IBM b-type Networks for High Performance and Scalable Computing Systems

Highlights

- Low latency—Low latency, cut-through switching enables minimal delay and jitter required for HPC applications.

- High bandwidth—Line-rate, full mesh, non-oversubscribed 10 Gbps supports the evolution from terascale to peta-scale architectures

- High density and scalability—Facilitate large-scale build-outs of hundreds or thousands of tightly-interconnected compute nodes with exceptional high-density switches

- Fault tolerant design—High availability, reliability, and serviceability features on every switch include hot-swappable and redundant components

- High power efficiency—Industry leading power efficiency dramatically reduces power and cooling expenses in large scale deployments

IBM® and Brocade® have strengthened their long-standing relationship to provide expanded network technology choices with the new IBM b-type Ethernet switches and IP routers to provide an integrated, end-to-end scalable and low latency framework. Combined with IBM's vast data center design experience and Brocade's networking expertise, this portfolio represents the ideal convergence of strength and intelligence. For organizations striving to transform their IT infrastructure, this portfolio helps reduce costs, manage risks, and prepare for the most demanding applications while improving service.

Speed, processing power and scalability-these are the common themes in High-Performance Computing (HPC) systems. What characterizes these turbo-charged installations is the need to connect hundreds to thousands of compute resources at extremely high speeds. The challenge is to handle trillions of operations per second.

As parallel programming models for HPC applications evolve, more and more HPC installations are designed based on standard processors and Graphics Processing Units (GPUs), tightly connected to form scalable systems. Networking solutions are helping these HPC systems move from terascale (10^{12}) to petascale (10^{15}) operations per second and beyond.

Networks for HPC systems perform a variety of roles: cluster interconnect, management, storage, and access. Networking technologies available today include: 1 and 10 GbE; proprietary interconnects; SDR, DDR, and QDR InfiniBand (IB); 4 and 8 Gigabit per second (Gbps) Fibre Channel (FC); and, in the future, Data Center Bridging or Converged Enhanced Ethernet (DCB or CEE).

As with any network configuration, the HPC network system design involves tradeoffs between overhead, manageability, and price. Of the above choices, Ethernet has the advantage of familiarity, ease of

operation, and integration into existing networks. Further, 1 and 10 GbE are the top interconnect technology choices among the world's Top 500 supercomputing sites in 2009.

HPC interconnects also need very low latency, high bandwidth, and extreme scalability. As Ethernet technologies evolve, standards-based Ethernet networks at 10 GbE and beyond deliver data rates, stability, and scalability that will readily meet the most demanding HPC requirements.

IBM b-type switches leverage Brocade technology to bring experience, innovation, and market leadership to all flavors of Ethernet, which make IBM network solutions the logical choice for HPC deployments.

IBM b-type networking solutions for HPC environments

HPC systems are utilized in a variety of applications-with different interconnect bandwidth requirements-and for simulations of various sizes in R&D and industry environments. System size varies from Windows-based small clusters for a department or a workgroup, to very large installations that run weather or physics simulations in national labs. Depending on system size and application requirements, IBM b-type networking solutions for these HPC systems can include one or more of the components in the next table.

IBM b-type networking products for HPC systems

Product Name	Standard Ethernet	CEE Support	Product Summary
IBM x-series	10 / 1 GbE	--	Compact 1U, low-latency, cut-through, high-performance, 24 10/1 GbE dual-speed ports plus 4 10/100/1000 RJ45 ports. Ideal as a top-of-rack and aggregation switch.
IBM y-series	1 GbE	--	Compact 1U, high performance, low-latency, top-of-rack switch with up to 48 10/100/1000 RJ45 ports and up to 4 10 GbE uplink ports
IBM m-series	10 / 1 GbE	--	Advanced routers with non-oversubscribed 10 GbE (up to 128 ports) and 1 GbE (up to 1,536 ports), wire-speed performance. 4, 8, 16 or 32 slot chassis
IBM r-series	10 / 1 GbE	--	Advanced switches with high port density of 10 GbE (256 ports, 4:1 oversubscribed or 64 ports 1:1 non-oversubscribed) and 1 GbE (768 ports 1:1 non-oversubscribed). 8 or 16 slot chassis
IBM Converged Switch B32	10 GbE	Yes	Compact 1U, 24 port low-latency 10 Gbps CEE and 8 port 8 Gbps FC, cut-through, non-blocking. Top-of-rack switch.
IBM System Storage **SAN768B** & **SAN384B**	10 GbE (w/ the 10-24 FCoE blade)	Yes	SAN768B: up to 384 8 Gbps FC ports. SAN384B: up to 192 8 Gbps FC ports, up to 3 per rack. FC and/or CEE/FCoE blades
IBM Converged Network Adapters	10 GbE		Dual mode, NIC & CNA, high-performance adapters. 1 or 2 port SFP+ configurations

Brocade network innovation

Brocade has been addressing the network infrastructure needs of HPC solutions since its inception. As the demand for low latency and increased bandwidth continues, Brocade is on the front line, working with IEEE on 40 and 100 GbE standards. These 40 and 100 GbE high-speed links will enable customers to link multiple HPC clusters and to maximize the computational power of large, distributed systems.

Brocade continues to innovate by being the first to deliver end-to-end 10 GbE CEE solutions to further optimize network traffic by creating low-latency, lossless topologies that can also carry Fiber Channel traffic. Since CEE is a lossless protocol, networks can optimize traffic flow intelligently to ensure that data is not dropped. This is critical in Storage Area Networks (SANs) today and is now utilized in Ethernet networks, transforming them from "forward and forget" networks to predictable, guaranteed delivery systems.

