



Shave Time, Make Money: GigalO Delivers More Sophisticated Inference for High-frequency Trading

SOLUTION BRIEF



GigalO's FabreX™ memory fabric bypasses the CPU, eliminating round-trip data latencies with peer-to-peer FPGA communication at linear scale for more sophisticated inference, to identify multi-market opportunities with greater determinism and higher order-to-trade ratios.

Key Challenges

.....

The world's most active proprietary high-frequency trading (HFT) firms and leading electronic market makers all have one thing in common: maintaining a competitive technological advantage is a constant arms race. In a business where success is measured exclusively on the HFT time horizon, ongoing infrastructure upgrades that reduce latencies by a few nanoseconds can cost millions.

Yet besides cost, another challenge is scaling those upgrades without negatively affecting latency. Field programmable gate array (FPGA)-based trading systems are considered the state-of-the-art in HFT. The high levels of parallelism and determinism enabled by FPGA solutions significantly accelerate the computation of mathematical models and the transmission of data to exchanges' matching engines.

This fully programmable technology enables certain types of trading algorithms to operate up to 1,000 times faster than traditional software solutions designed to accommodate generic-purpose, CPU-based compute.

Scaling FPGA-based systems, however, presents its own challenges. Traditionally, data in transit to the FPGA must communicate via the CPU, introducing roundtrip latency that diminishes the value of the accelerator. Bypassing non-essential software and generic-purpose hardware to deliver true FPGA-to-FPGA communication requires highly specialized, custom hardware and software systems, and upgrades cannot be implemented ad hoc, which can lead to potential interruptions in critical trading activities.

Beat the Street with Ultra-Low Latency FabreX Memory Fabric

- ⬡ Improve bid-ask spreads, order-to-trade ratios, and overall trade volume and market share across ultra-low latency FabreX
- ⬡ Improve time series analysis on price, time duration
- ⬡ Increase liquidity for more dynamic trading
- ⬡ Support more ticker feeds across major equities, futures, and options trading venues
- ⬡ Execute book building and consolidation at the sub-microsecond level
- ⬡ Initiate up to 3x more trades per day by keeping all transactions in one shared memory space
- ⬡ Write larger, more complex AI algorithms to identify even the smallest market opportunities for higher returns
- ⬡ Predict stock movement across thousands of tickers
- ⬡ Identify actionable intelligence amid extreme market volatility

Deploy FabreX to:

Gain unprecedented market advantage by enabling direct FPGA-to-FPGA communications, bypassing the CPU

Make the arms race for small improvements in bandwidth and latency irrelevant by scaling up to 32 or more FPGAs to one node directly across FabreX memory fabric

Achieve FPGA-to-FPGA latency as low as 250 ns roundtrip with further performance improvements at scale

Achieve CXL 3.0-like latency and data speeds years before the protocol is in wide production

Design a simpler, software-defined FPGA trading system with FabreX technology

Build the leanest possible system by uncoupling FPGA, GPU, high-performance storage, intelligent networking, and other accelerators from the CPU

Architect non-latency tasks to different hardware, such as GPU, discreetly across FabreX

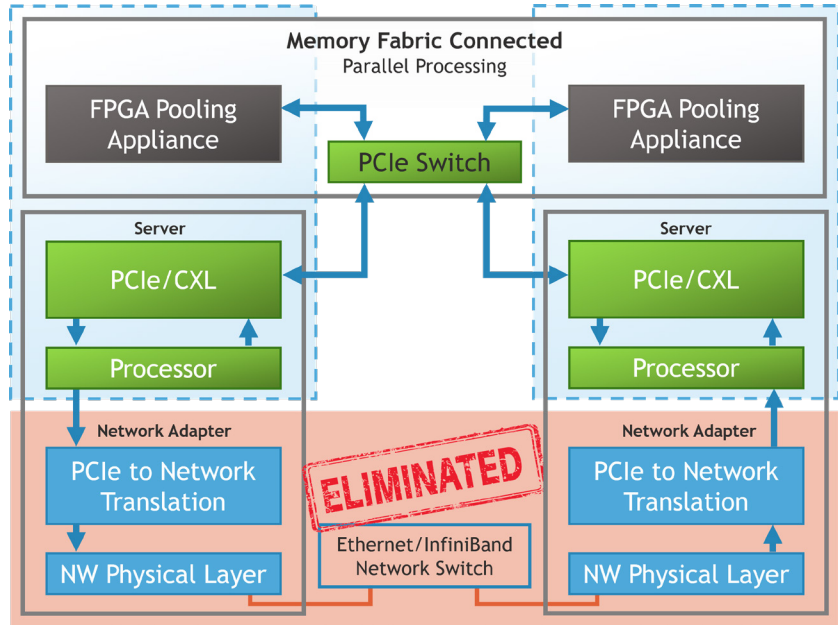
Deploy any additional PCIe-based technologies across FabreX in real time as required, eliminating unnecessary hardware and costly downtime

Gigalo's FabreX for Apex Performance in FPGA-based Trading

By enabling peer-to-peer communication between FPGAs and other networked devices, FabreX memory fabric from Gigalo bypasses data translations down the stack to the CPU entirely, eliminating the problem of roundtrip data latency even before it is introduced. This is accomplished by creating a memory interconnect between all fabric-connected FPGAs that provides direct memory access between all devices without the overhead of CPU and transports.

