SOLUTION BRIEF

Supercharge AI for Medical Imaging

Accelerate end-to-end imaging GPU workflows with modern AI infrastructure.

Medical and research imaging is at the verge of AI adoption. Even as physician skepticism fades and regulatory approvals progress, organizations are discovering challenges with their data infrastructure. AI models consume vast amounts of data from a wide variety of sources. When creating models for medical imaging, data comes from X-rays, MRIs, CTs, ultrasounds, microscopy, and digital pathology. Each of those studies can vary widely in size and complexity, plus effective model training can require multi-petabytes of images. These images are usually stored in picture archiving and communication systems (PACS) or vendor neutral archive (VNA) systems, where advanced imaging analytics capabilities are lacking and data retrieval is difficult. The complexities of legacy or do-it-yourself storage and compute can slow down data exchange between applications, throttling testing and training. As a result, many AI projects require completely new infrastructure for faster innovation and a competitive edge.

Simple and Fast AI at Scale

Architected by Pure Storage® and NVIDIA® to extend the power of NVIDIA DGX™ A100 systems, AIRI® provides speed-of-light AI-ready infrastructure to drive the transformative change needed in healthcare today. Powered by FlashBlade® storage and DGX A100 systems, AIRI (AI-Ready Infrastructure) offers a simple, fast, and future-ready infrastructure to researchers, clinicians, data scientists, and IT teams looking to put their imaging repositories to work—to train models that can help identify tumors or lesions, detect diseases earlier, or improve clinical decision support—and meet their AI demands at any scale.

AI for Medical Imaging

- Virtualized GPUs and petabyte-scale storage.
- Designed for AI workflows that are vital to the imaging industry.

Infrastructure for AI

- Future-ready infrastructure to meet AI demands at any scale.
- Powered by FlashBlade® storage and NVIDIA® DGX™ A100.

Single Platform

- Meet analytics, training, and inference needs.
- A single platform for multiple workloads and file types.

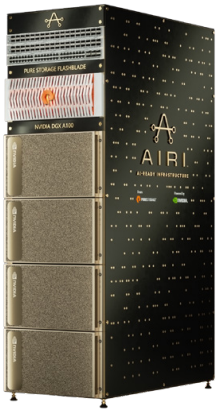
Maximum Performance for Key Imaging Use Cases

The combination of NVIDIA DGX A100 systems with the massive parallelism of FlashBlade delivers maximum performance for key imaging use cases. Data science is enabling an end-to-end transformation of the medical imaging industry in a wide variety of areas, including:

- **Enhanced imaging quality:** A recent survey by the ACR Data Science Institute shows that most radiologists expect near-term AI to improve lower-quality scans by reducing noise and artifacts, while enhancing contrast, resulting in a clearer view of the patient's pathology.
- **Improved patient experience:** Low-resolution images captured at low radiation doses can be enhanced by AI for clearer images for physicians to read. Plus, real-time AI can automatically detect motion on an image and determine whether the patient should be imaged again before leaving.
- **Optimized physician workflows:** AI is opening new frontiers in physician experience, including efficient worklist prioritization and automatic patient assignments to the most appropriate available physician.
- **Image triage:** AI systems can automatically triage diagnostic images so patients requiring immediate attention can be flagged and prioritized in real-time.
- **Clinical decision support:** When used with integrated information from electronic heath records (EHRs), PACS, or pathology systems, AI models can help physicians make better holistic decisions for their patients.
- **Improved diagnostics:** The ability to quickly identify, measure, and classify tumors, cardiovascular abnormalities, lung aberrations, or bone defects in scans can truly augment physician capabilities.
- **Clinical trials:** Aside from assisting with image reviews and measurements, AI can save precious time by automating quality control and the removal of protected health information (PHI) from images before they're transferred from trial sites to a central lab, contract research organization (CRO), or sponsor.

AIRI optimizes all these imaging use cases on a single platform by using virtualized GPUs and petabyte-scale storage.

