Western Digital.

Solution Test Report

Three Ways to Add Western Digital® OpenFlex™ Data24 3200 Storage to OpenStack®

Abstract

This document demonstrates three possible ways to use the Western Digital® OpenFlex™ Data24 3200 NVMe-oF™ enclosure to add block storage to OpenStack.

February 2025

Table of Contents

Introduction

The purpose of this report is to introduce several options for adding Data24 3200 block storage to an OpenStack environment. Future documents may cover object, NAS, and DBaaS storage options – please send feedback if there is interest in these options.

OpenStack is a virtualization system that can support just about any type of storage. There are many more than 3 options for adding Data24 3200 storage to OpenStack, but a basic understanding of these 3 methods will prepare you have a conversation with any OpenStack storage admin.

For the full configuration matrix, consult https://docs.openstack.org/cinder/latest/reference/support-matrix.html - for best results choose a storage product such as LinBit DRBD that supports both Cinder and the OpenFlex APIs.

This report provides a closer look at three potential ways to integrate Data24 3200 block storage into an OpenStack environment.

What is OpenStack?

OpenStack is a free, open infrastructure management system and cloud computing platform. It is typically deployed as Infrastructure-as-a-Service (IaaS) in both public and private clouds where virtual servers and other resources are made available to users (Wikipedia).

OpenStack began in 2010 as a joint project of Rackspace Hosting and NASA as a way to provide on-premise Amazon-EC2-like services. It has been particularly successful for telecom Network Function Virtualization (NFV) applications.

Summary

The following table provides a summary of the three methods used in this report. These methods are all open source and use “in box” packages on most Linux distributions:

Storage Type	Use Case	Advantages	Disadvantages
Ceph	General Purpose	Complete SAN/CoW functionality Excellent OpenStack integration Maturity	Poor Small Block performance
cinder-lvm	Quick & Dirty	Easiest method – just create a block device by hand and add to Cinder	“Monolithic” service – not easily scalable or clustered Careful monitoring required
Direct Fabric Connection from virtual machines to NVMe target NQNs	Highest Performance	High performance – direct connection from VMs to targets	VMs need direct network access to the Data24 3200 (Same subnet, Routed, or NAT) SR-IOV required for best results using ConnectX virtual functions

OpenFlex Data24 3200 NVMe-oF Storage Platform

Western Digital's OpenFlex Data24 3200 NVMe-oF Storage Platform is similar to a 2.5" SAS Enclosure. The Data24 3200 can be used as stand-alone storage but can also be a foundational block of a Software-Composable Infrastructure. It provides 24 slots for NVMe drives and a maximum capacity of 368 TB[1] when using Western Digital Ultrastar DC SN840 15.36 TB devices. Unlike a SAS enclosure, the Data24 3200's dual IO modules use Western Digital RapidFlex™ C2000 NVMe-oF Controllers. These controllers allow full access to all 24 NVMe drives over up to six ports of 100 Gb Ethernet.

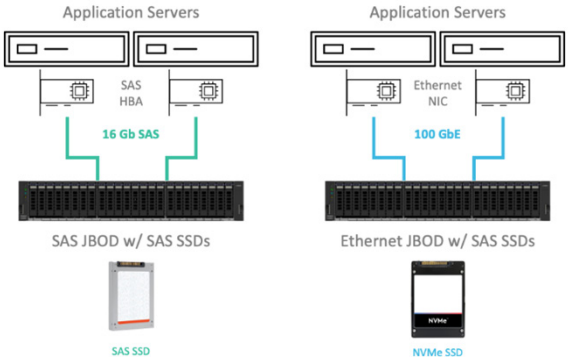
The Data24 3200 is a close replacement for the traditional SAS enclosures. However, the Data24 3200 offers a significant benefit over these enclosures: the ability to integrate directly into Ethernet fabric, allowing for an Any-to-Any mapping of Object Storage Targets to Object Storage Servers.

The OpenFlex Data24 3200 design exposes the full performance of the NVMe SSDs to the network. With 24 Western Digital Ultrastar DC SN840 3.2 TB devices, the enclosure can achieve up to 71.4 GB/s of 128K bandwidth and over 16.7 MIOPS at a 4K block size.

