IBM Storage for Data and AI

# Unlock your data at scale for AI

# Breaking down data barriers and creating an information supply chain with a *global data platform*

- **Optimized** data with higher performance, lower cost/energy usage and up to 380% ROI[1]
- **Secured** data that can be concurrently accessed in multiple locations with high availability and fast cyber secure recovery for PBs of data
- **Globally connected** data that can be accessed with up to 126GB/s in a single node and scalable to 1000s of nodes and optimization that turns data on or off when needed

1 https://www.ibm.com/downloads/cas/KY1EY12A

## Table of contents


**01** Introduction

**02** There is no AI without IA

- Collect
- Organize
- Analyze
- Infuse
- Modernize

**03** Modern IA for AI and hybrid cloud

- IBM Storage Scale
- IBM Stoprage Ceph
- IBM Storage Scale System

**04** Case studies: Creating a competitive advantage

- Continental Automotive
- University of Birmingham

**05** Conclusion


1.

Artificial intelligence (AI) is a journey that begins with data. But the requirements for this data are not the same as traditional legacy transactional storage or even for traditional file and object storage. Too often the focus for new workloads is only the compute resources or applications without the proper planning or strategy for the data or storage environment. We like to say there is no AI without an Information Architecture (IA) strategy. Customers are starting to realize that creating silos of data may be easy to get one project off the list but it is making it difficult to get a holistic view of all your information, limiting the value of AI. Without a strategy customers are chasing data requirements for each project and missing opportunities for overall optimization with new performance requirements, security, and agility of connecting data to more users and multiple applications.

Additionally, organizations are moving towards hybrid cloud to respond to evolving business needs. Infrastructure that was not built for AI and hybrid cloud is not flexible enough to respond to modern workloads and demands without adding complexity.

The best AI is built on a foundation of data that is collected, organized and analyzed carefully, then infused into the business. This foundation should also be open, flexible and allow access to data of every type, regardless of where it lives. Every successful AI project goes through a multistep process that starts with having the right data and progresses to using AI broadly.

# 76% of IT decision-makers

Surveyed said AI will be a key part of their digital Transformation strategy over the next one to two years.[1]

2b

1 . IDC White Paper, sponsored by IBM, Accelerating AI Modernization with Data Infrastructure, doc # US47460721, February 2021.

# AI IA There is no AI without IA

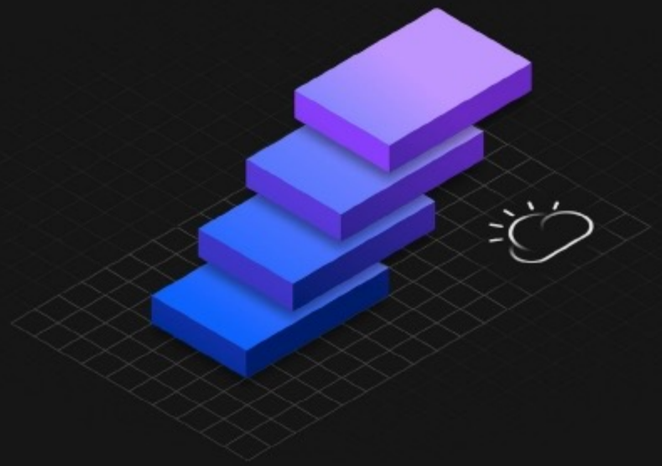
As companies begin to modernize, they seek to provide an architecture that will propel them into the future. The journey to AI is about moving data from ingest to insights with an IA that can easily be integrated throughout the organization. It is important that each part of the AI Ladder provides an integration to the entire journey. Starting a project on one part of the journey is fine, but it is critical to ensure the project considers an overall IA for AI to optimize resources and modernize your infrastructure for expanding AI workloads.

## Collect

Data is the fuel that powers AI, but it can become trapped or stored in a way that makes it difficult or cost-prohibitive to maintain or expand. You will need to unleash that data so it can expand from edge to core to public cloud within a simple and cost-efficient infrastructure. IBM Storage for data and AI makes data simple and accessible for hybrid cloud with AI storage solutions that fit your existing business model.

