

Basic Hadoop Commands (PART 2)

➤ Command used for conditional checking.

- We can use multiple parameters with test command they are d,s,f,e,r,w,z

Command: `hadoop fs -test -d directoryname`

- To check the result for the above command we have to use `echo $?`

Output:

```
cloudera@quickstart:~$  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop fs -test -d dir1  
[cloudera@quickstart ~]$ echo $?  
0  
[cloudera@quickstart ~]$
```

- -d → Check whether the path given by the user is a directory or not, return 0 if it is a directory.
- -f → Check whether the path given by the user is a file or not, return 0 if it is a file.
- -e → Check whether the path given by the user exists or not, return 0 if the path exists.
- -z → Checks whether the file size is 0 byte or not, return 0 if the file is of 0 bytes.
- -w → Return 0 if the path exists and write permission is granted.
- -r → Return 0 if the path exists and read permission is granted.
- -s → Check if the path is not empty, return 0 if a path is not empty.
- And the result will be in the form of 0 and 1.
 - 0 → we will get output as 0 if the directory is present.
 - 1 → we will get output as 1 if the directory is not present.

• To Move file from hdfs to local system and vice versa.

`hadoop fs -moveFromLocal filelocation (present in local system) location (of hadoop file system)`

`hadoop fs -moveToLocal filelocation (present in hadoop system) location (of local file system)`

Output:

```
[cloudera@quickstart ~]$ hadoop fs -moveFromLocal Desktop/TSRTC.pdf dir1  
[cloudera@quickstart ~]$ hadoop fs -ls dir1  
Found 3 items  
-rw-r--r-- 1 cloudera cloudera 8492119 2022-09-28 02:15 dir1/TSRTC.pdf  
-rwxr-xr-x 1 cloudera cloudera 14784 2022-09-07 01:29 dir1/country_wise_latest.csv  
-rwxr-xr-x 1 cloudera cloudera 56 2022-09-07 01:58 dir1/inptfile.csv  
[cloudera@quickstart ~]$
```

- **To Merge the files of Hadoop file system to Local file system.**

```
hadoop fs -getmerge (filename) file location (file name present in hadoop system) ~/localfile system
location/ merged file (as per our choice)
```

You can add 'n' number of files of local system to get merged with the hadoop file.

Output:

```
[cloudera@quickstart ~]$ hadoop fs -getmerge dir1/q2.txt dir1/q3.txt ~/Desktop/q1.txt
[cloudera@quickstart ~]$ cat Desktop/q1.txt
soumya soumya
abcdefghijkl
abcdefghijkl
lmnopqrstuvwxyz
```

- **To append files of local system to one of the file present in hadoop.**

```
hadoop fs -appendToFile ( files location of local file system ) (file of hdfs location)
```

Output:

```
[cloudera@quickstart ~]$ hadoop fs -appendToFile q1.txt q4.txt q2.txt
[cloudera@quickstart ~]$ hadoop fs -cat dir1/q2.txt
soumya soumya
soumya soumya
abcdefghijkl
abcdefghijkl
lmnopqrstuvwxyz
hello
hello
hello
hello
abcdefghijkl
hello
hello
hello
hello
hello
```

- **To check the integrity of a file.**

- Such as whether the file has been modified or not.
- Hashing concept is used in this command. (MD5 Algo is used—Message Digest)
- Output of this command will be in alpha-numeric.

```
hadoop fs -checksum file location
```

Output:

```
[cloudera@quickstart ~]$ hadoop fs -checksum dir1/q2.txt
dir1/q2.txt      MD5-of-0MD5-of-512CRC32C      000002000000000000000000fc67897e1b03f9d08bee061f26ddfc38
```

- **To check the health status of Hadoop System of a particular directory and entire root directory.**
- **fsck**→ file system check.

```
hdfs fsck - /
```

Output:

[illegible]

```
hdfs fsck - /dir1
```

```
[cloudera@quickstart ~]$ hdfs fsck - /dir1
Connecting to namenode via http://quickstart.cloudera:50070/fsck?ugi=cloudera&path=%2Fdir1
FSCK started by cloudera (auth:SIMPLE) from /10.0.2.15 for path /dir1 at Wed Sep 28 03:22:15 PDT 2022
.Status: HEALTHY
Total size:      42 B
Total dirs:      0
Total files:     1
Total symlinks:   0
Total blocks (validated): 1 (avg. block size 42 B)
Minimally replicated blocks: 1 (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 Wed Sep 28 03:22:15 PDT 2022 in 1 milliseconds

The filesystem under path '/dir1' is HEALTHY
```

- **To check count of files and directories present in the given location.**

```
hadoop fs -count directoryname
```

Output:

```
[cloudera@quickstart ~]$ hadoop fs -count dir1
1      5      8507124 dir1
[cloudera@quickstart ~]$ hadoop fs -count /
377    1158    911014237 /
```

- **From the above output of `hadoop fs -count dir1` explanation:**

1 → number of directories.

5 → number of files.

8507124 → size occupied by those files.

