IBM watsonx.data: An open, hybrid, governed data store

Scale AI workloads, for all your data, anywhere

Highlights

Choose fit-for-purpose query engines

Apply built-in data governance, security and automation

Share a single copy of data

Connect to data in minutes

Scaling AI requires trusted data, yet most organizations struggle with fundamental data challenges. According to IDC, by 2025 stored data will grow 250%[1] with data rapidly propagating on premises and across clouds, applications and locations, likely with questionable quality. This situation can create more data silos, higher costs and added complexities with governing an organization's AI and data workloads.

How do you combine the high performance and usability of a data warehouse with the flexibility and scalability of data lakes to address the challenges of today's complex data landscape and scale AI? Achieve this by optimizing your AI and analytics workloads and selecting the right engine for the right workload at the right cost-wherever your data resides.

The new IBM® watsonx.data™ platform does just that because it's an open, hybrid and governed data lakehouse optimized for all data and AI workloads. This offering helps organizations drive the greatest value from their analytics ecosystem supported by 3 core benefits.

Access all your data across hybrid cloud
Access all your data through a single point of entry with a shared metadata layer across all clouds and on-prem environments.

Get started in minutes
Connect to storage and analytics environments in minutes and enhance your trust in data with built-in governance, security and automation.

Reduce data warehouse costs by up to 50%[2]
Optimize data warehouses and modernize data lakes by using the right tool for the job such as Presto, Spark, IBM Db2®, Netezza® and others.

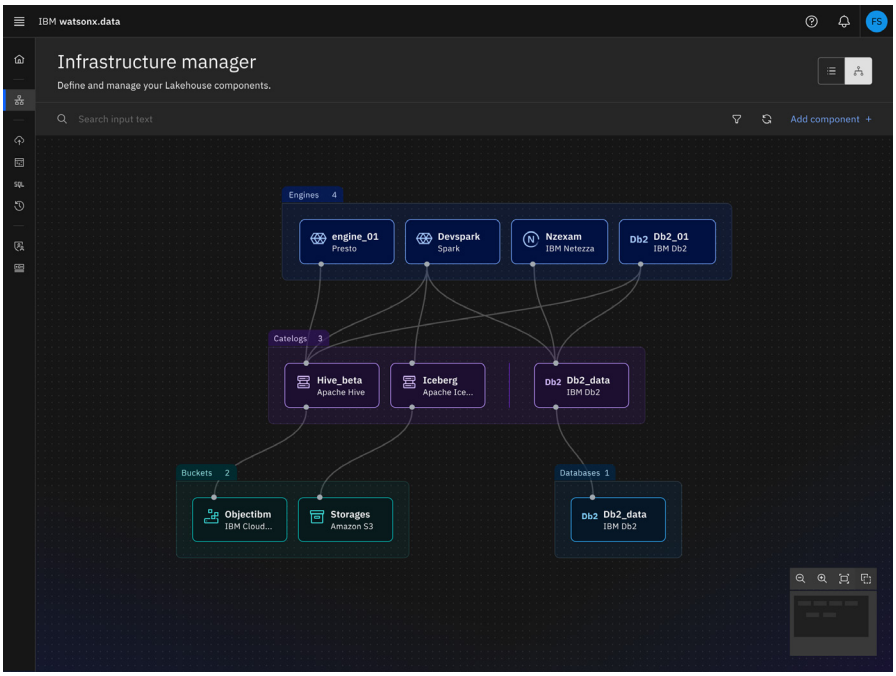

Figure 1. Fit-for-purpose query engines help drive analytics costs down with lower cost compute and storage and fit-for-purpose analytic engines, including Presto and Spark, that dynamically scale up and down.

Choose fit-for-purpose query engines
No single analytics engine can deliver on the breadth of demands that satisfy all analytics requirements. To fulfill such a wide variety of analytics requirements, multiple analytics engines are required.

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, that 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.

