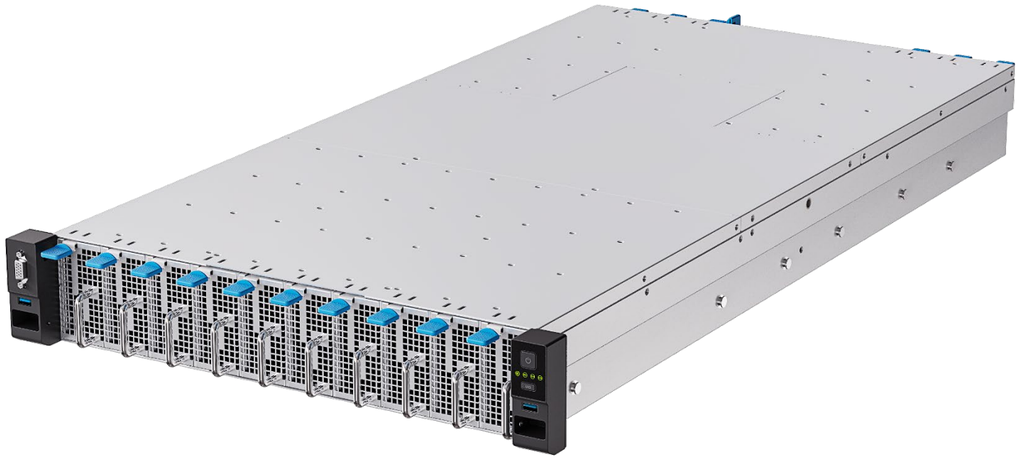
CSD2-N128 Power Server

- RK3588S, RK3576
- BM1688
- QCS8550
- SpacemiT K3



V1.0 2026-4-7

FIREFLY TECHNOLOGY

Product features

128 compute nodes, powerful computing power

The server is equipped with 16 built-in compute blades, supporting a total of 128 compute nodes. A single node can deliver computing power ranging from 6 TOPS to 60 TOPS, and the 16 compute blades are configurable.

3840 channels of video AI processing capabilities

It supports AI processing of 3840 video streams (the specific processing performance varies depending on the specifications of the core board and the type of AI model running), and has powerful multi-task concurrent processing capabilities, which can be widely used in AI application scenarios such as intelligent security and edge computing.

Supports the private deployment of large language models

It supports the privatization deployment of large models such as mainstream language models (Gemma, Llama, Qwen), large visual models (Qwen2.5-VL, InternVL3), and image generation models (Stable Diffusion).

Supports a variety of deep learning frameworks

Support traditional network architectures such as CNN, RNN, LSTM, etc. Supports multiple deep learning frameworks such as TensorFlow, PyTorch, PaddlePaddle, ONNX, Caffe, etc., and supports custom operator development.

Product features

Secure and high-speed network communication

Equipped with eight 10G SFP+ and an independent BMC management interface, ensuring secure, reliable and high-speed network communication.

Equipped with BMC management system

Equipped with BMC intelligent management system, it can easily complete real-time monitoring, software configuration, hardware management, troubleshooting, system upgrade, and can provide secondary development.

Standard 2U rack server

Highly dense and tightly deployed; The standard 2U rack server chassis design perfectly matches most types of cabinets in the data center.

