

HDFS 명령어

파일 목록 보기 - ls, lsr

ls

lsr

파일 용량 확인 - du, dus

du

dus

파일 내용 보기 - cat, text

cat

text

디렉터리 생성 - mkdir

mkdir

파일 복사 - put, get, getmerge, cp, copyFromLocal, copyToLocal

put

copyFromLocal

get

getmerge

cp

copyToLocal

파일 이동 - mv, moveFromLocal

mv

moveFromLocal

파일 삭제 - rm

rm

디렉터리 삭제 - rmdir

rmdir

카운트값 조회 - count

count

파일의 마지막 내용 확인 - tail

tail

권한 변경 - chmod, chown, chgrp

chmod

chown

chgrp

0바이트 파일 생성 - touchz

touchz

통계 정보 조회 - stat

stat

복제 데이터 개수 변경 - setrep

setrep

휴지통 비우기 - expunge

expunge

파일 형식 확인 - test

test

-
- 하둡은 사용자가 HDFS를 쉽게 제어할 수 있게 셸 명령어를 제공.

- 명령어 형식
 - cmd : 사용자가 설정한 명령어
 - args : 해당 명령어를 실행할 때 필요한 파라미터

```
.bin/hadoop dfs -cmd [args]
```

파일 목록 보기 -ls, lsr

ls

- 지정한 디렉터리에 있는 파일의 정보를 출력하거나, 특정한 파일을 지정해 정보를 출력하는 명령어.
- 파일 및 디렉터리의 권한 정보, 소유자, 소유 그룹, 생성일자, 바이트 수 등을 확인할 수 있음
- 경로를 지정하지 않을 경우 해당 계정의 홈 디렉터리를 조회
- 사용법

```
.bin/hadoop dfs -ls [디렉터리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -ls /user/hadoop/wordcount_output
```

lsr

- ls 명령어는 현재의 디렉터리만 출력하는 반면 lsr 명령어는 현재 디렉터리의 하위 디렉터리 정보까지 출력
- 사용법

```
.bin/hadoop -lsr [디렉터리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -lsr /user/hadoop/wordcount_output
```

파일 용량 확인 -du, dus

du

- 지정한 디렉터리나 파일의 사용량을 확인하는 명령어로, 바이트 단위로 결과를 출력
- 경로를 지정하지 않을 경우 사용자 계정의 홈 디렉터리에 있는 모든 디렉터리를 출력
- 사용법

```
.bin/hadoop dfs -du [디렉터리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -du /user/hadoop/wordcount_output
```

du

- du 명령어는 디렉터리와 파일별로 용량을 출력하지만, dus는 전체 합계 용량만 출력
- 경로를 지정하지 않을 경우 사용자 계정의 홈 디렉터리에 있는 모든 디렉터리의 사용량을 합산해서 출력
- 사용법

```
.bin/hadoop dfs -dus [디렉터리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -dus /user/hadoop/wordcount_output
```

파일 내용 보기 - cat, text

cat

- 지정한 파일의 내용을 화면에 출력
- 사용법

```
.bin/hadoop dfs -cat [파일 ...]
```

- 사용예

```
.bin/hadoop dfs -cat /user/hadoop/wordcount_output/hadoop-env.sh
```

text

- cat 명령어는 텍스트 파일만 출력 가능. 하지만 text 명령어는 zip파일 형태로 압축된 파일도 텍스트 형태로 화면에 출력.
- 사용법

```
.bin/hadoop dfs -text [파일 ...]
```

- 사용예

```
.bin/hadoop dfs -text /user/hadoop/wordcount_output/hadoop-env.sh
```

디렉터리 생성 - mkdir

mkdir

- 지정한 경로에 디렉터를 생성. HDFS에 이미 존재하는 디렉터를 생성할 경우 오류 발생
- 사용법

```
.bin/hadoop dfs -mkdir [디렉터리 ...]
```

- 사용예

```
.bin/hadoop dfs -mkdir testDir
```

파일 복사 - put, get, getmerge, cp, copyFromLocal, copyToLocal

put

- 지정한 로컬 파일 시스템의 파일 및 디렉터를 목적지 경로로 복사
- 타킷이 되는 디렉토리가 HDFS에 없을 경우 해당 디렉토리를 자동으로 생성하고, 하위 디렉터리가 아닌 디렉터리 자체에 파일을 복사.
- 사용법

```
.bin/hadoop dfs -put [로컬 디렉터리|파일 ...] [목적지 디렉터리|파일]
```

