watsonx.data

Scale AI workloads for all your data, anywhere with an open, hybrid, governed data store

Highlights
Connect to data in minutes

Share a single copy of data

Optimize workloads with fit-for-purpose query engines

Prepare and manage data for AI

The global data storage market is expected to more than triple by 2032.[1] Rapidly propagating data on premises and across clouds, applications and locations can result in more data silos, higher costs and added complexities when governing an organization's AI and data workloads.

To address the challenges of today's complex data landscape and scale AI, organizations can combine the high performance and usability of a data warehouse with the flexibility and scalability of data lakes with an open data lakehouse architecture.

The new IBM® watsonx.data™ platform is an open, hybrid and governed data lakehouse optimized for data and AI workloads, helping you derive the greatest value from your data landscape.

Access all your data across hybrid cloud
Connect to storage and analytics environments in minutes and access all your data through a single point of entry with a shared metadata layer across cloud and on-premises environments.

Optimize for price performance
Use watsonx.data to optimize your data workloads with multiple fit-for-purpose engines such as Presto C++, Presto, Spark and integrated data warehouse engines.

Prepare and manage data for AI
Unify, curate and prepare data efficiently for AI models and applications of your choice. Equip your AI with your trusted data.

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

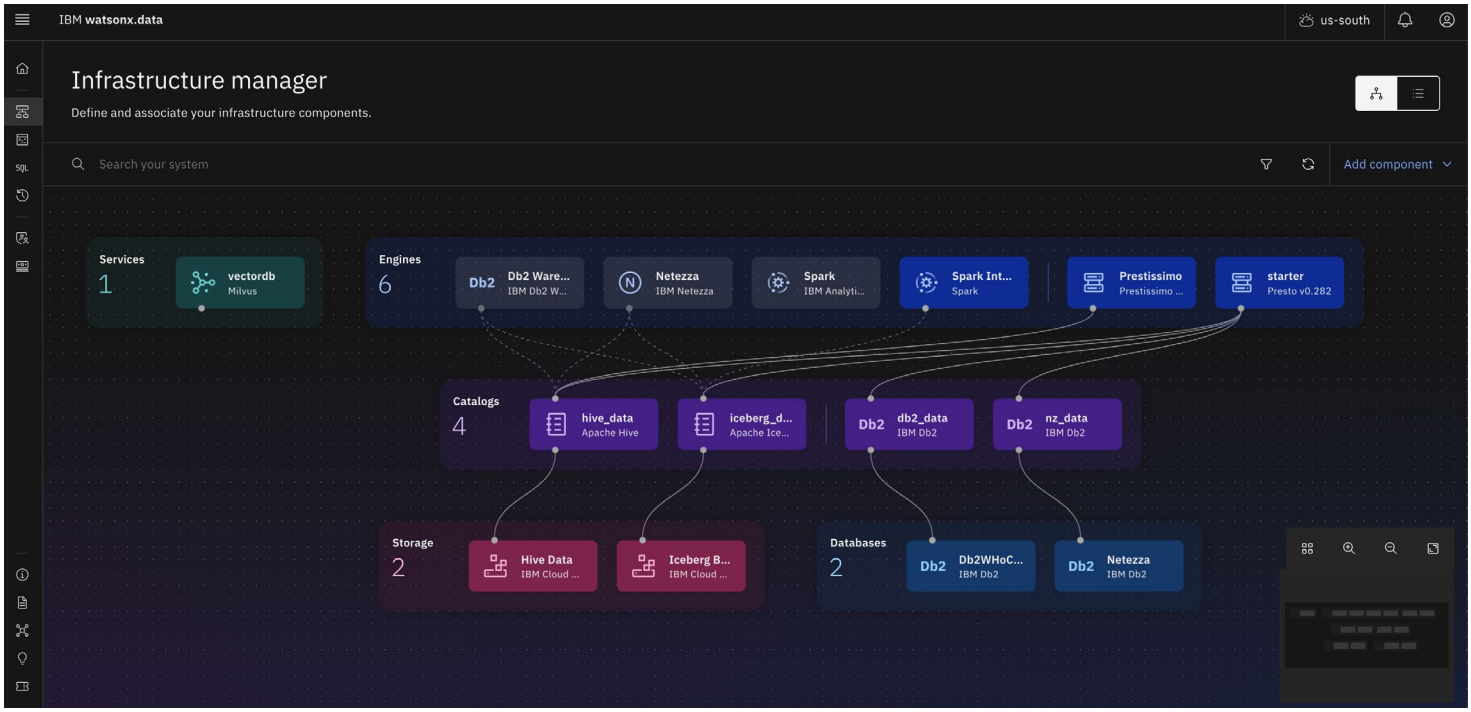

