

CIO 가이드

# 분석 및 기계 학습 성공을 위한 현대적 전략 수립



# 서문

**데이터와 분석**은 새로운 비즈니스 기회의 시대를 열었습니다. 클라우드 및 기계 학습을 기반으로 하는 현대적 분석은 비즈니스 전반의 가시성을 높여 큰 이익을 창출하는 의사 결정을 내릴 수 있도록 지원합니다.

경영진은 실시간 데이터 및 과거 데이터를 통해 수집한 인사이트를 활용하여 예상치 못한 변화에 신속하게 적응하고, 고객 경험을 개선하며, 예측력을 높이고, 프로세스를 간소화함으로써 운영 비용 및 시간을 절약할 수 있습니다.

이 eBook에서는 인공지능(AI) 및 기계 학습의 발전과 함께 클라우드 플랫폼 및 서비스로의 마이그레이션 증가가 어떻게 분석 이니셔티브를 가속화하고 데이터 가용성, 성능 및 확장성에 대한 기존의 어려움을 극복할 수 있도록 지원하는지 살펴봅니다.

## 데이터를 비즈니스 가치로 전환

다음은 조직에서 이미 비즈니스 이익과 공익을 위해 분석을 적용하고 있는 몇 가지 예입니다.

- ▶ **공급망 부문**은 시장 상황이나 고객 수요가 변동할 때 빠르게 전환할 수 있는 능력을 갖추고 있습니다. 예를 들어 파트너 네트워크에 중단이 발생하면 기계 학습 기반 분석이 네트워크 구성원에게 알림을 보내므로, 배송 일정이 순조롭게 진행될 수 있도록 대체 공급업체나 배송업체를 찾을 수 있습니다.
- ▶ **제조업체**는 고장난 장비를 탐지하여 조기에 시정 조치를 취함으로써 작업 현장의 가동 중단을 방지할 수 있습니다. 이러한 기능은 생산성과 비즈니스 연속성을 높임으로써 제조업체가 생산 및 배송을 개선하고 비용을 절감할 수 있도록 도와줍니다.
- ▶ **통신 회사** 및 기타 데이터 집약적 기업은 생성 및 수집한 데이터를 패키징하고 업종별 시장을 대상으로 판매함으로써 수익을 창출하고 있습니다. 이러한 방식으로 비즈니스를 위한 새로운 수익원을 창출하고 있습니다.

- ▶ **마케팅 부서**에서는 분석을 사용하여 디지털 발자국을 기반으로 소비자의 취향, 행동 및 경험을 파악합니다. 기업은 이 정보를 바탕으로 미시적으로 세분화된 그룹에 대한 제안을 정확하게 타겟팅하고 고객 기반을 확장할 수 있습니다. 또한 고객이 선호하는 의사 소통 및 거래 방법을 학습하여 비즈니스가 그에 따라 상호 작용하여 만족도를 높이고 충성도를 구축할 수 있도록 합니다.
- ▶ **과학 커뮤니티**는 위성, 우주 정거장 및 로봇 차량에서 수집한 방대한 양의 환경 데이터를 분석하고 공유하여 지구에서의 인간의 거주성을 개선하고 화성에서 거주를 실현할 수 있도록 하기 위한 조치에 대해 알려줍니다.

# 목차

- 3 **성공의 방해 요소**  
다양한 분석 기법의 올바른 조합은?
- 5 **장애 극복**  
기계 학습: 모두를 위한 기술
- 6 **현대적 분석 방식의 구현 방법**
- 8 **데이터의 활용**
- 8 **후원사 견해**



# 성공의 방해 요소

**데이터를 수익화**하면 잠재적으로 상당한 비즈니스 가치를 얻을 수 있습니다. 그러나 많은 조직에서 목표 달성에 필요한 정확하고 의미 있는 결과를 얻기 위해 집합적 데이터를 활용하는 데 어려움을 겪고 있습니다. 대표적인 몇 가지 문제는 다음과 같습니다.

