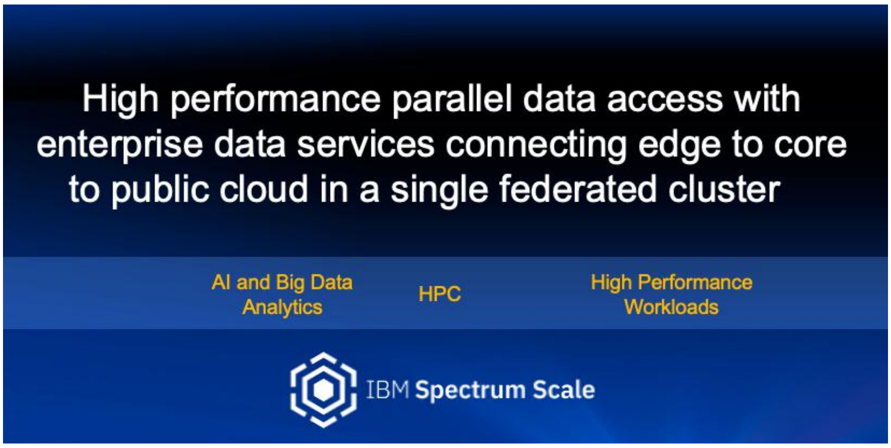
IBM Spectrum Scale

Performance and simplicity for the hybrid cloud

Highlights

- Consolidate storage across the hybrid cloud
- Achieve new operational efficiency and cost effectiveness
- Help lower data retention cost up to 90% with policy-driven automation
- Improve application performance with scale-out and flash-based acceleration
- Enable collaboration and efficient sharing of resources among global teams
- Transparently tier/achive to multi-cloud or local tape or low-cost media
- Built in enterprise security for compliance and business requirements
- Containerized for modern applicaitons
- Deploy in public cloud to create a true hybrid cloud data access



IBM Spectrum Scale has a simple message

Enterprises and organizations are creating, analyzing and keeping more data than ever before. Those that can deliver insights faster while managing rapid infrastructure growth are the leaders in their industry. To deliver those insights, an organization's underlying storage must support both new-era, big-data applications and traditional applications with high performance, reliability and security. To handle massive unstructured data growth, storage must scale seamlessly while matching data value to the capabilities and costs of different storage tiers and types. IBM Spectrum Scale meets these challenges and more. It is a high-performance parallel file system for managing data at scale with the distinctive ability to perform archive and analytics in place.

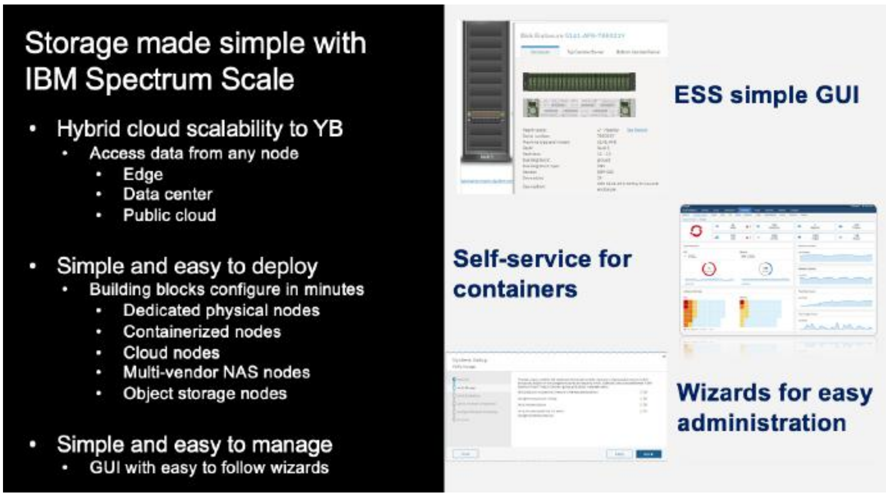
Part of the IBM Spectrum Storage family of solutions, IBM Spectrum Scale is an enterprise-grade parallel file system that provides superior resiliency, scalability and control. Based on IBM General Parallel File System (GPFS), IBM Spectrum Scale delivers scalable capacity and performance to handle demanding data analytics, content repositories and technical computing workloads. Storage administrators can combine flash, disk, cloud, and tape storage into a unified system with higher performance and lower cost than traditional approaches. With thousands of customers and nearly 20 years of demanding production deployments, IBM Spectrum Scale is a file system that can adapt to both application performance and capacity needs across the enterprise. By including IBM Spectrum Scale in their software-defined infrastructure, organizations can streamline data workflows, help improve service, reduce costs, manage risk and deliver business results today while positioning the enterprise for future growth.

