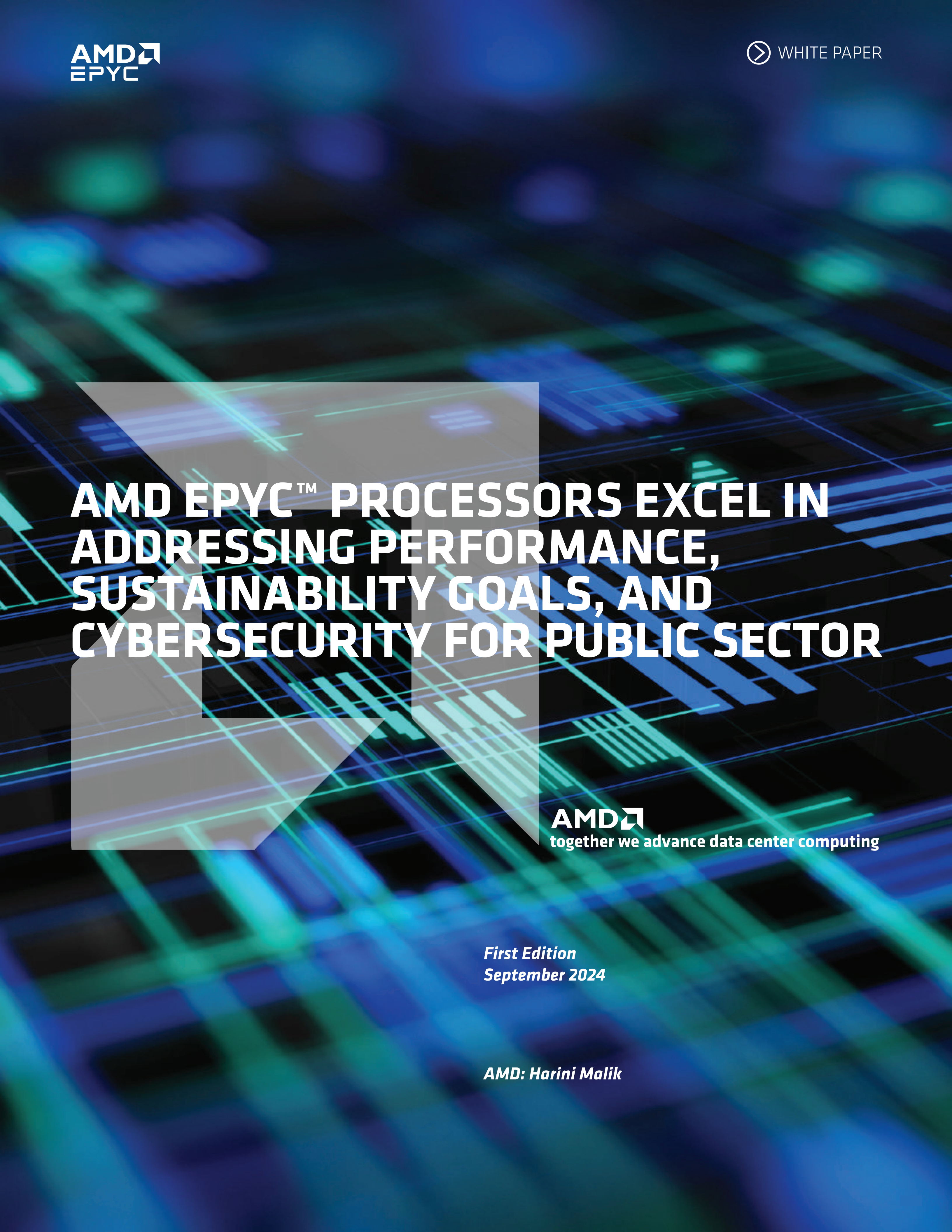
WHITE PAPER



AMD EPYC™ PROCESSORS EXCEL IN ADDRESSING PERFORMANCE, SUSTAINABILITY GOALS, AND CYBERSECURITY FOR PUBLIC SECTOR

together we advance data center computing

First Edition
September 2024

AMD: Harini Malik

AMD EPYC™ PROCESSORS EXCEL IN ADDRESSING PERFORMANCE, SUSTAINABILITY GOALS, AND CYBERSECURITY FOR PUBLIC SECTOR

CONTENTS

WHY INFRASTRUCTURE MATTERS

Public sector organizations face enormous challenges. Whether in government, education, healthcare, or research, organizations face growing demands on infrastructure and challenges related to cyberthreats, climate change, increasing complexity, and limited budgets. These challenges are compounded by factors such as increasing public expectations around online service delivery, an aging workforce, procurement challenges, and the critical need to meet sustainability goals which could include transitioning to low-carbon energy sources. Also top of mind is the emergence of generative AI and the challenges it brings related to public policy, ethics, and concerns about misinformation.

Despite these challenges, digital transition initiatives enabled by new technologies offer tremendous promise. By embracing transformative technologies such as AI for the public good, organizations can increase operational efficiencies, improve essential public services, offer new modes of service delivery, and provide new economic opportunities.

This paper explains why having the right technology foundation for public sector applications is critical. By deploying infrastructure powered by the latest AMD EPYC processors, organizations can increase application performance, address sustainability goals, and enhance their security posture against a wide range of cybersecurity threats while reducing their total cost of ownership (TCO).

CRITICAL INFRASTRUCTURE REQUIREMENTS FOR THE PUBLIC SECTOR

Whether operating on-premises or in public or private clouds, public sector organizations need high-performance infrastructure that is scalable, energy-efficient , cost-effective, and secure. Six critical infrastructure requirements illustrated in Figure 1 are presented below:

Performance is a critical driver of productivity across all workloads, from databases to virtualization solutions to emerging applications such as AI model training and inference. Public sector organizations need high-performance systems to increase capacity and effciency and deliver high-quality services to their constituents.

Given concerns about global warming fueled by greenhouse gas (GHG) emissions, **sustainability and energy efficiency** are major concerns for all organizations. According to Electricity 2024, published by the International Energy Agency (IEA), global electricity demand from data centers could double towards 2026 from a 2022 baseline of 460 TWh to between 620 and 1,050 TWh.[1] Moreover, the impact of AI on power consumption is only expected to grow. "Gartner® predicts more than 70% of government agencies will use AI to enhance human administrative decision-making by 2026."[2] Most public sector organizations have sustainability goals aimed at reducing direct and indirect GHG emissions, and energy-efficient infrastructure is critical to meeting these objectives.

While **security** is critical for all organizations, this is especially true for governments and public utilities that hold sensitive information, deliver essential services, and manage critical infrastructure. These organizations are high on the target list of many malicious actors.

The Center for Strategic & International Studies (CSIS), based in Washington, D.C., reports major worldwide cybersecurity incidents across governments, the defense industry, and high-tech companies. They have recorded hundreds of attacks with economic impacts of over **USD 1M per incident**—and these are just the reported incidents.[3]

Public sector organizations need to guard against a range of increasingly sophisticated threats, including malware, ransomware, phishing, social engineering, and distributed denial of service (DDoS) attacks. Adopting a "zero-trust" framework and security mindset, conducting regular security assessments and audits, and selecting infrastructure with end-to-end encryption features rooted in silicon are essential to help guard against these threats.

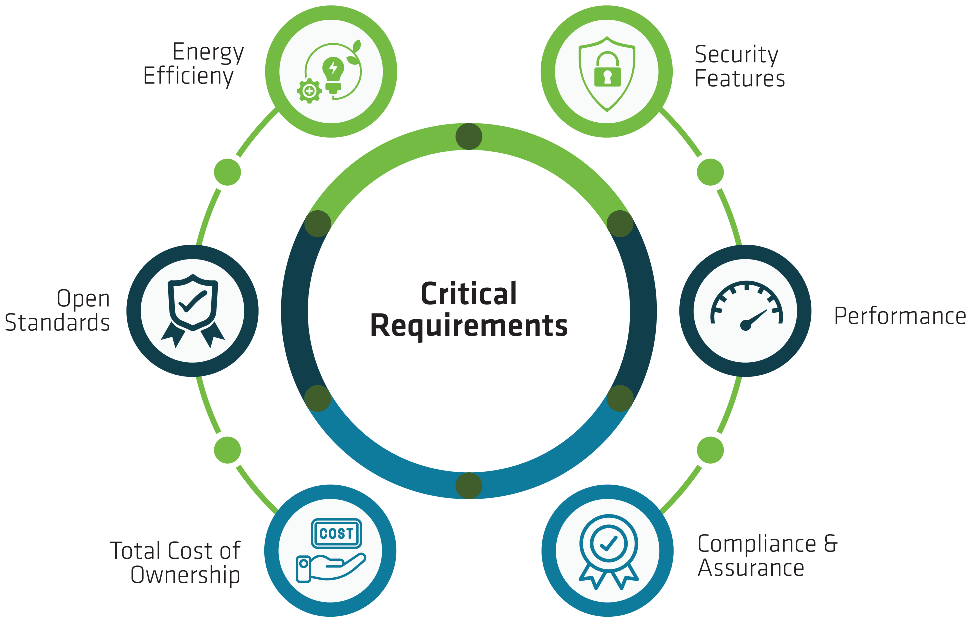


Figure 1 - Critical requirements for public sector workloads

With limited budgets, all organizations are concerned about **total cost of ownership (TCO)**, and performance and energy efficiency are key drivers. Next to infrastructure and staffing, power and cooling are the largest cost drivers in data centers. Organizations can meet operational requirements with fewer physical servers or cloud instances by deploying servers based on high-performance, energy-efficient processors. This can translate into less data center space, reduced power and cooling costs, and lower ancillary costs related to management, rack infrastructure, and network drops.

Compliance, assurance, and interoperability are also essential requirements for the public sector infrastructure. Whether operating on-premises or deploying services in the cloud, US federal departments have compliance mandates, including ITAR regulations, FedRAMP, Defense Federal Acquisition Regulation Supplement (DFARS), and FIPS requirements. European organizations have similar requirements. Public organizations may also be subject to privacy regulations depending on where they operate and the type of data they process.

Finally, selecting systems based on **open standards** is also critical. Open standards help ensure interoperability and portability and can reduce the risk of proprietary lock-in. They also help reduce costs and simplify procurement by allowing organizations to specify and source standards-based components from multiple vendors in a competitive marketplace. Open standards are not the same as open source, which refers to the availability of the source code for software. An open ecosystem does not require you to share your proprietary code, technology, or data but rather to adhere to common standards and protocols that enable seamless integration and communication among different systems. The greater the number of entities participating in an open ecosystem, the greater the interoperability and resulting benefits to innovation and security.

