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.

```
[venusp7663227@cxln4 ~]$ vi demo.txt
[venusp7663227@cxln4 ~]$
```

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

hadoop version

```
[venusp7663227@cxln4 ~]$ hadoop version
Hadoop 2.7.3.2.6.2.0-205
Subversion git@github.com:hortonworks/hadoop.git -r 721db98dcc87332e6a5c87ca1a5726b82d8a7fa0
Compiled by jenkins on 2017-08-26T09:20Z
Compiled with protoc 2.5.0
From source with checksum 90b73c4c185645c1f47b61f942230
```

2. **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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -mkdir 21012021001_Brijesh
[venusp7663227@cxln4 ~]$
```

3. **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

```
[venusp7663227@cxln4 ~]$ hadoop fs -ls
Found 181 items
drwx------   - venusp7663227 venusp7663227          0 2024-09-24 12:00 .Trash
drwx------   - venusp7663227 venusp7663227          0 2024-09-23 06:07 .staging
drwxr-xr-x   - venusp7663227 venusp7663227          0 2024-09-16 03:43 076
-rw-r--r--   3 venusp7663227 venusp7663227         80 2024-09-23 05:29 081
-rw-r--r--   2 venusp7663227 venusp7663227         40 2024-09-16 07:55 082.txt
```

4. **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>

```
[venusp7663227@cxln4 ~]$ touch brijesh.txt
[venusp7663227@cxln4 ~]$ hadoop fs -put brijesh.txt pr3demo
[venusp7663227@cxln4 ~]$
```

5. **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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -get pr3demo b.txt
get: `b.txt': File exists
[venusp7663227@cxln4 ~]$ hadoop fs -get pr3demo b2.txt
[venusp7663227@cxln4 ~]$
```

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

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

```
[venusp7663227@cxln4 practical_3]$
[venusp7663227@cxln4 practical_3]$ hadoop fs -cat demo.txt
Hello from uvpce
```

7. **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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -mv b brijeshhh
[venusp7663227@cxln4 ~]$
```

8. **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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -cp brijesh brijeshh
[venusp7663227@cxln4 ~]$
```

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

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

```
[venusp7663227@cxln4 ~]$ hadoop fs -copyFromLocal demo.txt brijeshhhh
[venusp7663227@cxln4 ~]$
```

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

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

```
[venusp7663227@cxln4 ~]$ hadoop fs -copyToLocal brijeshh t.txt
[venusp7663227@cxln4 ~]$
```

11. 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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -moveFromLocal brijesh.txt b
[venusp7663227@cxln4 ~]$
```

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

hadoop fs -rm <HDFS file path>

```
[venusp7663227@cxln4 ~]$ hadoop fs -rm brijeshh
24/09/24 17:41:55 INFO fs.TrashPolicyDefault: Moved:
usp7663227/brijeshh
[venusp7663227@cxln4 ~]$
```

13. 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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -tail -f b
Hello from Brijesh!
```

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

hadoop fs -expunge

```
[venusp7663227@cxln4 ~]$ hadoop fs -expunge
24/09/24 17:46:25 WARN hdfs.DFSClient: Cannot get all encrypted trash roots
org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.AccessContro
        at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkSu
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkSuperuser
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.listEncryption
```

15. 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>

16. 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>

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

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

```
[venusp7663227@cxln4 ~]$ hadoop fs -setrep 2 demo.txt
[venusp7663227@cxln4 ~]$
```

18. du: this command is used to check the amount of disk usage of the file or directory.
hadoop fs -du <HDFS file path>

```
[venusp7663227@cxln4 ~]$ hadoop fs -du b
20  b
[venusp7663227@cxln4 ~]$
```

19. 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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -df [-h] b
df: `[-h]': No such file or directory
Filesystem                                          Size          Used      Available  Use%
hdfs://cxln1.c.thelab-240901.internal:8020  1429370181120  166698459136  1026477439129   12%
[venusp7663227@cxln4 ~]$
```

20. 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.

