# InfiniBand -- Industry Standard Data Center Fabric is Ready for Prime Time

December 2005

Server and storage clusters benefit today from industry-standard InfiniBand's price, performance, stability, and widely available software leading to a convergence in the data center.

Overview

As the I/O technology with the largest installed base of 10Gb/s ports in the market (over 1 million ports projected by the end of 2005), InfiniBand has clearly delivered real world benefits as defined and envisioned by the InfiniBand Trade Association (www.infinibandta.org), an industry consortium formed in 1999. Like most nascent technologies, InfiniBand launched with unrealistic hype and expectations only to fall from grace as a result of the industry down-turn in the early 2000s. While many in the press and public eye deemed InfiniBand to never become a mainstream technology, the true industry believers of InfiniBand continued to mature both hardware and software solutions in preparation for a market recovery. The recent resurgence in IT infrastructure investment has been very timely for the adoption of mature InfiniBand fabrics and today's more realistic positive market perception is based on real world deployments. There are several factors that have enabled InfiniBand's adoption in data centers and technical compute clusters to quickly ramp and explain why it will continue to be the performance computing and storage fabric of choice.
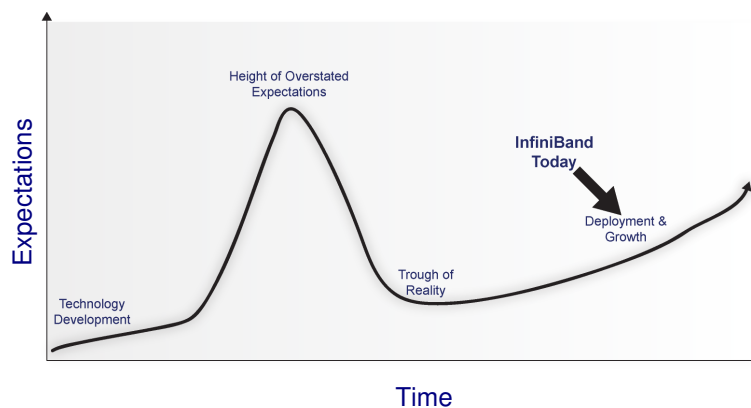
The Basics of the InfiniBand Fabric

InfiniBand fabrics are created with host channel adapters (HCA) and target channel adapters (TCA) that fit into commodity servers and storage nodes, interconnected by switches that tie all nodes together over a single high-performance network.
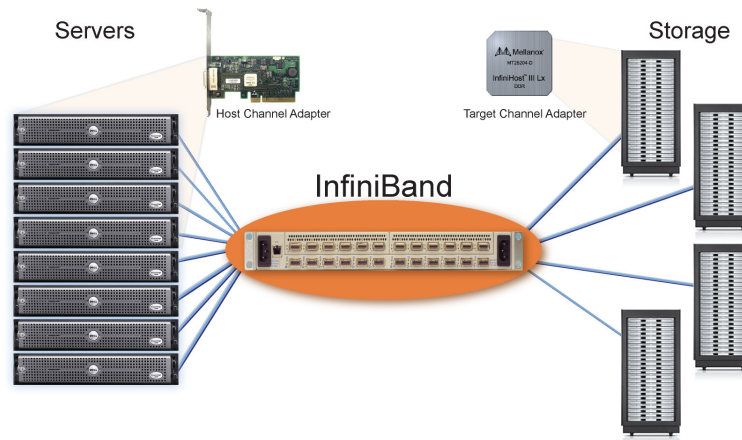
The InfiniBand Architecture (IBA) encompasses the entire fabric protocol designed to meet the following needs:

- High bandwidth, low-latency computing, storage and management over a single fabric
- Cost effective silicon and system implementations with an architecture that easily scales from generation to generation
- Highly reliable, available and scalable to thousands or even tens of thousands of nodes
- Exceptional utilization of compute processing resources
- Industry-standard ecosystem of cost-effective hardware and software solutions

InfiniBand is the only fabric that meets all of these criteria. Standard InfiniBand server-to-server and server-to-storage connections today deliver up to 20Gb/s of bandwidth. InfiniBand switch-to-switch connections deliver up to 60Gb/s. This bandwidth performance is matched with world-class application latency performance (well under 3µs) and switch latencies that are less than 200ns that enable efficient scale out of compute and storage systems. InfiniBand also defines an industry-standard implementation of hardware transport and remote direct memory access – protocols in the fabric hardware that enable the most efficient usage of computing resources on application processing rather than network communication. With these defined goals understood up front, IBA is clearly driving the most aggressive performance roadmap of any I/O fabric, while remaining affordable and robust for mass industry adoption.

