AIRI//S High-performance AI for Federal Agencies

Next-gen AI infrastructure to accelerate agency end-to-end GPU workflows.

Artificial intelligence (AI) has the potential to transform government, and we're already seeing its potential to improve health, safety, national security, and citizen services and drive better, more informed, data-driven decision-making. AI frees up millions of labor hours for critical tasks, leading to increased efficiency.

The power of AI can improve government operations, offering faster innovation and a better citizen experience. Federal agency spending on AI almost doubled from fiscal 2018 to fiscal 2022 and accounted for $8.4 billion of federal procurement. 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 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//S™ Technology Stack:

- Pure Storage FlashBlade//S
- NVIDIA DGX Systems
- NVIDIA Spectrum and Quantum[1] Networking
- Software includes: DGX software stack, Purity//FB OE, Pure1® Management, Pure RapidFile Toolkit.

AIRI//S Simplifies AI-at-scale

- Reduces racks of complexity into a complete solution
- Data scientists can focus on algorithms, not infrastructure

AI-at-scale Is an Advantage

- More compute = faster training
- More data = higher accuracy
- AIRI makes it simpler to run multi-node training

Uncomplicate Your Data

Fortunately, there is a much simpler path to realizing the benefits of AI. AIRI//S is the latest evolution of the industry's first complete AI-ready infrastructure. Architected by Pure Storage® and NVIDIA, and powered by the latest NVIDIA DGX systems and the new Pure Storage FlashBlade//S™, AIRI//S is a cutting-edge, disaggregated, and modular scale-out storage platform for unstructured data. The AI solution is out-of-the-box and simple-to-use, enabling you to reap the benefits of AI right away and delivering a simple, fast, and future-proof infrastructure to meet your AI demands at any scale.

Delivering the Next Generation AI Platform
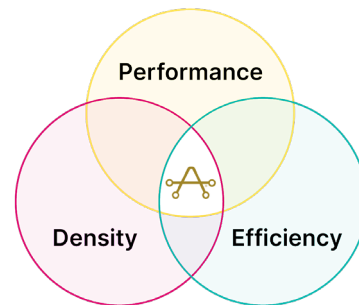
**Full Stack
Pre-integration for AI**

**Simplifies AI
Infrastructure**

**Maximizes Scalability
and Efficiency**

The goal of any AI factory is to maximize training performance. To achieve this, you need storage performance that meets the needs of the compute layer. Once this has been achieved, the next major focus should be on efficiency and density. Only AIRI//S achieves this sweet spot of performance, density, and efficiency that maximizes your GPU investment and optimizes TCO.

Performance

Density   Efficiency

Ease of Use

AIRI//S can be set up, deployed, and managed quickly as an end-to-end AI pipeline solution. It ensures that your data teams remain productive—spending more time delivering insights and less on worrying about their infrastructure.

Performance at Any Scale

Start your AI journey at any scale and seamlessly grow and adapt as needs evolve—all while delivering predictable performance to keep your GPU fed and data teams productive. FlashBlade//S scales to 1.92PB per single chassis and is greater than two times more performant than the previous generation FlashBlade®.

Sustainability

AIRI//S provides significant gains in data center efficiency. FlashBlade//S delivers industry leading 1.3 watts/TB which is 48% more efficient than another popular NAS solution. The DGX A100 MiG is seven times more performant than its predecessor the V100 GPU—in the same footprint.

Future Ready

AIRI//S is built for future upgrades to grow with your expanding AI needs. FlashBlade//S software and hardware upgrades are always included and non-disruptive. Future support includes Bright Cluster Management software, DGX H100, and more.

AIRI//S Technology Stack

AIRI represents a combination of groundbreaking solutions from two industry leaders. All-flash storage from Pure Storage is paired with NVIDIA's latest AI hardware and software. AIRI//S 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.

Data cleaning, exploration & accelerated workflows
Pure Storage RapidFile Toolkit

GPU Optimized Applications
NVIDIA CUDA-X
Chainer · mxnet · TensorFlow · Caffe2 · PyTorch · جاكس · PaddlePaddle · learn

Containerization
GPU-optimized NGC Library

Scale-out GPU compute
NVIDIA DGX A100

Networking Fabric
NVIDIA Spectrum
NVIDIA Quantum*

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

Maximum Performance for Key Agency Use Cases

The combination of NVIDIA DGX A100 systems with the massive parallelism of FlashBlade//S delivers maximum performance for key federal agency use cases.

- **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.

- **Social services and tax administration:** 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.

- **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.

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

AIRI//S from Pure Storage powered by NVIDIA accelerates and simplifies the process of deploying and running complex AI infrastructure, so agency data teams can focus on their core mission and discover critical insights and deliver mission outcomes faster.

Additional Resources

- Explore the AIRI//S solution.

- Learn more about DGX A100 systems.

- Read about how to leverage trustworthy AI in the Federal government.

- See how NASA accelerates new inventions and innovations.

---

[1] Quantum networking supported for larger DGX clusters