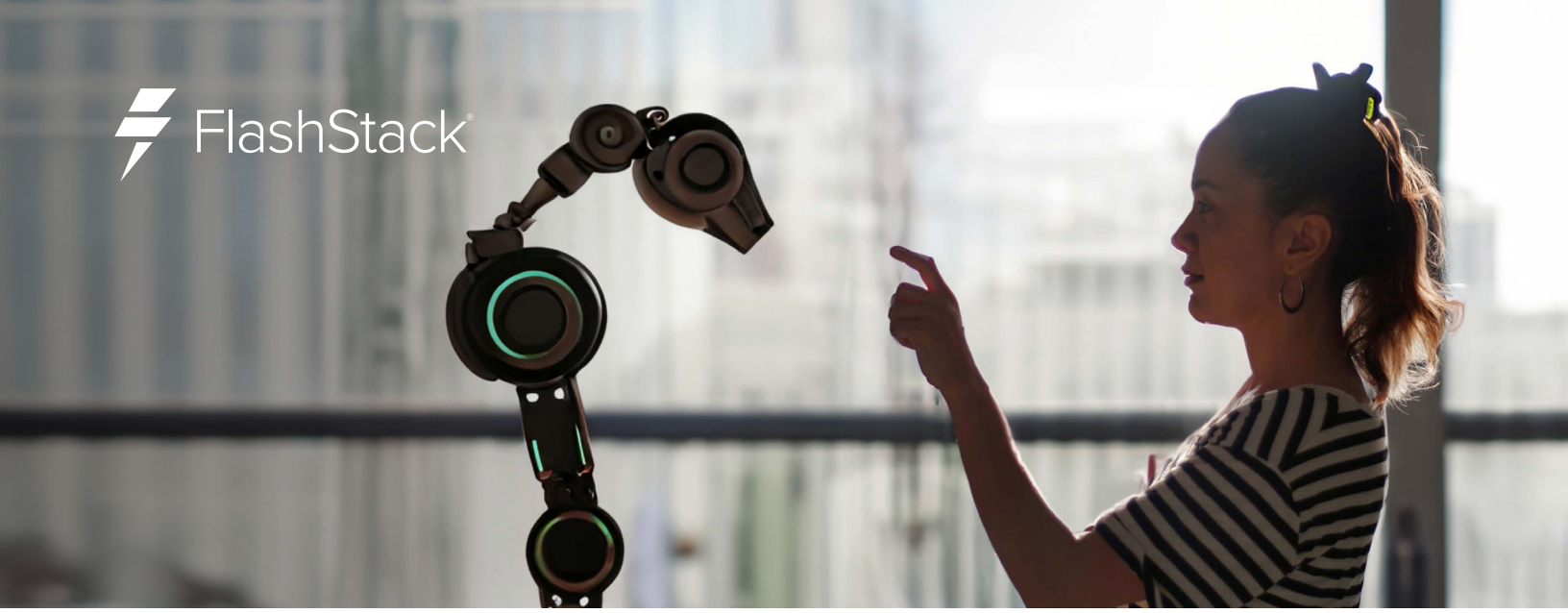


SOLUTION BRIEF

FlashStack for AI: The Foundation for Enterprise AI Factories

Pure Storage, Cisco, and NVIDIA deliver validated architectures and scale Enterprise AI Factories.

Enterprises need a faster, more reliable way to operationalize AI. FlashStack® for AI from Pure Storage, Cisco, and NVIDIA delivers an integrated, validated foundation for building and scaling Enterprise AI Factories. Pre-engineered FlashStack AI PODs accelerate time-to-value, while Cisco Validated Designs ensure predictable performance and risk-free operations. With unified management, built-in sustainability, and layered cyber resilience for added security, FlashStack helps organizations move from AI pilot projects to enterprise-wide innovation—turning data and compute into intelligence at scale.

Highlights

- Accelerate time-to-value with FlashStack AI PODs and Cisco Validated Designs for AI workloads.
- Build and scale Enterprise AI Factories powered by Pure Storage, Cisco, and NVIDIA.
- Operate simply with unified observability and automation via Intersight + Pure1®.
- Protect data with layered cyber resilience and SafeMode™ immutability.

NVIDIA PURESTORAGE®

Simplifying and Scaling Enterprise AI

AI is now central to business competitiveness. Organizations are racing to operationalize AI quickly, reliably, and at scale—but traditional infrastructure makes that difficult.

FlashStack for AI from Pure Storage, Cisco, and NVIDIA delivers a full-stack platform for building and scaling Enterprise AI Factories that integrate compute, networking, storage, and software into a unified, validated architecture.

