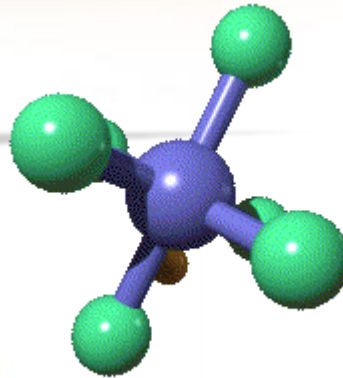


제1장 빅데이터 개요



Contents



1. 산업혁명과 빅데이터

- 21세기 빅데이터 시대
- 빅데이터 출현 배경
- 빅데이터와 산업혁명

2. 빅데이터 의의

- 빅데이터 개념
- 빅데이터 특성

3. 빅데이터 가치

- 빅데이터 가치 정의
- 빅데이터의 내재된 가치

1. 산업혁명과 빅데이터

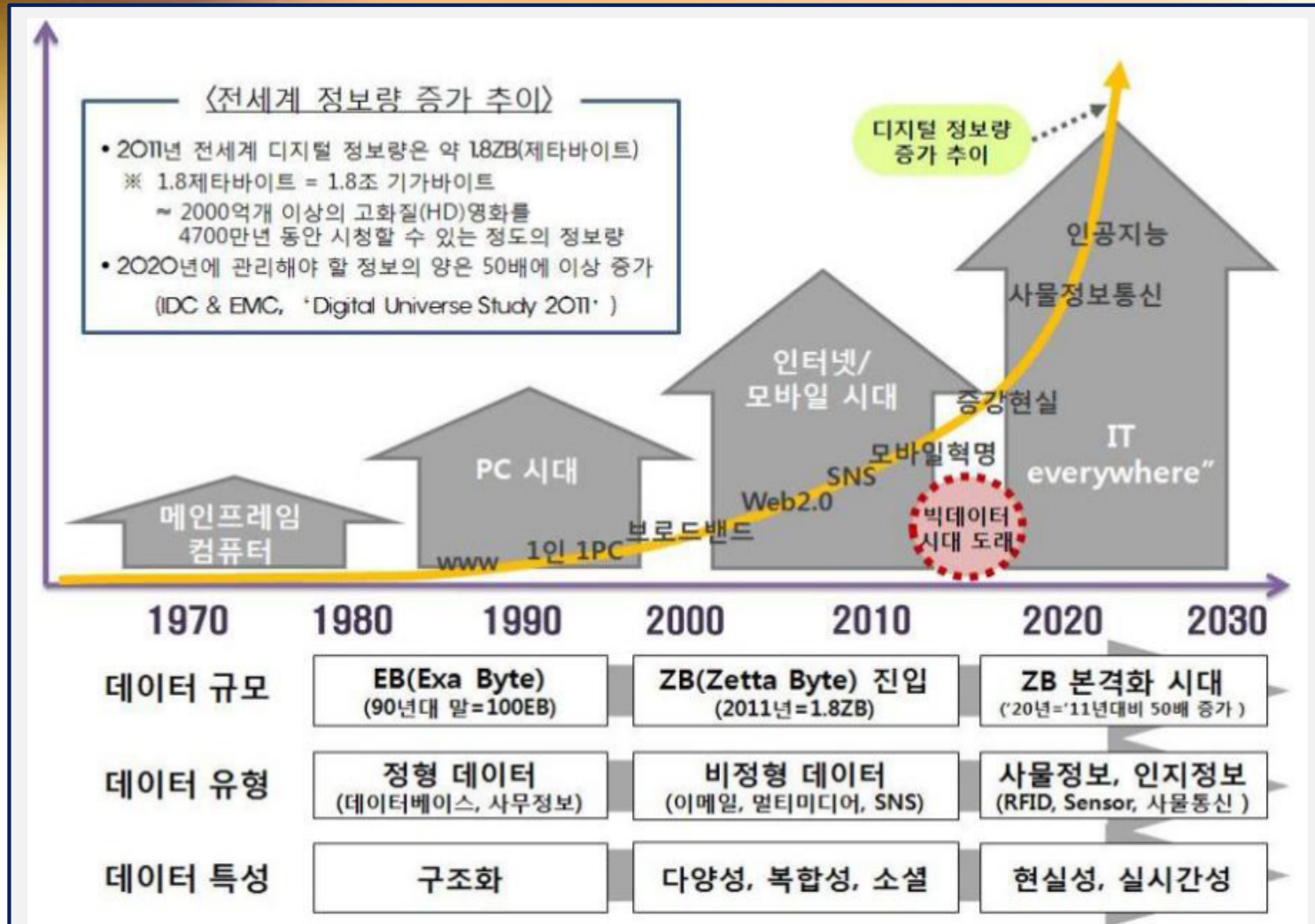
■ 21세기 빅데이터 시대

- 정보통신기술의 발달과 디지털 혁신 -> 데이터 증가
- 소셜 네트워크 서비스 -> 데이터 증가 가속화
- 향후 데이터의 폭발적 증가 -> 빅데이터 시대 도래
- 빅데이터 수집, 가공, 가치창출 -> 기업의 마케팅 전략 활용
- 빅데이터 활용(정보통신, 교육, 의료, 금융 등) 확산 -> 사회전반의 생산성 향상

