



# 소프트웨어세미나 3

빅데이터

빅데이터 직무

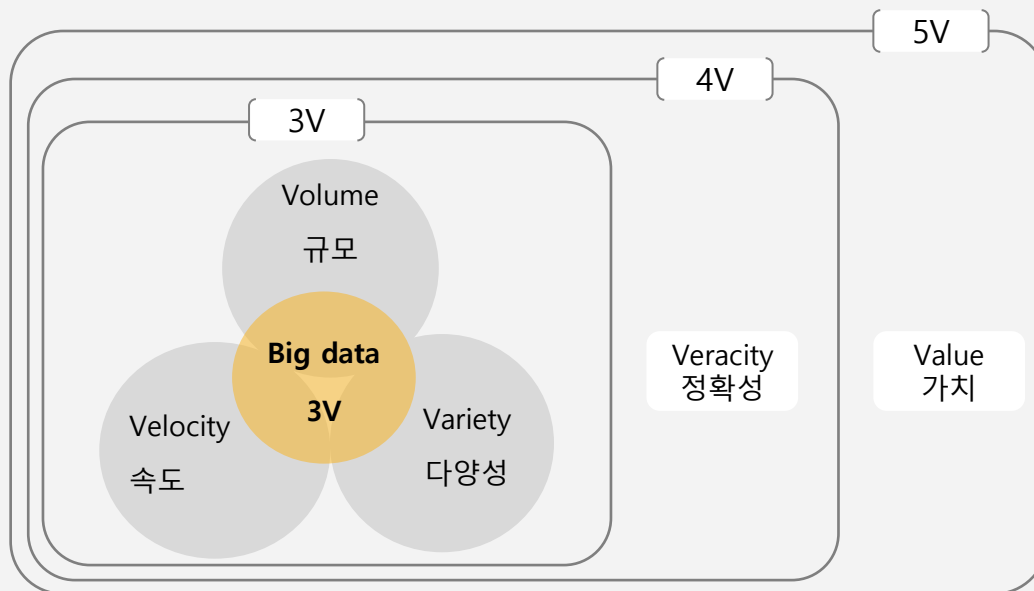
빅데이터 관련 오픈소스

# 빅데이터 개요

- 빅데이터

- 기존 데이터베이스 관리도구의 능력을 넘어서는 대량(수십 테라바이트)의 정형 또는 데이터베이스 형태가 아닌 비정형 데이터 집합을 포함한 데이터로부터 **가치를 추출**하고 **결과를 분석**하는 기술 –  
Wikipedia

- 특징



# 빅데이터 개요

- 빅데이터 분석 기술과 방법
  - 데이터 마이닝 - **데이터** 가운데 숨겨져 있는 유용한 상관관계를 발견
  - 기계 학습 - 컴퓨터에 학습할 수 있는 능력을 부여
  - 자연 언어 처리 - 인간의 언어를 컴퓨터에 인식시켜서 처리하는 일
  - 패턴 인식 - 사전에 정보를 기억시켜두고 그 정보와 대조하여 특징을 분석
  - 군집분석 - 관측 대상이 띠고 있는 다양한 특성의 유사성을 바탕
- 표현 기술
  - 시각적으로 표현하기 위한 대표 기술 - R 프로그래밍 언어

# 빅데이터 개요

- 빅데이터 직무 분류
  - 빅데이터 분석
  - 빅데이터 기획
  - 빅데이터 플랫폼 구축
  - 빅데이터운영·관리

# 빅데이터 직무 분류

- 빅데이터 분석

- 분류 : 정보통신 - 정보기술 - 정보기술전략
- 대용량의 데이터로부터 유용한 정보를 찾고 결과를 예측하기 위해 분석 기술과 방법론을 통해서 **정형/비정형 대용량 데이터를 구축, 탐색, 분석하고 시각화**를 수행

## 1.통계 기반 데이터분석

- 수집된 데이터(내·외부 데이터 및 정형·비정형 데이터)를 활용하여 분석 목적에 따라 **가설을 설정**하고 필요한 **데이터셋을 편성**하여 통계 기반 데이터 분석 모델을 만들고 평가

