



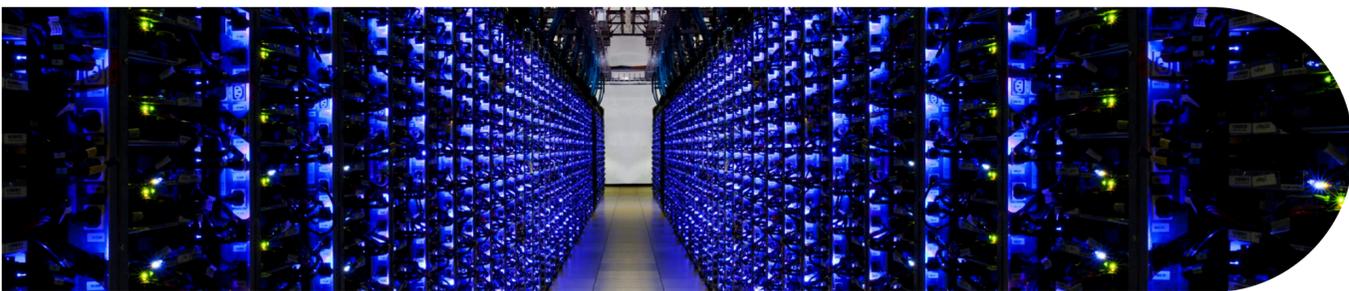
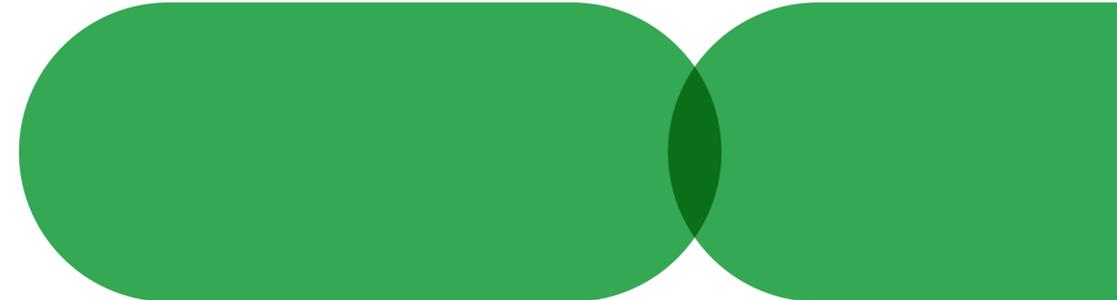
# 데이터를



# 자산으로



# 바꾸기



AI 시대에 선제적 데이터  
거버넌스를 구축하는 방법



# 목차

## 03 소개

## 04 AI 시대에 데이터 거버넌스가 중요한 이유

선제적인 데이터 거버넌스를 달성하기 위한 세 가지 방법

## 07 데이터 품질에 우선하여 ROI 증대하기

데이터에 대한 신뢰 구축

데이터 기반 의사 결정 강화

AI 성능 개선

다양한 업계에서 데이터 기반 의사 결정을 강화하는 방법

다음 단계

## 14 기술 전문가가 아닌 사용자도 데이터 기반 의사 결정권자가 되도록 지원

인사이트 발굴

안전하게 셀프 서비스 자동화

AI 성능 개선

다양한 업계에서 데이터 기반 의사 결정을 강화하는 방법

다음 단계

## 20 아키텍처를 뛰어넘기

규정 관련 업데이트에 손쉽게 적응

여러 아키텍처에 걸쳐 규정 준수 달성

메타데이터 관점으로 데이터 통합

자동화로 성과를 내는 방법

다음 단계

## 27 다크 데이터, 중복 데이터, 더티 데이터를 제거하여 ROI 극대화하기

## 32 미주



# 소개

## 데이터 가치를 극대화하기 위한 데이터 거버넌스 전략 가이드

데이터는 골칫거리가 될 수 있습니다. 데이터를 저장하는 데는 비용이 듭니다. 보호하는 데에도 비용이 듭니다. 정보 유출로 인한 금전적 피해가 평균 **435만 달러**로 추산되는 가운데, 정보 유출을 통제하지 못할 경우 손실 비용은 더욱 커집니다.

데이터는 조직의 규모나 업종에 관계없이 조직의 가장 강력한 자산이기도 합니다. 데이터는 인사이트를 발견하고, 우수한 고객 경험을 제공하고, 혁신을 이끄는 데 있어 핵심적인 요소입니다.

데이터 거버넌스는 수집에서 사용, 폐기에 이르는 데이터 수명 주기 동안 데이터 관리에 사용되는 원칙적인 접근법입니다. **견고한 데이터 거버넌스 프레임워크는 데이터 자산의 가치를 끌어냅니다.**

서로 다른 아키텍처에 걸쳐 데이터를 통합하는 프로세스부터 직원들이 필요할 때 정보에 액세스할 수 있게 해 주는 정책에 이르기까지, 데이터 거버넌스는 조직이 데이터에서 장기적인 ROI를 창출하고 AI 혁신을 구현할 수 있게 해 줍니다.

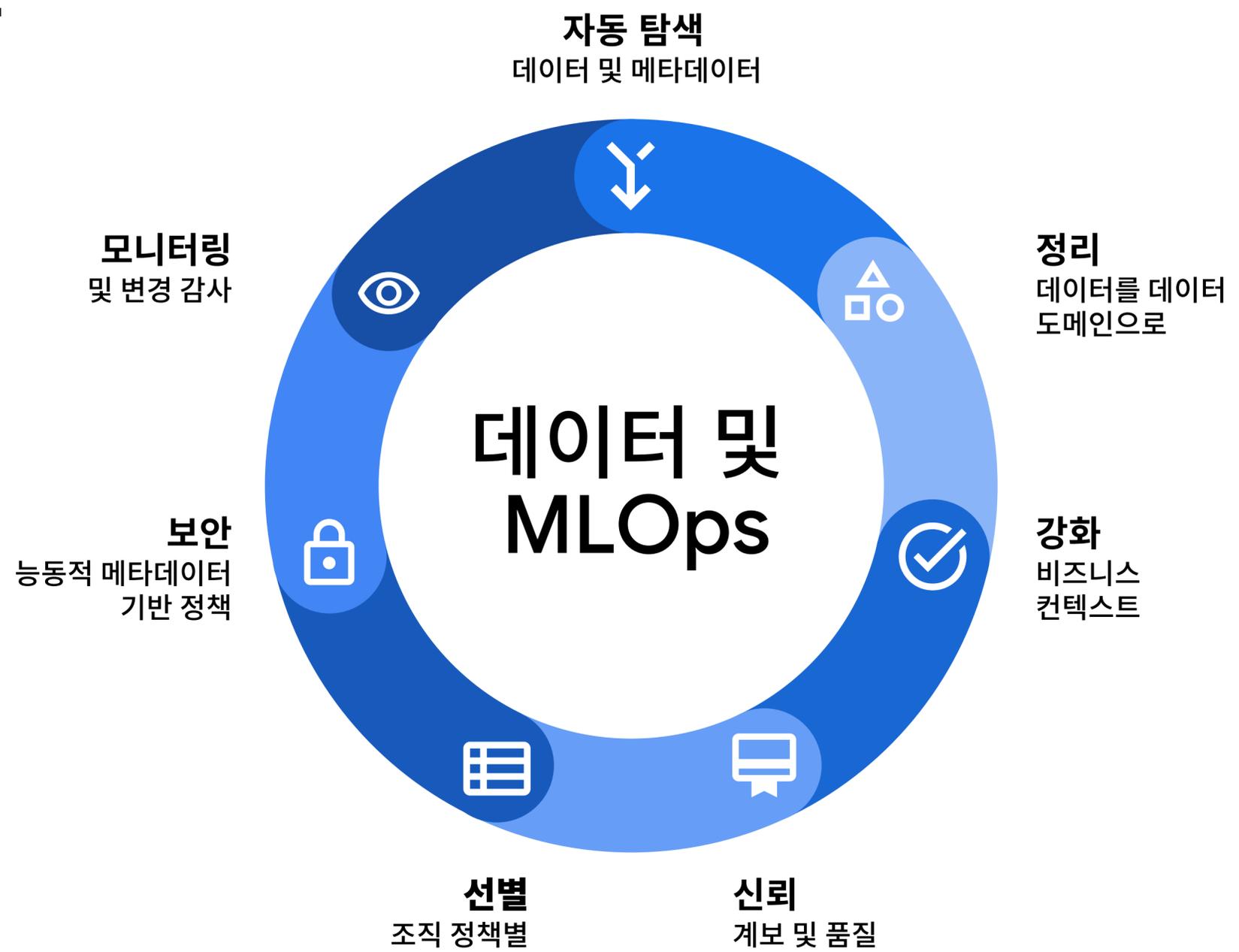
조직에서 데이터 자산을 극대화하기 위해서는 데이터 인텔리전스를 바탕으로 하는 선제적이고 자동화된 데이터 거버넌스가 필요합니다.