1. 산업혁명과 빅데이터



1. 산업혁명과 빅데이터



1. 산업혁명과 빅데이터

- 빅데이터의 데이터 단위

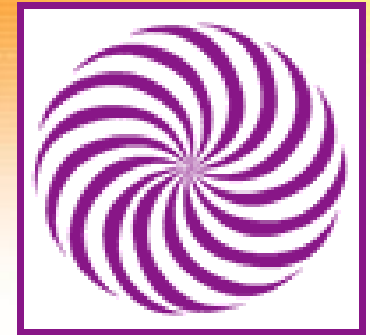
영문 단위	한글 단위	10승
YB(yotta byte)	1자(穰)	10^{24}
ZB(zetta byte)	10해(垓)	10^{21}
EB(exa byte)	100경(京)	10^{18}
PB(peta byte)	1,000조(兆)	10^{15}
TB(tera byte)	1조(兆)	10^{12}
GB(giga byte)	10억(億)	10^9
MB(mega byte)	100만(萬)	10^6
KB(kilo byte)	1,000	10^3

1. 산업혁명과 빅데이터

■ 빅데이터 출현 배경

- 세계적인 컨설팅 기관인 Mckinsey and Company는 'Big Data : The Next Frontier for Innovation, Competition, and Productivity'에서 빅데이터 출현배경을 아래와 같이 정리함.
- 기업의 고객 데이터 트래킹 및 수집행위의 증가.
- 저장매체와 카메라 모듈, 디스플레이 가격의 인하는 멀티미디어 콘텐츠 사용 확산과 이에 관한 정보의 증가를 가져옴.
- 트위터, 페이스북 등 SNS의 급격한 확산과 더불어 텍스트 등 비정형 데이터의 폭증.
- M2M(Machine to Machine) 및 사물인터넷(IoT; Internet of Things) 등의 통신 기술 발전에 따른 사물간 통신 네트워크에서 발생하는 데이터의 증가.
- 결론적으로 방대한 데이터의 축적, 실시간으로 생성되는 데이터 양의 지속적인 증가, 전 세계 데이터 양의 증가, 데이터 저장 장치의 발전과 데이터 처리 기술의 발전 등 복합적인 원인으로 인해 빅데이터 출현.

