

IT & Future Strategy

新가치창출 엔진, 빅 데이터의 새로운 가능성과 대응 전략

제18호 (2011. 12. 30)

목 차

- I. IT패러다임 변화와 빅 데이터의 의미
- II. 빅 데이터가 왜 대세인가?
- III. 빅 데이터의 특성과 미래사회에서의 역할
- IV. 빅 데이터 시대의 준비 과제

‘IT & Future Strategy’는 21세기 한국 사회의 주요 패러다임 변화를 분석하고 이를 토대로 미래 정보사회의 주요 이슈를 전망, IT를 통한 해결 방안을 모색하기 위해 NIA에서 기획·발간하는 보고서입니다.

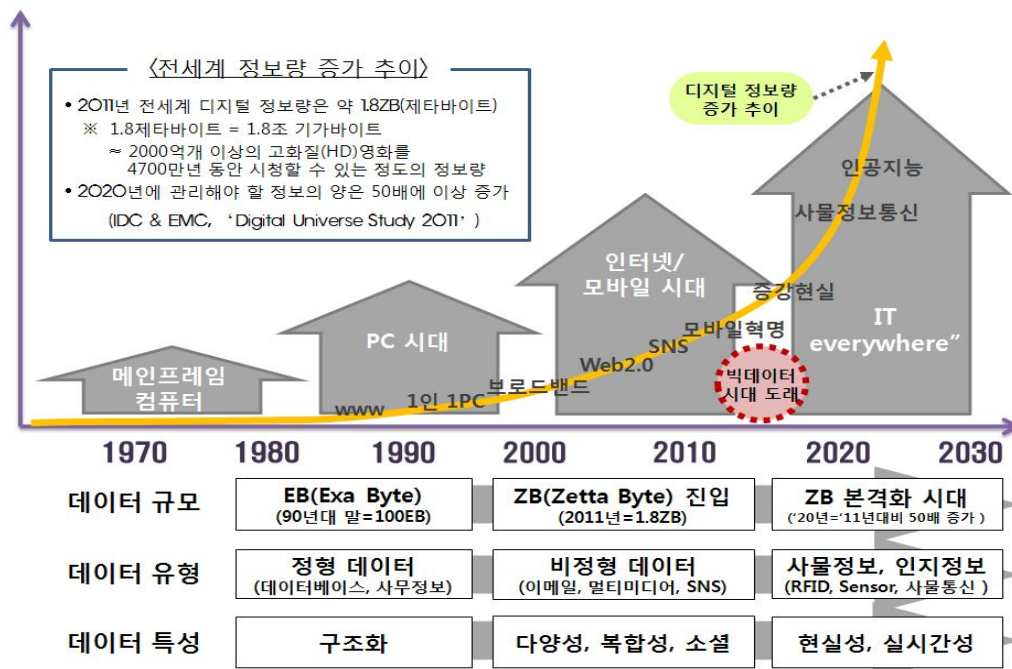
▶ 발행인 : 김성태

▶ 작 성 : 한국정보화진흥원 국가정보화기획단 정보화전략연구부
정지선 책임연구원 (02-2131-0533, jjs@nia.or.kr)

▶ 보고서 온라인 서비스 : www.nia.or.kr, www.itglobal.or.kr

요약

- ◇ IT의 일상화가 이뤄지는 스마트시대에는 소셜, 사물, 라이프로그 데이터 등이 결합되며 '빅 데이터'의 영향력이 증대
 - 데이터는 정보사회를 움직이는 핵심 연료인 만큼, 빅 데이터로의 발전은 정보사회의 패러다임을 견인할 정도의 큰 힘을 발휘
- ◇ 5년 이내에 데이터의 폭발적 증가로 인한 혼돈과 잠재적인 가능성이 공존하는 '빅 데이터(Big Data) 시대'가 도래



- ◇ 최근 글로벌 리서치 기관, 컨설팅 그룹 등은 차세대 키워드로 '빅 데이터'를 선정하고 경제적 가치에 주목

차세대 이슈로 빅 데이터가 떠오르는 이유

- ① ICT 주도권이 데이터로 이동
- ② 공간, 시간, 관계, 세상을 담는 데이터
- ③ 미래 경쟁력과 가치창출의 원천

◇ 빅 데이터는 미래사회의 가치창출 엔진

- 빅 데이터는 불확실성, 리스크, 스마트, 융합 등 미래사회의 특성에 대응하는 역할을 수행하며 기회요인을 창출하는 핵심 엔진으로 작용

< 미래사회의 특성과 빅 데이터의 역할 >

미래사회 특성	빅 데이터의 역할	
불확실성	→ 통찰력	<ul style="list-style-type: none"> • 사회현상, 현실세계의 데이터를 기반으로 한 패턴분석과 미래전망 • 여러 가지 가능성에 대한 시나리오 시뮬레이션 • 다각적인 상황이 고려된 통찰력을 제시 • 다수의 시나리오로 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 투명성 제고 및 낭비요소 절감
스마트	→ 경쟁력	<ul style="list-style-type: none"> • 대규모 데이터 분석을 통한 상황인지, 인공지능 서비스 등 가능 • 개인화, 지능화 서비스 제공 확대 • 소셜(니즈)분석, 평가, 신용, 평판 분석을 통해 최적의 선택 지원 • 트렌드 변화 분석을 통한 제품 경쟁력 확보
융합	→ 창조력	<ul style="list-style-type: none"> • 타분야와의 결합을 통한 새로운 가치창출(의료 정보, 자동차정보, 건물정보, 환경정보 등) • 인과관계, 상관관계가 복잡한 컨버전스 분야의 데이터 분석으로 안전성 향상, 시행착오 최소화 • 방대한 데이터 활용을 통한 새로운 융합시장 창출

◇ 성공적인 빅 데이터 활용을 위해서는 ‘연결과 협력’, ‘창의적 인력’, ‘신뢰 환경’ 등의 3가지 과제의 선행이 필수적

- ① 데이터 경제 시대를 대비하는 연결과 협력 : 공공, 민간 부문이 통합된 데이터 분석을 위한 플랫폼 필요
- ② 빅 데이터의 핵심 역량 ‘창의적 인력’의 양성 : 다학제적 이해와 통합적 사고, 직관력 등을 갖춘 데이터 사이언티스트의 양성
- ③ 데이터 신뢰 환경의 구축 : 개인 프라이버시를 위한 기법의 도입과 데이터 자원의 결합과 협력 촉진을 위한 신뢰기반 형성

I

IT패러다임 변화와 빅 데이터의 의미

□ 데이터의 발전과 IT패러다임 변화

- IT의 일상화가 이뤄지는 스마트시대에는 소셜, 사물, 라이프로그¹⁾ 데이터 등이 결합되며 ‘빅 데이터’의 영향력이 증대
 - 실시간 연결과 소통의 ‘스마트 혁명’은 데이터 폭증을 발생시켰고 기존의 데이터 저장·관리·분석기법은 한계와 도전에 직면
 - 데이터는 정보사회를 움직이는 핵심 연료인 만큼, 빅 데이터로의 환경변화는 정보사회의 패러다임을 견인할 정도의 큰 힘을 발휘
- 지능화, 개인화 등 스마트시대 주요 패러다임 선도를 위해서는 빅 데이터의 활용이 핵심이며, 그 수준이 경쟁력과 성패를 좌우
 - 빅 데이터의 가공과 분석에 따라 상황인식, 문제해결, 미래전망이 가능해지고, 데이터가 경제적 자산과 경쟁력의 척도로 부각

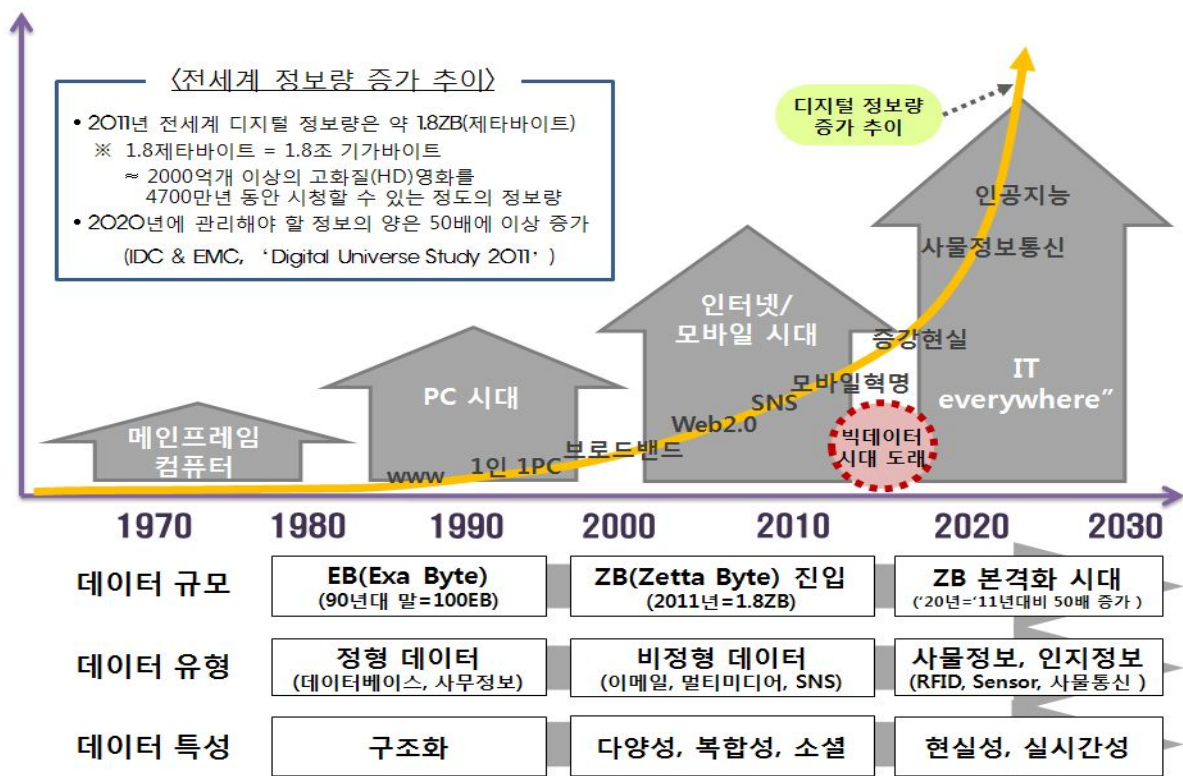
< 정보사회의 패러다임 변화와 힘의 이동 >

	PC시대	인터넷시대	모바일시대	스마트시대
패러다임 변화	디지털화, 전산화	온라인화, 정보화	소셜화, 모바일화	지능화, 개인화, 사물정보화
IT 이슈	PC, PC통신, 데이터베이스	초고속 인터넷, www, 웹서버	모바일 인터넷, 스마트폰	빅데이터 차세대PC, 사물네트워크(M2M)
핵심분야 (서비스)	PC, OS	포털, 검색엔진 Web2.0	스마트폰, 앱서비스, SNS	미래전망 상황인식, 개인맞춤형 서비스
대표 기업	MS, IBM 등	구글, 네이버, 유튜브 등	애플, 페이스북, 트위터 등	?
IT비전	1인 1PC	클릭 e-Korea	손안의 PC, 소통	IT everywhere, 新가치창출

1) 라이프로그(life-log) : 개인들의 일상생활의 기록

- o 5년 이내에 데이터의 폭발적 증가로 인한 혼돈과 잠재적인 가능성이 공존하는 '빅 데이터(Big Data) 시대'가 도래
 - 스마트 단말 확산, SNS 활성화, 사물네트워크(M2M) 확산으로 데이터 폭발이 더욱 가속화 되며 점차 빅 데이터 기반이 확대
 - 빅 데이터를 위한 고급분석 등 관련 기술은 현재 기술 발생단계(Technology Trigger)이며 향후 2~5년 후에 성숙될 것으로 평가²⁾

