



Best Practice Guide

Managing Data at Petabyte Scale with Object Storage

Best Practices for the Modern Data Center



Hewlett Packard
Enterprise



It's estimated that in the next five years, more than a trillion applications will exchange 58 zettabytes of data across more than 100-billion devices.¹

IT must be able to efficiently store and access the right data instantly. But traditional file and block storage architectures are being challenged by today's explosive growth of data. New data sources like social media, the Internet of Things, mobile apps, etc. have existing storage networks bursting at the seams. As data volumes grow, so does the complexity and cost of storing and managing that data.

In today's Idea Economy, a business's ability to grow is directly related to its ability to store, manage, and utilize data. Understanding how to handle this data explosion is no longer a worry specific to storage administrators or CIOs, but should be a top-of-mind concern for every executive.

Enterprise-class storage subsystems are designed to address storage requirements for business-critical transactional data speeds at terabyte (up to low-petabyte) scale. However, it's not uncommon for these unstructured data sources to quickly scale into the tens or even hundreds of petabytes and beyond. The traditional SAN-based storage approach may be unable to meet these needs in efficient and cost-effective ways. Traditional SAN-based storage is just too costly and complex to deploy and manage data at the petabyte scale. New types and sources of data require a modern approach with the latest storage technology.

Object storage

We're used to thinking about data in terms of files or blocks. SAN-based storage systems are organized that way. In object storage, data is treated as an object rather than as a block or file, with metadata and an object ID to uniquely identify each data object. Block-size and file-size limitations that exist in traditional storage systems are eliminated, as are limitations on the number of blocks or files that can be stored.

Unlike SAN (and especially flash SAN arrays), object storage solutions are not designed for mission-critical IOPS and latency performance. Instead, object storage is designed to provide consistent reliability for petabyte-scale data, supporting high bandwidth and parallel workloads with tolerable latency and adequate IOPS. How reliable is object storage? Vendors claim up to fifteen 9s reliability using techniques like erasure coding to rapidly reconstruct data that might have been corrupted or lost.

Object storage solutions take advantage of server-based local storage, allowing IT to easily scale out to hundreds of petabytes today while providing a clear and cost-effective path to expand to exabyte or even zettabyte scale in the future.



Capacity for scale-out file and object solutions will grow at over 40 percent compound annual growth through 2019.

¹ HP (now Hewlett Packard Enterprise) internal research, 2014; hpe.com/issue-no-1-june-2014/why-open-open-cloud-matters



Managing bulk content storage for rapidly growing data “oceans” made of structured or semi-structured data requires the move to scale-out file and object storage. Object storage solutions offer new breakthrough capabilities through software-defined storage that are simple, cost effective, scalable, and can be rapidly deployed.

Best practices for object storage

Object storage provides enterprises with a new alternative for storing both structured and unstructured data at petabyte scale. As with any technology, there are best practices to apply to using and deploying the technology to meet a specific set of business objectives.

Understand the use-cases for object storage

As with most technologies, there are a set of conditions and use-cases where object storage is the best option and others where traditional network-attached storage (NAS) is a better choice.

Object storage is optimal for:

- Web content and file downloads
- Data accessed at file/object-level like videos, medical images, etc.
- Data accessed infrequently like archival data and cold storage
- Massive scale-out data storage environments (10s or 100s of PB)
- Static content (i.e., data that is not actively modified) like sensor data, images, and video files with lower latencies

An increasing number of applications provide native support for object storage.



Some of the key object storage use-cases include:

- Enterprise backup and archive alternatives to tape, with lower latencies to access data and no long-term maintenance of tape-drive systems.
- Active archives of compliance, analysis, surveillance, and other file-and-object data where content scales into the petabytes.
- Communications, media, and entertainment content delivery including applications such as video-on-demand.
- Service providers offering storage as a service for consumer and business offerings such as hosted backup, archiving, disaster recovery, and sync/share.
- New Web and cloud-native services for public and on-premises private cloud deployments where distributed storage architectures are required.
- Emerging applications that tap into the Internet of Things, where unstructured and sensor-derived data need a repository that can scale to billions or even trillions of objects.

An object storage solution is not appropriate for:

- Transactional application data that is frequently modified (i.e., databases and root disks). Block storage that can provide low latency and synchronous data mirroring is more appropriate for these workloads.
- Shared file environments that require a single up-to-date view of a file to all personnel that have access, using file-level locking and file-sharing facilities (e.g., content development, CAD/CAM, actively updated documents, etc.) Network attached storage (NAS) or some scale-out cluster file systems solutions are the best option for these workloads.



Deploy a technology-agnostic, software-defined storage solution



RTL II is known for boosting video transfer speeds tenfold. German TV station RTL II needed to support a rapidly growing video archive expanding by 200 terabytes per year. Using HPE servers, the station was able to assure delivery of the latest broadcast information, increase the amount of online streaming video content, simplify support with a standard hardware design, and reduce transfer time for a 40 MB movie from 60–70 minutes to 6.5 minutes.

