THE STATE OF DATA CENTER MODERNIZATION IN AN AI-DRIVEN WORLD

End-to-End Solutions from Data Center to the Cloud

Artificial intelligence (AI) is transforming business, science, and technology. The latest developments offer a broad scope of opportunity to increase automation, enhance decision-making, and drive innovation—capabilities that are redefining industries and competitive landscapes. The success of AI and digital transformation initiatives is determined by the resilience of underlying IT infrastructure.

Modernizing IT and data center infrastructure has become a top priority as organizations scale their technology deployments, manage vast volumes of data, and meet increasing computational demands. AI-driven workloads are intensifying pressure on existing systems, necessitating solutions that are more scalable, high-performance, and efficient. Many organizations struggle to modernize effectively to meet these growing demands.

According to S&P Global Market Intelligence 451 Research, one-third of organizations have already transformed their data centers with technologies that are future-ready for the demands of AI. The remaining two-thirds are either actively modernizing or planning to do so. The data center modernization landscape can be segmented into three categories:[1]

- **Data center leaders:** 33% of organizations have fully modernized their data centers within the past two years.
- **Data center challengers:** 58% of organizations are currently undergoing data center modernization.
- **Data center observers:** 9% of organizations are planning to modernize their data centers within the next two years.

AI-readiness is what separates these groups. For example, data center leaders demonstrate a greater adoption of AI applications (9.2 applications in use out of 16 options) compared to their counterparts, and 51% of data center challengers are prepared to modernize compared to just 17% of data center observers.[1] The struggle stems from a growing perception gap between enterprise leaders, managers, and contributors about where they are in the modernization and AI journey.

Enterprise leadership typically has a broader view of all projects and tends to prioritize substantial investments in the physical IT footprint. In contrast, day-to-day managers and contributors of data center infrastructure tend to see bigger challenges in modernization and AI deployment with responsibilities and budgets centered around support, operations, and maintenance of IT systems. An organization that is misaligned with IT technology goals and lacks clarity regarding the state of its IT estate can encounter significant challenges. Common roadblocks include a lack of experience managing large and diverse workloads, fragmented adoption strategies, budgetary issues, and business culture resistance, among others.

To bridge the gap, organizations must create an environment that is AI-ready. The benefits of data center modernization maturity extend well beyond today's IT infrastructure, impacting the broader business for years to come. Those that prioritize data center modernization have the skills and resources to accelerate AI deployments and help ensure their longevity for whatever the future brings.

CRITICAL CHALLENGES OF MODERNIZATION

Leaders at the forefront of AI are leveraging their first-mover advantage to enable strategic growth and innovation. These forward-thinkers understand that AI is a business imperative to unlock deeper insights from their available data for increasing efficiency, profit margins, and competitiveness across all business functions. However, rapid change has exposed pitfalls for many organizations (such as outsized challenges in cloud adoption, disjointed IT and environmental systems, siloed data sources, and organizational tension). These factors create an opportunity for slower-progressing organizations to learn from their missteps and close the competitive gap.

Data center modernization presents a host of complex challenges spanning operational, strategic, financial, vendor, and organizational domains (Figure 1). Of these, business risks associated with mission-critical applications are the primary obstacle for many organizations.[1] Potential business risks may include data loss, security vulnerabilities, service disruptions, operational downtime, and increased implementation costs. Each of these obstacles must be addressed to capitalize on the full value of AI.
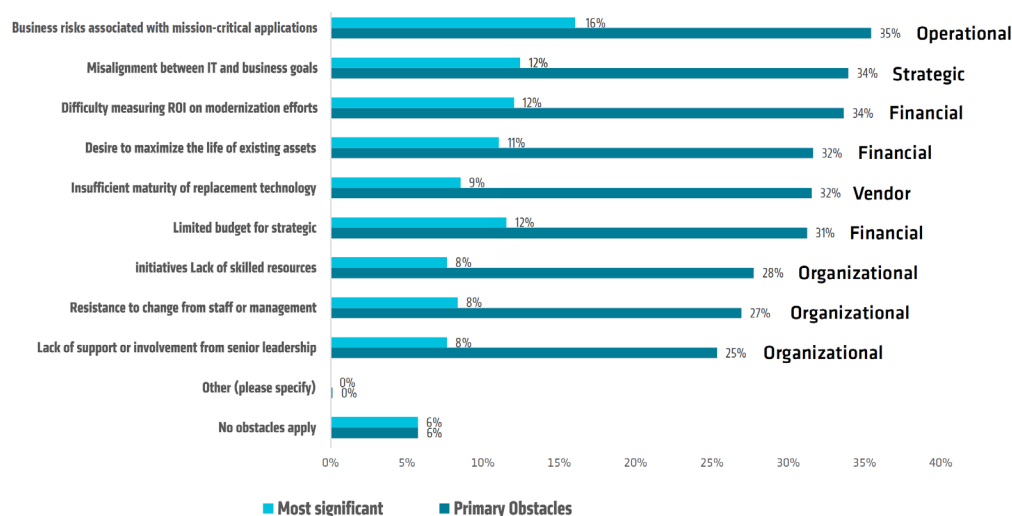
Figure 1: Obstacles to data center modernization and broader goals

