

Be as proud of Sogang as Sogang is proud of you

빅데이터 컴퓨팅 :: 개요



서강대학교
SOGANG UNIVERSITY

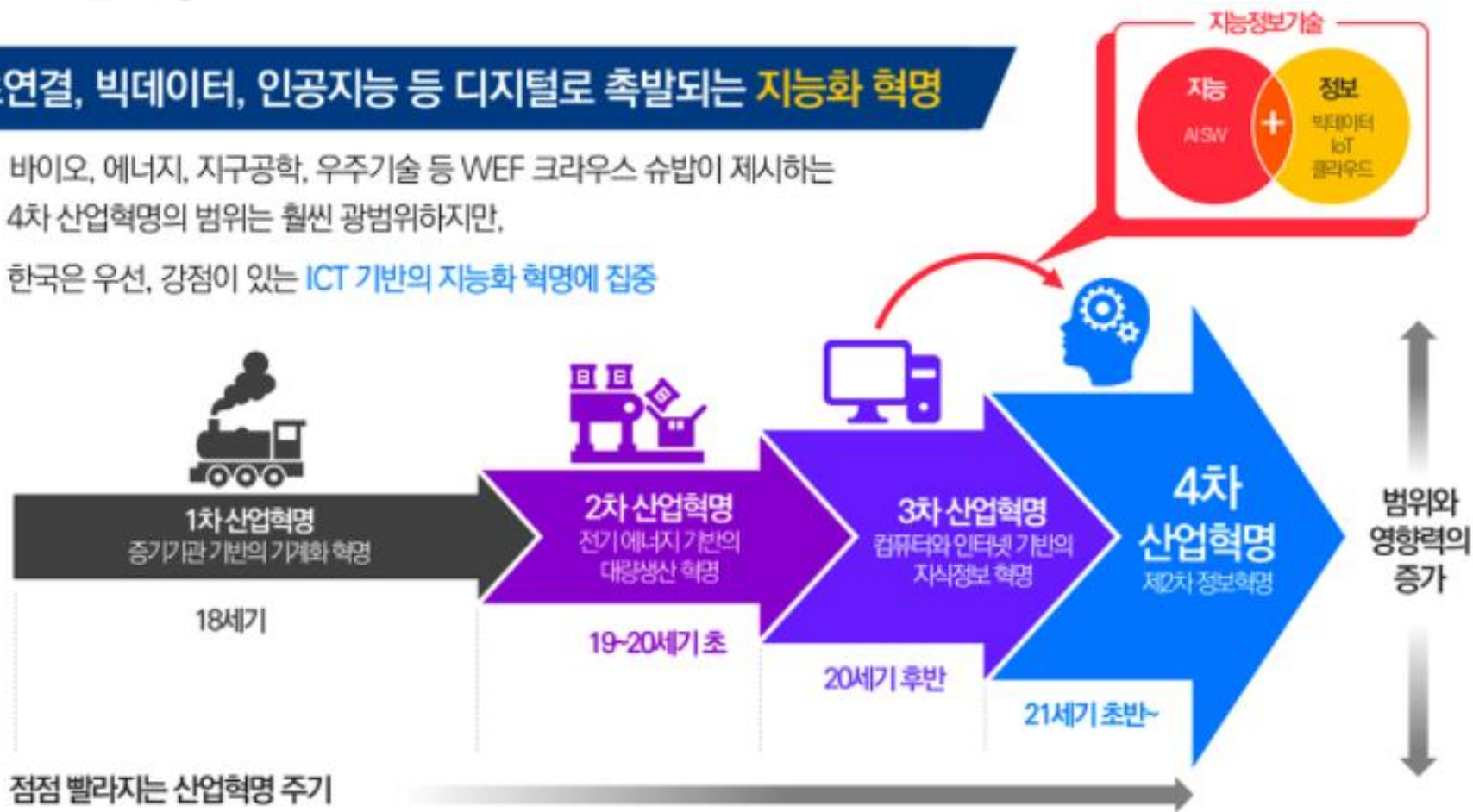
- 위키피디아: 기존의 소프트웨어로 처리하기에는 너무 크거나 복잡한 데이터를 분석하고 체계적으로 정보를 추출, 처리하는 방법을 다루는 분야
- 국가전략위원회: 대용량 데이터를 활용, 분석하여 가치 있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술
- 삼성경제연구소: 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합으로 대규모 데이터와 관계된 기술 및 도구(수집, 저장, 검색, 공유, 분석, 시각화 등)를 모두 포함하는 개념