Apply built-in data governance, security and automation
Help protect data, manage compliance and maintain trust in data used for AI with built in-governance, access controls and enterprise security in watsonx.data. Integrate with IBM's centralized governance capabilities for automatic policy enforcement and enable responsible, transparent, and explainable data and AI workflows across the enterprise. Finally, you can discover, augment, refine and visualize watsonx.data data and metadata through the power of watsonx.ai models and conversational user experience.

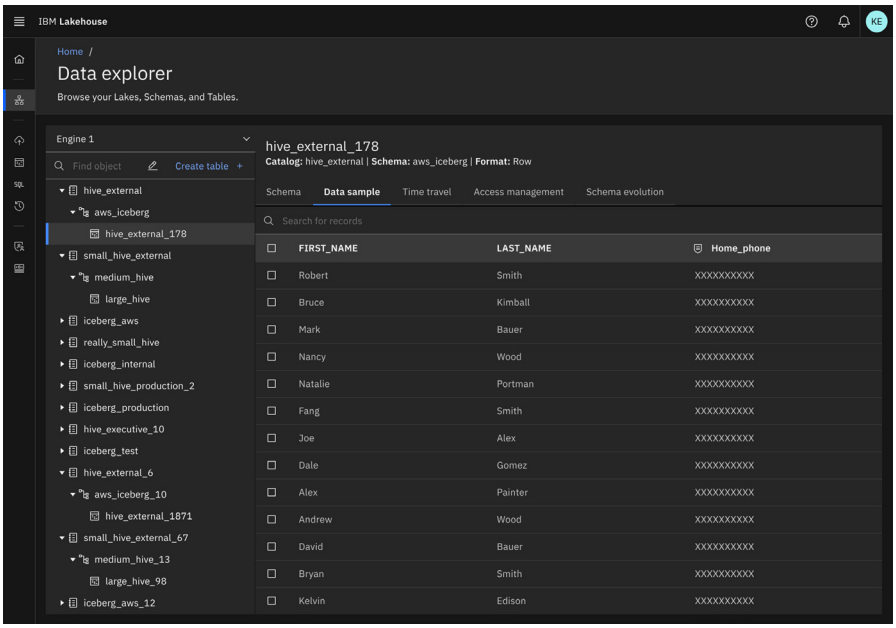

Figure 2. Built-in governance and security help address enterprise compliance and security with unified governance across your data ecosystem.

Share a single copy of data

Openness facilities 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 mean 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 data sets while allowing different engines to access and share the same data, at the same time using tools like Parquet, Avro, Apache Orc and more. Rely on watsonx.data to share metadata between multiple query engines using a single copy of data for all analytics and AI workloads.

