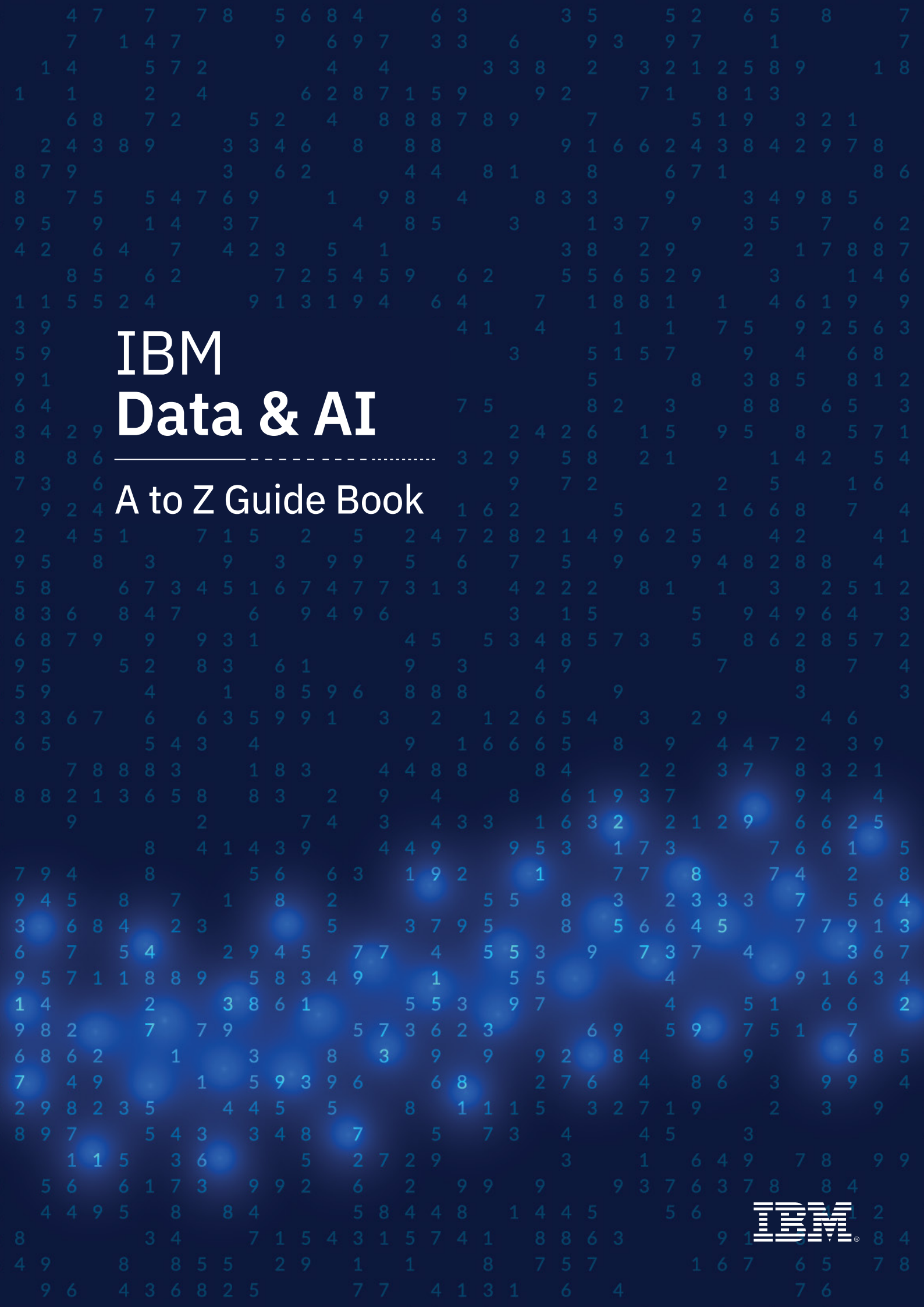
IBM
Data & AI

A to Z Guide Book



Contents

1. 인사이트 ---- 04

기업의 성공을 위한 데이터 혁신 ---- 04

CXO 를 위한 AI 관리 ---- 17

디지털 혁신을 위한 전략 가이드 ---- 25

2. IBM 솔루션 ---- 40

IBM AI 사다리 방법론 ---- 41

현대화 ---- 49

수집 & 구성 ---- 59

분석 ---- 63

적용 ---- 73

3. 고객 사례 ---- 76

1. 인사이트

기업의 성공을 위한 데이터 혁신 ---------- 04

CXO 를 위한 AI 관리 ---------- 17

디지털 혁신을 위한 전략 가이드 ---------- 25

기업의 데이터 혁신

성과를 지향하는 귀사의 비즈니스

성과 실현을 위한 전략

이제 기업에서는 현대화된 엔터프라이즈 분석의 시대에 적응해야 합니다. 개별 솔루션을 통해 특정 활용 사례를 해결하더라도, 전사적 차원에서 더 빨리 성과를 내기 위해서는 전체론적 관점의 사고가 필요합니다.

분석 에코 시스템의 구조와 현대화 전략을 연계함으로써 부분의 합보다 큰 시너지 효과를 내면서 성과 실현에 속도를 낼 수 있습니다.

성과 지향

수집(Collect)

소스의 데이터에 연결

현 상황 : 분석의 복잡성

분석의 복잡성 증대

1. 클라우드, 오픈소스, NoSQL 등 새로운 데이터 소스가 등장하면서, 서로 연결되지 않은 데이터 저장소가 많아졌습니다. 저마다 다른 데이터를 수용하면서 사용자를 위한 가용성에 걸림돌이 되곤합니다.
2. 비즈니스 사용자와 데이터 분석가는 데이터에 액세스해야 합니다. 그러나 데이터가 여러 환경에 흩어져 있으므로, 대개는 이 데이터를 이동해야 합니다. 다양한 데이터 소스를 통합하기 위해 사용자의 노트북으로 데이터를 추출해 오는 것이 일반적인 방법입니다.
3. 데이터가 여러 개별 시스템에 흩어져 있으므로, 사용자가 데이터를 찾아내고 분석에 사용 중인 데이터의 무결성을 보장하는 데 어려움이 따릅니다.

현 상황의 결과

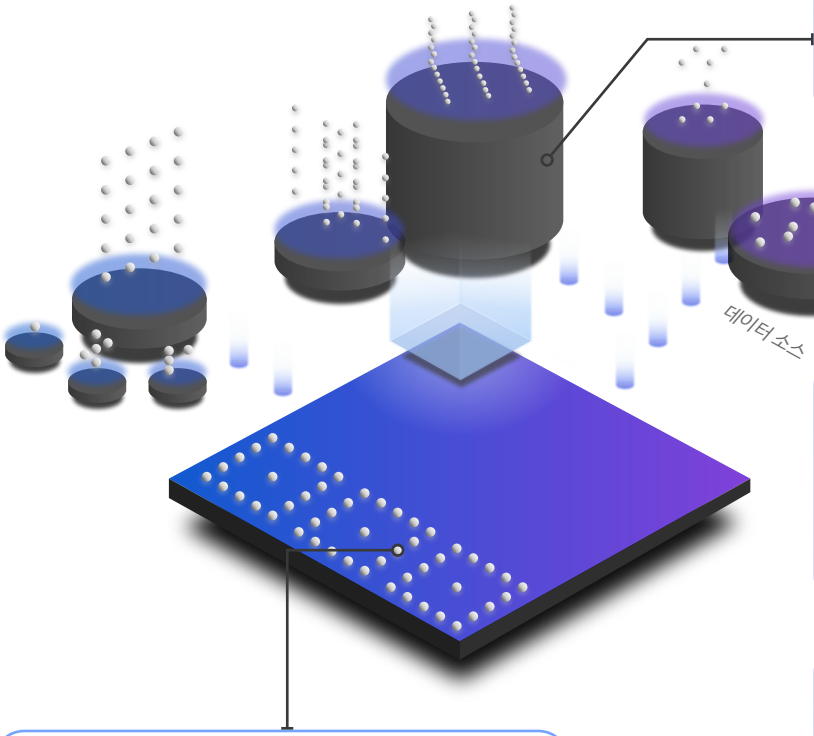
1. 비즈니스 사용자가 데이터를 찾고 각종 데이터 소스로부터 특정 병합 데이터 웨어하우스 또는 테이블을 생성하려면 IT 팀의 도움이 필요합니다. 상당한 인력이 필요하고 속도도 느립니다.
2. 데이터 이동은 보안 위험도 야기합니다. 데이터가 최종 소비자의 노트북으로 전송되고, 특별한 규제도 받지 않는다면 더욱 그렇습니다.
3. 데이터 조합 및 작업은 대개 수작업으로 이루어집니다. 이 경우 데이터의 무결성이 손상될 때가 많아 결국 데이터 및 그 데이터를 기반으로 한 분석에 대한 신뢰성 문제로 이어집니다.

분석 간소화

미래의 전략 : 수집

소스의 데이터에 연결

비즈니스 및 분석 사용자가 거버넌스 기반 셀프 서비스 방식으로 모든 데이터를 사용하면서 더 빠른 인사이트를 생성하고 비용 및 위험을 줄이게 합니다.



컴퓨팅 활용
데이터 페더레이션(federation) 방식보다는 저장소에서 분석을 실행하고 결과를 가져오게 합니다.

엔터프라이즈 가시성
모든 엔터프라이즈 데이터를 이동하지 않고 통합 모니터링하는 단일 뷰를 생성합니다. 전 세계에 분산된 여러 영구 저장소에 대해 유비쿼터스 액세스를 제공하므로, 하나의 저장소처럼 보입니다.

데이터의 100% 활용
모든 데이터를 현재 저장된 위치에서, 필요할때 액세스합니다. 중복되지 않은, 최신 버전의 데이터로 유지합니다.

IT/데이터 거버넌스 연계
최종 사용자가 셀프 서비스 데이터로 소비할 수 있도록 데이터 전 범위에, 서비스 프레임워크의 형태로 정책 및 표준을 적용합니다.

730만 달러
예상 편익
(TEI 연구)

65%
ETL 작업 감축
(TEI 연구)

430%
페더레이션보다 빠른 속도
(TEI 연구)

그리고 투명성까지!

흩어져 있는 사일로화된 데이터를 캡슐화하여 단일 가상 데이터 레이크에서 관리합니다. 기존 ETL 전략은 너무 느리고 비용이 많이 들어 데이터 소스의 급증 및 엔지니어에 대한 수요를 쫓아가기가 어렵습니다. **기업의 모든 데이터**를 수집하는다는 목표 아래 적응형 경로를 마련해야 합니다.

분석 간소화

IBM이 지향하는 투명성의 의미 :

사용자끼리 자산 또는 동료가 구현한 결과물을 재활용하여
조직 차원의 가치를 더 빨리 실현하게 해주는 협업 공간

페더레이션보다는 가상화를 통해 데이터 이동을 최소화할 뿐만 아니라 모든
영구 저장소의 컴퓨팅 리소스를 활용하여 인사이트 생성 가속화

지능형 데이터 카탈로그를 통해 모든 유형의 사용자가 조직의
모든 영구 저장소에 있는 정보에 액세스 가능

정책 엔진을 통해 자산에 대한 액세스를 제어할 뿐만 아니라 자산에 수록된
정보의 삭제/난독 처리를 통해 최고 수준의 규정 준수 및 자산 액세스 보장

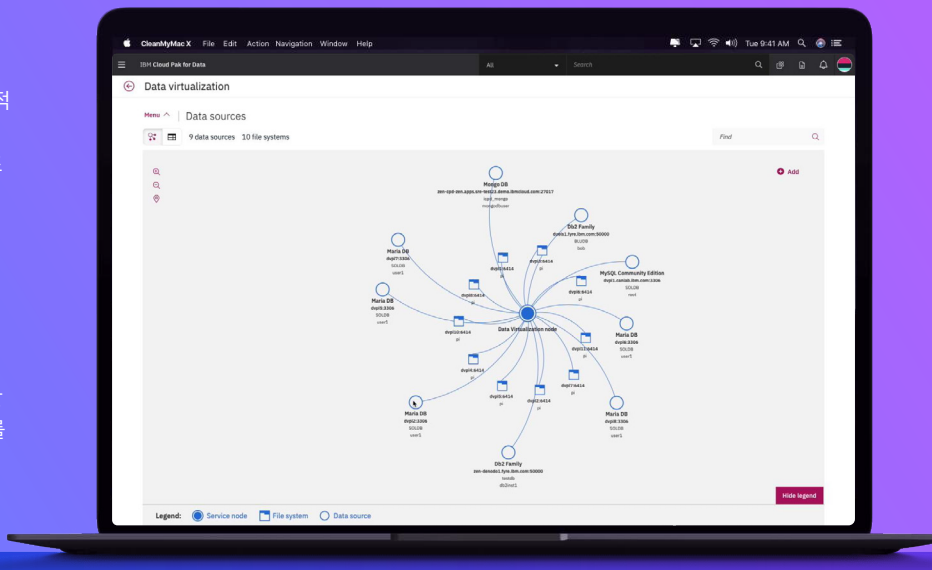
진정한 투명성의 기준입니다.

IBM의 제공 방식 :

첫째. 모든 데이터 자산을 소스에서 크롤링하고 분류합니다. 따라서 전사적 범위의 모든 자산을 빠짐없이 모니터링할 수 있습니다. 그런 다음 메타데이터를 사용자에게 전달합니다. 사용자는 기존 엔터프라이즈 데이터를 AI에 활용할 때 이를 길잡이로 활용할 수 있습니다.

둘째. 행 및 열 단위의 세분화된 수준으로 정책을 적용하여 자산의 올바른 사용을 보장합니다. 이러한 기능은 IBM 연구소에서 구현한 ML 기반 정책 엔진에서 제공합니다.

셋째. 데이터 소스에서 분석을 수행하여 데이터 이동을 최소화하는 한편 컴퓨팅 리소스를 최대한 활용합니다. 이 모든 것을 주관하는 AI 기반 엔진은 어디서 무엇을 프로세싱해야 효율성을 극대화할 수 있는지를 자동으로 결정합니다.



왜 IBM인가?

데이터 가상화와 지능형 카탈로그의 조합을 통해 기업의 모든 데이터 소스를 연결할 뿐만 아니라 중요 자산에 대한 올바른 거버넌스를 보장합니다.

독창적인 자율 조직형 P2P(self-organizing peer to peer) 아키텍처를 제공하므로, 사용자의 필요에 따라 애플리케이션에서 여러 데이터베이스, 플랫 파일, 빅데이터 저장소를 개별적으로, 또는 종합적으로 쿼리할 수 있습니다.

애플리케이션에서는 모든 데이터 저장소를 하나의 소스로 인식하므로 전사적 범위에서 장벽이 해소됩니다.

IBM만의 차별화 요소

IBM 고유의 특별한 차별화 요소인 데이터 카탈로그를 통해 모든(정형 및 비정형) 영구 저장소를 크롤링한 다음 세분화된 액세스 권한에 따라 정책을 적용하여 기업 사용자가 중요 정보를 오용하지않게 합니다.

컴퓨팅 가상화 메시를 제공하여 MPP(Massively Parallel Processing)를 통해 성능을 획기적으로 높임과 동시에 저장소 간 데이터 이동을 최소화합니다.

경쟁 차별화 요소

모든 데이터 저장소를 크롤링하는 기능으로 Forrester에 의해 최고의 데이터 인사이트 플랫폼에 선정된 지능형 데이터 카탈로그가 세부적인 정책 제어를 수행하면서 사용자를 위한 “데이터 쇼핑(shop for data)” 접근법을 제공합니다.

컴퓨팅 메시(가상화)를 통해 성능을 획기적으로 높이고 데이터 이동을 최소화합니다. 게다가 근본적으로 페더레이션보다 더 뛰어난 효율성을 발휘하는데, 이는 페더레이션 노드로 데이터를 취급하지 않고 소스에서 데이터를 처리하기 때문입니다.

전체 차별화 요소

IBM은 340억 달러 규모의 RedHat 인수를 통해 본격적으로 하이브리드-멀티클라우드 접근법을 지원하고 있습니다.
RedHat OpenShift를 사용하여 기업이 방화벽 내에서, 또는 어떤 클라우드에서든 엔터프라이즈급 애플리케이션을 배포할 수 있게 합니다.

IBM 플랫폼에서는 마이크로서비스 지향 아키텍처를 통해 필요한 서비스만 강력히 통합된 형태로 배포할 수 있습니다. 이 기능은 가장 유연한 방식으로 아키텍처를 현대화 합니다.

성과 지향

구성(Organize)

DataOps 거버넌스 기반

현 상황 : 각종 장애물

각종 장애물 발생

1. 대부분 기업에서 엔터프라이즈 환경의 데이터 저장소에 대한 잠금을 설정하고 사용자의 액세스를 제한하기 위한 거버넌스 제어 기능을 구현했습니다. 그로 인한 데이터 액세스 제약으로 분석 및 AI가 최적화되지 못하고 있습니다.
2. 이와 같이 사일로화된 곳이 대부분이므로, 표준화된 뷰에서 데이터 품질 및 데이터 계보를 관리하기 어렵습니다. 결국 상황에 따라 많은 비용을 들여 AI에 적합한 데이터로 만들어야 합니다.
3. 복잡한 통합 빌드 및 긴 데이터 액세스 요청 처리 시간으로 IT 팀의 부담이 커지고, 이는 비즈니스 사용자 불만, 애질리티 부재, IT 지출 급증 등의 문제를 야기합니다.

현 상황의 결과

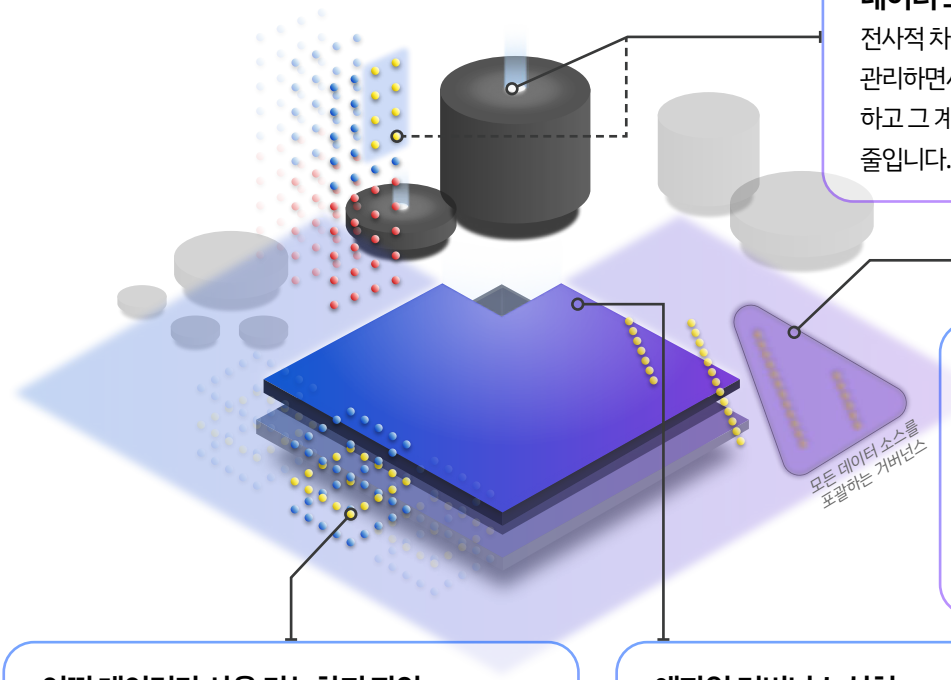
1. 중요 데이터 자산에 대한 액세스 제약에 따른 부정확하거나 불완전한 분석/AI 때문에 상당한 수익 손실이 발생합니다.
2. IT 비용은 꾸준히 늘고 있습니다. 비즈니스 팀에서 인사이트 발굴을 위한 데이터 액세스를 요청하는데도 불구하고, IT 팀은 최적의 수준이 아닌 통합을 제공하고 배포 시간도 길어질 수 밖에 없기 때문입니다.
3. 전사적 범위에서 협업이 저조합니다. 각 그룹이 저마다의 분석 및 AI를 가동하면서 독자적으로 움직이기 때문에 시간 및 리소스가 낭비됩니다.

셀프 서비스 실현

미래의 전략 : 구성

DataOps 거버넌스 기반

DataOps 방식으로 엔터프라이즈 환경에서 더 신속하게 가치를 창출합니다. 이렇게 품질 높은 자산을 기반으로 한 분석 및 AI를 통해 정확도를 높이고, 배포 시간을 단축하며, 전사적으로 위험을 줄일 수 있습니다.



데이터 소스까지 역추적

전사적 차원에서 데이터 계보 및 데이터 품질을 관리하면서 우수한 품질의 자산이 AI에 활용되게 하고 그 계보를 철저히 추적하여 기업의 위험 부담을 줄입니다.

데이터 보호

실행 가능한 정책 제어를 통해 중요 정보에 대한 액세스를 제한할 뿐만 아니라 정책 엔진을 기반으로 한 추천 엔진으로 사용자가 필요한 자산을 필요한 시점에 사용할 수 있게 합니다.

어떤 데이터가 사용 가능한지 파악

조직의 모든 자산을 한군데서 통합 모니터링하면서 사용자가 분석 및 AI를 빌드할 때 필요한 데이터를 '쇼핑'할 수 있게 합니다.

애자일 거버넌스 실현

데이터 정제(refinement), 품질, 계보를 통합 관리하여 데이터 운영을 간소화합니다. 더 반복적인 애자일 방식으로 DataOps 원칙을 적용합니다.

1,500만 달러

기업에서 부실한 데이터 품질 때문에 떠안는 연간 손실
(Gartner)

신뢰 - 필수 조건

사람, 프로세스, 기술을 최적화하여 신뢰할 수 있는 고품질의 데이터를 팀에서 필요한 시점에 제공합니다. 데이터를 검색, 액세스, 준비, 통합하고 모든 데이터/AI 이니셔티브에 사용할 수 있도록 제공하는 것과 관련된 비효율성을 해결합니다.

셀프 서비스 실현

IBM이 지향하는 신뢰성의 의미 :

지능형 데이터 카탈로그를 통해 모든 유형의 사용자가 조직의 모든 영구 저장소에 있는 정보에 액세스할 수 있게하고, 정책 엔진을 활용하여 자산에 대한 액세스를 제어합니다. 아울러 자산에 수록된 정보를 삭제/난독 처리하여 최고 수준의 규정 준수 및 자산 액세스를 보장할 수 있습니다.

데이터 자산에 포함된 콘텐츠를 자동으로 분류하여 중요 정보를 올바르게 표시하고, 정책에 따라 액세스 제한을 적용하거나 기밀 정보를 마스킹합니다.

데이터 사이언티스트, 데이터 엔지니어, 시민 분석가, 파워 유저 등 조직의 모든 페르소나에게 단일 통합 사용자 경험을 제공합니다.

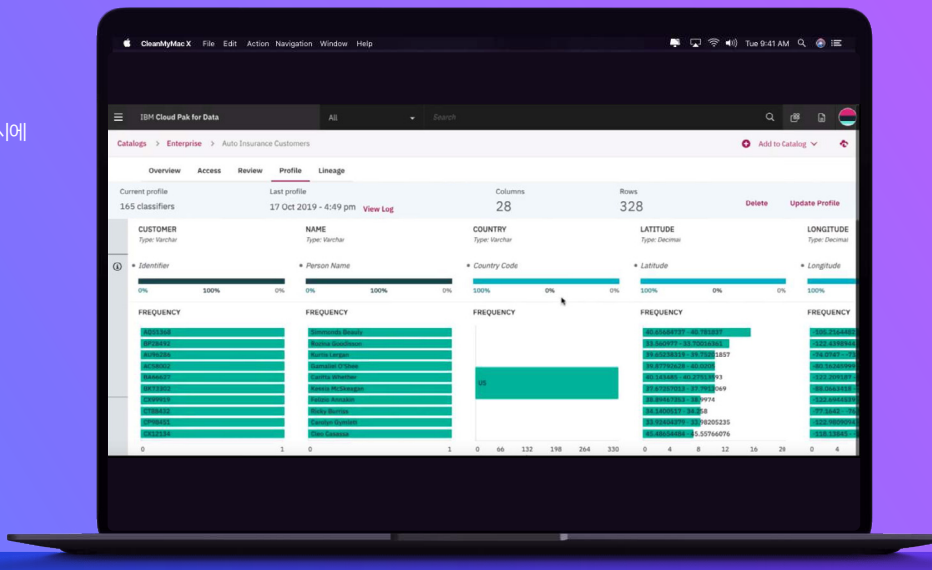
진정한 신뢰성의 기준입니다.

IBM의 제공 방식 :

첫째. 조직의 모든 데이터 자산을 크롤링하고 분류한 다음 사용자에게 메타데이터 뷰를 전달합니다. 정책을 적용하여 오용을 방지함과 동시에 자산에 대한 더 강력한 액세스를 지원합니다.

둘째. 모든 자산을 대상으로 품질을 지정하고 계보를 추적하며, 사용자가 데이터의 품질 및 출처를 확인할 수 있게 합니다.

셋째. AI 엔진을 활용하여 사용자에게 데이터 자산을 추천하고, 데이터 디스커버리 프로세스의 속도를 높여 AI 구축으로 순조롭게 진행할 수 있게 합니다.



왜 IBM인가?

정형/비정형 데이터를 자동으로 프로파일링하고 분류하여 사용자가 데이터 자산과 그 계보를 이해할 수 있게 합니다.

내장된 데이터 정제 기술을 통해 대화형 방식으로 데이터를 탐색, 정리, 준비합니다. 데이터 프로파일 시각화, 기본 제공된 차트, 통계 기능을 통해 데이터 품질 및 데이터 계보/분포를 이해합니다.

IBM만의 차별화 요소

Watson 및 조직 내 동료 구성원의 추천을 지능적으로 제공하면서 사용자가 더 유익한 자산을 더 빨리 찾도록 지원하는 곳은 IBM뿐입니다.

오로지 IBM에서만 자동으로 중요 데이터 요소를 동적 마스킹하여 데이터 오용을 방지하고 자신 있게 자산을 공유할 수 있게 합니다. 그리고 AI를 기반으로 한 능동적 정책 관리를 통해 거버넌스를 이행합니다.

경쟁 차별화 요소

사용자는 영구 데이터 저장소, BI 툴, 콘텐츠 리포지토리 등 40여 개 데이터 소스에 연결하여 사내의 어떤 데이터도 찾아 사용할 수 있습니다.

클라우드, 온프레미스, 하이브리드 저장소의 전 범위에서 AI를 활용하여 자동으로 자산을 분류하므로, 사용자가 데이터의 위치에 구애받지 않으면서 액세스하고 활용할 수 있습니다.

전체 차별화 요소

IBM은 340억 달러 규모의 RedHat 인수를 통해 본격적으로 하이브리드-멀티클라우드 접근법을 지원하고 있습니다.
RedHat OpenShift를 사용하여 기업이 방화벽 내에서, 또는 어떤 클라우드에서든 엔터프라이즈급 애플리케이션을 배포할 수 있게 합니다.

IBM 플랫폼에서는 마이크로서비스 지향 아키텍처를 통해 필요한 서비스만 강력히 통합된 형태로 배포할 수 있습니다. 이 기능은 가장

성과 지향

분석(Analyze)

DataOps 거버넌스 기반

현 상황 : 부정확한 ML/AI 구축

부정확한 ML/AI 구축

1. 사용자가 데이터/ML 자산을 찾아 재사용할 수 없고, 사용자 간 협업이 거의 또는 전혀 이루어지지 않으며, 사용자의 스킬 부족으로 AI 구축에 참여하지 못합니다.
2. AI 모델에 대한 책임 소재를 규명할 수 없고, 모델이 특정 결론에 도달한 이유 및 모델 편향 여부를 알지 못하며, AI 모델의 편향 또는 부정확한 결정을 바로잡을 수 없습니다.
3. 확장 가능한 방식으로 AI 모델을 배포하고 관리하기란 거의 불가능합니다. MLOps 유형 접근 방식의 부재로 버전 관리는 물론이고 모델을 이해하거나 프로덕션에서 모델을 승격/강등할 수 없습니다. 프로덕션 애플리케이션에 모델을 배포하는 것도 용이하지 않습니다.

현 상황의 결과

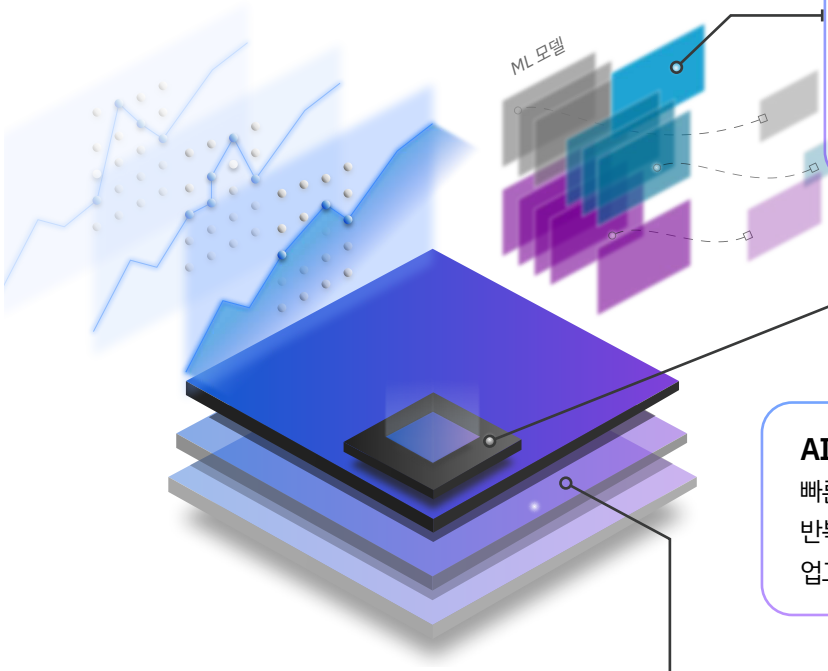
1. 기업에 스킬이 뛰어난 데이터 사이언스 인력이 있더라도 ROI를 거의 또는 전혀 거두지 못합니다. 모델 대부분이 프로덕션으로 가지 못하고, 설령 가더라도 신뢰성 부족으로 사용하지 못합니다.
2. 기업의 페르소나 대다수는 스킬 부족 때문에, 또한 모든 스킬 레벨을 수용하는 통합 환경의 부재로 인해 AI 구축 및 배포에 참여하지 못합니다.
3. 모니터링되지 않는 모델, 편향된 모델, 신뢰성이 떨어지는 모델에 대한 규제와 처벌이 엄격하여 그로 인한 불이익 또는 손실로 ROI가 상쇄되거나 모델 배포의 위험 부담이 너무 큽니다.

AI 확산

미래의 전략 : 분석

MLOps 전략 기반

모든 스킬 레벨의 사용자가 AI 구축 및 배포에 참여하게 해주는 통합 툴세트를 통해 AI 모델의 배포 속도를 개선하고 AI 모델의 정확성 및 신뢰성을 강화합니다.



강력한 협업

사용자는 원하는 ML 자산, 즉 데이터 자산을 확보하고 이를 토대로 협업할 수 있습니다. 게다가 모든 직원이 스킬 레벨에 구애받지 않고 AI 구축 및 배포에 참여할 수 있습니다.

AI 애플리케이션에 직접 반영

빠른 속도와 애자일 방식으로 ML 모델 및 분석을 반복하면서 엔터프라이즈 AI 애플리케이션을 업그레이드하거나 새롭게 개발합니다.

확고한 신뢰성

신뢰도 낮은 AI 모델로 인한 무거운 처벌 및 기타 부정적인 영향을 방지하고, 불리한 결과에 대한 걱정 없이 모델을 프로덕션 애플리케이션에 배포할 수 있게 합니다.

320만 달러

3년 예상 PV

+6%

통합 툴 세트 사용 시 모델 정확성 향상

정확성이 곧 가치와 속도

사람, 프로세스, 기술을 최적화하여 신뢰할 수 있는 고품질의 데이터를 팀에서 필요한 시점에 제공합니다. 데이터를 검색, 액세스, 준비, 통합하고 모든 데이터/AI 이니셔티브에 사용할 수 있도록 제공하는것과 관련된 비효율성을 해결합니다.

