Fabric Computing That Works

RDMA over Converged Ethernet

Bob Pearson

System Fabric Works, Inc.

OFA/IBTA Booth 6010

Outline

- InfiniBand ‘pro’s
- InfiniBand ‘con’s
- RoCE
- InfiniBand vs RoCE
- What should you be doing

InfiniBand ‘pro’s

- Well understood widely deployed HPC interconnect
  - Approximately 50% of Top 500 list
- Characteristics:
  - High bandwidth
  - Low latency
  - Strong reliability
  - Rich set of messaging protocols
  - Layered architecture (OSI model based)
  - Green low CPU utilization and power

InfiniBand ‘con’s

- It’s not a LAN
  - Not routable (at least in production)
  - Scaling limitations (approx. 10K nodes)
  - Not ubiquitous (can’t connect multiple clusters in one big fabric)
  - It’s not Ethernet (some folks just won’t make the leap)

RoCE

- Stands for RDMA over Converged Ethernet is an IBTA standard.
- InfiniBand L3-Ln over Ethernet Layer L1-L2
  - Effectively InfiniBand and Ethernet have converged their phy's at 40G and above; so really just InfiniBand with an Ethernet L2
- Has OFA (OFED & upstream) support

InfiniBand vs. RoCE

- IB is the performance leader for now.
  - IB making transition from QDR->FDR, Ethernet just hitting 40G now, it's a race to 100G.
- RoCE is the new kid on the block
  - It is ready for early deployments, it is getting traction outside of HPC
- RoCE addresses some (but not all) of the IB con's
  - It's Ethernet and can be ubiquitous☺, but it's not routable ☹
  - RoCE has hard and soft implementations, so it can run anywhere

What should you be doing?

- Do you have issues that are addressed by RoCE? If no, then thanks for coming, see you later.
- If you do
  - Make your Ethernet investments RoCE ready
    - Low latency, DCB enabled,
  - Think about what you could do with one continuous fabric
  - Ask your vendors for RoCE solutions
  - Do POC's
  - Come talk to us