

하둡 클러스터 환경 구성

이전 요구사항

- jdk설치
- 계정 생성

하둡 설치

```
cd /usr/local
sudo tar xzf hadoop-x.y.z.tar.gz

//압축이 풀린 하둡 파일 전체의 사용자와 그룹을 변경
sudo chon -R hadoop:hadoop hadoop=x.y.z

export HADOOP_HOME=/usr/local/hadoop-x.y.z
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

SSH구성

하둡 제어 스크립트는 SSH를 이용하여 전체 클러스터를 대상으로 수행하도록 개발되었다. 원활한 작업을 위해서는 클러스터에 있는 모든 머신에서 hdfs와 사용자 계정이 암호없이도 접속할 수 있는 SSH로그인 설정을 미리 해두는 것이 좋다.

```
ssh-keygen -t rsa -f ~/.ssh/id_rsa\

생성된 공개키를 클러스터에 참여할 노드들의 ~/.ssh/authorized_keys에 복사해야함
```

HDFS를 처음 설치했으면 사용전에 반드시 파일 시스템을 포맷하여야한다.
hdfs namenode -format

전체 클러스터 데몬 시작 중지는 스크립트를 통해 제공한다. ~~/sbin 하위

하지만 완전 분산모드에서는 스크립트 사용전에 설정파일을 건드려야한다.
커

컴퓨터의 목록을 저장하는 파일 ⇒ **slaves**, 한 행당 하나씩 호스트 명이나 IP주소를 기록하면된다.

이파일은 기본적으로 하둡의 설정 디렉터리에 있지만 , `hadoop-env.sh`파일의 `HADOOP_SLAVES` 속 성 값을 변경하면 다른 디렉터리로 옮길 수 있다.

이 파일은 네임노드에서만 사용됨으로 다른 노드들로 옮길 필요없다.

start-dfs.sh 스크립트에 사용되는 명령

- `hdfs getconf -namenodes` ⇒ `fs.defaultFS`에서 네임노드의 호스트명을 찾는다.
- `slaves` 파일의 목록에 있는 각 컴퓨터에서 데이터노드 시작
- `hdfs getconf -secondarynamenodes` , 보조네임 노드 시작

사용자 디렉터리 생성

하둡 클러스터를 구동했다면 사용자들이 하둡 클러스터에 접근할 수 있도록 해야한다.

각 사용자별로 홈 디렉터리를 생성하고 해당 디렉터리에 대한 소유권한을 부여한다.

```
hadoop fs -mkdir /user/<username>
```

```
hadoop fs -chown <username>:<username> /user/<username>
```

하둡 환경설정

hadoop-env.sh : 하둡을 구동하는 스크립트에서 사용되는 환경변수

mapred-env.sh : 맵 리듀스 구동하는 스크립트에서 사용되는 환경변수

yarn-env.sh : YARN을 구동하는 스크립트에서 사용되는 환경변수

`core-site.xml` : HDFS,mapreduce,yarn 공통적으로 사용되는 I/O설정 과 같은 하둡 코어를 위한 환경설정 구성

`hdfs-site.xml` : 네임 노드 ,보조 네임노드,데이터 노드등과 같은 HDFS 데몬을 위한 환경 설정 구성

mapred-site.xml:잡 히스토리 서버 같은 맵리듀스 데몬을 위한 환경 설정

yarn-site.xml: 리소스 매니저, 웹 애플리케이션 프록시 서버, 노드매니저와 같은 yarn 데몬
위한 환경 설정

slaves:: 데이터 노드와 노드매니저 구동할 컴퓨터 목록

hadoop-metrics2.properties:매트릭 표시

log4j.properties:시스템 로그 네임노드 감사 로그 jvm프로세스의 작업 로그

hadoop-policy.xml: 하둡을 보안 모드로 구동할 때 사용되는 접근제어 목록에 대한 환경 설
정 구성

위 파일들은 etc/hadoop 디렉터리에 있다.

환경설정 관리

하둡에서는 환경 설정 정보를 단일 또는 전역 장소에서 관리하지 않는다. 그대신 클러스터에
있는 각 하둡 노드는 자신만의 환경 설정 파일을 가

환경설정

hadoop-env.sh의 JAVA_HOME에 자바 라이브러리 위치 지정,

없으면 쉘 환경변수인 JAVA_HOME을 찾긴 함

하둡의 각 데몬은 1GB 메모리를 기본할당 , hadoop-env.sh 에서 HADOOP_HEAPSIZE
에 의해 제어

네임노드에 할당할 메모리는 HADOOP_NAMENODE_OPTS 속성에 JVM 옵션으로 메모
리 용량을 설정 ???

시스템 로그파일

\$HADOOP_HOME/logs에 저장된다.

hadoop-env.sh 에서 HADOOP_LOG_DIR 변경으로 변경가능

보통 로그파일은 /var/log/hadoop에 저장, hadoop-env.sh에 다음과 같이 추가

```
export HADOOP_LOG_DIR=/var/log/hadoop
```

반드시 정의해야하는 속성

core-site.xml

fs.defaultFS ⇒ hdfs://namenode/

위 속성은 모드에 따라 달라짐

hdfs-site.xml

yarn-site.xml

생략—

<https://newly0513.tistory.com/144>

<https://mungyo.tistory.com/21?category=980499> 도커를 이용