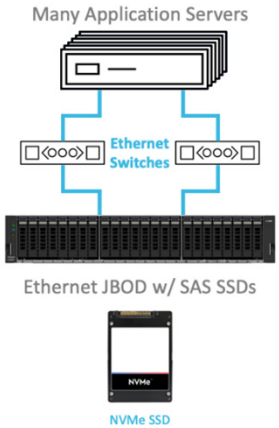
OpenFlex Data24 3200 NVMe-oF Enclosure

There are two main benefits an NVMe-oF enclosure provides to users. The first is that it allows users to upgrade from SAS SSDs to NVMe SSDs without changing their application architecture. Users can continue to deploy Active / Passive server clusters, parallel file systems, or other applications that required shared storage enclosures without any significant modifications.

Existing designs that use an external SAS attached enclosure can replace the SAS Host Bus Adapter (HBA) with an Ethernet Network Interface Card (NIC) and then replace the SAS cable with an Ethernet cable and maintain the same architecture model.



Identical Architectures for SAS & NVMe-oF



The second benefit is the ability to use an Ethernet switch between the end client hosts and the storage enclosure. SAS is a point-to-point protocol meaning an external storage enclosure is directly cabled to the server that is accessing it. External enclosures typically have dual IO Modules with multiple SAS ports and therefore can directly connect to a few hosts, but they cannot have any flexibility with regards to the host that access the drives—a drive in an external enclosure can only be accessed by a host that is directly cabled to the enclosure. With Ethernet as the transport layer, any host on the network can be configured to access any drive in the enclosure. This benefits greatly when a client node fails. When a node fails, the application can be restarted on any other node on the network and not just another node that is directly cabled to the external enclosure.

Many-to-one Connectivity with NVMe-oF Enclosures

[1] One gigabyte (GB) is equal to one billion bytes and one terabyte is equal to one trillion bytes. Actual user capacity may be less due to operating environment.

Evaluation Platform Overview

- The evaluation platform is a 6-node hyperconverged cluster with two Mellanox ConnectX-6 NICs in each server.
- The storage device is a Data24 3200 with 24 3.2TB Western Digital SN840s.
- Full server and network specifications and a bill of materials are available in Appendices A & B.
- The complete MAAS & Juju configuration process is available in another PFE document.

Test Procedure

The virtual machines used in these tests were created with two block devices / volumes – one volume for boot and a second 100GB volume for benchmarking.

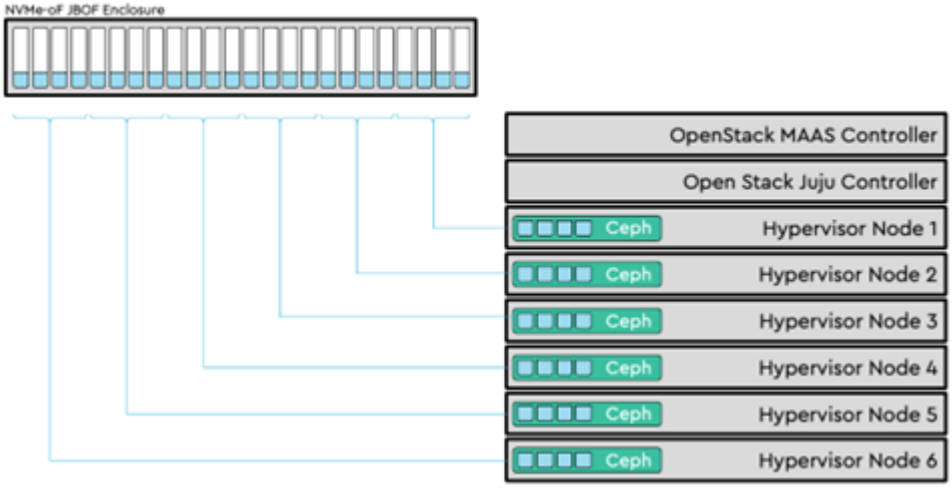
The test benchmarks used the fio engine with 128K and 4K tests similar to the Data24 3200 spec sheet.

For the Ceph tests, both volumes were created on one pool with 3x replication, using all 24 drives.

For the direct SR-IOV tests, the boot volumes used a Ceph pool backed by a single local drive on each server, and the benchmark volumes used direct routed NVMe fabric connections to the Data24 3200.

For cinder-lvm there are many possibilities since any block device can be used. It is possible to use a service VM or a dedicated physical server, and whatever data redundancy scheme is appropriate. In this test, we use a dedicated cinder-lvm server and 8 NVMe fabric drives from the Data24 3200. For redundancy we use mdadm RAID 10.

