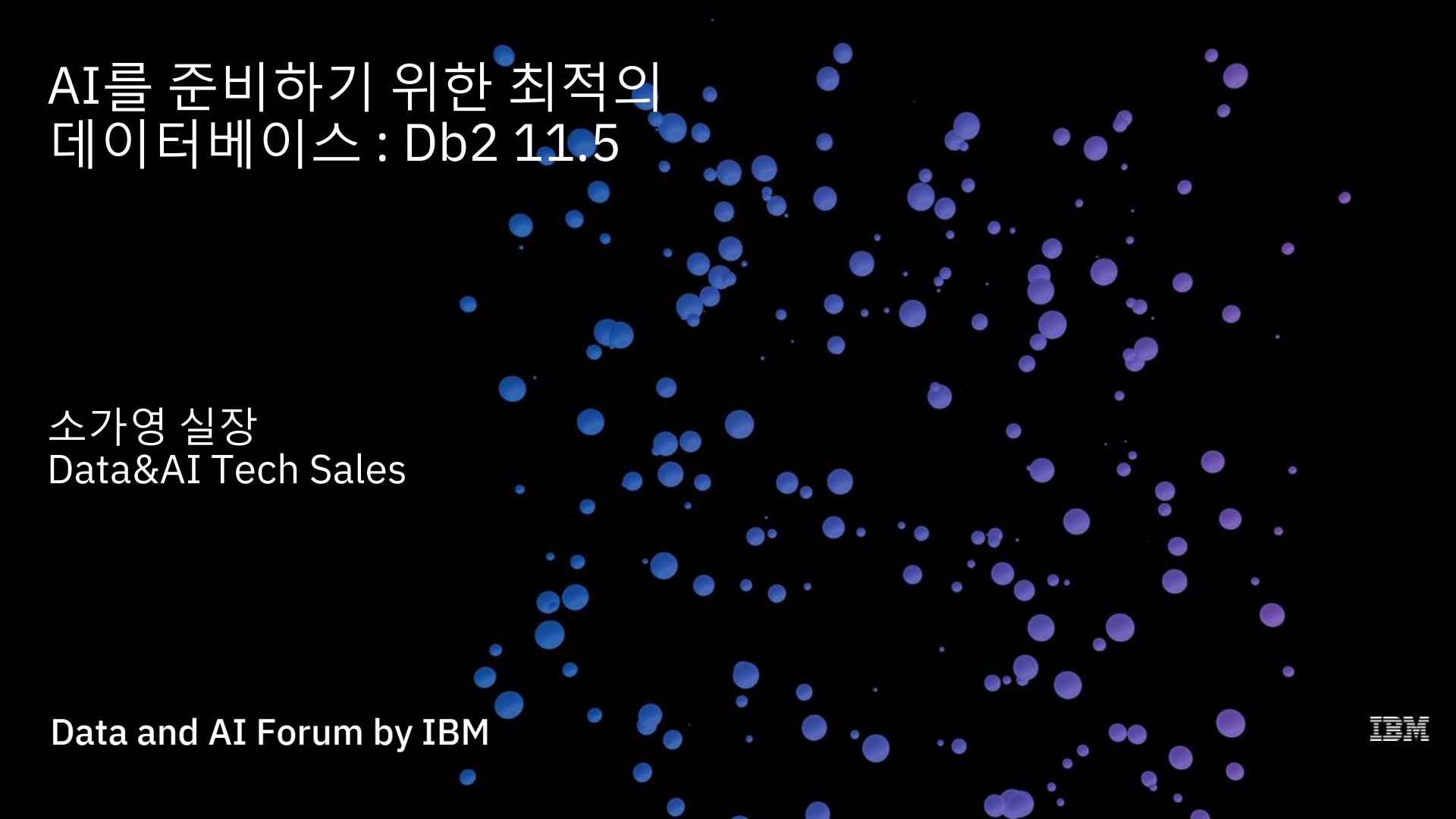
AI를 준비하기 위한 최적의 데이터베이스 : Db2 11.5

소가영 실장
Data&AI Tech Sales

Data and AI Forum by IBM



There is no AI without an IA

(information architecture)

99%

AI의 핵심은 학습입니다.

