SOLUTION BRIEF

NVMe In the Studio: For 8K and Beyond

The Digital Studios Data Challenge

Moore's law has overtaken modern film and video production in multiple dimensions. Resolution, frame-rate and color depth are all marching forward simultaneously. The arms race for clearer, crisper, more lifelike content is here and whatever image quality producers demand today is guaranteed to be obsolete tomorrow – driving the demand to create content now that will be future proof for years to come.

The Workflow Revolution

SDI (Serial Digital Interface) has not kept up with the accelerating video dynamics; higher-resolutions, increased color depths, more graphical animation and increases in frame rates. Proprietary SDI solutions are expensive, can't keep up with modern transitions and introduce challenges in the processes of rapid set up and tear down that is required when switching projects. All this has opened a door for the use of Internet Protocol technology (IP) as a video transport.

Concurrently, The revolution in high-resolution uncompressed workflows has overwhelmed the throughput capabilities of legacy Fibre Channel (FC) and Scale-Out NAS storage, making it increasingly unaffordable to use traditional RAID-based storage hardware to deliver performance. This has left left media and production companies searching for solutions to simplify operations and reduce costs. Fortunately, a simultaneous revolution is occurring in storage – delivering new levels of breakthrough performance.

NVMe in the Studio

The storage revolution is NVMe, the astonishingly fast interface protocol that bypasses legacy storage bottlenecks, getting data directly to servers and high-end workstations with no speed bumps. But left on their own, local workstation NVMe drives are under-utilized, non-redundant and individually managed. Pixit Media PixStor with Excelero NVMesh® takes NVMe drives in individual servers, or dedicated storage servers, and virtualizes them into aggregated volumes that can be rapidly accessed as a shared resource by multiple clients through NVMe compliant Mellanox ConnectX® network adapters. Excelero NVMesh with PixStor uniquely delivers centralized management, redundancy and high utilization rates, while maintaining local NVMe performance.

