

## Exercise 02: File management tasks in Hadoop

**This exercise is try to explain the command which is used to do file management tasks in HDFS (Hadoop).**

**1. Find the HDFS Command to check the health of the Hadoop file system.**

**Command:** `hdfs fsck /`

DEPRECATED: Use of this script to execute hdfs command is deprecated.  
Instead use the hdfs command for it.

```
Connecting to namenode via http://quickstart.cloudera:50070/fsck?ugi=cloudera&pa
th=%2F
FSCK started by cloudera (auth:SIMPLE) from /10.0.2.15 for path / at Mon Aug 09
02:17:27 PDT 2021
```

[illegible]

```
.....Status: HEALTHY
Total size: 1020670279 B (Total open files size: 166 B)
Total dirs: 199
Total files: 999
Total symlinks: 0 (Files currently being written: 3)
```

```

Total blocks (validated):      992 (avg. block size 1028901 B) (Total open file
blocks (not validated): 2)
Minimally replicated blocks:  992 (100.0 %)
Over-replicated blocks:      0 (0.0 %)
Under-replicated blocks:     0 (0.0 %)
Mis-replicated blocks:       0 (0.0 %)
Default replication factor:    1
Average block replication:    1.0
Corrupt blocks:               0
Missing replicas:             0 (0.0 %)
Number of data-nodes:         1
Number of racks:              1
FSCK ended at Mon Aug 09 02:17:27 PDT 2021 in 253 milliseconds

```

```
The filesystem under path '/' is HEALTHY
[cloudera@quickstart ~]$
```

**2. Attach the output screen shot for HDFS Command to display the list of Files and Directories in HDFS.**

**Command:** `hdfs dfs -ls /`

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
```

Found 6 items

drwxrwxrwx	-	hdfs	supergroup	0	2017-10-23 09:15	/benchmarks
drwxr-xr-x	-	hbase	supergroup	0	2022-07-29 00:47	/hbase
drwxr-xr-x	-	solr	solr	0	2017-10-23 09:18	/solr
drwxrwxrwt	-	hdfs	supergroup	0	2022-07-29 06:46	/tmp
drwxr-xr-x	-	hdfs	supergroup	0	2017-10-23 09:17	/user
drwxr-xr-x	-	hdfs	supergroup	0	2017-10-23 09:17	/var

### 3. Create HDFS Directory with the name 'temp'

**Command:** `hdfs dfs -mkdir /temp`

```

Found 12 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2021-08-09 01:48 /hbase
drwxr-xr-x - cloudera supergroup 0 2020-12-27 03:17 /infold01
drwxr-xr-x - cloudera supergroup 0 2020-12-27 03:23 /out1
drwxr-xr-x - cloudera supergroup 0 2020-12-27 03:14 /outfold01
drwxr-xr-x - cloudera supergroup 0 2020-12-27 04:43 /pig
-rw-r--r-- 1 cloudera supergroup 102 2020-12-27 15:05 /pigfile.txt
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxr-xr-x - cloudera supergroup 0 2021-08-09 02:27 /temp
drwxrwxrwt - hdfs supergroup 0 2021-01-20 21:49 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ █

```

#### 4. Create file in HDFS with 'touchz' and show the size of the created file. Attach the output screen shot.

**Command:** `hdfs dfs -touchz /temp/sample00`

```

[cloudera@quickstart ~]$ hdfs dfs -touchz /temp/sample00
[cloudera@quickstart ~]$ hdfs dfs -du /
0          0          /benchmarks
7          7          /hbase
0          0          /solr
0          0          /temp
0          0          /tmp
861282820  861282820  /user
0          0          /var

```

#### 5. Find the HDFS command for show file size in HDFS folder

**Command:** `hdfs dfs -du /temp`

```

0          0          /benchmarks
9207       268444497  /hbase
72         72         /infold01
41         41         /out1
0          0          /outfold01
0          0          /pig
102        102        /pigfile.txt
0          0          /solr
0          0          /temp
158974866  158974866  /tmp
861283067  861283067  /user
403090     403090     /var

```

```

0 0 /temp/sample00
[cloudera@quickstart ~]$ █

```

#### 6. Create file in local system using Linux command and display the content in that created file.

**Command:** `cat > te`

```

[cloudera@quickstart ~]$ cat > te
welcome to Big data world
^C
[cloudera@quickstart ~]$ cat te
welcome to Big data world _

```

## 7. Find HDFS command for copy locally created file into HDFS file system

**Command:** `hdfs dfs -copyFromLocal te /temp`

`hdfs dfs -ls /temp`