AIRI Hardware and Software



Architected by Pure Storage and NVIDIA, AIRI enables AI at scale for every enterprise.

Hardware:

- 4x NVIDIA DGX A100 systems and 20 PFLOPS of AI performance
- Pure FlashBlade with 15x 17TB blades and 1.5M NFS IOPS
- 2x NVIDIA 32x 200Gb EN Switches

Software:

- NVIDIA DGX software stack with NVIDIA Optimized Containers

One Platform for All AI Workloads

Traditional approaches to AI infrastructure result in server and storage silos that are over-spent on capacity or starve AI workloads. Medical imaging AI data centers need a platform that is suited to the unique demands of analytics, training, and inference. Built on DGX A100 systems as the compute building block, AIRI flexibly adapts to business demand as AI models move from prototyping to deployment, with one universal AI system that offers right-sized resources for every workload and consolidates silos into a single elastic AI infrastructure.

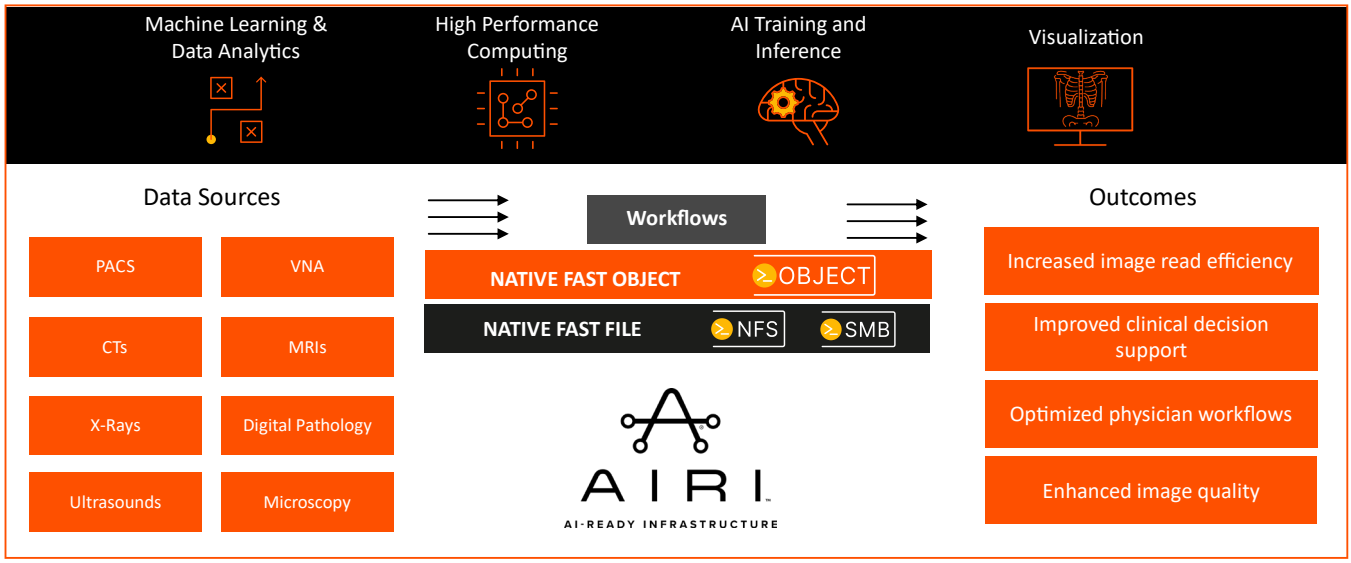


Figure 1. Facing an ever-growing number of data inputs, the medical imaging industry needs a high-performance analytics platform that helps ensure key outcomes.

Training AI models is the clearest interaction between data scientists and storage. As training jobs repeatedly read data sets, they drive a constant random read workload. DGX A100 systems process data quickly, so data must be accessible quickly as well. This requires a minimum non-sequential (random) read throughput from the underlying storage.

