IBM Systems Lab Services
Whitepaper

Building InfiniBand scaffolding for IBM Elastic Storage Server and IBM Spectrum Scale

By Ahmed Hazem Elsherif and Rakesh Chutke
IBM Systems Lab Services

Every day, storage disk infrastructure is becoming exponentially faster. Technologies like solid state drives (SSD) and Non-Volatile Memory Express (NVMe) are becoming the backbone for most modern storage infrastructures due to their massive data handling capability and sub-millisecond response times. To drive technologies that support huge bandwidth, businesses need a fast, reliable network like InfiniBand.

InfiniBand is a highly reliable low latency network for extremely high throughput systems such as high-performance computing (HPC) and analytics. In this paper, we will summarize the ways to use InfiniBand effectively for HPC and analytics environments, focusing on IBM Elastic Storage® Server (ESS) and IBM Spectrum® Scale.

What is InfiniBand?

InfiniBand is a type of communications link for data flow between processors and input/output (I/O) devices that offers throughput of up to 25 gigabytes per second for a single connection and supports up to 64,000 addressable devices. InfiniBand is highly scalable, supports quality of service (QoS) to meet expected performance criterion and is often used as a preferred way to connect computers in HPC environments.

How does InfiniBand work?

In InfiniBand, data is transmitted in packets that together form a communication called a message. A message can be:

- Remote direct memory access (RDMA) performing a read or write operation
- When a channel sends or receive message
- A reversible transaction-based operation or a multicast transmission

InfiniBand implements a messaging service for applications called channel I/O, which bypasses network operating systems in order to achieve high performance in specialized environments. It enables two InfiniBand-enabled applications to create a direct communication channel with send and receive queues called queue pairs. The queues map to memory spaces accessible to each application for data sharing (called RDMA).

The queue pair (QP) is one of the primary architectural elements of InfiniBand. In InfiniBand, communication occurs between queue pairs, instead of between ports. A queue pair is an addressable entity that consists of two work queues, a send work queue and a receive work queue. The channel adapter hardware arbitrates communication by multiplexing access to the send queue or demultiplexing messages on the receive queue.

