ION Accelerator™ Appliance 2.4.1

Configuration Guide

7.25.2014

FUSION-IO

Copyright Notice

The information contained in this document is subject to change without notice.

Fusion-io MAKES NO WARRANTY OF ANY KIND WITH REGARD TO THIS MATERIAL, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. Except to correct same after receipt of reasonable notice, Fusion-io shall not be liable for errors contained herein or for incidental and/or consequential damages in connection with the furnishing, performance, or use of this material.

Fusion-io: 2855 E. Cottonwood Parkway, Ste. 100 Salt Lake City, UT 84121 USA

(801) 424-5500

Document No. D0005447-003_2

この装置は、クラスA情報技術装置です。この装置を家庭環境で使用すると電波妨害を引き起こすことがあります。この場合には使用者が適切な対策を講ずるよう要求されることがあります。 VCCI－A

CONTENTS

Part One: Introduction

Overview

ION Accelerator™ software runs on a shared-storage appliance, based on ioMemory technology. It delivers an all-flash storage target designed to accelerate performance-sensitive applications. This extremely high-performance storage easily integrates into existing SAN environments without over-burdening rack space, power, or cooling.

ION Accelerator administration is simple and straightforward, and this guide provides the information you'll need for deployment and optimization. The basic ION Accelerator components include:

- High-speed networking – Fibre Channel or iSCSI
- 3TB of usable, persistent ioMemory flash storage in each of four PCIe slots. Usable capacity varies based on the number of slots used and the storage pool configuration. Up to 12TB of ioMemory capacity is supported, in High Availability (HA) or standalone (non-HA) mode.
- ION Accelerator software – Enables configuration and management of the ION Accelerator appliance
- Fusion-io VSL™ (Virtual Storage Layer) – An architecture that combines the persistent nature of ioMemory with a direct-access I/O interface

ION ACCELERATOR SOFTWARE FEATURES AND BENEFITS

Basic ION Accelerator software features and benefits include:

- Two-node, end-to-end High Availability (HA) configuration (clustering of ION Accelerator nodes) if the HA feature is purchased
- Seamless database integration with Oracle, Oracle RAC, SAP, SQL Server, MySQL, DB2, Informix, Sybase ASE, Sybase IQ and NoSQL
- Integration with the Fusion ioSphere GUI for full management capabilities
- An extensive Command-Line Interface (CLI) for configuration, node appliance management, and scripting

INTEROPERABILITY

Operating Systems and Multipathing

For information on specific OS software versions to use, refer to the Interoperability Matrix available from the Fusion-io Customer Support web site. For information on using multipathing, see *Multipathing Overview* later in this guide.

Browser Requirements

To operate the ION Accelerator GUI (based on ioSphere), you must use one of the supported browsers: Internet Explorer 10 or 11, or the latest version of Firefox or Chrome. If you are using the Firefox version that ships with SLES 11 or CentOS 6.0, you will need to update to the latest Firefox version in order for the management interface to work correctly.

Be sure to check the latest *ION Accelerator Release Notes* for information about the required firmware versions for ION Accelerator.

Clustering

ION Accelerator is interoperable with industry-standard clustering solutions. Refer to the Interoperability Matrix for details. For more information on clustering, see the *HA (High Availability) and Configuring Clusters* section.

Part Two: First Boot and Connection

First Boot Setup

The ION Accelerator software comes pre-installed on the appliance.

The following tasks are handled in the First Boot of the ION Accelerator:

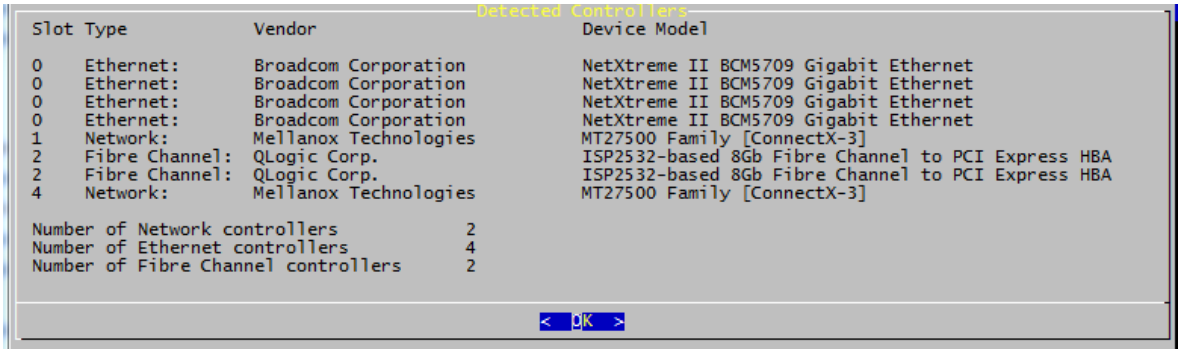
- License Agreement
- Network configuration, including management node and HA (if enabled)
- Date and time setup
- Cluster setup (if the HA feature is licensed)
- Password configuration

The screen examples shown in this section assume an HA configuration. If you are using ION Accelerator in standalone (non-HA) mode, skip over the steps that apply to HA configuration.

If you need to adjust First Boot settings after the First Boot process has completed, use the Network settings dialog (see *Network* in *Configuring ION Accelerator Settings*) or the `setup:lan` command in the CLI (see *setup:lan* in the *Command-Line Reference* section of the *ION Accelerator CLI Reference*).

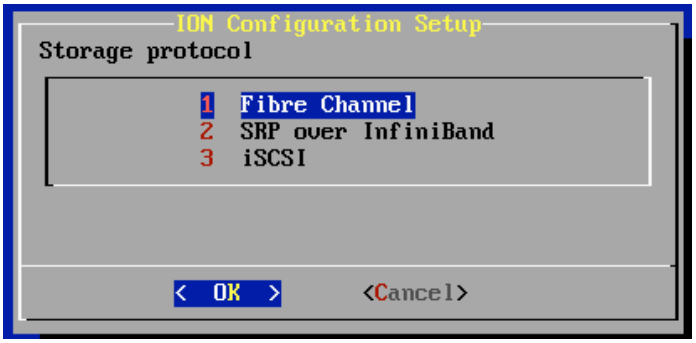
ION Accelerator Configuration

After installation, the appliance OS scans for existing network cards, and detected controllers are displayed. The sample screen below shows an example of detected controllers:



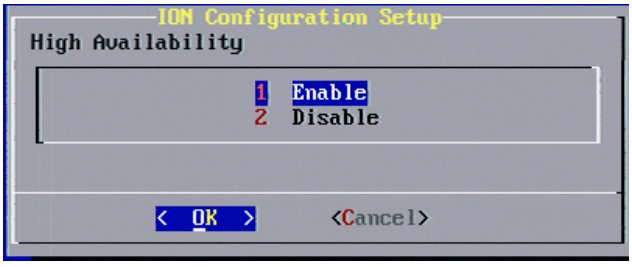
For standalone mode, ION Accelerator supports Fibre Channel and iSCSI protocols; however, *only one* of these protocols can be running at a time. For example, if a QLogic and a Mellanox card are installed in the same appliance, the standalone install of ION Accelerator will fail to recognize one of the two cards and will not display that card in the GUI.

1. Select **OK** to proceed. The **Storage protocol** dialog appears, with the detected protocol displayed as the default.



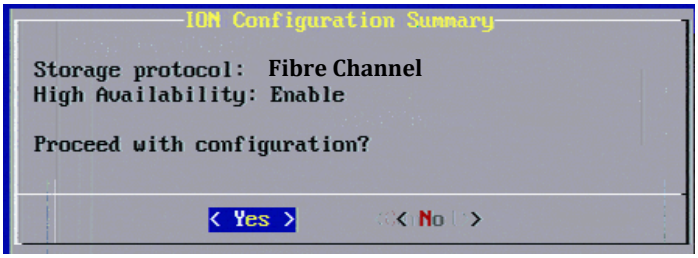
2. Select the storage protocol you want to use with ION Accelerator, and then select **OK**.

   The **High Availability** dialog appears, with the default determined by whether HA hardware was detected:



3. Select **Enable** if you want to enable HA mode (requires the ION Accelerator HA option), or **Disable** to use standalone mode, and then select **OK**.

The **ION Configuration Summary** dialog appears, with a summary of the selected protocol and mode (HA or standalone).



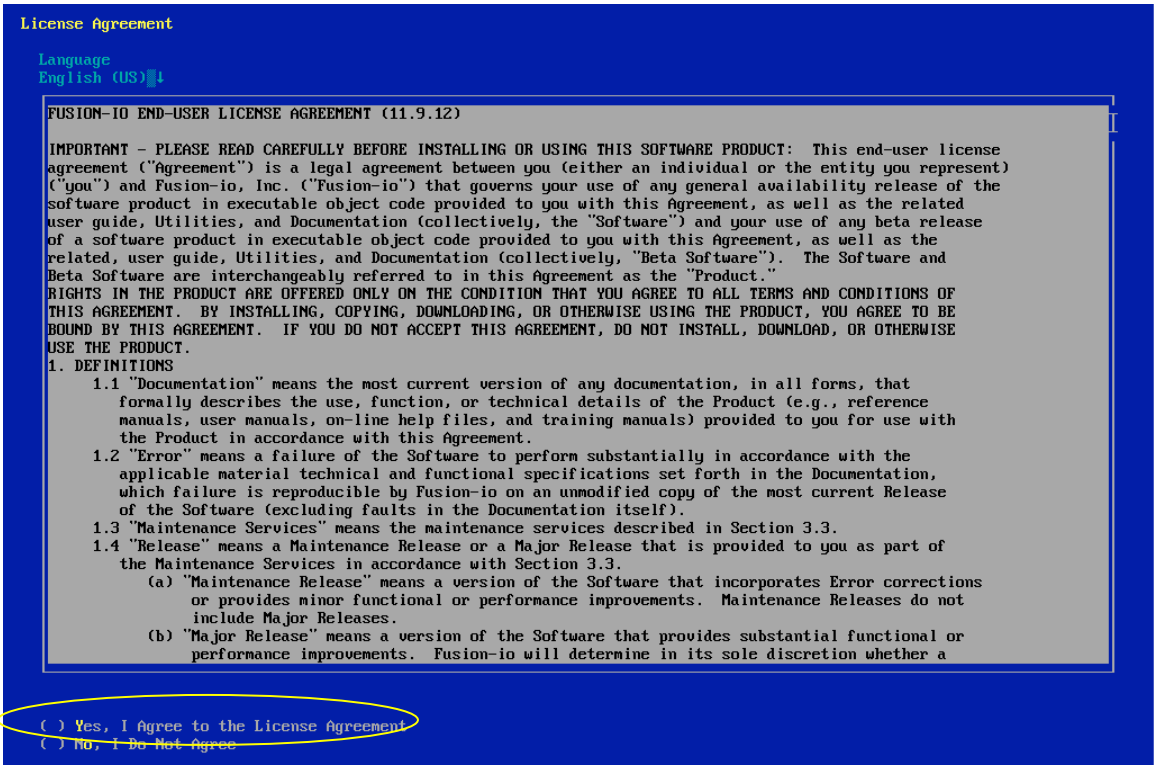
4. Select **Yes** to proceed with the ION Accelerator configuration.

   Or

   Select **No** to return to the **Detected Controllers** dialog (beginning of this section). This gives you the opportunity to review and correct configuration choices.

LICENSE AGREEMENT

The End-User License Agreement (EULA) screen appears, shown below (in part).



1. Check the "Yes" entry at the bottom of the screen to accept the agreement.
2. Select **Next** to continue.

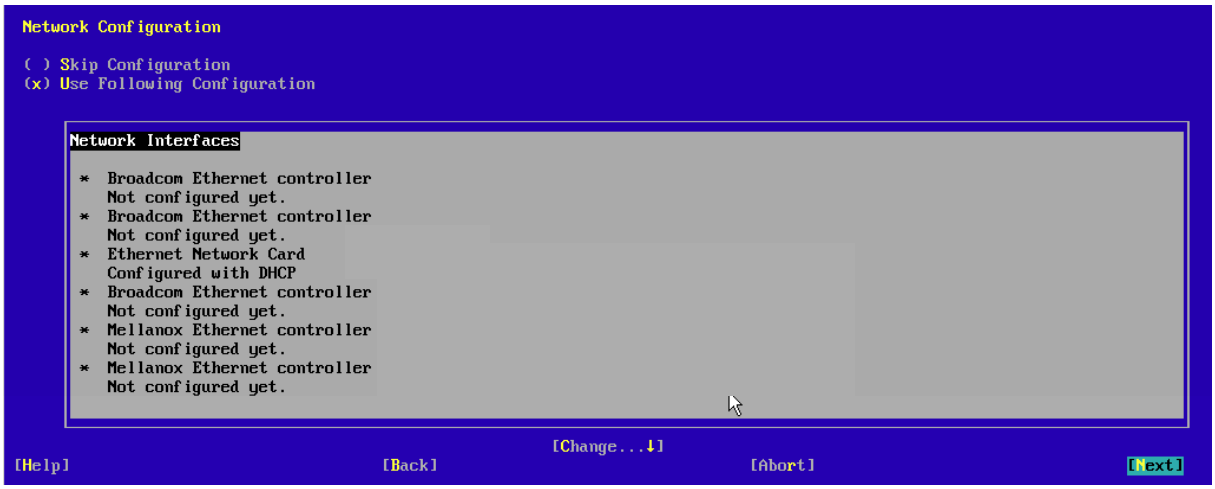
NETWORK CONFIGURATION OVERVIEW

The basic network configuration process for ION Accelerator is as follows:

1. Set the hostname and routing information for the appliance.
2. Configure the management node port that will be used to communicate with ION Accelerator.
3. If the HA feature is enabled, configure the HA port addresses for each HA network card.
4. If the adapter uses the iSCSI protocol, configure all iSCSI ports. (Fibre Channel adapters are automatically configured.)
5. Verify and save the configuration.

After the License Agreement is accepted, the appliance OS scans for existing network cards. The main **Network Configuration** screen then appears, with network interface cards listed.

