NVIDIA AI Enterprise with NetApp

∏ NetApp

⬢ NVIDIA.

Mainstream AI with proven best practices for a wide variety of enterprise workloads

The challenge

As the race to win with AI intensifies, the need for an easy-to-deploy, easy-to-scale, and easy-to-manage solution with simplified data management becomes increasingly urgent. Many organizations struggle with finding the strategy and platform that enables AI success. Unlike traditional enterprise applications, AI apps are a relatively recent development for many IT departments. They're rapidly evolving, open source, and leading edge, but they lack proven approaches that meet the rigors of scaled production settings in enterprises.

The solution

NetApp and NVIDIA are working together to enable customers to run their businesses using familiar infrastructure on many footprints—from virtualized containers with Kubernetes and private cloud deployments in corporate data centers to cloud-hosted AI development with services deployed on public clouds—all the way to the edge.

This platform delivers best-in-class data storage and data management with NetApp® all-flash cloud-integrated arrays, and AI software with the NVIDIA AI Enterprise suite. This suite consists of end-to-end, cloud-native AI and data analytics software that's optimized and certified for the industry's leading virtualization platform, VMware vSphere and VMware Cloud Foundation.

Ease of deployment and scaling

The rapid pace of AI innovation makes designing an effective AI infrastructure challenging. With NetApp, you can eliminate guesswork and get started faster by using a field-proven reference architecture.

Run all your workloads on a single adaptable infrastructure platform, so you can constantly optimize compute and storage usage across your hybrid cloud. Reduce TCO and simplify operations with flexible consumption. And with NetApp, your apps and data are always in the right place at the right time, whether on premises or in the cloud. Leverage the NetApp validated solution to deploy and manage AI workloads with optimized software.

Future-proof your investment

This solution is proven with design and deployment best practices for a wide variety of enterprise workloads. It lets you confidently scale and maintain your always-on apps with built-in, end-to-end security across control and data planes. And you can meet the needs of AI/ML workloads while continuing to serve your existing applications.

Boost performance

Investing in state-of-the-art compute demands state-of-the-art storage that can handle thousands of training images per second. You need a high-performance data services solution that keeps up with your most demanding deep learning (DL) training workloads. With NetApp all-flash storage,

Key benefits

Ease of deployment and scaling

- Get going faster by eliminating design complexity and guesswork with a jointly certified solution.

- Deploy and manage AI workloads in containers or VMs with optimized software.

- Simplify data science workspace orchestration with NetApp DataOps Toolkit.

Future-proof your investment

- Get a proven solution with design and deployment best practices for a wide variety of enterprise workloads.

- Meet the needs of artificial intelligence and machine learning (AI/ML) workloads while continuing to serve your existing applications.

- Take advantage of industry-leading NetApp data management and data protection capabilities.

Boost performance

- Accelerate and maximize performance for ML modeling and inferencing workloads.

- Constantly optimize compute and storage usage across your hybrid cloud.

- Leverage NVIDIA virtual GPU (vGPU) technology.

you can expect to get more than 2GBps of sustained throughput (5GBps peak). Further, there's well under 1 millisecond of latency, and the GPUs operate at over 95% utilization. A single NetApp AFF A800 system supports throughput of 25GBps for sequential reads and 1 million IOPS for small random reads, at latencies of less than 500 microseconds for NAS workloads.

The flexibility and scalability of the solution enable it to support and adapt to evolving workloads, making it a strong foundation to meet your future storage requirements. Modular storage building blocks give you a granular approach to growth and avoid AI silos. These building blocks also simplify management by using virtualization to fold AI deployments into existing enterprise infrastructure.
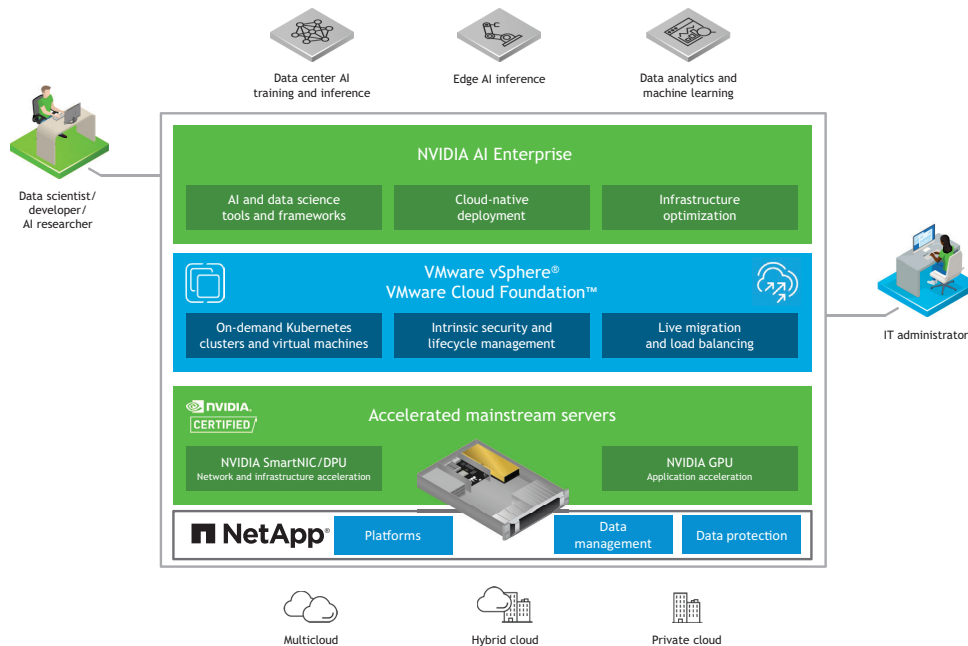
Figure 1: Simplified AI/ML and data analytics deployments.

Solution components
- NVIDIA-Certified Systems
- NVIDIA AI Enterprise Software suite
- VMware vSphere
- VMware Cloud Foundation
- NetApp DataOps Toolkit
- NetApp AFF A-Series systems

About NVIDIA
The invention of the GPU in 1999 by NVIDIA sparked the growth of the PC gaming market, redefined modern computer graphics, and revolutionized parallel computing. More recently, GPU deep learning ignited modern AI—the next era of computing—with the GPU acting as the brain of computers, robots, and self-driving cars that can perceive and understand the world.
More information at www.nvidia.com.

About NetApp
In a world full of generalists, NetApp is a specialist. We're focused on one thing, helping your business get the most out of your data. NetApp brings the enterprise-grade data services you rely on into the cloud, and the simple flexibility of cloud into the data center. Our industry-leading solutions work across diverse customer environments and the world's biggest public clouds.

As a cloud-led, data-centric software company, only NetApp can help build your unique data fabric, simplify and connect your cloud, and securely deliver the right data, services and applications to the right people—anytime, anywhere. www.netapp.com

🖪 NetApp

+1 877 263 8277