Figure 1. Fit-for-purpose query engines help drive analytics costs down with lower-cost compute and storage as well as fit-for-purpose analytic engines, such as Presto and Spark, that dynamically scale up and down.

Connect to data in minutes
Connect existing data with new data in minutes and unlock new, trusted insights without the cost and complexity of governing, duplicating and moving data. Users can explore and transform data using common SQL.

Watsonx.data also supports integration with a robust ecosystem of IBM and third-party technology to help simplify development and deployment of your analytics workloads and maximize value from existing data investments. Watsonx.data is readily accessible through SaaS on IBM Cloud® and Amazon Web Services or as containerized software. Seamlessly deploy across any cloud or on-premises environment in minutes with workload portability through Red Hat® OpenShift®.

Share a single copy of data
Openness facilitates collaboration. It can also improve data integrity and help address security risks by reducing the number of copies of data required to support different users and tools. And fewer copies means less software, reduced hardware requirements and lower storage costs. With watsonx.data, you can access all of your data across both databases and data lakes. Share large volumes of data through open table formats such as Apache Iceberg, built for high-performance analytics and large-scale data processing. Support multiple vendor open formats for analytic datasets while allowing different engines to access and share the same data at the same time using tools such as Parquet, Avro, Apache ORC and more. Rely on watsonx.data to share metadata among multiple query engines using a single copy of data for all analytics and AI workloads.

