SOLUTION TEST REPORT

APRIL 2022

Western Digital® OpenFlex™ Data24 and MinIO Object Storage Solution Test Report

Abstract

This document provides a record of the testing performed to validate the compatibility and performance of MinIO solutions with the Western Digital OpenFlex Data24.

Notices

Western Digital
5601 Great Oaks Parkway
San Jose, CA 95119

Table of Contents

Introduction

This Solution Test Report summarizes the results obtained during the validation of MinIO object storage application with a Western Digital® OpenFlex™ Data24. The objective of this exercise was to benchmark MinIO S3 performance on the OpenFlex Data24 using the Warp benchmark tool. This document includes guidelines for configuring MinIO with an OpenFlex Data24.

MinIO is high-performance, S3 compatible object storage. Kubernetes-native, MinIO object storage software is available on every public cloud, every Kubernetes distribution, the private cloud, and the edge. MinIO is software-defined and is 100% open source under GNU AGPL v3. Customers depend on MinIO for AI/ML, IoT and big data workloads.

All tests were completed utilizing MinIO over 100 Gb Ethernet interfaces to the OpenFlex Data24. These tests were performed by MinIO in Palo Alto, CA, USA.

This report is not an endorsement of MinIO by Western Digital, and no warranty of the product is either expressed or implied.

Objectives

There are several objectives for this test:

1. Validate environment configuration and compatibility between MinIO object storage software and OpenFlex Data24.
2. Generate an idealized hardware configuration with OpenFlex Data24 and MinIO to optimize S3 performance.
3. Benchmark S3 read & write performance of MinIO object storage when utilizing the OpenFlex Data24 as primary storage.

Prerequisite Table

Item	Notes
Servers with 100 GbE NIC supporting RoCE v2	Western Digital OpenFlex products require 50 GbE – 100 GbE network connection supporting RoCE V2
10-100 Gb Network Switch	Western Digital Storage Servers or Composable Storage require network connectivity ranging from 10 to 100 GbE

Technology Summary

Technology	Version
Dell® R650 Servers	Firmware V5.1
Linux® OS	CentOS 8.3.2011
Mellanox® ConnectX®-6 Dx	Dual port 100 GbE, PCIe Gen 4.0 x16, MOFED 5.3-1
Mellanox SN2700 Switches	32 port 100 GbE 1U switch
OpenFlex Data24	Firmware V3.0
MinIO object storage	minio-20220218015010.0.0.x86_64

Overview

MinIO is a pioneer in the development of high-performance, cloud-native object storage, refining and perfecting many of the features, protocols and APIs that have come to define best in class. MinIO is often used for AI workloads such as machine learning, IoT, data lake and other big data workloads. This solution combination of OpenFlex Data24 and MinIO object storage demonstrate some of the key benefits of composable disaggregated infrastructure (CDI): performance, flexibility, scalability, and maintainability.

MinIO distributed object storage server boasts a comprehensive implementation of the Amazon S3 API and the MinIO Console, a browser-based GUI that incorporates all the functionality of MinIO Client in a design that feels familiar for IT admins and developers alike. The MinIO Console is built to support cloud-scale deployments with minimal operational overhead, enabling administrators and users to provision multi-tenant object storage as a service, visually inspect the health of the system, perform key audit tasks, and simplify integration (via webhooks and API) with other components.

