



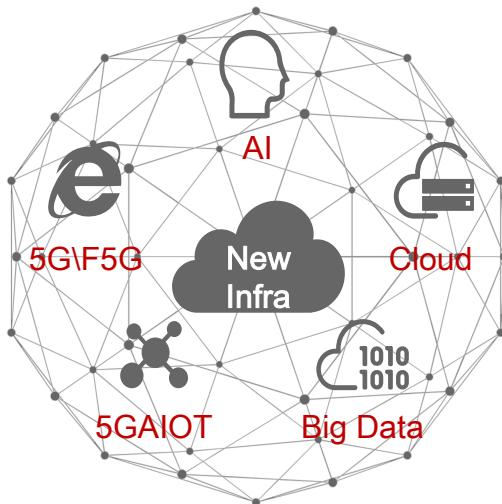
Indonesia National AI Platform Feasibility Study Report

- 1 Construction background**
- 2 As-Is Analysis**
- 3 Construction Requirements**
- 4 Construction Solution**
- 5 Operation Solution**
- 6 Budget and Plans**



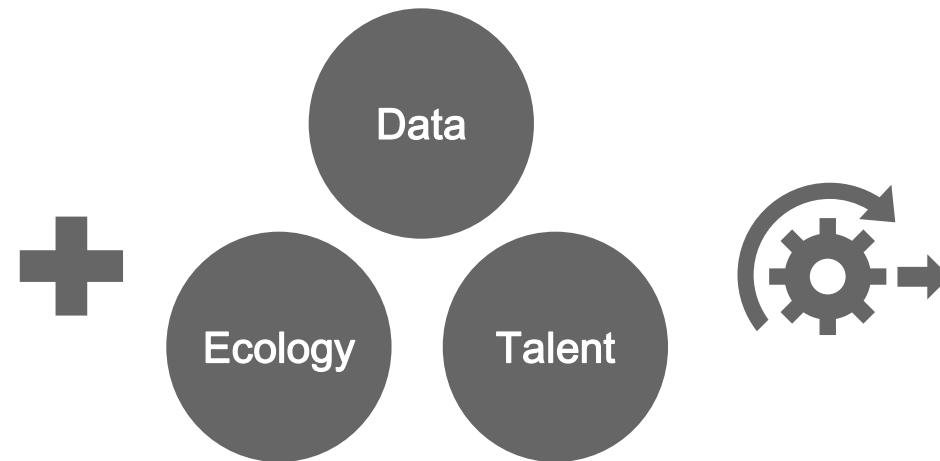
Background - 5G+Cloud+AI Acceleration Drives Digital Economy Transformation and Drives Rapid GDP Growth

5G+Cloud+AI Digital Economy Engine

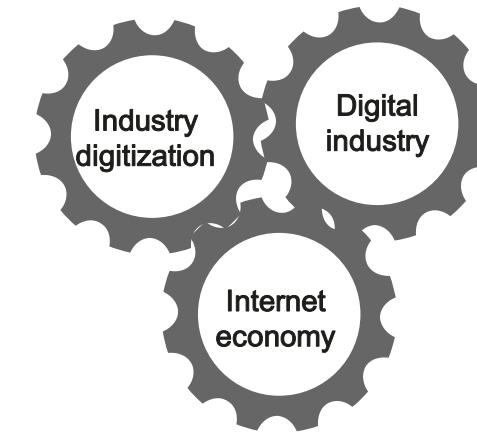


Digital economy production tools

Resource Elements for Digital Economy Transformation



Smart Digital Economy Drives GDP Growth



Digital Spillover USD 1 GDP USD 20



Digital economy changes the resource input and economic growth model, triggering the fourth industrial revolution.

Background - Many countries have released AI strategies, and AI has become the focus of global technological competition

More than 60 economies worldwide released AI strategies and increased investment in the AI industry



2020: AI become an important part of "New Infrastructure Construction" strategy
2019: Government Work Report Highlights "AI+X"
2019: AI Incorporated into Compulsory Curriculum for Primary and Secondary Schools
2018: AI Innovation Action Plan for Colleges and Universities – MOE
2017: Three-Year Action Plan for the Development of Next Generation AI Industry – MIIT
2017: New Generation AI Development Plan (State Council)
2016: Three-year Action Plan for "Internet+" AI



2021: The White House Launches the National Artificial Intelligence Initiative Office
2019: The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update
2019: The President signs the "National Artificial Intelligence Initiative"
2017: National Robotics Initiative 2.0
2016: The National Artificial Intelligence Research and Development Strategic Plan
2010: DARPA of the US began to prepare for the application of AI in various fields



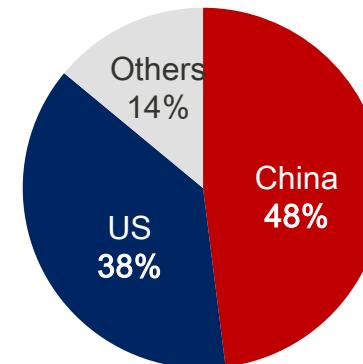
2017: Japan "Super Intelligent Society" (Society5.0), "Artificial Technology Strategy"
2015: New Robot Strategy



2021: European Union-Coordinated Plan on Artificial Intelligence 2021
2019: Germany announced the "The Industrial Strategy 2030" to the world
2018: France-The French National Strategy on Artificial Intelligence
2018: European Union-Release Coordinated Plan on Artificial Intelligence
2018: Germany-The Federal Government's Artificial Intelligence (AI) strategy
2013: European Union-Human Brain Project (HBP)

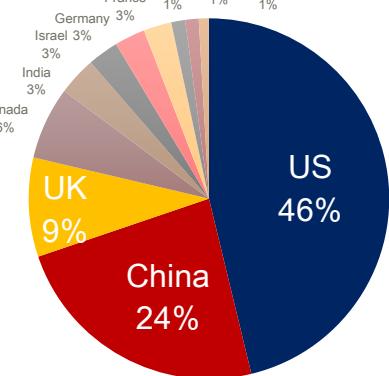
China and US are leading the AI industry

Share of Global AI Startup Financing



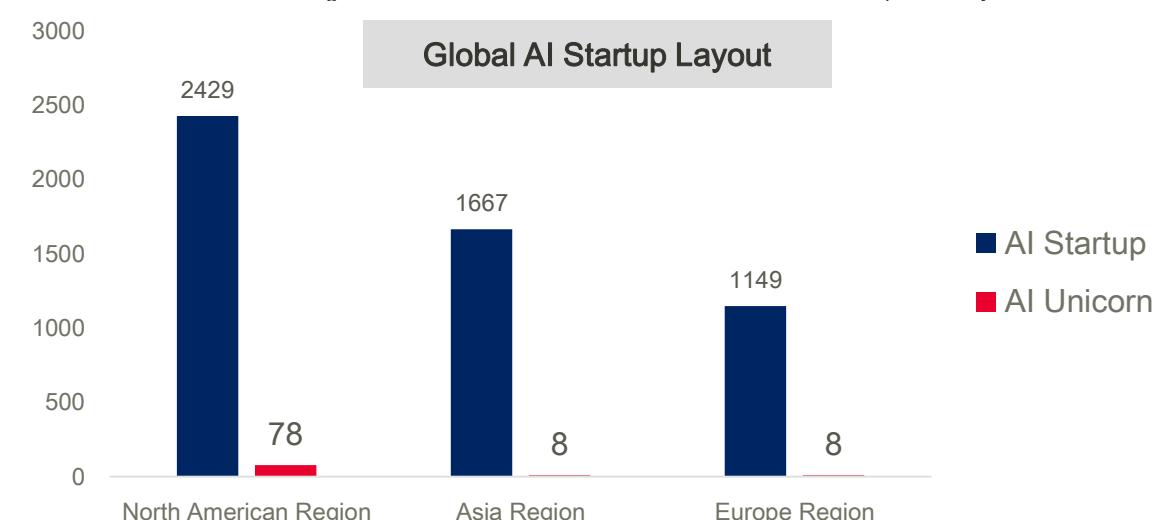
Source: 2018 Artificial Intelligence Investment

Share of Global AI Entity



Source: Global AI Enterprise Analysis 2018

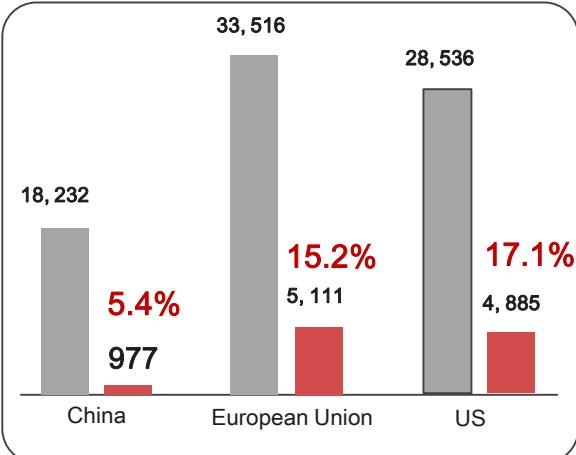
Global AI Startup Layout



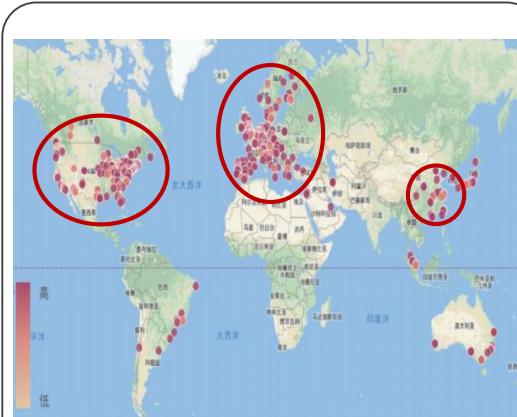
Source: The market status and competitive landscape of the global artificial intelligence industry in 2020

Background - AI talent is the key to industry development. Global top AI talent is mainly distributed in China, the US and Europe

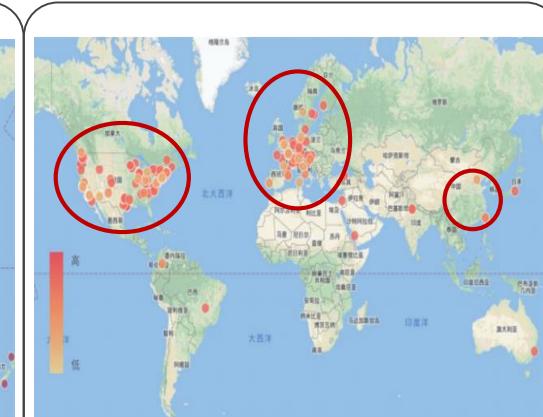
Number of people engaged in the AI industry



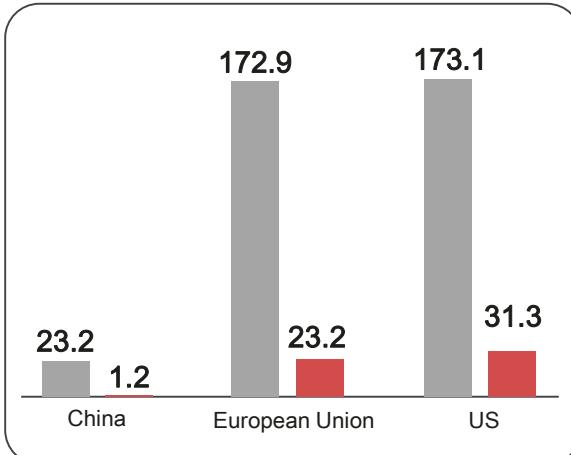
Global Top AI Talent Distribution



Talent Distribution of Global Top AI Companies



AI industry number per million population

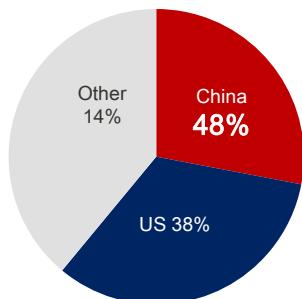


■ AI Talents

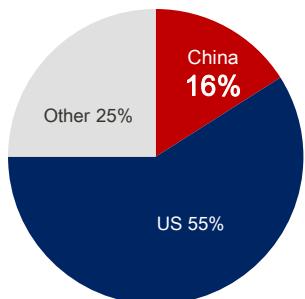
■ AI advanced talent

Global AI Paper Comparison, China Leads in Quantity and US Leads in Quality

Top 1% Papers in the AI Field



Top 100 highly cited articles in the AI field



China and the US are leading the world in terms of the number of top papers

- Compared with the top 1% of the 4130 papers cited in the AI field, China and the US are far ahead of other countries with 1166 and 1345 papers, accounting for 60.8% of the total. China and the US are leading the world in terms of theoretical basis and scientific research achievements
- The top 100 papers cited by the US accounted for 55, China accounted for 16 papers, and other countries accounted for 25 papers. The leading quality of these papers reflects the leading position of the US in scientific research and AI talent cultivation

Basic Background: AI Propels Transformation Across All Industries



Manufacturing

①②③

Quality inspection

①②③

Factory monitoring

Robots

Transportation

①②③④

Free-flow tolling

②④

Highway video cloud

Electric power

①②③④

Line inspection

①②③

Site monitoring

Factory monitoring

Safety supervision

Finance

①②③④⑤⑥

Ticket recognition

①②④

Smart branches

Smart customer

service

Smart risk controls

Internet

①②③④⑤⑥⑦

Content analysis

①②④⑤

Ad recommendations

⑦

Smart customer

service

Search and retrieval

Carriers

①②④⑤⑥

Smart customer service

⑤⑥

Smart branches

AI cloud services

Bill recognition



AI readiness for all industries

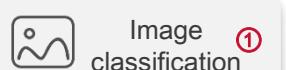


Image classification ①



Object detection ②

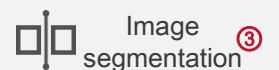
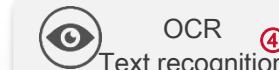


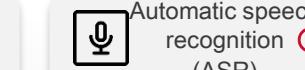
Image segmentation ③



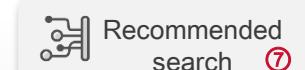
OCR Text recognition ④



Semantic comprehension ⑤



Automatic speech recognition (ASR) ⑥

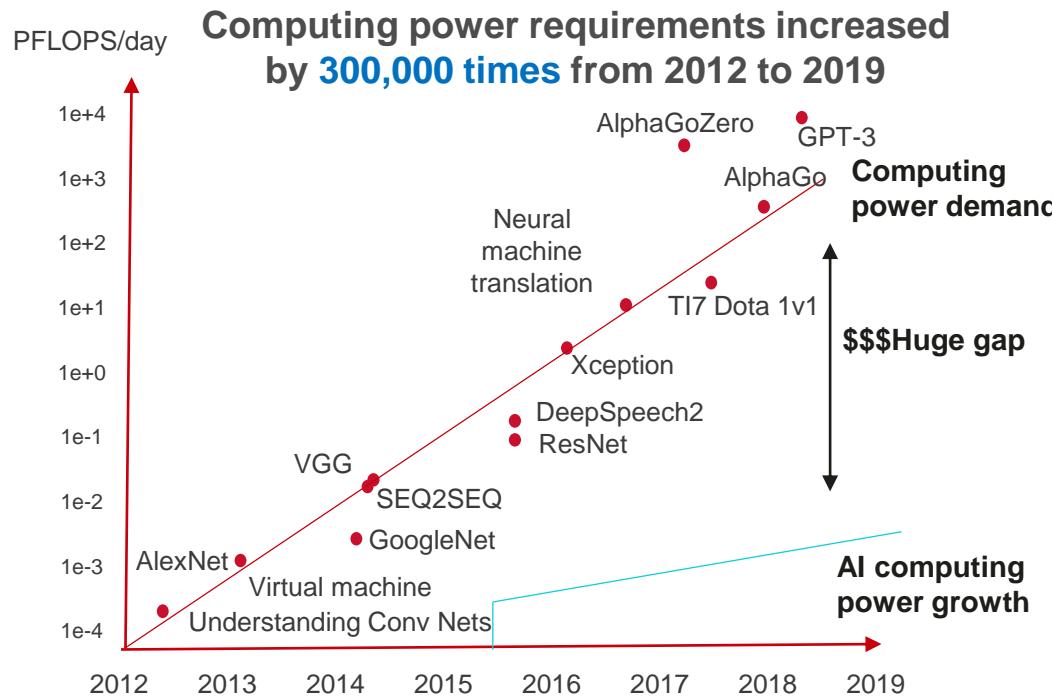


Recommended search ⑦

Basic Background: Computing Power Becomes the Basic Productivity of AI Enterprises

Expensive computing power hinders AI innovation and development

The computing power required for AI model training doubles **every 100 days** on average



AI model training is costly

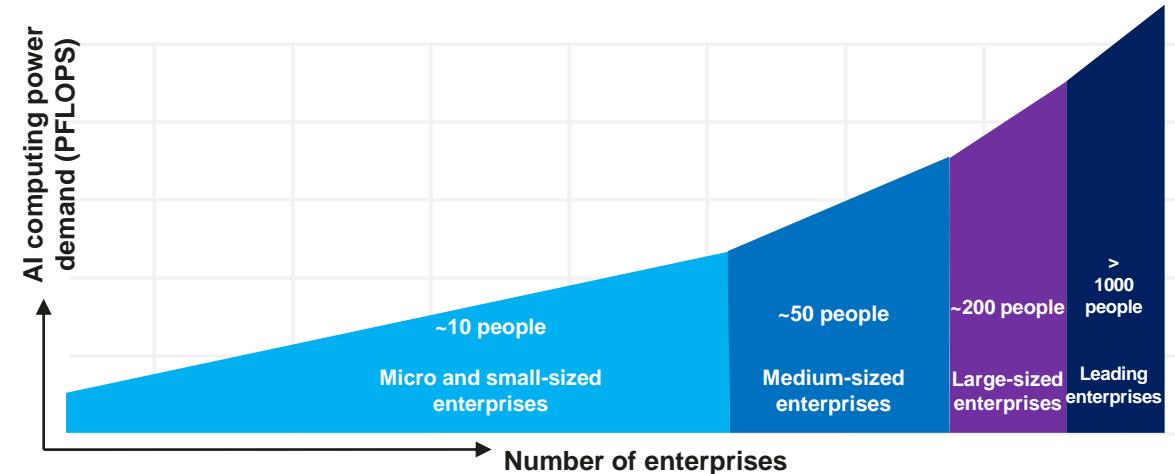
GPT-3

USD4.6 million

BERT

USD15,000 on average

Four stages for AI enterprises to mature, with high proportions of computing cost



Enterprise Type	Micro and Small-Sized	Medium-Sized	Large-Sized	Leading
AI Technology Phase	Nascent	With core AI products	Self-built AI	Combines basic AI and application AI
Number of Employees	10	50	200	>1000
Computing Power Demand (~PFLOPS)	1	5	30	200
Computing Power Acquisition Cost (USD1M/year)	0.06	0.5	5.4	35.7
Acquisition Method	Desktops + consumption cards	Computing power leased from clouds	Self-built or leased	Self-built
Percentage of Computing Power Expenditure	18%	25%	17.5%	14%

Basic Background: Indonesia incorporates AI in its new national strategy

Indonesia AI Strategy Formation Path



2019.6.28

In 2019, the G20 put **forward AI Principles** for the first time, and the Indonesian government subsequently formulated five recommendations for AI development.

2019.11.15

After being re-elected, President Joko proposed the "**Indonesia 2045**" vision, which defined the digital economy (including AI) as one of the five major policies.

2019.11.28

Indonesia's president has ordered government agencies to eliminate two civil service ranks by 2020 and replace them with AI in a bid to reduce the "red ribbon" that hampers investment.

2020.8.10

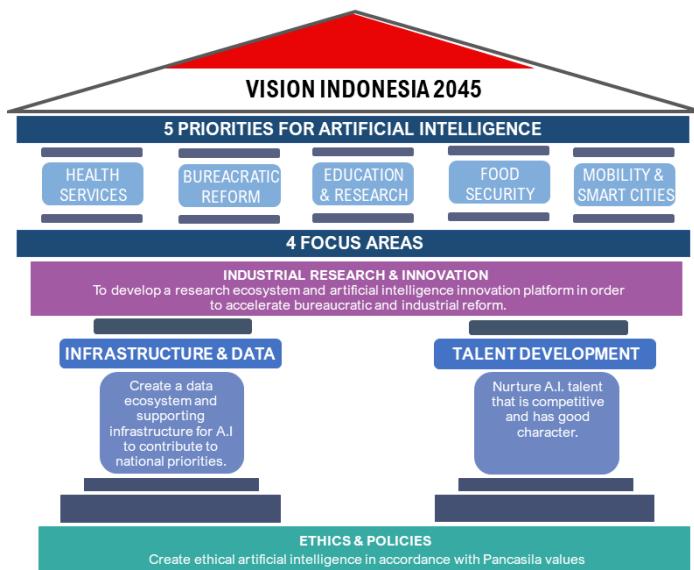
At the 25th Indonesian National Technology Day, the Vice President and the Minister of Scientific Research announced the launch of **AI Indonesia National Strategy**.

2020.11.14

President Joko delivered a speech at Indonesia's first **national AI summit**. This is Indonesia's largest and highest-profile AI summit to date.

2021.3.8

Indonesia's president stressed that AI competition is like the Cold War. Countries that control AI can rule the world, requiring **BPPT to be the driving force for AI development** in Indonesia.



Indonesia AI Strategic Framework

- One vision:** The AI national strategy supports the implementation of the "**Indonesia 2045**" strategy.
- Five priorities:** Health services, Bureaucratic reform, Education & Research, Food safety, and Mobility & smart cities are the top five application areas of AI.
- A platform:** technology-to-industry evolution and transition through industry research and innovation.
- Four Areas:** Supporting the development of the five priority industries are **Information Infrastructure & Data, Talent Development, Industry Research & Innovation, Ethics & Policy**.
- A base:** Creating AI that meets its Indonesian core values – **PANCASILA**

Basic Background: Some Focused Areas on AI Development in Indonesia

Health services

- Big Data Medical Assistance
 - Track COVID-19 epidemic and provide alerts
 - Improve the digital capabilities of medical institutions.
 - Patient diversion and resource management in medical institutions
 - epidemic disease prevention and control
- Telemedicine and Digital-Assisted Diagnostics
- Establishing a Secure National Medical Data Information System
- Assisting gene mapping and improving therapeutic capabilities



Bureaucratic Reform

- Building an E-Government Online Service Robot Platform
- Establishment of internal government automation system
- Budget management system
- Biology authentication system for government insiders, such as face, voiceprint, and fingerprint
- Public opinion analysis
- Government Big Data Analytics



Education & Research

- Smart online education
- AR VR Auxiliary Course
- Vision Lab
- Adaptive Learning System and Evaluation System
- Intelligent Evaluation System for Students' Courses
- Gamified teaching curriculum
- Precision Learning System



Food Security & Agriculture

- Improve productivity and land utilization and optimize resource allocation.
- Financial Services Support for Low-Income Farmers
- Forecast of Agricultural deficits
- Food Inventory Data Analysis and Recommendations
- Disease Prevention in Agricultural Products
- End-to-end tracking of Agricultural products



Smart City

- AI-powered intelligent traffic management solutions
- Intelligent sewage and garbage treatment system
- Auxiliary Natural Disaster Risk Management System
- Urban Resident Information Management
- Building quality assessment and urban space management
- Assisting Urban Public Facilities Management



Smart Ocean

- Marine Life Identification and Statistics
- Marine pollution detection
- Marine pollutant detection
- Marine weather monitoring
- Ocean Law Enforcement Management
- Regulation of Maritime Activities



Basic Background: 2020 - 2024, Establishing an AI Infrastructure in AI Strategy

Content	Timeline	Key Tasks	Responsible Department
AI ethics building and information security management strengthening	2021	AI Ethics Authority established in 2021 (1004)	BPPT, LIPI
Legal Preparations for AI Strategies	2022	Establishment of the AI Ethics Commission in 2022	RITEK
	2024	Complete cybersecurity and resilience-related bill changes in 2024	BSSN
AI Policy Supervision	2022 - 2024	Establishment of various regulatory organizations and regulations between 2022 and 2024	Departments and ethics committees, working with universities and external organizations
Human resources development and AI education ecosystem building	2021 - 2024	Promote talent cultivation and certification, compile teaching materials	MOE, LIPI, industry enterprise
Building the capacity of national institutions to access data	2023	Complete the data sharing architecture to enable AI to obtain data	RITEK
Building a shared data platform for AI learning	2023	Establish an AI learning supervision mechanism.	Kominfo
	2023	IoT Data Connected Architecture	RISTEK, BPPT
	2020 - 2024	2020-2024: Continuously promote cloud computing infrastructure construction	Kominfo
	2023 - 2024	Establish national supercomputer	Kominfo, RISTEK
	2024	Establish National Artificial Intelligence Data Council	RISTEK, BPPT

Although Indonesia's demographic dividend and its "unique" data resources are one of the strengths of AI development in Indonesia, Indonesia has neither the relevant regulations nor the official authority to oversee AI development. Therefore, Indonesia's AI strategy focuses on infrastructure and data, research and industry innovation, ethics and policy, and talent development, through the establishment of a Data Ethics Committee to oversee its development and to set regulations and national standards for AI innovation. The release of the **Indonesia AI strategy** lays the foundation for AI development in Indonesia.

