**Experiment 5:**

Implement the following file management tasks in Hadoop:

* Adding files and directories
* Retrieving files
* Deleting files

Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into HDFS using one of the above command line utilities

**Solution:**

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 we need to start the Hadoop services using the following command:

**start-all.cmd** (note that is sbin is already specified in the path then we can directly run the commands.

OR

**start-dfs.cmd**

**start-yarn.cmd**

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

Jps

On command prompt perform the following commands and note the output.

**Commands:**

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

**Syntax:**

hdfs dfs -ls <path>

**Example:**

hdfs dfs -ls / (/ for root directory)

It will print all the directories present in HDFS.

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

**Syntax:**

hdfs dfs -mkdir <folder name>

creating home directory:

hdfs -mkdir /user

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

1. **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:**

hdfs dfs -copyFromLocal <local file path> <dest(present on hdfs)>

**Example:** Let’s suppose we have a file a1*.txt* on Desktop which we want to copy to folder /user/XYZpresent on hdfs.

hdfs dfs -copyFromLocal ../Desktop/a1.txt /user/XYZ

(OR)

hdfs dfs -put ../Desktop/a1.txt /user/XYZ

1. **cat:** To print file contents.

**Syntax:**

hdfs dfs -cat <path>

**Example:**

// print the content of a1.txt present inside /user/XYZ folder.

hdfs dfs -cat /user/XYZ/a1.txt

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

**Syntax:**

hdfs dfs -copyToLocal <<srcfile(on hdfs)> <local file dest>

**Example:**

hdfs dfs -copyToLocal /user/XYZ/myfile.txt ../Desktop/PCE

(OR)

hdfs dfs -get /user/XYZ/myfile.txt ../Desktop/PCE

*myfile.txt* from /user/XYZ folder will be copied to folder PCE present on *Desktop*.

1. **cp:** This command is used to copy files within hdfs. This command allows multiple sources as well in which case the destination must be a directory.

**Syntax:**

hdfs dfs -cp <src(on hdfs)> <dest(on hdfs)>

**Example:**

hdfs -cp /user/XYZ/a1.txt /user/a2.txt

The destination can also be a folder name in which case the a1.txt will be copied with the same file name.

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

**Syntax:**

hdfs dfs -rmr <filename/directoryName>

**Example:**

hdfs dfs -rmr /user/mydir -> It will delete all the content inside the directory (mydir)then the directory itself.