Dell PowerEdge Acceleration-optimized servers for AI & HPC

Quick Reference Guide: Accelerators for PowerEdge servers

Businesses leaders today are faced with more data, decisions, and challenges than ever before, to grow the business and transform their operations for continued success. Faster decision-making, faster insights, faster real-time analysis are vital. Artificial Intelligence is the key, to helping businesses transform their data into insights and actionable results.

To design an infrastructure to deliver the capabilities which can make organizations successful with AI, the new potential of Generative AI and other demanding workloads requires a modern architecture approach where one of the biggest innovations is improved performance by accelerating insights, enabling a secure foundation for AI operations and the AI lifecycle with a trusted AI approach, and the addition of dense acceleration at scale for simplified operations and democratized AI.

Dell helps you put AI to work for you anywhere in any way to fast-track innovation, powering your AI workloads with accelerated insights, all from the new PowerEdge XE servers, powered by the Intel® Xeon® Scalable processors. Dell PowerEdge helps unleash your AI advantage with a modern compute foundation that continuously accelerates your journey, streamlining your operations securely.

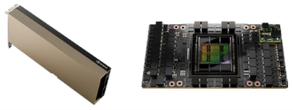
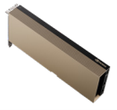
	PowerEdge XE servers are Acceleration-Optimized, purpose-built for complex compute and AI/ML/DL and HPC intensive workloads.				**PowerEdge rack servers are flexible, mainstream computing foundations for a wide range of applications, use cases and workloads.**				
Rack Server	**XE9680**	**XE9640**	**XE8640**	**XE8545**	**R760xa**	**R750xa**	**R940xa**	**R750 /7525 / 7515, R650 / 6525 / 6515**	**XR12**
									
Specifications	**No-compromise accelerated AI**	**Dense acceleration**	**Purpose-built performance**	**Superior outcomes for AI**	**Purpose-built flexibility**	**Purpose-built flexibility**	**Extreme acceleration**	**Mainstream performance**	**Edge performance**
Processor	Two 4th Generation Intel® Xeon® Scalable processors	Two 4th Generation Intel® Xeon® Scalable processors	Two 4th Generation Intel® Xeon® Scalable processors	Two 3rd generation AMD EPYC™ processors	Two 4th Generation Intel® Xeon® Scalable processors	Two 3rd Generation Intel® Xeon® Scalable processors	Up to four 2nd Generation Intel® Xeon® Scalable processors	One or Two 3rd Generation Intel® Xeon® Scalable or 3rd Generation AMD EPYC™ processors	One 3rd Generation Intel® Xeon® Scalable processor
Memory	32 DDR5 DIMMs, 4TB max	32 DDR5 DIMMs, 4TB max	32 DDR5 DIMMs, 4TB max	32 DDR4 DIMMs, 2TB max	32 DDR5 DIMMs, 4TB max	32 DDR4 DIMMs, 4TB max	48 DDR4 DIMMS, 12TB max	Up to 32 DDR4 DIMMs, 4TB max	8 DDR4 DIMMs, 1TB max
GPU support	8x NVIDIA(R) H100 Tensor Core SXM5 or 8x NVIDIA A100 SXM4 NVLInk connectivity	4x Intel Data Center Max 1550 OAM GPU GPU-GPU connectivity	4x NVIDIA H100 Tensor Core SXM5 NVLInk connectivity	4x NVIDIA A100 Tensor Core SXM4 NVLInk connectivity	4 x 350W Double-Wide or 12 x 75W Single-Wide	4 x 300W Double-Wide or 6 x 75W Single-Wide	4 x 300W Double-Wide	Up to 3 x 300W Double-Wide or 6 x 75W Single-Wide	Up to 2 x 300W Double-Wide or Single-Wide
Other features	Air-cooled operation (up to 35C) 6U rack height Up to 8 x 2.5” drives Up to 10 x PCIe Gen5	Liquid-cooled CPU and GPU operation 2U rack height Up to 4 x 2.5” drives 2 x PCIe Gen5	Air-cooled operation (up to 35C) 4U rack height Up to 8 x 2.5” drives Up to 4 x PCIe Gen5	Air-cooled operation (up to 35C) 4U rack height Up to 10 x 2.5” drives Up to 4 x PCIe Gen4	Air-cooled operation (up to 35C) 2U rack height Up to 8 x 2.5” drives Up to 4 x PCIe Gen5	Air-cooled operation (up to 35C) 2U rack height Up to 8 x 2.5” drives Up to 4 x PCIe Gen4	Air-cooled operation (up to 35C) 4U rack height Up to 24 x 2.5” drives Up to 12 x PCIe Gen3	Air-cooled operation (up to 35C) 1U or 2U rack height" Up to 8 x 2.5” drives Up to 8 x PCIe Gen4	*-5°C to 55°C 2U rack height Up to 4 x PCIe Gen4
Applications and use-cases	Large data set language models, Natural Language Processing, AI ML DL Training, HPC, CRISP, Healthcare, CSP/HPCaaS, Finance, Academia, Generative AI/GPT	AI ML DL Training, HPC, Modeling & Simulation, Healthcare, Life Sciences, Finance	Medium data set language Models, Modeling & Simulation, AI, ML/DL Training and Inferencing	AI ML Training and inferencing, small and medium data set langauge models	AI & ML training and inferencing, data analytics, HPC, VDI & Performance graphics	AI & ML training and inferencing, data analytics, HPC, VDI & Performance graphics	GPU database acceleration, data analytics, AI, machine learning	Light duty AI/ML/DL training, inferencing, VDI, Performance graphics, Edge	Edge AI training, Inferencing, Telco, rendering/modeling
Availability	now	1H, 2023	1H, 2023	now	1H, 2023	now	now	now	now

