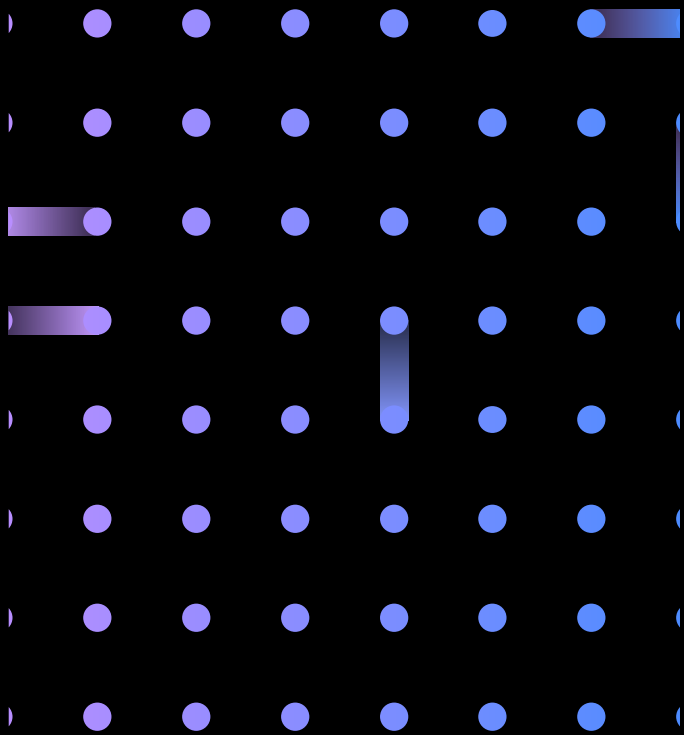


# DataOps로 비즈니스에 활용 가능한 데이터를 빠르게 제공

IBM DataOps 방법론 및 사례 소개



# 목차

소개	3
DataOps 정의	3
목표 비교	4
IBM DataOps 프로그램	7
DataOps를 충족하는 IBM Cloud Garage 방법론	7
성공적인 DataOps 사례의 영향	10
결론	11
부록: DataOps 파일럿 프로그램 템플릿	12

# 주요 내용

- DataOps는 데이터 사용자에게 신뢰할 수 있는 고품질 데이터를 빠르게 제공하기 위해 인력, 프로세스와 기술이 조화를 이룬 산물입니다.
- 자동화가 뒷받침하는 DataOps는 데이터에 대한 액세스, 데이터 준비 및 통합 그리고 데이터를 사용 가능하게 만드는 데 수반되는 비효율성과 관련된 문제를 해결합니다.
- IBM DataOps 사례는 극단적 자동화를 활용해 데이터 큐레이션 서비스, 메타데이터 관리, 데이터 거버넌스, 마스터 데이터 관리 및 셀프 서비스 상호 작용 등과 같은 기능에 눈에 띄게 큰 영향을 미칩니다.
- IBM에서는 규범적 방법론, AI 지원 자동화 및 IBM DataOps Center of Excellence를 통해 DataOps 사례를 구축 방안을 제시합니다.
- DataOps 워크숍은 DataOps 로드맵의 필수적인 부분으로, 조직에서 자체 DataOps 성숙도를 평가하고 파일럿 프로젝트 실행을 계획하도록 지원합니다.

## 소개

데이터는 혁신 및 경쟁 우위 지속에 있어 연료와 같습니다. 분석하고 비즈니스 추세 및 기회를 파악하는 핵심입니다. 새로운 방식으로 데이터의 가치를 찾아내면 조직의 AI 여정에 더욱 속도를 낼 수 있습니다.

그러나 데이터 관련 프로젝트에서 약속한 투자수익률(ROI)을 내지 못하면 주주들은 그 이유를 알고 싶어합니다. Experian의 2019년 Global Data Management Research 보고서에 따르면 89%의 기업이 데이터 관리에 어려움을 겪고 있다고 보고했습니다. 이러한 기업이 겪는 문제로는 인사이트 지연과 기초 데이터에 대한 신뢰의 부족을 들 수 있습니다.<sup>1</sup>

조직의 비즈니스 목표를 파악하는 것은 분석 및 AI에 대한 효과적인 데이터 전략 개발에 매우 중요합니다. 비즈니스 모델이 작동하기 위해서는 고객사의 필요를 충족해야 합니다. 성공은 비즈니스에 활용 가능한 통합된 데이터 파이프라인으로 데이터 작업을 간소화하는지 여부에 달려 있습니다. 해당 파이프라인은 항상 일관되고 완전한 비즈니스 관점을 제공하는 역할을 합니다.

보다 빠른 결과 창출에 대한 기대치는 계속해서 커지고 있는 항목 중 하나입니다. 기업은 특히, 조직 내에 존재하는 많은 사일로 때문에 최상의 의사결정을 위해 운영 효율성 및 효과를 개선하기 위한 방법을 장소를 가리지 않고 찾고 있습니다. 이러한 두 가지 요소로 인해 비즈니스 리더들은 단일 프레임워크 내에서 가장 큰 문제를 해결하기 위한 새로운 방법을 찾게 됩니다.

데이터 운영 내에서 혁신을 모색하는 조직의 경우 자동화 기술이 경쟁력을 발휘할 수 있습니다. 비즈니스에 활용 가능하고 신뢰할 수 있는 데이터가 조직에 차별화된 인사이트 및 운영 효율성을 견인할 때 데이터는 큰 가치를 지닙니다.

이 백서의 목표는 DataOps 방법론, 사례 및 로드맵의 이점을 집중 조명하는 것입니다.

