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 withHDFS. 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 retrieveexample.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

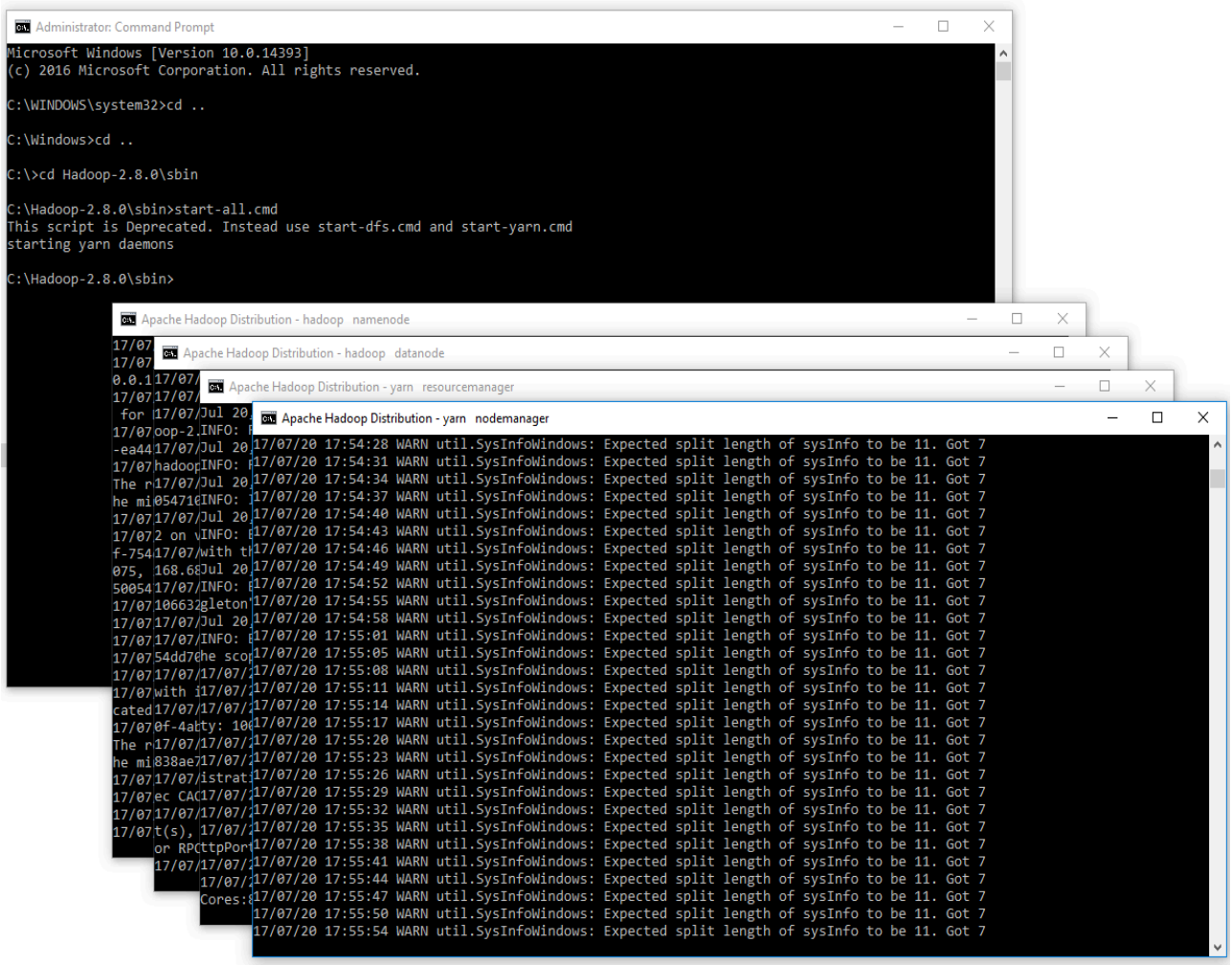
Command: hadoop fs -du

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



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

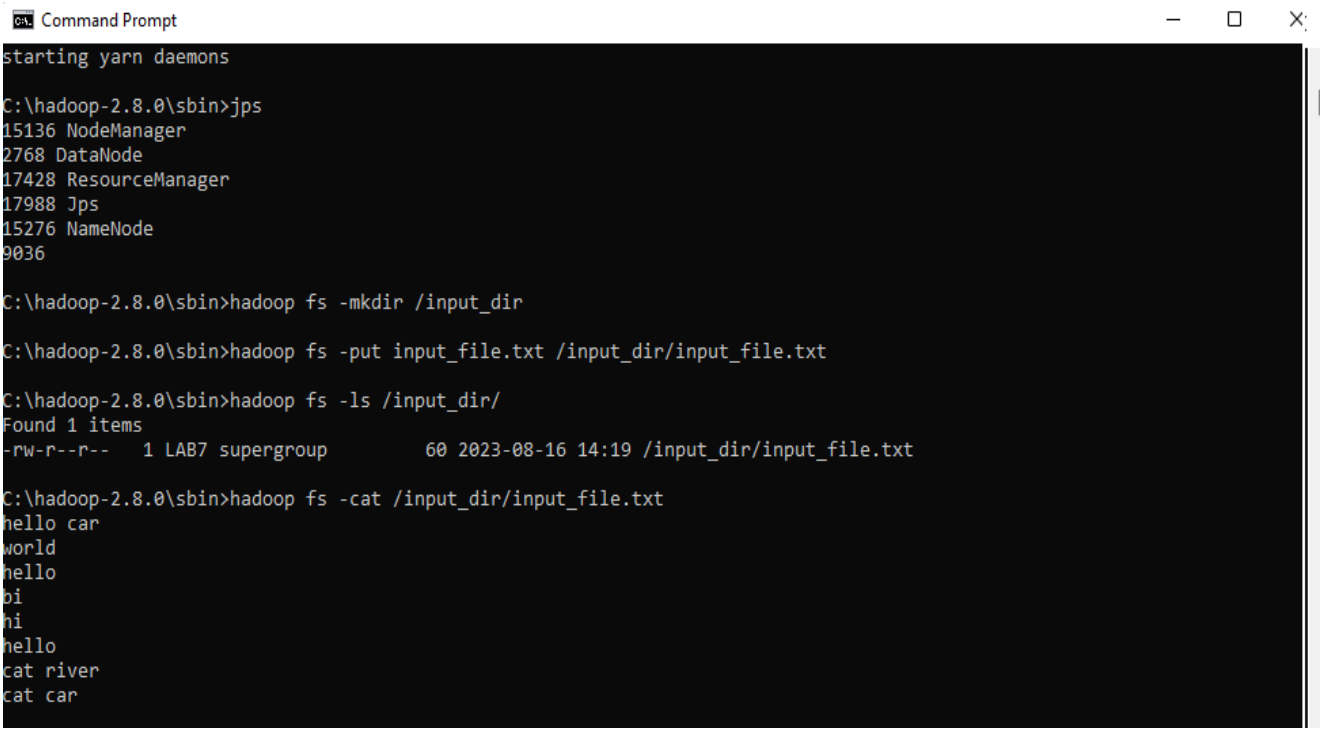
3. Verify input_file.txt available in HDFS input directory (input_dir).

   hadoop fs -ls /input_dir/

