

# 빅데이터 경영

## 빅데이터 개요

2016.3

권순범

Kookmin University

# 목 차

1. 빅데이터 개념
2. 빅데이터의 출현
3. 빅데이터의 경제적 가치
4. 빅데이터의 발전방향

# 빅데이터란?

- 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다 (<http://ko.wikipedia.org/>)

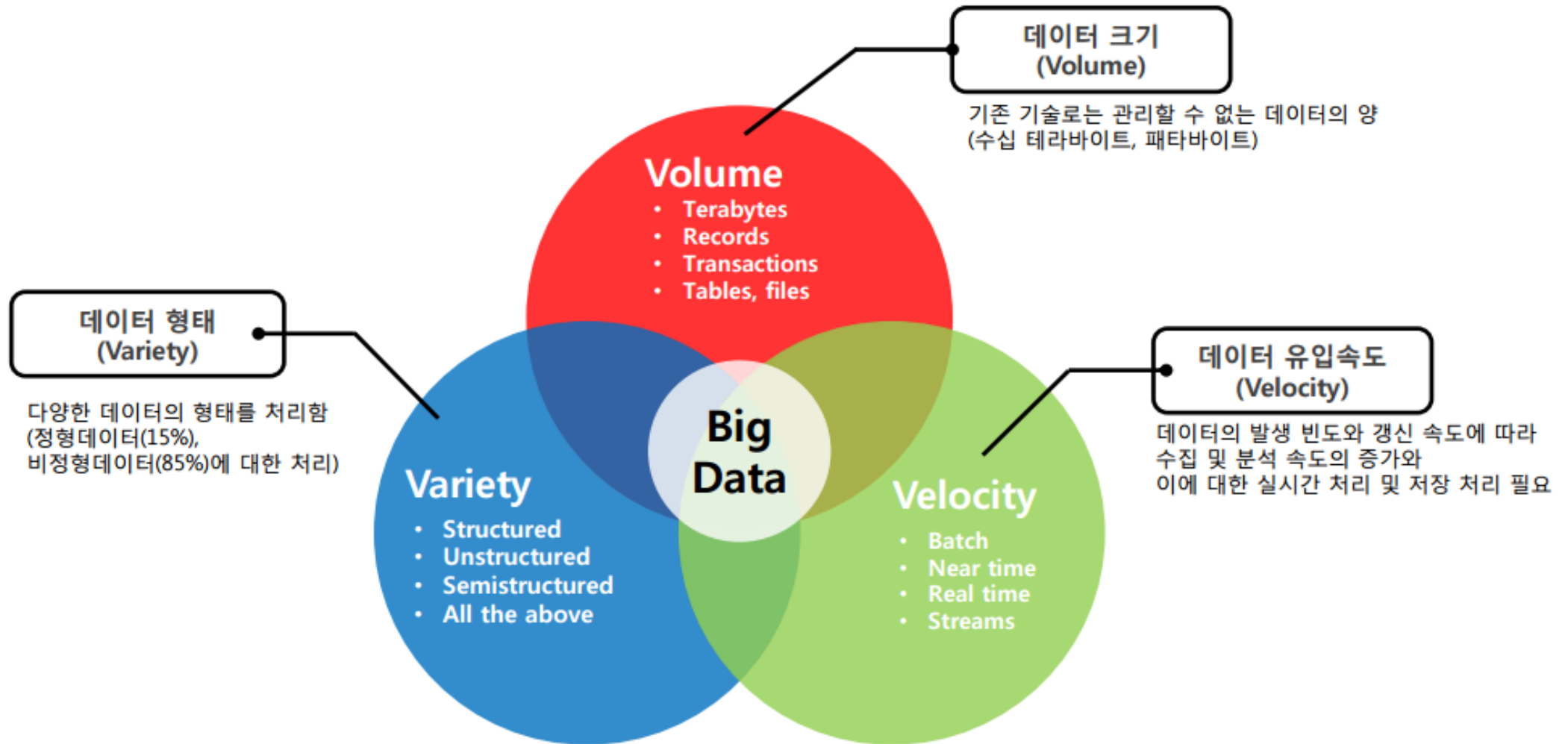


# 빅데이터란?

- 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다 (<http://ko.wikipedia.org/>)
- 데이터의 규모에 초점을 맞춘 정의  
기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석 역량을 넘어서는 데이터 (맥킨지 2011년 5월)
- 업무 수행 방식에 초점을 맞춘 정의  
다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고, 데이터의 빠른 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처 (IDC 2011년 6월)

# Big Data

- BI/DW 리서치 기관인 TDWI 가 빅데이터 3대 요소(3V)



**4V := Volume / Variety / Velocity +**

# 크기(Volume)

□ 일반적으로 수십 테라바이트(terabyte) 혹은 수십 페타바이트(petabyte) 이상이 빅데이터의 범위에 해당

- 1 Tera = 1024 GB, 1 Peta = 1024 TB
- 1 PB는 6 GB DVD 영화 17만 4천 편

**Byte, Kilo, Mega, Giga, Tera, Peta, Exa, Zetta**  
**1ZB(제타바이트) =  $10^{21}$  Byte = 1조 GB**

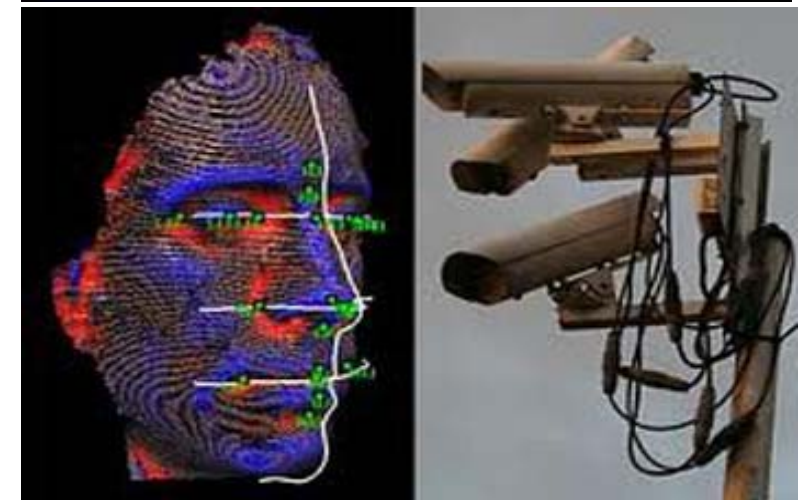
□ 기존 파일 시스템에 저장하기 어려울 뿐 아니라 데이터 분석을 위해 사용하는 기존 DW 같은 솔루션에서 처리하기 어려울 정도로 급격하게 데이터의 양이 증가하고 있음