## Organize

AI can only be as good as the data it relies on. Businesses must fully understand what data they have so they can leverage it for AI and other organizational needs, including compliance, data optimization, data cataloging and data governance. IBM Storage for data and AI provides insights into data from multiple sources by automatically and continuously indexing objects and files when changes are made and storing this information in the built-in storage catalog.



## Analyze

Analysis is critical to the AI journey and must provide high performance for fast analysis and seamless connection to both the data lake and the storage catalog. Organizations must plan for issues beyond the deployment of AI; you need to build AI infrastructures that give you confidence in your data and that enable you to access it wherever it resides. IBM Storage for data and AI provides high-performance access to data and an integrated AI infrastructure for analysis.

## Infuse

Business challenges can become an opportunity to explore, understand, predict and bring an AI infrastructure to every organization. IBM Storage for data and AI is empowering you to use data and AI storage to leverage that infrastructure in more ways that bring value to your organization. IBM provides more choices with software defined storage options to deploy when and where you need a solution for modernization of your data strategy.

## Modernize

A solid IA is the foundation for AI and hybrid cloud. Modernizing your infrastructure means building a foundation that not only takes advantage of cloud-native technologies, but also drives AI throughout the organization. IBM Storage for data and AI delivers the flexibility needed to respond to AI workloads, and integrates with Kubernetes and the Red Hat® OpenShift® platform, making it easier to deploy cloud-native applications.

Adopting AI is not without its challenges. The general-purpose storage infrastructure that organizations are accustomed to using needs to be replaced or supplemented with storage systems that are geared towards AI specific tasks.[1]

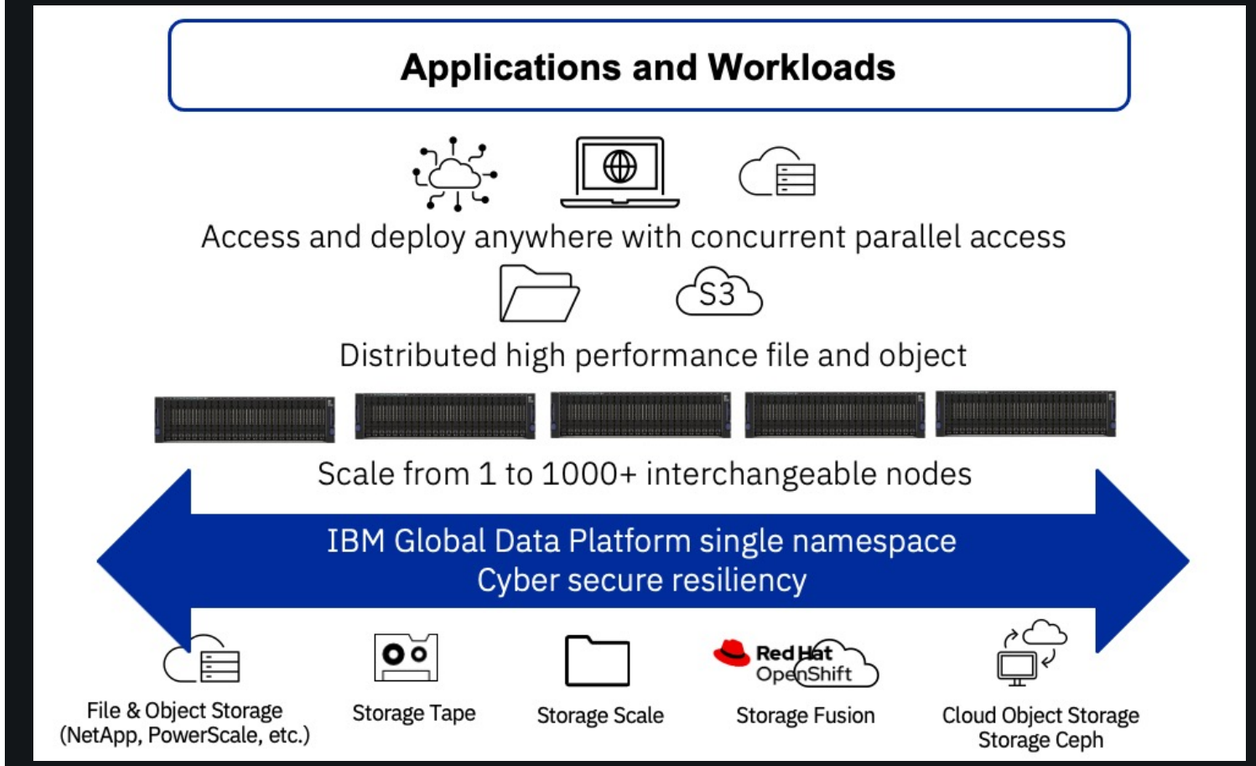
- **Moving from experiments to scaling AI for business value.** Modernizing for an AI-focused digital transformation requires expertise in new standards of developing, implementing and maintaining AI solutions at scale.
- **Legacy infrastructure/complexity.** Organizations can no longer use traditional, general-purpose computing or legacy storage infrastructure. This outdated infrastructure increases complexity and is not flexible enough to respond to AI workload demands. Instead, they must employ a scalable compute with an equally scalable, high-performing, integrated, flexible and secure storage infrastructure.
- **Data silos.** Storage is typically implemented with specific storage solutions that create silos of data. These silos are not integrated together, nor are they integrated with a comprehensive set of infrastructure solutions, resulting in a lack of global data access.

