

How big is Huawei's AI server



Article Overview

Huawei's AI servers, including the CloudMatrix 384 and Atlas 950/960 SuperPoD clusters, range from hundreds to over half a million AI chips, delivering up to 300 petaflops to multi-exaflop performance.

CloudMatrix 384

Huawei's CloudMatrix 384 is a high-performance AI server featuring 384 Ascend 910C AI chips connected via a proprietary "super-node" interconnect. This design allows the system to achieve approximately 300 petaflops of computing power, nearly double that of Nvidia's comparable NVL72 platform, though it consumes almost four times more energy. The server is optimized for large-scale AI model training and is priced around \$8 million per unit. Its architecture emphasizes chip stacking and high-speed interconnection to maximize aggregate performance despite individual chip limitations .

Atlas 950 and 960 SuperPoD Clusters

Huawei has also developed the Atlas 950 and Atlas 960 SuperPoD AI clusters, which are significantly larger in scale:

- Atlas 950 SuperPoD: Contains 8,192 Ascend 950 AI chips, delivering 8 EFLOPS FP8 and 16 EFLOPS FP16 performance with a total interconnect bandwidth of 16.3 PB/s .
- Atlas 950 SuperCluster: By combining multiple SuperPoD units, Huawei can scale up to 524,288 Ascend 950 AI chips, with potential expansion to 500,000-1 million processors, making it one of the largest AI compute clusters in the world

Article Content

Huawei Unveils World's Most Powerful SuperPoDs and

With the world's most powerful SuperPoDs and SuperClusters, Xu said Huawei has what it

Huawei challenges Nvidia with AI super server

Huawei has unveiled its most powerful AI server, the CloudMatrix 384, to challenge Nvidia's grip on the high

Inside Huawei's Atlas 960: How China's SuperPoDs Aim to ...

This article explores how Huawei's strategy of tightly coupling cutting-edge interconnect technology with large-scale compute nodes

Inside Huawei's plan to make thousands of AI chips think like one

Imagine connecting thousands of powerful AI chips scattered in dozens of server cabinets and making them work

Huawei's AI servers show tech advancing despite US curbs

TOKYO -- Huawei Technologies is steadily building up its own artificial intelligence infrastructure with homegrown chips and servers,

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,

Huawei AI server out-performs Nvidia's most powerful AI server

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

Huawei's Ascend AI chip ecosystem scales up as China

Huawei's AI server hardware, domestic packaging suppliers, and EDA partners now form

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

Huawei's Ascend-powered CloudMatrix platform tops

Huawei Technologies ' advanced data centre architecture powered by its Ascend chips

Huawei Unveils Atlas 950 AI SuperPoD to Compete

At MWC 2026, Huawei unveiled its Atlas 950 and TaiShan 950 SuperPoDs to a global AI server Market Size, Growth Report, 2026-2033

AI server market size was valued at \$131.7B in 2025 and is projected to grow from \$157.0B in 2026 to \$598.1B by 2033, at a CAGR

Huawei Unveils Next-Gen Atlas 950 & 960 SuperPoD AI

Huawei claims that the Atlas 950 SuperPoD will be the world's largest AI cluster, in terms of

Huawei Atlas 950 AI Supernode, 8192 Chips Challenge Nvidia

While Silicon Valley debates the ethics of AI, Huawei ships 8,192-chip supernodes. By 2027, they'll deploy million

AI Server Market report 2024-2030 [314 Pages & 252 Tables]

The global AI server market is expected to grow from USD 142.88 billion in 2024 to USD 837.83 billion by 2030, at a CAGR of 34.3%

Huawei's Ascend-powered CloudMatrix platform tops biennial tech ...

The system can support up to tens of thousands of interconnected chips to train and run AI models.

Atlas 500 Pro AI Edge Server User Guide (Model 3000) 18

The Atlas 500 Pro (model 3000) is a 2 U AI edge server powered by Huawei Kunpeng 920 processors, featuring superb computing

Atlas 800 Inference Server User Guide (Model 3000) 17

The Atlas 800 inference server (model 3000) is a data center server powered by Huawei Kunpeng 920 processors. It

Huawei's CloudMatrix 384 Roars Past Nvidia: China's AI Giant

With CloudMatrix AI clusters already deployed in 16 provinces across China, and Huawei's ecosystem gaining

Huawei AI server out-performs Nvidia's most powerful AI server

Although it costs three times more, and uses 3.9x the power of Nvidia's most powerful AI server the GB200 NVL72,

China AI Server Companies — Data Center Infrastructure Stocks | 2026

Quick Answers How big is China's AI server market? China's AI server market exceeded \$30 billion in 2025, driven by hyperscale AI

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

Huawei Launches Open-Access SuperPoD Architecture

Unleashing the potential of cluster-based compute with the SuperPoD architecture As AI

OceanStor AI Storage

As large AI models move into the multimodal era, enterprises face challenges in model training and

GPU Accelerated Cloud Server_ECS_HUAWEI CLOUD

GPU-accelerated cloud server (GACS) provides outstanding floating-point computing power that is great for real-time, highly

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

Huawei makes some stunning claims, saying that the CloudMatrix 384 delivers around 300 petaflops of compute

Huawei's CloudMatrix AI Supercomputer: A New Force in AI Compute ...

According to Huawei, it offers the scalability needed for foundation model training and is already being deployed by

Huawei Takes 60% of Chinese AI Market After NVIDIA Exit

Huawei is set to dominate 60% of China's AI market following NVIDIA's departure, powering a \$12 billion revenue surge.

Ascend AI Cloud Service | Huawei Cloud

Ascend AI Cloud Service offers instant access to immense yet cost-effective AI computing power, a reliable platform for training and

Huawei: SuperPoD Interconnect: Leading a New Paradigm for AI

This brings me to the topic of today's keynote: Groundbreaking SuperPoD Interconnect: Leading a New Paradigm for

Contact Us

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

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

Email: sales@thebucketlistbook.co.za

Phone: +33 1 42 78 39 16

Address: 10 Rue de la Paix, 75002 Paris, France

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

