Yandex

Big Data

HDFS Client


```
man         hdfs dfs -help
info        hdfs dfs -usage <utility_name>
            hdfs namenode
            hdfs datanode
            ...
```

-h : 파일 사이즈를 byte가 아닌 human readable하게

-R : recursive. 폴더 안에 있는 모든 폴더에 적용

```
$ hdfs dfs -ls -R -h /data/wiki
drwxr-xr-x - hdfs supergroup      0       2017-01-07 20:59 /data/wiki/en_articles
-rw-r--r-- 3 hdfs supergroup      11.5 G     2017-01-07 20:59 /data/wiki/en_articles/
Articles
```

-du : 파일 사이즈와 replica를 포함한 사이즈 용량 표시

-df : storage 사용량 정보

```
$ hdfs dfs -du -h /data/wiki

11.5 G 34.4 G /data/wiki/en_articles

73.3 M 219.9 M /data/wiki/en_articles_part

$ hdfs dfs -df -h /data/wiki
```

Filesystem	Size	Used	Available	Use
hdfs://virtual-master.atp-fivt.org:8020	1.6 T	389.0 G	1.1 T	23 %

```
$ hdfs dfs -mkdir deep/nested/path

$ hdfs dfs -mkdir -p deep/nested/path
```

-touchz : 파일 생성 (z의 의미는 zero lengt라는 뜻)

```
$ hdfs dfs –touchz file.txt
```

-skiptrash : 완전 삭제 (그냥 삭제는 trash로 이동한다.)

```
$ hdfs dfs –rm –r --skiptrash deep/nested/path
```

-put : 로컬에서 하둡으로 파일 이동


```
$ hdfs dfs –put ~/test.txt /data/test.txt
```

-get : 하둡에서 로컬로 파일 다운로드

```
$ hdfs dfs –get /data/test.txt <폴더위치>
```

- getmerge : 같은 이름으로 시작하는 파일들을 하나로 합쳐서 다운로드

```
$ hdfs dfs –getmerge hdfs_test* hdfs_merged.txt
```

```
$ hdfs dfs -cat hdfs_test_file.txt | head -4
test content #1
test content #2
test content #3
test content #4
```

hdfs에서 -cat은 전체 파일만 볼 수 있음

hdfs에서 head는 사용 불가하다.

대신, cat + 파이프라인으로 대체

```
$ hdfs dfs -cat hdfs_test_file.txt | tail -4
test content #12
test content #13
test content #14
test content #15
```

hdfs에서 tail은 1KB의 데이터만 보여준다.

```
$ hdfs dfs -tail hdfs_test_file.txt
test content #1
test content #2
...
test content #15
```

```
$ time hdfs dfs -setrep -w 1 hdfs_test_file.txt
Replication 1 set: hdfs_test_file.txt
Waiting for hdfs_test_file.txt ...
WARNING: the waiting time may be long for DECREASING the number of
replications...done
real 0m13.148s
user 0m4.232s
sys 0m0.156s
```

```
$ hdfs dfs -ls
Found 3 items
drwx------     -  adral adral 0       2017-04-05 13:00 .Trash
-rw-r--r--         1   adral adral 246   2017-04-05 13:03 hdfs_test_file.txt
-rw-r--r--         3   adral adral 246   2017-04-05 13:28 hdfs_test_file_copy.txt
```

```
$ time hdfs dfs -setrep -w 2 hdfs_test_file.txt
Replication 2 set: hdfs_test_file.txt
Waiting for hdfs_test_file.txt .... done
real 0m13.168s
user 0m4.092s
sys 0m0.148s
```

```
$ hdfs fsck /data/wiki/en_articles –files
Connecting to namenode via http://virtual-master.atp-fivt.org:50070
FSCK started by adral (auth:SIMPLE) from /138.201.91.190 for path /data/wiki/en_articles
at Wed Apr 05 14:01:15 CEST 2017
/data/wiki/en_articles <dir>
/data/wiki/en_articles/articles 12328051927 bytes, 92 block(s): OKsys 0m0.148s
```

```
$ hdfs fsck /data/wiki/en_articles -files –blocks
Connecting to namenode via http://virtual-master.atp-fivt.org:50070
FSCK started by adral (auth:SIMPLE) from /138.201.91.190 for path /data/wiki/en_articles
at Wed Apr 05 14:01:48 CEST 2017
/data/wiki/en_articles <dir>
/data/wiki/en_articles/articles 12328051927 bytes, 92 block(s): OK
0. BP-858288134-138.201.91.191-1481279621434:blk_1073808471_67650 len=134217728
Live_repl=3
1. BP-858288134-138.201.91.191-1481279621434:blk_1073808472_67651 len=134217728
Live_repl=3
2. BP-858288134-138.201.91.191-1481279621434:blk_1073808473_67652 len=134217728
Live_repl=3
```

```
$ hdfs fsck /data/wiki/en_articles -files -blocks –locations
Connecting to namenode via http://virtual-master.atp-fivt.org:50070
FSCK started by adral (auth:SIMPLE) from /138.201.91.190 for path /data/wiki/en_articles
at Wed Apr 05 14:01:56 CEST 2017
/data/wiki/en_articles <dir>
```

Summary

- you can request meta-information from Namenode and change its structure (ls, mkdir, rm, rm -r (-skipTrash), touch, mv)
- you can read and write data from and to Datanodes in HDFS (put, cat, head, tail, get, getmerge)
- you can change replication factor of files and get detailed information about data in HDFS (chown, hdfs groups; setrep; hdfs fsck; find)

BigDATAteam