

INSIGHTS | **TECH BRIEF**

Beyond Compute: Build a Secure Foundation for Enterprise AI

Accelerating AI starts with the right IT infrastructure.

Artificial intelligence is changing what businesses expect from their IT infrastructure. From training sophisticated AI models to delivering real-time inference, organizations are discovering that conventional server architectures may lack the computational power, memory bandwidth, and networking capabilities required for demanding AI workloads.

For IT leaders, selecting the right infrastructure for AI initiatives must deliver measurable business value while addressing rising infrastructure costs, power requirements, data protection, and cyber threats. The challenge is finding a platform that can support demanding workloads without compromising security or operational reliability.

Dell PowerEdge XE servers are purpose-built to meet these requirements. Combining advanced GPUs, high-performance processors, intelligent systems management, and cyber-resilient architecture, PowerEdge provides a foundation for organizations looking to accelerate AI initiatives from initial experimentation to production deployment.

Why AI Workloads Demand Specialized Servers

Not all AI workloads are created equal. Training a large language model presents different infrastructure requirements than deploying an AI-powered customer service application. Organizations must account for several factors:

- **Computational Performance:** AI training requires substantial parallel processing capabilities to analyze large datasets and perform intensive mathematical operations.
- **Memory and Data Throughput:** High-bandwidth memory and fast data movement are critical to keeping GPUs productive and minimizing processing delays.
- **Network Performance:** Distributed AI workloads require low-latency, high-bandwidth connections between servers and accelerators.
- **Power and Cooling:** GPU-intensive systems generate significant heat and demand infrastructure capable of supporting higher power densities.

77%

of organizations are either exploring potential use cases or investing significantly in GenAI technologies.[1]



DELL Technologies
PLATINUM PARTNER

These requirements make infrastructure selection a strategic decision. Without the right foundation, organizations risk investing in AI capabilities that cannot deliver the performance or business outcomes they expect.

Dell PowerEdge XE servers address these challenges with architectures designed specifically for AI training, inference, and high-performance computing.

PowerEdge Servers: Purpose-Built for AI Acceleration

Dell PowerEdge XE servers are engineered to deliver the processing capabilities required by today's most demanding AI applications. The portfolio combines advanced CPU and GPU architectures, expansive memory configurations, and high-bandwidth I/O to support data-intensive environments.

Unlike traditional general-purpose servers, these systems are designed to accommodate the unique demands of accelerated computing. This allows organizations to select infrastructure based on their specific AI requirements rather than forcing demanding applications onto hardware that may not be suited for the task.

Accelerating AI Model Training

AI model training requires processing enormous volumes of data through repeated computational operations. As models grow in size, organizations need greater GPU capacity, memory bandwidth, and high-speed communication between accelerators.

Dell PowerEdge XE servers support advanced GPU architectures from NVIDIA and AMD, providing options for intensive machine learning, generative AI, and high-performance computing workloads.

Models such as the PowerEdge XE9680, XE9680L, and XE9685L are designed for demanding AI environments. Dell's broader XE portfolio also includes newer platforms such as the XE9780L and XE9785L, supporting high-density accelerated computing configurations.[2]

These capabilities help organizations:

- Process large datasets for model development and training.
- Support computationally intensive AI and machine learning applications.
- Build infrastructure capable of accommodating larger models and growing processing requirements.

Supporting Scalable AI Inference

Training an AI model is only part of the equation. Once deployed, that model must process requests, generate responses, and deliver results to users and applications.

Inference workloads introduce their own performance requirements. Some applications prioritize low response times, while others must process large numbers of simultaneous requests. Selecting the appropriate GPU and server configuration is essential to balancing performance, capacity, and cost.

The PowerEdge portfolio supports a range of accelerators, allowing organizations to match infrastructure to their inference requirements. Options include NVIDIA L4 and L40S GPUs, alongside higher-performance accelerators for more demanding workloads.

Whether supporting generative AI applications, recommendation engines, or real-time analytics, the right PowerEdge configuration can help organizations move AI models into production while maintaining control over infrastructure investments.

Intelligent Infrastructure Management

Powerful hardware must also be manageable.

Dell's integrated Dell Remote Access Controller (iDRAC) provides secure remote server administration, while OpenManage Enterprise supports infrastructure monitoring and lifecycle management.

Together, these capabilities help IT teams better manage server operations, monitor system health, and maintain greater control over their respective AI environments.

High-Performance AI Requires a Resilient Foundation

Networking: The Critical Connection

Data must move quickly between accelerators, servers, and storage systems to prevent expensive computing resources from sitting idle.

