NVMe over Universal RDMA Fabrics

Build a Flexible Scale-Out NVMe Fabric with Concurrent RoCE and iWARP Acceleration

**EXECUTIVE SUMMARY**

Non-volatile memory express (NVMe) technology has made significant inroads within the storage industry over the past few years and is now recognized as a streamlined, high-performance, low latency, and low overhead interface for solid-state drives (SSDs). The industry and standard bodies have extended the specification beyond local attached server applications to deliver the same performance benefits across a network through the development of the NVMe over Fabrics (NVMe-oF) specification. This enables flash devices (such as PCIe SSDs and external storage arrays) to communicate over Ethernet-based remote direct memory access (RDMA) networks like RoCE and iWARP.

This paper presents a unique value proposition of Cavium FastLinQ NVMe-oF™ solution benefits:

- Reduce latency by 82% compared to traditional iSCSI
- Choice of 10/25/40/50/100GbE connectivity to host the most demanding storage workloads
- End-to-end NVMe-oF (NVMe™ over Fabrics) connectivity with both server and storage target solutions, including SPDK
- Delivers a flexible and future proof solution with support for concurrent RDMA over Converged Ethernet (RoCE), RoCEv2, and Internet Wide Area RDMA Protocol (iWARP)

**NVMe OVER FABRICS**

NVMe over Fabrics (NVMe-oF) enables the use of alternate transports to PCIe to extend the distance over which an NVMe host device and an NVMe storage drive or subsystem can connect. The NVMe-oF standard defines a common architecture that supports a range of storage networking fabrics for NVMe block storage protocol over a storage networking fabric. This includes enabling a front-side interface into storage systems, scaling out to large numbers of NVMe devices and extending the distance within a data center over which NVMe devices and NVMe subsystems can be accessed. The development of NVMe over Fabrics with RDMA is defined by a technical sub-group of the NVM Express organization.
Considering the ubiquity of Ethernet, RDMA fabrics like RoCE and iWARP are ideally suited for extending the reach of NVMe. With the availability of up to 100Gbps Ethernet connectivity in the data center and the highly scalable and deterministic nature of DCB-enabled Ethernet fabrics, NVMe over Fabrics solutions leveraging Ethernet based RDMA have the potential to deliver on cost, power, and performance characteristics required to scale out NVMe. (See Figure 1.)

**CAVIUM FastLinQ NVMe OVER FABRICS SOLUTIONS**

Cavium delivers the industry’s most comprehensive network adapter portfolio – FastLinQ Standard Ethernet Adapters and Converged Networking Adapters that cover the entire spectrum of customer Ethernet connectivity and offload requirements for architecting and deploying NVMe over Fabrics solutions.

Purpose built for accelerating and simplifying data center storage solutions, Cavium FastLinQ Ethernet technology delivers the following advantages (Figure 2):

1. **Broad Spectrum of Ethernet Connectivity** – 10/25/40/50/100GbE to host the most demanding storage workloads and deliver scalability, backed by a programmable architecture for future proof networking
2. **Universal RDMA** – Ultimate in choice and investment protection with concurrent support for RoCE, RoCEv2, and iWARP
3. **NVMe Direct** – Peer-to-peer flash access protocol enabling new efficiencies by enabling direct access to flash while offloading server CPU
4. **End-to-End Solutions** – Initiator and target mode solutions, including SPDK that leverage years of storage and storage networking experience
5. **Seamless Storage Migration** – Concurrent offload for NVMe-oF, iSER, iSCSI, and FCoE enables seamless upgrade paths to next gen storage

**Broad Spectrum of Ethernet Connectivity**

Cavium FastLinQ NICs deliver enterprise-class performance and reliability. With millions of Ethernet ports shipped, and a flexible architecture that delivers faster time to market, and adaptation to new and emerging technologies, FastLinQ Ethernet NICs are among the top choices for enterprise data centers.

Supporting a broad range of Ethernet connectivity speeds of 10/25/40/50/100GbE, Cavium FastLinQ NICs deliver scalability to drive business growth for the most demanding enterprise, telco, and cloud storage applications leveraging NVMe.
Universal RDMA

While RDMA has unique benefits of accelerating performance and offloading server CPU cycles and enabling a low latency fabric for scaling out NVMe, there exists multiple mutually incompatible standards for RDMA – two prominent Ethernet based such standards are RoCE and iWARP. RoCE requires a lossless Ethernet fabric and has a routable version called RoCEv2. iWARP relies on standard TCP offload, supports packet routing, and can operate on any Ethernet fabric. The Cavium FastLinQ 41000 and 45000 Series family is the only network adapter family that provides customers the technology choice and investment protection for NVMe-oF solutions with support for concurrent RoCE, RoCEv2 and iWARP. (See Figure 3.)