Further, because device aggregation is enabled across the memory fabric, traditional architectural roadblocks to pooling FPGAs as a single memory resource are also avoided, meaning the resource can be scaled linearly for unprecedented performance bandwidth. This provides the flexibility to deliver just one FPGA per node or as many as 32 or more FPGAs per node. These hardware changes can be made dynamically depending on the workload needs at any given time.

Enabling nanosecond-level inference at the edge of the compute cycle, FabreX provides proprietary trading firms and market makers game-changing performance advantages at the network level as well as the front and back office messaging levels.



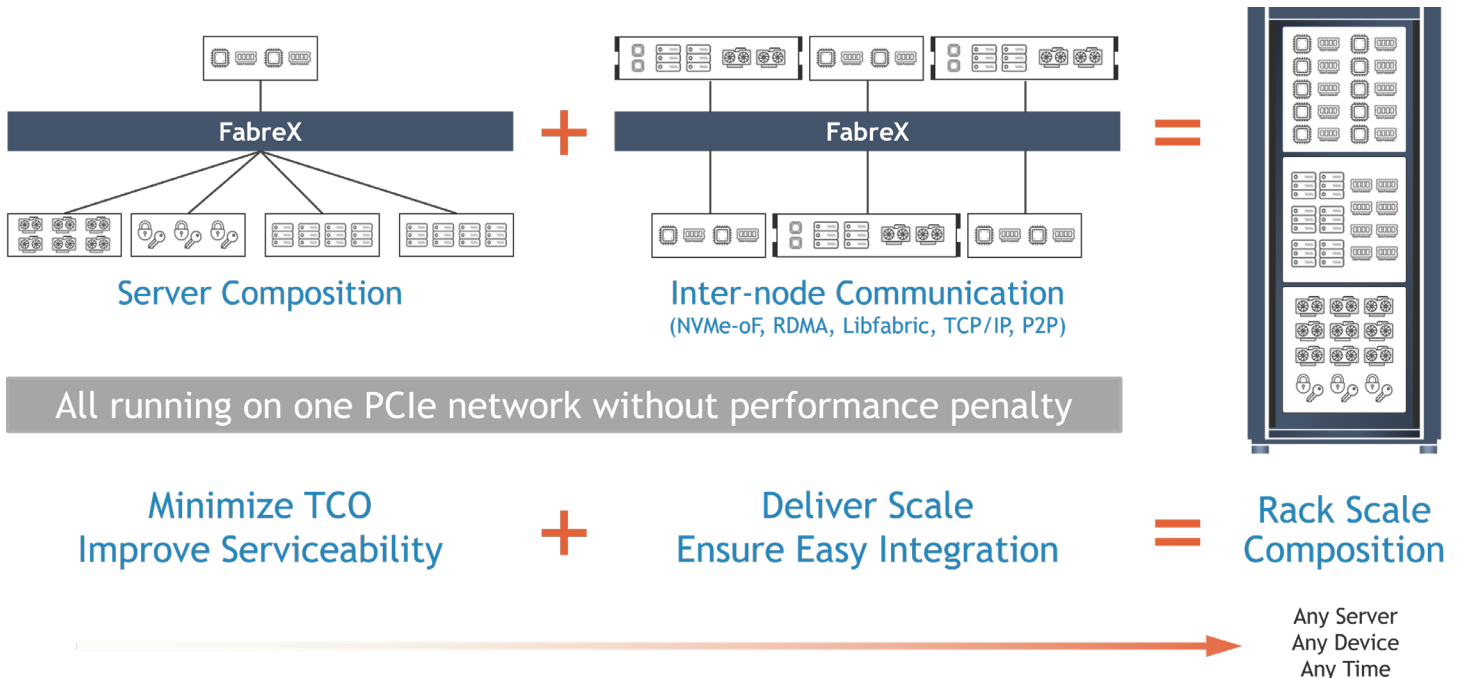
The Closing Bell: Architect FabreX for Superior HFT Performance

In addition to device-to-server, FabreX is unique in enabling server-to-server and device-to-device communication for FPGA and other accelerators, all across the same high performance PCIe memory fabric. FabreX can span multiple servers and multiple racks to scale up single-server systems and scale out multi-server systems, all unified via the FabreX software.



FabreX

One universal dynamic memory fabric carries all traffic.



FabreX delivers the industry's leadership execution with FPGA-based trading systems that set firms apart from their competition. Effortlessly connect new memory and storage products, FPGAs, and new or emerging accelerators and your choice of processors either directly attached or via server configurations like NVMe-oF.

With exceptional low latency and high bandwidth for FPGA-based trading, FabreX is built to be an open platform and offers robust, industry-standard APIs at all levels of

the software, making the integration of FabreX into your existing data center management systems easy and risk free.

Bypassing roundtrip data translation between CPUs and FPGAs, FabreX memory fabric from GigaIO delivers top execution for FPGA-to-FPGA-based trading with unprecedented linear scalability for inference on immense data sets at the edge of the compute cycle.

FabreX from GigaIO delivers the greatest depth of support in storage, compute, and FPGA-based transaction acceleration, with the highest density, industry-leading latency, and best performance for FPGA-based trading systems.

For proprietary trading firms and market makers looking to supercharge their infrastructure to gain a competitive edge for more effective execution, FabreX delivers a compelling way to implement state-of-the-art performance.