

Huawei AI Server



Overview

The Huawei CloudMatrix 384 is a high-density AI computing system featuring 384 Huawei Ascend 910C chips, designed to rival Nvidia's GB200 NVL72 (more below). The AI system employs a “supernode” architecture with high-speed internal chip interconnects. AI Compute Service offers instant access to immense yet cost-effective AI computing power, a reliable platform for training and running models and algorithms, E2E cloud-based toolchains, and a robust AI ecosystem, with support for all major open-source foundation models. Huawei has officially launched the Ascend 950PR, a 1. 56-petaflop AI inference chip that delivers 2. With 750,000. The rapid evolution of large language models (LLMs), driven by growing parameter scales, adoption of mixture-of-experts (MoE) architectures, and expanding context lengths, imposes unprecedented demands on AI infrastructure. threw down the gauntlet to Nvidia Corp. Saturday when it revealed its most powerful artificial intelligence server system to date. The company unveiled the CloudMatrix 384 system at the World Artificial Intelligence Conference in Shanghai, where dozens of local. 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. The Chinese AI firm has been at the forefront of competing with NVIDIA in China's AI market, particularly with rack-scale.

Article Content

Malaysia withdraws plan to deploy Huawei AI servers

The original plan included deploying 3,000 Huawei AI servers by 2026 and collaborating with DeepSeek.

Huawei updates AI WAN solution for autonomous

Huawei released an upgrade to its AI-driven WAN solution in the Chinese giant's latest development of autonomous network infrastructure. As

Serving Large Language Models on Huawei CloudMatrix384

Addressing these issues requires fundamentally redesigned hardware-software integration. This paper introduces Huawei CloudMatrix, a next-generation AI datacenter architecture,

Huawei launches CloudMatrix 384 AI System to rival

The Huawei CloudMatrix 384 is a high-density AI computing system featuring 384 Huawei Ascend 910C chips, designed to rival Nvidia's GB200

Huawei's Ascend 910C AI Chip Cluster "CloudMatrix"

Huawei's Ascend 910C AI chips can hit the markets anytime, and, the Chinese firm has managed to narrow down the gap with NVIDIA.

Foxconn Industrial Internet surpasses Huawei in 2025 revenue, driven

Foxconn's industrial arm, Foxconn Industrial Internet (FII), has ridden the AI server boom to become a revenue leader, illustrating to global manufacturers how advanced production capability

Why liquid cooling will dominate AI data centres in 2026

As AI power demands surge into 2026, liquid cooling is becoming the essential technology keeping data centres efficient, stable, and future-ready.

Huawei AI chip gains orders from ByteDance and Alibaba amid

Huawei is gaining renewed traction in the AI semiconductor market, with ByteDance and Alibaba reportedly planning to place orders for its latest AI chip, according to sources familiar with the

Huawei Unveils Next-Gen Atlas 950 & 960 SuperPoD

Huawei claims to be taking the competition directly to NVIDIA's ground, as their announcement includes unveiling AI clusters.

How is Huawei's Ascend 910B outperforming Nvidia's

Nvidia jumps to FOPLP for AI server chips amid CoWoS capacity strain Huawei defies US sanctions, surges in profit, and expands ecosystem in

Huawei targets 60% jump in AI chip revenue as DeepSeek drives

Huawei expects revenue from its AI chips to reach approximately \$12 billion this year, up from \$7.5 billion in 2025, a jump of at least 60% based on orders already secured, the Financial

Huawei Malaysia Launches First AI Lab & Innovation Centre At The ...

Huawei officially launched its first AI Lab and Innovation Centre outside of China to serve the whole Huawei Asia Pacific region at The Exchange 106 @ TRX. This also marks Huawei's first AI

Huawei AI CloudMatrix 384 – China's Answer to Nvidia

Meet China's newest and most powerful Chinese domestic solution, the CloudMatrix 384 built using the Ascend 910C. This solution competes

Huawei announces new AI infrastructure as Nvidia gets

Huawei's SuperPoD interconnect technology creates clusters of chips, including AI chips, to increase compute.

Huawei, operators outline AI-5G-A convergence at

The president of Huawei's ICT marketing and solution sales department said the AI revolution is pushing the mobile industry into what he

Huawei Ascend 950PR: The 1.56 PFLOP AI Chip vs Nvidia

At least two non-Chinese countries will sign major Huawei AI infrastructure deals by mid-2027. The most likely candidates are Saudi Arabia, the UAE, and Indonesia, all of which have

Huawei Launches Industry's First AI Core Network,

The AI Core Network will be an autonomous generative network capable of self-optimization and self-O& M, enabling the whole industry to evolve

Huawei AI server out-performs Nvidia's most powerful

At \$8 million, the price of the Huawei cluster is 3x the price of the Nvidia GB200 NVL72, while power consumption is almost 4x more, which

DeepSeek V4 triggers scramble for Huawei AI chips as US export

DeepSeek's endorsement changes that calculus. "Huawei's Ascend chips are the country's best homegrown alternative to Nvidia, and supporting DeepSeek V4 shows that top

Huawei Unveils New Ascend AI and Kunpeng Server

Huawei has publicly detailed its semiconductor strategy to develop advanced AI and server processors, aiming to reduce China's reliance on

Huawei Connect 2025: Building AI Infrastructure In A

ARM-Based Kunpeng processors: Huawei's in-house ARM CPUs now power laptops, servers, and AI infrastructure, reducing reliance on x86 and

Huawei Unveils Next-Gen Atlas 950 & 960 SuperPoD

Huawei claims to be taking the competition in the rack-scale segment directly to NVIDIA's ground, as their latest announcement includes

Chinese AI lab DeepSeek to run v4 on Huawei chips

One comparison pairs Huawei's CloudMatrix 384 (a large AI server cluster) with Nvidia's GB200 NVL72 (a high-end AI server system based on NVIDIA's Blackwell platform) 3.

Exclusive: Huawei readies new AI chip for mass

Huawei Technologies plans to begin mass shipments of its advanced 910C artificial intelligence chip to Chinese customers as early as next month,

Huawei launches CloudMatrix 384 server as an

CloudMatrix 384 is powered by 384 of Huawei's most sophisticated AI chips, the Ascend 910C graphics processing unit. That compares with just 72

Huawei Launches Its AI Data Platform to Power ...

At the Huawei AI DC Innovation Forum during MWC Barcelona 2026, Huawei unveiled its AI Data Platform, designed to address the key challenges in adopting AI agents and strengthen the

Contact Us

For more information, pricing, or custom solutions, please contact us:

Website: <https://www.truhope.co.za>

Email: sales@truhope.co.za

Phone: +27 64 987 3021

Address: 22 Loop Street, Cape Town, 8001, South Africa

This document is for informational purposes only. Specifications subject to change without notice.

