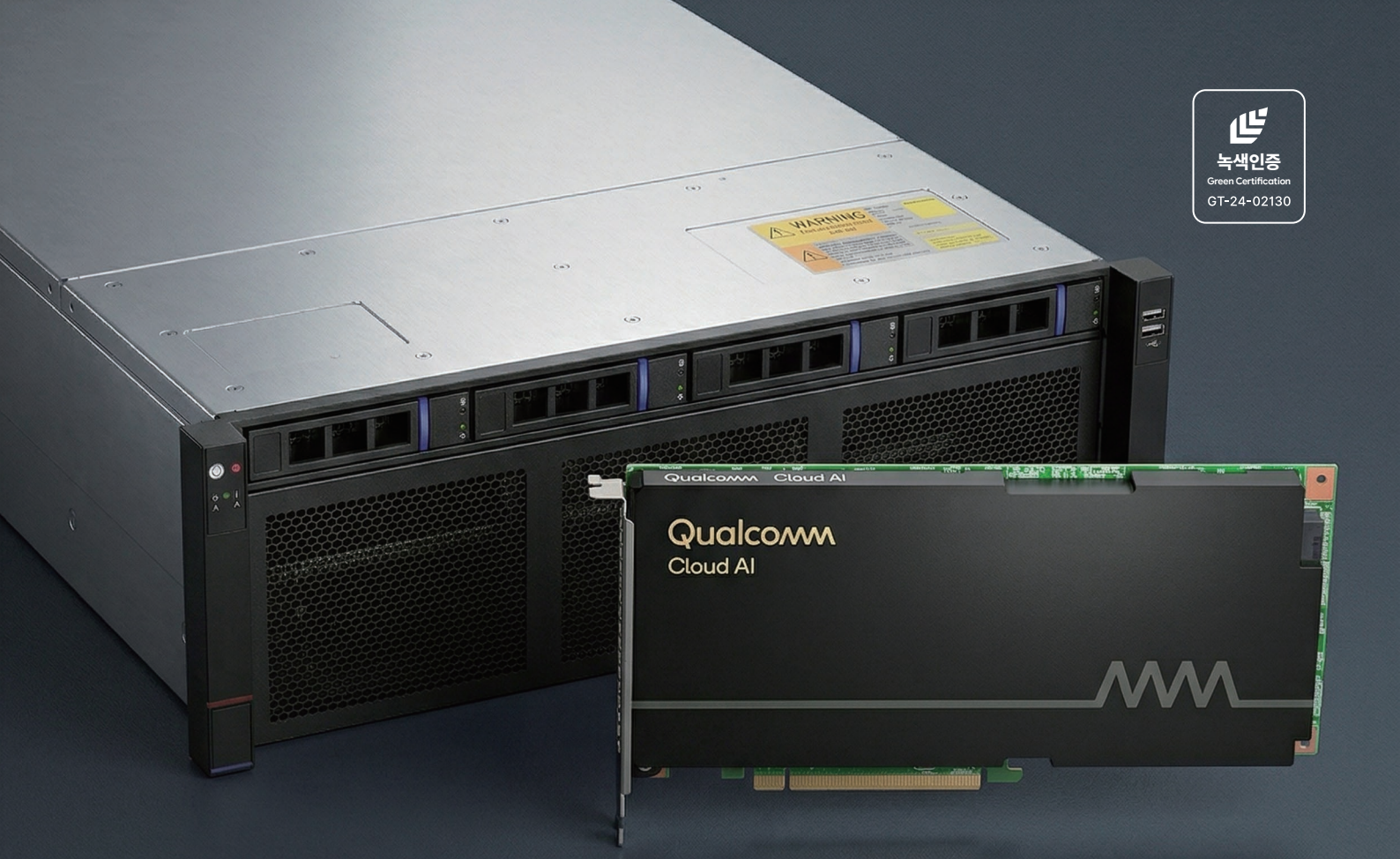


녹색인증
Green Certification
GT-24-02130

V-RAPTOR™ Q100

High Performance AI Inference Arm Server

GPU-less / Accelerator-based / Inference

Qualcomm

4U | 128C

The Hard Truth About AI GPU Servers

- GPU servers are **costly & overkill** for inference.
- Substantial **power & budget are wasted** on inference.
- In real AI service environments, **inference dominates** the overall workload.