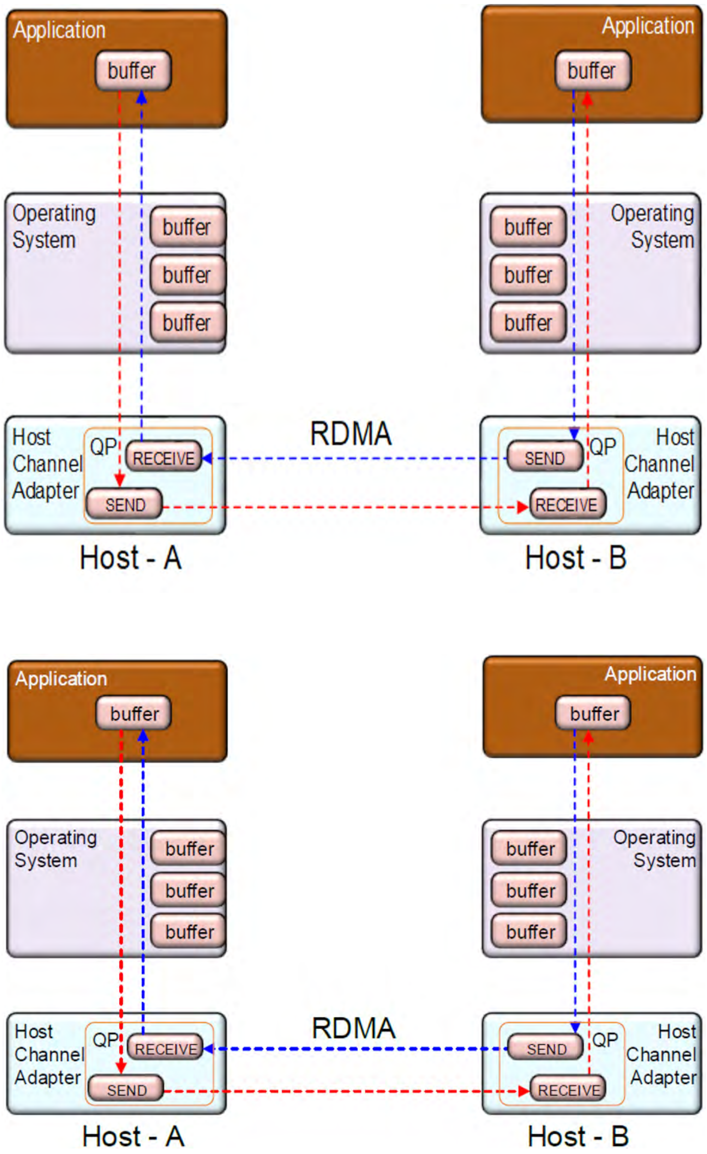

Figure 1: InfiniBand versus TCPIP operation

Why InfiniBand?

Greater speed: InfiniBand currently supports the following speeds:

- FDR at 56 Gbps
- EDR at 100 Gbps
- HDR at 200 Gbps

In 2021, it’s expected to support NDR at 1.2 Tbps.

IBM ESS storage systems currently support EDR technology of 100 Gbps speed.

Lower latency: Using advanced ASICs and silicon chip technologies, port latency on InfiniBand ranges between 0.5 μs and 1 μs.

Scalability: Multiple parallel processing applications in data centers or HPC, with scalability up to 48,000 nodes, can be accommodated in a single subnet without any performance penalties.

CPU off-load: CPU off-load is the key benefit of InfiniBand technology, where there is almost no use of system (server) hardware resources such as CPU or software resources such as operation system kernel and TCP stack.

In 2016, an offloading versus onloading test was performed by Mellanox that included send-receive data transfers at the maximum data speed supported by each interconnect (~100 Gbps) while measuring the CPU utilization. At the data speed of 100 Gbps, InfiniBand consumed only 1 percent CPU utilization, while other technologies such as Ethernet or Intel's Omni-Path consumed 59 percent CPU utilization for the same task. With InfiniBand, almost all the processor cores are available for applications' processing.

Also, it should be noted that Intel recently announced the sunset of the Omni-Path fabric technology, which means InfiniBand is the fastest technology available as of now for HPC workloads.

What are the different InfiniBand fabric components to consider in a bill of material?

InfiniBand fabric consists of four main components to run the small network and three additional components that are required for large-scale clusters.

The main components of InfiniBand fabric are:

1. Host Channel Adapter (HCA)

- Transfers messages from server into InfiniBand fabric
- Offers wide support for Linux distributions (Windows also available)
- Creates queue pairs (QP represents a communications endpoint, consists of a SEND queue and a RECEIVE queue)
- Provides RDMA support

2. InfiniBand switch

- Moves packets from one link to another in the same InfiniBand fabric
- Varients include Top of Rack (TOR) and Director Class switches

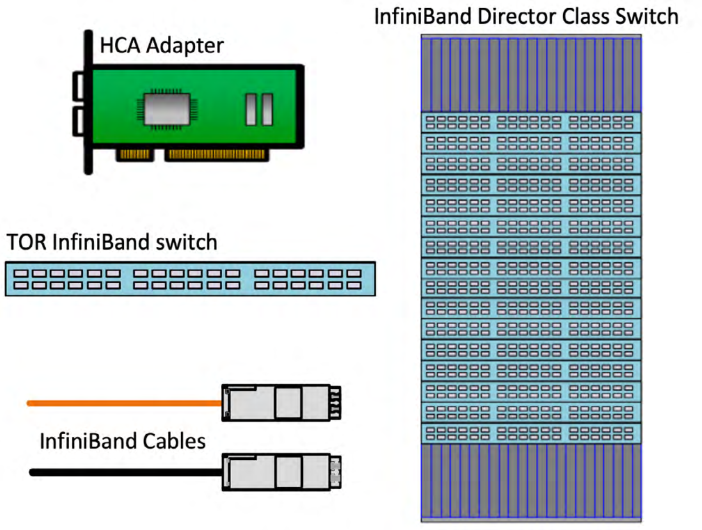

Figure 2: InfiniBand fabric main components

3. Cables

- Allow data transfer over copper or fiber media
- Connect with servers and switches using buit-in transceivers
- Support maximum length up to 100 m with multimode-fiber cables
- Permit extension up to 10 km using standalone transceivers

4. Subnet manager software (SMS)

The InfiniBand subnet manager (SM) is a centralized entity running in the switching system. The following are the different functions of subnet manager.

