

Db2

AI
데이터베이스

목차

- 3 AI 데이터베이스:
AI 기반, AI를 위한 설계
- 4 AI 기반
- 5 데이터 가상화
- 6 적응형 워크로드 관리 및 리소스 최적화
- 9 기계 학습 쿼리 최적화
- 10 확신도 기반 쿼리
- 11 AI에 맞는 설계
- 12 자연어 쿼리로 더 빠르게 데이터 탐색
- 14 인기 있는 언어와 코드 예제를 사용하여 AI
애플리케이션 구축
- 17 그래프 및 SQL을 사용하여 복잡한 관계
모델링
- 18 네이티브로 블록체인 데이터 분석
- 19 AI 기반 조직에서 Db2의 위치

AI 데이터베이스:

AI 기반, AI를 위한 설계

데이터의 볼륨과 다양성이 증가함에 따라 업계 전반적으로 디지털 혁신을 촉진하기 위한 새로운 전문 지식과 프로세스, 인프라가 필요한 상황입니다. 컴퓨터가 사람의 지속적인 개입 없이 데이터를 처리하고 경험을 통해 학습하는 기계 학습(ML)과 인공지능(AI)은 이를 위한 핵심 요소입니다.

AI와 ML은 분석, 물류부터 자연어 상호작용과 음악 작곡, 화학 설계와 사기 탐지에 이르기까지 전통적인 컴퓨팅 작업과 비전통적인 컴퓨팅 작업에 모두 도움이 됩니다. AI와 ML은 비즈니스에 데이터 중심의 인텔리전스를 구현하므로 엔터프라이즈는 과거에 무슨 일이 일어났고 현재 무슨 일이 일어나고 있으며 앞으로 어떤 일이 일어날 가능성이 높은지를 파악할 수 있습니다.

AI를 향한 여정은 각각의 구성 요소가 하나의 단을 형성하는 사다리와 같습니다. 이 여정을 실행하는 기업은 사다리를 강화해서 유용성을 높입니다.

적용(Infuse)

비즈니스 전반에서 AI 운영

분석(Analyze)

신뢰와 투명성을 바탕으로 AI 구축 및 확장

수집(Collect)

단순하고 액세스가 용이하도록 데이터 정제

현대화(Modernize)

데이터를 AI와 하이브리드 클라우드 환경에서 사용 가능하도록 준비

구성(Organize)

비즈니스에 적합한 분석 기반 생성

AI를 향한 여정의 기반은 데이터의 형식, 출처 또는 배포 형태와 관계없이 데이터를 액세스, 수집, 구성하고 사용할 수 있도록 하는 견실한 데이터 인프라입니다. ML 중심의 데이터 관리에서 AI는 목표인 동시에 그 목표를 달성하기 위한 수단입니다.

AI와 ML이 적용된 통합 솔루션을 사용하면 더 간편하고 빠르고 지능적인 데이터 관리가 가능하고, 개발자와 데이터 과학자는 새로운 방식으로 데이터를 사용하면서 운영부터 모델 구축, 애플리케이션 개발에 이르기까지 비즈니스 전반에 AI 기능을 구축할 수 있습니다.

AI 기반

IBM® Db2®에 내장된 기계 학습과 AI 기능은 다음을 포함한 여러 혁신적인 데이터 관리 기능을 가능하게 해 줍니다.

- 전체 스택 데이터 가상화
- 자동 워크로드 관리 및 리소스 최적화
- 10배 더 높은 쿼리 성능
- 확신도 기반 쿼리 결과

데이터 가상화



정의

데이터 페더레이션과 추상화 계층의 조합. 모든 사용자와 애플리케이션에 기반 데이터의 유형, 형식, 크기 또는 위치와 관계없이 하나의 액세스 지점에서 여러 데이터 소스와 동시에 상호작용할 수 있는 기능을 제공합니다. 이와 같은 액세스의 확대는 데이터 사일로를 허물고 더 빠르고 효율적인 데이터 사용을 통해 AI 도입을 확장하는 데 도움이 됩니다.

중요한 이유

데이터에 대한 하나의 진입 지점은 다음과 같은 이점을 자랑합니다.

- **데이터 전문가를 위한 간편한 데이터 액세스:** 모든 데이터에 대한 하나의 시야를 확보하면 데이터를 옮기는 데 시간과 리소스를 소비하지 않고도 구조적 또는 비구조적, 관계형 또는 NoSQL 등 다양한 데이터 소스를 포괄하는 데이터 검색과 액세스가 가능합니다. 개발자와 데이터 엔지니어는 과거와 현재의 데이터를 불문하고 가능한 모든 형식의 데이터에 액세스할 수 있으며 데이터의 예상치 못한 조합을 통해 인사이트를 이끌어낼 수 있습니다.
- **거버넌스와 보안을 위한 단일 제어 지점:** 관리자가 모든 사용자와 애플리케이션을 모든 데이터 저장소에 대한 단일 액세스 지점으로 간단히 보낼 수 있다면 보안과 거버넌스 작업이 훨씬 더 간편하고 견고해지며 오류 발생 가능성이 낮아집니다.
- **데이터 전송 감소:** 쿼리된 데이터를 로컬 저장소로 먼저 보내지 않고도 저장소 자체에서 데이터를 처리할 수 있습니다. 이렇게 하면 지연과 대역폭 비용을 크게 줄일 수 있습니다.

작동 방식

현대 엔터프라이즈 데이터 환경은 여러 데이터 저장소에 의존합니다. 여기에는 관계형 데이터 저장소의 트랜잭션 데이터, Hadoop 클러스터의 감성 데이터, 데이터 웨어하우스의 과거 고객 데이터, 고속 데이터 클러스터의 클릭스트림 전자상거래 데이터가 포함됩니다.

