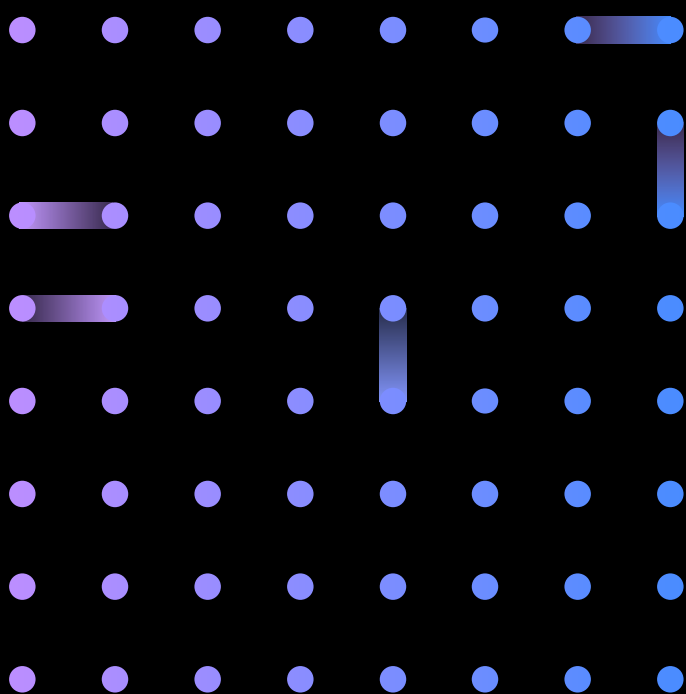


DataOps를 통해 즉시 활용 가능한 비즈니스 데이터를 빠르게 제공하세요

IBM DataOps 방법론 및 방식 소개



목차

소개	3
DataOps 정의	3
목표 비교	4
IBM DataOps 프로그램	7
IBM Cloud Garage 방법론과 DataOps의 결합	7
성공적인 DataOps의 영향	10
결론	11
부록: DataOps 파일럿 프로그램 템플릿	12

주요 내용

- DataOps는 데이터 사용자에게 신뢰할 수 있는 고품질 데이터를 신속하게 제공하기 위한 사람, 프로세스, 기술의 오케스트레이션입니다.
- 자동화 기반 DataOps는 비효율적인 데이터 액세스, 준비, 통합, 가용성 확보와 관련된 문제를 해결합니다.
- IBM DataOps는 극단적인 자동화를 통해 데이터 큐레이션 서비스, 메타데이터 관리, 데이터 거버넌스, 마스터 데이터 관리 및 셀프 서비스 상호 작용 등의 기능에 측정 가능한 영향을 줍니다.
- IBM은 규범적 방법론, 인공 지능(AI) 지원 자동화, IBM DataOps Center of Excellence를 통해 DataOps 도입 경로를 제공합니다.
- DataOps 워크숍은 DataOps 로드맵의 필수 요소로 조직이 DataOps 성숙도를 평가하고 파일럿 프로젝트 실행을 계획하도록 지원합니다.

소개

데이터는 혁신과 경쟁 우위를 유지하는 원동력으로, 분석과 비즈니스 동향 파악 그리고 기회 포착의 핵심 요소입니다. 새로운 방식으로 데이터의 가치를 실현하면 AI를 향한 조직의 여정 속도를 높일 수 있습니다.

그러나 데이터 관련 프로젝트가 약속한 투자 수익(ROI)을 제공하지 못하면 이해관계자들은 그 이유를 알고 싶어합니다. Experian의 2019 글로벌 데이터 관리 연구 보고서에 따르면 기업의 89%가 데이터 관리에 어려움을 겪고 있으며, 여기에는 인사이트의 지연과 기초 데이터에 대한 신뢰 부족도 포함됩니다.¹

조직의 비즈니스 목표를 이해하는 것은 분석 및 AI를 위한 효과적인 데이터 전략을 개발하는데 매우 중요합니다. 비즈니스 모델이 작동하려면 고객의 요구 사항을 충족해야 합니다. 성공 여부는 언제나 비즈니스에 대한 완전하고 일관된 가시성을 제공하고 즉시 활용 가능한 통합 비즈니스 데이터 파이프라인으로, 데이터 운영을 간소화하는 데 달려 있습니다.

더 빠른 결과를 얻으려는 기대치는 계속 높아지고 있습니다. 조직 내 만연한 사일로 때문에 모든 기업은 운영 효율과 효과를 개선하여 최상의 의사 결정을 지원할 방법을 찾고 있습니다. 이 두 가지 요인으로 인해 기업 경영진은 단일 프레임워크 내에서 가장 큰 문제를 해결할 수 있는 새로운 방법을 모색하게 됩니다.

데이터 운영을 혁신하려는 조직의 경우 자동화 기술이 경쟁 우위를 제공할 수 있습니다. 신뢰할 수 있고 즉시 활용 가능한 비즈니스 데이터로 조직을 위해 차별화된 인사이트를 확보하고 운영 효율을 높일 수 있을 때 데이터의 가치가 높아집니다.

이 백서의 목적은 DataOps 방법론, 방식, 로드맵의 장점을 조명하는 것입니다.

