

빅데이터 개요

2018년 1월 15일

유성준
세종대학교 컴퓨터공학과
인공지능-빅데이터연구센터 센터장
<http://abrc.or.kr>

목차

- 개요
- Why
- 데이터 과학
- 빅데이터 성공 요건

빅데이터 개요

- 빅데이터란 ?

- 기존 컴퓨팅 기술로는 저장, 관리, 분석이 불가능할 정도로 큰 데이터 집합과 관련 기술, 인력 등을 포괄하는 의미
- IT 기술에서 출발했으나 정치, 사회, 문화, 등 삶 전체의 이슈, 혁신 패러다임으로 부각 (Economist, Gartner, McKinsey, NYT)

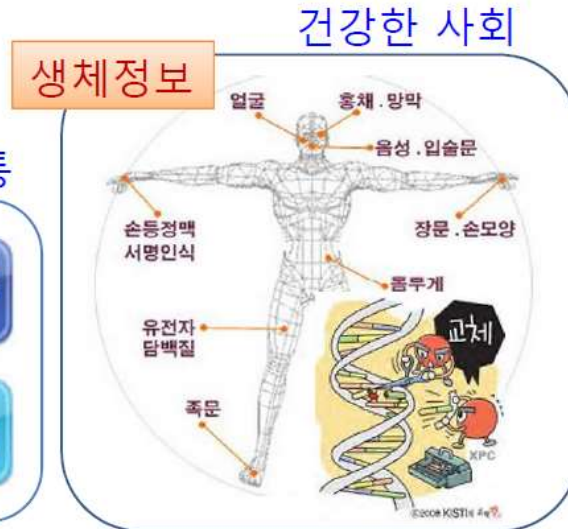
- 빅데이터의 특징 : 3Vs

	기존 데이터	빅데이터
크기(volume)	MB, GB	TB, PB 시대에 따라 다른 기준
다양성(variety)	정형 데이터 위주 (매출, 재고, 인사 생산, 고객 데이터)	비정형 데이터 (동영상, SNS, 음악, Stream data 등)
속도(velocity) (생성, 유통, 활용주기)	수시간~수주	수초, 수분 (실시간 분석 요구)

<출처: 조완섭, “빅데이터와 활용사례,” 2016년 5월 30일>

빅데이터 개요

- 빅데이터는 생성 원천은 ?



<출처: 조완섭, “빅데이터와 활용사례,” 2016년 5월 30일>

Big Data EveryWhere!

- Lots of data is being collected and warehoused
 - Web data, e-commerce
 - purchases at department/grocery stores
 - Bank/Credit Card transactions
 - Social Network



Why is “big data” a “big deal”?

- Private Sector
 - Walmart handles more than 1 million customer transactions every hour, which is imported into databases estimated to contain more than 2.5 petabytes of data
 - Facebook handles 40 billion photos from its user base.
 - Falcon Credit Card Fraud Detection System protects 2.1 billion active accounts world-wide

Why is “big data” a “big deal”?

- Google processes 20 PB a day (2008)
- Facebook has 2.5 PB of user data + 15 TB/day (4/2009)
- eBay has 6.5 PB of user data + 50 TB/day (5/2009)



640K ought to be
enough for anybody.

Why is “big data” a “big deal”?

- Government
 - Obama administration announced “big data” initiative
 - Many different big data programs launched
- Science
 - Large Synoptic Survey Telescope will generate 140 Terabyte of data every 5 days.
 - Biomedical computation like decoding human Genome & personalized medicine
 - Social science revolution
 - -...

Why is “big data” a “big deal”?

