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
   .....................................................................
   ...................
   .....................................................................
   ...................
   .....................................................................
   ...................
   .....................................................................
   ...................
   .....................................................................
   ...................
   .....................................................................
   ...................
   .....................................................................
   ...................
   .....................................................................
   ...................
   .....................................................................
   ...................
   ...................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 /**

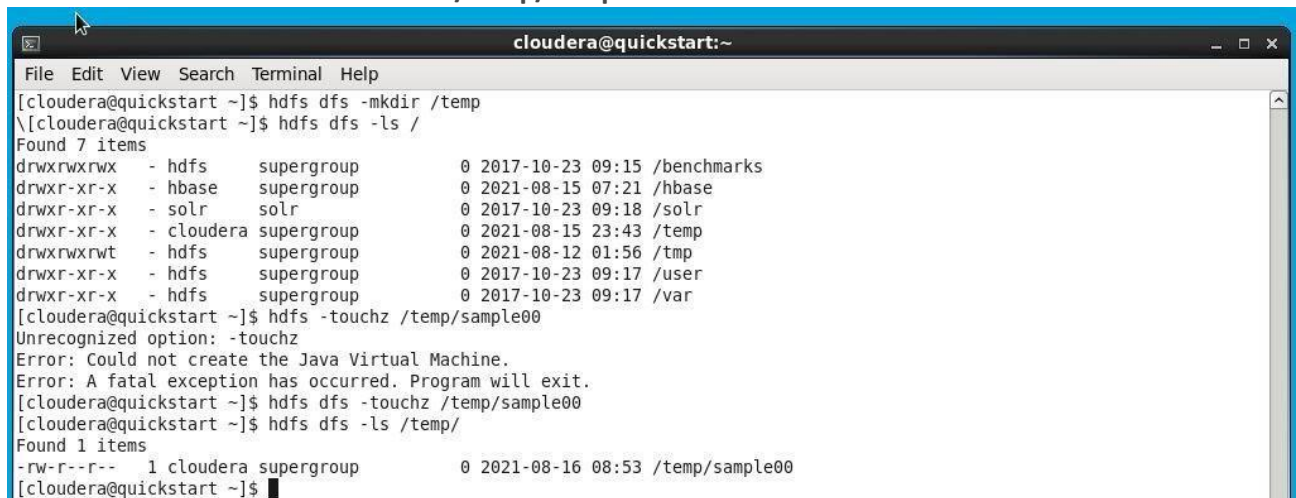   Paste the output screen shot

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:~                          _ □ ✕
File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ hdfs dfs -mkdir /temp
\[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2021-08-15 07:21 /hbase
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxr-xr-x   - cloudera supergroup          0 2021-08-15 23:43 /temp
drwxrwxrwt   - hdfs     supergroup          0 2021-08-12 01:56 /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 -touchz /temp/sample00
Unrecognized option: -touchz
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
[cloudera@quickstart ~]$ hdfs dfs -touchz /temp/sample00
[cloudera@quickstart ~]$ hdfs dfs -ls /temp/
Found 1 items
-rw-r--r--   1 cloudera supergroup          0 2021-08-16 08:53 /temp/sample00
[cloudera@quickstart ~]$ █
```

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

   Command: *hdfs dfs –df /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
```
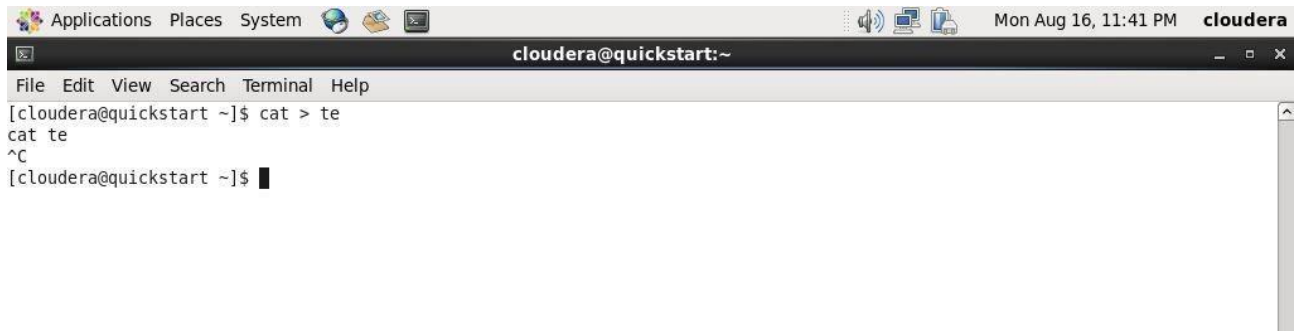
6. **Create file in local system using Linux command and display the content in that created file.**

   Command: **cat > te**

   Command: **cat te**

```
[cloudera@quickstart ~]$ cat > te
cat te
^C
[cloudera@quickstart ~]$
```