AMD EPYC – THE IDEAL PROCESSOR FOR PUBLIC SECTOR

Public sector organizations run a wide range of applications and workloads. Examples include:

- General-purpose application workloads in areas such as healthcare, education, taxation, social services, judicial, transportation & infrastructure related applications
- Virtualization and virtual desktop integration (VDI) solutions
- Cloud-native applications for online service delivery
- Database and analytic applications built on structured and unstructured databases
- Enterprise HPC & AI applications in areas such as research, aerospace & defense, life sciences, climate research, and AI model training and inference

Organizations need a secure, scalable, cost-effective technology platform that can run a full range of application software while also addressing other critical requirements.

THE EPYC ADVANTAGE

First introduced in June 2017, AMD EPYC processors bring together high core counts, large memory capacity, high bandwidth, and massive I/O to enable exceptional performance for a wide range of workloads.

EPYC is based on a hybrid multi-die architecture first introduced in 2nd Gen EPYC processors. Realizing that increasing core density in monolithic designs would become more difficult with time, AMD engineers pioneered an innovative design where CPU cores and I/O functions were implemented on separate dies using different fabrication technologies.

Known as "chiplets", this modular approach enabled AMD to mix and match CPU and I/O dies to address a variety of application requirements. In the latest 4th Gen EPYC processors, 'Zen 4' and 'Zen 4c' cores are produced with 5nm technology, and I/O dies are created with a 6nm process. Processor cores are combined with cache into a core complex (CCX), and these core complexes are fabricated onto a CPU die.[4]

While the number of CPU dies and cores varies by processor, Figure 2 shows the architecture of the top-of-stack AMD EPYC™ 9754 processor. In this configuration, 8 x 'Zen 4c' CPU dies, each with 16 'Zen 4c' cores and 32MB of shared L3 cache per core complex, are combined on a single AMD EPYC 9004 series processor providing 128 cores, 256 threads, and 256 MB of L3 cache for unprecedented density for hyper-converged infrastructure (HCI) and virtualization applications.

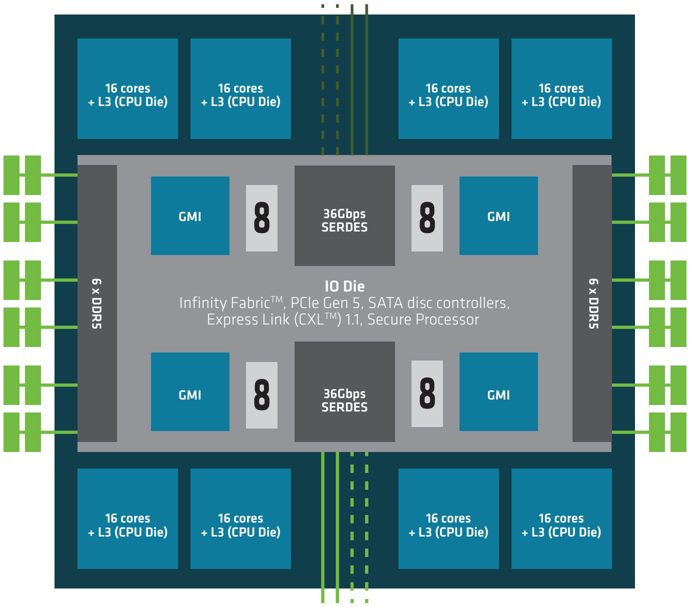


Figure 2 - AMD EPYC 9754 architecture diagram

4th Gen EPYC processors*

- Next generation 5nm technology
- Up to 128 cores/256 threads
- Up to 4.40 GHz Max boost clock speed[5]
- Up to 2 x 16MB = 32MB L3 cache per CPU die
- Up to 1152 MB L3 cache with 3D V-Cache[6]
- Infinity Fabric™ - Up to 72 Gb/s connections[7]
- 12 DDR5 memory channels per socket
- Up to 4800 MT/s DDR
- 128 PCIe® Gen 5 lanes per socket**
- CXL™ 1.1+ memory expansion
- Embedded AMD secure processor

* Specifications vary by processor SKU

** Max of 160 PCIe® lanes in 2P configurations

With this modular chiplet-based approach, AMD can mix and match IP elements from CPUs and GPUs to provide AMD Instinct™ MI300 APUs supporting both CPU- and GPU-intensive workloads. It also helps AMD achieve economies of scale to minimize costs. For applications that benefit from large amounts of cache per core, AMD offers 9004 series processors with 3D V-Cache™ technology, featuring up to 1152 MB of L3 cache per CPU.

The I/O dies used in all 4th Gen AMD EPYC processors support 12 Infinity Fabric connections. Depending on the processor, one or two connections are made between each CPU die and the I/O die via the Global Memory Interface (GMI), providing up to 36 Gb/s of connectivity.

The I/O die also supports extensive connectivity, including support for 128 PCIe Gen5 channels (up to 160 PCIe channels in a 2 socket configuration), 8 bonus PCI3 Gen 3 lanes, SATA disk controllers, and Compute Express Link (CXL™) 1.1 interconnects.

For public sector clients, AMD EPYC processors provide:

- Exceptional performance for a wide range of workload
- Cutting-edge security features with AMD Infinity Guard™[8]
- Exceptional energy efficiency
- Leadership TCO for an outstanding return on investment

EXASCALE LEADERSHIP DRIVES CONTINUED INNOVATION

While not every application requires supercomputer-class performance, public investments have been crucial in shaping the design of EPYC. AMD processors power 156 supercomputers in the latest Top500 list, a 29% increase year over year, showing increased adoption of AMD processors for the world's most demanding applications.[9] Noteworthy examples include Europe's LUMI, a powerful force in climate and cancer research, Frontier at Oak Ridge National Laboratory, the first computer to cross the exascale barrier with 1.6 exaflops of peak processing power, and El Capitan, a new exascale system being deployed at Lawrence Livermore National Laboratory.

More importantly, according to the latest Green500 list, **AMD powers 60% of the world's top fifty most energy-efficient supercomputers**, a testament to AMD leadership in designing energy-efficient processors and GPUs.[10]

UNRIVALED PERFORMANCE FOR THE FULL RANGE OF PUBLIC SECTOR WORKLOADS

AMD EPYC processors and AMD Instinct™ accelerators excel across a wide range of workloads. AMD and its partners have published **over 300 world records** using industry-standard consortium benchmarks and ISV workloads.[11] In addition, AMD has run several internal benchmarks comparing AMD processors with competitive processors.

Selected benchmarks are highlighted below, illustrating the advantage of AMD processors over those offered by competitors for a range of general-purpose and enterprise HPC workloads.

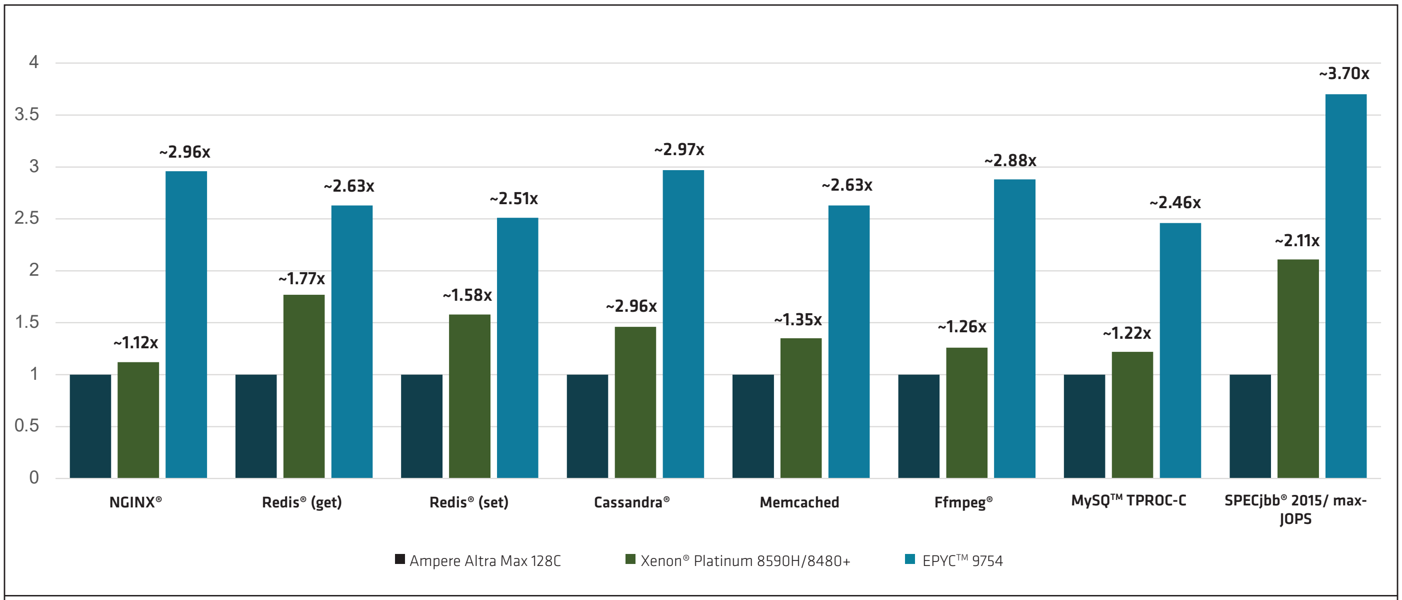
DATA CENTER	CLOUD	AI MODEL TRAINING & INFERENCE
60% faster general-purpose top-of-stack integer performance. 4th Gen AMD EPYC™ 9654 vs. 5th Gen Intel® Xeon® Platinum 8592+.[12]	**~70% average performance uplift and ~28% lower cloud OPEX.** Based on AMD internal testing – AWS M7a.4xlarge vs. AWS M7i.4xlarge instances.[15]	**~65% more AI test cases/minute on TPCx-AI SF30 based on AMD internal testing.** 4th Gen AMD EPYC™ 9654 vs. 5th Gen Intel® Xeon® Platinum 8592+ 2P server comparison.[17]
1.6x VMmark® 3 general-purpose top-of-stack virtualization performance. 4th Gen AMD EPYC™ 9654 vs. 5th Gen Intel® Xeon® Platinum 8592+.[13]	**37% average performance uplift and 31% lower Cloud OpEx on Google Cloud.** Based on AMD Internal testing – Google Cloud C3D-standard 16 vCPU vs N2-standard 16 vCPU instances.[16]	**~1.6x the performance on Bloom 176.** Based on AMD internal testing – 8x AMD Instinct™ MI300X vs. 8x NVIDIA® H100.[18]
2.25x better top-of-stack power efficiency with SPECpower_ssj® 2008. 4th Gen AMD EPYC™ 9754 vs. 5th Gen Intel® Xeon® Platinum 8592+.[14]		

