OpenNebula

From HPC Centers to AI Factories:

Operating the Next Generation of Scientific Infrastructure

Elastic, Multi-Tenant Infrastructure for HPC & AI

IBERGRID 2026

OpenNebula.io

Who Controls the Future of Your Infrastructure?

What Is Broken?

Public Cloud Dependency

“Dependency is easy to create…but very hard to escape”

Concentration

3 hyperscalers control **~70%** of the global cloud market

Source: Synergy Research Group

Dependency

US providers account for **~70–75%** of the European cloud market

Source: European Commission (Digital Strategy / Cloud)

Lock-In

60% say moving workloads between clouds is difficult

Source: Flexera

AI Reality Check

“AI doesn’t run where you want…it runs where the infrastructure exists.”

Concentration

Hyperscalers control the **majority** of global AI compute capacity

Source: Synergy Research Group

Readiness Gap (Europe)

Only **~20–25%** of EU data centers are AI-ready

Source: European Commission / IDC Europe

Investment Gap

US invests **~10×** more in AI infrastructure than Europe

Source: McKinsey Global Institute

Full Software Stack Dependency

“This is not just cloud—we do not control the software stack”

Infrastructure Dependency

European data centers rely on **~80–90%** non-EU software technology

Source: Synergy Research Group

Enterprise Dependency

More than 80% of enterprise software used in Europe is developed outside Europe

Source: European Commission / IDC Europe

Governance Dependency

Over 80% of key open source projects are governed outside Europe

Source: McKinsey Global Institute

What Are the Risks?

External Control. Real Consequences

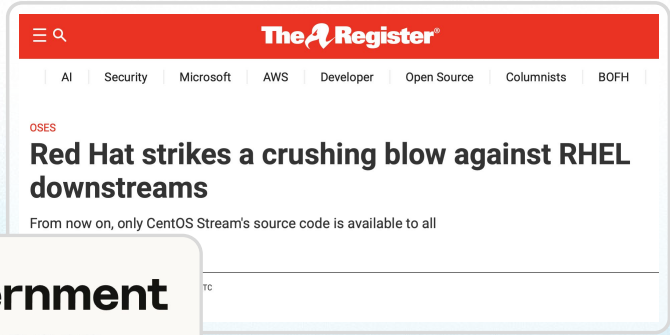
How foreign government decisions and vendor actions can disrupt critical services



IMPOSING SANCTIONS ON THE INTERNATIONAL CRIMINAL COURT

The White House | February 6, 2025

By the authority vested in me as President by the Constitution and the law
States of America, including the International Emergency Economic Powe
1701 *et seq.*) (IEEPA), the National Emergencies Act (50 U.S.C. 1601 *et seq*
212(f) of the Immigration and Nationality Act of 1952 (8 U.S.C. 1182(f)), and
of title 3, United States Code,



The Register

AI | Security | Microsoft | AWS | Developer | Open Source | Columnists | BOFH

OSES

Red Hat strikes a crushing blow against RHEL downstreams

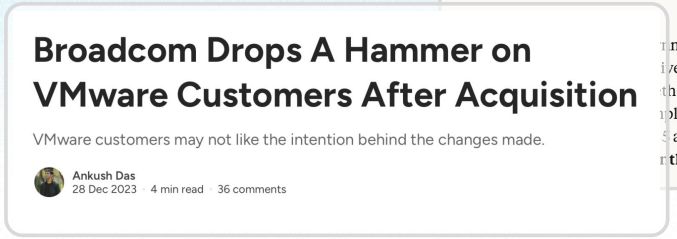
From now on, only CentOS Stream's source code is available to all



Statement on the US government directive to suspend access to Fable 5 and Mythos 5

12 Jun 2026

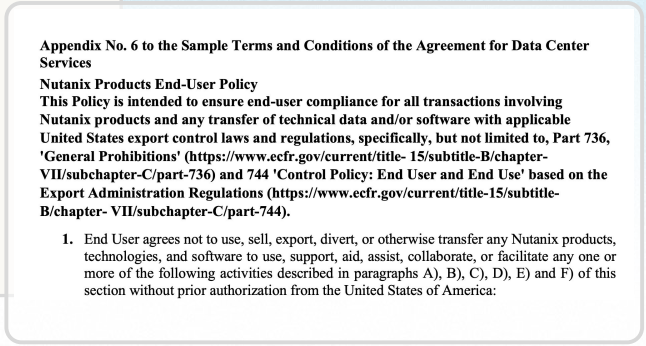
ment, citing national security authorities, has issued an export
ve to suspend all access to Fable 5 and Mythos 5 by any foreign
her inside or outside the United States, including foreign national
loyees. The net effect of this order is that we must abruptly
and Mythos 5 for **all** our customers to ensure compliance. **Access**
thropic models will not be affected.



Broadcom Drops A Hammer on VMware Customers After Acquisition

VMware customers may not like the intention behind the changes made.

Ankush Das
28 Dec 2023 · 4 min read · 36 comments



Appendix No. 6 to the Sample Terms and Conditions of the Agreement for Data Center Services

Nutanix Products End-User Policy

This Policy is intended to ensure end-user compliance for all transactions involving Nutanix products and any transfer of technical data and/or software with applicable United States export control laws and regulations, specifically, but not limited to, Part 736, 'General Prohibitions' (https://www.ecfr.gov/current/title- 15/subtitle-B/chapter-VII/subchapter-C/part-736) and 744 'Control Policy: End User and End Use' based on the Export Administration Regulations (https://www.ecfr.gov/current/title-15/subtitle-B/chapter- VII/subchapter-C/part-744).

1. End User agrees not to use, sell, export, divert, or otherwise transfer any Nutanix products, technologies, and software to use, support, aid, assist, collaborate, or facilitate any one or more of the following activities described in paragraphs A), B), C), D), E) and F) of this section without prior authorization from the United States of America:

Infrastructure Platform Sovereignty Risks

Sovereignty is ultimately about who retains control when circumstances change.



Economic Control

Costs and commercial dependence

- **Vendor lock-in**
  Migration becomes costly and complex.
- **Licensing risk**
  The vendor can change prices or terms unilaterally.



Technology Control

Platform and innovation dependency

- **Support dependency**
  Access to fixes and expertise depends on the vendor.
- **Roadmap dependency**
  Platform evolution follows the vendor's priorities.
- **Ecosystem risk**
  Changes in APIs, certifications or partner programs can have wide impact.