분석 간소화

IBM이 지향하는 정확성의 의미 :

데이터를 찾고 AI 모델을 구축하며 신뢰할 만한 AI 모델을 배포하는 전 과정에서 일관성 있고 강력하게 통합된 사용자 경험을 제공합니다.

AI를 통해 AI를 개발하는 기능으로 데이터 사이언티스트는 물론 다른 기술 레벨의 사용자도 더 신속하게 빌드하면서 타임라인을 효율적으로 활용할 수 있습니다. 다양한 빌드 옵션, 이를테면 코딩, 캔버스 접근법, 클릭 스루 GUI 등을 지원합니다.

설명 가능한 모델을 구축하고 편향을 추적하므로, 모델을 프로덕션에서 사용하더라도 비윤리적 결정을 내리지 않습니다.

진정한 정확성의 기준입니다.

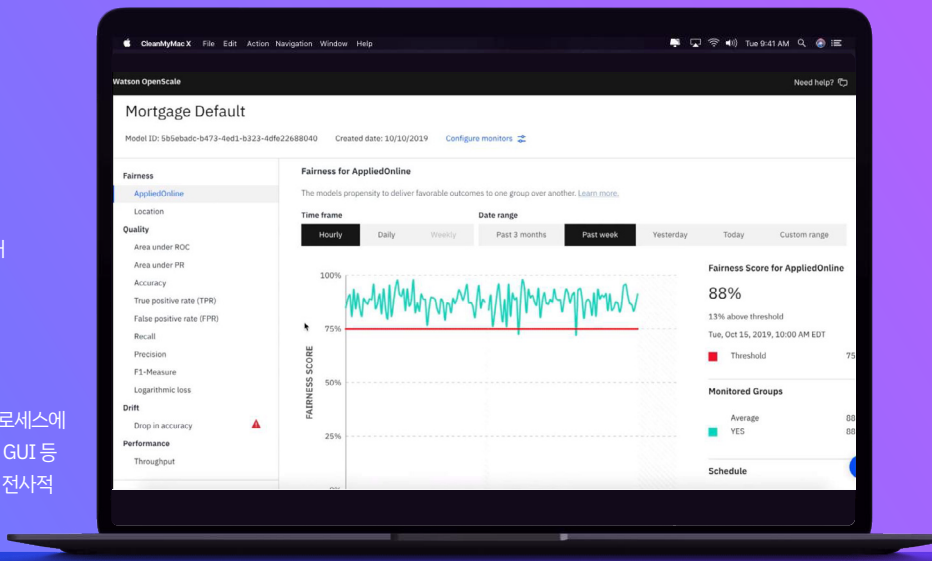
IBM의 제공 방식 :

첫째. 단일 콘솔에서 자산 검색, AI 구축, AI 배포, AI 모니터링을 모두 해결하면서 통합의 걸림돌 없이 성공적으로 배포할 수 있게 합니다.

둘째. IBM은 대조적 섭동(contrastive perturbation)이라는 기법을 통해 모델을 깊이 탐구하면서 어떤 결정을 내리게 된 이유를 확인합니다. 이 기능은 비즈니스 사용자가 접하는 인터페이스에서 제공되어 이해 관계자가 현재 조회 중인 메트릭을 제대로 이해하도록 돕습니다.

셋째. IBM은 AI 구축에 AI를 활용하면서 스킬 레벨에 상관없이 자동으로 ML 모델을 구축합니다.

넷째. 다양한 툴세트를 활용하면서 어떤 스킬 레벨의 사용자도 AI 구축 프로세스에 참여할 수 있게 합니다. 코딩, 드래그 앤 드롭 캔버스 방식, 클릭 스루 GUI 등 다양한 방식이 지원됩니다. 따라서 조직에 속한 모든 사용자가 AI를 전사적 범위로 확장하는 데 기여할 수 있습니다.



왜 IBM인가?

자동 데이터 준비, 피처 엔지니어링, 하이퍼 파라미터 최적화, 앙상블 기법을 통해 클릭 한 번으로 최상의 모델을 선택하고 배포할 수 있습니다.

비즈니스 결과를 활용하는 연속 피드백 루프를 통해 AI 성과를 유지하고 개선합니다.

IBM만의 차별화 요소

프로덕션 AI의 성능 및 비즈니스 목표에 미칠 영향을 단일 콘솔에서, 실행 가능한 메트릭으로 추적할 뿐만 아니라 성능, 편향 및 설명 가능성에 대한 위험 모델도 모니터링합니다. 따라서 규제를 위반할 위험이 최소화되고, 고객에게 더 공정하고 설명 가능한 결과가 나옵니다.

자동 데이터 준비, 피처 엔지니어링, 하이퍼 파라미터 최적화, 앙상블 기법을 통해 클릭 한 번으로 최상의 모델을 선택하고 배포할 수 있습니다.

경쟁 차별화 요소

오픈소스 툴(예: Jupyter Notebook)을 협업 프로젝트 기능 (예: 프로그래밍 기반 액세스 및 버전 관리)과 함께 사용하면서 프로덕션에서 사용할 만한 이상적인 모델이 나올 때까지 매개변수를 정밀 조정합니다.

드래그 앤 드롭 툴을 사용하여 데이터를 탐색하고 머신러닝을 활용합니다. 그와 더불어 Neural Network Modeler에서 코딩이 필요 없는 인터페이스, 즉 직관적인 드래그 앤 드롭 방식으로 딥러닝을 위한 시각적 프로그래밍도 지원합니다.

전체 차별화 요소

IBM은 340억 달러 규모의 RedHat 인수를 통해 본격적으로 하이브리드-멀티클라우드 접근법을 지원하고 있습니다. RedHat OpenShift를 사용하여 기업이 방화벽 내에서, 또는 어떤 클라우드에서든 엔터프라이즈급 애플리케이션을 배포할 수 있게 합니다.

IBM 플랫폼에서는 마이크로서비스 지향 아키텍처를 통해 필요한 서비스만 강력히 통합된 형태로 배포할 수 있습니다, 이 플랫폼은 가장 유연한 방식으로 아키텍처를 현대화하게 해줍니다.

CxO를 위한 최신 AI 활용 가이드: 규모에 맞게 성장 가속화

플랫폼 시대에 획기적인
AI 투자 수익 창출

IBM

요약

인공 지능(AI)의 도입은 이제 선도적인 조직이 눈부신 결과를 실현하고 마켓플레이스를 재편성하며 업계 경쟁자와 자사를 차별화하는 변곡점이 되었습니다. AI의 중심에는 자동화, 예측, 최적화라는 전략적인 동력원이 있습니다. 조직이 성장하려면 일상적인 작업을 자동화하고 결과를 예측하며 리소스를 최적화하는 능력이 필수적입니다. 실제로 성장 가도를 달리는 기업의 경우 비즈니스의 필수 요소, 즉 우수한 고객 경험, 신속한 제품 및 서비스 출시, 원활한 운영, 에코시스템 활용을 모두 갖추었을 뿐만 아니라 규모에 맞는 규제 준수와 위험 관리 요구사항도 충족하고 있습니다.

이 문서에서 다룰 주제는 다음과 같습니다.

– 빠르게 성장하는 선도 기업들의 AI 투자 특성
– 데이터 및 AI 플랫폼이 필요한 이유
– 데이터 및 AI 플랫폼 활용의 목적: 자동화, 예측, 최적화
– 신뢰와 투명성이 뒷받침되는 AI 구축 및 확장의 효과

이 문서에서는 업계 선두 주자들이 AI의 장점을 활용하는 방법과 플랫폼 접근법의 중요성, 그리고 데이터 및 AI 플랫폼이 선사하는 장점에 대해 설명합니다. 또한 도입할 수 있는 방법들을 중점적으로 알아보고 비즈니스의 성공에 도움을 줄 전략을 안내합니다.

AI가 가져오는 기회 및 해결 과제

기업들은 비즈니스 성과 예측, 운영 간소화, 효율성 향상, 사이버 위험과 범죄로부터의 보호, 새로운 시장 기회 발견을 위해 AI를 활용합니다. 이러한 예측을 통해 선도 기업들은 경쟁과 시장 변동성에 맞설 돌파구를 마련할 수 있습니다. 또한 기업 임원들은 주주의 기대를 만족시키는 동시에 AI 구현 시 공정, 기술, 조직에 대한 획기적인 변화를 실현해야 한다는 부담을 느낍니다.

나아가 이사회 수준에서 이행하는 데이터에 대한 정밀 조사와 AI 모델과 관련된 거버넌스를 거쳐야 합니다. IBM® Institute of Business Value에서 발표한 Csuite Study[1] 중 브랜드명 및 기관에 대한 고객의 신뢰를 조사한 "Build Your Trust Advantage—Leadership in the era of data and AI everywhere"가 빠르게 주목받고 있습니다. 고객은 제품 및 서비스 관련 데이터의 투명한 관리를 요구하며, 개인 데이터의 안전한 보관 및 합법적인 사용이 보장되기를 원합니다.

"데이터 거버넌스와 AI 모델 검증과 사용 방법에 대한 거버넌스가 이제는 이사회의 관심이 되고 있습니다. 데이터의 윤리적 사용도 마찬가지입니다."[1]

네덜란드 금융 분야
COO

고속 성장 중인 선도 기업들의 성공에 기여하는 AI

Forrester Research에 따르면 빠르게 성장하는 선도 기업은 AI에 대규모로 투자중입니다. Forrester 설문 조사 응답자의 절반 이상은 AI를 통해 5배 이상의 투자 수익을 실현할 예정인 것으로 나타났습니다.[2] 이를 고려할 때 고도 성장 중인 선도 업체가 1천만 달러를 투자했을 경우 6천만 달러에 이르는 ROI를 기대할 수 있습니다. 또한 선도 업체는 성장률이 낮은 기업에 비해 데이터 및 분석 예산은 2배, AI 및 머신 러닝(ML) 플랫폼에는 2.5배 더 투자합니다.

Forrester Research는 또한 예측 기반 학습, ML, 딥 러닝, 자연어 처리(NLP), 컴퓨터 시각(computer vision) 및 기타 모델 유형과 같은 핵심 기술을 연구하는 데이터 과학자 육성에 투자하는 기업이 그렇지 않은 기업보다 빠르게 성장한다는 사실을 알아냈습니다.

Forrester 조사에서는 인사 관리, 잠재 고객 평가, 비용 관리 및 부정 행위 감지를 비롯한 일부 AI 애플리케이션 분야에서 조직은 패키지 형태의 AI 솔루션 구매를 선호하는 것으로 나타났습니다.

– 46%는 일부 애플리케이션에서 AI가 기본 제공되는 패키지형 솔루션을 구입합니다.

– 20%는 기업 자체적으로 AI를 개발합니다.

실제로 기업에서는 비즈니스 요구사항에 따라 쉽고 빠르게 구매하고 구축할 수 있는 데이터 및 AI 플랫폼이 필요합니다.

데이터 및 AI 플랫폼 평가 시 고려해야 할 요건

효과적인 AI 활용의 핵심은 데이터 과학, 즉 기업이 의미 있는 패턴을 파악하고 결과를 예측하며 의사결정을 간소화할 수 있도록 하는 규칙입니다. 데이터에서 얻은 새로운 통찰, 패턴, 그 외에 새로 발견한 가치 있는 요소들을 활용하면 다음 단계를 예상하고 의사결정을 간소화할 수 있습니다. 엄선된 시나리오를 바탕으로 가능한 한 최고의 결과를 이끌어 내기 위해 올바른 제안과 접근법을 적용해 최적의 실행 방안을 마련할 수 있습니다. 이러한 잠재력이 바로 아이디어를 실행 방안으로 실현시키고 데이터 과학을 적극적으로 활용할 수 있는 플랫폼이 필요한 이유입니다.

그렇다면 플랫폼이란 무엇일까요? 플랫폼은 "참가자들 간에 가치를 창출하는 인터랙션을 지원하는 인프라입니다. 플랫폼은 해당 인터랙션을 위한 개방형 환경 역할을 하며 이를 위한 거버넌스 조건을 조성합니다."[3] AI로 비즈니스 현대화에 성공하려면 알맞은 플랫폼 선택이 관건입니다.

인재, 데이터, 신뢰 관리를 위한 기반 구축

기업의 목표는 예측 및 최적화 프로세스를 반복적인 혁신과 지능적인 워크플로로 바꾸는 것입니다. 이러한 목표를 이루고 AI가 조직 운영에서 그 능력을 십분 발휘하려면 기업 인력이 개발 자동화를 도와줄 데이터 및 AI 플랫폼에 쉽게 적응할 수 있어야 합니다. 또한 일회성 도구를 추가하는 것이 아니라 새로운 플랫폼을 통해 기존의 기술 투자를 활용하는 방법도 찾아야 합니다.

인재: 기술과 인력 배치는 AI 성공의 핵심 요소입니다. PwC가 진행한 2019 AI Predictions 조사를 통해 나타난 바와 같이 AI에는 능숙한 인력이 필요하며, 이를 위해서는 기업 인재가 보유한 기술을 재교육하고 향상시킬 수 있는 지속적인 학습 프로그램을 마련해야 합니다. 그리고 직무의 내용이 달라지면서 점점 역량이 향상되는 인력을 도울 조직 구조를 새롭게 구상해야 합니다.[4]

데이터: 데이터는 모델의 성능 측면에서 동력과도 같습니다. AI 모델은 데이터를 기반으로 구축되며, 올바른 운영 데이터에 따라 모델의 성능이 좌우됩니다. 그렇기 때문에 플랫폼은 지속적인 데이터의 흐름에 대한 액세스를 지원해야 합니다. 또한 훈련을 받은 데이터와는 다른 실제 운영 데이터를 접했을 때 모델에서 불가피하게 발생하는 정확도의 감소를 감지 및 완화해야 하며, AI 수명주기 전체에 걸쳐 사용된 모델과 데이터에 대해 감사용 레코드를 마련해야 합니다.

신뢰: 담당 팀은 AI 모델에서 바이어스를 감지 및 완화할 수 있는 방법을 마련하고 개별적인 결과에 대한 근거를 제시할 수 있어야 합니다. 또한 플랫폼은 비즈니스 KPI를 기준으로 한 결과를 추적할 수 있어야 합니다. AI 관련 노력을 확대하고 꾸준히 유지할 수 있도록 신뢰성과 모든 설명 가능성을 준비해 두어야 합니다.

AI 예측 및 최적화를 신속히 애플리케이션에 통합

데이터 및 AI 플랫폼은 AI 모델에서 최신 앱으로 통찰을 통합할 수 있도록 지원해야 합니다. 대부분의 기업은 이미 애플리케이션 개발에 대규모로 투자해 왔습니다. 유연한 개방형 데이터 및 AI는 기업의 애플리케이션 개발 및 비즈니스 팀의 모델 운영(ModelOps) 구축의 토대가 됩니다. ModelOps는 DevOps와의 원활한 협업을 통해 AI 기반의 최신 애플리케이션을 성공으로 이끄는 데 도움이 됩니다.

AI 수명주기 관리

데이터 과학 및 AI 투자는 일반적으로 비즈니스 관련 질문에 대한 해답을 얻고 소규모 프로세스를 자동화하는 데 필요한 예측 기반 분석과 ML 사용 위주로 이루어졌습니다. 하지만 대부분의 선도 기업은 이제 좀더 폭넓게 AI를 활용할 방법을 모색 중입니다. 이를 위해서는 플랫폼이 비즈니스 전체에 걸쳐 시작부터 끝까지 모델 및 도구의 관리를 운영하고 자동화할 수 있도록 설계되어야 합니다.

– 데이터 준비
– 기능 엔지니어링
– 머신 러닝 알고리즘 선택
– 하이퍼파라미터 최적화로 가능한 한 최고의 ML 모델 선택

이러한 일련의 단계를 진행할 때는 프로세스의 각 단계마다 가장 효율적인 단계를 찾아내는 AI 시스템의 안내에 따라야 합니다. 이것이 바로 AI를 통한 AI 구축이며 IBM Research™ 기반 기능인 AutoAI가 바로 그 예시입니다.[5]

예측 결과를 바탕으로 의사결정 최적화

성장 가도를 달리는 선두 기업 중 일부는 AI로 다양한 사용 사례를 연구하고 효과적인 의사결정을 내리는 능력이 탁월합니다. AI의 통찰과 AI 오케스트레이션을 사람의 능력과 결합할 수 있습니다. 기업에서 최고의 ROI를 달성하기 위해서는 데이터 및 AI 플랫폼이 예측 결과를 활용하고 이를 통해 실행 방안을 마련해야 합니다.

최신 데이터 및 AI 플랫폼은 기업에서 팀의 최적화 문제 해결을 위한 데이터 선택 및 편집 워크플로를 간소화할 수 있어야 합니다. 자연어 인터페이스로 팀원들이 최적화 모델을 실행하고 간트(Gantt) 차트가 포함된 보고서, 일정표, 리소스 계획 등을 작성 및 공유하며 수요와 공급을 배분할 수 있도록 지원해야 합니다. 데이터 및 AI 플랫폼의 일부로 의사결정 최적화 기능이 있으면 처방적 분석을 예측 결과에 손쉽게 적용할 수 있습니다. 이것이 바로 의사결정 인텔리전스입니다.[6]

다양한 기술 및 인재 대상으로 기술 재교육 및 숙련도 향상을 통한 생산성 향상

고속 성장하는 기업에 존재하는 성공적인 데이터 과학 팀은 다양한 인재로 구성되어 있습니다. 한 사람이 컴퓨팅, 데이터 관리, 수학 적용, 비즈니스 사용 사례, 의사결정 과학 등 여러 AI 분야에서 전문성을 갖추기는 어렵습니다. AI를 이용하려면 다양한 기술을 가진 사람들이 팀을 구성해 여러 관점에서 연구하고 협업해야 합니다.

따라서 모델 구축에 시각적 데이터 과학 도구와 개방형 소스 도구 및 Python 또는 R과 같은 프로그래밍 언어를 모두 사용하여, 시각적이고 프로그래밍 방식의 접근법을 모두 지원하는 플랫폼이 필요합니다. 또한 이 플랫폼은 시각적 데이터 과학자가 신속하게 텍스트를 구조화된 데이터와 결합하여 숨겨진 패턴을 찾아내고 이를 예측 및 최적화에 활용할 수 있도록 지원해야 합니다.

기존에 투자한 빅데이터를 함께 활용

또한 플랫폼이 분석에 빅데이터를 제대로 활용하여 더욱 안전하고 빠른 통찰을 대규모로 확보하는 것도 중요합니다. 예를 들어 최신 데이터 및 AI 플랫폼은 Hadoop 클러스터의 데이터에 액세스하고 데이터의 대규모 이동 없이 실행이 이루어지도록 할 안전한 방법을 마련할 수 있습니다. 또한 이 플랫폼은 기존에 보유한 Hadoop 기술을 활용해 확장 및 고가용성을 위한 분산 컴퓨팅 환경의 장점을 누리면서 모델 개발과 통찰 생성을 촉진할 수 있습니다.

AI 모델의 신뢰성 및 투명성 강화

데이터 및 AI 모델에서는 신뢰도 구축이 그 무엇보다 중요합니다. 성공적인 신뢰도 구축을 위해서는 수명 주기 전반에서 AI의 결과를 추적하고 측정해야 합니다. 또한 모델이 어디에서 개발되었는지 혹은 어떤 클라우드에서 실행 중인지 여부와 관계없이 공정성, 근거 제시 및 규정 준수를 보장할 수 있어야 합니다. 기업과 규제 당국은 잠재적인 편향을 완화하고 결과를 설명할 수 있기를 요구하기 때문에 이러한 기능이 최신 데이터 및 AI 플랫폼의 일부로 제공되어야 합니다.

모든 클라우드에서 오픈 소스 및 동급 최강 도구를 갖춘 에코시스템 제공기업의 팀 조직이 전 세계 각지에 위치해 있듯 데이터도 마찬가지입니다. 전 세계에서 일어나는 혁신의 물결에 합류하려면 데이터가 어느 위치에 있든 활용할 수 있어야합니다. 데이터 및 AI 플랫폼은 개방형이어야 하며 활발하게 운영되는 에코시스템의 장점을 활용하는 동시에 다양한 클라우드에서 실행되는 모델과 데이터를 지원해야 합니다. 또한 데이터 이동이 필요할 경우 규정 위반 또는 법률상의 문제를 유발할 수 있으므로 이로 인한 비용과 위험을 플랫폼에서 완화해 줄 수 있어야 합니다.

뿐만 아니라, 업계 특화 도구들로 AI 프로젝트를 즉시 시작할 수 있는 기능을 제공하고 사전에 정의된 비즈니스 조건과 데이터 과학 아티팩트가 포함된 사전 구축앱을 제공해야 합니다.[7]

유연한 라이센싱 모델의 구축 및 구매 옵션 조합기술 관련 분야에서 구축이냐 구매냐를 결정하기가 쉽지 않을 때가 많습니다. 빠르게 성장하는 대부분의 기업은 데이터 과학 및 AI의 측면에서 구매 및 구축 시기를 선택합니다. 기업의 요구사항은 수명 주기가 진행되면서 달라질 수 있습니다. 그렇기 때문에 패키지 형태의 AI 애플리케이션을 수용할 수 있는 유연한 데이터 및 AI 플랫폼을 찾아야 할 뿐만 아니라 오픈 소스 및 독점 소유 기술이 혼합된 엔터프라이즈급 도구를 지원해야 합니다.

통합 도구와 함께 작동하는 민첩한 AI 절차 수립데이터 과학과 AI의 가치를 실현하려면 민첩한 AI를 구현해야 합니다. 수백여 가지에 달하는 성공적인 AI 구현에서 나온 모범 사례를 토대로 삼을 수 있습니다.[8]

이러한 기본적인 기술 원칙을 검토하면 AI에 관해 다른 비즈니스 리더에게 조언을 하는 데 도움이 됩니다. 고도로 성장하는 기업의 AI 활용 방법을 설명한 다음, 개인 및 조직이 협업을 통해 생산성을 대폭 향상할 수 있음을 강조합니다. 경영진이 지속적으로 AI CoE(center of excellence)를 지원할 경우 회사의 AI 투자에 대한 높은 수익을 보장하는 데 도움이 될 수 있습니다.

신뢰와 투명성이 뒷받침되는 AI 구축 및 확장 효과

귀하는 AI 성공을 향해 조직을 이끌어 나아가고 있습니다. AI를 현대화하려면 올바른 데이터 및 AI 플랫폼에 투자해야 하며 혁신적인 방법으로 데이터, 인재 그리고 신뢰와 관련된 어려움을 헤쳐 나가야 합니다. 플랫폼은 예측, 자동화, 최적화를 AI 구현의 일부로 통합해야 합니다. 신뢰와 투명성으로 AI를 구축하고 확장함으로써 다음과 같은 영역에서 AI의 능력을 발전시킬 수 있습니다.

– AI 수명 주기 관리 자동화
– 예측 결과를 바탕으로 의사결정 인텔리전스 향상
– 모델 성능 모니터링
– 통합된 멀티클라우드 데이터 및 AI 플랫폼을 기반으로 협업 데이터
– 과학 지원

AI는 많은 혜택을 가져다줍니다. 조직은 매출과 수익성, 운영 우수성 및 혁신의 측면에서 경쟁자를 뛰어넘을 수 있습니다. 신뢰도가 높은 방법으로 데이터와 모델을 처리할 수 있고 고객 및 비즈니스 파트너와 차별화된 방식으로 소통할 수 있습니다. 또한 데이터 및 AI 모델에 대한 확신을 구축할 수 있으며 이를 활용해 가치 사슬과 더불어 고객 및 파트너 경험을 향상하고 성장을 촉진할 수 있습니다.

추가 정보

IBM Cloud Pak for Data가 비즈니스를 변화시키는 방법을 알아보십시오.

자세히 알아보기

데이터와 AI를 클라우드 환경에 맞게 현대화 - Cloud Pak for Data

모든 유형의 클라우드에서 실행되는 확장 가능한 개방형 데이터 및 AI 플랫폼(Cloud Pak for Data)을 통해 비즈니스 운영 방식을 혁신하고, AI로의 전환을 가속화하는 방법에 대해 알아보세요. Cloud Pak for Data는 통합형 데이터 및 AI 플랫폼으로서 기업들의 데이터를 수집, 구성 및 분석하고 AI를 시작하는 방법을 지원합니다. 또한 IBM 뿐만 아니라 타사 마이크로서비스 어레이를 통해서도 단 몇 시간 만에 구현 가능하며, 모든 클라우드에서 실행되므로 기업에서 더 쉽게 분석하고 애플리케이션을 통합하여 혁신을 가속화할 수 있습니다.

자세히 알아보기

부록: 업종별 AI 사용 사례

빠른 성장을 거듭하는 선도 기업들은 규모에 맞는 구체적인 비즈니스 문제를 해결해 줄 고유의 AI 모델을 구축하기 위해 어떤 투자를 하고 있을까요? 업종별로 사례를 살펴보겠습니다.

은행
– 시장 및 거래 상대방 위험 평가
– 대출 신청 시 신용 위험 평가
– 실시간으로 사기 거래 감지
– 대출 승인 및 고객 서비스 소요 시간을 단축하고신속한 신용 위험 평가로 위험 통제

보험
– 사기성 있는 청구 적발
– 각 가입 신청자의 관련 위험을 평가해 견적 및 보험료 최적화
– 악천후 이벤트를 예측해 자동차 보험 청구 감소

에너지 및 공공설비
– 방대한 물리적 자산 네트워크 관리
– 생산 및 수요 패턴 예측
– 사전에 정전 발생 예측
– 수요 및 공급 계획 수립
– 수요 및 공급 최적화 프로세스 간소화

정부기관
– 복지 혜택 부정 수급 감지
– 공공 서비스 사용 패턴 예측
– 쓰레기 관리 및 교통량 최적화

제조업
– 제품 품질 모델링 및 오류 감지로 원활한 생산 라인 가동
– 창고 관리 및 물류 최적화
– ML 모델을 사용해 자율주행 차량용 센서 개발
– 공급망 운영 최적화

소매
– 고객 충성도 프로그램 관리
– 고객 프로필 및 정교한 취향 모델을 바탕으로 지정 대상별 추천 항목을 다르게 선정해 교차 판매 및 상향 판매 촉진
– 정확한 수요 예측 지원

식품
– 식료품 상태에 대한 데이터 수집 및 분석 자동화
– 상태 악화 가능성 예측 및 경고로 신속한 개입이 가능하도록 지원
– 민감한 데이터 보호로 경쟁업체들과의 안전한 협업 환경 마련

의료
– ECG 및 기타 의료 기기의 데이터 흐름 모니터링
– 환자의 상태에 변화 시기 예측
– 의료 연 구 및 조사 수행
– 실시간 환자 데이터 흐름 분석

미디어 및 엔터테인먼트
– TV 시청자에 대한 신속하고 심층적인 통찰 제공
– 더욱 다양하고 복잡한 시청자 데이터에 대한 신속한 통찰 확보
– 유연한 리소스로 수요 변화 시 빠르고 손쉬운 확장 지원
– 기술 또는 운영보다 비즈니스 전략에 집중

교육
– 학생 성취도 및 유지 예측
– 목표 달성을 위해 추가적인 지원이 필요한 학생 파악
– 기부자 관계 강화
– 잦은 결석을 줄일 수 있도록 학생 동선 추적

컴퓨터 서비스
– 회사 업무 처리 절차에 대한 즉각적인 통찰 제공
– 로봇 프로세스 자동화(RPA)로 새로운 차원의 효율성 실현
– 직원들이 전략적 활동에 집중할 수 있도록 지원
– 고객 행동 예측 및 이탈 모델을 기반으로 더 높은 수준의 지원 경험을 조성함으로써 고객 만족도 향상

용어집

알고리즘(Algorithm) 작업의 순서를 정의하는 규칙 세트이며 데이터에 적용해 특정 문제를 해결할 수 있습니다.

인공 지능(Artificial intelligence, AI) 컴퓨터 시스템이 데이터를 해석하고 학습하는 능력입니다. 이 용어는 일반적으로 머신 러닝 또는 딥 러닝 모델을 사용해 구축된 시스템을 설명하는 데 사용됩니다. 컴퓨터가 기존에는 해결하기 어려운 것으로 여겨졌던 다양한 문제를 해결하도록 지원하는 데 AI 기술을 사용할 수 있습니다.

편향(Bias) 모델 설계, 훈련, 테스트 시 부정확한 예측을 발생시킬 수 있는 보편적인 문제를 말합니다. 기업이 다양한 분야에서 AI 도입을 고려하면서, 런타임중에 모델을 모니터링하고 감사를 실시해 편향을 완화하는 방법에 대한 관심이 점점 더 커지고 있습니다.

분류 모델(Classification model) 이미 분류된 일련의 데이터 요소와 비교하여 데이터 요소를 범주로 분류하기 위해 사용됩니다. 결과로 별개의 값이 산출되는데, 이는 점수가 아니라 제한된 옵션들 중 하나를 의미합니다. 예를 들어 분류 모델은 고객이 구매를 할지 여부 또는 고객이 신용 불량의 위험이 있는지 여부에 대해 예 또는 아니요로 답을 제시할 수 있습니다.

컨텐츠 분석(Content analytics) 텍스트, 이미지, 음성, 영상 파일과 같은 다양한 형식의 문서에 포함된 비정형 데이터를 분석하는 것입니다. 머신 러닝 기술은 기존에 직원들이 수십만 시간을 들여 검토 및 분류해야 했던 대규모 컨텐츠의 저장소 분석 시간을 대폭 단축할 수 있습니다.

데이터 과학(Data science) 통계, 데이터 분석, 머신 러닝의 다양한 측면을 통합해 데이터를 처리함으로써 비즈니스 문제를 해결하는 광범위한 학문입니다. 새로운 패턴을 인식하고 예측을 구축하고 의사결정을 최적화하는 데 유용합니다.

의사결정 최적화(Decision optimization) 고급 수학 및 인공 지능 기술을 사용해 수백만 가지 의사결정 변수, 비즈니스 제약, 거래와 관련된 의사결정 문제를 해결합니다.

딥 러닝(Deep learning) 수많은 숨겨진 계층이 포함된 신경망을 사용하는 머신러닝의 한 분야입니다. 이렇게 고도로 정교한 네트워크는 컴퓨터 시각, 기계 번역, 음성 인식과 같은 딥 러닝의 첨단 분야에 활용됩니다.

배포(Deployment) 모델을 비즈니스 애플리케이션에 통합하고 실제 데이터로 모델을 실행하는 프로세스입니다. 테스트 환경, 준비 환경, 운영 환경을 거치면서 모델을 구축 및 이동하려면 데이터 과학자, 애플리케이션 개발자, IT 운영 팀 간에 협업이 필요합니다.

예측 모델링(Predictive modeling) 예측 모델링에서는 일반적인 통계 기법 또는 머신 러닝 알고리즘을 사용해 데이터 세트에 대해 모델을 훈련 및 테스트하면서 모델을 생성 및 개량해 나갑니다. 개발 프로세스는 매우 반복적인 작업입니다. 원하는 정확도 수준을 얻기 위해 수십 또는 수백 개의 모델을 훈련해야 할 수도 있습니다.

설명 가능성(Explainability) AI 모델을 투명하고 감사가 가능하도록 하여 각 의사결정의 맥락을 제공하는 역할을 합니다. 추천 항목을 만들고 비즈니스 의사결정을 지원하기 위해 예측 모델을 사용하는 모든 시스템이 갖추어야 하는 중요한 특성입니다. 복잡하고 이해하기 힘든 예측 모델은 비즈니스 이해관계자, 규제 당국 및 고객의 신뢰를 얻지 못합니다.

탐색(Exploration) 데이터 탐색은 모델 구축 프로세스에서 중요한 영역입니다. 이 활동의 목적은 지정한 데이터 세트에서 흥미로운 특징을 찾아내고 숨겨진관계를 밝히며 예측 모델이 비즈니스 가치를 제공할 수 있는 사용 사례를 가려내는 것입니다.

