Legal Notices

Warranty

The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

The information contained herein is subject to change without notice.

Restricted Rights Legend

Confidential computer software. Valid license from HP required for possession, use or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Copyright Notices

© Copyright 2009 Hewlett-Packard Development Company, L.P.
Building a high performance, highly available IP storage network can be done in a variety of ways. Implementing SAN/iQ storage clusters along with standard networking and iSCSI technology allow a customer to create Enterprise-class storage systems at a reasonable price point. Given the options available to the customer, LeftHand Networks does recommend certain configurations for ease-of-deployment and functionality.

**General Recommendations**

- Implement a managed Enterprise-class switch infrastructure (see following section and Table 1).
- Implement Adaptive Load Balancing (ALB) NIC bonding on the storage node for GigE networks. For storage nodes with 10Gb adapters, implement active/passive bonding between the 10Gb NIC and a single GigE NIC.
- For Microsoft Windows Server environments, implement MPIO along with the LeftHand Networks DSM for NIC fault-tolerance and superior performance.
- For other operating systems, where supported, implement NIC bonding in the host software for NIC fault-tolerance and performance.

**Specific Network Recommendations**

- Implement a separate subnet or VLAN for the IP storage network for dedicated bandwidth.
• Implement a fault-tolerant switch environment as a separate VLAN through a core switch infrastructure or multiple redundant switches.

• Set the individual Gigabit ports connected to the storage nodes and servers at 1000-full-duplex at both the switch and host/node port level.

• Implement switches with full-duplex non-blocking mesh backplanes with sufficient port buffer cache (at least 512KB per port).

• Implement Flow Control on the storage network switch infrastructure.

• Connect an IP route from the IP storage network to the rest of the infrastructure so that the storage modules can be managed and alerts can be sent.

• (Optional) Implement Jumbo Frames support on the switch, storage nodes and all servers connected to the IP-SAN.

**Recommended Switch Infrastructure for an IP-SAN**

LeftHand Networks does not recommend any particular switch for use with SAN/iQ. However, there are a set of minimum switch capabilities that make building a high performance fault-tolerant storage network a relatively easy and cost effective task. As a rule of thumb, any Enterprise-class managed switch typically has the necessary capabilities most customers require. Table 1 summarizes the minimum switch capabilities LeftHand Networks recommends.
Switch Capability | Description
--- | ---
Gigabit Ethernet Support | Each storage node comes equipped with at least two copper gigabit Ethernet ports (802.3ab). To take advantage of full duplex gigabit capabilities the cabling infrastructure must be Cat5e or Cat6 cabling to the storage nodes. Servers connections and switch interconnects can also be done via fiber cabling, in addition to Cat5e or Cat6 cabling.
Fully Subscribed non-Blocking Backplanes | In order to achieve maximum performance on the IP-SAN it is important to select a switch that has a fully subscribed backplane. This means that the backplane is capable of supporting all ports at full-duplex mode. For instance, if the switch has 24 gigabit ports, it will have to have a 48 gigabit backplane to support full duplex gigabit communications.
Adequate per-port Buffer Cache | For optimal switch performance, LeftHand Networks recommends that the switch have at least 512KB of buffer cache per port. Consult your switch manufacturer specifications for the total buffer cache. For example, if the switch has 48 gigabit ports, the recommendation is to have at least 24MB of buffer cache dedicated to those ports. If the switch aggregates cache among a group of ports (ie 1MB of cache per 8 ports) space your storage modules and servers appropriately to avoid cache oversubscription.
Flow Control Support | IP storage networks are unique in the amount of sustained bandwidth that is required to maintain adequate performance levels under heavy workloads. Gigabit Ethernet Flow Control (802.3x) technology should be enabled on the switch to eliminate receive and/or transmit buffer cache pressure. The storage nodes should also be set to have flow control enabled. Note: some switch manufacturers do not recommend configuring Flow Control when using Jumbo Frames, or vice versa. Consult the switch manufacturer documentation for guidance on this issue. LeftHand Networks recommends implementing Flow Control over Jumbo Frames for optimal performance.
| Flow control is required when using DSM/MPIO.

