I D C  A N A L Y S T  C O N N E C T I O N

*Jie Wu*
*Research Manager, High-Performance Computing*

# InfiniBand: Poised for Market Growth

*May 2009*

*Recent trends have combined to pose greater challenges for interconnect instrastructures in high-performance computing (HPC) and enterprise datacenters. In HPC, these challenges include the rapid increase in HPC system sizes, and the proliferation of multicore processors and accelerators, all of which increase the need for high bandwidth and low latency. In enterprise datacenters, virtualization and nascent cloud computing make increased demands on interconnect fabrics. An additional trend is toward the convergence of enterprise and HPC data centers and workflows. InfiniBand and Ethernet are the chief standards-based interconnect technologies competing in these arenas.*

The following questions were recently posed by representatives of the InfiniBand Trade Association (IBTA) to Jie Wu, Research Manager in IDC's High-Performance Computing practice, on behalf of enterprises considering Infiniband technology.

**Q.** **What features and benefits of InfiniBand (IB) position it to provide strong I/O management for today's high-performance computing (HPC) and enterprise datacenters — particularly in light of the latest major technical developments in virtualization, multi-core, and energy efficiency? How do these features and benefits play into the emerging push for a "Unified Wire" infrastructure within future datacenters?**

A. IB has already established a strong footprint in HPC, especially at large sites with challenging workloads that require both high bandwidth *and* low latency. In particular, IB's low latency characteristics stand out versus alternatives today. HPC clustered servers and networks have been growing in physical size and complexity, making low latency an even more valuable attribute than high bandwidth in many cases.

The rise of multicore processors and accelerators (GPGUs, FPGAs, and others) has added to the bandwidth and latency challenge for system- and network-level I/O. IB has demonstrated an ability to handle I/O efficiently across large numbers of processor cores. For example, in a 2009 IDC worldwide study of 110 HPC datacenters of all sizes, the core count for a typical run of a top IB application was about 7 times larger on average (500 vs. 72 cores) than for a top non-IB application. The multi-core efficiency of the applications running on IB systems was also substantially greater than for the non-IB apps.

InfiniBand and other standards-based interconnects have largely replaced proprietary interconnects in the HPC market. The general consensus among the storage and I/O panellists at the IDC-operated HPC User Forum conference in April 2009 was that IB will

remain a force in the HPC market for the foreseeable future because of cost, scalability, and simplicity.

One of the most important trends in the future datacenter is "Unified Wire," a.k.a. converged fabric, which refers to a new approach of consolidating all datacenter communications onto a single fabric, in contrast to today's more complex and expensive implementation of using different fabrics for inter-process and I/O. IDC believes that IB already possess the attributes needed to successfully implement such infrastructures — i.e., compliance with standards, highly secured lossless fabric, high bandwidth, and low latency. Hence, IB should be a strong contender for Unified Wire infrastructures, particularly for a datacenter where high-performance workloads dominate.

Among the benefits envisioned for these future infrastructures are simplified administration (e.g., single switch), higher application and storage performance, and lower total cost of ownership. As a strategy that relies on making more efficient use of datacenter IT resources, virtualization will also benefit from the attributes of Unified Wire infrastructures for advanced QoS functions such as dynamically reallocating resources and monitoring performance.

The latest IB offerings are equipped with a software intelligence layer to automate basic capabilities such as QoS, virtual HCAs, and virtual NICs to provide a comprehensive I/O management system for virtualization. Specifically, IB's channelized architecture design can dynamically allocate high bandwidth to each virtual server, while consolidating LAN and WAN traffic from the virtual servers over the same IB pipe. This greatly reduces network complexity while still delivering low-latency and high-bandwidth capabilities.

Managing power consumption and cooling costs is a major challenge for most datacenters today. IB is designed to enable I/O consolidation through RDMA protocol implementation. The result is fewer adapters and simplified cabling, all leading to less energy consumption and greater ease of management for users.

