

A PURE PRIMER

인공지능(AI), 데이터 분석 그리고 기업의 미래



보다 스마트한 세상

빅데이터가 많은 사람의 입에 회자되고 있으며, 기업의 혁신을 주도하는 빅데이터의 역할에 대한 기대가 매우 높아졌습니다. 그러나 빅데이터의 진정한 혜택을 실현하고 있는 기업은 소수에 불과합니다.

지금 데이터를 잘 활용하고 있는 기업에서는 놀라운 일들이 벌어지고 있습니다. 데이터는 비즈니스뿐 아니라, 다양한 업계 및 현실 세계를 혁신적으로 변화시키고 있습니다. 오늘날 조직들은 빅데이터를 활용하여 헬스케어 서비스를 혁신하고, 시각 장애인들에게 시력을 되찾아 주며, 자율주행 자동차를 점점 현실화하고 있습니다.

- AI 스타트업 **Enlitic**은 의료진이 데이터를 활용해 더 빠르고 정확하게 진단을 내리도록 지원합니다. 이를 통해 보다 신속한 치료 결정을 가능하게 하며 잠재적으로 생명을 구할 수 있습니다.
- **Aipoly**와 **Microsoft**의 **Seeing AI**는 시각 장애인 전용 스마트폰 앱으로 이미지 인식 기술을 이용해 시각 장애인이 주변 환경을 이해하고 상호 작용하도록 지원합니다.
- 또 다른 AI 스타트업 **Drive.ai**는 **Lyft**와 협력해 딥러닝 접근방법을 자율주행차에 적용하기 위한 파일럿 프로그램을 진행 중입니다.

물론, 새로운 알고리즘, 병렬 시스템 및 빅데이터가 함께 모여 새로운 기회의 세상을 열어주는 "통찰력의 빅뱅"이 일어나지 않았다면, 이러한 기술들은 존재하지 않았을 것입니다. 통찰력 도출을 위한 기술들은 모두 결합되어 제 4차 산업 혁명을 가속화하고 있습니다.

그러나 많은 이들은 아직 4차 산업혁명의 영향을 실감하지 못하고 있습니다. 조직들은 빅데이터라는 거대한 숲 속에서 길을 잃고 유용한 통찰력을 도출하는데 어려움을 겪고 있습니다. 또한, 레거시 인프라는 데이터의 자유로운 활용과 조직의 혁신을 방해하며 소중한 시간과 리소스를 낭비하게 만들고 있습니다.

그렇다면, 데이터의 스마트한 활용을 위해 토론을 하는 단계에서 실질적인 결과를 도출하는 단계로 어떻게 이동할 수 있을까요? 데이터는 기업의 가장 큰 자산입니다. 이제 데이터를 최대한 활용해야 할 때입니다.



개요

빅데이터를 잘 활용한다는 것은 처음 시작하기 어려운 일일 수 있습니다. 그래서, 퓨어스토리지는 기업이 집중할 수 있는 주요 전략적 영역에 대해 살펴 보았습니다. 최근 부상하고 있는 빅데이터의 트렌드에는 다음과 같은 세 가지의 도전과제 및 기회가 있습니다.