Legal Control

Governance and regulatory exposure

- **Governance risk**
  Strategic decisions can be taken outside European control.
- **Export-control risk**
  U.S. regulations may restrict access to software, updates, AI capabilities or support.



Operational Control

Operations and supply chain dependence

- **Supply-chain risk**
  Critical components depend on a single foreign provider.
- **Operational risk**
  Infrastructure may be hosted in Europe, but key control mechanisms remain under foreign authority.



Sovereignty is ultimately about who retains control when circumstances change.

What Is Changing?

Sovereignty Becomes Real

“Sovereignty is no longer a discussion—it’s an industry and regulation requirement”

Regulation & Policy



Strategy Adoption

75% of enterprises outside the US will have digital sovereignty strategies by 2030

Source: World Economic Forum / IDC

Active Demand

63% are already increasing adoption of sovereign cloud

Source: Sumo Logic / ITPro

Market Growth

Sovereign cloud market growing at **~25% CAGR**

Source: Fortune Business Insights

Cloud and AI Development Act

European Commission

Defining cloud and AI sovereignty

The Cloud and AI Development Act defines cloud and AI sovereignty comprising four assurance levels, to be used by public sector bodies based on their risk assessments. Cloud service providers can be recognized under this framework by Member States, after undergoing an audit.

The **four levels** are:

- **Level 1:** where data is processed and stored in infrastructure located in the Union;
- **Level 2:** where providers must demonstrate independence from third countries and transparency over their software supply chain;
- **Level 3:** where providers must be owned and controlled from the EU and meet additional criteria, such as personnel citizenship. The Commission can recognize third-country providers;
- **Level 4:** where providers have full transparency and control over their software supply chain and no interference from a third country.


EUROPEAN COMMISSION

Brussels, 3.6.2026
COM(2026) 502 final

2026/0138 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

establishing a framework of measures for strengthening Europe's cloud and AI ecosystem (Cloud and AI Development Act)

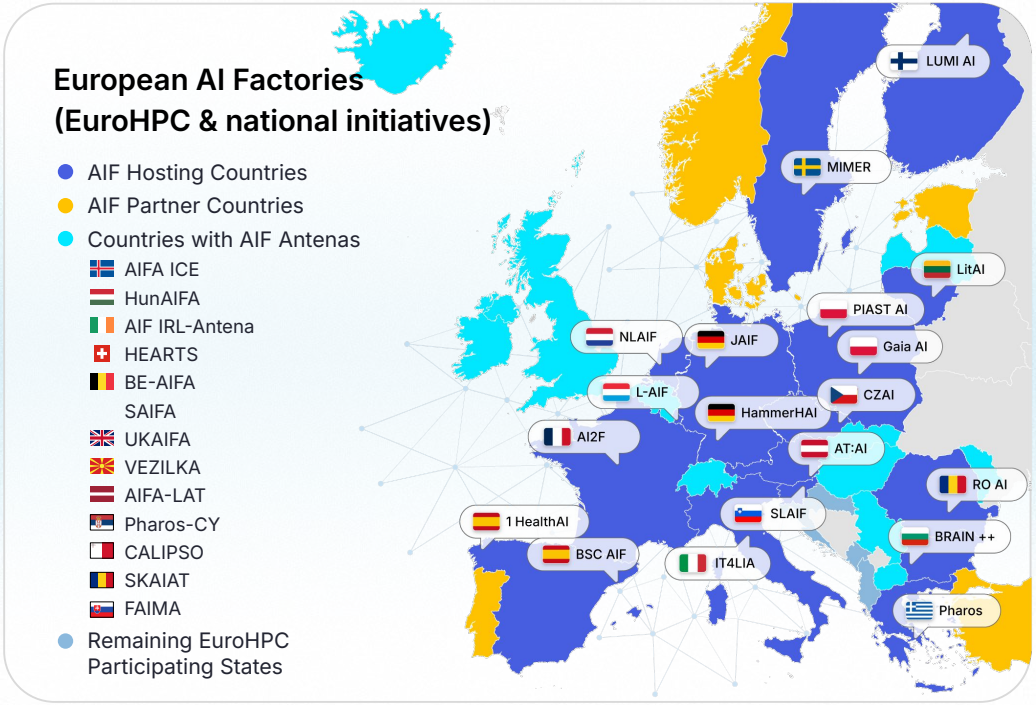
{SEC(2026) 502 final} - {SWD(2026) 502 final} - {SWD(2026) 503 final}

(Text with EEA relevance)

EN EN

Europe Is Moving From Debate to Deployment

Sovereignty is no longer a concept. It is becoming infrastructure



✕ Problem 1 - Scale Gap

Europe is investing billions in AI Factories...

but hyperscale AI infrastructure requires tens of billions per deployment.

✕ Problem 2 - Control Gap

Many AI factories are built on proprietary platforms...

...recreating dependency instead of sovereignty.

EuroHPC AI Gigafactories

A new generation of large-scale AI infrastructure for Europe

- **Large-scale AI infrastructure**
  Develop, train and deploy next-generation AI models.
- **Public-private partnerships**
  Industry-led consortia combining EU and Member States funding with significant private investment.
- **Scientific and industrial impact**
  Supporting research, innovation, startups and European industry.
- **European AI ecosystem**
  Connecting compute, data, talent and services across Europe.
- **Technological sovereignty**
  Strengthening Europe's capacity to develop and operate advanced AI infrastructure.

> €30 billion
total investment

Up to **€10 billion** EU + Member States and at least **€20 billion** private investment.

Up to 7
AI Gigafactories

4 in Lot 1
(large scale)
3 in Lot 2
(medium scale).

19
AI Gigafactories

in the existing **EuroHPC network** (supporting SMEs, research and industry).

2027
Expected awards

Selected projects expected in early 2027, operational within **18 months of contract signature**.

Two deployment lots with different scales

Lot 1 – Large scale | 4 AI Gigafactories

- Larger-scale infrastructure
- Higher computing capacity
- For training and operation of next-generation AI models

Lot 2 – Medium scale | 3 AI Gigafactories

- Significant computing capacity
- Focus on a broad range of AI use cases
- Strong regional impact

What to Do?

Basic Principles for Next Generation Infrastructure

“We are entering a new phase. Where infrastructure is not just technical…it’s strategic”



Principle 1

Design for control—not convenience



Principle 2

Build open—not dependent



Principle 3

Think distributed—not centralized

This is not cloud evolution. It’s a shift in control.

What Is OpenNebula?