특정 지역 분석(Geospatial analytics) 위도와 경도, 우편 번호, 주소를 비롯한 지리적 데이터에 대한 분석입니다. 이 분석은 여러 실질적인 데이터 과학 문제를 해결하는 데 매우 유용합니다. 최신 데이터 과학 플랫폼은 특정 지역 정보를 손쉽게 감지, 구문 분석 및 계산할 수 있어야 하며 결과를 시각적으로 나타낼 수 있도록 지도와 간단히 통합할 수 있어야 합니다.

추론(Inference) 인공 지능에서 추론은 지식 기반에 논리 규칙을 적용해 불확실성이 존재할 때 결론을 도출합니다. 추론을 통해 사용자는 실시간 성능에 대한 압축된 형태의 간략하고 최적화된 예측을 얻을 수 있습니다.

선형 회귀(Linear regression) 하나의 독립 변수를 사용해 값 또는 점수를 설명하거나 예측하는 통계 프로세스입니다. 예를 들면 특정 주에 판매되는 제품의 SKU 수 또는 고객이 계정을 닫을 위험도(%) 등입니다.

로지스틱 회귀(Logistic regression) 결과 예측에 사용되는 통계 프로세스입니다. 이 프로세스는 하나의 독립 변수에 가능성이 무한하지 않고 제한된 수의 가능한 값만 포함된다는 점에서 선형 회귀와 다릅니다. 응답이 첫째, 둘째, 셋째 등 숫자 순서와 같은 범주에 해당되는 경우에 로지스틱 회귀가 적용됩니다.

머신 러닝(Machine learning, ML) 머신 러닝은 대규모 데이터 세트에서 통계적 기법을 사용해 명시적인 프로그래밍 없이 정교한 예측 모델 및 알고리즘을 도출해냅니다. 이 반복적인 프로세스에서는 일반적으로 데이터 세트를 두 개의 하위 세트로 나누어 훈련 및 테스트를 시작합니다. 수십 또는 수백 가지 변형을 가지고 훈련 세트를 대상으로 모델을 훈련하고 테스트 세트를 대상으로 성능을 테스트하면서 예측의 정확성을 평가합니다. 이 프로세스를 실행하고 각 반복 시마다 성과가 가장 좋은 사례를 찾아 다음 변형의 기초로 활용하면서 해당 모델이 점진적으로 학습하게 되고 성과가 향상됩니다.

관리(Management) 시간이 지나도 정확성을 보장하기 위해서는 모델 관리가 필수적입니다. 정기적인 모델 재훈련을 통한 새로운 데이터 학습이 매우 중요하기 때문에 모델 개발, 구현, 배포 및 관리가 연속적인 주기를 형성해야 합니다.

ModelOps 모델 운영(model operation)이라고도 하며, AI 모델을 애플리케이션 개발에 통합하고 배포해 DevOps 및 애플리케이션 개발 주기를 늘리는 기능입니다. 운영할 모델은 ML 모델, 의사결정 최적화 모델 또는 데이터 변형 모델일 수 있으며 일반적인 ML 모델에 국한되지 않고 다양한 모델로 구성할 수 있습니다.

자연어 처리(Natural language processing, NLP) 기본적으로 비정형 텍스트 데이터의 분석이 중심이 되는 AI 분야입니다. 일반적인 사용 사례로는 음성인식, 자연어 이해 및 감정 분석(sentiment analysis) 등이 있습니다.

신경망(Neural network) 신경망은 훈련 모델의 프레임워크 역할을 하며, 최적의 모델을 최적의 모델을 판별할 수 있도록 수많은 머신 러닝 알고리즘 간의 복잡한 인터랙션을 지원합니다. 인간과 동물의 뇌에서 신경 세포를 상호 연결하는 구조에서 착안해 인공 신경망의 구조가 고안되었습니다. 계층(layer)들이 인공 신경 세포를 연결합니다. 데이터가 하나 이상의 숨겨진 계층을 통해 입력 계층에서 출력 계층으로 구조를 통과합니다. 이 때 수학 함수가 데이터를 예측으로 변환하며 예측의 정확성을 평가할 수 있습니다.

오픈 소스 소프트웨어(Open source software) 오픈 소스 소프트웨어는 통계적 모델링과 머신 러닝의 다양한 영역에서 점점 더 지배적인 패러다임이 되었습니다. R, Python, Scala 등의 언어, Apache Hadoop 및 Spark와 같은 빅데이터 아키텍처, 그리고 TensorFlow 및 Spark MLlib를 비롯한 머신 러닝 프레임 워크가 예측 분석 및 데이터 과학에서 중요한 역할을 수행 중입니다.

예측 분석(Predictive analytics) 예측 분석에서는 과거의 데이터를 사용해 특정 범위(domain) 또는 문제를 모델링하거나 과거에 특정한 결과를 일으킨 핵심 요인을 고립시킵니다. 이 프로세스로 구축된 모델은 새로운 데이터를 이용해 향후 발생 가능한 결과를 예측합니다. 예측 분석에는 일반적인 통계 모델링에서 머신 러닝 알고리즘에 이르는 다양한 기술이 사용될 수 있습니다.

예측 모델(Predictive model) 입력을 매핑하는 알고리즘입니다. 즉, 데이 터베이스 레코드, 텍스트 샘플 또는 이미지 등의 데이터 조각을 출력 또는 예측에 매핑합니다. 일반적으로 출력은 숫자 또는 백분율과 같은 연속적인 변수이거나 “예” 또는 “아니요”와 같은 별개의 범주입니다. 회귀 모델과 분류 모델이 예측 모델의 두 가지 주요 유형입니다.

데이터 준비(Preparation of data) 데이터 과학 프로세스의 첫 번째 단계 중 하나입니다. 대부분의 프로젝트는 상세 분석 후에도 가치 있는 결과를 얻을 수 있을 만큼 높은 품질을 보장하기 위해 데이터 세트 정제를 가장 먼저 실행합니다. 소스 데이터는 대체로 모델링 및 분석에 좀더 유리한 형식으로 정리 및 변형이 필요합니다. 머신 러닝 모델을 구축 중일 경우 지도 학습(supervised learning)에 사용하기 위해 수동으로 데이터 레이블을 지정해야 할 수도 있습니다.

회귀 모델(Regression model) 회귀 모델은 여러 가지 변수가 포함된 데이터 세트가 있고 이들 사이의 관계를 분석하려는 경우에 유용합니다. 특히 다른 변수가 변경될 때 하나의 특정 변수가 어떻게 변경될 것인지를 밝힐 수 있습니다. 선형 회귀는 값 또는 점수를 예측하는 데 사용할 수 있습니다. 예를 들어 특정 주에 제품의 재고 유지 단위(SKU) 판매량 또는 고객이 계정을 닫을 위험도(%) 등을 예측할 수 있습니다.

통계 모델링(Statistical modeling) 데이터 세트에 관한 확률적 가정을 근거로 모델을 생성하는 것과 관련된 수학 영역입니다. 기업에서는 데이터 세트의 중요 특징을 분석하고, 데이터 분류 또는 예측 생성에 활용할 수 있는 상관관계를 찾아내는 데 통계 모델을 사용해 왔습니다.

지도 학습(Supervised learning) 데이터가 이미 정확하게 레이블 지정이 되어있을 때 데이터 세트를 사용해 머신 러닝 모델을 훈련하는 방법입니다. 모델에서는 출력 변수로 보통 범주 또는 값이 생성됩니다. 따라서 레이블이 지정된 입력과 출력을 비교해 정확성을 쉽게 평가할 수 있습니다. 선형 회귀, 무작위 예측 및 SVM(Support Vector Machine)이 지도 학습 알고리즘 중 가장 널리 알려져 있으며 대부분의 예측 모델은 이러한 기술로 구축됩니다.

테스트(Testing) 테스트는 훈련과 함께 AI 프로세스에서 데이터의 정확성을 결정하는 필수 단계입니다. 예측 모델은 지속적인 테스트를 통해 정확성을 높여야 합니다. 모델이 실패할 경우 분석가는 근본 원인을 식별하고 재훈련과 테스트를 거쳐 모델을 개선해야 합니다.

텍스트 분석(Text analytics) 텍스트 분석에서는 언어 규칙, 자연어 처리 및 머신러닝을 사용해 비정형 컨텐츠를 측정합니다. 이 프로세스는 인간의 뇌와 유사한 접근법을 사용하되 훨씬 빠른 속도로 데이터를 검토합니다. 텍스트 분석을 통해 전체 데이터의 약 90%를 차지하는 비정형 컨텐츠에서 다양한 통찰과 새로운 내용을 알아낼 수 있습니다.

훈련(Training) 예측 모델 훈련은 머신 러닝, 딥 러닝 및 유용한 데이터를 선별하는 다른 AI 프로세스의 핵심 요소입니다. 정확한 예측을 제공하도록 훈련된 모델은 실시간 데이터에 사용할 수 있습니다. 변화하는 행동 패턴에 따라 모델을 조정하기 위해 정기적인 재훈련이 반드시 필요합니다.

비지도(Unsupervised) 비지도 학습은 레이블이 지정되지 않은 데이터로 머신러닝 모델을 훈련하는 방법입니다. 데이터 내에서 관심 있는 패턴 또는 구조를 모델링하고 강조하기 위한 목적으로 자주 사용됩니다. 고객을 분류하거나 고객 간의 유사성을 식별하는 새롭고 흥미로운 방법을 찾는 것과 같이 문제를 한데 모아 연결하는 것이 비정형 학습의 일반적인 영역입니다.

시각화(Visualization) 데이터를 그래픽으로 표현하는 프로세스이며 주로 차트와 다이어그램이 사용됩니다. 사람이 데이터를 이해하기 위해서는 이를 시각적으로 나타낼 수 있어야 합니다. 이 프로세스는 비즈니스 이해관계자에게 결과를 설명할 때와 프로젝트의 초기 단계 중에 새로운 데이터 설정을 탐색할 때 모두 중요합니다.

30 Questions with IBM TECH AI 편

A to Z Guide:
디지털 혁신을 위한
차세대 기술 플랫폼 전략

IBM

“뉴노멀 시대, AI를 활용하여 일하는 방식을 바꾸어야 하지 않을까요?”

글로벌 기업과 앞선 국내 기업들이 AI를 어느 곳에 활용하고 있고
향후 어느 영역에 도입하고자 하는지에 대해 구체적인 성공사례를 통해 소개합니다.

또한 막상 AI 프로젝트를 하려고 하면 데이터에서부터 스킬 역량에 이르기까지 많은 극복해야 하는 과제가 있는데요. 성공적인 AI 도입을 위해 고려할 부분은 무엇이며, 향후 Data & AI 로드맵과 미래 방향성은 무엇인지 IBM 최고의 전문가의 답변을 통해 확인해 보세요.

포스트 코로나 시대에는 언택트(Untact) 비즈니스 확대 등 많은 변화가 예상되는데, 이에 대응하기 위해 IBM은 어떤 AI 기술을 제안하고 있나요?

김세한 상무
Big Data/AI 전문 데이터 사이언티스트

IBM은 코로나19 경제 시대를 맞아 아래의 4가지 도전과제 해결에 필요한 7가지 고객의 요구사항에 대응하고 있습니다.

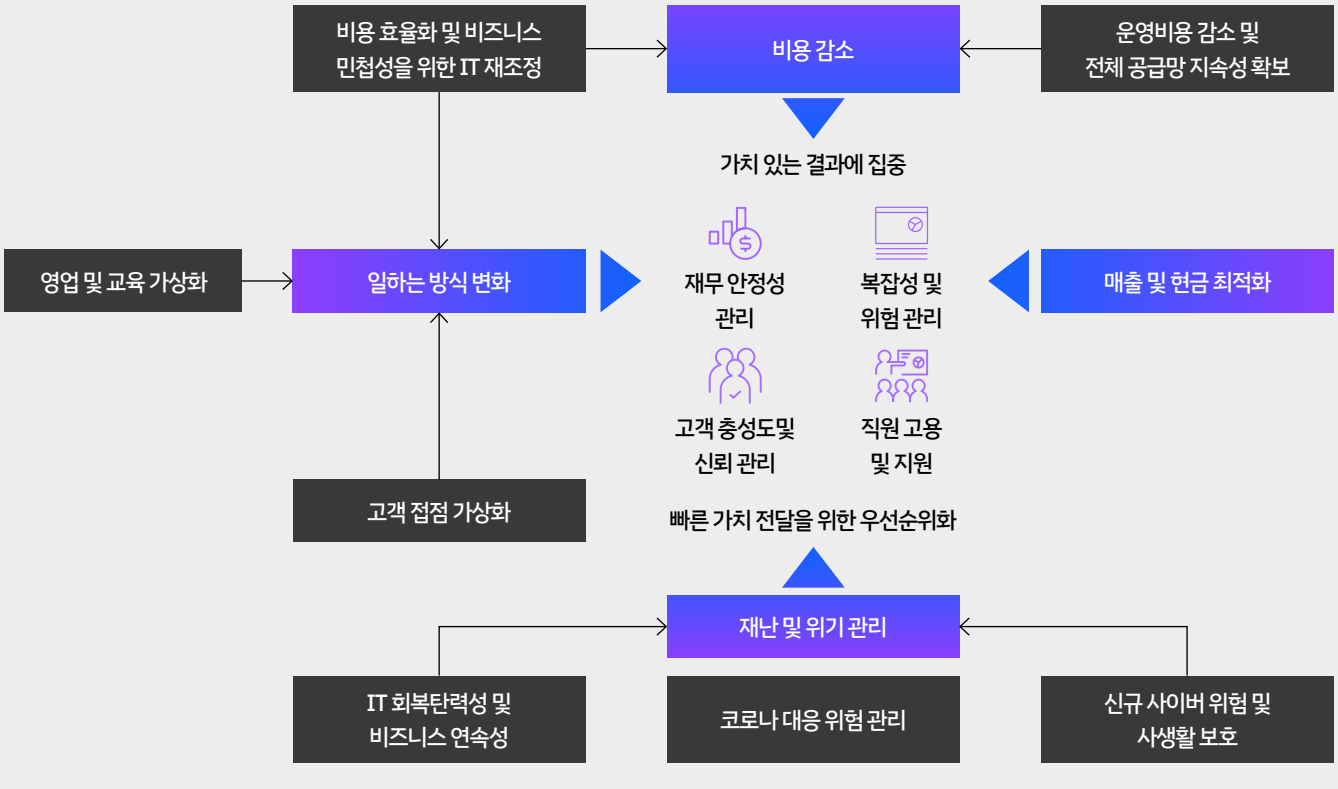


그림 1. 코로나19 경제 하에서 대응이 필요한 4대 영역 7가지 고객 요구사항

일하는 방식의 변화	• Untact 상황 하에서의 업무 생산성 유지, 원격/가상 업무담당자와의 협업 • 원격근무 시 보다 높은 생산성을 제공하는 통합 커뮤니케이션/협업 도구 제공 • 근무 방식 변화에 대한 새로운 정책 실행 및 지원 • 변동적인 환경에서의 인력 획득 및 구조 최적화
재난/위기 관리	• 주요 비즈니스/운영 위험에 대한 지속적 모니터링(임직원, 고객 및 공급자) • 민첩한 의사결정에 정보 제공을 돕는 실시간 외부/시장 데이터 분석 • 변동성 있는 환경에서의 위험관리를 관리하기 위한 재무/운영 시나리오 모델링 • 학습 및 우수 사례를 통한 비즈니스 연속성 계획 및 위험 관리 적용
비용 감소	• 진화하는 고객/시장 요구를 충족시키기 위한 공급망/운영 비용 간소화 • 비용 지출에 있어서의 우선순위화 및 최적화 • 구매 및 판매관리비 최적화, 공급자 및 협력사 관계 관리 강화 • 프로세스 효율성과 서비스 비용 감소를 위한 AI 기술 활용
매출/현금흐름 최적화	• 고객 수요 충족을 위한 운영 안정화(특히 주요 제품 공급) • 재무/유동성 수준을 확인하고 관리 • 신용, 현금 수금, 운전 자본에 대한 실시간 분석 및 엄격한 관리 • 제품 구색과 고객 기반에 대한 지속적 최적화

구체적으로 코로나19 환경에서 AI가 적용될 수 있는 솔루션을 본다면 우선 사회적 거리두기로 인해 직원들의 원격근무/재택근무가 본격화되면서 디지털 기반 내부 협업 및 고객과의 소통이 보편적인 업무 형태의 옵션으로 자리잡게 될 것입니다. 이러한 움직임은 직원의 디지털 활동에 대한 인사이트를 통해 업무 성과에 반영할 수 있도록 해주는 AI Employee Insight와 같은 새로운 형태의 인사 솔루션에 대한 필요로 이어질 것입니다. 이미 IBM은 활용 중에 있습니다.

또한 기업 경영의 불확실성으로 업무생산성 향상 및 자동화를 위한 스마트컨택센터, AI 전문가 지원 시스템, AI 기반의 운영 자동화를 위한 AIOps 솔루션 등의 필요성이 높아질 것으로 보입니다. 이외에도 코로나19 사태로 인한 경제의 변동성 확대는 변화하는 시장에 대한 소셜 데이터 모니터링 및 마켓 분석 솔루션 도입을 요구할 것입니다.

이러한 상황에서 단순 정형화 데이터뿐만 아니라 사람과 상호작용할 수 있는 AI 기술을 기업 의사결정에 적용하고 있기에, AI 비중은 점점 높아져가는 추세입니다. 요즘 운전자들이 지도책이나 자신의 감에 의존하는 대신 내비게이션을 사용하는 것과 같이 AI Enterprise 내에서 사람과 AI의 상호작용 비중이 지속적으로 증가할 것입니다.

AI 성공사례 중 기억에 남는 것과 해당 프로젝트를 통해 얻은 경험은 무엇인가요?

[항공사 사례]

이형기 상무
빅데이터 & AI 솔루션 아키텍트
(Senior Technical Staff Member)

빅데이터와 AI 기술의 초기 단계에서 진행했던 국내 한 항공사 사례가 기억에 남습니다. 빅데이터와 AI 성공의 가장 중요한 요소는 데이터의 존재와 품질 관리 상태입니다. 당시 항공기 센서 데이터 기반의 예지 정비에 큰 기대를 두고 프로젝트를 시작했는데, 목표했던 설비는 철저한 예방 정비로 인해 실제로 결함이 발생한 건수와 보관 데이터 건수가 기계 학습을 하기엔 너무 적어서 AI 모델링을 접었던 기억이 있습니다.

당시 다른 정비 빅데이터 대상을 찾던 중, 정비사들이 작성하는 설비 결함 정비 내역에 대한 기록이 정비 DB 내에 오랜 기간 잘 기록되고는 있지만 제대로 활용되지 못하고 있다는 사실을 발견했습니다. 그래서 이들 비정형 정비 기록에 대한 자연어 분석을 통해 정비 지식을 자산화하는 프로젝트를 시도했습니다. 고객사의 정비사들이 10여 년 동안 꼼꼼하게 기록한 결함 및 정비에 대한 기록들은 항공기 제조사가 제공하는 정비 매뉴얼과는 차원이 다른, 자사 직원들이 실제로 경험했던 Best Practice 정비 지식을 고스란히 담고 있었습니다. 그동안 잠자고 있던 데이터베이스 내의 비정형 데이터들이 엄청난 정비 지식 자산으로 바뀌는 감격의 순간이었습니다.

고객도 상당히 만족했던 성공 사례였는데, 데이터와 품질 외에 또 하나의 중요한 성공 요인은 바로 참여한 현업들의 열정이었습니다. 오랜 기간 수많은 정비 기술자들이 기록해 놓은 암묵지를 모든 사람이 공유할 수 있는 지식으로 자산화할 수 있다는 현업들의 확신과 열정이 없었다면, 현장의 정비사들이 너도 나도 쓰고 싶다고 이야기하게 만드는, 결함에 대한 정비지식 탐색 시간을 90% 이상 줄이게 한, 그런 정비 지식 시스템이 탄생할 수는 없었을 것입 니다. 결론적으로 데이터와 품질, 인력, 그리고 신뢰할 수 있는 기술력이 AI의 성공 요인이라고 할 수있습니다.

[항공사 사례]

김세한 상무
Big Data/AI 전문 데이터 사이언티스트

국내 금융지주사 중 한 곳과 AI 프로젝트를 수행하였습니다. 글로벌 빅데이터를 기반으로 어떤 투자 시장이 가장 유망한 지 예측하고, 이를 기반으로 자산 배분 방안을 찾고, 가장 큰 성과가 기대되는 글로벌 투자 상품을 찾아 이를 금융 소비자에게 제공하는 시스템을 개발하는 것이었습니다.

프로젝트 수행을 통해 첫째, 글로벌 15개 시장지수(주식/채권/원자재)에 대한 과거 트렌드 및 인공지능 모델이 산출한 단기, 중기, 장기 지수예측 값을 표시하여 각 시장에 대한 전망을 제공할 수 있었습니다. 둘째, 글로벌 뉴스를 분석하여 시장지수에 영향을 미칠 수 있는 주요 이벤트와 각 이벤트들의 연관관계(Relationship Network)를 제시하였습니다. 셋째, 시장 예측을 반영한 최적 분산 투자 방안을 제시하고, 지역(시장)별 및 투자유형별 추천상품을 추출, 개별펀드 현황 및 과거수익률 추이와 인공지능 알고리즘의 평가점수를 펀드판매 시 참고할 수 있도록 하였습니다.

프로젝트 수행 후 고객은 인공지능이 분석한 결과를 바탕으로 객관적인 맞춤형 투자전략과 추천상품을 고객에게 제시할 수 있게 되었으며, 현재 별도의 AI 투자자문사를 설립하여 운영 중에 있습니다.

[제과업계 사례]

김세한 상무
Big Data/AI 전문 데이터 사이언티스트

제과업계는 한 해에 무려 100여개 신제품을 내놓지만, 시장에 안착하는 비율은 3%대에 그치곤 합니다. 경쟁사 성공 제품에 미투(Me-Too) 전략으로 대응하다가 트렌드에 뒤늦은 제품을 내놓기도 합니다. 그래서 제과업계에서 연 매출 1천억원 이상인 히트 상품을 지속적으로 출시하는 방안에 대한 고민이 있었습니다. 이러한 고민에 대한 대응으로 그동안 카테고리 매니저들의 경험적인 리서치에 의존하였던 상품 개발을 내부 및 외부 빅데이터 기반으로 의사결정을 진행할 수 있는 체계를 구축하고자 하였습니다.

이러한 고민을 해결하기 위하여 고객사 내부 및 외부 데이터, 정형 및 비정형 데이터를 포함하여 연간 6.5억 건의 데이터를 분석할 수 있도록 데이터의 원천을 확장하였습니다. 또한 기존 분석 방법이나 경쟁사 대비 깊이 있는 분석이 가능하도록 시장 및 제품에 대한 속성을 정의하고 비정형 데이터에 대한 객관적 정량화를 통해 분석 관점을 정교화하였습니다. 이를 통해 신제품 기획에서 출시 리드타임을 고려한 예측 중심 데이터를 제공하는 체계를 갖출 수 있었습니다.

프로젝트 수행 후 고객은 일회성 분석이 아닌 매일 업무 수행에 인텔리전스를 접목한 Intelligent Advisor를 활용하고 있으며, 제과 및 식품 산업의 트렌드와 유행에 대한 (예: 깔라만시 소재, 혼맥/편맥 트렌드 등)신속한 감지가 가능하게 되었습니다. 이는 시장 선도 레시피를 적용한 신제품 출시로 이어져 직접적인 매출 기여로 발전했습니다.

AI 프로젝트에서 데이터가 왜 중요한가요?

이지은 상무
Cloud & Cognitive 기술영업 총괄

막상 AI 프로젝트를 하려면 가지고 있는 데이터는 많아도 실제로 필요한 데이터와 활용할 만한 데이터가 부족한 경우가 많습니다. 그럼 어떤 데이터가 필요할까요? AI 프로젝트가 데이터 제약 없이 성공하려면 기업 내부에 가용한 데이터와 함께 외부의 다양한 정형, 비정형 데이터를 탐색하고 활용할 수 있어야 합니다. 상품 개발에 AI를 적용한 사례의 경우 내부 데이터뿐 아니라 외부 기관의 시장 분석 데이터, 소셜 데이터 등으로 필요한 데이터의 범위를 확장하였습니다. 이에 따라 AI 분석에 고객성향을 표현하는 다양한 속성을 추가하여 시장의 트렌드를 민감하게 상품 개발에 반영할 수 있었습니다.

데이터를 지속해야 잘 관리하려면 어떻게 해야 할까요? AI 프로젝트 결과가 비즈니스에 지속해서 활용되려면 데이터의 범위와 깊이를 확장하는 과정이 시스템화되어야 합니다. 사일로로 존재하는 내부 데이터와 외부로부터 필요한 데이터를 제공하기 위하여 일회성 수작업이 아닌 데이터 획득 및 저장 그리고 필요한 속성을 만드는 것이 자동화 기반의 프로세스로 자리잡아야 합니다. 자동화 프로세스가 매달, 매주, 매일 등 필요한 주기로 이루어져야 AI 분석은 물론 AI 적용 및 효과 모니터링까지 원활하게 이루어집니다. 이외에 가상화 기술도 고려해야 합니다. AI를 위한 다양하고 복잡한 데이터 환경에 대응하기 위하여 물리적인 데이터 통합뿐 아니라 데이터 양, 필요한 주기 등을 고려하여 필요시 데이터 가상화 기술도 함께 도입하여 데이터에 대한 접근성을 높이고 다양한 데이터를 AI 프로젝트에 적극적으로 활용해야 합니다.

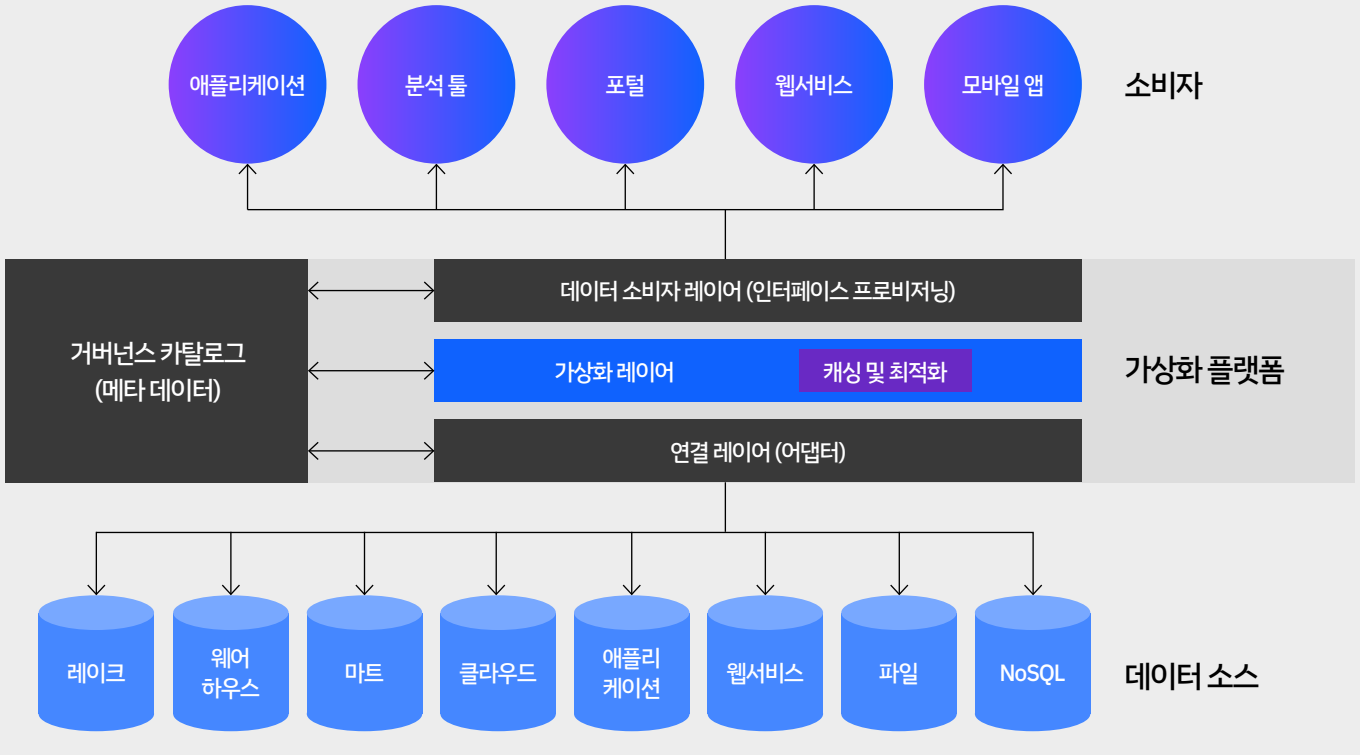


그림2. 데이터를 이동 및 복사하거나, 물리적 형식이나 위치를 알거나 이해할 필요없이 데이터를 보고, 엑세스하고, 조작 및 분석할 수 있는 가상화 기술

AI를 업무에 잘 활용하기 위해 지속적으로 신뢰할 수 있는 데이터를 제공할 방법이 있을까요?

이지은 상무
Cloud & Cognitive 기술영업 총괄

AI 프로젝트에서는 여러 AI 적용사례에 공통으로 필요한 데이터를 정의하고 일회성이 아니라 지속해서 이를 제공해야 합니다. 따라서 데이터를 사일로가 아닌 전사 관점으로 정리하여 신뢰할 수 있는 데이터를 투명한 프로세스로 제공할 수 있어야 합니다.

신뢰할 수 있는 데이터를 제공하고 유지하려면 데이터 카탈로그를 생성하고 관리하는 것이 중요합니다. 데이터 카탈로그에는 모든 데이터에 대한 위치와 DB, 테이블, 컬럼 같은 메타 데이터가 반영되어 있습니다. 또한 데이터의 비즈니스 의미를 모르면 분석 결과를 활용할 수 없으므로 비즈니스 용어를 정리하여 실제 데이터와 매핑하여 보여줄 수 있어야 합니다. 그래야 AI 결과를 신뢰하고 잘 활용할 수 있습니다.

또한, 비즈니스에 의미 없는 데이터가 원천적으로 들어오지 못하도록 특정 컬럼에 들어올 수 있는 데이터에 대한 규칙과 정책을 카탈로그에 정의해야 합니다. 더불어 계속 들어오는 데이터에도 이를 적용해 데이터 품질을 지속 관리합니다. 많은 고객이 데이터 거버넌스에서 비즈니스 용어 관리의 중요성은 알고 있으나, 비즈니스 용어 정리가 잘 안 되어 있거나, 정리되어 있다고는 하나 계속 변화하는 데이터 환경에 대응하기 위한 노력이 너무 많이 들어 지속적인 관리의 어려움을 호소하기도 합니다. 이런 고충은 AI 기술로 해결할 수 있습니다. 자동 메타 데이터 생성과 관리 기술이 점점 중요해지고 있는데, 최근에는 머신러닝, 딥러닝 기술을 접목하여 비즈니스 용어를 자동 추천하고, 채택 여부가 다시 학습되어, 추천의 정확도를 계속 높여 카탈로그에 지속해서 반영할 수 있습니다.

