Thursday, January 29. 2009

Future datacenter fabric: QDR IB or 40/100GbE?

There are several people thinking that the search for future data center fabric technology is already decided for Ethernet. I don´t think so. I still think of Infiniband as a quite capable contender in this discussion. 40 GB/s Infiniband is already available whereas 40 GBit/s Ethernet is still in discussion. With the current surge of virtualization the datacenter is in need of faster protocols with lower latency right now connecting lots of servers to build a compute fabric ... today, not in 2010. Thus the enterprise datacenter of tommorow isn´t much different to the HPC datacenter of today and HPC is a stronghold of Infiniband.
Seeking Alpha puts another viewpoint into this discussion: David Gross notes in "Will New QDR InfiniBand Leap Ahead of 40 Gigabit Ethernet?, that the costs for respective 40 GBit/s variants of both technologies are really in favour of Infiniband:What's been impressive about QDR InfiniBand is not the fact that it's here, but its cost, under $500 a port, or less than a 10GBASE-SR transceiver module. OC-768 data ports still cost the same as they did seven years ago, about $600,000, and will need a lot than the 100+ ports AT&T (T) has purchased for its MPLS core in order to come down in price. Moreover, OC-192 Packet-over-SONET ports still go for well over $100,000, even with intermediate range 1310nm optics.In the case of 40 GBit/s i´m not really sure if we will see a similar cost reduction like in other Gigabit variants. Optical 40 and 100 GBits are based on wavelength division multiplexing Even the less complex CWDMs are quite expensive, not to speak about DWDMs nescessary for 100 GBit/s. The cabling concepts for both doesn´t like really much different in terms of complexity and ease of use than at Infiniband. The Ethernet Alliance Technology Overview states on page 11:The effective data rate per lane is 10 gigabit per second, which when 64B/66B encoded results in a signaling rate of 10.3125 gigabaud per second.. Thus, the 40GBASE-KR4 and 40GASE-CR4 PMDs support transmission of 40 Gigabit Ethernet over 4 differential pair in each direction over either a backplane or twin axial copper cabling medium, while the 100GBASE-CR10 PMD will support the transmission of 100 Gigabit Ethernet over 10 differential pair in each direction for at least 10m over a twin axial copper cable assembly. Ethernet can´t play with the advantages of former generations of Ethernet (simplicity, smaller acquisition costs) that killed ATM, FDDI et al in the 90ies. So Mr. Gross annotation speaks a true word:Sharing so many components, including LVDS signaling, 8B/10B encoding, SerDes transceivers, integrated Clock Data Recovery, and now Clos switching, the economics of deploying a particular link layer protocol are increasingly a function of connection distance, configuration, and transceiver reach, not the name of the framing technology.At 40 GBit/s we have a different field, we have an leveled field for the next round of the game. The discussion about the future datacenter fabric will be an interesting one.

Posted by Joerg Moellenkamp in English, The IT Business at 16:41