贏向 AI 創新
佈局永續先機

AMD 資料中心解決方案事業群
台灣區資深業務副總經理

林建誠 **Ken Lin**

AMD

together we advance_

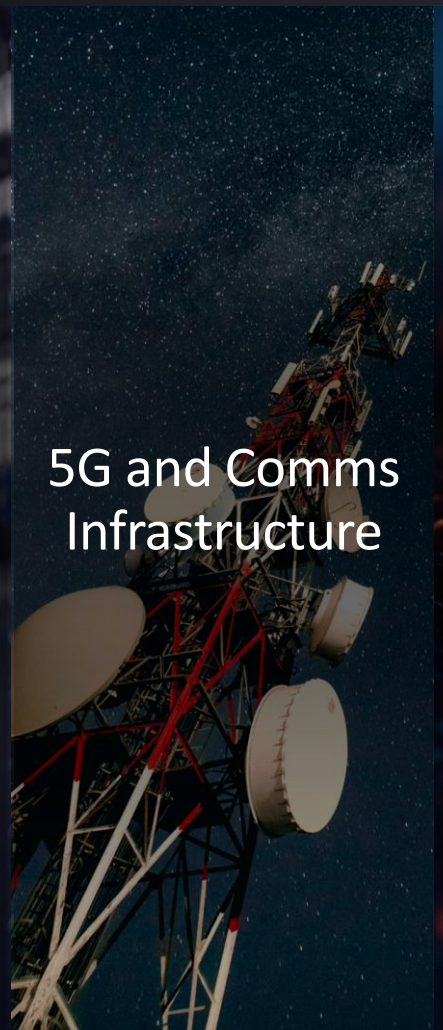
CAUTIONARY STATEMENT

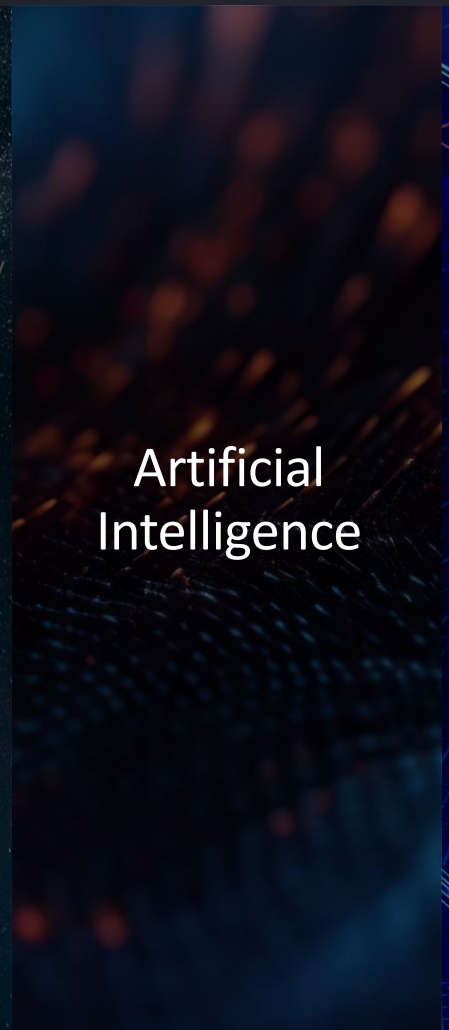
This presentation contains forward-looking statements concerning Advanced Micro Devices, Inc. (AMD) such as the features, functionality, performance, availability, timing and expected benefits of current and future AMD products, which are made pursuant to the Safe Harbor provisions of the Private Securities 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 presentation are based on current beliefs, assumptions and expectations, speak only as of the date of this presentation 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. 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.

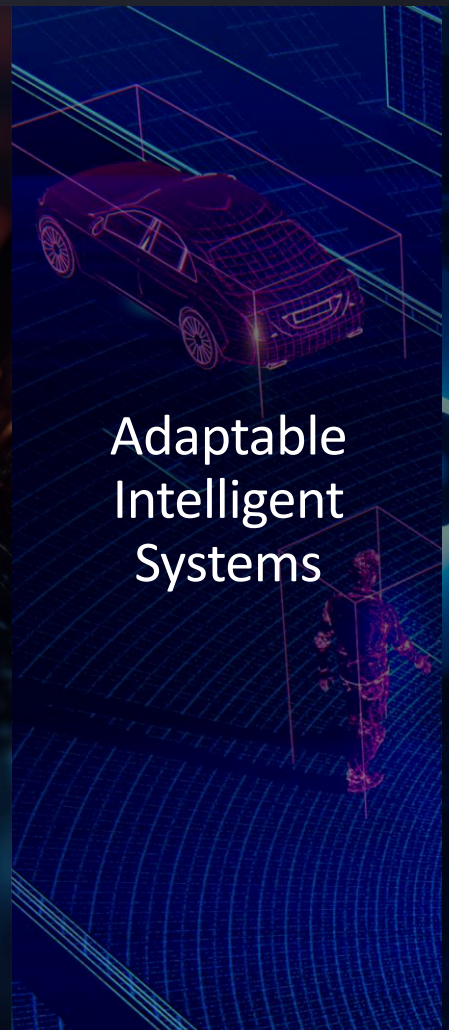
AMD does not assume, and hereby disclaims, any obligation to update forward-looking statements made in this presentation, except as may be required by law.

