AI that’s ready for business starts with data ready for AI

Build an open and trusted data foundation for scaling and governing generative AI

Highlights

- Modernize your data infrastructure to hybrid cloud
- Power applications, analytics, and AI with the right databases and open data lakehouse
- Automate data quality for AI
- Manage and deliver data for AI
- Build and create value from trusted data with AI Assistants, AI Apps and BI

The quality and quantity of data can make or break AI success, and organizations that effectively harness and manage their data will reap the most benefits. Establishing a robust foundation of high-quality, secure and well-governed data is essential for driving business value across data use cases, including generative AI (genAI), machine learning (ML), analytics and reporting, and data products.

Organizations that build an open and trusted data foundation to access trusted data will be best positioned to unlock the full potential of their data assets to drive meaningful business outcomes, regardless of the specific use case or application. But it’s not so simple. Leaders must re-think prohibitive on-premises approaches and monolithic data ecosystems, all while reducing costs and ensuring proper data governance and self-service access to more data across disparate data sources.

IBM can help enterprises drive more value out of their data, faster – wherever that data resides. More specifically, IBM helps clients to:

- Eliminate data siloes and accelerate data-driven transformation
- Prepare and deliver secure, governed data for AI models and applications
- Unlock faster operational and business outcomes with data across the hybrid cloud

This solution brief will walk through each of the components of building an open and trusted data foundation, and how IBM can help realize the full potential of your data.

95% of new digital initiatives will be developed on cloud-native platforms

Components of building an open and trusted data foundation for AI:

Modernize your data infrastructure to hybrid cloud

Gartner predicts that 95%[1] of new digital initiatives will be developed on cloud-native platforms, which are essential for AI technologies that require massive data storage and scalability. IBM helps organizations scale applications, analytics, and AI across any cloud and on-premises environment, with workload portability through Red Hat® OpenShift® and strong partnerships with AWS and Microsoft for fully managed deployments.

Power data-driven applications, analytics and AI with the right databases and open data lakehouse

For storing and analyzing data, you need to use the right database for the right workload, data types, and for the right price-performance, that ensures you have a data foundation that grows with your data needs, wherever your data resides. Your data strategy should incorporate databases designed for open and integrated components, which allows for seamless unification and access to data for advanced analytics and AI applications within a data platform.

Our database portfolio supports multiple data types, open standards like Apache Iceberg, and native integrations with watsonx and IBM Data Fabric. We address various workload and price-performance needs with the right engines and offer database deployments across any cloud, including IBM Cloud, AWS, Azure and Google Cloud. For transactional workloads, tens of thousands of customers rely on IBM Db2 to run high performance, AI-optimized workloads across the hybrid cloud, helping to modernize legacy apps, build new cloud-native applications, and power AI Assistants. Customers like the Masters Tournament rely on Db2 fully managed on Amazon Relational Database Service (RDS) to store and process years of player and hole data insights, helping to power AI-narration and commentary.

“Data readiness for AI is not something you can build once and for all nor that you can build ahead of time for all your data. It is a process and a practice based on availability of metadata to align, qualify and govern the data.”

- Gartner[2]

