Cyfuture AI Expands GPU Cloud Portfolio with GPU as a Service Featuring H100, L40S, V100 and A100 GPUs

Noida, Uttar Pradesh Jan 22, 2026 (Issuewire.com) - Cyfuture AI, a leading AI infrastructure and cloud solutions provider, today announced the expansion of its high-performance compute offerings with GPU as a Service (GPUaaS). The enhanced portfolio provides organizations on-demand access to advanced GPU cloud infrastructure including NVIDIA H100 GPU, L40S GPU, V100 GPU, and A100 GPU, enabling enterprises to scale AI innovation faster and more efficiently.

As demand rises for GPU for Gen AI use, enterprises are seeking reliable compute to train models, deploy inference at scale, and support AI-driven applications. Cyfuture AI's GPUaaS platform addresses this demand by offering flexible, secure, and performance-optimized GPU compute for AI workloads across industries.

Meeting the Growing Demand for GPU Compute in GenAI

Organizations worldwide are accelerating investment in Generative AI, large language models (LLMs), automation, and advanced analytics. However, GPU access remains constrained due to long procurement cycles, high capital expenditure, and infrastructure complexity.

With this expanded GPU cloud portfolio, Cyfuture AI is enabling customers to access Nvidia GPU for AI Apps instantly through a scalable cloud platform - reducing time to deployment and helping businesses focus on innovation instead of infrastructure.

"AI adoption is moving fast, but GPU availability is still a critical bottleneck for many businesses," said Animesh Anand, Project Manager, Cyfuture AI. "By expanding our GPU Cloud portfolio with GPU as a Service, we are making enterprise-grade GPU compute more accessible, supporting training and inference for GenAI workloads with H100, L40S, V100 and A100 GPUs."

GPU Options Tailored for AI Training, Fine-Tuning & Inference

Cyfuture AI GPUaaS offers multiple GPU instance options to support diverse AI requirements - from high-end LLM training to efficient inference workloads:

NVIDIA H100 GPU Server

Cyfuture AI offers NVIDIA H100 GPU server designed for advanced model training and high-compute GenAI workloads. It is ideal for LLM training, GenAI training, and foundation model fine-tuning, enabling organizations to accelerate large-scale AI development. The H100 GPU Cloud is best suited for enterprises requiring maximum performance for GPU for Gen AI use.

NVIDIA L40S GPU

The NVIDIA L40S GPU Cloud is built for high-efficiency inference and scalable production deployments. It supports real-time inference workloads including AI copilots and multimodal AI deployments, making it a strong choice for organizations looking to deploy AI systems at scale. The L40S GPU Cloud is best suited for teams deploying Nvidia GPU for AI Apps in production environments.

NVIDIA A100 GPU Server

Cyfuture AI provides NVIDIA A100 GPU server as a proven option for enterprise AI training, analytics, and HPC workloads. These instances are ideal for deep learning training, HPC simulations, and data science workloads, delivering consistent performance for demanding applications. A100 GPUs are best suited for reliable compute at scale for GPU for AI work.

NVIDIA V100 GPU Cloud

The NVIDIA V100 GPU Cloud is a widely used platform for AI and ML workloads, including development and training pipelines. It is ideal for ML training, inference workloads, and AI development environments, offering stable performance across a range of enterprise use cases. V100 GPUs are best suited for organizations needing dependable, cost-effective GPU compute.

Why Cyfuture AI GPU as a Service

Cyfuture AI GPUaaS is built for teams looking for speed, security, and scalability for real-world AI deployments:

- On-demand provisioning with fast delivery timelines
- Dedicated GPU environments for predictable performance
- Scalable infrastructure for training and inference workloads
- Enterprise security controls for sensitive AI workloads
- Support for containers, Kubernetes and AI frameworks
- Flexible commercial models for startups, enterprises, and public sector

Use Cases

The expanded GPU cloud portfolio is designed to support high-demand AI deployments such as:

- LLM training and fine-tuning for GenAI applications
- High-scale inference for chatbots, copilots, and AI assistants
- Computer vision for surveillance analytics, retail intelligence, and manufacturing
- Healthcare imaging workloads and diagnostics AI
- Financial AI for fraud detection, forecasting, and risk modeling
- Research workloads and HPC simulations

Availability

Cyfuture AI's expanded GPU as a Service offering is now available for enterprises, AI startups, research institutions, and government projects. Customers can rent GPU resources by choosing from H100, L40S, V100 and A100 GPU cloud instances as per workload, performance, and budget requirements.

About Cyfuture AI

Cyfuture AI is a next-generation AI infrastructure and cloud solutions provider delivering scalable compute, GPU cloud services, cloud storage, and enterprise-ready hosting platforms. With a focus on performance, security, and reliability, Cyfuture AI empowers organizations with GPU for Gen AI use and Nvidia GPU for AI Apps, helping them train, deploy, and scale AI systems efficiently.

Cyfuture AI also enables end-to-end AI adoption through advanced capabilities such as serverless inferencing, fine tuning, RAG AI (Retrieval-Augmented Generation), and a ready-to-use model library to accelerate enterprise AI deployment. The company supports rapid development and deployment of AI chatbot, voicebots, and other AI solutions through its AI Apps Builder, alongside production-ready tools for AI agents and automation workflows.

With offerings like AI Lab as a Service, Cyfuture AI provides an integrated environment for experimentation, development, and enterprise rollout of AI solutions - helping businesses build, test, and launch innovation faster.

Media Contact

Cyfuture AI

*********@cyfuture.cloud

0120-6619504

Plot no 126, nsez, noida phase 2, uttar pradesh-201305

Source : Cyfuture

See on IssueWire