NVMe-Direct

Cavium FastLinQ Non-Volatile Memory Express (NVMe) Direct (NVMe-Direct) is a solution that efficiently supports high-speed, low-latency NVMe solid state drive (SSD)-based external storage over Ethernet networks. NVMe Direct offload is an advanced software technology for the Cavium FastLinQ 45000 and 41000 Series of Universal RDMA enabled Ethernet Adapters.

Using the programmable engines of Cavium FastLinQ Adapters, NVMe-Direct allows the network interface card (NIC) to execute read/write commands directly on the NVMe SSD, bypassing the host CPU. NVMe-Direct dramatically reduces CPU utilization, since the host CPU is bypassed during read/write transactions. In addition, using NVMe Direct, no host software is used in any part of the read/write operations, which further minimizes latency. The FastLinQ Series Adapters are designed to support read/write operations at wire speed on 100Gb Ethernet (100GbE) with I/Os as small as 4KB and performance that scales linearly when adding more Ethernet ports.

NVMe-Direct is an extension of a mature target technology that is transparent and compatible with existing initiator solutions. It significantly reduces risks and streamlines the deployment of NVMe over Fabrics, making it ideal for any application requiring Ethernet-based, high-performance flash storage.

End-to-end Solutions

Storage architects and customers require a solution that delivers maximum performance, is easy to implement, and is an all-in-one solution. Cavium FastLinQ end-to-end NVMe-oF solutions provide just that with best-in-class performance, functionality, and reliability.

For today’s high-performance environments, the Cavium FastLinQ 10/25/40/50/100GbE Universal RDMA-enabled adapters provide a true end-to-end NVMe-oF RDMA solution. This solution consists of PCI Express Universal RDMA enabled NICs, Kernel mode and user mode (SPDK) integrations (Figure 4), initiator and target mode frameworks and enterprise-class management software.

1. Read the NVMe Direct white paper.
Seamless Storage Migration

FastLinQ 10/25/40/50GbE Converged Network adapters support full protocol concurrent offload for NVMe-oF, iSER, iSCSI, and FCoE. FastLinQ adapters deliver up to 3M IOPS while consuming the fewest server CPU cycles, leaving headroom for virtual applications and higher ROI on server investments with seamless upgrade paths for next gen storage connectivity. Customers connecting to legacy SSDs or hybrid flash arrays considering a transition to NVMe-based all flash array would continue to leverage the seamless concurrent storage protocol offloads of FastLinQ adapters as they transition to next gen storage paradigms – without a rip and replace of their networking interconnects.

PERFORMANCE BENEFITS OF NVME OVER FABRICS UNIVERSAL RDMA

Cavium conducted a head-to-head performance analysis, comparing the efficiencies of existing/legacy storage transports to NVMe over Fabrics RDMA transports to a single NVMe drive. The results clearly indicate that to leverage the best potential of data center investments in NVMe, RDMA is the clear and consistent choice.

Figures 6 and 7 show the results from the NVMe transport comparisons. While the iSCSI software initiators added 82% more latency (Figure 6), NVMe-oF delivers near native performance (Figure 7).
SUMMARY

NVMe over Fabrics is a standard that allows new high-performance SSD interface, Non-Volatile Memory Express (NVMe), to be connected across RDMA-capable networks. Deploying high-performance Cavium FastLinQ 10/25/40/50/100GbE networking solutions with universal RDMA, enable an efficient, flexible, and future-proof deployment of NVMe over Fabrics.

With broad end-to-end support for NVMe over Fabrics and a decade of storage and storage networking expertise, Cavium technologies are perfectly suited to bridge the gap between storage performance and storage network performance.

ABOUT CAVIUM

Cavium, Inc. (NASDAQ: CAVM), offers a broad portfolio of infrastructure solutions for compute, security, storage, switching, connectivity and baseband processing. Cavium’s highly integrated multi-core SoC products deliver software compatible solutions across low to high performance points enabling secure and intelligent functionality in Enterprise, Data center and Service Provider Equipment. Cavium processors and solutions are supported by an extensive ecosystem of operating systems, tools, application stacks, hardware reference designs and other products. Cavium is headquartered in San Jose, CA with design centers in California, Massachusetts, India, Israel, China and Taiwan.

Figure 7. Comparison of IOPS Between iSCSI and RoCE as a Fabric for NVMe