ZITHA TECH SPECIALTY CABLES (Pty) Ltd

Huawei AI Intelligent Server



Overview

Huawei Cloud Announces Agentic AI Products, Shaping the Moving forward, Huawei Cloud will drive software-hardware synergistic innovation to build the found.

Key Takeaways

Huawei's Intelligent AI Server, exemplified by the CloudMatrix 384, delivers high-performance AI compute with 384 Ascend 910C processors, achieving up to 300 petaflops for large-scale model training and inference.Overview

Huawei's Intelligent AI Server is designed to meet the growing demands of AI workloads, including foundation model training, inference, and enterprise agentic AI applications. The flagship system, CloudMatrix 384, integrates 384 Ascend 910C AI processors into a single cluster, forming a “supernode” architecture that enables massive parallel processing and high-speed interconnectivity between chips . This design allows Huawei to scale AI compute power beyond individual GPU limitations, providing a domestic alternative to Nvidia's high-end AI servers .

Performance and Architecture
• Compute Power: CloudMatrix 384 delivers approximately 300 petaflops, surpassing comparable Nvidia systems in aggregate performance .

• Network and Interconnect: The system uses high-speed network interconnection technology to ensure low-latency communication between processors, supporting large-scale distributed AI workloads .

• Energy Consumption: The server consumes 559 kW per hour, reflecting the high energy demands of ultra-large AI clusters .

© 2026 ZITHA TECH SPECIALTY CABLES (Pty) Ltd - All rights reserved

• Cooling and Efficiency: Huawei employs fully liquid-cooled AI data centers capable of handling up to 80 kW heat dissipation per cabinet, achieving a power usage effectiveness (PUE) of 1.1 .

Software and AI Integration

Huawei's AI servers are tightly integrated with Huawei Cloud's Agentic AI ecosystem, including:

• AI Cluster Service (AICS): Supports clusters with over 100,000 cards and up to 200 EFLOPS, with token generation latency under 10 milliseconds and throughput of 5 million tokens per second .

• ModelArts Next: Provides model training and inference with Reinforcement Learning as a Service (RLaaS), confidential inference, and dynamic model routing, optimizing performance and cost .

• AgentArts Platform: Enterprise-grade agent platform for long-running tasks, security, and industry-specific AI deployment, with an open-source edition called openJiuwen .

• Elastic Memory Service: Expands video RAM to reduce latency in multi-round AI model interactions .

Applications

Huawei Intelligent AI Servers are deployed across multiple industries, including:

• Healthcare: Smart diagnostics and AI-assisted treatment planning.

• Manufacturing: Predictive maintenance and process optimization.

• Scientific Computing: Large-scale simulations and research computing.

• Content Generation: AI-generated content (AIGC) and natural language processing tasks .

Key Advantages

• Scalability: Supports ultra-large AI clusters for foundation model training.

• Hardware-Software Synergy: Optimized integration between Ascend processors and Huawei Cloud AI platforms.

• Enterprise-Ready: Security, observability, and long-running task management for industrial deployment.

• Domestic AI Independence: Provides a high-performance alternative to restricted foreign GPU solutions . Huawei's Intelligent AI Server represents a comprehensive AI compute solution, combining cutting-edge hardware, advanced software platforms, and industry-specific applications to accelerate AI innovation and enterprise adoption.

© 2026 ZITHA TECH SPECIALTY CABLES (Pty) Ltd - All rights reserved

Article Content

Huawei Cloud Announces Agentic AI Products, Shaping the

Moving forward, Huawei Cloud will drive software-hardware synergistic innovation to build the foundation for enterprise

Huawei Unveils End-to-End Computing Platform

By providing an end-to-end, one-stop computing solution, Huawei empowers customers to

Huawei launches CloudMatrix 384 AI System to rival

The Huawei CloudMatrix 384 is a high-density AI computing system featuring 384 Huawei

OceanStor AI Storage

In response to the demands of large AI model training and inference, Huawei redefines distributed file

Huawei Launches Its AI Data Platform to Power Faster AI Adoption for ...

