The Fabric for Software Defined Infrastructure
The Complete Network and Fabric Solution

Software

FabreX Link Cards

FabreX™ Switches

Cables

Resource Boxes
Advanced Scale Computing

Utilizes a “cluster” of resources

...whether a ½ rack, a full rack or a complete row

AI/ML/DL  Biotech, Genomics  Oil & Gas  Edge  Media & Entertainment  Financials
**Old Paradigm**
Compute with servers
- Rigid configurations

**New Paradigm:**
Compute at Rack Scale
- Flexible Configurations

- CPUs
- Computational storage
- Application specific offloads
- Network
- NVME/Storage

**InfiniBand**
Fibre Channel

---

Ethernet evolution from an interconnect to a communications network
The FabreX™ Advantage

• Tremendous performance advantage in latency and bandwidth

• Superb flexibility in both scaling systems up and scaling systems out – with the exact same interconnect
FabreX Switch

Fabric Manager
- Creates resilient fabric capable of sharing all attached resources
  - APIs to user management interface (Rest APIs (ex. Redfish)
- Fabric set-up & management; data plane control

Data Plane
- PCIe Switch Chips
- PCIe Endpoints
- 128Gb/sec
- 256Gb/sec

Management Network

Server 1
- CPU
- Memory
- Root Complex
- PCIe Switch
- PCIe Endpoint

Server N
- CPU
- Memory
- Root Complex
- PCIe Switch
- PCIe Endpoint

128Gb/sec
256Gb/sec

FabreX Card

FabreX Card

www.gigaio.com
GigaIO Company Confidential
FabreX: Unlimited Flexibility

- Ultra-High Performance Network
FabreX: Unlimited Flexibility

- **Ultra-High Performance Network**
- **Super Server creation**
FabreX: Unlimited Flexibility

- **Ultra-High Performance Network**
- **Super Server creation**
- **Memory semantic capable – today!**
SDSC Testing
## Project #1 – Benchmarks and Real World Application

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SimpleP2P</td>
<td>NVIDIA program which measures bandwidth of data transfer using simple CUDA GPU to GPU data movement commands. Demonstrates that GPU Direct functions.</td>
</tr>
<tr>
<td>Memory bandwidth</td>
<td>A NVIDIA utility. The results reported are for: Host to Device (H2D) and Device to Host (D2H).</td>
</tr>
<tr>
<td>Monte Carlo</td>
<td>A broad class of highly parallel algorithms that rely on repeated random sampling to obtain numerical results used in many scientific and business applications.</td>
</tr>
<tr>
<td>AWP-ODC</td>
<td>AWP simulates the dynamic rupture and wave propagation that occurs during an earthquake. Developed by Dawei Wu – SDSC.</td>
</tr>
<tr>
<td>cuNCC</td>
<td>Memory bandwidth benchmark, demonstrates performance of LEX vs SGI UV 300 NUMAl ink7 interconnect.</td>
</tr>
</tbody>
</table>
Project #1 - Summary

• **FabreX Price/Performance is excellent**
  • At list: Rincon - $64K; Comet (2 nodes) - $140K
  • Rincon performance ~ 25% better
    • FabreX drives higher utilization
    • FabreX bandwidth was superior

• **Systems are very different**
  • Head to head comparison not realistic
Configuration Detail (Look Inside the Box):

Comet Node – 2x Intel CPUs 2.5GHz, 2x 160 GB SSD, 4x P100 GPU

GigaIO “Rincon” – Intel CPU 2.4GHz, NVMe JBOF, 12 x GTX1080Ti

Approx. price -- $140K

Approx. price -- $64K
SimpleP2P

- Measures bandwidth of data transfer GPU to GPU using simple CUDA
- Results favor Rincon
- Demonstrates that GDR works and FabreX bandwidth
Memory Bandwidth