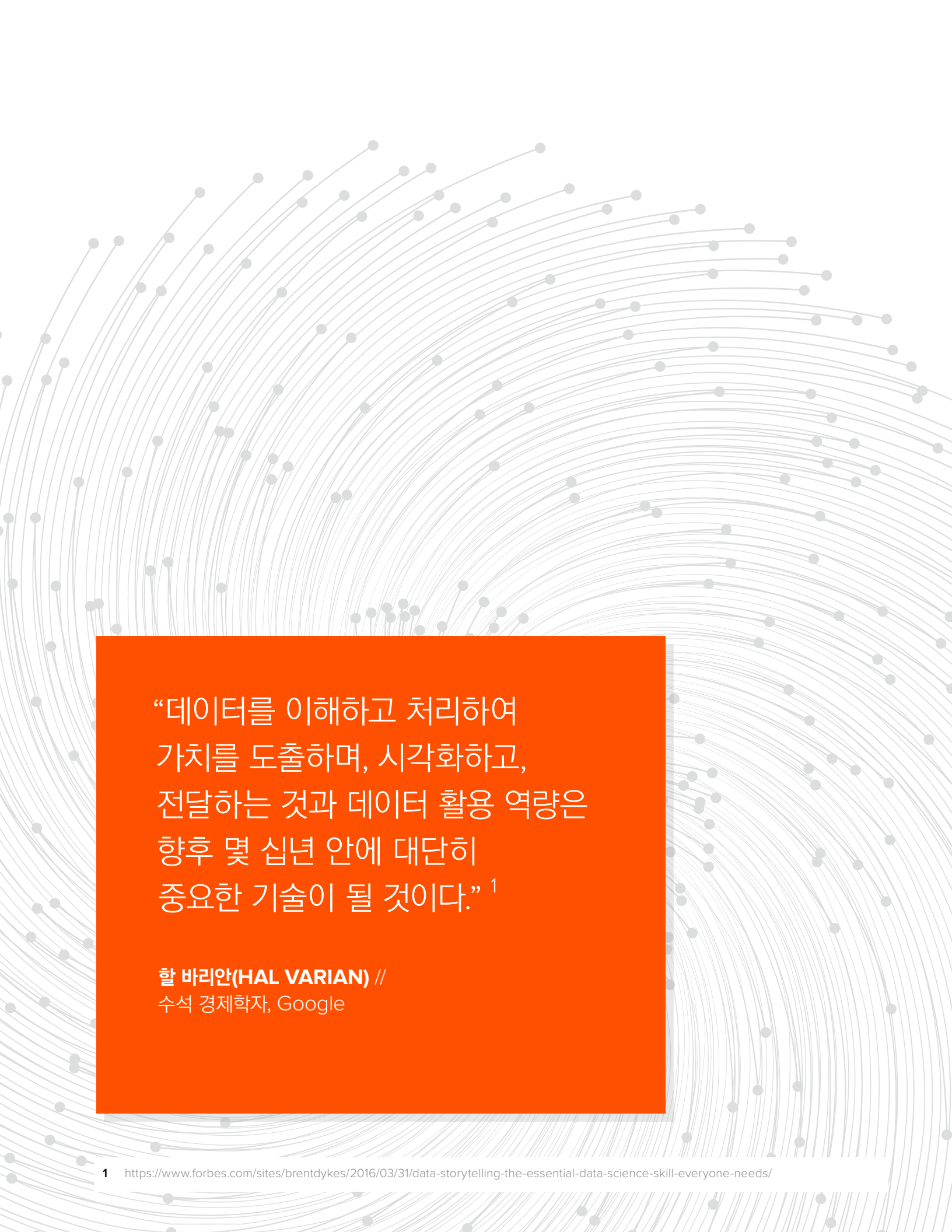
1. 분석이 필요한 방대한 양의 데이터
2. 강력하지만 복잡한 툴 및 기술의 급증
3. 빅데이터 전용 인프라 구축 및 새로운 기술에 대한 필요성

기업은 이 세 가지 영역에서 역량을 강화해야 하며, 데이터를 최대한 활용하는 능력을 갖춘 기업만이 자사의 미래 비즈니스를 주도해 나갈 것입니다.

우리는 50~100년마다 산업 혁명이 일어나는 것을 봐왔으며, 이러한 혁명에서 전환이 느렸던 기업들은 도태되어 왔습니다. 이제, 인공지능(AI) 및 데이터 분석의 급속한 부상으로 지원되는 4차 산업 혁명이 다가오고 있습니다.

새로운 데이터 중심의 세계에서 성공하려면, 조직들은 현대적인 데이터 플랫폼 및 스토리지 인프라를 포함한 올바른 기술, 접근방식 및 툴이 필요합니다. 스토리지는 데이터가 저장되는 곳이며, 기업의 비즈니스에 최적화 된 속도로 데이터를 제공해야 합니다. 또한 새로운 빅데이터 시대에 맞춰 특별히 설계되어야 합니다.



The background of the slide features a series of thin, light gray concentric arcs that sweep across the frame from the bottom left towards the top right. Scattered throughout this pattern are numerous small, solid gray dots, some of which are connected to the arcs by thin lines, creating a sense of data points or a network. The overall effect is a modern, tech-oriented aesthetic.

“데이터를 이해하고 처리하여
가치를 도출하며, 시각화하고,
전달하는 것과 데이터 활용 역량은
향후 몇 십년 안에 대단히
중요한 기술이 될 것이다.”¹

할 바리안(HAL VARIAN) //
수석 경제학자, Google

¹ <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>

1장

데이터는 모든 곳에 존재합니다.

“데이터 더미 속에 환자의 삶, 아니 전 세계를 바꿀 수 있는 지식이 숨어 있다.”

아틀 버트(ATUL BUTTE) // 의학 박사, 디렉터, THE INSTITUTE OF COMPUTATIONAL HEALTH SCIENCES, UNIVERSITY OF CALIFORNIA, SAN FRANCISCO (UCSF) ²

도전 과제

데이터에는 최고의 비즈니스 기회에 대한 통찰이 담겨 있습니다. 그러나 그 방대한 크기 때문에 많은 기업들이 쉽게 다가가지 못하고 있습니다.

데이터가 많을수록 이를 활용하는 일은 점점 복잡해집니다. 조직들이 처리하는데 어려움을 겪고 있는 데이터의 5가지 “V”는 다음과 같습니다.

