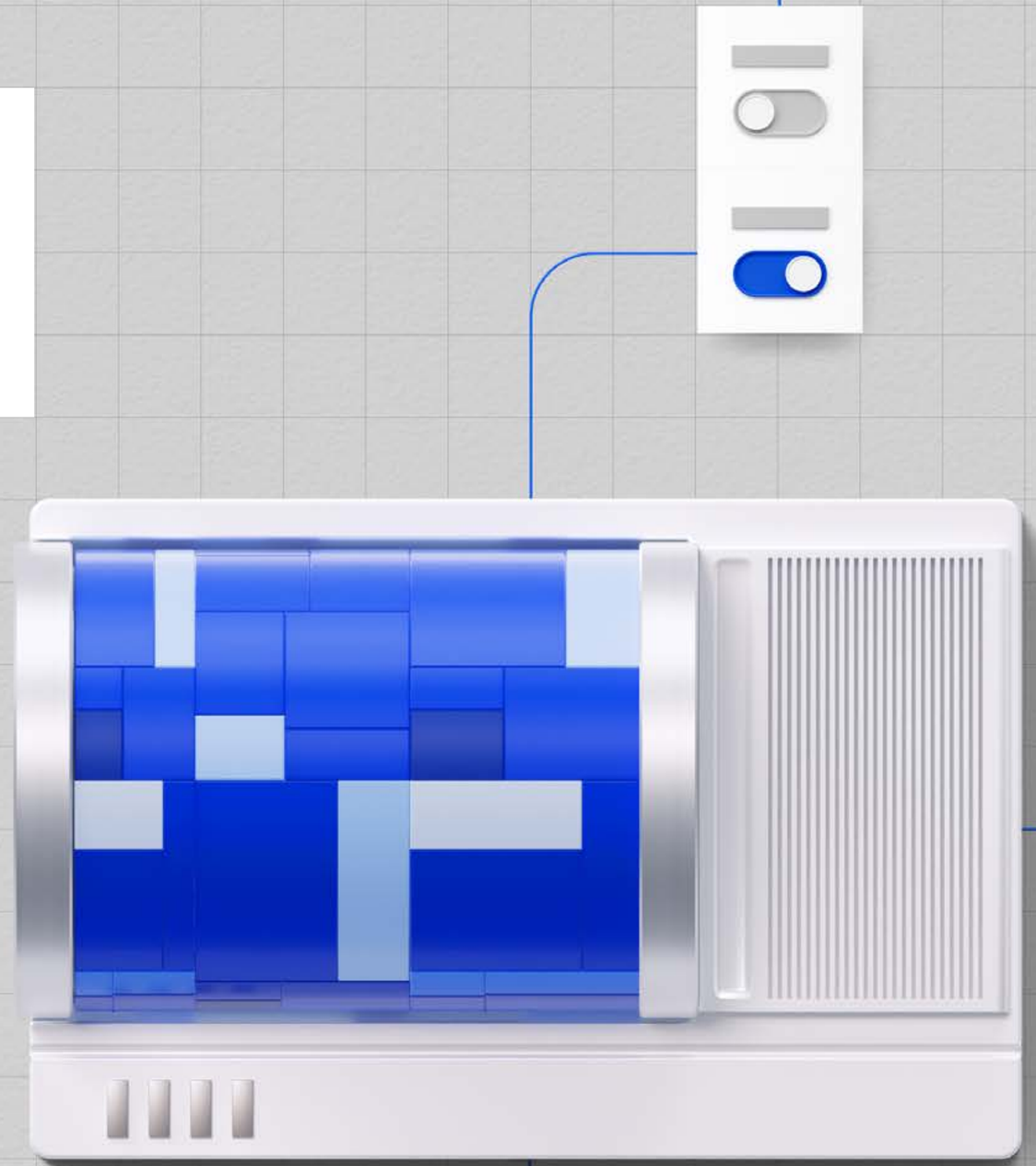
Unlock your data’s potential

AI-powered insights with IBM Storage technology



IBM

Tri-Delta Resources

Contents

AI, high-performance computing (HPC) and advanced analytics are transforming business—driving the development of new applications that can extract actionable insights from massive global datasets.