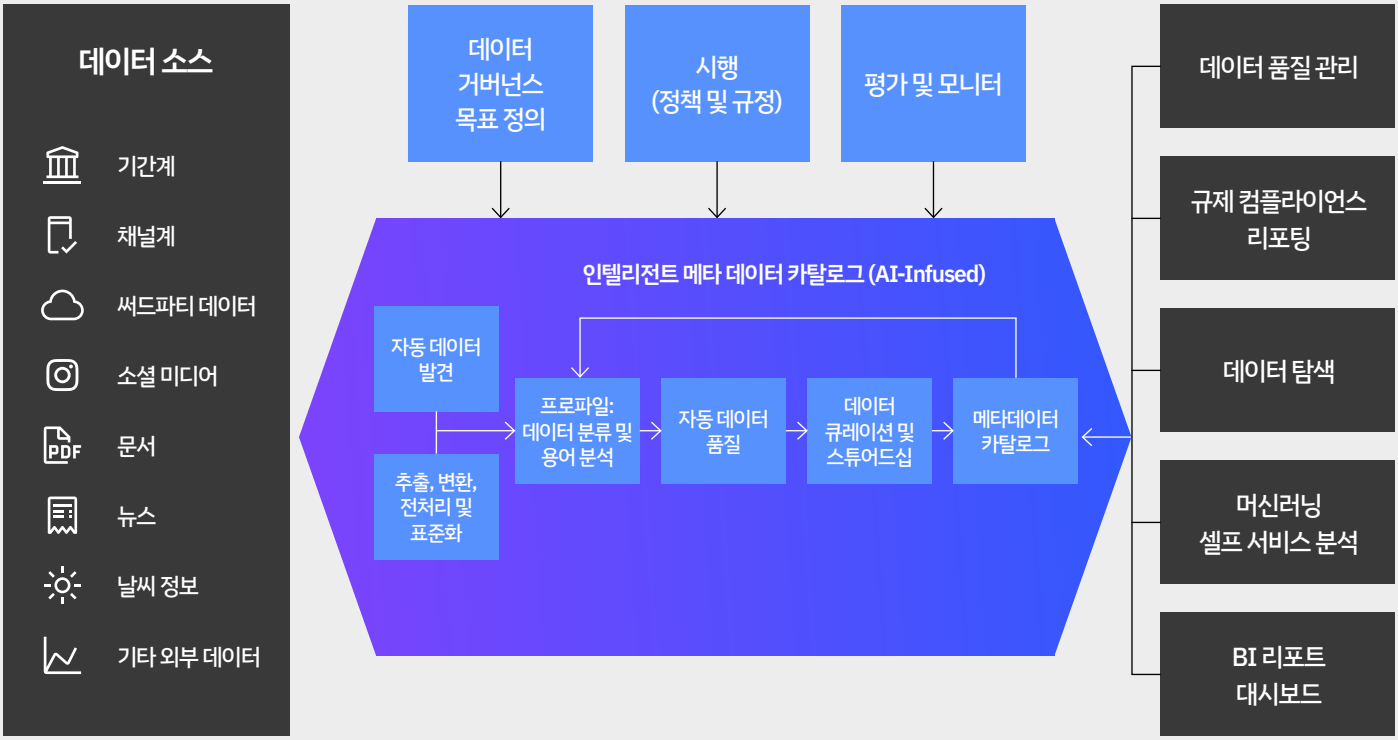


그림3. 데이터 카탈로그의 데이터 소스에 대한 자동 디스커버리 기능을 통해 데이터 검색과 거버넌스를 위한 카탈로그를 지원

기업이 필요로하는 데이터 및 AI 역량을 갖춘 전문가에는 어떤 유형이 있을까요?

김혜경 실장
Data & AI 이니셔티브 리더

먼저 데이터와 관련하여 AI 모델을 만드는 데이터 사이언티스트, 데이터 엔지니어, 데이터 분석가가 있습니다. 기업의 AI 로드맵과 적합한 기술을 선정하는 AI 전략 수립 및 기획자와 AI 플랫폼 전문가도 있어야 하고, 활용을 위한 AI 비즈니스 도메인 전문가를 포함한 다양한 영역의 전문가도 필요합니다. 또한 AI 기술이 오픈 소스 중심으로 다양한 기술들이 나오고 있음으로 오픈 소스 관련 역량도 확보해야 합니다. 그 중에서도 딥러닝 및 머신러닝 모델을 이해하고 구축할 수 있는 기술은 기업들이 반드시 갖추고 싶어하는 필수 역량입니다.

가트너에 따르면, 2025년에는 대부분의 일상에 AI가 접목되어 일반 사용자들도 AI 기술을 모바일 기술처럼 쉽게 활용하게 된다고 합니다. 대학 강의를 해보면 IT 지식이 별로 없는 학생들이 AI를 이해하고, AI 기술을 활용한 결과물을 만들어내는 것을 심심찮게 볼 수 있습니다. 이를 볼 때마다 AI 활용 역량은 기업에서 반드시 키워야 할 중요한 역량임을 체감합니다. 아직 완전하지 않지만, 데이터로부터 AI 모델을 스스로 만들어내는 Auto AI 기술도 등장하고 있습니다. 현재는 AI 모델을 만드는 데이터 사이언티스트가 제일 중요한 스킬 중 하나지만, 머지않아 이 영역은 AI 기술이 흡수할 수도 있습니다. 데이터와 AI 영역은 기술 진화 속도가 무척 빨라서 기업이 필요로 하는 필요 역량들이 함께 변할 수 있음을 염두에 두어야 합니다.

그림4. 기업의 데이터 & AI 스킬 전문가

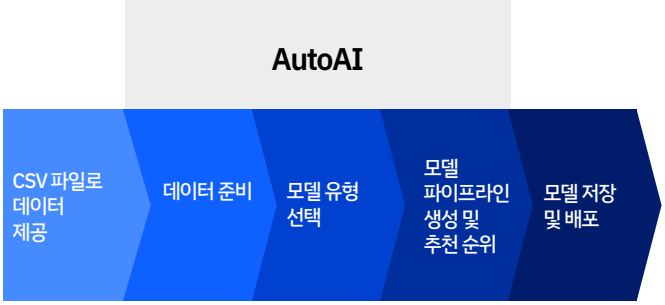


그림5. Auto AI란 데이터가 주어지면 자동으로 ML 알고리즘을 적용하여 AI 모델을 만들고 결과를 보여주는 기술

AI 프로젝트에서 데이터가 왜 중요한가요?

김혜경 실장
Data & AI 이니셔티브 리더

인공지능 전쟁은 곧 인재 전쟁이라고 합니다. 기업들은 국내뿐 아니라 해외로부터 우수한 AI 관련 인재를 채용하거나, 기업 내부에 AI 스터디 그룹이나 TFT 조직을 통해서 AI 인재를 양성하는 등 다양한 방식으로 AI 프로젝트를 위한 인재확보 노력을 기울이고 있습니다. 하지만 AI는 기술 영역이 다양하고, 기술도 빠르게 변하다 보니 AI 프로젝트를 기업 내부 인력으로만 진행하기는 사실 쉽지 않습니다.

내부 역량을 강화하고 싶다면 우선 AI 원천 기술을 보유하고 다양한 산업군에서 경험이 많은 업체와 프로젝트를 진행하는 것이 좋습니다. 이 과정에서 자연스럽게 기술 이전이 일어나서 내재화가 진행됩니다. 이것이 가장 쉽고 보편적인 방법입니다.

일례로 국내 한 유통그룹은 2017년부터 IBM AI 기술을 기반으로 대규모 AI 프로젝트를 진행하면서 디지털 데이터 자산화와 함께 AI 기술 인력 양성을 중요 목표로 설정하였습니다. 실제 해당 그룹은 지금까지도 전사 AI 전문가 조직을 중심으로 계열사별 AI 전문 인력을 지속해서 양성하고 있습니다. 국내 보험사의 경우 AI 콜봇 프로젝트에서 고장차 관련 견인 서비스를 IBM과 함께 진행하였는데, 추가적인 AI 서비스 개발은 자체 인력으로 진행하고 있습니다. AI 기술은 완성형이 아닌 현재형으로 진화하고 발전합니다. 따라서 기업 내부 인력의 AI 역량도 지속해서 함께 업그레이드해야 할 것입니다.

AI는 프로그램으로 해결하기 어려운 문제들을 해결하는 등 많은 도움이 되고 있는데요. 그 결정의 근거나 과정은 마치 블랙박스 같아서 막상 중요한 비즈니스 의사결정에 적용을 주저하게 되는 경우가 있습니다. 좀더 투명하고 신뢰할 수 있는 AI는 없나요?

이형기 상무
빅데이터 & AI 솔루션 아키텍트
(Senior Technical Staff Member)

모든 개인은 금융 기관 대출에 있어 평등한 기회가 주어져야 한다는 미국의 ECOA(Equal Credit Opportunity Act) 규제, 개인 정보가 자동화된 의사결정에 어떻게 사용되는지를 설명할 수 있어야 한다는 유럽의 GDPR 같은 규제들이 있습니다. 이러한 규제를 떠올리면 은행의 대출 심사나 CRM을 위한 개인 프로파일링 업무에 단순히 테스트 데이터 세트와 검증 데이터 세트를 가지고 측정한 모델의 성능과 정확도 수치만 보고 그 모델을 실무에 적용할 수 있을까요? 아마도 이미 그런 업무에 AI 모델을 적용한 경우라면 지금 바로 내려야 한다는 생각이 들 겁니다.

그 이유를 좀 생각해보면 AI 모델의 대부분은 히든 계층이 많아 사람이 이해하기 어려운 딥러닝 기계 학습 기법을 사용합니다. 따라서 이들이 낸 결과에 대한 근거를 쉽게 설명하기 어렵습니다. 그리고 이러한 기계 학습 방식의 AI 모델은 학습 데이터를 입력으로 해서 자동 작성되는 프로그램이라고 말할 수 있을 정도로 학습 데이터에 따라 그 결과가 달라집니다. 결과적으로 AI 모델의 학습에 사용한 데이터가 실제 상황을 얼마나 잘 대변하느냐는 AI 모델의 가장 중요한 요소입니다. 학습 및 테스트 데이터 세트를 통해 검증한 모델의 정확도와 커버리지와 같은 모델 성능은 높게 나왔을지라도 실제 데이터에서는 특정 필드에 있어 편중된 학습이 되거나 무관한 필드가 결과에 영향을 미치는 현상이 생길 수 있습니다. 따라서 기계 학습 이후에 실제 데이터가 AI 모델에 주입되는 실행을 지속해서 모니터링하면서, 어떤 입력값에 대해 어떤 결과를 내는 지 파악함으로써 그 모델의 동작을 설명하는 기술이 발전하고 있습니다.

바로 이러한 AI 모델의 동작을 설명하고 해석해주는 기술을 eXplainable AI, 한국어로는 설명 가능한 AI라고 합니다. 가령 왜 내가 신용 등급이 1등급이 아니라 3등급이 나왔는지, 신용 등급 판정에 영향을 주는 각각의 인자 영향도가 어떻게 되는지, 특정 인자값에 편파성이 얼마나 되는지를 설명할 수 있게 하는 것이 설명 가능한 AI입니다. 이렇게 설명을 통해 AI 모델의 신뢰성과 투명성을 제공합니다. 현재 이러한 기술은 오픈 소스는 물론 제품으로 제공되고 있으므로 지금 바로 사용이 가능합니다.

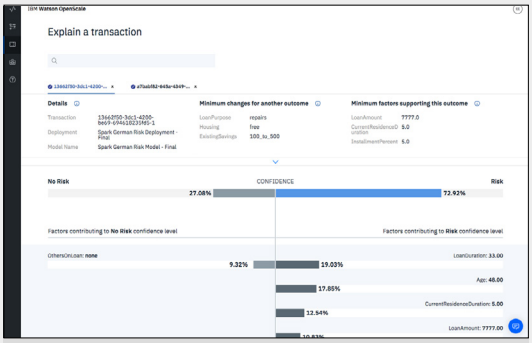
설명 가능한 AI 기술을 현재 나의 업무에 적용할 수 있을까요?

이형기 상무
빅데이터 & AI 솔루션 아키텍트
(Senior Technical Staff Member)

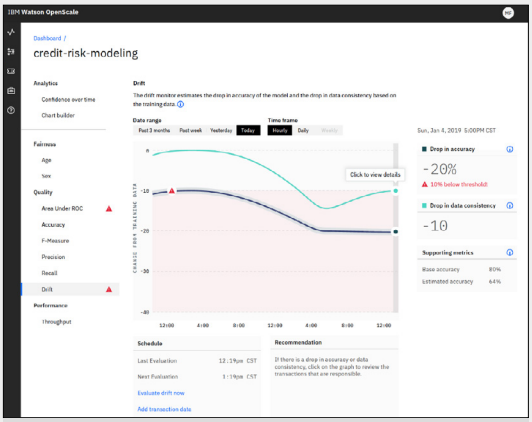
수년 전부터 이러한 블랙박스 머신러닝 모델의 해석을 위한 설명 가능한 AI 관련 연구 논문이 나오기 시작했습니다. 그리고 대표적인 오픈 소스인 LIME(Local Interpretable, Model-agnostic Explanations)을 포함 Skater, ELi5, SHAP 등의 해석 목적에 따라 선택해서 사용할 수 있는 다양한 프레임워크가 Python 또는 TensorFlow 등에서 사용할 수 있도록 이미 개발되어 있고 계속 발전하고 있습니다. 이들은 각 프레임워크에 따라 조금씩 차이는 있지만, 기존의 AI 모델에 약간의 학습 데이터/실행 데이터용 마트, 메타 데이터 설정 등을 추가로 해주면 사용이 가능합니다. 다만 실제 기업의 운영 환경에서는 하나의 모델에 대해서도 다양한 측면의 모델 해석 시도가 필요하기 때문에, 다수의 AI 모델들을 실제로 운영 중인 기업 입장에서는 각 AI 모델들의 라이프사이클 관리 자체도 버겁기에, 모델 해석을 위해 일일이 AI 모듈 소스를 수정하기는 쉽지 않습니다.

이러한 기업 사용자를 위한 상용 제품도 있어서 기존 AI 모델의 소스 코드 수정 없이도 외부 관리자 콘솔을 통한 간단한 설정과 학습 데이터 제공만으로도 보다 쉽게 검증된 explainable AI 기술 적용이 가능합니다. 이는 사용 중인 AI 모델들이 편파성 없고, 신뢰성이 확보되었음을 지속해서 모니터링하고 개선 발전시킬 수 있게 해주는 기업의 AI 모델의 프로덕션 시스템의 운영에서 꼭 필요한 환경을 제공합니다. AI 모델 실행 트랜잭션에서 사용된 입력 변수값(payload) 및 결과 기여도 모니터링, 특정 변수에 치우치는 결과들에 대한 바이어스 통계 모니터링 및 경고, 여러 기간 단위의 AI 모델 품질 지수 통계 모니터링 등에 영향을 미칩니다.

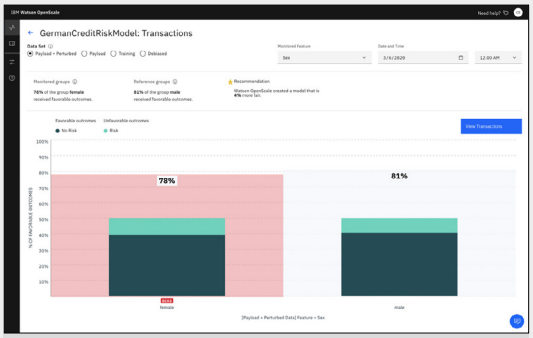
지금 바로 투명하고, 신뢰성 있고, 설명 가능한 AI 운영 환경을 적용해 보십시오. 비즈니스 프로세스 곳곳에 AI-Infused가 되는 코그너티브 엔터프라이즈가 바로 눈앞에 그려질 것입니다.



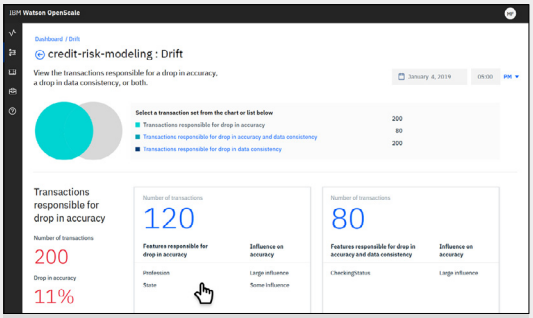
AI 판정 결과에 대한 패러미터 기여도 분석



모델의 정확도 Drift 모니터링



모델의 공정성(편파성) 모니터링



정확도 Drift 요인이 된 트랜잭션 및 패러미터 분석

그림6. 설명 가능한 AI(eXplainable AI) 기술 적용사례

참고자료
- Explain and govern AI. https://www.ibm.com/watson/explainable-ai
- IBM Watson OpenScale https://www.ibm.com/kr-ko/cloud/watson-openscale

데이터, 역량, 신뢰 등 AI 도전과제를 극복하기 위해 준비해야 할 기술기반은 무엇일까요?

이지은 상무
Cloud & Cognitive 기술영업 총괄

클라우드 기반의 현대화된 데이터, AI 플랫폼 기술이 필요합니다. 사용자 관리, 보안, 모니터링, 필요한 분석 환경에 대한 프로비저닝과 같은 공통 서비스와 함께 데이터 액세스부터 데이터 거버넌스, AI 분석과 활용까지 AI 라이프사이클을 공통 플랫폼 기반으로 유기적으로 지원해야 합니다.

먼저 데이터 관련 도전 과제 해결책을 알아보겠습니다. 플랫폼 상에 가용한 모든 데이터에 대해 유형과 위치, 저장 방식에 상관없이 접근할 수 있도록 하는 장치가 필요합니다. 데이터 전처리 과정을 자동화하고, 가용 데이터 검색이 가능하도록 데이터 카탈로그를 관리하면 다양한 AI 적용사례(Use Case)가 필요로 하는 데이터 액세스에 대한 노력과 비용이 감소합니다.

다음으로 역량에 대한 도전을 살펴보겠습니다. 다수의 사용자를 위한 협업 플랫폼으로 AI를 개개인의 역량에 의존하기보다 AI 경험을 자산으로 관리하여 기업 내 기술 역량 내재화로 이어지게 하면 안정적인 AI 기술 공급이 가능하게 됩니다.

마지막으로 신뢰에 대한 도전과제를 알아보겠습니다. AI 결과를 업무에 적용하고 플랫폼상에서 AI 결과의 정확도에 대한 주기적인 모니터링과 피드백을 하게 되면 신뢰할 수 있는 AI를 지속해서 실제 업무에 적용할 수 있습니다.

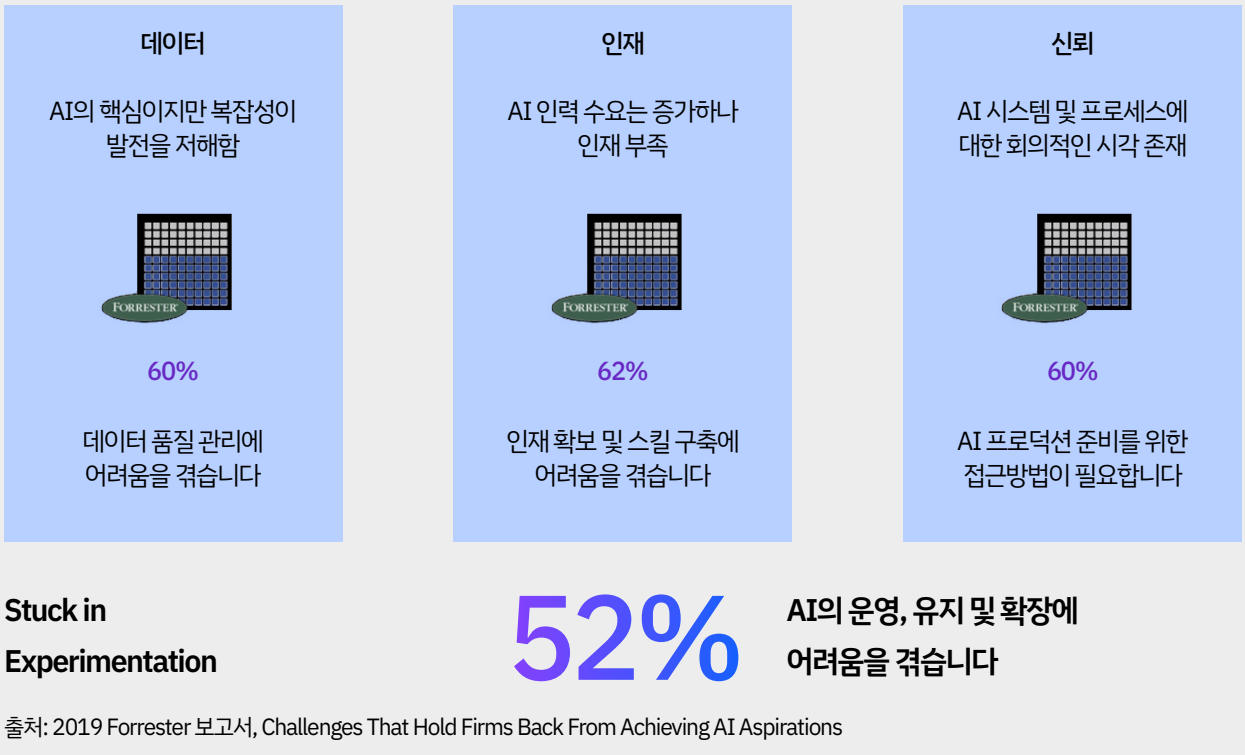
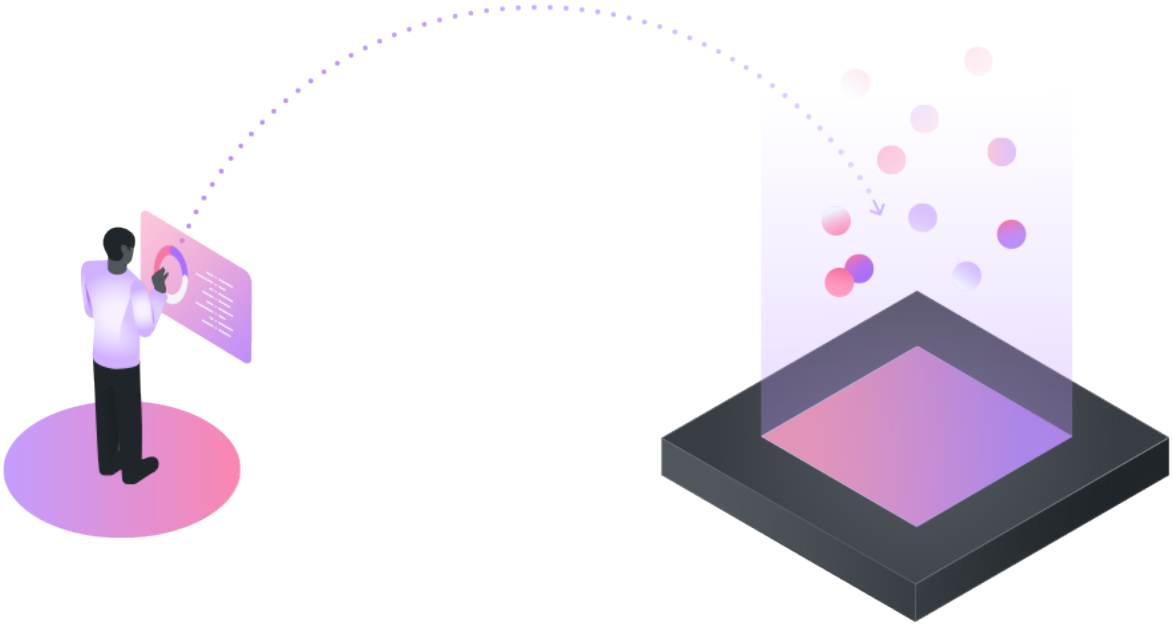


그림7. AI 도전 과제



최근 AI 기술은 오픈 소스 중심으로 다양한 기술들이 제공되므로 플랫폼에서 각종 오픈 소스를 지원해야 합니다. 다만 오픈 소스를 개인 용도가 아닌 기업 업무에 사용해야 하므로, 안정적으로 안전하게 기업에서 활용하기 위한 사전 테스트와 기술 지원 체계가 갖춰진 플랫폼이 필요합니다. 또한, 특정 AI 기술을 사용하기 위하여 특정 클라우드 플랫폼만을 사용해야 한다면 구축 후 다른 환경으로의 이동이 쉽지 않아 큰 제약이 되기 때문에 프라이빗, 온프레미스, 퍼블릭을 포함한 다양한 멀티 클라우드 환경에서 운영할 수 있는 쿠버네티스 기반의 컨테이너 환경과 같은 개방형 클라우드 기술을 고려해야 합니다.

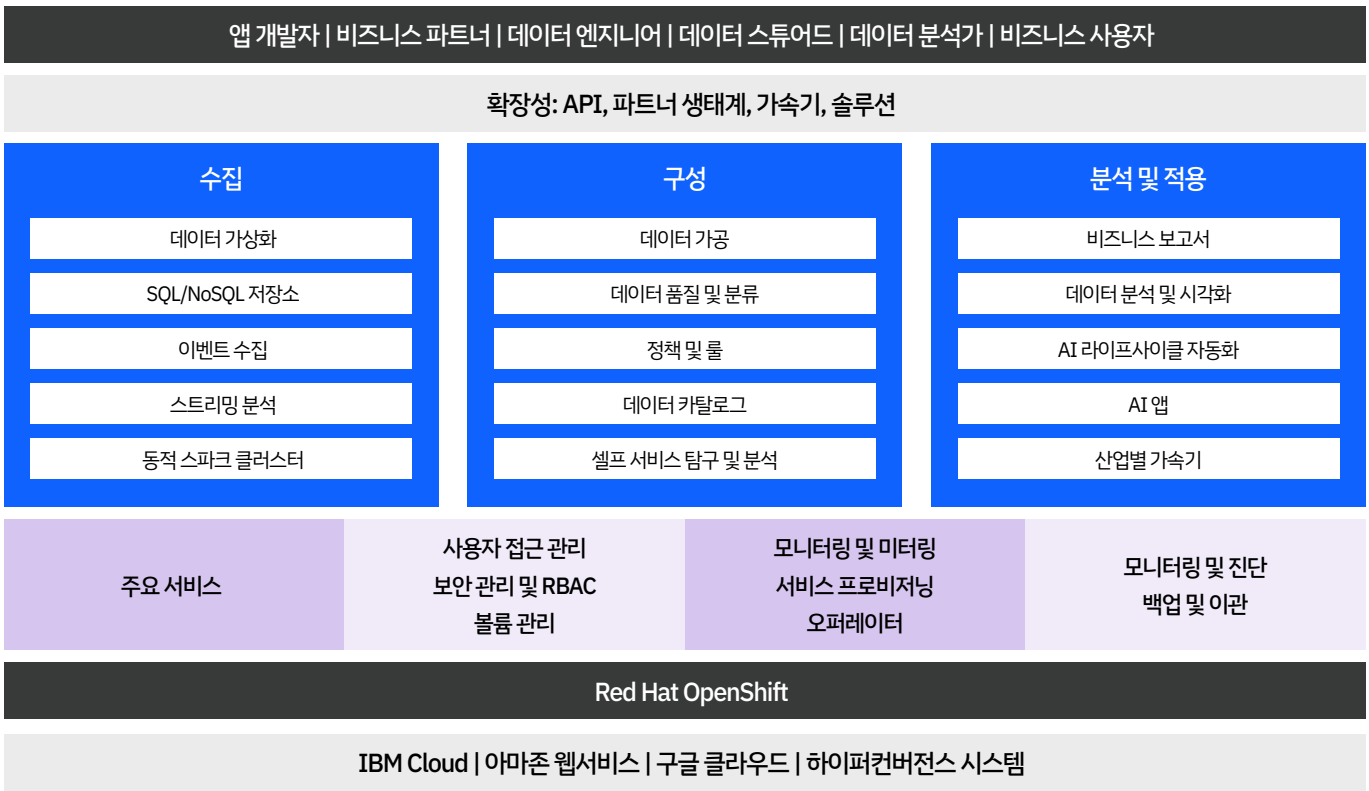


그림8. 클라우드 기반의 현대화된 데이터 및 AI 플랫폼

기업의 디지털 혁신을 지원하기 위한 IBM의 AI 기술 진화 방향은 무엇일까요?

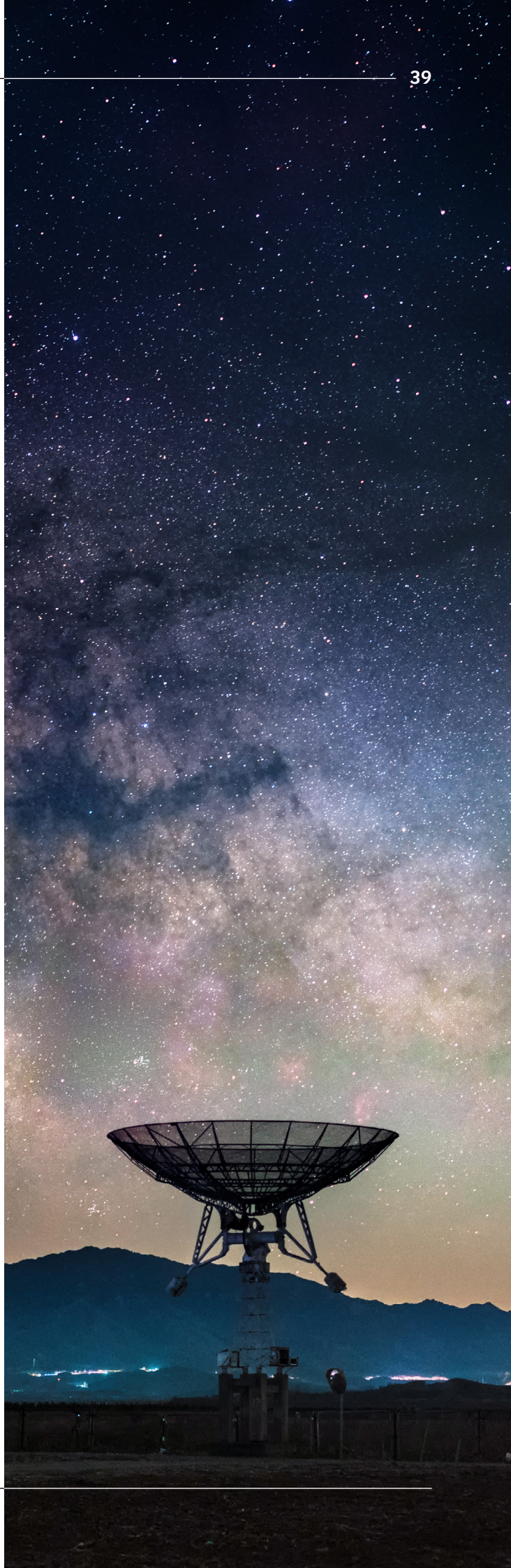
김혜경 실장
Data & AI 이니셔티브 리더

2025년에는 인공 지능 관련 산업이 2000조 원에 이르는 시장을 창출할 것으로 IBM은 예측하고 있습니다. 양자 컴퓨팅 기술이 양자 우위를 달성하게 되면, 우리가 불가능하다고 여겼던 영역에 AI 알고리즘이 접목될 것입니다. 예를 들면 주가 예측이나 전기 자동차의 배터리 기술 개발 등 기업들의 최적화 및 자동화 영역에서 상당한 진전이 이루어질 것입니다. IBM은 AI가 기업의 혁신을 주도하고, 주요 업무에 AI를 접목하여 비즈니스 효과를 내는데 초점을 맞추어 3가지 전략과 비전을 제시하고 있습니다.

첫 번째는 AI 기술을 어디서든 활용할 수 있도록 AI Everywhere를 지향합니다. 클라우드 환경뿐만 아니라 네트워크와 단절된 엣지 컴퓨팅 환경에서도 AI 엔진이 가동될 수 있도록 하는 것을 목표로 합니다.

두 번째는 IBM Watson을 기반으로 지속적으로 AI의 전체 라이프사이클, 즉 AI 모델을 개발하고 개발된 AI 서비스를 실행하고 AI 서비스가 제대로 작동하고 있는지 지속해서 모니터링까지 하는 오픈 소스 기반 AI 플랫폼을 지원하는 것입니다.

세 번째로 IBM은 AI 기술의 신뢰성 강화를 위해 Trust AI에 힘을 쏟을 것입니다. 착한 AI, 윤리적인 AI를 기업들이 활용할 수 있도록 신뢰할 만한 AI 제공자로 세상을 바꾸기 위해 노력할 것입니다.



