Netris® NAAM Extends NVIDIA® Mission Control™ with Hard Multi-Tenancy

Ethernet Fabric Automation and Hard Multi-Tenancy for NVIDIA Mission Control Environments

Whitepaper

Abstract

NVIDIA® Mission Control™ is the integrated management platform for NVIDIA AI factories — combining Base Command Manager, Run:ai, UFM, and NetQ to manage compute, workloads, and InfiniBand. Netris® NAAM (Network Automation, Abstraction, and Multi-Tenancy) extends Mission Control with Ethernet fabric automation, VPC-based multi-tenancy, and hard isolation (enforced on networking hardware) — integrating with UFM, NMX-C, and NVIDIA DOCA™ so operators can deliver cloud-provider-grade GPU infrastructure in weeks instead of years.

4701 Patrick Henry Drive, Bldg. 25
Santa Clara, California, 95054, USA

Copyright Notice, Authorization to Use and Disclaimers

Legal Notice

Document History

Version	Date	Authors	Description of Change
1.0	07/01/2026	Andrey Khomyakov (Netris)	Initial Version

Executive Summary

This whitepaper is intended for AI infrastructure architects, cloud service providers, and GPU cloud operators evaluating multi-tenant networking for NVIDIA Mission Control environments.

NVIDIA Mission Control™ is the integrated AI factory management platform built on NVIDIA operational best practices. It includes NVIDIA Base Command Manager (BCM) for cluster and node lifecycle management, Run:ai for workload orchestration, NVIDIA Unified Fabric Manager (UFM) for NVIDIA Quantum™ InfiniBand fabric management, NVIDIA NetQ for network observability, and more. Together, these components deliver a unified control plane for the compute and workload layers of an AI factory. For AI factory operators deploying Spectrum-X™ East–West Ethernet fabrics alongside Mission Control, or seeking to accelerate time to revenue and deliver cloud-provider-grade multi-tenancy, the Netris NAAM (Network Automation, Abstraction, and Multi-Tenancy) platform adds a complementary Ethernet fabric management layer. Netris integrates with UFM for InfiniBand PKey management, with the NMX Controller (NMX-C) for automated NVLink™ partition orchestration on NVL72 systems, and with NVIDIA DOCA™ for BlueField® DPU provisioning — extending the Mission Control stack with Ethernet automation, VPC-based multi-tenancy, and hard isolation (enforced on networking hardware) across North–South Ethernet-based fabrics, NVIDIA Spectrum-X-based East-West fabrics, NVIDIA BlueField DPUs, and NVLink Multi-Node fabrics. Operators add Netris to their Mission Control environment without changing any existing Mission Control component.

In One Sentence

Netris adds automation for East-West and North-South Ethernet-based fabrics, automated NVLink partition management via NMX-C, BlueField DPU lifecycle management via NVIDIA DOCA, and VPC-based multi-tenancy with hard isolation (enforced on networking hardware) to NVIDIA Mission Control environments — integrating with UFM so network isolation remains consistent across InfiniBand, Ethernet, and NVLink from a single tenant intent.

Table of Contents

NVIDIA Mission Control: The AI Factory Management Platform

NVIDIA Mission Control (NMC) is NVIDIA's integrated AI factory management platform and is the required management stack for NVIDIA GB200 and GB300 NVL72 systems. NMC also supports B200 and B300 architectures, as well as Vera Rubin. It combines several NVIDIA software components and adds new capabilities exclusive to the platform:

- **Base Command Manager (BCM)** — cluster management and node provisioning: BCM bootstraps and manages the lifecycle of DGX™ nodes, handling OS imaging, inventory, and hardware health. BCM supports both Run:ai and Slurm workload environments, with Grafana dashboards providing unified observability across GPU, NVLink, power, and Slurm metrics.
- **Run:ai** — workload orchestration: Run:ai schedules and manages GPU workloads across shared compute clusters, including multi-tenant quota management via a Department/Project/Workload hierarchy.
- **NVIDIA Unified Fabric Manager (UFM)** — InfiniBand fabric management: UFM manages the NVIDIA Quantum InfiniBand fabric, handling topology discovery, port management, and PKey-based partition administration for East–West inter-GPU communication.
- **NVIDIA NetQ** — network and NVLink observability: NetQ provides telemetry and observability for Ethernet and NVLink fabrics. The NetQ NVLink component (previously NMX-M) runs on Admin Service Nodes and surfaces NVSwitch health, GPU domain status, and port metrics.
- **NVLink Partition Management** — NVL72 systems include the NMX Controller (NMX-C) running on the NVSwitch hardware, which manages NVLink partition lifecycle via the NVOS interface. Partitions split a 72-GPU NVLink domain into isolated GPU collectives, with Run:ai detecting the resulting partitions automatically via node labels.
- **LaunchPad** — a browser-based access gateway providing unified RBAC and SSO authentication into NMC components. Individual component UIs (BCM, Run:ai, UFM, NetQ) each retain independent interfaces.
- **Net-new platform capabilities** — including Autonomous Hardware Recovery (AHR), Autonomous Job Recovery (AJR), power management (Domain Power Service), BMS integration for liquid cooling and leak detection, and many more. Please, see NVIDIA Mission Control documentation for more information.

