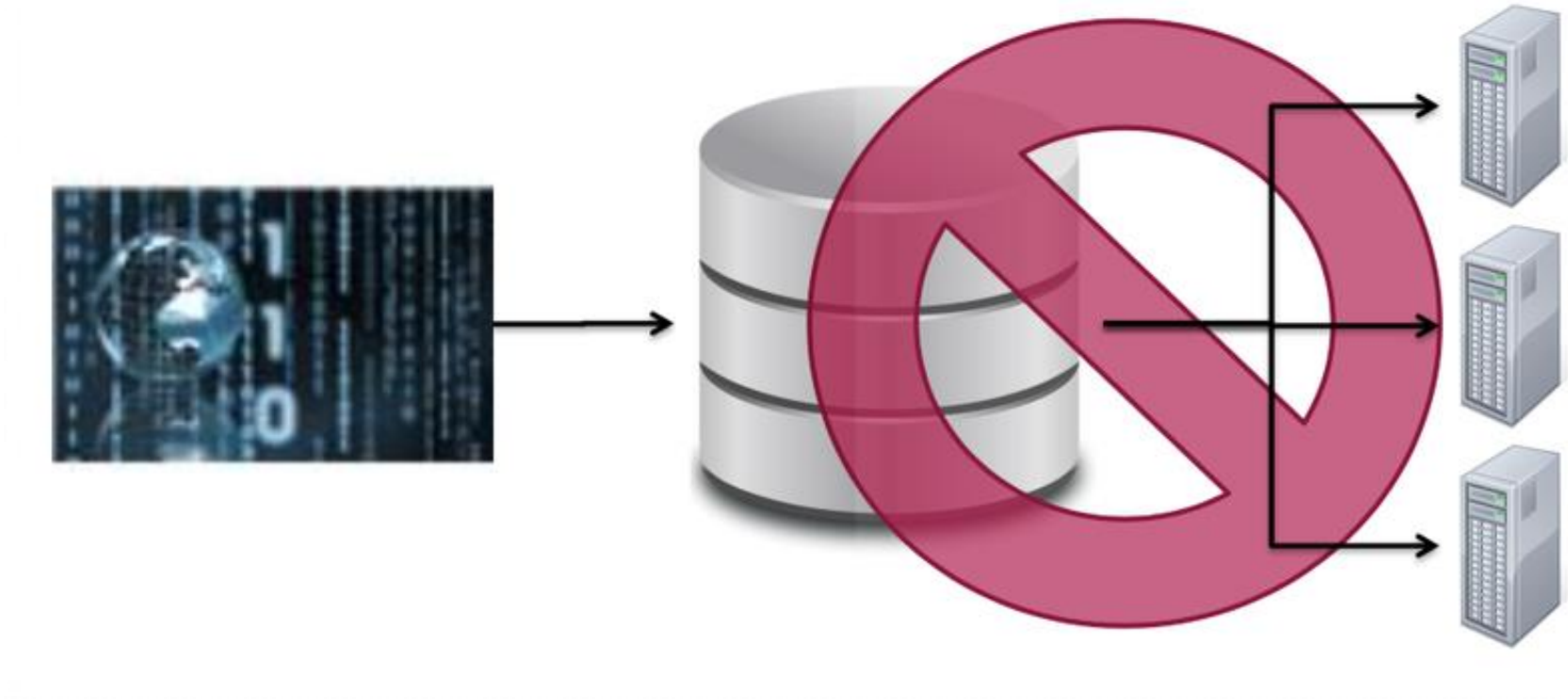


*2018 Winter
Team
Introduction*

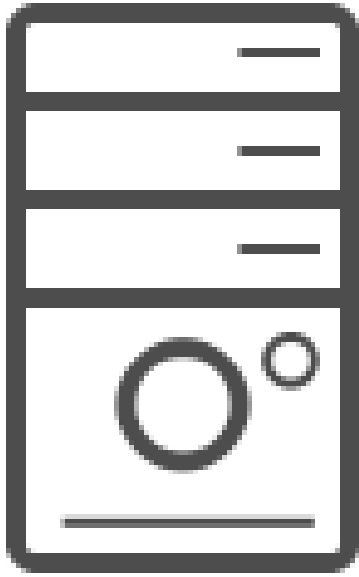


분산 처리 시스템의 필요성



점차 필요한 데이터의 양이 커짐에 따라 여러 프로세서가 중앙 저장소에 접근하는 것은 심각한 성능 저해 초래

분산 처리 시스템의 필요성



<Scale Up>

VS



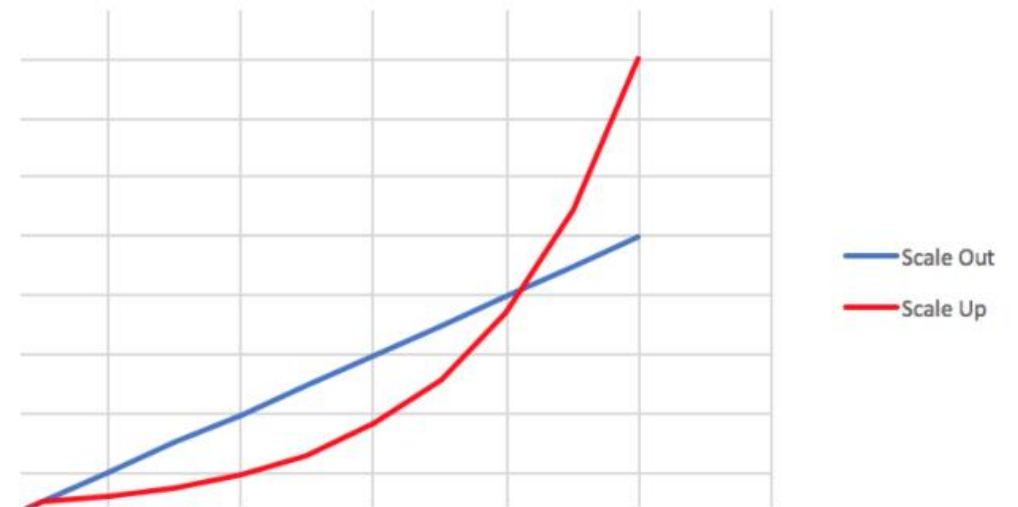
<Scale Out>

분산 처리 시스템의 필요성

