binary testing

The UK's premier independent network testing laboratory

Western Digital OpenFlex™ Data24

A review and performance validation test of Western Digital's NVMe-oF Storage Platform

Key findings

- Performance scales easily with more hosts
- Three servers returned a cumulative bandwidth of 34.1GB/sec for sequential reads and 7.63Million IOPS for random reads
- End-to-end NVMe-oF storage solution
- Up to 368TB in 2U of rack space
- Easy deployment and management
- Ready for the next generation of CDI data centres
- Western Digital OpenFlex architecture with Open Composable RESTful API

NVMe™ SSDs in the data centre

The performance and low latency of NVMe (Non-Volatile Memory Express) SSDs are driving exponential uptake as enterprises move to all-Flash in their data centres. These attributes combined with energy conservation and environmental concerns will make SSDs the predominant storage technology selected as enterprises upgrade their digital IT infrastructure.

Binary Testing labs conducts hands-on tests and reviews of enterprise-class servers from all the blue chips and we are now seeing all of them featuring integrated options to support NVMe SSDs with many now shipping with these high-performance storage devices. It's the same story for all-Flash arrays (AFAs) as the majority of enterprise products supplied to us for testing have over the past year, moved from SATA or SAS to NVMe SSDs.

NVMe will undoubtedly be the future of data centre storage as it's a mature protocol and the PCIe Gen4 specification will enhance performance even further with PCIe Gen5 also on the horizon. By connecting directly with the host system's PCIe bus, NVMe SSDs reduce CPU overheads for lower latency and hugely increased throughput.

IOPS are performed in parallel and a major advantage over legacy AHCI/SATA devices is the NVMe protocol which supports queue depths of 65K (as opposed to 32) for further acceleration of performance. Unlike spinning disks, NVMe SSDs are also available in multiple form factors such as U.2, U.3 and M.2.

There are, however, many challenges that must be overcome for data centres deploying AFAs using NVMe SSDs in large scale-out environments.

Instead of locking all that NVMe performance inside the server chassis, these external arrays are designed to present shared pools of storage to multiple host systems but the protocol would need to be translated from SCSI which has high overheads for flash devices to the much leaner and faster NVMe.

The fabric of storage

The NVMe-oF™ (NVMe over fabrics) standard was announced in 2016 and is specifically designed to overcome all these issues. It defines a protocol that disaggregates storage and compute and allows the performance benefits of NVMe to be delivered over standard local area networks.

The benefits of NVMe-oF are manifold as high-performance NVMe SSD storage arrays can now be presented as truly shared resources to multiple hosts without the inherent latency losses of legacy network transports. Unlike protocols such as iSCSI, no translation is required for fabrics as they use the same NVMe message-based command sets

This allows businesses to unlock their not-inconsiderable investment in Flash storage and unleash its full performance potential. Another key feature of NVMe-oF is it gives businesses a choice of how they implement it as it can be deployed over Ethernet, Fibre Channel, RoCE or InfiniBand.

As we've come to expect, the storage industry is once again responsible for creating more acronyms than any other and is does it again with RoCE (Remote Direct Memory Access over Converged Ethernet). RDMA provisions access to the memory of one computer with that of another without using their CPUs, cache or operating system and the RoCE protocol defines how RDMA is performed over Ethernet.

Implementing RoCE requires RDMA-compliant network interface cards (NICs) and the good news is the vast majority of vendors support this. For example, Intel offers its E810 and X772 series, Mellanox has the Connect-X family while Broadcom offers its NetXtreme E-Series - and this means NVMe-oF is ready for 10GbE, 25GbE, 100GbE and even 200GbE Ethernet networks.

Enter Western Digital

Western Digital's OpenFlex Data24 is one of the first NVMe-oF storage platforms to market and a key feature is it offers a complete solution from a single vendor. The Data24 is a 2U rack mount JBOF (Just a Bunch Of Flash) array with room up front for 24 NVMe SSDs. There's no need to shop around for storage as it is built to support Western Digital's new Ultrastar DC SN840 NVMe SSDs allowing it to deliver an impressive maximum capacity of 368TB.