With Cisco Validated Designs (CVDs) and modular AI PODs, FlashStack reduces design risk, speeds deployment, and simplifies day-2 operations—while ensuring performance, sustainability, and security. Integrated NVIDIA AI Enterprise software and unified management with Cisco Intersight and Pure1® turn complex AI infrastructure into a single, intelligent platform that supports generative AI, inference, and MLOps workloads.

Simple. Sustainable. Secure.

Simple

Deploying AI infrastructure is complex—but it doesn't have to be. FlashStack for AI simplifies deployment through Cisco Validated Designs that integrate Cisco UCS compute, NVIDIA GPUs, and Pure Storage FlashBlade//S™ unified storage into an enterprise-ready AI platform. FlashStack AI PODs provide pre-validated, workload-tuned configurations that scale predictably into full AI Factory architectures, enabling faster time-to-value and reliable outcomes. And with Cisco Intersight and Pure1, IT teams gain unified visibility, automation, and proactive insights that streamline day-2 operations and ongoing optimization.

Sustainable

AI processing is power-intensive. FlashStack was redesigned to deliver industry-leading efficiency, with up to 85% lower power consumption and an 80% smaller footprint than traditional infrastructure. Enterprises can achieve more AI performance per watt while meeting ESG and sustainability objectives.The modular design found in FlashStack enables you to scale non-disruptively and adopt new technologies without forklift upgrades, preserving capital and reducing environmental impact.

Secure

AI often involves the most sensitive and regulated data within an enterprise. FlashStack safeguards that data with a cyber resilience foundation from Pure Storage and Cisco. A layered security model with immutable and indelible Pure Storage SafeMode™ Snapshots; integrated recovery with Veeam, Rubrik, and Commvault; and comprehensive cybersecurity from Cisco including XDR, SOAR, and SIEM integrations helps organizations maintain AI data integrity from core to cloud to edge.

NVIDIA. PURESTORAGE®

FlashStack AI PODs and the Enterprise AI Factory

FlashStack AI PODs are pre-validated, workload-specific configurations that make AI deployment faster and more predictable. Each POD combines compute, networking, storage, and NVIDIA AI Enterprise software tuned for use cases such as retrieval-augmented generation (RAG), multimodal inference, or LLM fine-tuning. Our AI PODs act as modular building blocks that scale seamlessly into the Enterprise AI Factory architecture on FlashStack.

Validated integration with NVIDIA AI Enterprise ensures optimized performance for generative and inferencing workloads, giving IT and data teams a consistent operational model across on-premises, hybrid, and cloud environments.

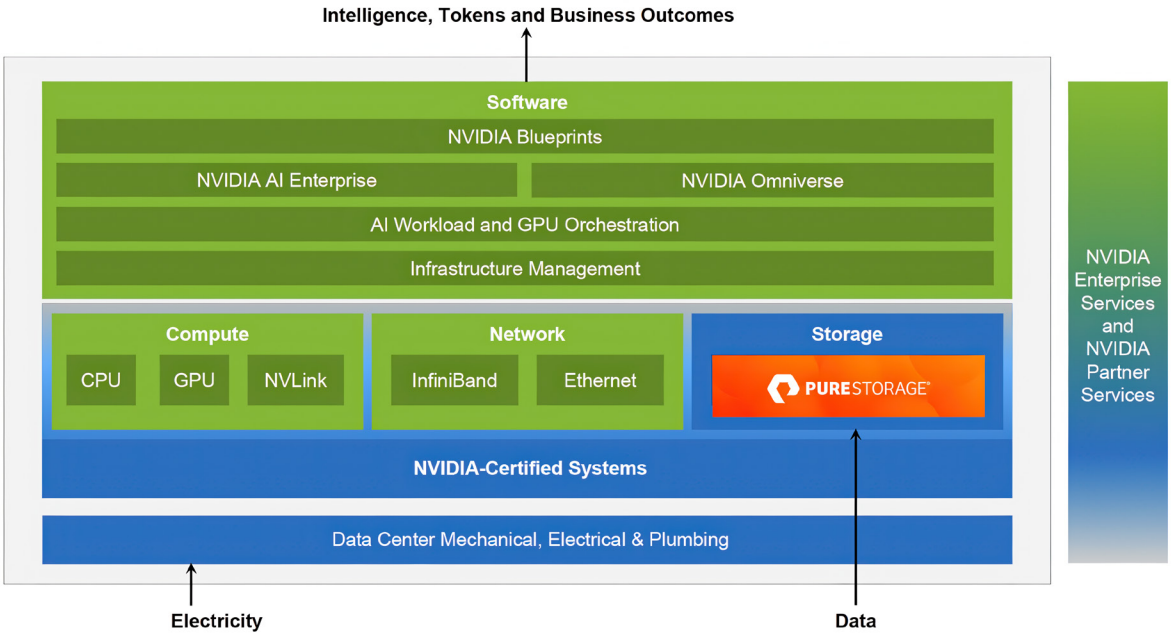
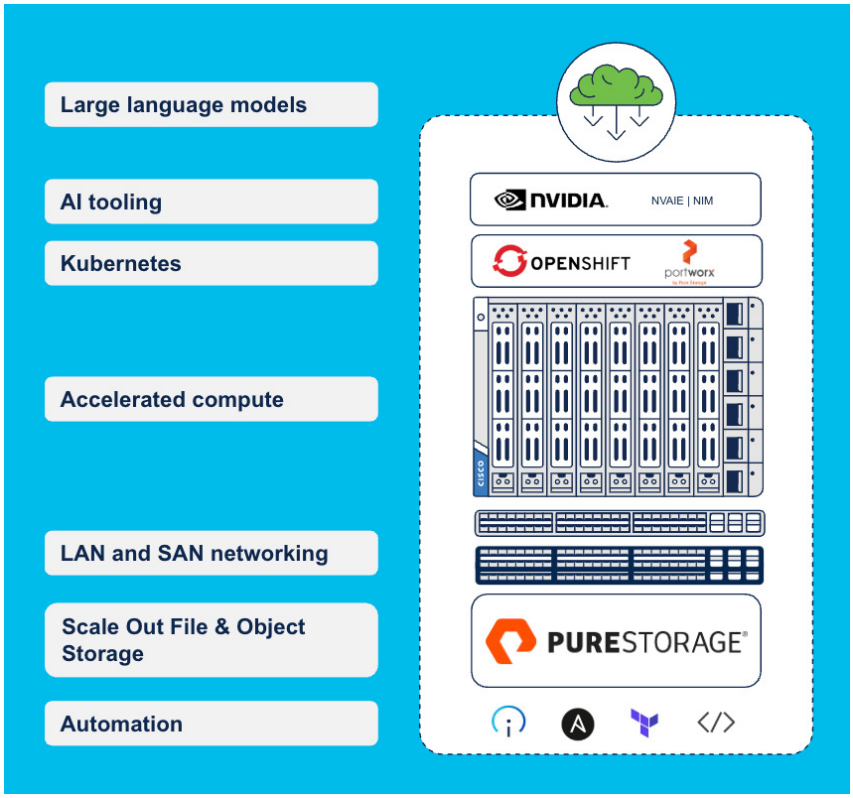