IBM Spectrum Scale enables the unification of virtualization, analytics, file and object use cases into a single scale-out storage solution. It can provide a single namespace for all of this data, offering a single point of management with an intuitive graphical user interface (GUI). Using storage policies transparent to end users, data can be compressed or tiered to tape or cloud to help cut costs; data can also be tiered to high-performance media, including server cache, to lower latency and improve performance. Intelligent caching of data at remote sites ensures that data is available with local read/write performance across geographically distributed sites using Active File Management (AFM).

Simplified data management at scale

IBM Spectrum scale is a parallel file system, where the intelligence is in the client and the client spreads the load across all storage nodes in a cluster, even for individual files. In traditional scale-out network-attached storage (NAS), one file can only be accessed through one node at a time by an individual client, limiting performance and scalability. By contrast, the IBM Spectrum Scale architecture allows it to seamlessly handle tens of thousands of clients, billions of files and yottabytes of data.

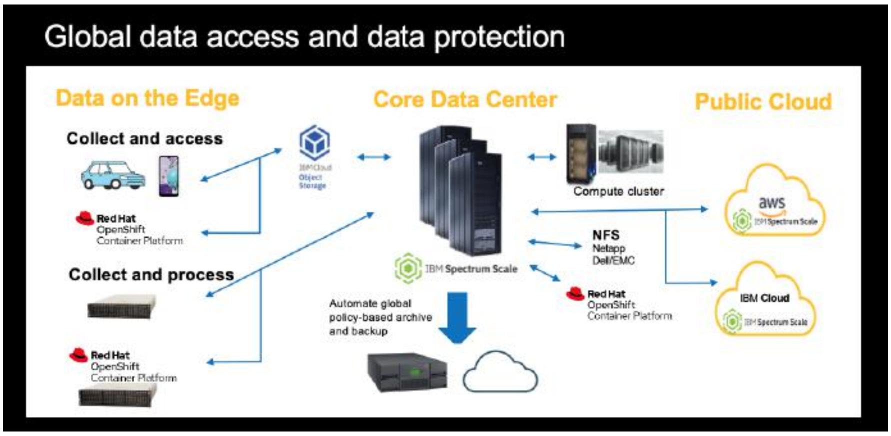
IBM Spectrum Scale allows different applications or services to access the same data without moving or altering it. Data can be written and retrieved as either files or objects. Rather than use a copy and change gateway, IBM Spectrum Scale supports both protocols natively for higher performance and simplified administration. The common storage layer enables most IBM Spectrum Scale features, including authentication, encryption and tiering, for both object and file storage.



Simple to manage and configure

IBM Spectrum Scale includes integrated management tools and an intuitive GUI to help manage data at scale. The system can span multiple storage environments and data centers to eliminate data silos and “filer sprawl.” It can automatically spread data across multiple storage devices—optimizing available storage utilization, reducing administration and delivering high performance where needed. IBM Spectrum Scale has multiple deployment options and configurations to incorporate current NFS filers, block storage and storage-rich servers into a global namespace with universal access. Its file system supports interfaces for file (POSIX, NFS, CIFS), object (S3, SWIFT) and Hadoop Distributed File System (HDFS) for in-place analytics. IBM Spectrum Scale is the caretaker of business-critical data with the ability to replicate, encrypt, compress, and distribute data across different hardware platforms, systems and data centers.

