Deploy high-performance Generative AI solutions for the enterprise

Unlock the power of large language models for accelerated enterprise AI outcomes with PowerEdge, PowerScale, NVIDIA AI Enterprise with NVIDIA NeMo framework.

Accelerate business outcomes and increase productivity.

Move quickly with GenAI: Innovation, performance and comprehensive solutions enable faster decision making and insights from your proprietary data and operations.

Streamline deployments: Scale purpose-built, custom models, reduce enablement complexity and democratize business-wide with consistent architectures.

Drive trusted AI: Deploy secure, high-quality on-premise foundation, with data privacy, boosting trusted decision-making from your data.

Use cases for Generative AI are expanding as the potential emerges to automate more repetitive tasks for employees, driving up productivity:

- Digital Assistants including retail, business operations, sales, Legal, HR and hiring, call center.
- Developer: Code generation and code efficiency, UI/UX design.
- Creative, sales and marketing scripting, creative content generation.
- Healthcare, research sciences, medical.

With Project Helix, in collaboration with NVIDIA, enterprises small and large across every industry and geography can harness the potential of Generative AI on their proprietary data business-wide, to accelerate productivity and enhance their business transformation strategy.

This jointly engineered solution of leading, innovative Dell infrastructure and software, combined with NVIDIA acceleration innovation, NVIDIA AI Enterprise software which includes NeMo, a framework of resources designed specifically to streamline Generative AI workstreams, gives AI data scientists, developers and decision-makers the solution they need to deliver critical guidance, recommendations and insights from billions of data points.

Streamline AI adoption for business transformation

Businesses have already kicked off and scaled many of their traditional AI initiatives into production, though many still encounter roadblocks to success to getting there, including deriving real value and ensuring data quality for accurate outcomes, the perceived complexity of AI from project planning to production and lifecycle, and technology deployment complexity. With new AI approaches like Generative AI (GenAI), the potential opportunities to extract and produce value from existing business data, enabling faster decision making and increasing productivity are tremendous.

While generative AI has the power to revolutionize industries, businesses are looking to ensure their approaches, data strategies, privacy, security and infrastructure are appropriately aligned. The complexity of harnessing AI results from these systems traditionally has been a barrier to deploying AI in production, with IT teams spending more time on deployment and lifecycle management than actual model development and innovation.

Project Helix delivers a series of full-stack solutions with technical expertise and pre-built tools based on Dell and NVIDIA infrastructure and software, including a complete blueprint to help enterprises use their proprietary data and more easily deploy generative AI on-premises responsibly and accurately. Solutions based on Project Helix helps companies to move quickly and deploy customized Generative AI applications that drive trusted decisions from their own data to grow and scale their businesses.

nVIDIA

Boost productivity business-wide

Project Helix will support the complete generative AI lifecycle from infrastructure provisioning, modeling, training, fine-tuning, application development and deployment to deploying inference and streamlining results. The validated designs based on high quality, AI-accelerated foundation solutions help enterprises quickly build on-premises generative AI infrastructure for scale. Businesses can now deliver the applications, models and operations that drives more insights and faster decisions, more cost-effectively and boosting productivity.

Technical Specifications

Configurations are based on the newest, AI-acceleration-optimized Dell PowerEdge XE and rack servers, leveraging the latest NVIDIA GPUs and the NVIDIA AI Enterprise with the NeMo framework to build, customize, and deploy generative AI models with billions of parameters. Fast, ample data lake storage for Generative AI and large language models is provided by Dell PowerScale all-flash or hybrid storage arrays.

Compute	Accelerators	Networking	Software	Storage
Training: PowerEdge XE9680, XE8640, R760xa Inferencing: XE9680, XE8640, R760xa Management and other: R660	Training: NVIDIA H100 Tensor Core GPUs with NVLink and NVSwitch. Inferencing: NVIDIA H100, L40, L4	NVIDIA Networking, Dell PowerSwitch Z9432F	Dell OpenManage Enterprise, Power Manager, CloudIQ. NVIDIA AI Enterprise, Base Command and NeMo framework for LLMs	Dell PowerScale F900, F600; Dell ECS Dell ObjectScale

As you seek to unlock the value of your proprietary business data into intelligent and smarter business outcomes, Project Helix delivers a strategic approach to adopt Generative AI validated foundations with PowerEdge, PowerScale and NVIDIA innovation which helps you simplify and speed LLM, GPT and NLP AI projects to production-AI results.

Dell Technologies and NVIDIA

Dell Technologies and NVIDIA work together **to enable and accelerate Generative AI workloads**, deliver engineering-validated hardware and software to accelerate AI, ML and DL workloads to meet customer needs across all businesses and verticals. With Dell Technologies and NVIDIA, you can deploy AI solutions to accelerate your digital transformation through real-time data that improves key decision-making, with solutions optimized for fastest time to value from your AI initiatives.