*An Open Cloud Management Platform—**Established in 2010***

1 Kubernetes Management Layer

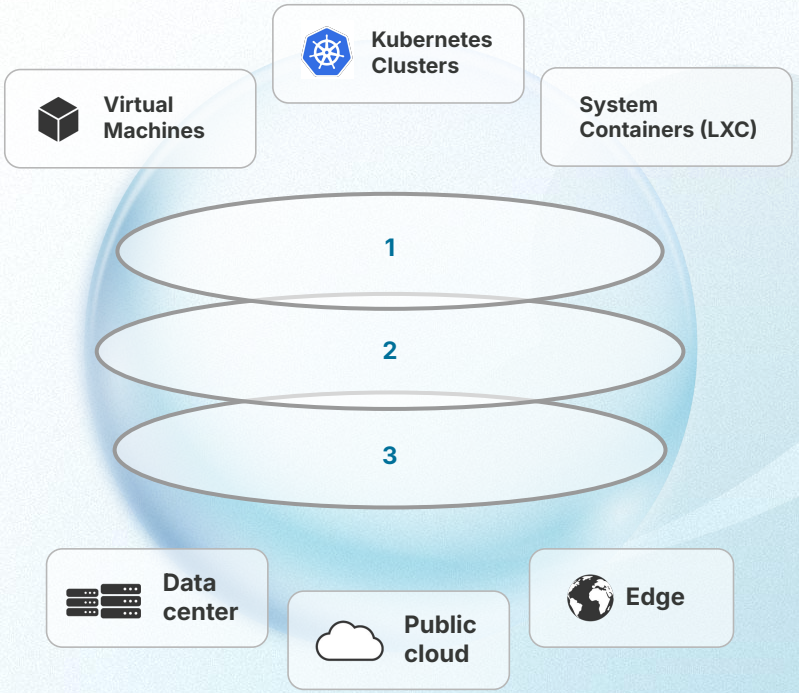
- Kubernetes lifecycle and orchestration
- Autoscaling and resource optimization
- Multi-cluster policy control
- Workload multi-tenancy

2 Cloud Orchestration Layer

- Physical resource provisioning
- Self-service and automation
- Elasticity and service catalog
- Lifecycle management

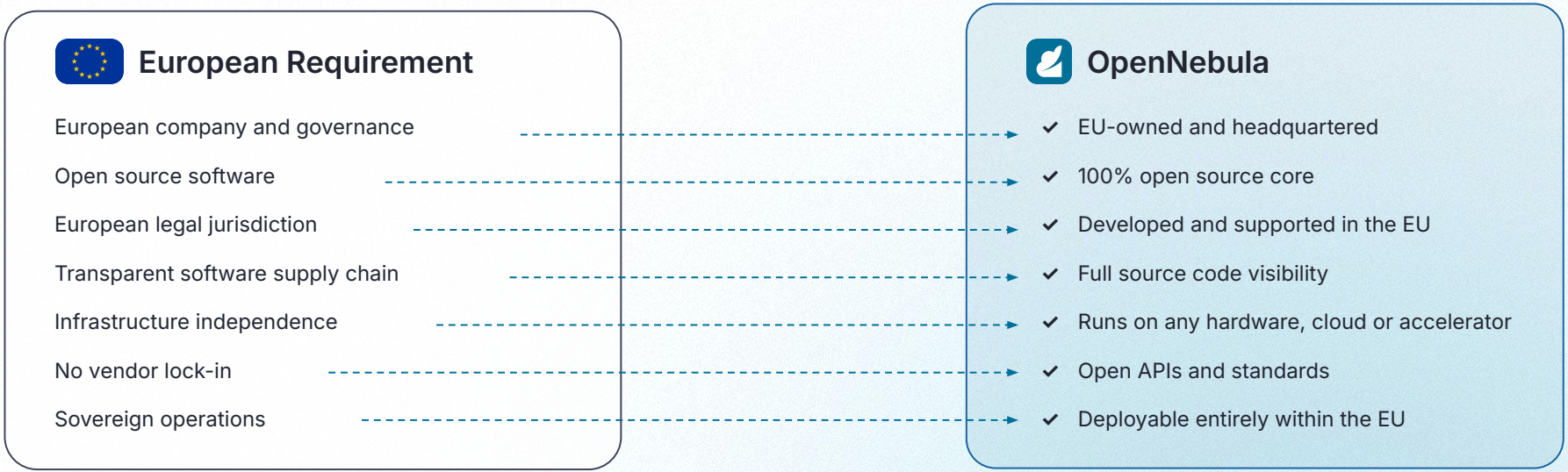
3 Virtual Infrastructure Management Layer

- VM, network, and storage orchestration
- Monitoring and scheduling
- Infrastructure optimization



European Sovereignty by Design

A European-Controlled, Open Infrastructure Control Plane



Enabling Higher Sovereignty Levels under the New Cloud & AI Development Act (CADA)

Level 3

European ownership, governance and operational control for sovereign cloud and AI infrastructure.

Level 4

Complete software supply-chain transparency and freedom from third-country control through open source software and European governance.

European IPCEIs: From Cloud Infrastructure to AI

Strategic European industrial initiatives to develop key digital technologies

IPCEI-Cloud (CIS)

Next-generation European cloud-edge infrastructure and services.

2023 - 2031

Approved. Implementation underway.

€2.6 billion
total investment

Up to **€1.2 billion** public funding
€1.4 billion expected private investment

- ✓ **Sovereign cloud infrastructure**
  European-controlled cloud and edge
- ✓ **Interoperability and open standards**
  Multi-provider ecosystem
- ✓ **European industrial ecosystem**
  Strengthening Europe's supply chain



IPCEI-AI

European AI technologies and infrastructure.

From 2027

In preparation. Subject to approval.

More than €10 billion
expected total investment

EU + Member States + private
Public/private breakdown to be confirmed

- ✓ **European AI technologies**
  From AI computing to applications
- ✓ **Innovation and deployment**
  Bridging research and real-world use cases
- ✓ **Strategic European value**
  Strengthening Europe's capabilities across the AI stack



OpenNebula

Direct Participant in IPCEI-Cloud and IPCEI-AI

Developing European cloud and AI technology through strategic industrial collaboration.

What Problems We Solve

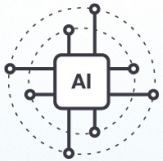
“We are not building another cloud. We are enabling the next generation cloud.”



