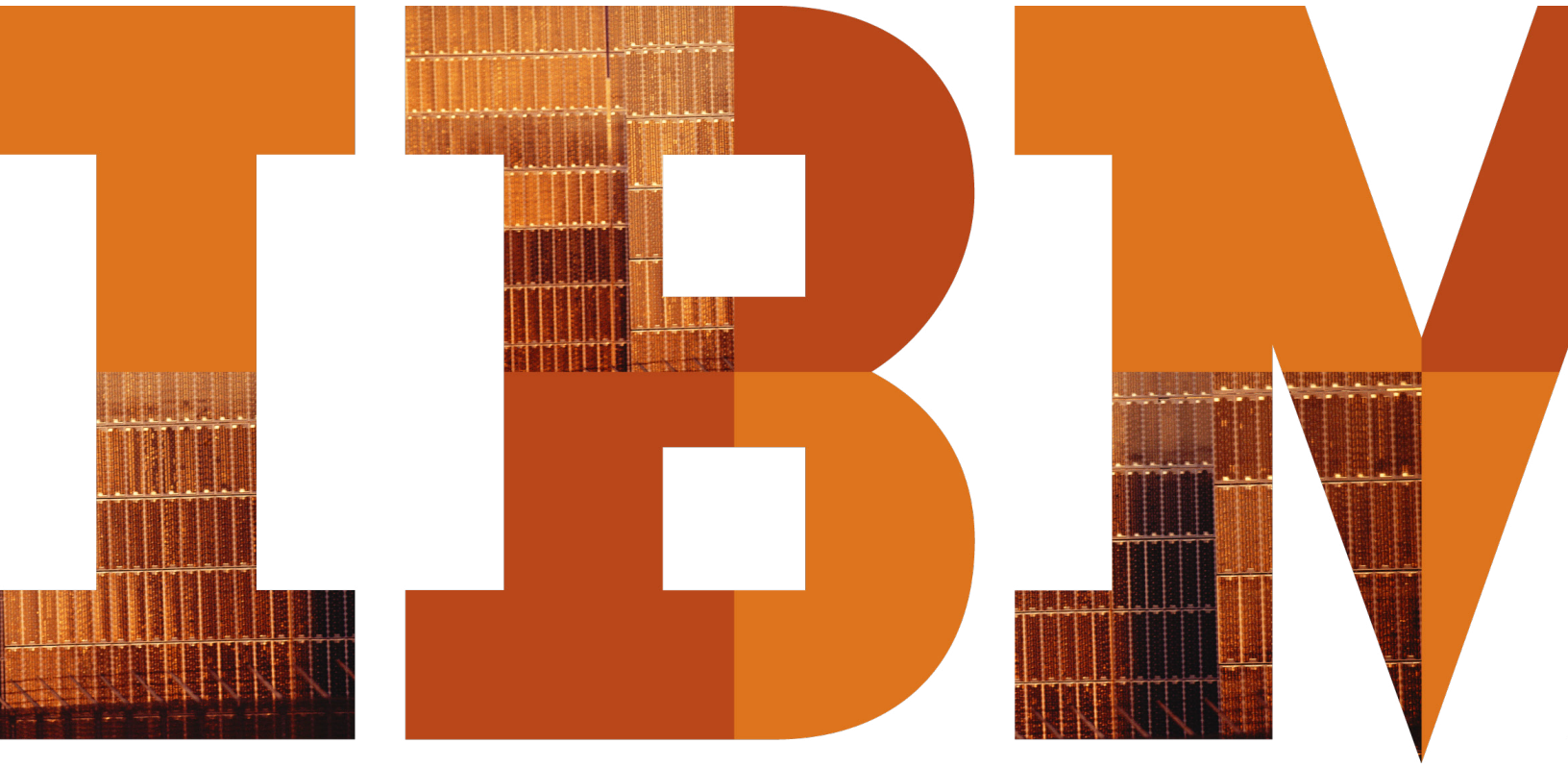
IBM Systems
White Paper

June 2019

Implement high-performance object storage with MinIO and IBM

Achieve robust performance for AI, IoT and more using MinIO and IBM Power Systems servers with POWER9 processors



IBM

Executive summary

Object storage presents several important benefits for accommodating fast-growing volumes of unstructured data. With the right object storage solution and hardware infrastructure, organizations can also achieve the robust performance required for supporting computationally intensive workloads, including artificial intelligence (AI)/machine learning, Internet of Things (IoT), and big data analytics.