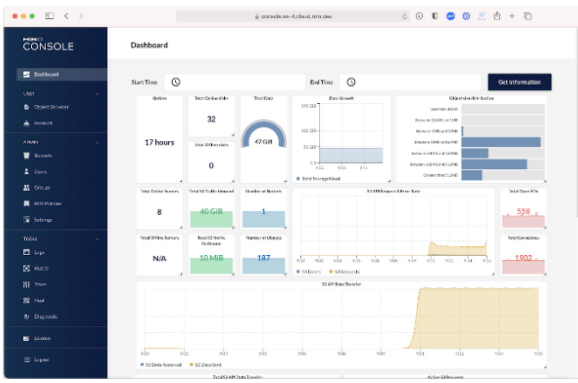

Diagram 3: Minio Console Dashboard

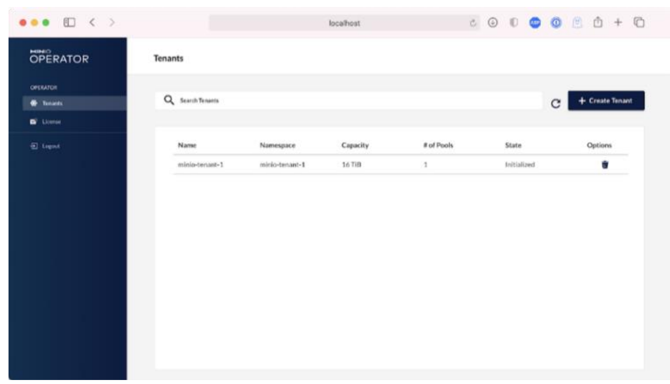

Diagram 4: Minio Kubernetes Operator Dashboard

Using 100 GbE as the fabric between the MinIO servers and the OpenFlex Data24 storage provides not only high performance with the NVMe-oF™ RDMA interconnect, but also great flexibility. Ethernet connectivity enables organizations to independently scale the compute and storage aspects of their storage environment across sites and racks. Whereas traditional storage architectures are limited by the amount of storage that can be connected via SAS enclosures or the number of media slots in the storage servers, a CDI architecture enables organizations to expand capacity without limits using NVMe-oF and industry standard 100 GbE connectivity. Combined with MinIO, organizations can dynamically deploy and expand their scale-out storage with zero downtime using a heterogenous mix of server hardware.

MinIO centrally focuses on object storage and as a result, has a broad range of features that are designed to create a persistence layer that is performant, resilient, secure, and scalable across the hybrid cloud with features such as:

- Per-object inline erasure coding, written in assembly code delivering high performance. MinIO uses a Reed-Solomon code to stripe objects into data and parity blocks, configurable to a desired redundancy level.
- SIMD accelerated implementation of the HighwayHash algorithm ensures that it will never return corrupted data by capturing and healing corrupted objects on the fly. Integrity is ensured by computing hash on WRITE and verifying on every READ.
- Integration with the OpenID connect and LDAP compatible IDP providers and fine grained and highly configurable API level access policies meet the most advanced standards in identity management for multi-tenant and multi-instance deployments.
- Multiple server-side encryption schemes protect data, ensures confidentiality, integrity, and authenticity with negligible performance overhead. Server and client-side encryption using AES-256-GCM, ChaCha20-Poly1305, and/or AES-CBC. Encrypted objects are tamper-proofed with AEAD server-side encryption.
- Object locking, retention, legal holds, governance and compliance for objects and buckets. Combined with versioning to eliminate the risk of data tampering or destruction and retention rules to ensure that an object is WORM protected for a configurable period.

- Lifecycle management tools protects data within and across deployments with a wide range of policies built on object and tag filters to declare expiry rules and programmatically tier objects across different storage mediums and cloud types to optimizes your data for performance and cost.
- Object-level versioning provides data protection and serves as the foundation for data lifecycle management, tiering and locking. Following Amazon's S3 structure/implementation to independently versioned objects allow users to retain multiple versions, or snapshots, of every object at the bucket level.
- Server pools are a group of nodes with their own compute, network, and storage resources making the foundation of horizontal scaling for multi-tenant configurations. A single namespace is distributed across multiple server pools, assigned to a single tenant, and allows scaling by creating a new server pool.
- Active-active bucket level near-synchronous replication enables object storage to be deployed, configured, and accessed across multiple data centers and clouds in a resilient and scalable manner to withstand site failure without down time.
- No separate metadata store so all operations are performed atomically at object level with strong consistency. This approach isolates any failures to be contained within an object and prevents spillover to larger system failures.
- A Kubernetes-native design to simplify the provision and management a multi-tenant object-storage-as-a-service infrastructure from an operator service. This provides each tenant their own isolated namespace while sharing the underlying hardware resources that scales seamlessly from TBs to any size.

For this report, Dell R650 servers were used for the MinIO storage servers, but MinIO object storage software only requires a minimum specification requirement enabling IT organizations to pair OpenFlex with their server platform of choice.

Architecture

Our testing environment was comprised of six MinIO object storage servers connected to one OpenFlex Data24 through a Mellanox SN2700 fabric switch. Both the Switch and servers were configured following the Western Digital Lossless Best Practice Guide. The latest release version of MinIO as of February 2022 was used for this report and was configured using the standard setup procedures. Additional per server configuration steps were taken to enable lossless networking and connect to OpenFlex Data24 devices on boot and prior to MinIO services starting. To closely replicate a real customer deployment, we paired the storage cluster (Servers + Data24) configuration with six load generation clients running the WARP S3 benchmarking tool to generate S3 traffic and measure the performance of the overall solution.