VMware Lock-In

→ **Revirtualization**

- ✓ Co-exit VMware safely
- ✓ Migrate without downtime
- ✓ Preserve your ecosystem



AI Factories

→ **Enablement**

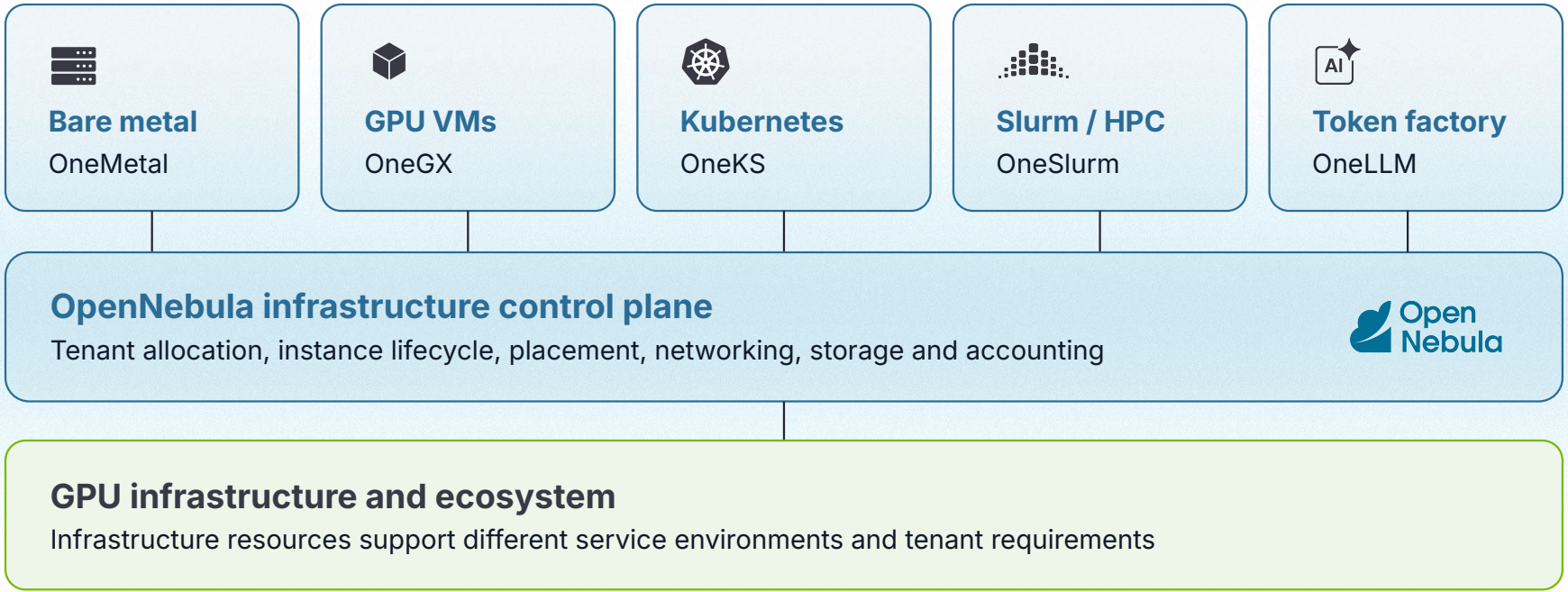
- ✓ Multi-tenant GPU at scale
- ✓ Open AI stack integration
- ✓ Production-ready AI infrastructure

We are trusted by Leading Organizations:

AIRBUS · cewe · Nintendo · ACTIVISION · MX · encore TECHNOLOGIES · BlackBerry · Qt Group · Dustin · Every Matrix

