



하둡 명령어



목차

1

블록 관련 명령어

2

디렉토리 관련 명령어

3

데이터 입/출력 관련 명령어

4

검색 관련 명령어

5

권한 관련 명령어

6

기타 명령어



PART 1. 블록 관련 명령어

하둡에서 사용하는 명령어 중 블록과 관련된 명령어에 대해 학습



PART1. 블록 관련 명령어

```
hadoop@hadoop-name: ~/bigdata (1)
hadoop@hadoop-name:~/bigdata$ hadoop fs -D dfs.block.size=30 -put stocks.csv /
19/07/23 21:10:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
put: Specified block size is less than configured minimum value (dfs.namenode.fs-limits.min-block-size): 30 < 1048576 (2)
hadoop@hadoop-name:~/bigdata$
```

- 실습에 필요한 파일(stocks.csv)은 수업시간에 제공
- (1) - 로컬에서 stock.csv를 블록사이즈 30으로 두고 하둡공간으로 옮기기 실행
hadoop fs -D dfs.block.size=30 -put stocks.csv /
- (2) - 에러메시지 발생:
30 바이트의 블록사이즈는 유효하지 않습니다.
블록사이즈는 최소 1048576 이상이 되어야 한다.

HORTONWORKS HDP 3.0.1 설정

```
[root@sandbox-hdp files]# hadoop fs -D dfs.block.size=30 -put stocks.csv /  
put: Permission denied: user=root, access=WRITE, inode="/":hdfs:hdfs:drwxr-xr-x
```

```
[root@sandbox-hdp files]# hadoop fs -getfacl /  
# file: /  
# owner: hdfs  
# group: hdfs  
user::rwx  
group::r-x  
other::r-x
```

- **Hortonworks** HDP 3.0.1를 이용하여 첫 번째 예제 (hadoop fs -D dfs.block.size=30 -put stocks.csv /) 실행 시 Permission denied 오류가 나는 경우
 - 루트("/") 디렉토리에 파일을 쓰기(Write)할 수 없어서 생기는 문제
 - 오류 메시지의 inode를 확인하면 루트 디렉토리의 소유자 및 그룹은 hdfs로 설정이 되어 있음
- hadoop fs -getfacl / 명령어로 루트의 권한을 확인하면, 소유자와 그룹은 hdfs로 설정이 되어 있음. 즉, hdfs 그룹에 root 사용자를 포함하고, 권한을 부여하여야 함

HORTONWORKS HDP 3.0.1 설정

```
[root@sandbox-hdp files]# usermod -aG hdfs root
[root@sandbox-hdp files]# hdfs dfsadmin -refreshUserToGroupsMappings
Refresh user to groups mapping successful
```

```
[root@sandbox-hdp files]# cat /etc/group | grep hdfs
hdfs:x:1000:hdfs,root
```

```
[root@sandbox-hdp files]# hadoop fs -setfacl -m user:root:rwX /
[root@sandbox-hdp files]# hadoop fs -getfacl /
# file: /
# owner: hdfs
# group: hdfs
user::rwX
user:root:rwX
group::r-x
mask::rwX
other::r-x
```

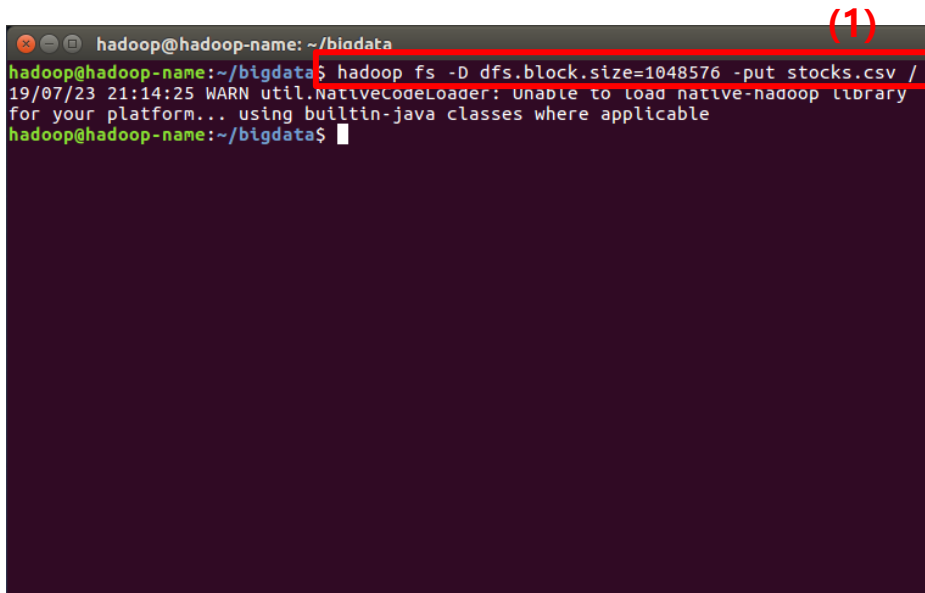
- hdfs 그룹에 root 사용자를 추가하고, 권한 부여
 - hdfs 그룹에 root 사용자 추가
 - usermod -aG hdfs root
 - -a 옵션 : append를 의미하여, 유저를 추가
 - -G 옵션 : Group 명을 지정
 - hdfs dfsadmin -refreshUserToGroupsMappings
 - 그룹 및 사용자 매핑 설정 새로고침 명령어
 - cat /etc/group | grep hdfs
 - hdfs 그룹에 속해 있는 사용자 확인
 - 권한 부여 및 확인
 - hadoop fs -setfacl -m user:root:rwX /
 - hadoop fs -getfacl /

PART1. 블록 관련 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -D dfs.block.size=2000000 -put stocks.csv /
19/07/23 21:11:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
-put: Invalid values: dfs.bytes-per-checksum (=512) must divide block size (=2000000).
usage: hadoop fs [generic options] -put [-f] [-p] [-l] <localsrc> ... <dst>
hadoop@hadoop-name:~/bigdata$
```

- (1) - 로컬에서 stock.csv를 블록사이즈 2000000으로 두고 하둡공간으로 옮기기 실행
hadoop fs -D dfs.block.size=2000000 -put stocks.csv /
- (2) - 에러메시지 발생:
2000000 바이트의 블록사이즈는 유효하지 않습니다. 블록사이즈는 checksum 의 제곱이어야 합니다. checksum = 512

PART1. 블록 관련 명령어

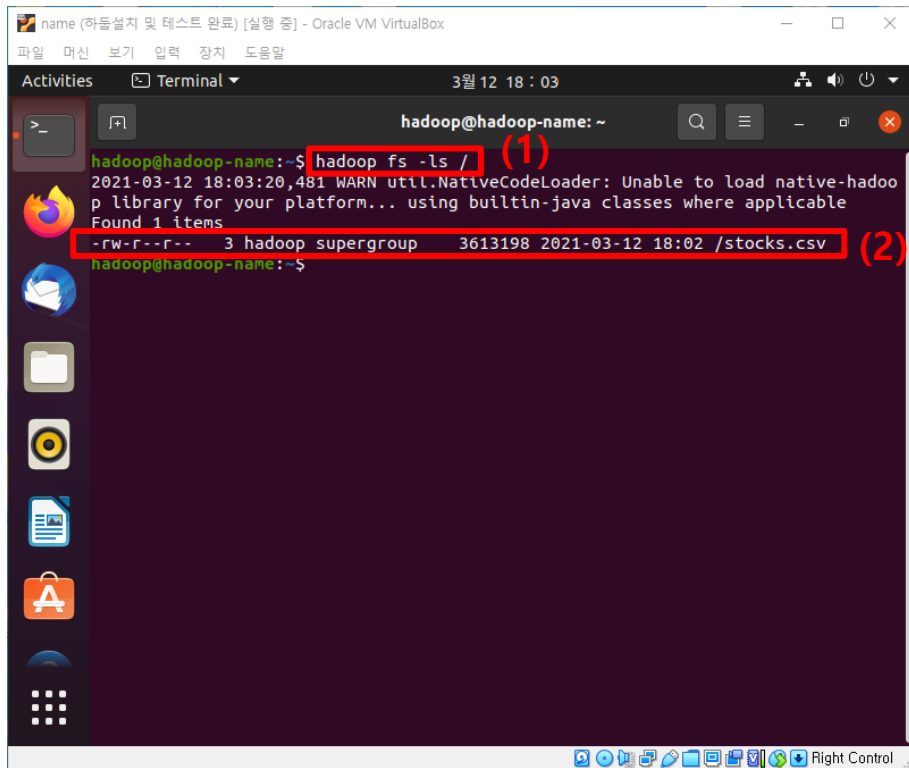


A terminal window titled 'hadoop@hadoop-name: ~/bigdata' with a red box highlighting the command 'hadoop fs -D dfs.block.size=1048576 -put stocks.csv /'. A red circle with the number '(1)' is next to the command. The output shows a warning message: '19/07/23 21:14:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable'.

