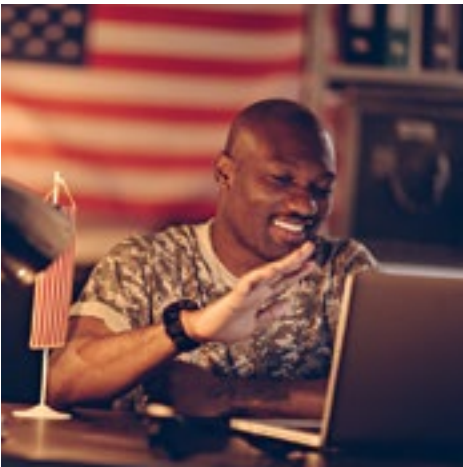


The
Software
Alliance

BSA

데이터는 왜 중요한가?



내용 요약

소프트웨어 혁신은 세상을 바꾸고, 개인에게 권한을 부여하며, 경제를 성장시키는 등 예상을 뛰어넘는 수준으로 발전을 가속화하고 있습니다.

하지만 이러한 디지털 변환의 진정한 잠재력은 이러한 혁신이 제공하는 데이터의 잠재력을 이해할 때에만 파악할 수 있습니다. 사실 우리는 데이터 혁신과 함께 살아왔습니다. 방대한 양의 데이터뿐만 아니라 정보를 수집하고, 저장하고, 분석하고, 변환하는 방식을 바꾸는 기본적인 기술도 이러한 데이터 혁신을 주도하고 있습니다.

오늘날 기업의 리더 중에서 90%가 토지, 인재, 자본 등과 마찬가지로 데이터도 가장 중요한 리소스인 동시에 가장 기본적인 차별화 요소라고 인식하고 있습니다.

얼마 전까지만 해도 강수 패턴을 파악하기 위해 데이터를 수집하려면 수백년 동안의 기상 패턴을 관찰해야 했습니다. 마찬가지로, 교통망을 파악하려면 도로변에 앉아 차량의 속도를 직접 측정해야 했으며, 질병이 전파된 방식과 그 치료법을 알려면 손으로 작성한 엄청난 양의 기록을 수집해야 했습니다.

하지만 이제 데이터는 수 많은 장치, 기계, 차량 등은 물론 가로수의 센서에서도 수집되고 있습니다. 이러한 방대한 양의 데이터를 보관하는 것은 비용도 많이 들고 어려운 일이었지만, 저장소 용량이 확대되고 비용이 급격히 저렴해지면서 저장된 데이터는 재활용이 가능한 리소스가 되었습니다. 데이터의 재활용성과 전용성 덕분에 시간과 비용을 절약하고 위험성도 적으며 가치 있는 통찰력을 제공하는 새로운 방식으로 데이터를 분석하고 변환할 수 있게 되었습니다.

캡처한 데이터 중에서 일부는 개인 정보이므로 첨단 보안을 갖추고 책임감 있는 관리 모델을 사용하여 안전하게 보호되고 올바르게 사용되도록 해야 합니다. 하지만 대부분의 데이터는 서로 통신하면서 관리자에게 수집 결과를 전달하는 다양한 장치와 기계가 수집하고 있습니다. 제조 공장의 조립 라인에서 여객기에 이르기까지 다양한 장소에서 다량의 데이터가 수집되고 분석되고 있는 것입니다. 그 결과 예전에는 상상할 수 없을 정도로, 성과가 개선되고 생산성이 급증하게 되었습니다.

이렇듯이 어디에나 존재하는 데이터의 보편성과 유익성이 우리의 생활을 다양하게 개선하고 있지만 데이터가 도대체 무엇이고 어디에서 오는지, 어떻게 사용되며 얼마나 큰 잠재력을 가지고 있는지 알고 있는 사람은 많지 않습니다.

본 자료에는 데이터 혁신으로 세계에서 가장 어려운 문제점이 어떻게 획기적인 성과로 이어지고 있는지 보여주는 몇 가지 예가 요약되어 있습니다. 또한 데이터의 수집, 저장, 분석 및 변환 방법의 근본적인 변화가 21세기 디지털 경제의 모든 가능성을 어떻게 열어주는지 설명하고 있으며, 사람들이 급성장하는 데이터 혁신 산업을 완전히 이해하기 위해 노력하는 동안, 만연해진 몇 가지 오해에 대해서도 설명하고 있습니다. 마지막으로 데이터 경제에 대해 알지 못하는 사람들에게 지침이 되도록 데이터 혁신과 관련된 용어를 정의하는 용어집도 제공하고 있습니다.

데이터 혁신이 가져온 기회는 지금까지 존재한 적이 없던 기회입니다. 혁신적인 소프트웨어는 이미 우리 생활을

놀라울 정도로 바꾸고 있으며 급증하는 데이터 리소스 속에 숨겨진 해결책을 찾는 데 도움이 되고 있습니다. 이 새로운 전환적 도구는 데이터를 새로운 제품, 새로운 솔루션, 새로운 기술혁신으로 바꾸면서 우리 생활도 바꾸고 있습니다. 경제학적 관점에서 볼 때 데이터를 잘 활용하면 향후 4년 동안에만 1조 6000억 달러의 “데이터 배당금”을 획득할 수 있습니다. 경제학자들은 데이터로 인한 효율성 개선으로 2020년 전세계 GDP에 약 15조 달러를 추가할 수 있을 것으로 예상하고 있습니다.

우리가 지금 올바른 선택을 한다면 “데이터 중심적” 경제는 새로운 일자리와 산업, 혁신 기술, 새로운 치료법을 창출하는 데 있어서 핵심적 역할을 할 것이며 미래의 경제 성장을 가속화할 것입니다.

‘데이터 혁신’의 정의

데이터 혁신의 “4가지 V”, 즉 데이터의 양을 나타내는 볼륨(Volume), 데이터가 생성되는 속도인 속도(Velocity), 관련된 데이터의 종류를 나타내는 다양성(Variety), 데이터의 정확성을 나타내는 진실성(Veracity)에 대한 이야기는 많이 들어보셨을 것입니다. 하지만 원시 데이터의 가치가 얼마나 보잘것 없으며 데이터 활용을 극대화하기 위해 우리가 공유하는 혁신적 기회가 얼마나 희박한지 따져본 적은 별로 없었습니다.

본 자료에서도 언급하고 있지만 실용적 혜택에서 필수적인 혜택에 이르는 다양한 혜택을 제공할 수 있도록 데이터를 수집하고, 저장하고, 분석하고, 변환해야 합니다. 이러한 절차는 데이터 혁신의 핵심이며 방대한 양의 비생산적인 정보에서 엄청난 가치를 발견하는 일과 같습니다.

목차

- 5 서론
- 7 데이터를 해결책으로 변환
- 14 데이터 중심적 경제
- 17 데이터에 대한 오해와 진실의 구분
- 28 디지털 담화: 데이터 언어에 대한 이해
- 34 미주
- 40 BSA 소개



기원전 6000년

15세기



1850년대

21세기



데이터 이정표

서론

인류 역사에서 문명의 발전을 알려주는 척도는 정보를 관찰하고 수집하는 능력의 발전으로 결정되어왔습니다. 우리 조상들은 거리, 무게, 부피, 온도, 시간, 장소를 측정하는 도구를 개발했고 개선을 거듭했는데, 이러한 각각의 도구는 수렵 생활에서 농경 생활로, 그리고 도시에 거주하는 문명인으로 발전하는데 중요한 역할을 했습니다.

기원전 6000년 전부터 작물 수율과 휴경에 대한 데이터를 사용하여 생산량과 수확량을 늘렸습니다. 15세기에는 천문학 데이터를 활용하여 세상을 탐험하고 해외 무역의 길을 열었습니다. 1850년대에는 데이터를 사용하여 오염된 물과 콜레라의 관계를 밝히고 인간의 생명을 구했습니다.

근대사에서는 제한된 양의 데이터라도 가장 어려운 도전과제에 대한 해결책을 찾아내는 통찰력을 제공하는 데 기여했습니다. 비석이나 파피루스 두루마리, 계몽 책자, 인쇄물 등에 기록되었는지에 관계 없이 데이터와 데이터의 보편성 및 중요성은 경제 성장 및 인류 발전의 주요 원동력이었습니다.

21세기에는 이러한 과정이 급속하게 가속화되고 있습니다. 데이터의 양은 더 많아지지만 데이터 저장 비용이 급감하면서 데이터 과학자들은 새로운 기술로 개발된 첨단 도구로 무장하여 다량의 데이터에서 중요한 통찰력을 얻고 있습니다. 데이터를 처리하는 기술이 점점 다양화되면서 그러한 기술의 영향도 더욱 커지고 있고 기회도 더 많아지고 있습니다.

우리는 정보의 제한이 없고, 가능성이 무한한 세상을 향해 발전하고 있습니다. 데이터의 활용 방식을 연구한다면 일상 생활을 개선할 방법도 예측할 수 있을 것입니다. 예측 가능한 데이터는 출근길에 우산을 가져갈지, 버스를 탈지 미리 결정하는 데 도움이 됩니다. 또한 교통 데이터는 신호등을 동기화하고, 열차 도착 시간을 예측하고, 자녀의 리허설 시간에 도착하기 위해 지름길을 찾는 데 도움이 되고 있습니다. 웨어러블 디바이스는 더 건강하고 더 오래 살기 위해 현명한 선택을 하도록 개인별 건강상태를

추적할 수 있게 하며, 과학자들은 새로운 치료법을 찾고 더 효과적이며 개별화된 치료법을 개발하기 위해 엄청난 양의 유전자 정보를 분석하고 있습니다.

세상을 바꾸는 데이터

- + **바르셀로나**는 데이터를 활용하여 더 편리한 도시로 변하고 있는데 그 결과 관광객의 이동 패턴을 파악하고, 공용 자전거 대여소를 더 배치할 장소를 찾고, ATM이 더 필요한 장소를 파악하고 있습니다.
- + **아랍 에미리트**에서는 새로운 데이터 도구를 사용하여 세계 최초로 소비하는 에너지보다 생산하는 에너지가 많은 건물을 설계하고 있습니다.
- + **케냐**에서는 모바일 데이터를 사용하여 말라리아 감염 패턴을 파악하고 정부가 방역활동을 집중적으로 할 장소를 파악하고 있습니다.
- + **아이오와주나 인도**의 농민들 모두 종자, 위성, 센서, 트랙터에서 수집한 정보를 사용하여 재배할 작물, 파종 시기, 식품의 신선도 추적 방법, 변화하는 기후에 적응하는 방법 등을 결정하고 있습니다.

수집

저장

분석

변환 및 전환

자동차를 구입할 때 가격표가 제공하는 정보보다는 연비, 유지보수, 보험, 사고 이력 등의 데이터가 더 정확한 선택을 하는 데 도움이 됩니다. 사실 이제는 자동차 자체가 움직이는 슈퍼 컴퓨터입니다. 차량의 성능을 분석하는 센서와 대화하는 프로세서가 있어서 운전자에게 오일 교환 시기, 전기 모터로 전환할 시기, 후진할 때 도로에서 놓고 있는 아이가 있는지의 여부 등을 알려줍니다.

이렇게 방대한 양의 데이터 덕택에 손가락 하나로 필요한 정보를 찾을 수 있게 되었습니다.

그런데 "데이터"란 정확하게 무엇일까요? 누가 데이터를 만드는 것일까요? 우리 생활을 얼마나 개선할 수 있을까요? 데이터의 장점을 극대화하려면 어떻게 활용해야 할까요? 데이터를 우리의 가치와 문제에 부합하는 방식으로 사용하려면 어떻게 해야 할까요?

데이터는 부족했던 자원에서 방대하고 가치가 높은 재생 가능한 자원으로 전환되었으며, 경제 및 사회적 혜택을 더 많이 제공해주는 주요 소스가 되고 있기 때문에 이러한 질문은 매우 중요한 질문입니다. 역사적으로 불 때 경제적으로 성공한 사람과 실패한 사람들 간의 차별화 요소는 토지, 노동력, 자본과 같은 리소스에 대한 사용 가능성이었습니다. 오늘날 기업의 리더 중에서 90%가 토지, 인재, 자본 등과 마찬가지로 데이터도 가장 중요한 리소스인 동시에 가장 기본적인 차별화 요소라고 말하고 있습니다.¹

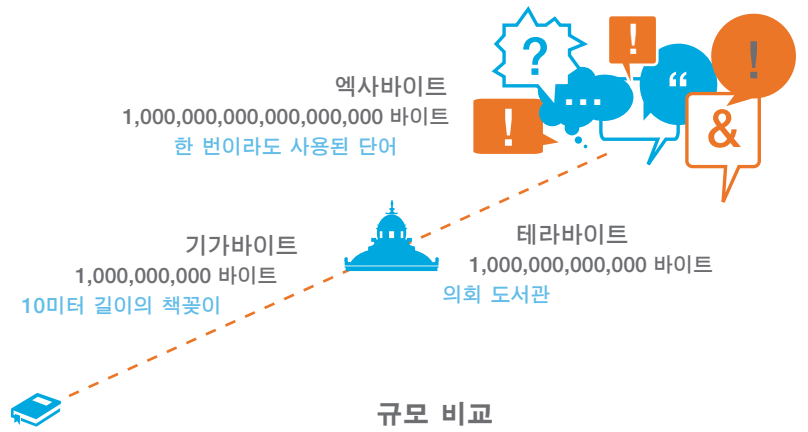
예를 들어 경제학자들은 데이터를 더 효율적으로 활용하여 업계의 효율성을 1% 개선할 경우, 보수적으로 보더라도 2030년 전세계 GDP가 거의 15조 달러가 증가할 것이라고 전망하고 있습니다.² 인터넷으로 연결된 수 많은 작은 시장이 "차세대 거대 시장"을 형성하여 세상에 대한 더 나은 데이터를 생성함으로써 더욱 데이터 중심적인 솔루션을 찾을 수 있습니다.³ 우리는 우리가 모르는지 깨닫지도 못했던 궁금한 사항들에 대한 답을 찾아내고 있습니다.

이러한 거대한 변화가 지금 일어나고 있습니다. 우리가 하는

모든 일에서 데이터가 생성되고 완전히 새로운 데이터의 흐름이 매일 생기고 있습니다. 사실 전세계 데이터의 90%는 지난 2년 동안 생성된 것이며 2년마다 데이터 생성 속도가 2배로 증가하고 있습니다. 생성되는 데이터의 대부분은 개인 데이터가 아닙니다. 데이터 프라이버시를 보호하는 것은 필수지만 우리의 생활을 개선해주는 데이터는 대부분 기계에 부착된 센서가 생성한 것이기 때문에 이것은 중요한 사항입니다.

우리의 과제는 데이터를 활용하고, 독창성을 활용하여 데이터에 내재하는 중요한 가치를 찾아내는 것입니다. 데이터를 처리하고, 관찰한 내용을 통해 통찰력을 얻으며, 통찰력을 통해 해답을 찾는 능력은 오늘날 직면한 중요한 과제에 대한 의미 있는 해결책을 찾는 데 도움이 될 것입니다.

지금이야말로
기업과 정부가 적극
협력하여 혁신 엔진을
가동해야 할 시점입니다.



데이터를 해결책으로 변환시키는

네 가지 핵심적 단계

오늘날의 데이터 혁신을 주도하는 것은 방대한 양의 데이터가 아니라 우리가 데이터를 수집하고, 저장하고, 분석하고 변환하는 기본적인 기술입니다. 이러한 기술 덕분에 데이터에 포함된 사실로부터 강력한 통찰력을 얻을 수 있으며, 새로운 지식을 발견하고 새로운 관계를 찾으며 새롭게 예측을 할 수 있습니다.

1

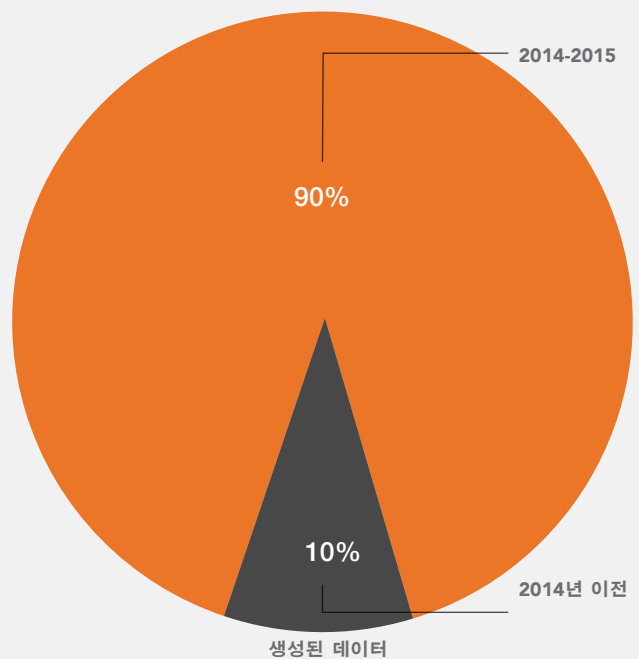
데이터 수집

역사를 돌아볼 때 우리는 이미 데이터를 수집하여 사회를 발전시키는 데 활용해왔습니다. 하지만 데이터는 그렇게 흔한 것이 아니었습니다. 이제 우리는 더 많은 전세계의 데이터를 활용할 수 있게 되었습니다. 데이터는 더 많은 의미를 담게 되었고, 더 유용한 형태로 수집되고 있으며, 더 신중하게 결과가 도출되고 있습니다. 예전에는 흔하지 않았던 데이터가 방대하고 중요하며 재사용이 가능한 자원이 되었습니다.

이것은 개인 정보를 통합하고 수집하는 능력이 아니라 다양한 장치와 센서를 인터넷에 연결하여 폭발적인 속도로 증가하는 새로운 데이터를 생성하는 우리의 능력으로 가능해진 일입니다. 그 결과 데이터는 교통 흐름을 측정하는 도로변의 센서, 우리가 만드는 디지털 음악과 영화, 공중에 떠있는 인공위성, 공장과 금융 시장을 통제하는 센서와 시스템, 차세대의 새로운 시장을 디지털 방식으로 설계하기 위해 사용하는 도구 등에 의해 모든 장소에서 생성되고 있습니다.

오늘날의 데이터 성장

전세계 데이터의 90%가 지난 2년 동안 만들어진 것.



자료 출처: IBM
<http://www.ibm.com/software/data/bigdata/what-is-big-data.html>

이러한 정보는 급속도로 증가하고 더 멀리 전파되며, 중요성도 커지고 있습니다.