Table 1   Minimum Recommended Switch Capabilities
Individual Port Speed and Duplex Setting

LeftHand Networks recommends that all ports on the switch, servers and storage nodes be configured to auto-negotiate duplex and speed settings. Although most switches and NICs will auto negotiate the optimal performance setting, if a single port on the IP storage network negotiates a sub-optimal (100mbit or less and/or half-duplex) setting, the entire SAN performance can be impacted negatively. Make sure to check each switch and NIC port to make sure the auto-negotiation resolved to be 1000Mb/s with full-duplex.

Link Aggregation / Trunking Support

Link Aggregation and/or Trunking support is important to enable when building a high performance fault-tolerant IP storage network. LeftHand Networks recommends implementing Link Aggregation and/or Trunking technology when doing Switch to Switch Trunking, Server NIC Load Balancing and Server NIC Link Aggregation (802.3ad). If using a Cisco infrastructure, Port Aggregation Protocol (PagP) for Etherchannel is not supported between switches and storage modules.

VLAN Support

LeftHand Networks recommends implementing a separate subnet or VLAN for the IP storage network. If implementing VLAN technology within the switch infrastructure, typically one will need to enable VLAN Tagging (802.1q) and/or VLAN Trunking (802.1q or InterSwitch Link [ISL] from Cisco). Consult your switch manufacturer configuration guidelines when enabling VLAN support.

<table>
<thead>
<tr>
<th>Switch Capability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Port Speed and Duplex Setting</td>
<td>LeftHand Networks recommends that all ports on the switch, servers and storage nodes be configured to auto-negotiate duplex and speed settings. Although most switches and NICs will auto negotiate the optimal performance setting, if a single port on the IP storage network negotiates a sub-optimal (100mbit or less and/or half-duplex) setting, the entire SAN performance can be impacted negatively. Make sure to check each switch and NIC port to make sure the auto-negotiation resolved to be 1000Mb/s with full-duplex.</td>
</tr>
<tr>
<td>Link Aggregation / Trunking Support</td>
<td>Link Aggregation and/or Trunking support is important to enable when building a high performance fault-tolerant IP storage network. LeftHand Networks recommends implementing Link Aggregation and/or Trunking technology when doing Switch to Switch Trunking, Server NIC Load Balancing and Server NIC Link Aggregation (802.3ad). If using a Cisco infrastructure, Port Aggregation Protocol (PagP) for Etherchannel is not supported between switches and storage modules.</td>
</tr>
<tr>
<td>VLAN Support</td>
<td>LeftHand Networks recommends implementing a separate subnet or VLAN for the IP storage network. If implementing VLAN technology within the switch infrastructure, typically one will need to enable VLAN Tagging (802.1q) and/or VLAN Trunking (802.1q or InterSwitch Link [ISL] from Cisco). Consult your switch manufacturer configuration guidelines when enabling VLAN support.</td>
</tr>
</tbody>
</table>
Basic IP Routing
The storage nodes support accessing external services like DNS, SMTP, SNMP, Syslog, etc. In order to get this traffic out of the IP storage network, an IP route will have to exist from the IP storage network to the LAN environment. Also, if the storage nodes are going to be managed from a remote network, there must exist an IP route to the storage nodes from the management station. Finally, if Remote Copy is going to be used, the Remote Copy traffic must be routable end-to-end between the Primary and Remote sites.

Spanning Tree / Rapid Spanning Tree
In order to build a fault-tolerant IP storage network multiple switches are typically connected into a single Layer 2 (OSI Model) broadcast domain using multiple interconnects. In order to avoid Layer 2 loops, the Spanning Tree protocol (802.1D) or Rapid Spanning Tree protocol (802.1w) must be implemented in the switch infrastructure. Failing to do so can cause numerous issues on the IP storage networks including performance degradation or even traffic storms.
LeftHand Networks recommends implementing Rapid Spanning Tree if the switch infrastructure supports it for faster Spanning Tree convergence.
If the switch is capable, consider disabling spanning tree on the storage node and server switch ports so that they do not participate in the spanning tree convergence protocol timings.

