GIGAIO FabreX™

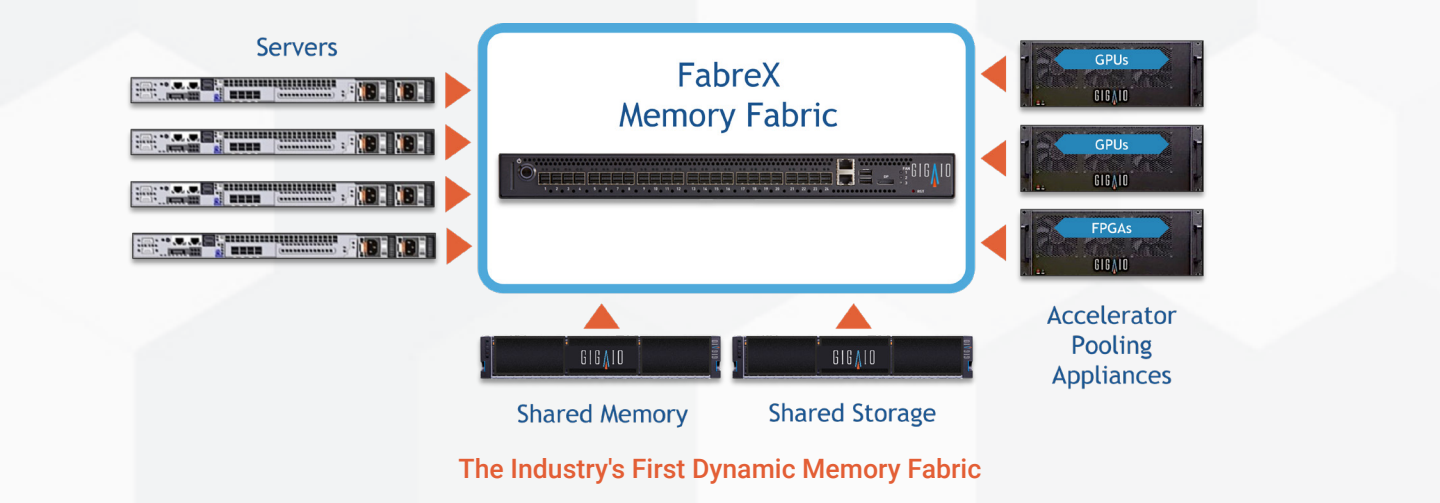
The Industry's First Dynamic Memory Fabric

- Works With Your Existing Schedulers and Tools
- Doubles GPU and FPGA Utilization
- Precise CPU/Accelerator Configurability
- World's Lowest Latency Interconnect
- Lowers Lifecycle Costs

DATA SHEET

The Industry's First Dynamic Memory Fabric

At the heart of GigaIO is FabreX, our patented dynamic memory fabric. FabreX disaggregates CPUs, GPUs, storage, and other processing resources trapped in confined servers into a single unified rack-scale system connected by a high-speed, low-latency memory fabric. FabreX transforms all resources connected to the fabric into unique memory resources, using the same PCI Express (PCIe) and CXL technology used within every industry-standard server. What makes FabreX unique is its use of Direct Memory Access (DMA) to transfer data between the memory of the various processing elements connected to the fabric.



The Industry's First Dynamic Memory Fabric

FabreX is the only technology available today that enables DMA memory-to-memory transfers between server nodes, between servers and devices like NVMe storage, GPUs, and FPGAs, and between devices. GPUs and FPGAs can transfer data directly to other GPUs and FPGAs via DMA – a compelling use case for today's heavily accelerator-dependent applications. GigaIO FabreX breaks the server barrier to enable the components trapped within the chassis to be connected at similar latency and bandwidth outside the server enclosure. This new architecture enables a unified, software-defined, standards-based, memory-centric composable infrastructure that can be configured to exactly meet dynamically changing, accelerator-rich application requirements.

Native Integration with Your Existing Tools to Simplify Administration through Open APIs

Accessing all the power of GigaIO FabreX is as simple as using the same cluster management, workload scheduling managers, and DevOps scripting tools as you always have. Instead of developing yet another management interface, GigaIO's strategically developed native integrations with the industry's leading cluster scheduling and management tools using Open Standard Redfish APIs. NVIDIA's Bright Cluster Manager is bundled with every GigaIO system to compose discrete computational elements into highly functional and reconfigurable systems, allowing multiple workflows to co-exist on the same infrastructure.

