NetApp

NetApp Powers the Future of AI with Intelligent Data Infrastructure

2024-09-24

SAN JOSE, Calif. – September 24, 2024 – NetApp® (NASDAQ: NTAP), the intelligent data infrastructure company, today announced new developments in its collaboration with industry leaders to accelerate AI innovation. By providing the intelligent data infrastructure required to make GenAI work, NetApp is helping organizations tap into one of the most important developments for business and IT in the last decade.

GenAI powers practical and highly visible use cases for business innovation such as generating content, summarizing large amounts of information, and responding to questions. **Gartner research** predicts that spending on AI software will grow to $297.9 billion by 2027 and that GenAI will account for over one-third of that. The key to success in the AI era is mastery over governable, trusted, and traceable data.

Yesterday, NetApp CEO George Kurian kicked off **NetApp INSIGHT 2024** with an expansive vision of this era of data intelligence. A large part of the AI challenge is a data challenge, and Kurian laid out a vision for how intelligent data infrastructure can ensure the relevant data is secure, governed, and always updated to feed a unified, integrated GenAI stack.

Today at NetApp INSIGHT, NetApp will be unveiling further innovations in intelligent data infrastructure, including a transformative vision for AI running on NetApp ONTAP®, the leading operating system for unified storage. Specifically, NetApp's vision includes:

- **NVIDIA DGX SuperPOD Storage Certification for NetApp ONTAP:** NetApp has begun the NVIDIA certification process of NetApp ONTAP storage on the AFF A90 platform with **NVIDIA DGX SuperPOD** AI infrastructure, 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 (LLMs), eliminating the need to compromise data management for AI training workloads.
- **Creation of a global metadata namespace** to explore and manage data in a secure and compliant fashion across a customers' hybrid multi-cloud estate to enable feature extraction and data classification for AI. **NetApp separately announced** today a new integration with NVIDIA AI software that can leverage the global metadata namespace with ONTAP to power enterprise retrieval augmented generation (RAG) for agentic AI.
- **Directly integrated AI data pipeline**, allowing ONTAP to make unstructured data ready for AI automatically and iteratively by capturing incremental changes to the customer data set, performing policy driven data

classification and anonymization, generating highly compressible vector embeddings and storing them in a vector DB integrated with the ONTAP data model, ready for high scale, low latency semantic searches and retrieval augmented generation (RAG) inferencing.

- **A disaggregated storage architecture** that enables full sharing of the storage backend, which maximizes utilization of network and flash speeds and lowers infrastructure cost, significantly improving performance while economizing rack space and power for very high-scale, compute-intensive AI workloads like LLM training. This architecture will be an integral part of NetApp ONTAP, so it will get the benefit of a disaggregated storage architecture but still maintain ONTAP's proven resiliency, data management, security and governance features.
- **New capabilities for native cloud services to drive AI innovation** in the cloud. Across all its native cloud services, NetApp is working to provide an integrated and centralized data platform to ingest, discover and catalog data. NetApp is also integrating its cloud services with data warehouses and developing data processing services to visualize, prepare and transform data. The prepared datasets can then be securely shared and used with the cloud providers' AI and machine learning services, including third party solutions. NetApp will also announce a planned integration that allows customers to use Google Cloud NetApp Volumes as a data store for BigQuery and Vertex AI.

"Organizations of all sizes are experimenting with GenAI to increase efficiency and accelerate innovation," said Krish Vitaldevara, Senior Vice President, Platform at NetApp. "NetApp empowers organizations to harness the full potential of GenAI to drive innovation and create value across diverse industry applications. By providing secure, scalable, and high-performance intelligent data infrastructure that integrates with other industry-leading platforms, NetApp helps customers overcome barriers to implementing GenAI. Using these solutions, businesses will be able to more quickly and efficiently apply their data to GenAI applications and outmaneuver competitors."

NetApp continues to innovate with the AI ecosystem:

- **Domino Data Labs chooses Amazon FSx for NetApp ONTAP**: To advance the state of machine learning operations (MLOps), NetApp has partnered with Domino Data Labs, underscoring the importance of seamless integration in AI workflows. Effective today, Domino is using Amazon FSx for NetApp ONTAP as the underlying storage for Domino Datasets running in Domino Cloud platform to provide cost-effective performance, scalability, and the ability to accelerate model development. In addition to Domino using FSx for NetApp ONTAP, Domino and NetApp have also begun joint development to integrate Domino's MLOps platform directly into NetApp ONTAP to make it easier to manage the data for AI workloads.
- **General Availability of AIPod with Lenovo for NVIDIA OVX**: Announced in May 2024, the NetApp AIPod with Lenovo ThinkSystem servers for NVIDIA OVX converged infrastructure solution is now generally available. This infrastructure solution is designed for enterprises aiming to harness generative AI and RAG capabilities to boost productivity, streamline operations, and unlock new revenue opportunities.
- **New capabilities for FlexPod AI**: NetApp is releasing new features for its FlexPod AI solution, the hybrid infrastructure and operation platform that accelerate the delivery of modern workloads. FlexPod AI running RAG simplifies, automates, and secures AI applications, enabling organizations to leverage the full potential of their data. With Cisco compute, Cisco network, and NetApp storage, customers experience lower costs, efficient scaling, faster time to value, and reduced risks.

"Implementing AI requires a collection of finely tuned pieces of technology infrastructure to work together perfectly," said Mike Leone, Practice Director, Data Analytics & AI, Enterprise Strategy Group, part of TechTarget. "NetApp delivers robust storage and data management capabilities to help customers run and support their AI

data pipelines. But storage is one piece of the puzzle. By collaborating with other industry-leading vendors in the AI infrastructure space, NetApp customers can be confident that their compute, networking, storage, and AI software solutions will integrate seamlessly to drive AI innovation."

Additional Resources

- **Elevating Enterprise AI: NetApp's Latest Ecosystem Advancements for Superior Innovation**
- **Next Generation Data Management for AI**
- **AI Data Solutions**
- **AIPod: Delivering the Promise**
- **FlexPod**

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.