October 10, 2024

AMD Launches 5th Gen AMD EPYC CPUs, Maintaining Leadership Performance and Features for the Modern Data Center

— New EPYC processors deliver record breaking performance and efficiency for a wide range of data center workloads —

— AMD EPYC CPUs continue momentum, with more than 950 AMD EPYC-powered public instances available globally and more than 350 platforms from OxMs —

SAN FRANCISCO, Oct. 10, 2024 (GLOBE NEWSWIRE) -- AMD (NASDAQ: AMD) today announced the availability of the 5th Gen AMD EPYC™ processors, formerly codenamed "Turin," the world's best server CPU for enterprise, AI and cloud[1].

Using the "Zen 5" core architecture, compatible with the broadly deployed SP5 platform[2] and offering a broad range of core counts spanning from 8 to 192, the AMD EPYC 9005 Series processors extend the record-breaking performance[3] and energy efficiency of the previous generations with the top of stack 192 core CPU delivering up to 2.7X the performance[4] compared to the competition.

New to the AMD EPYC 9005 Series CPUs is the 64 core AMD EPYC 9575F, tailor made for GPU powered AI solutions that need the ultimate in host CPU capabilities. Boosting up to 5GHz[5], compared to the 3.8GHz processor of the competition, it provides up to 28% faster processing needed to keep GPUs fed with data for demanding AI workloads.

"From powering the world's fastest supercomputers, to leading enterprises, to the largest Hyperscalers, AMD has earned the trust of customers who value demonstrated performance, innovation and energy efficiency," said Dan McNamara, senior vice president and general manager, server business, AMD. "With five generations of on-time roadmap execution, AMD has proven it can meet the needs of the data center market and give customers the standard for data center performance, efficiency, solutions and capabilities for cloud, enterprise and AI workloads."

The World's Best CPU for Enterprise, AI and Cloud Workloads

Modern data centers run a variety of workloads, from supporting corporate AI-enablement initiatives, to powering large-scale cloud-based infrastructures to hosting the most demanding business-critical applications. The new 5th Gen AMD EPYC processors provide leading performance and capabilities for the broad spectrum of server workloads driving business IT today.

The new "Zen 5" core architecture, provides up to 17% better instructions per clock (IPC) for enterprise and cloud workloads and up to 37% higher IPC in AI and high performance

computing (HPC) compared to “Zen 4.”[6]

With AMD EPYC 9965 processor-based servers, customers can expect significant impact in their real world applications and workloads compared to the Intel Xeon® 8592+ CPU-based servers, with:

- Up to 4X faster time to results on business applications such as video transcoding.[7]
- Up to 3.9X the time to insights for science and HPC applications that solve the world’s most challenging problems.[8]
- Up to 1.6X the performance per core in virtualized infrastructure.[9]

In addition to leadership performance and efficiency in general purpose workloads, 5th Gen AMD EPYC processors enable customers to drive fast time to insights and deployments for AI deployments, whether they are running a CPU or a CPU + GPU solution.

Compared to the competition:

- The 192 core EPYC 9965 CPU has up to 3.7X the performance on end-to-end AI workloads, like TPCx-AI (derivative), which are critical for driving an efficient approach to generative AI.[10]
- In small and medium size enterprise-class generative AI models, like Meta’s Llama 3.1-8B, the EPYC 9965 provides 1.9X the throughput performance compared to the competition.[11]
- Finally, the purpose built AI host node CPU, the EPYC 9575F, can use its 5GHz max frequency boost to help a 1,000 node AI cluster drive up to 700,000 more inference tokens per second. Accomplishing more, faster.[12]

By modernizing to a data center powered by these new processors to achieve 391,000 units of SPECrate®2017_int_base general purpose computing performance, customers receive impressive performance for various workloads, while gaining the ability to use an estimated 71% less power and ~87% fewer servers[13]. This gives CIOs the flexibility to either benefit from the space and power savings or add performance for day-to-day IT tasks while delivering impressive AI performance.

AMD EPYC CPUs – Driving Next Wave of Innovation
The proven performance and deep ecosystem support across partners and customers have driven widespread adoption of EPYC CPUs to power the most demanding computing tasks. With leading performance, features and density, AMD EPYC CPUs help customers drive value in their data centers and IT environments quickly and efficiently.