Recent benchmark testing shows that MinIO object storage running on IBM Power Systems servers with IBM POWER9 processors can deliver exceptional throughput performance—up to 25 GB/s in aggregate for four servers—plus linear scalability as clusters grow. That level of performance enables organizations to unlock the full value of their data while also capitalizing on the scalability, accessibility, data protection, and cost-effectiveness of object storage.

Launching data-intensive initiatives

Across industries, organizations are launching new technology initiatives that require them to store, access, and analyze large, fast-growing volumes of data. Whether they are implementing artificial intelligence (AI)/machine learning, capitalizing on Internet of Things (IoT) technology, or employing other big data solutions, these organizations might need to store and analyze tens—or hundreds—of petabytes of data.

Much of that data is unstructured. From multimedia files and text documents to web pages and log files, unstructured data can be difficult to query, making it challenging for organizations to work with all of the data they are collecting. Traditional hierarchical file storage systems and block storage are not the best fit for these unstructured data volumes.

Object storage offers an important alternative to file- and block-based storage for big data, as proven by organizations with hyperscale environments. Object storage provides the right combination of cost-effective scalability, data integrity, and accessibility that many organizations need. It also offers the flexibility to disaggregate storage from compute resources, enabling organizations to optimize compute and storage for specific workflows. As a result, object storage is fast becoming the default storage option for these organizations.

Using the right hardware infrastructure, object storage can also provide a fundamentally different performance profile than other types of storage, enabling organizations to implement new use cases and launch more ambitious projects. High-performance object storage can support workloads ranging from training AI algorithms to analyzing IoT data. Running MinIO object storage with IBM Power Systems servers based on IBM POWER9 processors can deliver this level of performance, opening important opportunities for enterprises deploying workloads in private cloud or multicloud environments.

Recognizing the advantages of object storage

For storing large, rapidly expanding volumes of unstructured data, object storage can present your organization with several advantages over more traditional file- or block-based storage.

Scalability

Object storage is designed to scale. Instead of the nested files and folders used by hierarchical file systems, object storage uses a flat structure. That structure enables you to store billions of files without the complexity and performance issues that can develop as you scale hierarchical environments. Object storage also lets you scale incrementally: you can scale performance or capacity simply by adding racks of clusters.

Fast retrieval

With MinIO object storage, each object has metadata and uses the URL as a unique identifier. These tags and ID numbers help eliminate the need to know the exact location of data within the storage environment. Every object is accessible from anywhere through its unique URL—only standard IP routing and DNS mechanisms are required. The right object storage solution can also avoid the bottleneck of a centralized metadata server, storing the metadata alongside objects.

Data protection and preservation

Object storage solutions protect and preserve data more efficiently than other types of storage architectures. By using data protection capabilities such as erasure coding, object storage can protect data using far less raw storage capacity than RAID-based architectures. Data protection capabilities can also help quickly repair problems on a per-object basis, instead of on a per-disk basis, helping to avoid data loss and to maintain high availability of data.

Cost-effectiveness

The ability of object storage to scale incrementally, without forklift upgrades, can help you control storage costs. In addition, object storage data protection capabilities help eliminate the need for numerous copies of files, reducing the raw storage capacity required to safeguard data and driving down capital expenditures.

Unlocking the full value of data with high-performance object storage

Object storage has not always been used for high-performance workloads. In fact, some organizations employ object storage as a backup environment or a long-term disk-based archive.

Object storage does have advantages for these use cases. By storing objects along with metadata, object storage can make it easier for users to find and retrieve the files, media clips, or entire projects they need among millions or billions of files. At the same time, data protection capabilities can help securely preserve data over the long term.

Yet to maximize the value of data residing in object storage, you need to be able to consume it quickly. High-performance object storage solutions can help you extend the benefits of object storage to new use cases and extract more value from your stored data. If you can achieve sufficient throughput, you can use object storage for big data and IoT analytics, as well as AI/machine learning workloads.

Until recently, big data, IoT, and AI workloads often drove organizations to employ Hadoop Distributed File System (HDFS) storage. With HDFS, you bring the algorithm to the data. Each node computes a part of the algorithm using local storage and then sends the results back to a centralized server, where results are aggregated. This approach can work well for some algorithms, and it can offer scalability for large-scale collections of data.

However, object storage presents several advantages over HDFS. For example, object storage can provide greater flexibility for balancing compute and storage across your environment. Using high-speed networking with your object storage environment, you can consume your compute and storage resources in the optimal way for each particular workload.

Object storage also requires less capacity than HDFS to ensure data protection for the same amount of data. While HDFS stores multiple copies of each file, object storage can use data protection capabilities such as erasure coding to protect data more efficiently. Object storage also helps eliminate the risk of using a single master node, which can become a single point of failure. Overall, high-performance object storage provides a more efficient and reliable way to support data-intensive workloads than HDFS.