- 데이터의 **양**(volume)
- 데이터의 처리 **속도**(velocity)
- 데이터 유형의 **다양성**(variety)
- 데이터의 **진실성**(veracity) – 데이터의 안정성 및 유용성
- 데이터로부터 **가치**(value)를 추출하는 방법

이제는 기존의 데이터 소스(예: 고객 데이터베이스 및 공공 기록 등)뿐 아니라, 점차 증가하는 새로운 데이터 소스(소셜 미디어, 웨어러블, IoT)를 통해 수집되는 데이터를 처리할 필요가 있습니다. 이는 정형 데이터 및 비정형 데이터 모두로부터 가치를 도출해야 한다는 것을 의미합니다.

정형 데이터는 정리가 잘 되어 있으며 기존 비즈니스 툴로 쉽게 분석이 가능합니다. 데이터베이스에서 수집 및 정렬되는 데이터가 대부분 여기 속합니다. 반대로, 비정형 데이터는 분석이 좀 더 어렵습니다. 그 일례로 소셜 미디어의 게시물을 들 수 있습니다. 트위터나 페이스북과 같은 게시물에서 기업은 비즈니스에 도움이 되는 통찰력 있는 가치를 찾아낼 수 있다는 것은 알지만, 그것을 분석하는 것은 특히나 까다롭습니다. 또한 비정형 데이터의 형태는 이미지, 오디오, 비디오 및 다양한 텍스트(예: 워드프로세스 문서, 이메일 및 인스턴트 메시지) 등으로 다양합니다.

90%

전세계의 모든 데이터 중 지난 2년간 생성된 데이터가 차지하는 비중이 90%입니다. ³ 2016년 처음으로 전세계 인터넷 트래픽의 양은 1제타바이트(ZB)를 초과한 것으로 나타났으며, 이는 지난 5년 동안 5배 증가한 결과입니다. ⁴

2 <http://sm.stanford.edu/archive/stanmed/2012summer/article3.html>

3 IBM, 2017년 10가지 주요 마케팅 트렌드

4 Cisco, Cisco 비주얼 네트워킹 인덱스: 예측 및 방법, 2016년-2021년

“데이터가 계속해서 폭발적으로 증가한다면, 데이터 분석 모델이 근본적으로 바뀌어야 한다는 사실을 깨달았다. 기업들은 쓰나미처럼 몰려오는 데이터를 분석해야 한다.”

아누쿨 라키나(ANUKOOL LAKHINA) // 창립자, GUAVUS(스트리밍 데이터 분석을 위한 플랫폼)

비정형 데이터⁵는 풍부한 잠재력을 지니고 있지만, 처리하기 까다롭습니다. 중요한 사실은, 비정형 데이터가 전세계 데이터의 80%를 차지한다는 것입니다. 빅데이터에서 통찰을 얻고자 하는 기업이라면 반드시 비정형 데이터를 활용할 방법을 찾아야 합니다.

기회

분석해야 할 어마어마한 양의 데이터에 직면해 있는 조직들은 전략적으로 집중 공략하는 접근방식을 취해야 합니다. 모든 데이터를 사용할 수도 없고 그런 시도를 할 필요도 없습니다. 대신, 어떤 데이터가 가장 큰 비즈니스 혜택을 가져다 줄지를 판단하고 그러한 데이터에 집중해야 합니다. 실질적인 가치를 창출하고 경쟁업체와 차별화할 수 있도록 경영진이 추구하는 비즈니스 이니셔티브에 맞는 데이터 전략을 수립해야 합니다.

동영상 스트리밍 서비스 Netflix는 기존 TV 방송국과 비교해 자사가 갖춘 경쟁력에 집중했습니다. 전세계에 수백만 명의 사용자를 둔 Netflix는 사람들이 무엇을 즐겨 시청하는지에 대한 풍부한 데이터를 보유하고 있습니다. 물론 TV 방송국들이 더 높은 시청률을 기록하고 있지만, Netflix는 사람들이 시청하는 동영상뿐 아니라 언제 어디서 이를 시청하는지에 대해 분석합니다. 또한, 영화 또는 비디오를 끝까지 보는 비율과 시청률이 올라가고 떨어지는 순간을 모니터링합니다.

Netflix는 이 데이터를 기반으로 프로그램을 자체 제작합니다. 정치 드라마가 인기를 끌면, 유사한 콘텐츠를 외주 제작하여 시청자들이 지속적으로 구독하고 관심을 갖도록 합니다.⁶

현대적인 분석 및 AI 툴을 통해, 비정형 데이터의 복잡성을 극복하고 새롭고 핵심적인 비즈니스 통찰을 발견할 수 있습니다. 최근 Aberdeen Group⁷이 발표한 보고서에 따르면, 비정형 데이터를 사용하는 기업들이 그렇지 않은 기업들과 비교해 데이터 품질 및 사용편의성에 대한 만족도가 두 배에 이르는 것으로 나타났습니다.⁷

마지막으로, 쏟아지는 데이터는 스트리밍 분석으로 해결할 수 있습니다. 실시간 또는 스트리밍 분석 툴은 데이터가 수집되는 순간에 작동하여, 초기에 불필요한 정보를 제거하고 중요한 데이터로부터 통찰을 확보하는데 보다 집중할 수 있도록 해줍니다.

