Cisco UCS C200 M2 Server Powered by Intel Xeon Processors: 10 Gigabit Ethernet Matches QDR InfiniBand for Clustered HPC

Performance Brief
November 2011

Highlights

Performance for High-Performance Computing

- A 10 Gigabit Ethernet cluster consisting of 32 Cisco UCS™ C200 M2 High-Density Rack-Mount Servers powered by Intel® Xeon® processors delivers equivalent or better performance than competing servers configured with quad data rate (QDR) InfiniBand running LS-DYNA, Schlumberger ECLIPSE®, and ANSYS FLUENT benchmarks.

10 Gigabit Ethernet with Remote Direct Memory Access Delivers Competitive Latency to QDR InfiniBand

- 10 Gigabit Ethernet with remote direct memory access (RDMA) capabilities delivers latency low enough to provide application performance that meets or beats QDR InfiniBand.

Simple, Standard, and Cost-Effective Ethernet Technology

- 10 Gigabit Ethernet lets organizations save through familiar Ethernet semantics, tools, and the application of existing management skill sets, while using existing multivendor management solutions.

Cisco UCS C200 M2 High-Density Rack-Mount Servers powered by Intel Xeon processors and connected by Cisco Nexus switches matched or exceeded InfiniBand performance across a range of high-performance computing applications.

Equipped with remote direct memory access (RDMA)–capable adapters, Cisco® 10 Gigabit Ethernet fabric delivered outstanding scalability and performance parity running benchmarks with LS-DYNA, Schlumberger ECLIPSE®, and ANSYS FLUENT, allowing organizations to replace complex InfiniBand fabrics with standard 10 Gigabit Ethernet technology while maintaining peak job throughput.

Challenging the Conventional Wisdom for High-Performance Computing (HPC) Interconnects

With horizontal scaling and clusters now dominating high-performance computing (HPC), leading applications require a balanced approach. Fast processors must be built into balanced systems that can handle demanding computational challenges. Those systems must be interconnected with cluster fabrics that offer not only substantial throughput, but also the lowest levels of latency to accelerate message passing interface (MPI) applications. In an industry first, Cisco's latest HPC benchmarks demonstrate that 10 Gigabit Ethernet HPC performance can match and even exceed performance on clusters configured with quad data rate (QDR) InfiniBand.

Cisco 10 Gigabit Ethernet with RDMA Compared to QDR InfiniBand

For at least the past decade, InfiniBand has been regarded as the fabric of choice for HPC clusters. Unfortunately, the addition of InfiniBand to traditional Ethernet networks can bring additional complications to an HPC cluster. InfiniBand requires its own switches and host adapters in addition to the standard Ethernet network that is typically already in place for tasks such as management, OS provisioning, and the launching and monitoring of MPI jobs. Moreover, because InfiniBand is not a TCP/IP network, it requires a different set of management skills and associated training than the ubiquitous Ethernet ecosystem. For example, InfiniBand cannot be managed by traditional Ethernet IP-based management tools, and InfiniBand-connected servers require gateway devices to access standard Ethernet-based storage.

While popular QDR InfiniBand has an impressive-sounding bandwidth of 40 Gbps, it is usually chosen more for its low latency than for its raw bandwidth capabilities. In fact, InfiniBand's high bandwidth is seldom fully utilized because of limitations

© 2011 Cisco Systems, Inc. All rights reserved. This document is Cisco Public Information.