가상화가 없으면 데이터베이스 관리자(DBA)부터 앱 개발자와 데이터 과학자에 이르기 까지, 이 데이터를 다루고자 하는 모든 사람은 각각의 데이터 저장소를 위한 맞춤형 액세스 방법을 구현해야 합니다. 최선의 경우라고 해도 관여하는 모든 사람에게 큰 골칫거리가 될 뿐입니다. 이 경우 사용자들이 다양한 형식의 데이터를 서로 비교하여 얻을 수 있는 인사이트가 제한됩니다(예를 들어 고객의 과거 이력과 실시간 탐색 행동 비교). 최악의 경우 데이터 손상과 새로운 보안 침해가 발생할 위험이 있습니다.

다양한 데이터에 대해 기업 전체에서 하나의 액세스 지점을 제공한다는 점은 데이터 페더레이션이 높은 인기를 얻게 된 이유입니다. 그러나 IBM 데이터 가상화는 페더레이션에서 한 걸음 더 나아가, 추상화 계층을 제공하고 사용자가 데이터 액세스를 위한 스스로의 용어와 방법론을 정의하도록 허용합니다(그림 1 참조).

적용 사례

데이터 가상화

유럽 은행 ING는 전 세계 모든 사용자를 위한 단일 데이터 액세스 지점을 구현하기 위해 IBM과 협력하고 있습니다. 이를 통해 ING는 전 세계 데이터 인프라를 관리하면서 필요한 성능과 확장성, 보안, 거버넌스를 제공할 수 있습니다. 또한 다양한 비즈니스 영역 간의 단절도 줄입니다.

IBM 솔루션을 통해 세계 어디서나 새로운 데이터를 ING 플랫폼에 추가하고, 개별 데이터 액세스 체계나 보안 권한을 수정할 필요 없이 전 세계 은행 사용자 누구나 이 데이터에 액세스할 수 있습니다.

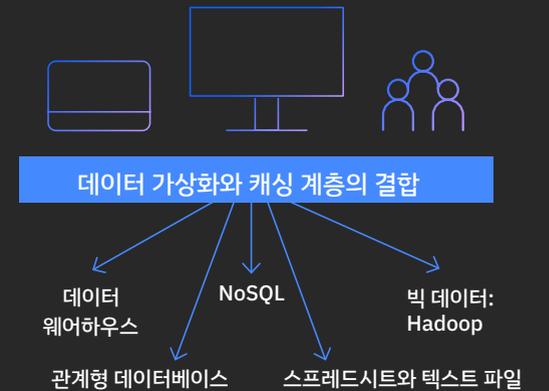


그림 1. 가상화를 통해 여러 데이터 소스의 단일화된 액세스 및 거버넌스가 어떻게 가능한지를 보여 주는 그림

적응형 워크로드 관리 및 리소스 최적화

정의

다양한 워크로드를 처리하기 위한 데이터 리소스를 자동으로 할당하는 인텔리전트 기술. 이 기술은 자동화된 적응형 기술로, 과거 조건에 대응하며 미래 수요도 예측합니다.