▶ **데이터 증가.** 전 세계에서 생성되는 데이터의 양은 계속해서 증가하고 있습니다. IDC는 향후 5년 동안 생성되는 디지털 데이터의 양이 2020년 64제타바이트에서 2025년 180제타바이트로 증가할 것으로 예측하고 있습니다. 이는 디지털 스토리지의 등장 이후 생성된 모든 데이터의 두 배 이상입니다. 빠르게 증가하는 데이터의 양과 다양한 유형 때문에 오래된 온프레미스 인프라와 수작업 프로세스를 사용하여 데이터를 관리하는 것은 점점 더 어려워지고 있습니다.

*Fortium Partners의 파트너이자 HBO와 전미 농구 협회(NBA)에서 글로벌 CIO 직책을 맡고 있는 Michael Gabriel은 “데이터가 클라우드에 저장되어 있던 로컬에 저장되어 있던 상관없이 모든 데이터를 제대로 이해하는 회사는 거의 없습니다.”라고 말합니다. “따라서 사용 가능한 데이터가 무엇인지, 또는 사용할 수 있었던 데이터가 무엇인지를 파악하고, 어떻게 데이터를 활용하여 분석을 지원할 수 있는지를 아는 것은 쉽지 않습니다.”*

▶ **오래된 데이터 인프라.** 3계층 구조의 온프레미스 데이터 인프라는 증가하는 데이터를 관리하는 데 필요한 확장성과 성능이 부족하며, 지속적인 구성, 관리 및 용량 계획이 필요합니다. 또한 인프라가 노후화되면 전사적으로 실행 가능한 인사이트를 제공하는 데 중요한 기계 학습 및 기타 고급 분석 기능을 지원하기 어렵습니다. 조직에는 요구 사항의 변화와 데이터 볼륨 증가에 따라 크기를 조정할 수 있는 데이터 시스템이 필요합니다.

*AWS 분석 담당 부사장인 Rahul Pathak은 “처리할 수 있는 용량의 한계에 도달한 온프레미스 분석 시스템 보유 고객을 많이 보았습니다.”라고 말합니다. “이러한*

*고객의 데이터 웨어하우스는 유입되는 데이터의 증가와 규모를 감당하지 못하며, 따라서 자체 데이터 센터에서 이를 유지 관리하는 데 드는 규모, 성능 및 운영 비용 문제로 어려움을 겪고 있습니다.”*

▶ **데이터 사일로.** 회사 데이터는 부서별로 구분되어 있으며 일부 사용자 그룹만 액세스할 수 있습니다. 또한 데이터는 데이터 웨어하우스, 데이터 레이크 및 데이터베이스를 포함한 여러 위치에 저장됩니다. 이러한 한계로 인해 불완전한 데이터에 의존하는 분석 프로젝트는 설계자가 기대하는 인사이트와 결과를 창출하지 못하는 경우가 많습니다. 데이터로부터 새로운 가치를 얻기 위해서는 사일로를 제거하여 데이터 저장 위치에 상관없이 모든 관련 데이터를 결합하고 분석할 수 있는 기능이 필요합니다.

*Gabriel은 “공유하지 않으면 데이터와 비용이 중복되고 전체 상황을 보지 못할 수 있습니다.”라고 말합니다. “이는 개별 부서뿐만 아니라 회사에도 불이익을 초래합니다.”*

▶ **데이터 중력.** 데이터가 계속 증가함에 따라 데이터를 이동하기가 더 어려워지고 있습니다. 이 데이터 ‘중력’은 특정 영역에 대해 실행하는 분석에서 포괄성을 떨어뜨리며, 이를 기반으로 하는 인사이트 및 의사 결정의 안정성을 저하시킬 수 있습니다.

*“신속하고 민첩하게 의사 결정을 내리려면 고객이 필요한 데이터를 안전하고 통제된 방식으로 데이터 스토어 간에 쉽게 이동함으로써 데이터 중력을 인식할 수 있어야 합니다.”라고 Pathak은 말합니다.*

▶ **일관성 없는 데이터 거버넌스.** 일부 조직은 오랜 시간에 걸쳐 여러 거버넌스 정책을 결합했습니다. 그 결과 데이터 업데이트 누락, 보안 취약성 및 규정 미준수 등의 문제가 발생할 수 있는 부문별로 상이한 액세스 제어 정책을 초래했습니다. 다른 조직들은 제한이 많은 거버넌스 정책으로 과도하게 적용하여 생산성을 떨어뜨리고 있습니다.

