

# LESSON 1

## Hadoop2의 개선사항



# Hadoop2의 개선사항

## • Hadoop1의 약점

### □ 배치(Batch)에 최적화

- 스트리밍과 같은 작업은?
- 맵/리듀스 형태의 작업에만 최적화

### □ 자원활용방법의 문제점

- 잡트래커가 맵/리듀스작업이 잘 되고 있는지 관리
- 잡트래커가 시스템 사용효율도 같이 관리
  - 맵작업이 바쁠경우/리듀스 작업이 바쁠 경우

### □ 네임노드의 SPOF의 문제점

- 네임노드와 세컨더리 네임노드가 동시에 장애를 일으키면?



# Hadoop2의 개선사항

## ▣ Hadoop2의 해결책

### ▣ 아키텍처의 변화

### ▣ 양(YARN)의 도입

- YARN을 통해서 맵/리듀스작업 처리
- 맵/리듀스작업은 YARN의 하나의 애플리케이션이 됨
  - 다른 형태의 서비스 지원(스트리밍 포함)
  - 애플리케이션 마스터(AM:Application Master)
    - 애플리케이션별 작업관리
- 잡트래커 + 태스크트래커 → 노드매니저 + 리소스매니저
- 시스템 리소스활용과 잡(작업) 진행사항관리를 분리



# Hadoop2의 개선사항

## • Hadoop2의 해결책

- 아키텍처의 변화
- 주키퍼(Zookeeper)의 도입

- 네임노드 고가용성(HA:High Availability)가 가능하도록 구성

- Active-Standby

- 리소스매니져 이중화



# Hadoop2의 개선사항

## HADOOP 1.0

*Single Use System  
Batch Apps*

Data Processing  
Frameworks  
(Hive, Pig, Cascading, ...)

MapReduce  
(distributed data processing  
& cluster resource management)

HDFS 1  
(redundant, reliable storage)

## HADOOP 2.0

*Multi Use Data System  
Batch Apps*

Standard SQL  
Processing  
*Hive*

Batch  
*MapReduce*

Online Data  
Processing  
*Hbase, Accumulo*

Interactive  
*Tez*

Real Time Stream  
Processing  
*Storm*

Others  
...

