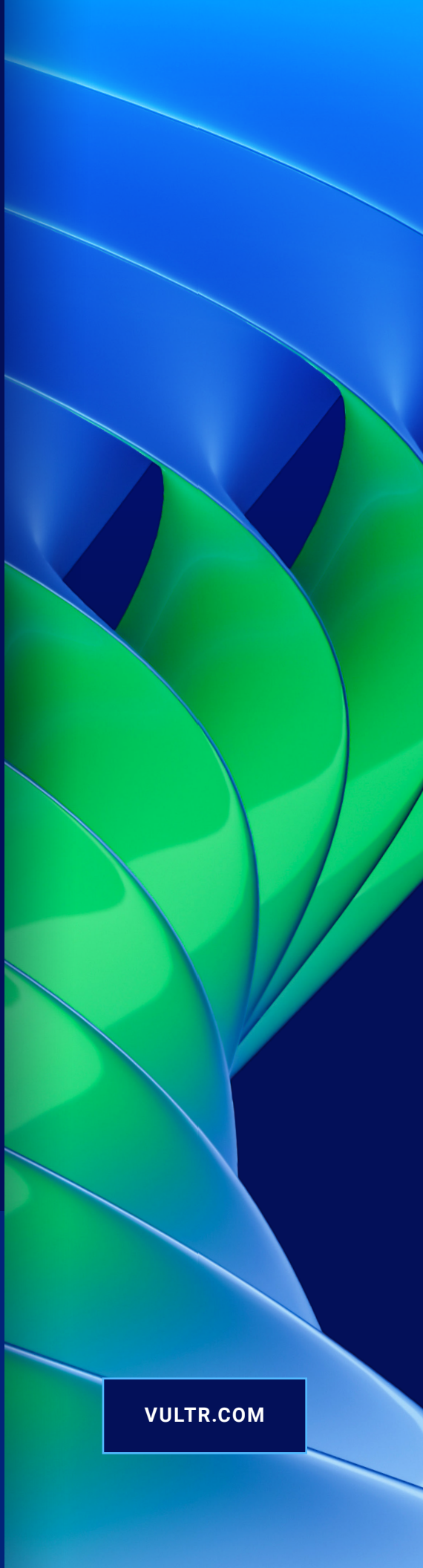
VULTR AMD

SOLUTION BRIEF

AMD + Vultr: High-Performance AI and HPC Without Vendor Lock-In



Combine AMD Instinct™ GPUs with Vultr's global cloud platform to provide scalable and efficient AI inference, model training, or HPC solutions, enabling enterprises to accelerate performance, reduce costs, and leverage the ROCm™ open software ecosystem across industries.

VULTR.COM

VULTR

Run High-Performance AI and HPC Applications on Vultr with the AMD Open Software Ecosystem

As part of the Vultr Cloud Alliance program, AMD brings its industry-leading Instinct™ MI355X, MI325X, and MI300X GPUs and ROCm™ open software to Vultr's global cloud infrastructure. This collaboration provides a powerful platform for AI and HPC, allowing enterprises to process massive datasets while controlling costs efficiently. It enables businesses to manage extensive data, reduce costs, save energy, and deploy AI solutions without vendor lock-in.

Vultr and AMD provide enterprises with scalable and composable infrastructure tailored to specific AI and HPC needs across industries. With Vultr's highly adaptable infrastructure, businesses can deploy workloads across multiple regions, ensuring low latency, high availability, and regulatory compliance. The integration with powerful AMD GPUs allows organizations to streamline complex processes and significantly enhance operational efficiency, all while maintaining complete control over their infrastructure. This empowers companies to rapidly scale and remain competitive in an increasingly demanding digital environment.

From training to inference

AI's journey from training to inference has distinct demands at each stage. During training, high computational power, large memory capacity, and flexible model iteration are required. Once the model is ready for inference, the focus shifts to efficient resource usage, low latency, and seamless scalability during inference.

Current state of the inference models

Inference challenges include fragmented and complex deployment pipelines, lack of interoperability across frameworks, limited optimization tools, and insufficient support for custom hardware and edge devices. The deployment process requires a unified API, advanced profiling tools, and seamless integration with business logic.

Effective inference requires developing unified and automated systems, enhancing support for diverse hardware, improving community resources, and streamlining workflows with standardized APIs and tools to ensure scalability, ease of use, and cross-platform compatibility.

Vultr Serverless Inference

Bridge the gap with a streamlined, serverless solution for easy deployment and scaling of AI models, leveraging Vultr's global infrastructure for low latency and optimal performance. With pre-trained models and the option to leverage proprietary data through a private vector database, Vultr simplifies AI deployment with an OpenAI-compatible API and secure data management. AMD GPUs enhance performance and scalability to meet the demands of AI workloads.

Key advantages

Freedom with open-source innovation

The AMD ROCm™ open software ecosystem eliminates vendor lock-in. Integrated with Vultr, it supports AI frameworks like PyTorch and TensorFlow, enabling flexible, rapid innovation. ROCm™ future-proofs AI solutions by ensuring compatibility across hardware, promoting adaptability and scalability.