- Provides L2 addressing to all connected InfiniBand devices in the fabric (like MAC address in Ethernet)
- Provides routing packets within the fabric (to select the best route between the source and destination)
- Has SMS that runs on either switch's management modules or on external standalone Linux servers connected to same network
- Is license-free software, can be installed either on the switch or an external Linux server (customers will have to pay for licenses if they are using UFM [Unified Fabric Manager], which is an optional fabric management software)

The following are additional components for expanding ESS and Spectrum Scale connectivity with an Ethernet network:

1. Gateway

Gateway is a switch in the network, where some of its ports are configured as Ethernet and others as InfiniBand. This is essentially a bridge to enable Ethernet devices to communicate with InfiniBand devices and vice versa.

For a hybrid network of Ethernet and InfiniBand, Proxy-ARP is used to forward IPv4 packets from the Ethernet network to the InfiniBand network and vice versa. Proxy-ARP is not an IP router; rather, it acts as a bridge that forwards the IPoETH (IP over Ethernet) packets to IPoIB (IP over InfiniBand) in Unreliable Datagram (UD) mode. The Proxy-ARP forwards the traffic within a single subnet.

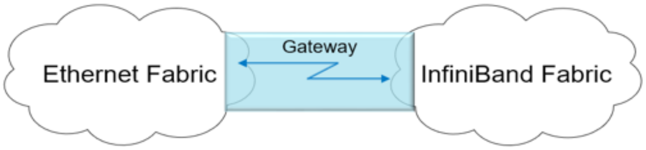

Figure 3: InfiniBand to Ethernet Gateway

In InfiniBand to Ethernet Gateway, the Ethernet switching function operates in layer 2 only, which means IPoIB nodes on InfiniBand fabric must be on the same IPv4 subnet as Ethernet connected external nodes. If different IPv4 subnet address ranges need to connect with the InfiniBand network, then Layer 3 switching must be in place to have Ethernet fabric communicate with the InfiniBand network through the Gateway switch.

In the IBM portfolio, IBM 8831-NF2 or IBM 8831-F36 supports the gateway function through a Mellanox VPI (Virtual Protocol Interconnect) license, which must be purchased separately, as it is not part of IBM e-config.

2. Router

InfiniBand routers are mainly used to segment a very large network into smaller subnets connected by an InfiniBand router. The segmentation may be useful for isolating some of the subnets from each other, or for building a very large network (40K+ nodes).

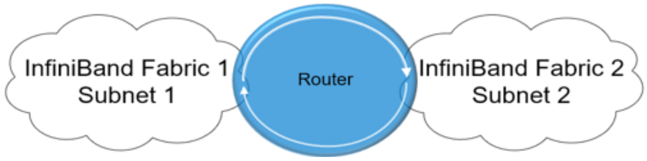

Figure 4: InfiniBand router

With a single hop topology, as shown in figure 4, where Subnet 1 and Subnet 2 have their own subnet managers, they can communicate with each other using Layer 3 (which is InfiniBand Layer 3). One should not relate InfiniBand Layer 3 function with the Layer 3 switching used in Ethernet technology, as they are separate entities.

3. Fabric management software (UFM)
Mellanox's comprehensive suite of management software provides a management solution that allows users to manage small to extremely large fabrics and enables fabric monitoring and performance optimization at the application-logical level rather than merely at the individual port or device level.

Mellanox's Unified Fabric Manager (UFM) performs the following functions:

- Monitoring historical events
- Advance alerting
- Advance optimization
- Fabric monitoring
- Performance management
- Automated provisioning

InfiniBand network design and topologies

InfiniBand offers versatile fabric connectivity topologies, depending on size and needs.

Back-to-back host connections
Direct connectivity between two Linux-based hosts is the smallest topology that you can configure with InfiniBand. You can connect one or more HCA adapter ports to their neighbors directly without having an InfiniBand switch. You need to additionally enable subnet manager on one of the hosts. The necessary software to configure the subnet manager function on the compute node is included in the OFED software package at no cost.