Project Participant Introduction - Kura Kura Island Bali (KKI)

Kura Kura Island Bali

- Kura Kura Island Bali Island is a pioneering development based on the Balinese concept of Tri Hita Karana – Three Ways to Happiness – the harmony of Creation, Nature and Culture in all aspects of life
- It is a vision of an Island of Happiness offering a sustainable ecosystem for a thriving international creative community
- Kura Kura Island Bali lies just 15 minutes' drive from Ngurah Rai International Airport, connected seamlessly to the main Bali island by a short bridge
- Kura Kura Island Bali development is under the Giti Group founded in the 1930s with interests in real estate, manufacturing, retail and development and natural resource development in the Asia Pacific Region



Project Participant Introduction - Badan Pengkajian dan Penerapan Teknologi (BPPT)

Badan Pengkajian dan Penerapan Teknologi (BPPT)

- BPPT is a non-ministerial government agency under the coordination of the Ministry for Research and Technology of Republic of Indonesia, which has the tasks of carrying out government duties in the field of assessment and application of technology
- President Joko Widodo (Jokowi) assigned BPPT as the center of Indonesia's artificial intelligence (AI)



Project Participant Introduction - Del Institute of Technology

Del Institute of Technology

- Del Institute of Technology is a private college established in the village of Sitoluama, the acquired Laguboti, Toba County, Northern Sumatra. It was founded by the Mr Luhut Binsar Panjaitan who is Coordinating Minister for Maritime Affairs and Investment. IT DEL started conducting academic activities in 2001 and it had already put down 15 class to date in September 2019



- 1 Construction background**
- 2 As-Is Analysis**
- 3 Construction Requirements**
- 4 Construction Solution**
- 5 Operation Solution**
- 6 Budget and Plans**



Four Challenges for AI Development in Indonesia

1

Data cannot be monetized



Internet data



Industry data



Government data

A large amount of digital application data accumulated in Indonesia **cannot be monetized**, hindering Indonesia's further digital development.

2

Slow growth of AI enterprises

AI-hungry SMEs

Face difficulty and high costs
in accessing AI computing
power

More than 70% of Indonesian
AI Startup still use Consumer
PC or Bitcoin Mining Machine,
No AI unicorns emerge

3

Limit AI talent development

<300

Indonesia new AI
Talents per annum

Cultivating high-end resources requires massive computing power.

4

High barriers to research

GPT-3

● 175 billion parameters
USD4.6 million per training

BERT

● 300 million parameters
USD15,000 per training



Indonesia still can not do such
research for such top AI model,
will **increase Indonesia Digital
Divide**



4%

AI penetration in all industries

Indonesia AI Development Status: Small-Scale AI Vendors Focus on Application and Technology Layers



- ✓ All local AI vendors in Indonesia focus on the **technology layer and application layer**. Many start-ups are in the **natural language processing area**, because local vendors are more competitive than foreign vendor on this area.
- ✓ **The development of local AI vendors in Indonesia is slow**. Most local AI vendors have not received round B financing and lack sufficient technology, capital, and industry ecosystem aggregation support.
- ✓ Local AI vendors in Indonesia **lack computing infrastructure**, and most of them use PCs with consumer GPUs or Bitcoin Mining Machine.

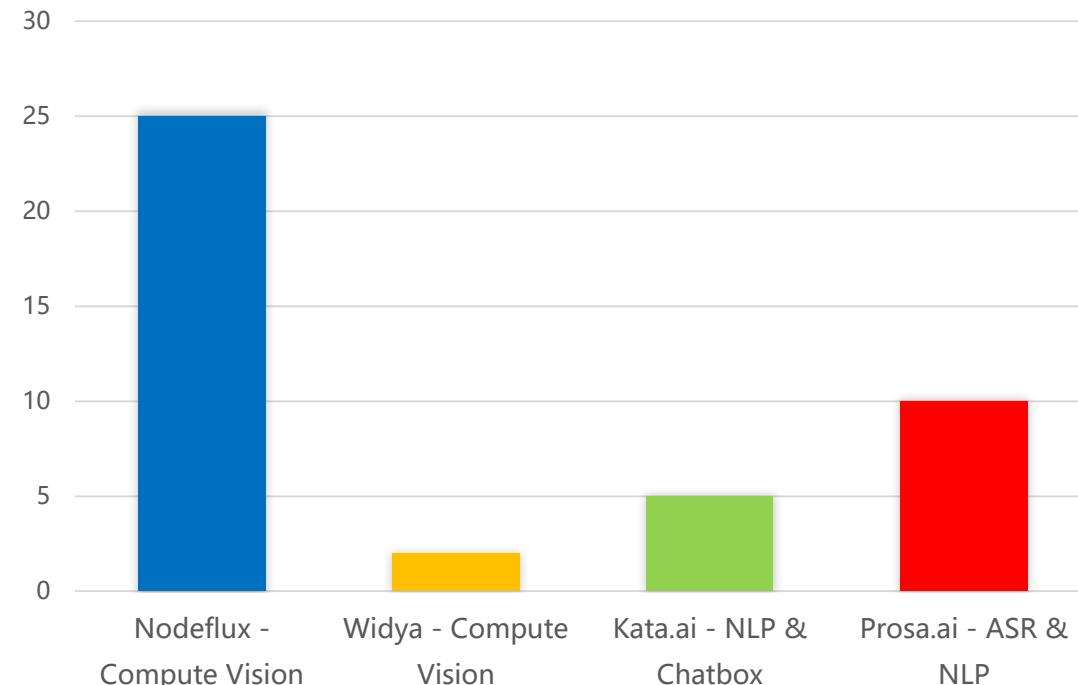
AI Development Status in Indonesia: Training Computing Power is Severely Insufficient

AI Computing Capacity in Some Research Institutions and Universities in Indonesia

institutions	CPU computing power (core)	Memory (GB)	GPU computing power (core)		Computing Capability Evaluation
			Cuda	Tensor	
BPPT/BJIK	480	4,964	0	0	NA
BPPT/PTIK	400	3,328	3,584X2	640X12	0.02 PFLOPS
LIPI(PUI HPC)	2,648	13,500	3,584X10	0	0.1 PFLOPS
UI	240	1,184	91,392	7,296	0.4 PFLOPS
ITB	2,712	13,992			NA
UGM	220	868	61, 440	7, 620	0.2 PFLOPS
ITS	596	1, 896			NA
Malang	100	1, 690			NA
Binus	40	512	3,584X8		0.08PFLOPS

Remarks: NVIDIA V100 5120 cuda cores/GPU, P100 3849 cuda cores/GPU

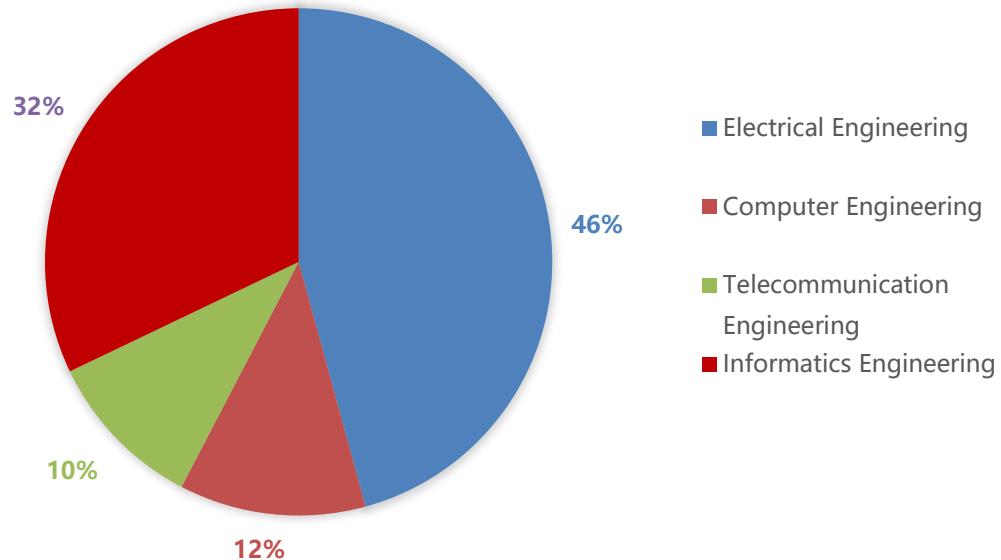
Computing Power Requirements for Indonesia Typical AI Startup (PFLOPS)



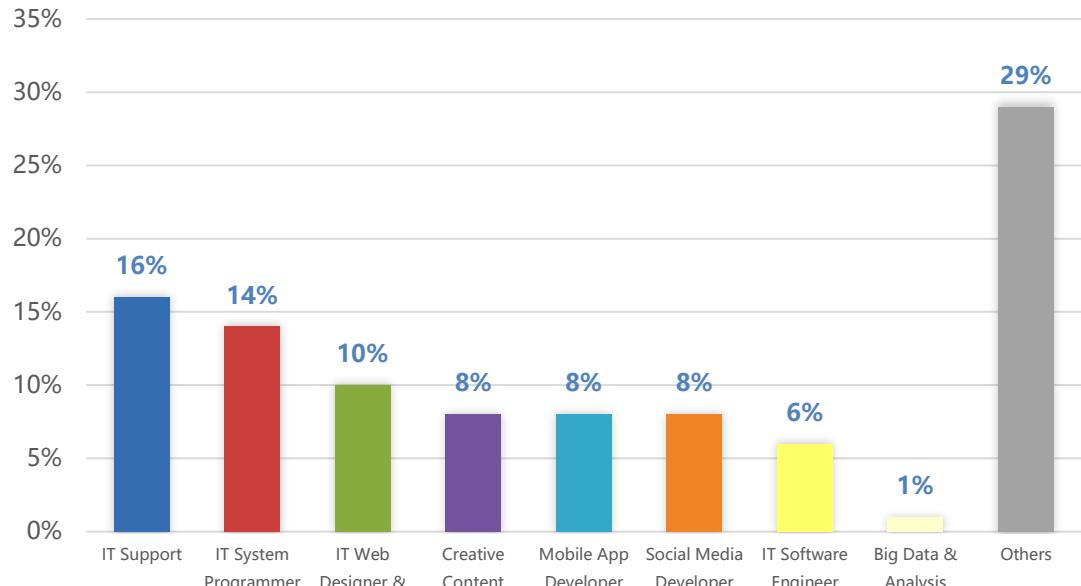
The AI computing power deployment in Indonesian research institutions and universities is seriously insufficient, which does not match the computing power requirements of the latest AI model training and has a big gap for Indonesia startup computing requirement, hindering the development of basic AI research in Indonesia.

AI Development Status in Indonesia: Insufficient AI Talents

DISTRIBUTION OF ICT STUDENTS MAJORING IN INDONESIAN
TOP 30 UNIVERSITIES



DISTRIBUTION OF IT-RELATED PROFESSIONAL EMPLOYMENT IN
INDONESIA

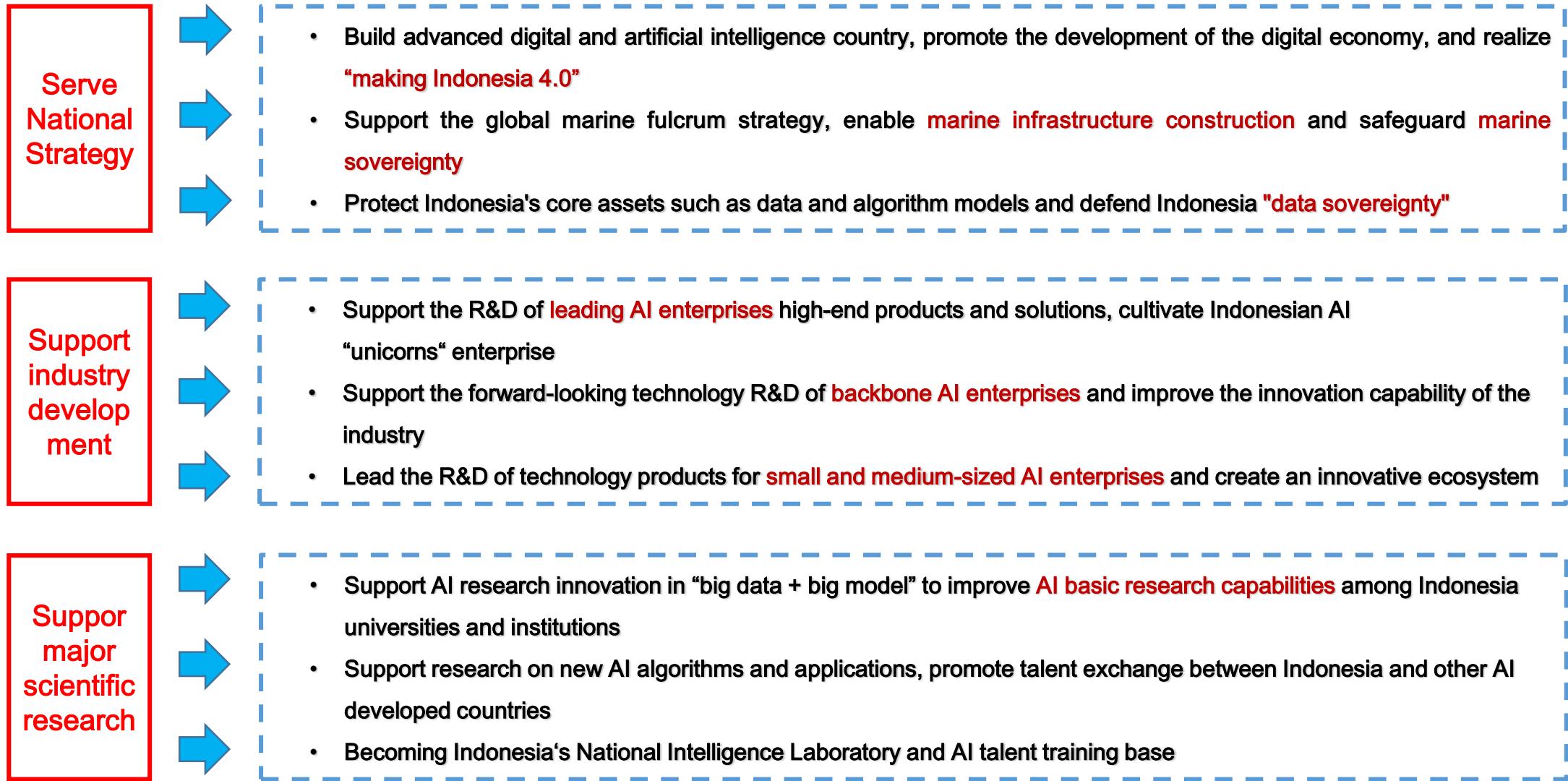


According to APTIKOM (Indonesian Association for Higher Education in Informatics and Computing), there are 40,000 to 50,000 ICT graduates in Indonesia every year. According to the statistics of ICT majors in Indonesian top 30 universities, about 44% of the total number of graduates in IT software-related majors each year. Further more, according to the distribution of IT employment in Indonesia, the number of people engaged in AI-related professions (big data and analytics) accounts for about 1% of the total number of employees employed in IT professions. **It can be inferred that there are about 200 to 250 new AI talents in Indonesia every year**, and there are very few AI talents.

- 1 Construction background**
- 2 As-Is Analysis**
- 3 Construction Requirements**
- 4 Construction Solution**
- 5 Operation Solution**
- 6 Budget and Plans**



Overall Requirements



Platform Construction Mode



Public AI Computing

- Provide inclusive public computing services for local enterprises, reducing initial investment costs.

Application Innovation

- Provides easy-to-use AI services and supporting technical support to implement platform-based enablement.



Government
of Indonesia
Investment
Entity



BPPT Operating
Subsidiaries

Business
Operation Entity

Kura Kura Bali



AI Industry Campus



Institut Teknologi Del
AI Research Entity

Industry Aggregation Development

- Aggregate various companies in the AI industry chain, to promote the transformation and upgrade of traditional industries to form a complete industry chain.

Research and Talent

- Build "AI Craftsman"
- Comprehensive improvement of AI literacy

Typical Service Scenarios



Smart
Agriculture



Smart Ocean



Smart
Healthcare

Government Agencies and Public Services : Smart agriculture, Smart ocean, and Smart healthcare

Typical Service Scenarios



Local enterprises: R&D of high-end products, high-value applications, and innovation ecosystem

Typical Service Scenarios

MLOps, low-code or no-code cluster learning, network-scale content analysis, simulated empathy and emotion, machine vision, massive translation systems, artificial emotional intelligence, digital twins, robotic process automation, autonomous driving, industrial manufacturing, scientific computing, facial recognition intelligent networks



Research institutions and universities: Automatic machine learning, Cross-media super-large-scale model pre-training, Digital twins, Robotic process automation, Autonomous driving, and Massive translation systems

Typical Service Scenarios

National AI Platform Enable Center for Future Knowledge

Computing Power Support

20P FLOPS Super Computing for Science Research

A Cloud Service Platform of high accessibility, security, and efficiency

DevOps Platform for Research and Innovation



Tri Hita Karana Center for Future Knowledge

Focuses on national strategic, futuristic, and fundamental technology research

HIDC Subcenter

DCR Subcenter

AMB Subcenter

ME Subcenter

LHN Subcenter

GES Subcenter

WFL Subcenter

HDX Subcenter

ECU Subcenter

EMS Subcenter

Data Support

Aggregate Public Sector Datasets

Powerful Big Data governance ability

A hub for AI and digital resources and talents across Indonesia

Big Data Action Lab



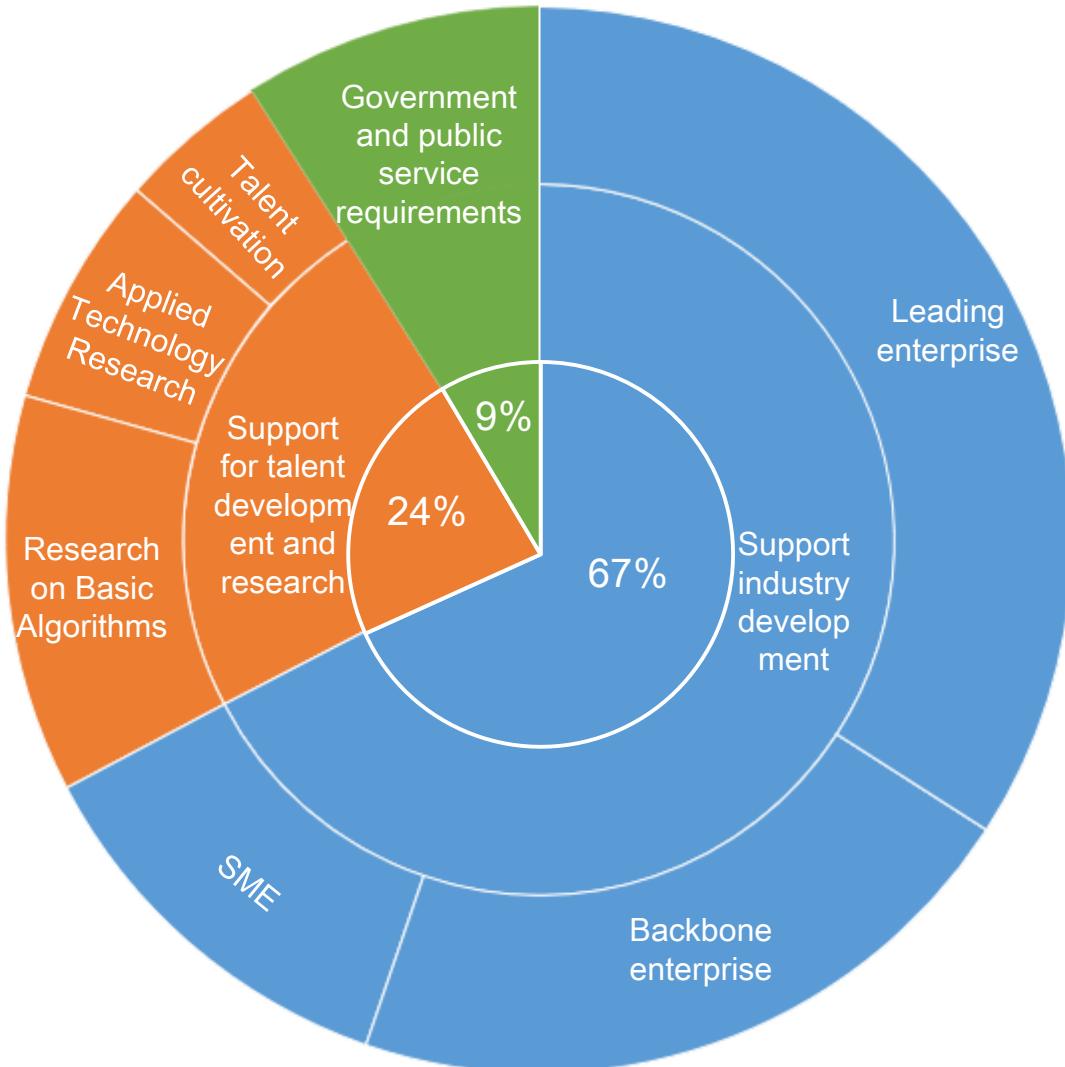
Computational
Science Lab

A critical AI facility

Indonesia National AI Platform

Computing power and storage requirements

Potential user demand statistics

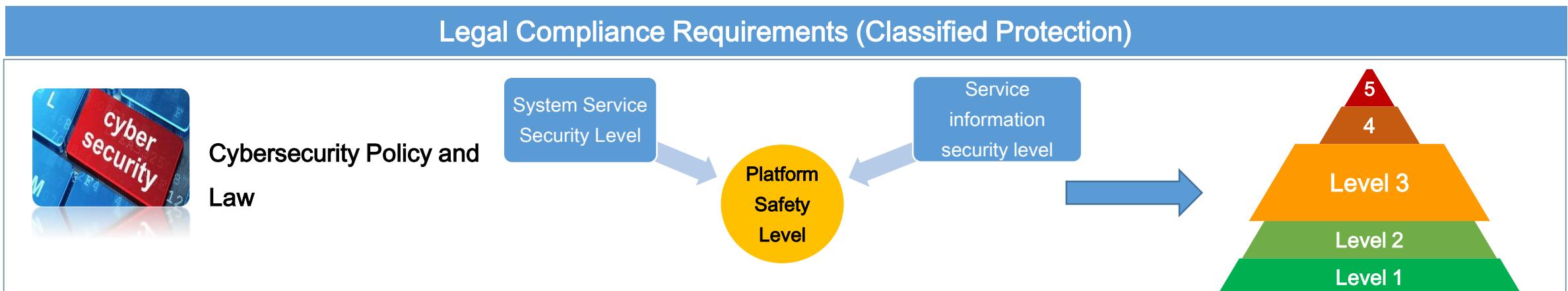


- **Industry requirements:** **14.7P FLOPS** for continuous computing and **1.5PB** storage
- **Research requirements:** **5.3P FLOPS** for continuous computing and **0.5PB** storage
- **Government and public service requirements:** **2P FLOPS** for continuous computing and **0.5PB** for storage

