Table of Contents

Cluster Networking Introduction ..........................................................3
  Assumptions ..........................................................................................3
  Cluster Networking Features ............................................................3
Networking Guidelines .............................................................................8
  Internal Network ..................................................................................8
  External Network .................................................................................8
Isilon Supported Network Topologies .....................................................8
  Standard Internal and External Configuration ....................................8
  High Availability Internal Network ....................................................9
  Dual-Subnet External Network ............................................................9
  10 Gigabit Networking ......................................................................10
Isilon Configuration Details .....................................................................11
  FlexNet Configuration .........................................................................11
  Using Link Aggregation ......................................................................14
Appendix A – Isilon Recommended Switches .........................................15
  Ethernet Switches ..............................................................................15
  Infiniband Switches ..........................................................................15
1. Cluster Networking Introduction

This configuration guide details the basic networking concepts and considerations when deploying an Isilon cluster. An Isilon cluster supports a wide variety of network topologies and configuration options to support large, complex and dynamic networking requirements.

Assumptions

It is assumed the reader:

- Has a basic understanding of network-attach storage (NAS) designs and implementations.
- Has a working knowledge of network topologies and implementations.
- Has a basic understanding of Isilon hardware and OneFS® operating system and applications.

Cluster Networking Features

This section introduces the high-level networking concepts relevant to an Isilon cluster. A cluster uses two types of network connectivity, internal and external, which are explained below.

Figure 1 depicts how an Isilon cluster connects to its internal and external networks.

![Figure 1 - Isilon Cluster Network Connectivity](image)

Internal Interconnect

An Isilon cluster uses a switched Infiniband (IB) fabric for the internal interconnect (also referred to as the back-end network), on which intra-cluster data and messages are transferred. The use of Infiniband delivers the following advantages:
• Extremely low latency between cluster nodes. Using a switched star topology, each node in the cluster is one IB hop away from any other node.
• Excellent performance, which is increased in X-Series hardware with the implementation of double data rate (DDR), which doubles the IB bandwidth to 20 gigabits per second.
• Performance is also maximized through the low-latency protocols used for intra-cluster communication.

The minimum requirements for the intra-cluster network are:

• Use of an Isilon recommended Infiniband switch (see Appendix A for Isilon recommended IB switches)
• Infiniband cables with CX4 copper connectors

External Interconnect
An Isilon cluster communicates with applications and users over an external network, also called the front-end network. Normally, each node on the cluster has at least one connection to the network, although in high-throughput applications, all clients may connect through an Accelerator-x™. Thus, in a typical configuration, the total bandwidth of a cluster is the aggregation of the bandwidth of the link(s) on each node. The network performance of each node depends primarily on the data access pattern and the protocol (usually NFS or CIFS) used for network file access.

Isilon storage nodes, e.g. IQ3000x, each have two Gigabit Ethernet (GigE) ports for external network connectivity. The minimum requirements for GigE connectivity are:

• Use of an Isilon recommended Ethernet switch (see Appendix A for Isilon recommended Ethernet switches)
• CAT-5e or CAT-6 cables with RJ-45 copper connectors

Figure 2 illustrates the locations of internal and external network ports on Isilon storage nodes.
The Isilon Accelerator-x has two GigE ports, and also two 10GigE ports for applications requiring high-throughput, such as uncompressed HD ingest, editing and playback. The minimum requirements for 10GigE connectivity are:

- Isilon recommended 10GigE Ethernet switch (see Appendix A for Isilon recommended Ethernet switches)
- 10GigE cables with CX4 copper or SR optical connectors

Figure 3 illustrates the locations of internal and external network ports on the Isilon Accelerator-x.

![Figure 3 – Isilon Accelerator-x Network Ports](image)

**Isilon Cluster Node Port Summary**

Table 1 lists the internal and external ports on Isilon cluster nodes.

<table>
<thead>
<tr>
<th>Port</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal</strong></td>
<td></td>
</tr>
<tr>
<td>Internal-A</td>
<td>Primary internal interconnect port. This port must be configured.</td>
</tr>
<tr>
<td>Internal-B</td>
<td>Secondary internal interconnect port. The use of this port results in a redundant back-end switch configuration for increased availability.</td>
</tr>
<tr>
<td><strong>External</strong></td>
<td></td>
</tr>
<tr>
<td>External-1</td>
<td>One of two GigE ports for client connectivity</td>
</tr>
<tr>
<td>External-2</td>
<td>Second of two GigE ports for client connectivity</td>
</tr>
<tr>
<td>10GigE-1 (Accelerator-x only)</td>
<td>One of two 10GigE ports for client connectivity</td>
</tr>
<tr>
<td>10GigE-2 (Accelerator-x only)</td>
<td>Second of two 10GigE ports for client connectivity</td>
</tr>
</tbody>
</table>

**Table 1 - Cluster Node Network Ports**

*Note: Isilon storage nodes also have three (3) Serial Attach SCSI (SAS) ports, which are similar in appearance to IB ports. These are clearly labeled and care should be taken to connect the correct port(s).*
FlexNet 2.0

FlexNet™ is the OneFS subsystem used for configuring and managing network interfaces. With the introduction of OneFS 5.0, major improvements have been made to FlexNet, now in version 2.0.