- ICT(Information & Communication Technology) 인프라의 고도화 및 성숙
- 소셜 네트워크에 기반한 대규모 데이터
- 기업의 축적된 자산데이터, 이의 활용을 통한 분석 수요 증가
- 4차 산업혁명 시대로의 패러다임 전환
- 하드웨어 기술 발전에 따른 데이터 저장 및 처리 비용 하락 (클라우드의 등장)

■ 4차 산업혁명

초연결, 빅데이터, 인공지능 등 디지털로 촉발되는 **지능화 혁명**

- 바이오, 에너지, 지구공학, 우주기술 등 WEF 크라우드 슈밥이 제시하는 4차 산업혁명의 범위는 훨씬 광범위하지만,
- 한국은 우선, 강점이 있는 **ICT 기반의 지능화 혁명에 집중**



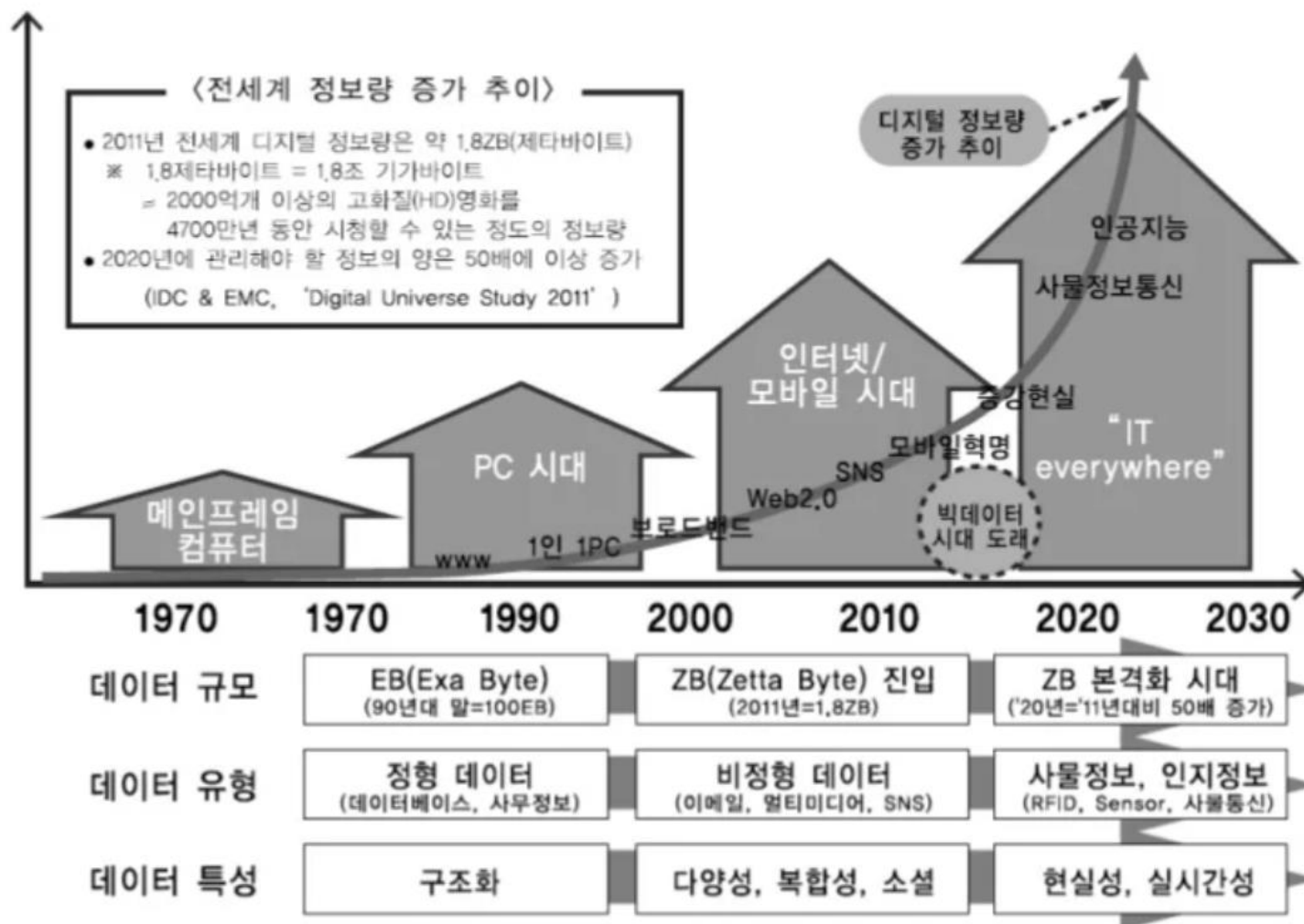
* 대통령 직속 4차산업혁명위원회

■ 정보기술 패러다임의 변화

구분	PC 시대	인터넷 시대	모바일 시대	스마트 시대
패러다임 변화	디지털화, 전산화	온라인화, 정보화	모바일화, 소셜화	지능화, 개인화, 사물정보화
정보 기술 이슈	PC, PC통신, 데이터베이스	인터넷, WWW	모바일 인터넷, 스마트폰	빅데이터, 차세대PC, 사물인터넷(IoT)
핵심 분야	PC, OS	포털, 검색엔진, Web2.0	앱서비스, SNS	미래전망, 상황인식, 개인맞춤 서비스
대표기업	MS, IBM, APPLE	구글, 네이버, 유튜브	페이스북, 트위터, 애플, 구글	구글, 네이버, 카카오, 페이스북, AWS, MS, IBM
정보기술 비전	1인 1PC	클릭 e-Korea	손안의 PC, 소셜 네트워크	빅데이터로 새로운 가치 창출