IBM has been engineering software storage for over 30 years to help customers solve some of the most difficult data problems. Software is the key to solving these problems quickly and efficiently. IBM has built a platform on four essential data services and the ability to add on other data services as requirements grow. This software defined platform which IBM calls the global data platform consists of a set of core software services that each contain specific capabilities that help break data barriers and solve client application requirements.

Organizations face a few core challenges when adopting AI, including scaling AI for business value, the use of legacy infrastructure and elimination of data silos.

1 . IDC White Paper, sponsored by IBM, Accelerating AI Modernization with Data Infrastructure, doc # US47460721, February 2021.

# The Global Data Platform



# Modernize IA for AI and hybrid cloud

AI initiatives are easier and more likely to succeed if they are built on a solid foundation. IBM Storage provides that foundation with a collection of offerings that modernize your IA and address the top business challenges associated with deploying AI workloads.

## Data Cataloging Services

IBM Storage data catalog is a multisource (IBM and non-IBM storage) data analysis and data governance tool that automatically and continuously indexes objects and files whenever changes are made using the metadata. It can also be used to create custom tags and policy-based workflows to orchestrate content inspection and activate data in AI, machine learning(ML) and analytics workflows. The data catalog helps enable faster AI analysis, compliance classification, image and video indexing, personal data identification, AI data pipeline integration, real-time data discovery and new insights to optimize data and find bad or duplicate data. ***The data catalog is included with IBM Storage Scale*** DME and connects data in real-time to Storage Scale, IBM Cloud Object Storage and Storage Ceph.

## IBM Storage Scale

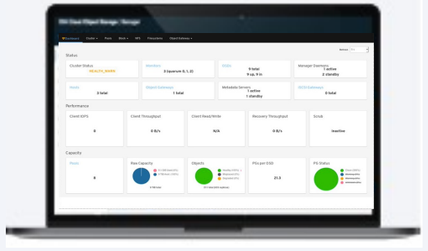
IBM Storage Scale is a highly scalable, data-efficient, high-performance storage solution with enterprise security and a global data platform for both file and object storage data. IBM Storage Scale enables the unification of data across a hybrid cloud into a single scale-out storage solution that includes multi-vendor and multi-cloud support. Available as software that can be deployed on most any commodity hardware on the edge or the core data center or even in the public cloud. IBM Storage Scale is also available as a fully integrated appliance (IBM Storage Scale System) for an easy and optimized solution.