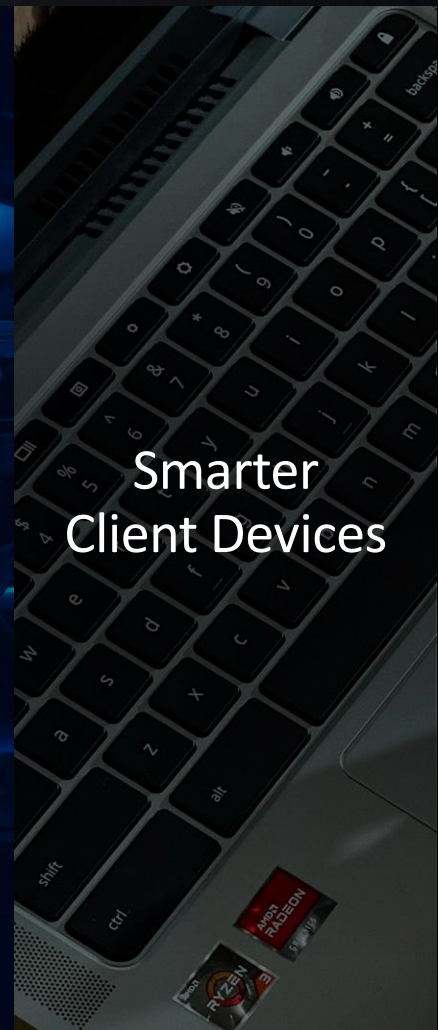
AMD

High-performance and adaptive computing powers our world

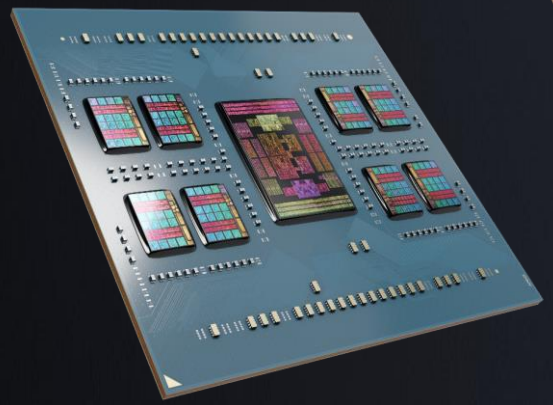






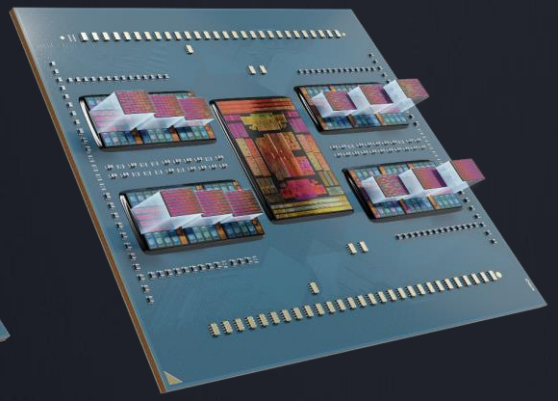


Computing infrastructure
optimized for data center workloads



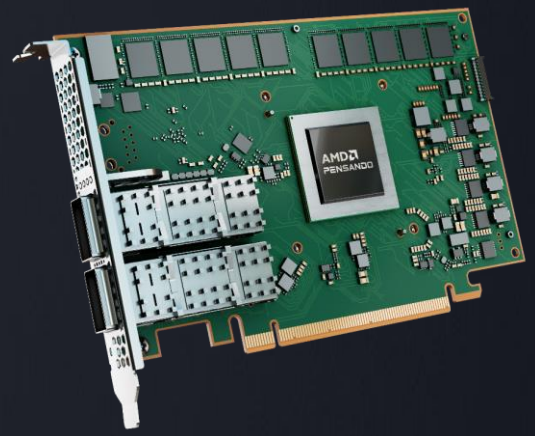
Cloud Native Computing

AMD EPYC



Technical Computing

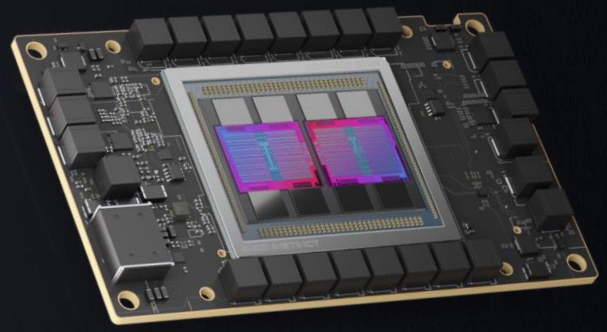
AMD EPYC



Networking

AMD ALVEO

AMD PENSANDO



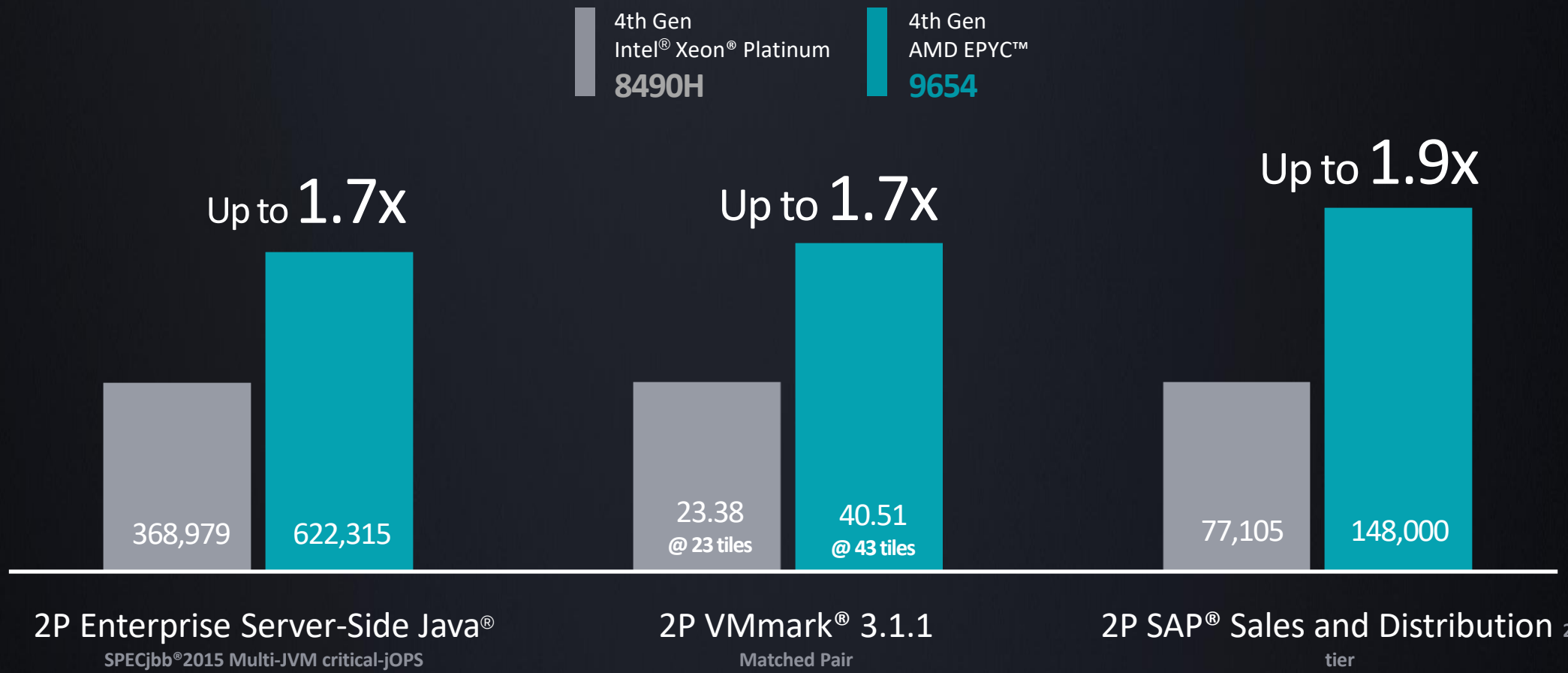
AI

AMD INSTINCT

AMD EPYC

AMD VERSAL

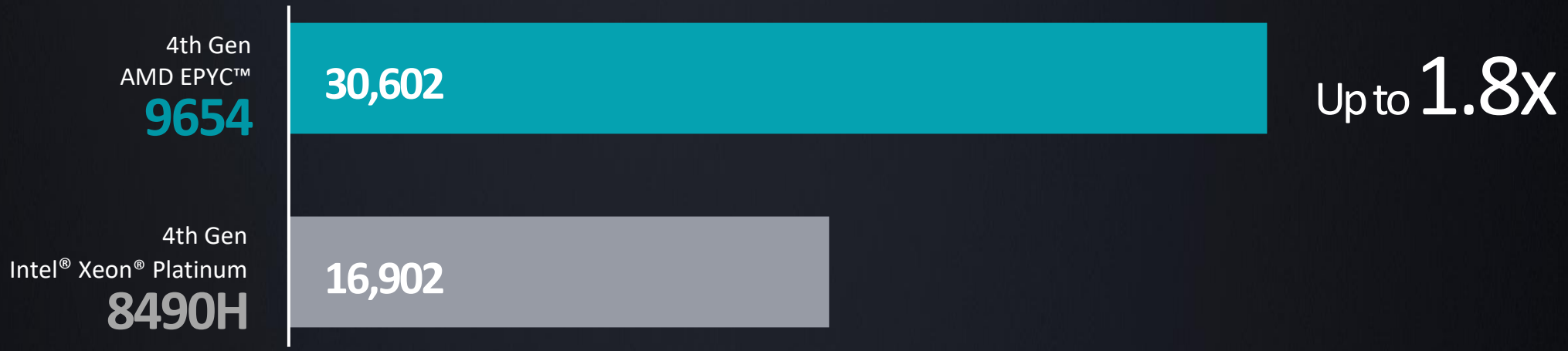
Enterprise leadership



Claims: SP5-104A, -049C, -056B

AMD
together we advance_

Efficiency leadership

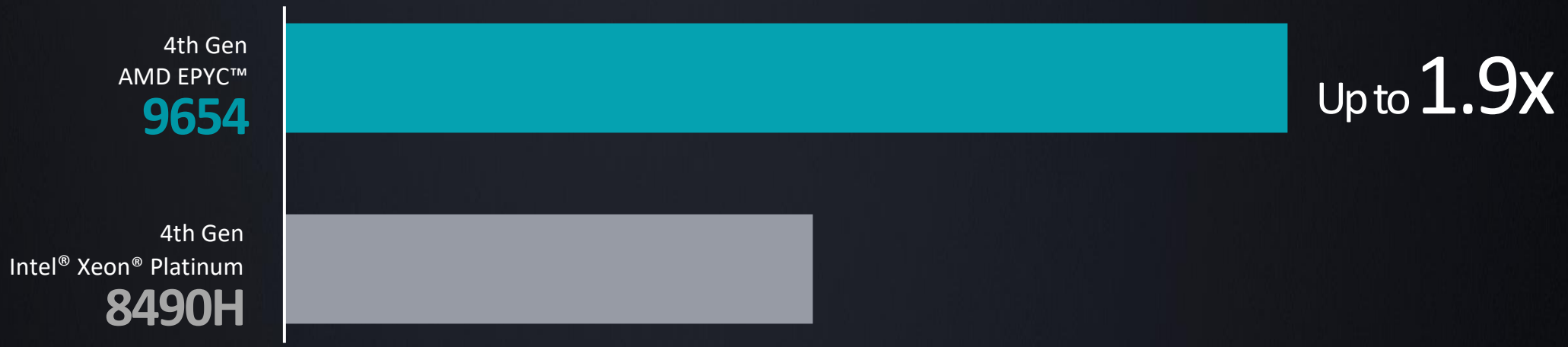


2P Energy Efficiency

SPECpower_ssj®2008

Claim: SP5-011E

CPU AI leadership



TPCx-AI

End-to-end workload derived from TPC® Express AI
Comparison run at SF3

Claim: SP5-051

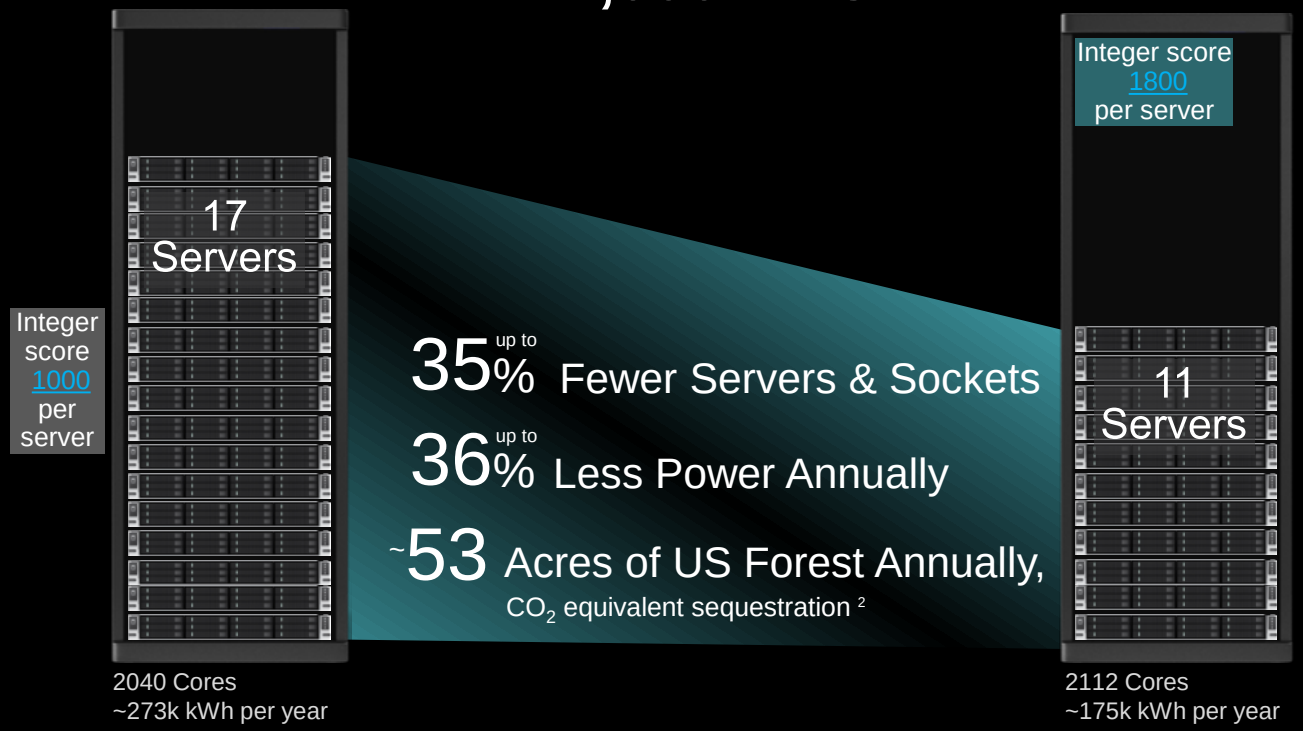
EPYC 96c 9654 vs. Intel 60c 8490H CPUs

EPYC Savings
(Estimated)

2P INTEL® Platinum 8490H

2,000 VMs

2P AMD EPYC™ 9654



35% up to Fewer Servers & Sockets

36% up to Less Power Annually

~53 Acres of US Forest Annually, CO_2 equivalent sequestration [2]

Concerned with

- Hdwr & SW License Costs?
- Space?
- Power?