## DataOps 정의

데이터 운영(DataOps)은 데이터 사용자에게 신뢰할 수 있는 고품질 데이터를 빠르게 제공하기 위해 인력, 프로세스와 기술이 조화를 이룬 산물입니다. 이 방식은 민첩성, 속도 및 새로운 데이터 이니셔티브를 대규모로 견인하기 위해 조직 전반에서 협업을 가능하게 하는 부분에 중점을 둡니다. DataOps는 자동화를 사용하여 데이터에 대한 액세스, 데이터 준비 및 통합 그리고 데이터를 사용 가능하게 만들 때 수반되는 비효율성과 관련된 문제를 해결하도록 설계되어 있습니다.

DataOps가 지닌 잠재적인 이점을 꼽자면 개인에게 정보 및 데이터를 제공하고, 효율성과 최적화를 얻기 위해 프로세스를 개선하는 과정에서 얻을 수 있는 상당한 생산성 이점이 있습니다. AI 데이터가 주도하는 이니셔티브를 포함한 자동화된 데이터 운영은 다음과 같은 결과를 제공합니다.

- 대규모로 분석 및 AI를 견인하며 비즈니스에 활용 가능한 통합된 데이터 제공
- 운영 효율성 달성
- 데이터 개인 정보 보호 및 규정 준수 가능

# 89%

에 달하는 기업에서 데이터 관리에 어려움을 겪고 있습니다.<sup>1</sup> 조직의 비즈니스 목표를 파악하는 것은 분석 및 AI에 대한 효과적인 데이터 전략 개발에 매우 중요합니다.



# DataOps는 DevOps가 아닙니다

대다수 조직은 자체 개발 분야 내에서 어느 정도의 DevOps를 구현했습니다. DevOps 사례에 대한 광범위한 친숙성과 이름의 유사성 때문에 새롭게 등장한 DataOps 사례와 비교해 보았습니다. 양쪽 모두 운영 모범사례를 촉진하기 위한 방법론이기는 하지만 각각 조직 내에서 고유한 위치를 차지하고 있습니다.

아래 표에서 조직을 위한 목표 및 이점이라는 관점에서 두 가지 사례를 어떻게 비교할 수 있는지 살펴보십시오.

## 목표 비교

	DataOps	DevOps
중점 항목	신속하게 사용할 수 있는 비즈니스에 활용 가능하고 신뢰할 수 있는 고품질 데이터	애플리케이션 및 소프트웨어 개발
혁신 목표	<ul style="list-style-type: none"> <li>- 모든 데이터 사용자를 위해 신뢰할 수 있는 고품질 데이터에 대한 셀프 서비스 액세스를 가능하게 하여 기업의 지속적이고 빠른 혁신 유도</li> <li>- 규제에 대한 우려에 효과적으로 대응하면서 데이터 거버넌스, 통합을 자동화하여 지속적 데이터제공 실현</li> <li>- 데이터 파이프라인을 모니터링 및 최적화하여 모든 데이터 사용자가 계속 학습할 수 있는 피드백 루프 제공</li> </ul>	<ul style="list-style-type: none"> <li>- 공동 개발을 가능하게 하고 가치사슬 전반에서 테스트하여 아이디어의 지속적인 혁신 가속화</li> <li>- 소프트웨어 제공 프로세스를 자동화하고, 폐기물을 줄여 규제에 대한 우려를 해결하도록 지원하면서 지속적인 혁신 전달 실현</li> <li>- 소프트웨어 기반 혁신을 모니터링 및 최적화하여 고객이 계속해서 학습할 수 있는 피드백 루프 제공</li> </ul>
효율성 목표	<ul style="list-style-type: none"> <li>- IT 시스템 지원, 운영 및 비즈니스 간에 더욱 긴밀하게 연계하도록 촉진해 인력과 목표의 잘못된 조정 교정</li> <li>- 데이터 전달 사이클 전반에 자동화를 도입하여 변화 가속화 및 전달 품질 개선</li> <li>- 결과를 사용해 최적화를 구현하여 메타데이터 및 데이터의 실제 가치에 대한 인사이트 개선</li> </ul>	<ul style="list-style-type: none"> <li>- 개발자, 운영 및 비즈니스 간에 더욱 긴밀하게 연계해 인력과 목표의 잘못된 조정 교정</li> <li>- 개발 사이클 전반에 자동화를 도입하여 변화를 가속화하고 변화 전달 시 오류 제거</li> <li>- 고객 피드백을 통해 최적화를 구현하여 애플리케이션의 실제 가치에 대한 인사이트 개선</li> </ul>

DataOps는 인력, 프로세스 및 기술이 조화를 이룬 산물로, DataOps 사례에 계속해서 집중하기 위해서는 모든 부서 전반에 걸쳐 긴밀한 협업이 필요합니다. 또한 분석 속도 및 정확도를 개선하는 프로세스와 데이터 관리 사례를 구축하는 데 집중해야 합니다.

**인력 및 프로세스**

DataOps는 자동화 기술로 생산성이 뛰어난 팀을 지원하여 프로젝트 결과와 결과 전달에 걸리는 시간 측면에서 효율성 관련 이점을 제공하도록 돕습니다. 그러나 이러한 이점을 누리기 위해서는 진정으로 데이터에 기반한 내부 문화를 발전시켜야 합니다. 컨텍스트에 기반한 인사이트를 구축하기 위해 데이터를 관리하고자 하는 비즈니스 부문이 늘어남에 따라 지금이 다음을 수행할 적기입니다.

