# Huawei begins deliveries of CloudMatrix 384 AI chip cluster amid US export restrictions



Huawei has initiated deliveries of its advanced artificial intelligence chip cluster, the CloudMatrix 384, to Chinese clients amid increasing orders following the company’s exclusion from Nvidia’s semiconductor supply due to US export restrictions. This development was reported by the Financial Times, citing two sources familiar with the matter.

The CloudMatrix 384 chip cluster, developed by Huawei’s Shenzhen-based technology conglomerate, comprises a system that links 384 AI processors together. It is designed to offer the substantial computing power required for creating and running sophisticated AI models and services. Early recipients of these chip clusters include data centres servicing Chinese technology companies.

Industry analysts have expressed surprise and admiration at the rapid pace with which Huawei developed and began shipping the CloudMatrix 384. According to Dylan Patel, founder of the chip consultancy firm SemiAnalysis, “The development of Huawei’s CloudMatrix 384 means China now has an AI system capable of beating Nvidia’s.”

Huawei’s product introduction comes amid further tightening of Chinese companies’ access to Nvidia’s AI chips. Nvidia recently announced it would incur a $5.5 billion reduction in earnings after new US export controls imposed by President Donald Trump required special licences for sales of its H20 AI chip to Chinese customers. These restrictions have pushed Huawei to accelerate its in-house chip development efforts to support Chinese technology enterprises competing on a global scale without access to leading international semiconductor technologies.

Huawei has informed customers that its CloudMatrix 384 cluster outperforms Nvidia’s NVL72 cluster, a widely used system among US tech companies that includes 72 Nvidia GB200 chips. According to company presentations and insider sources reviewed by the Financial Times, CloudMatrix 384 surpasses NVL72 by 67 percent in computing power and offers more than three times the aggregate memory capacity.

However, this performance is achieved despite Huawei’s individual chips, the Ascend 910C, lagging behind Nvidia’s GB200 chip in raw processing power. Huawei compensates for this by using a larger number of connected chips, linked via its proprietary “super node” technology. This system connects all processors optically to enhance overall cluster performance. Patel commented, “It is compensating for weaker individual chip performance with advanced networking to boost the chip performance in a cluster.”

Huawei’s deep expertise in telecommunications infrastructure has reportedly enabled the company to optimise networking capabilities within its AI chip clusters to gain competitive advantages.

Notwithstanding its performance metrics, the CloudMatrix 384 system has several drawbacks when compared to Nvidia’s NVL72. The Huawei cluster’s larger number of chips results in significantly higher energy consumption, leading to increased electricity costs. Additionally, its software infrastructure is considered less mature and requires more maintenance, demanding skilled engineers and thereby inflating operational manpower expenses by three to five times relative to Nvidia’s CUDA ecosystem.

Nonetheless, insiders emphasised that these disadvantages are offset by the relative abundance of electrical power and engineering talent within China, making CloudMatrix a practical alternative for customers deprived of access to Nvidia’s premium AI technology.

Pricing for the CloudMatrix 384 is reported to be around RMB 60 million (approximately $8.2 million) per set, with variations depending on specific contract terms. In contrast, Nvidia’s NVL72 cluster is estimated by analysts to cost roughly $3 million, though Nvidia maintains that their pricing varies based on manufacturer and client specifications.

Huawei declined to comment on the matter, while Nvidia acknowledged a range of pricing based on customer and original equipment manufacturer requirements.

This advancement underscores the growing capabilities of domestic Chinese technology companies in AI hardware development, especially in response to evolving global trade restrictions impacting the semiconductor sector.

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.ft.com/content/cac568a2-5fd1-455c-b985-f3a8ce31c097> - This Financial Times article reports on Huawei's delivery of the CloudMatrix 384 AI chip cluster to Chinese clients, highlighting its development and performance compared to Nvidia's NVL72 system.
2. <https://www.businesstimes.com.sg/companies-markets/telcos-media-tech/baidu-placed-ai-chip-order-huawei-shift-away-nvidia-sources> - This Business Times article discusses Baidu's order of Huawei's Ascend 910B chips, indicating a shift away from Nvidia's products due to U.S. export restrictions.
3. <https://www.tomshardware.com/tech-industry/artificial-intelligence/huaweis-new-ai-cloudmatrix-cluster-beats-nvidias-gb200-by-brute-force-uses-4x-the-power> - This Tom's Hardware article details Huawei's CloudMatrix 384 AI chip cluster, emphasizing its performance surpassing Nvidia's GB200 NVL72 system and its higher power consumption.
4. <https://www.huaweicentral.com/new-us-ai-chip-ban-on-china-could-become-a-golden-opportunity-for-huawei/> - This Huawei Central article explores how the U.S. AI chip ban on China presents an opportunity for Huawei to expand its AI hardware offerings domestically.
5. <https://www.tomshardware.com/pc-components/gpus/huawei-introduces-the-ascend-920-ai-chip-to-fill-the-void-left-by-nvidias-h20> - This Tom's Hardware article reports on Huawei's introduction of the Ascend 920 AI chip, designed to fill the market gap left by Nvidia's H20 chip due to U.S. export restrictions.
6. <https://www.chinadaily.com.cn/a/202409/23/WS651c8b8da310d2dce4b8c8b0.html> - This China Daily article discusses Huawei Cloud's AI initiatives, including the launch of the CloudMatrix 384 super node, highlighting its role in reshaping industries with AI.
7. <https://www.ft.com/content/cac568a2-5fd1-455c-b985-f3a8ce31c097> - Please view link - unable to able to access data