CSA1-N8S1684

8-node 140.8TOPS Cluster Server

V1.1 2025-2-19

T-CHIP INTELLIGENCE TECHNOLOGY

Product features

140.8TOPS powerful computing power

The SOPHON AI processor is BM1684 and supports up to 140.8TOPS (INT8) peak computing power or 17.6TFLOPS (FP32) high-precision computing power, which can meet the application requirements of deep learning model development.

Powerful multi-channel video full-process processing

It supports up to 256 channels of H.265/H.264 1080p@30fps video decoding, 128 channels of 1080P HD video processing (decoding + AI analysis), and 16 channels of H.265/H.264 1080p@25fps video encoding, meeting the needs of various AI application scenarios such as face detection, license plate recognition, and smart cities in video streaming.

Standard 1U rack server

Highly dense and tightly deployed. The server can be configured with up to 8 BM1684 computing modules to reduce TCO. The standard 1U rack server chassis design is designed to fit most types of racks in the data center.

Product features

Rich algorithms and strong practicality

It supports the transplantation of various algorithms such as human/vehicle/object recognition, video structuring, trajectory behavior analysis, etc., with high security and high reliability, and can be flexibly applied to various product research and development.

Open SDK, one-stop AI development kit

SOPHON SDK (BMNNSDK2) is a one-stop deep learning development toolkit that provides a series of software tools such as underlying driver environment, compiler, inference deployment tool, etc. Support mainstream frameworks such as Caffe/TF/PyTorch/Mxnet/ Paddle, support mainstream network models and custom operator development, support Docker containerization, and enable rapid deployment of algorithm applications.

A wide range of applications

Efficiently adapt to AI algorithms in the market to empower AI for industries such as visual computing, edge computing, general computing services, artificial intelligence, smart construction sites, smart transportation, and security.

Specifications

Specifications		
Basic Specifications	Name	Algorithmic Cluster Server
	Model	CSA1-N8S1684
	AI	140.8T@INT8, 17.6TFLOPS@FP32
	Encoding/ Decoding	Video Decoding: 1080@11520fps H.265 & H.264 Video Encoding: 1080P@600fps H.265 & H.264 Image Decoding: JPEG: 4800 pictures/sec @1080P
	Nodes	8 compute nodes(up to 64 ARM cores) + 1 control node
	CPU	Compute node: BM1684, 8-core (A53×8) 64-bit processor, up to 2.3GHz Control node: RK3588S, 8-core (4×Cortex-A76+4×Cortex-A55) 64-bit CPU, up to 2.4GHz
	RAM	12GB LPDDR4 × 8
	Storage	32GB eMMC × 8 Expandable 3.5" SATA3.0 HDD × 1
	Power	300W AC power (input: 100V AC ~ 240V AC)
	OS	Linux
	BMC	Integrated BMC management system provides management interface based on Web. Supports secondary development.
	Framework	Deep Learning Framework: TensorFlow / PyTorch / PaddlePaddle / Caffe / ONNX / MXNet / DarkNet
	Dimension	Standard 1U rack-mounted server: 490.0mm × 390.0mm × 44.4mm
	Environment	Operating temperature：0°C ~ 50°C, Storage Humidity：8% ~ 95%RH(non-condensing)
Interface Specifications	Internet	SFP 10GE × 2, Gigabit Ethernet (RJ45) × 2 (1 control node port, 1 common network port), 4G LTE/5G (optional)
	Display	HDMI2.0 (4K@60Hz, main core board display)
	USB	USB3.0 HOST × 2, Type-C × 1 (for processor core board debugging)
	Fan module	6 high-speed cooling fans

Interface description

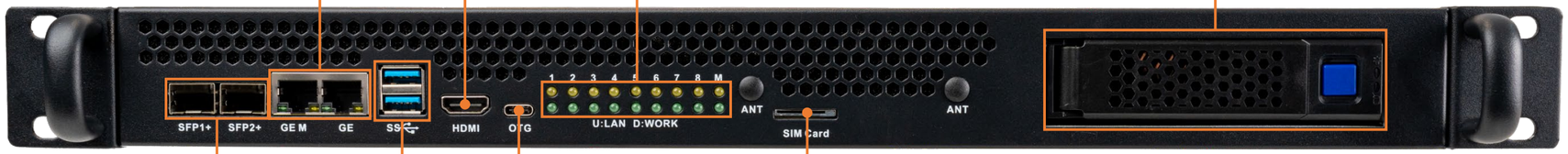
Front view

2×Gigabit Ethernet
RJ45/1000Mbps
The left network port is the control node port

HDMI2.0
4K@60Hz

Work indicator

3.5-inch SATA3.0 HDD



2×SFP 10GE

2×USB3.0

Type-C
OTG

SIM Card

Rear view

Power module
300W AC power supply (input: 100V AC~240V AC)

Fan module

Dimension


390mm
44.4mm
490mm
SFP1+
SFP2+
GE M
GE
HDMI
OTG
U:LAN
D:WORK
ANT
SIM Card
ANT

T-CHIP INTELLIGENCE TECHNOLOGY

Contact Us
(+86)18688117175

E-mail
global@t-firefly.com

Website
https://en.t-firefly.com/

Address
Room 2101, Hongyu Building, #57 Zhongshan 4Rd, East District, Zhongshan, Guangdong, China.