Capitalizing on MinIO high-performance object storage with enterprise capabilities

MinIO high-performance distributed object storage is designed for large-scale data environments. It is a well-suited Amazon S3–compatible replacement for HDFS, especially when used for AI/machine learning, IoT, and other big data workloads.

MinIO object storage comprises a server, optional client, and optional software development kits (SDKs):

- MinIO Server is a distributed object storage server that includes an array of enterprise-grade capabilities.
- MinIO Client ("mc") is a modern alternative to UNIX commands that supports web-scale object storage deployments.
- MinIO Client SDKs include simple APIs for accessing any Amazon S3–compatible object storage.

MinIO is an open source solution that offers several enterprise capabilities for protecting data, maintaining data integrity, tightening security, and maximizing flexibility.

Data protection and integrity

Per-object, inline erasure coding protects against data loss and maintains availability of data—even if multiple drives or devices are lost. Bitrot protection avoids reading corrupted data caused by aging drives, firmware bugs, accidental overwrites, and other problems.

Security

MinIO supports multiple, sophisticated server-side encryption schemes to protect data wherever it resides. MinIO Server encrypts each object with a unique object key. Even if an individual object is compromised, the same decryption key cannot be used with any other object. In addition, MinIO offers a write-once, read-many (WORM) mode, which disables all APIs that can potentially mutate the object data and metadata: once written, data becomes tamperproof.

Support for advanced standards in identity management creates centralized access with temporary and rotated passwords. Fine-grained, configurable access policies facilitate simple support of multitenant and multi-instance deployments.

Flexibility

MinIO allows you to combine multiple data instances to form a unified global namespace. As a result, you can support geographically distributed users while accommodating a variety of applications from a single console. By using an Amazon S3 API, MinIO also gives you the flexibility to support multiple clouds—and incorporate existing storage—while ensuring that your view of data looks exactly the same.

Achieving robust object storage performance with MinIO optimized for POWER9

IBM Power Systems servers based on POWER9 processors provide the high-performance infrastructure required by MinIO high-performance object storage software. Together, these solutions can support demanding workloads such as AI, IoT analytics, and big data analytics.

For many organizations, Power Systems servers offer the right combination of performance, reliability, cloud flexibility, and security.

- **Robust performance:** Outstanding core performance plus high memory bandwidth help deliver industry-leading performance.
- **Reliability:** IBM Power Systems servers provide dependable on-premises infrastructure to meet around-the-clock user demands.
- **Cloud flexibility:** These servers integrate easily into private cloud and multicloud strategies.
- **Security:** Strong security capabilities—such as accelerated encryption built into the chip—help ensure data remains protected.

To achieve the object storage performance needed for AI, IoT, and big data workloads, the POWER9-based servers take advantage of PCIe 4.0 technology. PCIe 4.0 doubles the bandwidth offered by PCIe 3.0, which remains the standard used by other CPU architectures.

In addition, these servers support nonvolatile memory express (NVMe) storage technology, through which each processor core communicates directly with storage devices using the PCIe bus. NVMe drives can deliver superior performance compared to previous-generation, flash-based storage. These drives also enable you to achieve that performance in dense environments that help control infrastructure costs.

Fast networking is critical for maximizing bandwidth across object storage clusters. By supporting multiple 100 Gb/s Ethernet networking links per server, the Power Systems servers help eliminate networking bottlenecks.

Several POWER9-based servers also feature a storage-rich design that supports processing and analysis of very large data volumes. The Power Systems LC922—which offers the highest storage capacity in the Power Systems portfolio—supports up to 120 TB of capacity in a 2U form factor.

Benchmarking IBM POWER9-based servers with MinIO

MinIO engineers conducted benchmark testing to demonstrate the extreme performance that is possible using MinIO Server with POWER9-based systems. The testing deployed four IBM Power Systems LC922 servers, equipped with POWER9 processors, along with four POWER8-based servers as clients. The POWER9 servers included NVMe-based flash drives in addition to hard-disk drives. The environment used a high-speed 100 Gb private network.

To fully capitalize on the throughput performance of POWER9-based servers, the MinIO team optimized and accelerated MinIO Server for the POWER9 architecture using the Golang (Plan 9) assembly feature.

100 GbE top-of-rack switch

4x IBM Power Systems S822LC servers (clients)

4x IBM Power Systems LC922 servers

Figure 1: The test environment included four IBM Power Systems LC922 POWER9 servers (right), four IBM Power Systems S822LC servers as clients, and 100 GbE networking.