1. 산업혁명과 빅데이터



■ 빅데이터 출현 배경

- 빅데이터의 새로운 가능성과 대응 전략

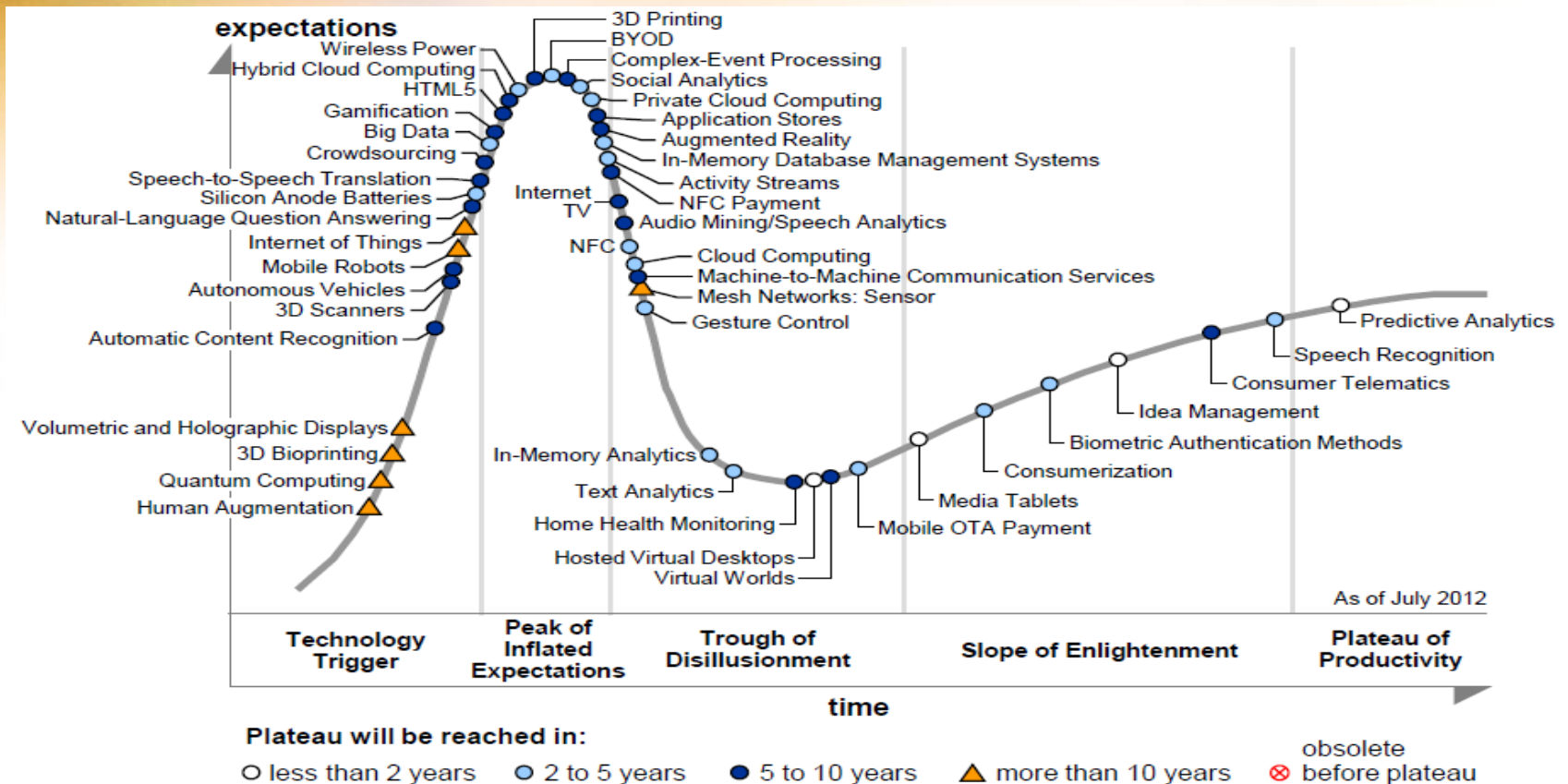
구분	PC시대	인터넷시대	모바일시대	스마트시대
패러다임 변화	디지털화, 전신화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물정보화
IT 이슈	PC, PC 통신, 데이터베이스	초고속 인터넷, www, 웹서버	모바일 인터넷, 스마트폰	빅데이터, 차세대PC, 사물네트워크
핵심분야(서비스)	PC, OS	포털, 검색엔진, Web2.0	스마트폰, 앱서비스, SNS	미래전망, 상황인식, 개인맞춤형서비스
대표기업	MS, IBM	구글, 네이버, 유튜브 등	애플, 페이스북, 트위터 등	?
IT 비전	1인 1PC	클릭 e-Korea	손안의 PC, 소통	lteverywhere, 가치창출

출처 : 한국정보화진흥원, 빅 데이터의 새로운 가능성과 대응 전략, 2011, 재인용

1. 산업혁명과 빅데이터

■ 빅데이터 출현 배경

- IT기술 측면에서 빅데이터가 중요한 기술로 발전할 것으로 예상

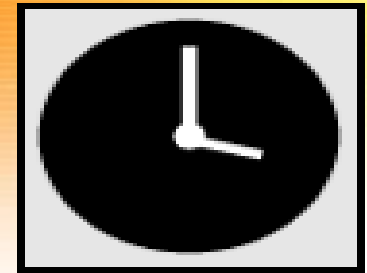


출처 : Gartner, 2012

1. 산업혁명과 빅데이터

■ 빅데이터와 산업혁명

- 산업혁명의 단계별 특징



1. 산업혁명과 빅데이터

■ 빅데이터와 산업혁명

- 4차 산업혁명의 용어는 2016년 다보스 포럼(World Economic Forum, WEF)에서 처음 등장.
- 포럼에서는 "산업혁명이 일어나는 주기는 점점 짧아지고 있으며, 이제는 4차 산업혁명의 시대" 라고 밝히며 '4차 산업혁명' 이 우리 사회의 새로운 지평을 열 것으로 내다봄.
- 4차 산업혁명에서는 모든 사물이 정보통신기술과 융합되고 정보는 빅데이터로 분석.
- 4차 산업혁명의 중심인 빅데이터는 인공지능으로 진화하고 있음.
- 4차 산업혁명은 데이터 혁명.

