IBM Cloud Pak for Data

Intelligently automate your data and AI strategy to connect the right data, to the right people, at the right time, from anywhere.

Contents

Introduction

In today's uncertain environment every organization must become smarter and more responsive in order to operate with intelligence and respond to market changes with flexibility and resilience. Fueled by data, AI is empowering leading organizations to transform and deliver value. A recent study found that data-driven organizations are 178% more likely to outperform competitors in terms of revenue and profitability.[1]

However, to successfully scale AI throughout your business, you must overcome data complexity. Companies struggle to manage and maintain vast amounts of data spanning public, private and on-premises clouds. 75% of global respondents stated their company is pulling from over 20 different data sources to inform their AI, BI, and analytics systems. In addition, one third cited data complexity and data siloes as top barriers to AI adoption. Further compounding the complexity of these fragmented data landscapes is the fact that the lifespan of that data—the time that it is most relevant and valuable—is shrinking.[1]

The solution? An agile and resilient cloud-native platform that enables clients to predict and automate outcomes with trusted data and AI.

IBM Cloud Pak for Data: Any data. Any cloud. Anywhere.

IBM Cloud Pak® for Data is a fully integrated data and AI platform that enables organizations to accelerate AI-powered transformation by unleashing productivity and reducing complexity. Collect, organize and analyze data; then infuse AI throughout your business within a collaborative platform experience. Cloud-native by design, IBM Cloud Pak for Data is built on and takes advantage of the underlying resource and infrastructure optimization and management in the Red Hat® OpenShift® Container Platform. The solution can be deployed on any cloud and fully supports multicloud environments such as AWS, Azure, Google Cloud Platform, IBM Cloud® and private cloud deployments. Its key integrated features span the entire analytics lifecycle, from data management and DataOps to business analytics and AI.

Key benefits include:

- **Single, unified platform**
  Bring data management, data governance, data science and AI capabilities together on an intuitive integrated platform based on your needs.
- **Built-in governance**
  Use automated end-to-end governance to help enforce policies and rules across your organization and quickly respond to changing regulations.
- **Extensible and customizable**
  Flexibly deploy data and AI services from a growing catalog of proprietary, 3rd party and open source services to build the platform that best suits your needs.
- **Pre-built AI and industry applications**
  Innovate at speed thanks to industry solutions for IT operations, customer service, risk and compliance, and financial operations.
- **Designed for hybrid cloud**
  Deploy the platform in almost any environment, whether on-premises or on the cloud, due to its cloud-native design and Red Hat OpenShift foundations.

The latest version of IBM Cloud Pak for Data further infuses intelligent automation throughout the platform with new AI-powered capabilities that are the core components of a new data fabric architecture within the platform.

This data fabric automates complex data management tasks and enables you to universally discover, integrate, catalog, secure and govern data across multiple environments, providing a trusted common data foundation for data science and AI.

What is a data fabric?

In the past, organizations have attempted to address data access problems either through point-to-point integration or introduction of data hubs. Neither of those are suitable when data is highly distributed and siloed. Point-to-point integrations add exponential cost for any additional end point that needs to be connected, making it difficult to scale this approach. The data fabric is an emerging architecture that aims to address the data challenges arising out of a hybrid data landscape. Its fundamental idea is to strike a balance between decentralization and globalization by acting as the virtual connective tissue between data endpoints.

Through technologies such as automation and augmentation of integration, federated governance as well as activation of metadata, a data fabric architecture enables dynamic and intelligent data orchestration across a distributed landscape, creating a network of instantly available information to power a business. A data fabric is agnostic to deployment platforms, data processes, geographical locations and architectural approach. It facilitates the use of data as an enterprise asset. A data fabric ensures your various kinds of data can be successfully combined, accessed and governed.

Cornerstones of IBM's data fabric:

- **Seamless data access and orchestration:** Unlock siloed data at scale by automating how you access, update and unify data spread across distributed stores and clouds, with a solution optimized for minimal data movement and high automation through intelligent orchestration.
- **Intelligent data catalog:** Automate the discovery, linking and semantic enrichment of your metadata to provide your data consumers with self-service access to trusted, business-ready data from across your enterprise.
- **Pervasive policy-based data privacy:** Automate how you enforce universal data and usage policies across your data ecosystems in a hybrid cloud landscape to reduce risk while enabling data use.

Learn more about the benefits of the data fabric architecture within IBM Cloud Pak for Data by reading the detailed paper.

Top use cases for IBM Cloud Pak for Data

Optimize data access and availability

Unify and simplify access to all your data, on any cloud, anywhere
Make data available to any data user with a universal query engine that intelligently works across any cloud, warehouse, data lake, database or any open file formats without moving, replicating, migrating or creating new copies. Minimize the complexity of multiple query engines with a distributed query engine and virtualized access to facilitate total data utilization and reduced data engineering workloads. Minimize resource-heavy data migration processes and expensive data repositories by creating virtualized data objects, intelligently optimized to perform at a petabyte scale. With modernized data access, you can:

- Work across disparate data types, structures, volumes, velocities and locations, leveraging an intelligent universal query engine to provide ease of access, and faster discovery of insights from all data.
- Monitor query performance information over time to correlate with algorithms that automatically create machine learning models.
- Optimize access paths for faster query execution and reduced resource consumption, yielding significant performance improvements.
- Apply the same enterprise data quality and governance policies to data accessed virtually. Utilize a common governance catalog to provide consistent enterprise data quality measures needed for audit or regulatory compliance inquiries across all data, anywhere.

