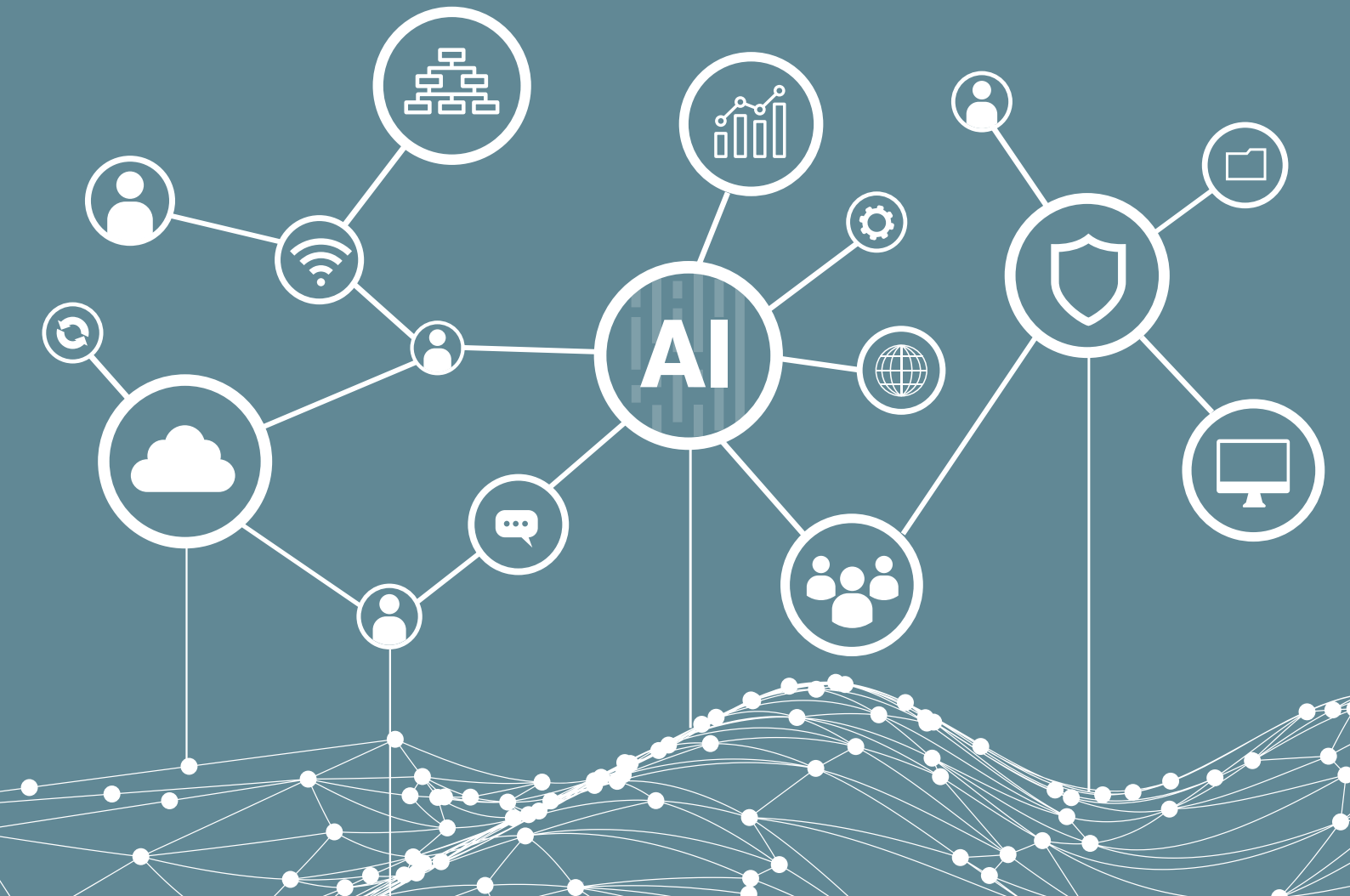


고성능 데이터 구축 + AI 조직 체계화





서문

“고성능 데이터 구축 + AI 조직 체계화”는 Databricks가 후원하는 MIT Technology Review Insights 보고서입니다. MIT Technology Review Insights는 이 보고서를 작성하기 위해 최고 데이터 책임자, 최고 분석 책임자, 최고 정보 책임자 및 기타 고위 기술 경영자 351명을 대상으로 글로벌 설문 조사를 실시했습니다. 응답은 북미, 유럽 및 아시아 태평양 지역에서 고르게 분포되어 있습니다. 표본은 14개 부문을 대표하며, 모든 응답자는 연간 매출 10억 달러 이상인 조직에 속해 있습니다. 조직의 데이터 관리, 분석 및 관련 인프라를 담당하는 경영진과의 인터뷰 또한 연구에 포함되었습니다. 이 보고서의 저자는 Denis McCauley, 편집자는 Francesca Fanshawe이고, 제작자는 Nicola Crepaldi입니다. 이 연구는 독립적으로 편집되었으며, MIT Technology Review Insights의 견해로만 구성되어 있습니다.

귀중한 시간을 내어 인사이트를 제공해주신 다음 분들께 감사를 표합니다.

McDonald's(미국) 데이터 과학 수석 이사 **Patrick Baginski**

CVS Health(미국) 최고 데이터 책임자, CVS Pharmacy 최고 분석 책임자 **Bob Darin**

L'Oréal(미국) 데이터, CRM & 분석 부사장 **Naveen Jayaraman**

Total(프랑스) 그룹 최고 데이터 책임자 **Michel Lutz**

Nielsen(미국) 최고 데이터 및 연구 책임자 **Mainak Mazumdar**

Hivery(호주) CTO **Andy McQuarrie**

Estée Lauder Companies(미국) 최고 분석 책임자 **Sol Rashidi**

Macquarie Bank(호주) 최고 데이터 및 분석 책임자 **Ashwin Sinha**

Northwestern Mutual(미국) 최고 데이터 책임자 **Don Vu**



01 개요.....	4
02 성장과 복잡성.....	6
Databricks 관점: 레이크하우스 효과의 부상	7
03 전략 조정 및 실행	9
데이터로 높은 성과를 내는 조직.....	11
Nielsen: 데이터 의존 비즈니스를 위한 데이터 변환	13
04 분석 및 머신 러닝 확장	14
CVS Health의 패러다임 전환	15
확장의 장벽	16
투자 수익 보호	17
기술, 민주주의, 그리고 문화	18
05 미래 비전	19
새로운 아키텍처를 위한 CDO 위시리스트	19
06 결론.....	21



01 개요

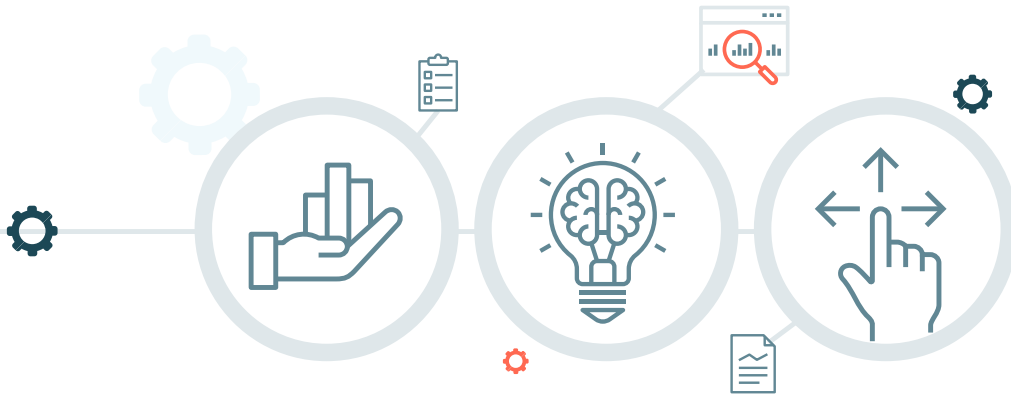


초고위 경영진과 이사회는 데이터로부터 실행 가능한 인사이트를 실시간으로 뽑아내는 조직의 능력이 전략적으로 가장 중요하다는 것을 알고 있습니다. 지난 해 위기 속에서 소비자들이 디지털로 이동하는 움직임이 가속화되면서 이 점에 대한 일말의 의심조차 사라졌습니다. 기업들은 데이터 기반 기업이 되기 위해 머신 러닝(ML) 기능이 있는 분석 도구를 비롯한 지능적 클라우드 기반 기술을 배포하고 있습니다. 그러나 쉽게 액세스할 수 있는 고품질의 풍부한 데이터가 없다면 이러한 도구들은 가치를 제대로 발휘할 수 없습니다.

이러한 맥락에서 효과적인 데이터 관리는 데이터 중심 조직의 기반이지만, 기업에서 데이터를 관리하기란 매우 복잡합니다. 새로운 데이터 기술이 등장하면서, 통합이나 링 펜싱이 불가능한 레거시 시스템과 데이터 사일로의 부담이 커집니다. 아키텍처 단편화는 사일로를 일으킬 뿐만 아니라 많은 조직에서 여러 가지 온프레미스 및 클라우드 기반 도구를 사용하게 되어서 대부분 최고 데이터 책임자(CDO)들에게 골칫거리가 되었습니다. 낮은 데이터 품질에 이런 문제까지 결부되면 조직의 데이터 플랫폼과 그들이 지원하는 머신 러닝 및 분석 모델에서 원하는 비즈니스 성과를 내는 데 필요한 속도와 규모에 도달할 수 없습니다.

MIT Technology Review Insights는 이러한 어려움 속에서 데이터 관리와 그 기술이 어떻게 발전하고 있는지 알아보기 위해 CDO, 최고 분석 책임자(CAO, 보고서의 여러 부분에서 이들을 “데이터 리더”로 지칭함), CIO(최고 정보 책임자), CTO(최고 기술 책임자) 및 기타 고위 기술 리더 351명에게 설문을 실시했습니다. 또한 여러 다른 고위 기술 리더들과 심층 인터뷰도 진행했습니다. 이 연구의 주요 결과는 다음과 같습니다.

- **조직의 13%만이 데이터 전략을 실행하는 능력이 탁월합니다.** 이 선별된 “고성과” 집단이 기업 전체에 측정 가능한 비즈니스 성과를 제공합니다. 이들은 데이터를 “민주화”하고 머신 러닝에서 가치를 도출할 수 있는 탄탄한 데이터 관리와 아키텍처의 기본이 무엇인지에 관심을 둔 덕분에 성공을 거두었습니다. 이렇게 기본을 지키면 데이터 중복이 감소하고, 관련 데이터에 손쉽게 액세스할 수 있으며, 대용량 데이터를 고속 처리할 수 있고, 데이터 품질도 개선됩니다. 또한 이들은 지능적 클라우드 도입에 앞장 섭니다. 그중 74%가 클라우드 환경에서 데이터 서비스 또는 인프라의 절반 이상을 실행했습니다.