2. 빅데이터 의의

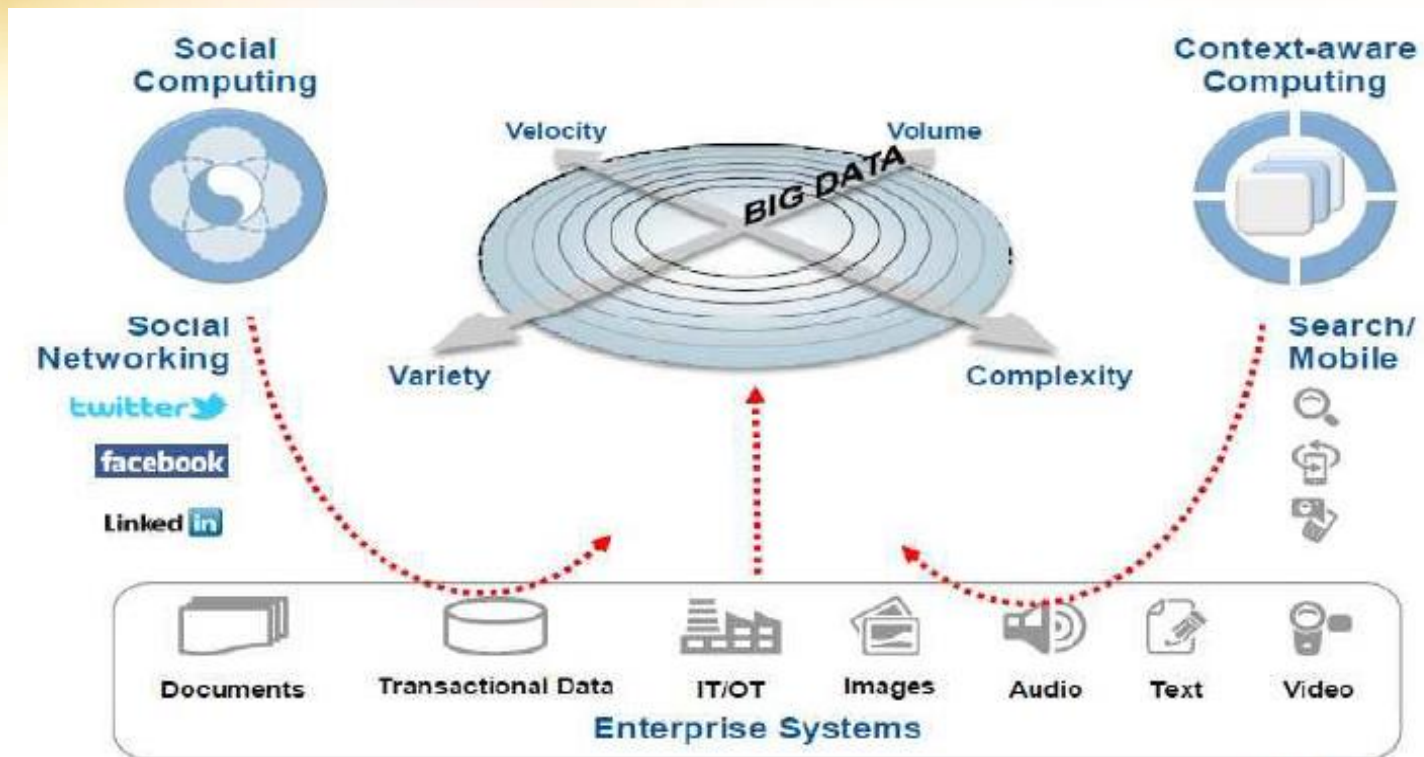
■ 빅데이터 개념

- 빅데이터라 함은 기존 데이터베이스의 데이터 수집, 저장, 관리, 분석의 역량을 넘어서는 대량의 정형 또는 비정형 데이터 및 이러한 데이터로부터 의사결정에 필요한 정보와 지식을 추출하고 결과를 분석하는 기술을 포괄적으로 뜻함.
- 맥킨지는 데이터의 규모에 초점을 맞추어 "기존 데이터 베이스 관리도구의 데이터 수집, 저장, 관리, 분석하는 역량을 넘어서는 데이터 " 로 정의함.
- IDC(internet data center)는 업무 수행 방식에 초점을 맞춰 "다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 빠른 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처 " 로 정의함.

2. 빅데이터 의의

■ 빅데이터 특성

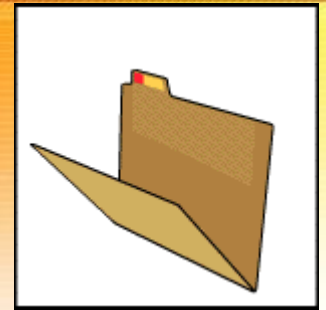
- 규모(Volume)의 증가, 다양성(Variety) 증가, 복잡성(Complexity) 증가, 속도(Velocity) 증가



출처 : Gartner(2011)

2. 빅데이터 의의

■ 빅데이터 특성



구분	특징
규모(Volume)	<ul style="list-style-type: none">• 정보기술의 발전과 IT의 생활화가 진행되면서 디지털 정보량이 기하급수적으로 폭증• 아날로그 자료의 디지털화에 따른 데이터 급증
다양성(Variety)	<ul style="list-style-type: none">• 로그기록, SNS, 위치정보, 소비정보, 현실데이터 등 기업정보 외 취합 데이터 종류의 증가• 텍스트 이외 멀티미디어(이미지, 음성, 영상 등) 등 비정형화된 데이터 유형의 다양화
복잡성(Complexity)	<ul style="list-style-type: none">• 구조화되지 않은 기업 외부데이터, 데이터 저장방식의 차이로 인한 정형화의 어려움• 데이터 종류의 확대, 외부데이터의 활용으로 인한 관리 대상의 증가• 데이터 관리 및 처리의 복잡성 심화로 새로운 분석 기법이 요구됨
속도(Velocity)	<ul style="list-style-type: none">• 사물정보(센서, 모니터링), 스트리밍 정보 등 실시간성 정보의 증가• 실시간성 정보의 증가로 데이터 생성, 이동(유동) 속도가 기존 데이터에 비해 증가• 대규모 데이터 처리 및 가치 있는 현재정보(실시간) 활용을 위해 데이터 처리 및 분석 속도가 중요