# AI 시대에 데이터 거버넌스가 중요한 이유

조직은 그 어느 때보다 많은 데이터를 보유하고 있습니다. 그리고 데이터에서 가치 있는 인사이트를 도출하는 동시에 데이터를 보호하고 규정을 준수하는 일은 갈수록 복잡해지고 있습니다.

데이터 거버넌스는 기업 조직 내의 효과적인 데이터 관리를 위한 일련의 프로세스 및 정책입니다. 또한 AI 기반의 자동화된 데이터 거버넌스는 오늘날의 데이터 품질, 민주화 규정 준수 문제를 대규모로 해결하기 위한 핵심 요소입니다. 이를 활용하지 않는 조직은 곧 뒤처지게 될 것입니다.

[데이터 거버넌스에 대해 자세히 알아보기 →](#)



2030년까지 AI를 완전히  
도입한 기업의 현금 흐름은  
두 배로 높아지는 반면 그렇지  
않은 기업의 현금 흐름은 20%  
줄어 들 수 있습니다.<sup>2</sup>



조직은 데이터를 민주화하는 동시에 분야별 거버넌스 및 기업 수준의 거버넌스를 실현해야 합니다. 데이터 인텔리전스를 내장한 인텔리전스를 바탕으로 하는 도구를 통해 조직은 데이터에 대한 엔드 투 엔드 가시성과 관리 역량을 확보할 수 있습니다. 또한 자동화된 도구와 정책은 조직 전체의 직원들이 관련된 인사이트에 액세스할 수 있도록 하는 동시에 빠르게 변화하는 규정을 준수할 수 있게 해 줍니다.

데이터 거버넌스라고 하면 일반적으로 데이터 보호를 떠올리지만, 조직이 미래 역량을 기를 수 있게 해 주는 핵심 요소이기도 합니다. 양질의 데이터는 비즈니스 인텔리전스의 동력이 되고, 생성형 AI 및 대규모 언어 모델(LLM)과 같은 기술에 정보를 제공합니다. 선제적인 데이터 거버넌스 관행은 데이터를 비즈니스의 근간으로 삼고 혁신을 촉진합니다.

이 가이드에서는 자동화된 원칙, 관행, 기술을 사용하여 데이터의 수명 주기 전반을 지원하는 미래 지향적인 데이터 거버넌스 전략을 구축하는 방법을 소개합니다.

시작하는 방법은 세 가지입니다.

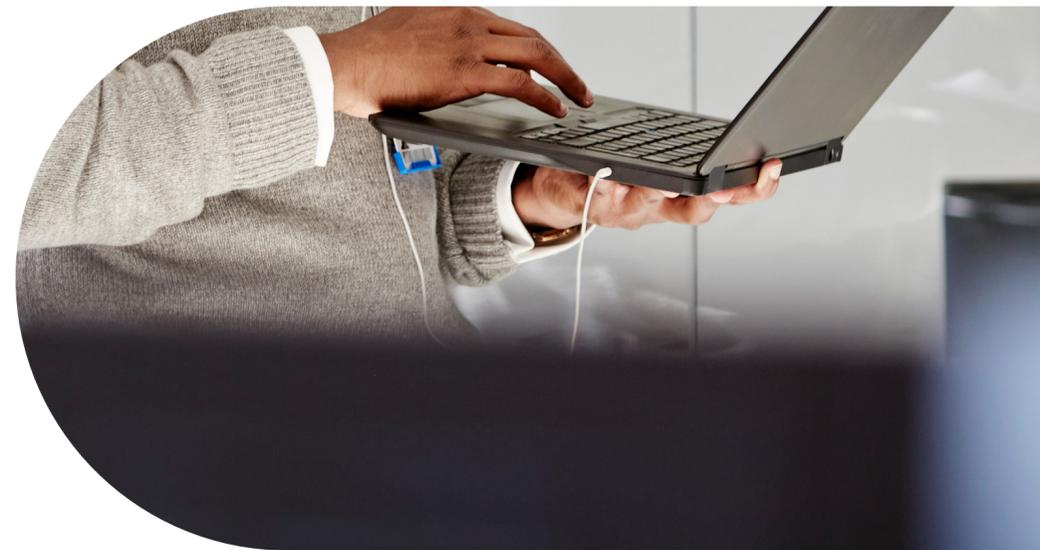
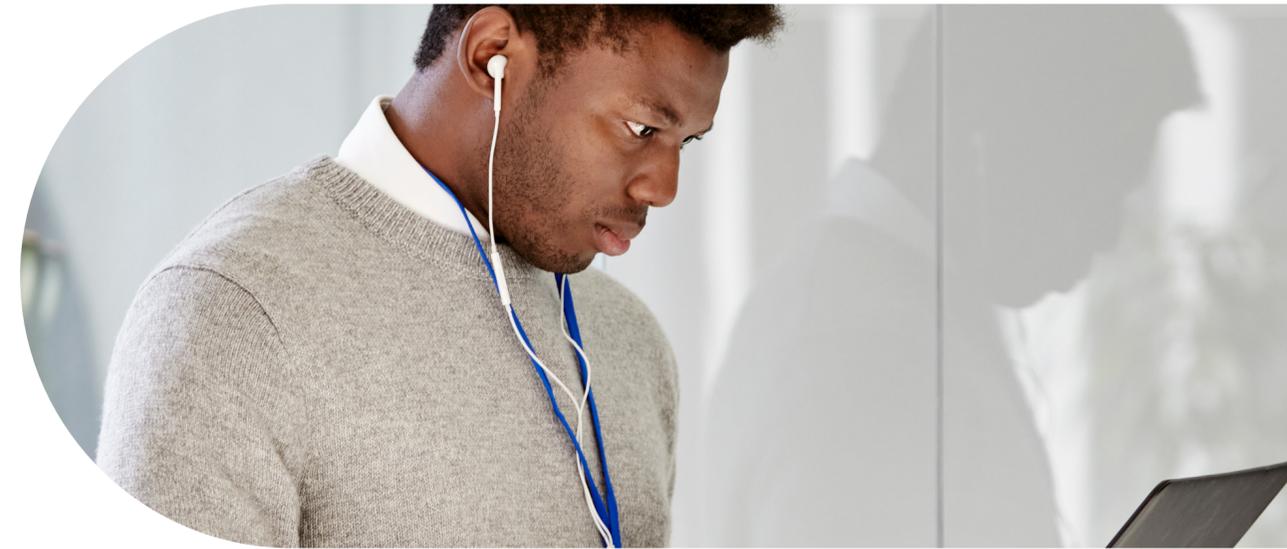
## 데이터를 대규모로 관리

