

# 3가지 즉각적인 이점을 제공하는 데이터 패브릭 아키텍처

더 나은 비즈니스 결과를 위해  
조직의 모든 데이터 통합 및 활용

## 목차

3 데이터의 현재 당면 과제

5 데이터 패브릭의 비즈니스 이점

6 IBM이 데이터 패브릭을 제공하는 방식

7 데이터 패브릭 적용 사례

7 기업의 혁신적인 다음 단계는 데이터 패브릭

# 데이터의 현재 당면 과제

데이터는 기업의 디지털 혁신에 필수적인 요소입니다. 그러나 기업은 데이터를 활용하려고 할 때 다양한 데이터 소스, 유형, 구조, 환경 및 플랫폼으로 인한 문제에 직면하게 됩니다. 이러한 다차원 데이터 문제는 기업이 하이브리드 및 멀티 클라우드 아키텍처를 채택할 때 더욱 복잡해집니다. 오늘날 많은 기업에서 운영 데이터는 대부분 격리되고 숨겨져 막대한 양의 다크 데이터로 이어지고 있습니다.

복미의 한 에너지 기업이 데이터 중심 조직으로 재탄생하여 여러 사업부에서 데이터 과학 기능에 쉽게 액세스할 수 있는 기업으로 거듭나려고 노력한 사례를 살펴 보겠습니다. 이들은 곧 격리된 데이터, 일관성 없는 톨 및 다양한 기술 레벨로 인해 디지털 전환이 방해받고 있으며 이 모든 것이 데이터 역량 간의 심각한 격차를 초래했다는 사실을 깨달았습니다.

직면한 문제는 이들의 사업에만 국한된 것이 아니라 데이터 관리 아키텍처를 능가하는 데이터 환경의 일반적인 결과입니다.

이 에너지 회사가 디지털 혁신을 위한 성공적인 길로 다시 들어선 것은 데이터 패브릭이라는 새로운 데이터 아키텍처 개념을 채택했기 때문입니다.

이 백서를 통해 데이터 패브릭이란 정확히 무엇이고 기존 아키텍처와 어떻게 다르며 기업에서 달성할 수 있는 것은 무엇이고 데이터 패브릭을 구현하는 IBM의 역할이 무엇인지에 대해 안내해드립니다

## 데이터 패브릭이란?

과거에 조직은 지점간 통합 또는 데이터 허브 도입을 통해 데이터 액세스 문제를 해결하려고 시도했습니다. 데이터가 고도로 분산되고 격리된 경우에는 이러한 방법 중 어느 것도 적합하지 않습니다. 지점간 통합의 경우 추가 엔드포인트를 연결해야 하는 관계로 기하급수적인 비용이 추가로 발생하므로 확장 불가능한 접근 방식입니다. 데이터 허브를 사용하면 애플리케이션과 소스를 더 쉽게 통합할 수 있지만 허브 내에서 데이터의 품질과 신뢰를 유지하기 위해 비용과 복잡성이 악화됩니다.

데이터 패브릭은 하이브리드 데이터 환경에서 발생하는 데이터 문제의 해결을 목표로 하는 새로운 아키텍처입니다. 데이터 엔드포인트 간의 가상 연결 조직 역할을 수행함으로써 분산과 세계화 간의 균형을 맞추는 것이 기본 구상입니다(4페이지 참조).

데이터 패브릭 아키텍처는 통합의 자동화 및 강화, 연합 거버넌스 및 메타데이터 활성화와 같은 기술을 통해 분산 환경 전반에서 동적이고 지능적인 데이터 조정을 가능하게 하여 즉시 사용 가능한 정보의 네트워크를 구축하여 비즈니스에 활력을 불어넣습니다.

데이터 패브릭은 배치 플랫폼, 데이터 프로세스, 지리적 위치 및 아키텍처 접근 방식에 구애받지 않습니다. 따라서 데이터를 엔터프라이즈 자산으로 쉽게 사용할 수 있습니다. 데이터 패브릭은 다양한 종류의 데이터를 효율적이고 효과적으로 결합, 액세스 및 제어할 수 있도록 보장합니다.

## 데이터 패브릭의 기능 및 원칙

데이터 패브릭 아키텍처의 핵심은 검색, 거버넌스, 큐레이션 및 오케스트레이션을 비롯한 광범위한 통합 데이터 관리 기능을 지원하는 데이터 관리 플랫폼입니다.

그러나 데이터 패브릭은 데이터 운영 수준을 높이기 위해 관행 수립에만 중점을 둔 DataOps와 같은 기존 데이터 관리 개념에서 발전하고 진화합니다. 데이터 자산의 과도한 다양성과 분산으로 인해 발생하는 요구사항을 해결할 수 있는 분산 아키텍처와 고급 기술을 기반으로 구축됩니다.

데이터 패브릭은 논리적으로 다음 네 가지 기능(또는 구성요소)으로 나눌 수 있습니다.

**지식, 인사이트 및 시맨틱**

- 데이터 마켓플레이스 및 쇼핑 환경을 제공합니다.
- 검색된 데이터 자산에 지식과 시맨틱을 자동으로 보강하여 소비자가 데이터를 찾고 이해할 수 있도록 합니다.

**통합 거버넌스 및 규제 준수**

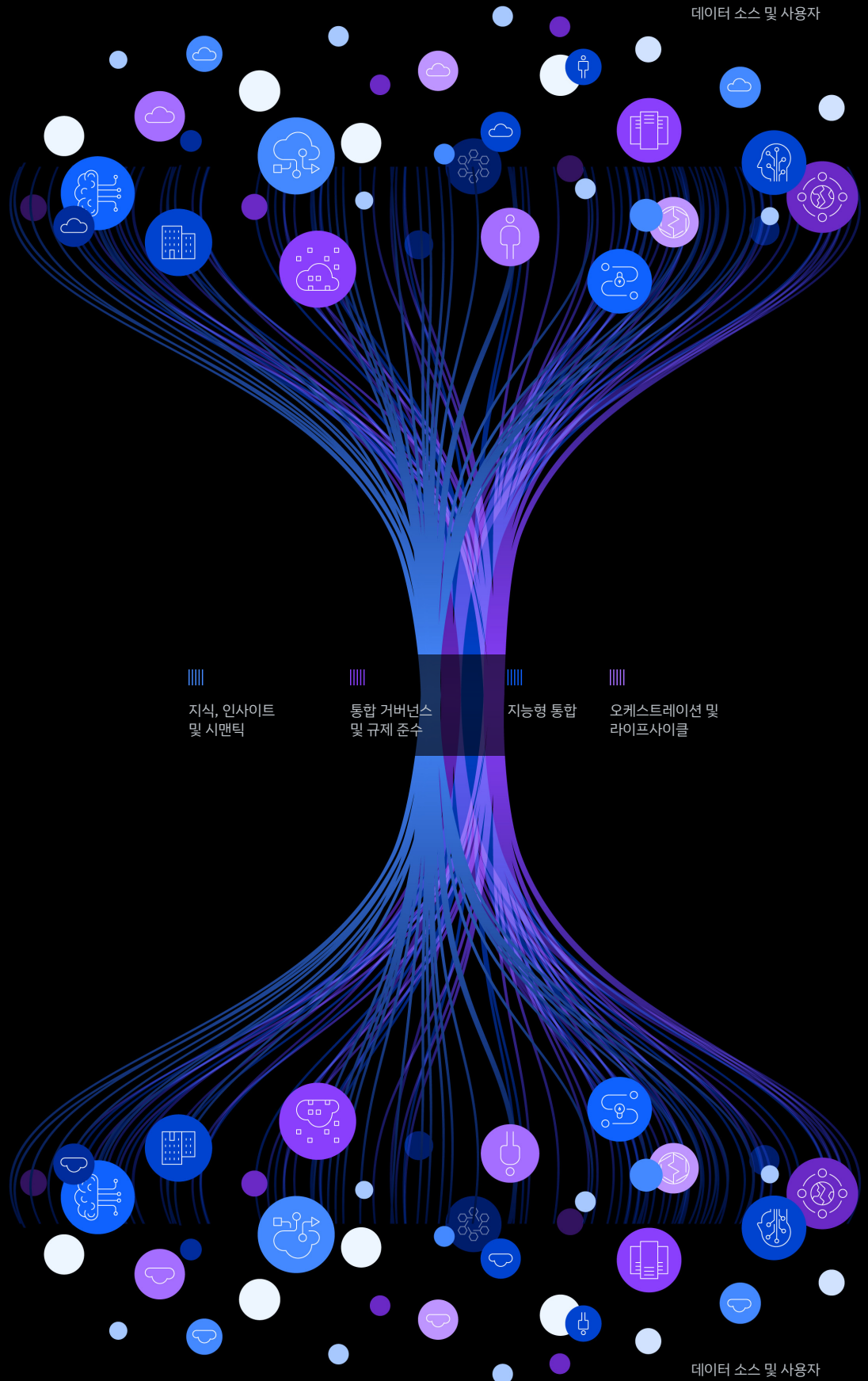
- 메타데이터의 로컬 관리 및 거버넌스를 허용하지만 글로벌 통합 보기 및 정책 시행을 지원합니다.
- 글로벌 및 로컬 규칙에 따라 데이터 자산에 정책을 자동으로 적용합니다.
- 고급 기능을 활용하여 데이터 자산 분류 및 큐레이션을 자동화합니다.
- 데이터 활성화 증대를 위해 카탈로그화된 자산에 대해 쿼리 가능한 액세스 경로를 자동으로 설정합니다.

**지능형 통합**

- 분산 데이터 소스 전반에 걸쳐 자동화된 플로우 및 파이프라인 생성을 통해 데이터 엔지니어의 태스크를 가속화합니다.
- 데이터 보호 정책의 로컬 및 글로벌 실행을 통해 모든 데이터에 대한 셀프 서비스 수집 및 데이터 액세스를 지원합니다.
- 최적화된 워크로드 분배 및 스키마 드리프트의 자가 튜닝 및 수정을 통해 최적의 실행을 자동으로 결정합니다.

**오케스트레이션 및 라이프사이클**

- 데이터 파이프라인을 구성, 테스트, 작동 및 모니터링할 수 있습니다.
- 데이터 라이프사이클에 AI 기능을 접목하여 태스크를 자동화하고, 자가 튜닝 및 자가 치유를 수행하며, 소스 데이터 변경을 감지하여 자동 업데이트를 지원합니다.



지식, 인사이트 및 시맨틱      통합 거버넌스 및 규제 준수      지능형 통합      오케스트레이션 및 라이프사이클

# 데이터 패브릭의 비즈니스 이점

데이터는 컨텍스트화되고 조직의 모든 사용자 또는 애플리케이션이 액세스할 수 있을 때만 비즈니스 가치를 제공합니다. 데이터 패브릭을 올바르게 구현하면 조직 전체에서 이러한 가치를 최대한 효율적이고 자동화된 방식으로 사용할 수 있습니다. 따라서 패브릭은 다음과 같은 세 가지 주요 이점을 제공합니다.

1. 셀프 서비스 데이터 사용 및 협업 지원
2. 활성 메타데이터를 통해 거버넌스, 보호 및 보안 자동화
3. 데이터 엔지니어링 태스크 자동화 및 하이브리드 클라우드 리소스 간 데이터 통합 강화

## 셀프 서비스 데이터 사용 및 협업 지원

여러 소스의 데이터를 통합하고 매일 생성되는 엄청난 양의 데이터를 분석함으로써 조직은 더 나은 통찰력을 얻고 변화하는 비즈니스 요구에 더 빠르게 대응할 수 있습니다. 데이터 패브릭은 데이터를 필요로 하는 사람들에게 신속하게 제공합니다. 셀프 서비스를 사용하면 조직 전체가 적절한 데이터를 더 빨리 찾을 수 있으며 해당 데이터를 사용하여 실질적인 통찰력을 제공하는 데 더 많은 시간을 할애할 수 있습니다.

셀프 서비스 데이터 사용을 위한 데이터 패브릭의 이점:

- 비즈니스 사용자가 조직 전체에서 데이터를 찾고, 이해하고, 웨이핑하고, 사용할 수 있는 단일 액세스 지점이 제공됩니다.
- 중앙 집중식 데이터 거버넌스 및 계보를 통해 사용자가 데이터의 의미, 데이터 출처 및 다른 자산과의 관계를 이해할 수 있습니다.
- 광범위하고 사용자 정의 가능한 메타데이터 관리가 손쉽게 확장되며 API를 통해 액세스할 수 있습니다.
- 신뢰할 수 있고 관리되는 데이터에 대한 셀프 서비스 액세스를 통해 다른 사용자와 LOB(Line-of-Business) 협업이 가능합니다.

Forrester Total Economic Impact 연구<sup>1</sup>에 따르면 이러한 기능은 다음을 의미할 수 있습니다.



## 활성 메타데이터를 통해 거버넌스, 데이터 보호 및 보안 자동화

모든 데이터 이니셔티브에 대한 분산된 활성 거버넌스 계층이 신뢰와 투명성을 제공하여 규제 준수 및 규제 위험을 줄입니다. 따라서 데이터 액세스에 대한 자동 정책 시행을 가능하게 하여 높은 수준의 데이터 보호 및 준수를 제공합니다.

AI 및 머신 러닝 기술을 활용하면 데이터 패브릭 사용자가 규제 문서의 언어 및 정의에 따라 데이터 거버넌스 규칙을 자동으로 추출하는 등 자동화 수준을 높일 수 있습니다. 이를 통해 조직은 몇 분 안에 산업별 거버넌스 규칙을 적용하여 값비싼 벌금을 피하고 데이터가 어디에 있는 데이터를 윤리적으로 사용할 수 있습니다.

관리되는 가상화를 위한 데이터 패브릭의 이점:

- 데이터 엔지니어, 데이터 과학자 및 비즈니스 분석가의 민첩성, 보안 및 생산성이 향상됩니다.
- 여러 글로벌 데이터 소스가 하나의 데이터베이스로 나타납니다.
- 업계 최고의 새로운 개인 식별 정보(PII) 및 중요한 데이터 요소 검색이 대규모로 가능합니다.

이러한 기능은 다음을 의미합니다.



## 데이터 엔지니어링 태스크 자동화 및 데이터 통합 강화

고급 데이터 엔지니어링이란 거의 모든 데이터 액세스 또는 전달 프로세스가 자동화되어 지루하거나 오류가 발생하기 쉬운 코딩 프로세스가 필요하지 않음을 의미합니다. 통합 강화를 통해 메타데이터 데이터를 활용하여 데이터 전달 및 액세스를 최적화할 수 있습니다.

데이터 엔지니어링 및 통합을 위한 데이터 패브릭의 이점:

- 자동으로 최적화된 데이터 통합을 통해 데이터 전달을 가속화할 수 있습니다.
- 자동 워크로드 밸런싱 및 탄력적 확장 기능은 모든 환경 및 데이터 볼륨에 대해 작업을 수행할 수 있음을 의미합니다.
- 복원력 및 CI/CD 자동화가 내장되어 있습니다.
- 변경사항을 실시간으로 캡처하는 자동화된 프로세스가 비즈니스 프로세스에 대한 고품질 데이터 전달을 지원합니다.
- 머신 러닝으로 사용자 정의 데이터 검색, 분류 및 큐레이션 프로세스를 자동화하고 확장할 수 있으므로 가치 실현 시간을 단축할 수 있습니다.
- 데이터가 있는 곳이면 어디서든 실시간으로 연속 분석을 자동으로 수행할 수 있습니다.

선도적인 소매업체의 결과<sup>3</sup>는 다음과 같습니다.



# IBM이 데이터 패브릭을 제공하는 방식

IBM Cloud Pak® for Data는 이러한 데이터 패브릭 개념을 가능하게 합니다. IBM Cloud Pak for Data는 데이터 수집, 구성 및 분석을 단순화 및 자동화하고 비즈니스 전반에 걸쳐 AI 주입을 가속화하는 인사이트 플랫폼입니다.

어디서나 데이터를 연결하고, 어디서나 워크로드를 실행하며 하이브리드 클라우드 환경에서 규모에 맞는 AI를 구축, 배치 및 관리할 수 있는 기능을 갖춘 IBM Cloud Pak for Data는 비즈니스 디지털 혁신을 가능하게 합니다.

이 플랫폼은 하이브리드 엔터프라이즈 전반에서 다음에 대한 원활한 통합을 제공합니다.

- IBM Cloud Pak for Data 서비스
- 외부 애플리케이션 및 데이터 소스
- 데이터 관리 및 거버넌스를 위한 고급 AI 기반 기능

이 기반은 비용, 성능, 규제 준수의 최적 균형과 워크로드, 데이터 지역성 및 데이터 정책을 기반으로 데이터 처리를 조정하고 최적화할 수 있는 인텔리전스를 통해 소비자가 큐레이션된 데이터를 사용할 수 있도록 합니다.

IBM Cloud Pak for Data는 다음 기능을 활용하여 데이터 패브릭이 요구하는 비즈니스용 데이터를 제공합니다. 이러한 모든 기능은 데이터 패브릭 아키텍처를 지원하는 데 중요한 역할을 합니다.

## 1. 메타데이터 기반 지식 코어

데이터 소스 및 카탈로그를 쉽게 검색할 수 있으며, 데이터 자산을 강화하고, AI를 사용하여 더 많은 자동화를 위한 통찰력을 추출하기 위한 분석을 수행합니다. 지식 코어는 시맨틱 검색 기능으로 마켓플레이스를 강화하는 데 사용됩니다.

## 2. 셀프 서비스 데이터 마켓플레이스

비즈니스 분석가와 같은 데이터 소비자가 기업의 데이터 환경 전반에서 데이터를 검색하는 데 도움이 되는 차세대 데이터 카탈로그입니다.

## 3. 스마트 통합

데이터를 추출, 가상화, 변환 및 스트리밍하여 데이터를 사용할 수 있습니다. 지식 코어와 통합되어 데이터 통합을 자동화하며, 워크로드 및 데이터 정책을 기반으로 가장 적합한 통합 접근 방식을 결정할 수 있는 인텔리전스를 제공합니다. 또한 데이터 엔지니어링 워크로드의 일부로 데이터를 준비하거나 데이터 제품을 만드는 데도 사용할 수 있습니다. 끝으로, 데이터 제품에 대한 업데이트를 게시하는 기능을 제공합니다.

## 4. 거버넌스

메타데이터를 카탈로그화 및 큐레이션하고, 개인정보 보호를 위한 데이터 정책을 정의하며, 데이터를 큐레이션하고, 데이터 계보를 캡처하며, 보안 및 규제 준수와 관련된 기타 태스크를 수행합니다. 이 계층은 데이터 형식(예: 정형 대 비정형) 및 데이터 중요도(예: 공개 대 보호)를 이해하고 각 데이터 비트와 각 잠재 사용자에게 올바른 정책을 적용합니다. 이 통합 플랫폼 기능을 사용하면 표준과 규칙을 데이터에 수동으로 적용하는 대신 조직 수준에서 적용하고 필요에 따라 다양한 데이터 리소스 전체에 전파할 수 있습니다. 서로 다른 톨의 분석 모델은 서로 통신할 수 있으며, 세부적인 레벨에서 데이터 정책 시행을 대부분 자동화할 수 있습니다.

## 5. 통합 개발 및 운영 통합 라이프사이클을 통해 운용 중인 데이터 플랫폼의 모든 측면을 구성하고 실행할 수 있습니다.







데이터 소비자/비즈니스 프로세스					
<b>통합 라이프사이클</b> - 설계 - 구성 - 구축 - Test - 조정 - 전개 - 운영 - 모니터링	<b>데이터 셀프 서비스</b> 탐색, 협업, 데이터 마켓플레이스, 프로비저닝	<b>통합 데이터 거버넌스, 보안, 규제 준수, 재무 거버넌스</b> - 데이터 정책 - 데이터 계보 - 데이터 품질 - 데이터 개인정보 보호 - 비용 최적화			
	<b>스마트 통합</b> 데이터 수집, 변환, 준비, 가상화, 실시간 데이터				
	<b>증강된 지식</b> 검색, 카탈로그, 시맨틱, 온톨로지, 그래프, 메타데이터, 마스터 데이터, 참조 데이터, 이벤트 기반 데이터				
IBM 퍼블릭 클라우드 	AWS 	Microsoft 	Google 	에지 	개인용 