향후 2년간 조직의 데이터 최우선순위는 **데이터 관리 개선, 데이터 분석 및 ML 향상, 그리고 스트리밍 및 비정형 데이터를 포함한 모든 유형의 엔터프라이즈 데이터 사용 확대**의 세 가지 영역으로 분류되며, 이 모든 것은 클라우드 플랫폼의 광범위한 적용이 뒷받침되어야 합니다.

- **기술 지원에 힘입은 협업으로 효과적인 데이터 문화를 조성합니다.** 연구를 위해 인터뷰한 CDO들은 분석 및 ML 기능의 민주화를 매우 중요하게 생각했습니다. 고급 데이터 기술을 통해 한계를 넓힌다면 최종 사용자가 정보에 입각한 비즈니스 결정을 내리는 데 도움이 됩니다. 이는 곧 강력한 데이터 문화의 상징이기도 합니다. 그러려면 최신 데이터 아키텍처가 반드시 필요합니다. 어느 CDO는 이를 일컬어, 적절한 사용자가 적절한 데이터에 액세스하여 비즈니스 가치를 창출하는 인사이트를 신속하게 만들어낼 때 성공적인 데이터 관리가 달성된다는 말로 정리했습니다.
- **ML이 비즈니스에 미치는 영향은 전체적 수명 주기 관리가 어려워져 제한됩니다.** 많은 조직에서 ML 사용 사례를 확장하려면 복잡한 과정을 거쳐야 합니다. 응답자의 55%는 ML 모델을 저장하고 검색할 중앙 위치가 없다는 것을 가장 중요한 문제로 지적했습니다. 즉, 이러한 문제는 데이터 과학과 프로덕션 간 핸드오프에서 오류가 발생하기 쉬운 데다, 숙련된 ML 리소스가 부족한 문제까지 겹쳐(둘 모두 응답자의 39%가 언급함) ML, 데이터 및 비즈니스 사용자 간의 팀 협업을 현실화하는 데 심각한 어려움을 발생시킵니다.
- **기업은 데이터 관리, 분석 및 머신 러닝을 지원하는 클라우드 네이티브 플랫폼을 찾고 있습니다.** 향후 2년간 조직의 데이터 최우선순위는 데이터 관리 개선, 데이터 분석 및 ML 향상, 그리고 스트리밍 및 비정형 데이터를 포함한 모든 유형의 엔터프라이즈 데이터 사용 확대의 세 가지 영역으로 분류되며, 모두 클라우드 플랫폼의 광범위한 적용이 뒷받침되어야 합니다. 데이터 전략 실행에 어려움을 겪는 “저성과” 조직의 경우 59%가 데이터 관리 개선이 가장 시급하다고 답했습니다. 이와 대조적으로 대부분의 고성능 그룹은 53%가 ML 사용 사례를 발전시키는 데 중점을 두었습니다.
- **개방형 표준은 미래 데이터 아키텍처 전략의 최우선 조건입니다.** 응답자들이 비즈니스를 위한 새로운 데이터 아키텍처를 구축할 수 있다면, 기존 아키텍처에 비해 오픈 소스 표준과 오픈 데이터 형식을 더 많이 수용할 수 있다는 것이 가장 큰 장점입니다. 데이터 리더들은 이제 혁신을 가속화하고 동급 최고의 타사 도구를 활용할 수 있는 오픈 소스 표준의 가치를 깨닫고 있습니다. 이에 따라 응답자들은 더 강력한 보안 및 거버넌스를 가장 시급한 것에 가까운 요구 사항으로 꼽았습니다.

02 성장과 복잡성

조직에서 데이터를 관리하는 방식을 변화하는 속도는 놀랍고도 실망스럽습니다. 한때 고위 경영진이 운영의 부산물 정도로 생각했던 데이터는 이제 비즈니스 가치를 창출하는 가장 중요한 요소로 간주됩니다. 이제 기업에서 저장할 수 있고 분석해야 하는 구조적, 반구조적 및 비구조적 데이터 유형 전반에 걸쳐 데이터 생성량이 빠른 속도로 계속 증가하고 있습니다.

얼마 전까지만 해도 조직은 소수의 기술 대기업에 의존해 데이터 인프라 및 도구에 대한 요구 사항을 충족했지만, 요즘 기업 고객은 광대한 데이터 에코시스템 내의 수백 개 업체 중에서 선택할 수 있습니다. 이러한 업체들은 전례 없는 속도, 깊이 및 정교함으로 데이터를 구문 분석하면서도, 종종 머신 러닝까지 동원하는 새로운 분석 도구를 지속적으로 개발합니다. 끝없이 확장되는 클라우드는 조직에 데이터를 저장할 수 있는 방대한 공간과 더불어, 데이터를 더욱 비용 효율적인 방식으로 처리할 수 있는 엄청난 힘을 제공합니다. 마지막으로, CDO(최고 데이터 책임자) 및 CAO(최고 분석 책임자)의 부상에서 볼 수 있듯이 전략적 목표에 부합하는 새로운 비즈니스 가치를 창출하는 데 조직의 데이터 역량을 활용하는 새로운 역할과 구조가 다양한 수준에서 등장했습니다.

“예전에는 고객 경험의 다양한 요소에 대한 데이터를 얻는 것이 어렵고 비용이 많이 들었습니다.”라고 CVS Health의 최고 데이터 책임자이자 CVS Pharmacy의 최고 분석 책임자인 Bob Darin은 말합니다. “이제 이전에는 포착할 수 없었던 고객, 공급망, 사람들이

한때 선택적 기술 환경으로 여겨졌던 클라우드는 오늘날 데이터 관리 현대화의 기반이 되었습니다. **응답자의 63%**는 데이터 아키텍처에서 클라우드 서비스나 인프라를 광범위하게 사용합니다.

일하는 방식에 대한 인사이트를 얻을 수 있습니다. 우리는 이런 데이터를 규모에 따라 분석할 수 있는 모든 도구를 보유하고 있으며, 도구 사용 비용이 감소하고 있습니다. 이를 통해 우리는 대규모로 인사이트를 개발하고 고객 및 고객 워크플로에 통합할 수 있고, 따라서 고객에게 더욱 개인화되고 관련성 있는 경험을 제공할 수 있습니다.”

한때 선택적 기술 환경으로 여겨졌던 클라우드는 오늘날 데이터 관리 현대화의 기반이 되어, 비용은 절감하면서 더 큰 스토리지와 컴퓨팅 성능을 제공합니다. 설문 조사에 참여한 회사 중 63%는 데이터 아키텍처에서 클라우드 서비스 또는 인프라를 광범위하게 사용합니다. 이 중 1/3 이상(34%)이 여러 클라우드를 운영합니다.

그런데도 데이터 관리에 대한 불만이 많습니다. 데이터 플랫폼 업그레이드를 모색하는 동안 많은 기업이 통합이 어렵거나, 높은 비용을 발생시키거나 데이터 중복 및 품질 저하로 인해 문제를 일으키는 일으키는 레거시 온프레미스 사일로에 부담을 느끼고 있습니다. 이는 데이터 인프라와 관련하여 상당한 복잡성을



Databricks 관점: 레이크하우스 효과의 부상

어느 회사나 데이터 회사가 되고자 하는 욕구가 있으며, 엄청난 비즈니스 잠재력을 발휘할 수 있는 AI를 점점 중요하게 생각하고 있습니다. 그러나 이 보고서에서 알 수 있듯이, 오늘날 기업의 13%만이 기업 데이터 전략을 성공적으로 이행하고 있습니다. 데이터 및 분석 리더는 그들의 성공이 상당 부분 데이터 관리 기본 사항을 확실히 지켰던 덕분이라고 말합니다. 그렇다면 다른 사람들은 왜 어려움을 겪을까요?

문제는 데이터 아키텍처에서부터 시작됩니다. 연구에 따르면 비즈니스 분석, 데이터 엔지니어링, 스트리밍 및 ML 등 모든 데이터 워크로드를 처리하려면 4가지 스택을 구축해야 합니다. 이 네 가지 스택은 모두 매우 다른 기술이 필요하며, 불행히도 함께 작동하지 않는 경우가 있습니다.

데이터 웨어하우스와 데이터 레이크를 아우르는 기술 에코시스템으로 인해 아키텍처는 더욱 복잡해집니다. 결국

관리하는 데 비용과 리소스가 많이 들어갑니다. 이러한 복잡성은 데이터 팀에 영향을 미칩니다. 데이터와 조직의 사일로는 의도치 않게 의사 소통 속도를 늦추고 혁신을 방해하며 각 팀의 목표가 분산될 수 있습니다. 그 결과 중복된 데이터 사본, 일관적이지 못한 보안/거버넌스 모델, 폐쇄형 시스템, 데이터팀 생산성 저하 등의 문제가 발생합니다.

한편 ML은 여전히 구현하기 쉽지 않습니다. 레이크하우스 아키텍처의 등장으로 조직은 더 이상 레거시 아키텍처의 한계와 복잡성에 얽매이지 않습니다. 데이터 웨어하우스의 성능, 안정성 및 거버넌스를 데이터 레이크의 확장성, 저비용 및 워크로드 유연성과 결합함으로써 레이크하우스 아키텍처는 실시간 스트리밍, 일괄 처리, SQL 분석, 데이터 과학 및 ML을 포함해 다양한 데이터 애플리케이션을 위한 유연한 고성능 설계를 제공합니다.

다음 페이지에 계속

발생시킵니다. 클라우드 는 온갖 혁신적인 효과를 제공함에도 불구하고, 조직이 공급업체 종속에 대비하고 지역적 요구 사항을 충족하거나 동급 최고의 솔루션을 최적화하려는 목적으로 여러 공급업체에 데이터를 지속적으로 저장하면서 복잡성을 증가시킬 수 있습니다. 그리고 데이터 아키텍처는 비교적 짧은 시간에 발전했기 때문에 조직에서는 환경 설정이나 거버넌스, 기타 기능을 수행하는 다양한 클라우드 기반 도구와 더불어 온프레미스 데이터베이스, 데이터 웨어하우스, 데이터 레이크 및 기타 새로운 데이터 아키텍처를 동시에 사용할 수 있습니다.