선제적인 데이터 거버넌스 전략을 갖춘 조직은 다음을 달성할 수 있습니다.

- ✓ 데이터 품질 보장
- ✓ 신뢰할 수 있는 데이터를 손쉽게 발견
- ✓ 데이터 출처와 계보에 대한 인사이트 제공
- ✓ 내부와 외부에서 데이터 공유를 안전하게 관리
- ✓ 규정 요구사항 준수

선제적인 데이터 거버넌스를 달성하기 위한 세 가지 방법

# 1 데이터 품질에 우선하여 ROI 증대하기





# 데이터의 가치는 이를 잘 활용할 때 얻을 수 있습니다.

조직의 데이터로 ROI를 달성하려면 실시간 의사 결정을 내리고, 효과적인 AI 모델을 구동하고, 더 나은 고객 경험을 위한 기회를 발견할 수 있게 해 주는 신뢰할 수 있는 고품질 데이터를 확보해야 합니다.

그러면 다음과 같은 결과로 이어집니다.

- 데이터에 대한 신뢰 구축
- 데이터 기반 의사 결정 강화
- AI 성능 개선

2025년 기준

전 세계 데이터 생성 규모는  
180제타바이트에  
이를 전망입니다.<sup>3</sup>

(1제타바이트 = 1조 기가바이트)



## 데이터에 대한 신뢰 구축

여러 사일로에 분산된 데이터의 양이 증가하는 상황에서 조직이 데이터에 대한 엔드 투 엔드 뷰를 확보하기란 어려울 수 있습니다.

**수집하고 분석하는 데이터가  
많을수록 보안과 거버넌스의  
중요성도 더 커집니다.**

그러나 조직은 데이터 품질을 종합적으로 이해할 때 데이터를 신뢰할 수 있게 됩니다. 이에 따라 사용자는 자신이 찾던 데이터를 발견했을 때 그 데이터가 정확하고 완전한 최신 데이터임을 확신할 수 있습니다.

데이터 관측 가능성 도구는 시스템에서 데이터가 어떻게 이동하는지에 대한 자동화된 모니터링(데이터 계보)을 통해 데이터 이상치를 감지하고 복구할 수 있습니다. 데이터에서 급격한 상승 및 하락이 발생하는 경우 이러한 도구를 사용하면 데이터가 완전한지, 또는 누락된 레코드가 있는지 여부를 확인할 수 있습니다. 또한 데이터 문제가 발생하면 생산자나 데이터 엔지니어가 즉시 알림을 받을 수 있습니다.

## 데이터 기반 의사 결정 강화

데이터 품질은 조직이 가치 있는 사용자 경험을 제공할 수 있게 해 주는 비즈니스 인텔리전스(BI)의 기반입니다.

비즈니스 인텔리전스를 통해 영업, 고객 서비스, 생산, 보안을 비롯하여 조직의 거의 모든 영역에서 나타나는 중요한 패턴과 트렌드를 발견할 수 있습니다.

조직은 실시간 분석을 통해 실행 가능한 인사이트를 찾아내고 장단기 성과 개선 및 조직 전반의 ROI를 높일 수 있습니다.

소매업에서 데이터는 실시간 타겟팅 추천을  
통해 맞춤형 온라인 쇼핑 경험을 제공합니다.  
또한 애플리케이션에 대해 더 빠른 액세스를  
제공하여 온라인 거래의 지연을 줄입니다.



## AI 성능 개선

AI 개발에 대한 신뢰(또는 신뢰의 부족)는 도입의 주된 장벽입니다. [더 신뢰할 수 있는 AI 모델](#) 구축을 위해 고려해야 할 사항은 많지만 그 기반이 되는 것은 데이터 품질입니다. 예를 들어 학습 데이터에 불공정한 편향이 포함된 경우 모델에도 이러한 편향이 반영됩니다. 개인에게 큰 영향을 미치는 의사 결정을 AI 시스템이 내리는 경우(예를 들어 대출 승인, 학생 성적 또는 직원 채용 결정), 학습 데이터를 선별하는 조직은 이러한 AI 시스템의 의사 결정에 대한 책임을 지고 불법적인 차별을 방지해야 합니다.

데이터 품질은 책임감 있는 AI 모델을 만들기 위한 선행 조건입니다. 데이터 거버넌스 관행은 데이터가 품질 표준을 준수하고 신뢰할 수 있도록 보장하는 방법입니다.

데이터의 신뢰성은 AI 혁신을 위한 기반입니다. AI 모델에 대한 모든 고려사항을 이 자료에서 다룰 수는 없지만 몇 가지 생각해 볼 만한 부분은 다음과 같습니다.

-  생성형 AI 기반 모델은 대용량의 데이터 코퍼스로 사전 학습됩니다. 생성형 AI를 배포할 때는 이 기반 모델이 편향되지 않았음을 신뢰하는 상태에서 배포하는 것입니다. 따라서 사용하는 모델의 학습 데이터와 관련된 모든 분쟁에 대한 손해배상 책임을 제공하는 제공업체를 찾는 것이 현명합니다.
-  고품질 AI 모델을 만드는 것은 학습 데이터가 전부가 아닙니다. 데이터를 중심에 두고 [이를 기반으로 모델을 구축](#)하는 것이 중요합니다. 임베딩은 데이터 간의 시맨틱 관계를 활용하여 모델이 LLM의 일반적인 결과물인 확률적 답변이 아닌 정확한 답변을 제공하도록 합니다.
-  인간의 지능은 AI 시스템의 편향을 모니터링하는데 중요한 역할을 하며, 포괄적인 데이터 거버넌스 접근 방식은 신뢰할 수 있는 데이터를 기반으로 한 AI 학습이 이뤄지도록 보장합니다.

