Use Cases for IBM Storage Ceph

Optimize the enterpise with flexibility and accelerate your cloud native data modernization

Highlights

- Single efficient, unified platform for object block and file storage
- Integrated with the IBM global data platform for a comprehensive data strategy
- Use off the shelf servers and hardware as you deploy with speed on your resources
- Open Source, massivly scalable, software defined storage for investment protection

As data requirements outpace current storage resources and the cost of storing the data continues to grow, organizations need reliable and flexible software-defined storage that can deploy quickly on any server and be flexible for multiple applications. Unstructured data brings challenges to an organization and 95% of business cite managing unstructured data as a problem.1 Legacy storage technologies cannot keep up with demand. The problem of managing costs while maintaining performance at scale is exacerbated by dwindling storage specialist skills. Organizations need a way to scale their storage without increasing cost, head count, or security issues while adapting to modern data management modalities that accelerate time to value.

IBM Storage Ceph provides an open, scalable and software defined multi-protocol storage solution designed to consolidate data anywhere and with the global data platform consolidate data everywhere.

Efficiently scaling to support petabytes of data and tens of billions of objects[2], the storage platform is designed to be self-healing and self-managing for many tasks and is also engineered with no single point of failure. IBM Storage Ceph is object storage optimized for enterprise, unified for simplicity, and software defined for flexibility.

[1] https://techjury.net/blog/big-data-statistics/#gref

[2] https://www.redhat.com/en/resources/data-solutions-overview

IBM Storage Ceph offers the option to start small, but it can cope with massive installations. It is simple to use and increasingly self-managed, so existing IT teams can manage more storage without needing to add more staff.

With a file, object or block front end and BlueStore (a high-performance storage architecture) back end, IBM Storage Ceph delivers up to double the performance of previous versions.[2] IBM Storage offers a range of reference architectures based on IBM Storage Ceph that are optimized for reliability, cost, and performance to meet any business's needs—even as demands change.

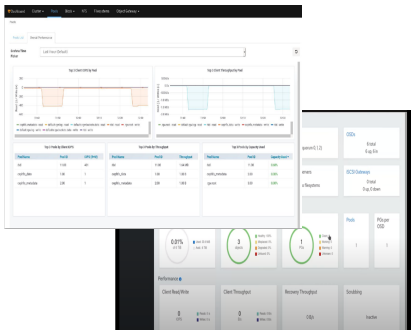
The integrated monitoring dashboard offers graphical visualizations of entire clusters or single components. Users can maintain quality of service with new features like noisy neighbor monitoring, which can help you visualize IOPS, throughput, and latency outliers to pre-emptively identify issues and mitigate them.

IBM Storage Ceph capabilities

Scalability

Businesses need to store more data than ever, but they cannot afford to compromise on performance. IBM Storage Ceph offers reliable performance and improved utilization of cluster hardware built on industry-standard servers and disks.

- Increases scalability, tested with more than 1 billion objects
- Improves performance with BlueStore back end with automatic re-balancing
- Automates workload redistribution as you expand your cluster
- Offers user-enabled dynamic block device expansion
- Provides rolling expansion and hardware updates without requiring downtime



Simplicity

IBM Storage Ceph offers improved monitoring and management tools to reduce the administrative burden. Power users can delegate tasks, enhancing skill development of junior users and increasing operational efficiency.

- Maintains compatibility with existing IBM Storage Ceph or previous Red Hat Ceph installations
- Makes startup quicker and easier with a new install cockpit and lower base configurations
- Accelerates deployment with Red Hat Ansible® Automation Platform
- Monitors entire clusters or single components with ease
- Provides actionable insights with an integrated, on-premises monitoring dashboard
- Cuts deployment time and cost-to-performance ratio with recommended reference architectures

Security

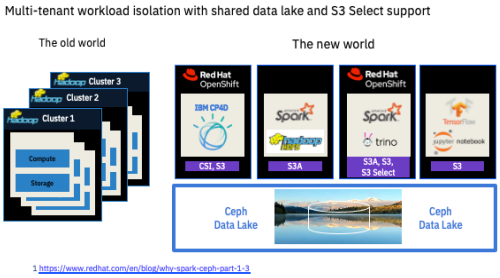
IBM Storage Ceph has features to protect data from malicious and accidental threats, including hardware failures, employee errors, and cyberattacks.

- Provides a continuum of resiliency and data durability options from erasure coding to replication
- Supports at-rest and end-to-end encryption, including National Institute of Standards and Technology (NIST)-approved cryptography
- IBM Storage Ceph has features to protect data from malicious and accidental threats, including hardware failures, employee errors, and cyberattacks.
- Provides a continuum of resiliency and data durability options from erasure coding to replication
- Supports at-rest and end-to-end encryption, including National Institute of Standards and Technology (NIST)-approved cryptography

IBM Storage for AI and Capacity Data

Solution Brief

Where can I use IBM Storage Ceph?



