

Name: Chloe Tee Rouyi

Student ID: 0354731

Section: 04

a) fsck (HDFS Command to check the health of the Hadoop file system):

i) **Command: hdfs fsck /**

```
[cloudera@quickstart ~]$ hdfs fsck /
Connecting to namenode via http://quickstart.cloudera:50070
FSCK started by cloudera (auth:SIMPLE) from /192.168.56.101 for path / at Mon Oct 14 01:59:19 PDT 2024
..
/hbase/data/hbase/meta/.tabledesc/.tableinfo.0000000001: CORRUPT blockpool BP-171906535-127.0.0.1-1491391569011 block blk_1073742696

/hbase/data/hbase/meta/.tabledesc/.tableinfo.0000000001: MISSING 1 blocks of total size 372 B..
/hbase/data/hbase/meta/1588230740/.regioninfo: CORRUPT blockpool BP-171906535-127.0.0.1-1491391569011 block blk_1073742695

/hbase/data/hbase/meta/1588230740/.regioninfo: MISSING 1 blocks of total size 32 B...
/hbase/hbase.id: CORRUPT blockpool BP-171906535-127.0.0.1-1491391569011 block blk_1073742694

/hbase/hbase.id: MISSING 1 blocks of total size 42 B..
/hbase/hbase.version: CORRUPT blockpool BP-171906535-127.0.0.1-1491391569011 block blk_1073742693

/hbase/hbase.version: MISSING 1 blocks of total size 7 B.....
.....
.....
.....
s: CORRUPT
Total size:      800475450 B
Total dirs:      61
Total files:     875
Total symlinks:   0
Total blocks (validated): 873 (avg. block size 916924 B)
*****
UNDER MIN REPL'D BLOCKS: 4 (0.45819014 %)
dfs.namenode.replication.min: 1
CORRUPT FILES:      4
MISSING BLOCKS:     4
MISSING SIZE:       453 B
CORRUPT BLOCKS:     4
*****
Minimally replicated blocks: 869 (99.54181 %)
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:   0.9954181
Corrupt blocks:              4
Missing replicas:            0 (0.0 %)
Number of data-nodes:        1
Number of racks:             1
FSCK ended at Mon Oct 14 01:59:20 PDT 2024 in 279 milliseconds

The filesystem under path '/' is CORRUPT
```

b) ls (HDFS Command to display the list of Files and Directories in HDFS):

i) Command: `hdfs dfs -ls /`

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx - hdfs supergroup 0 2017-04-05 04:27 /benchmarks
drwxr-xr-x - hbase supergroup 0 2024-09-27 05:50 /hbase
drwxr-xr-x - cloudera supergroup 0 2024-10-08 19:46 /mydata
drwxr-xr-x - solr solr 0 2017-04-05 04:29 /solr
drwxrwxrwt - hdfs supergroup 0 2017-04-05 04:29 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-04-05 04:29 /user
drwxr-xr-x - hdfs supergroup 0 2017-04-05 04:29 /var
```

c) help (HDFS Command that displays help for given command or all commands if none is specified):

i) Command: `hdfs dfs -help`

```
[cloudera@quickstart ~]$ 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>]
```

Task 5:

- a) Create a directory with name “mydata” in the hdfs

i) **hdfs dfs -mkdir /mydata**

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /mydata2
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxrwxrwx - hdfs supergroup 0 2017-04-05 04:27 /benchmarks
drwxr-xr-x - hbase supergroup 0 2024-09-27 05:50 /hbase
drwxr-xr-x - cloudera supergroup 0 2024-10-08 19:46 /mydata
drwxr-xr-x - cloudera supergroup 0 2024-10-14 02:49 /mydata2
drwxr-xr-x - solr solr 0 2017-04-05 04:29 /solr
drwxrwxrwt - hdfs supergroup 0 2017-04-05 04:29 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-04-05 04:29 /user
drwxr-xr-x - hdfs supergroup 0 2017-04-05 04:29 /var
```