DEFENSE, CLIMATE RESEARCH	AEROSPACE, MANUFACTURING	LIFE SCIENCES
Up to 47% better on SPEChpc® 2021 Tiny OMP Comparison. 2P AMD EPYC™ 9654 vs 2P Intel® Xeon® Platinum 8490H.[19]	**~50% better on ANSYS® LS-DYNA®.** Based on AMD internal testing – 2P AMD EPYC 9374F vs. 2P Intel Xeon 8562Y+.[22]	**~63% better general-purpose top-of-stack performance on GROMACS.** Based on AMD internal testing – 2P AMD EPYC 9654 vs. Intel Xeon Platinum 8592+.[25]
~50% faster on Weather Research and Forecasting (WRF® 2.1). Based on AMD internal testing – 2P AMD EPYC™ 9654 vs 2P Intel® Xeon® Platinum 8592+.[20]	**~48% better on ANSYS® CFX®.** Based on AMD internal testing – AMD EPYC 9374F vs. Intel Xeon 8562Y+.[23]	**~62% better general-purpose top-of-stack performance on Quantum Chemistry with CP2K.** Based on AMD internal testing – 2P AMD EPYC 9654 vs. Intel Xeon Platinum 8592+.[26]
19% better general-purpose top-of-stack performance on SPECrate® 2017 fp_base. 4th Gen AMD EPYC™ 9654 vs. 5th Gen Intel® Xeon® Platinum 8592+.[21]	**~25% better on ANSYS® FLUENT®.** Based on AMD internal testing – 2P AMD EPYC 9374F vs. 2P 32-core Intel Xeon Scalable 8562Y+.[24]	

Governments worldwide are increasingly embracing modern software development techniques to build scalable,containerized applications. These applications rely on a variety of cloud-native SQL and NoSQL data management solutions, and performance is critical for response times and quality of service.

In AMD testing on a series of benchmarks shown in Figure 3, 2P systems powered by 128-core 4th Gen AMD EPYC 9754 CPUs showed strong performance uplifts on key cloud-native workloads compared to comparable Intel® Xeon® Platinum 8490H and Ampere® Altra® Max CPUs.[27]

On average, a 2P 128-core 4th Gen AMD EPYC 9754 system demonstrated a **~2.84x** performance uplift on key cloud-native workloads vs. a top-of-stack 2P Ampere® Altra® Max system. It also significantly outperformed a prior top-of-stack Intel® Xeon® 8490H system by **~1.92x**.



2P servers: 128C Ampere Altra® Max M128-30 vs. 128C AMD EPYC™ 9754 vs. 60C Intel® Xeon® Platinum 8490H

Figure 3 - AMD EPYC delivers superior throughput for cloud-native application workloads

AMD EPYC processors also deliver leadership performance on industry-standard bare-metal benchmarks. Figure 4 shows a comparison of the latest top-of-stack 5th Gen Intel® Xeon® Platinum processors in a 2P configuration versus comparable AMD EPYC processors across a series of benchmarks.[28] Results are shown for SPECpower_ssj®, SPECrate® 2017_int_base, SPECrate® 2017_fp_base, SPECjbb® 2015, TPC Benchmark™ H, SAP SD, and VMmark® 3.

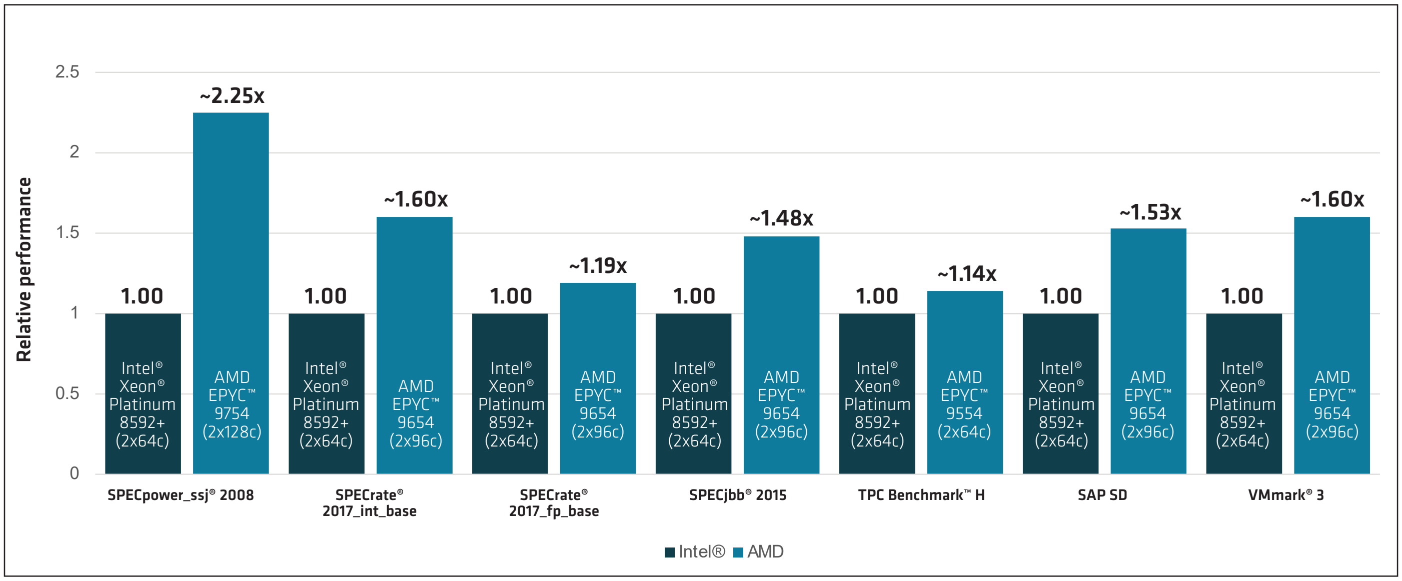


Figure 4: AMD EPYC™ delivers superior performance on a range of benchmarks compared to top-of-stack competitors

ACCELERATING AI MODEL TRAINING AND INFERENCE

AMD EPYC processors can accelerate your entire AI journey, providing a host processor for GPU-accelerated machine learning and an efficient processor for AI inferencing. Servers powered by 4th Gen AMD EPYC processors and the latest AMD Instinct MI300 accelerators offer the bandwidth needed to train predictive models and arrive at the necessary parameter weights to make models function with the speed and accuracy you need. AMD Instinct accelerators provide exceptional AI performance compared to the competition. For instance, based on AMD internal testing and looking at commonly used AI precisions, **the AMD Instinct MI300X can provide up to 1.3X the performance of NVIDIA's H100 accelerators.**[29]

Once models are trained, inferences can often be made on the CPU without GPU acceleration. With a 2-socket server powered by 128-core EPYC 9754 processors running select OpenVINO™ workloads from the Phoronix Test Suite, public sector customers can obtain up to 72% higher inference FPS per watt on average over 2P servers powered by 64-core 5th Gen Xeon 8592+ CPUs.[30]

A Zen Deep Neural Network (ZenDNN) plugin optimizes execution at the primitive level, accelerating models across diverse application types.[31] Whether you are using a computer vision model to recognize product defects, a natural language application to respond to customer prompts, or an engine to direct proactive maintenance, ZenDNN can help accelerate your inference workloads.

TAKING SUSTAINABILITY TO THE NEXT LEVEL

While performance is important, energy efficiency and data center sustainability are also critical. **AMD EPYC processors power the industry's most energy-efficient x86 servers**, delivering exceptional performance and helping lower energy consumption.[32] In the SPECPower® benchmark widely used to measure energy efficiency, a dual-socket system powered by **128-core AMD EPYC 9754 processors delivered ~2.25x the energy efficiency of a dual-socket system powered by Intel Xeon Platinum 8592+ processors.**[33]

Many federal and state governments worldwide have set greenhouse gas (GHG) reduction goals and require suppliers to disclose GHG emissions and climate risks as part of their procurement strategies.[34] Using energy-efficient AMD EPYC processors can not only help achieve ongoing cost reductions from operations, but it can also help achieve scope 2 GHG emission reduction targets related to the consumption of purchased non-renewable electricity for data center power and cooling. Similarly, for organizations operating in the cloud, using instances powered by AMD EPYC can help reduce direct cloud expenses and simultaneously reduce scope 3 emissions resulting from the purchase of cloud services.

A STEADFAST COMMITMENT TO ENVIRONMENTAL SUSTAINABILITY

AMD embraces the role of protecting the planet, helping to save energy, and reducing GHG emissions. Based on science-based goals aligned to a 1.5C° warning scenario, **AMD has pledged to reduce emissions from its internal operations (scopes 1 and 2) by 50% by 2030.**[35]

AMD is tackling environmental sustainability by focusing on three pillars:

- **Addressing Environmental Impacts at AMD and in Our Supply Chain:** Reducing AMD operational GHG emissions in line with science-based targets, while working closely with Manufacturing Suppliers to advance environmental performance.
- **Advancing Environmental Performance for IT Users:** Optimizing system-level energy effciency to help our customers and end users save energy and advance their sustainability goals.
- **Innovating on Collaborative Solutions to Address Environmental Challenges:** Enabling innovative solutions that optimize renewable energy generation, enhance smart solutions, and power cutting-edge climate and scientific research.

AMD is advancing its climate and sustainability programs through multiple initiatives, including water recycling and capture, investments in renewable energy and responsible sourcing. AMD has also set targets for its supply chain, including that 100% of direct manufacturing suppliers have a public GHG reduction goal and 80% source renewable energy by 2025. AMD shares its progress every year in a published corporate responsibility report.[36]

Collaboration is central to the company's approach on environmental sustainability. For example, AMD serves on the Responsible Business Alliance's Senior Environmental Advisory Taskforce, in addition to our continued role serving on the Alliance's Board of Directors, to shape the sustainability strategy for the world's largest industry coalition dedicated to corporate responsibility in global supply chains. Additionally, AMD is a founding and active member of the Semiconductor Climate Consortium and its Energy Collaborative, which focus on industry approaches to addressing Scope 3 emission "hot spots" across the electronics value chain, such as accelerating clean energy access in the Asia-Pacific region.