81%

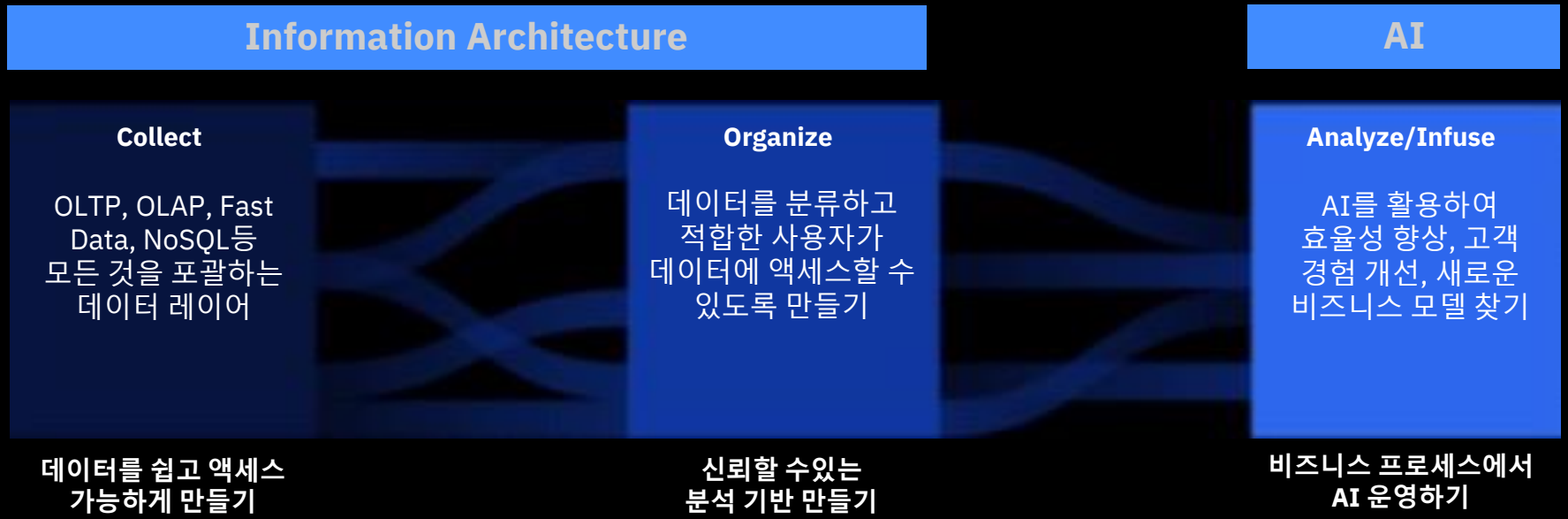
그러나, AI의 학습에 필요한 데이터를 얻기 어렵습니다.

"AI 에 대한 가장 일반적인 오해는 정교한 AI 알고리즘만으로도 충분한 데이터 없이 귀중한 비즈니스 솔루션을 제공 할 수 있다는 것입니다.

아무도 정교한 알고리즘으로 데이터 부족을 극복 할 수는 없습니다. 데이터 품질은 또 다른 과제입니다."