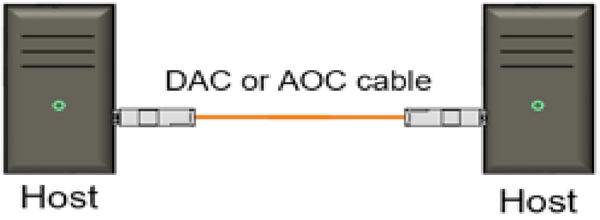

Figure 5: Host-to-host InfiniBand direct connection

Single switch
With a small number of nodes, it's possible to deploy a single InfiniBand TOR switch, running a subnet manager instance.

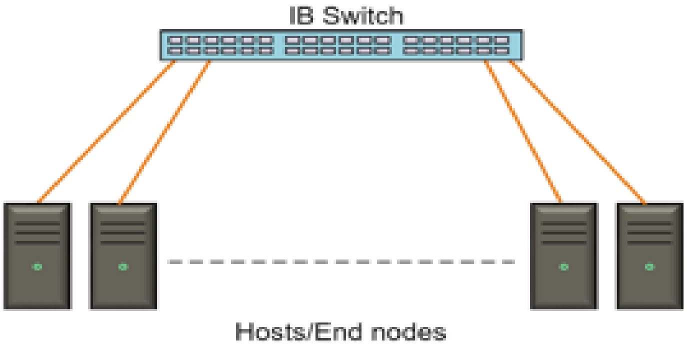

Figure 6: Single switch topology

Dual star
The dual star topology is suitable for an environment that has ESS storage and a small number of Spectrum Scale nodes. With two switches, you can have an InfiniBand fabric with high-availability subnet manager instances (each instance runs on a switch). Dual switch connectivity is implemented to achieve redundant paths and RDMA load balancing. It is important to interconnect the two switches through two to four connections (as shown in diagram, a direct link that connects both the switches together) to maintain IPoIB connectivity between ESS system nodes and client nodes connected to an ESS Spectrum

Scale cluster. This topology is suitable for one to two ESS systems with 10 nodes with dual connections or 20 nodes with a single connection.

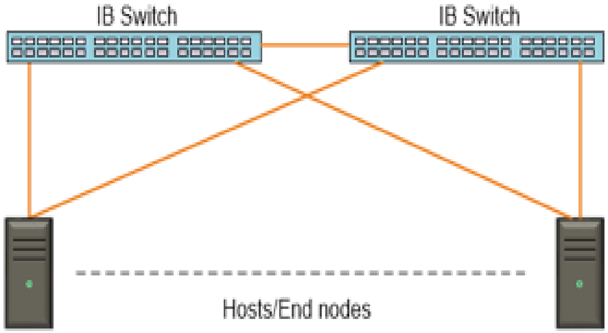

Figure 7: Dual switches topology

Spine and leaf
For scaling further to a larger number of nodes, spine and leaf topology is the best choice. Up to 648 nodes can be connected through this topology. Spine switches only provide connectivity to the underlying leaf switch, while leaf switches are serving nodes and host connections.

To achieve non-blocking fabric, the number of downstream links of a leaf switch must equal the number of upstream links to spine switches. For example, if you have 18 nodes with single connections, you need to connect 18 upstream links to spine switches.

For a 2:1 blocking ratio, you need to connect 8 upstream links to spines, and for 3:1, you need to connect 6 upstream links.

Spine and leaf topology is typically suitable for IBM offerings in HPC and Hortonworks Data Platform, with IBM ESS as backend storage.

Note: Non-blocking switch internal bandwidth is capable of handling all the port bandwidths, at the same time, at full capacity.