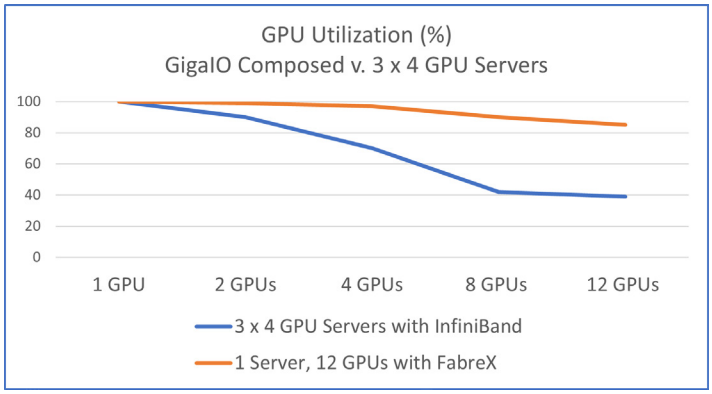
Bright Cluster Manager
(comes bundled)
includes integrations with:
- Slurm
- OpenPBS
- PBS Pro
- LSF
- Altair Grid Engine
- Kubernetes

CTRL IQ
- Fuzzball

Quali
- Cloudshell

Double GPU and FPGA Utilization

GPUs and FPGAs drive computational performance today, but they are trapped in servers, frozen in their configuration, lowering their utilization to as little as 15% – a significant waste of expensive resources. GigaIO frees GPUs and FPGAs by connecting dozens directly, peer-to-peer, economically pooled within dedicated enclosures and composed on the fly over low-latency FabreX. Results show over 100% improvement in GPU utilization, cutting time to completion in half.



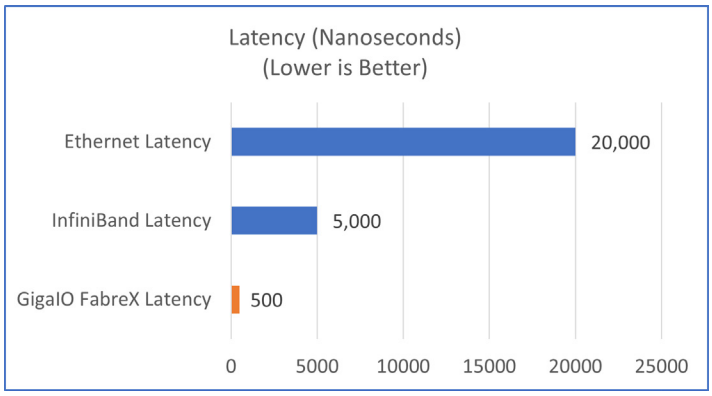
GPU Utilization More Than Doubles with GigaIO

Precise CPU/Accelerator Configurability

Whether an application needs large numbers of GPUs or FPGAs connected to a single CPU node or many CPUs connected to a few GPUs or FPGAs, GigaIO composes the exact configuration needed to meet dynamically changing, accelerator-rich, mixed workload requirements.

World's Lowest Latency Interconnect

GigaIO delivers the industry's lowest latency interconnect, accelerating performance-sensitive applications like deep learning AI, quantitative trading, genomic analysis, analytics, signal processing, electronic design automation, and computational fluid dynamics.



Ultra Low Latency FabreX is 10X Faster Than InfiniBand

Lower Lifecycle Costs

A complete computing platform for accelerated HPC can be built out of a few simple FabreX building blocks. Buy the precise GPU and CPU configurations you need, knowing you are doubling your performance compared to conventional GPU servers connected with InfiniBand. Eliminate stranded and underutilized resources. And down the road, you can upgrade or add individual servers, storage, and accelerators at the component level that plug-n-play with your environment. Every major subsystem can now operate on its own upgrade cycle. And the system's total cost is optimized over its lifecycle, as FabreX drives much higher utilization of all resources.

Our Remarkable Customers

GigaIO customers include leaders in HPC and AI research, financial technology, aerospace, analytics, and national defense.

Collins Aerospace · U.S. Dept of Defense · EmpowerAI · Ericsson · Fujitsu · IBM · intel

Lawrence Livermore National Laboratory · MITRE · Morgan Stanley · NeuroBlade · Oxford Nanopore Technologies · NVIDIA · Oak Ridge National Laboratory

SDSC San Diego Supercomputer Center · Supermicro · Sysnet Data Solutions · TACC Texas Advanced Computing Center · Wisconsin University of Wisconsin-Madison · VISA · 元大證券 Yuanta Securities

About GigaIO

GigaIO provides a workload-defined infrastructure through its dynamic 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 — from storage to accelerators to memory — at a fraction of cloud TCO, by optimizing the utilization and efficiency of their existing hardware, allowing them to run more workloads faster at lower cost through higher utilization of resources and more agile deployment.