In the example below, "Ethernet Network Card" refers to the management IP, and "Mellanox Ethernet controller" refers to the cluster interconnect controller(s).



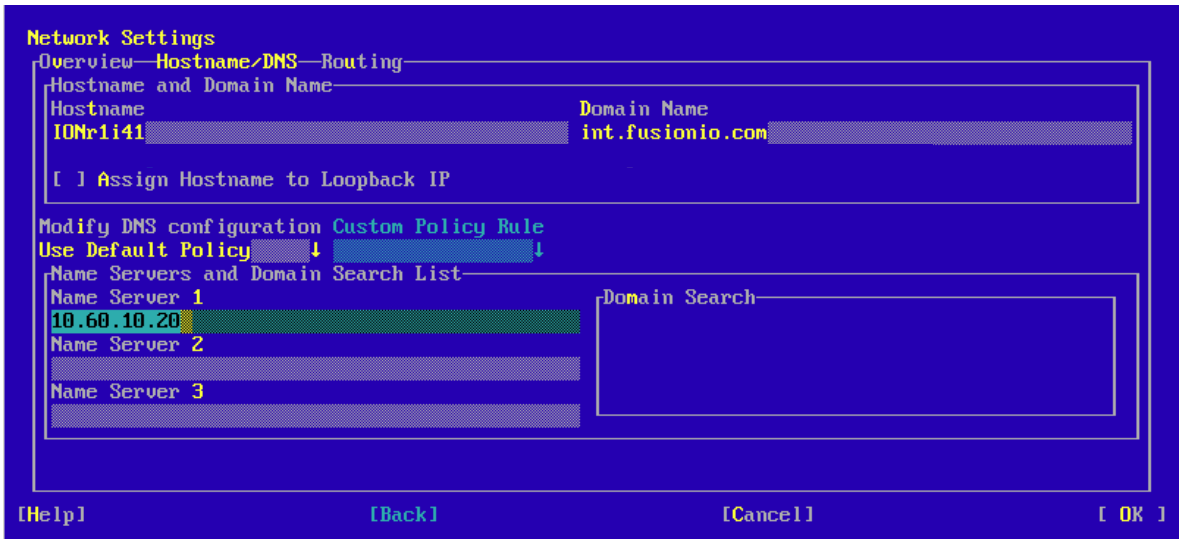
To make changes to the Network Configuration after First Boot finishes, use the CLI command `setup:lan` to invoke First Boot screens (see the *ION Accelerator Appliance CLI Reference* for details).

SETTING THE HOSTNAME AND ROUTING INFORMATION

Once you have configured all the network devices, you need to set the hostname and routing information for ION Accelerator.

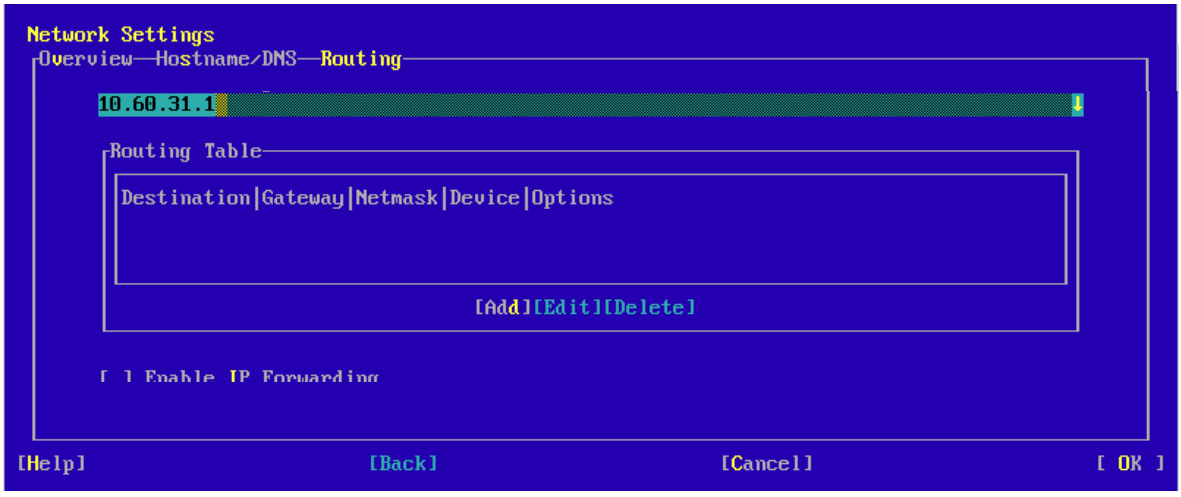
1. Select **Change** at the bottom of the Network Configuration screen to display the Network Settings screen.

2. Select **Hostname/DNS** at the top of the Network Settings screen.
3. Clear the selection for "Change Hostname via DHCP".
4. Type the hostname and domain name into their fields on the screen. For example:



When you choose a host name, be sure it conforms to RFC 1035/ARPANET standards, in order to avoid possible browser cookie issues. (Names must start with a letter, end with a letter or digit, and have as interior characters only letters, digits, and hyphens.)

5. Select **OK**.
6. Select **Routing** at the top of the Network Settings screen.
7. Type the Default Gateway address, as shown below, and select **OK**.



8. Select **OK**.

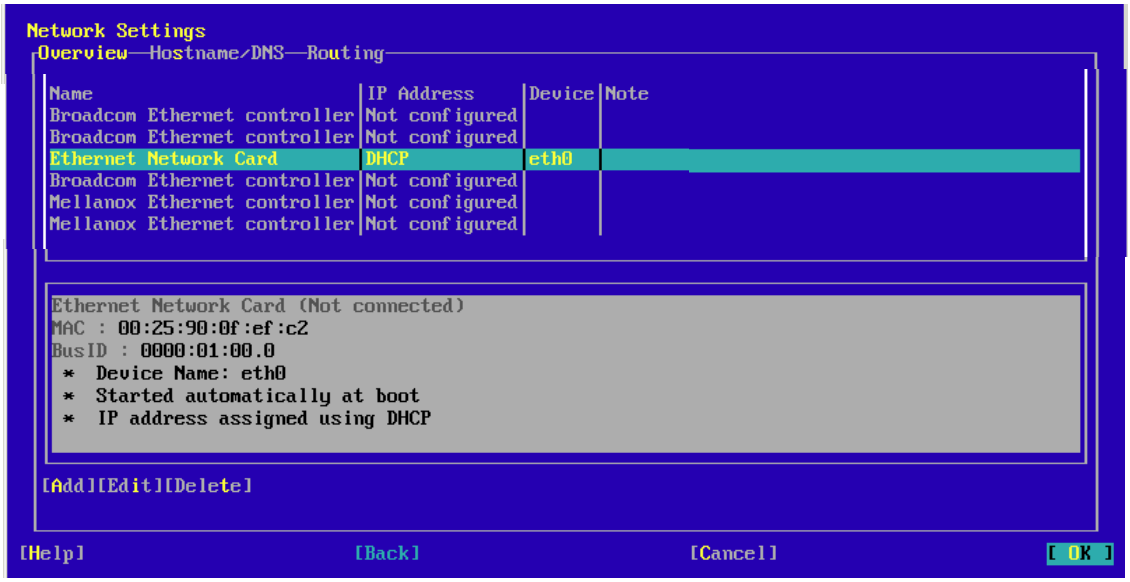
CONFIGURING THE MANAGEMENT PORT

The management port is connected by external cable and is publicly visible. This port needs to be configured so your network hardware communicates with ION Accelerator.

1. Select **Change** (bottom of the Network Configuration screen) and then **Network Interfaces**.

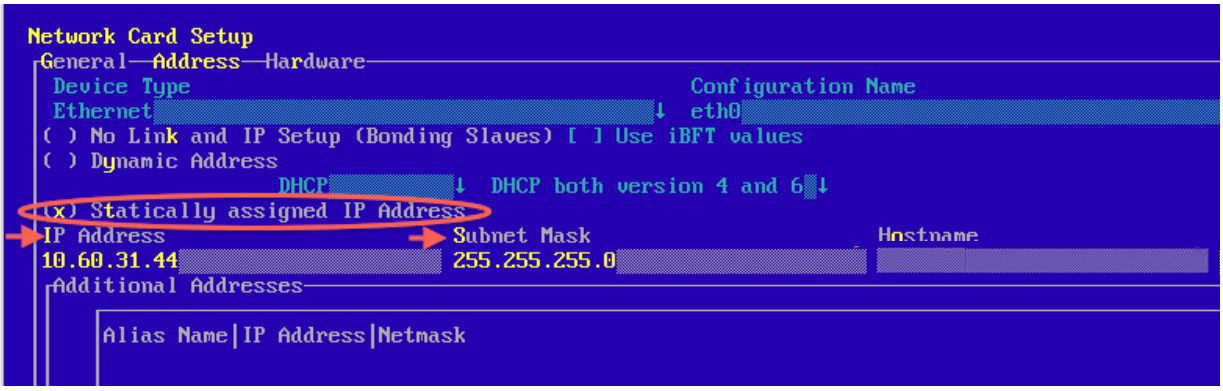
   The Network Settings – Overview screen appears, with the detected network adapters listed.

2. Select the management IP card from the list.



3. Select **Edit** (near the bottom of the screen). 

   The Network Card Setup – Address screen appears.



4. Select "Statically assigned IP Address" as shown above.
5. Type the IP address and subnet mask for your network, using the screen above as a guide.
6. Select **OK**.
7. Click **Next** to return to the Network Settings – Overview screen.

CONFIGURING ADDITIONAL DEVICES

The process for configuring additional network devices is essentially the same as what was done for the management port:

1. Select the next unconfigured device.
2. Complete the fields in the Network Card Setup – Address screen as described in the previous section.
3. Click **Next** to return to the Network Configuration Overview screen.

Cluster Interconnect Guidelines

- It is strongly recommended to use well-known static IP addresses for the cluster interconnect, such as 192.168.1.1 and 192.168.2.1 for one node, and 192.168.1.2 and 192.168.2.2 for the other.
- Ensure that the IP addresses are correct for HA connectivity.
- Ensure that two NIC ports have been configured, one for each HA node.

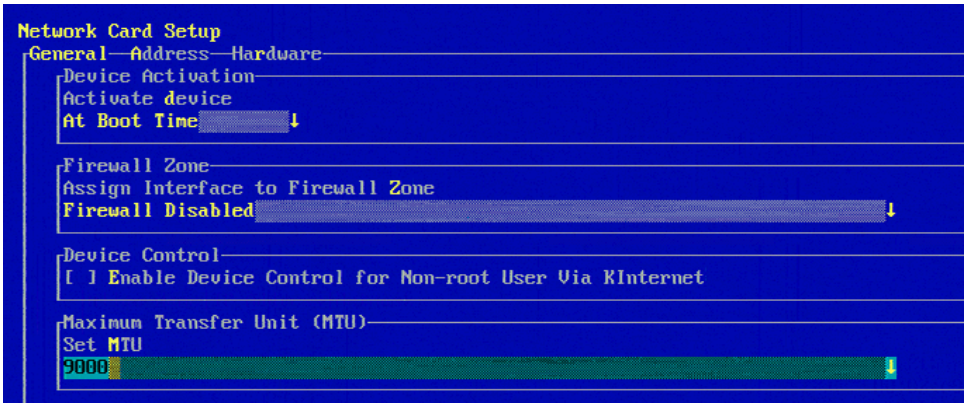
iSCSI Guidelines

If you are configuring multiple cards for iSCSI, ensure each card is placed on a *different* subnet. If both iSCSI ports are configured for the same subnet, the configuration will fail.

In the ION Configuration Setup screens, ensure you selected the iSCSI protocol (and enabled HA, if applicable).

If you need to set the MTU (Maximum Transmission Units) value for an iSCSI port,

1. Select **General**.
2. Press **Tab** until you reach the "down arrow" icon and then choose an MTU value from the dropdown list.



Configuring iSCSI Split-Port Capability

For additional information on port splitting for iSCSI, refer to to *Split-Function Ports for iSCSI – Dell R720* in *Appendix A: Dell PowerEdge R720 Configuration*.

In the **Network Settings** screen, each network card has its own unique Bus ID and MAC address. (The second port on the same card does not show the Bus ID.)