MITSloan Management Review

AI 활용의 라이프 사이클



The AI Ladder

AI로의 여행을 가속화하기 위한 기반

AI

INFUSE – 신뢰성과 투명성을 확보한 AI 운영

ANALYZE - 어디서나 AI를 활용하여 통찰력 확장

ORGANIZE - 신뢰할 수있는 분석 기반 만들기

COLLECT - 데이터를 쉽고 액세스 가능하게 만들기

데이터가 발생되고 저장되는
위치에 관계 없는 모든
데이터 유형

MODERNIZE
AI 및 멀티 클라우드
세상을위한 데이터 자산

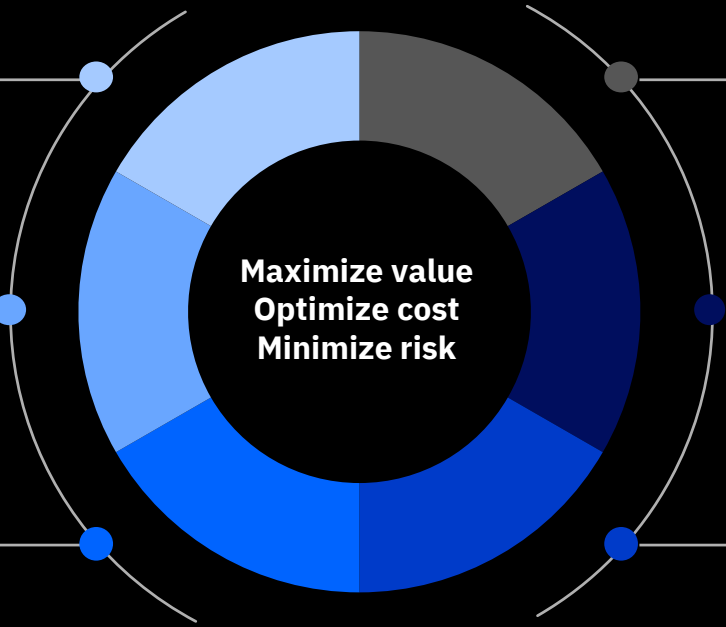
Hybrid Data Management 아키텍처

새로운 데이터 관리 전략

Multi-Cloud Availability
프라이빗 & 퍼블릭 클라우드

All Data
정형과 비정형 데이터
트랜잭션 & 분석 데이터

At Rest & In-Motion
이벤트 기반 시스템
부터 전통적 시스템까지



Self-service
모든 사람이
데이터 중심 의사 결정

Cloud Native
탄력적이고 자원 최적화

Built-in AI
보다 스마트한 의사 결정 및
최적화된 운영

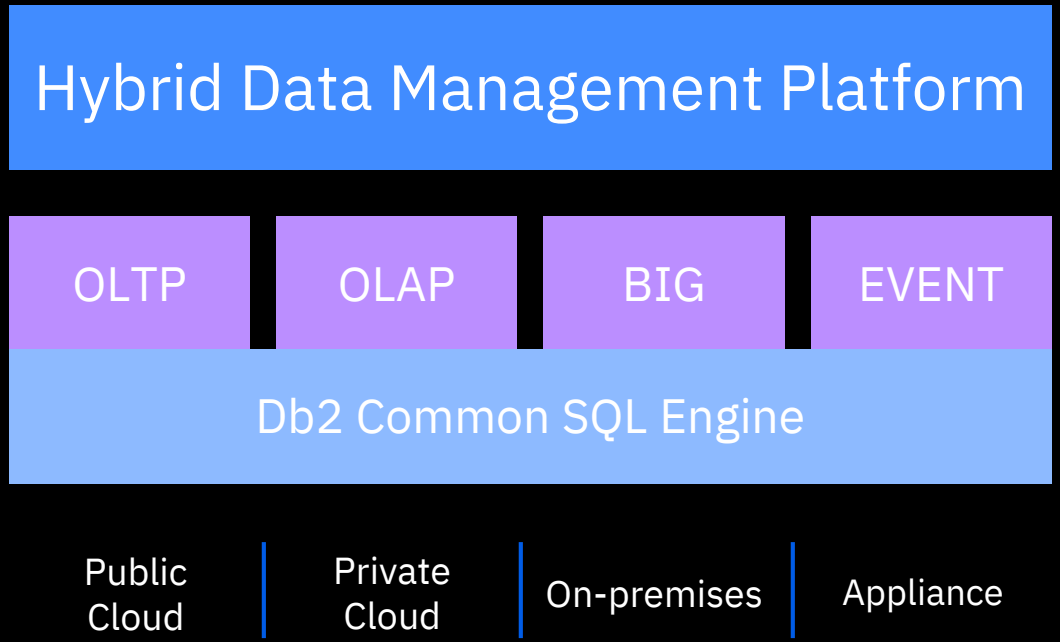
Collect : Hybrid Data Management

One entitlement

All workloads

One engine and experience

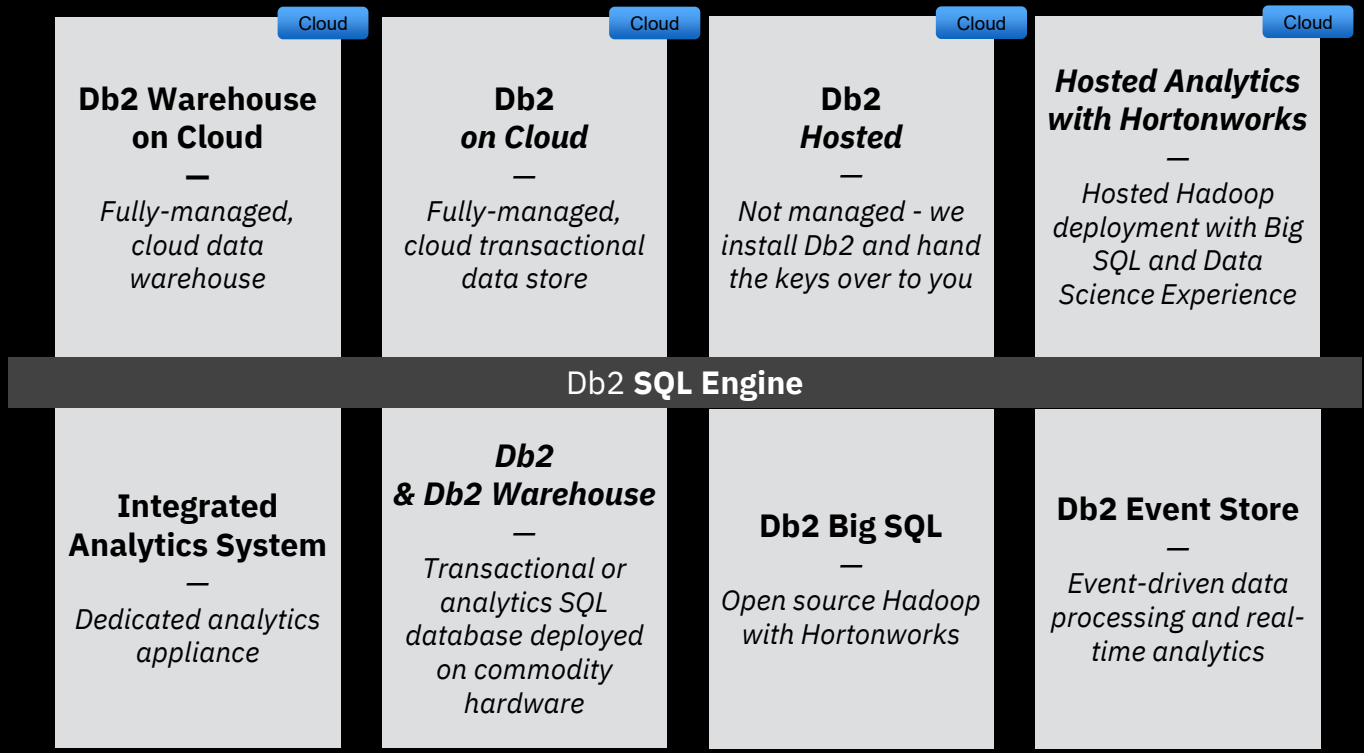
All deployment targets



HDM Solution Family

Db2 Common SQL Engine 에 기반으로 구축된 **Hybrid Data Management** 솔루션 패밀리

한번의 SQL 작성*으로 모든 플랫폼에 배포* ***어디서나 실행***



Db2 11.5

the AI Ready Database

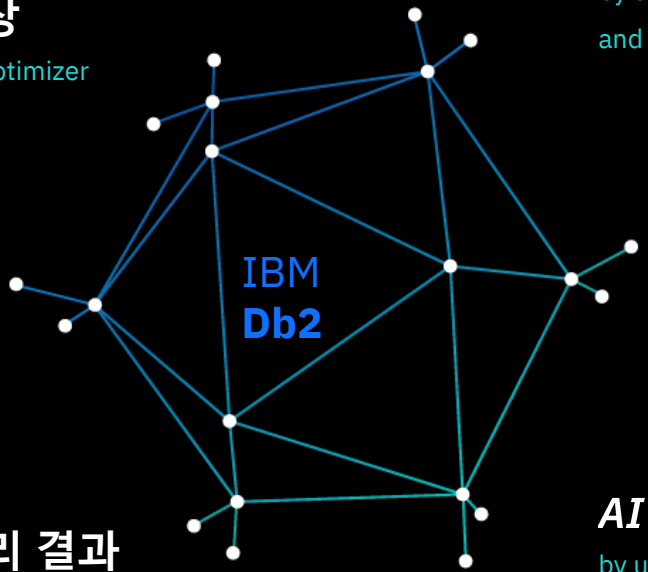
최대 10 배의 쿼리 성능 향상

powered by an ML-Optimizer