Software-defined storage solutions provide enterprise-grade availability and reliability, ease of use, and the ability to scale in a cost-effective manner. There are many viable object storage architectures and software-defined storage management solutions available, including Scality RING, Cleversafe dsNet, Ceph, and OpenStack® Swift. A software-defined storage solution abstracts logical storage from physical resources, allowing IT to select commodity hardware based on efficiency and cost effectiveness rather than support for proprietary architectures. To best meet the needs of object storage, a software-defined storage solution should:

- Run in a technology-agnostic, industry-standard x86 server environment
- Enable storage consolidation that supports mixed applications and workloads
- Provide access through a range of file, object, and OpenStack interfaces, all operating at the same time
- Enable data availability and durability through a variety of data protection mechanisms and the ability to build geo-distributed storage infrastructures
- Provide lower TCO and higher ROI than public-cloud offerings like Amazon S3

One solution that meets all of these criteria is Scality RING. RING can easily scale to hundreds of petabytes capacity and trillions of objects and supports millions of users while enabling global data protection and durability at massive scale. While you can utilize a choice of native file, object, and OpenStack interfaces, the RING serves, manages, and protects all data as objects, enabling global policies around replication, erasure coding, and geo-replication for incredibly efficient and durable storage.



BIGLOBE a leading Japanese Internet and cloud service provider, adopted HPE servers to achieve a low-cost, 2 PB high-capacity storage environment. They deployed 88 TB of disk capacity and 192 GB memory per server node and substantially reduced operating and maintenance costs.

Scality RING deployments support a wide variety of use-cases, including Web and cloud storage, content distribution, distributed computing, enterprise private cloud, and active archiving. Today, Hewlett Packard Enterprise and Scality have an enhanced partnership created to improve customer experience and value and to accelerate adoption of Object Storage solutions for petabyte-scale unstructured and semi-structured data with the Scality RING solution on HPE servers.

Leverage servers optimized for scale-out architectures

The shift in storage paradigms to object storage requires a new style of infrastructure, with traditional rack solutions giving way to ultra-dense nodes capable of storing more data in less physical space at a lower cost-per-GB than general-purpose platforms. Enterprises should deploy high-density servers designed specifically to provide the most scalable, efficient, and cost-effective object storage solutions. Ideally, optimized object storage nodes should provide:

- Higher density storage capacity
- Enhanced I/O bandwidth and memory capacity
- Lower power and cooling requirements
- Modular infrastructure
- Flexible configuration options
- Unlimited scale-out capacity
- Support for a variety of object storage platforms like Scality RING, Cleversafe dsNet, Ceph, and OpenStack Swift

HPE Apollo Systems provide optimized hardware built to meet the specific requirements of object storage solutions and work well with partner applications from Scality, Cleversafe, and other open source vendors. HPE Apollo 4200 Servers and HPE Apollo 4510 Systems are built to deliver performance, reliability, and massive scalability while answering data center challenges of space, energy, and time.



Bring the cloud home

Many enterprises have responded to the crush of data by moving data to the cloud.

Depending on your industry and the types of data stored, public cloud may or may not be the right option. For industries like financial services, healthcare, and many government applications, public cloud can be a non-starter due to strict data privacy and compliance issues. Other industries may be more open to deploying public cloud solutions, but these come with a level of risk:

- Security and compliance risks associated with managing data offsite
- Loss of control of sensitive enterprise data
- Long-term business viability of public-cloud operator
- Costs can rapidly escalate as data volumes grow
- Unplanned expenses from needing to retrieve large amounts of “cold” archive data

Many IT groups face demands from their business to provide storage as a service in the same way they see storage being offered by cloud services like Amazon. Some IT groups have even seen business users create their own “solutions,” bypassing IT to take data to the cloud.

Object storage solutions allow IT to bring the cloud home, providing business users with a “public cloud” experience within their own data center. An on-premises, private cloud content depot gives business users the same ease-of-use and agility they can get from public-cloud providers. Object storage provides an efficient and cost-effective solution, driving down costs through the use of commodity hardware. Software tools that provide ease-of-use and advanced implementation of reliability features like erasure coding can give business users the level of data availability and durability that they’re looking for in a secure, on-premises environment.



Gartner predicts that more than half of Global 1000 companies will store customer-sensitive data in the public cloud by year-end 2016.²

² Gartner, “Hybrid Cloud Storage Can Be an Antidote to Rapid Data Growth,” July 2013

Bottom line

Traditional file and block storage architectures are being challenged by the explosive growth of static unstructured data. Emerging storage architectures, based on object storage, are helping businesses deal with this challenge, providing cost-effective storage solutions that keep up with expanding storage-capacity demands.

Hewlett Packard Enterprise for Object Storage

One popular solution includes Scality RING running on purpose-built **Object Storage and Apollo System**, which provide petabyte scale with high reliability at a cost-point that provides a lower TCO and higher ROI than traditional on-premises storage or public cloud options.

HPE Helion Content Depot reference architecture provides the framework for an on-premises, private cloud solution composed of purpose-built server and network infrastructure along with installation and support services to deliver a lower total cost-of-ownership and higher return on investment.

Get more details on **HPE Object Storage solutions**, and learn more about **how to get started with Software-defined and Object Storage** with help from HPE Consulting.

Learn more at
[**hpe.com/info/Apollo**](http://hpe.com/info/Apollo)



Sign up for updates

★ Rate this document