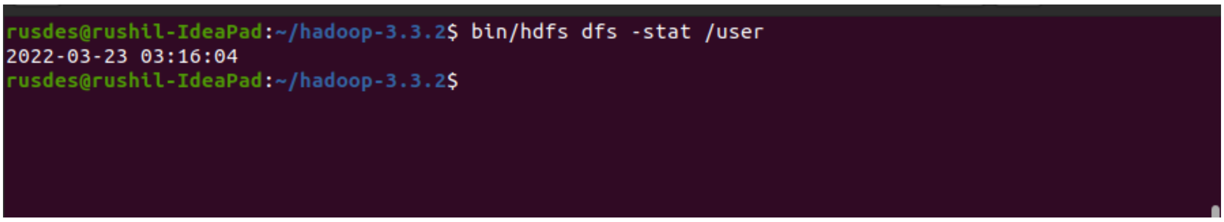


```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -dus /user
dus: DEPRECATED: Please use 'du -s' instead.
24  24  /user
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ 
```

13. **stat:** It will give the last modified time of directory or path. In short it will give stats of the directory or file.

Syntax:
bin/hdfs dfs -stat <hdfs file>

Example:
bin/hdfs dfs -stat /user



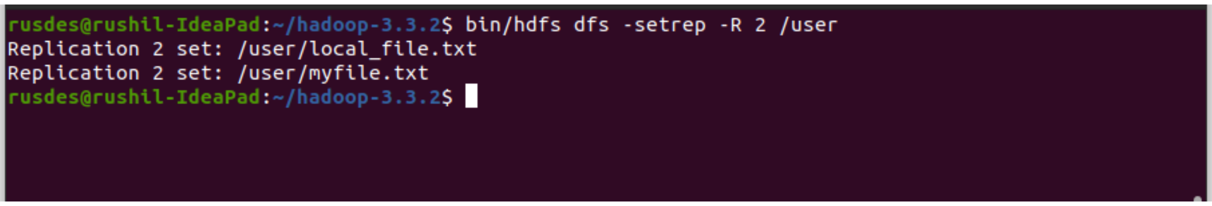
```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -stat /user
2022-03-23 03:16:04
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$
```

14. **setrep:** This command is used to change the replication factor of a file/directory in HDFS. By default it is 3 for anything which is stored in HDFS (as set in hdfs *core-site.xml*).

Example 1: To change the replication factor to 6 for *geeks.txt* stored in HDFS.
bin/hdfs dfs -setrep -R -w 6 local_file.txt

Example 2: To change the replication factor to 4 for a directory *geeksInput* stored in HDFS.
bin/hdfs dfs -setrep -R 4 / local_file.txt

