Scale AI workloads for all your data, anywhere

Kevin Shen
Product Manager | Data & AI Software
Yuankai.shen@ibm.com

Anson Kokkat
Product Manager | Data & AI Software
ansonk@ibm.com

Sonia Mezetta
Product Management Director | Data & AI Software
somezz@us.ibm.com

Adam Learmonth
Advisory, Learning Content Development
Adam.Learmonth@ca.ibm.com

watsonx.data™

IBM

The speed, scope, and scale of generative AI impact is unprecedented

Massive early adoption

80%

of enterprises are working with or planning to leverage foundation models and adopt generative AI

Broad-reaching & deep impact

Generative AI could raise global GDP by

7%

within 10 years

Critical focus of AI activity & investment

Generative AI expected to represent

30%

of overall market by 2025

Sources: Statista; Reuters; Goldman Sachs; IBM Institute for Business Value; Gartner. Scale Zeitgeist: AI Readiness Report, a survey of more than 1,600 executives and ML practitioners

However, AI is only as good as your data.

IT and data leaders are faced with unprecedented data challenges to scale AI

There’s more data

Exploding data growth

The aggregate volume of data stored is set to grow over 250% in the next 5 years.

In more locations

Multiple locations, clouds, applications, and silos

82% of enterprises are inhibited by data silos.

In more formats

Documents, images, video

80% of time* is spent on data cleaning, integration, and preparation.

With less quality

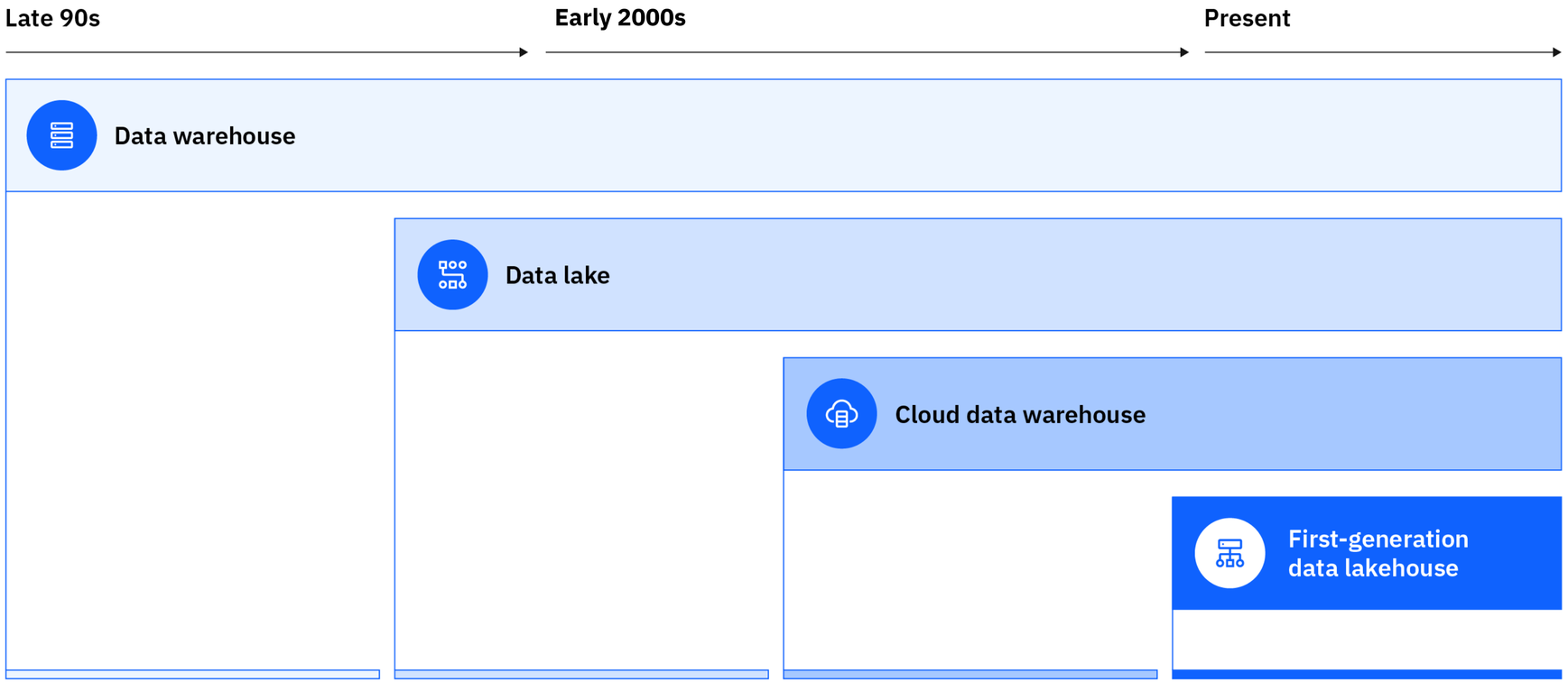
Stale and inconsistent

82% of enterprises say data quality is a barrier on their data integration projects.

This environment leads to more cost and complexity for those who seek to govern data for AI.

* Source: https://www.idc.com/getdoc.jsp?containerId=US49018922)

Traditional approaches to addressing these challenges have created more overall complexity and cost, which has led to the emergence of data lakehouse architectures



Today, leaders at most large enterprises manage their data and workloads using a mix of data repositories and data stores in hybrid environments.

The overall cost across all these repositories remains high.

It's difficult for leaders to effectively leverage and govern the data across multiple environments and use enterprise data for analytics and AI.

Introducing …

watsonx

What IBM offers

The platform for AI and data

watsonx

watsonx.ai

Train, validate, tune, and deploy AI models

A next generation enterprise studio for AI builders to train, validate, tune, and deploy both traditional machine learning and new generative AI capabilities powered by foundation models. It enables you to build AI applications in a fraction of the time with a fraction of the data.

watsonx.data

Scale AI workloads, for all your data, anywhere

Fit-for-purpose data store, built on an open lakehouse architecture, supported by querying, governance and open data formats to access and share data.