- **To delete a Directory.**

```
hadoop fs -rmr directoryname
```

Output:

```
[cloudera@quickstart ~]$ hadoop fs -rmr mydirectory1
rmr: DEPRECATED: Please use 'rm -r' instead.
Deleted mydirectory1
```

- **To delete a File**

```
hadoop fs -rm file location
```

Output:

```
[cloudera@quickstart ~]$ hadoop fs -rm foldr/file1.txt
Deleted foldr/file1.txt
```

- **To know Statistics of a file**

Output:

```
[cloudera@quickstart ~]$ hadoop fs -stat %r /f2.txt
1
[cloudera@quickstart ~]$ hadoop fs -stat %b /f2.txt
42
[cloudera@quickstart ~]$ hadoop fs -stat %g /f2.txt
cloudera
[cloudera@quickstart ~]$ hadoop fs -stat %u /f2.txt
cloudera
[cloudera@quickstart ~]$ hadoop fs -stat %y /f2.txt
2022-09-02 06:13:29
```

hadoop fs -stat %b file location

we can use multiple parameters:

%r→to know replication factor of a file

%b→to know number of bytes of file

%g→to know the group name

%u→to know username of a file

%y→to know when the file is modified

- To change group of a file

Output:

```
[cloudera@quickstart ~]$ hadoop fs -ls /
Found 16 items
drwxr-xr-x  - cloudera supergroup          0 2022-09-22 02:43 /-ext-10000
-rwxr-xr-x  1 cloudera supergroup    55351 2022-09-20 08:39 /AgentLoggingReport.csv
drwxrwxrwx  - hdfs supergroup           0 2017-10-23 09:15 /benchmarks
-rwxr-xr-x  1 cloudera supergroup     135 2022-09-06 00:55 /csv1.csv
-rwxr-xr-x  1 cloudera supergroup      42 2022-09-01 23:14 /dir1
-rwxr-xr-x  1 cloudera supergroup      42 2022-09-01 23:13 /f2.txt
-rwxr-xr-x  1 cloudera supergroup       5 2022-09-01 09:13 /folder1
-rwxr-xr-x  1 cloudera supergroup     119 2022-09-01 01:19 /foldr
drwxr-xr-x  - hbase supergroup           0 2022-09-28 01:13 /hbase
drwxr-xr-x  - cloudera supergroup          0 2022-09-22 04:07 /home
drwxr-xr-x  - cloudera cloudera          0 2022-09-07 02:53 /inptfile.csv
drwxr-xr-x  - cloudera cloudera          0 2022-09-27 10:20 /mydirectory1
drwxr-xr-x  - solr solr                  0 2017-10-23 09:18 /solr
drwxrwxrwt  - hdfs supergroup           0 2022-09-01 01:26 /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 fs -chgrp cloudera /f2.txt
[cloudera@quickstart ~]$ hadoop fs -ls /
Found 16 items
drwxr-xr-x  - cloudera supergroup          0 2022-09-22 02:43 /-ext-10000
-rwxr-xr-x  1 cloudera supergroup    55351 2022-09-20 08:39 /AgentLoggingReport.csv
drwxrwxrwx  - hdfs supergroup           0 2017-10-23 09:15 /benchmarks
-rwxr-xr-x  1 cloudera supergroup     135 2022-09-06 00:55 /csv1.csv
-rwxr-xr-x  1 cloudera supergroup      42 2022-09-01 23:14 /dir1
-rwxr-xr-x  1 cloudera cloudera      42 2022-09-01 23:13 /f2.txt
-rwxr-xr-x  1 cloudera supergroup       5 2022-09-01 09:13 /folder1
-rwxr-xr-x  1 cloudera supergroup     119 2022-09-01 01:19 /foldr
drwxr-xr-x  - hbase supergroup           0 2022-09-28 01:13 /hbase
drwxr-xr-x  - cloudera supergroup          0 2022-09-22 04:07 /home
drwxr-xr-x  - cloudera cloudera          0 2022-09-07 02:53 /inptfile.csv
drwxr-xr-x  - cloudera cloudera          0 2022-09-27 10:20 /mydirectory1
drwxr-xr-x  - solr solr                  0 2017-10-23 09:18 /solr
drwxrwxrwt  - hdfs supergroup           0 2022-09-01 01:26 /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 change the permissions of a file or directory.

hadoop fs -chmod 777(any rule you want to give) filelocation or directory

You can give 777,754,755...etc., any rule as per your convenience.

Output:

```
[cloudera@quickstart ~]$ hadoop fs -chmod 777 dir1/inptfile.csv
[cloudera@quickstart ~]$ hadoop fs -ls dir1
Found 5 items
-rw-r--r-- 1 cloudera cloudera 8492119 2022-09-28 02:15 dir1/TSRTC.pdf
-rwxr-xr-x 1 cloudera cloudera 14784 2022-09-07 01:29 dir1/country_wise_latest.csv
-rwxrwxrwx 1 cloudera cloudera 56 2022-09-07 01:58 dir1/inptfile.csv
-rw-r--r-- 1 cloudera cloudera 127 2022-09-28 02:48 dir1/q2.txt
-rw-r--r-- 1 cloudera cloudera 38 2022-09-28 02:38 dir1/q3.txt
[cloudera@quickstart ~]$ hadoop fs -chmod 754 dir1/inptfile.csv
[cloudera@quickstart ~]$ hadoop fs -ls dir1
Found 5 items
-rw-r--r-- 1 cloudera cloudera 8492119 2022-09-28 02:15 dir1/TSRTC.pdf
-rwxr-xr-x 1 cloudera cloudera 14784 2022-09-07 01:29 dir1/country_wise_latest.csv
-rwxr-xr-- 1 cloudera cloudera 56 2022-09-07 01:58 dir1/inptfile.csv
-rw-r--r-- 1 cloudera cloudera 127 2022-09-28 02:48 dir1/q2.txt
-rw-r--r-- 1 cloudera cloudera 38 2022-09-28 02:38 dir1/q3.txt
```