Jumbo Frames Support
Large sequential read and write workloads can benefit approximately 10-15% from a larger Ethernet frame. The storage nodes are capable of frame sizes up to 9KB. Jumbo frames must be enabled on the switch, storage nodes and all servers connected to the IP-SAN. Typically, jumbo frames are enabled globally on the switch or per VLAN and on a per port basis on the server. Jumbo frames are enabled individually on each gigabit NIC in the storage node.
Note: Some switch manufacturers do not recommend configuring Jumbo Frames when using Flow Control, or vice versa. Consult the switch manufacturer documentation for guidance on this issue. LeftHand Networks recommends implementing Flow Control over Jumbo Frames for optimal performance.

<table>
<thead>
<tr>
<th>Switch Capability</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic IP Routing</td>
<td>The storage nodes support accessing external services like DNS, SMTP, SNMP, Syslog, etc. In order to get this traffic out of the IP storage network, an IP route will have to exist from the IP storage network to the LAN environment. Also, if the storage nodes are going to be managed from a remote network, there must exist an IP route to the storage nodes from the management station. Finally, if Remote Copy is going to be used, the Remote Copy traffic must be routable end-to-end between the Primary and Remote sites.</td>
</tr>
<tr>
<td>Spanning Tree / Rapid Spanning Tree</td>
<td>In order to build a fault-tolerant IP storage network multiple switches are typically connected into a single Layer 2 (OSI Model) broadcast domain using multiple interconnects. In order to avoid Layer 2 loops, the Spanning Tree protocol (802.1D) or Rapid Spanning Tree protocol (802.1w) must be implemented in the switch infrastructure. Failing to do so can cause numerous issues on the IP storage networks including performance degradation or even traffic storms. LeftHand Networks recommends implementing Rapid Spanning Tree if the switch infrastructure supports it for faster Spanning Tree convergence. If the switch is capable, consider disabling spanning tree on the storage node and server switch ports so that they do not participate in the spanning tree convergence protocol timings.</td>
</tr>
<tr>
<td>Jumbo Frames Support</td>
<td>Large sequential read and write workloads can benefit approximately 10-15% from a larger Ethernet frame. The storage nodes are capable of frame sizes up to 9KB. Jumbo frames must be enabled on the switch, storage nodes and all servers connected to the IP-SAN. Typically, jumbo frames are enabled globally on the switch or per VLAN and on a per port basis on the server. Jumbo frames are enabled individually on each gigabit NIC in the storage node. Note: Some switch manufacturers do not recommend configuring Jumbo Frames when using Flow Control, or vice versa. Consult the switch manufacturer documentation for guidance on this issue. LeftHand Networks recommends implementing Flow Control over Jumbo Frames for optimal performance.</td>
</tr>
</tbody>
</table>
Storage Module Connect Options

LeftHand Networks storage modules come equipped with at least two Gigabit Ethernet network interfaces. When building high performance fault-tolerant IP storage networks, LeftHand Networks recommends implementing NIC bonding on every storage module connected to the network. SAN/iQ powered storage modules on the market today support a variety of network interface bonding techniques. Choose the appropriate NIC bonding solution for your environment. SAN/iQ supports each storage node represented as a single IP address on the network. SAN/iQ does not currently support storage nodes with multiple IP connections to the same network. Table 2 highlights the SAN/iQ NIC bonding types supported and associated recommended usages.