- 조직으로 흘러 드는 데이터의 품질 및 속도 증가
- 기업 전체에서 데이터 기반 비전을 지원 및 지속하기 위한 경영진 노력

이러한 유형의 혁신적인 변화는 기업의 실제 목표를 이해하는 것으로 시작됩니다. 데이터가 어떻게 고객에게 영향을 미치는 의사결정 및 서비스에 영향을 미칠 수 있을까요? 데이터가 어떻게 시장에서 경쟁우위를 유지하도록 지원할 수 있을까요? 데이터가 문제 해결에 도움이 될 수 있는 재무적 우선순위는 무엇일까요?

DataOps 리더는 문화 및 DataOps 사례의 발전을 위해 모든 데이터 사용자가 수행할 역할을 정의해야 합니다. 각 조직에는 IT, 데이터 사이언스 및 부서의 이해관계자가 성공적인 비즈니스를 견인하기 위해 가치를 더해야 하는 고유한 필요가 존재합니다. 또한 기존 데이터 거버넌스 위원회 및 이전 거버넌스 프로그램의 교훈을 활용하면 이러한 문화를 조성할 수 있으며, 거버넌스를 위한 노력은 DataOps 지원에 필요한 원동력 중 하나입니다.



그림 1: 역할별 DataOps 워크플로의 예

## 기술

DataOps의 핵심에는 조직의 **정보 아키텍처**가 있습니다. 데이터를 알고 있습니까? 데이터를 신뢰합니까? 오류를 신속하게 감지할 수 있습니까? 전체 데이터 파이프라인을 손상시키지 않고 점점 더 많은 변화를 이끌어 낼 수 있습니까?

이러한 질문에 답하기 위한 첫 번째 단계는 **데이터 거버넌스** 및 **데이터 통합** 도구 및 사례 목록을 만드는 것입니다. 자동화를 사용하는 모든 사례를 지원하기 위해서는 도구가 필요합니다.

조직 내에서 DataOps 사례를 지원하기 위한 도구를 고려할 때 다음 5가지 중요 영역에서 어떻게 자동화가 데이터 파이프라인을 혁신할 수 있는지 생각해 보십시오.

1. 데이터 큐레이션 서비스
2. 메타데이터 관리
3. 데이터 거버넌스
4. 마스터 데이터 관리
5. 셀프 서비스 상호작용

## DataOps 기능

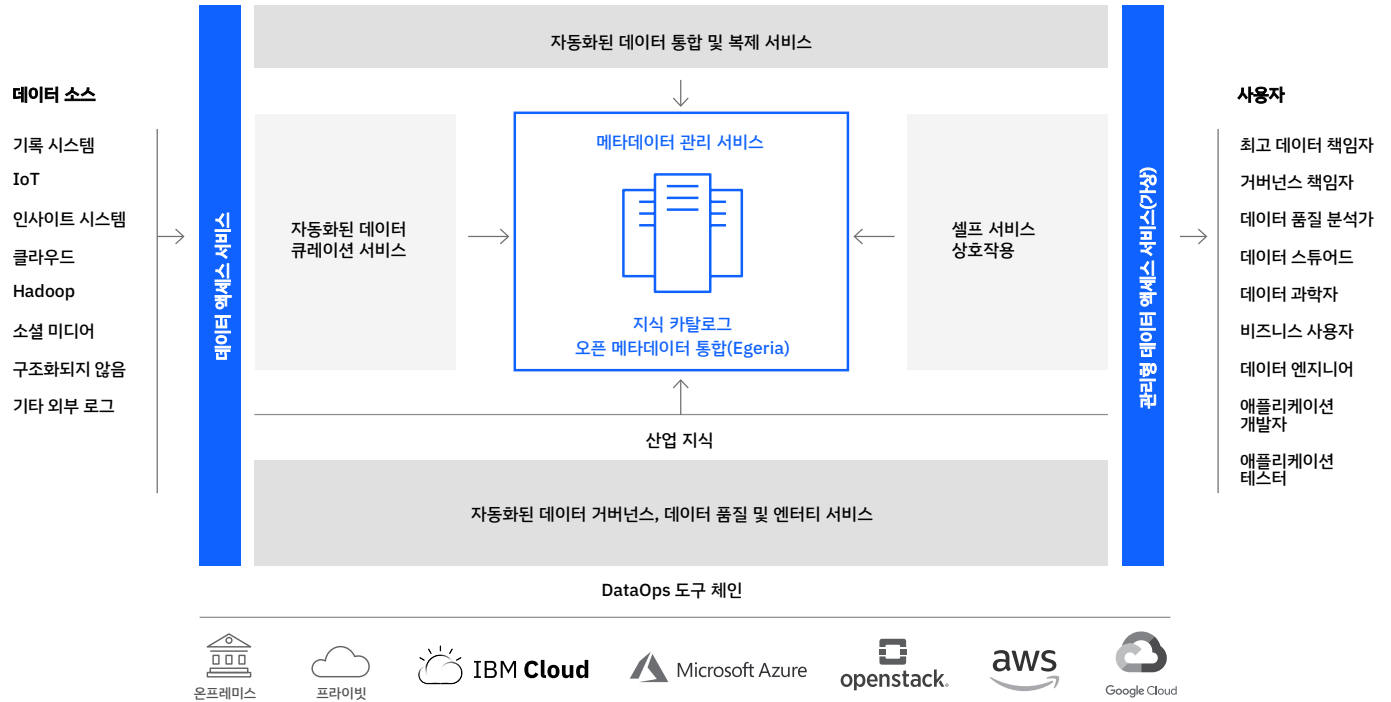


그림 2: DataOps를 지원하는 정보 아키텍처 살펴보기

비즈니스에 활용 가능한 데이터의 프로비전에는 이러한 모든 측면이 포함되며, 모든 DataOps 사례에는 5가지 측면을 모두 통합하는 전체적인 접근 방식을 포함해야 합니다. 다른 부분을 희생하면서 데이터 파이프라인의 한 가지 요소에 집중하는 조직은 DataOps 사례 구현의 이점을 실현하기 어렵습니다. 기술 대화 및 구현은 인력 및 프로세스에 관한 지속적인 계획과 동떨어져서는 안 됩니다. 도구는 문화를 지원 및 지속하도록 도와줍니다.

