

Chapter 1. 빅데이터의 이해

빅데이터 개요 및 활용

(1) 빅데이터 특징

빅데이터 개념 : 막대한 양(수십 테라바이트 이상)의 ____ 및 ____ 데이터

DIKW 피라미드 -> 데이터에서 가치 추출

D____	_____	같은 책인데 A 사이트에 3만원, B 사이트에서 4만원
I_____	_____	A 사이트에서 더 저렴하다
K_____	_____	서적은 더 저렴한 A 사이트 책을 구매하자
W_____	_____	아마 다른 책도 A 사이트가 더 저렴하겠지?

데이터의 양을 측정하는 단위 PEZY

빅데이터 특성 4V는 외우자!

V____(____), V____(____), V____(____), V____(____)

(2) 빅데이터 가치

빅데이터의 가치 : ____적 ____, ____ 제거, ____ 감소, 스마트한 경쟁력, 타 분야 ____

빅데이터 가치 산정이 어려운 이유 :

데이터 ____ ____의 ____화, ____ ____ 창출, ____ ____의 급속한 ____

빅데이터의 영향 기업 : ____ 수단 제공, ____력 강화, ____성 향상

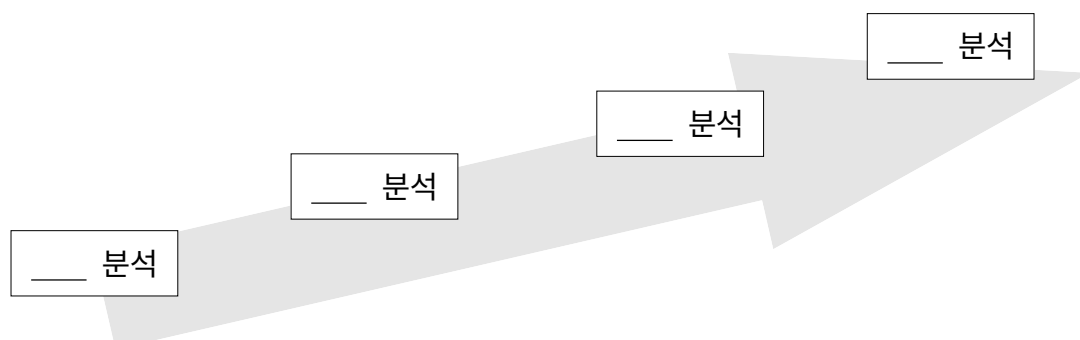
정부 : 환경 탐색, 상황 분석, 미래 대응 가능

개인 : 목적에 따른 활용

빅데이터 위기 요인과 통제 방안

위기 요인	통제 방안
____ ____	____의 ____
____ 원칙 훼손	____ 기반의 책임 적용
데이터 ____	____에 대한 ____ 허용

분석 가치 에스컬레이터



(3) 빅데이터 산업의 이해

_____ 기술의 발전으로 데이터 처리 비용이 급격하게 감소하여 빅데이터가 발전하고 있다.

(4) 빅데이터 조직 및 인력

빅데이터 업무 프로세스 : 도입 → 구축 → 운영

조직 구조 설계 요소 : _____, _____화, _____

조직 구조 유형

_____ 구조	_____ 구조	_____ 구조
전사 분석 업무를 분석 전담 조직에서 담당	해당 부서에서 분석 수행	분석조직 인력들을 현업 부서로 직접 배치

조직 역량 모델링 : 우수 성과자의 행동하는 특성을 파악하여 직무별 역량 모델을 만든다

데이터 사이언티스트 요구역량

_____ 스킬 (모든 직무)	협력 능력	커뮤니케이션 능력
	분석의 통찰력	논리적 _____ 능력, _____적 사고력, 호기심
	설득력있는 전달력	_____ 능력, 비주얼라이제이션
_____ 스킬 (빅데이터 업무)	분석기술의 숙련도	목적에 맞는 최적 분석 설계, 노하우 축적
	이론적 지식	빅데이터 관련 기법 및 다양한 방법론 습득