“우리가 목격하는 흔한 실수는 조직이 지나치게 경직된 거버넌스 시나리오를 생성하는 것입니다. 이러한 조직에서는 중앙 팀이 비즈니스 개선을 위해 데이터 작업을 수행하는 과정에서 병목 현상이 발생합니다.”라고 Pathak은 말합니다. “훌륭한 거버넌스의 핵심은 액세스를 정의하는 방법을 파악하여 문제를 해결하는 것입니다. 즉, 데이터가 필요할 때마다 중앙 조직의 다른 사람에게 요청해야 하는 방식이 아니라 예외 프로세스를 생성한다는 의미입니다.”



분석 및 기계 학습 전문가 위한 현대적 전략 수립

## 다양한 분석 기법의 올바른 조합은?

데이터 분석에는 많은 종류가 있으며, 기업은 다양한 성과를 달성하기 위해 이러한 데이터 분석을 적절히 혼합하여 적용하고자 할 것입니다. 기본 분석 유형 및 가장 일반적인 사용 사례는 다음과 같습니다.

- 실시간 분석**은 수집되는 데이터를 즉시 인사이트로 변환합니다. 이러한 유형의 분석은 온라인 거래 또는 차량 제어 시스템과 같이 시간에 매우 민감한 애플리케이션에 사용됩니다. 실시간 분석은 장비에 장애가 발생하는 시기를 예측하고, 자율 주행 차량이 사고를 피할 수 있도록 도와주며, 거래가 완료되기 전에 신용 카드 사기를 감지할 수 있습니다. 실시간 분석에는 크게 두 가지 유형이 있습니다. 온디맨드 분석은 사용자 또는 애플리케이션으로부터 쿼리를 수신한 후에 결과를 제공하는 반면, 스트리밍 분석은 알림 또는 결과를 지속적으로 제공합니다.
- 로그 분석**은 운영 분석이라고도 하며 컴퓨터, 네트워크, 애플리케이션, 운영 체제 또는 다른 IT 구성 요소에서 캡처될 수 있는 이벤트 데이터를 평가하는 작업입니다. 조직은 로그 분석을 사용하여 사용자 행동의 패턴을 파악하고, 문제 지점을 식별하고, 보안 활동을 감사하고, 규정 준수를 관리하며, 용량 또는 기타 IT 인프라 변경을 계획할 수 있습니다.
- 빅 데이터 분석**은 다양한 소스와 크기의 정형, 반정형 및 비정형 데이터를 포함하는 매우 크고 다양한 데이터 집합에 대한 고급 분석을 실행하는 것과 관련이 있습니다. 데이터는 센서, 컴퓨팅 또는 통신 디바이스, 비디오/오디오, 네트워크, 로그 파일, 트랜잭션 애플리케이션, 웹 콘텐츠 및 소셜 미디어 등에서 생성됩니다.
- 데이터 웨어하우징 분석**은 애플리케이션 로그 파일 및 트랜잭션 애플리케이션과 같은 다양한 소스에서 수집된 대량의 과거 데이터에 대해 쿼리를 수행합니다. 사용자는 이 유형의 분석을 활용하여 주제에 따라 쿼리를 실행하고 시간에 따른 변경 사항을 평가할 수 있습니다.
- 기계 학습**은 데이터의 패턴을 분석하고 해석하여 인간의 상호 작용 없이 학습과 의사 결정을 가능하게 합니다. 이를 통해 완전히 새로운 수익 기회를 창출하고, 더 나은 의사 결정을 더 빠르게 내리며, 운영 효율성을 개선할 수 있습니다.

## 장애 극복

**분석을 성공적으로 수행**하기 위해서는 예상 성과에 맞춘 명확한 비즈니스 조정과 이러한 목표를 방해하지 않으면서 제대로 지원하는 IT 인프라가 필요합니다.

Gabriel은 “어떤 목표를 달성하고자 하는지, 현재 인프라의 어느 부분에 개선이 필요한지를 알기 전에는 인프라를 제대로 파악하기가 어렵습니다.”라고 말합니다.