watsonx.governance

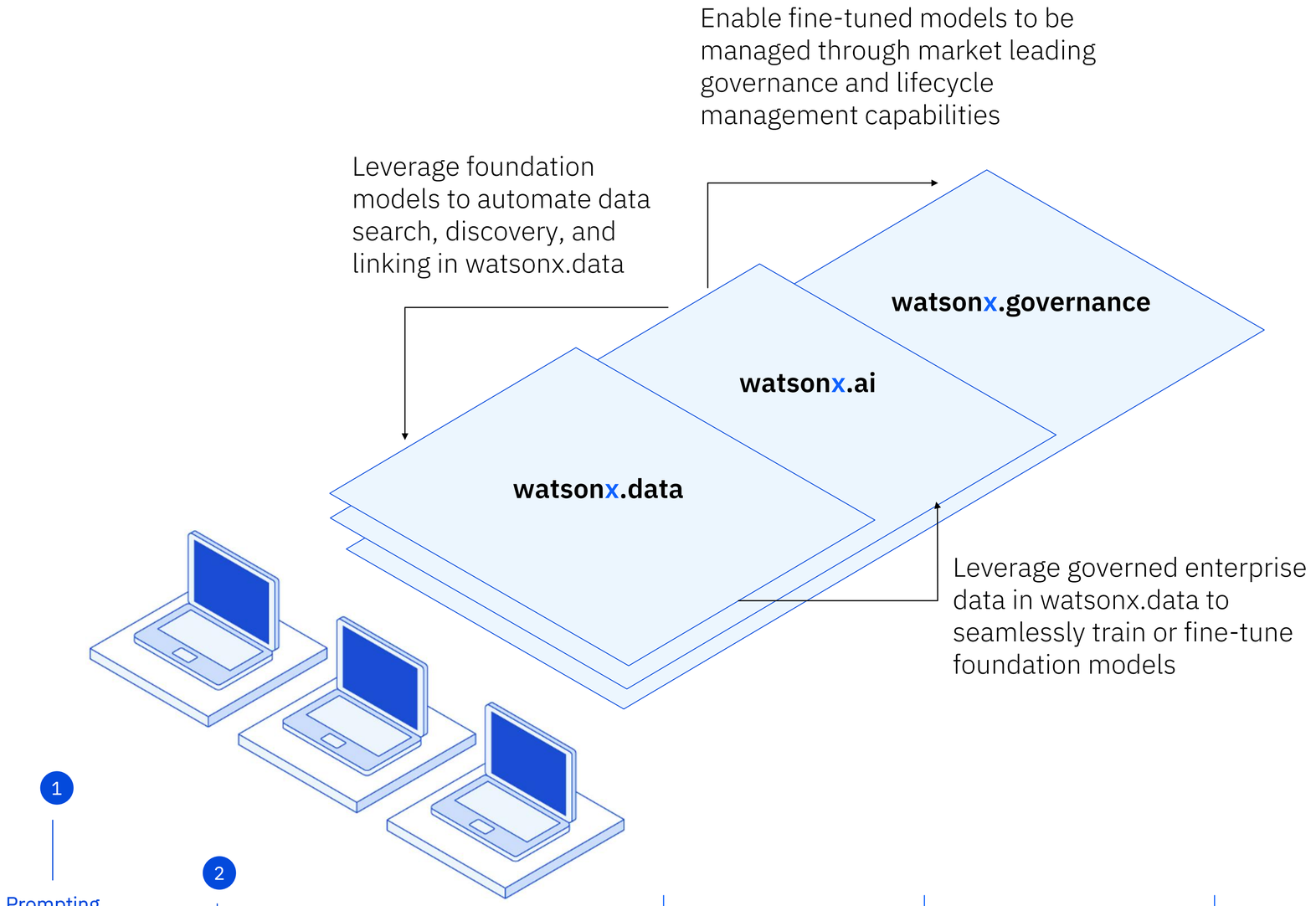
Accelerate responsible, transparent, and explainable AI workflows

End-to-end toolkit for AI governance across the entire model lifecycle to enable responsible, transparent, and explainable AI workflows.

The platform for AI and data |

watsonx

Scale and accelerate the impact of AI with trusted data.



watsonx.data

Scale AI workloads, for all your data, anywhere

watsonx.ai

Train, validate, tune and deploy AI models

watsonx.governance

Enable responsible, transparent and explainable AI workflows

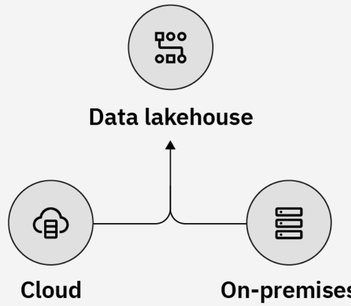
watsonx.data

Scale AI workloads, for all your data, anywhere

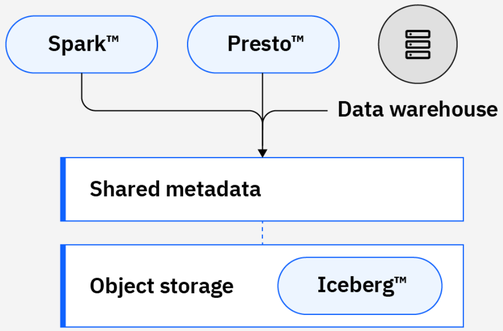
A fit-for-purpose data store, based on an open lakehouse architecture, supported by querying, governance, and open data formats to access and share data.

Seamlessly deploy across any cloud or on-premises environment in minutes with workload portability through Red Hat® OpenShift®.

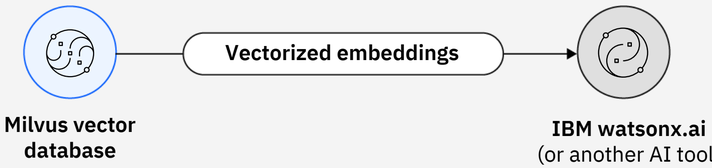
Access all your data through a single point of entry across all clouds and on-premises environments.



Reduce the cost of your data warehouse by up to 50%* through workload optimization across multiple query engines and storage tiers.



Unify, curate, and prepare data for AI

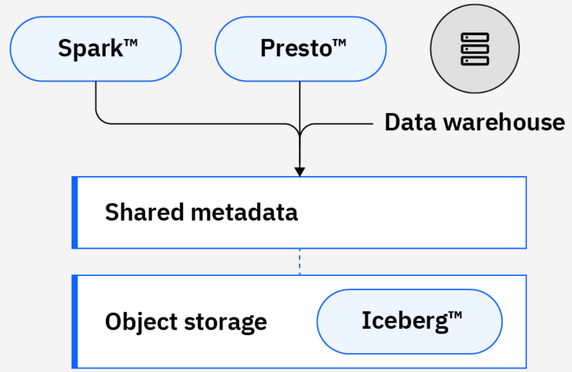


*When comparing published 2023 list prices normalized for VPC hours of IBM watsonx.data to several major cloud data warehouse vendors. Savings may vary depending on configurations, workloads and vendors.

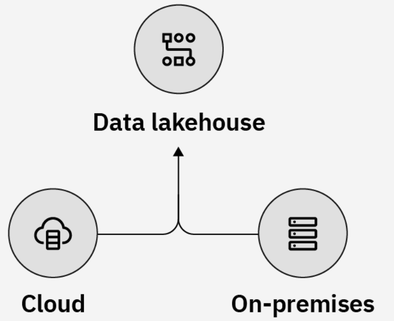
Access all your data across hybrid cloud through a single point of entry

Connect to storage and analytics environments in minutes with a shared metadata layer across clouds and on-premises environments

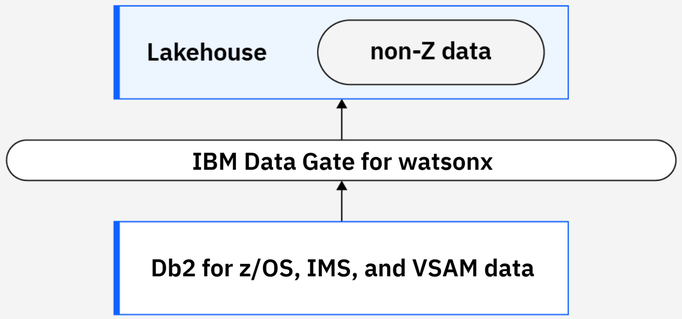
1 Share a single copy of data with tools that can read open data formats to minimize data duplication



2 Connect to and access data remotely across hybrid cloud with the ability to cache remote sources



3 Synchronize and incorporate Db2 for z/OS data for lakehouse analytics

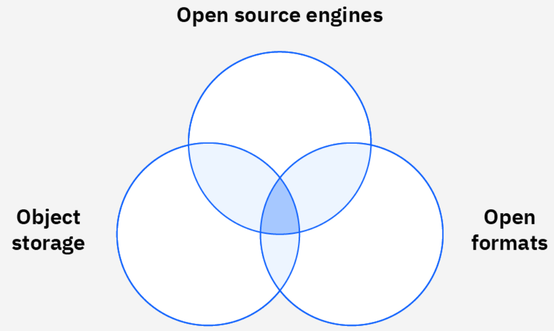


Reduce your data warehouse costs by up to 50%* by optimizing workloads

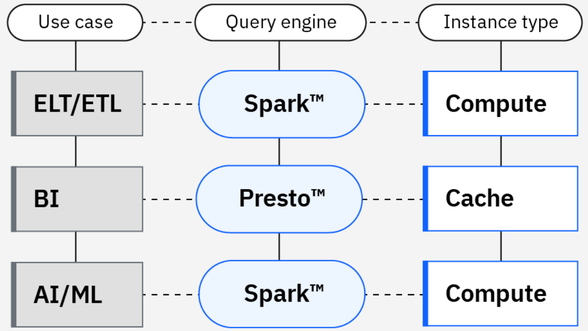
Optimize your data management workloads with multiple fit-for-purpose engines and shared metadata layer, so the right workload is seamlessly paired with the right engine

*When comparing published 2023 list prices normalized for VPC hours of IBM watsonx.data to several major cloud data warehouse vendors. Savings may vary depending on configurations, workloads and vendors.

