**EXPERIMENT 1**

**HDFS COMMANDS**

**Aim:** To understand the concept of HDFS and to execute HDFS commands

**Objectives:**

1. Create a file in local file system and copy it in HDFS

2. Copy a file from HDFS to local file system

3. Create a directory in HDFS and copy a file into it from local file system 4. Execute some other HDFS shell commands like: get, put, ls, cat, cp, mv, mkdir, rmdir, rm,appendToFile. tail, touchz, expunge etc.

**Key concept**

HDFS commands are used to access the Hadoop File System. HDFS stands for ‘**H**adoop **D**istributed **F**ile **S**ystem’.

The HDFS is a sub-project of the Apache Hadoop project. This Apache Software Foundation project is designed to provide a fault-tolerant file system designed to run on commodity hardware. HDFS is accessed through a set of shell commands.

For executing the HDFS commands , open the terminal in cloudera vm.

**Q1) How to check the version of Hadoop framework?**

Command: hadoop version

[cloudera@quickstart Desktop]$ hadoop version

Hadoop 2.6.0-cdh5.10.0

**Q2) List the files and subdirectories present in HDFS**

***Command:* hadoop fs –ls**

This command will list all the available files and subdirectories under default directory. For instance, in our example the default directory for Cloudera VM is /user/cloudera [cloudera@quickstart Desktop]$ hadoop fs -ls

**Q3: Return all the directories under root directory**

Variations of Hadoop ls Shell Command

[cloudera@quickstart Desktop]$ hadoop fs -ls /

**Q4) Copy a file from local to HDFS?**

**copyFromLocal**

HDFS Command to copy the file from a Local file system to HDFS.

***Usage:* hadoop fs -copyFromLocal <localsrc> <hdfs destination>**

**First I have created a file in local named student.txt**

[cloudera@quickstart Desktop]$ gedit students.txt

[cloudera@quickstart Desktop]$ hadoop fs -copyFromLocal students.txt /user/cloudera

**Q5: Check if the file is copied to HDFS?**

[cloudera@quickstart Desktop]$ hadoop fs -ls

**Q6: Check the contents of file that you copied in HDFS?**

cat: HDFS Command that reads a file on HDFS and prints the content of that file to the standard output.

***Usage:* hadoop fs –cat /path\_to\_file\_in\_hdfs**

[cloudera@quickstart Desktop]$ hadoop fs -cat students.txt

1,kriti,cse,45

2,neha,ece,56

3,jyothi,ce,78

4.priya,mechanical,89

**Q7: Achieve the same operation as above with put command?**

put: HDFS Command to copy single source or multiple sources from local file system to the destination file system.

***Usage:* hadoop fs -put <localsrc> <destination>**

[cloudera@quickstart Desktop]$ hadoop fs -put students.txt /user/cloudera/studentscopied.txt [cloudera@quickstart Desktop]$ hadoop fs -ls

**Q8) Copy any file from HDFS to Local File System**

∙ copyToLocal

HDFS Command to copy the file from HDFS to Local File System.

***Usage:* hadoop fs -copyToLocal <hdfs source> <localdst>**

[cloudera@quickstart Desktop]$ hadoop fs -copyToLocal students.txt studentscopiedtolocal.txt [cloudera@quickstart Desktop]$ ls

**Q9) Check the health of the Hadoop file system.**

∙ fsck

HDFS Command to check the health of the Hadoop file system.

***Command:* hdfs fsck /**

hdfs fsck /

The filesystem under path '/' is HEALTHY

**Q10) Create a directory in HDFS**

HDFS Command to create the directory in HDFS.

***Usage:* hadoop fs –mkdir /directory\_name**

[cloudera@quickstart Desktop]$ hadoop fs –mkdir kriti2018

[cloudera@quickstart Desktop]$ hadoop fs -ls

**Q11) Copy any file present in HDFS to a directory which is also present in HDFS** [cloudera@quickstart Desktop]$ hadoop fs –mkdir kriti2018

[cloudera@quickstart Desktop]$ hadoop fs -ls

[cloudera@quickstart Desktop]$ hadoop fs –cp kriti.txt kriti2018

**Q12): Create an empty file in HDFS?**

Touchz: HDFS Command to create a file in HDFS with file size 0 bytes. Usage: hadoop fs –touchz /directory/filename

[cloudera@quickstart Desktop]$ hadoop fs –touchz empty.txt

[cloudera@quickstart Desktop]$ hadoop fs -ls

**Q13) Check the file size of any file in HDFS?**

Du: HDFS Command to check the file size.