DataOps 정의

데이터 운영(DataOps)은 데이터 사용자에게 신뢰할 수 있는 고품질 데이터를 신속하게 제공하기 위한 사람, 프로세스, 기술의 오케스트레이션입니다. 이는 조직 전반의 협업을 지원하여 민첩성, 속도, 새로운 데이터 이니셔티브를 대규모로 추진하는 데 중점을 둡니다. DataOps는 자동화를 통해 비효율적인 데이터 액세스, 준비, 통합, 가용성 확보와 관련된 문제를 해결하도록 설계되었습니다.

DataOps의 잠재적인 이점에는 상당한 생산성 향상이 포함됩니다. 이러한 생산성 향상은 정보와 데이터를 개인에게 전달하고 프로세스를 개선하여 효율 향상과 최적화를 달성합니다. AI 데이터 주도 이니셔티브를 포함하는 자동화된 데이터 운영은 다음과 같은 결과를 달성하는데 기여합니다.

- 대규모 분석 및 AI를 견인하는 즉시 활용 가능하고 통합된 비즈니스 데이터 제공
- 운영 효율성 달성
- 데이터 개인 정보 보호 및 규정 준수 지원

89%

기업이 데이터 관리에 어려움을 겪고 있으며,¹ 조직의 비즈니스 목표를 이해하는 것은 분석 및 AI를 위한 효과적인 데이터 전략을 개발하는데 매우 중요합니다.



DataOps는 DevOps와 다릅니다

대부분의 조직은 자체적인 개발 원칙 하에서 일정 수준의 DevOps를 구현했습니다. DevOps는 그 익숙함과 유사한 명칭으로 인해 새로운 DataOps와 비교될 때가 많습니다. 둘 다 운영 모범 사례를 촉진하는 방법이지만 조직 내에서 각각 고유한 위치를 차지합니다.

아래 표에서 조직의 목표와 이점이라는 관점에서 이 둘을 어떻게 비교할 수 있는지 살펴보세요.

목표 비교

	DataOps	DevOps
중점 항목	빠르게 사용 가능하고 신뢰할 수 있는 고품질 비즈니스 데이터	애플리케이션 및 소프트웨어 개발
혁신 목표	<ul style="list-style-type: none">- 모든 데이터 사용자가 신뢰할 수 있는 고품질 데이터에 셀프 서비스로 액세스할 수 있도록 하여 비즈니스를 위한 지속적이고 빠른 혁신 촉진- 데이터 거버넌스와 통합을 자동화하고 규제 문제를 대응하면서 지속적인 데이터 전달을 실현함- 데이터 파이프라인을 모니터링하고 최적화하여 모든 데이터 사용자의 지속적 학습을 위한 피드백 루프 제공	<ul style="list-style-type: none">- 가치 사슬 전반에 걸쳐 협업 개발 및 테스트를 지원하여 아이디어의 지속적 혁신 가속화- 소프트웨어 제공 프로세스를 자동화하고 낭비를 해소함으로써 이런 혁신을 지속적으로 제공하는 동시에 규제 문제 해결에 기여함- 소프트웨어 기반 혁신을 모니터링하고 최적화하여 고객의 지속적 학습을 위한 피드백 루프 제공
효율성 목표	<ul style="list-style-type: none">- IT 시스템 지원, 운영 및 비즈니스 간 긴밀한 연결을 촉진하여 사람과 목표의 불일치 사항 수정- 데이터 전달 사이클 전반에 자동화를 도입하여 변화 가속화 및 전달 품질 개선- 결과를 활용하는 최적화를 통해 메타데이터 및 데이터의 실제 가치에 대한 인사이트 개선	<ul style="list-style-type: none">- 개발자, 운영 및 비즈니스 간 긴밀한 연결을 촉진하여 목표의 불일치 사항 수정- 개발 사이클 전반에 걸친 자동화를 도입하여 변화를 가속화하고 오류 제거- 고객 피드백을 활용하는 최적화를 통해 메타데이터 및 데이터의 실제 가치에 대한 인사이트 개선

DataOps는 사람, 프로세스, 기술의 오케스트레이션으로 DataOps 실행에 집중하기 위해서는 모든 부서의 긴밀한 협업이 필요합니다. 분석 속도와 정확성을 향상시키는 데이터 관리 방식과 프로세스 구축에 집중해야 합니다.

사람과 프로세스

DataOps는 자동화 기술로 생산성이 높은 팀을 지원해 프로젝트 결과와 소요되는 시간 측면에서 효율을 높여줍니다. 그러나 이런 이점을 경험하려면 조직의 내부 문화가 진정한 데이터 중심 문화로 진화해야 합니다. 컨텍스트 기반 인사이트를 확보하기 위해 데이터를 관리해야 하고, 관리하려는 기업이 증가할 수록 다음을 수행할 적기입니다.

- 조직으로 유입되는 데이터의 품질과 속도 개선
- 비즈니스 전반에 걸쳐 데이터 기반 비전을 지원하고 유지하기 위한 경영진의 약속 확보

이런 유형의 혁신은 비즈니스의 진정한 목표를 이해하는 것부터 시작됩니다. 데이터에서 얻는 정보는 고객에게 영향을 미치는 의사결정과 서비스에 어떤 영향을 줄까요? 시장에서 경쟁 우위를 유지하는데 데이터가 어떻게 도움이 될까요? 데이터가 문제 해결하는데 도움이 될 수 있는 재정적 우선순위는 무엇인가요?