Internal networking configuration and management has not changed from previous releases in OneFS 5.0.

FlexNet 1.0 was designed to support a few simple network topologies, and has a single management object, known as a profile. Profiles have limited flexibility, and can contain a single interface. An interface bound to a profile cannot exist in another profile.

FlexNet 1.0 was loosely coupled with SmartConnect™, with different configuration processes and operations.

FlexNet 2.0 has been designed to support complex and variable network topologies. It has several hierarchical and overlapping management objects which allow for extremely flexible configurations, simply defined and managed.

FlexNet 2.0 is tightly integrated with SmartConnect to provide improved flexibility and stability, and easy management.

The following FlexNet 2.0 terms are important to understanding its operation:

- **Pool** – Also referred to as an IP Address Pool, contains one or more interfaces (e.g. External-1) and a set of IP addresses to be assigned to them. SmartConnect settings, such as the zone name, and whether IPs are allocated statically or dynamically, are also configured at the Pool level. A Pool can be thought of as a layer-2 object.
- **Subnet** – Specifies a network subnet, netmask and other parameters related to layer-3 networking. VLAN tagging is configured here. A subnet contains one or more pool objects, which relates a set of IP addresses, bound interfaces, and allocation policies to the network.
- **Provisioning Rule** – Specifies actions to take when a node is added to the cluster, based on the node type and interface. For example, a rule could state that when a node of type storage is added, External-1 and External-2 are assigned to two different pools, which in turn belong to two separate subnets.

Figure 4 illustrates the relationship among Pools, Subnets, and Provisioning Rules.
In the figure, two pools, Pool0 and Pool1, have been created inside of Subnet0. Pool0 contains the 10GigE-1 interface from an Accelerator-x, and is used for HD editing. Pool1 contains the Ext-1 Gigabit interfaces from the storage nodes, and is used for SD editing.

A third pool, Pool2, has been created with the Ext-2 interfaces on the nodes and placed in a separate subnet, Subnet1. This subnet and IP address pool are used by IT to manage the cluster.

**FlexNet 2.0 and SmartConnect**

As noted, FlexNet 2.0 is tightly coupled with SmartConnect, the OneFS client load-balancing and failover application. SmartConnect options, which were previously configured separately from FlexNet, are now configured as part of FlexNet.

At the subnet level, the SmartConnect Service IP Address, formerly known as the Virtual IP (VIP), is specified. This is the IP address used primarily by a DNS server to forward SmartConnect zone lookups to the cluster.

SmartConnect options having to do with zone name, load balancing and failover are set at the pool level. Different pools inside the same subnet can have different configurations for different use cases.
2. Networking Guidelines

Internal Network
When an Isilon-recommended IB switch is used, no configuration of the internal network is required, aside from assigning IP addresses to the network during the initial cluster setup.

Please see Appendix A for a list of recommended IB switches.

External Network
Consider the following requirements when implementing external network connectivity for an Isilon cluster.

Ethernet Switch Manufacturer and Model
Use an Isilon-recommended Ethernet switch. Please see Appendix A for a list of recommended switches.

Non-blocking Switch Fabric
A switch used for external connectivity must use a non-blocking switch fabric. A switch is said to be non-blocking if the internal switching fabric is capable of handling the theoretical total of all ports, such that any routing request to any free output port can be established successfully without interfering with other traffic. All switches recommended by Isilon have this feature.

Switch Port Buffer Size
The switch must have a port buffer size of at least 1MB. At load, smaller port buffer sizes will fill up, resulting in dropped packets that must be retransmitted, negatively impacting performance.

Jumbo Frame Support
This is not a requirement, but for best performance with most applications, jumbo frames should be enabled on the external network. "Jumbo frames" refers to a Maximum Transmission Unit (MTU) size of 9000 bytes, compared to a standard MTU of 1500 bytes. Jumbo frames allow more data to be transferred between network endpoints with fewer operations, which will increase throughput for most applications.

3. Isilon Supported Network Topologies
The process for connecting an Isilon cluster to its internal and external networks is straightforward and simple. A cluster supports several different topologies, which differ primarily in their level of redundancy.

Standard Internal and External Configuration
The majority of Isilon clusters are deployed using single Ethernet and IB switches, which provides reasonable performance and availability. Figure 5 shows this type of configuration.
In this configuration, one external GigE port (usually External-1) is connected to the client network, and one internal IB port (must be Internal-A) is connected to the internal cluster network.

