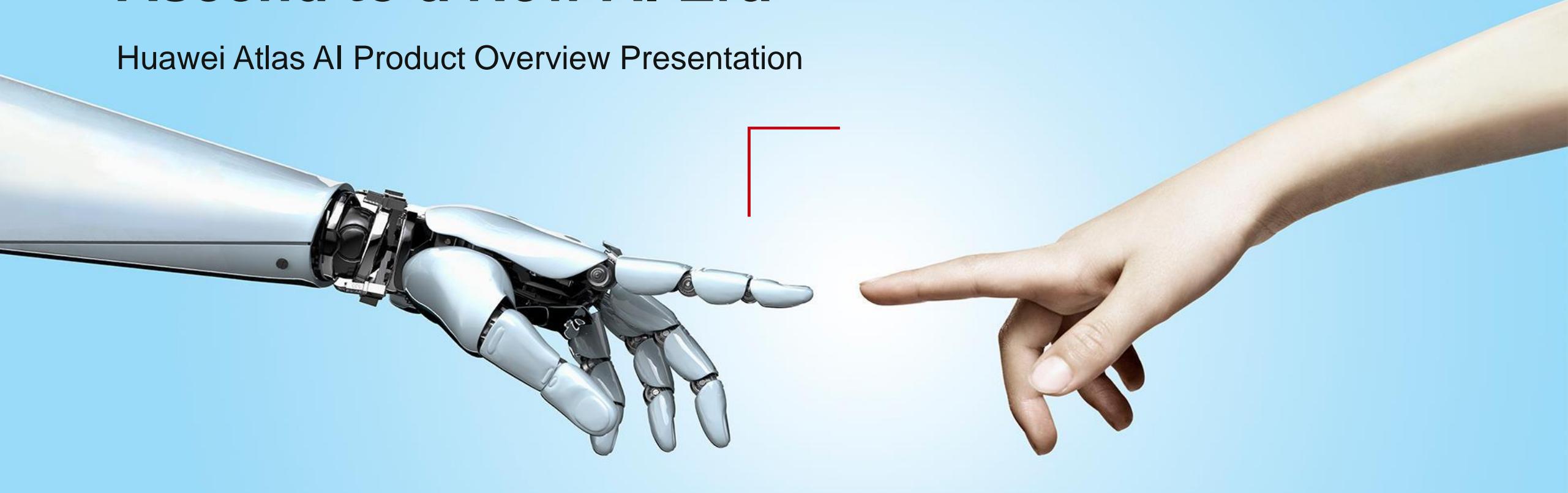


Ascend to a New AI Era

Huawei Atlas AI Product Overview Presentation



Huawei: Leading Global Provider of ICT Infrastructure and Smart Devices



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

Huawei's end-to-end portfolio of products, solutions and services are both competitive and secure. Through open collaboration with ecosystem partners, we create lasting value for our customers, working to empower people, enrich home life, and inspire innovation in organizations of all shapes and sizes.

At Huawei, innovation focuses on customer needs. We invest heavily in basic research, concentrating on technological breakthroughs that drive the world forward.



188,000

Employees



80,000

R&D employees



170+

Countries and regions



No. 68 in

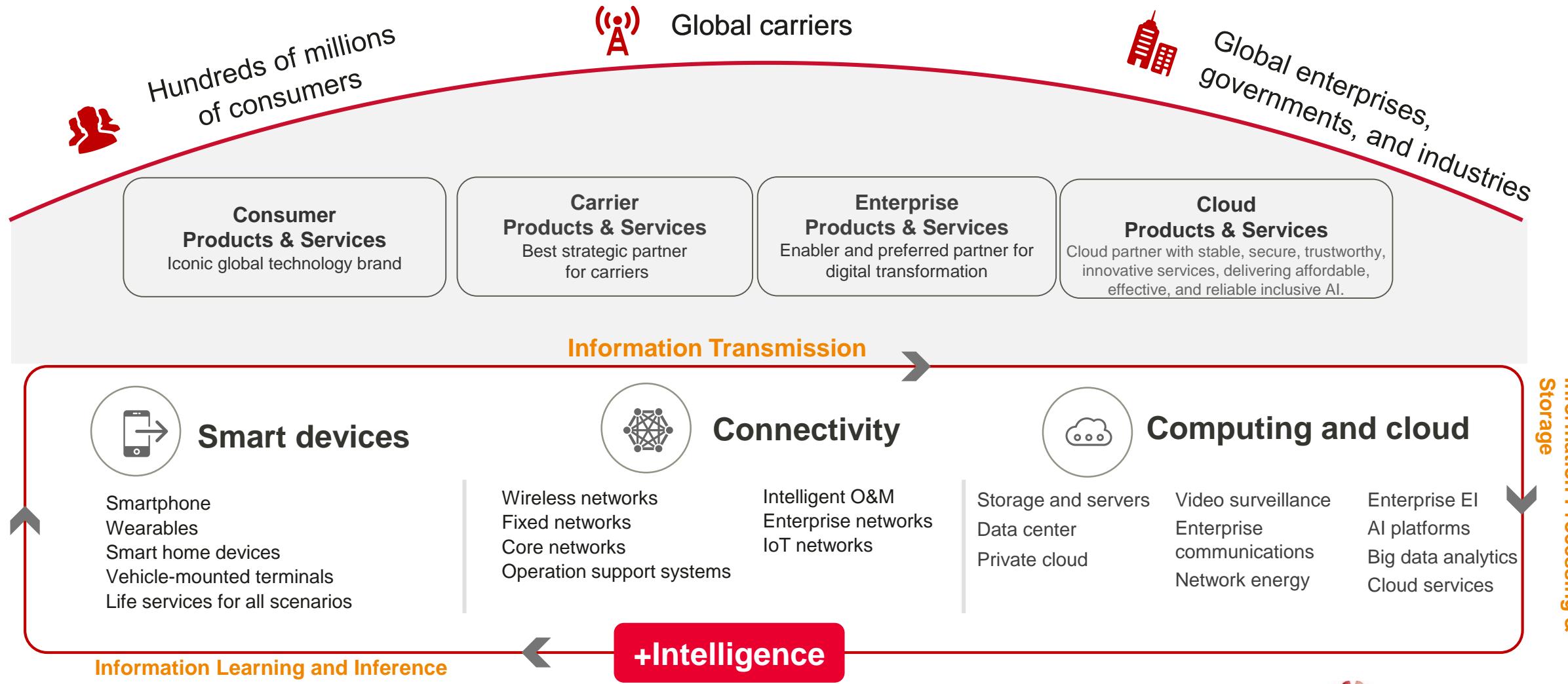
Interbrand's
Top 100
Best Global Brands



No. 61 in

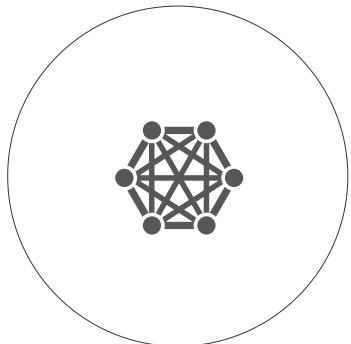
Fortune Global 500

Focusing on Smart Devices, Connectivity, Computing, and Cloud; Providing Products and Solutions for Three Customer Groups



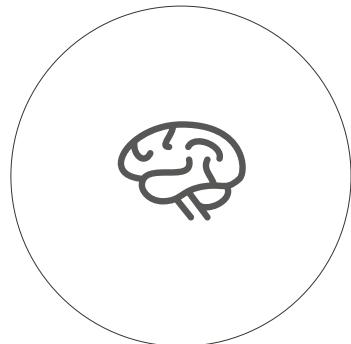
Huawei's Value Propositions

Building a fully connected, intelligent world



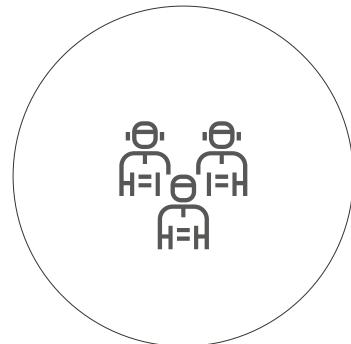
Ubiquitous Connectivity

- Wireless access
- Wired access
- Bearer network
- Core network
- Enterprise network



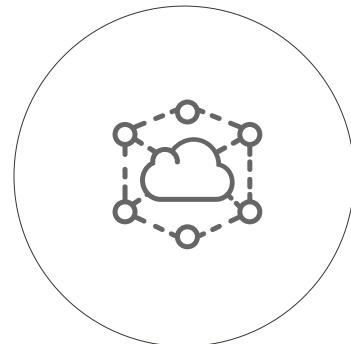
Pervasive Intelligence

- Mobile AI
- SoftCOM AI network
- Edge intelligence
- Enterprise intelligence (EI)
- Intelligent computing (servers + AI)
- Intelligent storage
- Intelligent security (management, analysis, and control platforms)