AMD SUSTAINABILITY GOALS

In 2021, AMD announced an aggressive goal to deliver a **30x increase in energy efficiency** for AMD processors and accelerators powering servers for AI training and HPC from 2020 to 2025. As illustrated in Figure 5, as of 2023, AMD achieved this goal **13.5x** improvement in energy efficiency from its 2020 baseline reporting.[30]

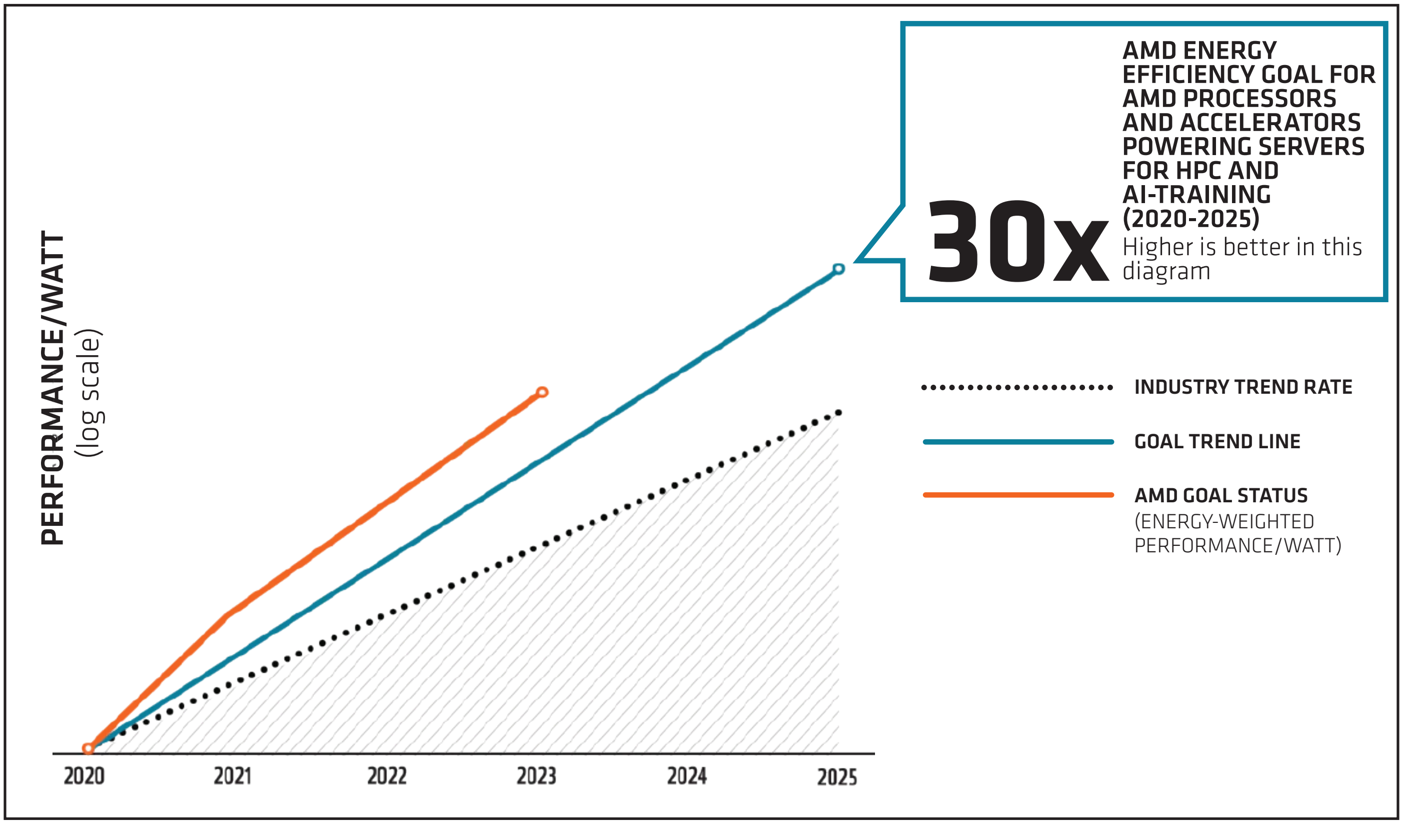


Figure 5 - AMD is on track to achieve an aggressive 30x energy efficiency improvement by 2025.

At the Imec Technology Forum (ITF) World 2024 conference in Belgium, AMD Chair and CEO, Dr. Lisa Su was honored with the prestigious imec Innovation Award for her work driving innovation in high-performance and adaptive computing. During her keynote, she announced that AMD is not only on track to meet its 30x energy efficiency goal but also sees a path to more than a **100x** improvement through 2027.[37]

PERFORMANCE AND ENERGY EFFICIENCY ENABLE LOWER TCO

Performance, energy efficiency, and total cost of ownership (TCO) are tightly coupled. AMD is focused on continually improving application performance, energy efficiency, and TCO across the public sector.

To provide a concrete example, virtualization with VMware vSphere® is a staple application across many public sector institutions.

Delivering 2,000 virtual machines takes an estimated 17 2P Intel® Xeon® Platinum 8490H-based servers. The same performance can be achieved by deploying an estimated 11 2P servers powered by AMD EPYC™ 9654 processors, representing a **~35% reduction in servers and a ~36% reduction in power consumption while reducing hardware CAPEX by up to 47%.[38]** These savings are illustrated in Figure 6.

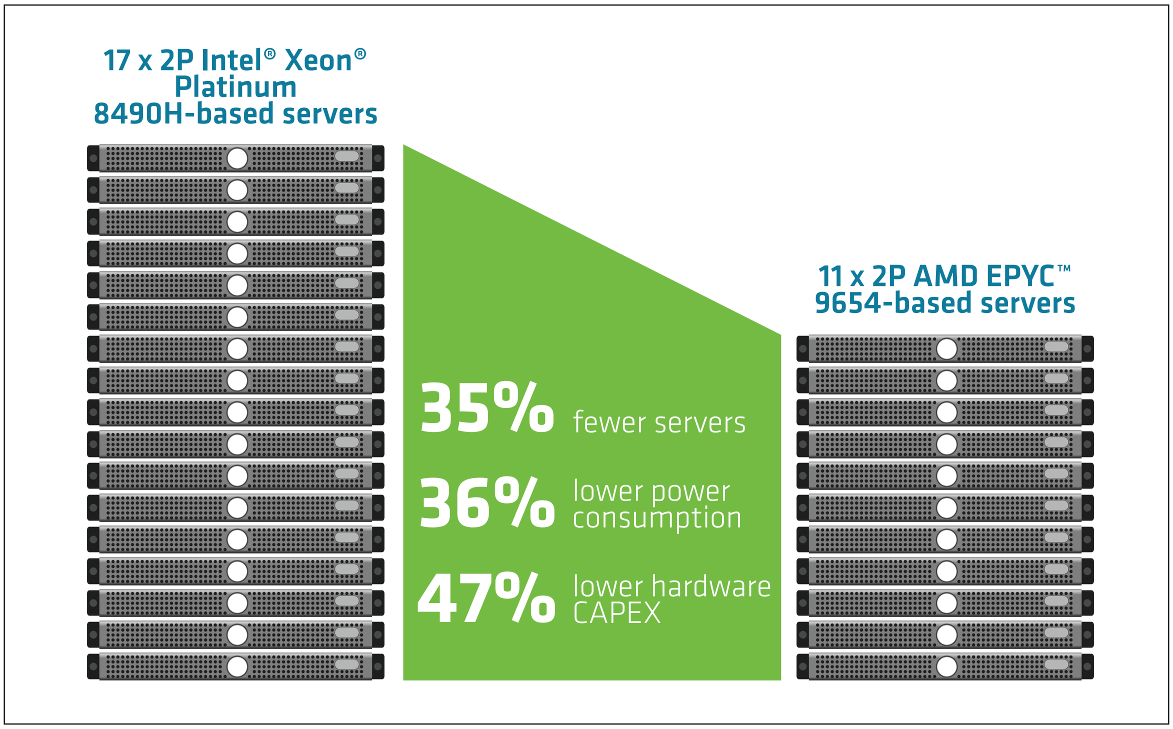


Figure 6 - Reducing cost, power consumption, and data center space with energy-efficient EP YC CPUs

For public sector organizations, the combination of high-performance processors with lower energy requirements is a game changer with cascading benefits. For organizations that operate on-premises infrastructure, lower server counts can translate into less packaging and waste, lower manufacturing and shipping costs, savings in data center space, rack and network costs, reduced power and cooling costs, and reduced management and administration overhead.

SUSTAINABLE BY DESIGN

As a semiconductor manufacturer, wafers are the most significant input to the company's supply chain. A powerful example of design impacting sustainability is the use of the chiplets described earlier, component IP building blocks connected via the AMD scalable Infinity Fabric™. AMD recently measured the sustainability benefits of chiplets in the wafer manufacturing process for one product line, and the results are compelling.

Producing 4th Gen AMD EPYC™ CPUs with up to 12 separate compute chiplets instead of one monolithic die saved ~132,000 metric tCO2e in 2023 through avoidance of wafers manufactured, 2.8 times more than the annual operational CO2e footprint of AMD in 2023.[39]

ADDRESSING 21ST-CENTURY SECURITY CHALLENGES

With the vast array of threats facing most organizations, cybersecurity is a top priority in the public sector. AMD works to continuously strengthen security protections in its products and continuously align its security controls with leading industry frameworks, such as CIS Critical Security Controls, the NIST Cybersecurity Framework, and COBIT. AMD also works to continuously align with European security frameworks, including the EU's ENISA Information Assurance Framework and Germany's BSI Act (Federal Information Technology Security Act).[40]

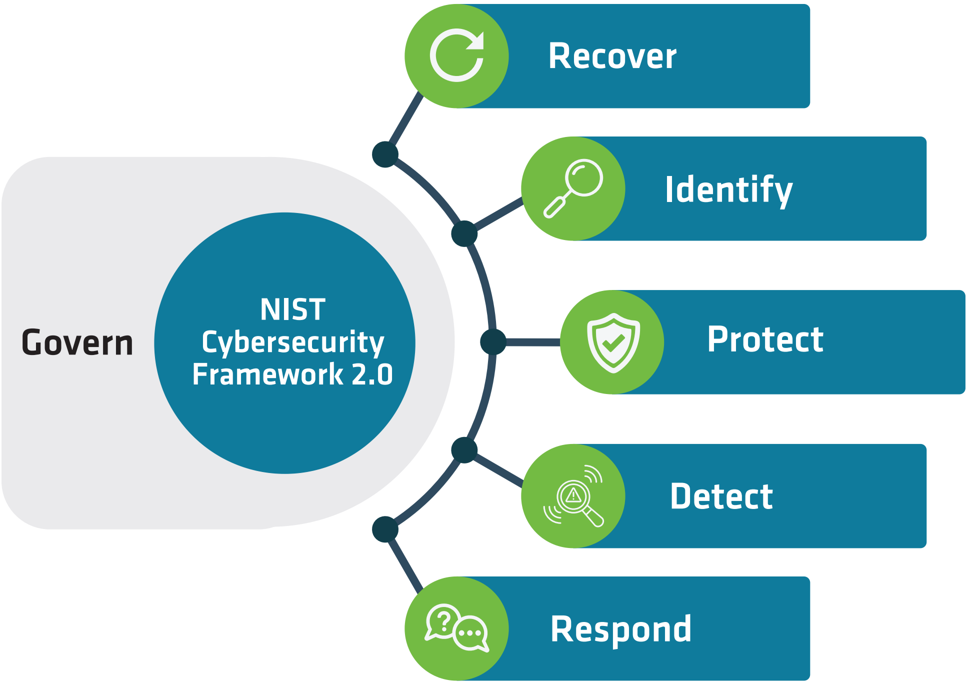


Figure 7 - NIST Cybersecurity Framework 2.0

The NIST Cybersecurity Framework (CSF 2.0), shown in Figure 7, is a set of guidelines for mitigating cybersecurity risks published by the US National Institute of Standards and Technology.[41] It is used by both public and private organizations. The framework provides a taxonomy of core cybersecurity functions, categories, and subcategories that organizations can use to build a tailored cybersecurity plan.