• 빅데이터 시대의 도래

세계는 2010년 zettabyte 시대에 돌입
1 Zettabyte : 美의회도서관 정보(235TB, 11/4 기준)
의 4백만 배, 16GB iPad로 축구장 넓이로 쌓아도
대기권 높이 두배

빅데이터 활용으로 연간
미국 의료분야에서 \$3,300억 가치생산
유럽 공공분야에서 €2,500억 절감효과

2018년까지 미국에서만 연간
분석전문가 14~19 만명,
데이터 기반 관리자 150 만명 추가 수요

빅 데이터 생성 속도

- 하루 **250경** 바이트 비정형 정보
- 매달 **10억** 여 개 트윗
- 매달 **300억** 여 개 페이스북 메시지
- 1조** 대 이상 모바일 기기로 가속화

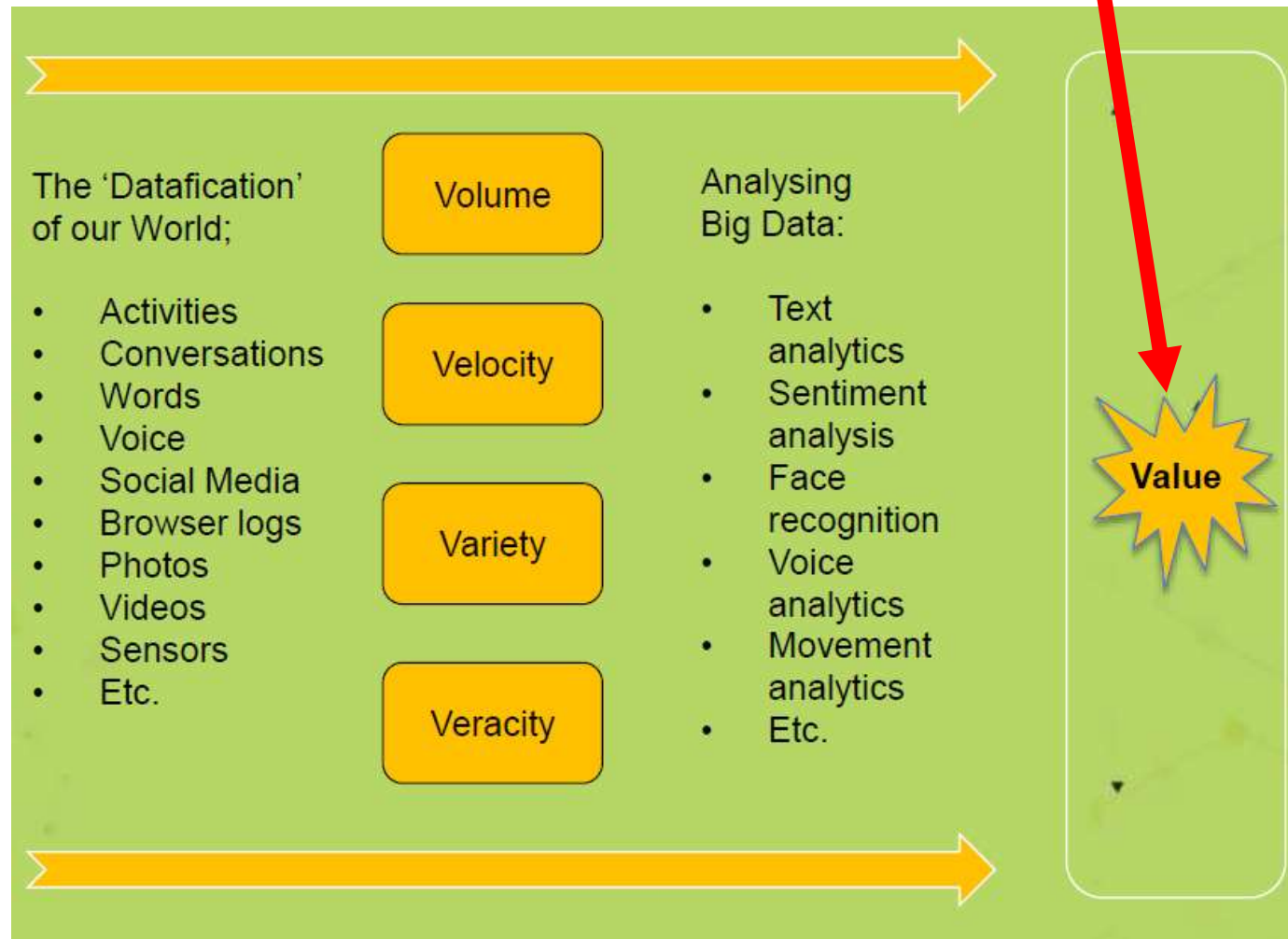
전세계 데이터는
매년 40% 증가

Global information created and available storage (Exabytes)