< ICT 발전에 따른 데이터의 변화 방향 >



< 데이터의 대폭발 >

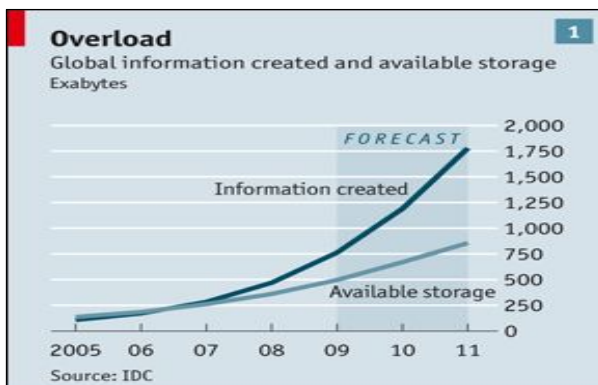
- ‘10년~’15년까지 전세계 모바일 데이터 트래픽은 연평균 92%, 인터넷 트래픽은 연평균 34% 증가할 것으로 예상(Cisco, 2011)
- 트위터(twitter)는 전 세계 1억명의 월간 이용자(active user)들이 이용하고 있으며, 하루 평균 2억 개의 트윗이 발생(Twitter, 2011)
- 오늘날 11억 인구가 소셜 네트워크를 이용하고 있고 2억 5000만명이 매일 페이스북에 사진을 업로드하고 있다(인텔 CEO 폴 오텔리니, 2011)

2) Gartner(2011), ‘Hype Cycle for Emerging Technologies, 2011’

□ 제타바이트 시대, '빅 데이터'가 화두

- 인터넷이 일상화된 최근 10년 사이, 인류는 디지털 데이터가 폭증하는 데이터 홍수(Data Deluge)³⁾ 현상에 직면
 - '07년부터 전 세계적으로 생성된 디지털 정보량이 사용가능한 저장 공간(available storage)을 초과하기 시작(Economist, 2010)
 - '11년, 전 세계 데이터에 생성될 디지털 정보량이 1.8ZB(제타바이트)에 달하는 '제타바이트 시대'로 진입⁴⁾
 - 정보량은 기하급수적으로 증가하여 2020년에는 관리해야할 정보량이 50배 급증하고, 10배 많은 서버가 필요할 것으로 전망
- ※ 1ZB(제타바이트)는 1조 GB(기가바이트)에 해당하는 양으로 美의회도서관 저장정보(235 테라바이트, '11.4월 현재)의 4백만 배에 해당

< 데이터 과부하 현상 >



※ 출처 : IDC, Economist(2010) 재인용

< 전 세계 데이터 증가 추세 >



※ 그림출처 : 중앙일보(2011)

- 디지털 정보량의 기하급수적인 증가에 따라 대규모 데이터가 중대 이슈로 부각하며 '빅 데이터(Big Data)'라는 용어가 등장
 - 최근 기술발전에 따른 데이터 저장 및 처리 비용의 하락, 소셜 네트워크 서비스 확대 등으로 막대한 데이터 폭발이 진행 중
 - 앞으로는 도로, 건축물 등에 내장된 센서 및 임베디드 시스템(embedded system)에서 막대한 데이터가 만들어질 것으로 전망

3) Economist(2010), 'The data deluge'

4) IDC(2011), 'Digital Universe study'

- ‘빅 데이터’란 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합을 지칭(SERI, 2010)
 - 대규모 데이터와 관계된 기술 및 도구(수집 · 저장 · 검색 · 공유 · 분석 · 시각화 등)도 빅 데이터의 범주에 포함
 - 과거 빅 데이터는 천문 · 항공 · 우주 정보, 인간게놈 정보 등 특수 분야에 한정됐으나 ICT의 발달에 따라 전분야로 확산
- 빅 데이터의 정의는 데이터 규모와 기술 측면에서 출발했으나, 빅 데이터의 가치와 활용효과 측면으로 의미가 확대되는 추세
 - 빅 데이터는 고객정보와 같은 정형화된 자산정보(내부) 뿐만 아니라 외부데이터, 비정형, 소셜, 실시간데이터 등이 복합적으로 구성
 - 빅 데이터는 규모, 다양성, 복잡성, 속도의 증가 특성을 갖고 있으며, 4개의 요소가 충족될수록 빅 데이터에 적합
 - 특정 규모(big volume) 이상을 빅 데이터로 칭하기 보다는 원하는 가치(big value)를 얻을 수 있는 정도로 상대적인 해석이 가능

< 빅 데이터의 4가지 구성 요소 >5)

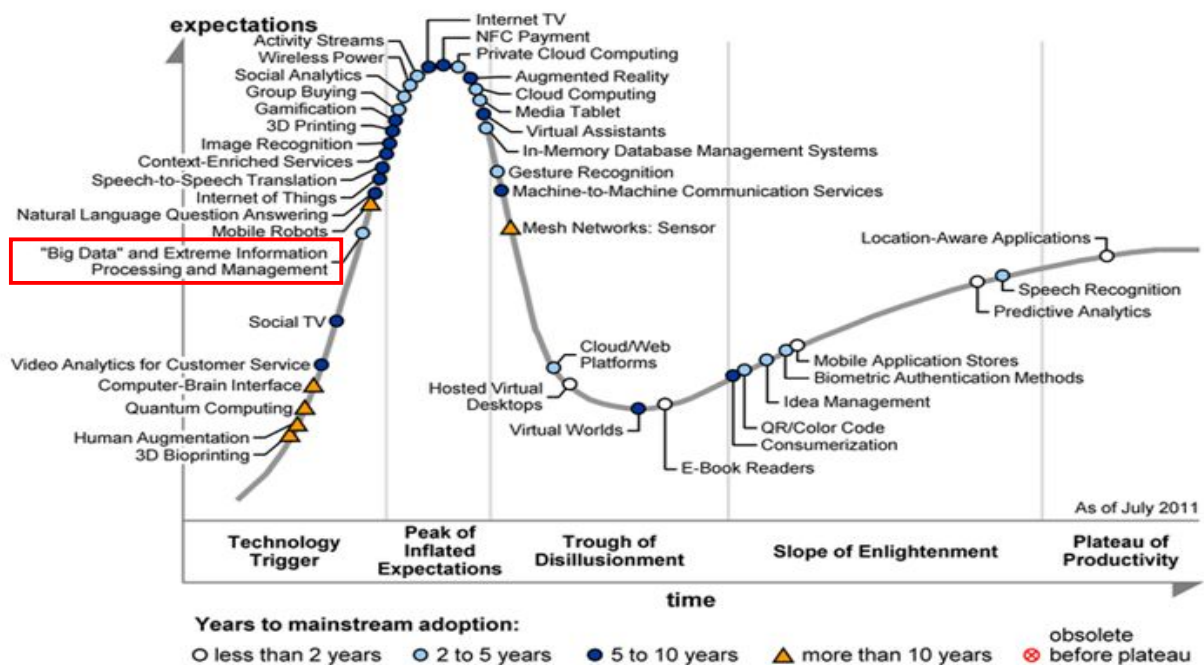
구 분	주요 내용
규모(Volume)의 증가	• 기술적인 발전과 IT의 일상화가 진행되면서 해마다 디지털 정보량이 기하급수적으로 폭증 ⇒ 제타바이트(ZB) 시대로 진입
다양성(Variety) 증가	• 로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터 종류의 증가 • 텍스트 이외의 멀티미디어 등 비정형화된 데이터 유형의 다양화
복잡성(Complexity) 증가	• 구조화되지 않은 데이터, 데이터 저장방식의 차이, 중복성 문제 등 • 데이터 종류의 확대, 외부 데이터의 활용으로 관리대상의 증가 • 데이터 관리 및 처리의 복잡성이 심화되고 새로운 기법 요구
속도(Velocity) 증가	• 사물정보(센서, 모니터링), 스트리밍 정보 등 실시간성 정보 증가 • 실시간성으로 인한 데이터 생성, 이동(유통) 속도의 증가 • 대규모 데이터 처리 및 가치 있는 현재정보(실시간) 활용을 위해 데이터 처리 및 분석 속도가 중요

5) Gartner(2011)가 제시한 빅 데이터의 4가지 구성 요소를 참고하여 정리함

□ IT기업들의 '빅 데이터' 시장 준비 현황

- 최근 글로벌 리서치 기관, 컨설팅 그룹 등은 차세대 키워드로 '빅 데이터'를 선정하고 경제적 가치에 주목
 - 가트너는 2011년 이머징 기술 전망에서 '빅 데이터'를 새롭게 포함시키고, 앞으로 주목해야할 기술로 소개⁶⁾
 - 빅 데이터는 현재 기술 발생단계(Technology Trigger)⁷⁾이며 향후 2~5년 후에 성숙될 것으로 평가
 - 맥킨지(McKinsey)는 비즈니스 지형을 바꿀 10가지 기술 트렌드 중 하나로 빅 데이터를 선정⁸⁾
 - 빅 데이터를 수집, 저장하고 이를 토대로 새로운 정보를 찾아내는 것이 경제성장을 위한 중요한 가치창출 효과를 가져 온다고 분석

< 2011년도 가트너의 이머징 기술 하이프 사이클(Hype Cycle) >



※ 출처 : Gartner(2011), 'Hype Cycle for Emerging Technologies, 2011'

6) Gartner(2011), 'Hype Cycle for Emerging Technologies, 2011'

7) Technology Trigger : 신기술, 신제품 출시 및 새로운 이벤트가 산업관계자들의 흥미를 유발하는 단계

8) McKinsey&Company(2010), Clouds, big data, and smart assets : Ten tech-enabled business trends to watch

- 국내외 IT기업들은 빅 데이터 시장을 선점하고 주도권을 잡기위해 데이터 분석 중심으로 조직을 재편하고 역량 강화 및 기술을 개발
- 빅 데이터 정보수집과 분석을 위한 대용량 데이터 처리 기술과 솔루션을 중심으로 빅 데이터 시장에 대비

< 글로벌 IT기업들의 빅 데이터 준비 현황 >

회사명	내용
EMC	<ul style="list-style-type: none"> • 데이터 저장부터 관리, 분석까지 빅 데이터와 관한 모든 것을 제공하기 위해 그린플럼, 아이실론 등 빅 데이터 솔루션 업체 및 데이터 관련 다수업체 인수 • 빅 데이터 스토리지 솔루션(아이실론, 아트모스), 콘텐츠 관리 솔루션(다큐멘텀) 등 제공 • 'EMC 애널리틱스 랩'을 운영하여 데이터 사이언티스트(Data Scientist) 육성
IBM	<ul style="list-style-type: none"> • 지난 5년간 140억 달러 이상을 투자하여 비즈니스 분석 관련업체 인수 <ul style="list-style-type: none"> - 분석용 데이터 저장관리 업체(네티자), 데이터 통합 업체(에센셜), 분석 솔루션 업체(코그너스) 등 • 빅 데이터 솔루션 : InfoSphere BigInsight(Hadoop), InfoSphere Streams • 지속가능한 지구를 만들기 위해 지구 데이터(기온, 토양상태, 교통 흐름 등)를 분석하는 '스마트 플래닛(smart planet)' 프로젝트 전개
오라클	<ul style="list-style-type: none"> • 세계적인 DB 업체, '하이퍼리온社'를 인수로 분석기술 확보 • 오라클 빅 데이터 어플라이언스 제품 출시
SAP	<ul style="list-style-type: none"> • 업무용 어플리케이션 업체에서 최근 DB 전문업체로 변신 • 메모리 기반 DB 어플라이언스(HANA) 제시 • BI 소프트웨어, 플랫폼을 제공하는 '비즈니스 오브젝트社' 인수
테라데이터	<ul style="list-style-type: none"> • 데이터웨어하우징 및 비즈니스 인텔리전스(BI) 전문 업체 • 비정형 데이터의 고급분석·관리 솔루션 업체 인수(애스터데이터) • '애스터 맵리듀스 플랫폼' 제시
HP	<ul style="list-style-type: none"> • BI 솔루션 업체 '버티카', 기업용 검색엔진 업체 '오토노미' 인수 • 버티카와 오토노미를 결합하여 빅 데이터 분석 시장에 진입 • '인스턴트-온 엔터프라이즈(Instant-On Enterprise)' 솔루션으로 기업경영 의사결정, 경영정보 분석 등 경영지원 전략 수립 서비스 제공
마이크로 스트레티지	<ul style="list-style-type: none"> • 비즈니스 인텔리전스(BI) 소프트웨어 공급업체 • BI(Business Intelligence)에 빅 데이터 분석 처리를 접목하여 사업 역량 강화
구글	<ul style="list-style-type: none"> • 대용량 데이터 처리 기술 발표 : GFS(Google File System, 2003년), MapReduce(2004년), Sawzall(2005년), Bigtable(2006년) • '빅쿼리(Big Query) 서비스 공개(2011년) : 이용자(기업 등)가 업로드한 거대한 양의 데이터 분석 처리를 지원하는 서비스
MS	<ul style="list-style-type: none"> • 윈도우 애저(Windows Azure)와 윈도우 서버 플랫폼용 아파치 하둡 개발 계획 • 하둡(Hadoop)⁹⁾ 기술 전문업체 '호튼웍스'와 협력
넥스알(국내)	<ul style="list-style-type: none"> • 빅 데이터 관리 솔루션(넥스알 빅 데이터 어널리틱스 플랫폼) 출시

