



NeoCloud GPU-as-a-Service AI-Ready, Multi-Tenant, Scalable

Accelerate AI with NeoCloud: GPU-as-a-Service, multi-tenant orchestration, SONiC networking, deep observability, and vendor-neutral scalability.

NeoCloud Demands Speed and Control

AI workloads push GPUs, DPUs, and high-bandwidth networks to their limits:

- **GPU Innovation Outpaces Deployment:** NVIDIA & AMD GPUs evolve faster than deployment timelines.
- **Managing Multi-Tenant Segmentation:** Tenants require isolated compute and network environments. Precise segmentation avoids idle GPUs and tenant conflicts.
- **Bandwidth & Scaling:** AI training and inference workloads demand massive east-west bandwidth.
- **End-to-End Visibility:** Fragmented monitoring delays troubleshooting, compliance, and operational decisions.
- **Limited Engineering Capacity:** Engineers spend too much time on L1/L2 support instead of scaling the cloud.

The result: Idle GPUs, slow deployments, operational risk, and delayed revenue.

How Aviz Helps NeoCloud Providers

SONiC/Cumulus-Based, Multi-Vendor Fabrics

High-performance, scalable fabrics for NVIDIA Spectrum-X and AMD GPUs. Scale 100–800 GbE without forklift upgrades and adopt the latest switches immediately.

End-to-End Observability with Aviz PacketBroker + Service Nodes

Unified visibility across GPUs, DPUs, hosts, and networks. Enables compliance reporting, operational insights, and strengthens security tool posture.

Network Copilot™ for E-Commerce

Private AI assistant ingests GPU, SONiC/Cumulus logs, and edge data (SD-WAN, firewalls, WAFs) to:

- Automate L1/L2 support
- Enforce compliance
- Detect anomalies
- Free engineers for high-value scaling initiatives

Fast Onboarding & GPU Utilization

Maximize GPU efficiency and accelerate tenant deployment with intelligent orchestration and operational automation.

Multi-Tenant Orchestration & Day-0/1/2 Operations

ONES manages provisioning, firmware upgrades, configuration drift, rollback, and RMA workflows — ensuring seamless operations with minimal disruption.



Achieve More with Aviz

Save More (Financial Benefits)

- **Maximize GPU Utilization:** Multi-tenant orchestration ensures GPUs are never idle.
- **Reduce CapEx & OpEx:** Avoid forklift upgrades with multi-vendor SONiC/Cumulus fabrics and FTAS automation.
- **Accelerate Revenue:** Rapid onboarding of GPUs enables quicker deployment of AI workloads.

Achieve More (Strategic Benefits)

- **Scalable AI Infrastructure:** Seamlessly grow GPU capacity using Spectrum-X and clusters without operational bottlenecks.
- **Operational Efficiency:** Network Copilot™ reduces L1/L2 support, freeing engineers for innovation.
- **Compliance & Risk Mitigation:** End-to-end observability supports audits and secure multi-tenant deployments.
- **Business Agility:** Respond to tenant demands, AI workload spikes, or cloud expansion without vendor lock-in.