Using 10 Gigabit Ethernet Interconnect for Computational Fluid Dynamics in Automotive Design and Engineering

What You Will Learn

Will your high-performance computing (HPC) cluster meet the growing demands for low latency and high performance while reducing costs and increasing manageability, simplicity, and flexibility?

Today's HPC clusters have multiple nodes—compute nodes, processor nodes, and I/O nodes—connected through multiple interconnect technologies. This design adds to the complexity and manageability challenges of the HPC environment and to the costs of low-latency, high-throughput interconnect fabrics such as InfiniBand. 10 Gigabit Ethernet, with its low latency and with performance similar to that of InfiniBand, is emerging as a viable technology for HPC clusters.

This document presents the results of a series of tests conducted by Cisco to compare the cluster performance of HPC applications over 10 Gigabit Ethernet and InfiniBand using different workloads (data set types and sizes) and network topologies.

Challenges of HPC Cluster Design

The automotive industry typically uses computational fluid dynamics (CFD) and finite element analysis (FEA) to implement high-performance modeling and simulation in a number of areas such as aerodynamics, crash testing, and engine design. The challenges in designing an HPC cluster for these applications include:

- **Large data sets measured in gigabytes and terabytes, sometimes approaching petabytes:** CFD and FEA used in the automobile industry use large data sets that run with ever-growing granularity.
- **Real-time visualization needs:** CFD and FEA require real-time visualization for hundreds of Mbps of streamed data, and CPU efficiency that demands a maximum network latency of approximately 10 milliseconds (ms).
- **Need to increase the computational power of the HPC cluster without compromising efficiency:** To reduce product life cycles when such large data sets are used, the cluster needs ever-growing computational power, forcing engineering to look more closely at both compute and network latencies and the overall efficiency of the HPC environment.
- **Need to balance open standards, high system availability, and reduced cost with compute power and latency:** The compliance to open standards lets automobile industries to deploy the right HPC infrastructure for their applications and adopt new technologies with confidence.

Solution: 10 Gigabit Ethernet

With its low latency and performance on a par with InfiniBand, 10 Gigabit Ethernet is emerging as a viable technology for HPC clusters. The Cisco Nexus™ 5000 Series Switches are Cisco® access switches that offer unified fabric, high-performance 10 Gigabit Ethernet port densities, and very low latency. The Cisco Nexus 5000 Series supports Fibre Channel over Ethernet (FCoE) and meets the needs of more than 95 percent of the applications in the HPC space. While the costs per port and per connection have dropped significantly, the application performance, latencies, and throughput are on a par with InfiniBand. The Cisco Nexus 5000 Series also supports adapters from multiple vendors. Small Form-Factor Pluggable Plus (SFP+) provides options for copper or optical connectivity for nodes and switches. In addition, the Cisco Nexus 5000 Series provides the capability to create virtual network interface cards (NICs) on a single interface, allowing you to connect several servers to an interface.

Performance Tests

Cisco conducted a series of performance tests to measure the scalability of different types and sizes of data sets over 10 Gigabit Ethernet using different topologies and methodologies for job distribution, in comparison with InfiniBand. Cisco also compared the effects of network latency on application performance across 10 Gigabit Ethernet and InfiniBand.

The topologies used for the test were:

- Double-data-rate (DDR) InfiniBand
- 10 Gigabit Ethernet with a single switch hop
- 10 Gigabit Ethernet with multiple switch hops
- Gigabit Ethernet

Cisco compared the speed increase, scaling, and latency of 10 Gigabit Ethernet and InfiniBand for Fluent and Abaqus applications. The tests for Fluent applications were based on two benchmarks: data set with 5 million cells and 20 million time steps and data set with 111 million cells. Cisco performed the performance tests by starting with a serial job run and then incrementing the job runs by a power of 2, up to 256 core job runs.

