

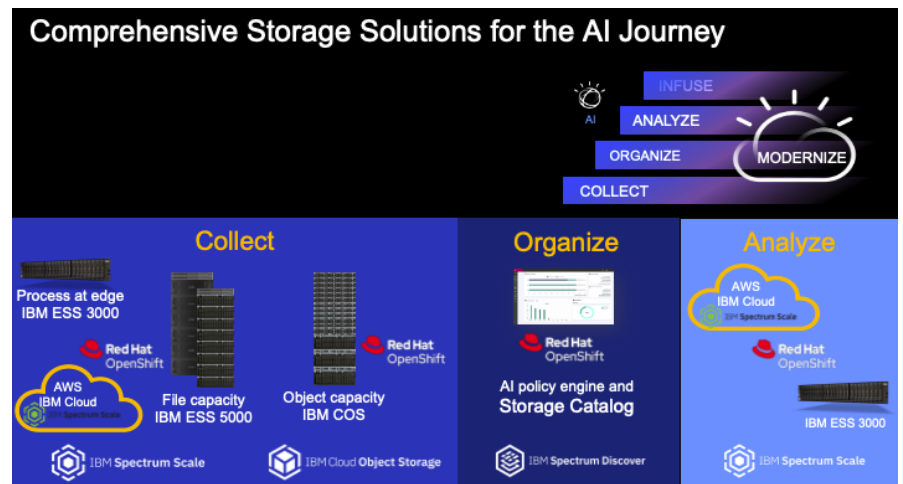
AI acceleration and data orchestration for object storage

Highlights

- IBM Cloud Object Storage accessed from IBM Spectrum Scale
 - Concurrent access to IBM Cloud Object Storage and file access
 - Automatically evict unused data to create space for the required data
 - Eliminate unnecessary data copies and associated management of them
 - Caching functionality allows enterprises to prefetch required data
 - Optionally use IBM Spectrum Discover for policy engine
-

High performance object storage access with IBM Spectrum Scale and IBM Cloud Object Storage

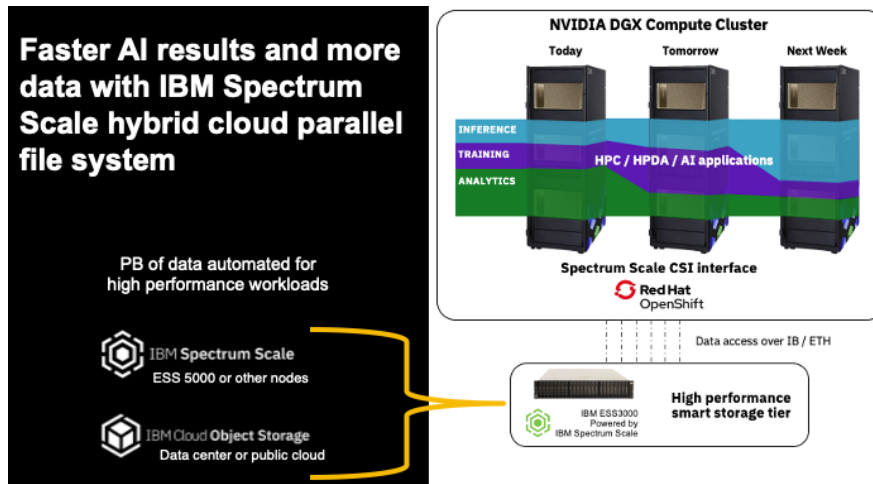
As adoption of AI becomes mainstream in organizations, storage infrastructure to support the large amounts of datasets fueling AI models become critical. IBM Storage for Data and AI has recently introduced a new capability for AI acceleration and automated access to object storage data from a high performance file system. IBM Spectrum Scale is an industry leader in high performance file access for HPC, AI, analytics and other high performance workloads. IBM Cloud Object Storage is a massively scalable and efficient storage solution for cloud native S3 applications. As object storage systems like IBM Cloud Object Storage have become an efficient way to storage massive amounts of data cost effectively and with high reliability, customers are looking for ways to connect object data to AI and analytics workloads with high performance and easy access.



IBM Storage for AI

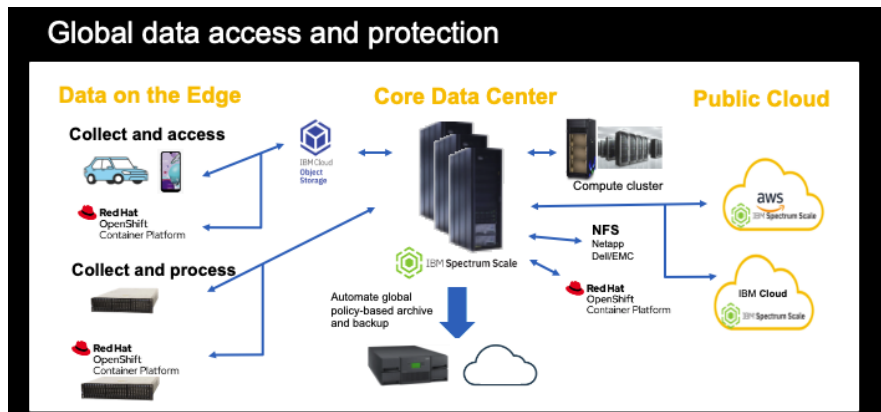
Most current scale-out NAS systems support using object storage or cloud storage as only an archive tier. What if a customer could leverage all the data stored in an object storage system on-prem or in the cloud directly in their file storage for high performance applications? Customers could simultaneously ingest data into the object storage system and leverage that data directly for NVIDIA workloads or other high-performance requirements. With the new AI acceleration using Spectrum Scale this is now possible. This is a key technology that bridges file and object storage together and leverages the best of both storage solutions into a single accessible namespace. This new capability is made possible due the advanced design of Spectrum Scale and AFM and will help propel IBM Storage for Data and AI into many new use cases and customer environments.