```
hadoop@hadoop-name: ~/bigdata$ hadoop fs -D dfs.block.size=1048576 -put stocks.csv /
19/07/23 21:14:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name: ~/bigdata$
```

- (1) - 로컬에서 stock.csv를 블록사이즈 1048576으로 두고 하둡공간으로 옮기기 실행
hadoop fs -D dfs.block.size=1048576 -put stocks.csv /
- 에러메시지가 발생하지 않았다.

PART1. 블록 관련 명령어



The screenshot shows a terminal window titled "name (하둡설치 및 테스트 완료) [실행 중] - Oracle VM VirtualBox". The terminal prompt is "hadoop@hadoop-name: ~". The first command entered is "hadoop fs -ls /", which is highlighted with a red box and labeled with a red "(1)". The output shows a warning message and "Found 1 items". The second line of output is "-rw-r--r-- 3 hadoop supergroup 3613198 2021-03-12 18:02 /stocks.csv", which is highlighted with a red box and labeled with a red "(2)". The terminal prompt is then "hadoop@hadoop-name:~\$".

```
hadoop@hadoop-name:~$ hadoop fs -ls / (1)
2021-03-12 18:03:20,481 WARN util.NativeCodeLoader: Unable to load native-hadoop
p library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 3 hadoop supergroup 3613198 2021-03-12 18:02 /stocks.csv (2)
hadoop@hadoop-name:~$
```

- (1) - 하둡공간을 확인
 - `hadoop fs -ls /`
- (2) - stock.csv가 들어간 것을 확인할 수 있다.

PART1. 블록 관련 명령어

```
hadoop@hadoop-name: ~/bigdata
hadoop@hadoop-name:~/bigdata$ hadoop fsck /stocks.csv (1)
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

19/07/23 21:15:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Connecting to namenode via http://master:50070/fsock?ugi=hadoop&path=%2Fstocks.csv
FSCK started by hadoop (auth:SIMPLE) from /192.168.56.101 for path /stocks.csv at T
ue Jul 23 21:15:44 KST 2019
FSCK Status: HEALTHY
Total size: 3613198 B
Total dirs: 0
Total files: 1
Total symlinks: 0
Total blocks (validated): 4 (avg. block size 903299 B)
Minimally replicated blocks: 4 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 3.0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
Number of data-nodes: 3
Number of racks: 1
FSCK ended at Tue Jul 23 21:15:44 KST 2019 in 9 milliseconds

The filesystem under path '/stocks.csv' is HEALTHY
```

- (1) - 블록 개수를 알아보기 위한 명령어 실행
hadoop fsck /stocks.csv
- 총 블록 수가 4개인 것을 확인할 수 있다.
Total blocks (validated): 4 (avg. block size 903299 B)

PART1. 블록 관련 명령어

```
hadoop@hadoop-name: ~/bigdata
hadoop@hadoop-name:~/bigdata$ hadoop fsck /stocks.csv -files -blocks (1)
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

19/07/23 21:17:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Connecting to namenode via http://master:50070/fsck?ugi=hadoop&files=1&blocks=1&path=%2Fstocks.csv
FSCK started by hadoop (auth:SIMPLE) from /192.168.56.101 for path /stocks.csv at T
ue Jul 23 21:17:23 KST 2019
/stocks.csv 3613198 bytes, 4 block(s): OK
0. BP-321230114-127.0.1.1-1562140979934:blk_1073741831_1009 len=1048576 repl=3
1. BP-321230114-127.0.1.1-1562140979934:blk_1073741832_1010 len=1048576 repl=3
2. BP-321230114-127.0.1.1-1562140979934:blk_1073741833_1011 len=1048576 repl=3
3. BP-321230114-127.0.1.1-1562140979934:blk_1073741834_1012 len=467470 repl=3

Status: HEALTHY
Total size: 3613198 B
Total dirs: 0
Total files: 1
Total symlinks: 0
Total blocks (validated): 4 (avg. block size 903299 B)
Minimally replicated blocks: 4 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 3.0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
Number of data-nodes: 3
Number of racks: 1
FSCK ended at Tue Jul 23 21:17:23 KST 2019 in 1 milliseconds

The filesystem under path '/stocks.csv' is HEALTHY
hadoop@hadoop-name:~/bigdata$
```

- (1) - 블록 개수들과 블록 각각의 ID를 출력하는 명령어 실행
hadoop fsck /stocks.csv -files -blocks
- 블록 아이디를 확인할 수 있다.

PART1. 블록 관련 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fsck /stocks.csv
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

19/07/23 22:38:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Connecting to namenode via http://master:50070/fsck?ugi=hadoop&path=%2Fstocks.csv
FSCK started by hadoop (auth:SIMPLE) from /192.168.56.101 for path /stocks.csv at T
ue Jul 23 22:38:28 KST 2019
. Status: HEALTHY
Total size:      3613198 B
Total dirs:      0
Total files:     1
Total symlinks:   0
Total blocks (validated): 4 (avg. block size 903299 B)
Minimally replicated blocks: 4 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 3.0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
Number of data-nodes: 3
Number of racks: 1
FSCK ended at Tue Jul 23 22:38:28 KST 2019 in 2 milliseconds

The filesystem under path '/stocks.csv' is HEALTHY
```

- /stocks.csv 파일을 fsck 해보면 현재 파일의 복제본은 3개 인것을 확인할 수 있음

PART1. 블록 관련 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -setrep 2 /stocks.csv
19/07/23 22:47:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Replication 2 set: /stocks.csv
hadoop@hadoop-name:~/bigdata$ hadoop fsck /stocks.csv
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

19/07/23 22:47:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Connecting to namenode via http://master:50070/fsck?ugi=hadoop&path=%2Fstocks.csv
FSCK started by hadoop (auth:SIMPLE) from /192.168.56.101 for path /stocks.csv at T
ue Jul 23 22:47:32 KST 2019
Status: HEALTHY
Total size:      3613198 B
Total dirs:      0
Total files:      1
Total symlinks:    0
Total blocks (validated): 4 (avg. block size 903299 B)
Minimally replicated blocks: 4 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 2.0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
Number of data-nodes: 3
Number of racks: 1
FSCK ended at Tue Jul 23 22:47:32 KST 2019 in 1 milliseconds