- 사용예

```
.bin/hadoop dfs -put /usr/local/hadoop/etc/hadoop/hadoop-env.sh testDir
```

- 사용예

```
.bin/hadoop dfs -put conf testDir2
```

copyFromLocal

- put 명령어와 동일한 기능을 제공

get

- HDFS에 저장된 데이터를 로컬 파일 시스템으로 복사

- HDFS는 파일의 무결성을 확인하기 위해 체크섬(checksum)기능을 사용하는데, 체크섬을 숨김 파일로 저장하고, 해당 파일을 조회할 때 체크섬을 이용해 확인.
- -crc 옵션을 사용하면 로컬 파일 시스템에 체크섬 파일도 복사
- -ignoreCrc 옵션을 이용하면 해당 파일의 체크섬을 확인하지 않음
- 사용법

```
.bin/hadoop dfs -get <-ignoreCrc> <-crc> [소스 디렉터리|파일 ...] [로컬 디렉터리|파일]
```

- 사용예

```
.bin/hadoop dfs -get /user/hadoop/wordcount_output/hadoop-env.sh usr/local/wordcount_output
```

getmerge

- 지정한 경로에 있는 모든 파일의 내용을 합친 후, 로컬 파일 시스템에 단 하나의 파일로 복사
- 사용법

```
.bin/hadoop dfs -getmerge [소스 디렉터리|파일 ...] [로컬 파일명]
```

- 사용예

```
.bin/hadoop dfs -get /user/hadoop/wordcount_output/_logs/history usr/local/wordcount_log
```

cp

- 지정한 소스 디렉터리 및 파일을 목적지 경로로 복사하는 기능을 제공
- 여러 개의 파일을 복사할 경우 반드시 디렉터리로 복사되도록 설정
- 사용법

```
.bin/hadoop dfs -cp [소스 디렉터리|파일 ...] [로컬 디렉터리|파일]
```

- 사용예

```
.bin/hadoop dfs -get /user/hadoop/wordcount_output/hadoop-env.sh usr/local/wordcount_output/hadoop_env.bak
```

copyToLocal

- get 명령어와 동일한 기능을 제공

파일 이동 -mv, moveFromLocal

mv

- 소스 디렉터리 및 파일을 목적지 경로로 옮김
- 여러 개의 파일을 이동할 경우 반드시 목적지 경로를 디렉터리로 설정해야 함
- 사용법

```
.bin/hadoop dfs -mv [소스 디렉터리|파일 ...] [목적지 디렉터리|파일]
```

- 사용예

```
.bin/hadoop dfs -mv conf conf2
```

moveFromLocal

- put 명령어와 동일하게 동작하지만 로컬 파일 시스템으로 복사된 후 소스 경로의 파일은 삭제
- 사용법

```
.bin/hadoop dfs -put [소스 디렉터리|파일 ...] [로컬 디렉터리|파일]
```

파일 삭제 - rm

rm

- 지정한 디렉터리나 파일을 삭제.
- 디렉터리는 반드시 비어 있는 경우에만 삭제 가능
- 사용법

```
.bin/hadoop dfs -rm [디렉터리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -rm conf2/hadoop-env.bak
```

디렉터리 삭제 - rmr

rmr

- 지정한 파일 및 디렉터를 삭제
- 비어있지 않은 디렉터리도 정상적으로 삭제 가능
- 사용법

```
.bin/hadoop dfs -rmr [디렉터리]
```

- 사용예
 - conf2의 하위 디렉토리 중 디렉터리 이름이 test로 시작하는 모든 디렉터리 삭제
 - test1 / test2 / testDir /testDir2 삭제

```
.bin/hadoop dfs -rmr conf2/test*
```

카운트값 조회 - count

count

- 기본적으로 지정한 경로에 대한 전체 디렉터리 개수, 전체 파일 개수, 전체 파일 크기, 지정한 경로명을 출력
- HDFS는 디렉터리에서 생성할 수 있는 파일 개수나 파일 용량을 제한할 수 있음
- -q 옵션을 사용할 경우 이러한 쿼터 정보를 조회할 수 있음
- 사용법

```
.bin/hadoop dfs -count <-q> [디렉터리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -count -q conf2
```