DataOps 리더는 문화와 DataOps 방식의 발전을 위해 모든 데이터 사용자의 역할을 정의해야 합니다. 각 조직마다 고유한 요구 사항이 있고, IT, 데이터 과학, 비즈니스 라인의 이해관계자는 성공적인 비즈니스를 추진하기 위해 가치를 더해야 합니다. 또한, 거버넌스가 DataOps를 지원하는데 필요한 원동력 중 하나이므로 기존 데이터 거버넌스 위원회와 종신 데이터 거버넌스 프로그램의 교훈을 활용하는 것은 이런 문화를 확립하는데 도움이 됩니다.

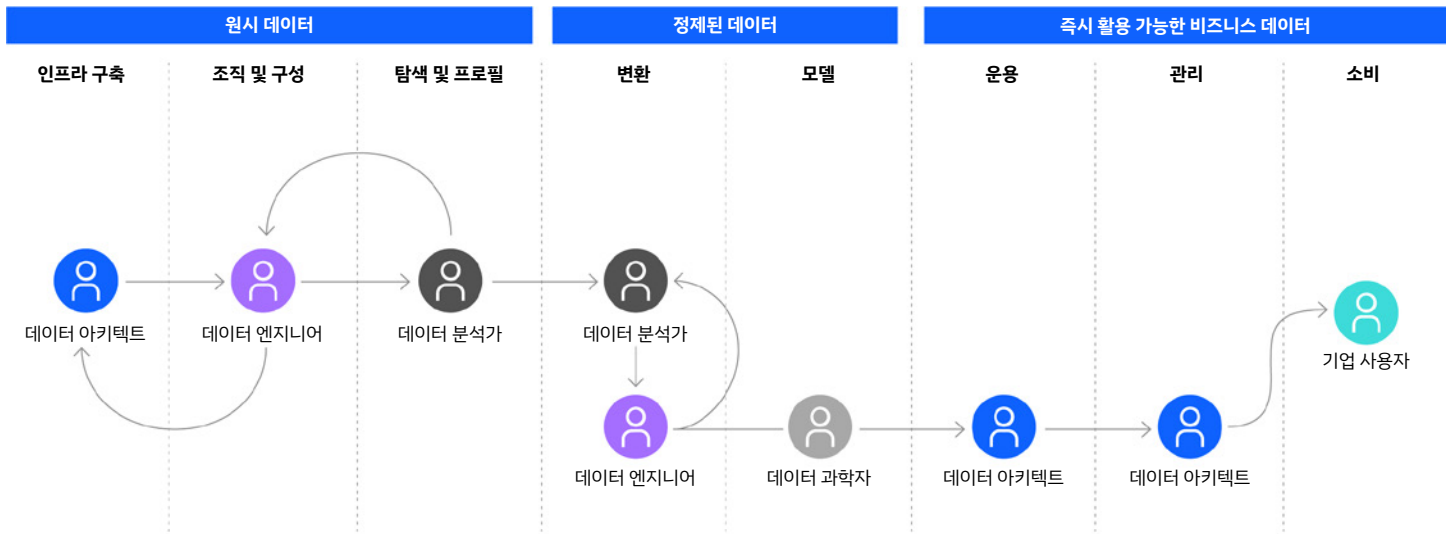


그림 1: 역할별 DataOps 워크플로우 예시

기술

DataOps의 핵심은 조직의 **정보 아키텍처**입니다. 귀사의 데이터를 알고 있습니까? 데이터를 신뢰합니까? 오류를 빠르게 감지할 수 있습니까? 전체 데이터 라이프라인을 '손상'시키지 않고 점진적으로 변경할 수 있습니까? 이러한 질문에 답하기 위한 첫 번째 단계는 **데이터 거버넌스**와 **데이터 통합** 도구 및 방식에 대한 인벤토리를 작성하는 것입니다. 자동화에 의존하는 방식을 지원하려면 도구가 필요합니다.

조직 내에서 DataOps 방식을 지원하기 위한 도구를 고려할 때 다음 5가지 핵심 영역의 자동화가 데이터 라이프라인을 어떻게 변화시킬 수 있는지 생각해 보십시오.