일찍이 이동통신 기업들은 스트리밍 분석의 혜택을 누려왔습니다. 이 기업들은 자사의 서비스 이용자들에게 고객의 위치 및 프로필 정보에 기반한 실시간의 타겟화된 서비스를 제공합니다. 스팸을 즉각적으로 탐지할 수 있으며, 실시간 인사이트를 활용해 고객 이탈을 방지할 수 있습니다.

5 IBM Watson, <https://www.ibm.com/blogs/watson/2016/05/biggest-data-challenges-might-not-even-know/>

6 <https://blog.kissmetrics.com/how-netflix-uses-analytics/>

7 Aberdeen Group, Hadoop의 마력, 2016년 5월



AT&T 는 AI로 지원되는 "자가치유 및 자가 학습" 유지보수 모델을 개발하고 있습니다. 이 기업은 이미 실시간 인사이트를 사용하여 네트워크 성능을 조사하고 있었으며, 장애에 취약한 네트워크 영역을 예측하기 위해 데이터를 사용하기 시작했습니다. 이러한 방식으로 사전 대응적으로 장애를 예방하고, 복구하며 시간과 비용을 절감하여 고객 만족을 보장하고 있습니다.

AT&T는 또한 기지국의 철탑을 점검하는데 카메라가 장착된 드론을 사용하고 있습니다. 카메라가 포착한 영상은 기술자가 실시간으로 분석합니다. 그러나 미래에는 인공지능 기계가 영상을 분석하고 필요한 보수 조치를 제안하게 될 것입니다.⁸

2장

현대적인 툴, 그 강력한 성능에 대한 이해

“모든 비즈니스 분석 툴의 목표는 데이터를 분석하여, 성과를 향상시키는데 사용할 수 있는 실행 가능하며, 비즈니스 가치가 있는 정보를 추출하는 것이다. 그러나 현재 시중에 너무 많은 툴이 출시돼 있어 언제 어떤 툴을 사용해야 할지 가능해가 어렵다.”

버나드 마(BERNARD MARR) // 빅데이터 분야 기초 연설자 겸 작가

도전 과제

데이터 분석 툴은 점점 더 스마트해지며, 그 수 또한 늘어가고 있습니다. Gartner(가트너)는 현대적 비즈니스 인텔리전스 및 분석 플랫폼이 우리 시대의 가장 빠르게 성장하는 소프트웨어 시장 중 하나가 될 것이라고 예측하고 있습니다.⁹ 실제로 시장에서는 다음에 설명하는 것과 같은 많은 기술들이 사용되고 있습니다.

그리고, 데이터베이스는 대용량 데이터를 처리하는 능력을 가지고 있습니다. 관계형 데이터베이스는 정형화된 테이블에 데이터를 저장합니다. 하지만 NoSQL 데이터베이스는(아래에 나열된 NoSQL DB들) 관계형 데이터베이스와 달리 보다 유연한 방식의 데이터 저장 및 처리를 합니다.

머신러닝 툴 및 딥러닝 툴은 인간의 두뇌에서 영감을 얻은 알고리즘을 통해 모델을 훈련시킵니다. 방대한 양의 데이터가 제공되면, 이 툴들은 패턴과 상관 관계를 찾아내어 유용한 비즈니스 통찰을 지원할 수 있습니다.



CASSANDRA



MONGODB



REDIS

또한, 수년간 기업들이 사용해 온 데이터 분석 및 관리 솔루션은 다음과 같습니다.

Caffe

CAFFE2/CAFFE



MICROSOFT
COGNITIVE
TOOLKIT

mxnet

MXNET



MATWORKS

sas

SAS

TERADATA

TERADATA



TENSORFLOW

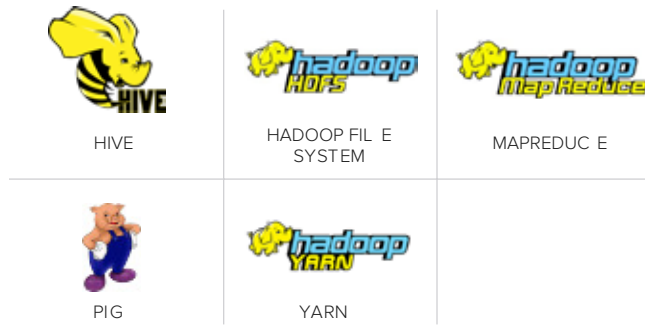


TORCH

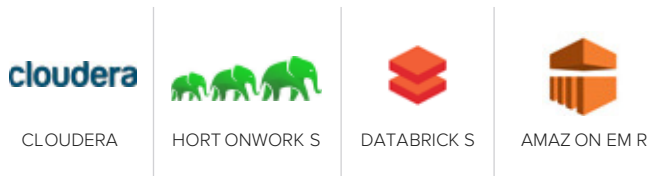
theano

THEANO

보다 광범위하게 사용되고 있는 Apache® Hadoop 프레임워크는 다음과 같은 요소들로 구성되어 있습니다.