NMC delivers a unified operational experience for the compute, workload, and InfiniBand layers of an AI factory. It represents NVIDIA's vision for how DGX-based AI factories should

be managed — with standardized tooling, integrated best practices, and Autonomous Recovery capabilities that significantly reduce mean time to recovery.

Netris Extends NVIDIA Mission Control

Netris extends NMC with automated Ethernet fabric management including BlueField DPU management, VPC-based multi-tenancy, and hard isolation (enforced on networking hardware) on the North–South and Spectrum-X (East-West) fabrics — extending NMC deployments for operators who need cloud-provider-grade networking.

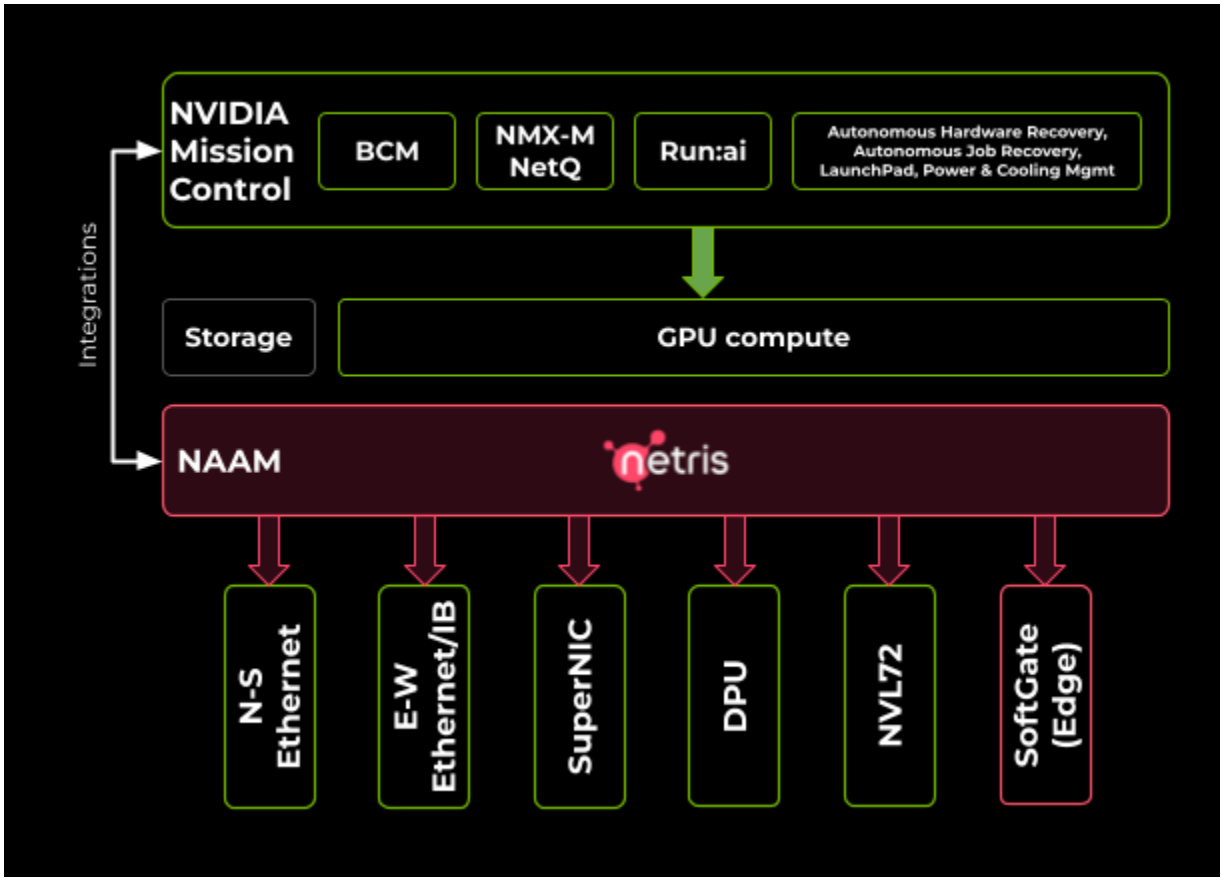

Figure 1. Netris and NVIDIA Mission Control combined architecture

Netris integrates directly with UFM — the InfiniBand fabric manager NMC includes — via the UFM API. This means Netris can manage InfiniBand PKey membership alongside Ethernet VRF and VXLAN configuration, keeping tenant boundaries consistent across both InfiniBand and Ethernet from a single high-level tenant definition. Operators define which servers belong to which tenant. Netris translates that intent into PKeys in UFM, VRFs/VXLANs on Ethernet switches, and NVL72 partitions simultaneously.

What Netris Adds to an NMC Environment

As AI cloud operators scale to hundreds or thousands of GPUs shared across tenants, dynamic hard isolation (enforced on networking hardware) — mandatory for securely sharing GPU resources across multiple customers — would not be operationally achievable through manual processes. Netris NAAM makes that isolation automatic:

- **Ethernet fabric automation and abstraction** — Netris functions as the fabric manager for North-South and Spectrum-X (East-West) Ethernet fabrics, automatically provisioning VRFs, VXLANs, BGP/EVPN, and switch port assignments. For Spectrum-X deployments, the Netris Host Networking (NHN) plugin applies required RoCEv2 settings on BlueField-3 SuperNICs, ensuring full compliance with the NVIDIA Spectrum-X Deployment Guide without manual per-node configuration.