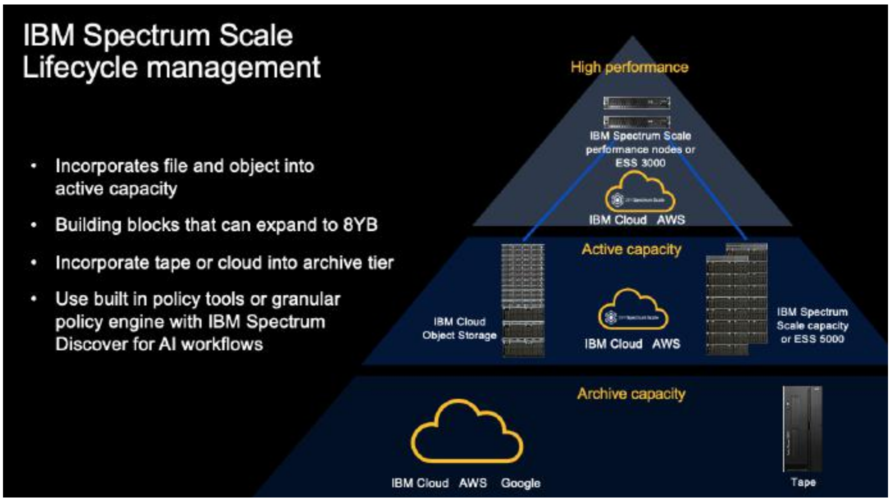
Global collaboration



One global file system with built-in data protection

IBM Spectrum Scale enables low-latency read and write access to data from anywhere in the world using AFM distributed routing and advanced caching technology. AFM expands the IBM Spectrum Scale global namespace across geographical distances, providing fast read and write performance with automated namespace management. As data is written or modified at one location, all other locations get the same data with minimal delays. AFM leverages the inherent scalability of IBM Spectrum Scale, providing a high-performance, location-independent solution that masks network failures and hides wide-area latencies and outages. These game-changing capabilities accelerate project schedules and improve productivity for globally distributed teams.

Advanced data management



Data lifecycle management

IBM Spectrum Scale can help improve performance, lower costs, add resiliency and simplify collaboration with algorithmic and policy-driven data movement, copying and caching. IBM Spectrum Scale catalogs data across multiple storage pools, including the cloud. It tracks usage profiles, storage latency and a broad range of standard and custom metadata from which data movement policies can be constructed.

Armed with the knowledge of the data usage and the underlying storage, IBM Spectrum Scale curates data across multiple tiers of storage, including tape and cloud. The powerful, data-aware intelligence engine can create optimized tiered storage pools by grouping devices—flash, solid-state drive (SSD), disk or tape—based on performance, locality or cost. Migration policies transparently move data from one storage pool to another without changing the file's location in the directory structure. Automated analysis of data usage patterns can help administrators pull data back up to higher performance tiers as needed. The information lifecycle management toolset built into IBM Spectrum Scale helps simplify data management by enabling additional control over data placement. The toolset includes storage pooling and a high-performance, scalable, rule-based policy engine.

IBM Spectrum Scale supports Hadoop workloads and HDFS—without requiring any changes to applications. With IBM Spectrum Scale HDFS Transparency connector, multiple IBM Spectrum Scale clusters or another HDFS repository can be federated into a single HDFS instance. This reduces the need to move data, simplifying the deployment and workflow of Hadoop, Apache Spark and related packages.

End-to-end data availability, reliability and integrity

IBM Spectrum Scale provides system scalability, very high availability and reliability with no single point of failure in large-scale storage infrastructures. Administrators can configure the file system so that it automatically remains available if a disk or server fails. The system is designed to transparently fail over metadata operations and other services, which can be distributed throughout the entire cluster. For additional reliability, IBM Spectrum Scale supports snapshots, synchronous and asynchronous replication, and asynchronous error diagnosis while affected input/output (I/O) operations continue. IBM Spectrum Scale offers protection of data at rest and secure deletion with file-level encryption. It can encrypt data in flight or at rest with independent key management that integrates with leading enterprise key management systems. IBM Spectrum Scale can be part of the enterprise disaster-recovery plans with the ability to quickly back up, copy and restore data as needed. With automatic fail-over and intelligent fail-back, IBM Spectrum Scale keeps businesses and organizations up and running.

Using Transparent Cloud Tiering, public cloud or on-premises IBM Cloud Object Storage can be added as a tier of storage. Ideal for adding active archive storage pools or taking advantage of storage as a service, the design leaves end users unaffected by data movement to and from the cloud. IBM Spectrum Scale manages the metadata, movement and caching to seamlessly tier to and from any Amazon S3 or OpenStack Swift storage without the inconvenience, complexity and performance hit of adding a separate cloud or object storage silo.

IBM Spectrum Scale Innovations

Innovation within the IBM Spectrum Scale platform continues at a brisk pace. IBM Spectrum Scale includes these new capabilities and enhancements, among many others:

