Building Scalable HPC Clusters with Voltaire InfiniBand

Asaf Wachtel, Senior Product Manager

© 2010 Voltaire Inc.

Voltaire 40Gb/s InfiniBand Portfolio

Fabric provisioning and performance monitoring

Application Acceleration

40Gb/s InfiniBand Switching Platforms

4036
36 x IB ports

4036E
34 x IB ports + 2 x 1/10GbE

4200
162 x IB ports

4700
324/648 x IB ports

Voltaire Grid Director™ - More than Just a Switch

- Lowest Latency
- 4th Generation Silicon

- Fully IBTA Compliant
- Automatic Signal Optimization

- Secure Device Management
- Advanced Cable Management
- SM on board
- SNMP-based monitoring

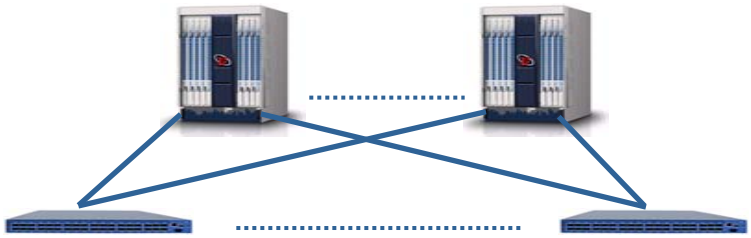
- Isolation & QoS
- Traffic Aware Routing
- Congestion Management
- Collective Offload

© 2010 Voltaire Inc.

Scalable Architectures

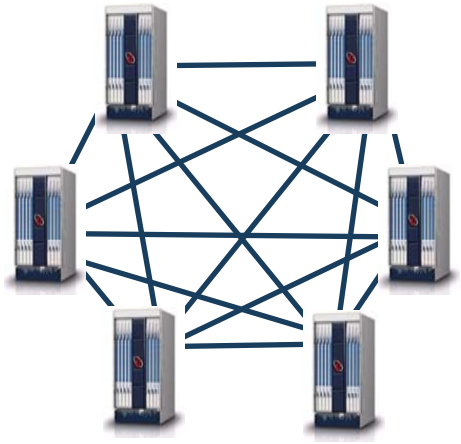
Fat Tree

- Full bi-sectional bandwidth at any node count
- Uniform oversubscription options



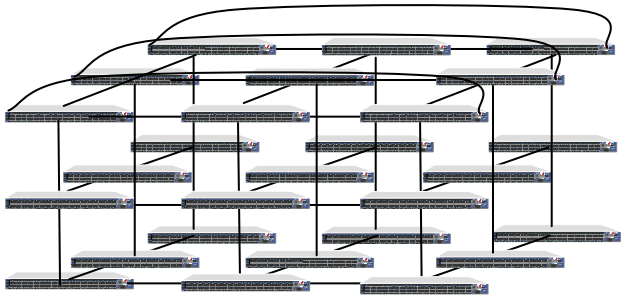
HyperScale

- Scale to thousands of nodes with linear performance
- Large non-blocking islands (more than 2,000 cores)
- 4-hops maximum latency to any port
- Lowest number of switches and cables



Torus

- Lowest cost solution
- Built entirely with edge switches and copper cables
- Voltaire actively involved in new projects leveraging Red-Sky (Torus2Qos going into UFM)



© 2010 Voltaire Inc.

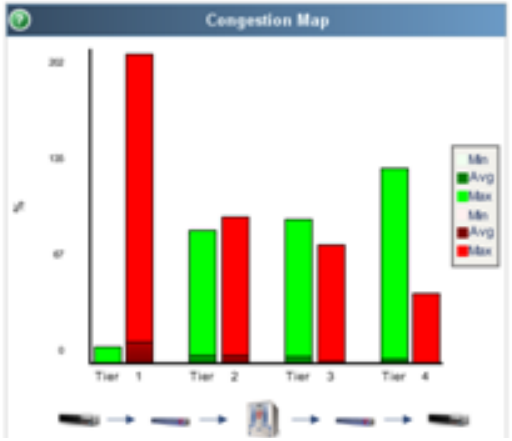
UFM – Unified Fabric Manager

Maximizing Cluster Utilization

- **Provides unprecedented fabric visibility and control**
- **Maximizes the performance of existing equipment**
- **Solves traffic bottlenecks**
- **UFM TARA (Traffic Aware Routing Algorithm) integrates with leading job schedulers to optimize routing based on actual communication patterns:**

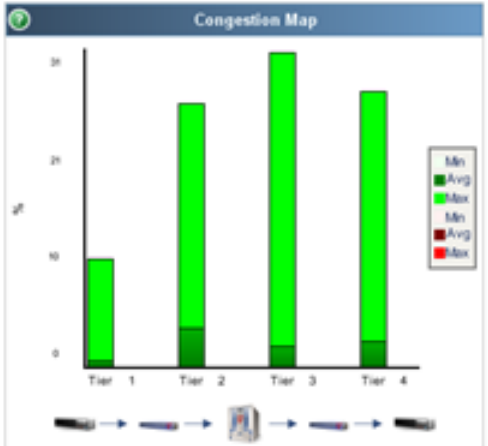
Adaptive COMPUTING | Platform Computing | PBS Works™

