5 REASONS 4TH GEN AMD EPYC™ 9004 PROCESSORS ADVANCE DATA CENTER AI

AT A GLANCE

Machine learning and AI technologies are presenting amazing new opportunities for businesses and organizations of all sizes, transforming workloads and even entire industries, but AI success requires data center modernization to both accommodate and make space for the computing requirements of AI applications. Now is the time to update data centers to maximize the potential of AI.

With their leadership performance and density, exceptional energy efficiency and built-in security capabilities, AMD EPYC 9004 processors can be a critical component of AI success, making room in existing data centers for new AI applications as well as supporting key elements of AI to extend and enrich business workloads.

1

CONSOLIDATE WORKLOADS

AMD EPYC 9004 CPU -based servers accommodate the demands of AI and analytics solutions by consolidating existing workloads efficiently, freeing up space and energy to deploy AI, delivering exceptional performance and helping reduce energy costs.

2

MIX WORKLOADS AND ENTERPRISE AI

Use AMD EPYC 9004 CPU -powered servers to provide the foundational business process computing and data sources needed to feed AI innovation and insights. As workload needs diversify and evolve, AMD EPYC 9004 CPU-powered servers provide compelling performance and efficiency for classical machine learning and batch and small-scale real-time inferencing.

3

POWER AI HOST NODES ON GPU HOSTS

For large AI training or inferencing deployments, AMD EPYC 9004 CPU-based servers are excellent hosts to support GPU acceleration, featuring the performance, scalability, compatibility and energy efficiency to support hosting of advanced GPU AI engines.

4

CREATE ENERGY-EFFICIENT SOLUTIONS

AMD EPYC processors power the world's most energy-efficient servers, delivering exceptional performance and helping reduce energy costs.[EPYC-028D] Discover new ways to optimize core usage, impact your TCO and advance your sustainability goals.

5

CONFIDENTLY NAVIGATE TODAY'S BUSINESS RISKS, COMPLEXITIES AND REQUIREMENTS

Compute with confidence, knowing that your business is addressing many of today's newest security challenges with the advanced features of AMD Infinity Guard,[1] and AMD commitments to compliance, corporate responsibility, supply chain resilience and a long-term processor roadmap.

TECHNICAL DEEP DIVE

#1 CONSOLIDATE WORKLOADS

- AMD EPYC 9004 CPU-based servers accommodate demanding workloads efficiently, freeing up space and energy to deploy AI. For instance, when comparing 2P servers to deliver 10,000 units of integer performance, it takes an estimated five fewer servers (six AMD servers vs. 11 Intel servers) and 168 fewer cores using 96-core AMD EPYC 9654 CPUs vs 60-core Intel Xeon Platinum 8490H CPUs.[SP5TCO-032] Over the 3-years of this analysis, the AMD powered server is estimated to deliver:
  - The same performance from 45% fewer servers
  - Up to 45% less power usage
  - Up to 58% CAPEX reduction
  - Up to 46% OPEX reduction
- AMD EPYC 9004 CPUs make server consolidation possible. In a refresh scenario, you could deploy 73% fewer servers and 70% fewer racks when using 4th Gen AMD EPYC 9334 CPUs to replace aging Intel Xeon 6143 Sky Lake CPUs to deliver 80,000 units of integer performance.[SP5TCO-055]
- Get outstanding performance across a wide range of workloads. The AMD EPYC Family of Processors holds 300+ world records for performance.[2]

#2 MIX WORKLOADS AND ENTERPRISE AI

- AMD EPYC 9004 processors are compatible out of the box with major x86 application vendors whether on premises or hosted by one of the major cloud providers making data center modernization smooth.
- AMD EPYC 9004 processors power servers that deliver cost-optimized performance across a wide range of diverse workloads, including enterprise AI projects. By offering offer high core counts, I/O throughput and memory bandwidth, they enable rapid digital transformations while simultaneously delivering efficiency with the possibility of deploying fewer servers to support the same number of workloads, users and jobs.
- Achieve outstanding end-to-end AI throughput performance on a wide variety of use cases. Using the TPCx-AI SF30 benchmark, 2P servers with 96C AMD EPYC 9654 processors deliver up to an aggregate of ~65% more AI test cases per minute vs. 2P servers with 64C Intel Xeon Platinum 8592+.[SP5-051A]