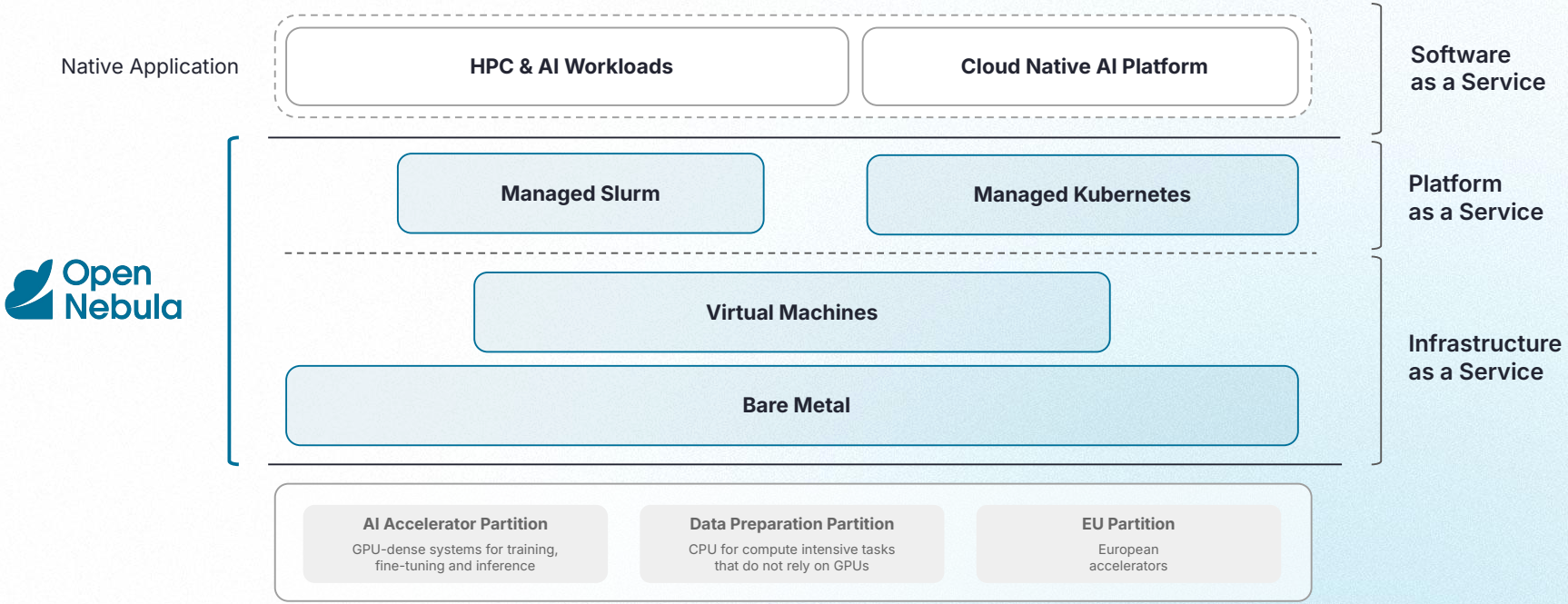
Operational Models for HPC AI Factories

A cloud operating model for infrastructure and AI service delivery



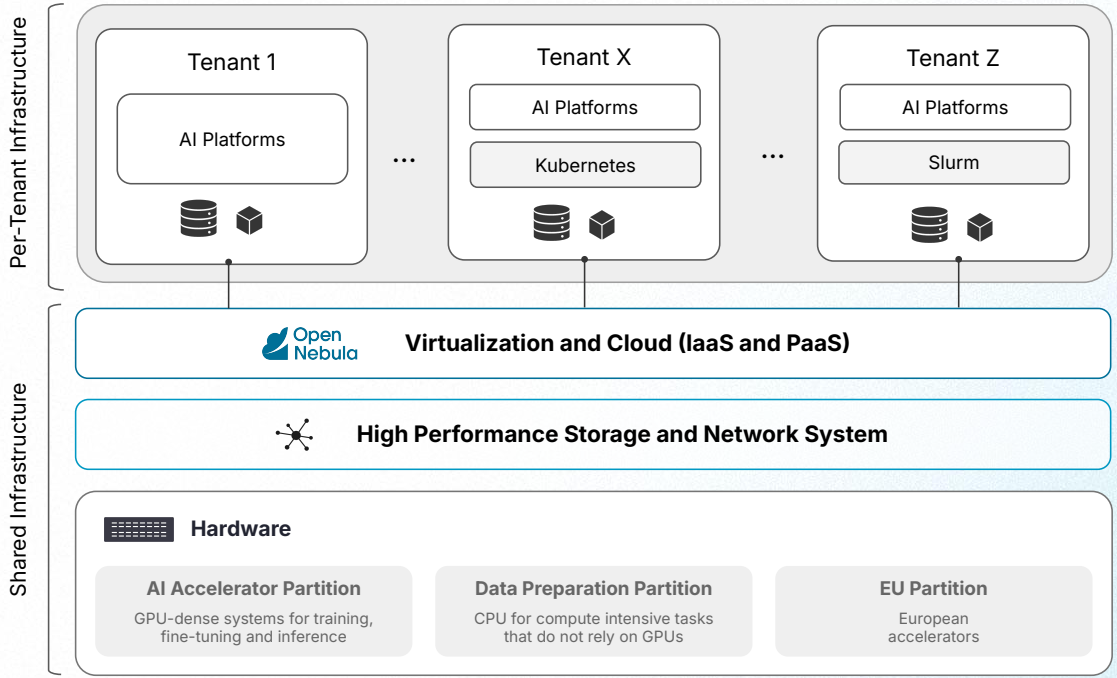
AI Factory Reference Architecture

From bare metal to HPC & AI services: Slurm → GPUaaS → K8SaaS → LLMaaS



Secure Multi-Tenancy

Isolated HPC & AI environments on shared infrastructure



Access Management

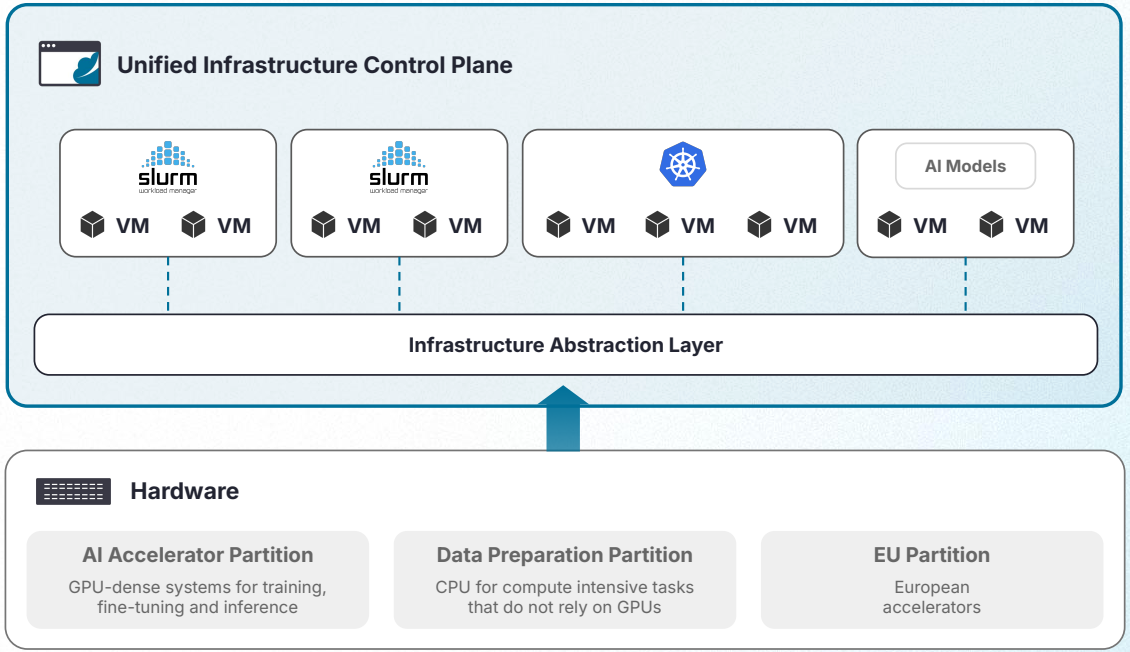
- **Users & Groups:** Simplifies access control by grouping users.
- **Permissions:** Defines access rights per resource for owner, group, and others.
- **ACLs (Access Control Lists):** Provides granular, user- and group-based access control to resources.

Infrastructure & Resource Management

- **Accounting & Showback:** Tracks resource usage by user and group.
- **Isolated Networking:** Ensures secure, tenant-isolated network traffic.
- **Virtual Data Centers:** Allocates physical resources into logical units for users and groups.