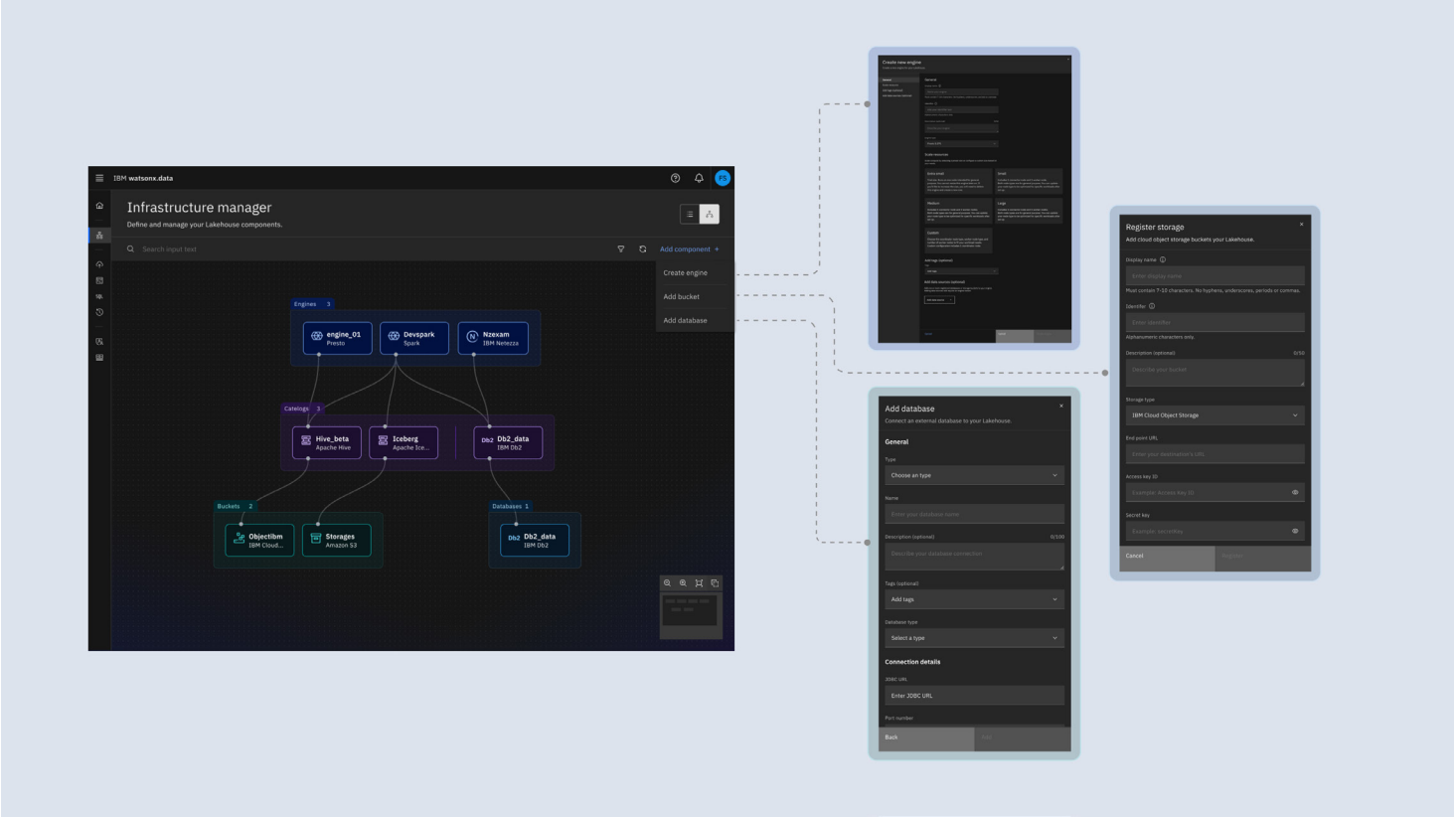

Figure 3. Easy-to-use integrated console lets you connect to your existing analytics data and deploy query engines in minutes.

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.

No matter how you decide to deploy watsonx.data, it's ready to go to work in minutes. It's readily accessible via SaaS on IBM Cloud® and AWS or as containerized software. Your teams can move faster with a simple UX and console to ingest, access and transform data as well as run workloads.