(출처) McKinsey (2011,05), “Big Data: The next frontier for innovation, competition, and productivity”

빅데이터 분석의 전제



빅데이터 활용 효과

- 빅데이터 효과

- 인류가 직면한 **의료, 환경, 식량, 에너지** 등의 분야에서 해결방안 제시

 <p>Economist (2010.05)</p> <ul style="list-style-type: none">• SNS와 M2M 센서 등을 통해 도처에 존재하는 데이터의 효과적 분석으로 전 세계가 직면한 <u>환경, 에너지, 식량, 의료문제</u>에 대한 해결책을 제시	 <p>Gartner (2011.03)</p> <ul style="list-style-type: none">• 데이터는 21세기의 원유이며 데이터가 미래 경쟁 우위를 좌우• 기업들은 다가온 데이터 경제시대를 이해하고 정보 공유를 늘려 Information silo를 극복해야 함	 <p>McKinsey (2011.05)</p> <p>Big data: The next frontier for innovation, competition, and productivity</p> <ul style="list-style-type: none">• 빅데이터의 활용에 따라 기업/공공 분야의 경쟁력 확보와 생산성 개선, 사업혁신/신규사업 발굴• 특히 의료, 공공행정 등 5대 분야에서 6천억불 이상의 가치 창출 예상
---	--	---

<출처: 조완섭, “빅데이터와 활용사례,” 2016년 5월 30일>

빅데이터 활용 효과

- 빅데이터의 효과 - 산업분야별 경제적 효과



의료 건강(美)

- 매년 \$3,300억 가치
 - 年 ~0.7%의 생산성 증가
- 스페인 의료비의 2배



공공, 행정 부문(유럽)

- 매년 €2,500억 가치
 - 年 ~0.5%의 생산성 증가
- 그리스 GDP 규모



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)

