



**S. B. JAIN INSTITUTE OF TECHNOLOGY,
MANAGEMENT & RESEARCH, NAGPUR.**

Practical No. 2

Aim: Implementing the basic Hadoop HDFS Commands like File/Directory creation, deletion, update and many more operations.

Name of Student: Shruti Pradeep Bagdi

Roll No: CS22130

Semester/Year: 7th / 4th

Academic Session: 2025 - 2026

Date of Performance: _____

Date of Submission: _____

AIM: Implementing the basic Hadoop HDFS Commands like File/Directory creation, deletion, update and many more operations.

OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- Students will be able to recall the fundamental HDFS commands used for file and directory operations.
- Students will comprehend the concept of a distributed file system and its benefits for handling big data.
- The importance of caution when using delete commands and the necessity of irreversible operation confirmation.
- Students will develop practical skills in managing files and directories within HDFS, preparing them for more advanced data processing tasks.

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement: High Configuration computer

Software Requirement: Hadoop-3.3.6, jdk1.8, notepad++.

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. With the help of the HDFS command, we can perform Hadoop HDFS file operations like changing the file permissions, viewing the file contents, creating files or directories, copying file/directory from the local file system to HDFS or vice-versa, etc.

Before starting with the HDFS command, we have to start the Hadoop services. To start the Hadoop services do the following:

Hadoop mkdir Command

```
hadoop fs -mkdir /path/directory_name
```

This command's unique feature is the creation of a directory in the HDFS filesystem cluster if the directory does not exist. Besides, if the specified directory is present, then the output message will show an error signifying the directory's existence.

Hadoop Touchz(Creating an Empty File:)

`hadoop fs -touchz /directory/filename`

This command allows the user to create a new file in the HDFS cluster. The “directory” in the command refers to the directory name where the user wishes to create the new file, and the “filename” signifies the name of the new file which will be created upon the completion of the command.

Hadoop AppendToFile Command

`hadoop fs -appendToFile <localsrc> <dest>`

It allows the user to append the content of one or many files into a single file on the specified destination file in the HDFS filesystem cluster. On execution of this command, the given source files are appended into the destination source as per the given filename in the command.

Hadoop ls Command

`hadoop dfs -ls /path`

The Hadoop dfs shell command ls displays a list of the contents of a directory specified in the path provided by the user. It shows the name, permissions, owner, size, and modification date for each file or directories in the specified directory.

Hadoop chmod Command

`hadoop fs -chmod [-R] <mode> <path>`

This command is used when there is a need to change the permissions to accessing a particular file. On giving the chmod command, the permission of the specified file is changed. However, it is important to remember that the permission will be modified when the file owner executes this command.

copyFromLocal

The hadoop fs -copyFromLocal command is used to copy files from your local file system to the Hadoop Distributed File System (HDFS). This command is particularly useful when you want to transfer files from your local machine to the HDFS for further processing or storage in a Hadoop cluster.

`hadoop dfs -copyFromLocal /home/user/data.txt /user/hadoop/input`

In this example:

/home/user/data.txt is the source file on your local machine.

/user/hadoop/input is the destination directory in HDFS.

put:

This command is used to copy files from the local file system to the HDFS filesystem. This command is similar to –copyFromLocal command. This command will not work if the file already exists unless the –f flag is given to the command. This overwrites the destination if the file already exists before the copy.

Deleting a File

hadoop fs -rm /hdfs/path/to/file

Deleting a Directory (Empty):

hadoop fs -rmdir /hdfs/path/to/directory

du:

This command is used to show the amount of space in bytes that have been used by the files that match the specified file pattern. Even without the –s option, this only shows the size summaries one level deep in the directory.

hadoop dfs -du [-s] [-h]

-s:- used to show the size of each individual file that matches the pattern, shows the total (summary) size

-h:- used to format the sizes of the files in a human-readable manner rather than the number of bytes.