사실 매일 2500경 바이트의 데이터가 생성된다고 추산되고 있습니다.⁴ 아직도 아날로그를 고집하는 사람들은 데이터가 얼마나 방대한지 헤아리기도 어려운 정도입니다. 작년에 전 세계적으로 생성된 디지털 데이터의 양은 지구에서 달까지 DVD를 포개서 왕복할 수 있는 거리만큼 된다고 합니다.⁵ 또한 데이터가 생성되는 속도도 점점 가속화되고 있다고 합니다. 전 세계 모든 기업의 비즈니스 데이터의 양은 1.2년마다 2배로 증가하고 있다고 합니다.⁶ 그 많은 데이터는 어디에서 오는 것일까요? 다음의 예는 몇 가지 데이터의 출처를 보여주고 있습니다.

대부분의 데이터에는 개인 식별 정보가 없습니다. 전세계의 연결된 장치들 덕분에 머지 않아 우리가 세상을 더욱 잘 이해하고 개선할 수 있을 것입니다.

- + 임상적 영상이 대부분인 병원의 디지털 정보는 2015년까지 하루에 665 테라바이트까지 증가할 것으로 예상되며 치료법을 찾고 생명을 구하는 데 기여할 것으로 예상됩니다.⁷
- + 현대의 대륙간 항공기에는 엔진, 덮개, 착륙 장치 등에 연결된 센서가 많아서 비행 성과를 개선하기 위해⁸ 비행할 때마다 0.5 테라바이트의 데이터를 생성함으로써 이전보다 2,000배나 빠른 속도로 난기류를 차단하고, 안전을 개선하고, 엔진 결함을 감지하고 있습니다.⁹ 매일 운항되는 25,000여 대의 항공기에 이 수치를 곱하면 상업용 항공기에서 생성되는 유용한 데이터의 양을 가늠할 수 있을 것입니다.
- + 기상 측정 인공위성, 기상관측소, 레이더 및 기타 센서에서는 시간당 15회에 걸쳐 22억 5000만 개 이상의 데이터 요소에서 데이터를 캡처하여 전세계적으로 더 정확한 기상 예측이 가능합니다.¹⁰
- + 금융 거래로 하루에 4-5 테라바이트의 데이터가 생성되어 실시간 분석 및 문제가 있는 거래 활동을 감지하는 데 사용되고 기업이 성장하고 경제가 활성화되는 데 기여합니다.¹¹
- + 수 많은 배송 차량의 컴퓨터 통신 센서는 엔진 성능을 추적하고, 라우팅을 개선하며 사전에 문제를 감지하고 있습니다. 맵핑 데이터 분석과 통합된 차량 센서 데이터를 통해 기업들은 엄청난 양의 연료를 절약하고 도로에서 수천대의 차량을 1년 동안 운행하여 발생하는 것과 같은 양의 배기가스를 줄일 수 있습니다.¹²

- + 유럽입자물리연구소(European Organization for Nuclear Research), CERN의 강입자충돌기(Large Hadron Collider)는 실험할 때 1초마다 40 테라바이트의 데이터를 생성하여 가장 심오한 우주 원리의 비밀에 대해 새로운 견해를 제공하고 있습니다.¹³ 칠레의 LSST(Large Synoptic Survey Telescope)도 매일 밤 하늘을 관찰할 때 우주에 대한 데이터를 30 테라바이트 정도 생성합니다.¹⁴
- + 단일 DNA 지놈의 염기서열은 200 기가바이트의 데이터를 생성합니다. DNA 염기서열 정보 처리 비용이 급감하면서 과학자들은 의학적 발전과 관련된 차이점 및 유사점을 발견하고 생명을 구하기 위해 수 많은 염기서열로 구성된 방대한 데이터베이스를 구축하고 있습니다.¹⁵

데이터의 양도 폭발적으로 급증했지만 데이터 생성 방식도 크게 증가했습니다. 인터넷과 세상을 연결하는 장치의 종류가 많아지고 “사물 인터넷”이 형성되면서 수 많은 센서들이 완전히 새로운 형태의 데이터를 매일 생성하고 있습니다. 2020년까지 강력한 센서를 갖춘 약 500억 대의 장치들이 인터넷에 연결될 것으로 예상되기 때문에 차세대 거대 시장은 수 많은 작은 시장들을 바탕으로 형성될 수 있습니다.¹⁶

이러한 장치들은 토양의 습도, 엔진 성능, 에너지 시스템의 효율성, 천식이 발병하는 장소 등을 측정하는 등의 일을 함으로써 데이터를 생성할 것입니다. 인간은 오감을 사용하여 주변을 감지합니다. 머지않아 전세계로 연결된 장치들이 세상을 더 많이 이해하고 개선할 수 있도록 하기 위해 물리적으로 세상에 대한 모든 특징들을 감지하게 될 것이며 엄청난 양의 새롭고 유익한 데이터를 생성할 것입니다.

데이터 저장 비용이 계속 하락하고 데이터의 양은 점점 늘어나고 있어 데이터 사용도 지속적으로 확대되고 있습니다.

2

데이터 저장

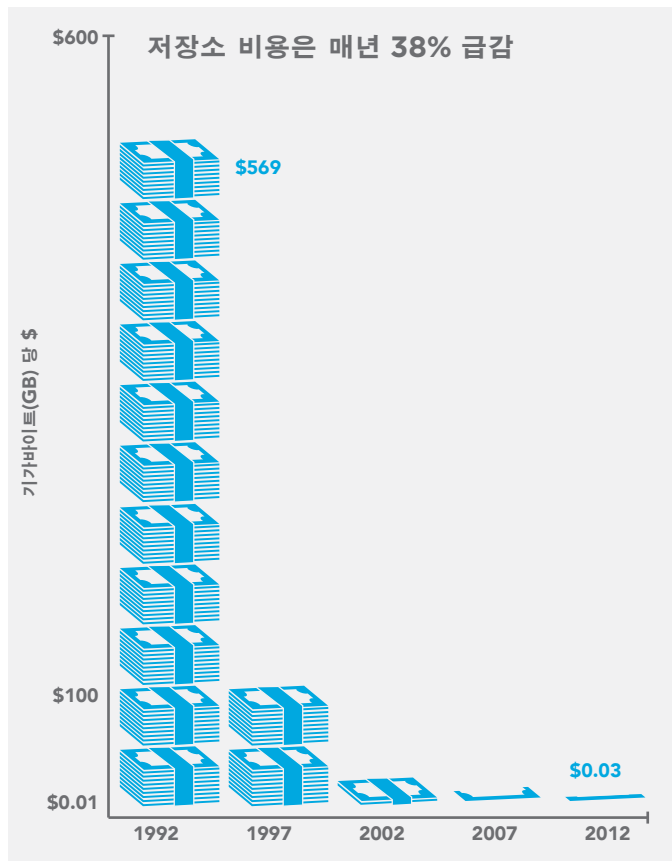
저장 비용의 급감으로 데이터 중심적인 혁신이 가능해지고 있습니다. 1980년에 1기가바이트의 데이터 저장소는 거의 찾아볼 수 없었고 비용도 막대했으며, 관리를 전담하는 직원이 필요했습니다.¹⁷ 하지만 오늘날에는 1기가바이트의 저장소 가격은 몇 푼에 불과하며 관리도 간편하고, 언제 어디서나 액세스할 수 있습니다.¹⁸ 1980년대 이후 저장소 비용은 1000만 분의 1 수준으로 낮아졌습니다.¹⁹ 같은 맥락에서 석유 가격이 같은 비율로 하락했다면 우리가 1980년에 지불한 연료 비용으로 지금 약 10,000배나 싼 가격에 차를 운행할 수 있을 것입니다.²⁰

저장 비용이 계속 저렴해지기 때문에 데이터의 양이 급증해도 저장할 수 있게 되었습니다. 1994년에는 전세계에서 3%의 데이터만 디지털 방식으로 저장되었습니다.²¹ 2007년에는 94%가 디지털 방식으로 저장되었습니다.²²

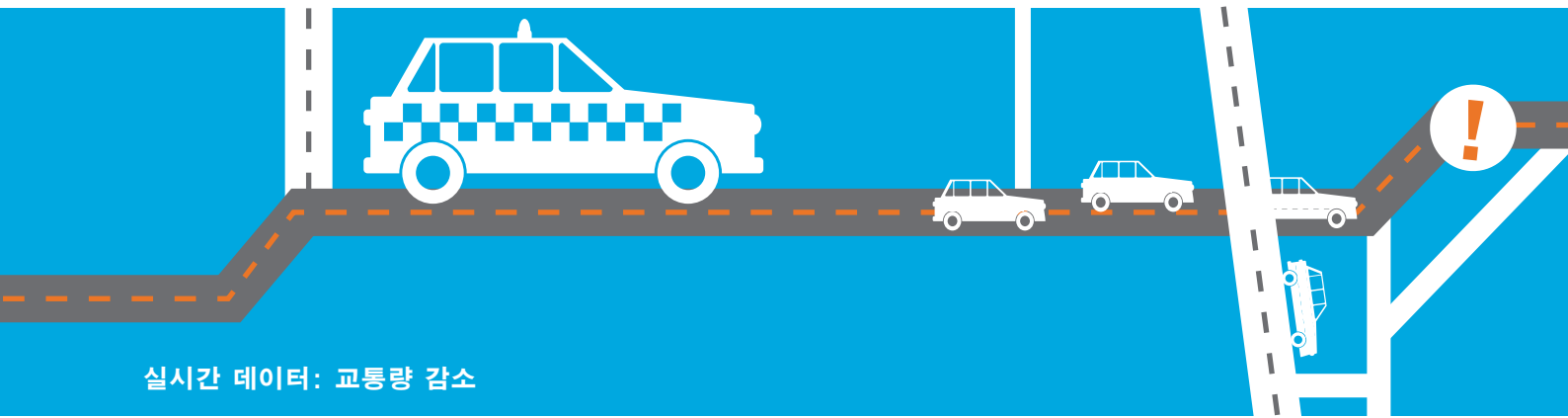
데이터 저장소의 비용이 계속 하락하고 데이터의 양은 점점 늘어나고 있으므로 데이터 사용도 지속적으로 확대되고 있습니다. 클라우드 기술로 데이터를 원격으로 저장할 수 있고 여러 장치에서 액세스할 수 있으며 데이터 저장소 가격이 급속도로 하락하기 때문에 데이터를 사용한 후 새로운 데이터를 저장할 공간을 만들기 위해 데이터를 삭제할 필요가 없어졌습니다.²³

그 결과, 데이터를 사용해도 다른 자원과 마찬가지로 고갈되지 않습니다. 데이터는 다른 데이터 집합과 결합할 수 있고 여러 차례 사용하여 데이터를 처음에 생성할 때 알지 못했던 문제점에 대한 해답도 찾을 수 있는 재생 가능한 자원입니다. 예를 들어, 날씨 데이터는 우산을 준비할지 결정하는 데만 사용하는 것이 아니라 작물 수율을 예측하는 데에도 도움이 될 수 있습니다.

클라우드 기술은 사용자가 더 신뢰성 있고 저렴하며 더 융통성 있게 데이터를 활용할 수 있도록 하며, 방대한 양의 데이터를 회사 내부의 IT 인프라에 보관할 수 있어서 보관에 대한 부담도 덜어줍니다. 데이터 저장소를 구입, 판매 및 배송하는 방식을 기본적으로 바꾸고, 데이터를 언제 어디서나 사용할 수 있게 하는 클라우드 기술은 이 시대의 가장 전환적인 기술 중 하나이며 데이터 중심적 솔루션 중에서도 최대 공신으로 떠오르고 있습니다.



자료 출처: Hagel III, John 외 From Exponential Technologies to Exponential Innovation. Deloitte University Press, 2013. 인��출. 2013 Shift Index Series.



실시간 데이터: 교통량 감소

3

데이터 분석

데이터는 이해할 수 있을 경우에만 가치가 있으며 그렇지 않을 경우, 임의적 관찰 결과에 불과합니다. 데이터에 포함된 의미를 파악하는 것은 인간의 창의성과 혁신적 소프트웨어를 결합해서만 가능합니다.

세상이 점점 자율적으로 변하고 있지만 데이터에서 답을 찾으려면 호기심, 인간의 능력, 그리고 집중적 노력이 필요합니다.

우선 원시 데이터가 유용한 데이터가 되려면 정화를 해야 합니다. 데이터 과학자들은 불규칙한 디지털 데이터에서 유용성을 발견하고 유용한 데이터로 만드는 데 전체 근무 시간의 50-80%나 투자해야 할 수 있습니다.²⁴

두 번째로, 창의력을 사용하여 올바른 질문을 한 다음, 불필요한 데이터를 분류하고 찾아내어 의미 있는 방식으로 결과를 해석하여 해답을 찾아야 합니다. 데이터 과학자의 역할은 분석가, 아티스트, 스토리텔러의 역할이 혼합된 형태입니다.²⁵ 각각의 데이터는 화면의 화소와 같습니다. 데이터 자체로는 약간의 정보만 제공할 뿐입니다. 하지만 충분한 양의 화소가 올바른 순서로 결합될 경우, 다양한 정보를 전달하는 화면을 형성할 수 있으며 새롭고 가끔은 예측하지 못했던 의미까지 전달할 수 있습니다.

분석 도구는 데이터를 분류하여 세분화함으로써 사용자가 새로운 패턴과 트렌드를 찾고, 서로 상관이 없어 보이는 데이터에서 예기치 않은 해결책을 발견하고, 통계학적으로 흥미있는 관계를 자동으로 찾아내는 데 도움이 될 수 있습니다. 우리는 이제 급증하는 데이터베이스와 첨단 통계 알고리즘을 갖춘 소프트웨어 분석 도구를 통해 유용한 정보를 찾기 위해 막대한 양의 데이터를 분류할 수 있게 되었습니다.

다행히도 오늘날 컴퓨터의 더욱 강력해진 처리 기능과 혁신적인 소프트웨어의 결합 덕분에 첨단 도구를 갖춘 데이터 과학자들이 방대한 양의 데이터를 이해하고 그 안에 포함된 중요한 의미를 찾아내고 있습니다.

오늘날의 네트워크가 발달된 것은 사실이지만 전체 네트워크에 있는 엄청난 양의 데이터를 한꺼번에 처리하기 위해 한 장소로 집결시키는 것은 경제적으로 비합리적이고 논리적으로도 불가능한 경우가 많습니다. 더욱 강력한 일부 분석 엔진은 수평 분산된 다수의 클라우드 컴퓨팅 플랫폼을 통해 그러한 작업을 수행하고 비용 부담도 적은 경우가 있습니다. 이러한 플랫폼으로 사용자는 여러 장소에 저장된 데이터에 대해 최고 수준의 데이터 분석 도구를 실행할 수 있습니다.

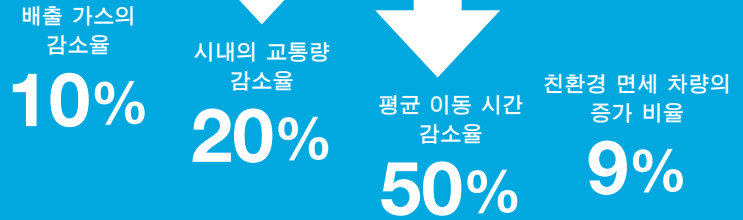
그럼 이러한 데이터 분석으로 무엇을 할 수 있을까요? 미래를 예측하는 것은 터무니 없는 일로 생각되었으나 이제는 불가피한 일이 되었습니다. 이제는 현대식 데이터 분석 덕분에 항상 신뢰성 있는 예측을 하고 있습니다. 즉, 10일 이후의 일기 예보도 상당히 신뢰성을 가질 수 있게 되었습니다. 차량 관리자는 차가 고장 나기 전에 수리해야 할 엔진이 어느 것인지 예측할 수 있습니다. 현재 상황에서 얻은 데이터를 과거의 데이터와 비교할 수 있다면 미래를 예측하는 데 도움이 되는 경우가 종종 있습니다.

경제 학자들은 시장, 취업, 인플레이션을 더욱 잘 예측할 방법을 찾고 있습니다. 하지만 정부의 의사결정자들은 오래된 경제 데이터 때문에 과거의 정보를 바탕으로 결정을 내려야 했습니다. 예를 들어 GDP 성장률과 같은 정부의 경제 통계학은 항상 몇 개월 전의 수치이기 때문에 미래를 위한 최선의 벤치마크는 과거의 경제 성과였습니다. 이제 경제학자들은 현황을 더욱 정확하게 파악하고 건전한 경제 구조를 형성하기 위해 새로운 구인 광고 및 업계 주문량과 같은 다양한 실시간 데이터를 통합하고, 더 나은 정책을 만들기 위해 과거의 데이터와 비교하고 있습니다.

또한 실시간 데이터 분석의 출현으로 자율적인 의사결정이 가능해져 더 신속하고 정확한 의사결정을 할 수 있게 되었습니다. 미국의 주요 자동차 제조업체들은 이미 수많은 센서, 컴퓨터 통신 장치, 자동 주차와 같은 첨단 기술을 가능하게 하는 실시간 연결성 등을 갖춘 새로운 차량을 설계하고 있습니다. 자동차 제조업체들은 또한 자율적으로



IBM은 스톡홀름시와 제휴하여 택시에 1600개의 GPS 시스템을 설치했습니다. GPS 장치의 데이터는 IBM 스트리밍 소프트웨어를 사용하여 처리하고 교통 흐름, 이동 시간, 최적의 통근 경로 등에 대한 통찰력을 제공하는 데 사용됩니다.



자료 출처: Bertolucci, Jeff. 'Dublin Points Big Data Tech At Traffic Jams'. InformationWeek 2013. 웹 사이트.
Nusca, Andrew. 'Stockholm Uses Real-Time GPS Data To Manage Traffic Congestion' ZDNet, 2010. 웹 사이트.

충돌을 예방하고 자동 운전이 가능한 차량을 위해 실시간 분석 도구를 개발하고 있습니다. 그러한 첨단 기술을 통해 사람이 운전할 때보다 상황에 더욱 신속하고 신뢰성 있게 반응함으로써 생명을 구할 수 있게 될 것입니다.

세상의 실시간 데이터가 급증하는 가운데 배운 내용을 즉시 이해할 수 있다면 데이터 분석 도구의 효과를 극대화할 수 있을 것입니다. 오늘날 최고의 도구가 가지는 장점은 새로운 상관 관계를 찾고, 사람들이 무엇을 질문해야 할지 모르는 상황에서도 데이터에서 예기치 않은 해답을 찾는 능력이 있어야 활용할 수 있습니다. 사실 분석 도구는 전 세계적으로 의미 있는 상관 관계를 찾고 예기치 않은 성과를 내고 있습니다. 그 예는 다음과 같습니다.