분석 현대화 전략에는 2가지 기본 구성 요소가 있습니다.

1) 데이터를 통합하여 인증된 사용자가 보편적으로 액세스할 수 있어야 하며, 2) 인프라가 데이터 관리, 보안 및 거버넌스에 대한 중앙 집중식 방식을 지원해야 합니다.

조직은 데이터 및 집계를 중앙에서 관리함으로써 현명한 의사 결정과 최적의 비즈니스 성과에 필요한 모든 데이터에 대해 분석을 실행할 수 있습니다.



## 기계 학습: 모두를 위한 기술



### AI 서비스

애플리케이션 및 비즈니스 프로세스에 AI가 도입되는 경우가 점차 늘어나고 있습니다. 즉시 사용할 수 있도록 제공되는 AI 서비스 활용 인텔리전스는 맞춤형 추천, 고객 센터 인텔리전스, 문서 처리, 지능형 검색, 비즈니스 지표 분석 등과 같은 일반 사용 사례를 해결합니다.



### 기계 학습 인프라 및 프레임워크

실무 전문가는 TensorFlow, MXNet, PyTorch 등 원하는 기계 학습 프레임워크를 실행할 수 있는 매우 강력한 인프라를 원합니다.



### 기계 학습 서비스

데이터 사이언티스트, 기계 학습 개발자, 그리고 점점 더 많은 데이터 분석가들이 기계 학습을 더 쉽고 빠르게 구축하고 구현할 수 있는 도구를 찾고 있습니다. 소프트웨어용 통합 개발 환경과 마찬가지로 레이블 지정, 데이터 준비, 특성 추출, 통계 편향 감지, 자동 기계 학습, 훈련, 튜닝, 호스트, 설명 가능성, 모니터링 및 워크플로를 포함한 모든 기계 학습 개발 단계를 포괄하는 서비스가 필요합니다.

# 현대적 분석 방식의 구현 방법

현대적 분석 방식에서 얻을 수 있는 **인사이트와 혁신을 활용**하기 위해 고려해야 하는 4가지 기본 단계는 다음과 같습니다.

## 1. 데이터 레이크를 사용하여 데이터 집계

사일로화된 데이터를 통합하는 중요한 단계는 하나 이상의 데이터 레이크로 통합하는 것입니다. 저장, 관리 및 분석할 데이터의 양이 많을수록 무한 확장 가능한 처리 리소스가 있는 퍼블릭 클라우드 기반에 데이터 레이크를 만드는 것이 좋습니다.

데이터 레이크에서는 통합 데이터 집합을 계속해서 사용할 수 있으므로 비용이 절감되고 데이터 가치가 극대화됩니다. 이러한 전체적인 방식을 사용하는 분석 프로그램은 최적의 결과를 얻기 위한 인사이트를 수집하고 반환할 때 전체 데이터를 파악할 수 있도록 지원합니다.

Mastercard의 자회사인 NuData는 **종합적이고 총체적인 데이터 분석의 이점**을 입증하는 사례입니다. 이 회사는 기계 학습을 기반으로 Amazon S3 데이터 레이크에 저장하는 대용량 데이터 집합에 대해 전문 분석을 실행하는 사기 탐지 서비스를 구축했습니다. NuData는 26개의 마이크로서비스를 운영하고 있으며, 각 마이크로서비스는 특정 유형의 사기를 탐지하기 위한 고객의 사용 사례를 기반으로 합니다. 각 마이크로서비스는 자체 데이터 레이크를 가지고 있지만 다른 마이크로서비스와 데이터를 공유할 수 있습니다. NuData 소프트웨어 엔지니어링 담당 이사인 Justine Fox에 따르면 회사는 이러한 기능을 사용하여 매일 페타바이트 규모의 데이터를 상호 연관시킴으로써 사용자 계정을 손상시킬 수 있는 악의적이거나 잘못된 로그인 시도를 식별할 수 있다고 합니다.