Looking ahead, Huawei will strengthen its investment in AI data infrastructure, empower industry transformation through ongoing

Huawei AI Data Center Reference Design

Intelligent iPower improves system reliability and simplifies operation and maintenance for customers. The S-ECO(Super ECO)

Huawei Cloud

Leverage cutting-edge technologies such as cloud computing, big data, AI, and 5G to empower digital transformation and AI-driven

Huawei and Cambricon lead China''s AI server chip surge, piling

Huawei and Cambricon are set to dominate China''s AI server market as geopolitics and self-reliance accelerate

AIDC Facility Solutions for High-Quality Intelligent Computing

Huawei Digital Power provides comprehensive artificial intelligence data center facility solutions designed for high-quality intelligent

Ascend AI Cloud Service | Huawei Cloud

© 2026 ZITHA TECH SPECIALTY CABLES (Pty) Ltd - All rights reserved

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

OceanStor AI Storage

The Next-Gen High-Performance Distributed File Storage is designed for AI, providing bandwidth in

Huawei Releases AI Data Lake Solution, Positioned to Accelerate ...

Peter Zhou, Vice President of Huawei and President of Huawei Data Storage Product Line, introduced the solution in

Huawei Launches FusionServer Pro Intelligent Servers

At today''s +Intelligence, Taking Computing to New Levels press conference, Huawei

Data Center Network

Huawei''s Xinghe Intelligent Ultra-Efficient Data Center Network Solution delivers ultra-large computing power, high efficiency, and

EXCLUSIVE: Huawei aims to mass-produce newest AI

China''s Huawei plans to start mass-producing its most advanced artificial intelligence chip

Huawei Ascend 950PR: The 1.56 PFLOP AI Chip vs

The Huawei Ascend 950PR packs 1.56 petaflops of FP4 compute and 112GB HBM,

Data Center Portfolio Solution | Huawei Enterprise

Huawei has launched a series of data center portfolio and individual solutions to underpin

Huawei Unveils Fully Upgraded AI WAN Solution with

Wang emphasized that Huawei looks forward to collaborating with more leading global

Huawei to announce new Ascend AI chip solutions at WAIC 2026

Huawei will once again introduce some dynamic Ascend AI chip solutions at the 2026 WAIC (World Artificial

Elastic Cloud Server (ECS)

Elastic Cloud Server (ECS) provides secure, scalable, on-demand compute resources for you to flexibly deploy applications and

© 2026 ZITHA TECH SPECIALTY CABLES (Pty) Ltd - All rights reserved

Huawei AI Data Center Reference Design

Preface This manual systematically presents Huawei''s data center facility solutions for AI intelligent computing scenarios. We

Huawei and Intel Jointly Launch the Next-Gen FusionServer Pro V6 ...

The intelligent server is equipped with the latest AI training and inference modules to unleash up to 560 TFLOPS

Video Intelligent Analysis Service

Video Intelligent Analysis Service (VIAS) analyzes camera video data and reports alarm events. By

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

Supercharge Your Intelligent Computing Center with AI-Ready Data ...

If you don''t want your intelligent computing center to be burdened with data silos or performance and capacity issues,

Huawei and Intel Jointly Launch the Next-Gen FusionServer Pro V6 ...

The intelligent server is equipped with the latest AI training and inference modules to unleash up to 560 TFLOPS

Huawei Cloud INSPIRE 2026

Huawei Cloud also announced four dedicated zones: the Smart Healthcare Zone, Embodied AI Zone, Smart Manufacturing Zone,

Huawei Atlas AI Computing Platform: A Gift for Developers

The Atlas AI computing platform provides powerful AI computing for customers to handle massive data volumes. It

Huawei could seize China''s AI chip crown in 2026 as Nvidia''s H200 ...

