

## HDFS 시작하기

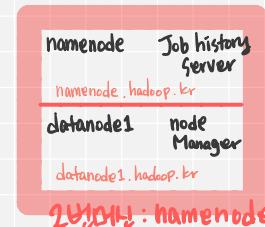
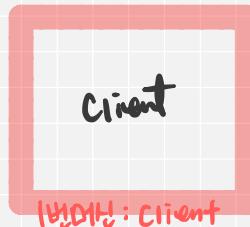
\* HDFS 터미널 실행시에만 가능합니다

- ① ssh namenode ⇒ 1번마シン (client)에서 2번마シン (namenode)으로 SSH 접속
- ② hostname ⇒ 호스트명 확인
- ③ hadoop/bin/hdfs namenode -format ⇒ 초기화
- ④ ls /data
- ls -R /data ⇒ 디렉토리별 파일 목록 확인
- ⑤ exit ⇒ 2번마シン에서 다시 1번마シン (client)로

## YARN 서비스 시작하기

- ① HDFS 시작하기
- ② YARN 시작하기
- ③ MapReduce 시작하기

마진에 영향 cf. Page 6에 응시, 노트와 대비 참고!



YARN은 모든 node에  
존재하는 Resource Manager  
2번마シン에만 Secondarynamenode  
존재하는

## 1 HDFS 시작하기

- ① ssh namenode ⇒ 2번 마シン으로 디스
- ② hostname ⇒ namenode.hadoop.kr 으나 터미널
- ③ jps ⇒ java进程 목록 터미널
- ④ hadoop/sbin/start-dfs.sh ⇒ hdfs 서비스 시작
- ⑤ jps ⇒ 2번 마シン에서 HDFS 서비스를 확인
- ⑥ ssh secondnode ⇒ 3번마シン으로 이동
- ⑦ jps ⇒ 3번 마シン에서 HDFS 서비스를 확인
- ⑧ exit ⇒ 2번마シン에 2번 마シン으로
- ⑨ exit ⇒ 2번마シン에 1번 마シン으로



HDFS의 시작과 종료는  
NameNode에서만 가능합니다

## 2 YARN 시작하기

- ① \$ ssh secondnode ⇒ 1번마シン에서 기본마シン으로
- ② \$ hostname ⇒ 호스트명 터미널
- ③ \$ jps ⇒ 자바 프로세스 목록
- ④ \$ hadoop/sbin/start-yarn.sh ⇒ YARN 서비스 시작
- ⑤ \$ jps ⇒ 1번 마シン에 YARN 서비스 터미널
- ⑥ \$ ssh namenode ⇒ 2번마シン에서 2번 마シン으로 SSH 접속
- ⑦ \$ jps ⇒ 2번 마シン에서 HDFS 서비스를 확인
- ⑧ \$ exit ⇒ 3번마シン, 2번 마シン으로
- ⑨ \$ exit ⇒ 3번마シン, 1번 마シン으로

YARN 서비스의 시작과 종료는  
NameNode에서만 가능합니다

## 3 MapReduce 시작하기

- ① \$ ssh namenode ⇒ 1번마シン (client)에서 2번마シン (namenode)으로 SSH 접속
- ② \$ hostname
- ③ \$ jps
- ④ \$ hadoop/sbin/mr-jobhistory-daemon.sh start historyserver ⇒ 감시하던 서비스 시작
- ⑤ \$ jps ⇒ 2번 마シン에서 기본마シン 서비스 터미널
- ⑥ \$ exit ⇒ 3번마シン, 1번마シン

감시하던 마シン의 시작과  
종료는 NameNode에서만 가능합니다

## ① HDFS 테스트하기

## ② Pig3 테스트하기

- pig3 HDFS / Map Reduce 테스트
- Grunt 데이터 테스트하기

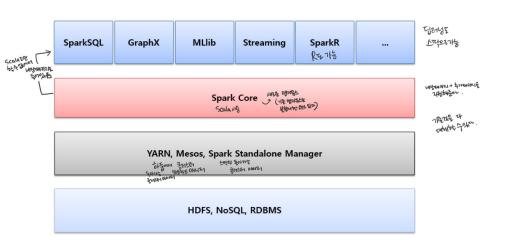
## ③ Hadoop 서비스 종단하기



## Apache Spark

↳ 병렬을 원한 사용된다. (병렬을 원하는 다른)  
HIVE, Pig는 MapReduce보다, Framework가 훨씬

- Hadoop의 Map Reduce와 유사한 API를 분산형 프레임워크
- 다양한 분석기능 제공, 예: HDFS, NoSQL, RDBMS
- 예: HIVE, Pig, Mahout, R, Storm 같은 모두 지원 가능
- 쇠퇴한 MapReduce를 대체할



- 병렬 처리를 원하는 경우에만