# 다양한 업계에서 데이터 기반 의사 결정을 강화하는 방법



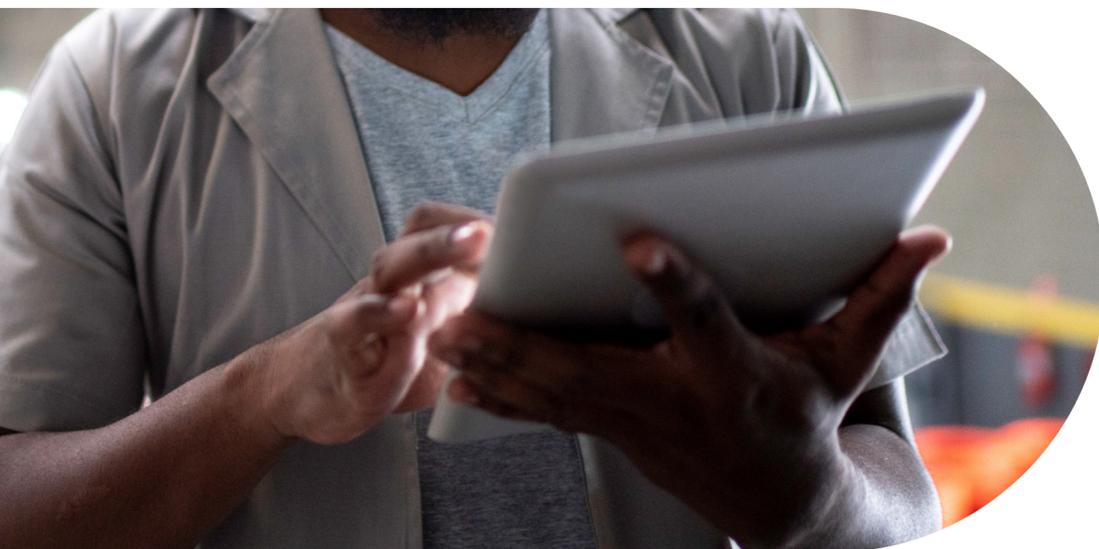
## 금융 서비스

[Macquarie Bank](#)는 소매 금융 고객에게 고도로 개인화된 서비스를 제공합니다. 여기에는 고객이 더 효과적으로 재정을 관리할 수 있도록 거래와 지출에 대한 전체적인 시야를 제공하는 자동 분류 기능이 포함됩니다.



## 의료 서비스

[존스 홉킨스 대학교 BIOS 사업부](#)는 뇌손상 환자를 위한 더 빠르고 정확한 의사 결정을 통해 500건의 스캔을 진단하는 데 소요되는 시간을 2,500시간에서 90분으로 단축했으며 연구 및 인프라 비용을 50% 절감했습니다.



## 다음 단계

데이터의 품질을 평가할 때 처음에는 많은 조직이 다음과 같은 질문과 관련된 어려움을 겪습니다.

- 신뢰할 수 있는 출처에서 추출된 데이터인가?
- 이 데이터가 마지막으로 갱신된 시점은 언제인가?
- 민감한 정보를 어떻게 이동 또는 복사하는가? 데이터 거버넌스 관행을 준수하는가?

조직은 **데이터 관측 가능성**을 통해 데이터를 모니터링하고 문제를 선제적으로 처리함으로써 데이터 품질을 개선할 수 있습니다.

- [데이터 계보 도구](#)를 구현하여 민감한 정보를 비롯한 데이터를 어떻게 가져오고 변환하는지 추적합니다.
- 세분화되고 광범위한 확인 과정을 통해 생태계 전반에 대한 뷰를 확보하고 적절한 알림과 보고서를 제공합니다.
- 데이터에 플래그가 지정되면 해당 데이터를 격리하고 조치를 취하여 문제가 되기 전에 해결합니다.

이러한 선제적인 관행은 사용자가 데이터에 액세스할 때 그 데이터의 신뢰성을 보장합니다. 이 수준의 신뢰는 ROI를 이끄는 비즈니스 인텔리전스 인사이트와 데이터 민주화를 위한 기반입니다.

선제적인 데이터 거버넌스를 달성하기 위한 세 가지 방법

2 기술 전문가가 아닌 사용자도 데이터 기반 의사 결정권자가 되도록 지원





데이터팀에서 병목 현상은 흔한 일입니다. 사람들이 업무에서 능숙하게 데이터를 다루는 데 필요한 인사이트와 노하우에 접근할 수 있도록 하는 데이터 민주화 통해 이 부담을 덜 수 있습니다.

그러나 많은 조직에는 데이터 셀프 서비스를 실현하는 동시에 일관적인 거버넌스를 유지하기 위한 도움이 필요합니다.

이들 조직에는 팀에 필요한 데이터에 대한 폭넓은 액세스를 제공하면서 높은 위험과 비용을 낮출 수 있는 프레임워크가 필요합니다. 강력한 데이터 거버넌스 정책으로 뒷받침되는 안전한 셀프 액세스 도구와 프로세스는 액세스와 보안 사이의 균형을 맞추는 열쇠입니다.

그러면 다음과 같은 결과로 이어집니다.

- 인사이트 발굴
- 데이터에 대한 셀프 서비스를 안전하게 자동화
- 혁신 강화

69%

데이터 전문가의 69%는 데이터 액세스 문제 대응, 관리, 해결에 매주 평균 6~10시간을 소비합니다. 월 단위로는 24~40시간, 연 단위로는 288~480시간에 해당하는 시간입니다.<sup>4</sup>



## 인사이트 발굴

비즈니스 인텔리전스는 단순히 빅데이터와 비정형 데이터에 액세스할 수 있도록 한다고 해서 구현되지 않습니다. **인사이트**(데이터의 의미와 함의)의 민주화는 가치 창출로 이어집니다.

갱신 관리자가 중요한 통화에 앞서 CRM에서 고객의 프로필을 검토한다고 상상해 보세요. 직원은 해당 고객에 대한 데이터에 액세스하여 데이터 테이블을 살펴보고 고객이 소프트웨어 플랫폼을 어떻게 사용하는지에 대한 정보를 확인하여 유용한 내용을 확보할 수 있습니다. 해당 계정에 대한 인사이트에 액세스할 수 있는 갱신 관리자의 경우 고객이 지속적으로 사용량 한도에 도달했으며 따라서 더 많은 라이선스를 업셀링할 만한 유력한 후보임을 즉각 확인할 수 있습니다.

AI 기반의 임베디드 분석 도구는 필요할 때 적절한 인사이트에 액세스할 수 있도록 해 줍니다. 또한 적절한 통합을 통해 직원들이 비즈니스 인텔리전스 플랫폼 외부에서도 익숙한 워크플로에 따라 이러한 분석에 액세스할 수 있습니다.

분석은 기술 사용자나 데이터 엔지니어의 전유물이 되어서는 안 됩니다. 오늘날의 로우 코드 도구는 기술적 역량이 부족한 직원도 인사이트에 액세스할 수 있도록 해줍니다. 가까운 미래에는 자연어 처리와 생성형 AI가 실시간 검색 기능의 표준이 될 것입니다.

데이터 기반 문화 구성에 대해  
자세히 알아보기



## 셀프 서비스를 안전하게 자동화

데이터는 조직의 고객, 시장, 경쟁사를 이해하는 데 매우 중요하며, 셀프 서비스는 인사이트의 잠재력을 더 높여줍니다. 그러나 적절한 보안 제어가 없으면 셀프 서비스로 인해 위험이 커질 수 있습니다.

자동화된 프로세스는 고객 데이터의 안전과 규정 준수를 유지하면서, 특정 데이터 유형에 가장 가까운 사용자에게 자율성을 부여할 수 있습니다.