- + 캐나다의 연구원들은 1000여 개의 데이터 요소를 순식간에 추적하고 상호 연결함으로써 미숙아가 이례적으로 안정적인 활력 징후를 가지고 태어났지만 다음 날 심각한 고열이 발생한 것과의 상호 연관성을 입증하였고, 의사들은 예방 조치를 취할 수 있었습니다.²⁶
- + 앙골라와 같은 국가에서 언제 어디에서 콜레라가 발생할지 예측하는 데 지난 20년 동안의 신문 기사를 사용하고 있습니다.²⁷

- + 경찰서에서는 원래 지진을 예측하기 위해 만들어진 알고리즘을 수정하여 500 피트 이내에서 범죄가 발생하기 쉬운 장소를 예측하는 데 사용하고 있습니다. 이 소프트웨어를 사용하는 장소에서는 절도 범죄가 33% 줄었고 폭력 범죄는 21%가 감소했습니다.²⁸
- + 연구원들은 파도, 해류, 기타 데이터 등을 모니터링하기 위해 데이터 분석 및 해양 센서를 사용하는데 쓰나미와 기타 자연 재해 및 그 영향을 예측하기 위해서도 그러한 장치를 사용하기도 합니다.²⁹
- + 의사 방문 및 처방약 정보에서 수집한 데이터로 자가면역 질병을 가진 환자들의 간질 위험이 더 큰 이유를 파악했습니다.³⁰
- + 신용평점 데이터는 처방약 복용을 "친절하게" 안내할 환자를 예측하는 데 사용되고 있습니다.³¹
- + 항공사는 날씨 패턴과 관련된 10년 동안의 비행 이력 데이터를 사용하여 비행 시간을 정확하게 준수할 가능성이 가장 높은 항공편을 파악할 수 있습니다.³²

오늘날 컴퓨터의 더욱 강력해진 처리 기능과 혁신적인 소프트웨어의 결합 덕분에 첨단 도구를 갖춘 데이터 과학자들이 방대한 양의 데이터를 이해하고 그 안에 포함된 중요한 의미를 찾아내고 있습니다.

우리는 데이터에 둘러싸여 있기 때문에 기회도 그만큼 많이 존재한다고 할 수 있습니다. 혁신가들이 책임감 있고 창의적인 방식으로 행동할 때, 데이터 혁신은 일상적 문제점과 세계에서 가장 어려운 도전과제에 대한 해답을 가져올 수 있습니다.

4

데이터 변환 및 전환

우리는 새롭고 강력한 소프트웨어 도구를 통해 “느낌”이나 직감이 아닌 사실을 기반으로, 더 나은 의사결정을 내리기 위해 데이터 집합을 사용할 능력을 갖추고 있습니다.

특히 새로운 도구들은 추론하고, 집중하고, 가시화하고, 반영하고, 재정의하고, 모형을 만들고, 예측할 수 있는 방식으로 데이터를 전환하여 데이터를 유용하게 사용할 수 있게 해줍니다. 이러한 도구에는 데이터에 더욱 잘 반응하도록 하기 위해 데이터를 이해하는 기계 학습 기술, 시나리오를 테스트하고 데이터를 실제 솔루션으로 변환할 수 있는 모델링 및 시뮬레이션 기술, 사운드, 이미지 또는 비디오를 새롭고 더 의미 있는 형태로 인식하고 전환하는 도구 등이 포함됩니다.

이러한 방식으로 데이터를 변환할 경우, 더 나은 계획 수립, 훌륭한 설계, 더 현명한 의사결정을 할 수 있습니다. 예를 들어, 의료 업무를 담당하는 의사들은 새로운 연구개발로 허덕이고 있기 때문에 최신 개발 정보를 지속적으로 파악하고 환자 데이터를 실시간으로 해석하는 것이 거의 불가능한 상태입니다.³³ 그 결과 병원들은 임상 의사결정 지원 시스템에 의존하고 있습니다. 기본적으로 복잡한 데이터 환경에서 더 신속하고 신뢰성 있는 진단을 내릴 수 있도록 소프트웨어 시스템이 서로 다른 소스에서 수집한 데이터를 분석하는 것이 70% 이상의 경우 효과가 있다는 것이 입증되었습니다.³⁴

기타 소프트웨어 도구들은 데이터를 더 의미 있는 형식으로 전환하는 것을 돕고 있습니다. 또한 오디오, 이미지, 비디오 데이터의 실시간 처리는 획기적인 발명으로 이어지고 있습니다. 사람들이 말하는 방식에 대해 더 많은 데이터를 수집하기 때문에 언어 인식 기술은 지속적으로 개선되고

있습니다. 그 결과 대륙간의 음성 대화를 실시간으로 양방향 통역하는 것과 같은 혁신 기술로 전세계 상거래와 무역에 새로운 기회를 열어줄 수 있습니다. 마찬가지로 청력 상실로 고통을 받는 3억 6000만 명을 대상으로 연구하는 중국의 연구원들은 3D Kinect 센서에서 수집한 데이터의 패턴을 인식하고 실시간으로 처리하는 기술에 주목하여 수화를 이해하여 실시간으로 구어와 문자 언어로 또는 그 반대로 전환하는 시스템을 개발했습니다.³⁵

신속한 이미지 처리도 암 발견, 인지 컴퓨팅, 신경 생물학, 로봇 공학 등과 같은 분야에 큰 영향을 주고 있습니다. 예를 들어, 뇌 종양의 경우 예측할 수 없는 외관과 모양 때문에 의학 영상으로는 파악하기가 정말 쉽지 않습니다. 클라우드 컴퓨팅과 첨단 이미지 분석 알고리즘 덕분에 과학자들은 더 정확하고 신속하게 뇌 종양을 찾아내는 최고의 소프트웨어 알고리즘을 찾기 위해 경쟁하고 있습니다.³⁶

유방조영술의 경우 2D에서 3D 이미지로 전환함으로써 유방암 발견율을 개선하고 있습니다. 3차원 유방조영술은 여러 엑스레이를 다양한 각도에서 통합하여 유방암 발견율을 높이고 오진율을 줄일 수 있는 3 차원 영상을 구성하고 있습니다.³⁷

사실 가시화 및 시뮬레이션을 위해 데이터를 사용하는 능력 덕분에 기술에 대한 이해를 높이고 활용을 간편하게 할 수 있습니다. 우리는 실제로 시스템을 만들지 않고도 복잡한 시스템의 모형을 만들고 시뮬레이션하며, 개선된 정확성과 속도로 설계를 테스트할 수 있게 되었습니다. 예를 들어 1980년대에 Boeing은 실제 풍동 터널을 사용하여 77가지 종류의 767 프로토타입을 테스트했습니다. 2005년에 Boeing은 시간과 에너지, 비용 등을 절약하고 생명을



구하기 위해 787에 대해 11 가지의 물리적 테스트를 했으며 가상 풍동 터널 및 슈퍼컴퓨팅을 사용하여 프로토타입을 테스트했습니다.

가상 풍동 터널은 다량의 데이터를 압축하여 보다 간편하고 신속하게 3D 전산 유체 역학을 구현한 도구의 한 예입니다. 이러한 도구를 통해 더 나은 성능을 얻기 위해 열 흐름, 유체 흐름, 공기 흐름, 프로세스 흐름에 대해 더 나은 모델을

만들 수 있습니다. 이러한 도구는 지하수에서 오염물질이 이동할 수 있는 장소에 대한 모형을 만들고, 풍력 터빈 성능을 개선하는 방법, 자연 재해를 견디는 더 나은 건물을 설계하는 방법을 모델링할 때 사용할 수 있습니다.

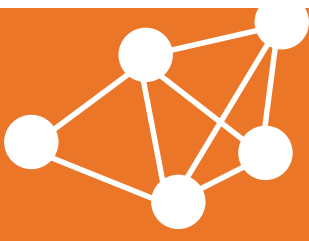
간단하게 말해서 이러한 도구는 데이터를 솔루션으로 전환하는 것입니다.

데이터를 수집하고, 보관하고, 분석하고 변환하는 능력을 수렴하여 더 나은 솔루션을 위해 새로운 기회를 모색.

실제로 이 4가지 임무를 서로 다른 사람들이 분산된 장소에 저장된 다양한 데이터 집합에 대해 수행하는 경우가 많습니다. 하지만 이것은 데이터 혁신의 장점 중에서 일부에 불과합니다. 이전에 서로 관련성이 없었던 데이터 집합이 다른 장소에 보관된 경우에도 통합하고 분석할 수 있으며, 체계화되지 않은 경우에도 확대할 수 있고, 데이터 생성자조차도 질문해야 하는지 몰랐던 질문에 대한 기본적인 답변을 찾을 수 있게 되었습니다. 특정 데이터 집합에서 추출할 수 있는 정보가 나중에 상관성이 적은 통합된 데이터 집합에게 중요해질 수 있으며 이러한 통찰력을

활용할 알고리즘이 발명되지 않았기 때문에 그러한 정보의 힘을 누구도 정확하게 예측하지 못한 것이 사실입니다.

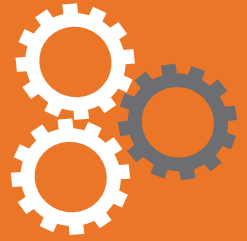
우리는 데이터와 기회에 둘러싸여 있기 때문입니다. 혁신가들이 책임감 있고 창의적인 방식으로 행동할 때 데이터 혁신은 일상적 문제점과 세계에서 가장 어려운 도전과제에 대한 해답을 가져올 수 있습니다. 새로운 생산성, 경제 성장, 개인의 혜택 등을 가속화할 방법으로 데이터 활성화 기술을 극대화하려면 혁신 엔진을 다시 가동하여 첨단 솔루션의 길을 열고 그 효과를 극대화해야 합니다.



데이터를 최대한 활용하는 제조업체는 4년 동안

3710억 달러

정도를 절약 할 수도 있습니다.



비즈니스 데이터: 비용 절약

데이터 중심적 경제

이제 데이터는 경제적 수익을 창출하는 가장 역동적인 원동력으로 부상하고 있습니다. 데이터가 경제에 미치는 영향은 기술 수준에 관계 없이 이미 여러 분야에서 확인되고 있습니다. 데이터를 더욱 잘 활용한다면 향후 4년 동안에만 전세계적으로 1조 6000억 달러의 "데이터 배당금"을 달성할 수 있을 것으로 예상하고 있습니다.³⁹

데이터 혁신은 생산성을 높인다는 것이 입증되었습니다. 이미 데이터를 바탕으로 의사결정을 하는 기업들은 생산성이 5-6% 향상되었다고 보고되어 있습니다.⁴⁰ 데이터를 더욱 효율적으로 활용한다면 다양한 업계에서 조금이라도 수익이 향상되어 효율성을 1% 높인다면 2030년에는 전세계 GDP에 약 15조 달러를 추가할 수 있을 것으로 경제학자들은 전망하고 있습니다. 이 수치는 미국 경제 규모의 2배에 달하는 수준입니다. 생산성이 1% 증가하는 것은 대단하지 않게 보일 수 있지만 GE(General Electric)의 Jeff Immelt CEO는 "정유 업계의 직원에게 소프트웨어로 수익성을 1% 개선할 수 있는 방법을 알려준다면 그 사람에게 평생의 은인으로 남게 될 것"이라고 말할 정도로 대단한 것"이라고 말했습니다.⁴¹

데이터 혁신은 일자리 창출의 지름길.

데이터 혁신은 경제 성장을 촉진하는 것 이외에도, 강력한 신규 일자리 창출 엔진인 동시에 일자리 제조기의 역할도 합니다. 데이터 혁신을 통해 데이터 분석가에서부터 소프트웨어 개발자, 데이터 혁신을 가능하게 하는 데이터 웨어하우스를 운영하는 직업까지 수 많은 신규 일자리가 이미 만들어지고 있습니다. 또한 미국의 최고 경영진의 61%와 유럽 경영진의 58%는 더 많은 직원을 채용하는 회사의 계획에 있어서 데이터 분석이 중요하다고 말했습니다.⁴² 데이터와 관련된 IT 직업이 생길 때마다 IT 이외 분야에서 3개의 일자리가 생기므로 전체 경제 차원에서 볼 때 엄청난 파급 효과가 있는 것입니다.⁴³ 궁극적으로는 또 다른 신기술 중심의 일자리가 급증하여 광범위한 경제적 혜택을 얻게 될 것입니다.

미국에서
데이터 관련
일자리 하나는 추가로
3개의
일자리를
간접적으로 창출합니다.

61%의
미국 기업 경영진은
데이터 분석이 회사의
채용 계획에서
매우 중요하다고 언급하였습니다.



기업들은 스마트 빌딩만으로 연간 에너지 비용을

250억 달러

절약할 수 있습니다.

항공 업계에서 데이터 중심의 생산성을 1% 개선할 경우

300억 달러

상당의 연료를 15년 동안 절약할 수 있습니다.



데이터는 기업의 민첩성, 대응 능력, 경쟁력을 향상.

우리 경제를 성장시키고 생활 수준을 높여준 IT 중심의 생산성 향상은 전체 경제의 일부 분야에 국한되었던 것이 사실입니다. 하지만 더 많은 분야에서 더 많은 형태의 데이터를 생성하면서 데이터는 이제 농업에서 보건, 운송에서 교육, 에너지에서 금융에 이르기까지 전통적으로 수익성이 좋지 않았던 분야에서 수익을 창출하는 역할을 하고 있습니다. 사실 최근 미국에서 실시한 설문조사에서는 최고 경영진의 79%, 유럽의 경우 80%가 고객 요구 사항을 더욱 잘 해결하기 위한 회사 계획에서 데이터 분석이 중요하다고 말했습니다.⁴⁴ 또한 최고 경영진 중에서 미국의 70%, 유럽의 72%는 신제품 또는 서비스를 만드는 회사 계획에 있어서 데이터 분석이 중요하다고 말했습니다.⁴⁵

의료 서비스

데이터 혁신은 더 나은 통찰력을 제공하여 수명을 연장하고 더 건강하게 생활하는 데 기여.

의료 서비스는 오늘날 데이터가 가장 많은 환경 중 하나로서 매일 각 병원에서 수백 테라바이트의 데이터를 생성하고 있습니다.⁴⁶ 이 데이터의 사용을 극대화한다면 치료법을 개선하고 완치 기간을 단축하는 비결이 될 수 있을 것입니다. 오늘날 데이터는 의사가 환자의 건강을 파악하고 더 나은 의사결정을 할 수 있는 능력을 제공하고 있습니다. 소프트웨어 분석 덕분에 의사들은 미래의 모든 환자들이 이용할 수 있는 치료 옵션을 확보하기 위해 이전 환자의 모든 치료 결과를 활용할 수 있습니다. 하지만 의료 데이터는 제대로 활용하지 않는 경우가 많습니다. 의료 서비스 부문이 효율성과 품질을 개선하기 위해 데이터를 더 효율적으로 활용한다면 매년 3000억 달러를 절약하고 지출을 8%나 절감할 수 있을 것으로 예상됩니다.⁴⁷ 데이터 혁신의 효과는 금전적 부문에서 뿐만 아니라 생명을 구하는 차원에서도 확인할 수 있습니다. 한 가지 예로써 연구진들은 심장 마비를 4시간 일찍 예측할 수 있는 기계 학습 알고리즘을 개발했는데 이 알고리즘은 실시간 데이터와 환자의 병력을 결합하여 66%의 경우 정확합니다.⁴⁸

운송

시간, 비용 및 연료를 절약하고, 생명을 구하는 데이터 혁신.

데이터는 더 지능적인 운송의 미래를 열어주었습니다. 즉, 항공, 철도 및 도로에서 실시간으로 데이터를 분석함으로써

재화와 사람들을 안전하고 효율적으로 운송하는 능력을 개선하는 더 지능적인 운송 시스템을 만들고 있습니다. 새 자동차는 1시간에 25 기가바이트까지 데이터를 생성할 수 있으며 이 데이터를 처리하기 위해 1000만 행 이상의 소프트웨어 코드를 포함하고 있습니다.⁴⁹ 이 데이터는 안전 벨트만큼 효과적으로 안전성을 높일 수 있는 새로운 안전 및 충돌 방지 시스템을 만드는 데 사용함으로써 부상자 및 사망자 수를 50%나 줄일 수 있습니다.⁵⁰ 도로에 차량이 넘쳐나는 상황에서 교통 체증을 줄이고 더 효율적으로 교통량을 배치하기 위해 데이터를 혁신적인 방법으로 활용함으로써 시간, 연료, 배기가스를 모두 줄이고 있습니다.

항공 업계의 데이터에 대한 관심도 급증하고 있습니다. 오늘날 현대식 항공기는 기체에 곳곳에 부착된 센서에서 1회 운항할 때마다 0.5 테라바이트의 데이터를 생성할 수 있는데 이 데이터는 운항 성능을 개선하고, 난기류를 차단하며, 안전을 개선하고, 이전보다 2,000배나 빠른 시간에 엔진 결함을 파악하는 데 사용되고 있습니다.⁵¹ 비행 데이터는 또한 운항 경로를 개선하고 고장 나기 전에 교체할 부품을 승무원들에게 알리는 데 도움이 되고 있습니다. 그로 인한 경제적 혜택도 계속 증가하고 있습니다. 항공기에서 데이터로 인해 생산성이 1% 개선될 경우 전세계적으로 300억 달러에 달하는 연료를 절약할 수 있습니다.

에너지/환경

데이터 혁신으로 에너지 소비량을 줄이고 환경을 개선.

데이터는 에너지 절약에 크게 기여하고 있습니다. 데이터를 잘 활용하면 다양한 업계에서 수십억 달러의 에너지 비용을 절약하면서 환경도 개선할 수 있습니다. 데이터를 중심으로 한 통찰력은 전기 격자, 건물, 주택, 공장, 농장 등은 물론 환경 전체의 에너지 사용을 줄여줍니다. 예를 들어, 건물에 데이터 설계 도구를 사용하여 에너지를 덜 사용하도록 설계할 수 있으며, 다량의 센서 데이터를 분석기 및 액추에이터와 결합시킨 지능적 시스템으로 통합하여 더 효율적으로 운영할 수 있습니다. 모든 점을 고려할 때 스마트 빌딩만으로 기업들은 연간 에너지 비용에서 250억 달러를 절약할 수 있다고 추산됩니다.⁵²

의료 서비스 부문에서 데이터를 더 효율적으로 사용하여 효율성과 품질을 개선할 경우, 매년 **3000억 달러**를 절약할 수 있으며 지출은 **8%**나 절감 가능합니다.

제조업

데이터 혁신은 제품의 설계, 제조 및 배포 방식을 개선.

제조 부문은 다른 부문보다 데이터를 더 많이 저장합니다.⁵³ 그 결과, 제조업체들은 효율성을 높이고, 품질을 개선하며, 제품의 설계, 제조 및 배포 방식을 개선함으로써 상당한 혜택을 누리고 있습니다. 제조업체에서 데이터를 잘 활용하면 제품 개발 시간과 조립 비용을 50%까지 줄일 수 있다고 합니다.⁵⁴ 사실 IDC(International Data Corporation)는 데이터를 충분히 활용하는 제조업체는 4년 동안 3710억 달러의 데이터 배당금을 획득할 수 있을 것이라고 예측하고 있습니다.⁵⁵ 기업들은 실시간 데이터를 사용함으로써 전세계 공급망을 더욱 잘 추적하고 관리하며 제품 결함도 줄일 수 있습니다.