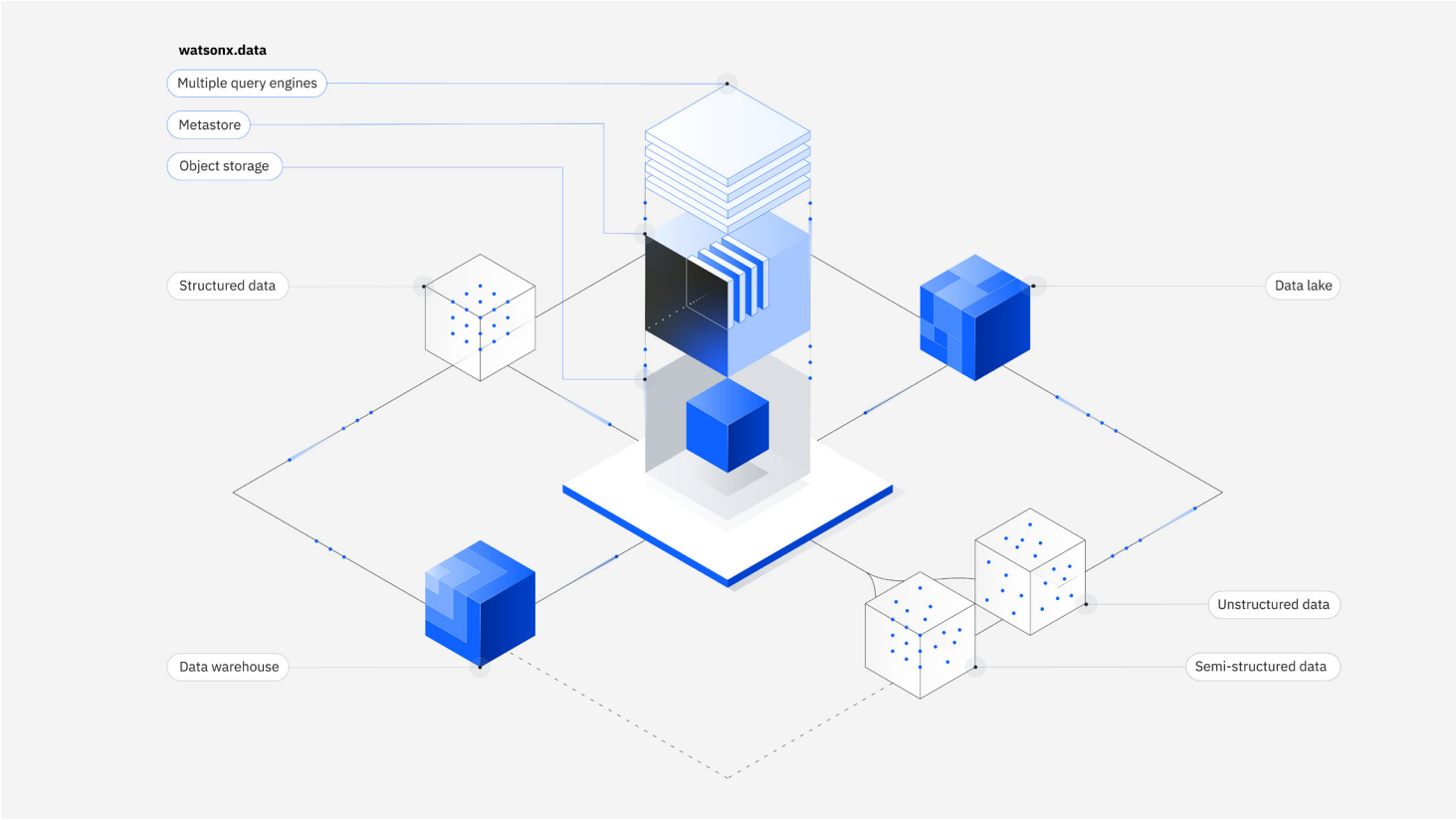

Figure 4. IBM watsonx.data is the only data lakehouse optimized for all data and AI workloads.

Conclusion

Watsonx.data allows you to access all your data across cloud and on-premises environments. It lets you connect to data and get started in minutes with built-in governance, security and automation. Leverage multiple query engines to run analytics and AI workloads, reducing your data warehouse costs by up to 50%.[2] As an open, hybrid, and governed data store optimized for all data and AI workloads, get greater value from your analytics ecosystem and put AI to work with watsonx.data.

Why IBM?

IBM is trusted to manage the most mission-critical data and applications for our clients. Our experience of 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 client's data belongs to them 100%.

For more information

To learn more about watsonx.data, contact your IBM representative or IBM Business Partner.

1. IDC Global DataSphere Forecast 2022-2026
   https://www.idc.com/getdoc.jsp?containerId=US49018922
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, workloads.

To learn more, contact your IBM Business Partner.

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
June 2023