출처 : Gartner, 2011, 한국정보화진흥원, 2011, 재인용

2. 빅데이터 의의

■ 빅데이터 특성

- 빅데이터 3가지 특징 정리



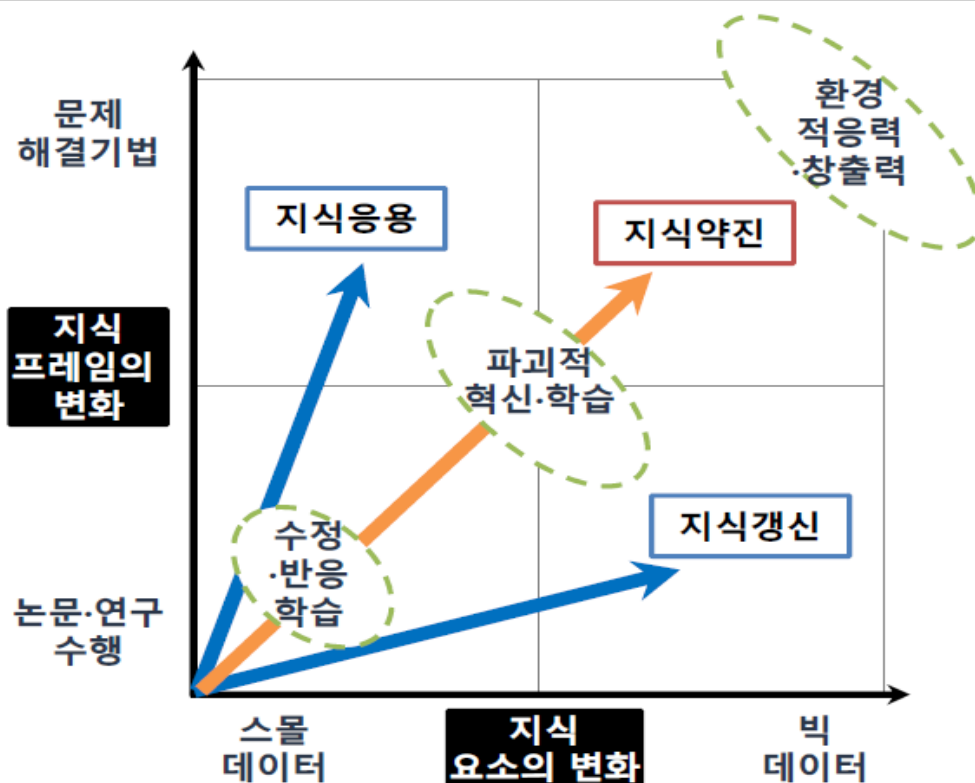
출처 : 함유근, 채승병, 빅데이터 경영을 바꾸다 , 삼성경제연구소, 2012

2. 빅데이터 의의

■ 빅데이터와 지식

빅데이터 등장에 따른 분석스킬의 변화와 분석에 대한 관점의 전환을 통해 실무문제해결에 대한 환경적응력 및 환경창출력이 강화될 수 있음

지식점프 매트릭스



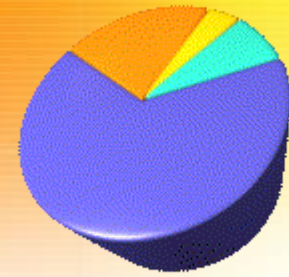
시사점

- 지식점프의 필요성
 - 환경적응력과 환경창출력을 강화
- 지식요소의 변화
 - 현행 사용자대상 서베이기반 스몰데이터를 활용한 추론통계에서 다양한 종류의 내·외부 데이터를 활용한 마이닝 적용
- 지식프레임의 변화
 - 단순한 통계분석을 이용한 연구·논문작성이라는 산출(output)차원에서 벗어나야 함
 - 실무현안들에 대한 해결대안이라는 결과(outcome), 영향(impact) 차원의 문제해결사 관점의 정립