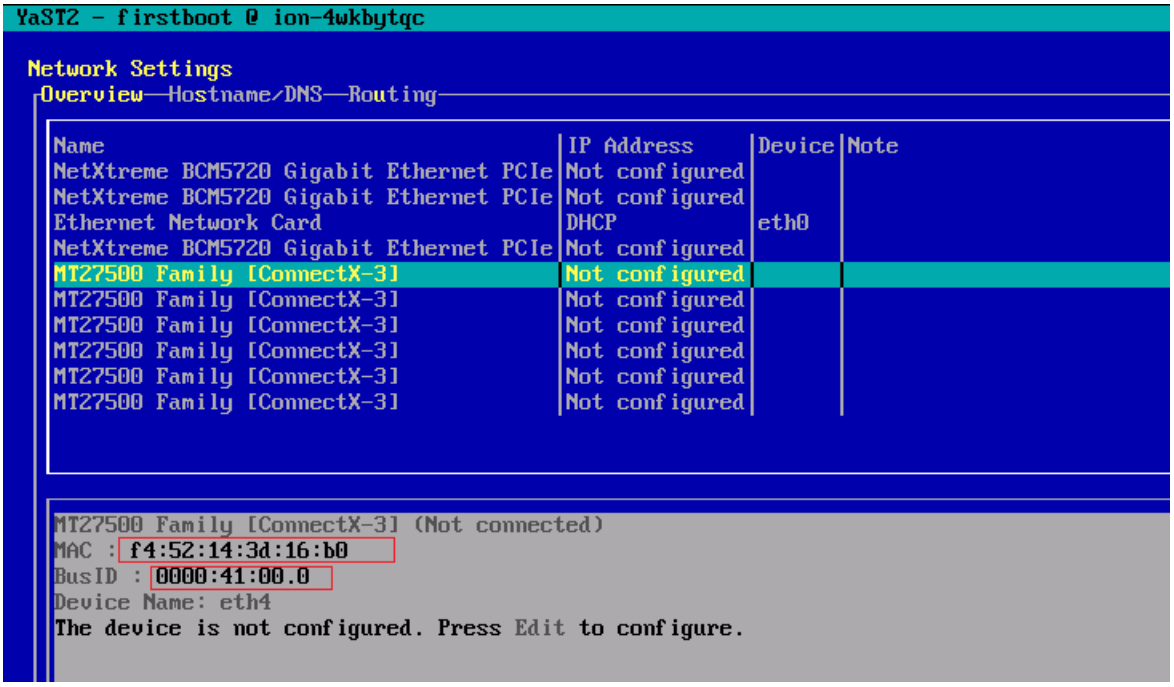
With three Mellanox ConnectX-3 cards installed in slots 1, 2, and 3, the corresponding bus IDs are as follows:

- Slot# 1 → bus ID 0000:41:00.0
- Slot# 2 → bus ID 0000:44:00.0
- Slot# 3 → bus ID 0000:43:00.0

Below is a sample two-node HA configuration, using a cluster interconnect setup in split-port mode:

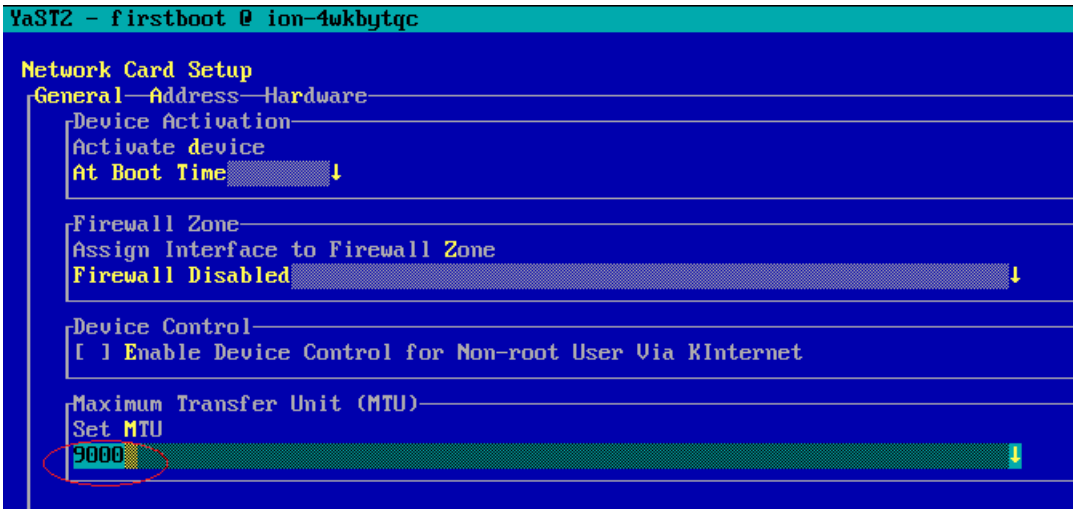
- Node1, slot 1, port1 (192.168.1.1) → Node2, slot1, port1 (192.168.1.2)
- Node1, slot 2, port1 (192.168.2.1) → Node2, slot2, port1 (192.168.2.2)

All other ports are configured as normal for iSCSI. Below is a sample screen *before* the port configuration is done:



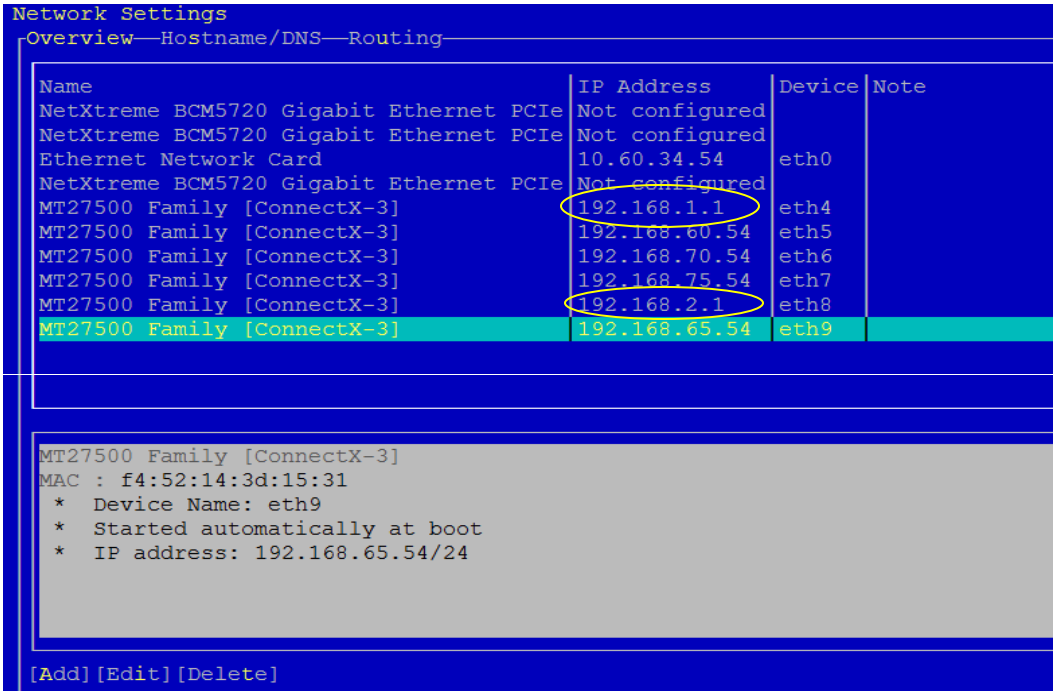
1. Edit the first port with the correct IP address for port splitting, as explained in the bullet list above.

2. Set the MTU value for the port:

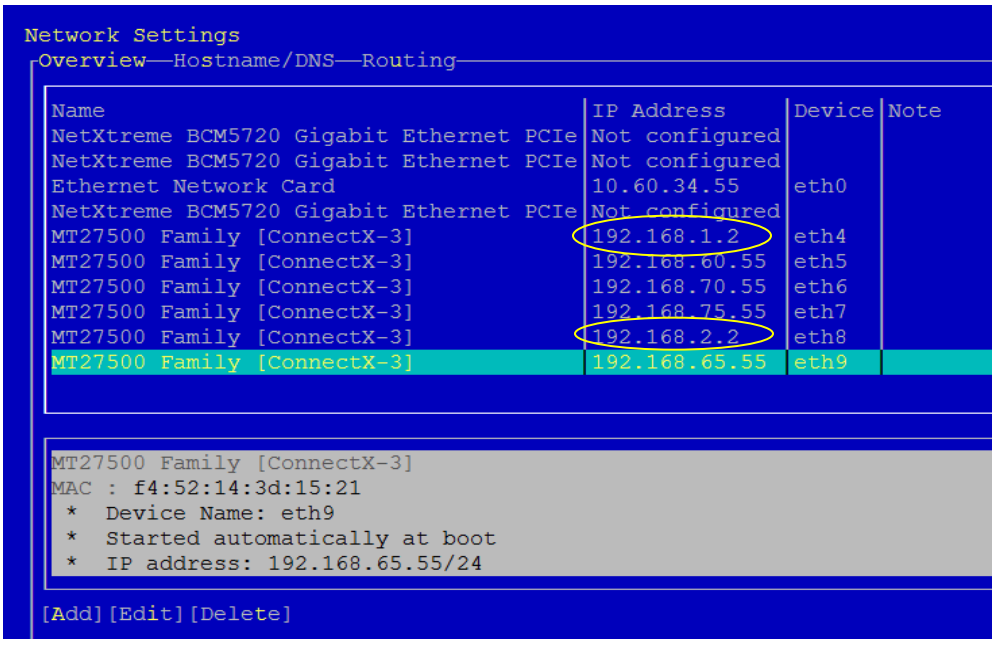


3. Repeat the previous two steps for each of the ports.

After split-port configuration, results for the *first* node will look similar to this:



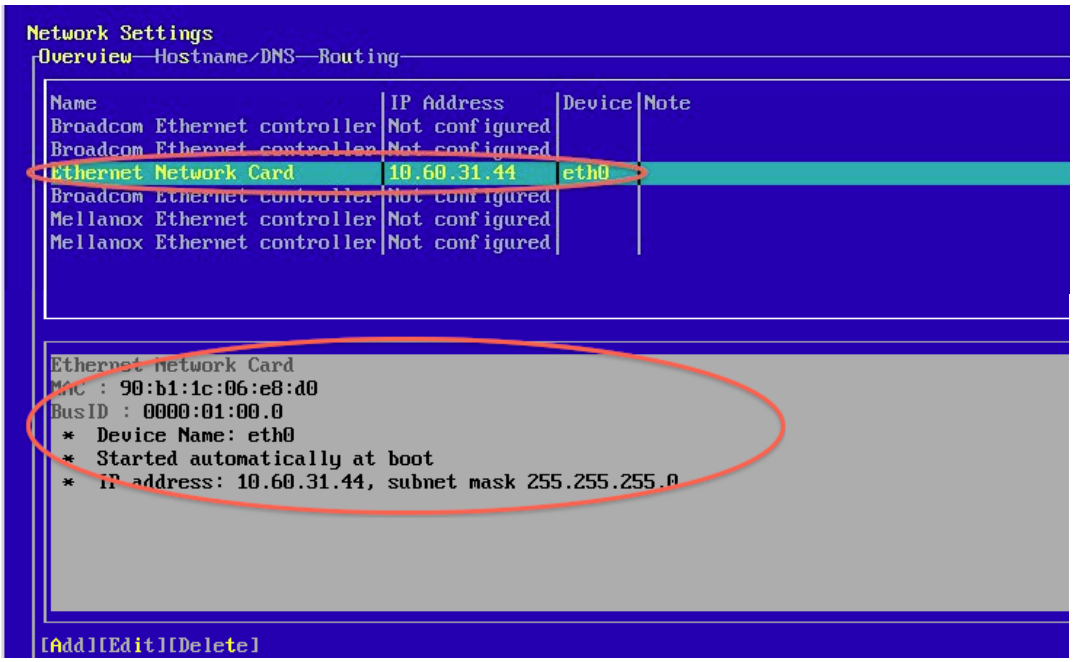
After split-port configuration, results for the *second* node will look similar to this:



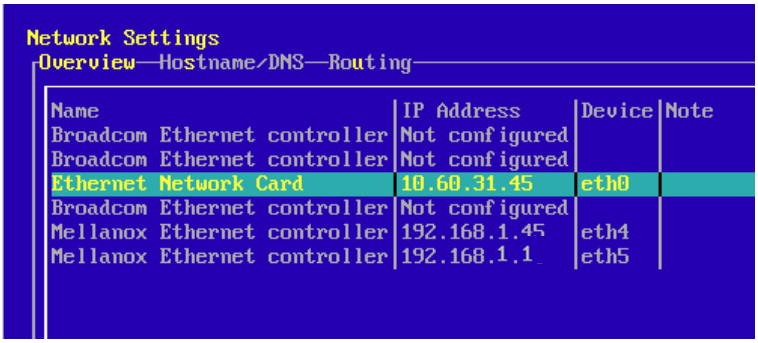
VERIFYING THE NETWORK SETTINGS

With your configuration choice completed (default DHCP, modified DHCP, or static IP address), the Network Settings screen shows your device for verification.

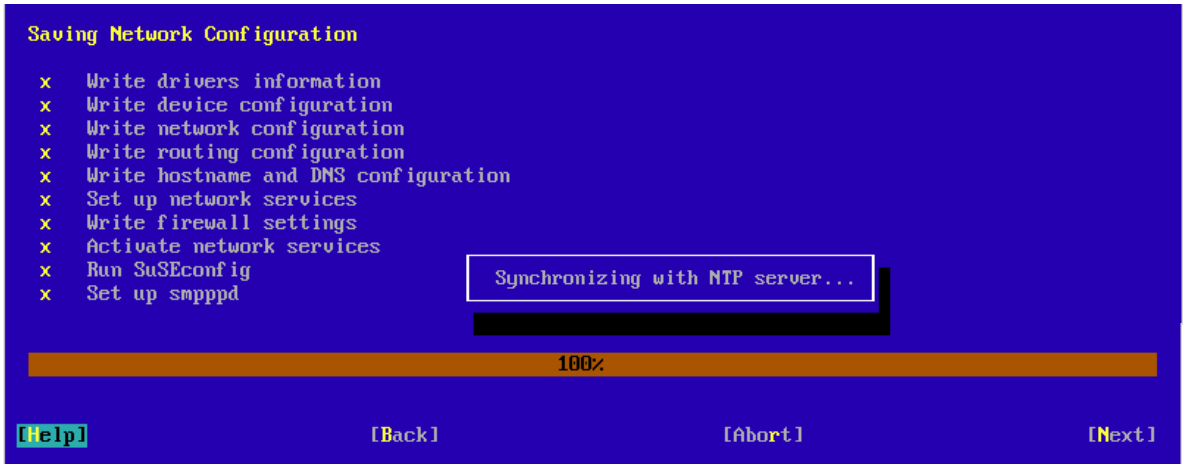
1. Review the settings (circled below) and select **OK** to finish Network Configuration.