5th Gen AMD EPYC Features
The entire lineup of 5th Gen AMD EPYC processors is available today, with support from Cisco, Dell, Hewlett Packard Enterprise, Lenovo and Supermicro as well as all major ODMs and cloud service providers providing a simple upgrade path for organizations seeking compute and AI leadership.

High level features of the AMD EPYC 9005 series CPUs include:

- Leadership core count options from 8 to 192, per CPU

- “Zen 5” and “Zen 5c” core architectures
- 12 channels of DDR5 memory per CPU
- Support for up to DDR5-6400 MT/s[14]
- Leadership boost frequencies up to 5GHz[5]
- AVX-512 with the full 512b data path
- Trusted I/O for Confidential Computing, and FIPS certification in process for every part in the series

Model (AMD EPYC)	Cores	CCD (Zen5/Zen5c)	Base/Boost[5] (up to GHz)	Default TDP (W)	L3 Cache (MB)	Price (1 KU, USD)
9965	192 cores	“Zen5c”	2.25 / 3.7	500W	384	$14,813
9845	160 cores	“Zen5c”	2.1 / 3.7	390W	320	$13,564
9825	144 cores	“Zen5c”	2.2 / 3.7	390W	384	$13,006
9755 9745	128 cores	“Zen5” “Zen5c”	2.7 / 4.1 2.4 / 3.7	500W 400W	512 256	$12,984 $12,141
9655 9655P 9645	96 cores	“Zen5” “Zen5” “Zen5c”	2.6 / 4.5 2.6 / 4.5 2.3 / 3.7	400W 400W 320W	384 384 384	$11,852 $10,811 $11,048
9565	72 cores	“Zen5”	3.15 / 4.3	400W	384	$10,486
9575F 9555 9555P 9535	64 cores	“Zen5” “Zen5” “Zen5” “Zen5”	3.3 / 5.0 3.2 / 4.4 3.2 / 4.4 2.4 / 4.3	400W 360W 360W 300W	256 256 256 256	$11,791 $9,826 $7,983 $8,992
9475F 9455 9455P	48 cores	“Zen5” “Zen5” “Zen5”	3.65 / 4.8 3.15 / 4.4 3.15 / 4.4	400W 300W 300W	256 192 192	$7,592 $5,412 $4,819
9365	36 cores	“Zen5”	3.4 / 4.3	300W	256	$4,341
9375F 9355 9355P 9335	32 cores	“Zen5” “Zen5” “Zen5” “Zen5”	3.8 / 4.8 3.55 / 4.4 3.55 / 4.4 3.0 / 4.4	320W 280W 280W 210W	256 256 256 256	$5,306 $3,694 $2,998 $3,178
9275F 9255	24 cores	“Zen5” “Zen5”	4.1 / 4.8 3.25 / 4.3	320W 200W	256 128	$3,439 $2,495
9175F 9135 9115	16 cores	“Zen5” “Zen5” “Zen5”	4.2 / 5.0 3.65 / 4.3 2.6 / 4.1	320W 200W 125W	512 64 64	$4,256 $1,214 $726
9015	8 cores	“Zen5”	3.6 / 4.1	125W	64	$527

Supporting Resources

- Watch the full AMD Advancing AI Keynote
- Learn more about 5th Gen AMD EPYC Processors
- Follow AMD on X
- Connect with AMD on LinkedIn

About AMD

For more than 50 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 and X pages.

