PURESTORAGE®

NVIDIA. | NVIDIA ELITE PARTNER

High-performance AI for Federal Agencies

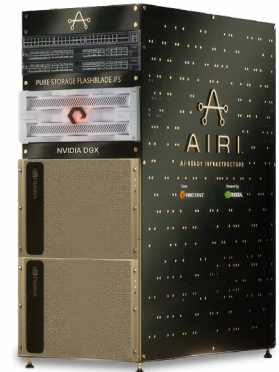Simple, accelerated, and scalable AI infrastructure from Pure Storage and NVIDIA

While artificial intelligence (AI) has the potential to deeply transform federal agencies, we're already seeing its potential to improve healthcare, public safety, national security, and citizen services. AI also helps agencies drive better, more informed, data-driven decision-making, and can free up millions of labor hours for critical tasks, leading to increased operational and administrative efficiency.

A recent report estimated that AI could boost productivity by $519 billion a year across all U.S. governments and they are poised to be the biggest AI spenders by industry with an estimated 19% CAGR in AI investment between 2022 and 2027. Government investments in artificial intelligence and data analytics will play a crucial role in shaping real-time operational decisions and outcomes. AI requires completely new infrastructure, and the complexities of legacy solutions create roadblocks that prevent agencies from tapping into the full potential of the new era of intelligence.

"Do-It-Yourself" Causes Roadblocks

The IT infrastructure historically used to support legacy data science efforts can't provide data access fast enough to unleash the advantage of GPU-powered AI-based data science. Agencies need solutions to enable AI simply, quickly, and cost effectively. When implementing home-grown AI solutions, agencies contend with a variety of roadblocks including:

- **Unexpected hardware expenses** due to the high-performance, dedicated, and optimized hardware required to successfully process each stage of the AI pipeline and store massive amounts of data. Legacy servers are not up to the task.

- **Complex AI software** requiring careful, precise, and frequent configuration by highly specialized developers.

- **Endless maintenance cycles** demanding constant software patching, tuning, and compiling, along with ongoing tweaking of the AI algorithms and frequent hardware upgrades to keep up with the unending deluge of new data.

AIRI™ Technology Stack

Pure Storage FlashBlade//S

Pure Storage Software: Purity// FB OE, Pure1® Management, Portworx® Kubernetes Data Platform, Pure RapidFile Toolkit

NVIDIA NVIDIA DGX H200, NVIDIA DGX™ B200

NVIDIA DGX BasePOD™

NVIDIA Spectrum™ and Quantum1 Networking

NVIDIA Software: NVIDIA AI Enterprise, NVIDIA Base Command™, NVIDIA NGC™, CUDA-X™, NVIDIA Magnum IO™, NVIDIA RAPIDS™

AIRI Simplifies AI at Scale

Reduces the complexity of integration with a proven complete DGX BasePOD solution

Data scientists can focus on algorithms and outcomes, not infrastructure

AI at Scale Is an Advantage

More compute = faster training

More data = higher accuracy

AIRI makes it simpler and faster to run multi-node training

AI at Scale, Made Simple and Fast

AIRI®, from Pure Storage®, is a reference architecture for NVIDIA DGX BasePOD™, representing the next evolution of the innovative and efficient AI Ready Infrastructure (AIRI) from Pure Storage and NVIDIA. AIRI combines Pure Storage FlashBlade//S™ with the NVIDIA DGX™ platform to provide the underlying infrastructure and software to accelerate deployment and execution of agency AI workloads. This includes support for NVIDIA Magnum IO™ GPUDirect® Storage, which provides the high-performance IO directly to the GPUs powering the AI infrastructure.

FlashBlade//S is the ideal data storage platform for AI, as it was purpose-built from the ground up for modern, unstructured workloads and accelerates AI processes with the most efficient storage platform at every step of your data pipeline. A centralized data platform in a deep learning architecture increases the productivity of data scientists and makes scaling and operations simpler and more agile for the data architect.

FlashBlade//S is also a certified ethernet storage solution for NVIDIA DGX SuperPOD™. Together, they provide powerful accelerated compute, a high-performance and efficient storage platform, and AI software and tools to address the most demanding AI and HPC workloads.

For full technical details, please consult the AIRI Pure Storage NVIDIA DGX BasePOD™ Reference Architecture and FlashBlade//S with NVIDIA DGX SuperPOD™ Reference Architecture.

Delivering the Next-generation AI Platform

The Pure Storage platform accelerates and simplifies AI adoption by helping agencies maximize performance and efficiency, unify their data, simplify data storage management, and solve the unpredictability of AI growth.