※ 출처 : 디지털데일리(2011. 11.), '빅 데이터 전쟁의 승자는 누가될까', 아이뉴스24 (2011. 12), '글로벌 IT기업들 한국 빅 데이터 시장을 잡아라' 외 재구성

9) 하둡(Hadoop) : 대용량 데이터를 처리하기 위해 대규모 분산처리를 지원하는 오픈소스 방식의 프레임워크

II

왜 '빅 데이터'가 대세인가?

□ 차세대 이슈, '빅 데이터(Big Data)!'

- ICT 전체를 흐르는 데이터의 변화·발전은 IT와 비즈니스 영역에 새로운 공급체제와 수요를 창출하는 원동력으로 작용
 - ICT 인프라의 성숙에 따른 신규 투자처에 대한 니즈 증가로 새로운 공급체제와 시장을 창출하는 빅 데이터에 대한 관심 증가
 - 기업은 그간 축적된 자산 데이터(asset data)와 소셜 네트워크에 기반한 대규모 외부 데이터의 활용으로 혁신과 기회 포착을 기대
- 인적 관계 데이터, 위치, 행태, 인식 등과 같이 사람과 사회현상을 이해할 수 있는 비정형의 대규모 데이터의 증가
 - 지극히 개인적인 부분부터 사회적 논의까지 이뤄지는 소셜 네트워크는 여론과 현안을 파악할 수 있는 소통의 장소
 - 표출된 감정, 인식, 경험의 수집과 분석이 가능해져 전 사회 분야에서 소셜 데이터를 포함한 빅 데이터 분석 활용도가 증대
- 정보(지식)가 중요한 자산인 정보시대에는 대규모 데이터에서 의미를 찾고 정보와 지식을 만들어내는 능력이 경쟁력
 - 사용자 참여와 정보 공유가 늘어날수록 지식의 가공이 중요해지며 데이터 분석 능력이 경쟁자들과 가장 큰 차별화 요소로 부상



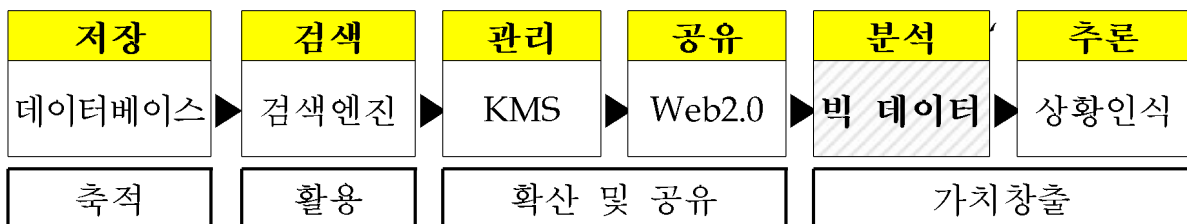
차세대 이슈로 빅 데이터가 떠오르는 이유

- ① ICT 주도권이 데이터로 이동
- ② 공간, 시간, 관계, 세상을 담는 데이터
- ③ 미래 경쟁력과 가치창출의 원천

① ICT 주도권이 데이터로 이동

- 모바일, 클라우드, SNS의 등장으로 근본적인 인프라와 데이터의 변화가 일어나며 데이터를 바라보는 시각이 변화
 - IT의 주도권이 인프라, 기술, SW 등에서 데이터로 전이되고 있으며 데이터가 IT에서 분리된 독립적인 주체로 발전
 - 데이터 폭증에 대한 대응과 데이터 분석이 ICT의 중요 이슈로 부각되며 빅 데이터가 ICT 시장과 기술발전의 핵심 주제로 인식
- 축적과 공유를 통해 유의미하게 분석할 수 있는 데이터 자원이 쌓이자 (빅 데이터) 데이터의 역할은 '분석과 추론(전망)'의 방향으로 진화
 - 빅 데이터는 사회 환경변화를 신속하게 감지하고 대응하는 역량, 스마트한 지능형/개인화 서비스의 창출하는 원천요소로 작용

< 데이터의 과거-현재-미래 >



- 스마트 시대에는 데이터의 저장-검색-관리-공유-분석-추론의 전체적인 과정이 업그레이드되며 정보화 시대와 차별적으로 발전

< 정보화 시대 vs. 스마트 시대의 데이터 관련 이슈 변화 >

구 분	정보화 시대(1세대)	스마트 시대(2세대)
저장	관계형/정형 데이터베이스, 데이터웨어하우스	비관계형/비정형 데이터베이스, 가상화, 클라우드 서비스
검색	검색엔진(text), 포털 서비스	자연어/음성·영상/시맨틱 검색서비스
관리·공유	KMS, Web 2.0	플랫폼, 소셜 네트워크, 집단지성
분석	경영정보/고객정보/자산정보 분석 (ERP, CRM, 데이터마이닝 등)	빅 데이터 분석 (소셜 분석, 고급 분석, BI, 시각화)
추론	-	상황인식 서비스(미래전망, 사전대응, 자동화 서비스), 개인화 서비스

- 빅 데이터는 대용량 데이터, 비정형화된 데이터의 수집, 검색, 데이터 전처리 및 분석 기술, 시각화 기술 등이 중요

< 빅 데이터와 연계된 기술들 >

용어	설명
빅 테이블 (Big Table)	<ul style="list-style-type: none"> • 구글 파일 시스템 상에 구축된 상용 분산 데이터베이스 시스템 • H베이스에 영향을 미침
카산드라 (Cassandra)	<ul style="list-style-type: none"> • 분산 시스템에서 방대한 분량의 데이터를 처리할 수 있도록 디자인 된 오픈소스 데이터베이스 관리시스템 • 이 시스템은 원래 페이스북에서 개발했으며 지금은 아파치 소프트웨어 재단의 한 프로젝트로 관리되고 있음
데이터웨어하우스 및 분석 어플라이언스	<ul style="list-style-type: none"> • 데이터웨어하우스를 위해 서버, 스토리지, 운영체제, 데이터베이스, BI, 데이터마이닝 등 기타 여러 가지 소프트웨어가 최적화되어 설치된 통합제품
분산 시스템	<ul style="list-style-type: none"> • 동시에 일을 처리하기 위해 네트워크로 연결된 컴퓨터들의 집합 • 단일 또는 다수의 컴퓨터 리소스를 부분적으로 활용함으로써 시스템의 가격 대비 성능비, 안정성, 확장성을 향상시킬 수 있음
구글 파일 시스템	<ul style="list-style-type: none"> • 구글에서 개발한 분산파일 시스템, 하둡(Hadoop)과 관련 있음
하둡(Hadoop)	<ul style="list-style-type: none"> • 분산 시스템 상에서 대용량 데이터 처리 분석을 지원하는 오픈소스 소프트웨어 프레임워크 • 구글이 개발한 맵리듀스를 오픈소스로 구현한 결과물 • 야후에서 최초 개발되었으며 지금은 아파치 소프트웨어 재단의 한 프로젝트로 관리되고 있음
H베이스 (Hbase)	<ul style="list-style-type: none"> • 구글의 '빅 테이블'을 참고로 개발된 오픈소스 분산 비관계형 데이터베이스 • 파워셋에서 개발했으며 현재는 아파치 소프트웨어 재단에서 하둡(Hadoop)의 일환인 프로젝트로 관리되고 있음
맵리듀스 (MapReduce)	<ul style="list-style-type: none"> • 분산 시스템 상에서 대용량 데이터 세트를 처리하기 위해서 구글이 제안한 소프트웨어 프레임워크, 하둡(Hadoop)에도 구현되었음
비관계형 데이터베이스	<ul style="list-style-type: none"> • 비관계형 데이터베이스는 데이터를 테이블에 저장하지 않는 데이터베이스이며 관계형 데이터베이스와는 대조적인 개념 • 이를 사용하면 스키마 없는 엔티티(NoSQL)를 관리할 수 있음

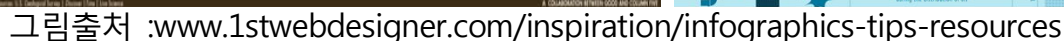
※ 출처 : IDC(2011), 아이뉴스24 재인용

< IT 분야의 새로운 주제, "Making Data Work" >

- 최근 IT 시장에서는 데이터를 IT의 한 분야가 아닌 IT에서 분리된 독립 분야임을 선언하며 데이터가 핵심요소로 급부상
- IT 두뇌들의 회의, 실리콘밸리 '스트라타(Strata)' 포럼은 '데이터를 가능하게 하라(Making Data Work)'를 올해의 회의 주제로 정하고, 3일간 60여개의 포럼에서 '데이터'를 키워드로 논의

출처 : 주간조선(2011.02.21), '데이터를 돈으로 바꿔라!'

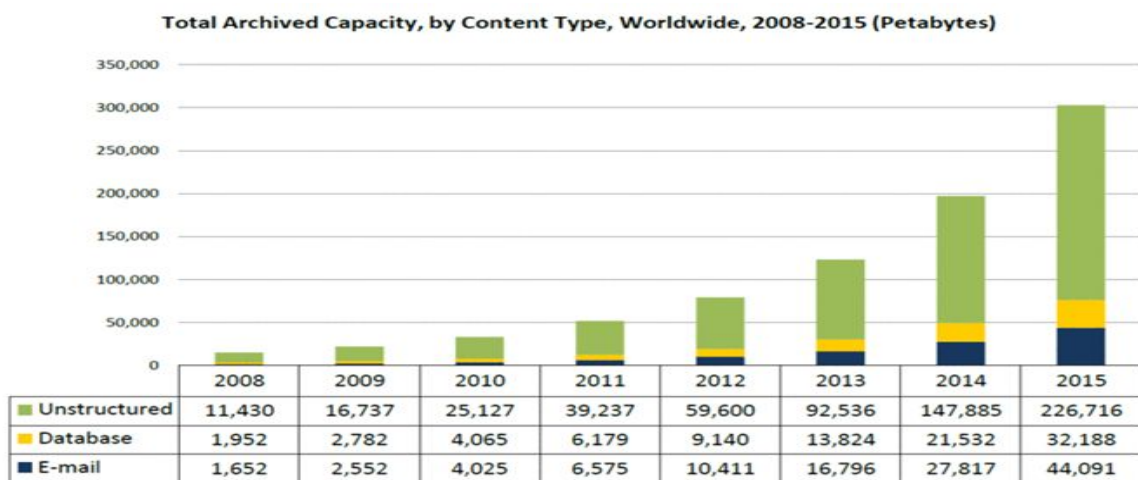
- ## < 정보 시각화(Infographics)의 예시 >

[illegible]

② 공간, 시간, 관계, 세상을 담는 데이터

- 2010년 이후 본격적인 스마트 모바일 확산과 함께 사용자들이 자발적으로 참여하고, 정보를 생성하는 소셜 데이터 혁명이 발생
 - 소셜 데이터 혁명은 정보의 생성자, 규모, 파급효과 등에서 1990년대에 기업들이 고객의 정보를 축적했던 정보혁명과 구분
 - 페이스북, 트위터 등 SNS 이용 확산과 커뮤니케이션 방식의 변화는 데이터 변혁을 야기하는 가장 중요한 요인
- 소셜 데이터는 세상을 이해할 수 있게 도움을 주며, 기업들은 고객과의 시장 공감의 맥을 찾는 쌍방향 소통 수단으로 활용
 - SNS를 통해 제공되는 정보는 지식정보와 함께 정서적인 공감에 바탕을 둔 감성적 정보가 큰 비중을 차지¹⁰⁾
 - SNS에서는 개인의 취향이 보다 직접적으로 반영되고, 진실성과 진정성, 관련성이 증가되어 데이터로서의 가치가 높음¹¹⁾
- 소셜 데이터 이외에도 이메일, 동영상 등과 같은 비정형 정보들이 향후 10년 동안 생성되는 전체 데이터의 90%에 달할 것으로 전망

< 정형-비정형 데이터 유형의 변화 >



※ 출처 : Enterprise Strategy Group(2010)

Source: Enterprise Strategy Group, 2010.