HOW TO MODERNIZE FOR AI SUCCESS

Data center modernization challenges must be addressed with smart decision-making and strong technology partnerships. Today, numerous industries are taking steps to implement AI-ready infrastructure and seeking external support from providers to deliver end-to-end solutions that simplify and accelerate the use of AI.

Technology partners that meet these requirements can deliver critical ROI for modernization projects, which translates into long-term business value. Consistent value creation can benefit key areas and operations of the business:

- **Workforce productivity:** Increased IT efficiency, reduced operational overhead.
- **IT infrastructure:** Reduced downtime, improved workload efficiency.
- **IT footprint:** Lower real estate costs.
- **Energy efficiency:** Reduced energy consumption, more efficient power and cooling.
- **Growth:** Faster time-to-market, enhanced scalability, rapid feature deployment.
- **Risk mitigation:** Reduced cybersecurity, brand, and operational risks, minimized regulatory penalties.

BUILDING THE DATA CENTER OF THE FUTURE

Modern AI workloads typically demand greater compute power, storage, and flexibility than outdated infrastructure can deliver. In many organizations, substantial portions of IT budgets are allocated to maintaining these legacy systems rather than investing in AI transformation. One thing is certain: harnessing the power of AI requires business goals and IT initiatives as well as a strategic roadmap to bring organizations from planning to production. Choosing the right technology partner is a critical step for enabling the journey and building a modern data center for AI. An effective partnership brings decades of experience, proven solutions, and a track record of success.

AMD envisions AI evolving into a pervasive force across enterprises, spanning hybrid cloud, edge, and endpoint environments. As AI becomes increasingly task-specific and industry-specialized, AMD anticipates that organizations will need a holistic AI integration strategy, embedding AI into both IT and business processes for maximum impact. Instead of a one-size-fits-all approach, AMD prioritizes deep domain expertise and high-quality data, which will define competitive advantage.

AMD differentiates itself as the partner of choice by offering the broadest portfolio of AI solutions, an open ecosystem approach, and proven leadership in innovation. The achievements are significant:

- AMD powers the two fastest supercomputers in the world.[2]
- AMD EPYC™ 9005 CPU-powered servers deliver 2.7x integer performance compared to leading competitive offerings.[3]
- AMD is one of only two major providers offering GPU accelerated computing technology to fuel AI.
- AMD sells the best processor for next gen AI Enterprise PCs.[4]

AMD provides a range of CPU, GPU, and adaptive computing solutions, enabling workload-optimized architectures with freedom of vendor choice. AMD has made a commitment to open standards and flexibility that is augmented by the company's reputation for delivering enterprise-grade performance, security features, and scalability for any type of environment. Organizations can expect optimized total cost of ownership and energy efficiency with the exceptional performance that AI demands. These capabilities make AMD the partner of choice for public cloud (T1 CSPs), on-premises servers, neo-cloud CSPs, and PCs with hundreds of platforms and thousands of instances worldwide.

OPTIMIZING THE ENTIRE AI JOURNEY

AMD has a uniquely broad portfolio of technologies and support to address the challenges of data center modernization and AI. These include integrated technologies, open-standards development and enablement software stack, as well as innovative silicon holding world records in performance and efficiency.[5] Together, these technologies offer a complete solution for the entire AI lifecycle, from data acquisition to model deployment and ongoing optimization.