풍부하고 빠른 데이터 탐색

by using Natural Language Query

and Db2 Augmented Data Explorer



데이터 이동 없이 모든 데이터에 대한 단일 뷰

delivered by data virtualization

복잡한 관계 모델링

by using Db2 Graph & SQL

신뢰 기반 쿼리 결과

leveraging ML-SQL

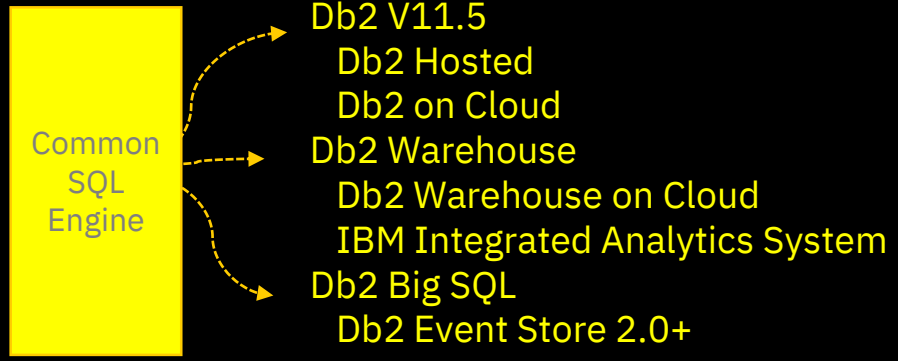
***AI* 기반 애플리케이션 구축**

by using Python, GO , JSON

and Jupyter notebooks

Db2 Common SQL Engine으로의 통합

- 모든 Db2 제품군을 동일한 버전의 Db2 Common SQL Engine 기반으로 통합하고 Continuous Delivery Model 형태로 제공됩니다.
  - 일관된 경험
  - Write-once Run-anywhere 애플리케이션 패러다임
  - 투자 보호
  - 좀 더 빠른 기능 배포



ML Optimizer and Cost Optimizer

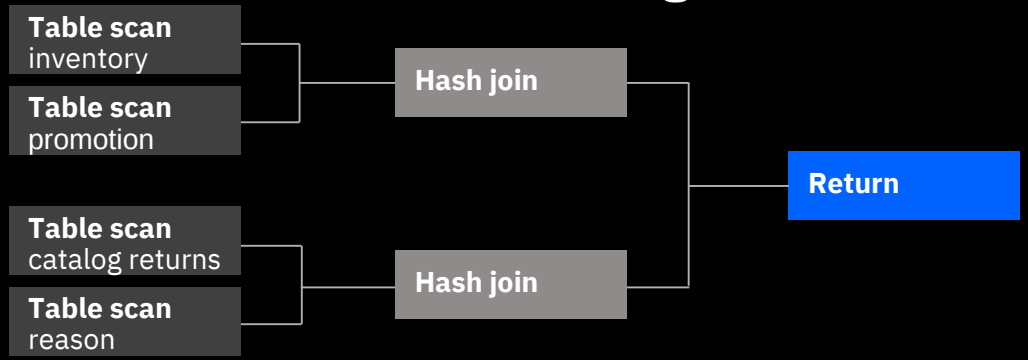
New

Machine Learning Optimizer
: 경험 기반으로 자동 향상되며, 쿼리 최적화를 단순하고 견고하며 안정적으로 지원

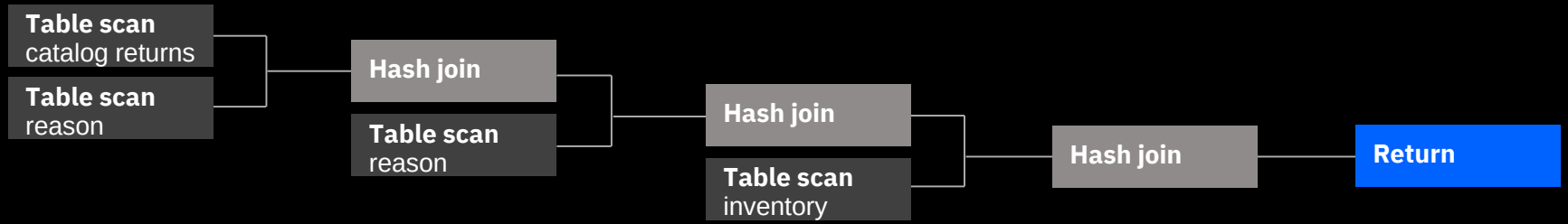
- 쿼리 실행 단계마다의 결과 행 수 예측은 쿼리 실행 성능에 영향을 미치는 가장 일반적인 이슈이며, 옵티마이저의 카디널리티 추정을 통해 계산됩니다.
- 기본적으로 이 카디널리티 추정의 정확도를 향상시키는 것이 튜닝에 큰 영향을 미칩니다.
- **Phase 1 – Cardinality Estimation**
  - Initial phase – expression 이 없는 equal 연산과 로컬 range 연산 지원
  - Future phases – expression, predicate 및 BLU에 대한 카디널리티 지원, 조인 계산 지원

