

Unleash AI Potential

Accelerate artificial intelligence on your terms with Dell Technologies and AMD



Open source is the engine of AI innovation

Dell Technologies and AMD break down barriers to AI and generative AI (GenAI) with the power and choice of open-source solutions. By combining AMD Instinct™ GPUs, AMD EPYC™ processors, and the open-source AMD ROCm™ software platform on Dell PowerEdge servers, we empower you to scale AI capabilities efficiently, drive innovation and unlock new growth opportunities with ease.



Overcoming AI barriers to entry

AI is becoming essential due to its ability to enhance efficiency, drive innovation and improve decision-making. However, despite the advantages, many organizations run into early roadblocks.

Cost

Rigid hardware commitments and vendor lock-in continue to inflate AI costs — undermining agility and anchoring businesses to legacy systems. Dell Technologies helps organizations sidestep these constraints with flexible consumption models and scalable infrastructure that align spending with outcomes. With Dell PowerEdge XE9785 and XE9785L servers featuring energy-efficient, high-performance AMD Instinct MI355X GPUs, teams can reduce operating costs and automate routine workloads. Together, they help partners focus on outcomes, not obstacles.

Complexity

AI complexity — across architecture, deployment and evolution — slows adoption and raises costs. [Dell Enterprise Hub](#) streamlines AI deployment with curated, containerized models optimized for the PowerEdge XE9680 and AMD MI300X. Combined with Dell's open systems and automation, and AMD's native software compatibility, it accelerates the AI lifecycle with performance and simplicity. Together, they reduce friction, boost control and deliver scalable AI that turns complexity into competitive advantage.

System integration

Dell Technologies supports full rack-level integration and can include full-stack testing and deployment via [Dell AI Services](#) offerings. These services are designed to fit seamlessly into existing enterprise IT environments, with interoperability across open standards and platforms like Kubernetes.

Skills gaps

As AI advances, the demand for specialized skills can stall adoption and slow innovation. Dell Technologies helps bridge this divide with [Dell AI Services](#) designed to reduce complexity with full stack, end-to-end managed solutions. Paired with open-source AMD Instinct GPUs and AMD ROCm software, organizations can reduce reliance on niche expertise and empower broader participation. Together, Dell and AMD help you overcome skills gaps to create momentum for scalable AI growth.

¹ QuantumBlack by McKinsey, [The state of AI: How organizations are rewiring to capture value](#), March 2025.

² McKinsey, [How COOs maximize operational impact from gen AI and agentic AI](#), March 20, 2025.

AI optimism faces opposition

78%

of organizations use AI in at least one business function.¹

63%

use AI in more than two business functions¹

71%

of organizations regularly use GenAI¹

70%

of GenAI high performers find capturing value challenging²

Learn More

[QuantumBlack by McKinsey State of AI report](#)

Seize your AI advantage

Whether you're just getting started or scaling deep learning (DL) across your enterprise, Dell Technologies and AMD are ready to meet you at every stage. With a comprehensive portfolio designed for flexibility and growth, Dell helps you seize new AI-driven opportunities faster.

Dell PowerEdge servers deliver the performance, density and efficiency AI demands. Through strategic collaborations with leading innovators like [Hugging Face](#)® and [Meta](#)®, Dell and AMD offer a unified, open ecosystem that powers every phase of your AI journey – from edge to cloud, workstation to data center.

It's not just infrastructure. It's your competitive edge.