Unfortunately, most storage solutions are tuned to specific file sizes and access patterns. That is because it is often more efficient to read a small number of large files rather than millions of tiny files. AI workloads, however, are dependent on performant random read access for a range of file types and sizes. They also may need to write large amounts of data at parts of the workflow. Storage that's optimized for only one access pattern or file type is likely to become a bottleneck during some parts of the AI workflow.

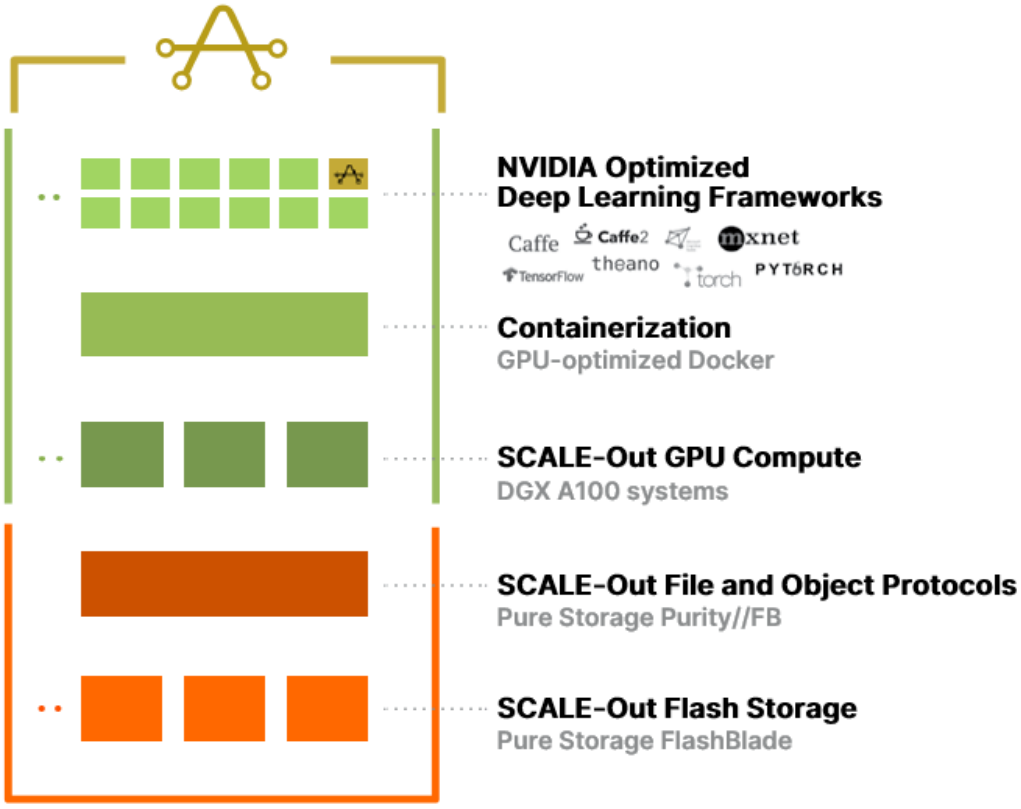
FlashBlade is uniquely designed to handle both large and small files and any range of access patterns. As such, it is ideally matched with NVIDIA to feed the GPUs at a pace that maximizes performance; GPUDirect can feed data-hungry AI models directly from storage arrays.

Eliminate Data Bottlenecks

Do-it-yourself infrastructure requires constant tuning. As one bottleneck is resolved, another often shows up somewhere else in the system, resulting in weeks to months of lost productivity. AIRI is an end-to-end, integrated infrastructure solution, tuned from software to hardware to keep the GPUs busy for workloads at any scale.

AIRI Technology Stack

AIRI is built with a complete software stack that enables data scientists to get up and running in a few hours, not weeks or months.



Additional Resources

- Read the AIRI reference architecture and performance benchmarks.
- Find out how Paige.AI uses AIRI to augment the clinical diagnosis of cancer.
- Explore more enterprise imaging solutions from Pure Storage.
- Learn more about NVIDIA Clara Imaging and DGX A100 systems.

PS2146-01 09/2021