The filesystem under path '/stocks.csv' is HEALTHY
```

- setrep 명령어를 사용하면 파일의 복제본 개수를 조절할 수 있음
- 복제본의 개수를 2개로 설정 후 fsck로 확인

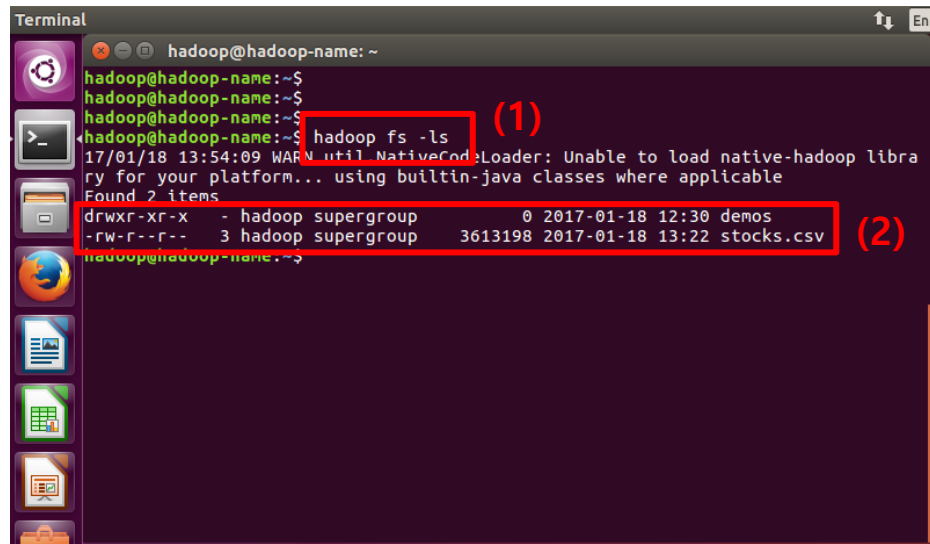


PART 2. 디렉토리 관련 명령어

하둡에서 사용하는 명령어 중 디렉토리와 관련된 명령어에 대해 학습



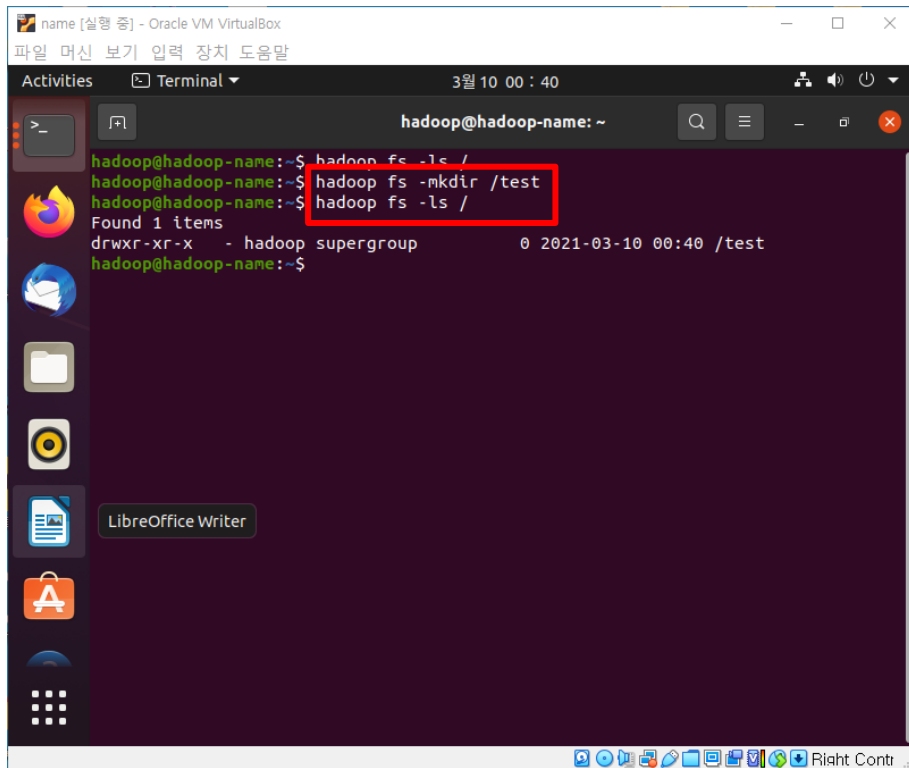
PART2. 디렉토리 관련 명령어



```
Terminal
hadoop@hadoop-name: ~
hadoop@hadoop-name:~$
hadoop@hadoop-name:~$
hadoop@hadoop-name:~$ hadoop fs -ls (1)
17/01/18 13:54:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x - hadoop supergroup 0 2017-01-18 12:30 demos
-rw-r--r-- 3 hadoop supergroup 3613198 2017-01-18 13:22 stocks.csv (2)
hadoop@hadoop-name:~$
```

- (1) - 하둡 공간을 확인하는 명령어
hadoop fs -ls /
- (2) - 결과

PART2. 디렉토리 관련 명령어



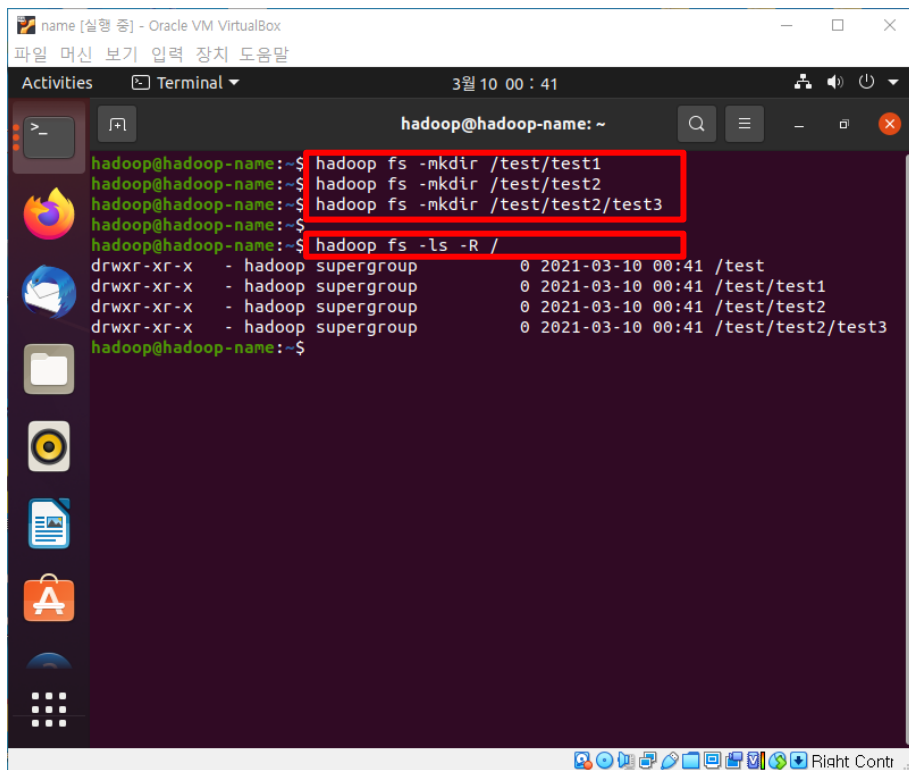
The screenshot shows a terminal window titled "name [실행 중] - Oracle VM VirtualBox". The terminal prompt is "hadoop@hadoop-name: ~". The following commands and output are shown:

```
hadoop@hadoop-name:~$ hadoop fs -ls /
hadoop@hadoop-name:~$ hadoop fs -mkdir /test
hadoop@hadoop-name:~$ hadoop fs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup          0 2021-03-10 00:40 /test
hadoop@hadoop-name:~$
```

The commands "hadoop fs -mkdir /test" and "hadoop fs -ls /" are highlighted with a red box.

- 하둡 공간에 루트 아래 test라는 디렉토리 만들고, 확인
 - `hadoop fs -mkdir /test`
 - `hadoop fs -ls /`

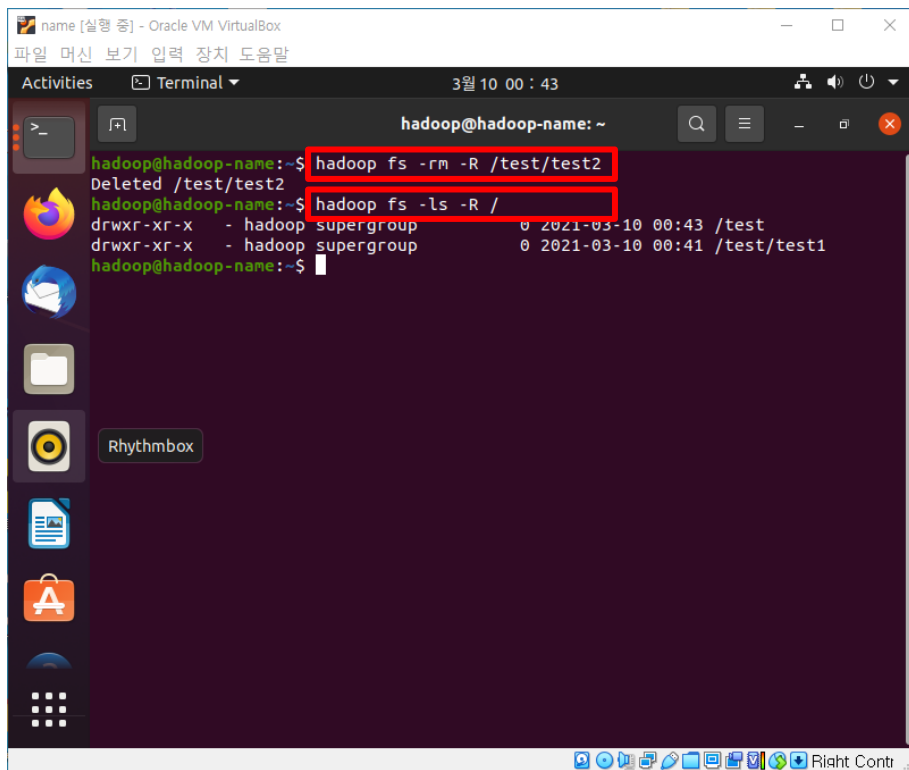
PART2. 디렉토리 관련 명령어