Configuring a cluster in this manner requires no additional steps beyond those taken when the cluster is first created. See the FlexNet Configuration section for more details.

**High Availability Internal Network**

An Isilon cluster can be configured to use redundant IB switches for internal interconnect. Adding a second IB switch does not improve performance (which is unnecessary in any case); rather, it provides a failover switch in case of hardware or cabling issues. If one node fails to the second IB switch, all nodes in the cluster will do so as well.

When configuring a redundant internal switch, the IP addresses used must be on a different subnet than those used on the primary switch.

Figure 6 shows a redundant internal network configuration.

**Dual-Subnet External Network**

It is also possible to deploy a dual-subnet external network. In this topology, each port (GigE or 10GigE) is connected to a separate Ethernet switch, each on a different subnet. This can provide access to clients on separate subnets, and when used with multi-homed hosts, this topology prevents data inaccessibility due to switch, interface or cabling issues. Note that this topology is not a failover or high-availability design. Multi-homed clients connected to a subnet on which
there is a failure will lose their connections, and must re-establish connections on the second subnet.

Figure 7 illustrates a dual-subnet external network configuration (with a redundant internal network configuration).

**Figure 7 – Dual-Subnet External Network**

**10 Gigabit Networking**

The Isilon Accelerator-x supports two 10GigE ports using either CX4 copper or SR optical connectors. 10GigE ports are configured in an identical manner to GigE ports. Figure 8 shows a common network topology used with Accelerator-x nodes.

**Figure 8 - 10GigE Networking with Accelerator-x**
4. Isilon Configuration Details

This section provides details and steps for configuring internal and external networking on an Isilon cluster.

**FlexNet Configuration**

**Internal network**

Configuration of the internal interconnect is straightforward. Set up of the primary internal interface (Internal-A) for each node is a required step in the initial configuration of the cluster. It's also possible to configure the Internal-B failover interface at this time, or the failover interface can be added at a later time.

For the primary internal network, the following parameters must be provided:

- Netmask value
- IP address range

MTU does not need to be configured for IB networks.

Internal-A can be edited using the WebUI or command-line interface (CLI). Figure 9 shows the WebUI page for editing Internal-A.

![Figure 9 - Edit Primary Internal Interface](image)

When configuring a redundant internal network using the Internal-B interface, the steps are the same as for the primary interface, with the following differences:
• IP address range must be on a different subnet from Internal-A
• A third IP address range, on a separate subnet from both Internal-A and Internal-B, must be specified. This is the subnet used for the Failover interface, which is part of the IP failover mechanism.

The Internal-B/Failover interfaces can also be edited using the WebUI, as illustrated in Figure 10.

**Networking > Edit Internal Interface**

**Interface int-b / Failover**

**Settings**

State:  
- Disabled
- Enabled

Before changes to the interface take effect, the cluster must be rebooted.

Netmask:*  

**IP Ranges**

**Interface int-b**  

Adding or deleting IP ranges is effective immediately and does not require a cluster reboot.

10.11.10.1 - 10.11.10.9

**Interface Failover**  

Adding or deleting IP ranges is effective immediately and does not require a cluster reboot.

10.12.10.1 - 10.12.10.9

**Figure 10 - Edit Failover Internal Interface**

**External Network**

FlexNet 2.0 introduces a new process for configuring external networking, and as noted above, a new architecture and terminology.

When initially configuring a cluster, the first external network interface (External-1, typically) is setup as part of the configuration process. In order for this process to complete successfully, the following information is required:

- Netmask
- IP address range
- Default gateway
- Domain name server list (optional)
- DNS search list (optional)
- SmartConnect zone name (optional)
- SmartConnect service address (optional)

When this information is provided, the following actions occur:
• A default external subnet is created, named subnet0, with the netmask and optional SmartConnect service address.
• A default IP address pool is created, named pool0, with the specified IP address range, the gateway, the optional SmartConnect zone name, and the initial external interface with the first node in the cluster as the only member.
• A default network provisioning rule is created, named rule0, which automatically assigns the first external interface for all newly added nodes to pool0.
• pool0 is added to subnet0 and configured to use subnet0 as its SmartConnect service address.
• The global outbound DNS settings are configured with the optional domain name server list and DNS search list, if provided.

External network settings can be edited using the WebUI or CLI. Figure 11 shows the Edit Subnet page from the WebUI.