The illustration below shows an ideal solution configuration, where all networks are physically segmented for maximum availability and performance.

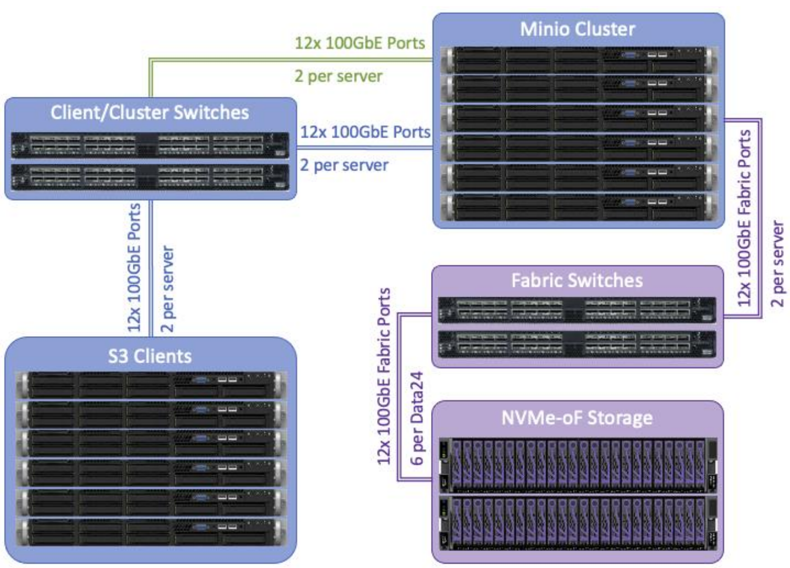

Diagram 5: Idealized minimal deployment configuration

Network Topology

A standard MinIO object storage configuration was set up with both a client and a cluster network. The client network is for servicing IO requests of clients and applications. The cluster network is for replicating or erasure encoding data and distributing it to storage nodes. For this solution, which incorporates the OpenFlex Data24, we required an additional network to support NVMe-oF storage fabric. The key to achieving performance is to make all network segments as symmetrical as possible. To this point we recommend using 100 GbE networking across the entire solution.

Testing Environment

The OpenFlex Data24 was configured with two namespaces per NVMe SSD, and MTU of 2200, and all six 100 GbE fabric ports onto a single subnet. This was done to provide 8 devices per server, over a single 100 GbE fabric port/IP. Since there were no redundant connections and multipathing, only a single VLAN and subnet was required.

The Fabric VLAN was configured following version 1.2 of the OpenFlex Lossless Best Practice Guide. Ask a Western Digital sales and support specialist to obtain a copy of this guide. The settings below were adjusted to accommodate the available hardware at the time of testing:

- Two namespaces were created on each NVMe-oF device of the Data24.
- Each namespace was only attached a single fabric port of the Data24.
- All six Data24 fabric ports were configured with only one subnet and cabled to a single switch.
- Each server accessed the NVMe-oF namespaces over only one Data24 Fabric Port.
- Three VLANs were configured to separate storage fabric, cluster, and client network traffic.
- Both Client and Cluster Network were configured to use the same server NIC port.

This illustration shows our testing environment.