* NIA(한국정보화진흥원) 참고

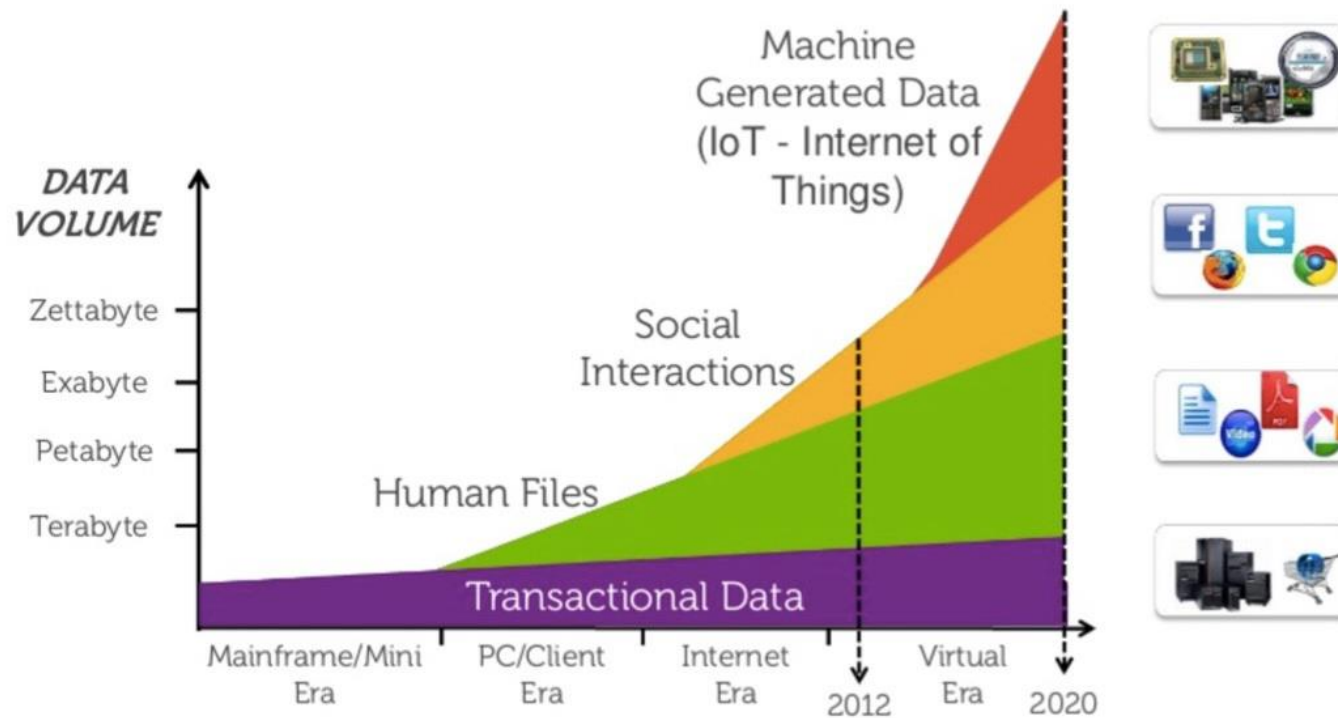
■ ICT 발전에 따른 데이터 증가



* NIA(한국정보화진흥원)

- 새로운 물결: 생산성 향상, 비용 절감, 새로운 마케팅 기법을 넘어서 정치, 경제, 사회, 문화, 예술 등 다양한 분야에서 혁명을 불러일으킴
- 빅데이터의 정의는 이전에는 규모와 기술만을 고려했으나, 현재는 빅데이터의 활용과 이를 통해 창출되는 가치를 고려하는 방법적 정의까지 포함하도록 변화
 - 규모적 정의: 전통적인 방법으로 저장, 관리 분석할 수 있는 범위를 초과하는 규모의 데이터
 - 기술적 정의: 다양한 종류의 데이터로부터 새로운 가치를 추출하고, 수집, 탐색, 분석을 지원하도록 고안된 차세대 기술
 - 방법적 정의: 빅데이터는 거대한 데이터 자체만을 지칭했으나, 현재는 관련 도구, 플랫폼, 분석기법, 활용, 창출되는 가치까지 포괄하는 의미로 변화

1. Volume: zettabyte의 시대



■ Zettabyte?

WHAT'S A ZETTABYTE?

1 kilobyte	1,000,000,000,000,000,000
1 megabyte	1,000,000,000,000,000,000,000
1 gigabyte	1,000,000,000,000,000,000,000,000
1 terabyte	1,000,000,000,000,000,000,000,000,000
1 petabyte	1,000,000,000,000,000,000,000,000,000,000
1 exabyte	1,000,000,000,000,000,000,000,000,000,000,000
1 zettabyte	1,000,000,000,000,000,000,000,000,000,000,000,000

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice

Gigabyte : 3 Semi trucks

Terabyte : 2 Container Ships

Petabyte : Blankets Manhattan

Exabyte : Blankets west coast states

Zettabyte : Fills the Pacific Ocean

Yottabyte : A EARTH SIZE RICE BALL!