Hadoop 및 Spark™ 에코시스템과 워크플로우 간소화를 위한 전문 서비스는 다음과 같습니다.



분석해야 할 데이터의 양도 많지만, 시중에 나와 있는 툴 및 기술의 수도 이에 못지 않게 매우 많습니다. 많은 기업들에게 이러한 복잡하고 다양한 툴을 기존 시스템에 구현하고 통합하는 것은 어려운 일일 수 있습니다. 예를 들어, 딥러닝 알고리즘의 컴퓨트 요구사항을 충족시키지 못하는 레거시 스토리지 인프라는 혁신 속도를 매우 지연시킵니다.

아직까지 많은 기업들에게, 데이터 분석 및 AI 분야에서 역량을 강화하고 이를 활용해 실제 측정 가능한 비즈니스 가치를 실현하는 것은 쉬운 일이 아닙니다.

어려움 속에 찾는 기회

하지만 너무 어렵고 복잡하다고 포기하지 마십시오. 과장된 광고에 현혹되어서도 안되며, 비현실적인 목표를 설정해서도 안됩니다.

68%

의 CEO들은 데이터 분석이 R&D와 고객 관리 시스템 보다 비즈니스에 훨씬 더 높은 수익을 제공한다고 평가합니다.¹⁰

상황을 해결할 수 있는 키는 활용할 데이터의 전략적 집중과 우선 순위 선정입니다. 현재 어떤 비즈니스 도전과제가 기업에 가장 중요한지를 판단하고, 이러한 도전과제를 가장 잘 해결할 수 있는 기술을 선택하고 투자 및 통합해야 합니다.

올바른 분석 툴은 비즈니스에 그만한 투자 가치가 있습니다. eHarmony, Google, MyFitnessPal과 같은 기업들은 특정한 핵심 영역에만 분석 기술을 집중적으로 사용하여 그 역량을 심분 발휘했습니다.

eHarmony는 Hadoop을 통해 수백만 명에게 보다 개별화된 결혼 배우자 매칭 서비스를 제공하고 있습니다. 이 기업은 복잡한 분석을 수행하여 보다 개별화된 결과를 생성합니다. 그 결과, 매칭 성공의 가능성이 높아졌으며, 궁극적으로는 회원들에게 더 나은 고객 경험을 제공할 수 있게 되었습니다.¹¹

Google은 모두가 탐내는 자산인 검색 엔진을 AI로 현대화했습니다. 과거에 구글 검색 엔진은 엔지니어가 직접 수정하고 정의한 규칙에 의해 구현되는 알고리즘을 통해 개발되어 왔습니다. 하지만 이제, 사람이 했던 모든 알고리즘 개발 과정이 딥러닝으로 통합되었습니다. 또한 AI를 통해, 검색 알고리즘은 스스로 학습하며 새로운 규칙들이 자체적으로 생성되게 되었습니다.¹²

¹⁰ PWC의 19세기 연간 글로벌 CEO 설문조사

¹¹ <https://www.cio.com/article/2943599/big-data/how-7-companies-bring-power-to-hadoop-big-data-applications.html#slide4>

¹² <https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/>

MyFitnessPal은 다이어트 및 피트니스 추적 기능을 제공하는 애플리케이션으로, Spark를 사용하여 핵심적인 서비스를 제공합니다. 이 앱은 칼로리 소모량을 계산하는데 이를 위해 정확한 영양 정보를 필요로 합니다. MyFitnessPal은 정교한 알고리즘 세트로 정확도 높은 식품 데이터베이스를 구축했습니다. 그리고 이제 “검증된 식품”들의 정보까지 제공합니다. 앱은 데이터베이스에 있는 식품 목록 중 완전한 영양 정보를 보유한 식품에 체크 표시를 하기까지 합니다.¹³

지금까지 살펴본 데로, 흥미롭고 현대적인 분석 기술은 이러한 툴을 활용하는 기업들에게 중대한 기회를 제공합니다. 이러한 툴은 지속적으로 증가하는 데이터를 통해 더욱 발전합니다. 더 많은 데이터에 노출되어 더 많은 학습으로 얻어진 인사이트를 통해 기업의 비즈니스는 혁신이 가능하며, 그러한 혁신은 우리가 사는 세계를 빠르게 변화시키고 있습니다

WALMART는 매장 안에서 발생하는 실질적인 문제를 해결하기 위해 분석 툴을 이용하고 있습니다. 2011년부터 Walmart Labs에서 자체 툴을 개발하기 시작했으며, 최근 ‘Data Café’를 열었습니다. 왜 어떤 제품은 판매가 되고 또 어떤 제품은 판매가 되지 않는지에 대한 인사이트를 확보합니다. 이를 기반으로 얻은 실시간 해결책으로 경쟁력을 강화할 수 있게 되었습니다.