Personalized Experience

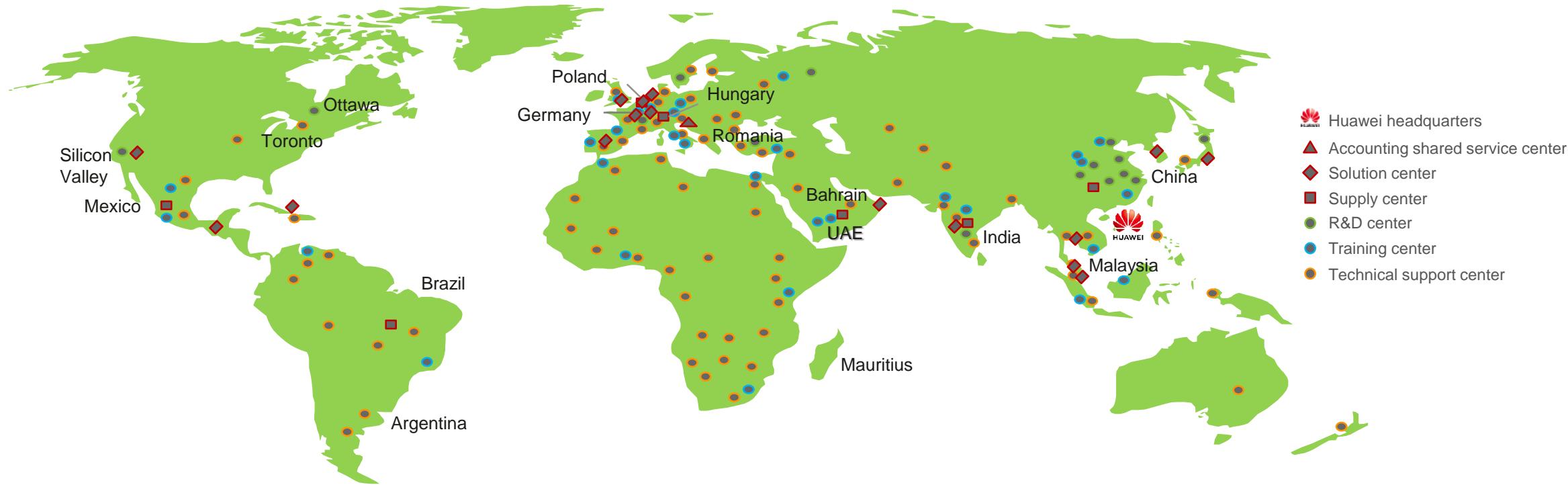
- Intelligent experience for all scenarios: Personal, car, home, and office scenarios
- Smart experience
- Chip-device-cloud synergy
- Intuitive interaction
- Multi-level security protection: Chips, devices, and cloud



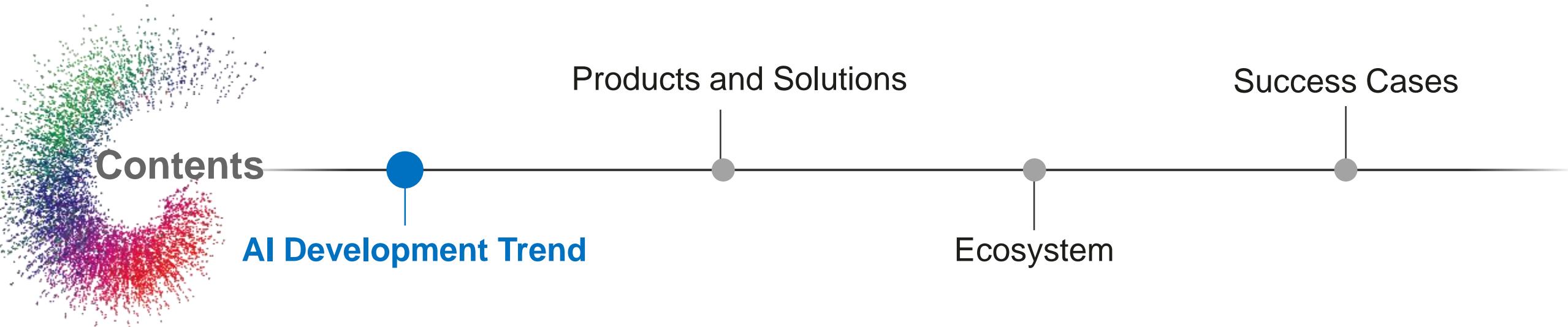
Digital Platform

- Open, reliable cloud platform (Public cloud and private cloud)
- Open hybrid cloud architecture
- ICT infrastructure platform with device-network-cloud synergy

Globalized Positioning, Localized Operations



- Operations in 170+ countries; approximately 188,000+ employees of 160+ nationalities; 70% of staff hired locally
- Huawei is increasing its "glocalization", delegating more end-to-end business responsibility and authority to local teams
- We aim to develop a healthy industry in which every player can succeed, and we pride ourselves on being a responsible local corporate citizen



With the Advent of the 4th Industrial Revolution, AI Is Driving Global Development



Steam engine



Electric power



Information technology



Artificial intelligence

Global macro trends



Sustainability



New growth of digital economy



Upgrade to smart city



Digital transformation of industries



Better consumer experience



Scenario-specific solutions

Two Key Technologies in the Intelligent World

Building a fully connected,
intelligent world

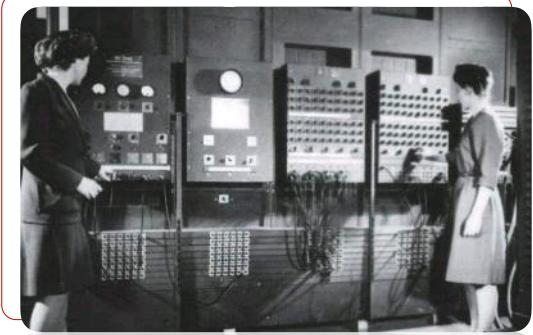
Connectivity

100 billion connections

Computing

\$2 trillion computing market

Computing Extends Human Capabilities



Mainframe computers



PCs

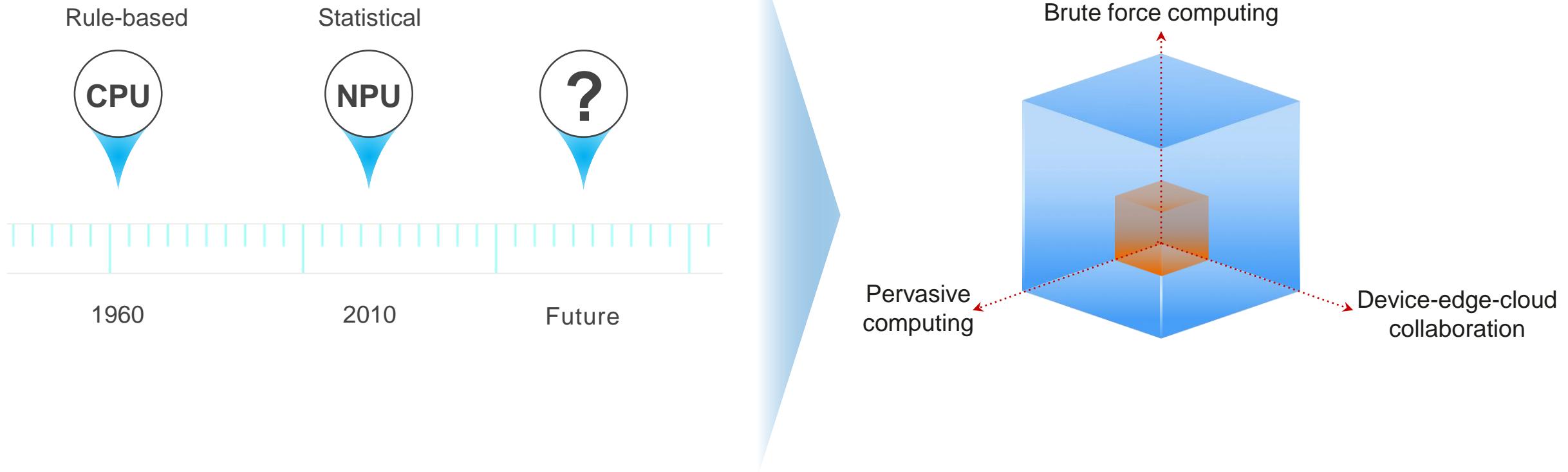


Mobile phones



Wearables

Computing Modes Constantly Evolve, Propelling the Industry into the Intelligent Era



Many Countries Have Released AI Strategies



2019: "Smart +" was reported on the work of the governments, highlighting smart manufacturing, intelligent transformation of industries, and deepened big data application
2018: Artificial Intelligence Innovation Action Plan for Higher Schools; 2019: AI included in the compulsory courses of primary and secondary schools
2017: Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry
2017: Next Generation Artificial Intelligence Development Plan
2016: Internet + Artificial Intelligence Three-Year Action Plan



2019: The President signed the "AI Initiative" (the federal government increases investment and protects the US leading edge)
2017: National Robotics Initiative 2.0, NRI-2.0
2016: National AI Research and Development Strategic Plan Preparing for the Future of AI
2013: Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative
2010: US's DARPA started to support AI application in various fields



2019: South Korea released the Dataset AI Development Plan, announcing the increase in AI investment and becoming the leader of global data and AI by 2030
2017: Society 5.0, Artificial Intelligence Technology Strategy
2015: Japan's Robot Strategy



2019: Germany released Germany's Industrial Strategy 2030 in February, emphasizing the use of AI technology to maintain the global industrial position.
2018: France – French Intelligence Artificielle
2018: UK – Industrial Strategy: Artificial Intelligence Sector Deal
2018: EU – intends to release an AI strategy and coordination plan; Germany – Key points for a Federal Government Strategy on Artificial Intelligence, AI development strategy to be released by the end of the year
2013: EU – Human Brain Project (HBP)

Source: Huawei EBG New ICT, 2019



Contents

Products and Solutions

AI Development Trend



Ecosystem



Success Cases

Huawei's AI Strategy

Invest in basic research

Develop fundamental capabilities for data & power-efficient (i.e., less data, computing, and power needed), secure & trusted, automated/autonomous machine learning in computer vision, natural language processing, decision/inference, etc.



Build a full-stack AI portfolio

Deliver abundant and affordable computing power.
Provide an efficient and easy-to-use AI platform with full-pipeline services.
Adaptive to all scenarios, both standalone and cooperative scenarios between cloud, edge, and device.



Develop open ecosystem and talents

Collaborate widely with global academia, industries, and partners.