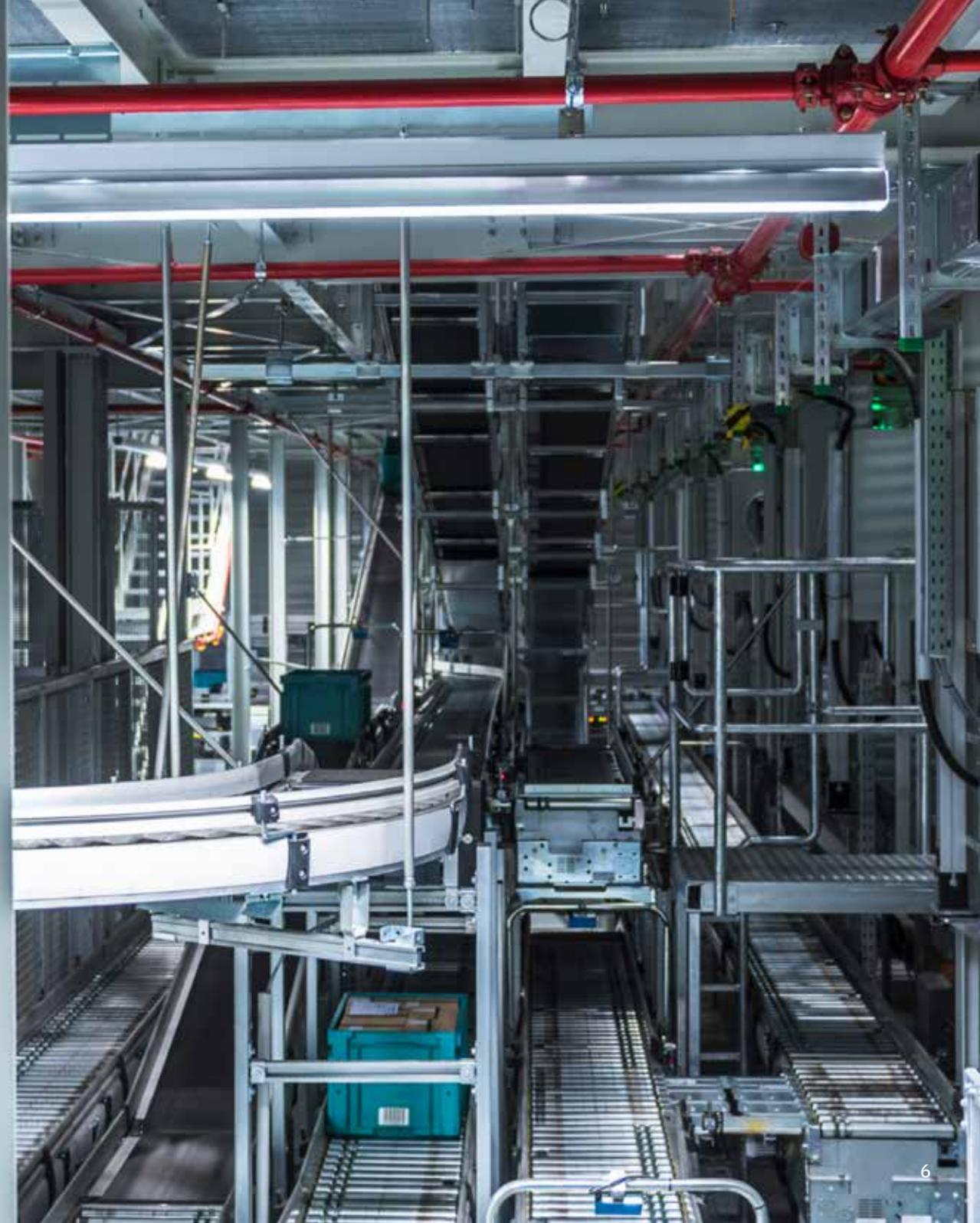
중요한 이유

적응형 워크로드 관리를 사용하면 많은 노동이 투입되는 조정 및 설정이 불필요한 안정적인 고성능 시스템을 기본적으로 확보할 수 있습니다. 현대 엔터프라이즈 데이터 플랫폼의 복잡함과 수동 워크로드 관리 전략의 높은 장애 비율을 감안하면 이는 필수적인 요소입니다.

너무 많은 워크로드가 동시에 도달하면 데이터베이스는 추가적인 수요를 처리할 방법을 결정해야만 합니다.

이러한 "개방형" 또는 수동 스키마에는 피드백 메커니즘이 없습니다. 즉 워크로드의 수나 크기에 대해 사전에 정해진 제한을 사용자가 설정해야 합니다. 그러나 이러한 제한은 복잡한 워크로드 집합에 직면할 경우 제대로 작동하지 않을 수 있으며 최선의 환경이라 해도 지속적인 수동 모니터링과 조정이 필요합니다. 이는 데이터베이스 성능 문제, 사용률 저하 또는 장애로 이어집니다.

반면 Db2는 사용률 추세를 감지 및 예측하므로 이 추세가 심각한 문제로 발전하기 전에 사용자에게 알려거나 자동으로 교정할 수 있습니다. 구성과 튜닝에 필요한 작업 시간이 단축되므로 DBA의 업무가 한결 쉬워지고 각 데이터베이스의 소유 비용이 절감됩니다. IBM 테스트 결과 이와 같이 사용률이 개선되면 전체적인 데이터베이스 성능이 최대 30% 향상되는 것으로 나타났습니다.





작동 방식

적응형 워크로드 관리는 피드백 메커니즘 형태로 기계 학습을 수용하므로 데이터베이스는 다양한 워크로드의 예상 런타임과 실제 런타임을 지속적으로 모니터링하면서 가용 리소스를 조정하여 부족분을 보충합니다. Db2 데이터베이스는 기본 상태 그대로 실행 가능하며 대부분의 워크로드에서 최소한의 설정만 하면 됩니다.

특히 복잡한 상황에서 수동 설정이 필요하더라도 이 적응형 기술을 통해 쉽게 완료할 수 있습니다. DBA는 여러 워크로드 클래스를 만들고 각각의 성능 목표를 지정할 수 있습니다. 그러면 적응형 관리자가 수신되는 워크로드를 모니터링하고 목표 달성을 위한 리소스를 할당합니다.

적용 사례

자동화된 워크로드 관리

자동화된 적응형 워크로드 관리는 엔터프라이즈 데이터베이스 관리에 수반되는 여러 공통적인 골칫거리를 해결합니다.

예를 들어 데이터 웨어하우스 환경에서 데이터베이스는 스트리밍 흡수부터 응답 시간이 1초 미만인 저지연 보고서(예를 들어 웹 페이지 로드), 그리고 완료를 위해 몇 분이 걸리는 무거운 일괄 보고서에 이르기까지 다양한 작업을 수신할 수 있습니다. 관리자는 시스템이 큰 작업에 막혀 긴급한 작업을 완료하지 못하는 상황을 방지해야 합니다.

적응형 워크로드 관리자는 이 작업을 매우 쉽게 해 주므로 관리자는 리소스 그룹(워크로드 또는 서비스 클래스)을 만들고 각 그룹에 시스템의 일정 부분을 할당할 수 있습니다. 그러면 데이터베이스는 지능적으로 리소스를 할당하여 성능 목표를 달성할 수 있습니다. 이러한 모든 과정은 개별 워크로드와 리소스를 수동으로 모니터링 및 튜닝할 필요 없이 실행됩니다.

AI는 목표이자
이 목표를
달성하기 위한
수단이 될 수
있습니다.



기계 학습 쿼리 최적화

정의

Db2 Machine Learning Optimizer는 자율 기계 학습을 사용하여 일반적인 비용 기반 쿼리 최적화에서 더 개선된 쿼리 최적화 전략을 구현하는 부가적인 지능적 최적화 수준을 선사합니다.

중요한 이유

기본적인 워크로드 비용 최적화 도구는 특정 쿼리에 대한 실행 전략을 제안할 수 있지만 데이터베이스의 최근 변경을 인식하지 못하고 경험을 통해 학습할 수도 없습니다. 가장 빠른 실행 전략을 추천하지만 그 전략이 예상한 대로 작동하지 않더라도 계속 같은 전략을 추천합니다.

반면 Db2 Machine Learning Optimizer는 실제 쿼리 성능으로부터 피드백을 받아 실제 데이터 인프라를 사용한 작업에서 최선의 결과를 제공하는 실행 전략을 추천하고, 각각의 실행에서 쿼리 경로를 조정합니다.