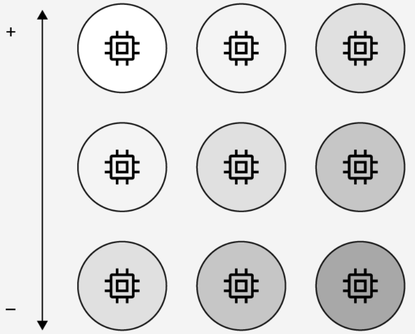
1 Share data between multiple analytics engines



2 Use fit-for-purpose compute and cache-optimized instances



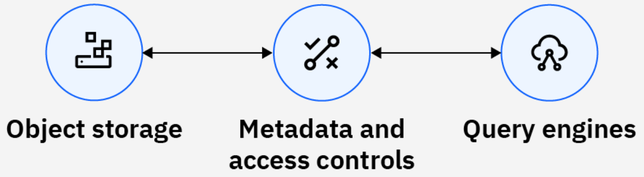
3 Dynamically scale up and scale down



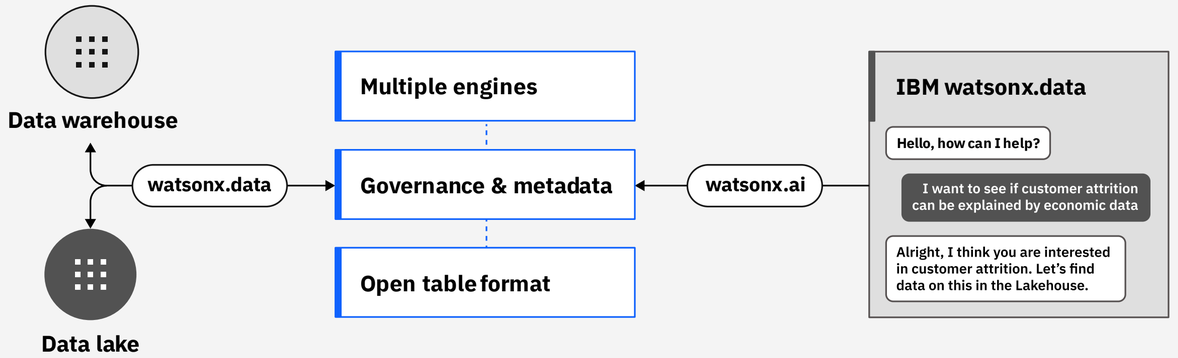
Unify, curate and prepare data for AI

Unify your data, accelerate data discovery and insights with Generative AI, and store vectorized embeddings to enable RAG use cases at scale across large sets of your trusted, governed data.

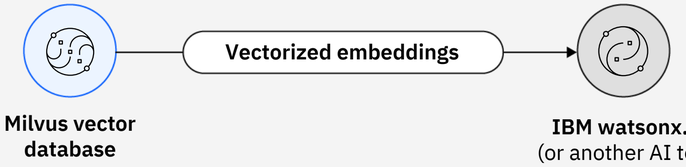
Connect to your existing analytics data and deploy fit-for-purpose query engines in minutes



Use foundation models to accelerate data discovery and insights in natural language

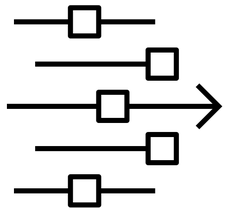


Store vectorized embeddings to improve the relevance and precision of your AI



Organizations need to meet evolving data management needs without being slowed down by the cost and complexity of data siloes and migration

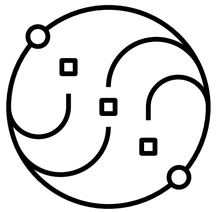
Traditional path



Migrate workloads to realize value

1. Migrate the data and workloads
2. Understand the trade-offs of new technology and adopt new workflow
3. Plan for the next migration

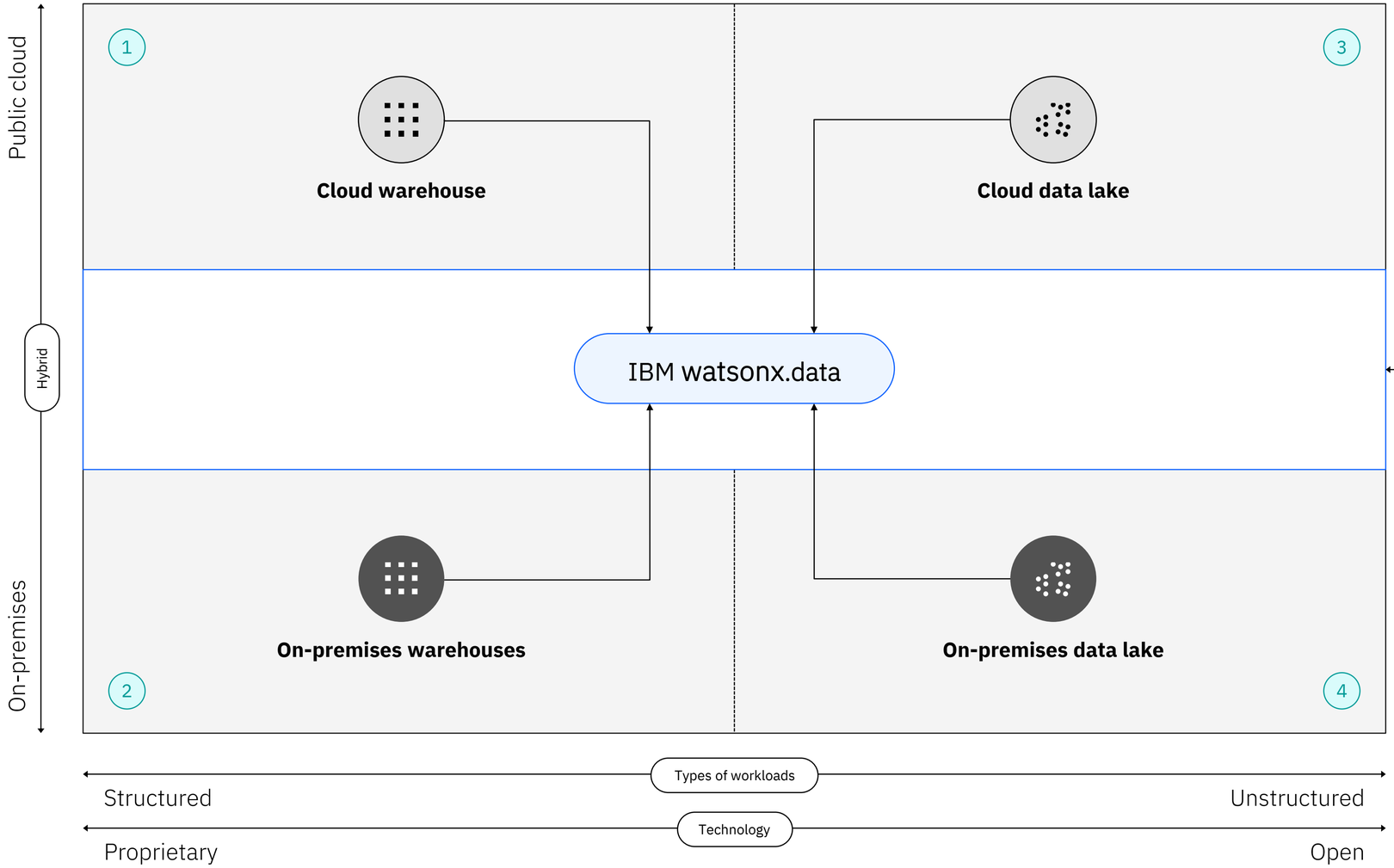
With **IBM watsonx.data**



Future-proof and unlock value from day one

1. Use new technology with **existing data investments** through a shared metadata layer
2. Move data and workloads **at your own pace**
3. Eliminate data siloes and end constant migration with **open formats and standards** to support interoperability with future technology stacks