Learn More

Blog

[Dell Technologies and AMD:
Powering Next-Gen AI](#)

Dell and AMD help you harness AI to solve business challenges



Bring AI to your data

Purpose built AI infrastructure



Make smarter decisions

Intelligent AI infrastructure



Future-proof and scale

Flexible AI infrastructure



Maximize seamless uptime

Efficient AI infrastructure



Get started in 4 steps

1. Identify workloads
2. Evaluate validated design
3. Select industry-specific AI agents
 - Insurance claims processing agent
 - Data center infrastructure agent
 - Telecommunications infrastructure agent
 - Healthcare scribe and diagnostics agent
 - Earnings call analysis agent
 - Manufacturing operations agent
 - Legislative bill analysis agent
4. Launch with Dell Services

Strike the best path forward

Dell Technologies and AMD are committed to fostering an open ecosystem that democratizes AI. At the heart of this vision is the Dell AI Platform with AMD. Part of the Dell AI Factory framework, this is a simplified, end-to-end approach that brings together the data, infrastructure, software ecosystem and services to accelerate AI outcomes. From uncovering insights to powering real-time decisions, the AI Factory helps deploy models in modern data centers or at the edge — meeting your organization where it is and taking it where it needs to go.

Simple

Accelerate AI success by removing deployment barriers. Dell Technologies offers turnkey infrastructure and integrated AI services that streamline implementation, reducing complexity and time to value. Paired with cutting-edge hardware and powered by the open-source AMD ROCm software stack, these solutions deliver the performance and flexibility needed to accelerate training and inference across diverse workloads.

Tailored

Maximize profitability and impact with solutions that align tightly with your business goals. The Dell AI Platform with AMD supports flexible scaling, while AMD Instinct GPUs provide efficient, high-throughput acceleration to enhance workload adaptability. This portfolio empowers you to customize AI deployments to optimize resource use and improve long-term return on investment.

Trusted

Enterprise-grade AI that safeguards what matters most gives you the confidence to take control. Secure, on-premises and edge-ready Dell architectures are engineered to protect data sovereignty and intellectual property — critical in regulated and competitive industries. Designed to enhance performance without compromising control, building trust through resilience, transparency and governance.

Learn More

[Dell AI Factory web page](#)

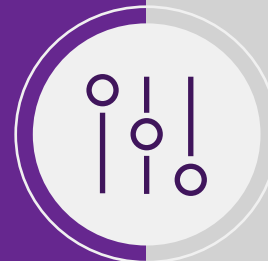
[Dell AI Platform with AMD](#)

[Dell AI web page](#)



Simple

Reduce time to value with an integrated AI ecosystem, infrastructure and services.



Tailored

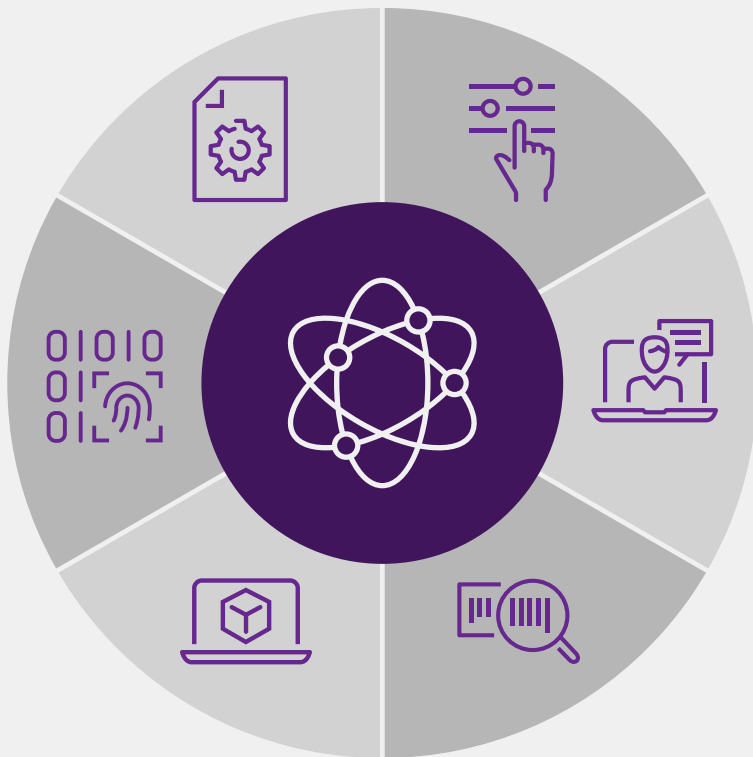
Right-size and optimize for your workloads and business goals.



