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 to intelligently operate and engage the world, resiliently respond to market changes, flexibly optimize costs, and innovate. 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 in terms of revenue and profitability.

However, to successfully scale AI throughout your business, you must overcome data complexity. Companies today are struggling to manage and maintain vast amounts of data spanning public, private and on-premises clouds. 70 percent 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 hybrid data landscapes is the fact that the lifespan of that data—the time that it is most relevant and valuable—is drastically 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 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 next generation of IBM Cloud Pak for Data

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 intelligent data fabric within the platform.

This intelligent data fabric uses AI to automate complex data management tasks and universally discover, integrate, catalog, secure and govern data across multiple environments.

Automate workflows with an intelligent data fabric

AutoAI
Automate data preparation, model development and feature engineering to find and deploy top-performing models in minutes. Simplify AI lifecycle management to build models faster, accelerate deployment and open up AI to broader skill sets.

Learn more about AutoAI here.

AutoSQL
Automate how you access, update and unify data spread across distributed stores and clouds without the need for data movement or replication. AutoSQL is a high performance, universal query engine. It simplifies your data landscape by allowing you to use the same query in conjunction with the platform's existing data across disparate data sources including data warehouses, data lakes and streaming data, all without manual changes.

Learn more about AutoSQL here.

AutoCatalog
Automate how data throughout a hybrid data and cloud landscape is discovered, cataloged and enriched for user relevancy. Provide business-ready data to more people.

AutoPrivacy
Automate how you enforce universal data and usage policies across hybrid data ecosystems in a hybrid cloud landscape to reduce risk while enabling data use.

Learn more about the unified privacy framework IBM Cloud Pak for Data can deliver here.

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
- IBM Cloud Pak for Data System: A preconfigured, hyper-converged system that combines storage, compute, networking and software and reduces private cloud deployment times to a matter of hours
- IBM Cloud Pak for Data as a Service: A "pay-as-you go" subscription model for a starter set of IBM Cloud Pak for Data services, fully managed on IBM Cloud infrastructure.

Top use cases for IBM Cloud Pak for Data

Data store convergence

Simplify your data landscape to enable faster extraction of value from your data, cost effectively.
Data drives digital transformation; organizations need to extract accurate and actionable insights from their data quickly enough to impact business outcome. Data-driven insights spur organization-wide innovation, uncover opportunities for new products or markets, empower salespeople to have more meaningful discussions, and identify internal process that can be improved.

Data store convergence involves the migration of data from legacy to modern data architectures to facilitate and accelerate digital transformation; it involves optimizing data estates to better leverage varying data types, structures, volumes and velocities.

IBM data store convergence enables organizations to collect multi-dimensional data, across all data landscapes to provide business ready data that supports mission critical applications, fuels advanced analytics and enables real-time operational efficiencies while reducing infrastructure.

DataOps

Deliver business-ready data for AI, fast.
As business and market needs change, enterprises need to ensure that they are able to respond quickly to support remote work, cost cutting initiatives, and other imperatives while maintaining their competitive advantage. More enterprises are uncovering that data is their biggest asset. To capture the value of enterprise data, enterprises must face the realities of current hybrid and multicloud data sources and design a modern information architecture to accommodate the wide variety of data sources and formats that are the reality for most.

DataOps data integration and governance capabilities enable organizations to derive value from all their data, which requires organizations to access and trust both structured and unstructured data, anywhere and anytime. Allowing validation rules to run automatically as part of the ingestion process helps reduce silos and ensure that meaningful and valuable information is used in near real-time.

The revolutionary cloud-native IBM Cloud Pak for Data platform unifies market-leading services spanning the entire data and analytics lifecycle and serves as an enterprise data fabric that helps stitch data together for analytics and AI. It includes the capabilities previously provided by the IBM Information Server platform which are now available as DataStage® and IBM Watson® Knowledge Catalog cloud-ready services on IBM Cloud Pak for Data.

Data privacy and security

Build a pervasive privacy framework for hybrid cloud.
Most businesses are faced with two, often competing imperatives: make data more accessible for analysis and AI while ensuring security and compliance. Managing these seemingly contradictory priorities is further complicated by the recent explosion of data types and sources and the corresponding rules and regulations on how that information can and cannot be used. Smarter businesses are looking to new strategies and technologies to help transition their disparate data security, privacy and governance practices into a more holistic approach to understand and police sensitive data throughout their organization.

IBM is simplifying this journey by helping customers to fully understand and manage how sensitive data is utilized throughout their organization. Integrating leading solutions spanning data security, hybrid data management, governance and risk and compliance within a collaborative platform that can be deployed on any cloud, IBM provides a real time view of PII across customers' businesses. In addition to this universal view of who has access to sensitive data, why and what outcomes it has impacted, customers also gain pervasive policy and regulation enforcement tools required to accelerate risk mitigation for all data and AI assets.

Capitalize on intelligent automation to accelerate the collection, cataloging and masking of sensitive data to build a foundation of trusted, compliant data that can be easily accessed by your teams and models. Turn every data consumer into an expert in risk and compliance with built in AI that understands the latest regulations as well as natural language and context to augment your team's skills and understanding.

ModelOps

Automate the AI lifecycle and improve returns on AI investments.
By 2023, 70% of AI workloads will use application containers or be built using a serverless programming model necessitating a DevOps culture.[2]

Model operations (ModelOps) is a principled approach to operationalizing a model in apps. ModelOps synchronizes cadences between the application and model pipelines. With ModelOps you can optimize your data and AI investments using data, models and resources from edge hybrid clouds. ModelOps covers the end-to-end lifecycles for optimizing the use of models and applications across clouds, targeting machine learning and deep learning models, optimization models and other operational models to integrate with continuous integration and continuous delivery (CICD).

To build a ModelOps practice, you need a platform to simplify and automate AI lifecycles of organizing data; building, running and managing models; and optimizing decisions. IBM Cloud Pak for Data is the ideal platform to build your ModelOps practice and accelerate the time to value for your AI investments across any cloud.

Deliver innovation with trustworthy, compliant AI.
AI is starting to deliver on its potential. However, as adoption rises, new gaps in the data lifecycle are emerging as organizations try to govern, manage and protect data and models trained on that data. The term AI governance refers to the organizational approach toward developing policies to effectively design, deploy, and monitor AI-powered models and algorithms with a major focus on fairness, accountability, transparency, safety, and privacy to help ensure fair outcomes.

IBM has been at the forefront of AI for decades and the IBM Cloud Pak for Data platform now addresses every stage of the data and AI lifecycle. It comes with built-in governance and quality tools such as IBM Watson Knowledge Catalog as well as model automation with Watson Studio and purpose-built AI model risk management tools including Watson OpenScale™. The foundation for AI is data. With a unified platform that builds trust and ensures consistent compliance, AI can truly be used to deliver innovation.

Next steps

Contact your IBM Business Partner to learn more about IBM Cloud Pak for Data.

World Wide Technology

https://www.wwt.com/partner/ibm/overview

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
May 2021

1 IBM Global AI Adoption Index 2021 Executive Summary, 2021.

2 Gartner Artificial Intelligence, Data Science & Machine Learning, 2020.