

**EX.NO: 2**

## **HADOOP FILE MANAGEMENT TASKS**

**DATE:**

**AIM:**

Implement the following file management tasks in Hadoop:

- Adding files and directories
- Retrieving files
- Deleting files

**DESCRIPTION:**

HDFS is a scalable distributed filesystem designed to scale to petabytes of data while running on top of the underlying filesystem of the operating system. HDFS keeps track of where the data resides in a network by associating the name of its rack (or network switch) with the dataset. This allows Hadoop to efficiently schedule tasks to those nodes that contain data, or which are nearest to it, optimizing bandwidth utilization. Hadoop provides a set of command line utilities that work similarly to the Linux file commands, and serve as your primary interface with HDFS. We're going to have a look into HDFS by interacting with it from the command line. We will take a look at the most common file management tasks in Hadoop, which include:

- Adding files and directories to HDFS
- Retrieving files from HDFS to local filesystem
- Deleting files from HDFS

**ALGORITHM:**

### **SYNTAX AND COMMANDS TO ADD, RETRIEVE AND DELETE DATA FROM HDFS**

#### **Step-1: Adding Files and Directories to HDFS**

Before you can run Hadoop programs on data stored in HDFS, you'll need to put the data into HDFS first. Let's create a directory in and put a file in it. HDFS has a default working directory of /user/\$USER, where \$USER is your login user name. This directory isn't automatically created for you, though, so let's create it with the mkdir command.

**Note: input\_file.txt is created in sbin with some contents**

```
C:\hadoop-2.8.0\sbin>hadoop fs -mkdir /input_dir
```

```
C:\hadoop-2.8.0\sbin>hadoop fs -put input_file.txt /input_dir/input_file.txt
```

#### **Step 2: List the contents of a directory.:**

```
C:\hadoop-2.8.0\sbin>hadoop fs -ls /input_dir/
```

### **Step 3: Retrieving Files from HDFS**

The Hadoop command get copies files from HDFS back to the local filesystem. To retrieve example.txt, we can run the following command:

```
C:\hadoop-2.8.0\sbin>Hadoop fs -cat /input_dir/input_file.txt
```

**Output:** Hello world hello hi (which is stored in input\_file .txt)

### **Step 4: Download the file:**

Command: `hadoop fs -get`: Copies/Downloads files to the local file system Example:

```
hadoop fs -get /user/saurzcode/dir3/Samplefile.txt /home/
```

### **Step 5: Copy a file from source to destination**

This command allows multiple sources as well in which case the destination must be a directory.

Command: `hadoop fs -cp`

Example: `hadoop fs -cp /user/saurzcode/dir1/abc.txt /user/saurzcode/ dir2` **Step**

### **6: Copy a file from/To Local file system to HDFS copyFromLocal**

Command: `hadoop fs -copyFromLocal` URI

Example: `hadoop fs -copyFromLocal /home/saurzcode/abc.txt /user/ saurzcode/abc.txt`

### **copyToLocal**

Command: `hadoop fs -copyToLocal [-ignorecrc] [-crc] URI`

### **Step 7: Move file from source to destination**

Note:- Moving files across filesystem is not permitted.

Command: `hadoop fs -mv`

Example: `hadoop fs -mv /user/saurzcode/dir1/abc.txt /user/saurzcode/ dir2`

### **Step 8: Deleting Files from HDFS**

```
C:\hadoop-2.8.0\sbin>hadoop fs -rm input_file.txt /input_dir/input_file.txt
```

### **Recursive version of delete:**

Command: `hadoop fs -rmr`

Example: `hadoop fs -rmr /user/saurzcode/`

### **Step 9: Display last few lines of a file**

Similar to tail command in Unix.

Usage : `hadoop fs -tail`

Example: `hadoop fs -tail /user/saurzcode/dir1/abc.txt`

### **Step 10: Display the aggregate length of a file**

Command: `hadoop fs -du`

Example: `hadoop fs -du /user/saurzcode/dir1/abc.txt`

## HADOOP OPERATION:

1. Open cmd in administrative mode and move to “C:/Hadoop-2.8.0/sbin” and start cluster

Start-all.cmd

The screenshot shows a Windows Command Prompt window titled "Administrator: Command Prompt" with the following text:

```
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\WINDOWS\system32>cd ..
C:\Windows>cd ..
C:\>cd Hadoop-2.8.0\sbin
C:\Hadoop-2.8.0\sbin>start-all.cmd
This script is deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
C:\Hadoop-2.8.0\sbin>
```

Overlaid on top of the Command Prompt are four windows from the "Apache Hadoop Distribution" package:

- `hadoop namenode`
- `hadoop datanode`
- `yarn resourcemanager`
- `yarn nodemanager`

The `yarn nodemanager` window displays a log of warnings from `util.SysInfoWindows`, indicating that the expected split length of `sysInfo` is 11, but only 7 was received. The log includes timestamps and various system identifiers.