FIGURE 1 NVIDIA AI factory architecture on FlashStack

NVIDIA PURESTORAGE®

MLOps with Red Hat OpenShift AI, validated for NVIDIA AI Enterprise

As part of the Enterprise AI ecosystem, FlashStack also supports Red Hat OpenShift AI and Portworx® by Pure Storage for MLOps orchestration. This combination enables containerized model development, training, and inference with persistent storage and data mobility across clouds. Integrated Cisco Intersight automation and NVIDIA AI Enterprise software provide an open, portable, and production-ready platform to operationalize AI pipelines efficiently. FlashStack architectures are tested and validated with NVIDIA AI Enterprise, ensuring performance and support for generative AI, MLOps, and inference workloads.



NVIDIA PURESTORAGE®

FlashStack®

SOLUTION BRIEF

"Pure Storage and Cisco customers realized significant value by leveraging FlashStack to optimize IT and business operations. Through staff efficiency gains, reduced IT costs, better performance, and business enablement, IDC found that interviewed organizations achieved benefits averaging $4.8 million per year, resulting in a 3-year ROI of 343%.

IDC BUSINESS VALUE SNAPSHOT
(APRIL 2025, #US53137425)

FlashStack for the Enterprise AI Factory

FlashStack for AI delivers a converged, enterprise-grade foundation for NVIDIA AI Factories—validated by Cisco, Pure Storage, and NVIDIA. It eliminates DIY risk through CVD-enabled deployment, unifies operations with Intersight and Pure1, and scales from AI PODs to full AI Factories that combine performance, resiliency, sustainability, and data confidence.

Our joint solution features validated NVIDIA RTX Pro 6000 Generation GPUs, BlueField-3 DPUs, and NVIDIA AI Enterprise (NVAIE), NIMs, and Blueprints to provide a full-stack, future-proof platform for enterprises to immediately operationalize AI at scale.

Business benefits include:

- Accelerate innovation and reduce risk with Cisco Validated Designs for NVIDIA AI Factory architectures
- Scale from FlashStack AI PODs to full AI platforms with built-in sustainability and AI data confidence powered by Pure FlashBlade//S and Portworx
- Modern virtualization for AI—run workloads on VMware, alternative hypervisors, or container-based environments and extend to the cloud
- Improve economics and resilience with a sustainable architecture that lowers energy costs, reduces footprint, and increases operational uptime

flashstack@purestorage.com | www.cisco.com/go/flashstack | www.flashstack.com

NVIDIA.

PURESTORAGE®

ID9311-01-en-10/25