

## Gryf – The First Ever Suitcase-sized AI Supercomputer

### HIGHLIGHTS

#### PERFORMANCE

AI/ML and HPC at the Edge

#### AGILE

Compact case, just roll to the field

#### CAPABILITY

Modular slots for compute, storage, and accelerators

#### ENABLE THE EDGE

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, 733 TFLOPS of FP16 performance, and 1,466 TFLOPS of FP8 performance. By combining up to five Gryf units into a "SWARM" configuration, used when applications run best on multiple servers, the total performance scales to an impressive 458 TFLOPS of FP32, 3,665 TFLOPS of FP16, and 7,330 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.

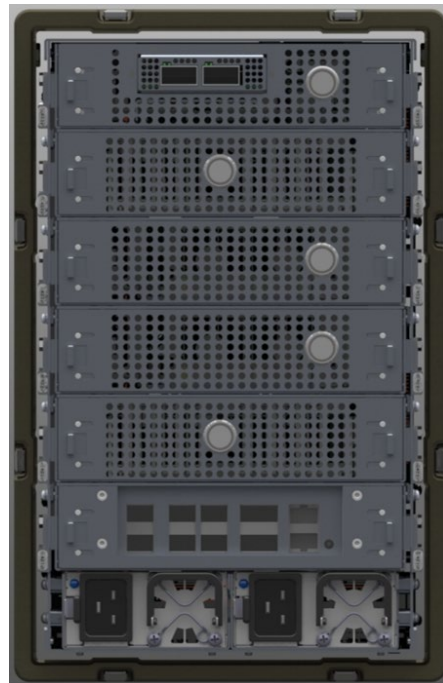


## Capability

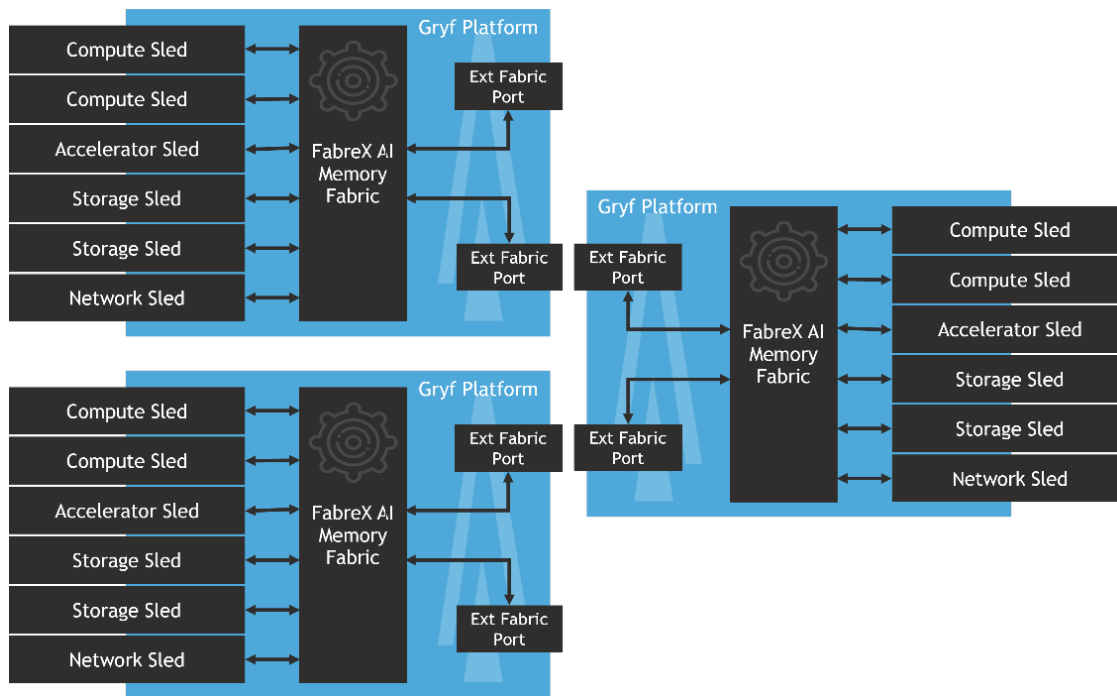
The modular architecture of the Gryf system allows it to be tailored to specific mission requirements. The platform features multiple slots that can be populated with a variety of compute, storage, accelerator, and network sleds. Users can mix and match these components to create the optimal configuration 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.