2. IBM 솔루션

IBM AI Ladder 소개

현대화

- Cloud Pak for Data
- IBM Db2 on Cloud

수집 & 구성

- IBM Db2 Warehouse
- IBM Cloud Pak for Data
  Netezza Performance Server

분석

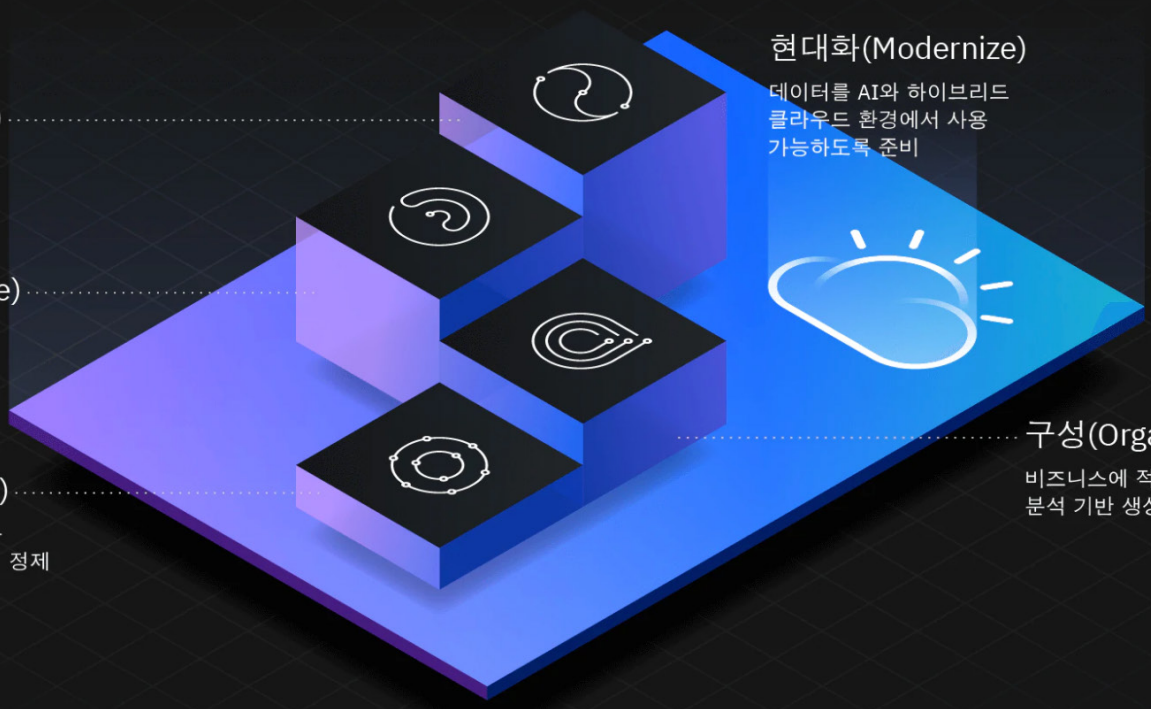
- IBM DataOps

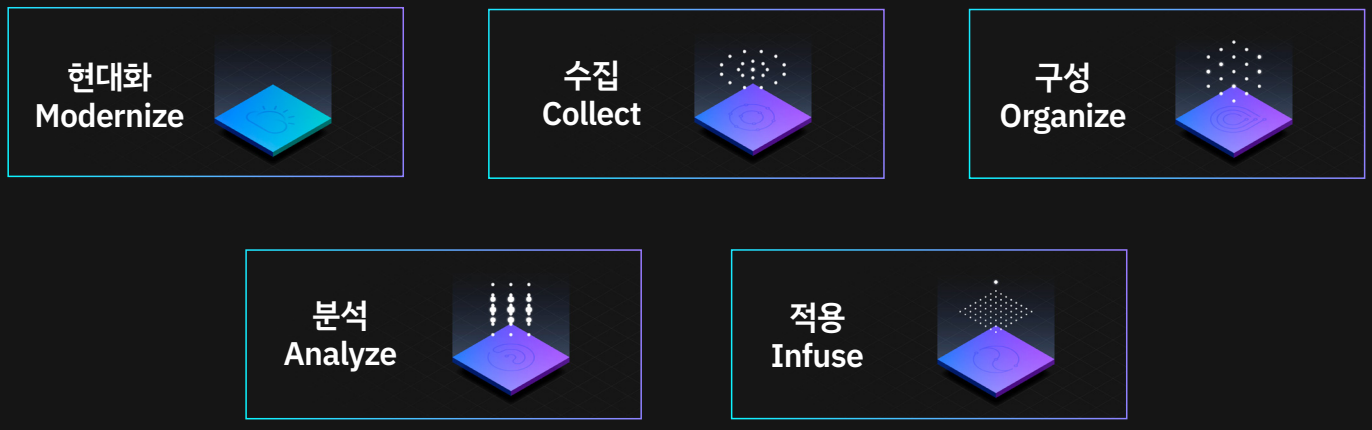
적용

- IBM Watson

AI 사다리 방법론을 통해, Data and AI로의 여정을 가속화하세요!

자세히 알아보기





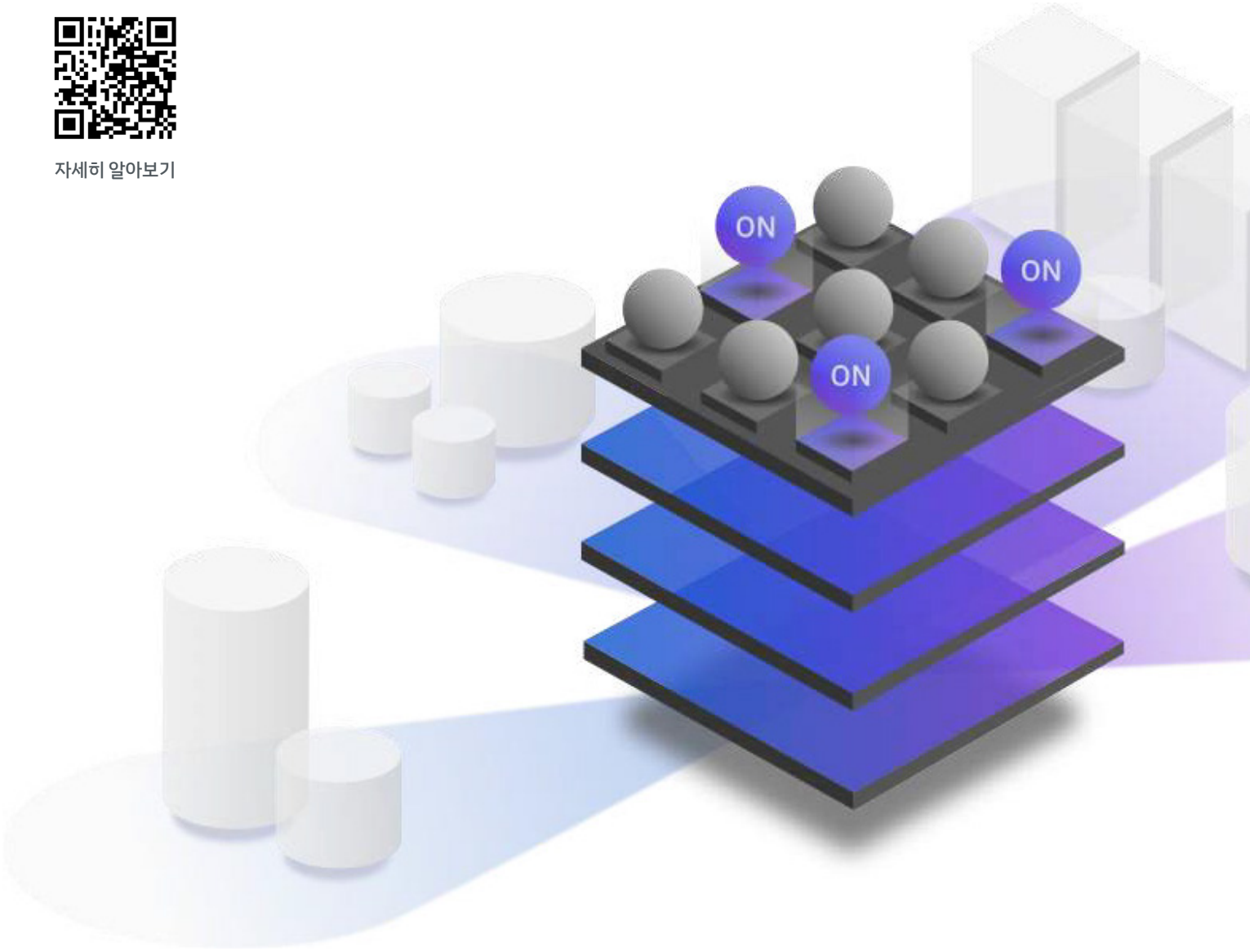
IBM의 AI 사다리는 기업의 데이터 수집에서 AI 모델 개발 및 운영까지의 end to end 전체 라이프 사이클을 지원하는 통합 플랫폼을 구현합니다. 나아가 컨테이너 기반의 최신 아키텍처 반영 플랫폼으로 온프레미스/프라이빗 클라우드 / “Any” 퍼블릭 클라우드에 동일한 분석환경 적용이 가능한 **‘플랫폼 에니웨어’**를 지원합니다.

현대화

Cloud Pak for Data

혁신의 가속화를 위한 데이터 및 AI 플랫폼

자세히 알아보기



※ 자세한 정보 : ibm.com/kr-ko/ products/cloud-pak-for-data

혁신을 위해 준비된 데이터와 AI

오늘날의 데이터 및 AI 환경에서 우리는 컴퓨팅 용량, 주요 툴의 성숙도, 데이터 소스 가용성, 자동화된 AI 기능과 관련된 문제를 해결해왔습니다. 단순한 제한 요인이었던 것들이 이제는 불필요하게 복잡성을 가중시키고 빠른 속도와 애자일 방식을 활용한 현대화의 성과를 달성하지 못하도록 방해하고 있습니다. 오늘날의 비즈니스에서는 개별 툴이 발전하는 방식 대신 다양한 기능이 함께 발전하는 방식을 강조해야 합니다.

데이터 및 AI 분야는 전체적인 플랫폼 접근법을 통해 혁신을 위한 준비를 마쳤습니다. 비용 절감, 툴 통합, 프로세스 효율화의 기회는 그야말로 엄청납니다.

상호 운용성은 현재 및 미래의 투자를 극대화하여 중복되고 비효율적인 지출을 방지합니다.

기능 제공 속도를 향상하여 현재 무언가가 필요할 때 지체하지 않도록 팀을 지원합니다.

경험을 통합함으로써 작업의 장애물, 작업 이관, 작업의 중복 수행을 방지하고 프로세스 효율성과 혁신의 속도를 중시합니다.

비즈니스의 성과와 우선순위는 변하지 않았습니다. 이러한 성과를 더 빨리 더 정확하게 실현하는 일이 어렵기 때문에 혁신이 필요합니다. 커넥티드 데이터 개발 라이프사이클 접근법인 AI 사다리(AI Ladder) 전략을 통해 IBM은 조직이 데이터로부터 목표한 가치를 더 빨리 실현하는 방법을 새롭게 만들었습니다. Cloud Pak for Data는 모든 데이터 관련 과제를 해결하는 데 필요한 모든 기능을 제공하는 데이터 및 AI 플랫폼입니다.

AI 사다리

전체론적 데이터 전략

AI 여정에 필요한 다섯 가지 중요 단계.

현대화(Modernize)
AI 및 하이브리드 멀티클라우드 환경에 맞게 데이터를 준비

수집(Collect)
데이터를 단순화하고 접근성을 높이기

구성(Organize)
비즈니스에 바로 활용할 수 있는 분석 기반을 마련

분석(Analyze)
신뢰와 투명성을 바탕으로 AI를 구축하고 실행

적용(Infuse)
비즈니스 전반에서 AI를 활용

2021년에는 AI 기술의 확대로

2조9천억 달러의

비즈니스 가치를 창출할 것입니다*

**Gartner, INC.*

데이터 라이프사이클 각 단계의 요건

수집

비즈니스 및 분석 사용자가 거버넌스 기반 셀프 서비스 방식으로 모든 데이터를 사용하면서 더 빨리 인사이트를 생성하고 비용 및 위험을 줄이게 합니다.

- 모든 컴퓨팅 능력 활용
- 엔터프라이즈 전반에 대한 가시성 확보
- 필요한 데이터의 100% 활용
- IT 및 데이터 거버넌스 연계

구성

DataOps 방식으로 엔터프라이즈 환경에서 더 신속하게 가치를 창출합니다. 이렇게 품질 높은 자산을 기반으로 한 분석 및 AI를 통해 정확도를 높이고, 배포 시간을 단축하며, 전사적으로 위험을 줄일 수 있습니다.

- 데이터 소스까지 역추적
- 애자일 데이터 거버넌스 실현
- 정책 제어를 통해 데이터 보호
- 어떤 데이터가 사용 가능한지 파악

분석

모든 스킬 레벨의 사용자가 AI 구축 및 배포에 참여하게 해주는 통합 툴세트를 통해 AI 모델의 배포 속도를 개선하고 AI 모델의 정확성 및 신뢰성을 강화합니다.

- 강력한 협업
- AI 앱에 직접 반영되는 라이프사이클
- 신뢰할 수 있는 방식으로 공급된 모델

적용

진정으로 AI를 활용하세요. 인간과 기계가 협력을 통해 훨씬 더 복잡한 과제를 적정 규모 및 빠른 속도로 완료할 수 있습니다.

- 비즈니스 프로세스 자동화
- AI 지원 인텔리전스 활용
- 계획 수립, 예산 책정 및 예측 분석 자동화

AI 여정에서 귀사가 어느 단계에 있는지 확인해 보십시오.

https://www.ibm.com/analytics/journey-to-ai/assessment/

비용 효율성, 생산성

2020 Forrester Total Economic Impact 연구의 통계 발표

ibm.biz/cloudpakfordata-tei

65-85% 감소

컨테이너 및 컨테이너 관리 효율성

애플리케이션 관리 및 효율성 향상을 통해 인프라 및 관련 관리 작업의 65~85%가 줄어들 것입니다.

26-65% 감소

데이터 가상화 및 거버넌스

거버넌스를 통한 자동화 및 액세스 속도 향상으로 생산성이 향상되었고 ETL 요청이 25%~65% 줄어듭니다.

약 300만 달러 감소

데이터 사이언스, 머신러닝, 인공지능이 제공하는 이점

데이터 사이언티스트 생산성 50% 향상 (120만 ~ 340만 달러의 절감액 기록)

730만 달러

예상 편익

(TEI 연구)

430%

페더레이션보다 빠른 속도

(IBM 테스트)

65%

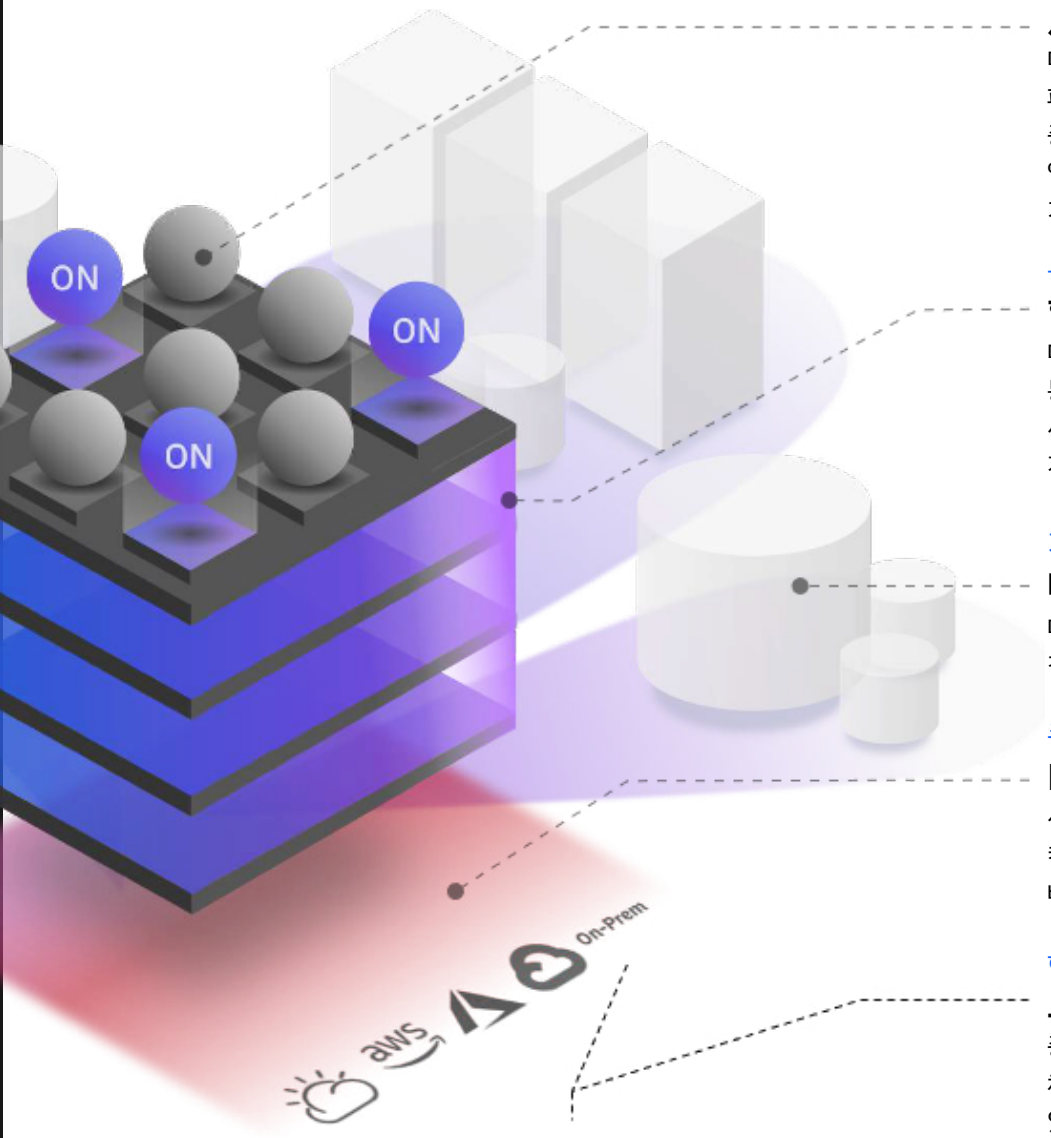
ETL 작업 감축 비율

(TEI 연구)

Cloud Pak for Data

부분의 합보다 큰 시너지 효과 제공

Cloud Pak for Data는 AI 사다리를 자동화하며, 현대 엔터프라이즈의 데이터 전략을 실현합니다. 즉, 일종의 연결 조직의 역할을 하면서 이 과정에 포함되는 모든 툴 및 페르소나를 위해 End-to-End 데이터 라이프사이클을 제공합니다. 50가지가 넘는 IBM 서비스를 통합하고, 훨씬 더 큰 규모의 파트너 에코시스템에 연결하며, 모든 유형의 클라우드 환경 모델을 활용하여 원활한 사용자 경험을 제공함으로써 **부분의 합보다 큰 시너지 효과를 냅니다.**



필요할 때 이용 가능한 기능

서비스 에코시스템

데이터 및 AI 서비스, IBM 산업 액셀러레이터, 파트너 오퍼링으로 구성되어 있고 계속 성장 중인 에코시스템을 클릭 한 번으로 검색 및 액세스하여 Cloud Pak for Data의 포괄적인 기능을 통해 확장할 수 있습니다.

통합된 사용자 경험

협력적 데이터 라이프사이클

데이터 관리, 데이터 거버넌스 및 분석을 통합하는 단일 사용자 경험으로 가치 실현 시간을 단축하고 효율성과 리소스 사용 방식을 개선할 수 있습니다.

기존 인프라 활용

데이터 가상화

데이터를 이동하지 않고 여러 데이터 소스에 걸쳐 빠르고 쉽게 쿼리를 수행할 수 있습니다.

쿠버네티스 실행

Red Hat OpenShift

선도적인 하이브리드 클라우드 엔터프라이즈 컨테이너 플랫폼을 활용하여 혁신적이고 빠른 배포 전략을 구사할 수 있습니다.

하이브리드 멀티클라우드 환경

모든 클라우드 지원

종속 현상을 방지하고 멀티클라우드 접근법을 채택하여 모든 클라우드 인프라를 활용할 수 있습니다.

01 기본 서비스

하나의 플랫폼에 통합된 서비스의 전체 데이터 라이프사이클

수집

Db2 Eventstore
실시간 분석을 위한 메모리 최적화 데이터베이스, 이벤트 중심 앱 및 머신러닝 기반 AI를 사용하여 하루에 2,500억 건 이상의 이벤트를 저장하고 분석합니다.

Data Virtualization
데이터를 이동하지 않고 여러 데이터 소스에서 빠르고 쉽게 쿼리를 수행할 수 있습니다.

Db2 Warehouse
프라이빗 클라우드와 컨테이너화된 배포 환경을 위해 탁월한 운영상의 유연성을 제공하는 클라이언트 관리형 데이터 웨어하우스입니다.

BigSQL
엔터프라이즈급 하이브리드 ANSI 호환 SQL-on-Hadoop 엔진으로, MPP(Massively Parallel Processing) 및 고급 데이터 쿼리 기능을 제공합니다.

구성

Master Data Connect
조직의 중요 데이터를 정의하고 관리하는 포괄적인 방법입니다. 전사적 범위의 데이터에 대해 신뢰할 수 있는 단일 뷰를 제공하며, 애자일 셀프 서비스 액세스, 분석 그래프 기반 탐색, 거버넌스 및 사용자 친화적 대시보드도 제공합니다.

Watson Knowledge Catalog
Data Refinery 포함
비즈니스에 바로 활용할 수 있는 데이터를 카탈로그로 작성하고 이해, 관리, 분석하며 데이터 시민에게 제공합니다.

Information Governance Catalog
WKC에 포함됨
IBM InfoSphere Information Governance Catalog(IGC)는 정보를 탐색, 이해, 분석하도록 지원하는 웹 기반 툴입니다.

Open Source Management
오픈소스 소프트웨어 관리(OSS)를 제공합니다.

Instascan
네이티브 인덱스를 사용하여 클라우드 데이터에 대한 가시성을 제공합니다. 업계에서 인정받는 샘플링 알고리즘을 활용하여 신속하고 신뢰할 수 있는 방식으로 위험 평가 및 규정 준수 확인 작업을 수행합니다.

InfoSphere Information Analyzer
WKC에 포함됨
데이터의 내용과 구조를 정확하게 평가하여 일관성과 품질을 유지할 수 있도록 데이터 프로파일링과 분석을 제공합니다.

Regulatory Accelerator
WKC에 포함됨
머신러닝의 강력한 기능을 활용하여 규정 준수 절차를 효율화합니다.

분석

Watson OpenScale
AI에 신뢰성과 투명성을 구현합니다. AI 모델이 의사결정을 내리는 방식을 이해하여 편향을 탐지하고 완화합니다.

Watson Studio
데이터 사이언스를 간소화하고 확장하여 비즈니스 성과를 예측하고 최적화합니다.

Watson Machine Learning
Auto AI 포함
머신러닝 모델을 적정 규모로 프로덕션 환경에 구축합니다.

Guardium(통합 구성요소)
외부 SIEM(Security Information and Event Management) 시스템과 통합합니다.

Analytics Engine for Apache Spark
대규모 분산 데이터 처리 및 머신러닝을 위한 강력한 통합 분석 엔진입니다.

IBM Streams
기회를 더 빨리 발굴할 수 있도록, 모든 소스에서 얻은 지속적으로 사용 가능한 데이터를 활용합니다.

Cognos Dashboard
API 기반 솔루션으로, 개발자가 쉽게 애플리케이션에 End-to-End 데이터 시각화 기능을 추가하도록 지원합니다.

관리 간소화

권한 부여
관리 권한과 편집 권한을 적합한 사람에게 부여하여 분석 기능을 편리하게 활용할 수 있도록 합니다.

시각화
데이터 세트에 지정된 데이터 거버넌스 규칙을 신속하게 확인하는 것부터 다양한 기능의 컴퓨팅 성능 소비 현황을 추적하는 것까지 가능하게 해줍니다. 전체 데이터 자산을 볼 수 있는 단일 창구를 제공합니다.

협업
팀 인사이트 획득 시간을 단축합니다. 여러 역할과 팀이 하나로 운영될 수 있는 상호 운용 가능한 단일 플랫폼에서 워크플로우를 중앙집중화 합니다.

모델 개발

구축
AI 라이프사이클의 이 단계에서 기업들은 적합한 알고리즘을 사용하여 예측을 위한 모델을 구축해야 합니다.

실행
맞춤형 모델을 애플리케이션 또는 비즈니스 프로세스의 프로덕션 환경에 배포합니다. 모델을 구축한 후 실행할 때 중요한 것은 '모델을 신뢰성과 투명성을 유지하면서 어떻게 확장할 수 있을까?'라는 질문이 될 것입니다. 아울러, 이것이 바로 관리가 중요한 역할을 하는 이유이기도 합니다.

관리
조직은 관리가 적절하게 이루어지도록 함으로써 모델을 변경한 사람, 모델 배포 시기, 그리고 모델상의 계보 등을 추적할 수 있습니다. 또한, 조직은 이러한 모든 사항을 추적하여 모델이 편향되지 않았고 설명 가능하며 투명하다는 것을 확인할 수 있습니다.

02 확장 서비스

활용 가능한 추가 기능으로 구성된, 성장 중인 에코시스템

Cognos Analytics
Cognos Analytics는 데이터 디스커버리부터 작동까지 전체 분석 주기를 지원하는 AI 기반 비즈니스 인텔리전스 플랫폼입니다. 데이터에 대한 실행 가능한 인사이트를 시각화 및 분석하고 조직의 모든 사람과 함께 공유할 수 있습니다.

Db2
IBM Db2®는 온프레미스와 프라이빗 및 퍼블릭 클라우드 환경에서 정형/비정형 데이터를 모두 관리하도록 지원하는 AI 기반 기능 모음을 제공하는 하이브리드 데이터 관리 제품군입니다. Db2는 확장성 및 유연성을 위해 설계된 지능형 공통 SQL 엔진을 기반으로 구축되었습니다.

Decision Optimization
IBM Decision Optimization은 조직이 더 나은 결정을 내리고 비즈니스 성과를 향상하도록 지원하는 처방적 분석(prescriptive analytics) 기능을 제공하는 최적화 소프트웨어 제품군입니다.

Financial Crimes Insights(FCI)
AI 및 규정 준수 관련 전문성을 활용하여 금융 범죄를 탐지하고 줄이는 프로세스를 간소화합니다.

Financial Services Workbench
AI 뱅킹 앱을 더 빨리 구축하도록 지원하는 마이크로서비스를 설계, 구축, 제공합니다.

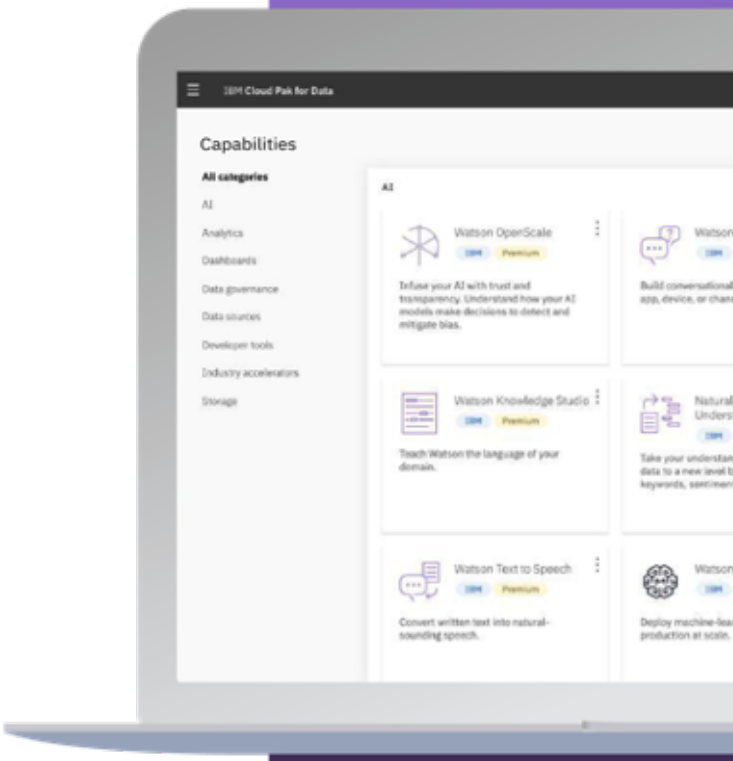
Information Server
IBM InfoSphere® Information Server는 데이터를 이해, 정리, 모니터링, 변환, 제공하고 비즈니스와 IT 부문 간의 간극을 협업을 통해 좁히도록 지원하는 제품군을 포함하는 시장 선도적 데이터 통합 플랫폼입니다. InfoSpherer Inforamtion Server는 MPP(Massively Parallel Processing) 기능을 통해 규모에 상관없이 모든 데이터 볼륨을 처리하는 확장성과 유연성이 뛰어난 통합 플랫폼을 제공합니다.

InfoSphere DataStage
IBM® InfoSphere® DataStage®는 소스 시스템에서 타겟 시스템으로 데이터를 이동하는 작업을 개발하기 위한 그래픽 프레임워크를 제공합니다. 변환된 데이터는 데이터 웨어하우스, 데이터 마트, 운영 데이터 스토어, 실시간 웹 서비스 및 메시징 시스템, 그리고 기타 엔터프라이즈 애플리케이션에 제공될 수 있습니다.

Master Data Management(MDM)
Master Data Management(MDM)는 조직의 중요한 데이터를 정의하고 관리하기 위한 포괄적 방법을 제공합니다. 전사적 범위의 데이터에 대해 신뢰할 수 있는 단일 뷰를 제공하며 애자일 셀프 서비스 액세스, 분석 그래프 기반 탐색, 거버넌스 및 사용자 친화적 대시보드를 제공합니다.

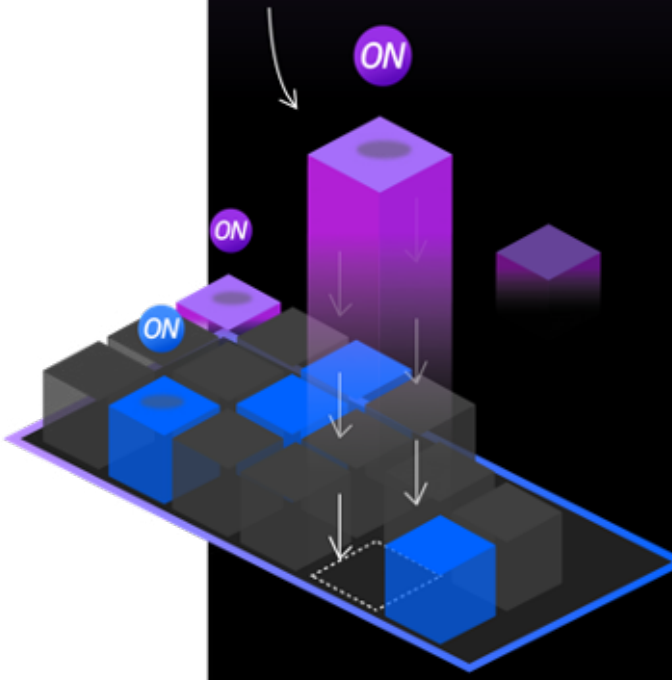
Planning Analytics
IBM Planning Analytics는 계획 수립, 예산 책정, 예측을 자동화하고 더 지능적인 워크플로우를 만들기 위해 AI를 사용하는 통합 계획 솔루션입니다.

SPSS Modeler
드래그 앤 드롭 방식의 직관적인 데이터 사이언스 툴을 활용하여 ROI를 증대하고 가치 실현 시간을 단축합니다.



부분의 합보다 큰 시너지 효과

Cloud Pak for Data의 상호 운용성을 활용하여 더 많은 기능을 연결합니다.



Virtual Data Pipeline(Actifio)
데이터 사일로를 데이터 파이프라인으로 전환하여 데이터 중심 엔터프라이즈에 강력한 지원을 제공합니다.

Watson Studio Premium
Watson Studio Premium은 AI에 대한 투자를 통해 가치 실현 시간을 단축하도록 돕습니다. 이 솔루션에는 IBM Decision Optimization, SPSS Modeler, Hadoop Execution Engine이 포함됩니다. 이와 같은 강력한 기능의 결합으로 단일 통합 플랫폼에서 비즈니스 성과를 예측하고 최적화하는 데 도움을 줍니다. 시민 데이터 사이언티스트들이 드래그 앤 드롭 방식의 인터페이스를 사용하여 데이터 준비 및 모델 생성을 수행할 수 있도록 지원할 수 있습니다. 또한, 데이터가 위치한 곳에서 분석을 수행하므로 데이터를 이동할 필요가 없으며 Hadoop에 대한 투자를 다시 활용할 수 있습니다.