Strengthen existing portfolio

Bring an AI mindset and techniques into existing products and solutions to create larger value and enhance competitive strengths.

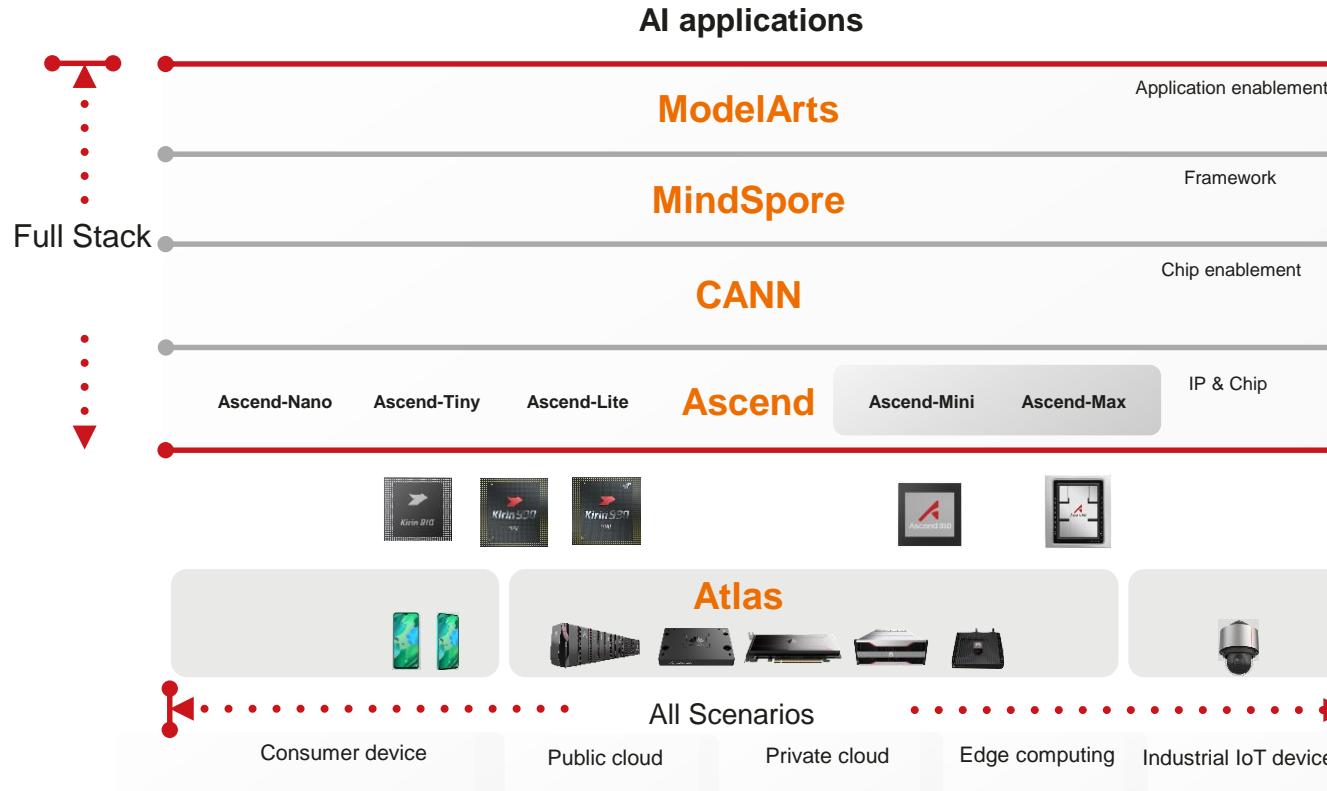


Drive operational efficiency

Apply AI to massive volumes of routine business activities for better efficiency and quality.



Huawei Full-Stack, All-Scenario AI Solution



Application enablement: provides E2E services (ModelArts), layered APIs, and pre-integrated solutions

MindSpore: a unified training and inference framework for independent and collaborated operation across the device-edge-cloud

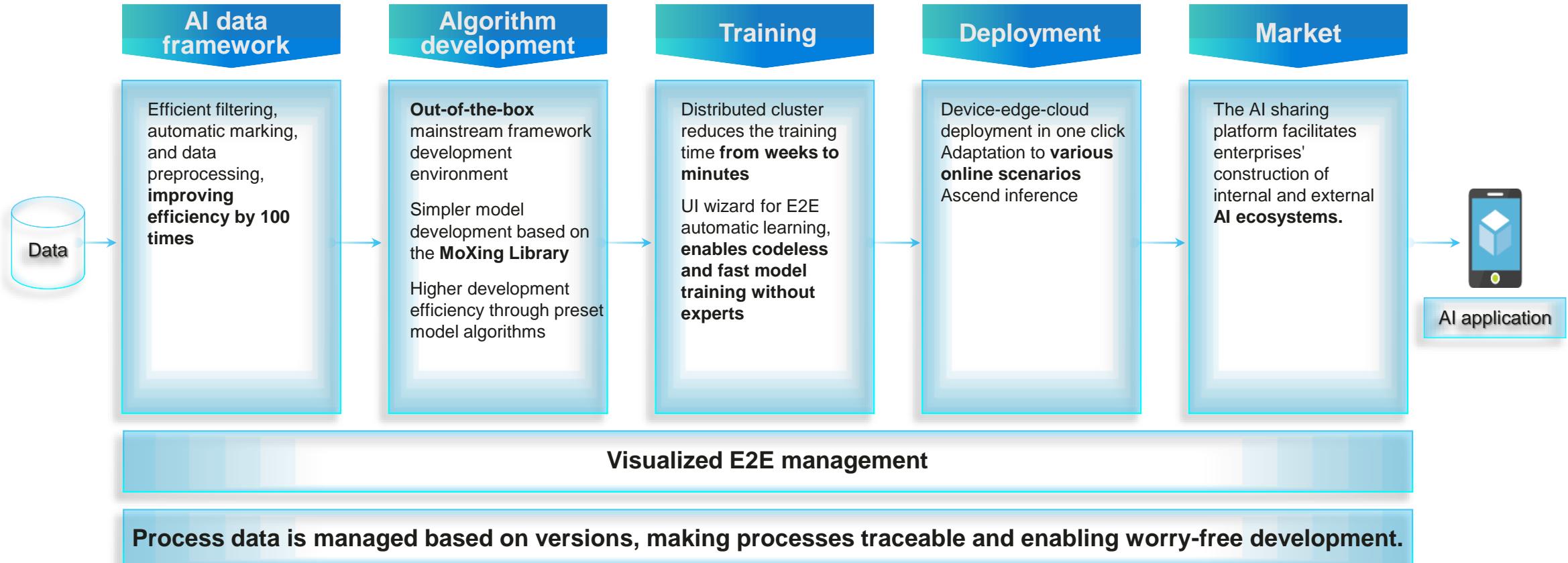
CANN: chip operator libraries and highly automated operator development tools

Ascend: unified and scalable AI IPs and chips

Atlas: an all-scenario AI infrastructure solution for the device-edge-cloud based on the Ascend series AI processors and various product form factors

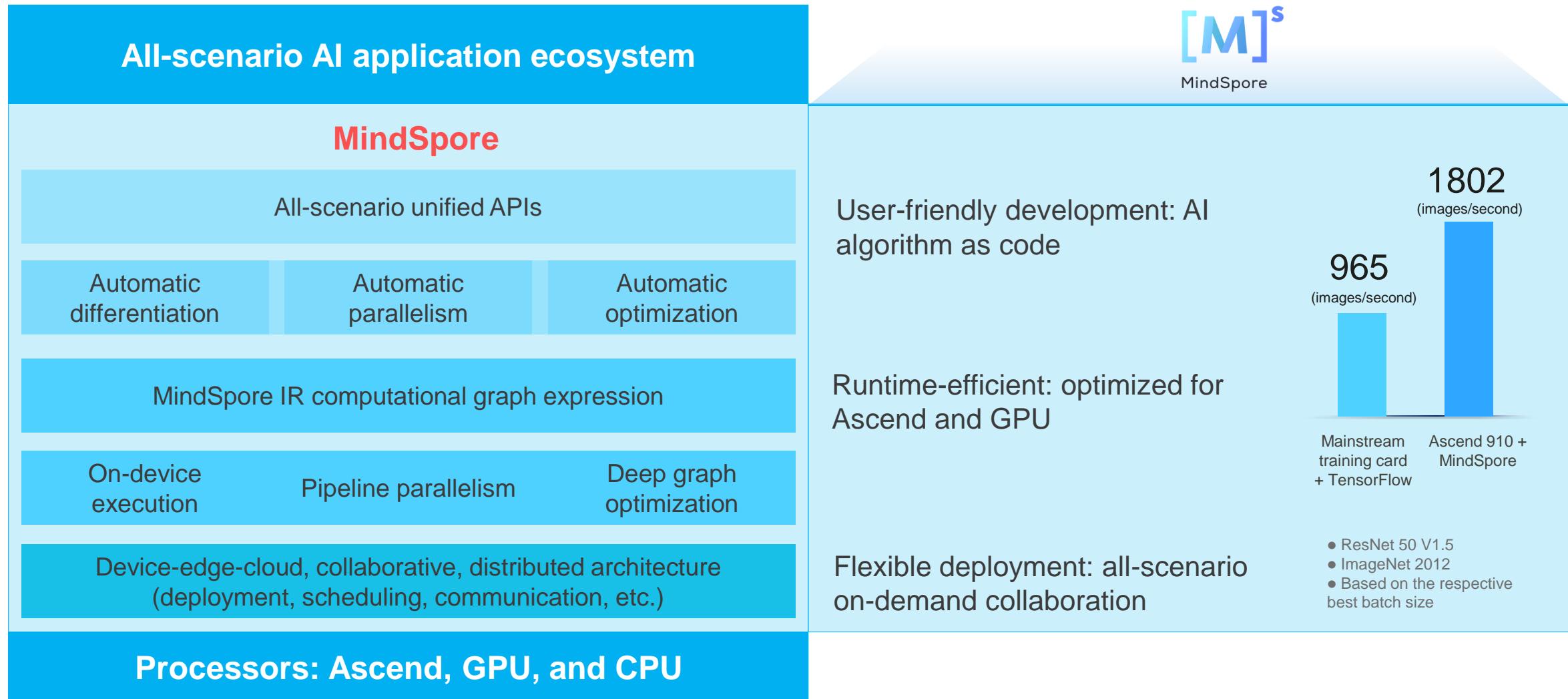
ModelArts: A Faster, Inclusive AI Development Platform

