Dell Technologies Expands Dell AI Factory with NVIDIA to Turbocharge AI Adoption

May 20, 2024

LAS VEGAS, May 20, 2024 /PRNewswire/ -- **DELL TECHNOLOGIES WORLD** --

News summary

- Dell PowerEdge XE9680L server offers direct liquid cooling and eight NVIDIA Blackwell Tensor Core GPUs for fast processing in a compact form factor
- Dell delivers industry's densest and energy efficient turnkey rack-scale solutions to accelerate large Blackwell GPU deployments
- Dell NativeEdge is the first edge orchestration platform that automates the delivery of NVIDIA AI Enterprise software, including NVIDIA NIM
- Dell solutions and services with NVIDIA technology speed AI application deployment and digital assistant implementation

Full story

Dell Technologies (NYSE: DELL) expands the Dell AI Factory with NVIDIA to include new server, edge, workstation, solutions and services advancements that speed AI adoption and innovation.

"Organizations are moving quickly to capture the AI opportunity, which is why our collaboration with NVIDIA is so important," said Michael Dell, founder and CEO, Dell Technologies. "Our expansion of the Dell AI Factory with NVIDIA continues our joint mission – we're making it easy for organizations to implement AI so they can move boldly into this next technological revolution."

"Generative AI requires a new type of computing infrastructure – an AI factory that produces intelligence," said Jensen Huang, founder and CEO, NVIDIA. "Together, NVIDIA and Dell are providing the world's industries with a full-stack offering – including computing, networking and software – that drives the copilots, coding assistants, virtual customer service agents and industrial digital twins of the digital enterprise."

Dell AI Factory with NVIDIA transforms data into insights and outcomes

The Dell AI Factory with NVIDIA integrates Dell's leading AI portfolio with NVIDIA AI Enterprise software platform, underpinned by NVIDIA Tensor Core GPUs, NVIDIA Spectrum-X Ethernet networking fabric and NVIDIA Bluefield DPUs. Customers can purchase integrated capabilities tailored to their needs or pre-validated, full stack solutions to get them started on AI use cases that require accelerated performance like RAG, model training and inferencing. Advancements to the Dell AI Factory with NVIDIA allow organizations to:

Use advanced compute power to handle large scale AI deployments

- The new **Dell PowerEdge XE9680L** delivers high performance with support for eight NVIDIA Blackwell GPUs in a smaller 4U form factor. The server provides the highest possible rack-scale density for NVIDIA GPUs in an industry standard x86 rack, offering 33% more GPU density per node.[1] The platform offers 20% more PCIe Gen. 5 slots and double the North/South network expansion capacity.[2]

  Direct liquid cooling (DLC) improves overall efficiency with greater cooling capacity for CPUs and GPUs. The PowerEdge XE9680L is designed for easy serviceability and will be available fully configured with advanced factory integration for rack-scale deployments and onsite installation.

  Dell is also announcing industry leading density and highly energy efficient turnkey rack-scale solutions that speed time to value for large GPU-accelerated deployments. Multiple variants will be available including an air-cooled design supporting 64 GPUs in a single rack, or a liquid cooled format featuring 72 NVIDIA Blackwell GPUs in a single rack.

Accelerate edge AI application deployment with Dell NativeEdge and NVIDIA

- **Dell NativeEdge** is the first edge orchestration platform that automates the delivery of NVIDIA AI Enterprise software,[3] helping developers and IT operators easily deploy AI applications and solutions at the edge. Businesses from manufacturers to retailers can quickly and accurately analyze their edge data with new Dell NativeEdge deployment blueprints which include NVIDIA Metropolis video analytics, NVIDIA Riva speech and translation capabilities and NVIDIA

NIM inference microservices.

Simplify AI application development and deployment for faster time to value

- The new **Dell Generative AI Solution for Digital Assistants** helps speed deployment of digital assistants that deliver a personalized self-service experience for end users on a full-stack Dell and NVIDIA solution. **Implementation Services for Digital Assistants** help organizations design, plan, implement, test and scale the solution.