To help further scale and reduce latency in the network, Brocade is a leader in the development of Transparent Interconnection of Lots of Links (TRILL), which enables network administrators to overcome the shortcomings of Spanning Tree in their networks today. TRILL can leverage primary and redundant links by load balancing traffic over all available network links. This allows a network architect to design fully meshed, resilient, and scalable systems and maximize the available bandwidth.

Brocade is committed to the HPC market and will continue to design Ethernet networking solutions that have very low latency, are extremely scalable, have optimal power design, are easy to manage and will support CEE.

b-type network solutions examples

Small cluster with 1 GbE server uplinks

A small cluster with a non-oversubscribed interconnect topology is shown in Figure 1. It uses the IBM y-series Top-of-Rack (ToR) switches and IBM Ethernet Switch B24X for aggregation to provide a non-oversubscribed Ethernet interconnect that can move data to the data center network or to a storage network at line rate. The IBM B48Y switch provides 48 GbE ports and four 10 GbE uplink ports in a compact 1U design. The industry-leading 4-port 10 GbE density in the IBM B48Y provides up to 40 Gbps of uplink bandwidth to the core switch of the network. This bandwidth enables a

near 1:1 subscription ratio throughout HPC network.

To aggregate the 10GbE uplinks from the Top or rack switches, an IBM B24X can provide a non-oversubscribed Ethernet interconnect that can move data at line-rate. The IBM B24X is an ultra-low-latency, cut-through, non-blocking 1U switch with 24 port of 10 with four additional 1 GbE ports. The cut-through speed of the IBM B24X, aggregates the flows quickly and moves the data where it can be used or stored.
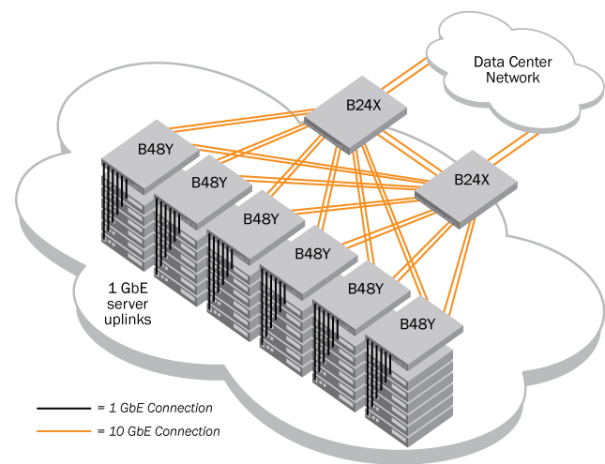

Figure 1

Medium cluster with 1/10 GbE server uplinks

A medium cluster with a non-oversubscribed interconnect topology is shown in Figure 2. When server attachments are 10 GbE, one or multiple IBM B24X switches are optimal ToR switches. They offload data from servers and send it to the core of the network at cut-through speeds. This non-blocking design ensures that all, not just some, of the data is delivered quickly and efficiently. The IBM B24X can support either 1 GbE or 10 GbE on all 24 ports, enabling a mid-sized data center cluster to easily migrate to 10 GbE servers all at once or a few at a time and still be configured for 1:1 cut-through delivery of traffic from the rack to the core of the network. If server links are all 1 GbE, the IBM y-series can be used as a ToR switch.

Figure 2 shows a split core using the highly scalable IBM m-series or r-series chassis which are capable of providing up

to 128 ports of 10GbE per chassis for non-oversubscribed topologies. When the traffic arrives at the m-series switch, the industry's most advanced hardware-based buffering system utilizing a multi-stage Clos fabric ensures that no traffic is lost and that all traffic arrives safely. The m-series switch is a fully redundant, non-blocking, full-bandwidth chassis capable of handling the most intense data flows. For an over-subscribed topology, the r-series chassis can provide up to 256 10 GbE ports.
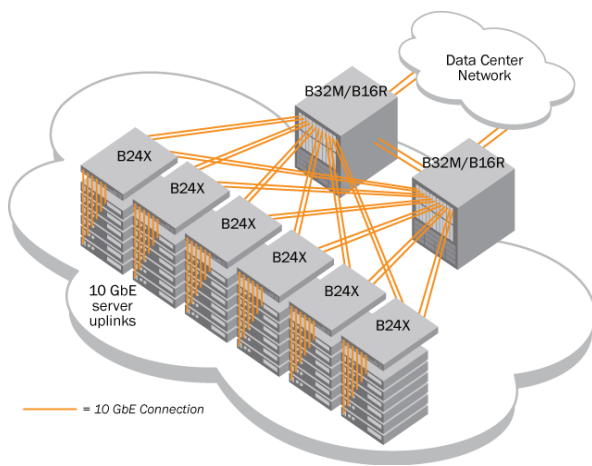
Figure 2

High-density, large cluster with 10 GbE server uplinks

A large cluster with a non-oversubscribed interconnect topology is also shown in Figure 2. Similar to the medium cluster, in a large cluster, the IBM B24X ToR switch is running 1:1 subscription; non-blocking, ultra-low latency, cut-through switching that moves data out of the rack and onto the core at line rate. Once at the core, in a large IBM m-series chassis such as the B32M, traffic is switched or routed to its destination. The advanced Quality of Service (QoS), buffering, fully non-blocking fabric, and the fully redundant features of this platform guarantee reliable and predictable flows at peak performance.

Many other network designs for HPC systems, with or without oversubscription and depending on individual customer requirements, are possible using IBM b-type Ethernet switches and IP routers.

For more information

To learn more about IBM b-type Ethernet switches from Brocade, please contact your IBM marketing representative or IBM Business Partner, or visit the following Web site:

• www.smarternetworks.com