Huawei AI Server Cluster



Overview

Now, at the Huawei Connect 2025, the firm has announced new iterations of its 'SuperPoD' AI clusters. These will be the Atlas 950 and the Atlas 960, with the earlier one featuring the new Ascend AI chips, and interestingly, will compete with NVIDIA's Rubin lineup. The company unveiled the CloudMatrix 384 system at the World Artificial Intelligence Conference in Shanghai, where dozens of local companies showed off their latest AI hardware. At the recent World AI Conference in Shanghai, Huawei unveiled the CloudMatrix 384, a massive AI cluster designed to serve China's growing demand for large-scale model training—at a time when access. Huawei announced its CloudMatrix 384 AI system a few months ago, which was reportedly to have surpassed NVIDIA's Blackwell AI system. The latency for token generation is reduced to less than 10 milliseconds, with a throughput of 5 million tokens. On April 10, 2025, at the "Huawei Cloud Ecosystem Conference 2025" held in Wuhu, Anhui, Zhang Pingan, executive director of Huawei and CEO of Huawei Cloud Computing, announced a major breakthrough in AI infrastructure architecture - the launch of CloudMatrix 384 super-node cluster based on a new. Although it costs three times more, and uses 3. So China can resource internally all the computing power it needs to pursue AI development.



© 2026 BAGYOU FIBRE NETWORKS - All rights reserved

Article Content

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

At the recent World AI Conference in Shanghai, Huawei unveiled the CloudMatrix 384, a massive AI cluster designed to serve China''s growing demand for large-scale model training—at a

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 the chips onboard, since systems with 500K to 1

Huawei Cloud announces agentic AI products

At this event, Huawei Cloud officially introduced Agentic Infra, a new paradigm featuring an efficient token factory, continuous learning, unified general & AI compute scheduling, and secure

Huawei releases CloudMatrix 384 supernode: a milestone ...

CloudMatrix 384 supernode is Huawei''s latest masterpiece in the field of AI infrastructure. Its core innovation is to achieve the transformation of resource supply mode from traditional server

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 almost 4x more, which shows that Huawei has developed the cluster to

Huawei launches CloudMatrix 384 AI chip cluster

Today, Huawei launches the CloudMatrix 384 AI chip cluster at the WAIC 2025 event. After days of talks, the company has officially introduced its

Huawei Unveiled the Latest SuperPoD, Making an AI Infrastructure

For AI computing, the Atlas 950 SuperPoD, powered by UnifiedBus, integrates 64 NPUs per cabinet and can scale up to 8,192 NPUs, delivering superior performance for large-scale AI

Huawei''s CloudMatrix 384: The new AI cluster that challenges Nvidia

© 2026 BAGYOU FIBRE NETWORKS - All rights reserved

Huawei unveils CloudMatrix 384, its most powerful AI cluster that competes directly with Nvidia in performance.

Huawei delivers advanced AI chip ''cluster'' to Chinese clients cut off ...

Huawei has started the delivery of its advanced artificial intelligence chip “cluster” to Chinese clients who are increasing orders after being cut off from Nvidia''s semiconductors because of...

Huawei Connect 2025: Building AI Infrastructure In A

Ascend AI NPUs: The Ascend 910C powers Huawei''s CloudMatrix 384 “supernode” cluster, rivaling NVIDIA''s GB200 NVL72 in compute

Huawei''s new computing cluster, world''s first AI agent phone to debut ...

The coming World Artificial Intelligence Conference (WAIC) is set to feature major new product releases, such as Huawei Technologies'' next-generation computing cluster, as China

Huawei outlines roadmap for Ascend AI chips

Huawei has said the new chips will strengthen the foundation of AI computing power, both in China and around the world In sum – what to know:

Key products in Huawei''s AI chips and computing power roadmap

China''s Huawei (HWT.UL) has ended years of secrecy to detail its product roadmap for chips and computing power systems, discussing for the first time how it intends to challenge global

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.

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

Ascend clusters apply that mindset to AI, with the emphasis on predictable scaling behavior and integration with Huawei''s own AI frameworks rather than leading benchmarks.

Huawei Unveils AI Agent Phone, Computing Cluster at WAIC

Huawei Technologies is set to unveil its next-generation computing cluster and what has been described as the world''s first artificial intelligence agent phone. This will occur at the upcoming World Artificial

Huawei Unveils Atlas 950 AI SuperPoD to Compete with Nvidia and

© 2026 BAGYOU FIBRE NETWORKS - All rights reserved

At MWC 2026, Huawei unveiled its Atlas 950 and TaiShan 950 SuperPoDs to a global audience for the first time, expanding its largest AI computing clusters beyond the Chinese market.

Contact Us

Continue your research online:

Website: https://bagyou.es
Technical inquiry: https://bagyou.es/project-enquiry.html

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

© 2026 BAGYOU FIBRE NETWORKS - All rights reserved