Trusted

Protect sensitive workloads on-premises by supporting data control, governance and resilience at scale.

Top AI workloads



Content creation

Generate reports and high-quality, on-brand, hyper-personalized content at speed and scale.

Digital assistants

Enhance communication and productivity with more accurate, intuitive and human-like AI interactions.

Design and data creation

Generate synthetic data using algorithms that mimic real-world structures and patterns to enable scenario testing where real data is limited or unavailable.

Code development

Accelerate software development by generating code from natural language prompts and descriptions.

Computer vision

Automate image and video data analysis to detect patterns, extract insights and support decision-making in real time.

Digital twins

Simulate, test and monitor products or processes before deployment through digital replicas.

Learn More

Dell AI solutions web page
Dell.com/AI

eBook
[Drive Powerful AI Insights with an Open Ecosystem](#)

White paper
[Empowering Enterprises with AI: Entering the Era of Choice](#)

Solution brief
[Dell Technologies and AMD for Deep Learning Analytics](#)

White paper
[Dell and AMD for Deep Learning Analytics](#)

Unleash your AI advantage

Dell enables you to harness the power of AI across various environments and applications, accelerating innovation with the help of new Dell PowerEdge servers accelerated by AMD Instinct GPUs. These solutions empower your AI workloads, delivering faster insights, improved inferencing and advanced visualization capabilities for superior AI workload performance.

Accelerate transformation anywhere with Dell PowerEdge AI servers



Accelerated

Advanced processing power

Drive faster results with servers equipped with the latest multi-core AMD EPYC processors and AMD Instinct GPUs.



Efficient

Innovative cooling and power

Drive sustainability with innovative air cooling and direct liquid cooling technologies and smart power tools.



Future-ready

Simple scalability

High-density configurations and flexible storage options make it easy to expand as business demands grow.



Tailored

Optimized for your workloads

Choose from a diverse range of configuration and GPU options to meet workload-specific needs.



Dell PowerEdge AI servers



PowerEdge XE9785

Air-cooled architecture for maximum AI performance.



PowerEdge XE9785L

Direct liquid-cooled architecture for dense AI performance.

Applications and use cases	<ul style="list-style-type: none">• GenAI, AI/ML/DL, high performance computing (HPC)• Large language models (LLMs), recommendation engines, molecular dynamics and genome sequencing	<ul style="list-style-type: none">• GenAI, AI/ML/DL, high performance computing (HPC)• Large language models (LLMs), recommendation engines, molecular dynamics and genome sequencing
Processor	<ul style="list-style-type: none">• 2x 5th Generation AMD EPYC 9005 Series processors with up to 192 cores per processor	<ul style="list-style-type: none">• 2x 5th Generation AMD EPYC 9005 Series processors with up to 192 cores per processor
GPU	<ul style="list-style-type: none">• 8x AMD Instinct MI355x GPUs	<ul style="list-style-type: none">• 8x AMD Instinct MI350X GPUs
Features	<ul style="list-style-type: none">• 10U rack height. 19" rack• Air-cooled• 24x DDR5 DIMM slots• Up to 16x E3.S drives	<ul style="list-style-type: none">• 30U 21 inch IR7000 rack integration• Direct liquid cooling• Up to 24x DDR5 DIMM slots• Up to 16x E1.S drives

Learn More

Dell PowerEdge servers web page
Dell.com/PowerEdge

Spec sheet
[Dell PowerEdge XE9785](#)

Spec sheet
[Dell PowerEdge XE9785L](#)



2TB HBM3E memory
to train the largest LLMs³



Up to 2.3TB with HBM3E
accessible with 8 GPUs via
AMD Infinity Fabric™³



256 high-throughput CUs
per GPU³



Up to 1.3X
AI performance compared
to competitors³



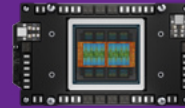
~6.8X
AI training performance compared
to MI250 GPUs³

³ AMD, AMD Instinct™ MI300 Series GPUs, accessed July 2025.