**Q. With the emerging compute clouds, there is increasing importance being placed on high-performance networking-based solutions that are capable of providing the required real-time response and fast dynamic modifications. How do you see the role of HPC solutions, in particular InfiniBand, in clouds?**

A. There is not strong agreement yet in the IT sector on the definition of cloud computing. However, some things are clear. First, cloud computing is still in its early stages and has the potential to significantly transform the way many end-users access enterprise and HPC IT resources. Second, private (internal) clouds have stronger momentum at this point than public clouds.

Regardless of how you define cloud computing, bandwidth and latency are among its major limitations today. These limitations, along with others such as security and SLA management, play an important role in determining which applications (or potions of applications) and other functions datacenter administrators assign to clouds. IB's high-bandwidth and low-latency characteristics have the potential to accelerate whatever portion of the data pathway they reside on, which is a major benefit in conjunction with cloud computing, including providing scalability benefits to meet the mixed-workload requirement in the cloud.

Compared to other alternatives, IB offers price/performance advantage. Today, IB offers node-to-node latency of 1us, up to 40Gb/sec bandwidth, and scalability of over 10,000 nodes, plus reliability features such as redundant paths between nodes — and it costs less than 10GE, its closest competition. This price/performance advantage makes IB a good candidate for network infrastructure consideration in cloud computing environments.

In summary, IB could play an important role in helping to broaden cloud adoption, starting at sites that already employ IB.

**Q. What do you think IB's strongest advantages are, and how can those advantages be leveraged to expand the adoption and role of IB, specifically in competition with 10-gigabit Ethernet (10GE)?**

A. Lossless, secured, high bandwidth, and low latency are the most distinct value propositions of InfiniBand today, and those features have advanced performance on many computationally intensive applications in the technical computing space. IDC sees a trend of increased convergence between technical (HPC) and enterprise datacenters and workflows.

The latest HP-Oracle database server is one of those examples where InfiniBand was chosen to accelerate the system speed for running large, multi-terabyte data warehouse applications. The result is a 10x performance boost compared to conventional database solutions. In the financial space, more and more financial services firms are moving to unify the back- and middle-office server-based HPC work performed by their "quants," with the front-office work done by their traders on desktop computers. Analogous workflow unification is occurring in other HPC markets, such as digital content creation, oil and gas exploration, and the bio-sciences. And, at a higher level, more organizations are siting their enterprise and HPC resources within a single datacenter, although unified datacenters like this are still a small minority.

Together, these trends present important opportunities for IB to expand its presence. To accomplish this, IB intends to maintain its existing advantages over alternatives, especially Ethernet. IB has a substantially larger footprint in HPC than 10GE today, but 1GE has an HPC footprint on a par with IB, and 10GE is working its way through the HPC market. Advances in multicore technology, the proliferation of virtualization in commercial datacenters, and the emergence of cloud computing all create new requirements for a more efficient network infrastructure. Today, in addition to delivering high bandwidth and low latency, IB offers a number of new features including improved scalability, I/O consolidations/simplicity, and improved price/performance.

As the closest competing technology, 10GE is still considered more costly compared to IB, and the feature set is not as rich and mature as that for IB. There is also a debate on how much it would cost to upgrade from 1GE to 10GE, all of which make 10 GE a less-appealing choice for certain industries and applications. At least for now, IB leads 10GE in this battle and enjoys a first-mover advantage.

Ethernet is here to stay. To thrive and grow in this battle, IB will have to continuously renew and add to its value propositions for the HPC and enterprise datacenter markets, while building a strong ecosystem with various players involved to help grow the community.

## ABOUT THIS ANALYST

*Jie Wu, Research Manager within IDC's Technical Computing team, provides market analysis, research, and consulting in the Technical Computing space. Specific research areas include market sizing, segmentation, competitive analysis and forecasting of the Technical Computing market. With the expertise in building results from field studies, Ms. Wu provides insightful guidance and actionable recommendations to the Technical Computing community.*