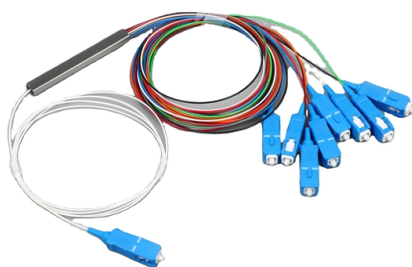


AI server shipment



Overview

Global AI Server Shipments Forecast to Grow Over 28% YoY in 2026, The rapid growth of AI inference services is boosting demand for general-purpose servers.

Key Takeaways

Global AI server shipments are projected to grow over 28% year-on-year in 2026, driven by GPU and ASIC-based systems, with North America leading the market. Market Growth and Forecast

Global AI server shipments are expected to surge by more than 28% YoY in 2026, outpacing the overall server market growth of 12.8% YoY . This growth is fueled by the rapid expansion of AI inference services, generative AI applications, and large-scale cloud infrastructure investments. The global AI server market is projected to reach USD 352.28 billion by 2034, growing from USD 39.23 billion in 2025 at a CAGR of 27.62% . North America remains the largest market, while Asia-Pacific is the fastest-growing region .

Hardware and Technology Trends

- GPU-based servers dominate the market, accounting for approximately 69.7% of shipments, with NVIDIA's GB300 and VR200 platforms driving most deployments .
- ASIC-based AI servers are gaining traction, projected to reach 27.8% of shipments, with Google and Meta expanding in-house ASIC development for AI inference .
- Cooling technologies are evolving, with air-cooled systems still widely used, while liquid and hybrid cooling are increasingly adopted for high-density AI workloads .

- Form factors: Rack-mounted servers dominate due to scalability and modularity, while blade servers are growing rapidly for converged AI infrastructure .

Key Drivers of Shipment Growth

- Cloud Service Provider (CSP) investments: Google, Microsoft, AWS, Meta, and Oracle are increasing capital expenditures by around 40% YoY in 2026 to expand AI infrastructure and replace older servers .

- Generative AI and LLM workloads: The rise of AI agents, Copilot, and LLaMA-based applications is shifting focus toward inference services, increasing demand for both dedicated AI servers and general-purpose servers .

- Edge AI deployment: Organizations are deploying AI servers closer to data sources for real-time processing, boosting demand for compact, energy-efficient servers .

- High-performance computing needs: AI training and inference require GPUs, TPUs, and high-bandwidth memory, driving specialized server adoption .

Regional Insights

- North America: Largest market share, driven by hyperscale cloud providers and government AI initiatives .

- Asia-Pacific: Fastest-growing region, supported by industrial automation, smart city projects, and semiconductor manufacturing capacity .

- Europe: Emerging growth due to AI adoption in enterprise and public sectors .

Market Challenges

- High infrastructure costs and power consumption remain significant barriers .

- Skilled workforce shortage for AI server deployment and maintenance is a limiting factor .

- Supply chain risks from geopolitical instability and component shortages can affect shipment timelines .

Summary

AI server shipments are experiencing rapid expansion, with GPUs leading current deployments and ASICs growing fastest in adoption. North American CSPs are the primary drivers, while Asia-Pacific is emerging as a key growth region. The market is shaped by generative AI, edge computing, and high-performance AI workloads, with continued innovation in server architecture, cooling, and form factors supporting this surge .

Article Content

Global AI Server Shipments Forecast to Grow Over 28% YoY in 2026,

The rapid growth of AI inference services is boosting demand for general-purpose servers, supporting both

AI Server Market Size & Share, Statistics Report 2025-2034

The AI server market was valued at USD 128 billion in 2024 and is expected to grow at a CAGR of 28.2% between 2025 and 2034,

AI Server Shipments Forecast Raised to Nearly 31% YoY in 2026 as

Surging investment in AI infrastructure is expected to drive the combined CapEx of the world's nine largest CSPs up by

Geopolitical Tensions and US Export Restrictions

However, geopolitical tensions and US export restrictions impacting the Chinese market have

Global AI Server Shipments Forecast to Grow Over 28% YoY in 2026,

North American CSPs' continued investments in AI infrastructure are expected to increase global AI server shipments

TrendForce: AI Server Market Will Keep Growing Through 2025, with ...