AMD data center products can help organizations align with core requirements of the NIST CSF 2.0 and other frameworks through security features, including secure boot and firmware, memory encryption, and hardware-accelerated cryptography. These features are important for aligning with NIST standards for data protection.

SECURITY STARTS WITH HARDWARE

As a semiconductor manufacturer, wafers are the most significant input to the company's supply chain. A powerful example of design impacting sustainability is the use of the chiplets described earlier, component IP building blocks connected via the AMD scalable Infinity Fabric™. AMD recently measured the sustainability benefits of chiplets in the wafer manufacturing process for one product line, and the results are compelling.

AMD INFINITY GUARD™

With hardware-based security features such as AMD Infinity Guard™, encryption keys are generated by the secure processor at runtime. This can be advantageous since a hostile actor has no central key management center to exploit. In addition, cryptographic algorithms are implemented in hardware, making them exceptionally fast and helping maintain a secure environment, ideally with minimal impact on the performance of production applications. Additionally, there is not a need to change x86 application software to take advantage of these robust security features. AMD hardware can support encryption with both 128- and 256-bit keys and complies with the US Federal Information Processing Standard (FIPS) 140-3 security requirements for cryptographic modules.[42]

In addition to helping address cybersecurity concerns, this hardware-based approach to security can help customers, hardware manufacturers, and cloud providers comply with various privacy regulations by preventing unauthorized access to sensitive information. These include the EU's General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), Canada's Personal Information Protection and Electronic Documents Act (PIPEDA), and industry-specific privacy requirements such as the Health Insurance Portability and Accountability Act (HIPAA).

AMD Infinity Guard is a set of advanced capabilities implemented by the AMD Secure Processor designed to help defend against internal and external threats to keep data safe. **AMD Infinity Guard serves as the technical foundation of confidential computing**.[43] **Key features include:**

- **Secure Encrypted Virtualization (SEV)** can help protect data while it is being processed. Keys are issued dynamically by the AMD secure processor for each VM, so that not even cloud providers have visibility to data—only the person or service that "owns" the data can encrypt it.
- **SEV Encrypted State (SEV-ES)** can extend SEV's functionality, helping protect register contents even when virtual machines are offline .
- **SEV Secure Nest Paging (SEV-SNP)** can provide additional levels of security against malicious hypervisors and optional defenses against interrupt injection, specific speculative side-channel attacks, and TCB rollback attacks.
- **Transparent Secure Memory Encryption (TSME)** can shield main memory from snooping in enterprise bare-metal environments and in the cloud by using hardware-based encryption between processors and memory, even helping to guard against certain cold-boot attacks.

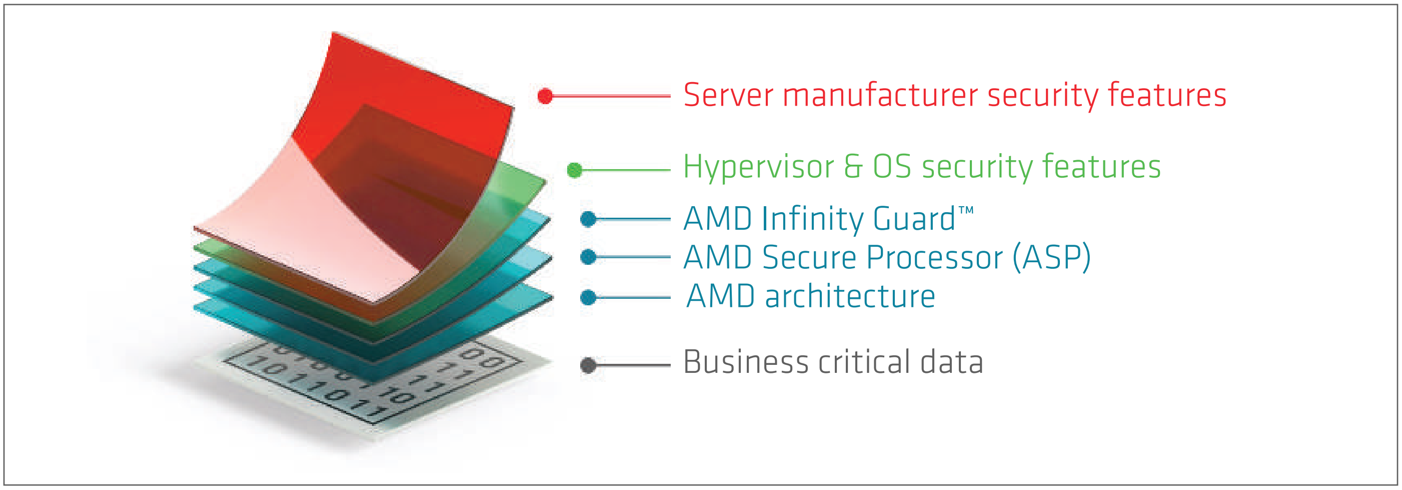


Figure 8 - A layered approach to security

An advantage of AMD Infinity Guard is its performance — because encryption is performed at the hardware level, and no changes are required to applications, performance impact may be lessened.[44]

TRANSPARENCY PROMOTES ENTERPRISE SECURITY

For government agencies, transparency is important. Unlike proprietary "black box" designs, AMD offers an approach to security based on transparency and open standards. For example, AMD makes its SEV implementation freely available on GitHub for various Linux® operating environments.[45] This transparency helps enable peer review by stakeholders such as public sector clients, OS & hypervisor providers, system manufacturers, and cloud providers. Additionally, transparency can also help stakeholders more quickly identify and patch potential vulnerabilities and exploits.

AMD works closely with partners, academics, and researchers to quickly respond to security incidents. As a CNA (CVE Numbering Authority) member, AMD follows coordinated vulnerability disclosure practices and seeks to respond quickly and publish security bulletins and security briefs to inform of potential vulnerabilities and communicate mitigation strategies.[46] AMD is also a member of FIRST, the Forum of Incident Response and Security Team, and responds systematically to potential issues.

SECURITY IN THE SUPPLY CHAIN

To help deliver tamper-free products designed to be secure, AMD focuses on its supply chain, aiming for visibility, transparency, and traceability. A focus on product integrity helps avoid tampering and counterfeiting, and AMD also takes strong cybersecurity measures to protect its data and intellectual property across its ecosystem to help maintain product security.[47]

"If integrity in supply chains is not maintained, you can risk loss of data and customer privacy. It can halt or jeopardize enterprise or government operations."

-Akash Malhotra, Director of Security Product Management, AMD

CONFIDENTIAL COMPUTING AND SOVEREIGN CLOUD

With the widespread use of cloud computing, containers, and on-premises virtualization solutions, helping ensure confidentiality in shared computing environments has become a critical concern. As enterprises have invested in firewalls, network-level & data encryption, and various other security measures to protect against malware and other threats, sophisticated actors have increasingly focused their attacks on data-in-use while being processed by the CPU, a historically underappreciated vulnerability.

Privacy regulations such as the GDPR require safeguarding personal information with stringent penalties for non-compliance. Decisions by the Court of Justice of the EU (CJEU) have invalidated both the Safe Harbor and Privacy Shield provisions previously relied on by European entities to store data in the US data centers, and these measures have been replaced by a new Data Privacy Framework (DPF) agreed to by the EU and the US effective July 10, 2023. As a result of these changes, sovereign cloud is high on the agenda for both public and private sector organizations.[48]

Confidential computing, enabled by AMD Secure Encrypted Virtualization (SEV) technology, follows a "zero trust" model. In this model, not even the hypervisor is trusted, so customers can be assured that their information will remain confidential even when operating in public, private, or hybrid clouds. Cloud providers favor this operating model because it helps overcome obstacles related to data privacy and sovereign cloud.

By deploying confidential computing cloud instances powered by AMD EPYC processors, organizations can comply with stringent privacy requirements by simply "ticking a box" when provisioning VMs to protect data in use. Today, most major cloud providers, including AWS, Microsoft Azure, and Google Cloud Platform, offer confidential computing instances powered by AMD EPYC processors.[49]

In on-premises, hybrid, and private cloud environments, public institutions running AMD EPYC processors can realize similar privacy protections to those offered in the cloud by taking advantage of VMware Sphere® support for AMD Infinity Guard™ protection features.[50]

HIGH-PERFORMANCE SYSTEMS ENHANCE SECURITY

In addition to the silicon-based security features described above, most organizations deploy software for identity management, network monitoring, and detecting malware. Examples include Okta, Checkpoint, CrowdStrike, Palo Alto Networks, and various security and monitoring services offered by cloud providers.

Increasingly, these security solutions leverage AI techniques to train predictive models to recognize patterns in data and detect and neutralize threats. With high-performance AMD EPYC processors and AMD Instinct accelerators, users can train ML models to recognize and respond to potential security incidents quickly, helping minimize risk and the amount of time organizations are exposed to threats.

A COMMITMENT TO OPEN SOFTWARE AND STANDARDS

Above, we explained how open-source software and transparency help ensure security, but open software and adherence to standards are also beneficial for other reasons. They help foster innovation, ensure portability and interoperability, and simplify procurement by encouraging competition around open standards.

