IBM Storage Scale

Software-defined storage for building a global data platform for AI / analytics / hybrid cloud

Highlights
Global data abstraction services provide connectivity across multiple data sources and locations

Support for NFS and SMB file protocols, S3 object-based protocol, HDFS, POSIX, NVIDIA GPU Direct Storage, Container Native and Container CSI interfaces

Robust data resilience services to help protect against ransomware and other threats

Also available: IBM Storage Scale System appliance for rapid deployment and building-block expandability

Organizations around the world face many challenges today as they reorganize their data resources to modernize and to take advantage of the opportunities presented by artificial intelligence (AI):

- *More data*: Enterprises are creating and storing more data than ever, a trend that is not expected to subside any time soon.
- *More places*: Modern data strategies focus on distributed storage architectures for optimal performance, cost, and data resilience
- *More formats*: Structured data (i.e.: SQL databases), semi-structured data (web pages, social media posts and log files) and unstructured data (text, video, audio, and IoT sensors)

Much of this is *unstructured* data created in GenAI / AI / ML workloads, analytics, data lakes, IoT, cloud-native applications, and backup and archive applications. The data from these data-intensive apps must be stored in distributed file and object storage systems to make it accessible to geographically dispersed applications, services, and devices.

IBM Storage Scale is designed to address these requirements, with global data abstraction services that provide connectivity from multiple data sources and multiple locations to bring together data from IBM and non-IBM storage environments. It's based on a massively parallel file system and can be deployed on multiple hardware platforms including x86, IBM Power, IBM zSystem mainframes, ARM-based POSIX client, virtual machines, and Kubernetes.

The main use cases for Storage Scale have traditionally been high-performance computing (HPC), data-intensive technical computing, big data, and analytics, but these applications have recently been complemented by a flood of machine learning and other AI workloads.

Storage Scale also provides the foundation for IBM Fusion and the IBM Fusion HCI appliance, which enable organizations to quickly and flexibly deploy OpenShift applications and IBM watsonx.

The 2023 Gartner Magic Quadrant for Distributed File Systems and Object Storage lists IBM Storage Scale in the leader's quadrant for the eighth consecutive year.

According to Gartner, "distributed file systems and object storage platforms serve as the foundation for where the majority of enterprise unstructured data is stored and analyzed to drive business insights and outcomes."

Distributed File and Object Storage

The IBM Storage Scale global data platform is built around four key sets of capabilities:

- Data access services
- Data abstraction and acceleration services
- Data orchestration services
- Data resilience services.

Data access services

Data access services enable organizations to maintain a single source of truth for all their data, even with multiple workloads accessing the same data sets simultaneously across multiple locations. Storage Scale provides robust multi-protocol data access, with support for the NFS and SMB file protocols used in Linux and Windows environments respectively, as well as the AWS S3 object-based protocol and HDFS. It also supports POSIX, NVIDIA GPU Direct Storage (GDS), and Container Native and Container CSI interfaces.

This architecture provides multiple paths to the same data and is designed for linear scaling of performance and capacity, with sub-millisecond latency and hundreds of GB/s throughput and tens of millions of IOPs while addressing datasets in the exabyte range. It helps eliminate data silos because multiple application APIs can access the same data, and AI and analytics workloads can use output from one application to drive results to another.

Data abstraction and acceleration services

IBM Storage Scale provides transparent data abstraction with active file and object management, enabling tiering and sharing of data across clusters and the cloud. This provides data centers with faster access to remotely stored data by transparently caching remote data locally as needed, increasing application agility across edge, core, and cloud environments. It also helps organizations rapidly scale by adding storage capabilities wherever they can be most efficiently deployed, without regard to geography.

Deploying Storage Scale can also help organizations protect their existing infrastructure investments and lower their costs moving forward with an open ecosystem of storage options that leverage multi-vendor and multi-cloud resources.

Data orchestration services

Data orchestration is essential both for optimizing performance across an organization's entire IT environment and for controlling data storage costs. Storage Scale provides outstanding data management and orchestration capabilities thanks to its global parallel file system with advanced file management (AFM), which can integrate with a wide range of external data services and sources.

Storage Scale can automate many routine operations with built-in policy tools and can be integrated into IBM Storage Discover for AI workflows. It also provides several cost-effective options for archiving data to tape or to the cloud.

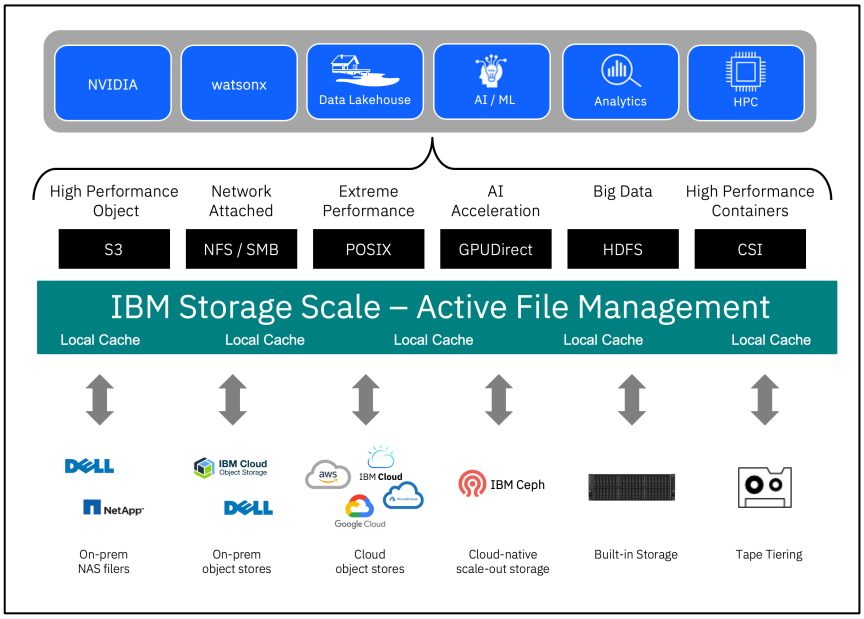

Figure 1. IBM Storage Scale provides a global data platform for your organization's geographically dispersed devices, data sources, and workloads.

