Smart
Our Switchtec™ Gen 4 PCIe switches are GPU-optimized with low pin-to-pin latency and low latency variation for optimal machine learning workload performance.

Connected
Switchtec features include up to 100 line-rate capable PCIe lanes as well as an integrated high-performance cut-through Direct Memory Access (DMA) engine.

Secure
Switchtec's high-reliability features include hot- and surprise-plug support, end-to-end data integrity and debugging with the ChipLink diagnostics tool.

GigaIO™ Introduces PCI Express® (PCIe®) Network Fabric Solution for the Age of AI

The emergence of Artificial Intelligence (AI) and the associated Machine Learning (ML) and Deep Learning (DL) applications are fueling demand for fundamental changes in the creation of compute and storage clusters. Faster and larger storage arrays and a rapid proliferation of specialized compute accelerators, like Graphics Processing Units (GPUs), Field-Programmable Gate Arrays (FPGAs) and Application-Specific Integrated Circuits (ASICs), are creating bottlenecks and configuration problems for the interconnect systems. Traditional networks were never designed to handle the performance requirements of these workloads and devices.

MICROCHIP

SMART | CONNECTED | SECURE

microchip.com
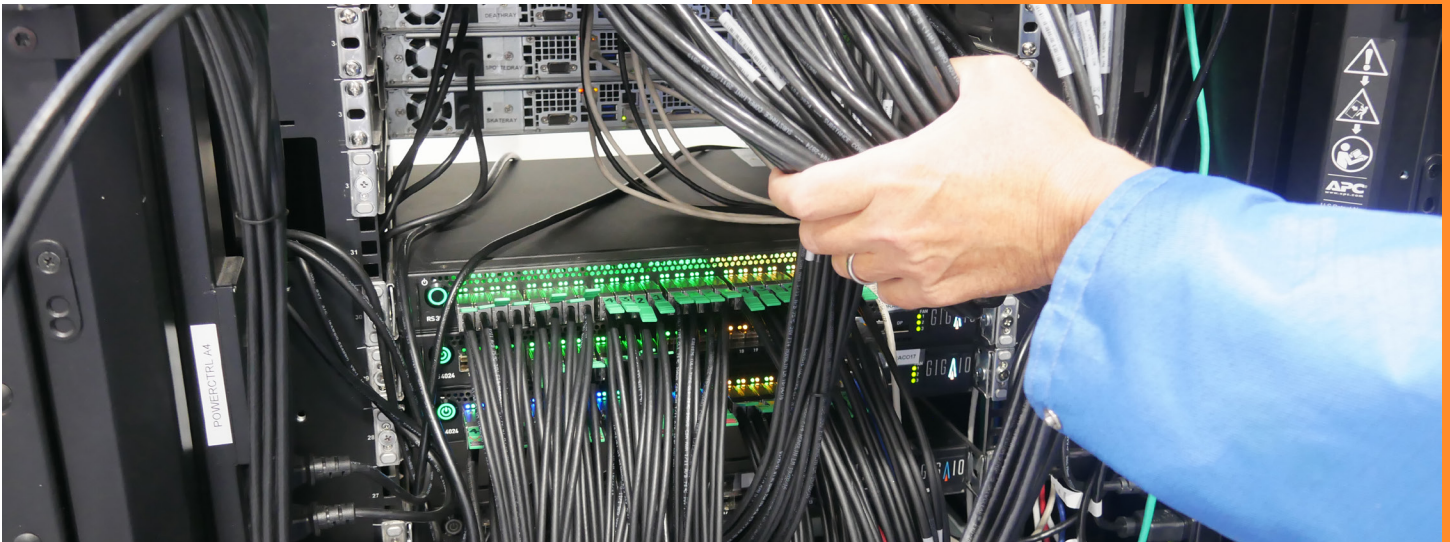
Today's data streams are far wider and deeper than ever, with workflows so varied that they may combine different infrastructure setups within a single workflow. Today, a single workload may require different types of accelerators: NVIDIA® V100 Tensor Core V100 GPUs for compute projects, NVIDIA GeForce® RTX graphics cards for visualization or FPGAs for encryption, all within the same workflow.