도메인 기반 아키텍처는 개별 비즈니스팀이 특정 사용 사례 또는 데이터 세트를 소유할 수 있도록 합니다. 강력한 데이터 거버넌스 프레임워크는 포괄적인 규정 준수 제어를 유지하면서 이러한 팀에 데이터에 대한 액세스 권한과 거버넌스를 제공합니다.

96%

디지털 리더의 96%에게 팀에서 아이디어를 내고 직접 의사 결정을 내리도록 지원하는 최고 책임자 경영진이 있었던 반면, 후발 주자의 73%는 리더의 지시를 따르도록 요구받았습니다.<sup>5</sup>

## 혁신 강화

현대의 조직은 데이터를 기반으로 합니다. 더 나은 고객 경험과 성과를 실현하고 미래에 대비하여 조직을 재구성하기 위해서입니다.

빠르게 변화하는 환경 속에서 혁신하기 위한 핵심 요소인 협업은 선도적인 조직이 다른 조직과 구분되는 주된 요인이기도 합니다.<sup>6</sup>

데이터 민주화는 조직 사일로를 허물고 더 많은 직원들이 데이터를 사용하여 문제를 해결하고 새로운 아이디어를 개발할 수 있게 해 줍니다. 또한 협업과 공유 및 상호 작업 간의 이점을 얻기가 쉬워지므로 직원들은 함께 협력하여 더 큰 영향력을 발휘할 수 있습니다.

69%

직원의 69%가 지난 12개월 동안 조직의 사이버 보안 지침을 우회했습니다. 직원의 74%는 자신 또는 팀이 비즈니스 목표를 달성하는 데 도움이 된다면 사이버 보안 지침을 우회할 수 있다고 답했습니다.<sup>6</sup>



# 데이터 민주화 일으키는 대대적인 변화

맞춤 보고서에 소요되는 시간이 몇 주에서 며칠로 단축되고, 이 과정에 들어가는 상당한 엔지니어링 리소스를 다른 곳에 투입할 수 있다고 상상해 보세요. 이것이 민주화가 가진 힘입니다. 데이터 민주화는 많은 기업의 미래상이지만 [Subskribe](#)는 이미 그 지점에 도달했습니다.

Subskribe는 최신 SaaS 플랫폼을 위한 견적에서 매출까지의 통합 플랫폼을 제공하여 주문형 인사이트를 통해 의사 결정을 개선하고 성장할 수 있도록 지원합니다. 이전에는 신규 SaaS 고객을 온보딩하려면 맞춤 보고서 개발에 몇 주간의 엔지니어링 작업이 필요했습니다. Looker Embedded의 통합으로 이제

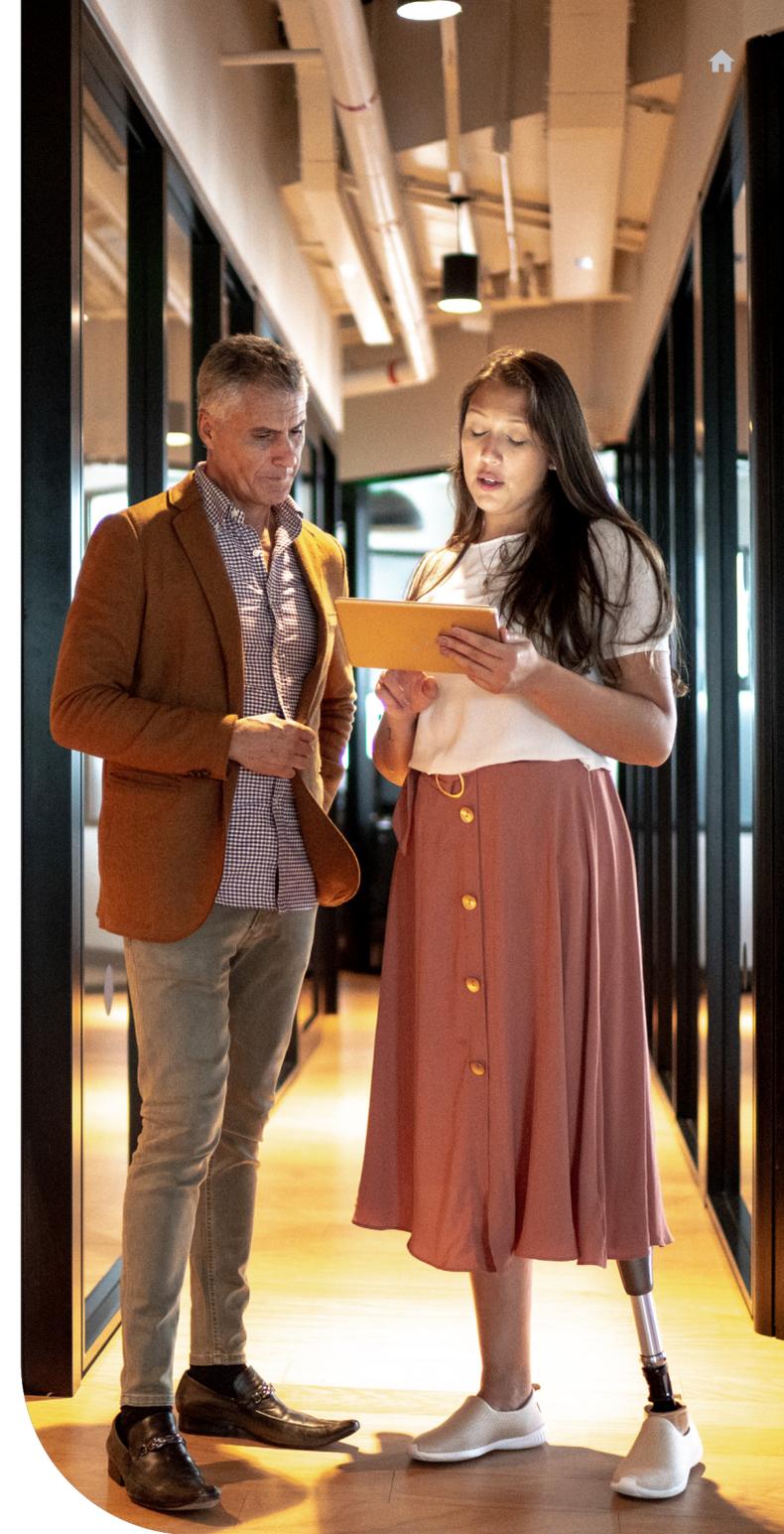
기술 분야 이외의 팀도 코드 없이 정교한 보고서를 개발할 수 있고, 따라서 혁신 기회에 엔지니어링 리소스를 투입할 수 있습니다. 그러나 이는 시작에 불과합니다.

현재 Subskribe는 Looker로 고객이 자체 대시보드와 보고서를 만들 수 있게 해 줄 임베디드 분석 솔루션을 구축하고 있습니다. 또한 확장된 맞춤설정 옵션으로 경영진, 재무 담당자, 고객 성공팀과 같은 사용자 역할에 맞는 신뢰할 수 있는 분석을 제공하여 제품 도입 및 고객 성공에도 이바지할 수 있습니다.