“아키텍처가 정말 복잡해진 건 우리가 과도하게 복잡하게 만드는 경향이 있기 때문입니다.”라고 Estée Lauder Companies의 최고 분석 책임자인 Sol Rashidi는 말합니다. “그 이유는 핵심을 비켜가기 때문입니다. 우리는 문제를 해결해 줄 것을 기대하며

최신 기술과 플랫폼을 너무나 쉽게 도입합니다. 그러나 도구들을 활용할 준비가 되어 있고, 인사이트를 추출할 만큼 성숙도를 갖추고, 프로세스와 논리가 일치하지 않는 한 중구난방인 아키텍처가 더욱 악화할 뿐입니다.”

조직이 복잡성을 관리할 수 없는 경우, 일반적으로 기회 상실(예: ML 사용 사례에서 수익 제공 실패), 비용 증가(예: 여러 중복 시스템 관리 및 지원)를 경험하는 데다, 점점 엄격해지는 데이터 규제를 준수하기 어렵고 궁극적으로는 상당한 경쟁에 노출됩니다.

그럼에도 불구하고, 우리 조사에서 여실히 알 수 있듯이, 데이터 및 기술 리더들은 비즈니스를 위해 데이터를 효과적으로 관리할 수 있는 현재와 미래의 능력에 관해서는 좌절감보다는 열정과 낙관적인 생각이 더 큼니다.

파트너 관점

Databricks에서는 Databricks Lakehouse Platform을 통해 레이크하우스 아키텍처를 구현합니다. 이 혁신의 핵심은 Delta Lake의 지원에 있습니다. Delta Lake는 이 플랫폼의 핵심이며 선별된 데이터 레이크를 생성해 데이터 웨어하우스에서 기존 데이터 레이크로 직접 안정성, 성능 및 거버넌스를 더해줍니다. 기업에서는 전사적 데이터 관리를 좀 더 잘 이해할 수 있게 됩니다.

Databricks Lakehouse Platform은 3가지 면에서 탁월합니다.

간단합니다. 하나의 공통 플랫폼에서 모든 워크로드를 지원하기 위해서 데이터가 하나만 있으면 됩니다.

개방적입니다. 오픈 소스 및 개방형 표준을 기반으로 하므로 기존 도구로 작업하기 쉽고 독점적 형식을 피할 수 있습니다.

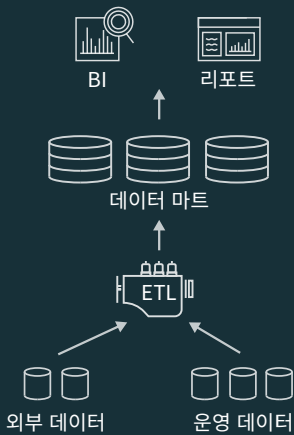
협력적입니다. 데이터 엔지니어, 애널리스트 및 데이터 과학자들이 더 효율적으로 함께 작업할 수 있습니다.

Databricks Lakehouse Platform이 제공하는 비용 절감, 효율성 및 생산성 향상은 이미 모든 업종과 지역의 기업에 실질적인 영향을 미치고 있습니다. Databricks는 지나치게 복잡한 아키텍처에서 벗어나, 모든 주요 클라우드 공급업체의 모든 데이터 및 워크로드에 대해 하나의 공통 클라우드 기반 데이터 기반을 제공합니다. 데이터 및 분석 리더는 예측적 유지 관리를 통해 매일 계획해야 하는 부담과 온갖 복잡성을 낮추고, 가치를 더하는 데 중점을 둔 데이터 중심 문화를 조성할 수 있습니다.

비디오 스트리밍 분석에서 고객 평생 가치에 이르기까지, 그리고 질병 예방에서 화성에서 생명체 찾기에 이르기까지 데이터는 솔루션의 일부입니다. 데이터에 대한 이해가 문을 여는 열쇠입니다.

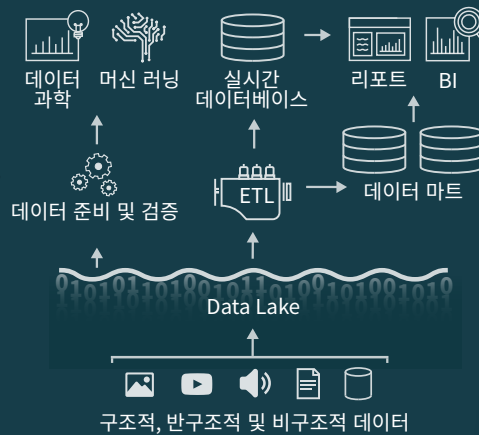
1980년대 후반

데이터 웨어하우스



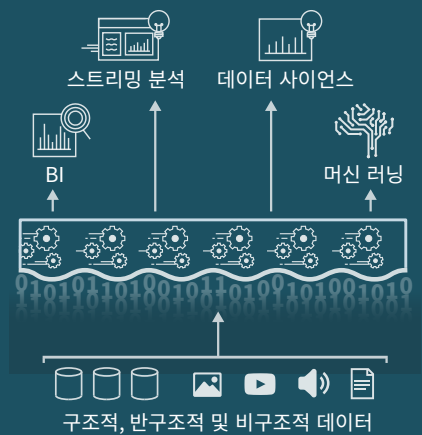
2011

데이터 레이크



2020

레이크하우스



03 전략 조정 및 실행

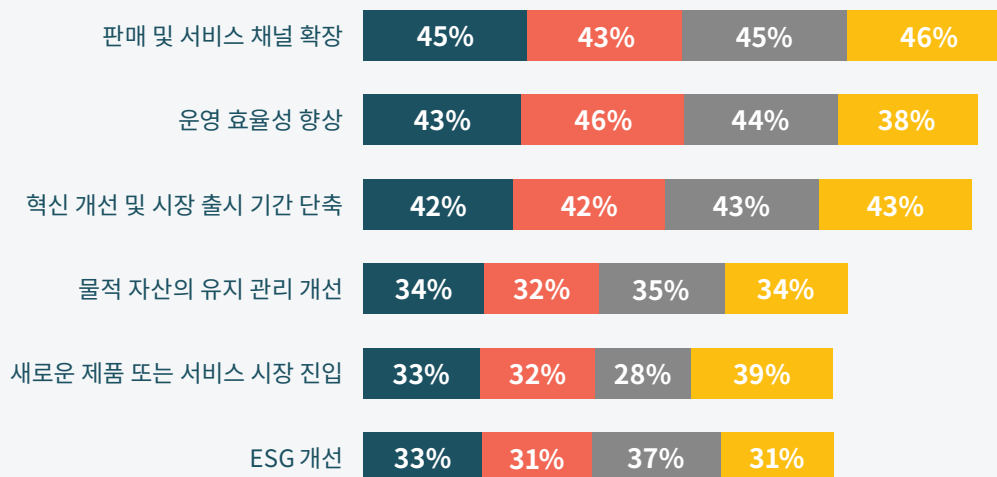


100 년 가까이 볼 수 없었던 규모의 글로벌 경제 침체 속에서 기업들은 야망을 누르고 수익을 올리는 데 집중할 것으로 예상했습니다. 그러나 이번 설문 조사 응답자 대부분이 성장을 지향하는 것으로 나타났습니다. 향후 2년 내 엔터프라이즈 데이터 전략에 설정한 가장 중요한 비즈니스 목표가 무엇인지 질문했을 때 효율성 향상을 지정한 응답자(43%) 보다 확장된 판매 및 서비스 채널의 형태로서 매출 성장을 강조하는 응답자(45%)가 더 많았습니다. 그다음은 혁신 개선과 신제품 및 개선된 제품의 출시 기간 단축으로 나타났습니다(42%).

설문 조사에 참여한 기업의 향후 2년간 주요 데이터 이니셔티브를 살펴보면 성장 지향적인 비즈니스 전략과 상당한 수준 일치하는 것을 알 수 있습니다. 또한 이러한 비즈니스 목표를 지원하려면 데이터 관리를 시급히 개선해야 한다는 필요성을 느끼고 있습니다. 이런 이니셔티브는 클라우드 네이티브 플랫폼을 광범위 도입으로 뒷받침됩니다.

가장 자주 언급되는 우선순위는 데이터 품질 및 처리 개선을 통한 데이터 관리 개선입니다. 응답자의 48%가 이 내용을 언급했습니다. (석유 및 가스 회사는 74%, 소비재 회사의 67%입니다.) 이러한

그림 1: 엔터프라이즈 데이터 전략을 위해 향후 2년간 기업의 가장 중요한 비즈니스 목표(최고 응답, 응답자 %)



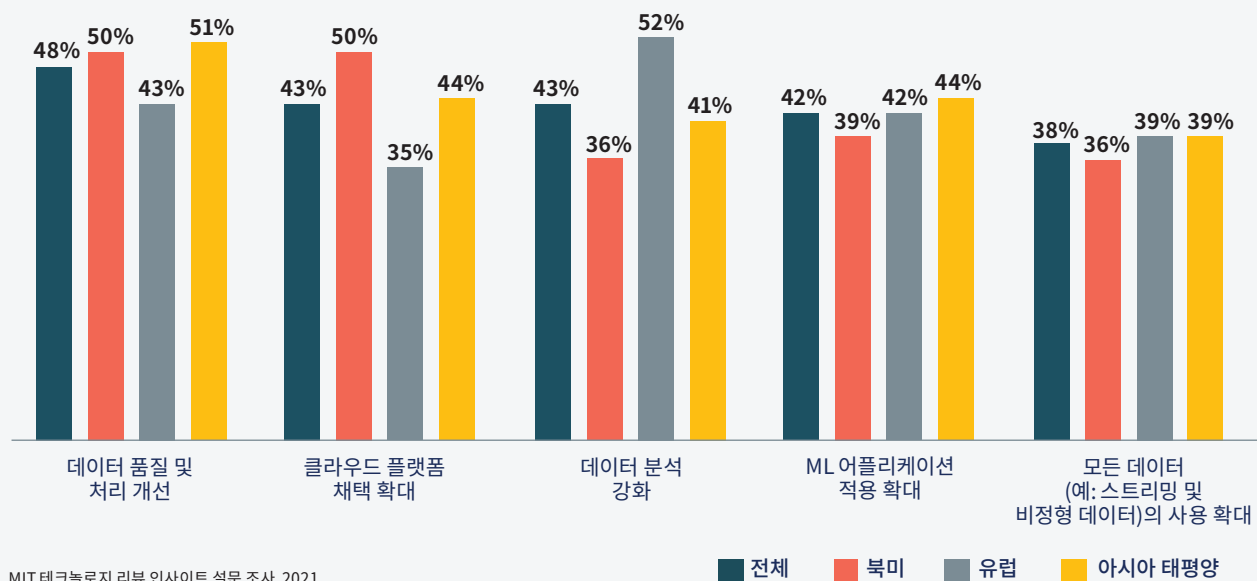
ML이 중심이 되는 성장 지향적인 활동을 빠르게 추진하려면 데이터 품질 및 처리를 개선하여 더 나은 데이터 관리를 달성하는 것이 중요합니다.