```
[cloudera@quickstart ~]$ ls
cloudera-manager  eclipse      lib          pig_1609071035413.log  pig_1609073285453.log  pig_1611207545890.log  Templates
cm_api.py         enterprise-deployment.json  Music        pig_1609071868480.log  pig_1609073645121.log  pigfile.txt           Videos
Desktop          expig       parcels      pig_1609072460700.log  pig_1609110421108.log  Public                WordCount.jar
Documents        express-deployment.json    pfile.txt   pig_1609072756374.log  pig_1609111318432.log  te                    workspace
Downloads        kerberos    Pictures     pig_1609073082983.log  pig_1611186283740.log  temp

21/08/09 03:00:03 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1281)
    at java.lang.Thread.join(Thread.java:1355)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
[cloudera@quickstart ~]$ hdfs dfs -ls /temp/
Found 2 items
-rw-r--r--  1 cloudera supergroup          0 2021-08-09 02:31 /temp/sample00
-rw-r--r--  1 cloudera supergroup        24 2021-08-09 03:00 /temp/te
[cloudera@quickstart ~]$
```

## 8. Copy locally created file into HDFS file system using 'put' command

**Command:** `hdfs dfs -put temp /temp/`

```
[cloudera@quickstart ~]$ hdfs dfs -ls /temp
Found 2 items
-rw-r--r--  1 cloudera supergroup          0 2022-07-31 09:20 /temp/sample00
-rw-r--r--  1 cloudera supergroup        26 2022-07-31 09:30 /temp/te
[cloudera@quickstart ~]$ cat temp
welcome to the department of data science
[cloudera@quickstart ~]$ hdfs dfs -put temp /temp/
[cloudera@quickstart ~]$ hdfs dfs -ls /temp
ls: `/temp': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -ls /temp
Found 3 items
-rw-r--r--  1 cloudera supergroup          0 2022-07-31 09:20 /temp/sample00
-rw-r--r--  1 cloudera supergroup        26 2022-07-31 09:30 /temp/te
-rw-r--r--  1 cloudera supergroup        42 2022-07-31 09:39 /temp/temp
[cloudera@quickstart ~]$
```

## 9. Remove Local file using Linux command

**Command:**

```
[cloudera@quickstart ~]$ rm temp
rm: remove regular file `temp'? y
[cloudera@quickstart ~]$ ls
cloudera-manager  Desktop  Downloads  enterprise-deployment.json  kerberos  Music  Pictures  te  Templates  Videos
cm_api.py         Documents  eclipse    express-deployment.json    lib        parcels  Public    temp  temp.txt  workspace

cloudera-manager  eclipse      lib          pig_1609071035413.log  pig_1609073285453.log  pig_1611207545890.log  Videos
cm_api.py         enterprise-deployment.json  Music        pig_1609071868480.log  pig_1609073645121.log  pigfile.txt           WordCount.jar
Desktop          expig       parcels      pig_1609072460700.log  pig_1609110421108.log  Public                workspace
Documents        express-deployment.json    pfile.txt   pig_1609072756374.log  pig_1609111318432.log  temp
Downloads        kerberos    Pictures     pig_1609073082983.log  pig_1611186283740.log  Templates

cloudera-manager  eclipse      lib          pig_1609071035413.log  pig_1609073285453.log  pig_1611207545890.log  WordCount.jar
cm_api.py         enterprise-deployment.json  Music        pig_1609071868480.log  pig_1609073645121.log  pigfile.txt           workspace
Desktop          expig       parcels      pig_1609072460700.log  pig_1609110421108.log  Public
Documents        express-deployment.json    pfile.txt   pig_1609072756374.log  pig_1609111318432.log  Templates
Downloads        kerberos    Pictures     pig_1609073082983.log  pig_1611186283740.log  Videos
[cloudera@quickstart ~]$
```

## 10. Copy HDFS file to Local File System using copyToLocal command

**Command:** `hdfs dfs -copyToLocal /temp/test /home/cloudera/`  
 [cloudera@quickstart ~]\$ `cat > test`  
 Helooooo  
 ^C  
 [cloudera@quickstart ~]\$ `hdfs dfs -copyFromLocal test /temp/`  
 [cloudera@quickstart ~]\$ `hdfs dfs -ls /temp`  
 Found 4 items  
 -rw-r--r-- 1 cloudera supergroup 0 2022-07-31 09:20 /temp/sample00  
 -rw-r--r-- 1 cloudera supergroup 26 2022-07-31 09:30 /temp/te  
 -rw-r--r-- 1 cloudera supergroup 42 2022-07-31 09:39 /temp/temp  
 -rw-r--r-- 1 cloudera supergroup 9 2022-07-31 09:46 /temp/test

## 11. Find another way to copy HDFS files into Local File System.

**Command:** `hdfs dfs -get /temp/sample00 /home/cloudera/`

