PowerEdge XE8640

DELLTechnologies

Product Message: Drive Generative AI, traditional AI, HPC modeling and simulation workloads with superior performance

Positioning: Purpose-built performance for AI & HPC workloads with optimized compute and accelerated throughput to automate analysis into insights.

Key Fact: GPU Optimized, air-cooled design supporting 4 x NVIDIA H100 accelerators with exceptional GPU bandwidth (2-socket, 4U rack server)

Product Positioning

Target Customers

Customers looking for accelerated compute to maximize performance in a dense air-cooled, scalable server architecture for workloads like
- High Performance Computing
- AI, ML & DL training
- Generative AI

Ask This / Listen For....

- How far along is your organization in deploying and scaling out production-level HPC, AI, Machine Learning (ML) and Deep Learning training projects?
- Are you looking to train models or run HPC applications
- Do you need NVLink GPUs, and is time to value critical for your business?
- Is your power per rack between 6kW-40kW consumption and you need air cooled servers?

Opportunities

- Adoption of demanding AI, ML, DL, HPC workloads and urgency to deploy is adding complexity to data center operations. Time to results (TTR) is of primary importance
- AI, ML, DL and HPC applications need optimal performance in a production environment and to deploy at scale while driving outcomes

Product Advantages

- Expect 2-4x[1] performance improvement (training and inferencing) with XE8640 over current XE8545 platforms, from H100 NVLink SXM5 GPUs, having the highest peer to peer bandwidth (900Gb/s) in a low latency switchless design.
- Support for up to 4 PCIe Gen 5 slots, with 1:1 GPU:NIC ratio combined with PCIe Gen5 switch architecture enables GPU Direct that minimizes latency by direct data transfers
- Air cooled design supporting the highest wattage CPUs and GPUs, up to 35C ambient.
- XE8640 simplifies data center operations as it fits in standard rack, doesn't need extended rack rails, and has comprehensive systems management support.

Customer Challenges/Pain Points

- Requirements to implement demanding HPC and AI workloads.
- Looking for SXM (NVLink) GPUs for training medium to large models to achieve faster time to value
- Current HPC infrastructure requires specialized equipment such as liquid cooling, extended rack space, and lacks management software
- Looking for full standard PowerEdge management, capability and connectivity features.

Product

[1]Dell Internal testing, results based on upcoming MLCommons Training and Inferencing benchmark comparisons to PowerEdge XE8545

Technical Specifications – **XE8545, XE8640**

Features	PowerEdge XE8545	PowerEdge XE8640
CPU	2x 3rd Generation AMD EPYC ™ processors with up to 64 cores per processor Support for up to 2 x 280W processors	Two 4th Generation Intel® Xeon® Scalable processors with up to 56 cores per processor Support for up to 2 x 350W processors
Memory	16x or 32x DDR4 DIMM Speed: Up to 3200 MT/s	Up to 32 x DDR5 RDIMMs DIMM Speed: Up to 4800 MT/s
Storage	Up to 10x 2.5" Hot Plug SAS/SATA or up to 8x NVMe BOSS	Up to 8 E3.S Gen5 NVMe or 8 2.5" SAS/NVMe SSDs Optional Hot-Plug BOSS-N1 (2 x M.2 NVMe) for boot
Storage Controller	S1.5, H745, H755	PERC12, SW RAID
Network	Optional OCP NIC 3.0, Fixed 2 x 1GbE LOM	Optional OCP NIC 3.0, Optional 2 x 1GbE LOM
PCIe slots	1 x PCIe Gen4 LP (x16) 1 x PCIe Gen4 FH (x16) 1 x PCIe Gen4 FH (x16) or 2 x PCIe Gen4 FH (x8)	4 x PCIe Gen5 FH (x16)
GPU	4 x A100 GPUs with NVLink. Options: 40GB/400W or 80GB/500W	4 x NVIDIA H100 SXM5 700W 80GB NVLink GPUs
Integrated Ports	Front: 1 x USB 2.0, 1 x iDRAC Direct micro-USB + front VGA Rear: 1 x USB 3.0 + 1 x USB 2.0, Dedicated iDRAC port, VGA	Front: 1 x USB 2.0, 1 x iDRAC Direct micro-USB + front VGA Rear: 1 x USB 3.0 + 1 x USB 2.0, Dedicated iDRAC port
System Management	iDRAC9 Enterprise, Datacenter license options; OpenManage Enterprise and Plugins (Power Manager, SupportAssist, and Update Manager) iDRAC Direct, Quick Sync 2.0	16G iDRAC Express(base), Enterprise, Datacenter Full Support; Open Manage Enterprise and Plugins (Power Manager, SupportAssist, and Update Manager) iDRAC Direct
High Availability	Hot Plug redundant drives, BOSS, PSUs	Hot Plug redundant drives, BOSS-N1, PSUs
Power Supplies	2 + 2 2400W Redundant Hot Plug	4x 2800W Titanium 208-240Vac PSUs with 3+1 Full redundancy or 2+2 FTR Redundancy
Thermals	Air Cooled Up to 35C	Air Cooled (Liquid Assisted Air-Cooled GPUs) up to 35C
Dimensions	H x W x D: 4U x 477mm x 810mm	H x W x D: 4U x 482mm x 866mm
Form Factor	4U Rack Server 810mm (1070mm rack capable)	4U Rack server 866m (1070mm rack capable)