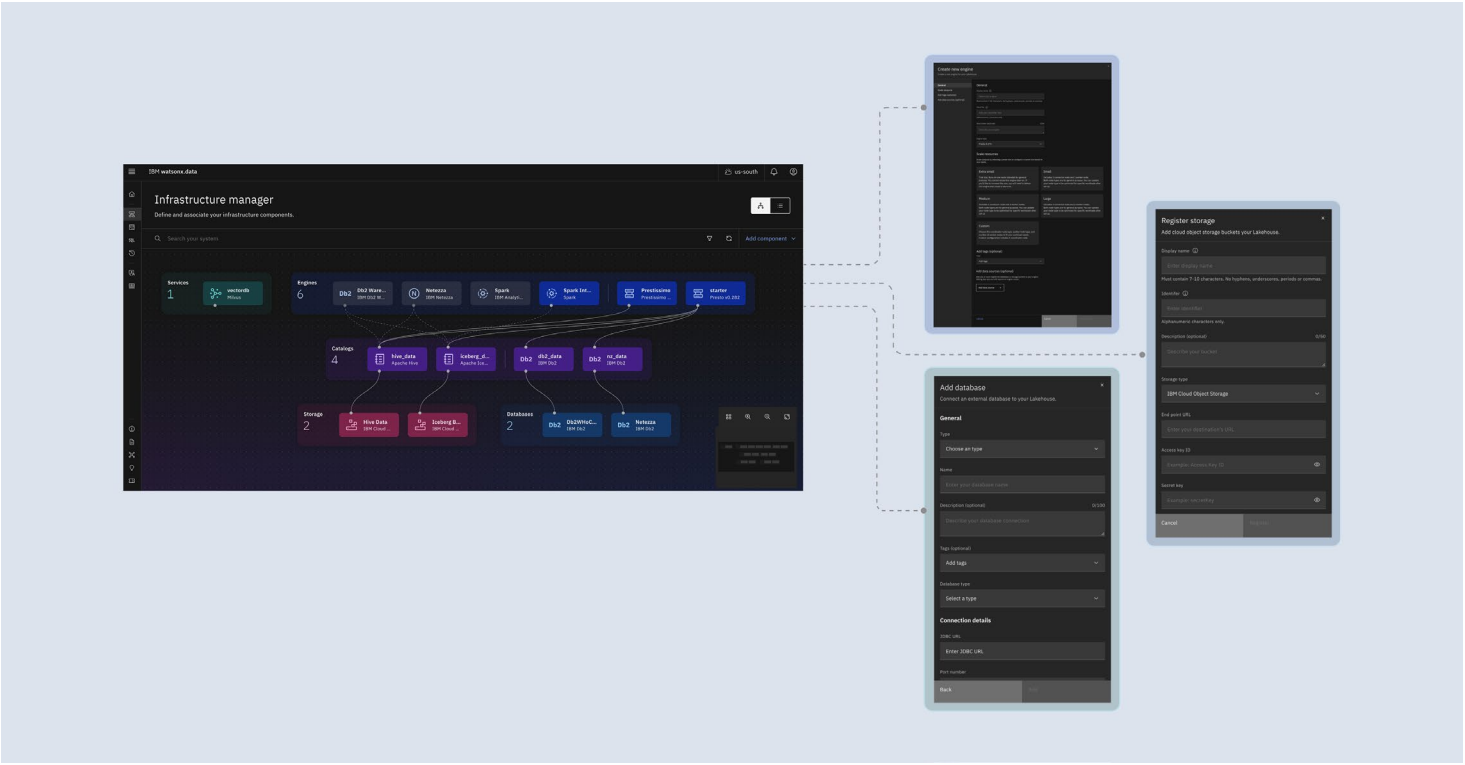

Figure 2. A user-friendly integrated console lets you connect to your existing analytics data and deploy query engines in minutes.

Optimize workloads with fit-for-purpose query engines

No single analytics engine can deliver on the breadth of demands that satisfy all analytics requirements. Fulfilling such a wide variety of analytics requirements requires multiple analytics engines.

You can optimize costly warehouse workloads and help reduce data warehouse costs by up to 50% through workload optimization using cost-effective object storage and fit-for-purpose query engines.[2] These include Presto, optimized for BI workloads, and Spark, optimized for machine learning and data science (ML/DS) workloads—both of which scale up or down automatically as your needs change. With just a few clicks, you can quickly add a new query engine of your choice to meet your price-performance requirements.