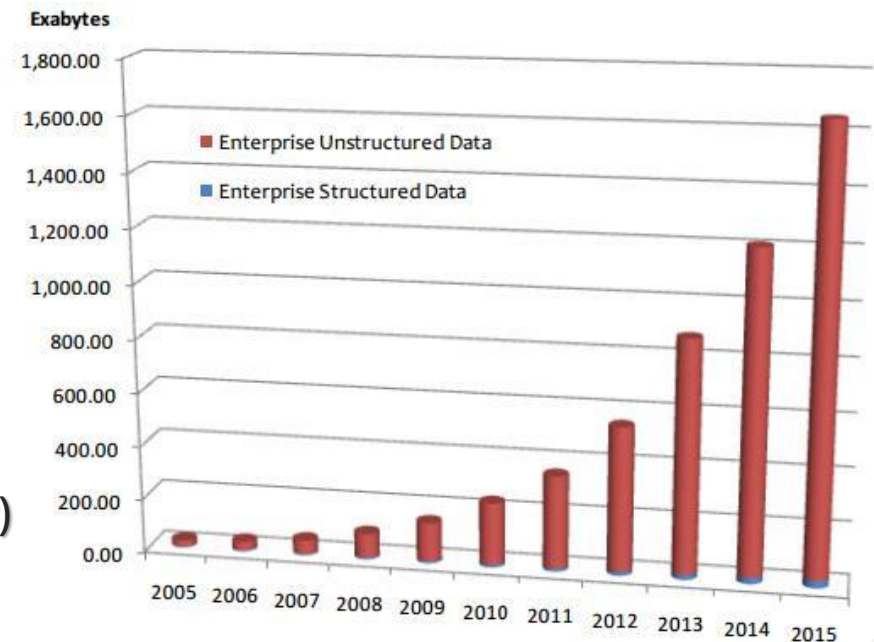


Graphic by Karl Tate

TechNewsDaily

2. Variety: 데이터의 종류 증가

- 정형화된(structured) 데이터
 - 일정한 규칙을 가지고 체계적으로 정리된 데이터
 - 통계청의 통계자료, 과학적 데이터 등
 - 데이터베이스, 엑셀
- 반정형화된(semi-structured) 데이터
 - 표나 그림도 포함하지만 일반적으로 문자로 서술된 데이터