- A developer-centric and full-lifecycle platform covering data preparation, model training, and inference execution
- Facilitating AI adoption by enabling developers of different needs with hierarchical APIs and pre-integrated solutions

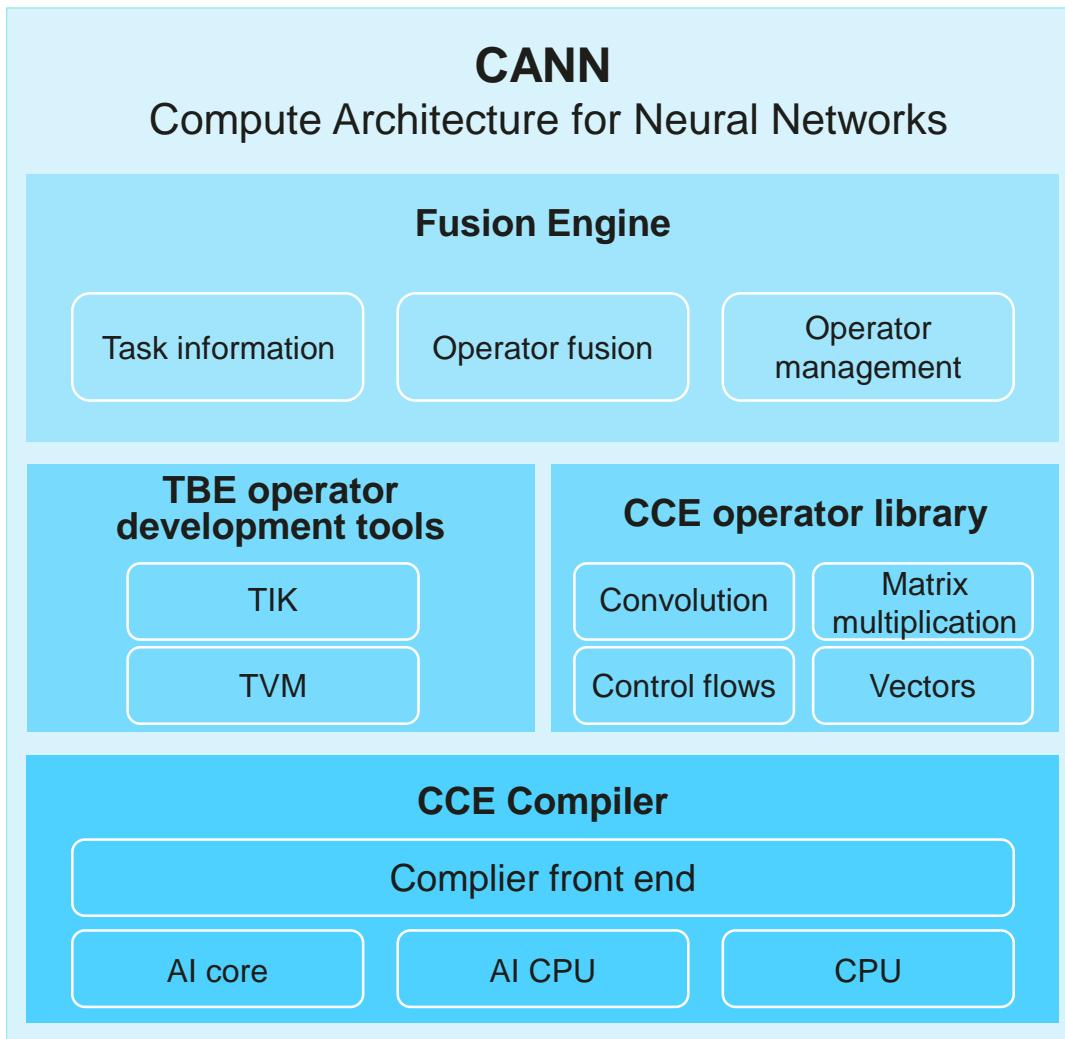


In the latest training ranking of Stanford DAWN Bench, Huawei cloud ModelArts takes 10 minutes and 28 seconds, making ModelArts the world's fastest, which is 8 minutes faster than the Amazon cloud.

MindSpore: All-Scenario AI Computing Framework

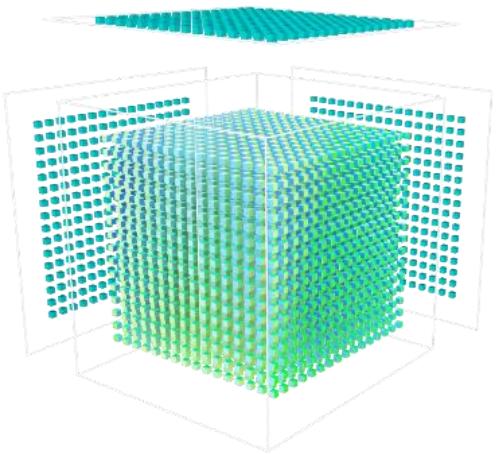


CANN: High-Performance Chip Operator Library and Automated Operator Development Tools



- **CANN:** Includes the chip operator library and highly automated operator development tools for optimal development efficiency and Ascend performance matching
- **Fusion Engine:** Ascend-based memory architecture implements operator fusion, reduces operator memory data transfer, and improves performance by 17%
- **Tensor Boost Engine (TBE) operator development tools:** various preset APIs for custom operator development and automatic optimization, **shortening project cycle by 53% and reducing labor by 42%**
- **CCE operator library:** high-performance operator library based on in-depth collaborative optimization of the Ascend chip, **for 10% performance boost**
- **CCE compiler:** compiler and binary tool set using heterogeneous hybrid programming language (C/C++ extension) to optimize performance and programming, enabling Ascend to support all scenarios

Da Vinci Architecture, for Tensor Computing



3D Cube: Made for AI Matrix Operation



Human Intelligence

Obtain complete semantics from holographic information input

Artificial Intelligence

Data, computing power, and algorithms

3D Cube:

16*16*16 3D Elastic Cube

- High computing power: 4096 FP16 MAC operations completed within a clock cycle
- Efficiency: from IP modules of dozens of mW to chips of hundreds of watts, adapting to the smoothly scalable architecture of device-edge-cloud

Ascend AI Processors Infuse Superior Intelligence into Computing



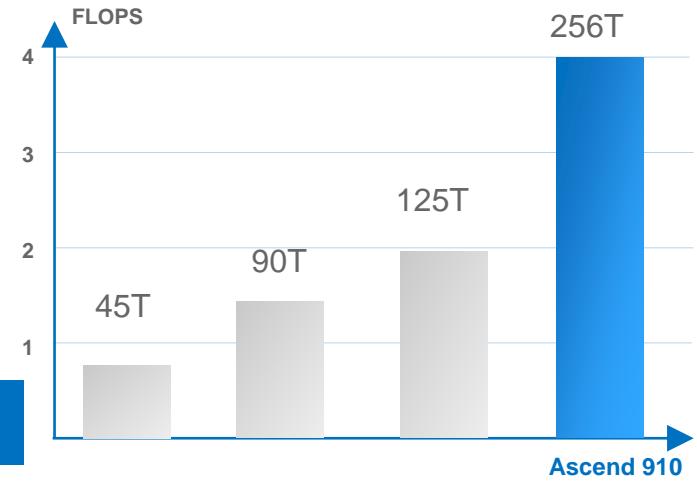
Ascend 310

AI SoC with Ultimate Perf./Watt



Ascend 910

World's Most Powerful AI Processor



* Normalized to 16-bit

Ascend-Mini
Architecture: Da Vinci

Half precision (FP16): 8 TFLOPS
Integer precision (INT8): 16 TOPS
16-channel full-HD video decoder: H.264/265
1-channel full-HD video encoder: H.264/265

Max. power consumption: 8 W

Ascend-Max
Architecture: Da Vinci

Half precision (FP16): 256 TFLOPS
Integer precision (INT8): 512 TOPS
128-channel full-HD video decoder: H.264/265

Max. power consumption: 310 W

Atlas AI Computing Platform Portfolio



Superior computing power



All-scenario deployment



Cloud-edge-device collaboration



Atlas 900 AI cluster



Atlas 800 AI server
Model: 9000/9010



Atlas 800 AI server
Model: 3000/3010



Atlas 300 AI accelerator card
Model: 3000/9000



Atlas 500 AI
edge station



Atlas 200 AI
accelerator module

Cloud ----- Edge ----- Device -----

Cloud

Edge

Device

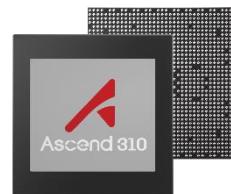


Ascend 310
AI processor



Ascend 910
AI processor

Atlas Accelerates AI Inference



Ascend 310
AI processor

Intelligence into terminal devices
7x higher performance



Atlas 200 AI accelerator module
Model: 3000

64-channel video inference
Industry's highest density



Atlas 300 AI accelerator card
Model: 3000

Edge intelligence and cloud-edge collaboration



Atlas 500 AI edge station
Model: 3000

AI inference platform with ultimate computing power



Atlas 800 AI server
Model: 3000/3010

Atlas 200: Smartening up Terminal Devices, with a 7x Performance Leap

Cameras Drones Robots



30 → 200

Number of faces in a
single frame



AI camera

Run multiple algorithms
in parallel

Pixel-level image
segmentation

- + Facial recognition
- + Posture recognition
- + Vehicle recognition

- **16 TOPS INT8, 9.5 W**
- **16-channel HD video real-time analytics, and JPEG decoding**
- **4 GB/8 GB memory, PCIe 3.0 x4 interface**
- **Operating temperature: -25°C to +80°C**

Atlas 200 Empowers Robotic Arms for Garbage Sorting

Challenges

- China has to deal with more than 400 million tons of domestic waste every year. Garbage sorting involves intensive workload and labor, and endangers the health of the workers.