In a traditional data center, managers must guess and decide at server refresh time what hardware their users will need for the next three to five years and provision for peak demand, which means resources may sit idle—the GPU industry estimates that the average utilization rate is around 15 percent—despite being some of the most expensive resources in the rack.

Having made sizable hardware and software investments to pursue their High-Performance Computing (HPC) and AI data center initiatives, managers often look to allocate expensive new resources such as GPUs and FPGAs found inside servers to other workloads during idle periods.

The philosophy of resource sharing is nothing new. Today's network architects and administrators have practically grown up with the concept of virtualization, where data center resources have transitioned beyond physical servers in a fixed location to cloud-based virtual environments. Highly inefficient designs of the past where server hardware resources were allocated only to specific operating systems and enterprise applications seem severely outdated. Modern approaches not only allow servers to operate near capacity but they also reduce power requirements, operating costs and downtime. Beyond these efficiencies and economies of scale, virtual environments even make enterprise software applications perform better.

But today's advanced AI-driven data center models come with a tax on efficiency. High-dollar investments in compute resources often result in ultra-powerful accelerators remaining idle during down periods when they aren't attacking today's AI workloads.

Echoes of the past are being heard again.

Half a century ago, large corporations and universities were making multi-million-dollar bets on mainframe computers. The "big iron" managers who had to justify the unprecedented spending to the executive suite continually needed to prove that their goliath machines were producing economies of scale. Resources that sat idle were their enemies as well.

"Buy the largest computer you could afford and find ways to maximize the amount of work done on it" was a common refrain. [1]
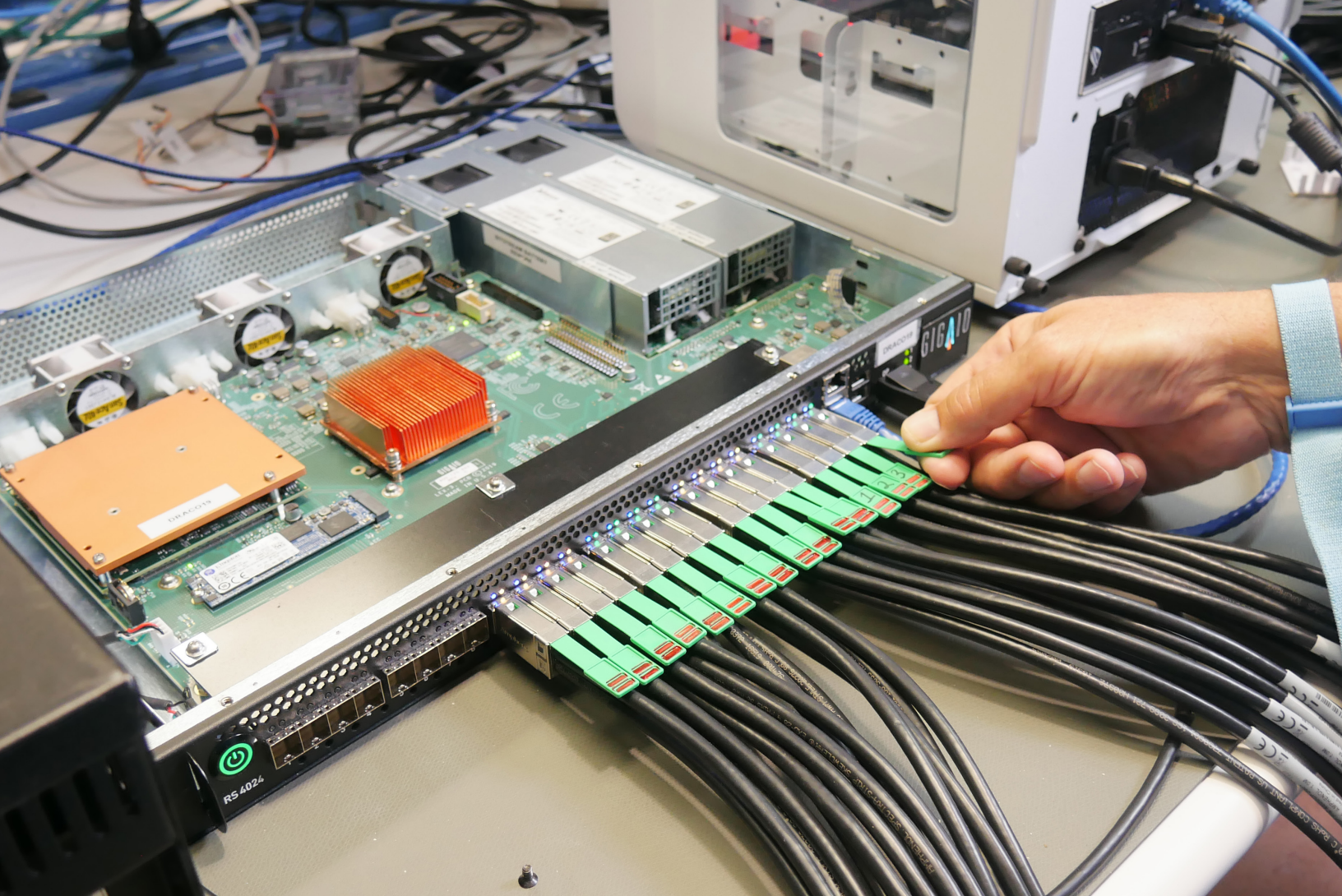
The Rise of Composable Disaggregated Infrastructure (CDI)