Access all your data, quickly and optimize your data architecture with multi-engine support and hybrid deployment of analytics and AI workloads



1

Optimize costly cloud warehouses
Make the most of fit-for-purpose query engines and compute resources

2

Optimize & access on-premises warehouses
Use low-cost object storage and fit-for-purpose engines

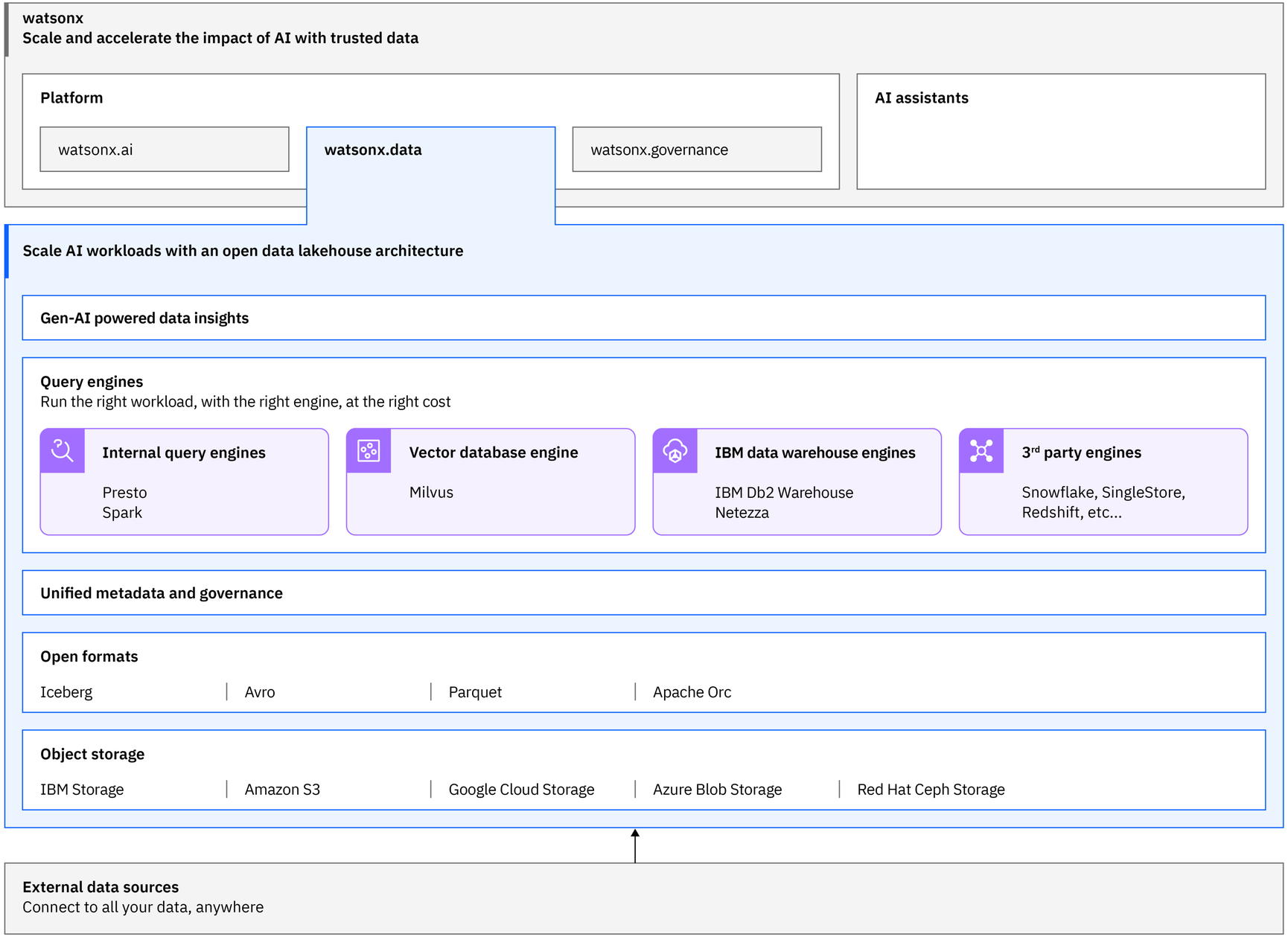
3 4

Modernize data lakes
Run existing reporting and enable new AI workloads without the cost and complexity of Hadoop

1 2 3 4

Deploy across hybrid cloud and multicloud
Seamlessly deploy to both the public cloud and to your existing on-premises investment

Overview of the key components of the IBM watsonx.data



Key Features:

Multiple engines such as Presto, Presto C++, and Spark that provide fast, reliable, and efficient processing of big data at scale

Integrated vector database built on open-source Milvus to store vectorized embeddings for AI applications at scale

Built-in governance that is compatible with existing solutions such as watsonx.governance and IBM Knowledge Catalog

Vendor agnostic open formats, allowing different engines to access and share the same data, at the same time and **unlocking data** from existing data investments for AI, such as z/OS mainframe data

Cost effective, simple object storage available across hybrid-cloud and multicloud environments

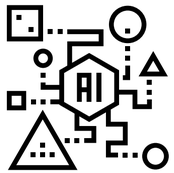
Hybrid-cloud deployments and workload portability across hyperscalers and on-prem with Red Hat OpenShift

watsonx.data

Benefits

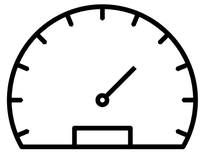
Minimize business disruption

- Metadata sharing between engines
- Compatibility with legacy open table formats
- Avoid forced migration and duplication



Unlock multi-modality data for AI and analytics

- Integrated data management for generative AI
- Prepare multi-modal data for AI use cases (such as Retrieval Augmented Generation)



Accelerate time to value

- Reduce ETL and data duplication
- Unlock mainframe data for AI with IBM Data Gate for watsonx integration
- Accelerate data discovery and enrichment

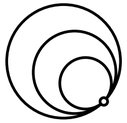
Infuse openness and interoperability

- Iceberg open table standard
- Presto open-source query engine
- Metadata store based on open-source Hive Metadata Store
- Plug-and-play engines
- HDFS access

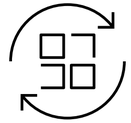
The IBM approach to a data lakehouse architecture combines the best of IBM with the best of open source

