

Atlas AI Computing Solution



CONTENTS

Atlas Intelligent Computing Solutions	01
Atlas 200 Al Accelerator Module	02
Atlas 200DK AI Developer Kit	03
Atlas 300I Inferance Card	04
Atlas 300T Training Card	05
Atlas 500 AI Edge Station	06
Atlas 800 inference Server	07
Atlas800 Training Server	08
Huawei Intelligent Computing Portfolio	09

Atlas AI Computing solutions





Atlas 200 Al Accelerator Module

Model: 3000



The Atlas 200 Al accelerator module (model: 3000) integrates the Ascend 310 Al processor to implement facial recognition and image classification on the device side. Atlas 200 is widely used in Al scenarios such as intelligent cameras, robots, and drones.

Specifications

AI Processor Ascend 310 22 TOPS INT8 AI Computing 16 TOPS INT8 Power 8 TOPS INT8 LPDDR4X, 8 GB/4 GB, total bandwidth 51.2 Memory GB/s H.264 hardware decoding, 16-channel 1080p 30 FPS (2-channel 3840 x 2160 @ 60 FPS) H.265 hardware decoding, 16-channel 1080p 30 FPS (2-channel 3840 x 2160 @ 60 FPS) H.264 hardware encoding, 1-channel Encoding/ 1080p 30 FPS Decoding H.265 hardware encoding, 1-channel 1080p 30 FPS JPEG decoding: 1080p 256 FPS; encoding: 1080p 64 FPS; maximum resolution: 8192 x 4320 PNG decoding: 1080p 24 FPS; maximum resolution: 4096 x 2160 PCIe x4 Gen3.0 Port 1 USB 2.0/USB 3.0 1 RGMII Serial Bus UART/I2C/SPI Interface 144-pin BTB connector Specifications **Typical Power** 4 GB: 5.5 W Consumption 8 GB: 8 W Operating -25°C to +80°C Temperature Weight 30 g Dimensions 8.5 mm x 52.6 mm x 38.5 mm $(H \times W \times D)$

Ultimate performance

- 22 TOPS INT8 in the size of half a credit card, supporting real-time analysis of 20-channel HD videos (1080p 25FPS)
- Multi-level computing power configuration: 22/16/8 TOPS

Ultra-low consumption

 Hibernation at milliwatts and wakeup in milliseconds, typical power consumption of 5.5 W, enabling edge AI applications

Application Scenarios

Embedded in edge intelligence



Building a Fully Connected, Intelligent World

Atlas 200 DK AI **Developer Kit**

Model: 3000



The Atlas 200 DK (model: 3000) is a highperformance AI application developer board that integrates the Ascend 310 AI processor to facilitate quick development and verification. It has been widely used in scenarios such as developer solution verification, higher education, and scientific research.

High integration

Powered by the Huawei Ascend 310 AI processor, and integrates various peripheral interfaces and the Mind Studio, facilitating access to the development environment and enabling quick development

Easy-to-use software environment

Mind Studio provides a user-friendly programming interface and GUI-based debugging, allowing automatic management of offline models with a simulation environment

Application Scenarios



Developer solution verification

Model verification Solution verification



Higher education

Entry-level AI education Talent cultivation



Scientific research

Application research Algorithm research

Specifications

AI Processor	Ascend 310		
AI Computing Power	22 TOPS INT8 16 TOPS INT8 8 TOPS INT8		
Memory	LPDDR4X, 8 GB, total bandwidth 51.2 GB/s		
Encoding/ Decoding	 H.264 hardware decoding, 16-channel 1080p 30 FPS (2-channel 3840 x 2160 @ 60 FPS) H.265 hardware decoding, 16-channel 1080p 30 FPS (2-channel 3840 x 2160 @ 60 FPS) H.264 hardware encoding, 1-channel 1080p 30 FPS H.265 hardware encoding, 1-channel 1080p 30 FPS JPEG decoding: 1080p 256 FPS; encoding: 1080p 64 FPS; maximum resolution: 8192 x 4320 PNG decoding: 1080p 24 FPS; maximum resolution: 4096 x 2160 		
Port	 Network: 1 GE RJ45 port USB: 1 USB 2.0/USB 3.0 port Camera: two 15-pin Raspberry Pi camera connectors Others: 1 40-pin I/O connector 		
Power Supply	5–28 V DC; 12 V/3 A adapter configured by default		
Power Consumption	Typical: 20 W		
Operating Temperature	0°C to 45°C		
Dimensions (H x W x D)	32.9 mm x 137.8 mm x 93.0 mm		



