SOLUTION BRIEF

SUPERMICRO SUPERSERVER® STORAGE APPLIANCES DELIVER ELASTIC NVME FOR AI WORKLOADS

BigTwin™, Ultra, or All-Flash Storage Servers with Direct-Attached Low-Latency NVMe Expand your GPU Storage Pool

EXECUTIVE SUMMARY

Supermicro SuperServers running Excelero's Elastic NVMe enable AI data scientists to process massive amounts of data in minutes, rather than hours or days. Optimized to eliminate compute performance bottlenecks and storage overhead, the BigTwin, Ultra, and All-Flash Storage Servers supporting NVMesh deliver up to 20x faster data processing for GPU compute nodes. This reference architecture enables you to swiftly add direct-attached NVMe storage capacity to your GPU server, as one single pool of storage without performance penalty.

SUPERMICRO® - *July 2020*

NVMESH BIGTWIN DIRECT-ATTACHED REFERENCE ARCHITECTURE SPECS

Supermicro has been a leader in introducing NVMe support and advanced NVMe features in our product portfolio and BigTwin continues that tradition. BigTwin supports 24 hot-swappable 2.5" U.2 NVMe drives, or mixed configurations with SAS3 and SATA3 drives (varies by different BigTwin models).

The key benefit of BigTwin is the no-compromise design. Historically multi-node systems traded off features and capacity for higher density. They were deployed for workloads that did not require the highest performance or the highest memory density on a single node.

The new 2U BigTwin design is a breakthrough multi-node system that supports the highest performing CPUs, full 24 DIMMs of memory and up to 24 all-flash NVMe SSD drives.

This NVMesh direct-attached Elastic NVMe GPU storage reference architecture leverages the Supermicro BigTwin™ architecture, featuring a 2U 4-node system that delivers the highest performance and efficiency per rack unit. The system comes with 24 All-Flash NVMe drives, totaling 368.64 TB of raw capacity. NVMesh enables users to scale infrastructure by adding more systems.

CONFIGURATION WITH HIGH AVAILABILITY (HA):			
2U 4-Node server with 24 NVMe drives that enables an HA solution with MeshProtect-0, 1, 10 & 6 @ 8+2			
	Supermicro SKU	**DESCRIPTION**	**QTY**
SERVER	SYS-2029BZ-HNR	BigTwin 3UPI 2U 4-Node, 6x2.5" NVMe per node, X11DPT-BH, 217BHQ+	1
CPU	P4X-CLX4216-SRFBB	CLX 4216 2P 16C/32T 2.1G 22M 9.6GT 100W 3647 L1	8
MEMORY	MEM-DR416L-CL02-ER29	16GB DDR4-2933 1RX4 LP ECC RDIMM	48
M.2 CARRIER	AOC-SMG3-2H8M2-B-O	2x Hybrid NVMe/SATA M.2 Carrier for Big Twin,HF,RoHS	4
BOOT DRIVE	HDS-IMT0-SSDSCKKB960G8	Intel D3 S4510 960GB M.2 SATA 6Gb/s 3D TLC 22x80mm 1DWPD	8
NVME DRIVES**	HDS-MUN-MTFDHAL15T3TDP1A	Micron 9300 PRO 15.3TB NVMe PCIe 3.0 3D TLC U.2 15mm	24
IB/NIC*	AOC-MCX556A-ECAT	MCX556A-ECAT, CX-5 VPI,EDR IB,100GbE,2p,QSFP28,PCIe3x16	8
SIOM NIC	AOC-MGP-I4M-O	SIOM 4-port GbE RJ45, Intel i350 with 1U bracket	4
SUPERMICRO SYSTEM MANAGEMENT SOFTWARE	SFT-DCMS-SINGLE	Supermicro System Managment Software Suite node license,HF,RoHS/REACH,PBF	4
NVMESH	SFT-EX-SWCCI24-3YR	NVMesh per-chassis subscription license, up to 24 NVMe devices. Includes unlimited Clients. 3 year subscription including Premium Support.	1
	SVC-EX-PSINSTALL	On-site, NVMesh Installation Services, per day, per 100 Client and/or Target Nodes	1