Watson API Kit
Watson™ API Kit는 AI 애플리케이션을 구축하거나 다른 애플리케이션을 개선하는 것을 목표로 하나의 제품 안에 여러 Watson API를 결합하고 선택할 수 있는 기능을 제공합니다.

Watson Assistant
IBM Watson™ Assistant를 사용하면 어느 디바이스, 애플리케이션 또는 채널에서든 자체 브랜드 어시스턴트를 구축할 수 있습니다. 자체 어시스턴트는 이미 사용 중인 고객 관계 리소스에 연결할 수 있으므로 고객에게 매력적이고 통합된 문제 해결 경험을 제공할 수 있습니다.

Watson Discovery
IBM Watson™ Discovery는 수상 경력에 빛나는 AI 검색 기술로, 데이터 사일로를 제거하고 엔터프라이즈 데이터 안의 숨겨진 정보를 찾아냅니다. Watson Discovery는 자연어 처리 기능을 포함하여 머신러닝 분야의 최신 기술적 성과를 적용함으로써 사용자의 분야에 맞게 쉽게 훈련할 수 있습니다. IBM Cloud의 유연성 덕분에 사용자는 어디에서나 작업할 수 있습니다.

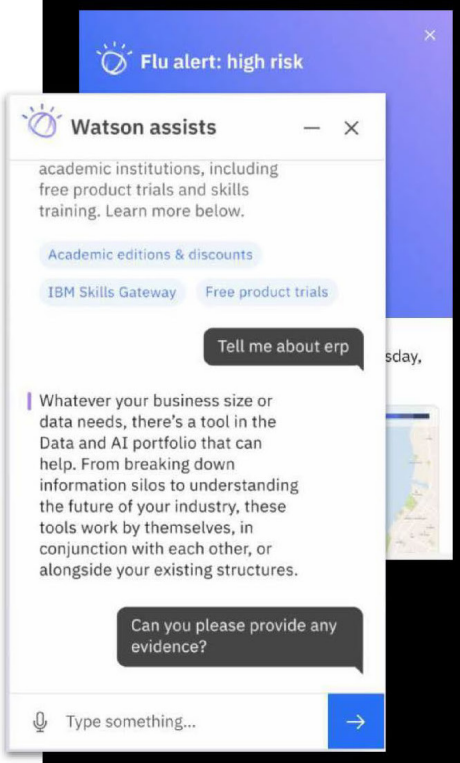
03 두 개의 에디션

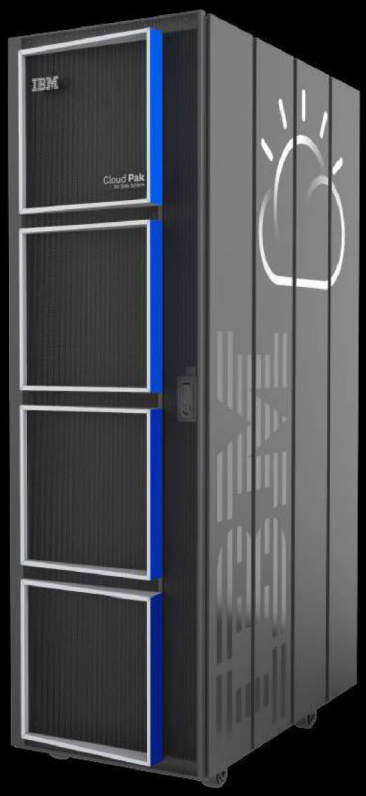
Standard Edition
Standard Edition은 몇 가지 유즈 케이스에 대해 Cloud Pak for Data를 시작하기에 좋은 방법입니다. 모든 기능을 사용할 수 있지만 최대 64 VPC로 사용이 제한됩니다.

Enterprise Edition
Enterprise Edition은 엔트리 레벨 유즈 케이스에서 전사적 범위의 적용까지 완벽하게 확장됩니다. 더 큰 컴퓨팅 용량을 요구하는 기능을 활용하고, 시스템과 함께 방화벽 내의 플랫폼을 활용하고, 엔터프라이즈 데이터 전략을 한 단계 더 향상할 수 있습니다.

04 Cloud Pak for Data System

Cloud Pak for Data System
스토리지, 컴퓨팅, 네트워킹, 소프트웨어를 결합하는 하이퍼컨버지드 시스템입니다. Cloud Pak for Data가 사전 설치된 경우, 이 모듈형 시스템으로 방화벽 뒤에 End-to-End 플랫폼을 구축하는 효과를 얻을 수 있습니다.





Cloud Pak for Data System

현대화

하이브리드 데이터 관리

IBM Db2 on Cloud

탄력적 확장성 및 자율 장애 조치 기능과
IBM Cloud 및 Amazon Web Services를
완전히 관리하는 관계형 데이터베이스

자세히 알아보기

IBM

IBM® Db2® on Cloud는 완전히 관리되는 SQL 클라우드 데이터베이스로, IBM Cloud™ 및 Amazon Web Services에서 프로비저닝해 하드웨어 및 소프트웨어 설치와 일반적인 유지관리에 소요되는 시간 및 비용을 없앨 수 있습니다. Db2 on Cloud는 Db2 제품군에 원활한 호환성, 통합 및 라이선싱 기능을 제공해 데이터의 이동성과 유연성을 크게 향상합니다. Db2 제공 에코시스템을 통해 기업에서는 기록 시스템을 사용하지 않고 데이터 소스 또는 유형에 상관없이 진정한 분석 정보를 얻을 수 있습니다. Db2 on Cloud와 Db2 제품군은 하이브리드 멀티클라우드 아키텍처를 지원하여 데이터 소스에서 지능형 분석에 대한 액세스, 비즈니스 전반에 대한 정보 그리고 변화하는 워크로드 및 소비 사례 지원을 위한 유연성을 선사합니다. 클라우드 네이티브 애플리케이션 빌드, Db2의 완전히 관리되는 인스턴스로의 이행 또는 재해 복구 워크로드 제거를 모색하든 상관없이 Db2 on Cloud는 신속한 쿼리 실행 및 엔터프라이즈급 애플리케이션 지원에 필요한 유연성 및 기민성을 자랑합니다.

주요 내용

- 고가용성 계획 및 원격지 재해 복구를 통한 99.99% 가동 시간 SLA
- 장애 조치를 위해 Db2 HADR, 휴대용 IP 및 자동 클라이언트 라우팅 변경 활용
- 가동 중지 시간 없이 90초 이내에 모든 크기의 데이터베이스 확장
- 사용한 만큼만 요금을 지불하는 스토리지 및 컴퓨팅의 탄력적 확장성
- SQL 편집기 및 기본 도구를 갖춘 내장형 웹 콘솔
- Db2 Advanced Enterprise 엔진으로 최대한 완전히 관리됨

Db2 on Cloud의 기능 및 이점

보안 및 재해 복구

클라우드 데이터베이스는 애플리케이션을 보호해야 하고, 기능, 인프라, 운영, 네트워크 및 물리적 보안을 제공하는 플랫폼에서 실행할 수 있는 기술을 제공해야 합니다. IBM Db2 on Cloud는 저장 및 전송 중인 데이터를 모두 암호화해 이를 달성하므로 데이터가 수명 주기 전체에서 더욱 안전해집니다. IBM Db2 on Cloud는 플랫폼 서비스 및 리소스 액세스 제어에 대한 사용자 인증을 사용해 승인된 당사자에게만 데이터 사용을 허용합니다. 이러한 보호의 일환으로, 모든 Db2 on Cloud 데이터베이스 인스턴스는 IBM ID 및 액세스 제어 관리 도구를 사용할 수 있습니다.

조직에서 다양한 데이터 및 보안 규정을 준수할 수 있도록, IBM Cloud 및 AWS에 대한 Flex 및 다양한 Precise Performance 요금제에는 다음 보안 표준 및 감사 프로세스가 포함되어 있습니다.

- ISO
- HIPAA
- SOC 2 Type 2
- GDPR
- Privacy Shield

Geo-Replicated Disaster Recovery 노드가 포함된 고가용성 요금제
데이터를 악의적인 외부자로부터 보호하고, 액세스할 필요가 없는 내부 사용자의 액세스를 제한하더라도 오류 또는 재해로 인해 데이터가 손실되거나 데이터 액세스 권한이 침해될 수 있습니다. Db2 데이터베이스는 강력한 백업 및 가용성 기능을 사용하여 이러한 위험을 완화하는 것을 목표로 삼고 있습니다. IBM Db2 on Cloud는 데이터베이스 복사 옵션을 사용해 14일 동안 보관할 수 있는 일일 백업 기능을 제공합니다. 이러한 백업을 통해 원치 않는 이벤트가 발생하기 전의 특정 시점으로 Db2 인터페이스를 복원할 수 있습니다.

Db2 on Cloud는 99.99%의 가동 시간 SLA, 원활한 장애 조치 및 롤링 업데이트 기능이 제공되는 고가용성 요금제를 제공합니다. 이 모든 기능이 사용자를 대신해 ACR(Automatic Client Reroute) 및 휴대용 IP를 사용하여 관리됩니다. 이는 1차 서버에서 장애 발생 시 즉시 1차 서버의 역할을 인계하는 2차 서버를 사용하여 애플리케이션에 고가용성을 쉽게 통합하도록 도와주는 탄력적 IP 기술을 포함해 달성할 수 있습니다. 원격지 복구에 추가 Geo-Replicated Disaster Recovery노드를 사용할 수 있어 원격지 IBM Cloud 데이터 센터에 있는 데이터베이스 노드로 사용자의 데이터를 실시간으로 동기화할 수 있습니다. 따라서 지역에서 계속해서 이어지는 정전 또는 재해에도 불구하고 운영 중단이 발생하지 않습니다.

진정한 하이브리드 아키텍처를 위한 데이터 페더레이션
Db2 on Cloud는 데이터 가상화를 지원하여 조직 내 어떤 위치에서든 모든 데이터에 단일 쿼리로 액세스할 수 있습니다. 이는 Db2 on Cloud를 기반으로 하는 공통 SQL 엔진 덕분입니다. 따라서 데이터베이스, 데이터 웨어하우스, 오픈 소스 프로젝트 및 기존 Netezza® 서비스를 비롯해 Db2 제품군 전체에서 애플리케이션이 클라우드 또는 온프레미스에서 작업할 수 있습니다. 또한 공통 SQL 엔진은 내장 데이터 가상화를 지원하고, Oracle Application Compatibility 계층을 포함해Oracle 애플리케이션이 IBM Db2 제품군과 통합할 수 있습니다. 간단히 말해, 공통 SQL 엔진을 활용하면 조직에서 SQL 코딩을 한 번 작성한 후 모든 위치에 가상으로 배포할 수 있습니다.

이러한 코드 유연성은 여러 가지로 매우 유용합니다. 무엇보다, 여러 데이터 관리 솔루션에서 코드를 다시 작성할 필요가 없으므로 시간을 절약해 부가가치가 있는 프로젝트에 투입할 수 있습니다. 마찬가지로, 조직 전체에서 데이터에 대한 액세스가 개선되어 더 많은 변수를 고려하고 보다 종합적인 정보를 얻을 수 있는 기회를 선사합니다. 또한 공통 코드 기반을 통해 데이터 전송의 복잡성에 대한 우려 없이 추가 클라우드 용량으로 아키텍처를 확장할 수 있어 확장성 역시 더욱 커집니다. 이는 클라우드 및 온프레미스 배포 전반에서 부하 분산의 경우에도 마찬가지입니다. 마지막으로, 공통 SQL 엔진은 코드 정렬을 위한 광범위한 연결을 구축할 필요 없이 마이크로서비스가 함께 더욱 원활하게 작동할 수 있도록 합니다.

여러 퍼블릭 클라우드 공급자에 배포 가능
Db2 on Cloud는 현재 IBM Cloud 및 AWS(Amazon Web Services)에서 모든 핵심 기능과 함께 배포할 수 있어 Db2 엔진에 고유한 기능을 모두 제공하고 동시에 선택한 데이터 인프라에서 실행할 수 있습니다. 여러 클라우드에서 작동하기 때문에 표준화된 한 가지 클라우드에서는 찾을 수 없는 더 많은 데이터 분석 기능을 활용할 수 있습니다. AWS에 배포된 Db2 on Cloud 워크로드는 고가용성 옵션, 롤링 업데이트, 프라이빗 VPC(요청 시), 매우 빠르게 프로비저닝되는 IOP 및 온프레미스에서 복제를 지원합니다.

AWS에서 사용할 수 있는 완전히 관리되는 Db2 on Cloud 버전 이외에 Amazon AWS Marketplace를 사용하여 전체 루트 액세스 권한을 가지고 AWS 가상 머신에서 Db2를 실행할 수 있습니다. 기존 Db2 라이선스도 사용할 수 있습니다.

또한 AWS에서 IBM Cloud Pak™ for Data를 사용하여 프라이빗 클라우드에서Db2를 실행할 수도 있습니다. 이전에 AWS에서 실행할 수 없었던 Watson™, Cognos® 및 기타 솔루션 내의 IBM 제품 역시 IBM Cloud Pak for Data 및 Db2를 활용하는 강력한 솔루션의 일부가 될 수 있습니다. 이는 데이터가 이미 AWS에 있는 경우 특히 중요해, 다른 클라우드 공급자로 비즈니스를 이전하지 않아도 됩니다.

현재 프로세스에 대한 변경이 최소화되어 기존 Db2 온프레미스 워크로드를 클라우드로 간단하게 마이그레이션할 수 있습니다. 현재 Db2 자격이 있는 경우 할인된 비용으로 Db2 on Cloud 서비스에 액세스할 수 있습니다.

스토리지 및 컴퓨팅 능력의 독립적 확장
Db2 on Cloud는 직관적인 콘솔에서 동적 슬라이더를 사용하여 조정할 수 있는 RAM, 스토리지 및 컴퓨팅 코어에 대한 독립적인 확장을 제공합니다. 독립적인 확장을 통해 조직에서는 항상 동일한 컴퓨팅 용량을 유지하는 비용을 감당할 필요 없이 필요한 경우 피크 워크로드 수요를 충족할 수 있습니다. 사용자는 컴퓨팅 코어를 슬라이더를 사용해 바로 늘리고, 번거롭게 요금제를 변경할 필요 없이 또는 컴퓨팅 코어 증가 결정에 따르는 추가 스토리지 또는 스토리지 비용 없이 나중에 줄일 수 있습니다. 따라서 유연성이 크게 개선되어 기업에서는 조직 및 시장의 고유한 상황에 따라 보다 정확한 맞춤 정보를 생성할 수 있습니다. 최초 Flex 요금제는 코어 1개, 4GB RAM 및 2GB의 디스크 공간으로 배포됩니다.

애플리케이션 요구사항을 충족하는 안전한 연결 옵션
JDBC, ODBC, CLI, .NET, PHP 및 REST API 등과 같은 공통 프로그래밍 언어를 사용해 Db2 on Cloud에 연결하는 애플리케이션을 생성할 수 있습니다. DataStage®, Segment, Data Studio, Cognos Analytics, Microsoft Excel 및 SPSS® 등과 같은 외부 애플리케이션 및 도구를 Db2 on Cloud에 연결해 데이터의 추가 관리 또는 분석에 사용할 수 있습니다. 애플리케이션은 서비스 프로비저닝 시 제공된 공용 호스트 이름 또는 요청한 경우 VPN(가상 사설망)을 통해 액세스할 수 있는 격리 네트워크에서 비공개 호스트 이름을 통해 연결할 수 있습니다. 두 가지 경우 모두 데이터 과학자 및 개발자가 원하는 도구 및 언어에 액세스해 다시 학습하는 데 시간을 허비할 필요 없이 기존 기술을 활용할 수 있고 최상의 성과를 내는 데 필요한 유연성을 가질 수 있음을 의미합니다.

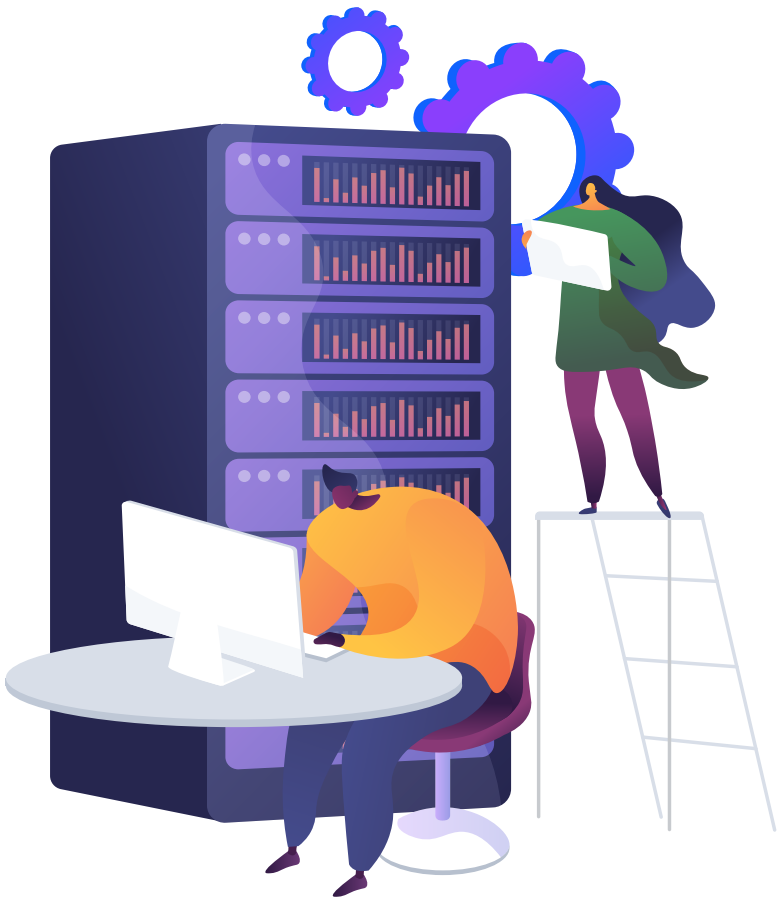
수집&구성

분석 플랫폼

IBM Db2 Warehouse

프라이빗 클라우드의 소프트웨어 정의 환경을 사용하는 하이브리드 데이터 웨어하우징

자세히 알아보기



IBM

데이터 웨어하우스의 진화

오늘날의 증가하는 분석 수요에 대응하여 대규모 온프레미스 데이터 웨어하우스 환경을 관리하는 일은 복잡하고 많은 비용이 들 수 있습니다. 이제 분석 비용을 낮추고 전례 없는 유연성을 실현하고 더 심층적인 인사이트를 제공하는 하이브리드 데이터 웨어하우스 설계를 고려할 시점입니다.

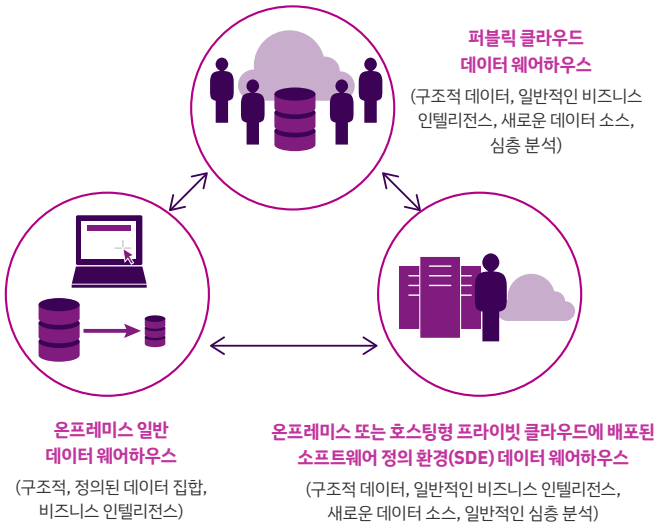


그림 1: 하이브리드 데이터 웨어하우스 아키텍처

하이브리드 데이터 웨어하우스는 전통적인 데이터 웨어하우스 기능을 확장하여 데이터, 분석, 위치의 새로운 조합을 지원하는 데 필요한 주요 기능을 제공하는 동시에 다음과 같은 IT 과제를 해결합니다.

- 시간에 민감한 비즈니스 이니셔티브를 충족하기 위한 새로운 분석 서비스 및 데이터 집합 제공
- 새로운 데이터 소스와 분석 기능, 사용자의 급증으로 인해 높아지는 비용 관리
- 민감한 비즈니스 데이터에 대한 데이터 웨어하우스 탄력성과 민첩성 달성

IBM Db2 Warehouse는 소프트웨어 정의 환경을 사용하여 하이브리드 데이터 웨어하우징 실현

IBM® Db2® Warehouse는 사전 구성된 클라이언트 관리형 데이터 웨어하우스로, 프라이빗 클라우드, 가상 프라이빗 클라우드 및 기타 컨테이너 지원 인프라에서 실행됩니다. 이 데이터 웨어하우스는 데이터에 대한 통제력을 유지해야 하지만 클라우드와 같은 유연성을 원하는 기업에 이상적인 솔루션을 제시합니다. 쿼리에 빠르게 응답을 제공하기 위한 인메모리 처리, 그리고 수요 증가에 따른 수평 및 수직 확장에 유용한 대규모 병렬 처리(MPP)가 포함됩니다. 분석의 경우 Db2 Warehouse 제품을 사용하여 익숙한 구조적 쿼리 언어(SQL), 통합 R 및 Python, 또는 지리공간 분석을 포함한 강력한 데이터베이스 내 분석을 활용할 수 있습니다.

Db2 Warehouse는 전체적인 IBM 하이브리드 데이터 웨어하우스 전략, 즉 역동적이고 움직임이 빠른 빅 데이터 및 클라우드 컴퓨팅 세계에 필요한 고유연성 아키텍처를 제공한다는 전략을 보완합니다. Db2 Warehouse와 Db2 Warehouse on Cloud는 공통 분석 엔진을 사용하므로 퍼블릭 클라우드와 프라이빗 클라우드 간에 애플리케이션 변경 없이 분석 워크로드를 이동할 수 있습니다. Db2 Warehouse 기술은 Db2 및 IBM PureData® System for Analytics(IBM Netezza® 기술 기반), 그리고 Oracle SQL과 호환됩니다. 애플리케이션에 따라 Db2 Warehouse 또는 클라우드로 더욱 쉽게 분석 워크로드를 옮길 수 있는 호환성을 제공합니다.

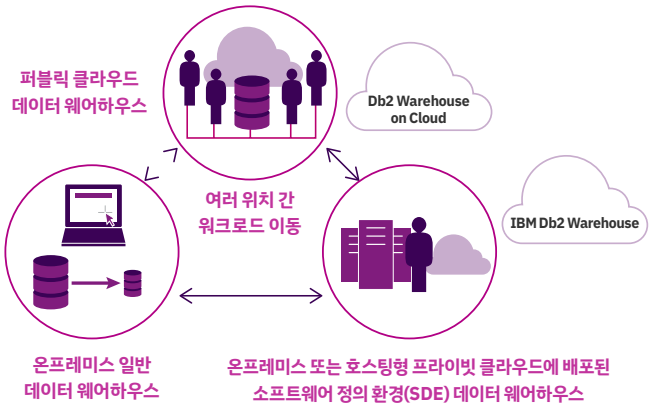


그림 2: Db2 Warehouse 제품이 있는 하이브리드 데이터 웨어하우스

Db2 Warehouse는 관리를 간소화하고 배포 시간을 분 단위로 줄여 주는 Docker 컨테이너 기술을 사용합니다. 또한 탄력적인 확장 기능을 제공하며 업데이트와 업그레이드도 용이합니다. 이러한 모든 기능은 IT 또는 클라우드 관리자에게 유용하도록 고안되었습니다. 사용자 관점에서 Db2 Warehouse는 데이터 집합을 신속하게 획득하고, 분석을 적용하여 특정 비즈니스 문제를 해결하고, 최적의 배포 옵션을 사용하여 인사이트를 실제 운용하는 데 필요한 성능을 제공하도록 해 줍니다.

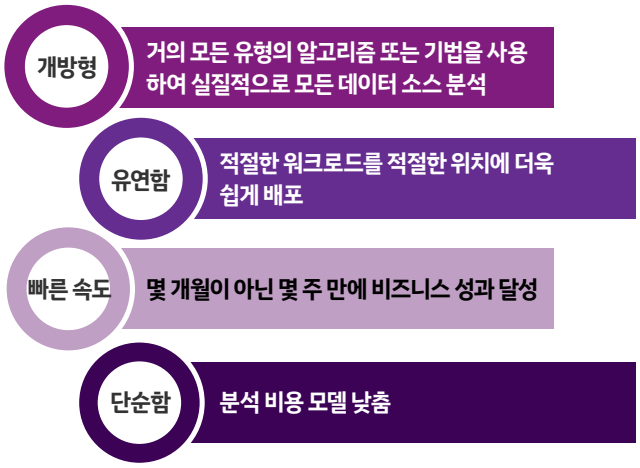


그림 3 : Db2 Warehouse 주요 차별화 요소

개방형

사용자들은 IT 부서에 R, Python Spark와 같은 새로운 분석 서비스를 제공하고 다양한 데이터 형식을 지원할 것을 요구합니다. 이러한 요청에 대응하지 못할 경우 분석이 사일로 형태로 난립하게 되고 데이터에 대한 통제력도 약화됩니다. Db2 Warehouse는 다양한 데이터 소스에 걸쳐 거의 모든 유형의 알고리즘 또는 기법을 사용하여 보다 쉽게 데이터를 처리하도록 설계된 개방형 솔루션을 제공함으로써 IT 부서가 통제력을 확보하도록 합니다. 또한 Linux, 클라우드, Apple OSX 및 Microsoft Windows 플랫폼상의 Docker 설치를 지원하는 표준 하드웨어에서 최소한의 사전 요건으로 Db2 Warehouse를 실행할 수 있습니다.

Db2 Warehouse의 개방적인 특성 덕분에 다음이 가능합니다.

- 데이터 요청에 따라 다양한 소스에서 더욱 쉽게 지리공간 및 객체 스토리지 데이터를 포함한 폭넓은 구조적 데이터와 비구조적 데이터를 로드할 수 있습니다.
- 익숙한 비즈니스 인텔리전스 도구와 오픈 소스 R, Python 및 Spark를 데이터베이스 내 처리에 사용할 수 있습니다.
- Esri ArcGIS를 연결해 지리공간 분석을 수행할 수 있습니다.
- 이미 데이터 센터에서 사용 중인 하드웨어를 활용 가능합니다.

유연성

조직 전반의 다양한 팀에서 다양한 데이터 소스와 도구를 사용하여 분석 솔루션을 만듭니다. 이러한 모든 솔루션을 하나의 애플리케이션으로 가져오기 위해서는 여러 번의 런타임이 필요할 수 있고, 온프레미스와 클라우드 배포를 오가는 최적의 데이터 흐름이 필요할 수도 있습니다. 또한 IT 부서는 비즈니스에 민감한 데이터, 예측할 수 없는 요구와 같은 최신 비즈니스 요구 사항에 대처하기 위해 효율적으로 워크로드를 관리해야 합니다.

Db2 Warehouse 기술은 하이브리드 웨어하우스 아키텍처를 통해 이러한 과제를 해결합니다. Db2 Warehouse는 클라우드와 같은 민첩성과 탄력성을 자랑하는 동시에 최신 프로그래밍 모델과 데이터 소스를 지원하기 위한 고급 분석을 제공합니다. Db2 Warehouse는 공통 데이터베이스 기술 제품군의 일부이므로 애플리케이션을 한 번만 작성하면 적절한 위치로 이 워크로드를 옮길 수 있습니다. 위치에는 퍼블릭 클라우드, 프라이빗 클라우드 또는 온프레미스가 포함되며 필요한 애플리케이션 변경은 최소 수준이거나 없습니다.

Db2 Warehouse의 유연성 덕분에 다음이 가능합니다.

- 분석 처리를 위해 Spark 또는 SQL을 실행할 수 있습니다.
- 퍼블릭 클라우드 또는 프라이빗 클라우드, 온프레미스 데이터웨어하우스를 포함한 위치 간에 워크로드를 이동할 수 있습니다.
- 페더레이션된 쿼리용으로 내장된 IBM Fluid Query를 사용하여데이터를 현재 위치에 그대로 둘 수 있습니다.
- 폭넓은 인프라 리소스에 걸쳐 탄력적 확장을 도입할 수 있습니다.

빠른 속도

확고한 데이터 웨어하우스를 보유한 조직도 분석을 통해 더 빠르게 비즈니스 성과를 달성하는 혜택을 누릴 수 있습니다. 버튼만 누르면 30분 이내에 배포가 완료되므로 사용자는 핵심 온프레미스 데이터 웨어하우스를 보완 및 확장하는 최적화된 프라이빗 클라우드 데이터 웨어하우스를 제공할 수 있습니다.

이 하이브리드 아키텍처를 통해 새로운 분석을 최소한의 위험으로신속하게 시작할 수 있습니다. SDE 프라이빗 클라우드 배포는 적절한 데이터와 분석 서비스를 조합한 셀프 서비스 프로비저닝을 통해 사용량이 낮은 리소스를 활용합니다.

사용자는 IBM BLU® Acceleration® 인메모리 처리를 위한 단일 노드와 MPP 아키텍처 및 Netezza 데이터베이스 내 분석과 Spark를 활용함으로써 대량의 데이터를 대상으로 신속하게 분석 모델을 구축하고 테스트할 수 있습니다.

단순함

SDE는 컴퓨팅, 스토리지, 네트워크 리소스를 포함한 전체 컴퓨팅 인프라를 최적화하도록 설계되었습니다. 또한 SDE는 필요한 워크로드의 요구 사항에 자동으로 스스로를 맞춥니다.

Db2 Warehouse는 Docker 컨테이너 기술을 통해 제공되며 SDE를 활용합니다. 예를 들어 리소스를 자동으로 프로비저닝하여 변화하는 워크로드 요구에 대처합니다.

또한 Db2 Warehouse는 탄력적인 확장과 손쉬운 업데이트 및 업그레이드로 배포와 관리의 효율성을 높입니다. Db2 Warehouse는 Spark를 포함한 전체 데이터 웨어하우스 스택을 몇 분 내에 프로비저닝하여 기존 운영 및 보안 프로세스를 그대로 유지하면서 퍼블릭 또는 프라이빗 클라우드의 서비스를 관리하도록 지원합니다.

Db2 Warehouse는 다음을 통해 간소함을 갖추도록 설계되었습니다.

- 배포와 관리를 쉽게 해 주는 컨테이너 기술
- 동적인 조정을 통해 변화하는 워크로드 요구를 충족하는 시스템 리소스
- Spark가 기본 포함되어 있으므로 따로 Spark를 설치 및 구성할 필요가 없음
- 내장된 고가용성과 재해 복구

Db2 Warehouse가 적절한 선택일까요?

Db2 Warehouse는 다음을 포함한 중요한 요구사항 충족에 도움이 될 수 있습니다.

- 탄력성 개선을 위해 SDE를 사용한 더 많은 데이터 웨어하우스 용량으로 서비스 수준을 지속적으로 충족하고 일반 하드웨어와 같은 기존 리소스 사용을 극대화
- 분석과 데이터 서비스의 적절한 조합에 신속하게 액세스하기 위해동적으로 프로비저닝할 수 있는 리소스
- 빠르고 쉬운 배포를 위해 설계되어 튜닝이나 관리가 거의 불필요한 웨어하우스 또는 데이터 마트
- 내부 요구 사항 및 기타 의무 사항으로 인해 데이터를 더 직접적으로 기업 통제하에 두거나 온프레미스에 보관하는 클라우드 전략
- 모바일, 웹, 사물 인터넷(IoT) 애플리케이션에 의해 생성되는 방대한양의 데이터에서 더 심층적인 인사이트를 얻기 위한 경제적인 고성능 처리 엔진
- 특히 구조적 데이터와 일반 하드웨어를 사용해 작업할 때 Hadoop에서 사용하기 위해 애플리케이션 코드를 다시 작성하는 것보다 경제적인 대안

시작하기: 사용 사례 예시

다음 사용 사례는 Db2 Warehouse를 시작하는 방법을 안내하기 위한 용도입니다.