일례로, 데이터 분석을 통해 일부 매장에서는 새로 나온 특정 쿠키들이 팔리지 않고 있다는 사실을 발견했습니다. 이에 해당 매장의 관리팀에게 경보가 전송되었고, 곧 해당 제품이 매장 진열대에 진열되어 있지 않게 되었습니다. 실시간 데이터 분석은 이러한 상황에서 의사결정을 할 수 있는 핵심적인 역할을 수행합니다.¹⁴

¹³ <http://go.databricks.com/case-studies/myfitnesspal>

¹⁴ <https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-clou>

3장

이제는 실행으로 옮겨야 할 때

“많은 조직들이 데이터 분석을 통해 비즈니스에 실질적인 가치를 확보하고자 인력, 비즈니스 프로세스 및 조직적 역량 개발에 애쓰고 있다. 이는 데이터 분석 역량이 점점 더 기업 경쟁의 우위를 만들고, 데이터 분석에 강한 리더 기업들은 이미 시장에서 많은 혜택을 받고 있기 때문이다. 데이터 분석은 이제 기업에겐 긴박한 과제이다.

MCKINSEY AND CO. // 분석의 시대: 데이터 중심 세상에서의 경쟁'

도전 과제

조직들이 빅데이터 혁명을 위해 결단을 내려야 하는 순간이 다가왔습니다. 비즈니스를 운영하는 방식에 큰 변화가 일어나고 있습니다. 시장에서 뒤처지거나 경쟁력을 잃고 싶지 않다면 기업들은 결정적 조치를 취해야 합니다.

결정적 조치는, 기업의 핵심 인프라의 변화를 의미하며, 현대화된 변화를 위하여 필요한 영역들은 다음과 같습니다.

제품 및 서비스 – 데이터 인사이트를 통해 제품과 서비스를 지속적으로 혁신하고 개발해야 합니다. 관련성 있는 사용자 데이터를 활용해 현재의 제품 및 서비스를 향상시키거나 새로운 제품 및 서비스를 개발할 수 있는 기업은 성공할 것입니다.

인력 및 기술 – 조직들은 끊임 없이 진화하고 적응하는 동시에 민첩성을 갖춘 분석팀이 필요합니다. 디스럽션(Disruption)으로 정의되는 환경에서 확실한 한가지 사실은 모든 비즈니스가 변화한다는 것입니다. 이런 관점에서 기업들은 예측 가능성, 일관성, 정확성 만 추구해서는 안되고, 좀 더 창의적이고 실험적인 태도를 추구해야 합니다.

인프라(컴퓨터 및 스토리지) – 빅데이터 혁명은 현대적인 병렬 시스템으로 가속화 될 수 있습니다. 병렬 시스템은 다수의 컴퓨팅 프로세스가 동시에 실행될 수 있도록 합니다. 큰 규모의 문제를 여러 개의 작은 문제로 나눠 동시에 해결할 수 있습니다. 이러한 방식으로, 방대한 양의 데이터 속에서 빠르게 인사이트를 발견할 수 있습니다. 그러나 데이터를 관장하는 스토리지가 방대한 병렬 구조를 지원하도록 설계되지 않았다면, 기업의 핵심적인 인사이트를 얻는 것은 어려울 수 밖에 없습니다.

앱 및 툴 – 2장에서 설명한 것처럼, 올바른 기술들을 선택하고 이에 투자를 하는 것은 매우 중요합니다. 비즈니스에서 데이터 툴을 활용해 경쟁력을 확보할 수 있는 영역을 찾고, 입증된 분석 솔루션을 통해 집중 공략해야 합니다.

기업이 조직과 인프라의 구조를 데이터 인사이트 중심으로 재조정한다는 것은 엄청난 과업처럼 느껴질 수도 있습니다. 그러나 이 과정을 꼭 수행하여야만 빅데이터가 주는 최상의 혜택을 도출해 낼 수 있습니다.

“기업이 데이터 분석을 통한 인사이트를 통해 완전한 혜택을 얻기 위해서는 데이터 관리와 비즈니스 운영에 인사이트를 통합할 수 있는 내부 역량을 개발하는 것이 필수적이다.”

BAIN & CO // INSIGHTS BLOG "더 나은 인사이트를 제공하기 위한 IT 역량 개발"

기획

비즈니스 성과 향상을 위해 데이터 분석을 도입하여 활용하는 조직들은 경쟁에서 우위를 점할 가능성이 높습니다.

하룻밤 사이에 변화를 기대하긴 어렵습니다. 대신, 반복적 개선과 쉬운 목표 달성에 기반해 비즈니스 인텔리전스를 확보해야 합니다.