Dynamic Capacity Management

Multi-tenant Slurm, Kubernetes, and VMs on the same infrastructure



Hybrid HPC–Cloud Experience

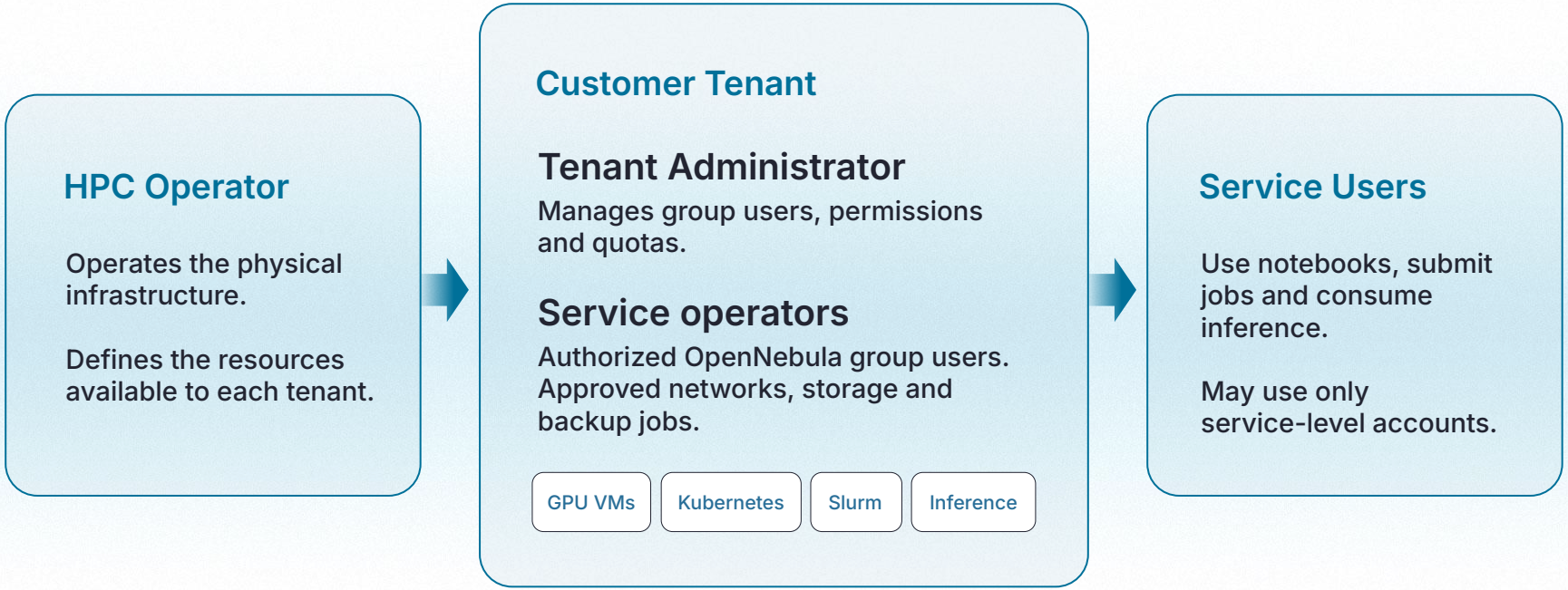
- HPC performance with cloud flexibility.
- On-demand GPU, VM, and Kubernetes environments.
- Supports both batch HPC and interactive AI workloads.

Unified Platform (No Silos)

- Slurm, Kubernetes, and Virtual Machines on the same infrastructure.
- Consistent policies, quotas and governance.
- Higher utilization with lower operational complexity.

Delegate Infrastructure to Tenant Teams

Give customers an infrastructure allocation they can manage themselves



VDCs define resource access. Group permissions define who operates it.

Combine Dedicated and Shared Infrastructure

Match infrastructure consumption to customer isolation, performance and cost requirements

Dedicated infrastructure

Physical GPU servers
for exclusive hardware access

Customer operates its environment

Shared GPU capacity

GPU VMs for development
and enterprise workloads

Provider applies allocation policies

Managed environments

Kubernetes, Slurm
and inference services

Service teams operate the offering

Common tenant allocation and infrastructure governance

Capacity, networks and storage. Workload schedulers manage jobs inside environments

Bare Metal Performance with Cloud Agility

PCI passthrough combines native GPU performance with virtualization flexibility and security.

OpenNebula → **The Best of Both Worlds (Bare Metal Performance + Cloud Agility + Secure Multi-Tenancy)**

1 NVIDIA Recommended Architecture

NVIDIA NCP Software Reference Guide (November 2025)

> *"The IaaS model should evolve toward a virtual machine-based architecture rather than relying on bare metal deployments."*

Why Virtual Machines?

- **Multi-Tenant & Better GPU Utilization**
  Efficiently share GPUs across tenants and workloads.
- **Strong Isolation**
  Ensures secure separation between tenants and workloads.
- **Faster Provisioning & Simpler Operations**
  Deliver new tenants and machines in minutes.

+ 2 Bare Metal Performance



Based on IBM studies and OpenNebula validation on NVIDIA GB200 following the NVIDIA Grace I/O Virtualization Guide.

Optimized for AI Workloads

- PCI Passthrough
- CPU Pinning
- Huge Pages
- NUMA Awareness
- SR-IOV
- NVLink/ NVSwitch

Bare Metal as a Service

Virtual machines by default. Bare metal when required

Alternative Option: NVIDIA Infrastructure Controller Integration

3 Integration with NVIDIA Infrastructure Controller (NICo)

NVIDIA Infrastructure

- Automatic discovery, telemetry, and lifecycle management of NVIDIA infrastructure.

Bare Metal Servers

- OpenNebula integrates with NVIDIA Infrastructure Controller to provision bare metal servers when required.

Use of bare metal only when the workload explicitly requires it.



Sovereign ≠ Isolated

European control. Open infrastructure. Global technology

European Control

European Infrastructure Ecosystem

AXELERA · ear · MULTIVERSE · NOKIA · Open Nebula · semidynamics · SUSE

- European Infrastructure Providers
- EU Data Centers
- European Software Ecosystem
- Compliance with EU Regulations

Sovereign Infrastructure Control Plane

One Platform. Full Control

Open Nebula

- Multi-Tenancy
- Automation
- Isolation & Security
- Monitoring & Operations
- Open APIs & Standards

Global Technology and Innovation

Best-in-Class Technologies Integrated Under Your Control

AMD · Canonical · Everpure · intel. · NetApp · NVIDIA · Red Hat · SaturnCloud · VAST

Open, Interoperable, Sovereign Infrastructure. Combine European and global technologies under your control

- **Freedom of choice:** Integrate and replace technologies as needed.
- **No lock-in:** Open standards, open APIs.
- **European control:** Deploy and operate in the EU.
- **Production ready:** Validated with global technology leaders.

NVIDIA-Validated AI Cloud Platform

One of the six AI cloud ISV platforms officially validated by NVIDIA

NVIDIA AI Cloud Ready Validated Partners

The NVIDIA AI Cloud Ready validation initiative brings together infrastructure and platform software providers—aligning their solutions with NVIDIA's AI infrastructure reference architecture—to provide infrastructure and platform capabilities that support scalable AI factory deployments on AI clouds.

