하둡의 명령어 기본구조

localhost:50070

**hdfs dfs -cmd <args>**

hdfs dfsadmin –report

safe mode 시에

hdfs dfsadmin -safemode leave

**cat**

hadoop fs –ls /

hdfs dfs -cat FILE [FILE ...]   
파일의 내용을 나타낸다. 압축된 파일을 읽기 위해서는 fs 대신에 text 명령어를 사용해야 한다.

**chgrp**

hdfs dfs -chgrp [-R] GROUP PATH [PATH ...]   
파일과 디렉토리에 대한 그룹을 변경한다. -R 옵션은 하위 디렉토리까지 재귀적으로 해당 변경을 적용한다. 사용자는 해당 파일의 소유자이거나 슈퍼유저여야 한다.

**chmod**

hdfs dfs -chmod [-R] MODE [, MODE ...] PATH [PATH ...]   
파일과 디렉토리의 권한을 변경한다. 유닉스와 유사하게 권한 모드는 세 자리 8진수 모드 또는 +/- {rwxX}이다. -R 옵션은 변경을 재귀적으로 적용한다. 사용자는 해당 파일의 소유자이거나 슈퍼유저여야 한다.

**chown**

hdfs dfs -chown [-R] [OWNER] [ :[GROUP]] PATH [PATH ...]   
파일과 디렉토리의 소유자를 변경한다. -R 옵션은 재귀적으로 하위 디렉토리에 설정을 반영한다. 사용자는 슈퍼유저여야 한다.

**copyFromLocal**

hdfs dfs -copyFromLocal LOCALSRC [ LOCALSRC ...] DST   
put과 동일하다. 로컬 파일 시스템으로부터 파일들을 복사한다.

ex) hdfs dfs -copyFromLocal aa.txt /data 

**copyToLocal**

hdfs dfs -copyToLocal [-ignorecrc] [-crc] SRC [SRC ...] LOCALDST   
get과 동일하다. 파일들을 로컬 파일 시스템으로 복사한다.

**count**

hdfs dfs -count [-q] PATH [PATH ...]   
PATH에 있는 모든 파일과 디렉토리에 대한 이름, 사용된 바이트 수, 파일 개수, 하위 디렉토리 개수를 출력한다. -q 옵션은 할당 정보를 나타낸다.

**cp**

hdfs dfs -cp SRC [SRC ...] DST   
소스에 있는 파일들을 목적지로 복사한다. 만약 다수의 소스 파일들이 지정되면, 목적지는 디렉토리여야 한다.

**du**

hdfs dfs -du PATH [PATH ...]   
파일 크기를 나타낸다. 만약 PATH가 디렉토리이면, 그 디렉토리에 있는 각 파일의 크기가 리포트된다. 파일명 앞에 전체 URI 프로토콜이 붙는다. 비록 du는 디스크 사용량을 나타내지만 있는 그대로 받아들여서는 안 된다. 디스크 사용량은 블록 크기와 복제 요소들에 따라 다르기 때문이다.

**dus**

hdfs dfs -dus PATH [PATH ...]   
du와 비슷하다. 그런데 디렉토리에 대해서 dus는 개별적으로 나타내기보다는 합계 (파일 크기)를 리포트한다.

**expunge**

hdfs dfs -expunge   
휴지통을 비운다. 만약 휴지통 기능이 활성화 되어 있으면, 파일이 삭제되었을 때 우선 임시 .Trash/ 폴더로 이동하게 된다. 그리고 사용자가 설정한 시간 후에 .Trash/에 있는 파일이 완전히 삭제된다. expunge 명령어는 .Trash/ 폴더에 있는 파일들을 강제적으로 삭제한다. .Trash/ 폴더에 파일이 존재하는 한 그것을 원래 위치로 이동시켜 해당 파일을 복구 할 수 있다.

**get**

hdfs dfs -get [-ignorecrc] [-crc] SRC [SRC ...] LOCALDST   
파일들을 로컬 파일 시스템으로 복사할 때, 만약 다수의 소스 파일들이 지정되면 로컬 목적지는 디렉토리여야 한다. 만약 LOCALDST가 -이면, 그 파일들은 stdout으로 복사된다. HDFS는 파일에 대한 각 블록의 체크섬을 계산한다. 파일에 대한 체크섬은 숨김 파일에 저장되고, 해당 파일이 HDFS에서 읽힐 때 숨김 파일에 있는 체크섬들은 해당 파일의 무결성을 확인하는데 사용된다. get 명령어에서 -crc 옵션을 사용하면, 숨김 파일도 복사하고, -ignorecrc 옵션은 복사할 때 체크섬을 확인하는 과정을 건너뛴다.

