

Huawei AI Server Series



AI Overview

Huawei's AI server series includes the Atlas and TaiShan lines, powered by Ascend NPUs, designed for large-scale AI training and inference with SuperPoD architecture.

Overview of Huawei AI Servers

Huawei's AI servers are built around Ascend NPUs (Neural Processing Units), which are purpose-designed for AI workloads rather than adapted from graphics processing. The main server series include:

- Atlas 950 SuperPoD: Integrates up to 8,192 Ascend 950DT NPUs per SuperPoD, delivering up to 16 EFLOPS FP16 and 8 EFLOPS FP8 performance. It uses Huawei's UnifiedBus interconnect to operate thousands of NPUs as a single logical machine, enabling large-scale AI training and high-concurrency inference .
- Atlas 960 SuperPoD: A next-generation cluster designed to scale even further, supporting up to 500,000-1 million Ascend NPUs in a SuperCluster configuration, targeting the largest AI compute workloads globally .
- TaiShan 950 SuperPoD: Focused on general-purpose AI computing with low-latency memory pooling and high bandwidth, suitable for both training and inference in carrier and enterprise environments .
- TaiShan 500 and TaiShan 200 series: Cover medium and low computing power requirements, providing flexible deployment options for diverse AI workloads .
- Atlas 850E: Supports air-cooled data centers, scaling from 8 to 1,024 NPUs, ideal for smaller-scale inference clusters .

Ascend NPU Roadmap

Article Content

Huawei FusionServer Pro Intelligent Server Is the Perfect Partner for ...

Huawei and Intel jointly launched the next-generation FusionServer Pro 2488H V6 intelligent server. The FusionServer Pro 2488H V6

Atlas 800 AI Server-Ascend Community

The Atlas 800 inference server (Model 3000) supports up to eight Atlas 300I/V Pro cards, delivering powerful real-time inference and

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,

A Holistic View at Huawei's Full-Stack, All-Scenario AI Strategy

Huawei has also developed dozens of partners for the Atlas series cards and servers and implemented AI solutions in

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

Huawei Launched Intelligent Servers at HUAWEI CONNECT 2018

[Shanghai, China, October 19, 2018] Huawei launched its Intelligent Server series at HUAWEI CONNECT 2018. The servers offer

Groundbreaking SuperPoD Interconnect: Leading a New Paradigm

With the Atlas 950 SuperPoD and Atlas 960 SuperPoD, we are confident in our ability to provide abundant computing

Overview

The Atlas 800 training server (model 9010) is an AI training server running on the Huawei Ascend 910 AI Processor and Intel

OceanStor AI Storage

Huawei AI Inference Acceleration Solution is built on OceanStor A series storage and comes equipped with Unified Cache Manager

Huawei AI Servers Test Effectiveness of US Curbs

The recent Huawei unveiling in Anhui province is one example of China's continued push. The new server cluster

AI-accelerated ECSs_x86 ECS Specifications_ECS Types and ...

AI-accelerated ECSs, powered by Ascend processors and software stacks, are dedicated to accelerating AI

Huawei set to dominate China's AI server market, phasing out Nvidia

A new report states that Huawei and Cambricon, the leading chip suppliers in China's AI segment, will rule over the

Ascend AI Cloud Service | Huawei Cloud

Ascend-native, cloud-based AI agents, such as search agents and big data agents, allow users to integrate the capabilities of AI

Huawei unveils new AI-centric products, cloud services

At its annual Cloud Connect conference in Shanghai last week, Huawei put the spotlight squarely on artificial

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

Server Huawei: AI & Cloud Trends 2026

AI Training and Inference Servers (Ascend Series) Snapshot: High-performance computing nodes equipped with

Huawei Enterprise Servers

Huawei RH Series Rack Servers Check out Huawei's next-generation FusionServer V3 and versatile Tecal V2 servers to meet

Buy Huawei Rack Servers – G5500, 5885H, 5288H, 2488H, 2288H,

FusionServer AI Server Huawei's FusionServer series includes a wide range of rack servers—from high-density AI platforms to

Groundbreaking SuperPoD Interconnect: Leading a New Paradigm ...

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

Huawei Launches FusionServer Pro V6 Intelligent Server Based on

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

Huawei FusionServer Pro Intelligent Servers Fully Support Intel

The Huawei FusionServer Pro series provides a wide range of intelligent servers, including 2-, 4-, and 8-socket rack

Huawei launches new AI server solution – Trivium China

What happened: On April 10, Huawei launched its new AI server system, CloudMatrix 384. Huawei claims it

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

Purpose This document describes the Atlas 500 Pro AI edge server (model 3000), Atlas 500 Pro (model 3000) for short, in terms of

Huawei Unveiled the Latest SuperPoD, Making an AI Infrastructure

Huawei's latest computing portfolio, including Atlas 950 SuperPoD, Atlas 850E, TaiShan 950 SuperPoD, and TaiShan

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

Contact Us

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

Website: <https://www.roof-repairs-johannesburg.co.za>

Email: info@roof-repairs-johannesburg.co.za

Phone: +27 63 592 4187

Address: 2nd Floor, The Leonardo, 75 Maude Street, Sandton, Johannesburg, 2196, South Africa

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

