

SOFTWARE-DEFINED BLOCK STORAGE FOR HYPERSCALE APPLICATIONS

SOLUTION BRIEF



SCALE-OUT SERVER SAN WITH DISTRIBUTED NVME, POWERED BY HIGH-PERFORMANCE NETWORK TECHNOLOGY

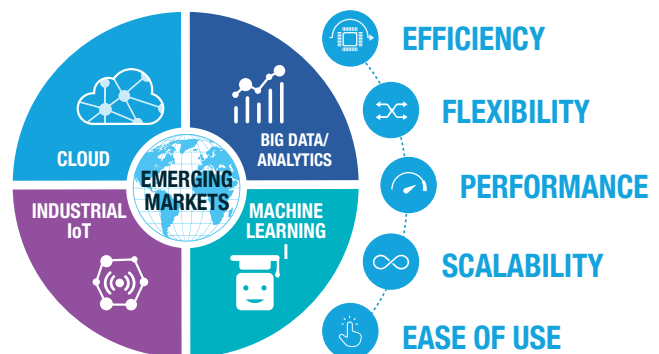
INTRODUCTION

The evolution in data-centric applications, such as Industrial IoT, machine learning and artificial intelligence have a massive impact on the volumes of data that need to be transferred, stored and processed. Moreover, real-time analytics demand the lowest levels of latency; hyper-scale web applications need unlimited scalability. The result is new application workloads across industries, which are driving the need for new storage architectures. To meet these storage requirements, enterprises and service providers are seeking to optimize their infrastructures in the same way as the Tech Giants have done. For storage, this means they want to deploy scale-out storage infrastructures using standard servers and Software-Defined Storage (SDS).

Broadcom and Excelero have teamed up to deliver a Software-Defined Block Storage solution for Hyperscale Applications leveraging industry standard servers with NVMe flash, Broadcom's BCM57414 RDMA-enabled Ethernet SmartNIC and Excelero NVMesh® Server SAN software, which allows unmodified applications to utilize pooled NVMe storage devices across a network at local speeds and latencies.

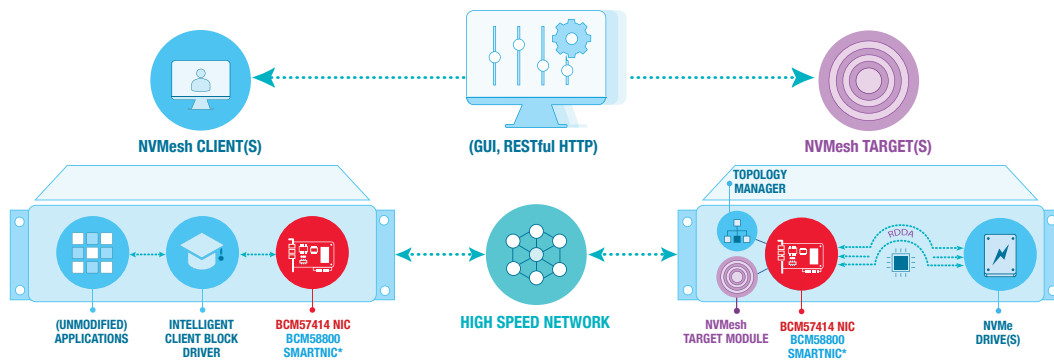
KEY BENEFITS:

- USE STANDARD SERVERS WITH STATE-OF-THE-ART FLASH COMPONENTS
- SHARE NVME ACROSS THE NETWORK AT LOCAL SPEEDS TO MAXIMIZE UTILIZATION AND AVOID DATA LOCALIZATION
- SCALE PERFORMANCE & CAPACITY LINEARLY, ACROSS DATA CENTERS WITHOUT LIMITS
- NO NOISY NEIGHBORS THROUGH FULL LOGICAL DISAGGREGATION OF STORAGE AND COMPUTE



EXCELERO NVMesh

Excelero’s Software-Defined Storage platform NVMesh enables customers to design Server SAN infrastructures for the most demanding enterprise and cloud-scale applications. NVMesh is a Software-Defined Block Storage solution that forms a Non-Volatile Mesh, a distributed block layer that allows unmodified applications to utilize pooled NVMe storage devices across a network at local speeds and latencies. Distributed NVMe storage resources are pooled with the ability to create arbitrary, dynamic block volumes that can be utilized by any host running the NVMesh block client. These virtual volumes can be striped, mirrored or both while enjoying centralized management, monitoring and administration. In short, applications can enjoy the latency, throughput and IOPs of a local NVMe device while at the same time getting the benefits of centralized, redundant storage.