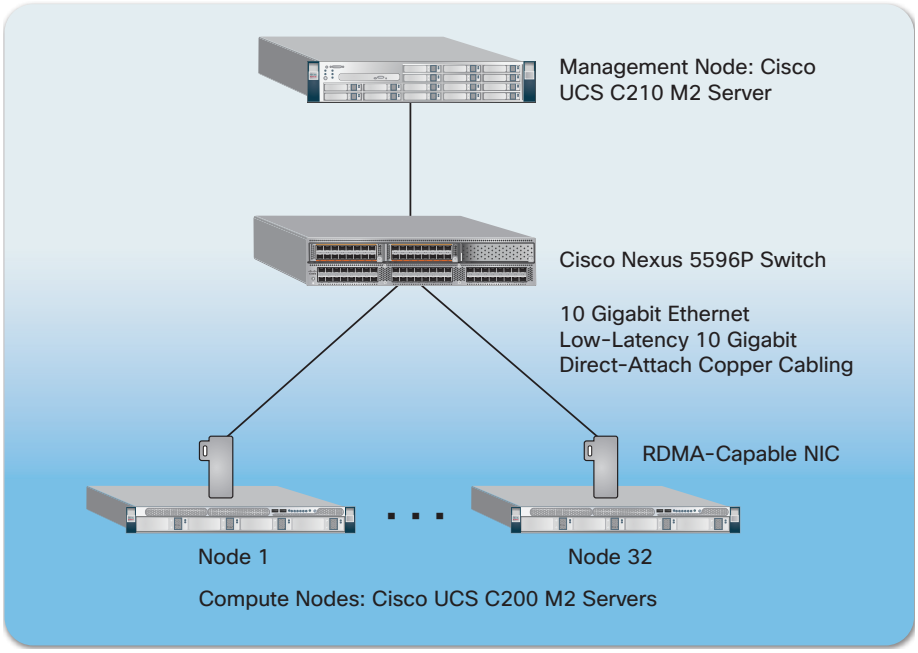

Figure 1. Cisco HPC Cluster Configured with Cisco UCS C200 servers, Cisco Nexus 5000 Series Switch, and 10 Gigabit Ethernet RDMA NIC Cards.

in host adapter PCI Express (PCIe) interfaces or application limitations, or both. However, low node-to-node latency is critical for clustered HPC application performance, especially for MPI applications that send significant amounts of communication over the network. Historically, Gigabit Ethernet has compared poorly to InfiniBand in terms of latency, being more than an order of magnitude slower

10 Gigabit Ethernet now presents dramatically better latency and offers a viable alternative technology to QDR InfiniBand. Like InfiniBand, 10 Gigabit Ethernet supports RDMA, allowing direct, zero-copy data transfer between application memory on servers with RDMA-capable adapters. RDMA bypasses the TCP/IP software stack in the Linux kernel, eliminating the need for data to be copied multiple times between buffers and also reducing context switching. With RDMA adapters, the resulting latency for 10 Gigabit Ethernet interconnects is close to that of InfiniBand, essentially removing latency as a factor in application performance in most cases.

10 Gigabit Ethernet Cluster Configuration

Cisco achieved similar or better results than competitors' QDR InfiniBand solutions using a cluster of Cisco UCS™ C200 M2 High-Density Rack-Mount Servers interconnected with a Cisco Nexus® 5596P Switch for MPI traffic (Figure 1). A Cisco Nexus 2000 Series Fabric Extender Gigabit Ethernet switch was used for management (not shown). Each server was connected to the Cisco Nexus 5596UP Switch using 10 Gigabit Ethernet network interface cards (NICs) with RDMA support.

A Cisco UCS C210 M2 Rack-Mount Server was used as a head node, running Platform Cluster Manager (PCM) and Red Hat Enterprise Linux 5.5, and using RAID10 for shared file systems, using Network File System (NFS). Because MPI is a standards-based API, no changes were required to the applications. Only a simple environment variable setting was required at runtime to select the interconnection type.

Cisco UCS C200 M2 High-Density Rack-Mount Server

The cluster was configured with 32 Cisco UCS C200 M2 servers, each with two top-of-the-line Intel® Xeon® processor X5670 CPUs and 96 GB of memory. Each server was configured with a 10 Gigabit Ethernet RDMA NIC, such as the Chelsio T420 adapter or the Mellanox MNPH29D-XTR 10 Gigabit Ethernet ConnectX2 NIC

Intel Xeon Processor 5600 Series

Ideal for HPC applications, six-core Intel Xeon 5600 series processors provide advanced features such as error-correcting code (ECC) memory and a large Level 3 cache. An integrated double data rate 3 (DDR3) memory

© 2011 Cisco Systems, Inc. All rights reserved. This document is Cisco Public Information.

controller provides up to 32 GBps of memory bandwidth. Depending on conditions, Intel Turbo Boost Technology allows processor cores to automatically run faster than the base frequency.