□ 분산 컴퓨팅 방식

- 확장 가능한 방식으로 데이터를 저장하고 분석
- 분산 컴퓨팅 솔루션 - 구글의 GFS, 아파치의 하둡
- 대용량 병렬 처리 데이터베이스 - EMC의 GreenPlum, HP의 Vertica, IBM의 Netezza(신세계몰 2012년 구축), 테라데이터의 Kickfire 등

# 다양성(Variety)

휴대폰 제조업체 매출현황								
				2007년	2008년	2009년	2010년	당국핵심업
비즈니스	제조업체	종류	최소금액	2007년	2008년	2009년	2010년	당국핵심업
스카이	LG	미국	RM-7100	26,000	24,000	20,000	24,000	27.1%
호반미	WCO	일본	MS-150	16,000	16,000	20,000	24,000	19.6%
얼리큐	WCO	일본	SCP-401	12,000	16,000	15,000	18,900	16.3%
귀멸	GLS	한국	82	16,000	21,500	26,400	26,300	24.3%
세디안	GLS	한국	802101	21,000	30,000	30,000	41,000	31.0%
에드콜	SAMS	한국	E-171	25,000	45,000	56,000	56,400	40.0%
콜먼	SAMS	한국	E-2900	-	50,000	26,000	26,400	26.0%
스카이	LG	한국	RM-6100	-	24,000	26,000	24,000	21.0%
호반미	WCO	일본	SCP-402	12,000	16,000	15,000	18,900	16.3%
귀멸	GLS	한국	83	21,500	26,400	26,400	26,400	26.4%
세디안	GLS	한국	802101	21,000	30,000	30,000	30,000	29.2%
에드콜	SAMS	한국	E-171	45,000	45,200	52,100	55,000	36.3%
콜먼	SAMS	미국	E-3900	-	32,000	26,000	15,000	16.4%



<Books>

<Book ISBN="0553212419">

<title>Sherlock Holmes: Complete Novels...

<author>Sir Arthur Conan Doyle</author>

</Book>

<Book ISBN="0743273567">

<title>The Great Gatsby</title>

<author>F. Scott Fitzgerald</author>

</Book>

<Book ISBN="0684826976">

<title>Undaunted Courage</title>

<author>Stephen E. Ambrose</author>

</Book>

<Book ISBN="0743203178">

<title>Nothing Like It In the World</title>

<author>Stephen E. Ambrose</author>

</Book>

</Books>

Prof. SB Kwon

# 다양성(Variety)

- 데이터는 정형화 정도에 따라 정형(Structured), 반정형(Semi-Structured), 비정형(Unstructured)으로 구분
- 정형 데이터는 고정된 필드에 저장되는 데이터를 의미하며, 일정한 형식을 갖추고 있음
  - RDBMS의 테이블
- 반정형 데이터는 고정된 필드로 저장되지는 않지만, XML 이나 HTML 같이 메타데이터나 스키마(태그) 등을 포함
- 비정형 데이터는 고정된 필드에 저장되지 않는 데이터를 의미함.
  - 오디오(통화내용), 비디오(동영상), 텍스트(메신저로 주고 받은 대화 내용), 웹로그 데이터, 센서데이터, 생체데이터, 위치 정보 등
- 빅데이터는 90%를 차지하고 있는 비정형 데이터도 처리할 수 있어야 함



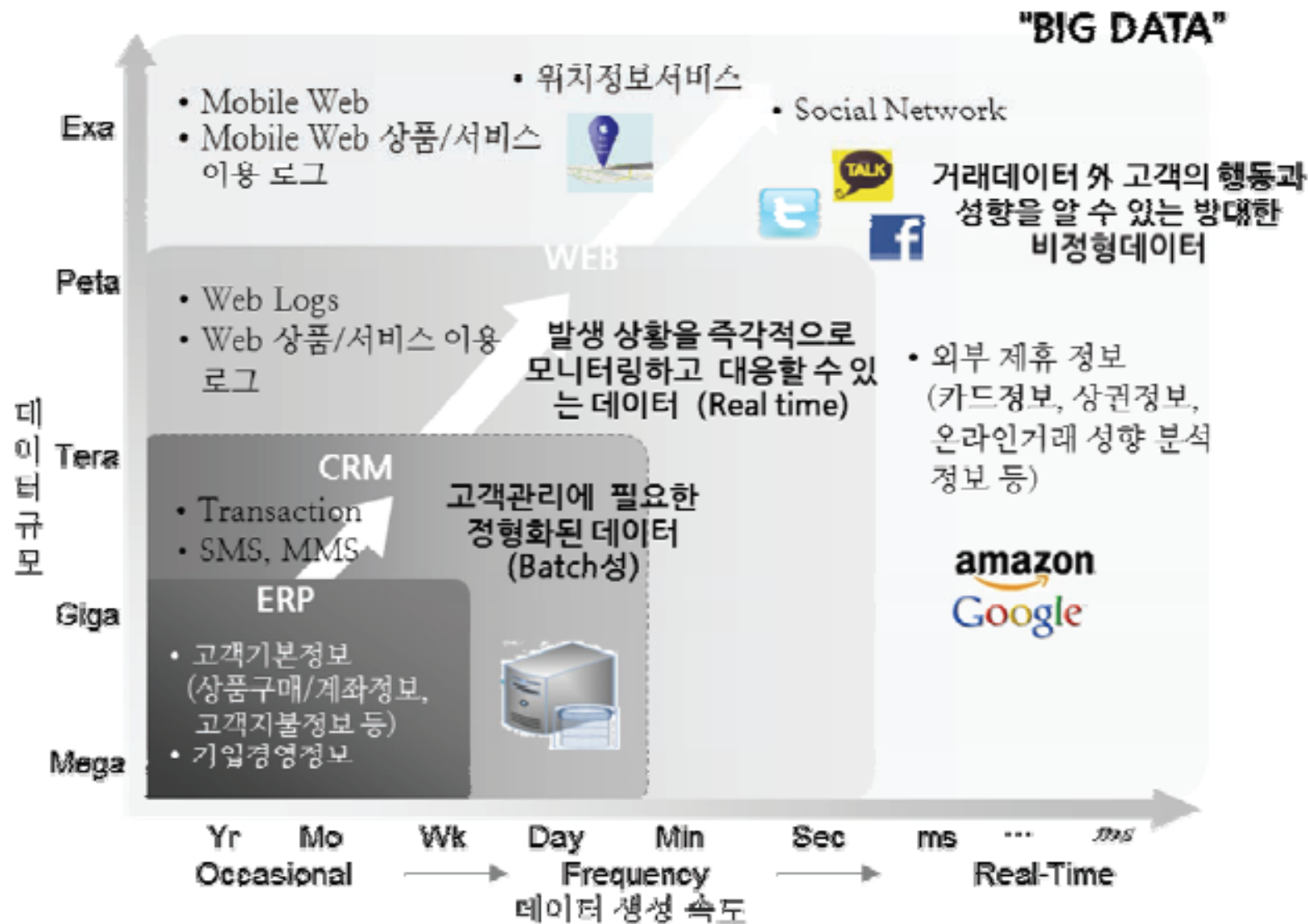
# 속도(Velocity)