A key component of Excelero’s NVMesh is the patented Remote Direct Drive Access (RDDA) functionality, which leverages RDMA to bypass the target CPU and, thus, avoids the noisy neighbor effect for the application. The shift of data services from centralized CPU to client side distribution enables unlimited linear scalability, provides deterministic performance for applications and enables customers to maximize the utilization of their flash drives. Elastic Virtual SAN is deployed as a virtual, distributed non-volatile array and supports both converged and disaggregated architectures, giving customers full freedom in their architectural design.

NVMesh FEATURES

Data Management & Protection

- Multiple Transports
- Logical Volumes
- Multiple Drive Types
- Failure Domains
- Patented RDDA, NVMe-ready
- Concatenated, RAID 0, RAID 1, RAID 10
- NVMe, NVmf, SATA, SAS
- Host, rack & row aware

Management & Monitoring

- Interactive Interfaces
- Automated Provisioning
- Web GUI & CLI commands
- RESTful API, Docker Persistent Volumes

Next-gen Data Center

- Flexible Topologies
- High Performance
- Scale-out Architecture
- Advanced Networking
- Physically converged, disaggregated or mixed
- Client, target and management scale independently
- Intelligent clients utilize multiple hosts, drives, network-paths and racks
- Ethernet (RoCE v2), InfiniBand

BROADCOM ETHERNET SOLUTION

Based on **Broadcom’s scalable 10/25/50/100G Ethernet controller architecture**, NetXtreme® E-Series standard PCIe Networking Interface Cards (NICs) can be used to build highly-scalable, feature-rich RoCE networking storage solutions in servers for Enterprise and Cloud-scale applications, including NVMe and **NVMe-oF**.

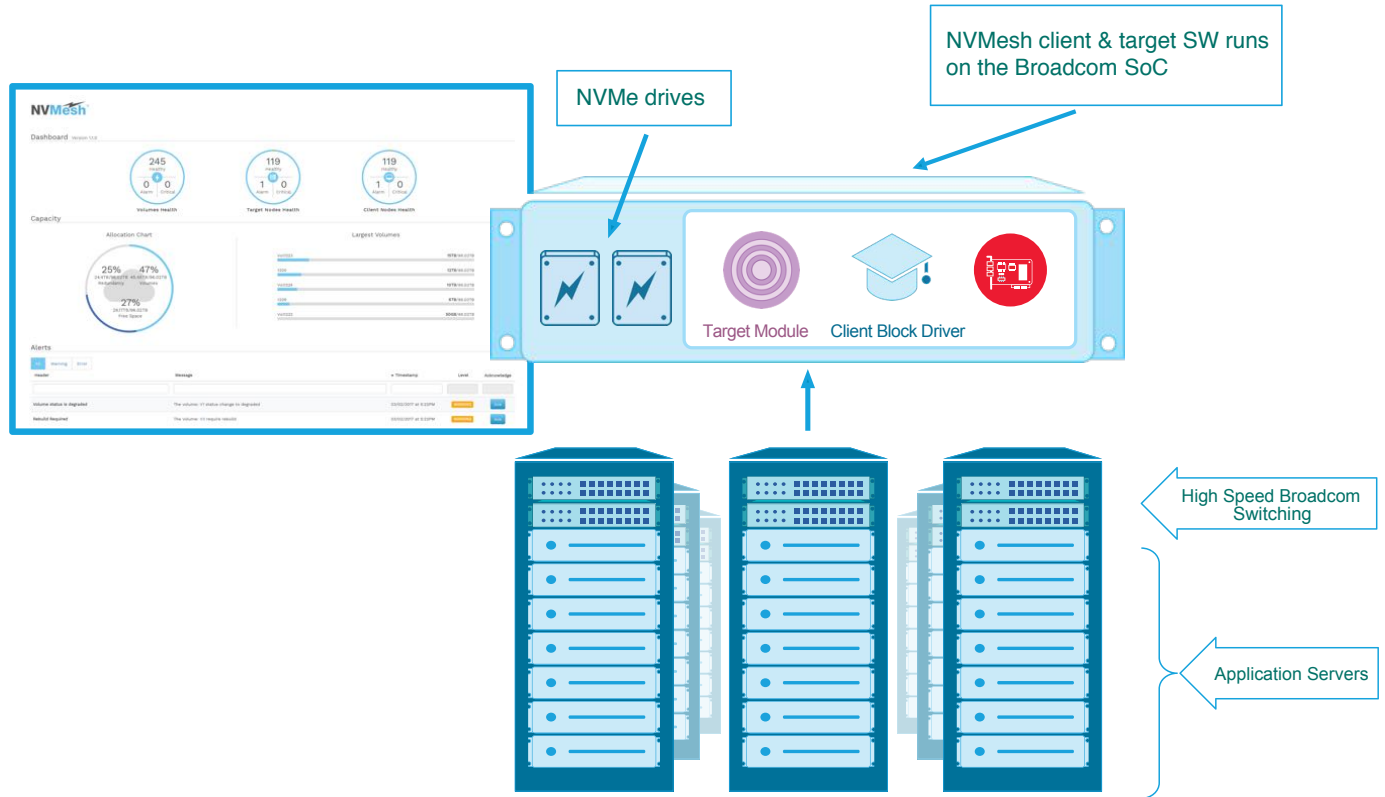
NetXtreme®-E Series addresses the performance and service demands of mega-scale data center networks for high throughput and advanced flow processing. Features such as TruFlow™ increase VM density up to 50%, freeing more CPU cycles for additional virtual machines. On-chip tunneling protocol processing for Geneve, VXLAN, and NVGRE provides up to a 5X throughput increase while lowering CPU utilization up to 90% compared to software-only solutions. In addition, NetXtreme®-E Series solutions support advanced networking technologies such as RoCE, SDN and NFV, to facilitate the implementation of high-performance networks and to enable the most robust storage solution.