Consequently, AI server shipments increased by 46% YoY for 2024. As for 2025, TrendForce forecasts that AI server

Geopolitical pressures are impacting AI server growth – but

Global AI server shipments are projected to rise 24.3% in 2025, slightly below forecasts due to U.S. export restrictions

AI Server Compute ASIC Shipments to Triple by 2027 as Custom

Global AI Server Compute ASIC server shipments for the top 10 players are set to triple between 2024 and 2027.

Global AI Server Shipments Forecast to Maintain Double-digit Growth

With sustained demand from North American CSPs and OEM customers, global AI server shipments are projected to

Energy demand from AI – Energy and AI – Analysis

Energy demand from AI What is a data centre? Artificial intelligence (AI) model training and deployment occur mainly in data centres.

Decoding the 2023 server landscape and anticipating a

After a lukewarm 2023, the server industry is poised for a prosperous 2024. Indications from

Global Server Shipments Expected to Increase by 2.05% in 2024, with AI ...

Global server shipments are estimated to reach approximately. 13.654 million units in 2024, an increase of about

Shipments of AI Servers Will Climb at CAGR of 10.8% from

Furthermore, shipments of AI servers are forecasted to increase at a CAGR of 10.8% from 2022 to 2026. TrendForce

AI servers to be 70% of server market in 2025 * Electronics Weekly

The server industry was worth \$306 billion in 2024. says TrendForce, with AI servers accounting for \$205 billion. AI

Global annual AI server shipments, 2024-2025

Global AI server shipments are expected to reach 1.81 million units in 2025, with high-end models no longer

Geopolitical Tensions and US Export Restrictions

However, geopolitical tensions and US export restrictions impacting the Chinese market have

Global AI Server Shipments Forecast to Grow Over 28% YoY in 2026,

North American CSPs'' continued investments in AI infrastructure are expected to increase global AI server shipments

TrendForce: AI Server Market Will Keep Growing Through 2025, with ...

Consequently, AI server shipments increased by 46% YoY for 2024. As for 2025, TrendForce forecasts that AI server

Geopolitical pressures are impacting AI server growth – but

Global AI server shipments are projected to rise 24.3% in 2025, slightly below forecasts due to U.S. export restrictions

AI Server Compute ASIC Shipments to Triple by 2027 as Custom

Global AI Server Compute ASIC server shipments for the top 10 players are set to triple between 2024 and 2027.

Global AI Server Shipments Forecast to Maintain Double-digit Growth

With sustained demand from North American CSPs and OEM customers, global AI server shipments are projected to

Global AI Server Shipments Forecast to Grow Over 28% YoY in 2026,

The rapid growth of AI inference services is boosting demand for general-purpose servers, supporting both

AI Server Market Size & Share, Statistics Report 2025-2034

The AI server market was valued at USD 128 billion in 2024 and is expected to grow at a CAGR of 28.2% between 2025 and 2034,

AI Server Shipments Forecast Raised to Nearly 31% YoY in 2026 as

Surging investment in AI infrastructure is expected to drive the combined CapEx of the world's nine largest CSPs up by

AI server Market Size, Growth Report, 2026-2033

The rapid adoption of generative AI applications across various industries is driving market growth, with

Server Shipment

Description Server Shipment Published by TRENDFORCE TrendForce Frequency Quarterly Next Update

Servers Market Insights

Worldwide Server Market grew by 30.7% in spending in the first quarter of 2026, driven by the continued mass

Global AI Server Shipments Forecast to Grow Over 28% YoY in 2026,

The rapid growth of AI inference services is boosting demand for general-purpose servers, supporting both

AI Server Market Size & Share, Statistics Report 2025-2034

The AI server market was valued at USD 128 billion in 2024 and is expected to grow at a CAGR of 28.2% between 2025 and 2034,

AI Server Shipments Forecast Raised to Nearly 31% YoY in 2026 as

Surging investment in AI infrastructure is expected to drive the combined CapEx of the world's nine largest CSPs up by

AI server Market Size, Growth Report, 2026-2033

The rapid adoption of generative AI applications across various industries is driving market growth, with

Contact Us

Continue your research online:

Website: <https://thegalleryselfcatering.co.za>

Technical inquiry: <https://thegalleryselfcatering.co.za/contact.html>

This document is for preliminary research. Verify specifications against current standards, product documentation and project requirements.