**Learn about IBM Storage Scale**

## IBM Storage Ceph

IBM Storage Ceph is a highly scalable cloud native storage solution designed for large inexpensive data lake repositories. IBM Storage Ceph uses an innovative approach for cost-effectively storing large volumes of file, object and block data. It delivers the capabilities required to provide continuous access to data assets while improving research outcomes, decision making, business responsiveness and regulatory or legal demands. Available as software that can be deployed on most any commodity hardware on the edge or the core data center or even in the public cloud. IBM Storage Ceph seamlessly connects to Storage Scale for a seamless global data platform solution.

**Learn about IBM Storage Ceph**



Faster results with ultra performance for AI or any file/object workloads from containers to GPUs

Better Economics with GUI based lifecycle management from ultra performance flash to cloud to tape

More secure with multi-site protection and cyber security automation that recovers PB data in minutes

Increased agility by breaking silos and created a global single source of truth for all file and object data

# IBM Storage Scale System

IBM Storage Scale System is an AI optimized and modern implementation of software-defined storage, making it easier to deploy fast, highly scalable storage for AI and the hybrid cloud.

[Learn about IBM Storage Scale System]

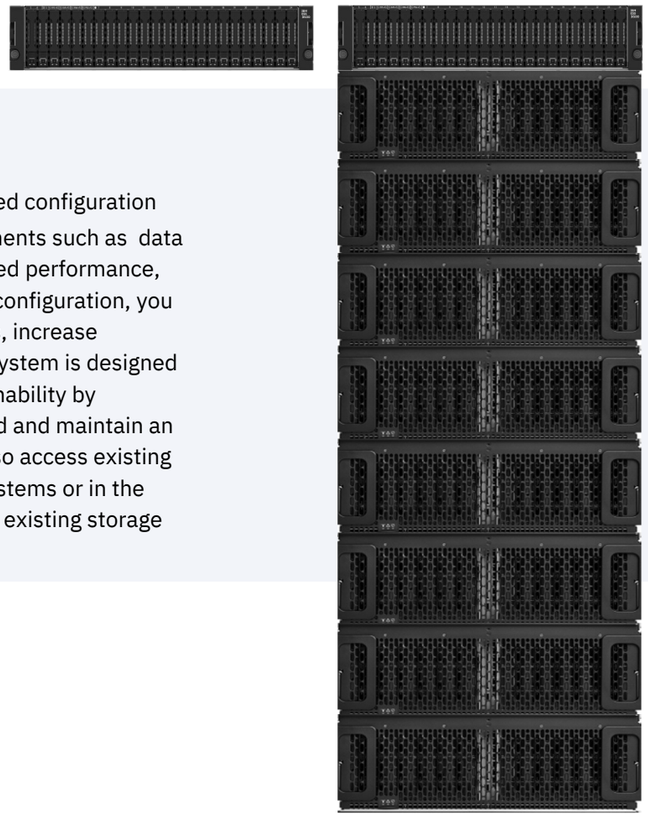
Up to

**1.2M IOPs**

**126GB/s**

per node[1]

## A single node from 2u to 34u

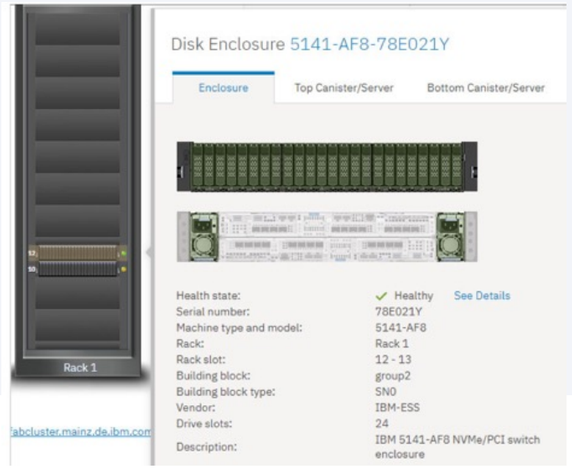


## Unified Configuration

IBM Storage Scale System with a unified configuration is designed for large capacity requirements such as data lakes or cloud scale data with increased performance, density and scalability. With a unified configuration, you can consolidate massive data volumes, increase simplicity and accelerate speed. The system is designed to lower cost and increase data sustainability by transparently archiving to tape or cloud and maintain an always on access. The system can also access existing storage from other vendors storage systems or in the cloud offering seamless integration to existing storage environments.

