Press release

Oracle and AMD Expand Partnership to Help Customers Achieve Next-Generation AI Scale

Beginning in calendar Q3 2026, Oracle will be the first hyperscaler to offer a publicly available AI supercluster powered by 50,000 AMD Instinct MI450 Series GPUs

Oracle AI World, Las Vegas and Sunnyvale, California—Oct 14, 2025 – Oracle and AMD (NASDAQ: AMD) today announced a major expansion of their long-standing, multi-generation collaboration to help customers significantly scale their AI capabilities and initiatives. Building on years of co-innovation, Oracle Cloud Infrastructure (OCI) will be a launch partner for the first publicly available AI supercluster powered by AMD Instinct™ MI450 Series GPUs—with an initial deployment of 50,000 GPUs starting in calendar Q3 2026 and expanding in 2027 and beyond.

This announcement builds upon the joint work of Oracle and AMD to deliver AMD Instinct GPU platforms on OCI to end customers, beginning with the launch of AMD Instinct MI300X powered shapes in 2024 and extending to the general availability of OCI Compute with AMD Instinct MI355X GPUs. These will be available in the zettascale OCI Supercluster.

Demand for large-scale AI capacity is accelerating as next-generation AI models outgrow the limits of current AI clusters. To train and run these workloads, customers need flexible, open compute solutions engineered for extreme scale and efficiency. OCI's planned new AI superclusters will be powered by the AMD "Helios" rack design, which includes AMD Instinct MI450 Series GPUs, next- generation AMD EPYC™ CPUs codenamed "Venice," and next-generation AMD Pensando™ advanced networking codenamed "Vulcano." This vertically-optimized, rack-scale architecture is designed to deliver maximum performance, scalability, and energy efficiency for large-scale AI training and inference.

"Our customers are building some of the world's most ambitious AI applications, and that requires robust, scalable, and high-performance infrastructure," said Mahesh Thiagarajan, executive vice president, Oracle Cloud Infrastructure. "By bringing together the latest AMD processor innovations with OCI's secure, flexible platform and advanced networking powered by Oracle Acceleron, customers can push the boundaries with confidence. Through our decade-long collaboration with AMD—from EPYC to AMD Instinct accelerators—we're continuing to deliver the best price-performance, open, secure, and scalable cloud foundation in partnership with AMD to meet customer needs for this next era of AI."

"AMD and Oracle continue to set the pace for AI innovation in the cloud," said Forrest Norrod, executive vice president and general manager, Data Center Solutions Business Group, AMD. "With our AMD Instinct GPUs, EPYC CPUs, and advanced AMD Pensando networking, Oracle customers gain powerful new capabilities for training, fine-tuning, and deploying the next generation of AI. Together, AMD and Oracle are accelerating AI with open, optimized, and secure systems built for massive AI data centers."

AMD Instinct MI450 Series GPUs Coming to OCI

AMD Instinct MI450 Series GPU-powered shapes are designed to deliver high-performance, flexible cloud deployment options and provide extensive open-source support. This provides the ideal foundation for customers running today's most advanced language models, generative AI, and high-performance computing workloads. With AMD Instinct MI450 Series GPUs on OCI, customers will be able to benefit from:

- **Breakthrough compute and memory:** Helps customers achieve faster results, tackle more complex workloads, and reduce the need for model partitioning by increasing memory bandwidth for AI training models. Each AMD Instinct MI450 Series GPU will provide up to 432 GB of HBM4 and 20 TB/s of memory bandwidth, enabling customers to train and infer models that are 50 percent larger than previous generations entirely in-memory.
- **AMD optimized "Helios" rack design:** Enables customers to operate at scale while optimizing performance density, cost, and energy efficiency via dense, liquid-cooled, 72-GPU racks. The AMD "Helios" rack design integrates UALoE scale-up connectivity and Ethernet-based Ultra Ethernet Consortium (UEC)-aligned scale-out networking to minimize latency and maximize throughput across pods and racks.
- **Powerful head node:** Helps customers maximize cluster utilization and streamline large-scale workflows by accelerating job orchestration and data processing on an architecture consisting of next-generation AMD EPYC CPUs, code named "Venice." In addition, these EPYC CPUs will offer confidential computing capabilities and built-in security features to help safeguard sensitive AI workloads end to end.
- **DPU-accelerated converged networking:** Powers line-rate data ingestion to improve performance and enhance security posture for large-scale AI and cloud infrastructure. Built on the fully programmable AMD Pensando DPU technology, the DPU-accelerated converged networking offers the security and performance required for data centers to run the next era of AI training, inferencing, and cloud workloads.
- **Scale-out networking for AI:** Enables customers to leverage ultra-fast distributed training and optimized collective communication with a future-ready open networking fabric. Each GPU can be equipped with up to three 800 Gbps AMD Pensando "Vulcano" AI-NICs, providing customers with lossless, high-speed, and programmable connectivity that supports advanced RoCE and UEC standards.
- **Innovative UALink and UALoE fabric:** Helps customers efficiently expand workloads, reduce memory bottlenecks, and orchestrate large multi-trillion-parameter models. The scalable architecture minimizes hops and latency without routing through CPUs and enables direct, hardware-coherent networking and memory sharing among GPUs within a rack via UALink protocol transported over a UALoE fabric. UALink is an open, high-speed