- **Dell AI Factory with NVIDIA solutions** help organizations quickly stand-up AI environments for a variety of use cases with full stack deployment automation engineered in collaboration with NVIDIA. Full stack automation reduces the time to value by up to 86% compared to doing it yourself.[4] When combined with NVIDIA NIM inference microservices, the overall time from delivery to running inferencing jobs is even further reduced. NIM microservices provide enterprise developers production-ready, optimized inference engines for popular AI models available from NVIDIA and its partner ecosystem.

- New **Dell Accelerator Services for RAG on Precision AI Workstations** help shorten the AI development cycle and quickly yield better performing AI applications through a tailored large language model using RAG on a Dell Precision workstation with NVIDIA AI Workbench, a development platform to experiment, test, and prototype AI and ML projects in a more secure environment.

"Our research shows that organizations seek IT simplification, but are challenged to achieve it. AI, while promising simplification, brings its own set of complexities including choice, data quality, consistency and security. Having a proven, one-stop shop to help stand up the infrastructure, software and services needed to realize value from AI can reduce risk and lower costs," said Dave Vellante, chief analyst, theCUBE Research. "Dell's end-to-end capabilities are a core differentiator that we believe will translate to AI. The Dell AI Factory with NVIDIA is a leading example of an AI solution designed to simplify adoption for emerging AI workloads."

Availability

- Dell PowerEdge XE9680L servers will be available in the second half of 2024.
- Dell NativeEdge deployment blueprints for NVIDIA will be available globally starting in the second half of 2024.
- Dell Generative AI Solution for Digital Assistants and Implementation Services for Digital Assistants are available now in North America.
- Dell AI Factory with NVIDIA automated deployment capabilities enabled through Dell professional services will be available in the second half of 2024. NVIDIA NIM microservices are available today.
- Dell Accelerator Services for RAG on Precision AI Workstations are available in late May in the United States.

Additional Resources

- **Blog:** Transform Innovation into Value: The Dell AI Factory with NVIDIA
- **Blog:** Announcing Dell PowerEdge XE9680L Featuring AI Factory Rack-Scale Architecture
- **Blog:** Dell NativeEdge Accelerates AI Innovation at the Edge
- **Blog:** Elevate the Self-Service Experience for Your Customers
- Connect with Dell on X and LinkedIn

About Dell Technologies
Dell Technologies (NYSE: DELL) helps organizations and individuals build their digital future and transform how they work, live and play. The company provides customers with the industry's broadest and most innovative technology and services portfolio for the data era.

Dell Technologies World
Join us May 20-23 for Dell Technologies World, the company's flagship event that brings together the latest trends, technology and thought leaders. During the event, customers and partners will explore technologies that accelerate and simplify AI adoption along with PCs, Dell APEX, multicloud and edge solutions that transform ideas into action.

Copyright © 2024 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies and Dell are trademarks of Dell Inc. or its subsidiaries.

[1] Dell PowerEdge XE8680L is 4U vs. 6U PowerEdge XE9680.

[2] Dell PowerEdge XE9680L (Q4CY24 RTS) has 4 additional PCIe 5.0 FHHL expansion slots for network I/O compared to 2 additional PCIe slots on the XE9680. Both the XE9680 and XE9680L has 8 dedicated PCIe slots for each GPUs AI fabric interface.

[3] Based on Dell analysis, May 2024.

[4] Estimate based on Dell analysis in May 2024 comparing time to setup a 2-node Kubernetes cluster for a general-purpose LLM using automated scripts vs deploying a common design manually. Setup time includes base installation only. Actual setup time will vary depending on solution configuration.

View original content to download multimedia:https://www.prnewswire.com/news-releases/dell-technologies-expands-dell-ai-factory-with-nvidia-to-turbocharge-ai-adoption-302150245.html

SOURCE Dell Technologies

Rosie Izzi, rosie.izzi@dell.com