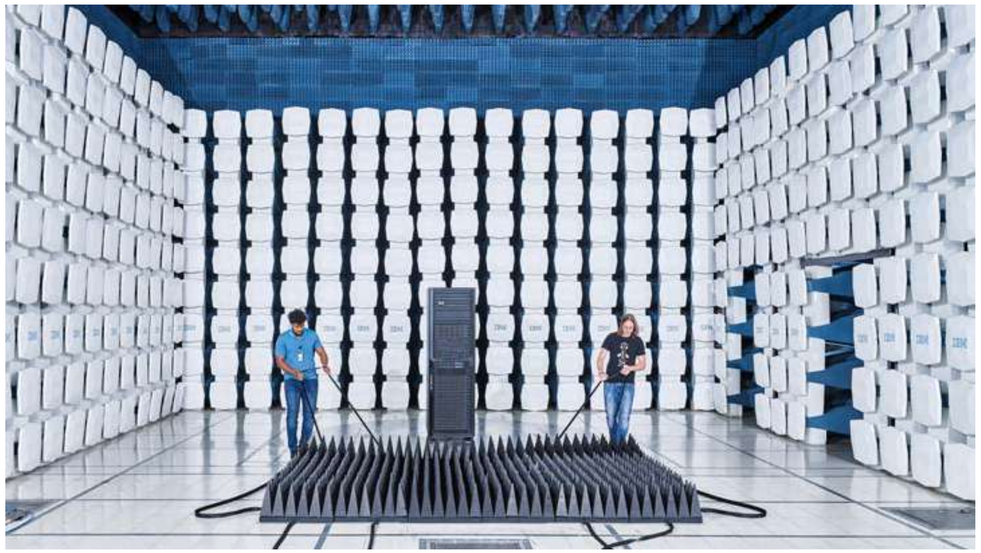
However, effectively leveraging these emergent capabilities requires a robust and adaptable storage infrastructure. Many traditional enterprise workloads operate on structured data in database fields, but many AI and analytics workloads work with file-based applications and unstructured data. This approach requires powerful storage solutions that help organizations maximize the value of their distributed file and object data.

01

The data dilemma in the age of AI

AI and advanced analytics thrive on data. But large organizations create tons of new data every day, and it’s often globally dispersed and in a variety of different forms.

This situation presents some major challenges:



AI, HPC and analytics workloads generate an immense and constantly expanding volume of data, necessitating scalable storage solutions.

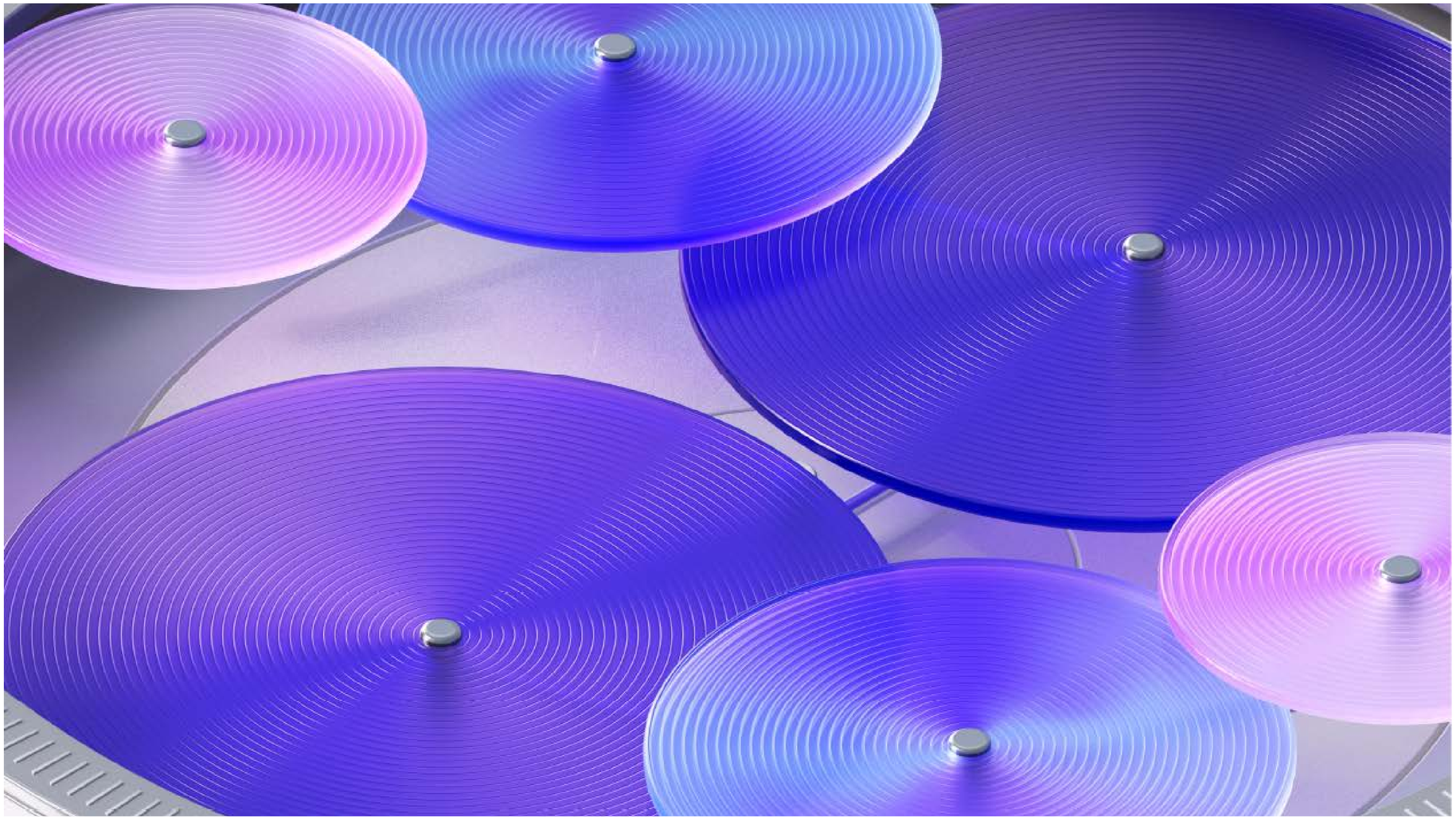
Data can often reside in isolated pockets or silos across different systems and locations, hindering a holistic view and limiting the scope of AI initiatives.

