E8 storage
Rack-Scale Flash. No Compromise.

Mellanox TECHNOLOGIES

Driving High Performance Centralized NVMe Storage Arrays

Solution Brief E8 Storage - Mellanox Technologies

ABSTRACT

NVMe drives were introduced in 2014, heralding a new era in PCIe-attached flash. But until now, NVMe drives could only be used as local drives. To be used as shared storage, a new and critical requirement for NVMe connected drives would be high availability (HA) and very efficient networking. E8 Storage has produced the world's first HA centralized NMVe storage solution, which can unlock the economics and architectural advantages of centralized storage with the high I/O and low latency advantages of PCIe NVMe SSDs. E8 Storage's unique and patented distributed software architecture is able to extract the full performance of remote NVMe drives. E8 Storage integrates its rack scale flash architecture with the Mellanox ConnectX-4 network interface cards (NICs), thus enabling converged networking with very low latency and very high throughput and bandwidth. The benefit of using NVMe is that it is built from the ground up to support fast solid-state storage and doesn't carry the weight of legacy storage protocols originally designed for spinning disks.

E8 Storage's NVMe HA enclosure with its distributed software stack, combined with the Mellanox ConnectX-4 NICs, allows for a whole new breed of storage, since it combines the benefits of local flash and NVMe with the benefits of centralized storage and high availability. This is especially suitable for hyperscale data centers and large data centers that are deploying NVMe devices today, as well as enterprise and private cloud efforts seeking next-generation all-flash arrays.

INTRODUCTION

NVMe SSDs have been revolutionizing the data center, bringing 10x the performance and density compared to SATA and SAS devices. This performance and density explosion has pushed data centers and enterprises to deploy NVMe only as local drives within the server. Connecting them remotely from the server as JBOF (Just-a-Bunch-Of-Flash) is the ultimate goal but will impact their latency, and there had been no products or solutions available before today that could extract the full bandwidth and throughput of many remote NVMe SSDs.

E8 Storage allows a customer to enjoy the benefits of local flash - including bandwidth, throughput and latency - with the benefits of centralized storage - including shared volumes, centralized provisioning, highly available storage and fault tolerance. E8 Storage's patented architecture has been designed to maximize the performance of NVMe for high-performance enterprise applications, including real-time market data analytics, hyperscale data centers and high-performance computing.

HIGH AVAILABILITY AND HA ENCLOSURES

For centralized storage, high availability is critical. It ensures that a single failure in any of the storage array components, e.g. front-end network ports, internal CPU/RAM, etc., will not result in the loss of a large and expensive group of SSDs. Also, maintaining 2 or 3 replicas (instead of relying on centralized storage high availability) is unaffordable when it comes to NVMe SSDs. Centralized NVMe therefore requires HA.

E8 storage
Rack-Scale Flash. No Compromise.

Mellanox TECHNOLOGIES

THE CHALLENGE

Using local NVMe drives introduces a problem that shared storage solved long ago: local SSDs are islands of storage and their capacity must be determined upfront when buying a SSD for the lifetime duration of the server's life. Customers normally overprovision local flash, purchasing extra SSDs to avoid locking out any server as the SSDs fill up. This results in a lot of stranded capacity; on average up to 70% of the capacity of SSDs is unallocated.

Moving to shared NVMe has several advantages. First of all, it eliminates the need to determine up front how much SSD capacity is required for each server when building a new server farm. Additional SSDs can be procured when the capacity of the existing SSDs runs out instead of procuring them in advance. This means that customers can benefit from the declining price of flash rather than paying higher SSD prices up front. Secondly, it was common practice up until now that servers were maintained separately from storage. This means that with shared NVMe server downtime does not require storage downtime and vice versa. Thirdly, the lifecycles of servers and storage in the data center are kept separated, e.g., if servers need to be replaced every 2 years due to TCO demands, the new servers can share the existing storage instead of replacing the local storage. However, if the NVMe SSDs reside inside the servers, either they are not replaced at all (leading to an inefficient data center), or the SSDs need to be migrated from server to server (which is expensive). This separation of storage and compute leads to higher efficiency and lower costs in the data center.