**Networking > Edit Subnet**

**ext1** – ext1 [Delete subnet]

**Settings** Edit

- **Subnet:** 192.168.0.0
- **Netmask:** 255.255.255.0
- **MTU:** 9000
- **Gateway:** 192.168.0.1, Priority 1
- **SmartConnect service IP:** not set
- **VLAN tagging (802.1Q):** Disabled
- **VLAN ID:** not set
- **Hardware load balancing:** not set

**IP Address Pools** Add pool

- **ext1-Static-pool** – ext1 -- Static pool

  **Basic settings** Edit
  
  - **IP range (low-high):** 192.168.0.101-192.168.0.109

  **SmartConnect settings** Edit
  
  - **Zone name:**
  - **Connection policy:** Round Robin
  - **SmartConnect service subnet:**
  - **IP allocation method:** Static
  - **IP failover policy:** Round Robin
  - **Rebalance policy:** Automatic Failback

  **Pool members** Edit
  
  - **ext-1, Node 01 (192.168.0.101)**
  - **ext-1, Node 02 (192.168.0.102)**
  - **ext-1, Node 03 (192.168.0.103)**
  - **ext-1, Node 04 (192.168.0.104)**
  - **ext-1, Node 05 (192.168.0.105)**
  - **ext-1, Node 06 (192.168.0.106)**

**Figure 11 - WebUI Subnet Configuration**
VLAN Tagging

Virtual LANs (VLANs) are used to logically group together network endpoints, and to partition network traffic, e.g. for security. VLANs are tagged with a unique identifier to segregate traffic. FlexNet 2.0 supports VLAN tagging for use in external networks using VLANs. In FlexNet, VLANs are configured at the Subnet level.

Using Link Aggregation

Isilon OneFS supports the use of redundant NICs to provide layer-2 failover. OneFS link aggregation supports the IEEE 802.3ad protocol, and works with switches and clients that also support this protocol.

Note that OneFS uses link aggregation primarily for Network Interface Card (NIC) failover purposes. Both NICs are used for client I/O, but the two channels are not ‘bonded’ into a single 2 Gigabit link. Each NIC is serving a separate stream.

Link aggregation, which can be configured for a new subnet or an existing one, involves creating an IP pool with the aggregated interface on each node as the pool’s members:

1. On the Edit Subnet page, at the top of the IP Address Pools section, click the Add pool link.
2. In the Create Pool wizard, enter a name for the pool, and optional description, and a range of IP addresses to use for this pool. Click Next.
3. If SmartConnect is used, options for the pool can be set on the next page of the wizard. Once these options have been selected, click Next.
4. On the next page, the interfaces to be members of this pool are selected. To use link aggregation, select the ‘ext-agg’ interface for each node to be in the pool. The interface type is also listed as AGGREGATION.
5. Click Submit to complete the wizard.
### Ethernet Switches

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cisco</strong></td>
<td>Catalyst 3560</td>
</tr>
<tr>
<td></td>
<td>Catalyst 3750</td>
</tr>
<tr>
<td></td>
<td>Catalyst 6509</td>
</tr>
<tr>
<td><strong>Extreme Networks</strong></td>
<td>Summit 1i</td>
</tr>
<tr>
<td></td>
<td>Summit 5i</td>
</tr>
<tr>
<td></td>
<td>Summit 7i</td>
</tr>
<tr>
<td></td>
<td>BlackDiamond</td>
</tr>
<tr>
<td><strong>Force 10</strong></td>
<td>E600</td>
</tr>
<tr>
<td><strong>10 Gigabit Ethernet</strong></td>
<td>6500 Series</td>
</tr>
<tr>
<td><strong>Fujitsu</strong></td>
<td>XG700 (12 and 24 port models)</td>
</tr>
<tr>
<td><strong>Force 10</strong></td>
<td>C-Series, E-Series, S-Series</td>
</tr>
</tbody>
</table>

#### Minimum Requirements for Ethernet Switches
- Non-blocking switch fabric
- Minimum 1MB buffer per switch port
- Jumbo frame support

### Infiniband Switches

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cisco</strong></td>
<td>SFS 7008</td>
</tr>
<tr>
<td></td>
<td>SFS 7000</td>
</tr>
<tr>
<td><strong>Flextronics</strong></td>
<td>F-X430061</td>
</tr>
<tr>
<td></td>
<td>F-X430066</td>
</tr>
<tr>
<td><strong>Mellanox</strong></td>
<td>MTS2400</td>
</tr>
</tbody>
</table>
About Isilon Systems

Isilon Systems is the worldwide leader in clustered storage systems and software for digital content and unstructured data, enabling enterprises to transform data into information and information into breakthroughs. Isilon's award-winning family of IQ clustered storage systems combines Isilon's OneFS® operating system software with the latest advances in industry-standard hardware to deliver modular, pay-as-you-grow, enterprise-class storage systems. Isilon's clustered storage solutions speed access to critical business information while dramatically reducing the cost and complexity of storing it. Information about Isilon can be found at http://www.isilon.com.

© 2001-2008 Isilon Systems, Inc. All rights reserved. Isilon, Isilon Systems and OneFS are registered trademarks, and SyncIQ, SnapshotIQ, SmartConnect, and SmartQuotas are trademarks of Isilon Systems, Inc.