



Solution Brief

NetApp AI Solutions for Federal Government

Accelerate your journey to AI

Key Benefits

Accelerate Your Data Pipeline

- Get up and running faster with the NetApp® ONTAP® AI validated reference architecture from NVIDIA and NetApp.
- Increase training throughput—up to four times faster than competing solutions.

Streamline DataOps

- Automate the deployment of Kubernetes on NVIDIA DGX supercomputers with Ansible modules for NetApp.
- Enable persistent storage for Kubernetes with Trident.

Simplify Version Control for Machine Learning

- Monitor and track every version of your experiments using fast, space-efficient NetApp Snapshot™ copies.

Tap into the Cloud

- Take advantage of innovative AI software in the world's largest clouds, while retaining NetApp data management and NFS capabilities.

Opportunity

Artificial intelligence (AI) has tremendous potential to improve the lives of citizens and help government organizations deliver on their missions—reducing pollution, saving and generating energy, improving safety and privacy, improving public health, enhancing agricultural outcomes, enabling smart transportation, and optimizing infrastructure.

But AI isn't just a technology of the future. Most government organizations are using AI in some way today, starting with foundational applications such as voice assistants, chatbots, and big data analytics. For example, [the U.S. Postal Service uses handwriting recognition technologies](#) to sort mail by ZIP code. Some machines can process over 18,000 pieces of mail an hour. [In Pittsburgh, AI-enabled traffic lights](#) have helped cut travel times by 25% and idling times by 40%. In the military, AI is used to guide drones and robot soldiers, enable intelligent systems, and improve cybersecurity.

And with new focus on AI coming from the highest levels of the current administration, organizations are reporting acceleration in the adoption and application of AI. In a recent [MeriTalk survey](#) of federal, state, local, and higher education IT managers, 73% reported plans to increase spending on AI. Yet, at the same time, most of these organizations also reported feeling unprepared for AI.

Government IT leaders say that data silos and technology complexity are the two biggest challenges to moving AI projects into production. Government organizations need to move quickly without the limitation of where data exists. They need a true data fabric.

Become AI Ready

NetApp helps you tailor your data fabric to accelerate your journey to AI. Only NetApp enables you to integrate your data fabric and streamline the flow of data from ingestion and collection at the edge, to preparation, training, and inference at the core, to analysis and tiering using the world's biggest clouds. Our unified data management supports seamless, cost-effective data movement across the hybrid multicloud environment.

NetApp provides deep technical integrations with leading AI technologies so you can simplify, accelerate, and protect your data pipeline for AI. With NetApp, you can build a data fabric that enables you to deploy AI with confidence.

Accelerate your data pipeline

[NetApp ONTAP AI](#) brings together NVIDIA DGX supercomputers, NetApp cloud-connected all-flash storage, and Cisco Nexus switches. This proven architecture simplifies, integrates, and accelerates your data pipeline for machine learning (ML) and deep learning (DL).

With ONTAP AI, you can eliminate guesswork and get started faster with a validated reference architecture that detangles design complexity. ONTAP AI allows you to start small and grow as needed, adding compute, storage, and networking to clustered configurations without disrupting ongoing operations. NetApp AFF systems keep data flowing to DL processes with the industry's fastest and most flexible all-flash storage, featuring the world's first end-to-end NVMe technologies. The AFF A800 array can feed data to NVIDIA DGX-1 systems up to four times faster than competing solutions.¹

ONTAP AI also enables you to unify data management across the pipeline with a single data fabric. Use the same tools to securely control and protect your data in flight, in use, or at rest, and meet compliance requirements with confidence—whether your data is on the premises or in the cloud.

Simplify DataOps

Today's world is containerized and cloud native. By deploying Kubernetes on top of NVIDIA DGX supercomputers, you can use the same hardware and container images across an orchestrated, distributed, and highly available platform. In fact, many AI applications, such as [Kubeflow](#) and [JupyterHub](#), require Kubernetes to run. With Kubernetes running on DGX supercomputers, you can give data scientists access to a self-service platform with massive compute power for AI pipelines and ML.

Using [Ansible modules for NetApp](#), you can fully automate the deployment of Kubernetes on your DGX systems. One of the biggest data challenges with containers is the issue of storage persistence—if a container is lost or deleted, all data inside is lost. With Trident, you can create persistent storage for Kubernetes with multiple storage classes (for example, Gold, Silver, and Bronze) for different performance and availability levels.

Trident allows Kubernetes users to create, manage, and interact with persistent storage volumes in the standard Kubernetes format that they are already familiar with, while still taking advantage of the advanced data management capabilities of NetApp technology. Trident abstracts away the complexities of persistent storage and makes it simple to consume.

Simplify version control for ML

AI is all about experimentation. Each experiment is composed of different hyperparameters, code, and input data. You must monitor and track every version of your experiments to determine the best models, manage the complex nature of ML and associated data, revert to previous working versions if necessary, and provide clear data provenance for regulatory audits.

This experimentation can be time consuming and frustrating if not done properly. NetApp Snapshot technology lets you capture point-in-time versions of the data, trained models, and logs. Because Snapshot copies capture only incremental changes, they don't consume any extra space and are extremely fast to create (~1 second for petabyte-scale data).

This industry-leading feature of NetApp ONTAP makes data versioning efficient when you are dealing with large amounts of data. Snapshot copies integrate with GitHub for ML versioning, making it easy to share and collaborate with others by using a normal git pull operation to get the latest complete ML version—not just code. To roll back to the previous best-known model, you need to pull only the specific version from the repository, and associated models are automatically updated. The same is true when trained models are accidentally deleted and need to be recovered.

Tap into the cloud

DL success requires an agile infrastructure with an ecosystem of technologies and solutions. The cloud offers an ideal environment that enables small and large organizations to get started with a pay-as-you-go structure, keeping costs low. Public clouds can be used to run proofs of concept (POCs), testing and development, data processing before beginning training, and full-scale production training.

Today's leading hyperscale clouds also provide a rich software ecosystem for building AI applications and training DL models. Using a storage solution with fast file services is essential to maintaining high GPU utilization, completing the training runs faster, and keeping costs to a minimum. NetApp cloud data services deliver instant productivity in the cloud. You can bring the superior data management and NFS capabilities or NetApp technology to Azure, Amazon Web Services, and Google Cloud with Azure NetApp Files, NetApp Cloud Volumes Service for AWS, and Cloud Volumes Service for Google Cloud Platform. The NetApp AI Control Plane provides full-stack data and experiment management across the hybrid cloud.

About NetApp

NetApp is the leader in cloud data services, empowering global organizations to change their world with data. Together with our partners, we are the only ones who can help you build your unique data fabric. Simplify hybrid multicloud and securely deliver the right data, services and applications to the right people at the right time. Learn more at www.netapp.com.

1. Read throughput of up to 300GBps per all-flash cluster versus 75GBps from a leading competitor.