The MinIO team first evaluated throughput performance for accelerated versions of two computationally intensive algorithms: erasure coding and HighwayHash (for bitrot detection).

Erasure coding

With MinIO, erasure coding is designed to take place inline on a per-object basis. When you store 1 GB of data, MinIO splits up that data across a large number of drives and creates the appropriate amount of parity data on separate drives. Depending on the parity configuration you choose, you can afford to lose up to half of the servers and half of the drives—you will still be able to reconstruct all of your data. Running erasure coding inline—instead of offline—enables you to start protecting data the moment you store it, but it inherently demands high-performance object storage, which MinIO is able to provide.

In the benchmark testing, the optimized erasure coding algorithm running on POWER9 systems achieved throughput of 7–9 GB/s per core, which is critical for saturating the fast 100 Gb network. This level of throughput for the optimized algorithm reflects the robust performance of the POWER9 system architecture, which is particularly well suited for this type of high-throughput workload.

Bitrot detection

Similar to erasure coding, MinIO is designed to run bitrot detection on the fly. MinIO's implementation of the HighwayHash algorithm helps prevent the reading of corrupt data. The algorithm computes a hash on read and verifies the hash on write from the application. Any change in the hash fingerprint indicates data corruption and requires the use of parity data instead of the corrupted data.

Hashing operations require considerable CPU resources, but the POWER9-based servers can deliver the required performance. In the benchmark testing, the optimized HighwayHash algorithm running on the POWER9 servers achieved throughput of 5 GB/s per core, which can saturate the 100 Gb network.

COSBench

The team also ran COSBench, a commonly used open source benchmarking tool, to measure the performance of object storage services. COSBench testing used four POWER9-based systems, each with four NVMe drives and connected with 100 Gb/s networking.

The team ran COSBench on the four clients with 256 threads per client (1024 total). Each test typically took about an hour, with a prepare (WRITE) stage of 20–30 minutes, a 20-minute main (READ) stage, and a final cleanup stage. The team uploaded and downloaded more than 10 TB of data to mitigate any memory caching effects that could inflate the performance numbers.

Object-size benchmarks: The team used the four-node cluster to benchmark MinIO object storage read and write throughput for objects of increasing size. Read performance reached 18 GB/s and stayed constant through 32 MB and 64 MB object sizes. For larger objects, the write performance achieved 50 percent of the read performance, which is a strong result.

Object Size	10 MB	20 MB	32 MB	64 MB
Read (GB/s)	14.9	18.1	18.7	18.0
Write (GB/s)		5.7	7.3	10.1

Figure 2: Read performance reached 18 GB/s for objects of 20 MB or larger.

Cluster scaling benchmarks: The team also benchmarked MinIO cluster scaling by increasing the number of nodes used in the test. The COSBench test demonstrated a maximum read performance of nearly 25 GB/s in aggregate for the four POWER9-based servers.

Expanding the cluster could also boost read performance. Because MinIO clusters can grow to 32 servers, and overall throughput increases as cluster size increases, the total read performance could be higher than 25 GB/s.

Number of Servers	1	2	3	4
Throughput (GB/s)	10.5	19.4	24.1	25.4

Figure 3: MinIO Server performance increases as the cluster size expands.

Benchmarking summary

Results from the erasure coding, bitrot, and COSBench testing all show the impressive throughput performance that can be achieved with MinIO Server on POWER9-based systems. The results of the erasure coding and bitrot detection algorithm testing highlight how well this architecture handles these two specific computationally intensive processes. But the results also suggest that this architecture could deliver strong results for computationally intensive AI, IoT, and big data workloads.

The COSBench testing illustrates how this distributed object storage architecture can deliver outstanding aggregate throughput performance across a cluster, enabling clients to take full advantage of the high-performance nature of MinIO object storage. Whether your organization is running a private or multicloud environment, you can use this architecture to gain the performance you need for parallel processing of large sets of unstructured data.

Moving forward with MinIO and IBM

Object storage provides an important alternative to file and block storage for large and growing volumes of unstructured data. By selecting high-performance object storage, your organization can extend the benefits of object storage to new use cases, including AI/machine learning, IoT, and other big data workloads. Employing MinIO in combination with IBM Power Systems servers based on POWER9 processors can deliver the performance to support those workloads and unlock greater value from data.

Learn more

To discover more about MinIO benefits for AI, IoT, and additional big data workloads, visit: https://min.io

To learn more about the complete line of the IBM Power Systems family, visit: **ibm.com**/it-infrastructure/power

IBM Global Services
Route 100
Somers, NY 10589
USA

Produced in the United States of America
June 2019

Please Recycle

MINIO