CDI, under the umbrella of SDI (Software-Defined Infrastructure), has garnered growing interest in the data center. Composable infrastructure disaggregates compute, storage and networking resources into shared resource pools that can be available for on-demand allocation (i.e., "composable").

CDI is a data center paradigm with a software layer that abstracts a hardware layer consisting of IT resources across a virtual fabric and organizes them into logical resource pools, not dissimilar to a Hyper-Converged Infrastructure (HCI) cluster. But rather than having a linked-up series of self-contained nodes of compute, storage, network and memory as in HCI, composable infrastructure is a rack-scale solution like Converged Infrastructure (CI). Unlike CI, however, composable infrastructure is, at its core, an infrastructure-as-code delivery model with an open and extensible unifying API layer that communicates and controls hardware resources.

The remarkable increase in the amount of data being collected, analyzed and stored is driving the rapid adoption of advanced data analytics and AI that is challenging the fundamental architectures of today's data centers. This challenge explains much

of the move to the public cloud, which promises the ultimate in resource flexibility but comes with a high price tag, especially for HPC and AI workloads.

The rapid pace of change in acceleration technology and AI software fuels the necessity for highly flexible and easily upgradeable architectures capable of incorporating new hardware technology without forklift upgrades to expensive equipment. This means breaking the server box and disaggregating elements of the traditional server into separate pieces that can be easily shared. But to effectively disaggregate storage and accelerators, the interconnects must support both exceptional low latency and high bandwidth.

Data center managers also want to drive high utilization of expensive new storage and acceleration components to keep the costs of Capital Expenditure (CapEx) and Operating Expenses (OpEx) down.

Today's advanced scale computing, enterprise, cloud and edge data centers need both scale-up and scale-out resources across the cluster and require a network technology that will scale accordingly.

The PCIe Standard: Evolution at Work

PCIe is a widely deployed bus interconnect interface commonly used in server platforms. It is also being increasingly used as a storage and application accelerator interconnect solution.

With the cooperation and support from more than 800-member companies that represent the modern connectivity ecosystem, the PCI Special Interest Group (PCI-SIG) manages the development of PCI specifications as open industry standards. The organization has seen the PCIe standard grow to become a widespread solution over the course of two decades. PCI-SIG advocates like to point out that Input/Output (I/O) bandwidth doubles every three years, and over the course of five generations of the standard, PCIe is now being deployed to serve devices and applications that could not have been imagined during the early days of its development.