**getmerge**

hdfs dfs -getmerge SRC [SRC ...] LOCALDST [addnl]   
SRC에서 확인된 모든 파일을 가져와 합치고, 로컬 파일 시스템에 존재하는 하나의 LOCALDST 파일에 기록한다. addnl 옵션은 각 파일의 끝을 나타내는 개행 문자를 LOCALDST 파일에 추가한다.

**help**

hdfs dfs -help [CMD]   
해당 명령어(CMD)에 대한 사용법을 나타낸다. 만약 CMD 없이 해당 명령어를 사용하면 모든 명령어에 대한 사용법을 출력한다.

**ls**

hdfs dfs -ls PATH [PATH ...]   
파일과 디렉토리를 조회한다. 각 엔트리는 명칭, 권한, 소유자, 그룹, 크기, 수정일을 나타낸다. 파일 엔트리는 복제 요소도 포함해서 보여준다.

**lsr**

hdfs dfs -lsr PATH [PATH ...]   
ls의 재귀적 버전이다.

**mkdir**

hdfs dfs -mkdir PATH [PATH ...]   
디렉토리를 생성한다. 유닉스의 mdir -p와 같이 상위 디렉토리를 지정하지 않아도 해당 디렉토리까지 포함해서 생성한다.

**moveFromLocal**

hdfs dfs -moveFromLocal LOCALSRC [LOCALSRC ...] DST   
LOCALSRC가 성공적으로 HDFS에 복사된 후에 해당 파일이 삭제되는 것만 빼고는 put과 유사하다.

**moveToLocal**

hdfs dfs -moveToLocal [-crc] SRC [SRC ...] LOCALDST   
아직 구현되어 있지 않다.

**mv**

hdfs dfs -mv SRC [SRC ...] DST   
SRC에 있는 파일들을 DST로 옮긴다. 만약 다수의 소스 파일들이 지정되면, 목적지는 디렉토리여야 한다. 파일 시스템들 간 이동은 허가되지 않는다.

**put**

hdfs dfs -put LOCALSRC [LOCALSRC ...] DST   
로컬 시스템으로부터 파일과 디렉토리를 목적지 파일 시스템으로 복사한다. 만약 LOCALSRC가 -로 설정되어 있으면 입력은 stdin으로 지정되고, DST는 파일이어야 한다.

**rm**

hdfs dfs -rm PATH [PATH ...]   
파일과 빈 디렉토리를 삭제한다.

**rmr**

hdfs dfs -rmr PATH [PATH ...]   
rm의 재귀적 버전이다.

**setrep**

hdfs dfs -setrep [-R] [-w] REP PATH [PATH ...]   
주어진 파일들에 대한 대상 복제 계수를 REP로 설정한다. -R 옵션은 PATH에 의해 확인된 디렉토리들에 대한 파일들의 대상 복제 계수를 재귀적으로 적용한다.

**stat**

hdfs dfs -stat [FORMAT] PATH [PATH ...]   
파일의 통계 정보를 나타낸다. 아래 포맷 지정자에 따른 FORMAT 문자열에 맞게 형식화되어 해당 정보가 출력된다.

%b 파일 크기 (블록 단위)

%F 파일 형식에 따른 "디렉토리" 또는 "일반 파일"

%n 파일명

%o 블록 크기

%r 복제

%y UTC 날짜 yyyy-MM-dd HH:mm:ss 형식

%Y 1970년 1월 1일 (UTC) 이후 경과한 밀리세컨드

**tail**

hdfs dfs -tail [-f] FILE   
파일의 마지막 1KB의 내용을 출력한다.

**test**

hdfs dfs -test -[ezd] PATH   
PATH에서 다음의 형식 점검을 수행한다.

-e PATH 존재 유무, PATH가 존재하면 0을 반환한다.

-z 빈 파일 여부, 파일 길이가 0이면 0을 반환한다.

-d PATH가 디렉토리이면 0을 반환한다.

**text**

hdfs dfs -text FILE [FILE ...]   
파일의 텍스트 내용을 나타낸다. 만약 파일이 텍스트 파일이라면 cat 명령과 동일하다. 압축 형식(gzip과 하둡의 바이너리 시퀀스 파일 포맷)으로 알려진 파일들은 우선 압축을 해제한다.

**touchz**

hdfs dfs -touchz FILE [FILE ...]   
길이가 0인 파일을 생성한다. 만약 길이가 0이 아닌 파일이 이미 존재하면 에러가 발생한다.