1 https://www.redhat.com/en/blog/why-spark-ceph-part-1-3

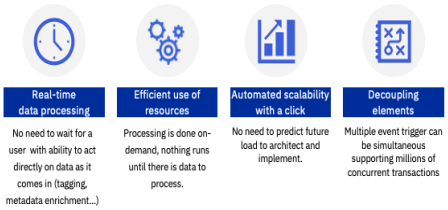
Data analytics and artificial intelligence machine learning (AI/ML)

With its simplicity, flexibility, and cost-efficient characteristics of object platforms like IBM Storage Ceph, object storage has become one of the preferred platforms for collecting, analyzing, and retaining today's growing mountain of diverse and disjointed enterprise data.

IBM Storage Ceph can help organizations of every size are leverage the S3 object interface to underpin a variety of use cases, such as:

- Running cloud-native and mobile applications/web services,
- Archiving data for regulatory and compliance purposes,
- Enabling disaster recovery from the cloud, and
- Establishing enterprise data lakes to facilitate big data analytics, unlocking insights.

Object storage as-a-service

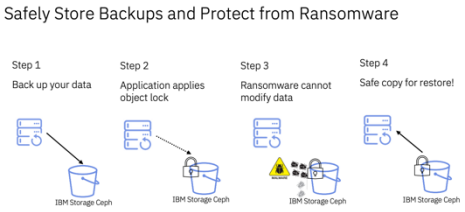


Implementing an object storage service, with proven scalability and performance for both small and large object storage. IBM Storage Ceph is ideal for implementing an object storage service, with proven scalability and performance for both small and large object storage.

Cloud native applications

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

Disaster Recovery and Backup

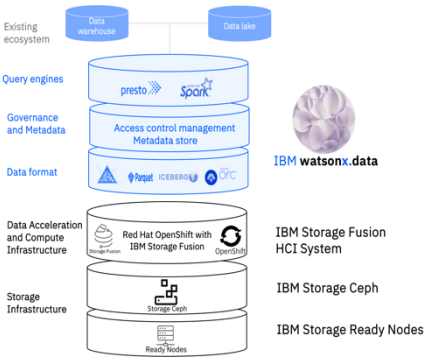


A growing list of software vendors have certified their backup applications with IBM Storage Caph, to serve a wide variety of performance-optimized workloads. Benefits of using IBM Storage Ceph insead of traditional backup servers include:

- In a Storage Ceph system, objects can reside on multiple servers, enabling easy scalability.
- Objects are used to store and retrieve entire unstructured data rather than single blocks or entire files.
- An object can include image files, HTML pages, binary files, videos, executables, and user-generated content.

- Applications do not use namespaces and directory structures, but instead process objects based on their ID and simple HTTP requests like PUT, UPLOAD, GET, and DELETE.
- Applications such as backup software can store and retrieve objects from object storage systems with a simple request, without incurring large amounts of file and block storage overhead.

Storage Infrastructure with IBM watsonx.data



Watsonx.data makes it possible for enterprises to scale AI workloads using all their data with a fit-for-purpose data lakehouse architecture optimized for governed data and AI workloads, supported by querying, governance, and open data formats to access and share data. This is based on open-source technologies, including Presto and Iceberg. IBM Storage Ceph can provide the storage infrastructure for a watsonx.data on-premises deployment. IBM watsonx.data includes 768TB of IBM Storage Ceph software license and support. The easiest way to start with IBM Storage Ready Nodes is with 4 nodes or 7 nodes depending if you desire perfromance or storage efficiency. The 4 node configuration provides the least number of nodes and the fastest perfromance with the ability to loose 2 nodes without incident. The 7 node configuration provides more nodes nodes but yield greater storage efficiency with the ability to loose 2 nodes without incident and the configuration can scale one node at a time as capacity needs grow. With watsonx.data and IBM Storage Ceph, you can access all your data across both databases and data lakes as each configuration can be optimized in the same Storage Ceph cluster. Share large volumes of data through open table formats, such as Apache Iceberg, built for high performance analytics and large-scale data processing and at the same time store large amount of data for other large data set analysis. IBM Storage Ceph also supports multiple vendor open formats for analytic data sets while allowing different engines to access and share the same data, at the same time using tools like Parquet, Avro, Apache Orc and more.

Why IBM?

Data matters. When planning a data strategy for new or existing applications it's easy to focus on compute resources and applications without proper planning for the data that will drive the results for the applications. Our products are all about solving hard problems faster with data. IBM helps customers achieve business value with a clear data strategy. Our strategy is simple, unlock data to speed innovation, de risk data to bring business resilience and help customers adopt value based data to bring cost and energy efficiencies. Value needs to be delivered by connecting the multiple organizational data sources with business drivers to create business value that mean something to the organization. Many organizations focus on a single driver with a storage solution, but the best solution is driven by an infrastructure strategy than can accomplish most if not all the drivers for maximum benefits. Our story is not just about another storage product but is about innovation and a comprehensive storage portfolio that is helping businesses drive more value throughout the organization.

For more information
To learn more about IBM Storage Ready Nodes with any of the IBM Storage software offerings please contact your IBM representative or IBM Business Partner, or visit our web pages and ask to chat with a representative: IBM Storage Ceph Web Page

IBM Storage for AI and Capacity Data
Solution Brief

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the
United States of America
September 2023