통계전공자들이 만들어 놓은 프레임을 깨고 해당업무분야의 문제전문가라는 새로운 관점디자인 필요

2. 빅데이터 의의

■ 빅데이터 분석비교



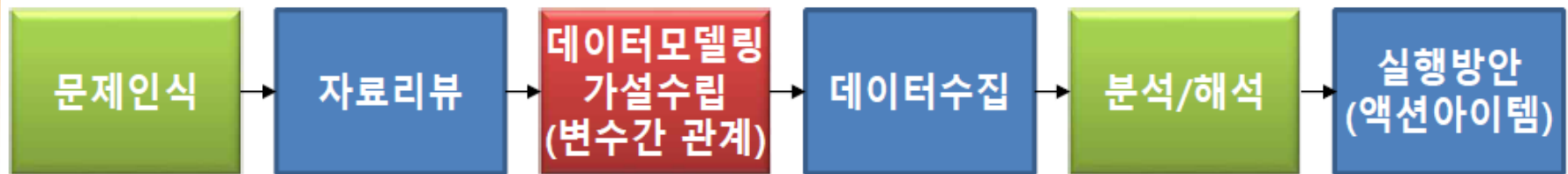
구분	스몰데이터 분석	빅데이터 분석
분석 목적	<ul style="list-style-type: none"> ■ 표본분석을 토대로 모집단의 특성을 추론하여 해석함 ■ 비교집단간 특성차이, 조치수단 간 효과차이 비교를 통해 관심대상 집단·수단을 선별하는데 중점 	<ul style="list-style-type: none"> ■ 대규모 데이터에 숨어있는 패턴을 발견하고 규칙을 도출함 ■ 도출된 패턴 및 규칙을 이용해 개별대상별 (고객·제품 등) 액션방안 마련에 중점
데이터 수집	<ul style="list-style-type: none"> ■ 사용자를 대상으로 한 서베이나 포커스그룹 인터뷰 활용으로 상당한 조사기간이 수반됨 	<ul style="list-style-type: none"> ■ 대규모 내부거래처리 데이터, 음성, 동영상, 외부공공 및 소셜데이터 활용으로 실시간에 가까운 분석
분석 기법	<ul style="list-style-type: none"> ■ 기술통계 및 차이분석, 회귀분석, 구조방정식 등 추론통계 중심 	<ul style="list-style-type: none"> ■ 군집분석, 연관분석, 분류분석, 예측분석 등 정형 및 텍스트 마이닝 기법중심
분석 도구	<ul style="list-style-type: none"> ■ 전통적 SPSS, SAS 등 상용 통계분석 패키지 활용 	<ul style="list-style-type: none"> ■ R 등의 오픈소스 중심의 분석도구 활용
적용 분야	<ul style="list-style-type: none"> ■ 제조, 금융, 유통, 관광, 보건, 행정, 국방 등 공공 및 민간 전분야를 망라함 	<ul style="list-style-type: none"> ■ 동일한 적용분야를 가지며, 보다 다양한 데이터소스 및 분석기법 활용으로 분석근거의 정확성과 예측력이 향상됨
분석 사용자	<ul style="list-style-type: none"> ■ 통계학 및 마이닝 전문가에게 의뢰를 통한 분석 	<ul style="list-style-type: none"> ■ 업무실무자 스스로 셀프분석을 통한 의사결정에 활용

2. 빅데이터 의의

■ 빅데이터 분석절차

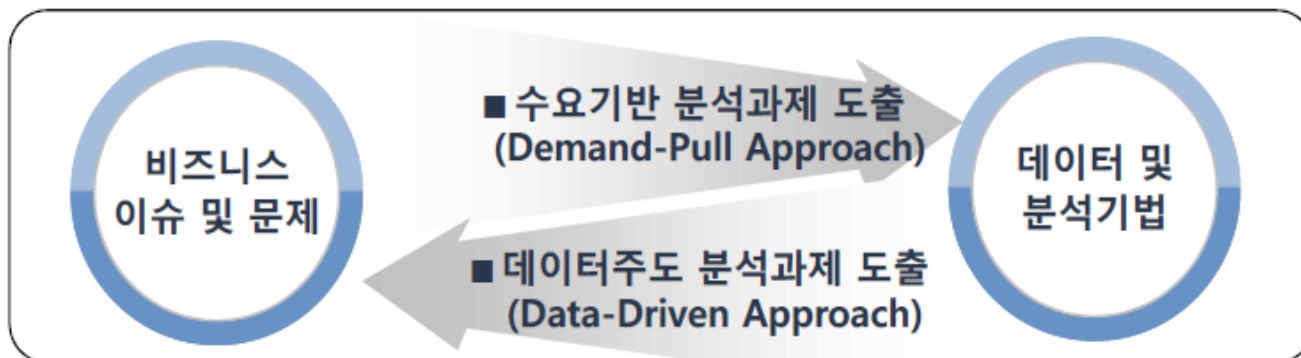
■ 데이터분석

- 기대상태와 현재상태를 조절하기 위하여 또는 이런 활동에 영향을 미치는 여러 가지 제약조건을 파악하고, 잠재적 원인을 진단하고, 가설적 해결방안을 수립
- 이와 관련된 데이터를 수집/가공/분석하는 활동