IBM has been working with NVIDIA to build validated reference architectures using NVIDIA DGX systems and IBM Spectrum Scale storage. One example of leveraging the comprehensive portfolio of IBM Storage for Data and AI is with NVIDIA. The combination of NVIDIA and IBM are simplifying enterprise AI environments and delivering the required storage and GPU performance to bring faster business results. IBM has demonstrated in the past that IBM Spectrum Scale Storage and the ESS 3000 2U NVMe building block, delivers up to 40GB/s throughput with linear scalability as you attach more ESS 3000 nodes to a DGX cluster. If data is ingested or archived into an object storage system then IBM Spectrum Scale can now transparently include that data into the IBM Spectrum Scale global parallel file system for high performance access to NVIDIA AI analysis.



An example to expand access to more data

IBM Storage for Data and AI now brings a unique data orchestration capability that is differentiated from other storage options. This data orchestration capability transforms ESS 3000 from a “high performance storage” to “high performance smart storage tier”. It provides the ability to connect ESS3000 to your data lakes (File or Object Stores) in your organization and intelligently cache required data for your AI modeling into your DGX Pod.



An AI information architecture

Only IBM can intelligently cache required data from file and object storage from a global federated namespace that can span up to 8YB. This approach allows our customers to optimize TCO and speed productivity while they enjoy superior high performance for their AI workloads.

To deliver this differentiated value, we have enhanced IBM Spectrum Scale's AFM (Active File Management) capability to become IBM's native data mover. This data mover can now connect to any NAS or S3 based data stores like IBM Cloud Object Storage along with the IBM Spectrum Scale global federated namespace.

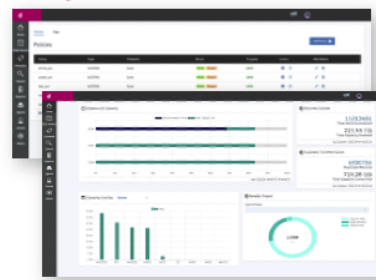
IBM Spectrum Discover

- **Data insights:**
Multi-vendor connections including kubernetes
- **Better AI:**
Search billions of records < second
- **Optimize data workflows:**
Policy based automation and auto tag data
- **Data security and compliance**
Discover security, compliance and governance issues before they become problematic
- **Business Value**
Link data to IBM Watson solutions and IBM Cloud Pak for Data

Comprehensive search and AI analysis with a data catalog and policy engine



IBM Spectrum Discover



IBM Spectrum Discover

IBM Spectrum Discover is a multi-source data catalog that automatically and continuously indexes objects and files whenever changes are made using the metadata in real-time. The result is a powerful and customizable database with a user-friendly interface that allows users to locate and identify the most relevant data regardless of its type or location. Using either a simple SQL query command or actionable API scripts or commands, users are empowered with comprehensive insight into the data in a fast and efficient manner. Spectrum Discover can also be used to create custom tags and policy-based workflows to orchestrate content inspection and activate data in artificial intelligence (AI), machine learning (ML), and analytics workflows. Spectrum Discover can be used for faster AI analysis, compliance classification, image and video indexing, identifying personal data, AI data pipeline integration, real-time data discovery, and providing new insights to optimize data and find bad or duplicate data. Data sources include IBM Spectrum Scale, IBM COS, AWS S3, NFS or SMB data sources including Netapp and Isilon and Red Hat (Ceph and OCS).

Why IBM?

IBM Spectrum Discover becomes the brain behind this data orchestration by providing ability to select the right dataset for movement based on its metadata indexing. IBM Spectrum Scale data mover combined with Spectrum Discover is designed to offer a comprehensive data orchestration solution that can serve use cases like active archiving, data migration, caching or AI acceleration. This allows IBM Spectrum Scale and ESS to deliver value for not just traditional on-prem AI / HPC / HPDA workloads but also hybrid cloud and edge computing workloads.

Next steps

- [IBM Spectrum Scale web page](#)
- [IBM Spectrum Discover web page](#)
- [IBM Cloud Object Storage web page](#)

For more information

<https://www.ibm.com/it-infrastructure/storage/ai-infrastructure>

© Copyright IBM Corporation 2020.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at <https://www.ibm.com/legal/us/en/copytrade.shtml>, and select third party trademarks that might be referenced in this document is available at https://www.ibm.com/legal/us/en/copytrade.shtml#section_4.

This document contains information pertaining to the following IBM products which are trademarks and/or registered trademarks of IBM Corporation:
IBM®



All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.