Wide application scenarios

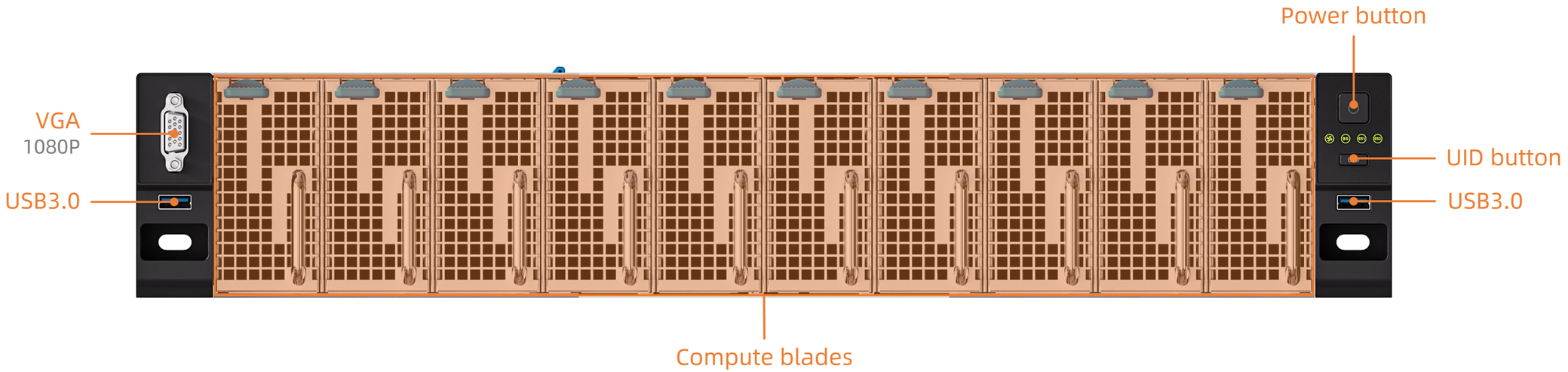
Widely used in edge computing, on-premises large model deployment, smart city, smart healthcare, smart industry, intelligent security and other fields.

Specifications

		QCS8550	RK3588S	RK3576	BM1688	K3
Technical Specifications	Product model	CSD2-N128Q8550	CSD2-N128R3588S	CSD2-N128R3576	CSD2-N128S1688	CSD2-N128STK3
	Launch Status	Launch in April 2026				
	Server form	2U rack-mounted computing power server				
	Architecture	ARM architecture				
	Number of nodes	16 compute blades (128 distributed compute nodes) + 1 control node				
	Compute nodes	Octa-core 64-bit processor Qualcomm QCS8550, up to 3.36GHz	Octa-core 64-bit processor RK3588S2, up to 2.4GHz	Octa-core 64-bit processor RK3576, up to 2.2GHz	Octa-core 64-bit processor BM1688, up to 1.6GHz	Octa-core 64-bit processor SpacemiT Key Stone K3, up to 2.4GHz
	Video encoding	8K@30fps/4K@120fps H.265/H.264	H.265&H.264: 1×8K@30fps, 16×1080P@30fps	H.265&H.264: 1×4K@60fps	H.265&H.264: 10×1080P@30fps	4K@60fps H.265/H.264
	Video decoding	8K@60fps/4K@240fps H.265/H.264/VP9/AV1	8K@60fps/4K@120fps (H.265/VP9/AVS2) 8K@30fps (H.264/AVC/MVC) 30×1080P@30fps (H.265& H.264)	1×4K@120fps (H.265/HEVC,VP9,AVS2,AV1) 1×4K@60fps (H.264/AVC)	H.265&H.264: 10×1080P@30fps	4K@120fps H.265/H.264/VP9
	Control nodes	Octa-core 64-bit processor RK3588, main frequency up to 2.4GHz, the highest computing power is 6TOPS				
	AI computing power	6144TOPS (48T × 128，INT8)	768TOPS (6T × 128，INT8)	768TOPS (6T × 128，INT8)	2048TOPS (16T × 128，INT8)	7680TOPS (60T× 128，INT4)
	RAM	16GB LPDDR5X × 128	16GB LPDDR5 × 128 (4/8/16/32GB)	8GB LPDDR4/LPDDR5 × 128 (4/8/16GB)	8GB LPDDR4 × 128 (4/8/16GB)	32GB LPDDR5 × 128 (8/16/32GB)
	Storage	256GB UFS4.0 × 128	256GB eMMC × 128 (16/32/64/128/256GB)	64GB eMMC × 128 (16/32/64/128/256GB)	32GB eMMC × 128 (16/32/64/128/256GB)	128GB UFS2.2 × 128
	Power	2 × 1300W hot-swappable power supplies, 1+1 redundancy support (Input: 100V AC ~ 240V AC)				