OUTPUT (SCREENSHOTS):

```
C:\>cd hadoop
C:\hadoop>cd sbin
C:\hadoop\sbin>start-dfs.cmd
C:\hadoop\sbin>jpd
'jpd' is not recognized as an internal or external command,
operable program or batch file.

C:\hadoop\sbin>jps
25488 NameNode
39988 DataNode
29996 Jps

C:\hadoop\sbin>start-yarn.cmd
starting yarn daemons

C:\hadoop\sbin>jps
25488 NameNode
39988 DataNode
40884 ResourceManager
39736 NodeManager
42092 Jps
```

```
C:\> Administrator: Command Prompt
C:\hadoop\sbin>hdfs fsck
Usage: hdfs fsck <path> [-list-corruptfileblocks | [-move | -delete | -openforwrite] [-files [-blocks [-locations | -racks | -replicaDetails | -upgradedomains]]]] [-includesnapshots] [-showprogress] [-storagepolicies] [-maintenance] [-blockId <blk_Id>] [-replicate]
    <path> start checking from this path
    -move move corrupted files to /lost+found
    -delete delete corrupted files
    -files print out files being checked
    -openforwrite print out files opened for write
    -includesnapshots include snapshot data if the given path indicates a snapshottable directory or there are snapshottable directories under it
    -list-corruptfileblocks print out list of missing blocks and files they belong to
    -files -blocks print out block report
    -files -blocks -locations print out locations for every block
    -files -blocks -racks print out network topology for data-node locations
    -files -blocks -replicaDetails print out each replica details
    -files -blocks -upgradedomains print out upgrade domains for every block
    -storagepolicies print out storage policy summary for the blocks
    -maintenance print out maintenance state node details
    -showprogress Deprecated. Progress is now shown by default
    -blockId print out which file this blockId belongs to, locations (nodes, racks) of this block, and other diagnostics info (under replicated, corrupted or not, etc)
    -replicate initiate replication work to make mis-replicated
blocks satisfy block placement policy
```

```
C:\> Administrator: Command Prompt
C:\hadoop\sbin>hdfs fsck /
Connecting to namenode via http://localhost:9870/fsck?ugi=user&path=%2F
FSCK started by user (auth:SIMPLE) from /127.0.0.1 for path / at Sat Aug 16 20:43:19 IST 2025

Status: HEALTHY
Number of data-nodes: 1
Number of racks: 1
Total dirs: 4
Total symlinks: 0

Replicated Blocks:
Total size: 0 B
Total files: 0
Total blocks (validated): 0
Minimally replicated blocks: 0
Over-replicated blocks: 0
Under-replicated blocks: 0
Mis-replicated blocks: 0
Default replication factor: 1
Average block replication: 0.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 0
Blocks queued for replication: 0
```

Administrator: Command Prompt

```
C:\hadoop\sbin>hdfs dfs -ls /
Found 1 items
drwxr-xr-x  - user supergroup          0 2025-08-16 20:05 /user

C:\hadoop\sbin>hdfs dfs -mkdir /myfolder

C:\hadoop\sbin>hdfs dfs -mkdir /bigdatatest

C:\hadoop\sbin>hdfs dfs -ls /
Found 3 items
drwxr-xr-x  - user supergroup          0 2025-08-16 20:46 /bigdatatest
drwxr-xr-x  - user supergroup          0 2025-08-16 20:46 /myfolder
drwxr-xr-x  - user supergroup          0 2025-08-16 20:05 /user
```

Administrator: Command Prompt

```
C:\hadoop\sbin>hdfs dfs -touchz /bigdatatest/test.dat

C:\hadoop\sbin>hdfs dfs -ls /
Found 3 items
drwxr-xr-x  - user supergroup          0 2025-08-16 20:48 /bigdatatest
drwxr-xr-x  - user supergroup          0 2025-08-16 20:46 /myfolder
drwxr-xr-x  - user supergroup          0 2025-08-16 20:05 /user

C:\hadoop\sbin>hdfs dfs -ls /bigdatatest /
Found 1 items
-rw-r--r--  1 user supergroup          0 2025-08-16 20:48 /bigdatatest/test.dat
Found 3 items
drwxr-xr-x  - user supergroup          0 2025-08-16 20:48 /bigdatatest
drwxr-xr-x  - user supergroup          0 2025-08-16 20:46 /myfolder
drwxr-xr-x  - user supergroup          0 2025-08-16 20:05 /user

C:\hadoop\sbin>hdfs dfs -du -s /bigdatatest/test.dat
0 0 /bigdatatest/test.dat
```