Designed for data centres, the DC SN840 U.2 form factor NVMe SSDs are PCIe 3.1 models with dual ports for high availability. The highest capacity model is 15.36TB which has a 1 DW/D endurance for read intensive workloads and all support configuration of up to 128 NVMe namespaces

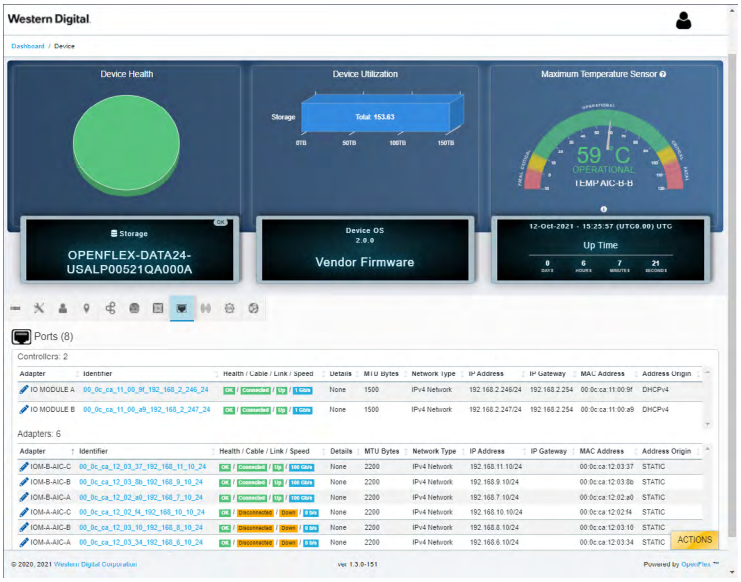
The Data24 employs two hot-swap I/O controller modules each providing three PCIe x16 slots for Western Digital's RapidFlex™ C1000 NVMe-oF controller cards each with a single 100GbE port. These controller cards support the latest RoCE v2 and come courtesy of Western Digital's acquisition of Kazan Networks allowing the company to offer an end-to-end high-performance storage solution comprising the array, storage and NVMe-oF adapters.

The I/O modules offer a lot of flexibility as each pair of ports provides connected hosts with access to eight NVMe SSDs. This makes it possible to directly attach up to six hosts without requiring a network switch while more basic environments can specify the Data24 with two adapters and use it as a replacement for an external SAS array.

Easy deployment

Western Digital's OpenFlex architecture means the Data24 is ready for the next generation of composable disaggregated infrastructure (CDI) data centres.

Supporting Western Digital's publicly available Open Composable API via a RESTful interface, the Data24 presents a composable storage system where, for example, NVMe-oF RDMA storage can be discovered, monitored and dynamically connected to scale-out storage clusters.



The Data24 is easily managed from the OpenFlex web browser GUI

For more general storage management and configuration, the Data24 provides dedicated Gigabit ports on each controller and presents Western Digital's OpenFlex web browser GUI. It opens with a smart dashboard view that provides vital health and performance statistics of all fabric devices on the same subnet.

We could drill down to individual storage devices and for the Data24, view information on its controllers, PSUs, cooling fans and all ports. Integral storage sensors relay information of the temperatures and health of all NVMe SSDs while the media section list shows more detail on individual devices and provides options to change their power state.

A ribbon menu provides quick access to details of all network ports such as connection status and health, connection speeds, IP addresses and MTUs. Our system was supplied with a full house of 100GbE RapidFlex C1000 NVMe-oF controller cards and it was a simple process to configure them ready for our performance testing.

Testing environment setup

For our performance tests, we used three Xeon Scalable servers each equipped with 100Gbps Mellanox ConnectX adapter cards which provide native hardware support for RDMA over Ethernet. We chose to install CentOS 8.4 on each server as it supports the Mellanox OFED (OpenFabrics Enterprise Distribution) driver and, unlike Windows, includes a native NVMe-oF initiator.