Training and inferencing AI is computationally intensive and requires high-speed data access for optimal performance. Delays in data retrieval may significantly hinder results.

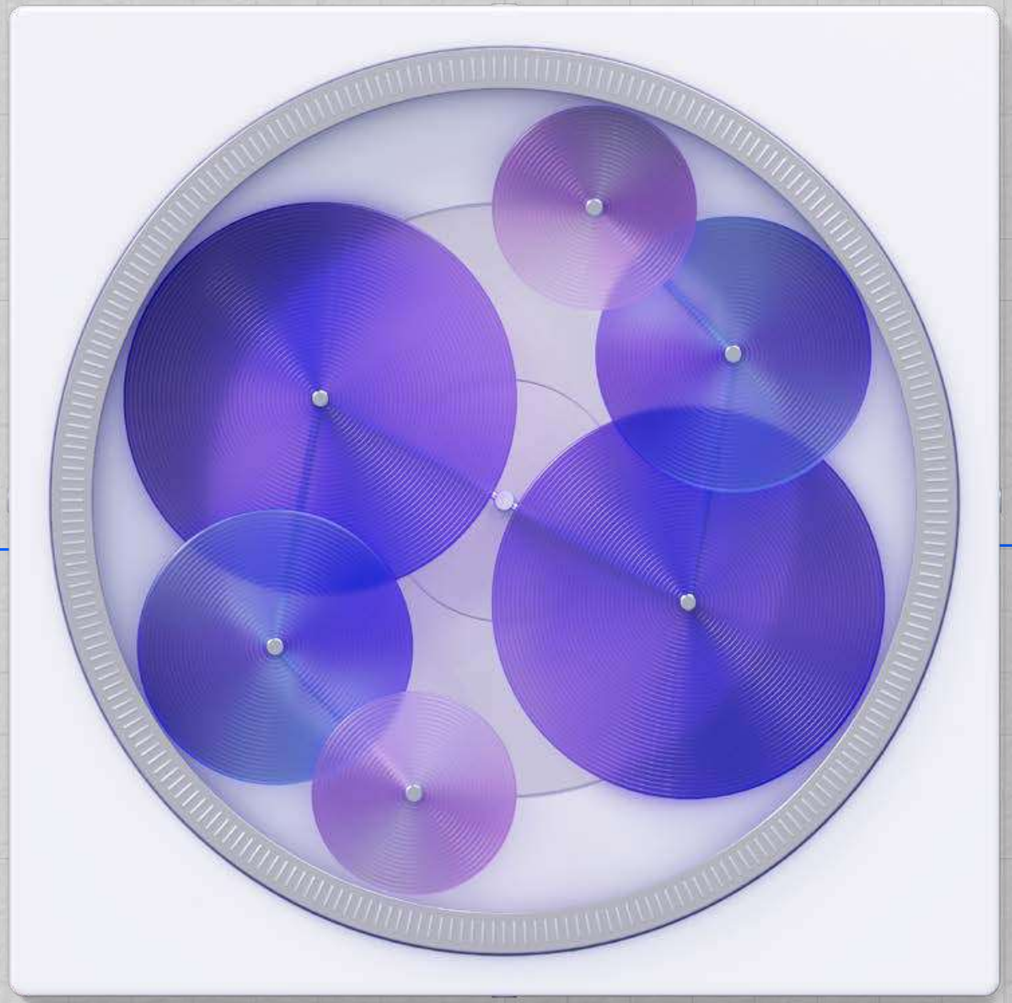
Specialized resources, such as NVIDIA GPUs, are costly and scarce—a significant challenge for organizations seeking to scale their AI infrastructure.

