Research Note: Weka's New WEKApod Nitro & Extreme

STEVE MCDOWELL, CHIEF ANALYST
10/30/24

CONTEXT

WEKA expanded its footprint in the AI data infrastructure space with the release of two new data platform appliances designed to meet diverse AI deployment needs.

The new products, WEKApod Nitro and WEKApod Prime, are WEKA's latest offerings for high-performance data solutions that support accelerated AI model training, high-throughput workloads, and enterprise AI demands. The solutions address the rapid growth of generative AI, LLMs, and RAG and fine-tuning pipelines across industries.

WEKAPOD NITRO

WEKApod Nitro is a high-performance, enterprise-grade data platform appliance specifically engineered for large-scale AI deployments and AI solution providers who require advanced infrastructure for tasks like training, tuning, and inferencing LLMs and other foundational AI models.

The new offering integrates WEKA's advanced Data Platform software with industry-leading hardware to create a turnkey solution for organizations with demanding, performance-intensive workloads.

KEY FEATURES AND CAPABILITIES

1. **High-Performance Density**: WEKApod Nitro handles substantial data volumes with exceptional speed and low latency, delivering over 18 million IOPS in a single cluster. This is among the highest IOPS densities available in enterprise data solutions.
2. **Scalability**: WEKApod Nitro's starting capacity is 0.5 petabytes of usable data, which can be expanded in 0.5-petabyte increments to accommodate the growth of AI and data-intensive operations.

3. **NVIDIA DGX SuperPOD Certification**: WEKApod Nitro is certified for NVIDIA DGX SuperPOD, demonstrating that WEKApod Nitro is fully optimized for high-density GPU environments.
4. **Efficient AI Model Training and RAG Optimization**: The WEKA Data Platform enables WEKApod Nitro to support RAG and fine-tuning pipelines, enhancing data throughput for AI model checkpointing and other data-intensive AI processes.
5. **Flexibility and Deployment Versatility**: WEKApod Nitro's cloud-native architecture enables hybrid cloud deployment and seamless data portability between on-premises and cloud environments.

WEKAPOD PRIME

WEKApod Prime is Weka's versatile, high-performance data platform appliance that supports smaller-scale AI deployments and multi-purpose HPC use cases.

Designed for organizations that require powerful AI capabilities without the extensive scale of WEKApod Nitro, WEKApod Prime balances performance with cost efficiency, making it an attractive choice for enterprises scaling AI infrastructure or managing diverse data-intensive workloads.

KEY FEATURES AND CAPABILITIES

1. **Optimized Performance**: WEKApod Prime offers robust data throughput, supporting up to 320 GB/s of read bandwidth, 96 GB/s of write bandwidth, and up to 12 million IOPS, ensuring smooth handling of AI training, inference, and other high-performance data workflows.
2. **Flexible Configuration Options**: WEKApod Prime's modular design allows enterprises to customize configurations based on specific performance needs, enabling cost-effective infrastructure scaling. Optional add-ons allow users to tailor the appliance to their workload without overprovisioning, an attractive option for companies managing variable AI or HPC requirements.
3. **Scalability**: WEKApod Prime starts with a capacity of 0.4 petabytes of usable data and can be configured up to 1.4 petabytes. Its flexible capacity range makes WEKApod Prime a great choice for organizations that anticipate growth but do not require the extreme scalability of WEKApod Nitro.

4. **Cost Efficiency**: WEKApod Prime offers a balanced price-performance ratio, allowing enterprises to scale their AI infrastructure without over-investing in components that may not be immediately necessary.

5. **Cloud-Native and Hybrid Deployment**: WEKApod Prime's cloud-native architecture offers seamless deployment flexibility, supporting both on-premises and hybrid cloud environments.

ANALYSIS

The rapid adoption of generative AI and RAG-based applications has increased the demand for flexible, high-performance data platforms. WEKApod Nitro and Prime are WEKA's response to this market shift, offering customizable and scalable AI-native infrastructure to meet both current and future needs. By providing solutions that range from the enterprise scale of WEKApod Nitro to the more versatile and cost-effective WEKApod Prime, WEKA effectively covers a broad spectrum of AI infrastructure requirements.

WEKApod Nitro's IOPS density and compatibility with NVIDIA DGX SuperPOD differentiate WEKA from other high-performance data appliances in the market. Its scalability in half-petabyte increments ensures that WEKA can efficiently serve the largest enterprise AI needs. WEKApod Prime, on the other hand, aligns well against competitors focused on flexible, high-performance data solutions by offering scalable configurations without compromising on IOPS or bandwidth, a significant advantage in the mid-market HPC and AI segments.

WEKApod Prime brings WEKA's performance benefits to smaller or more cost-conscious deployments. Competing solutions often require a significant upfront investment or have less flexible scalability options, whereas WEKApod Prime's modular approach allows companies to add features as needed without overprovisioning.

The launch of WEKApod Nitro and WEKApod Prime reinforces WEKA's position as a leader in scalable AI-native data platforms, providing enterprises with flexible, high-performance solutions that adapt to a wide range of AI and HPC requirements. The platforms offer valuable options for organizations at different stages of AI maturity, from advanced large-scale deployments to cost-efficient, flexible configurations. WEKA has long been a leader in providing performant, flexible storage. That continues with these new offerings that address the complexities of modern AI workloads while supporting enterprise AI innovation across industries.

NAND Research is a technology-focused industry analyst firm providing research, customer content, market and competitive intelligence, and custom deliverables to technology vendors, investors, and end-customer IT organizations.

Contact NAND Research via email at info@nand-research.com or visit our website at nand-research.com.