데이터는 또한 설계를 개선하는 데에도 도움이 됩니다. 데이터가 활성화된 디지털 설계는 반복적인 설계 공정을 개선하여 설계자들이 최적의 제품 설계를 찾을 때까지 다양한 아이디어를 테스트하고 가상 환경에서 수정할 수 있습니다. 오늘날 Fiat와 Nissan은 데이터와 모델링 기술을 협력적으로 사용하여 신모델 개발 시간을 30%에서 50%까지 줄였습니다.⁵⁶

금융

효율성을 강화하고, 준법 이행을 향상시키며 사기 행위를 근절하는 데이터 혁신.

금융 서비스 부문에서는 방대한 양의 데이터를 잘 활용하면 운영 효율성을 개선하고, 준법 이행을 강화하며, 사기 행위를 파악할 수 있습니다. 어느 설문조사에서는 은행 및 금융 서비스 업체의 71%가 정보의 활용 및 분석으로 회사에 경쟁력 있는 장점을 얻게 되었다고 보고했습니다.⁵⁷ 즉 데이터에 투자하면 상당한 배당금을 받을 수 있는 것입니다. 어떤 신용카드 회사는 데이터 분석으로 매년 20억 달러의 신용카드 사기 피해를 예방할 수 있었습니다.⁵⁸

농업

데이터 혁신으로 적은 자원으로도 더 좋은 식량을 더 많이 생산.

데이터는 농장의 도처에서 생성되면서 증가하는 인구를 위해 더 맛있고 영양가 높은 식량을 생산하도록 도와주고 있습니다. 정밀 농업이라는 개념 덕택에 전세계 농장에서는 묘종, 인공위성, 센서, 트랙터 등에서 수집한 데이터를 사용하여 수확량을 늘리고, 비용은 줄이며, 더 많은 사람에게 혜택이 돌아가는 현명한 결정을 내릴 수 있게 되었습니다. 농업 및 식량 생산과 관련된 전세계 인구의 1/4은 데이터를 더 많이 활용하여 큰 이득을 얻고 있습니다. 예를 들어, 농부들은 데이터 분석 도구를 활용하여 원가, 살충제 사용, 화학물질 사용 등은 줄이면서 1 에이커당 5-10부셀까지 수확량을 늘릴 수 있게 되었습니다.⁵⁹ 낙농업의 경우, 크로아티아의 신흥 낙농업체는 각각의 가축에서 사료의 품질이 생산량과 수태율에 미치는 영향과 같은 실시간 데이터를 제공하는 클라우드 기반 소프트웨어 분석 플랫폼 분야를 개발했는데 그 결과 생산성과 효율성을 50%까지 높였습니다.⁶⁰

데이터는 전세계인의 삶의 수준을 개선합니다.

캐나다의 과학자들은

**1,000개 이상의
데이터 요소**

를 1초만에 추적하여

이례적으로 안정적인 활력 징후를 가지고 태어난 미숙아에게
고열이 발생할 위험이 높다는 사실을 확인하여 의사들이 조기에
조치를 취하고 생명을 구할 수 있도록 했습니다.

캐나다



데이터에 대한 오해와 사실의 구분

최근의 데이터 혁신과 데이터 경제에 대해 다음과 같이
몇 가지 잘못된 오해가 있습니다.

- + 개인 정보 및 데이터 보호,
- + 데이터 경제의 경제적 효과,
- + 데이터 신뢰성,
- + 21세기 데이터 혁신,
- + 데이터 혁신의 전세계적인 혜택,
- + 데이터 규제에 대한 정부의 역할.

개인 정보 및 데이터 보호

오해

데이터 혁신은 개인이 아닌 IT 업계에만 유리하다.

진실

데이터 혁신은 소비자들이 더 나은 의사결정을 내리게
하고 기업들은 상품과 서비스를 맞춤화할 수 있게 합니다.
또한 개별화된 학습과 의술을 통해 획기적인 발전을
주도할 수 있으며, 개별화된 엔터테인먼트를 통해 고객
경험을 개선할 수 있고, 대량 생산 중심의 경제에서 대량
맞춤화된 경제로 전환해줍니다. 데이터 혁신을 위해 필요한
혁신적인 신기술을 개발하고 제공하는 주도적 기업들이
일부 있습니다. 하지만 지금 현명한 판단을 한다면 이러한
데이터 중심적 경제는 새로운 일자리와 산업을 만들고
소비자에게 유익한 세상을 만들 수 있을 것입니다.

오해

모든 데이터는 개인 데이터이다.

진실

일부 데이터는 개인 정보일 수 있습니다(예: 모바일
장치에서 만들거나 소셜 네트워크를 사용하여 생성한
데이터). 하지만 대부분의 데이터는 개인 정보가 아닙니다.

인도에서는 인터넷 키오스크를 통해

400만명 이상의 농부들이

곡물 가격, 날씨, 기타 정보를

현지 언어로 이용합니다.

인도



매일 생성되는 다량의 데이터에는 인공위성의 기상 모니터링, 제트 엔진 성능, 컴퓨터가 생성한 주식시장 거래량, 개인과 관련이 없는 센서 등의 정보가 포함됩니다. 데이터가 개인의 소유물인 경우에도 다른 사람이 액세스하지 못하고 비식별화될 가능성이 있으므로 기본적으로 관련된 개인의 신원을 나타내는 정보 없이도 저장 및 사용이 가능합니다.

오해

기업들은 개인 데이터 보호를 중요하게 생각하지 않는다.

진실

개인 데이터가 생성되면 적절하게 보호해야 합니다. 데이터 이용 기회를 확대하려면 데이터에 대한 일반인의 신뢰와 확신이 높아야 합니다. 데이터를 사용하는 기업과 조직도 데이터에 대한 주인 의식을 가지고 있어야 합니다. 이러한 노력은 책임감 있는 데이터 사용에 대한 자발적 지침을 수립하려는 업계 중심적 노력을 통해 표준화할 수 있을 것입니다. 업계의 많은 리더들은 소비자의 데이터를 공유할 경우, 데이터 수집 방법을 명확하게 이해시키기 위한 노력을 하고 있습니다. 즉 개인 정보를 익명화하도록 하는 모범 사례를 따르는 기업들이 많이 있습니다.

오해

데이터 혁신으로 모든 프라이버시가 침해될 것이다.

진실

데이터 경제의 성공 여부는 소비자의 신뢰로 결정되며, 소비자는 개인 정보가 안전하다고 느껴야 합니다. 주요 소프트웨어 개발업체는 이미 처음부터 시스템에 "프라이버시 중심 디자인(privacy by design)"이라고 하는 프라이버시 보호 기능을 담고 있습니다. 또한 익명화, 비식별화, 암호화 도구를 사용하여 데이터 침해의 피해를 최소화할 수 있도록 하고 있습니다. 개별 사용자를 식별할 수 없는 방법으로 데이터를 집계할 경우, 사용자의 신뢰나 프라이버시를 침해하지 않고 행동 패턴을 파악하기 위해 데이터를 분석할 수 있습니다. 게다가 법적 효력이 있는

개인정보 보호정책은 데이터의 노출이나 남용과 관련된 위험을 고려하여 가장 중요한 데이터(금융 또는 의료 데이터)에 대해 가장 엄격하게 프라이버시를 보호해야 합니다. 즉 기상 데이터나 업무 분석과 같이 개인 정보와 관련이 없는 데이터는 개인의 의료 데이터와 같은 수준의 보호가 필요하지는 않습니다.

오해

데이터를 완전히 비식별화할 수 없다. 데이터의 비식별화는 효과가 없다.

진실

데이터의 비식별화는 개인의 신원이 정보와 연결되는 것을 방지할 때 사용되는 프로세스입니다. 데이터를 비식별화하면 데이터를 개인과 연결시키지 않고 이용할 수 있습니다. 전문가들은 프라이버시와 데이터 품질을 모두 극대화하는 방식으로 데이터를 비식별화할 수 있는 기술을 개발했습니다.⁶¹ 전문가들에 의하면 올바르게 비식별화할 경우, 익명화된 데이터에서 개인을 다시 식별할 수 있는 위험은 대부분의 경우 1% 미만이라고 합니다.⁶²

오해

데이터를 사용하는 회사는 신뢰할 수 없다.

진실

기업들은 프라이버시 문제에 귀를 기울이고 있습니다. 오늘날 주요 기업들은 프라이버시 보호를 개선하기 위해 경쟁하고 있는 분위기입니다. 예를 들어, 전세계 스마트폰의 96.4%를 차지하는 운영체제를 운영하는 기업들은 최근에 프라이버시 설정을 개선한다고 발표했습니다. 이 기업들은 사용자에게 추가적인 제어권을 제공하고 개인의 프라이버시를 보호하기 위해 데이터 암호화를 더 많이 하고 있습니다.⁶³ 이제 주요 소프트웨어 기업들은 프라이버시 보호를 위한 노력을 광고하며 소비자에게 직접적으로 접근하고 있습니다.⁶⁴ 또한 기업들은 광고를 위해 고객의 의사소통 내용을 검열하지 않는 등 더욱 프라이버시를 존중하는 방향으로 서비스와 정책을 변경하고 있습니다.⁶⁵

미국의 주요 자동차 제조업체들은
자동 주차 및 충돌 방지와 같은 첨단 기술을 쓰기 위해

수 많은 센서

가
장착된 새로운 차량을 설계하고 있습니다.

미국



오해

개인은 자신의 데이터를 제어할 수 없다.

진실

개인이 자신의 데이터를 제어할 수 없는 것처럼 보일 수 있지만 데이터를 제어하는 데 도움이 되는 다양한 도구에 대해 알지 못하여 그럴 수 있습니다. 예를 들어, 일부 주요 소프트웨어 회사들은 자사의 웹브라우저에서 기본적으로 "추적 금지" 기능을 활성화시켜 사용자가 방문하는 웹사이트와 기업이 사용하는 제3의 광고업체들에게 자동으로 "추적 금지" 요청을 보냅니다. 이 기능은 웹사이트에게 사용자가 추적을 원하지 않는다는 사실을 알리는데 이것은 사용자가 웹 상에서 추적 당하는 것을 방지하는 데 도움이 될 수 있습니다. 또한 일부 데이터 브로커들은 소비자가 자신에 대해 수집된 정보를 보고 추가적인 데이터 수집을 하지 못하게 하고, 잘못된 정보는 수정할 수 있는 웹 페이지를 만들었습니다.⁶⁶ 집합적으로 이러한 도구들을 통해 소비자는 자신의 정보의 수집 및 사용 방법을 더욱 잘 제어하거나 특정 마케팅 용도로 사용하지 못하게 할 수 있습니다.

데이터 경제의 경제적 효과

오해

데이터 혁신으로 신규 일자리가 창출되지 않고 오히려 일자리가 없어질 수 있다.

진실

데이터 혁신은 경제 성장의 강력한 원동력이 될 수 있습니다. 사실 최근에 설문조사에 응한 미국의 최고 경영진의 61%와 유럽 최고 경영진의 58%는 회사의 직원을 채용하는 계획에 있어서 데이터 분석이 중요하다고 답변했습니다.⁶⁷ 데이터 혁신은 IT 부문의 일자리 창출에만 국한된 것은 아닙니다. 데이터와 관련된 하나의 일자리는 IT 이외의 분야에서 3개의 일자리를 창출하므로 전체 경제 차원에서 더 많은 일자리를 만들게 되는 셈입니다.

데이터 혁신은 이전의 업무를 수행할 새로운 방식을 찾는 데 도움이 되기 때문에 일자리에 변화가 있을 수 있지만, 이러한 변화는 일자리가 하나 사라질 때마다 2.6개의 일자리가 창출된다고 추산되었던 인터넷 확산의 시기와 비슷합니다.⁶⁸ 일자리 증가가 가장 많을 것으로 예상되는 분야 중 하나는 데이터 분석입니다. 데이터에 포함된 해결책을 찾아내는 것은 인간의 창의력을 통해서만 가능할 수 있습니다. 사실 전세계적으로 데이터를 이해하도록 도와주는 유능한 분석가와 데이터 관리자가 부족한 것이 사실입니다. McKinsey에 따르면 미국에서만 데이터에 숨겨진 잠재력을 파악하는 데 필요한 분석 능력을 가진 인재와 함께 데이터 경제를 이해하고 데이터 분석을 바탕으로 의사결정을 내리는 능력을 가진 관리자와 분석가는 12만 명에서 19만 명이 부족한 상태라고 합니다.⁶⁹

데이터 과학자의 2014년 평균 기본 급여는 \$120,000이며, 관리자의 경우 \$160,000라고 합니다.⁷⁰ 데이터 중심적 혁신의 가능성을 완전하게 활용하려면 전세계 기업과 정책 담당자들은 인재 부족 현상을 극복하기 위한 조치를 취해야 한다고 제안하는 전문가가 많았습니다.

브라질에서, 최대 규모의

대두 생산업체

중 하나가 소프트웨어와 데이터 분석에 투자하여 피해 대책 기술의 효율성을 높이고, 비용을 절감하며, 곡물 재배 기술을 향상시켰습니다.

브라질



오해

데이터 분석은 분석 과정을 통해 판단을 내리는 일이다.

진실

데이터를 사용하여 답변할 수 있는 질문(예: 우리 마을의 인구는 증가할까요 감소할까요?)도 일부 있지만 현명한 답변 중 다수는 명확하지 않은 경우도 있습니다. 다양한 데이터 요소가 서로 어떻게 관련이 있는지 항상 파악할 수 없기 때문입니다. 어떤 질문을 해야 할지 미리 알지 못할 수 있기 때문에 분석가들은 성공적인 질문을 하는 반복적인 과정을 통해 답을 찾는 경우가 많습니다. 이러한 이유로 차이점을 조정하고 잠재적 불일치 사항을 찾아내기 위해서는 인간의 판단과 의견을 무시할 수 없습니다.

데이터만으로는 만병통치약이 될 수 없고 기적을 만들 수 없습니다. 사실 데이터는 그 자체만으로는 가치가 거의 없습니다. 복잡하고 질서가 없거나 깨끗하게 정리되어 있지 않은 경우가 많습니다. 무엇보다 데이터를 이해하여 데이터에 포함된 연관성을 찾는 것이 어렵습니다. 데이터가 문제를 해결할 가능성은 더 빠르고 효과적인 솔루션으로 연결될 수 있는 데이터 전략을 효과적으로 실행하는지의 여부에 달려 있습니다. 또한 데이터에 대한 올바른 질문을 하는가에 따라서도 결정됩니다. 하지만 데이터를 올바른 방법으로 활용한다면 우리 사회의 가장 심각한 일부 문제들에 대한 답변을 찾고, 혁신의 기반을 마련하고, IT 중심의 일자리와 경제 성장을 위한 강력하고 새로운 원동력을 만드는 데 도움이 될 수 있습니다.

오해

데이터 혁신은 대기업만을 위한 것이다.

진실

데이터는 점점 더 흔해지고 저장 비용은 저렴해지며, 분석 도구는 더 강력해지고 저렴해지는 가운데 규모가 작은 기업들도 한때 대기업만 활용할 수 있었던 도구인 첨단 데이터 분석의 혜택을 누릴 수 있게 되었습니다. 예를 들어, Intuits의 QuickBooks Online에 있는 Trends 기능으로 소규모 기업들은 다른 Intuit 사용자들의 집합적인 지식을 활용하여 기회를 찾기 위해 수입과 지출을 비교하는 방법을 알 수 있습니다. 즉, 운영 방식에 대해 현명한 판단을 할 수 있는 것입니다. 비즈니스 인텔리전스와 분석 솔루션을 사용하는 것은 중소기업에게 보편적이지 않지만 이를 채택하는 기업이 신속하게 증가할 것으로 예상됩니다.⁷¹ 최근의 연구 결과에 의하면 데이터 분석은 소규모 기업의 60%에게 중요하다고 합니다.⁷² 기업의 최고 의사결정권자들에 따르면 직원이 50명 미만인 미국 기업의 57%와 유럽의 소규모 기업 62%가 여기에 포함된다고 합니다. 중간 규모의 기업(직원 수가 51명에서 500명 사이)의 경우, 미국 경영진의 87%와 유럽 경영진의 79%는 데이터 분석이 중요하다고 답변했습니다.

말라리아 감염 패턴에

대해 연구하는 과학자들은 최근에 케냐에서 질병 전파가 이루어진 장소를 파악하기 위해 휴대폰 데이터를 사용하여

정부의 **방역활동** 노력에 도움을 주었다고 합니다.

케냐



오해

데이터는 IT 부문에만 혜택이 되고 다른 경제 부문에는 혜택이 없다.

진실

데이터는 거의 모든 경제 분야를 전환하고 개선할 수 있는 주요 기술 전환의 중심에 있습니다. 사실 데이터 혁신이 전체 경제 분야에서 생산성을 개선할 능력을 가지고 있다고 믿는 사람들이 많습니다. 데이터 중심적인 의사결정을 하는 기업들이 생산성이 5-6% 증가했다고 보고하고 있지만 데이터 혁신으로 단 1%의 효율성만 개선되어도 전체 경제 부문에 미치는 영향은 상당할 것이며 에너지와 연료를 절약하고, 더 저렴한 비용으로 건강을 개선하며, 물리적 자산의 성과와 수명이 증가할 것입니다.⁷³ 예를 들어 항공업계의 경우, GE는 엔진 효율성을 개선하고 승객들을 더욱 효율적으로 연결하기 위해 데이터를 실시간으로 캡처할 수 있기 때문에 연료비를 1%만 절감하더라도 15년 동안 300억 달러의 비용을 절약할 것이라고 예상하고 있습니다.⁷⁴

오해

연관성은 항상 인과 관계를 의미한다.

진실

“아직 알려지지 않은 모르는 것들”을 찾는 것은 데이터에서 얻을 수 있는 여러 해결책 중에서 하나입니다. 하지만 연관성을 이해할 수 없는 경우도 있으며 연관성이 인과 관계가 아닌 경우도 있습니다. 예를 들어, 폭력 범죄와 살인 범죄의 비율은 아이스크림 판매량이 증가할 때 급증하지만 아이스크림을 사는 사람이 살인자가 될 가능성은 극히 적습니다.⁷⁵ 하지만 데이터의 일반적인 관계를 찾는 것은 데이터에서 우리가 발견할 수 있는 가장 가치 있는 통찰력 중 하나입니다. 일상 생활에서 점점 더 많은 흥미로운 상관관계와 인과 관계를 찾고 있으며 중요성도 더 커지고 있습니다. 단순한 상관관계와 인과 관계 사이의 차이를 찾는 능력은 데이터를 활용하기 위해 인과 관계에서 연관성을 구분하는 방법을 이해하고 비개연성 테스트에 부합하지 않는 결과를 삭제할 자격을 갖춘 데이터 과학자가 필요한 이유 중 하나입니다.