Best-in-class cost and performance optimizations for compute and storage

Built-in integrations with IBM data repositories and data fabric

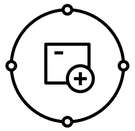


Deep expertise and capabilities in data and storage



Open and vendor-agnostic across architectural tiers

Enables hybrid, multicloud deployments with the Red Hat OpenShift platform



The best of open source

presto

Apache Spark

AVRO

HIVE

ICEBERG

Parquet

Apache orc

watsonx.data

The power of open source

Why **Apache Iceberg?**

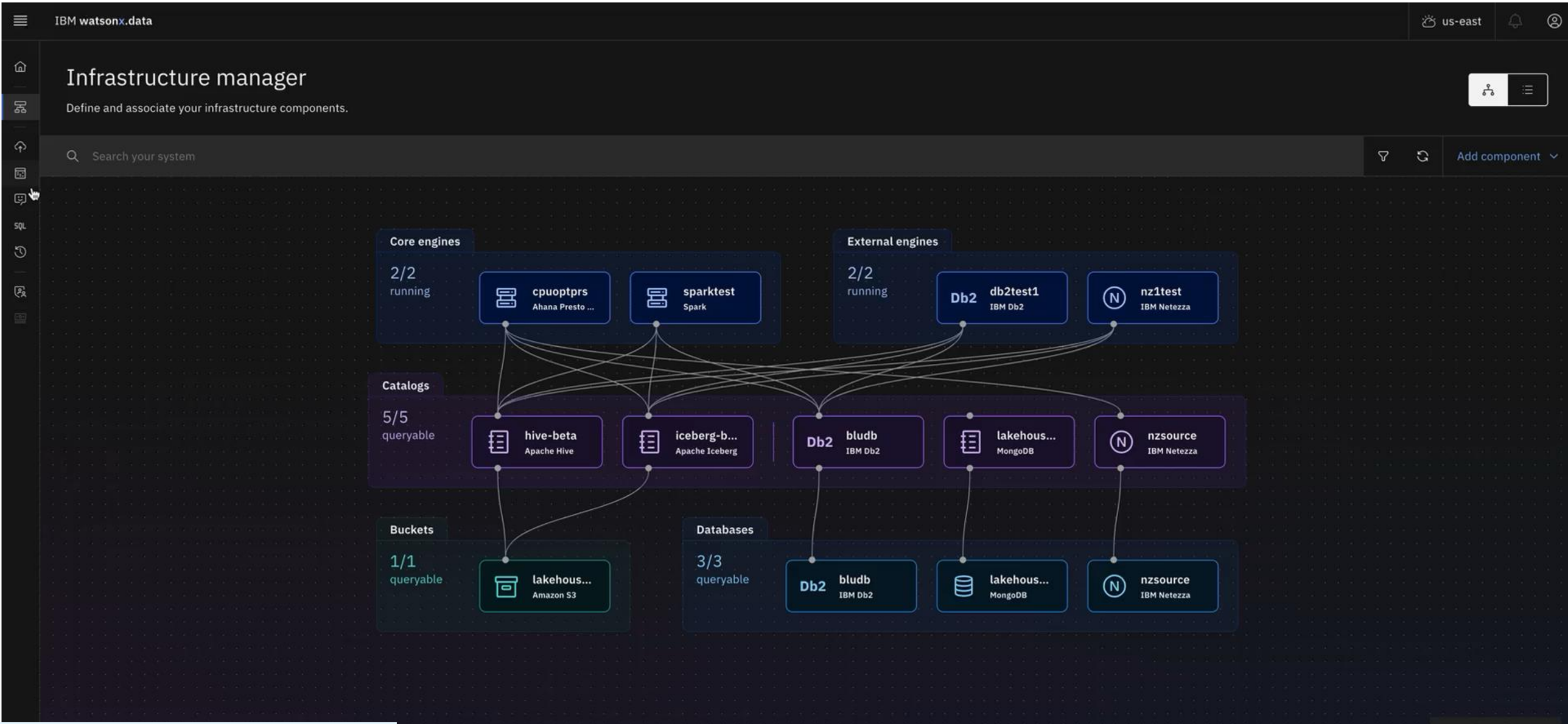
Apache Iceberg is an **open-source data table format** that helps simplify data processing on large datasets stored in data lakes

- SQL — Use it to build the data lake and perform most operations without learning a new language
- Data Consistency — ACID compliance (not just append data operations to tables)
- Schema Evolution — Add/remove columns without distributing underlying table structure
- Data Versioning — Time travel support that lets you analyze data changes between updates and deletes
- Cross Platform Support — Supports variety of storage systems and query engines (Spark, Presto, Hive, +++)

Why **Presto?**

Presto is an an open-source distributed SQL engine suitable for querying large amounts of data

- Ease of use — Easy to use with data analytics and business intelligence tools
- Cost effective — Engines are ephemeral and can be spun up and shut down as needed
- No-copy data access — Presto connectors allow access to data in-place, allowing for **no-copy data access and federated querying**
- Support — Supports relational and non-relational sources, interactive and batch workloads, and a wide variety of data sources, including:



watsonx.data

watsonx.data

Use Cases

Data warehouse optimization

Optimize workloads from your data warehouse by choosing the right engine for the right workload, at the right cost. Replace ETL jobs and reduce costs of your data warehouse by up to 50% through workload optimization.

Data lake modernization

Augment Hadoop data lakes using watsonx.data and access better performance, security, and governance, without migration or ETL

Mainframe data for AI

Unleash the power of mainframe data for AI and analytics in watsonx.data with integration to IBM Data Gate for watsonx and Data Virtualization Manager for z/OS. Readily virtualize or replicate data to Iceberg for analytics and AI.

Datastore for Generative AI

Unify, curate, and prepare data efficiently for AI models and applications. Integrated vectorized embedding capabilities enable RAG use cases at scale across large sets of your trusted, governed data.

Generative AI powered data insights

Leverage Gen-AI infused in watsonx.data to find and understand data and unlock new data insights through semantic search — no SQL required. Unleash cryptic structured data using auto-generated semantic metadata in natural language for easy self-service access to data.

watsonx.data is helping companies scale their AI workloads

intel

“We look forward to partnering with IBM to optimize the watsonx.data stack and contributing to the open-source community.”

Das Kamhout
VP and Senior Principal Engineer
Intel

capital bank

"We’re excited to see how watsonx can help us drive predictive analytics, identify fraud, and optimize our marketing.”

Bahaa’ Awartany
Chief Data Officer
Capital Bank of Jordan

CLOUDERA

“Customers will benefit from a truly open and interoperable hybrid data platform that fuels the adoption of AI.”

Paul Codding
EVP of Product Management
Cloudera

NucleusTeq

“We believe watsonx.data will help enterprises lower storage costs, optimize compute, and ensure seamless data management.”

Ashish Baghel
CEO and Founder
NucleusTeq

Powered by

presto

Adobe

Digital advertising platform

Over 2000 daily reports and 100s of pipelines on a 7 PB data lake with over 400 billion records

Bolt

Ride-hailing, micromobility rentals, and food delivery in Europe and Africa

Up to 100,000 daily queries (over 1.5 million queries per month) with over 2000 active internal users on 2 PB data lake

Meta

Social media

30,000 queries per day with 1000 daily active users on a 300 PB data lake

Uber

Ride-hailing, food delivery

Over 100 million queries per day with 7000 weekly active users on a 50 PB data lake

ByteDance

Internet technology

Over 2 million queries per day for business intelligence and one-off use cases

twilio

Communications API technology

Over 2700 active internal users running 1 million queries scanning 40 PB of data per month

Analysts agree, IBM is a leader in the data management market

Forrester Wave: Data Management for Analytics

Gartner Magic Quadrant for Cloud Database Management Solutions

Forrester: The Total Economic Impact™ for IBM Data Management

Sources: Forrester, Gartner

What IBM offers

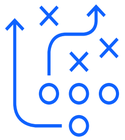
Why IBM?