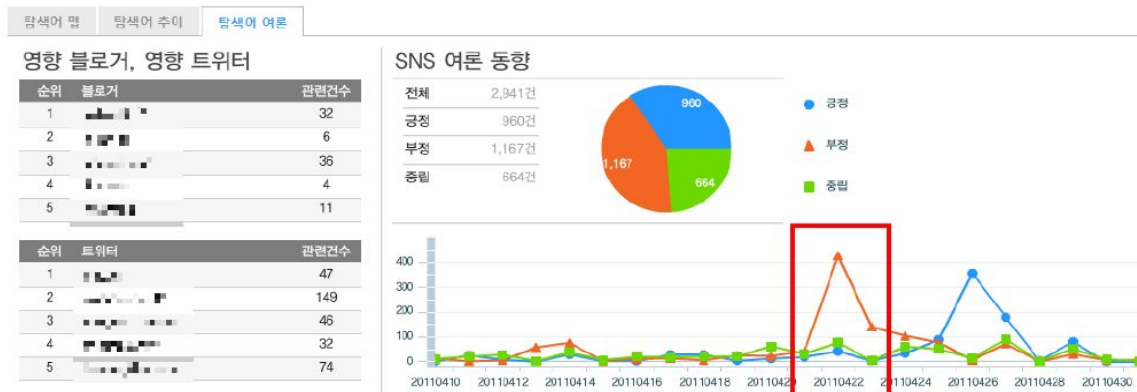
10) 11) 허지성(2011), '정보의 소셜화 시대, 공감의 맥을 찾아라, LG경제연구원 LG Business Insight, Weekly 포커스

- 최근 소셜 미디어에서 나타나는 행동·이용패턴을 분석하여 기업, 제품에 대한 인식 및 의견을 확인하기 위한 소셜 분석이 이슈
- 소셜 미디어는 실시간성과 가속성이라는 특징을 지녔기 때문에 어떠한 매체보다 이슈의 확산 속도가 빠름
- 국내는 '사이람', '다음소프트', '그루터'와 같은 소셜 분석 전문 업체들이 소셜 미디어 데이터를 기반으로 사회현상을 분석
- 소셜 분석은 마케팅 비용은 줄이면서 서비스 피드백을 받을 수 있고, 설문조사 방식과는 차별성이 있어 기업들이 관심을 갖기 시작

< 국내 소셜 분석 전문기업 현황 >

회사명	내용
사이람 (netminer.com)	소셜 네트워크 분석 소프트웨어 넷마이너(NetMiner) 개발 소셜 네트워크 분석 응용솔루션 및 컨설팅 제공
소셜메트릭스 (socialmetrics.co.kr)	SNS 정보 기반 여론 진단 서비스, 소셜미디어 트위터, 블로그 트렌드 주제 검색 및 동향 제공
그루터 (gruter.com)	소셜 네트워크 분석, 데이터 분석, 모니터링, 마케팅 성과측정 서비스 제공
트윗트렌드 (tweettrend.com)	트윗을 수집, 분석하여 트위터 트렌드 분석
트윗믹스 (tweetmix.net)	관심 키워드 모니터링, 관심 키워드 중심의 실시간 커뮤니케이션 (리트윗, 리플라이, 팔로잉 등), 키워드별 통계 정보 제공

< 소셜 데이터 분석의 사례 >

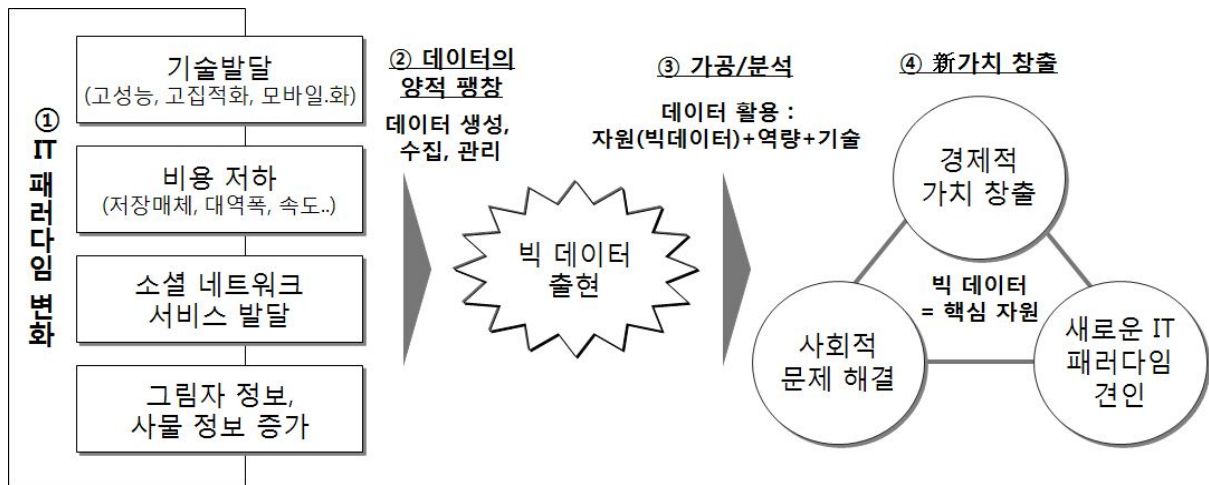


※ 출처 : 강학주(2011), 'Social Big Data & Collective Intelligence'

③ 미래 경쟁력과 가치창출의 원천

- 빅 데이터는 잠재적 가치와 잠재적 위험이 공존하며, 사회·경제적으로 성패를 좌우하는 핵심 원천이 될 것으로 평가
- IT패러다임의 변화는 데이터의 양적 팽창을 가져왔으며, 데이터 범람이 새로운 기회와 편익을 창출해내기 시작

< 빅 데이터의 출현과 新가치창출의 흐름 >



< 빅 데이터의 사회·경제적 의미¹²⁾>

구분	주요 내용
천연 자원 (Natural Resources)	<ul style="list-style-type: none"> • 데이터에 내포된 가치와 가능성에 대해 주목 • 사회적으로 현안과 위험을 해결할 수 있는 잠재력에 기대 • 새로운 경제적 가치의 원천으로 활용 ※ 새로운 원유, 데이터 골드러쉬, 데이터 금맥 찾기(data mining)
새로운 재난 (Natural Disasters)	<ul style="list-style-type: none"> • 정보의 범람으로 기회를 파악하기가 모호해지고, 규정 준수가 어려움 • 늘어나는 데이터로 인해 현 상태를 유지하는데 IT 예산이 사용되어 혁신을 위한 새로운 동력에 투자가 어려워짐 • 데이터 처리의 낮은 응답속도가 기업의 생산성 저하로까지 이어질 우려가 있음 ※ 데이터 토네이도(data tornado), 데이터 홍수(data deluge)
산업적 도구 (Industrial Devices)	<ul style="list-style-type: none"> • 데이터 효율적인 관리와 분석을 통해 기업의 경쟁 우위 확보 • 데이터를 신속하게 처리해 실시간 의사결정에 지원 • 데이터 분석 역량이 기업 경쟁력을 좌우 ※ 데이터 산업혁명(Industrial Revolution)

12) Tyler Bell(2011), 'Big Data: An opportunity in search of a metaphor'를 기준으로 제작성
(radar.oreilly.com/2011/02/big-data-metaphor.html)

- 세계 각 국의 정부와 기업들은 빅 데이터가 향후 기업의 성패를 가늠할 새로운 경제적 가치의 원천일 될 것으로 기대
 - 빅 데이터에서 유용한 정보를 찾아내고 잠재된 정보를 활용할 수 있는 기업들이 경쟁에서 시장을 선도
 - McKinsey, Economist, Gartner 등은 빅 데이터를 활용한 시장 변동 예측, 신산업 발굴 등 경제적 가치창출 사례 및 효과 제시

< 빅 데이터의 경제적인 가치 전망 >

기관명	주요 내용
Economist (2010)	<ul style="list-style-type: none"> • 데이터는 자본이나 노동력과 거의 동등한 레벨의 경제적 투입 자본, 비즈니스의 새로운 원자재 역할 • 비즈니스 트렌드 파악, 질병 예방, 범죄 해결 등 효과
MIT Sloan (2010)	<ul style="list-style-type: none"> • 데이터 분석을 잘 활용하는 조직일수록 차별적 경쟁력을 갖추고 및 높은 성과 창출 • 조직 분석역량 3단계(열망-숙련-변혁 단계) 특징 제시
PwC (2010)	<ul style="list-style-type: none"> • 빅 데이터는 이전까지는 다루지 못하고, 시도하지 못했던 데이터의 활용을 가능하게 하며 잠재적 가치와 영향력이 높음 • 빅 데이터의 중요성에 대해 기업들이 주목하고 있으며, 새로운 비즈니스의 가치창출의 핵심 키가 될 것
Gartner (2011)	<ul style="list-style-type: none"> • 데이터는 21세기 원유, 데이터가 미래 경쟁 우위를 좌우 • 기업은 다가올 '데이터 경제 시대'를 이해하고 정보 고립(Information Silo)을 경계해야 성공 가능 • 빅 데이터는 향후 주목해야할 이머징 기술(2~5년후 성숙)
McKinsey (2011)	<ul style="list-style-type: none"> • 글로벌 비즈니스 지형을 뒤바꿀 기술 트렌드의 3가지 핵심은 '클라우드', '빅 데이터', '스마트 자산(smart assets)' • 빅 데이터는 혁신, 경쟁력, 생산성의 핵심 요소 • 의료, 공공행정 등 5대 분야에서 6천억불 이상 가치 창출

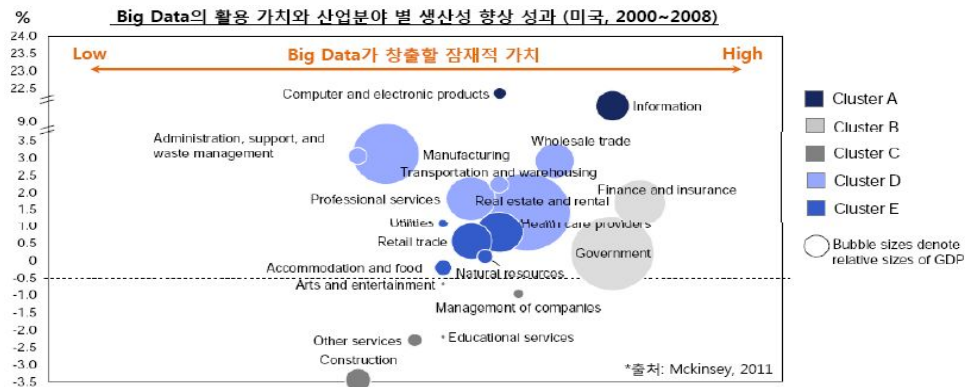
- 한편 '정보의 바다'에서 '데이터 홍수의 시대'로 변화하는 시대에 빅 데이터에 대처하지 않으면 상당한 위험에 노출될 것으로 우려
 - 기업들은 폭증하는 데이터 관리와 처리의 어려움과 유용한 데이터의 선별, 운영·유지비용 증가 문제에 직면
 - 저하된 데이터 처리속도가 기업의 생산성 하락으로 이어질 우려 증가