“

누구나 몇 시간 만에 Looker를 사용하여 자신이 볼 수 있도록 승인된 데이터로 자신만의 대시보드와 보고서를 만들 수 있도록 시맨틱 모델을 설계했습니다.”

우구르칸 아크테페  
Subskribe 소프트웨어 엔지니어



# 다음 단계



데이터 민주화 규정을 준수하면서 안전한 데이터에 대한 셀프 서비스 액세스를 지속 가능한 방식으로 지원하는 조직 전반의 문화와 프로세스를 구축하는 것을 의미합니다."

프라작타 담레  
Google Cloud 제품 관리 부문 책임자

협업은 혁신으로 나아가기 위한 핵심 요소입니다. 또한 협업을 가로막는 오래된 장벽인 분산된 팀과 사일로화된 정보는 연결되고 민주화된 현대의 업무 환경에서 사라지고 있습니다.

그러나 보안은 여전히 필수적인 관심사입니다. 직원들은 협업을 할 때 누가 어느 정보에 액세스할 수 있는지에 대해 걱정할 필요가 없어야 합니다. 조직은 보안 내재화 설계된 자동화된 프로세스와 도구를 사용함으로써 협업을 촉진할 수 있습니다.



업계 리더 중 97%가 조직 전체에서 데이터 및 분석에 대한 액세스를 민주화하는 것이 비즈니스 성공에 중요하다고 답했습니다.<sup>7</sup>

보안이 내장된 데이터 민주화 프레임워크:

- 모든 단계에서 데이터를 보호합니다(데이터 액세스, 관리, 보관 방법에 대한 정책 사용).
- 각 데이터 작업에 대한 가시성을 유지합니다(데이터 로그 및 계보 도구 사용).
- 사용자가 알맞은 작업을 손쉽게 수행할 수 있도록 합니다 (직원이 정보를 안전하게 공유할 수 있게 해 주는 자동 보호 기능 및 도구 사용).

[Looker의 임베디드 분석](#)은 사용자의 기술, 요구사항, 선호도에 맞춰 작동합니다.

이 Google Cloud 웹 세미나에서 자세히 알아보기

선제적인 데이터 거버넌스를 달성하기 위한 세 가지 방법

# 3 아키텍처를 뛰어넘기





전 세계적으로 데이터 규정이 강화되면서 조직은 더 복잡한 아키텍처 및 데이터 액세스 수요를 보호하기 위해 고심하고 있습니다.

많은 기업이 온프레미스와 클라우드를 혼합한 하이브리드 환경에서 일하거나 여러 클라우드를 사용해 작업을 수행하고 있는 만큼, 규정 준수는 더 이상 개별 아키텍처에 국한될 수 없습니다.

규정 준수를 위해서는 인프라에 무관한 포괄적인 정책과 함께 선제적인 엔드 투 엔드 전략이 필요합니다.

그러면 다음과 같은 결과로 이어집니다.

- 규정 관련 업데이트에 손쉽게 적응
- 여러 아키텍처에 걸쳐 규정 준수 달성
- 메타데이터 관점으로 데이터 통합

**각 국가에서 데이터 규정을 빠른 속도로 도입함에 따라, 기업들은 규정 준수 상태를 유지하는 데 어려움을 겪고 있습니다.<sup>8</sup>**

## 규정 관련 업데이트에 손쉽게 적응

데이터 규정의 범위는 여러 지역과 산업에 걸쳐 넓어지고 있습니다. 한때는 데이터 집약적인 산업의 영역이었지만 이제는 소규모 기업도 고객 데이터를 안전하게 보호하기 위해 정부의 개인 정보 보호 및 보안 규정을 준수해야 합니다.

규정 준수가 부가적(사후 대응적)인 경우 새로운 규정이 생길 때마다 데이터 거버넌스 프레임워크는 더 복잡해집니다. 그러나 규정 준수가 기본적(선제적)이고 비즈니스 중심의 프로세스에 기반하면 조직은 엔드 투 엔드 규정 준수 상태를 유지하고 규정 관련 업데이트를 손쉽게 관리할 수 있습니다.



**클라우드 데이터 관리 기능(CDMC)**  
프레임워크는 클라우드 환경의 데이터를 보호하고 전 세계의 데이터 개인 정보 보호 요건을 충족하는 데 필요한 주요 제어 기능을 정의합니다.

## 디지털 주권에 대해 알아보기

Google Cloud의 **디지털 주권 탐색기**는 조직의 디지털 주권 요구사항에 대한 일련의 질문을 통해 사용자를 안내하는 온라인 도구입니다.

## 여러 아키텍처에 걸쳐 규정 준수 달성

많은 조직이 여러 클라우드와 아키텍처에 걸쳐 시스템을 운영하면서 중요한 서비스를 제공하고 민감한 정보를 보호합니다. 확장성은 조직이 보안과 규정 준수 상태를 유지할 수 있게 해 줍니다. 또한 오픈소스는 엔드 투 엔드 확장성을 제공할 수 있는 파트너와의 통합을 가능하게 해 줍니다.

**오픈소스는 확장성을 개선합니다.** 독점 소프트웨어는 상호 운용성 문제가 만연하고 생존 가능성 측면의 문제도 있습니다. 오픈소스 도구와 개방형 표준을 기반으로 하는 솔루션은 통합이 쉽고 개발 비용의 상당 부분을 절감해 주며,

핵심 워크로드를 여러 퍼블릭 클라우드 플랫폼에 걸쳐 또는 그 외부에 배포(필요한 경우 마이그레이션)할 수 있는 유연성을 제공합니다. 규정 준수를 유지하면서 기능적인 인사이트를 수집하기 위해서는 여러 아키텍처에 걸쳐 연결되는 도구를 보유하는 것이 중요합니다.



기업의 99%가 오픈소스를 사용 중이며 기업 소프트웨어의 35%는 오픈소스 코드를 기반으로 합니다.<sup>11</sup> 2022년에는 오픈소스 데이터 다운로드가 사상 최고를 기록했습니다.<sup>12</sup>

수집되었지만 사용은 되지 않는 데이터를 '다크 데이터'라고 합니다. 이 용어에는 발견되지 않은 데이터, 활용도가 낮은 데이터, 부적절하게 분류된 개인 식별 정보(PII)가 포함됩니다.

66%

조직의 66%는 데이터의 절반 이상이 다크 데이터이며, 상당한 위험 요소라고 응답했습니다.