Cautionary Statement

This press release contains forward-looking statements concerning Advanced Micro Devices, Inc. (AMD) such as the features, functionality, performance, availability, timing and expected benefits of AMD products including AMD EPYC™ processors, which are made pursuant to the Safe Harbor provisions of the Private Securities Litigation Reform Act of 1995. Forward-looking statements are commonly identified by words such as "would," "may," "expects," "believes," "plans," "intends," "projects" and other terms with similar meaning. Investors are cautioned that the forward-looking statements in this press release are based on current beliefs, assumptions and expectations, speak only as of the date of this press release and involve risks and uncertainties that could cause actual results to differ materially from current expectations. Such statements are subject to certain known and unknown risks and uncertainties, many of which are difficult to predict and generally beyond AMD's control, that could cause actual results and other future events to differ materially from those expressed in, or implied or projected by, the forward-looking information and statements. Material factors that could cause actual results to differ materially from current expectations include, without limitation, the following: Intel Corporation's dominance of the microprocessor market and its aggressive business practices; Nvidia's dominance in the graphics processing unit market and its aggressive business practices; the cyclical nature of the semiconductor industry; market conditions of the industries in which AMD products are sold; loss of a significant customer; competitive markets in which AMD's products are sold; economic and market uncertainty; quarterly and seasonal sales patterns; AMD's ability to adequately protect its technology or other intellectual property; unfavorable currency exchange rate fluctuations; ability of third party manufacturers to manufacture AMD's products on a timely basis in sufficient quantities and using competitive technologies; availability of essential equipment, materials, substrates or manufacturing processes; ability to achieve expected manufacturing yields for AMD's products; AMD's ability to introduce products on a timely basis with expected features and performance levels; AMD's ability to generate revenue from its semi-custom SoC products; potential security vulnerabilities; potential security incidents including IT outages, data loss, data breaches and cyberattacks; uncertainties involving the ordering and shipment of AMD's products; AMD's reliance on third-party intellectual property to design and introduce new products; AMD's reliance on third-party companies for design, manufacture and supply of motherboards, software, memory and other computer platform components; AMD's reliance on Microsoft and other software vendors' support to design and develop software to run on AMD's products; AMD's reliance on third-party distributors and add-in-board partners; impact of modification or interruption of AMD's internal business processes and information systems; compatibility of AMD's products with some or all industry-standard software and hardware; costs related to defective products; efficiency of AMD's supply chain; AMD's ability to rely on third party supply-chain logistics functions; AMD's ability to effectively control sales of its products on the gray market; long-term impact of climate change on AMD's business; impact of government actions and regulations such as export regulations, tariffs and trade protection measures; AMD's ability to realize its deferred tax assets; potential tax liabilities; current and future claims and litigation; impact of environmental laws, conflict minerals related provisions and other laws or regulations; evolving expectations from governments, investors, customers and other stakeholders regarding corporate responsibility matters; issues related to the responsible use of AI; restrictions imposed by agreements governing AMD's notes, the guarantees of Xilinx's notes and the revolving credit agreement; impact of acquisitions, joint ventures and/or investments on AMD's business and AMD's ability to integrate acquired businesses; impact of any impairment of the combined company's assets; political, legal and economic risks and natural disasters; future impairments of technology license purchases; AMD's ability to attract and retain qualified personnel; and AMD's stock price volatility. Investors are urged to review in detail the risks and uncertainties in AMD's

Securities and Exchange Commission filings, including but not limited to AMD's most recent reports on Forms 10-K and 10-Q.

AMD, the AMD Arrow logo, EPYC and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other names are for informational purposes only and may be trademarks of their respective owners.

[1] EPYC-029C: Comparison based on thread density, performance, features, process technology and built-in security features of currently shipping servers as of 10/10/2024. EPYC 9005 series CPUs offer the highest thread density [EPYC-025B], leads the industry with 500+ performance world records [EPYC-023F] with performance world record enterprise leadership Java® ops/sec performance [EPYCWR-20241010-260], top HPC leadership with floating-point throughput performance [EPYCWR-2024-1010-381], AI end-to-end performance with TPCx-AI performance [EPYCWR-2024-1010-525] and highest energy efficiency scores [EPYCWR-20241010-326]. The 5th Gen EPYC series also has 50% more DDR5 memory channels [EPYC-033C] with 70% more memory bandwidth [EPYC-032C] and supports 70% more PCIe® Gen5 lanes for I/O throughput [EPYC-035C], has up to 5x the L3 cache/core [EPYC-043C] for faster data access, uses advanced 3-4nm technology, and offers Secure Memory Encryption + Secure Encrypted Virtualization (SEV) + SEV Encrypted State + SEV-Secure Nested Paging security features. See the AMD EPYC Architecture White Paper (https://library.amd.com/l/3f4587d147382e2/) for more information.

[2] AMD EPYC™ 9005 processors utilize the SP5 socket. Many factors determine system compatibility. Check with your server manufacturer to determine if this processor is supported in systems configured with previously launched AMD EPYC 9004 family CPUs.

[3] EPYC-022F: For a complete list of world records see: http://amd.com/worldrecords.

[4] 9xx5-002C: SPECrate®2017_int_base comparison based on published scores from www.spec.org as of 10/10/2024.