<table>
<thead>
<tr>
<th>NIC Bonding Support</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive Load Balancing (Recommended)</td>
<td>Adaptive Load Balancing (ALB) is the most flexible NIC bonding technique that can be enabled on the storage nodes and provides for increased bandwidth and fault-tolerance. There is typically no special switch configuration required to implement ALB. Both NICs in the storage nodes are made active, and they can be connected to different switches for active-active port failover. ALB operates at 2 gigabits of aggregated bandwidth. Adaptive Load Balancing is only supported for NICs with the same speeds – i.e., two GigE NICs. ALB is not supported between a 10Gb and a GigE NIC.</td>
</tr>
<tr>
<td>Active / Passive</td>
<td>Active / Passive (aka Active / Backup) NIC bonding is the most simple NIC bonding technique that can be enabled on the storage nodes and provides for fault-tolerance only. There is typically no special switch configuration required to implement Active / Passive bonding. Only a single NIC is made active, while the other NIC is made passive. Therefore, an Active / Passive bond operates at 1 gigabit bandwidth.</td>
</tr>
<tr>
<td>Link Aggregation - 802.3ad</td>
<td>Link Aggregation (LACP / 802.3ad) NIC bonding is the most complex NIC bonding technique that can be enabled on the storage nodes and provides for link aggregation only. Link Aggregation bonds must typically be built on both the storage node and switch as port pairs. Both NICs in the storage nodes are made active, however they can only be connected to a single switch. A Link Aggregation bond operates at 2 gigabits of aggregated bandwidth. 802.3ad is only supported for NICs with the same speeds – i.e., two GigE NICs. 802.3ad is not supported between a 10Gb and a GigE NIC.</td>
</tr>
</tbody>
</table>

Table 2 SAN/iQ NIC Bonding Support
Other recommendations when implementing NIC bonding on the storage modules:

1. Implement the same NIC bonding type on all storage modules in the Cluster.

2. Implement NIC bonding before joining the storage module to the Management Group.

Sample Configurations

Exhibit 1 Recommended IP Storage Network Configuration for Windows environments: Dual Redundant Switches w/ ALB and MPIO.
Exhibit 2  Recommended IP Storage Network Configuration for non-Windows environments: Dual Redundant Switches w/ ALB and MPIO.

Key configuration notes:

Switch Infrastructure: Dual Redundant Gigabit (10Gb when applicable) switches trunked together for bandwidth and fault-tolerance.

Storage Node connectivity: Adaptive Load Balancing NIC bond with a single port connected to each switch.

Host Server connectivity: Dual NICs connected to the IP storage network with a single port connected to each switch. For Windows 2003/2008, use the DSM for MPIO for multiple NIC support (Exhibit 1). Otherwise, use ALB NIC bonding on the host server (Exhibit 2) where supported.

**Note: Flow control is required when using DSM for MPIO. Without flow control, performance may be negatively impacted.
Sample IP Storage Network Configuration Setup

Regardless of switch manufacturer, the typical recipe for configuration is very similar across switch manufacturers. Consult the switch manufacturer configuration reference guides for specific configuration tasks. As a general rule, the following are the configuration tasks that LeftHand Networks recommends to get your IP storage network implemented:

1. Cable the switches with multiple inter-connect cables for redundancy and performance. If the switches do not auto create the trunk between the switches, configure the trunking protocols.

2. Enable Flow Control support on the switches. This is must be done globally on the switch or Flow Control will not be negotiated on the network. Verify flow control on the servers and storage nodes.

3. (Optional) Enable Jumbo Frames support on the switches. This is must be done globally on the switch or per VLAN, otherwise jumbo frames will not be negotiated on the network.

4. Setup the appropriate VLAN configurations as necessary.

5. Enable Rapid Spanning Tree for the VLAN configured in Step #4.

6. (Optional) Enable rapid spanning-tree convergence or portfast (Cisco) on the storage node and server switch ports.

7. Configure the IP routes necessary to get traffic routed out of the IP storage network into the LAN.

8. Connect up all storage nodes and servers to the switches with Cat5e or Cat6 cabling.

9. Set all gigabit ports on the switches, servers and storage nodes to auto negotiate duplex and speed. Verify that all ports negotiated to 1000-full duplex.