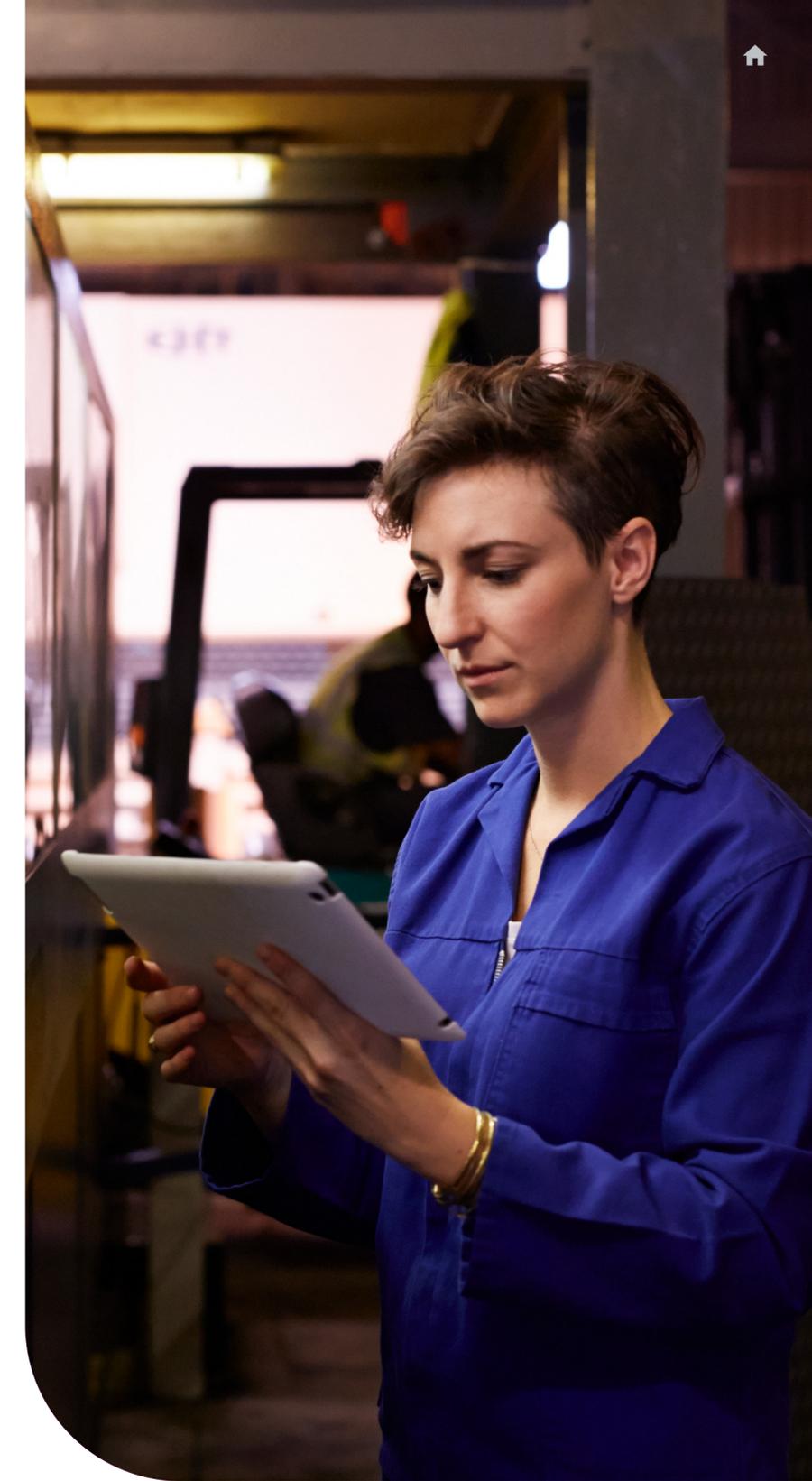
## 메타데이터 관점으로 데이터 통합

데이터는 중요합니다. 그러나 먼저 조직은 데이터가 어디에 있는지 알아야 합니다. AI 기반 메타데이터 태깅은 조직에서 구조화된 데이터와 구조화되지 않은 데이터를 발견 및 활용하는 동시에 데이터 계보에 대한 이해도를 높일 수 있게 해 줍니다.

전통적으로 분석은 데이터 웨어하우스에 최적화되어 있었습니다. 데이터 웨어하우스는 데이터를 추출, 변환해서 특정 형식으로 로드해야 하는 구조화된 아키텍처입니다. 현재 조직은 덜 경직된 아키텍처를 활용할 수 있지만, 엔드 투

엔드 발견 가능성이 없으면 다크 데이터는 새로운 골칫거리입니다. 자동화는 데이터를 적절히 태깅, 저장, 폐기하여 데이터 주권 규정 및 데이터 민주화 규정을 준수하기 위해 반드시 필요합니다.

메타데이터 태깅 및 주석 전략을 구현하기 위해서는 기술 그 이상이 필요합니다. 자동화는 잘 알려진 데이터 세트를 발견하는 데 도움이 되지만 조직은 여전히 구체적인 비즈니스 컨텍스트를 파악, 정의하고 데이터 거버넌스 전략의 일부로 포함해야 합니다.



Google Cloud는 포트폴리오 전반에서 보관 정책을 지원합니다. 따라서 일정 기간이 지나면 되면 Google Cloud는 어떤 상황이 발생했는지 나타내는 로그를 비롯한 데이터를 옮기거나 삭제합니다.

# 자동화로 성과를 내는 방법



[Deutsche Bank](#)는 제어를 자동화함으로써 혁신 가속화를 위한 핵심 동력을 확보했습니다.



[Achievers](#)는 190개국에서 360만 명 이상의 사용자를 보유하고 있으며, 빠른 성장 중에도 데이터 품질을 유지하면서 조직 전반에 데이터 신뢰성을 제공하고 팀 전체에 걸쳐 수백 시간을 절약했습니다.



[Casa Dos Ventos](#)는 빈번한 '급한 불 끄기'식 데이터 처리 방식에서 자동 데이터 품질 검사로 전환했습니다. 자동 데이터 품질 검사는 데이터에 대한 신뢰를 구축하고 적절한 효율성 목표를 설정하는 데 필요한 컨텍스트를 제공하며 브라질 재생 에너지 시장에 현명하게 투자할 수 있도록 해 줍니다.

# 다음 단계

데이터 거버넌스는 데이터 폐기를 포함한 데이터의 전체 수명 주기 관리에 영향을 미칩니다. 많은 관할권에서는 정해진 기간 이후 개인 정보를 파기할 것을 요구합니다. 그러나 규정의 복잡성과 조직 데이터의 단편화된 특성으로 인해 기술적으로 규정 준수가 어려울 수 있습니다.<sup>2</sup>

조직에서 규정 준수 상태를 유지하기 위한 3단계는 다음과 같습니다.



## 데이터 발견

과거에 비해 데이터가 훨씬 더 많아지면서 데이터가 어디에 위치하는지 파악하는 것뿐 아니라, 데이터가 더 엄격한 보안 및 규정 준수 정책이 적용되는 민감한 정보에 해당하는지 여부를 파악하는 것도 중요해졌습니다. 이에 대한 가시성을 확보하려면 엔드 투 엔드 발견 가능성과 데이터 계보 도구가 필요합니다.



## 주석

비즈니스 컨텍스트로 데이터를 분류해서 의미 있는 비즈니스 도메인을 만듭니다. 그러면 인프라가 아닌 비즈니스를 보는 관점에서 데이터 거버넌스 전략을 수립할 수 있습니다.



## 자동화된 거버넌스 정책 적용

주석을 기반으로 거버넌스 인텐트를 표현하고 분산된 데이터 전반에 대규모로 이러한 정책을 적용하여 데이터를 보호 및 관리하고 다양한 규정에 대한 준수를 보장합니다. 비즈니스 컨텍스트를 기반으로 이러한 정책을 정의하고 자동화하면 데이터가 필요에 따라 안전하게 액세스 및 삭제되도록 하고, 규제 기관에 보고할 감사 로그도 유지할 수 있습니다.