1. 데이터 큐레이션 서비스
2. 메타데이터 관리
3. **데이터 거버넌스**
4. **마스터 데이터 관리**
5. 셀프 서비스 상호 작용

DataOps 기능

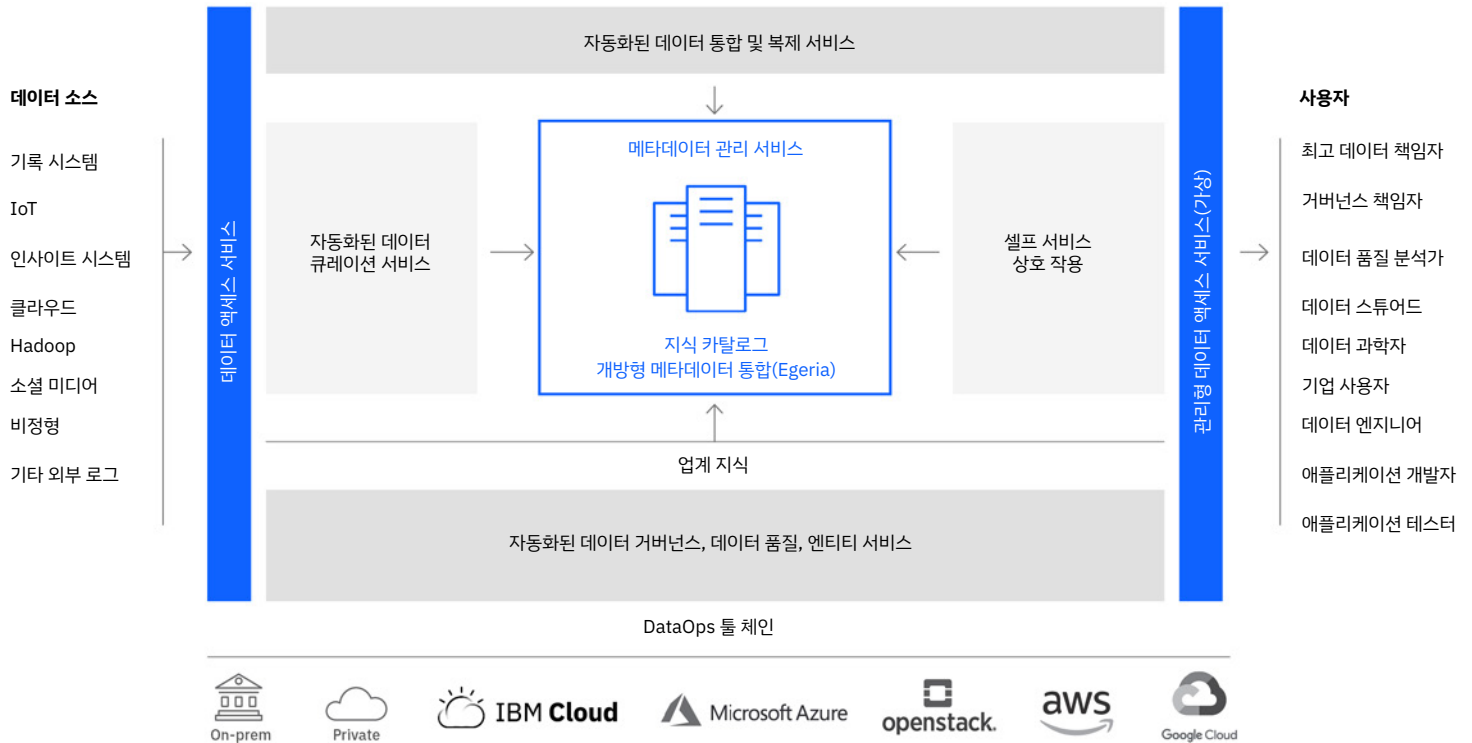


그림 2: DataOps를 지원하는 정보 아키텍처 검사

즉시 활용 가능한 비즈니스 데이터를 제공하는 데는 이러한 모든 측면이 포함되며 모든 DataOps 방식에는 5가지 측면을 모두 통합하는 전체론적 접근 방식이 포함되어야 합니다. 다른 요소를 희생하면서 데이터 라이프라인의 한 가지 요소에 집중하는 조직은 DataOps 방식을 구현해서 얻을 수 있는 이점을 실현하지 못할 것입니다. 첨단 기술에 대한 대화와 구현은 사람과 프로세스에 관한 지속적 계획과 분리되어서는 안 됩니다. 둘은 조직의 문화를 지원하고 유지하는데 도움이 됩니다.

결론

조직이 데이터 관리자의 역할 정의 또는 데이터 유효성 검사 규칙 생성과 같은 기본적인 문제로 여전히 어려움을 겪고 있는 경우 DataOps가 어렵게 보일 수 있습니다. 그러나 DataOps

방식은 조직이 디지털 혁신 이니셔티브에서 경험한 수 많은 실패에 대한 솔루션을 제공합니다.

조직이 인정하는 실패의 가장 일반적인 예는 데이터 레이크 내에 있습니다. 많은 사람들이 기술적 성공을 달성하기 위해 두 번째, 세 번째 또는 네 번째 시도를 하고 있으며 리더십 팀이 필요한 문화의 변화를 이끌도록 하고 있습니다. 그러나 과거에 이런 데이터 레이크 구현에 실패한 이유는 무엇일까요?

이러한 프로젝트 중 다수는 정제되지 않고 관리되지 않는 데이터를 데이터 레이크로 수집하는 데만 집중했습니다. 사람, 프로세스, 기술 문제를 효과적으로 해결하는 데 한계에 직면하여 실패했을 가능성이 큼니다.

IBM DataOps 프로그램

DataOps 도입을 위한 전환은 현실입니다. 최근 설문 조사에 따르면 **기업의 73%가 DataOps에 투자할 계획입니다.**² IBM은 규범적 방법론, 선도적인 기술, **IBM DataOps Center of Excellence(CoE)**를 통해 **DataOps 방식** 도입을 위한 경로를 제공합니다. CoE에서 IBM 전문가는 조직과 협력하여 비즈니스 목표를 기반으로 맞춤형 접근 방식을 정의하고 이해관계자를 위한 가치를 창출하는 올바른 파일럿 프로젝트를 파악합니다.

IBM DataOps 기능은 AI 지원 자동화, 통합 거버넌스, 강력한 지식 카탈로그와 함께 업계 최고의 기술을 제공하여 기업 전반에 걸쳐 지속적인 고품질 데이터를 운영함으로써 즉시 활용 가능한 비즈니스 데이터를 제공하는데 기여합니다. 효율, 데이터 품질, 검색 가능성을 높이고 기본적으로 모든 소스에서 적시에 적절한 사람들에게 셀프 서비스 데이터 파이프라인을 제공하기 위해 규칙을 적용합니다.