 - HTML, XML, 한글 문서, 워드 문서 등
 - 데이터 처리 기술이 요구됨 (난이도 중)
- 비정형화된(unstructured) 데이터
 - 형식이 정해지지 않은 데이터
 - 개인이 SNS로 교류되는 정보
 - 카카오톡, 페이스북, 인스타그램 등
 - 정형화된 데이터 형태로 파싱이 요구됨 (난이도 상)

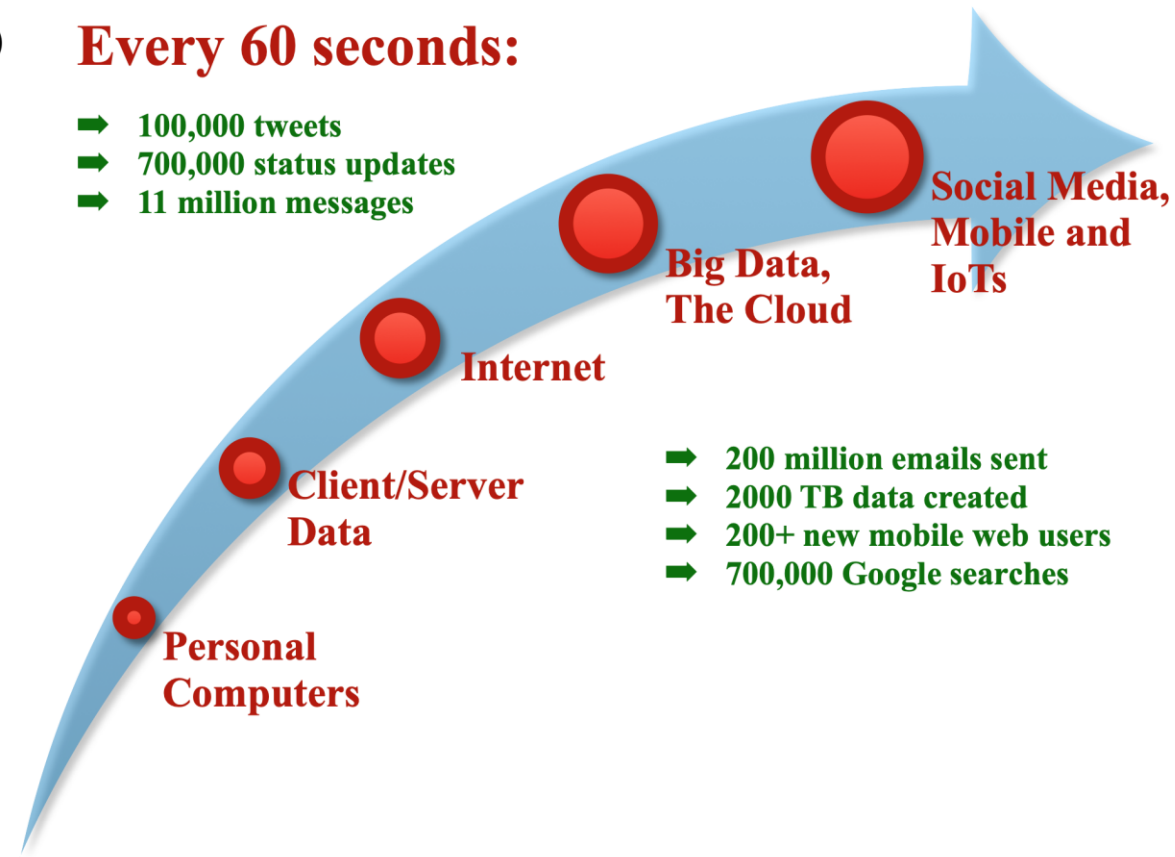


3. Velocity: 데이터 생성 속도의 증가, 실시간 분석 필요



Every 60 seconds:

- ➔ 100,000 tweets
- ➔ 700,000 status updates
- ➔ 11 million messages



- ➔ 200 million emails sent
- ➔ 2000 TB data created
- ➔ 200+ new mobile web users
- ➔ 700,000 Google searches

4. Veracity: 데이터의 불확실성

Uncertain data/Inconsistent data

Min	Max	Mean	SD
4.3	?	5.84	0.83
2.0	4.4	3.05	50000000
15000	7.9	1.20	0.43
0.1	2.5	?	0.76

Uncertainty and inconsistencies in the data

5. Value: 유의미한 정보 추출, 가치 창출

The Value of Big Data in Service-Oriented Organizations

Over 75% of organizations in professional services; engineering, construction, and operations; cargo transportation and logistics; passenger travel and leisure; media; sports and entertainment; and telecommunications indicate:

Big Data is critical to meet strategic objectives.



Sharing insights is a must-have capability for businesses.



Big Data will amplify other technology innovations.



Nearly **67%** have started to use Big Data in specific cases

... but only **6%** consider themselves mature.

62%

view data quality and huge data volumes as major challenges.

01001010111010010
10100101000101010
01001010001011101
101010100010100010
10100100101000101

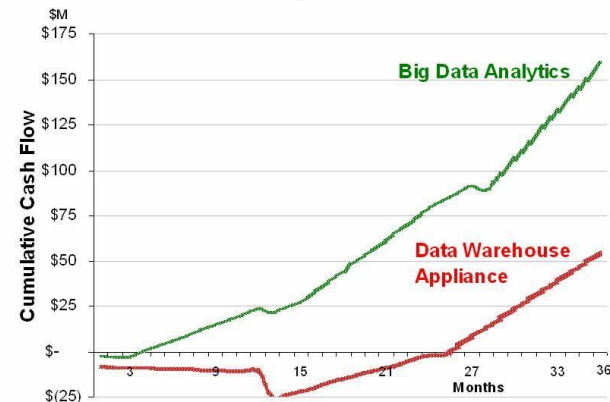
70%

of executives consider improved customer engagement and performance across lines of business as high priorities.



Value of Big Data Analytics

Accelerating Time-to-Value



- Projected cash flow data from actual business case analysis and real world experiences
- Breakeven 4 mos vs. 25
- 3-year Project Value \$158M vs. \$54M

© Wikibon 2011

March 1 Peer Incite:

Big Data Explosion - Is Your Data Warehouse a Dinosaur?

www.wikibon.org

■ 빅데이터 시장 현황

- 많은 국가에서 4차 산업혁명의 핵심으로 빅데이터와 AI를 차세대 산업으로 선정
- 세계적인 기업들도 빅데이터와 AI 핵심 기술과 인재를 확보하기 위해 노력
- 빅데이터와 AI를 도입한 기업의 만족도가 높음



◇ 10개 중 9개 기업이 빅데이터 사용을 통한 사업 성과에 만족하는 것으로 나타났다.
[인포그래픽=엑센츄어]