According to Dell Technologies, up to 33% of AI and machine learning processing time can be spent waiting on network resources.[3] This underscores why networking is an essential component of AI infrastructure design.

Dell PowerEdge XE servers can be paired with Dell PowerSwitch Z-Series and SN-Series networking solutions, as well as supported NVIDIA networking technologies, to deliver the high-bandwidth, low-latency connectivity required by demanding AI workloads.

Dell networking technologies support features such as adaptive routing, dynamic load balancing, and RoCEv2 to help manage network traffic and reduce congestion across AI clusters.

For organizations deploying distributed AI training or expanding inference environments, the combination of PowerEdge compute and high-performance networking helps establish a foundation for sustained processing performance.

Addressing Power, Cooling, and Deployment Challenges

As GPU density increases, so do the demands placed on data center infrastructure. Traditional cooling and power configurations may not be sufficient for large AI deployments.

Dell addresses these requirements through advanced thermal designs, air-cooled and liquid-cooled server options, and integrated rack-scale solutions.

Dell PowerRack brings together compute, networking, power, and cooling within factory-integrated and tested rack configurations. This approach can help organizations reduce deployment challenges while supporting high-density AI infrastructure.

Cyber Resilience Remains Essential

While performance is critical, AI infrastructure must also protect the systems and data that power business operations.

AI environments can contain sensitive business information, proprietary models, and valuable intellectual property. Compromised infrastructure can expose these assets while disrupting mission-critical workloads.

Dell PowerEdge servers incorporate security throughout the product lifecycle, from supply chain controls to firmware protection and secure system management.

Key capabilities include:

- **Silicon-Based Root of Trust:** Helps establish a trusted foundation for validating critical system firmware during startup.
- **Secure Component Verification:** Provides factory-to-site integrity assurance to help detect unauthorized changes to supported system components.
- **Role-Based Access Controls:** Helps restrict administrative privileges according to user responsibilities.
- **Multi-Factor Authentication:** Adds protection to supported management access workflows.
- **Secure Remote Management:** Dell iDRAC supports protected administrative access and system management.

These capabilities support a Zero Trust approach in which access must be explicitly authorized rather than automatically trusted.

Cyber resilience does not mean eliminating every potential threat. It means building infrastructure with security controls that help organizations reduce risk, protect critical assets, and maintain trusted operations.

Build Your AI Infrastructure Strategy

The promise of AI is compelling, but realizing its value requires more than purchasing powerful hardware.

Organizations must determine which workloads they intend to support, how much computational capacity those workloads require, and whether their existing infrastructure can accommodate the associated networking, storage, power, and cooling demands.

There is also the question of cost. Overprovisioning AI infrastructure can result in expensive, underutilized resources. Underprovisioning can create performance bottlenecks that limit the value of AI investments.

That is where WEI can help.

WEI works closely with customers to assess their technology requirements, evaluate infrastructure options, and design solutions aligned with business objectives.

Our experienced engineering team can help organizations explore Dell PowerEdge server configurations and supporting technologies based on their specific AI workloads.

Through WEI's Proof of Concept Lab, Integration & Testing Lab, and customized deployment services, customers can evaluate solutions before introducing them into production environments.

WEI also offers predeployment configuration, integration, testing, and ongoing technical support to help customers prepare their infrastructure for operational use.

Talk to WEI Today

AI is placing new demands on enterprise infrastructure, but the underlying objective remains the same: delivering technology that supports measurable business outcomes.

Dell PowerEdge XE servers provide the advanced computing capabilities required for AI training, inference, and data-intensive workloads, while incorporating the security and management features organizations need to protect their investments.

Whether you're beginning your AI journey, expanding an existing deployment, or planning your next infrastructure refresh, WEI can help you evaluate the right combination of Dell PowerEdge compute, networking, and supporting technologies.

Before making your next AI infrastructure investment, talk to WEI about building a secure, high-performance foundation for your business.

Sources:

1. IDC Future Enterprise Resiliency and Spending Survey, Wave 6, July 2023.
2. Dell Technologies, PowerEdge XE Series, product specification sheet, pp. 1–4.
3. Dell Technologies, PowerEdge and Dell Networking for Accelerating AI Workloads, solution brief, pp. 1–2.

About WEI

WEI is an innovative, full service, customer centric IT solutions provider.

Why WEI? Because we care. We go further.

WEI is an expert in business technology improvement, helping clients optimize their technology environments and work efficiently. WEI works with clients to understand goals, integrate strategy with technology solutions, and leverage their current IT environment into one company-wide model to increase utilization and efficiencies around their unique business processes.