Open	IBM’s AI is based on the best open technologies available
Trusted	IBM’s AI is transparent, responsible, and governed
Targeted	IBM’s AI is designed for enterprise and targeted at business domains
Empowering	IBM’s AI is for value creators, not just users

Getting started

Three ways to get started with watsonx.data today

IBM’s investment in partnering with clients



Free trial

Experience watsonx.data and test out core capabilities with a free trial.

[Try our free trial]

Client briefing

Discussion and custom demonstration of IBM’s generative AI watsonx point-of-view and capabilities. Understand how watsonx.data can be leveraged in any businesses AI strategy.

2-4 hours

Pilot program

Watsonx pilot developed with IBM AI engineers. Prove watsonx.data value for the selected use case(s) with a plan for adoption.

1-4 weeks

Thank you

Appendix

IBM’s generative AI technology and expertise

AI assistants

Empower individuals to do work without expert knowledge across a variety of business processes and applications.

watsonx Code Assistant
watsonx Assistant
watsonx Orchestrate
watsonx Orders

SDKs and APIs

Embed watsonx platform in 3rd party assistants and applications using programmatic interfaces.

Ecosystem integrations

AI and data platform

Leverage generative AI and machine learning — tuned with your data — with responsibility, transparency, and explainability.

watsonx
watsonx.ai
watsonx.governance
watsonx.data

Foundation models

Granite	*IBM*
Open Source	*Hugging Face*
Llama 2	*Meta*
Geospatial	*IBM + NASA*
...	

Data services

Define, organize, manage, and deliver trusted data to train and tune AI models with data fabric services.

Cloud Pak for Data
watsonx Discovery

Hybrid cloud AI tools

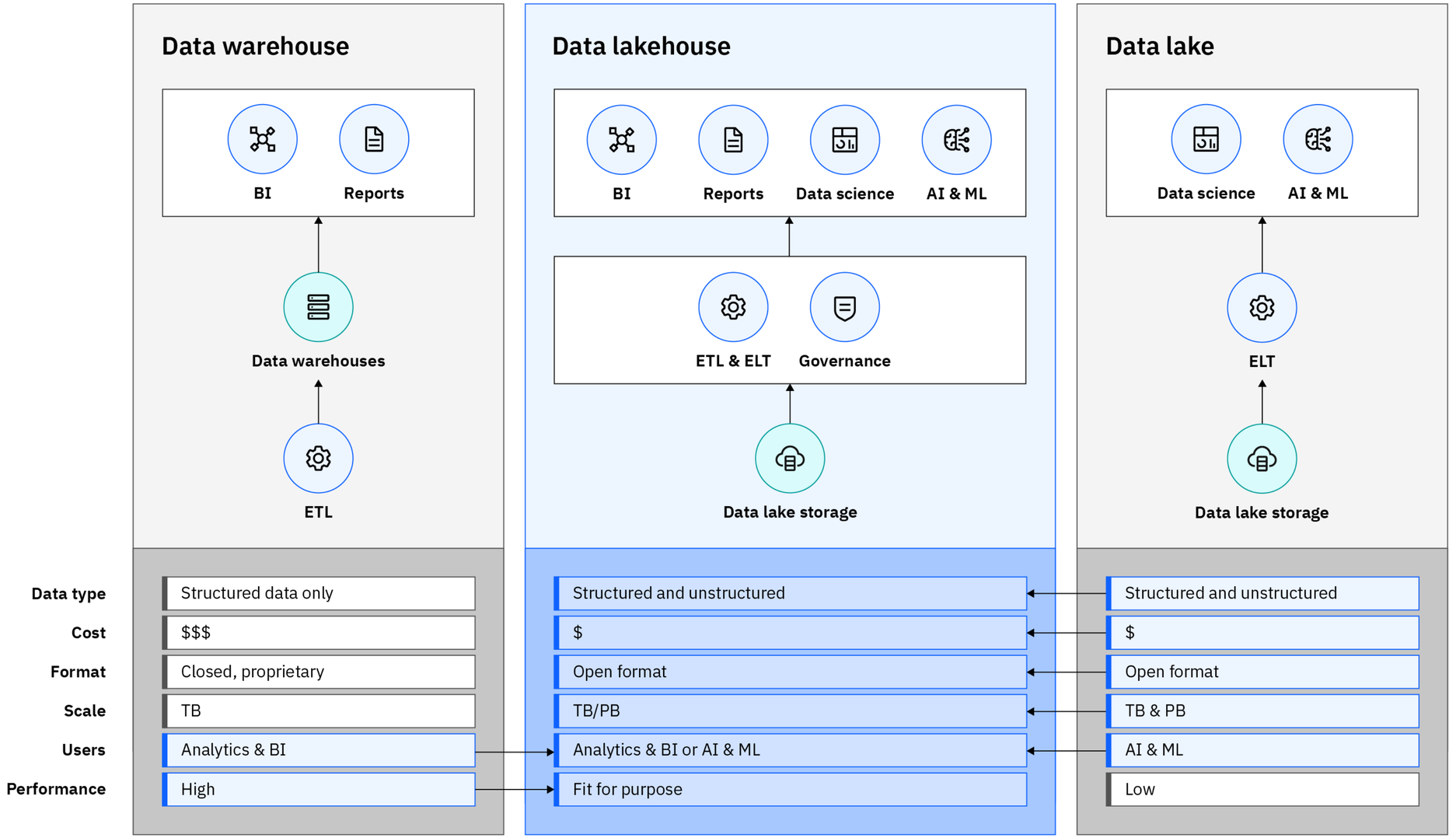
Build on a consistent, scalable, foundation based on open-source technology.

Red Hat OpenShift AI
(*e.g.,* Ray, PyTorch)

Consulting
Generative AI strategy, experience, technology, operations

Ecosystem
System Integrators, Software and SaaS partners, Public Cloud providers

Lakehouses are a new approach meant to combine the advantages of data warehouses and data lakes, but first generation lakehouses still have key constraints



First generation lakehouses are still limited by their ability to address cost and complexity challenges:

- Single query engines set up to support limited workloads –typically just BI or ML
- Typically deployed over cloud only with no support for multi-/hybrid -cloud deployments
- Minimal governance and metadata capabilities to deploy across your entire ecosystem

Investments in an open and trusted data foundation will accelerate and scale your AI initiatives

Databases

Store, unify, and share data with open formats and native integrations

Database
IBM Db2
IBM Informix

Data Warehouse
IBM Db2 Warehouse
IBM Netezza

Data Lakehouse
IBM watsonx.data

3rd party databases
SingleStore, C, mongoDB., PostgreSQL, DATASTAX, snowflake, aws, Any many more

Data Fabric

Simplify data access across the enterprise and hybrid cloud environments with a data fabric architecture