2P AMD EPYC 9965 (3000 SPECrate®2017_int_base, 384 Total Cores, 500W TDP, $14,813 CPU $), 6.060 SPECrate®2017_int_base/CPU W, 0.205 SPECrate®2017_int_base/CPU $, https://www.spec.org/cpu2017/results/res2024q3/cpu2017-20240923-44833.html)

2P AMD EPYC 9755 (2720 SPECrate®2017_int_base, 256 Total Cores, 500W TDP, $12,984 CPU $), 5.440 SPECrate®2017_int_base/CPU W, 0.209 SPECrate®2017_int_base/CPU $, https://www.spec.org/cpu2017/results/res2024q4/cpu2017-20240923-44837.pdf)

2P AMD EPYC 9754 (1950 SPECrate®2017_int_base, 256 Total Cores, 360W TDP, $11,900 CPU $), 5.417 SPECrate®2017_int_base/CPU W, 0.164 SPECrate®2017_int_base/CPU $, https://www.spec.org/cpu2017/results/res2023q2/cpu2017-20230522-36617.html)

2P AMD EPYC 9654 (1810 SPECrate®2017_int_base, 192 Total Cores, 360W TDP, $11,805 CPU $), 5.028 SPECrate®2017_int_base/CPU W, 0.153 SPECrate®2017_int_base/CPU $, https://www.spec.org/cpu2017/results/res2024q1/cpu2017-20240129-40896.html)

2P Intel Xeon Platinum 8592+ (1130 SPECrate®2017_int_base, 128 Total Cores, 350W TDP, $11,600 CPU $) 3.229 SPECrate®2017_int_base/CPU W, 0.097

SPECrate®2017_int_base/CPU $, http://spec.org/cpu2017/results/res2023q4/cpu2017-20231127-40064.html)

2P Intel Xeon 6780E (1410 SPECrate®2017_int_base, 288 Total Cores, 330W TDP, $11,350 CPU $) 4.273 SPECrate®2017_int_base/CPU W, 0.124 SPECrate®2017_int_base/CPU $, https://spec.org/cpu2017/results/res2024q3/cpu2017-20240811-44406.html)

SPEC®, SPEC CPU®, and SPECrate® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information. Intel CPU TDP at https://ark.intel.com/.

[5] GD-150: Boost Clock Frequency is the maximum frequency achievable on the CPU running a bursty workload. Boost clock achievability, frequency, and sustainability will vary based on several factors, including but not limited to: thermal conditions and variation in applications and workloads. GD-150.

[6] 9xx5-001: Based on AMD internal testing as of 9/10/2024, geomean performance improvement (IPC) at fixed-frequency.

- 5th Gen EPYC CPU Enterprise and Cloud Server Workloads generational IPC Uplift of 1.170x (geomean) using a select set of 36 workloads and is the geomean of estimated scores for total and all subsets of SPECrate®2017_int_base (geomean), estimated scores for total and all subsets of SPECrate®2017_fp_base (geomean), scores for Server Side Java multi instance max ops/sec, representative Cloud Server workloads (geomean), and representative Enterprise server workloads (geomean).

“Genoa” Config (all NPS1): EPYC 9654 BIOS TQZ1005D 12c12t (1c1t/CCD in 12+1), FF 3GHz, 12x DDR5-4800 (2Rx4 64GB), 32Gbps xGMI;

“Turin” config (all NPS1): EPYC 9V45 BIOS RVOT1000F 12c12t (1c1t/CCD in 12+1), FF 3GHz, 12x DDR5-6000 (2Rx4 64GB), 32Gbps xGMI

Utilizing Performance Determinism and the Performance governor on Ubuntu® 22.04 w/ 6.8.0-40-generic kernel OS for all workloads.

- 5th Gen EPYC generational ML/HPC Server Workloads IPC Uplift of 1.369x (geomean) using a select set of 24 workloads and is the geomean of representative ML Server Workloads (geomean), and representative HPC Server Workloads (geomean).

“Genoa” Config (all NPS1) “Genoa” config: EPYC 9654 BIOS TQZ1005D 12c12t (1c1t/CCD in 12+1), FF 3GHz, 12x DDR5-4800 (2Rx4 64GB), 32Gbps xGMI;

“Turin” config (all NPS1): EPYC 9V45 BIOS RVOT1000F 12c12t (1c1t/CCD in 12+1), FF 3GHz, 12x DDR5-6000 (2Rx4 64GB), 32Gbps xGMI

Utilizing Performance Determinism and the Performance governor on Ubuntu 22.04 w/ 6.8.0-40-generic kernel OS for all workloads except LAMMPS, HPCG, NAMD, OpenFOAM, Gromacs which utilize 24.04 w/ 6.8.0-40-generic kernel.