# 빅데이터 직무 분류

- 빅데이터 분석

- 2. 머신 러닝 기반 데이터 분석

- 복잡한 데이터 구조 패턴을 기계(컴퓨터)로 하여금 스스로 학습하게 하는 **머신러닝 알고리즘 기술을 활용**하여 데이터를 분석

- 3. 텍스트 데이터 분석

- 텍스트 데이터로부터 고품질의 정보를 도출하기 위해 텍스트 데이터를 변환 및 정제하여, 추출된 **단어 관계 및 패턴, 규칙을 분석**

# 빅데이터 직무 분류

- 빅데이터 분석

- 4. 빅데이터 분석 결과 시각화

- 사용자가 분석 결과를 이해하기 쉽게 **그래픽 의미를 이용하여 시각적으로 표현**

- 5. 분석용 데이터 구축

- 수집 저장된 데이터를 **분석용 데이터로 정제, 변환, 적재, 검증**

- 6. 탐색적 데이터 분석

- 유의미한 데이터간의 관계를 찾고 검증하기 위하여 **데이터의 기본 통계와 데이터 분포를 분석하고 데이터 변수간 관계를 확인**

# 빅데이터 직무 분류

- 빅데이터 기획

- 분류 : 정보통신 - 정보기술 - 정보기술전략
- 가치 창출을 위하여 빅데이터 서비스, 환경분석, 분석 모델 및 플랫폼을 정의하고 이에 대한 활용과 성과 및 운영관리를 수립

1. 빅데이터 서비스 기획

- 빅데이터 분석 목적과 목표를 정의하고 목표를 달성하기 위한 비즈니스 서비스 모델을 도출하며, 이를 실행하기 위한 빅데이터 추진 전략 및 정책을 수립

2. 빅데이터 환경분석

- 서비스 모델과 관련된 비즈니스 내·외부 환경을 분석하고 데이터, 기술, 인력 관점에서 필요한 자원을 분석



# 빅데이터 직무 분류

- 빅데이터 기획

- 3. 빅데이터 분석 기획

- 데이터 분석 결과를 활용하기 위하여 빅데이터 분석 요건 정의, 데이터 확보, 데이터 탐색, 분석 모델링, 분석 결과 적용을 기획

- 4. 빅데이터 기술 플랫폼 기획

- 빅데이터 분석과 활용을 위한 빅데이터 아키텍처, 수집 기술, 저장 기술, 처리 기술에 필요한 하드웨어와 소프트웨어를 기획

- 5. 빅데이터 성과 관리 기획

- 조직의 목표달성을 위하여 추진된 결과를 평가하기 위하여 성과 기준, 성과측정 방법, 성과평가 실행, 피드백 등을 기획

# 빅데이터 직무 분류

- 빅데이터 기획

- 6. 빅데이터 활용 기획

- 빅데이터 분석으로 생산된 데이터를 조직의 목적에 맞는 분석결과 활용, 비즈니스 모델 활용, 시장 활용, 상품화를 기획

- 7. 빅데이터 운영 기획

- 빅데이터 운영의 전반적인 활동을 관리·통제하기 위한 품질관리, 보안관리, 조직 수립, 운영관리에 대한 계획을 수립

# 빅데이터 직무 분류

- 빅데이터 플랫폼 구축

- 분류 : 정보통신 – 정보기술 – 정보기술개발
- 다양한 형식의 대용량 데이터를 수집, 저장, 처리, 분석, 활용하기 위한 **인프라, 인터페이스 및 솔루션을 개발하고 적용**

1. 빅데이터 플랫폼 요구사항 분석

- 플랫폼 구축의 목적을 달성하기 위해 요구사항을 수집하고 수집된 자료로 요구사항을 도출 및 분석하여 빅데이터 플랫폼 범위를 명세하고 검증

2. 빅데이터 플랫폼 아키텍처 설계

- 빅데이터 수집, 저장, 처리, 분석 및 품질관리를 할 수 있도록 빅데이터 플랫폼 구축에 필요한 하드웨어와 소프트웨어의 구조를 설계

# 빅데이터 직무 분류

- 빅데이터 플랫폼 구축

- 3. 빅데이터 수집시스템 개발

- 활용 목적에 적합한 데이터를 수집하기 위하여 **빅데이터 수집시스템 구성**
    - 내·외부 데이터 수집모듈 개발, 데이터 변환모듈 개발 및 수집 데이터 검증모듈을 개발

- 4. 빅데이터 저장시스템 개발

- 수집 변환된 데이터를 조직의 활용 목적에 적합하도록 데이터 유형과 분석 목적을 고려하여 저장구조를 설계하고 고가용성을 제공하는 저장시스템을 개발

- 5. 빅데이터 처리시스템 개발

- 저장된 데이터를 처리목적에 따라 크기, 종류, 저장 구조를 고려하여 처리 및 가공하기 위한 분산처리, 실시간 처리, 이벤트 처리 모듈을 개발

# 빅데이터 직무 분류

- 빅데이터 플랫폼 구축

- 6. 빅데이터 분석시스템 개발

- 빅데이터에 숨겨진 패턴과 알려지지 않은 정보간의 관계를 찾아내기 위해 사용되는 통계기법, 머신러닝, 텍스트마이닝 등의 분석방법들을 효과적으로 적용하기 위한 분석시스템을 개발