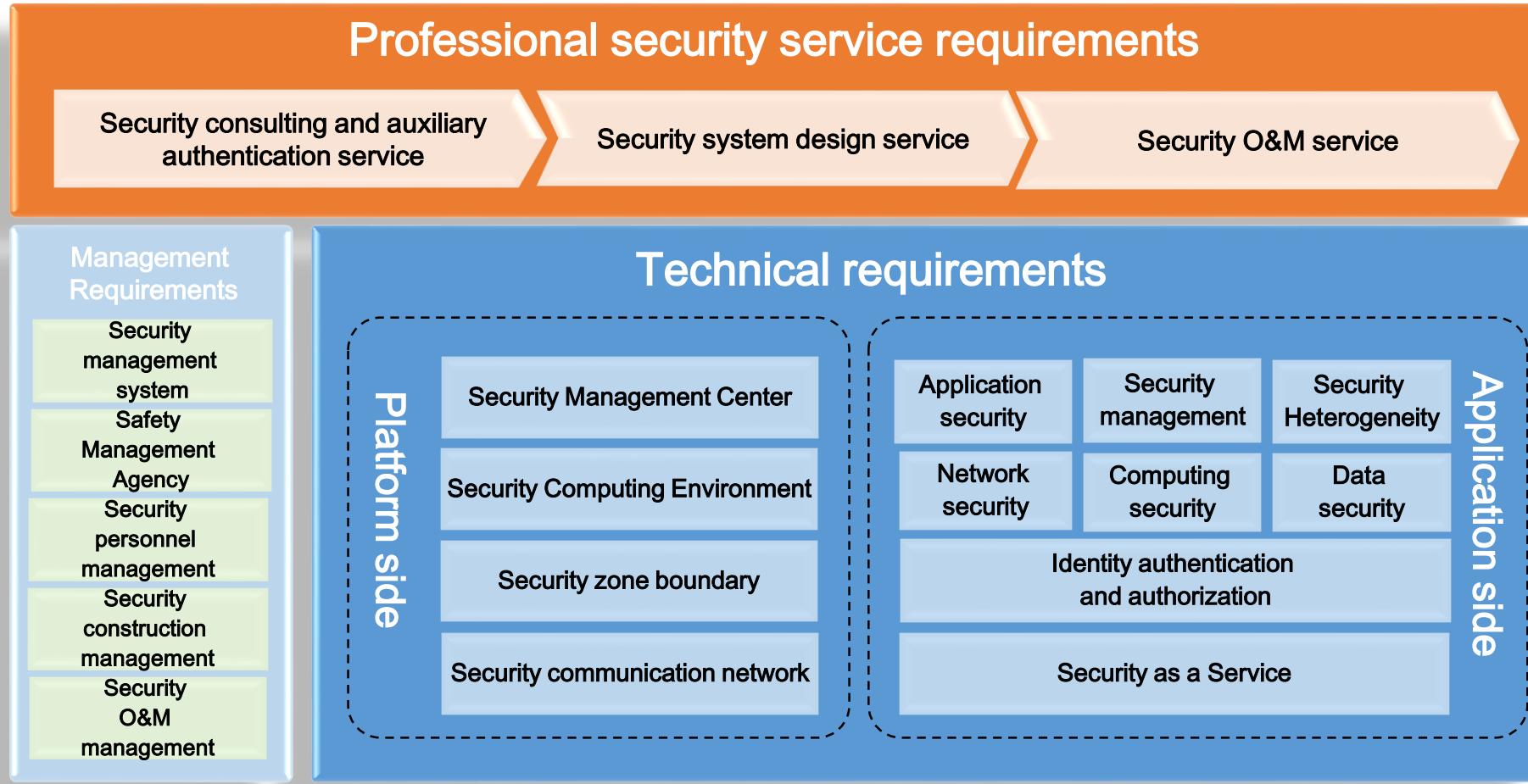
The total computing power requirement is about **22P FLOPS**, and the short-term storage requirement is **3PB**

Security Risks and Compliance Requirements

Security Threats to be Prevented	Key Security Requirements
<ul style="list-style-type: none">• Data Leakage• Inadequate identity, credentials, and access management• Insecure interfaces and application programming interfaces (APIs)• System flaw• Account hijacking• Malicious insiders <ul style="list-style-type: none">• Advanced Persistent Threat (APT)• Data loss• Inadequate due diligence• Abuse and Malicious Use of Cloud Services• Denial of Service (DoS)• Shared Technical Vulnerabilities	<p>Service continuity</p> <ul style="list-style-type: none">• Network attack prevention• Anti-hacker intrusion• Legal compliance and compliance <p>Manageable O&M process</p> <ul style="list-style-type: none">• Configuring Security Policies• Risk identification and handling• Auditable and traceable operations <p>Data confidentiality</p> <ul style="list-style-type: none">• Anti-external theft• Invisible to internal unauthorized employees• Invisible to cloud service providers



Security Protection Requirements



Management requirements:
avoid and reduce **security accidents**

Technical requirements:
Improve the security protection capabilities of networks, borders, and computing and provide a unified security management center

Overall Requirements for Equipment Room Construction

项目总图



xxxx

1# One-story computer room building (building area xxxm²)

Overall Requirements for Equipment Room Construction

Construction grade

- The GB50174-2017 Code for Design of Data Centers meets the A-class equipment room standard
- Refer to the Tier-III standard in ANSI/TIA942-A-2017, and consider the possibility of reserving a later level upgrade

Parameters and indicators

Equipment room layout	Liquid cooling equipment room (two layers)
Cabinet Type	Liquid-cooled cabinet
Number of cabinets	4
Power density (kW/R)	45
market capacity	Ultra-high power of a single cabinet, total IT power consumption is about xx MW, and total market capacitance is about xx MVA.
Cooling capacity	Liquid cooling equipment room area: xx KW of wind and liquid, xx KW of cooling board Cloud equipment room: xxRT*2-wind fluid xxkw
PUE	PUE≤1.25



- 1 Construction background**
- 2 As-Is Analysis**
- 3 Construction Requirements**
- 4 Construction Solution**
- 5 Operation Solution**
- 6 Budget and Plans**

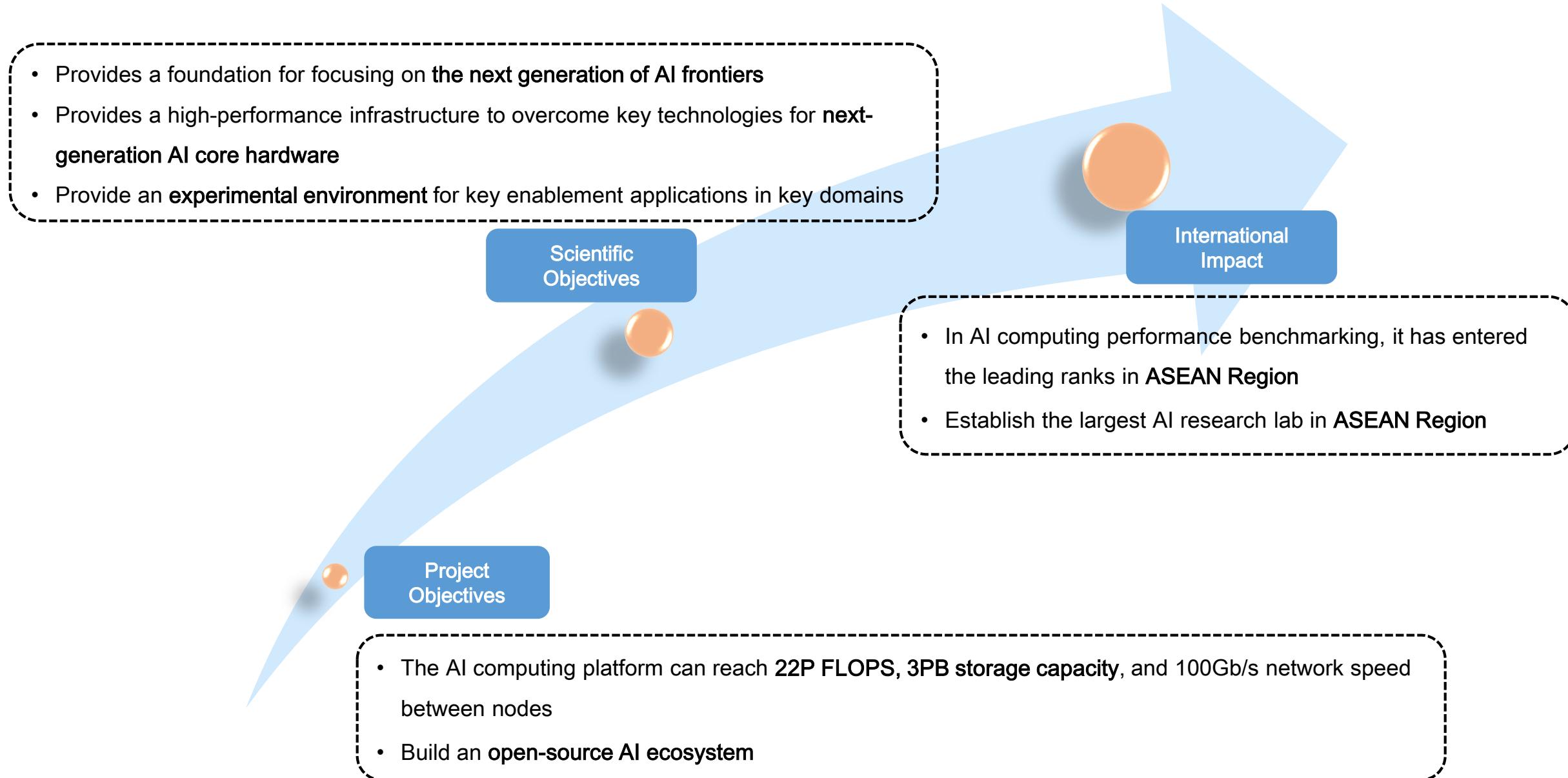


Construction Objectives

- Provides a foundation for focusing on the **next generation** of AI frontiers
- Provides a high-performance infrastructure to overcome key technologies for **next-generation AI core hardware**
- Provide an **experimental environment** for key enablement applications in key domains

Scientific
Objectives

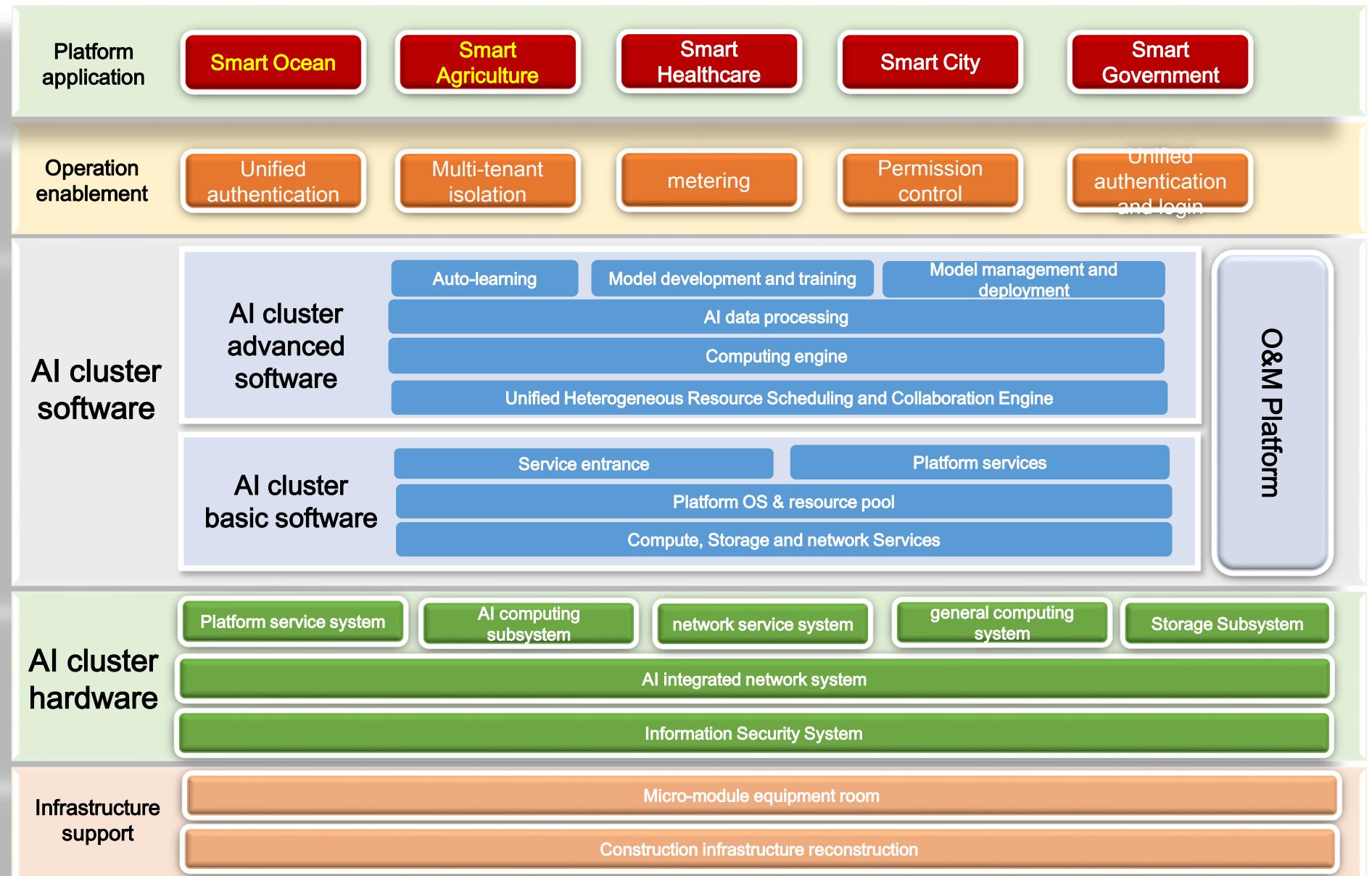
International
Impact



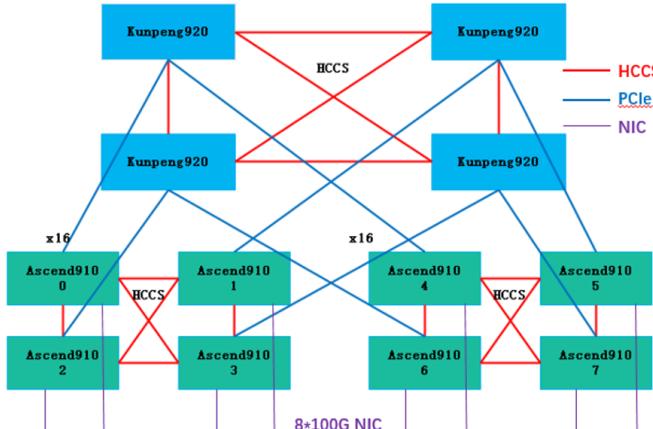
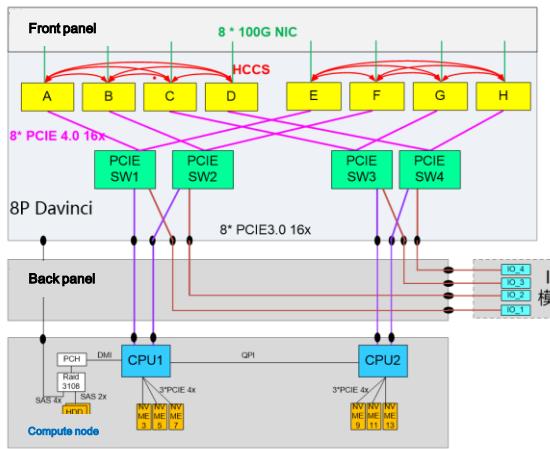
Overall Platform Architecture

Overall Platform Architecture

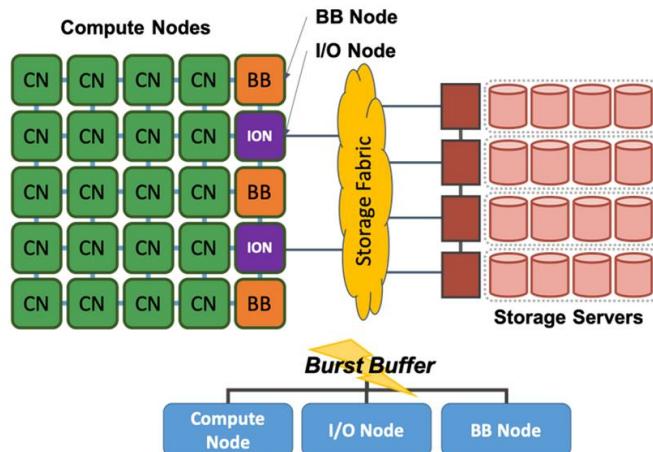
The overall main technical solution is divided into five levels: infrastructure support, AI cluster hardware, AI cluster basic software, AI cluster advanced software and Platform application



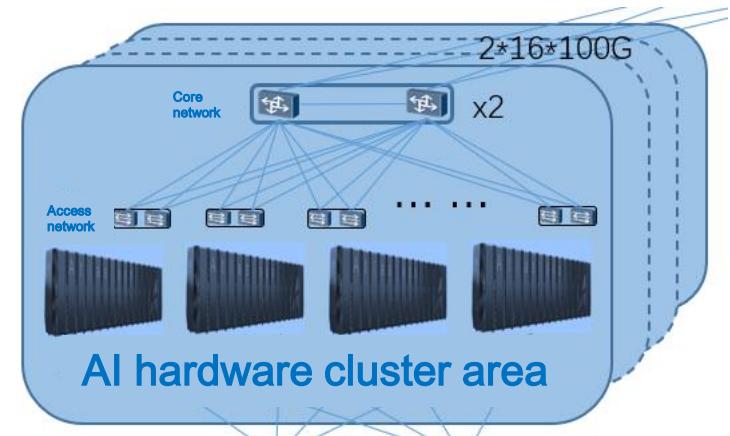
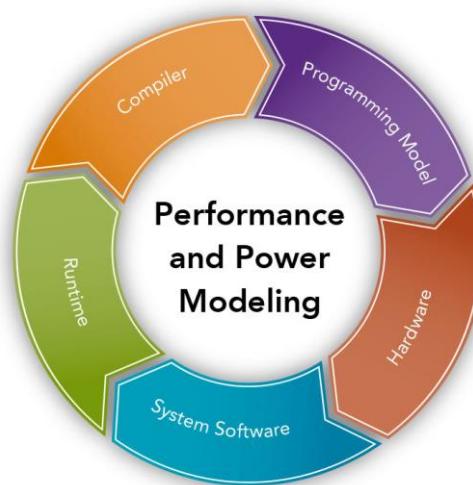
Principles for designing the cluster computing power balance solution



Enhanced chip interconnection topology within a node
2-socket x86 processor and 4-socket Kunpeng processor
Balance/Bandwidth/Latency Improvement



- I/O and Computational Balance: Burst Buffer Technology**
- ✓ Multi-layered storage is required to ensure performance
 - ✓ Burst Buffer is the most promising solution
 - ✓ Use non-volatile flash memory, etc. as high-speed data buffer
 - ✓ Between a regular storage system and a compute node
 - ✓ Provides temporary storage space for local data
 - ✓ Improves system aggregate IO bandwidth and greatly improves IOPS

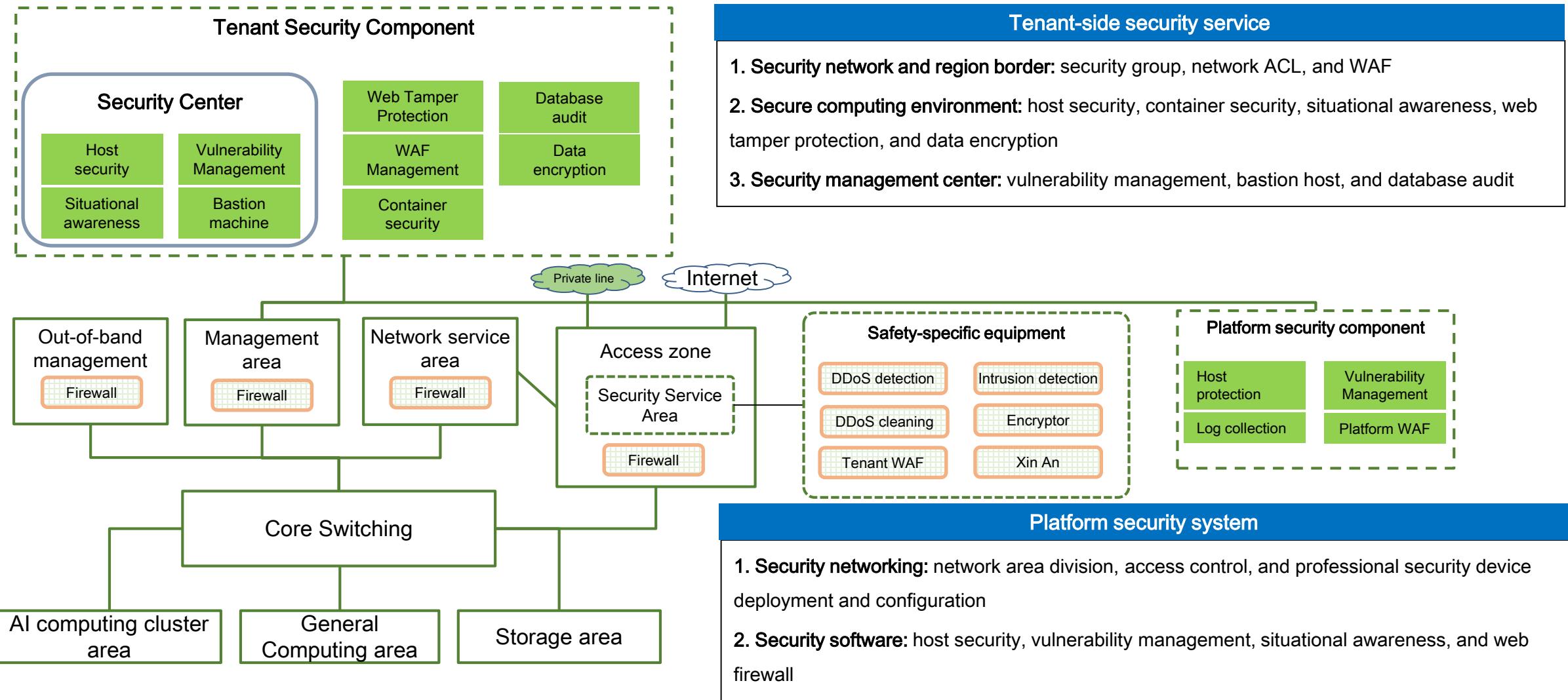


Enhanced network interconnection
Multiple 100 Gbit/s Network Designs
Independent networking of AI computing and general computing
Ensure all types of communication performance

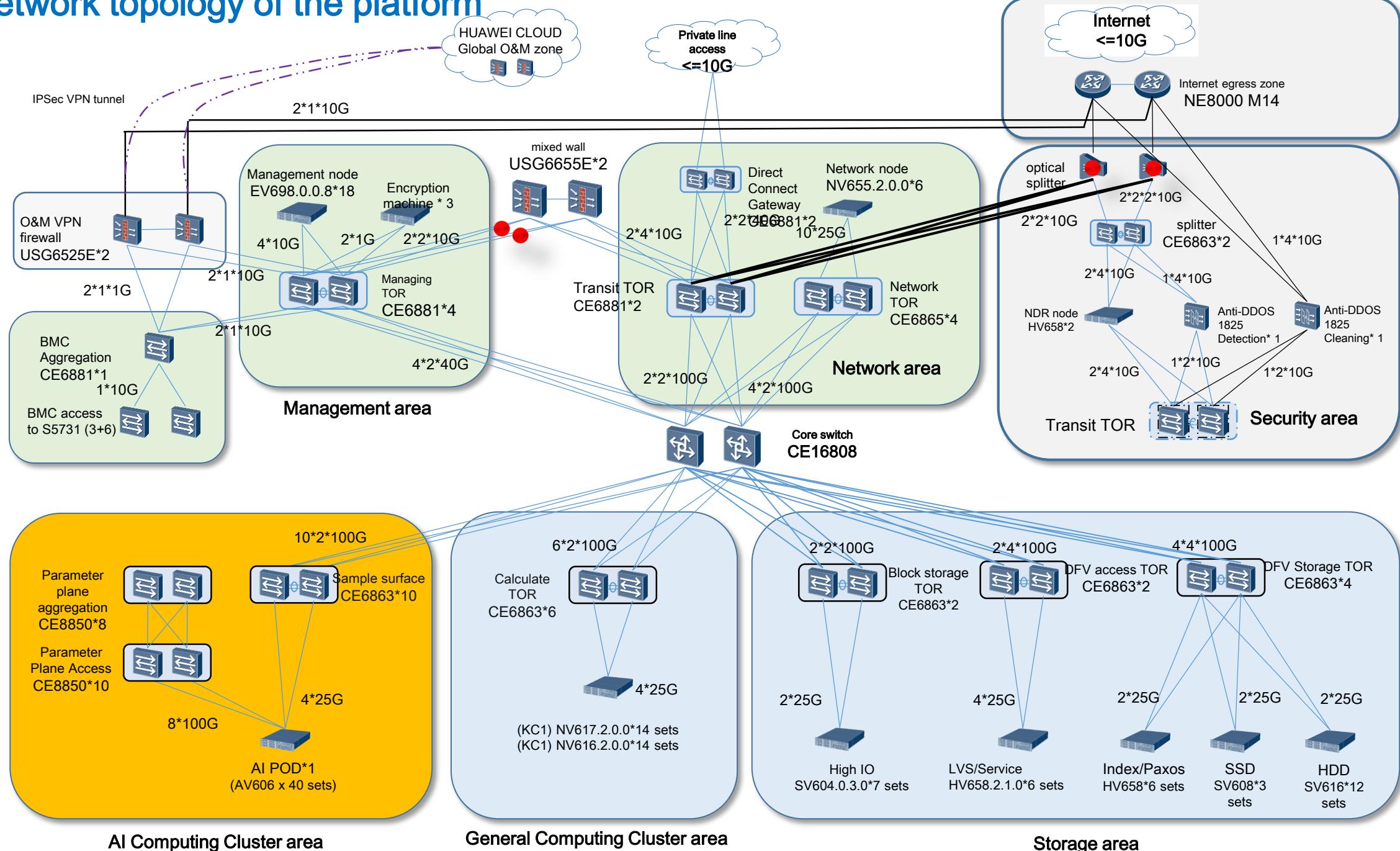
Co-Design for Software and Hardware Joint Design Optimization

- ✓ Multi-level optimization of integrated data framework, communication framework, and computing framework
- ✓ Extending the Adaptability of the System Architecture
- ✓ Adapts to a wider range of application types
- ✓ Improve system usage efficiency and scale of applications.