Don’t let your tech estate crumble like sand; build it on solid ground

To maximize the full potential of this new, innovative technology, businesses need a strong foundation to build upon. Such a foundation is the basis of these organizations trusting their AI systems and having the confidence to stride forward and push the envelope in their own space. But as you can see with the following study results, there needs to be a paradigm shift in how we approach IT and AI.



According to an IBM Institute for Business Value report:



16%

of surveyed executives say they're very confident their cloud and data capabilities are fully ready to support generative AI investments in 2024,

27%

say they're unsure of readiness.[1]

According to a recent Salesforce survey:

75%

of those who don't trust the data that trains AI also believe that AI lacks the information needed to be useful.[2]

56%

of AI users say it's difficult to get what they want out of AI.[2]

54%

of AI users don't trust the data used to train AI systems.[2]

68%

of those who don't trust the data that trains AI are hesitant to adopt it.[2]

There's not effective AI without a solid IA

Organizations need more than algorithms to generate value from AI, HPC and analytics applications—they need an information architecture (IA) that efficiently handles the unique ways these workloads manage data. This IA encompasses:

Data Collection

Consolidating a variety of data forms from disparate sources, including edge devices, on-prem systems and cloud environments, helps create a more unified data infrastructure.

Data Organization

Indexing, tagging and cataloging data helps ensure discoverability and facilitates efficient analysis.

Data Analysis

AI training and inferencing workloads are computationally intensive, necessitating storage systems that can handle high throughput and low latency demand.

Data Infusion

A computationally intensive storage system is required to handle the high throughput and low-latency demands of AI training and inferencing workloads.

