HADOOP COMMANDS

To check the version of Hadoop:

> hadoop version

```
[cloudera@quickstart ~]$ hadoop version
Hadoop 2.6.0-cdh5.13.0
Subversion http://github.com/cloudera/hadoop -r 42e8860b182e55321bd5f5605264da4adc8882be
Compiled by jenkins on 2017-10-04T18:08Z
Compiled with protoc 2.5.0
From source with checksum 5e84c185f8a22158e2b0e4b8f85311
This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.13.0.jar
```

Command for the help 😡

> hadoop fs -help

To store that commands into one file, below is the command:

```
[cloudera@quickstart ~]$ hadoop fs -help >> /home/cloudera/Desktop/BigData/Hadoop/savehelpcommand.txt
```

To check the files and directories in hdfs 👍

> 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 2024-06-09 01:15 /hbase
drwxr-xr-x   - solr  solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs  supergroup          0 2024-06-09 01:14 /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
```

To create the directory in Hadoop:

> hdfs dfs -mkdir komi-hadoop

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /komi-hadoop
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2024-06-09 01:15 /hbase
drwxr-xr-x   - cloudera supergroup          0 2024-06-09 03:41 /komi-hadoop
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs     supergroup          0 2024-06-09 01:14 /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 ~]$
```

To override if the directory exits:

> hdfs dfs -mkdir -p komi-hadoop

To get the HDFS admin report:

> hadoop dfsadmin -report

```
[cloudera@quickstart ~]$ hadoop dfsadmin -report
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Configured Capacity: 58531520512 (54.51 GB)
Present Capacity: 47003013120 (43.77 GB)
DFS Remaining: 46130372608 (42.96 GB)
DFS Used: 872640512 (832.21 MB)
DFS Used%: 1.86%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 2789

-------------------------------------------------
Live datanodes (1):

Name: 10.0.2.15:50010 (quickstart.cloudera)
Hostname: quickstart.cloudera
Decommission Status : Normal
Configured Capacity: 58531520512 (54.51 GB)
DFS Used: 872640512 (832.21 MB)
Non DFS Used: 8548454400 (7.96 GB)
DFS Remaining: 46130372608 (42.96 GB)
DFS Used%: 1.49%
DFS Remaining%: 78.81%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 2
Last contact: Sun Jun 09 04:04:23 PDT 2024
```

To copy the file from local FS to HDFS:

> hadoop dfs -put /sourcepath /hdfspath

```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/BigData/Hadoop/test.txt /komi-hadoop
[cloudera@quickstart ~]$ hdfs dfs -ls /komi-hadoop
Found 1 items
-rw-r--r--   1 cloudera supergroup         70 2024-06-09 04:24 /komi-hadoop/test.txt
```

> hadoop dfs -copyFromLocal /sourcepath /hdfspath

```
[cloudera@quickstart Hadoop]$ hdfs dfs -copyFromLocal /home/cloudera/Desktop/BigData/Hadoop /komi-hadoop
[cloudera@quickstart Hadoop]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2024-06-09 01:15 /hbase
drwxr-xr-x   - cloudera supergroup          0 2024-06-09 05:25 /komi-hadoop
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs     supergroup          0 2024-06-09 01:14 /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 Hadoop]$ hdfs dfs -ls /komi-hadoop
Found 2 items
drwxr-xr-x   - cloudera supergroup          0 2024-06-09 05:25 /komi-hadoop/Hadoop
-rw-r--r--   1 cloudera supergroup        215 2024-06-09 04:38 /komi-hadoop/test.txt
[cloudera@quickstart Hadoop]$
```

Behavior:

- **Retention of Local Files**:
  - `-put` can be thought of as potentially moving files (although in practice it copies).
  - `-copyFromLocal` explicitly copies files and leaves the original files in place.

Checksum 🙂:

> hdfs dfs -checksum /hadoop directory

The checksum is a value used to verify the integrity of data. By comparing checksums before and after a file transfer, you can ensure that the file has not been altered or corrupted.

```
[cloudera@quickstart ~]$ hadoop dfs -checksum /komi-hadoop/test.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

/komi-hadoop/test.txt   MD5-of-0MD5-of-512CRC32C        000002000000000000000000076196532a35ba9877a4086d4535aa8e8
[cloudera@quickstart ~]$
```

In this example:

- `/user/hadoop/example.txt` is the file path.

- `MD5-of-0MD5-of-512CRC32C` is the type of checksum algorithm used.
- `0000020000000000000000000e4b102ec` is the actual checksum value.

