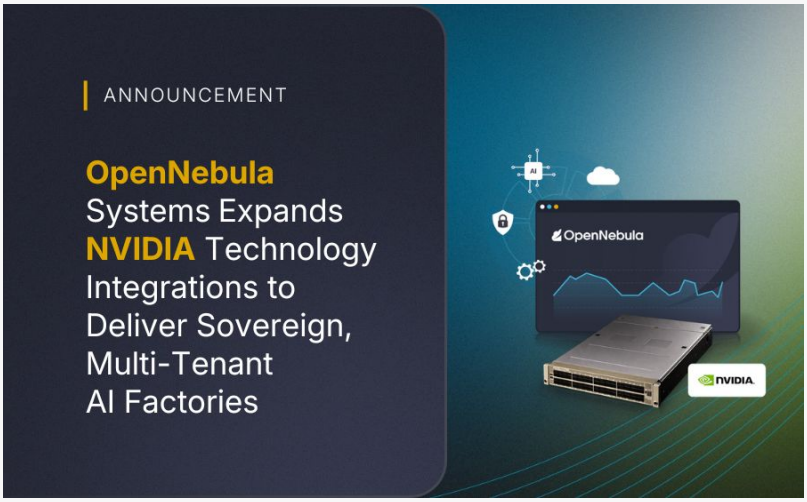
OpenNebula Systems Expands NVIDIA Technology Integrations to Deliver Sovereign, Multi-Tenant AI Factories

Demonstrations at NVIDIA GTC will showcase end-to-end automation from bare metal to production-ready AI cloud services.



OpenNebula Systems Expands NVIDIA Technology Integrations

SAN JOSE, CA, UNITED STATES, March 17, 2026 /EINPresswire.com/ -- OpenNebula Systems has expanded its integration with NVIDIA technologies, enabling organizations to deploy secure, production-ready AI Factories at scale. By combining OpenNebula's cloud management and virtualization platform with NVIDIA accelerated computing, networking, and bare-metal lifecycle management solutions, enterprises, HPC centers, and neocloud providers can deliver high-performance AI infrastructure with full tenant isolation and automated lifecycle control.

The integrations cover GPU virtualization, network offload, and automated bare-metal provisioning. OpenNebula supports NVIDIA GB200 NVL4 GPUs via PCI passthrough, providing virtual machines with direct hardware access and preserving native performance. Multi-Instance GPU (MIG) support enables hardware-level partitioning, allowing multiple tenants to share accelerators securely while maintaining predictable resource allocation.

NVIDIA BlueField DPUs are fully managed within OpenNebula's control plane, offloading network and security tasks from host CPUs. Per-tenant traffic enforcement, containerized network functions, and programmable switching policies reduce CPU load, strengthen isolation, and improve throughput across AI, Telco, and edge environments.

Integration with NVIDIA NCX Infra Controller automates provisioning, OS deployment, and environment configuration for GPU servers, creating isolated AI Factory instances on demand. This unified workflow combines lifecycle automation with strong governance, supporting rapid

scaling and consistent operations for enterprise, HPC, and sovereign AI initiatives.

“AI factories require tight integration across compute, networking, and infrastructure automation,” said Warren Barkely, VP of DGX Cloud, NVIDIA. “OpenNebula’s work with NVIDIA technologies, including NCX Infra Controller, helps organizations build scalable, multi-tenant AI environments with greater automation and operational control”.

“Our focus is to provide a unified control plane for AI infrastructure that preserves hardware performance while introducing strong governance and lifecycle automation,” said Ignacio M. Llorente, Managing Director at OpenNebula Systems. “Through deep integration with NVIDIA technologies, we enable organizations to deploy AI Factories that meet enterprise, HPC, and sovereign requirements with predictable performance and operational control.”

At NVIDIA GTC, OpenNebula Systems will showcase live demonstrations at Booth 7031, South Market Lot, highlighting secure multi-tenancy for GPU- and AI-as-a-Service workloads, DPU-based network offload and data-plane acceleration, NVIDIA Spectrum-X Ethernet networking integration validated on NVIDIA BlueField DSX Air, and end-to-end AI Factory deployment using NVIDIA NCX Infra Controller. These demos illustrate how AI infrastructure can be deployed from bare metal to fully operational, multi-tenant AI cloud services while maintaining strict performance, isolation, and governance standards.

About OpenNebula Systems

OpenNebula Systems is the developer of OpenNebula, an open and vendor-neutral cloud management and virtualization platform for private, hybrid, and edge infrastructures. The company provides SLA-based support and professional services to enterprises, service providers, and public-sector organizations worldwide.

OpenNebula is increasingly adopted as a VMware replacement platform, enabling organizations to modernize and consolidate their infrastructure while preserving existing hardware investments and operational practices. At the same time, the platform has evolved to support AI Factories and AI Gigafactories, acting as a unified control plane for GPU-accelerated infrastructure, Kubernetes platforms, and large-scale AI workloads.

Focused on simplicity, flexibility, and long-term digital sovereignty, OpenNebula combines cloud-native agility with the control and predictability of on-premises environments. It enables secure multi-tenancy, high-performance segmentation of AI infrastructure, and federation across multiple sites, supporting both enterprise and service-provider operating models.

With a global presence and offices in Europe and the United States, OpenNebula Systems is a trusted partner for large-scale, mission-critical deployments, supporting more than 5,000 cloud deployments worldwide and infrastructures that scale to thousands of hosts and tens of thousands of GPUs.

Anastasiia Rachkova
OpenNebula Systems
community-manager@opennebula.io
Visit us on social media:
LinkedIn
Bluesky
Facebook
YouTube
X