A summary screen appears, listing the settings for the HA nodes (both `eth4` and `eth5` are configured in this example):

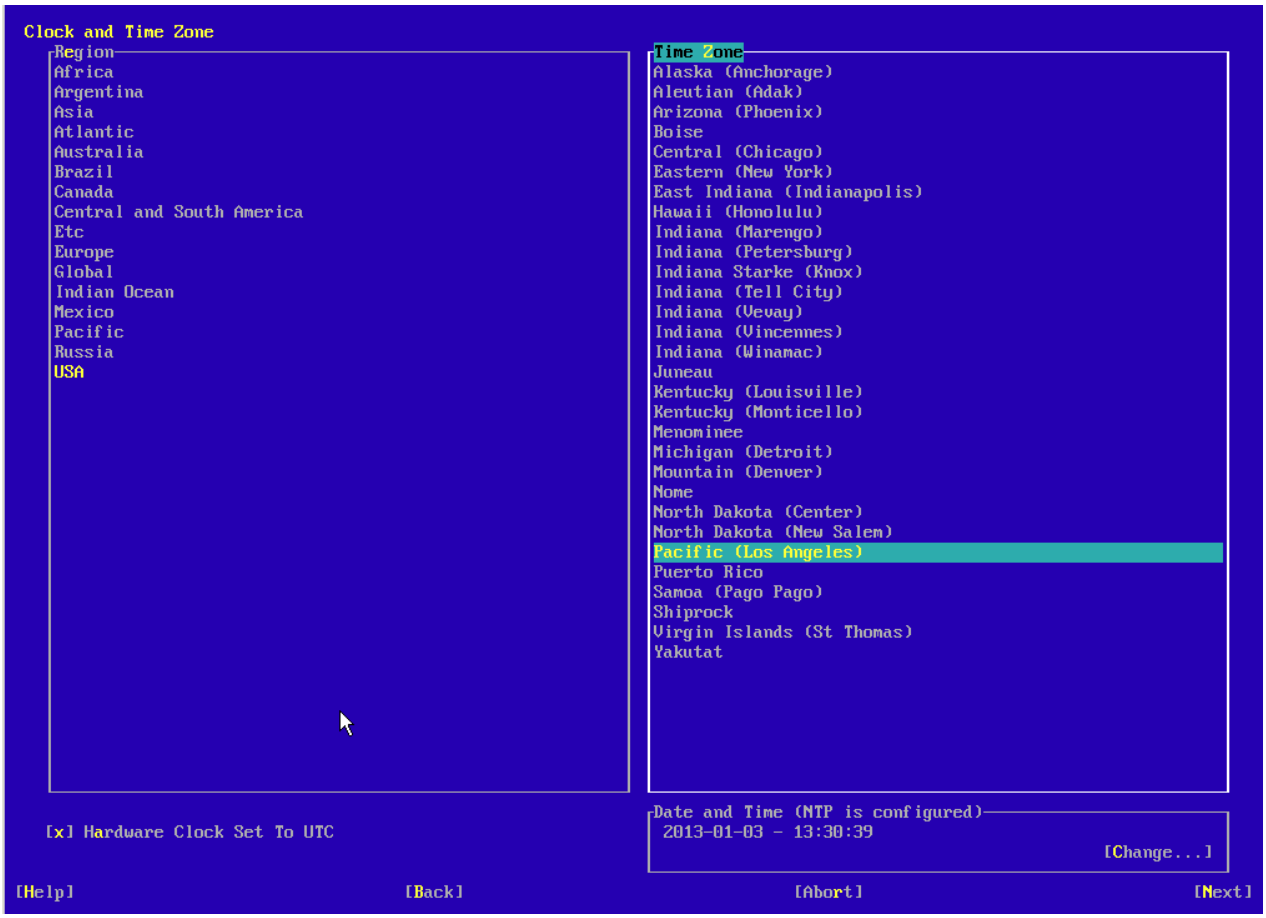


2. Select **OK** to save the configuration. A progress bar appears, and each completed configuration task is listed. At the end of the save process, the appliance is automatically synchronized with the NTP server, as shown below.



3. Select **Next** to continue to the Clock and Time Zone screen (shown below).

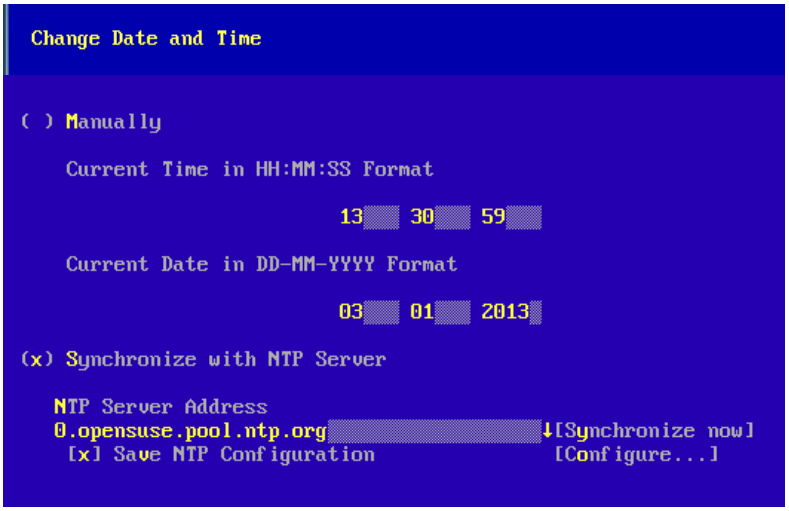
SETTING THE CLOCK AND TIME ZONE



1. To select the Time Region, select **Region** (left side of the screen) and press the Down arrow to select your region.
2. To select the Time Zone, select **Zone** (right side of the screen) and press the Down arrow to select your time zone. You can also manually change the current date and time, or disable the setting of the hardware clock to UTC.
3. Select **Next** to proceed to the Change Date and Time screen (shown below).

SETTING THE DATE AND TIME

In this screen you can manually change the current date and time for the server, or you can synchronize the server time with an existing NTP server.



If you plan on configuring HA/clustering with the ION Data Appliance, you *must* use the "Synchronize the NTP Server" option.

To manually set the date and time, select "Manually" and edit the Current Time and Current Date fields.

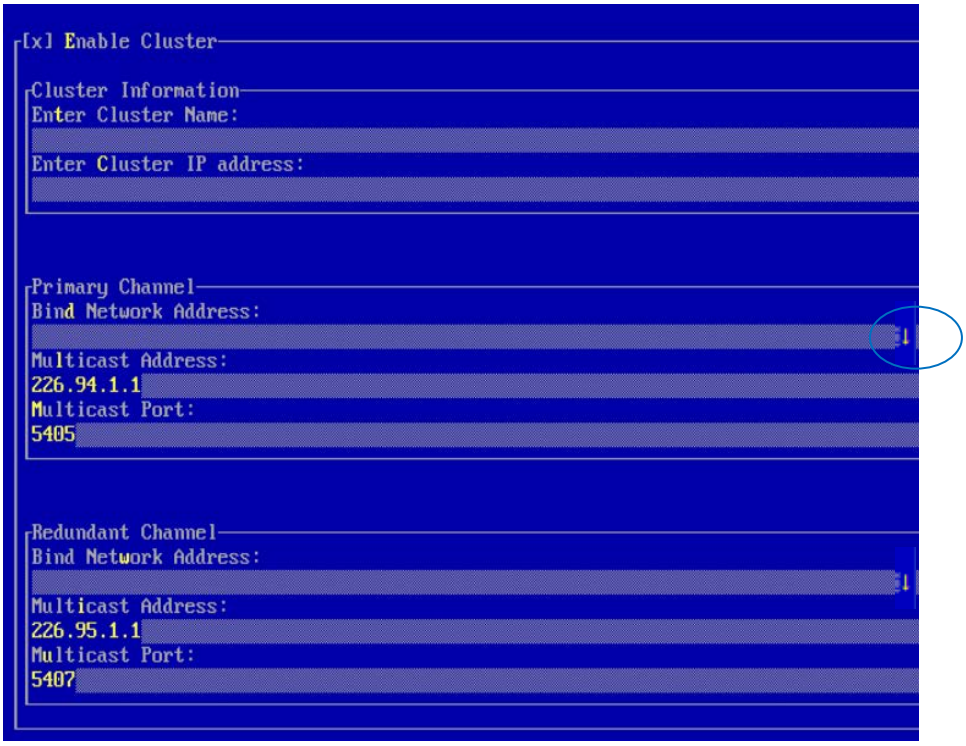
To synchronize the current date and time with the NTP server,

1. Select "Synchronize with NTP Server".
2. Type the URL of the NTP server to synchronize with.
3. You can also select "Synchronize now", "Save NTP Configuration", and "Configure" to further control NTP settings.
4. Select **Accept** to save the NTP changes and continue.

If you are *not* using HA, then the Password Setup screen appears; in that case, skip ahead to *Password Setup* below.

ENABLING THE HA CLUSTER

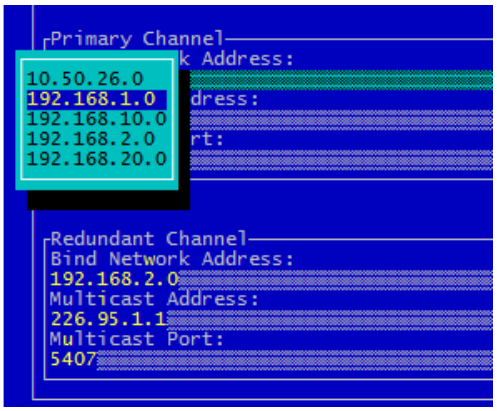
If you are using the HA feature, the Enable Cluster screen appears, as shown below. (If you are not using HA, skip ahead to *Password Setup* below).



1. Enter the cluster name and cluster IP address in the Cluster Information fields.

Do *not* choose the name or IP address of one of the currently existing nodes. Also, the cluster name must be a DNS entry, not DHCP, and must be on the same subnet as the management IP address for the ION Accelerator nodes.

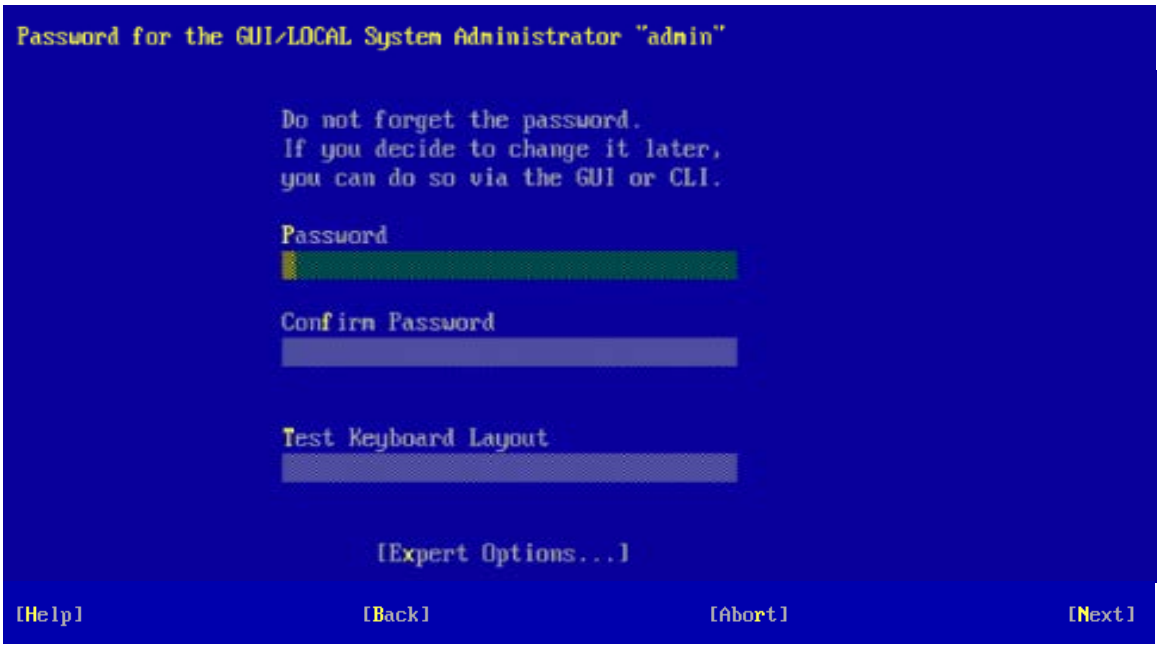
2. In the Primary Channel area, tab to the drop-down arrow (circled above) corresponding to the Bind Network Address.



3. Choose the correct IP subnet to assign to the primary channel (192.168.1.0 in the example above).
4. In similar manner, choose the correct IP subnet for the Redundant Channel, from its drop-down menu.
5. Select **OK** to save the changes.

SETTING UP PASSWORDS

A sample Password Setup screen is shown below. If you decide to change this password later, you will need to change both the admin and GUI passwords separately. See *Changing the Admin Password* and *Local Accounts* for more information.



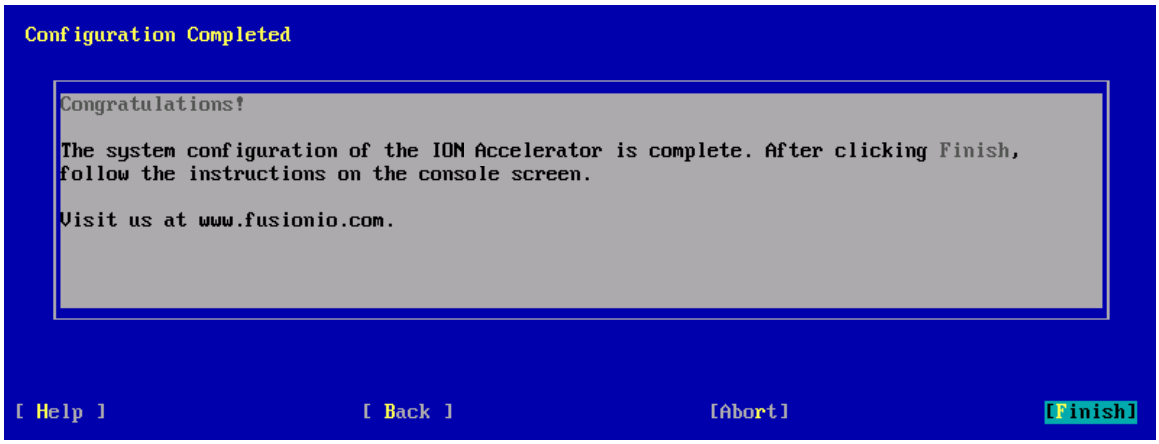
1. Type the ION Accelerator password for the admin user.

If the password you selected is not sufficiently strong, a warning message appears so you can change the password.