노력은 ML이 주도하는 것과 같은 성장 지향적인 활동을 빠르게 추진하는 데 중요합니다. 제품이 인공지능(AI)으로 구동되는 호주 기반 소매 기술 회사인 Hivery에게는 대량 수집 능력보다 고객 데이터의 품질이 훨씬 더 중요합니다. Hivery의 CTO인 Andy McQuarrie는 사실, 고객 데이터가 깨끗할수록 발생하는 수집 문제가 줄어든다고 말합니다.

클라우드 플랫폼 채택 증가(43% 응답), 데이터 분석 강화(43%), ML 적용 확대(42%) 등, 설문에 응한 기업의 다른 최우선 데이터 우선순위가 충족되면 무엇보다도 데이터팀이 새로운 판매 및 서비스 기회를 신속하게 활용하고 새로운 데이터 제품 개발을 지원하는 데 필요한 추가 용량, 전력 및 규모를 얻을 수 있습니다. 물론 운영 효율성 향상이라는 목표도 전적으로 지원합니다. 또 다른 우선순위(38% 응답)는 스트리밍, 비정형 및 기타 다양한 데이터의 사용을 확대하는 것입니다.

지금 데이터 전략이 전체 비즈니스 목표와 밀접하게 연계되어야 한다는 자명한 것처럼 보이지만, 조정의 중요성에 대해서는 잘 모르는 경우가 있습니다. 미국 금융 서비스 회사인 Northwestern Mutual의 최고 데이터 책임자인 Don Vu에 따르면, CDO가 영향력을 행사하고 데이터 책임이 간소화된 조직 구조로 통합되면서 많은 기업에서 데이터와 비즈니스 전략이 훨씬 긴밀하게 연결되었습니다. “조정이 중요하다는 것은 이미 알고 있었지만, 우리 팀이 다양한 비즈니스 전략 이니셔티브를 실제로 제공하는 방법에 대해 더욱 깊이 고민하면서 분명해졌습니다. 기본적인 정보 출처에 대한 신뢰나 데이터 사용의 민주화와 같은 개념에서 비즈니스 전략과의 관계가 훨씬 더 명확해졌습니다.” 라고 Don Vu는 말합니다.

그림 2: 향후 2년간 기업의 가장 중요한 전사적 데이터 전략 이니셔티브(최고 응답, 응답자 %)



데이터로 높은 성과를 내는 조직

데이터 관리에 탁월한 대기업은 많지 않습니다. 이는 설문조사에서 데이터 전략 제공과 관련하여 1~10점을 기준으로 했을 때 최고점(9-10)을 매겨 조직의 성과를 높게 평가한 응답자가 13%에 불과한 데서도 알 수 있습니다. 이들의 경영진은 데이터 “고성과자”들은 여러 사업부에 걸쳐 측정 가능한 비즈니스 영향을 제공한다고 말합니다. 이들은 6점 이하로 평가된 비슷한 규모의 “저성과자” 그룹(표본의 12%)과 대조됩니다.

이 두 집단은 목적뿐만 아니라 특정 속성에서도 큰 차이가 있습니다. 예를 들어 클라우드를 높은 성취자들의 데이터 아키텍처에서 더 두드러지게 나타납니다. 이 집단의 74%가 클라우드 환경에서 데이터 서비스나 인프라의 절반 이상을 실행하는 반면, 낮은 성취자 그룹은 60%에 불과합니다. 데이터 우선순위와 관련하여, 대부분의 성취도가 낮은 기업(59%)은 향후 2년 동안 데이터 관리(데이터 품질 및 처리) 개선에 중점을

그림 3a: 조직의 엔터프라이즈 데이터 전략 성공적 이행 정도(10을 성공이라고 했을 때 1-10 기준의 자체 평가 등급)

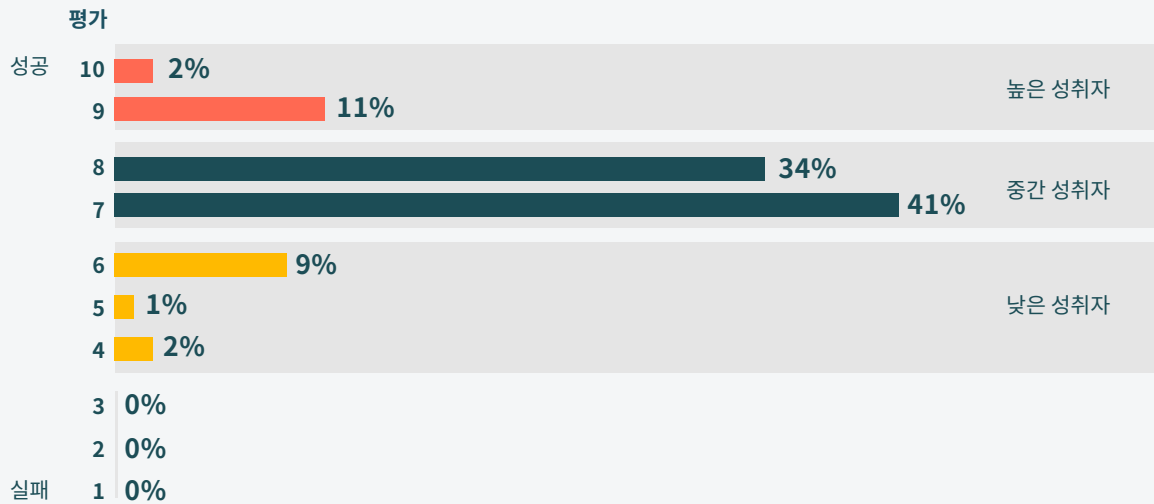


그림 3b: 높은 성취자: 여러 비즈니스 단위에 걸쳐 측정 가능한 비즈니스 영향과 함께 기업 데이터 전략 전달에 대해 조직을 9 또는 10으로 평가한 응답자(전체, 지역 및 섹터 산업)

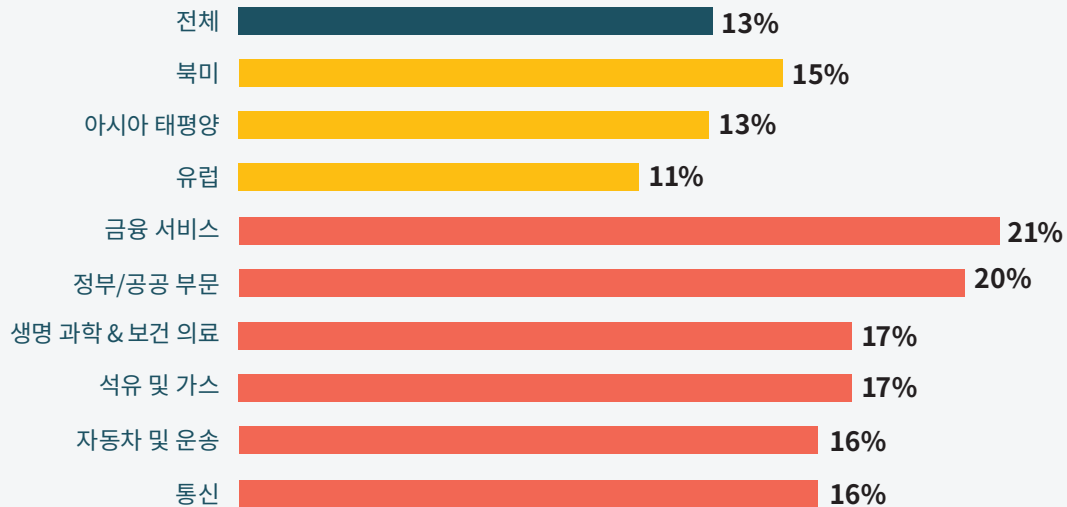
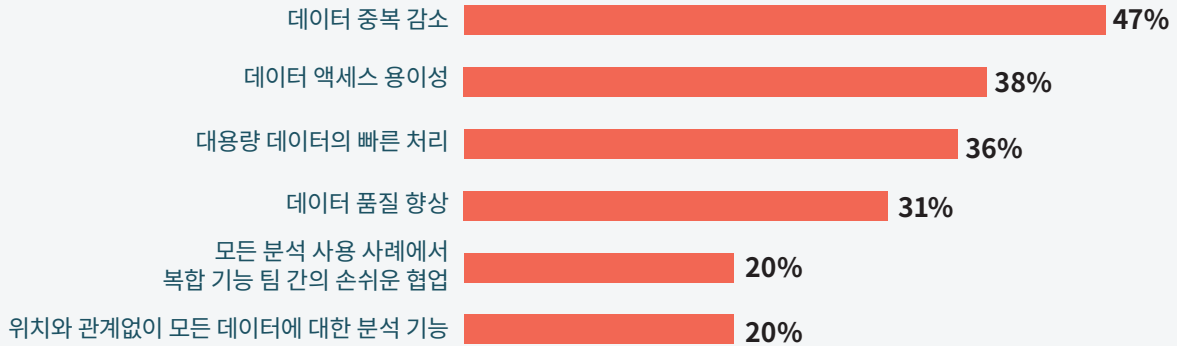


그림 4: “높은 성취자” 조직이 데이터 전략 이니셔티브를 이행하는 주요 성공 요인(최고 응답, 응답자 %)



MIT 테크놀로지 리뷰 인사이트 설문 조사, 2021

두었지만, 높은 성취를 달성한 기업은 ML 적용의 확대를 가장 자주 언급했습니다(53%).

높은 성취를 이룬 집단은 기본을 당연하게 여기지 않고, 성공의 비결이 건전한 데이터 관리의 기초에 세심한 주의를 기울인 덕분이라고 생각합니다. 여기에는 엔터프라이즈 최종 사용자가 데이터에 쉽게 액세스하도록 지원하고 대량의 데이터를 고속으로 처리하는 등, 데이터 중복 감소와 데이터 민주화의 특징도 포함됩니다.

대규모 조직은 데이터 웨어하우스, 운영 체제, 리포트, 대시보드 및 데스크톱 도구와 같은 여러 수준에서 데이터 복제가 발생합니다.

