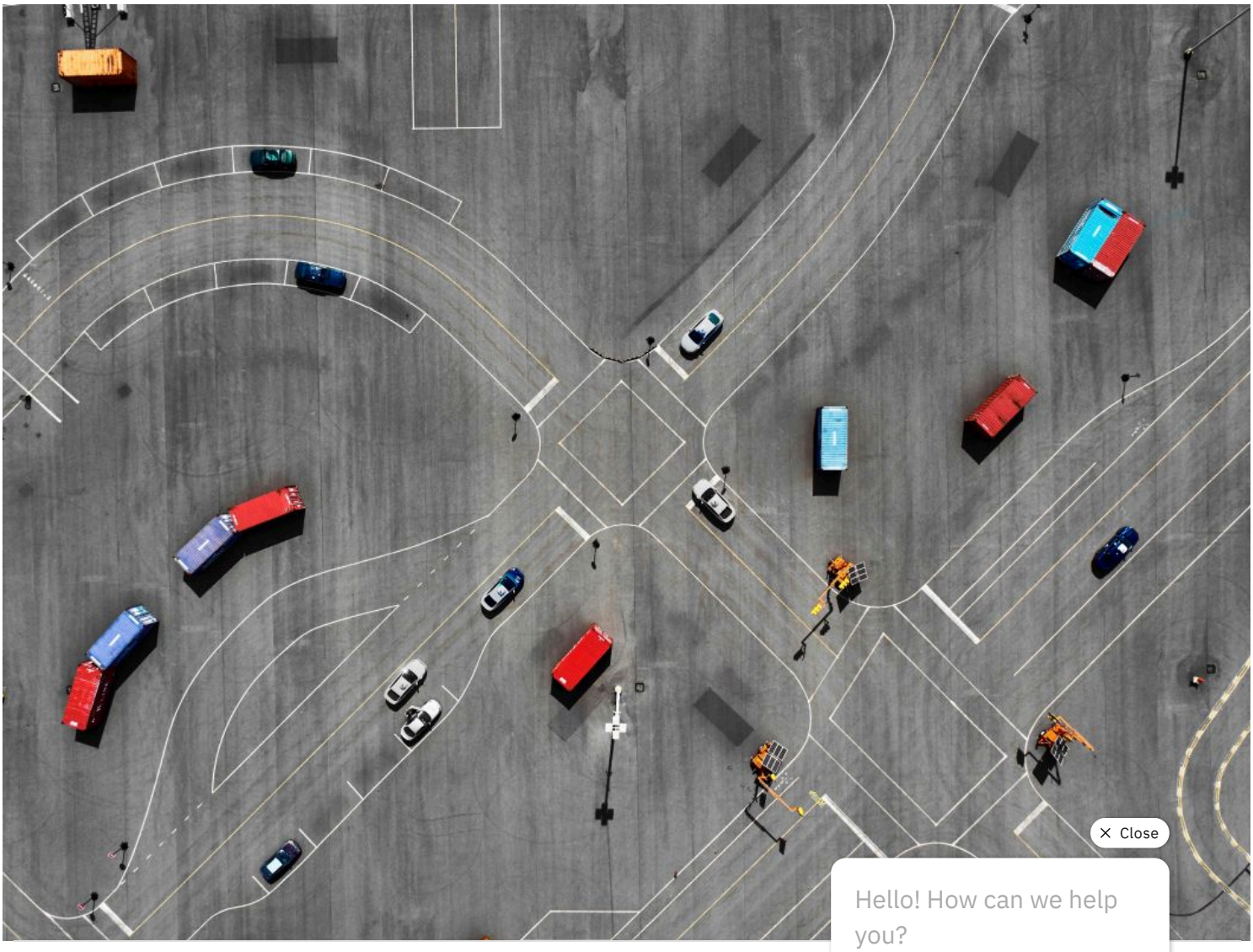
Improving AI workloads u Cloud

IBM Chief Information Officer organization optimizes models with the help of watsonx

Learn more about IBM Cloud →

Learn more about Re



Business challenge Transformation Outcomes About The IBM Chief Informat

A challenge to meet the demands of AI applications

The IBM hybrid cloud platform is confronted with a complex challenge: hosting new artificial intelligence (AI) applications alongside nearly 2,000 existing internal workloads that comprise business operations at IBM.[1] When hosting AI applications on a hybrid cloud platform, these requirements need to be considered:

- AI applications must comply with the company's intellectual property (IP) requirements.
- Developer tools need to be simple and foster productivity when used to create applications that utilize generative AI content.
- Sensitive data must be securely managed throughout ingestion and usage.
- Hosting environments must be cost-effective in terms of hardware resources, development and operations.
- Some AI workloads need to be hosted on-premises and require support for the same AI environments on local hardware.

