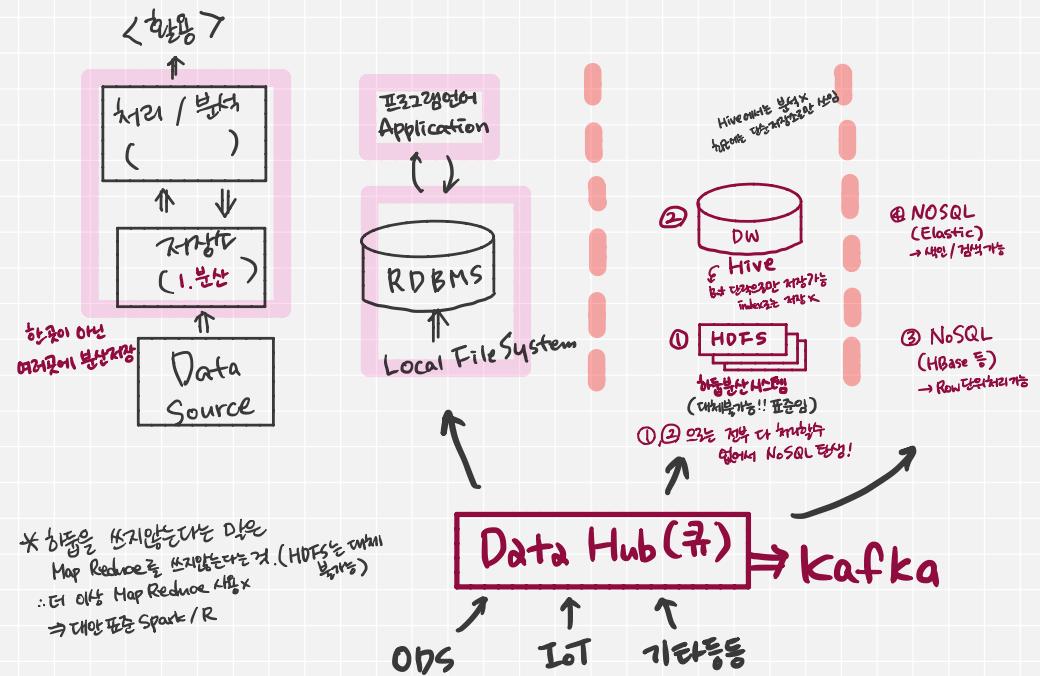


스파크 아키텍처



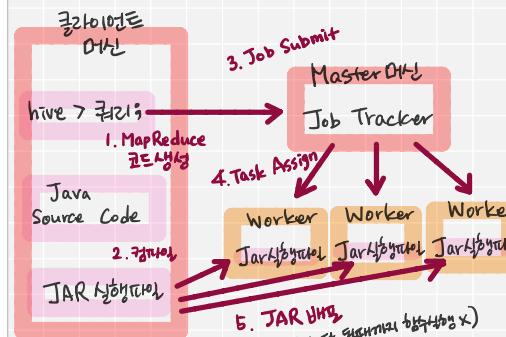
MySQL MySQL
MySQL이 기능을 가지고 있어서 MySQL은 할 수 있지만
MySQL과 다른 부분이 있어 MySQL도 제공된다.

부산병历처리에 높이학원 함수형 030

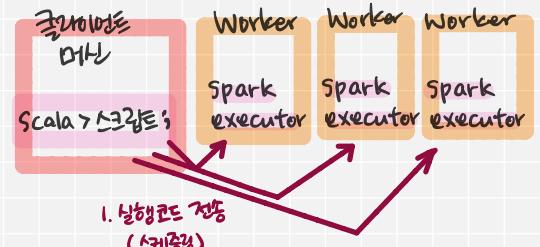
- Spark Core 프레임워크 : Scala / Java⁸ 대체로 JVM!

- 대화적인 험수집방법에서는 SQL로 알맞다!
 - 일정이 같으면 평균값을 끝나면 \Rightarrow 평균값이나가 있으므로 이용된다
 - 평균값이나가를 사용하면 평균이 평균값을 이용된다. \Rightarrow 평균을 100/100으로 이용된다