Macquarie Bank의 최고 데이터 및 분석 책임자인 Ashwin Sinha는 이런 행위가 상당한 비용, 위험 관리 및 안정성에 영향을 미친다고 말합니다. 데이터 복제는 조직 전체에서 머신 러닝을 확장하고 효과적으로 사용하는 능력에도 영향을 미칩니다.

발전을 가로막는 요인이 무엇인지 묻는 질문에 낮은 성취자들은 데이터 관리 플랫폼의 제한된 확장성을 가장 많이 지목했습니다. 그 외에도 대용량 데이터의 느린 처리와 협업 촉진의 어려움도 장애 요인으로 자주 언급되었습니다. 앞으로 살펴보겠지만, 규모, 속도 및 협업은 데이터 운영 전반에 걸쳐 조직이 해결해야 할 과제이기도 합니다.

그림 5: “낮은 성취자” 조직이 데이터 전략 이니셔티브를 이행하지 못하게 하는 주요 성공 요인(최고 응답, 응답자 %)



MIT 테크놀로지 리뷰 인사이트 설문 조사, 2021

Nielsen: 데이터 중심 비즈니스를 위한 데이터 변환



처

음부터 데이터를 비즈니스 모델의 중심으로 삼았던 장수 조직 중 하나인 Nielsen은 건전한 데이터 관리의 중요성을 무엇보다도 강조합니다. Nielsen의 패널은

소비재 회사에게 고객이 어떤 제품을 구매하는지, 고객 행동이 어떻게 변화하는지 알려줍니다. 또한, 텔레비전 광고를 언제 어디에 배치해야 하는지도 조언합니다. 회사의 최고 데이터 책임자로 2년차, 최고 연구 책임자로 5년차에 접어든 Mainak Mazumdar는 데이터 관리 및 인프라 혁신을 주도했습니다. 그는 그 때의 기억을 이렇게 회상합니다. “몇 년 전만 해도 우리는 데이터 볼륨이 매일 페타바이트씩 증가하는 가운데 데이터 자체와 데이터 액세스에 필요한 집단 지식이 사일로화되는 단편화 문제와 메타데이터 및 거버넌스 부족 문제로 어려움을 겪었습니다.”

Mazumdar는 이제 상황이 달라졌다고 말합니다. “이제 우리는 플랫폼에서 보유하는 전문가를 20~30명에서 300명 이상으로 빠르게 확장할 수 있습니다. 데이터가 큐레이팅되고, 레이블이 지정 및 정의되고, 메타스토어가 있으며, 통합되는 데이터 레이크가 있는 클라우드 플랫폼을

사용합니다. 자체적인 분석 엔진도 구축했습니다. 사실, 과거에 소프트웨어 엔지니어가 하던 일의 대부분은 이제 우리 팀이 수행하여 프로덕션에 직접 배포합니다.” Mazumdar는 이러한 변화로 인해 팀의 사이클 시간이 50% 단축되었다고 말합니다. “우리 모델의 속도는 이제 약 50배 향상되었습니다. 20분이 걸리던 일을 1분 안에 할 수 있죠. 동시에 우리는 데이터 과학에 쉽게 접근할 수 있는 방대한 양의 데이터를 수집하고 처리하고 있습니다. 이것은 엄청나 변화입니다.”

예를 들어, Nielsen이 이러한 능력을 사용하여 미국의 약 200개 지역 시장에서 새로운 등급 제품을 출시하는 등의 방법으로 성장을 지원했습니다. “고인식 딥러닝 모델”은 TV 셋톱박스에서 대량의 데이터를 처리함으로써 Nielsen에서 시청자가 특정 시간에 시청하는 프로그램뿐만 아니라, 이전에는 불가능했던 각 가정에서 누가 시청하고 있는지까지도 예측할 수 있도록 지원합니다. Mazumdar는 “데이터 관리 방식을 변경하지 않았더라면 이 제품을 출시할 수 없었을 뿐만 아니라 이만한 양의 데이터를 수집할 수조차 없었을 것입니다.”라고 말합니다. “모델과 제품이 개선되는 건 그 어느 때보다 많은 양의 데이터를 수집한 덕택입니다.”

04 분석 및 머신러닝 확장

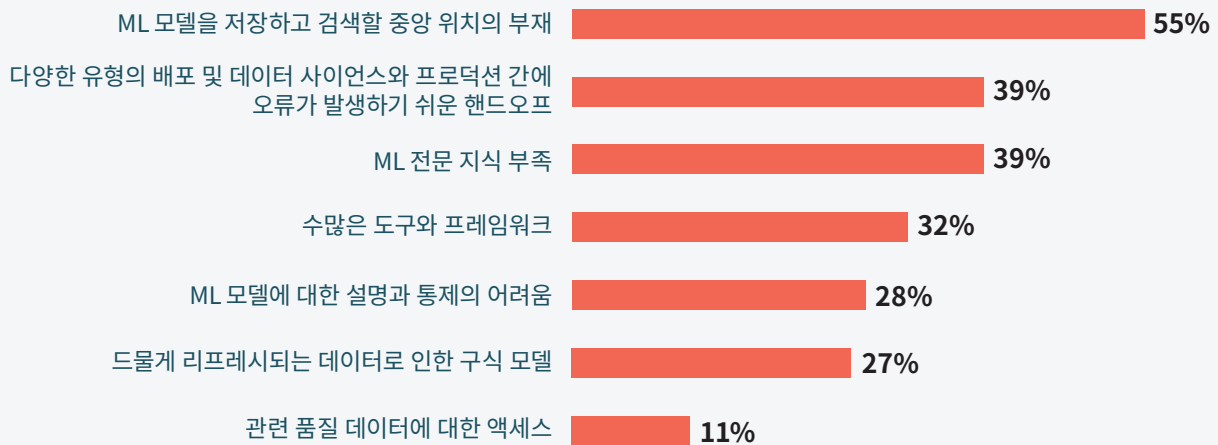


비즈니스 리더는 수요에 보조를 맞추고 예측하고, 경쟁 압력을 관리하고, 효과적으로 혁신하고, 효율적으로 운영할 수 있는 회사의 능력이 분석 및 ML의 속달에 달려 있다는 것을 알고 있습니다.

거의 모든 산업 분야의 조직은 비즈니스에 더 큰 영향을 가져올 분석 및 ML 사용 사례를 개발하느라 분주합니다. 대부분 대기업에서 프로덕션 및 대규모 사용 사례를 아우르는 광범위한 포트폴리오는 더 이상 없으면 좋은 것이 아니라 반드시 있어야 하는 것입니다. CDO와 그 팀은 점점 이러한 사례를 제공하는 데 기여한 것을 척도로 평가받고 있습니다.

많은 조직이 이 문제, 특히 상당한 영향력을 발휘하는 데 필요한 규모를 달성하는 데 어려움을 겪고 있습니다. Estée Lauder의 Sol Rashidi에 따르면, 욕심이 과하기 때문입니다. “ML을 사용하면서 기본을 마스터하지 못한 채 기기와 걷기는 건너뛰고 곧바로 달리고 싶어 하는 기업이 많습니다.” Northwestern Mutual의 Don Vu 등 다른 CDO의 경우, 프로덕션에 배포할 적절한 사용 사례를 선택하는 것을 가장 어려워합니다. 그는 비즈니스 사용자 입력이 없으면 비즈니스 목표에 명확하게 매핑되지 않는 사례를 선택할 가능성이 높아진다고 말합니다.

그림 6: ML 사용 사례를 확장할 때 기업이 직면하는 주요 문제(최고 응답, 응답자 %)



CVS Health의 패러다임 전환



미국은 항상 사회에서 필수적인 역할을 해왔지만, 작년에는 그 어느 때보다도 많은 역할을 했습니다. 미국에서 매출이 가장 큰 CVS와 같은 약국 체인은 고객이 약물 및 기타 의료 서비스에 대한 최신 정보를 받을 수 있도록 더욱 발전된 데이터 기능을 사용하여 기저 질환이 심각한 건강 문제로 악화되지 않도록 관리해줍니다. Bob Darin은 CVS Health의 최고 데이터 책임자 및 소매 약국 비즈니스의 최고 분석 책임자로서 이런 여러 가지 활동을 이끌고 있습니다.

이 회사는 오래전부터 데이터 시스템을 사용하여 고객이 최신 약물을 사용하도록 유도해왔습니다. 여기에는 전화나 문자를 통한 환자 지원, 약국 카운터에서의 안내, 환자가 특정 후속 조치 또는 약물 검토에 대해 의료 제공자와 상담하도록 권장하는 것 등이 포함됩니다. Darin은 최근 몇 년 사이에 이러한 이니셔티브가 데이터 과학에 포함되어 더욱 개인 맞춤형으로 변모하고 있다고 말합니다. “우리는 이제 어떤 알림, 프로그램, 커뮤니케이션 모드 및 메시지가 사람들이 약 복용을 기억하고, 픽업 일정을 잡고, 비용 절감 기회를 이해하도록 돕는 데 가장 효과적인지 알고 있습니다.” 과거에는 획일적인 프로세스였으나 이제는 모두 분석 모델에 의해 이루어지고 있다고 말합니다.

Darin은 “우리에게는 패러다임의 전환이었습니다.”라고 말합니다. “분석을 사용하는 방법과 더불어 데이터 플랫폼 및 아키텍처를 관리하는 방법에서도 말입니다.” 그는 여기에는 데이터 웨어하우징, 운영 보고 및 분석에 대한 투자가 포함되어 있다고 말합니다. 그러나 그의 팀은 데이터 웨어하우징 및 설명 분석 사용 사례를 지원할 수 있을 뿐만 아니라 고급 머신 러닝, 알고리즘 개발, 대규모 최적화 기술을 사용할 수 있는 하이브리드 다차원 데이터 환경으로 더 많은 비즈니스 기능을 이동하고 있습니다.

이런 환경은 “사람들이 매일 비즈니스 결정을 내리는 데 필요한 정보인 설명과 쿼리와 같은 일상적인 비즈니스 분석은 물론이고, 고급 데이터 과학 애플리케이션까지 지원해야 합니다. 우리는 두 가지 모두를 지원하는 아키텍처와 데이터 플랫폼이 필요합니다.”라고 말합니다.

전력 부문에서 일하는 직원의 3/4, 그리고 정부 및 소비재 부문 종사자의 2/3 이상이 ML 모델을 저장하고 발견할 수 있는 중앙 위치의 부족이 확장에 큰 걸림돌이라고 말합니다.