Data Governance
IBM Knowledge Catalog
Manta

Data Integration
IBM DataStage
IBM Data Replication
IBM Data Virtualization

Data as a Product
Delivering data products

Data Observability
IBM Databand

Master Data Management
IBM Master Data Management

IBM Cloud Pak for Data

Data
Delivering quality data to consumers and users

AI
Enabling data teams to be more productive with automated insights

watsonx™

Scale and accelerate the impact of AI with trusted data

AI and Data Platform

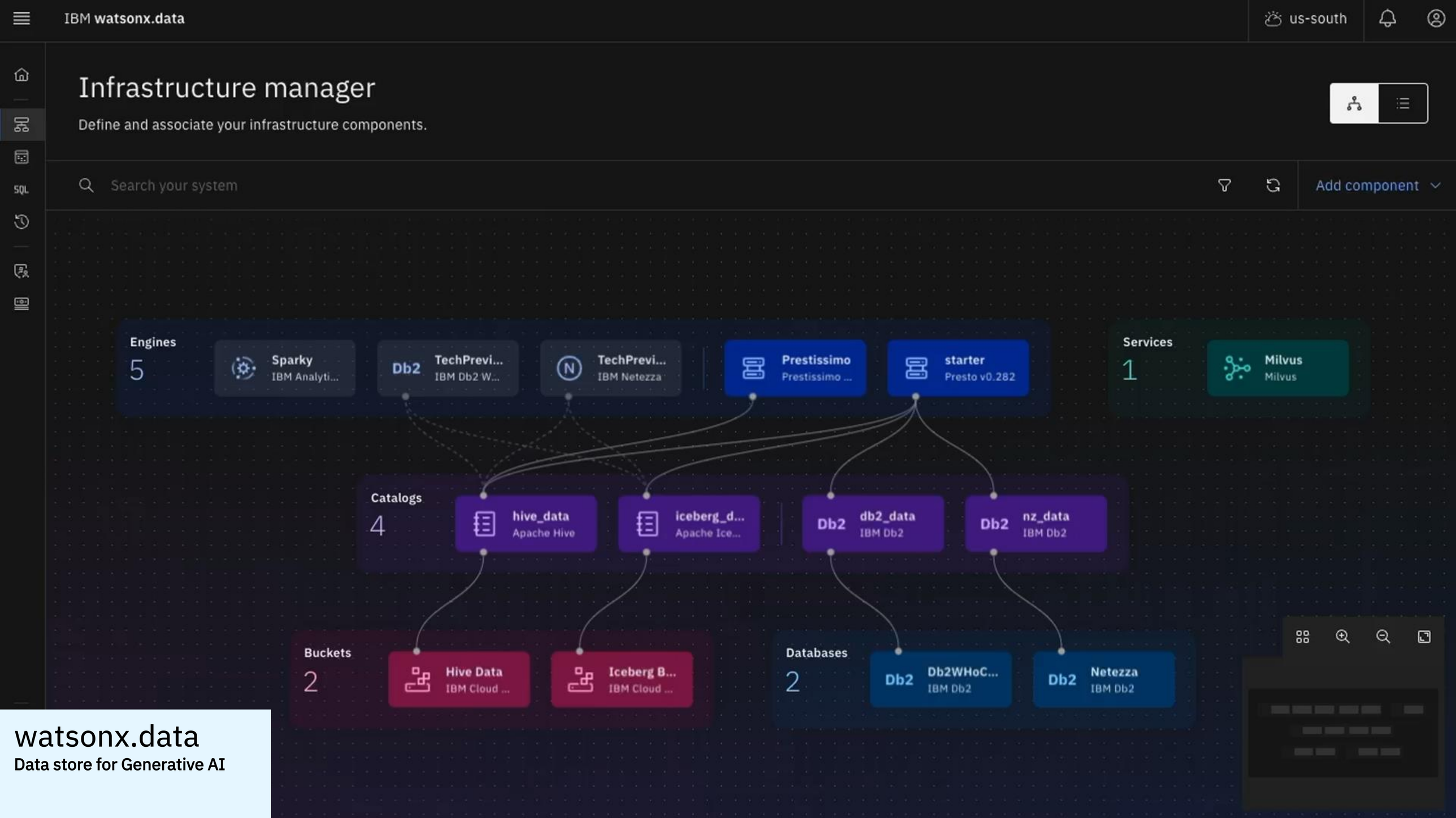
watsonx.ai
watsonx.data
watsonx.governance

AI Assistants

watsonx Orchestrate
watsonx Assistant
watsonx Code Assistant

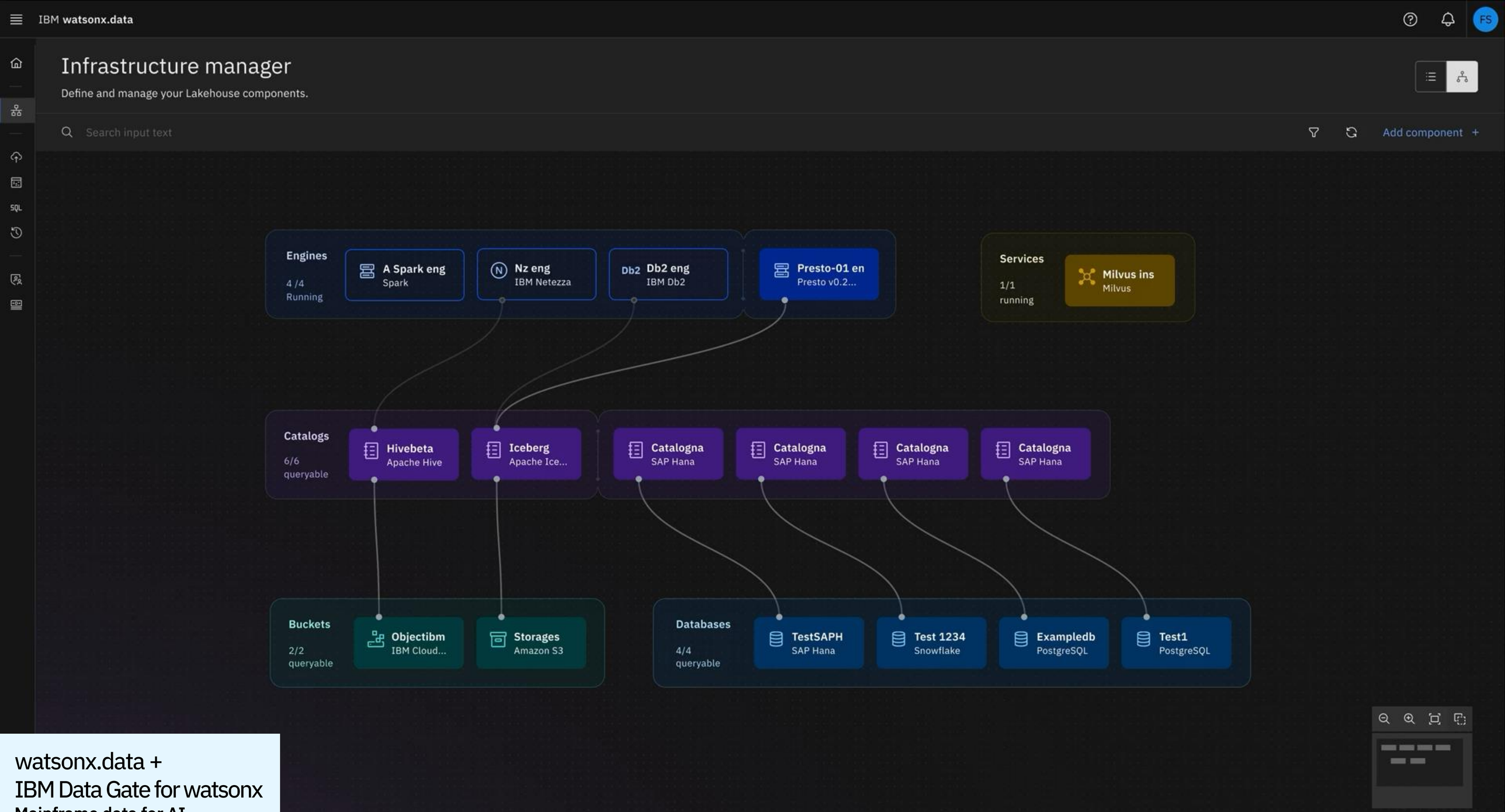
Automated data lineage

Gain deeper visibility into your data and its journey from source to end-use for regulatory compliance and AI use cases with Manta, an IBM company