- 빅데이터의 속도 특징은 크게 실시간 처리와 장기적인 접근으로 나눌 수 있음
  - 디지털 데이터는 매우 빠른 속도로 생성되기 때문에 데이터의 수집, 저장, 분석 등이 실시간으로 처리돼야 함
  - 예) 사기감지(Fraud Detection)
- 모든 데이터가 실시간 처리만을 요구하는 것은 아님
- 수집된 대량의 데이터를 다양한 분석 기법과 표현 기술로 분석해야 하는데, 이는 장기적이고 전략적인 차원에서 접근할 필요가 있음
- 인간 게놈 프로젝트 (교과서 p. 15 사례연구)
  - 2003년: \$30억, 13년 만에 완성
  - 2012년: \$1,000, 2시간

# 비즈니스 데이터

	비즈니스 데이터	
사용 데이터베이스	<ul style="list-style-type: none"> <li>● 관계형 데이터베이스 (RDBMS) 이용</li> </ul>	
데이터 저장 형태	<ul style="list-style-type: none"> <li>● 테이블</li> </ul>	
데이터의 가치	<ul style="list-style-type: none"> <li>● 개개의 데이터가 중요</li> </ul>	
주 활용 분야	<ul style="list-style-type: none"> <li>● 소비자, 생산, 매출, 자산, 재고 등 ERP나 SCM 등 내부기간 시스템에서 활용</li> <li>● 운영을 위한 개별 데이터 관리</li> </ul>	
필수 요건	<ul style="list-style-type: none"> <li>● 정합성 (consistency): 모든 노드는 같은 시간에 같은 데이터를 보여줘야 함</li> <li>● 유효성 (availability): 일부 노드가 다운되어도 다른 노드에 영향을 주지 않아야 함</li> </ul>	

# 비즈니스 데이터



출처 : Gartner(2011) 빅데이터의 4가지 구성 요소

## ❑ Dark Data문제

- 데이터 품질 문제이며 기존의 DB에서도 동일한 문제
- 실시간 데이터의 폭증에 따라서 노이즈도 증가하며, 이런 데이터에서 의미 있는 분석이나 패턴을 찾는 것은 어려움. (빅노이즈)
- 빅데이터 수집단계에서 검색/수집/변환 과정으로 정제된 데이터 확보가 중요
- 예 1: CJ 오쇼핑의 데이터 정제과정(1년)
  - 같은 제품에 대해서
  - 단일, 묶음 판매 / 이벤트 판매 / 공급자, 바이어에 따라 모든 다른 코드 부여
- 예 2: 네슬레의 데이터정제 (18개월)
  - 바닐라 가격 데이터: 같은 판매자로부터 전 세계 29개의 가격
  - 같은 원료인 바닐라에 대해서 전 세계 모든 공장이 다른 코드 부여
  - 18개월간의 데이터 정합성 확보 작업
- 참고
  - 좋은 정보: 정확성 / 신뢰성 / 유용성
  - 공개 공공정보의 품질 문제 ([www.data.go.kr](http://www.data.go.kr))






DATA 공공데이터포털  
.GO.KR

# 데이터 품질 – Five Stars of Linked Data

## ❑ 데이터 품질과 재이용성 평가 / 개방데이터 공개를 위한 점진적 모델

- data.go.kr(2013.12) – XLS 2368, HWP 1999, PDF 425, JPG 350, TXT 366

## To Get to 5-Stars With Open Data

Star	Definition	Example / Tool*	
	Make your stuff available on the Web (whatever format) under an open license	<a href="#">This Story</a> / MindTouch	
	Make it available as structured data (e.g., Excel instead of image scan of a table)	<a href="#">Spreadsheet</a> / Excel	
	Use non-proprietary formats (e.g., CSV instead of Excel)	<a href="#">Table</a> / MindTouch and Spotfire	
	Use URIs to identify things, so that people can point at your stuff	<a href="#">Table of Contents</a> / MindTouch and Spotfire	RDF (Resource Description Framework)
	Link your data to other data to provide context	<a href="#">Table</a> / MindTouch and Spotfire	

\* Examples of tools used.