- 7. 빅데이터 품질관리시스템 개발

- 빅데이터 품질목표를 달성할 수 있도록 빅데이터 품질관리시스템 설계 및 구성, 품질관리 모듈을 개발

- 8. 빅데이터 플랫폼 테스트

- 구축된 플랫폼이 정확하고 완전하게 개발되었는지 확인하기 위하여 테스트 시나리오를 작성하고 플랫폼에 대한 기능 및 비기능 요소를 검증

# 빅데이터 직무 분류

- 빅데이터운영·관리

- 분류 : 정보통신 - 정보기술 - 정보기술운영
- 구축된 빅데이터 플랫폼을 안정적으로 운영하기 위하여 운영정책을 수립하고 수립한 정책에 따라 빅데이터 서비스 및 솔루션, 데이터를 유지 관리

1. 빅데이터 플랫폼 운영 정책 수립

- 빅데이터 서비스 운영 지침을 수립하며 성과와 비용을 관리하고 품질을 개선

2. 빅데이터 서비스 운영 계획

- 빅데이터 서비스에 대한 사용자 만족과 서비스 품질 개선을 위해 수준의 정의, 목표설정, 평가측정, 지속적인 개선활동을 기획, 관리

# 빅데이터 직무 분류

- 빅데이터운영·관리

- 3. 빅데이터 서비스 운영 관리

- 빅데이터 서비스 운영 지침에 따라 서비스 성능, 가용성, 연속성 및 구성 변경을 관리

- 4. 빅데이터 솔루션 운영 계획

- 빅데이터 서비스 제공을 위한 빅데이터솔루션의 운영 현황을 분석하여 목표를 설정하고 이를 수행하기 위한 운영 자원을 확보

- 5. 빅데이터 솔루션 운영 관리

- 빅데이터 서비스 운영 요구사항을 충족할수 있도록 빅데이터 솔루션에 대한 변경 관리와 환경 설정을 관리

# 빅데이터 직무 분류

- 빅데이터 운영·관리

## 6. 빅데이터 플랫폼 모니터링

- 원활한 빅데이터 플랫폼 운영을 위해 빅데이터 플랫폼 점검과 성능 모니터링을 수행한 후, 빅데이터 플랫폼의 상태와 문제점들을 분석하여 효과적인 운영을 위해 장애유형별 대응방안을 수립하고 관리

## 7. 빅데이터 품질 관리

- 데이터 운영관리 시 기대수준의 품질을 확보하기위해 빅데이터 품질 기준을 수립하고 빅데이터 품질을 진단하여 빅데이터 품질을 개선

## 8. 빅데이터 플로우 관리

- 정의된 데이터 플로우를 점검하고, 적합한 방식으로 데이터 플로우를 변경하여 유효한 데이터가 처리되도록 관리



# 빅데이터 직무 분류

- 빅데이터 운영·관리

## 9. 빅데이터 모델 운영

- 타겟 값을 획득할 수 있는 데이터를 파악하기 위하여 메타 데이터를 설계하고 타겟 데이터의 모델을 관리하는 능력

## 10. 빅데이터 처리 운영

- 운영 중인 빅데이터 처리 솔루션과 프로그램을 이해관계자의 변경 요구 사항에 부합할 수 있도록 변경

# 빅데이터 오픈소스 SW

- 빅데이터 오픈소스 SW
  - OpenStack Swift
    - 분산 구조의 Object 데이터의 저장 스토리지 체계
  - 쿠버네티스(Kubernetes)
    - 구글에 의해 설계되었으며, 컨테이너(Container)를 쉽고 빠르게 배포-확장하고 관리를 자동화해주는 오픈소스 소프트웨어
  - HDFS ( Hadoop Distributed File System)
    - 대용량 데이터를 분산 처리할 수 있는 자바(Java) 기반의 오픈소스 프레임워크
  - KFS ( Kosmos File System )
    - Kosmix에서 만든 Google File System의 오픈소스 구현체
  - 아파치 스파크( Apache Spark )
    - 오픈 소스 클러스터 컴퓨팅 프레임워크
  - MongoDB
    - 오픈소스 DBMS
  - R 언어
    - 빅데이터 통계 분석 시각화 언어

# 빅데이터 오픈소스 SW

LOTTE  
VALLEY에 산다

일-리크루터비  
RecruiTV

**Q** 데이터 분석팀은 어떤 팀인가요?

 **T h a n k      y o u**