THE ESSENTIAL BUILDING BLOCKS FOR GENERATIVE AI

Modernize your data center with AMD, Red Hat, and Supermicro

THE RAPID GROWTH OF GENERATIVE AI IS BRINGING THE AVERAGE DATA CENTER TO ITS KNEES.

In fact, **82%** of enterprises have experienced AI workload performance issues in the last year.[1] One reason? Most data centers are built for traditional IT systems, lacking the capacity or resources for escalating AI demands.

Enter AMD, Red Hat, and Supermicro. Here's how to accelerate your AI initiatives with AI-ready infrastructure.

THE CHALLENGES OF GENERATIVE AI WORKLOADS

Generative AI is transforming the workplace, but it's also putting immense pressure on the data center.

70% of senior IT leaders say their current IT infrastructure isn't ready for future AI demands.[2]

3X Global demand for data center capacity could more than triple by 2030.[3]

61% of IT leaders report skills shortages in managing specialized computing infrastructure.[4]

AS-8126GS-TNMR

HOW AI-READY INFRASTRUCTURE CAN HELP

AMD, Red Hat, and Supermicro are working together to propel AI capabilities in the modern data center. The strategic collaboration combines:

- Industry-leading open-source technologies
- Next-generation hardware accelerators
- High-performance servers optimized for AI

The result is ***scalable, cost-efficient, and production-ready infrastructure,*** purpose-built for AI-enabled workloads.

AS-4126GS-NBR-LCC

Red Hat OpenShift AI: Unleash Intelligent Applications and Generative AI

As an add-on to Red Hat OpenShift, OpenShift AI provides a trusted platform for handling the most demanding workloads.

End-to-end AI lifecycle management
Rapidly build, deploy, and manage AI-enabled applications.

Hybrid cloud flexibility
Move workloads to the cloud, on-prem, or the edge, as required by the business.

MLOps integration
Boost efficiency with a consistent user experience for data scientists, developers, and IT teams.

OPTIMIZED FOR AI WORKLOADS
Streamline deployment at scale of the largest AI models, via high-density GPU configurations

ADVANCED COOLING OPTIONS
Lower costs with liquid-cooling and energy-efficient performance

INDUSTRY-PROVEN SYSTEM DESIGNS
Choose from CPU, memory, and networking options tailored for AI environments

SUPERMICRO H14 SERVERS DRIVE AI INNOVATION WITH EXCEPTIONAL TCO

Powered by AMD Instinct™ MI350X GPUs, Supermicro H14 servers are built to supercharge AI performance, while saving energy in the data center.

AMD INSTINCT MI350X GPUS: FUEL THE NEXT GENERATION OF AI INFERENCE

AMD Instinct MI350X accelerators reduce bottlenecks and improve throughput for open-source AI models, like Red Hat's enterprise-grade distribution of vLLM.

Massive Memory for the AI Lifecycle
Accelerate the most demanding workloads with 288GB of HBM3E memory and 8TB/s of bandwidth.

High-compute Performance
Power leading-edge performance of transformer-based models and large language models (LLMs).

Advanced Interconnects
Reduce latency with AMD Infinity Fabric for fast GPU-to-GPU and CPU-GPU data movement.

DISCOVER REAL-TIME VALUE WITHOUT COMPROMISE

Chart your path to AI-ready infrastructure with AMD, Red Hat, and Supermicro.

LEARN MORE

[1] https://www.flexential.com/resources/report/2024-state-ai-infrastructure
[2] https://www.spglobal.com/market-intelligence/en/news-insights/research/ai-infrastructure-trends-thoughts-and-a-2025-research-agenda
[3] https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-power-expanding-data-center-capacity-to-meet-growing-demand
[4] https://www.flexential.com/resources/report/2025-state-ai-infrastructure