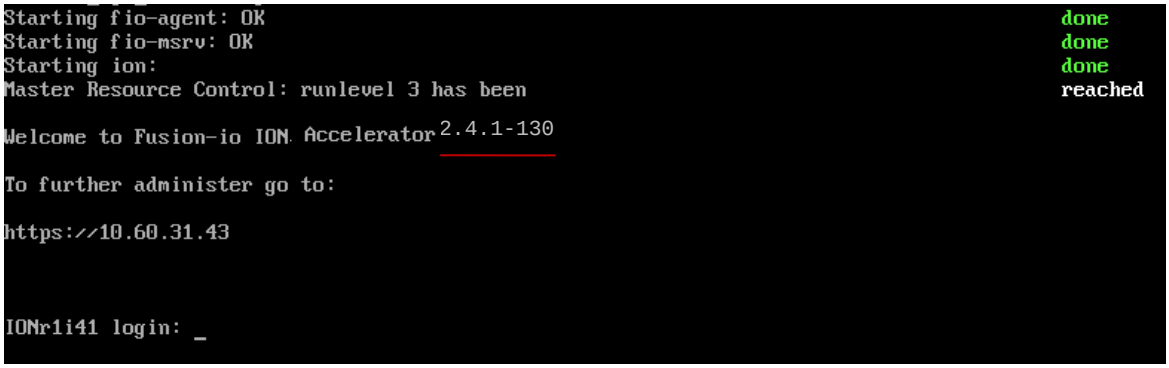
2. Confirm the password you entered. **Important**: Be sure to record this password in a secure location in case it needs to be retrieved.
3. If you want to test the keyboard layout or use the Expert Options, select those choices on the screen.
4. Select **Next** to continue.

COMPLETING FIRST BOOT

The First Boot configuration is now complete, as shown below.



1. Select **Finish** to prepare for login. The console will display results similar to the following:



Connecting to ION Accelerator

If you are setting up a permanent HA cluster, it is recommended to connect to each node. That way, if failback occurs, first-time setup on the second node won't be necessary.

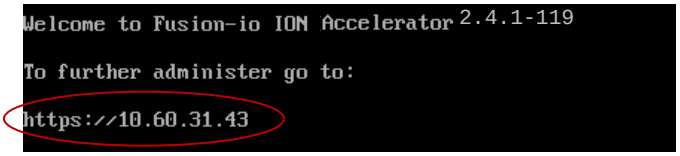
You can access and use the ION Accelerator software in two ways: via the command line or via the GUI (Graphical User Interface).

Logging In via the CLI (Command-line Interface)

1. At the login prompt, type **admin** as the username and press **Enter**.
2. At the password prompt, type the password you created in First Boot and press **Enter**.
3. Once you are logged in to ION Accelerator, proceed to the *ION Accelerator CLI Reference* for instructions on how to set up and manage your ioMemory storage.

Logging In via the GUI

1. To log in with your browser, access the URL indicated in the console screen, which is the management port you configured in First Boot. (The example below uses 10.60.31.43.)

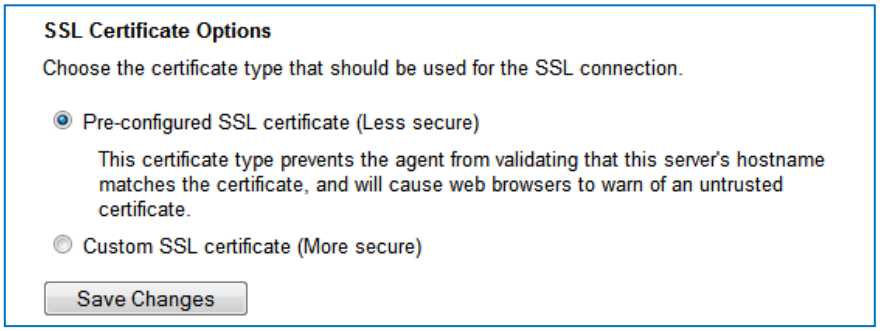


2. At the login prompt, type **admin** as the username and press **Enter**.
3. At the password prompt, type the password you created in First Boot and press **Enter**.

When you access the ION Accelerator URL, the Admin Password login prompt appears:

1. Type **admin** for the admin username.
2. Type the password in the second field (this is the password you created in First Boot).
3. Click **Login**.

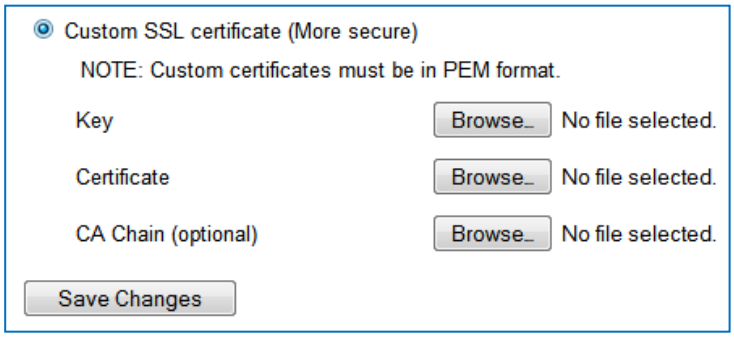
   The SSL Certificate Options dialog appears. By default, the "pre-configured" option is selected.



4. To use the less-secure pre-configured SSL certificate, click **Save Changes**.

   Or,

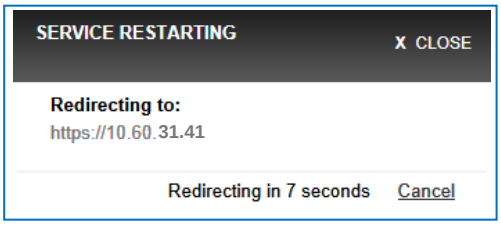
   To use a custom SSL certificate, select the "Custom SSL certificate (more secure)" option and then click **Save Changes**.



The default "Use pre-configured Fusion-io SSL Certificate" option is less secure, as it causes "untrusted certificate" warnings to appear.

5. Click **Save Changes**.

If the port, host name, or SSL information has changed, the service restarts in a few seconds, and the browser is redirected to the port you specified.



If the host name has changed, the service restarts, and the login process is initiated again. After re-direction, the login dialog appears again.

6. Log in with the same username and password as before.

   The software is now configured so you can begin to manage your ION Accelerator appliance. The Overview screen appears, as shown in the *Basic Tasks and Overview Tab* section.

   For items not covered in the steps above, see *Remote Access* in the *Configuring ION Accelerator Settings* section.

To change the admin password, see *Changing the Admin Password* in the *Tabs Bar* section. For details on managing admin and user passwords, see *Users* in the *Configuring ION Accelerator Settings* section of this guide.

The admin account allows up to 10 concurrent sessions to run on the ION Accelerator appliance.

7. Once you are logged in to ION Accelerator, proceed to the *ION Accelerator Appliance GUI Guide* for instructions on how to set up and manage your ioMemory storage.

CHANGING NODE NAMES AND IP ADDRESSES

This section explains how to change the names or IP addresses for ION Accelerator nodes once they have been configured in the First Boot process. For details on using the `system:setup` command referred to in this section, refer to the *ION Accelerator CLI Reference*.

Changing a Node Name in a Cluster

Changing a node name requires cluster downtime.

1. Ensure that both nodes are up.
2. Close all active sessions by disabling or disconnecting all target ports.
3. In the CLI, run `system:setup lan` to display the Setup dialog for LAN configuration.
4. Change the name of the node in the dialog.
5. Select **OK** and press **Enter**. Both nodes will restart, one at a time.
6. Repeat the previous two steps for the second node in the cluster.

Changing the Management IP Address

1. Ensure that both nodes are up.
2. In the CLI, run `system:setup cluster` to display the Setup dialog for cluster configuration.
3. Change the management IP address in the dialog.
4. Select **OK** and press **Enter**. The node with the cluster IP address will fail over to the other node.

Changing the ION Accelerator Cluster Name or Cluster IP Address

1. Ensure that both nodes are up.
2. In the CLI, run `system:setup cluster` to display the Setup dialog for cluster configuration.
3. Change the cluster name in the dialog.
4. Select **OK** and press **Enter**.

Changing the Gateway IP Address

1. Ensure that both nodes are up.
2. In the CLI, run `system:setup lan` to display the Setup dialog for LAN configuration.
3. Click the **Routing** tab in the dialog.
4. Change the gateway IP address.
5. Select **OK** and press **Enter**.

Changing the IP Addresses for Cluster Interconnect Ports


You can change interconnect ports *only* when no cluster resources are configured.

1. Ensure that both nodes are up.
2. In the CLI, run `system:setup lan` to display the Setup dialog for LAN configuration.
3. Change the cluster interconnect IP address in the dialog.
4. Select **OK** and press **Enter**.
5. Restart both nodes.

Changing the iSCSI Port IP Address

1. Ensure that both nodes are up.
2. In the CLI, run `system:setup lan` to display the Setup dialog for LAN configuration.

3. Change the iSCSI port IP address in the dialog.
4. Select **OK** and press **Enter**.
5. For Linux initiators, add the new target session to the initiator by running the following command:

   ```
   iscsiadm -m discovery -t st -p <new iSCSI IP address> <new iSCSI IP address> <iqn:eth#>
   ```

6. For Linux initiators, log in to the target (all sessions) from the initiator:

   ```
   iscsiadm -m node –l
   ```

7. Run the CLI command `initiators –dt` to display the new initiator.
8. Add this initiator to an existing group: `initiator:update –a <inigroup name> <initiator iqn>`

Part Three: HA and Host Configuration

HA (High Availability) and Configuring Clusters

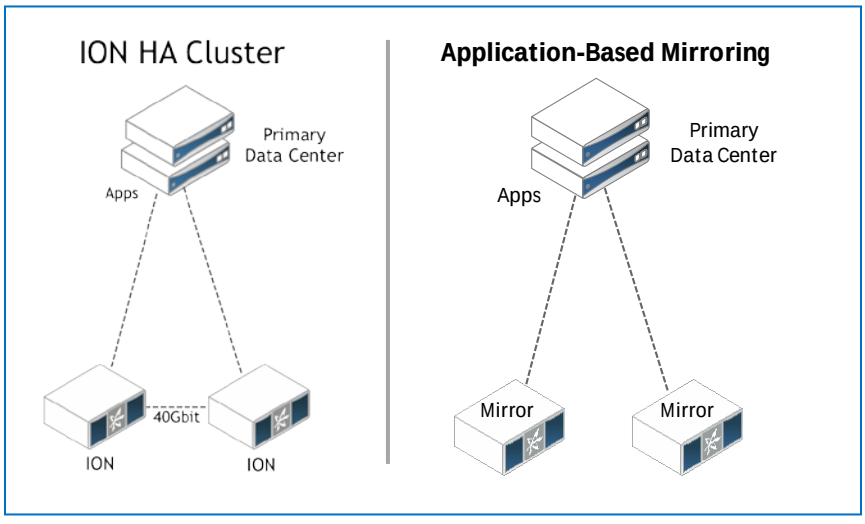
If you are not planning to implement the high-availability feature of the ION Accelerator software, skip ahead to the next section.

ABOUT ION ACCELERATOR HIGH AVAILABILITY

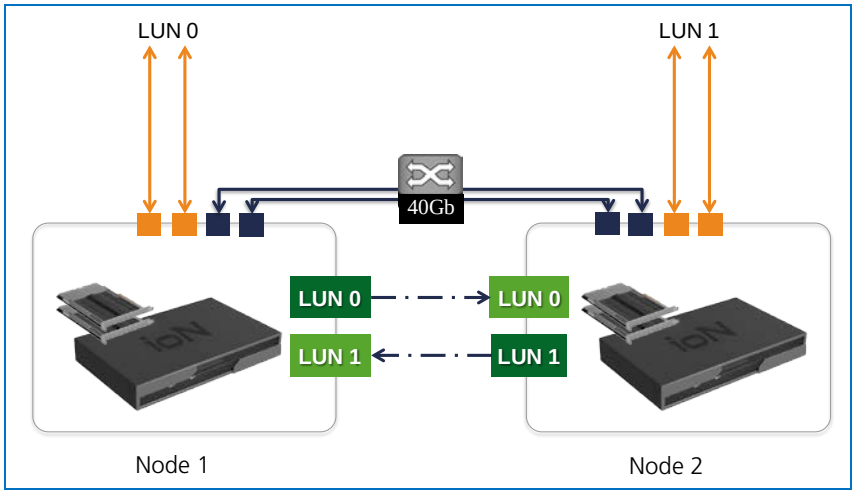
ION Accelerator enables a powerful and effective High Availability (HA) environment for your shared storage. HA clustering provides an important option for customers who prefer array-based HA over application-based mirroring. This can be especially useful if your application does not provide logical volume management, such as with all VMware environments and most implementations of Microsoft Clustering.

Clustering relies on point-to-point connections – HA networking across geographically distributed sites is not supported.

The diagram on the left shows a simple HA clustering setup, using a 40Gb Ethernet connection between appliances. The diagram on the right shows an ION Accelerator mirroring configuration without clustering, an alternative approach.



The diagram below shows basic LUN access (exported volumes) in an HA configuration.



In this simplified example, Node 1 presents LUN 0 to the application, while Node 2 presents LUN 1 to the application. All writes to LUN 0 are synchronously replicated to Node 2. All writes to LUN 1 are synchronously replicated to Node 1. When the replication is complete, the original node acknowledges the write by the application.

Replication occurs over the 40Gb Ethernet interconnect between the two nodes, which consists of two dual-ported ConnectX®-3 adapters. In the event a node fails, all data will be available from the remaining active unit.

ISCSI CONNECTIONS FOR HA

iSCSI systems use dual-port ConnectX®-3 cards for cluster interconnections.

For HA systems, the two ports on the interconnect card are "split" – one port on each of the first two ConnectX®-3 cards is used for the cluster interconnect, and one is used for the fabric connection. See the *First Boot Setup* section for details and dialogs regarding cluster setup, and refer to *Split-Function Ports for iSCSI – Dell R720* in *Appendix A: Dell PowerEdge R720 Configuration* for details on split-port configuration.

FABRIC AND CLUSTER INTERCONNECT CARDS (ISCSI)

The photos below show common fabric and cluster interconnect cards, with ports labeled. As you configure your ION Accelerator HA system, compare these port locations with the ones shown in *Appendix A: Dell PowerEdge R720 Configuration*. This will ensure that the optimal port connections are made for a high-performance HA configuration.