Before



Random Routing Causes Congestion

After



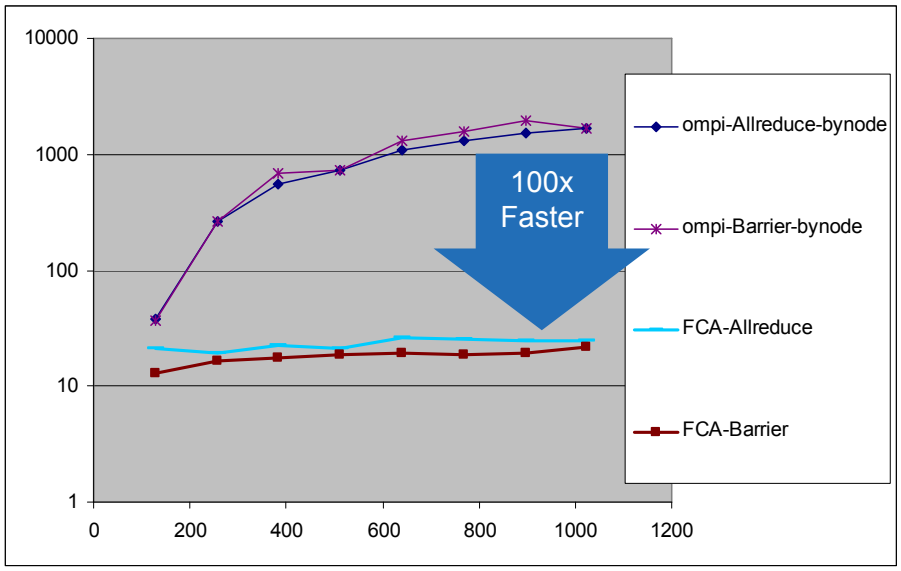
UFM Traffic Aware Routing™ Algorithm (TARA) Applied - No Congestion

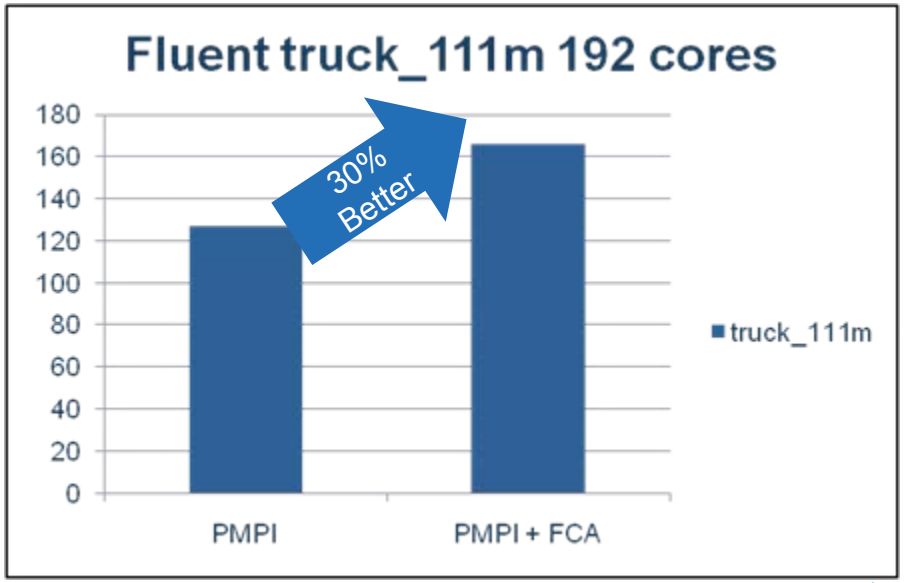
© 2010 Voltaire Inc.

FCA– Fabric Collective Accelerator

Unmatched Application Scalability

- First and only system-wide solution for offloading MPI collectives
- Accelerates MPI collective computation by as much as 100X
- 10-40% improvement in application runtime
- Integrated with leading MPI implementations





© 2010 Voltaire Inc.

Meeting Strict Quality Requirements

- **Voltaire Grid Director™ QDR switches selected by Fujitsu after months of extensive tests, including:**
  - Minimal BER over days of full load traffic (Zero Errors)
  - Extreme operating temperature conditions for more than 24 hours (> 45° C)
  - Beyond standard ESD (Electrostatic Discharge) test (9kV)
  - Full Shock & Vibration tests
  - Beyond standard Drop Test (2m)
  - Visual inspection of assembly and soldering
  - Force measurements on mechanical insertion of cables
  - Full software monitoring capabilities – real-time error reporting and statistics collection

The End Result:
More Reliable Products for Voltaire, The most Reliable Solution for Fujitsu

© 2

Summary

- **Most Reliable, Scalable InfiniBand-based HPC Clusters**

- **The foundation for maintaining Japan's supercomputer leadership**

JAEA

TOP500 SUPERCOMPUTER SITES

- **Continuing the partnership into FDR InfiniBand…**

FUJITSU

VOLTAIRE

Thank You!

© 2010 Voltaire Inc.