```

cloudera-manager  eclipse  lib  pig_1609071035413.log  pig_1609073285453.log  pig_1611207545890.log  WordCount.jar
cm_api.py        enterprise-deployment.json  Music  pig_1609071868480.log  pig_1609073645121.log  pigfile.txt  workspace
Desktop          expig  parcels  pig_1609072460700.log  pig_1609110421108.log  Public
Documents        express-deployment.json  pfile.txt  pig_1609072756374.log  pig_1609111318432.log  Templates
Downloads        kerberos  Pictures  pig_1609073082983.log  pig_1611186283740.log  Videos

cloudera-manager  eclipse  lib  pig_1609071035413.log  pig_1609073285453.log  pig_1611207545890.log  Templates
cm_api.py        enterprise-deployment.json  Music  pig_1609071868480.log  pig_1609073645121.log  pigfile.txt  Videos
Desktop          expig  parcels  pig_1609072460700.log  pig_1609110421108.log  Public
Documents        express-deployment.json  pfile.txt  pig_1609072756374.log  pig_1609111318432.log  te  WordCount.jar
Downloads        kerberos  Pictures  pig_1609073082983.log  pig_1611186283740.log  temp  workspace
[cloudera@quickstart ~]$ █

```

## 12. Count number of files in particular HDFS directory using 'count' command

```

cloudera@quickstart ~]$ hdfs dfs -count /temp
1 4 77 /temp

```

## 13. Remove the entire directory and all if its content from HDFS

**Command:** `hdfs dfs -rmrmdir /demo`

`hdfs dfs -ls /demo`

```

16/10/20 11:54:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
16/10/20 11:54:54 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval =
0 minutes, Emptier interval = 0 minutes.
Deleted /new edureka
[edureka@localhost ~]$ █

```

## 14. Copy HDFS file from one folder to another folder using 'cp' command **Command:**

`hdfs dfs -cp /user/hadoop/file1 /user/hadoop/file2 /user/hadoop/dir`

```

[cloudera@quickstart ~]$ hdfs dfs -cp /tmp/ /temp/

[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hdfs dfs -ls /temp
Found 6 items
drwxr-xr-x - cloudera supergroup 0 2022-07-31 09:58 /temp/new
-rw-r--r-- 1 cloudera supergroup 0 2022-07-31 09:20 /temp/sample00
-rw-r--r-- 1 cloudera supergroup 26 2022-07-31 09:30 /temp/te
-rw-r--r-- 1 cloudera supergroup 42 2022-07-31 09:39 /temp/temp
-rw-r--r-- 1 cloudera supergroup 9 2022-07-31 09:46 /temp/test
drwxr-xr-x - cloudera supergroup 0 2022-07-31 09:57 /temp/tmp

```

## 15. Write the HDFS command to move files form source to destination.

**Command:** `hdfs dfs -mkdir /new`

`hdfs dfs -mv /new//temp/`

`hdfs dfs -ls /temp`

```
16/10/20 12:20:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$ hdfs dfs -ls /user
16/10/20 12:20:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
Found 5 items
drwxrwxrwx - edureka supergroup          0 2016-10-12 09:46 /user/edureka
drwxrwxrwx - edureka supergroup          0 2014-09-24 18:27 /user/flume
drwxr-xr-x - edureka supergroup          0 2016-10-20 12:20 /user/hadoop
-rw-r--r-- 1 edureka supergroup        58 2016-10-20 12:14 /user/test
-rw-r--r-- 1 edureka supergroup          0 2016-10-20 12:18 /user/test1
[edureka@localhost ~]$
```

## 16. Catch HDFS command for empty the trash

**Command:** `hdfs dfs -expunge`

```
16/10/20 11:59:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
16/10/20 11:59:14 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval =
0 minutes, Emptier interval = 0 minutes.
[edureka@localhost ~]$
```

## 17. Remove HDFS directory using 'rmdir' command