Solution

- Huawei provides a solution composed of the Atlas 200 AI accelerator module, development interface, and HUAWEI CLOUD resources, and works with ABB to develop a vision-enabled robotic arm solution for garbage sorting.
- Atlas 200 is integrated into the robotic arm control board to identify garbage types and control the robotic arm, which boosts sorting efficiency and frees up human labor.
- Atlas 200 delivers the leap from programming to learning in industrial robotic arms, bridging the "last mile" in industrial automation. Robotic arms now act with real-time decision making instead of moving as programmed in advance.

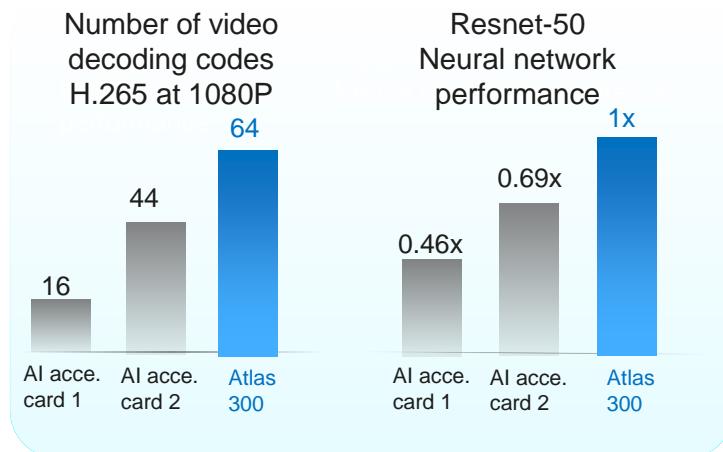
Atlas 300: Industry's Highest-Density AI Accelerator Card, Supporting 64-Channel Video Inference



Model: 3000

- **64 TOPS of INT8, 67 W**
- **32 GB memory**
- **64-channel HD video real-time analytics**

Video Analytics | OCR | Speech Recognition |
Precision Marketing | Medical Image Analysis



Industry-leading performance

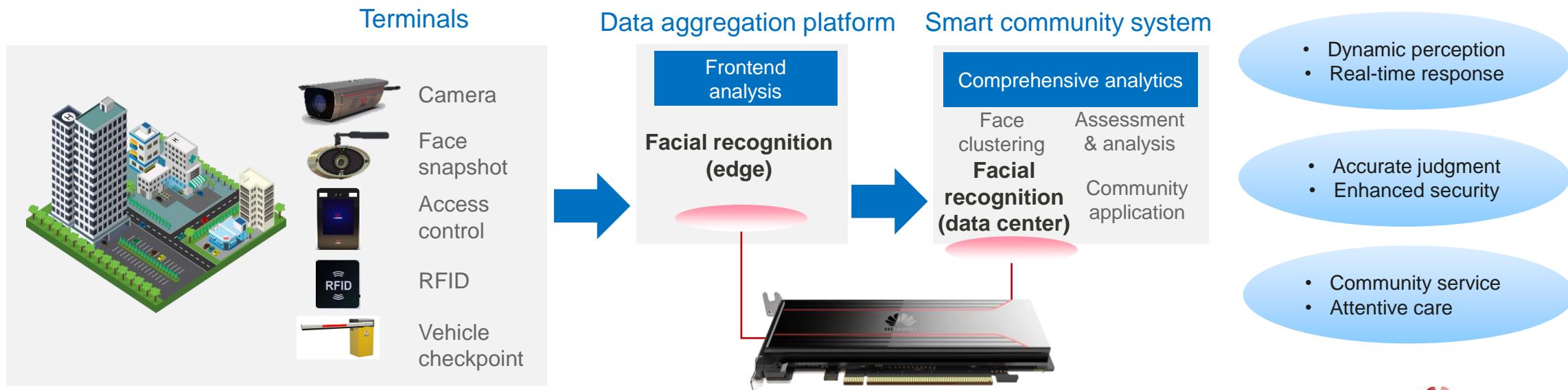
Atlas 300 Helps Build a Safe Smart Community

Challenges

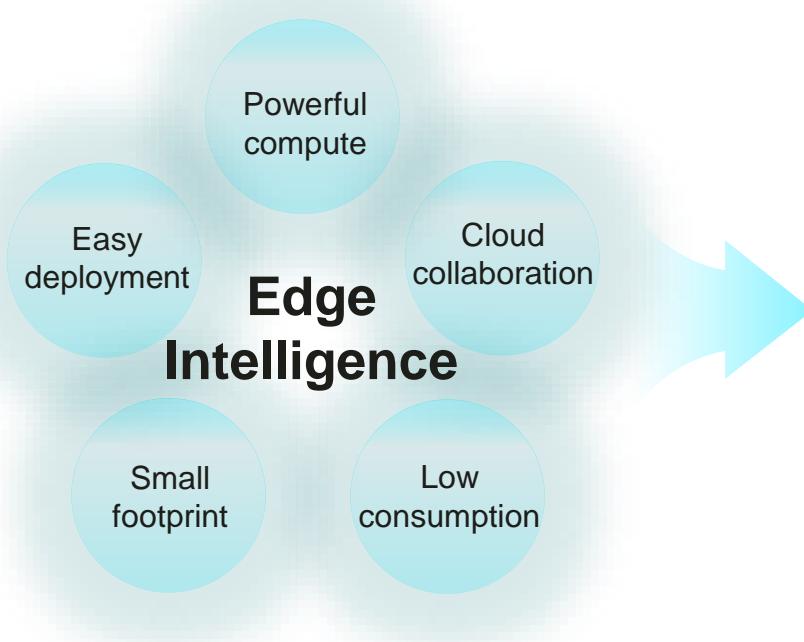
- Community information collection is slow and difficult, especially in real population management and basic information collection.
- Data resources are isolated, making it impossible to apply digitalized statistics or big data methods to the population information of most communities.
- Current methods cannot manage floating population in time or based on networks.

Solution

- The smart community solution based on the Atlas 300 AI inference accelerator card automatically identifies community residents, facilitates community residents' access, and helps communities deploy security warnings for suspicious persons, improving community security.
- The web interface supports facial retrieval, frequency analysis, statistics report, resource management, and system management. The statistics data can be uploaded to the management platform for display.



Atlas 500 AI Edge Station



- **16 TOPS INT8**
- **25–40 W**
- **Wi-Fi & LTE**
- **16-channel HD video real-time analytics**
- **Fanless, stable at -40°C to +70°C**