Benefits of Adopting XSLAB's AI Inference Servers

- **Lower initial investment** for inference servers compared to GPUs.
- Significantly boost **operational efficiency** with high inference throughput.
- Ensure stable service operations through architecture **purpose-built for AI inference.**

AI Inference: The Foundation of A Sustainable Long-term Cost Structure.

XSLAB 엑세스랩

System Specifications

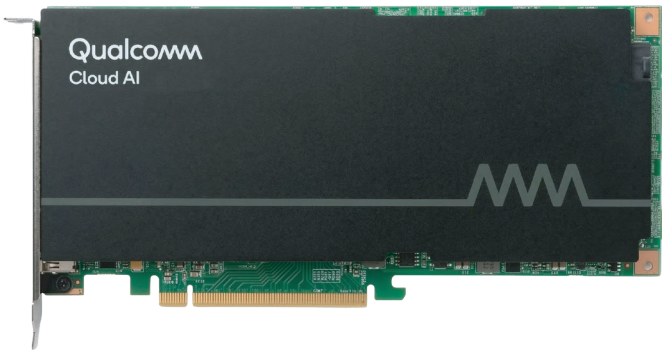
AI Compute Architecture					
CPU	Model	Ampere Altra® / Altra® Max	NPU	Model	Qualcomm® Cloud AI Series
	Core	Up to 128 Cores, 3.0 GHz		Type	AI 100 / AI 080, (Ultra / Std.)
	Instructions	Arm v8.2+, SBSA Lv.4		Performance	Up to 870 TOPs @INT8
	Etc.	1 MB L2 Cache / Core, 7 nm		Etc.	Up to 10x Equipment
System I/O					
RAM	Type	DDR4, 3,200 MT/s, ECC	Storage	Type 1	2.5″ / 3.5″, SATA / SAS
	Channels	8 Ch., 16 DIMM (2 DPC)		Type 2	M.2 NVMe®
	Capacity	Up to 4 TB	I/O	Network	10 GbE Ethernet
Manage	Remote	Self-developed BMC based XMAS™		External	USB 3, PCIe® Gen 4
OS & System Reliability					
OS	Ubuntu	Server 20.04 / 22.04 / 24.04	Reliability	Temperature	0°C - 90°C
	Rocky	Linux 8.10, 9.6, 10.0		Usage	Enterprise-grade Reliability

AI Inference PoC Support

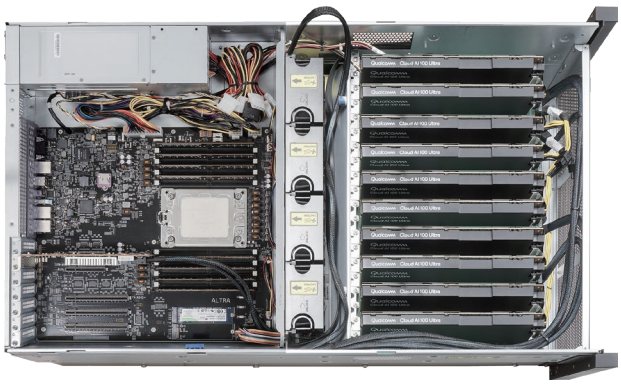
Configuration & Architecture-Consulting

Performance Validation & Benchmarking

Technical Support & Operational Guidance



Qualcomm® Cloud AI Series



V-Raptor Q100

XSLAB 엑세스랩
엑세스랩 주식회사 | XSLAB Inc.

xslab.co.kr

Seoul Office
R-1701, 02, 03, Daeryung Post Tower 8th,
43, Digital-ro 26-gil, Guro-gu, Seoul, Korea.

Jeju Office
402, Dongyeon Tower,
4, Sinhyeong-ro, Jeju-si, Jeju, Korea.

T. +82 02.6952.9974
E. sales@xslab.co.kr