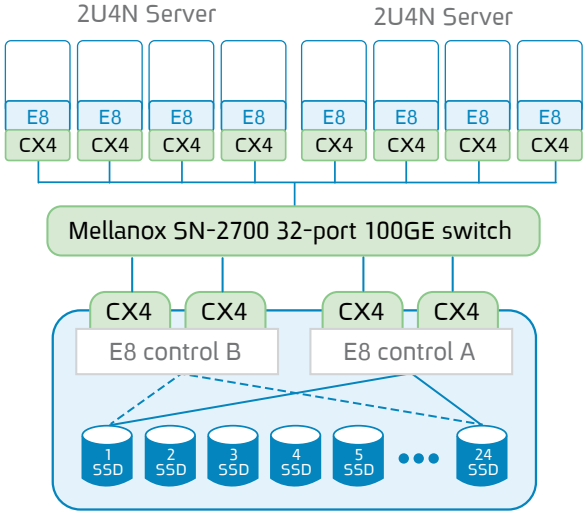
In order for NVMe drives to be used as shared storage, an NVMe enclosure and software stack are required. The storage technology used must enable full performance (throughput and bandwidth) without impacting latency. Shared storage also requires redundancy - not only in connectivity, e.g., via dual ports, but also redundancy around SSD failures, e.g., via RAID.

THE SOLUTION

NVMe SSDs with the E8 Storage HA NVMe enclosure delivers NVMe shared storage. The distributed architecture and software stack allow customers to achieve full NVMe performance of their remote drives (both throughput and bandwidth) and with minimal impact on NVMe latency (as compared to local NVMe drive usage). The high performance storage solution scales with capacity while maintaining the same level of performance over the network as expected from local SSDs inside the servers. The disaggregated architecture allows storage capacity to be dynamically allocated, augmented or replaced, without impacting the rack's performance or requiring maintenance downtime.

"Mellanox ConnectX-4 RDMA NICs combined with E8 Storage's enclosure enables a whole new breed of NVMe HA storage. Leveraging the benefits of local flash and NVMe with those of centralized storage and high availability is especially suitable for hyperscale data centers and large data centers that are deploying NVMe today." **Zivan Ori, CEO & Co-Founder @ E8 Storage**

"Mellanox agrees that high-availability and enhanced performance are vital ingredients to support deployments of NVMe. E8 Storage provides a complete NVMe solution with innovative software which leverages our RoCE RDMA solutions over 25, 50, and 100GbE networks to deliver a highly-available, high performance, shared flash storage solution." **Rob Davis, vice president of storage technology @ Mellanox Technologies**



E8 D24 NVMe HA Enclosure

E8 storage
Rack-Scale Flash. No Compromise.

Mellanox TECHNOLOGIES

Features

Hardware design without a single point of failure:

- 2 redundant front-end canisters, hot swappable
- Passive mid-plane design
- 2 PSUs
- RAID on NVMe SSDs, dual-parity support
- Dual-port SSDs support
- Power failure protection

Benefits

The combination of Mellanox ConnectX-4 NICs with E8 Storage HA NVMe enclosure addresses the storage needs in enterprise and hyperscale data centers by providing NVMe shared storage that has high performance, availability and reliability.

Additional Benefits Include:

- TCO is lowered by deploying NVMe storage where and when needed, without a need to predict storage consumption and pay for its purchase and maintenance before it is used.
- In hyperscale deployments, data is persistent (even in the event of compute node failure) achieving 100% uptime without performance degradation due to service rebalancing in the cluster.
- Storage investment is gradual instead of upfront. Customers can now buy NVMe storage when they need it, instead of when they start building the data center.
- Supply chain simplification - all servers are the same: any server in the rack can be provisioned with as much storage capacity and performance as needed, removing the need for multiple configurations and inefficiency in the supply chain.

Specifications*

Density	24 2.5" NVMe drives in a 2U enclosure	
Disk redundancy	Double parity RAID	
Connectivity	4-8 ports of 40GE / 50GE / 100GE	
Performance	**Read**	**Write**
Latency (4KB, QD=1)	100us	40us
IOPS (4KB)	10M	2M
Throughput (GB/s)	40	20
Power consumption	1200W max; 800W typical