

5 REASONS 4TH GEN AMD EPYC™ 9004 PROCESSORS ADVANCE ENTERPRISE AI IN GPU-ACCELERATED SOLUTIONS

AT A GLANCE

GPU accelerators have become the workhorse for modern AI, excelling in training large, complex models and real-time inference at scale. However, maximizing the potential of your GPU investment requires a powerful CPU partner.

By choosing an AMD EPYC 9004 CPU, you can be confident that your AI workload will benefit from its high performance, efficiency, and compatibility. AMD EPYC CPUs are designed to accelerate AI training and inference, enabling you to achieve faster results and optimize your AI project outcomes.

1

HIGH FREQUENCY & CORE COUNT

High frequency (3 GHz+) allows the CPU to process instructions faster, while a high core count (8+ cores) enables parallel processing of multiple tasks simultaneously. This is crucial for handling extensive data tasks. Additionally, it facilitates faster post-processing of results generated by the AI model.

Several AMD EPYC 9004 Series CPUs offer all core boost speeds greater than 3.5GHz with 16 – 96 Cores.
See the AMD EPYC 9004 Series CPUs data sheet: [Datasheet](#)

2

LARGE CACHE SIZE

A large cache size (20MB+) reduces the need to fetch data from slower main memory, improving performance. This is essential for facilitating fast access to massive datasets in training and running AI models and helps minimize latency, leading to quicker training times and faster execution.

AMD EPYC 9004 CPUs offer large L3 cache ranging from 64MB to 1152MB.

3

HIGH MEMORY BANDWIDTH & HIGH PERFORMANCE I/O

CPUs with high memory bandwidth (50GB/s+) and high-performance I/O interfaces are well-suited for AI tasks enable efficient data flow, enabling rapid data transfer between the CPU and GPU, and helping minimize bottlenecks during AI model training and inference.

With support for up to 12 channels of DDR5 memory and 128 PCIe® 5 I/O lanes per CPU, AMD EPYC 9004 processors deliver exceptional memory and I/O bandwidth.

4

ENERGY-EFFICIENT CORES

AMD EPYC processors have energy-efficient cores that enable efficient GPU-accelerated systems, optimizing high performance, energy efficiency, memory, and IO scalability. AMD EPYC processors power the world's most energy-efficient servers, delivering exceptional performance and helping reduce energy costs.¹

2P 4th Gen AMD EPYC 9754 (128C)-powered servers have up to 2.4x (avg. 72% higher) inference FPS per watt vs. 2P 5th Gen Xeon 8592+ (64C)-powered servers when comparing select published Phoronix Test Suite OpenVINO™ workloads.²

5

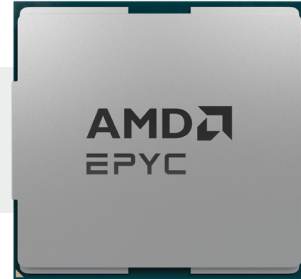
GPU ECOSYSTEM COMPATIBILITY

GPU compatibility enables seamless communication and optimized performance between different components. This allows for efficient utilization of GPU resources and facilitates smooth software integration, leading to faster training times and better model performance.

AMD EPYC CPUs are compatible with a wide range of GPUs, including those from Nvidia and AMD. This allows for flexibility in choosing the best GPU for your specific AI needs. Additionally, AMD collaborates closely with major AI software vendors to optimize their software for AMD EPYC processors, enabling smooth integration and efficient utilization of resources.

4TH GEN AMD EPYC™ PROCESSORS ADVANCE DATA CENTER AI

TOGETHER WE ADVANCE AI



1 - EPYC-028D: SPECpower_ssj® 2008, SPECrate®2017_int_energy_base, and SPECrate®2017_fp_energy_base based on results published on SPEC's website as of 2/21/24. VMmark® server power-performance / server and storage power-performance (PPKW) based results published at <https://www.vmware.com/products/vmmark/results3x.1.html?sort=score>. The first 105 ranked SPECpower_ssj®2008 publications with the highest overall efficiency overall ssj_ops/W results were all powered by AMD EPYC processors. For SPECrate®2017 Integer (Energy Base), AMD EPYC CPUs power the first 8 top SPECrate®2017_int_energy_base performance/system W scores. For SPECrate®2017 Floating Point (Energy Base), AMD EPYC CPUs power the first 12 SPECrate®2017_fp_energy_base performance/system W scores. For VMmark® server power-performance (PPKW), have the top 5 results for 2- and 4-socket matched pair results outperforming all other socket results and for VMmark® server and storage power-performance (PPKW), have the top overall score. See <https://www.amd.com/en/claims/epyc4#faq-EPYC-028D> for the full list. For additional information on AMD sustainability goals see: <https://www.amd.com/en/corporate/corporate-responsibility/data-center-sustainability.html>. More information about SPEC® is available at <http://www.spec.org>. SPEC, SPECrate, and SPECpower are registered trademarks of the Standard Performance Evaluation Corporation. VMmark is a registered trademark of VMware in the US or other countries.

2 - SP5-252: Third-party testing OpenVINO 2023.2.dev FPS comparison based on Phoronix review <https://www.phoronix.com/review/intel-xeon-platinum-8592/9> as of 12/14/2023 of select OpenVINO tests: Vehicle Detection FP16, Person Detection FP16, Person Vehicle Bike Detection FP16, Road Segmentation ADAS FP16 and Face Detection Retail FP16. Road Segmentation ADAS FP16 was max uplift of 2.36x. Testing not independently verified by AMD. Scores will vary based on system configuration and determinism mode used (Power Determinism used). OpenVINO is a trademark of Intel Corporation or its subsidiaries.

©2024 Advanced Micro Devices, Inc. all rights reserved. AMD, the AMD arrow, EPYC, and combinations thereof, are trademarks of Advanced Micro Devices, Inc. Intel, the Intel logo and Xeon are trademarks of Intel Corporation or its subsidiaries. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies. Certain AMD technologies may require third-party enablement or activation. Supported features may vary by operating system. Please confirm with the system manufacturer for specific features. No technology or product can be completely secure.

PID: 242852451-A August 2024