ML Optimizer

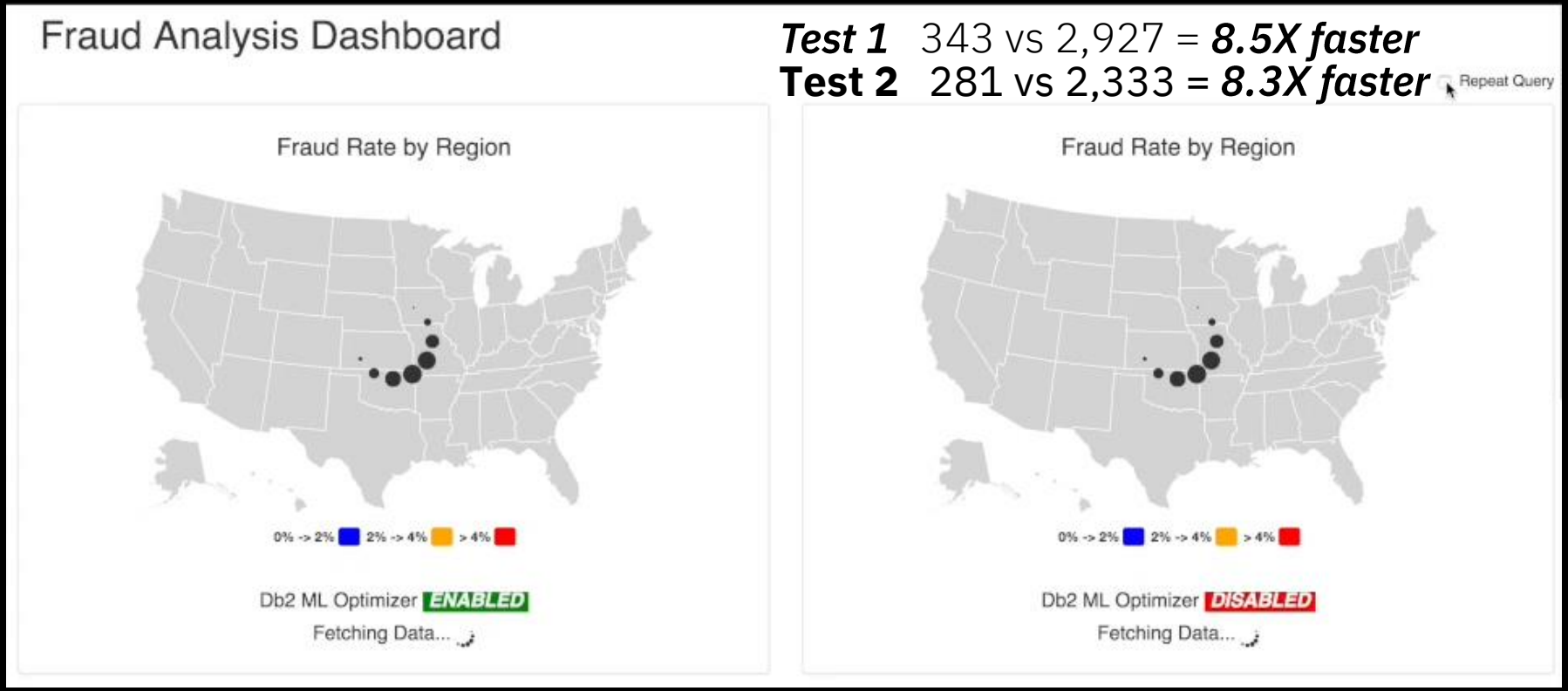
- **Without Machine Learning**



- **With Machine Learning**



ML Optimizer Performance Test



Db2 Graph Capability

실시간 시각화 및 모델링

Post-GA

하나의 데이터에 그래프 쿼리 및 SQL 지원

- 그래프 어플리케이션과 Db2 사이의 그래프 레이어
- 관계형 데이터를 위한 Db2 런타임의 변경 없음
- Gremlin 언어 및 Tinkerpop 프레임 워크를 통한 오픈 소스 그래프 지원
- SQL 엔진이 쿼리 할 수 있도록 그래프 데이터를 테이블에 저장
- ACID 트랜잭션은 기존 관계형 응용 프로그램을 방해하지 않으며 그래프를 실시간으로 업데이트
  - 트랜잭션 발생 시점에 그래프를 실시간으로 볼 수있는 기능
- Gremlin API를 사용하여 데이터의 그래프 보기 제공
- SQL은 그래프 데이터에서 완벽하게 작동하므로 그래프 데이터의 SQL 분석 수행

Natural Language Query (NLQ)

데이터 분석을 위한 새로운 툴

New

ANSI SQL 구문을 알아야 할 필요성 제거

- 대화 분석은 미래입니다.
  - Alexa, Google Home 등이 트렌드를 입증하고 있습니다.
  - 데이터와의 대화의 자연스러운 확장입니다.
  - 빠르고 쉽게 적응가능하며 모든 스키마 / 테이블에서 작동 할 수 있습니다.
  - 비즈니스 사용자를 위한 대부분의 사용 사례에 적용될 수 있습니다.
- Developer Works에서 무료로 다운로드 할 수있는 도구입니다.
  - Db2와의 완벽한 플러그 앤 플레이를 지원합니다.
  - 현재 Beta 지원