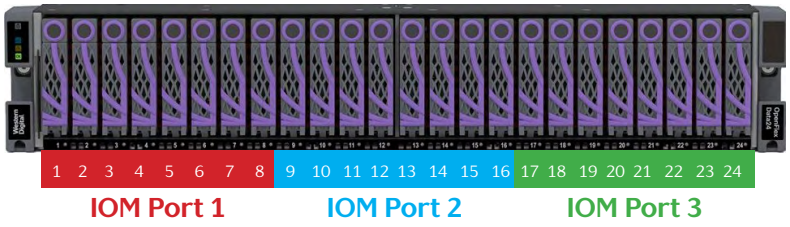
Using 1-metre DACs (direct attach cables), each server was connected to a dedicated 100GbE network port on IOM A on the Data24 allowing each one to access their own set of eight NVMe SSDs. All tests were performed using the open-source FIO (Flexible I/O) disk benchmarking tool.

Starting with one server, we ran FIO tests to determine maximum bandwidth in GB/sec and throughput in IOPS across the eight assigned NVMe SSDs in the Data24. Four tests were run using 128K block sizes to measure sequential read and write bandwidth while 4K blocks were used to determine IOPS for random read and write operations.

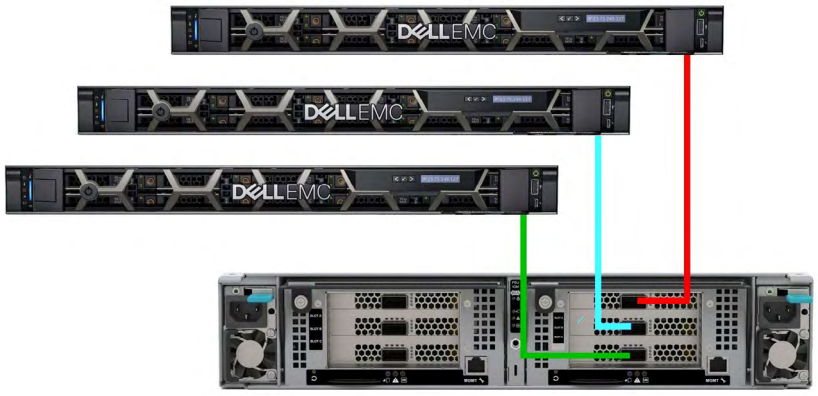
The same tests were then run simultaneously on two servers each with their own dedicated bank of eight NVMe SSDs. Finally, all tests were run together on three servers to determine if there was any contention for resources on the Data24.

Hardware environment

The Data24 physically assigns eight NVMe SSDs to each IOM so Server 1 sees drives 1-8, Server 2 sees drives 9-16 and server 3 sees drives 17-24. The FIO tests were configured to run on all eight drives for each attached server.



We used three Xeon Scalable servers to run our performance tests. Each was equipped with a Mellanox 100GbE NVMe-oF adapter card and cabled to a dedicated IOM port on the Data24 using DACs.



FIO test parameters

Multiple tests were initially run to get baseline numbers. We then modified FIO parameters for each test to achieve the optimum bandwidth and IOPS throughput (see Table 1).

