GIGAIO

No Performance Penalty with Composed Architecture

Configuration

GigaIO FabreX Composed Configuration 2S 1VE (use 2 VE)

Vector Engines inside the Accelerator Pooling Appliance and shared between all servers on FabreX
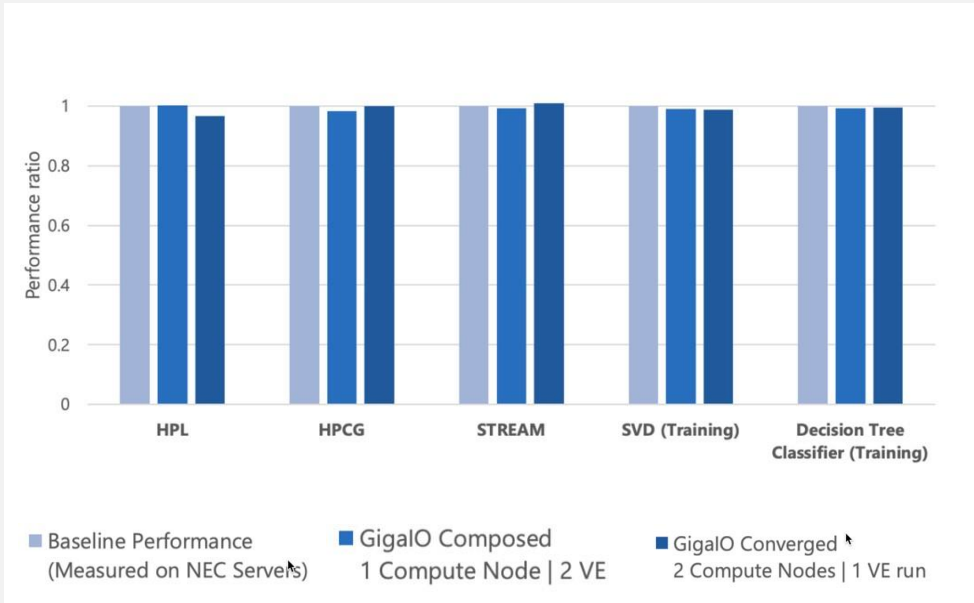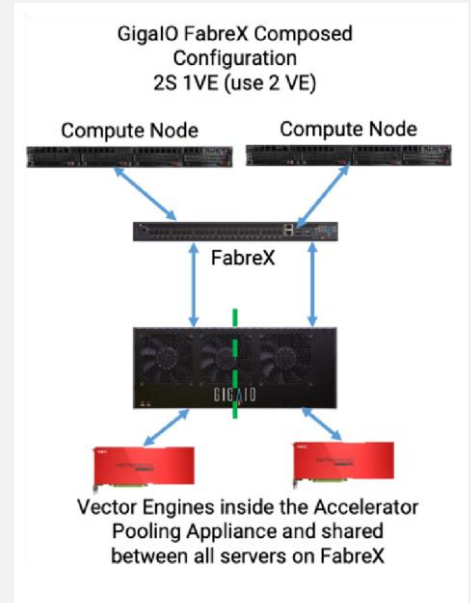
Native PCIe Network Fabric Optimizes Utilization of NEC Vector Engine Cards for Accelerated Data Center Workloads

Converged configuration – The converged system uses a single server with two NEC SX-Aurora cards on the internal PCIe bus. The cards cannot be shared.

Composed configuration – The GigaIO composed architecture tested had a server connected to an Accelerator Pooling Appliance with two NEC Vector Engine accelerator cards. The server accessed the accelerators via the FabreX™ network fabric at PCIe latency and bandwidth.

Benchmarking software – Five benchmarks were used: the HighPerformance Computing LINPACK (HPL), the High-Performance conjugate Gradients (HPCG), STREAM, Singular Value Decomposition (SVD), and Decision Tree Classifiers.

Result summary – The composed configuration enabled the SXAurora Vector Engines to be shared between multiple servers and reconfigured dynamically. Performance results were identical in all configurations. FabreX composable architecture with NEC Vector Engines delivers performance without compromise.

Modern AI workloads perpetually grow and change, so data center architecture needs to stay flexible to support changing business needs. Deploying a FabreX composed architecture, easily reconfigured via software orchestration, makes it possible to accommodate those changing requirements, without incurring a performance penalty for the added flexibility.

GigaIO FabreX is a Rack-Scale composable infrastructure solution that delivers the unlimited flexibility and agility of the cloud, at a fraction of the cost. Benefits include:

Improved system agility by disaggregating system resources on the fly and creating shared resource pools that can then be dynamically composed in real-time.

Slashed Total Cost of Ownership by enabling device sharing which increases resource utilization and eliminates over provisioning, resulting in reduced CapEx and OpEx.

Simplified and automated system set-up, administration and serviceability with freedom of choice for management tools from powerful CLI and Redfish APIs to ready-to-run, off-theshelf enterprise-class orchestration software. Seamless support for any PCIe-compliant device including servers, CPUs, memory, 3D-XPoint, storage, GPUs, FPGAs, specialty ASICS and NICs.

Blazing system performance with industry-leading PCIe latency and bandwidth throughout the rack and beyond. As PCIe resources are added they immediately benefit from the native PCIe performance as all data transfers and buffers are completely eliminated.

Visit www.gigaio.com to discover more about GigaIO and FabreX, the industry's only pure PCIe Network Fabric.