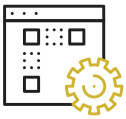
MEET AIRI™
YOUR FIRST AI-READY INFRASTRUCTURE AT SCALE

THE ERA OF AI IS HERE

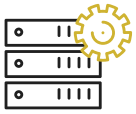
Enterprises will tap the power of AI for faster innovation and a competitive edge. Yet AI requires a completely new infrastructure, and the complexities of legacy solutions are holding enterprises back from moving into the new era of intelligence.

"DO-IT-YOURSELF" OFTEN YOUR ONLY OPTION

AI requires new, modern technologies like GPUs, scale-out flash, and RDMA fabric to move tremendous amounts of data. Unfortunately, too many AI initiatives are stalled with the complexities of a "do-it-yourself" approach using legacy technologies, leading to months of delays and idle time.



NEVER-ENDING CYCLES OF COMPILING & TUNING OPEN SOURCE SOFTWARE



MONTHS OF SYSTEM BUILDING AND TUNING, CONSTANT MAINTENANCE

LEGACY SOLUTIONS FULL OF DATA BOTTLENECKS, FROM STORAGE TO GPU TO APPS

AI-AT-SCALE MADE SIMPLE AND FAST

AIRI™ is the industry's first complete AI-ready infrastructure, architected by Pure Storage® and NVIDIA® to extend the power of NVIDIA® DGX™ systems. Powered by FlashBlade™ storage and NVIDIA DGX-1 servers, AIRI offers enterprises a simple, fast, and future-proof infrastructure capable of growing from AIRI "Mini" to rack-scale – and meeting AI demands at any scale, without downtime.

AIRI AI SOLUTION

ARCHITECTED BY PURE STORAGE AND NVIDIA, ENABLING AI-AT-SCALE FOR EVERY ENTERPRISE

AIRI

4x NVIDIA® DGX-1™ SYSTEMS
4 PFLOPS of DL Performance

PURE FLASHBLADE
15x 17TB Blades
1.5M NFS IOPS

CONVERGED FABRIC
2x 100Gb Ethernet Switches with RDMA

AIRI MINI

2x NVIDIA® DGX-1™ SYSTEMS
2 PFLOPS of DL Performance

PURE FLASHBLADE
7x 17TB Blades
700K NFS IOPS

CONVERGED FABRIC
2x 100Gb Ethernet Switches with RDMA

SOFTWARE

NVIDIA GPU CLOUD DEEP LEARNING STACK
NVIDIA Optimized Frameworks

AIRI SCALING TOOLKIT
Multi-Node Training Made Simple

PURESTORAGE®

AIRI™
Powered by NVIDIA

AIRI 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

AI IS DIVERSE AND EVOLVING

- From CNNs to GANs, AI is constantly changing
- AI is a data pipeline, requiring an integrated platform

50 RACKS UNDER 50 INCHES

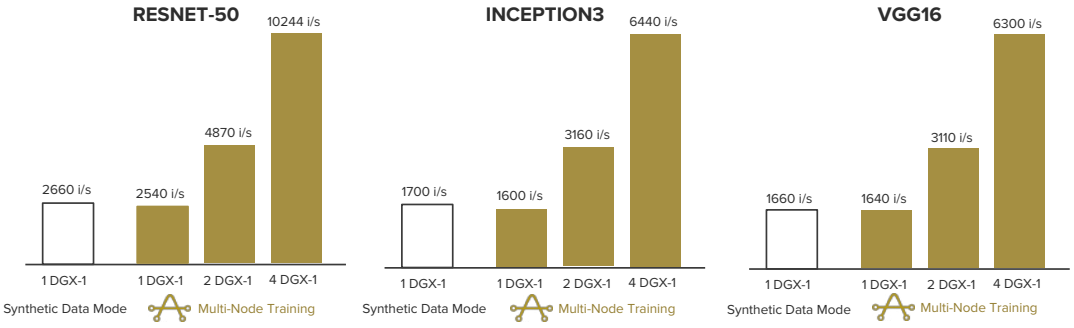
AI pushes beyond the reach of legacy technologies like serial CPUs and spinning disks. At the core of AIRI are NVIDIA® DGX-1™ servers and FlashBlade, industry-leading solutions architected for AI. Each replaces the performance of many racks of legacy technology, offering data scientists the power of a large supercomputer for any AI initiative – in a converged solution with a fraction of the physical space and power and cooling costs of legacy infrastructure.