Table 1: FIO workload parameters used for all tests

FIO tests	Block size	IO depth	No. of jobs	Runtime (s)
100% Sequential read	128K	16	16	300
100% Sequential write	128K	32	1	300
100% Random read	4K	32	8	300
100% Random write	4K	4	8	300

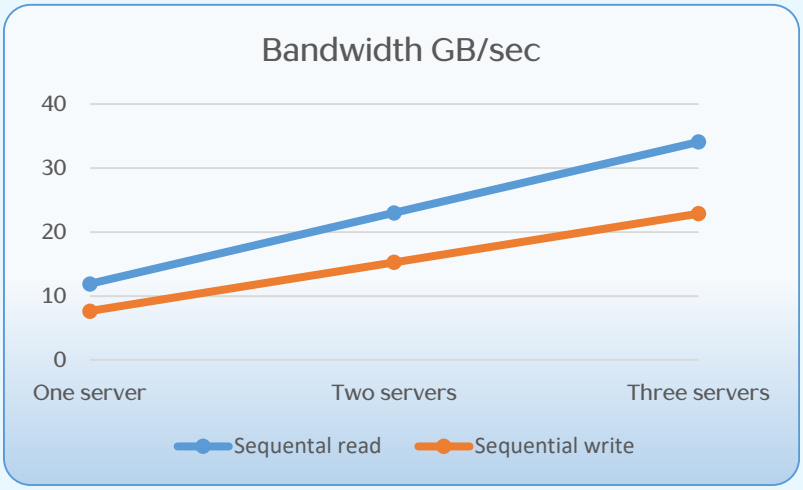
Western Digital OpenFlex Data24

A review and performance validation test of Western Digital's NVMe-oF Storage Platform

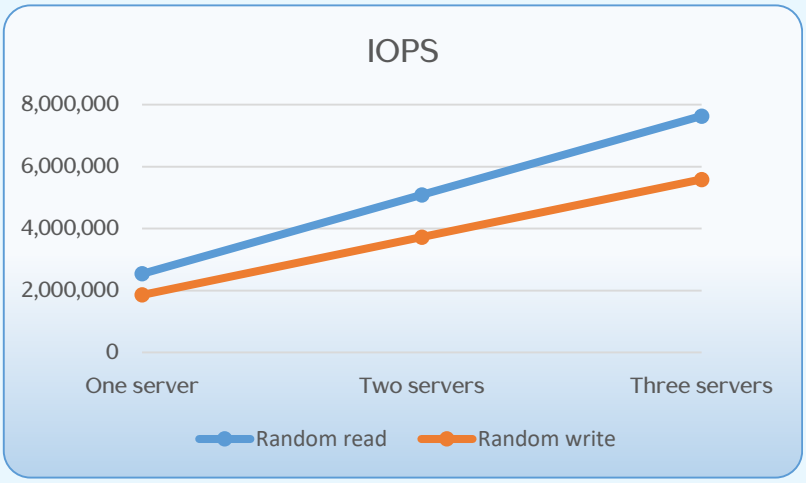
Performance results

Cumulative raw numbers

Bandwidth GB/s	Seq read	Seq write
One server	11.90	7.63
Two servers	23.00	15.26
Three servers	34.10	22.89



IOPS	Rand read	Rand write
One server	2,541,000	1,863,000
Two servers	5,090,000	3,726,000
Three servers	7,631,000	5,589,000



Conclusion

NVMe-oF is a huge game changer for data centres as it allows them to disaggregate storage from compute resources and release the full performance potential of their investment in Flash storage. Western Digital's OpenFlex Data24 is perfectly poised for the transition as this storage platform delivers a complete NVMe-oF solution from a single vendor.

It is clearly capable of scaling easily with demand as our 100GbE lab performance tests showed no degradation in bandwidth and IOPS as we increased the pressure with more host servers. As demonstrated in our performance graphs, all servers delivered the same maximum speeds and feeds regardless of whether one, two or three hosts were running the FIO benchmark tests.

Our three servers returned a cumulative bandwidth of 34.1GB/sec for sequential reads and an impressive 7.63Million IOPS throughput for random reads. With these validated numbers, there's no reason for us to doubt Western Digital's claim that the Data24 can deliver a maximum 71.3GB/sec bandwidth and a massive 15.2Million IOPS with six direct-attached servers.

binary testing

The UK's premier independent network testing laboratory

w www.binarytesting.com
e info@binarytesting.com
t +44 (0)1825 817370

This report was commissioned by Western Digital