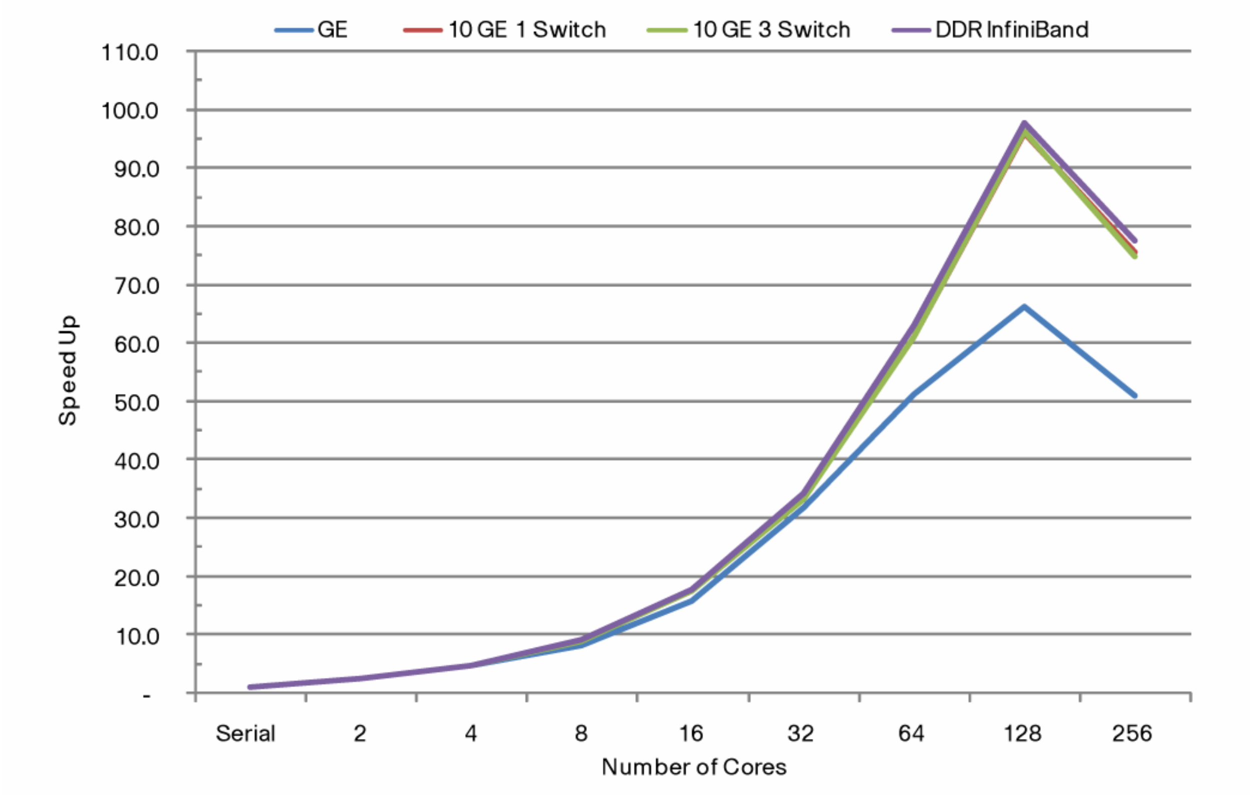
List of Terms

- **Computational fluid dynamics (CFD):** Fluid mechanics that uses a set of numercial methods for solving and analyzing the problems of liquid and gases flow.
- **Compute latency:** The time that a message takes to travel from one message core to another.
- **Efficiency:** The balance between the compute and communicate times of nodes.
- **Finite element analysis (FEA):** Solid mechanics that you can use to simulate processes in applied forces, deformations, internal stresses, thermal expansions, and other physical properties.
- **Latency:** The time taken to encode a packet for transmission, the time taken for the serial data to traverse the network equipment between the nodes, and the time taken to get the data off the circuit.
- **Rating:** The primary metric used to report performance results for the Fluent benchmarks; it is defined as the number of benchmarks that can be run on a given machine (in sequence) in a 24-hour period, and it is computed by dividing the number of seconds in a day (86400 seconds) by the number of seconds required to run the benchmark: a higher rating means faster performance.
- **Scale-up performance:** The number of components into which you can divide a job before a significant effect on efficiency occurs.
- **Speed-up performance:** A comparison showing how much faster a job runs as you add more cores or machines in contrast to the speed at which the application runs with a single processor core; the speed-up calculation is based on the wall-clock time and the rating values of a full run.
- **Tightly coupled applications:** Applications that involve high degrees of internodal and interprocessor communication.
- **Wall-clock time:** The compute time of a job.