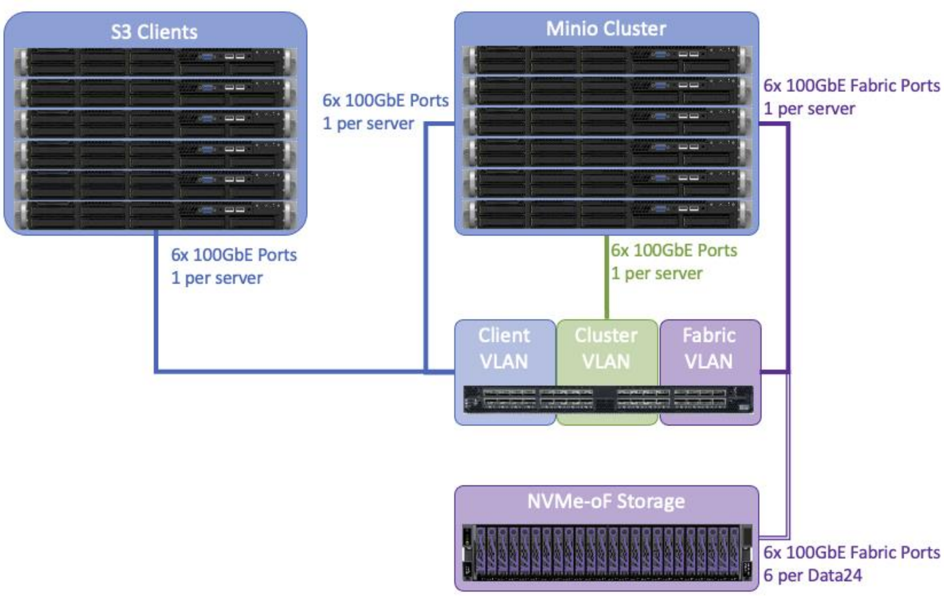

Diagram 6: Tested Configuration

OpenFlex Data24 and MinIO Configuration

1. Configure the fabric NICs for lossless networking on servers running MinIO.
2. Configure all fabric switch ports for lossless networking.
3. Verify Data24 is configured with the same subnet as the servers running MinIO.
4. Verify all Data24 namespaces are connected to the appropriate servers running MinIO.
5. Verify all network and fabric settings/connections are complete before MinIO services start.
6. Create and mount filesystems on all Data24 namespaces on each server running MinIO.
7. Configure MinIO to utilize all locally mounted filesystems for the Data24 devices.
8. Configure MinIO for erasure coding EC:4.

Workflow Testing

A 6 node MinIO cluster was created utilizing storage from one OpenFlex Data24. Each server was provisioned with CentOS 8.3.2001, Mellanox ConnectX-6 100 GbE Network cards, Intel Xeon Gold 6354 CPU, and 512 GiB RAM.

We ran the WARP S3 Benchmark (https://github.com/minio/warp) for our performance tests. This tool conducts benchmark tests from one or more clients to one or more hosts. WARP is an open source S3 performance benchmark tool developed and maintained by MinIO.

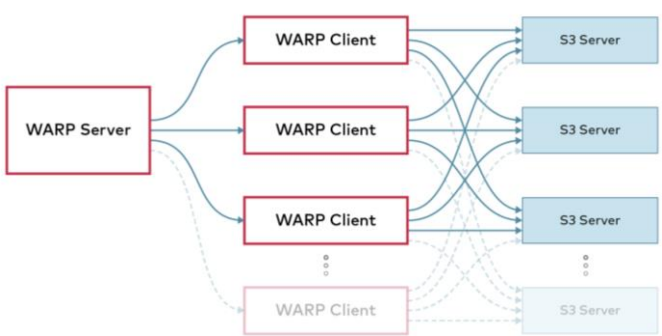

Diagram 7: WARP multiclient connection distribution

We ran the WARP benchmark with 6 clients. Each client was provisioned with CentOS 8.3.2011, Mellanox ConnectX-5 100 GbE Network cards, AMD Epyc 7542 CPU, and 512 GB RAM.

The benchmark ran with objects of various sizes from 4 to 128 KiB.

Results Summary

We obtained the following results from the WARP benchmark.

As you can see, performance maxed out at 101,640 GET objects/second and 26,213 PUT objects/second.