<출처: 조완섭, "빅데이터와 활용사례," 2016년 5월 30일>

빅데이터 시장 전망



빅데이터 시장 전망

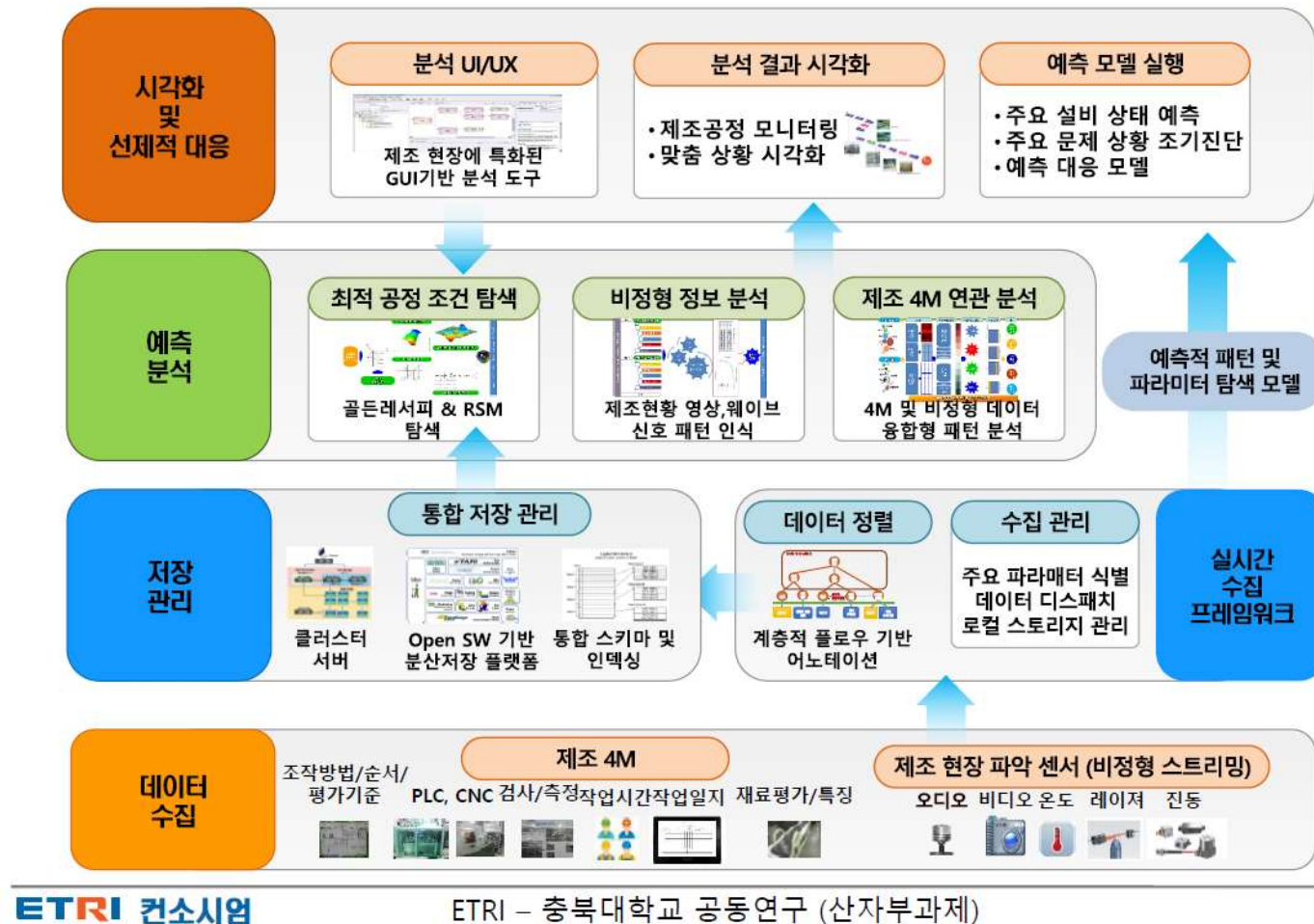
(단위 : 백만달러, %)

구분	2010	2011	2012	2013	2014	2015	비중
하드웨어	919	1,371	2,269	3,368	4,474	5,756	34
SW	1,062	1,415	1,851	2,476	3,376	4,625	27
서비스	1,236	1,979	2,721	3,883	5,099	6,538	39
합계	3,217	4,765	6,841	9,727	12,949	16,919	100

IDC, Worldwide Big Data Technology and Service market Forecast, 2011

<출처: 조완섭, "빅데이터와 활용사례," 2016년 5월 30일>

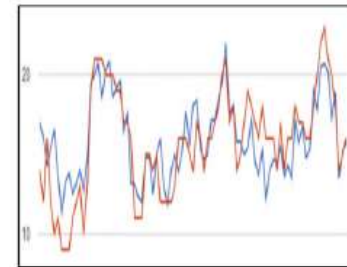
빅데이터 수집 및 분석 과정



<출처: 조완섭, "빅데이터와 활용사례," 2015년 6월 5일>

데이터 과학(자)

- 데이터 분석을 통해 가치(data product)를 창출하는 전문가
 - 데이터의 다양한 소스를 찾아 데이터를 추출하고,
 - 크고 복잡한 데이터를 구조화/단순화하여 분석이 가능하게 만들며,
 - 이상한 데이터를 실시간으로 탐지해 내고,
 - 적절한 통계적 모델링을 통하여 예측을 수행하며,
 - 인사이트를 발견한 후 시각화하여 비즈니스의 방향을 제시함