With: 6 Fewer Servers[1]
12 Fewer Sockets[1]
2 Fewer Licenses[1]

EPYC Solutions Enable
(Estimated)

47% up to Lower Hardware Cost[1]

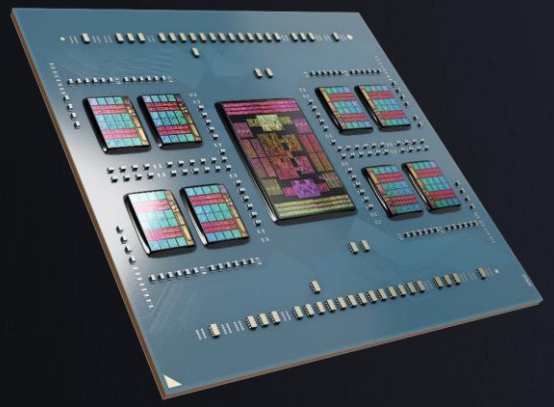
21% up to Lower 1st yr Cost / VM[1]

Analysis based on the AMD EPYC™ Server Virtualization & Greenhouse Gas Emission TCO Estimation Tool - version 12.15 as of 05/19/2023.
AMD processor pricing based on 1KU price as of Jan 2023. Intel® Xeon® Scalable CPU data and pricing from https://ark.intel.com as of Jan 2023. All pricing is in USD.
Use of third-party marks / logos/ products is for informational purposes only and no endorsement of or by AMD is intended or implied. GD-83
Virtualization license cost are retail price for VMware® vSphere Enterprise Plus w/ Production support - 24x7 3yr support, calculated with one software license for every 32-core increment in a socket. VMware is a registered trademark of VMware in the US or other countries.
[1] TCO time frame of 3-year and includes estimated costs for hardware, virtualization software, real estate, admin and power with power @ $0.128/kWh with 8kW / rack and a PUE of 1.7. Networking and storage power external to the server are not included in this analysis. [2] Values are for USA.

See endnote SP5TCO-036A.

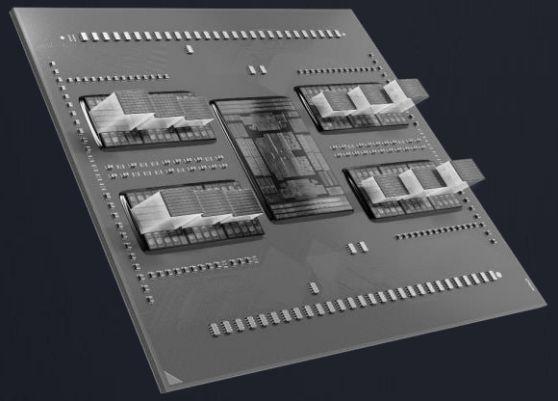
AMD
together we advance_

Computing infrastructure
optimized for data center workloads



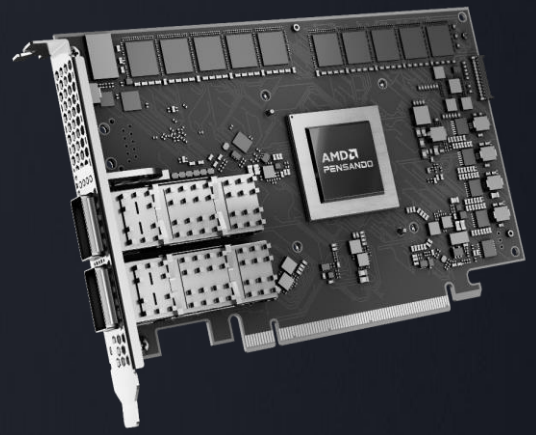
Cloud Native Computing

AMD EPYC



Technical Computing

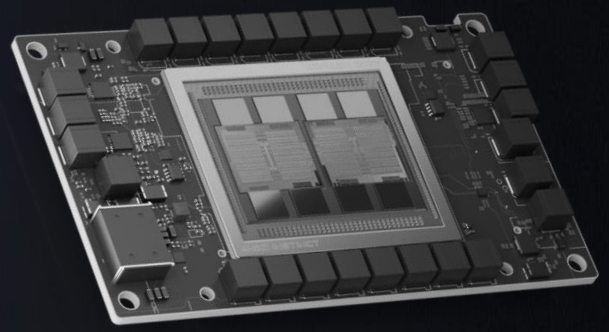
AMD EPYC



Networking

AMD ALVEO

AMD PENSANDO



AI

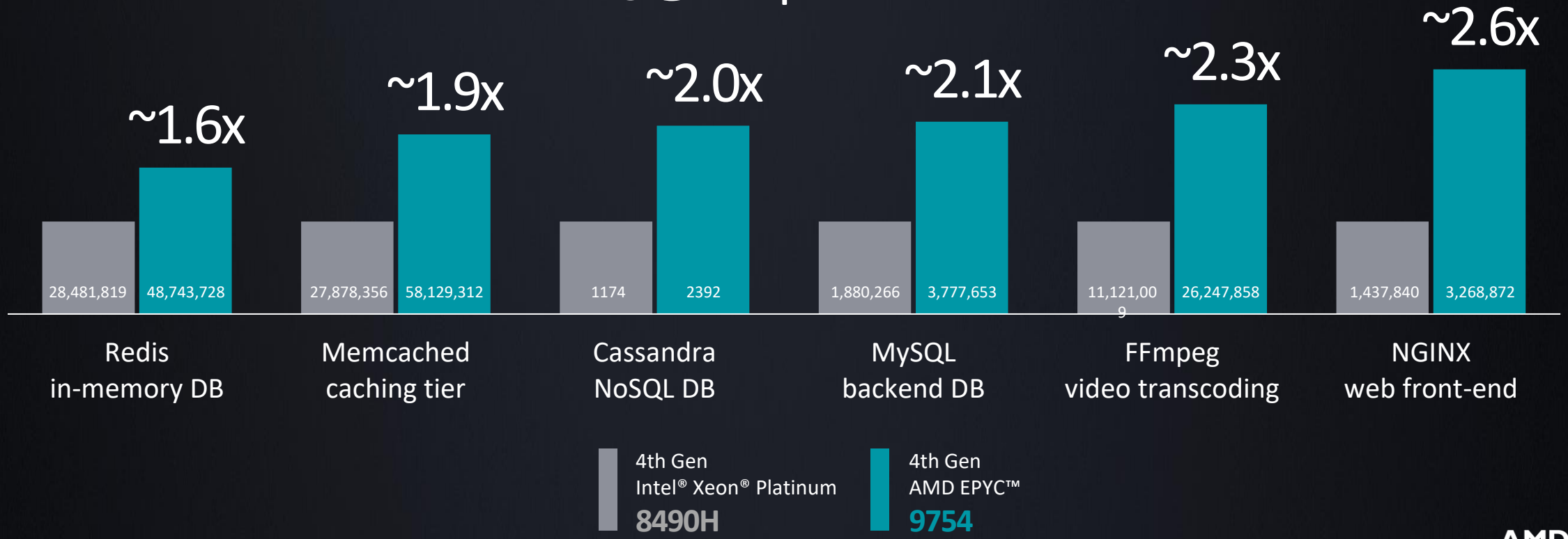
AMD INSTINCT

AMD EPYC

AMD VERSAL

Cloud native leadership

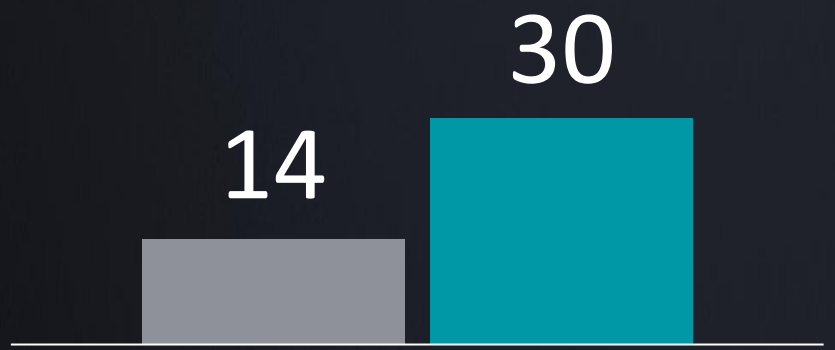
Up to **2.6x** more performance



Source: https://www.amd.com/system/files/documents/amd-epyc-9004-pb-cloud-native-workloads.pdf.

Optimized for cloud native deployments

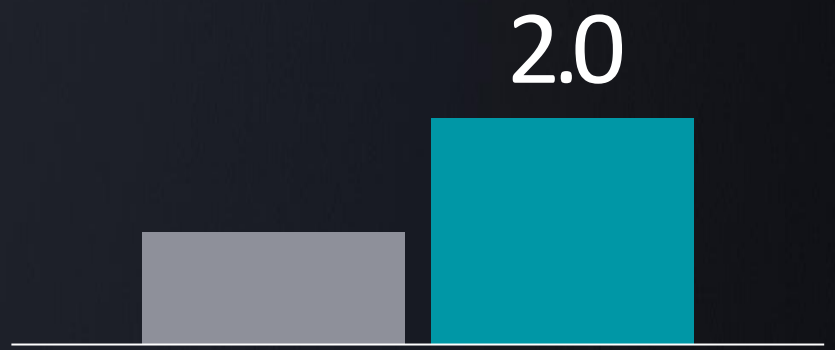
Up to **2.1x** container density per server



Total Containers

Equivalent Ops Per Sec Per Container

Up to **2.0x** the efficiency



Overall Server-Side Java® Ops Per Watt

SPECpower_ssj® 2008

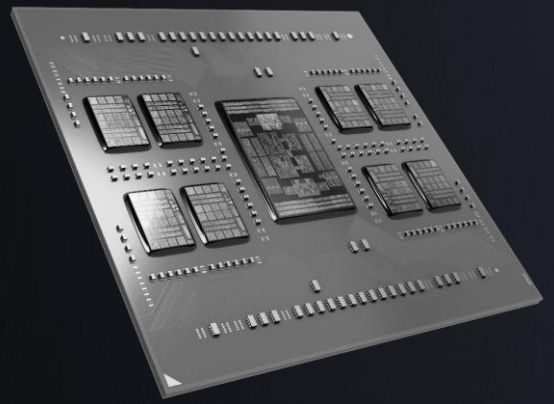
4th Gen Intel® Xeon® Platinum **8490H**

4th Gen AMD EPYC™ **9754**

Claims: SP5-149

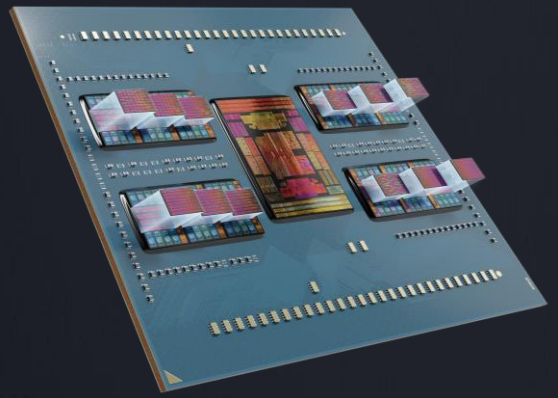
Source: https://www.amd.com/system/files/documents/amd-epyc-9004-pb-spec-power.pdf.

