



Nutanix Gives Enterprises the Freedom to Run Production Agentic AI Their Way

August 26, 2026

NEWS SUMMARY

- *Nutanix Cloud Platform (NCP) is being enhanced and expanded for production agentic AI with its dual-native architecture, including the introduction of Nutanix Enterprise AI (NAI) 2.8 and Nutanix Kubernetes Platform (NKP) 2.19.*
- *The new capabilities give customers a flexible cloud operating model designed to consistently manage and govern AI across environments, supported by a dual-native architecture for virtual machines and containers.*
- *[NAI 2.8](#) is available now. It provides centralized control for AI inference and agentic AI, including [Nutanix Agent Gateway](#), now with a generally available Model Context Protocol (MCP) gateway for governing how agents connect with apps and data via MCP. Nutanix Private Inference also provides enhanced capabilities for high-performance fine tuning and inference, along with improved security and governance.*
- *NKP 2.19 will be available soon. It is expected to provide [streamlined container management for bare metal and virtualized environments](#), with a built-in AI catalog designed for building and running agentic AI applications.*
- *The new capabilities in NAI 2.8 and NKP 2.19, combined with [Cloud Native Computing Foundation \(CNCF\) certification](#), enhance NCP dual-native and AI capabilities.*
- *Nutanix: Verified Services program and [Service Provider \(SP\) Central](#) are now available to help partners drive new AI opportunities.*

WHY IT MATTERS

Enterprises often struggle to deploy AI alongside the apps and data it needs to access. Nutanix aims to solve this with a governed, dual-native architecture that runs both virtual machines (VMs) and containers side-by-side. This allows customers to bring AI directly to their apps and data regardless of how they are deployed, helping accelerate ROI without adding architectural complexity.

SAN JOSE, Calif., Aug. 26, 2026 (GLOBE NEWSWIRE) -- [Nutanix](#) (NASDAQ: [NTNX](#)), a hybrid cloud leader and AI innovator, today announced the general availability of Nutanix Enterprise AI (NAI) 2.8, and the upcoming general availability of Nutanix Kubernetes Platform (NKP) 2.19, along with new incentives, programs, and resources designed to help partners accelerate growth on emerging AI opportunities.

Many enterprises face a major roadblock when deploying AI: AI is accelerating the shift to containers, while critical applications and data remain spread across both virtualized and containerized environments. This divide can force enterprises to add infrastructure silos, move data or rearchitect existing workloads to support AI alongside the applications and data they already run. Nutanix addresses this with a governed, dual-native architecture that runs traditional applications and modern AI side by side, allowing the infrastructure to flexibly support the workload, integrated with leading silicon partners to provide choice and flexibility to customers. By bringing AI to where enterprise data already lives, Nutanix helps customers reduce silos and accelerate ROI without costly rearchitecting or added networking and data layer complexity.

Expanding on NCP capabilities, NAI and NKP are designed to enable organizations to securely run, manage, and govern AI, containerized applications and virtualized workloads through a consistent control plane. These complement the core platform capabilities for [near-bare metal performance for AI on virtualized infrastructure](#) introduced with NCI 7.6. Together, they give enterprises a flexible alternative to infrastructure stacks that limit architectural choice, drive up costs, and require disruptive platform changes.

For partners, the new Powered by Nutanix: Verified Services program aims to enable them to capitalize on major industry shifts and next-generation AI deployments by empowering them to build validated, high-margin services practices that span the full customer lifecycle. To operationalize these flexible new offerings, Service Provider (SP) Central provides an adaptable multitenant cloud foundation, giving service providers the control they need to grow infrastructure, platform, cloud-native, and AI services on their own terms.

"Enterprise AI should not require customers to rebuild the systems that already run their business," said **Thomas Cornely, Executive Vice President, Product Management**, Nutanix. "With our dual-native architecture, customers can bring AI to their existing applications and data while running each workload on the infrastructure best suited to it, with consistent operations and governance across VMs, containers and AI. This gives organizations a practical path to production without creating new silos or limiting future choice."

Nutanix Enterprise AI: Helping Enterprises Bring AI to Their Data Without Rebuilding Everything

NAI delivers a unified and secure platform to deploy, manage, and scale AI workloads across hybrid environments. It enables enterprises to enforce governance over their agents and models, gain total visibility over token usage, and streamline AI development with a simple interface with built-in observability metrics and easy to use model-as-a-service while ensuring security, scalability, and integration with existing infrastructure.

Key NAI updates include:

- **Agent Gateway:** This now includes a generally available MCP Gateway which serves as a secure, unified front door for AI agents to access tools and data without custom engineering. To complement this, Nutanix has also released [MCP Server](#) for NCP to help customers build agentic AI applications with secure access to the infrastructure managed by Nutanix.
- **Private Inference:** New advanced inference and fine tuning capabilities enable scalable, multiGPU inference for LLMs via tensor parallelism, delivering high-throughput serving and low-latency response times for enterprise LLM workloads. In

addition, this enables batch inference and speculative decoding. Key features include: 1) Parameter-Efficient Fine-Tuning which supports Low-Rank Adaptation (LoRA) fine-tuning for smaller models (<8B parameters), helping organizations to cost-effectively customize open LLMs on private domain data using single-GPU compute while seamlessly deploying adapters straight to serving pipelines; 2) Scalable multiGPU serving which enables high-throughput multiGPU inference via tensor parallelism, delivering fast, distributed serving across enterprise hybrid cloud environments; and 3) Speculative decoding which accelerates LLM inference token generation by up to 2.5x using lightweight draft models, cutting output latency without sacrificing model accuracy.