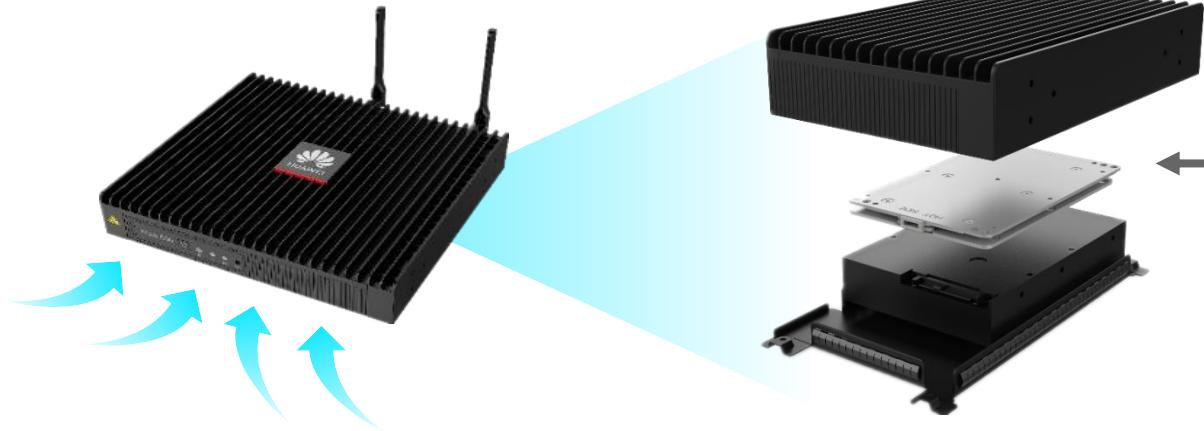
Built for Harsh Edge Environments

Extreme cold | Extreme hot | Heavy dust



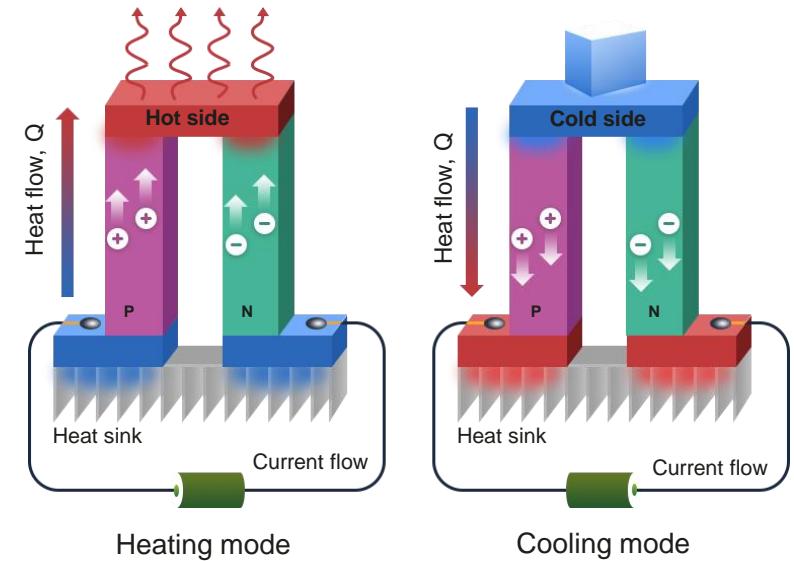
TEC Semiconductor Cooling Technology

Atlas 500



Thermoelectric
cooling (TEC)
semiconductor

Peltier Effect, Semiconductor Cooling



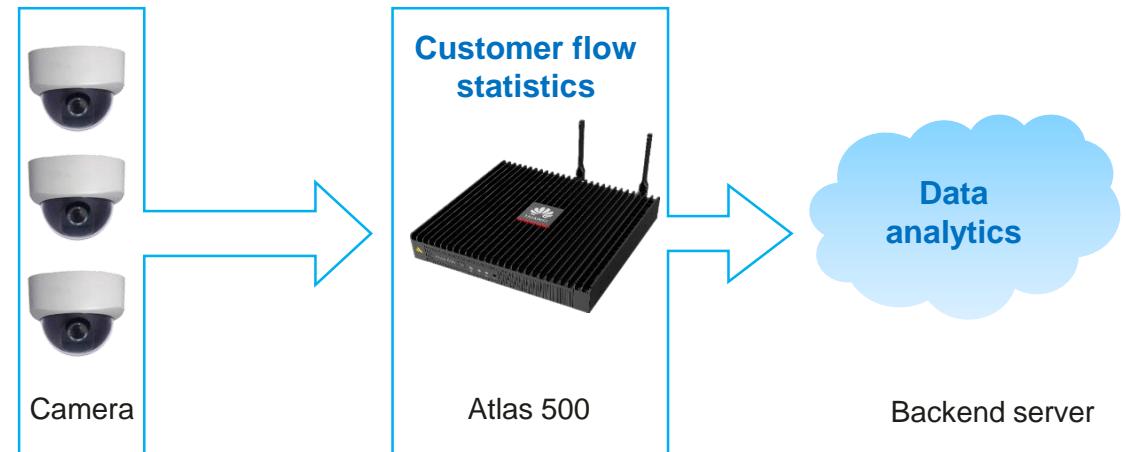
AI Edge Station Enables Accurate Customer Flow Analysis for Smart Shopping Malls

Challenges

Shopping mall companies cannot quantify customer turnover, and therefore fail to analyze the direct cause of sales fluctuation. They cannot adapt agilely to customer needs.

Solution

- Atlas 500 smart shopping mall solution: Uses AI algorithms to analyze crowd characteristics such as human bodies, postures, and traffic, and uses big data analytics to obtain operation status and identify customer interest.
- Quantify and visualize operation data, and clearly display business operation problems.
- Identify VIP customers and recommend products based on customer interest to improve the sales success rate.
- Adjust the rent based on the passenger flow data to ensure shop owners' rights and interests.



Atlas 800 AI Server – 3000 Model: Efficient Inference Platform Powered by Kunpeng



- **2 Kunpeng 920 processors in a 2U space**
- **8 PCIe slots, supporting up to 8 Atlas 300 AI accelerator cards**
- **Up to 512-channel real-time HD video analytics**
- **Air-cooled, stable in 5°C to 40°C environments**

Powerful Real-Time Inference

- Supports **8 Atlas 300 AI accelerator cards**, industry-leading 64 TeraOPS per card; **512-channel real-time HD video analytics** per system
- Runs on 2 **Kunpeng 920 processors**, unlocking supreme performance and compute power; leverages a multi-core architecture for application acceleration

Efficient Computing

- Provides an AI computing platform with high efficiency, low consumption for inference scenarios, fully leveraging the multi-core and energy-efficient advantages of Kunpeng
- Atlas 300 runs at an ultra-low power of 67 W, fueling the AI server with faster computing and **higher performance per watt**

Atlas 800 AI Server – 3010 Model: Flexible Inference Platform Based on Intel Processors



- 2 Intel® Xeon® SP Skylake or Cascade Lake processors in a 2U space
- 8 PCIe slots, supporting up to 7 Atlas 300/NVIDIA T4 AI accelerator cards
- Up to 448-channel real-time HD video analytics
- Air-cooled, stable in 5°C to 35°C environments

Flexible Configuration, Workload-Adaptive

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

Intelligent Video Analytics

- Supports **7 AI accelerator cards**, adapting to various video and image analytics scenarios
- Uses the Atlas 300 dedicated decoding engine for transcoding and inference of up to 448-channel HD video streams

Atlas 200 DK AI Developer Kit: Superb Computing and Ease of Use



**Full-stack AI development on
and off the cloud**

Developers



Set up a dev environment with 1 laptop
Ultra low cost for local independent environment, with multiple functions and interfaces to meet basic requirements

Researchers



Local dev + Cloud training collaboration
The Developer Kit and HUAWEI CLOUD use the same protocol stack, allowing for training on the cloud and local deployment without any modification

Startups



Code-level demo
Based on the reference architecture, algorithm functions are completed with only 10% code modification; support developer community interaction; seamless porting for commercial products

Atlas Accelerates AI Training



Ascend 910
AI processor

The ultimate training card



Atlas 300 AI accelerator card
Model: 9000

World's most powerful training server



Atlas 800 AI server
Model: 9000/9010

World's fastest AI training cluster



Atlas 900 AI cluster

Atlas 300 AI Accelerator Card: Industry's Most Powerful Training Card



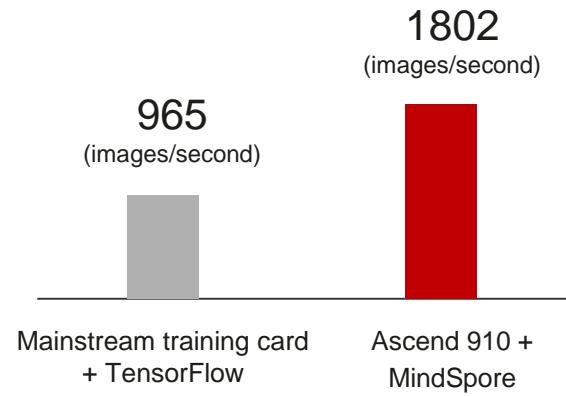
Atlas 300

Model: 9000

2x ↑

Computing power
per card

256T FLOPS at FP16



Test benchmark:
• ResNet 50 V1.5
• ImageNet 2012
• Optimal batch size respectively

