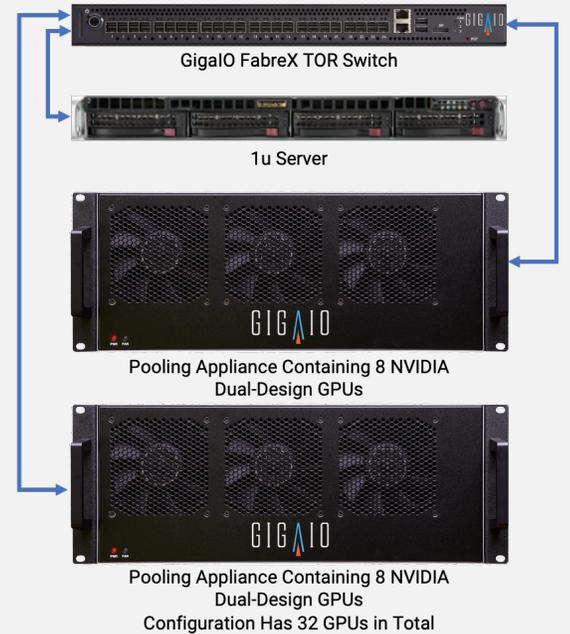
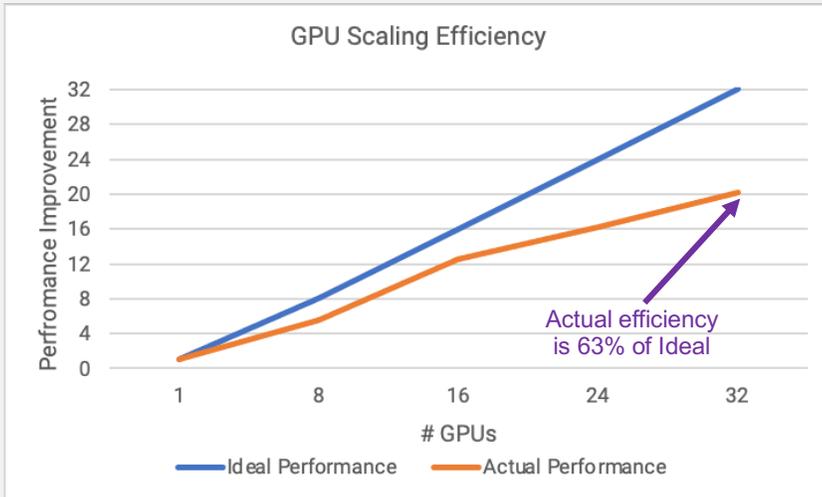


FabreX Demonstrates Exceptional Performance, Scaling and Efficiency with Support for 32 GPUs

Configuration



FabreX™ Industry First: Supporting 32 GPUs With Near Linear Performance, Scaling and Efficiency

Configurations – The GigaIO composed architecture is a single 1U dual-socket AMD Rome server connected to the FabreX PCIe Network fabric with two Accelerator Pooling Appliance (APA). Each APA contains 8 NVIDIA Dual-Design GPUs for a total of 32 GPUs.

Software – The GPU software used is NVIDIA Collective Communication Library (NCCL) supporting multi-GPU and multi-node communication primitives optimized for NVIDIA GPUs and FabreX.

Application tested is the industry standard ML/DL image processing ResNet-50.

Objectives – The objective is two-fold 1) can a single server support 32 GPUs, and 2) what is the scaling efficiency of a single server 32 GPU farm.

Results – For the first objective this benchmark clearly demonstrates that GigaIO FabreX can support a GPU farm of 32 GPUs, **an industry first**. For the second objective, the ideal scaling for a 32 GPU farm would be 32x single GPU, representing perfect linear scalability. The FabreX configuration's actual performance for 16 GPUs is over 12.5x, a 78% efficiency, and for 32 GPUs the actual performance is over 20x a single GPU, a 63% efficiency when compared to the ideal scaling efficiency of 32 or 100%.

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.