Enterprise Strategy Group™

BRIEF

AUGUST 2025

WEKA Unveils NeuralMesh: A Data Foundation for the Age of AI Reasoning

Simon Robinson, Principal Analyst

Abstract: As the AI landscape continues to evolve at a staggering pace, a range of challenges are emerging at the data and infrastructure level. In response, WEKA's NeuralMesh storage system is designed to handle a range of advanced AI workloads in a highly flexible manner.

Key Highlights

- Data management issues are already a top AI implementation challenge.
- WEKA's new NeuralMesh offers a full AI storage foundation within a flexible architecture.
- The underlying data infrastructure will grow in importance as the focus of AI pivots to advanced inference.

Running AI At Scale Quickly Becomes a Data (and Storage) Challenge

The AI landscape is developing at astonishing speed and, though this continues to open exciting opportunities for innovation, it also presents substantial challenges to those building AI infrastructure. Much of the initial effort here has necessarily focused on the compute layer—the GPUs. But building and deploying AI workloads at scale also needs—and creates—huge amounts of data. It therefore stands to reason that any AI strategy must be built on a successful data strategy.

Unfortunately, this is where many AI implementations struggle. According to Enterprise Strategy Group research, data management issues are the second most cited challenge when implementing an AI initiative, behind only high costs. While the specific challenges can vary—depending on the workload, and the stage of the AI lifecycle (e.g. training vs inference)—the issues often boil down to challenges at the storage infrastructure level. In short, traditional enterprise storage approaches were not designed for either the performance levels (e.g. throughput, latency) or data volumes generated by large-scale AI workloads.

Market Insight

84%

of organizations are already leveraging generative AI for initial or multiple use-cases.[1]

It goes without saying that GPUs are quick at processing data; so fast that traditional storage approaches, even those utilizing fast non-volatile memory express (NVMe) storage media, often cannot keep up. This is a particularly pressing challenge because GPUs are also expensive. If they are sitting idle because of latency in the storage system, that can quickly become an immense waste.

Another growing problem, especially as the emphasis shifts to inferencing approaches such as advanced reasoning, is rapidly growing context windows. In agentic AI workloads, hundreds of thousands or even millions of

[1] Source: Enterprise Strategy Group Research Report, *Navigating Build-versus-buy Dynamics for Enterprise-ready AI*, January 2025. All research references in this Brief have been taken from this report.

tokens need to be stored to avoid recomputing the same data. This can easily overwhelm GPU-based memory, so alternative approaches are required.

A final challenge is the sheer unpredictable nature of how AI workloads are evolving. Organizations are building large-scale AI factories and are looking for a storage infrastructure that can form a capable, yet resilient flexible foundation that can support their future AI requirements, however they might evolve.

WEKA Unveils NeuralMesh as a Foundation for AI innovation

WEKA's NeuralMesh is a software-defined storage system, built on a dynamic mesh architecture that offers an intelligent, adaptive foundation for AI and agentic AI deployments within large-scale AI environments. Built on WEKA's legacy of software innovation—including its HPC parallel file system software that is optimized for demanding high-throughput, low-latency AI use-cases—the company has re-architected and containerized the entire stack to create NeuralMesh. The result is a microservices-based architecture that is flexible and composable, much like modern AI applications themselves.

NeuralMesh spans a full range of data and storage capabilities, offering consistent, high-performance data access with microsecond latency, intelligent monitoring, enterprise-grade security and a self-healing infrastructure approach that the company says improves overall resiliency as it scales. Notably, NeuralMesh is designed to run on a variety of validated NVMe-based storage hardware. Though WEKA's software was initially designed to run on an external storage system connected to a cluster of GPU-based servers, WEKA has also created a version, NeuralMesh Axon, that runs entirely within a cluster of GPU servers themselves. NeuralMesh is currently available in limited release, with general availability scheduled for Fall 2025.

Analyst Insight

The pending release of NeuralMesh is a significant development for WEKA and its customers, for a number of reasons. Perhaps most significant is the containerization of its entire software stack into a microservices architecture, enabling a range of customer types—including AI specialists, hyperscale and neocloud service providers, and enterprises—to deploy a storage infrastructure environment that best meets their own individual AI needs. This is critical given both the fast pace of development and fragmentation of AI workloads over a growing number of use cases, spanning model training and various forms of advanced inferencing (reasoning, agentic, etc). There's no 'one size fits all' when it comes to AI infrastructure, and a flexible approach that spans on-premises, hybrid, cloud-based, and bare metal deployments should boost WEKA's appeal to the broadest set of customers and prospects. A further attractive aspect of NeuralMesh is enhanced support for multi-tenancy, enabling service providers to securely build a shared infrastructure across multiple clients.

The development of NeuralMesh is also significant in the context of some of the other efforts that WEKA is working on and, hence, should not be viewed in isolation. One aspect is the NeuralMesh Axon variant; the ability to implement WEKA's software stack entirely in GPU servers could be significant since it enables customers to take advantage of the potentially large amounts of NVMe storage that is often included with those servers but is often not fully utilized. When aggregated across large numbers of servers this could amount to a potentially significant pool of shared NVMe storage across what is likely already a sunk hardware cost.

Finally, WEKA is also developing an additional capability, Augmented Memory Grid, that lets AI builders create large-scale 'token warehouses' utilizing cost effective NVMe storage as a complementary tier to limited, high-cost GPU memory. This is particularly important in advanced inferencing use cases, especially as content windows grow exponentially to support increasingly sophisticated AI prompts. WEKA has specifically designed Augmented Memory Grid to work in harmony with NeuralMesh and NeuralMesh Axon, in the process addressing both a memory problem and a storage efficiency problem at the same time.

Conclusion

The relentless pace of development of the AI landscape continues to present challenges for infrastructure builders, but the supplier ecosystem also continues to rise to that challenge. As we enter the age of reasoning, the emphasis will increasingly focus on data and hence will bring the underlying data—and storage—infrastructure into even sharper focus. The ability to deliver storage performance that can keep up with increasingly capable GPUs is becoming table stakes. What AI builders also require is a flexible, agile, software-defined storage foundation that can grow as they grow and is versatile enough to support potentially numerous use cases, and potentially changing priorities in a fast-changing market. With NeuralMesh, WEKA seems to be developing just that foundation. Organizations contemplating their AI data infrastructure options are encouraged to take a close look at NeuralMesh.

> "As we enter the age of reasoning, the emphasis will increasingly focus on data, and hence will bring the underlying data—and storage—infrastructure into even sharper focus."
>
> **- Simon Robinson,** *Principal Analyst, Enterprise Strategy Group*

About Enterprise Strategy Group
Enterprise Strategy Group, now part of Omdia, provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com