	Fan module	7 high-speed cooling fans				
Physical Specifications	Size	495.60mm(L) × 928.52mm(W) × 88.80mm(H)				
	Installation requirements	IEC 297 Universal Cabinet Installation: 19 inches wide and 800 mm deep and above Retractable slideway installation: The distance between the front and rear holes of the cabinet is 543.5mm~848.5mm				
	Environment	Operating Temperature: 0°C ~ 30°C, Storage Temperature: -40°C ~ 60°C, Operating Humidity: 5% ~ 80%RH (non-condensing)				
Software Specifications	BMC	The BMC management system is integrated with the web-based management interface, supporting Redfish, VNC, NTP, monitoring advanced and virtual media, and the BMC management system can be redeveloped				
	Large language models	All models support private deployment of ultra-large-scale parameter models under the Transformer architecture, such as large language models including Deepseek-R1 Series, Gemma Series, Llama Series, ChatGLM Series, Qwen Series, Phi Series, etc.				
	Visual large model	K3: Supports private deployment of all vision large models QCS8550: Supports private deployment of vision large models including Qwen2.5-VL, InternVL3, etc.				
	AI Painting	K3: Supports private deployment of all image generation models QCS8550: Supports private deployment of the Stable Diffusion image generation model				
	Deep learning	All models: Support traditional network architectures such as CNN, RNN, LSTM, and support various deep learning frameworks such as TensorFlow, PyTorch, PaddlePaddle, ONNX, and Caffe. Support custom operator development and Docker containerization management technology				
Interface Specifications	Internet	8 × 10Gbps SFP+, 1 × Gigabit Ethernet (RJ45, MGMT is used as BMC management network)				
	Console	1 × Console (RJ45, BMC debug serial port, baud rate 115200)				
	Display	1 × VGA (maximum resolution 1080P, BMC management display)				
	USB	3 × USB3.0, 1 × Type-C (OTG)				
	Button	1 × Power, 1 × UID, 1 × Recovery, 1 × Reset				