Cisco Nexus 5000 Series Switch

The servers were networked with a Cisco Nexus 5596UP Switch, which contributed to extremely low latency. For in-rack or adjacent-rack cabling, the Cisco Nexus 5000 Series Switches support Enhanced Small Form-Factor Pluggable (SFP+) direct-attach 10 Gigabit Ethernet copper cabling, an innovative solution that integrates transceivers with Twinax cables into an energy-efficient, low-cost, and low-latency solution.

Comparing 10 Gigabit Ethernet to InfiniBand for HPC Workloads

The three benchmarks described in the sections that follow all compare Cisco's 10 Gigabit Ethernet configurations with similar available QDR InfiniBand results.

LS-DYNA Car2car

As the leading crash code, LS-DYNA provides the opportunity for engineers to numerically model automobile crashes while reducing the cost and time related to real-world crash tests. As shown in Figure 2, the LS-DYNA Car2car test running on the Cisco 10 Gigabit Ethernet cluster matches and even surpasses benchmark results from Intel that used QDR InfiniBand as posted on http://www.topcrunch.org.

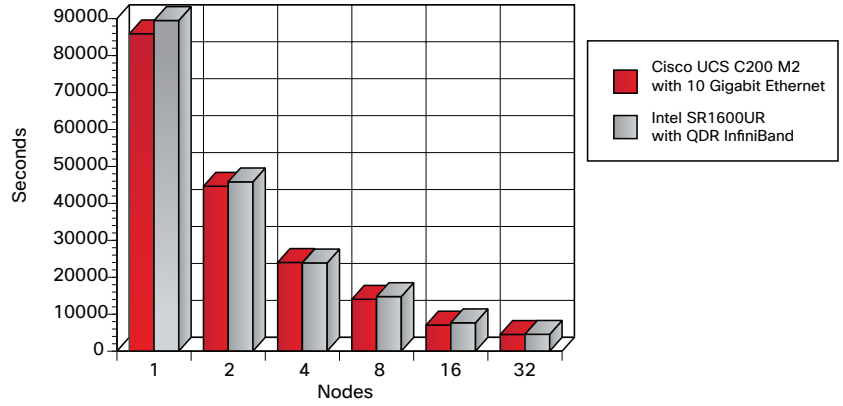

Figure 2. A Cisco 10 Gigabit Ethernet-Based Cluster Generally Exceeds the Performance of a QDR InfiniBand Cluster Running LS-DYNA Car2Car (Lower is Better).

ECLIPSE

Schlumberger ECLIPSE helps optimize oil and gas recovery, reducing the uncertainty of production and its impact on the environment. As shown in Figure 3, Cisco's 10 Gigabit Ethernet cluster again matched and often exceeded results provided by original equipment manufacturer (OEM) blade servers on QDR InfiniBand.

ANSYS FLUENT

As the leading computational fluid dynamics (CFD) application, ANSYS FLUENT software provides the physical modeling capabilities needed to model flow, turbulence, heat transfer, and reactions for a broad range of industrial applications. The truck-14m benchmark simulates external flow over a truck body using approximately 14 million

