

MACx GPU Accelerator

Model: MACx-01 PCIe AI Accelerator

Scale AI Without Vendor Lock-In

Enterprise-grade AI inference acceleration built for production LLM workloads and **scalable data center infrastructure.**



Why MACx?

- 64GB HBM2e high-bandwidth memory
- Up to 340 TFLOPS AI performance
- Optimized for production AI inference
- Native PyTorch and TensorFlow support

MACx GPUs provide a platform for AI inference and workloads, supporting frameworks such as PyTorch and inference engines like SGLang and vLLM, while enabling flexible and cost-efficient AI infrastructure deployments.

MACx GPUs are a general-purpose accelerator architecture developed by an international consortium of companies from Europe and Asia led by Torriatte Labs (Japan), and distributed in Europe by Toraks d.o.o.

Deployment & Pricing

MACx GPU Pricing

List price: 5.490 € + VAT per unit

GPU Server Deployment & Pricing:

Pre-configured GPU servers available – pricing on request

***Standard warranty: 2 years. For larger volumes please contact us.**

GPU Card Specifications

Performance

FP64: 21 TFLOPS

FP32: 42 TFLOPS

TF32: 170 TFLOPS

BF16: 340 TFLOPS

FP16: 340 TFLOPS

INT8: 680 TOPS

Video Decode

90 streams H.264 / VP9

80 streams H.265

110 streams AVS2

Memory

Capacity: 64GB HBM2e

Bandwidth: 1.8 TB/s

High-bandwidth memory optimized for AI inference workloads

Interface

PCIe 5.0 x16

Form Factor

Full-height, full-length PCIe card

Power

400W air-cooled

Server Specifications

CPU: AMD EYPC 9474F*2

Memory: DDR5 64G*8

HDD: 480G 2.5 SATA 6Gb R SSD*2

Raid Controller: 4GB 8ports Raid Controller

Ethernet Controller: Dual-port 1GB Ethernet

Controller+Dual-port 200G NDR

Controller+Dual-port 25G SFP28 Fiber Ethernet Controlle



“ MACx AI accelerators are already **successfully deployed in production at Arctur**, powering LLM inference workloads with outstanding reliability. They deliver **exceptional price-performance, remarkable efficiency, and true supply independence**—demonstrating a compelling, real-world alternative to traditional GPU-based solutions.

Tomislav Šubić, Head of AI, Arctur

Test before you buy - free access to production environment available

Testing capacity: up to 3–4 servers simultaneously

Toraks d.o.o.

sales@toraks.si

www.toraks.si

Toraks