기계 학습 최적화 도구를 사용하면 DBA는 시스템 성능을 모니터링하고 쿼리를 최적화하는 데 시간을 소비할 필요가 없습니다. 그 대신 AI 애플리케이션 구현, 데이터 사용 전략 수립, 더 효과적인 데이터 활용을 위해 조직 전반의 비즈니스 사용자를 돕는 것과 같은, 조직에 더 많은 가치를 제공하는 활동에 집중할 수 있습니다.

작동 방식

Db2 Machine Learning Optimizer는 신경망 패턴을 모방하여 쿼리 경로를 최적화하며, 이를 통해 일부 쿼리는 8~10배 더 빠르게 완료됩니다(IBM 내부 테스트 결과).



적용 사례 ML 쿼리 최적화

일반적인 비용 최적화 도구는 통계 및 리소스 모델링을 사용해서 주어진 쿼리의 실행 전략을 평가합니다. 이러한 도구는 아래 그림 2의 첫 번째 옵션을 반환합니다. 즉, 테이블 두 쌍을 조인한 다음 결과를 조인해서 반환합니다. 그러나 기계 학습 알고리즘은 경험을 통한 학습으로 더 나은 실행 전략이 있음을 인식합니다. 즉, 테이블 한 쌍을 조인하고, 그 결과에 세 번째 테이블을 조인한 다음 그 결과에 네 번째 테이블을 조인해 같은 내용을 반환하는 것입니다.

기계 학습 미사용



기계 학습 사용



그림 2. 비용 최적화(위)와 기계 학습 최적화(아래) 쿼리 실행 전략의 예

이러한 실행 전략의 변화는 실제 쿼리 실행에 대응해서 실시간으로 적용할 수 있습니다. 사기 탐지와 같이 시간에 민감한 애플리케이션에서 이 실시간 개선은 중요한 역할을 할 수 있습니다.

확신도 기반 쿼리

정의

가능한 예-아니오 방식의 답 대신 가능성 또는 "가장 일치하는 항목" 관점에서 SQL 쿼리 결과를 제공하는 고급 기능.

중요한 이유

확신도 기반 일치는 기계 학습 환경에서 예전부터 제공되었지만 Db2 기능은 그 적용 범위를 SQL 식까지 확장합니다. SQL 엔지니어에게 이 기능은 데이터 과학자 없이 수행할 수 있는 데이터 작업의 범위를 비약적으로 넓혀 줍니다. 따라서 비즈니스에서 SQL 엔지니어의 가치가 더 높아지고 이미 과도한 업무에 시달리는 데이터 과학자가 짊어지는 부담을 덜 수 있습니다.

작동 방식

확신도 기반 쿼리는 심층 피드 포워드 신경망을 구현함으로써 SQL에 기계 학습 확장을 추가합니다. 자율 심층 학습의 이 변형 버전은 "유사성"과 일치 가능성이 모두 높은 곳에서 활용됩니다. 확신도 기반 쿼리는 다음 분야에 적용할 수 있습니다.

- 유사성/차이점 쿼리
- 귀납적 추론 쿼리(예를 들어 의미 군집화, 유추, 예외적 사례 배제)
- 의미론적 그룹화 작업
- 패턴 이상(예: 사기 탐지)
- 이미지, 오디오, 비디오

적용사례

확신도 기반 쿼리

경찰 데이터베이스의 경우 신장, 체중, 연령, 피부색, 눈 색, 특징적인 표지를 비롯한 변수가 많은 데다, 목격자 진술을 완전히 신뢰할 수 없으므로 이러한 변수 모두 사실상 불확실합니다. 일반적인 데이터베이스 검색은 이 불확실성을 감안하기 위해 주어진 목격자 진술과 일치하는 값 "범위"를 수동으로 생성하는 매우 복잡한 SQL 쿼리를 사용합니다. 예를 들어 신고된 체중이 195lb라면 실제 체중이 185~205lb 사이인 용의자와 일치할 수 있습니다. 이러한 범위는 사용하기 불편하고, 목격자가 진술한 값이 범위에서 지나치게 벗어나는 경우 용의자를 잘못 배제하는 상황으로 이어질 수 있습니다. 반면 목격자 진술을 용의자 프로필과 대조하는 경찰관은 확신도 기반 쿼리를 사용하여 개별 데이터 포인트가 수용 가능한 일치 범위를 벗어나는 경우에도 전체적인 목격자 프로필과 가장 일치하는 용의자를 찾는, 간단한 확률적 SQL 문을 생성할 수 있습니다.

AI에 맞는 설계

또한 Db2는 AI에 맞게 설계되었습니다. 즉, 데이터 전문가는 강력한 AI 기반 워크로드 업그레이드와 워크플로 개선을 더욱 쉽게 실현할 수 있습니다. 여기에는 다음이 포함됩니다.

- 자연어 쿼리로 더 빠르게 데이터 탐색
- 인기 있는 언어와 코드 예제를 사용하여 AI 애플리케이션 구축
- 그래프 및 SQL을 사용하여 복잡한 관계 모델링
- 네이티브로 블록체인 데이터 분석

자연어 쿼리로 더 빠르게 데이터 탐색

정의

Augmented Data Explorer는 복잡한 비즈니스 데이터 탐색을 위한 "Google과 비슷한" 웹 기반 인터페이스입니다. 사용자를 위한 자연어 검색, 개발자를 위한 REST API를 제공합니다. 자연어로 질문을 하면 모든 가용 데이터 소스를 활용한 데이터 시각화와 요약이 생성됩니다.