[빅 데이터의 활용 가치, McKinsey]

◇ 의료, 공공행정, 소매, 제조, 개인정보 부문에 적용시 1%의 추가 생산성 향상 가능

- 각 부문별로 적게는 \$1,000억에서 \$7,000억 규모의 경제적 효과 창출 예상
- (생산성 향상 정도에 따라 나누어 볼 때) 컴퓨터, 전자제품 및 정보통신분야에서 빅 데이터의 적용 효과가 클 것으로 분석됨
- 미국에서는 2018년까지 14만~19만명의 전문가 및 150만명 정도의 데이터 관리자와 분석 인력이 부족할 것으로 예상

< Big Data 활용가치 >



◇ Big Data의 활용은 산업 부문별로 약 0.5~1% 정도의 생산성 증가

- 미 의료 부문에서는 연간 \$3,300억, 유럽의 공공부문에서는 €2,500억 절감 가능
- McKinsey가 Big Data 활용가치를 투사해 본 5개 부문은 글로벌 GDP \$57.5조의 40%(\$22.3조)를 차지

< Big Data 활용가치 >



※ 출처 : McKinsey(2011), KT경제경영연구소(Big Data, 미래를 여는 비밀 열쇠) 재인용

- 빅 데이터는 미래 국가 경쟁력에도 큰 영향이 미칠 것으로 기대
 - 국가별로 안전을 위협하는 글로벌 요인이나 테러, 재난재해, 질병, 위기 등에 선제적으로 대응하기 위해 데이터 분석을 활용¹³⁾
 - ※ 미국, 영국, 일본 등 방재선진국들은 첨단장비나 센서를 활용 다양한 감시체제를 연계, 구축하여 재난 예방 상호 협력(SERI, '08)
 - ※ 미국은 홈랜드 안보를 위해 정보(intelligence) 부문에 연간 200조원 이상 투입

< 각국의 빅 데이터 관련 전략과 추진 방향 >

기관명	주요 내용
대한민국, 국가정보화전략위원회 (2011)	<ul style="list-style-type: none"> • '빅 데이터를 활용한 스마트 정부 구현(안)' 제시('11.11) • 빅 데이터 시대에 대비하여 융합 지식과 분석 행정을 통한 스마트 정부 실현을 위한 세부과제 도출
美 대통령 과학기술자문위 (2010)	<ul style="list-style-type: none"> • 'Designing a Digital Future' 보고서 발표('10.12) • 대규모 데이터(large-scale data)의 가치에 주목하고 데이터 취합 및 관리, 분석의 중요성 제시 • 미국 정부기관들이 데이터를 지식으로, 지식을 행동으로 변환하는 전략에 집중해야 함을 주장
싱가포르 RAHS 시스템	<ul style="list-style-type: none"> • 데이터를 기반으로 싱가포르를 위협하는 리스크에 대한 평가와 환경변화를 탐지하는 국가위협관리 시스템(RAHS) 구축 • 호라이즌 스캐닝(Risk Assessment Horizontal Scanning) 시스템 : 데이터 수집, 분류, 분석, 관계성, 예측 지원
OECD	<ul style="list-style-type: none"> • 빅 데이터를 비즈니스 효율성을 제공하는 새로운 자산으로 인식 • 제15차 WPIIS 회의¹⁴⁾에서 빅 데이터의 경제학 측정(Measuring the Economics of "Big Data")을 의제로 채택 : 무형 자산인 Big Data의 경제적 가치를 실증하고 정책적 함의 제공이 목적

< 우리나라 공공분야 빅 데이터 활용의 경제 효과는 10.7조원 이상 >


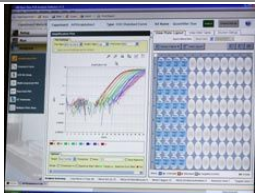
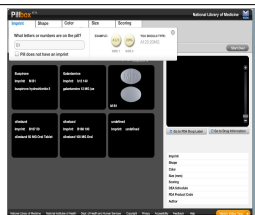
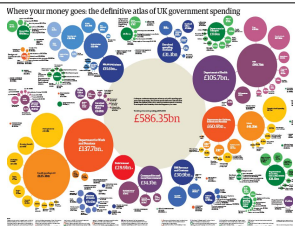

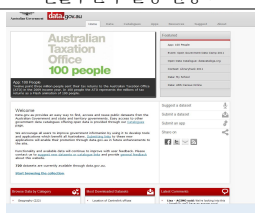
- EU는 비용절감, 부정 및 오류에 따른 손실 감소, 세수 증대에 따른 비용 효과가 1,500~3,000억 유로 규모
- 오바마 정부의 'Phillbox 프로젝트'로 연간 5000만 달러 비용 절감 효과 (미국 연간 100만건의 알약 문의 X 건당 확인 비용 평균 50달러)
- 독일 연방 노동기관에서 빅 데이터 활용 맞춤형 고용으로 3년간 백억 유로 비용 절감
- 우리나라는 최소 비율 적용 시 10.7조원의 경제 효과 추산

출처 : 국가정보화전략위원회(2011), '빅 데이터를 활용한 스마트 정부 구현'

13) 국가정보화전략위원회(2011), '빅 데이터를 활용한 스마트 정부 구현(안)'

14) OECD, 15TH MEETING OF THE WORKING PARTY ON INDICATORS FOR THE INFORMATION SOCIETY, 7-8 June 2011.

[주요국 빅 데이터 활용 현황]

국가	활용 분야	내용	비고
미국	국토 보안	<ul style="list-style-type: none"> 9.11 이후 미국은 국토안보부를 중심으로 테러·범죄 방지를 위한 범정부적 빅 데이터 수집, 분석 및 예측체계를 도입 부시행정부의 국토안보부 장관인 Michael Chertoff는 국토보안을 위한 빅 데이터 추진현황 언급 국내외 금융 시스템의 개인, 기관의 금융거래 감시로 자금 세탁 및 테러 자금 조달 색출 강화 	<p>Big data analysis can thwart sec</p> <p>By Ryan Kim Jun 22, 2011, 1:13pm PT Comments Off</p>  <p>Michael Chertoff, Department of Homeland Security, says the exploitation of Internet and the government security is to for the kinds of signals and traffic that yield important insight</p> <p>< 국토 보안을 위한 빅 데이터 추진현황 보도 ></p>
	치안	<ul style="list-style-type: none"> FBI의 종합 DNA 색인 시스템(CODIS) DNA포렌직, 클라우드DNA분석 등 “빅DNA데이터”의 활용을 통해 2007년 45,400건의 범인 DNA Hit rate 달성 1시간 안에 범인 DNA 분석을 위한 주정부 데이터 연계 및 빅 데이터 실시간 분석 솔루션 확보 	 <p>< Arkansas주 CODIS 분석 시스템 ></p>
	의료	<ul style="list-style-type: none"> 오바마 Health.20 - 필박스 프로젝트(Pillbox) 국립보건원(National Library of Medicine)의 사이트로 약 검색을 서비스 Pillbox를 통해 수집된 빅 데이터를 통해 후천성면역결핍증(HIV) 등 관리대상 주요 질병의 분포, 연도별 증가 등에 대한 통계치 확보 가능 	 <p>< PillBox 사용 화면 ></p>
영국	정보 공개	<ul style="list-style-type: none"> 영국은 정부 사이트(data.gov.uk)를 통해 공공부문의 정보 공유 및 활용을 위한 데이터 원스톱 서비스 제공 정부의 투명성 제고, 국민의 권리 향상, 데이터의 공개를 통한 경제적 사회적 가치 증대, 차세대 웹(web of data)에서 주도권 획득 목표 일반인들의 참여를 장려하고 아이디어 수렴, 앱 개발, 데이터 공개 등의 주제에 대한 커뮤니티 제공 	 <p>< Where did my tax go? ></p>
싱가포르	국가 위험 관리	<ul style="list-style-type: none"> 싱가포르 정부는 빈번히 발생하는 테러 및 전염병으로 인한 불확실한 미래 대비를 위하여 2004년부터 빅 데이터 기반 위험 관리 계획을 추진 RAHS(Risk Assessment & Horizon Scanning) 시스템을 통해 질병, 금융위기 등 모든 국가적 위험을 수집 및 분석하여 위험을 선제적으로 관리 수집된 위험 정보는 시뮬레이션, 시나리오 기법 등을 통해 분석되어 사전에 위험을 예측하고 대응 방안을 모색함 	 <p>< RAHS 2.0 시스템: 세계 신종 인플루엔자 발병 현황 ></p>
호주	정보 공개	<ul style="list-style-type: none"> 호주 정보관리청은 정부 2.0을 통한 정보 개방 방대한 양의 정보를 검색하고 분석 및 재사용할 수 있도록 자동화된 툴을 활용하여 시간과 자원을 절감 AGIMO 산하 정부 2.0 전략/서비스 팀에서는 정부 데이터에 대한 리포지터리 및 검색 툴을 서비스하는 data.gov.au 웹사이트 운영 	 <p>< data.gov.au ></p>

※ 출처 : 국가정보화전략위원회(2011), '빅 데이터를 활용한 스마트 정부 구현'



빅 데이터의 특성과 미래사회에서의 역할

□ 빅 데이터는 미래사회의 현안과를 해결하는 신 가치창출 엔진

- ‘롱테일 법칙’으로 유명한 크리스 앤더슨은 데이터 홍수로 인해 기존 과학적 방법이 필요 없어질 것이라고 주장¹⁵⁾
 - 현실의 복잡도로 인해 일부 데이터로 전체를 예측하는 샘플링 기반의 귀납적 모델링은 극단적인 예외 케이스를 놓치는 문제 발생
 - 데이터를 수집하고, 처리하는 능력이 높아지면 현실 데이터를 기반으로 상관모델을 구하는 새로운 추론 방법이 도입·보완될 것
 - ※ 그 근거로 수많은 번역 데이터를 토대로 통계적인 상관 규칙을 이용한 구글의 번역 시스템이 다른 번역 시스템보다 우수하다고 주장
- 기존 데이터의 개념에서 진일보한 빅 데이터의 특성과 컴퓨팅 파워의 발달로 빅 데이터의 실생활 적용이 빠르게 확산될 전망

< 빅 데이터의 특성과 효과 >

빅 데이터의 특성	효 과
대규모(Huge Scale)	<ul style="list-style-type: none"> • 기술 발전으로 데이터를 수집, 저장, 처리 능력 향상 • 현실세계 데이터를 기반으로 한 정교한 패턴분석 가능 • 데이터가 많을수록 유용한 데이터, 전혀 새로운 패턴의 정보를 찾아낼 수 있는 확률도 증가
현실성(Reality)	<ul style="list-style-type: none"> • 우리사회 일상에서의 데이터 기록물의 증가 등 현실 정보, 실시간 정보의 축적이 급증될 전망 • 개인의 경험, 인식, 선호 등 인지적인 정보 유통 증가
시계열성(Trend)	<ul style="list-style-type: none"> • 현시점 뿐만 아니라 과거 데이터의 유지로 시계열적인 연속성을 갖는 데이터의 구성 • 과거, 현재, 미래 등 시간 흐름상의 추세 분석 가능
결합성(Combination)	<ul style="list-style-type: none"> • 의료, 범죄, 환경, 안보 등 타분야, 이종 데이터간의 결합으로 새로운 의미의 정보 발견 • 실제 물리적인 결합 이전에, 데이터의 결합을 통한 사전 시뮬레이션, 안전성 검증 분야 발전 가능