Cluster Resource Management  
*YARN*

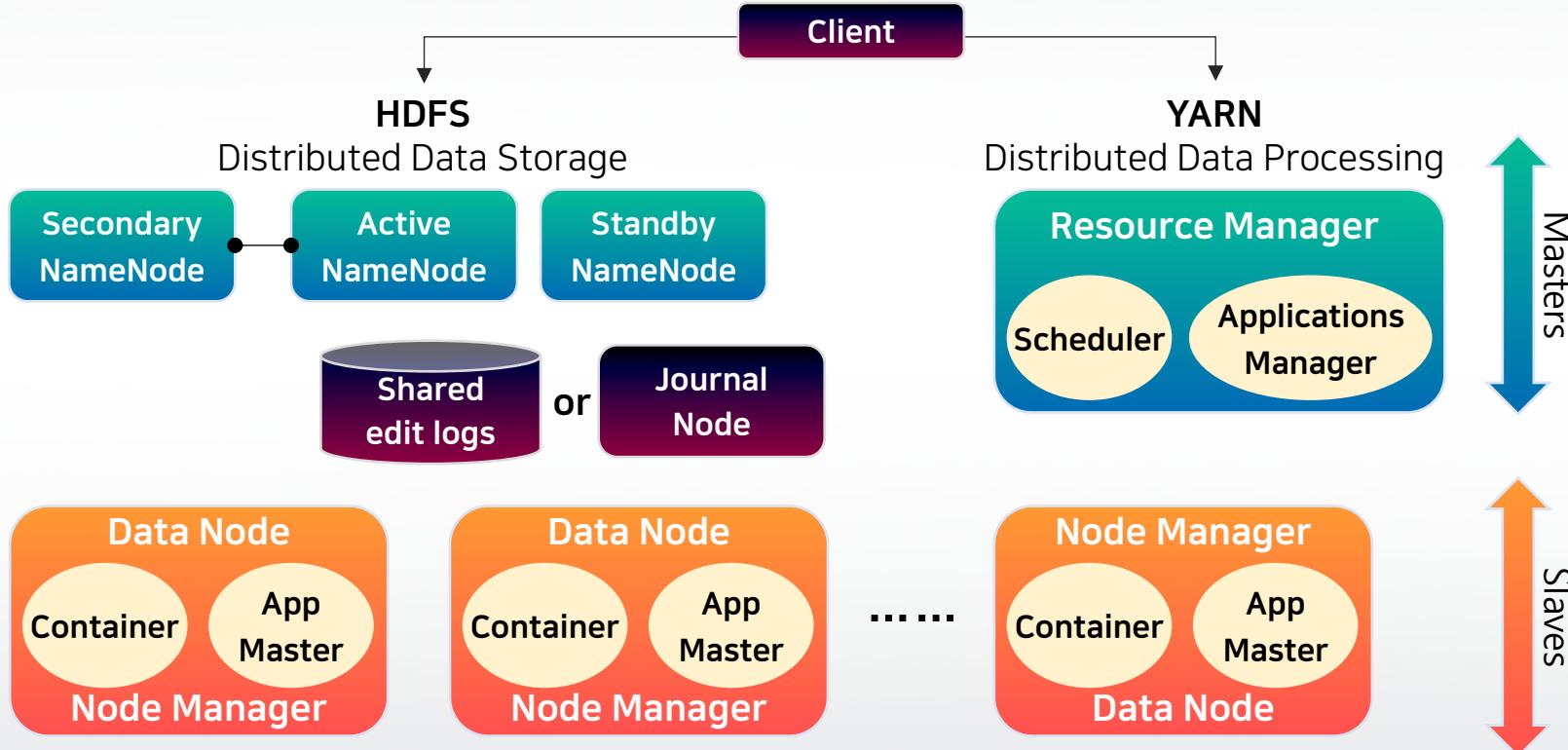
Redundant, Reliable Storage  
*HDFS 2*

Interact with all data in  
Multiple ways simultaneously



# Hadoop2의 개선사항

## Apache Hadoop 2.0 and YARN



# LESSON 2

## Hadoop2 설치/세팅



# Hadoop2 설치/세팅

## ☞ 하둡 1.x와 2.x 폴더구조/사용법 차이점

하둡 1.x

하둡 2.x

- conf 폴더 안에 설정 파일들이 있음
- 예제가 앱축 풀 폴더(\$HADOOP\_HOME)에 있음
- 1.x 초기 버전은 \$HADOOP\_HOME/bin에 실행파일 및 스크립트파일 존재
- 1.x 후기 버전부터 2.x 부터는 \$HADOOP\_HOME/bin과 \$HADOOP\_HOME/sbin 에 나눠 저장됨
- sbin에는 .sh와 같은 스크립트가 저장
- 1.x 초기버전은 start-all.sh/stop-all.sh에 HDFS/MapReduce 모두 수행/정지
- 1.x 후기 버전부터 hdfs의 수행/정지는 start-dfs.sh/stop-dfs.sh로,
- 맵리듀스 수행/정지는 start-mapred.sh/stop-mapred.sh로 나눠짐



# Hadoop2 설치/세팅

## ▣ 하둡 1.x와 2.x 폴더구조/사용법 차이점

하둡 1.x

하둡 2.x

- 설정파일위치
  - \$HADOOP\_HOME/etc/hadoop 폴더
  - \$HADOOP\_CONF\_DIR
- 예제의 위치
  - \$HADOOP\_HOME/share/hadoop/mapreduce
- 패키지명 변경
  - hadoop-core(1.x) -> hadoop-common(2.x)
- 맵리듀스 수행/정지는 start-yarn.sh/stop-yarn.sh로 나눠짐



실습동영상 추가  
hadoop-2.7.2 pseudo distributed