Data resilience services

Data resilience is fundamental to the IBM Scale architecture, with support for a variety of multi-site configurations using either synchronous or asynchronous communication.

The software can be deployed as software-defined storage, as an integrated hardware and software solution, or as a cloud service. It can be deployed on NAS, SAN, iSCSI, NVMeoF, or storage rich servers, and can be pooled to create a common namespace with data protection using erasure coding or replication. These configurations can be clustered together into a larger cluster to create a global data platform with exabytes of capacity.

To enhance data resilience, a single Storage Scale stretch cluster can be configured using nodes and storage across two data centers. The cluster is essentially stretched between the two sites, connected by a WAN (wide area network). File systems in such a cluster are available to systems at both sites and can be actively used concurrently by both sites.

The file systems can use "failure group" replication to ensure both sites have a current replica of all the data, so that one site always remains active, even if the other site or link fails. The result is a highly available file system with active-active synchronous replication. With AFM, any single Storage Scale cluster can be configured using nodes and storage across two or more data centers.

IBM Storage Scale also includes IBM Safeguarded Copy capability, which can create isolated immutable snapshots of data on a regular schedule so that operational data can be rapidly recovered in the event of a cyber-attack or other potential data loss event.

IBM Storage Scale Editions

IBM Storage Scale offers different editions that are based on functional levels.

IBM Storage Scale Developer Edition
Available on Red Hat Enterprise Linux on x86_64. This edition provides all the features of the Data Management Edition but is limited to 12 TB per cluster. It is available at no charge at https://www.ibm.com/products/storage-scale to enable customers to use IBM Storage Scale in test environments. It cannot be used in production environments.

IBM Storage Scale Data Access Edition
This edition provides the core IBM Storage Scale functions and is available on AIX®, Linux®, and Windows. On AIX and Linux, the available features include Information Lifecycle Management (ILM), Active File Management (AFM), and Clustered NFS (CNFS). On Windows, the available features include limited Information Lifecycle Management (ILM).

On Red Hat Enterprise Linux 7.x and 8.x, SLES 15, Ubuntu 20.04, and Ubuntu 22.04, the available features include the ability to enable and use the additional protocol functionality integration (NFS and SMB). On Red Hat Enterprise Linux 8.x, the object protocol functionality is also supported.

IBM Storage Scale Data Management Edition
Available on AIX and Linux. This edition provides all the features of the Data Access Edition and certain additional features, per the table on Page 5.

IBM Storage Scale Erasure Code Edition
This edition provides identical functionality to IBM Storage Scale Data Management Edition plus the support for storage rich servers. IBM Storage Scale Erasure Code Edition provides network-dispersed erasure coding, distributing data and metadata across the internal disks of a cluster of servers. This allows IBM Storage Scale to use internal disks as reliable storage with low overhead and high performance.

Figure 2. IBM Storage Scale System 6000 is a hardware appliance that allows you to deploy IBM Storage Scale on thousands of nodes with TB/s performance, low latency, and tens of millions of IOPS per node.

IBM Storage Scale System

Storage Scale is also available as an appliance, IBM Storage Scale System, for streamlined, rapid deployment complete with IBM support services. This option is designed for organizations wanting to build high-performance global data storage capabilities in their own data centers or co-location facilities.

Optimized for AI / NVIDIA workloads
Storage Scale System 6000 is the simplest and fastest way to deploy a global data platform for AI and NVIDIA GPU infrastructure, with exceptional performance that includes GPUDirect Storage support.

For more information, visit ibm.com/products/storage-scale-system.

Features in IBM Storage Scale Editions

Feature	Data Access Edition	Data Management Edition	Erasure Code Edition
Multi-protocol scalable file service with simultaneous access to a common data set	√	√	√
Facilitate data access with a global namespace, massively scalable file system, quotas and snapshots, data integrity and availability, and filesets	√	√	√
Simplify management with a GUI	√	√	√
Improved efficiency with QoS and compression	√	√	√
Create optimized tiered storage pools based on performance, locality, or cost	√	√	√
Simplify data management with information lifecycle management tools that include policy-based data placement and migration	√	√	√
Enable worldwide data access using AFM asynchronous replication	√	√	√
Asynchronous multi-site disaster recovery		√	√
Protect data with native software encryption and secure erase, NIST compliant and FIPS certified		√	√
File audit logging		√	√
Watch folder		√	√
Erasure coding	Storage Scale System only	Storage Scale System only	√
IBM Fusion Data Cataloging service		√	√

For more information
To learn more about IBM Storage Scale, contact your IBM representative or IBM Business Partner, or visit ibm.com/products/storage-scale.

Partner für DACH:

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
May 2024