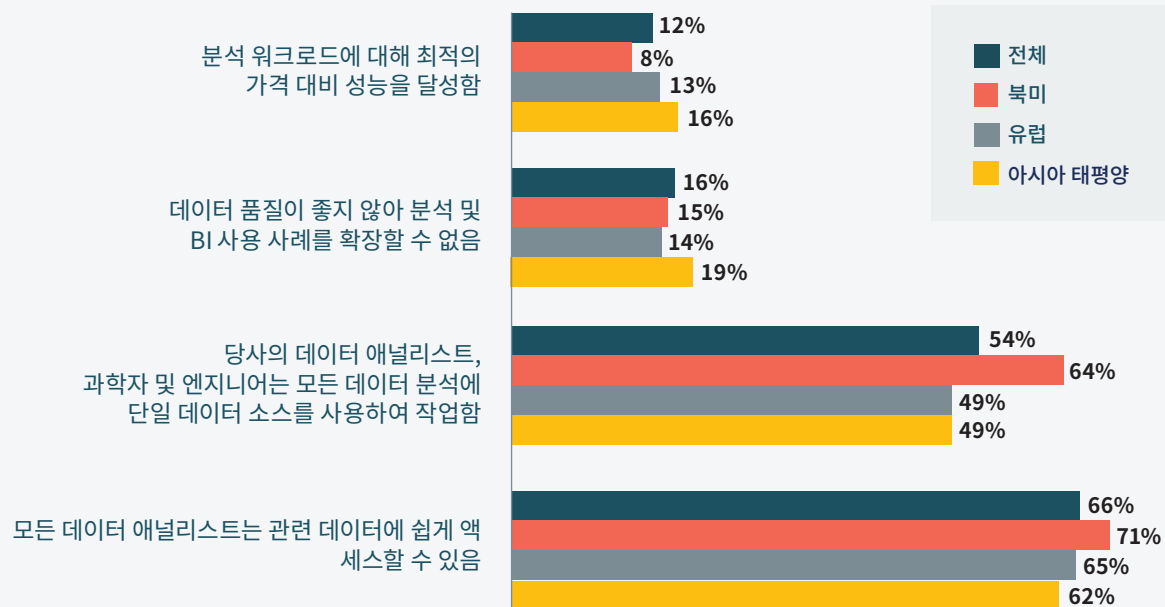
ML 모델을 구축하는 담당자와 비즈니스 사용자가 긴밀한 대화를 나누면 “빛나는 알고리즘” 효과를 방지할 수 있습니다. 이 효과는 데이터 과학팀에게는 가장 흥미롭지만 비즈니스 사용자에게는 거의 쓸모가 없는 ML 사용 사례를 우선순위로 지정하는 것을 말합니다. Hivery의 Andy McQuarrie는 “머신 러닝 사용 사례가 비즈니스 사용 사례와 매우 밀접하게 매핑되어 있지 않다면, 대체 무슨 소용이 있는지 궁금합니다.”라고 말합니다.

확장의 장벽

기업은 ML 수명 주기를 전체적으로 관리하는 데 복잡성이 커서 어려움을 겪고 있습니다. 이는 응답자 조직이 사용 사례를 확장할

때 직면하는 어려움으로 언급한 내용에도 나타나 있습니다.(그림 6). 응답자의 55%는 ML 모델을 저장하고 검색할 중앙 위치가 없다는 것을 가장 일반적인 문제로 언급했습니다. 전력 및 기타 유틸리티 분야에서 일하는 사람들의 3/4, 그리고 정부 기관과 소비재 회사 직원의 2/3 이상이 규모 확장의 큰 장애물이라고 답했습니다. 응답자의 39%는 데이터 과학과 프로덕션 간의 협업이 부족해서 여러 번 배포하고 핸드오프에서 오류가 발생하기 쉬운 것도 큰 장애물이라고 답했습니다. 생명 과학 및 의료 기관 응답자는 50%가 여기에 동의했습니다. 그리고 도구의 다양성과 프레임워크, 드물게 리프레시되는 데이터로 인해 어려움을 겪는 조직도 많습니다.(각각 32%, 27% 언급)

그림 7: 조직의 분석 및 비즈니스 인텔리전스(BI)에 대한 설명에 전적으로 또는 다소 동의하는 응답자(응답자 %)





“데이터 사이언스와 프로덕션 사이에서 여러 번 배포가 필요하고 핸드오프에서 오류가 발생하기 쉬운 것은 큰 문제입니다. 데이터 과학의 결과와 실질적 결과 사이에는 차이가 있는 경우가 많습니다.”

– L’Oréal 데이터, CRM & 분석 부사장 Naveen Jayaraman

우리가 인터뷰한 CDO들은 데이터 사이언스와 프로덕션 사이에서 여러 번 배포하고 핸드오프에 오류가 발생하기 쉬운 문제가 있다는 데 공감했습니다. “이것은 엄청난 문제입니다”라고 L’Oréal의 데이터, CRM & 분석 부사장인 Naveen Jayaraman은 말합니다. 데이터 과학의 결과와 실질적 결과 사이에는 차이가 있는 경우가 많습니다.”

Bob Darin에 따르면, CVS Health는 핸드오프 문제를 줄이는 데 성공했습니다. “과거에는 데이터 과학자가 모델을 구축하고 인사이트를 제공했습니다. 다른 팀이 이를 받아서 구현했습니다. 그 ‘폭포수’ 방식은 효과가 없었고 모델도 확장이 어려웠습니다.” CVS Health의 경우 IT와 협력하여 데이터 과학과 프로덕션을 하나의 팀으로 통합하는 것이 이러한 어려움을 완화하는 데 도움이 되었다고 합니다. “한 팀은 인사이트를 개발하고 다른 팀은 확장을 수행하는 것보다 훨씬 낫습니다.”

기술 격차는 여전히 성공적인 ML 사례를 구축하려는 조직의 시스템 문제로 남아 있습니다. 응답자의 39%가 사용 사례 확장의 주요 장벽으로 ML 전문 지식의 부족을 언급했고, 설문 조사에 참여한 제조업체(64%)도 이를 중요하게 꼽았습니다. 높은 성취자들은 이러한 격차를 극복할 수 있는 것으로 보이며 그중 27%만이 ML의 문제로 언급했습니다. 반면, 낮은 성취자들은 이 문제를 크게 느꼈고 59%가 중요한 문제라고 말했습니다.

투자 수익 보호

ML 및 분석 사용 사례를 의도한 규모로 구축하는 어려움은 투자 수익(ROI)에 영향을 미칩니다. 예를 들어, 설문 응답자의 12%만이 분석 워크로드에 대해 최적의 가격 대비 성능을 달성했다고 말했습니다. 앞서 언급한 데이터 중복 문제는 성능을 저하시킬 수 있습니다. 낮은 품질의 데이터를 정제하는 수준 낮은 작업에 안 그래도 부족한 리소스를 할당해야 하기 때문입니다. 그리고 Jayaraman은 전사적 규모로 확장하려면 클라우드 인프라 및 데이터 관리에 대한 투자가 수반되는 경우가 많다고 말합니다. “이는 까다로운 영역입니다.”라고 그는 말합니다. “사용 사례에서 매출이나 효율성이 크게 향상되지 않는 한 항상 ROI에서 큰 문제가 발생합니다.”

반면, 패스트푸드 대기업 McDonald’s의 데이터 과학 수석 이사인 Patrick Baginski는 성과 기대치를 충족하거나 초과하면 ROI가 제공하는 단순한 이점 이상의 것을 얻을 수 있다고 말합니다. ML과 데이터 과학의 가치를 더 빨리 입증할수록 사용자의 동의를 더 빨리 얻고, 경영진도 두 가지 모두 조직에 얼마나 중요한지 확신할 수 있게 됩니다.” 그러면 데이터 문화가 발전하는 데도 중요한 역할을 합니다.

기술, 민주주의, 그리고 문화

데

이터 관리는 시스템, 리더십 및 비즈니스 가치에 대한 인식에 그동안 많은 변화를 일으켰지만, 여전히 대부분 조직에서는 최종 사용자, 프런트오피스나 일상적인 결정을

내리는 데 데이터 인사이트가 필요한 백오피스 직원 사이에 상당한 의견 차이가 존재합니다. 많은 CDO가 정기적으로 사용자와 상호작용하는 사업부에 데이터 과학자를 참여시켜 이 격차를 줄이려고 합니다. 격차를 메우는 또 다른 방법은 사용자가 분석을 직접 처리하여 필요에 따라 인사이트를 스스로 얻을 수 있도록 하는 것입니다. McDonald's의 Patrick Baginski는 사용자가 있는 위치 가까이로 데이터를 밀어주어야 한다고 말합니다.

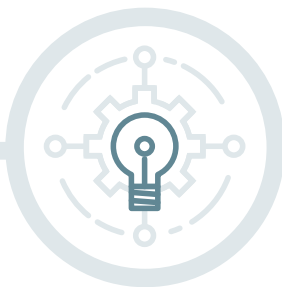
우리가 인터뷰한 모든 CDO와 CAO는 모두 같은 목표를 추구했습니다. Bob Darin은 “도구가 발전하고 사람들이 고급 분석 및 데이터 과학에 익숙해질수록, 사용자에게 다른 사람이 생성한 분석을 사용하지 않고 스스로 분석을 실행할 수 있는 기능을 제공해야 합니다.”라고 말합니다.

이것이 데이터 민주화의 실제입니다. 또한 일부 CDO는 데이터 활용 능력을 구축할 뿐만 아니라 조직의 데이터 문화 성장을 장려하는 수단으로 활용하기도 합니다. 데이터 인프라와 도구는 여기에 직접적으로 기여합니다. Don Vu는 “중앙 데이터 팀으로서 우리의 목표는 대부분 책임과 권한을 다른 사람에게 연결해주는 도구를 만드는 것입니다.”라고 말합니다. “그래서 우리는 데이터 플랫폼의 기능을 옛지까지

제공하려고 노력하고 있습니다. 우리는 기능적 관점에서 애플리케이션 팀과 사용자 사이에 구심점을 만들고 싶습니다. 사람들이 이 플랫폼에서 데이터 품질을 관리하고 개인 정보를 보호하며 액세스를 관리하기를 원합니다.”