CONFIGURATION WITHOUT HIGH AVAILABILITY (NON-HA): 1U 10 drive bay server with 10 NVMe drives that allows a non-HA solution with MeshProtect-0 & 6 @ 8+2			
	Supermicro SKU	**DESCRIPTION**	**QTY**
SERVER	SYS-1029U-TN10RT	X11DPU, 119UTS-R1K02P-N10T, AOC-URN6-I2XT	1
CPU	P4X-CLX4216-SRFBB	CLX 4216 2P 16C/32T 2.1G 22M 9.6GT 100W 3647 L1	2
MEMORY	MEM-DR416L-CL01-ER29	16GB DDR4-2933 2RX8 LP ECC RDIMM	6
BOOT DRIVE	HDS-IMT0-SSDSCKKB960G8	Intel D3 S4510 960GB M.2 SATA 6Gb/s 3D TLC 22x80mm 1DWPD	1
NVME DRIVES**	HDS-MUN-MTFDHAL15T3TDP1A	Micron 9300 PRO 15.3TB NVMe PCIe 3.0 3D TLC U.2 15mm	10
IB/NIC*	AOC-MCX556A-ECAT	MCX556A-ECAT, CX-5 VPI,EDR IB,100GbE,2p,QSFP28,PCIe3x16	2
SUPERMICRO SYSTEM MANAGEMENT SOFTWARE	SFT-DCMS-SINGLE	Supermicro System Managment Software Suite node license,HF,RoHS/REACH,PBF	1
NVMESH	SFT-EX-SWCCI12-3YR	NVMesh per-chassis subscription license, up to 12 NVMe devices. Includes unlimited Clients. 3 year subscription including Premium Support.	1
	SVC-EX-PSINSTALL	On-site, NVMesh Installation Services, per day, per 100 Client and/or Target Nodes	1

EDSFF-SHORT E1.S SERVER			
	Supermicro SKU	**Description**	**Qty**
SERVER	SSG-1029P-NES32R	CSE-121EF + X11DSF-E-P	1
CPU	P4X-CLX4210-SRFBL	CLX 4210 2P 10C/20T 2.2G 13.75M 9.6GT 85W 3647 R1	2
MEMORY	MEM-DR416L-SL04-ER26	16GB DDR4-2666 1Rx4 LP ECC REG DIMM,HF,RoHS	12
BOOT DRIVE	HDS-MMT-MTFDDAV960TDS1AW	Micron 5300 PRO 960GB, SATA, M.2, 22x80mm,3D TLC,1.5DWPD	2
AOC	AOC-SLG2-2TM2-T-O	OEM LP PCI-E 2 2x 7pin SATA3 RAID 1	1
NIC CARD*	AOC-MCX556A-ECAT	Mellanox ConnectX-5 VPI adapter card, EDR IB(100Gb/s) & 100GbE, dual port QSFP28	2
NVME	HDS-IEN0-SSDPEYKX040T8	Intel DC P4511 EDSFF E1.S TLC 5.9mm 4TB	32
SUPERMICRO SYSTEM MANAGEMENT SOFTWARE	SFT-DCMS-SINGLE	Supermicro System Managment Software Suite node license,HF,RoHS/REACH,PBF	1
NVMESH	SFT-EX-SWCCI32-3YR	NVMesh per-chassis subscription license, up to 32 NVMe devices. Includes unlimited Clients. 3 year subscription including Premium Support.	1
	SVC-EX-PSINSTALL	On-site, NVMesh Installation Services, per day, per 100 Client and/or Target Nodes	1

EDSFF-LONG E1.L SERVER

|	Supermicro SKU	Description	Qty
SERVER	SSG-1029P-NEL32R	1U DP 32 EDSFF Long	1
CPU	P4X-CLX4210-SRFBL	CLX 4210 2P 10C/20T 2.2G 13.75M 9.6GT 85W 3647 R1	2
MEMORY	MEM-DR416L-SL04-ER26	16GB DDR4-2666 1Rx4 LP ECC REG DIMM,HF,RoHS	12
BOOT DRIVE	HDS-MMT-MTFDDAV960TDS1AW	Micron 5300 PRO 960GB, SATA, M.2, 22x80mm,3D TLC,1.5DWPD	2
NIC CARD*	AOC-MCX556A-ECAT	Mellanox ConnectX-5 VPI adapter card, EDR IB(100Gb/s) & 100GbE, dual port QSFP28	2
NVME	HDS-IEN0-SSDPEWNV153T8	Intel D5-P4326 E1.L PCI-E 3. 1x4QLC9.5mm15.36T 3D2 <0.5DWPD	32
SUPERMICRO SYSTEM MANAGEMENT SOFTWARE	SFT-DCMS-SINGLE	Supermicro System Managment Software Suite node license,HF,RoHS/REACH,PBF	1
NVMESH	SFT-EX-SWCCI32-3YR	NVMesh per-chassis subscription license, up to 32 NVMe devices. Includes unlimited Clients. 3 year subscription including Premium Support.	1
	SVC-EX-PSINSTALL	On-site, NVMesh Installation Services, per day, per 100 Client and/or Target Nodes	1