Speed-Up Performance

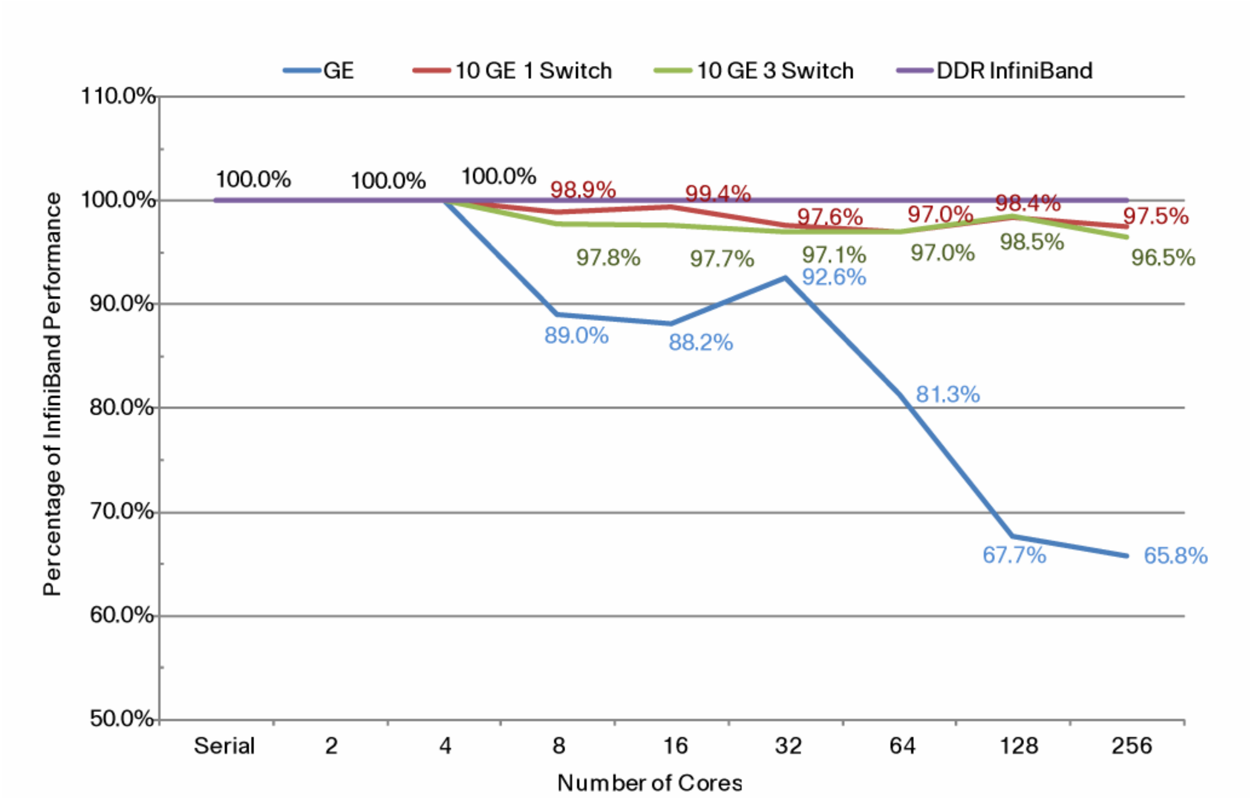
The performance tests indicate that the speed-up performance of 10 Gigabit Ethernet is within 3.5 percent of InfiniBand at 56 core job runs and within 3 percent at 256 core job runs (Figure 1).

Figure 1. Fluent Water Jacket 5-Million-Cell Data Speed Up



The variation in the speed-up performance of 10 Gigabit Ethernet for single and multiple switch hops was negligible. With very little TCP offload and no Remote Direct Memory Access (RDMA), the performance of 10 Gigabit Ethernet was very close to that of InfiniBand: less than or equal to 2.5 percent for a single switch hop and less than or equal to 3.5 percent for three switch hops. These results indicate that you do not need to be overly concerned about the topology when you schedule jobs. Traditionally, the approach has been to reduce the number of switch hops to reduce network use. However, the results indicate that adding switch hops to the message path has little effect on the application performance, primarily because the application uses large messages varying in size between 64 and 128 kilobytes (KB) to communicate (Figure 2).