■ 빅데이터 시장 규모의 증가

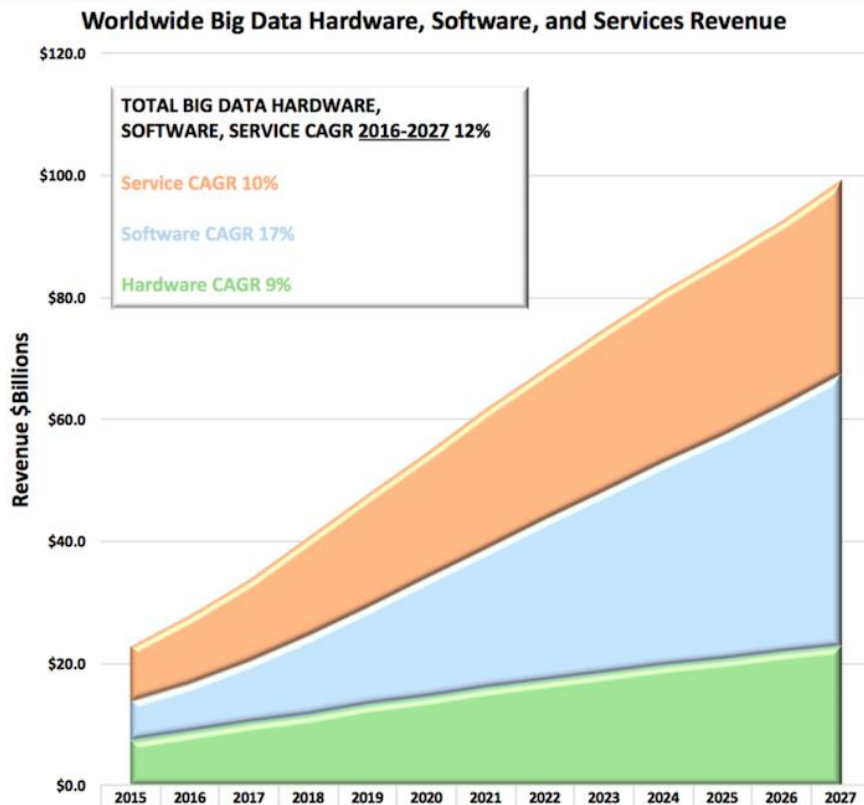
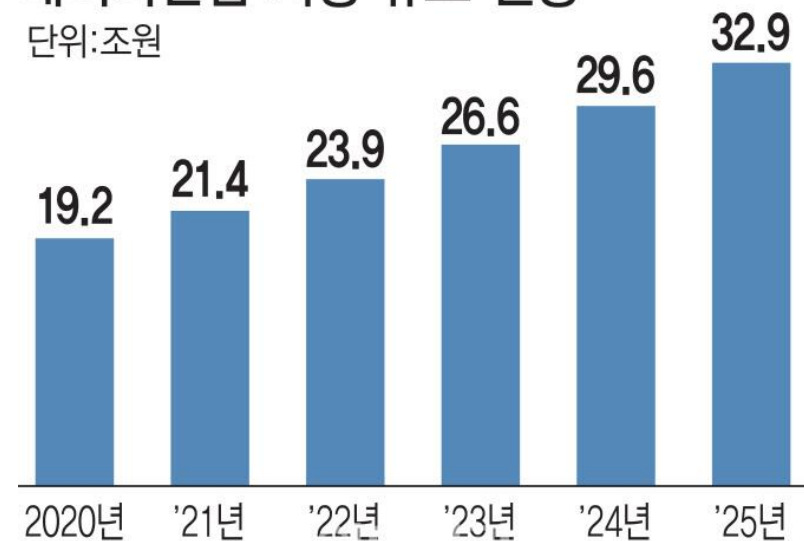


Figure 1. Worldwide Big Data Hardware, Software, and Services Revenue

데이터산업 시장 규모 전망

단위:조원



자료:한국데이터산업진흥원

■ 자동차

- 운전자의 센서 데이터를 분석하여 운전 배터리, 운전 스타일, 운전 기록, 이산화탄소 배출량 등 다양한 정보를 제공
- 운전 환경 데이터를 분석하여 다양한 환경에 대한 차량의 반응 연구
- 자동차의 진단 데이터를 수집하여 불량 파악, 고객의 보상 정확도, 고객의 요구 사항 파악, 새로운 차량에 반영

■ 스포츠

- 경기의 빅데이터를 활용하여 성적이 가장 좋은 파트너 선정
- 상대 선수에 따라 승률이 좋은 선수 파악
- 팬서비스 지원을 위한 각종 리포트

■ 은행

- 주요 고객의 신용카드, 대출을 파악하여 고객별 맞춤 서비스
- 콜센터 이력 정보 및 지점 방문 정보를 통해 온라인 앱이나 오프라인 판매점으로 상품 제안
- 빅데이터 분석 시스템으로 신용관리 및 손실 예측 처리 시간 단축