- **Enhanced Security against Rogue AI:** With the rise of agentic AI and the risk of models breaking out of sandboxes, security is paramount. NAI provides robust protection against rogue models through our platform and APIs, featuring fine-grained Identity and Access Management (IAM), custom roles and seamless model sharing. This enforces least-privilege security, helping ensure agents operate securely and restricting access to only authorized roles, as well as support for air-gapped NVIDIA NIM deployment.

Nutanix Kubernetes Platform: A Trusted, Production-Ready Foundation for Modern Apps and AI

With NKP, organizations can simplify container operations across containers running on bare metal and VMs without piecing together complex, custom stacks. The upcoming release will provide an AI-optimized platform for building and running agentic applications at scale, including the following features coming soon:

- **NKP Metal:** Built to bring HCI-grade simplicity to bare-metal Kubernetes, with automated OS, firmware, and container deployment, and persistent, enterprise-grade storage natively, eliminating the complexity of patchwork platforms.
- **NKP Full Stack:** While NKP Metal is intended to bring simplicity to bare-metal deployments, NKP on AHV remains the cornerstone for organizations requiring robust, agile virtualized environments. Combined with Nutanix Flow, NKP on AHV is designed to deliver stronger network-level sandboxing for AI agents, helping provide essential isolation to mitigate the risk of rogue attacks and lateral movement.
- **AI Applications Catalog:** Offers a one-click deployment path for curated, validated AI/ML software (Kubeflow, Milvus, Slurm) to help bypass manual integration challenges.
- **Hardware and Compliance:** Planned expansion of ecosystem support with validated GPU integrations, alongside dynamic resource allocation for modern AI workloads.

CNCF Certified Kubernetes AI Conformant Platform: NKP has attained formal CNCF certification to validate that NKP provides the standardized APIs and capabilities required to reliably operate enterprise AI workloads.

AI Storage Performance and Validated Certifications with NVIDIA

Demanding AI workloads require infrastructure that keeps data moving, maximizes GPU utilization, and reduces deployment risk. Nutanix Unified Storage (NUS) recently achieved [NVIDIA-Certified](#) Storage validation at the enterprise level, providing a trusted, interoperable foundation that helps eliminate data bottlenecks. NUS establishes a low-latency, high-throughput data path directly to GPUs, maximizing GPU utilization and ensuring linear scalability for large-scale production AI workloads.

Helping Partners Build, Monetize and Grow AI Services

According to Gartner®, “By 2029, 55 percent of enterprises will migrate 100 percent of workloads from VMware to alternative infrastructure delivery solutions”. To help partners capitalize on this market shift, next-generation AI deployments and other emerging opportunities, Nutanix recently launched the Powered by Nutanix: Verified Services program. This program provides the framework to support the broader partner ecosystem in transitioning from traditional, one-time deals to high-margin, recurring revenue streams by building validated services practices. Backed by streamlined onboarding, comprehensive service delivery kits, and exclusive badging across hybrid cloud infrastructure, Kubernetes, and VM migration, the program equips partners to own the full customer lifecycle, drive faster time-to-value, and maximize long-term retention.

Nutanix is also announcing the general availability of SP Central, a unified multitenant control plane that gives displaced VMware service provider partners greater choice in how they build and monetize services. Service Provider Central gives providers one consistent foundation to build and monetize a broad portfolio of infrastructure, application, cloud-native, and AI services, with the flexibility on deployment location and licensing they need. This is designed to help providers improve utilization and protect margins while giving customers more choice in how and where they run applications and AI workloads.

Customer Comment

“The Nutanix Cloud Platform gives us a dual-native architecture that is governed and brings AI to our applications and data across virtual machines and containers. This helps speed AI innovation across the production environments our organization depends on.”

Jasim Abdul Rahman, Group Chief Information Officer, Power International Holding

Industry Analyst Commentary

“By introducing a “dual native” model that unifies services such as networking, security, and data protection across both VMs and containers, Nutanix is leaning into a competitive landscape where consistent policy enforcement and lifecycle management are becoming key differentiators. Enterprises are increasingly prioritizing platform consistency as application portfolios span legacy, refactored, and cloud-native designs.”

- Matt Flug, IDC

Ecosystem Partners Commentary

"As enterprises deploy agentic AI to automate complex workflows, they need infrastructure that delivers the performance, security and control required for production environments. Cisco is bringing that foundation. Together with Nutanix's governed, dual-native cloud operating model, we're giving organizations greater choice in how they deploy and operate AI while maintaining the control they need across their environment."