```
hadoop@hadoop-name: ~  
hadoop@hadoop-name:~$ hadoop fs -mkdir /test/test1  
hadoop@hadoop-name:~$ hadoop fs -mkdir /test/test2  
hadoop@hadoop-name:~$ hadoop fs -mkdir /test/test2/test3  
hadoop@hadoop-name:~$ hadoop fs -ls -R /  
drwxr-xr-x - hadoop supergroup 0 2021-03-10 00:41 /test  
drwxr-xr-x - hadoop supergroup 0 2021-03-10 00:41 /test/test1  
drwxr-xr-x - hadoop supergroup 0 2021-03-10 00:41 /test/test2  
drwxr-xr-x - hadoop supergroup 0 2021-03-10 00:41 /test/test2/test3  
hadoop@hadoop-name:~$
```

- 하둡 공간 내 test 디렉토리 내에 추가로 디렉토리를 생성
 - `hadoop fs -mkdir /test/test1`
`hadoop fs -mkdir /test/test2`
`hadoop fs -mkdir /test/test2/test3`
- ※ -R 을 붙여주면 디렉토리 내의 디렉토리 까지 출력
력이 된다.
 - 하둡 공간 확인
 - `hadoop fs -ls -R /`
 - Hortonworks의 경우 다른 디렉토리가 많이 있기 때문
에 아래 명령어로 수행
 - `hadoop fs -ls -R / | grep test`

PART2. 디렉토리 관련 명령어



```
hadoop@hadoop-name:~$ hadoop fs -rm -R /test/test2
Deleted /test/test2
hadoop@hadoop-name:~$ hadoop fs -ls -R /
drwxr-xr-x - hadoop supergroup 0 2021-03-10 00:43 /test
drwxr-xr-x - hadoop supergroup 0 2021-03-10 00:41 /test/test1
hadoop@hadoop-name:~$
```

- test2 디렉토리와 디렉토리 내부를 삭제
 - `hadoop fs -rm -R /test/test2`
- 삭제 후 확인
 - `hadoop fs -ls -R /`
 - Hortonworks의 경우 다른 디렉토리가 많이 있기 때문에 아래 명령어로 수행
 - `hadoop fs -ls -R / | grep test`

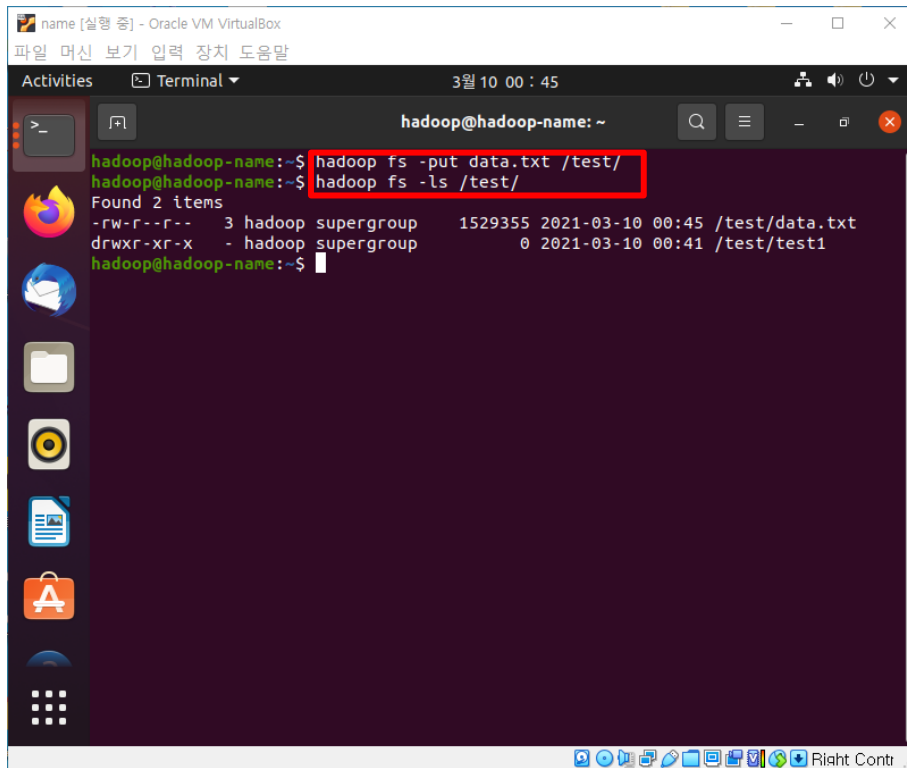


PART 3. 데이터 입/출력 관련 명령어

하둡에서 사용하는 명령어 중 데이터 입/출력과 관련된 명령어에 대해 학습



PART3. 데이터 입/출력 명령어

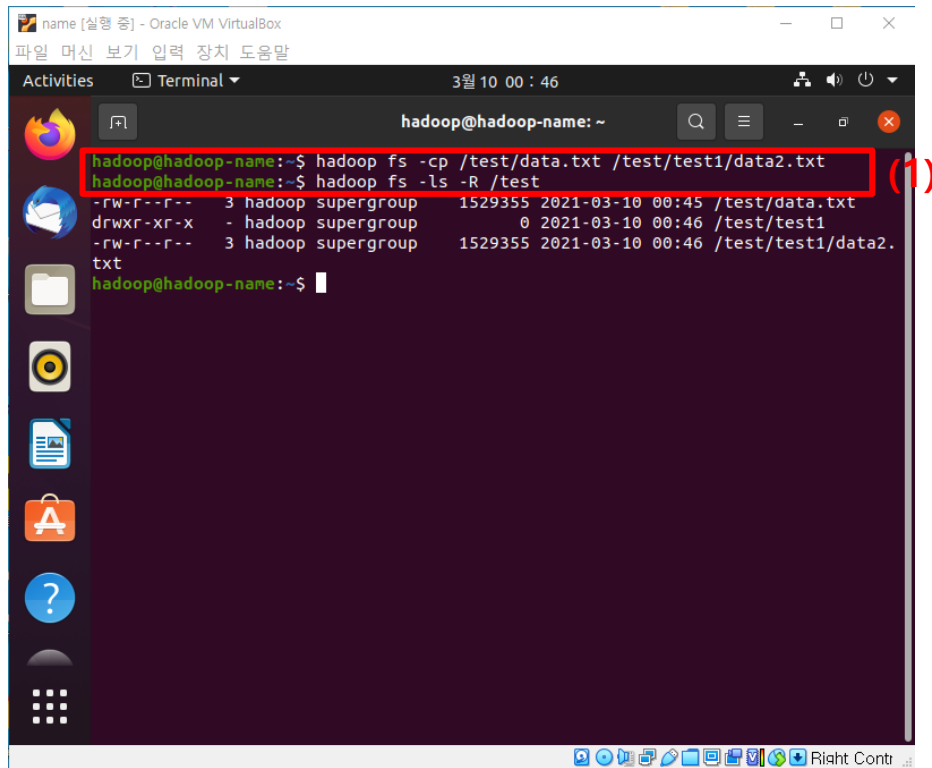


The screenshot shows a terminal window titled 'name [실행 중] - Oracle VM VirtualBox'. The terminal prompt is 'hadoop@hadoop-name: ~'. The user enters the command 'hadoop fs -put data.txt /test/' and then 'hadoop fs -ls /test/'. The output shows two items in the /test directory: 'data.txt' and 'test1'. The command 'hadoop fs -put data.txt /test/' is highlighted with a red box.

