

NVIDIA DGX SUPERPOD WITH NETAPP ONTAP STORAGE



Accelerating AI innovation at scale



As organizations embrace artificial intelligence (AI) and machine learning (ML) initiatives, the demand for robust, scalable, and efficient infrastructure solutions has never been greater. The AI Center of Excellence (AI CoE), or AI Factory, has emerged as a crucial component in enterprise AI strategy, serving as a centralized hub for developing and deploying AI solutions. At the heart of these initiatives lies the challenge of managing and training increasingly complex AI models while ensuring data security, accessibility, and resource optimization.

The evolution of agentic AI—AI that can make decisions, plan, and adapt to accomplish specific goals without much human input—has created unprecedented demands on computational and storage infrastructure. Organizations must now handle massive datasets, support multiple concurrent training workloads, and maintain high-performance computing environments while protecting data and complying with regulations.

Traditional infrastructure solutions can struggle to meet these demands, leading to operational inefficiencies and delayed time to value for AI projects.

The solution

NVIDIA DGX SuperPOD™ with NetApp® ONTAP® Storage combines the computing performance of NVIDIA DGX™ systems with the world-class enterprise AI functionality of cloud-connected NetApp AFF A90 storage systems to enable data-driven workflows for ML, AI, and high-performance technical computing (HPC). This integrated solution provides organizations with a scalable, secure, and efficient platform for AI development and deployment. The architecture can eliminate traditional infrastructure silos, enabling seamless collaboration among data scientists, engineers, and other stakeholders.

Data management and access

NVIDIA DGX SuperPOD with NetApp ONTAP Storage implements advanced data reduction technology that optimizes storage utilization while maintaining performance. Through the NetApp Snapshot™ and FlexClone® capabilities of ONTAP, teams can instantly create space-efficient copies of datasets for parallel development and testing. Multiple AI teams can work simultaneously on different versions of the same dataset without consuming additional storage space or compromising data integrity.

Furthermore, with the solution's unified data access framework, teams can easily integrate existing workflows and tools. Data scientists and engineers can leverage familiar protocols and interfaces while benefiting from enterprise-grade data protection and management capabilities. The platform's intelligent data tiering automatically optimizes data placement across storage tiers, so frequently accessed datasets remain readily available while less frequently used data moves to more cost-effective storage. This intelligent approach to data management, combined with the global namespace capabilities of ONTAP, provides AI teams with a truly unified and efficient data access experience.

Scalability

NVIDIA DGX SuperPOD with NetApp ONTAP Storage features intelligent workload orchestration that automatically balances resources, optimizing performance for concurrent AI training jobs and inference workloads. ONTAP storage efficiency technologies like deduplication and compression maximize storage capacity as data grows.

By breaking down infrastructure silos and optimizing resource management within AI Centers of Excellence, organizations can eliminate redundant systems and reduce operational overhead. This unified infrastructure accelerates AI project timelines and improves the return on technology investments.

KEY BENEFITS

Unified Data Management

A centralized platform that provides comprehensive control over your AI data lifecycle.

Accelerated Data Access

Rapid, secure access to data across your organization.

Seamless Infrastructure scaling

A unified platform that grows with your needs, from proof of concept to enterprise-wide deployment.

Comprehensive data protection

Enterprise-grade data protection that safeguards sensitive information throughout the entire AI lifecycle

Zero-Trust Security

Security with a zero-trust architecture that protects your AI infrastructure at every layer

Organizations can seamlessly scale by adding DGX compute nodes and ONTAP storage systems without disruption. The architecture supports both scale-up and scale-out approaches for computational power, memory, and storage. The unified fabric maintains high performance and low latency at any scale, optimizing efficiency for AI workloads.

Security

Through NetApp's robust security measures such as role-based access control (RBAC), multifactor authentication, and audit logging, organizations can confidently pursue AI initiatives while protecting intellectual property and maintaining compliance. With this protection, only authorized personnel and AI agents can access sensitive models, training data, and enterprise data.



The solution protects data through comprehensive encryption both at rest and in transit, while providing integrated security monitoring for real-time threat detection and automated response. Built-in Snapshot and backup capabilities enable quick recovery from security incidents, minimizing both downtime and data loss.

This strong security foundation allows businesses to innovate faster, collaborate more effectively with partners, and scale their AI operations while maintaining excellent data protection.

Conclusion

NVIDIA DGX SuperPOD with NetApp ONTAP Storage solution represents a significant advancement in AI infrastructure solutions. By addressing key challenges around security, data management, resource use, and scalability, it enables organizations to accelerate their AI initiatives while maintaining operational efficiency, data protection, and collaboration. The solution's integrated approach eliminates common bottlenecks in AI development pipelines, enabling data scientists and engineers to focus on innovation rather than infrastructure management. As AI continues to evolve and drive digital transformation across industries, the need for strong, scalable, and secure infrastructure solutions will only grow. NVIDIA DGX SuperPOD with NetApp ONTAP Storage provides organizations with a foundation for current and future AI initiatives, enabling them to maintain competitive advantage in an increasingly AI-driven world.



Contact Us

About NetApp

NetApp is the intelligent data infrastructure company, combining unified data storage, integrated data services, and CloudOps solutions to turn a world of disruption into opportunity for every customer. NetApp creates silo-free infrastructure, harnessing observability and AI to enable the industry's best data management. As the only enterprise-grade storage service natively embedded in the world's biggest clouds, our data storage delivers seamless flexibility. In addition, our data services create a data advantage through superior cyber resilience, governance, and application agility. Our CloudOps solutions provide continuous optimization of performance and efficiency through observability and AI. No matter the data type, workload, or environment, with NetApp you can transform your data infrastructure to realize your business possibilities. www.netapp.com