- **VPC-based network multi-tenancy** — Netris provides cloud-provider-grade VPC constructs including VPC peering, elastic IPs, NAT, and Layer-4 load balancing, with hard isolation (enforced on networking hardware) at the switch and DPU level. This extends NMC's workload-layer tenancy to the physical network fabric.
- **Cross-fabric tenant lifecycle** — Netris Server Cluster feature simultaneously aligns Ethernet VRFs/VXLANs, InfiniBand PKeys (via UFM), and NVLink partitions (via NMX-C), coordinating tenant changes across all fabrics from one API call.
- **BlueField DPU management** — Netris manages BlueField DPUs configured in zero-trust mode as network fabric devices. Customers deploy the Netris DPU agent via NVIDIA DOCA Platform Framework (DPF) and Netris extends the EVPN fabric into the server adding concurrent multi-tenancy with hard isolation (enforced on networking hardware) at sub-server granularity. NMC monitors the overall infrastructure through NetQ while Netris handles the DPU's Host-Based Networking (HBN) configuration and network policy enforcement.

You can learn more about hard multi-tenancy (enforced on networking hardware) in *NVIDIA DGX-based AI Clouds by downloading our Multi-Tenancy in AI Clouds* whitepaper at https://netris.io/

Separation of Responsibilities

The division of responsibilities between NMC and Netris is clear.

- **NMC manages compute and workloads:** BCM provisions nodes, Run:ai schedules GPU workloads and enforces quota-based tenancy, UFM manages InfiniBand, and NetQ NVLink monitors NVLink partition health.
- **Netris manages the Ethernet networking layer, cross-fabric automation, multi-tenancy:** provisioning VRFs, VXLANs, and BGP/EVPN on Ethernet switches,

enforcing network-layer tenant isolation across switches and DPUs, orchestrating InfiniBand (PKeys) via UFM API and NVLink partitions via the NMX-C gRPC API, and coordinating all fabric changes simultaneously from a single tenant definition.

Together, NMC and Netris own the full stack: NMC owns compute orchestration and Netris owns network automation, abstraction, and hard isolation for true multi-tenancy.