Administrator: Command Prompt

```
C:\hadoop\sbin>hdfs dfs -appendToFile - /bigdatatest/test.dat
My name is Shruti Bagdi
This is My Hadoop File.
^D
^Z

C:\hadoop\sbin>hdfs dfs -du -s /bigdatatest/test.dat
55 55 /bigdatatest/test.dat

C:\hadoop\sbin>hdfs dfs -ls /bigdatatest/test.dat
-rw-r--r--  1 user supergroup          55 2025-08-16 20:50 /bigdatatest/test.dat

C:\hadoop\sbin>hdfs dfs -ls /
Found 3 items
drwxr-xr-x  - user supergroup          0 2025-08-16 20:48 /bigdatatest
drwxr-xr-x  - user supergroup          0 2025-08-16 20:46 /myfolder
drwxr-xr-x  - user supergroup          0 2025-08-16 20:05 /user
```

```
C:\hadoop\sbin>hdfs dfs -chmod 777 /bigdatatest /  
  
C:\hadoop\sbin>hdfs dfs -ls /  
Found 3 items  
drwxrwxrwx  - user supergroup          0 2025-08-16 20:48 /bigdatatest  
drwxr-xr-x  - user supergroup          0 2025-08-16 20:46 /myfolder  
drwxr-xr-x  - user supergroup          0 2025-08-16 20:05 /user  
  
C:\hadoop\sbin>hdfs dfs -ls /bigdatatest /  
Found 1 items  
-rw-r--r--  1 user supergroup      55 2025-08-16 20:50 /bigdatatest/test.dat  
Found 3 items  
drwxrwxrwx  - user supergroup          0 2025-08-16 20:48 /bigdatatest  
drwxr-xr-x  - user supergroup          0 2025-08-16 20:46 /myfolder  
drwxr-xr-x  - user supergroup          0 2025-08-16 20:05 /user
```

```
C:\hadoop\sbin>hdfs dfs -rm /bigdatatest/test.dat  
Deleted /bigdatatest/test.dat  
  
C:\hadoop\sbin>hdfs dfs -ls /bigdatatest /  
Found 3 items  
drwxrwxrwx  - user supergroup          0 2025-08-16 20:56 /bigdatatest  
drwxr-xr-x  - user supergroup          0 2025-08-16 20:46 /myfolder  
drwxr-xr-x  - user supergroup          0 2025-08-16 20:05 /user  
  
C:\hadoop\sbin>hdfs dfs -ls /bigdatatest /
```

CONCLUSION:

Hadoop commands are fundamental for efficient data management in the Hadoop ecosystem. These commands empower users to create, copy, move, and delete files and directories within HDFS, laying the groundwork for effective distributed data handling and processing.

DISCUSSION AND VIVA VOCE:

- Explain the purpose of the -mv command in Hadoop.
- What is the purpose of the hadoop fs -mkdir command in HDFS?
- How does the hadoop fs -copyFromLocal command differ from hadoop fs -put?
- How can you remove a non-empty directory in HDFS? Provide the command.
- What is the outcome of the hadoop fs -rmdir command if the specified directory is not empty?

REFERENCE:

- https://www.google.com/search?q=hadoop+basic+commands+in+windows+10&sca_esv=560946880&ei=x7jtZKaCCvKQseMP6NKL0Ao&oq=hadoop+basic+commands+in+windo&gs_lp=EgxnD3

Big Data Analysis (PECCS602P)

[Mtd2l6LXNlcnAiHmhZG9AYFhgewAAgQYgGAZAGCA&sclient=gws-wiz-serp](#)

Observation book: (3)	Viva-Voce (3)	Quality of Submission and timely Evaluation (4)
Total:		Sign with date: