

# DataOps로 AI 여정을 앞당기는 방법

인력, 프로세스, 기술을 통한  
비즈니스 성과 혁신

IBM



# 목차

3

DataOps의 정의 및  
AI에서의 역할

6

사람:  
데이터 사용자의 역량  
강화

8

프로세스:  
DataOps의 방법론

10

기술:  
AI 및 자동화 구축

- 데이터 품질 개선
- 사일로된 데이터 연결
- 신뢰할 수 없는 데이터 해결
- 복잡한 데이터 간소화

16

DataOps 실행의 가치

22

DataOps 시작하기

24

다음 단계

# DataOps의 정의 및 AI에서의 역할

인공 지능(AI) 기술에는 데이터에서 새로운 가치를 발견하고 비즈니스 혁신을 실현할 새로운 길을 개척하는 잠재력이 있습니다. 실제로 Gartner는 기업에서 AI를 통해 더 많은 인텔리전스와 자동화된 워크플로를 구축하고 미래의 성과를 예측 및 구상하며 직원들이 더 높은 가치를 창출하는 업무를 수행하도록 지원하면서, 1 2022년이 되면 AI에서 파생되는 가치가 3조 9천억 달러에 이를 것으로 예상했습니다.

하지만, 우수한 정보 아키텍처(IA)가 없이는 AI도 존재할 수 없습니다. 상호 연결되고 정확하고 빠른 AI를 구현하려면 데이터가 필요하며, 이로 인해 정보 아키텍처의 약

점이 그대로 드러나게 됩니다. 데이터의 품질이 떨어지고 액세스가 불가능하며 단절되어 있고 믿을 수 없다면 AI의 잠재력을 실현하거나 데이터 기반 통찰 또는 운영의 가치를 극대화할 수 없습니다. 이것이 바로 IBM이 조직에서 AI 기술을 준비하고 구현하는 방법에 대한 IBM만의 관점을 보여주는 AI 사다리(그림 1 참조)를 개발한 이유입니다. 조직에서는 먼저 데이터 플랫폼을 현대화하고 이를 기반으로 데이터를 수집, 구성, 분석해 결과적으로 비즈니스에 AI를 적용해야 합니다. 이러한 기능을 구현하려면 IBM Cloud Pak® for Data와 같이 확장성을 갖추고 모든 클라우드에서 실행이 가능한 플랫폼이 필요합니다.

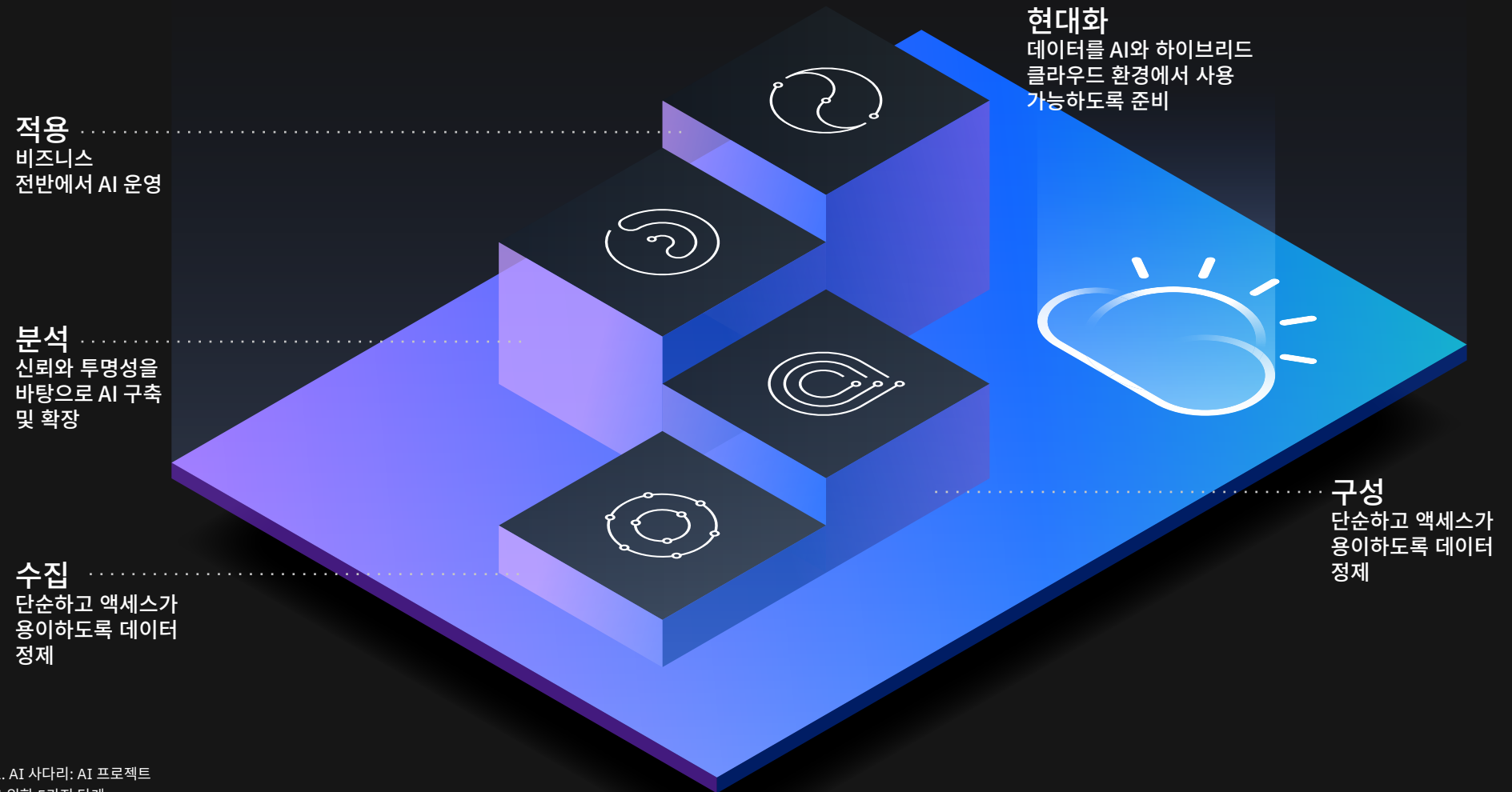


그림 1. AI 사다리: AI 프로젝트 성공을 위한 5가지 단계

## 비즈니스에 적합한 분석 기반 구축

데이터가 AI의 핵심이기 때문에 조직에서는 데이터의 정확성, 일관성, 신뢰성 보장을 위해 데이터의 품질과 거버넌스 향상에 최선을 다해야 합니다. 하지만 데이터 파이프라인 및 흐름을 개발하는 과정에서 운영 모델을 간소화하는 데 어려움을 겪는 경우가 많습니다.

데이터에 관한 가장 보편적인 어려움은 다음과 같습니다.

- 데이터 품질 및 거버넌스 수준 미달
- 구조화 및 비구조화 데이터의 신뢰성
- 관련 데이터 검색 능력
- 조직 내에서 고립된 데이터
- 실시간으로 다루어야 하는 문제에 대한 통찰 확보 지연
- 규정 준수, 개인 정보 보호 및 규제 준수 부담
- 셀프 서비스를 통한 데이터 액세스 기능 제공

이렇게 다양한 데이터 문제를 해결하기 위해 조직들은 지속적으로 애플리케이션을 현대화하고 데이터 가치망을 개선함으로써 새로운 혁신을 가속화하는 동시에 규제를 준수하도록 데이터에 대한 접근 방식을 바꾸는 중입니다. 전사적으로 신뢰할 수 있는 데이터를 제공하려면 새로운 방법과 자동화 기술의 도입으로 데이터 운영 측면에서 운영 최적화를 실현해야 하는데, 이를 DataOps라고 합니다.

DataOps는 AI 사다리의 “구성” 단계에 해당하며 다음과 같은 질문에 대한 해답을 얻을 수 있습니다.

- 조직에서 어떤 데이터를 보유하고 있으며 소유자는 누구인지
- 데이터의 위치는 어디인지
- 데이터가 어떤 시스템에서 사용되고 있으며, 사용 목적은 무엇인지
- 데이터가 각종 규제 및 규정 준수 요구사항을 충족하는지

DataOps는 또한 데이터 사용자와 비즈니스 사용자의 협업 효율성 및 효과를 높일 수 있도록 애자일 개발 프로세스를 데이터 분석에 도입해 협업 데이터 관리 절차를 구축합니다. 또한 자동화의 장점을 활용하여 데이터 액세스, 온보딩, 준비, 통합 및 사용 가능한 상태로의 가공 등을 아우르는 데이터 관리의 비효율성과 관련한 문제를 해결하는 데 도움을 줍니다.

IBM DataOps 기능은 다음과 같은 방법으로 모든 소스로부터 비즈니스에 적합한 데이터를 적시에 필요한 인력에게 전달할 수 있습니다.

- AI 지원 자동화 및 거버넌스 적용
- 시장을 선도하는 엔터프라이즈 지식 카탈로그
- 뛰어난 효율성, 데이터 품질 및 검색 성능
- 업계 유일의 6단계 DataOps 방법론

DataOps는 데이터르 필요로 하는 이에게 신뢰할 수 있는 고품질 데이터르 제공하기 위한 인력, 프로세스, 기술의 집약체입니다.

# 사람: 데이터 사용자의 역량 강화

현대의 기업에는 최고 데이터 책임자에서 데이터 과학자, 분석가, 설계자, 엔지니어 외에도 데이터에서 추출한 통찰을 필요로 하는 개별 사업 부문 사용자에게 이르기까지 서로 다른 다양한 "데이터 사용자"가 있습니다. DataOps는 다양한 조직이 협업을 통해 민첩성과 속도, 그리고 대규모의 신규 데이터 이니셔티브를 추진할 수 있도록 지원하는 데이터 기반 문화를 조성하고 유지할 방법을 모색합니다.

이를 지원하기 위해, DataOps 리더들은 모든 데이터 사용자가 DataOps 문화 및 적용 절차를 발전시키기 위해 수행할 역할을 정의해야 합니다. 주제 영역, 기능, 비즈니스 프로세스, 시스템 또는 프로젝트 등의 하위

데이터 세트 각각을 관리하는 "데이터 책임자"를 지정하는 것도 여기에 포함됩니다. 데이터 책임을 위한 구체적인 모델은 조직의 현재 상태와 미래의 목표에 따라 다를 수 있습니다. 각 조직에는 IT, 데이터 과학 및 사업 부문의 이해관계자가 비즈니스에서 성공을 거두기 위해 가치를 더해야 하는 고유한 요구사항이 존재합니다. 뿐만 아니라, 거버넌스도 DataOps를 지원하는 데 필요한 핵심 요소 중 하나이기 때문에 기존의 데이터 거버넌스 위원회 및 널리 사용되는 데이터 거버넌스 프로그램의 주요 내용을 활용하여 이러한 문화와 책임을 수립할 수 있습니다.



DataOps의 효과란 올바른 기술을 구현하고 신뢰할 수 있는 고품질 데이터를 올바른 인력과 프로세스에 최대한 빠르게 제공하는 셀프 서비스 기능을 개발해 비즈니스 기능의 효율을 높이는 것을 의미합니다. 그림 2에 DataOps의 워크플로우가 나와 있습니다. 인프라와 원시데이터 프로파일링을 위해 설계자, 엔지니어, 분석가가 협업합니다. 분석가, 엔지니어, 과학자는 분석 모델 구축을 위해 협업하며 여기서 AI 사용 여부 등이 결정됩니다. 설계자는 모델 운영, 데이터 거버넌스 및 필요한 곳에 통찰 제공을 위해 비즈니스 사용자와 협업합니다.

각각의 역할을 수행하는 개인은 특정한 하위 데이터 세트에 대한 "데이터 책임자"로 지정됩니다. DataOps 방법론의 요점은 이러한 여러 역할 각각이 정확하고 포괄적이며 안전하고 거버넌스가 보장된 데이터에 의지할 수 있도록 하는 것입니다.

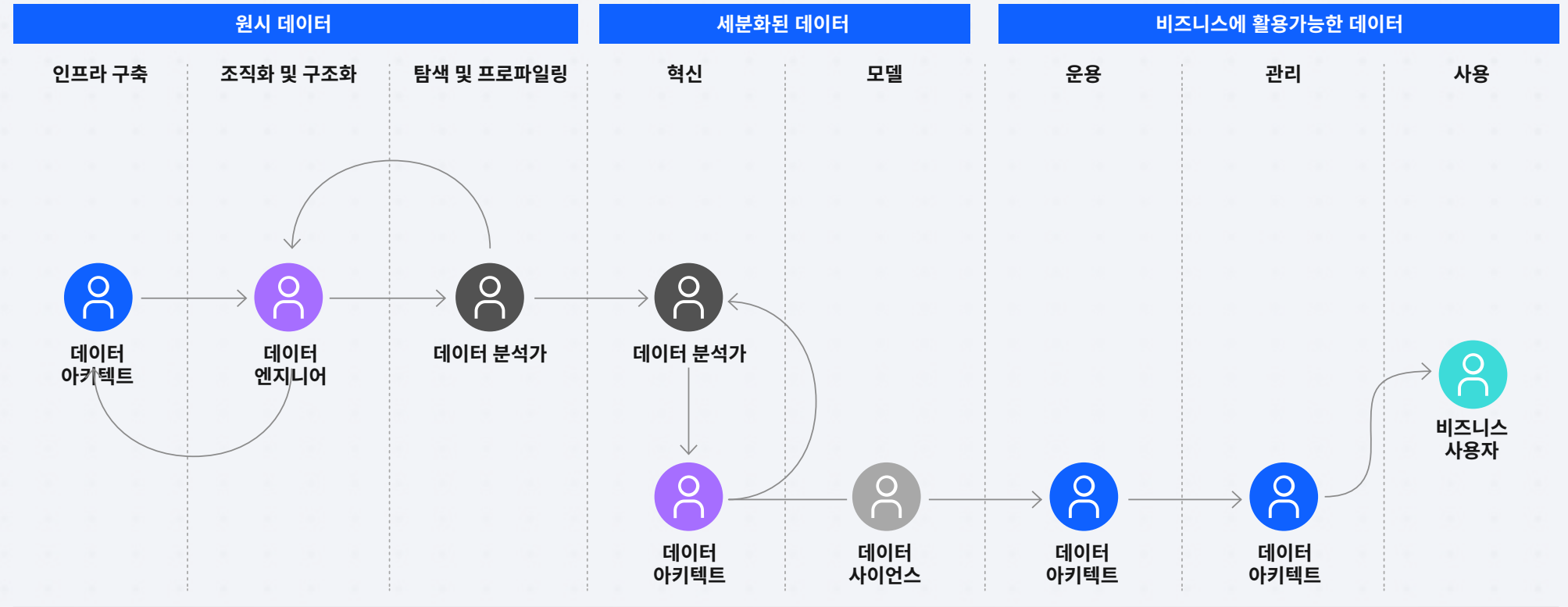


그림 2. 역할별 DataOps 워크플로의 예

# 프로세스: DataOps의 방법론

DataOps 프로세스는 조직에서 디지털 혁신 과제를 추진하면서 경험해온 다양한 문제를 해결할 수 있습니다. 바로 이 점이 86%의 기업에서 DataOps 투자를 늘리려는 이유이며,<sup>2</sup> 동시에 Gartner가 2019년 7월 Hype Cycle for Data Management에서 DataOps를 떠오르는 “유망” 기술로 지정한 이유입니다.<sup>3</sup> 효과적인 DataOps 프로세스의 중요성을 고려하여, IBM은 DataOps 실행 절차를 배포하고 비즈니스 전략에 맞춘 데이터 전략을 수립하는 데 있어서의 주요 장애물을 없앤 구현 방법론을 마련했습니다.

IBM은 DataOps 구현을 6단계로 나누었으며, 각각의 단계는 모두 기업의 데이터 문화를 바꾸고 적절한 프로세스와 기술을 구현하는 데 초점을 두고 있습니다.

## 통찰

신뢰할 수 있는 데이터에 액세스하는 시간을 단축함으로써 얻을 수 있는 효과를 조사합니다. 데이터 거버넌스 및 데이터 통합 톨과 실행 절차에 대한 인벤토리를 수집합니다. 조직의 DataOps 성숙도를 지속적으로 평가하고 이를 비즈니스 목표에 맞추어 조율합니다.

## 코드

강력한 자동 테스트 제품을 버전 관리 시스템 및 소스 제어 관리(SCM)와 함께 사용합니다.

## 전달

머신 러닝, AI, 거버넌스 적용과 같은 데이터 기술을 통해 DataOps 프로세스와 워크플로를 자동화하고 추적 및 감사 등의 기능을 구현합니다.

## 실행

지속적인 통합과 배포를 수행하도록 Red Hat® OpenShift®와 같은 배포 플랫폼 및 이를 위한 개발 환경을 도입합니다.

## 관리

지속적이고 규칙적인 배포를 통해 애자일 개발 기술을 현업에 도입합니다.

## 학습

적절한 커뮤니케이션 톨 및 자동 통지 기능으로 소통 및 프로세스 관리를 강화합니다.



주요 비즈니스 목표가 결정되면 해당 목표를 추진할 하나 이상의 프로젝트를 개시하기에 앞서 데이터의 범위를 지정해야 합니다. DataOps 프로세스는 다음과 같은 핵심 질문에 대한 답을 제시하기 때문에 데이터 범위 지정에 유용하게 활용할 수 있습니다. 데이터를 알고 있습니까? 데이터를 신뢰합니까? 오류를 신속하게 감지할 수 있습니까? 전체 데이터 파이프라인을 '손상'시키지 않고 점점 더 많은 변화를 이끌어 낼 수 있습니까?

뿐만 아니라, DataOps는 단일 뷰를 통해 보통은 서로 다른 출처에 고립되어 있는 애플리케이션에 저장되고 또한 중복되어 있을 중요 데이터 자산을 모두 파악할 수 있습니다. 또한 마스터 데이터 엔터티, 이들의 상호 관계 및 계층구조에 대한 일관된 정보도 제공합니다.

AI 데이터 중심 추진 과제를 포함한 자동화된 데이터 프로세스를 함께 사용해 조직에서 얻을 수 있는 이점은 다음과 같습니다.

- 분석 및 AI를 견인하며 비즈니스에 활용 가능한 데이터 제공
- 운영 효율성 및 셀프 서비스 데이터 액세스 확보
- 더욱 안전하게 데이터 개인 정보 보호 및 규정 준수 가능



# 기술: AI 및 자동화 구축

조화롭게 작동하는 데이터 시스템이 구현된 조직은 미래의 기회를 포착하는 동시에 운영, 성장 또는 통찰 확보를 저해하는 현재의 어려움에 대응할 장치를 갖춘 것과 같습니다. DataOps를 위한 전용 기술은 성공적인 DataOps 전략의 핵심 요소입니다. 올바른 DataOps 기술은 엔터프라이즈 데이터로 인해 발생하는 일반적인 4가지 문제를 극복할 수 있도록 해줍니다.

## 01

저품질 데이터를 카탈로그 작성 및 메타데이터 관리를 비롯한 큐레이션 조치와 거버넌스 그리고 품질을 통해 개선해야 합니다.

## 02

고립된 데이터를 가상화, 통합, 복제를 통해 연결해야 합니다.

## 03

신뢰할 수 없는 데이터를 마스터 데이터 관리를 통해 해결해야 합니다.

## 04

복잡한 데이터를 셀프 서비스 분석 및 테스트가 가능하도록 자동 데이터 준비를 통해 간소화해야 합니다.



# 01

## 거버넌스, 품질, 큐레이션을 통한 데이터 품질 향상

조직에서 여러 출처로부터 서로 다른 유형의 데이터가 넘쳐나는 경우, 해당 데이터가 무엇을 의미하는지, 무엇을 얻어낼 수 있는지, 그리고 무엇보다 어떤 책임과 결과를 불러올지 추적하기가 어려워 집니다. Gartner에 따르면 데이터 품질은 “신뢰할 수 있는 데이터로 비즈니스 혁신을 가속화하는 동시에 규제 요건을 반드시 충족함으로써 운영 위험을 낮추고 효율성을 높이기 위한 성공 요인”입니다.<sup>4</sup> 데이터를 수동으로 카탈로그화하기 위해 고군분투해본 사람이라면 저품질 데이터의 문제를 잘 알고 있습니다. 데이터에 대한 관점이 서로 다르고 데이터를 수동으로 정리하는 데 많은 시간을 소비할 경우 데이터의 생산성이 저하되어 다음과 같은 문제가 발생합니다.

- 낮은 품질의 교육 또는 운영 데이터로 인해 실패한 AI 이니셔티브
- Apache Hadoop 또는 기타 데이터 레이크 솔루션 마이그레이션을 진행할 때 잘 관리되지 않아 실패한 데이터 레이크로 인해 수동 코딩 증가
- GDPR, CCPA 및 Basel Banking Supervision의 표준번호 239(BCBS 239)와 같은 규정 준수 압력 급증

비즈니스는 이러한 어려움을 해결하기 위해 모든 데이터 유형의 카탈로그 작성, 보호 및 거버넌스, 자동 용어 지정 지원 및 데이터 리포지토리 관리를 위한 머신 러닝 기능을 필요로 합니다. 올바른 기술을 갖춘다면, 이러한 모든 기능이 성공적인 DataOps 구현의 일부가 될 수 있습니다.

DataOps 기술을 [2019 Forrester Wave](#)에서 최우수 엔터프라이즈 통찰 플랫폼으로 인정받은 [IBM Cloud Pak for Data](#)와 같은 플랫폼을 기반으로 구현하는 것이 이상적입니다. Cloud Pak for Data는 하이브리드 및 멀티클라우드 지원, 개방형 메타데이터 지원, 데이터 과학 및 AI 툴과의 간소한 통합, 엔터프라이즈급 확장성을 통해 적응형 배포를 가능하게 해 줍니다. 또한 Cloud Pak for Data는 업계 최고인 Red Hat OpenShift 플랫폼을 기반으로 구축되어 완전히 컨테이너화된 플랫폼이기 때문에 인프라 부담이 큰 폭으로 감소합니다. 뿐만 아니라 [IBM Watson® Knowledge Catalog](#)와 같은 서비스가 포함되어 데이터 카탈로그화의 속도와 정확성 및 거버넌스가 향상되기 때문에 AI 및 머신 러닝을 대규모로 배포하는 데 매우 효과적입니다. 이러한 기능 덕분에 Watson Knowledge Catalog는 Gartner로부터 [Customers' Choice 2020](#)을 수상하는 동시에 [Forrester Wave](#)로부터 리더로 선정되었습니다.

데이터 거버넌스, 데이터 품질 및 활성 정책 관리 기능을 자동화하고 통합하면 조직에서 더욱 효과적으로 민감한 데이터를 보호 및 관리하고 데이터 계보를 추적하고 데이터를 관리할 수 있습니다. 그 결과로, 조직은 전사적으로 신뢰할 수 있고 품질이 우수하며 연속성이 보장되는 운영 데이터를 보유하게 됩니다.

### 무료 체험판

[IBM Cloud Pak for Data](#)

[Watson Knowledge Catalog](#)

# 02

## 가상화, 통합, 복제를 통해 단절된 데이터 연결

데이터가 단절되면 통찰을 얻는 데 제한이 발생합니다. 데이터와의 연결이 끊기면 예를 들어 쇼핑 고객의 기존 거래 데이터와 소셜 미디어의 정서 및 실시간 클릭 검색 동작을 결합해 얻어지는 데이터 등 다양한 종류의 데이터가 제공하는 추가적인 뉘앙스, 즉 미묘한 차이를 놓치게 됩니다. 데이터의 단절은 조직이 여러 사업 단위 구조로 이루어져 있고 서로 다른 플랫폼을 사용하며 업무 수행 장소도 다양하기 때문에 자연스럽게 발생하는 경우가 많습니다. 또한 멀티클라우드 아키텍처로 인해 더 많은 단절이 생기기도 하고 데이터 볼륨, 거버넌스, 일반 제어 워크플로의 부족 및 통일된 보안 및 관리 필요성과 관련한 여러 통합 문제점이 발생할 수 있습니다. 동시에, 처리해야 하는 데이터 볼륨은 빠르게 증가하는 반면 IT 예산은 늘릴 수 없거나 오히려 줄어드는 상황에 대처해야 합니다. 즉, 단절을 스마트하게 연결하지 않으면 통찰에 제약이 발생하고 귀중한 시간과 리소스가 낭비됩니다.

하지만 임베디드 머신 러닝과 함께 적절한 데이터 가상화, 통합, 복제 기능을 활용하면 이러한 어려움을 극복할 수 있습니다. 가상화는 IBM의 DataOps 접근법, 그 중에서도 특히 Cloud Pak for Data 플랫폼과 **IBM® Db2®** 제품군에 속한 전체 데이터 관리 제품의 기반입니다. 가상화는 데이터 페더레이션에 추상화 계층을 결합한 것으로, 사용자와 애플리케이션이 기반 데이터의 종류, 형식, 크기 또는 위치와 관계없이 단일 액세스 지점에서 동시에 여러 데이터 소스와 상호작용할 수 있도록 해줍니다. 가상화가 단절된 데이터를 통합하는 강력한 수단이지만 **IBM InfoSphere® DataStage®**와 같은 통합 솔루션을 구현하면 유효성 확인 규칙을 수집 프로세스의 일부로 자동 실행하고, 로그 기반 복제 및 배치 처리와 같은 기존 데이터 전달 스타일을 지원하고, 실시간 데이터 동기화 및 라인 데이터 품질 내에서 스트림 데이터 통합 등의 복잡한 데이터 전달 스타일을 지원하며, 정책 기반 데이터 액세스를 위한 메타데이터 규칙을 자동화하여 그 영역을 확장할 수 있습니다. 조직에서 이 솔루션이 제공하는 풍부한 기능과 사전

구축된 다양한 기능 및 커넥터를 활용하면 가치 실현 시간을 단축하고 IT 위험을 줄이며 데이터 레이크, 데이터 웨어하우스, 데이터 마트 및 PoI(point-of-impact) 솔루션에서 올바른 데이터를 사용할 수 있도록 보장할 수 있습니다.

데이터를 연결하면 비즈니스에서 다음과 같은 효과를 얻을 수 있습니다.

- 시간과 장소에 구애받지 않고 각종 클라우드에서 거의 실시간으로 구조화 및 비구조화 데이터를 액세스 및 신뢰
- 변경 데이터 캡처 기술 및 스트림 데이터 통합으로 데이터 복제, 배치 처리 및 이벤트 기반 실시간 캡처를 비롯한 다양한 데이터 전달 스타일을 통합하여 본질을 꿰뚫는 실시간 분석 지원
- 다양한 IT 분야의 강력한 공유 및 협업을 지원하는 셀프 서비스 기능 배포
- 애플리케이션을 한 번 작성하여 기업의 모든 데이터에 대해 실행

IBM DataOps 플랫폼은 탁월한 성능과 확장성, 그리고 실시간 업데이트를 제공합니다. 이러한 모든 기능을 지원하는 DataOps는 여러 사이트에서 멀티클라우드 워크로드 배포 및 연속적인 가용성에 이상적입니다.

### 추가 리소스

[DataStage 데모 가이드](#)

[IBM Db2 자세히 알아보기](#)

# 03

## 마스터 데이터 관리를 통해 신뢰할 수 없는 데이터 해결

데이터가 올바르게 관리되고 동시에 모든 데이터에 액세스할 수 있다고 해도 무결성을 신뢰할 수 없다면 무용 지물입니다. 신뢰할 수 없는 데이터는 거버넌스가 우수한 데이터 시스템에도 존재할 수 있으며 이를 해결하려면 구체적인 전략이 필요합니다. 각종 캡처 애플리케이션, 접점, 인수 합병 등으로 인해 많은 기업에서 데이터 세트의 홍수를 경험하고 있습니다. 이러한 데이터는 체계적이지 못하고 단편적이며 중복되어 기업에서 고객, 제품 또는 에코시스템을 파악하려고 할 때 병목 현상을 일으킵니다.

데이터 자산에 대한 신뢰할 수 있는 단일 뷰를 구축하고 비즈니스 분석가, 마케팅 전문가, 정보 설계자가 데이터에 대한 고유한 신원을 파악하고 종합적으로 이해하기 위해서는 마스터 데이터 관리 솔루션이 필요합니다. 이러한 단일 뷰를 통해 고객이 어떠한 옴니채널을 사용하는지 여부와 상관없이 적시에 적절한 비즈니스 사용자에게 올바른 데이터를 제공함으로써 고객 서비스와 관계를 최적화할 수 있습니다.

[Master Data Management on Cloud](#) 및 [InfoSphere Master Data Management](#)를 비롯한 IBM 마스터 데이터 관리 제품군은 IBM DataOps, 마이크로서비스 및 API 사이에서 머신 러닝을 폭넓은 데이터 거버넌스 및 규정 준수 전략과 통합합니다. 그 결과, 고객, 제품 및 거래처에 대한 깊은 이해를 통해 조직의 총 수익이 증가하고 데이터 관리와 연관된 비용과 위험 감소로 순 수익에도 영향을 미칩니다.

### 추가 리소스

[마스터 데이터 관리 제품 알아보기](#)

[데모 가이드: InfoSphere Master Data Management](#)

# 04

## 준비 자동화를 통해 복잡한 데이터 간소화

DataOps 기술의 마지막 장점은 셀프 서비스 분석 및 테스트를 위한 데이터 준비 간소화 및 자동화입니다. 오늘날의 조직은 내부 운영을 관리하고 모든 데이터와 관련된 규정 준수를 유지 관리할 뿐만 아니라 관련 애플리케이션과 툴을 그 어느 때보다 신속하게 개발, 정제, 배포 및 개선해야 한다는 부담에 시달리고 있습니다.

따라서, 이러한 조직에서는 효율성 향상, 비용 절감 및 경쟁력 확보를 위해 애플리케이션 현대화에 많은 투자를 아끼지 않습니다. 개발 후에 수작업을 통해 테스트를 진행하던 많은 수의 기업이 이제는 테스트와 개발이 동시에 이루어지는 애자일 방법을 도입했습니다. 하지만 이로 인해 다음과 같은 프로세스 문제가 발생합니다.

테스트 시간을 줄여 시장 출시 시간을 단축해야 합니다. 조직들은 이 문제를 해결하기 위해 자동화 및 다양한 테스트 데이터 관리 절차를 찾고 있습니다.

개발 팀은 오류를 줄이고 신뢰성과 속도를 높이기 위해 연속적인 전달 모델에 의존합니다. 이러한 변화를 해결하고자 조직들은 애자일 작업 흐름을 지원하는 업체 및 솔루션을 필요로 합니다.

[InfoSphere Optim™ Test Data Management](#)와 같은 IBM 솔루션은 온 프레미스 또는 클라우드에서 테스트 데이터 환경에 대한 프로비저닝 및 새로 고침을 수행하여 클라이언트가 더 짧은 시간에 더 우수한 품질의 애플리케이션을 배포할 수 있도록 지원합니다.

뛰어난 DataOps 플랫폼은 대규모의 원시 데이터를 양질의 소비 가능한 정보로 신속하게 변환할 수 있도록 자동화 기능을 갖추고 있습니다. 모든 출처로부터 적시에 올바른 대상에게 연결되는 셀프 서비스 데이터 파이프라인이 조직에 구축되어 있다면 보고, 분석 또는 테스트 요구사항에 맞는 신뢰할 수 있는 고품질 데이터를 제공할 수 있습니다. 데이터 준비 기능을 DataOps 절차에 맞게 통합하면 손쉽게 자동화를 추가하여 데이터 여정의 마지막 단계를 손쉽게 달성할 수 있습니다.

### 추가 리소스

[InfoSphere Optim Test Data Management 데모 가이드](#)  
[InfoSphere Advanced Data Preparation 자습서](#)

# DataOps 사용자가 거둔 성과

95%

메타데이터 속도 향상

20일 → 24시간

처리 소요 기간

# DataOps 실행의 가치

DataOps의 가치에 대해 고품질 데이터에 더 빨리 액세스하여 비즈니스에 혜택을 가져다 준다고 막연하게 이야기 하기 쉽습니다. 하지만 IBM은 최고의 기업들과 오랜 기간 함께 해왔기 때문에 다양한 업종에서 DataOps의 혁신적인 효과를 확인할 수 있는 실제 구현 사례를 제시할 수 있습니다.





## 통찰 가속화

예를 들어, **미국 현지 소매업체**는 최근 300개 매장에서 가용성, 주문 처리 및 온라인 장바구니를 포함한 공급망 관리를 현대화하기 위해 Google Cloud Platform(GCP)에 **Onera Decision Engine**을 구현하기로 결정했습니다. Onera는 AI를 통해 실제 행동을 예측하고 미묘한 차이를 반영한 통찰과 의사 결정을 생성하여 최신 상거래 공급망을 지원하는 코그니티브 운영 체제입니다.

하지만 이 Onera 엔진이 올바르게 작동하려면 소매업체가 시간당 9백만 ~ 2천 백만 개의 메시지를 제공해야 하며 이를 위해 온 프레미스 Oracle 시스템에서 데이터를 추출하여 이를 GCP로 옮긴 후에 낮은 지연율과 실시간에 가까운 성능을 유지하면서 해당 데이터를 GCP 게시/구독 서비스에 게시해야 했습니다. 이 과정이 쉽지 않았기 때문에 Talend 플랫폼을 사용하는 이전의 시도에서는 GCP로 게시하는 메시지가 시간당 120만 개를 넘지 못했습니다.

이 문제를 해결하기 위해 IBM DataOps 팀에서는 DataStage에 Python 작업을 구축하여 이러한 목표를 초과 달성했을 뿐만 아니라 필요 시에는 더 많은 확장성을 지원할 수 있게 되었습니다. DataOps 팀은 또한 고객의 문제를 해결하는 프로토타입 솔루션을 자신들이 얼마나 신속하고 민첩하게 개발할 수 있는지에 대해서도 직접 시연했습니다. 이들의 경험은 IBM DataOps 방법론을 사용하는 효과적인 데이터 플랫폼이 획기적인 처리량을 지원함으로써 모든 클라우드에서 AI 기반 분석 애플리케이션에 데이터를 제공하고 실시간 통찰을 얻을 수 있음을 보여 줍니다.



## 프로세스 개선

또 다른 사례로, 60개국에서 근무하는 4만 명의 직원들이 매일 쏟아내는 폭발적인 데이터 볼륨을 처리하고 분석하는 데 어려움을 겪고 있는 **유럽의 유명 소매업체**를 살펴 보겠습니다. 이 기업은 기존에 보유 중인 분석 인프라로는 필요한 통찰을 원하는 속도와 주기로 얻을 수 없었습니다. 하지만 **IBM BigIntegrate**, **IBM BigQuality** 및 **Watson Knowledge Catalog**를 활용한 빅데이터 통합, 품질 및 거버넌스 솔루션을 구현하자 전보다 훨씬 빠른 속도로 고품질 데이터를 분석 환경으로 전달할 수 있게 되었습니다.

솔루션 구현의 결과는 드라마틱했습니다. 회사 대변인의 말에 따르면 기존에는 최대 20일이 소요되던 데이터 처리 작업이 이제는 단 24시간이면 가능합니다. 이는 구매가 끝난 후에도 고객의 재방문을 유도할 수 있도록 기존 고객의 구매가 예상되는 품목을 예측하는 고객 관련성 분석을 신속히 수행하는 데 유용하게 활용됩니다.

이러한 동일한 데이터 플랫폼 개선을 통해 이 소매업체는 3주가 아니라 단 몇 분만에 **소스 시스템에서의 영향 분석(데이터 소스 변경이 다운스트림 애플리케이션에 영향을 미치는지 여부 판별)**을 수행하고 하루 종일이 아니라 4시간 만에 일별 재고 포지션을 계산할 수 있으므로 각 매장에 올바른 제품 재고를 유지할 수 있습니다.

회사는 “우리 회사가 Hadoop용 IBM 툴을 이 조합으로 선택한 전 세계 최초의 기업입니다. IBM 팀은 당사 비즈니스 인텔리전스 팀과의 긴밀한 협업을 통해 실행 가능한 비즈니스 사례를 구축하고 최고의 교육을 제공해 주었기 때문에 저희는 굳은 확신과 마음의 안정을 느낄 수 있었습니다.”라고 언급했습니다.

궁극적으로, 이 DataOps 구현은 회사가 모든 매장의 선반에 올바른 제품을 계속 진열하는 데 필요한 통찰을 확보하고 합리적인 가격으로 제품을 제공함으로써 고객 만족과 브랜드 충성도라는 두 마리 토끼를 모두 잡는다는 것을 의미합니다.



## 비용 및 법적 책임 감소

약간 다른 계기로 접근한 사례도 있습니다. **Fortune 선정 50대 글로벌 은행**은 ETL 워크로드를 Teradata와 AIX®에서 Hadoop으로의 마이그레이션을 통해 큰 폭의 비용 절감을 실현하고자 했습니다. IBM DataOps 방법론에 따른 통합 기반 덕분에 이 은행은 코드를 한 줄도 다시 작성하지 않고 18,000개 이상의 작업을 마이그레이션할 수 있었습니다. 마이그레이션된 작업의 99% 이상이 첫 번째 테스트부터 완벽하게 실행되었고 가장 규모가 큰 작업도 새 인프라에서는 5배 빨리 실행되었습니다. 덕분에 저장 공간 및 컴퓨팅 비용 수백 만 달러가 절감되었습니다. 은행 관리자 중 한 명은 모든 과정과 결과를 경험한 후에 "BigIntegrat는 정말 대단한 제품입니다!"라고 말했습니다.

DataOps 방법론 및 툴은 기존 데이터에 대한 이해도를 높여 예측하지 못한 법적 책임을 식별하고 이를 바로잡는 데 도움을 주는 효과도 있습니다. IBM Information Analyzer를 사용하여 이미 삭제되어야 하는 PHI(Protected Health Information)와 PII(Personally Identifiable Information)가 포함된 데이터 테이블을 검사한 **미국 유명 헬스케어 조직**의 사례를 살펴보겠습니다. 이 조직은 해당 테이블에서 모든 PHI/PII를 제거하도록 명시된 법적 요구사항에 대한 규정 준수를 확인하고자 했지만, 검사 결과 21개 테이블 중 15개(71%)가 올바르게 삭제되지 않아 PII가 여전히 남아 있음을 알게 되었습니다. 이러한 부주의로 인해 막대한 법적 책임을 지게 될 가능성이 있었으며 IBM 솔루션은 그 가치를 제공함으로써 사용을 시작한지 2시간 만에 ROI를 달성했습니다.



## 개입을 없애 효율성 발견

IBM은 DataOps 원칙을 **전세계 공급망 관리**와 같은 내부 데이터 프로세스에도 적용했습니다. 수십 개국 수백 개 공급업체가 포함된 IBM 공급망은 운영 중단 발생 가능성이 있는 수천 가지 변수들의 영향을 받습니다. IBM은 이러한 가동 중단을 예측하고 구성요소의 위치를 추적하며 빠르게 변화하는 이벤트가 비즈니스에 미칠 영향까지도 예측하는 정교하고 정밀한 분석을 개발했습니다. 하지만, 이러한 통찰은 AI를 적용한 DataOps 플랫폼을 통해 어마어마하게 다양한 소스로부터 제공되는 신속하고 신뢰할 수 있는 데이터가 있었기 때문에 얻을 수 있습니다.

이러한 성공 사례는 효과적인 AI 활용을 비롯하여 DataOps 방법론을 철저히 적용하면 실제 비즈니스에서 막대한 효과를 가져다 줄 수 있음을 보여줍니다. 하지만 DataOps 프로세스와 기술의 선별적인 적용

만으로도 적지않은 효과를 얻을 수 있습니다. 예를 들어, 미국의 통신사, **재무 서비스 공급업체** 그리고 **보험사**는 각각 DataOps 툴로 데이터 준비를 자동화하자 메타데이터 용어집 생성 시간이 95%, 그리고 메타데이터를 찾아 용어를 지정하는 시간이 90%까지 단축되었습니다.

그밖에, DataStage 입력을 사용하여 수십만 SKU를 추적하고 일별 소진 예상치를 업데이트하는 재고 예측 시스템 덕분에 절감 효과를 누리게 된 **글로벌 소매업체**의 사례도 있습니다. 이렇게 다양한 사례에서 알 수 있듯이 DataOps는 올바른 데이터를 올바른 담당자에게 혹은 올바른 프로세스로 적시에 전달함으로써 가치를 제공합니다.



# IBM DataOps Center of Excellence(CoE)

# DataOps 시작하기

DataOps가 필요한 이유는 간단합니다. 믿을 수 있고 완전하며 일관성 있는 데이터가 없으면 분석이나 AI에 확신을 가질 수 없기 때문입니다. 처음 도입을 고려 중이시라면, IBM이 전문성, 서비스, 기술이 집약된 제품들을 오랜시간 다양한 기업에 제공해 온 전통있는 업체인 동시에 분석 통찰과 데이터 현대화에 필요한 완벽한 정보 관리 및 거버넌스를 제공할 수 있는 유일한 DataOps 공급업체임을 기억해 주시기 바랍니다.

IBM DataOps Center of Excellence(CoE)와 함께라면 강력한 DataOps 전략을 구축하고 특화된 방법론, 리소스, 기술로 설계된 독보적인 DataOps 프로그램의 장점을 누릴 수 있습니다. 이러한 독점 리소스를 활용하여 전문가가 고객과 함께 협력해 비즈니스 목표를 바탕으로 DataOps 접근 방식을 개별 조율하고 경영진을 설득할 가치를 창출하는 올바른 파일럿 프로젝트를 식별합니다.

DataOps CoE는 새로운 기능을 지속적으로 설계, 제공 및 검증하기 위한 비즈니스, 개발 및 운영을 가능하게 하는 접근 방식인 **IBM Garage™**를 활용합니다. DataOps Discovery Workshop에서 이 Garage 프레임워크를 사용하여 조직의 DataOps 준비도 및 전사적 필요성을 평가해볼 수 있습니다. 협업 업무 세션에서는 팀별로 데이터 파이프라인 개발, 업데이트, 관리와 관련해 엄선된 비즈니스 목표를 채택하고 중점 영역을 식별한 후에 구체적인 다음 단계를 개발합니다.

## 4시간 Discovery Workshop 어젠다

● 연합 활동   ● 고객 주도   ● IBM 주도

8:30–8:45	● 소개
8:45–9:30	● <b>비즈니스 목표 및 원하는 결과</b> – 비즈니스 측면: 단기 및 장기적 관점, 우선 순위 – 기술 측면: 단기 및 장기적 우선 순위 – 목표 달성을 위한 주요 “필수” 요구사항
9:30–10:15	● <b>“현재”</b> – 사용자, 데이터 소스, <b>데이터 흐름</b> , 시스템, 종속 요소 – 목표에 맞추어 기회 및 해결 과제(차이) 조정 – 조직 및 팀 구조, 이해관계자
10:15–10:45	● <b>사용 사례 및 쟁점</b>
10:45–11:30	● <b>“미래”</b> – 비즈니스 목표의 맥락에서 원하는 미래 역량 및 지표 – 비즈니스 목표 조정 – 데이터, 기능, 구성요소
11:30–12:00	● <b>실천 항목 / 다음 단계</b>

그림 3. 반나절 일정의 DataOps Discovery Workshop 샘플 어젠다

# 모든 유형의 클라우드에서 통합 기술 스택 구축

## 데이터 품질 향상

IBM은 포괄적인 데이터 수집 및 발견을 위해 고급 머신러닝 및 더욱 스마트한 메타데이터가 통합 카탈로그 생성을 지원하는 유일한 공급업체입니다.

**관련 제품:**

[Watson Knowledge Catalog](#)  
[IBM Cloud Pak for Data](#)

## 사일로된 데이터 연결

서로 다른 플랫폼 요소를 통합하면 애플리케이션을 한 번 구축해 온 프레미스 또는 클라우드에 설치된 Apache Hadoop, Red Hat OpenShift, IBM Cloud Pak for Data에서 실행할 수 있습니다.

**관련 제품:**

[InfoSphere DataStage](#)  
[InfoSphere Information Server](#)  
[IBM Db2](#)

## 신뢰할 수 없는 데이터 해결

실시간에 가까운 검색 결과와 함께 정확도가 높은 매칭 엔진을 활용하고 이를 통해 얻은 통찰을 800개 이상의 업종별 모델을 통해 비즈니스에 맞게 조정합니다.

**관련 제품:**

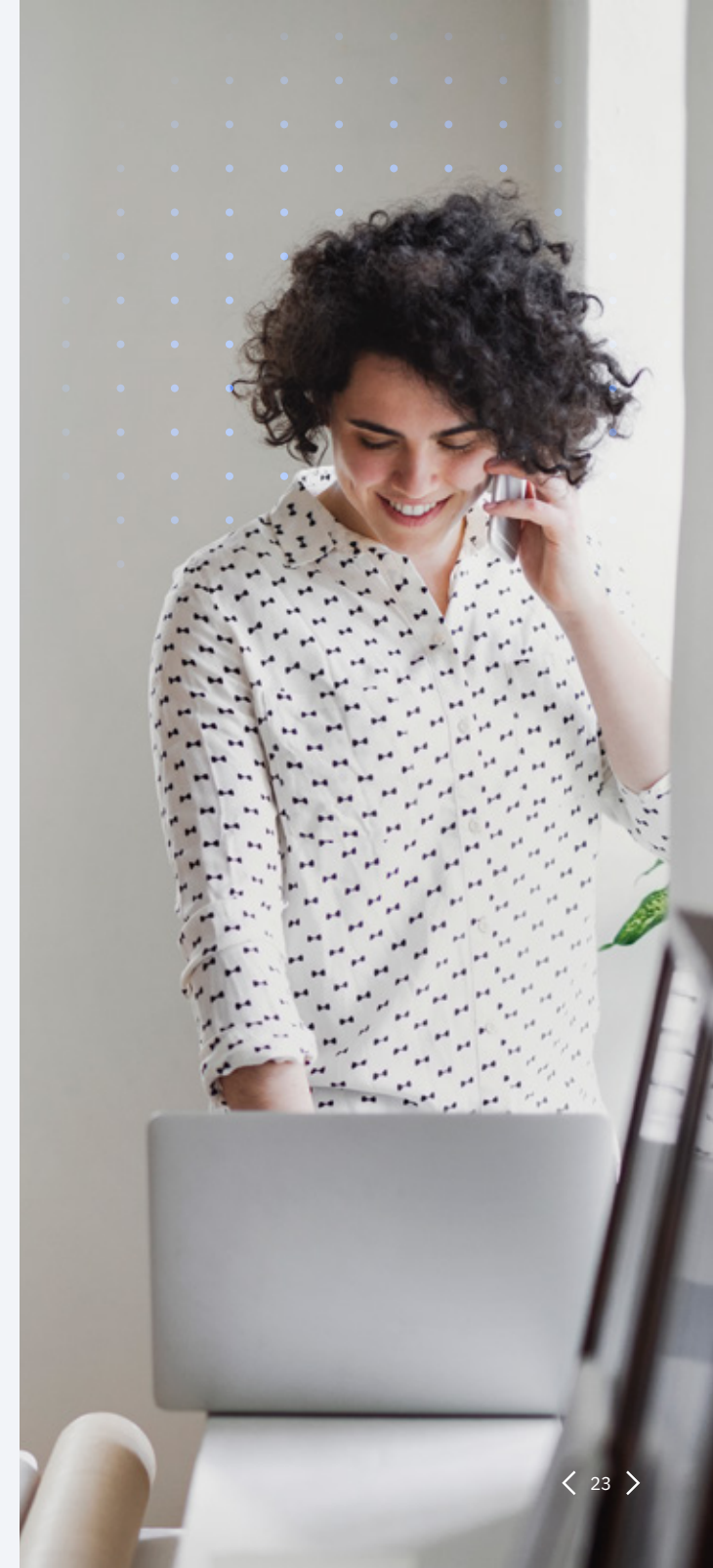
[InfoSphere Master Data Management on Cloud](#)  
[InfoSphere Master Data Management](#)

