



NEWS RELEASE

NetApp Teams with NVIDIA to Redefine Enterprise RAG and Power Agentic AI

2024-09-24

SAN JOSE, Calif.–September 24, 2024–NetApp® (NASDAQ: NTAP), the intelligent data infrastructure company, today unveiled an advanced generative AI data vision and end-to-end integrated solutions that combine NVIDIA AI software and accelerated computing with NetApp intelligent data infrastructure for enterprise retrieval augmented generation (RAG) to power the future of agentic AI applications.

This will bring new capabilities to the NetApp ONTAP unified storage operating system that can leverage a new **NetApp global metadata namespace** to unify data stores for the tens of thousands of enterprises that trust NetApp for their data infrastructure. It opens up exabytes of enterprise data stored across clouds and on-premises infrastructure to drive RAG capabilities that can put enterprises' entire data estate to work, accelerating next-generation agentic AI applications.

The solution brings together proven NetApp AIPod architecture with NetApp ONTAP and the NetApp BlueXP unified control plane, with **NVIDIA NeMo Retriever** and **NIM microservices**, which are part of the **NVIDIA AI Enterprise** software platform.

"To power AI applications and drive transformative progress for their business, enterprises must unlock the potential of their data," said Harv Bhela, Chief Product Officer at NetApp. "Combining the NetApp data management engine and NVIDIA AI software empowers AI applications to securely access and leverage vast amounts of data, paving the way for intelligent, agentic AI that tackles complex business challenges and fuels innovation."

"Data is fundamental to the evolution of generative AI," said Manuvir Das, vice president, Enterprise Computing at NVIDIA. "By combining NVIDIA AI software and accelerated computing with NetApp intelligent data infrastructure, enterprises can turn their data into knowledge, and AI agents can turn that knowledge into action."

With the new NetApp AI capabilities built into NetApp AIPod – certified for **NVIDIA DGX BasePOD** infrastructure and **NVIDIA OVX** solutions— and managed through BlueXP, NetApp customers will be able to easily discover, search, and curate data on-prem and in the public cloud based on a set of criteria, honoring existing policy-based governance criteria.

Once the data collection has been established through NetApp BlueXP, it can be dynamically connected to

NVIDIA NeMo Retriever, where the dataset will be processed and vectorized to be accessible for enterprise GenAI deployments with appropriate access controls and privacy guardrails. This creates the foundation for a generative AI flywheel to power next-generation agentic AI applications that can autonomously and securely tap into data to complete a broad range of tasks to support customer service, business operations, financial services and more.

The end-to-end integration unlocks enterprise data for AI and takes a responsible approach by preserving the security and policy guardrails throughout the AI data and model lifecycle. This integration was first referenced as a proof-of-concept shown by Huang in his NVIDIA GTC 2024 keynote address. This secure and compliant GenAI integration will be available for customers to experience at NetApp INSIGHT today and is targeted to be released as a technology preview to customers later this calendar year.

NetApp has also begun the NVIDIA certification process of NetApp ONTAP storage on the AFF A90 platform with **NVIDIA DGX SuperPOD**, which will enable organizations to leverage industry-leading data management capabilities for their largest AI projects. This certification will complement and build upon NetApp ONTAP's existing certification with NVIDIA DGX BasePOD. NetApp ONTAP addresses data management challenges for large language models, eliminating the need to compromise data management for AI training workloads.

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. Learn more at www.netapp.com or follow us on [X](#), [LinkedIn](#), [Facebook](#), and [Instagram](#).

NETAPP, the NETAPP logo, and the marks listed at www.netapp.com/TM are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.