AMD is a technology leader offering a cohesive, end-to-end solution for AI and general-purpose compute needs across data center, cloud, and at the edge. AMD makes it possible to integrate AI technologies with existing infrastructure, delivering high agility and performance for diverse requirements. Enterprises can tailor deployments to specific workloads, budget constraints, and strategies while easing the management burden of AI hardware. AMD solutions are compatible with existing x86 hardware to help ensure interoperability and a seamless path to AI innovation. Enterprises can choose between cloud-based AI to preserve CAPEX or on-premises AI to drive lower OPEX, often adopting a hybrid model to balance cost and security. This approach can create infrastructure that is both right-sized and AI-ready.

The AMD Enterprise AI portfolio is designed from the ground-up to empower AI in the enterprise. This breakthrough offering is gaining recognition as it enables organizations everywhere to achieve three critical outcomes:

1. **Accelerating business outcomes:** AMD solutions can help maximize space, power, and OPEX savings when modernizing existing data centers. For example, on-premises data centers can replace seven legacy servers with one 2P AMD EPYC 9965-based server for an 87% reduction in physical space while delivering the same 391,000 units of integer performance as a legacy Intel® Xeon® 8280 solution.[6] Existing x86 applications can be retained without needing re-write (unlike Arm-based options), so organizations can train select AI models out-of-the-box to gain immediate ROI.

2. **Providing energy efficiency and security features:** AMD supports this goal with leadership data center CPU and GPU performance-per-watt, which can result in lower space and power utilization. Enterprise AI solutions scale seamlessly to optimize operating costs and accelerate results. AMD maintains an open ecosystem, which enables simplified validation and integration, day zero support from key partners, and compliance with many regulatory guidelines and industry best practices. Organizations also benefit from multilayered security capabilities throughout the entire AI environment. It starts with built-in security features, focused on shared, open-standard security protocols and a robust security feature-set in AMD Infinity Guard as well as extends beyond the chip with AMD Secure Encrypted Virtualization.[7]

3. **Partnering with confidence:** Having an end-to-end product portfolio is a true advantage for enterprises due to its access to experts across the full technical ecosystem. AMD focuses on collaborative partnerships for tailoring solutions to specific workloads. Furthermore, our partnerships are driven by long-term planning to help ensure that AI infrastructure is predictably executed and purpose-built for today's POCs and tomorrow's projects.

END-TO-END SOLUTIONS FROM DATA CENTER TO CLOUD

The AMD Enterprise AI portfolio offers leading solutions that power AI across the data center and cloud. Each technology is purposefully chosen and integrated to power modern data centers. Key benefits of the AMD portfolio include:

Data center CPU performance and security features: AMD EPYC processors deliver the data center computing performance that enterprises need to keep pace with industry demands. They can reduce rack footprint, enable energy-efficient performance, and scale AI workloads effectively. Built into the silicon of AMD EPYC processors, AMD Infinity Guard provides a unique set of security features that helps to protect data from internal and external threats.

GPU acceleration and efficiency: Powered by the 4th Gen AMD CDNA architecture, the 8-GPU AMD Instinct MI355X Platform offers up to 1.03X or 3% the training throughput (tokens per second per GPU) for the LLaMA 3-8B model running Megatron-LM when comparing published LLaMA 3-8B results on NVIDIA HGX B200 8-GPU platform running NeMo, both using FP8 datatype.[8]

Optimized data center networking: AMD provides unique DPUs and NICs to help alleviate the strain that AI workloads put on networks, balancing network throughput with compute capabilities for optimal performance. As a steering member of the Ultra Ethernet consortium, AMD is on the forefront of open networking solutions for AI.

Enterprise AI PCs and workstations: AI is not reserved only for large-scale cloud and data center server deployments. For many end users, AI workloads can be run locally for efficient results. AMD Ryzen™ PRO processors power AI workloads with a range of laptops, desktops, and workstations from OEMs. The AMD Ryzen AI PRO 300 is the best processor for next gen AI Enterprise PCs, offering 53% higher performance in Cinebench R24 nt and delivering up to 9.2 hours of battery life.[9,10] Enterprises can also adopt the AMD Ryzen Threadripper™ PRO to deliver V-ray performance uplift from 45% to 223% on AI-integrated applications.[11]

SUMMARY

AI is radically changing how businesses operate, compete, and innovate, but success hinges on having a high-performance data center. As AI workloads continue to grow in scale and complexity, integrating specialized storage, compute, and efficient networking capabilities are paramount for enabling long-term sustainability and growth.