Tech Industry Artificial Intelligence Huawei could seize China''s AI chip crown in 2026 as Nvidia''s H200 shipments stall

Huawei Launches Xinghe AI Network Security Agentic SOC to Build

Huawei''s newly launched Xinghe AI Network Security Agentic SOC integrates three intelligent agents (Sensing,

OceanStor AI Storage

© 2026 ZITHA TECH SPECIALTY CABLES (Pty) Ltd - All rights reserved

The Next-Gen High-Performance Distributed File Storage is designed for AI, providing bandwidth in

Huawei Releases AI Data Lake Solution, Positioned to Accelerate ...

Peter Zhou, Vice President of Huawei and President of Huawei Data Storage Product Line, introduced the solution in

Huawei Launches FusionServer Pro Intelligent Servers

At today''s +Intelligence, Taking Computing to New Levels press conference, Huawei

Data Center Network

Huawei''s Xinghe Intelligent Ultra-Efficient Data Center Network Solution delivers ultra-large computing power, high efficiency, and

EXCLUSIVE: Huawei aims to mass-produce newest AI

China''s Huawei plans to start mass-producing its most advanced artificial intelligence chip

Huawei Ascend 950PR: The 1.56 PFLOP AI Chip vs

The Huawei Ascend 950PR packs 1.56 petaflops of FP4 compute and 112GB HBM,

Huawei Cloud Announces Agentic AI Products, Shaping the

Moving forward, Huawei Cloud will drive software-hardware synergistic innovation to build the foundation for enterprise

Huawei Unveils End-to-End Computing Platform

By providing an end-to-end, one-stop computing solution, Huawei empowers customers to

Huawei launches CloudMatrix 384 AI System to rival

The Huawei CloudMatrix 384 is a high-density AI computing system featuring 384 Huawei

OceanStor AI Storage

In response to the demands of large AI model training and inference, Huawei redefines distributed file

Huawei Launches Its AI Data Platform to Power Faster AI Adoption for ...

Looking ahead, Huawei will strengthen its investment in AI data infrastructure, empower industry transformation through ongoing

© 2026 ZITHA TECH SPECIALTY CABLES (Pty) Ltd - All rights reserved

Huawei AI Data Center Reference Design

Intelligent iPower improves system reliability and simplifies operation and maintenance for customers. The S-ECO(Super ECO)

Huawei Cloud

Leverage cutting-edge technologies such as cloud computing, big data, AI, and 5G to empower digital transformation and AI-driven

Huawei Cloud Announces Agentic AI Products, Shaping the

Moving forward, Huawei Cloud will drive software-hardware synergistic innovation to build the foundation for enterprise

Huawei Unveils End-to-End Computing Platform

By providing an end-to-end, one-stop computing solution, Huawei empowers customers to

Huawei launches CloudMatrix 384 AI System to rival

The Huawei CloudMatrix 384 is a high-density AI computing system featuring 384 Huawei

OceanStor AI Storage

In response to the demands of large AI model training and inference, Huawei redefines distributed file

Huawei Launches Its AI Data Platform to Power Faster AI Adoption for ...

Looking ahead, Huawei will strengthen its investment in AI data infrastructure, empower industry transformation through ongoing

Huawei AI Data Center Reference Design

Intelligent iPower improves system reliability and simplifies operation and maintenance for customers. The S-ECO(Super ECO)

Huawei Cloud

Leverage cutting-edge technologies such as cloud computing, big data, AI, and 5G to empower digital transformation and AI-driven

Related Video Reference

This video was associated with the source search result. Verify technical details against current product documentation and project requirements.

© 2026 ZITHA TECH SPECIALTY CABLES (Pty) Ltd - All rights reserved

Contact Us

Continue your research online:

Website: https://zithatech.co.za
Technical inquiry: https://zithatech.co.za/contact.html

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

© 2026 ZITHA TECH SPECIALTY CABLES (Pty) Ltd - All rights reserved