데이터 분야 직무별 업무

데이터 엔지니어	_____ 세트 _____ _____ 도구 개발, 데이터 _____ 처리 시스템 개발 프로그래밍 언어 사용 스킬 필수
데이터 분석가	비즈니스 _____ 제공 데이터 _____ 및 보고서 작성 전략 수립, 데이터 수집 및 분석
데이터 사이언티스트	정형, 비정형 데이터에서 인사이트 창출 _____ 모델링, _____ 모델링, _____ 시스템 개발
데이터 아키텍트	데이터 _____를 위한 청사진 데이터 _____ 프로세스 정립, 모니터링, 테스트

데이터 거버넌스 구성요소 : _____, _____, _____

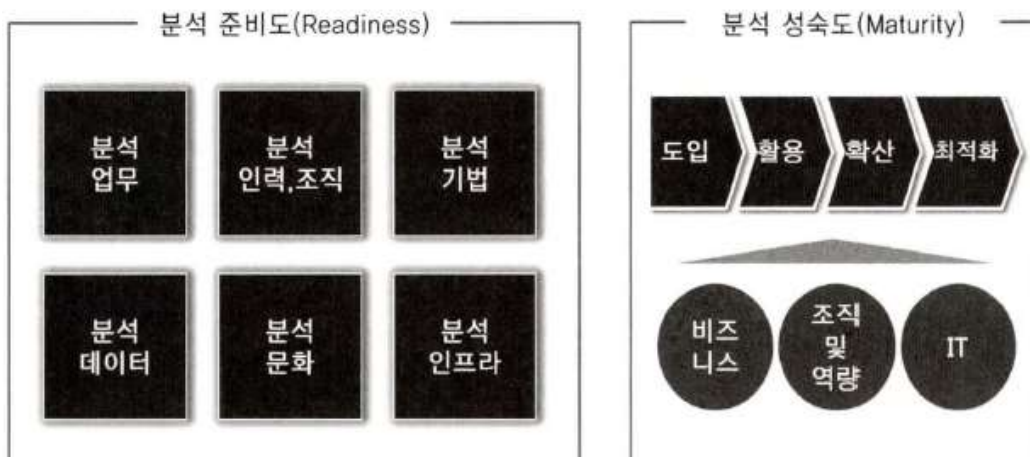
데이터 거버넌스 체계

데이터 표준화	데이터 표준 __ __, 명명 __, 메타데이터 __ 데이터 __ __ (data dictionary in advance 아님) 데이터 표준 준수 진단, 논리·물리 모델 표준에 맞는지 검증
표준화 활동	__ __ 여부를 주기적으로 점검 및 모니터링
데이터 관리 체계	메타데이터와 데이터 사전의 __ __ __
데이터 저장소 관리	메타데이터 및 표준 데이터를 관리하기 위한 __ __

조직 평가를 위한 성숙도(Maturity) 단계

단계 1)	__ 단계	__ 단계	__ 단계	__ 단계
설명	분석을 시작하여 __ 과 __ 구축	분석 결과를 실제 업무에 적용	전사 차원에서 분석을 관리하고 __	분석을 진화시켜 혁신 및 성과 향상에 기여
비즈니스	<ul style="list-style-type: none"> 실적분석 및 __ 정기보고 수행 운영 데이터 기반 	<ul style="list-style-type: none"> 미래 결과 예측 시뮬레이션 운영 데이터 기반 	<ul style="list-style-type: none"> 전사 성과 실시간 분석 프로세스 혁신 3.0 분석규칙 관리 이벤트 관리 	<ul style="list-style-type: none"> 외부 환경분석 활용 최적화 업무 적용 실시간 분석 비즈니스 모델 진화
조직 · 역량	<ul style="list-style-type: none"> __ 부서에서 수행 __ 역량에 __ 	<ul style="list-style-type: none"> 전문 담당부서에서 수행 분석기법 도입 관리자가 분석 수행 	<ul style="list-style-type: none"> 전사 모든 부서 수행 분석 COE 조직 운영 데이터 사이언티스트 __ 	<ul style="list-style-type: none"> 데이터 사이언스 그룹 경영진 분석 활용 전략 연계
IT	<ul style="list-style-type: none"> 데이터 웨어하우스 데이터 마트 ETL/ EAI OLAP 	<ul style="list-style-type: none"> 실시간 대시보드 통계 분석 환경 	<ul style="list-style-type: none"> 빅데이터 관리 환경 시뮬레이션·최적화 비주얼 분석 분석 전용 서버 	<ul style="list-style-type: none"> 분석 협업환경 분석 Sandbox 프로세스 내재화 빅데이터 분석