Security Solution Design Principles



Overall network topology of the platform



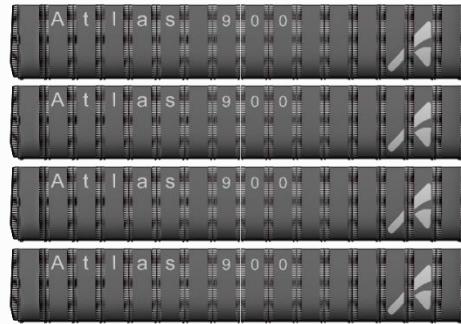
AI Cluster Hardware Solution

Atlas 900 AI Cluster : Best AI Scalability for the World's Fastest Training Cluster

Atlas 900 PoD is AI Cluster Basic Unit of Atlas 900



Up to **64** racks
Atlas 900 AI Cluster

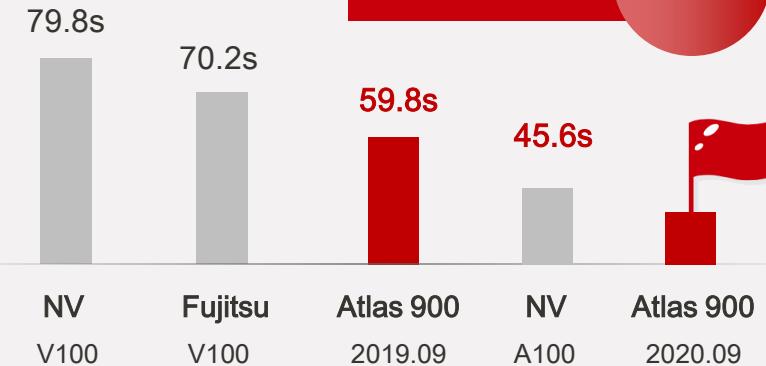


Atlas 900 PoD

- **Industry's highest computing density**
The 47U rack provides 20.48 PFLOPS FP16 powerful AI computing
- **Superior AI efficiency**
Up to 20.48 PFLOPS/43 kW ultra-high energy efficiency
- **Best AI scalability**
Supports scaling by basic units to an AI cluster of up to 4096 Ascend 910 processors, delivering up to 1.31 EFLOPS FP16

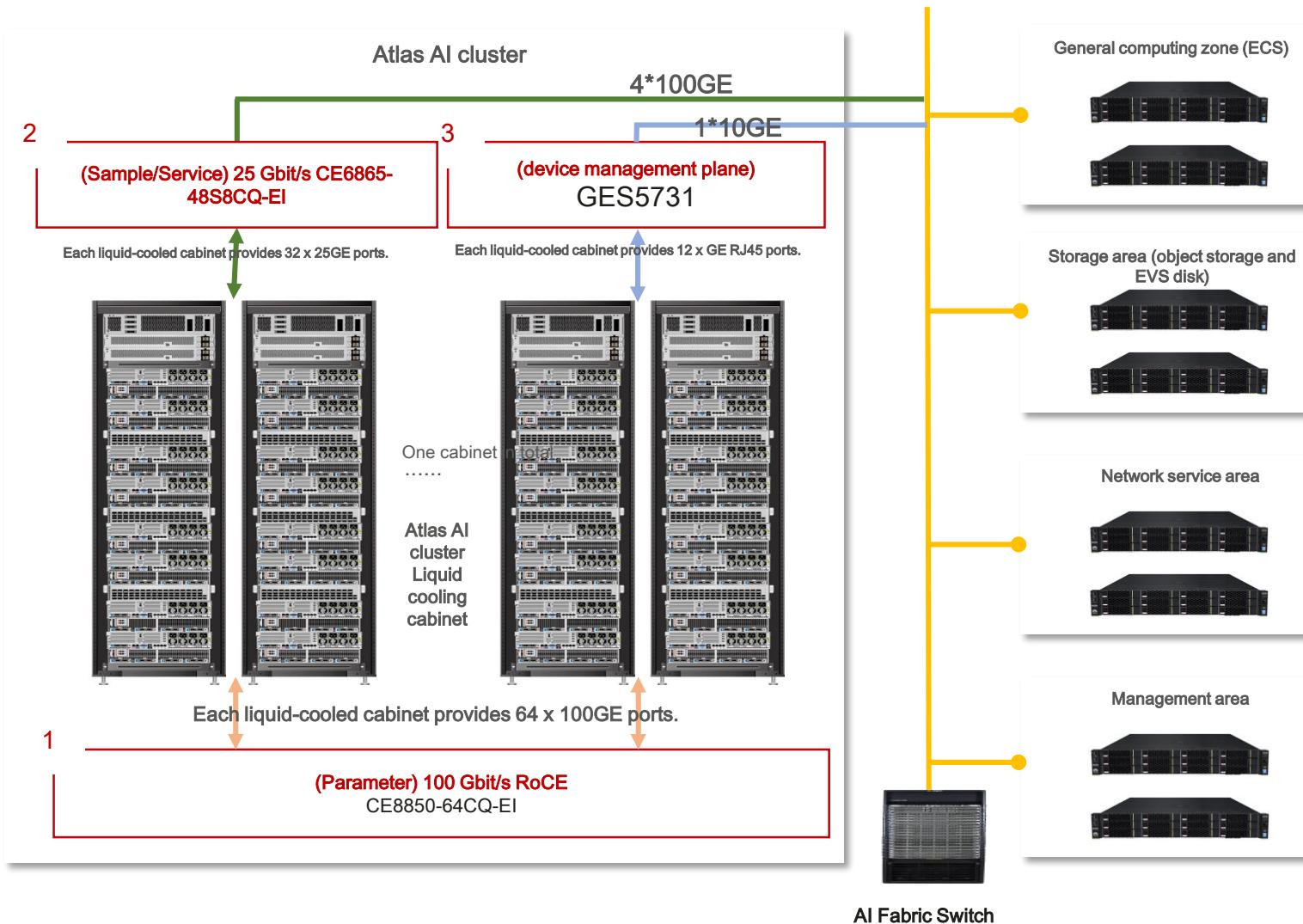
Atlas 900 AI Cluster

World's fastest AI training cluster



- Test benchmark: ResNet-50 V1.5 model, ImageNet-1k dataset
- Test time: September 2020

AI computing power cluster networking - 20P FLOPS computing power



[Atlas AI Cluster Liquid Cooling Cabinet Solution]

The AI cluster is delivered in high-density and integrated cabinet mode. Each liquid-cooled cabinet provides 20.48P FLOPS AI computing power, with a total power of 46 kW.

[Network Solution Design]

The AI cluster network consists of three planes: 100G parameter plane, 25G sample/service plane, and GE device management plane.

① Parameter plane: Each liquid-cooled cabinet provides 64 100GE parameter plane ports, which are connected through two access switches in the cabinet and converged through aggregation switch outside the cabinet, implementing high-bandwidth and low-latency cluster communication for AI training tasks

② Sample plane/service plane: Each liquid-cooled cabinet provides 32 25GE sample plane/service plane ports, which are accessed through access switches. These ports are mainly used to access high-speed and large-bandwidth interconnection storage systems and service scheduling and management

③ Device management plane: Each device provides one GE network, which is used for device management and operation

Atlas AI Cluster liquid-cooled whole cabinet design

AI Cluster Liquid Cooling Cabinet Design

47U		Power box	47U
46U			46U
45U			45U
44U		假面板2	44U
43U		假面板2	43U
42U		假面板2	42U
41U		理线架	41U
40U		管理面S5731交换机	40U
39U			39U
38U		Atlas800-9000	38U
37U			37U
36U			36U
35U			35U
34U		Atlas800-9000	34U
33U			33U
32U			32U
31U		CE8850	31U
30U			30U
29U			29U
28U		Atlas800-9000	28U
27U			27U
26U			26U
25U			25U
24U		Atlas800-9000	24U
23U			23U
22U			22U
21U		CE6865 业务面	21U
20U		CE6865 业务面	20U
19U			19U
18U			18U
17U		Atlas800-9000	17U
16U			16U
15U			15U
14U		Atlas800-9000	14U
13U			13U
12U			12U
11U		CE8850	11U
10U			10U
9U			9U
8U		Atlas800-9000	8U
7U			7U
6U			6U
5U			5U
4U			4U
3U		Atlas800-9000	3U
2U			2U
1U		假面板	1U

Strongest computing power

A cabinet provides a maximum of 20.48P FLOPS@FP16 AI computing power.
Maximum power consumption: 43 kW

One-stop cabinet delivery

Built-in switch, intra-cabinet access, fixed networking

- Parameter plane access switch CE8850-64CQ-EI
- Sample plane/service plane access switch CE6865-48S8CQ-EI
- BMC management plane access switch S5731-H48T4XC

Full liquid cooling for the ultimate energy efficiency



Board-level liquid cooling dissipates 70% heat



Rack-scale enclosed adiabatic design



dissipates 30% heat by air-to-liquid heat exchange

- 200 test items and 2000-hour strict test, ensuring high reliability for liquid cooling
- High-performance fan modules and VC heat sinks combined with field synergistic heat exchangers to improve heat dissipation efficiency by 10%
- Supports 50°C high-temperature water inlet (30°C by industry counterparts), improving cooling efficiency
- Real-time leakage detection and quick automatic shutdown

About Ascend 910

AI development trend: Model scale/calculator demand growth far exceeds Moore's law, and computing demand is increasing with each passing day

- OpenAI analysis of the most advanced models requires doubling the computing power every 3.4 months
- Model Parallelism Brings More Challenges to Cluster Training
- The expansion of model size is an important driver of the surge in demand for computing power



Future AI processors need to meet the preceding development trends

- Improve the chip dominant frequency and AI computing capability of a single chip.
- Improved model parallel capability
- HBM Bandwidth Upgrade
- Improve energy efficiency ratio



Maximum computing power
256 TFLOPS FP16 Powerful

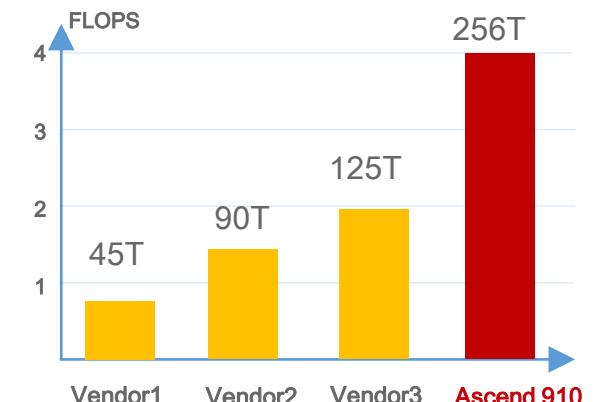


High integration
32 Huawei DaVinci AI Cores + 16 TaiShan cores + 100GE RoCE v2
NICs, all-in-one AI computing, general computing, and I/O



High-speed connection
Integrates three high-speed interfaces: HCCS, PCIe 4.0, and 100G
RoCE, improving training data and gradient synchronization efficiency.

- Architecture: Da Vinci
- Power calculation: 256 TFLOPS FP16
- Maximum power consumption: 310 W



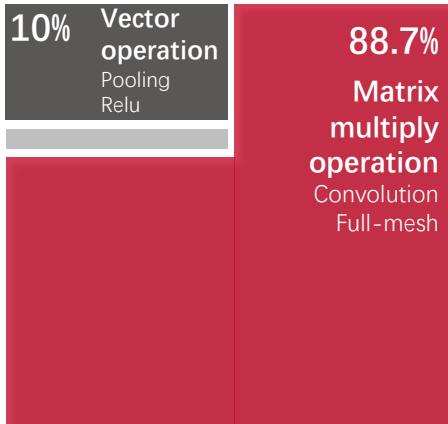
* Normalized to 16-bit

Based on Huawei Da Vinci Chip Architecture, Unlocks the Ultimate AI Computing Power

AI computing characteristics

~90% are matrix multiply operation

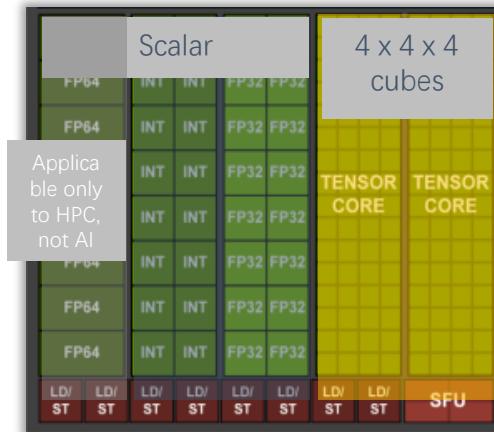
- AI computing is mainly based on the convolutional neural network (CNN) model.
- About 90% of the CNN model is based on matrix multiply operations.
- The **cube computing unit** is the most suitable.



Da Vinci architecture: best-fit for AI computing

GPU

- The GPU is not designed for AI computing. Therefore, it is inefficient in matrix multiply computing.

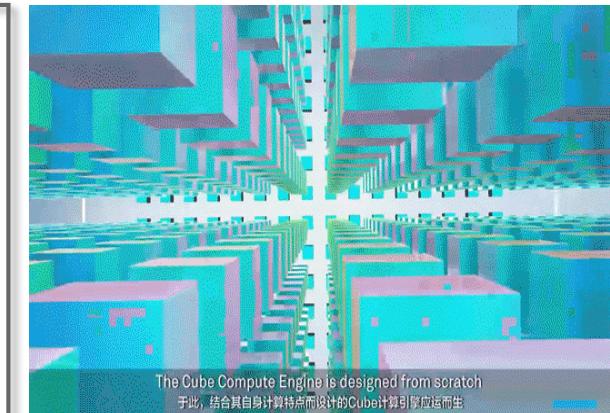
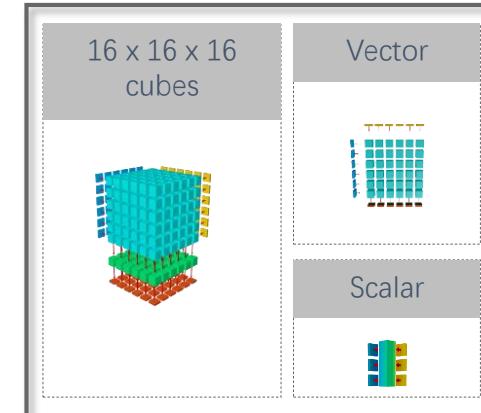


2x area-to-efficiency ratio

On the same area, the Da Vinci architecture delivers 2x computing power.

NPU

- The Da Vinci architecture is specially designed for AI computing.
- Provides cube, vector, and scalar computing units for AI computing
- The Da Vinci architecture has a large proportion of cubes, enabling **high matrix multiply computing efficiency, and optimal area-to-efficiency ratio**.



	Other AI Chip Architectures	Da Vinci Architecture
Chip area (12nm)	5.x mm ²	13.x mm ²
Computing power	1.7 TOPS FP16	8 TOPS FP16
Area-to-efficiency ratio	~0.3	~0.6

Atlas 900 AI Clusters Lead in Power, Performance, and Linearity

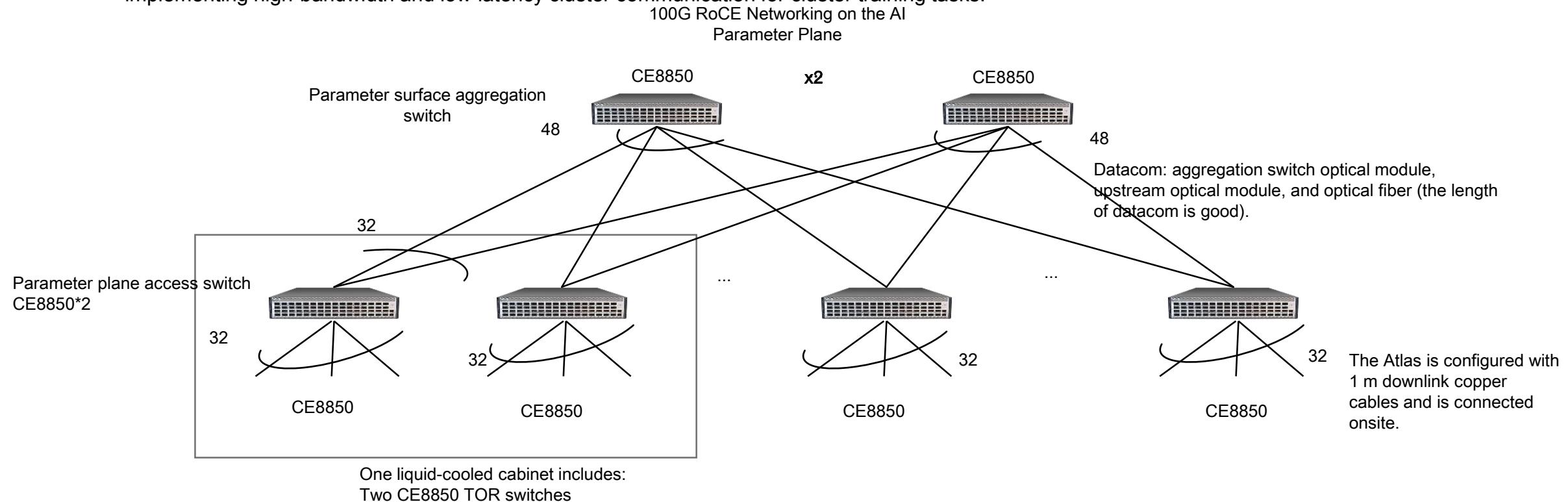
Higher computing power than Vendor 1				Better performance, 2.2 times that of Vendor1			
Calculate power@FP16				Resnet50 Network Performance			
Vendor1 Ascend 910				BERT-LARGE Network Performance			
Number of chips	Vendor1	Atlas 900 Compute node	Relative performance	Number of chips	Vendor1	Atlas 900 Compute node	Relative performance
8	8675	14060	1.6x	8	445	861	1.9x
16	13191	27731	2.1x	16	800	1643	2.1x
32	24302	54150	2.2x	32	1360	3050	2.2x

Higher linearity, 1.5 times higher than vendor 1				Higher computing power density: 30%+ less cabinets with the same computing power			
Number of Cluster Chips	Vendor1	Atlas 900 Compute node	relative linearity	Model	Single-cabinet computing power	1024P computing power Number of cabinets	Difference in the number of cabinets
8	85%	95%	1.1x	Atlas 900 Basic unit	16P	64	/
16	65%	94%	1.4x	Vendor1	6P	172	Multi 108 cabinets
32	60%	92%	1.5x	Vendor2	12.5P	84	More than 20 cabinets
1024	< 60%	85%	>1.5x				

Network topology of the 100G parameter plane of the AI cluster

AI cluster parameter plane network:

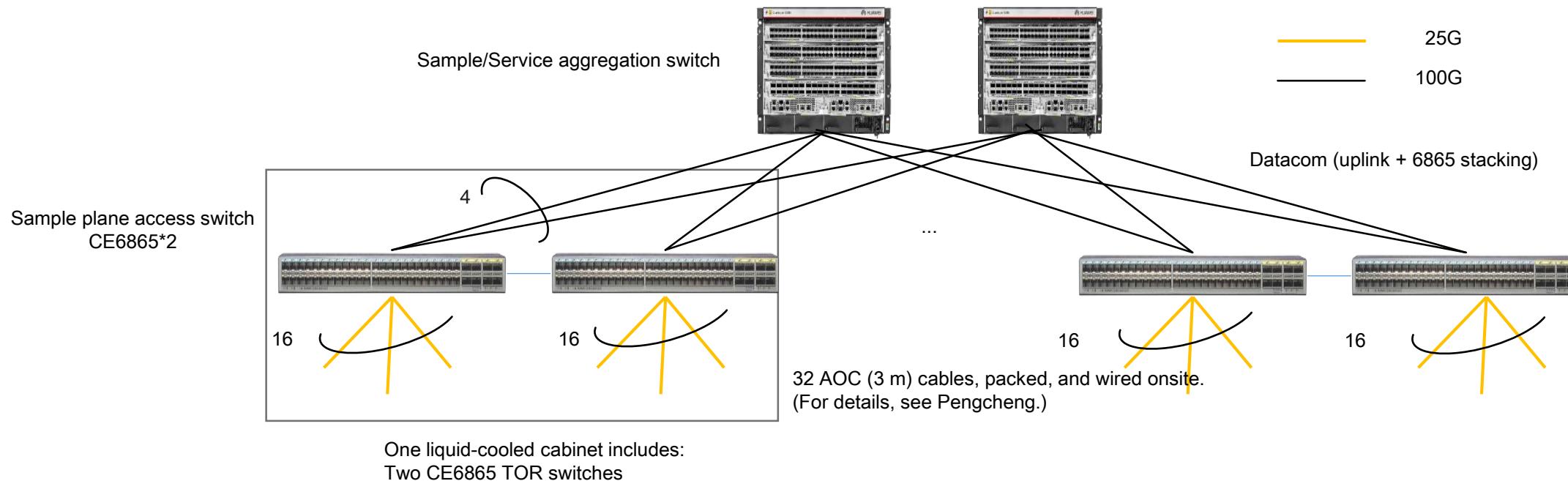
- Each AI processor has a 100 Gbit/s onboard network port connected to a high-speed switch and uses the RoCE v2 technology to form a high-performance AI training parameter network.
 - Leaf-spine networking is used. The convergence ratio is 1:1, ensuring optimal performance of the high-performance computing network and implementing high-bandwidth and low-latency cluster communication for cluster training tasks.