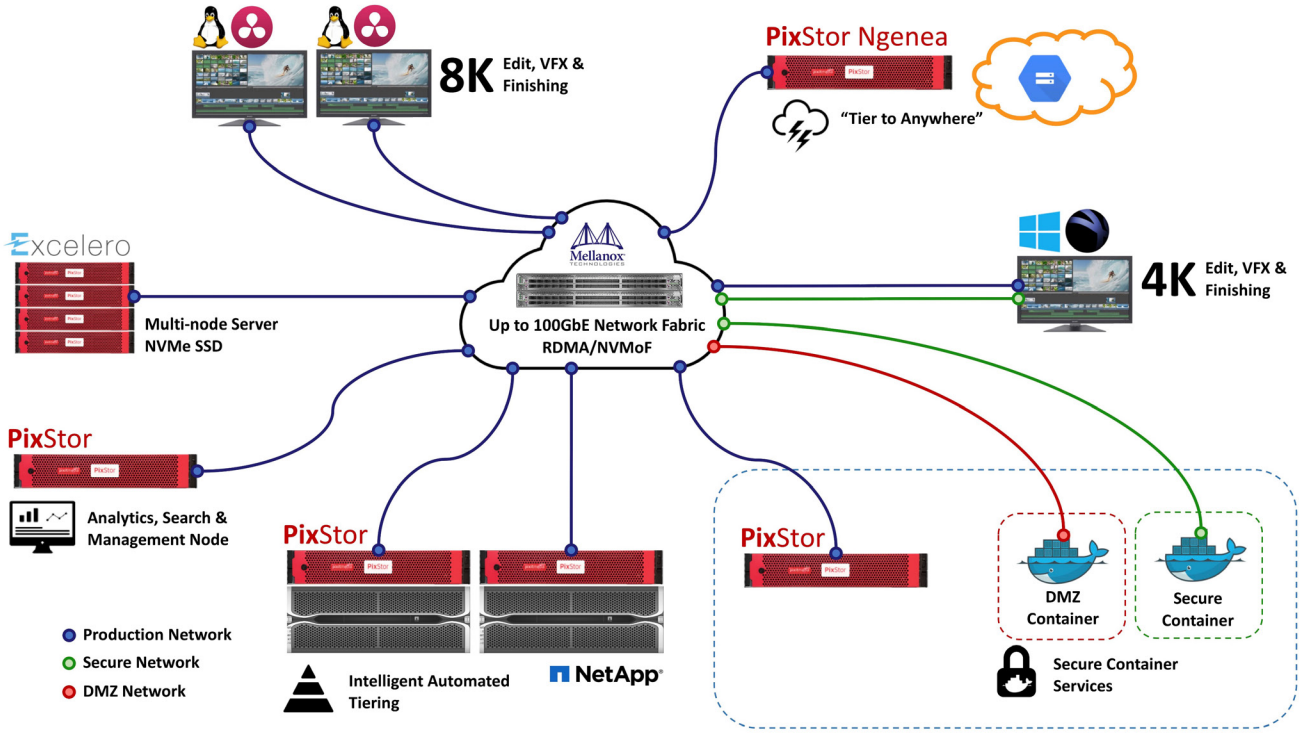
Highlights

- Mellanox Ethernet Video Fabric creates a network fabric that unleash the maximal power of the NVMe storage pool
- Bypass storage/server bottlenecks to deliver the highest possible performance, accelerating uncompressed 8K and beyond workflows
- Over 20 GB/s single stream performance to a workstation or server over Ethernet
- Almost zero overheard on either the client or NVMesh target server delivers performance to the client with almost 100% CPU and memory available for applications

Editing, finishing and VFX applications can now enjoy the low latency, high throughput and IOPS of a local NVMe device while getting the benefits of centralized, redundant storage. The combined solution provides a single namespace for applications that demand extreme performance. As shown below, a near-line tier can be added using traditional storage arrays with PixStor's data placement policy engine to automatically tier data to lower cost storage for high capacity requirements – with workflows extending to the cloud.

congestion management are implemented in the ConnectX NIC hardware to deliver reliability even with UDP over a lossy network. Mellanox Spectrum Ethernet switches provide up to 100GbE line rate performance and consistent low latency with zero packet loss. With its high performance, low latency, intelligent end-to-end congestion management and QoS options, Mellanox Spectrum switches are ideal to implement NVMeoF at scale. Combined with the Mellanox LinkX® they form a high-speed Ethernet Video Fabric (EFV) infrastructure that enable customers to extract additional value from their hardware and software investments by moving on to a fast and reliable NVMeoF infrastructure, providing a significant performance boost in the form of reduced latency and increased throughput.

Fig 1. Pixit Media Excelero Single Namespace 8K+ Cloud Integrated Architecture with Mellanox Ethernet Video Fabric



Ethernet Video Fabric

Excelero's NVMesh leverages and greatly benefits from Remote Direct Memory Access (RDMA)– running on RDMA over Converged Ethernet (RoCE)–a key feature of Mellanox ConnectX® adapters, and the Mellanox Spectrum® switches. RoCE provides direct memory access from the memory of one host (storage or compute) to the memory of another host without involving the Operating System and CPU. This boosts network and host performance by lowering latency, reducing CPU load and as a result increases bandwidth. In contrast, TCP/IP communications typically require copy operations, which add latency and consume significant CPU and memory.

Mellanox ConnectX adapters incorporate RoCE which is utilized by NVMe over Fabrics (NMVeoF) to provide best of breed performance. RoCE and

Removing Legacy Bottlenecks

Media and Entertainment companies are dealing with an explosion of data resulting from a move to higher video dynamics including increased resolutions and color depths, more graphical animation and an increase in frame rates.

This is quite the technology race and is driving increasing networking requirements since data rates directly affects the network required and drive the amount of storage that will be necessary.

Due to these new requirements, the networking infrastructure for the entire creative, pre and post-production and distribution processes within the media and entertainment industry has become a mission critical resource. These innovations are driving the need for high-speed networking infrastructure and fast access to storage. This in turn is driving a move to

fast and efficient open Ethernet networks, and file-based storage that can easily scales as content grows into the petabytes.

PixStor Excelero NVMesh volumes are served out over 100 Gb Ethernet, through open standards 100 GbE Mellanox Spectrum Ethernet switches. Spectrums flexible switch buffers and zero-packet loss provide predictable network performance, consistent and very low port-to-port latency and minimize jitter with QoS & DSCP features. In concert with Excelero NVMesh, there are no network bottlenecks nor from RAID controllers or proprietary hardware. With all legacy layers removed, all that remains is the raw speed of the NVMe SSD, the speed of the PCI bus on the server and the speed of the network connection. As these industry standard technologies improve, both PixStor and Excelero can exploit immediate performance boosts with small, incremental upgrades. Incremental improvements to commodity server, network and storage technology means incremental improvements to NVMesh performance 'for free', without having to tear up everything and begin again.

An Industrial Strength, Robust, Easy to Manage Storage Platform

PixStor is an easy to manage, scalable, high performance storage system with centralized control of all file system components from a single interface employing graphical interfaces for common tasks, and Python and REST APIs for file system automation and workflow integration. PixStor monitoring, alerting and performance analytics are tailored to the exact needs of the administrative workflow and empower the organization with unprecedented insight into data usage and trends.

Scale with Virtually No CPU Overhead

High end workstations connected to the NVMe network fabric see and address volumes directly, without having to go through the storage servers. If a NVMesh node or drive fails, redundancy is handled transparently without the need for administrative intervention. What you see on the clients is not the actual NVMe drives, it is the virtualized volumes. Tens of thousands of NVMe drives can scale capacity and performance linearly within a single namespace, making management and usage easy. Unique to the combined solution, there is almost zero overhead on the client or NVMesh target server delivering the performance to the client, leaving CPU and memory almost 100% available for applications.