```
[venusp7663227@cxln4 ~]$ hadoop fsck b
DEPRECATED: Use of this script to execute hdfs command is depr
Instead use the hdfs command for it.

Connecting to namenode via http://cxln1.c.thelab-240901.intern
FSCK started by venusp7663227 (auth:SIMPLE) from /10.142.1.4 f
.Status: HEALTHY
 Total size:    20 B
 Total dirs:    0
 Total files:   1
 Total symlinks:                0
 Total blocks (validated):      1 (avg. block size 20 B)
 Minimally replicated blocks:   1 (100.0 %)
 Over-replicated blocks:        0 (0.0 %)
 Under-replicated blocks:       0 (0.0 %)
 Mis-replicated blocks:         0 (0.0 %)
 Default replication factor:    3
 Average block replication:     3.0
 Corrupt blocks:                0
 Missing replicas:              0 (0.0 %)
 Number of data-nodes:          3
 Number of racks:               1
FSCK ended at Tue Sep 24 17:52:51 UTC 2024 in 0 milliseconds


The filesystem under path '/user/venusp7663227/b' is HEALTHY
[venusp7663227@cxln4 ~]$
```

21. Touchz: this command creates a new file in the specified directory of size 0.
hadoop fs -touchz <HDFS file path>

```
[venusp7663227@cxln4 ~]$ hadoop fs -touchz bri
[venusp7663227@cxln4 ~]$
```

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

hadoop fs -test <HDFS file path>

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

hadoop fs -text <HDFS file path>

```
[venusp7663227@cxln4 ~]$ hadoop fs -text b
Hello from Brijesh!
[venusp7663227@cxln4 ~]$
```

24. 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'.

```
[venusp7663227@cxln4 ~]$ hadoop fs -stat b
2024-09-24 17:18:19
[venusp7663227@cxln4 ~]$
```

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

hadoop fs -usage <command>

```
[venusp7663227@cxln4 ~]$ hadoop fs -usage ls
Usage: hadoop fs [generic options] -ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [<path> ...]
[venusp7663227@cxln4 ~]$
```

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

hadoop fs -help <command>

```
[venusp7663227@cxln4 ~]$ hadoop fs -help ls
-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [<path> ...] :
  List the contents that match the specified file pattern. If path is not
  specified, the contents of /user/<currentUser> will be listed. For a directory a
  list of its direct children is returned (unless -d option is specified).

  Directory entries are of the form:
        permissions - userId groupId sizeOfDirectory(in bytes)
  modificationDate(yyyy-MM-dd HH:mm) directoryName

  and file entries are of the form:
        permissions numberOfReplicas userId groupId sizeOfFile(in bytes)
  modificationDate(yyyy-MM-dd HH:mm) fileName

    -C  Display the paths of files and directories only.
    -d  Directories are listed as plain files.
```

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

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

```
[venusp7663227@cxln4 ~]$ hadoop fs -chmod -r demo.txt
[venusp7663227@cxln4 ~]$
```

28. 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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -appendToFile b.txt demo.txt b
[venusp7663227@cxln4 ~]$ hadoop fs -cat b
Hello from Brijesh!
Hello Brijesh here!
:
[venusp7663227@cxln4 ~]$
```

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

hadoop fs -checksum <HDFS file Path>

```
[venusp7663227@cxln4 ~]$ hadoop fs -checksum b
b       MD5-of-0MD5-of-512CRC32C        000002000000000000000002175b4a5d6c3547e083dacb1b45c4e45
[venusp7663227@cxln4 ~]$
```

30. 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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -count b
           0            1                 42 b
[venusp7663227@cxln4 ~]$
```

31. 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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -find b
b
[venusp7663227@cxln4 ~]$
```

32. 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>

```
[venusp7663227@cxln4 ~]$ hadoop fs -getmerge b demo.txt
[venusp7663227@cxln4 ~]$ hadoop fs -cat b
Hello from Brijesh!
Hello Brijesh here!
:
[venusp7663227@cxln4 ~]$ cat demo.txt
Hello from Brijesh!
Hello Brijesh here!
:
[venusp7663227@cxln4 ~]$
```