

Vultr to Offer AMD Instinct MI355X GPUs for AI and HPC Workloads

June 13, 2025

By: [Dave McCarthy](#)

IDC'S QUICK TAKE

Vultr's forthcoming offering of the AMD Instinct MI355X GPU is an important announcement that reinforces Vultr's position as a competitive and forward-thinking cloud provider and the company's strategic partnership with AMD.

By making this high-performance, energy-efficient GPU accessible, Vultr is enabling its customers to push the boundaries of AI and HPC innovation, providing the necessary tools to develop, train, and deploy next-generation models and simulations. This move is expected to attract a wide range of users, from start-ups to large enterprises, seeking powerful and cost-effective cloud infrastructure for their most demanding workloads.

PRODUCT ANNOUNCEMENT HIGHLIGHTS

Vultr's early [adoption](#) of the AMD Instinct MI355X GPU demonstrates the company's agility and responsiveness to the evolving needs of the AI and HPC markets. By collaborating with AMD and Supermicro, Vultr aims to seamlessly integrate these powerful new GPUs into its global datacenter infrastructure. This initiative will allow Vultr customers to access advanced hardware without the typical complexities associated with deploying such sophisticated systems, enabling them to scale their operations and accelerate their computational tasks more effectively.

The AMD Instinct MI355X GPU represents a significant leap forward in accelerator technology, particularly for AI and HPC applications. Key features and benefits include:

- **New AMD CDNA 4 architecture:** The MI355X is built on AMD's new CDNA 4 architecture, which is designed to deliver superior performance and energy efficiency for AI and HPC. This architectural foundation enables enhanced computational throughput and optimized power consumption.
- **Massive HBM3E memory capacity and bandwidth:** Each MI355X GPU comes equipped with 288GB of high bandwidth memory 3E (HBM3E) and a high memory bandwidth of 8TB/s. This substantial memory capacity and bandwidth are crucial for developers working with large-scale AI models, as they allow for bigger models to be loaded entirely into memory and processed rapidly. This reduces the necessity of distributing models across multiple GPUs, thereby lowering

infrastructure costs and significantly decreasing energy consumption, especially during inference tasks.

- **Support for new low-precision formats:** The MI355X introduces support for new low-precision formats, including FP4 and FP6. These formats are vital for modern AI workloads, as they enhance computational throughput and memory efficiency. The ability to utilize lower precision data types makes the MI355X particularly well suited for deploying large AI models more quickly and efficiently, accelerating the development and deployment cycles of advanced AI applications.
- **Advanced cooling solutions:** For higher-density environments, the MI355X supports direct liquid cooling (DLC) options. This feature delivers enhanced thermal efficiency, which in turn unlocks higher-performance per rack and supports advanced cooling infrastructure at scale, crucial for sustained high-performance operations.
- **Robust ROCm software ecosystem:** Complementing the powerful hardware, the latest ROCm software enhancements further optimize AI inference, training, and framework compatibility. The ROCm software provides the performance, flexibility, and security needed to deliver tailored AI solutions, ensuring high throughput and ultra-low latency for a diverse range of modern AI workloads.
- **Optimized for AI training, inference, and HPC:** The design and feature set of the MI355X are specifically tailored to excel in computationally intensive tasks such as deep learning training, rapid AI inference, and complex HPC simulations. Its capabilities empower users to tackle more ambitious projects with greater speed and accuracy.

IDC'S POINT OF VIEW

Vultr's partnership with AMD extends beyond hardware integration; it is reinforced by a significant strategic investment. In December 2024, AMD Ventures coled a \$333 million equity financing round for Vultr, valuing the company at \$3.5 billion. This marked Vultr's first external capital raise and underscores AMD's commitment to fostering Vultr's growth in the cloud infrastructure space.

This investment is mutually beneficial:

- **Accelerated AI infrastructure expansion:** The capital infusion will enable Vultr to significantly expand its global cloud GPU fleet, including the establishment of supercompute clusters featuring thousands of AMD GPUs. This expansion directly addresses the escalating demand for high-performance AI compute.
- **Strategic market positioning for AMD:** For AMD, this partnership and investment provide a critical platform to enhance its competitive standing against rivals like NVIDIA in the burgeoning AI chip market. By having its Instinct GPUs readily

available through Vultr's widely accessible cloud platform, AMD gains increased visibility and adoption among AI innovators and enterprises.

- **Open ecosystem development:** The collaboration also includes the integration of AMD Instinct MI300X GPUs and the ROCm open source software ecosystem into Vultr's cloud infrastructure, promoting an open and flexible environment for AI development. AMD is also a member of the [Vultr Cloud Alliance](#), further cementing their strategic relationship.
- **Full-stack AMD support:** Vultr offers a robust lineup of AMD solutions, including AMD EPYC 9004 Series, AMD EPYC 7003 Series CPUs, and AMD Instinct MI325X and MI300X GPUs. When coupled with AMD EPYC 4005 Series CPUs, customers benefit from full-stack support, delivering high-performance compute and seamless integration from processor to accelerator.

IDC believes the availability of the AMD Instinct MI355X GPU on Vultr's platform is a critical development for the broader AI cloud infrastructure landscape. As AI models continue to grow in complexity and size, the demand for highly efficient and powerful compute resources intensifies. The MI355X's immense memory capacity and bandwidth, along with support for low-precision formats, directly address bottlenecks often encountered in large-scale AI deployment and training. The inclusion of direct liquid cooling also points to Vultr's commitment to supporting high-density, performant AI operations.

Subscriptions Covered:

[Public and Dedicated Cloud IaaS](#)

Please contact the IDC Hotline at 800.343.4952, ext.7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC or Industry Insights service or for information on additional copies or Web rights. Visit us on the Web at www.idc.com. To view a list of IDC offices worldwide, visit www.idc.com/offices. Copyright 2025 IDC. Reproduction is forbidden unless authorized. All rights reserved.