NVIDIA and Dell Technologies Expand AI Portfolio

Full NVIDIA AI stack now available on 15 new Dell PowerEdge servers, helping enterprises to build and deploy AI workloads faster and more efficiently.

Author: Manuvir Das

In their largest-ever joint AI initiative, NVIDIA and Dell Technologies today launched a wave of Dell PowerEdge systems available with NVIDIA acceleration, enabling enterprises to efficiently transform their businesses with AI.

A total of 15 next-generation Dell PowerEdge systems can draw from NVIDIA's full AI stack — including GPUs, DPUs and the NVIDIA AI Enterprise software suite — providing enterprises the foundation required for a wide range of AI applications, including speech recognition, cybersecurity, recommendation systems and a growing number of groundbreaking language-based services.

The news was released at Dell's PowerEdge .Next event, where NVIDIA founder and CEO Jensen Huang joined Dell Technologies founder and CEO Michael Dell in a fireside chat.

Commenting on how they've celebrated a 25-year history of collaboration, the two CEOs looked at solving enterprise challenges through the lens of AI.

"As the amount of data in the world expands, the majority of information technology capacity is going to be in service of machine intelligence," said Dell. "Building systems for AI first is a huge opportunity for Dell and NVIDIA to collaborate."

"AI has the power to transform every business by accelerating automation across every industry," said Huang. "Working closely with Dell Technologies, we're able to reach organizations around the globe with a powerful, energy-efficient AI computing platform that will boost the IQ of modern enterprise."

A key highlight among Dell's portfolio is Dell PowerEdge systems featuring NVIDIA BlueField-2 DPUs .

BlueField data processing units can offload, accelerate and isolate the networking and operating system stacks of the data center, which means businesses using NVIDIA DPUs could cut data center energy use by close to 25%, potentially saving them millions of dollars in energy bills. Dell PowerEdge servers with NVIDIA BlueField DPUs optimize performance and efficiency for private, hybrid and multi-cloud deployments, including those running VMware vSphere.

Additionally, systems featuring NVIDIA H100 GPUs have shown they are able to process data 25x more efficiently to deploy diverse AI models into production, and that NVIDIA-accelerated Dell PowerEdge servers are up to 300x more energy efficient for running inference on large language models — those exceeding 500 billion parameters — when compared to prior-generation non-accelerated servers.

To help customers get their AI projects up and running fast, Dell PowerEdge servers accelerated with NVIDIA H100 GPUs come with a license for NVIDIA AI Enterprise software.

An end-to-end, secure, cloud-native suite of AI software, NVIDIA AI Enterprise streamlines the development and deployment of predictive AI and includes global enterprise support for a wide range of domain- and industry-specific workloads. NVIDIA AI Enterprise includes more than 50 frameworks and pretrained models as well as a set of AI workflows, all which can help organizations speed time to deployment while reducing costs of production-ready AI.

NVIDIA AI frameworks included in NVIDIA AI Enterprise 3.0 are NVIDIA Clara Parabricks for genomics, MONAI for medical imaging, NVIDIA Morpheus for cybersecurity, NVIDIA Metropolis for intelligent video analytics, NVIDIA DeepStream for vision AI, NVIDIA Merlin for recommender systems, and many others. Additionally, it includes new AI workflows for building contact center intelligent virtual

assistants, multi-language audio transcriptions and digital fingerprinting for cybersecurity threat detection.

Enterprises can immediately experience NVIDIA AI Enterprise in dozens of hands-on labs at no charge on NVIDIA LaunchPad with new AI workflow labs expected to debut next week.

Original URL: https://blogs.nvidia.com/blog/2023/01/17/dell-ai-portfolio/