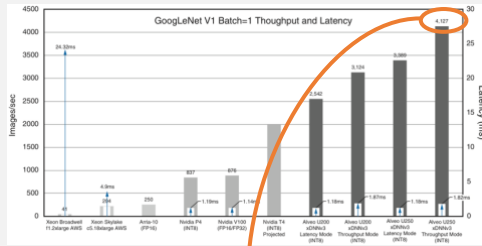
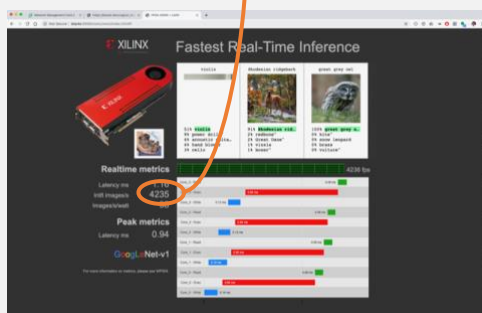


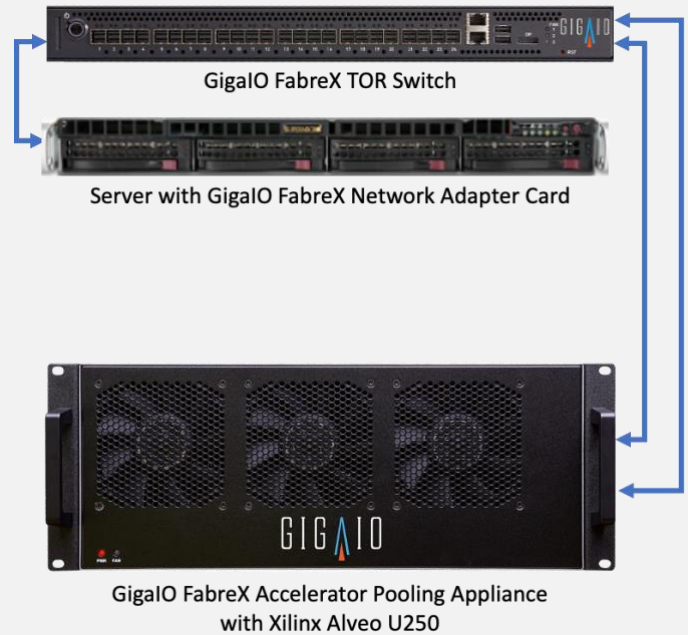
No Performance Penalty with Composed Architecture



Same >4,000 Images/Sec Performance Internal PCIe (top) vs. FabreX (bottom)



Configuration



Native PCIe Network Fabric Optimizes Utilization of Accelerator Cards for AI Data Center Workloads

Converged configuration – In the converged configuration: the FPGA accelerator is trapped inside the server and cannot be shared. This system uses a 1U server with a Xilinx Alveo U250 FPGA card inside, connected through the internal PCIe bus.

Composed configuration – GigaIO composed architecture is a single 1U server connected to a single Accelerator Pooling Appliance that houses the Xilinx FPGA accelerator card, which accesses the server via the FabreX™ network fabric, at PCIe latency and bandwidth.

Benchmarking software – The benchmark uses GoogLeNet V1, a 22-layer deep convolutional neural network (CNN). The network trained on ImageNet classifies images into 1000 object categories, such as keyboard, mouse, pencil, and many animals.

Result summary –The converged result is 4,127 images/sec and the composed result is 4,235 images/sec, essentially identical. GigaIO FabreX composed architecture delivers higher performance compared to converged architecture – 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-the-shelf 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.