Layer	NVIDIA Mission Control	Netris
Node provisioning	BCM	-
Workload management	Run:ai, Slurm	-
InfiniBand fabric management	UFM	PKey orchestration via UFM API
NVLink partition lifecycle (cross-fabric)	NMX-C via NVOS interface	GPU partition orchestration via NMX-C gRPC API
NVLink observability	NetQ NVLink	-
Spectrum-X Ethernet fabric management	-	Native Ethernet fabric management
North-South Ethernet automation	Planning worksheets	Automated VRF/VXLAN provisioning
Inventory Source-of-Truth	Compute and Network resources	Network resources and tenancy
Network-layer multi-tenancy (VPCs)	-	VPC abstraction, hard isolation
BlueField DPU lifecycle & segmentation	DPU observability	Netris DPU agent
Edge services (NAT, LB)	-	Netris SoftGate
Network observability & drift detection	NetQ	Topology validation & drift detection

Architecture: Multi-Tenant AI Cloud with Netris

The Netris NAAM (Network Automation, Abstraction, and Multi-Tenancy) platform implements multi-tenancy using three foundational practices:

1. **Define tenant intent once.** Tenant isolation is expressed as a high-level abstraction — the set of servers assigned to a tenant. Operators do not configure individual switches, PKeys, or NVLink partitions directly.
2. **Translate intent across fabrics.** Netris converts tenant intent into Ethernet VRFs/VXLANs, InfiniBand partitions (via UFM), NVLink GPU partitions (via NMX-C), and DPU policies — simultaneously and consistently.
3. **Continuously validate the isolation state.** Netris continuously validates topology and configuration state across Ethernet fabrics to ensure tenant boundaries remain consistent as infrastructure evolves.

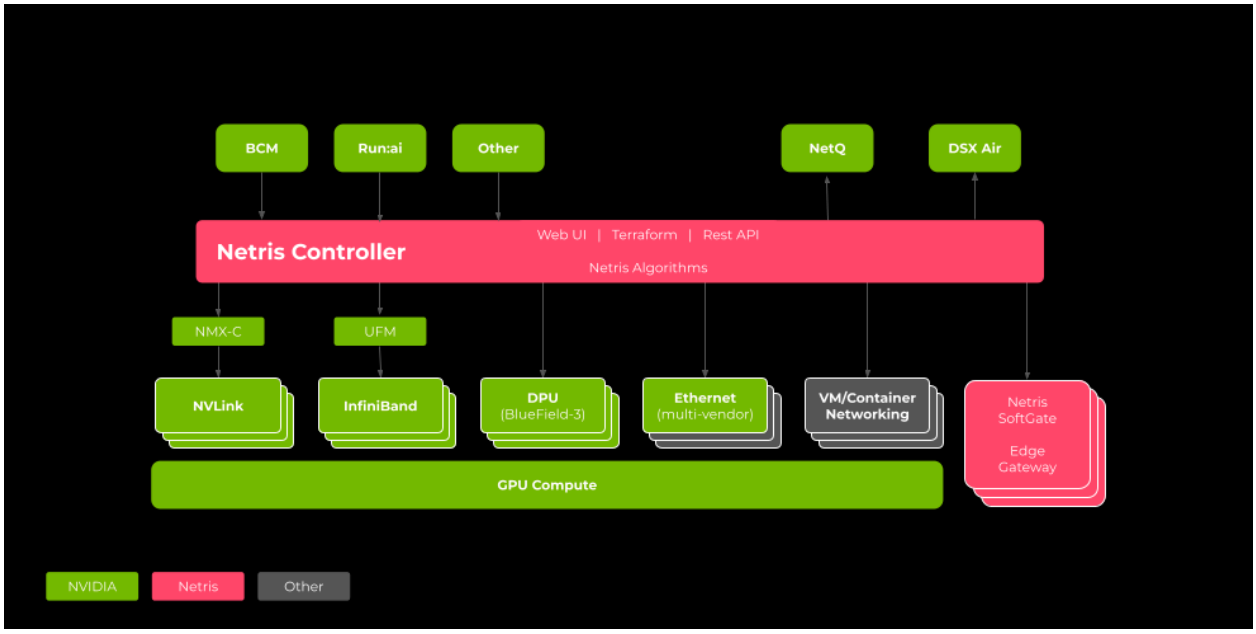

Figure 2. Detailed Netris and NVIDIA Mission Control Architecture

Netris sits between the upstream platform (IaaS/PaaS or an administrative portal) and the individual fabric control planes. The upstream platform interacts with Netris and Netris translates tenant intent to UFM, NMX-C, and the Ethernet fabric independently. This keeps the upstream platform decoupled from individual fabric controllers and ensures changes are applied consistently.