그림 1. AI 개발 및 데이터 패브릭을 지원하는 IBM Cloud Pak for Data 플랫폼의 기능

# 데이터 패브릭 적용 사례

데이터 패브릭의 가치를 이해하는 가장 좋은 방법은 인사이트 플랫폼을 기반으로 데이터 패브릭을 구현하여 실제 조직이 달성한 비즈니스 이점을 파악하는 것입니다.

## 에너지 회사

본 백서의 앞부분에서 다룬 당면 과제 절에서는 디지털 혁신을 겪고 있는 북미 에너지 기업의 사례를 소개했습니다.

이 고객은 IBM Cloud Pak for Data를 기반으로 데이터 패브릭 아키텍처를 구현한 IBM과 계약했습니다.

유연하고 통합된 데이터 패브릭을 통해 이 회사는 다음과 같이 다양한 사업부 전반에서 중요한 데이터 프로젝트를 구현할 수 있었습니다.

- eMobility
- 가스 운영 문서 검색(필기 추출 포함)
- 전기 고객 세분화 및 부하 예측
- 자산 관리
- COVID-19 부하 영향
- 업무 복귀 위험 모델

### 데이터 패브릭을 통해 이 조직이 얻을 수 있는 이점

- 셀프 서비스 인사이트 플랫폼을 통해 여러 사업부에서 직접 데이터에 액세스할 수 있도록 지원합니다.
- 목적에 맞는 컴퓨팅 용량을 사용하여 수십억 개의 데이터 행에서 모델을 효율적으로 실행합니다.
- 모델을 공동으로 개발하고 이러한 모델을 쉽게 배치하여 회사 전체에 통찰력을 풀어냅니다.

## 이동 감소 및 감독 개선

대규모 산업 서비스 제공업체는 데이터 거버넌스를 개선하고 규제 준수를 촉진했습니다.

이 조직은 대량의 데이터를 클라우드 기반 데이터 레이크로 이동하는 데 어려움을 겪었습니다. 모든 데이터 액세스 인스턴스에 강력한 거버넌스 계층을 적용하고 비즈니스 사용자 액세스를 제공하기 전에 데이터 품질을 결정해야 하는 두 가지 요구 사항이 있었습니다.

이 조직은 IBM Cloud Pak for Data를 기반으로 데이터 패브릭을 구현함으로써 데이터 거버넌스, 데이터 규제 준수 및 데이터 변환 프로세스를 대폭 개선했습니다. 데이터 패브릭을 사용하면 클라우드 기반 데이터 레이크 및 온프레미스 SAP 데이터 소스에 있는 수백 개의 데이터 소스에 대한 쉽고 안전한 액세스를 비즈니스 사용자에게 제공할 수 있습니다.

또한 업계 규제 액셀러레이터 툴을 활용하여 데이터 소스에서 GDPR 및 CCPA 규제를 적용할 수 있는 PII 데이터를 스캔하고 있습니다. 자동 데이터 플래그 지정은 온프레미스와 클라우드 모두에서 PII 처리를 위해 관리해야 하는 데이터 소스를 식별하는 데 도움이 됩니다.

더 나은 액세스와 더 빠른 PII 스크리닝을 통해 비즈니스 사용자는 데이터 액세스를 오래 기다리거나 보호된 데이터를 노출할 위험 없이 중요한 통찰력을 얻기 위해 관련 데이터를 마이닝할 수 있습니다.

### 데이터 패브릭을 통해 이 조직이 얻을 수 있는 이점

- 적절한 데이터 거버넌스를 보장하는 동시에 회사 전체의 데이터를 활용합니다.
- 신뢰할 수 있는 데이터를 확보하고 데이터 준비량을 줄일 수 있습니다.
- GDPR 및 CCPA와 같은 개인정보 보호 규제를 준수하도록 지원합니다.