## 2. 통합 데이터 거버넌스 및 액세스 활성화

조직 내에서 데이터에 액세스해야 하는 모든 직원은 필요할 경우 언제 어디서나 데이터에 액세스할 수 있어야 합니다. 이를 위해서는 데이터 레이크와 목적별 데이터 스토어 및 서비스에 안전하고 원활하게 액세스할 수 있어야 하며, 데이터가 어디에 있든 상관없이 데이터에 액세스할 수 있어야 합니다. 또한 데이터는 애플리케이션, 시스템 및 서비스 간에 원활한 이동이 가능해야 합니다. 예를 들어, 웹 애플리케이션의 클릭스트림 데이터를 데이터 레이크에서 직접 수집할 수 있으며, 일일 보고를 위해 해당 데이터의 일부를 데이터 웨어하우스로 옮길 수도 있습니다. 또는 영업 데이터를 웨어하우스에서 데이터 레이크로 옮겨, 기계 학습을 사용하여 저장하고 분석할 수 있습니다.

데이터 통합과 더불어 중앙 집중식 데이터 거버넌스 및 보안이 필요합니다. 통합 데이터 거버넌스에는 데이터, 서비스 및 애플리케이션 전반에 걸쳐 일관된 정책을 설정하는 작업이 포함됩니다. 데이터 보안과 작업자 생산성 사이에서 적절한 균형을 유지하는 것이 중요합니다. 이때 고려할 단계는 다음과 같습니다.

▶ **데이터 리소스의 전사적 인벤토리**를 확보하고, 데이터 리소스의 담당자와 액세스가 필요한 사용자를 파악합니다. 누가 어떤 리소스에 액세스할 수 있는지를 설정하기 위한 맵을 생성합니다. "클라우드에서 이 작업을 수행하는 이점 중 하나는 바로 이러한 모든 상호 작용을 추적할 수 있다는 것입니다." 라고 Pathak은 말합니다.

▶ 액세스 제어를 위해 필요한 사용자와 애플리케이션으로만 액세스를 제한하는 **제로 트러스트 보안 방식**을 채택합니다. 보안상의 이유로 액세스를 엄격하게 제어하면서 여러 사일로의 데이터를 통합하려면 섬세한 데이터 관리 균형이 필요합니다.



통합 데이터 거버넌스에는 데이터, 서비스 및 애플리케이션 전반에 걸쳐 일관된 정책을 설정하는 작업이 포함됩니다. 데이터 보안과 작업자 생산성 사이에서 적절한 균형을 유지하는 것이 중요합니다.

## 2. 계속

▶ **데이터 카탈로그를 배포**하거나 데이터를 자동으로 검색, 태그 지정 및 카탈로그화하는 기타 중앙 집중식 관리 메커니즘을 배포하여 정책을 한 곳에서 관리하고 감사할 수 있습니다. 이를 통해 적시에 올바른 사용자에게 데이터에 대한 세분화된 액세스를 제공하고, 규제 거버넌스 및 규정 준수 요구 사항을 효과적으로 충족할 수 있습니다.

▶ **클라우드 서비스 제공업체(CSP)와 협력**하여 여러 지역에 걸쳐 규정 준수를 관리할 수 있습니다. 특히, 클라우드는 이론적으로 어디에나 위치할 수 있는 가상 머신을 사용하므로 데이터가 실제로 상주하는 위치를 CSP가 제어할 수 있는지 확인합니다. 규정 준수 데이터베이스를 생성하고 유지 관리하면 도움이 되며, 국가별 디지털 규정 준수 표준을 준비하면 명확하고 능동적인 규정 준수 구조를 구축할 수 있습니다.

## 3. 최상의 가격 대비 성능을 위한 목적별 데이터 및 분석 서비스 배포

분석 및 상관 관계 파악이 필요한 데이터 포인트의 양이 급증함으로 인해 많은 기업이 더 많은 데이터와 분석 인프라를 클라우드로 마이그레이션하고 있습니다. 클라우드에는 대량의 데이터를 분석하고, 의미 있고 실행 가능한 인사이트를 제공하며, 정확한 기계 학습 기반 자동화에 필요한 풍부한 훈련 데이터를 제공하는 데 필요한 무한 확장 가능한 컴퓨팅 및 스토리지 리소스가 있습니다.

