

The most authentic Huawei AI server model



Overview

CloudMatrix 384 is powered by 384 of Huawei's most sophisticated AI chips, the Ascend 910C graphics processing unit. That compares with just 72 B200 GPUs found in Nvidia's GB200 NVL72 system. 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. This supernode is the industry's first to implement peer-to-peer interconnection of 384 proprietary NPUs and 192 Kunpeng CPUs through a high-speed MatrixLink network to form a super AI server, which increases the inference throughput of a single card to 2,300 tokens/s, a near 4-fold improvement. Scaled-out MoE LLMs and scaled-up SuperPods create new systems challenges for production Model-as-a-Service (MaaS), requiring disaggregation, low-latency communication, and decentralized serving. This report presents xDeepServe, the production serving system behind Huawei Cloud's MaaS offering on. The Atlas 500 Pro (model 3000) is a 2 U AI edge server powered by Huawei Kunpeng 920 processors, featuring superb computing performance, strong environmental adaptability, easy deployment and maintenance, and cloud-edge collaboration. The AI system employs a "supernode" architecture with high-speed internal chip interconnects. 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.

Article Content

May 10, 2026

Weekly news roundup: Huawei found to use outdated

Google has added a new 270-million-parameter model to its Gemma 3 series of open-weight AI models, targeting lower power consumption for mobile

Jul 09, 2025

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 supports AI accelerator cards and provides powerful real-time inference and AI training

Sep 30, 2025

Huawei launches CloudMatrix 384 server as an

China's Huawei Technologies Co. Ltd. threw down the gauntlet to Nvidia Corp. Saturday when it revealed its most powerful artificial intelligence

Jun 22, 2026

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

Nov 06, 2025

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 directly

Mar 29, 2026

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 million

Sep 08, 2025

Huawei Cloud Model-as-a-Service on the CloudMatrix384 SuperPod

This report presents xDeepServe, the production serving system behind Huawei Cloud's MaaS offering on CloudMatrix384, a 48-server SuperPod with 384 Ascend 910C chips connected by a high

Jul 28, 2025

Huawei pushes back on AI model plagiarism claims

Huawei has publicly denied reports that its Pangu Pro MoE open-source model is a "recycled product" based on work from Alibaba.

Dec 18, 2025

Atlas 800 Training Server User Guide (Model 9010) 17

The Atlas 800 training server (model 9010) is an AI training server running on the Huawei Ascend 910 AI Processor and Intel Cascade Lake processor. It is widely used in deep learning model development

Dec 06, 2025

ModelArts AI Development Platform | Huawei Cloud

Huawei Cloud ModelArts is a one-stop AI development platform that enables developers and data scientists of any skill level to rapidly build models.

Jan 23, 2026

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

Oct 01, 2025

Exclusive: Big Chinese tech firms scramble to secure Huawei AI chips ...

Demand for Huawei's Ascend 950 AI chips has surged following the release of DeepSeek's V4 artificial intelligence model that runs on the Shenzhen-based tech firm's chips, with

Mar 31, 2026

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 performance, strong environmental adaptability, easy deployment and

Jul 15, 2025

Huawei vs. Alibaba: The AI Model Controversy and Its Implications

Explore the Huawei vs. Alibaba AI feud over model originality. Learn the facts & implications—read the full blog to stay informed!

Sep 30, 2025

HUAWEI MatePad Pro Max

HUAWEI MatePad Pro Max, features a 13.2-inch flexible OLED PaperMatte Display, an ultra-light and thin body, comes equipped with PC-level WPS Office, and supports HUAWEI Glide

Feb 13, 2026

Ascend AI Cloud Service

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

Jul 18, 2025

Huawei's AI CloudMatrix 384 is big, expensive and

Built for trillion-parameter AI models, it delivers real-time inference at speeds up to 30 times faster than previous systems by avoiding traditional data

Oct 20, 2025

Huawei denies copying Alibaba's AI model

The controversy between Huawei and Alibaba coincides with landmark developments in Chinese copyright law addressing AI-generated content, creating new precedents for how model

Mar 31, 2026

Huawei showcases CloudMatrix 384 AI system to rival

Initially announced in April, CloudMatrix 384 has been closely watched by industry observers, with some analysts positioning it as Huawei's answer to

Jul 13, 2025

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

Jun 28, 2025

Atlas 800 Inference Server User Guide (Model 3000) 17

About This Document Purpose This document describes the appearance, structure, components, and specifications of the Atlas 800 inference server (model 3000), and provides guidance for installing,

Apr 29, 2026

Huawei Launches FusionServer Pro V6 Intelligent

Today, the Huawei Computing Product Line officially released the latest member to their family - the next-generation FusionServer Pro V6

Jan 17, 2026

Prices of Nvidia's B300 server at 7 million yuan in ...

Strong demand for artificial intelligence computing equipment in China has nearly doubled prices for Nvidia's B300 servers to about seven million yuan (\$1.3 million) each,

Apr 29, 2026

Huawei Cloud Announces Pangu Models 5.5 and All

Today, the Huawei Cloud AI Cloud Service has emerged as the preferred choice for AI infrastructure. It provides robust AI compute for more than 1,300 customers,

Jun 01, 2026

Huawei Denies Copying Alibaba AI Model Amid GitHub

Updated 8 July 2025 at 17:52 IST Huawei Denies Copying Alibaba AI Model Amid GitHub Controversy Huawei argues that its model was independently trained and

Feb 28, 2026

DeepSeek Unveils V4 at Rock-Bottom Prices With Full Support From Huawei ...

DeepSeek unveiled its V4 model on April 24, 2026—1.6 trillion parameters, \$3.48 per million tokens, open-source, and trained on Huawei Ascend chips. Here's what it means for the US

Contact Us

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

Website: <https://elagage-lorrain.fr>

Email: sales@elagage-lorrain.fr

Phone: +33 6 47 82 19 35

Address: 15 Rue de la République, 69002 Lyon, France

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