The workloads were generated using fio. Results are from fio, not the raw block devices.



Reference Architecture Under Test

Option 1 - Ceph

The most common general purpose storage software used with OpenStack is Ceph. The main advantages of Ceph include:

- Complete SAN/CoW functionality such as snapshots
- Many Replication and Erasure Coding Options
- Excellent OpenStack Integration
- Native Block, NAS, and Object Interfaces
- Native support for Linux, OSX, and Windows clients
- Maturity

Ceph's primary "Achilles Heel" is OLTP environments due to slow small block performance.

For this evaluation we chose Ubuntu 20.04 and their deployment environments MAAS and Juju. However, these concepts apply to any OpenStack or Ceph vendors such as OS Nexus.

The test cluster uses 6 hyperconverged nodes with two 100Gb ports for OpenStack, and two 100Gb fabric ports for the Data24 3200. Each node connects to 4 drives using NVMe native multipathing. The Ceph pool used replication with

Ceph Performance Results

The fio test used 12 virtual machines, 2 per OpenStack node. See the Data24 3200 spec sheet for test details

Ceph RBD Results from Test ID 887cede3	
128K Sequential Write Bandwidth	8,051 MB/sec
128K Sequential Read Bandwidth	27,549 MB/sec

Option 2 - cinder-lvm

Cinder can use any block device to create volumes using the cinder-lvm driver.

There are many ways to connect to a set of Data24 3200 drives to create a block device. One such method would be to add a dedicated hardware node for cinder-lvm. In our cluster design, this node would need 4 100Gb ports–two for OpenStack and two for the storage fabric.

Then you can connect to 8 Data24 3200 drives and use mdadm RAID 10 to provide redundancy. cinder-lvm will create volumes and export them to the virtual machines using iSCSI or NVMeoF.

Cinder-lvm Estimated Performance Results

Estimated cinder-lvm Bandwidth from Data24 3200 1x8n Test IDs 3102ee2a and 6f12b03b	
128K Sequential Read Bandwidth	11,633 MB/sec
128K Sequential Write Bandwidth	7,450 MB/sec
4K Random Read IOPs	2,545,636
4K Random Write IOPs	1,860,564

Option 3 – Direct Fabric Connection To Virtual Machines

For maximum performance, enable SR-IOV/Virtual Functions on the NIC and attach each VM directly to the Data24 3200 targets using the Neutron “sriovnicswitch” driver.

This allows for near-native performance.

This configuration uses routed access from VMs to the fabric VLANs, using the Mellanox SN2700 as the router. There were 24 VMs, two per node, with one SN840 each. The fabric connection used RoCEv2 and NVMe native multipath.

Direct Fabric Virtual Function Performance Results

Bandwidth from Data24 3200 Direct Fabric Test IDs feca1764 and a2db00b4	
128K Sequential Read Bandwidth	64,215 MB/sec
128K Sequential Write Bandwidth	36,291 MB/sec
4K Random Read IOPs	11,483,411

Appendix A - Evaluation Hardware Bill of Materials

Storage System	Western Digital OpenFlex Data24 3200 NVMe-oF Enclosure with (24) 3.2 TB Ultrastar® SN840 SSDs
MAAS, Juju, and (optional) cinder-lvm Controller Nodes	2x Dell R650 – low powered CPUs are OK
OpenStack Nodes	Dell® R750 servers with dual Intel® Xeon® Gold 6354 18-core processors and 512 GB of DDR4-3200 Memory, and dual Mellanox® ConnectX-6 Dx Network Adapters
Network Switch	Mellanox® SN2700
Operating System	Ubuntu® 20.04

Appendix B – Network VLANs

1 Gigabit Ethernet	PXE deployment, all nodes are on this VLAN
100 Gigabit VLAN #1	All cluster traffic, Juju services, and fixed IP addresses. Each node has two interfaces on this VLAN. In a production environment, this would be split into multiple VLANs
100 Gigabit VLAN #2	Data24 3200 IOMA
100 Gigabit VLAN #3	Data24 3200 IOMB

Appendix C – Reference Information

Contributors

Name	Company	Title
Mark Lehrer	Western Digital	Technologist, Applications Engineering
Barrett Edwards	Western Digital	Sr. Director, Platforms Field Engineering

References

- OpenFlex Data24 3200 Series NVMe-oF Storage Platform Product Brief

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