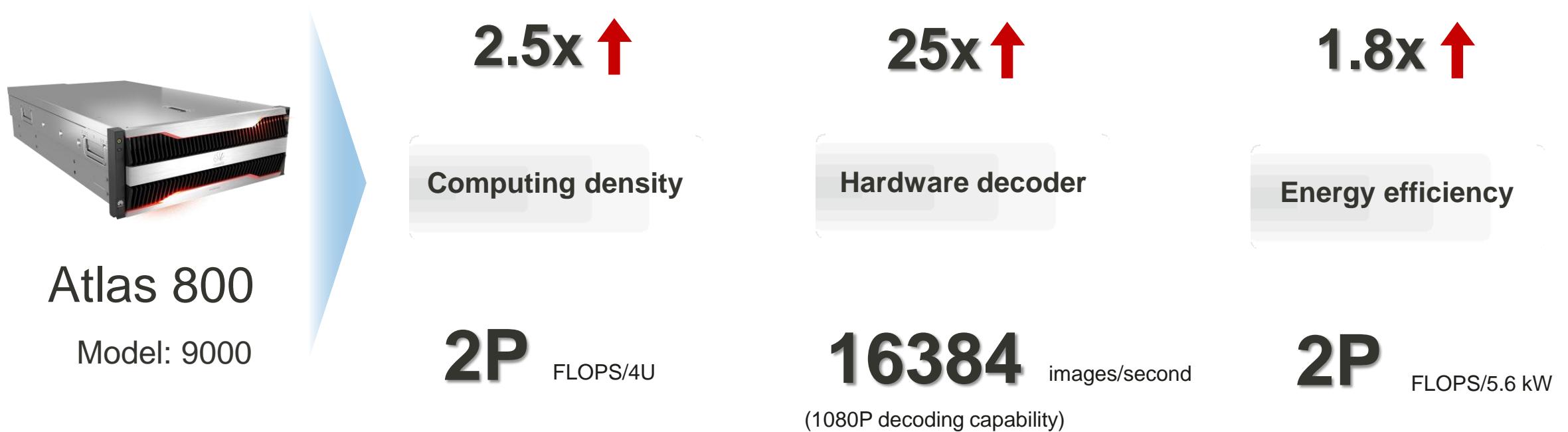
70% ↓

Gradient
synchronization latency

Direct 100G RoCE



Atlas 800 AI Server: World's Most Powerful AI Training Server



Atlas 900 AI Cluster: World's Fastest AI Training Cluster



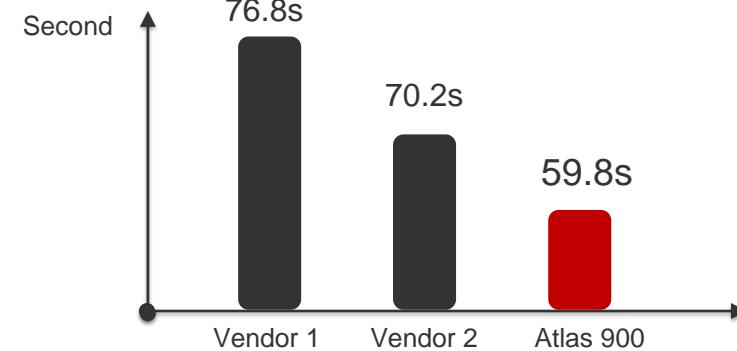
Atlas 900

256–1024 PFLOPS at FP16

Leading computing power | Best cluster network |
Ultimate heat dissipation



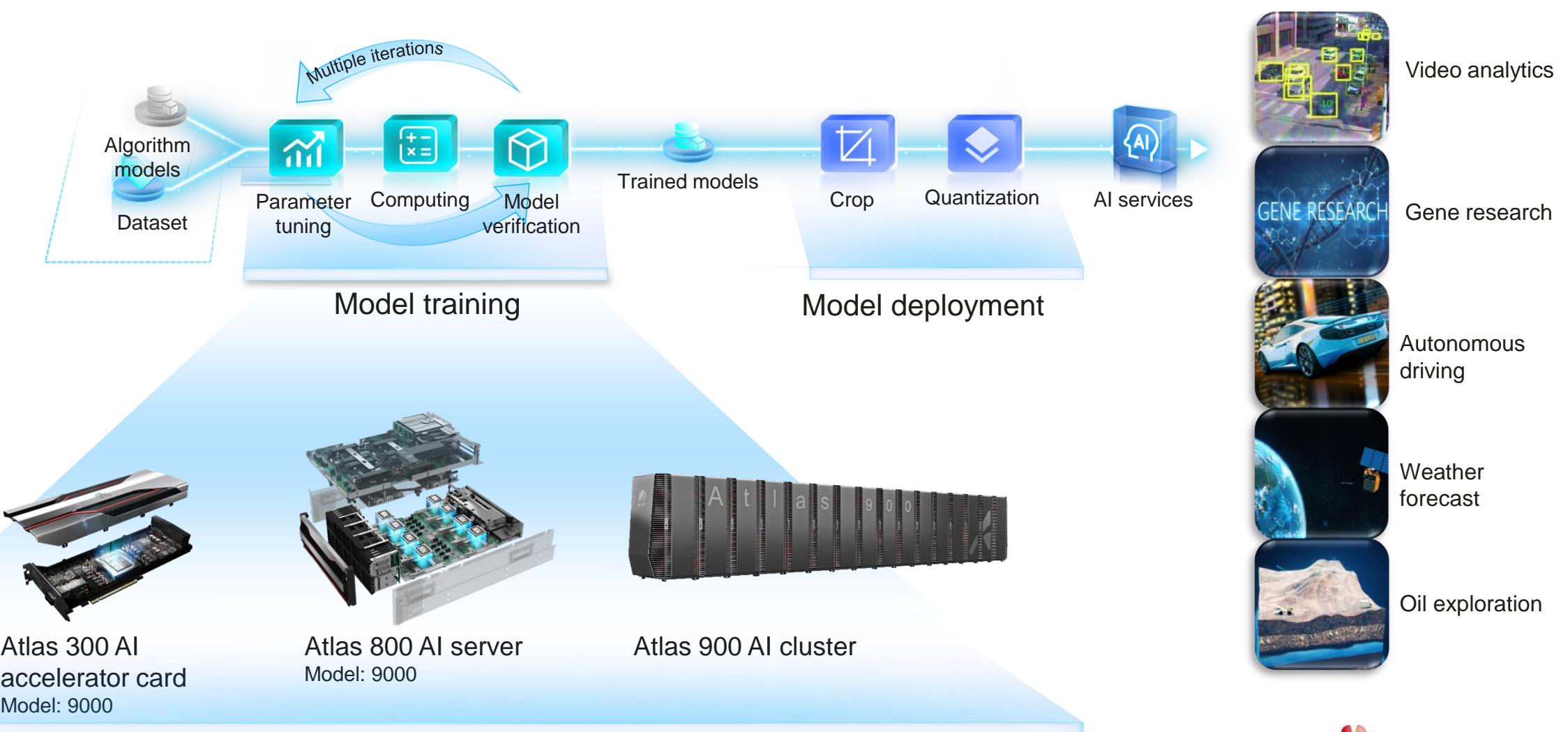
World's #1: 59.8s

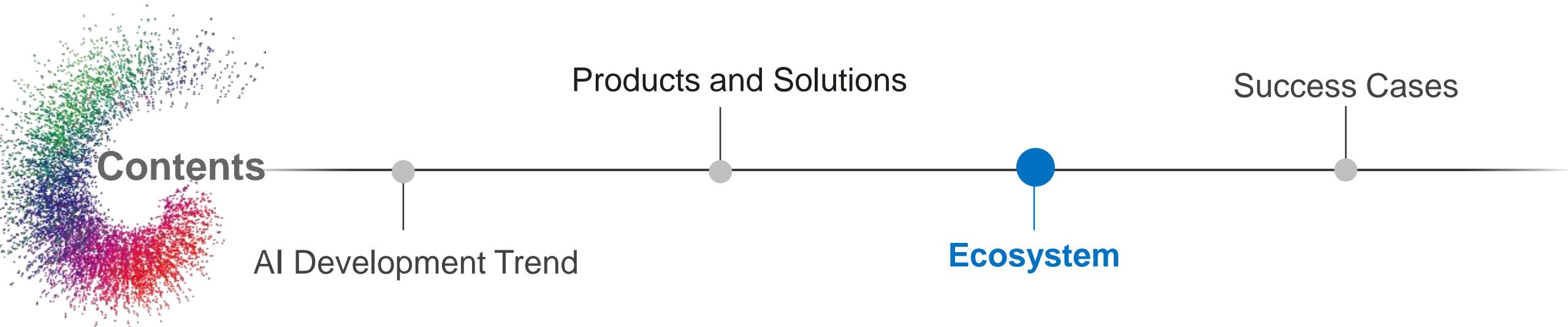


Test benchmark:

- Benchmark: ResNet-50 V1.5 model, ImageNet-1k dataset
- Cluster: 1024 Ascend 910 AI processors
- Accuracy: 75.9%

Atlas Accelerates AI Model Training for Various Applications





1 Platform + 3 Plans Support Ascend Industry Partners and Developers

2000 partners, 1 million developers in 5 years



Business
partners



Developers



Universities

Solution Partner
Program

Developer
Enablement Plan

AI Talent
Development Plan

Platform for Ascend industry development

Industry cooperation

Open source

Technical support

Marketing support

Work Closely with Industry Organizations to Incubate Industry Standards



Huawei tasked with building China's Next-Generation AI Open Innovation Platform for basic software and hardware

Industry organization

Research institute

Industry association

Serves as the governing unit in 10+ AI industry organizations

- AIIA governing unit
- AITISA governing unit
- Governing unit of the Information Technology and Application Innovation Committee

Strategic cooperation with 20+ research institutes

- Peng Cheng Laboratory
- Shanghai Astronomical Observatory
- National Natural Science Foundation of China

In-depth participation and technical cooperation in global industry associations

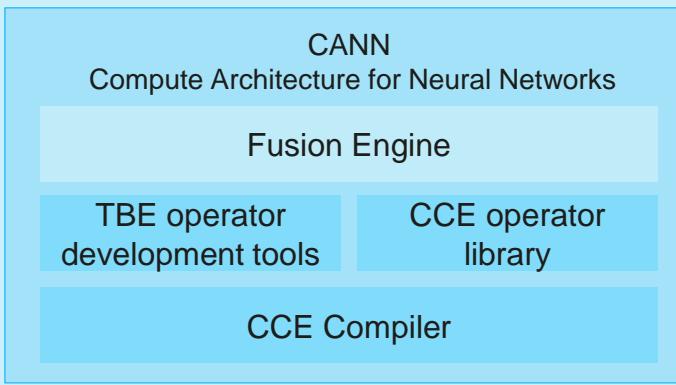
- National Smart City Standardization & Integration Team
- Smart Security Industry Association of Shenzhen
- China Intelligent Transportation Systems Association

Open the CANN Toolchain to Fully Unleash the Potential of Ascend

Fully Open CANN Toolchain

TBE: custom operator development tools, enabling optimal development efficiency and allowing operators to best match the performance of Ascend

CCE: optimal performance and programming efficiency for all scenarios



Full Promotion of Operator Development

Initiate the operator development plan and cooperate with universities, research institutes, and partners to fully unlock the potential of Ascend processors

Q4 2019

Tool release

Q1 2020

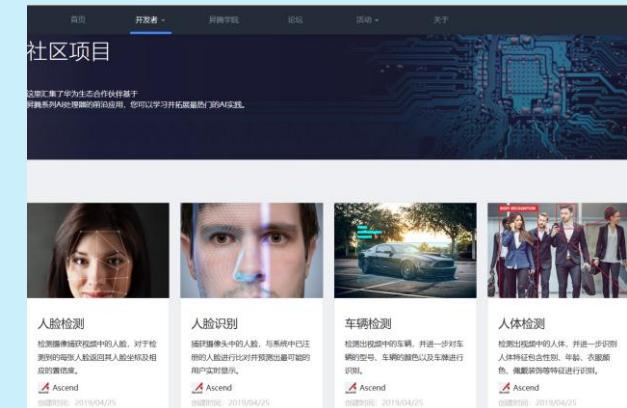
Operator Development Contest

Fully Open Mature Operators & Models

Open-source mature operators



Open network models in the developer community



Work with Partners to Develop Scenario-Specific Solutions

Work with partners to develop 20+ joint solutions in 2019. The annual plan is 50+.



Create a joint solution display platform to demonstrate certification solutions.



In 2020, Huawei will further enhance partners' rights and provide 1000+ training cards.