AMD ROCm™ software is an open software stack that includes programming tools, compilers, libraries, and runtimes for development on select AMD GPUs.[51] Because ROCm is freely available and open source, developers have quickly embraced the platform, leading to an extensive ecosystem of AI models and HPC applications optimized for AMD hardware.[52] ROCm is powered by the Heterogeneous-computing Interface for

Portability (HIP)—an API and runtime environment that enables developers to create applications that are portable across GPUs from multiple manufacturers.[53]

Supported by leading ML frameworks such as PyTorch and TensorFlow, ROCm is being incorporated into models from Hugging Face, Lamini, and other open AI providers.[54] These innovations make it easier for public sector organizations to incorporate off-the-shelf, standard models into their own own applications. Also, customers can leverage dozens of curated, off-the-shelf containers for HPC and AI applications and frameworks used in government from AMD Infinity Hub—a collection of advanced software containers.[55]

AMD is also working to establish open standards for the interconnect technologies critical to modern AI and HPC infrastructure. AMD is a founding member of the **Ultra Ethernet Consortium (UEC)**, which establishes open standards for high-bandwidth, low-latency networking.[56] Along with Google, Microsoft, Meta, and other industry leaders in AI, AMD has joined the **Ultra Accelerator Link (UA Link)** promoter group to develop an open industry standard for connecting the GPUs widely used in AI model training.

CUSTOMER DEPLOYMENTS

AMD EPYC processor-based servers are widely deployed for a wide range of public and private sector applications. They deliver a unique combination of leadership performance, energy effciency, and security features critical to meeting regulatory and privacy requirements.

A good example is Zonar Systems, a leading fleet management and mobility solutions company headquartered in Seattle, Washington. **By leveraging the confidential computing capabilities of AMD EPYC CPU-based N2D instances in the Google Cloud, Zonar was able to provide a higher level of protection against various types of cyber threats while providing GDPR and Schrems II compliance for their European clients.**[57]

Europe's largest semiconductor company, ST Microelectronics, deployed AMD EPYC CPU-based systems and Microsoft Azure cloud instances to provide additional capacity and performance for its demanding semiconductor design environment.

To achieve its objective of carbon neutrality by 2027, ST Micro needed to decrease its total energy consumption by 150 GW hours per year—a tall order for a business experiencing double-digit growth. By deploying high-performance AMD EPYC CPU-based systems for demanding simulation and regression tests, **ST Micro was able to realize 12% better performance, achieve up to a 30% lower cost per core, and reduce its total energy consumption by 33%**.[58]

Other large deployments of AMD EPYC CPU-based systems in government and education include the National Oceanic and Atmospheric Administration (NOAA)[59] , CERN's Large Hadron Collider[60] , and Cornell[61] and Purdue[62] Universities.

SUMMARY

Public sector organizations need high-performance infrastructure that is cost-effective, energy-efficient, and that helps secure their environment against a wide range of cyber threats. AMD EPYC processors address all the critical infrastructure requirements for public sector applications.

- AMD offers **leadership performance**, having achieved over **300 world records** along with partners running industry standard consortium benchmarks and ISV workloads.[11]
- AMD EPYC processors power the **industry's most energy-efficient x86 servers**32, delivering up to **2.25x** the energy efficiency compared to comparable servers.[14]
- AMD Infinity Guard™ serves as a technical foundation of **confidential computing**, helping customers employ a zero-trust model and align to privacy and security mandates while supporting best practices such as the NIST cybersecurity framework.
- High-performance, energy-efficient processors can **significantly reduce TCO**. For the virtualized environments increasingly common in public sector applications, in a 2000 VM environment, customers can achieve an estimated **35% reduction in server footprint, a 36% reduction in power consumption, and hardware CAPEX reductions of up to 47% compared to competing processors**.[38]
- Finally, AMD is committed to transparency and open standards, helping ensure interoperability and investment protection, fostering innovation, improving security features, and simplifying public sector procurement.

GETTING STARTED WITH AMD SOLUTIONS

AMD EPYC CPU-based servers are available from most major computer manufacturers, including CISCO, Dell, HPE, Lenovo, and Supermicro. AMD EPYC processor-powered machine instances are also available in most clouds, including Amazon Web Services (AWS), Google Cloud Platform, Microsoft Azure, and Oracle Cloud Infrastructure (OCI).

To learn more about AMD solutions for the public sector, visit https://www.amd.com/en/solutions/public-sector.html.

To estimate how a move to AMD EPYC may help reduce your TCO and total greenhouse gas (GHG) emissions, visit https://www.amd.com/en/processors/epyc-VirtTCOtool.

FOOTNOTES

1. International Energy Agency, Electricity 2024 report, January 2024. See page 31: "Global electricity demand from data centres could double towards 2026".

2. Gartner Press Release, Gartner Announces the Top Government Technology trends for 2024, April 16, 2024
GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

3. Center for Strategic & International Studies Significant Cyber Incidents – Significant cybersecurity incidents against government agencies, defense and high-tech companies or economic crimes with losses of more than USD 1M.

4. See the 4th Gen AMD EPYC™ Processor Architecture Whitepaper

5. Refers to the AMD EPYC 9174F part. Max boost speeds vary by processor. EPYC-018 – Max. boost for AMD EPYC processors is the maximum frequency achievable by any single core on the processor under normal operating conditions for server systems.

6. Applies to the AMD EPYC™ 9684X processor with 3D V-Cache ™ technology. L3 cache varies by processor.

7. Applies only to AMD EPYC™ 9004 and 8004 series processors with four or fewer CPU dies.

8. GD-183A: 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.

9. Press release: AMD Remains the Partner of Choice for World's Fastest and Most Efficient High Performance Computing Deployments – 05/13/2024.

10. See June 2024 Green500 list at https://top500.org/lists/green500/2024/06/.

11. See AMD EPYC™ Processor World Records.

12. SPECrate® 2017_int_base results @ top-of-stack comparisons as of 08/09/2024:

- 2P EPYC 9654, Score 1810, https://www.spec.org/cpu2017/results/res2024q1/cpu2017-20240129-40896.html
- 2P Xeon 8592+, Score 1130, https://www.spec.org/cpu2017/results/res2023q4/cpu2017-20231127-40064.html

SPEC® and SPECrate® are registered trademarks of Standard Performance Evaluation Corporation. See www.spec.org for more information.

13. VMmark 3 performance @ general purpose top-of-stack:
- 2P EPYC 9654, 40.66 @ 42 tiles, https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmmark/2023-06-13-Lenovo-ThinkSystem-SR665V3.pdf
- 2P Xeon 8592+, https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/vmmark/2024-02-20-Dell-PowerEdge-R760.pdf.

14. SP5-011F: SPECpower_ssj® 2008 comparison based on published 2P server results as of 1/12/2024. Configurations: 2P 128-core AMD EPYC 9754 (36,210 overall ssj_ops/W, 2U, https://spec.org/power_ssj2008/results/res2024q1/power_ssj2008-20231205-01347.html) is 2.25x the performance of best published 2P 64-core Intel Xeon® Platinum 8592+ (16,106 overall ssj_ops/W, 2U, https://spec.org/power_ssj2008/results/res2024q1/power_ssj2008-20231205-01349.html). SPEC®, SPECpower®, and SPECpower_ssj® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information.

15. SP5C-004 - AWS M7a.4xlarge max score and Cloud OpEx savings comparison to M7i.4xlarge running six common application workloads using on-demand pricing US-East (Ohio) Linux as of 10/9/2023.
- FFmpeg: ~1.9x the raw to vp9 encoding performance (52.3% of M7i runtime) saving ~40% in Cloud OpEx
- NGINX™: ~1.6x the WRK performance (61.7% of M7i runtime) saving ~29% in Cloud OpEx
- Server-side Java® multi-instance max Java OPS: ~1.4x the ops/sec performance (71.4% of M7i runtime) saving ~18% in Cloud OpEx
- MySQL™: ~1.4x the TPROC-C performance (70.4% of M7i runtime) saving ~18% in Cloud OpEx
- SQL Server®: ~1.3x the TPROC-H performance (76.0% of M7i runtime) saving ~13% in Cloud OpEx
- Redis™: ~2.2x the rps performance (44.6% of M7i runtime) saving ~49% in Cloud OpEx

Cloud performance results presented are based on the test date in the configuration. Results may vary due to changes to the underlying configuration, and other conditions such as the placement of the VM and its resources, optimizations by the cloud service provider, accessed cloud regions, co-tenants, and the types of other workloads exercised at the same time on the system.

16. SP5C-006 - MySQL™, Redis®, NGINX®, server-side Java multi-instances, and FFmpeg™ comparison of Google Cloud C3D-standard 16 vCPU to N2-standard 16 vCPU based on AMD testing on 11/02/23. OpEx savings calculated based on on-demand pricing at https://cloud.google.com/compute/vm-instance-pricing for us-central1 (Iowa) as of 11/01/2023. Configurations both with 64GB running Ubuntu 22.04.3 LTS. Comparisons:

- MySQL 8.0.28 HammerDB 4.2 TPROC-C (~1.2x tpm, 22% Cloud OpEx savings),
- Redis 7.2 get/set: (~1.4x rps, 32% Cloud OpEx savings),
- NGINX 1.1.9-2 WRK 4.2: (~1.2x ops/sec, 23% Cloud OpEx savings),
- server-side Java® multi instances max-OPS (~1.7x OPS, 45% Cloud OpEx savings) and
- FFmpeg 4.4.2.0 Ubuntu 22.04.1 h264-vp9, raw_h264, raw_vp9, vp9_h264 at 1080p (~1.4x frames/hr, 32% Cloud OpEx savings).

Cloud performance results presented are based on the test date in the configuration. Results may vary due to changes to the underlying configuration, and other conditions such as the placement of the VM and its resources, optimizations by the cloud service provider, accessed cloud regions, co-tenants, and the types of other workloads exercised at the same time on the system.