10. (Optional) Set all ports on the servers and storage nodes to use jumbo frames. Use a 9KB (9000Byte) frame size.

11. Setup ALB NIC bonding on all the storage nodes. Assign static IP addresses to the bond interfaces.

12. Setup the NICs on the host servers. For Windows, setup individual static IP addresses on each NIC. For other operating systems, use the native NIC bonding configuration.

13. Verify the host servers can communicate with the storage nodes.
Proceed with the rest of the SAN configuration.

**Multi-site SAN Networking Requirements**

It is important to pay special attention to the networking configuration when building a SAN/iQ Multo-Site SAN. The two primary factors that contribute to the health of the Multi-Site SAN are network latency and network bandwidth.

**Network Latency**

High network latency can be the primary cause of slow I/O performance, or worse iSCSI drive disconnects. It is important to keep network latency on your Multi-Site SAN subnet below 2 milliseconds. Many factors can contribute to increasing network latency, but the two most common factors are distance between storage cluster modules and router hops between storage cluster modules. Configuring the Multi-Site SAN on a single IP subnet with layer-2 switching will help to lower the network latency between storage cluster modules.

Attenuation is a general term that refers to any reduction in the strength of a signal. Sometimes called loss, attenuation is a natural consequence of signal transmission over long distances. As the distance between storage cluster modules increases, attenuation will increase network latency because data packets will not pass checksum and must be resent. Networking vendors do offer various products to reduce the impact of distance, but most Multi-Site SAN configurations will be built based on the network switch vendor distance recommendations.

**Network Bandwidth**

Network bandwidth required for a Multi-Site SAN depends on the server applications, maintenance utilities and backup/recovery processes. Most I/O intensive applications, like Microsoft Exchange and SQL Server, will not consume much network bandwidth and are more sensitive to network latency issues. Bandwidth becomes much more important when you are performing maintenance operations, like ESEUtil.exe for Exchange, or backup/recovery. Any sequential read/write stream from the Multi-Site SAN could consume significant bandwidth.
**Note:** Storage data transfer rates are typically measured in BYTES while network data transfer rates are measured in BITS. A 1 Gb/sec (lower case “b” means bits) network connection can transfer a maximum of 120-130 MB/sec (upper case “B” means bytes).

Microsoft Windows provides performance monitor counters that can help to determine the data-path bandwidth requirements. **Disk Bytes/sec** is the rate bytes are transferred to or from the disk during write or read operations. M requirements after the data on that disk has been transferred to the Multi-Site SAN. The bandwidth requirements for every volume on the Multi-Site SAN must be accounted for in the total bandwidth calculation.

Allocate 50 MB/sec bandwidth per storage module pair, and continue to allocate 50 MB/sec (400 Mb/sec) to each additional storage module pair as the storage cluster grows. Monitoring this value will provide some insight into the future bandwidth.

---

**Suggested Best Practices for Multi-site SAN**

**Use a single IP subnet with Layer-2 switching** – Layer-2 switching will help to reduce the network latency introduced by a traditional Layer-3 router.

**Configure redundant network paths** – configure the network topology to eliminate all single points of failure between storage modules on the Multi-Site SAN.

**Configure for fast network convergence** – make sure your Multi-Site SAN network is configured so the network will converge quickly after a component or path failure. Consider rapid spanning tree for layer-2 switching with the goal to achieve network convergence in less than 15 seconds.

**Monitor network latency and bandwidth** – consistently monitor network latency and bandwidth on the Multi-Site SAN network. Make sure to plan properly before adding additional storage capacity or data volumes to the Multi-Site SAN storage cluster.

**Run data backups against a Remote Copy of the data** – SAN/iQ Remote Copy can be used to create copies of your production data on a storage cluster that is not stretched across the Multi-Site SAN. Data backups and data mining can be done from this isolated secondary copy. This will reduce the impact of these operations on network bandwidth and latency.
Automated data failover (Recommended) – configure an equal number of SAN/iQ Managers on both sides of the Multi-Site SAN and add a Failover Manager at a third location. If an event disables any of the three locations, quorum will still exist with the managers running at the two remaining locations and automatic failover will occur and the volumes will stay online.