JOINT SOLUTION WITH NVMesh ON NetXtreme

The joint Broadcom-Excelero solution provides remote, high-speed, low-latency shared storage with in-server flash performance. Customers can deploy the solution on existing application clusters and turn them into converged infrastructure with compute and a high-performance storage without affecting the application performance. With this "just add software" approach, customers remove storage bottlenecks without the need to purchase external storage arrays or servers. Alternatively, customers can choose for a disaggregated architecture and add NVMe storage nodes that can be accessed by application servers with the same performance as in-server NVMe.

ARCHITECTURE

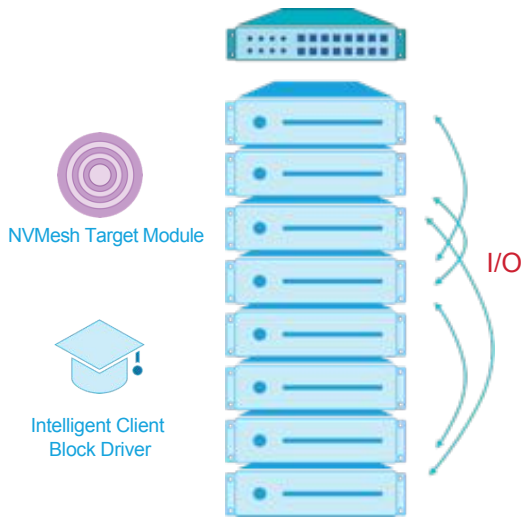


The solution was designed to scale naturally. Targeting larger data centers with upwards of 10,000 converged nodes, NVMesh was designed to avoid any centralized capability for the data path. This eliminates centralized metadata and lock management. Client, target and management functionalities are built as scale-out technologies. As it is critical to ensure that the networking pattern is suitable for scale, clients are completely independent. Rare knowledge sharing occurs indirectly and anonymously via targets. Client independence enables easy client scaling.

The Broadcom-Excelero solution was designed to let applications enjoy full performance, capacity and processing power of underlying servers and storage. Leveraging the NVMesh RDDA technology, the solution adds a mere 5µs latency over local NVMe drive latency, allowing customers to leverage full performance of their hardware at data center scale. The solution enables customers to achieve higher performance levels than proprietary all-flash-arrays at a much lower \$/IOP, or \$/GB/s.