LinkedIn은 이러한 접근 방법을 사용해 성공한 좋은 사례입니다. 처음 플랫폼에 그룹 추천 기능을 만들 때 LinkedIn은 소규모로 시작했습니다. 회원들의 프로필에 기반해 이와 연관된 기본적 제안을 제공하는 간단한 프로그램을 생성했습니다. 회원이 특정 학교를 다닌 경우, 프로그램이 그 학교 동문회에 초대하는 식이었습니다. 시간이 지나면서, 이 프로그램은 회원들이 참여한 행사와 다른 정보들을 추가로 고려하게 되었습니다. 시작은 간단했지만, 마침내 대규모 프로젝트로 성장할 수 있었습니다.

LinkedIn 데이터 과학자들은 작은 규모의 성공에 기반해 시간이 흐르면서 더 많은 가치를 추가해 나갔습니다. 유기적이고 민첩하게 성장하는 점진적 방식은 훌륭한 성과를 이끌어 냅니다.¹⁵

인력과 기술의 현대화에 관한 한, 팀에 능력 있는 데이터 과학자를 보유하는 것이 핵심입니다. LinkedIn의 수석 과학자인 DJ 파틸(DJ Patil)은 이상적인 데이터 과학자란 "회사 창립을 같이 할 수 있는 사람"이라고 말합니다. 데이터 과학 또는 분석 팀을 구성할 때는 창업 정신과 창의력을 갖춘 다재다능한 사람, 그리고 컴퓨터 과학, 수학 및 예술 분야에서 골고루 재능을 나타내는 사람을 뽑아야 합니다.

마찬가지로, 인프라 변경은 장기적인 관점에서 미래의 유연성을 염두에 두고 계획되어야 합니다. 기업과 함께 성장하고 변화할 수 있는 기술을 설계하고 지속적인 성공을 위해 길을 마련해야 합니다.

스토리지는 기업의 가장 가치 있는 자산인 데이터가 저장되는 장소이기 때문에, 인프라라는 퍼즐의 매우 중요한 조각입니다. 최상의 조각을 확보해야 합니다. 데이터 분석에 대한 미래 수요를 예측한다는 것은 어려운 일이지만, 현대적인 분석을 위해 특별히 설계된 "모든 것이 튜닝된" 스토리지 솔루션을 통해, 기업은 확신을 갖고 데이터의 역량을 활용할 수 있습니다.

15 O'Reilly Radar 보고서: 데이터 과학이란 무엇인가?



AMERICAN EXPRESS

이 기업이 성공적인 빅데이터 여정은 몇 년 전부터 시작되었습니다. 이러한 여정을 통해 그간 얻은 교훈과 인사이트를 발표했습니다. American Express(AMEX)에 따르면, 조직들은 이제 빅데이터에 적응할 준비를 꼭 갖춰야 합니다. Global Credit Risk and Information Management의 사장 애쉬 굽타(Ash Gupta)는 “빅데이터는 마음가짐”이라고 말합니다.

기업은 또한 새로운 인재를 고용 해야 합니다. 현재 AMEX 팀은 전세계에 800명의 데이터 과학자를 보유하고 있습니다. 지속적인 개선에 대한 집중과 시행 착오를 용인하는 환경 또한 매우 중요합니다. 마지막으로, 오늘날의 변화 무쌍한 비즈니스 환경에서는 속도와 민첩성이 필수입니다.¹⁶

현대적인 데이터 플랫폼은 빠르고(FAST), 크며(BIG), 간단합니다(SIMPLE).

“플래시블레이드(FlashBlade)는 특별히 설계된 시스템으로 기업의 데이터 분석에 방해가 되는 요소들을 제거하여, 비즈니스가 발전 할 수 있는 스토리지 인프라를 제공한다”

에릭 버게너(ERIC BURGNER) //
스토리지 담당 연구 부장, IDC

인사이트의 시대에서 데이터는 기업의 가장 가치 있는 자산입니다. 오늘날의 빅데이터 도전과제를 극복하는 기업은 더 스마트하고, 더 빠르며, 더 강력하게 미래 비즈니스를 성장시킬 수 있습니다.

이러한 잠재력을 발휘하려면 적절한 인프라가 필요합니다. 플래시블레이드(FlashBlade™)는 현대적인 분석과 SI를 위해 특별히 설계된 새로운 클래스의 스케일-아웃 스토리지입니다. 기본적으로 방대한 병렬성을 활용하도록 설계되어, 데이터 과학자들에게 전례 없는 성능과 단순함을 제공합니다. 현대적인 알고리즘 및 컴퓨트 시스템들과 마찬가지로, 플래시블레이드(FlashBlade)는 방대한 병렬 아키텍처로 이루어져 스케일-아웃 확장이 가능하고, 믿을 수 없을 만큼 빠르며, 간단합니다.