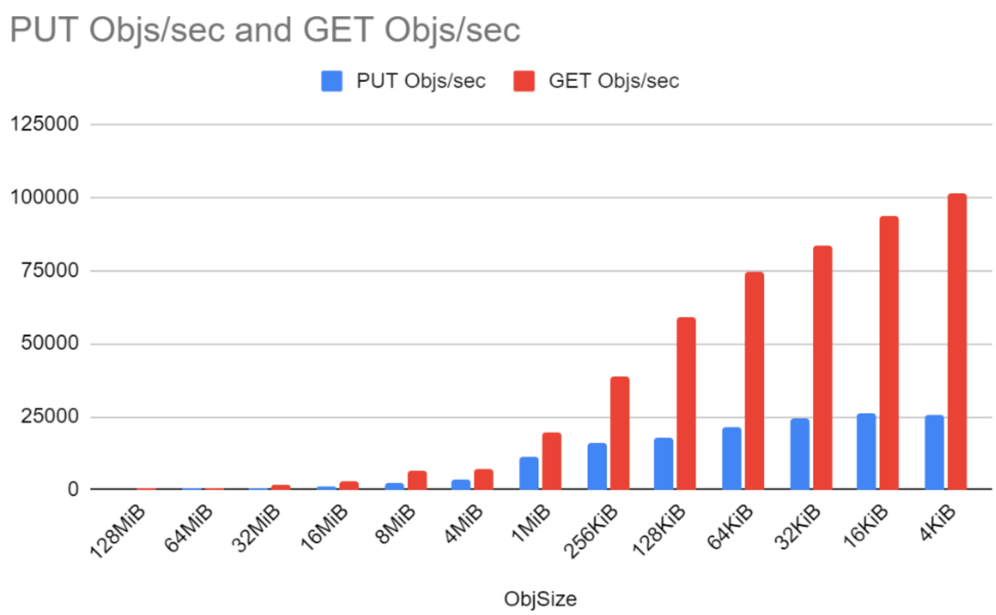

Diagram 8: Performance characteristics Objects/second

Looking at bandwidth, we recorded maximums of 22 GiB/sec for PUT traffic and 51 GiB/sec for GET.

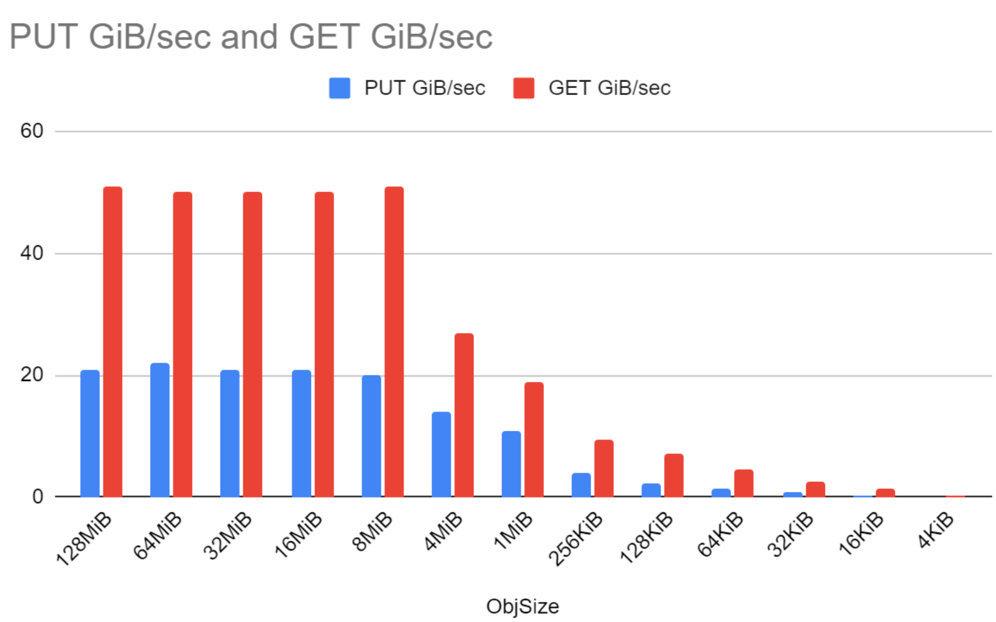

Diagram 9: Performance characteristics in GiB/second

Testing was conducted by MinIO.

Considerations

Scaling options	Hardware changes	Notes
High-Capacity HA	1x Data24 with 2 Fabric NICs For each server	Each server can achieve same storage performance as tested, but with 24 drives available per server capacity is drastically increased
High Performance HA	1x Data24 with 6 Fabric NICs For every 3 servers	This allows for the realization of nearly the full drive performance as tested above, but with the addition of multiple paths for added resiliency
Minimal Deployment	1x Data24 with 2 Fabric NICs for every 4 servers with 25 Gb NICs or ever 2 servers with 50 Gb NICs	Reduces network port cost, requires break-out cables, lowers CPU requirements, can be replicated to reach capacity & compute targets

Appendix

Contributors

Name	Company	Title
David Tobin	Western Digital	Sr. Technical Marketing Engineer
Matt Sarrel	MinIO	Director of Technical Marketing

References

Document Title	Date
OpenFlex Data24 User Manual	February 14, 2022
MinIO Administrator Guide	February 14, 2022
MinIO Installation Guide	February 14, 2022

Document Feedback

For feedback, questions, and suggestions for improvements to this document send an email to the Data Center Systems (DCS) Technical Marketing Engineering (TME) team distribution list at pdl-dcs-tm@wdc.com.

Version History

Version	Date	Notes
1.0	8 April 2022	Initial release

5601 Great Oaks Parkway
San Jose, CA 95119, USA
www.westerndigital.com