DATA BOTTLENECKS, ELIMINATED

"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 a complete infrastructure, tuned from software to hardware to keep the GPUs busy for workloads at any scale.

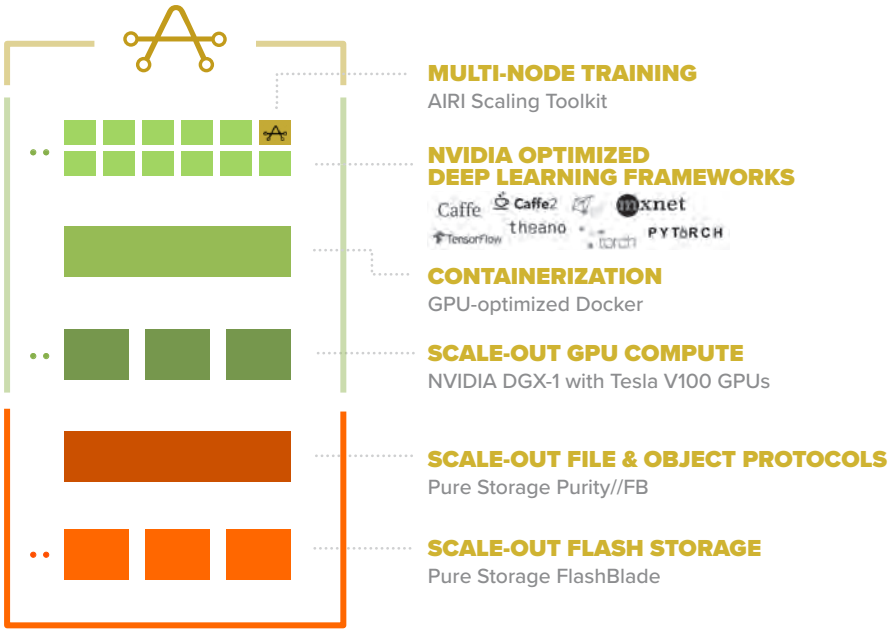
AIRI DELIVERS LINEAR, SCALE-OUT PERFORMANCE

KEEPING GPUs BUSY WITH TENSORFLOW AND 100Gb ETHERNET WITH RDMA



AIRI TECHNOLOGY STACK

EXTENDING THE POWER OF NVIDIA DGX-1 SYSTEMS FOR AI-AT-SCALE



AIRI is built with a **complete software stack** enabling data scientists to get up and running in a few hours, not weeks or months. **AIRI Scaling Toolkit** allows users to run their first multi-DGX training within hours, with a few commands, and to slash training time from weeks to days for their most critical AI workloads.

CUSTOMER HIGHLIGHT

LEADING SOLUTIONS PROVIDER BRINGING AI TO ALL ENTERPRISES

ElementAI is an artificial intelligence solutions provider that gives organizations unparalleled access to cutting-edge technology. It is founded by leading AI experts, including Yoshua Bengio, who is widely considered to be one of the three pioneers of deep learning.

ELEMENT AI

"AIRI represents an exciting breakthrough for AI adoption in the enterprise, **shattering the barrier of infrastructure complexities** and clearing the path to jumpstart any organization's AI initiative. AIRI is built with the same core solutions ElementAI uses extensively both internally and with customers – the NVIDIA® DGX-1™ and Pure Storage FlashBlade."

Jeremy Barnes, Chief Architect, ElementAI

VISIT **PURESTORAGE.COM/AIRI** TO LEARN MORE

NVIDIA PREFERRED PARTNER

NVIDIA DGX-1
ESSENTIAL INSTRUMENT FOR AI RESEARCH

The Challenges of Building a Platform for AI

Data scientists depend on computing performance to gain insights and innovate faster, using the power of deep learning and analytics. GPU technology offers a faster path to AI, but building a platform goes well beyond deploying a server and GPU's.