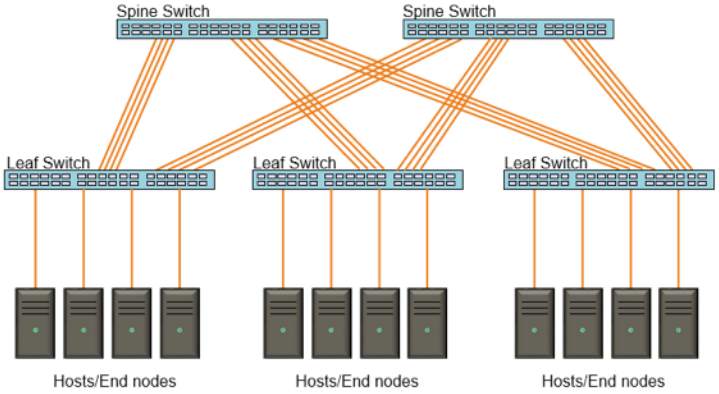

Figure 8: Spine and leaf topology

Torus
For large Spectrum Scale cluster deployment with thousands of nodes (such as supercomputers or grid computing environments), Torus topology is the right choice. Torus interconnect is a switchless topology forming a mesh interconnect with nodes arranged in a rectilinear array of N = 2, 3 or more dimensions, with nodes connected to their nearest neighbors, and corresponding nodes that are connected on opposite edges of the array.

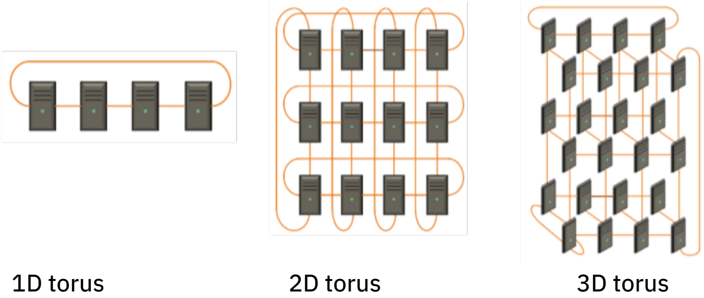

Figure 9: Torus topologies

Torus topology is ideally suitable for the cluster having thousands of compute nodes running parallel jobs involving massive processing and huge node to node network communication.

What are the different InfiniBand connectivity options for Spectrum Scale and ESS clusters?

Spectrum Scale and ESS support two types of InfiniBand communication topologies.

1. IPoIB (IP over InfiniBand): For IPoIB, the only available common communication channel between I/O nodes and Spectrum Scale clients is the InfiniBand network, hence the InfiniBand network is configured as a common source of communication for both administrative commands and daemon traffic. Here InfiniBand adapters are configured with independent IP addresses for I/O servers and clients. The RDMA can then be enabled on InfiniBand adapters of both clients and I/O servers. Additionally, you can configure bonding in the Linux OS, aggregating multiple InfiniBand ports and thereby enabling high availability configuration and support of higher bandwidth. When you are dealing with IPoIB mode with bonding configured, IP traffic always follows active-passive mode, which means out of two InfiniBand ports in the bond only one will be used for sending and receiving IP traffic. On the other hand, RDMA traffic follows active-active transmission mode, meaning it will try to use all available ports in the bond to send and receive RDMA traffic. When IPoIB topology is used along with RDMA enabled, Spectrum Scale will automatically send all administrative communication using the IP protocol and all daemon communication using the RDMA protocol by injecting both protocols to the same communication channel, which will dramatically improve performance as compared to a conventional Ethernet IP network when configured for both daemon and administrative communication.

Pros:

- No separate Ethernet adapters and Ethernet switches are required, as the entire communication is over InfiniBand.
- The RDMA performance on the InfiniBand network is not compromised due to use of the IPoIB configuration option.
- System performance is not compromised even though IP and RDMA are configured on the same InfiniBand port, as the same HCA (InfiniBand adapter) buffers will be serving both (IP and RDMA) of them and not just RDMA.
- Host System CPU utilization is very minimal when RDMA is enabled and hence more processing power is available for application workloads.
- Performance tuning for optimum efficiency is relatively simple in the case of an InfiniBand network as compared to a pure Ethernet IP network.

Cons:

- Connecting an external Ethernet landscape with an InfiniBand network is a difficult task, as you need additional gateway switches to extend the InfiniBand network to the external Ethernet world. The typical use case would be if you want to do remote mount of an InfiniBand cluster file system on an external cluster running over an Ethernet network, then you may need gateway switches for bridging connectivity between the InfiniBand and Ethernet network.