Deliver quality governed data

Connect the right data to the right people at the right time
Deliver high-quality, governed, secure, business-ready data to the right people at the right time across the enterprise to drive business outcomes at speed and scale. Understand what data you have and where you have it with AI-powered data discovery and profiling to set governing rules around data quality, define business taxonomy, access rights, privacy and protection to trust your data. Onboard discovered data rapidly with AI-powered cataloging, and make it available to users across the enterprise with a graph-like intuitive search, encouraging self-service use. Automate virtual or physical data integration needs automatically based on data usage patterns, reducing data engineering workloads. Evolve from using pointed tools to an AI-powered, integrated, modular and re-usable data and AI platform that connects the right data to the right people when they need and where they need it across a hybrid landscape, so you can:

- Automatically create comprehensive and dynamic metadata knowledge about all data you have with AI-powered auto-discovery and auto-cataloging. Understand what data you have, where you have it, and what controls are needed. Make data understood in business terms to enable self-service consumption, empowering data consumers to derive value from data anywhere across the enterprise.
- Enable intelligent data integration through automated data engineering and integration. Use AI augmentation to decide what integration approach is best based on workloads, data policies and geographical data access rules, so that you can accelerate delivery of business-ready data for the enterprise.
- Automate governance with active metadata, define policies for privacy and security, taking geographical and global legislation such as GDPR into account. Understand the data format and significance to apply the correct policies to data, and each prospective user, to facilitate automated data policy enforcement at a granular level. Capture technical and business lineage to easily respond to compliance, audit, and privacy requests.

Minimize risk and ensure compliance

Empower your organization with a pervasive data privacy framework for hybrid multicloud
Discover, understand and manage the sensitive or high-risk data that exists throughout your organization, with a unified privacy framework that enables risk mitigation to protect your brand reputation and preserve customer trust. Deliver a real-time view of sensitive data and AI assets such as PII or AI models across hybrid multicloud environments, and enforce protective policies automatically. See who has access to high-risk data or AI artifacts, and what outcomes are impacted, helping business leaders and auditors take corrective actions. Provide self- service access of the right data to the right data consumers, without sacrificing security or compliance, with an end-to-end data privacy framework and solutions that helps you:

- Discover high-risk data or assets to eliminate compliance blind spots and minimize risk, spanning the entire data and analytics landscape so that you avoid compliance cost overruns, shield your brand from competitive attacks or disputes, and protect customer trust.
- Provide data consumers with self-service access to a trusted foundation of quality data on which risks are proactively managed to accelerate insights and innovation.
- Turn your security, compliance and data governance teams into strategic partners through a collaborative platform to stay collectively confident during challenging audit or regulatory inquiries.
- Empower everyone on your team to become a risk expert with the help of intelligent risk identification, remediation, utilizing automation and AI throughout.
- Deliver the flexibility IT teams demand with a true cloud-agnostic hybrid multicloud platform that can be deployed anywhere.

Deepen customer engagement

Drive customer satisfaction by leveraging insights from a comprehensive client view.
Enable your enterprise to create rich customer profiles and extract key insights that deliver products and services that delight your customers. With Customer 360, organizations can connect data on a customer from disparate sources across the data landscape. The resulting insights—derived from that 360º view of the customer—can improve customer satisfaction and reduce churn.

- Compile a customized and trusted customer view.
- Enable workflow capabilities to implement policies and processes for governance.
- Operationalize master data for deeper analysis and enhanced insights.
- Manage your customer relationships and hierarchies and get more accurate reporting.

Deployment models for IBM Cloud Pak for Data

Since the release of IBM Cloud Pak for Data more than three years ago, IBM has continued to advance new features and additional deployment and consumption models, including:

- IBM Cloud Pak for Data: A client-managed software platform that runs on any cloud. The details of this solution brief highlight the core components of this deployment model.
- IBM Cloud Pak for Data System with Netezza® Performance Server: A preconfigured, hyper-converged system that combines storage, compute, networking and software, and reduces private cloud deployment times to a matter of hours, allowing teams to provision and deploy data services flexibly and rapidly to specific needs.
- IBM Cloud Pak for Data as a Service: A "pay-as-you go" subscription model for a set of integrated IBM Cloud Pak for Data services, fully managed on IBM Cloud infrastructure. IBM Cloud Pak for Data as a Service eliminates underlying IT management challenges and helps organizations quickly scale the tools and processes needed for enterprise AI in the cloud. Integrated with IBM Cloud Satellite™ services, IBM Cloud Pak for Data as a Service can run across distributed cloud environments.

Next steps

Take the next steps to learn more about IBM Cloud Pak for Data.

Read the data fabric white paper

Sign up for the free trial

Visit the platform page

IBM