Wrap up



빅데이터 기술 및 제도

(1) 빅데이터 플랫폼

빅데이터 플랫폼 구성 요소 : 수집, 저장, 분석, 활용

하둡 에코시스템 기술은 서브 프로젝트 기술과 기능만 매치시킬 수 있으면 된다

비정형 데이터 수집			정형 데이터 수집	
_____	_____	_____	_____	_____

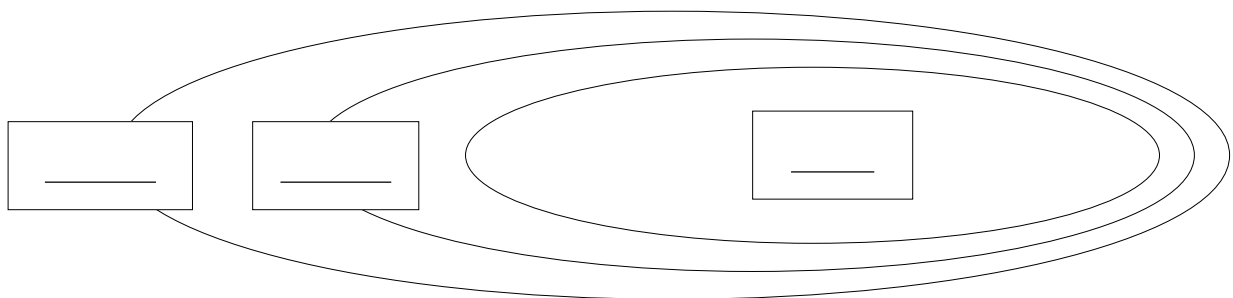
분산 데이터 저장	분산 데이터베이스	분산 데이터 처리
_____ (Hadoop distributed File System)	_____	_____

리소스 관리	인메모리 처리	데이터 가공		데이터 마이닝
_____	_____	_____	_____	_____

실시간 SQL 질의		워크플로우 관리	분산 코디네이션
_____	_____	_____	_____

(2) 빅데이터와 인공지능

인공지능의 개념



(3) 개인정보보호법·제도

빅데이터 개인정보보호 가이드라인

개인정보 _____화	개인정보가 포함된 정보는 비식별화 조치 후 이용 가능
재식별 시 조치	재식별 시 파기 및 비식별화 조치
_____성 확보	빅데이터 _____, _____ 등의 공개를 통한 투명성 확보

개인정보의 수집·이용이 가능한 경우 : ____, ____, ____, ____, 제_자의 ____,
 한 ____, 개인정보__의 __한 __

개인정보의 수집·이용을 위해 동의를 받을 때 고지할 사항 : 목적, 항목, 기간, 거부 불이익

데이터 3법

____ 법	__명 정보 개념 도입 동의 없이 처리할 수 있는 개인정보 합리화
____ 법	온라인상 관련 규제와 감독 주체를 '개인정보보호위원회'로 변경
____ 법	금융 분야 빅데이터 분석·이용의 법적 근거 명확화 개인정보보호위원회 기능 강화 금융 분야 ____ 산업 도입

정보 비교

개인정보	살아 있는 개인에 관한 정보	
__명 정보	추가 정보의 사용 없이는 특정 개인을 알아볼 수 없게 조치한 정보	통계작성, 연구, 공익적 기록보전 목적으로 동의 없이 활용 가능
__명 정보	더 이상 개인을 알아볼 수 없게 조치한 정보	개인정보가 아니기 때문에 제한 없이 자유롭게 활용