Manual data failover – If a third location is not available to run the Failover Manager, configure an equal number of SAN/iQ Managers on both sides of the Multi-Site SAN and add a SAN/iQ virtual manager. If an event disables one side of the Multi-Site SAN the second site can be brought online quickly by starting the virtual manager.

Sample Switch Configurations and Commands

Note: Consult the switch vendor specific documentation for the most up-to-date command syntax.

Table 3 Cisco 3750

<table>
<thead>
<tr>
<th>Capability</th>
<th>Sample Configuration Commands</th>
</tr>
</thead>
</table>
| Port Trunking                  | Configure an interface as a trunk port for VLANs 1 and 100 and auto negotiate the trunk protocol:  
|                                | Cisco(config-if)# switchport trunk allowed vlan 1,100 encapsulation negotiate |
| Enable Jumbo Frames            | To setup the switch to support 9K jumbo frames:  
|                                | Cisco(config)# system mtu jumbo 9000 |
| Enable Flow Control            | Use the flow control interface command to enable flow control on the SAN interfaces:  
|                                | Cisco(config-if)# flowcontrol receive {desired | off | on} |
**Note: Make sure to save the configuration when done.**

<table>
<thead>
<tr>
<th>Capability</th>
<th>Sample Configuration Commands</th>
</tr>
</thead>
</table>
| Setup VLAN                        | Configure a storage network VLAN (100)  
Cisco(config)# interface vlan 100  
Cisco(config-if)# ip address x.x.x.x 255.255.255.0  
Assign the physical gigabit Ethernet interface(s) to the VLAN:  
Cisco(config-if)# switchport access vlan 100 |
| Enable Spanning Tree / Rapid Spanning Tree | To enable spanning tree globally on the switch use the following command:  
Cisco(config)# spanning-tree mode {mst | pvst | rapid-pvst} |
| Enable Portfast                   | Enable portfast mode on the server and storage node switchport interfaces only.  
Cisco(config-if)# spanning-tree portfast |
| Configure IP Routing              | Setup a default IP route (assumes basic ip support on the switch):  
Cisco(config)# ip default-gateway X.X.X.X  
Or, use a static default route:  
Cisco(config)# ip routing  
Cisco(config)# ip route 0.0.0.0 0.0.0.0 X.X.X.X |

### Table 3 - Cisco 3750

**Note:**
<table>
<thead>
<tr>
<th>Capability</th>
<th>Sample Configuration Commands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Port Trunking</td>
<td>To configure ports C1 and C2 into a trunk use the following command:</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) int c1-c2 lacp active</td>
</tr>
<tr>
<td>Setup VLAN</td>
<td>To configure a VLAN (100) for the storage network a command similar to the following:</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) int vlan 100 ip address 10.1.1.254 255.255.255.0</td>
</tr>
<tr>
<td></td>
<td>To specify ports C1 thru C8 to be part of VLAN 100 use the following command:</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) int 100 tagged C1-C8</td>
</tr>
<tr>
<td>Enable Jumbo Frames</td>
<td>To configure Jumbo Frames on the storage network VLAN (100):</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) vlan 100 jumbo</td>
</tr>
<tr>
<td>Enable Flow Control</td>
<td>First, you must enable flow control globally on the switch as follows:</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) flow-control</td>
</tr>
<tr>
<td></td>
<td>To configure flow control on ports C1 thru C8 use the following command:</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) int c1-c8 flow-control</td>
</tr>
<tr>
<td>Enable Spanning Tree / Rapid</td>
<td>To enable rapid spanning tree (RSTP) use the following command:</td>
</tr>
<tr>
<td>Spanning Tree</td>
<td>ProCurve(config) spanning-tree protocol-version rstp</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) spanning-tree</td>
</tr>
<tr>
<td></td>
<td>To enable spanning tree (STP) use the following command:</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) spanning-tree protocol-version stp</td>
</tr>
<tr>
<td>Enable Spanning Tree Fast Mode</td>
<td>Enable Spanning Tree Fast Mode support on server and storage node switchport interfaces C1 thru C8:</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) spanning-tree c1-c8 mode fast</td>
</tr>
<tr>
<td>Configure IP Routing</td>
<td>To configure a default gateway for the switch use the following command</td>
</tr>
<tr>
<td></td>
<td>ProCurve(config) ip default-gateway X.X.X.X</td>
</tr>
</tbody>
</table>