AI cluster sample plane/service plane network topology

AI cluster sample plane/service plane network:

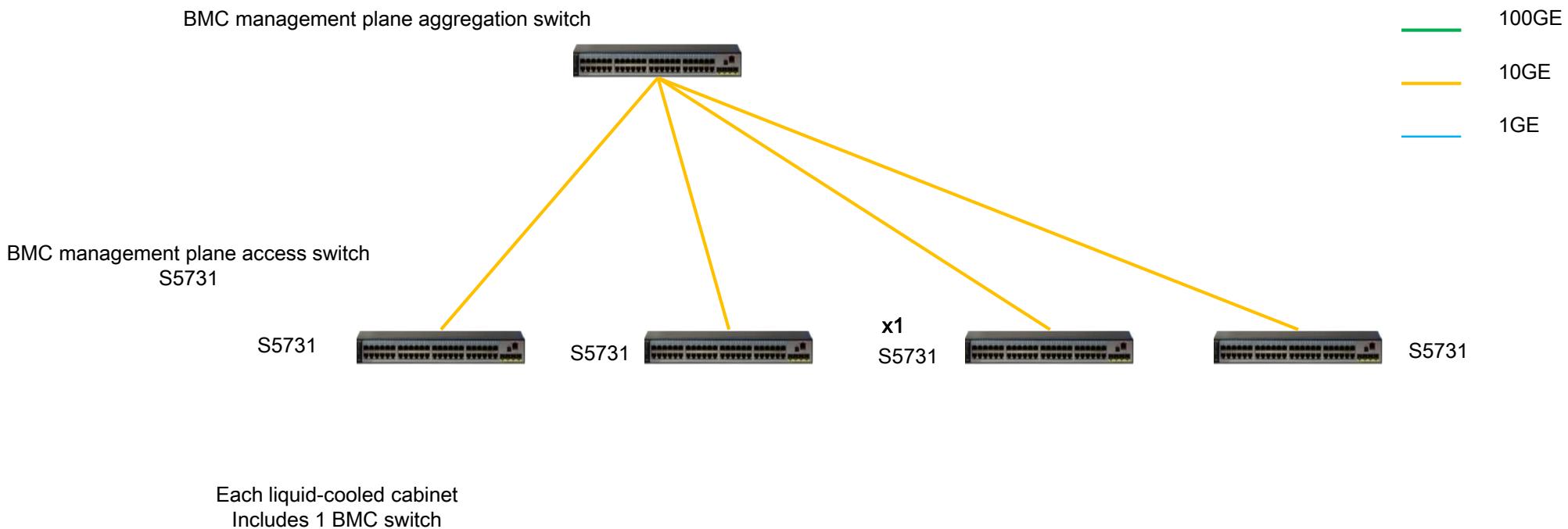
- The Spine-Leaf two-layer architecture is used to construct the sample/service plane network of the AI computing cluster to access the high-speed and large-bandwidth interconnected storage system and service scheduling and management
- Each group of two access switches is configured on the sample plane and service plane. The stacking technology is used to form a logical device that shares the same control plane, simplifying management



BMC management plane network topology of the AI cluster

BMC management plane network of the AI cluster:

Each device provides one GE network, which is accessed through access switches. It is mainly used to manage and operate cluster computing nodes



AI Cluster Advanced Software Solution

AI Cluster advanced Software: One-stop AI platform

● AI Cluster advanced Software : One-stop AI platform

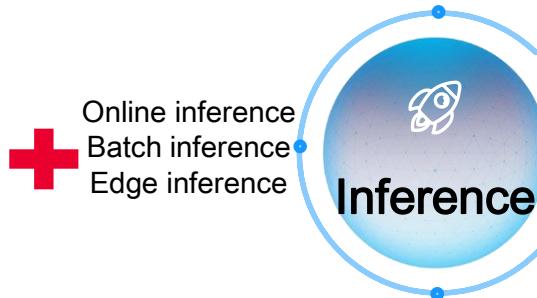
- ✓ E2E AI development pipeLine
- ✓ Unified resource scheduling and collaboration engine
- ✓ Mainstream computing engine
- ✓ Mature Industry AI development suites
- ✓ Abundant knowledge compute service
- ✓ Intelligent Data Lake
- ✓ Industrial Intelligent Twins Template

Distributed training



Industry-leading
training performance

All-scenario inference



Online inference
Batch inference
Edge inference

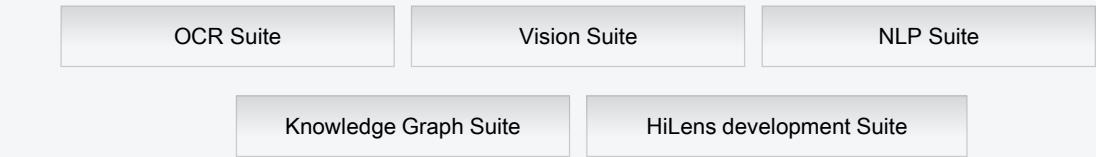
Model lifecycle management

Low cost and high reliability

7 days → 3 minutes

Development of the API for extracting the structured information from forms with new layouts

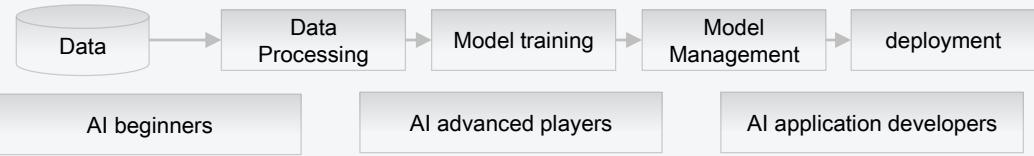
ModelArts Pro: Know-How Prefabricated in the AI Development Field



Arabic supported workflows: ID Card, General Text, General Table, Passport, Scene Text, Customized-template, ASR-AR, Speech suite,

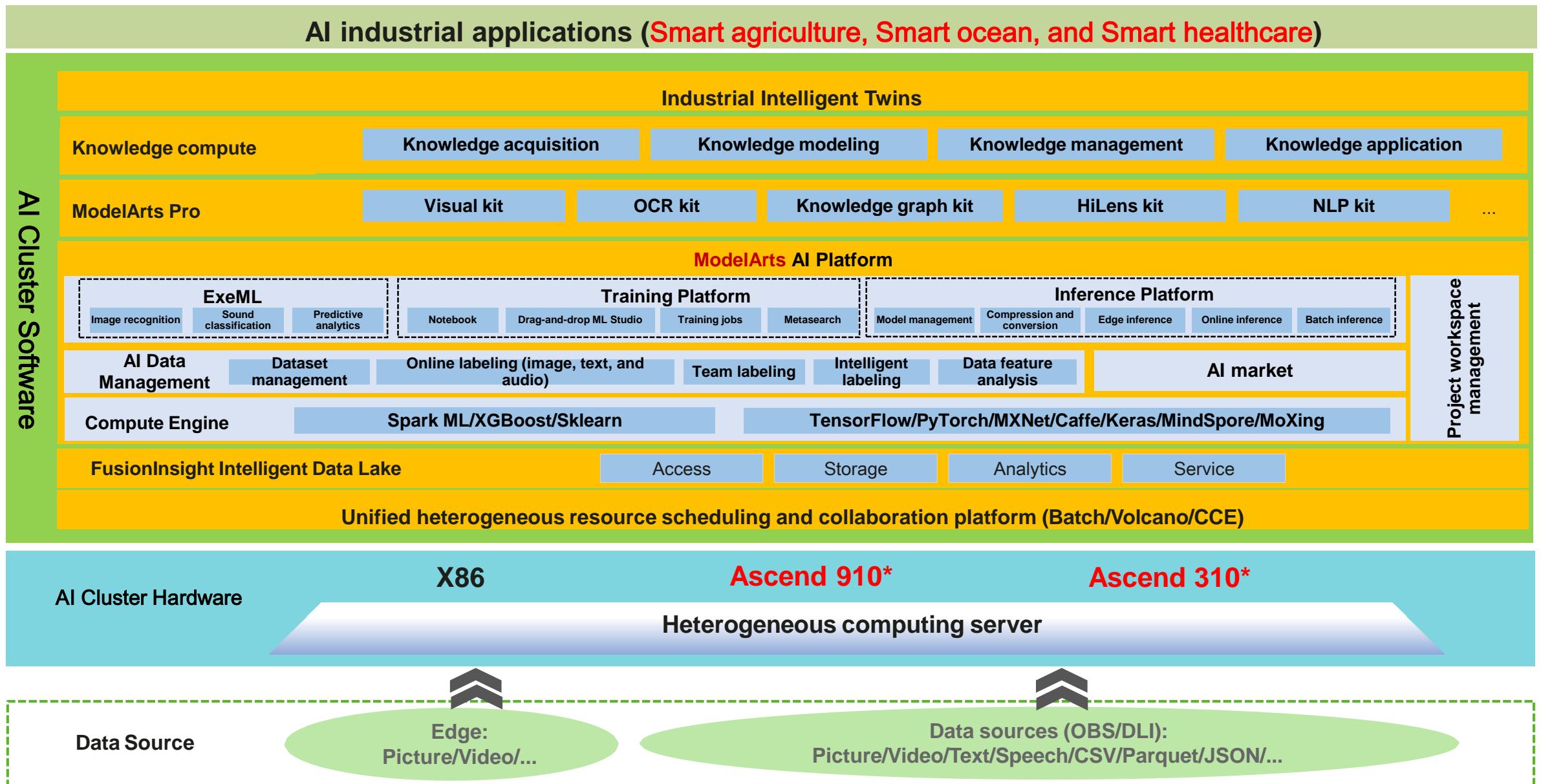
Preset algorithm | Process-based development | Workflow orchestration | Continuous iteration

ModelArts Fundamental: One-stop and Easy Development

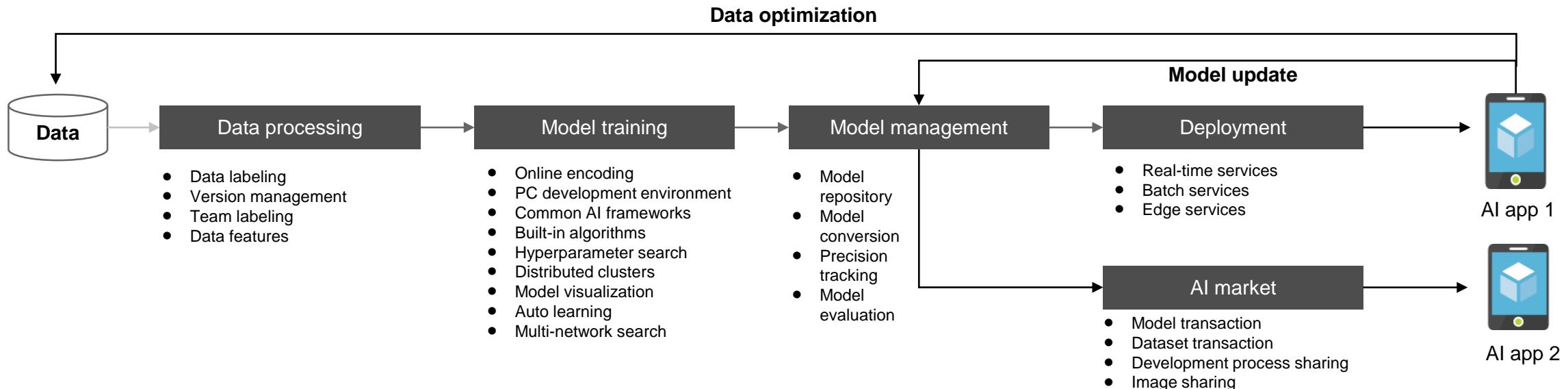


Automatic Annotation | Auto-learning | SDK/Pycharm | Few-shot learning | Federated Learning | Difficult case discovery

AI Cluster Software Architecture



ModelArts: E2E AI development pipeline



- ❑ A **one-stop** AI platform, covering the **entire process** of machine learning, deep learning, and reinforcement learning
- ❑ **Distributed** parallel execution
- ❑ Interactive **notebook** based on the **Python** language
- ❑ Built-in AI engine libraries including **TensorFlow**, **Pytorch**, **MindSpore**, **Scikit-Learn**, and **XGBoost**
- ❑ Mining of **hard examples** and AI-based quick closed-loop
- ❑ **Unified management** of model assets, facilitating application implementation

High efficiency

- ❑ **Auto learning (ExeML)**: quick customization of AI models with **zero AI experience and zero coding**
- ❑ **50+** built-in machine learning operators, including input, data transform, modeling, evaluation, and output operators
- ❑ **Visual** display of data/model/evaluation results
- ❑ TensorBoard integrated to support neural network visualization
- ❑ **10+** built-in **high-precision deep learning models**, covering image recognition, object detection, segmentation, and NLP

Low requirements

- ❑ Self-developed **MoXing** and **AIBox** deep learning training and inference acceleration frameworks with high performance
- ❑ Unified management of heterogeneous resources and container-based **scheduling of massive tasks in seconds**
- ❑ **Distributed** parallel training, with a linear acceleration ratio of 0.8 on 1,000+ GPUs
- ❑ Inference acceleration, increasing the speed by **x times**
- ❑ Support for model conversion and compression, and model deployment on devices
- ❑ **Dynamic scaling**, load balancing, and high availability of tasks in high-concurrency online inference scenarios

High performance

- ❑ Offline SDK/Pycharm plug-in, which can be called by local PCs
- ❑ Unified management of **multi-vendor, multi-framework, and multi-function** models by model repositories
- ❑ Service details monitoring (**CPU/Memory/GPU**)
- ❑ **Gray release** and A/B test
- ❑ **Online rolling upgrades of models**, ensuing service continuity
- ❑ Support for **standard RESTful APIs**, which can be easily encapsulated in AI platforms
- ❑ **Model evaluation and diagnosis**, facilitating quick model rollout

Easy O&M

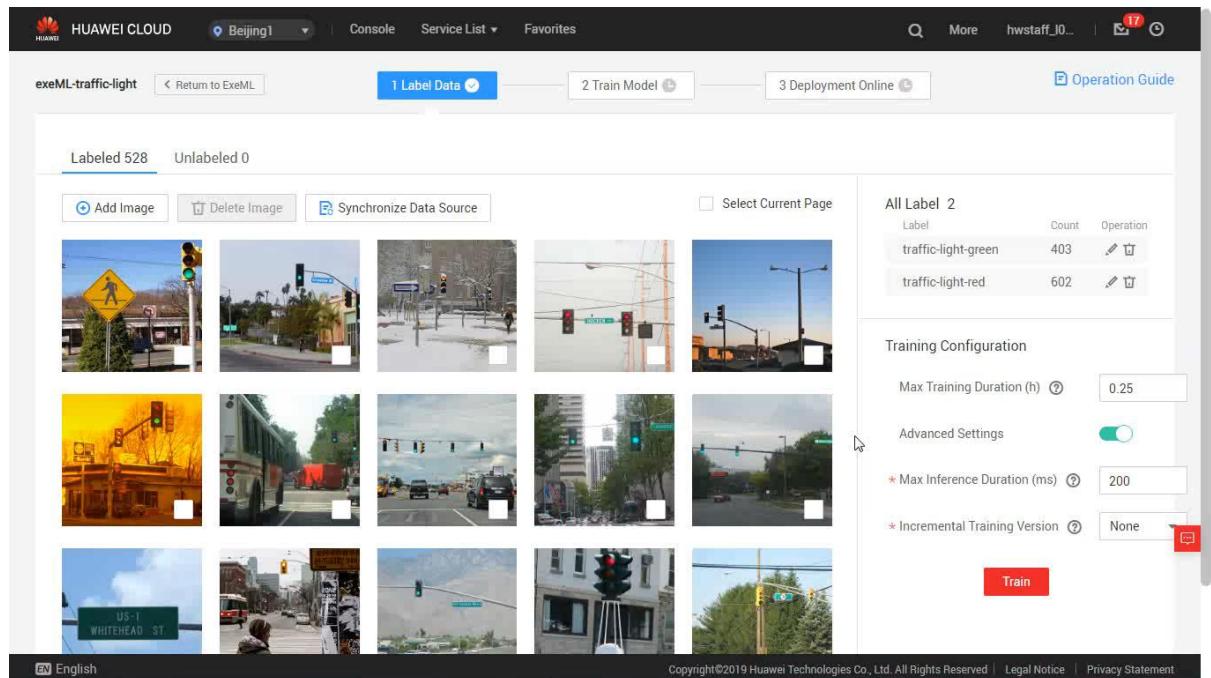
Note: For more information about ModelArts, see the *ModelArts Technical White Paper* or *ModelArts Product Documentation*.

ModelArts ExeML: The execution-orientated auto-machine learning Platform

Zero
Coding

Zero
AI Experience

- Step 1:
Upload and label data.
- Step 2:
Train a model.
- Step 3:
Verify and release the model.



The training can be completed in 20 minutes.

ModelArts : Built-in industry algorithms on the platform enable Industry innovative application development

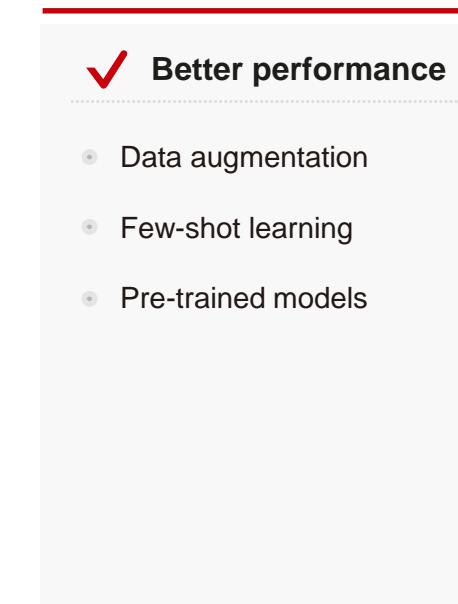
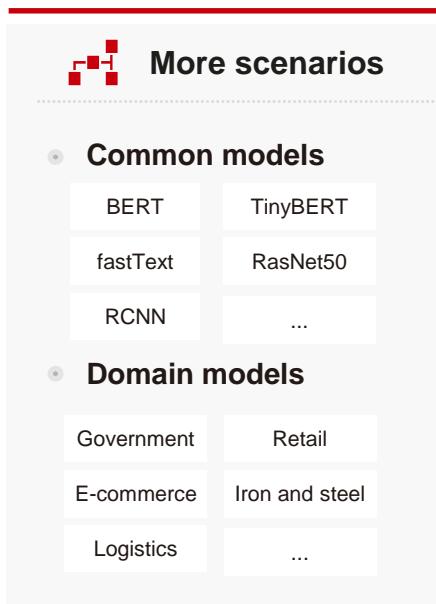
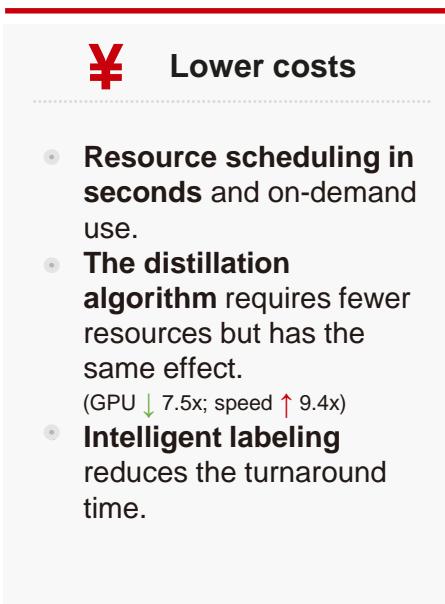


ModelArts Pro
pre-trained industry
algorithms

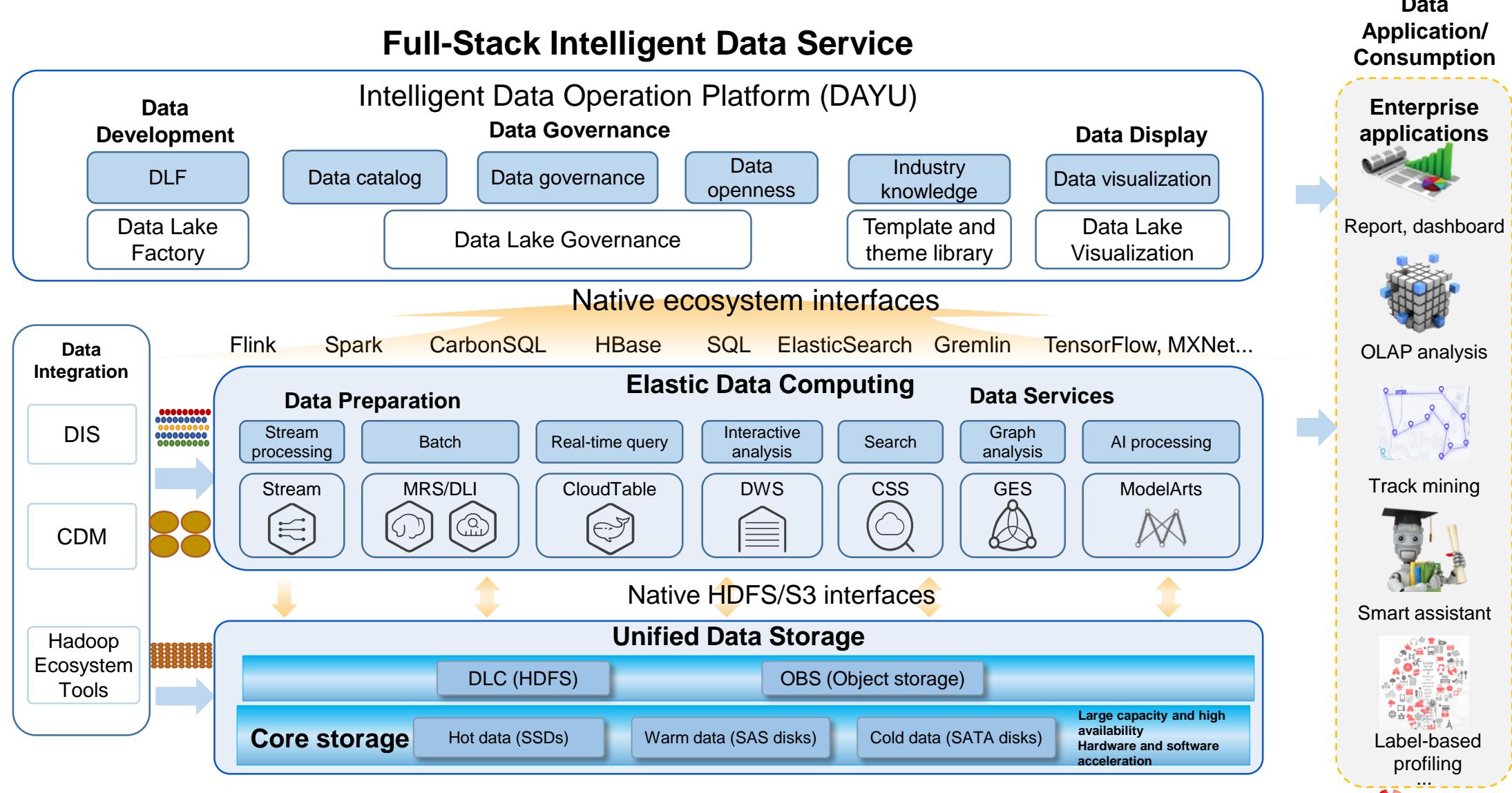


› High precision with a small amount of data

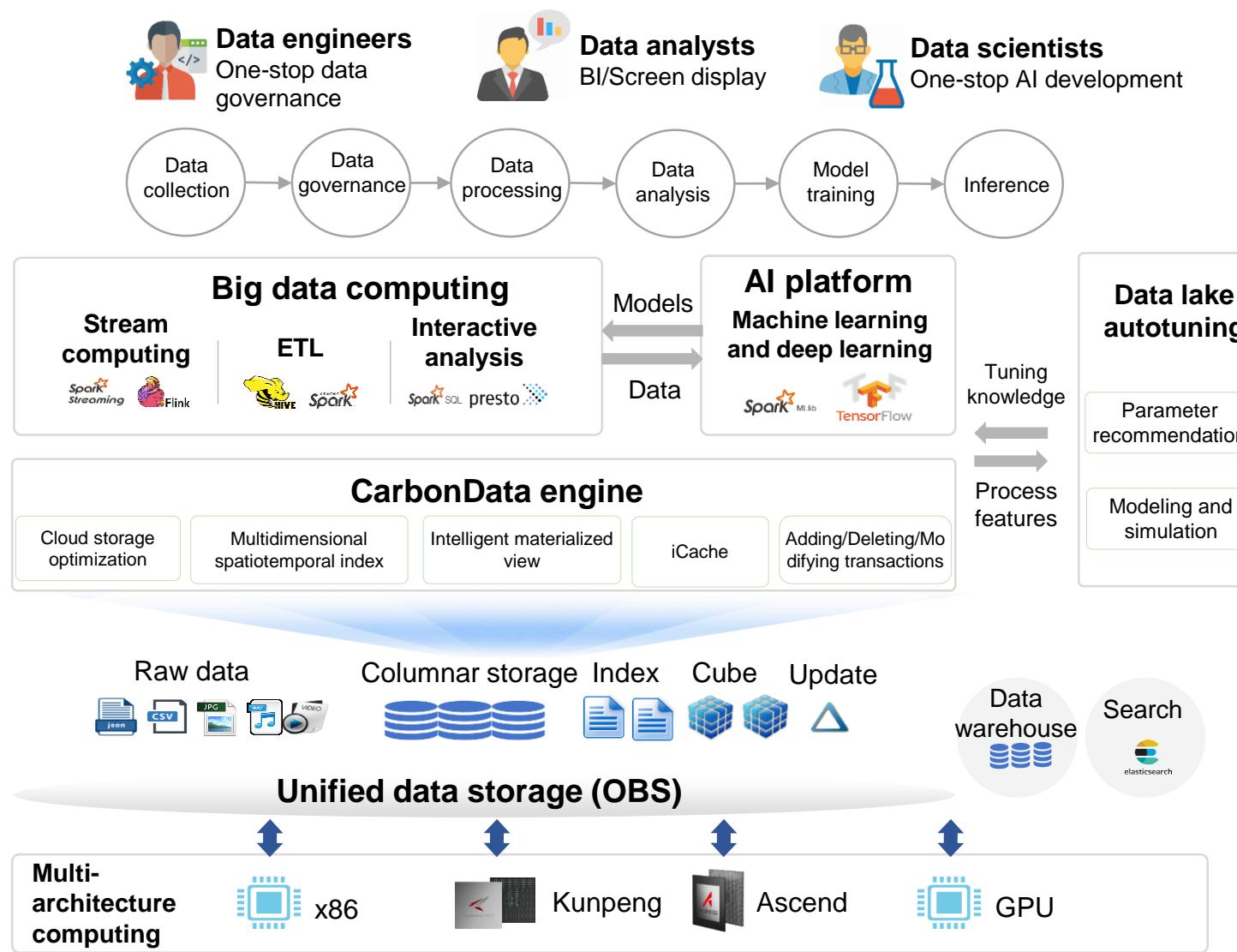
› Oriented to more vertical scenarios



Intelligent Data Lake: Full-Stack Big Data Solution



Intelligent Data Lake: Featuring Interactive Analysis, Transactional Updates, and All-Scenario