Scale Out의 단점

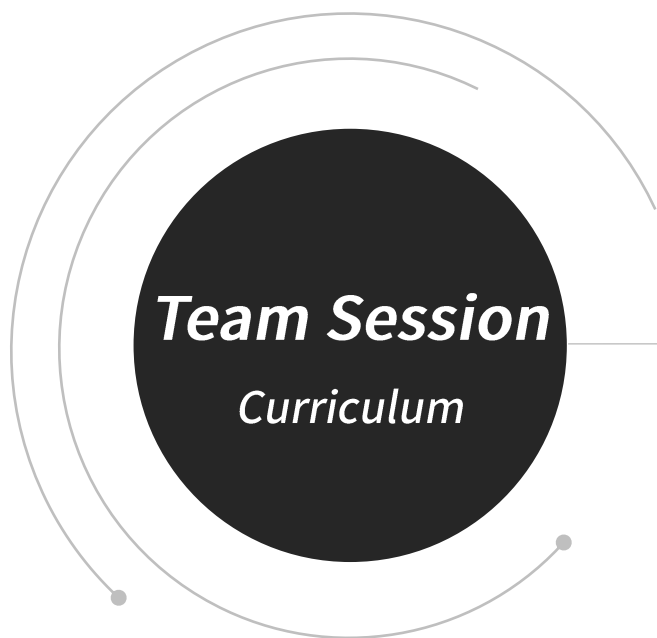
- 주어진 일을 수행하기 위해 머신 간에 통신하는 비용이 발생
- 더 많은 머신을 관리해야 하기 때문에 불안정
- 단순히 N개의 머신을 이어 붙였다고 N배의 성능이 나오지 않음

Scale Out의 장점



* X축은 성능, Y축은 비용을 의미

Data Engineering Session Curriculum



Hadoop

모든 분산 처리 시스템의
기저에서 사용되는
하둡의 파일 시스템 구조,
분산 처리 시스템의 구조 등에
대하여 공부합니다.



