

빅데이터 이해 강의 소개

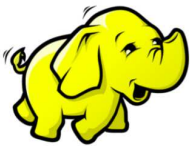
강의 목표 강의 내용 교재 및 강의 방법 과제

순천향대학교 컴퓨터공학과
이 상 정

순천향대학교 컴퓨터공학과

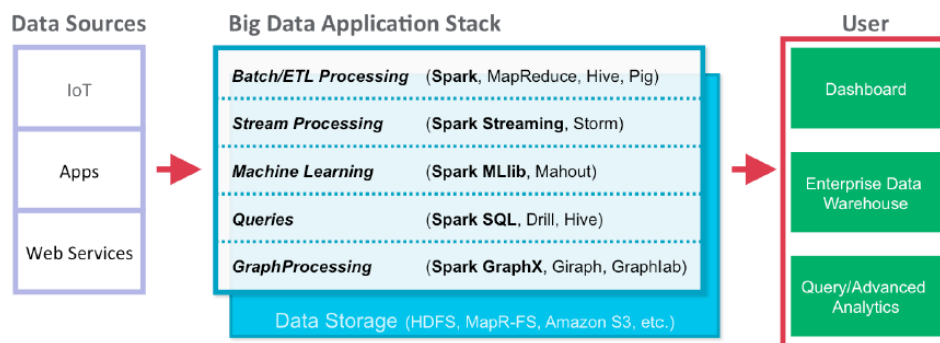
1

빅데이터 이해



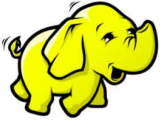
강의 목표

- 빅데이터 컴퓨팅의 기본 개념 및 원리 이해
- 리눅스 클러스터 서버 구축 및 활용
- 하둡 분산파일 시스템 및 맵리듀스 기본 개념 이해
- 파이썬 스파크 분산 데이터 처리 통합 엔진 활용
- 빅데이터 적용 사례를 통한 분산 데이터 처리 및 분석 기법



순천향대학교

강의소개



강의 내용 (1)

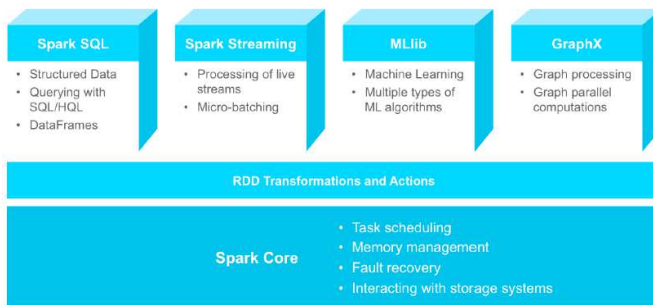
빅데이터 컴퓨팅 소개

- 빅데이터 컴퓨팅 개념
- 아파치 하둡 소개
- 맵리듀스 소개



스파크 분산 데이터 처리 통합 엔진

- 고급 질의, 그래프 처리, 스트림 처리, 머신러닝 등과 같은 고수준 분석도구를 위한 **통합 프레임워크**



순천향대학교 컴

강의소개

강의 내용 (2)

스파크를 사용한 빅데이터 분석 및 처리 사례

- SFPD 범죄 사건 데이터
- 단어 카운트
- 미국 주식 데이터
- 항공기 운항 데이터
- 석유 시추 데이터
- 영화 추천 데이터

User Item Rating Matrix

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



- [PySpark 공식 문서](#)
- [[Spark by {Examples}](#)] [PySpark Tutorial](#)
- [Learning Apache Spark with Python](#)
- [[databricks](#)] [Introduction to Apache Spark](#)
- [[MAPR Academy](#)] [Hadoop Developer Courses & Certification](#)
- [[MAPR Academy](#)] [Spark Developer Courses and Certification](#)

- 강의 홈페이지의 참고 사이트
 - 논문 및 블로그

강의 방법, 사이트 및 평가

- 강의 방법
 - VirtualBox 가상머신 클러스터 상에서 예제 실습
 - 하둡 분산 파일 시스템과 매퍼듀스의 빅데이터 처리
 - 스파크를 사용한 분산 데이터 처리 및 분석 기법 등을 강의
- 강의 사이트
 - <http://cs.sch.ac.kr/> => 빅데이터 이해
 - 강의 노트 및 과제 제출
 - 순천향 대학교 학습 플랫폼:(<https://eclass.sch.ac.kr/>)
 - > 빅데이터이해
 - > 강의자료실
- 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 **PPT**로 작성하여 **학습 플랫폼**에 업로드
순천향대학교 학습플랫폼 -> 운영체제 -> 주차 별 학습활동
- 업로드 파일 이름: **학과-학번-이름-과제이름.pptx**
- 과제 내용 및 발표 등을 고려하여 평가
 - **제출 기한이 지나면 학습 플랫폼 업로드 안됨**