Source of Star and Definition: <http://www.w3.org/DesignIssues/LinkedData.html>

# 빅데이터의 출현 - 저장과 전송기술

## □ 저장공간

년도	제품	가격
1980년	1GB HDD	5억 원
2005년	80GB HDD	15만 원
2010년	1TB HDD	15만 원

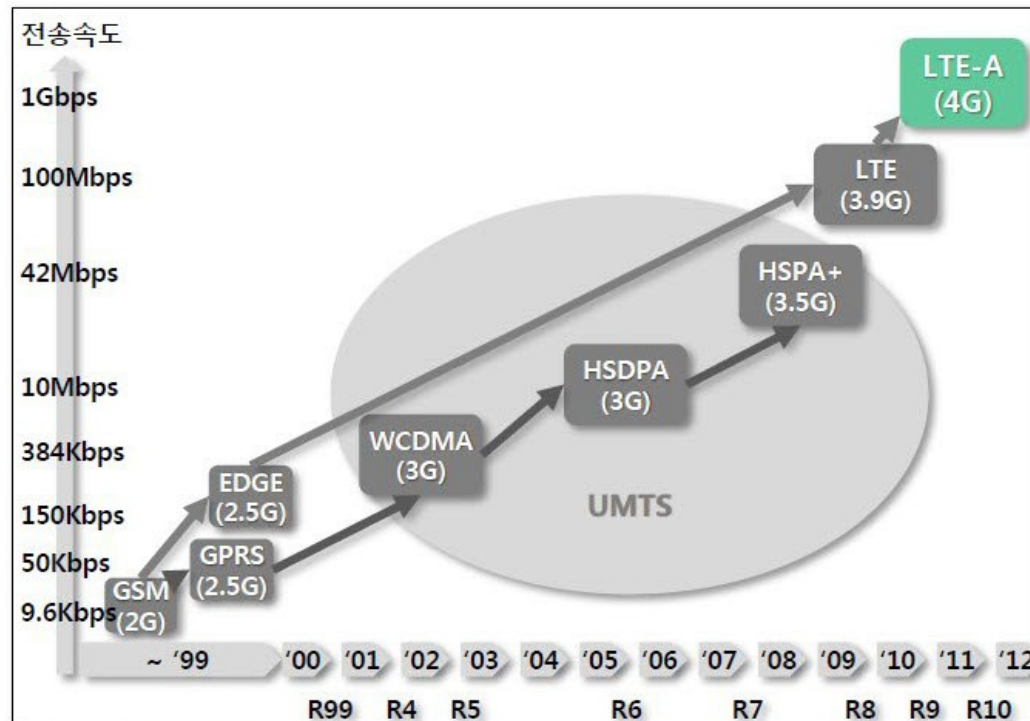


[씨게이트 1TB 7200RPM 64M...](#)

57,650원 ↓ 최저

가격비교 339개

## □ 네트워크 전송속도



출처: 3GPP 및 TTA, 2012

# 빅데이터의 출현 - SNS

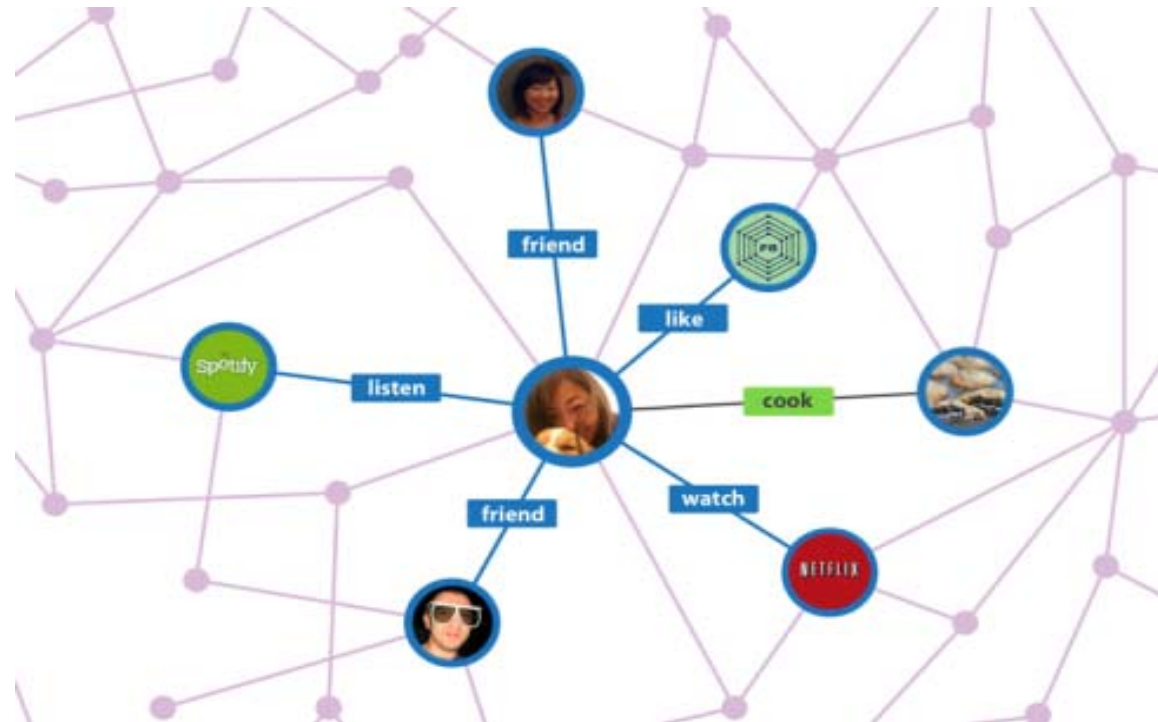
## □ SNS의 확산

- **Exhaust Data** (간접 데이터)

- 다른 활동의 부산물로 생겨나는 데이터에 대한 데이터
- 페이스북의 사진에 태그를 붙이도록 도와주는 안면 인식 루틴
- 댓글, 선호도, 웹 로그

- **소셜 플랫폼**

- 외부 웹 사이트/앱들과 sns와의 연동
- OpenGraph API





# 빅데이터의 출현 - IoT

## ❑ Internet of Things – 사물 인터넷

- 스마트폰 = 개인 정보 + 위치 정보
- 임베디드 시스템
  - M2M (Machine to Machine) : 사물이 감지장치를 통해 정보를 전달
- **IoT + 클라우드**
  - 3천 만개의 사물 센서가 연결
    - 차량 센서
    - 스마트 에너지 미터
    - 스마트 가전
    - 헬스/의료 모니터링 기기
    - 건축물 진동 센서
    - RFID 등
  - InterCloud (개방형 플랫폼)
  - CSB(Cloud Service Brokerage)
- **44ZB(2020, IDC)**