**Note: Make sure to save the configuration when done.**
10 GbE Best Practices

CX4 Cables

It is important to make sure that the right cables are used for the 10Gb connection. At the time this document was written, the NSMs were using CX4 10Gb cards which required CX4 specific cables. Make sure that the cables being used are the CX4 Ethernet cables. Infiniband cables look very similar to CX4 cables and can physically connect to a CX4 network adapter, but they will not meet the specifications of CX4.

Switches

Make sure the switches are non-blocking switches that allow 10Gb/s bidirectionally on every port. If the switch does not offer this level of performance, you may see unexpected performance from your 10Gb SAN.

Flow Control

Flow control can have a dramatic impact on performance in a 10Gb environment. This is especially true with a mixed GigE and 10GigE environment.

When a network port becomes saturated, excess frames can be dropped because the port can't physically handle the amount of traffic it is receiving. This causes the packets to be re-sent, and the overhead of re-sending the packets can cause a performance decrease. An example of this is a 10GbE link sending data at 10 Gb/s to a single GigE link. Flow control eliminates this problem by controlling the speed at which data is sent to the port. For this reason, best practices dictate that flow control is always enabled.

Flow control must be enabled on both the switches and NICs/iSCSI initiators for it to function properly. If it is not enabled everywhere, the network defaults to the lowest common denominator, which would be to have flow control disabled.

Bonding

It is also possible to bond 10Gb NICs. At the time of this document’s creation, the only supported 10Gb NIC for NSMs is a dual-port 10Gb CX4 card. While it is possible to bond the two ports on that card to each other, that does not protect against a fault on the card itself. Therefore, the best practices dictate that a active-passive bond be created with one port of the 10GbE card and one 1GbE port on the NSM. The 10GbE will be the preferred interface for this bond.
Bonding of three or four interfaces (such as all the NICs on the NSM) is not allowed. The bonding of bonds is also not allowed (for example, creating two active-passive bonds as described above, then bonding those together with ALB).

The active-passive bond described above provides protection against a single NIC hardware failure. There is no performance benefit to bonding both 10GbE ports on the 10GbE card as traffic does not get high enough to saturate even a single 10GbE port.

Server Side 10 GbE NICs

Use your server vendor’s recommendations for maximum compatibility and support. LeftHand does not recommend or endorse any particular vendor or 10GbE card for the servers.

Multi-Site SAN

In a Multi-Site SAN, each half of the NSMs in a cluster (or each half of a SAN) lives in a different physical space (whether a separate rack, floor, building, etc.). The link speed between the two locations must be adequate for 10GbE benefits to be noticed. Specifically:

- The link speed between the two locations may need to be increased to take advantage of the performance benefit of the 10Gb NICs (if this a new installation, the link speed should be 200 MB/s per storage node)
- The latency between the two locations must be 2ms or less

If these conditions are not met, the benefits of 10GbE will be minimized.

Contacting Support

Contact LeftHand Networks Support if you have any questions on the above information:

North America:
Basic Contract Customers
1.866.LEFT-NET (1.866.533.8638)
303.217.9010
http://support.lefthandnetworks.com

EMEA:
All Customers
00.800.5338.4263 (Int’l Toll-Free number)
+1.303.625.2647 (US number)
http://support.lefthandnetworks.com

Premium Contract Customers
1.888.GO-SANIQ (1.888.467.2647)
303.625.2647
http://support.lefthandnetworks.com