- **프로토타입 제작, 개발 또는 테스트 생태계.** 프로덕션 구현에 앞서 빠르고 더 쉽게 새로운 애플리케이션과 데이터 소스를 테스트합니다.
- **부서별 또는 가속화된 분석 프로젝트.** 다양한 데이터 소스, 고급분석 및 애플리케이션 개발을 처리하기 위한 요구 사항을 충족할수 있는 분석 서비스를 신속하게 시작합니다.
- **서비스형 데이터 웨어하우스(DWaaS) 또는 하이브리드 데이터웨어하우스.** 애플리케이션, 데이터 또는 두 가지 모두를 온프레미스 웨어하우스에서 클라우드로 부분적으로, 또는 전체적으로마이그레이션합니다.

오랫동안 전통적인 온프레미스 데이터베이스 서비스를 제공해 왔고 약 5년 동안은 IaaS클라우드 서비스도 제공했습니다. 최근 처음으로 IBM 클라우드 데이터 및 분석 서비스를 사용하기 시작했으며 추가 고객 프로젝트와 애플리케이션에 활용하기 위해 Db2 Warehouse의 일반 출시를 기대하고 있습니다.

T-Systems

추가 정보

Db2 Warehouse에 대한 자세한 정보는 IBM 담당자 또는 IBM 비즈니스 파트너에게 문의하거나 이 웹사이트를 참조하십시오.

Db2 Warehouse를 사용해 보려면 호스트 서버에 Docker 엔진을 설치하고 Global File System 2(GFS2) 및 IBM GPFS™ 와 같은 POSIX 준수 클러스터 파일 시스템 스토리지를 구성하십시오.

또한 IBM Global Financing은 다양한 지불 옵션을 통해 회사가 비즈니스 성장에 필요한 기술을 확보할 수 있도록 도와드립니다. IBM은 구매에서 폐기까지 IT 제품 및 서비스의 전체 수명 주기 관리 서비스를 제공합니다. 자세한 정보는 다음 사이트에서 확인할 수 있습니다.
ibm.com/kr-ko/financing

자세히 알아보기

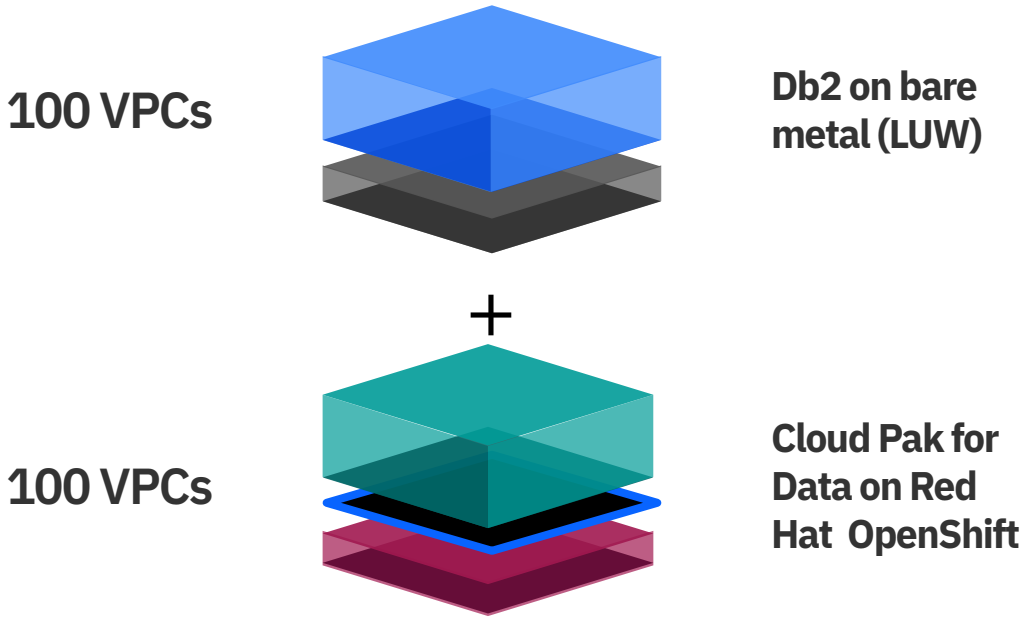
Db2 현대화 업그레이드

기존 사용중인 Db2 는 현대화 업그레이드를 통해 컨테이너 환경으로 이동할 수 있습니다. 고객은 원하는 속도로 현대화를 진행하면서 동시에 **Cloud Pak for Data 플랫폼** 사용의 추가 혜택을 누릴 수 있습니다.

Db2 라이센스	현대화 업그레이드
Db2 Advanced VPC(Virtual Processor Core)	Db2 Advanced Modernization
Db2 ESE (Enterprise Server Edition) 70 PVU	
Db2 AESE(Advanced ESE) 70 PVU	
HDMP(Hybrid Data Management Platform) VPC	
HDMP FP(Flex Point)	
Db2 Standard Modernization	
Db2 Standard VPC	Db2 Standard Modernization

※ Db2 ESE, Db2 AESE : pre 11.5 version, Db2 Advanced, Db2 Standard : 11.5 version

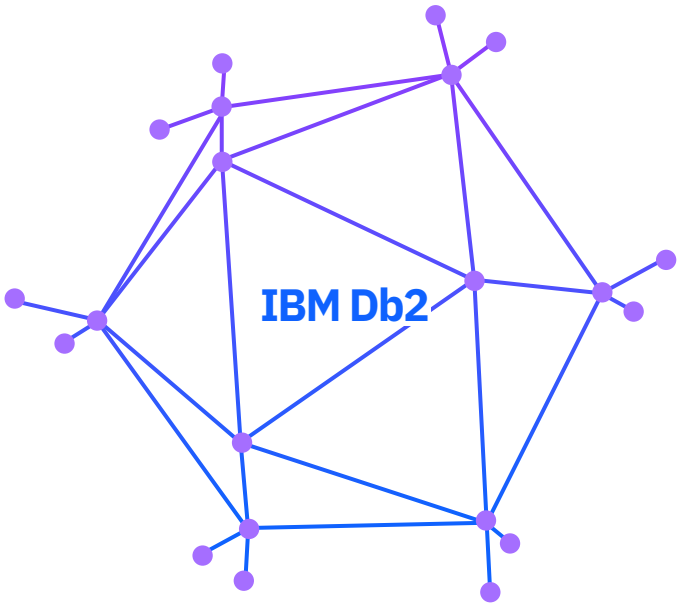
예시. Db2 라이센스를 100 VPC 업그레이드



Db2 11.5 = AI Database

우수한 쿼리 성능
- 머신러닝 Optimizer, 인메모리 분석

성능 저하 없는 실시간 데이터 압축
- 적응형 압축 기술



이기종 데이터 실시간 액세스
- 데이터 가상화 기술

무중단 아키텍처 (pureScale)
- 온라인 유지보수, 무중단 DBMS패치

폭넓은 SQL 개발
- ANSI + Oracle SQL

최신 AI 애플리케이션 개발
- Python, GO , JSON , Jupyter notebooks

수집&구성

하이브리드 데이터 관리

IBM Cloud Pak for Data Netezza Performance Server

- 온프레미스, 클라우드, 하이브리드 솔루션 환경 등 어디서나 이용 가능

자세히 알아보기

주요 내용

- nz_migrate anywhere라는 단일 명령어로 업그레이드 가능
- 특허받고 입증된 데이터 분석 기술
- 속도, 사용 용이성, 가치를 구현하도록 설계된 안전한 하이브리드 멀티클라우드 인사이트 플랫폼 시스템
- 규모에 맞는 데이터 사이언스 및 인데이터베이스(in-database) 머신러닝
- 데이터 가상화 — 다수의 데이터베이스를 하나의 데이터베이스인 것처럼 사용하여 어디에서나 무엇이든 쿼리 가능
- 레질리언스와 비즈니스 연속성을 위한 설계

데이터 홍수(Data deluge)는 데이터의 양, 다양성, 속도, 진실성을 관리하는 일과 연관 지으면 복잡성의 새로운 동의어가 됩니다. 이러한 데이터는 온프레미스와 클라우드의 다양한 저장소에 복제되면서 고비용, 프로젝트 지연, 보안 노출을 초래하는 데이터 사일로를 유발하고 있습니다. 따라서, 각기 다른 데이터 소스들로부터 가치를 얻는 일이 훨씬 더 복잡해집니다.

오늘날 조직은 이러한 복잡성을 관리하기 위해 스스로를 혁신하거나 역사 속으로 사라지는 것 중 어느 하나를 선택해야 하는 기로에 놓여 있습니다.

데이터 민주화와 데이터를 통한 수익 창출이 데이터의 가치를 실현하기 위한 핵심 요소이지만 이를 위해서는 지능적 분석이 효과적으로 수행되도록 지원하는 탄력적인 정보 아키텍처가 필요합니다. 인사이트 플랫폼 시스템인 IBM Cloud Pak® for Data는 이러한 복잡성을 해소하여 신뢰할 수 있게 관리되는 클라우드 네이티브 데이터 및 AI 플랫폼을 제공합니다. 클라우드에서건 온프레미스에서건 (하이퍼컨버지드 아키텍처로서), 이 시스템은 데이터 관리, 거버넌스, 분석을 간소화하고 통합합니다. Cloud Pak for Data를 기반으로 구축된 Netezza® Performance Server로 IBM은 다시 한 번 고객의 데이터 및 AI 여정의 판도를 바꾸고 있습니다.

Netezza Performance Server

Netezza Performance Server를 움직이는 것은 첨단 Cloud Pak for Data입니다. Cloud Pak for Data는 IBM Red Hat® OpenShift®를 기반으로 구축되어 모듈성, 레질리언스 및 확장성이 뛰어난 클라우드 네이티브, AI용 데이터 플랫폼입니다.

이 서버는 구성 가능한 마이크로서비스 기반 인사이트를 구축하도록 지원하는 진정한 클라우드 지원 소프트웨어 정의 환경을 제공합니다. 클라우드 네이티브 방식이므로 어디에서나 Netezza를 배포할 수 있습니다. 여기에는 IBM Cloud™와 Amazon Web Services(AWS)와 같이, 온프레미스와 프라이빗/퍼블릭 클라우드가 조합된 환경도 포함됩니다.

Netezza Performance Server는 데이터베이스, 서버, 스토리지, 첨단 분석 기능을 온프레미스 또는 IBM Cloud 및 Amazon Web Services에서 관리하기 쉬운 단일 플랫폼으로 통합하도록 특별히 설계된 표준 기반 데이터 및 AI 시스템입니다. 페타바이트급 규모로 확장되는 데이터 볼륨에 대해 신속하고 심층적인 분석을 수행하도록 설계된 이 시스템은 낮은 소유 비용으로 결코 불가능하다고 생각되었던 인사이트를 제공합니다. Netezza Performance Server 제품군은 엔트리 레벨부터 멀티 페타바이트 규모까지 다양하며, 광범위한 데이터 용량 요구사항을 충족합니다.

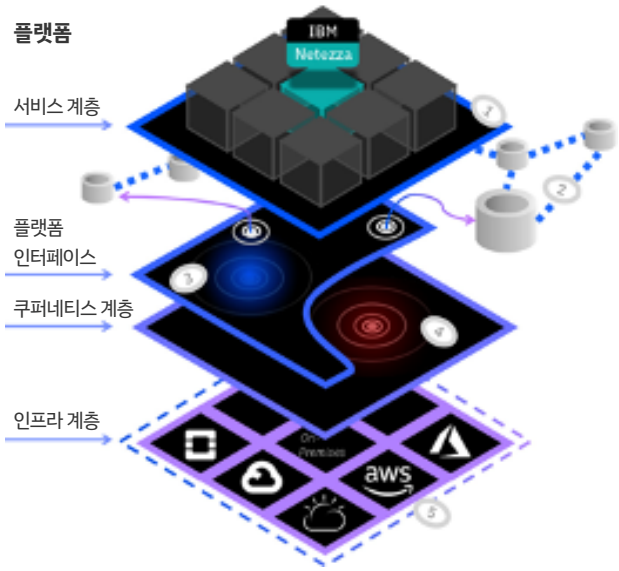
Netezza Performance Server는 비즈니스 니즈에 맞는 입증된 성능, 확장성, 인텔리전스, 단순성을 제공합니다. 이 시스템은 최초 배포 후 최소한의 처리 작업과 튜닝으로 지속적 관리가 가능하므로 총 소유 비용(TCO, Total Cost of Ownership)이 낮아집니다.

분석 능력을 보유할 수 있는 기회

경쟁이 심한 경제 환경에서는 다량의 데이터를 심층적으로 정교하게 분석할 수 있는 능력이 기업에게 반드시 필요합니다. 이러한 분석 능력은 경쟁사보다 더 강력한 경쟁력을 부여합니다. 그러나 대부분의 조직은 새로운 분석 기능의 경우 시장 출시 시간 때문에, 기존 분석 기능의 경우 서비스 수준 계약 유지 때문에 제약을 받습니다.

전략적 분석은 제공하기에 간편하고 관리가 쉬워야

데이터 분석에 대한 IBM의 접근법은 특허를 보유한 이미 입증된 기술입니다. 아울러, 이번 Netezza Performance Server 릴리스에서는 이전 시스템에 비해 훨씬 더 우수하게 여러 사항이 개선되었습니다. 대규모 병렬 처리 시스템이 지원하는 멀티 페타바이트 규모의 확장 기능은 고객이 기존 시스템을 사용할 때보다 훨씬 더 빨리 방대한 데이터 볼륨에서 인사이트를 얻도록 지원합니다. 그리고 이러한 작업은 데이터를 별도의 분석 서버에 복제하지 않고도 완료할 수 있습니다.



1. 서비스 생태계
클릭 한 번으로 IBM과 타사가 제공하는 45가지 이상의 분석 서비스와 템플릿을 이용하고 배포할 수 있습니다.

2. 데이터 가상화
데이터를 이동하지 않고도 여러 데이터 소스에서 쉽고 빠르게 쿼리를 수행할 수 있습니다.

3. 플랫폼 인터페이스
데이터 관리, 데이터 거버넌스 및 분석 기능을 통합하여 효율성과 리소스 활용도를 향상하는 단일 사용자 경험을 통해 가치 실현 시간을 단축합니다.

4. Red Hat OPENSHIFT®
혁신적이며 빠른 배포 전략을 지원하는 선도적인 하이브리드 클라우드 엔터프라이즈 컨테이너 플랫폼을 활용할 수 있습니다.

5. 모든 클라우드 지원
멀티클라우드 접근법으로 종속성을 피하고 모든 클라우드 인프라를 활용할 수 있습니다.

빠른 속도, 확장성, 단순성에 스마트함까지 겸비

Netezza Performance Server는 경쟁업체의 솔루션보다 더 짧은 실행 시간으로 대규모 데이터 볼륨에 대한 복잡한 분석을 실행하도록 특별히 설계되었습니다. 이 시스템은 조직이 데이터에서 인사이트를 얻는 데 필요한 검증된 성능, 확장성, 인텔리전스, 그리고 단순성을 제공합니다.

단순성

온프레미스, 클라우드 또는 이 모두가 결합된 하이브리드 환경 등 어디에서나 nz_migrate anywhere 라는 간단한 명령어로 즉시 업그레이드를 수행할 수 있습니다. Netezza Performance Server는 사용자의 속도에 맞춘 업그레이드를 통해 클라우드 또는 하이브리드 여정을 현대화하는 동안, 기존 Netezza (PureData® System for Analytics) 솔루션을 함께 사용할 수 있습니다.

이 시스템은 표준 ODBC, JDBC 및 OLE DB 인터페이스를 통해 선도적인 ETL, 데이터 및 AI 애플리케이션과 통합되며 기존 데이터 통합 및 BI 툴과 100% 호환성을 갖췄습니다. 업그레이드 시 기존 ETL 또는 타사 애플리케이션을 재지정하기만 하면 됩니다. Netezza와 관련하여 사용자가 선호하는 동일한 예측 가능한 선형 성능 확장성을 구현하는 경우에도 변경이 필요하지 않습니다.

Netezza Performance Server는 또한 관리자가 시스템 리소스를 모니터링 및 관리하고, 데이터베이스 오브젝트를 관리하며, 워크로드 관리를 구성하고, 활성 세션을 확인하며, 용량 계획을 위해 시스템 리소스 활용도를 모니터링하도록 지원하는 현대화된 통합 GUI를 제공합니다. 이 UI는 사용하기 편리하도록 하나의 액세스 지점에서 Netezza Performance Server를 지원하는 통합 관리 인터페이스를 제공합니다.

Netezza Performance Server는 고가용성과 레질리언스를 구현하도록 설계되었습니다. 시스템 내 하나의 프로세싱 노드 또는 구성요소에 장애가 발생해도 성능이 크게 저하되지 않습니다. 이러한 특징 덕분에 데이터센터에 시스템이 설치되는 순간부터 강력한 프로덕션 지원 환경이 조성됩니다.

개발 및 배포의 단순성과 용이성은 Netezza를 진정으로 차별화합니다. 이러한 특징은 최초 배포와 지속적인 유지보수를 위해 최소한의 관리 노력만을 요구하며 인덱싱과 튜닝 없이 즉시 성능을 발휘하는 핵심적인 차별화 요소입니다. 이를 통해 총 소유 비용을 낮추고 탁월한 가치 실현 시간을 달성할 수 있습니다. IBM이 모든 단계에서 복잡성을 해소해 주므로 사용자는 귀중한 리소스를 수익에 긍정적 영향을 주는 이니셔티브에 활용하기만 하면 됩니다.

빠른 속도

Netezza Performance Server는 다른 옵션과 비교할 수 없는 성능상의 이점을 제공합니다. 이러한 성능은 IBM의 특허 받은 하이브리드 컬럼식 가속화 지원(hybrid columnar acceleration assist)을 사용하는 고유한 AMPP(Asymmetric Massively Parallel Processing)로부터 구현됩니다. 온프레미스에서건, IBM Cloud와 Amazon Web Services에서건, 이 시스템은 데이터 이동을 최소화하면서 더 빠른 코어 및 고급 NVMe 플래시 드라이브와 같은 첨단 하드웨어 기술을 활용할 수 있습니다. 이러한 역량 덕분에 수천 명의 사용자를 지원하면서도 상당한 속도로 데이터를 처리하는 분석 워크로드가 가능한 것입니다. 이는 결국, 빠르지만 비용은 적게 드는 시스템임을 뜻합니다.

하이브리드 멀티클라우드 확장성

Netezza Performance Server 솔루션을 사용하면 데이터 볼륨과 워크로드에 적절한 크기의 환경을 구축할 수 있습니다. 또한, 이 아키텍처는 데이터 볼륨이 커짐에 따라 확장 요구사항이 변해도 쉽고 빠르게 확장할 수 있도록 지원하므로 안심할 수 있습니다.

설계상으로 클라우드 데이터 웨어하우스는 빠르고 탄력적이며 거의 무제한의 용량으로 확장할 수 있습니다. IBM은 이러한 이점을 한 차원 끌어올리기 위해 클라우드에서 Netezza를 구축했습니다. IBM Cloud와 Amazon Web Services에서 모두 지원되는 내장형 자동화 및 자동 복구 기능은 인간의 개입을 최소화하고 가용성과 내결함성이 뛰어난 Netezza 배포 환경을 제공하므로 데이터 웨어하우스를 연중무휴, 하루 24시간 운영할 수 있습니다.

Netezza는 또한 워크로드 크기가 증가함에 따라 선형적 성능 확장성을 제공하므로 조직은 적절한 크기의 시스템을 선택하여 데이터 볼륨과 성능 요구사항을 모두 충족할 수 있습니다. 이것은 예측 가능하고 확장 가능한 성능 덕분에 가능합니다. 또한, 데이터 볼륨이 커짐에 따라 배포 환경을 관리하고 유지하기 위해 상당한 리소스를 추가하지 않아도 됩니다.

클라우드 데이터 레이크가 IBM Cloud Object Store에 있건 AWS S3에 있건 고속 Netezza 외부 테이블로 액세스할 수 있습니다. 편리한 시간에 백업을 예약할 수 있고, 백업을 오브젝트 스토어에 보관하거나 여러 가용 영역 (availability zone)에 복제할 수도 있습니다.

물론, 모든 조직이 전적으로 클라우드 또는 온프레미스 환경만을 사용하려고 하지는 않습니다. 추가적인 선택 옵션을 제공하기 위해, IBM은 Netezza 클라우드 경험을 위한 원활한 Netezza 토대를 제공함으로써 조직이 하이브리드 아키텍처를 관리할 수 있도록 지원합니다. 데이터 가상화와 더불어 이 두 가지 배포 옵션 간에 완전한 호환성이 구현되므로 항상 가장 유리한 위치에 데이터를 배치하고 워크로드를 실행할 수 있습니다. 또한, 데이터를 다른 곳에서 이용해야 할 경우 비용이 많이 드는 ETL이나 마이그레이션을 수행할 필요가 없습니다.

스마트한 기능

Netezza Performance Server는 모든 분석 활동을 데이터가 보관되어 있는 한 장소에 통합하여 분석을 대폭 간소화합니다. IBM의 내장형 분석 플랫폼을 사용하면 분석 기능을 Netezza Performance Server로 쉽게 이동할 수 있습니다. 또한, IBM Cloud Pak for Data에서 제공되는 데이터 가상화 기능을 통해 데이터를 이동하지 않고 어디서나 데이터 소스를 쿼리할 수 있습니다. 즉, 여러 데이터베이스를 마치 하나의 데이터베이스인 것처럼 취급할 수 있습니다.

IBM Cloud Pak for Data용 Netezza Performance Server와 함께 제공되는 IBM Watson® Studio는 데이터 사이언티스트와 분석가가 데이터를 준비하고 어느 클라우드에서든 규모에 맞게 모델을 구축하도록 지원합니다. Watson Studio는 유연한 개방형 멀티클라우드 아키텍처를 통해 기업이 엔터프라이즈 데이터 사이언스 및 AI를 간소화하는 데 보탬이 되는 기능을 제공합니다.

데이터 사이언티스트는 모든 엔터프라이즈 데이터를 사용하여 모델을 구축한 다음 다른 모델들을 통해 훨씬 더 빠르게 반복하여 최고의 솔루션을 얻을 수 있습니다. 일단 모델이 개발되면, 해당 어플라이언스 내에서 관련 데이터를 대상으로 원활하게 실행할 수 있습니다. 예측 및 스코어링(scoring)도 데이터가 보관된 곳에서 수행할 수 있습니다. 사용자가 예측 점수를 거의 실시간으로 얻을 수 있으므로 고급 분석을 수행하는 데 도움이 되며 전사적으로 이를 이용할 수 있도록 지원할 수 있습니다.

Netezza Analytics는 내장형 분석 인프라와 통계 및 수학 함수로 구성된 광범위한 라이브러리를 제공하여, Open Source R을 포함한 다양한 분석 툴과 프로그래밍 언어를 지원합니다. Netezza Analytics는 200가지 이상의 사전 구축된 확장형 인데이터베이스 분석 함수로 구성된 라이브러리가 함께 제공되는데, 이러한 분석 함수는 분석 작업을 병렬적으로 수행하는 한편, 개발자, 사용자, DBA가 경험해온 병렬 프로그래밍의 복잡성을 해소해 줍니다.

Netezza Analytics 기능에는 또한 산업 표준 ESRI GIS 형식과 호환되는 선도적인 인데이터베이스 지리공간 분석 기능이 포함됩니다. 따라서 기존의 지리공간 분석 환경과 쉽게 통합할 수 있습니다.

최고의 경제성

Netezza Performance Server는 비용 효율적인 데이터 및 AI 시스템입니다. 이 시스템은 지속적 관리와 튜닝을 최소한으로 요구하므로 매우 낮은 총 소유 비용으로 내부 리소스 및 실행 비용을 최소화합니다. 튜닝, 인덱싱 또는 집계 테이블 없이 Netezza Performance Server의 성능과 확장성을 즉시 이용할 수 있습니다.

IBM의 오퍼링을 통해 중요 데이터 및 AI 이니셔티브의 가치 실현 시간을 단축할 수 있으므로 귀사는 더욱 정확한 인텔리전스를 바탕으로 기회와 위험이 나타날 때 신속하고 정확하게 대응할 수 있습니다. 변화하는 시장 상황과 증가하는 데이터 및 AI에 대한 수요에 대응할 수 있는 유연성이 기업에 요구되는 이 시점에, 증가하는 데이터 볼륨을 빨리 실행하고 분석할 수 있고 관리가 간편한 시스템을 채택하는 것은 매우 타당한 일일 것입니다.

Netezza Performance Server에 대해 자세히 알아보기

Netezza 제품은 범용 시스템의 유연성과 클라우드의 탄력성, 그리고 어플라이언스의 단순성을 모두 구현합니다. 이 제품은 설계 단계부터 통합되며, 간소화된 IT 경험을 수십년간 제공해온 경험을 바탕으로 한 전문성이 반영되어 있습니다.

온프레미스, IBM Cloud나 Amazon Web Services 또는 이 모든 환경에서 Netezza Performance Server는 관리하기 쉬운 대규모 병렬 처리 환경을 제공하며, 설치 및 지속적인 관리 작업은 최소한으로만 요구됩니다. 또한, 더 빠르고 일관된 분석 성능을 구현합니다. Netezza Performance Server는 데이터가 보관되어 있는 위치에서 바로 어플라이언스의 모든 분석 활동을 통합하여 데이터 및 AI를 지능적으로 대폭 간소화함으로써 업계를 선도하는 성능을 구현합니다.

이 전문 통합 시스템 제품군이 모든 단계에서 어떻게 복잡성을 제거하고 조직을 위해 진정한 비즈니스 가치를 실현하도록 지원하는지 알아보려면 ibm.com/kr-ko/analytics/Netezza를 방문하거나, IBM 영업대표 또는 비즈니스 파트너에게 문의하세요.

자세히 알아보기

분석

DataOps로 비즈니스에 활용 가능한 데이터를 빠르게 제공

IBM DataOps 방법론 및 사례 소개

자세히 알아보기

IBM

주요 내용

- DataOps는 데이터 사용자에게 신뢰할 수 있는 고품질 데이터를빠르게 제공하기 위해 인력, 프로세스와 기술이 조화를 이룬산물입니다.
- 자동화가 뒷받침하는 DataOps는 데이터에 대한 액세스, 데이터준비 및 통합 그리고 데이터를 사용 가능하게 만드는 데 수반되는비효율성과 관련된 문제를 해결합니다.
- IBM DataOps 사례는 극단적 자동화를 활용해 데이터 큐레이션서비스, 메타데이터 관리, 데이터 거버넌스, 마스터 데이터 관리및 셀프 서비스 상호 작용 등과 같은 기능에 눈에 띄게 큰 영향을미칩니다.
- 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 실행을계획하도록 지원합니다.

속도

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.[1]

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 집중조명하는 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. 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	- 모든 데이터 사용자를 위해 신뢰할 수 있는 고품질데이터에 대한 셀프 서비스 액세스를 가능하게 하여기업의 지속적이고 빠른 혁신 유도 - 규제에 대한 우려에 효과적으로 대응하면서 데이터거버넌스, 통합을 자동화하여 지속적 데이터 제공 실현 - 데이터 파이프라인을 모니터링 및 최적화하여 모든데이터 사용자가 계속 학습할 수 있는 피드백 루프 제공	- 공동 개발을 가능하게 하고 가치사슬 전반에서테스트하여 아이디어의 지속적인 혁신 가속화 - 소프트웨어 제공 프로세스를 자동화하고, 폐기물을줄여 규제에 대한 우려를 해결하도록 지원하면서지속적인 혁신 전달 실현 - 소프트웨어 기반 혁신을 모니터링 및 최적화하여고객이 계속해서 학습할 수 있는 피드백 루프 제공
효율성 목표	- IT 시스템 지원, 운영 및 비즈니스 간에 더욱 긴밀하게연계하도록 촉진해 인력과 목표의 잘못된 조정 교정 - 데이터 전달 사이클 전반에 자동화를 도입하여 변화가속화 및 전달 품질 개선 - 결과를 사용해 최적화를 구현하여 메타데이터 및데이터의 실제 가치에 대한 인사이트 개선	- 개발자, 운영 및 비즈니스 간에 더욱 긴밀하게연계해 인력과 목표의 잘못된 조정 교정 - 개발 사이클 전반에 자동화를 도입하여 변화를가속화하고 변화 전달 시 오류 제거 - 고객 피드백을 통해 최적화를 구현하여 애플리케이션의 실제 가치에 대한 인사이트 개선

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.

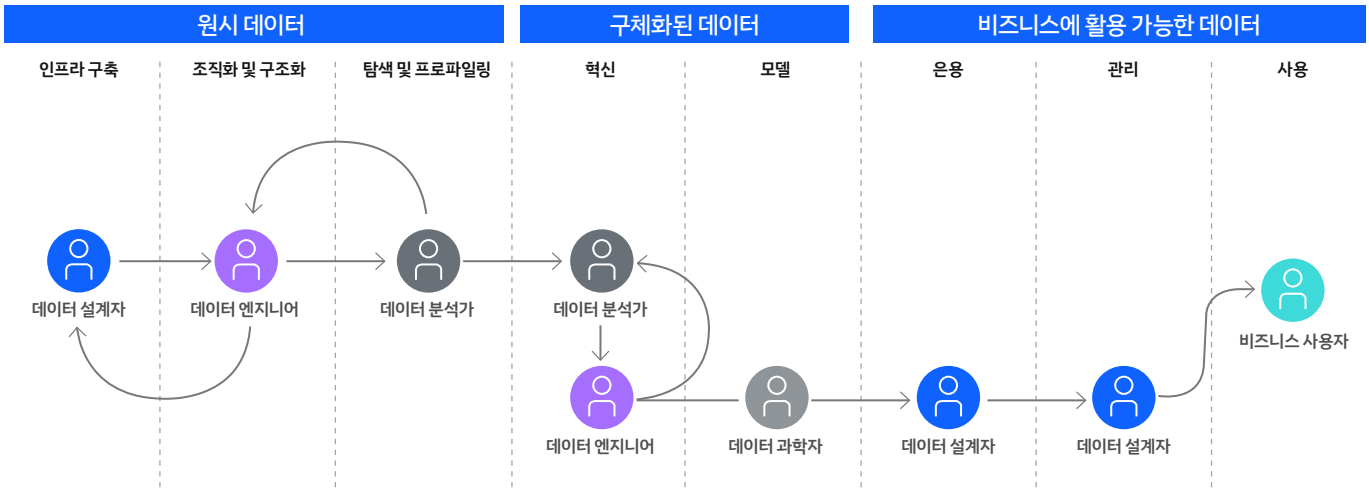


그림 1: 역할별 DataOps 워크플로의 예

기술

DataOps의 핵심에는 조직의 정보 아키텍처가 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

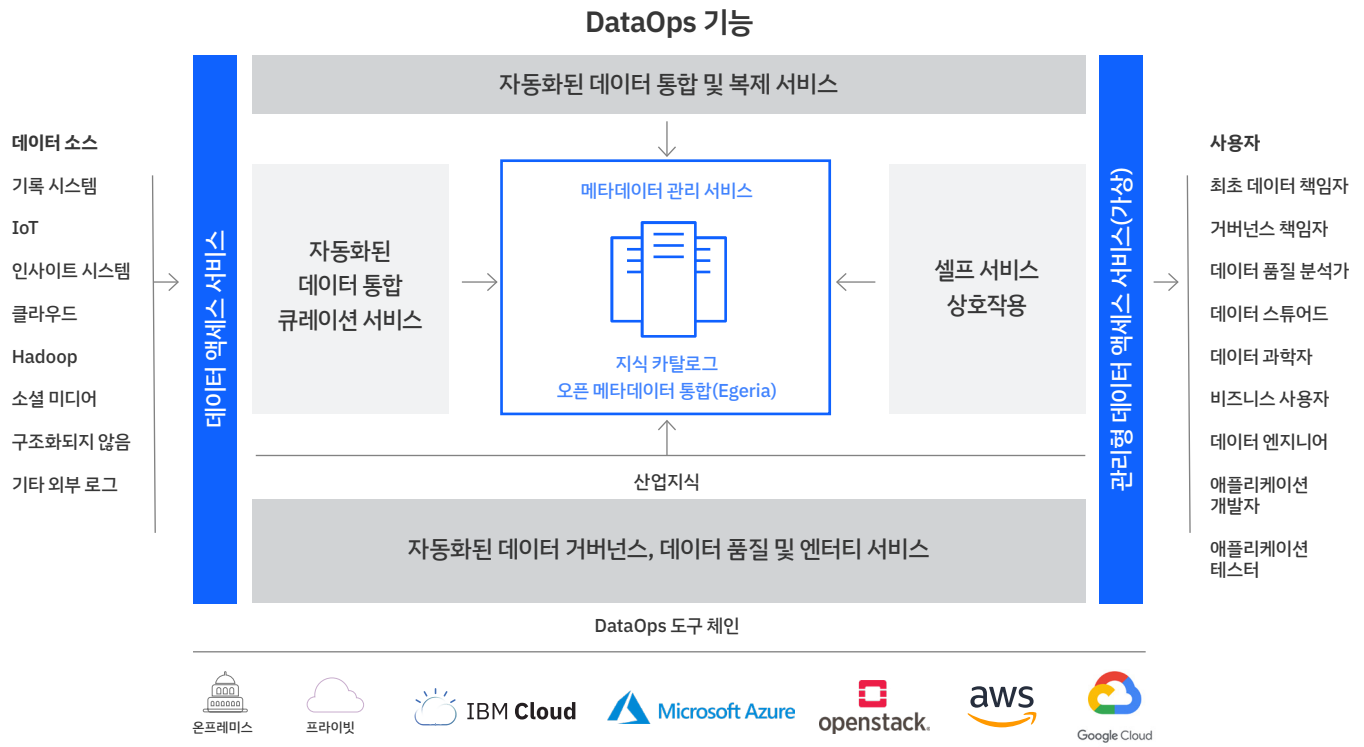


그림 2: DataOps를 지원하는 정보 아키텍처 살펴보기

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 힘듭니다. 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 벅차게 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.2 IBM에서는 규범적 방법론, 첨단 기술 및 IBM DataOps Center of Excellence(CoE)를 통해 DataOps 사례를 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 적임자에게 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 위해데이터를 파악하고, 신뢰하고, 사용하는 경우 고품질 엔터프라이즈데이터를 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 고려사항에 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 지속하든 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 작업을 수행하고, 데이터 클래스에 정책을 할당하고, 데이터 품질을 평가해 점수를 매기고, 스프레드시트, 조직 내에만 머무르는 지식 또는 수동 코딩과 달리 데이터를 통합하기 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 질문해 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.