2. Separate Ethernet and InfiniBand network for the same Spectrum Scale cluster: You can also choose to separate admin and daemon traffic on two different networks. In this case, a separate Ethernet network can be configured with an IP address to carry cluster administrative communication, while an InfiniBand network with only RDMA (VERBS) protocol enabled will carry daemon traffic. When Spectrum Scale daemon starts, it will automatically detect if there is any InfiniBand path available with RDMA configured. It will then start using that path for transporting Spectrum Scale daemon traffic over RDMA while all admin traffic goes over the Ethernet IP network.

Here, you are mainly using the InfiniBand network for RDMA base communication, and hence you don't need to configure IP address and bonding aggregating multiple InfiniBand ports on Spectrum Scale clients and ESS I/O servers.

There is not much performance difference when you compare IPoIB and separate Ethernet and InfiniBand network topology because in both the cases daemon traffic will automatically use the RDMA protocol to carry data from the compute node to ESS I/O nodes and vice versa.

Pros:

- When an entire InfiniBand network goes down, Spectrum Scale will start transporting admin and daemon traffic over the Ethernet network. Hence, application data access continuity is maintained with reduced performance.
- Some types of communication are more efficient over TCP/IP. For example, if packets are less than 8 KB, by default, GPFS uses the TCP/IP connection to move the data. This value can be configured using the verbsRdmaMinBytes parameter. The default value of verbsRdmaMinBytes is 8192.
- Small writes for ESS will benefit if a separate InfiniBand network is used along with RDMA enabled. ESS uses a parameter "nsdRAIDFastWriteFSDataLimit=512k," which means any write of size 512k or less hitting to ESS will be absorbed by NVRAM (log tip disk). The same NVRAM contents are further replicated to the neighboring ESS I/O node of the same building block to maintain data consistency. Any application that generates a huge number of smaller writes than the "nsdRAIDFastWriteFSDataLimit" value will get absorbed immediately by ESS NVRAM to hide the latency of slower back-end NLSAS disks. When you have a separate InfiniBand network for daemon communication, it will tangibly improve application write response time as most of the small writes will hit to the preferred I/O node NVRAM, further replicating the same writes to a neighboring I/O node NVRAM through the low-latency InfiniBand network.

Cons:

- You need two different switching fabrics, one for Ethernet traffic and the other for InfiniBand traffic, with separate Ethernet and InfiniBand adapters connected to respective switching networks. The cost of the solution will be higher as compared to IPoIB topology.
- Your I/O server may not have enough PCI slots to accommodate a redundant number of Ethernet and InfiniBand adapters (for example, some POWER9™ servers such as AC922 as well as ESS GLxC models do not have PCI slots available to add more Ethernet and InfiniBand adapters).

Whether to select InfiniBand or Ethernet technology for ESS and IBM Spectrum Scale

When sizing workloads that demand very high performance and low latency, and where capacity is not the prime focus, it is always a good approach to select InfiniBand over Ethernet. The InfiniBand as Spectrum Scale clustering interconnect over Ethernet provides tangible boost in performance due to transport offloading in hardware, Linux kernel bypass and RDMA capabilities.

Advantages of InfiniBand

- You need lower compute processing to achieve the same performance in teraflops (Tflops). In order to drive 1 Tflops performance with an

InfiniBand network, you need 256 CPU (3Ghz Xeon processor) as compared to 1024 CPU (3 Ghz processor) with gigabit Ethernet.

- You need a smaller number of switch ports and network adapters in an InfiniBand network as compared to an Ethernet network and therefore the cost of InfiniBand becomes more competitive in comparison to Ethernet.
- Optimum max performance can be achieved with minimal tuning efforts in InfiniBand technology as compared to Ethernet.

Advantages of Ethernet

- Today's Ethernet network interface cards are equipped with kernel bypass and TCP acceleration in the hardware, which can deliver the same capability of bypassing the OS kernel and delivering a low latency host interface.
- RoCE (RDMA over converged Ethernet) support in Ethernet makes it equally competent by providing the same performance as InfiniBand at a lesser cost.
- Ethernet technology is progressing faster than InfiniBand, and there are lots of developers contributing and vendors manufacturing Ethernet technologies. The roadmap of Ethernet technology looks more promising. Ethernet technology still is a viable option for workloads that are capacity driven where performance requirements can easily be met within the given capacity.