데이터 레이크 관리를 용이하게 하는 솔루션부터 애플리케이션 개발 및 규정 준수 보장까지, IBM DataOps는 조직이 의사 결정 및 시간 최적화와 관련해 데이터의 가치를 보여줄 수 있도록 지원합니다. 조직이 클라우드와 모든 중요한 환경에서 가치를 창출하기 위해 데이터를 파악하고 신뢰하여 사용할 때, AI를 지원하기 위한 고품질 엔터프라이즈 데이터를 제공하는 것이 가능해 집니다.

IBM Cloud Garage 방법론과 DataOps의 결합

IBM Cloud Garage 방법론은 비즈니스, 개발, 운영에 있어 지속적으로 새로운 기능을 설계, 제공, 검증할 수 있도록 해주는 접근 방식입니다. 실무 방식, 아키텍처, 톨 체인은 처음부터 고객 피드백 및 시장 변화 수집 및 대응을 통해 전체 제품 수명 주기를 다룹니다. Open Toolchain 아키텍처는 Continuous Delivery(CD)와 같은 IBM Cloud™ Platform 서비스를 오픈 소스 및 선도적인 타사 톨과 쉽게 결합하여 DataOps에 맞춘 톨 체인에 통합할 수 있도록 설계되었습니다. 이러한 패턴은 팀 간 템플릿으로 공유하여 조직 전반에서 DataOps의 성공적인 도입을 촉진할 수 있습니다.

IBM은 성공적인 DataOps 구현을 위해 DataOps 라이프사이클의 6단계와 필요한 문화적 고려사항을 파악했습니다. 이는 혁신 여정의 일환으로 DataOps를 도입하기 위해 IBM이 겪은 과정을 기반으로 합니다.

IBM Cloud Garage 방법론에서는 이 6단계를 다음과 같이 설명합니다.

- **사고:** 기능 개념화, 개선, 우선 순위 지정
- **코드:** 기능 생성, 향상, 최적화, 테스트
- **전송:** 자동화된 오퍼링 생산과 전달
- **실행:** 실행에 필요한 서비스, 옵션, 기능
- **관리:** 오퍼링에 대한 지속적 모니터링, 지원, 복구
- **학습:** 실험 결과에 따른 지속적 학습과 피드백



그림 3. IBM Cloud Garage 방법론의 6단계

사고: DataOps 성숙도를 지속적으로 평가하고 비즈니스 목표에 맞춥니다.

DataOps는 기존 조직과 확립된 프로세스를 혁신할 수 있습니다. DataOps의 목적은 기존의 많은 수작업을 자동화하고 데이터 파이프라인 생성 프로세스를 간소화하는 것입니다. 기본적인 DataOps 업무 방식을 시작하거나 유지하는 단계에 관계없이, 즉시 활용 가능한 비즈니스 데이터를 빠르게 제공하고 비즈니스 가치 창출에 부합하는 개선 계획을 세우는 팀의 능력을 평가하는 것이 중요합니다.

DataOps의 성공은 메타데이터를 캡처하고, 데이터 클래스에 정책을 할당하고, **데이터 품질**을 평가하고 점수를 매기고, 스프레드시트, 부족한 지식 또는 수작업 코딩 대신 **데이터 통합** 톨을 활용하여 데이터 자산을 **카탈로그화**하는 것으로 시작됩니다. 팀의 성숙도가 정의되면 목표와 목적은 가능한 많은 DataOps 측면에서 기능을 개선하는 것이 되어야 합니다.

DataOps 팀은 필요한 데이터를 제공하고, 이 데이터가 비즈니스에 가져올 수 있는 가치를 일치화 하는데 집중해야 합니다. 이 정보가 더 빨리 제공된다면 얼마나 많은 금액을 절약하거나 벌 수 있는가를 질문해 보시기 바랍니다.

코드: 버전 제어 시스템 사용 - 소스 제어 관리

데이터 파이프라인은 원시 콘텐츠를 유용한 정보로 변환하는 소스 코드입니다. 이 파이프라인은 데이터 분석의 핵심이며 재생산 가능한 형태의 사용할 수 있는 소스 코드를 생성하기 위해 엔드 투 엔드로 자동화할 수 있습니다. 분석과 관련된 다양한 파일, 구성, 매개변수가 통제 제어 없이 조직 내 다양한 장소와 환경에 배포되어 일관되지 않은 배포로 이어집니다. GitHub과 같은 수정 제어 톨은 코드 및 구성에 대한 모든 변경 사항을 저장하고 관리하는 데 도움이 됩니다. 또한, 중앙 집중식 리파지토리는 사고나 재해 발생 시 신뢰할 수 있는 복구 기능을 포함해 기업이 모든 환경에서 항상 일관되고 신뢰할 수 있는 정보를 보유할 수 있도록 지원합니다. 또한, 수정 제어는 팀이 개발 노력을 병렬화하고 분기 및 병합을 사용하여 전달 파이프라인의 민첩성을 높이도록 도와줍니다.

데이터 분석 파이프라인이 제대로 작동하는지 확인하려면 테스트해야 합니다. 매개변수화로 보완된 지속적 통합/지속적 개발(CI/CD)을 통해 배포 및 테스트를 완전히 자동화할 수 있습니다. 입력, 출력 및 비즈니스 로직 테스트는 데이터 분석 파이프라인의 각 단계에서 적용해야 하며, 배포 전에 오류 또는 경고가 발생하는지 여부 또는 잠재적인 편차를 확인하여 일관된 품질을 보장해야 합니다. 메뉴얼 테스트는 오류가 발생하기 쉽고 시간이 많이 소요되고 노동 집약적이기 때문에 고성능 조직에는 부적합합니다. 강력하고 자동화된 테스트는 CI/CD를 달성하는 핵심 요소이며 온디맨드 경제에 필수적입니다.