## 요점

조직이 여전히 데이터 스튜어드의 역할 정의 또는 데이터 검증 규칙 생성 등 기본적인 문제를 해결하는 데 어려움을 겪고 있다면 DataOps는 박차게 느껴질 수 있습니다. 그러나 DataOps 사례는 여러 조직에서 디지털 혁신 이니셔티브에서 경험한 다양한 실패에 대한 솔루션을 제공합니다.

조직에서 인정하는 가장 일반적인 실패의 예는 데이터 레이크 내에 있습니다. 많은 조직에서 기술적 성공을 이루기 위해 두 번째, 세 번째, 네 번째 시도를 하고 있으며, 필요한 문화적 변화를 수용할 수 있도록 리더십 팀에 다시 활기를 불어넣고 있습니다. 그런데 과거에 이러한 데이터 레이크 구현에 실패한 이유는 무엇일까요?

데이터 레이크 구현 프로젝트 중 다수는 정화되지 않거나 관리되지 않은 데이터를 수집하는 데 전적으로 집중했습니다. 아마 그러한 실패는 인력, 프로세스 및 기술 문제의 효율적 처리를 방해하는 제한사항 때문에 발생했을 것입니다.

## IBM DataOps 프로그램

DataOps 도입으로의 변화는 현실입니다. 최근의 조사에 따르면 **기업의 73%가 DataOps에 투자할 계획입니다.**<sup>2</sup> IBM에서는 규범적 방법론, 첨단 기술 및 **IBM DataOps Center of Excellence(CoE)**를 통해 **DataOps 사례**를 구축할 수 있는 길을 제시합니다. CoE에서 IBM 전문가는 조직과 함께 협력해 비즈니스 목표를 바탕으로 접근 방식을 맞춤화하고 이해관계자를 위한 가치를 구축하는 데 올바른 파일럿 프로젝트를 파악합니다.

IBM DataOps 기능은 AI 지원 자동화, 주입된 거버넌스 및 강력한 지식 카탈로그와 함께 작동하는 업계 최고의 기술을 제공하여 비즈니스에 활용 가능한 데이터를 제공하도록 도와 기업 전반에서 지속적인 고품질 데이터를 운용할 수 있도록 합니다. 또한 효율성, 데이터 품질, 검색성을 개선하고, 기본적으로 모든 소스에서 적시에 책임자에게 셀프 서비스 데이터 파이프라인을 제공하기 위한 관리 규칙을 적용합니다.

데이터 레이크 관리를 촉진하는 솔루션에서부터 애플리케이션 개발 및 규제 준수 보장 지원까지 IBM DataOps는 조직에서 의사결정 및 시간을 최적화하는 데 데이터의 가치를 알 수 있도록 돕습니다. 조직이 클라우드 환경 및 모든 중요한 환경에서 가치를 끌어 올리기 위해 데이터를 파악하고, 신뢰하고, 사용하는 경우 고품질 엔터프라이즈 데이터를 제공해 AI를 지원하는 것은 그리 어렵지 않은 일입니다.

## DataOps를 충족하는 IBM Cloud Garage 방법론

IBM Cloud Garage 방법론은 새로운 기능을 지속적으로 설계, 제공 및 검증하기 위한 비즈니스, 개발 및 운영을 가능하게 하는 접근 방식입니다. 실무 방식, 아키텍처 및 도구 체인은 처음부터 고객 피드백 및 시장 변화 수집 및 대응을 통해 전체 제품 수명 주기를 다룹니다. Open Toolchain 아키텍처는 지속적인 제공(CD) 등과 같은 IBM Cloud® Platform 서비스를 최고의 오픈 소스 타사 도구와 함께 통합하고, DataOps 사례에 맞춘 도구 체인으로 손쉽게 결합하도록 설계되어 있습니다. 이러한 패턴은 팀 간에 템플릿으로 공유해 조직 전반에서 DataOps의 성공적인 도입을 촉진할 수 있습니다.

IBM은 DataOps 사례의 성공적인 구현을 위해 필요한 문화적 고려 사항에 더해 DataOps 수명 주기에서 6가지 단계를 식별했습니다. 이때 IBM의 혁신 여정의 일부로 내부 DataOps 도입도 기본 요소로 고려했습니다.

IBM Cloud Garage 방법론은 이러한 6단계를 다음으로 설명합니다.

