**HADOOP**

아파치 **하둡**(Apache **Hadoop**, High-Availability Distributed Object-Oriented Platform)은 대량의 자료를 처리할 수 있는 큰 컴퓨터 클러스터에서 동작하는 분산 응용 프로그램을 지원하는 프리웨어 자바 소프트웨어 프레임워크이다. 하둡은 여러 대의 서버에 데이터를 저장하고, 저장된 각 서버에서 동시에 데이터를 처리하는 방식이다. 하둡은 기존의 RDBMS(Oracle,MS-SQL, MySQL등)을 대치하는 것이 아니다. 즉 트랜젝션이나 무결성을 보장해야하는 데이터처리에는 적합하지 않다. 하둡은 배치성으로 데이터를 저장하고 처리하는데 적합한 시스템이다. 이러한 저장기능을 맞는 시스템이 HDFS이다.

**HDFS**

Hadoop Distributed File System ( HDFS ) is a distributed file system designed to run on commodity hardware로 하둡은 대용량 데이터를 분산 처리할 수 있는 자바 기반의 오픈 소스 프레임워크이다. 하둡은 구글이 논문으로 발표한 GFS(Google File System)과 맵리듀스(MapReduce)를 2005년 더그커팅이 구현한 결과물이다. 하둡은 분산시스템인 HDFS(Hadoop Distributed File System)에 데이터를 저장하고, 맵리듀스를 이용해 데이터를 처리한다

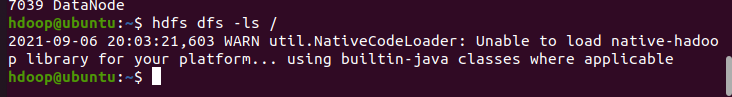
|  |  |
| --- | --- |
|  |  |
| HDFS를 실행하여 NAME NODE를 통해 DN(DATA NODE)에 분산저장 그리고 읽는다. | |

시작은 start-all로 시작하였다가, stop-all로 실행을 종료한다.

리눅스 터미널에서 hdfs dfs - + 명령어(리눅스 명령어)를 통해 실행한다.

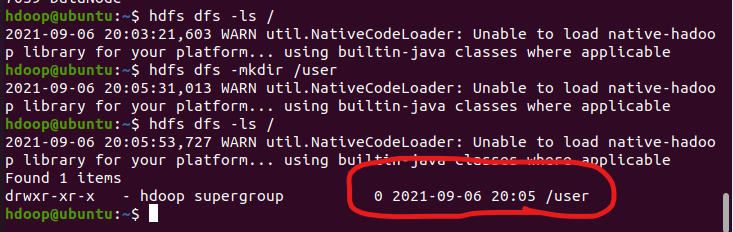
예제

1) hdfs 명령어사용 원칙 : hdfs dfs - + 리눅스 명령어



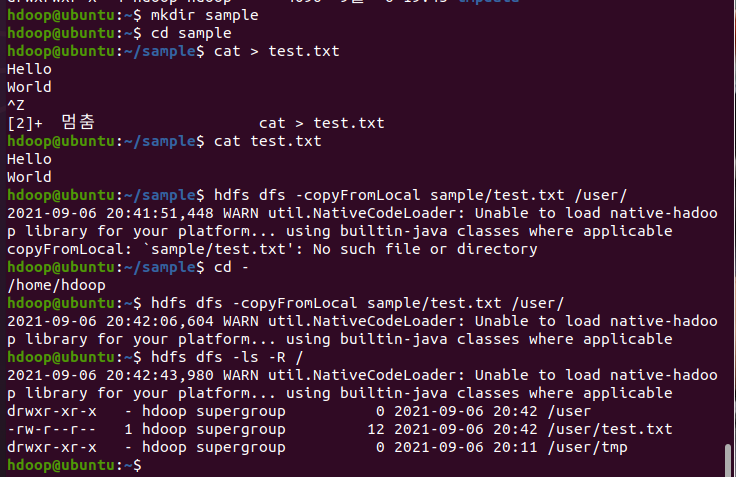
2) 디렉토리 만들기

$ hdfs dfs -mkdir /user



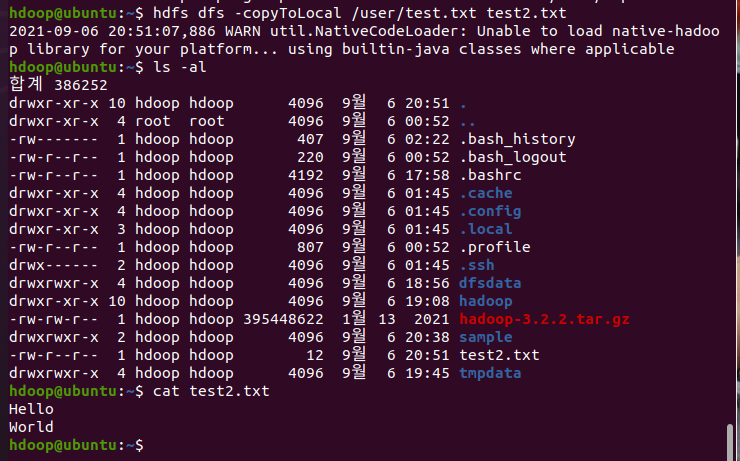
3) 로컬 파일을 HADOOP의 USER 폴더에 저장하는 방법

|  |  |
| --- | --- |
| ~$ mkdir sample  ~$ cd sample  ~/sample$ cat > test.txt  ~/sample$ cat test.txt  ~$ hdfs dfs -copyFromLocal sample/test.txt/user/  ~$ hdfs dfs -ls -R / | (로컬) sample 폴더 만들기  (로컬) sample 폴도 이동  (로컬) test.txt 파일 생성 + 작성  (로컬) test.txt 파일 읽기  (하둡) 로컬파일을 user폴더로 이동  (하둡) 디렉토리 검색 |



4) 하둡 파일을 로컬로 저장

|  |  |
| --- | --- |
| ~$ hdfs dfs -copyToLocal /user/test.txt test2.txt  ~$ ls -R / | (하둡) 하둡파일을 로컬로 이동  (하둡) 확인 |



5) [이어쓰기] 새로작성한 test2.txt를 test.txt 기존 내용에 이어 씀

|  |  |
| --- | --- |
| ~$ cat test2.txt  ~$ hdfs dfs -appendToFile test2.txt /user/test.txt | (하둡) 하둡파일을 로컬로 이동  (하둡) 확인 |