```
hadoop@hadoop-name:~$ hadoop fs -put data.txt /test/
hadoop@hadoop-name:~$ hadoop fs -ls /test/
Found 2 items
-rw-r--r-- 3 hadoop supergroup 1529355 2021-03-10 00:45 /test/data.txt
drwxr-xr-x - hadoop supergroup 0 2021-03-10 00:41 /test/test1
hadoop@hadoop-name:~$
```

- 로컬에 있는 data.txt파일을 하둡 공간의 루트 아래 test 디렉토리에 복사하는 명령
 - `hadoop fs -put data.txt /test/`
- 복사 후 확인
 - `hadoop fs -ls /test/`

PART3. 데이터 입/출력 명령어

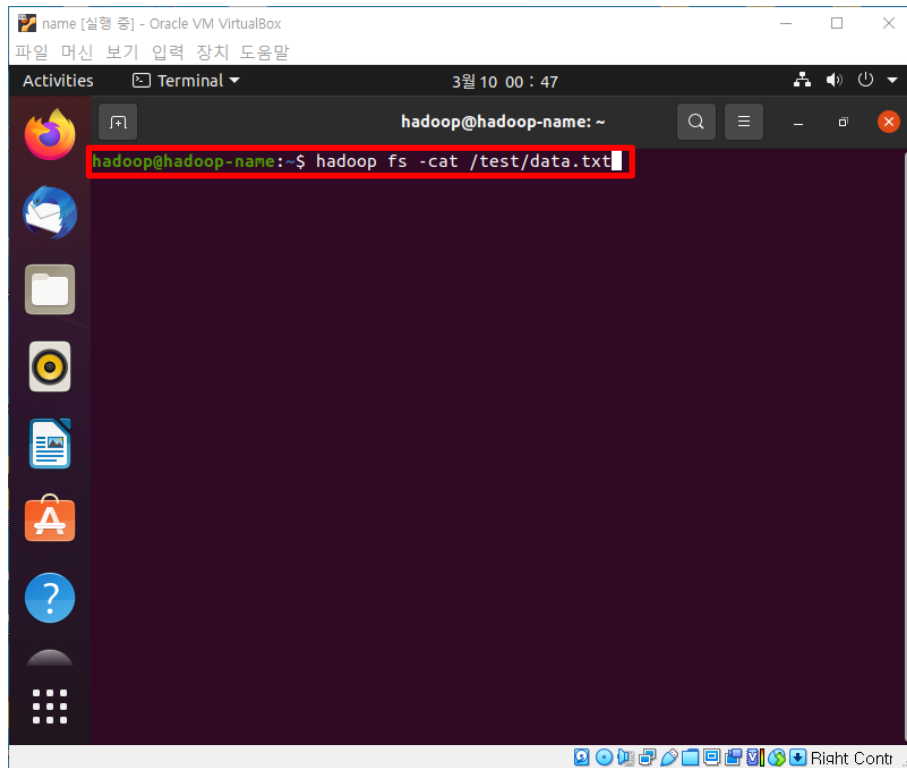


The screenshot shows a terminal window titled "name [실행 중] - Oracle VM VirtualBox". The terminal prompt is "hadoop@hadoop-name: ~". The first command entered is "hadoop fs -cp /test/data.txt /test/test1/data2.txt", which is highlighted with a red box. The second command is "hadoop fs -ls -R /test", which lists the contents of the /test directory. The output shows three entries: /test/data.txt, /test/test1, and /test/test1/data2.txt. A red circle with the number (1) is next to the first command.

```
hadoop@hadoop-name:~$ hadoop fs -cp /test/data.txt /test/test1/data2.txt
hadoop@hadoop-name:~$ hadoop fs -ls -R /test
-rw-r--r-- 3 hadoop supergroup 1529355 2021-03-10 00:45 /test/data.txt
drwxr-xr-x - hadoop supergroup 0 2021-03-10 00:46 /test/test1
-rw-r--r-- 3 hadoop supergroup 1529355 2021-03-10 00:46 /test/test1/data2.txt
hadoop@hadoop-name:~$
```

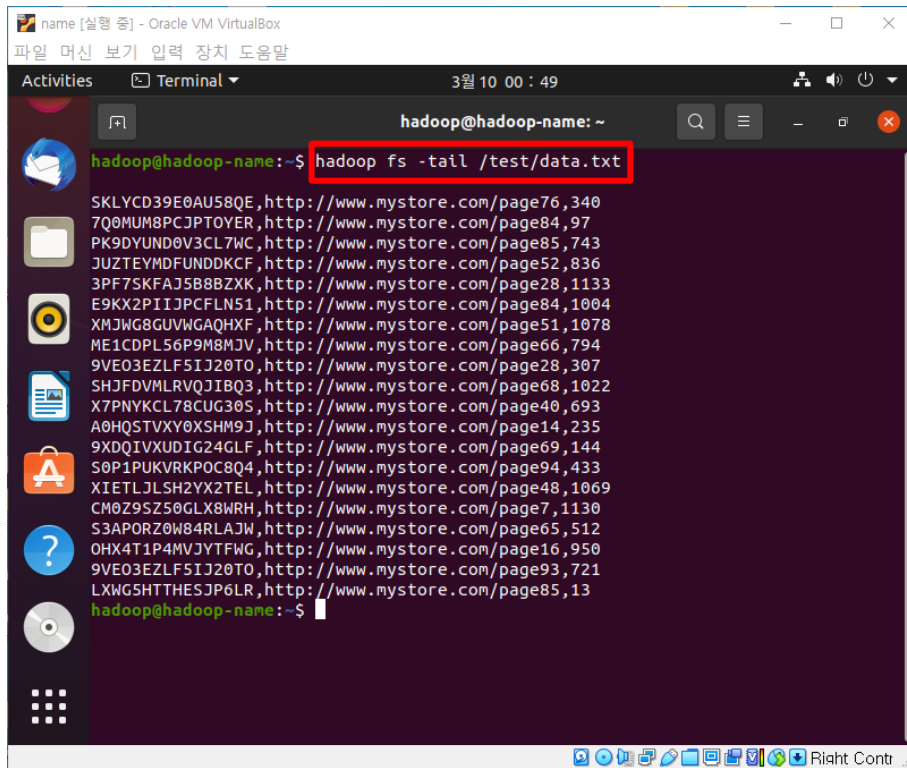
- (1) - 전에 옮겼던 test디렉토리 안의 data.txt파일을 같은 하둡공간의 test1디렉토리 내부로 옮기고, 이름은 data2.txt로 지정
 - `hadoop fs -cp /test/data.txt /test/test1/data2.txt`

PART3. 데이터 입/출력 명령어



- data.txt의 내용을 cat 명령어로 출력
 - `hadoop fs -cat /test/data.txt`

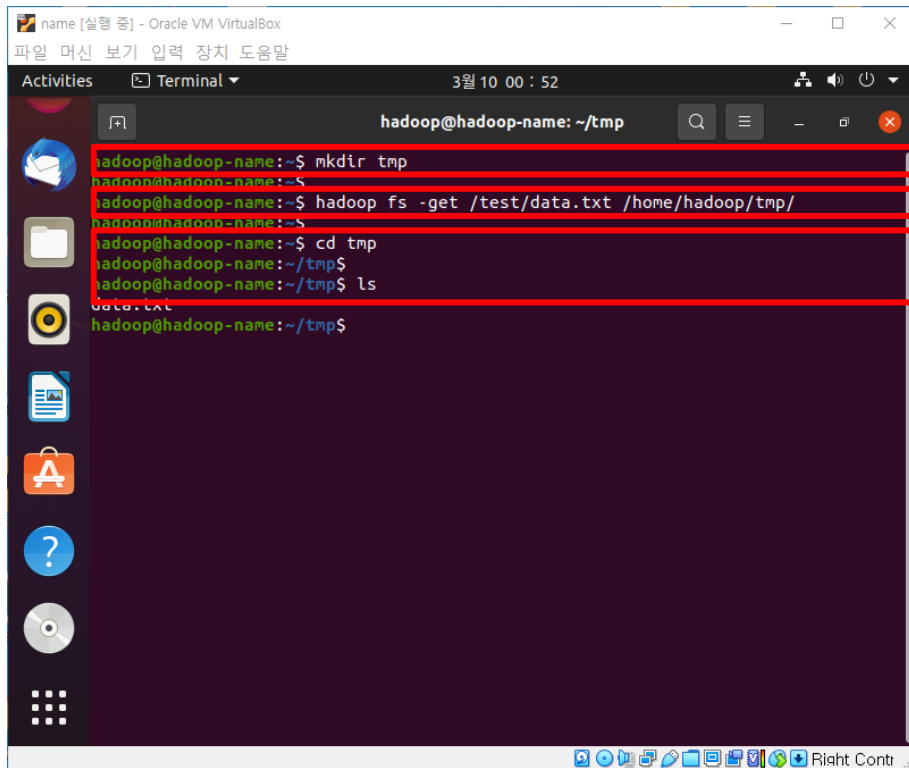
PART3. 데이터 입/출력 명령어