오해

데이터 혁신을 위해서는 막대한 예산이 필요하다.

진실

데이터를 활용하기 위해 반드시 막대한 예산이 필요한 것은 아닙니다. 약간의 투자만 필요하며 이러한 투자는 상당한 수익으로 이어질 수 있습니다. 예를 들어, 데이터에 대해 더 전체론적 접근방법을 사용하는 조직은 향후 4년 동안에만 가장 큰 수익을 얻을 것이며 추가적인 매출, 비용 절감, 생산성 개선 등으로 막대한 “데이터 배당금”을 달성할 것입니다.⁷⁶ 사실, 작은 데이터 집합과 이미 보유하고 있는 데이터를 사용하는 것부터 시작할 수 있습니다. 1주일 분량의 금융 거래 데이터를 검토하여 사기 행위를 찾는 것이 과거 5년간의 데이터를 검토하는 것보다 더 간편하게 동향을 파악하는 방법일 수 있습니다. 데이터 경제 분석을 효과적으로 활용하는 기업은 경쟁사보다 수익을 26% 많이 올리며, 직원과 물리적 자산을 통해 매출을 9% 더 창출하고, 12% 높은 시장 가치율을 달성한다는 분석이 있습니다.⁷⁷

소비하는 에너지보다 생산하는 에너지가 많은 건물

을 설계하고 있습니다.

아랍 에미리트



데이터 신뢰성

오해

데이터를 통해 얻은 해결책은 항상 정확하다.

진실

데이터를 통해 얻은 해결책은 기본적인 데이터가 정확하고 도구가 효과적이며, 왜곡된 결과를 얻지 않도록 유능한 데이터 과학자가 모델 개발에 참여할 경우에만 정확합니다. 예를 들어 보스턴의 혁신적인 Street Bump 앱은 가속도계와 GPS 위치를 결합시켜 도로의 돌출부에 차량이 충돌할 때마다 사용자로부터 구멍이 위치를 찾기 위해 스마트폰을 사용했습니다. 하지만 스마트폰 사용자는 젊고 경제력이 있는 경우가 많으므로 이 앱에서 얻은 데이터는 도시 전역에서 구멍이 위치 관련 데이터를 정확하게 파악하지 못할 수 있으며 구멍이 위치에 대한 정보가 왜곡될 위험이 있습니다. 그러한 이유로 데이터 과학자들이 관여하여 통계학적 관련성을 확인하고 데이터와 관련된 왜곡을 방지해야 합니다.

오해

직관적으로 내린 결정은 항상 올바른 결정이다.

진실

우리는 태어날 때부터 직감에 따라 의사결정을 내리는 방법을 배웁니다. 사실 전세계 관리자의 19%는 자신이 거의 본능과 직감에만 의존하는 "본능적인 의사결정자"라고 묘사합니다.⁷⁸ 하지만 오늘날 데이터는 정보를 바탕으로 하여 더 정확하고, 신속하며 더 효과적인 결정을 내리는 데 도움이 될 수 있습니다. 다양한 업계의 IT 의사 결정자를 대상으로 한 어떤 설문조사에 따르면 응답자 중에서 59%는 데이터 기술에 대한 투자를 이끄는 주요 목적이 더 올바른 의사결정을 하는 것이라고 답했습니다.⁷⁹ 이것은 직감에만 국한된 것이 아니라 일상 생활에서도 직관적 결정을 내리기 때문에 중요한 점입니다. 경우에 따라 더 신속하고 더 나은 결정을 내리는 데 도움이 될 수 있는 도구를 매우 유용하게 활용할 수 있습니다. 예를 들어 차량 충돌 사고의 93%는 인재가 원인이지만 주변 환경에 대한 방대한 양의 데이터와

실시간 분석 자료를 수집하는 새로운 차량 센서와 자동화된 의사결정이 결합될 경우(완전히 자율화되지 않더라도) 차량 관련 부상과 사망률을 50%나 줄일 수 있습니다.⁸⁰

오해

데이터의 양이 많을수록 좋은 데이터이다.

진실

데이터 집합이 크다고 항상 좋은 것은 아닙니다. 사실 크기는 별로 중요하지 않은 요인 중 하나에 불과합니다. 올바른 도구를 사용하여 분석한다면 소량의 데이터라도 날카로운 통찰력으로 이어질 수 있습니다. 이러한 통찰력을 찾아내는 것은 기본적인 데이터의 품질과 사용한 도구의 품질에 의존하는 경우가 많습니다. 하지만 데이터가 많을수록 더 많은 진실을 찾고 데이터 집합이 클수록 객관적이라고 믿는 경우가 있습니다. 동네 가게가 언제 문을 닫으며, 차량이 마지막으로 오일 교환을 한 후 몇 마일을 달렸는지, 이번 달 전화요금은 얼마인지 등과 같이 가장 즉각적인 영향을 줄 수 있는 단순 데이터를 이용하는 경우가 있습니다. 질문에 따라 웹 페이지, Excel 스프레드시트 또는 CRM 데이터베이스에 포함된 데이터는 적을 수도 있지만 빅 데이터만큼의 답변을 제공할 정도로 강력할 수 있습니다. 가장 중요한 것은 데이터의 크기에 관계 없이 강력한 데이터를 생성하여, 안전하게 저장하고, 데이터에 대한 액세스 권한을 가지고, 데이터를 처리할 능력을 갖추어 문제 해결에 필요한 경우 활용할 수 있어야 한다는 것입니다.

기상 데이터

는 우산을 준비할지 예측하는 데만 사용하지 않고 작물 수율과 독감 발생률을 예측하는 데에도 사용할 수 있으며 모로코에서 몇 개의 담수 공장을 지을지 예측하는 데에도 도움이 됩니다.

모로코



오해

체계화되지 않은 데이터는 쓸모 없다.

진실

체계화되지 않은 데이터도 유용할 수 있습니다. 사실 가장 강력한 데이터 중 일부는 체계화된 구조에 완벽하게 맞지 않을 수도 있습니다. 즉, 데이터는 일반 문서나 엑스레이 사진과 같이 체계화되지 않을 수 있습니다. 중요한 것은 데이터를 가장 유용한 형태로 전환하는 것입니다. 평균적인 포춘 1000대 기업이 데이터의 유용성을 10%만 향상시켜도 매출을 20억 달러 이상이나 향상시킬 수 있습니다.⁸¹

오해

데이터는 수집할 당시의 원래 목적으로만 사용해야 한다.

진실

클라우드 기술 덕분에 데이터 저장 비용이 크게 감소하여 데이터를 처음 사용한 후에 새로운 데이터를 위한 공간을 만들기 위해 삭제할 필요가 없어졌습니다. 그 결과, 데이터를 사용해도 다른 자원과 마찬가지로 고갈되지 않습니다. 데이터는 재사용이 가능하고 다른 데이터 집합과 결합할 수 있고 여러 차례 사용하여 데이터를 처음에 생성할 때 알지 못했던 문제점에 대한 해답도 찾을 수 있는 재생 가능한 자원입니다. 예를 들어 기상 데이터는 우산을 준비할지 예측하는 데만 사용하지 않고 작물 수율과 독감 발생률을 예측하는 데에도 사용할 수 있으며 모로코에서 몇 개의 담수 공장을 지을지 예측하는 데에도 도움이 됩니다.

21 세기 데이터 혁신

오해

데이터의 가장 큰 혜택은 먼 미래가 되어야 누릴 수 있다.

진실

데이터 혁신은 먼 미래에도 혜택을 주겠지만 막강한 효과는 오늘날 거의 모든 분야에서 확인할 수 있습니다. 이미 막대한 양의 일상적 데이터가 더 정확한 기상 정보를 제공하는데 사용되며, 수 많은 GPS 신호를 통해 얻은 데이터는 정체 시간과 도착 시간을 예측하는 데 사용되며, 수 많은 의료 데이터 요소는 잠재적인 질병의 원인을 파악하는 데 사용되고, 방대한 재무 데이터는 신용카드 사기 범죄를 예방하는 데 유용하게 사용되고 있습니다.

사실 Economist Intelligence Unit는 설문조사 응답자들에게 지난 5년 동안 데이터가 회사에 미친 영향에 대해 설명하도록 했더니 거의 10%가 데이터를 통해 업무 방식이 완전히 바뀌었다고 답변했다고 합니다.⁸² 응답자 중에서 46%는 데이터가 업무 결정을 주도하는 중요한 요소가 되었다고 답변했다고 합니다. 어느 새로운 연구는 데이터가 새로운 기회와 해결책을 찾게 해주므로 데이터를 잘 활용하면 향후 4년 동안에만 전세계적으로 "데이터 배당금"으로 1조 6000억 달러까지 혜택을 볼 수 있다는 사실을 제시하고 있습니다.⁸³ 현재 상황에 대해 조사하는 또 다른 설문조사에서는 미국의 최고 경영진 중에서 33%, 유럽의 경우 24%가 회사 성장 중에서 10% 이상이 데이터 분석과 관련이 있을 것이라고 예상했습니다.⁸⁴ 향후 5년을 예측한 경우에도 미국의 최고 경영진의 58%와 유럽의 43%가 동일하게 예상하고 있었습니다.

2004년 남아시아에서 쓰나미가 발생한 후 인도네시아의 어부에게 휴대폰을 제공했습니다. 이들은 난생 처음으로 생선의 시장 가격에 대한 데이터를 얻게 되었기 때문에 수입이

30%나 증가했다고 합니다.

인도네시아



오해

데이터는 너무 과대 포장되었다.

진실

인류의 문제를 해결하기 위해 데이터를 활용한 것은 새로운 일이 아닙니다. 근대 문명이 시작된 이래로 더 나은 의사결정을 위해 데이터를 사용해 왔으며 더 많은 인구에게 혜택을 주는 휴경 기술, 글로벌 차원의 무역을 가속화한 항해 기술, 콜레라로 인한 사망을 예방하게 한 의료 기술 등으로 이어졌습니다. 하지만 과거에 데이터는 저장 비용이 많이 들고 다루기 힘든 희귀한 자원이었습니다. 오늘날 달라진 것은 데이터의 양이 많아졌으며, 저장 비용은 크게 떨어지고 데이터를 관리하는 도구가 훨씬 강력해졌다는 것입니다. 그 결과, 우리가 새로운 문제에 직면했을 때 강력하고 새로운 데이터 분석 도구는 급증하는 데이터를 분류하여 중요한 정보를 발견하고 가장 어려운 문제점에 대해 예기치 않은 해결책을 찾는 데 도움이 될 수 있습니다.

오해

IT 중심의 경제 성장의 시대는 끝났고 데이터 혁신은 생산성을 높일 수 없다.

진실

IT 혁신과 이로 인한 경제 성장, 일자리 창출, 전세계 생활 수준의 향상 등의 능력은 1990년대 1%-2%의 생산성 증가에서 입증된 바와 같이 생산성 향상에 바탕을 두고 있습니다.⁸⁵ Northwestern University의 저명한 경제학자는 IT 혁신이 가져다 줄 가장 위대한 혜택은 아직 모습을 드러내지 않았다고 말했습니다.⁸⁶ 하지만 생산성의 시대는 아직도 건재하고 있습니다. 사실, 데이터 중심적인 의사결정을 하는 기업들은 생산성이 5%-6% 증가했다고 보고하고 있습니다.⁸⁷ 이러한 데이터 기회는 미국의 생산성을 1.5% 정도 향상시키는 데 그친다고 해도 향후 20년 동안 평균 국가 소득을 30%나 향상시킬 정도의 비용을 절감해줄 것으로 예상됩니다.⁸⁸ 데이터를 더욱 효율적으로 활용하여 광범위한 업계에서 효율성을 1%만 올리더라도 2030년에는 전세계 GDP에 약 15조 달러를 추가할 수 있을 것이라고 경제학자들은 내다보고 있는데 이것은 전세계 경제에 미국의 경제 규모가 또 하나 추가되는 것과 같은 수준입니다.⁸⁹

더 지능적인 도시

를 만들고,

정부 서비스를 개선하며,

더 지속 가능한 교통

솔루션을 제공하고 있습니다.

스페인



데이터 혁신의 글로벌 차원의 혜택

오해

미국만 데이터의 혜택을 누리고 있다.

진실

전세계적으로 중요한 과제를 해결하기 위해 이미 데이터가 활용되고 있습니다. 그 예는 다음과 같습니다.

- + 캐나다의 과학자들은 1,000개 이상의 데이터 요소를 1초 만에 추적하여 이례적으로 안정적인 활력 징후를 가지고 태어난 미숙아에게 고열이 발생할 위험이 높다는 사실을 확인하여 의사들이 조기에 조치를 취하고 생명을 구할 수 있도록 했습니다.
- + 브라질에서 최대 규모의 대두 생산업체 중 하나가 소프트웨어와 데이터 분석에 투자하여 피해 대책 기술의 효율성을 높이고, 비용을 절감하며, 곡물 생산성을 높였습니다.
- + 바르셀로나는 데이터를 활용하여 더 지능적인 도시를 만들고, 정부 서비스를 개선하며, 더 지속 가능한 교통 솔루션을 제공하고 있습니다.⁹⁰
- + 아랍 에미리트에서는 새로운 데이터 도구를 사용하여 세계 최초로 소비하는 에너지보다 생산하는 에너지가 많은 건물을 설계하고 있습니다.⁹¹
- + 인도의 농민들은 종자, 위성, 센서, 트랙터에서 수집한 정보를 사용하여 재배할 작물, 재배 방법, 식품의 신선도 추적 방법, 변화하는 기후에 적응하는 방법 등을 결정하고 있습니다.⁹²

오해

개발도상국은 아직 데이터 분석을 활용할 준비가 되지 않았다.

진실

데이터 혁신과 이를 통한 혜택은 전세계적으로 일반적인 현상이 되었습니다. 데이터의 가장 중요한 혜택과 가장 큰 기회 중에서 일부는 기술이 뒤떨어지기 쉬운 개발도상국에 존재하기도 합니다. IDC에 따르면 디지털 세상에서 신흥 시장은 2017년에 선진국 시장을 앞지를 것이며 2012년과 2020년 사이에 디지털 세상은 36%에서 62%까지 확장될 것이라 합니다.⁹³ 마찬가지로 개발도상국의 NGO가 실시간 설문조사에 의하면 응답자 중에서 90% 이상은 최종 수혜자를 도와줄 더 나은 통찰력을 제공하는 데 있어서 데이터 분석이 가장 중요한 도구가 될 것으로 생각한다고 합니다.⁹⁴

개발도상국에서의 데이터 혁신의 예는 다음과 같습니다.

- + 말라리아 감염 패턴에 대해 연구하는 과학자들은 최근 케냐에서 질병 전파가 이루어진 장소를 파악하기 위해 휴대폰 데이터를 사용하여 정부의 방역활동 노력에 도움을 주었다고 합니다.⁹⁵
- + 2004년 남아시아에서 쓰나미가 발생한 후 인도네시아의 어부에게 휴대폰을 제공했습니다. 이들은 난생 처음으로 생선의 시장 가격에 대한 데이터를 얻게 되었기 때문에 수입이 30%나 증가했다고 합니다.⁹⁶
- + 페루에서는 개발 때문에 유적지가 위험한 상태에 있습니다. 페루는 항공 기술과 영상을 서로 통합하는 강력한 소프트웨어를 사용하여 위험에 처한 유적지를 찾고, 모니터링하고, 보호하기 위해 3차원 데이터 요소 클라우드를 만들었습니다.⁹⁷
- + 베트남에서는 문화 유적을 보존하기 위해 3D 스캐너를 사용하여 향후 5년 동안 4만개의 역사 유물을 디지털화한다는 목표를 세웠습니다.⁹⁸
- + 인도에서는 인터넷 키오스크를 통해 400만명 이상의 농부들이 곡물 가격, 날씨, 기타 정보를 현지 언어로 사용하고 있습니다. 이들은 첨단 분석 및 모바일 기술을 사용하여 개별 농장의 데이터를 추적하고 분석한 후 비료와 종자와 같은 농부들의 요구사항을 바탕으로 용품을 제공할 수 있게 되었습니다.⁹⁹

베트남에서는 문화 유적을 보존하기 위해
3D 스캐너를 사용하여 향후 5년 동안

40,000개의 역사 유물을

디지털화한다는 목표를 세웠습니다.

베트남



데이터 규제에 대한 정부의 역할

오해

데이터는 저소득층과 취약 계층이 직면한 불이익을 보강하는 능력을 통해 배타적 도구로 사용될 것이다.

진실

데이터가 특정인이나 계층에게 불공평하게 사용될 것이라는 우려를 해소하는 것은 업계와 정부가 최우선적으로 해결해야 하는 과제입니다. 동시에 불공평한 차별을 근절하고 그룹들에게 권한을 부여하기 위해 데이터를 사용해야 하는 경우를 간과해서는 안 됩니다. 사실 데이터를 책임감 있게 사용할 경우 일자리, 금융, 교육, 기획 등의 활용을 막는 기존의 차별관행을 찾아내는 새롭고 강력한 도구가 될 수도 있습니다.¹⁰⁰

오해

정부는 아무런 역할을 하지 않는다.

진실

원격 컴퓨팅, 저장소, 분석, 소프트웨어 서비스 등의 급성장으로 새로운 정책 문제가 나타나고 있습니다. 대부분의 조약, 법률, 규정이 만들어질 당시에는 이러한 가능성을 예측할 수 없었습니다. 이러한 모호한 상황은 정부와 데이터 경제의 성장에 있어서 문제가 될 수 있습니다. 이러한 기술이 급성장하면서 프라이버시와 보안의 균형, 자유 무역 흐름, 기본 연구, 인재 개발 등과 같은 오랜 정책 문제에 대해 다시 고려해야 합니다.

사실 정책 입안자들이 데이터 혁신의 잠재력을 완전히 파악하고 생산성, 경제 성장, 소비자 혜택의 새로운 움직임을 가속화하기 위해 취할 수 있는 확고한 절차가 몇 가지 있습니다. 정책 입안자들은 사회의 가장 심각한 문제들을 해결할 수 있는 데이터 솔루션을 발전시키고 교육, 식량 생산, 생활 영위, 에너지 절약, 여행, 경제 성장 등의 방식을 개선할 수 있는 데이터 솔루션의 발전에 투자할 수 있습니다. 또한 국제법 집행 협력을 위한 명확한 규칙을 세울 수도 있습니다. 기술 업계가 투자하고, 혁신하고, 거래하고 성장하는 능력에 영향을 주는 시장의 자유를 촉진하는 방법도 있습니다. 물리적 경계에 방해 받지 않고 데이터가 자유롭게 흐르도록 할 수도 있으며, 인재를 확보하고 데이터 과학자들의 능력 차이를 극복하도록 도와줄 수도 있습니다. 데이터의 장점을 기업체와 소비자에게 전달하는 것을 가속화할 혁신의 기반을 지원할 수도 있습니다. 정책 입안자들은 인간의 능력과 집념을 활용하고, 혁신과 투자를 활용하며, 용량과 기능을 확장하는 실용적인 정책을 통해 데이터 중심적 경제가 제공할 수 있는 혜택을 극대화하는데 일조할 수 있을 것입니다.