Usage: hadoop fs –du –s /directory/filename

[cloudera@quickstart Desktop]$ hadoop fs –du students.txt

67 67 /user/cloudera/students.txt

**Q14) Print the contents of a file stored in HDFS?**

∙ cat

HDFS Command that reads a file on HDFS and prints the content of that file to the standard output.

***Usage:* hadoop fs –cat /path/to/file\_in\_hdfs**

[cloudera@quickstart Desktop]$ hadoop fs -cat students.txt

1,kriti,cse,45

2,neha,ece,56

3,jyothi,ce,78

4.priya,mechanical,89

**Q15) Count the number of directories and files inside a directory in HDFS?**

count: **HDFS Command to count the number of directories, files, and bytes under the paths that match the specified file pattern.**

Usage: hadoop fs -count <path>

[cloudera@quickstart Desktop]$ hadoop fs –count kriti2018

1 1 13 /user/cloudera/kriti2018

**Q16) Remove a file from HDFS?**

rm: HDFS Command to remove the file from HDFS.

Usage: hadoop fs dfs –rm <path>

[cloudera@quickstart Desktop]$ hadoop fs -rm empty.txt

Deleted empty.txt

**Q17) Delete a directory completely in HDFS?**

rm -r

HDFS Command to remove the entire directory and all of its content from HDFS. Usage: hadoop fs -rm -r <path>

[cloudera@quickstart Desktop]$ hadoop fs -rm –r kriti2018

**Q18) Copy a file or multiple files in a directory in HDFS**

Cp: HDFS Command to copy files from source to destination. This command allows multiple sources as well, in which case the destination must be a directory**.**

***Usage:* hadoop fs -cp <src> <dest>**

***Command:* hadoop fs -cp /user/hadoop/file1 /user/hadoop/file2**

[cloudera@quickstart Desktop]$ hadoop fs -mkdir /user/cloudera/kriti2018 [cloudera@quickstart Desktop]$ hadoop fs –cp dummy3.txt kriti2018

**Q19)Move a file into a directory in HDFS?**

mv: HDFS Command to move files from source to destination. This command allows multiple sources as well, in which case the destination needs to be a directory.

Usage: hadoop fs -mv <src> <dest>

[cloudera@quickstart Desktop]$ hadoop fs –mv emptyfile.txt kriti2018

**Q20) Find a help for an individual command?**

Usage command gives all the options that can be used with a particular hdfs command. HDFS Command that returns the help for an individual command.

Usage: hadoop fs -usage <command>

Command: hdfs dfs -usage mkdir

[cloudera@quickstart Desktop]$ hdfs dfs -usage mkdir

Usage: hadoop fs [generic options] -mkdir [-p] <path> ...

**Q21) Find the help for a given or all commands?**

help

HDFS Command that displays help for given command or all commands if none is specified. Command: hadoop fs -help

**Q22)Check the memory status?**

Check memory status:

Usage: hadoop fs -df hdfs :/

[cloudera@quickstart Desktop]$ hadoop fs -df

Filesystem Size Used Available Use%

hdfs://quickstart.cloudera:8020 58531520512 1245229056 46116413440 2% [cloudera@quickstart Desktop]$

**Q23) Check for cluster balancing in HDFS?**

Cluster Balancing

Usage: hadoop balancer

Type command

hadoop balancer

**Q24) Change permission for a file to 777?**

chmod: Changes the permissions of files.

[cloudera@quickstart Desktop]$ hadoop fs -ls -r

[cloudera@quickstart Desktop]$ hadoop fs -chmod 777 /user/cloudera/students.txt [cloudera@quickstart Desktop]$ hadoop fs -ls -r

**Q25) Empty the trash in HDFS?**

expunge: Empties the trash. When you delete a file, it isn’t removed immediately from HDFS, but is renamed to a file in the /trash directory. As long as the file remains there, you can undelete it if you change your mind, though only the latest copy of the deleted file can be restored.

[cloudera@quickstart Desktop]$ hadoop fs -expunge

**26) Display the last kilobyte of the particular file?**

tail

This hadoop command will show the last kilobyte of the file to stdout.

**[cloudera@quickstart Desktop]$ hadoop fs -tail /user/cloudera/n.txt**

**27) Append the contents of a file present in local to a file present in HDFS** [cloudera@quickstart Desktop]$ gedit first.txt

[cloudera@quickstart Desktop]$ gedit second.txt

[cloudera@quickstart Desktop**]$ hadoop fs -copyFromLocal second.txt /user/cloudera/**

[cloudera@quickstart Desktop**]$ hadoop fs -appendToFile** /home/cloudera/Desktop/first.txt /user/cloudera/second.txt

[cloudera@quickstart Desktop]$ hadoop fs -cat /user/cloudera/second.txt