Spark

분산 처리 시스템과
Spark를 다루는 방법,
Spark에서 제공하는 여러 라
이브러리를 학습합니다.



Spark

Spark 세션에서 직접 강의하고
다른 팀원들을 위한
과제를 출제하며 Spark에 대한
더 깊은 이해를 할 수
있습니다.



Web Crawling

인터넷에서 데이터를
수집하는 방법을 알아보고
다양한 사이트를 크롤링하며
실습하는 시간을 갖습니다.

MySQL

관계형 데이터베이스 관리
시스템 (RDBMS) 의 구조와
SQL를 사용하는 방법을
배웁니다.



NoSQL

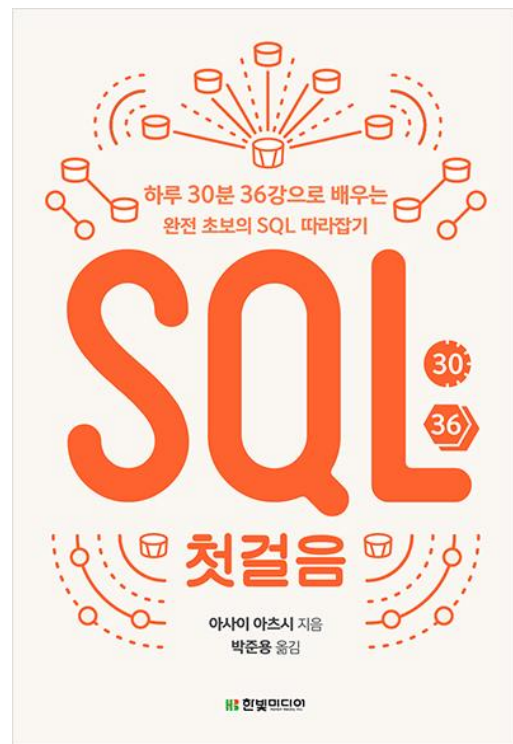
빅데이터와 웹 어플리케이
션을 다룰 때에 자주 사용되
는 NoSQL을 배웁니다.

Study Session
Curriculum



정규 팀 세션

- 토요일 12 : 30 – 15 : 00
- 하둡과 스파크 등 분산 처리 시스템 학습
- 팀원들이 준비한 강의로 구성
- 2주 간 프로젝트를 진행하여 실습
- 매주 주어진 과제에 대한 질문 시간 및 피드백 有



기수 별 스터디

- 크롤링, SQL, NoSQL 등 다양한 주제로 학습
- 스터디 팀원 간 돌아가며 발제
- 기수 별로 시간을 조정하여 모임
- 교재와 스터디 주제는 팀원 들의 상의에 따라 변경 가능

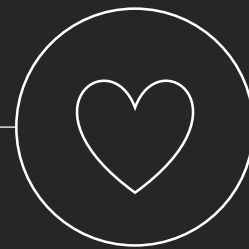
학기 중 선배 특강

- 현직자를 초빙하여 학기 중 1-2회 특강 ex) SQL을 이용한 추천 시스템 구축 특강

자유 스터디

- 공부하고 싶은 주제가 있을 시 팀 내에서 자유롭게 스터디 구성
- 현재 Spark 보충 스터디, 딥러닝 스터디 등 다양한 스터디 진행 중

Advantage of Team Engineering



접하기 어려운 기술 학습

학부생이 혼자 공부하기
쉽지 않고 타 학회에서도
잘 다루지 않는 분산 처리 기술을
접할 수 있습니다.