Figure 2. Speed Up as a Percentage of InfiniBand Performance

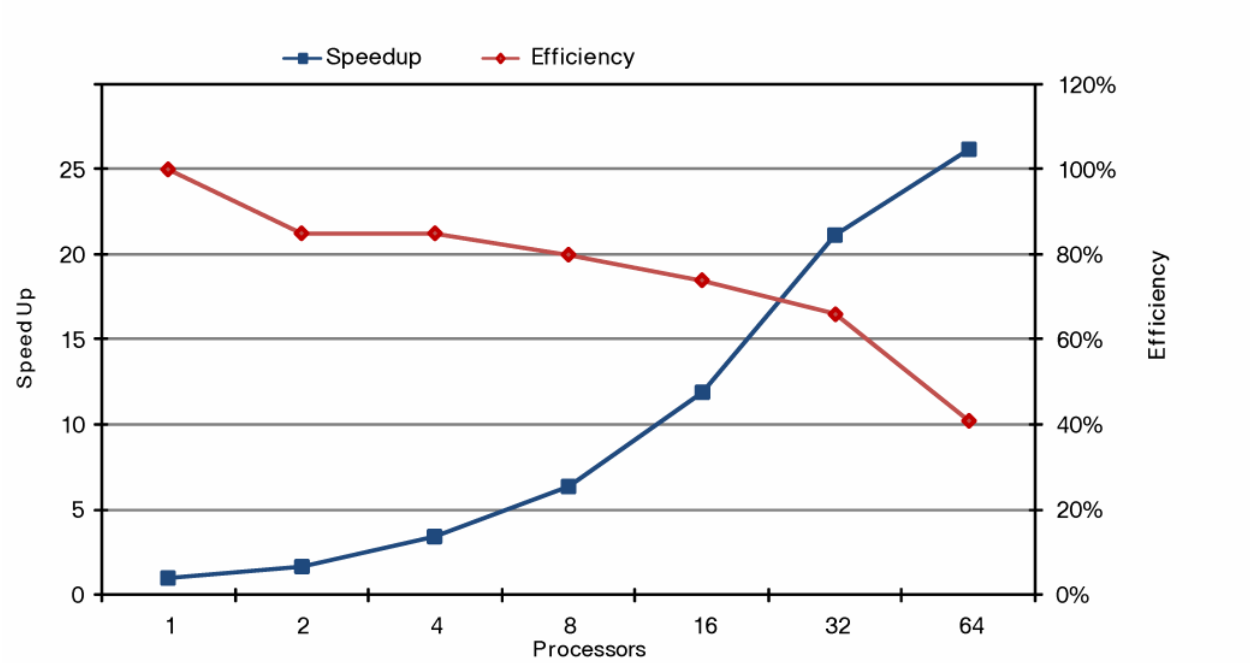


Scale-Up Performance

The challenge is to increase the computational power of the cluster without decreasing the efficiency. The scale-up performance of Fluent applications on a cluster is based on the number of times a given data set can be run in a 24-hour period, what Fluent calls a rating. A higher rating indicates higher computational power of the cluster.

Cisco compared the rating variation between 10 Gigabit Ethernet and InfiniBand by using a very large benchmark of a 111-million-cell Fluent data set. The performance variation was less than 7 percent (Figure 3).