Fast: performance improved by over tenfold

- Industry's first multidimensional index based on the data lake, supporting second-level query of billions of data records
- Ascend and Kunpeng, addressing optimal cost-effectiveness

Reliable: strong consistency, modifiable, deletable, and deduplicated

- Industry's first transaction management based on the data lake
- Data addition/deletion/modification allowed; the highest isolation level in ACID supported
- Automatic deduplication of stream data, ensuring no duplicate data or missing data

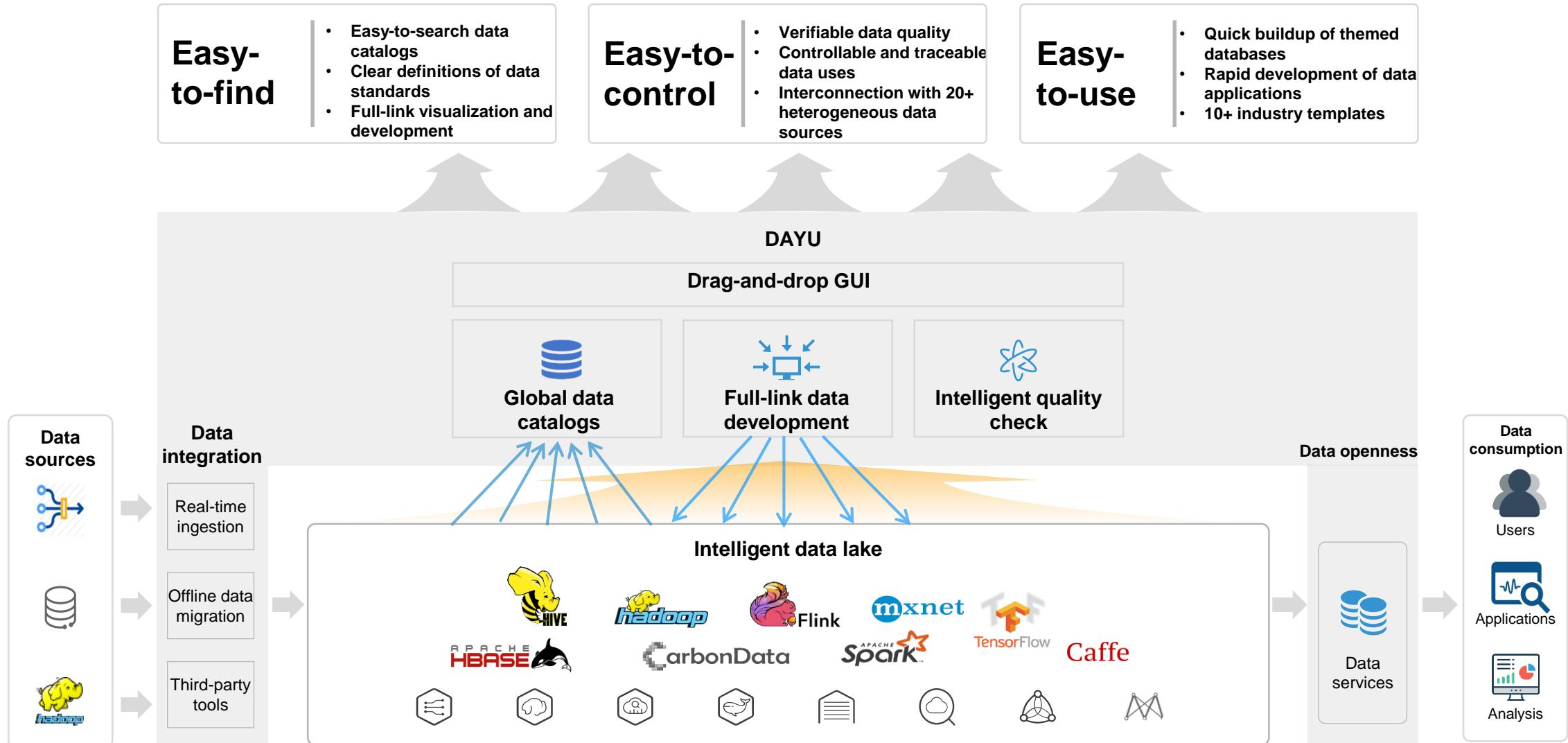
Intelligent: automatic parameter adjustment for better ease of use

- Industry's first Spark + Carbon automatic tuning system
- Learning process features for automatic tuning and parameter recommendation

All-scenario: all in one

- All types of data and roles, and complete service process
- Big data, AI, time sequence, real-time, and cross-source applications

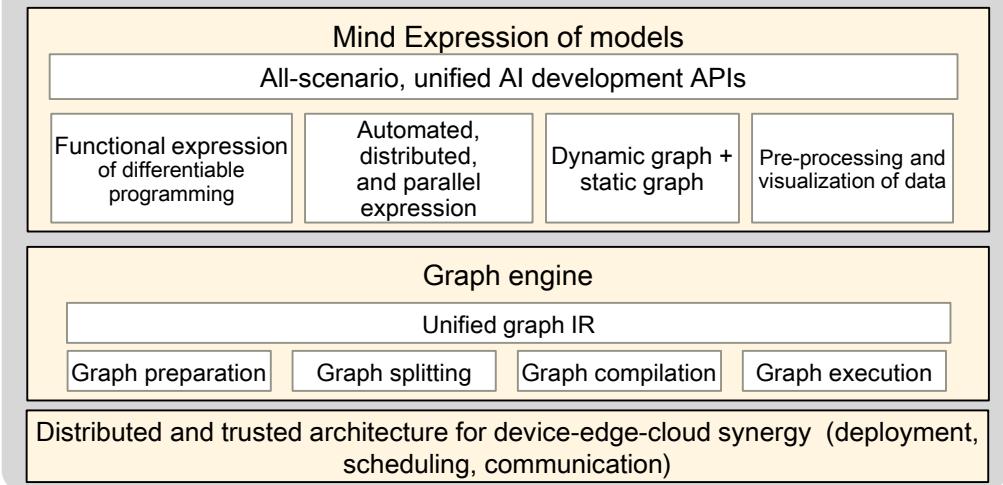
Intelligent Data Lake: One-Stop Data Operation Platform, Expert in Enterprise Data Governance



AI cluster software computing engine

✓ AI Computing framework-Algorithm development

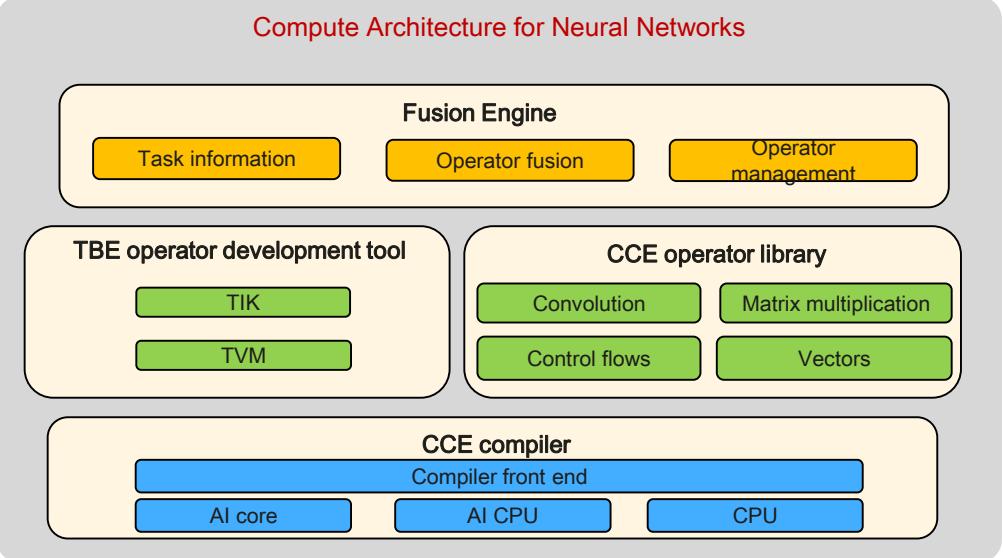
MindSpore: a unified training and inference framework



Maximize performance and efficiency in AI development.

Flexible deployment	Efficient and easy to use	Performance optimization	Open source and openness
Unified architecture for device-edge-cloud scenarios, enabling device-cloud collaboration	Functional Differential Programming Paradigms for Data Scientists, Native Mathematical Expressions	Multi-dimensional pursuit of ultimate performance, soft-hard collaboration, and full play of chip computing power	Open and scalable open source architecture, meeting flexible customization requirements

✓ AI Chip Enabler Library-Operator Development

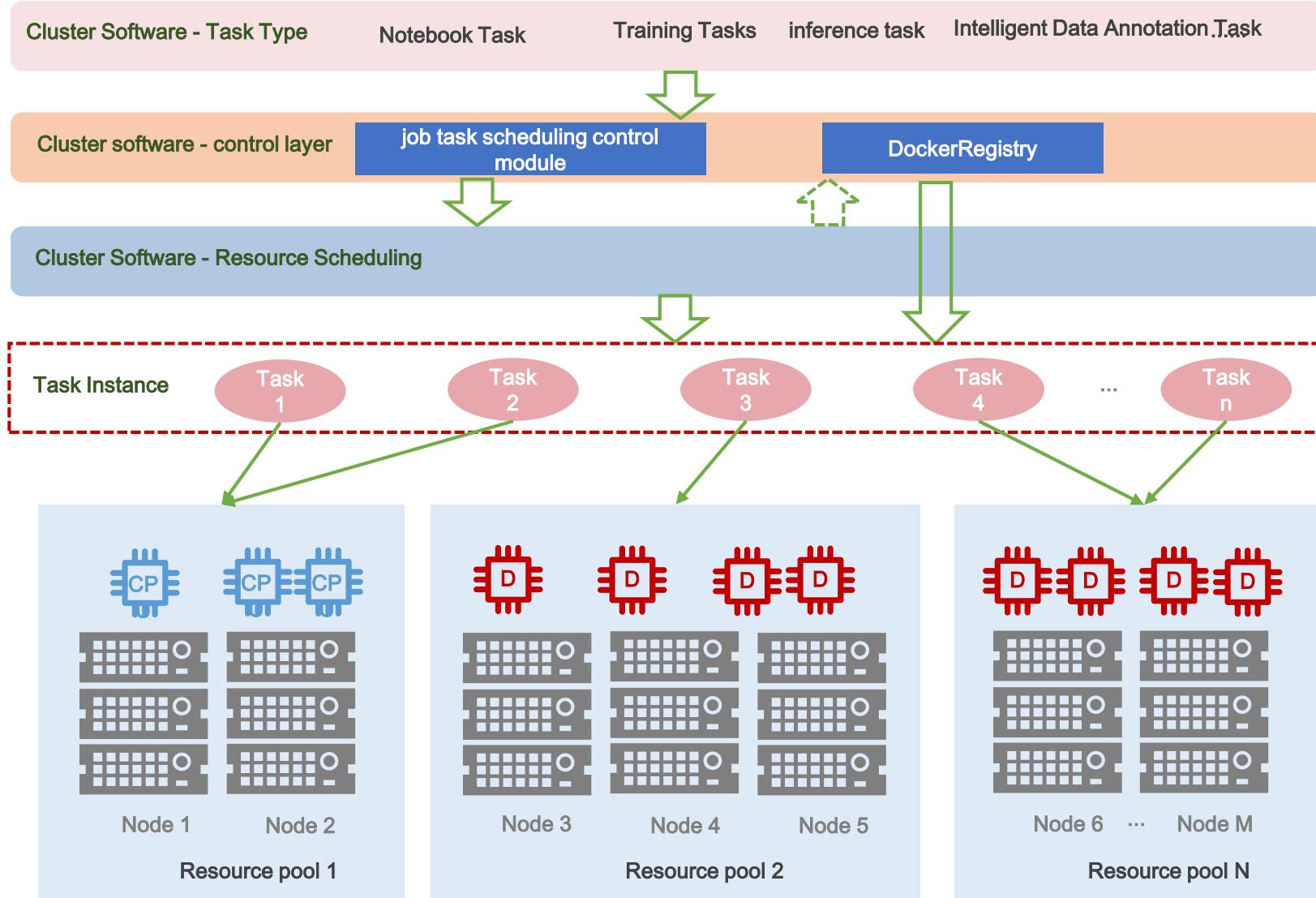


Optimal Development Efficiency and Operator Optimal Matching of NPU Chip Performance

Convergence engine	Operator development tool	Operator library	Compiler
The memory architecture based on the NPU processor implements operator convergence, reduces memory transfer, and improves operator performance.	Preconfigured rich APIs to support custom operator development and optimization	High-performance operator library based on deep optimization of NPU processor	Compilers and Binary Toolset for Heterogeneous Programming Languages

Compute framework and chip energy library are combined to achieve the ultimate performance of software and hardware collaboration.

Unified heterogeneous resource scheduling Engine



Value Features

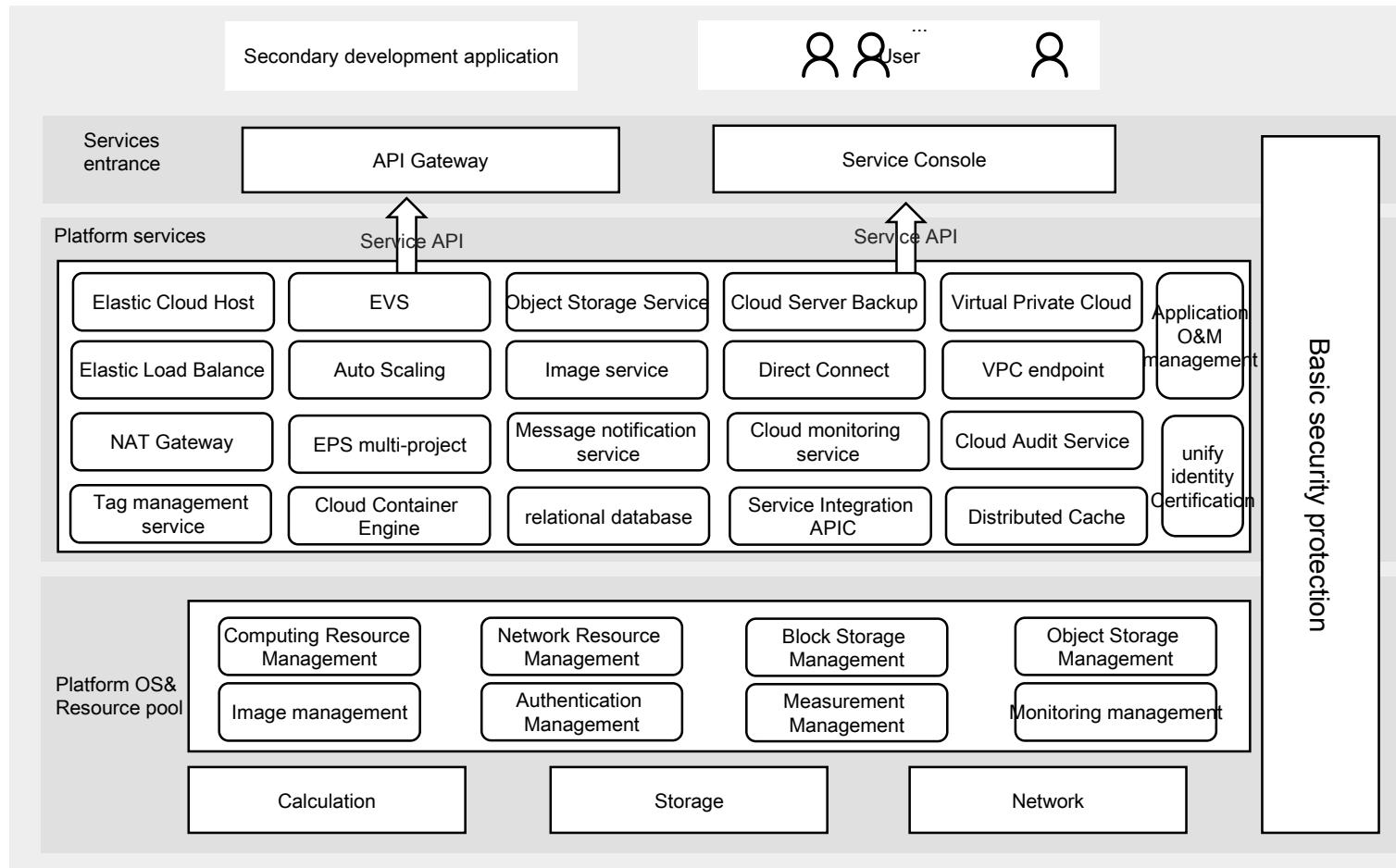
- ✓ Virtual pooling of CPU and NPU resources, and extra-large EB-level computing scheduling and management capabilities
- ✓ Task instances are run and scheduled using Docker images, and are isolated from each other
- ✓ Leading computing power and value
- ✓ Supports reliable task scheduling management authentication.
- ✓ **Unified resource management and containerized scheduling of massive tasks.**



AI cluster basic software

Basic AI computing platform

Includes service entry, platform service, platform OS, and resource pool modules



Service entrance

- **Service console:** Create, use, and delete resources on the console.
- **API interface:** orchestrates and schedules resources using service APIs.

Platform services

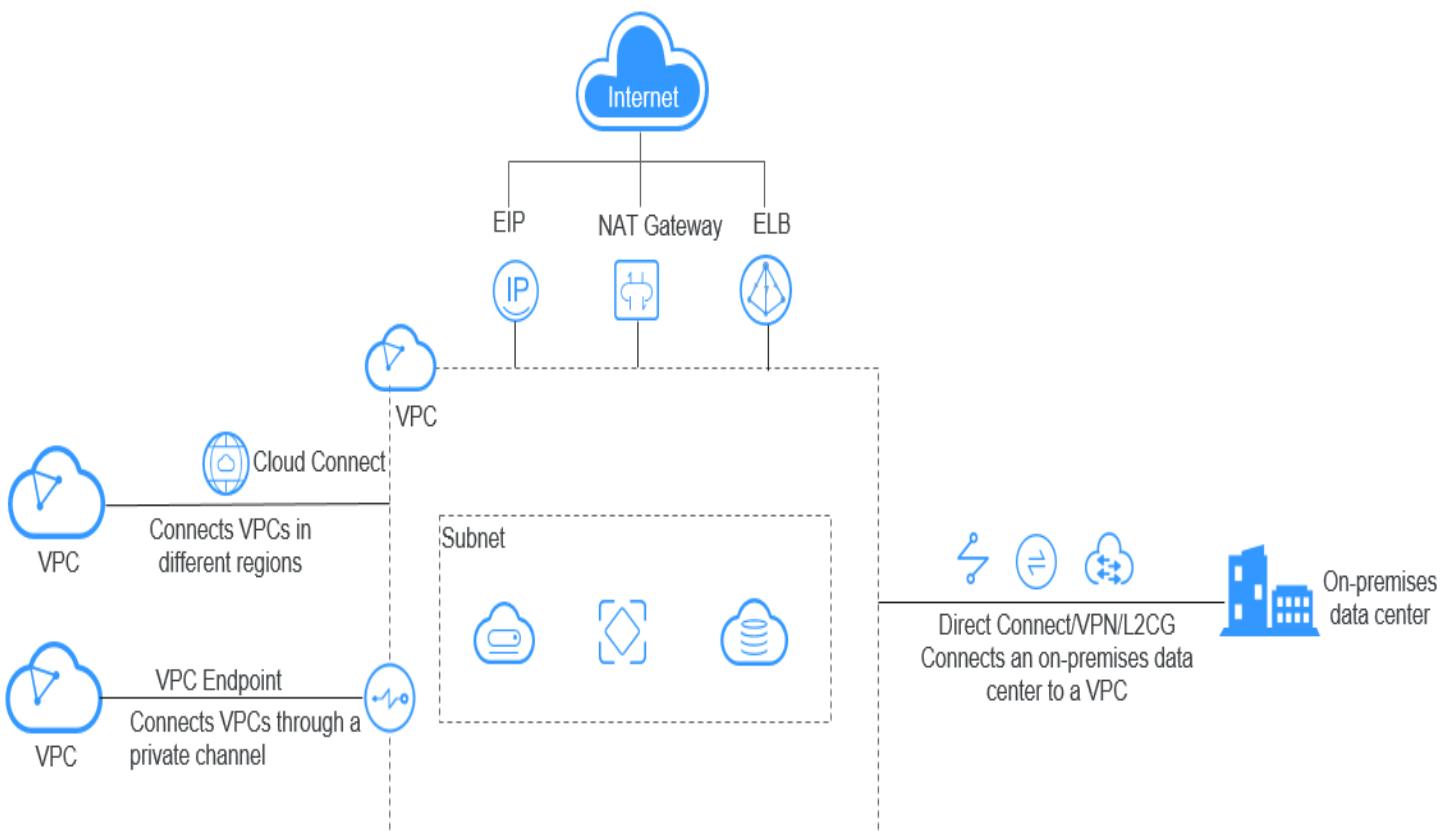
- **Various types:** Provides various computing, storage, network, and management services.
- **Multi-tenant:** Services can be used by tenants that are isolated from each other.
- **Elasticity:** supports on-demand scaling and elastic scaling.

Platform OS and resource pool

- **High reliability:** physical redundancy, mutual backup of components, and plane isolation
- **Resource management:** supports computing, storage, and network resource pools, authentication, and monitoring management.

Network service system

The network service system builds a virtual network environment that is isolated, configured, and managed by users for cloud servers, cloud containers, and cloud databases, improving the security of users' cloud resources and simplifying network deployment.

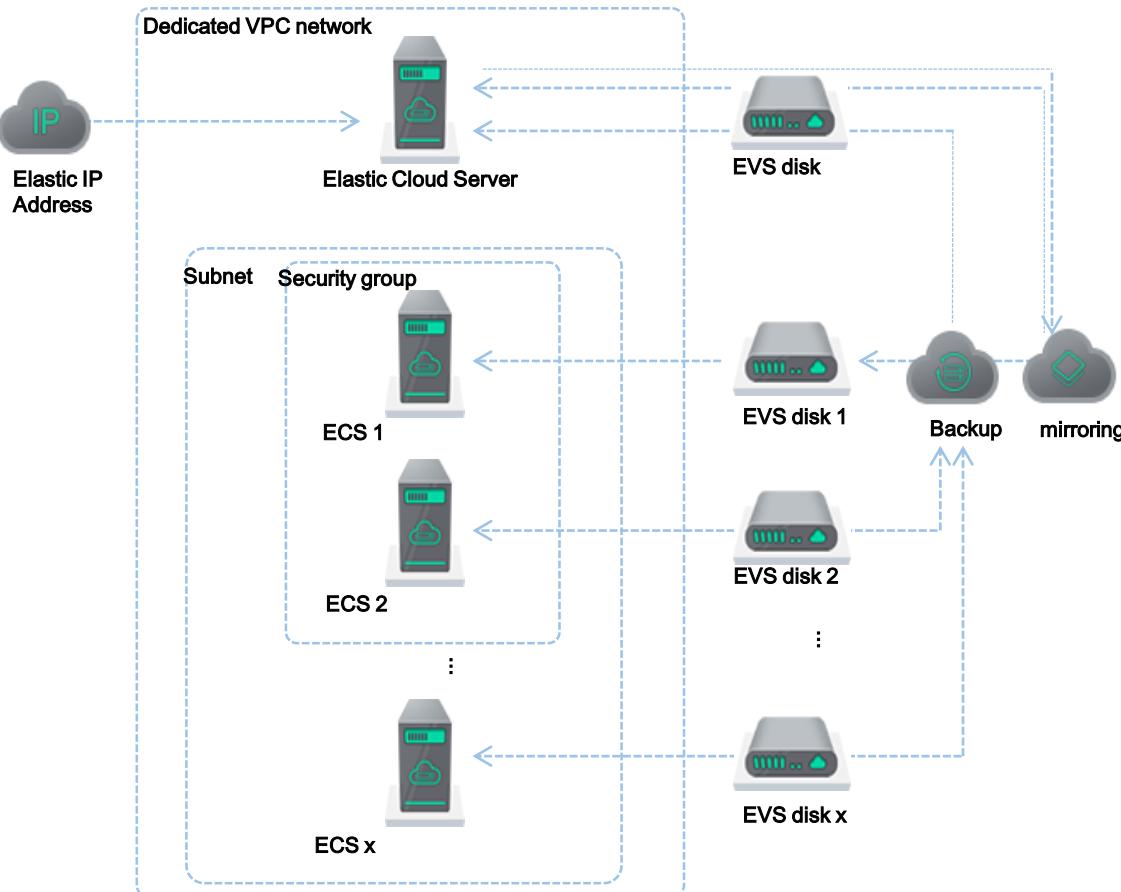


- Flexible configuration:** Subnets are divided as required, IP address segments, DHCP, and routing tables are configured.
- Secure and reliable:** Logical isolation between VPCs, network ACLs for subnets, and security groups for ECSs
- Interconnection:** Supports multiple interconnection modes, such as elastic public IP addresses, Elastic Load Balance, NAT gateways, virtual private networks, and direct connections.