많은 조직이 각 사용 사례에 최적화된 방식으로 데이터를 저장하거나 처리함으로써 분석 사용 사례를 보다 효과적으로 해결하기 위해 목적별 데이터베이스, 분석 및 기계 학습 서비스를 사용하고 있습니다. 예를 들어, 문서 데이터베이스는 뛰어난 확장성과 성능이 필요한 모바일 애플리케이션에 적합한 반면, 그래프 데이터베이스는 개발자가 밀접하게 관련된 레코드 간의 숨겨진 연결을

탐색하는 데 도움이 될 수 있습니다. 따라서 모든 워크로드에 최고의 기능을 사용할 수 있으므로 성능, 확장성 또는 비용을 모두 충족할 수 있습니다. 조직은 목적별 데이터 서비스를 사용하여 모든 애플리케이션 및 분석 요구 사항에 맞는 최상의 가격 대비 성능을 얻을 수 있습니다.

## 4. 기계 학습 및 AI를 사용하여 비즈니스 과제 해결

조직의 목표가 고객 경험 개선, 생산성 향상, 비즈니스 프로세스 최적화, 혁신 가속화 및 확장 등 어느 것이든 상관없이 기계 학습 및 AI 서비스를 활용하여 비즈니스 요구 사항을 충족할 수 있습니다.

조직은 AI 및 기계 학습 기술을 활용하여 이전에는 거의 사용할 수 없었던 데이터 집합을 통해 더 많은 작업을 수행할 수 있습니다. 예를 들어 PDF, 오디오, 비디오, 수익 기록 및 보고서와 같은 콘텐츠에서 볼 수 있는 비정형 데이터를 이제 기계 학습 프로세스를 통해 실행하여 새로운 인사이트를 얻을 수 있습니다.

혁신과 성장을 위한 기본 요소로 **클라우드, 데이터, AI 및 기계 학습을 수용**한 Nasdaq의 투자 인텔리전스 수석 부사장 겸 AI 및 기술 부문 책임자인 Michael O'Rourke는 "분석가들이 수십만 개의 문서를 읽는 대신 기계 학습을 사용하여 문서를 살펴보고, 정형 데이터를 생성하고, 이를 토대로 애플리케이션을 구축할 수 있습니다."라고 말합니다. AI와 기계 학습은 Nasdaq의 데이터 비즈니스뿐만 아니라 조직 전체에서 점점 더 중요한 역할을 하고 있습니다.

"금융 업계에서 AI의 기회는 엄청납니다."라고 O'Rourke는 말합니다. "Nasdaq 내 모든 비즈니스 라인은 기계 학습과 AI를 활용하여 더 나은 제품을 만들고 생산성을 개선하며 새로운 솔루션을 만드는 방법을 모색하고 있습니다."



## 데이터의 활용

**오늘날의 디지털 환경에서** 데이터는 모든 규모의 조직에서 감당하기 어려운 부담이 될 수 있습니다. 현대적 분석 방식을 사용하면 조직 전체가 데이터에 액세스할 수 있으며, 데이터를 활용하고 비즈니스를 혁신할 수 있는 새로운 방법을 모색할 수 있습니다.

## 후원사 견해

AWS의 목적별 서비스 포트폴리오를 기반으로 하는 데이터 전략에 대한 현대적 접근 방식을 통해 조직이 미래를 계획하고, 업계를 주도하며, 무한한 혁신 가능성을 실현하도록 지원할 수 있습니다. 분석을 현대화하면 모든 데이터에서 신속하게 인사이트를 얻고 모든 사용자에게 액세스를 제공할 수 있습니다.

AWS는 현대적 데이터 및 분석 전략을 실현하기 위해 매우 광범위하고 심층적인 목적별 분석 서비스 포트폴리오를 제공합니다. 이러한 포트폴리오에는

매우 확장성이 뛰어난 데이터 레이크, 매우 광범위한 목적별 분석 및 기계 학습 서비스 세트, 원활한 데이터 액세스 및 통합 거버넌스가 포함되어 있으며, 이 모든 것이 저렴한 비용으로 최고의 성능을 제공합니다. 이미 수만 개의 조직이 AWS를 사용하여 데이터 레이크를 구축했으며, 지금 바로 간단히 시작할 수 있습니다.

**▶ 자세히 알아보기**