ConnectX®-3 cluster interconnect card

Port 1 Port 2

Multipathing Overview

For information on managing initiators, such as creating initiator groups, or adding or deleting initiators, refer to the *ION Accelerator Appliance GUI Guide* or the *ION Accelerator Appliance CLI Reference*.

NETWORK FABRIC AND HOSTS SUPPORTED

ION Accelerator supports Fibre Channel and iSCSI network fabrics for initiators. Supported operating systems include

- Windows Server 2008 with iSCSI
- Windows Server 2012 with Fibre Channel
- SLES
- CentOS
- OL,
- ESX
- RHEL
- AIX
- Solaris
- HP_UX

Also supported are VMware, Hyper-V, and Oracle VM hypervisors.

Refer to the *ION Accelerator Interoperability Matrix*, available from Fusion-io Customer Support, for complete details on hardware compatibility and multipathing support.

ABOUT MULTIPATHING

Multipath I/O (MPIO) establishes multiple routes and connections to a storage array, using redundant physical paths (adapters, cables, switches). That way, when a component fails, an alternate I/O path is used. Multipathing provides redundancy of I/O paths and can improve overall system performance.

Before multipathing is installed, you see one drive for each path setup (so with two paths, you see two drives). After multipath installation, you will see a single drive for all paths of the same drive. The drive Properties dialog, available by right-clicking on the drive in the Disk Management tool, now has a multipath tab, where the load-balance policy is set.

If you are using Fibre Channel, make sure you have the latest HBA driver installed before configuring multipathing.

LOAD-BALANCE POLICIES FOR MPIO

Below are some basic load-balance policies typically used with multipathing:

- Round-robin, subsets: Standby paths are used only if all primary paths fail.

For Windows MPIO, the "round-robin with subsets" method is required for HA configurations; for Windows standalone configurations, round-robin or dynamic least queue depth methods may be used, with the latter generally preferred.

- Failover: No balancing; standby paths are used.
- Failback: I/O is rerouted to preferred path when available.
- Round-robin: All available paths are used for balanced I/O.
- Dynamic least queue depth: I/O to path with fewest outstanding requests (in Linux, this is "queue-length"). This enables multipathing to compensate for an unbalanced load on the fabric. This may be advantageous for standalone configurations.
- Weighted path: Paths are assigned priority weights.

Configuring Multipathing on Windows

INSTALLING MULTIPATH ON WINDOWS

When MPIO is installed, the Microsoft device-specific module (DSM) and the MPIO Control Panel are also installed.

For additional information on Windows MPIO settings, refer to http://blogs.msdn.com/b/san/archive/2011/12/02/updated-guidance-on-microsoft-mpio-settings.aspx

In Windows 2008 R2 SP2, disk I/O operations may fail even when valid failover paths exist. To resolve this problem, apply the Microsoft hotfix available at http://support.microsoft.com/kb/2752538/en-us

Installing Multipath on Windows Server 2008

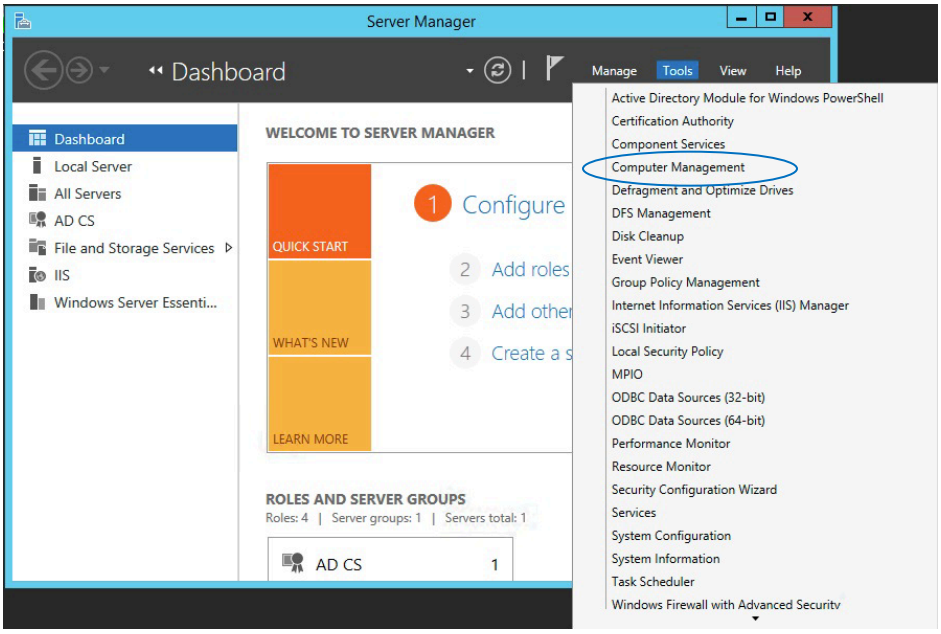
To install MPIO on a server running Windows Server 2008 R2,

1. Open Server Manager (**Start** > **Administrative Tools** > **Server Manager**).
2. In the **Server Manager** tree, click **Features**.
3. In the Features area, click **Add Features**.
4. In the Add Features Wizard, on the Select Features page, select the Multipath I/O check box, and then click Next.
5. On the Confirm Installation Selections page, click **Install**.
6. When the installation finishes, on the Installation Results page, click **Close**.
7. When prompted to restart the computer, click **Yes**. After restarting, the computer finalizes the MPIO installation.
8. Click **Close**.

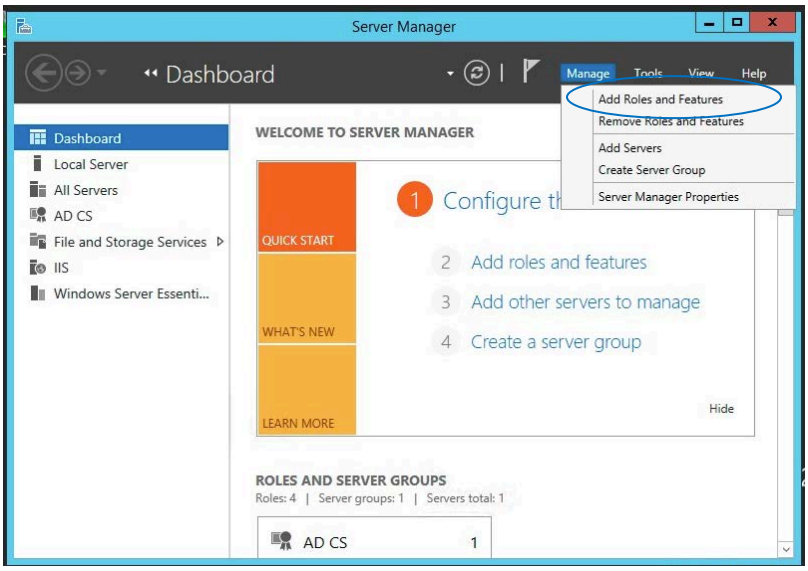
Installing Multipath on Windows Server 2012

To install MPIO on a server running Windows Server 2012 R2,

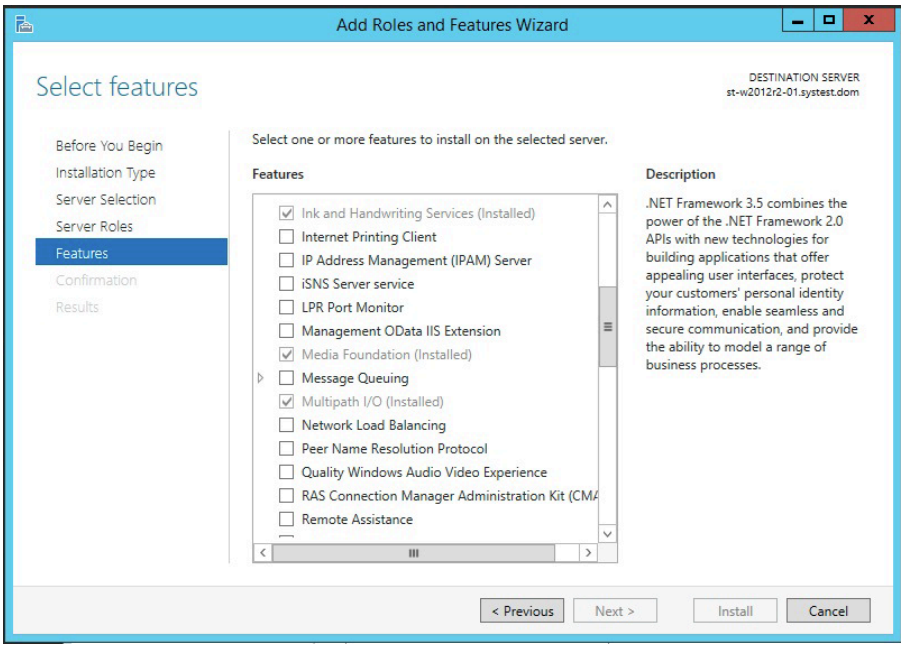
1. Start **Server Manager**.
2. Click **Tools** and select **Computer Management** from the pull-down menu.



3. In the Manage menu, click **Add Roles and Features**.



4. Click **Next** on "Before You Begin".
5. Select Role-based or Feature-based installation and click **Next**.
6. In the left pane, click Features and select **Multipath I/O** (already installed in this example).



7. Click **Next** to proceed with the installation
8. Click **Yes** if prompted to restart the computer.

CONFIGURING WINDOWS MULTIPATHING

The MPIO Control Panel enables you to:

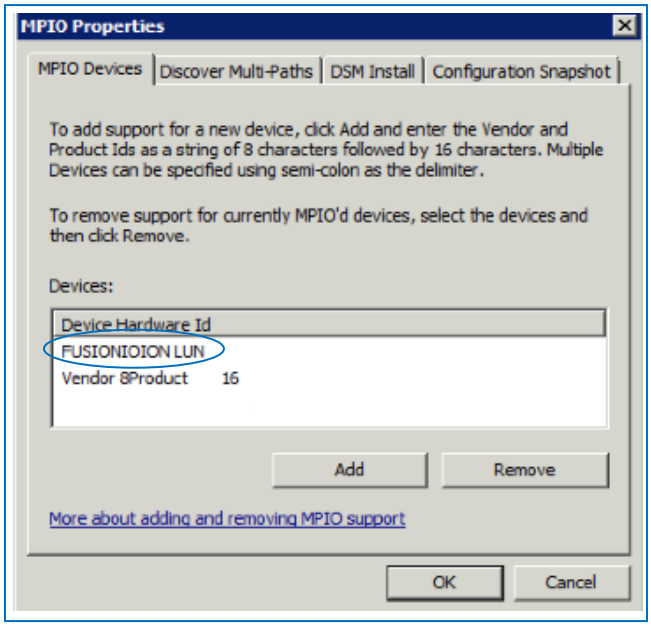
- Configure MPIO functionality
- Install additional storage DSMs
- Create MPIO configuration reports

Configuring Multipathing on Windows Server 2008

To configure Windows multipathing on Windows Server 2008,

1. Open the MPIO Control Panel (**Start** > **Control Panel** > **Views list** > **Large Icons** > **MPIO**.
2. On the User Account Control page, click **Continue**.

   The MPIO Control Panel opens to the Properties dialog, MPIO Devices tab. This tab displays the hardware IDs (such as FUSIONIOION LUN in the example below) of the devices managed by MPIO whenever they are present.



3. To add a hardware ID,

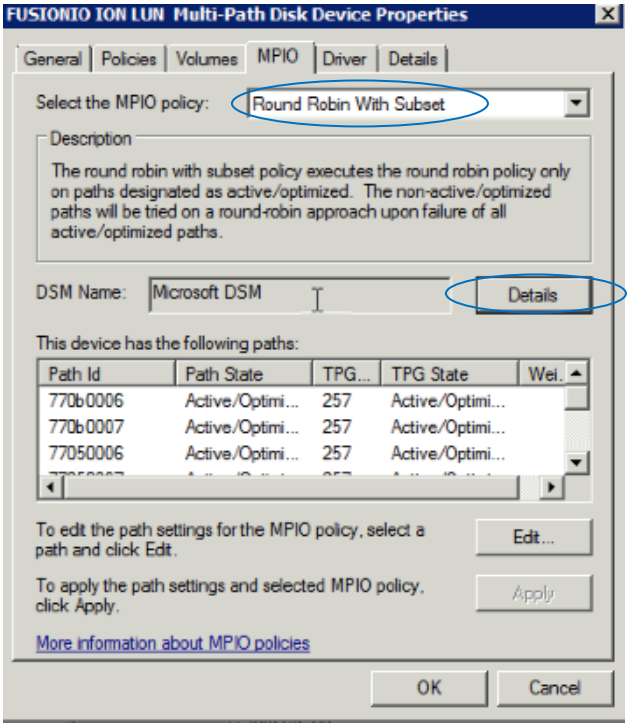
   a) Click **Add** in the dialog.

   b) Type the hardware ID, which is Vendor ID (eight characters) plus Product ID (16 characters).

   c) Click **OK**.

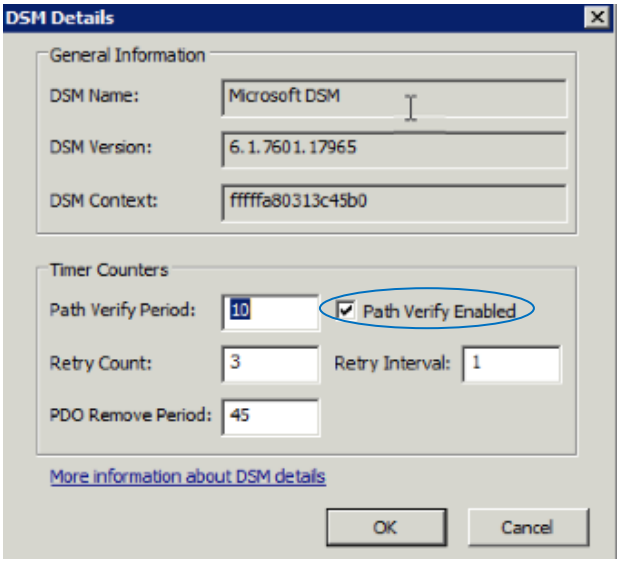
4. Click **OK**.
5. Open the Windows Disk Management tool (**Storage** >**Server Manager** >**Disk Management**).