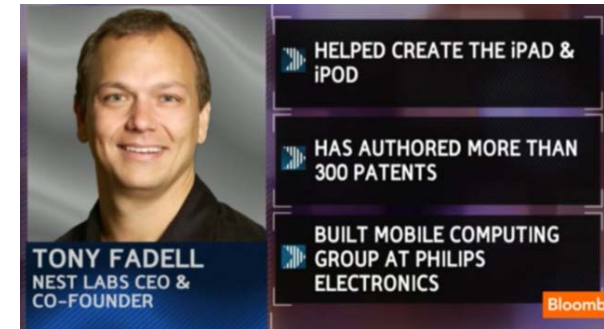
Nest Labs Rush Hour Reward



# 빅데이터의 출현 - IoT

2013. 6. 27, 텍사스 오스틴의 온도가 38°C에 육박했는데, Austin Energy는 에너지 러시 아워가 발생할 수 있음을 예측하고, Nest Labs에 중앙 온도 조정 요청

- 러시 아워 시간 전에 프리 쿨링을 실시하고 러시 아워 시간에 실내 설정 온도를 평소보다 1°C 정도 높임.
- 러시 아워 리워드에 가입한 건물의 에어컨은 가장 더운 시간에 가동 시간이 평균 56% 감소
- 약 10% 정도만 러시 아워 시간 동안 수동으로 온도를 조절하고, 나머지 90%는 자동 온도 조절을 불편 없이 사용
- Austin Energy의 고객들은 이후 85\$ 짜리 수표를 사례로 받았고, Nest Labs는 고객당 30\$의 수익을 창출



- Centralized or Local Sub-network Control (클라우드)
- IoT의 사업성이 높아져 적용분야가 늘어 날 수록 빅데이터 발생

## □ 구글의 NestLab 인수(3.4B, 3조4천억원, 2014.1)의 의미 @ 실리콘밸리

- S/W vs. H/W(Social/Mobile/Cloud -> IoT), 매출 10배 가격, 인재영입(토니파델+ 300명)
- 기술적: wireless mesh/multi-hop (802.15.4기반) / 같은 기술을 사용하는 다른 제품들과 Nest Weave Network 형성 가능 / 인터넷이 끊기더라도 네트워크 유지 / Machine Learning
- Android@Home

# 빅데이터의 출현

빅데이터 소스 및 그 비즈니스 가치에 대한 네가지 유형



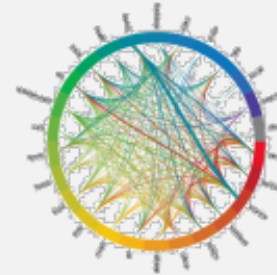
## 전통적인 소스의 크기가 커짐

SOURCE

기업, 정부기관, 금융사, 비즈니스와 소비자 연구, 설문조사, 투표

VALUE

비즈니스 퍼포먼스 측면 - 운영 효율, 매출 관리, 전략 기획



## 인터랙션들의 디지털 데이터

SOURCE

Online click-stream, Application logs, Call/service records, ID scans, Security cameras

VALUE

새로운 매출 영역, 소비자 프로모션, Risk management, Fraud detection



## Web 2.0 현상

SOURCE

소셜 미디어 포스트들에서 생성된 콘텐츠들, 트위터, 블로그, 사진, 비디오, 평가

VALUE

고객 접점, 고객 서비스, 브랜드 관리, 폭넓은 마케팅



## 물건들의 인터넷

SOURCE

기계들이 생성하는 센서 데이터 및 서로 연결된 기기들 사이의 통신 정보

VALUE

운영 효율 제고, 비용 최적화, 발생할 수 있는 리스크를 미리 파악하여 회피

MicroStrategy

# 빅데이터의 출현 – 비즈니스 측면

## □ 비즈니스 혁신과 New BM

- 기존 BM의 개선

- 기존 비즈니스의 운영효율화 및 고도화
- 기업 의사결정의 질과 속도 향상에 의한 성과향상
  - 예: 글로벌 패션기업인 자라(ZARA)는 빅데이터 분석을 통해 상품 공급 프로세스를 혁신하고 재고 수량을 최적화함으로써 매출과 영업이익을 크게 증대시켰음
  - 예: 은행의 리스크 관리 - 대출자 채무불이행 가능성 측정을 위해서 개인 신용도 + 재정상태 + 고객 행동 데이터 분석

- 새로운 BM <- 고객가치의 창출

- 소비자들의 일상적인 디지털 생활의 데이터(Life Log)로부터 고객 가치 창출

# 빅데이터의 출현 – 비즈니스 측면

## 미국 기업들이 기대하는 빅데이터 활용 시의 장점



• eMarketer 2012

# 빅데이터의 출현 – 기업 의사결정

## □ 미국 은행의 빅데이터 활용 예

- Segmentation
- 이탈고객 관리 / 상품개발
- 고객의 거래 패턴 파악 -> 조직구성에 반영

은행	특징
JPMorgan Chase	방대한 양의 고객 신용카드 이용 정보와 정부를 통해 확보한 금융소비자 재무 정보를 통합하여 새로운 소비 트렌드를 발견하고 세그멘테이션 등에 활용
Bank of America	자사 웹사이트 및 콜센터 통화 내역에서 고객 기록을 수집하여 이탈한 고객들의 특성을 파악, 이 정보를 All-in-One 상품 개발에 이용하여 온라인 상품으로 제공한 결과 이익이 크게 증가
Citigroup	전 세계적으로 고객 정보를 수집하여 개별 고객의 거래 패턴을 발견하기 위해 노력, 그룹 내 새로운 조직 구성에 반영
Wells Fargo	2년 전부터 각기 다른 시스템에서 관리되어왔던 고객 기록을 통합하는 작업을 추진해왔으며 IT 투자비용의 80%는 사업부문 단위에서 새로운 서비스와 역량 개발을 위해 필요한 분야에 지출

## □ 푸르덴셜 생명의 우량채 판단 모델

- 과거 10년 데이터로 고객 집단별 평균사망률을 추정하여 언더라이팅 업무에 적용
  - Logistic 모델, Generalized Additive 모델 사용
  - 모델 사용 결과 건강상태(혈당, 간수치 등)보다 직업/소득에 따른 생활습관과 라이프스타일이 더 중요함을 발견