Macquarie Bank의 Ashwin Sinha에 따르면, 최종 사용자가 데이터 및 분석에 쉽게 액세스할 수 있도록 개선하면 데이터 문화를 향상하는 데 직접적으로 기여합니다. “애널리스트에게만 좋은 건 아닙니다.”라고 그는 말합니다. “분석적 사고방식을 가진 고위 리더는 간단한 인터페이스가 있는 데이터 시각화 도구로 데이터를 분석하고 인사이트를 얻습니다.” Sinha는 액세스를 향상하는 데 시각화 도구가 가장 중요한 역할을 했지만, 팀에서 클라우드 인프라를 사용하여 안정성과 성능을 개선하는 것도 중요하다고 말합니다. “규제, 머신 러닝, 분석 등의 데이터 워크로드는 모두 지난 3년 동안 클라우드 데이터 플랫폼으로 전환되었습니다. 데이터 플랫폼은 다양한 클라우드 서비스, 오픈 소스 패키지 및 데이터 관리 도구를 통해 데이터의 전체적 계보 및 추적 기능을 사용하여 조직 전체의 데이터가 관리 목적에 맞게 표준화, 통합 및 공급되도록 합니다.”

Sinha는 강력한 데이터 거버넌스, 액세스 용이성, 단순화가 모두 결합되어야 데이터 및 분석에 대한 사용자의 신뢰가 커진다고 말합니다. 데이터 문화는 데이터 플랫폼에 대한 신뢰와 신뢰성 없이는 발전할 수 없습니다.



“어떤 데이터가 가장 중요한지 파악하고, 우선순위를 정하고, 데이터를 보호 및 관리하는 원칙을 구축한 후, 데이터를 민주화하여 데이터 전문가와 최종 사용자가 혁신에 필요한 인사이트를 추출할 수 있도록 하십시오.”

– Estée Lauder Companies 최고 분석 책임자 Sol Rashidi

05 미래 비전



실

문조사에 응한 임원의 절반(기술 및 제조 회사의 임원 중 약 2/3)은 현재의 데이터 문제를 해결하기 위해 현재 새로운 데이터 플랫폼을 평가하거나 구현하고 있다고 답했습니다. 9%는 그럴 계획이 있지만

아키텍처를 변경하는 데 방해가 되는 요소를 해결해야 한다고 답했습니다. 여기서도 높은 성취자들과 낮은 성취자들이 뚜렷하게 구분됩니다. 대부분의 높은 성취자(56%)들은 현재 아키텍처에 만족하고 있지만, 그중 약 1/3이 새로운 플랫폼을 연구하거나 공급업체를 평가하고 있습니다. 이와 대조적으로, 낮은 성취자들은 24%만이 기존 아키텍처에 만족하고 59%는 적극적으로 변화를 모색하고 있습니다. 기술 회사와 제조업체는 다른 부문보다 플랫폼을 업그레이드할 가능성이 훨씬 더 큼니다.

새로운 아키텍처를 위한 CDO 위시리스트

데이터 및 기술 리더가 비즈니스를 위한 새로운 데이터 아키텍처를 구축할 수 있다면 어떨까요? 우리는 그들에게 기존 제품에 비해서

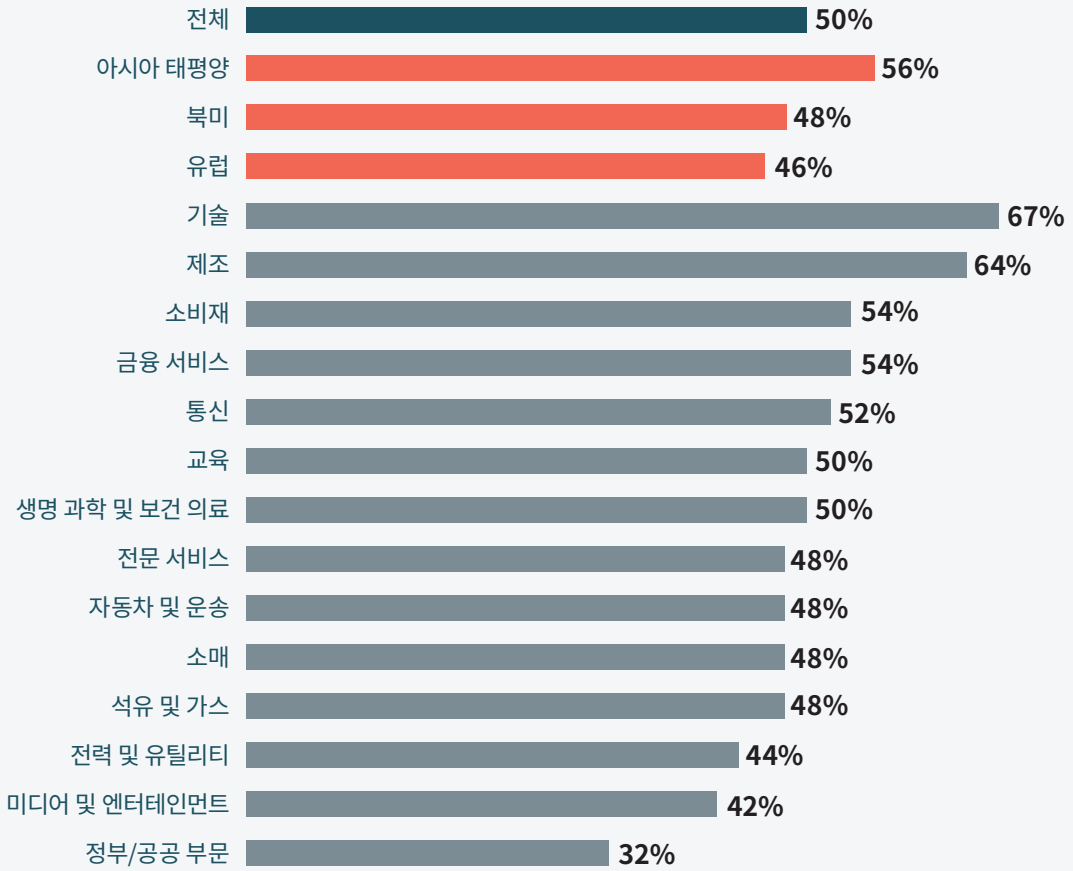
중요한 이점이 무엇이 있는지 물었습니다. 위시리스트에서 1위는 50%(기술, 유틸리티 및 정부 응답자 가운데 60% 이상)가 언급한 오픈 소스 표준과 오픈 데이터 형식이 차지했습니다. Hivory의 Andy McQuarrie는 그다지 놀랍지 않다는 반응을 보였습니다. “개방형 표준을 사용하면 성숙도를 상당히 쉽게 높일 수 있습니다. 단기간에 관리형 서비스의 서비스를 이용할 수 있죠. 성숙된 상태에서 구성 요소를 추가해야 하는 경우, 오픈 소스를 사용하면 전체 아키텍처에 영향을 주지 않으면서 해당 옵션을 추가할 수 있습니다. 과거에 했던 방식으로 기술을 변경할 필요가 없습니다.”

그다음 순위로는 새로운 기능이 아니라 인프라, 운영, 유지 관리 및 기타 아키텍처 요소에 대한 보안 강화, 거버넌스 강화, 가격/성능 향상 등 기술 리더가 부단히 개선하고자 노력하는 영역이 두드러지게 나타났습니다. 또한, 설문조사 응답자들은 새로운 아키텍처가 ML, 데이터 과학 또는 비즈니스 인텔리전스를 기반으로 하는 모든 분석 사용 사례를 지원한다고 주장합니다.

오픈 소스 표준 및 오픈 데이터 형식은 기술 리더들이 원하는 최우선순위이며, 아키텍처 요소에 대한 보안 및 거버넌스 강화, 가격/성능 향상 등이 그 뒤를 따릅니다.

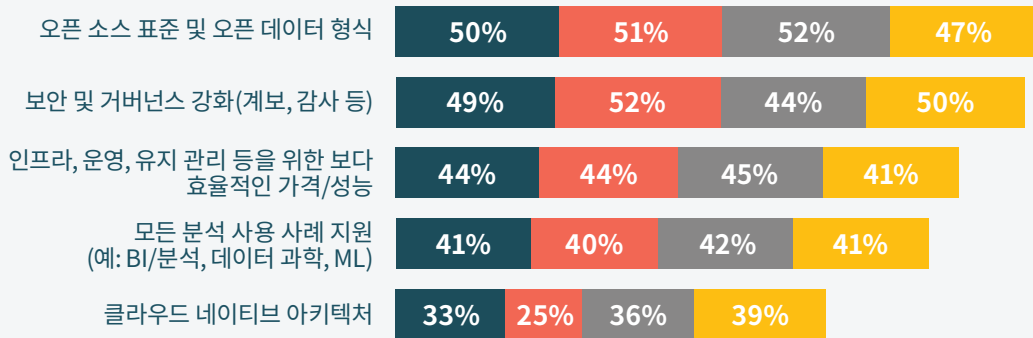


그림 8: 현재 당면 과제를 해결할 전사적 데이터 전략 또는 솔루션을 구현하기 위해 현재 새로운 데이터 플랫폼을 평가하거나 구현하는 응답자(응답자 %)



MIT 테크놀로지 리뷰 인사이트 설문 조사, 2021

그림 9: 기존 아키텍처 대비 응답자의 이상적인 새 아키텍처의 가장 중요한 장점(최고 응답, 응답자 %)



MIT 테크놀로지 리뷰 인사이트 설문 조사, 2021

■ 전체 ■ 북미 ■ 유럽 ■ 아시아 태평양

“우리는 고객 시스템과의 기본 통합에 정말로 집중하고 있습니다. 고객이 우리에게 데이터를 더 쉽게 전송할 수 있도록 하는 방법을 고민 중이죠.”

– Hivery CTO Andy McQuarrie

우리가 인터뷰한 CDO들은 데이터 관리 및 인프라가 발전함에 따라 새로운 희망과 계획을 꿈꾸었습니다. 주요 에너지 회사인 Total의 그룹 최고 데이터 책임자인 Michel Lutz는 메타데이터 관리의 ML 기반 자동화를 확장하는 것을 목표로 합니다. 이미 지구과학 데이터에 대해서는 수행되었지만, 내년에는 그룹의 모든 데이터 도메인으로 확장할 계획입니다. Lutz는 데이터 메시 아키텍처의 잠재력도 살펴볼 계획입니다. “조직 확장성을 높이고 도메인 전문화를 강화하기 때문에, 확실히 우리 데이터 아키텍처가 다음으로 나아갈 단계라고 생각합니다.”라고 그는 말합니다.