Task 6:

- a) Copy a file from the dataset folder onto mydata folder in the HDFS:

i) **hdfs dfs -put /home/cloudera/Downloads/datasets/allsales/mydata2/**

```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Downloads/datasets/allsales /mydata2/
```

Task 7:

- a) Check the file if it is copied to HDFS storage:

i) **hdfs dfs -ls /mydata2/**

```
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata2/
Found 1 items
-rw-r--r-- 1 cloudera supergroup 86 2024-10-14 02:51 /mydata2/allsales
```

Task 8:

- a) View the content of the files on the local storage

i) **cat Downloads/datasets/allsales**

```
[cloudera@quickstart ~]$ cat Downloads/datasets/allsales
Alice 2999 us
Bob 3625 ca
Carlos 2764 mx
Dieter 1749 de
Étienne 2368 fr
```

Task 9:

- a) View the content of the copied file in mydata folder in the HDFS storage and compare it with the content of the original file to ensure they are the same.

i) **hdfs dfs -cat /mydata2/allsales**

```
Fredo 5637 it[cloudera@quickstart ~]$ hdfs dfs -cat /mydata2/allsales
Alice 2999 us
Bob 3625 ca
Carlos 2764 mx
Dieter 1749 de
Étienne 2368 fr
```

Task 10:

- a) Create an empty file in mydata folder in hdfs:

i) **hdfs dfs -touchz /mydata2/test1.txt**

```
Fredo 5637 it[cloudera@quickstart ~]$ hdfs dfs -touchz /mydata2/test1.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata2
Found 2 items
-rw-r--r-- 1 cloudera supergroup 86 2024-10-14 02:51 /mydata2/allsales
-rw-r--r-- 1 cloudera supergroup 0 2024-10-14 03:00 /mydata2/test1.txt
```

Task 11:

a) Add content (append) to a file in hdfs

- i) **hdfs dfs -appendToFile - /mydata3/test1.txt (to append text into the file)**
- ii) **hdfs dfs -ls /mydata3 (check files in mydata3 to make sure its created)**
- iii) **hdfs dfs -cat /mydata3/test1.txt (to check contents in the file)**

```
[cloudera@quickstart ~]$ hdfs dfs -appendToFile - /mydata3/test1.txt
this is first line.
this is the second line.
12
12
1
2
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata3
Found 1 items
-rw-r--r--  1 cloudera supergroup      55 2024-10-15 19:27 /mydata3/test1.tx
t
[cloudera@quickstart ~]$ hdfs dfs -cat /mydata3/test1.txt
this is first line.
this is the second line.
12
12
1
2
[cloudera@quickstart ~]$ hdfs dfs -appendToFile - /mydata3/test1.txt
this is new update 1
new update 2
new update 3
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata3
Found 1 items
-rw-r--r--  1 cloudera supergroup     102 2024-10-15 19:28 /mydata3/test1.tx
t
[cloudera@quickstart ~]$ hdfs dfs -cat /mydata3/test1.txt
this is first line.
this is the second line.
12
12
1
2
this is new update 1
new update 2
new update 3
```

Task 12:

a) copy the new created file from hdfs to local file system

- i) **mkdir testDir2**
- ii) **hdfs dfs -get /mydata3/test1.txt /home/cloudera/testDir2/**
- iii) **ls /home/cloudera/testDir2/**

```
[cloudera@quickstart ~]$ mkdir testDir2
[cloudera@quickstart ~]$ ls
bdtsec03  cloudera-manager  Desktop  Downloads  enterprise-deployment.json  lib  parcels  Public  testdir  Videos
chloetee  cm_api.py          Documents  eclipse     kerberos  Music  Pictures  Templates  testDir2  worksp
[cloudera@quickstart ~]$ hdfs dfs -get /mydata3/test1.txt /home/cloudera/testDir2/
[cloudera@quickstart ~]$ ls /home/cloudera/testDir2/
test1.txt
[cloudera@quickstart ~]$ █
```

cloudera

b) Giving a new name:

- i) **hdfs dfs -get /mydata3/test1.txt /home/cloudera/testDir2/renamedTest1.txt**

```
[cloudera@quickstart ~]$ hdfs dfs -get /mydata3/test1.txt /home/cloudera/testDir2/renamedTest1.txt
[cloudera@quickstart ~]$ ls /home/cloudera/testDir2/
renamedTest1.txt  test1.txt
[cloudera@quickstart ~]$ █
```

Downloads

Task 13:

a) Create a new file in local storage

i) **cat >/home/cloudera/Downloads/test3.txt**

b) Then move it to hdfs

i) **hdfs dfs -moveFromLocal /home/cloudera/Downloads/test3.txt /mydata2**

```
[cloudera@quickstart ~]$ cat >/home/cloudera/Downloads/test3.txt
this is the first line
[cloudera@quickstart ~]$ cat >/home/cloudera/Downloads/test3.txt
this is the second line
```

```
[cloudera@quickstart ~]$ ls /home/cloudera/Downloads
datasets  datasets.zip  test1.txt  test3.txt
[cloudera@quickstart ~]$ hdfs dfs -moveFromLocal /home/cloudera/Downloads/test3.txt /mydata2
[cloudera@quickstart ~]$ ls Downloads/
datasets  datasets.zip  test1.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata2
Found 4 items
-rw-r--r--  1 cloudera supergroup      86 2024-10-14 02:51 /mydata2/allsales
-rw-r--r--  1 cloudera supergroup      43 2024-10-15 18:58 /mydata2/test1.txt
-rw-r--r--  1 cloudera supergroup       0 2024-10-15 19:11 /mydata2/test2.txt
-rw-r--r--  1 cloudera supergroup      24 2024-10-15 19:40 /mydata2/test3.txt
```

Task 14:

a) Create a new folder called 'myfolder'

i) **hdfs dfs -mkdir /myfolder**

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /myfolder
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 9 items
drwxrwxrwx - hdfs supergroup 0 2017-04-05 04:27 /benchmarks
drwxr-xr-x - hbase supergroup 0 2024-09-27 05:50 /hbase
drwxr-xr-x - cloudera supergroup 0 2024-10-08 19:46 /mydata
drwxr-xr-x - cloudera supergroup 0 2024-10-15 19:11 /mydata2
drwxr-xr-x - cloudera supergroup 0 2024-10-15 19:14 /myfolder
drwxr-xr-x - solr solr 0 2017-04-05 04:29 /solr
drwxrwxrwt - hdfs supergroup 0 2017-04-05 04:29 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-04-05 04:29 /user
drwxr-xr-x - hdfs supergroup 0 2017-04-05 04:29 /var
```

b) Use -cp and -mv

i) **hdfs dfs -cp /mydata2/test3.txt /mydata4/**

ii) **hdfs dfs -mv /mydata2/test2/test1.txt /mydata4**

```
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata2/
Found 4 items
-rw-r--r-- 1 cloudera supergroup 86 2024-10-14 02:51 /mydata2/allsales
-rw-r--r-- 1 cloudera supergroup 43 2024-10-15 18:58 /mydata2/test1.txt
-rw-r--r-- 1 cloudera supergroup 0 2024-10-15 19:11 /mydata2/test2.txt
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:40 /mydata2/test3.txt
[cloudera@quickstart ~]$ hdfs dfs -cp /mydata2/test3.txt /mydata4/
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata4/
Found 1 items
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:44 /mydata4/test3.txt
[cloudera@quickstart ~]$ hdfs dfs -cp /mydata2/test3.txt /mydata4/mm.txt
^[A^[[A^[[B[cloudera@quickstart ~]$ ls /mydata4/
Found 2 items
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:44 /mydata4/mm.txt
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:44 /mydata4/test3.txt
[cloudera@quickstart ~]$ hdfs dfs -mv /mydata2/test1.txt /mydata4
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata4/
Found 3 items
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:44 /mydata4/mm.txt
-rw-r--r-- 1 cloudera supergroup 43 2024-10-15 18:58 /mydata4/test1.txt
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:44 /mydata4/test3.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata2/
Found 3 items
-rw-r--r-- 1 cloudera supergroup 86 2024-10-14 02:51 /mydata2/allsales
-rw-r--r-- 1 cloudera supergroup 0 2024-10-15 19:11 /mydata2/test2.txt
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:40 /mydata2/test3.txt
```


Task 15:

a) Removing File

i) `hdfs dfs -rm /mydata4/mm.txt`