15) Chris Anderson(2008), ‘The End of Theory: The Data Deluge Makes the Scientific Method Obsolete’, WIRED MAGAZINE : 16.07, www.wired.com/science/discoveries/magazine/16-07/pb_theory

- 빅 데이터를 잘 활용하면 미래사회에서 새로운 기회를 창출하고, 위험을 해결하는 사회 발전의 엔진 역할을 수행 할 것으로 기대
 - 사회 발전의 속도가 빨라지고, 위험요인과 복잡성이 증가할수록 시스템적으로 신속하게 환경 변화를 감지할 필요성 증가
 - 빅 데이터의 분석을 통해 미래의 통찰력, 대응력, 경쟁력, 창조력을 향상시키며 국가 지속발전을 이룰 수 있는 전략 수립 필요

< 미래사회의 특성과 빅 데이터의 역할 >

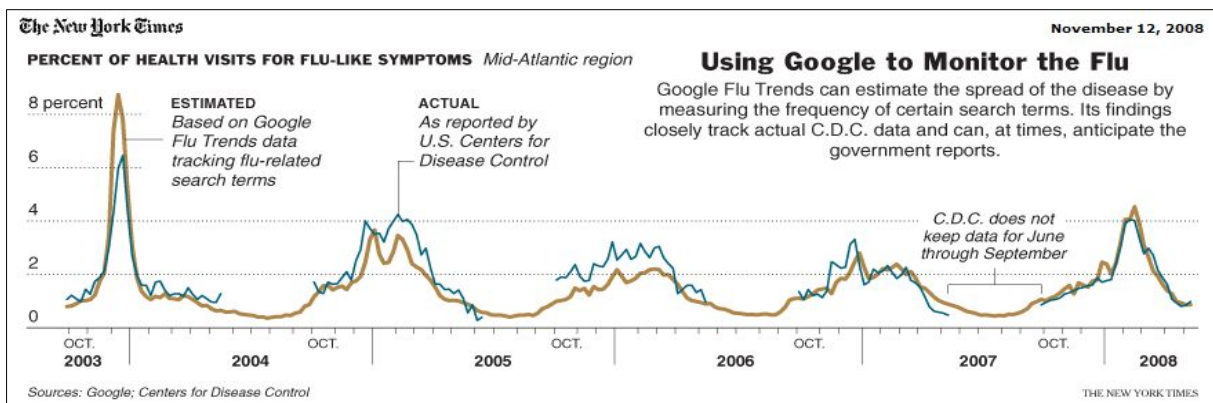
미래사회 특성	빅 데이터의 역할	
불확실성	→ 통찰력	<ul style="list-style-type: none"> • 사회현상, 현실세계의 데이터를 기반으로 한 패턴분석과 미래전망 • 여러 가지 가능성에 대한 시나리오 시뮬레이션 • 다각적인 상황이 고려된 통찰력을 제시 • 다수의 시나리오로 상황 변화에 유연하게 대처
리스크	→ 대응력	<ul style="list-style-type: none"> • 환경, 소셜, 모니터링 정보의 패턴 분석을 통한 위험징후, 이상 신호 포착 • 이슈를 사전에 인지·분석하고, 빠른 의사결정과 실시간 대응 지원 • 기업과 국가 경영의 투명성 제고 및 낭비요소 절감
스마트	→ 경쟁력	<ul style="list-style-type: none"> • 대규모 데이터 분석을 통한 상황인지, 인공지능 서비스 등 가능 • 개인화, 지능화 서비스 제공 확대 • 소셜(니즈)분석, 평가, 신용, 평판 분석을 통해 최적의 선택 지원 • 트렌드 변화 분석을 통한 제품 경쟁력 확보
융 합	→ 창조력	<ul style="list-style-type: none"> • 타분야와의 결합을 통한 새로운 가치창출(의료 정보, 자동차정보, 건물정보, 환경정보 등) • 인과관계, 상관관계가 복잡한 컨버전스 분야의 데이터 분석으로 안전성 향상, 시행착오 최소화 • 방대한 데이터 활용을 통한 새로운 융합시장 창출

- 급변하는 글로벌 환경 및 미래사회에 선제적으로 대응하기 위해서는 대규모 데이터 분석 기반의 국가미래전략수립 체계 마련이 선결 과제

□ 데이터 분석을 통한 성공사례 및 동향

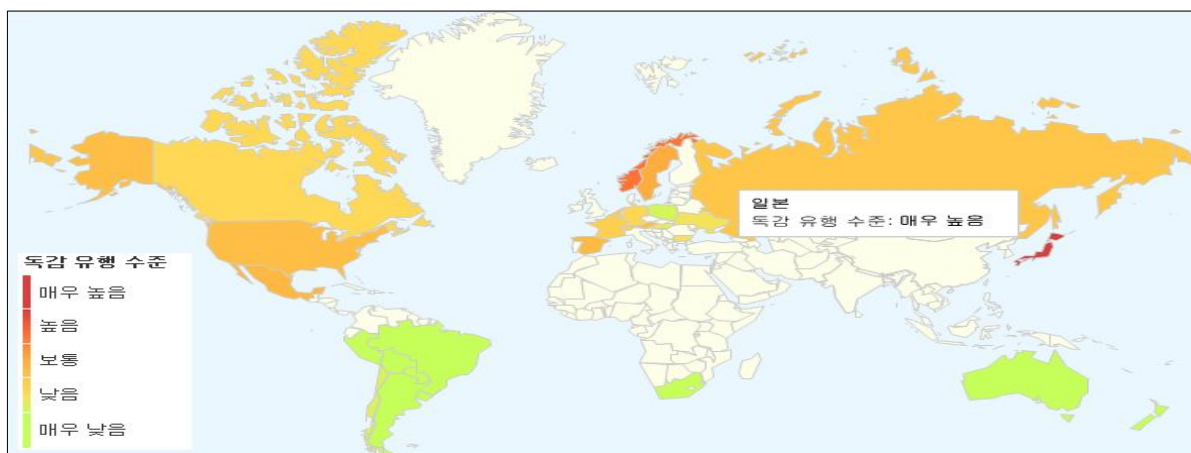
- 독감 유행 수준을 파악하는 ‘구글 독감 트렌드 서비스’, 인터넷에서의 집단행동과 의학이 만나서 맺은 결실¹⁶⁾
 - 구글은 독감에 걸리면 나타나는 증상들에 관한 검색어가 얼마나 자주 검색됐는지를 파악해 독감 확산을 예측
 - 이는 미국질병통제예방센터(Centers for Disease Control and Prevention)의 공표보다 일주일에서 열흘 앞서 독감의 창궐을 탐지¹⁷⁾
- ※ 2008년 2월, 미국질병통제예방센터가 애틀란타 주 중부의 독감 발생을 공표했었는데, 구글에서는 이미 2주 전에 독감 유행을 예보함

< 구글의 독감 모니터링 예측과 실제 확산 추세 비교 >



※ 그림출처 : 뉴욕타임즈(2008)

< 전 세계 독감 확산 현황 서비스(www.google.org/flutrends) >

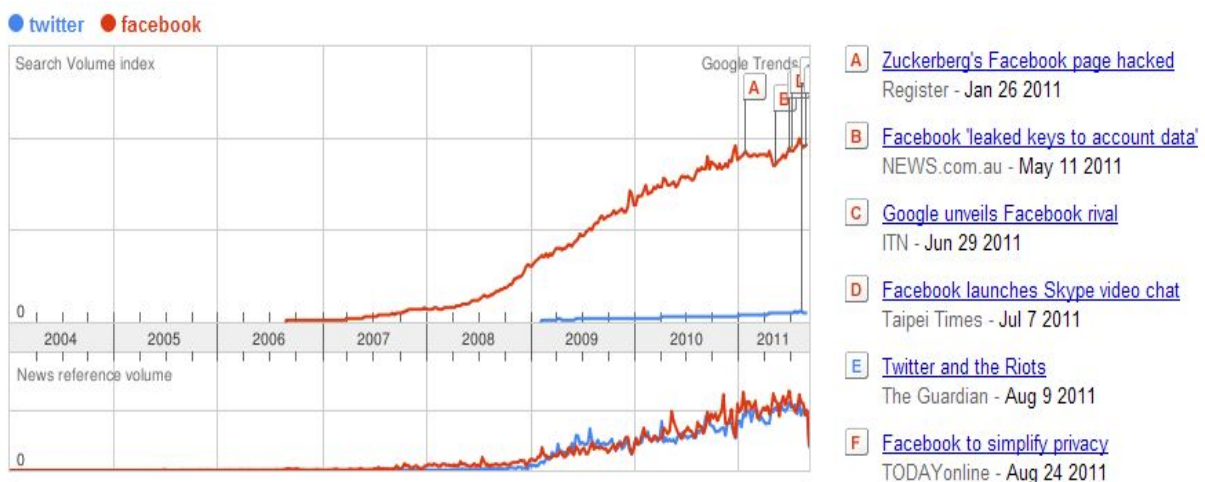


16) <http://pmsid.egloos.com/1093187>

17) <http://www.nytimes.com/2008/11/12/technology/internet/12flu.html?em>

- 오바마 정부의 ‘노후차량 보상 프로그램(Cash for Clunkers Program)’과 구글의 예측¹⁸⁾
 - 경제위기 극복과 친환경 정책의 일환으로 실행한 ‘노후차량 보상 프로그램’에 배정한 예산이 약 일주일 만에 거의 소진
 - 프로그램 계획 당시 정부는 경제위기 상황에서 새 차 구입의 수요가 적을 것으로 보고, 4개월 시행 목표로 10억 달러 예산 편성
 - ※ 일주일 만에 예산이 소진되자 의회는 20억 달러 긴급 추가예산 편성(‘09.07)
 - 그러나 구글은 노후차량 보상 프로그램에 관한 검색 폭주를 보고 호응도를 분석하여 정부 예산이 부족할 것을 정확히 예측
- 과거, 현재, 미래를 예측할 수 있는 ‘구글 트렌드 서비스’
 - 웹사이트나 키워드의 트래픽 성향을 비교해 볼 수 있게 해주는 구글 웹 서비스(www.google.com/trends)
 - 구글은 자신들이 수집한 검색 및 인덱싱 자료들을 축적해서 그것을 다양하게 활용할 수 있도록 지원
 - 1개 이상의 검색어를 입력하면, 각 검색어의 연도별 검색 추이를 비교하며 분석 가능

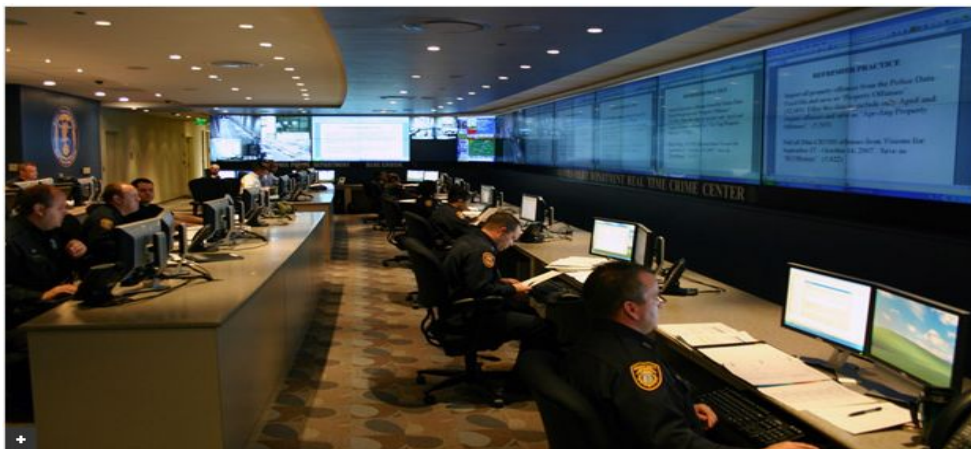
< 구글 트렌드 서비스 >



18) 장영재(2010), ‘경영학 콘서트’, 비즈니스북스

- 미국 CIA는 빅 데이터 해석과 예측분석기술 확보에 주력
 - CIA는 산하의 벤처 투자기업인 'In-Q-Tel社'를 통해 안전보장 관련 IT 활동의 일환으로 이른바 '빅 데이터'의 해석에 주력¹⁹⁾
 - In-Q-Tel社는 대량의 텍스트 데이터의 문맥을 해석하는 SW를 개발하는 Digital Reasoning社와 업무 제휴('11.2월)
 - CIA는 세계 각지의 인원이 수집하는 첩보 데이터에서 암시적 관련성 발견을 목표로 기술 활용
 - 미국 CIA와 구글이 합작으로 예측분석기술을 주력으로 삼은 기술회사인 'Recorded Future社'에 투자('10년)²⁰⁾
 - 웹 기반 분석회사인 'Recorded Future社'는 웹 상에서 벌어지는 일(개인, 그룹, 이벤트 등) 들을 모니터할 수 있는 기술력을 보유
 - '시간분석엔진'을 통해 사람들의 과거, 현재, 미래 예측의 정보를 시각화해서 제공
- 미국 테네시주 멤피스시 실시간 범죄감시 센터는 각종 데이터를 분석하고 사건 패턴을 수집해 범죄 예방에 활용²¹⁾
 - 범죄율이 30%나 감소했으며, 향후 범죄발생이 높은 지역과 시기를 예측하여 적시적소에 인력을 배치해 범죄를 예방