페루에서는 개발 때문에 유적지가 위험한 상태에 있습니다. 페루는 항공 기술과 영상을 서로 통합하는 강력한 소프트웨어를 사용하여 위험에 처한 유적지를 찾고, 모니터링하고, 보호하기 위해

3 차원 데이터 요소 클라우드를 만들었습니다.

페루



오해

데이터 지역화는 프라이버시를 보호하고 보안을 개선하는 데 도움이 된다.

진실

데이터를 특정 국가 내에만 저장하도록 제한하는 데이터 지역화가 프라이버시와 보안을 개선할 수 있다고 생각될 수 있을 것입니다. 하지만 인터넷과 같은 글로벌 차원의 기술과 서로 다른 데이터 센터 사이를 이동하는 데이터 덕분에 오늘날의 기술 혜택을 얻을 수 있었습니다. 사실 국경을 초월한 인터넷 트래픽은 2005년 이후 50%나 증가했습니다.¹⁰¹ 데이터가 국경을 넘어 자유롭게 이동하도록 한 결과 규모가 작은 기업들도 제품, 서비스, 아이디어를 전세계에 판매하고 소싱하기 때문에 전세계적인 활동 무대가 되고 있습니다. 하지만 전세계 정부는 데이터의 자유로운 흐름을 제한하거나 시장 활동의 조건으로서 데이터 서버를 관할 구역 내에 비치하도록 하는 정책을 고려하는 경우가 종종 있습니다.¹⁰² 이러한 제한은 데이터 혁신을 통해 얻을 수 있는 막대한 효율성과 경제적 혜택은 물론 급증하는 데이터에서 유용한 정보를 찾기 위해 다양한 장소의 다양한 데이터 집합을 통합하는 능력을 저해하고 있습니다. 또한 자연 재해나 기술 오류가 발생할 경우에 대비하여 가치 있는 데이터를 여러 장소에서 백업하는 것을 막아 보안을 해칠 수도 있습니다. 데이터가 제공할 수 있는 모든 혜택을 누리기 위해서 모든 국가의 법률이 같을 필요는 없지만 서로 호환될 필요성이 있습니다. 국경을 초월하여 데이터의 자유로운 흐름을 가능하게 하는 것은 데이터 중심적 혜택을 활성화하는 기본적인 개념입니다.

오해

데이터를 보호할 수 있는 유일한 방법은 정부가 데이터 보호에 관여할지의 여부이다.

진실

기존의 정부 프라이버시 규칙은 데이터의 안전과 개인정보를 보호하기 위해 강력하고 혁신적인 프라이버시 기술 및 자발적인 업계 모범 사례와 통합될 수 있습니다. 이와 반대로 데이터를 현지에서 저장하도록 함으로써 프라이버시 보호와 보안을 실행하려는 정부 방침은 기술혁신을 막고 데이터 혁신으로 얻을 수 있는 사회적 혜택을 제한할 수 있습니다.



디지털 담론

데이터 언어에 대한 이해

풍부한 데이터

예전에는 데이터가 흔하지 않았지만 저장 비용이 급격히 저렴해지고 새로운 방식으로 가치를 창출하게 되면서 의미 있는 형태의 방대한 디지털 데이터를 완전히 새로운 방식으로 수집할 수 있게 되었습니다.

적응형 인텔리전스

적응형 인텔리전스는 통계학적 데이터 처리에만 관련된 것이 아니라 특정 영역의 인텔리전스를 포함하는 데이터와 통합하는 컴퓨터 인텔리전스입니다. 지능적인 행동 모델을 전문 지식과 통합함으로써 시스템은 예제에서 더 많은 것을 학습하고 새로운 상황에 적응할 수 있습니다.

알고리즘

알고리즘은 문제 해결을 위해 데이터를 분석하는 데 수학을 활용하는 단계별 절차 또는 일련의 컴퓨터 지침입니다. 알고리즘은 거의 모든 소프트웨어 프로그램에서 사용되고 있습니다.

엠펬트 인텔리전스

엠펬트 인텔리전스는 일반적이고 비용이 저렴한 컴퓨팅 기술이 모든 주변 환경에 내재되면서 기술이 더욱 더 가시화되지 않게 되는 비전을 말합니다. 엠펬트 인텔리전스는 모든 사물이 서로 연결되고 지능화되고 반응성을 가지면서 우리가 필요로 할 때마다 사용할 수 있게 될 것입니다.

분석

분석은 통계학 및 소프트웨어 기반 알고리즘을 동시에 사용하여 데이터에서 의미 있는 해결책, 패턴 및 관계를 찾는 것을 말합니다.

이상 탐지

이상 탐지는 데이터 집합에서 예상된 패턴과 일치하지 않는 데이터를 파악하는 것입니다. 이상은 데이터의 이상치, 예외 또는 이물질이라고도 하며 중요하고 유용한 정보를 제공하는 경우가 있습니다.

익명화

데이터를 익명화하는 것은 개인 식별로 이어질 수 있는 모든 개인 식별 정보를 제거하는 일입니다.

잘못된 데이터

잘못된 데이터는 누락되었거나 잘못된 데이터를 말합니다. 잘못된 주소와 같이 간단할 수도 있지만 포춘 1000대 기업들은 잘못된 데이터 때문에 매년 거액의 경제적 부담을 안고 있습니다.

빅 데이터

빅 데이터는 새로운 해결책을 발견하고 의사결정을 개선하기 위해 컴퓨터 분석을 방대한 데이터에 적용하는 프로세스를 말하는 포괄적 용어입니다. 양이 너무 많으며, 다양하고, 기존의 데이터 처리 도구로는 처리하기 어려운 속도로 변하는 데이터 집합을 말하는 경우도 종종 있습니다.



브론토바이트

브론토바이트는 비정상적으로 방대한 양의 데이터에 대한 비공식적 측정 용어입니다. 브론토바이트는 일반적으로 1,000 요타바이트와 같다고 간주되며 숫자 0이 27개가 붙어서 표시됩니다.

비즈니스 인텔리전스(BI)

비즈니스 인텔리전스는 원시 데이터를 업무 성과와 의사결정을 개선할 수 있는 운영적 통찰력으로 변환하는 일련의 기술 및 애플리케이션을 말합니다.

군집 분석

군집 분석은 서로 비슷한 데이터 개체를 파악하여 데이터 사이의 차이점과 유사점을 더욱 잘 이해하기 위해 함께 묶는 프로세스입니다.

클라우드

클라우드에는 모든 애플리케이션, 서비스 또는 원격으로 호스트되는 데이터를 칭하는 광범위한 용어입니다. 일반적으로 컴퓨팅 또는 저장소 리소스에 대한 일반적인 주문형 네트워크 액세스를 활성화하기 위해 서로 네트워크화된 다수의 원격 서버 그룹을 통해 클라우드가 가능해졌습니다.

인지 컴퓨팅

인지 컴퓨팅은 인간 두뇌의 작용 방식을 흉내내기 위해 다량의 정보를 기계 학습 기술, 패턴 인식 기술, 자연 언어 처리와 통합하는 과정입니다. 이러한 시스템은 정보 소스와 컨텍스트 및 통찰력을 통합하여 학습하고 사람들과 대화할 수 있는 경우가 많습니다.

컴퓨터 생성 데이터

컴퓨터 생성 데이터는 컴퓨터 로그 파일, 인공지능 원격 측정 데이터, 산업용 기계에서 수집한 센서 데이터 등 인간이 개입하지 않고도 컴퓨터에서 자동으로 생성되는 데이터를 말합니다.

다크 데이터

다크 데이터는 보관하고 있지만 분석이나 처리되지 않았으며 어떠한 방식으로든 방치되거나 활용하지 않은 것으로 간주되는 체계화되지 않고 개발되지 않은 데이터로 구성되어 있습니다.

데이터

데이터는 디지털 방식으로 처리하여 상태, 개체 또는 아이디어를 표현할 수 있는 체계화되지 않은 원시 형태의 정보입니다. 데이터의 일반적인 유형에는 매출액, 마케팅 연구 결과, 기상 측정기의 측정치 또는 도시 목록과 도시의 인구 등이 포함됩니다. 우리는 매일 약 2500경 바이트의 데이터를 생성하고 있습니다.

데이터 집계

데이터 집계는 고차 분석을 제공할 목적으로 여러 소스에서 데이터를 모으는 행동을 말합니다.

데이터 집계 도구

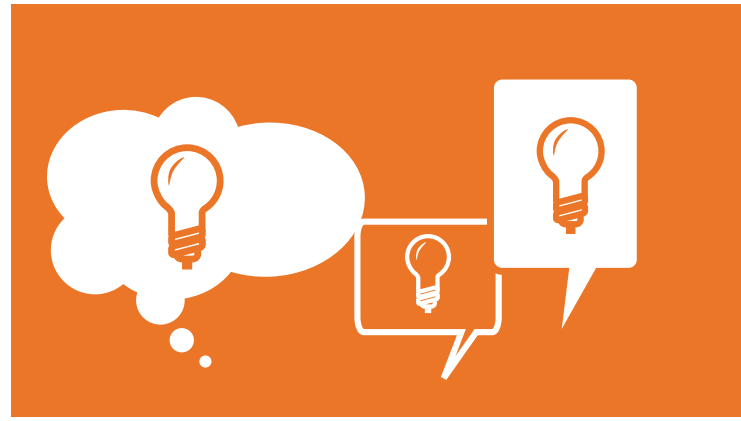
데이터 집계 도구는 여러 소스에 분산된 데이터를 하나의 새로운 데이터 집합으로 변환합니다.

데이터 분석

데이터 분석은 데이터로부터 유용한 정보, 통찰력 또는 의미를 얻기 위해 데이터를 변환하고 모델링하는 하나의 방법으로서 소프트웨어를 활용하는 것입니다. 또한 숨겨진 패턴이나 알려지지 않은 상관 관계를 찾고 의사 결정에 활용하기 위해 사용하는 경우도 종종 있습니다.

데이터 분석가

데이터 분석가는 데이터를 준비하고, 정화하고, 처리하는 일을 담당하는 사람입니다.



데이터 아키텍처 및 설계

데이터 아키텍처는 데이터를 처리하고 저장하고 사용하고 액세스하는 방법을 설계하고 조직화하기 위해 새로운 시스템의 기획 단계에서 실시합니다. 특정 데이터를 서로 연결하여 작동시킬 방법을 처음부터 정의함으로써 데이터가 시스템 전체에서 보호되도록 데이터가 흐르는 방식과 이를 제어할 방법을 정할 수 있습니다.

데이터베이스

데이터베이스는 내부의 데이터를 신속하게 검색하고 액세스하고 업데이트할 수 있도록 설계한 디지털 데이터의 체계적인 대규모 집합을 말합니다.

데이터 센터

데이터 센터는 네트워크화된 서버와 데이터 저장소가 위치하는 물리적 시설로서 일반적으로 원격 저장과 원격으로 액세스하여 다량의 데이터를 처리하는 데 사용됩니다. 전세계에 50만 개의 데이터 센터가 있는 것으로 추산되며 이 중에서 다수가 클라우드를 형성합니다.

데이터 정리/정화

데이터 정리는 중복을 삭제하고, 오류를 수정하고, 누락된 데이터를 추가하고, 손상된 데이터를 제거하며 일관성을 보강하기 위해 원시 데이터를 검토하고 수정하는 과정입니다.

데이터 중심적 의사결정

데이터 중심적 의사결정을 하는 기업들은 중요한 결정을 뒷받침하기 위해 데이터를 수집하고, 처리하고 분석합니다. MIT(Massachusetts Institute of Technology) Sloan School of Management의 경제학자인 Eric Brynjolfsson 교수가 실시한 연구에 따르면 데이터 중심적 의사결정을 하는 기업들은 생산성을 5% - 6% 향상할 수 있다고 합니다.

데이터 마이닝

데이터 마이닝은 강력한 컴퓨터 알고리즘을 사용하여 대규모 데이터 집합에서 패턴이나 지식을 찾는 과정입니다.

데이터 품질

데이터 품질은 사용자에게 데이터의 가치를 정의하는데 사용되는 측정 기준입니다. 또한 의사결정, 계획 또는 운영을 위한 데이터의 신뢰성, 효율성, 가치를 말합니다.

데이터 과학

데이터 과학은 복잡한 문제를 해결할 수 있는 의미 있는 통찰력을 발견하기 위해 통계학, 데이터 시각화, 컴퓨터 프로그래밍, 데이터 마이닝, 기계 학습, 데이터베이스 공학 등을 통합하는 규율입니다.

데이터 과학자

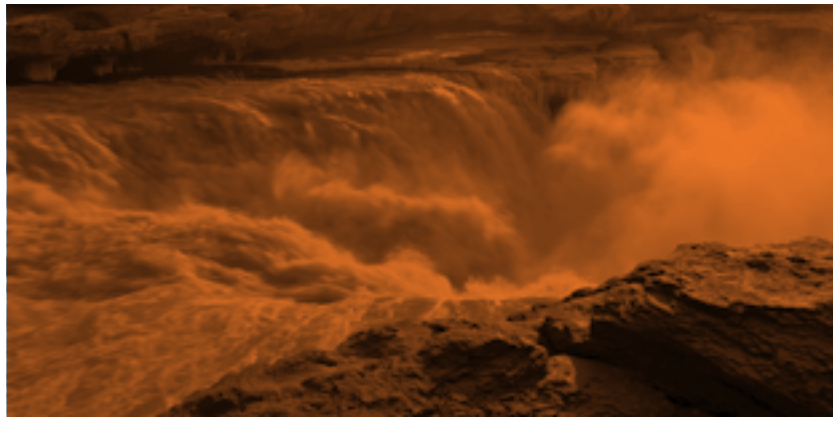
데이터 과학자는 인간의 견해, 수학적 노하우 및 기술 도구를 결합시켜 컴퓨터 알고리즘을 발전시키고 배치함으로써 데이터에 의미를 부여하는 사람을 말합니다.

데이터 보안

데이터 보안은 데이터가 파괴되거나, 남용되거나, 무단 액세스되는 것을 방지하는 것을 말합니다. 적절한 데이터 보안 방법은 데이터 누출을 방지하고 데이터 무결성을 확보하며, 프라이버시를 보호하는 데 도움이 될 수 있습니다. 따라서 인재, 프로세스, 기술을 모두 중시해야 하는 경우가 많습니다.

데이터 집합

데이터 집합은 하나의 단위로 조작될 수 있으며 표 형식으로 된 독립된 요소로서의 정보 집합을 말합니다.



데이터 소스

데이터 소스는 데이터베이스, 스프레드시트 또는 데이터 스트림과 같이 데이터를 얻는 주요 장소입니다.

데이터 시각화

데이터 시각화는 의미를 추론하거나 정보를 더욱 효과적으로 전달하기 위해 데이터를 시각적으로 표현하는 것을 말합니다.

데이터 가상화

데이터 가상화는 데이터가 위치하는 장소나 포매팅되는 방식에 대한 기술적 상세정보를 알지 못해도 다양한 데이터 소스를 검색하고 조작하는 과정입니다.

비식별화

데이터를 비식별화하는 것은 개인을 특정 정보와 연결시키는 정보를 제거하는 과정을 말합니다.

파괴적인 변화

파괴적인 변화는 사회와 기업의 근본적인 대규모 변화로서 업무 수행, 생활, 여가 활동, 가치 창출 등에 대해 완전히 새로운 방식을 정하는 전환적 신기술을 통해 가능해지는 경우가 많습니다. 데이터 혁신은 파괴적인 변화를 가능하게 하는 기술로 묘사되기도 합니다.

엑사바이트

엑사바이트는 데이터 저장의 단위로서 10의 18 제곱을 말합니다. 의미상으로 볼 때 매일 우리는 1 엑사바이트의 새로운 정보를 생성합니다.

하둡

하둡은 분산된 파일 시스템에서 막대한 양의 데이터를 처리하고 저장하도록 만들어진 오픈 소스 소프트웨어 프레임워크입니다.

사물 인터넷

사물 인터넷은 사람들의 손에서부터 세상과 연결되는 혁신을 확장하기 위해 일반적인 장치가 더욱 지능적으로 만들어져 인터넷에 연결되는 세상을 말합니다. 서로 연결될 수 있는 모든 사물은 연결되므로 만물 인터넷이라고 말하기도 합니다. 현재는 세상에서 연결될 수 있는 사물의 1% 밖에 연결하지 못한 상황이라고 합니다. 2020년에는 500억 대의 장치가 인터넷으로 연결될 것으로 예상되고 있습니다.

레거시 시스템

레거시 시스템은 낙후된 컴퓨터, 애플리케이션 또는 기술이지만 필요한 기능을 적절하게 수행하기 때문에 계속 사용됩니다.

기계 학습

기계 학습은 컴퓨터가 특정 패턴이나 사건이 발생할 때 취하는 행동 경험에서 "학습"할 목적으로 데이터를 분석할 수 있도록 알고리즘을 사용하는 것입니다.

메타데이터

메타데이터는 데이터에 대한 데이터입니다. 데이터 제작자, 데이터 생성 날짜, 파일 크기, 마지막 수정 날짜 등과 같이 데이터에 대한 기본 요약 정보가 포함됩니다.

이상값 감지

이상값은 대규모 데이터 집합 내의 일반적인 평균값에서 크게 벗어나는 데이터입니다. 나머지 데이터와 숫자상 차이가 있는 이상값은 잘못된 것이 있음을 암시하는 것이므로 추가적인 분석을 해야 합니다. (이상 탐지를 참조하십시오.)



패턴 인식

패턴 인식은 데이터 내의 패턴을 찾고 파악하는 과정입니다. DNA 서열에서 반복되는 패턴을 파악하는 것처럼 간단할 수 있으며 하나의 이벤트를 다른 이벤트와 연결하는 패턴의 존재 여부를 발견하기 위해 두 개의 데이터 세트가 상호작용하는 방법에서 패턴을 찾거나, 사진에 포함된 숫자를 찾는 것처럼 기계 학습의 도움으로 더 복잡한 패턴을 찾을 수도 있습니다.

페타바이트

페타바이트는 10의 15 제곱 또는 100만 기가바이트로 표시되는 저장소의 용량 단위입니다. 페타바이트는 미국 의회 도서관에 포함된 데이터의 약 4배에 해당하는 양입니다.