전송: DataOps 프로세스 및 워크플로 자동화 - 데이터 기술

DataOps 방법론을 성공적으로 도입하기 위해서는 자동화가 필수이며, 런타임 유연성을 가지고 설계된 데이터 분석 파이프라인이 필요합니다. 메타데이터 및 데이터 샘플링 기술을 사용하는 데이터 큐레이션, 데이터 수집, 카탈로그 및 분류 기반의 일관된 관리형 데이터 파이프라인은 신뢰할 수 있는 데이터를 전달하기 위한 핵심 요구사항입니다.

신뢰할 수 있고 관리되는 데이터를 전달하기 위한 반복 가능하고 강력한 데이터 파이프라인은 다음을 수행하는 메커니즘이 필요합니다.

- 데이터 거버넌스와 데이터 보호정책의 일관된 정의와 시행
- 효율적인 데이터 이동 지원
- 수정을 시작하거나 사전 정의된 용어집을 갖춘 산업별 모범 사례 및 템플릿 확보

이 프로세스는 소스 코드나 구성 변경 없이 일관되게 여러 플랫폼에 걸쳐 관리형 데이터 파이프라인을 배포하고 신뢰할 수 있는 완전 관리형 데이터를 제공할 수 있습니다. 또한 예외 처리 및 관리를 지원하기 위해서는 수정을 위한 적절한 도구로 DataOps 프로세스를 지원해야 합니다. 변경 사항에 대한 역추적 및 감사 가능성은 이러한 관리형 데이터 파이프라인의 실질적 요구 사항입니다.

IBM은 기업 전반에서 지속적인 고품질 데이터를 운용할 수 있도록 임베디드 머신러닝(ML), AI 자동화, 주입된 거버넌스와 강력한 데이터 카탈로그를 포함하는 혁신적인 새로운 기능을 제공합니다. DataOps 효율성은 데이터 파이프라인에 사용되는 데이터 기술 구성 요소의 극단적 자동화에 달려 있습니다.

- IBM InfoSphere® 및 IBM Watson® Knowledge Catalog(WKC)는 효율적이고 강력하며 자동화되고 반복 가능한 방식으로 이러한 요구사항을 해결할 수 있습니다.
- IBM InfoSphere Information Server는 데이터 품질 및 정책 시행을 보장하는 동시에 데이터 파이프라인 내에서 데이터의 이동, 게시, 사용의 필요성을 해결할 수 있습니다. 효율적인 소스 제어 관리를 통해 CI/CD 파이프라인 내에서 효율적으로 자동화하고 실행할 수 있습니다.
- IBM WKC 및 IBM Information Server에 내장된 ML은 자동화 프로세스를 보완하고 강력한 파이프라인을 위해 반복할 때마다 자동화 프로세스를 최적화합니다.
- IBM Cloud™ DevOps Insights는 데이터 파이프라인에 대한 운영 인사이트를 제공하고 시각화 하는데 도움이 될 수 있습니다. 지속적으로 모니터링되는 보안 및 품질 측정을 시행하고 예기치 않은 변동을 감지하며 IBM WKC와 IBM Information Server간의 자동화와 맞춤형 통합을 기반으로 운영 통계를 생성하도록 지원합니다.
- Apache Airflow 및 NiFi는 워크플로우 설계 및 오케스트레이션에 도움이 될 수 있습니다.
- 매개변수화와 함께 REST 엔드포인트를 사용하는 극단적인 자동화를 사용하면 특정 데이터 세트 또는 환경을 동적으로 선택하고, 파이프라인의 소스 코드에 영향을 주지 않으면서 동작을 수정하고, 데이터 분석 전문가의 일상적인 요구를 수용할 수 있습니다.

실행: 지속적 통합 및 배포

데이터 파이프라인 엔지니어 또는 소유자는 파이프라인을 언제든지 업데이트하거나 변경할 수 있으며, 개발 분기 또는 개인 분기 내의 개인 사본으로 개정 제어 시스템 내에 보관할 수 있습니다. 여러 엔지니어가 동시에 작업하고 개발 또는 사설 분기에 변경 사항을 동시에 전달해 생산성을 몇 배로 높일 수 있습니다. 파이프라인 변경을 완료하고 분기 내 테스트가 끝나면 소스 코드를 기본 코드 베이스 또는 트렁크에 병합하고 프로덕션 라인으로 전달할 수 있습니다. 병합된 코드가 작동하지 않는 경우 데이터 파이프라인은 항상 파이프라인 소스 코드의 이전 작업 버전으로 복구할 수 있습니다. 데이터 분석 팀은 분기 및 병합을 통해 자체 테스트를 실행, 변경하고 위험을 감수, 실험하여, 일련의 변경 사항이 실패할 경우 폐기할 수 있습니다.

지속적 배포

데이터 분석 전문가는 이러한 파이프라인을 실행하기 위한 소스 코드 및 환경의 개인 사본과 별도로 파이프라인에서 사용하는 관련 데이터가 필요합니다. 프로덕션 데이터베이스 또는 환경에서 직접 작업하는 것은 효율적이지 않으며 종종 충돌로 이어집니다. 충돌과 의존성을 줄이려면 데이터 파이프라인에 다음이 필요합니다.