interconnect standard purpose built for AI accelerators and supported by a broad industry ecosystem. As a result, customers gain the flexibility, scalability, and reliability needed to run their most demanding AI workloads on open standards-based infrastructure.

- **Open-source AMD ROCm™ software stack:** Enables rapid innovation, offers freedom of vendor choice, and simplifies the migration of existing AI and HPC workloads by providing customers with an open, flexible programming environment, including popular frameworks, libraries, compilers, and runtimes.
- **Advanced partitioning and virtualization:** Enables customers to safely share clusters and allocate GPUs based on workload needs by facilitating the secure and efficient use of resources via fine-grained GPU and pod partitioning, SR-IOV virtualization, and robust multi-tenancy.

To give customers that build, train, and inference AI at scale more choice, OCI also announced the general availability of OCI Compute with AMD Instinct MI355X GPUs. These will be available in the zettascale OCI Supercluster that can scale to 131,072 GPUs. AMD Instinct MI355X-powered shapes are designed with superior value, cloud flexibility, and open-source compatibility. Learn more here and here.

###

Additional Resources

- Learn more about Oracle Cloud Infrastructure
- Learn more about OCI Compute

Contact Info

Drew Smith

Oracle

drew.j.smith@oracle.com

415.336.1103

Aaron Grabein

AMD

aaron.grabein@amd.com

+1 737.256.9518

About Oracle

Oracle offers integrated suites of applications plus secure, autonomous infrastructure in the Oracle Cloud. For more information about Oracle (NYSE: ORCL), please visit us at www.oracle.com.

About Oracle AI World

Oracle AI World is where customers and partners discover the latest product and technology innovations, see how AI is being applied across industries, and connect with experts and peers. Attendees will gain practical tips and insights to drive immediate impact within their organizations and explore how Oracle is helping unlock the full potential of cloud and AI. Join the event to see new capabilities in action and hear from thought leaders and industry movers. Register now at oracle.com/ai-world or follow the news and conversation at oracle.com/news and linkedin.com/company/oracle.

About AMD

For more than 55 years AMD has driven innovation in high-performance computing, graphics and visualization technologies. Billions of people, leading Fortune 500 businesses and cutting-edge scientific research institutions around the world rely on AMD technology daily to improve how they live, work and play. AMD employees are focused on building leadership high-performance and adaptive products that push the boundaries of what is possible. For more information about how AMD is enabling today and inspiring tomorrow, visit the AMD (NASDAQ: AMD) website, blog, LinkedIn, Facebook and X pages.

Future Product Disclaimer

The preceding is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

Forward-Looking Statements Disclaimer

Statements in this press release relating to Oracle's and AMD's future plans, expectations, beliefs, and intentions are "forward-looking statements" and are subject to material risks and uncertainties. Many factors could affect Oracle's and AMD's current expectations and actual results, and could cause actual results to differ materially. A discussion of such factors and other risks that affect Oracle's and AMD's business is contained in Oracle's and AMD's Securities and Exchange Commission (SEC) filings, including Oracle's and AMD's most recent reports on Form 10-K and Form 10-Q under the heading "Risk Factors." These filings are available on the SEC's website or on Oracle's website at http://www.oracle.com/investor and AMD's website at https://ir.amd.com. All information in this article is current as of October 14, 2025 and Oracle and AMD undertakes no duty to update any statement in light of new information or future events.

Trademarks