Atlas 300I Inference Card

Model: 3000/3010



Powered by the Ascend 310 AI processor, the Atlas 300I inference card (model: 3000/3010) unlocks superior AI inference performance. A single card provides up to 88 TOPS INT8 computing power and supports 80-channel realtime HD video analytics, making it an ideal option for intelligent scenarios such as smart city, transportation, and finance.

Specifications

Form Factor	Half-height half-length PCIe standard card		
AI Processor	Ascend 310		
AI Computing Power	88 TOPS INT8		
Memory	LPDDR4X, 32 GB, total bandwidth 204.8 GB/s		
Encoding/ Decoding	 H.264 hardware decoding, 64-channel 1080p 30 FPS (8-channel 3840 x 2160 @ 60 FPS) H.265 hardware decoding, 64-channel 1080p 30 FPS (8-channel 3840 x 2160 @ 60 FPS) H.264 hardware encoding, 4-channel 1080p 30 FPS H.265 hardware encoding, 4-channel 1080p 30 FPS J.265 hardware encoding, 4-channel 1080p 40 FPS; maximum resolution: 4096 x 2160 		
PCle	PCIe x16 Gen3.0		
Power Consumption	Maximum: 67 W		
Operating Temperature	0°C to 55°C		
Dimensions (W x D)	169.5 mm x 68.9 mm		

Superior computing

 A single card provides 88 TOPS INT8 computing power and supports 80-channel HD video real-time analytics (1080p 25 FPS), providing powerful support for edge inference

Hardware encoding/decoding

 Supports JPEG and video hardware codecs, improving image and video application performance

Low latency

 Supports large-capacity and high-bandwidth memory for feature matching scenarios, reducing application latency

Application Scenarios

Integrated in servers and industrial computers for AI inference





Atlas 300T Training Card

Model: 9000



Ultimate computing power

- 32 built-in Da Vinci Al Cores
- Industry-leading 280 TFLOPS FP16 computing power

Highest integration

- Al computing, general computing, and I/O 3-in-1
- Integrates 32 Huawei Da Vinci Al Cores, 16 TaiShan Cores, and 2 100GE RoCE v2 NICs

Highest bandwidth

- Supports PCIe 4.0 and 2 100 Gbit/s RoCE highspeed ports, with a total egress bandwidth of 56.5 Gbit/s
- Boosts the efficiency of data training and gradient synchronization by 10–70% without the need for external NICs

Application Scenarios



The Huawei Atlas 300T training card (model: 9000) is based on the Ascend 910 AI processor and works with servers to provide powerful computing for data centers. A single card provides 280 TFLOPS FP16 computing power, accelerating deep learning and training. Atlas 300T features the highest computing power, integration, and bandwidth, meeting the AI training and high-performance computing requirements of the Internet, carriers, and finance.

Specifications

Form Factor	Full height 3/4 length, dual-slot				
AI Processor	Ascend 910				
AI Computing Power	280 TFLOPS FP16 (Pro) 256 TFLOPS FP16				
Encoding/ Decoding	 16-channel 4K (or 64-channel 1080p) 60 FPS H.264/H.265 JPEG decoding: 1080p 2048 FPS, or equivalent decoding capability; maximum resolution: 8192 x 4320 PNG decoding: 1080p 240 FPS, or equivalent decoding capability; maximum resolution: 4096 x 2160 JPEG encoding: 1080p 256 FPS, or equivalent encoding capability; maximum resolution: 8192 x 4320 				
Memory	 32 GB HBM DDR4 16GB 2933 MT/s 				
Network	2 100GE QSFP-DD ports				
PCle	PCle x16 Gen4.0				
Power Consumption	Maximum: 330 W ¹				
Cooling Mode	Passive air cooling				
Environment Conditions	 Operating temperature: 5°C to 35°C Storage temperature: -40°C to +75°C Operating humidity: 5% to 90% RH (non- condensing) Storage humidity: 5% to 95% RH (non- condensing) Altitude: Between 900 m and 3000 m, the maximum operating temperature decreases by 1°C as the altitude increas every 300 m. 				