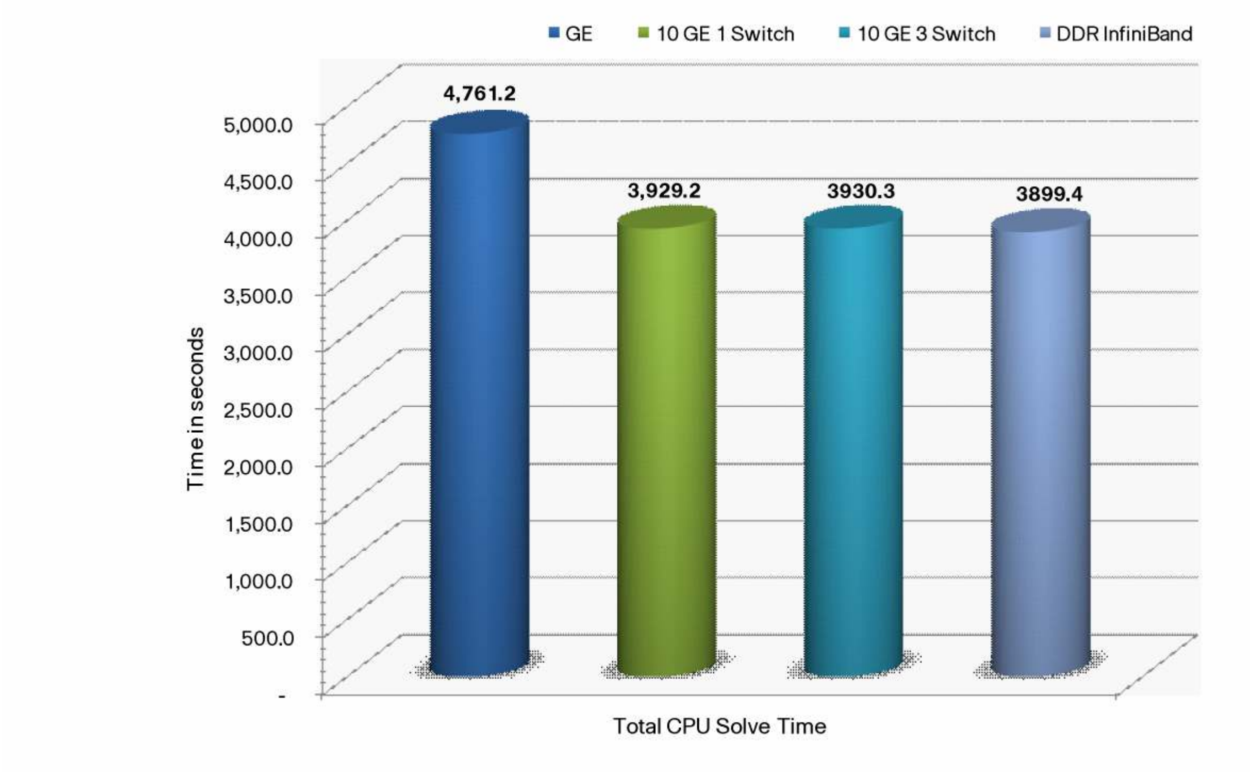
Figure 3. Speed Up Compared to Efficiency