■ 분석과제기획

- 일련의 데이터분석 업무수행을 위한 계획을 수립하는 활동



2. 빅데이터 의의

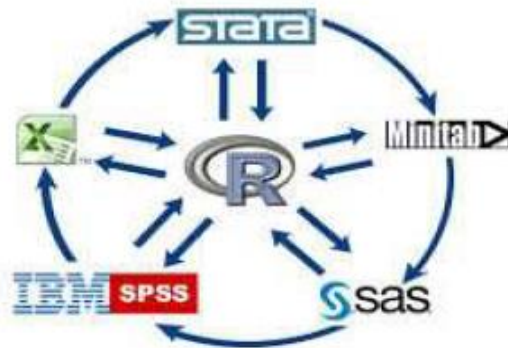
■ 빅데이터 접근방법

데이터분석 기법중심



- 가장 일반적인 접근방법
- 기초통계학, 확률분포, 데이터마이닝, 텍스트마이닝, 최적화 등 분석기법을 하나씩 학습함
- 통계학, 산업공학, 경영공학 등 전공특성에 따라 사용하는 분석기법 종류에 차이가 있음
- 전통적 통계학·마이닝 접근법

분석도구 중심



- R, SAS, SPSS, Matlab 등 분석도구의 기능코드나 메뉴를 하나씩 익혀나감
- 도구마다 특화된 학문분야 및 산업분야가 있음
- 추론통계 및 마이닝 등 기법별로 개별 소프트웨어를 익혀야 함
- 개발자 중심 접근법

문제해결 중심

보조활동	하부구조	하부지원: 기획, 재무, MIS, 법무, 총무				
	기술개발	연구, 설계, 개발, 디자인				
	인적자원	직무관리, 보상관리, 평가관리, 조직관리				
주요활동	기술획득	제품설계	제조	마케팅	유통유통	서비스
	·발전기술 ·기술경요 ·지적재산권 ·생산공정	·제품개발 ·특성 ·디자인 ·품질	·입지 ·공수 ·원재료조달 ·부품/조립	·상품·가격 ·광고, 홍보 ·판매/판매점 ·요청, 채널	·수송 및 저장 ·유통망 ·재고 ·통관	·보장제도 ·직접/특정 ·속도 ·가격

Source: Michael E. Porter, Competitive Advantage, 1985

- 내가 속해 있는 산업·업종, 기능부서 등의 고유한 이슈 중심
- 전통적인 중시되어오는 핵심 성과지표의 현황 및 목표를 통해 분석과제 접근이 가능
- 또는 경쟁사·벤치마킹을 통해 새로운 KPI도입가능
- 문제해결사 중심 접근법

3. 빅데이터 가치



■ 빅데이터 가치 정의

- 빅데이터 분석을 통해 가치 있는 정보를 실시간으로 도출해서, 트렌드 파악, 마케팅, 의사결정 등 다양한 분야에서 활용 증가.
- 빅데이터 시대에는 데이터를 수집, 저장, 처리, 분석하는 것뿐만 아니라 이로부터 새로운 가치를 창출하는 전 과정을 포괄하는 의미로 확장하고 있음.
- 빅데이터를 통해 얻은 혁신적인 통찰력(insight)은 기업의 브랜드 가치를 향상시킴.
- 빅데이터의 전략적 가치는 불확실성과 리스크가 존재하는 미래사회에는 통찰력, 대응력이 필요하며, 스마트화된 고객, 사회, 융복합이 이루어지고 있는 상황에서 미래사회는 경쟁력과 창조력을 필요로 함.

3. 빅데이터 가치

■ 빅데이터 가치 정의

- 미래사회에서의 빅데이터 역할

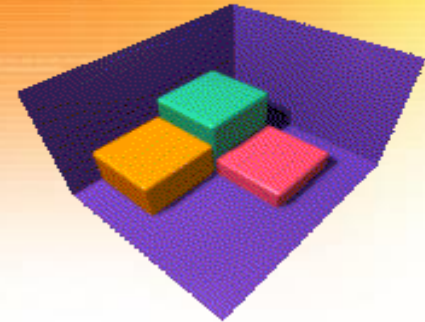
구분	빅데이터의 역할	
불확실성	통찰력	<ul style="list-style-type: none">• 사회현상, 현실세계의 데이터를 기반으로 한 패턴 분석과 미래전망• 여러 가지 가능성에 대한 시나리오 시뮬레이션• 다각적인 상황이 고려된 통찰력을 제시• 다수의 시나리오의 상황 변화에 유연하게 대처
리스크	대응력	<ul style="list-style-type: none">• 환경, 소셜, 모니터링 정보의 패턴 분석을 통한 위험징후, 이상 신호 포착• 이슈를 사전에 인지, 분석하고 빠른 의사결정과 실시간 대응 지원• 기업과 국가 경영의 명성 제고 및 낭비요소 절감
스마트	경쟁력	<ul style="list-style-type: none">• 대규모 데이터 분석을 통한 상황인지, 인공지능 서비스 가능• 개인화, 지능화 서비스 제공 확대• 소셜분석, 평가, 신용, 평판 분석을 통해 최적의 선택 지원• 트렌드 변화 분석을 통한 제품 경쟁력 확보
융합	창조력	<ul style="list-style-type: none">• 타 분야와의 융합을 통한 새로운 가치 창출• 인과관계, 상관관계가 컨버전스 분야의 데이터 분석으로 안정성 확보, 시행착오 최소화• 방대한 데이터 활용을 통한 새로운 융합시장 창출