6. Right-click on an ION LUN to display its properties.



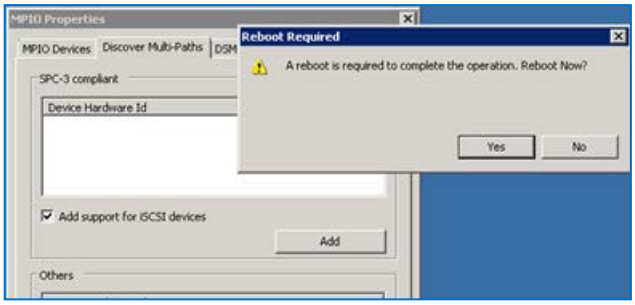
7. In "Select the MPIO Policy" on the MPIO tab, make sure "Round Robin With Subset" is selected.

8. Click **Details** to display the DSM Details dialog:



9. Make sure that "Path Verify Enabled" is checked as shown above.

10. Click **OK** to exit each dialog, until prompted to restart the system.



11. If you are running in standalone mode, click **Yes** to reboot, and then skip ahead to *Discovering Multipaths* below.

12. If you are running in HA mode, click **No** at the reboot prompt, as there are still additional steps to do.

13. Use `regedit` to edit the Registry with the following changes:

    ```
    HKLM\SYSTEM\CurrentControlSet\services

        \Disk\TimeOutValue = 60 secs
        \mpio\Parameters\UseCustomPathRecoveryInterval = 1
    ```

    If you are using Fibre Channel, use the 30-second interval shown below; if you are using iSCSI, set the interval to 5:

    ```
        \mpio\Parameters\PathRecoveryInterval = 30 secs
    ```

14. For the QLogic HBA (Fibre Channel), make the following changes in the driver, using the QLogic CLI or GUI in Windows:

    ```
    QLogic Port Link/Down Timeout = 15 secs
    QLogic Port Retry/Down Timeout = 15 secs
    ```

    Or if you are using Emulex drivers with Fibre Channel, make the following changes:

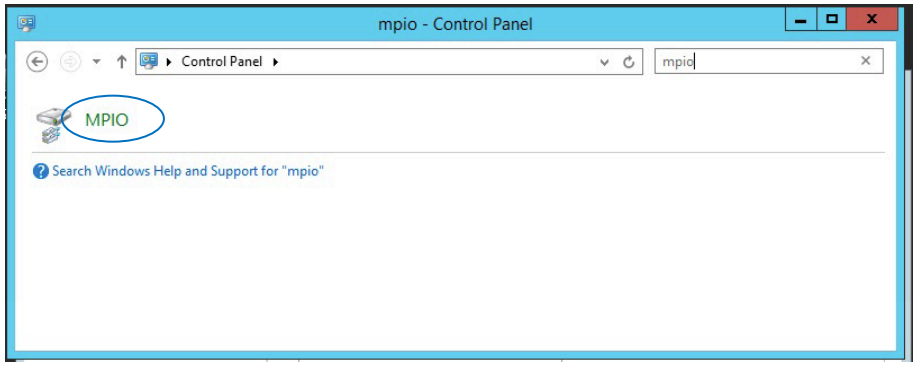
    ```
    Emulex LinkTimeout = 15 secs
    Emulex NodeTimeout = 60 secs
    ```

15. Restart the Windows system.

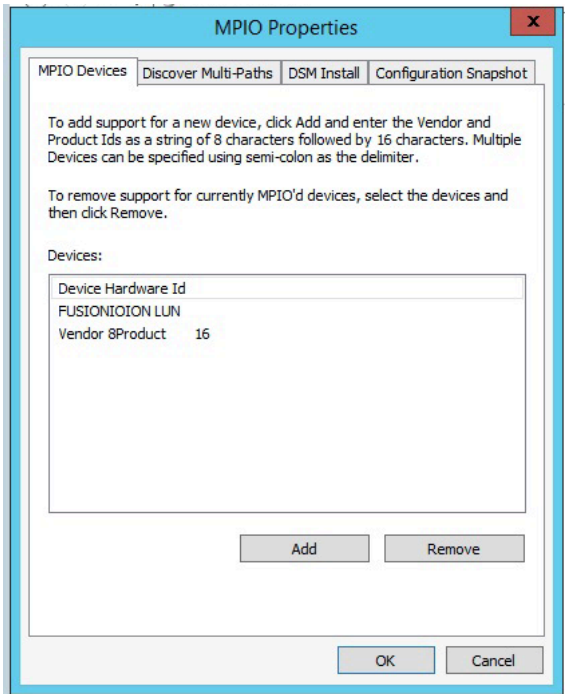
Configuring Multipathing on Windows Server 2012

To configure multipathing on Windows Server 2012,

1. Open the MPIO Control Panel and type **mpio** in the search box.
2. Click **MPIO** to start the mpio utility.

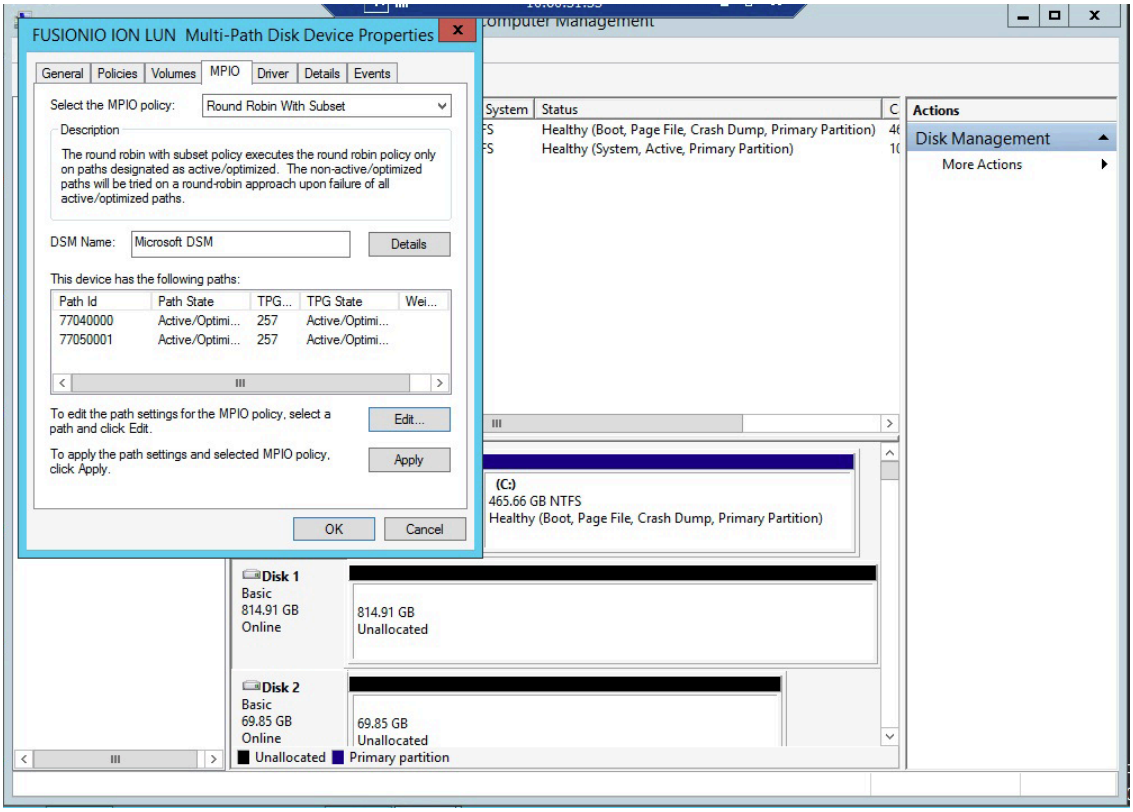


The Properties dialog appears, with the MPIO Devices tab active. This tab displays the hardware IDs (such as FUSIONIOION LUN in the example below) of the devices managed by MPIO whenever they are present.



3. Click **Add** in the dialog to add a hardware ID.

4. Type the hardware ID, which is Vendor ID (eight characters) plus Product ID (16 characters).
5. Click **OK** exit each dialog.
6. Open the **Windows Disk Management** tool (**Server Manager** > **Tools** > **Computer Management** > **Disk Management**).
7. Right-click an ION LUN to display its properties.



8. In “Select the MPIO Policy” on the MPIO tab, select “Round Robin With Subset”.

9. Click **Details** to display the DSM Details dialog.

10. Make sure that "Path Verify Enabled" is checked, as shown above.

11. Click **OK** to exit each dialog, until prompted to restart the system.

Discovering Multipaths

To discover multipaths that exist in your configuration,

1. Verify that multiple instances actually represent the same Logical Unit Number (LUN) through different paths. Hardware IDs for those devices are displayed for use with MPIO.

2. Add a second initiator to a volume.

3. With multipathing running, open a disk management tool, such as Windows Disk Manager.

4. View the volume from the initiator and verify that now only one volume shows up. (Before installing multipathing, the volume would show up twice).

CONFIGURING ISCSI INITIATORS FOR WINDOWS

The instructions in this section apply to multipathing for Windows Server 2008 R2.

Each initiator iSCSI port should be in a separate subnet that is the same as that of the target iSCSI port.

Standalone Mode

1. Make sure the following Windows hot fix module and settings are applied: http://support.microsoft.com/kb/2752538/en-us
2. Use `regedit` to disable the delayed ACK in the registry:

   ```
   HKLM\SYSTEM\CurrentControlSet\Services\Tcpip\Parameters\Interfaces\<I
   nterface GUID>\TcpAckFrequency = 1
   ```

   To create this registry setting:

   a) Right-click in the interface you are using for iSCSI.

   b) Select a new DWORD 32-bit value.

   c) Name the new value `TcpAckFrequency`.

   d) Set the value to 1.

3. Increase the reconnect retries value in the registry as follows:

   ```
   HKLM\SYSTEM\CurrentControlSet\Control\Class\{4D36E97B-E325-11CE-BFC1-
   08002BE10318}\<Instance Number>\Parameters\PortalRetryCount =
   ffffffff
   ```

4. After setting these registry values, restart the system.

HA Mode

1. Follow steps 1 to 3 in the Standalone Mode instructions above.
2. Go to Control Panel > MPIO > Add Device Hardware ID.
3. Specify the device as `FUSIONIOION LUN`.
4. Use `regedit` to set the timeout value to 60 seconds:

   ```
   HKLM\SYSTEM\CurrentControlSet\services\Disk\TimeOutValue = 60 secs
   ```

5. Use `regedit` to set the custom path recovery interval to 1:

   ```
   HKLM\SYSTEM\CurrentControlSet\services\mpio\Parameters\UseCustomPathR
   ecoveryInterval = 1
   ```

6. Use `regedit` to set the path recovery interval to 5 seconds:

   ```
   HKLM\SYSTEM\CurrentControlSet\services\mpio\Parameters\PathRecoveryIn
   terval = 5 secs
   ```

7. After setting these registry values, restart the machine.

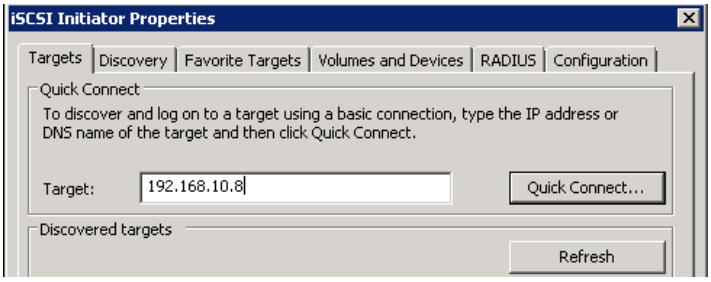
SETTING UP ISCSI IN WINDOWS

For the example used in this configuration, the following items should be noted:

- The Windows 2008 R2 initiator has an iSCSI card that uses IP addresses 192.168.10.125 and 192.168.11.125.
- The target is ION Accelerator in HA mode. Each node has a card installed and configured for iSCSI. The IP addresses are 192.168.10.8 and192.168.11.8 for the first node, and 192.168.10.9 and 192.168.11.9 for the second node.
- An ION Accelerator volume of 100GB is used, with LUNs already created and presented to the Windows initiator.

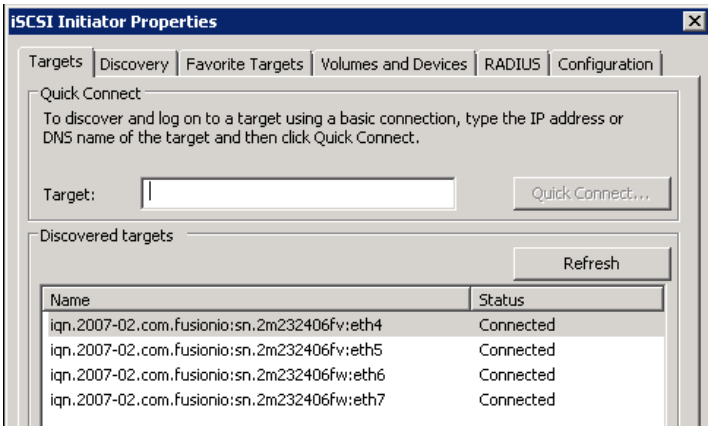
1. To open the Microsoft iSCSI initiator, click **Start** > **Administrator Tools** > **iSCSI Initiator**.

