Atlas AI Computing Platform

The Huawei Atlas intelligent computing platform is powered by the Huawei Ascend series AI processors and mainstream heterogeneous computing components. It provides a rich array of product form factors, such as modules, cards, edge stations, and appliance, enabling all-scenario AI infrastructure solutions for Device-Edge-Cloud. It is widely applied to domains such as safe city, smart transportation, smart healthcare, and AI inference.

Superior Compute Power

Powered by the Huawei Ascend series AI processors, a single chip delivers 16 teraOPS (TOPS) of INT8 and supports real-time analytics of 16-channel HD videos with a power consumption less than 8 W.

All-Scenario AI

The all-scenario AI infrastructure solution is designed and tuned for Device-Edge-Cloud, fully meeting diversified AI application scenario requirements in the intelligent era.

Open Ecosystem

The Atlas intelligent computing platform supports mainstream frameworks, and provides easy-to-use code porting and model conversion tools, allowing for flexible collaboration with industry ISVs to build an open industry ecosystem.

Atlas 200 AI Accelerator Module

- 16 TOPS INT8 @ 9.5W
- Supports real-time analytics of 16-channel HD videos, and JPEG encoding/decoding
- 4 GB/8 GB memory | PCIe 3.0 x4 interface
- Operating temperature: -25°C to +80°C
- Dimensions: 52 mm x 38 mm x 8.5 mm

Cameras | Drones | Robots

Large Capacity in a Compact Size

- Powered by the Huawei Ascend AI processor, it supports real-time analytics of 16-channel HD videos in the size of half a credit card

Built for Intelligent Devices and Edge

- Positioned for terminal devices such as cameras and drones, and edge devices such as AI edge stations, it has a thermal design power (TDP) of only 9.5W

Atlas 200 DK AI Developer Kit

- 16 TOPS INT8 @ 24 W
- 1 USB type-C | 2 CCM interfaces | 1 GE network port | 1 SD card slot
- 8 GB memory
- Operating temperature: 0°C to 45°C
- Dimensions: 125 mm x 80 mm x 24 mm

Quickly build development environment in 30 minutes

High Integration

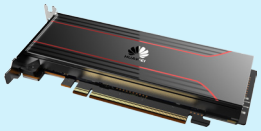
- Inspired by the Huawei Ascend AI processor, it integrates rich peripheral interfaces and the MindStudio software to help developers quickly familiarize with the development environment

Easy-to-Use Software Environment

- The MindStudio provides a user-friendly programming interface and graphics-based debugging capability, allowing automatic management of offline models with a simulation environment

Atlas 300 AI Accelerator Card

- 64 TOPS INT8 / 32 TFLOPS FP16 @ 67 W
- Supports real-time analytics of 64-channel HD videos, and JPEG encoding/decoding
- 32 GB memory, 204.8 GB/s memory bandwidth, with support for ECC
- PCIe 3.0 x16, half-height half-length card



64 channels per card
Best choice for high-density video inference

Superior Compute Power

- Provides 64 TOPS of INT8 and 64-channel HD video real-time analytics on a single card, fueling deep learning and inference with superior compute power

Hardware Encoding/Decoding

- Supports JPEG and video hardware encoding/decoding, delivering a leap in image and video application performance
- Large memory capacity and high bandwidth, meeting memory requirements in feature matching scenarios and reducing application latency

Atlas AI Computing Platform

Atlas 500 AI Edge Station

Safe City | Smart Transportation | Unattended Retail

Intelligent Edge
- An industry-leading edge product that integrates AI processing capabilities
- Works stably at -40°C to +70°C outdoors without fans

Large Capacity in a Compact Size
- Capable of processing 16-channel HD videos in the size of an STB
- Delivers a 4x performance lead over competing products in the industry

Edge-Cloud Collaboration
- Works with Huawei private cloud and HUAWEI CLOUD, receiving applications and updated algorithms pushed from the cloud
- Unified device management and firmware upgrade on the cloud

Atlas 600 AI Inference Appliance

Safe City | Smart Transportation | Smart Finance

Rich Scenarios
- Out-of-the-box, rapid deployment in various scenarios, including facial recognition, vehicle identification, target structurization, and OCR, highly versatile for industries such as safe city, smart transportation, and smart finance

Flexible Configuration
- Provides deeply tuned hardware for customers to rapidly deploy software, enabling optimal hardware and software combinations with the best economics for customers' scenarios

Atlas 800 Deep Learning System

Deep Learning | Model Training | Recommendation System

Out-of-the-Box
- Ready to work in 2h with pre-installed AI development environment, underlying software library, and development framework
- Automatic generation of AutoDL models, hyperparameter tuning, and one-click model deployment

Ultimate Performance
- Provides an optimized AI environment based on the standard framework and programming environment
- Unlocks ultimate performance with GPU scheduling algorithms, with over 15% resource utilization

Integrated Management
- Comprehensive management of GPU utilization, health monitoring, and job scheduling
- Easy-to-use WebUI, more intuitive and efficient than the command line interface (CLI)

Atlas G5500 Heterogeneous Server

AI | HPC | Cloud GPU Acceleration

Outstanding Heterogeneous Computing Performance
- Supports 8 FHFL training or 32 HHHL inference accelerator cards
- GPUDirect RDMA, P2P, and NVLink interconnects

One-Click Switching of Heterogeneous Topologies
- One-click topology switching for AI and HPC
- Supports diversified applications and reduces hardware investment

Fully Modular Architecture
- Decoupled design for CPU and heterogeneous resources, supporting long-term technology evolution
- Efficient heat dissipation design, industry-unique to run stably at 35°C with full specifications

Atlas G2500 Intelligent Video Analytics Server

Safe City | Facial Recognition | Smart Transportation

Super Intelligence
- Supports up to 16 HHHL inference cards in a 4U chassis
- Provides 256-channel intelligent video analytics of people/vehicles/things

Ultra-Large Storage
- Supports 24 x 3.5'' hard drives, up to 240 TB
- Enables massive data storage and real-time information retrieval

Edge Deployment Capability
- Supports 55°C working temperature
- 675 mm deep chassis, supporting installation in short cabinets

For more information, please visit
https://e.huawei.com/en/solutions/business-needs/data-center/atlas