# 빅데이터의 출현 – Customer Experience

## □ 패션산업

### • Nordstrom TrueFit

- 사이즈가 소비자들이 온라인에서 의류 및 풋웨어 구매를 망설이는 요인
  - 온라인 패션 구매 후 반품 이유 중 70% 이상
  - 무료 배송 및 반품 서비스를 제공하는 자포스(Zappos)
- 빅데이터 활용
  - 1000개 이상의 매장에서의 정보수집
  - 수백만 개의 고객 프로필

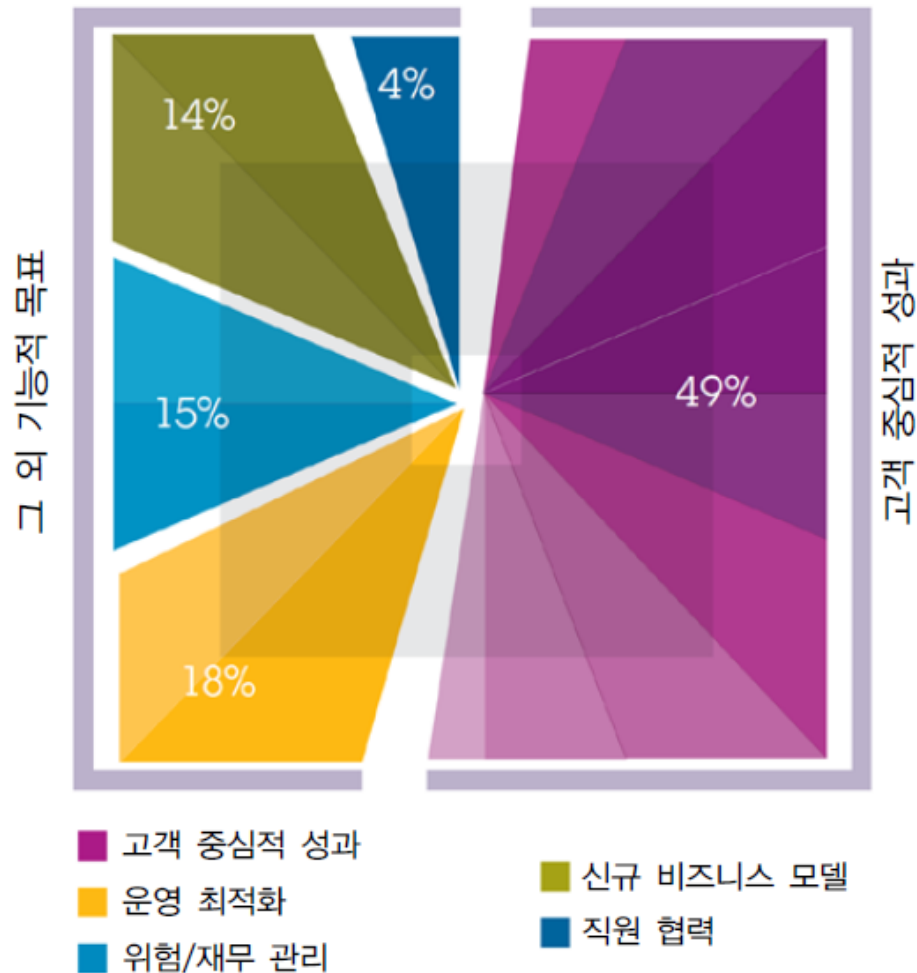
### • 트렌드 예측

- 전문가의 직관에 기반 (Worth Global Style Network , Stylesight)
- 데이터에 기반한 예측 (Editd)
  - 현재 시장에 존재하는 상품들이 어떤 것이 있고,
  - 얼마나 팔렸으며,
  - 얼마나 빨리 상품이 팔렸는지
  - 소셜 미디어와 블로그를 모니터 소프트웨어
  - 주요고객: 갭(Gap), 타겟(Target), 길트그룹(Gilt Group)



# 빅데이터의 출현 – 비즈니스 측면

빅 데이터의 도입을 통해 기업이 추구하는 성과



빅데이터 투자를 2년 내에 계획중인  
생명보험 기업의 도전/장애요인  
(Text p. 97)

빅 데이터 도입을 통해 조직이 달성하고자 하는 목표  
총 응답자 수 = 1,067 명

IBM 보고서

# 빅데이터의 출현 - 국가정책

- 국가 경쟁력 제고를 위한 빅데이터 정책
- 미국 백악관
  - 대통령 과학기술자문위원회(PCAST), 과학기술정책실
  - Big Data R&D Initiative : 연방정부기구 주도 (NSF, NIH, DoD, DARPA, DoE, NASA ...)
  - Data.gov
- 일본 총무성과 국가전략회의
  - Active Data(2012.5)
  - 10조 엔의 부가가치 창출
  - 12 - 15조 엔의 사회비용 절감
- EU
  - Open Data Strategy
    - Open-data.Europa.eu
- 영국
  - Data Strategy Board
  - Data.gov.uk

●● 각국의 빅데이터 관련 전략과 추진 방향 ●●

기관명	주요 내용
대한민국, 국가정보화전략위원회(2011)	<ul style="list-style-type: none"> <li>• '빅데이터를 활용한 스마트 정부 구현(안)' 제시('11.11)</li> <li>• 빅데이터 시대에 대비하여 융합 지식과 분석 행정을 통한 스마트 정부 실현을 위한 세부과제 도출</li> </ul>
美 대통령 과학기술자문위 (2010)	<ul style="list-style-type: none"> <li>• 'Designing a Digital Future' 보고서 발표('10.12)</li> <li>• 대규모 데이터(large-scale data)의 가치에 주목하고 데이터 취합 및 관리, 분석의 중요성 제시</li> <li>• 미국 정부기관들이 데이터를 지식으로, 지식을 행동으로 변환하는 전략에 집중해야 함을 주장</li> </ul>
싱가포르 RAHS 시스템	<ul style="list-style-type: none"> <li>• 데이터를 기반으로 싱가포르를 위협하는 리스크에 대한 평가와 환경 변화를 탐지하는 국가위험관리 시스템(RAHS) 구축</li> <li>• 호라이즌 스캐닝(Risk Assessment Horizontal Scanning) 시스템 : 데이터 수집, 분류, 분석, 관계성, 예측 지원</li> </ul>
OECD	<ul style="list-style-type: none"> <li>• 빅데이터를 비즈니스 효율성을 제공하는 새로운 자산으로 인식</li> <li>• 제15차 WPIIS 회의<sup>4)</sup>에서 빅데이터의 경제학 측정(Measuring the Economics of "Big Data")을 의제로 채택 : 무형 자산인 Big Data의 경제적 가치를 실증하고 정책적 함의 제공이 목표</li> </ul>