< 멤피스시의 실시간 범죄감시 센터 >



※ 그림출처
: 조선비즈

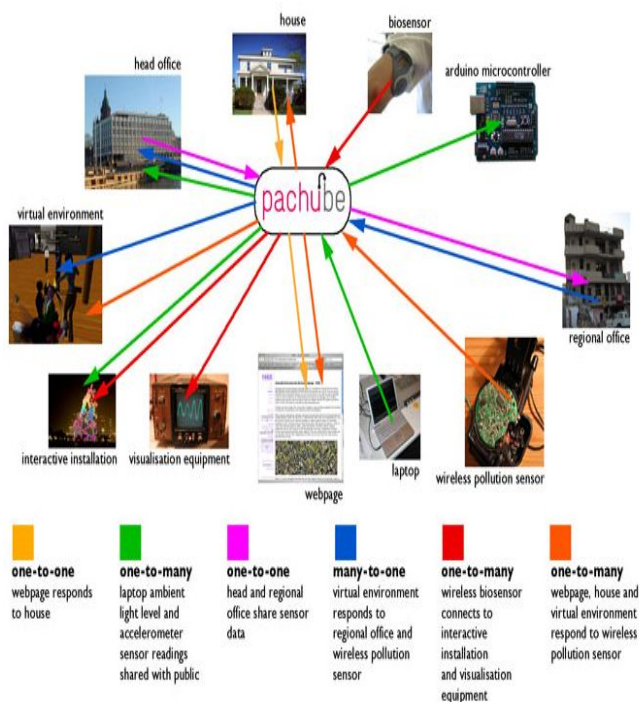
19) KIET(2012), '미국의 국방체제에서의 IT 이활용 동향'

20) http://news.cnet.com/8301-27080_3-20012213-245.html

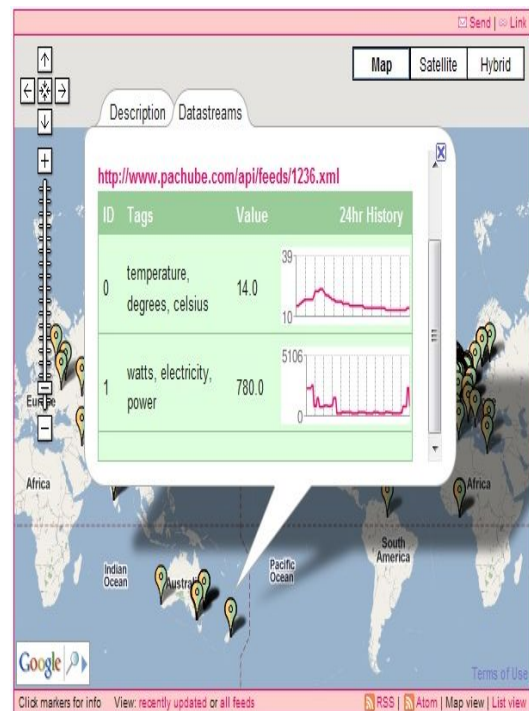
21) 조선비즈(2011), '21세기는 '정보전쟁'...정보분석에 기업생존 달렸다'

- 위키리크스 데이터 분석을 통한 효과적인 전술 정보 제공²²⁾
 - NYU 박사 과정 학생 드루 콘웨이는, 위키리크스에 저장돼 있는 데이터를 분석해 미국과 아프가니스탄 연합군의 병력 동향 파악
 - 아프가니스탄 주요 5곳을 적, 중립, 동맹지역으로 나눠 정보를 분류하고 각 지역별 활동 패턴 분석
 - 탈레반의 활동 지역, 미국 동맹 지역을 한눈에 파악할 수 있고, 시간 흐름에 따른 전쟁 양상 변화 확인 가능
- 영국의 파큐브(Pachube)는 가전·휴대폰·가로등 등에 부착된 센서를 통해 수집된 정보를 저장·분석·제공하는 전문회사
 - 센서 모니터링 정보 등 일반 데이터를 무료로 공개하며, 고급 정보를 원하는 고객에게는 유료로 서비스 제공

< 파큐브(Pachube)의 개요도 >



< 파큐브 정보 제공 예시 >



※ 출처 : www.pachube.com

22) 이지영(2011), '빅 데이터 분석이 세상을 바꾼다', www.bloter.net/archives/68798

IV

빅 데이터 시대의 준비 과제

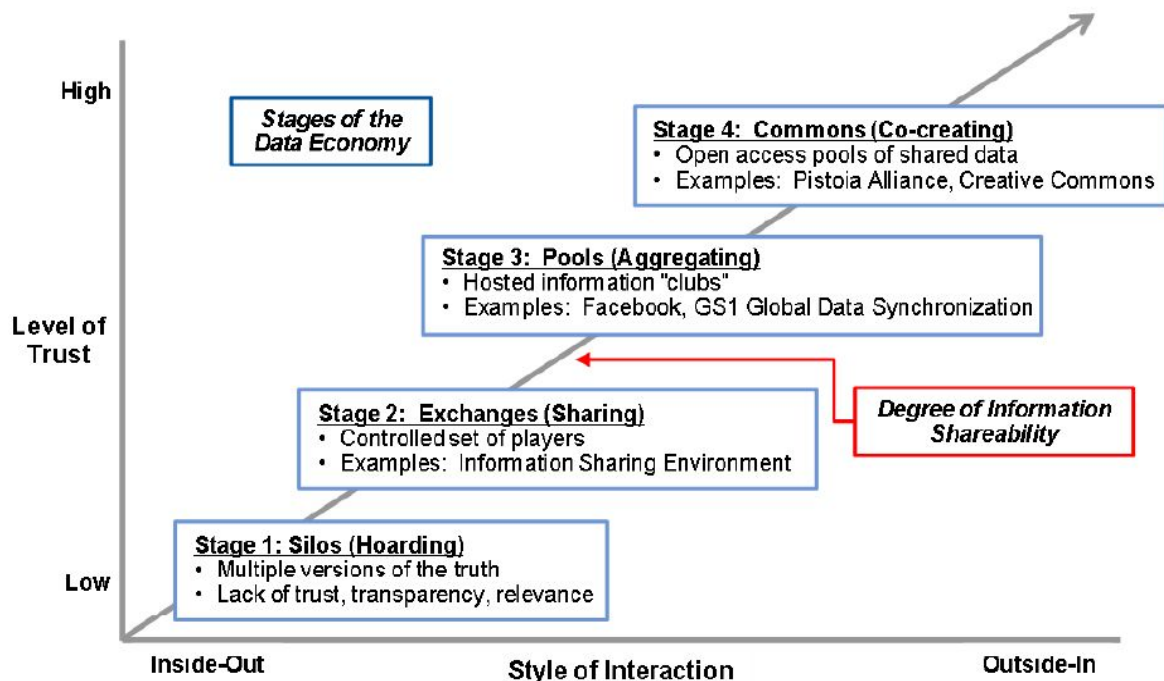
1

데이터 경제 시대를 대비하는 '연결과 협력'

- o 기업은 다가올 '데이터 경제 시대'를 이해하고 정보 고립 상태 (Information Silo)를 경계해야 성공 가능²³⁾
 - 데이터는 무한한 자원이 될 수 있으나 활용 가능한 자원의 영역은 상호 연결과 협력을 통해서 더욱 확장될 수 있음
 - 데이터 경제시대에는 플랫폼(App.), 오픈소스(SW), 초고속 컴퓨팅 파워(HW)의 영향력이 커지며 상호 연결과 협력이 핵심 전략으로 부각

< 데이터 경제 시대의 단계 >

Figure 2. The Stages of the Data Economy Era



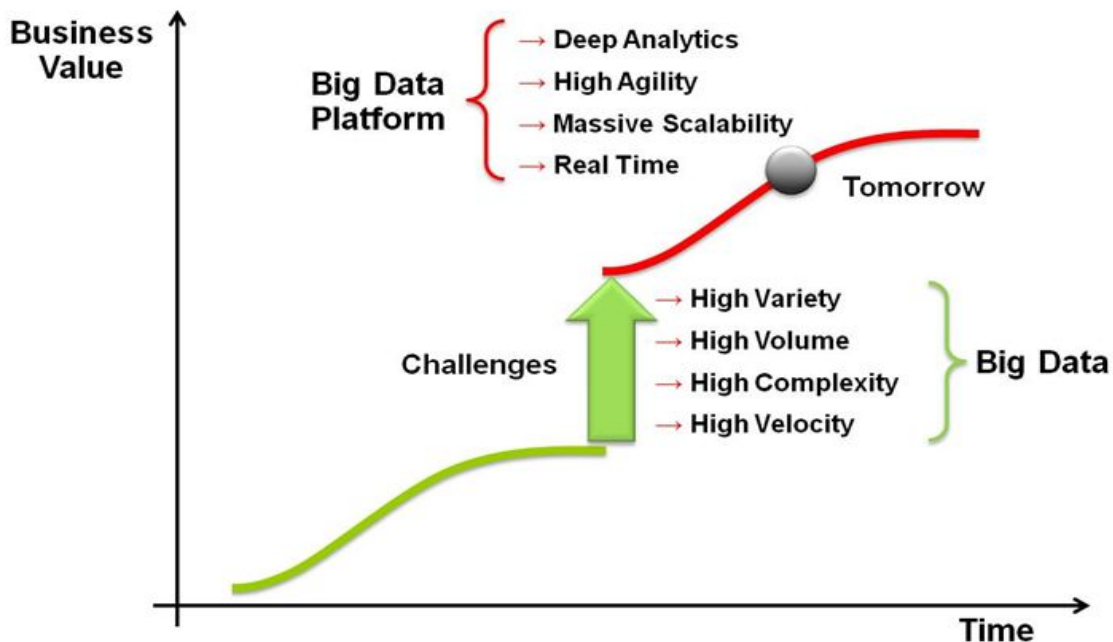
Source: Gartner (March 2011)

※ 출처 : Gartner(2011), 'How to Plan, Participate and Prosper in the Data Economy'

23) Gartner(2011), 'How to Plan, Participate and Prosper in the Data Economy'

