

Modern Storage for Neocloud Service Providers

NeuralMesh™ by WEKA® delivers the scalable, secure, and multi-tenant infrastructure modern Cloud Solution Providers (CSPs) need to stay competitive.

Challenges

- Legacy storage lacks isolation, elasticity, and AI readiness
- Manual provisioning slows tenant onboarding and wastes resources
- Metadata bottlenecks and latency degrade performance at scale

Solution

- Containerized, service-oriented architecture purpose-built for multitenancy
- Built-in physical and logical tenant isolation for secure operations
- Automated scaling and orchestration via native Kubernetes integration

Benefits

- Seamlessly scale to 150K+ tenants without performance loss
- Rapid tenant provisioning with zero tuning or disruption
- Secure, compliant, high-performance multitenant infrastructure by design

As service providers evolve into neocloud platforms—offering AI, GPU, and high-performance services on demand, they must rethink the foundations of their infrastructure. Modern tenants expect more than just storage, they require ultra-low latency, zero downtime, ironclad isolation, and elastic scalability. Traditional storage systems, built for pre-AI workloads, fall short when it comes to enabling multi-tenant, AI-native cloud services at scale.

As Neocloud providers race to deliver GPU-accelerated services for AI, ML, and high-performance computing, storage becomes a critical, and often limiting, layer of the stack. Legacy systems were never designed for the speed, scale, and isolation that modern multitenant environments demand, slowing down service providers and costing them money. To stay competitive and deliver clients the capabilities they need, like low-latency access, secure isolation, and elastic scale, service providers must first overcome these fundamental infrastructure roadblocks:

- **Legacy storage can't keep up with GPU speed**
Traditional architectures introduce I/O bottlenecks and latency that leave expensive GPUs underutilized—hurting client performance and provider ROI.
- **Multitenancy at scale strains existing systems**
Offering secure, high-performance GPU services to multiple tenants requires robust isolation, dynamic resource allocation, and fine-grained control—features most storage solutions weren't built to handle.

- **Manual tuning and provisioning slows down operations**

Provisioning environments, staging data, or managing tiering by hand creates operational drag and delays onboarding, especially as customer demand grows.

- **Unpredictable Inference, training, and RAG pipelines**

Unpredictable metadata access patterns from AI workloads overwhelm conventional file systems and introduce unpredictable latency.

NeuralMesh by WEKA is purpose-built for the demands of Neocloud service providers delivering GPU-powered AI services at scale. It replaces the rigid, monolithic storage architectures of the past with a fully containerized, service-oriented design—one that treats every storage function as an independently orchestrated microservice. These services, including data access, metadata processing, and telemetry, are distributed across every node in the cluster, delivering consistent, low-latency performance even as infrastructure grows.

NeuralMesh eliminates the operational friction common in legacy environments and generates better return on investment (ROI). It integrates natively with container orchestrators enabling automated provisioning, elastic scaling, and seamless tenant lifecycle management. Its zero-copy, zero-tuning architecture removes the need for manual scripting or file system configuration, allowing service providers to deliver high-performance environments on demand with out-of-the-box readiness for any workload, including unknown or emerging AI use cases that future tenants may bring. Whether you're enabling AI startups, hosting agentic AI pipelines, or powering enterprise-grade inference workloads, NeuralMesh provides the agility, scalability, and performance needed to stay competitive.

With NeuralMesh, Neocloud providers gain more than just a modern storage solution—they gain a strategic advantage. Its architecture not only meets today's performance and scalability demands but also future-proofs operations against the rising complexity of AI workloads and multitenant infrastructure. The result is faster time to market, better resource efficiency, and a superior experience for every customer.

- **Multi-tenant Isolation by Design**

Each tenant gets dedicated resources—drives, memory, CPU—with both physical and logical separation, eliminating noisy neighbors and ensuring secure, compliant operations.

- **Linearly Scalable Performance**

NeuralMesh maintains peak throughput even as workloads grow, thanks to intelligent rebalancing and distributed service delivery across all nodes.

- **Out-of-the-Box Performance for Any Workload**

NeuralMesh requires no custom tuning or workload-specific engineering, enabling providers to confidently support unpredictable and evolving tenant needs.

- **Faster Time to Revenue**

CSPs can provision secure, high-performance tenant environments in minutes—not days—accelerating time to market and boosting ROI.

- **Cloud-Native Deployment**

Fully compatible with container orchestration tools, enabling fast provisioning, elastic scaling, and real-time infrastructure adaptation.

True Multitenancy with Built-In Physical Isolation

NeuralMesh by WEKA delivers enterprise-grade multitenancy that scales—without compromising security, performance, or simplicity. Unlike virtualized or soft-partitioned approaches, NeuralMesh provides hard resource isolation by design.

Each tenant is assigned physically isolated infrastructure components—including dedicated drives, CPU cores, and memory allocations—ensuring there is no resource contention, no noisy neighbor interference, and no data bleed. This physical separation is reinforced by:

- Cluster-level namespace isolation
- Unique authentication credentials per tenant
- End-to-end encryption and role-based access control
- Dedicated IPs, NICs, and metadata processes where needed

With support for up to 110 isolated tenants per server and 150,000+ per cluster, NeuralMesh enables neocloud providers to deliver secure, performant, multi-tenant services with zero compromise. No manual tuning, no complex overlay management—just built-in, elastic multitenancy that's ready for AI-era scale.