Latency Performance

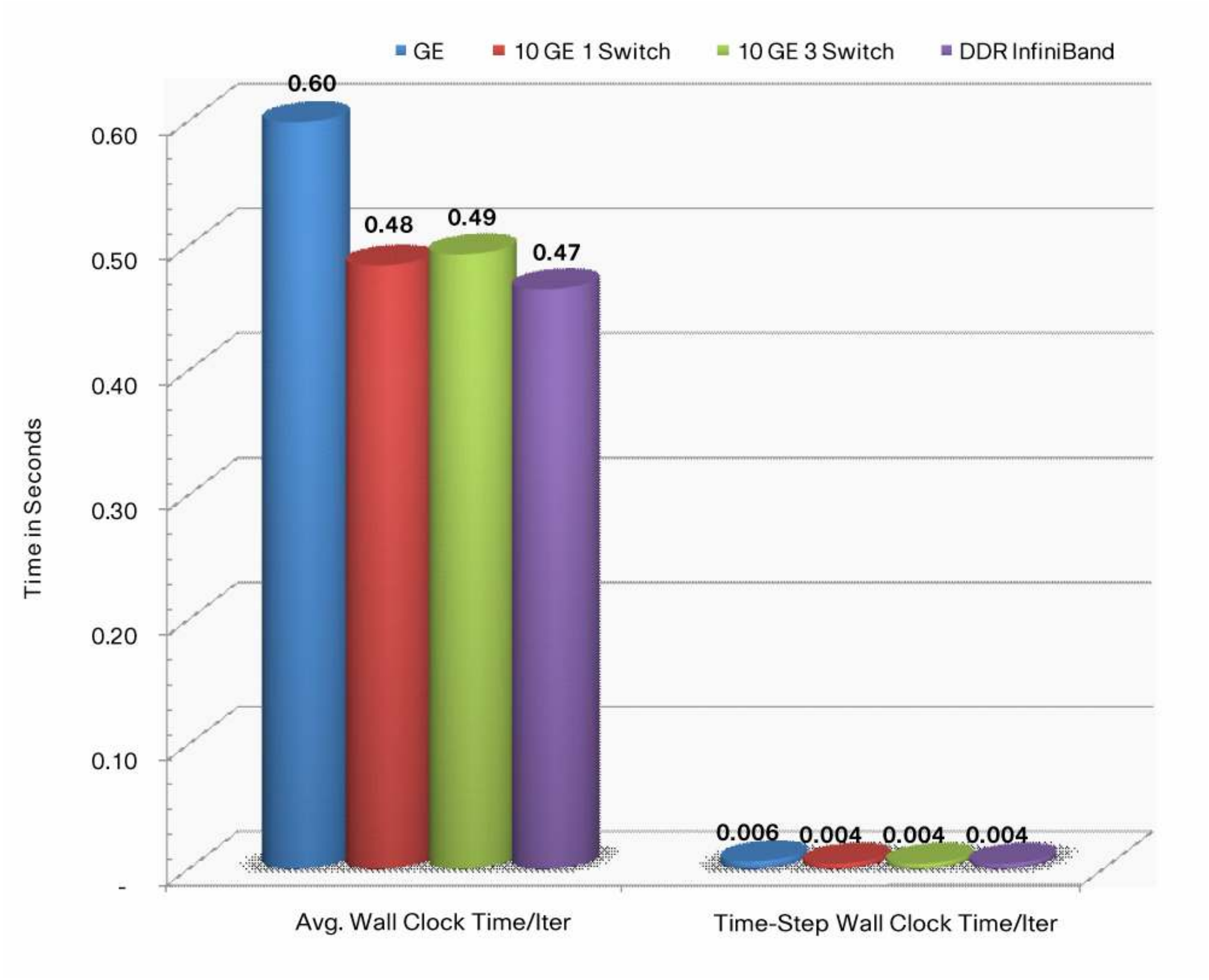
A rule of topology design suggests that, for tightly coupled applications, lower message latency reduces the run time for the application. To validate this theory, Cisco compared the network latency of 10 Gigabit Ethernet and InfiniBand and obtained a performance variance in the CPU solve time of less than 1 percent, clearly indicating that the latency performance of 10 Gigabit Ethernet is on par with InfiniBand. Even when the network latency component was tripled, there was little effect on the total application run time, with an effect on compute times of less than 1 percent. Network latencies are a much smaller time component than the compute latencies of the applications (Figure 4).

Figure 4. Total CPU Solve Time



The performance of 10 Gigabit Ethernet was equivalent to that of InfiniBand for the wall-clock time of each individual time step and for each iteration comprising more than 100 time steps (Figure 5).