17. SP5-051A: TPCx-AI SF30 derivative workload comparison based on AMD internal testing running multiple VM instances as of 4/13/2024. 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. AMD system configuration: Processors: 2 x AMD EPYC 9654; Frequencies: 2.4 GHz | 3.7 GHz; Cores: 96 cores per socket (1 NUMA domain per socket); L3 Cache: 384MB/socket (768MB total); Memory: 1.5TB (24) Dual-Rank DDR5-5600 64GB DIMMs, 1DPC (Platform supports up to 4800MHz); NIC: 2 x 100 GbE Mellanox CX-5 (MT28800); Storage: 3.2 TB Samsung MO003200KYDNC U.3 NVMe; BIOS: 1.56; BIOS Settings: SMT=ON, Determinism=Power, NPS=1, PPL=400W, Turbo Boost=Enabled; OS: Ubuntu® 22.04.3 LTS; Test config: 6 instances, 64 vCPUs/instance, 2663 aggregate AI use cases/min vs. Intel system configuration: Processors: 2 x Intel® Xeon® Platinum 8592+; Frequencies: 1.9 GHz | 3.9 GHz; Cores: 64 cores per socket (1 NUMA domain per socket); L3 Cache: 320MB/socket (640MB total); Memory: 1TB (16) Dual-Rank DDR5-5600 64GB DIMMs, 1DPC; NIC: 4 x 1GbE Broadcom NetXtreme BCM5719 Gigabit Ethernet PCIe; Storage: 3.84TB KIOXIA KCMYXRUG3T84 NVMe; BIOS: ESE124B-3.11; BIOS Settings: Hyperthreading=Enabled, Turbo boost=Enabled, SNC=Disabled; OS: Ubuntu® 22.04.3 LTS; Test config: 4 instances, 64 vCPUs/instance, 1607 aggregate AI use cases/min. 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.

18. MI300-34: Token generation throughput using DeepSpeed Inference with the Bloom-176b model with an input sequence length of 1948 tokens, and output sequence length of 100 tokens, and a batch size tuned to yield the highest throughput on each system comparison based on AMD internal testing using custom docker container for each system as of 11/17/2023.
Configurations:
2P Intel Xeon Platinum 8480C CPU powered server with 8x AMD Instinct™ MI300X 192GB 750W GPUs, pre-release build of ROCm™ 6.0, Ubuntu 22.04.2. vs. An Nvidia DGX H100 with 2x Intel Xeon Platinum 8480CL Processors, 8x Nvidia H100 80GB 700W GPUs, CUDA 12.0, Ubuntu 22.04.3.
8 GPUs on each system were used in this test.
Server manufacturers may vary configurations, yielding different results. Performance may vary based on use of latest drivers and optimizations.

19. SP5-060A: SPEChpc® 2021 Tiny OMP comparison based on published results as of 2/15/2023. Configurations: 2P AMD EPYC 9654 (192 cores, 24 base ranks, OMP parallel mode, https://spec.org/hpc2021/results/res2022q4/hpc2021-20221016-00135.html) scores 13.9 SPEChpc®2021_tny_base versus 2P Intel Xeon Platinum 8490H (120 cores, 30 base ranks, OMP parallel mode, https://spec.org/hpc2021/results/res2023q1/hpc2021-20230108-00156.html) scores 9.45 SPEChpc®2021_tny_base for 1.47x the performance. SPEChpc® is a registered trademark of Standard Performance Evaluation Corporation (SPEC). Learn more at www.spec.org.

20. SP5-241: AMD testing as of 04/23/2024. The detailed results show the average uplift of the performance metric (Mean Time/Step) of this benchmark for the AMD EPYC™ 9654 96-Core Processor, and the INTEL® XEON® PLATINUM 8592+ running the following test on Open-Source WRF® 4.2.1: * conus2.5km: ~1.50x.
AMD System Configuration: Server: AMD Titanite; Processors: 2 x 96-Core AMD EPYC™ 9654 ; Memory: 24x 64GB DDR5-4800; Storage: SAMSUNG MZQL21T9HCJR-00A07; BIOS: RTI1009C; BIOS Settings from Default; SMT=Off, NPS=4, Determinism=Power; OS: RHEL 9.3; Kernel: Linux 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: amd_iommu=on, iommu=pt, mitigations=off ; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance, Disable C2 States.
Intel System Configuration: Server: Lenovo Thinksystem SR650 V3; Processors: 2 x 64-Core INTEL® XEON® PLATINUM 8592+; Memory: 16x 64GB DDR5-5600; Storage: KIOXIA KCMYXRUG3T84; BIOS: ESE122V-3.10; BIOS Settings from Default: Hyperthreading=Off, Profile=Maximum Performance Profile; OS: RHEL 9.3; Kernel: 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: processor.max_cstate=1, Intel_idle.max_cstate=0, iommu=pt, mitigations=off ; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance. Results may vary based on system configurations, software versions, and BIOS settings.

21. SPECrate® 2017_fp_base results @ top-of-stack comparison as of 08/09/2024:
- 2P EPYC 9654, Score 1480, https://www.spec.org/cpu2017/results/res2024q1/cpu2017-20240111-40517.html

- 2P Xeon 8592+, Score 1240, https://www.spec.org/cpu2017/results/res2023q4/cpu2017-20231127-40063.html SPEC® and SPECrate® are registered trademarks of Standard Performance Evaluation Corporation. See www.spec.org for more information.

22. SP5-246: AMD testing as of 04/23/2024. The detailed results show the average uplift of the performance metric (Elapsed Time) of this benchmark for the AMD EPYC™ 9374F 32-Core Processor, and the INTEL® XEON® PLATINUM 8562Y+ running select tests on Ansys® LS-DYNA® R13_1_1: * ls-3cars: ~1.61x, * ls-car2car: ~1.48x, * ls-neon: ~1.67x, * ls-odb10m-short: ~1.25x. AMD System Configuration: Server: AMD Titanite; Processors: 2 x 32-Core AMD EPYC™ 9374F ; Memory: 24x 64GB DDR5-4800; Storage: SAMSUNG MZQL21T9HCJR-00A07; BIOS: RTI1009C; BIOS Settings from Default; SMT=O° , NPS=4, Determinism=Power; OS: RHEL 9.3; Kernel: Linux 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: amd_iommu=on, iommu=pt, mitigations=o° ; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance, Disable C2 States. Intel System Configuration: Server: Lenovo Thinksystem SR650 V3; Processors: 2 x 32-Core INTEL® XEON® PLATINUM 8562Y+; Memory: 16x 64GB DDR5-5600; Storage: KIOXIA KCMYXRUG3T84; BIOS: ESE122V-3.10; BIOS Settings from Default: Hyperthreading=O° , Profile=Maximum Performance Profile; OS: RHEL 9.3; Kernel: 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: processor.max_cstate=1, Intel_idle.max_cstate=0, iommu=pt, mitigations=o° ; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance. Results may vary based on system configurations, software versions, and BIOS settings.

23. SP5-245: AMD testing as of 04/23/2024. The detailed results show the average uplift of the performance metric (Elapsed Time) of this benchmark for the AMD EPYC™ 9374F 32-Core Processor, and the INTEL® XEON® PLATINUM 8562Y+ running select tests on Ansys® CFX® V231. Uplifts for the performance metric normalized to the INTEL® XEON® PLATINUM 8562Y+ follow for each benchmark: * Airfoil 100: ~1.54x, * Airfoil 10: ~1.53x, * Airfoil 50: ~1.54x, * LeMans Car: ~1.40x, * Automotive Pump: ~1.37x. AMD System Configuration: Server: AMD Titanite; Processors: 2 x 32-Core AMD EPYC™ 9374F ; Memory: 24x 64GB DDR5-4800; Storage: SAMSUNG MZQL21T9HCJR-00A07; BIOS: RTI1009C; BIOS Settings from Default; SMT=O° , NPS=4, Determinism=Power; OS: RHEL 9.3; Kernel: Linux 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: amd_iommu=on, iommu=pt, mitigations=o° ; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance, Disable C2 States. Intel System Configuration: Server: Lenovo Thinksystem SR650 V3; Processors: 2 x 32-Core INTEL® XEON® PLATINUM 8562Y+; Memory: 16x 64GB DDR5-5600; Storage: KIOXIA KCMYXRUG3T84; BIOS: ESE122V-3.10; BIOS Settings from Default: Hyperthreading=O° , Profile=Maximum Performance Profile; OS: RHEL 9.3; Kernel: 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: processor.max_cstate=1, Intel_idle.max_cstate=0, iommu=pt, mitigations=o° ; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance. Results may vary based on system configurations, software versions, and BIOS settings.

24. SP5-244: AMD testing as of 04/23/2024. The detailed results show the average uplift of the performance metric (Core Solver Rating) of this benchmark for the AMD EPYC™ 9374F 32-Core Processor, and the INTEL® XEON® PLATINUM 8562Y+ running select tests on Ansys® Fluent®: * aircraft_14m: ~1.33x, * aircraft_2m: ~1.26x, * combustor_12m: ~1.22x, * combustor_71m: ~1.27x, * exhaust_system_33m: ~1.22x, * f1_racecar-140m: ~1.29x, * fluidized_bed_2m: ~1.10x, * Fluent®-ice2: ~1.19x, * landing_gear_15m: ~1.23x, * LeMans_6000_16m: ~1.21x, * oil_rig_7m: ~1.03x, * f1_racecar_280m: ~1.24x, * pump_2m: ~1.43x, * rotor_3m: ~1.34x, * sedan_4m: ~1.39x. AMD System Configuration: Server: AMD Titanite; Processors: 2 x 32-Core AMD EPYC™ 9374F ; Memory: 24x 64GB DDR5-4800; Storage: SAMSUNG MZQL21T9HCJR-00A07; BIOS: RTI1009C; BIOS Settings from Default; SMT=O° , NPS=4, Determinism=Power; OS: RHEL 9.3; Kernel: Linux 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: amd_iommu=on, iommu=pt, mitigations=o° ; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance, Disable C2 States. Intel System Configuration: Server: Lenovo Thinksystem SR650 V3; Processors: 2 x 32-Core INTEL® XEON® PLATINUM 8562Y+; Memory: 16x 64GB DDR5-5600; Storage: KIOXIA KCMYXRUG3T84; BIOS: ESE122V-3.10; BIOS Settings from Default: Hyperthreading=O° , Profile=Maximum Performance Profile; OS: RHEL 9.3; Kernel: 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: processor.max_cstate=1, Intel_idle.max_cstate=0, iommu=pt, mitigations=o° ; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance. Results may vary based on system configurations, software versions, and BIOS settings.

25. SP5-243: AMD testing as of 04/23/2024. The detailed results show the average uplift of the performance metric (ns/day) of this benchmark for the AMD EPYC™ 9654 96-Core Processor, and the INTEL® XEON® PLATINUM 8592+ running select tests on Open-Source GROMACS. * benchPEP: ~1.70xm, * gmx_water1536K_PME: ~1.56x. AMD System Configuration: Server: AMD Titanite; Processors: 2 x 96-Core AMD EPYC™ 9654 ; Memory: 24x 64GB DDR5-4800; Storage: SAMSUNG MZQL21T9HCJR-00A07; BIOS: RTI1009C; BIOS Settings from Default; SMT=O° , NPS=4, Determinism=Power; OS: RHEL 9.3; Kernel: Linux 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: amd_iommu=on, iommu=pt, mitigations=o° ;

Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance, Disable C2 States. Intel System Configuration: Server: Lenovo Thinksystem SR650 V3; Processors: 2 x 64-Core INTEL® XEON® PLATINUM 8592+; Memory: 16x 64GB DDR5-5600; Storage: KIOXIA KCMYXRUG3T84; BIOS: ESE122V-3.10; BIOS Settings from Default: Hyperthreading=Off, Profile=Maximum Performance Profile; OS: RHEL 9.3; Kernel: 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: processor.max_cstate=1, Intel_idle.max_cstate=0, iommu=pt, mitigations=off; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance. Results may vary based on system configurations, software versions, and BIOS settings. AMD testing as of 04/23/2024.