가명처리 절차 : 사전 준비 → 가명처리 → 적정성 검토 및 추가처리 → 사후관리

프라이버시 보호 모델

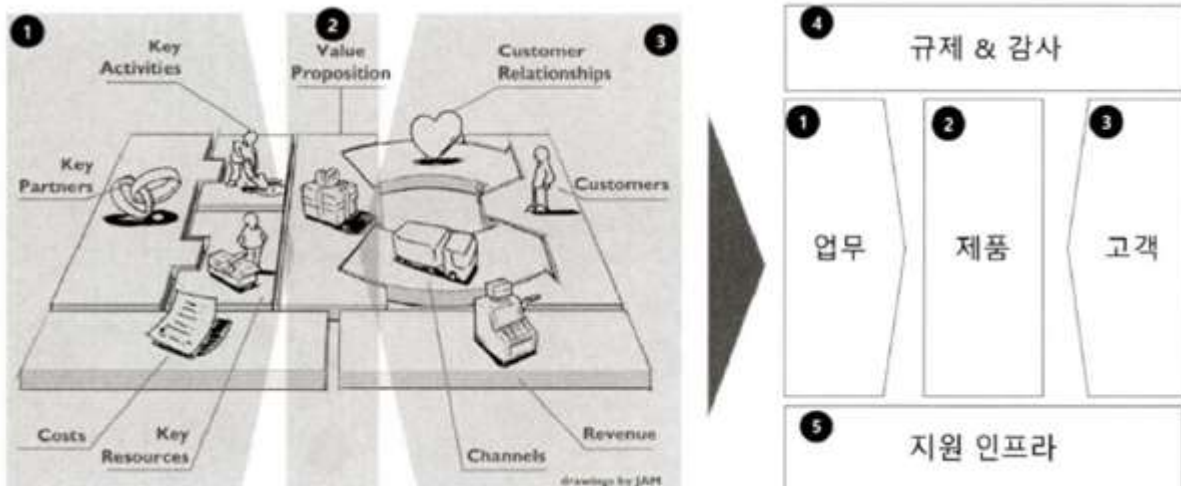
k-__성	같은 값이 적어도 k개 이상 존재하도록	
l-__성	l개 이상 다양한 정보를 갖도록	
t-__성	정보의 분포가 t 이하의 차이를 보이도록	
m-__성	원본 데이터와 동일한 속성 값의 조합이 최소 m개 이상 존재하도록	

Chapter 2. 데이터 분석 계획

(1) 분석 로드맵 설정

(2) 분석 문제 정의

하향식 접근 방식 - _____



Key activities에서 과제 발굴 (무엇을 분석해야 할까?)

하향식 분석 과제 발굴 절차

: 문제 _____ → 문제 _____ → _____ 탐색 → _____성 검토 → _____

상향식 접근 방식 - _____

상향식 접근 방식 절차

: 프로세스 _____ → 프로세스 _____ → 분석 _____ → 분석 _____

분석 기획 유형

_____ / _____	안다	모른다
안다	_____ (Optimization)	_____ (Insight)
모른다	_____ (Solution)	_____ (Discovery)

1)

단계	도입단계	활용단계	확산단계	최적화단계
설명	분석을 시작하여 환경과 시스템을 구축	분석 결과를 실제 업무에 적용	전사 차원에서 분석을 관리하고 공유	분석을 진화시켜서 혁신 및 성과 향상에 기여
비즈니스 부문	<ul style="list-style-type: none"> 실적분석 및 통계 정기보고 수행 운영 데이터 기반 	<ul style="list-style-type: none"> 미래 결과 예측 시뮬레이션 운영 데이터 기반 	<ul style="list-style-type: none"> 전사 성과 실시간 분석 프로세스 혁신 3.0 분석규칙 관리 이벤트 관리 	<ul style="list-style-type: none"> 외부 환경분석 활용 최적화 업무 적용 실시간 분석 비즈니스 모델 진화
조직·역량 부문	<ul style="list-style-type: none"> 일부 부서에서 수행 담당자 역량에 의존 	<ul style="list-style-type: none"> 전문 담당부서에서 수행 분석기법 도입 관리자가 분석 수행 	<ul style="list-style-type: none"> 전사 모든 부서 수행 분석 COE 조직 운영 데이터 사이언티스트 확보 	<ul style="list-style-type: none"> 데이터 사이언스 그룹 경영진 분석 활용 전략 연계
IT 부문	<ul style="list-style-type: none"> 데이터 웨어하우스 데이터 마트 ETL/ EAI OLAP 	<ul style="list-style-type: none"> 실시간 대시보드 통계 분석 환경 	<ul style="list-style-type: none"> 빅데이터 관리 환경 시뮬레이션·최적화 비주얼 분석 분석 전용 서버 	<ul style="list-style-type: none"> 분석 협업환경 분석 Sandbox 프로세스 내재화 빅데이터 분석