SPEC® and SPECrate® are registered trademarks for Standard Performance Evaluation Corporation. Learn more at spec.org.

[7] 9xx5-006: AMD internal testing as of 09/01/2024, on FFMPEG (Raw to VP9, 1080P, 302 Frames, 1 instance/thread, video source: https://media.xiph.org/video/derf/y4m/ducks_take_off_1080p50.y4m).

System Configurations: 2P AMD EPYC™ 9965 reference system (2 x 192C) 1.5TB 24x64GB DDR5-6400 running at 6000MT/s, SAMSUNG MZWLO3T8HCLS-00A07, NPS=4, Ubuntu 22.04.3 LTS, Kernel Linux 5.15.0-119-generic, BIOS RVOT1000C (determinism enable=power), 10825484.25 Frames/Hour Median

2P AMD EPYC™ 9654 production system (2 x 96C) 1.5TB 24x64GB DDR5-5600, , SAMSUNG MO003200KYDNC, NPS=4, Ubuntu 22.04.3 LTS, Kernel Linux 5.15.0-119-generic, BIOS 1.56 (determinism enable=power) , 5154133.333 Frames/Hour Median

2P Intel Xeon Platinum 8592+ production system (2 x 64C) 1TB 16x64GB DDR5-5600, 3.2 TB NVME, Ubuntu 22.04.3 LTS, Kernel Linux 6.5.0-35-generic), BIOS ESE122V-3.10, 2712701.754 Frames/Hour Median

For 3.99x the performance with the AMD EPYC 9965 vs Intel Xeon Platinum 8592+ systems

For 1.90x the performance with the AMD EPYC 9654 vs Intel Xeon Platinum 8592+ systems

Results may vary based on factors including but not limited to BIOS and OS settings and versions, software versions and data used.

[8] 9xx5-022: Source: https://www.amd.com/content/dam/amd/en/documents/epyc-technical-docs/performance-briefs/amd-epyc-9005-pb-gromacs.pdf

[9] 9xx5-071: VMmark® 4.0.1 host/node FC SAN comparison based on “independently published” results as of 10/10/2024.
Configurations:

2 node, 2P AMD EPYC 9575F (128 total cores) powered server running VMware ESXi8.0 U3, 3.31 @ 4 tiles,
https://www.infobellit.com/BlueBookSeries/VMmark4-FDR-1003

2 node, 2P AMD EPYC 9554 (128 total cores) powered server running VMware ESXi 8.0 U3, 2.64 @ 3 tiles,
https://www.infobellit.com/BlueBookSeries/VMmark4-FDR-1002

2 node, 2P Intel Xeon Platinum 8592+ (128 total cores) powered server running VMware ESXi 8.0 U3, 2.06 @ 2.4 Tiles,
https://www.infobellit.com/BlueBookSeries/VMmark4-FDR-1001

VMmark is a registered trademark of VMware in the US or other countries.

[10] 9xx5-012: TPCxAI @SF30 Multi-Instance 32C Instance Size throughput results based on AMD internal testing as of 09/05/2024 running multiple VM instances. The aggregate end-to-end AI throughput test is derived from the TPCx-AI benchmark and as such is not comparable to published TPCx-AI results, as the end-to-end AI throughput test results do not comply with the TPCx-AI Specification.

2P AMD EPYC 9965 (384 Total Cores), 12 32C instances, NPS1, 1.5TB 24x64GB DDR5-6400 (at 6000 MT/s), 1DPC, 1.0 Gbps NetXtreme BCM5720 Gigabit Ethernet PCIe, 3.5 TB Samsung MZWLO3T8HCLS-00A07 NVMe®, Ubuntu® 22.04.4 LTS, 6.8.0-40-generic (tuned-adm profile throughput-performance, ulimit -l 198096812, ulimit -n 1024, ulimit -s 8192), BIOS RVOT1000C (SMT=off, Determinism=Power, Turbo Boost=Enabled)

2P AMD EPYC 9755 (256 Total Cores), 8 32C instances, NPS1, 1.5TB 24x64GB DDR5-6400 (at 6000 MT/s), 1DPC, 1.0 Gbps NetXtreme BCM5720 Gigabit Ethernet PCIe, 3.5 TB Samsung MZWLO3T8HCLS-00A07 NVMe®, Ubuntu 22.04.4 LTS, 6.8.0-40-generic (tuned-adm profile throughput-performance, ulimit -l 198096812, ulimit -n 1024, ulimit -s 8192), BIOS RVOT0090F (SMT=off, Determinism=Power, Turbo Boost=Enabled)