```
[cloudera@quickstart ~]$ hdfs dfs -rm /mydata4/mm.txt
Deleted /mydata4/mm.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /mydata4/
Found 2 items
-rw-r--r-- 1 cloudera supergroup 43 2024-10-15 18:58 /mydata4/test1.txt
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:44 /mydata4/test3.txt
```

Task 16:

a) Removing folder (removing recursively) -rmr

i) `hdfs dfs -rm /mydata4/`

```
^[[A[cloudera@quickstart ~]$ hdfs dfs -ls /mydata4/
Found 3 items
-rw-r--r-- 1 cloudera supergroup 86 2024-10-15 19:50 /mydata4/as
-rw-r--r-- 1 cloudera supergroup 43 2024-10-15 18:58 /mydata4/test1.txt
-rw-r--r-- 1 cloudera supergroup 24 2024-10-15 19:44 /mydata4/test3.txt
[cloudera@quickstart ~]$ hdfs dfs -rmr /mydata4/
rmr: DEPRECATED: Please use 'rm -r' instead.
Deleted /mydata4
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 10 items
drwxrwxrwx - hdfs supergroup 0 2017-04-05 04:27 /benchmarks
drwxr-xr-x - hbase supergroup 0 2024-09-27 05:50 /hbase
drwxr-xr-x - cloudera supergroup 0 2024-10-08 19:46 /mydata
drwxr-xr-x - cloudera supergroup 0 2024-10-15 19:46 /mydata2
drwxr-xr-x - cloudera supergroup 0 2024-10-15 19:23 /mydata3
drwxr-xr-x - cloudera supergroup 0 2024-10-15 19:18 /myfolder
drwxr-xr-x - solr solr 0 2017-04-05 04:29 /solr
drwxrwxrwt - hdfs supergroup 0 2017-04-05 04:29 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-04-05 04:29 /user
drwxr-xr-x - hdfs supergroup 0 2017-04-05 04:29 /var
[cloudera@quickstart ~] Computer
```

Task 17:

a) Get the size of each file in the directory.

i) `hdfs dfs -du /mydata2/`

```
[cloudera@quickstart ~]$ hdfs dfs -du /mydata2/
86 86 /mydata2/allsales
0 0 /mydata2/test2.txt
24 24 /mydata2/test3.txt
```

Task 18:

- a) Get the total size of directory/file

i) **hdfs dfs -dus /mydata2/**

```
[cloudera@quickstart ~]$ hdfs dfs -dus /mydata2/
dus: DEPRECATED: Please use 'du -s' instead.
110 110 /mydata2
```

Task 19:

- a) Get the last modified time of directory or path. In short it will give stats of the directory or file

i) **hdfs dfs -stat /mydata2/**

```
[cloudera@quickstart ~]$ hdfs dfs -stat /mydata2/
2024-10-16 02:46:34
```

Task 20:

- a) Change the replication factor of a file/directory in HDFS.

i) **hdfs dfs -setrep -R 4 /mydata2/**

```
[cloudera@quickstart ~]$ hdfs dfs -setrep -R 4 /mydata2/
Replication 4 set: /mydata2/allsales
Replication 4 set: /mydata2/test2.txt
Replication 4 set: /mydata2/test3.txt
[cloudera@quickstart ~]$
```