Computing infrastructure
optimized for data center workloads



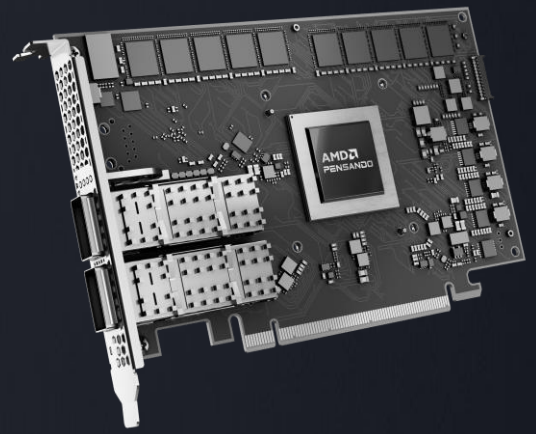
Cloud Native Computing

AMD EPYC



Technical Computing

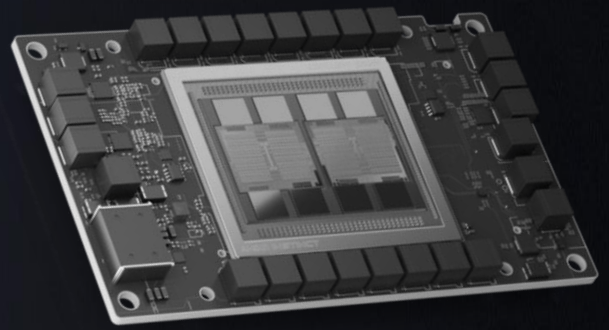
AMD EPYC



Networking

AMD ALVEO

AMD PENSANDO



AI

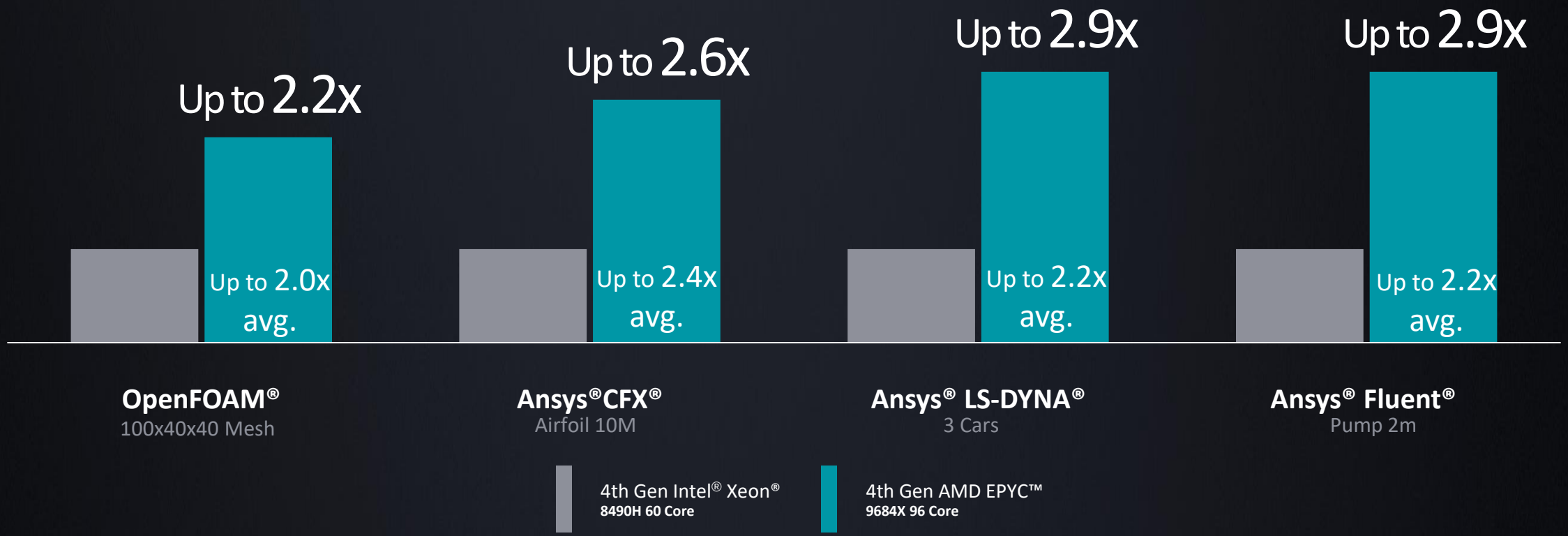
AMD INSTINCT

AMD EPYC

AMD VERSAL

Performance leadership for technical computing

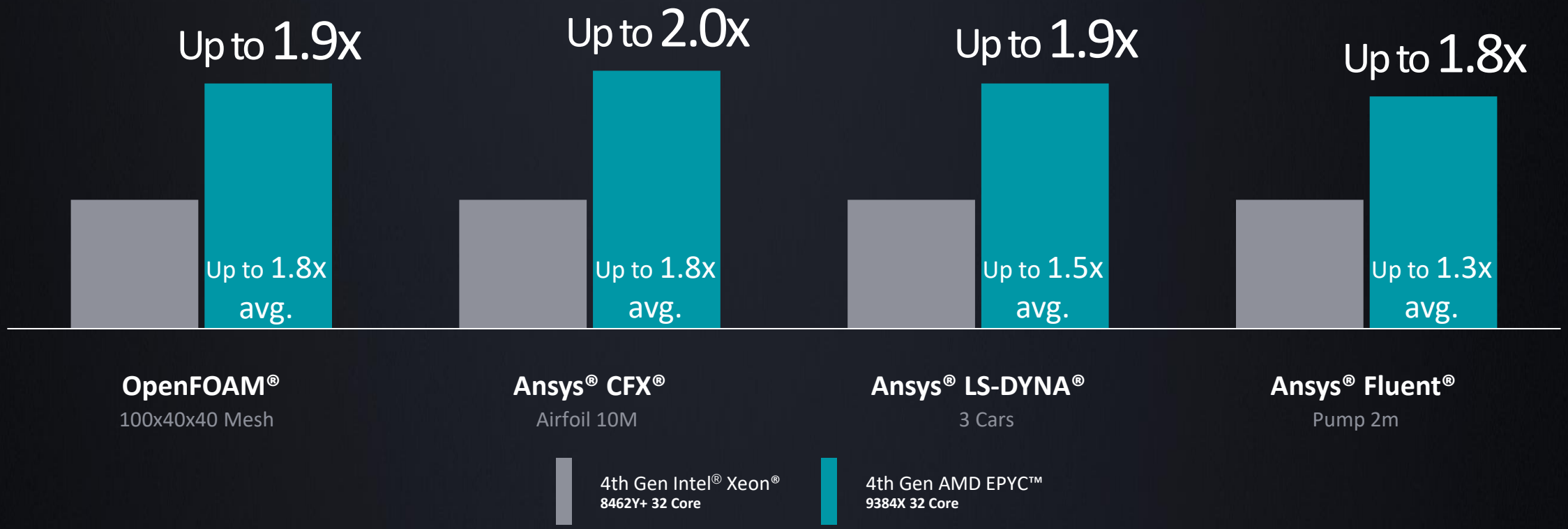
CFD and **FEA** | **Top-of-stack** Comparison



Sources: ANSYS CFX https://www.amd.com/system/files/documents/amd-epyc-9004x-pb-ansys-cfx.pdf, ANSYS LS-DYNA https://www.amd.com/system/files/documents/amd-epyc-9004x-pb-ansys-ls-dyna.pdf, ANSYS Fluent https://www.amd.com/system/files/documents/amd-epyc-9004x-pb-ansys-fluent.pdf, OpenFOAM https://www.amd.com/system/files/documents/amd-epyc-9004x-pb-openfoam.pdf.

Performance leadership for technical computing

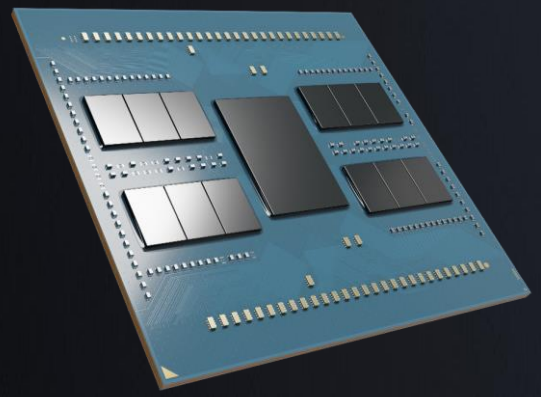
CFD and **FEA** | **32-Core** Comparison



Sources: ANSYS CFX https://www.amd.com/system/files/documents/amd-epyc-9004x-pb-ansys-cfx.pdf, ANSYS LS-DYNA https://www.amd.com/system/files/documents/amd-epyc-9004x-pb-ansys-ls-dyna.pdf, ANSYS Fluent https://www.amd.com/system/files/documents/amd-epyc-9004x-pb-ansys-fluent.pdf, OpenFOAM https://www.amd.com/system/files/documents/amd-epyc-9004x-pb-openfoam.pdf.