파일의 마지막 내용 확인 - tail

tail

- 지정한 파일의 마지막 1KB에 해당하는 내용을 화면에 출력
- -f 옵션을 사용하면 해당 파일에 내용이 추가될 때 화면에 출력된 내용도 함께 갱신
- 사용법

```
.bin/hadoop dfs -tail <-f> [파일]
```

- 사용예

```
.bin/hadoop dfs -tail -f wordcount_output/part-r-0000
```

권한 변경 - chmod, chown, chgrp

chmod

- 지정한 경로에 대한 권한을 변경
- 권한 모드는 숫자로 표기하는 8진수 표기법 또는 영문으로 표기하는 심볼릭 표기법으로 설정 가능.
- 해당 파일의 소유자이거나, 슈퍼유저일 때만 가능
- -R 옵션을 사용할 경우 권한 변경을 재귀적으로 실행하여 하위 디렉토리의 설정도 모두 변경됨
- 사용법

```
.bin/hadoop dfs -chmod <-R> [권한모드 ...] [디렉토리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -chmod 777 sample.csv
```

chown

- 지정한 파일과 디렉토리에 대한 소유권을 변경
- -R 옵션을 사용할 경우 재귀적으로 명령어가 실행되어 하위 디렉토리의 설정도 모두 변경됨
- 사용법

```
.bin/hadoop dfs -chown <-R> [변경 사용자명:변경 그룹명] [디렉토리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -chown hduser:hadoop sample.csv
```

chgrp

- 지정한 파일과 디렉토리에 대한 소유권 그룹만 변경
- -R 옵션을 사용할 경우 재귀적으로 명령어가 실행되어 하위 디렉토리의 설정도 모두 변경됨
- 사용법

```
.bin/hadoop dfs -chgrp <-R> [변경 그룹명] [디렉토리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -chgrp hadoop sample.csv
```

0바이트 파일 생성 - touchz

touchz

- 크기가 0바이트인 파일을 생성
- 지정한 파일명이 이미 0바이트 이상인 상태로 저장돼 있다면 오류가 발생
- 사용법

```
.bin/hadoop dfs -touchz [파일 ...]
```

- 사용예

```
.bin/hadoop dfs -touchz touchzExample
```

통계 정보 조회 - stat

stat

- 지정한 경로에 대한 통계 정보를 조회
- 별도의 옵션을 설정하지 않을 경우 해당 디렉터리 혹은 파일이 최종 수정된 날짜를 출력

stat 명령어에서 사용가능한 옵션

Aa 옵션	≡ 내용
%b	블록 단위의 파일 크기
%F	디렉터리일 경우 "directory", 일반 파일일 경우 "regular file" 출력
%n	디렉터리명 혹은 파일명
%o	블록 크기
%r	복제 파일 개수
%y	디렉터리 및 파일 갱신일자를 yyyy-MM-dd HH:mm:ss 형식으로 출력
%Y	디렉터리 및 파일 갱신일자를 유닉스 타임스태프 형식으로 출력

- 사용법

```
.bin/hadoop dfs -stat <출력 포맷> [디렉토리|파일 ...]
```

- 사용예

```
.bin/hadoop dfs -stat %b-%F-%n-%o wordcount_output
```

복제 데이터 개수 변경 - setrep

setrep

- 설정한 파일의 복제 데이터 개수를 변경할 수 있음.
- -R 옵션을 설정할 경우 재귀적으로 명령어가 실행되어 하위 디렉터리의 복제 데이터 개수도 모두 변경됨

- 사용법

```
.bin/hadoop dfs -setrep <-R> -w [복제데이터개수] [디렉토리|파일]
```

- 사용예

```
.bin/hadoop dfs -setrep -w 1 sample.csv
```

휴지통 비우기 - expunge

expunge

- 설정된 휴지통 데이터 삭제 주기와 상관없이 휴지통을 비움
- HDFS에서는 삭제한 파일을 .Trash/라는 임시 디렉터리에 저장한 다음 일정 시간이 지난 후에 완전히 삭제.
- 휴지통 기능과 휴지통 데이터 삭제 주기를 환경설정 파일에서 설정할 수 있음
- 사용법

```
.bin/hadoop dfs -expunge
```

파일 형식 확인 - test

test

- 지정한 경로에 대해 [-ezd] 옵션으로 경로가 이미 존재하는지, 파일 크기가 0인지, 디렉터리인지 확인
- 체크 결과가 맞을 경우 0을 출력
- 사용법

```
.bin/hadoop dfs test -defsz [디렉토리|파일]
```