

BDMA_EXERCISE 02 : FILE MANAGEMENT TASKS IN HADOOP

NAME : ANNAPOORNIMA S

ROLL NO : 225229101

1) [cloudera@quickstart ~]\$ hdfs fsck /

Connecting to namenode via http://quickstart.cloudera:50070/fsck?ugi=cloudera&path=%2F

FSCK started by cloudera (auth:SIMPLE) from /127.0.0.1 for path / at Thu Jul 13 20:28:07 PDT 2023

.....
.....
.....
.....
.....
.....
.....
.....
.....

.....Status: HEALTHY

Total size: 858123172 B (Total open files size: 166 B)

Total dirs: 79

Total files: 931

Total symlinks: 0 (Files currently being written: 3)

Total blocks (validated): 929 (avg. block size 923706 B) (Total open file blocks (not validated):2)

Minimally replicated blocks: 929 (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 Thu Jul 13 20:28:07 PDT 2023 in 145 milliseconds

The filesystem under path '/' is HEALTHY

2) cloudera@quickstart ~]\$ hdfs dfs -ls /

Found 6 items

```
drwxrwxrwx - hdfs supergroup    0 2017-07-19 05:34 /benchmarks
drwxr-xr-x - hbase supergroup    0 2023-07-13 20:16 /hbase
drwxr-xr-x - solr solr          0 2017-07-19 05:37 /solr
drwxrwxrwt - hdfs supergroup    0 2023-07-13 20:17 /tmp
drwxr-xr-x - hdfs supergroup    0 2017-07-19 05:36 /user
drwxr-xr-x - hdfs supergroup    0 2017-07-19 05:36 /var
```

3) [cloudera@quickstart ~]\$ hdfs dfs -ls /temp/

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

Found 7 items

```
drwxrwxrwx - hdfs supergroup    0 2017-07-19 05:34 /benchmarks
drwxr-xr-x - hbase supergroup    0 2023-07-13 20:16 /hbase
drwxr-xr-x - solr solr          0 2017-07-19 05:37 /solr
drwxr-xr-x - cloudera supergroup 0 2023-07-13 20:36 /temp
drwxrwxrwt - hdfs supergroup    0 2023-07-13 20:17 /tmp
drwxr-xr-x - hdfs supergroup    0 2023-07-13 20:33 /user
drwxr-xr-x - hdfs supergroup    0 2017-07-19 05:36 /var
```

4) [cloudera@quickstart ~]\$ hdfs dfs -du /temp/sample00

0 0 /temp/sample00

5) [cloudera@quickstart ~]\$ hadoop fs -du /temp/

0 0 /temp/sample00

6) [cloudera@quickstart ~]\$ cat > te

i am annapoornima

^C

[cloudera@quickstart ~]\$ cat te

i am annapoornima

8) [cloudera@quickstart ~]\$ hdfs dfs -ls /temp/

Found 1 items

-rw-r--r-- 1 cloudera supergroup 0 2023-07-13 20:40 /temp/sample00

9) [cloudera@quickstart ~]\$ hdfs dfs -put te

[cloudera@quickstart ~]\$ hadoop fs -ls

Found 1 items

-rw-r--r-- 1 cloudera cloudera 12 2023-07-13 21:02 te

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

Found 7 items

drwxrwxrwx - hdfs supergroup 0 2017-07-19 05:34 /benchmarks

drwxr-xr-x - hbase supergroup 0 2023-07-13 20:16 /hbase

drwxr-xr-x - solr solr 0 2017-07-19 05:37 /solr

drwxr-xr-x - cloudera supergroup 0 2023-07-13 20:40 /temp

drwxrwxrwt - hdfs supergroup 0 2023-07-13 20:17 /tmp

drwxr-xr-x - hdfs supergroup 0 2023-07-13 20:33 /user

drwxr-xr-x - hdfs supergroup 0 2017-07-19 05:36 /var

[cloudera@quickstart ~]\$ rm te

rm: remove regular empty file `te'? yes

```
[cloudera@quickstart ~]$ hadoop fs -ls
```

Found 2 items

```
-rw-r--r--  1 cloudera cloudera    12 2023-07-13 21:02 te
-rw-r--r--  1 cloudera cloudera    12 2023-07-13 21:08 temp
```

```
10) [cloudera@quickstart ~]$ hdfs dfs -get /temp /home/cloudera/
```

get: `/home/cloudera/temp/sample00': File exists

```
11) [cloudera@quickstart ~]$ hdfs dfs -copyToLocal /temp /home/cloudera/
```

get: `/home/cloudera/temp/sample00': File exists

```
12) [cloudera@quickstart ~]$ hdfs dfs -count /temp
```

```
      1      1      0 /temp
```

```
13) [cloudera@quickstart ~]$ hdfs dfs -rm -r /temp
```

Deleted /temp

```
14) [cloudera@quickstart ~]$ hdfs dfs -cp /user/hue /user/hive
```

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

Found 9 items

```
drwxr-xr-x  - cloudera cloudera    0 2023-07-13 21:08 /user/cloudera
drwxr-xr-x  - mapred  hadoop       0 2017-07-19 05:34 /user/history
drwxrwxrwx  - hive   supergroup    0 2023-07-13 21:23 /user/hive
drwxrwxrwx  - hue    supergroup    0 2017-07-19 05:35 /user/hue
drwxrwxrwx  - jenkins supergroup    0 2017-07-19 05:35 /user/jenkins
drwxrwxrwx  - oozie  supergroup    0 2017-07-19 05:35 /user/oozie
drwxrwxrwx  - root   supergroup    0 2017-07-19 05:35 /user/root
drwxr-xr-x  - hdfs   supergroup    0 2017-07-19 05:36 /user/spark
drwxr-xr-x  - cloudera supergroup    0 2023-07-13 20:33 /user/temp
```

16) [cloudera@quickstart ~]\$ hdfs dfs -rm -skipTrash /user/temp

rm: `/user/temp': Is a directory

17) [cloudera@quickstart ~]\$ hdfs dfs -rmdir /temp

rmdir: `/temp': No such file or directory

18) [cloudera@quickstart ~]\$ hadoop fs -help touchz

-touchz <path> ... :

Creates a file of zero length at <path> with current time as the timestamp of that <path>. An error is returned if the file exists with non-zero length

19) hadoop fs

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> ...]
 [-setfacl [-R] [{-b|-k} {-m|-x <acl_spec>} <path>]| [--set <acl_spec> <path>]]
 [-setfattr {-n name [-v value] | -x name} <path>]
 [-setrep [-R] [-w] <rep> <path> ...]
 [-stat [format] <path> ...]
 [-tail [-f] <file>]
 [-test [-defsz] <path>]
 [-text [-ignoreCrc] <src> ...]
 [-touchz <path> ...]
 [-usage [cmd ...]]

Generic options supported are

- conf <configuration file> specify an application configuration file
- D <property=value> use value for given property
- fs <local|namenode:port> specify a namenode
- jt <local|resourcemanager:port> specify a ResourceManager
- files <comma separated list of files> specify comma separated files to be copied to the map reduce cluster
- libjars <comma separated list of jars> specify comma separated jar files to include in the classpath.

-archives <comma separated list of archives> specify comma separated archives to be unarchived on the compute machines.

The general command line syntax is

bin/hadoop command [genericOptions] [commandOptions]

20) [cloudera@quickstart ~]\$ hdfs dfs -text /new_edureka/test

text: `/new_edureka/test': No such file or directory