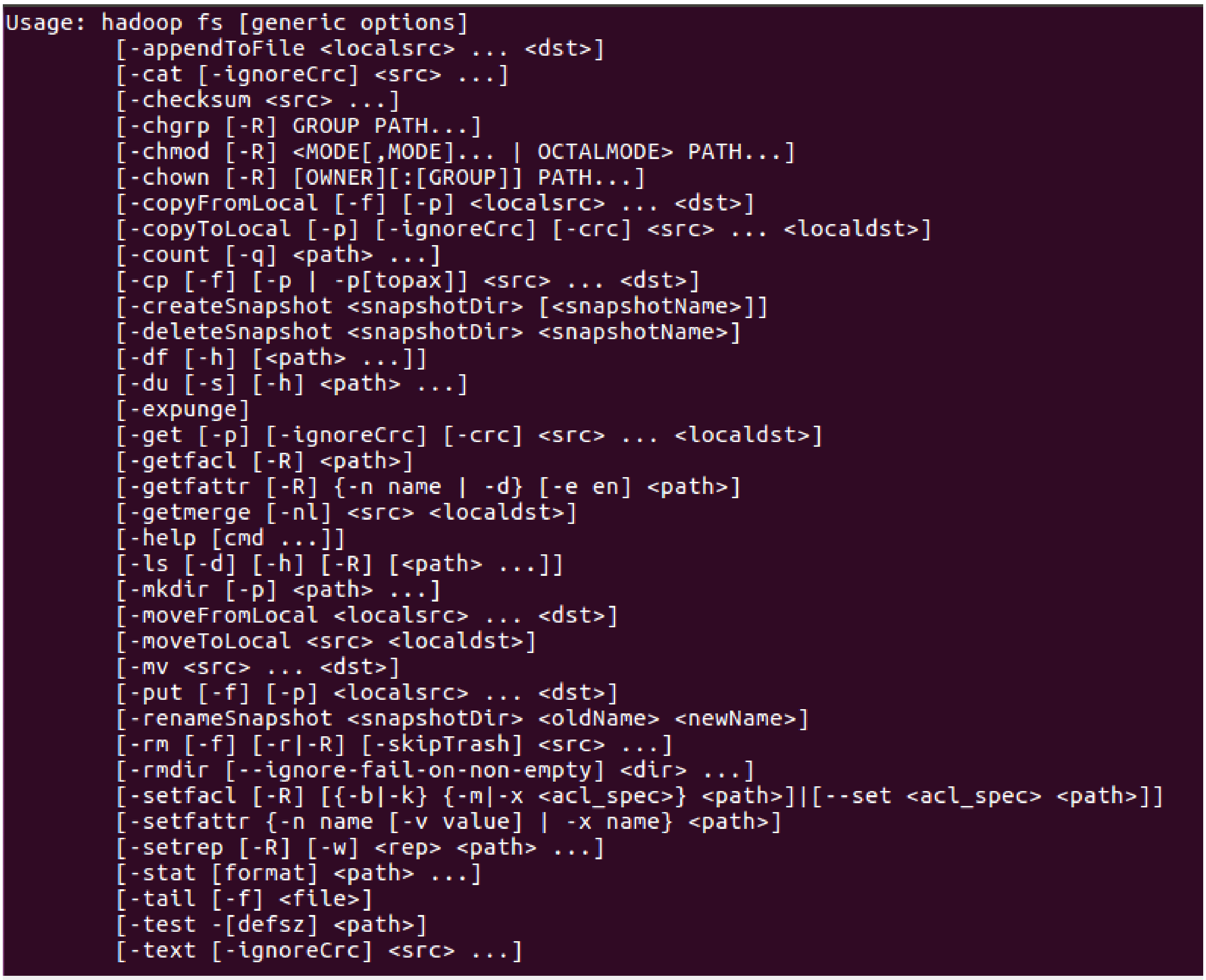
Note: The **-w** means wait till the replication is completed. And **-R** means recursively, we use it for directories as they may also contain many files and folders inside them.



```
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$ bin/hdfs dfs -setrep -R 2 /user
Replication 2 set: /user/local_file.txt
Replication 2 set: /user/myfile.txt
rusdes@rushil-IdeaPad:~/hadoop-3.3.2$
```

Note: There are more commands in HDFS but we discussed the commands which are commonly used when working with Hadoop. One can check out the list of *dfs* commands using the following command:

bin/hdfs dfs



```
Usage: hadoop fs [generic options]
        [-appendToFile <localsrc> ... <dst>]
        [-cat [-ignoreCrc] <src> ...]
        [-checksum <src> ...]
        [-chgrp [-R] GROUP PATH...]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER][:[GROUP]] PATH...]
        [-copyFromLocal [-f] [-p] <localsrc> ... <dst>]
        [-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
        [-count [-q] <path> ...]
        [-cp [-f] [-p | -p[topax]] <src> ... <dst>]
        [-createSnapshot <snapshotDir> [<snapshotName>]]
        [-deleteSnapshot <snapshotDir> <snapshotName>]
        [-df [-h] [<path> ...]]
        [-du [-s] [-h] <path> ...]
        [-expunge]
        [-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
        [-getfacl [-R] <path>]
        [-getfattr [-R] {-n name | -d} [-e en] <path>]
        [-getmerge [-nl] <src> <localdst>]
        [-help [cmd ...]]
        [-ls [-d] [-h] [-R] [<path> ...]]
        [-mkdir [-p] <path> ...]
        [-moveFromLocal <localsrc> ... <dst>]
        [-moveToLocal <src> <localdst>]
        [-mv <src> ... <dst>]
        [-put [-f] [-p] <localsrc> ... <dst>]
        [-renameSnapshot <snapshotDir> <oldName> <newName>]
        [-rm [-f] [-r|-R] [-skipTrash] <src> ...]
        [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
        [-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>]|[--set <acl_spec> <path>]]
        [-setfattr {-n name [-v value] | -x name} <path>]
        [-setrep [-R] [-w] <rep> <path> ...]
        [-stat [format] <path> ...]
        [-tail [-f] <file>]
        [-test -[defsz] <path>]
        [-text [-ignoreCrc] <src> ...]
```

CONCLUSION:

In this Experiment, I have performed various HDFS commands on the terminal. Through these commands one can access, write or update data or files in the HDFS system and also create or delete files and directories. To get a full list one can simply type *bin/hdfs dfs* command. Thus, I have learnt about HDFS and it's commands and successfully implemented them.