예측 분석

예측 분석은 하나 이상의 데이터 집합에 소프트웨어 알고리즘을 사용하여 동향이나 미래의 이벤트를 예측하는 것입니다. 현재 상황에서 얻은 데이터를 과거의 데이터와 비교할 수 있다면 미래를 예측하는 데 도움이 되는 경우가 종종 있습니다.

예측 모델링

예측 모델링은 오늘의 이벤트와 과거의 이벤트를 비교하여 동향, 미래 행동 또는 결과를 예측할 수 있을 가능성이 가장 큰 모델을 만드는 과정입니다.

실시간 데이터

실시간 데이터는 생성된 목적에 따라 작용하는 데이터입니다. 밀리 초 이내에 생성되고, 처리되고, 저장되고 분석되는 경우가 많습니다. 실시간 데이터에는 주가에서 전자 제어식 제동 장치에 사용된 바퀴의 속도에 이르기까지 모든 것이 포함될 수 있습니다.

추천 엔진

추천 엔진은 기계 학습 기술에서 파생된 다양한 패턴을 바탕으로 사용자를 위해 추천하거나, 제안하거나, 개별화할 수 있는 컴퓨터 알고리즘입니다.

회귀 분석

회귀 분석은 2가지 이상 변수의 관계를 예측하기 위해 데이터를 사용하는 통계학적 과정입니다.

위험 분석

위험 분석은 프로젝트, 행동 또는 의사결정의 위험을 파악하기 위해 소프트웨어 데이터 분석 도구를 사용하는 것입니다. 새로운 데이터 도구는 잠재적 위험을 미리 파악하고, 조직이 직면한 위험을 줄일 수 있도록 일련의 시나리오에 대해 더 나은 모델을 만들고, 원래의 경로를 벗어나기 시작할 경우 문제점을 파악하기 위해 시스템을 모니터링하는 데 도움이 될 수 있습니다.

근본 원인 분석

근본 원인은 원인과 결과 사이의 관계를 살펴 오류나 문제의 근본 원인을 파악하는 데 초점을 맞추는 문제 해결 방법입니다. 특정 원인을 이벤트 순서에서 제거할 경우, 부적절한 이벤트가 반복되지 않으면 그 제거된 원인은 근본 원인이라고 할 수 있습니다.

준구조적 데이터

준구조적 데이터는 데이터베이스에 사용된 것과 같은 공식적인 데이터 모델로 구성되지 않으며 데이터와 계층구조를 설명할 다른 방법을 제공합니다. 준구조적 데이터는 자기기술 구조라고 하는 방식으로 태그 또는 기타 데이터 표식을 사용하는 경우가 많습니다.



작은 데이터

작은 데이터는 고객 설문조사에 포함된 데이터처럼 소량의 데이터를 활용하여 실행 가능한 결과를 달성하는 것입니다. 일반적으로 인간이 충분히 이해하거나 분석할 수 있는 작은 크기의 데이터를 말합니다.

구조적 데이터

구조적 데이터는 행과 열로 잘 정리되어 있어서 검색과 조작이 간편합니다.

테라바이트

테라바이트는 10의 12 제곱으로 표시되는 데이터의 단위입니다. 테라바이트 하드 드라이브는 가정과 직장의 컴퓨터에 일반적으로 사용되거나 클라우드를 통해 액세스할 수 있습니다. 테라바이트는 300시간 분량의 고해상도 비디오를 저장할 수 있습니다.

텍스트 분석

텍스트 분석은 의미를 도출하고, 개념을 추출하거나, 통찰력을 얻기 위해 텍스트 기반 데이터에 통계학적, 언어학적, 기계 학습 기술을 사용합니다. 텍스트 분석은 일반적으로 문서, 대화 내용, 웹 게시물, 코멘트 또는 양식에 포함된 것과 같은 자연어 텍스트에 대해 실시합니다. 콘텐츠의 요약, 발견 또는 분류를 위해 유용하게 사용할 수 있습니다.

트랜잭션 데이터

트랜잭션 데이터는 재무상 구매, 송장, 지급 및 배송 데이터와 같은 특정 이벤트에서 파생된 데이터입니다. 일반적으로 타임스탬프가 포함되며 조직의 일상적 업무를 지원합니다.

비구조화 데이터

비구조화 데이터에는 예를 들어 회의록과 같이 사전 정의된 구조가 없습니다. 경우에 따라서는 비구조화 정보는 조직의 모든 데이터의 70%-80% 이상을 차지할 수 있다고 합니다.

다양성

데이터 혁신을 정의하는 4가지 V 중 하나인 다양성(Variety)은 통찰력을 도출하기 위해 통합하고 분석되는 다양한 소스에서 얻은 다양한 데이터를 나타냅니다. 오늘날 애플리케이션에서 처리되는 다양한 종류의 데이터에는 문자식 데이터베이스, 트랜잭션 데이터, 스트리밍 데이터, 이미지, 오디오, 비디오 등이 포함될 수 있습니다.

속도

데이터 혁신을 정의하는 4개의 V 중 하나인 속도(Velocity)는 데이터가 생성되고, 저장되고, 분석되고, 가시화되는 속도를 말합니다. 예를 들어 대규모 데이터 웨어하우스는 매일 수 많은 새로운 정보를 수신할 수 있습니다. 시간별 데이터는 가치를 극대화하기 위해 흐르는 대로 사용해야 합니다.

진실성

데이터 혁신을 정의하는 4개의 V 중 하나인 진실성(Veracity)은 데이터의 정확성, 확실성, 정교함을 의미하는 데 사용됩니다.

볼륨

데이터 혁신을 정의하는 4개의 V 중 하나인 볼륨(Volume)은 메가바이트에서 브론도바이트에 이르기까지 처리한 데이터의 양을 말합니다.

요타바이트

요타바이트는 10의 24 제곱으로 표시되는 매우 방대한 데이터 저장소의 단위입니다. 요타바이트는 250 조개의 DVD에 저장되는 데이터의 양을 나타냅니다.

제타바이트

제타바이트는 10의 21 제곱으로 표시되는 데이터 저장소의 단위입니다. 2013년을 기준으로 볼 때 WWW(World Wide Web)는 4 제타바이트에 도달했다고 합니다. 2016년에는 전세계 네트워크에서 매일 1 제타바이트 이상이 사용될 것이라고 합니다.

미주

- 1 Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. 웹 사이트. 견해. <https://www.capgemini.com/resources/the-deciding-factor-big-data-decision-making>
- 2 생산성이 1% 증가하는 것은 대단하지 않게 보일 수 있지만 GE의 Jeff Immelt CEO는 “정유 업계의 직원에게 소프트웨어로 수익성을 1% 개선할 수 있는 방법을 알려준다면 그 사람에게 평생의 은인으로 남게 될 것이라고 말할 정도로 대단한 것”이라고 말했습니다. Evans, Peter C. 및 Marco Annunziata 공저. *Pushing the Boundaries of Minds and Machines*. GE, 2012. 웹 사이트. <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- 3 2020년에 약 500억 대의 장치가 인터넷에 연결되어 사물 인터넷이 가능하게 될 것으로 예상됩니다. Evans, Dave. *The Internet Of Things How The Next Evolution Of The Internet Is Changing Everything*. Cisco Internet Business Solutions Group (IBSG), 2011. 웹 사이트. 백서. http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf
- 4 “Bringing big data to the enterprise.” IBM. <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>
- 5 Golub, Ben. “Enough Data to Fill a Stack of DVDs to the Moon (and Back).” *ComputerWorld* 2011. 웹 사이트. http://blogs.computerworld.com/18351/a_stack_of_dvds_to_the_moon_and_back
- 6 Corry, Will. “BIG Data / The Volume Of Business Data Worldwide, Across All Companies, Doubles Every 1.2 Years, According To Estimates.” *The Marketing Blog* 2012. 웹 사이트. <http://www.themarketingblog.co.uk/2012/10/big-data-the-volume-of-business-data-worldwide-across-all-companies-doubles-every-1-2-years-according-to-estimates/>
- 7 “Digital Imaging in the Cloud.” *There Magazine* 2012: 16. 웹 사이트. http://www.agfahealthcare.com/he/global/en/binaries/THERE_12_tcm541-95647.pdf
- 8 787은 데이터 센서를 사용하여 연료 사용을 줄이고, 시스템을 모니터링하며 난기류에 저항하기 위해 기체 정면에 가속도계를 사용하고 있습니다. 센서가 급강하를 감지할 경우, 즉시 날개 플랩에 알려 (나노초의 속도로) 조정하도록 하며 구식 항공기에서는 9피트 강하였던 것이 787에서는 3피트로 줄어 더 원활한 비행을 하도록 하고 있습니다. Humphries, Matthew. “The Boeing 787 produces over 500GB of Data during Every Flight.” *Geek.com* 2013. 웹 사이트. Gosling, Kevin. “E-Enabled Capabilities of the 787 Dreamliner.” *Aero Quarterly* 2009: 22-24. http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/pdfs/AERO_Q109_article05.pdf
- 9 제트 엔진 제조업체인 GE는 엔진 데이터로 이전보다 2,000배 빠른 속도로 결함을 감지할 수 있다고 합니다. Hardy, Quentin. “What Cars Did for Today’s World, Data May Do for Tomorrow’s.” *New York Times* 2014. B7. 인쇄물. http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?_php=true&_type=blogs&_php=true&_type=blogs&module=BlogPost-Title&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs®ion=Body&r=1&
- 10 Henschen, Doug. “Bad Winter Weather Meets Big Data Prediction.” *InformationWeek* 2014. 웹 사이트. <http://www.informationweek.com/big-data/software-platforms/bad-winter-weather-meets-big-data-prediction/d-d-id/1114192>
- 11 Groenfeldt, Tom. “At NYSE, The Data Deluge Overwhelms Traditional Databases.” *Forbes* 2013. 웹 사이트. <http://www.forbes.com/sites/tomgroenfeldt/2013/02/14/at-nyse-the-data-deluge-overwhelms-traditional-databases> “NYSE Euronext Adapting To Market Changes with Near-Real-Time Insight into Information.” IBM Corporation, 2013. 인쇄물. NYSE Euronext. <http://www.ibmbigdatahub.com/sites/default/files/document/NYSE-Euronext-IMC14787USEN.PDF>
- 12 Mayyasi, Alex. “Why UPS Trucks Don’t Turn Left.” *Priceonomics.com* 2014. 웹 사이트. <http://priceonomics.com/why-ups-trucks-dont-turn-left/>
- 13 *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper*. Cisco Visual Networking Index, 2015. 인쇄물. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html
- 14 “LSST General Public FAQs.” LSST.org 2015. <http://www.lsst.org/lsst/faq>
- 15 Quazi, Ed. “Next Generation Sequencing and the Big Data Bottleneck.” *Arkivum* 2014. 웹 사이트. <http://arkivum.com/next-generation-sequencing-and-the-big-data-bottleneck/>
- 16 Valerio, Pablo. “Internet Of Things: 50 Billion Is Only The Beginning.” *EE Times* 2014. 웹 사이트. http://www.eetimes.com/document.asp?doc_id=1321229
- 17 1980년대에는 1GB의 저장소를 관리하기 위해 데이터 관리자가 필요한 것이 일반적이었습니다. GB 디스크의 비용이 수백만 달러 하던 시절이므로 디스크 공간을 최적화하고 사용을 모니터링할 사람이 필요했을 것입니다. Gray, Jim 및 Prashant Shenoy 공저. “Rules Of Thumb in Data Engineering.” Redmond, WA: Microsoft Research Advanced Technology Division, 2009. 인쇄물.

- 기술 보고서. http://research.microsoft.com/pubs/68636/ms_tr_99_100_rules_of_thumb_in_data_engineering.pdf
- ¹⁸ Wohlsen, Marcus. "Dropbox Slashes Its Price As The Cost Of A Gigabyte Nears Zero." *Wired* 2014. 웹 사이트. <http://www.wired.com/2014/08/dropboxs-plan-to-stay-relevant/>
- ¹⁹ 1980년대에 1 기가바이트에 \$200,00 (백만 달러까지 했음)이었던 것이 2013년에는 기가바이트당 \$0.02 까지 하락. Meeker, Mary. "Kleiner Perkins Internet Trends 2014." 2014. 프리젠테이션. <http://cryptome.org/2014/05/internet-trends-2014.pdf>
- ²⁰ 1980년대 승용차의 평균 연비(24.3 mpg)를 바탕으로 생각할 때 같은 비용으로 1000만배나 많은 연료를 구입할 수 있기 때문에 휘발유를 1000만 갤런이나 구입하여 2억 4300만 마일을 달릴 수 있는 것입니다. 지구의 둘레가 24,901 마일이라면 지구를 9,758 바퀴 또는 거의 10,000 바퀴 돌 수 있을 것입니다. U.S. Department of Transportation. 표 4-23: Average Fuel Efficiency Of U.S. Light Duty Vehicles. Washington, DC: Bureau of Transportation Statistics, 2013. 인쇄물. http://www.rita.dot.gov/bts/sites/rita.dot.gov/bts/files/publications/national_transportation_statistics/html/table_04_23.html
- ²¹ Savitz, Eric. "Big Data: The Hidden Opportunity." *Forbes* 2012. 웹 사이트. <http://www.forbes.com/sites/ciocentral/2012/05/01/big-data-the-hidden-opportunity/>
- ²² Martin Hilbert와 Priscila López가 실시한 연구에 따르면 정보를 저장하는 전세계의 1인당 기술 용량은 1980년대 이후 40개월마다 약 2배로 증가했습니다. Hilbert, M., and P. Lopez. "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science* 332.6025 (2011) : 60-65. 웹 사이트. <http://www.sciencemag.org/content/332/6025/60>
- ²³ 저장소 비용은 1992년 기준으로 기가바이트당 \$569에서 2013년 기가바이트당 \$0.02로 줄어 매년 38%씩 감소했습니다. Meeker, Mary. "Kleiner Perkins Internet Trends 2014." 2014. 프리젠테이션. <http://cryptome.org/2014/05/internet-trends-2014.pdf>
- ²⁴ Lohr, Steve. "For Big-Data Scientists, 'Janitor Work' Is Key Hurdle To Insights." *New York Times*. 2014: B4. 인쇄물. http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html?_r=0
- ²⁵ "Data, data everywhere," *The Economist*, 2010년 2월 25일자. <http://www.economist.com/node/15557443>
- ²⁶ Crovitz, L. Gordon. "Why 'Big Data' Is A Big Deal." *Wall Street Journal*. 2013. 인쇄물. <http://online.wsj.com/news/articles/SB10001424127887324077704578364632408717740>
- ²⁷ 2006년 앙골라의 가뭄 보고서는 이전의 콜레라 발병 사례로 볼 때 발병하고 몇 년 후에 다시 재발한다는 교훈을 얻었기 때문에 콜레라 발병 가능성을 경고했습니다. 이 시스템의 경고는 70% ~ 90% 정도 정확했습니다. Simonite, Tom. "Software Predicts Tomorrow's News by Analyzing Today's And Yesterday's." *MIT Technology Review* 2013. 인쇄물. <http://www.technologyreview.com/news/510191/software-predicts-tomorrows-news-by-analyzing-todays-and-yesterdays/>
- ²⁸ "Ten Big Data Case Studies in a Nutshell." *TechTarget*, 2013. 인쇄물. Essential Guide. <http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell>
- ²⁹ "Big Data to Predict Offshore Accidents, Tsunamis and Other Natural Disasters." *Predictive Analytics Today*. 2013 웹 사이트. <http://www.predictiveanalyticstoday.com/big-data-predict-shore-accidents-tsunamis-natural-disasters/>
- ³⁰ "New Developments in Big Data Visualization." *USTelecom Media* 2014. 웹 사이트. <http://www.ustelecom.org/blog/new-developments-big-data-visualization#sthash.HefD5H52.dpuf>
- ³¹ Quinn, Tom. "New and Unexpected Uses for Scoring Technology." *Credit Score Blog* 2011. 웹 사이트. <http://blog.credit.com/2011/06/new-and-unexpected-uses-for-scoring-technology/>
- ³² Crovitz, L. Gordon. "Why 'Big Data' Is A Big Deal." *Wall Street Journal* 2013: 인쇄물. <http://online.wsj.com/news/articles/SB10001424127887324077704578364632408717740>
- ³³ 데이터 과부하: 오늘날의 경험 많은 임상 의사는 의료 서비스를 제공하기 위해 약 200만 건의 정보가 필요하며 의사는 평균 7개의 저널을 구독하여 매년 2,500개 이상의 기사를 읽고 있는데 진단, 예후, 치료 및 관련된 의료 문제에 대해 최신 정보를 파악하는 것이 거의 불가능한 상태입니다. "Clinical Decisions Support Systems: The Time Has Come...." Frost & Sullivan, 2009. 인쇄물. 시장 전망. <http://www.frost.com/prod/servlet/cio/181298788>
- ³⁴ "Clinical Decisions Support Systems: The Time Has Come...." Frost & Sullivan, 2009. 인쇄물. 시장 전망. <http://www.frost.com/prod/servlet/cio/181298788>
- ³⁵ "Kinect Sign Language Translator Expands Communication Possibilities." *Microsoft Research* 2013. 웹 사이트. <http://research.microsoft.com/en-us/collaboration/stories/kinect-sign-language-translator.aspx>
- ³⁶ "Brats 2012 - Multimodal Brain Tumor Segmentation Challenge." *CodaLab*, 2012. 인쇄물. <https://www.codalab.org/competitions/191>
- ³⁷ Grady, Denise. "3-D Mammography Test Appears To Improve Breast Cancer Detection Rate." *New York Times* 2014: p. A1 인쇄물. <http://www.nytimes.com/2014/06/25/health/breast-cancer->

3d-mammography-test-x-ray.html?emc=edit_th_20140625&nl=todaysheadlines&nid=435891&r=0