## GUI Configuration

IBM Storage Scale System is designed to meet and beat the challenges of modern organizations who are managing more resources and require less complexity. IBM provides a simplified interface, so management is easier and simple to use.




## Flash with Turbo-Performance Tier

IBM Storage Scale System with an all NVMe flash configuration is designed to be the most efficient part of your Global Data Platform. In just a 2u node configuration this system maintains full redundancy of key components and provides the fastest interface to all the data contained in the global data platform. Start with a single node and scale to 1000s all in the same cluster and all nodes maintaining the same parallel access to all the data. Any data, any where, all the time.

1 Based on internal benchmarks from IBM engineering with IBM Storage Scale System

# Continental Automotive

## Accelerating insight into vehicle safety at Continental

For many people, driving is simply a series of automatic decisions. Training AI to make those same decisions even a 10th of a second faster requires petabytes of data. To develop autonomous driving solutions that potentially make driving safer, Continental used IBM Elastic Storage System, IBM Spectrum Scale and NVIDIA DGX™ systems to:

- Modernize its application development without giving up on infrastructure requirements like performance, scalability or simplicity.
- Ensure that its infrastructure will support the growth required, whether in the cloud or on premises.
- Optimize for deep learning with multi-node training, enabling it to increase model accuracy for higher levels of safety without impacting time to production.

Read the case study →


"The collaboration between Continental, IBM Storage and NVIDIA is bringing a promise to life in terms of safety."

**Robert Thiel**
Head of AI, Advanced Driver Assistance, Continental Automotive AG

## Results

**150 years**

Continental has pushed the boundaries of automotive innovation for 150 years.

**70%**

Continental improved AI training time 70% using IBM Spectrum® Scale and NVIDIA DGX systems.

**14x**

Continental has the ability to run at least 14x more deep learning experiments per month at the same time.

# University of Birmingham

## Driving innovative research forward by taking control of data

Today ’s research simulations generate more data than ever before. To meet this ever-increasing demand, the University of Birmingham deployed IBM Spectrum Scale and IBM Spectrum Protect to:

- Provide a single data management plane across multiple storage systems.
- Enable price-performance decisions when matching workloads to platforms, without causing complexity to spiral out of control.
- Allow researchers to deploy applications where it makes sense with immediate data availability.

Read the case study →

“We support research in a wide range of areas, including applying and developing techniques to use AI and deep learning. For example, we’re collaborating with the University of Nottingham on the Centre of Membrane Proteins and Receptors project. By analyzing the super high-resolution images produced by the latest generations of microscopes, the project will shed light on how cardiovascular disease, respiratory disorders and cancer can be better prevented and treated.”

**Simon Thompson**
Research Computing Infrastructure Architect, University of Birmingham

## Results

Supports compliance with data protection regulations at low cost and without disruption.

### Up to 2

Up to 2 FTEs estimated saving due to enhanced operational efficiency.

### 5,000

5,000 researchers supported by infrastructure that helps them find solutions to key issues faster.

# Conclusion

The decisions you make as you build your AI foundation have far-reaching implications that will impact your organization every step of the way and, ultimately, determine business outcomes. That is why having the right partner from the outset is critical.

IBM Storage for data and AI is more than a set of storage products and solutions. It consists of a storage strategy that will help you on your journey to AI and hybrid cloud. IBM continues to drive leadership for scalable, high-performance workloads as well as efficient, secure capacity storage for file and object-based solutions. Additionally, IBM Storage for data and AI solutions come ready with broad support and integration with Kubernetes and the Red Hat OpenShift platform.

Our solutions provide a flexible, high-performance IA for AI that modernizes your infrastructure with global data access and services that are simple to manage, faster to access and optimized to scale with cost efficiencies to drive down expenses and bring more value to your organization.

Learn more about IBM Storage for data and AI →