

빅데이터 컴퓨팅 응용 강의 소개

강의 목표
강의 내용
교재 및 강의 방법
과제

순천향대학교 컴퓨터공학과
이 상 정

순천향대학교 컴퓨터공학과

1

빅데이터 컴퓨팅 응용



강의 목표

- 빅 데이터 컴퓨팅의 기본 개념, 원리 및 응용 기법 강의
 - 빅 데이터 컴퓨팅 소개
 - 아파치 하둡 (Apache Hadoop), 분산 파일 시스템
 - 맵리듀스 (MapReduce), 분산 처리 시스템
 - 아파치 스파크 (Apache Spark), 분산 데이터 처리 통합 프레임워크
 - 아파치 HBase (Apache HBase), 열-지향 분산 데이터베이스



순천향대학교 컴퓨터공학과

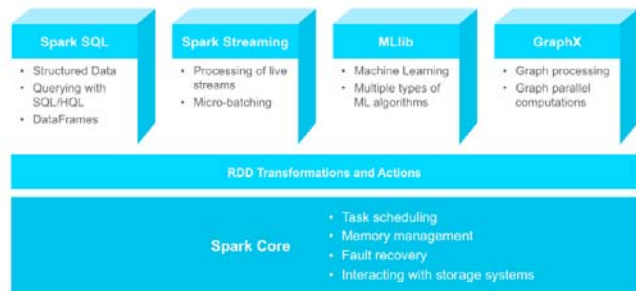
2

강의소개

강의 내용 (1)

빅 데이터 컴퓨팅의 기본 개념, 원리 및 응용 기법 강의

- 빅 데이터 컴퓨팅 소개
 - 아파치 하둡 소개
 - 맵리듀스 소개
- 아파치 스파크 코어
 - 스파크 프로그래밍 기초
 - RDD, Pair RDD, 데이터프레임
 - 응용 모니터링
- 스파크 통합 응용 프레임워크
 - 스파크 SQL
 - 스파크 스트리밍 (Spark Streaming)
 - 스파크 GraphX
 - 스파크 Mlib
- 아파치 HBase



강의 내용 (2)

빅 데이터 컴퓨팅의 응용 예

- 온라인 경매 조회 (Online Auction)
- 샌프란시스코 경찰 사건 조회 SFPD (San Francisco Police Department)
- 석유 시추 설비의 센서 모니터링
- 미국 교통국 항공기 운항 조회 및 분석
- 영화 추천



User Item Rating Matrix

| | A | B | C |
|-------|---|---|---|
| Ted | 4 | 5 | 5 |
| Carol | | 5 | 5 |
| Bob | | 5 | ? |

MapR Academy 강의 자료



- <http://learn.mapr.com>
- Introduction to Big Data
 - <http://learn.mapr.com/ess-100-introduction-to-big-data>
- Developing Hadoop Applications
 - <http://learn.mapr.com/dev-301-developing-hadoop-applications>
- Build and Monitor Apache Spark Applications
 - <http://learn.mapr.com/dev-361-build-and-monitor-apache-spark-applications>
- Create Data Pipelines Using Apache Spark
 - <http://learn.mapr.com/dev-362-create-data-pipelines-using-apache-spark>

강의 홈페이지의 참고 사이트

- 논문 및 블로그

강의방법

- 강의노트는 강의 홈페이지에 게시
 - <http://cs.sch.ac.kr/> => 강의 => 빅데이터 컴퓨팅 응용
 - <http://cs.sch.ac.kr/lecture/BigData/2017-Grad/17-BigDataComp.htm>
- 제출 과제 발표

평가

- 출석 10%, 과제 및 발표 60%, 시험 30%

□ 과제 종류

- 각 단원 학습 후 복습의 일환으로 **실습 과제**
 - **자신만의 데이터**를 사용하여 적용하고 실행
 - 팀 프로젝트와 관련하여 연관성이 있는 데이터
- 실습 과제 종합하여 **팀 프로젝트**로 제출 및 종합 발표



□ 과제 제출

- 과제는 PPT로 작성하여 강의 페이지 과제 제출 게시판에 업로드
 - http://cs.sch.ac.kr/xe/index.php?mid=big_grad17
 - 과제 제출 시 회원 가입 필수
 - 회원 가입 방법은 게시판 공지 참조
- 과제 내용 및 발표, 제출기한 등을 고려하여 평가