1. This specification item is in continuous optimization. The value is dynamically updated based on the optimization result.

Building a Fully Connected, Intelligent World



Atlas 500 AI Edge Station

Model: 3000



Intelligent edge

- State-of-the-art edge product with AI processing capabilities
- Fan-free heat dissipation, stable outdoor at –40°C to +70°C

Superb capacity in a compact size

- 22 TOPS INT8 computing power in the size of an STB
- 20-channel HD video processing (1080p 25 FPS)

Edge-cloud collaboration

- LTE wireless transmission
- Cloud-edge collaboration for real-time model update
- Unified device management and firmware update
 on the cloud

Application Scenarios

Independently deployed for edge intelligence



The Atlas 500 AI edge station (model: 3000) is designed for edge applications. It features superb computing performance in a compact size, strong environmental adaptability, easy maintenance, and cloud-edge collaboration, and can be widely deployed at the edge. The Atlas 500 AI edge station meets complex requirements in scenarios such as security, transportation, community, campus, shopping malls, and supermarkets.

Specifications

AI Processor	Ascend 310
AI Computing Power	22/16 TOPS INT8
Memory	LPDDR4X, 8 GB/4 GB, up to 51.2 GB/s
Encoding/ Decoding	H.264 hardware decoding, 16-channel 1080p 30 FPS (2-channel 3840 x 2160 @ 60 FPS) H.265 hardware decoding, 16-channel 1080p 30 FPS (2-channel 3840 x 2160 @ 60 FPS) H.264 hardware encoding, 1-channel 1080p 30 FPS H.265 hardware encoding, 1-channel 1080p 30 FPS JPEG decoding: 1080p 256 FPS; encoding: 1080p 64 FPS; maximum resolution: 8192 x 4320 PNG decoding: 1080p 24 FPS; maximum resolution: 4096 x 2160
Port	Network: 2 GE RJ45 ports Other I/O ports: 1 HDMI port 1 input and 1 output (stereo), 3.5 mm audio connector 2 external USB 2.0 ports and 1 internal USB 2.0 port (Type-A)
Typical Power Consumption	Without disks: 25 W With disks: 40 W
Environment Conditions	Without disks: -40°C to +70°C With disks: -40°C to +60°C
Dimensions (H x W x D)	Without disks: 220 mm x 45 mm x 235 mm With disks: 220 mm x 45 mm x 355



Atlas 800 Inference Server

Model: 3010



Flexible configuration for various workloads

- Supports any combination of SAS/SATA/NVMe/M.2 SSD drives
- Supports LAN on motherboard (LOM) and FlexIO cards, providing rich network interface options

Smart video analysis

• Supports up to 7 Atlas 300I inference cards and 560channel real-time HD video analytics (1080p 25 FPS)