AI and deep learning can require a substantial commitment in software engineering. An investment that could delay your project by months as you integrate a complex stack of components and software including frameworks, libraries, and drivers. Once deployed, additional time and resources are continually needed as you wait for the ever-evolving open source software to stabilize. You'll also be waiting to optimize your infrastructure for performance, along with administrative costs that increase as the system scales.

The Fastest Path to Deep Learning

Inspired by the demands of AI and data science, NVIDIA® DGX-1™ fast-tracks your AI initiative with a solution that works right out of the box so that you can gain insights in hours instead of months. With DGX-1 you can simply plug in, power up, and get to work, thanks to the integrated NVIDIA deep learning software stack. In addition to leveraging the Ubuntu Linux Host OS, popular among developers, DGX-1 also supports Red Hat for organizations that require seamless integration within their existing enterprise IT management tools. Now you can start deep learning training in as little as a day, instead of spending months integrating, configuring, and troubleshooting hardware and software.

Effortless Productivity

NVIDIA DGX-1 removes the burden of continually optimizing your deep learning software and delivers a ready-to-use, optimized software stack that can save you hundreds of thousands of dollars. It includes access to today's most popular deep learning frameworks, NVIDIA DIGITS™ deep learning training application, third-party accelerated solutions, the NVIDIA Deep Learning SDK (e.g. cuDNN, cuBLAS, NCCL), CUDA® toolkit, Docker Engine Utility for NVIDIA GPU.

SYSTEM SPECIFICATIONS

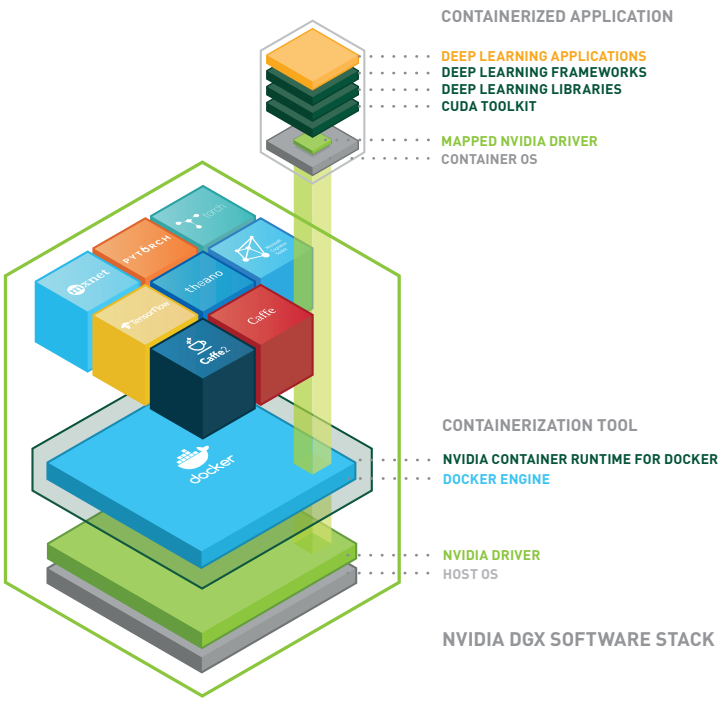
GPUs	**8X Tesla V100**
Performance (Mixed Precision)	**1 petaFLOPS**
GPU Memory	**256 GB total system**
CPU	**Dual 20-Core Intel Xeon E5-2698 v4 2.2 GHz**
NVIDIA CUDA® Cores	**40,960**
NVIDIA Tensor Cores (on V100 based systems)	**5,120**
Power Requirements	**3,500 W**
System Memory	**512 GB 2,133 MHz DDR4 RDIMM**
Storage	**4X 1.92 TB SSD RAID 0**
Network	**Dual 10 GbE, 4 IB EDR**
Operating System	**Canonical Ubuntu, Red Hat Enterprise Linux**
System Weight	**134 lbs**
System Dimensions	**866 D x 444 W x 131 H (mm)**
Packing Dimensions	**1,180 D x 730 W x 284 H (mm)**
Operating Temperature Range	**5–35 °C**