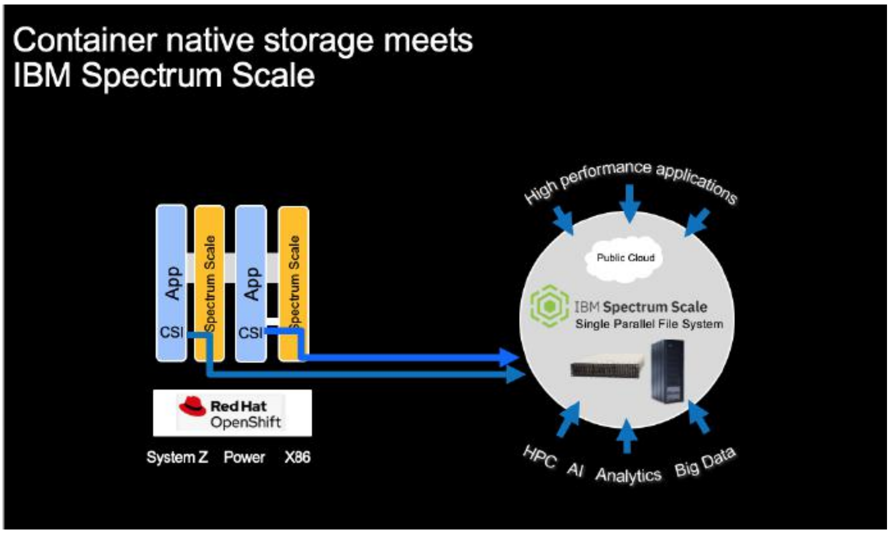
- A new level of storage performance and efficiency
  - Dramatic improvements in I/O performance
  - Significantly reduced inter-node software path latency to support the newest low-latency, high-bandwidth hardware such as NVMe
  - Improved performance for many small and large block size workloads simultaneously from new 4 MB default block size with variable sub-block size based on block size choice
  - Improved metadata operation performance to a single directory from multiple nodes
- Simpler, more powerful system administration
  - Faster and simpler out-of-the-box experience with tuning of an additional 20+ communication protocol and buffer management parameters now handled automatically, aiding setup for optimal performance
  - Enhanced GUI features for many capabilities, including performance, capacity, network monitoring, AFM (multi-cluster management), Transparent Cloud Tiering, and enhanced maintenance and support, including interaction with IBM remote support
- Security enhancements for management of sensitive data
  - New file audit logging capability to track user accesses to filesystem and events supported across all nodes and all protocols
  - Parsable data stored in secure retention-protected fileset
- A third-generation Representational State Transfer (REST) application programming interface (API) that allows modern, highly automated, cloud-ready management and monitoring, along with remote operation of IBM Spectrum Scale clusters
- Expanded Transparent Cloud Tiering support with better granularity of data and support for multiple cloud accounts and multiple containers
- Better backup/restore via new support of snapshots, better performance via improved load balancing, and better production vs. recovery prioritization
- Faster implementation and deployment in IBM Elastic Storage Server environments with enhanced Call Home configuration and network pre-checking

IBM Spectrum Scale in Public Cloud

IBM Spectrum Scale in AWS

IBM Spectrum Scale is available in AWS marketplace and includes a supported reference architectures. IBM Spectrum Scale on AWS is configured in a high-performance tier in AWS for AI and other high-performance applications and can be configured with the ESS 5000 and ESS 3000 or any IBM Spectrum Scale configuration.

IBM Spectrum Scale for Kuberneties Containers



Containerized for simplicity

IBM Spectrum Scale has containerized the storage access inside a Kuberneties container. Storage deployed inside containers, alongside applications running in containers, is an important innovation that benefits both developers and administrators. By containerizing storage services and managing them under a single management plane such as Kubernetes, administrators have fewer housekeeping tasks to deal with, allowing them to focus on more value-added tasks. Developers benefit by being able to provision application storage themselves that's both highly elastic and developer-friendly. With IBM Spectrum Scale container native storage access there are no limits of localized disks and performance is not effected by the application server but is optimized by external storage resources. IBM Spectrum Scale provides most of the benefits of container native storage without the limits of using only local disk and application server performance.

More than a product - a hybrid cloud data modernization strategy

IBM Spectrum Scale

IBM Spectrum Scale has a simple message, Secure access. Anywhere. Enterprise data services. Everywhere. Hybrid cloud. Anyone. IBM Spectrum Scale is the center of IBM Storage for Data and AI information architecture. Its a global hybrid cloud file system with parallel access and is the glue that creates a hybrid cloud storage solution for an AI information architecture for any enterprise or organization.

IBM Spectrum Scale at a glance

Why IBM?

The message for IBM Spectrum Scale is simple: High performance parallel data access with enterprise data services connecting edge to core to public cloud in a single cluster. IBM Spectrum Scale is a marketing leading storage solution for AI, Analytics, HPC and any high performance or capacity file storage workload.

For more information

To learn more about IBM Spectrum Scale, please contact your IBM representative or IBM Business Partner, or visit: http://www.ibm.com/products/spectrum-scale