Augmented Data Exploration

풍부하고 깊은 통찰력을 위한 새로운 도구

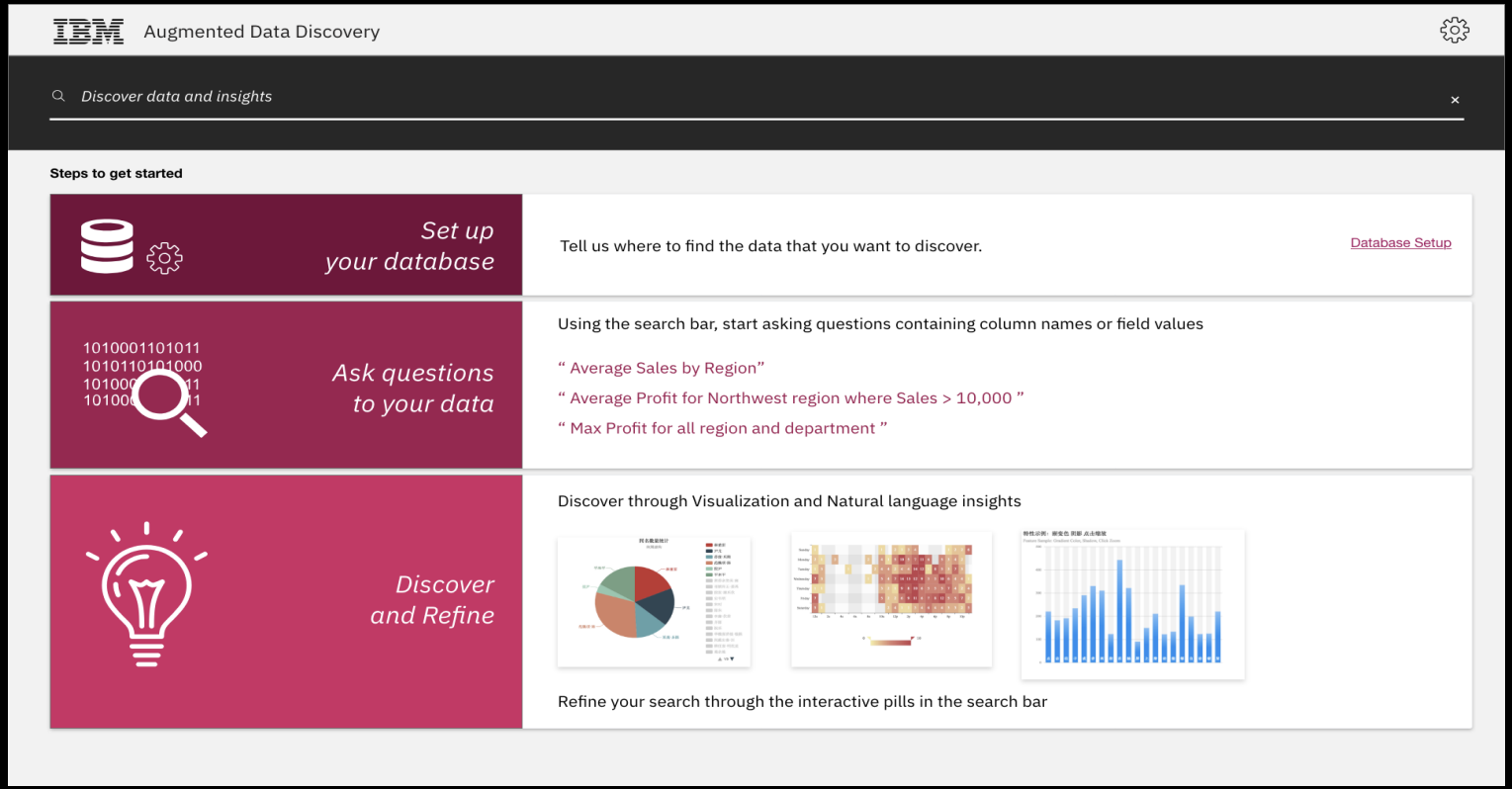
New

익숙하지 않은
데이터 세트
탐색에서
더 빠른 통찰력 확보

- 데이터 사이언티스트는 종종 익숙하지 않은 데이터 세트를 탐색해야하는 어려움에 직면합니다.
  - 많은 시간이 소모됩니다.
  - 무엇을 찾아야할지 알기 어렵습니다.
  - 데이터 세트의 내용에 대한 이해 및 지식 이전이 필요합니다.
- 직관적 인 데이터 탐색 도구가 필요합니다
  - 최소한의 터치로 데이터에 대한 여러 보기 형태를 제공합니다.
- Developer Works에서 무료로 다운로드 할 수있는 도구입니다.
  - Db2와의 완벽한 플러그 앤 플레이를 지원합니다.
  - 현재 Beta 지원