4th Gen AMD EPYC™

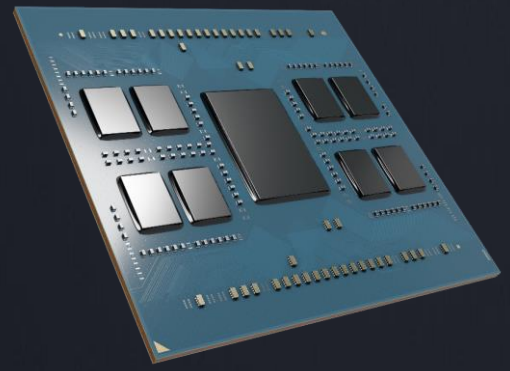
Leadership across segments



General Purpose Computing

“Genoa”

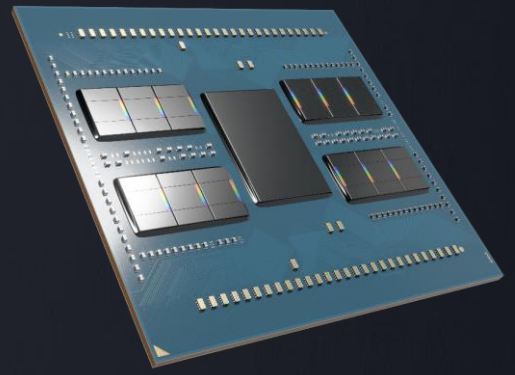
Available Now



Cloud Native Computing

“Bergamo”

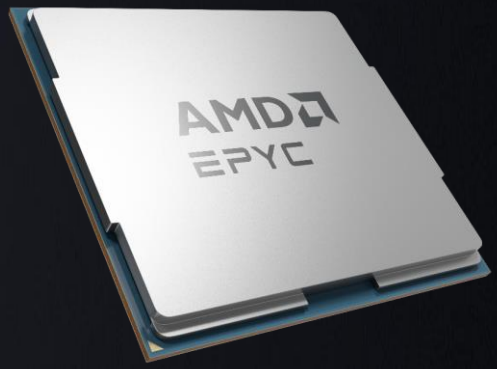
Available Now



Technical Computing

“Genoa-X”

Available Now



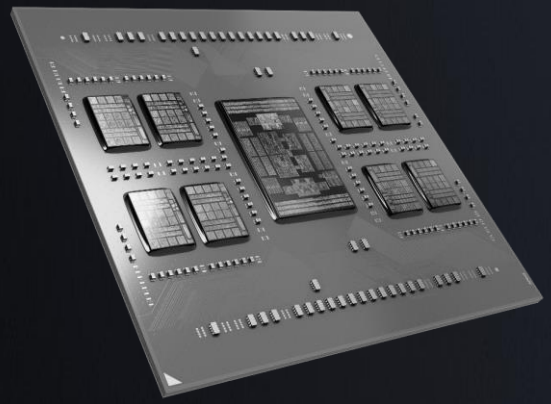
Telco/Edge Computing

“Siena”

Available 2H23

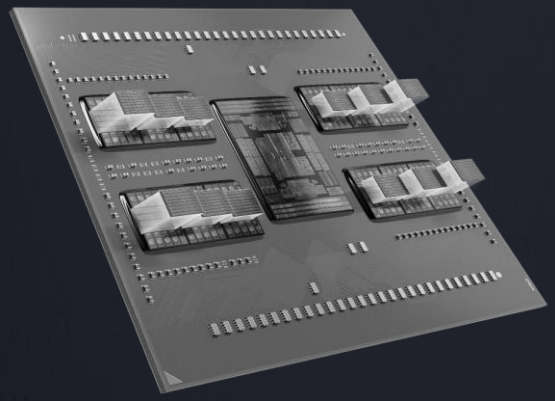
Dates are subject to change

Computing infrastructure
optimized for data center workloads



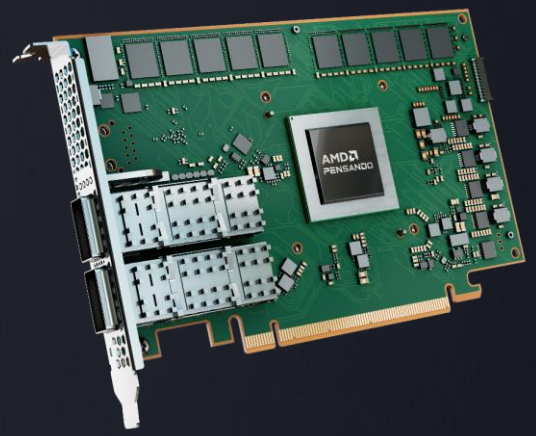
Cloud Native Computing

AMD EPYC



Technical Computing

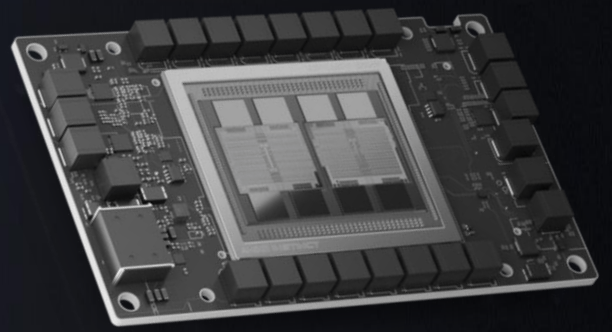
AMD EPYC



Networking

AMD ALVEO

AMD PENSANDO

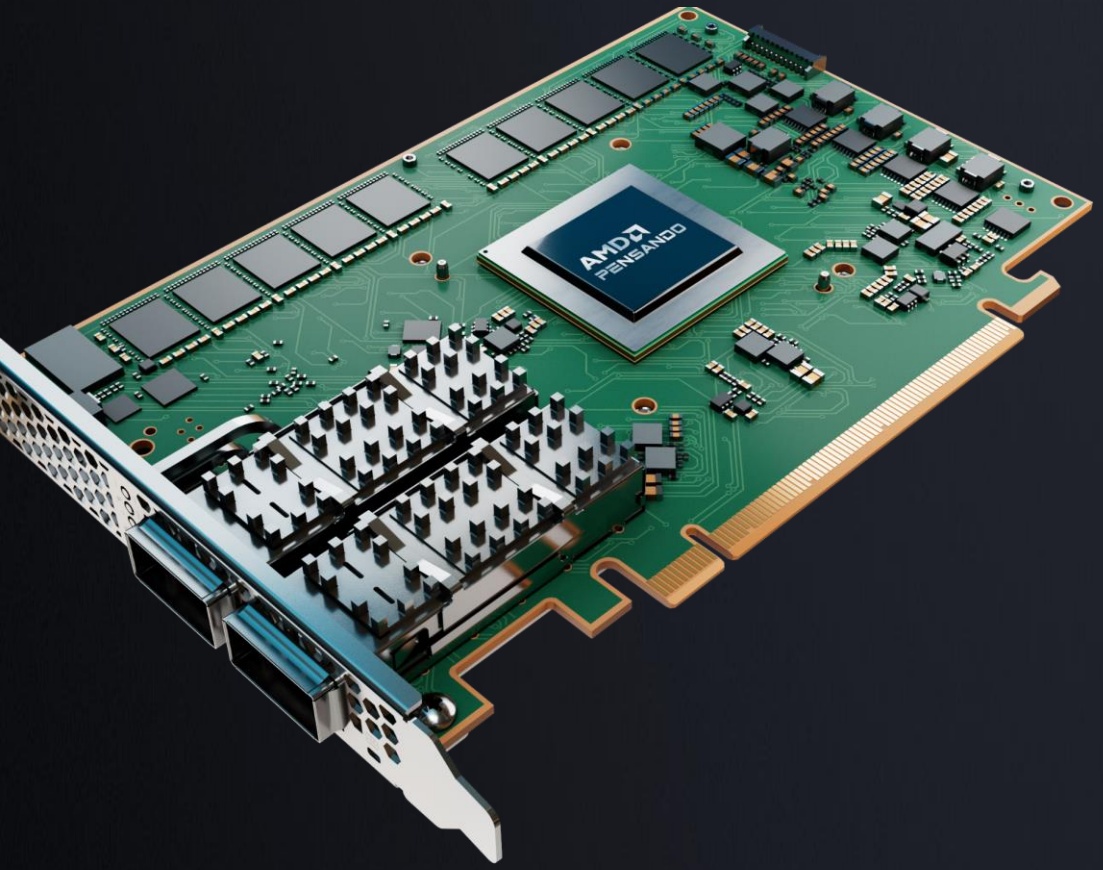


AI

AMD INSTINCT

AMD EPYC

AMD VERSAL



AMD Pensando™ SmartNICs

Offloads cloud / virtualization overhead

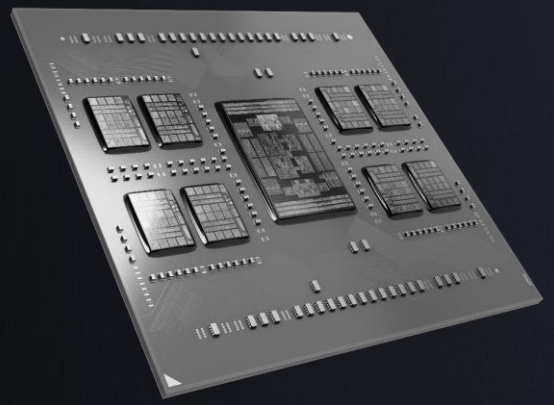
Dramatically enhances security and visibility

Enables broad range of infrastructure services offload

Deployed in major public clouds; available as VMware® vSphere® solutions

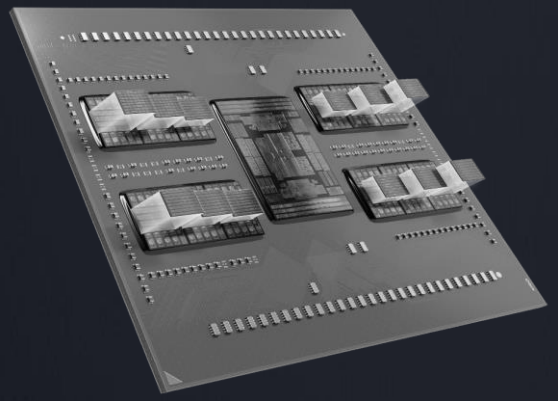
DELL Technologies Goldman Sachs Hewlett Packard Enterprise IBM Cloud ORACLE CLOUD