The popularity of display adapters for 3D graphics and video and the evolution of new storage paradigms in the early 2000s led PC-based developers to seek alternatives for predecessor connectivity standards that couldn't deliver the computer-to-peripheral communications speed and performance that users required. The very first PCIe standard offered a common protocol for transferring data at 250 Mb/s.

As industry demands for higher-performing I/O are being met by each succeeding generation of the standard, the PCI-SIG mandates backwards compatibility, cost-efficiency, high performance, a processor-agnostic stance and a high degree of scalability.

> *"PCIe is seen as a dominant data transfer protocol that has expanded device support to include Add-In Cards (AICs), network adapters, Network Interface Cards (NICs), acceleration cards, FPGAs and NVMe™ Flash drives."*

The breadth of its influence can be seen well beyond traditional compute environments as its footprint now extends from PC, server, storage and cloud sectors into mobile, IoT and automotive connectivity.

Enter GigaIO

With any fundamental shift in computing, networking or storage technology, new entrants typically emerge to take advantage of market opportunities before they reach mass adoption. Such is the case with GigaIO Networks, Inc. of Carlsbad, CA. The company was established in 2017 by networking and high-performance computing veterans who assembled a highly skilled team with deep network architecture, hardware, software and silicon development capabilities. The team invented the first truly composable cloud-class, software-defined universal infrastructure fabric, which empowers users to accelerate workloads on demand, using industry-standard PCIe technology. GigaIO's patented technology optimizes cluster and rack system performance in a way that also reduces the Total Cost of Ownership (TCO) for customers. They have quite literally taken PCIe outside the box with a software-centric approach to enable on-demand resources for AI workloads.
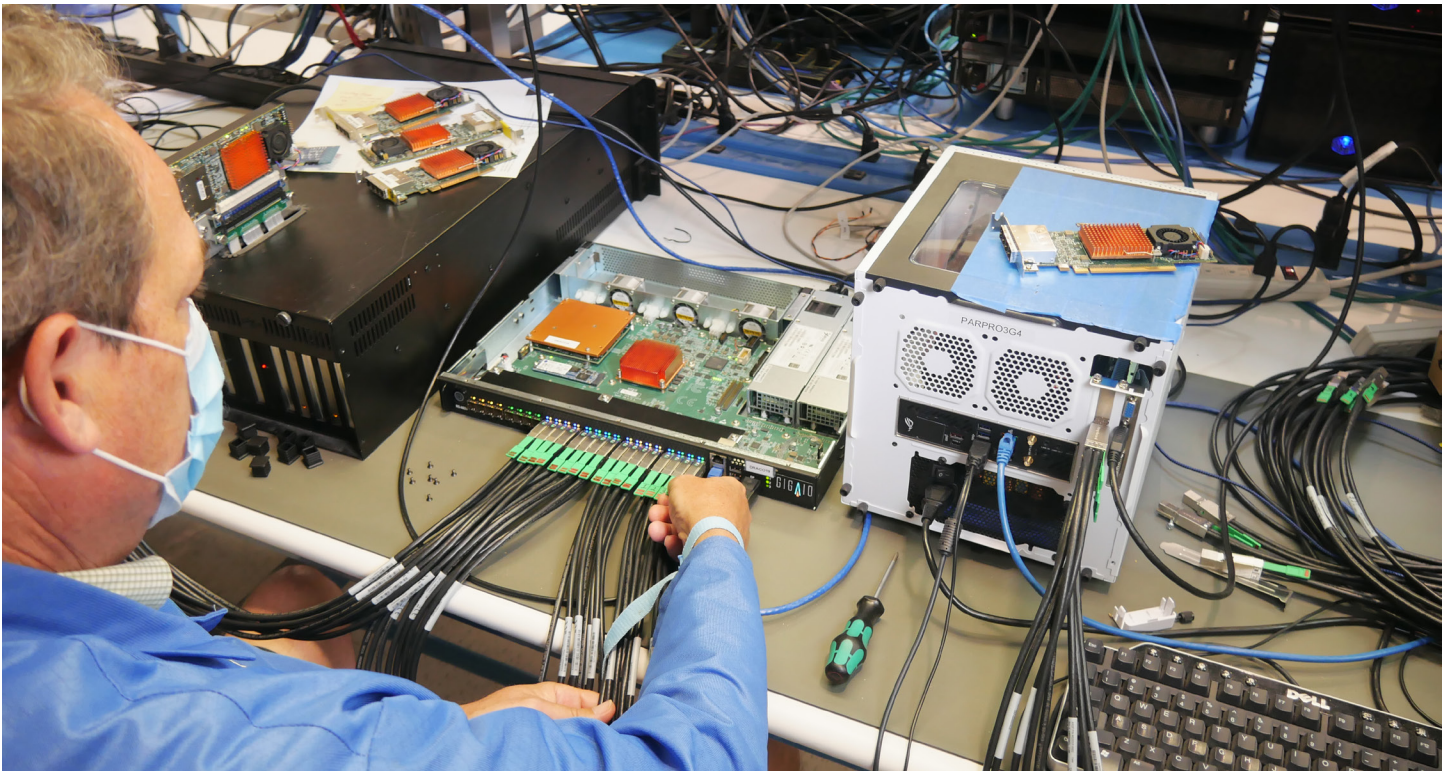
The Challenge