Sample Device Support

- 1000+ AI training cards
- 1000+ Developer Kits
- Cloud training service

Joint Innovation Technical Support

- Onsite technical support
- Special technical training
- Global OpenLabs

Develop 1M Developers Based on the Ascend Developer Community



Developer-Centric enablement platform
<https://ascend.huawei.com/home>

Annual technical salon



- Held in **10+** cities
- 20+ senior trainers
- Dozens of speeches

Developer Challenge



- **1500+** teams
- Over CNY1 million annual prize
- Equal opportunities for enterprises and universities

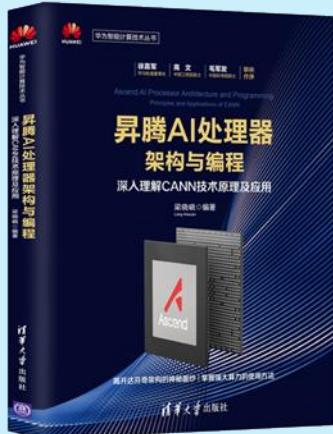
Developer support



- Public cloud vouchers
- Free certification course tickets
- Free Atlas Developer Kits

Huawei Developer Program for AI Talent, Covering 100 Colleges and Universities

Joint university curriculum development & lab building



- New generation of AI textbooks of the Department of Higher Education
- Authored by professor Liang Xiaoyao, Shanghai Jiao Tong University
- Published by Tsinghua University Press

Teaching & Training

Joint development of teaching materials
Practice platform construction

University Contest

University developer contest
Incubation and innovation

Research & Joint Innovation

Scientific research cooperation
Industry-education joint innovation

Joint curriculum development with 30+ universities in 2019

First universities for joint course development



Shanghai Jiao Tong University



Zhejiang University



Tsinghua University



Peking University



Tianjin University



Renmin University of China



Beijing University of Posts and Telecommunications



Huazhong University of Science and Technology



Southern University of Science and Technology



Xidian University

Cooperation with Training Partners

Tech Education

Open Education

Simple Education

Xunfang Technology

Strengthen Cooperation with Universities and Support the Implementation of Advanced Technologies with Ascend

The first 15+ partner universities have achieved many innovations in robotics, machine vision, and image processing.



Fundus retinal vessel segmentation

Humanoid

Thmart

Protein subcellular localization prediction

Super-resolution image algorithm based on deep learning

Age recognition

Intelligent robot cluster for desert restoration

Handwritten Chinese character recognition

Single-image facial expression animation based on image warping and GAN

Expand cooperation and build Ascend into an innovative platform for AI research in universities

● Natural language processing

Natural language generation

Information extraction

Knowledge management

Machine translation

Natural language expression

Speech recognition

Conversational system

Q&A system

● Graphics

3D body segmentation

3D reconstruction

● Medical image

Segmentation of tissue and organs

Lesion location identification

Atlas Products: Built on Ascend 310 and Serving Many Industries

50+ Atlas industry solutions



Smart City



Finance



Electric power



Transportation



Internet



Carriers

Safe city

Smart branch

Unmanned
inspection of high-
voltage lines

Free flow at
province toll
stations

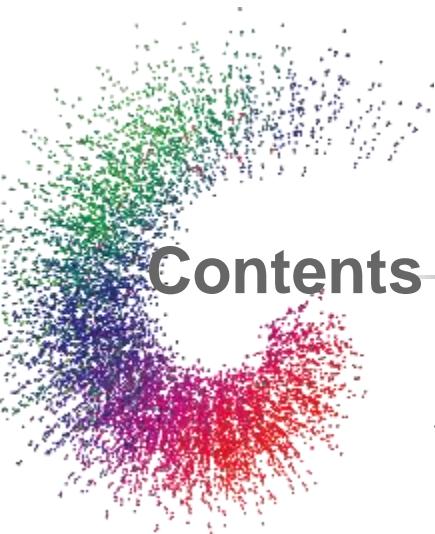
Intelligent
recommendations

Smart customer
service center

...

Joint Innovation in Segmented Scenarios to Facilitate Intelligent Transformation of Industries

Smart City	Transportation	Finance	Electric power
 Facial recognition  Video structurization	 Vehicle identification  Highway toll collection	 Smart branch  OCR	 Line monitoring  Device status monitoring
 云从科技 CLOUDWALK  以萨技术 YISA  格灵深瞳 DEEPCINT  流媒体科技 Seemo	 以萨技术 YISA  流媒体科技 Seemo  中控·SUPCON	 高重科技 GRANDHONOR  合合	 金三立 SANTACHI  智洋创新 ZHY INNOVATION  众芯汉创 Alcorehotress  亚联发展 Asia Link Technology  四方四方  铁越电气 TEAMWAY ELECTRIC
Carriers	Internet	Governments and corporations	
 Family cloud album  AI training platform	 New retail  Smart community Ads recommendation	 Archive identification  Logistics & new retail	
 中移互联网 China Mobile Internet  中国移动 CMRI 中国移动研究院	 爱华盈通 Aiwin  中国电信 CHINA TELECOM	 JD AI 京东人工智能	 汇纳科技 WINNER  华宇 THUNISOFT



Contents

AI Development Trend

Products and Solutions

Ecosystem

Success Cases

Atlas Intelligent Computing Platform Recognition



Spearhead China's Next-Generation AI Open Innovation Platform

Two-Time Winner of the Best of Show Award at Interop Tokyo

Winner of Two Red Dot Awards 2019

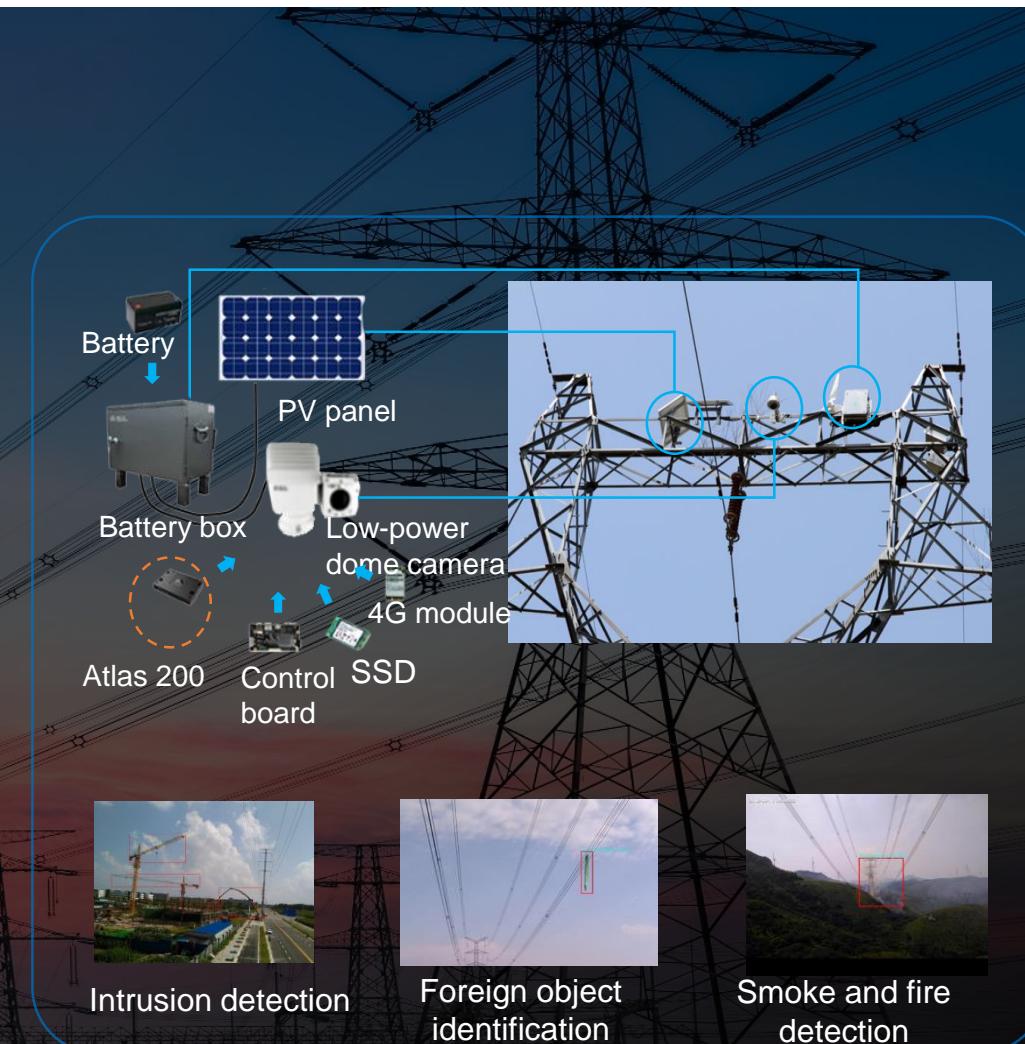
China International Industry Fair Award

World Leading Internet Scientific and Technological Achievements Award

Huawei Atlas Enables Intelligent Inspection of Power Transmission Lines

Authorized by this cust...

- The Atlas 200 AI accelerator module is embedded into traditional tower cameras. Atlas 200 uses AI inference algorithms to **analyze images and videos locally, and reports alarms in a timely manner, implementing unattended power line inspection.**
- The Atlas deep learning solution and Atlas 300 AI accelerator card are deployed at the master station to implement cloud-based training and inference, remotely deliver and deploy models using management software.
- Atlas 200 has a low-power design. The entire camera consumes 8 W, **runs on solar power**, and is maintenance-free throughout its lifecycle.



Thank you.

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

**Copyright©2019 Huawei Technologies Co., Ltd.
All Rights Reserved.**

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 an acceptance. Huawei may change the information at any time without notice.

