IBM Storage Ceph

One software-defined storage solution for block, file, and object data.

Highlights

High-availability fault-tolerant architecture engineered to support petabytes of data and tens of billions of objects

Select the deployment option that best suits your IT strategy–as a service, software only, or Ceph-certified servers

Integrates with IBM watsonx.data to provide a data lakehouse for AI workloads

Supports NVMe/TCP, which makes it a competitive block storage solution in many virtualized environments

IT leaders today are grappling with the challenges posed by explosive data growth, increasing complexity, and rising storage costs. These challenges are compounded by the complexity of managing multiple protocols and storage systems at a time when skilled IT operators are an increasingly scarce and expensive resource.

IBM Storage Ceph is enterprise-class software-defined storage for data-intensive applications, and it's precisely aimed at meeting these challenges. It's based on a scale-out architecture with an autonomous, distributed object store that is self-healing and self-managing, providing a highly elastic and reliable storage experience with minimal user intervention. Storage Ceph is massively scalable – it's engineered with no single point of failure and can scale to support petabytes of data and tens of billions of objects.

Storage Ceph consolidates support for block, file, and object storage protocols to help organizations eliminate data silos and deliver a cloud-like experience while retaining the cost benefits and data sovereignty advantage of on-premises IT. Data stored in Storage Ceph can be accessed via block protocols (for structured data) and via AWS S3 compatible REST APIs (for unstructured data), making it ideal for both application storage and as back-end storage for data lakehouses. It also supports NVMe/TCP, which makes it a competitive alternative in many virtualized environments.

Storage Ceph is entirely software-defined – it delivers enterprise-grade scalability and flexibility without reliance on proprietary hardware. Designed for data-intensive workloads and cloud-native applications, it can be deployed in a variety of options to suit your IT strategy and business needs– as a service, as software only on commodity x86 hardware, and on Ceph-certified servers, with the potential for significant cost savings when compared with dedicated storage hardware systems.

Key Features and Benefits

Key capabilities of IBM Storage Ceph include:

- **Scalability**: Non-disruptively grow from as few as four nodes to thousands, capable of addressing billions of pieces of information; scale capacity and throughput to adapt to evolving storage requirements.
- **Affordability**: Storage Ceph is software-defined storage built with open standards, which helps keep CAPEX and OPEX costs in line with underlying industry-standard hardware prices.
- **Availability**: Using a fault tolerant architecture, data is distributed across multiple disks over multiple servers into a storage cluster with no single point of failure, helping ensure that data is always accessible.
- **Data services**: Includes data reduction for disk usage optimization, partial or complete reads/writes with atomic transactions, replication and erasure coding for data protection, policy-based optimization, and much more.
- **Optimization**: Self-healing and self-managing rebalancing data distribution throughout the cluster handles failures without interruption, automatically recovering to the desired predefined data resilience level.
- **Foresight**: Built-in storage analytics help organizations monitor capacity utilization and growth so they can plan for near-term and long-term capacity needs.
- **Security**: Certified object lock for write-once-read many (WORM) data governance and protection; optional FIPS 140-2 certified cryptography; key management service integration, and server-side encryption.
- **Automation**: Storage Ceph automates many processes, such as data distribution and replication, data placement, and failure recovery.
- **Easy management**: The Storage Ceph Dashboard provides point-and-click interactivity for common tasks like managing and adding storage capacity and configuring storage services and access for file, block, and object data. A command line interface (CLI) is available for more advanced users, plus REST APIs for lights-out data center operations.

Deployment Options

As a managed service: IBM Storage Ceph as a Service provides the flexibility and simplicity of cloud storage in your datacenter, managed by IBM experts. Designed to simplify and expedite the deployment of Storage Ceph clusters for faster time-to-value in all your projects. With Storage Ceph as a Service, IT teams can focus on business priorities and manage applications, not infrastructure.

Software only: A highly customizable and scalable way to leverage new or existing infrastructure, including commodity x86 hardware, fine-tuning your systems to meet workload requirements with the potential for significant cost savings when compared with dedicated storage hardware systems.

Ceph-certified servers: Easy building-block infrastructure that can be rapidly deployed and expanded, IBM Storage Ready Nodes provide an integrated solution – A simple, flexible and cost-effective way to deploy storage on purpose-built servers that have been optimized, tested and certified to work with IBM Storage Ceph.

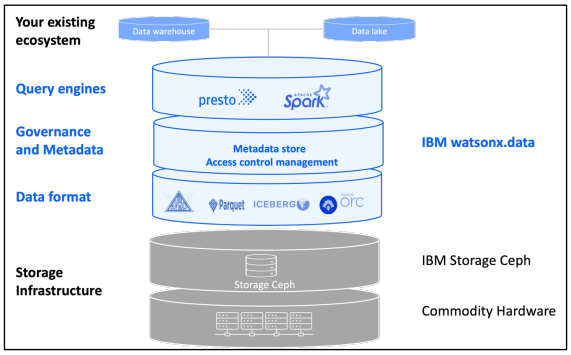

Figure 1 – IBM Storage Ceph and IBM watsonx.data together provide a performant data lakehouse designed to accelerate AI workloads.