Interpreting Operator Intent and Cross-fabric Alignment

Scaling AI infrastructure means keeping tenant intent aligned across fabrics that were never designed to coordinate with each other — Ethernet, InfiniBand, and NVLink each have their own configuration model. Netris enables operators or upstream platforms to define a tenant using a high-level abstraction — a Netris Server Cluster, essentially a set of servers allocated to a tenant. The Netris algorithm automatically translates this into the fabric-specific configurations required by each participating connectivity fabric:

- **Spectrum-X East–West Ethernet:** Netris functions as the fabric manager. It automatically generates tenant VRFs and L3VPN VXLAN configurations compliant with the NVIDIA Spectrum-X Deployment Guide, assigns switch ports to the appropriate VXLANs, and applies required RoCEv2 settings on BlueField-3 SuperNICs through the NHN plugin.
- **InfiniBand East–West:** Netris integrates with UFM — the same UFM instance that NMC manages — to create, modify, and remove InfiniBand partitions (PKeys) and associate the relevant HCA GUIDs with the appropriate PKeys based on the Server Cluster definition.
- **NVLink Multi-Node Fabrics (NVL72):** Netris integrates with the NVIDIA NMX Controller (NMX-C) to discover GPU UIDs and create, modify, and delete NVLink partitions using the NMX-C gRPC API.
- **North–South Ethernet Fabric**: Netris functions as the fabric manager. It configures tenant VRF and VXLAN-based routing domains for ingress, egress, storage access, and service communication. BlueField DPU configurations inherit the same tenant definition.
- **Edge Connectivity:** Netris SoftGate, a multi-tenant edge gateway, provides tenant-aware NAT, Layer-4 Load Balancing, and other edge services automatically aligned within tenant boundaries.

Network Services for Tenants

Multi-tenant AI environments require network services at tenant boundaries, interconnections between tenant VPCs, and reachability to and from external networks. Netris provides these as cloud-provider-grade constructs that apply automatically within tenant isolation boundaries.

VPC Peering

VPC peering provides controlled Layer-3 connectivity between two or more VPCs. Each VPC retains its own routing context, and only explicitly permitted IP prefixes are advertised across the peering relationship. Netris implements VPC peering using VRF route leaking. The most common use case is the Shared Services VPC model, in which storage endpoints are placed into a dedicated VPC and tenant VPCs peer with it — preserving strict tenant isolation while enabling controlled, routed storage access.

VPC Connect

VPC Connect provides a mechanism for connecting tenant VPCs with external networks, including non-Netris-managed data center networks, WAN routers, and security appliances. Route exchange is explicitly controlled using route-maps and prefix-lists.

Edge Services (NAT, Layer-4 Load Balancing)

Netris SoftGate is an optional, horizontally scalable edge services layer that delivers tenant-aware NAT and Layer-4 load balancing as integrated extensions of the Netris-managed North–South fabric. These services are automatically aligned with tenant isolation boundaries and require no per-tenant manual configuration. SoftGate is deployed as a cluster of dedicated bare-metal nodes and provides NAT for controlled ingress and egress, Layer-4 load balancing for TCP and UDP traffic with health checks, and DHCP services where required.

Access Control Lists

Access Control Lists in Netris are expressed as high-level policies associated with a tenant VPC. Netris determines the enforcement point automatically — fabric switches, DPUs, host-based enforcement points, or edge service components — reducing cognitive load and misconfiguration risk while ensuring consistent enforcement aligned with tenant network boundaries.

Concurrent Multi-Tenancy and Host-Level Segmentation

The multi-tenancy model described above treats the whole bare-metal server as the atomic unit of tenancy. Operators may also enable concurrent multi-tenancy — scenarios in which bare-metal servers are shared across multiple tenants simultaneously — to improve hardware utilization. Two mechanisms are available:

DPU-Based / Hardware-Accelerated Host Segmentation (HBN)

NVIDIA BlueField DPUs operating in Zero-Trust mode with Host-Based Networking (HBN) enforce tenant isolation outside the host operating system, even when multiple tenants share a single physical host.

Netris configures the DPU's DOCA HBN service as a VTEP, including BGP/EVPN adjacencies, and applies tenant-specific networking configuration, mapping tenant definitions to DPU-managed Virtual (VF) or Physical (PF) Functions. NVIDIA DPF (DOCA Platform Framework) manages DPU device lifecycle and deploys the Netris agent onto each DPU.