# 다크 데이터, 중복 데이터, 더티 데이터를 제거하여 ROI 극대화하기





# 조직은 전례 없는 수준으로 데이터를 수집하고 있습니다.

이 자산을 활용하면 탁월한 고객 경험을 제공하고 미래를 위해 혁신하는 데이터 기반 조직이 될 수 있습니다.

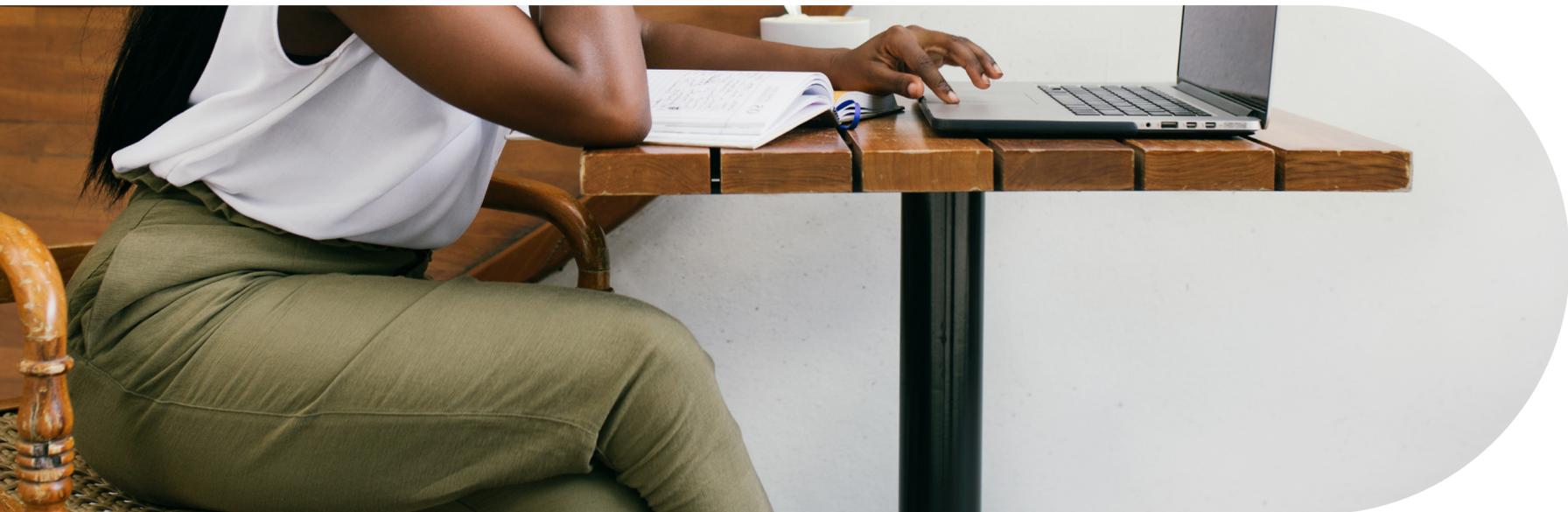
그러나 데이터가 골칫거리인 경우도 많습니다. 다크 데이터, 중복된 데이터, 더티 데이터 모두 혼란을 일으키고 조직에서 데이터의 잠재력을 현실화하지 못하도록 합니다.

포괄적인 데이터 거버넌스 전략은 데이터 기반 조직을 향한 길을 가로막는 장애물을 제거합니다. 조직은 데이터 품질, 데이터 보안, 데이터 규정 준수를 보장함으로써 조직의 잠재력을 끌어낼 수 있습니다.

"채팅 애플리케이션이나 로그 파일의 구조화되지 않은 데이터에 PII와 같은 민감한 정보가 예기치 않게 포함되는 경우 조직에 심각한 문제를 일으킬 수 있습니다. 사람들이 어떤 정보를 제출할지 알 수 없는 고객 지원 스크립트가 그러한 예에 해당합니다.

누군가 고객 지원팀과 채팅을 하면서 '약을 받지 못했습니다. 제 이름과 필요한 약, 주민등록번호를 알려드립니다.'라고 입력한다면, 이러한 민감한 PII 데이터가 적절하게 보호 및 분류되지 않을 수도 있는 데이터베이스에 저장될 수 있는 것입니다.

안톤 슈바킨, 수석 보안 컨설턴트,  
Google Cloud([데이터 및 AI 트렌드](#))



현대 조직은 데이터 거버넌스에 대해 기본 제공되는 AI 기반의 선제적인 접근 방식을 채택하고 있습니다. 이를 통해 직원들은 필요한 인사이트에 액세스하고 고객 데이터는 안전하게 유지되며 조직은 AI 혁신을 수용할 수 있습니다.

**데이터는 이미 주어져 있습니다. 발견하고, 민주화하고, 활용하세요.**

**데이터를 골칫거리에서 자산으로 바꿀 준비가 되셨나요?**

지금까지 AI 시대에 조직이 데이터를 최대한 활용하는 방법을 살펴보았습니다. 여기에는 다음이 포함됩니다.

- 데이터 사일로 허물기
- AI 성능 개선
- 데이터에 대한 신뢰 구축
- 인사이트에 대한 액세스 민주화
- 협업 및 혁신 개선
- 아키텍처 전반의 규정 준수 보장

Google Cloud가  
어떤 도움을 제공할 수 있는지  
자세히 알아보고 싶으시다면  
전문가에게 문의해 보세요.



전문가에게 문의하기

[cloud.google.com/contact](https://cloud.google.com/contact)

# 참고문헌

- 1 [정보 유출에 따른 비용 보고서\(Cost of a data breach report\)](#)(2023) IBM.
- 2 [소비자 데이터 기회와 개인 정보 보호 의무\(The consumer-data opportunity and the privacy imperative\)](#)(2020) McKinsey.
- 3 [2023년 데이터 엔지니어링 현황 설문조사\(2023 State of Data Engineering Survey\)](#)(2023) Immuta.
- 4 [Gartner, 사이버 보안 리더의 약 절반이 2025년까지 직업을 바꿀 것으로 전망\(Gartner Predicts Nearly Half of Cybersecurity Leaders Will Change Jobs by 2025\)](#)(2023) Gartner.
- 5 [디지털 가치 확장 비결\(The Keys to Scaling Digital Value\)](#)(2022) Boston Consulting Group, Google 후원
- 6 [데이터에서 탁월한 비즈니스 가치 도출하기\(Turning Data into Unmatched Business Value\)](#)(2023) Harvard Business Review, Google 후원
- 7 [기업 관점에서 데이터 보호법 증가가 갖는 의미\(This is what increasing data protection laws mean for your company\)](#) (2023) 세계 경제 포럼.
- 8 [디지털화의 꿈이 데이터 가시성에 가로막히지 않게 하기\(Don't Let Data \(In\)Visibility Limit Your Digitization Dreams\)](#)(2022) Gartner.
- 9 [소비자 데이터 기회와 개인 정보 보호 의무\(The consumer-data opportunity and the privacy imperative\)](#)(2020) McKinsey.
- 10 [앞서가는 조직이 오픈소스에서 슈퍼 파워를 얻는 방법\(How leading organizations are making open source their super power\)](#)(2022) Google Next '22.
- 11 [소프트웨어 공급망 현황\(State of the Software Supply Chain\)](#)(2023) Sonatype.