1. Create an input directory in

HDFS. `hadoop fs -mkdir`

`/input_dir`

2. Copy the input text file named input\_file.txt in the input directory (input\_dir) of HDFS.

```
hadoop fs -put C:/input_file.txt /input_dir
```

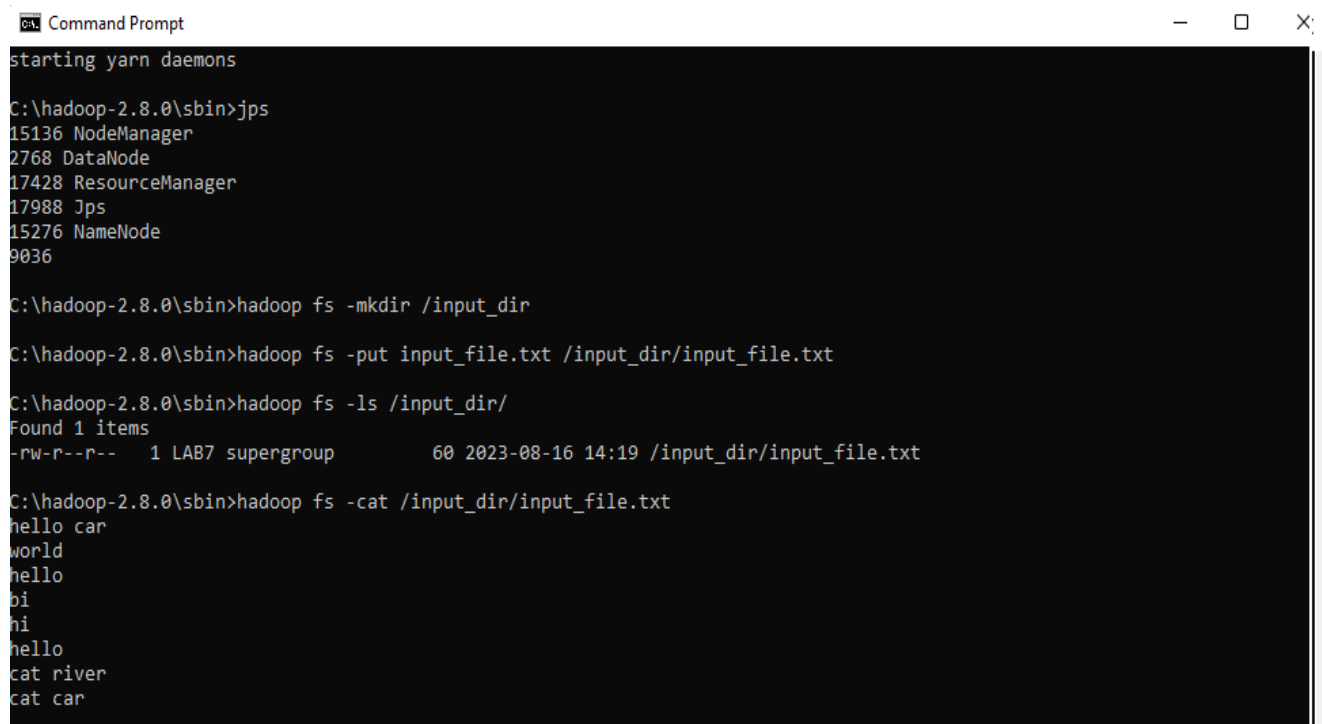
3. Verify input\_file.txt available in HDFS input directory (input\_dir).

```
hadoop fs -ls /input_dir/
```

### Verify content of the copied file.

```
hadoop dfs -cat /input_dir/input_file.txt
```

### OUTPUT:



```
Command Prompt
starting yarn daemons
C:\hadoop-2.8.0\sbin>jps
15136 NodeManager
2768 DataNode
17428 ResourceManager
17988 Jps
15276 NameNode
9036

C:\hadoop-2.8.0\sbin>hadoop fs -mkdir /input_dir

C:\hadoop-2.8.0\sbin>hadoop fs -put input_file.txt /input_dir/input_file.txt

C:\hadoop-2.8.0\sbin>hadoop fs -ls /input_dir/
Found 1 items
-rw-r--r-- 1 LAB7 supergroup      60 2023-08-16 14:19 /input_dir/input_file.txt

C:\hadoop-2.8.0\sbin>hadoop fs -cat /input_dir/input_file.txt
hello car
world
hello
bi
hi
hello
cat river
cat car
```

### OTHER COMMANDS:

1. To leave Safe mode

```
hadoop dfsadmin -safemode leave
```

2. To delete file from HDFS directory

```
hadoop fs -rm -r /input_dir/input_file.txt
```

3. To delete directory from HDFS directory

```
hadoop fs -rm -r /input_dir
```

## OUTPUT

```
C:\>hadoop dfsadmin -safemode leave
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Safe mode is OFF

C:\>hadoop fs -rm -r /input_dir/input_file.txt
Deleted /input_dir/input_file.txt

C:\>hadoop fs -rm -r /input_dir
Deleted /input_dir

C:\>
```

## OUTPUT:

Browsing HDFS - Mozilla Firefox

localhost:50070/explorer.html#/

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

### Browse Directory

/

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	lendi	supergroup	0 B	Wed 17 Aug 2016 02:44:00 AM EDT	0	0 B	<a href="#">lendi_english</a>
drwxr-xr-x	lendi	supergroup	0 B	Wed 17 Aug 2016 02:17:48 AM EDT	0	0 B	<a href="#">sadhana</a>
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:31:42 AM EDT	0	0 B	<a href="#">shakes</a>
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:35:59 AM EDT	0	0 B	<a href="#">shakes1</a>
drwx-----	lendi	supergroup	0 B	Sat 13 Aug 2016 01:19:03 AM EDT	0	0 B	<a href="#">tmp</a>

**RESULT:**

Thus, the implementation for file management in Hadoop was successfully executed.