With ESS supporting RoCE over 100 Gbps, Ethernet will improve the throughput of Ethernet network technology to multifold level, making it as competent as InfiniBand.

Selecting InfiniBand switches and HCA

When selecting an InfiniBand adapter and switches, more weight is given to the throughput requirement and type of workload.

InfiniBand switches selection:

Always select the 100 Gbps EDR switch models, either as TOR (top of the rack) (8828-E36) or Director Class switch (8828-ED0 / 8828-ED1 / 8828-ED2). To select either EDR TOR switch design or EDR Director Class switch design, consider the following important factors:

- Fabric oversubscription (node bandwidth versus upstream bandwidth)
- Total number of nodes' connections
- Cabling and total cost of ownership

Fabric oversubscription, with total number of connected nodes, determines the total number of switches in the fabric.

Examples:

- 40 x nodes with single connection, 1 x ESS system, 1:1 fabric oversubscription = 5 x 8828-E36 switches, OR 1 x 8828-ED0 director switch with 2 x spine modules and 2 x leaf modules
- 40 x nodes with dual connections, 1 x ESS system, 1:1 fabric oversubscription = 8 x 8828-E36 switches, OR 1 x 8828-ED0 director switch with 3 x spine modules and 3 x leaf modules
- 100 x nodes with single connection, 10 x ESS system, 1:1 fabric oversubscription = 17 x 8828-E36 TOR switches OR 1 x 8828-ED1 director switch with 6 x spine modules and 6 x leaf modules
- 100 x nodes with dual connections, 10 x ESS system, 1:1 fabric oversubscription = 24 x 8828-E36 switches OR 1 x 8828-ED1 director switch with 9 x spine modules and 8 x leaf modules

A winning solution with optimum design at minimal cost, requires an exercise of around six to eight hours to brainstorm and to sketch possible network schematics, taking into consideration HPC or AI solutions workloads, 1:1 oversubscription must be used to design the networking fabric, while in the case of big data and analytics workloads, you may go with either 1:1 or 2:1 or even lower to 3:1.

InfiniBand host channel adapter selection:

- Use 100 Gbps EDR adapters instead of 56 Gbps FDR adapters; however, if there is no need for higher throughput, you can still use FDR adapters to connect to either FDR/EDR switches by selecting FDR cables instead of EDR cables.
- Consider using PCIe4 adapters instead of PCIe3, especially on POWER9 servers.
- Consider ConnectX-5 adapters for HPC environments, since it provides a message rate of 200 million messages per second, which is 33 percent higher than the ConnectX-4 adapters. ConnextX-5 comes with feature code E6xx, while ConnectX-4 comes with feature code E3xx.
- In HPC environments where AC922 or LC922 are proposed, use PCIe4 CAPI enabled EDR InfiniBand adapters for the best performance.
- For Power® nodes with mixed Ethernet and InfiniBand adapter connectivity, do not use VPI (Virtual Protocol Interface) enabled adapters to operate as an Ethernet adapter, because it is not supported or tested on Power Systems™. Use VPI adapters for InfiniBand only, which is the default function of such adapters.
- ESS has its own e-config to select the right adapters, but it still uses POWER8® nodes with PCIe3 adapters; therefore, consider a single connection per adapter instead of two ports connections per adapter to avoid PCIe bus congestion. It is recommended not to select Mellanox Connectix-3 adapters as they do not provide stable performance, and throughput can be shaky when heavy traffic is pumped.

ESS configuration best practices for InfiniBand fabrics

The following are some best practices:

1. Regularly update your InfiniBand switches with the latest software images from the Mellanox support site. The software will automatically update switch firmware along with the Mellanox operating system.

2. Configure subnet manager for high availability if you have two or more InfiniBand switches in the fabric.

3. Install/upgrade the complete suite of Mellanox InfiniBand OFED on the compute nodes as per the installation procedure given in the OFED installation document.

4. After install/upgrade of OFED, it is highly recommended to run the mlnx_tune utility on InfiniBand client nodes. mlnx_tune is a static system analysis and tuning tool. It has two main functions: to “report” and to “tune.” The reporting function is used for running a static analysis of the system. The tool checks current performance relevance and system properties and tunes the system to maximum performance.