중요한 이유

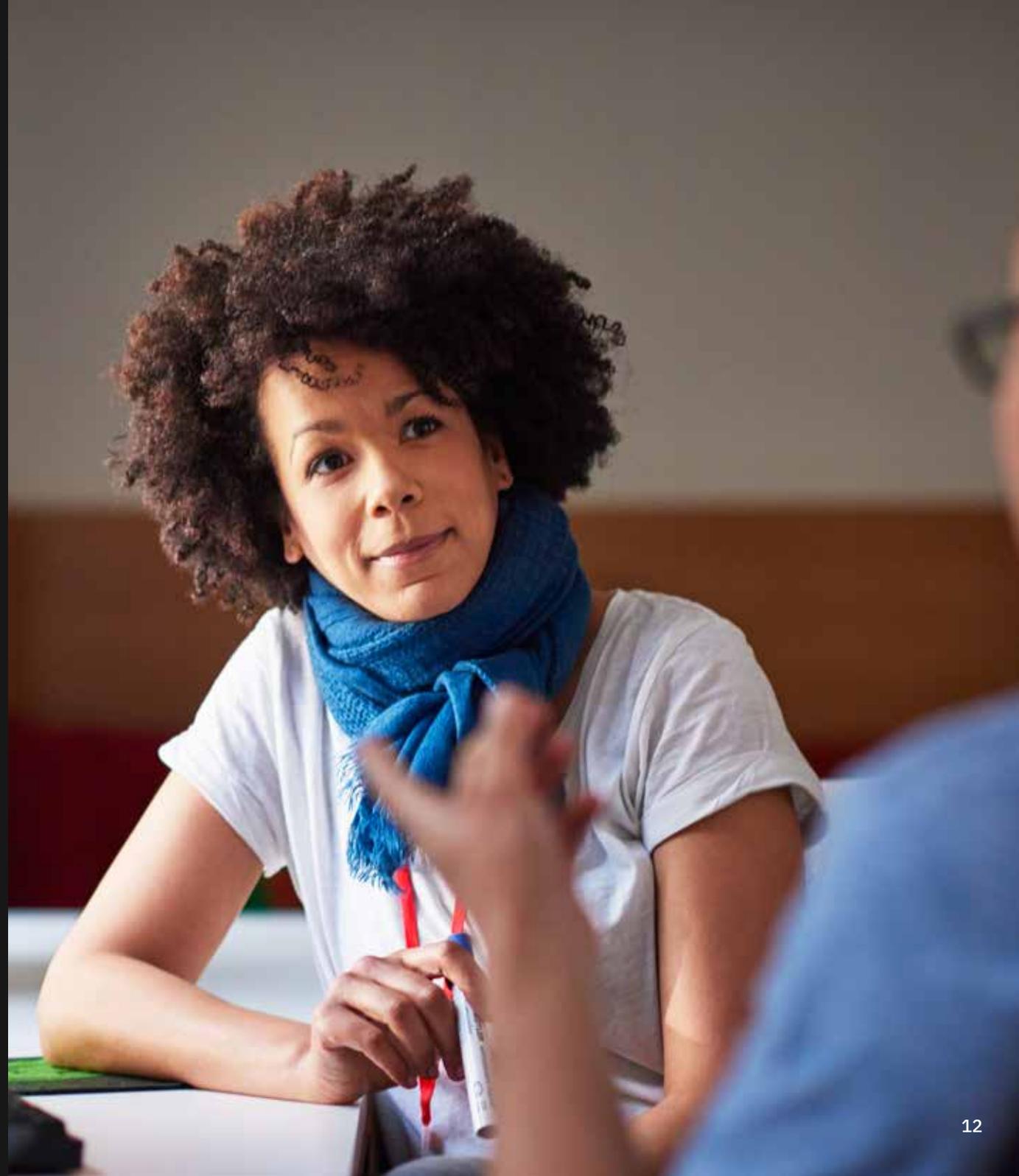
데이터 환경이 갈수록 복잡해지면서 많은 사용자들이 방대한 규모의 데이터 집합에 직면하고 있습니다. SQL 쿼리 또는 분석 앱을 개발하기 위한 기술 인력 공급은 부족하지만 데이터에서 빠르게 답을 구하고자 하는 사람들은 더 많아졌습니다.

Augmented Data Explorer는 초기 단계의 데이터 탐색 도구입니다. 이 도구는 기술적 지식이 없는 사용자에게 정교한 데이터 쿼리의 문을 열어 줍니다. "2019년 판매량" 또는 "10월에 비가 오는 빈도"와 같은 자연어 쿼리를 입력하기만 하면 됩니다.

이렇게 하면 모든 관련 데이터를 쿼리해서 원하는 결과를 자연어 형식, 동적인 데이터 시각화 또는 두 가지 모두로 반환하는 SQL 문이 생성됩니다.

과거에는 비즈니스 사용자가 이러한 쿼리를 하려면 일반적으로 자연어 요청을 필요한 SQL 쿼리로 변환하기 위한 수작업, 그리고 여기서 필연적으로 발생하는 문제에 도움을 줄 전문가의 조언이 필요했습니다. Augmented Data Explorer는 두 가지 요구 사항을 모두 해소하고, 이미 보유한 데이터에서 새로운 가치를 찾기 위한 새로운 길을 제시하도록 설계되었습니다.

데이터 과학자와 기술에 정통한 사용자는 Augmented Data Explorer를 사용하여 익숙하지 않거나 용량이 큰 데이터 집합에 초기 쿼리를 수행해 데이터 집합에 어떤 내용이 포함되어 있는지 감을 잡은 후 본격적으로 조사를 시작할 수 있습니다. Augmented Data Explorer는 REST API도 제공하므로 개발자가 솔루션 코딩에 시간을 낭비할 필요 없이 자체 애플리케이션에 자연어 검색 및 예측 기능을 손쉽게 내장할 수 있습니다.



