Dell Server AI

Article Overview

Dell's AI servers, including the PowerEdge XE and R-series, are purpose-built for high-performance AI workloads with advanced GPU acceleration and scalable infrastructure.

Overview of Dell AI Servers

Dell Technologies offers a range of AI-optimized servers designed to accelerate machine learning, deep learning, and generative AI workloads. These servers leverage NVIDIA GPUs, high-core-count CPUs, and advanced interconnects to deliver maximum throughput and efficiency for AI training, inference, and analytics . Dell's AI server portfolio includes both rack-scale and air-cooled systems, supporting flexible deployment options for enterprises and research institutions .

Key Server Models

o PowerEdge XE9680: A high-performance 6U server featuring eight NVIDIA H100 GPUs, NVIDIA AI software, and dual-socket Intel Xeon processors. It is optimized for large-scale AI model training, NLP, recommender systems, and data analytics, offering nine times more performance and twice the networking speed with NVIDIA ConnectX-7 SmartNICs .
o PowerEdge XE8640: A 4U air-cooled server with four NVIDIA H100 Tensor Core GPUs and NVIDIA NVLink, designed for AI model development, training, and deployment. It supports 4th Gen Intel Xeon Scalable processors and is optimized for PCIe GPU workloads .
o PowerEdge R760xa: A 2U dual-socket server optimized for PCIe GPUs, supporting AMD, Intel, and NVIDIA accelerators. It is suitable for AI training, inferencing, analytics, virtualization, and performance rendering .
o PowerEdge XE8712: Offers the highest GPU density in a standard rack, supporting up to 144 Blackwell GPUs per Dell IR7000 rack, with advanced thermal controls and integrated management software for large-scale AI deployments .
o PowerEdge XE9785/XE9785L: Rack-scale systems purpose-built for next-generation AI and HPC workloads, featuring AMD Epyc processors, AMD Instinct MI355X GPUs, and options for air or direct liquid cooling .

Performance and Features

Dell AI servers are engineered for high GPU density, bandwidth, and power efficiency, enabling faster time-to-insight and scalable AI operations . They support NVIDIA SuperPOD integration, advanced networking, and AI software stacks to accelerate generative AI, deep learning, and analytics workloads. Dell also provides iDRAC, OpenManage Enterprise, and Integrated Rack Controller for monitoring, management, and thermal optimization .

Market Impact

Dell has experienced a surge in AI server demand, reporting a 757% increase in AI-optimized server revenue and a record $51.3 billion backlog, reflecting the growing enterprise adoption of AI infrastructure . Despite strong growth, margins remain thin due to high GPU costs and complex engineering requirements, with NVIDIA handling much of the GPU node design . Dell currently serves over 4,000 enterprise and sovereign AI customers, highlighting its leadership in AI infrastructure .

Summary

Dell's AI servers combine high-performance GPUs, scalable CPU architectures, and advanced management tools to support demanding AI workloads. From large-scale model training to generative AI applications, Dell provides flexible, high-density solutions that enable enterprises to accelerate AI adoption while managing operational efficiency and scalability .

This design guide describes the architecture and design of the Dell Validated Design for Generative AI Inferencing with NVIDIA to

AI Racks and Servers: Delivers a broad portfolio of products including Dell PowerRack and

Dell''s $14.4B AI server backlog shows how its infrastructure is reshaping enterprise IT in 2025. Explore the systems

Featuring liquid-cooled Dell PowerRack systems with Dell PowerEdge XE9712 servers and NVIDIA GB200 NVL72

Drive faster results with servers equipped with the latest multi-core processors. Leverage CPUs to deliver

Dell offers solutions with NVIDIA hardware and software engineered to support AI workloads from PCs with AI

Dell Technologies offers a range of acceleration-optimized servers and an extensive acceleration portfolio with NVIDIA GPUs. Two

Executive summaryAs the digital revolution accelerates, the vision of an AI-powered future

Dell zählt zu den weltweit führenden Anbietern von Servern für KI- und Machine-Learning-Anwendungen. Das Unternehmen hat mit

Dell Server AI

Dell has announced a number of updates to its server, storage, and networking offerings to better support

Dell AI Data Platform updates improve support for full AI lifecycle from ingestion to

Dell AIOps is an AI-management tool, available across the PowerEdge range. It employs adaptable, smart

PowerEdge XE series Unlock the power of AI and high-performance computing with the Dell PowerEdge XE portfolio. A

Boost AI, generative AI, and compute-intensive workloads with servers that offer a variety of powerful GPU accelerators.

Dell"s new high-end Nvidia-based server is a centerpiece for its integrated AI platform aimed at enterprise customers

Dell Technologies rides surging AI server demand with $64B in orders and a $43B backlog, positioning its AI

Web: https://www.morwarreconst.co.za