General Computing Services



ECSs are self-service and scalable ECSs. ECSs help you build a reliable, secure, flexible, and efficient application environment, ensure stable service running, and improve operation efficiency

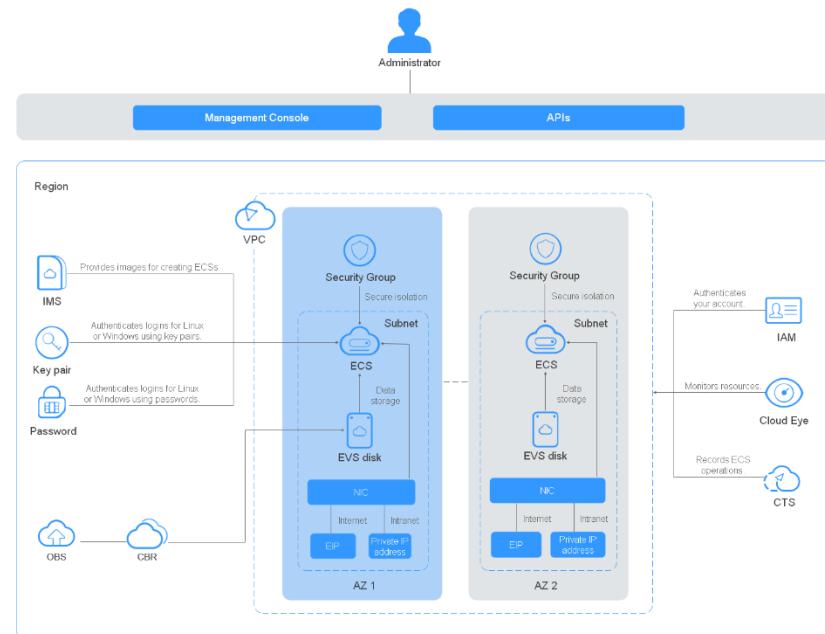


- Reliable data: backup and restoration of the entire system, and multiple data copies are redundant.
- Security protection: network isolation, anti-DDoS traffic cleaning, WAF, and vulnerability scanning
- Elastic and easy-to-use: Automatically adjusts elastic computing resources based on service requirements.
- Load balancing: Automatically distributes traffic to multiple ECSs to expand application service capabilities.
- Various disks: High I/O and ultra-high I/O disks
- Efficient O&M: Various management modes, such as console and APIs, are provided
- Cloud-based monitoring: Provides timely and effective resource information monitoring and alarm reporting

Storage System Services



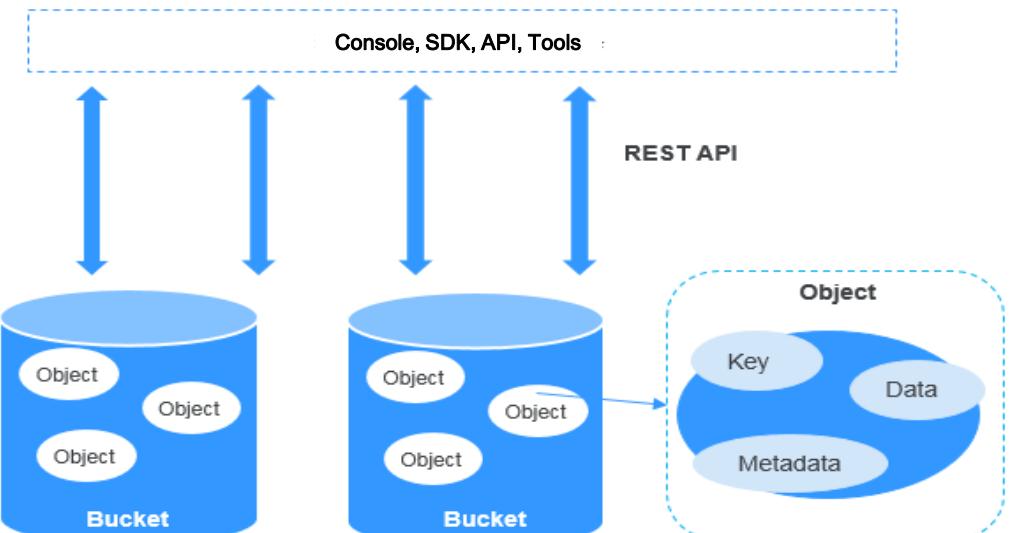
Elastic Volume Service (EVS) is a virtual block storage service based on the distributed architecture and scalable



- **High performance:** 33000 IOPS and 350 MB/s per disk
- **Distributed architecture:** no single-point bottleneck and offline expansion with scale
- **Elastic expansion:** On-demand online capacity expansion without service awareness
- **Secure and reliable:** supports multiple copies, data encryption, and snapshot backup.



Object Storage Service (OBS) is a service for storing unstructured data, such as AI samples, documents, images, and videos



- **High scalability:** hierarchical independent expansion, DHT routing
- **High performance:** pre-read, distributed cache, hierarchical, and 1.5 TB/s read throughput
- **High reliability:** component and data redundancy, cabinet and server security, and durability of 11 9s
- **High security:** multiple protection and authorization management

AI industrial applications

Smart Agriculture AI+ Management

Agricultural production AI+ visualized management



Mobile terminal operation is easy for farmers to use



① Identification unhealthy plants:

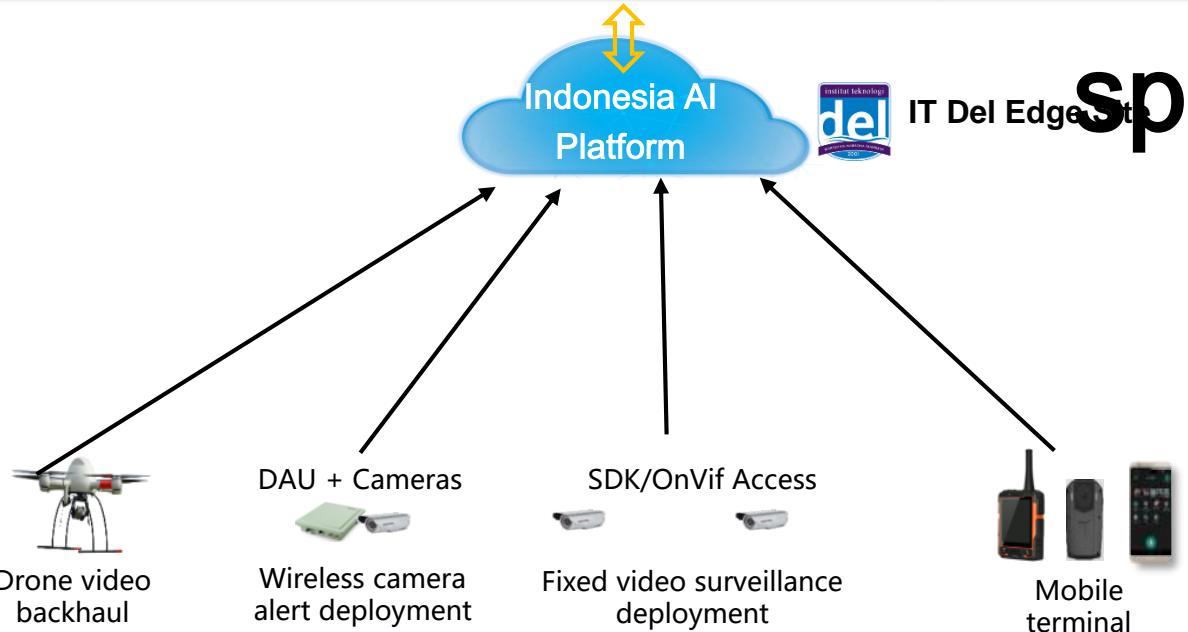
- Identify plant disease & pest with just a photo. Reduce risk of crops failure and increase your harvest.

② Expert knowledge service:

- Provide agronomist consulting service through apps

③ Cultivation Guide :

- Provide expert knowledge, help farmers to learn how to cultivate dozens of commodities and manage hundreds of plant disease and pests. Be a super farmer



Need to be more specific



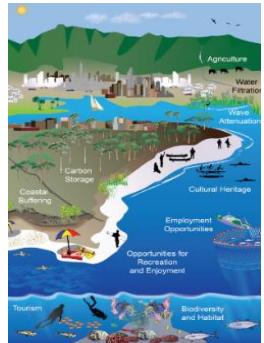
Visual image of different pests infections sent for training to build model inventory for identification of pest infection for early intervention to prevent widespread infection affecting the valuable crops

Smart Ocean: Fishing Management



Management of Fisheries

Indonesia Fishery sector: \$27B -3% GDP
Jobs: 7M
Provides for 50% of protein diet for locals
Overfishing @38% fisheries in Indonesia



Conservation of Coastal Ecosystem Services

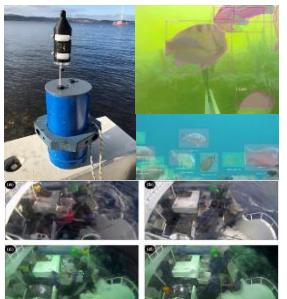
Huge archipelago with ~100K km of coastal line subjected to high level of flooding (source: Global Facility for Disaster Reduction and Recovery)
Strengthen Blue Carbon economy



- Prediction of fish stocks sustainability for fishery management
- Identification of illegal fishing activities
- Monitoring use of harmful fishing gear

Need to be more specific

- Historical data to model and mitigate against coastal mangroves destruction for prevention of storm surge related flooding
- Modeling and Prediction of storm surge levels
- Enhance attractiveness of marine biodiversity for development of hospitality industry



Monitoring of Fishing Stocks

The current exchange of information between fishermen and authorities is not optimal for managing the sustainability of fishing stocks (overfishing, illegal fishing etc) Fishing stock recovery increases 30% over traditional methods when applying AI to model fisheries management (MIT Darwin project)

- Combining robotic camera and underwater sound sensors to monitor illegal fishing during both day and night, and to integrate multiple types of data to build a more complete picture of ocean activities

Infrastructure construction

Features of the Micro-module Equipment Room

Rapid deployment

Modular design and engineering
productization

Deployment in 4 - 10 months

Micro-module DC



Intelligent O&M

AI digital O&M,

Visualized monitoring of all links and predictive
maintenance

Green and environmental protection

Fully utilize natural cooling and AI optimization,
and PUE ≤ 1.4

Ultra-high power

Reliable liquid cooling, maximum power
density

About 45kW/cabinet

Equipment List

Site	Model Name	Qty	Power consuming
KKI	Management Node-x86 2U	18	12,132
	Network Node-x86 2U	2	700
	General Purpose Computing Server-x86 2U	24	10,848
	General Purpose Computing Server-x86 2U	182	78,260
	General Purpose Computing Server-Taishan 2U	3	1,170
	Memory-Optimized Computing Server-x86 2U	2	946
	Dense Storage Computing Server-x86 2U	7	4,039
	Dense Storage Computing Server-x86 4U	6	4,896
	Accelerated Computing Server-Taishan 2U	7	8,891
	AV606.0.0.2-physical.pi6.3xlarge-AI model	11	50,424
	High I/O-x86 2U	10	3,900
	Ultrahigh I/O-x86 2U	29	12,702
	OBS-Object storage server-x86 4U	10	5,680
	LVS Server-x86 2U	2	948
	Security service zone-south-north&management plane NDR-x	2	948
	Scattered public cloud package-Server Accessory-Interconnect	1	0
	Atlas 900 POD	1	43,000
Total			239,484

Datacenter Facility List

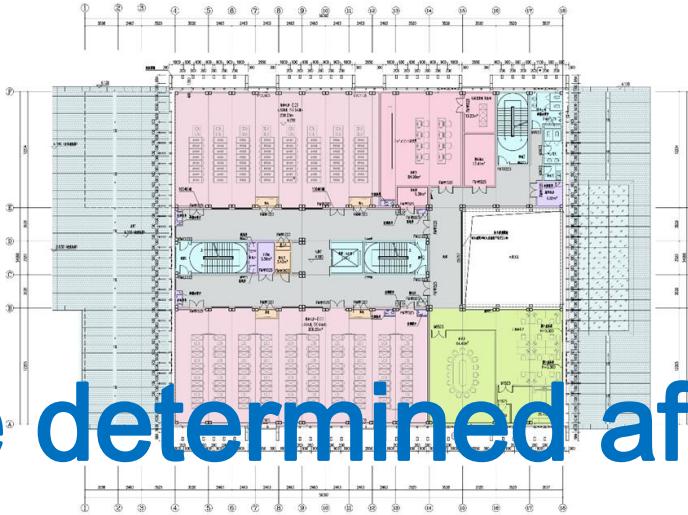
Model Name	Sub Model	Description	Power consuming
Datacenter Module	Power supply and distribution system	Transformer 2500KVA	2
		low voltage distribution-1 5000A*3+3200A+1250A*3+3*630A/3P+2*400A+320A+250A+250KVAR	2
		low voltage distribution-2 3*1250A	2
		low voltage distribution-3 630A/3P+250A/3P*6	2
		Other power distribution	16
		UPS-1.5MW,UPS5000H	0
		UPS-600K,UPS5000H	4
		UPS-300K,UPS5000E	2
		UPS-200K,UPS5000E	0
		Battery - full cabinet	20
	Refrigeration system	Cables inside the container	1
		70U	10
		LCU-65kw	12
		LCU-30KW	4
		POD Internal HVAC System - Air System	1
Peripheral cooling system	Management system	POD Internal HVAC System - Water System	1
		DCIM	1
		Fire protection system	1
		Cabinets	26
		R pdu	52
	Smart module	CDU	7
		Smart busbar	1
		Lighting/Draining	1
		Internal installation/commissioning	1
	Chiller	Air-cooled screw600KW	3
		Refrigeration module	1
		Liquid cooling module	1
		Cooling tower	2+1
		Chiller teamwork control	1
		Cold storage tank	1

Datacenter room design- TBD

一层



二层



三层



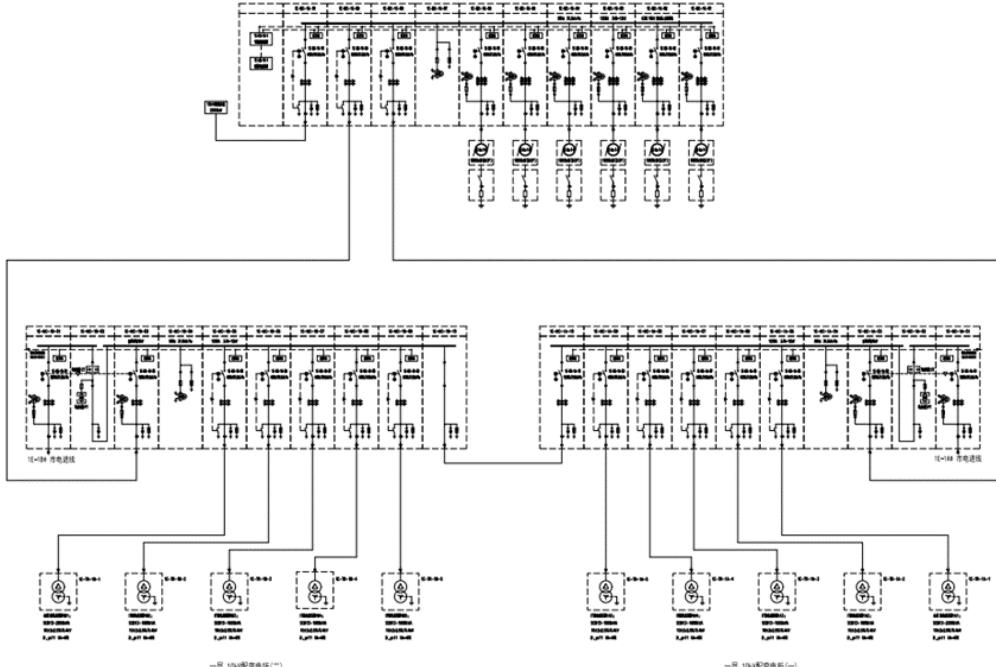
To be determined after on site survey

功能房间	数量	面积
展厅	1	121
大堂	1	125
供电局用房	2	24
运营商机房	2	17.5
10kV配变电室	2	614
电池室	2	143
柴发并机控制室	1	63
消防控制室	1	38
报警阀室	1	10
备件室	2	68
工具间	1	16
卫生间	2	18
卸货区	1	32

功能房间	数量	面积	备注
液冷机房	1	250	液冷机柜64台、平均功率约45KW
普通机房	1	306	47U云机柜90台、平均功率约9KW
ECC监控中心	1	85	拼接屏、8人以上操作台
运维监控室	1	13	
测试机房	1	17	
小会议室	2	50	可容纳不低于20人开会
开放办公区	1	50	可容纳不低于15人办公
值班休息室	1	26	可容纳2人休息、含冲淋间
茶水间	1	3	
卫生间	2	18	
新风机房	1	6	
工具间	1	5	

功能房间	数量	面积	备注
普通机房	2	612	可容纳47U云机柜180台、平均功率约9KW
大会议室	1	65	可容纳不低于45人开会
开放办公区	2	140	可容纳不低于30人办公
开放讨论区	1	26	
茶水间	1	4	
卫生间	2	20	
新风机房	1	6	
工具间	1	5	

Power supply and distribution architecture



- The load devices in the data center equipment room use dual power supplies, and the end switches automatically. Use the battery as the backup power supply (the backup time is 15 minutes)
- The air conditioning system in the data center equipment room uses dual power supplies, and the end switches automatically. The storage battery is used as the backup power supply (the backup time is 15 minutes), and the external cold storage tank is configured as the backup uninterrupted cold source



Smart lithium battery UPS

- Long life: no replacement is required.
- High Reliability: Most Stable Lithium Battery Architecture
- Ultra-high density: industry's highest density UPS 3U@100 kVA
- Flexible: The new and old batteries can be used together.



Power module

- AI-based intelligent visualization of all links and node temperature
- Predictive AI maintenance and proactive warning
- Ultra-efficient, up to 95.5% link efficiency
- Industry's highest density UPS: 3 U@100 kVA

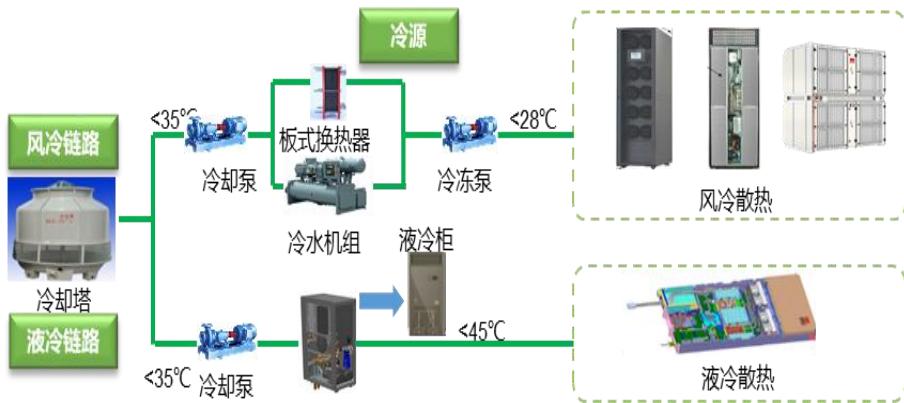


Intelligent busbar

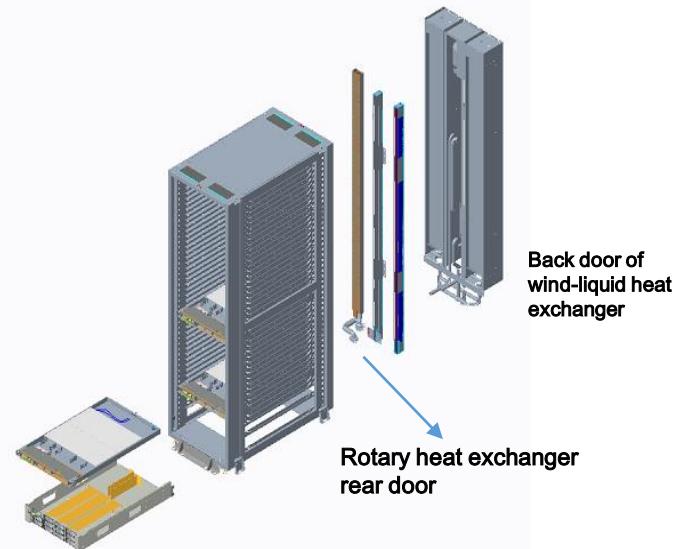
- Reliable: hot swappable for easy maintenance
- Intelligent control: All-round electrical parameter monitoring and risk warning; Configure FE interface, strong networking capability, and remote signaling, telemetry, and teleregulation. Flexible management, PUE management and monitoring

Cooling System Architecture and Advanced Technologies

Multi-system natural cooling



fully liquid-cooled cabinet, extreme heat dissipation



AI-based All-Link Energy Efficiency Optimization



- The chilled water system in the cloud equipment room uses three operating modes: mechanical cooling, partial natural cooling, and full natural cooling. The water supply and return temperature is 17/23°C, prolonging the natural cooling duration and saving energy.

- The plate-level cooling system of the liquid cooling equipment room uses the cooling tower as the cooling source, which can realize the natural cooling throughout the year. The wind-hydraulic heat exchange system adopts the natural cooling + chiller to make up the cooling to further improve the energy utilization efficiency.