작동 방식

Augmented Data Explorer는 ML 기술을 사용하여 데이터를 크롤링하고 색인화해서 숨겨진 인사이트를 이끌어 냅니다. 이 도구의 AI 기능은 자연어 지원(다양한 영역과 분야의 문맥, 유의어 및 구문 이해)과 검색어를 기반으로 한 관련 관심 주제 예측에 중점을 둡니다. 이 도구는 손쉬운 배포와 관리를 위해 컨테이너화되며, 사용이 용이한 웹 기반 인터페이스를 갖추고 있습니다.

또한 그래프 기능이 포함되어 있어 탐색기 엔진이 실제로 액세스하고 크롤링하는 데이터 집합을 일목요연하게 보여 줍니다. 사용자는 시스템이 제공한 답을 산출하기 위해 어떤 데이터(예를 들어 스키마와 테이블)가 사용되었는지를 손쉽게 볼 수 있습니다.



Augmented Data Explorer의 가치

Augmented Data Explorer는 데이터 사용자가 어떤 질문을 해야 할지 확실히 모르는 상황에서 가장 유용합니다. 이 도구를 통해 사용자는 질문을 제기하고 데이터 내에서 관계를 발견하고 실행 가능한 인사이트를 구할 수 있습니다. 또한 이와 같은 비즈니스 과제를 해결할 수 있는 인력 풀이 확장되는 효과도 있습니다. 사용자에게 많은 책임이 주어지며 지속적으로 변화하는 비즈니스 환경에서 Augmented Data Explorer는 막대한 양의 데이터를 다루고 이 데이터에서 새로운 가치를 창출하기 위한 첫 번째 단계입니다.

소매 기업의 영업 관리자를 예로 들어 보겠습니다. 이 영업 관리자에게는 지역별 판매량 분석에서 고객 유지율 파악, 구매 포기 유발 원인 찾기, 단골 고객이 방문을 하지 않는 이유 예측, 날씨가 재고 배송 리드 시간에 미치는 영향에 이르기까지 매일 또는 매월 물어야 하는 여러 질문이 있을 것입니다.

Augmented Data Explorer는 새로운 질문을 식별하고 신속하게 답변하도록 설계되었으므로 AI 기능을 훨씬 더 광범위한 비즈니스 영역으로 확장할 수 있습니다.

인기 있는 언어와 코드 예제를 사용하여 AI 애플리케이션 구축

정의

Db2 11.5는 Python, JSON, GO, Ruby, PHP, Java, Node.js, Sequelize, Jupyter Notebook을 포함한 인기 있는 언어 및 라이브러리를 네이티브로 지원합니다. 데이터 전문가는 이미 알고 있는 언어로 AI 애플리케이션을 구축한 다음 이를 Db2의 데이터에 원활하게 연결할 수 있습니다.

중요한 이유

IBM은 인기 있는 기계 학습 언어에 대한 네이티브 지원을 통해 ML 및 AI 중심 데이터 생태계로 데이터 관리와 빠른 앱 개발 간의 간극을 해소합니다. 개발자와 데이터 과학자는 엔터프라이즈 데이터베이스에 데이터를 그대로 두면서 기존의 인기 있는 개발 언어와 라이브러리를 사용할 수 있습니다.

네이티브 언어와 라이브러리 지원은 더 빠르고 쉬운 AI 애플리케이션 개발로 이어지는 여러 이점을 자랑합니다.

- 개발자는 인기 있는 언어 및 라이브러리와 함께 GitHub에 있는 산업별 코드 예제를 활용하여 적용 사례에 맞는 앱을 신속하게 구축할 수 있습니다.
- 개발자는 Db2 데이터를 네이티브로 사용하는 애플리케이션을 손쉽게 작성할 수 있습니다. Db2 데이터에 액세스하기 위한 맞춤형 백 엔드 코드가 불필요하며, 앱을 위한 하나의 공통 실패 또는 관리 지점도 필요 없습니다.
- 필요한 경우 새로운 데이터베이스를 클라우드에 신속하게 배포 및 확장한 다음 원활하게 앱에 연결할 수 있습니다.
- 개발자는 동일한 데이터베이스 생태계 내에서 기계 학습 개발 환경(IBM Watson® Studio)에 액세스할 수 있습니다.
- 데이터 과학자는 데이터베이스 액세스 문제 해결이 아닌 기계 학습 모델을 사용한 작업에 시간을 투자할 수 있습니다.
- 기업은 단순히 데이터에 액세스하기 위해 새로운 기술이나 언어를 배울 필요 없이 데이터를 분석하는 데 필요한 기술을 갖춘 인력을 채용할 수 있습니다.

작동 방식

AI 애플리케이션 개발 프로세스에는 다음과 같은 여러 중요한 질문이 포함됩니다.

- 관련된 데이터가 어디에 저장되어 있는가?
- 이 데이터를 어떻게 액세스하고 탐색하는가?
- 앱을 구축하면서 어떤 방법으로 기계 학습을 적용하는가?
- 어떤 언어를 사용할 수 있는가?

IBM 생태계 내에는 엔터프라이즈 수준에서 이러한 모든 질문에 대한 답이 있습니다. 클라우드나 온프레미스의 Db2 또는 Db2 Warehouse, Event Store 또는 Hadoop에 데이터를 저장할 수 있습니다. 거의 모든 데이터베이스 조합이 가능합니다. 모든 데이터베이스의 공통점은 완전한 지원과 안정성, 성능으로 뒷받침되는 엔터프라이즈급이라는 것입니다. 공통 Db2 코드 베이스와 기반 공통 SQL Engine은 애플리케이션이 모든 생태계 구성요소의 데이터를 쿼리할 수 있으며, 생태계의 다른 부분에서 실행되기 위해 다시 작성할 필요가 없음을 의미합니다. 또한 주요 오픈 소스 프로그래밍 언어와 프레임워크를 위한 일련의 드라이버가 새로 제공되므로 이제 개발자는 Db2를 사용하여 손쉽게 기계 학습 모델을 분석하고 애플리케이션 내에 구축할 수 있습니다.