## 복잡한 데이터 간소화

데이터 준비 툴 확장을 통해 전사적 분석 요구사항을 충족하고 보안이 철저한 테스트 데이터를 위한 강력한 마스킹 및 데이터 생성 툴을 제공합니다.

**관련 제품:**

[InfoSphere Optim Test Data Management](#)  
[InfoSphere Advanced Data Preparation](#)



# 다음 단계

클라우드, 온 프레미스, 하이브리드 환경에서 구조화 및 비구조화 데이터를 통합, 분석 및 관리할 수 있는 솔루션을 갖춘 기업은 각종 기술, 프로그램, 방법론 및 모범 사례를 활용하여 더욱 짧은 시간에 데이터의 민주화와 현금화를 실현하고 기회를 창출할 수 있습니다. 올바른 데이터를 필요한 곳에 신속히 제공하는 것이 바로 DataOps입니다.

IBM DataOps Center of Excellence와 함께 지금 바로 시작하십시오.



IBM DataOps 전문가와 무료 일대일 컨설팅 예약하기





IBM Corporation  
New Orchard Road  
Armonk, NY 10504

Produced in the United States of America  
2020년 7월

IBM, IBM 로고, ibm.com, IBM Cloud Pak, IBM Watson, Db2, InfoSphere, DataStage, Optim, AIX 및 IBM Garage는 전 세계에 등록되어 있는 International Business Machines Corp.의 상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표일 수 있습니다. 최신 IBM 상표 목록은 웹 "저작권 및 상표 정보(www.ibm.com/legal/copytrade.shtml)"에 있습니다.

Red Hat® 및 OpenShift®는 미국 및 기타 국가에서 Red Hat, Inc. 또는 그 자회사의 상표 또는 등록상표입니다.

이 문서는 처음 발행될 당시의 날짜를 기준으로 업데이트되었으며 IBM은 언제든지 문서 내용을 변경할 수 있습니다. IBM이 사업을 운영하는 국가라도 일부 제품은 공급되지 않을 수 있습니다.

여기서 소개된 성능 데이터와 고객 사례는 오로지 예시를 목적으로 작성된 것입니다. 실제 성능 결과는 특정 구성 및 운영 환경에 따라 달라질 수 있습니다. 본 자료에 발췌 또는 수록된 고객 사례는 일부 고객사가 IBM 제품을 사용한 결과 긍정적인 성과가 있었다는 맥락으로 예시된 것입니다. 실제 환경 비용 및 성능 특징은 개별 고객의 구성 및 조건에 따라 다릅니다. 어떤 도움을 받을 수 있는지 궁금하시면 IBM으로 문의해 주십시오. IBM 제품 및 프로그램과 함께 사용되는 기타 제품 또는 프로그램을 평가 및 검증하는 것은 사용자의 책임입니다. 이 문서의 정보는 상품성에 대한 보증, 특정 목적의 적합성 여부 및 저작권을 침해하지 않는다는 보증 또는 조건을 포함해 명시적 또는 암묵적 보증 없이 "있는 그대로" 제공됩니다. IBM 제품은 제공된 약정에 명시된 조항 및 조건에 따라 보증됩니다.

고객은 관련 법령과 규정을 반드시 지켜야 할 책임이 있습니다. IBM은 법률 상담을 제공하지 않으며 IBM 서비스 또는 상품이 고객의 법령 또는 규정 준수를 보장한다고 주장하거나 보증하지 않습니다.

모범 보안 사례문: IT 시스템 보안은 예방, 탐지 및 기업 내부와 외부로부터의 부정 접근에 대한 대응을 통해 시스템과 정보를 보호하는 활동에 관여합니다. 부정 접근은 정보의 변개, 파괴, 오용 또는 남용을 초래하거나 사용자 시스템이 다른 시스템에 대한 공격에 이용되는 것을 포함하여 오용 또는 손상을 주는 결과를 가져올 수 있습니다. 어떠한 IT 시스템이나 제품도 완전하게 안전하다고 간주하지 않아야 하고, 어떠한 단일 제품, 서비스 또는 보안 대책도 잘못된 사용 또는 액세스를 완전히 효과적으로 방지할 수 없습니다. IBM 시스템, 제품 및 서비스는 포괄적인 보안 접근 방식의 일부로 설계되어 있으므로 추가적인 운영 절차에 참여하거나 최상의 효과를 위해 다른 시스템, 제품 또는 서비스를 필요로 할 수 있습니다. IBM은 시스템, 제품 또는 서비스가 악의적이거나 불법적인 행위로부터 영향을 받지 않는다는 것을 보증하지 않으며 귀사가 이러한 행위로부터 영향을 받지 않음을 보증하지 않습니다.

- 01 Gartner press release, 25 April 2018. <https://www.gartner.com/en/newsroom/press-releases/2018-04-25-gartner-says-global-artificial-intelligence-business-value-to-reach-1-point-2-trillion-in-2018>
- 02 451 Research, "DataOps Lays the Foundation for Agility, Security and Transformational Change" February 2019
- 03 <https://www.gartner.com/en/documents/3955768/hype-cycle-for-data-management-2019>

APM14B8J