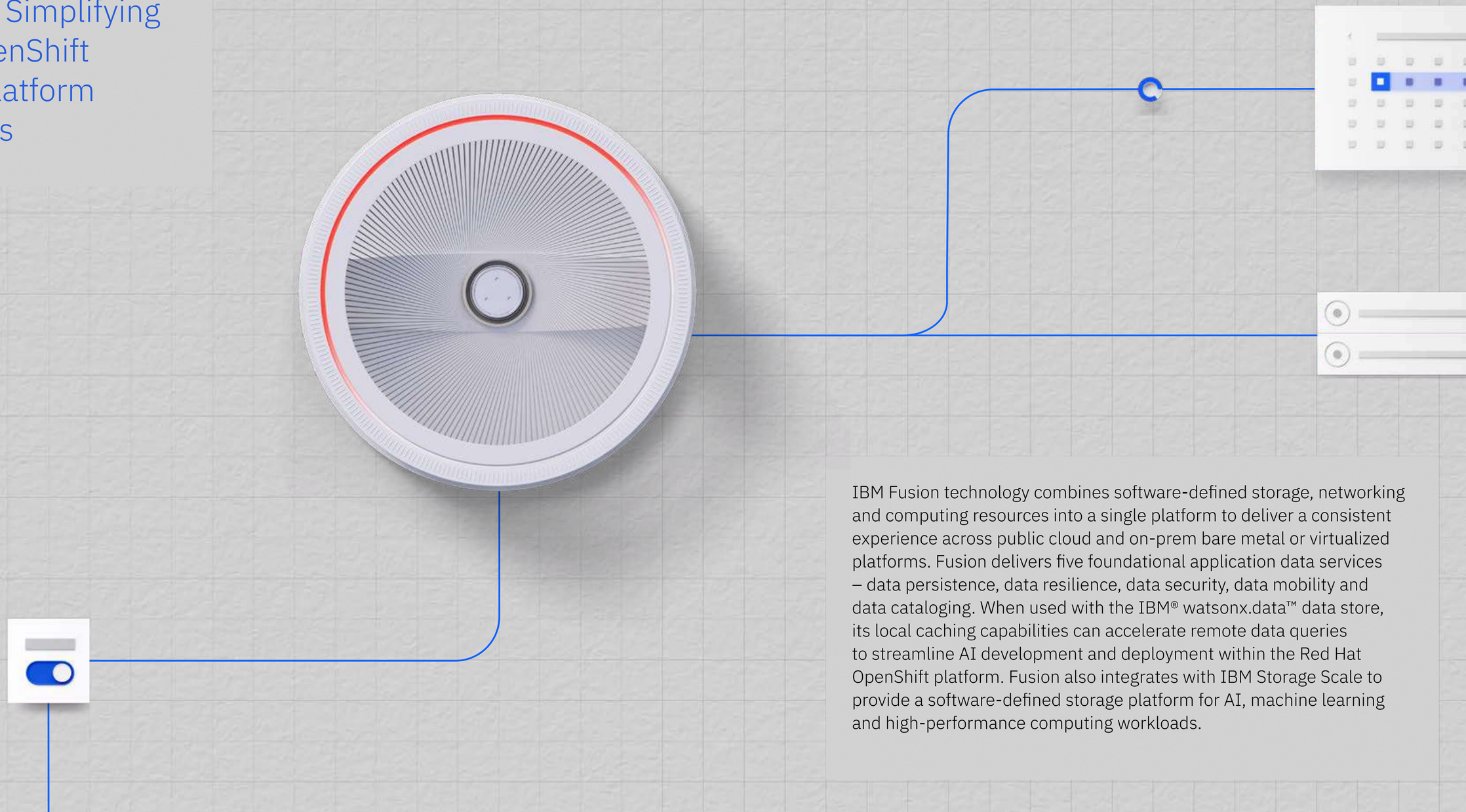
Data Modernization

Effective IA leverages cloud-native technologies and supports hybrid cloud deployments to help ensure scalability, flexibility and agility in managing AI workloads.

02

Drive AI innovation with scalable IA storage solutions

IBM Fusion: Simplifying Red Hat OpenShift Container Platform deployments



IBM Fusion technology combines software-defined storage, networking and computing resources into a single platform to deliver a consistent experience across public cloud and on-prem bare metal or virtualized platforms. Fusion delivers five foundational application data services – data persistence, data resilience, data security, data mobility and data cataloging. When used with the IBM® watsonx.data™ data store, its local caching capabilities can accelerate remote data queries to streamline AI development and deployment within the Red Hat OpenShift platform. Fusion also integrates with IBM Storage Scale to provide a software-defined storage platform for AI, machine learning and high-performance computing workloads.

IBM Storage Ceph: A scalable data lake foundation

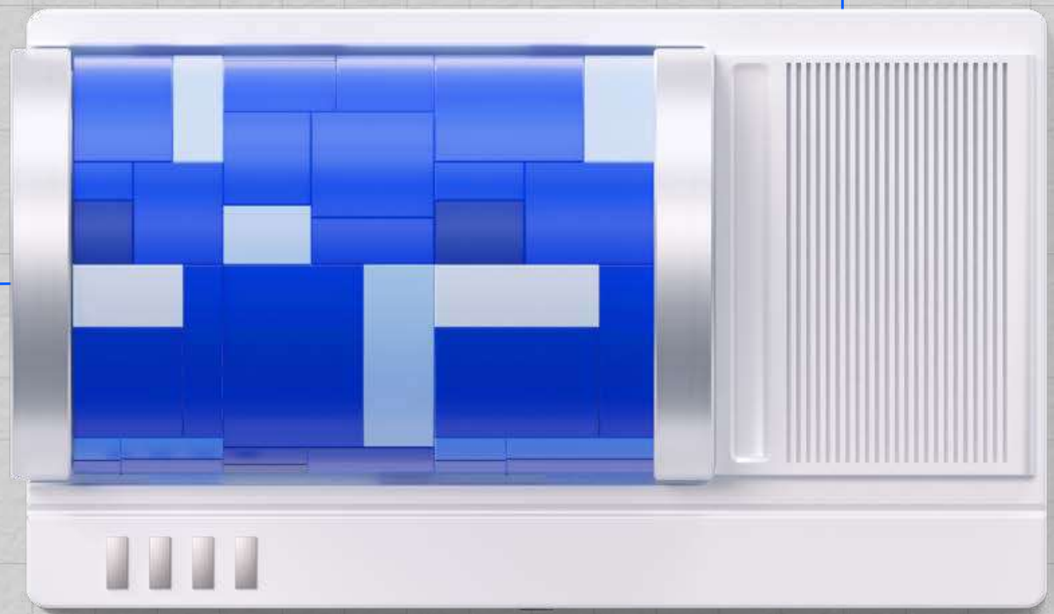
IBM Storage Ceph is a software-defined and scalable storage platform, facilitating an ideal environment for data lakes.