Once the content of the file changes, we can see the checksum value changes:

```
[cloudera@quickstart Hadoop]$ hadoop dfs -checksum /komi-hadoop/test.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

/komi-hadoop/test.txt   MD5-of-0MD5-of-512CRC32C        000002000000000000000007ea542b4bfc4ac05ec99d54c927ae8ff
```

To copy file or directory from HDFS to local:

> hdfs dfs -copyToLocal <hdfs path> <local FS path>

```
[cloudera@quickstart Hadoop]$ hdfs dfs -copyToLocal /komi-hadoop/Hadoop /home/cloudera/Desktop/
```

> hdfs dfs –get <hdfs path> <local FS path>

```
[cloudera@quickstart Hadoop]$ hdfs dfs -get /komi-hadoop/Hadoop /home/cloudera/Desktop/
```

To append the Hdfs File:

hdfs dfs -appendToFile /home/cloudera/Desktop/BigData/Hadoop/test.txt /komi-hadoop/test.txt

---

Check whether the files and directories in HDFS are healthy or corrupted:

> hdfs fsck /

```
[cloudera@quickstart Hadoop]$ hdfs fsck /
Connecting to namenode via http://quickstart.cloudera:50070/fsck?ugi=cloudera&path=%2F
FSCK started by cloudera (auth:SIMPLE) from /10.0.2.15 for path / at Sun Jun 09 05:00:33 PDT 2024
.....................................................................................................
.....................................................................................................
.....................................................................................................
.....................................................................................................
.....................................................................................................
.....................................................................................................
.....................................................................................................
.....................................................................................................
.....................................................................................................
................................Status: HEALTHY
 Total size:    861287149 B
 Total dirs:    79
 Total files:   935
 Total symlinks:                0
 Total blocks (validated):      933 (avg. block size 923137 B)
 Minimally replicated blocks:   933 (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 Sun Jun 09 05:00:33 PDT 2024 in 157 milliseconds


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

To Count the number of files:

> hdfs dfs -count -h /

```
[cloudera@quickstart Hadoop]$ hdfs dfs -count /
          80          939          861305269 /
[cloudera@quickstart Hadoop]$ hdfs dfs -count -h /
          80          939            821.4 M /
```

To copy the data within hdfs 😣:

> hdfs dfs -cp <hdfs-path1-source> <hdfs-path1-destination>

```
[cloudera@quickstart Hadoop]$ hdfs dfs -cp /komi-hadoop/Hadoop /komi-hadoop1
[cloudera@quickstart Hadoop]$ hdfs dfs -ls /komi-hadoop1
Found 2 items
-rw-r--r--   1 cloudera supergroup      17975 2024-06-09 08:20 /komi-hadoop1/savehelpcommand.txt
-rw-r--r--   1 cloudera supergroup        145 2024-06-09 08:20 /komi-hadoop1/test.txt
```

To find the disk free space in HDFS 👍

> hdfs dfs -df /

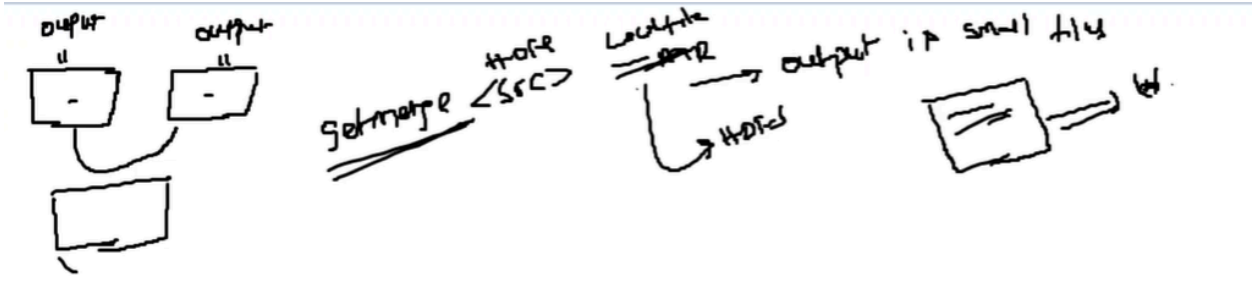
```
[cloudera@quickstart Hadoop]$ hadoop fs -df /
Filesystem                                Size       Used    Available  Use%
hdfs://quickstart.cloudera:8020  58531520512  872699754  46030139392    1%
[cloudera@quickstart Hadoop]$ hadoop fs -df -h /
Filesystem                           Size     Used  Available  Use%
hdfs://quickstart.cloudera:8020  54.5 G  832.3 M     42.9 G    1%
[cloudera@quickstart Hadoop]$
```

To find the disk usage in HDFS 🙂:

> hdfs dfs -du -h /