Optimizing Both High-Performance Computing and Data Center Applications

While InfiniBand has gained significant traction in clustered computing for government, academic, scientific, and research oriented environments (traditionally considered High-Performance Computing or HPC), it is also expanding into mainstream retail, financial and commercial data center markets for use in compute networking. The reason InfiniBand is quickly being adopted in both of these markets is the value proposition clusters of commodity servers have over monolithic, expensive symmetric multi-processing (SMP) servers. Not only is clustered computing a fraction of the price of single SMP servers from an infrastructure investment standpoint, but the ability to incrementally add compute power to a cluster (referred

to as scaling out) is advantageous versus forklift upgrades with SMP servers (referred to as scaling up) as requirements increase over time.
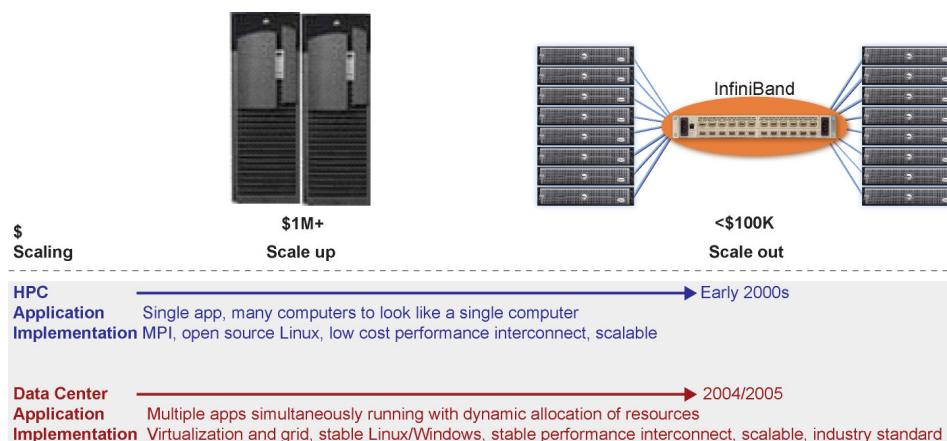

Typically, the goal of HPC applications is to make many separate compute nodes function like a single computer. The performance and characteristics of InfiniBand fabrics are ideal for this application, and have been in production usage for several years. On the other hand, compute networks in data centers have a different set of objectives, which have taken more time to mature in the market. For example, typical data centers simultaneously run multiple applications, and require the flexibility to dynamically allocate compute resources to those applications depending on immediate requirements. The network must support compute, storage, communications, and management traffic seamlessly. In addition, software application stability is a requirement for commercial environments. The recent emergence of virtual and grid computing solutions, in addition to mature software solutions that have evolved over the last several years, have set the stage for InfiniBand's mass deployment in business and utility computing environments. Industry-standard InfiniBand has the price, performance, stability, and widely available software that makes it ready for prime time.

Industry Standard

The importance of an industry-standard open specification for IBA cannot be understated. By gaining acceptance from industry leading solution providers from its initial inception, InfiniBand has garnered wide support for both hardware and software based solutions. Nearly every server vendor in the industry is shipping InfiniBand adapter solutions in the form of add-in PCI-X and PCI Express cards, as well as Landed-on-Motherboard (LOM) and Blade Server architectures.

3

Server Vendors Shipping InfiniBand Solutions

*Partial List

Hitachi   Fujitsu   NEC   SGI

Dell   HP   IBM   Sun
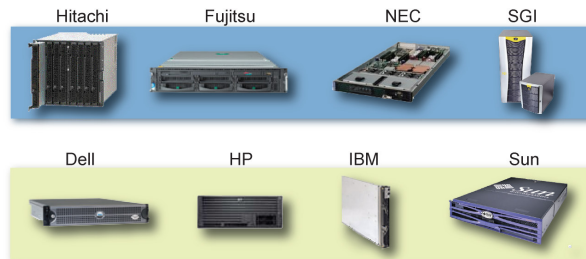
Not only servers and storage solutions demand a high performance interconnect. InfiniBand is widely used in embedded and communication applications and is becoming a ubiquitous performance fabric. Several InfiniBand solutions are available in industry-standard form factor chassis that houses these applications, such as VME and Advanced TCA. Other specialized chassis take advantage of the performance of InfiniBand fabrics for networking, industrial, medical and military applications.