– **Jeremy Foster, Cisco GM & SVP, Cisco Compute**

"At Continent 8 Technologies, we've worked closely alongside Nutanix to help shape the next generation of cloud services for regulated industries. Our continued collaboration with Nutanix, including the development of Service Provider Central, is built around helping our customers adopt modern cloud and AI capabilities without adding unnecessary complexity. Together, we're enabling organizations in highly regulated industries to innovate faster and focus on their core business, while maintaining the security, compliance, and operational control they require."

— **Edward O'Connor, Chief Technology Officer at Continent 8 Technologies**

"The rise of agentic AI is redefining how AI services are delivered, making security and sovereignty foundational requirements. Through our collaboration with Nutanix, we are enabling enterprises and service providers to build high-performance, multitenant AI platforms with hardware-enforced isolation and security. This ensures that sensitive data and models remain protected in use, delivering the secure scale and efficiency required for next-generation AI factories."

– **Bill Pearson, VP, Data Center Software, Intel**

Availability

NAI 2.8 and SP Central are generally available now. NKP 2.19 will be available soon.

Frequently Asked Questions

What is Nutanix's dual-native architecture?

Nutanix treats virtual machines and containers as first-class infrastructure, giving organizations a consistent way to run and govern traditional applications, modern apps and AI workloads.

How does Nutanix give customers more choice in where they run AI and why is this important?

Nutanix enables organizations to place AI workloads on the infrastructure best suited to their performance, data and business requirements without forcing every workload into the same architecture, including across VMs and containers, CPUs and GPUs (and other accelerators), private and public model providers, and more. For example, agentic AI applications often involve agents that are compute-intensive (running on CPUs) and intelligence functions that require GPU acceleration. Nutanix allows agents as well as Inference workloads to run on the optimal substrate with a unified AI cloud control plane.

How does Nutanix help enterprises connect AI with existing applications and data?

Nutanix brings AI closer to the applications, workflows and proprietary data already running in enterprise environments, helping customers reduce silos and minimize the need for costly rearchitecture.

How does Service Provider Central help partners deliver cloud and AI services?

SP Central gives service providers a self-service tenant lifecycle management platform to build and deliver infrastructure, platform, cloud-native, and AI services, supporting their efforts to monetize these offerings while maintaining centralized operations and governance, built on top of NCP's multitenant architecture.

*Gartner, 2026 Strategic Roadmap for VMware Modernization, Julia Palmer, et al., 7 April 2026

GARTNER is a trademark of Gartner, Inc. and/or its affiliates.

IDC, Nutanix .NEXT 2026: Enabling Customer Choice in Compute and Deployment Architectures, April 10, 2026, Matt Flug

About Nutanix

Nutanix is a hybrid cloud leader and AI innovator, offering organizations a unified infrastructure software platform to safely run applications, data, and AI anywhere. Trusted by customers worldwide, Nutanix empowers more than 50% of the Global 2000 to innovate faster with AI, while modernizing infrastructure, simplifying operations, and controlling costs. Learn more at www.nutanix.com or follow us on social media.

©2026 Nutanix, Inc. All rights reserved. Nutanix, the Nutanix logo, and all Nutanix product and service names mentioned herein are registered trademarks or unregistered trademarks of Nutanix, Inc. ("Nutanix") in the United States and other countries. Other brand names or marks mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s). This press release is for informational purposes only and nothing herein constitutes a warranty or other binding commitment by Nutanix. This content contains express and implied forward-looking statements, including but not limited to statements regarding the availability and capabilities of the upcoming NKP 2.19 release, its features, performance, integrations, certifications and ecosystem support, the functionality and strategic direction of NAI and SP Central, and our plans and expectations regarding the Powered by Nutanix: Verified Services program. Such statements are not historical facts and are instead based on our current expectations, estimates, and beliefs. The accuracy of such statements involves risks and uncertainties and depends upon future events, including those that may be beyond our control, and actual results may differ materially and adversely from those anticipated or implied by such statements, including, among others: potential development or distribution delays; risks inherent in the adoption of new agentic AI technologies and multi-tenant infrastructure solutions; the effectiveness of security and governance frameworks; our ability to successfully launch and gain market acceptance for new partner programs and service delivery models; and evolving industry standards for AI-driven automation. Any forward-looking statements included speak only as of the date hereof and, except as required by law, we assume no obligation to update or otherwise revise any such forward-looking statements to reflect subsequent events or circumstances. Customer statements and

reports on results, benefits, savings or other outcomes depend on a variety of factors including their use case, individual requirements, and operating environments, and should not be construed to be a promise or obligation to deliver specific outcomes. Certain information contained in this content may link or refer to, or be based on, studies, publications, surveys, and other data obtained from third-party sources. While we believe these third-party studies, publications, surveys, and other third-party data are reliable as of the date of publication, they have not independently verified unless specifically stated, and we make no representation as to the adequacy, fairness, accuracy, or completeness of any information obtained from a third-party. Our decision to publish, link to or reference third-party data should not be considered an endorsement of any such content.

Media Contact: Ken Lotich pr@nutanix.com