Keep your AI options open

Power AI and HPC with AMD GPUs

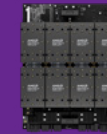
Dell Technologies and AMD are coming together to expand the Dell Technologies GenAI Solutions portfolio, continuing the work of accelerating advanced workloads and offering more choice to continue your unique AI journey. This new technology highlights the pivotal role an open ecosystem and silicon diversity plays in empowering your organization with simple, trusted and tailored solutions to bring AI to your data.



AMD Instinct MI350 GPUs

Leadership AI and HPC performance
for data center infrastructures

- 256 compute units (CU)
- 288GB dedicated HBM3E memory per GPU
- 8TB/s peak memory bandwidth
- Air-cooled



AMD Instinct MI355X GPUs

Leadership AI and HPC performance
enabling high-density infrastructures

- 256 CUs
- 288GB dedicated HBM3E memory per GPU
- 8 TB/s peak memory bandwidth
- Liquid cooling for denser compute environments

Learn More

Dell Technologies AI Servers web page
Dell.com/GPU

Infographic
[Accelerate Insights with
Advanced AI Capabilities](#)

eBook
[Generative AI business value:
Turning POC into ROI](#)

Video
[Dell PowerEdge XE9680
with AMD Instinct MI300X](#)

AMD technologies are built in

Dell PowerEdge servers support AMD technologies that help speed AI and HPC workloads — and results.

AMD Instinct GPUs

AMD Instinct GPUs excel at handling the most challenging AI and HPC workloads, delivering impressive performance, large memory density, high bandwidth memory and compatibility with specialized data formats to drive exceptional results.

- **MI200 Series GPUs:** Enterprise, research and academic AI and HPC workloads at scale.
- **MI300 Series GPUs:** Leadership AI and HPC performance and cost efficient.



Learn More

Blog

[Silicon Diversity: Deploy GenAI on the PowerEdge XE9680 with AMD Instinct MI300X GPUs](#)

Blog

[Delivering Choice for Enterprise AI: Multi-Node Fine-Tuning on Dell PowerEdge XE9680 with AMD Instinct MI300X](#)

Blog

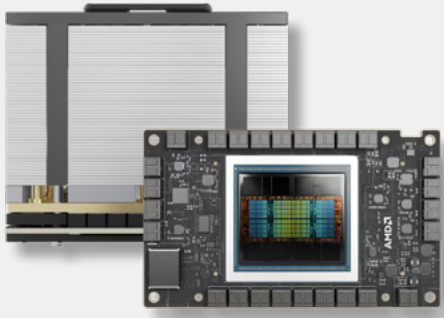
[vLLM Meets Kubernetes-Deploying Llama-3 Models using KServe on Dell PowerEdge XE9680 with AMD MI300X](#)

Blog

[Run Llama 3 on Dell PowerEdge XE9680 and AMD MI300X with vLLM](#)

White paper

[Deploy and Finetune Llama2 70B Chat on PowerEdge XE9680 with AMD Instinct MI300X](#)



Dell, AMD and Hugging Face collaborate to simplify LLMs

This partnership leverages AMD Instinct GPUs for AI, Dell Technologies compute and storage infrastructure, and Hugging Face expertise in open-source models to help you accelerate your AI innovation journey with simplified on-premises deployment of customized LLMs.

AMD ROCm software

As AI evolves at a record pace, the AMD ROCm software platform helps developers move just as fast. Built to be open, scalable and performance-focused, AMD ROCm software empowers teams to push boundaries with fewer constraints. Benefits include:

- Leadership inference performance, with >3.5x gains expected in AMD ROCm 7⁴
- Expanded training capabilities for large-scale models
- Day-one support for major AI workloads, including Llama[®] and DeepSeek[™]
- Frequent updates and support for advanced data types like FP4
- Integration with cutting-edge algorithms such as FAv3
- Deep alignment with open-source frameworks like vLLM and SGLang — advancing innovation faster than closed platforms

AMD Infinity Fabric

Socket-to-socket AMD Infinity Fabric increases CPU-to-CPU transactional speeds with dedicated lanes that allow multiple sockets to communicate directly to one another. Next-gen AMD Infinity architecture, along with AMD Infinity Fabric technology, enables coherent, high-throughput unification of AMD GPUs and CPU chiplet technologies with stacked HBM3 memory in single devices and across multi-device platforms. It also offers enhanced I/O with PCIe 5 compatibility.