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

Jenkins Pipeline은 IBM의 전달 파이프라인을 보완하는 툴입니다. Red Hat® OpenShift®는 런타임 시 제공되는 다양한 값을 사용하여 데이터 파이프라인의 동시 인스턴스를 검증하기 위한 반복 가능하고 일관된 배포 플랫폼을 제공합니다.

관리: 일관되고 빈번한 배포 작업

데이터 분석 전문가는 프로덕션 환경에서 현재 데이터 파이프라인을 손상시키는 변경 사항을 배포하는 것을 피하고 싶어합니다. 두 가지 주요 워크플로우가 상황을 해결할 수 있습니다.

- **가치 파이프라인.** 데이터가 프로덕션에 유입돼 조직을 위한 가치를 창출합니다.
- **혁신 파이프라인.** 개발 중이고 프로덕션 파이프라인에 추가되는 새로운 분석의 미래를 이어집니다.

이 두 파이프라인은 DataOps 조직이 완벽한 품질을 유지하면서 프로덕션에 대한 데이터 오케스트레이션 그리고 새로운 기능 배포를 완벽히 마스터하는 지점에서 교차합니다. 개발 팀에서는 프로덕션 시스템 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 최소한의 시간과 노력으로 비즈니스 요구사항을 분석 아이디어로 전환하고, 반복 가능하고 재사용 가능한 프로덕션 프로세스로 릴리스할 수 있습니다.

학습: 커뮤니케이션 및 프로세스 관리

효율적이고 자동화된 알림은 DataOps 방법론 내 커뮤니케이션 및 수정 프로세스의 핵심입니다. 소스 코드가 변경되거나 파이프라인이 실행, 실패, 완료 또는 배포되면 알림을 받을 수 있습니다. 장애 발생 시 알림과 함께 문제 해결을 위한 정보를 표시할 수 있습니다. 파이프라인을 검증하고, 다음 단계로 배포하고, 최신 정보 및 데이터 품질로 대시보드를 업데이트하기 위해 후속 수정 프로세스를 자동으로 트리거할 수 있습니다. DataOps 툴 체인의 일부로 이해관계자 간 커뮤니케이션, 협업, 피드백 확인 및 공유 촉진을 위해 사용하는 툴로는 Slack, Apache Kafka, PagerDuty, Trello 등이 있습니다.



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

성공적인 DataOps의 영향

한 소매업체는 DataOps를 채택해 데이터 파이프라인 전반을 개선하여 이전에 3주나 걸리던 데이터 변경 작업을 2분 이내에 조직 전체에 적용할 수 있었습니다. 그 결과 이 소매업체는 즉시 활용 가능한 비즈니스 데이터를 활용하여 이전에 20일 소요되었던 고객 선호도 분석 프로세스를 하루 안에 수행할 수 있었습니다. 더 나아가 인벤토리 재고 위치 보고에 필요한 시간을 1/6로 줄였습니다.

성공적인 DataOps 적용의 기준은 다음과 같습니다.

1. **데이터 오피스 설립:** 이 프로세스에는 데이터를 बैं크에 리소스로 제공할 때 역할의 범위를 명확히 정의하고, 주요 이해 관계자를 식별하여, 데이터 파이프라인의 모든 이해 관계자가 협업 작업과 문화를 달성하기 위해 해야 할 노력을 이해하는 것이 포함됩니다.
2. **비즈니스 목표와 일치화:** 시장에서 경쟁력을 유지하려면 정보와 데이터 기반 접근 방식을 통해서만 제공할 수 있는 새로운 기회에 대한 신속하게 대응할 수 있어야 합니다. 한 마디로, 비즈니스와 데이터 전달 사이에 원활한 의사소통이 이루어지지 않는다면 경영진은 조직이 성장하지 못할 것을 알게 됩니다.
3. **성공적인 데이터 운영 확대:** 모든 데이터 주도 이니셔티브에서 리더는 생성된 데이터를 지속적으로 사용 및 재사용할 수 있도록 보장해야 하며, 매번 데이터의 가치가 커져야 합니다. 중앙에서 공유되고 검색 가능하며 비즈니스 언어와 일치할 때만 이 목표를 달성할 수 있습니다.



결론

DataOps를 성공적으로 도입한 조직은 어떤 데이터 자산을 액세스할 수 있는지 알고, 데이터의 의미와 품질을 신뢰하며 데이터를 최대한 활용합니다. 신뢰할 수 있고 즉시 활용 가능한 비즈니스 데이터로 차별화된 인사이트, 운영 우수성, 협업, 경쟁 우위를 달성할 수 있을 때 데이터는 가치가 있습니다.

DataOps를 달성하려면 다음이 필요합니다.

- 파일럿 프로젝트를 실행하여 조직 고유 역량과 과제를 이해하는데 투자
- 파일럿 프로젝트의 성공을 활용하는 DataOps 기술 및 조직의 확장 성장
- DataOps에 참여할 더 많은 팀을 모집하여 성공을 촉진
- 습득한 교훈 공유 및 DataOps CoE 구축 시작

이제 다음 단계로 나아가 자체 IBM DataOps Garage Workshop을 예약하고 즉시 활용 가능한 비즈니스 데이터 확보를 위한 여정의 속도를 높일 때입니다. 지금 Jiseon.Lim@ibm.com으로 문의하세요.

