**PRACTICAL :- 3**

**HDFS Commands**

HDFS is the main hub of the Hadoop ecosystem, responsible for storing large data sets both structured & unstructured across various nodes & thereby maintaining the metadata in the form of log files. Thus, to work with such a system, we need to be well versed or at least should be aware of the common commands and processes to ease our task. In that matter, we have consolidated some of the most commonly used HDFS commands that one should know to work with HDFS.



1. **Version**: this command is used to know the version of Hadoop, with additional local file system location and compilation information.

hadoop version



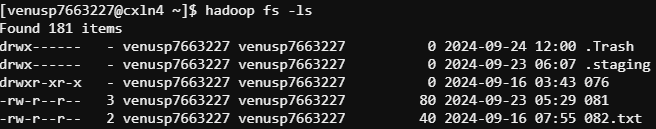
1. **mkdir**: this command is used to create a new directory, if it does not exists. If the directory exists, it will give a “File already exists” error.

hadoop fs -mkdir <Directory Path/Name>



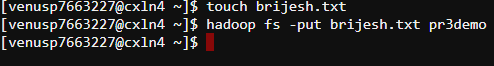
1. **ls**: this command is used to check the files or directory in the HDFS. It shows the name, permissions, owner, size, and modification date for each file or directory in the specified directory.

hadoop fs -ls



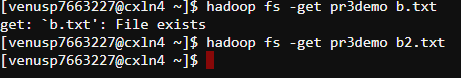
1. **put**: this command is used to copy the data from the local file system to HDFS.

hadoop fs -put <Local File Path> <HDFS file path>



1. **get**: this command is used to copy the data from HDFS to the local file system. This command is the reverse of the ‘put’ command.

hadoop fs -get <HDFS file path> <Local File Path>



1. **cat**: command used to view the data from the file in HDFS

hadoop fs -cat <HDFS file path with file name>



1. **mv**: this command is used to move a file from one location to HDFS to another location in HDFS.

hadoop fs -mv <Source HDFS path> <Destination HDFS path>



1. **cp**: this command is used to copy a file from one location to HDFS to another location within HDFS only.

hadoop fs -cp <Source HDFS path> <Destination HDFS path>



1. **copyFromLocal**: this command is used to copy data from the local file system to HDFS.

hadoop fs -copyFromLocal <Local File Path> <HDFS file path>



1. **copyToLocal**: this command is used to copy data from HDFS to the local file system.

hadoop fs -copyToLocal <HDFS file path> <Local File Path>



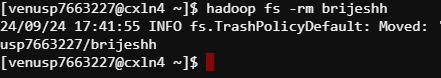
1. **moveFromLocal**: this command is used for moving a file or directory from the local file system to HDFS.

hadoop fs -moveFromLocal <Local File Path> <HDFS file path>



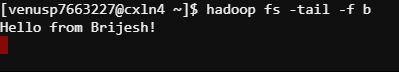
1. **rm**: removes, this command is used to delete/remove a file from HDFS.

hadoop fs -rm <HDFS file path>



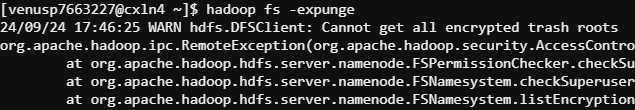
1. **tail**: this command is used to read the tail/end part of the file from HDFS. It has an additional parameter “[-f]”, that is used to show the appended data to the file.

hadoop fs -tail [-f] <HDFS file path>



1. **expunge**: this command is used to make the trash empty.

hadoop fs -expunge



1. **chown**: we should use this command when we want to change the user of a file or directory in HDFS.

hadoop fs -chown <HDFS file path>

1. **chgrp**: we should use this command when we want to change the group of a file or directory in HDFS.

hadoop fs -chgrp <HDFS file path>

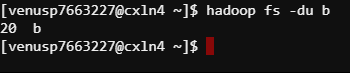
1. **setrep**: this command is used to change the replication factor of a file in HDFS.

hadoop fs -setrep <Replication Factor> <HDFS file path>



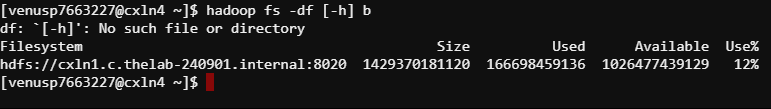
1. **du**: this command is used to check the amount of disk usage of the file or directory.

hadoop fs -du <HDFS file path>



1. **df**: this command is used to shows the capacity, free space and size of the HDFS file system. It has an additional parameter “[-h]” to convert the data to a human-readable format.