MIRANTIS

Open Nebula

RAFAY

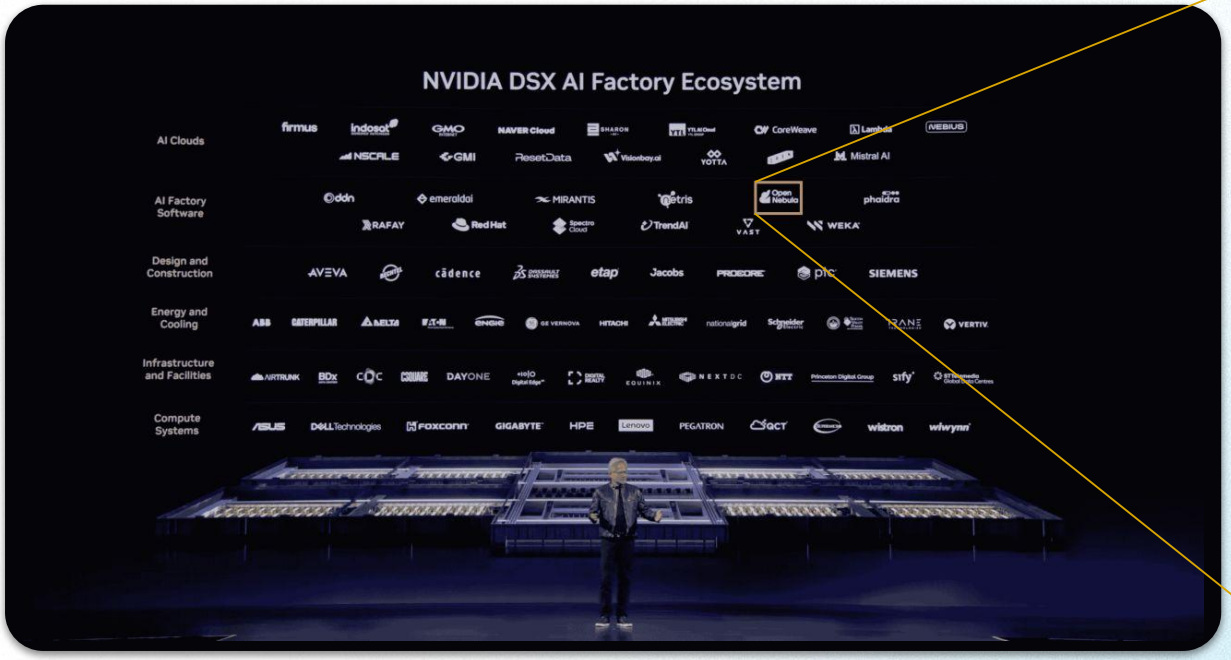
Red Hat

Spectro Cloud

zadara

NVIDIA-Validated AI Cloud Platform

Validated to work with the key NVIDIA hardware and software building blocks



Open Nebula

NVIDIA Validated:

- ✔ GB200/GB300
- ✔ Spectrum-X
- ✔ Infiniband
- ✔ DPU
- ✔ Run:AI
- ✔ NICo

Vendor-Neutral AI Infrastructure Ecosystem

Choose the infrastructure technologies that best fit your AI factory

Compute Platforms

Support the latest CPU, GPU and accelerator technologies.

CPUs

- Intel Xeon, AMD EPYC
- NVIDIA Grace
- NVIDIA Vera

GPUs

- NVIDIA Ampere (A100)
- NVIDIA Hopper (H100/H200)
- **NVIDIA Blackwell (B200/B300)**
- NVIDIA Rubin (Vera Rubin platform)
- **AMD Instinct (MI300X/MI325X)**
- **Axelera AI (Europa AIPU)**

High-Speed Interconnects

- **NVLink and NVSwitch**



> Upcoming Integrations

- *Semidynamics*
- *Fujitsu MONAKA*

Networking Infrastructure

Support modern AI fabrics and accelerated networking.

AI Fabrics

- NVIDIA Spectrum-X Ethernet
- **NVIDIA InfiniBand**

Smart Networking

- NVIDIA BlueField DPUs



> Upcoming Integrations

- *Kalray DPUs*
- *Nokia Network Fabric*

Storage Infrastructure

Integrate with enterprise and AI-optimized storage platforms.

Storage Infrastructure

- **NetApp ONTAP**
- Everpure
- **VAST Data**
- **WEKA**
- IBM Storage Scale (GPFS)
- **Lustre (Parallel File System)**
- Ceph
- NFS / NAS
- SAN: iSCSI, Fibre Channel, NVMe-oF
- Local NVMe / SSD



> Upcoming Integrations

- *Onedata*
- *MinIO*

Vendor-Neutral AI Software Ecosystem

Choose the software ecosystem that best fits your AI factory

Infrastructure Foundation

Built on open standards.

Operating Systems

- Red Hat Enterprise Linux
- Ubuntu Pro
- SUSE Linux Enterprise

Observability

- **Prometheus, Grafana**
- Loki, Tempo, Open Telemetry

Infrastructure as Code

- **Terraform & Ansible**

Identity

- LDAP / AD, OIDC, SAML, MyAccessID

EuroHPC

- **Waldur, Open OnDemand, EESSI**

Kubernetes & HPC

Choose the orchestration layer that best fits each workload.

Kubernetes Platforms:

- **SUSE RKE2**
- Red Hat OpenShift
- Canonical Kubernetes

Schedulers:

- NVIDIA Run:AI / KAI
- **Slurm**

GPU Management:

- NVIDIA GPU Operator
- NVIDIA Network Operator



AI Platforms & Inference

Deploy any AI platform without infrastructure lock-in.