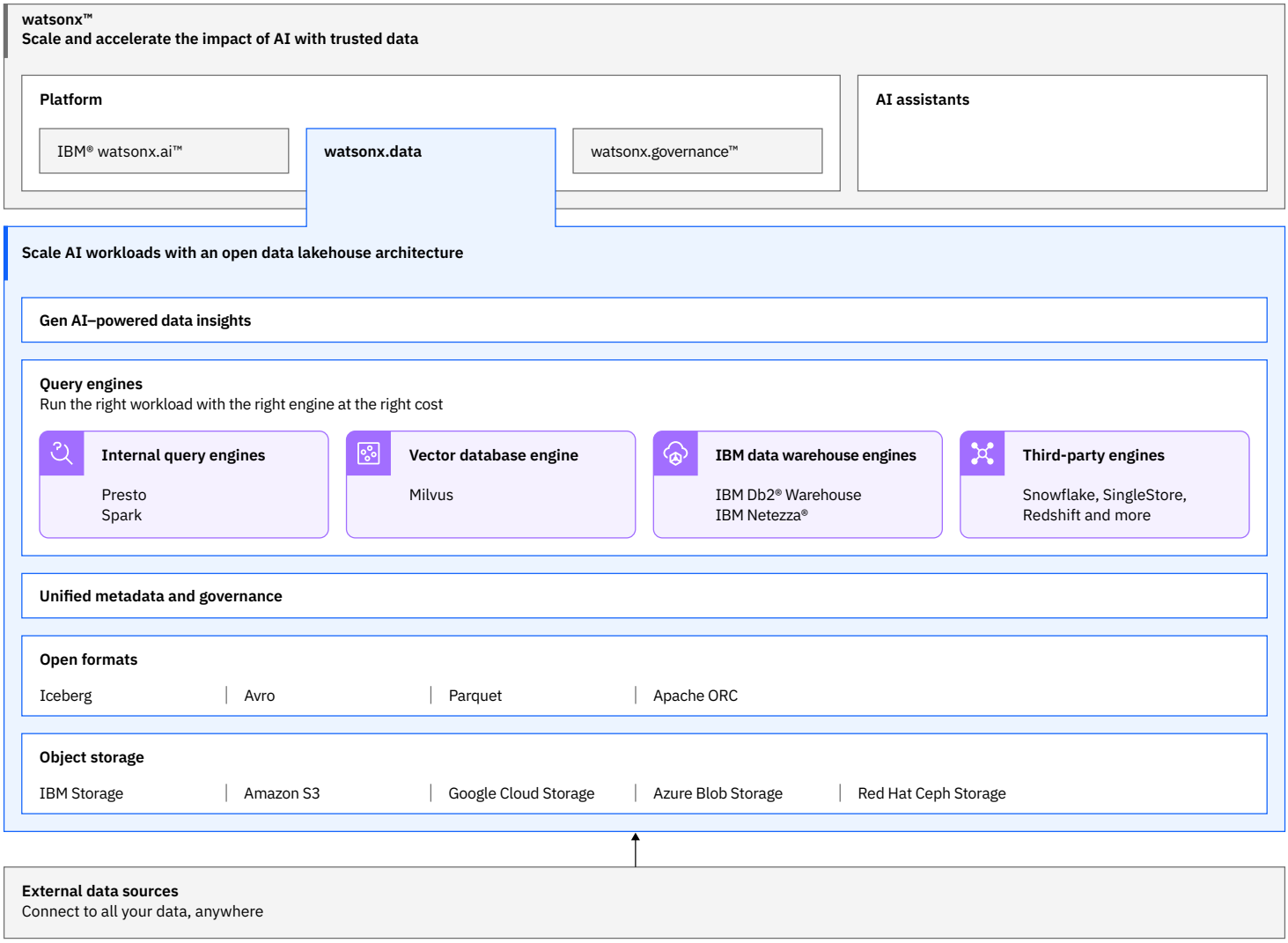

Figure 3. The IBM watsonx.data data lakehouse is optimized for all data and AI workloads.

Prepare and manage data for AI

Trusted, governed data is essential for ensuring the accuracy and relevance of AI applications. One way to prepare data for AI is by creating vectorized embeddings for low-latency queries. This unlocks large volumes of enterprise data for gen AI and retrieval-augmented generation (RAG) use cases at scale. Watsonx.data features embedded Milvus vector databases that let you store and query vectorized embeddings for RAG use cases. This feature helps ground AI applications in trusted data, enhancing the relevance and precision of your outputs.

Conclusion
With watsonx.data, you can access all your data across cloud and on-premises environments. The platform lets you connect to storage and analytics environments in minutes and access all your data through a single point of entry with a shared metadata layer. You can use multiple query engines to optimize analytics and AI workloads for price performance and prepare your data for AI with an integrated vector database. Get greater value from your data investments with an open, hybrid, governed data store that's optimized for all data and AI workloads, and put AI to work.

Why IBM?
IBM is trusted to manage the most mission-critical data and applications for our clients. Our experience with innovation in enterprise data solutions includes market-making database solutions and enterprise-ready AI. We help our clients run solutions in almost any cloud or on-prem environment and believe that our clients' data belongs to them, 100%.

For more information
To learn more about watsonx.data, contact your IBM representative or IBM Business Partner, or visit ibm.com/products/watsonx-data.

Try watsonx.data for free →

Request a live demo →

1. Data Storage Market, Fortune Business Insights, 29 April 2024.
2. When comparing published 2023 list prices normalized for VPC hours of watsonx.data to several major cloud data warehouse vendors. Savings may vary, depending on configurations and workloads.

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
May 2024