* InfiniBand or RoCE

** The architecture supports NVMe drives from all major manufacturers, including Micron, Intel, Kioxia, Samsung and Western Digital. Different drive sizes are supported. For Erasure-Coding configurations, NVMe drive families, including Western Digital SN200, Samsung 1725b & 1733, Kioxia CM5 & CM6 with 3 DWPD or more endurance are supported.

FULLY FLEXIBLE STORAGE ARCHITECURES DESIGNED WITH NVMESH

The essence of data-driven sciences such as artificial intelligence (AI), including machine learning (ML) and deep learning (DL), is the ability to process as much collected and simulated data as possible in the shortest amount of time. As data sets continue to grow exponentially, processing massive volumes of data at high speed has become one of the major challenges for modern data centers.

GPUs have become the go-to compute resources behind AI workloads, and NVMe flash has become the standard for high-performance, low-latency storage. By providing GPUs with direct access to an elastic pool of NVMe, AI data scientists and HPC researchers can feed far more data to the applications – they no longer face data bottlenecks and can get to better results faster.

Designed as a 100% software-defined storage solution, NVMesh gives customers full flexibility in designing the storage architecture that best meets their business and application requirements. This direct-attached NVMe storage reference architecture leverages the flexibility of NVMesh but aims to provide AI customers with the simplest and easiest-to-deploy solution for low-latency GPU storage.
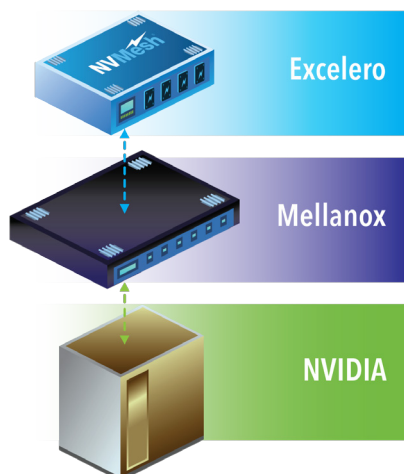
DIRECT-ATTACHED NVME STORAGE APPLIANCE

Add capacity to your GPU server(s) with an all NVMe storage appliance. Local (in GPU server) and remote (NVMe appliance) NVMe drives are joined in one single pool of high-performance storage with local latency, accessible through a local file system. This ensures no changes are required to your existing workflow.
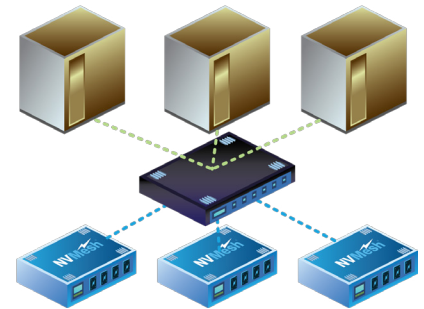
SOLVING YOUR GPU BOTTLENECK

BENEFITS OF ELASTIC NVME FOR MACHINE LEARNING AND AI:

- Exceeds the performance and capacity limits of local flash on GPU servers

- Accesses remote NVMe at local speed

- Shares NVMe resources across multiple GPU servers

- Eliminates the need to copy data locally, conserving time and drive endurance

- Datasets can exceed the actual capacity of the GPU Server

- Full CPU offload with Excelero's patented Remote Direct Drive Access (RDDA) technology

When data scientists are training machine learning models, they process hundreds of terabytes of data. A training dataset may contain thousands of satellite images, for example, to map the effects of climate change. The faster training runs are completed, the more they can be repeated and the better the training model becomes or the more models can be trained on the infrastructure.