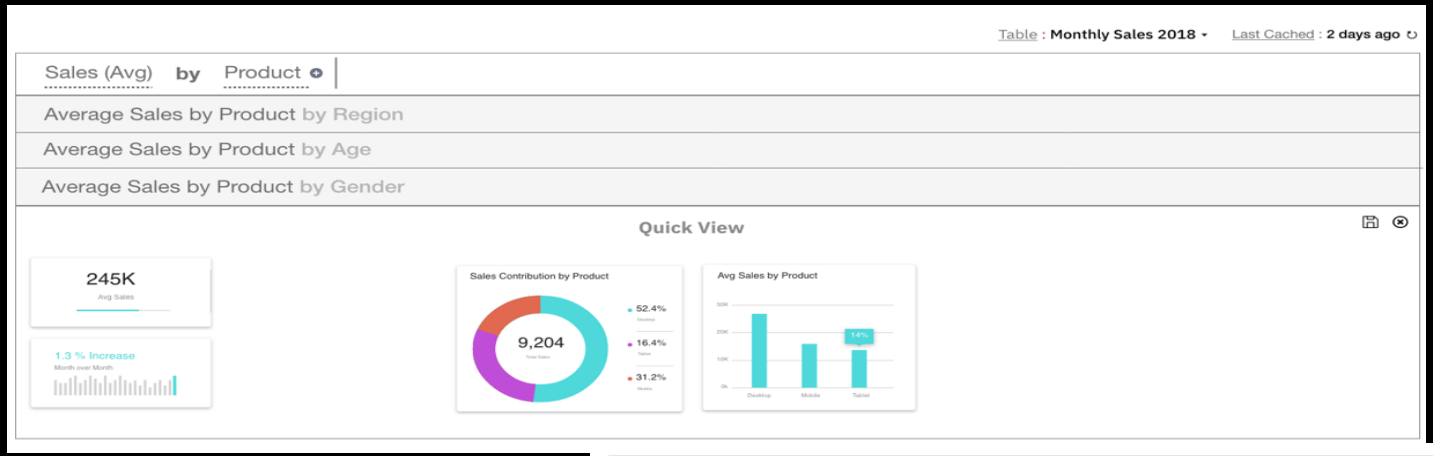
새로운 Db2 Augmented Data Explorer 테스트 해보기

https://www.ibm.com/us-en/marketplace/db2-augmented-data-explorer

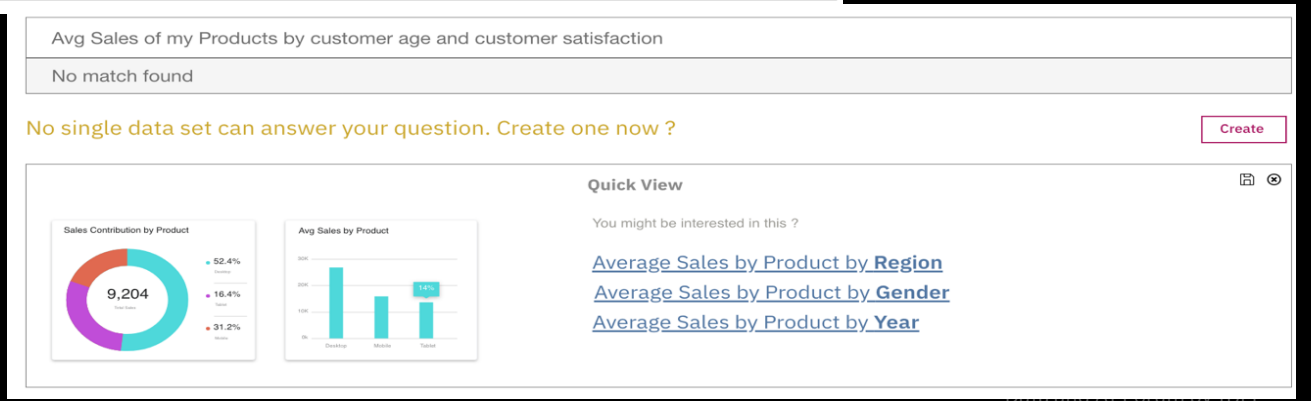


새로운 Db2 Augmented Data Explorer 테스트 해보기

https://www.ibm.com/us-en/marketplace/db2-augmented-data-explorer



Answer known and unknown questions



IBM Db2 Augmented Data Explorer

Beta

Start your data discovery here... e.g. avg sales by region

Discoveries

Live updates

Welcome to IBM Db2 Augmented Data Explorer!

Take a quick tour to get started.

Skip

Next (1/4)

Getting started

You may want to check out the tour!

Point to your database

ections by clicking

k Configure, and then crawl your data.

You can also start by interacting with our sample data.

Search your data

Use the search bar to ask questions to your data. Below are some sample questions:

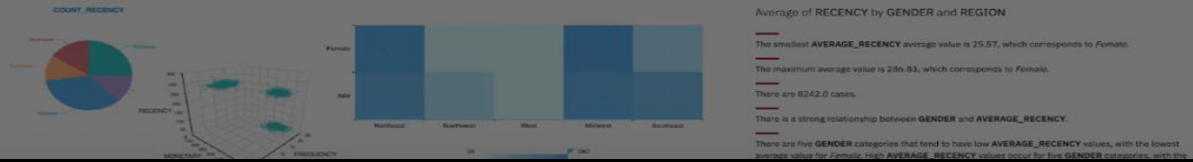
"Average Avg_Money_Spent by Region"

"Average Number_of_Purchases by Age where Days_Since_Last_Purchase > 30"

"Days_Since_Last_Purchase, Number_of_Purchases, Avg_Money_Spent, Region"

Discover insights and refine

Discover insights about your data through automatically generated visualizations and natural language summaries