Accelerate and Simplify AI Adoption with the Pure Storage Platform			
Maximize Performance and Efficiency	**Unify Your Data to Speed Insights**	**Simplify AI Storage Management**	**Solve the Unpredictability of AI Growth**
• Accelerate AI data pipeline with parallel architecture & multi-dimensional performance • 85-95% less power and space for storage[1] = more for compute and GPUs • Improve data preparation with 64x faster file management with RapidFile Toolkit	• Consolidate data silos on shared, unified platform • Integrate diverse data sources, irrespective of protocol, file size, or file/object count • Single consistent platform for your entire AI ecosystem	• Improve productivity with simple & automated management and provisioning • Improve storage utilization with dynamic workload movement and AI intelligent recommendations • Fast-track deployment with validated designs	• Grow easily with SLA-driven storage-as-a-service • Meet AI growth requirements with non-disruptive capacity & performance expansion • Zero downtime deployment for new software upgrades

AIRI Technology Stack

AIRI represents a combination of groundbreaking solutions from two industry leaders. All-flash storage architected for AI from Pure Storage is paired with the latest AI hardware and software from NVIDIA. AIRI is optimized out-of-the-box and can seamlessly scale storage and AI compute performance, allowing agencies to begin their AI journey at any scale and grow and adapt as their needs evolve.

NVIDIA Base Command™ software and NVIDIA AI Enterprise software coupled with the Purity Operating Environment and Pure Storage RapidFile Toolkit provide a strong software foundation to jump-start and accelerate your AI journey.

Data cleaning, exploration & accelerated workflows
Pure Storage RapidFile Toolkit, Ansible, API's

NVIDIA AI Enterprise
GPU-accelerated AI frameworks, pretrained models, libraries, NVIDIA NIM, NVIDIA NeMo

Kubernetes Data Platform, Job Scheduling and Orchestration
portworx by Pure Storage   slurm

Cluster Management
NVIDIA Base Command Manager

Scale-out GPU compute
NVIDIA H200, B200

Unified Fast File and Object, Scale-out Data Services and Protocols
FlashBlade //S   Purity FB

Networking Fabric
NVIDIA Spectrum

NVIDIA Quantum*
*Quantum supported for large DGX clusters

Maximum Performance for Key Agency Use Cases

The combination of NVIDIA DGX systems with the massive parallelism of FlashBlade//S delivers maximum performance efficiency for key federal agency use cases such as:

- **National security:** Collecting, analyzing, and responding to intelligence; improving logistics, cyber operations, information operations, and command and control; used in a variety of semiautonomous and autonomous vehicles.

- **Cybersecurity:** Automating prevention, protection, and investigation of threats; safeguarding critical infrastructure; and enhancing response to incidents and threats.

- **Citizen services:** Implementing chatbots for citizen services can enable agencies to perform a variety of tasks including scheduling appointments, answering common questions, and directing requests to appropriate areas within the government.

- **Fraud identification:** Identifying fraudulent benefits claims that cost agencies billions. AI-powered fraud detection can enable governments to track down tax, benefit, and welfare corruption by identifying patterns in claims.

- **Healthcare:** Improving diagnosis, treatment, and delivery of patient care; tracking and preventing disease spread; leveraging large amounts of health data for research; classifying risks by drawing on population data; increasing operational capacity; identifying fraudulent health claims; and boosting productivity by streamlining manual tasks.

- **Predictive analytics:** Analyzing extensive datasets to make predictions, optimize services, and initiate appropriate actions, such as forecasting energy consumption patterns for optimal resource planning.

- **Operational efficiency:** Leveraging AI for expenditure analysis, process enhancements, and budget planning.

- **Document automation:** Extracting and inputting invoices, forms, legal, and other documents. Analyzing millions of tax documents and financial records to identify patterns that would indicate tax evasion or financial wrongdoing.

AIRI can optimize all these agency use cases and many more. With petabyte-scale storage, AIRI can support multiple use cases on a single platform.

Sustainable AI Adoption

AI adoption is on the rise across agencies, yet most organizations lack the necessary infrastructure to handle the high-performance data demands and energy requirements essential for maximizing its benefits. In fact, legacy systems often cannot support the massive AI data pipelines required to get the most from machine learning models. And while AI brings immense promise, its impact on energy requirements can be surprising. These hidden costs of AI pose a challenge to the successful implementation of critical government initiatives, including those aimed at achieving environmental goals.

According to a recent survey, AI adoption has risen from about 50% in previous years to 72%. This surge in AI implementation has led to a significant increase in energy requirements. Research estimates that AI could drive a 160% increase in data center power consumption, amounting to an additional 200 terawatt-hours per year.

Right now, the vast majority of data in data centers—over 80%—remains trapped on magnetic hard disks. Shifting to flash-optimized systems is a preliminary step agencies can take to reduce power usage between five and ten times. Pure Storage arrays use up to 85% less power than competitive all-flash systems. Pure Storage delivers a unified all-flash platform that helps maximize agency energy and space efficiency while reducing e-waste and operational costs.

Accelerate and Simplify AI Infrastructure for Faster Mission Outcomes

As AI continues to accelerate in adoption, IT teams require a meaningful data strategy to ensure they can efficiently and effectively operationalize AI through the right infrastructure. AIRI accelerates and simplifies the process of deploying and running complex AI infrastructure, allowing agency data teams to concentrate on their core mission, gain crucial insights, and achieve faster outcomes.

Additional Resources

- Learn more about the AIRI solution.
- Read the white paper: The Pure Storage Platform for AI.
- Explore Pure Storage solutions for federal agencies.

1 Vs. competitive all-flash storage