Data lakes are centralized repositories designed to store massive amounts of raw data in their native format. They excel in scenarios requiring open-ended investigation, such as AI, machine learning and exploratory analytics.

Data lakes also enable experimentation with large datasets without being restricted by predefined schemas.

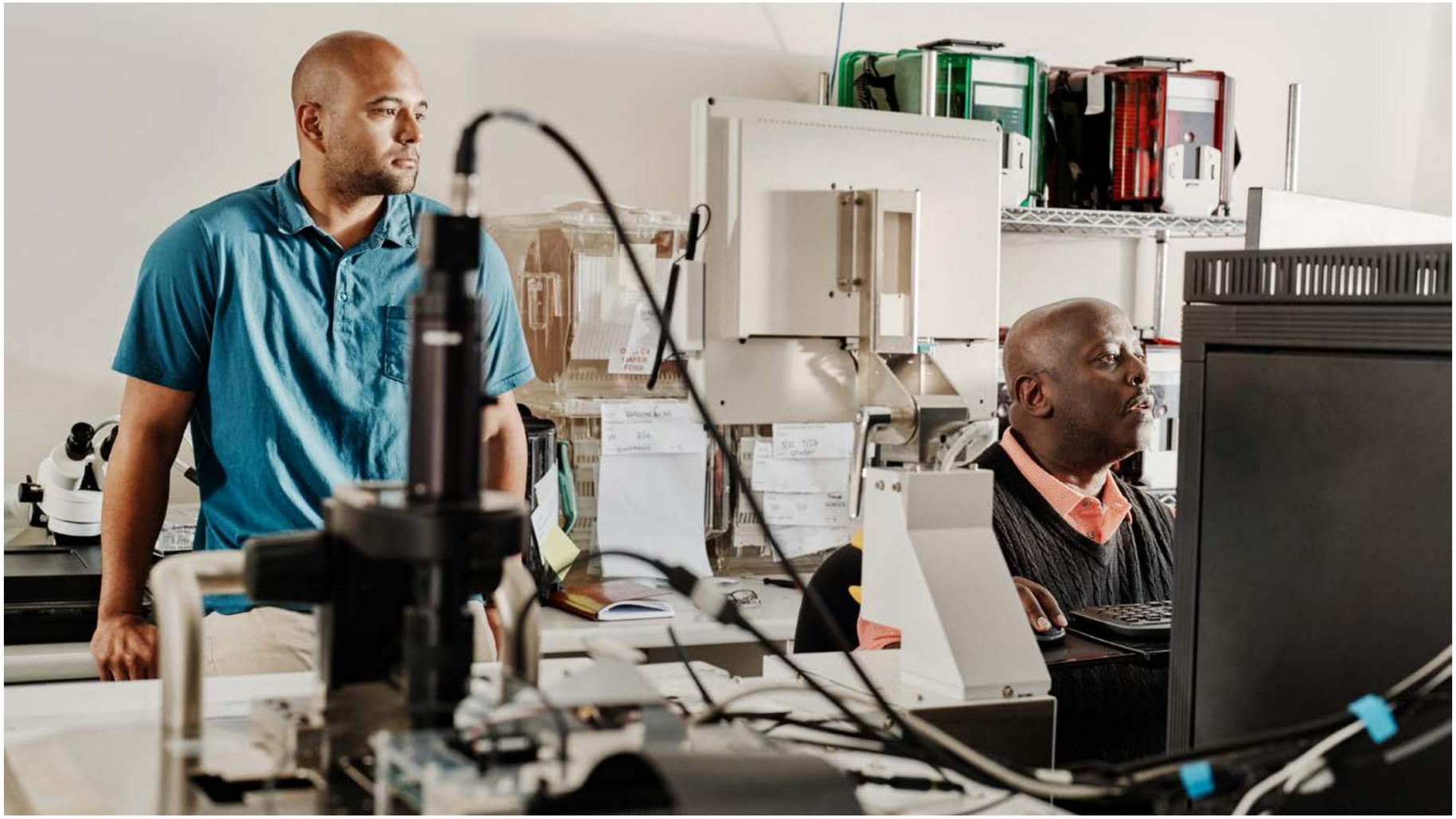
IBM Storage Ceph technology supports block, file and object storage, allowing organizations to eliminate data silos and consolidate their storage infrastructure. This technology can self-heal and self-manage to help ensure high availability and reduce overhead. Integrated with the watsonx.data data store, IBM Storage Ceph is well-equipped for AI workloads requiring a data lakehouse architecture.