# 기업의 혁신적인 다음 단계는 데이터 패브릭

IBM Cloud Pak for Data 기술을 기반으로 구축된 데이터 패브릭을 사용하면 하이브리드 및 멀티클라우드 데이터 환경에서 데이터 검색, 데이터 거버넌스 및 데이터 소비를 초차동화할 수 있습니다. 데이터 패브릭을 사용하면 비즈니스 사용자의 가치 실현 시간을 단축하고, 데이터 엔지니어링 및 운영을 위한 생산성을 높이며, 거버넌스 및 규제 준수 충실도를 높일 수 있습니다.

시작하기전에 IBM Cloud Pak for Data를 무료로 사용해 보십시오.

IBM Data and AI 담당자와 무료 일대일 상담을 예약하십시오.



© Copyright IBM Corporation 2021

(07326) 서울특별시  
영등포구 국제금융로  
10 서울국제금융센터(3IFC)  
Tel. 02-3781-5114

Produced in the United States of America  
2021년 5월

IBM, IBM 로고 및 IBM Cloud Pak은 미국 및/또는 기타 국가에서 사용되는 International Business Machines Corporation의 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 [ibm.com/trademark](http://ibm.com/trademark)에 있습니다.

이 문서는 최초 발행일을 기준으로 하며, 통지 없이 언제든지 변경될 수 있습니다. IBM이 영업하는 모든 국가에서 모든 오픈링이 제공되는 것은 아닙니다.

인용된 성능 데이터와 고객 예제는 예시 용도로만 제공됩니다. 실제 성능 결과는 특정 구성과 운영 조건에 따라 다를 수 있습니다. 그러나 IBM 제품 및 프로그램과 함께 사용한 기타 다른 제품이나 프로그램의 운영에 대한 평가와 검증은 사용자의 책임입니다. 이 문서의 정보는 상품성, 특정 목적에의 적합성에 대한 보증 및 타인의 권리 침해에 대한 보증이나 조건을 포함하여(단, 이에 한하지 않음) 명시적이든 묵시적이든 일체의 보증 없이 "현상태대로" 제공됩니다. IBM 제품은 제공 조건으로 체결된 계약의 이용 약관에 따라 보증됩니다.

우수 보안 관리제도에 대한 설명: IT 시스템 보안은 귀사 내/외부로부터의 부적절한 접근을 방지, 감지, 대응함으로써 시스템과 정보를 보호하는 일을 포함합니다. 부적절한 접근은 정보의 변경, 파괴 또는 유출을 초래하거나, 타 시스템에 대한 공격을 포함한 귀사 시스템에 대한 피해나 오용을 초래할 수 있습니다. 어떠한 IT 시스템이나 제품도 완벽하게 안전할 수 없으며, 단 하나의 제품이나 보안 조치만으로는 부적절한 접근을 완벽하게 방지하는 데 효과적이지 않을 수 있습니다. IBM 시스템과 제품은 합법적이며 종합적인 보안 접근방법의 일부로서 고안되며, 이러한 접근방법은 필연적으로 추가적인 실행절차를 수반하며 가장 효과적이기 위해서는 다른 시스템, 제품 또는 서비스가 필요할 수도 있습니다. IBM은 시스템과 제품이 임의의 당사자의 악의적 또는 불법적 행위로부터 영향을 받지 않는다는 것을 보장하지는 않습니다.

법률과 규정을 준수하는지 확인해야 할 책임은 고객에게 있습니다. IBM에서는 해당 서비스나 제품을 통해 고객이 법률 또는 규정을 준수한다고 나타내거나 보증하거나 법률 자문을 제공하지 않습니다.

- 1 A Forrester New Technology: Projected Total Economic Impact™ Study Commissioned By IBM, 2020년 2월
- 2 성능 측정값은 다양한 100TB 데이터 소스에 대해 IBM 데이터 가상화를 사용하여 IBM Silicon Valley Labs의 통제된 테스트 환경 내에서 수집되었습니다. 측정값(2019년 5월에 측정)과 성능 향상은 IBM Federation과 비교됩니다.
- 3 IBM 고객의 경험을 기반으로 합니다.