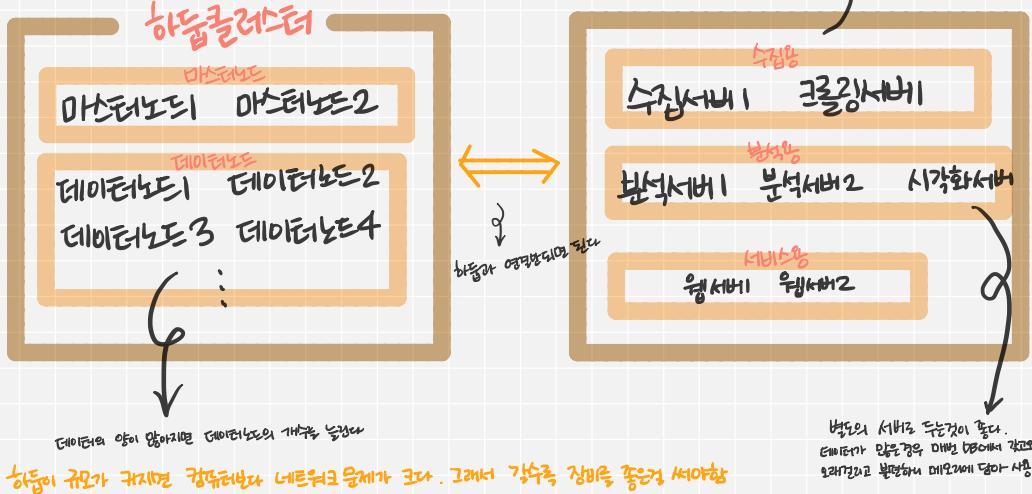


## 인프라 & 설치

### Infra 아키텍처

- PC 1대 → 의사분산모드 (이서모드): 개발용, 교육용
- PC 2대 → 하둡을 활용할 수 있는 커리큘럼에 대해서도 활용하는데 문제x
- PC 3대 이상 → 분산모드

데이터를 대량으로 처리하는  
수집과 분석을 위한 시스템



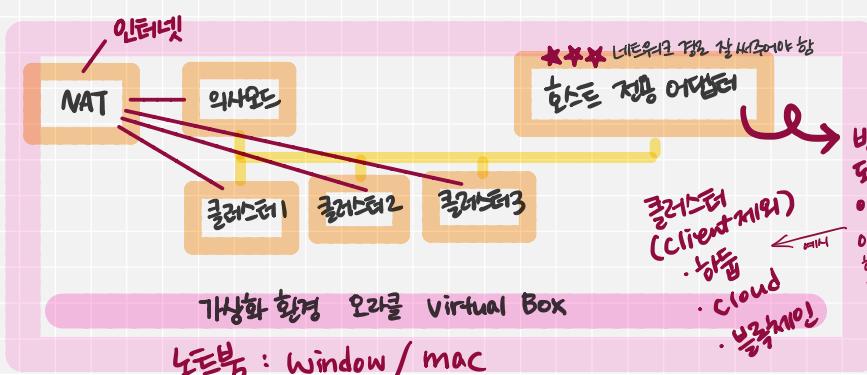
## hadoop 설치 가이드

### 설치환경 설정

- ① 가상화 환경구축 → virtual Box 설치
- ② 가상머신 만들기 → 이서모드 1대, 클라우드 3대
- ③ 리눅스설치 → 이서모드: CentOS 6.x 32비트 / 클라우드 CentOS 7.5 64비트
- ④ 설치준비 → 네트워크설정, JDK설정, 필요한 패키지 다운로드
- ⑤ hadoop설치 → Apache Hadoop 2.6.4
- ⑥ Pig/Hive 설치 → Pig: 0.16.0 / Hive: 1.2.1
- ⑦ Spark 설치 → Spark 1.6.2

노트북 운영체제: Windows(권장), Linux는 개인집으로 수업하시는 분들은  
CPU 2GHz 이상, RAM 4GB 64bit  
SSD 30GB, 여유공간 50GB 이상

## 가상화 환경 네트워크 구성



## 데몬

데몬이란?

↳ 행정 데이는 운영진을 경제 (웹서버=웹데온, DB서버=DB데온), 모든 데온은 포트를 가지고 있다

- 서비스
  - 1. HDFS
    - ↳ 네임노드 1, 2 (마스터노드)
    - ↳ 데이터노드 1, 2, 3...
  - 2. YARN
    - ↳ 리소스마니저 (1개의 포트)
    - ↳ 노드매니저

application → 시작과 끝이 존재

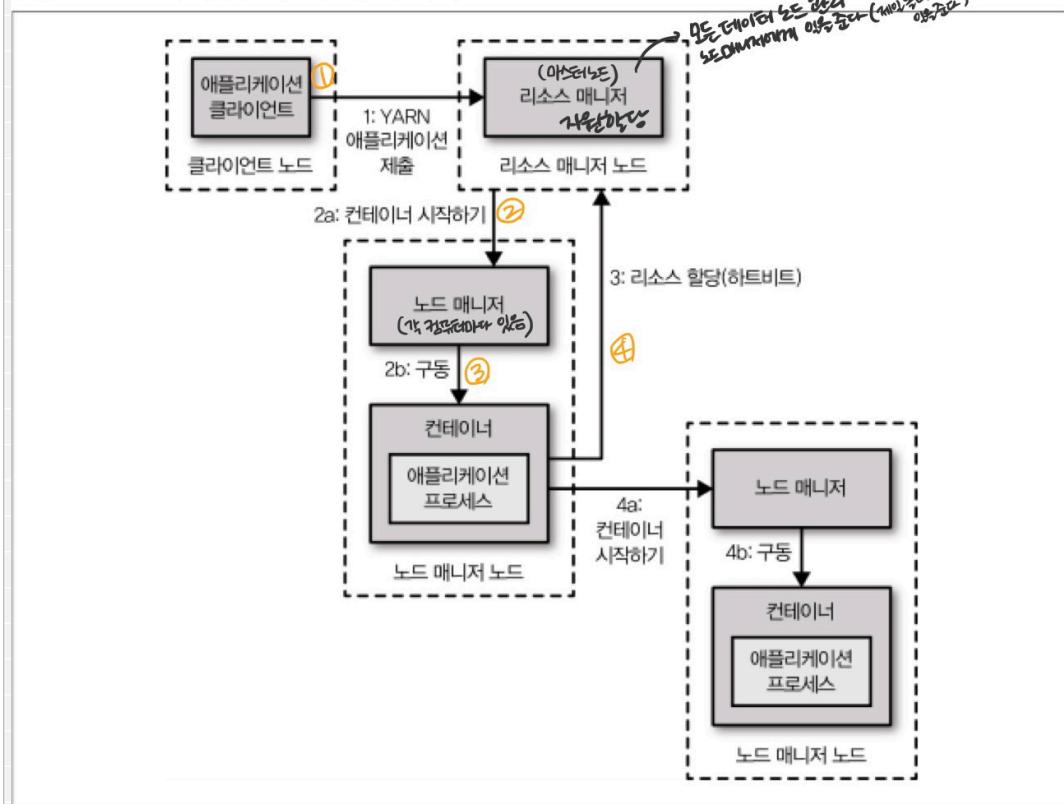


## YARN

YARN이란?

↳ 병렬의 클러스터의 자원 관리 시스템

그림 4-2 YARN이 애플리케이션을 구동하는 방식



## 주요 허둡 설정 파일