Software-Based Host Segmentation (EVPN on Host)

In environments without DPUs, Netris supports software-based host segmentation. The Netris EVPN-on-Host agent configures the host networking stack to implement tenant-specific VRFs, VXLAN bridges, and routing tables aligned with tenant VPCs. From the platform's perspective, a host configured with EVPN-on-Host behaves like any other EVPN endpoint in the fabric. Virtual machines or containers can be connected directly to tenant VPCs by attaching them to the appropriate Linux bridges established by the Netris agent.

Network Validation

Operating a multi-tenant AI infrastructure requires coordinating multiple independent systems — network fabrics, accelerators, and control planes — that were not designed to collectively validate tenant intent. Netris, together with NMC, provide continuous validation across the Ethernet fabric layer.

Topology-Aware Validation

Netris continuously verifies that the live Ethernet fabric topology matches the declared topology maintained in the Netris topology manager. This validation covers switch-to-switch links and switch-to-server links. When observed physical connectivity does not match the declared topology, Netris raises a notification event.

Detection of Partial States and Drift Detection

Tenant lifecycle changes are applied across multiple systems. Validation must show whether these changes were fully applied.

Netris tracks the status of the configuration application during tenant operations. It surfaces cases where the configuration was not applied, failed to apply, or was only applied to part of the environment.

This allows operators to identify incomplete changes directly.

Data-Plane Reachability Validation in EVPN Fabrics

In EVPN VXLAN-based Ethernet fabrics, Netris tracks reachability between all VTEPs in each managed EVPN fabric, verifying that every VTEP can reach every other VTEP and surfacing issues when reachability is lost.

Simulation with NVIDIA DSX Air

NVIDIA DSX Air is NVIDIA's SaaS platform for simulating AI factory infrastructure, introduced at GTC 2026. It creates high-fidelity digital twins of the full hardware stack — GPUs, SuperNICs, DPUs, and switches — and integrates with partner solutions through open API-based connectivity. A complete AI factory environment can be constructed, configured, and validated in DSX Air before a single server is unboxed. Netris integrates with DSX Air, allowing operators to include network automation and multi-tenant configurations in the simulation alongside compute and storage systems. Tenant isolation constructs, VPC boundaries, and Ethernet fabric configurations built in Netris can be tested in DSX Air exactly as they would be deployed in a live environment. Operators use simulation across three phases: pre-deployment validation before hardware arrives; proof-of-concept work to explore new architectures or tenant models without risk to live environments; and ongoing change management to rehearse upgrades and tenant changes before applying them to live infrastructure.

Summary

NVIDIA Mission Control provides a comprehensive management platform for the compute and workload layers of an AI factory — node provisioning, workload orchestration, InfiniBand fabric management, and observability. Netris extends that foundation with NAAM (Network Automation, Abstraction, and Multi-Tenancy), adding automated Ethernet fabric management, network-layer abstraction, and cloud-provider-grade multi-tenancy with hard isolation (enforced on networking hardware) to the NMC stack.

The result is a complete AI factory stack where compute, workload, and network layers operate as a single, cohesive system. NMC and Netris NAAM together give AI cloud providers unified management from bare metal to tenant including node provisioning, workload orchestration, InfiniBand, automated Ethernet fabric, NVLink partitioning, DPU management, and edge services, all with hard isolation (enforced on networking hardware) configured automatically across every fabric from a single tenant definition. Operators can get GPU cloud business operational in weeks instead of years, provision tenants immediately, and maximize GPU utilization by dynamically reallocating capacity across tenants.

To learn how Netris integrates with your NMC environment, contact the Netris team at sales@netris.io or visit netris.io.

About Netris

Netris is the leading provider of network automation and multi-tenancy for AI infrastructure.

The Netris NAAM (Network Automation, Abstraction, and Multi-Tenancy) platform is the most widely deployed platform — trusted by high-growth neoclouds, NVIDIA Cloud Partners, AI factories, and leading AI platform providers. Netris provides a unified control plane and native integrations across the complete AI infrastructure networking stack — Ethernet (including NVIDIA Spectrum-X), NVIDIA Quantum InfiniBand, NVL72, NVIDIA BlueField DPUs, and virtual and edge networking.

Netris enables operators to deploy and operate shared GPU infrastructure in weeks instead of years, provision tenants immediately with hard isolation (enforced on networking hardware) configured automatically, maximize GPU utilization by dynamically reallocating capacity across tenants, ensure network stability, and future-proof AI infrastructure.

Learn more at https://netris.io/