InfiniBand Embedded Solutions

VMEbus CompactPCI   Advanced TCA®

Video Server   Military, Industrial Data Acquisition   Communications, Military, Medical   Networking, Telecom

## The InfiniBand Price Advantage

On the adapter side, InfiniBand is now in production with low priced, single-chip solutions (including integrated SerDes), enabled by the extremely fast adoption of PCI Express based servers. PCI Express enables fast access to system memory with low latency. As a result, a migration is taking place from the earlier InfiniBand adapters, which include local memory chips on-board for storing network context information (when utilizing PCI-X), to the memory-free (MemFree) adapters which now utilize system memory for storing that information utilizing PCI Express. Not only does this significantly reduce the price of InfiniBand adapter solutions, but it also reduces power to a mere 3.5W for a 20Gb/s InfiniBand port.

Reducing the solution down to a single HCA chip also reduces the real estate required making InfiniBand optimal for LOM and blade server designs. At the time of this document's publication, over six motherboard manufacturers provide InfiniBand LOM solutions, and over ten server manufacturers are shipping InfiniBand fabrics for their blade servers.
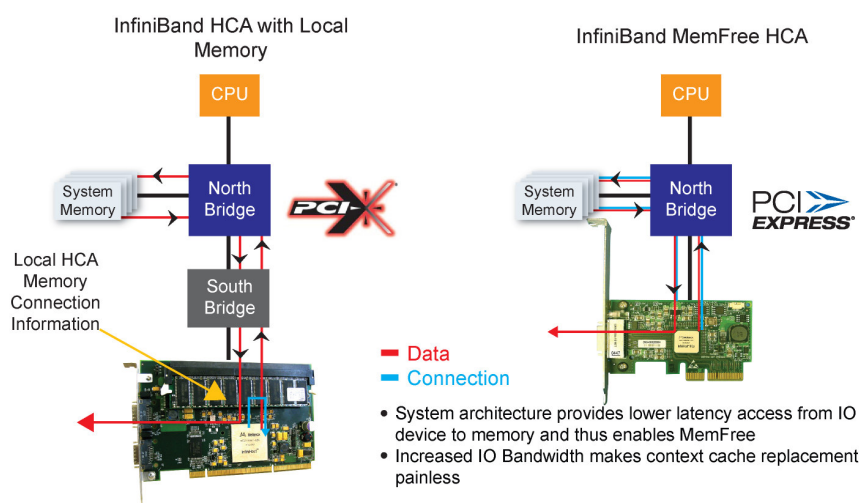
With a true cut-through forwarding architecture and a well defined end-to-end congestion management protocol, IBA also defines the most cost-effective and scalable switch solutions in the market. A single switch silicon device supports twenty-four 10Gb/s or 20Gb/s InfiniBand ports, which equates to nearly a terabit per second of aggregate switching bandwidth. Switches and adapters support up to 16 virtual lanes per link to enable granular segregation and prioritization of traffic classes for delivering Quality of Service (QoS) and guaranteed bandwidth. With integrated SerDes on all ports, InfiniBand is generations ahead of other 10Gb/s switching solutions, and has enabled the industry's most dense switching systems (up to 288 ports in a single chassis), significantly reducing the cost per port for large node-count fabrics.
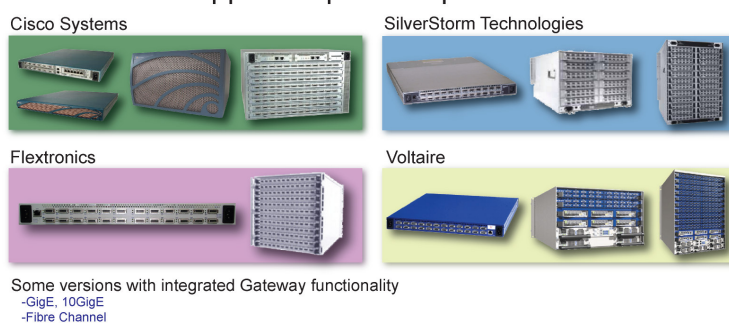
InfiniBand Performance