- 단절된 정보의 제한적 활용을 개선하고, 사회 전반적·통합적 데이터 수집 및 활용을 위한 협의와 공동의 참여가 필요
 - 국가적으로 유용한 데이터 자원의 창출을 위해서는 공공부문과 민간부분이 통합되는 데이터 수집·분석 플랫폼 등의 기반 조성
 - 이를 위해서는 참여자들 간의 신뢰, 상호연결성, 공동 활용 방안마련 등이 우선되어야 하며 상호 합의를 통해 발전 가능
- 빅 데이터 플랫폼은 고급분석, 빠른 처리속도, 확장성, 실시간성을 지원할 수 있도록 구성
 - 대규모 데이터 측면에서는 대용량 데이터 저장 기술, 분산/병렬 처리 기술 등이 핵심
 - 비정형 데이터가 대부분인 빅 데이터는 복잡성이 높아지는 만큼 확장성에 대한 이슈가 중요

< 빅 데이터 플랫폼의 핵심 구성요소 >

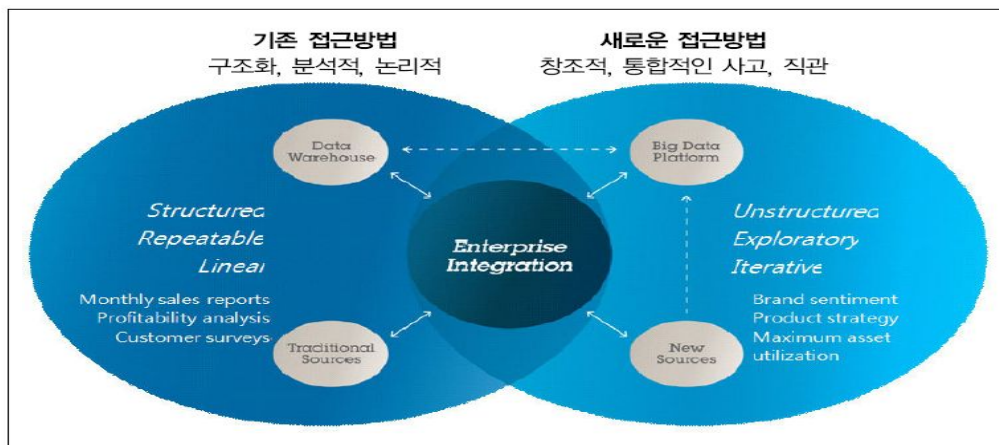


※ 출처 : Jean-Pierre Dijcks(2011), 'Big Data: Achieve the Impossible in Real-Time', nosql.mypopescu.com/post/6312810458/big-data-achieve-the-impossible-in-real-time

2 빅 데이터의 핵심 역량 '창의적 인력'의 양성

- 빅 데이터는 수많은 데이터를 수집·축적하는 것 보다 무엇을 분석할 것인지 분명한 목적의식과 통합적 사고, 해석력이 중요
 - 기존 데이터베이스의 분석은 구조화, 분석적, 논리적 접근이었으나 빅 데이터 시대에는 창조적, 통합적인 사고, 직관력이 필요
 - 수많은 데이터 속에서 목적에 부합하는 데이터를 찾고, 효과적인 질문과 분석 결과를 공급하는 창의적 분석가가 빅 데이터의 핵심

< 빅 데이터의 새로운 접근방식 >



출처 : 한국IBM(2011), 'Big Data Special Report DW 선택 가이드1', IDG Korea

- 구조화되지 않은 대규모 데이터 속에서 숨겨진 패턴을 찾아내는 데이터 사이언티스트(Data Scientist)는 스마트 시대에 최고의 인재
 - 데이터 사이언티스트는 통계학, 경제학, IT기술, 수학 등 다학제적 이해가 필요하며 학문적인 지식 외에 통합적 사고, 직관력 등이 요건
 - 글로벌 IT 업체도 데이터 사이언티스트 확보에 심혈을 기울이며 인재확보와 내부역량 강화에 노력
 - ※ 이베이에는 고객 데이터를 분석하고 가공하는 일을 맡은 직원만 5000명에 이름
 - ※ EMC는 경제학, 통계학, 심리학 등을 전공한 박사급 인재들인 데이터 사이언티스트로 구성된 '애널리틱스' 랩을 운영
 - ※ 미국 IBM은 사내에 200명 이상의 수학자들이 '분석학(analytics)'을 집중 연구, 500개 이상의 관련 특허를 취득하면서 미래 사업을 준비

3 데이터 신뢰 환경의 구축

- 소셜 미디어 상에 존재하는 메시지, 흔적이나 개인의 정보가 담긴 빅 데이터의 분석은 프라이버시 침해의 위험이 상존
 - 접속기록, 검색패턴, 데이터 속성(메타데이터)이 기록된 그림자 데이터(Digital Shadow)²⁴⁾의 증가는 프라이버시 침해를 위협
 - 소셜 네트워크 기술 발전은 개인의 사적공간을 훼손하고 있지만, 장기적으로는 사적공간의 경계가 점차 모호해지며 분쟁 증가
- 개인의 프라이버시를 보호하는 문제는 개인정보 제공자나 개인정보의 활용자 모두에게 매우 중요한 이슈
 - 데이터에 민감한 개인 사용자 정보의 노출 없이도 타당한 수준의 분석을 도출할 수 있는 방안에 대한 고려 필요

< Privacy-preserving data mining >

- 분석 대상 데이터에 포함되어 있는 민감한 개인 사용자 정보의 노출 없이도 타당한 수준의 분석 결과를 도출하는 연구 기법
- 분석하기 전에 민감한 사용자 정보를 보호하기 위해 암호화 등으로 변경하여 개인정보 노출을 최소화
- 사용자 정보가 여러 개의 저장장소에 분산되어 저장되어 데이터가 직접 노출되지 않으며, 분석도 상호 협력하에 분산 환경에서 수행

출처 : 한국정보화진흥원(2011), 11대 이머징 기술의 현황과 전망

- 데이터 생성에 참여하는 국가, 개인, 기업 등 주체들 간에 지식·자원의 결합과 협력 촉진을 위한 신뢰기반 형성
 - 데이터의 안전한 사용을 보장하고, 데이터의 신뢰성을 높이기 위한 법적, 사회적 측면의 환경 조성이 필수적

24) 그림자 데이터(Digital Shadow)란 모바일 어플리케이션, 소셜 네트워크 서비스 등 정보서비스 이용 시 생성되는 위치정보, 검색패턴, 접속기록과 같은 숨겨진 데이터를 말함(IDC, 2011)



참고 자료

- [1] 강학주(2011), 'Social Big Data & Collective Intelligence'
- [2] 국가정보화전략위원회(2011), '빅 데이터를 활용한 스마트 정부 구현'
- [3] 양지현(2011), 'Advanced Analytics 경쟁 우위 확보를 위한 차세대 핵심 과제', LG CNS, Technology Inside 제 12호
- [4] 이지영(2011), '빅 데이터 분석이 세상을 바꾼다', www.bloter.net/archives/68798
- [5] 장영재(2010), '경영학 콘서트', 비즈니스북스
- [6] 채승병(2011), '정보홍수 속에서 금맥 찾기: '빅 데이터(Big Data)' 분석과 활용', 삼성경제연구소, SERI 경영노트, 제 91호
- [7] 한국정보화진흥원(2011), 11대 이머징 기술의 현황과 전망
- [8] 허지성(2011), '정보의 소셜화 시대, 공감의 맥을 찾아라', LG경제연구원, LG Business Insight, Weekly 포커스
- [9] KIET(2012), '미국의 국방체제에서의 IT 이활용 동향'
- [10] IDG(2011), '빅 데이터 시대 BI 성공의 조건', IDG Tech Focus
- [11] IDG(2011), 'Big Data Special Report : DW 선택 가이드 1'
- [12] IDG(2011), 'Big Data Special Report : DW 선택 가이드 2'
- [13] KT경제경영연구소(2011), 'Big Data, 미래를 여는 비밀 열쇠'
- [14] Saltlux(2011), 'White Paper : 검색의 진화, 시맨틱 검색'
- [15] Chris Anderson(2008), 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', WIRED MAGAZINE : 16.07, www.wired.com/science/discoveries/magazine/16-07/pb_theory
- [16] Cisco(2011), 'Cisco Visual Networking Index : Global Mobile Data Traffic Forecast Update, 2010-2015'
- [17] Gartner(2011), 'Hype Cycle for Analytic Applications, 2011'
- [18] Gartner(2011), 'Hype Cycle for Emerging Technologies, 2011'

- [19] Gartner(2011), 'How to Plan, Participate and Prosper in the Data Economy'
- [20] Gartner(2011), 'CEO Advisory : 'Big Data' Equals Big Opportunity'
- [21] IDC(2011), 'Extracting Value from Chaos'
- [22] IDC(2011), 'Digital Universe Study'
- [23] McKinsey&Company(2010), Clouds, big data, and smart assets : Ten tech-enabled business trends to watch
- [24] McKinsey Global Institute(2011), 'Big Data : The next frontier for innovation, competition, and productivity'
- [25] OECD(2011), 'MEASURING THE ECONOMICS OF "BIG DATA"', Working Party on Indicators for the Information Society, DSTI/ICCP/IIS(2011)4
- [26] PwC(2010), 'Tapping into the power of Big Data', PwC Technology Forecast 2010 Issue 3
- [27] Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S. Hopkins and Nina Kruschwitz(2010), 'Big Data, Analytics and the Path From Insights to Value', MIT Sloan Management Review, WINTER 2001, Vol. 52, NO.2
- [28] The Economist(2010), 'The data deluge'
- [29] The Economist(2010), 'Data, data everywhere'
- [30] Tyler Bell(2011), 'Big Data: An opportunity in search of a metaphor', radar.oreilly.com/2011/02/big-data-metaphor.html
- [31] U.S. President's Council of Advisors on Science and Technology(2010), 'Designing a digital future : federally funded research and development in networking and information technology'
- [32] 각 언론사 보도자료

2011년도 IT & Future Strategy 보고서

- 제1호(2011. 4), 「IT강국 코리아의 퀀텀점프를 이끄는 플랫폼 전략」
- 제2호(2011. 4), 「미래 통일한국 시나리오와 국가정보화 전략」
- 제3호(2011. 5), 「미래 정부를 위한 지식플랫폼 구축방향」
- 제4호(2011. 6), 「미래 사회의 新학습모델, 소셜 러닝의 부상」
- 제5호(2011. 6), 「미래 사회 메가트렌드로 본 10대 미래기술 전망」
- 제6호(2011. 8), 「미래 정부를 위한 SNS 기반 지식경영시스템 구축방향」
- 제7호(2011. 9), 「중국의 미래와 정보화 정책」
- 제8호(2011. 9), 「미래 정부를 위한 NFC 기술이 공공부문 적용방향」
- 제9호(2011. 11), 「서비스 패러다임의 변화와 미래정부 혁신방향」
- 제10호(2011. 12), 「복지 패러다임 변화에 따른 新복지 이슈와 스마트 복지전략」
- 제11호(2011. 12), 「2011 국가미래준비지수」
- 제12호(2011. 12), 「소셜 네트워크 시대의 개인정보 활용과 보호」
- 제13호(2011. 12), 「미디어 패러다임 변화에 따른 정부의 소셜미디어 커뮤니케이션 방향」
- 제14호(2011. 12), 「2011 해외 신간도서로 보는 미래 정보사회」
- 제15호(2011. 12), 「2011 한국의 지역별 미래준비인식도 조사」
- 제16호(2011. 12), 「미래 국가경쟁력 강화를 위한 멘토링 네트워크 구축 전략」
- 제17호(2011. 12), 「주요국의 정보화 전략을 통해 본 미래비전」
- 제18호(2011. 12), 「新가치창출 엔진, 빅 데이터의 새로운 가능성과 대응 전략」

1. 본 보고서는 정보통신진흥기금으로 수행한 정보통신연구개발사업의 연구결과입니다.
2. 본 보고서의 내용을 발표할 때는 반드시 행정안전부 정보통신연구개발사업의 연구결과임을 밝혀야 합니다.
3. 본 보고서의 내용은 한국정보화진흥원의 공식 견해와 다를 수 있습니다.
4. 본 보고서 내용에 대해 무단전재를 금하며, 가공·인용할 때는 반드시 「한국정보화진흥원」이라고 출처를 밝혀 주시기 바랍니다.