- 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
- **혁신 파이프라인 :** 현재 개발 중이고 생산 파이프라인에 추가되는새로운 분석 기능의 미래로 이어짐.

이러한 두 파이프라인은 DataOps 조직이 뛰어난 품질을 유지하면서 새로운 기능의 생산 및 배포에 대한 데이터 오케스트레이션에 숙달된 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.

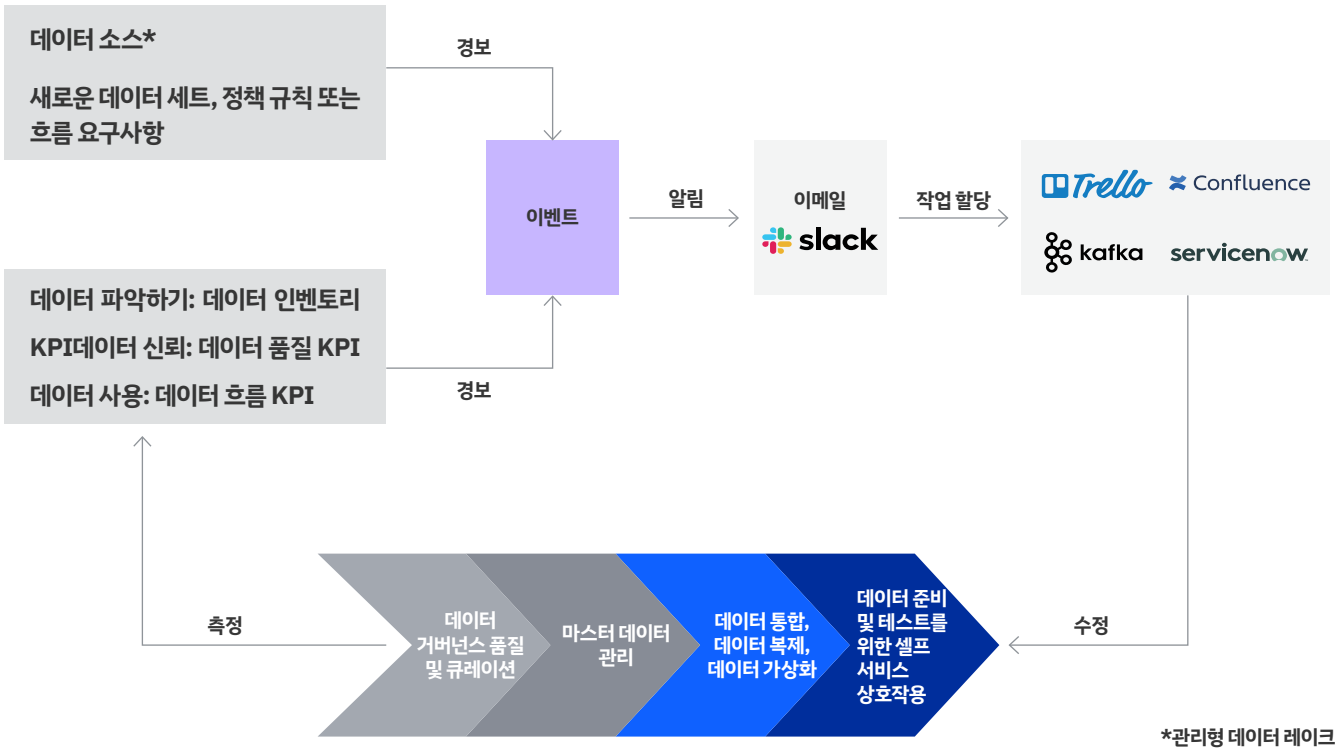


그림 4: 관리형 데이터 레이크 환경 내에서 커뮤니케이션 및 프로세스 관리 시각화

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 안에 수행할 수 있었습니다. 게다가 재고 위치에 대한 보고에 소요되는 시간을 1/6로 줄였습니다.

성공적인 DataOps 사례의 기준은 다음과 같습니다.

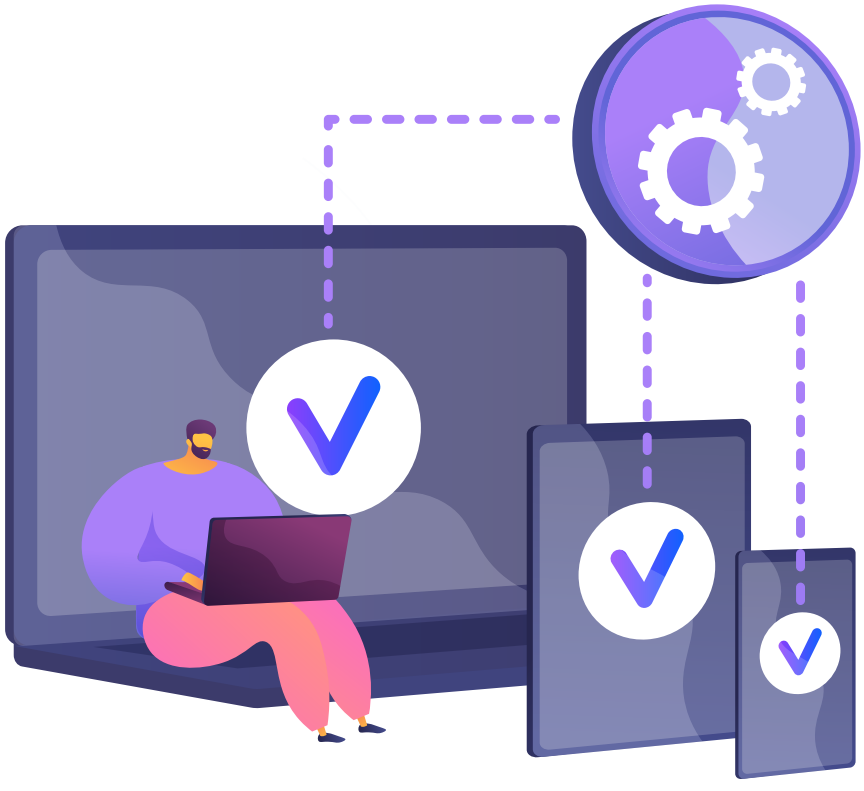
1. **데이터 오피스 설정** 이 프로세스에는 뱅크에 데이터를 리소스로 제공할 때 역할의 범위를 확실하게 정의하고, 핵심 이해관계자를 식별하고, 데이터 파이프라인의 모든 이해관계자가 협업 작업 및 문화에 대해 수행하는 노력을 파악하는 것이 포함됩니다.
2. **비즈니스 목표에 맞추기** 경쟁우위를 유지하기 위해 시장에서는 정보를 기반으로 한 데이터 중심 접근 방식을 통해서만 제공할 수 있는 새로운 기회에 신속하게 반응할 수 있어야 합니다. 간단히 말해, 비즈니스 및 데이터 전달 간에 뛰어난 커뮤니케이션이 빠져 있다면 경영진은 조직이 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 수있는 경우에만 이러한 결과를 얻을 수 있습니다.

적용

IBM Watson

머신 러닝의 혁신적인 최신 기술을 기반으로 하는
Watson은 AI 라이프사이클을 자동화합니다.

자세히 알아보기



IBM

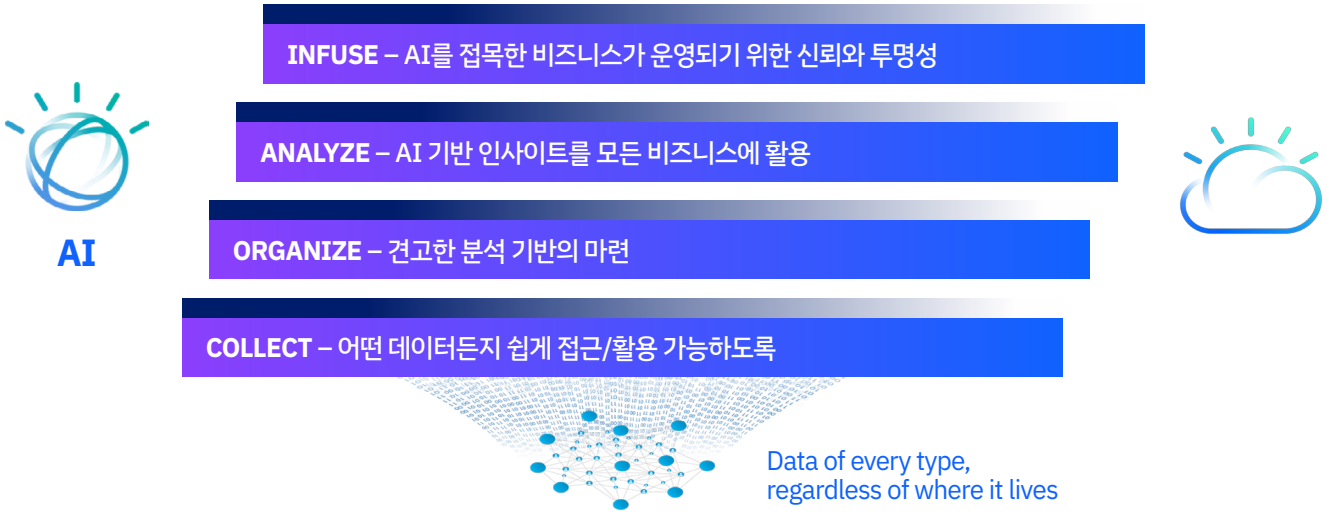
New Normal !

AI를 활용한 업무 및 일상의 Needs 가속화

콜센터 대고객 응대 서비스	*대직원 Digital Help Desk*	*업무 프로세스 심사•모니터링 자동화*	*개인 맞춤형 추천 서비스*
콜센터 확대	**재택근무 대중화 속도 증가**	**컴플라이언스 위험도**	**비대면 구매 확대**
- 온라인 소비량 늘면서 문의•클레임 함께 증가 예상 - 상담원 1명당 처리 건수 늘면서 고객 대기시간 증가, 불만 및 호 포기율 증가 예상	- 재택근무가 빠르게 운영 적용 중 - Pandemic 사태 종식 후에도 재택근무를 일정 수준 운영 할 것이라고 전망	- 비대면 금융거래 수요 늘면서 서비스 속도•접근성 향상 위해 서비스 절차 간소화 - 이 과정에서 불공정•불완전 판매 위험도 증가	- 온라인 상품 / 식품 구매 증가에 따라 시장 경쟁 심화 - 동일 상품에도 차별화된 서비스 제공 및 소비 패턴 분석에 따른 선 제안 고도화 필요

AI는 단계적 접근이 필요합니다

A prescriptive approach to accelerating the journey to AI

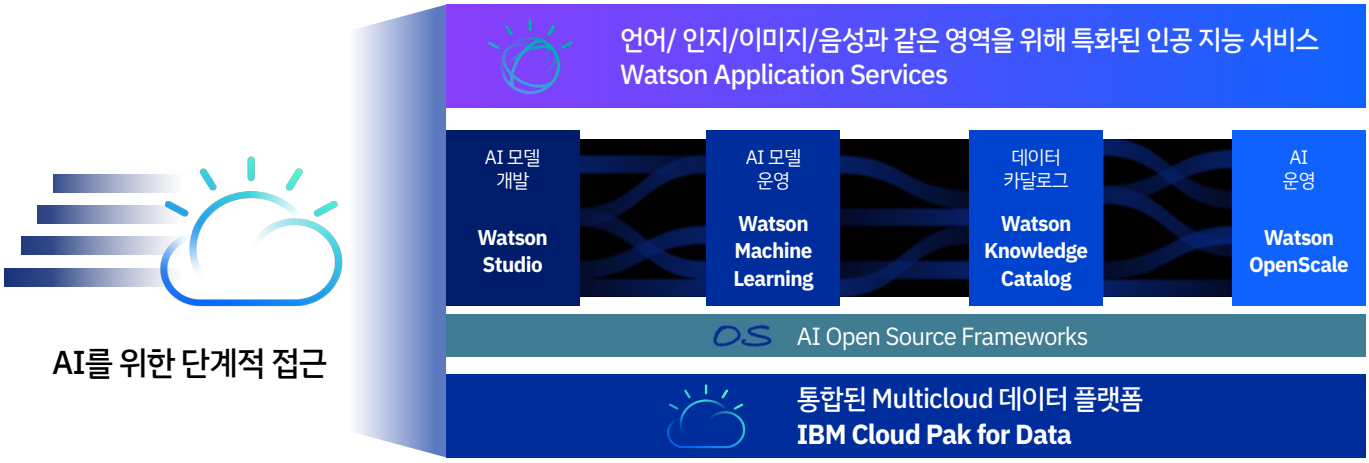


AI의 단계적 접근의 장점

1) 확장성 있는 오픈 인포메이션 아키텍처의 구축
2) Hybrid Cloud 기반의 데이터 자산 고도화
3) AI를 위한 데이터 준비
4) 오픈소스의 적극적인 활용
5) 신뢰할 수 있는 AI 모델의 비즈니스 접목

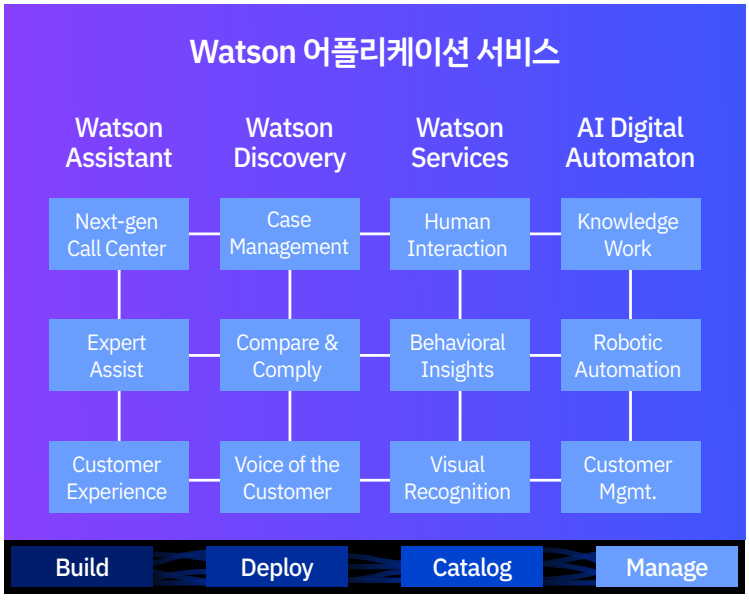
AI를 위한 다양한 기술요소를 고려한 IBM Watson

IBM AI portfolio - Everything you need for Enterprise AI, on any cloud



비즈니스에 바로 활용하는 AI 서비스, Watson

Speed time-to-value with pre-built application services



- 비즈니스 예측과 정교한 의사결정 지원
- 현업 담당자가 더 가치 있는 일을 할 수 있도록 최적화
- 의사결정, 비즈니스 프로세스의 자동화
- 새로운 비즈니스 모델, 전략의 실행

Expert Accelerators

- Industry Business Consulting Services
- Applied Data Science & AI Services
- Business Partner Ecosystem

고객 사례

IBM AI Ladder 방법론의 단계별 사례를 확인해보세요.

자세히 알아보기

현대화 사례 (Modernize)

#AI사다리 (현대화) - 데이터를 AI와 하이브리드 클라우드 환경에서 사용 가능하도록 준비

IBM Data and AI

#Data
#AI
#Modernize

런던시 자전거 공유서비스

머신러닝 알고리즘으로 자전거 이동, 고객수요 및 행동 예측

영국 런던의 최대 공유 자전거 시스템인 산탄데르 사이클 스테이션은 IBM AI 및 데이터 기술을 적용하여 운영 효율화와 개선된 대민 서비스를 제공합니다.

인도 캘커타 지역보건 센터 iKure

AI를 통한 농촌 의료 활동 보강

IBM의 AI 및 클라우드는 의사를 대체하는 것이 아니라 의사가 올바른 판단을 내리도록 돕는다는 철학을 갖고 있는 iKure의 목표에 따라 농촌 의료 활동을 보강했습니다.

글로벌 마케팅 대행사 Wunderman Thompson

데이터 및 AI로 머신 러닝 수준 향상

Wunderman Thompson는 IBM Cloud Pak for Data를 사용하여 규모에 맞는 정보를 최대한 수집하여 예측력을 높이고 고객사에 더 정확한 인사이트를 제공합니다.

덴마크 뮤직 페스티벌 Smukfest

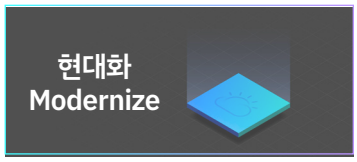
덴마크 스칸데르보르그 뮤직 페스티벌인 Smukfest는 IBM 제품을 이용해 페스티벌 데이터를 수집, 구성 및 적용하고 AI 기능을 추가하여 참석 관객에게 안전함과 편안함을 제공했습니다.

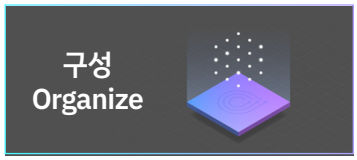
수집 사례 (Collect)

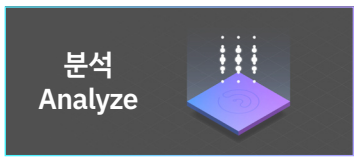
> #AI사다리 (수집) - 단순하고 액세스가 용이하도록 데이터 정제

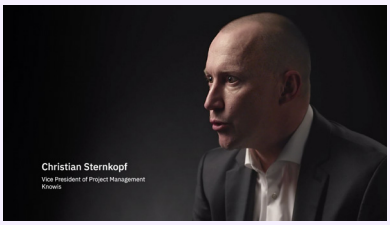
IBM Data and AI

#Data
#AI
#Collect









금융 솔루션 회사 Knowis

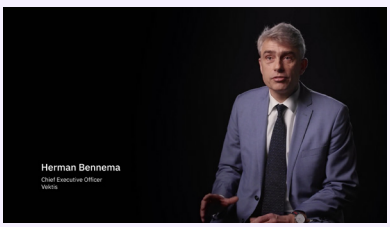
Db2 on Cloud로 소규모 은행까지 비즈니스 영역 확대

Knowis의 부사장 Christian Sternkopf는 IBM Db2가 온프레미스와 클라우드에서 배포 장벽을 넘어, 어떻게 새로운 시장 개척에 도움을 주었는지에 대해 이야기합니다.

미디어 테크 회사 RSG Media

IBM Db2 on Cloud로 미디어 플랜에 대한 혁신 시도

미디어 회사 RSG Media는 IBM Db2로 더 크고 더 복잡한 데이터 세트로 부터 심층 인사이트를 얻기 위한 자체 분석 플랫폼을 마이그레이션 했습니다.



네덜란드 의료서비스 제공업체 Vektis

데이터를 활용해 의료보험 서비스 향상

Vektis의 CEO인 Herman Bennema으로부터 IBM Db2 Warehouse on Cloud가 의료 보험 데이터에 대한 깊은 통찰력과 함께 저렴하고 수준 높은 서비스가 가능하게 한 이야기를 들어보세요.

세계 최대 유리용기 제조업체 Owens- Illinois

Oracle에서 IBM Db2로 마이그레이션하여, 총소유비용 절감

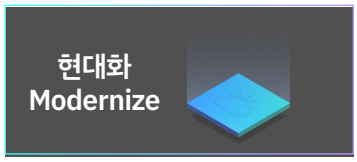
세계 최대 유리용기 제조업체 Owens-Illinois는 IBM Db2 도입을 통해 총소유비용을 절감하면서 쿼리 성능을 향상했고, 서버 설치 공간 감소 등의 부가적인 이점을 누리면서도 IT 예산을 절감하여, 전략적인 영역에 더 많은 시간과 비용을 투자할 수 있었습니다.

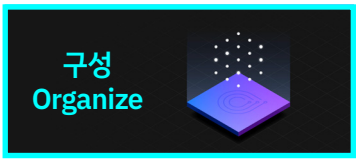
구성 사례 (Organize)

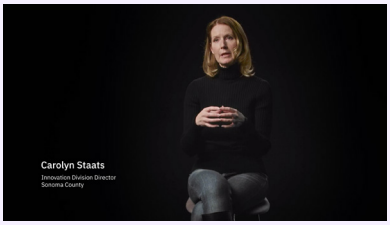
#AI사다리 (구성) - 비즈니스에 적합한 분석 기반 생성

IBM Data and AI

#Data
#AI
#Organize



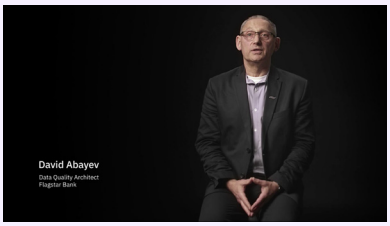




캘리포니아 소노마 카운티

IBM MDM으로 정보 장벽을 극복해 향상된 대민 서비스 실현

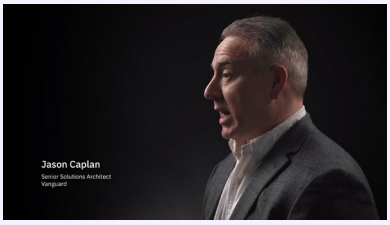
미국 캘리포니아주 소노마군(Sonoma County) 공무원들이 어떻게 주민의 민원을 360도로 파악하고 정보 장벽을 극복하여 더 나은 대민 서비스를 실현하고 있는지 확인하실 수 있습니다.



미국 플래그스타 은행

비즈니스 데이터를 위한 정보 아키텍처 현대화

Flagstar 은행의 데이터 품질 아키텍트 David Abayev가 우수한 데이터 품질 및 데이터 거버넌스에 중점을 두면서, 고객 동향을 분석하고 새로운 솔루션을 설계하기 위해 필요한 데이터 현대화에 관한 전략을 안내합니다.



글로벌 자산운용사 뱅가드

데이터 거버넌스로 데이터과학자 역량 강화 및 데이터 품질 향상

Vanguard 수석 솔루션 아키텍트인 제이슨 캐플란(Jason Caplan)이 데이터 거버넌스와 개인정보보호에 중점을 두고, 금융업계의 디지털 혁신을 향한 놀라운 여정을 이끌었던 방법을 공유합니다.

노르웨이 비료회사 YARA

세계 식량에 대한 책임과 지구 환경 보호를 위해 디지털 농업 플랫폼 구축

노르웨이 비료회사 Yara는 IBM과 협력하여 전 세계 농가에 종합적인 디지털 서비스와 농경법 자문을 제공하고, 궁극적으로 기존 농지에서 식품 생산량을 증대하여 산림 파괴를 최소화하는 플랫폼을 구축하고 있습니다.

분석 사례 (Analyze)

#AI사다리 (분석) -신뢰와 투명성을 바탕으로 AI 구축 및 확장

IBM Data and AI

#Data
#AI
#Analyze

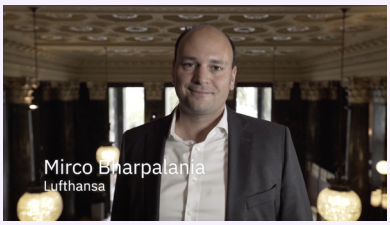
현대화
Modernize

수집
Collect

구성
Organize

분석
Analyze

적용
Infuse



Lufthansa

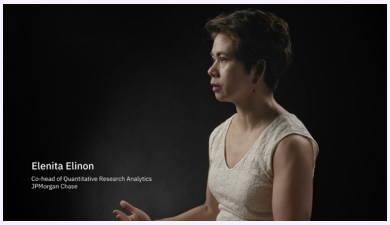
데이터 및 사이언스
능력 향상을 통해 AI 확장

루프트한자 항공사가 Watson을 활용한 AI기반 지식 시스템 구축을 통해 매출 증대 및 비용 절감을 이뤄낸 방법을 확인해보세요.

글로벌 컨설팅 회사 KPMG

Watson OpenScale을 활용해
AI 기반으로 금융권 고객들의 대출, 카드, 보험 서비스에 혁신 제공

KPMG LLP의 신기술 위험 부문이사 Kelly Combs가 AI 적용의 걸림돌이 되는 요소와 그 해결방안에 대해 설명합니다. KPMG는 Watson OpenScale을 이용해 자사 금융권 고객의 대출, 카드, 보험 서비스에 AI 기반의 혁신을 제공합니다.



미국 최대 은행 JP 모간 체이스

IBM Watson Studio로 리스크
관리 모델 개선

JPMC의 Quant 모델 위험 관리 그룹 Elenita Elinon으로부터 IBM 데이터 과학 및 AI 엘리트 팀과의 협업에 대한 소감을 듣고 Watson Studio를 활용하여 위험 관리 모델을 개선하세요.

미국 석유회사 ExxonMobil

AI기반 데이터 플랫폼 현대화를 통해
시추 작업 효율성을 극대화

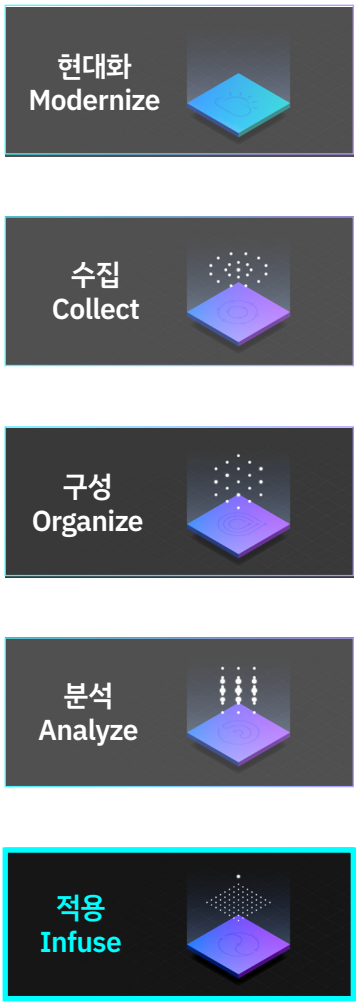
미국 ExxonMobil의 Digital Transformation 수석 고문인 Huang Xiaojun 박사가 IBM과 협업하여 회사 내 다양하고 사일로화된 데이터를 AI에 적합한 데이터로 다루어 의사결정을 내리는 시간을 단축한 방법에 대해 안내해드립니다.

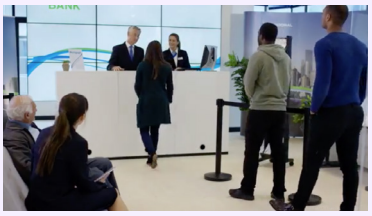
적용 사례 (Infuse)

#AI사다리 (적용) - 비즈니스 전반에서 AI 운영

IBM Data and AI

#Data
#AI
#Infuse

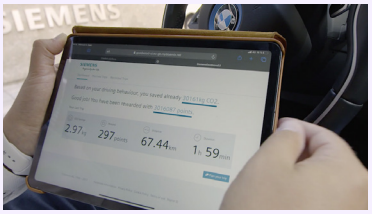




호주 최초 온라인 은행 Ubank

AI 챗봇으로 고객 경험 개선

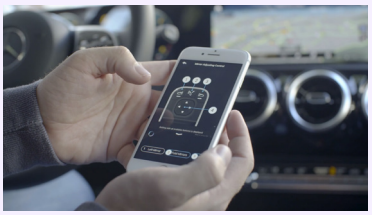
UBank는 IBM Cloud Garage를 통해 IBM Cloud 플랫폼에서 Watson Assistant로 구현 가능한 챗봇 시스템 RoboChat을 구상 및 계획했습니다. RoboChat을 사용하여 UBank는 이미 온라인 대출 신청 완료 비율을 15% 이상 높였습니다.



친환경 기업 지멘스

인공지능, 블록체인, 사물인터넷 기술을 통해 탄소 배출량 감소

이산화탄소 배출량 감소는 Siemens가 중요하게 생각하는 기업 문화입니다. 직원들은 인공지능 솔루션을 활용하여 이산화탄소가 적게 배출되는 친환경적인 여행 가이드를 만들고, 탄소 제로의 미래를 향한 길을 열었습니다.



Watson Assistant가 만든 자동차 혁명, 메르세데즈-벤츠

모바일 챗봇으로 차량 취급설명서 대체 증강현실(AR)로 차량기능검색

기계, 전기, 전자, 소프트웨어 등의 영역에서 사일로 기반으로 제품을 개발하고 있는 벤츠 자동차가 시스템이 서로 간섭하며 생긴 고도의 복잡성을 어떻게 관리하는지 확인해보세요.

건축자재 제조업체 Mueller

세계지역 기반 회사의 데이터 활용이 만드는 소도시 인재 활용

Mueller의 전략, 데이터 및 인공지능 관리자인 Mark Lack은 신뢰할 수 있는 데이터가 어떻게 지역에 토대를 둔 기업의 성장을 지원하고, 문화 유실을 방지하고, 소도시 인재 활용을 지원하는지 이야기합니다.

© **Copyright IBM Corporation 2020**

한국아이비엠주식회사
(07326) 서울시 영등포구 국제금융로10
서울국제금융센터 (Three IFC)

TEL : (02) 3781-7114
www.ibm.com/kr

2021년 03월

Printed in Korea

All Rights Reserved

이 문서는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다. IBM은 본 문서에서 검토한 제품이나 기능, 서비스를 다른 국가에서는 제공하지 않을 수도 있습니다. 본 정보는 통지 없이 변경될 수 있습니다. 귀하가 사용할 수 있는 제품, 기능 및 서비스에 대한 정보는 한국 IBM 비즈니스 담당자에게 문의하십시오.

IBM이 제시하는 방향 및 의도에 관한 모든 언급은 특별한 통지 없이 변경될 수 있습니다. 이는 SOD에 의해 명시됩니다.

IBM, IBM 로고, Active Memory, AIX, BladeCenter, EnergyScale, Power, POWER, POWER8, PowerHA, PowerVM, POWER SYSTEMS 및 POWER SYSTEMS Software는 미국 또는 기타 국가에서 사용되는 International Business Machines Corporation의 상표 또는 등록상표입니다. IBM이 소유한 미국 상표의 전체목록은 다음 사이트에 있습니다.
ibm.com/legal/copytrade.shtml