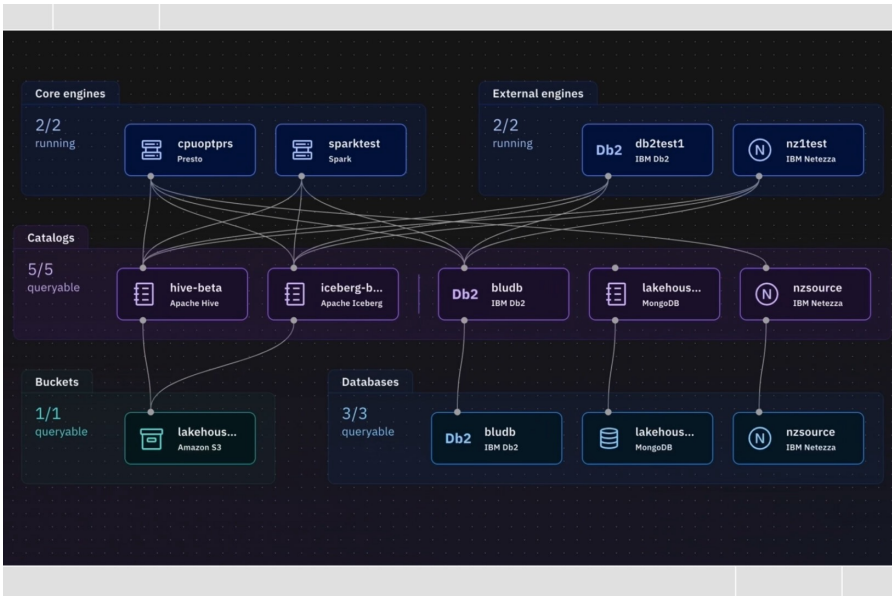

Figure 1: IBM® watsonx.data is the open, hybrid data lakehouse to power AI and analytics with all your data, anywhere it resides.

IBM Informix is designed for seamless embeddability and high-performance throughput in applications at the edge and beyond. We offer purpose-built cloud databases such as document, in-memory, key-value, full-text search and messaging for specific applications and third-party partner databases like Cloudera, SingleStore, MongoDB, EDB and DataStax where we consolidate your servicing, licensing and support, and provide indemnification of open-source solutions.

However, without a way to unify and analyze this data, organizations face stale and inconsistent data for insights. Rapidly propagating data on-premises and across clouds, and on various applications and locations can result in more data silos, higher costs and added complexities.

IBM® watsonx.data is the open, hybrid data lakehouse that helps power AI and analytics with all your data, anywhere it resides. Access and share a single copy of data across your organization, breaking down data siloes and reducing data duplication. Watsonx.data is built with open standards for interoperability with your existing data estate, without needing to migrate or recatalog – including native integrations across your databases such as Db2, Netezza, Informix, MongoDB, SingleStore, Cloudera, MySQL, PostgreSQL and more.

With watsonx.data you can connect to your storage and analytics environments across hybrid cloud in minutes and optimize AI and analytics workloads for superior price-performance with multiple, fit for purpose query engines such as Presto C++, Presto, Spark and integrated data warehouse engines, like Db2 Warehouse and Netezza, for powering real-time analytics and BI. Db2 Warehouse and Netezza support Apache Iceberg to enable sharing of data seamlessly with the lakehouse and to optimize workloads for price performance. Watsonx.data is seamlessly deployable across any cloud or on-premises environment.

Some of the most important data for supporting analytics and building predictive AI models is derived from the critical transaction data that often originates on IBM Z in Db2 for z/OS, IMS for z/OS and native VSAM data sets (Virtual Sequential Access Method).

80%

of enterprises are using or intending to use AI models

250%

growth in data volumes by 2025

Traditional approaches to providing access to this data include writing complex mainframe programs, ETL processes and using replication technologies. IBM Data Gate for watsonx, revolutionizes the way organizations synchronize, analyze and build AI models from the critical transactional data originating on IBM Z®. By bringing transactional data originating on the mainframe into an open, governed data lakehouse like watsonx.data, enterprises can build AI models to help grow revenue, enhance productivity and manage cost.

Automate data quality for AI

According to IDC, by 2025, stored data will grow 250%[3] across on-prem and cloud storages. In today's complex data landscape, ensuring data quality is a significant hurdle for organizations seeking to harness the power of data and leverage innovative AI technologies. The consequences of poor data quality are severe, including erosion of customer trust, non-compliance with regulatory requirements and reputational losses. To mitigate these risks, it is essential to embrace a data architecture strategy that enables effective data quality management.

IBM Data Fabric helps organizations simplify curation of high-quality data by monitoring the quality of data pipelines and sources to detect and resolve issues before they affect downstream operations. Founded on a secure, trusted and governed data architecture, IBM's data quality solutions empower organizations to uncover critical data elements (CDEs) and automates enforcement of data quality rules and service level agreements (SLAs).