- The results reported are for:
  - Host to Device (H2D)
  - Device to Host (D2H)
- Results favor Comet
Monte Carlo – Rincon vs. Comet

- Strong and Weak scalability tests favors Rincon
  - Lower bar is better (faster)
- 12 GPUs on Rincon very efficient
AWP-ODC Strong Scalability—Rincon vs. Comet

### Performance vs. Ideal

- **Category 1**: Comet Ideal, Rincon Ideal, Comet, Rincon
- **Category 2**: Comet Ideal, Rincon Ideal, Comet, Rincon
- **Category 3**: Comet Ideal, Rincon Ideal, Comet, Rincon

### Efficiency vs. Ideal

<table>
<thead>
<tr>
<th></th>
<th>1 GPU</th>
<th>2 GPUs</th>
<th>4 GPUs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comet</td>
<td>100%</td>
<td>89%</td>
<td>68%</td>
</tr>
<tr>
<td>Rincon</td>
<td>100%</td>
<td>102%</td>
<td>92%</td>
</tr>
</tbody>
</table>
AWP-ODC Weak Scalability – Rincon vs. Comet

Performance vs. Ideal

Scaling Efficiency
cuNCC – Rincon vs. SGI Bandwidth Benchmark

- I/O bandwidth versus SGI UV300 Big Memory Supercomputer
  - 32 socket Intel
  - NUMAlink 7 interconnect
  - 8 PCIe Gen3
  - 24 TBytes DRAM
  - 128 TBytes Flash
  - 2 x Tesla M40 GPUs
  - $1M + system
- Rincon almost 2x SGI writing to memory
PEARC18 Scientific Event
Practice and Experience in Advanced Research Computing

- Dawei presented at PEARC18
  - “Deep Learning for Seismic Template Recognition”

- Rincon achieved 24.8 TFLOPS with 12x GTX1080Ti GPUs

- IO achieved 13.3 GB/s bandwidth on 12x NVMe SSD
  - Sustained bandwidth 83% of peak
Project #2 - Researching Visualization Over a High Performance Interconnect Fabric

- Investigate the performance of large-scale, post-simulation visualizations using a visualization node, SSD array, and GPUs connected over a FabreX interconnect fabric.

- System configuration and setup the same as Project 1
  - Open Source software
    - Visualization Toolkit (VTK)
    - ParaView - data analysis and visualization application which built on VTK.
    - ParaView Catalyst - *in situ* use case library orchestrating simulation and analysis and/or visualization tasks
Project #2 - Results

- Demonstrated flexibility to rearrange the computing resources according to different type of jobs.
  - Using FabreX, were able to offload the computational workload to the NVidia GPUs while assigning the visualization workload to the server GPU.
- GigaIO FabreX cluster provided disaggregation and composability, avoiding computational resource waste and boosting efficiency.
Results -- Visualization in Real-time
Next Steps

• Overlap IO with computation using OpenMPI

• Implement GPU direct output to SSDs with GDR

• Significant increase in dataset size – multi-TBytes
SDSC / GigaIO Relationship

• GigaIO has been an Industrial Partner for a year
• Successfully completed 2 projects
• Starting to show that a new approach to architecture enables better results to be achieved in less time with dramatically reduced cost
• Demonstrates the value that academic supercomputer centers can deliver in working with early stage technology companies in the testing and evaluation of new technologies.
  • contributing to the overall “innovation economy”
  • bringing new capabilities to advance scientific research.
• GigaIO looking forward to continued collaboration.
Scale to One – The Right Tool

BREATHTAKING PERFORMANCE
Super low latency and high bandwidth
Eliminates overheads

MEMORY SEMANTICS
Load and Store across the fabric

EXTREME FLEXIBILITY
True disaggregation
Dynamic composability

DO MORE WITH LESS
Less complexity, power, cooling, CAPEX, OPEX

STANDARDS-BASED
100% compliant with PCI SIG
Investment protection
For more information visit:

www.gigaio.com