One of the key reasons that data centers are deploying industry-standard InfiniBand is the total cluster performance the fabric enables. First, InfiniBand is the only shipping solution that supports 20Gb/s node bandwidth (with a measured 3GB/s of bi-directional server-to-server throughput) and switch aggregation through 60Gb/s links. Second, InfiniBand has world-class application latency with measured delays of less than 3µs end-to-end. Third, InfiniBand enables the most efficient use of all processors in the network by performing all of the data transport

5

mechanisms in the adapter card. These three metrics combine to make InfiniBand a powerful interconnect.

The performance benefits are echoed in the trends of the Top500.org list that tracks the world's most powerful supercomputers. Published twice a year, this list is increasingly used as an indication of what technologies are emerging in the clustered and supercomputing arena. The November 2005 published list shows that the number of InfiniBand clusters reporting on the Top 500 grew almost 200% from the November 2004 report. Even more impressive, InfiniBand is the only interconnect technology that is increasing in rate of usage for new clusters introduced in the report—a sign of the overwhelming momentum InfiniBand fabrics have for large cluster deployments.

InfiniBand's Stable and Widely Available Software Solutions

From a software perspective, InfiniBand has garnered support from virtually every mainstream operating system including Linux, Windows, Solaris, Unix, HPUX, AIX, MAC OSX, and VxWorks. Open source and community-wide development of interoperable and standards-based Linux and Windows stacks are managed through an industry alliance called OpenIB.org. This alliance, consisting of InfiniBand solution providers, InfiniBand end-users and any programmer interested in furthering development of the Linux or Windows stacks, has successfully driven InfiniBand support into the Linux kernel distribution, and is driving towards support for Windows Server based solutions. The successful inclusion of InfiniBand drivers and upper layer protocols in the Linux kernel insures interoperability between different vendor solutions, and will ease the deployment of InfiniBand fabrics in heterogeneous environments.

OpenIB.org is also responsible for developing open source virtualization software support that utilizes the InfiniBand fabric based on the Xen virtualization solutions from University of Cambridge and XenSource. Other key virtualization software providers, including VMWare and Virtual Iron, have announced their support for InfiniBand for usage in enterprise data center environments. I/O channel-based communication available in the InfiniBand architecture is perfectly suited for I/O virtualization enabling unprecedented levels of resource utilization and flexibility in the allocation of compute, storage and network resources based on dynamic data center demands.

From an application point of view, InfiniBand has support for a plethora of clustered parallel processing applications in both enterprise and technical computing environments. In the enterprise environment, InfiniBand is being used for grid computing and clustered database applications driven by market leaders Oracle and IBM DB2 for retail, enterprise resource planning and customer relationship management. In addition, financial data centers are quickly transitioning from large, expensive SMP servers typically priced at over $1M, to clusters of low-cost servers (less than $5K per server) interconnected with InfiniBand. This clustered compute power combined with high performance InfiniBand interconnect is required for heavy processing related tasks such as customer risk analysis and real time forecasting.

In the technical computing field, InfiniBand provides the fabric connecting servers and storage to address a wide range of applications including oil and gas exploration, fluid dynamics, drug research, weather forecasting and molecular modeling—just to name a few. Over 20 independent software vendors have ported their custom applications to run on top of InfiniBand, and more are continually being added to the list.

InfiniBand Enables Convergence of Computing and Storage

As rapid adoption of InfiniBand continues in both technical computing and data center environments, and computing resources are able to handle higher data processing throughputs, storage access has now become a bottleneck. Legacy interconnects (Fibre Channel and Ethernet) are insufficient. As a result, the industry is demanding storage solutions that connect directly into the 10 or 20Gb/s InfiniBand fabric. The convergence of computing and storage access onto a single InfiniBand fabric has significant infrastructure investment savings and total cost of ownership (TCO) benefits, in addition to improved overall storage access performance.

Storage vendors are quickly introducing solutions based on IBA to address the strong market demand led by financial and clustered database applications as well as parallel technical computing applications. While the ultimate performance and price benefits of convergence are based on native, front-end InfiniBand storage solutions, some vendors use InfiniBand for back-end clustered storage solutions with a strategy to migrate InfiniBand interconnected solutions to the front end. Both block level and file level storage solutions are available in the market today with mass deployment planned in the 2006 timeframe.