End-to-end solutions from AMD are designed for diverse AI objectives and long-term success. AMD believes that open ecosystems delivering comprehensive offerings will play a key role in shaping AI, providing greater choice, flexibility, transparency, and the ability to rapidly adapt to evolving market demands. That is why the AMD Enterprise AI portfolio puts enterprises at the forefront of AI data center innovation by offering solutions for any path, any outcome, and any workload.

Together, we advance the future of AI. Let AMD empower your journey.

DISCLAIMER

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. In addition, AMD PRODUCTS may include errata which cause the processor to deviate from published specifications and AMD will occasionally identify such product errata without notice but is under no obligation to do so. 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.

ENDNOTES

1. https://www.amd.com/en/solutions/data-center/insights/data-center-modernization-in-an-ai-driven-world.html
2. TOP500 List, June 2025
3. SPECrate®2017_int_base comparison based on published scores from www.spec.org as of 10/10/2024. 2P AMD EPYC 9965 (3100 SPECrate®2017_int_base, 384 Total Cores, 500W TDP, $14,813 CPU $), 6.200 SPECrate®2017_int_base/CPU W, 0.200 SPECrate®2017_int_base/CPU $ (https://www.spec.org/cpu2017/results/res2024q4/cpu2017-20241004-44979.html); 2P AMD EPYC 9755 (2720 SPECrate®2017_int_base, 256 Total Cores, 500W TDP, $12,984 CPU $), 5.440 SPECrate®2017_int_base/CPU W, 0.209 SPECrate®2017_in t_base/CPU $ (https://www.spec.org/cpu2017/results/res2024q4/ cpu2017-20240923-44824.html); 2P AMD EPYC 9754 (1950 SPECrate®2017_int_base, 256 Total Cores, 360W TDP, $11,900 CPU $), 5.417 SPECrate®2017_int_base/CPU W, 0.164 SPECrate®2017_int_base/ CPU $ (https://www.spec.org/cpu2017/results/res2023q2/cpu2017-20230522-36617.html); 2P AMD EPYC 9654 (1810 SPECrate®2017_int_base, 192 Total Cores, 360W TDP, $11,805 CPU $), 5.028 SPECrate®2017_int_base/CPU W, 0.153 SPECrate®2017_int_base/CPU $ (https://www.spec.org/cpu2017/results/res2024q1/cpu2017-20240129-40896.html); 2P Intel Xeon Platinum 8592+ (1130 SPECrate®2017_int_base, 128 Total Cores, 350W TDP, $11,600 CPU $) 3.229 SPECrate®2017_int_base/CPU W, 0.097 SPECrate®2017_int _base/CPU $ (https://spec.org/cpu2017/results/res2023q4/ cpu2017-20231127-40064.html); 2P Intel Xeon 6780E (1410 SPECrate®2017_int_base, 288 Total Cores, 330W TDP, $11,350 CPU $) 4.273 SPECrate® 2017_int_base/CPU W, 0.124 SPECrate®2017_int_base/ CPU $ (https://spec.org/cpu2017/results/res2024q3/cpu2017-20240811-44406.html). SPEC®, SPEC CPU®, and SPECrate® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information. Intel CPU TDP at https://ark.intel.com/. 9xx5-002E.
4. Based on AMD product specifications and competitive products announced as of March 2025. AMD Ryzen™ AI PRO 300 Series processors' NPU offers up to 55 peak TOPS. This is the most TOPS offered on any system found in enterprise today. AI PC is defined as a laptop PC with a processor that includes a neural processing unit (NPU). STXP-06a.
5. AMD EPYC™ Processor World Records, AMD, 2025 https://www.amd.com/en/products/processors/server/epyc/epyc-world-records.html