With deep ties to the high-performance computing and storage communities and a vision for expanding the role of PCIe connectivity, the team at GigaIO recognized a gap in the market. The demand for GPU-based solutions and accelerator hardware to address AI-driven workloads was increasing. Applications for intense data processing were crossing over from research institutions and into the enterprise. Geophysical modeling, weather forecasting, bioinformatics, physical simulations, edge computing, fintech trading platforms, advanced visualizations, analytics and cloud computing all use advanced heterogeneous computation architectures to enable faster execution of compute-intensive tasks. The rise of real-world AI and ML has boosted GPU and related hardware sales. But what if these new assets could be leveraged on-demand to address unprecedented workloads? While they offer peak performance when in use, they often sit idle.

By asking the right questions, the GigaIO team addressed the challenge to architect a rack-scale composable infrastructure solution that delivers the unlimited flexibility and agility of the cloud but at a fraction of the cost. They identified a series of foundational requirements:

- The solution would need to orchestrate any compute, acceleration (GPUs, FPGAs, ASICs), storage, memory or networking resource for any workload using an enterprise-class, easy-to-use and open standard, high-performance network.

- It would need the ability to deploy, expand, reduce or replace all rack resources dynamically in real time.

- It would need to fully automate bare-metal, cloud-class infrastructure, delivering higher throughput at a fraction of the cost.

The Solution

GigaIO collaborated with Microchip on a Gen 4 PCIe switch appliance targeting HPC and AI data center managers looking to free expensive resources (such as GPUs and FPGAs) trapped inside servers. In the past, to share these resources between their users, IT managers had to install not one but two or three networks in a rack. Even when servers were communicating to storage and accelerators over PCIe, IT managers had to incur the additional expense, overhead and latency hit of another network, like Ethernet or InfiniBand, to communicate server to server. With FabreX™, a single universal PCIe fabric can connect servers without resorting to InfiniBand or Ethernet. As a result, the Microchip and GigaIO collaboration delivers a solution with the industry's lowest latency.

Microchip's Switchtec™ Gen 4 PCIe switches enable customers to build next-generation, high-performance, low-latency interconnect solutions in high-growth markets, including ML, data center servers and storage equipment.

This new generation of PCIe switches, which offer high density, reliability and low power, provide customers a quick time-to-market solution with field-proven Switchtec technology firmware and a chip architecture enabling significant re-use of customers' investments in Switchtec technology management software, drivers, firmware and system design.