데이터 전문가는 Augmented Data Explorer를 비롯해 이 전자책에 설명된 다른 기능을 사용하여 데이터에 포함된 내용을 신속하게 이해 가능합니다. Db2에 저장된 데이터는 Watson Studio의 기계 학습 개발 환경에 직접 연결됩니다. 업종별 라이브러리, 코드 예제 및 템플릿은 개발자가 기능적인 앱을 신속하게 만들 수 있게 해 줍니다. 마지막으로, 네이티브 드라이버 집합을 통해 SQL 문을 ML 언어별 모델(예를 들어 Python)로 변환하여 이후 원하는 언어에서 추가 개발이 가능합니다.

네이티브 언어 지원의 가치

AI 지원 애플리케이션의 잠재 시장은 막대합니다. 몇 가지 예만 들어 보면 다음과 같습니다.

소매: 추천 엔진, 대상 지정 콘텐츠, 프로모션

보험: 사기 분석, 신청 처리

물류: 운송 선단 관리, 예측 유지보수

교통: 교통량, 기상 등을 기반으로 실시간 경로 안내

시장의 많은 개발자 및 데이터 과학자들이 일반적으로 Python, Java 또는 그 외의 인기 있는 언어를 사용하여 스마트 앱과 알고리즘을 개발하고 있습니다. 이러한 앱에 공급되는 데이터는 하나 이상의 데이터베이스에 위치합니다. Python에 네이티브로 연결되는 데이터베이스가 있지만 일반적으로 이러한 데이터베이스는 엔터프라이즈 규모, 성능 또는 안정성을 갖추지 못했습니다. 그러나 선택한 ML 개발 언어와 기반 엔터프라이즈 데이터베이스 간의 네이티브 통합이 지원되지 않으면 개발자가 애플리케이션 자체를 필요한 데이터에 연결하기가 어렵고 불편합니다.

Db2와 널리 사용되는 개발 언어 간의 네이티브 통합을 통해 개발자는 DBA 또는 SQL 전문가의 개입 없이 데이터에 액세스할 수 있습니다. 그 결과 개발자와 데이터 과학자의 생산성이 높아지고 앱은 더 견고하고 빠르게 개발됩니다.

AI는 사용자가
데이터를
사용하여 할 수
있는 일을
과거에 비해
훨씬 더
늘려줍니다.



그래프 및 SQL을 사용하여 복잡한 관계 모델링

정의

Db2는 그래프 기능을 관계형 데이터 및 SQL과 심층적으로 통합하므로 그래프 애플리케이션을 관계형 데이터에서 바로 실행할 수 있으며, Db2 SQL 엔진은 그래프 데이터를 직접 쿼리할 수 있습니다.

중요한 이유

그래프는 중요한 인사이트를 제공하는 강력한 도구지만 이전에는 트랜잭션 및 OLTP 시스템과 호환되지 않았습니다. 조직에서 이 비호환성 문제를 피하기 위해서는 원래의 관계형 인프라를 유지하면서 특정 데이터를 가져와 그래프 데이터베이스에 넣고 거기서 그래프 앱을 실행해야 합니다. 이러한 종류의 중복, 지연, 오버헤드는 더 이상 용납되지 않습니다. 조직은 관계형데이터에서 초 단위로 그래프 기반의 인사이트를 얻을 수 있어야 합니다.

많은 업계의 사용자들은 기존 분석 인사이트 툴킷에 그래프 인사이트를 추가하고자 합니다. 단 그 과정에서 데이터 인프라에 중단을 초래하거나 오로지 그래프를 지원하기 위한 새 데이터베이스를 추가할 필요는 없어야 합니다. Db2로 이러한 조건을 충족 가능합니다.

작동 방식

그래프 데이터는 Db2 또는 Db2 Event Store의 관계형 프레임워크의 테이블에 저장됩니다. 이는 SQL 엔진이 그래프 데이터를 직접 쿼리할 수 있고 그래프 애플리케이션은 관계형 테이블의 데이터를 직접 쿼리할 수 있음을 의미합니다. 이 아키텍처는 Gremlin 및 Tinkerpop과 같은 오픈 소스 그래프 쿼리 언어도 지원합니다.

관계형 데이터와 Db2 그래프 기능 간의 심층적인 통합은 다음과 같은 여러 중요한 이점을 선사합니다.

- 그래프를 기존 관계형 데이터에 대해 직접 실행할 수 있으며 이는 SQL 기반 분석 위에 새로운 수준의 인사이트를 더합니다.
- 관계형 테이블에 저장된 그래프 데이터를 대상으로 직접 SQL 분석을 실행할 수 있습니다.
- 그래프 애플리케이션에서 맞춤형 API 또는 Spark를 사용하여 Db2에 연결할 수 있습니다.
- ACID 트랜잭션은 기존 관계형 애플리케이션을 중단시키지 않으면서 실시간으로 그래프를 업데이트합니다. 이는 분석 외에 실시간 트랜잭션 처리 용도로 그래프를 사용할 수 있음을 의미합니다.
- 의료 보건 및 금융과 같이 그래프를 많이 사용하는 업종에 업종별 솔루션을 사용할 수 있습니다.



네이티브 그래프 지원의 가치

그래프 기반 인사이트를 통해 관계형 데이터로 새로운 효율성과 워크플로를 실현할 수 있는 사례는 많습니다. 예를 들어 소매점은 실시간 그래프 관계를 사용하여 여러 위치의 특정 항목에 대한 반품 패턴을 파악할 수 있습니다. 그러면 이 인사이트로 소매점의 트랜잭션 데이터베이스를 즉시 업데이트해서 당일엔 진열대에서 해당 항목을 뺄 수 있게 됩니다.