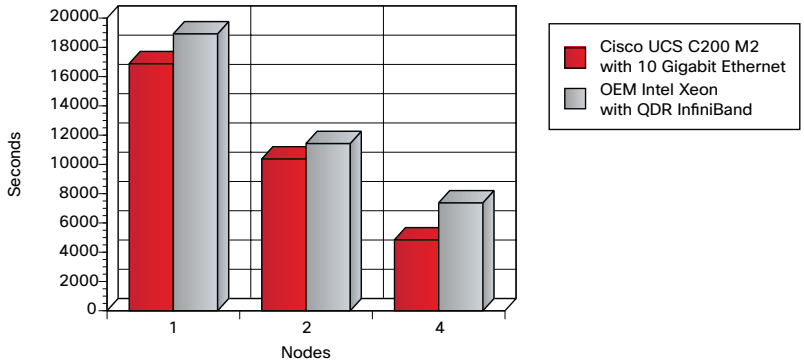

Figure 3. A Cluster Based on Cisco 10 Gigabit Ethernet Matches or Exceeds the Performance of a QDR InfiniBand Cluster for ECLIPSE (1 Million Cell Workload, Lower Values Are Better).

© 2011 Cisco Systems, Inc. All rights reserved. This document is Cisco Public Information.

cells. As shown in Figure 4, testing on the Cisco 10 Gigabit Ethernet cluster provided results that matched and even exceeded similar results submitted by SGI for a cluster equipped with QDR InfiniBand.

Conclusion

Despite the high bandwidth and performance claims of InfiniBand, Cisco has demonstrated that 10 Gigabit Ethernet is a viable solution for HPC clusters. The powerful Cisco C200 M2 server combined with Cisco Nexus 5000 Series Switches and RDMA NICs provides a low-latency solution that meets or beats QDR InfiniBand in real application performance for leading HPC applications.

At a time when simplicity and economy are essential, 10 Gigabit Ethernet presents an attractive proposition, providing essential low-latency performance with familiar Ethernet management, tools, and semantics. With 10 Gigabit Ethernet solutions from Cisco, organizations can continue to exploit industry-leading performance for their HPC applications without having to complicate their network infrastructure and management with different network technologies.

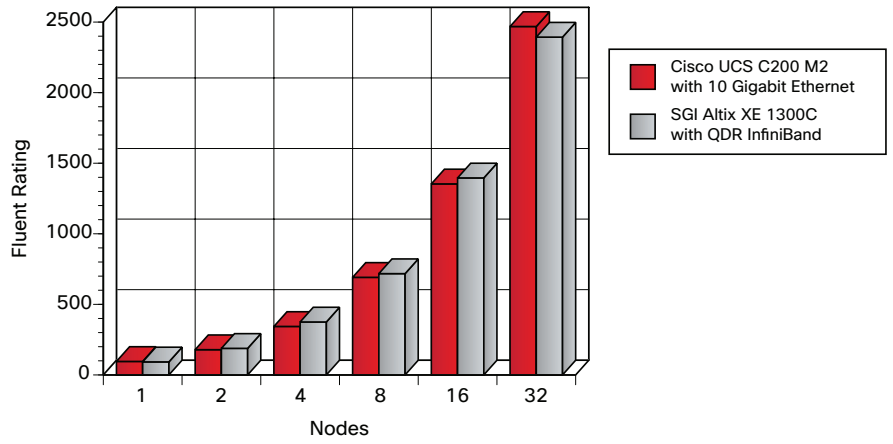

Figure 4. Cisco 10 Gigabit Ethernet Results Match or Exceed FLUENT Ratings for a QDR InfiniBand Cluster Running the ANSYS FLUENT truck-14m Benchmark (Higher is Better)

For More Information

- For more information about Cisco Unified Computing System™ (Cisco UCS) servers, please visit http://www.cisco.com/go/ucs
- For more information about Cisco UCS performance, please visit http://www.cisco.com/go/ucsatwork
- For more information about Cisco Nexus 5000 Series Switches, please visit http://www.cisco.com/en/US/products/ps9670/index.html

Disclosures

QDR InfiniBand comparative results used in this document were obtained from the following sources:

- LS-DYNA car2car results for the Intel SR1600UR system were provided by http://www.topcrunch.org and are current as of November 3, 2011.
- ECLIPSE results for OEM blade servers were provided by Schlumberger and are current as of November 3, 2011.
- Fluent truck-14m results for SGI Altix XE 1300C servers were provided by http://www.ansys.com and are current as of November 3, 2011.

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**.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)

LE-32804-00 11/11