2P AMD EPYC 9654 (192 Total cores) 6 32C instances, NPS1, 1.5TB 24x64GB DDR5-4800, 1DPC, 2 x 1.92 TB Samsung MZQL21T9HCJR-00A07 NVMe, Ubuntu 22.04.3 LTS, BIOS 1006C (SMT=off, Determinism=Power)

Versus 2P Xeon Platinum 8592+ (128 Total Cores), 4 32C instances, AMX On, 1TB 16x64GB DDR5-5600, 1DPC, 1.0 Gbps NetXtreme BCM5719 Gigabit Ethernet PCIe, 3.84 TB KIOXIA KCMYXRUG3T84 NVMe, , Ubuntu 22.04.4 LTS, 6.5.0-35 generic (tuned-adm profile throughput-performance, ulimit -l 132065548, ulimit -n 1024, ulimit -s 8192), BIOS ESE122V (SMT=off, Determinism=Power, Turbo Boost = Enabled)

Results:

CPU Median Relative Generational
Turin 192C, 12 Inst 6067.531 3.775 2.278
Turin 128C, 8 Inst 4091.85 2.546 1.536
Genoa 96C, 6 Inst 2663.14 1.657 1
EMR 64C, 4 Inst 1607.417 1 NA

Results may vary due to factors including system configurations, software versions and BIOS settings. TPC, TPC Benchmark and TPC-C are trademarks of the Transaction Processing Performance Council.

[11] 9xx5-009: Llama3.1-8B throughput results based on AMD internal testing as of 09/05/2024.

Llama3-8B configurations: IPEX.LLM 2.4.0, NPS=2, BF16, batch size 4, Use Case Input/Output token configurations: [Summary = 1024/128, Chatbot = 128/128, Translate = 1024/1024, Essay = 128/1024, Caption = 16/16].

2P AMD EPYC 9965 (384 Total Cores), 6 64C instances 1.5TB 24x64GB DDR5-6400 (at 6000 MT/s), 1 DPC, 1.0 Gbps NetXtreme BCM5720 Gigabit Ethernet PCIe, 3.5 TB Samsung MZWLO3T8HCLS-00A07 NVMe®, Ubuntu® 22.04.3 LTS, 6.8.0-40-generic (tuned-adm profile throughput-performance, ulimit -l 198096812, ulimit -n 1024, ulimit -s 8192) , BIOS RVOT1000C, (SMT=off, Determinism=Power, Turbo Boost=Enabled), NPS=2

2P AMD EPYC 9755 (256 Total Cores), 4 64C instances , 1.5TB 24x64GB DDR5-6400 (at 6000 MT/s), 1DPC, 1.0 Gbps NetXtreme BCM5720 Gigabit Ethernet PCIe, 3.5 TB Samsung MZWLO3T8HCLS-00A07 NVMe®, Ubuntu 22.04.3 LTS, 6.8.0-40-generic (tuned-adm profile

throughput-performance, ulimit -l 198096812, ulimit -n 1024, ulimit -s 8192), BIOS RVOT1000C (SMT=off, Determinism=Power, Turbo Boost=Enabled), NPS=2

2P AMD EPYC 9654 (192 Total Cores) 4 48C instances , 1.5TB 24x64GB DDR5-4800, 1DPC, 1.0 Gbps NetXtreme BCM5720 Gigabit Ethernet PCIe, 3.5 TB Samsung MZWLO3T8HCLS-00A07 NVMe®, Ubuntu® 22.04.4 LTS, 5.15.85-051585-generic (tuned-adm profile throughput-performance, ulimit -l 1198117616, ulimit -n 500000, ulimit -s 8192), BIOS RVI1008C (SMT=off, Determinism=Power, Turbo Boost=Enabled), NPS=2

Versus 2P Xeon Platinum 8592+ (128 Total Cores), 2 64C instances , AMX On, 1TB 16x64GB DDR5-5600, 1DPC, 1.0 Gbps NetXtreme BCM5719 Gigabit Ethernet PCIe, 3.84 TB KIOXIA KCMYXRUG3T84 NVMe®, Ubuntu 22.04.4 LTS 6.5.0-35-generic (tuned-adm profile throughput-performance, ulimit -l 132065548, ulimit -n 1024, ulimit -s 8192), BIOS ESE122V (SMT=off, Determinism=Power, Turbo Boost = Enabled).
Results:

CPU 2P EMR 64c 2P Turin 192c 2P Turin 128c 2P Genoa 96c
Average Aggregate Median Total Throughput 99.474 193.267 182.595 138.978
Competitive 1 1.943 1.836 1.397
Generational NA 1.391 1.314 1

Results may vary due to factors including system configurations, software versions and BIOS settings.