다른 분야로, 보험 사기 조사관은 그래프를 사용해서 보험금 청구의 수와 규모, 청구인 간의 관계, 서비스 공급자의 신원과 같은 여러 데이터 집합을 결합할 수 있습니다. 그래프에서 드러난 관계는 사기 가능성의 지표가 될 수 있습니다.

네이티브로 블록체인 데이터 분석

정의

Db2의 네이티브 Blockchain Connector API는 블록체인 원장에 저장된 고도로 압축된 데이터에 대한 전례 없는 투명성과 인사이트를 자랑합니다.

중요한 이유

블록체인 기술은 많은 산업에서 빠른 속도로 도입되고 있지만 이로 인해 분석 병목 현상이 발생하고 있습니다. 각 블록체인 원장에는 방대한 양의 잠재적 가치를 지닌 정보가 압축된 형식으로 포함되어 있지만, 그 동안에는 이 정보를 보고 분석할 간단한 방법이 없었습니다. Db2 Blockchain Connector를 사용하면 가능합니다. 작고 투명하고 제어 가능하며 변경 불가능한 이러한 블록체인 원장을 이제 엔터프라이즈의 다른 데이터 소스와 마찬가지로 분석할 수 있습니다. 기업은 블록체인만을 위한 특별 보고 솔루션을 개발할 필요 없이 Db2 엔진의 모든 이점, 즉 성능, 유연성, 확장성, 보안을 얻을 수 있습니다.

블록체인 커넥터는 네이티브 연결 및 Db2에서 사용 가능한 다른 데이터 형식과의 상호운용성으로 인해 그 가치가 더욱 높습니다. 기업은 다른 데이터 저장소의 추가 데이터를 손쉽게 통합하여 맥락을 파악하고 블록체인 데이터에 대한 새로운 분석 가능성을 열 수 있습니다. 마지막으로, 이 커넥터는 AI 앱에 블록체인 데이터를 사용할 수 있는 가능성을 선사합니다. 이전에는 이 적용사례가 매우 번거로웠습니다. 블록체인 데이터가 사실상 자체 사일로에 갇혀 있었기 때문입니다. AI 개발자는 이제 블록체인 데이터 집합을 손쉽게 수용하여 앱을 위한 주 데이터 소스로 사용하거나 부가적인 세부 정보를 제공하는 용도로 사용할 수 있습니다.

작동 방식

Blockchain Connector는 압축되어 블록체인 원장 내에 저장된 트랜잭션 데이터를 꺼내 이를 Db2의 관계형 테이블로 제공합니다. 견고한 캐싱 방식을 통해 기존 Db2 기능을 사용하여 블록체인 데이터를 위한 캐시 테이블을 생성해 사용자가 필요할 때 최신 데이터를 불러올 수 있도록 하는 동시에 쿼리 성능을 개선합니다.



네이티브 블록체인 지원의 가치

블록체인은 이미 보험, 금융, 물류 및 의료 보건의료 같은 업계에서 변경 불가능한 거래 기록 저장에 사용되고 있지만, 이러한 업계에서는 데이터를 더 잘 이해하기 위해 분석도 실행하고자 합니다. 예를 들어 보험사는 특정 지역에서 고객별로 접수된 보험금 청구의 수를 알고자 하며, 배송 업체는 화물과 차량을 추적하고자 하고, 의료 보건 기업은 환자의 진료 기록을 파악하고자 합니다.

이때 이 데이터를 추출하기 위한 맞춤형 백 엔드를 구축하는 대신 간단히 Db2를 블록체인 데이터 저장소에 연결해서 다른 데이터를 다룰 때와 같은 방식으로 분석을 실행하면 됩니다. Db2에 내재하는 연결과 유연성 덕분에 부가적인 맥락 데이터를 블록체인 트랜잭션 데이터와 연계해서(예를 들어 특정 배송 동안의 기상, 비슷한 청구를 했을 가능성이 있는 보험 청구인의 관계인) 각 트랜잭션이 비즈니스에 의미하는 바를 훨씬 더 풍부하게, 전체적으로 이해할 수 있습니다. 블록체인 데이터에서 직접 실행되는 AI 앱도 이제 가능성 있는 이야기입니다.

AI 기반 조직에서 Db2의 위치

전통적인 온프레미스 배포, 클라우드 관계형 데이터베이스 또는 웨어하우스 인스턴스, 구독 기반 Hybrid Data Management Platform, 새로운 IBM Cloud Pak™ for Data 또는 다른 배포 모델 등 Db2 기능을 어떤 형태로 배포하든 관계없이 기반 기술이 가진 잠재력은 크고 비즈니스 혜택은 막대합니다.

Db2는 규모를 불문하고 모든 조직에서 AI를 사용자가 실제 활용할 수 있게 해 줍니다. 가치 제안은 명확합니다. 온프레미스와 클라우드를 가리지 않고 어떤 인프라에서든 엔터프라이즈 AI를 구축, 실행, 관리하는 것입니다. 물리적으로 어디에 위치하든 모든 데이터 소스를 이용할 수 있습니다. 데이터베이스에 적용된 AI 기능을 활용하여 데이터 기반의 의사 결정을 위한 예측적이고 선제적인 인사이트를 실현할 수 있습니다. 기업은 Db2를 데이터 기반으로 두면 AI 애플리케이션을 더 효율적으로 구축 및 연결하여 실행 가능한 인사이트를 신속하게 얻고 더욱 많은 사람들이 AI를 이용하도록 할 수 있습니다.

AI 데이터베이스, Db2로 시작하십시오.

[상담 예약➔](#)

[자세히 알아보기➔](#)



© Copyright IBM Corporation 2020. IBM Corporation, New Orchard Road, Armonk, NY 10504

미국에서 제작, 2020년 1월.

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