Huawei procures AI servers

Overview

A new report states that Huawei and Cambricon, the leading chip suppliers in China''s AI segment, will rule over the country''s AI server market, phasing out Nvidia from the game.

Key Takeaways

Huawei is actively deploying advanced AI servers, including the Atlas 950 and 960 SuperPoD clusters powered by Ascend AI chips, capturing a significant share of China's AI server market.

Latest AI Server Models

Huawei recently unveiled the Atlas 950 and Atlas 960 SuperPoD AI clusters, designed to compete with NVIDIA's AI servers. The Atlas 950 SuperPoD features 8,192 Ascend 950 AI chips, delivering 8 EFLOPS FP8 and 16 EFLOPS FP16 performance with 16. 3 PB/s interconnect bandwidth.

These clusters can scale into Atlas SuperClusters with up to 524,288 Ascend 950 chips, making them among the largest AI compute clusters globally. The Atlas 960 SuperCluster is projected to reach zettaFLOP-level compute power with nearly a million NPUs by 2027.

A

© 2026 ALFA PRO COMMUNICATION SYSTEMS (Pty) Ltd - All rights reserved

I Chip Usage and Market ShareHuawei's Ascend AI chips have become a cornerstone of China's domestic AI server ecosystem. In 2025, Huawei shipped approximately 812,000 AI chip units, contributing to Chinese firms capturing 41% of the domestic AI accelerator server market, while NVIDIA's share declined to 55%. The Ascend 910B and 910C chips are widely used for inference workloads, and the performance gap with NVIDIA's H100 and H200 chips is narrowing.

Enterprise Adoption and ApplicationsHuawei's AI servers are deployed across multiple industries, including manufacturing, healthcare, agriculture, aviation, and automotive. The CloudMatrix384 supernode enables full pooling of compute, memory, and storage resources, supporting distributed parallel AI tasks and delivering 3–4 times higher inference performance per card compared to previous models. Huawei also offers Elastic Memory Service to reduce latency in multi-round AI model interactions.

Additionally, “all-in-one” servers with embedded DeepSeek AI models allow enterprises to quickly deploy AI applications on-premises, meeting security requirements for sensitive data. Infrastructure and EcosystemHuawei's AI infrastructure integrates Kunpeng ARM processors, TaiShan servers, and HarmonyOS, providing a vertically integrated ecosystem that reduces reliance on x86 hardware.

The Unified

Bus 2. 0 interconnect links thousands of NPUs in SuperPoDs and scales to millions in future SuperClusters, surpassing traditional NVLink and RoCE interconnects. Huawei's open-source AI stack, including MindSpore, CANN, and openEuler, offers a full-stack alternative to NVIDIA's CUDA, supporting enterprise AI adoption globally.

Summary

Huawei is aggressively expanding its AI server procurement and deployment, leveraging Ascend AI chips, large-scale SuperPoD clusters, and integrated software-hardware ecosystems. These efforts position Huawei as a major player in China's AI server market, enabling enterprises to deploy high-performance AI workloads while reducing dependence on Western hardware providers.

© 2026 ALFA PRO COMMUNICATION SYSTEMS (Pty) Ltd - All rights reserved

Article Content

Huawei set to dominate China''s AI server market, phasing out Nvidia

A new report states that Huawei and Cambricon, the leading chip suppliers in China''s AI segment, will rule over the country''s AI server market, phasing out Nvidia from the game. Both firms

China Telecom''s $1.7bn server deal hands Huawei a big win

China Telecom''s 40,000-server, $1.7bn order went mostly to Huawei''s Kunpeng ecosystem, as Beijing swaps US chips out of its digital backbone.

Huawei Unveils Next-Gen Atlas 950 & 960 SuperPoD AI Servers

Huawei claims to be taking the competition in the rack-scale segment directly to NVIDIA''s ground, as their latest announcement includes unveiling cutting-edge AI clusters.

Huawei announces new AI infrastructure as Nvidia gets locked out of ...

Tech giant Huawei unveiled new AI infrastructure meant to help boost compute power and allow the company to better compete with rival chipmaker Nvidia.

The New Cold War Is Over GPUs: How Sovereign AI Drives 2026

Sovereign AI server procurement is reshaping global IT infrastructure as nations build national AI supercomputing centers to avoid reliance on foreign cloud models. NVIDIA''s 2026 Q1 earnings

Huawei Advances AI Servers Amid US Restrictions

TOKYO -- Huawei Technologies is steadily building up its own artificial intelligence (AI) infrastructure with homegrown chips and servers, underscoring China''s progress on AI development

Huawei''s Ascend and Kunpeng progress shows how China is

Huawei used its New Year message to highlight progress across its Ascend AI and Kunpeng CPU ecosystems, pointing to the rollout of Atlas 900 supernodes and rapid growth in

China''s AI Compute Domestic Substitution

Domestic AI accelerators already account for 41% of China''s shipment market and is approaching 60% this year. China''s domestic substitution machine is compouding, with Huawei

Tech war: Huawei wields growing influence in China''s

© 2026 ALFA PRO COMMUNICATION SYSTEMS (Pty) Ltd - All rights reserved

Huawei Technologies, which was forced to divest its server subsidiary three years ago due to US tech sanctions, is regaining its influence in

Chinese Nvidia Cloud Partner procured 300 servers with banned AI

Chinese Nvidia Cloud Partner procured 300 servers with banned AI GPUs worth $92 million — shares of data center supplier Sharetronic plummet following Super Micro smuggling arrest

Huawei Connect 2025: Building AI Infrastructure In A

At this year''s Huawei Connect topics such as export controls and supply chain constraints dominated the conversation. Find out more in this event

Huawei launches new AI server solution – Trivium China

What happened: On April 10, Huawei launched its new AI server system, CloudMatrix 384. Huawei claims it outperforms Nvidia''s GB200 NVL72 system. But the comparison isn''t apples-to

Huawei, Inspur champion DeepSeek-powered all-in-one

With AI adoption accelerating across industries, Chinese tech firms are simplifying AI infrastructure to meet growing demand. Many are rolling out all

Huawei Atlas AI Computing Platform: A Gift for

The Atlas AI computing platform provides powerful AI computing for customers to handle massive data volumes. It plays an important role in the

Huawei Releases the Full-Stack Data Infrastructure Solution of AI

Against this backdrop, Huawei unveiled the full-stack data infrastructure solution of AI DCs to accelerate AI DC construction and large-scale AI adoption for enterprises. AI data lake High

Huawei launches CloudMatrix 384 server as an alternative to Nvidia''s AI ...

Huawei launches CloudMatrix 384 server as an alternative to Nvidia''s AI infrastructure stack - SiliconANGLE SiliconANGLE Media is a recognized leader in digital media innovation, uniting ...

Atlas: Opening the door to AI with massive computing power

Huawei is accelerating its research in AI, and continues to embrace the data processing challenges of the AI era with a full-stack AI solution. A key part of Huawei''s full-stack AI solution, the Atlas

Huawei challenges Nvidia with AI super server

© 2026 ALFA PRO COMMUNICATION SYSTEMS (Pty) Ltd - All rights reserved

Huawei challenges Nvidia with AI super server The AI server race heats up as Huawei counters US chip export restrictions. Huawei has unveiled its most powerful AI server, the

Contact Us

Continue your research online:

Website: https://alfa-pro.co.za
Technical inquiry: https://alfa-pro.co.za/contact.html

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

© 2026 ALFA PRO COMMUNICATION SYSTEMS (Pty) Ltd - All rights reserved