- **통찰.** 기능의 개념화, 개선 및 우선순위 지정
- **코딩.** 기능의 생성, 개선, 최적화 및 테스트
- **제공.** 제품의 자동화된 생산 및 제공
- **실행.** 실행해야 할 서비스, 옵션 및 기능
- **관리.** 제품의 지속적인 모니터링, 지원 및 복구
- **학습.** 실험의 결과를 바탕으로 한 지속적인 학습 및 피드백



그림 3. IBM Cloud Garage 방법론의 6가지 단계

### 통찰: DataOps 성숙도의 지속적인 평가 및 비즈니스 목표에 맞추기

DataOps는 기존 조직 및 확립된 프로세스에 혁신적일 수 있습니다. DataOps의 목적은 기존의 많은 수동 작업을 자동화하고, 데이터 파이프라인 생성 프로세스를 간소화하는 것입니다. 기본 DataOps 사례를 시작 또는 지속하던 상관없이 비즈니스에 활용 가능한 데이터를 빠르게 제공하고 비즈니스 가치 창출에 부합하는 개선을 위한 계획을 마련하는 팀 능력을 평가하는 것이 중요합니다.

DataOps의 성공은 메타데이터를 수집하여 데이터 자산에 대한 **카탈로그 로그 작업을 수행**하고, 데이터 클래스에 정책을 할당하고, 데이터 품질을 평가해 점수를 매기고, 스프레드시트, 조직 내에만 머무르는 지식 또는 수동 코딩과 달리 **데이터를 통합**하기 위한 도구를 활용하는 것으로 시작합니다. 팀의 성숙도가 정의되면 목표는 여러 DataOps 측면 전반에서 기능을 최대한 많이 개선하는 것이 되어야 합니다.

DataOps 팀은 필요한 데이터의 전달이 비즈니스에 가져올 수 있는 가치와 부합하도록 하는 데 집중해야 합니다. 이러한 정보를 더욱 빠르게 사용할 수 있다면 비용을 얼마나 절감할 수 있는지 스스로에게 질문에 보시기 바랍니다.

## 코딩: 버전 제어 시스템 사용—소스 제어 관리

데이터 파이프라인은 원시 콘텐츠를 유용한 정보로 변환을 담당하는 소스 코드입니다. 이러한 파이프라인은 데이터 분석에 매우 중요하고, 재현할 수 있는 형식으로 사용할 수 있는 소스 코드를 생성하도록 종단 간에 자동화할 수 있습니다. 분석과 관련된 여러 파일, 구성 및 매개변수는 관리 제어 없이 조직 내 다양한 위치 및 환경에 분산되어 있어 일관되지 않은 배포로 이어집니다. 수정 제어 도구(예: GitHub)는 코드 및 구성에 대한 모든 변경 사항의 저장 및 관리에 도움이 됩니다. 또한 중앙 집중화된 리포지토리는 기업이 신뢰할 수 있는 복구 기능을 사용하여 우발성 또는 재해를 비롯한 환경 전체에서 항상 일관되고 신뢰할 수 있는 정보를 사용할 수 있도록 지원합니다. 수정 제어 역시 팀이 배포 노력을 병렬화하고, 브랜치 및 병합을 사용하여 전달 파이프라인에서 민첩성을 발휘할 수 있도록 돕습니다.

데이터 분석 파이프라인이 제대로 작동하는지 확인하려면 테스트가 반드시 필요합니다. 매개변수화로 보완되는 지속적 통합/지속적 개발(CI/CD)을 통해 완전히 자동화된 방식으로 배포 및 테스트할 수 있습니다. 입력, 출력 및 비즈니스 로직 테스트는 데이터 분석 파이프라인의 각 단계에서 적용해야 하며 배포 전에 오류 또는 경고가 발생하는지 여부와 함께 정확성 또는 잠재적인 편차를 확인해 일관된 품질을 보장해야 합니다. 오류가 발생할 수 있고, 시간이 걸리며, 노동 집약적이기 때문에 고성능 조직에서는 수동 테스트를 수행하지 않습니다. CI/CD를 구현하기 위해서는 자동화된 강력한 테스트가 핵심 요소이며, 온디맨드 경제에 필수적입니다.

## 전달: DataOps 프로세스 및 워크플로 자동화—데이터 기술

DataOps 방법론을 성공적으로 도입하기 위해서는 자동화가 필수이며, 런타임 유연성을 가지고 설계된 데이터 분석 파이프라인이 필요합니다. 신뢰할 수 있는 데이터를 전달하기 위한 핵심 요구사항은 메타데이터 및 데이터 샘플링 기술을 사용하여 데이터 큐레이션, 데이터 수집, 카탈로그 및 분류를 바탕으로 하는 일관된 관리형 데이터 파이프라인입니다.

신뢰할 수 있는 관리형 데이터의 전달에 필요한 반복 가능하고 강력한 데이터 파이프라인에는 다음 작업을 수행하기 위한 메커니즘이 필요합니다.

- 데이터 거버넌스 및 데이터 정보 보호 정책을 일관되도록 정의 및 실행
- 효과적인 데이터 이동 지원
- 수정 시작 또는 사전 정의된 용어를 사용한 산업별 모범사례 및 템플릿 구축

이러한 프로세스는 소스 코드 또는 구성 변경 없이 여러 플랫폼 전체에서 일관되게 관리형 데이터 거버넌스를 배포할 수 있으며, 신뢰할 수 있는 완전 관리형 데이터를 제공할 수 있습니다. 또한 예외 처리 및 관리를 지원하기 위해 DataOps 프로세스는 수정에 적절한 도구로 뒷받침해야 합니다. 모든 변경에 대한 역추적 및 감사 가능성은 이러한 관리형 데이터 파이프라인에 대한 실질적인 요구사항입니다.

IBM은 기업 전반에서 지속적인 고품질 데이터를 운용할 수 있도록 임베디드 기계 학습(ML), AI 자동화, 주입된 거버넌스와 강력한 데이터 카탈로그를 포함한 혁신적인 새로운 기능을 제공합니다. DataOps의 효율성은 데이터 파이프라인에 사용되는 데이터 기술 구성 요소의 극단적 자동화에 따라 달라집니다.