Data lakehouses, such as watsonx.data, combine the scalability and flexibility of data lakes with the management and performance capabilities of data warehouses. Unlike data lakes, which store data in its raw format, data lakehouses transform, cleanse and structure data to make it more suitable for high-performance analytics and business intelligence applications.



IBM Storage Scale: Building a global data platform

The IBM® Storage Scale solution offers software-defined file and object storage technology that provides a robust foundation for managing unstructured data in AI and HPC environments.
Its key features include:



Global data platform
IBM Storage Scale technology seamlessly connects data from multiple sources and protocols, so you can introduce new datasets without refactoring applications. This helps eliminate data silos and accelerate no-code application modernization.

Multi-protocol access
IBM Storage Scale technology supports a wide range of data access protocols, such as NFS, SMB, Amazon S3, HDFS, POSIX, NVIDIA GPUDirect Storage and Container Native and Container CSI interfaces. This flexibility enables diverse applications to access and process data concurrently.

Parallel file system
Built on a massively parallel file system, IBM Storage Scale technology helps ensure broad compatibility and scalability with deployment possible across a variety of platforms, including x86, IBM Power® servers, IBM zSystems® mainframes, ARM-based POSIX clients, virtual machines (VMs) and Kubernetes container technology.

Automated data tiering
The IBM Storage Scale offering can optimize cost and performance through automated data tiering. Active file management (AFM) dynamically moves data across different tiers—flash, disk, tape or cloud—based on predefined policies, ensuring the most frequently accessed data is stored on the fastest media.

IBM Storage Scale System 6000: Unleashing high-performance storage

To address the challenges of large-scale AI workloads, IBM® Storage Scale System 6000, a hardware implementation of IBM Storage Scale software, offers organizations a solution with exceptional performance and scalability. Key features include:

High throughout and IOPS:

The system provides up to 310 GB/s throughput and 13 million IOPS using NVMeoF, helping ensure rapid data access to accelerate AI training and inferencing.

Direct GPU access:

IBM Storage Scale System 6000 supports the NVIDIA GPUDirect Storage protocol—facilitating a seamless data journey between GPU memory and storage. This functionality bypasses the CPU and DRAM, reducing latency and boosting processing speed.

Massive scalability:

IBM Storage Scale System 6000 grows with data demands, scaling to thousands of nodes. It also can support up to nine expansion enclosures, with each enclosure holding up to 91 HDDs (20 TB or 22 TB), enabling a maximum storage capacity of 18 PB per rack.

Enhanced data resilience:

Robust data resilience measures, such as erasure coding, data replication and immutable snapshots with Safeguarded Copy, protect businesses against hardware failures, data corruption and cyberattacks. Safeguarded Copy can also create established snapshots for rapid data recovery in case of security incidents.

03

Unlocking the value of data with IBM Storage

IBM storage technology provides a comprehensive range of scalable storage solutions for organizations seeking to build a future-focused storage infrastructure for their AI and analytics workloads.

By overcoming the challenges of managing vast and complex datasets, these solutions pave the way for businesses to maximize value from their data and accelerate their AI-driven transformations.

1. From chaos to cash: How hybrid by design creates business value, IBM Institute of Business Value, 7 May 2024.

2. “New Survey: Data Will Make or Break Workers’ Trust in AI”, Salesforce survey report, April 2024

IBM Corporation
New Orchard Road
Armonk, NY 10504