```
hadoop@hadoop-name: ~  
hadoop@hadoop-name:~$ hadoop fs -tail /test/data.txt  
SKLYCD39E0AU58QE,http://www.mystore.com/page76,340  
7Q0MUM8PCJPTOYER,http://www.mystore.com/page84,97  
PK9DYUND0V3CL7WC,http://www.mystore.com/page85,743  
JUZYTEYMDFUNDDKCF,http://www.mystore.com/page52,836  
3PF7SKFAJ5B8BZXK,http://www.mystore.com/page28,1133  
E9KX2PIIJCFLN51,http://www.mystore.com/page84,1004  
XMJWG8GUVWGAQHXF,http://www.mystore.com/page51,1078  
ME1CDPL56P9M8MJV,http://www.mystore.com/page66,794  
9VE03EZLF5IJ20TO,http://www.mystore.com/page28,307  
SHJFDVMLRVQJIBQ3,http://www.mystore.com/page68,1022  
X7PNYKCL78CUG30S,http://www.mystore.com/page40,693  
A0HQSTVXY0XSHM9J,http://www.mystore.com/page14,235  
9XDQIVXUDIG24GLF,http://www.mystore.com/page69,144  
S0P1PUKVRKPOC8Q4,http://www.mystore.com/page94,433  
XIETLJLSH2YX2TEL,http://www.mystore.com/page48,1069  
CM0Z9SZ50GLX8WRH,http://www.mystore.com/page7,1130  
S3APORZ0W84RLAJW,http://www.mystore.com/page65,512  
OHX4T1P4MVJYTFWG,http://www.mystore.com/page16,950  
9VE03EZLF5IJ20TO,http://www.mystore.com/page93,721  
LXWG5HTTHESJP6LR,http://www.mystore.com/page85,13  
hadoop@hadoop-name:~$
```

- data.txt의 내용을 tail 명령어로 출력
 - `hadoop fs -tail /test/data.txt`

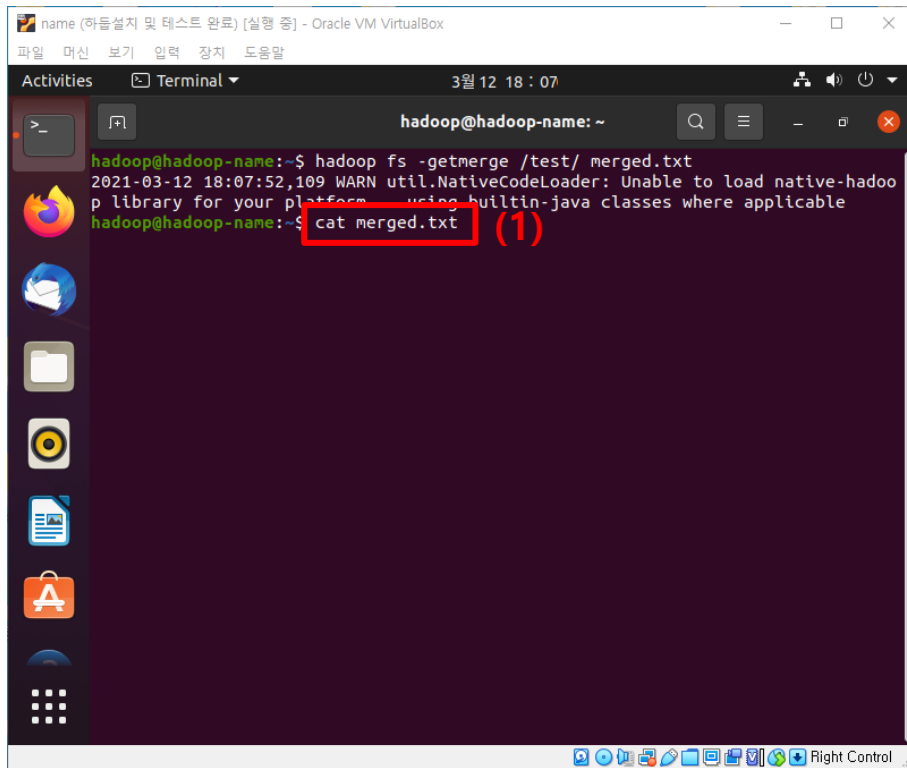
PART3. 데이터 입/출력 명령어



```
hadoop@hadoop-name: ~/tmp
hadoop@hadoop-name:~$ mkdir tmp
hadoop@hadoop-name:~$
hadoop@hadoop-name:~$ hadoop fs -get /test/data.txt /home/hadoop/tmp/
hadoop@hadoop-name:~$
hadoop@hadoop-name:~$ cd tmp
hadoop@hadoop-name:~/tmp$
hadoop@hadoop-name:~/tmp$ ls
data.txt
hadoop@hadoop-name:~/tmp$
```

- 로컬에 tmp 디렉토리를 생성
 - `mkdir tmp`
- hadoop 공간에 있는 data.txt 파일을 로컬의 tmp 디렉토리로 복사
 - `hadoop fs -get test/data.txt home/hadoop/tmp/`
- tmp 디렉토리 진입
 - `cd tmp`
- tmp 디렉토리 내부 확인
 - `ls`

PART3. 데이터 입/출력 명령어



```
hadoop@hadoop-name: ~  
hadoop@hadoop-name:~$ hadoop fs -getmerge /test/ merged.txt  
2021-03-12 18:07:52,109 WARN util.NativeCodeLoader: Unable to load native-hadoop  
p library for your platform... using builtin-java classes where applicable  
hadoop@hadoop-name:~$ cat merged.txt (1)
```

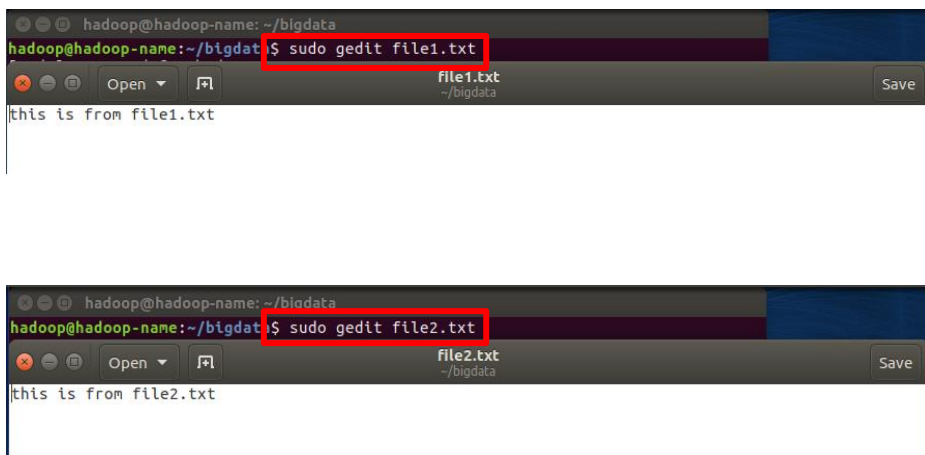
- (1) – getmerge를 통해 HDFS /test 디렉토리 내
부의 파일을 하나로 합침
 - `hadoop fs -germerge /test/ merged.txt`
- (2) - 생성된 merged.txt를 cat 명령어를 이용하
여 출력
 - `cat merged.txt`

PART3. 데이터 입/출력 명령어

```
hadoop@hadoop-name: ~/bigdata
hadoop@hadoop-name:~/bigdata$ hadoop fs -touchz /test1.txt (1)
19/07/23 21:36:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
19/07/23 21:36:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 3 hadoop supergroup 3613198 2019-07-23 21:14 /stocks.csv
-rw-r--r-- 3 hadoop supergroup 0 2019-07-23 21:36 /test1.txt
hadoop@hadoop-name:~/bigdata$
```

- 리눅스 기본 명령어인 touch와 유사하게 0바이트의 크기를 가진 빈 파일을 생성
 - `hadoop fs -touchz /test1.txt`

PART3. 데이터 입/출력 명령어



- 리눅스 로컬에서 file1.txt 파일을 생성 후 좌측 그림과 같이 내용 입력
 - `sudo gedit file1.txt`
 - Hortonworks의 경우 vi 명령어로 수행
 - `sudo vi file1.txt`
 - 키보드 "i"를 입력하면, 문자를 타이핑할 수 있음. 타이핑 후 키보드 "ESC" 키를 입력하고, `:wq`를 입력하면 자동 저장 및 종료 됨
- 리눅스 로컬에서 file2.txt 파일을 생성 후 좌측 그림과 같이 내용 입력
 - `sudo gedit file2.txt`
 - Hortonworks의 경우 vi 명령어로 수행
 - `sudo vi file2.txt`
 - 키보드 "i"를 입력하면, 문자를 타이핑할 수 있음. 타이핑 후 키보드 "ESC" 키를 입력하고, `:wq`를 입력하면 자동 저장 및 종료 됨

PART3. 데이터 입/출력 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -appendToFile file1.txt /test.txt
19/07/03 17:13:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
19/07/03 17:13:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 3 hadoop supergroup 24 2019-07-03 17:13 /test.txt
hadoop@hadoop-name:~/bigdata$ hadoop fs -cat /test.txt
19/07/03 17:14:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
this is from file1.txt
```