- IBM Watson® Knowledge Catalog(WKC)를 비롯하여 IBM Cloud Pak for Data®는 이러한 요구사항을 효율적이고, 강력하며, 반복 가능한 자동화된 방식으로 처리할 수 있습니다.
- IBM Cloud Pak for Data 서버는 데이터 품질 및 정책 실행을 보장하도록 지원하면서 동시에 데이터 파이프라인 내에서 데이터 이동, 게시 및 사용 필요성을 처리할 수 있습니다. 효율적인 소스 제어 관리 기능을 사용해 CI/CD 파이프라인 내에서 효율적으로 자동화 및 실행할 수 있습니다.
- IBM Cloud Pak for Data용 IBM Watson Knowledge Catalog 내에 내장된 ML은 자동화 프로세스를 보완하고 강력한 파이프라인에 대해 반복할 때마다 자동화 프로세스를 최적화합니다.
- IBM Cloud® DevOps Insights는 데이터 파이프라인에 대한 인사이트 및 시각화 제공에 도움이 될 수 있습니다. 또한 지속적으로 모니터링되는 보안 및 품질 수단을 이행하고, 예기치 않은 변동을 감지하고, 극단적 자동화 및 IBM Cloud Pak for Data와의 맞춤형 통합을 기반으로 한 운영 통계를 생성하도록 지원합니다.
- Apache Airflow 및 NiFi는 워크플로 설계 및 오케스트레이션에 유용할 수 있습니다.
- 매개변수화와 함께 REST 엔드포인트를 통해 자동화를 적극적으로 사용하면 특정 데이터 세트 또는 환경의 동적 선택에 도움이 되며, 파이프라인의 소스 코드에 영향을 미치지 않고 동작을 수정하고, 데이터 분석 전문가에 대한 일상적인 필요를 수용할 수 있습니다.

## 실행: 지속적 통합 및 배포

### 지속적 통합

데이터 파이프라인 엔지니어 또는 소유자는 언제든지 파이프라인을 업데이트 또는 변경할 수 있으며, 수정 제어 시스템 내에서 개발 브랜치 또는 사설 브랜치 내 개인 사본으로 유지할 수 있습니다. 여러 엔지니어가 함께 작업하고 개발 또는 사설 브랜치에 변경사항을 동시에 전달하여 생산성을 몇 배나 증대할 수 있습니다. 파이프라인 변경사항이 완료되어 브랜치 내에서 테스트되면 소스 코드를 기본 코드 베이스 또는 트렁크에 병합하고 생산 라인에 전달할 수 있습니다. 병합된 코드가 작동하지 않는 경우에는 데이터 파이프라인을 이전에 작동했던 파이프라인 소스 코드 버전으로 언제든지 되돌릴 수 있습니다. 브랜치 및 병합을 통해 데이터 분석 팀은 자체 테스트를 실행하고, 변경하고, 위험을 감수하고, 실험하고, 변경사항이 성공적이지 않은 것으로 확인된 경우 폐기할 수 있습니다.

### 지속적 배포

이러한 파이프라인을 실행하기 위해 데이터 분석 전문가는 소스 코드 및 환경의 개인 사본과 별도로 파이프라인에서 사용하는 관련 데이터가 필요합니다. 프로덕션 데이터베이스 또는 환경에서 직접 작업하는 것은 효율성이 떨어지고 흔히 충돌로 이어집니다. 충돌 및 의존성을 줄이기 위해 데이터 파이프라인에는 다음이 필요합니다.

- 효율적인 소스 제어 관리
- 유연한 환경 배포 옵션 가용성
- 데이터 동작 테스트

Jenkins 파이프라인은 IBM 전달 파이프라인을 보완하는 도구입니다. Red Hat® OpenShift®는 런타임 시 제공되는 여러 값을 사용하여 데이터 파이프라인의 동시 인스턴스 확인에 사용되는 반복 가능하고 일관된 배포 플랫폼을 제공합니다.



**관리: 일관되고 빈번한 배포 작업**

데이터 분석 전문가는 프로덕션 환경에서 현재 데이터 파이프라인을 손상시키는 변경사항은 배포하지 않고자 합니다. 다음 두 가지 핵심 워크플로가 이러한 상황을 처리할 수 있습니다.

- **가치 파이프라인:** 프로덕션을 향하는 데이터 흐름을 통해 조직을 위한 가치 생성
- **혁신파이프라인:** 현재 개발 중이고 생산 파이프라인에 추가되는 새로운 분석 기능의 미래로 이어짐.

이러한 두 파이프라인은 DataOps 조직이 뛰어난 품질을 유지하면서 새로운 기능의 생산 및 배포에 대한 데이터 오케스트레이션에 속달된 프로덕션 환경에서 교차합니다. 개발 팀에서는 프로덕션 시스템에 대한 손상을 걱정할 필요 없이 데이터 파이프라인 품질 제어(예: 데이터 및 새로운 배포 파이프라인을 모니터링하는 통계 프로세스 제어)를 배포할 수 있습니다. 민첩한 개발 및 DevOps를 통해 새로운 분석 속도가 극대화되어 더욱 빨라졌습니다.

따라서 비즈니스 필요를 분석 아이디어로 변환할 때 필요한 시간 및 노력이 최소화되고, 반복 가능하고, 재사용 가능한 프로덕션 프로세스로 릴리스할 수 있습니다.