Computing infrastructure
optimized for data center workloads



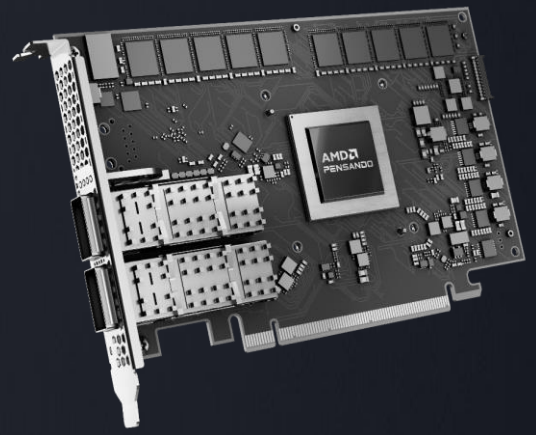
Cloud Native Computing

AMD EPYC



Technical Computing

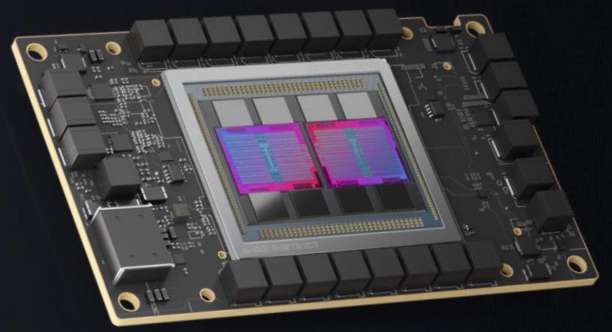
AMD EPYC



Networking

AMD ALVEO

AMD PENSANDO



AI

AMD INSTINCT

AMD EPYC

AMD VERSAL

AMD AI Platforms

Open
software approach

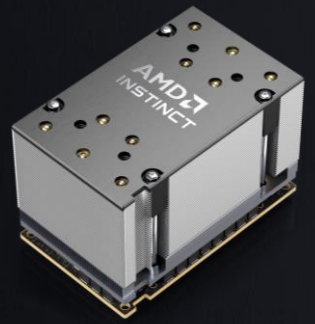
Proven
AI capabilities

Ready
support for AI models

AMD
AI Platforms

Training and inference portfolio

Data center | Edge | End point

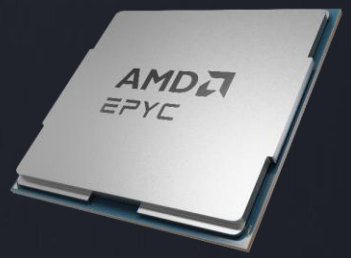


AMD Instinct™ Accelerators

HPC and data center training and inference

AMD Alveo™ Accelerators

Data center and edge inference



4th Gen AMD EPYC™ Processors

CPU AI leadership

AMD Embedded Versal™ AI Edge

AI + sensor embedded inference

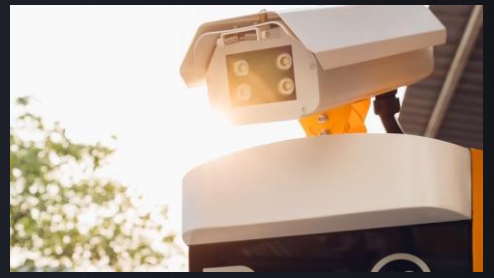
AMD Ryzen™ 7040 Mobile Processors

Ryzen™ AI inference engine for select Windows PCs

AMD
AI Platforms

Powering inference from edge to endpoint




Aerospace | Automotive | Healthcare | Industrial | PCs

acer ABB AISIN ASUS CERN clarius DELL hp innodisk Lenovo

kakao i cloud Microsoft NASA Orchestrating a brighter world NEC SK telecom SUBARU Tattile Custom Vision Solutions veoneer ZF

AMD EPYC | INSTINCT

Powering datacenter AI at scale

TOP500 The List. | #1 Frontier

National Cancer Institute and DOE accelerating cancer research and treatment

TOP500 The List. | #3 LUMI

Largest Finnish language model (TurkuNLP-13B)

AI2 OLMo

Allen Institute scientific LLM

TOP500 The List. | #11 Explorer

WUS3 running AI and HPC workloads

1st Korean LLM

T5 NLP with 11B parameters

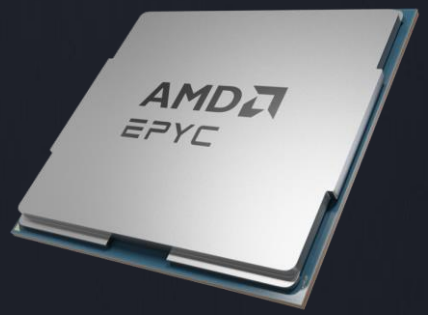
AMD AI Platforms

ROCm

Data center GPU

ZenDNN

Data center CPU



Vitis AI

Edge and end points

Optimized AI software stack

AI Models and **Algorithms**

PyTorch TensorFlow ONNX

AI Ecosystem optimized for AMD

Libraries

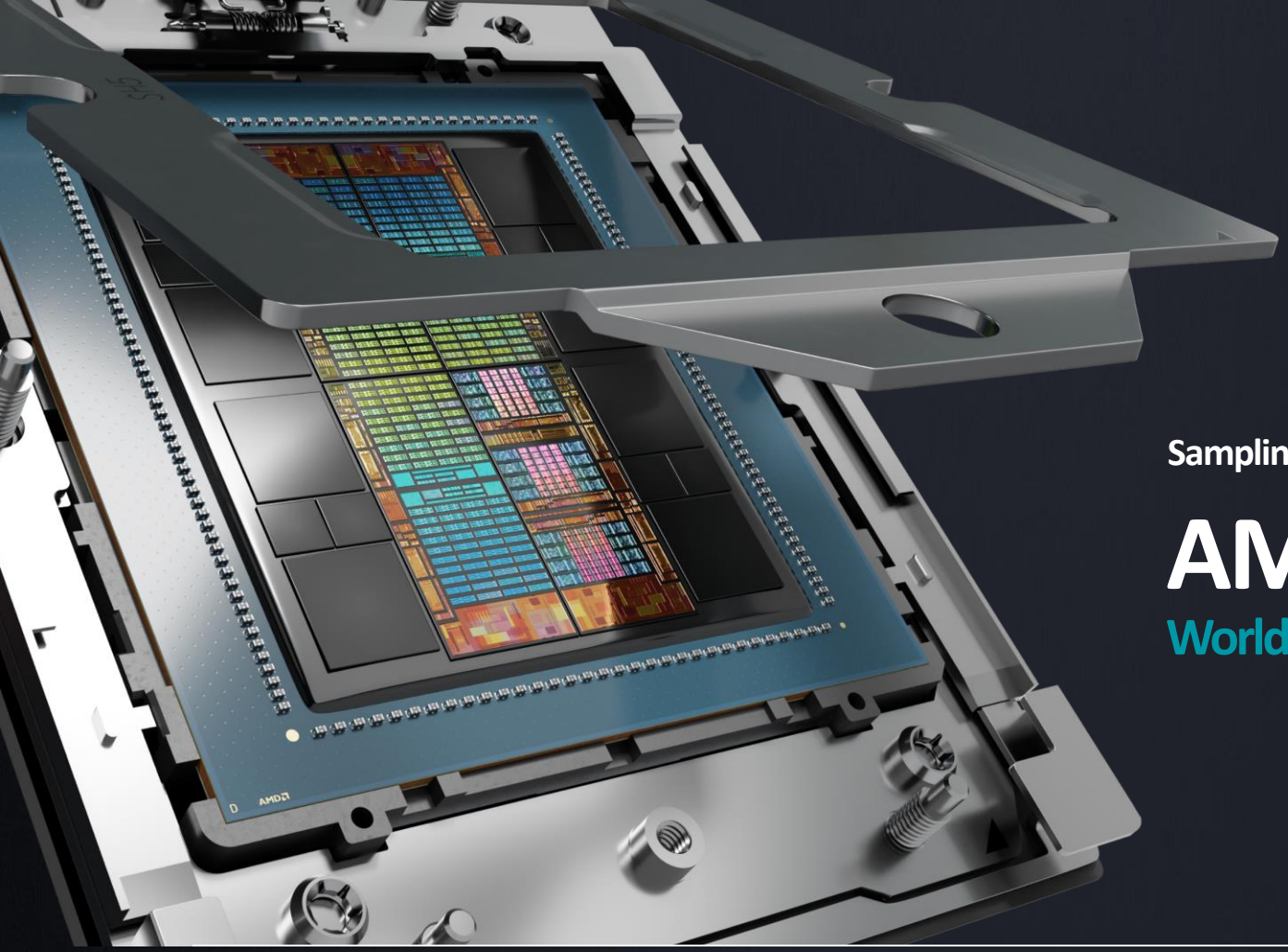
Compilers and **Tools**

Runtime

AMD ROCm | A proven software stack

AMD Instinct™ GPU

AMD INSTINCT | Leadership performance



Sampling now

AMD Instinct™ MI300A

World's first APU accelerator for AI and HPC

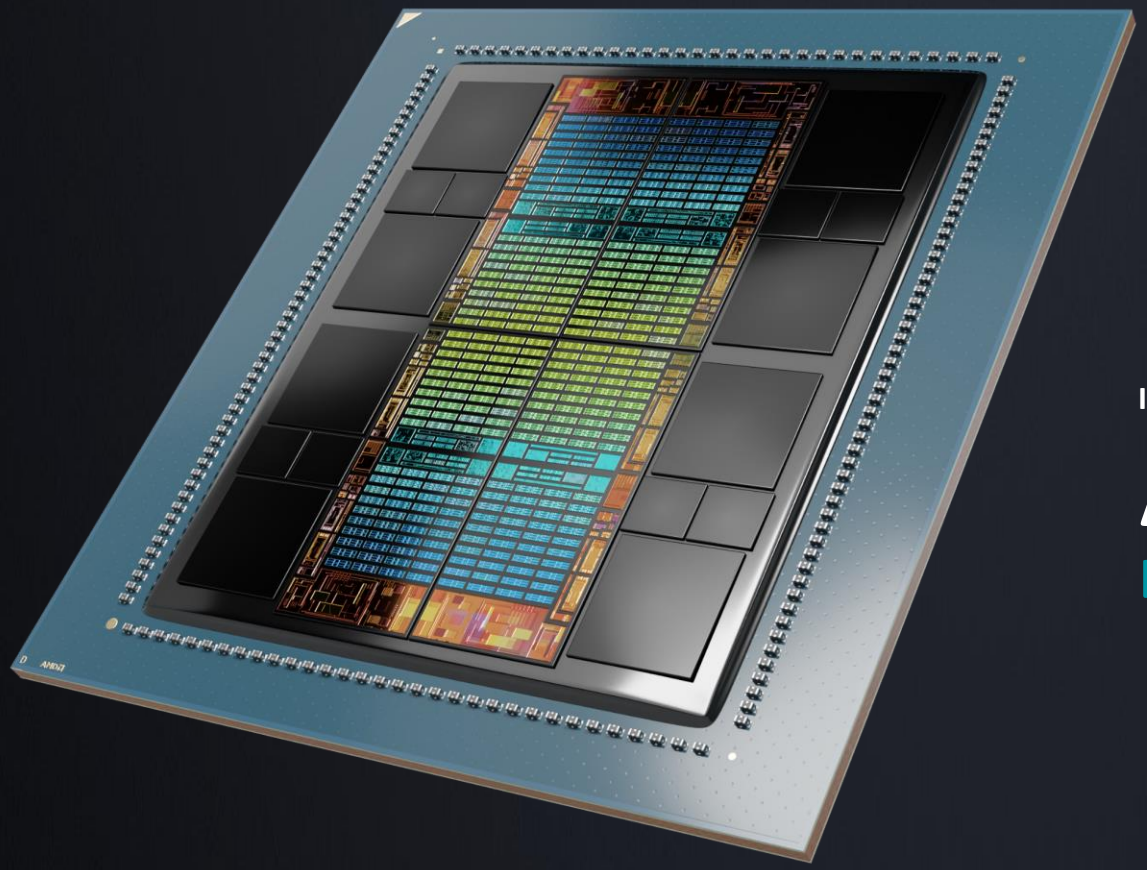
AMD CDNA 3 Next-Gen Accelerator Architecture