Switchtec Gen 4 PCIe switches are GPU-optimized with low pin-to-pin latency and low latency variation for optimal ML workload performance. Other features include up to 100 line rate-capable PCIe lanes and an integrated high-performance cut-through Direct Memory Access (DMA) engine. High-reliability features include hot- and surprise-plug support, end-to-end data integrity and best-in-class debugging with the ChipLink diagnostics tool.

Switchtec PCIe switches are ideal for GigaIO's FabreX design as a result of their low latency, high bandwidth, low power and dynamic configurability.

The GigaIO team also chose open standard Redfish® APIs (instead of vendor-proprietary APIs as other CDI vendors do), enabling programmatic and template-driven control and automation so that admins could rapidly configure and reconfigure disaggregated IT resources to meet the requirements of specific workloads and DevOps. This flexibility, low TCO and speed are, in part, the reason many analysts expect the composable infrastructure market to explode over the next few years.
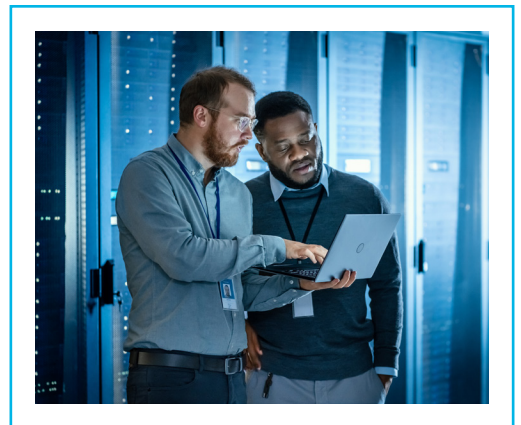
The Result

With its innovative GigaIO FabreX open architecture, data centers can scale up or scale out the performance of their systems, enabling their existing investment to adapt as workloads and business change over time.

GigaIO has introduced the first native PCIe Gen4 network fabric, which supports GDR, MPI, TCP/IP and NVMe over Fabrics (NVMe-oF). FabreX technology transforms rack-scale architectures to enable complete software-defined, dynamically reconfigurable rack-scale systems, eliminate system waste, reduce maintenance burdens and improve performance.

FabreX is a fundamentally new fabric architecture that integrates computing, storage and other communication I/O into a single-system cluster fabric. GigaIO enables true server-to-server communication across PCIe and makes true cluster scale networking possible with DMA by an individual server to system memories of all other servers in the cluster fabric for the industry's first in-memory network.

This new architecture enables a hyper-performance network with a unified, software-defined, composable infrastructure.

With its exceptional low latency and high bandwidth, FabreX makes the disaggregation of storage and accelerators possible. And by offering standard Redfish APIs, leading third-party orchestration and composition software that can easily run on FabreX, GigaIO creates a true software-defined infrastructure that dynamically assigns resources to match changing workloads.

By embracing this approach, IT managers are finding that they do not need to make massive, costly overhauls to server resources simply to incorporate new technology. With FabreX, a single unified PCIe fabric can connect servers without resorting to InfiniBand or Ethernet. As a result, the Microchip and GigaIO collaboration delivers a solution that democratizes AI and HPC by delivering the industry's lowest latency to more users for the same investment.

> *"GigaIO's FabreX technology is revolutionizing both data centers and edge computing by enabling the applications to dynamically change the infrastructure to match the requirements of specific workflows. Customers get the flexibility of the cloud, at a fraction of the cost."*
>
> *Alan Benjamin, CEO*

Endnotes
1 Arms, William Y., The Early Years of Academic Computing. The Internet-First University Press. June 2014.
www.cs.cornell.edu/wya/AcademicComputing/text/earlytimesharing.html and www.ecommons.cornell.edu/handle/1813/36926?show=full page 1.11. Accessed August 20, 2021.

Microchip Technology Inc. | 2355 W. Chandler Blvd. | Chandler AZ, 85224-6199 | microchip.com

MICROCHIP