**학습: 커뮤니케이션 및 프로세스 관리**

효율적이고 자동화된 알림은 DataOps 방법론 내 커뮤니케이션 및 수정 프로세스에 매우 중요합니다. 소스 코드가 변경된 경우 또는 파이프라인이 실행, 실패, 완료 또는 배포된 경우 이러한 이벤트를 알릴 수 있습니다. 실패한 경우 문제 해결을 위해 알림과 함께 정보를 푸시할 수 있습니다. 파이프라인을 검증한 후 다음 단계로 배포하고, 최신 정보 및 데이터 품질로 대시보드를 업데이트하기 위해 후속 수정 프로세스를 자동으로 트리거할 수 있습니다. Slack, Apache Kafka, PagerDuty, Trello 등과 같은 도구가 DataOps 도구 체인의 일부로 이해관계자 간 커뮤니케이션, 협업, 피드백 파악 및 공유를 촉진하는 데 일반적으로 사용됩니다.

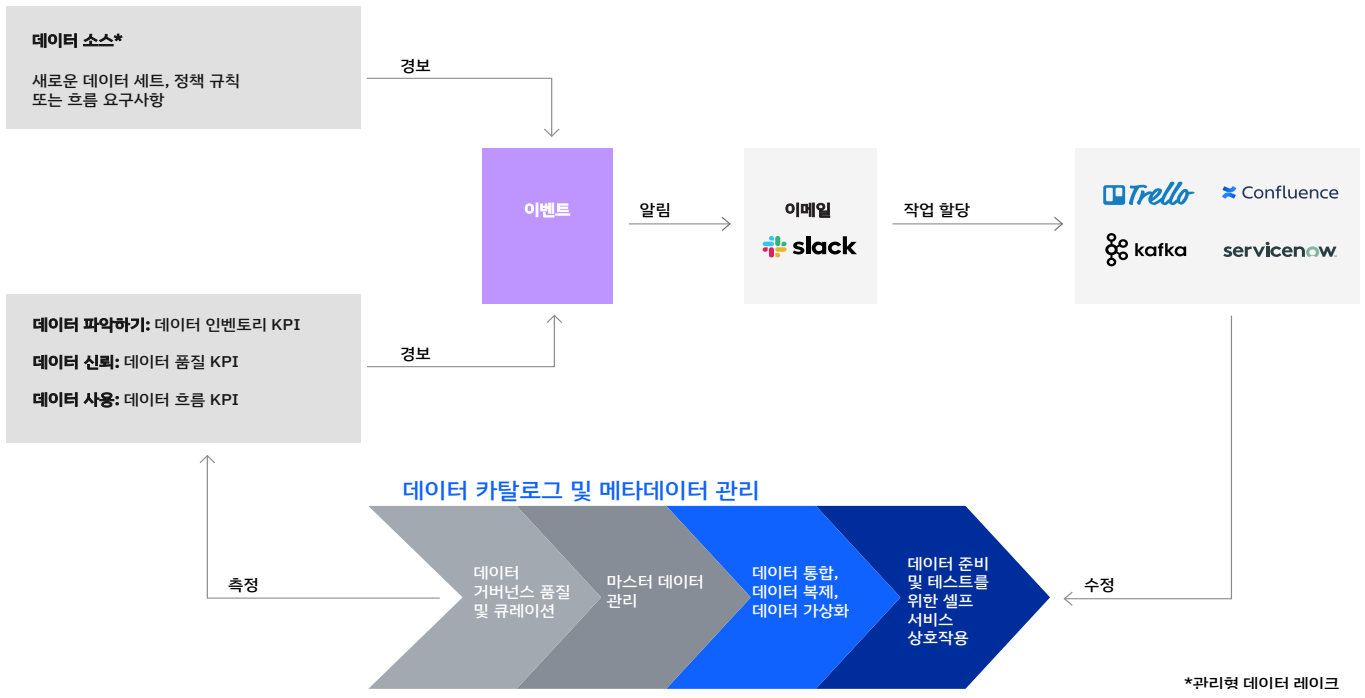


그림 4: 관리형 데이터 레이크 환경 내에서 커뮤니케이션 및 프로세스 관리 시각화

## 성공적인 DataOps 사례의 영향

한 소매업체는 DataOps 사례를 채택하여 데이터 파이프라인 전반을 개선하여 이전에 3주나 걸리던 데이터 변경 작업을 2분 이내에 적용할 수 있었습니다. 그 결과 이 소매업체는 비즈니스에 활용 가능한 데이터를 사용하여 이전에는 20일이나 걸리던 고객 선호도 분석 프로세스를 하루도 안채 수행할 수 있었습니다. 게다가 재고 위치에 대한 보고에 소요되는 시간을 1/6로 줄였습니다.

성공적인 DataOps 사례의 기준은 다음과 같습니다.

1. **데이터 오피스 설정.** 이 프로세스에는 은행에 데이터를 리소스로 제공할 때 역할의 범위를 확실하게 정의하고, 핵심 이해관계자를 식별하고, 데이터 파이프라인의 모든 이해관계자가 협업 작업 및 문화에 대해 수행하는 노력을 파악하는 것이 포함됩니다.
2. **비즈니스 목표에 맞추기.** 경쟁우위를 유지하기 위해 시장에서는 정보를 기반으로 한 데이터 중심 접근 방식을 통해서만 제공할 수 있는 새로운 기회에 신속하게 반응할 수 있어야 합니다. 간단히 말해, 비즈니스 및 데이터 전달 간에 뛰어난 커뮤니케이션이 빠져 있다면 경영진은 조직이 성장할 수 없다는 사실을 알게 됩니다.
3. **데이터 확장 성공.** 모든 데이터 중심 직관력을 통해 리더는 생성된 데이터를 계속해서 사용 또는 재사용할 수 있음을 보장해야 하고 매번 데이터의 가치는 커져야 합니다. 중앙에서 데이터를 공유하고, 검색할 수 있고, 비즈니스 언어에 데이터를 맞출 수 있는 경우에만 이러한 결과를 얻을 수 있습니다.