- ³⁸ “The game-changing technology that’s transforming manufacturing.” *Manufacturing Weekly*, 2014년 1월 31일. <http://web.archive.org/web/20140131233544/http://www.manufacturingweekly.com/supercomputers/>
- ³⁹ “The Return on the Data Asset in the Era of Big Data: Capturing the \$1.6 Trillion Data Dividend.” *Cloud Platform News Bytes Blog* 2015. 웹 사이트. <http://blogs.technet.com/b/stbnewsbytes/archive/2014/04/15/the-return-on-the-data-asset-in-the-era-of-big-data-capturing-the-1-6-trillion-data-dividend.aspx>
- ⁴⁰ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. 웹 사이트. Point Of View. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- ⁴¹ 생산성이 1% 증가하는 것은 대단하지 않게 보일 수 있지만 GE의 Jeff Immelt CEO는 “정유 업계의 직원에게 소프트웨어로 수익성을 1% 개선할 수 있는 방법을 알려준다면 그 사람에게 평생의 은인으로 남게 될 것이라고 말할 정도로 대단한 것”이라고 말했습니다. Evans, Peter C. 및 Marco Annunziata 공저. *Pushing the Boundaries of Minds and Machines*. GE, 2012. 웹 사이트. <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- ⁴² BSA/IPSOS Global Data Analytics Poll, 2014년 11월, www.bsa.org/datasurvey
- ⁴³ Gartner, “Gartner Says Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally To Support Big Data By 2015.” 2012. 인쇄물. <http://www.gartner.com/newsroom/id/2207915>
- ⁴⁴ BSA/IPSOS Global Data Analytics Poll, 2014년 11월, www.bsa.org/datasurvey
- ⁴⁵ BSA/IPSOS Global Data Analytics Poll, 2014년 11월, www.bsa.org/datasurvey
- ⁴⁶ ESG 연구에 따르면 각 병원에서 관리하는 데이터는 2010년 168 테라바이트에서 2015년 665 테라바이트로 증가할 것으로 예상되고 있습니다. “Digital Imaging in the Cloud.” *There Magazine* 2012: 16. 인쇄물. http://www.agfahealthcare.com/he/global/en/binaries/THERE_12_tcm541-95647.pdf
- ⁴⁷ Manyika, James 외 *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 인쇄물. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- ⁴⁸ 연구원들이 133,000명의 환자들에서 수집한 데이터에 대해 기계 학습 알고리즘을 교육했습니다. 이 모델은 허위 양성률을 줄이기 위해 더 노력할 필요가 있습니다. Rutkin, Aviva. “Machine Predicts Heart Attacks 4 Hours Before Doctors - New Scientist.” *New Scientist*. 2014. 웹 사이트. <http://www.newscientist.com/article/mg22329814.400-machine-predicts-heart-attacks-4-hours-before-doctors.html>
- ⁴⁹ Ford의 현대식 하이브리드 Fusion 모델은 1 시간당 최고 25GB의 데이터를 생성합니다. Hemsoth, Nicole. “How Ford Is Putting Hadoop Pedal To The Metal.” *Datanami*. 2013. 웹 사이트. http://www.datanami.com/2013/03/16/how_ford_is_putting_hadoop_pedal_to_the_metal/
- Chevy Volt에는 1000 만개 이상의 소프트웨어 코드가 포함되어 있으며 소프트웨어 개발업체는 오래 동안 제조업의 중심지로 알려진 미시건주의 남동부에서 가장 급성장하는 기술 전문기업 중 하나입니다. Trop, Jaclyn. “Detroit, Embracing New Auto Technologies, Seeks App Builders.” *The New York Times*. 2013년 6월 30일. <http://www.nytimes.com/2013/07/01/technology/detroit-embracing-new-auto-technologies-seeks-app-builders.html>
- ⁵⁰ Miller, Claire Cain. “If Robots Drove, How Much Safer Would Roads Be?” *New York Times* 2014: A3. 인쇄물. <http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html?ref=technology&r=0>
- ⁵¹ 787은 데이터 센서를 사용하여 연료 사용을 줄이고, 시스템을 모니터링하며 난기류에 저항하기 위해 기체 정면에 가속도계를 사용하고 있습니다. 센서가 급강하를 감지할 경우, 즉시 날개 플랩에 알려 (나노초의 속도로) 조정하도록 하며 구식 항공기에서는 9피트 강하였던 것이 787에서는 3피트로 줄어 더 원활한 비행을 하도록 하고 있습니다. Gosling, Kevin. “E-Enabled Capabilities of the 787 Dreamliner.” *Aero Quarterly* 2009: 22-24. http://www.boeing.com/commercial/aeromagazine/articles/qtr_01_09/pdfs/AERO_Q109_article05.pdf
- ⁵² 제트 엔진 제조업체인 GE는 엔진 데이터로 이전보다 2,000배 빠른 속도로 결함을 감지할 수 있다고 합니다. Hardy, Quentin. “What Cars Did for Today’s World, Data May Do for Tomorrow’s?” *New York Times* 2014: B7. 인쇄물. http://bits.blogs.nytimes.com/2014/08/10/g-e-creates-a-data-lake-for-new-industrial-ecosystem/?_php=true&_type=blogs&_php=true&_type=blogs&module=BlogPost-Title&version=Blog%20Main&contentCollection=Big%20Data&action=Click&pgtype=Blogs®ion=Body&r=1&
- ⁵³ Long, Jessica 및 William Brindley 공저. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. 인쇄물. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf

- ⁵⁴ 제조 부문은 2010년에만 2 엑사바이트의 새로운 데이터를 저장했습니다. Manyika, James 외 *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 인쇄물. http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
- ⁵⁵ Manyika, James 외 *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 인쇄물. http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
- ⁵⁶ “‘Data Smart’ Strategies for Customers Are Yielding ‘Early But Impressive Returns.’” *Microsoft Research the Fire Hose* 2014. 웹 사이트. <http://blogs.microsoft.com/firehose/2014/05/22/data-smart-strategies-for-customers-are-yielding-early-but-impressive-returns/>
- ⁵⁷ Somers, Dan. “Manufacturing 4.0 – From Industrialization to Data-Driven Product Lifecycle.” *Citizenekkk*. 2013. 웹 사이트. <http://citizenekkk.com/2013/11/05/manufacturing-4-0-industrialisation-data-driven-product-lifecycle/>
- ⁵⁸ McKinsey는 Toyota가 이러한 데이터 중심적 설계 기술을 사용하여 최초의 프로토타입을 제작하기 전에 결함의 80%를 제거할 수 있었다고 보고했습니다. Manyika, James 외 *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 인쇄물. http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
- ⁵⁹ Findings of the New Intelligent Enterprise Study. IBM 2010 *New Intelligent Enterprise Global Executive Study*. 2010. 인쇄물.
- ⁶⁰ Geron, Tomio. “Cows in the Cloud: The Hot Startup Moving Farmers into the Cloud.” *Forbes* 2012. Web; Helmer, Jodi. “Get Ready For Robot Farmers.” Yahoo 2014. 웹 사이트. <https://www.yahoo.com/tech/get-ready-for-robot-farmers-100613764059.html>
- ⁶¹ De-Identification of Personally Identifiable Information, National Institute of Science and Technology, DRAFT NISTIR 8053 (2015년 4월) .
- ⁶² Cavoukian, Ph.D., Ann, and El Emam, Ph.D., Khaled, Dispelling the Myths Surrounding De-Identification: Anonymization Remains a Strong Tool for Protecting Privacy, Information and Privacy Commissioner of Ontario, (2011년 6월); Cavoukian, Ph.D., Ann, and Daniel Castro Castro. Big Data And Innovation, *Setting The Record Straight: De-Identification Does Work*. ITIF, 2014. 인쇄물. <http://www2.itif.org/2014-big-data-deidentification.pdf>
- ⁶³ 예를 들어 Microsoft가 최우선적으로 생각하는 추가적인 프라이버시 보호 참조 https://www.youtube.com/watch?feature=player_embedded&v=bt51MWII1oY
- ⁶⁴ *Government Information Requests*는 Apple이 첨단 암호화 기술을 iPhone 운영체계에 사용하여 “사진, 메시지(첨부 파일 포함), 이메일, 연락처, 통화 내역, iTunes 콘텐츠, 메모, 알림 등과 같은 개인 데이터가 암호로 보호되도록 했다”고 www.apple.com/privacy/government-information-requests/에서 언급.
- ⁶⁵ Hachman, Mark, “Microsoft’s updated privacy policy makes it clear it’s not selling ads against your words,” *PCWorld*, 2014년 6월 11일, <http://www.pcworld.com/article/2362130/microsofts-updated-privacy-policy-makes-it-clear-its-not-selling-ads-against-your-words.html>
- Timberg, Craig. “Newest Androids Will Join iPhone In Offering Default Encryption, Blocking Police.” *Washington Post* 2014; 인쇄물. <http://www.washingtonpost.com/blogs/the-switch/wp/2014/09/18/newest-androids-will-join-iphones-in-offering-default-encryption-blocking-police/>
- ⁶⁶ 데이터 브로커 Acxiom은 소비자를 위한 데이터 웹 사이트를 열고 탈퇴 기능을 제공 <http://cir.ca/news/acxiom-gives-consumers-data-peekaboutthedata.com>
- ⁶⁷ BSA/IPSOS Global Data Analytics Poll, 2014년 11월, www.bsa.org/datasurvey
- ⁶⁸ McKinsey Global Institute. *Internet Matters: The Net’s Sweeping Impact On Growth, Jobs, And Prosperity*. McKinsey & Co., 2011. 인쇄물.
- ⁶⁹ Manyika, James 외 *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011. 인쇄물. http://www.mckinsey.com/~media/McKinsey/dotcom/Insights20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx
- ⁷⁰ Burtch Works의 2014년 4월 연구인 Salaries of Data Scientists.
- ⁷¹ Bagley, Rebecca. “How The Cloud And Big Data Are Changing Small Business.” *Forbes* 2014. 웹 사이트. <http://www.forbes.com/sites/rebeccabagley/2014/07/15/how-the-cloud-and-big-data-are-changing-small-business/>
- ⁷² BSA/IPSOS Global Data Analytics Poll, 2014년 11월, www.bsa.org/datasurvey
- ⁷³ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. 웹 사이트. Point Of View. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- ⁷⁴ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. 웹 사이트. 견해.

- <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- ⁷⁵ Gerbis, Nicholas. "10 Correlations That Are Not Causations." *HowStuffWorks*. 2015. 웹 사이트. <http://science.howstuffworks.com/innovation/science-questions/10-correlations-that-are-not-causations.htm>
- ⁷⁶ Vesset, Dan, Henry D. Morris 및 John F. Gantz 공저. *Capturing the \$1.6 Trillion Data Dividend*. IDC, 2014. 인쇄물. IDC White Paper.
- ⁷⁷ Westerman, George, Didier Bonnet 및 Andrew McAfee 공저, "The Advantages of Digital Maturity." 2012년 11월, MIT Sloan.
- ⁷⁸ "Meeting the Big Data Challenge: Don't Be Objective." *Forbes* 2013. 웹 사이트. <http://www.forbes.com/sites/darden/2013/02/01/meeting-the-big-data-challenge-dont-be-objective/>
- ⁷⁹ IDG Enterprise의 2014 Big Data 연구. IDG. *CEOs Call for Big Data and IT Continues To Lead Investment Decisions*. 2014. 인쇄물. <http://www.idgenterprise.com/press/ceos-call-for-big-data-and-it-continues-to-lead-investment-decisions>
- ⁸⁰ Miller, Claire Cain. "If Robots Drove, How Much Safer Would Roads Be?" *New York Times* 2014: A3. 인쇄물. <http://www.nytimes.com/2014/06/10/upshot/if-robots-drove-how-much-safer-would-roads-be.html?ref=technology&r=0>
- ⁸¹ Clemens, Samuel. "7 Facts about Data Quality [Infographic]." *InsightSquared*. 2012년 1월 3일 화요일. 웹 사이트. <http://www.insightsquared.com/2012/01/7-facts-about-data-quality-infographic/>
- ⁸² Economist Intelligence Unit. *Big Data Harnessing a Game-Changing Asset*. SAS, 2011. 웹 사이트. http://www.sas.com/resources/asset/SAS_BigData_final.pdf
- ⁸³ "The Return on the Data Asset in the Era of Big Data: Capturing the \$1.6 Trillion Data Dividend." *Cloud Platform News Bytes Blog* 2015. 웹 사이트. <http://blogs.technet.com/b/stbnewsbytes/archive/2014/04/15/the-return-on-the-data-asset-in-the-era-of-big-data-capturing-the-1-6-trillion-data-dividend.aspx>
- ⁸⁴ BSA/IPSOS Global Data Analytics Poll, 2014년 11월, www.bsa.org/datasurvey
- ⁸⁵ 고성장하는 1990년대까지 소매를 포함한 미국 전체 경제에 대한 IT 투자는 미국 생산성의 연간 성장률을 1%-2% 높였습니다. Farrell, Diana 외 *How IT Enables Productivity Growth*. San Francisco: McKinsey Global Institute High Tech Practice, 2002. 인쇄물. http://www.mckinsey.com/insights/business_technology/how_it_enables_productivity_growth
- ⁸⁶ National Bureau of Economic Research. NBER Working Paper No. 18315: *Is U.S. Economic Growth Over?* *Faltering Innovation Confronts The Six Headwinds*. 2012. 인쇄물. <http://www.nber.org/papers/w18315>
- ⁸⁷ Economist Intelligence Unit. *The Deciding Factor: Big Data & Decision Making*. Cap Gemini, 2012. 웹 사이트. Point Of View. <http://bigdata.pervasive.com/Solutions/Telecom-Analytics.aspx>
- ⁸⁸ Gertner, Joey. "GE for Making the 'Internet of Things' Real." *Fast Company* 2014. 웹 사이트. <http://www.fastcompany.com/most-innovative-companies/2014/ge>
- ⁸⁹ Evans, Peter C. 및 Marco Annunziata 공저. *Pushing the Boundaries of Minds and Machines*. GE, 2012. 웹 사이트. <http://files.gereports.com/wp-content/uploads/2012/11/ge-industrial-internet-vision-paper.pdf>
- ⁹⁰ "City Of Barcelona Realizes Vision of Innovative City Governance with Cloud, Devices, and Apps." Customers. microsoft.com. 2014. 웹 사이트. <https://customers.microsoft.com/Pages/Home.aspx>
- ⁹¹ "Autodesk the Gallery Masdar Headquarters Positive Energy Building." *Autodesk.com*. 2015. 웹 사이트. <http://www.autodesk.com/gallery/exhibits/currently-on-display/adrian-smith-gordon-gill-architecture-masdar-headquarters>
- ⁹² Bunge, Jacob. "Big Data Comes To The Farm, Sowing Mistrust." *Wall Street Journal* 2014. 인쇄물. <http://online.wsj.com/news/articles/SB10001424052702304450904579369283869192124>
- "Supply Chain Management Solution for Fast Moving Consumer Goods & Food Industries - Farm to Fork Tech Mahindra." *Techmahindra*. 2015. 웹 사이트. http://www.techmahindra.com/en-US/www/solutions/Pages/Enterprises/retail_farm_fork.aspx
- ⁹³ 2013년에서 2020년 사이에 선진국 시장과 신흥 시장(중국) 사이의 디지털 세상은 선진국 시장이 60%를 차지했던 것에서 신흥 시장 디지털 세상의 데이터가 60% 차지하는 것으로 바뀌게 될 것입니다. EMC Digital Universe. *Executive Summary Data Growth, Business Opportunities, and the IT Imperatives*. IDC, 2014. 인쇄물. <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>
- ⁹⁴ Long, Jessica 및 William Brindley 공저. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. 인쇄물. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf
- ⁹⁵ Long, Jessica 및 William Brindley 공저. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. 인쇄물. Accenture Development Partnerships Insights into the Role of Technology in Addressing

Development Challenges. https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf

⁹⁶ Levy, Stephen. "Bill Gates and President Bill Clinton on the NSA, Safe Sex, and American Exceptionalism." *Wired* 2013: 인쇄물. <http://www.wired.com/2013/11/bill-gates-bill-clinton-wired/2/>

Chhachhar, Abdul Razaque 및 Siti Zobidah Omar 공저. "Use of Mobile Phone among Fishermen for Marketing and Weather Information." *Archives Des Sciences* 65.8 (2012): 107-119. 인쇄물. http://www.academia.edu/4592505/Use_of_Mobile_Phone_among_Fishermen_for_Marketing_and_weather_information

⁹⁷ Neuman, William 및 Ralph Blumenthal 공저. "New to the Archaeologist's Tool Kit: The Drone." *New York Times* 2014. 인쇄물. http://mobile.nytimes.com/2014/08/14/arts/design/drones-are-used-to-patrol-endangered-archaeological-sites.html?_r=1&referrer

⁹⁸ "Forty Thousand Relics to Be Digitized In Five Years." *Thanhien News*. 2010. 웹 사이트. <http://www.thanhiennews.com/entertainment/forty-thousand-relics-to-be-digitized-in-five-years-22816.html>

⁹⁹ Long, Jessica 및 William Brindley 공저. *The Role of Big Data and Analytics in the Developing World*. Accenture, 2013. 인쇄물. Accenture Development Partnerships Insights into the Role of Technology in Addressing Development Challenges. https://www.accenture.com/us-en/~media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Strategy_5/Accenture-ADP-Role-Big-Data-And-Analytics-Developing-World.pdf

¹⁰⁰ Future of Privacy Forum. *Big Data: A Tool for Fighting Discrimination and Empowering Groups*. Future of Privacy Forum and Anti-Defamation League, 2014. 인쇄물. <http://www.futureofprivacy.org/wp-content/uploads/Big-Data-A-Tool-for-Fighting-Discrimination-and-Empowering-Groups-Report1.pdf>

¹⁰¹ Wladawsky-Berger, Irving. "The Changing Nature of Globalization in Our Hyperconnected, Knowledge-Intensive Economy." *Wall Street Journal* 2014. 인쇄물. http://blogs.wsj.com/cio/2014/06/20/the-changing-nature-of-globalization-in-our-hyperconnected-knowledge-intensive-economy/?mod=wsj_ciohome_cioreport

¹⁰² 예를 들어 아르헨티나, 호주, 브라질, 캐나다, 칠레, 중국, 콜롬비아, 코스타리카, 그리스, 홍콩, 인도, 인도네시아, 한국, 멕시코, 페루, 러시아, 스위스, 베트남은 기업들이 개인 정보를 해외로 전송하는 것을 엄격하게 금지 또는 제한하는 규칙을 채택하거나 제안했습니다.

BSA | THE SOFTWARE ALLIANCE 소개

BSA | The Software Alliance(www.bsa.org)는 정부 정책과 해외 시장으로부터 전세계 소프트웨어 업계를 옹호하는 단체입니다. 세계에서 가장 혁신적인 기업들이 구성원이며 경제를 활성화하고 현대의 삶의 질을 향상시키는 소프트웨어 솔루션을 만들고 있습니다.

워싱턴 DC에 본부를 두고 60여 개 국가에서 활약하는 BSA는 합법적인 소프트웨어 사용을 촉진하고 기술혁신을 권장하며 디지털 경제의 성장을 주도하는 공공 정책을 옹호하는 준법 프로그램의 선도자입니다.



www.bsa.org

BSA 전세계 본부
20 F Street, NW
Suite 800
Washington, DC 20001
전화: +1.202.872.5500
팩스: +1.202.872.5501

BSA 아시아 태평양 지역
300 Beach Road
#25-08 The Concourse
Singapore 199555
전화: +65.6292.2072
팩스: +65.6292.6369

BSA 유럽, 중동 및 아프리카 지역
2 Queen Anne's Gate Buildings
Dartmouth Street
London, SW1H 9BP
United Kingdom
전화: +44.207.340.6080
팩스: +44.207.340.6090