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

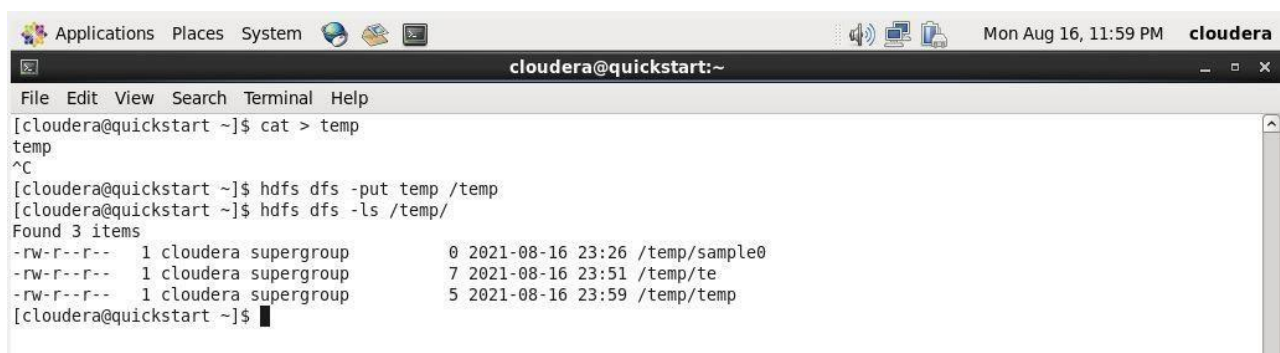   *Command:* *hfds dfs –copyFromLocal Te /temp/*

```
[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 ~]$ cat > temp
temp
^C
[cloudera@quickstart ~]$ hdfs dfs -put temp /temp
[cloudera@quickstart ~]$ hdfs dfs -ls /temp/
Found 3 items
-rw-r--r--   1 cloudera supergroup          0 2021-08-16 23:26 /temp/sample0
-rw-r--r--   1 cloudera supergroup          7 2021-08-16 23:51 /temp/te
-rw-r--r--   1 cloudera supergroup          5 2021-08-16 23:59 /temp/temp
[cloudera@quickstart ~]$
```

9. **Remove Local file using Linux command**

   Command: rm temp

```
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 ~]$ ls
cloudera-manager  Documents  enterprise-deployment.json  lib      Pictures  Templates  test1  workspace
cm_api.py         Downloads  express-deployment.json     Music    Public    tempp      test2
Desktop           eclipse    kerberos                    parcels  te        test       Videos
[cloudera@quickstart ~]$ rm test2
rm: remove regular file `test2'? y
[cloudera@quickstart ~]$ hdfs dfs -copyToLocal /temp1/test2 /home/cloudera/
[cloudera@quickstart ~]$ ls
cloudera-manager  Documents  enterprise-deployment.json  lib      Pictures  Templates  test1  workspace
cm_api.py         Downloads  express-deployment.json     Music    Public    tempp      test2
Desktop           eclipse    kerberos                    parcels  te        test       Videos
[cloudera@quickstart ~]$
```

`cloudera@quickstart:~`

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

    Command: hdfs dfs –get /temp/sample0 /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                 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
[cloudera@quickstart ~]$
```

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

    Command: **hdfs dfs –count /temp**

```
Applications  Places  System                          Tue Aug 17, 12:27 AM   cloudera
                            cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ hdfs dfs -count /temp
        1            3              12 /temp
[cloudera@quickstart ~]$
```

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

Command: *hadoop fs –rm –r /book*

```
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 -mkdir /tweet
[cloudera@quickstart ~]$ hdfs dfs -cp /tweet /temp/
[cloudera@quickstart ~]$ hdfs dfs -ls /temp
Found 3 items
-rw-r--r--   1 cloudera supergroup          7 2021-08-16 23:51 /temp/te
-rw-r--r--   1 cloudera supergroup          5 2021-08-16 23:59 /temp/temp
drwxr-xr-x   - cloudera supergroup          0 2021-08-19 06:08 /temp/tweet
[cloudera@quickstart ~]$
```

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

Command: *hdfs dfs –mkdir /tw*

hdfs dfs –mv /tw /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

Command: hdfs dfs –rmdir /temp

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /temp2
[cloudera@quickstart ~]$ hdfs dfs -rmdir /temp2
[cloudera@quickstart ~]$ █
    cloudera@quickstart:~
```

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 ~]$ █
```
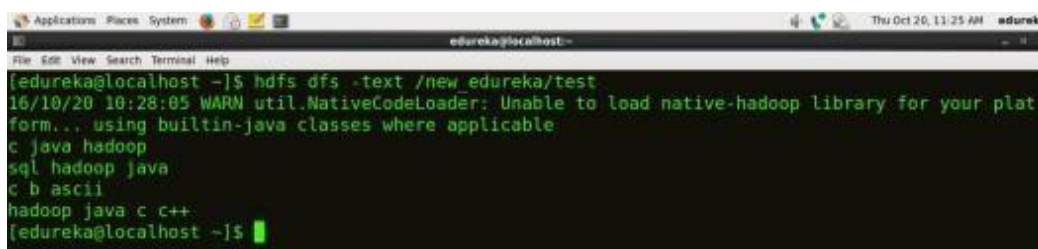
19. Find HDFS Command that displays help for given command or all commands if none is specified

Command: hdfs dfs -usage

```
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> ...]
        [-rmdir [--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