Turnkey NVIDIA L40S GPU AI Cluster Servers with Integrated Shared NVMe Storage by StoneFly

StoneFly launches NVIDIA GPU AI Cluster Servers to boost AI, ML, LLMs, and DL with scalable multi-GPU power, optimized storage, and energy efficiency.

CASTRO VALLEY, CA, UNITED STATES, September 18, 2024 /EINPresswire.com/ -- — StoneFly, Inc., a leading innovator in data storage, hyperconverged infrastructure (HCI), backup and disaster recovery (DR), and cloud solutions, is excited to announce the release of its new line of NVIDIA GPU AI Cluster Servers.

These high performance servers are engineered to accelerate Artificial Intelligence (AI), Machine Learning (ML), generative AI, Large Language Models (LLMs), and Deep Learning (DL) workloads using shared NVMe-based storage, providing enterprises with scalable multi-GPU per node computing power for their most demanding in-house data processing tasks.

StoneFly's NVIDIA GPU AI servers are designed to deliver performance with the integration of powerful NVIDIA GPUs, including the NVIDIA L40S, NVIDIA A100 Tensor Core, NVIDIA RTX A6000, and NVIDIA T4 GPUs.

Key Features of StoneFly's NVIDIA GPU AI Servers:

Unparalleled GPU Performance: Equipped with NVIDIA GPUs, these servers deliver high performance for AI and ML workloads. The NVIDIA L40S GPU, in particular, delivers up higher performance compared to previous-generation GPUs, allowing enterprises to run their AI models faster and more efficiently.

Scalability and Flexibility: StoneFly's NVIDIA GPU AI servers are designed to scale out and up without forklift upgrades, enabling enterprises to expand their computing capabilities as their data needs grow. With support for multiple GPU configurations per server, organizations can customize the servers to meet specific workload requirements and performance goals.

Optimized for AI and Data Analytics Workloads: These AI servers provide high-speed optional interconnects, including NVIDIA NVLink and NVIDIA NVSwitch, ensuring low-latency communication between GPUs for faster data processing and model training. They are ideal for high-performance computing (HPC) environments, AI research, and enterprise-level AI and DL

deployments.

High-Performance Storage Integration: The NVIDIA GPU AI servers support optional integration with StoneFly's high-performance low-latency storage solutions, including All NVMe Flash, SAS, SAN, NAS, S3 object appliances, and unified storage, to ensure seamless data flow and storage management for AI workloads. This integration minimizes bottlenecks, delivering a high-throughput, low-latency environment ideal for real-time data processing and analytics.

Advanced Cooling and Power Efficiency: Built with enterprise-grade components, StoneFly's NVIDIA GPU AI servers feature advanced cooling systems and optimized power management to handle intensive AI workloads while maintaining energy efficiency. This makes them an excellent choice for data centers looking to minimize their carbon footprint while maximizing computational power.

Enhanced Security and Reliability: StoneFly servers come with optional integrated advanced security and ransomware protection features to protect data and AI models from unauthorized access and cyber threats. The robust design ensures reliable operation with high availability, enabling continuous operation for critical AI applications.

Cost-Effective AI Infrastructure: StoneFly's NVIDIA GPU AI servers provide a cost-effective solution by combining powerful GPU performance with scalable storage options. Enterprises can optimize their investments by starting with a smaller configuration and scaling up as needed, reducing upfront costs and avoiding the need for frequent, costly hardware upgrades.

“StoneFly's NVIDIA GPU AI cluster servers are designed for high performance, low-latency, flexibility, and scalability. With multiple GPUs per server and modular performance and storage scalability through additional controllers or storage nodes, our solutions are ideal for demanding AI and deep learning workloads” said John Harris, Director Technical Sales at StoneFly, Inc. “Using StoneFly's patented technology, we have the option to create a storage pool that spans across servers, allowing GPUs to access high-speed storage efficiently, optimizing AI processing and data analytics.”

Availability
StoneFly’s NVIDIA GPU AI servers are available for order starting today. For more information on specifications, configurations, and pricing, please visit the StoneFly NVIDIA GPU AI Cluster Servers page or contact our sales team at sales@stonefly.com.

About StoneFly, Inc.
StoneFly, Inc. is a leading provider of innovative data storage, hyperconverged infrastructure, and backup and disaster recovery solutions. With a focus on high-performance, scalability, and security, StoneFly delivers robust and reliable storage and data management solutions tailored to meet the needs of enterprises across all industries.

Lilly Tahmasebi
StoneFly, Inc.
+15104158991 ext.
email us here
Visit us on social media:
Facebook
X
LinkedIn
Instagram
YouTube
