Overview of IBM HPC Interconnect Strategy

ScicomP 7/06

Clustering Configurations

- **Parallel Application Connectivity**
  - **‘Tightly Coupled’ S&TC**
  - **Parallel Data Base**
  - **IP connected**
- **Shared Data S&TC and BI**
  - **Multi-Cluster or Shared Data Base**
  - **Machine Room Floor**
  - **Physically Dispersed (Grid)**
- **Data/File System**
  - **SANS**
- **HA**
  - **Redundancy and Failover**

Interconnect Characteristics

- **Single Cluster – Performance**
  - **Scalability**
  - **BW**
  - **Latency**
  - **Collectives/Barriers**
  - **Bisection**
  - **CPU Utilization**
  - **Failover/Striping**

Interconnect Characteristics

- **Other Considerations**
  - Heterogeneous/Multi-Cluster Networks
    - Standards Based
    - Multi-Vendor Solution
  - Emphasis
    - Price
    - Price/Performance
    - Performance
  - Multiple Source
  - APIs
- **Large Range of Requirements**

One Size does not Fit All

Adapter Types

- **Adapter (HCA) Server Attachment Method**
  - Internal ‘Proprietary’ Bus Attachment
    - Optimizing Performance for the Server
  - Open/Multi-Vendor Slot Attachment
    - Facilitates Heterogeneous System Solutions
- **Interconnect Fabric Type**
  - ‘Proprietary’ Protocol and/or Network
    - Value Add over current Industry Standards
  - Industry Standard Carrier and APIs

IBM System Interconnect Types

- P-Series
  - Internal Slot/Proprietary Network – SP-Switch-2 – P3, HPS/Federation
  - Internal Slot/Standard Network – IB-12x (SDR/DDR)
  - Open Slot/Standard Network –IB-4x (DDR/QDR), Ethernet
- X-series
  - Open Slot/Standard Networks
- eServer 1350
  - Open Slot/Standard Networks
- Z-series
  - Internal Connection/Proprietary Network for Coupling Facility
  - IB may provide a way to offer an Open Coupling Facility
- BlueGene
  - Internal Connection/Proprietary Networks

Interconnect Type Evolution

- P-series
  - High End - Focus on advancing Standard Networks Following HPS
    - IB-DDR/QDR
    - Low Latency (User Space) Ethernet
    - Collective Offload/Acceleration
    - Combination of Internal Attachment Point and Open Slot HCAs
  - Standard Networks
    - IP: Ethernet; 1Gb, 10Gb, 40-100Gb, Lower Latency
    - MPI: IB and (LL)Ethernet
    - Parallel DB and ‘commercial’: IB4x/12x
  - Work with Industry to extend PCI-Express to DDR (QDR)
    - To match Internal I/O Bus capability
- Other Series Deep Computing
  - Use Standard Networks and Standard Attachment points
- Research
  - Continue to look at new ways to use networks especially for large Scale-out Solutions

IBM HPC Interconnect Strategy

- **Focus on Scaling up and improving Interconnects that are standards**.
  - InfiniBand: make production ready and scale-up. Integrate higher link speeds into products over time. Work with OpenIB/OpenMPI groups
  - Ethernet (1Gb then10Gb, 40Gb): Work on lowering latencies, partner with Ethernet vendors to improve its high performance capabilities.
  - Collaborating with switch vendors for both InfiniBand and Ethernet

Infiniband: GX attached Std Slot attached	Developing 12x/4x InfiniBand HCA (adapter) for Power6 and Follow-ons; Industry is working toward 4x DDR/QDR
HPC Ethernet: Std Slot based	Working with Partners to develop lower latency solutions

- Continue to provide “production ready” HPC software enabled on key interconnect. IBM base and Open Source Base