*Interconnect up to five Gryfs across FabreX, enabling any resource on one Gryf to access resources on the other connected Gryfs.*



## Configurations

Gryf can be easily configured and reconfigured to meet your ongoing deployment needs. Simply populate the platform with powerful sleds to accomplish any task:

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

Combine up to five Gryfs for SWARM mode, when applications run best on multiple servers, to achieve unprecedented edge results:

	Gryf*	x5 Gryfs*
<b>CPU Cores</b>	64	320
<b>System Memory</b>	0.51TB	2.56TB
<b>Storage</b>	246TB	1.23PB
<b>GPU Memory</b>	48TB	240TB
<b>GPU Performance</b>	FP32: 91.6 TFLOPS FP16: 733 TFLOPS FP8: 1,466 TFLOPS	FP32: 458 TFLOPS FP16: 3,665 TFLOPS FP8: 7,330 TFLOPS
<b>AI Fabric Performance</b>	256Gb/s	1.28Tb/s

\* based on AI/ML Configurations

## Specifications

Compute Sled	
<b>CPU</b>	(1) AMD EPYC 7003 series, 7313, 16 cores/32 threads, 155W, 3Ghz base/3.7Ghz turbo
<b>System Memory</b>	(4) 64 GB DDR4 DIMMs (256 GB total), 3,200Mhz max
<b>OS Storage</b>	(1) 512GB NVMe-M.2 SSD
<b>OS Support</b>	Linux Rocky 8/9 or Ubuntu 20/24
<b>Networking</b>	(2) QSFP56/QSFP28/QSFP+ 100GbE, copper/optical
<b>BMC / IPMI</b>	Via platform RJ45
Accelerator Sled	
<b>Accelerator Slot</b>	Single/Double-wide PCIe-FHFL form factor, up to 350W
<b>Accelerator</b>	(1) Nvidia L40S-48GB (other options to be qualified)



<b>Storage Sled</b>	
<b>Storage</b>	(8) 30TB NVMe-E1.L SSD (246TB total)
<b>Network Sled</b>	
<b>Ports</b>	(2) QSFP56-100GbE, copper/optical (6) SFP28-25GbE, copper/optical
<b>Platform</b>	
<b>Sled Slots</b>	(6) for Compute Sled, Accelerator Sled, Storage Sled, Network Sled in any location
<b>AI Memory Fabric</b>	Internal: PCIe 256Gb/s board-to-board External: (8) FabreX Mini-SAS-HD-32G 32Gb/s for Gryf daisy-chaining or data offload at home base
<b>Management Network</b>	(5) RJ45-10GbE/1GbE for FabreX Fabric Manager and out-of-band (OOB) baseboard management controller (BMC) / IPMI
<b>Software</b>	GigaIO FabreX Fabric Manager Preloaded Linux & libraries: GPUDirect RDMA (GDR), Libfabric, and NVMe-oF
<b>Power</b>	(2) IEC-320-C13 power inlets, 100-240 VAC @ 50 to 60Hz, dual AC/DC 2,500W 1+1 power supplies
<b>Fans</b>	(6) 60mm fans dynamically optimized for system workloads
<b>Fan Filters</b>	Removable 45 PPI filters
<b>Dimensions</b>	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)
<b>Weight</b>	Less than 55 lbs (24.95 kgs) max
<b>Enclosure</b>	Ruggedized carbon fiber, detachable top with folding handle, detachable bottom with wheels
<b>Environment</b>	Operating temperature: 10°C to 32°C (50°F to 90°F)
<b>Serviceability</b>	Field replaceable units (FRUs): sleds, power supplies, fan tray with fan filtration, case top and bottom covers
<b>Support</b>	3 or 5 years; Service & Support: Basic - best effort warranty replacement Service & Support: Plus - next business day advanced shipment in USA, express air for international
<b>Compliance</b>	FCC Class A, CE

## 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](https://gigaio.com/products/gryf)