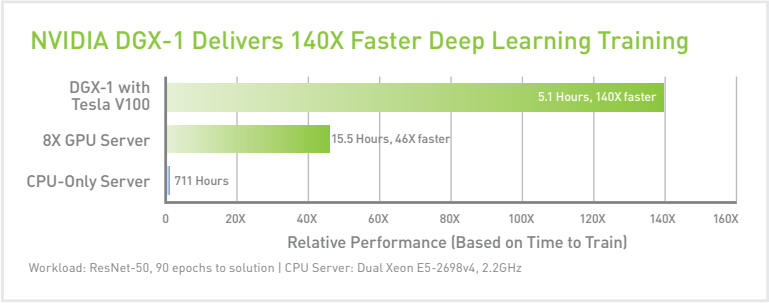
Revolutionary AI Performance

While many solutions offer GPU-accelerated performance, only NVIDIA DGX-1 unlocks the full potential of the latest NVIDIA GPU's like the NVIDIA Tesla® V100, including innovations like next generation NVLink™ and new Tensor Core architecture. With its performance-engineered deep learning software stack, DGX-1 delivers up to three times faster training speed than other GPU-based systems. With the computing capacity of 140 servers in a single system that integrates the latest NVIDIA GPU technology with the world's most advanced deep learning software stack, you can take advantage of revolutionary performance to gain insights faster than ever, powered by NVIDIA DGX-1.

Investment Protection

Your AI initiative is critical to your organization's success, and dependent on a frequently optimized software stack and integrated hardware infrastructure. With today's rapidly evolving open source software and the complexity of libraries, drivers, and hardware, it's good to know that NVIDIA's enterprise grade support and software engineering expertise are behind every DGX-1. This software stack is built on years of R&D, innovation, and deep learning expertise, and maintained by monthly optimized framework releases. Also, NVIDIA's support includes software upgrades and priority resolution of critical issues; you can have peace of mind that your environment is tuned for maximized performance and uptime.





ABC Product (Model) Name

ABC Product (Model) Name

Partner product description paragraph. One hundred words maximum. Xeris exeria nobis exerferis dolupt.

Spec 1: Some Data

Spec 2: Some Data

Spec 3: Some Data

Spec 4: Some Data

ABC COMPANY, INC.

NVIDIA SOLUTION PROVIDER

Company Name and/or Logo

Optional company brief description paragraph. No more than fifty words. Explia consequam il ilis escipiducium remd. Xeris exeria nobis exerferis dolupt, qui quo volores dolori blab iliquate il il excerum excesequi dolori manaianisi mintes.

www.abccompany.com | +1 (123) 555-678 | jdoe@abccompany.com

For more information, visit **www.nvidia.com/dgx-1**

NVIDIA

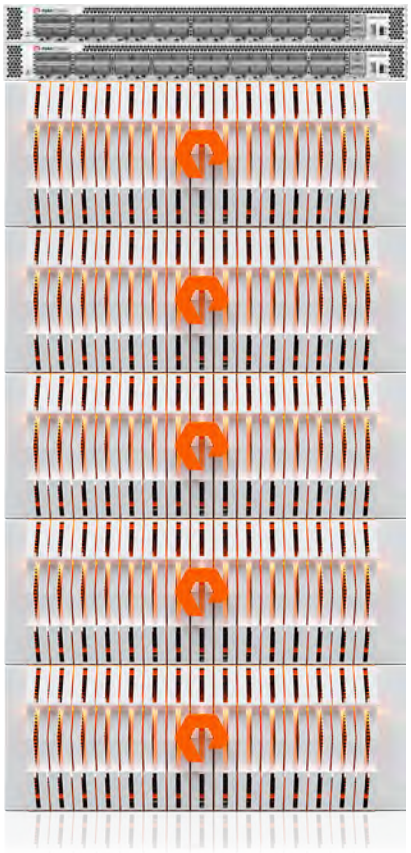
FLASHBLADE: SCALE-OUT STORAGE FOR MODERN DATA

SUMMARY

Data is the most valuable asset in an organization today. However, slow and complex legacy storage systems often hold data back from being put to use. FlashBlade is the industry's most advanced file and object storage platform ever, a data hub built to consolidate data silos like backup appliances and data lakes – to accelerate tomorrow's discoveries and insights.