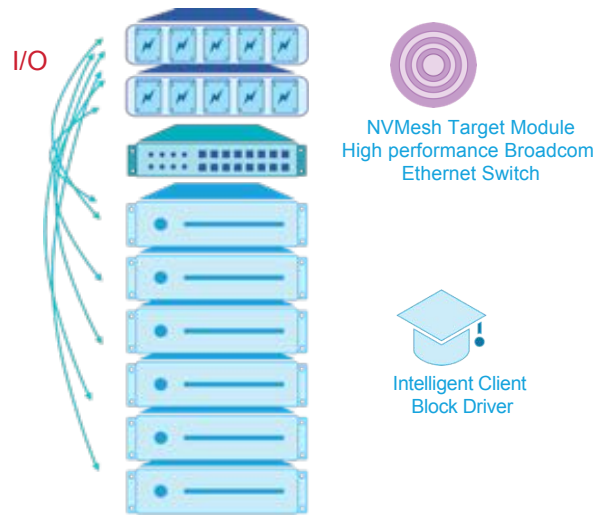
NVMe ON NetXtreme ARCHITECTURE

Local Storage in Application Server



- Storage is unified into one pool, linearly scalable
- NVMe Target Module & Intelligent Client Block Driver run on all nodes
- NVMe client & target SW runs on the Broadcom SoC
- NVMe bypasses server CPU

Storage is Centralized



- Storage is unified into one pool
- NVMe Target Module runs on storage nodes
- Intelligent Client Block Driver runs on server nodes
- NVMe client & target SW runs on the Broadcom SoC
- Applications get performance of local storage

PERFORMANCE:

As part of the integration and validation process, the Excelero integration team performed a series of tests to determine IOPs and bandwidth performance. The tests were performed remotely on test nodes that had 2x25Gbps connections and Samsung SM963 drives (280k iops in 4k randread). The tests consisted of various 4k random reads on Raid0 configurations. Utilizing both 25GbE ports, we were able to achieve a maximum of 1.339M 4K random read IOPs, from a RAID-0 NVMe logical volume striped across 6 NVMe SSDs.

Test Results:

Raid0-Wide=5
 100% randread 4k
 read: IOPS=1329k, BW=5190MiB/s
 (5442MB/s)(864GiB/170431msec)

Raid0-Wide=7
 100% randread 4k
 read: IOPS=1336k, BW=5218MiB/s (5471MB/s)(1440GiB/282580msec)

Raid0-Wide=6
 100% randread 4k
 read: IOPS=1339k, BW=5231MiB/s (5486MB/s)(9196GiB/1800002msec)

SOLUTION BENEFITS

VIRTUAL SAN

NVMesh is the only virtual SAN approach for shared NVMe, and it is the only solution that is 100% software-only. Customers can choose any standard servers and state-of-the-art flash drives to meet all their storage require-

UNIFIED NVMe

NVMesh allows unmodified applications to utilize pooled NVMe storage across a network at local speeds and latencies. A unified pool of NVMe enables customers to maximize NVMe utilization and avoids data locality issues for the application.

NO CPU

NVMesh enables 100% converged infrastructure by full logical disaggregation of storage and compute. The patented Remote Direct Drive Access (RDDA) bypasses the CPU, allowing applications to leverage the CPU to full capacity.

FLEXIBLE

NVMesh is the industry’s most flexible deployment model for Server SAN: it is deployed as a virtual, distributed non-volatile array and supports both converged and disaggregated architectures, and even mixed environments, giving customers full freedom in their architectural design.

100% EFFICIENCY

NVMesh is the only NVMe sharing technology that scales performance linearly at near 100% efficiency. The shift of data services from centralized CPU to complete client side distribution provides deterministic performance for applications and enables customers to maximize the utilization of their flash drives.

ABOUT EXCELERO

Excelero enables enterprises and service providers to design scale-out storage infrastructures leveraging standard servers and high-performance flash storage.

Founded in 2014 by a team of storage veterans and inspired by the Tech Giants’ shared-nothing architectures for web-scale applications, the company has designed a Software-Defined Block Storage solution that meets performance and scalability requirements of the largest web-scale and enterprise applications

With Excelero’s NVMesh, customers can build distributed, high-performance Server SAN for mixed application workloads. Customers benefit from the performance of local flash, with the convenience of centralized storage while avoiding proprietary hardware lock-in and reducing the overall storage TCO. The solution has been deployed for hyper-scale Industrial IoT services, machine learning applications and massive-scale simulation visualization.

With Excelero’s NVMesh, customers can build distributed, high-performance server SAN for mixed application workloads. Customers benefit from the performance of local flash, with the convenience of centralized storage while avoiding proprietary hardware lock-in and reducing the overall storage TCO. The solution has been deployed for hyper-scale Industrial IoT services, machine learning applications and massive-scale simulation visualization.