Dell AI Data Platform Advancements Help Customers Harness Data to Power Enterprise AI with NVIDIA and Elastic

August 11, 2025

- Dell AI Data Platform updates improve support for full AI lifecycle from ingestion to inferencing
- Dell collaborates with vector search leader Elastic to improve inferencing and analytics applications
- The Dell PowerEdge R7725 server with NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs will be the first 2U server platform to deliver the NVIDIA AI Data Platform reference architecture
- Dell and the NVIDIA AI Data Platform reference design offers a validated, GPU-accelerated reference architecture for AI—integrating storage, data, compute, networking and software

ROUND ROCK, Texas--(BUSINESS WIRE)--Aug. 11, 2025-- Dell Technologies (NYSE: DELL), the world's No. 1 provider of AI infrastructure, [1] today announced updates to the Dell AI Data Platform to help customers better support the full lifecycle of AI workloads from ingestion and transformation to agentic inferencing to AI-powered knowledge retrieval.

Why it matters

Enterprise data is massive, growing rapidly and increasingly unstructured, but only a fraction of it is usable for generative AI today. To unlock its value, organizations need continuous indexing and a vector retrieval engine that converts content into embeddings for fast, precise semantic search. As workloads grow, organizations need infrastructure that streamlines data preparation, unifies data access across silos and delivers end-to-end enterprise-grade performance.

The latest updates to the Dell AI Data Platform enhance unstructured data ingestion, transformation, retrieval and compute performance to streamline AI development and deployment – turning massive datasets into reliable, high quality real-time intelligence for generative AI.

Accelerating AI inferencing and analytics

The Dell AI Data Platform helps customers quickly move from AI experimentation to production by automating data preparation.

At the core of the Dell AI Data Platform's architecture are specialized storage and data engines that help seamlessly connect AI agents to high quality enterprise data. Together, the Dell AI Data Platform and the NVIDIA AI Data Platform reference design provide a validated, GPU-accelerated solution that integrates storage engines and data engines with NVIDIA accelerated computing, networking and AI software to power generative AI systems.

Expanding the capabilities of the Dell AI Data Platform is the new **unstructured data engine,** designed to provide real-time, secure access to large-scale unstructured datasets for inferencing, analytics, and intelligent search. This engine, made possible through a new collaboration with open-source Search AI leader Elastic, will offer customers advanced vector search, semantic retrieval and hybrid keyword search capabilities—key capabilities for powering AI applications. Additionally, the unstructured data engine will leverage built-in GPU acceleration to deliver breakthrough performance.

The unstructured data engine works alongside the platform's other tools, like a federated SQL engine for querying scattered structured data, a processing engine for handling large-scale data transformation, and storage designed for fast, AI-ready access.

Powering enterprise AI discovery

As AI becomes increasingly crucial for business-as-usual operations, Dell PowerEdge R7725 and R770 servers featuring NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs provide the mainstream computing foundation for accelerated enterprise workloads, from visual computing, data analytics and virtual workstations, to physical AI and agentic inference. These servers are ideal for running NVIDIA AI reasoning models such as the latest NVIDIA Nemotron models for agentic AI, as well as NVIDIA Cosmos world foundation models for physical AI.

Offering better price for performance for a wide range of enterprise use cases, these air-cooled systems make flexible high-density AI compute more attainable. The NVIDIA RTX PRO 6000 offers enterprises up to six times the token throughput for LLM inference,[2] double the capacity for engineering simulation performance[3] and can support four times the number of concurrent users compared to the previous generation with support for MIG.

The Dell PowerEdge R7725 server will also be the first 2U server platform to integrate the NVIDIA AI Data Platform reference design. When the Dell PowerEdge R7725 server featuring NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs is paired with the Dell AI Data Platform and its new unstructured data engine, enterprises can take advantage of a turnkey solution without the need to architect and test their own hardware and software platforms. The combination of the two delivers faster inferencing, more responsive semantic search and support for larger, more complex AI workloads.

See innovation in action at SIGGRAPH 2025

Dell Technologies is showcasing how customers can accelerate media production pipelines and power intelligent asset management at scale using the Dell AI Data Platform, NVIDIA Omniverse software and Dell infrastructure at this year's SIGGRAPH conference (August 10-14) in Vancouver, Canada. Dell will also feature the new Dell Pro Max high-performance PC portfolio, including laptops, desktops and the upcoming Dell Pro Max with GB10, a compact AI developer workstation.

Perspectives

“The key to unlocking AI’s full potential lies in breaking down silos and simplifying access to enterprise data,” said Arthur Lewis, president, Infrastructure Solutions Group, Dell Technologies. “Collaborating with industry leaders like NVIDIA and Elastic to advance the Dell AI Data Platform will help organizations accelerate innovation and scale AI with confidence.”

“Enterprises worldwide need infrastructure that handles the growing scale and complexity of AI workloads,” said Justin Boitano, vice president of enterprise AI at NVIDIA. “With NVIDIA RTX PRO 6000 GPUs in new 2U Dell PowerEdge servers, organizations now have a power efficient, accelerated computing platform to power AI applications and storage on NVIDIA Blackwell.”

“Fast, accurate, and context-aware access to unstructured data is key to scaling enterprise AI,” said Ken Exner, Chief Product Officer at Elastic. “With Elasticsearch vector database at the heart of the Dell AI Data Platform's unstructured data engine, Elastic will bring vector search and hybrid retrieval to a turnkey architecture, enabling natural language search, real-time inferencing, and intelligent asset discovery across massive datasets. Dell’s deep presence in the enterprise makes them a natural partner as we work to help customers deploy AI that’s performant, precise, and production-ready.”

Availability

- Unstructured data engine in Dell AI Data Platform will be available later this year.
- Dell PowerEdge R7725 and R770 servers with NVIDIA RTX PRO 6000 GPUs will be globally available later this year.

Additional Resources

- Blog: Dell AI Data Platform Gets a Boost with NVIDIA and Elastic
- 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 AI era.

---

[1] IDC Semiannual Artificial Intelligence Infrastructure Tracker, 2024H1 (Feb 2025)

[2] Llama3 70B Inference; 8K/256, 20 tokens/s/user ; RTX PRO 6000 Blackwell Server Edition vs NVIDIA L40S

[3] Ansys Fluent (2025.2), FP32; RTX PRO 6000 Blackwell Server Edition vs NVIDIA L40S

View source version on businesswire.com: https://www.businesswire.com/news/home/20250811003387/en/

Dell Technologies Media Relations: Media.Relations@Dell.com

Source: Dell Technologies