```
[cloudera@quickstart Hadoop]$ hadoop fs -du -h /
0        0        /benchmarks
4.0 K    4.0 K    /hbase
17.9 K   17.9 K   /komi-hadoop
17.7 K   17.7 K   /komi-hadoop1
0        0        /solr
0        0        /tmp
821.4 M  821.4 M  /user
0        0        /var
```

To Merge two files in HDFS ➕:



> hdfs dfs -getmerge <hadoop_path> <hadoop_path> <local file system_path>

```
[cloudera@quickstart Hadoop]$ hdfs dfs -getmerge /komi-hadoop1/test.txt /komi-hadoop1/savehelpcommand.txt /home/cloudera/Desk
top/BigData/Hadoop/getMerge.txt
[cloudera@quickstart Hadoop]$
```

If you want to change the replication factor size:

- Then go to a file called hdfs-site.xml and change the replication value
- Then restart the name node and data node

To set/change the replication factor for the loaded file:

> hdfs dfs -setrep -w <rep_value> <hdfs-path>

```
[cloudera@quickstart ~]$ hdfs dfs -setrep -w 3 /hadoop-user/savehelpcommands.txt
Replication 3 set: /hadoop-user/savehelpcommands.txt
```

Before the name node completes the replication, if it fails, you can see that 1 is written but the other one is failed.
So you can see the under replication is 1

```
[cloudera@quickstart ~]$ hadoop dfsadmin -report
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

Configured Capacity: 58531520512 (54.51 GB)
Present Capacity: 46693675129 (43.49 GB)
DFS Remaining: 45821007372 (42.67 GB)
DFS Used: 872667757 (832.24 MB)
DFS Used%: 1.87%
Under replicated blocks: 1
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0

-------------------------------------------------
Live datanodes (1):

Name: 10.0.2.15:50010 (quickstart.cloudera)
Hostname: quickstart.cloudera
Decommission Status : Normal
Configured Capacity: 58531520512 (54.51 GB)
DFS Used: 872667757 (832.24 MB)
Non DFS Used: 8589358483 (8.00 GB)
DFS Remaining: 45821007372 (42.67 GB)
DFS Used%: 1.49%
DFS Remaining%: 78.28%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 6
Last contact: Mon Jun 10 11:11:54 PDT 2024
```

If you dont care whether its successful or not, you just want to replicate it, then use -R :

> hdfs dfs -setrep -R <rep_value> <hdfs-path>

Expunge:

- To delete the trash permanently

To find the details about the files stored in block locations:

> hdfs fsck /hadoop-user/savehelpcommands.txt -files -locations -blocks

```
[cloudera@quickstart ~]$ hdfs fsck /hadoop-user/savehelpcommands.txt -files -locations -blocks
Connecting to namenode via http://quickstart.cloudera:50070/fsck?ugi=cloudera&files=1&locations=1&blocks=1&path=%2Fhadoop-use
r%2Fsavehelpcommands.txt
FSCK started by cloudera (auth:SIMPLE) from /10.0.2.15 for path /hadoop-user/savehelpcommands.txt at Mon Jun 10 12:07:48 PDT
2024
/hadoop-user/savehelpcommands.txt 17975 bytes, 1 block(s):  Under replicated BP-1067413441-127.0.0.1-1508775264580:blk_107374
2759_1935. Target Replicas is 3 but found 1 live replica(s), 0 decommissioned replica(s), 0 decommissioning replica(s).
0. BP-1067413441-127.0.0.1-1508775264580:blk_1073742759_1935 len=17975 Live_repl=1 [DatanodeInfoWithStorage[10.0.2.15:50010,D
S-621c9e78-caa3-4a7b-bf10-3c8a1245cb51,DISK]]

Status: HEALTHY
 Total size:    17975 B
 Total dirs:    0
 Total files:   1
 Total symlinks:                0
 Total blocks (validated):      1 (avg. block size 17975 B)
 Minimally replicated blocks:   1 (100.0 %)
 Over-replicated blocks:        0 (0.0 %)
 Under-replicated blocks:       1 (100.0 %)
 Mis-replicated blocks:         0 (0.0 %)
 Default replication factor:    1
 Average block replication:     1.0
 Corrupt blocks:                0
 Missing replicas:              2 (66.666664 %)
 Number of data-nodes:          1
 Number of racks:               1
FSCK ended at Mon Jun 10 12:07:48 PDT 2024 in 2 milliseconds


The filesystem under path '/hadoop-user/savehelpcommands.txt' is HEALTHY
```