RETHINK STORAGE IN THE ERA OF MODERN DATA

There are two types of storage systems. One is optimized to store data. The other is optimized to deliver it. One is engineered with legacy technologies, like spinning disk or retrofit software. The other is a modern system, architected from the ground up to be massively parallel, thus eliminating serial bottlenecks that hold data back. This modern storage is FlashBlade™ from Pure Storage.



DATA HUB MODERNIZE EVERYTHING – FROM BACKUP APPLIANCES TO DATA LAKES

From artificial intelligence to analytics, data is at the center of today's innovation. Organizations are often hindered by legacy infrastructure, which prevents their data from moving at the speed of their business. FlashBlade is the industry's most advanced scale-out storage, architected to accelerate modern workloads and simplify infrastructure.

*"Pure Storage FlashBlade is about **10 times faster out of the box, with no specific tuning or effort.** It enabled us to boost our GPU from about 20% average utilization to close to 100% utilization."*

— ***JEREMY BARNES**, CHIEF ARCHITECT*

ELEMENT AI

FLASHBLADE – POWERED BY THREE INDUSTRY FIRSTS

BLADE

Compute and network integrated with DirectFlash™ technology – hot-plug blades into the system to add capacity and performance.

PURITY//FB

The heart of FlashBlade, architected on a massively distributed key-value pair database for limitless scale and performance with simplicity.

ELASTIC FABRIC

Powered by an innovative converged fabric, FlashBlade delivers up to 1.5Tb/s aggregate bandwidth with 75 blades.

Accelerate AI data pipeline while keeping GPUs fully utilized

Restore data rapidly to meet SLAs for disaster recovery

Shorten design cycles in EDA by eliminating data bottlenecks

Consolidate data warehouses and data lakes for simplicity and real-time performance

Meet ever-growing demands of modern DevOps environments

Unleash data-intensive simulations from genomics and finance workloads

REPLACE RACKS OF LEGACY TECHNOLOGY WITH FLASHBLADE

FlashBlade delivers unprecedented performance in a small form factor. It is tuned to deliver multi-dimensional performance for any data size, structure, or access, delivering 10x or greater savings in power, space, and cooling costs.

"FlashBlade is just 4U, but provides the same performance as 50 racks of our legacy storage sitting in the same data center. It's extraordinary."

*— **JIM DOLAN**, MANAGER, HPC WORLDWIDE SUPPORT*

ion

EVERYTHING YOU EVER WANTED IN A DATA HUB

Some storage alternatives claim to be performant, but are complex to deploy. Others promise large capacity, but deliver data slowly. FlashBlade is the first scale-out storage solution to intersect on all three dimensions of big, fast, and simple.

FAST

- Elastic performance that grows with data, up to 75 GB/s
- Always-fast, from small to large files
- Massively parallel architecture from software to flash

BIG

- Petabytes of capacity
- Elastic concurrency, up to 10s of thousands of clients
- 10s of billions of objects and files

SIMPLE

- Evergreen™ – don't rebuy TBs you already own
- "Tuned for Everything" design, no manual optimizations required
- Scale-out everything instantly by simply adding blades

SPECIFICATIONS

PERFORMANCE

- 17 GB/s bandwidth with 15 blades
- 75 GB/s bandwidth with 75 blades
- 7.5M NFS IOPS with 75 blades

CONNECTIVITY

- 8x 40Gb/s or 32x 10Gb/s Ethernet ports / chassis
- 2x FlashBlade External Fabric Modules (XFM) to scale up to 75 blades

PHYSICAL

- 4U per chassis
- 1,800 watts per chassis (nominal at full configuration)

*"Our quants want to test a model, get the results, and then test another one – all day long. So a **10-20X improvement in performance is a game-changer** when it comes to creating a time-to-market advantage for us."*

*— **GARY COLLIER**, CO-CTO*

Man | AHL

PURESTORAGE®

VISIT **PURESTORAGE.COM/FLASHBLADE** TO LEARN MORE