# 결론

DataOps를 성공적으로 배포한 조직은 액세스할 수 있는 데이터 자산을 파악하고 있으며, 데이터 의미 및 그 품질을 신뢰하고, 데이터를 최대한 사용할 수 있습니다. 데이터는 비즈니스에 활용 가능한 신뢰할 수 있는 데이터가 차별화된 인사이트, 운영 효율성, 협업 및 경쟁 우위를 견인하도록 지원할 때 가치를 지닙니다.

DataOps 사례를 구축하려면 다음이 필요합니다.

- 파일럿 프로젝트를 실행하여 조직의 고유 역량 및 문제 파악에 투자
- 파일럿 프로젝트의 성공을 활용하여 DataOps 기술과 조직 확장 및 성장
- DataOps 사례에 참여할 수 있는 팀을 추가로 모집하기 위해 성공적 홍보
- 교훈 공유 및 DataOps CoE 구축 시작

[dataops@us.ibm.com](mailto:dataops@us.ibm.com)에 문의해 다음 단계를 수행하고 고유한 IBM DataOps Garage 워크숍을 예약하고 비즈니스에 활용 가능한 데이터를 얻기 위한 여정에 속도를 내보십시오.

데이터 카탈로그, 데이터 레이크 또는 마스터 데이터 이니셔티브에 대한 작업을 이미 수행 중인 경우 조직은 DataOps를 도입하기 위해 효과적으로 준비된 상태입니다. [ibm.com/DataOps](https://ibm.com/DataOps)에서 시장을 선도하는 기술을 활용한 DataOps 지원에 대해 자세히 알아보십시오.



© Copyright IBM Corporation 2020  
IBM Corporation  
New Orchard Road, Armonk, NY 10504  
Produced in the United States of America  
2020년 4월

IBM, IBM 로고, [ibm.com](https://ibm.com), IBM Cloud, IBM Cloud Garage, IBM Watson 및 Cloud Pak for Data는 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"([ibm.com/legal/copytrade.shtml](https://ibm.com/legal/copytrade.shtml))에 있습니다.

Microsoft 및 Windows는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

Red Hat 및 OpenShift는 미국 및 기타 국가에서 Red Hat, Inc. 또는 그 자회사의 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 고객 사례는 오로지 예시를 목적으로 작성된 것입니다. 실제 성능 결과는 특정 구성 및 운영 환경에 따라 달라질 수 있습니다. 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 변경, 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.

- 1 2019 Global data management research: Taking control in the digital age." Experian, 2019.
- 2 Jarah Euston, "The DataOps Trend is Real: 73% of Companies Plan to Invest in DataOps to Manage Data Teams in 2018," Nexla

10028810-KRKO-01

# 부록: DataOps 파일럿 프로그램 템플릿

프로젝트 이름:

날짜:

부서 또는 단위:

파일럿 프로그램 리더:

이름	역할	이메일	전화번호

회사 전체의 확장 이해관계자:

이름	역할	이메일	전화번호

문제 설명:

근본 원인 체크리스트:

문제	적용 가능 여부? Y/N	추가 메모

성공 지표:

시작 날짜:

스프린트 종료 날짜:

평가:

구현	현재 상태	원하는 스프린트 결과	원하는 결과를 얻기 위한 조치 단계
데이터 자산 수집, 자동화된 검색 및 분류			
데이터 품질 평가 및 수정			
비즈니스 용어 배정			
데이터 개인 정보 보호, 규제 준수 및 회사 정책 정의 및 이행			
데이터 소비자 요구사항 정의 요청 처리			
데이터 요청 커뮤니케이션 및 알림 (예외 및 오류 처리 및 수정 포함)			
큐레이션된 데이터를 카탈로그에 게시			
데이터 연계 및 보고			
협업, 피드백 및 감사			

**구현 감사 중 할 수 있는 질문의 예:**

데이터 자산 수집, 자동화된 검색 및 분류

- 비즈니스 연속성을 위해 대기 시간이 짧은 대용량 복제를 수행합니까?
- 대기 시간이 짧은 실시간 분석을 위해 고급 스트리밍 분석을 사용합니까?
- 모든 데이터 소스에 쉽게 연결하고 복잡한 데이터 변환 및 통합을 수행합니까?
- 소셜 미디어의 데이터, 날씨 데이터 또는 기타 퍼블릭 클라우드 데이터 소스의 데이터를 제공할 수 있습니까?
- 우리 회사의 데이터 소비자가 모든 데스크톱 애플리케이션에서 우리 메타데이터 리포지토리에 실시간으로 액세스할 수 있습니까?
- 우리 데이터 소비자가 데이터 카탈로그에 실시간으로 액세스해 자신의 업무와 관련이 있는 데이터 세트 검색 시 셀프서비스 및 지원을 이용할 수 있습니까?
- 데이터를 파악하고, 데이터 가치, 열 및 테이블 관계를 확인하고, 이상을 찾아 분석하기 위해 데이터 프로파일링 도구를 사용합니까?
- 비즈니스 규칙 관리가 메타데이터 관리 인프라에 통합되어 있습니까?