Augmented metadata enrichment capabilities included within IBM Knowledge Catalog -powered by large language models (LLMs) from IBM Research - provide a robust foundation for the effective application of data quality rules. By leveraging gen AI capabilities, you can accelerate the time it takes to understand, discover, augment and cleanse data. When embedded into watsonx.data, the semantic layer generates data enrichments that enable clients to find and understand previously cryptic, structured data across their data estate in natural language through semantic search. You can rapidly develop data quality rules and ensure the integrity of CDEs by enforcing data quality SLA rules. With IBM Databand, you can observe, monitor, and identify data quality issues closer to the source, to mitigate the impact on downstream processes and ensure data quality issues are addressed.

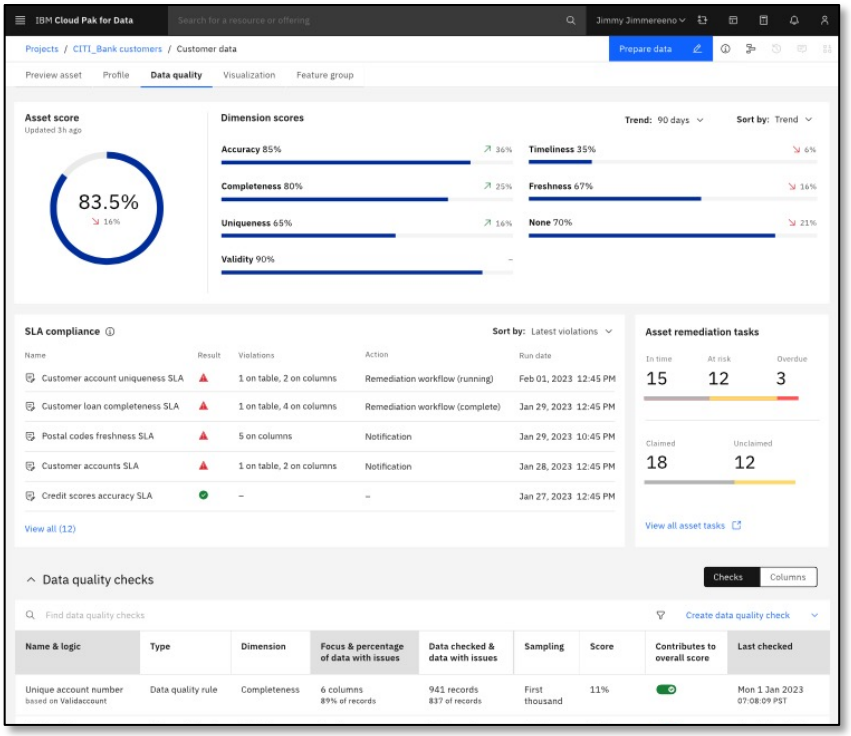

Figure 2: Gain transparency into the journey of your data with automated data lineage capabilities from IBM Manta Data Lineage.

90%

reduction in time to implement new use cases by organizations that manage data as a product [5]

Manage and deliver data for AI

According to Gartner, 80%[4] of enterprises are either using or intending to use foundation models and embrace gen AI. Against this backdrop, organizations are elevating the importance of data as they rush to harness its potential to tailor and enhance the performance of gen AI models faster than their competitors, despite fundamental challenges with data access, volumes, quality.

With IBM watsonx.data's open data lakehouse approach, you can access, curate and prepare data efficiently for AI models and applications at scale, wherever that data resides. Watsonx.data's embedded Milvus vector database enables you to store and query vectorized embeddings for RAG use cases to ground AI applications in trusted data, enhancing the relevance and precision of your outputs. Users can seamlessly connect to trusted data in watsonx.data from IBM® watsonx.ai™ or another AI tool

Once you have unified access to high quality, governed data for AI, cost-optimized your growing data workloads, and prepared data for AI, the next step is to scale the delivery of the right data for AI use cases in your organization. With IBM Data Product Hub, you can streamline the creation and delivery of business-ready data in a reusable manner, enabling self-service data access with governance. This eliminates the need for repeated, manual data request workflows, allowing data producers and consumers to efficiently share data products, accelerating the time to value of your AI initiatives.