AI Platforms

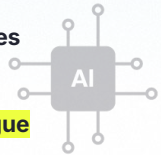
- OpenShift AI
- SUSE AI
- ClearML
- Flower AI
- Kubeflow

AI Inference & Serving

- NVIDIA Dynamo
- NVIDIA NIM
- vLLM
- **KServe**
- **llm-d**

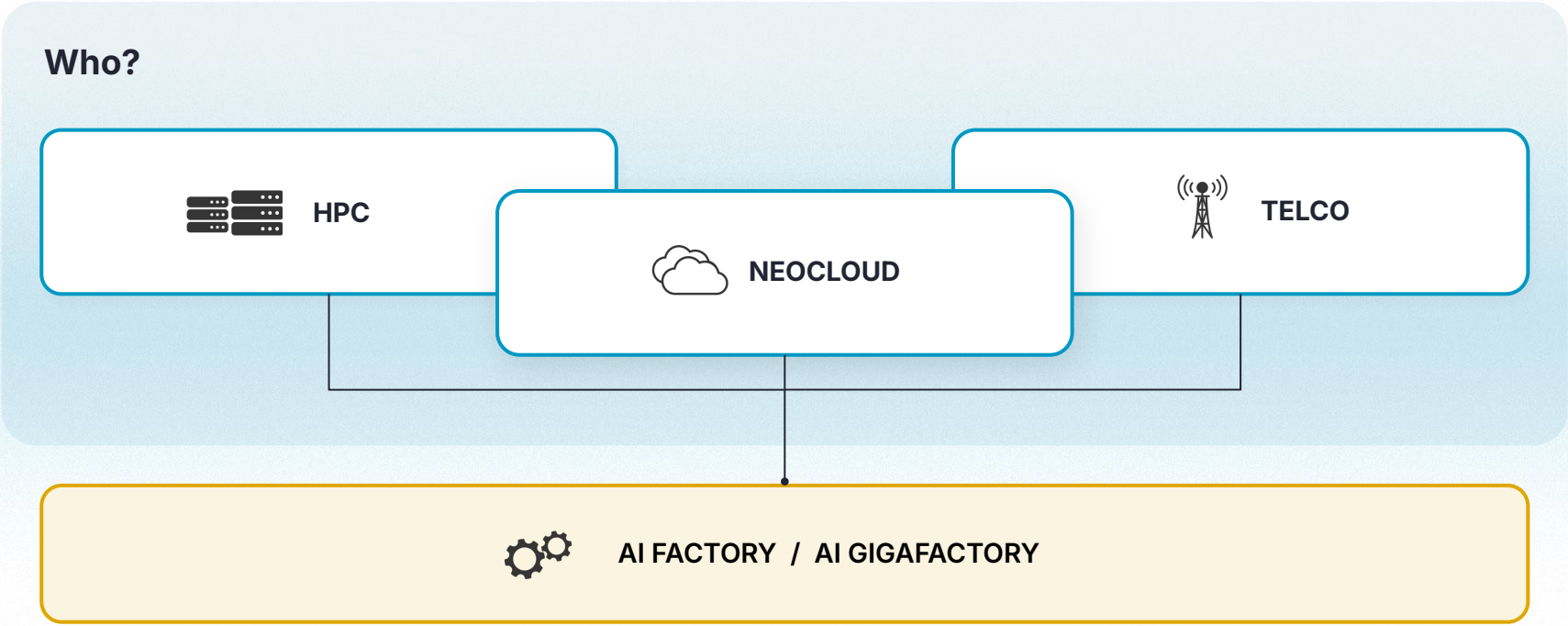
Model & Services Catalogues

- **Hugging Face**
- NVIDIA NIM Catalog
- **OneLLM Private Catalogue**



Different Operators. Converging AI Infrastructure Requirements

Bringing specific technological and sovereignty requirements



Building an AI Factory: Delivering GPUaaS and LLMaaS for Hybrid HPC-Cloud Environments

- Deliver GPU, Kubernetes, Slurm and LLM services on one platform
- Run HPC, cloud-native, and AI workloads on shared infrastructure
- Optimize GPU utilization through orchestration and automation

Meet us at our Tutorial

16:00 h - 18:00 h

OpenNebula Reference Architecture

From bare metal GPUs to cloud services: GPUaaS → K8SaaS → LLMaaS



▶ WATCH NOW

https://www.youtube.com/watch?v=le3Spyzvm8I



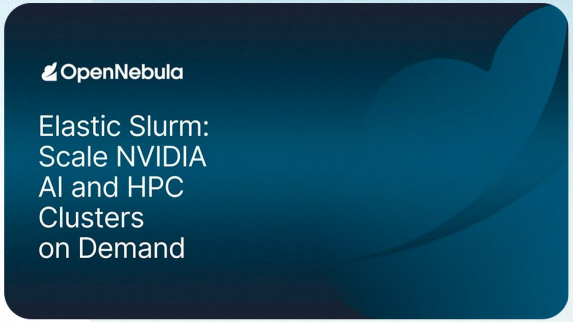
▶ WATCH NOW

https://www.youtube.com/watch?v=TGyRcV3XATc



▶ WATCH NOW

https://www.youtube.com/watch?v=wE9HBpPHpzU

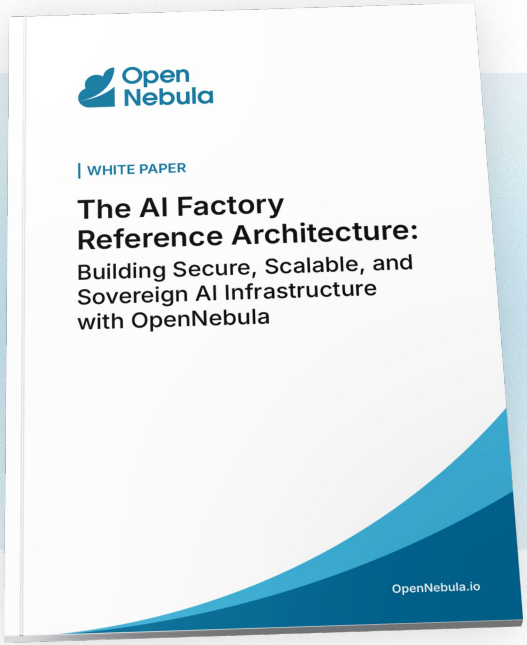


▶ WATCH NOW

https://www.youtube.com/watch?v=tI2ymex-YzM

White Paper: The AI Factory Reference Architecture

Building secure, scalable, and sovereign AI Infrastructure with OpenNebula



> DOWNLOAD NOW

Learn More About Sovereign AI Factories

Building secure, scalable, and sovereign AI infrastructure with OpenNebula



From Metal to a Production-Ready AI Cloud Service

› DISCOVER NOW

OpenNebula

OpenNebula.io

Thank You!

contact-eu@opennebula.io

+34 91 297 9741 (ES)

OpenNebula Systems Headquarters

EMEA

La Finca Business Park, Building 13
28223 Pozuelo de Alarcón, Madrid
Spain

USA

1500 District Avenue
Burlington, MA 01803
USA

OpenNebula Labs

Czech Republic

Cyrilská 7 – Impact Hub Brno
602 00 Brno
Czech Republic

Belgium

Brussels Manhattan Center, 5th Floor
Avenue du Boulevard 21, Brussels 1210
Belgium