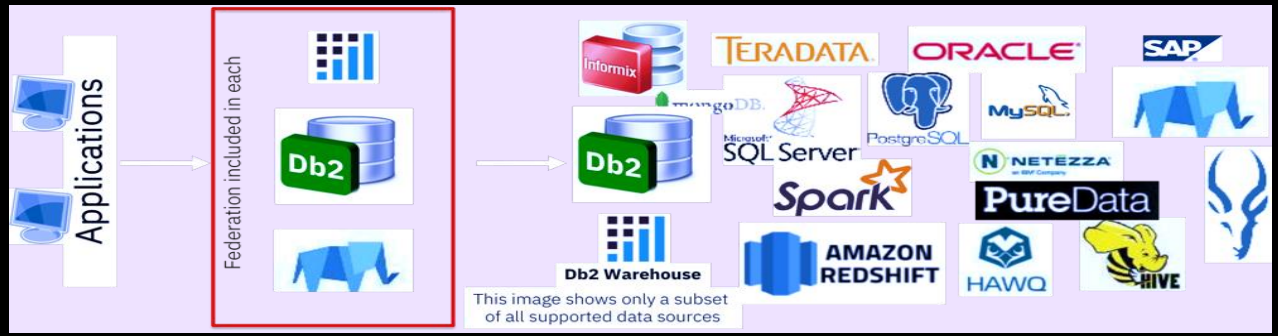


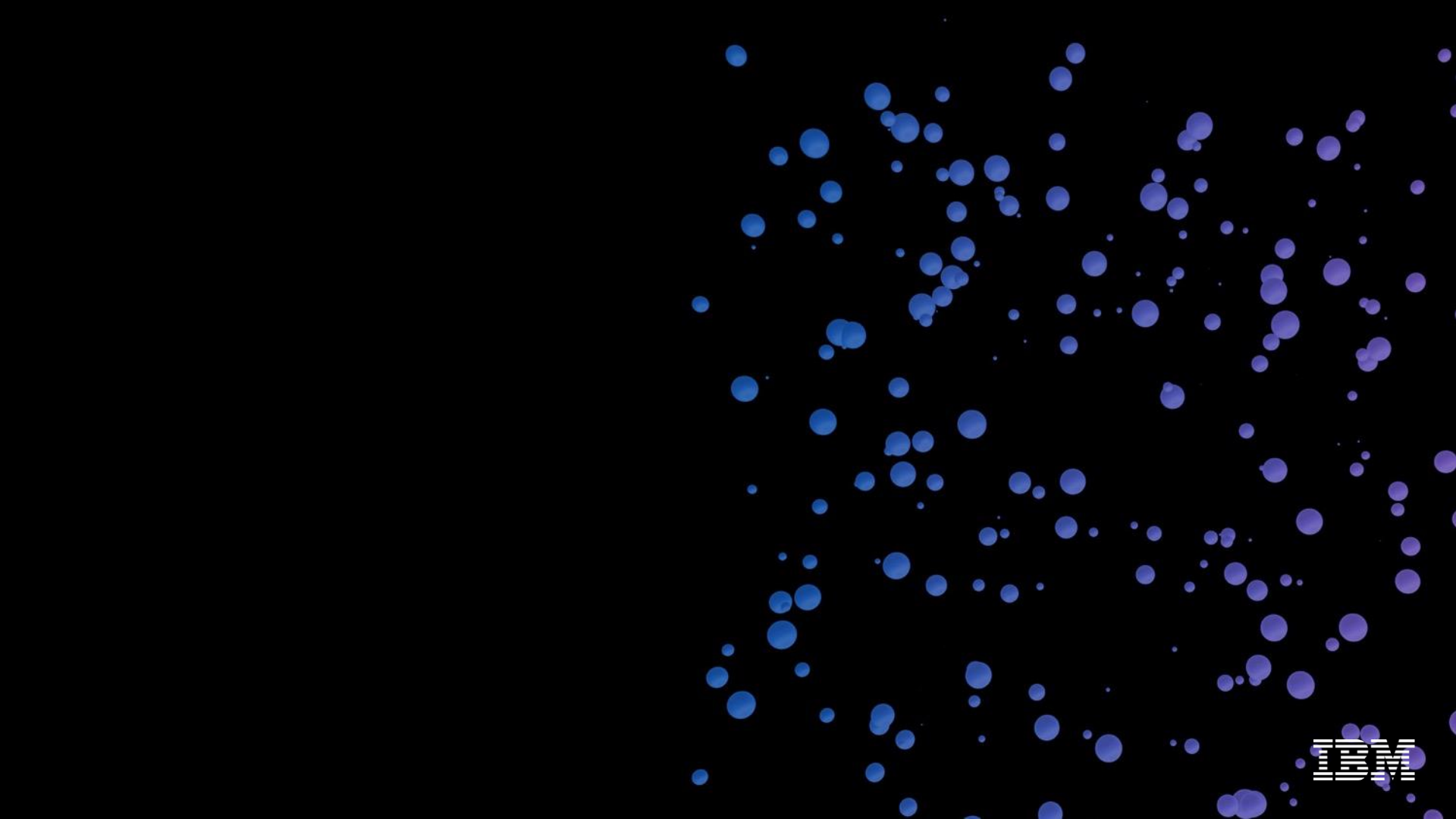
Data Virtualization

모든 비즈니스 데이터의 연계 : 차세대 Federation

여러 데이터 저장소에 분산 된 데이터로 작업 할 수있는 정교한 기능을 제공

- 페더레이션 기능 및 성능에 계속적 투자로 기능이 확장됩니다.
- 점점 더 많은 데이터 소스에 페더레이션 가능합니다.
- 점점 더 많은 드라이버가 Db2 설치 프로그램에 직접 통합되어 제공됩니다.
- SQL 및 NoSQL 기반 데이터 저장소 모두로 확장됩니다.
- 온-프레미스 및 클라우드 데이터 소스 모두로 확장됩니다.
- PostgreSQL 및 MySQL과 같은 오픈 소스 데이터베이스로 확장됩니다.




IBM