```
[cloudera@quickstart ~]$ hdfs dfs -rmdir /temp
rmdir: `/temp': Directory is not empty
[cloudera@quickstart ~]$ hdfs dfs -mkdir /temp00
hd[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 9 items
drwxrwxrwx - hdfs      supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - cloudera  supergroup          0 2022-07-31 09:54 /demo
drwxr-xr-x - hbase     supergroup          0 2022-07-29 00:47 /hbase
drwxr-xr-x - solr      solr              0 2017-10-23 09:18 /solr
drwxr-xr-x - cloudera  supergroup          0 2022-07-31 09:58 /temp
drwxr-xr-x - cloudera  supergroup          0 2022-07-31 10:02 /temp00
drwxrwxrwt - hdfs      supergroup          0 2022-07-29 06:46 /tmp
drwxr-xr-x - hdfs      supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs      supergroup          0 2017-10-23 09:17 /var
[cloudera@quickstart ~]$ hdfs dfs -rmdir /temp00
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxrwxrwx - hdfs      supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - cloudera  supergroup          0 2022-07-31 09:54 /demo
drwxr-xr-x - hbase     supergroup          0 2022-07-29 00:47 /hbase
drwxr-xr-x - solr      solr              0 2017-10-23 09:18 /solr
drwxr-xr-x - cloudera  supergroup          0 2022-07-31 09:58 /temp
drwxrwxrwt - hdfs      supergroup          0 2022-07-29 06:46 /tmp
```

## 18. How to returns the help for an individual command in HDFS



**Command:** `hdfs dfs -help`

Usage: `hadoop fs [generic options] -mkdir [-p] <path> ...`

Usage: `hadoop fs [generic options] -put [-f] [-p] [-l] <localsrc> ... <dst>`  
[cloudera@quickstart ~]\$

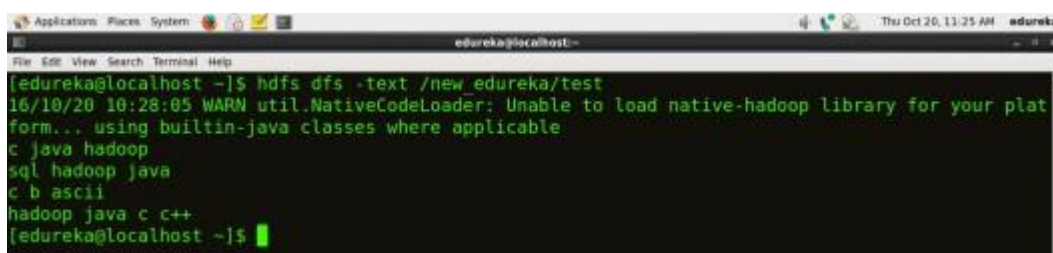
## 19. Find HDFS Command that displays help for given command or all commands if none is specified

**Command:** `hdfs dfs -help`

```
Usage: hadoop fs [generic options]
[-appendToFile <localsrc> ... <dst>]
[-cat [-ignoreCrc] <src> ...]
[-checksum <src> ...]
[-chgrp [-R] GROUP PATH...]
[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
[-chown [-R] [OWNER][:[GROUP]] PATH...]
[-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>]
[-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
[-count [-q] [-h] [-v] [-x] <path> ...]
[-cp [-f] [-p | -p[topax]] <src> ... <dst>]
[-createSnapshot <snapshotDir> [<snapshotName>]]
[-deleteSnapshot <snapshotDir> <snapshotName>]
[-df [-h] [<path> ...]]
[-du [-s] [-h] [-x] <path> ...]
[-expunge]
[-find <path> ... <expression> ...]
[-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
[-getfacl [-R] <path>]
[-getfattr [-R] {-n name | -d} [-e en] <path>]
[-getmerge [-nl] <src> <localdst>]
[-help [cmd ...]]
[-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [<path> ...]]
[-mkdir [-p] <path> ...]
[-moveFromLocal <localsrc> ... <dst>]
[-moveToLocal <src> <localdst>]
[-mv <src> ... <dst>]
[-put [-f] [-p] [-l] <localsrc> ... <dst>]
[-renameSnapshot <snapshotDir> <oldName> <newName>]
[-rm [-f] [-r|-R] [-skipTrash] <src> ...]
[-rmkdir [--ignore-fail-on-non-empty] <dir> ...]
```

## 20. Using HDFS command takes a source file and text formatted output file.

**Command:** `hdfs dfs -text /new_edureka/test`



```
Applications Places System Thu Oct 20, 11:25 AM edureka
edureka@localhost:~$ hdfs dfs -text /new_edureka/test
16/10/20 10:28:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
c java hadoop
sql hadoop java
c b ascii
hadoop java c c++
[edureka@localhost ~]$
```