InfiniBand: The Eight Myths

Here are the eight commonly perpetrated myths and misconceptions about InfiniBand:

1. ***Myth: InfiniBand is lower latency than Ethernet***

   Busted: InfiniBand vendors advertise latency measured via specialized micro benchmarks with two servers in a back-to-back configuration, which is meaningless to real applications. Application level latency in a real cluster is what matters. Since IB lacks congestion management and adaptive routing, it quickly hits hot spots even in clusters of moderate size. iWARP over Ethernet, in contrast, achieves reliability via TCP, which results in a lower effective latency for useful applications.

2. ***Myth: QDR-IB has higher bandwidth than 10GbE***

   Busted: InfiniBand marketers conveniently specify the raw link level bandwidth, as opposed to Ethernet, where the speed refers to user bandwidth: a 10Gb IB link is effectively 8Gb. Furthermore, InfiniBand cards, like Ethernet cards, are limited by PCIe-Gen2 x8. Independently of how many 10Gb or 40Gb ports an adapter exposes, the aggregate bandwidth is limited to about 26Gbps in each direction. Therefore, Chelsio's T4 based adapters and the leading IB adapters offer the SAME bandwidth.

3. ***Myth: InfiniBand scales better than 10GbE***

   Busted: Whereas a significant percentage of the large clusters installed in the past used InfiniBand, the reason is historical – the absence of Ethernet media faster than 1GbE – rather than any inherent IB scaling capabilities. In fact, with the advent of cost-reduced iWARP over 10GbE, there is no longer need to live with the shortcomings of IB in lack of congestion control, and susceptibility to hot spots for the sake of high speed connectivity. Large clusters using iWARP are now starting to be deployed, with the first 1000+ node cluster (first in the top-500) recently announced. With the versatility of Ethernet, such clusters can take on multi-purpose roles, and converge storage, networking and high-performance computing.

4. ***Myth: IB is cheap***

   Busted: IB and Ethernet switch port prices have reached parity. The same can be said about adapter prices. However, an IB cluster further requires an Ethernet switch for management, a gateway for routing, and expensive IB storage available from a limited set of suppliers, as well as specialized IT personnel.

5. ***Myth: IB enables convergence***

   Busted: Network convergence is happening over Ethernet. InfiniBand is not part of the solution, leaving IB vendors scrambling to run InfiniBand directly over Ethernet (called RDMA over CEE – RoCE), requiring "lossless" operation unlike iWARP, which benefits from TCP's congestion control, reliability and efficient recovery.

6. ***Myth: IB has a large ecosystem of vendor support***

   Busted: IB is effectively driven by a single switch/NIC vendor, which controls 90% of the market. Only one other vendor holds the remaining share. Ethernet is by nature familiar, multi-vendor and widely supported.

7. ***Myth: IB is low power***

   Busted: An IB cluster requires management switches and routers that burn additional power compared to Ethernet. Furthermore, an Ethernet iWARP adapter offloads the host CPU via the built-in TOE, thereby saving power for all applications.

8. ***Myth: IB has a stronger technology roadmap than Ethernet***

   Busted: Ethernet has a superior roadmap, with 4x10GbE today, going to 40GbE in 2011, and 100GbE in 2012, of fully terminated, overhead-free, user bandwidth. IB's roadmap ends at EDR (Eight Data Rate), which will only deliver 50Gb of user bandwidth.

Related links:

1. *Sandia Labs Research Tech Note*
2. *IBM Research Report on IB and 10GbE Performance for HPC Applications*
3. *NFSRDMA vs. IB from Sandia*
4. *IBM/Blade Networks Presentation*
5. *Purdue University 10GbE Coates Cluster Whitepaper*