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -appendToFile file2.txt /test.txt
19/07/03 17:21:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -cat /test.txt
19/07/03 17:21:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
this is from file1.txt
this is from file2.txt
```

- appendToFile 명령어를 사용하면 put과 동일하게 HDFS에 파일 저장이 가능
- file1.txt 파일을 HDFS에 test.txt로 저장
- appendToFile 명령어로 file2.txt를 test.txt에 저장하면 기존에 있는 test.txt 마지막에 file2.txt의 내용이 저장됨

PART3. 데이터 입/출력 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -appendToFile - /test.txt
19/07/03 19:24:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Hello This is the text typed in stdin, also command is appendToFile.
^Chadoop@hadoop-name:~/bigdata$ hadoop fs -cat /test.txt
19/07/03 19:25:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
this is from file1.txt
this is from file2.txt
Hello This is the text typed in stdin, also command is appendToFile.
```

- appendToFile 명령어로 입력할 파일을 지정하지 않고 "-" (하이픈)을 사용하면 사용자가 수동으로 입력 가능
- 입력후 컨트롤 + c 단축키로 입력을 종료
- 입력후 cat명령어로 확인하면 사용자가 입력한 내용을 확인할 수 있음

PART3. 데이터 입/출력 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -checksum /stocks.csv
19/07/23 21:56:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
/stocks.csv      MD5-of-2048MD5-of-512CRC32C  00000200000000000000000000000000ccddb0352c9
36af1caea25f14ba48b4e
hadoop@hadoop-name:~/bigdata$
```

- checksum 명령어는 해당 파일의 체크섬 관련 정보를 확인할 수 있음

PART3. 데이터 입/출력 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -touchz /test2.txt
19/07/23 22:17:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -touchz /test3.txt
19/07/23 22:17:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -touchz /test4.txt
19/07/23 22:17:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
19/07/23 22:17:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 5 items
-rw-r--r--  3 hadoop supergroup    3613198 2019-07-23 21:14 /stocks.csv
-rw-r--r--  3 hadoop supergroup      0 2019-07-23 21:56 /test1.txt
-rw-r--r--  3 hadoop supergroup      0 2019-07-23 22:17 /test2.txt
-rw-r--r--  3 hadoop supergroup      0 2019-07-23 22:17 /test3.txt
-rw-r--r--  3 hadoop supergroup      0 2019-07-23 22:17 /test4.txt
hadoop@hadoop-name:~/bigdata$
```

- touchz 명령어로 크기가 없는 임시 파일 생성
- test2.txt, test3.txt, test4.txt 생성
- 현재 총 파일이 5개
 - Hortonworks의 경우 아래 명령어로 조회

PART3. 데이터 입/출력 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -count /
19/07/23 22:19:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
      1      5      3613198 /
hadoop@hadoop-name:~/bigdata$ hadoop fs -mkdir /Dir1
19/07/23 22:21:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -mkdir /Dir2
19/07/23 22:21:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -count /
19/07/23 22:21:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
      3      5      3613198 /
```

- count 명령어를 이용하여 현재 디렉토리의 디렉토리 개수, 파일 개수, 전체 파일의 용량을 볼 수 있음
- 결과가 1 5 3613198이 출력되었는데
 - 1 : 디렉토리 개수(현재 디렉토리를 포함)
 - 5 : 파일 개수
 - 3613198 : 디렉토리 내의 파일 전체 용량
- Dir1, Dir2 디렉토리를 추가로 생성하고 count 명령어를 사용하면 3 5 3613198 결과를 볼 수 있음



PART 4. 검색 관련 명령어

하둡에서 사용하는 명령어 중 검색과 관련된 명령어에 대해 학습



PART4. 검색 관련 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -find / -name test* -print  
19/07/23 22:33:31 WARN util.NativeCodeLoader: Unable to load native-nadoop library  
for your platform... using builtin-java classes where applicable  
/test1.txt  
/test2.txt  
/test3.txt  
/test4.txt  
hadoop@hadoop-name:~/bigdata$
```

- find 명령어를 사용하여 HDFS 내의 파일을 찾을 수 있음
- 옵션으로 -name test*를 주고, test로 시작하는 파일을 찾으면 결과가 출력됨
 - Hortonworks의 경우 /user/root/.Trash ~로 시작하는 경로가 출력될 수도 있음.
 - Hortonworks의 HDFS에서 파일 또는 디렉토리를 삭제하면 로그 파일을 남기기 때문에 기록됨

PART4. 검색 관련 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -getfacl /
19/07/23 22:35:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
# file: /
# owner: hadoop
# group: supergroup
getfacl: The ACL operation has been rejected. Support for ACLs has been disabled b
y setting dfs.namenode.acls.enabled to false.
hadoop@hadoop-name:~/bigdata$
```

- getfacl 명령어는 해당 디렉토리의 경로와 디렉토리 소유자, 그룹명을 알려줌.



PART 5. 권한 관련 명령어

하둡에서 사용하는 명령어 중 권한과 관련된 명령어에 대해 학습



PART5. 권한 관련 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -chmod 777 /stocks.csv
19/07/23 23:22:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
19/07/23 23:22:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - hadoop supergroup      0 2019-07-23 22:21 /Dir1
drwxr-xr-x - hadoop supergroup      0 2019-07-23 22:21 /Dir2
-rwxrwxrwx 3 hadoop supergroup    12345 2019-07-23 22:56 /stocks.csv
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 21:56 /test1.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test2.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test3.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test4.txt
hadoop@hadoop-name:~/bigdata$
```

- chmod 명령어는 파일의 접근 권한을 변경할 수 있음
- 기존에 -rw-r--r-- (644)로 구성된 stocks.csv의 권한이 -rwxrwxrwx (777)로 변경된 것을 확인

PART5. 권한 관련 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -chown testowner /stocks.csv
19/07/23 23:24:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
19/07/23 23:24:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - hadoop supergroup      0 2019-07-23 22:21 /Dir1
drwxr-xr-x - hadoop supergroup      0 2019-07-23 22:21 /Dir2
-rwxrwxrwx 3 testowner supergroup 12345 2019-07-23 22:56 /stocks.csv
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 21:56 /test1.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test2.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test3.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test4.txt
hadoop@hadoop-name:~/bigdata$
```

- chown 명령어는 파일의 소유자명을 변경할 수 있음
- 기존에 hadoop으로 지정된 파일의 소유자명이 testowner로 바뀐 것을 확인할 수 있음

PART5. 권한 관련 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -chgrp testgroup /stocks.csv
19/07/23 23:25:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
19/07/23 23:25:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - hadoop supergroup 0 2019-07-23 22:21 /Dir1
drwxr-xr-x - hadoop supergroup 0 2019-07-23 22:21 /Dir2
-rwxrwxrwx 3 testowner testgroup 12345 2019-07-23 22:56 /stocks.csv
-rw-r--r-- 3 hadoop supergroup 0 2019-07-23 21:56 /test1.txt
-rw-r--r-- 3 hadoop supergroup 0 2019-07-23 22:17 /test2.txt
-rw-r--r-- 3 hadoop supergroup 0 2019-07-23 22:17 /test3.txt
-rw-r--r-- 3 hadoop supergroup 0 2019-07-23 22:17 /test4.txt
hadoop@hadoop-name:~/bigdata$
```

- chgrp 명령어는 파일의 그룹명을 변경할 수 있음
- 기존에 supergroup 으로 그룹명이 지정되어 있었지만, 명령어를 통해 testgroup으로 바뀐 것을 확인할 수 있음



PART 6. 기타 명령어

앞장에서 설명하지 않은 하둡 관련 명령어에 대해 학습



PART6. 기타 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop version  
Hadoop 2.7.5  
Subversion https://shv@git-wip-us.apache.org/repos/asf/hadoop.git -r 18065c2b6806ed  
4aa6a3187d77cbe21bb3dba075  
Compiled by kshvachk on 2017-12-16T01:06Z  
Compiled with protoc 2.5.0  
From source with checksum 9f118f95f47043332d51891e37f736e9  
This command was run using /home/hadoop/bigdata/hadoop/share/hadoop/common/hadoop-c  
ommon-2.7.5.jar
```

- `hadoop version` 명령어를 사용하면 현재 사용중인 하둡의 버전을 확인할 수 있음

PART6. 기타 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop dfsadmin -report
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

19/07/23 22:50:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Configured Capacity: 56669491200 (52.78 GB)
Present Capacity: 25831026574 (24.06 GB)
DFS Remaining: 25823584256 (24.05 GB)
DFS Used: 7442318 (7.10 MB)
DFS Used%: 0.03%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0

-----
Live datanodes (3):

Name: 192.168.56.102:50010 (slave1)
Hostname: hadoop-data1
Decommission Status : Normal
Configured Capacity: 18889830400 (17.59 GB)
DFS Used: 53134 (51.89 KB)
Non DFS Used: 13642162290 (12.71 GB)
DFS Remaining: 4264472576 (3.97 GB)
DFS Used%: 0.00%
DFS Remaining%: 22.58%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
```

- `hadoop dfsadmin -report` 명령어를 사용하면 현재 사용하는 노드들의 대한 정보를 한눈에 볼 수 있음.