[12] 9xx5-087: As of 10/10/2024; this scenario contains several assumptions and estimates and, while based on AMD internal research and best approximations, should be considered an example for information purposes only, and not used as a basis for decision making over actual testing.

Referencing 9XX5-056A: “2P AMD EPYC 9575F powered server and 8x AMD Instinct MI300X GPUs running Llama3.1-70B select inference workloads at FP8 precision vs 2P Intel Xeon Platinum 8592+ powered server and 8x AMD Instinct MI300X GPUs has ~8% overall throughput increase across select inference use cases” and 8763.52 tokens/s (9575F) versus 8,048.48 tokens/s (8592+) at 128 input / 2048 output tokens, 500 prompts for 1.089x the tokens/s or 715.04 more tokens/s.

1 Node = 2 CPUs and 8 GPUs.
Assuming a 1000 node cluster, 1000 * 715.04 = 715,040 tokens/s

For ~700,000 more tokens/s

Results may vary due to factors including system configurations, software versions and BIOS settings.

[13] 9xx5TCO-001a: This scenario contains many assumptions and estimates and, while based on AMD internal research and best approximations, should be considered an example for information purposes only, and not used as a basis for decision making over actual testing. The AMD Server & Greenhouse Gas Emissions TCO (total cost of ownership) Estimator Tool - version 1.12, compares the selected AMD EPYC™ and Intel® Xeon® CPU based server solutions required to deliver a TOTAL_PERFORMANCE of 39100 units of SPECrate2017_int_base performance as of October 10, 2024. This scenario compares a

legacy 2P Intel Xeon 28 core Platinum_8280 based server with a score of 391 versus 2P EPYC 9965 (192C) powered server with an score of 3030 (https://spec.org/cpu2017/results/res2024q3/cpu2017-20240923-44833.pdf) along with a comparison upgrade to a 2P Intel Xeon Platinum 8592+ (64C) based server with a score of 1130 (https://spec.org/cpu2017/results/res2024q3/cpu2017-20240701-43948.pdf). Actual SPECrate®2017_int_base score for 2P EPYC 9965 will vary based on OEM publications.

Environmental impact estimates made leveraging this data, using the Country / Region specific electricity factors from the 2024 International Country Specific Electricity Factors 10 – July 2024 , and the United States Environmental Protection Agency 'Greenhouse Gas Equivalencies Calculator'.

For additional details, see https://www.amd.com/en/claims/epyc5#9xx5TCO-001a

[14] 9xx5-083: 5th Gen EPYC processors support DDR5-6400 MT/s for targeted customers and configurations. 5th Gen production SKUs support up to DDR5-6000 MT/s to enable a broad set of DIMMs across all OEM platforms and maintain SP5 platform compatibility

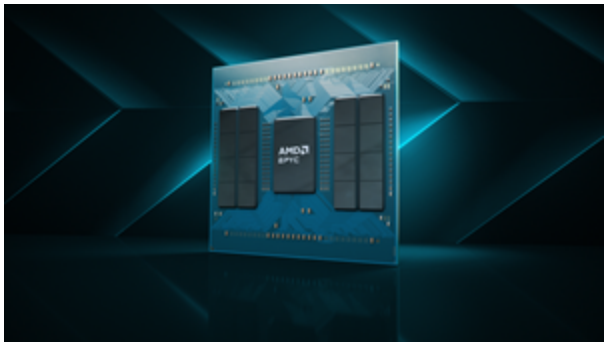
A photo accompanying this announcement is available at https://www.globenewswire.com/NewsRoom/AttachmentNg/3bb614ee-e307-43a7-a36b-f5bd02ed1335

Media Contacts:
Aaron Grabein
AMD Communications
+1 512-602-8950
aaron.grabein@amd.com

Mitch Haws
AMD Investor Relations
+1 512-944-0790
mitch.haws@amd.com

Source: Advanced Micro Devices, Inc.

5th Gen AMD EPYC CPU



5th Gen AMD EPYC CPU