Dell PowerEdge Acceleration-optimized servers for AI & HPC

Quick Reference Guide: Accelerators for PowerEdge servers

Demanding use cases require the optimal compute approach. With the emergence of AI, Machine learning, deep learning, data analytics and visualization, as well as increased workforce access to more business resources, IT can now choose to leverage GPU acceleration. PowerEdge servers are designed and built with accelerators boost graphics operations (for Simulation, Oil & Gas, HPC, life sciences, research and academia, financial and cybersecurity) and collaboration (virtualization, Power users and Knowledge worker VDI, rich media and performance graphics).

- NVIDIA offers a complete portfolio with Hopper, Ada Lovelace and Ampere GPUs from entry-level to mainstream to the highest performance, each providing the versatility to accelerate the widest range of AI applications, whether at the edge, in the cloud, or on-premise.
- Intel GPUs and the Data Center GPU Max Series is designed to take on the most challenging high-performance computing (HPC) and AI workloads.
- AMD Instinct™ family of accelerators can deliver industry leading performance for data center computing, supercharging HPC and AI workloads

	NVIDIA GPU accelerators										Intel GPUs	AMD GPUs
	H100		A100		L40 & A40	L4	A30	A16	A10	A2	Max 1550	MI210
												
Workload	HPC/AI/ML/DL Training		HPC/AI/Database Analytics		Performance graphics/VDI/ Modeling	Mainstream AI inferencing, VDI, virtualization, Edge	Mainstream AI	VDI, Virtualization	Mainstream graphics/VDI	Inferencing/Edge/ VDI	HPC/AI/ML/DL Training	HPC/Machine learning training
Memory	80 GB	80 GB	80 GB	40 / 80 GB	48 GB	24 GB	24 GB	32 GB	24 GB	16 GB	128 GB	64 GB
System interface	PCIe Gen5x16/ NVLink bridge	SXM5, NVLink bridge	PCIe Gen4x16/ NVLink bridge	SXM4, NVLink bridge	PCIe Gen4x16/ NVLink bridge	PCIe Gen4x16	PCIe Gen4x16/ NVLink bridge	PCIe Gen4x16	PCIe Gen4x16	PCIe Gen4x8	OAM, XGMI bridge	PCIe Gen4x16
Slot width	Double-wide	n/a	Double-wide	n/a	Double-wide	Single-wide	Double-wide	Double-wide	Single-wide	Single-wide	n/a	Double-wide
Max power cons.	350W	700W	350W	500W (80GB) 400W (40GB)	300W	72W	165W	250W	150W	60W	600W	300W
PowerEdge support	R760, R750xa, R750, R7525	XE9680 (8xH100), XE8640 (4xH100)	R760, R7625, R7615, R750xa, R750, R7525, XR12, R940xa, R740/XD, DSS8440	XE8545 XE9680 (8xA100, 80GB)	L40: R750xa, R750, R7525, A40: R750xa, R750, R7525, XR12, DSS8440, R740, R740xd, T550	R750, R7525, R650	R760, R7625, R7615, R750xa, R750, R7525, R7515, R740, R740xd, XR12, XE2420, T550	R760, R7625, R7615, R750xa, R750, R7525, R7515, R740, R740xd	R750xa, R750, R7525, R740, R740xd, XE2420	R760, R7625, R7615, R660, R6625, R6615, C6620, R750xa, R750, R7525, R7515, R650, C6520, R6525, R6515, C6525, XR12, XR11, R740, R740xd, R640, T550	XE9640	R7625, R7615, R750xa, R7525, R7515

Increase efficiency and accelerate operations with an autonomous infrastructure