Storage Ceph Use Cases

A Data Lakehouse for AI Workloads
As AI becomes integrated into more and more routine business processes, organizations increasingly require access to massive and continually expanding repositories of highly diverse datasets. And they need data management tools for control over the content of the datasets, where the data's stored, who has access, the regulatory and compliance requirements, and more.

The recommended solution these days is a data lakehouse, a modern data architecture that combines the key benefits of data lakes (large repositories of raw data in its original form) and data warehouses (organized sets of structured data). Now organizations can deploy quickly yet still scale massively by combining IBM Storage Ceph with IBM watsonx.data, an open, hybrid, governed, fit-for-purpose data store.

Together, IBM Storage Ceph and IBM watsonx.data provide a highly scalable data lakehouse that allows organizations to:
- Consolidate multiple data types into a single elastic repository that expands capacity online as requirements grow;
- Access data across through a single-entry point with a shared metadata layer leveraging open data and open table formats;
- Connect to storage and analytics environments in minutes and enhance trust in data with built-in governance, security, and automation;
- Reduce data lake and lakehouse costs with unified data storage options and fit-for-purpose compute engines that are clustered and capable of scaling automatically.

Cloud-native Applications Using S3 API
As a data lake, IBM Ceph Storage delivers massive scalability and high availability to support demanding multitenant analytics and AI/ML workloads. With high fidelity compatibility to the Amazon AWS S3 interface, applications can access their storage with the same application API in public, private, or hybrid clouds.

Object Storage-as-a-Service Use Cases
Storage Ceph is used to deliver object storage as-a-service for applications such as analytics and data pipelines, online archives, backup targets, disaster recovery, and enterprise file sharing. Many such applications now use object storage because of its cost-efficient and secure design, concurrent multi-user access, and the ability to scale to very large capacity and performance requirements. Storage Ceph is a great fit in these environments because it supports concurrent data access from multiple applications, multiple systems, and multiple locations in secure repositories or buckets.

Virtual Private Clouds and Container Management Platforms
Ceph is a proven cloud storage solution that has evolved into the leading storage platform for OpenStack-based virtual private Clouds (VPCs) and for the OpenShift Container Platform.

IBM Storage Insights

IBM Storage Ceph licenses include IBM Storage Insights, software that provides a birds-eye view of storage resources from the server, application, network, and file system perspective. In a few easy steps, users can set up application and department views for email, SAP workloads, recovery site storage, and other applications. These views enable reporting on application or departmental use of storage, as well as opportunities to optimize performance and control costs.

Storage Insights provides operators with an at-a-glance view of the entire storage environment including storage networking. It helps them quickly distinguish between healthy storage and systems requiring attention while also providing visibility into potential capacity and performance issues.

IBM Storage Ceph Editions

IBM Storage Ceph is available in two editions:

- **IBM Storage Ceph Pro Edition**. This edition includes all the components to enable a massive scalable object storage solution, including the Ceph software components and the predictive analytics add-on capabilities of IBM Storage Insights, the software license, and support. Clients purchasing this edition must provide their own license and support for Red Hat Enterprise Linux.
- **IBM Storage Ceph Premium Edition**. This edition includes all the components of the Pro addition and includes the Red Hat Enterprise Linux (RHEL) server operating system software license and support.

IBM Storage Ceph Protocol Support

Object Storage	**AWS S3**	**Supports all common AWS S3 bucket and object API calls**
		Advanced S3 features such as object versioning, object lock, S3-select with table format support (Apache Parquet, CSV and JSON)
	OpenStack Swift	Compatible with the OpenStack Swift object storage API
Block Storage	NVMe/TCP	Supports NVMe/TCP block protocol. Supports VMware ESXi 7.0U3 and 8.0+. Supports latency-sensitive applications running on operating systems that support NVMe/TCP
	Ceph RBD	Block storage access through the native Ceph RBD client part of RHEL or OpenStack. Used for KVM/QEMU virtual machine workloads and native Linux applications requiring TCP/IP-accessible block storage
File Storage	NFSv3 and NFSv4	Export of CephFS volumes and sub-volumes via NFSv3 and NFS v4.1
	NFSv4 gateway to object storage	Support for data ingest & export of object storage data via NFS shares
	CephFS	Export of CephFS volumes and sub-volumes via native Linux kernel or FUSE client running on RHEL or other Linux distribution
Container Storage	Kubernetes CSI-drivers	Supports CephFS and Ceph RBD CSI-drivers to provide persistent storage for containers. Also available as a fully integrated solution with IBM Storage Fusion.

For more information

To learn more about IBM Storage Ceph, contact your IBM representative or IBM Business Partner or visit ibm.com/products/storage-ceph

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
March 2025