<출처: 조완섭, “빅데이터와 활용사례,” 2016년 5월 30일>

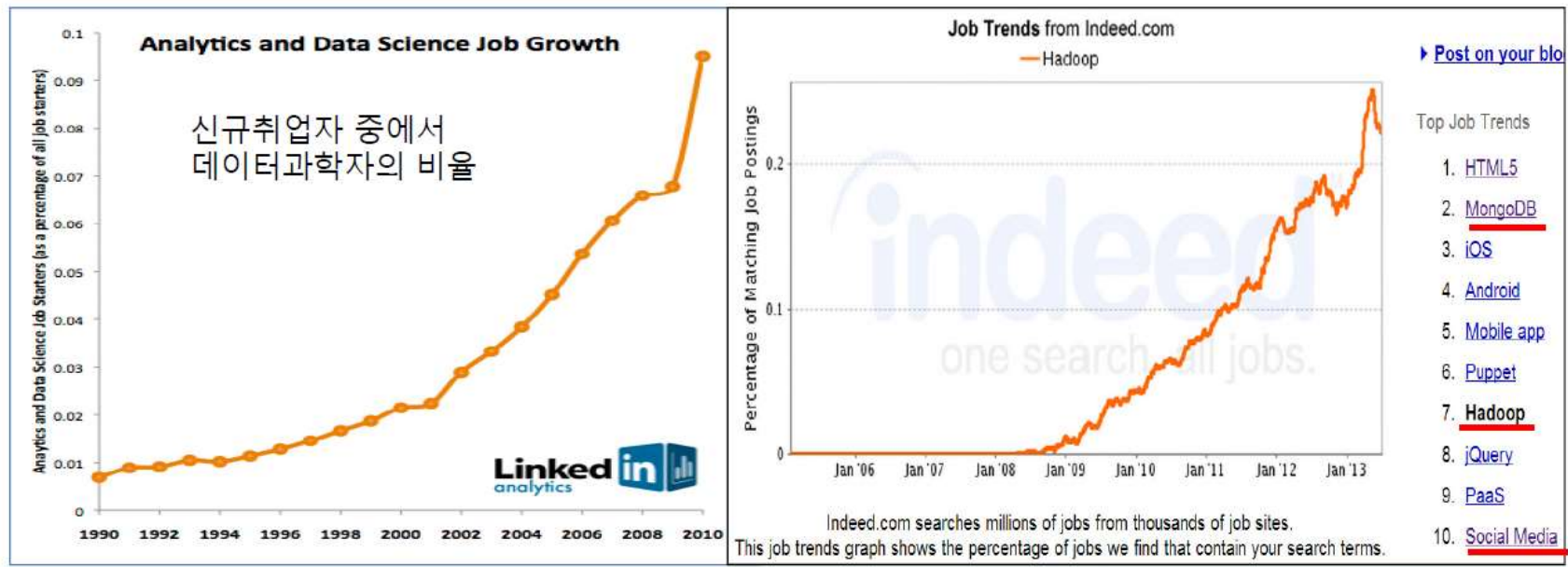
데이터과학자

- 데이터과학자 수요

- 국가별 향후 3-5년 이내 수요

- 미국 : 18만명의 분석전문가, 150만명의 데이터기반 관리자(맥킨지 2011년 보고서)
 - 영국: 5만명 예상
 - 우리나라 : 2017년까지 약 11,000명 예상

