

Gryf – The First Ever Suitcase-sized AI Supercomputer

HIGHLIGHTS

PERFORMANCE

AI/ML and HPC at the Edge

AGILITY

Compact case, just roll to the field

CAPABILITY

Modular slots for compute, storage, and accelerators

EDGE INTELLIGENCE

Brings unprecedented intelligence to field deployments



Gryf is a highly flexible and reconfigurable mobile data center system designed to bring powerful AI, machine learning, and high-performance computing capabilities directly to the tactical edge.

Co-designed by GigaIO and SourceCode for portability and scalability, Gryf allows users to quickly transform the vast amounts of sensor data collected at the edge into actionable solutions. It sets a new standard for on-demand configurability in the field, and is unique in its ability to scale to the performance of a field supercomputer, all in the TSA-friendly form factor of carry-on luggage.

This scaling is made possible by GigaIO's groundbreaking FabreX™ AI memory fabric, which unifies edge-to-core infrastructures to dynamically deploy any mission application for actionable real-time intelligence – anywhere.

Performance

The Gryf system is powered by high-performance compute, storage, and accelerator components optimized for AI/ML workloads. In a single Gryf unit, you can achieve up to 91.6 TFLOPS of FP32 performance, 362 TFLOPS of FP16 performance, and 733 TFLOPS of FP8 performance. By combining up to five Gryf units into the planned "SWARM" configuration – used when applications run best on multiple servers – the total performance scales to an impressive 458 TFLOPS of FP32, 1,815 TFLOPS of FP16, and 3,665 TFLOPS of FP8.

Agility

The Gryf system is designed for maximum mobility and ease of deployment. The compact, ruggedized platform features a detachable top with a folding handle and a detachable bottom with wheels, allowing the entire system to be easily transported and rolled into the field. This makes it ideal for rapid deployment in remote or challenging environments.

Capability

The modular architecture of the Gryf system allows it to be tailored to specific mission requirements. The platform features multiple slots that can be populated with a variety of compute, storage, accelerator, and network sleds. Users can mix and match these components to create the optimal configuration for their needs, whether that's focused on AI/ML, high-performance computing, or data-intensive storage.

Edge Intelligence

By bringing powerful AI and HPC capabilities directly to the edge of the network, the Gryf system enables unprecedented levels of real-time intelligence and decision-making in field deployments. This is enabled by GigaIO's FabreX AI memory fabric, which unifies the edge-to-core infrastructure and allows for the dynamic deployment of mission-critical applications.



The Gryf system is designed to provide a highly versatile and capable edge computing platform that can be rapidly deployed and reconfigured to address a wide range of mission requirements, from combat operations to disaster response to scientific research in remote locations. Its combination of performance, agility, and modularity makes it a valuable asset for organizations that need to push the boundaries of what is possible at the tactical edge.

Configurations

Gryf is easily configured and reconfigured to meet your deployment needs by populating the platform with sleds of your choice. Initial fixed configuration options at launch are as follows:

	AI/ML	Compute	Storage
Compute Sled	2	2	1
Accelerator Sled	1	0	0
Storage Sled	2	3	5
Network Sled	1	1	0



The table below illustrates the capabilities of the AI/ML Gryf. When applications run best on multiple servers, planned SWARM mode combines up to five Gryfs for unprecedented performance at the edge.

	Gryf*	x5 Gryfs*
CPU Cores	32	160
System Memory	512GB	2.56TB
Storage	492TB	2.5PB
GPU Memory	48GB	240GB
GPU Performance	FP32: 91.6 TFLOPS FP16: 362 TFLOPS FP8: 733 TFLOPS	FP32: 458 TFLOPS FP16: 1,810 TFLOPS FP8: 3,665 TFLOPS

* Based on AI/ML Configurations with NVIDIA L40S

Specifications

Compute Sled	
CPU	(1) AMD EPYC 7003 Family, 7313, 16-core tested 155W TDP, up to 64-core planned**
System Memory	(4) 64 GB DDR4 DIMMs (512GB total), 3,200Mhz max, 1TB planned (4x256GB)**
OS Storage	(1) 512GB NVMe-M.2 SSD
OS Support	Linux Rocky 8/9 or Ubuntu 20/24, planned support for Windows Server 2019 & 2022
Networking	(2) QSFP56/QSFP28/QSFP+ 100GbE, copper/optical
BMC / IPMI	Via platform RJ45

** Based on system configuration

Accelerator Sled	
Accelerator Slot	Single/Double-wide PCIe-FHFL form factor, up to 350W
Accelerator	(1) Nvidia L40S-48GB (other options to be qualified)
Storage Sled	
Storage	(8) 30TB NVMe-E1.L SSD (246TB total)
Network Sled	
Ports	(2) QSFP56-100GbE, copper/optical (6) SFP28-25GbE, copper/optical



Platform	
Sled Slots	(6) for Compute Sled, Accelerator Sled, Storage Sled, Network Sled in any location
AI Memory Fabric	Internal: PCIe 128Gb/s board-to-board Expansion: (8) FabreX Mini-SAS-HD-32G 32Gb/s (256Gb/s total) for Gryf daisy-chaining or data offload at home base
Management Network	(5) SFP+ 10GbE copper/optical for FabreX Fabric Manager and out-of-band (OOB) baseboard management controller (BMC) / IPMI
Software	GigaIO FabreX Fabric Manager
Power	Dual AC/DC 2,500W 1+1 power supplies IEC-320-C20 power inlet, 100-240 VAC @ 50 to 60Hz
Fans	(6) 60mm fans dynamically optimized for system workloads
Fan Filters	Removable 45 PPI filters
Dimensions	With wheels: 9.00" x 14.00" x 24.50" (228.6mm x 355.6mm x 622.3mm) Without wheels: 9.00" x 14.00" x 22.00" (228.6mm x 355.6mm x 558.8mm)
Weight	Less than 55 lbs (24.95 kgs) max
Enclosure	Ruggedized carbon fiber, detachable top with folding handle, detachable bottom with wheels
Environment	Operating temperature: 10°C to 32°C (50°F to 90°F)
Serviceability	Field replaceable units (FRUs): sleds, power supplies, fan tray with fan filtration, case top and bottom covers
Compliance	FCC Class A, CE, Trade Agreements Act (TAA) / Made in the USA, IP55 in transport

About GigaIO

GigaIO provides workload-defined infrastructure through its AI memory fabric, FabreX, which seamlessly composes rack-scale resources and integrates natively into industry-standard tools. FabreX lets customers build "Impossible Servers" for HPC + AI workflows.

Learn more at gigaio.com/products/gryf