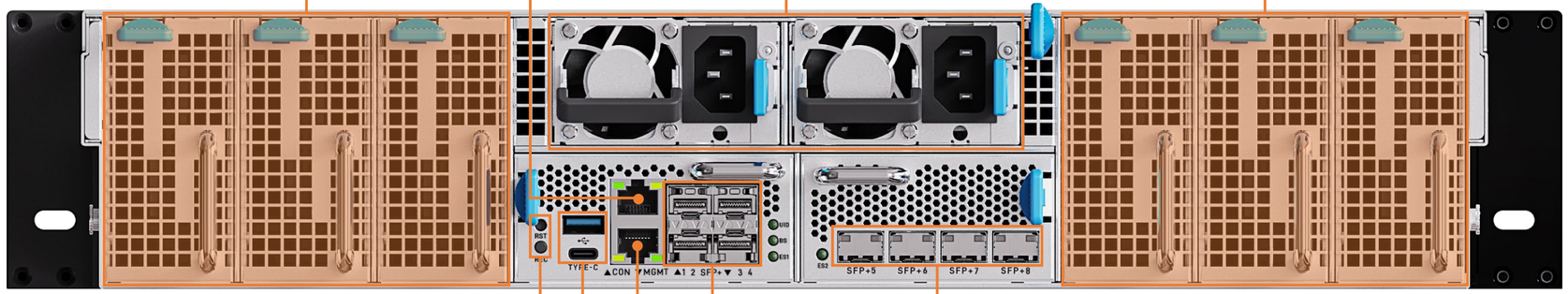
Interface description

Front view

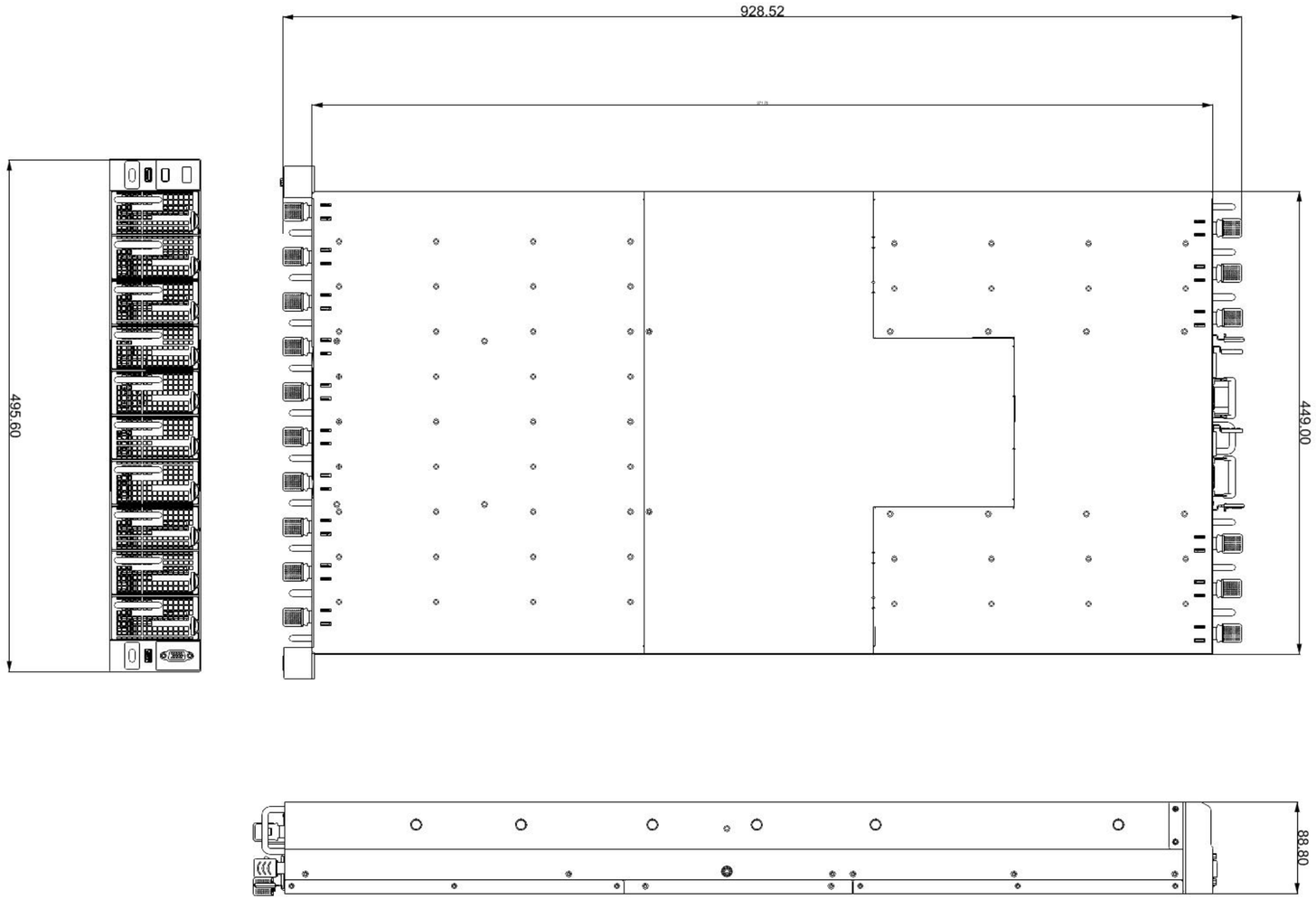


Interface description

Rear view



Dimension



FIREFLY 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.