Figure 5. Wall-Clock Times per Iteration

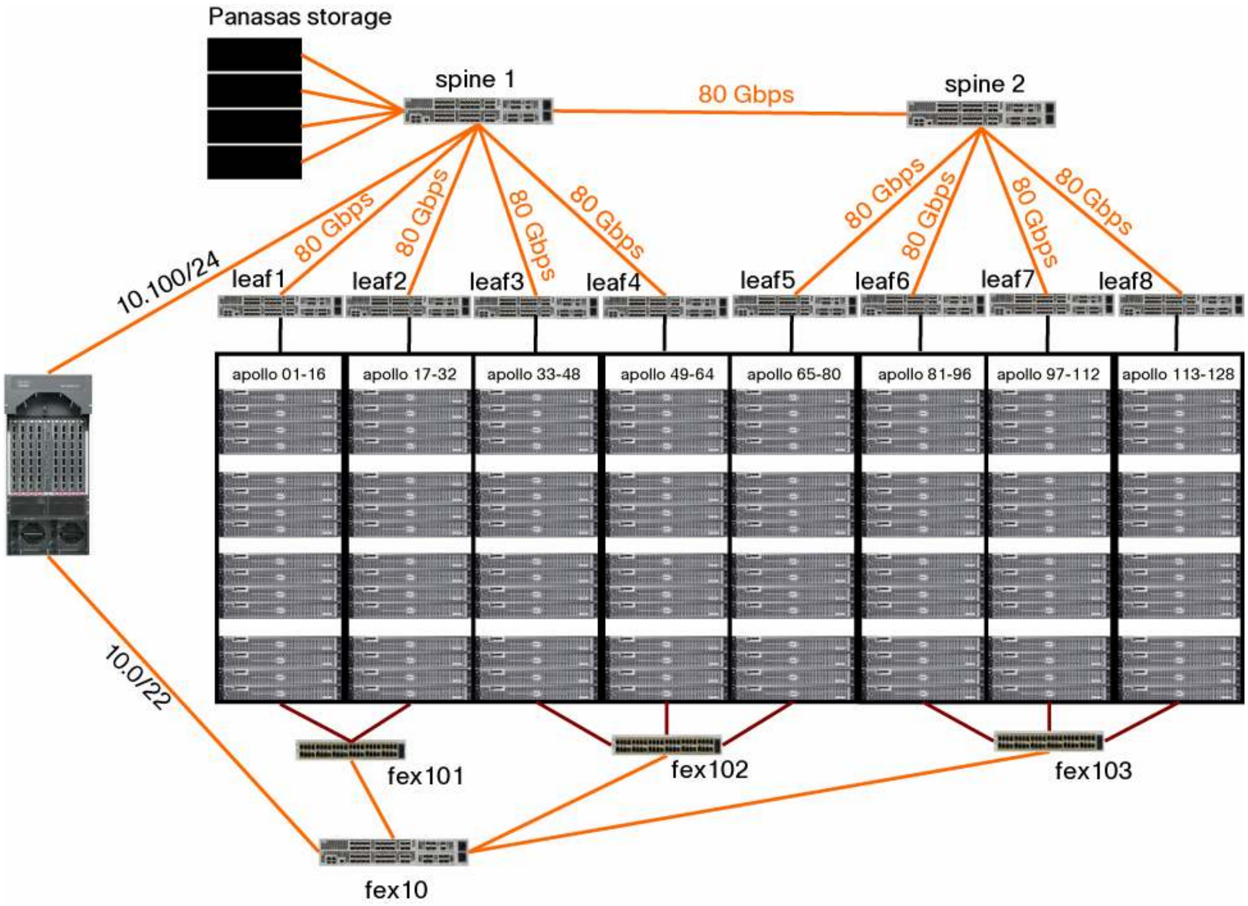


Test System Configuration

- Computer Platform
  - Dell PowerEdge 2950 III server
  - Dual socket with Intel Xeon E5430 quad-core CPU at 2.66 GHz
  - 16 GB of DDR2 667-MHz DIMMs
  - 1066-MHz front-side bus
  - 300-GB 7200 RPM SATA drive
  - Two 8-slot and one 4-slot PCIe 1.1
  - Two integrated Broadcom NetXtreme II 5708 Gigabit Ethernet LAN on motherboard ports
- Operating System
  - Red Hat Enterprise Linux 5.0
  - 2.6.18-8.el5 kernel
  - ServerEngines 10 Gigabit Ethernet dual-port, dual-controller adapter (http://www.serverengines.com/products/iocontroller.html)
- Network
  - Cisco Nexus 5020 52-port nonblocking 10 Gigabit Ethernet switch

  - Leaf and spine topology
  - Eight leaves and two spines
- Storage
  - 40-TB Panasas network-attached storage (NAS)
  - Four 10-TB 10 Gigabit Ethernet connected shelves
  - Network File System (NFS) and Panasas ActiveScale File System (PanFS)
- Cluster Topology
  - See Figure 6.

Figure 6. Cluster Topology



Conclusion

With the advent of low latency and performance rivaling that of InfiniBand, 10 Gigabit Ethernet is fast emerging as a viable technology for HPC clusters. The performance tests indicate that 10 Gigabit Ethernet provides a communication fabric that allows applications to perform at the highest levels. In the past two to three years, the cost per port has dropped from about US$12,000 to as low as US$500 today. The cost of 10 Gigabit Ethernet adapters is about 25 percent of what it was in the recent past. 10 Gigabit Ethernet technologies help reduce port costs, port densities, physical-layer (PHY) power draw, and heating costs. The Cisco Nexus 5000 Series consolidates multiple networks into a single network fabric that can concurrently handle LAN, SAN, and server clustering traffic, resulting in significant reductions of capital and operating expenses. It supports I/O consolidation at the rack level, which reduces the number of adapters, cables, switches, and transceivers that each server must support, all while protecting investment in existing storage assets.

For More Information

- Cisco Nexus 5000 Series Switches: http://www.cisco.com/en/US/products/ps9670/index.html

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-554431-00 08/09