ZEN 4 24 CPU Cores

128 GB HBM3

5nm and **6**nm Process Technology

Shared Memory CPU + GPU



Introducing today

AMD Instinct™ MI300X

Leadership generative AI accelerator

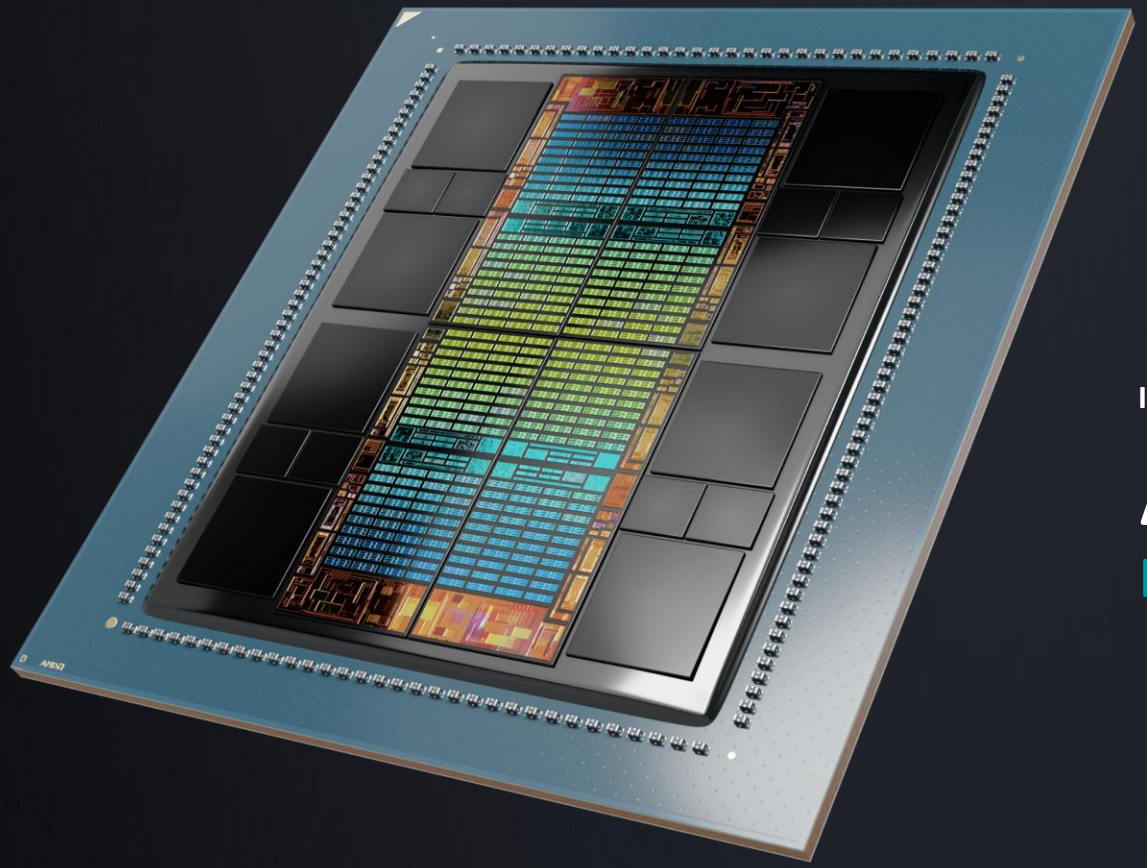
AMD CDNA 3

192 GB
HBM3

5.2 TB/s
Memory Bandwidth

896 GB/s
Infinity Fabric™ Bandwidth

153 B
Transistors



Introducing today

AMD Instinct™ MI300X

Leadership generative AI accelerator

Up to

2.4x HBM density

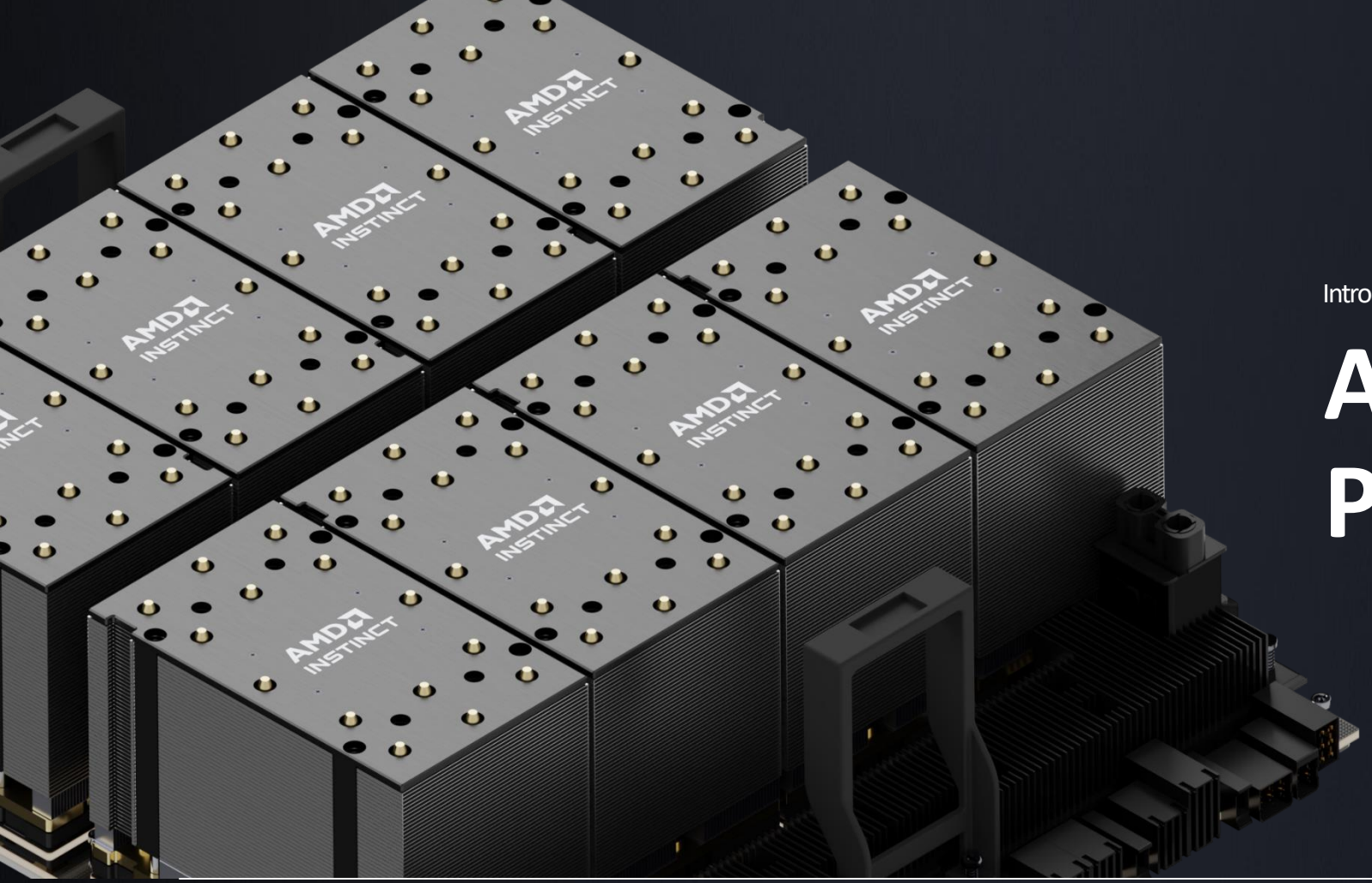
compared to Nvidia H100

Up to

1.6x HBM bandwidth

compared to Nvidia H100

See Endnotes: MI300-05



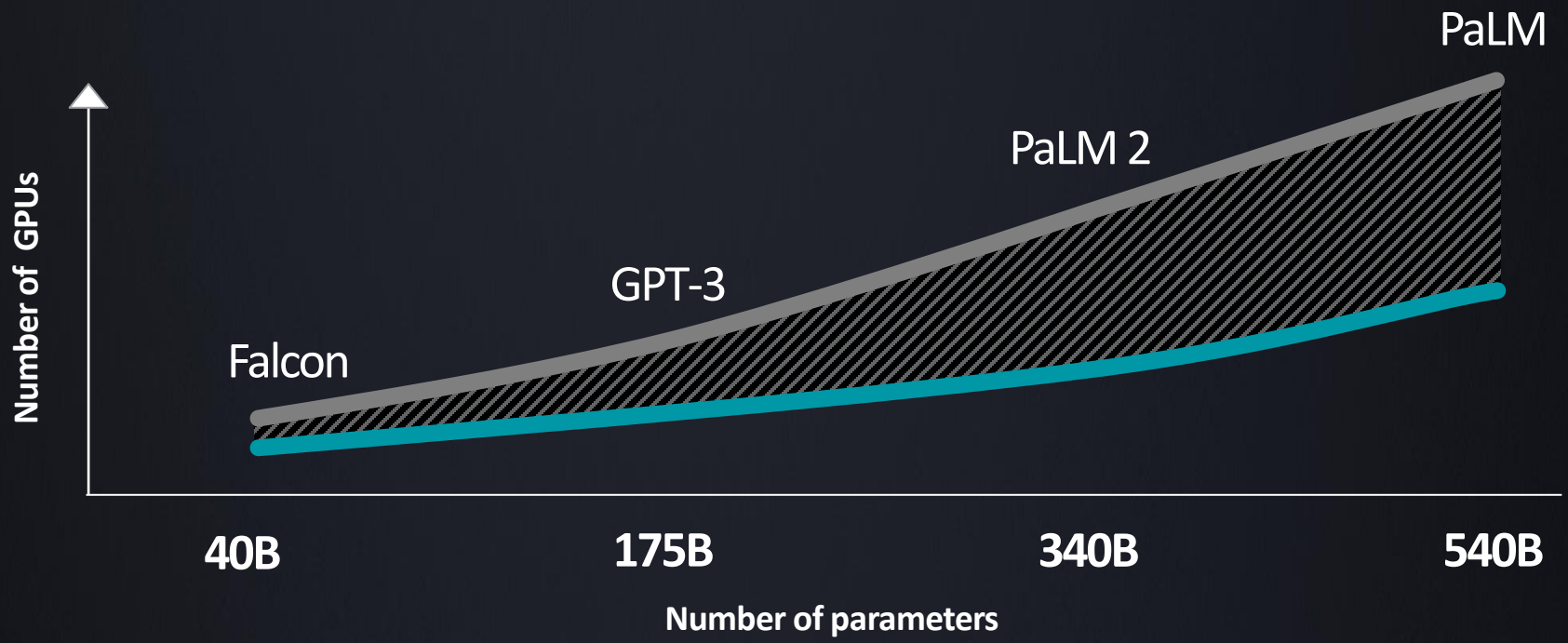
Introducing today

AMD Instinct™ Platform

8x MI300X | **1.5 TB** HBM3 Memory | Industry-Standard Design

AMD Instinct™ MI300X

Inference advantage

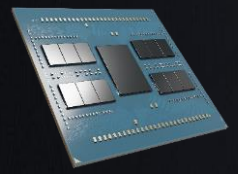


Competition 80GB

AMD Instinct™ MI300X 192GB

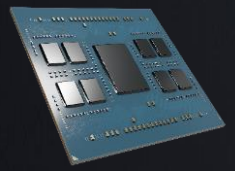
See Endnote: MI300-08K

AMD together we advance_



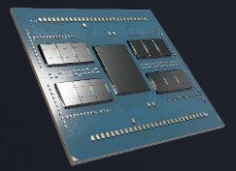
General Purpose Computing

4th Gen EPYC™ CPU "Genoa"