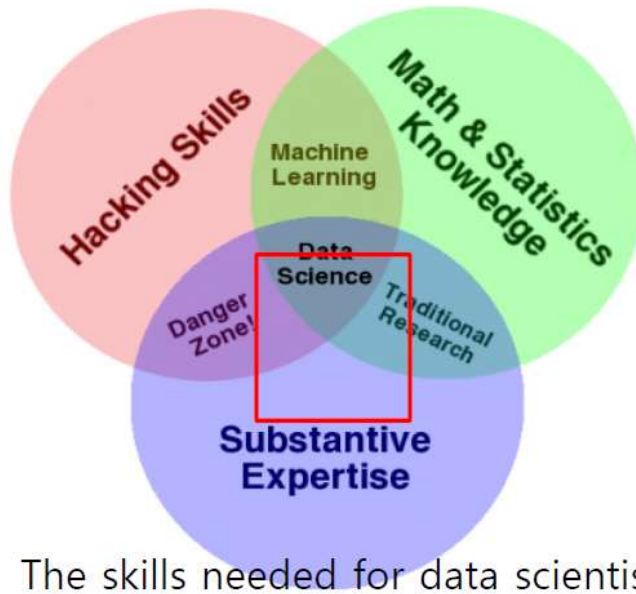
(KISDI Premium Report「(14-10)빅데이터 2.0시대, 주요 이슈와 정책적 시사점」)



<출처: 조완섭, “빅데이터와 활용사례,” 2016년 5월 30일>

데이터과학자

- Data Scientist – Venn Diagram



데이터과학자는
IT 기술 (hacking skills),
수학 및 통계 지식
현장 업무 지식 (business, science 등)을
모두 갖추어야 한다.

두 가지만 갖추면
기존 과학자나 기계학습 혹은
위험한 사람(danger zone)이 된다.

The skills needed for data scientists

출처 : <http://www.dataists.com/2010/09/the-data-science-venn-diagram/>

마윈 "IT시대 저물고 향후 30년간은 DT시대 온다"

DT 혁명에 기반을 둔 중소기업이 창의력을 발휘하도록 대기업과 정부가 지원해야 할 것이다.

<출처: 조완섭, "빅데이터와 활용사례," 2016년 5월 30일>

빅데이터 성공 요건

- 빅데이터가 성공하려면 ?
 - 데이터 기반 조직문화 형성
 - 직관보다 데이터 기반의 과학적 의사결정 중시 문화가 필요
 - CEO 의지가 가장 중요한 관건임
 - 조직의 분석지능 제고
 - 빅데이터 시대에 분석지능이 높은 조직(기업, 국가)일수록 혁신 능력과 리스크 관리 능력이 우수함(MIT Report) - 데이터과학자
 - 작은 업무에서라도 빅데이터 분석을 시작!
 - 재직자 교육의 필요성 (경기도 사례)
 - 분석결과가 업무혁신으로 연결, 지속되어야 함
 - 작은 데이터부터 빅데이터까지 모두 고려
 - 업무혁신으로 연결되어야 하고, 지속되어야 함
 - 1회성 => 분석결과가 살아 움직이게...

<출처: 조완섭, “빅데이터와 활용사례,” 2016년 5월 30일>

빅데이터 성공 요건

- **지속가능한 빅데이터가 되려면 ?**
 - 빅데이터의 **활용**과 함께 **데이터 거버넌스**가 구축되어야 !

빅데이터 거버넌스는

데이터의 품질보장, 프라이버시 보호, 데이터 수명관리, 전담조직과 규정정립, 데이터 소유권과 관리권 명확화 등을 통하여 빅데이터가 적시에 필요한 사람에게 제공되도록 체계를 확립하는 것

빅데이터 거버넌스가 확립되지 못하면

- 품질이 낮은 데이터를 중요한 의사결정에 사용함으로써 심각한 문제를 야기
- 개인 프라이버시 관련 데이터로 인해 빅브라더의 우려가 현실화
- 폭증하는 데이터에 대한 메타 데이터 & 수명 관리 소홀로 인한 IT 비용 급증
- 데이터 관련 문제를 전담하고 책임지는 조직과 인력이 없다면 빅데이터 효과는 일회성에 그칠 것임

<출처: 조완섭, “빅데이터와 활용사례,” 2016년 5월 30일>