InfiniBand Storage Products

Back-end and Failover

Clustered and Distributed File Systems

Native InfiniBand Storage
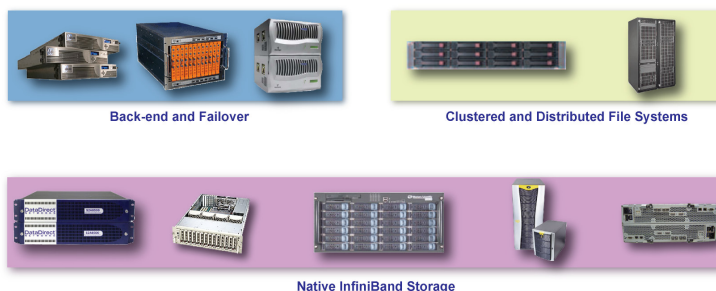
InfiniBand's Roadmap

One of the early goals of IBA was to define a fabric protocol that is cost effective to implement in silicon and system solutions while being easily scalable from generation to generation. Vendors are successfully meeting the milestones on what is the most aggressive I/O technology roadmap in the industry and are scheduled to introduce 40Gb/s node links in 2006 followed by 120Gb/s switch links in 2007. This exceptional roadmap will continue to depend on industry-standard, commodity based components to enable world-class supercomputing performance at mainstream interconnect prices. At the same time, these solutions will remain backward compatible to the eco-system of software being developed and deployed today.

The next generation InfiniBand adapters are perfectly aligned with the roadmap for PCI Express. Today, most servers ship with PCI Express x8 slots capable of transmitting 20Gb/s and receiving 20Gb/s of data simultaneously (20+20Gb/s). This performance level is well matched with today's dual active 10Gb/s InfiniBand ports or a single 20Gb/s InfiniBand port adapter. As PCI Express transitions to the next generation in the 2007 timeframe, a PCI Express x8 slot will double in performance to 40+40Gb/s. InfiniBand's roadmap has 40Gb/s adapters available in the same timeframe.

InfiniBand Roadmap

InfiniBand's roadmap outpaces all proprietary and standard based I/O technologies in both pure performance and price/performance
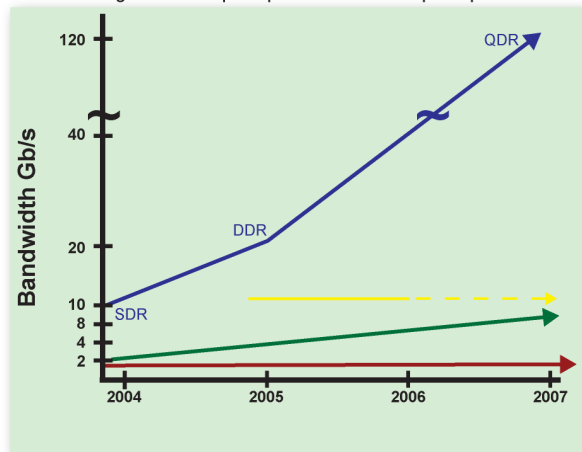
Number of IB Lanes	Per Lane Bandwidth		
	SDR 2.5Gb/s	DDR 5Gb/s	QDR 10Gb/s
4X	10Gb/s	20Gb/s	40Gb/s
8X	20Gb/s	40Gb/s	80Gb/s
12X	30Gb/s	60Gb/s	120Gb/s

- InfiniBand
- 10GigE, 10G iSCSI, Proprietary
- Fibre Channel*
- GigE, iSCSI

*FCIA estimates 2007/8 for 8Gb/s FC

InfiniBand – Proven for Real World Deployments Today

Surviving the industry down-turn in the early 2000s, InfiniBand technology has matured and is emerging as the preferred fabric for enterprise data center and performance compute cluster deployments. Shipping in production today with 10 and 20Gb/s adapters and switch solutions capable of up to 60Gb/s switch-to-switch links, the InfiniBand fabric is delivering on the value proposition defined by the industry back in 1999. With virtually every server vendor shipping InfiniBand solutions, availability of native InfiniBand storage systems, far reaching operating system support, and a wide variety of enterprise and technical computing applications, all of the pieces are in place for continued mass market deployment of InfiniBand fabrics.