**HDFS Commands**

NOTE:

Use **hadoop dfs** or **hadoop fs** or **hdfs dfs** or **hdfs fs** to write any HDFS Shell command listed below:

* **usage:**
  + HDFS Command that returns the help for an individual command.
  + By using usage command, you can get information about any command.
  + Syntax :
    - hdfs dfs –usage <command\_name>
  + Example:
    - hdfs dfs –usage mkdir
* **help:**
  + HDFS Command that displays help for given command or all commands if none is specified.
  + Syntax :
    - hdfs dfs –help

* **mkdir:**
  + **HDFS Command to create the directory in HDFS.**
  + **Syntax:**
    - **hdfs dfs –mkdir <directory\_name>**
  + **Example:**
    - **hdfs dfs –mkdir /new\_univo**
  + **NOTE:**
    - **Here we are trying to create a directory named “new\_univo” in HDFS.**
    - **In order to create sub directory, parent directory must exist else ‘No Such directory found’ Error would come.**
* **ls:**
  + **HDFS Command to display the list of Files and Directories in HDFS.**
  + **Syntax:**
    - **hdfs dfs –ls <dir\_path>**
  + **Example:**
    - **hdfs dfs –ls /**

**NOTE :**

**This command will list all the files and directories in HDFS root directory.**

* The information returned is very similar to that returned by the Unix command ls -l, with a few minor differences. The first column shows the file mode.
* The second column is the replication factor of the file (something a traditional Unix filesystem does not have). Remember we set the default replication factor in the site-wide configuration to be 1, which is why we see the same value here. The entry in this column is empty for directories because the concept of replication does not apply to them—directories are treated as metadata and stored by the namenode, not the datanodes.
* The third and fourth columns show the file owner and group.
* The fifth column is the size of the file in bytes, or zero for directories.
* The sixth and seventh columns are the last modified date and time.
* Finally, the eighth column is the name of the file or directory.
* **To list the content of the directories recursively:**
* **Example:**
  + - **hdfs dfs –lsr /**
* **chmod:**
  + Changes the permission of a file
  + This works similar to a Linux shell’s chmod command with a few exceptions.
  + –R: Recursively change the permission for all the files underlying the path you mention.
  + HDFS has a permissions model for files and directories that is much like the POSIX model.
  + There are three types of permission:
  + The read permission (r), the write permission (w), and the execute permission (x).
  + The read permission is required to read files or list the contents of a directory.
  + The write permission is required to write a file or, for a directory, to create or delete files or directories in it.
  + The execute permission is ignored for a file because you can’t execute a file on HDFS (unlike POSIX), and for a directory this permission is required to access its children.
  + Each file and directory has an *owner*, a *group*, and a *mode*.
  + The mode is made up of the permissions for the user who is the owner, the permissions for the users who are members of the group, and the permissions for users who are neither the owners nor members of the group.
  + Syntax:
    - hadoop dfs –chmod [–R] <MODE> <PATH>
  + Example:
    - hadoop dfs –chmod –R 754 /user/Hadoop
* r w x (User)
* r w x (Group)
* r w x (Mode For all the Users Members of Non-Members)
* **copyFromLocal or put:**
  + HDFS Command to copy the file from a Local file system to HDFS.
  + Syntax:
    - hdfs dfs -copyFromLocal <localsrc> <hdfs destination>
  + Example:
    - hdfs dfs –copyFromLocal /home/univo/test /new\_univo
  + NOTE :
    - Here the test is the file present in the local directory /home/edureka and after the command is executed the test file will be copied in /new\_edureka directory of HDFS.
    - Default path to destination would be home directory at local file system. That is by default : /home/<user\_name>

hdfs dfs –put /home/univo/test

* + - File name in Unix can be without extension as well.
* **copyToLocal or get:**
  + HDFS Command to copy the file from HDFS to Local File System.
  + Syntax :
    - hdfs dfs -copyToLocal <hdfs\_source> <localdst>

or

hdfs dfs -get <hdfs source> <localdst>

* + Example :
    - **hdfs dfs –copyToLocal /new\_univo/test /home/univo**

NOTE:

* Default path to destination would be home directory at local file system. That is by default : /home/<user\_name>

hdfs dfs –get /new\_univo/test

* File name in UNIX can be without extension as well.
* Note: Here test is a file present in the new\_univo directory of HDFS and after the command gets executed the test file will be copied to local directory /home/edureka
* **cat:**
  + HDFS Command that reads a file on HDFS and prints the content of that file to the standard output.
  + Syntax :
    - hdfs dfs –cat /path/to/file\_in\_hdfs
  + Example:
    - hdfs dfs –cat /new\_univo/test
* **du:**
  + HDFS Command to check the file size.
  + Syntax:
    - hdfs dfs –du –s /directory/filename
  + Example:
    - hdfs dfs –du –s /new\_univo/test
* **touchz:**
  + HDFS Command to create a file in HDFS with file size 0 bytes.
  + Syntax:
    - hdfs dfs –touchz /directory/filename
  + Example:
    - hdfs dfs –touchz /new\_univo/sample
* **count:**
  + HDFS Command to count the number of directories, files, and bytes under the paths that match the specified file pattern.
  + Syntax:
    - hdfs dfs -count <path>
  + Example:
    - hdfs dfs –count /user
* **rm:**
  + HDFS Command to remove the file from HDFS.
  + Syntax:
    - hdfs dfs –rm <path>
  + Example:
    - hdfs dfs –rm /new\_univo/test
* **rm –r:**
  + HDFS Command to remove the entire directory and all of its content from HDFS.
  + Syntax:
    - hdfs dfs -rm -r <path>
  + Example:
    - hdfs dfs -rm -r  /new\_univo
* **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.
  + Syntax:
    - hdfs dfs -cp <src> <dest>
  + Example:
    - hdfs dfs -cp /user/hadoop/file1 /user/hadoop/file2
    - hdfs dfs -cp /user/hadoop/file1 /user/hadoop/file2 /user/hadoop/dir
* **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.
  + Syntax:
    - hdfs dfs -mv <src> <dest>
  + Example:
    - hdfs dfs -mv /user/hadoop/file1 /user/hadoop/file2
* **rmdir:**
  + HDFS Command to remove the directory.
  + Syntax:
    - hdfs dfs -rmdir <path>
  + Example:
    - hdfs dfs –rmdir /user/Hadoop
* **setrep:**
  + This command is used to change the replication factor of a file to a specific instead of the default of replication factor for the remaining in HDFS. Remaining in HDFS.
  + If <path> is a directory then the command recursively changes the replication factor of all files under the directory tree rooted at path <path>
  + Syntax:
    - hadoop dfs –setrep <replication factor number> <file/path\_name>
  + Example:
    - hadoop dfs –setrep 2 /user/Hadoop