The Dell OpenManage™ systems management portfolio delivers a secure, efficient, and comprehensive solution for PowerEdge servers. Simplify, automate and centralize one-to-many management with the OpenManage Enterprise console and iDRAC.

Enable and accelerate AI workloads

Design, test and bring your vision to life at Customer Solution Centers worldwide

Tap into collaborative expertise in the AI Innovation Lab.

Leverage best practices from our worldwide AI Centers of Excellence.

Leverage suites to simplify deployments

NVIDIA AI Enterprise is an end-to-end, cloud-native suite of AI tools and frameworks optimized to run on VMware vSphere with NVIDIA-Certified Systems. It includes key technologies for the rapid deployment, management and scaling of AI workloads.

NVIDIA-Certified Dell Systems brings together NVIDIA GPUs and NVIDIA networking in servers and hyperconverged infrastructure from Dell Technologies in optimized configurations.

NVIDIA LaunchPad is a free curated lab experience that enables organizations to get immediate, short-term access to the hardware and software stacks for AI, data science, 3D-design collaboration and simulation, and more. NVIDIA Launch Pad is proudly built on Dell PowerEdge servers.

AMD ROCm™ delivers an open-source exascale-class platform for accelerated computing in HPC and cluster deployments.

Rest easier with Dell Technologies Services

Maximize your PowerEdge Servers with comprehensive services ranging from Consulting, to ProDeploy and ProSupport suites, Data Migration and more – available across 170 countries and backed by our 60K+ employees and partners.

Achieve more, deliver quick results and maximize efficiency

Dell Validated Designs are purpose-designed with IT's transformation journey in mind to run intelligent applications and processes in the digital business.

Along with Dell PowerEdge servers, Dell Technologies partners and collaborates with industry leaders including Intel, Microsoft, NVIDIA, and others to optimize IT for your critical business workloads together with emerging technologies such as AI, machine learning, and blockchain.

Validated Designs for AI

- AutoML, including Deep Learning with NVIDIA GPUs and Cloudera
- MLOps
- Conversational AI
- Validated Designs for Data Analytics
- Validated Designs for HPC
- Validated Designs for VDI

Learn more here

Discover more about PowerEdge servers

Learn more about our PowerEdge servers

Learn more about our systems management solutions

Learn more about our Services for PowerEdge

Search our Resource Library

Follow PowerEdge servers on Twitter

Contact a Dell Technologies Expert for Sales or Support

Follow PowerEdge servers on Linkedin

For a comprehensive list, visit dell.com/PowerEdge. Product availability may vary by region. Please contact your Dell representative for more information.

March 2023