6. This scenario contains many assumptions and estimates and, while based on AMD internal research and best approximations, should be considered an example for information purposes only, and not used as a basis for decision making over actual testing. The AMD Server & Greenhouse Gas Emissions TCO (total cost of ownership) Estimator Tool - version 1.12, compares the selected AMD EPYC™ and Intel® Xeon® CPU based server solutions required to deliver a TOTAL_PERFORMANCE of 391000 units of SPECrate2017_int_ base performance as of October 10, 2024. This estimation compares a legacy 2P Intel Xeon 28 core Platinum_8280 based server with a score of 391 versus 2P EPYC 9965 (192C) powered server with a score of 3000 (https://www.spec.org/cpu2017/results/res2024q4/cpu2017-20240923-44837.pdf) along with a comparison upgrade to a 2P Intel Xeon Platinum 8592+ (64C) based server with a score of 1130 (https://www.spec.org/cpu2017/results/res2023q3/cpu2017-20240701-43948.pdf). Actual SPECrate®2017_int_base score for 2P EPYC 9965 will vary based on OEM publications. Environmental impact estimates made leveraging this data, using the Country / Region specific electricity factors from the 2024 International Country Specific Electricity Factors 10 – July 2024, and the United States Environmental Protection Agency 'Greenhouse Gas Equivalencies Calculator'. For additional details, see https://www.amd.com/en/legal/claims/epyc.html#q=epyc4#SP9xxTCO-002A.
7. 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 http://www.amd.com/en/products/processors/ server/epyc/infinity-guard.html.
8. Based on calculations by AMD internal testing of 6/4/2025. Using 8 GPU AMD Instinct MI355X Platform for overall GPU-normalized Training Throughput (processed tokens per second) for text generation using both LLaMA3-70B and LLaMA3-8B chat models running Torchtitan (BF16) or Megatron-LM (BF16) where applicable when using a maximum sequence length of 8192 tokens compared to 8 GPU Nvidia B200 Platform performance running NeMo (BF16) when using a maximum sequence length of 8192 tokens. Server manufacturers may vary configurations, yielding different results. Performance may vary based on use of latest drivers and optimizations. MI350-031.
9. STXP-12: Testing as of Sept 2024 by AMD performance labs on an HP EliteBook X G1a (14in) (40W) with AMD Ryzen AI 9 HX PRO 375 processor, Radeon™ 890M graphics, 32GB of RAM, 512GB SSD, VBS=ON, Windows 11 Pro vs. a Dell Latitude 7450 with an Intel Core Ultra 7 165H processor (vPro enabled), Intel Arc Graphics, VBS=ON, 16GB RAM, 512GB NVMe SSD, Microsoft Windows 11 Pro in the application(s) (Best Performance Mode): Cinebench R24 nT. Laptop manufactures may vary configurations yielding different results. STXP-12.
10. Based on internal testing by AMD as of 9/23/24. Battery life results evaluated by operation of a nine-participant Microsoft Teams video conference on battery. Test configuration for AMD and Intel systems run from power level 90% > 45% @150nits brightness and power mode set to ""best power efficiency."" System config: HP EliteBook X G1a (14in) with an AMD Ryzen AI 9 HX PRO 375 processor (40W), Radeon™ 890M graphics, 32GB RAM, 512GB SSD, VBS=ON, Windows 11 Pro. System config: Apple MacBook Pro 14 with M3 Pro 12-core processor, Apple integrated graphics, 36GB RAM, 1TB SSD, MacOS 15.0. System Config: Dell Latitude 7450 with an Intel Core Ultra 7 165H processor (28W) (vPro enabled), Intel Arc Graphics, VBS=ON, 16GB RAM, 512GB NVMe SSD, Windows 11 Pro. Manufacturers may vary configurations yielding different results. Performance may also vary based on use of latest drivers. STXP-32.
11. Based on AMD performance lab testing as of August, 2023, using the Chaos V-Ray benchmark to compare the performance of the full stack of AMD Ryzen Threadripper PRO 7000 WX-Series processors and the AMD 5965WX and 5955WX processors in a reference system configured with 8x32GB DDR5, NVIDIA Quadro RTX A5000 graphics, 1TB SSD, Win 11 vs. similarly configured BOXX workstations with the full stack of Intel Xeon w-3400 series and the Intel w7-2495X and the Intel w7-2465X processors. Workstation manufacturers may vary configurations, yielding different results. Results may vary. SPP-11.