PART6. 기타 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
19/07/23 22:56:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - hadoop supergroup      0 2019-07-23 22:21 /Dir1
drwxr-xr-x - hadoop supergroup      0 2019-07-23 22:21 /Dir2
-rw-r--r-- 3 hadoop supergroup 3613198 2019-07-23 22:54 /stocks.csv
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 21:56 /test1.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test2.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test3.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test4.txt
hadoop@hadoop-name:~/bigdata$ hadoop fs -truncate 12345 /stocks.csv
19/07/23 22:56:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Truncating /stocks.csv to length: 12345. Wait for block recovery to complete before
further updating this file.
hadoop@hadoop-name:~/bigdata$ hadoop fs -ls /
19/07/23 22:56:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - hadoop supergroup      0 2019-07-23 22:21 /Dir1
drwxr-xr-x - hadoop supergroup      0 2019-07-23 22:21 /Dir2
-rw-r--r-- 3 hadoop supergroup 12345 2019-07-23 22:56 /stocks.csv
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 21:56 /test1.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test2.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test3.txt
-rw-r--r-- 3 hadoop supergroup      0 2019-07-23 22:17 /test4.txt
```

- truncate 명령어는 지정한 숫자 만큼 파일의 길이를 맞춰줌
- 12345 바이트의 크기만큼 파일이 수정된걸 확인할 수 있음

PART6. 기타 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -stat "type:%F|%u:%g|size:%b|mtime:%y|name:%n" /stocks.csv
19/07/23 23:02:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
type:regular file|hadoop:supergroup|size:12345|mtime:2019-07-23 13:56:15|name:stocks.csv
hadoop@hadoop-name:~/bigdata$
```

- stat 명령어는 해당 파일의 통계를 보여줌
- %F : 파일의 타입
- %u : 파일의 소유자
- %g : 파일의 그룹명
- %b : 파일의 크기
- %y : UTC 날짜 (yyyy—MM-dd HH:mm:ss)
- %n : 파일명

PART6. 기타 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -du /
19/07/23 23:09:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
0      /Dir1
0      /Dir2
12345  /stocks.csv
0      /test1.txt
0      /test2.txt
0      /test3.txt
0      /test4.txt
hadoop@hadoop-name:~/bigdata$ hadoop fs -du -h /
19/07/23 23:09:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
0      /Dir1
0      /Dir2
12.1 K /stocks.csv
0      /test1.txt
0      /test2.txt
0      /test3.txt
0      /test4.txt
hadoop@hadoop-name:~/bigdata$
```

```
hadoop@hadoop-name:~$ hadoop fs -du -h /
0 0 /Dir1
0 0 /Dir2
8 16 /test1.txt
29 87 /test2.txt
5 15 /test3.txt
0 0 /test4.txt
```

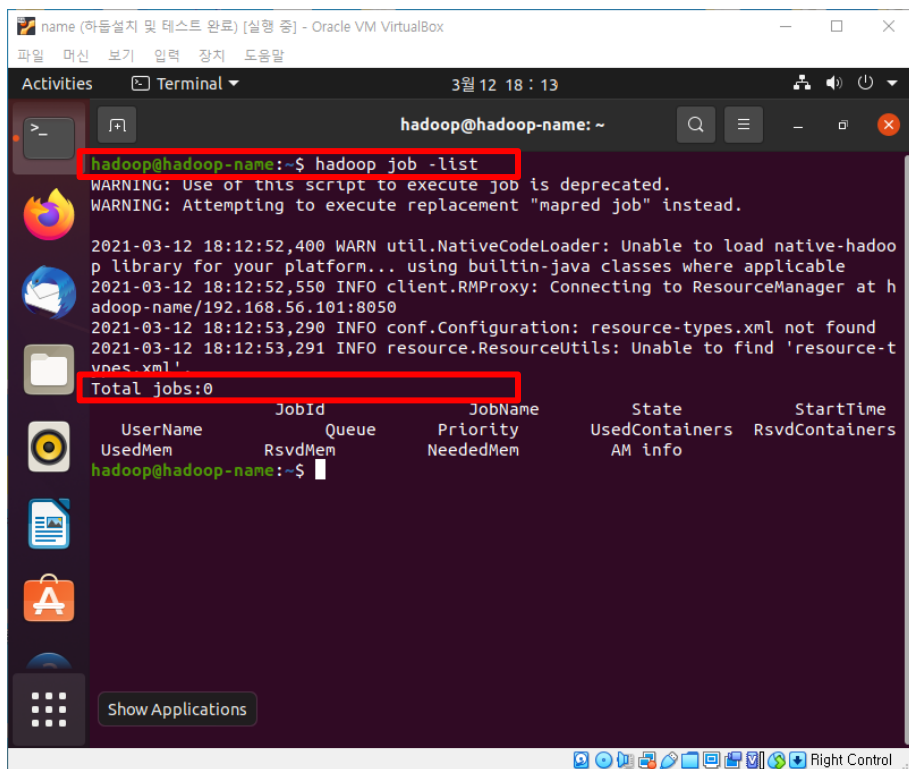
- **du** 명령어는 현재 디렉토리내 파일의 용량을 확인할 수 있음
- **-h** 옵션을 주면 보기 편하게 단위를 줄여줌
 - 그냥 숫자로 표현되면 바이트 크기를 의미하고, 그 이상은 KB, MB로 단위가 붙여 표현됨
- 좌측 아래 그림은 하둡 3.0.0 버전부터 바뀐 출력 결과임
 - 첫 번째 컬럼의 숫자는 파일의 크기를 보여주며, 두 번째 컬럼의 숫자는 HDFS에 저장되어 있는 파일의 복제본 전체 크기를 알려줌
 - EX) 29 / 87 / test2.txt는 파일의 크기가 29바이트이지만, 복제를 통해 3개의 블록으로 저장되었기 때문에 $29 \times 3 = 87$ 로 표현됨

PART6. 기타 명령어

```
hadoop@hadoop-name:~/bigdata$ hadoop fs -df /
19/07/23 23:11:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Filesystem                Size      Used Available Use%
hdfs://master:9000    56669491200  172032  25830801408    0%
hadoop@hadoop-name:~/bigdata$ hadoop fs -df -h /
19/07/23 23:11:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
Filesystem                Size      Used Available Use%
hdfs://master:9000    52.8 G  168 K    24.1 G    0%
hadoop@hadoop-name:~/bigdata$
```

- df 명령어는 파일시스템의 경로, 파일시스템 크기, 사용중인 크기, 사용가능 공간, 사용률을 보여줌
- -h 옵션을 주면 보기 편하게 단위를 줄여줌

PART6. 기타 명령어



```
hadoop@hadoop-name: ~  
hadoop@hadoop-name:~$ hadoop job -list  
WARNING: Use of this script to execute job is deprecated.  
WARNING: Attempting to execute replacement "mapred job" instead.  
  
2021-03-12 18:12:52,400 WARN util.NativeCodeLoader: Unable to load native-hadoop  
p library for your platform... using builtin-java classes where applicable  
2021-03-12 18:12:52,550 INFO client.RMProxy: Connecting to ResourceManager at h  
adoop-name/192.168.56.101:8050  
2021-03-12 18:12:53,290 INFO conf.Configuration: resource-types.xml not found  
2021-03-12 18:12:53,291 INFO resource.ResourceUtils: Unable to find 'resource-t  
ypes.xml'  
Total jobs:0  


| UserName | JobId   | Queue     | Priority       | JobName        | State   | StartTime |
|----------|---------|-----------|----------------|----------------|---------|-----------|
| UsedMem  | RsvdMem | NeededMem | UsedContainers | RsvdContainers | AM info |           |

  
hadoop@hadoop-name:~$
```

- job 명령어는 Map-Reduce 작업을 실시 하였을 때 관련된 정보를 확인할 수 있음
- -list 옵션은 하둡을 실행한 후 Map-Reduce의 모든 작업 목록을 볼 수 있으며, 현재까지 실행한 Map-Reduce가 없기 때문에 Total jobs가 0으로 표현됨