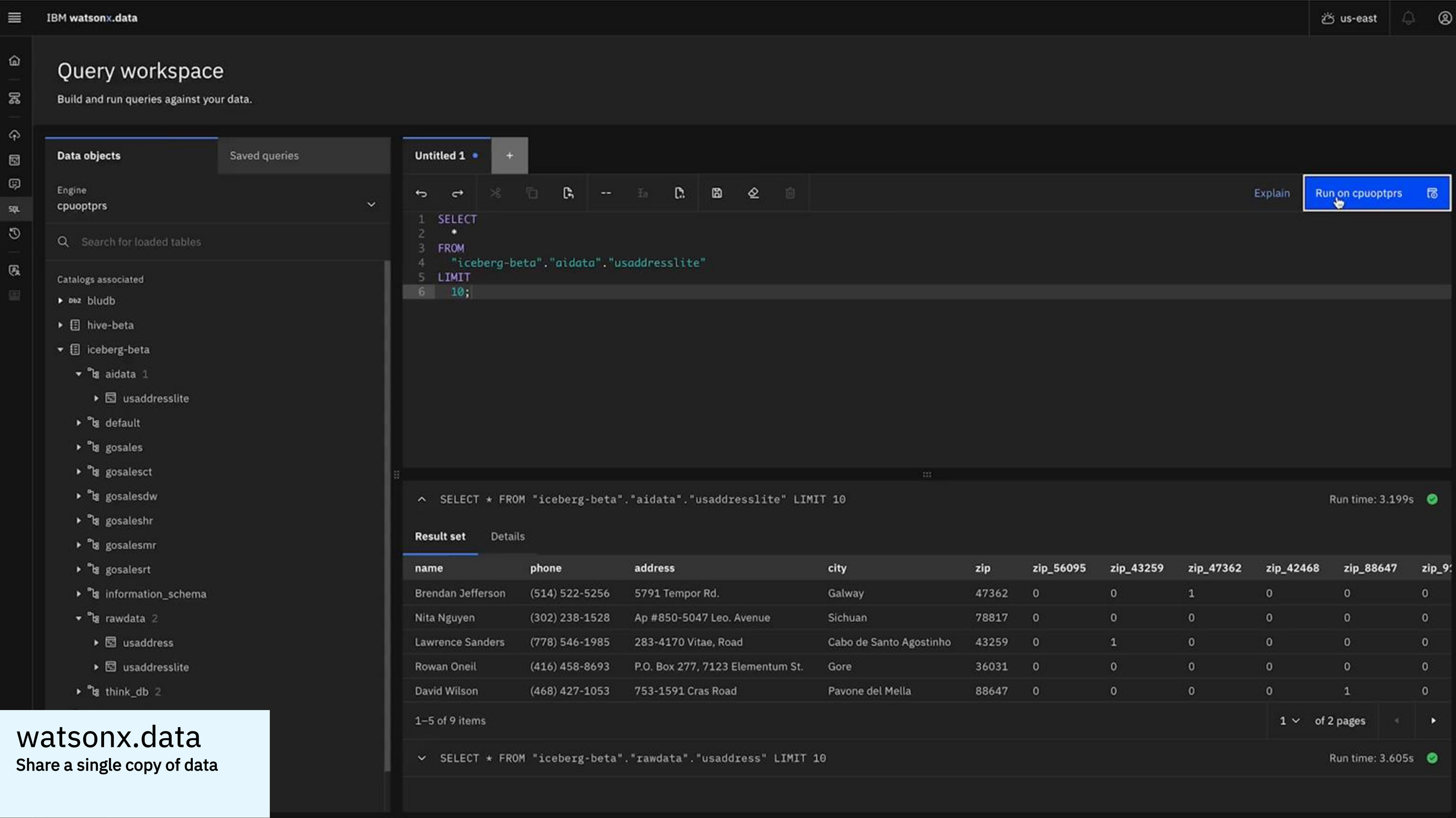


watsonx.data

Data store for Generative AI

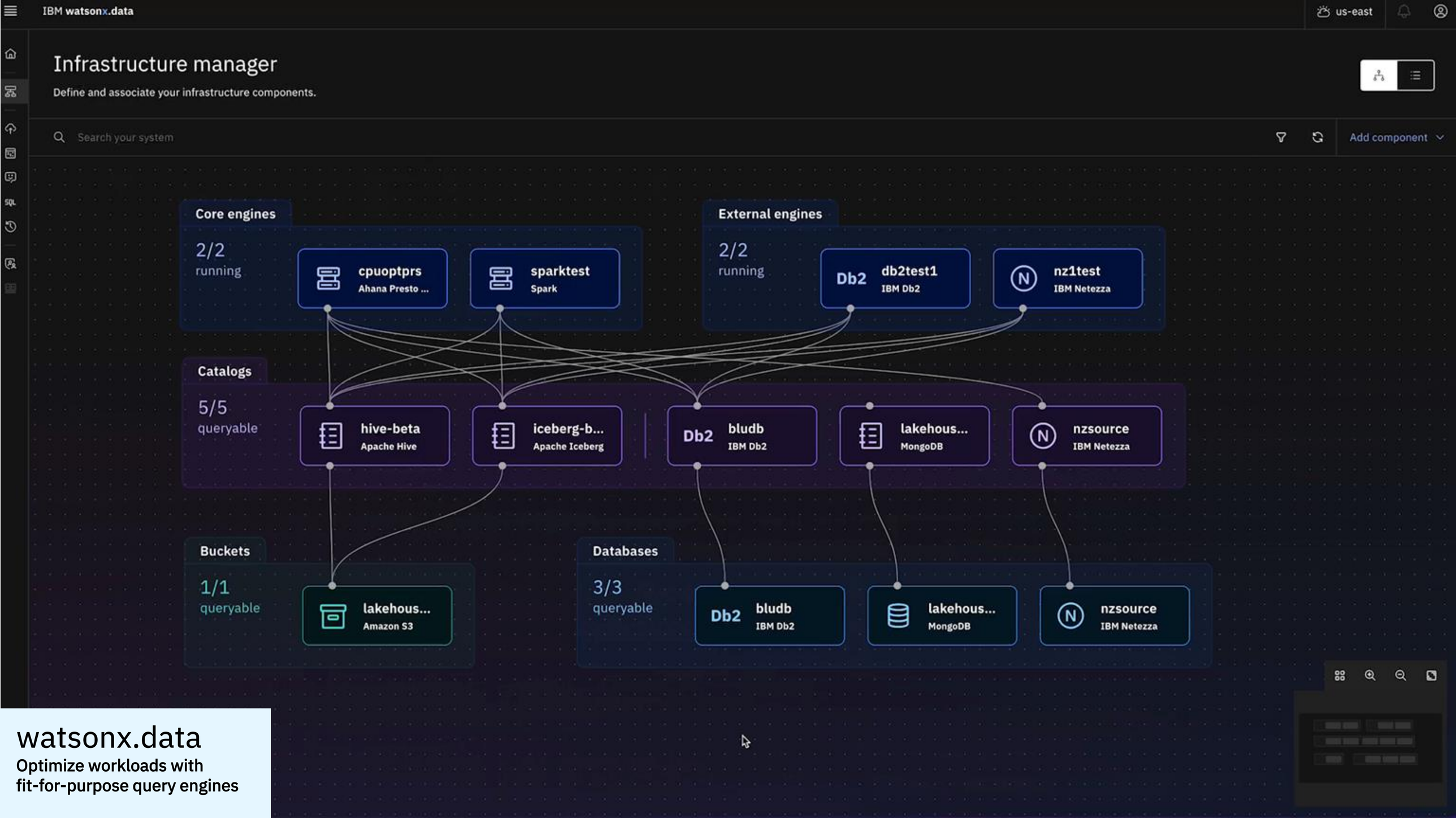


watsonx.data +
IBM Data Gate for watsonx
Mainframe data for AI



watsonx.data

Share a single copy of data



watsonx.data

Optimize workloads with fit-for-purpose query engines