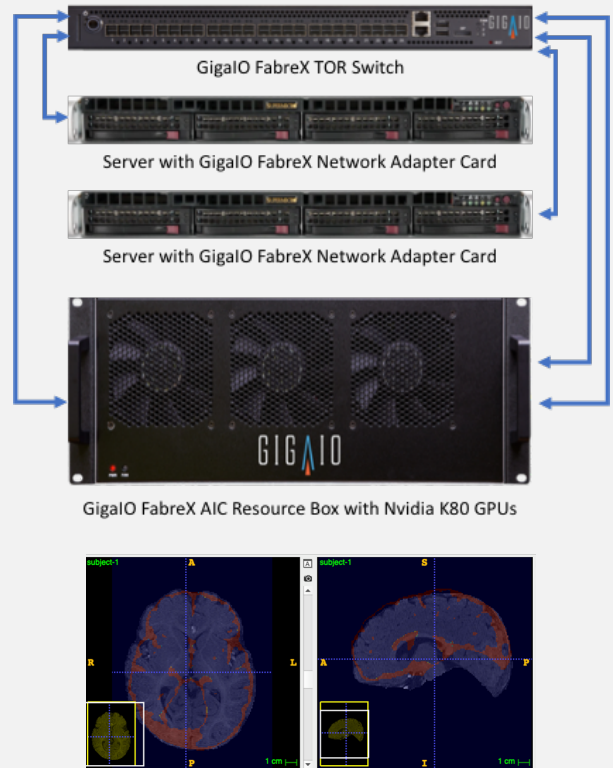
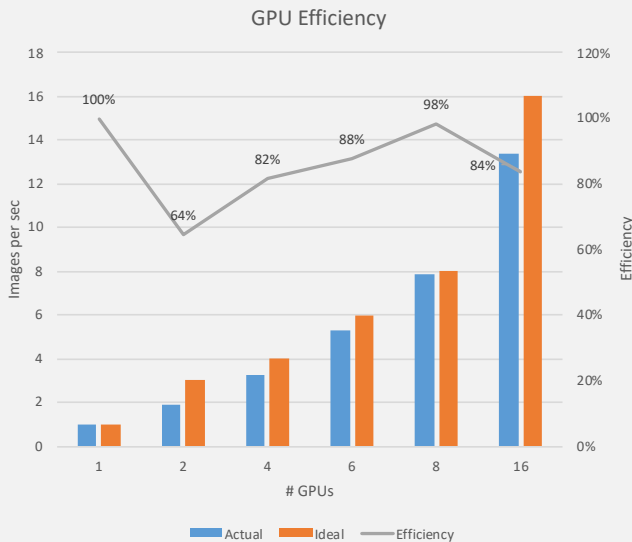


Near-Linear scaling of performance from 1-16 GPUs  
Highly efficient GPU RDMA across network



## PCIe Fabric Unleashes GPU Performance and Scale for Medical AI Training and Inference Applications

### Unrivaled NVMe PCIe Performance:

- GigalIO FabreX™ creates a native PCIe network fabric offering the same performance as internal PCIe slots

### Seamless Support for AI S/W Stack:

- NCCL and RCCL libraries for broad GPU vendor support
- AI Frameworks e.g. Distributed TensorFlow
- Minimal Application S/W rework

### Composable Infrastructure with Peer-to-Peer and RDMA:

- GPUs can communicate with each other in the same PCIe domain using peer-to-peer DMA transfers
- GPUs and SSDs can communicate with each other between PCIe domains using highly optimized RDMA transfers

**Scale without Limits:** Servers have limitations to how many GPUs they can accommodate before running into power, thermal, physical or BIOS limitations. FabreX enables disaggregation of GPUs and offers the ability to scale beyond the server's physical limitations. Leverage multiple servers and multiple GPUs for distributed applications.

**Extreme Performance:** Scale does not need to be realized at the expense of performance. FabreX enables near linear performance improvement as applications leverage an increasing number of H/W resources.

**Optimized Cost:** FabreX is fully PCIe standard compliant network and allows users to minimize costs by supporting all PCIe standard CPUs, GPUs and SSDs while maximizing their utilization.

*In collaboration with University of Southern California*