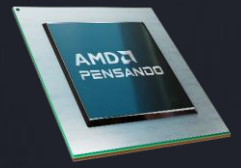
Cloud Native Computing

4th Gen EPYC™ CPU "Bergamo"



Technical Computing

4th Gen EPYC™ CPU "Genoa-X"



Networking Pensando P4 DPU

Cloud Efficiency for Enterprise

MI300A

Sampling now

MI300X

Sampling in Q3

CPU AI leadership

100s of embedded AI inference customers

Broadest AI-powered PC portfolio

Open | Proven | Ready

AI Software


AMD
together we advance_

Endnotes

SP5TCO-036A: As of 05/19/2023 based on AMD Internal analysis using the AMD EPYC™ Server Virtualization & Greenhouse Gas Emission TCO Estimation Tool - version 12.15 estimating the cost and quantity of 2P AMD 96 core EPYC™ 9654 powered server versus 2P Intel® Xeon® 60 core Platinum 8490H based server solutions required to deliver 2000 total virtual machines (VM), requiring 1 core and 8GB of memory per VM for a 3-year period. This includes VMware software license cost of $6,558.32 per socket + one additional software for every 32 CPU core increment in that socket. Environmental impact estimates made leveraging this data, using the Country / Region specific electricity factors from the '2020 Grid Electricity Emissions Factors v1.4 – September 2020', and the United States Environmental Protection Agency 'Greenhouse Gas Equivalencies Calculator'. 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. For additional details, see https://www.amd.com/en/claims/epyc4#SP5TCO-036A.

MI300-08K - Measurements by internal AMD Performance Labs as of June 2, 2023 on current specifications and/or internal engineering calculations. Large Language Model (LLM) run comparisons with FP16 precision to determine the minimum number of GPUs needed to run the Falcon (40B parameters); GPT-3 (175 Billion parameters), PaLM 2 (340 Billion parameters); PaLM (540 Billion parameters) models.Calculated estimates based on GPU-only memory size versus memory required by the model at defined parameters plus 10% overhead.

Calculations rely on published and sometimes preliminary model memory sizes. Tested result configurations: AMD Lab system consisting of 1x EPYC 9654 (96-core) CPU with 1x AMD Instinct™ MI300X (192GB HBM3, OAM Module) 750W accelerator Vs. Competitve testing done on Cirrascale Cloud Services comparable instance with permission.

Results (FP16 precision):	Model:	Parameters	Tot Mem. Reqd	MI300X Reqd	Competition Reqd
Falcon-40B	40 Billion	88 GB		1 Actual	2 Actual
GPT-3	175 Billion	385 GB		3 Calculated	5 Calculated
PaLM 2	340 Billion	748 GB		4 Calculated	10 Calculated
PaLM	540 Billion	1188 GB		7 Calculated	15 Calculated

Calculated estimates may vary based on final model size; actual and estimates may vary due to actual overhead required and using system memory beyond that of the GPU. Server manufacturers may vary configuration offerings yielding different results.

Endnotes

- SP5-011E: SPECpower_ssj®2008 comparison based on published 2P server results as of 6/13/2023. Configurations: 2P AMD EPYC 9654 (30,602 overall ssj_ops/W, 2U, https://spec.org/power_ssj2008/results/res2022q4/power_ssj2008-20221204-01204.html) is 1.81x the performance of best published 2P Intel Xeon Platinum 8490H (16,902 overall ssj_ops/W, 2U, https://spec.org/power_ssj2008/results/res2023q2/power_ssj2008-20230507-01251.html). SPEC® and SPECpower_ssj® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information.
- SP5-049C: VMmark® 3.1.1 matched pair comparison based on published results as of 6/13/2023. Configurations: 2-node, 2P 96-core EPYC 9654 powered server running VMware ESXi 8.0b (40.66 @ 42 tiles/798 VMs, https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmmark/2023-06-13-Lenovo-ThinkSystem-SR665V3.pdf) versus 2-node, 2P 60-core Xeon Platinum 8490H running VMware ESXi 8.0 GA (23.38 @ 23 tiles/437 VMs, https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmmark/2023-03-21-Fujitsu-PRIMERGY-RX2540M7.pdf) for 1.74x the score and 1.75x the tile (VM) capacity. 2-node, 2P EPYC 7763-powered server (23.33 @ 24 tiles/456 VMs, https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmmark/2022-02-08-Fujitsu-RX2450M1.pdf) shown at 0.98x performance for reference. VMmark is a registered trademark of VMware in the US or other countries.
- SP5-051: TPCx-AI SF3 derivative workload comparison based on AMD internal testing running multiple VM instances as of 6/13/2023. 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. Configurations: 2 x AMD EPYC 9754 on Titanite (BIOS and Settings: AMI Core Ver. 5.25, Project Ver. RTI1000F and Default BIOS settings (SMT=on, Determinism=Auto, NPS=1)), 1.5TB (24) Dual-Rank DDR5-4800 64GB DIMMs, 1DPC, SK Hynix SHGP31-500GM 500GB NVMe, Ubuntu® 22.04 LTS (8-instances, 30 vCPUs/instance, 1841 AI test cases/min); 2 x AMD EPYC 9654 on Titanite (BIOS and Settings: AMI Core Ver. 5.25, Project Ver. RTI1000F and Default BIOS settings (SMT=on, Determinism=Auto, NPS=1)), 1.5TB (24) Dual-Rank DDR5-4800 64GB DIMMs, 1DPC, Samsung SSD 983 DCT 960GB, Ubuntu 22.04.1 LTS (6-instance, 28 vCPUs/instance, 1554 AI test cases/min); 2 x Intel(R) Xeon(R) Platinum 8490H on Dell PowerEdge R760 (BIOS and Settings: ESE110Q-1.10 and Package C1E, Default BIOS settings (C State=Disabled, Hyper-Threading, Turbo boost= enabled (ALL)=Enabled, SNC (Sub NUMA)=Disabled)), 2TB (32) Dual-Rank DDR5-4800 64GB DIMMs, 1DPC, Dell 1.7TB NVMe, Ubuntu 22.04.2 LTS (4-instance, 30 vCPUs/instance, 831 AI test cases/min). Results may vary due to factors including system configurations, software versions and BIOS settings. TPC Benchmark is a trademark of the TPC.
- SP5-056B: SAP® SD 2-tier comparison based on published results as of 6/13/2023. Configurations: 2P 96-core EPYC 9654 powered server (148,000 benchmark users, https://www.sap.com/dmc/benchmark/2022/Cert22029.pdf) versus 2P 60-core Xeon Platinum 8490H (77,105 benchmark users, https://www.sap.com/dmc/benchmark/2023/Cert23021.pdf) for 1.92x the number of SAP SD benchmark users. 2P EPYC 7763 powered server (75,000 benchmark users, https://www.sap.com/dmc/benchmark/2021/Cert21021.pdf) shown at 0.98x the performance for reference. For more details see http://www.sap.com/benchmark. SAP and SAP logo are the trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and in several other countries.
- SP5-104A: SPECjbb® 2015-MultiJVM Critical based on published scores from www.spec.org as of 3/31/2023. Configurations: 2P AMD EPYC 9654 (664,375 SPECjbb®2015 MultiJVM max-jOPS, 622,315 SPECjbb®2015 MultiJVM critical-jOPS, 192 Total Cores, https://www.spec.org/jbb2015/results/res2022q4/jbb2015-20221019-00860.html) is 1.69x the critical-jOPS performance of published 2P Intel Xeon Platinum 8490H (458,295 SPECjbb®2015 MultiJVM max-jOPS, 368,979 SPECjbb®2015 MultiJVM critical-jOPS, 120 Total Cores, http://www.spec.org/jbb2015/results/res2023q1/jbb2015-20230119-01007.html).
- SP5-149: Container density throughput based on sustaining ~25k e-commerce Java Ops/sec/container until exceeding SLA utilizing >90% of the total cores on composite server-side Java workload as measured by AMD as of 6/13/2023. Common container settings: allocated 40GB memory, similar disks & NICs. 2P server configurations: 2P EPYC 9754 128C/256T SMT ON, Memory: 1.5TB = 24 x 64 GB DDR5 4800, OS Ubuntu 22.04, NPS Setting: L3 as NUMA running 16 vCPUs vs. 2P Xeon Platinum 8490H 60C/120T HT ON, Memory: 2TB = 32 x 64 GB DDR5 4800, OS Ubuntu 22.04, NPS Setting: NPS 2 running 16 vCPUs vs. 2P Ampere Altra Max 128-30, Memory: 1TB =16 x 64GB DDR3200, OS Ubuntu 22.04, NPS Setting: NPS 1 running 25C. Results may vary due to factors including system configurations, software versions and BIOS settings.
- MI300-005: Calculations conducted by AMD Performance Labs as of May 17, 2023, for the AMD Instinct™ MI300X OAM accelerator 750W (192 GB HBM3) designed with AMD CDNA™ 3 5nm FinFet process technology resulted in 192 GB HBM3 memory capacity and 5.218 TFLOPS sustained peak memory bandwidth performance. MI300X memory bus interface is 8,192 and memory data rate is 5.6 Gbps for total sustained peak memory bandwidth of 5.218 TB/s (8,192 bits memory bus interface * 5.6 Gbps memory data rate/8)*0.91 delivered adjustment. The highest published results on the NVidia Hopper H100 (80GB) SXM GPU accelerator resulted in 80GB HBM3 memory capacity and 3.35 TB/s GPU memory bandwidth performance.

Disclaimers and Attributions

The information contained herein is for informational purposes only and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD’s products are as set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale. GD-18