GPUs are a game changer for training on large datasets because they are so efficient for matrix multiplication and convolution. This is due to a combination of their parallel computing functionality and memory bandwidth. But the biggest advantage of modern GPU computing is also creating its biggest challenge: GPUs have an amazing appetite for data, and AI workloads need much more capacity than it is available locally in GPU systems.

NVMesh solves the 'storage bottleneck' that AI and ML users report they often experience. It allows bypassing CPUs all the way from GPU memory to NVMe devices achieving frictionless access to NVMe's superior performance for shared NVMe storage at local speed.

Excelero's elastic NVMe enables machine learning workloads to process more data, much faster. NVMesh delivers low-latency (5μs added to media's read latency), high bandwidth block storage for AI and ML workloads.

This direct-attached architecture enables shared NVMe with a local file system. GPU-based systems benefit from the performance of local NVMe flash with the convenience of centralized storage while avoiding proprietary hardware lock-in and maximizing the overall GPU return on investment (ROI).


NVMESH - ELASTIC NVME

Excelero's NVMesh eliminates any compromise between performance and practicality, and allows GPU optimized servers to access scalable, high-performance NVMe flash storage pools as if they were local flash. This technique ensures efficient use of both the GPUs themselves and the associated NVMe flash. The end result is higher ROI, easier workflow management, and faster time to results.
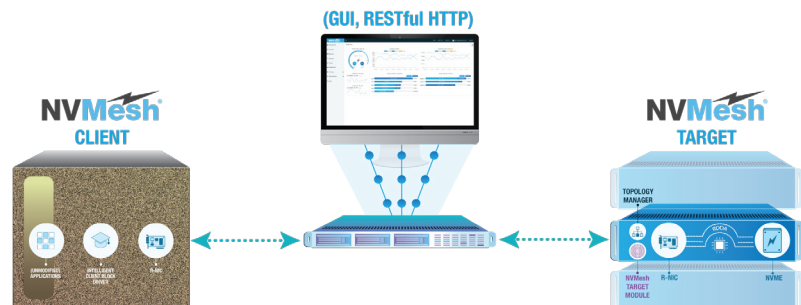
NVMESH FEATURES FOR GPU

- NVMesh unifies local and remote NVMe devices into a logical block pool that performs the same as local NVMe flash

- NVMesh allows full utilization of the I/Os and bandwidth capabilities of NVMe drives across a network

- GPU servers can use their massive network connectivity to access remote NVMe logical volumes, with redundancy if desired!

- MUCH faster than local SATA SSDs

- Larger shared pools than possible within the platform

- Random I/O characteristics of NVMe preserved, achieving tens of millions of potential I/Os at minimal latencies

The storage services in NVMesh are implemented as client-side services, with target-side CPU bypassed completely from the data path. Getting rid of any centralized entity on the data path for metadata or lock management means that the clients are totally independent and there is no inter-client communication. This allows GPUs to gain access to the full performance of NVMe media at any scale, starting from a few drives in a server to a massive scale of hundreds or thousands of storage nodes.

PERFORMANCE

The NVMesh direct-attached reference architecture provides 19M 4K IOPs and as many as 80GB/s of aggregate bandwidth, with only 5μs added latency (over the media's latency). This baseline reference architecture scales linearly as more systems are added.

The following configuration was used for performance testing:

NVMesh target Server: SuperServer SYS-2029BZ-HNR
NVMe drive: 24 Micron 9100 2.4TB NVMe U.2 drives
OS: RHEL 7.3
SW ver: NVMesh v1.2.1
Traffic was generated using FIO running on 12 clients with: 2x Intel E5-2690v4 & 1x Mellanox CX-5 per client (3x SuperMicro BigTwins SYS-2028BT-HNR+ servers with 4 nodes each)

ABOUT SUPER MICRO COMPUTER, INC.

Supermicro® (NASDAQ: SMCI), the leading innovator in high-performance, high-efficiency server technology is a premier provider of advanced server Building Block Solutions® for Data Center, Cloud Computing, Enterprise IT, Hadoop/Big Data, HPC and Embedded Systems worldwide. Supermicro is committed to protecting the environment through its "We Keep IT Green®" initiative and provides customers with the most energy-efficient, environmentally-friendly solutions available on the market.

www.supermicro.com