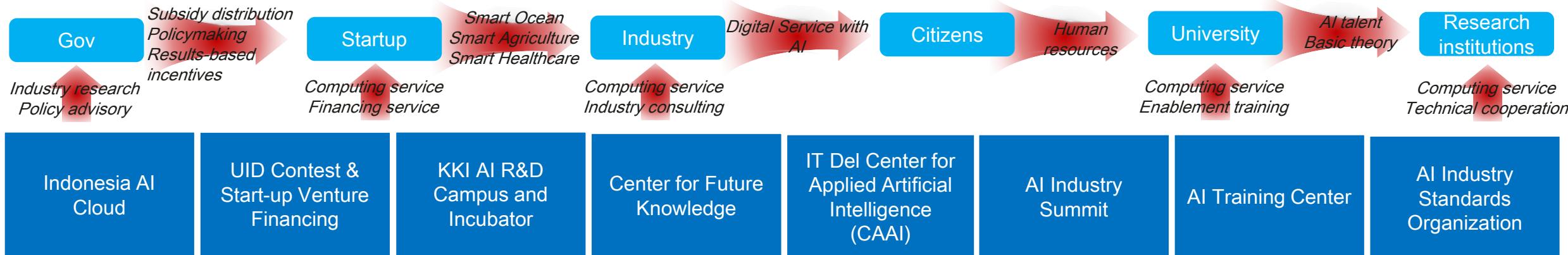
- AI algorithm adjustment, optimizing PUE energy efficiency
- 7 x 24 real-time online, precise on-demand cooling
- Continuously updated models and optimized energy saving effects

- 1 Construction background**
- 2 As-Is Analysis**
- 3 Construction Requirements**
- 4 Construction Solution**
- 5 Operation Solution**
- 6 Budget and Plans**



Indonesia National AI Platform Operation Full View

Making Indonesia 4.0 Digital GDP growth and Citizen Satisfaction growth



Support AI Innovation in Indonesia's Government, Industry, Education, and Research area with Four Ecosystem Platforms, Build a Long-Term Platform Operation System

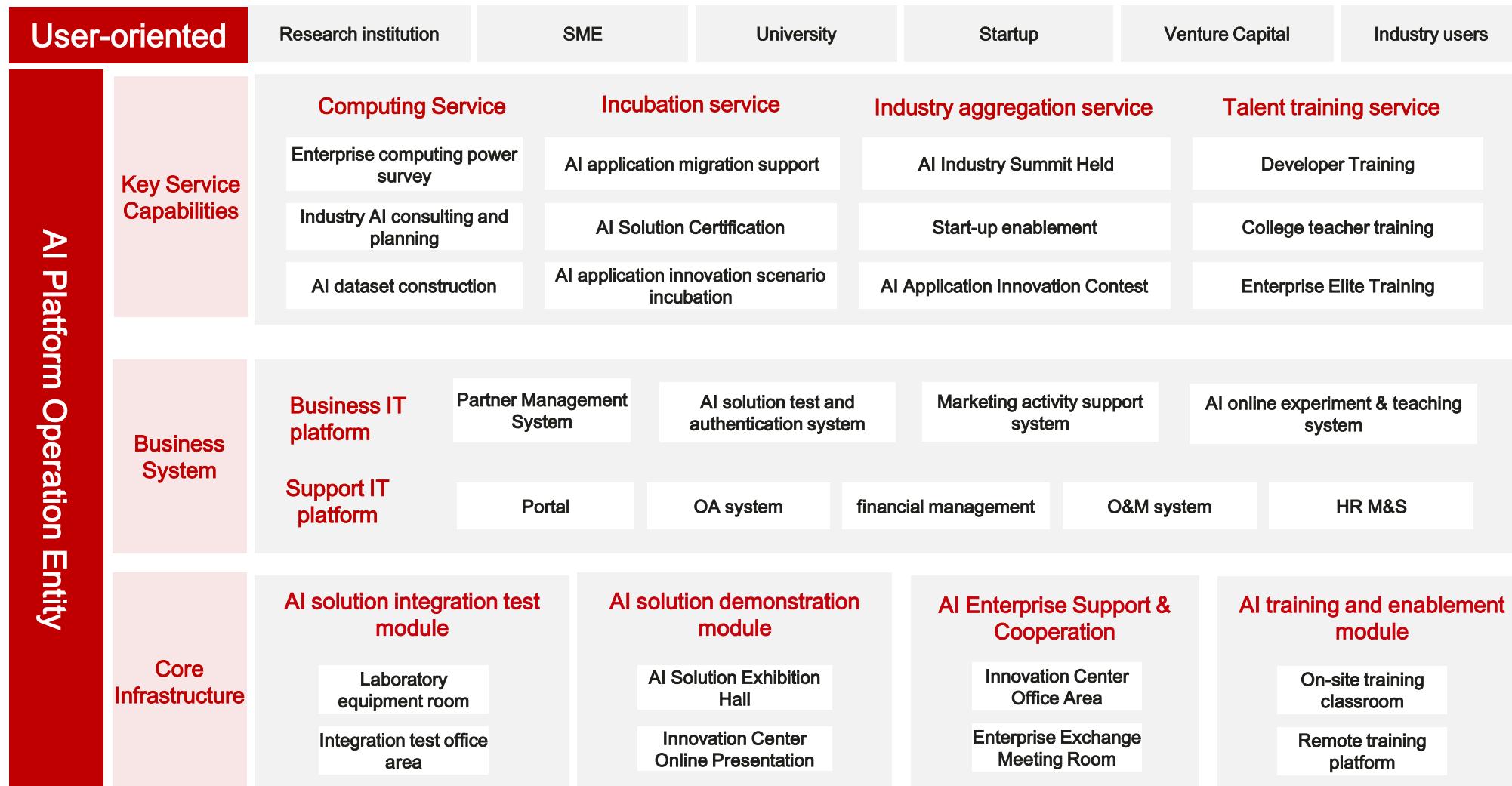
Public AI Computing Platform	Application Innovation Platform	Industry Aggregation Development Platform	Research and Talent Platform
Introduce computing power subsidy policies Providing inclusive computing services (Industry analysis, computing capability research, and subsidy support)	Considering the characteristics of Indonesia's industry Build an AI industry unicorn (Smart Ocean, Smart Agriculture, Smart Healthcare)	Carry out a series of eco-innovation activities Build KKI AI R&D Campus and Incubator (Industry Summits, Technology Salons, Startup Enterprise Connections, and Elite Special Training Camps)	Support scientific research, innovation, and AI talent cultivation, and make breakthroughs in common AI technologies (Local Universities and Research Institutes)

Build four national AI ecosystem platforms based on the national AI platform infrastructure, and enable industry clusters with AI computing power clusters

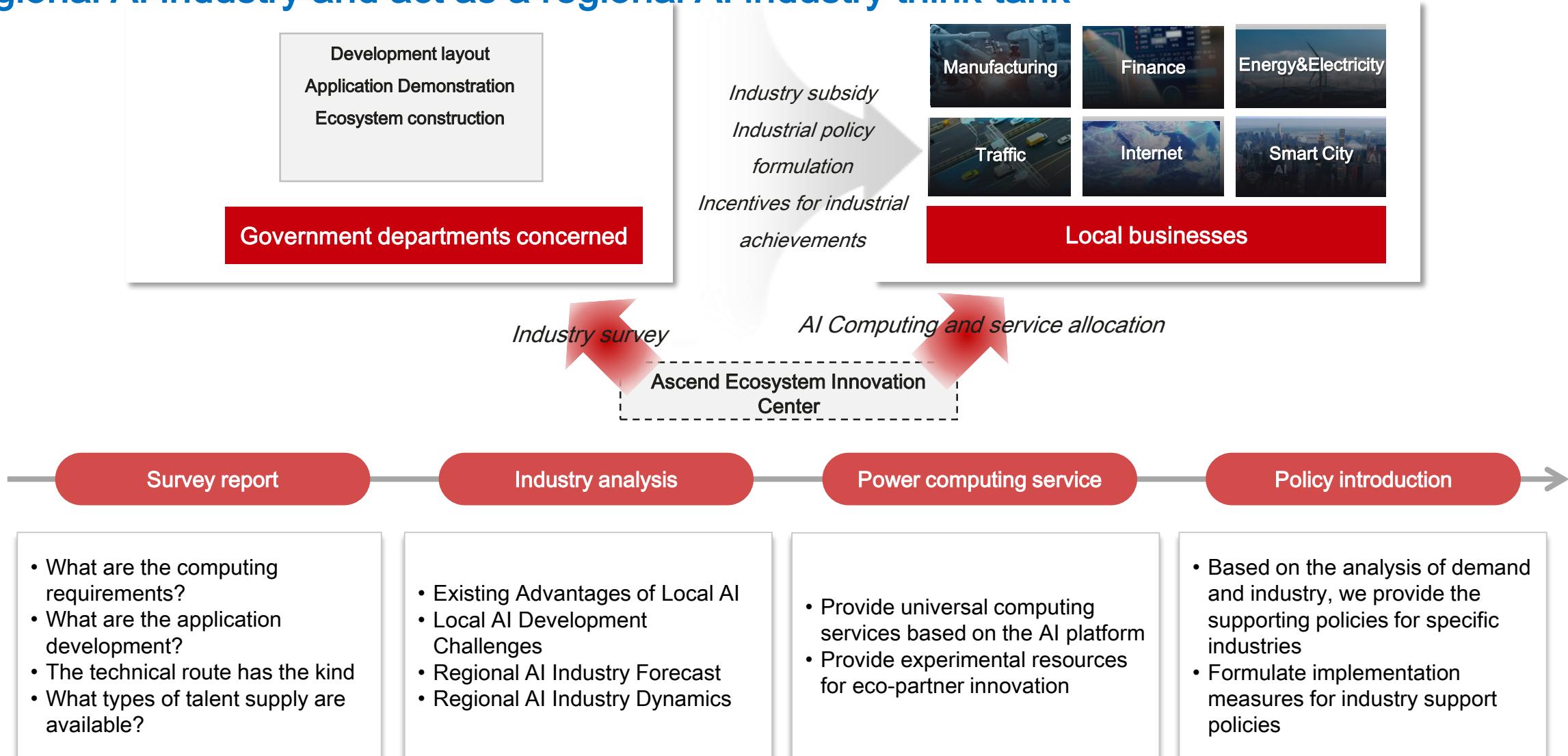
Indonesia National AI Platform Infrastructure

Indonesia National AI Platform Operation Entity Functional Architecture

Functional Architecture of the AI Platform Operation Entity



Public power computing service platform: Gain insight into the current situation of the regional AI industry and act as a regional AI industry think tank



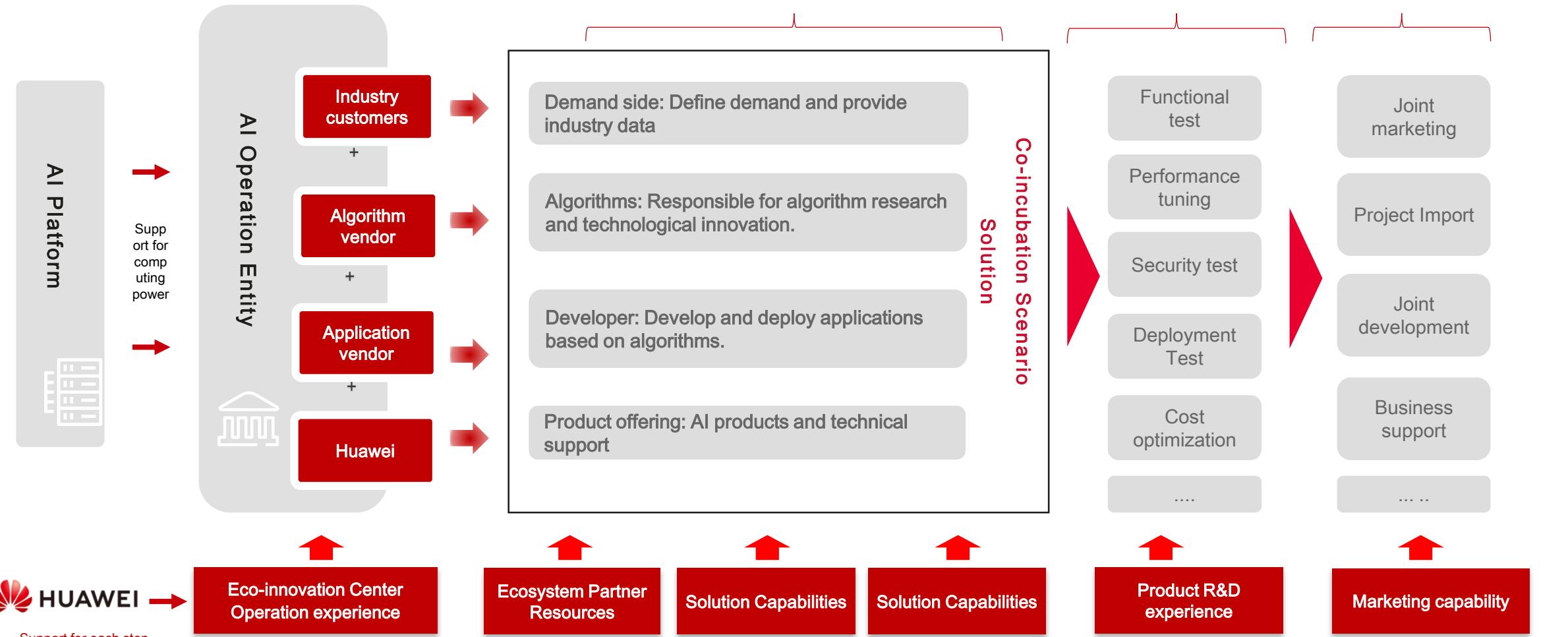
Innovation Application Incubation Platform: Integrates Multiple Resources for Collaborative Innovation, Helps Industry AI Application Incubation

Multi-party collaboration based on innovation centers

Solution incubation phase

Test Phase

Go-to-Market Promotion



Industry Aggregation Development Platform: Carry out Multi-Level Ecosystem Activities to Support Intensive Development of AI SMEs



Build Ecosystem Activities to Support Rapid Growth of AI Enterprises and Promote the Prosperity of AI R&D Campuses

Technology Enablement Activities

Provide initial training for developers and start-ups to help enterprises quickly understand

ISV Special support

Provide special support for ISVs in key industries and cultivate national AI unicorns

Industry Summit Promotion Activities

Promote solution implementation in local industries and attract key ISVs to enter and develop

Start-up Interconnection

Interconnect with and identify local AI startups and provide resource support to promote rapid growth

Developer contest

Build an AI innovation platform based on the AI platform to gather enterprises and achieve intensive development

Inclusive computing power



Research and Talent Platform: Collaborate with universities and research institutes to cultivate high-end AI talent and promote AI research innovation

Cooperation in education with universities and social training institutions

Ally government education sector promote the AI education



Enable social training institutions and expand the team of AI application engineers



Joint Research Institute Technical Cooperation

Operator, model, and SDK crowdfunding development



Academic Incentive Fund



Online and offline multi-form training

AI Academy Online Training



Activities such as colleges and universities and special training camps



AI Platform Operation Target : Three-year construction to improve AI strength and achieve AI industry prosperity

Build an innovative AI industry cluster



Cultivate a batch of competitive application scenarios



Cultivate a batch of high-end AI talent



Building AI Cities with Global/National Impact



Target 1 Develop X+ISV, help X + Startup Success

Cultivate star solutions, develop server partners, edge-side and device-side hardware enterprises, and cooperate with startups to incubate solutions, cultivate national AI unicorns

Target 2 Completed X + Application innovation

Focus on independent innovation in industries such as smart transportation, smart manufacturing, smart healthcare, urban governance, and smart finance to improve competitiveness

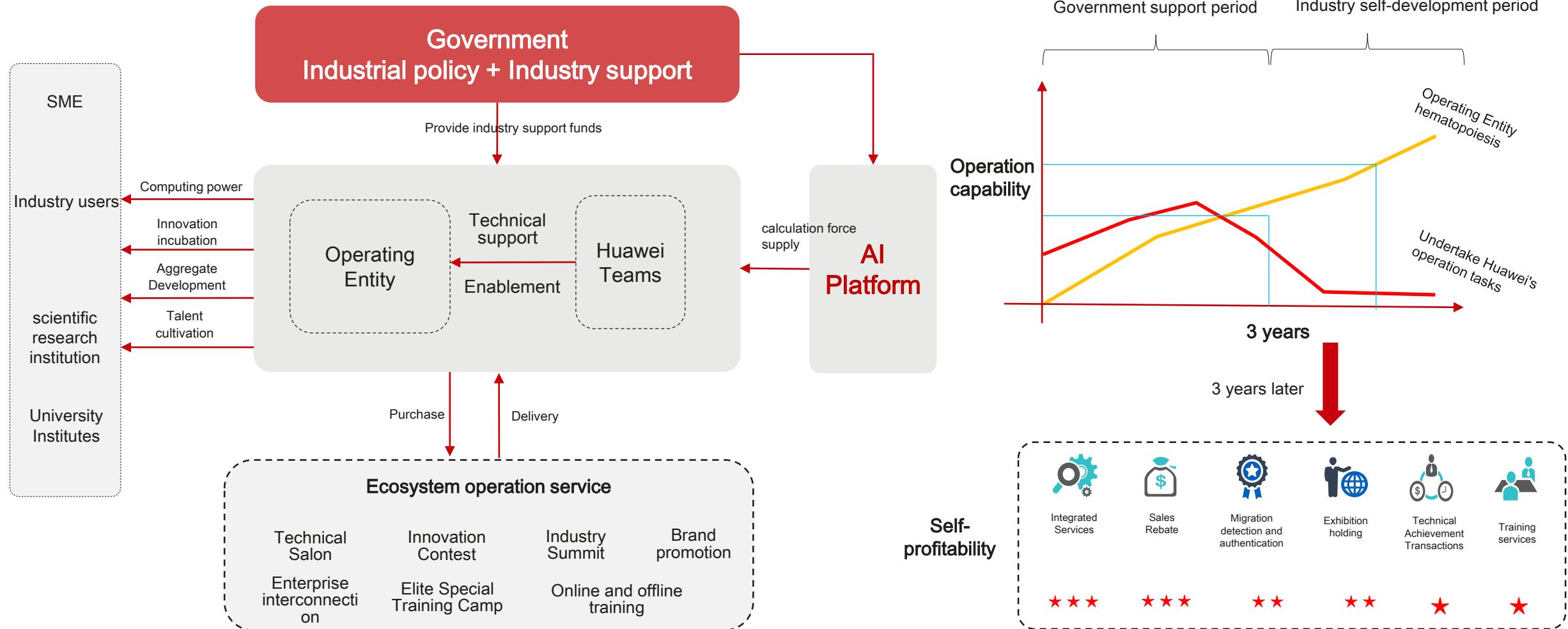
Target 3 Cultivate XXX+ AI talent

Cultivate advanced AI developers to lead research and development

Target 4 Build the brand of the world's top AI summit

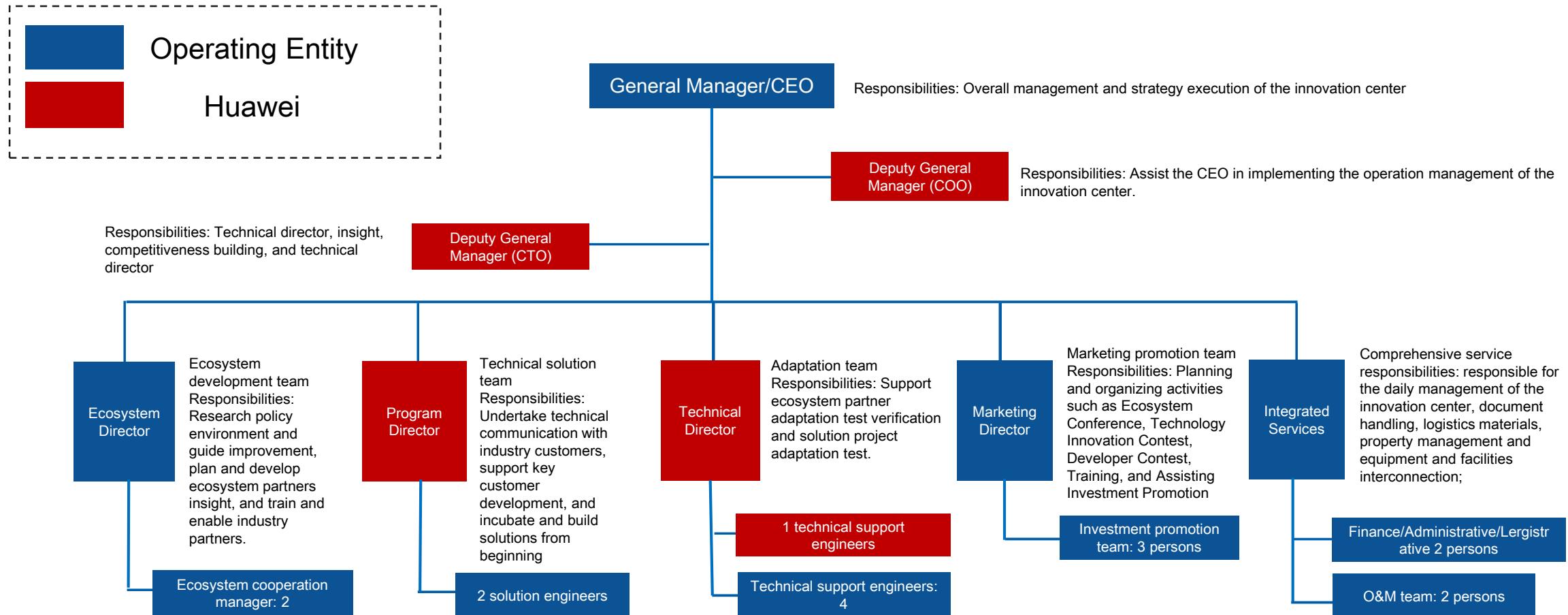
Annual summits that bring together top AI companies, academic leaders, and industry leaders around the world are held to spread AI achievements around the world, Increase the voice of AI standards organizations

Operation solution: Establish a third-party independent entity to operate and build a long-term sustainable development model



Independent Operation Entity Architecture Design

The Operation Entity has XX employees in the first year, including the operation team. (X persons, including CEOs, service, legal affairs, finance, and investment promotion teams) and Huawei's assigned team (X people, COOs, CTOs, and technical teams). The other XX employees will be recruited by the Operation Entity. The organizational structure will be adjusted based on the business development.



- 1 Construction background**
- 2 As-Is Analysis**
- 3 Construction Requirements**
- 4 Construction Solution**
- 5 Operation Solution**
- 6 Budget and Plans**



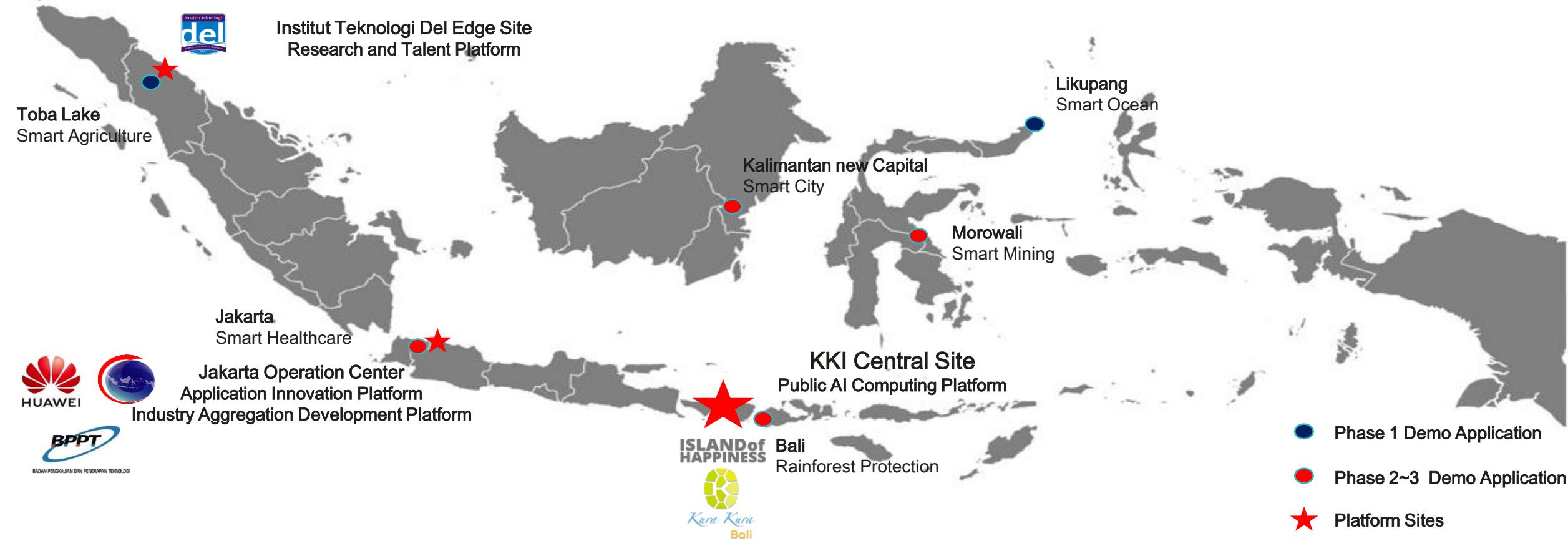
Indonesia National AI Platform Budget Planning

No.	Items	Investment (USD)
1	AI cluster Hardware	7,261,223
1.1	AI computing subsystem	5,097,833
1.2	Platform service system	39,257
1.3	Network service system	32,328
1.4	General computing system	98,452
1.5	Storage Subsystem	790,929
1.6	AI integrated network system	1,202,424
2	AI cluster software	1,712,498
2.1	AI cluster basic software	555,220
2.2	AI cluster advance software	1,036,171
2.3	Unified O&M management platform	108,414
2.4	Secondary development, deployment, and training	12,693

No.	Items	Investment (USD)
3	Smart Ocean Applications	2,590,767
3.1	Marine fisheries	1,541,495
3.2	Disaster prevention and mitigation	1,049,272
4	Information Security System	401,953
5	Energy Infrastructure construction	5,083,559
5.1	Construction and installation	508,356
5.2	Prefabricated micro-module equipment room	4,575,203
6	Operation	4,450,000
6.1	1st Year	1,560,000
6.2	2nd Year	1,810,000
6.3	3rd Year	1,080,000
Total Summary		21,500,000

The total investment of this project is 21,500,000 USD, including 17,050,000 USD for engineering construction and 4,450,000 USD for subsequent operation of the platform

Indonesia National AI Platform Construction Planning



Phase 1: 2021 ~ 2022

- Target: Building up AI Platform Infrastructure
- Total Capacity: **22P FLOPS** (KKI – 20FLOPS, IT Del – 2FLOPS)
- Business Case Trial: Smart Agriculture, Smart Ocean
- Fund support: **Maritime Cooperation MOU**
- Budget Planning: 16.5 Million USD

Phase 2: 2022 ~2023

- Target: Establish **Four Ecosystem Platforms, AI R&D Industry Campus and Center for Future Knowledge** in KKI and **CAAI** in IT Del
- Business Case: Smart Healthcare, Reinforce Protection
- Fund support: **Maritime Cooperation MOU**
- Budget Planning: 5 Million USD

Phase 2: 2023 ~ 2024

- Target: Open Service to all Indonesia entity, enable AI digitalization upgrade and transformation
- Business Case: Smart City, Smart Mining
- Fund support: Platform Self-operation Funds or Government Subsidies
- Budget Planning: xx Million USD

- Phase 1 Demo Application
- Phase 2~3 Demo Application
- ★ Platform Sites

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。
Bring digital to every person, home, and
organization for a fully connected,
intelligent world.

Copyright©2020 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.