# 빅데이터 경제적가치

## □ 빅데이터 새로운 기술혁신인 동시에 경제적 이익을 창출하는 산업

- 맥킨지(McKinsey) - 빅데이터기술이 미국에서 최대 7,000억 달러의 경제적 효과와 2018년까지 연간 150만 명의 새로운 일자리를 창출할 것으로 추정함.4)
- 영국의 비즈니스센터 - 빅데이터 관련 산업이 향후 5년 간 영국 내에 약 360조 원 이상의 경제적 파급효과를 유발할 것으로 예측함.5)
  - 4) McKinsey, “Big data : The next frontier for innovation, competition, and productivity”, 2011.06.
  - 5) 한국정보화진흥원, “새로운 미래를 여는 빅데이터 시대” , 2013-02.

# 빅데이터의 출현 - 개방데이터정책

## □ ODB(Open Data Barometer), p.294

- WWW재단 + Open Data Institute(영국)
- 국가별 공공데이터 개방정책에 대한 평가
  - 준비, 실행, 효과 3개 부분에 대한 측정지표
- 정부의 개방데이터 정책추진 방향과 발전단계 모델 제시

순위	국가명	세부 지표			ODB 평점
		준비도	실행력	영향력	
1위	영국	100	100	79.91	100
2위	미국	95.26	86.67	100	93.38
3위	스웨덴	95.20	83.14	71.95	85.75
4위	뉴질랜드	81.88	65.49	89.81	74.34
5위	노르웨이	91.88	70.98	46.15	71.86
12위	한국	77.19	54.90	24.56	54.21

- 데이터 자체의 개방 보다 애플리케이션을 개발할 수 있게 하는 플랫폼 환경 제공이 더 중요함.
  - 주요 개방 데이터 플랫폼 (p. 300)

# 빅데이터 경제적가치

## □ 국가별 ‘공공데이터’ 개방의 경제적 효과 사례 (12장 3절- 중국, 영국, EU)

- 독일의 지리정보시장 50% 성장
- 네덜란드는 지리정보 시장에서 1만 5천 명 고용창출 효과 달성
- 유럽연합(EU)은 1,860억 유로의 경제적 효과 추산
- 영국은 150억 파운드의 경제적 효과와 5천 8백개 이상의 일자리창출 기대
- 미국은 900억 달러의 경제적 가치 창출 및 기상데이터 활용으로 4천 개 이상의 일자리 창출 효과 기대
- 일본은 122억 달러의 시장과 500억 달러 이상의 경제적 효과 예상

# 빅데이터의 출현 - 국가정책

- 부처와 협력을 통해 선정된 이슈와 미래비전을 중심으로 데이터 분석 및 부처 정책수립 지원
  - 2014년도에는 공모를 통해 선정된 기상청/농진청, 여가부, 특허청, 관세청 등 최종 4개 부처를 선정하여 데이터 분석기반 미래전략 수립 지원

## 컨설팅 아젠다 1

### 기상기후 데이터 분석을 통한 과학적 농업경영 지원

농작물 생산에 영향을 미치는 기상요인 분석을 통해  
'농작물 생산량 예측 모형' 개발 및 국민경제 안정화 지원

[주관]



[연관 기관]



## 컨설팅 아젠다 2

### 특허기술정보를 활용한 미래 ICT 유망 기술 예측

특허정보 분석지표와 비정형 빅데이터를 매쉬업 분석하여  
국가특허전략 청사진 사업 부품분야기술의  
미래 유망성 예측

[주관] 특허청

[연관 기관] KIPSI  
한국지식재산전략원

## 데이터 기반 미래 전략 수립

## 컨설팅 아젠다 3

### 데이터 기반의 경력단절여성 맞춤형 재취업 지원

새일센터 데이터의 다차원 빅데이터 분석을 통해  
경력단절여성의 일자리 미스매치 원인을 진단하고  
원인별 맞춤형 취업 지원

[주관] 여성가족부

[연관 기관]



## 컨설팅 아젠다 4

### 전자상거래 수출입 통관 데이터를 통한 대국민 정보 제공

"전자상거래 유망 시그널 103選" 백서를 통해 건전한  
시장 경쟁 유도 및 신규 일자리 창출을 위한  
시그널 상품 정보 제공

[주관] 관세청  
KOREA CUSTOMS SERVICE

[연관 기관] 한국관세무역개발원

# 빅데이터의 출현 - 국가정책

## <분석결과 종합>

'기상청' 제공  
분석 적용 기상기후 변수

총 33 종류 기상기후 변수

1월 평균 일교차  
2월 평균 일교차  
3월 평균 일교차  
4월 누적강수량  
6월 평균 일조시간  
12월 평균 습도  
6월 최대누적강수량(태풍)  
7월 최대풍속 15km 이상 유/무(태풍)  
3월 최저기온 < 0도 월별 관측 횟수  
5월 20mm 이상 월별 강우관측 횟수  
8월 50mm 이상 월별 강우관측 횟수  
1월 일 최저기온 > 90h 백분위수 월별 관측 횟수  
12월 일 최저기온 > 90h 백분위수 월별 관측 횟수  
5월 일 평균기온 > 18도 월별 관측 횟수