Initially, the IBM Chief Information Officer (CIO) organization hosted a vector database and generative AI models that use provisioned, high-cost graphics processing units (GPUs). Though this option was effective, it wasn't optimal. During peak hours, competition for resources between Development and Production teams resulted in the acquisition of additional GPUs, which remained idle during off-peak hours.

Also, this complex environment required the operations team to dedicate extra time to manage AI clusters. This involved procuring resources, patching, and managing resiliency.

[1]IBM Hybrid Cloud applications data was obtained through the Superset IBM Application Portfolio Management (APM) ServiceNow Dashboard on September 30, 2024.

“IBM CIO needs to deliver AI work
scale, built on our data, with a strong

commitment to integrity and transparency. Building on hybrid cloud, with IBM cloud components, we scaled to meet the enterprise demands."

Matt Lyteson
Vice President, CIO Technology Platforms Transformation
IBM

Optimizing AI workload management with watsonx on IBM Cloud

The need to simplify the CIO organization environment was evident. So, the AI model and data management, development and production workloads were transferred to IBM watsonx™ as a Service on the IBM Cloud®.

Now, the tools and resources required by the application teams to support the development and training of AI models are supported by watsonx. Even deployment of AI models for application inferencing is handled by the watsonx integrated environment.

"The CIO Hybrid Cloud platform team has found the IBM Cloud environment highly effective to support production and development AI workloads."

Ben Pritchett
Platform Architect, CIO AI, Automation and Data Platform

Successful transition to IBM Cloud: efficient data management and AI services

The integration with IBM® watsonx.data™ and IBM watsonx.ai™ on the IBM Cloud enables the CIO organization to efficiently manage domain-specific data and AI services.

It also provides a series of advantages that are important to the business:

- The consumption-based pricing model is based on the resources used.
- The spikes in GPU usage are handled by the elasticity of watsonx as a Service on the IBM Cloud, with the operations team no longer managing GPU resources.
- The tools required for development.
- Secured deployment of models.
- The watsonx environment is available on Red Hat® OpenShift®, allowing flexibility, including workload placement on premises.
- When combined with watsonx.governance, under some scenarios, IBM has offered intellectual property indemnification to the output that is generated by covered models, including IBM-developed and some third-party foundation models that are available from watsonx.ai.

AI applications are revolutionizing the way IBM operates by integrating AI into its systems and processes. Applications like AskIBM, the AI-powered digital assistant that is automating day-to-day business tasks, are in production. By leveraging the IBM Cloud platform offerings, the CIO organization has the ability to scale solutions according to its specific needs.

About The IBM Chief Information Officer (CIO) organization

The IBM Chief Information Officer (CIO) organization leads the internal IBM IT strategy and is responsible for delivering, securing, modernizing and supporting the IT solutions that IBM employees, clients and partners use to do their jobs every day. The CIO organization's strategy encompasses creating an adaptive IT platform that makes IT tools, applications and systems easier to access across the enterprise, accelerates problem-solving and serves as an innovation engine for IBM, catalyzing business growth.

Solution components

- IBM® watsonx.ai™ →
- IBM watsonx.data™ →
- IBM watsonx™ →
- IBM Cloud® →
- Red Hat® OpenShift® →

The CIO organization optimizes AI workload management with watsonx on IBM Cloud

The CIO organization successfully transitioned the AI data management, development and production workload to watsonx on IBM Cloud to help increase efficient management of domain-specific data and AI services.

Learn more about IBM watsonx.ai →

Learn more about IBM watsonx.data →

Hello! How can we help you?

country in which IBM operates.

Red Hat® and OpenShift® are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

Client examples are presented as illustrations of how those clients have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.