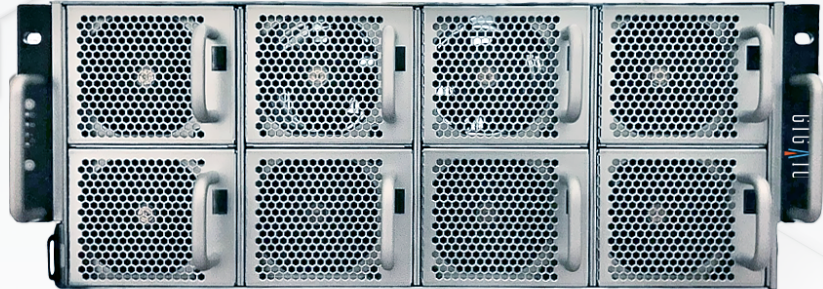


GigalIO™ Accelerator Pooling Appliance - PCIe



CAPACITY

8x double-wide PCIe slots for accelerators

HIGH PERFORMANCE

Ultra low latency 512Gb/s uplinks

ULTIMATE FLEXIBILITY

Any GPU, FPGA, ASIC, NIC or NVMe AIC

SIMPLIFIED DEPLOYMENT

RESTful APIs and WebGUI

IDEAL FOR

- AI/ML training and inferencing clusters
- High-performance computing environments
- Data analytics acceleration
- Composable infrastructure deployments
- Scale-up computing architectures

Specifications

Model	RB5008, Accelerator Pooling Appliance - PCIe, 4U, 8 slot double-wide 675Watt PCIe-slot accelerator	
	GPU to GPU	Fabric uplink
Data rate (each direction)	512Gb/s	4x 512Gb/s (2.048Tb/s total)
Bandwidth (bi-directional)	128GB/s	4x 128GB/s (512GB/s total)
Latency	<120ns*	<120ns
Form factor	4U 19" rack-mountable chassis	
Accelerator slots	8x double-wide x16 PCIe Gen5 FHFL slots Supports up to 675W cards Supports bridge cards	
Fabric ports	8x QSFP-DD PCIe x8 256Gb/s NRZ Supports 2 to 8 uplinks to fabric switches or head node servers	
Fabric port configuration	Each port configured as single x8 (256Gb/s) Aggregate two ports as x16 (512Gb/s)	
Fabric cable support	Direct attached copper (DAC) up to 2m Optical transceivers with passive fiber up to 30m**	
Silicon	Dual ExpressFabric PEX89144	
Management	DMTF Redfish® RESTful API WebGUI Role-based authentication and access control Baseboard Management Controller (BMC) Dedicated micro processor for real-time telemetry	
Power	4x redundant 2+2 or 3+1 3200W power supplies IEC-320-C19 power inlets, 100 to 240VAC, 50 to 60Hz	
Fans	8x heavy duty hotswappable fans Airflow front-to-rear (rack handle to power side)	
Dimensions	17.6" W x 6.9" H x 25.6" D (448mm W x 175mm H x 650mm D)	
Weight	86.6 lbs (39.3 kgs) net w/o accelerator cards	
Environmental	Operating temperature: 10°C to 35°C (50°F to 95°F) Relative humidity: 5% to 95% (non-condensing)	

READY TO GET STARTED?

Contact a GigalIO authorized representative today.

info@gigalio.com

* Best non-bridge GPU-GPU performance between GPU1 through GPU4 and GPU5 through GPU8.

** End-user application latency may shorten range; speak with a representative for more information