23개 종류  
**193개**



18개 종류  
**162개**



22개 종류  
**144개**



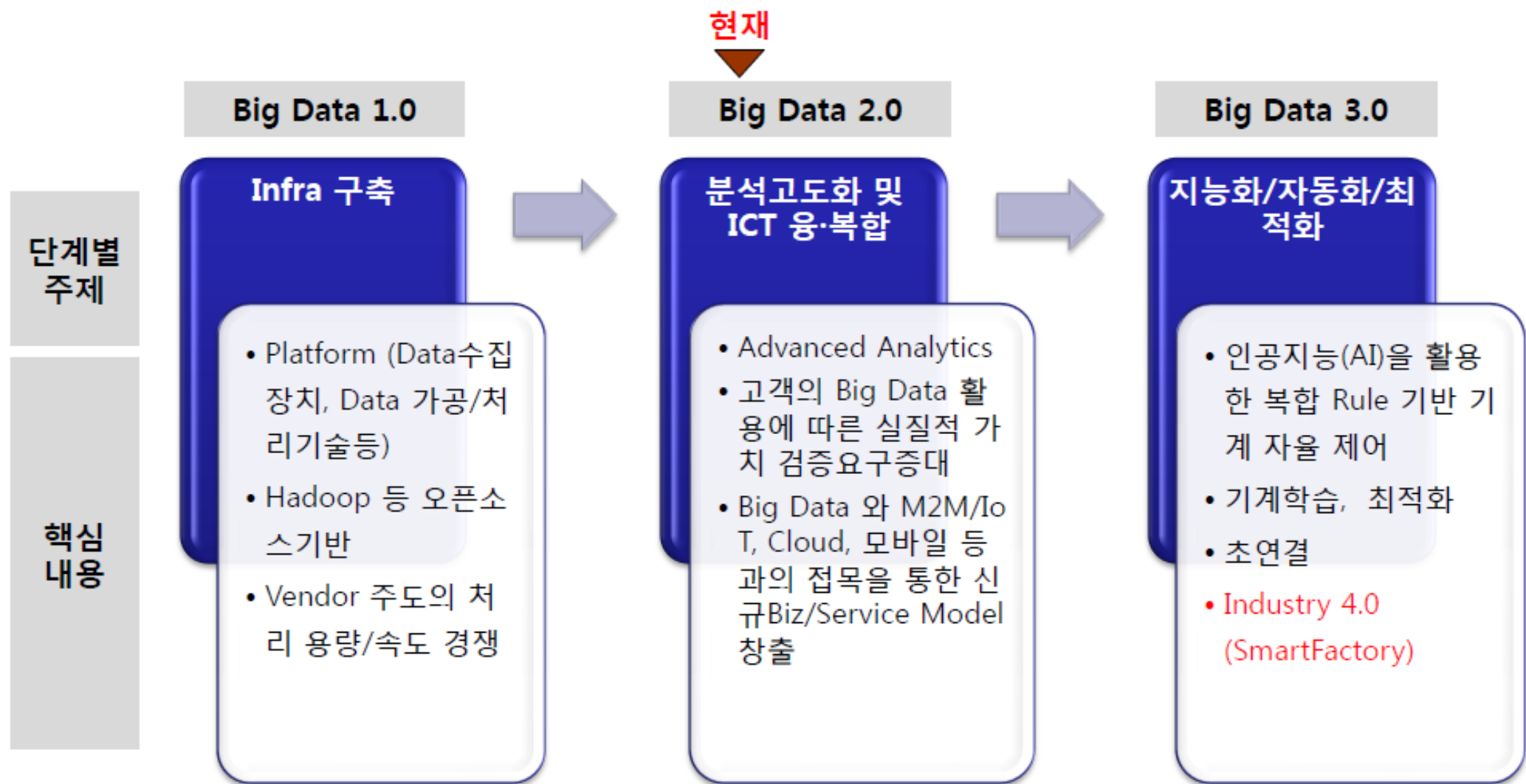
'통계청' 제공  
지역별 농산물 생산량 정보

'수급조절위원회' 활용  
지역별 단수 예측 모형



2000 2001 2002 2003 2004 2005 ... 2013 2014

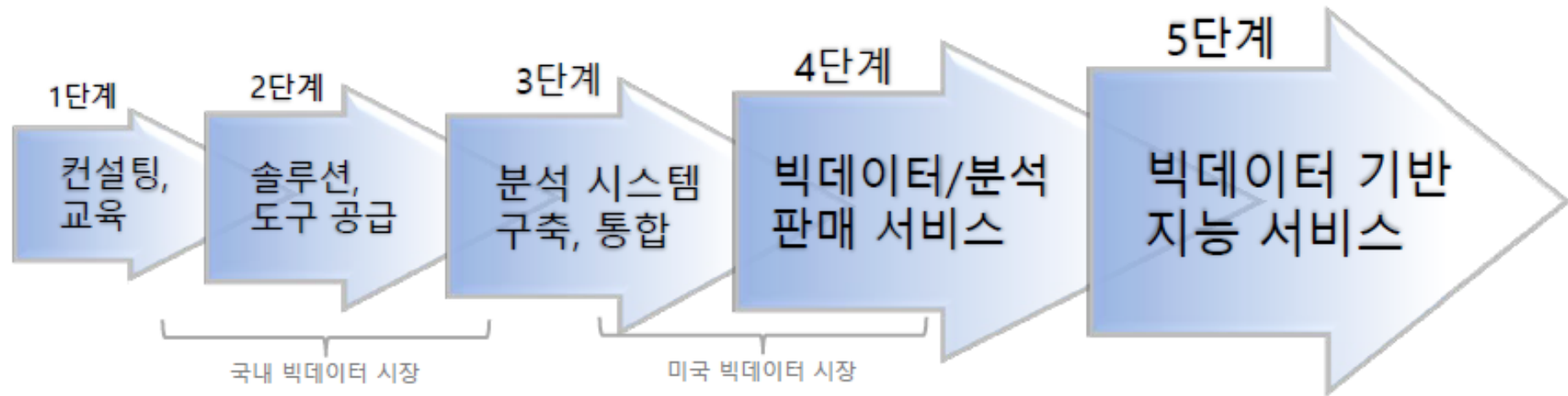
# 빅데이터의 발전방향



디지털 기회 선점을 위해 관점을 바꿔라, SK C&C 이원석 상무, CIO Summit 2015

# 빅데이터의 발전방향

## 빅데이터 산업 발전 5단계



- ❑ Big Data Predictions for 2016 – Information Week
- ❑ Top 8 Trends for Big Data 2016 – Information Age
- ❑ Big Data Trends – Tableau

# Homework #2

- 아마존 AWS와 MS Azure는 클라우드 상에서 빅데이터 인프라를 구축할 수 있는 서비스를 제공하고 있다.
  - 이에 대한 조사 분석
  - 학번 끝자리 홀수는 AWS, 짝수는 MS