MARKEN L		

Form Factor 2U AI server

1 or 2 Intel® Xeon® Skylake or Cascade Lake Processor Scalable processors, 205 W TDP Processor 24 DDR4 DIMM slots, up to 2933 MT/s Memory AI Accelerator Up to 7 Atlas 300I inference cards Card AI Computing Up to 616 TOPS INT8 Power 8 x 2.5" SAS/SATA drives 12 x 3.5" SAS/SATA drives Local 8 x 2.5" SAS/SATA + 12 x 2.5" NVMe 24 x 2.5" SAS/SATA drives Storage 24 x 2.5" NVMe 25 x 2.5" SAS/SATA drives RAID RAID 0, 1, 5, 6, 10, 1E, 50, or 60 10 PCIe Gen3.0 PCle (including 1 RAID controller card and 1 FlexIO) 2 hot-swappable PSUs, with support for 1+1 redundancy. Supported options include: 550 W AC Platinum PSUs, 900 W AC Power Platinum/Titanium PSUs, and 1500 W AC Supply Platinum PSUs 1500 W 380 V HVDC PSUs, 1200 W -48 V to -60 V DC PSUs 4 hot-swappable fan modules, supporting N+1 Fan Modules redundancy Operating 5°C to 45°C Temperature Chassis with 3.5" drives: Dimensions 748 mm x 86.1 mm x 447 mm $(H \times W \times D)$ Chassis with 2.5" drives: 708 mm x 86.1 mm x 447 mm

Powered by the Intel processors, the Atlas 800 inference server (model: 3010) supports up to 7 Atlas 300I inference cards for 560-channel real-time HD video analytics. It is widely used for AI

inference in data centers.

Specifications

Application Scenarios

Deployed in data centers to enable AI inference





Atlas 800 Training Server

Model: 9010



The Atlas 800 training server (model: 9010) is an Al training server based on the Intel processors and Huawei Ascend 910 processors. It features the industry's highest computing density and high network bandwidth. The server is widely used in deep learning model development and training scenarios, and is an ideal option for computingintensive industries, such as smart city, intelligent healthcare, astronomical exploration, and oil exploration.

Specifications

Form Factor	4U AI server		
Processor	2 Intel V5 Cascaded Lake processors		
Processor Memory	Up to 24 DDR4 DIMM slots, supporting RDIMMs		
AI Processor	8 Ascend 910 processors		
HBM	256 GB		
AI Computing Power	2.24 PFLOPS FP16 2 PFLOPS FP16		
Local Storage	 2 x 2.5" SATA + 8 x 2.5" SAS/SATA 2 x 2.5" SAS/SATA + +6 x 2.5" NVMe 		
RAID	RAID 0, 1, 10, 5, 50, 6, or 60		
Network	8 100GE 1 OCP NIC 3.0 standard card, supporting 2 25GE		
PCIe Expansion	Up to 6 PCIe 3.0 x8 slots		
PSUs	4 hot-swappable 3 kW AC PSUs, supporting 2+2 redundancy		
Power Supply	 200–240 V AC 240 V DC 		
Power Consumption	Maximum: 5.6 kW ¹		
Cooling Mode	Air cooling		
Fan Modules	8 hot-swappable fan modules, supporting N + 1 redundancy		
Temperature	 Operating: 5°C to 35°C Storage: -40°C to +70°C 		
Dimensions (H x W x D)	790 mm x 175 mm x 447 mm		

1. This specification item is in continuous optimization. The value is dynamically updated based on the optimization result.

8

The ultimate computing density

- Up to 2.24 PFLOPS FP16 in a 4U space
- 1.5x the computing density of industry peers

High-speed network

 8 100G RoCE v2 ports, slashing cross-server chip interconnect latency by 10–70%

Application Scenarios

Deployed in data centers to enable Al training Model training Model training Fraining Smart city Smart city Astronomical exploration Oil exploration

Building a Fully Connected, Intelligent World

Huawei Intelligent Computing Portfolio



Bring digital to every person, home and organization for a fully connected, intelligent world



Copyright © Huawei Technologies Co., Ltd. 2020. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademark Notice

, HUAWEI, and trademarks or registered trademarks of Huawei Technologies Co., Ltd. Other trademarks, product, service and company names mentioned are the property of their respective owners.

General Disclaimer

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor a n a cceptance. H uawei may change the information at a ny time without notice.

HUAWEI TECHNOLOGIES CO., LTD.

Huawei Technologies France S.A.S.U. 18 Quai du point du jour 92100, Paris Tel : +33 146206069 ascend.huawei.com

For more information To learn more about Huawei servers, contact Huawei sales representatives or business partners, or visit: http://e.huawei.com/en/products/cloud-computing-dc/servers