■ 카드사

- 고객의 카드 내역을 분석하여 고객에게 필요한 쿠폰 제안, 이를 선택하면 자동 혜택 적용
- 빅데이터 기반 고객 니즈를 분석하여 가맹점의 마케팅에 활용

■ 홈쇼핑

- 과거의 구매 이력을 활용하여 특정한 날에 상품 추천
- 고객이 방문하는 페이지 간 연관성을 계산하여 고객 행동 분석, 상품 추천

■ 유통

- 물류 빅데이터를 이용하여 재고관리 및 수요 예측 시스템을 적용
- 이를 이용하여 예약, 실시간 배송, 3시간 배송, 새벽 배송 등 다양한 서비스 품질 고도화
- 수요 예측으로 상품을 미리 창고에 구비하고 소비자에게 제시 (아마존, 알리바바)

■ 의료

- 미숙아 실시간 모니터링 빅데이터를 이용하여 의료진보다 24시간 전에 감염 사실을 발견 (온타리오 공과대)
- 독감 유행 예측 서비스, 중환자의 심실부 정맥 예측 서비스 (서울아산병원)
- 입원 병상 배정 최적화 시스템으로 입원환자의 대기시간, 장기 입원일 감소
- 에스토니아 국민의 DNA를 저장하고 분석하여 질병 가능성을 예측하고 맞춤형 진료에 사용

■ 공공

- 골목 상권 분석을 통해 영세 소상공인을 위한 창업위험도 예측 서비스 제공 (서울시)
- 유동인구 밀집도 분석을 통해 심야버스 노선 최적화 및 배차 간격 조정 시행 (서울시)
- 저축은행의 여신관련 통합 정보를 이용하여 불법 및 부실 혐의 여신을 사전에 탐지 (금융감독원)
- 조류독감 발병, 축산차량 이동 등의 데이터를 활용하여 조류독감 확산 예측 모델 개발 (농림축산검역본부)
- 지진발생 시 행동패턴, 이동패턴, 피해, 대피소 등의 데이터를 이용하여 시설/구호 제공 방안 도출 (국립재난안전연구원)

- 빅데이터가 왜 지금 이슈가 되는가?
 - 모바일 기기의 발전
 - 클라우드 서비스
 - 소셜 미디어의 보급
 - 개인의 정보와 밀접한 빅데이터가 발생
 - 기업은 방대한 빅데이터 자산 축적
 - 이는 새로운 기회와 경제적인 이익을 창출
- 빅데이터는 21세기의 원유

- 빅데이터는 기존 데이터베이스 시스템의 처리 용량을 초과하는 데이터이다. 데이터가 너무 크거나 빠르면 데이터베이스 구조에 맞지 않는다. 이 데이터에서 가치를 얻으려면 다른 처리 방법을 선택해야 한다. (O'Reilly)
- 빅데이터는 volume, speed, complexity 세가지 특징을 가진 데이터이다. (IBM)
- 2003년까지 인류가 쌓아 올린 데이터가 5엑사 바이트 수준인데, 이제는 단 하루만에 그 정도의 분량을 쏟아내는 시대가 되었다. (구글의 전 CEO, 에릭 슈미트)
- 데이터는 자본이나 노동력과 거의 동등한 레벨의 경제적 투입자본으로 비즈니스의 새로운 원자재 역할 수행한다. (Economist, 2010)
- 빅데이터는 미래 경쟁력을 좌우하는 21세기의 원유. 기업은 다가올 데이터 경제시대를 이해하고 정보 고립을 경계해야 생존할 수 있다. (Gartner, 2011)

- 데이터는 화폐나 금처럼 새로운 자산이 될 것이다. (Davos Forum)
- 빅데이터는 혁신, 경쟁과 생산성에 있어서 차세대 첨단 주자 (McKinsey, 2011)
- 데이터 분석을 잘 활용하는 조직일수록 차별적 경쟁력을 갖추고 높은 성과를 창출 (MIT Sloan, 2010)