#3 POWER AI HOST NODES ON GPU HOSTS

- For larger and more demanding workloads, GPUs may be the right choice for AI processing. However, to maximize the potential of these valuable GPUs, it is crucial to select the right host platform. AMD EPYC 9004 processor-powered servers deliver cost-optimized performance across a wide range of workloads, including enterprise AI, and they offer:
  - Up to 2X the cores vs. 5th Gen Intel® Xeon® processors for heterogenous and GPU processing[3]
  - Up to 29% more all-core turbo frequency per core[4, EPYC-022A]
  - Up to 50% more memory capacity and up to 29% more memory bandwidth for heterogeneous GPU workloads[EPYC-032B, EPYC-033B]
- Augment AMD EPYC 9004 CPUs with the acceleration performance of AMD Instinct™ GPUs to address the full spectrum of AI workloads.
- AMD EPYC 9004 Series CPUs are qualified for AI "Host Node" systems for GPU-accelerated systems working with HPE, Lenovo, Supermicro, Inspur and multiple ODMs.

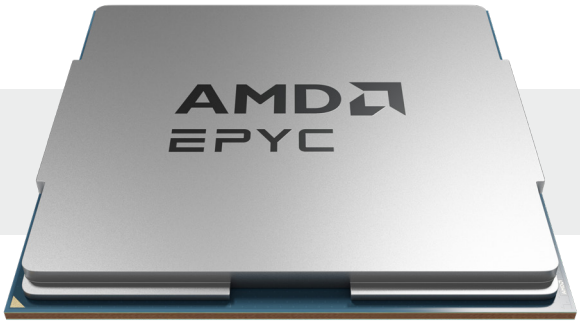
#4 CREATE ENERGY-EFFICIENT SOLUTIONS

- AMD delivers enterprise-ready IT and AI solutions that offer energy efficiency, technology innovation, and low TCO while contributing to reduced CO2 emissions.
- Comparing 2P servers, AMD EPYC 9754 delivers ~53% more integer performance/est. system W than Intel Xeon Platinum 8592+.SP5-175A
- A 2P 128-core EPYC 9754 powered server has 2.25x the overall ssj_ops/W of a 2P 64-core Xeon Platinum 8592+ base server running SPECpower_ssj®2008.SP5-011F

#5 CONFIDENTLY NAVIGATE TODAY'S BUSINESS RISKS, COMPLEXITIES AND REQUIREMENTS

- The AMD "Security by Design" approach includes state-of-the-art security features and a silicon embedded security subsystem. It starts with a foundation of data and cyberthreat management from AMD Infinity Guard[1] features that have been consistently added to AMD EPYC processors across its longstanding security roadmap.
- 4th Gen AMD EPYC processors add more layers for both physical and virtual security, addressing special security concerns about migrating sensitive applications and data.
- Compliance and corporate responsibility are more straightforward than ever. AMD partners with suppliers to advance human rights, drive environmental sustainability and support supply chain resilience.[5]
- AMD takes a collaborative approach with our suppliers to promote continuous improvement and drive positive change across our value chain.

4TH GEN AMD EPYC™ PROCESSORS ADVANCE DATA CENTER AI



TOGETHER WE ADVANCE_AI

LEARN MORE AT AMD.COM/AI.

1 AMD Infinity Guard features vary by EPYC™ Processor generations and/or series. Infinity Guard security features must be enabled by server OEMs and/or Cloud Service Providers to operate. Check with your OEM or provider to confirm support of these features. Learn more about Infinity Guard at https://www.amd.com/en/technologies/infinity-guard. GD-183A

2 For a full list of AMD EPYC world records, see https://www.amd.com/en/processors/epyc-world-records.

3 Core advantage comparison of highest core count 4th Gen EPYC, the 128C AMD EPYC 9754 (https://www.amd.com/en/product/13371) vs. highest core count 5th Gen Intel Xeon, the 64C Xeon 8592+.

4 All-core boost* advantage comparison of 64C AMD EPYC 9554 (https://www.amd.com/en/legal/claims/epyc-claims.html) vs. 64C Intel Xeon 8592+ (https://www.phoronix.com/image-viewer.php?id=intel-5th-gen-xeon-emeraldrapids&image=intel_xeon_11_lrg).

5 See amd.com/en/corporate-responsibility/supply-chain-responsibility.

For details on the claims used in this document, visit amd.com/en/legal/claims/epyc.

PID 242648751-A June 2024