CASE STUDY

TELCO Meets Datacenter Growth and Green IT Demands with InfiniBand-based Layered Grid Architecture

This European-based telecommunications (TELCO) leader provides wireless communication and value-added services to millions of consumers worldwide. Its datacenters provide online transaction processing and data warehousing services that directly impact the quality of service delivered to consumers and businesses.

Challenges

The company's existing IT environment hosts more than 30,000 servers deployed in multiple geographically distributed datacenters. Data storage facilities include a mix of network attached storage (NAS) and storage area networks (SAN). Connectivity is a mix of Gigabit Ethernet between server tiers and storage, and Fibre Channel within the SAN environments.

The company is experiencing rapid, sustained business growth that is straining its IT infrastructure. At the same time its IT organization, under pressure to reduce costs, has committed to deliver more transactions per dollar, reduce power consumption, and contract the space requirements of new datacenter build-outs. Unfortunately, the existing infrastructure won't scale affordably to accommodate the firm's growth-driven performance requirements, or support its increasingly ambitious cost-, energy- and space-reduction goals.

One problem is the large number of SMP (symmetric multi-processor) servers; fully 50 percent are 4-64 CPU systems that are inherently expensive, inefficient and inflexible. A second set of challenges spans the complexity, inflexibility and performance penalties that inevitably result from the use of two separate interconnect technologies within the datacenter.

The company seeks a new datacenter architecture that will meet its short-term challenges while establishing a reliable, risk-free growth path. This new architecture must be open standards-based and vendor-neutral. It must deliver a compact infrastructure footprint, maximize server utilization, reduce power usage and significantly improve the end user experience.

The Solution: A Layered Grid Approach

The company's architectural solution will divide the datacenter infrastructure into three grids—a compute grid, a data grid and a storage grid—each with its own performance targets. Commodity servers will be clustered to create an SMP-like operating environment in which performance will scale linearly as servers are added, allowing each grid to scale independently as business requirements dictate.

To enable incremental scalability while tightly controlling footprint, power and cost will require a high-bandwidth, low-latency, converged interconnect to link the grids and the server and storage nodes within them. The interconnect must support virtualized and converged I/O so that compute, data and storage traffic from each grid can flow through the same interconnect with appropriate quality of service provisioning. For the grids to scale linearly, the interconnect technology must be super efficient; node-to-node latency must be ultra low and network processing must be offloaded from the CPU to enable faster application processing. The interconnect must also be standards-based, be available from multiple suppliers, and use open source software and protocols supported by major operating system vendors.

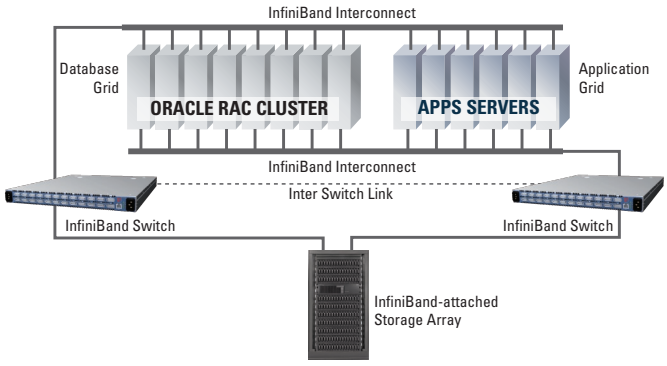
The InfiniBand interconnect architecture satisfies all of these requirements. Its layer two link protocol allows multiple, independent logical data flows to be multiplexed onto a single physical network with separate bandwidth allocation mechanisms to ensure quality of service. It provides hardware support for Virtual Interface architecture, decoupling CPU operation from I/O management to reduce overhead and improve processing performance. Its Reliable, In-order, Connection (RIC) transport service, implemented entirely in hardware, reduces transport-related CPU overhead to significantly improve performance while preserving full support for

© Copyright 2008. Mellanox Technologies. All rights reserved. • www.mellanox.com

legacy apps with TCP/IP communication stacks.

After careful considerations, the company has chosen a Mellanox InfiniBand-based interconnect solution and completed a proof-of-concept implementation.

The Test Configuration



The test configuration included an Oracle 11g RAC database with application and database servers from IBM. Each server blade supported two quad-core Intel Xeon processors, providing 8 cores per blade and 112 cores per chassis. Two Mellanox ConnectX IB adapters were used in a high-availability, DDR configuration (Double Data Rate—20Gb/s), and connected using a Voltaire InfiniBand switch system based on Mellanox InfiniScale III silicon. Backend storage was provided by InfiniBand-native IBM LSI DS5000 storage arrays with 256 drives and four InfiniBand DDR connections per array. Oracle ClusterWare (CRS) and Automatic Storage Management (ASM) solutions were used for advanced clustering and storage management functions.

Application servers and database servers were adequately loaded and transactions per second (TPS), I/O operations per second (IOPS) and I/O response times (IRT) were measured. The number of disks and server nodes were varied to determine scalability and bottlenecks.

Mellanox Makes the Difference

In the test setup, Mellanox InfiniBand delivered raw unidirectional bandwidth of 1397 megabytes per second, and bi-directional bandwidth of 2795 megabytes per second. Raw node-to-node latency measured 1.15 microseconds.

When application-level benchmarks were run and disk-related bottlenecks were eliminated using all 256 drives, the results showed that the TPM and IOPs increased almost linearly as the number of nodes in the database grid was doubled. IRT declined by 25 to 50 percent as node count increased.

These compelling results were achieved using Open Fabrics Enterprise Distribution (OFED) software. OFED is accepted industry-wide as the leading open source clustering and unified I/O software stack, and ships in leading Linux OS distributions.

Overall, the above results proved that the InfiniBand-based layered grid architecture met or exceeded the TELCO's goals, as shown in the table below:

Benefit	Metric	Results
Scalability, Server Utilization	Transactions per minute I/O operations per minute I/O response times	Near linear performance scalability, enables higher compute and storage performance with fewer grid nodes
Total Cost of Ownership	Amount of server, storage and I/O infrastructure hardware; cabling; power and cooling; ease of management	Close to 90 percent savings using InfiniBand-based unified I/O compared to separate Ethernet and Fibre Channel I/O solutions
Vendor neutral for deployment flexibility	Industry standard, open source software supported by the OS vendor	Use of OS vendor-certified OFED software with IBTA standards-compliant hardware enables unprecedented flexibility

Conclusions

Based on the test cluster results, the Telco has concluded that a layered grid datacenter architecture based on a converged InfiniBand interconnect will provide:

- Linear scalability for both its applications and Oracle database
- Significant power savings
- Simplified management
- Higher server utilization
- Improved application performance – The independent consultant who conducted the test work stated that, *"With the same SpecInt sizing as an SMP server system using Fibre Channel storage, the InfiniBand DDR Cluster provides 4X better application performance!"*
- Dramatically lower total cost of ownership – A financial consultant who evaluated the test results stated that, *"TCO savings will be close to **90 percent** if we replace our SMPs with IB Clusters."*

Mellanox TECHNOLOGIES
2900 Stender Way, Santa Clara, CA 95054
Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com

© Copyright 2008. Mellanox Technologies. All rights reserved.
Mellanox, ConnectX, InfiniBlast, InfiniBridge, InfiniHost, InfiniRISC, InfiniScale, and InfiniPCI are registered trademarks of Mellanox Technologies, Ltd. Virtual Protocol Interconnect is a trademark of Mellanox Technologies, Ltd. All other trademarks are property of their respective owners.