많은 조직들은 이미 데이터 카탈로그, 데이터 레이크 또는 마스터 데이터 이니셔티브를 활용하고 있다면 DataOps를 도입하기 유리한 위치에 있다는 것을 깨닫고 있습니다. ibm.com/DataOps를 방문해 시장을 선도하는 기술을 활용하는 DataOps 지원에 대해 자세히 알아보세요.



© Copyright IBM Corporation 2020
(07326) 서울특별시 영등포구 국제금융로 10
서울국제금융센터(3IFC)

미국에서 제작됨
2020년 1월

IBM, IBM 로고, ibm.com, IBM Cloud, IBM Cloud Garage, IBM Watson 및 InfoSphere는 전 세계 여러 관할 구역에 등록된 International Business Machines Corp.의 상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 기타 회사의 상표일 수 있습니다. IBM 상표의 현재 목록은 ibm.com/legal/copytrade.shtm의 '저작권 및 상표 정보' 웹에서 볼 수 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 등록 상표입니다.

이 문서는 최초 발행일 현재 기준의 내용이며 IBM은 언제든지 이를 변경할 수 있습니다. IBM이 운영되는 모든 국가에서 모든 제한을 이용할 수 있는 것은 아닙니다.

인용된 클라이언트 예제는 설명 목적으로만 제공됩니다. 실제 성능 결과는 특정 구성 및 작동 조건에 따라 다를 수 있습니다. 이 문서의 정보는 상품성, 특정 목적에의 적합성 및 비침해에 대한 보증을 포함하여, 명시적이든 묵시적이든 어떠한 보증도 없이 '있는 그대로' 제공됩니다. IBM 제품은 제공되는 계약 조건에 따라 보증됩니다.

고객은 해당 법률 및 규정을 준수할 책임이 있습니다. IBM은 법률 자문을 제공하거나 해당 서비스 또는 제품이 고객이 법률 또는 규정을 준수하도록 보장한다고 진술하거나 보증하지 않습니다.

우주 보안 관행 선언문: IT 시스템 보안에는 기업 내부의 부적절한 액세스에 대한 예방, 탐지, 대응을 통해 시스템과 정보를 보호하는 것이 포함됩니다. 부적절한 액세스로 인해 정보가 변경, 파괴, 남용, 오용될 수 있으며 다른 사람에 대한 공격에 사용하는 것을 포함하여 시스템이 손상되거나 오용될 수 있습니다. 어떤 IT 시스템이나 제품도 완전히 안전한 것으로 간주되어서는 안 되며 어떤 단일 제품, 서비스 또는 보안 조치도 부적절한 사용이나 액세스를 방지하는 데 완전히 효과적일 수 없습니다. IBM 시스템, 제품, 서비스는 합법적이고 포괄적인 보안 접근 방식의 일부로 설계되었으며, 여기에는 반드시 추가 운영 절차가 필요하며 가장 효과적인 다른 시스템, 제품 또는 서비스가 필요할 수 있습니다. IBM은 시스템, 제품 또는 서비스가 당사자의 악의적이거나 불법적인 행위로부터 면역이 되거나 귀하의 기업이 면역이 되지 않을 것이라고 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.

10028810USEN-01

부록: DataOps 파일럿 프로그램 템플릿

프로젝트 이름:

날짜

부서 또는 단위:

파일럿 프로그램 리더:

이름	역할	이메일	전화

회사 전반의 추가 이해관계자

이름	역할	이메일	전화

문제 설명:

근본 원인 체크리스트:

문제	적용 가능 여부 (예/아니오)	추가 참고 사항

성공 지표:

시작 날짜:

스프린트 종료 날짜:

평가:

이행 사항	현재 상태	목표 스프린트 결과	목표 결과를 달성하기 위한 조치
데이터 자산 수집, 자동 검색 및 분류			
데이터 품질 평가 및 수정			
비즈니스 용어 할당			
데이터 개인정보 보호, 규정 준수 및 기업 정책 정의 및 시행			
데이터 사용자의 요구사항 정의 요청 처리			
예외 및 오류 처리 및 수정을 포함한 데이터 요청 커뮤니케이션 및 알림			
카탈로그에 업선된 데이터 게시			
데이터 계보 및 보고			
협업, 피드백, 감사			

감사 중 생길 수 있는 질문의 예:

- 데이터 자산 수집, 자동 검색 및 분류
 - 중단 없는 업무 운영을 위해 대기 시간이 짧은 대용량 복제를 수행하나요?
 - 실시간, 로우 레이턴시 분석을 위해 고급 스트리밍 분석을 사용하나요?
- 모든 데이터 소스에 쉽게 연결하고 복잡한 데이터 변환 및 통합을 수행하나요?
- 소셜 미디어, 날씨 데이터 또는 기타 공용 클라우드 데이터 소스에서 얻은 데이터를 제공할 수 있나요?
- 데이터 사용자가 모든 데스크톱 애플리케이션에서 메타데이터 저장소에 실시간으로 액세스할 수 있나요?
- 데이터 사용자가 데이터 카탈로그에 실시간으로 액세스하여 작업과 관련된 데이터 세트를 검색하고 셀프 서비스 및 지원을 이용할 수 있나요?
- 데이터 프로파일링 툴을 사용하여 데이터를 이해하고, 데이터 값, 열, 테이블 관계를 검증하고 이상 징후를 찾고 분석하나요?
- 비즈니스 규칙 관리가 메타데이터 관리 인프라와 통합되어 있나요?