Advanced HPC and seamless AI integration

AMD Instinct™ MI355X, MI325X, and MI300X GPUs provide the computational power needed for memory-intensive HPC tasks, accelerating complex simulations and large-scale data processing on Vultr. Seamless ROCm™ integration streamlines AI model development and deployment, reducing backend complexity and enabling faster innovation and production.

Scalable, high-performance, and sustainable

Vultr's cloud, powered by AMD Instinct™ GPUs, delivers high performance per watt, reducing energy consumption and making AI and HPC workloads more sustainable. With Vultr's predictable pricing and seamless scalability, this solution optimizes resource use, accelerates data processing, and enhances price-to-performance, offering an affordable option for demanding applications while minimizing environmental impact.

© Vultr 2025

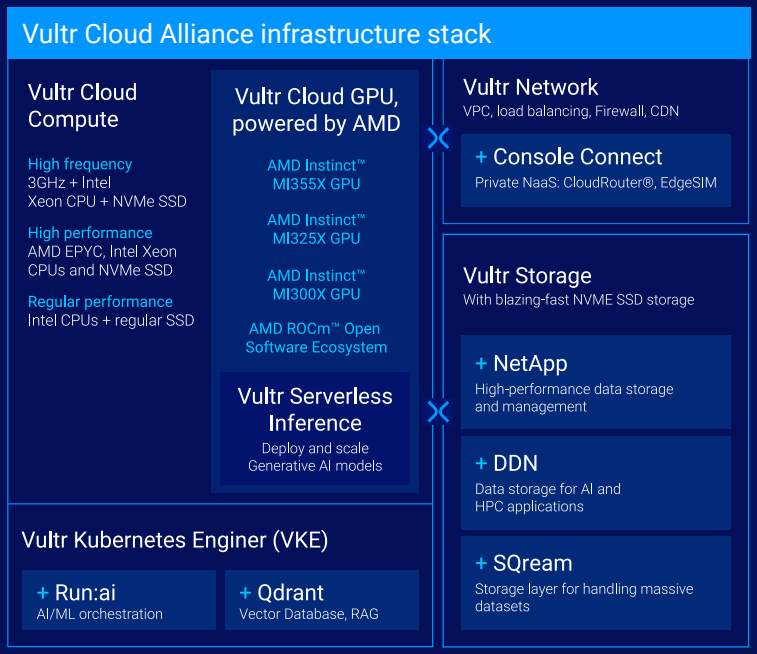
Vultr Cloud GPU powered by AMD

AMD Instinct™ MI355X GPU
Harness exceptional acceleration with breakthrough memory capacity and memory bandwidth.

AMD Instinct™ MI325X GPU
High-performance computing with 3rd Gen AMD CDNA™ for AI and HPC workloads, including deep learning and large-scale simulations.

AMD Instinct™ MI300X GPU
Experience the cutting-edge performance engineered to accelerate the most complex AI models and HPC tasks easily.

ROCm™ Open Software Ecosystem
An open software platform that drives rapid innovation and compatibility across leading AI frameworks, ensuring your AI projects are scalable and future-proof.



Driving efficiency across industries

Enterprises can tackle complex challenges in AI inference, model training, or high-performance computing (HPC) while streamlining operations, enhancing performance, and reducing costs. AMD and Vultr enable businesses to leverage advanced technology for faster insights and optimized workloads.

Healthcare & life sciences

Utilize Vultr's cloud platform and AMD Instinct™ GPUs to drive drug discovery and personalized medicine advancements. This setup ensures fast processing of complex models and simulations, meeting the industry's needs for speed and accuracy. Use cases include optimizing clinical trials, personalizing treatment plans, and performing genomic analysis to identify rare disease mutations.

Financial services

Enhance real-time analysis and risk management with AMD Instinct™ GPUs on Vultr's low-latency cloud. This solution ensures secure, fast AI inference for precise financial modeling and compliance. For example, detect and prevent fraudulent transactions in milliseconds, safeguarding against significant financial losses.

Manufacturing and energy

Optimize operational efficiency with AMD-powered AI on Vultr Cloud GPU, streamlining production and energy management. Use predictive maintenance and real-time simulations to minimize downtime and refine product design. AI can anticipate equipment failures in manufacturing and optimize power usage in energy plants to drive cost savings and operational stability.

Media and entertainment

Accelerate content creation with AMD on Vultr and benefit from real-time rendering, seamless video editing, and quicker development cycles. For example, game developers can create photorealistic graphics and immersive environments faster, while streaming platforms can deliver high-definition, low-latency video experiences, enhancing viewer engagement.

Retail

Optimize retail operations with AI-driven loss prevention and inventory management on Vultr Serverless Inference with AMD Instinct™ GPUs. Provide real-time, personalized customer experiences and streamline your supply chain. AI can analyze shopper behavior to offer personalized recommendations and help manage inventory levels, ensuring you stay fully stocked during peak sales periods.

Telecommunications

AI-powered solutions on Vultr's global cloud, using AMD Instinct™ GPUs, improve network reliability and security. They also improve network management and service quality while ensuring robust security. For example, AI can optimize network traffic to reduce peak-hour latency, improving customer experience in high-demand areas.

Learn more about AMD and Vultr

Contact us at vultr.com to get started. →

© Vultr 2025