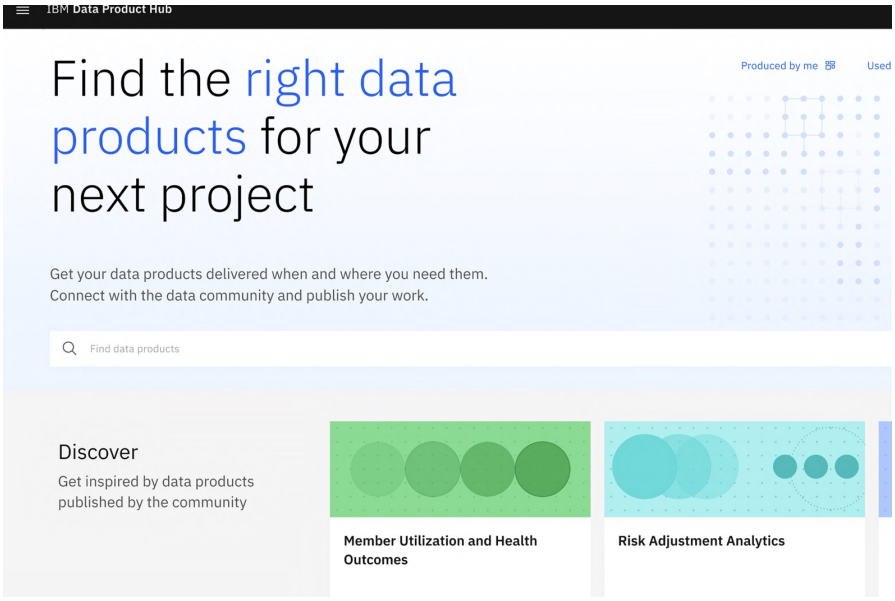

Figure 3: IBM Data Product Hub helps streamline data sharing and automates the delivery of data products to data consumers across the organization

To accelerate value realization, we use a range of pre-built methods, assets, and accelerators that address all three critical dimensions of transformation: technology, people, and process.

Build and Create Value From Trusted Data with AI Assistants, AI Applications, and Business Intelligence

With an open and trusted data foundation in place, you can unlock the full potential of your data and create value from it. This can be achieved through building AI assistants, AI applications and business intelligence solutions. Build, train, tune, and deploy ML models or fine tune LLMs for new gen AI applications with your own trusted data using our IBM watsonx AI and data platform. Enhance your employee and customer experience with watsonx AI assistants such as watsonx Assistant for customer care experiences, watsonx Orchestrate for digital labor and HR automation and watsonx Code Assistant for IT automation. Help business users make better decisions by providing insights, recommendations and predictions with watsonx BI Assistant. No matter your goals, with the right data ready for analytics and AI, you can create a data-driven organization that drives business value and innovation.

Getting Started

Scale data and AI with technology, people and processes

IBM Consulting's data transformation services can help leverage deep expertise in complex data landscapes to design and deliver a modern data foundation on any cloud, tailored to your organization's specific requirements. This is underpinned with your strategic objectives for data which has likely changed because of generative AI. To accelerate value realization, we use a range of pre-built methods, assets and accelerators that address all three critical dimensions of transformation: technology, people and process.

Meet with IBM Consulting's data services experts for a personalized session on how to design and build out your ideal data estate, regardless of existing technology investments.

For more information
To learn more about IBM data and AI solutions, contact your IBM representative or IBM Business Partner, or visit our website .

1. Strategic Roadmap for Migrating Data Management Solutions to the Cloud, Gartner, September 2023
2. Gartner, Quick Answer: What Makes Data AI-Ready?, 15 May 2024
3.IDC, Worldwide IDC Global DataSphere Forecast, 2023–2027
4. Gartner, Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026, 11 Oct 2023
5. Veeral Desai, Tim Fountaine, Kayvaun Rowshankish, A Better Way to Put Your Data to Work, Harvard Business Review, July 2022
6. The Masters, 2024

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
June 2024