NVMe over Fabrics

100G Ethernet NVMe Storage JBOF with Application Acceleration

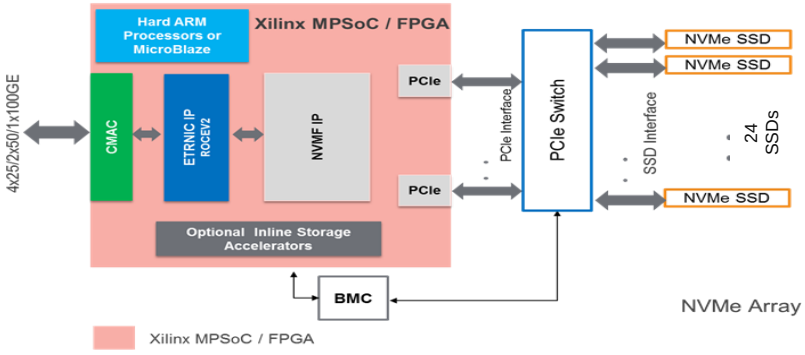
SOLUTION BRIEF

XILINX ALVEO™

- 100Gb/s NVMe-oF Alveo U50 based solution
- Support for up to 24 NVMe SSDs
- 100GbE line rate performance
- Ability to add Hardware Accelerators

INTRODUCTION

Unlike most of the JBOF designs in today's storage marketplace, this Xilinx® NVMe-over-Fabric (NVMe-oF™) reference was created with the idea of adding computational storage into next generation networked storage solutions. Using Remote Direct Memory Access (RDMA) this design provides a low latency, high performance, industry standard interconnect for up to 24 NVMe SSDs. This platform provides the flexibility to define custom acceleration functions within an NVMe-oF compliant environment and eliminates the need for an external processor or Network-Interface-Card (NIC) thus enabling a highly integrated and cost-effect JBOF storage solution.



SOLUTION OVERVIEW

This design transports NVMe packets over a standard RDMA Ethernet fabrics allowing the creation of remote storage namespaces, which enables the efficient pooling and sharing of storage resources across datacenter servers. These remote namespaces can dramatically reduce the storage cost, footprint and power within datacenters. Adding to this is the ability to incorporate acceleration services via name space management and this Alveo™ U50 base NVMe-oF solution become a true disaggregated computational storage accelerator.

This Xilinx solution provides reliable transport of NVMe frames with low latency, high throughput and massive scalability to remote hosts. A block diagram of a typical system solution is depicted above in Figure 1. The Xilinx NVMe-oF reference design implements the NVM express over fabric protocol and the RDMA NIC protocol in the single highly integrated Xilinx FPGA contained in an U50 add in card with a significant amount of programable logic left over for use as computational storage accelerators.

The key data transfer commands in the NVMe-oF protocol are offloaded to hardware, while the embedded CPUs in the Xilinx device process the control plane commands giving this Xilinx solution significant performance advantages over processor only implementations. The Xilinx ERNIC IP integrated into the reference design provides reliable transport, flexibility in network interconnect and the performance to support line speed bandwidth. An implementation supporting 24 drives can be accomplished using just over 200K LUTs leaving the rest of the U50s resources for customizable acceleration engines or other types of differentiation.

NVMe over Fabrics

100G Ethernet NVMe Storage JBOF with Application Acceleration

SOLUTION DETAILS

Features	Description
Supported RDMA Protocol	RoCEv2
Network Side Interface	Up to 100Gb Ethernet 1x100GbE, 2x50GbE, 2x40GbE, 4x25GbE and 4x10GbE
PCIe Side Interface	Up to one PCIe Gen3 x16 or two Gen4 x8 interfaces
Number of Hosts	A maximum of 128 hosts
Send and Receive Queue-Pairs (QPs)	Up to 255, which is QP1- QP255
Completion Queues (CQ)	Up to 255
Completion Queues (CQ) Queue Depth	64 entries per queue
Latency	Hardware design limited only
Performance (4KB IOPS)	2.5M
Management Interfaces	SMBus, NVMe-MI
Future support	NVMe 1.4+ NVMe 1.1+ NVMe/TCP
Inline Accelerator Examples	Storage services: (De)Compression, (De)Encryption, Data protection Database Acceleration: Scan, Filter, Aggregate
Resource Utilization	The full solution resources include NVMF-IP, ERNIC-IP, CMAC, AXI-DMA and DDR-MIG IP. The resource utilization depends on selected configuration

RESULTS

Operation	Typical 4KB NVM-oF 100GbE Performance	Typical NVM-oF Solution Latency
Random Read	2.4M	25.0us
Random Write	2.1M	31.0us
Sequential Read	2.4M	22.75us
Sequential Write	2.1M	29.9us

TAKE THE NEXT STEP

Learn more about Xilinx Alveo accelerator cards
Learn more about Xilinx's NVMEoF Solution