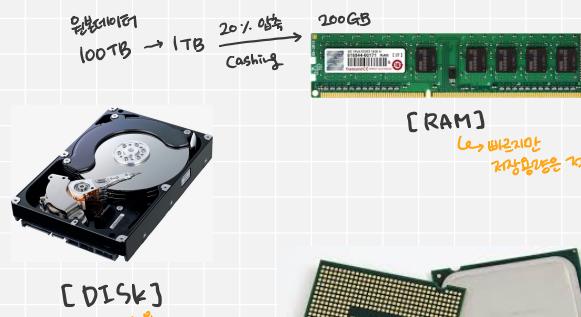
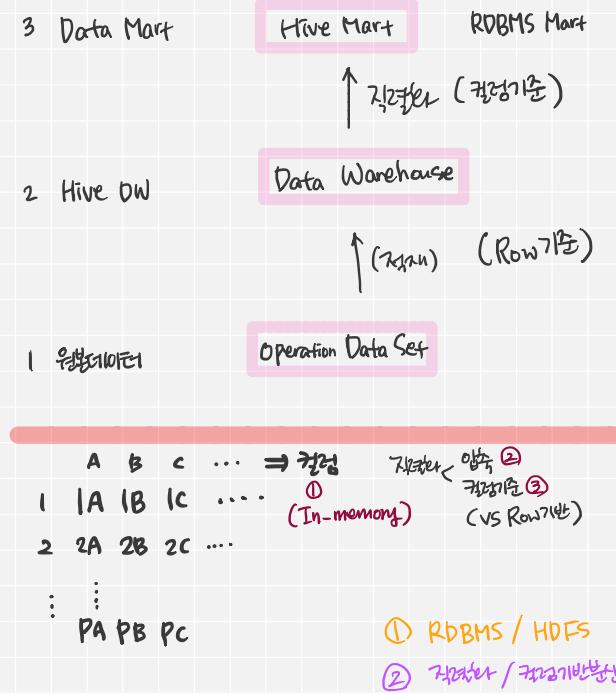
하둡 Map Reduce 방식



Spark API



cf. YARN이 있다
설정된 모듈은 그대로 실행된다면 YARN이 퍼포먼스
제한으로 인해 더디다



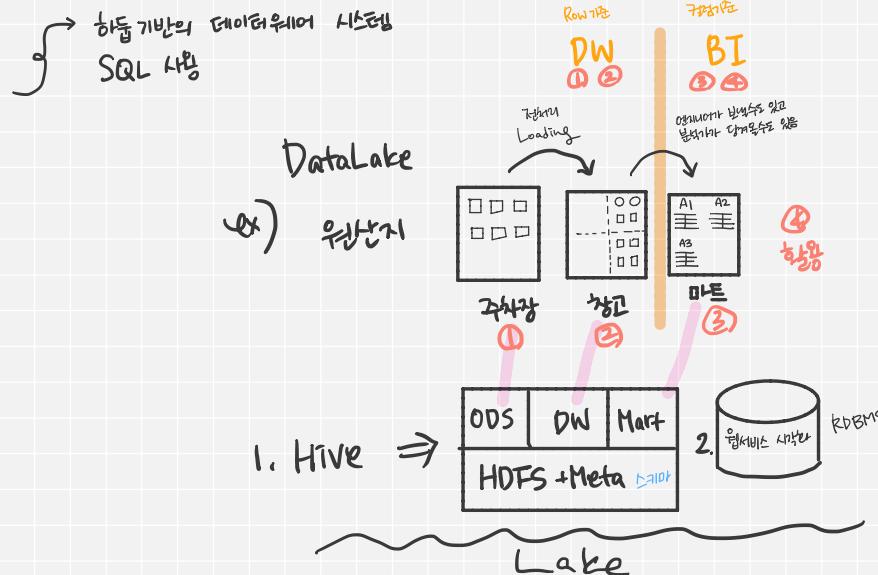
[DISK] HDFS 데이터베이스



[CPU]

하둡 컴퓨터 사양

- SSD를 사용하지 않는 것이 default
 - ↳ 데이터가 너무 많기 때문에 비용면에서 일어나므로 '데이터트래픽장'이 비용면으로 절대 쓰지 말라고 한다
 - 구성 전과 후로 바뀌는데 RAM이 많아질 때 고래는 SSD 사용을 권장한다
 - ↳ 고래는 SSD에 가능성을
프로그램 끌어온 SSD에 있는 내용까지
 - DISK을 잘쓰는 이유, 기록과 불리는 이유, RAM 또는 SSD이나 저장장치



HIVE 작업 절차

- ① [root 접근] 네임노드(24104번)에서 MySQL 접속
- ② [hadoop 접근] Hive Client 접속
ID: hadoop
- ③ 네임노드가 있는곳에서 (ssh namenode) HDFS 시작하기 ⇒ hadoop/sbin/start-dfs.sh
- ④ 리스트너노드가 있는곳에서 (ssh Secondarynode) YARN 시작하기 ⇒ hadoop/sbin/start-yarn.sh
- ⑤ 클라이언트로 명령어로 Hive CLI 시작하기!