Furthermore, RoCE which is utilized by the NVMeoF standard offloads data movement tasks from the CPU onto the ConnectX NIC hardware along with congestion management to deliver low latency and reliable performance even with UDP over a lossy network.

Table 2. 3 NVMesh Target Server Configuration Stream Count

Format	Resolution (width X height)	Frame Rate (fps)	Data Rates (MB/s)	Storage Capacity GB/Hour
SDTV (NTSC, 8-bit)	8192 x 4320	-30	31	112
HDTV (1080p, 8-bit) RGB	1920 x 1080	24	149	537
UHD-1 4K (10-bit) RGB	3840 x 2160	60	1,866	6,718
UHD-2 8K (12-bit) RGB	7680x 4320	120	17,916	64,497
Digital Cinema 2K (10-bit) YUV	2048 x 1080	24	199	717
Digital Cinema 4K (12-bit) YUV	4096 x 2160	48	1,910	6,880
Digital Cinema 8K (16-bit) YUV	8192 x 4320	120	25,480	91,729

Superior Performance on Industry Standard Hardware

Each NVMesh target can deliver 20 GB/s of read performance. An NVMesh target node fits within a simple, commodity server configuration that contains a CPU, a relatively small amount of memory, NVMe drives and Mellanox Ethernet connectivity. Off-the-shelf, Industry standard servers can fit four of these nodes in a 2U chassis, delivering over 80 GB/s of read performance to clients. This reduces power and space requirements by an order of magnitude, compared to conventional disk-based storage. This is an important benefit in many modern Studio environments where space and power are both very expensive and in very short supply.

Blazingly Fast Performance to a Client

Pixit Media has measured over 20 GB/s to a single server or workstation from an NVMesh system using only 2 X 100 Gb Ethernet connection. Comparably, throughput requirements under 10 GB/s can be delivered with only a single cable and card. With traditional Fibre Channel based solutions, up to 8 X 16 Gb FC connections would be required to meet this performance, with a much lower value per port compared with 100 Gb Ethernet.

Linearly Scalable Aggregate Performance

Web-scale and private cloud enterprises are undergoing a major transformation to increase agility and efficiency by introducing virtual networking.

These data center networks need to be much more responsive, efficient and offer agility to rapidly deploy new virtual network services. This must be architected into the network to enable the nimbleness necessary to meet growing demands all while conserving capital expenditure, reducing operating expenses, and accelerating time to delivery. Red Hat and Mellanox offer a tightly integrated Network Function Virtualization Infrastructure (NFVI) and cloud data center solution that removes deployment and interoperability barriers by combining Red Hat open cloud software and Mellanox Intelligent NICs as a joint solution. By offering best of the breed hardware offloads that improve server and network efficiency, performance, and scalability, deployment challenges are addressed for virtual networking workloads.

Table 2. 3 NVMesh Target Server Configuration Stream Count

Format / FPS	Aggregate Stream
10 Bit DPX, HD, 24 FPS	270 streams
10 Bit DP, 2K, 24 FPS	180 streams
10 Bit DPX, 4096 x 2160, 24 FPS	75 streams
16 Bit DPX, 4096 x 2160, 24 FPS	48 streams
10 Bit DPX, Full 4K, 24 FPS	48 streams
16 Bit DPX, Full 4K, 24 FPS	34 streams
10 Bit DPX, 8192 x 4320, 24 FPS	18 streams

Combined Workflows in a Single Namespace

PixStor with Excelero on top of a Mellanox EVF provide the performance needed to combine VFX, finishing and editing workloads all running at maximum speed without effecting the editors' or artists' experience. In addition to tiering data to the cloud with PixStor Ngenea HSM, PixStor PixCache facilitates efficient Burst Rendering onto unlimited compute nodes to take on extra projects or accelerate schedules. PixCache minimizes data transfer and maximizes data locality by caching on premises data close to cloud computing resources. Results are automatically transferred back to on-premises storage with no manual intervention.

Conclusion

The move to next-gen IP-based broadcast is inevitable and Mellanox end-to-end efficient Ethernet Video Fabric is radically changing the economics of broadcast applications. Combined with best-of-breed PixStor storage platform and Excelero NVMesh customers can build distributed, high-performance Server SAN for mixed application workloads. Customers benefit from local flash performance, with the convenience of centralized storage while avoiding proprietary hardware lock-in and reducing the overall storage TCO. This allows broadcasters to save time, money and reliably deliver new levels of breakthrough performance.

350 Oakmead Parkway, Suite 100
Sunnyvale, CA 94085
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com

MLNX-423558315-99363