Learn More

Blog

[AMD Instinct MI300X GPUs on PowerEdge XE9680 serving Cohere's Command R+ model using vLLM](#)

Blog

[Dell PowerEdge XE9680 servers with AMD Instinct MI300X GPUs: GenAI with Llama 3.1 405B LLMs](#)

White paper

[AMD CDNA[™] 4 Architecture](#)

Blog

[AMD ROCm[™] software blogs](#)

⁴ AMD, [AMD ROCm[™] Software](#), accessed July 2025.

Why Dell Technologies

Collaborate at worldwide Customer Solution Centers

Work with Dell Technologies engineering teams at one of our worldwide [Customer Solution Centers](#), tap into the resources of one of our HPC & AI Centers of Excellence, or test and tune real-world systems at the [HPC & AI Innovation Lab](#).

Consume AI-as-a-Service with Dell APEX

With simple and consistent cloud experiences delivered as-a-Service, [Dell APEX](#) can help you get the AI-optimized solutions you need to fast-track intelligent outcomes everywhere.

Speed success with services

[Dell Technologies Services](#) include consulting, deployment, support and education to help drive rapid adoption and optimization of AI environments. [Managed Services](#) and [Residency Services](#) can help reduce the cost, complexity and risk of managing IT.

Jump-start GenAI objectives

The Dell Accelerator Workshop for Generative AI is a focused half-day session designed to jump-start your strategy. Dell experts work with your team to assess readiness, explore use cases, and define a future-state vision using our proven “AS-IS/TO-BE” approach — all captured in a concise executive overview.

Learn More

Brochure

[Accelerate the power of AI for your data](#)

Accelerate intelligent outcomes with open options

Dell Technologies and AMD empower organizations of every size to unlock the full value of data with proven AI expertise and a flexible, hybrid-ready portfolio. With 35+ data science teams driving 450+ AI projects, Dell helps improve IT efficiency, reduce risk and accelerate insights — whether on-premises, in the cloud or at the edge.

With Dell and AMD, AI success doesn't require vast resources or rare GPUs. Through the Dell and AMD partnership, you gain access to optimized models, reliable software and high-performance infrastructure that's ready to deploy and built to scale.

Dell Technologies and AMD can help you win in the age of AI.

Dell Technologies and AMD Together we advance your world

[Dell Technologies and AMD](#) work together to Drive Innovation Forward for our valued customers. Together, we bring silicon diversity to the market in a viable solution that empowers GenAI LLMs, video analytics, recommendation engines or any other data-intensive AI workloads. Our joint solutions enable you to get to market faster with more options and fewer roadblocks, drive faster insights with an open ecosystem that doesn't hold you back, and choose what's best for you with the reliability of a supply chain that consistently brings innovative technology to market when it's most needed.

Learn more at Dell.com/GPU

© 2025 Dell Inc. or its subsidiaries. All Rights Reserved. Dell and other trademarks are trademarks of Dell Inc. or its subsidiaries. AMD, the AMD Arrow logo, Instinct™, ROCm™, EPYC™, Infinity Fabric™, CDNA™, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Kubernetes® is a registered trademark of The Linux Foundation. Hugging Face® is the registered trademark of Hugging Face, Inc. Meta® and Llama® are registered trademarks of Meta Platforms. DeepSeek™ is a trademark of Hangzhou DeepSeek Artificial Intelligence Co., Ltd. Other trademarks may be the property of their respective owners. Published in the USA 07/25 eBook

Dell Technologies believes the information in this document is accurate as of its publication date. The information is subject to change without notice.