출처 : 강만모 외, 빅데이터의 분석과 활용, 2012

3. 빅데이터 가치

■ 빅데이터의 내재된 가치

- 빅데이터 활용가치



< Big Data 활용가치 >



의료 건강(美)

- 매년 \$3,300억 가치
- 年 ~0.7%의 생산성 증가



공공, 행정 부문(유럽)

- 매년 €2,500억 가치
- 年 ~0.5%의 생산성 증가



개인 위치 정보(글로벌)

- 서비스 공급자 매출은 \$1,000억 이상
- 사용자 혜택은 \$7,000억



소매업(美)

- 이윤 60% 증가 가능
- 年 0.5~1% 생산성 증가



제조업

- 제품개발비 50% 감소
- 운전자본 7% 절감 가능

자료: McKinsey (2011.05)

3. 빅데이터 가치

기술과 인문학의 교차점



- 아이폰에 들어가는 각종 전자·센서부품의 기술적 작동원리를 난해한 공학이론, 수학기초 수준이 아닌 개념적으로 이해
- 사용자들의 생활양식, 업무환경 속에 ICT 기능특성이 어떻게 적용·활용될 수 있는지에 평생을 바침

과학상자 놀이



- 아이들이 과학상자를 가지고 놀 때 각 부속품들의 규격, 동작원리, 성분, 주의사항 등에는 관심이 없음
- 자동차, 비행기, 배, 장비 등 만들고 싶어하는 구체적인 대상(목표)을 정하고 각 부속품을 자유롭게 사용하는 데 중점을 둠

온도 데이터분석



- 온도를 측정하는 도구로 온도 변화시 액체물질의 열팽창성, 열기전성, 전기저항성, 방사열발생성을 측정하는 개념적 원리정도 이해하면 됨
- 업무에서 온도가 관련된 분석 대상·주제는 무엇인지, 온도의 변화에 영향을 미치는 선행변수와 결과변수에 대한 학습이 더 중요함

3. 빅데이터 가치

체온과 면역력



- 모든 질병과 노화는 체온저하와 밀접한 관련성이 있음
- 체온 1도를 올리면 면역력이 5배 정도 높아짐
- 36도 이상이면 효소활동이 활발해 면역력이 강화
- 낮으면 자율신경에 이상이 있고, 암세포 활동이 나타남

데이터센터 온도와 서버안전성



- 서버실은 22도라는 게 알려진 관행이었음
- 구글에서 온도와 서버의 성능, 수명과의 관계분석결과 27도에서 서버가 안정적임
- 온도 5도 증대로 냉각에너지를 반 이상 절감
- 폐북은 연평균기온이 낮은 스웨덴 북부에 IDC 예정

시설원예 온도와 작물생육



- 작물생육에 영향을 주는 요소 중의 하나는 온도임
- 작물의 종류에 따라, 계절에 따라, 성장단계에 따라 적정유지 온도범위가 존재함
- 저온장해시 탄수화물이나 원형질 합성에 어려움
- 고온장해시 기공폐쇄로 광합성 작용이 감소함

3. 빅데이터 가치

사람·내외부 고객



제품·서비스



시설·장비



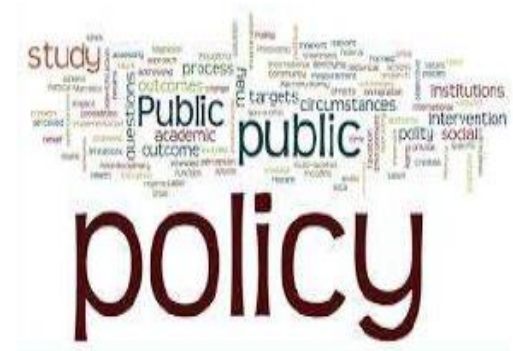
날씨·기상·자연



지역·장소·공간·위치



정책·법률·규제·제도



희망이란

본래 있다고도 할 수 없고
없다고도 할 수 없다.

그것은 마치 땅 위의 길과 같은 것이다.
본래 땅 위에는 길이 없었다.



한 사람이 먼저 가고
걸어가는 사람이 많아지면
그것이 곧 길이 되는 것이다.

-루쉰의
<고향>중에서-

감사합니다