26. SP5-239: AMD testing as of 04/23/2024. The detailed results show the average uplift of the performance metric (Elapsed Time) of this benchmark for the AMD EPYC™ 9654 96-Core Processor, and the INTEL® XEON® PLATINUM 8592+ running the following test on Open-Source cp2k. Uplifts for the performance metric normalized to the INTEL® XEON® PLATINUM 8592+ follow for each benchmark: * H2O-dft-ls: ~1.62x. AMD System Configuration: Server: AMD Titanite; Processors: 2 x 96-Core AMD EPYC™ 9654 ; Memory: 24x 64GB DDR5-4800; Storage: SAMSUNG MZQL21T9HCJR-00A07; BIOS: RTI1009C; BIOS Settings from Default; SMT=Off, NPS=4, Determinism=Power; OS: RHEL 9.3; Kernel: Linux 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: amd_iommu=on, iommu=pt, mitigations=off; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance, Disable C2 States. Intel System Configuration: Server: Lenovo Thinksystem SR650 V3; Processors: 2 x 64-Core INTEL® XEON® PLATINUM 8592+; Memory: 16x 64GB DDR5-5600; Storage: KIOXIA KCMYXRUG3T84; BIOS: ESE122V-3.10; BIOS Settings from Default: Hyperthreading=Off, Profile=Maximum Performance Profile; OS: RHEL 9.3; Kernel: 5.14.0-362.8.1.el9_3.x86_64; Kernel CMDLINE: processor.max_cstate=1, Intel_idle.max_cstate=0, iommu=pt, mitigations=off; Runtime Tunings: Clear caches, NUMA Balancing 0, randomize_va_space 0, THP ON, CPU Governor=Performance. Results may vary based on system configurations, software versions, and BIOS settings. AMD testing as of 04/23/2024.

27. Results may vary due to factors including system configurations, software versions, and BIOS settings. As of 6/13/2023. See Cloud Native Workloads on AMD EPYC™ 9754 Processors.

28. See 4th Gen AMD EPYC™ Processors Outshine the Latest 5th Gen Intel® Xeon® Processors.

29. MI300-17: Measurements conducted by AMD Performance Labs as of November 11th, 2023 on the AMD Instinct™ MI300X (750W) GPU designed with AMD CDNA™ 3 5nm | 6nm FinFET process technology at 2,100 MHz peak boost engine clock resulted in 653.7 TFLOPS peak theoretical TensorFloat-32 (TF32), 1307.4 TFLOPS peak theoretical half precision (FP16), 1307.4 TFLOPS peak theoretical Bfloat16 format precision (BF16), 2614.9 TFLOPS peak theoretical 8-bit precision (FP8), 2614.9 TOPs INT8 floating-point performance. The MI300X is expected to be able to take advantage of fine-grained structure sparsity providing an estimated 2x improvement in math efficiency resulting 1,307.4 TFLOPS peak theoretical TensorFloat-32 (TF32), 2,614.9 TFLOPS peak theoretical half precision (FP16), 2,614.9 TFLOPS peak theoretical Bfloat16 format precision (BF16), 5,229.8 TFLOPS peak theoretical 8-bit precision (FP8), 5,229.8 TOPs INT8 floating-point performance with sparsity. Published results on Nvidia H100 SXM (80GB) 700W GPU resulted in 989.4 TFLOPs peak TensorFloat-32 (TF32) with sparsity, 1,978.9 TFLOPS peak theoretical half precision (FP16) with sparsity, 1,978.9 TFLOPS peak theoretical Bfloat16 format precision (BF16) with sparsity, 3,957.8 TFLOPS peak theoretical 8-bit precision (FP8) with sparsity, 3,957.8 TOPs peak theoretical INT8 with sparsity floating-point performance. Nvidia H100 source: https://resources.nvidia.com/en-us-tensor-core.

30. SSP5-252: 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. 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.

31. See AMD Zen Deep Neural Network (ZenDNN).

32. 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.

33. SP5-011F: SPECpower_ssj®2008 comparison based on published 2P server results as of 1/12/2024. Configurations: 2P 128-core AMD EPYC 9754 (36,210 overall ssj_ops/W, 2U, https://spec.org/power_ssj2008/results/res2024q1/power_ssj2008-20231205-01347.html) is 2.25x the performance of best published 2P 64-core Intel Xeon® Platinum 8592+ (16,106 overall ssj_ops/W, 2U, https://spec.org/power_ssj2008/results/res2024q1/power_ssj2008-20231205-01349.html) SPEC® and SPECpower_ssj® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information

34. See Federal Supplier Climate Risks and Resilience Proposed Rule, 11/10/2022

35. In the AMD Corporate Responsibility Report and other AMD climate-related disclosures, AMD references this operational GHG reduction goal (market-based emissions) as a “science-based target” that is aligned with a 1.5˚C scenario. Per the California Climate and Carbon Disclosure Requirements (AB-1305), the basis for this statement is the AMD GHG goal is aligned with the Science-based Target initiative’s (SBTi’s) 1.5-degree minimum target ambition of 4.2% linear annual reduction. The SBTi criteria considers multiple climate scenario models from the IAMC and IEA. Interim progress by AMD toward the goal is disclosed in the AMD CR Report (page 35,37) as measured by metric tCO2e of Scope 1 and 2 GHG emissions (market-based). The AMD 2020 base year value is 61,754 metric tCO2e and the 2030 target year value is 30,877 metric tCO2e (a 50% reduction). AMD reports total Scope 1 and 2 GHG emissions for each interim year and undergoes third-party limited assurance of the Scope 1 and Scope 2 values, as well as the percentage completion toward the goal. For the most current AMD reported data, calendar year 2023, AMD reported 46,605 metric tCO2e Scope 1 and 2 emissions, approximately a 24.5% reduction from the 2020 baseline. The data received limited level assurance in accordance with the International Standard on Assurance Engagements (ISAE) 3000 Revised.

36. Download the 2023-24 Corporate Responsibility Report.

37. See AMD Lisa Su honored with the 2024 imec Innovation Award.

38. 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 approxima-tions, 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."

39. AMD estimation based on defect density (defects per unit area on the wafer), chip area and n-factor (manufacturing complexity factor) to estimate the number of wafers avoided in one year. Yield = (1 + A*D0)^(-n) where A is the chip area, D0 is the defect density and n is the complexity factor. The area is known from our design, D0 is known based our manufacturing yield data, and n is a number provided by a foundry partner for a given technology. The calculations are not meant to be precise, since chip design can have a large influence on yield, but it estimates the area impact on yield. The carbon emission estimates of 132,064 mtCO2e were calculated using the estimated number of 5 nm wafers saved in one year, based on the TechInsights’ Semiconductor Manufacturing Carbon Model. Comparison to AMD corporate footprint is based on AMD reported scope 1 and 2 market-based GHG emissions in 2023: 46,606 mtCO2e. Water savings estimates of 1,110 million liters were calculated using the estimated number of 5 nm wafers saved in one year times the amount of water use per 300mm wafer mask layer times the average number of mask layers. Comparison to AMD corporate water use is based on AMD 2023 reported value of 225 million liters.

40. AMD approach to Cybersecurity and adherence to standards is outlined in the CR Management and Governance section of its annual Corporate Responsibility Report.

41. See NIST Cybersecurity Framework 2.0

42. See 4th Gen AMD Infinity Guard Security Features on page 15 of the 4th Gen AMD EPYC Processor Architecture Whitepaper. The AMD ASP Cryptographic Coprocessor is FIPS 140-3 certified (certificate #4723) under the NIST Cryptographic Module Validation Program (CMVP).

43. 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 Provid-ers 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.

44. Note: security patches may impact performance.

45. See https://github.com/AMDESE/AMDSEV.

46. See AMD Security Bulletins and Briefs.

47. See An Effective Supply Chain Starts with Security, Wall Street Journal Business.

48. The EU-U.S. Data Privacy Framework (EU-U.S. DPF), the UK Extension to the EU-U.S. Data Privacy Framework (UK Extension to the EU-U.S. DPF), and the Swiss-U.S. Data Privacy Framework (Swiss-U.S. DPF) were developed to facilitate transatlantic commerce by providing U.S. organizations with reliable mechanisms for personal data transfers to the United States from the European Union / European Economic Area, the United Kingdom (and Gibraltar), and Switzerland that are consistent with EU, UK, and Swiss law. For details, see the Data Privacy Framework (DPF) Overview.

49. Specific hardware protections vary by cloud provider and the underlying AMD EPYC processor used.

50. vSphere and AMD Secure Encrypted Virtualization-Encrypted State is supported on AMD EPYC 7002 series processors (code named "Rome") and later CPUs. https://docs.vmware.com/en/VMware-vSphere/7.0/com.vmware.vsphere.vm_admin.doc/GUID-FB511CBA-4B89-469F-9799-D1347E1F2B0A.html

51. For a full list of Radeon parts supported by ROCm™ software as of 5/1/2024, go to https://rocm.docs.amd.com/en/latest/reference/gpu-arch-specs.html. GD-241.

52. See https://github.com/ROCm/ROCm.

53. For details on the Heterogeneous-computing Interface for Portability (HIP) see the AMD ROC™ Software documentation.

54. See Hugging Face and AMD partner on accelerating state-of-the-art models for CPU and GPU platforms. See also, Lamini & AMD: Paving the Road to GPU-Rich Enterprise LLMs.

55. See AMD Infinity Hub.

56. See AMD Founding Member of the Ultra Ethernet Consortium (UEC)

57. See Zonar helps ensure GDPR and Schrems II compliance by enhancing privacy and data protection.

58. See ST Microelectronics boosts chip design speed and enhances sustainability with AMD EPYC™ processors.

59. See NOAA completes upgrade to weather and climate supercomputer system.

60. See AMD EPYC™ CPUs Enable Rapid Quark Detection at CERN.

61. See Cornell Uses AMD Technology to Help Understand the Universe.

62. See Purdue University Breaks Research Computing Barriers.