Verify content of the copied file.

hadoop dfs -cat /input_dir/input_file.txt

OUTPUT:



OTHER COMMANDS:

1. To leave Safe mode
   hadoop dfsadmin –safemode leave

2. To delete file from HDFS directory
   hadoop fs -rm -r /iutput_dir/input_file.txt

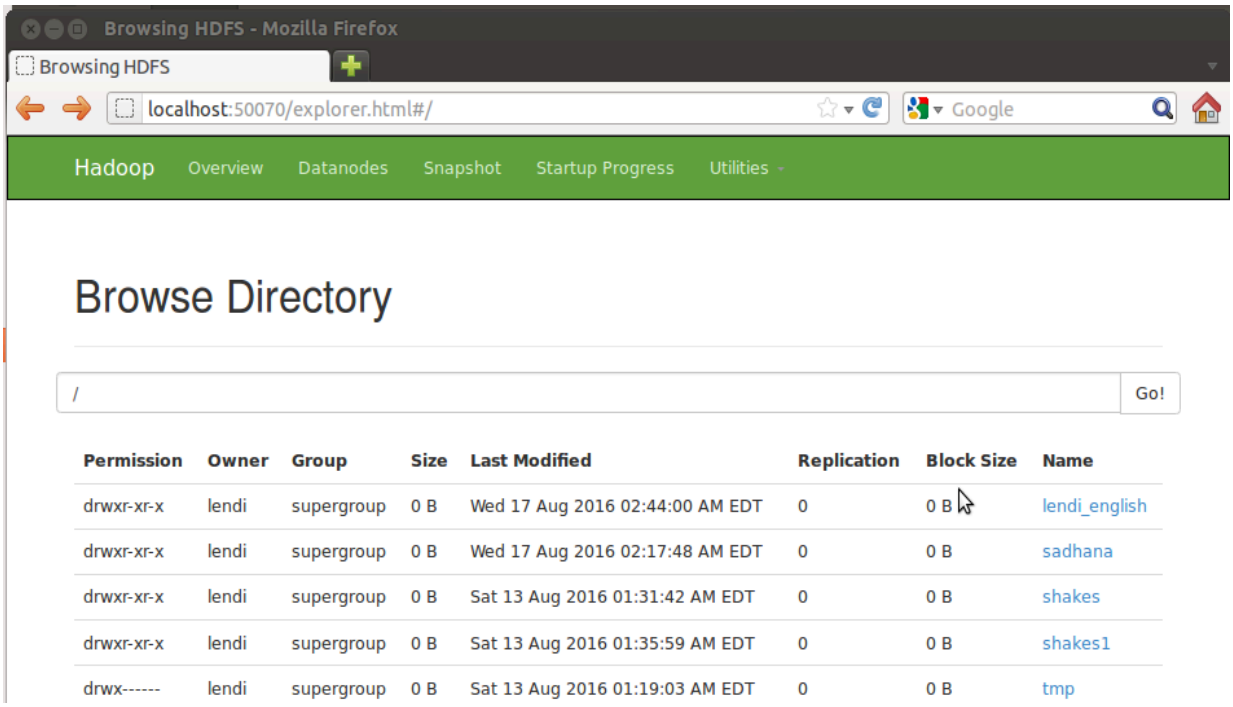
3. To delete directory from HDFS directory

   hadoop fs -rm -r /iutput_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:



RESULT:

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