마찬가지로, L’Oréal의 Naveen Jayaraman은 데이터 관리 기능을 널리 배포하려는 기업을 위한 도메인 지향 데이터 제품의 이점을 살펴보고 있습니다. Jayaraman은 데이터 메시지를 사용하는 아키텍처 접근 방식은 다중 언어 저장소와 같은 일부 요소가 중앙 집중화되지만 한다면 그의 목적을 이루어줄 가능성이 있다고 말합니다.

Nielsen의 Mainak Mazumdar는 말하자면 AI 모델을 자동으로 선택할 수 있는 “AI를 위한 AI”를 구현하기를 기대합니다. “우리는 사용할 모델에 대해 최선의 결정을 내릴 수 있는 또 다른 인텔리전스 계층을 구축하고 싶습니다.”

Ashwin Sinha는 거시적인 관점에서 데이터 엔지니어링의 미래 혁신에 적응할 수 있을 만큼 기민한 “애자일 및 실시간 아키텍처”의 가능성에 대해 생각합니다. 적어도 중기적으로는 최근에 설계된 아키텍처를 미래에 대비할 수 있는 방법이 된다고 말합니다. “우리는 2~3년 안에는 구식이 되지 않을 만큼 충분히 유연한 아키텍처 접근 방식을 도입하고 싶습니다.”

06 결론

우 리 연구, 특히 CDO들과의 토론을 통해서 데이터 과학을 관통하는 데이터 인프라에서부터 조직의 모든 직원이 보유하고 있는 데이터에 대한 태도에 이르기까지, 직접적인 라인이 형성되어 있으며 이 모든 것을 비즈니스 효과와 다시 연결하는 것, 즉 데이터 문화에 대해 인정하게 되었습니다. 이 연구는 자산을 사용하여 비즈니스에서 데이터 문화를 키우려는 CDO들에게 몇 가지 교훈을 제공합니다.

- **단순성과 유연성을 추구하십시오.** 지나치게 복잡한 아키텍처를 단순화하기를 원하는 CDO가 많지만, 분석 최종 사용자를 위한 인터페이스를 사용하기 쉽고 최대한 흥미롭게 구현하는 것도 중요합니다. 또한 아키텍처는 새로운 기술을 추가하기 위한 대규모 마이그레이션이나 재설계 없이도 미래의 비즈니스 요구 사항을 충족할 수 있을 만큼 충분히 유연해야 합니다. 더 많은 사용자가 자신이 선호하는 도구를 사용하여 지속해서 리프레시되는 데이터를 가져와서 사용하고 요청에 따른 인사이트를 얻을 수 있으므로 참여가 늘어날 것입니다.
- **잘 관리하십시오.** 데이터에서 오류가 반복적으로 발견되는 것은 그 무엇보다도 직원의 분석 경험에 악영향을 미칩니다. CDO들은 이미 건전한 데이터 거버넌스의 중요성을 잘 알고 있지만, 거버넌스 실패가 조직에 미치는 영향이 어느 정도인지 미처 깨닫지는 못할 수도 있습니다. 단일 정보 소스를 구축하고 유지하는 것은 모두에게 중요합니다.

- **설명하고 전파하십시오.** 일부 CDO는 팀원들이 사업부 및 기타 직원에게 데이터 과학에 대해 교육하고 분석 및 기타 도구 사용법을 훈련시키도록 합니다. 또한, 많은 CDO가 데이터와 분석을 사용하여 얻은 긍정적인 결과와 얻은 가치를 널리 알리면 직원들의 데이터와 분석 사용 욕구가 커질 수 있다는 것을 발견했습니다. 효과적인 거버넌스를 통해 달성한 데이터 정확성을 알리는 것도 신뢰를 키우는 데 도움이 됩니다.

- **중역들과 유대를 형성하십시오.** CDO 및 CAO에게는 동료 중역들과의 긴밀한 유대 관계가 당연하게 느껴지겠지만, CIO 및 CTO와 데이터 관리가 단절될 경우에도 불신의 원인이 될 수 있습니다. 데이터 전략은 기존 기술 인프라를 고려하고 항상 기업 우선순위를 실행하는 데 중점을 두어야 합니다.



MIT Technology Review Insights 소개

MIT Technology Review Insights는 세계 최고의 기술 기관의 지원을 받는 세계 최장기 발행 기술 잡지인 *MIT Technology Review*의 맞춤형 출판 부서로, 라이브 이벤트를 만들고 당대의 주요 기술 및 비즈니스 과제에 대한 연구를 진행합니다. 인사이트는 미국 및 해외에서 정성적, 정량적 연구 및 분석을 수행하고 기사, 보고서, 인포그래픽, 동영상, 팟캐스트 등 다양한 콘텐츠를 발행합니다. 또한 성장하는 MIT Technology Review 글로벌 패널을 통해, Insights는 전 세계의 고위 경영진, 혁신가 및 사상가 리더들과 설문 조사 및 심층 인터뷰를 하기 위한 최고의 네트워크를 보유하게 되었습니다.

후원 기업

Databricks는 데이터+AI 기업입니다. 최신 레이크하우스 아키텍처를 기반으로 구축된 Databricks는 최고의 데이터 웨어하우스와 데이터 레이크를 결합하여 모든 데이터 워크로드를 위한 단순하고 개방된 협업 플랫폼을 제공합니다. Shell, Comcast, CVS Health, HSBC, T-Mobile, Regeneron을 비롯한 전 세계 5,000개 이상의 조직에서 Databricks를 사용하여 대규모 데이터 엔지니어링, 탐색적 데이터 사이언스, 전체 수명 주기 머신 러닝 및 비즈니스 분석을 지원합니다. Microsoft, Amazon, Google, Tableau를 비롯한 수백 개의 파트너사와 글로벌 입지를 갖춘 Databricks는 데이터 팀이 세계에서 가장 어려운 문제들을 해결할 수 있도록 지원하는 사명을 가지고 있습니다.

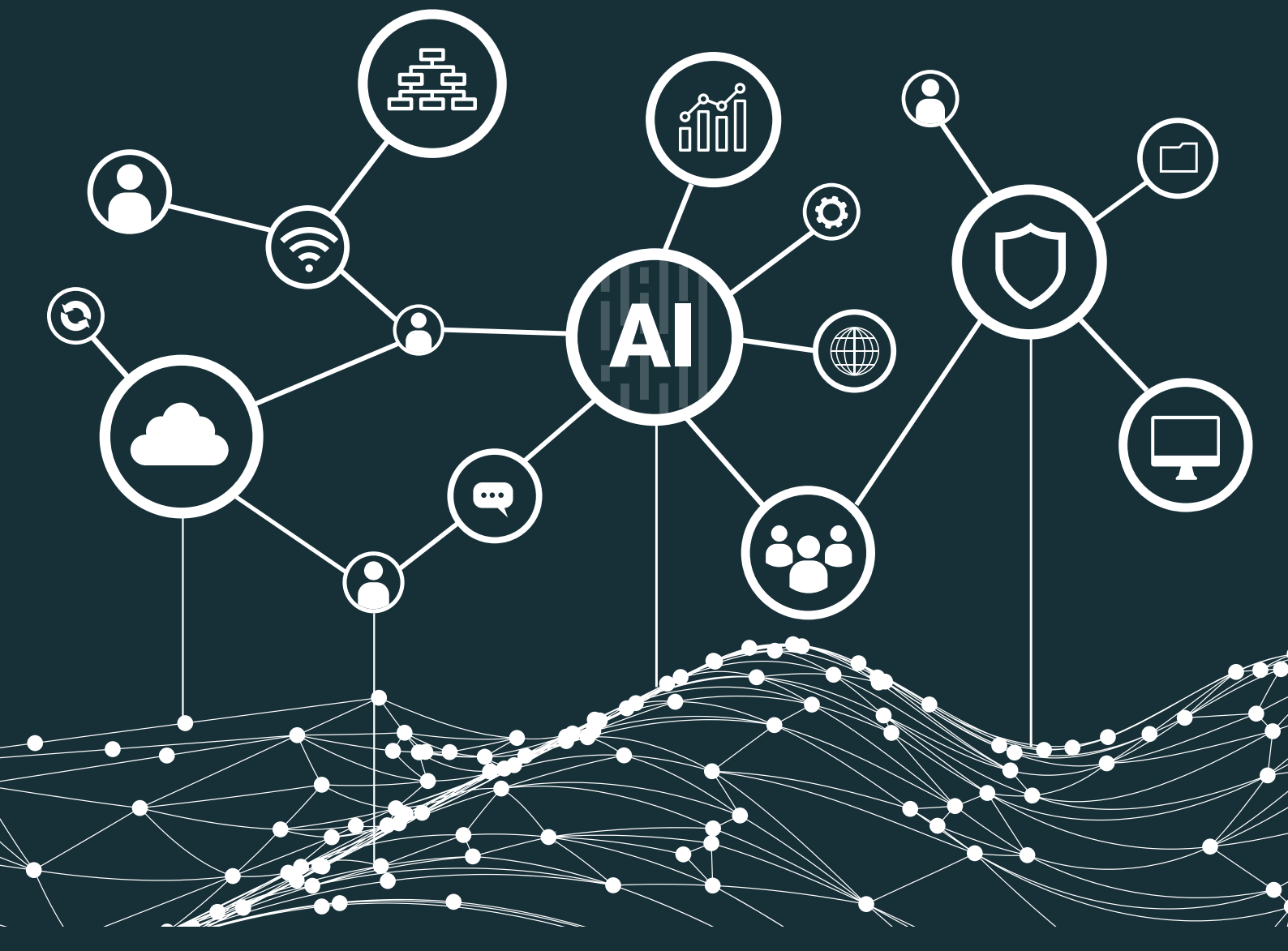


삽화

Davooda: 5, 9, 13, 14; fad82:13; Friday Studio: 2, 3, 6, 15, 17, 22; Nadiinko: 18; ninamalin:18, papipo:15; Prostock Studio: 표지 앞면, 뒷면, 2, 3, 4, 5, 13, 19, Rashes Ashor: 19; Starline: 표지 앞면, 뒷면, 2. 모든 삽화는 Shutterstock 제공, Scott Shultz Design 제작

정확한 정보를 제공하기 위해 모든 노력을 기울였지만, MIT Technology Review Insights는 이 보고서에 포함된 사람 또는 이 보고서에 명시된 정보, 의견 또는 결론에 대한 신뢰에 대해 어떠한 책임도 지지 않습니다.

© Copyright MIT Technology Review Insights, 2021. All rights reserved.



MIT Technology Review Insights

 www.technologyreview.com

 [@techreview@mit_insights](https://twitter.com/techreview@mit_insights)

 insights@technologyreview.com