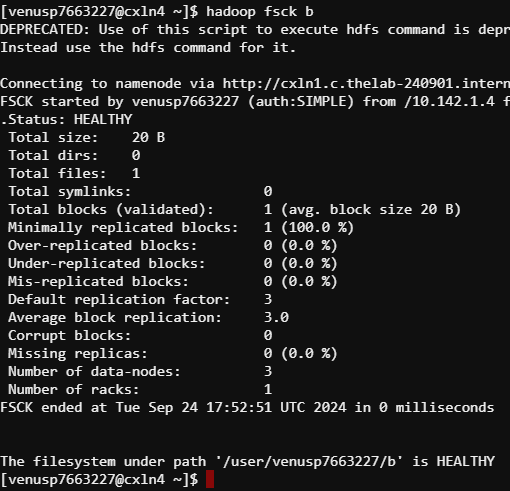
hadoop fs -df [-h] <HDFS file path>



1. **fsck**: this command is used to check the health of the files present in the HDFS file system.

hadoop fsck <HDFS file path>

It also has some attributes/options to modify the command use.



1. **Touchz**: this command creates a new file in the specified directory of size 0.

hadoop fs -touchz <HDFS file path>

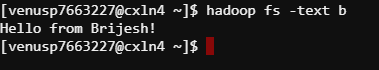


1. **test**: this command answer various questions about <HDFS path>, with the result via exit status.

hadoop fs -test <HDFS file path>

1. **Text**: this is a simple command, used to print the data of an HDFS file on the console.

hadoop fs -text <HDFS file path>



1. **stat**: this command provides the stat of the file or directory in HDFS.

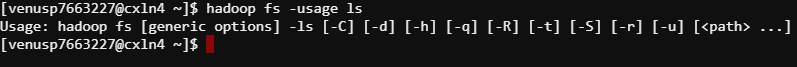
hadoop fs -stat <HDFS file path>

It can provide data in the following formats. By default, it uses ‘%y’.



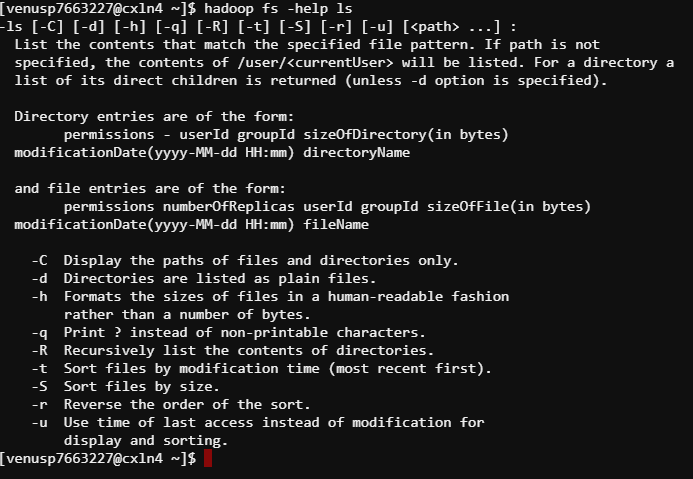
1. **Usage**: Displays the usage for given command or all commands if none is specified.

hadoop fs -usage <command>



1. **help**: Displays help for given command or all commands if none is specified.

hadoop fs -help <command>



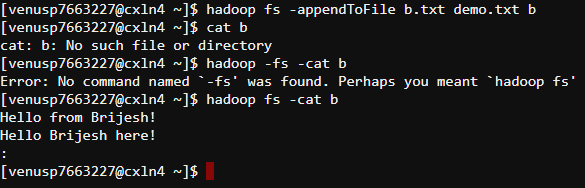
1. **chmod**: It is used to change the permission of the file in the HDFS file system.

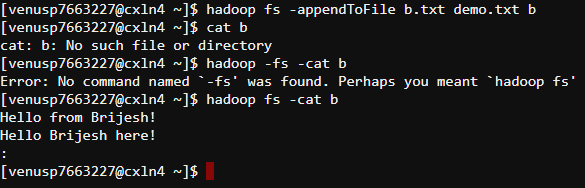
hadoop fs -chmod [-r] <HDFS file path>



1. **appendToFile**: this command is used to merge two files from the local file system to one file in the HDFS file.

hadoop fs -appendToFile <Local file path1> <Local file path2> <HDFS file path>





1. **checksum**: this command is used to check the checksum of the file in the HDFS file system.

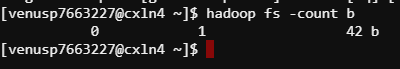
hadoop fs -checksum <HDFS file Path>



1. **count**: it counts the number of files, directories and size at a particular path.

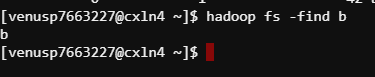
This function also has few functions to modify the query as per need.

hadoop fs -count [options] <HDFS directory path>



1. **Find**: this command is used to find the files in the HDFS file system. We need to provide the expression that we are looking for and can also provide a path if we want to look for the file at a particular directory.

hadoop fs -find <HDFS directory path> <Expression>



1. **getmerge**: this command is used to merge the contents of a directory from HDFS to a file in the local file system.

hadoop fs -getmerge <HDFS directory> <Local file path>

