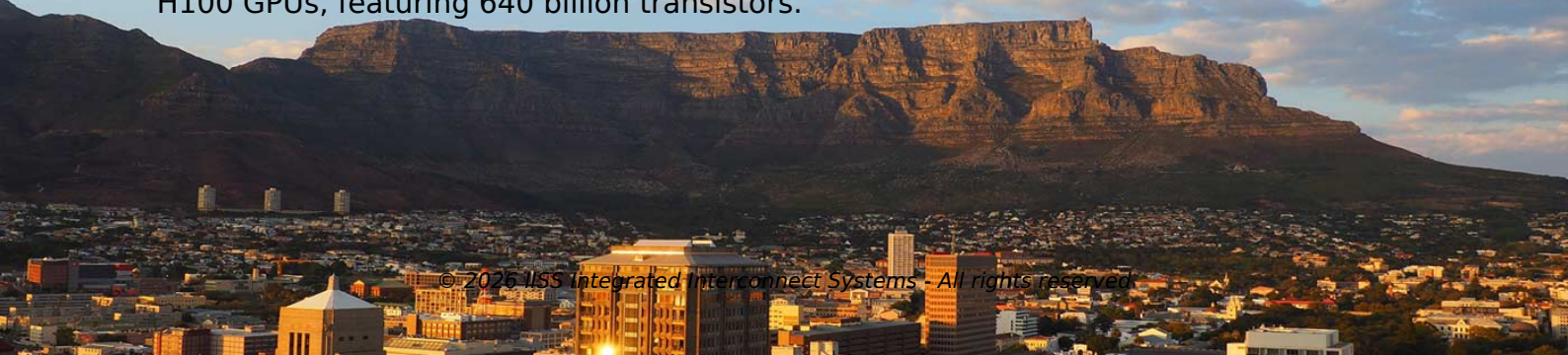


What chips are used in domestic AI servers

Overview

Detailing and benchmarking hardware architectures for current market players and emerging AI chips for server CPUs, custom AI ASICs, and other AI chips, including AI accelerators using heterogeneous matrix-based systems and spatial AI accelerators. China's domestic AI chips took 41% of the accelerator server market in 2025. New data shows Huawei alone shipped roughly 812,000 AI chip units last. These include GPUs, custom AI application-specific integrated circuits (ASICs) used by hyperscalers and cloud service providers (CSPs), AI-capable central processing units (CPUs), and other AI ASICs developed by both AI chip-focused startups and large vendors.

Graphics Processing Units (GPUs) The. In October 2023, the US government imposed new export controls that effectively limited the computing performance of advanced chips exported to China. While intended to hinder China's technological progress, these restrictions may have had an unintended consequence – fueling even greater demand for. DeepSeek's latest models highlight China's growing AI ecosystem built on domestic hardware (AI generated). Summary DeepSeek unveils its V4 series, focusing on large-scale mixture-of-experts (MoE) architectures to improve efficiency and scalability. The models are optimized for Huawei Ascend AI. Huawei has officially launched the Ascend 950PR, a 1.8 times the FP4 performance of Nvidia's H20 — marking the most aggressive challenge yet to American semiconductor dominance from a Chinese chipmaker operating under heavy US sanctions. With 750,000. The DGX H100, launched in 2022, is NVIDIA's latest DGX system iteration and a foundation for DGX SuperPOD deployments. A DGX server is equipped with eight H100 GPUs, featuring 640 billion transistors.



Article Content

Sep 28, 2025

Huawei Ascend 950PR: The 1.56 PFLOP AI Chip vs Nvidia

Huawei has officially launched the Ascend 950PR, a 1.56-petaflop AI inference chip that delivers 2.8 times the FP4 performance of Nvidia's H20 — marking the most aggressive challenge ...

Aug 17, 2025

Huawei launches Ascend AI chips, Kunpeng server chips

The announcement, breaking years of secrecy around its chip operations, outlined timelines for its Ascend artificial intelligence chips and Kunpeng server processors, potentially raising ...

Mar 08, 2026

Daily: Alibaba and Baidu use domestic chips for training AI models

Alibaba and Baidu use domestic chips. In this vein, Chinese tech giants Alibaba and Baidu both report that they are using domestic chips to train their AI models. This is in line with ...

Mar 06, 2026

Nvidia market share in China falls to less than 60% — Chinese chip ...

Tech Industry Nvidia market share in China falls to less than 60% — Chinese chip makers deliver 1.65 million AI GPUs as the government pushes data centers to use domestic chips

Jun 28, 2026

Domestic Chips Break into China Server Ecosystem with 820000 ...

Currently, GPUs are still dominating the market, accounting for 69% of the market. However, the market share of non-GPU-accelerated servers, such as ASICs and FPGAs, has ...

Jun 22, 2026

AI Chips for Data Centers and Cloud 2025-2035: Technologies, ...

Graphics processing units (GPUs) and other AI chips have been instrumental in driving this growth of artificial intelligence, providing the compute needed for deep learning within data centers and cloud ...

Dec 31, 2025

AI Chips and Servers in China: Surging Demand Amid US Restrictions

In addition to GPUs and ASICs, more networking chip units like retimers, switches, and smart NICs are also needed to support the increased number of AI accelerators per system.

Aug 06, 2025

Deepseek's v4 breakthrough: chinese ai pushes efficiency and ...

DeepSeek unveils its V4 AI model series optimized for Huawei Ascend chips. Explore how China is advancing efficient, large-scale AI systems.

Oct 22, 2025

AI Chips and Servers in China: Surging Demand Amid ...

In addition to GPUs and ASICs, more networking chip units like retimers, switches, and smart NICs are also needed to support the increased ...

Jan 21, 2026

Memory Interface Chips for AI Servers

In AI servers, a Retimer chip is typically required between CPU and GPU to ensure signal integrity. Many AI servers deploy multiple Retimers; for example, Astera Labs configures four ...

Mar 20, 2026

Huawei Chips Claim 41% of China's AI Server

Huawei Chips Now Claim 41% of China's AI Server Market as NVIDIA's Grip Loosens
China's AI hardware landscape shifted dramatically in 2025, with domestic chip makers claiming ...

Contact Us

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

Website: <https://www.iiss.co.za>

Email: sales@iiss.co.za

Phone: +27 63 174 2895

Address: Unit 8, Innovation Park, 15 Rivonia Road, Sandton, Johannesburg,
2196, South Africa

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