5. Use various built-in Mellanox commands and tools to the check latencies between ESS I/O nodes and target clients to check whether read/write latencies are within permissible limits.

- `ib_send_lat`
- `ib_write_lat`
- `ib_read_lat`

6. Use bandwidth measurement commands between ESS I/O nodes and end clients to check if you can achieve the required read/write bandwidth.

- `ib_send_bw`
- `ib_write_bw`
- `ib_read_bw`

7. Use ibdiagnet InfiniBand command to scan the InfiniBand fabric on ESS I/O nodes and client nodes using the directed route packets

method, extracting all the available information regarding the connectivity and devices. This command produces a set of files in the output directory located in /tmp which is default directory location unless you explicitly specified while running command.

Example:
The following example shows how to test the InfiniBand fabric with the `ibdiagnet` command. The command checks for 4x link width and 25 Gb/sec speed, and then dumps the Performance Manager counters and then clears them.

```
# ibdiagnet -lw 4x -ls 25 -pm -pc
```

8. An MTU size of either 2048 or 4092 on a Mellanox InfiniBand switch delivers similar performance numbers when it is operated as IPoIB with RDMA enabled; however, with IPoIB without RDMA, the bigger MTU size would improve throughput. The maximum size of MTU that the InfiniBand switch can support is 4092. It is advisable to set consistent MTU size across ESS I/O nodes/clients and on InfiniBand switches.

9. Enabling IPoIB is mandatory for ESS or Spectrum Scale when a cluster is formed using the InfiniBand network.

10. Consider moving all supported InfiniBand devices to the Datagram mode (CONNECTED_MODE=no) for more information refer to the RHEL Networking Guide.

11. Make sure all InfiniBand host adapters' port "0" is connected to one fabric and InfiniBand adapters' port "1" is connected to another fabric. Follow these connectivity guidelines meticulously to obtain optimum perfromnace and availability.

12. Do not configure InfiniBand port bonding when communication is pure RDMA without IP (for IPoIB configuration bonding is needed but, in that case, only one port will be active and the other port in the bond will be standby, which means IP traffic will flow through only one port while RDMA traffic use both the ports in a balanced way).

13. Connect InfiniBand cables to each port of ESS I/O nodes when you have only one InfiniBand adapter. It is a good idea to use only one port of a dual InfiniBand adapter when you have enough adapters for connectivity with the InfiniBand switch; this will avoid PCI bus saturation, providing steady and consistent performance output.

14. Do the InfiniBand performance testing using Spectrum Scale and ESS natively available tools and commands such as nsdperf and gssnettest to ensure that you get the full end to end read/write bandwidth. This test must be run before file system creation, as it generates a very high load on the network and hence may not be advisable to run when production workload is on.

15. Use the `mmnetverify` command to validate a cluster network before going into production or when you suspect network problems.

Conclusion

There are different views within the technical community about the consumption of InfiniBand technology for various workloads. Some architects are shying away from using InfiniBand technology, thinking it is too expensive as compared to Ethernet, but the fact is that there is not much difference between their costs, and therefore InfiniBand makes a strong use case for most modern workloads that demand extremely low response times and higher throughput since it is lossless fabric technology by its nature.

Through this paper, we have tried to emphasize the fact that InfiniBand networks are very reliable, easy to set up and require very little tuning to achieve the targeted performance. This paper also covered some of the data points including how effectively InfiniBand technology can be used in an IBM ESS and Spectrum Scale environment. Some of the points are deliberately accommodated to help architects make effective decisions on various aspects such as topology selection, sizing and selection of InfiniBand components for building networking scaffolding around an ESS and Spectrum Scale environment.

About the Authors

Ahmed Hazem Elsherif

Ahmed is a Lab Services Spectrum Scale and Networking consultant with more than 27 years of experience in products management and professional services. Having a wide domain in different technologies, his specialization in the past decade has focused on infrastructure transformation and application performance management technologies.

Rakesh Chutke

Rakesh is IBM Storage and Spectrum Scale consultant with IBM System Lab Services for the last 13 years. He has more than 19 years of IT industry experience working with clients across industries, including banking, insurance, telecommunications, media and entertainment, oil and gas.

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
February 2020

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.