FAST: 비정형 데이터는 기존의 처리 방식으로 성능을 예측하기 어렵습니다. 하지만 플래시블레이드(FlashBlade)는 이러한 비정형데이터 환경에서도 전례 없는 데이터 처리 성능을 제공합니다.



BIG: 테라바이트(TB)에서 페타바이트(PB)까지, 모든 데이터를 처리 할 수 있는 파워를 지닌 플래시 블레이드는, 전자레인지 하나만한 크기로 기존 데이터 센터의 전체 레거시디스크 스토리지들을 교체 할 수 있는 솔루션 입니다.

SIMPLE: 현대적인 데이터 분석은 끊임 없이 변화하고 있으며, 매우 복잡합니다. 플래시블레이드(FlashBlade)는 여러분이 간단하게 사용하는 클라우드 서비스처럼 쉽고 민첩하여 모든 데이터를 손쉽게 처리할 수 있습니다.

플래시블레이드(FlashBlade)는 데이터의 가치를 극대화하고자 하는 기업들을 위해, 현대적인 데이터 분석용으로 특별히 설계된 업계 최초의 스토리지 시스템입니다. 헬스케어에서 소비자 가전, 유전체학, 게임, 제조, 소셜 미디어에 이르기까지, 플래시블레이드(FlashBlade)는 다양한 분야에서 사용되고 있으며, 세계 최대 기업들의 데이터 분석을 지원하고 있습니다. 플래시블레이드(FlashBlade)는 세계에서 성능이 가장 뛰어난 AI 슈퍼컴퓨터 와 현재 함께 구성되어 머신러닝의 혁명을 가속화 하고 있습니다. 또한, 업계에서 가장 뛰어난 역량으로 전세계 기업과 조직들에게 미래로 발전할 수 있는 새로운 기회를 보여주는 역할을 하고 있습니다. 퓨어스토리지 플래시블레이드(FlashBlade)가 여러분이 속한 기업을 위해서는 무엇을 할 수 있는 지 궁금하지 않으세요?

플래시블레이드(FlashBlade)가 여러분의 어떻게 여러분의 비즈니스 인사이트를 창출 을 도와드릴 수 있는지, 자세한 정보는 <http://www.purestorage.com/kr/solutions/industries/analytics.html> 를 통해 확인할 수 있습니다.

체계적인 금융 거래를 위해 10배 더 빠른 속도를 제공하는 SPARK

“금융 시장 분석가들은 여러가지 금융 공학 모델을 테스트하고 결과를 확인한 후 또 다시 다른 금융 공학 모델을 테스트하는 일을 종일 반복한다. 그렇기 때문에 분석 성능이 10배에서 20배 향상되면, 시장 출시 속도에서 판도를 바꿀만한 경쟁력을 확보할 수 있다.

스토리지에 대한 수요는 매 18개월마다 두 배로 늘어난다. 겁이 나기도 하는 상황이다. 경쟁업체 제품은 퓨어스토리지의 플래시블레이드(FlashBlade)보다 4배 많은 상면 공간이 필요했으며, 용량을 추가할 수 없었다. 플래시블레이드(FlashBlade)는 우리 팀의 생산성을 획기적으로 향상시켰으며, 새로운 주식 거래 아이디어를 보다 신속하게 시장에 선보일 수 있도록 해주었다.”

게리 콜리어(GARY COLLIER) //
CTO, MAN AHL



유전체학을 위해 SPARK 속도 3배 향상

“높은 성능, 관리 편의성 및 확장성을 제공하는 데이터 플랫폼, 플래시블레이드(FlashBlade)를 활용하여 유전체 염기서열 분석 프로세스를 가속화시킬 수 있었다.”

A large offshore oil rig stands in the ocean under a dramatic sunset sky. The rig's complex steel structure is silhouetted against the colorful horizon. The water is dark blue with gentle waves. An orange graphic element is on the right side of the page.

5배 빨라진 석유 및 가스 탐사

“정말 놀라운 사실은 확장 가능한 시스템의 성능을 크게 향상시켰으며, 앞으로 더욱 향상 될 수 있다는 점이다. 또한 플래시블레이드(FlashBlade)의 용량이 5배나 더 커지는 동시에 성능도 5배나 더 빨라질 것이라고는 생각지도 못했다. 데이터 집약적이며 성능에 의해 좌우되는 석유, 가스 업계와 같은 산업군에서는 이미징 프로세스의 지속적인 성능 개선에 대한 요구가 점차 증가하고 있다. 더 큰 규모의 플래시블레이드(FlashBlade)는 공급 시간을 단축시키고, 리소스 생산성을 향상시키며, 비즈니스 가치를 직접적으로 높여줄 것이다.”

짐 돌란(JIM DOLAN) //

고성능 컴퓨팅(HPC) 담당 글로벌 책임자,
ION GEOPHYSICAL