체계적인 커리큘럼

전체 팀 세션과 기수별 스터디에
대한 자세한 커리큘럼을 계획,
실행하고 있습니다.

화목한 분위기

정기적인 친목 도모 행사와
기수 별 스터디로
팀원 간 관계가 돈독합니다.

Advantage of Team Engineering



접하기 어려운 기술 학습

학부생이 혼자 공부하기
쉽지 않고 타 학회에서도
잘 다루지 않는 분산 처리 기술을
접할 수 있습니다.

체계적인 커리큘럼

전체 팀 세션과 기수별 스터디에
대한 자세한 커리큘럼을 계획,
실행하고 있습니다.

화목한 분위기

정기적인 친목 도모 행사와
기수 별 스터디로
팀원 간 관계가 돈독합니다.

Advantage of Team Engineering 흔히 접할 수 없는 기술 학습

Java[™]
ubuntu[®]



- 분산 처리 시스템인 하둡이나 스파크를 학습하기 위해서는 Java와 Linux에 대한 사전 지식이 요구됨
➡ Java와 간단한 Linux 명령어를 가르치는 세션을 마련하여 진입 장벽을 최소화

Advantage of Team Engineering



접하기 어려운 기술 학습

학부생이 혼자 공부하기
쉽지 않고 타 학회에서도
잘 다루지 않는 분산 처리 기술을
접할 수 있습니다.

체계적인 커리큘럼

전체 팀 세션과 기수별 스터디에
대한 자세한 커리큘럼을 계획,
실행하고 있습니다.

화목한 분위기

정기적인 친목 도모 행사와
기수 별 스터디로
팀원 간 관계가 돈독합니다.

Advantage of Team Engineering 체계적인 커리큘럼

일시	강의자	내용	Cloudera_Developer_Training	시작하세요! 하둡 프로그래밍
2018-01-20	이한솔	Introduction to Apache Hadoop and its Ecosystem - Hadoop Basic Concepts and HDFS - Yarn - Introduction to MapReduce - Hadoop Clusters and Hadoop Ecosystem - Hadoop 활용 사례	16 ~ 142	Part 01 하둡 기초 - 01. 하둡 살펴보기 - 02. 하둡 개발 준비 - 03. 하둡 분산 파일 시스템 - 04. 맵리듀스 시작하기

*2018년 겨울방학 Hadoop 2주차 강의 계획

- 장기적인 계획부터 매주 세션에 대한 자세한 정보까지 문서화하여 공유
- 매주 강의자가 제공하는 과제와 학기 말 진행되는 프로젝트로 배운 내용에 대한 충분한 복습 진행
- 보충 스터디 세션과 온라인 Q&A 시간을 통해 부족한 부분 보충

Advantage of Team Engineering



접하기 어려운 기술 학습

학부생이 혼자 공부하기
쉽지 않고 타 학회에서도
잘 다루지 않는 분산 처리 기술을
접할 수 있습니다.

체계적인 커리큘럼

전체 팀 세션과 기수별 스터디에
대한 자세한 커리큘럼을 계획,
실행하고 있습니다.

화목한 분위기

정기적인 친목 도모 행사와
기수 별 스터디로
팀원 간 관계가 돈독합니다.

Advantage of Team Engineering 화목한 분위기



- 정기적인 친목 도모 행사 진행 (신년회, 정기 회식 등)
- 팀 세션과 다양한 스터디로 인한 깊은 친밀감 형성

엔지니어링이 원하는 인재



대용량 데이터를 처리하는 것에 관심이 있는 분



긴 호흡으로 공부할 수 있는 분



여러 가지 기술을 배우고자 하는 의지가 있는 분

12기 오픈 세션



2018년 1월 20일 토요일 12시 30분

“Introduction to Apache Hadoop and its Ecosystem”을 주제로 오픈 세션 진행

Thank You!