2. Log in to each target port and click **Quick Connect**.

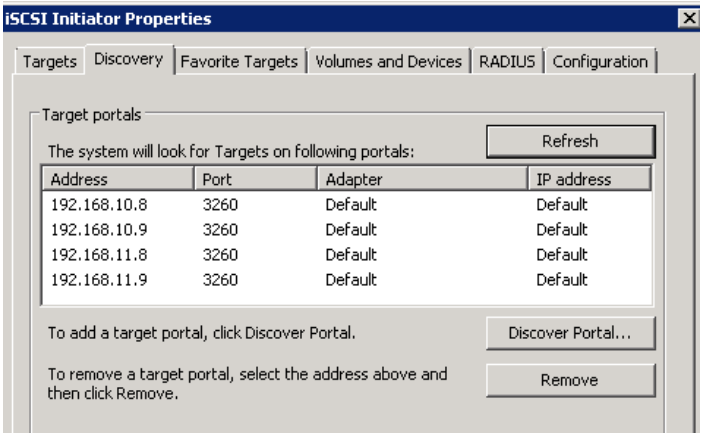


After successful login to all four targets, the target connection and discovery should be seen, as in the examples below:

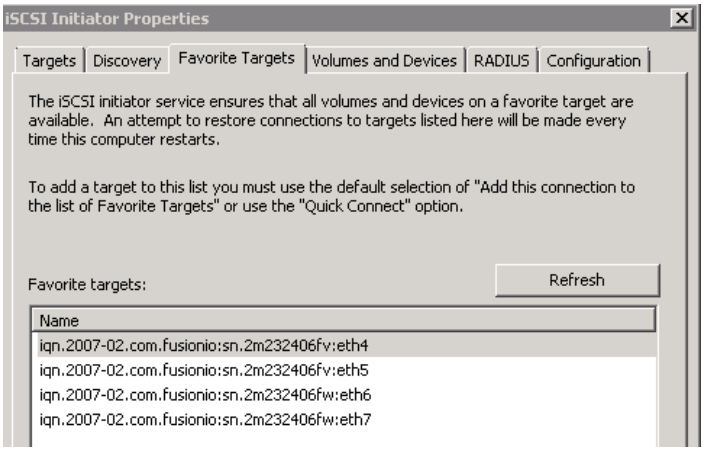
Discovered Targets (Targets Tab)



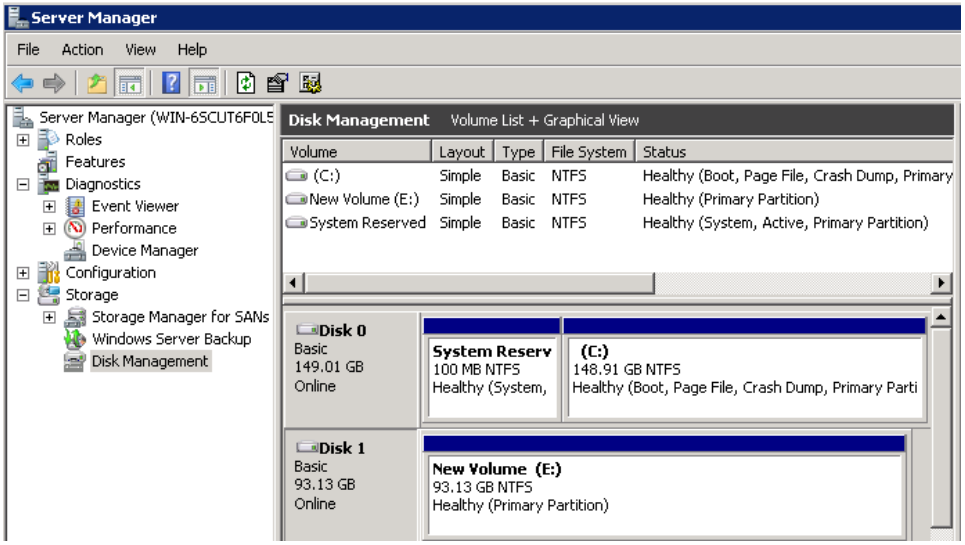
Target Portals (Discovery Tab)



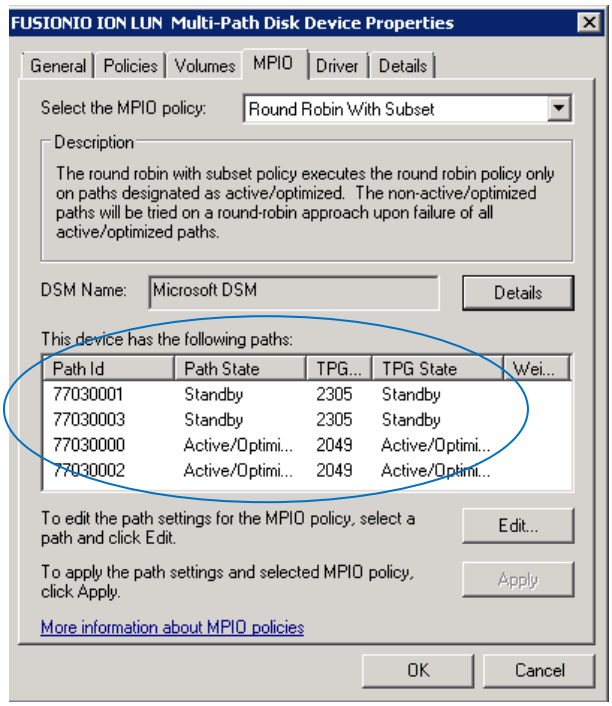
Favorite Targets Tab



3. To view the LUN discovery in Server Manager, open the Disk Management volume list. Disk 1 in the example below shows the discovered volume formatted and partitioned as drive E:



4. To see the discovered LUN as an MPIO multi-path configuration disk device, with two active and two standby paths, right-click Disk 1 and select **Properties**.



Creating Windows Initiators in ION Accelerator

1. From ION Accelerator (standalone or HA), create an initiator group in the CLI:

```
inigroup:create W2K8

Id  W2K8
Parent
Initiators  []
UUID  79f7b844-5bbf-11e3-acce-0015178fbc10
```

2. For each target, create an initiator. The examples below are for four iSCSI targets in ION Accelerator HA:

```
initiator:create -a W2K8 iqn.1991-05.com.sqa.microsoft:win-
ru3kmomq0ov#192.168.10.8 win_1

id  win_1
UUID  iqn.1991-05.com.sqa.microsoft:win-ru3kmomq0ov#192.168.10.8
Protocol  iSCSI
Discovered  false
Initiator Group  79f7b844-5bbf-11e3-acce-0015178fbc10
```

```
initiator:create -a W2K8 iqn.1991-05.com.sqa.microsoft:win-
ru3kmomq0ov#192.168.11.8 win_2

id  win_2
UUID  iqn.1991-05.com.sqa.microsoft:win-ru3kmomq0ov#192.168.11.8
Protocol  iSCSI
Discovered  false
Initiator Group  79f7b844-5bbf-11e3-acce-0015178fbc10

initiator:create -a W2K8 iqn.1991-05.com.sqa.microsoft:win-
ru3kmomq0ov#192.168.10.9 win_3

id  win_3
UUID  iqn.1991-05.com.sqa.microsoft:win-ru3kmomq0ov#192.168.10.9
Protocol  iSCSI
Discovered  false
Initiator Group  79f7b844-5bbf-11e3-acce-0015178fbc10

admin@ionr8i48/> initiator:create -a W2K8 iqn.1991-
05.com.sqa.microsoft:win-ru3kmomq0ov#192.168.10.9 win_4

id  win_4
UUID  iqn.1991-05.com.sqa.microsoft:win-ru3kmomq0ov#192.168.11.9
Protocol  iSCSI
Discovered  false
Initiator Group  79f7b844-5bbf-11e3-acce-0015178fbc10
```

Configuring Multipathing on Linux

INSTALLING MPIO ON LINUX

The basic components for Linux multipathing are described below.

- `dm_multipath` kernel module – Reroutes I/O, supports failover for paths & path groups.
- `mpathconf` utility – Configures and enables device mapper multipathing
- `multipath` command – Lists & configures multipath devices. Start with `/etc/rc.sysinit`, or `udev` when a block device is added.
- `multipathd` daemon – Monitors paths; as paths fail and come back, it may initiate path group switches. Provides for interactive changes to multipath devices. This must be restarted for any changes to the `/etc/multipath.conf` file
- `kpartx` command - Creates device mapper devices for the partitions on a device. Use this command for DOS-based partitions with DM-MP. The `kpartx` is provided in its own package, but the device mapper-multipath package depends on it.

The basic steps for installing DM Multipath are listed here:

1. Install `device-mapper-multipath RPM`.
2. On OL 6.4 or RHEL, run `mpathconf --enable` to enable multipathing.
3. If necessary, edit the `multipath.conf` configuration file to modify default values and save the updated file.
4. Start the multipath daemon.

Common mpathconf Settings

Here are some common multipathing configuration settings:

```
mpathconf --find_multipaths y
mpathconf --with_module y
mpathconf --with_chkconfig
```

Installing Multipathing in HA Mode

To install multipathing in an ION Accelerator HA environment, use the `multipath.conf` file as explained below.


These instructions apply to OL 6.4 and RHEL 6.4+.

1. Follow the instructions on the first line of the code about copying the file.
2. Follow the "uncomment" instructions near the end of the file.
3. Restart `multipathd` as indicated on the first line.

```
# copy this file to /etc/multipath.conf and restart multipathd

defaults {
   user_friendly_names     yes
   queue_without_daemon    no
}
devices {
   device {
           vendor                  "FUSIONIO"
           features    "3 queue_if_no_path pg_init_retries 50"
           hardware_handler        "1 alua"
           path_grouping_policy    group_by_prio
           path_selector           "queue-length 0"
           failback                immediate
           path_checker            tur
           prio                    alua
      # Uncomment if using FC. Do not use for SRP and iSCSI
      #fast_io_fail_tmo       15
      #dev_loss_tmo           60
   }
}
```

Installing Multipathing with Host-Based Mirroring (RAID/LVM/ASM) in Standalone Mode

To install multipathing with RAID/LVM/ASM (host-based mirroring) in a standalone ION Accelerator environment, use the `multipath.conf` file as explained below.


These instructions apply to OL 6.4, RHEL 6.4+, and SLES 11.

1. Follow the instructions on the first line about copying the file.
2. Follow the "uncomment" instructions at the end of the file.

3. Restart `multipathd` as indicated on the first line.

```
# copy this file to /etc/multipath.conf and restart multipathd

defaults {
   user_friendly_names     yes
   queue_without_daemon    no
}

devices {
   device {
           vendor                  "FUSIONIO"
           features                "0"
           hardware_handler        "1 alua"
           path_grouping_policy    group_by_prio
           path_selector           "queue-length 0"
           failback                immediate
           path_checker            tur
           prio                    alua
           no_path_retry           3

           # Uncomment if using FC. Do not use for SRP and iSCSI
           #fast_io_fail_tmo       15
           #dev_loss_tmo           60
   }
}
```

Restarting Multipathing

If your multipath configuration changes, you may need to restart multipathing, as in this example:

```
# multipathd -k
multipathd> reconfigure
ok
multipathd> show config
multipathd> exit
```

When using the show config command, you will need to search through the output and make sure it matches what was entered in `/etc/multipath.conf`.

If the restart is not successful, you can try the following commands:

```
# multipath -F
# service multipathd stop
# service multipathd start
// confirm the new settings are used for existing multipath devices
# multipath –ll
```

(See *Creating RHEL Initiators in ION Accelerator* for sample `multipath -ll` output.)

Setting the Node Session Timeout for iSCSI

If you are using the iSCSI protocol, you need to set the `node.session.timeo.replacement_timeout` value. This should be done for both HA and standalone systems.

 This should not be done while applications are using the iSCSI or dm-multipath devices.

The `node.session.timeo.replacement_timeout` setting controls the amount of time a path is tried before it is failed. When using `dm-multipath`, a value of 15 seconds is safe. If your applications require faster failovers, you may need to set the value lower, such as to 5 seconds.

To set this value so it can be used for both current and new sessions,

1. Edit the /etc/iscsid.conf file and set an appropriate value for `node.session.timeo.replacement_timeout`.
2. Log out of current sessions by running `iscsiadm -m session –u`.
3. Set the replacement timeout for currently discovered portals, as follows:

   ```
   iscsiadm -m node -o update -n node.session.timeo.replacement_timeout -
   v 15
   ```

4. Log in to the targets again:

   ```
   iscsiadm -m node -T target -p IP -l
   ```

LINUX ISCSI INITIATOR NOTES

Below are some tips on using Linux iSCSI initiators with ION Accelerator.

- The `SCSI H:C:I:L` address assigned by a Linux initiator system consists of four components: Host, Channel, Target ID and LUN. Of these components, only the LUN number is assigned by the ION Accelerator system. The other three components are assigned by the initiator system.
- An HBA reset can cause the first three components of an `H:C:I:L` address to change.
- An HBA reset can cause a new `/dev/sd*` device node to be assigned to a path.
- A LUN rescan initiated by the `rescan-scsi-bus.sh` script changes neither the `H:C:I:L` iSCSI address nor the `/dev/sd*` device node assigned to a path.

HA Mode

1. Copy the following file to `/etc/multipath.conf`, or copy the settings from this file to the existing `multipath.conf` file.

2. Follow the "uncomment" instructions near the end of the file.

```
# copy this file to /etc/multipath.conf and restart multipathd

defaults {
   user_friendly_names     yes
   queue_without_daemon    no
}
devices {
   device {
           vendor                  "FUSIONIO"
           features    "3 queue_if_no_path pg_init_retries 50"
           hardware_handler        "1 alua"
           path_grouping_policy    group_by_prio
           path_selector           "queue-length 0"
           failback                immediate
           path_checker            tur
           prio                    alua
      # Uncomment next 2 lines for FC; do not use for SRP or iSCSI.
      #fast_io_fail_tmo       15
      #dev_loss_tmo           60
   }
}
```

3. Run `service multipathd restart`.

Standalone Mode

1. Copy the following file to `/etc/multipath.conf`, or copy the settings from this file to the existing `multipath.conf` file.

```
# copy this file to /etc/multipath.conf and restart multipathd

defaults {
   user_friendly_names      yes
   queue_without_daemon     no
}

devices {
   device {
         vendor             "FUSIONIO"
         features           "0"
         path_grouping_policy     multibus
         path_selector            "queue-length 0"
         failback           immediate
         path_checker       tur
         no_path_retry            3
```

```
                # Uncomment next 2 lines for FC; do not use for SRP or iSCSI.
                # fast_io_fail_tmo       15
                # dev_loss_tmo           60
        }
    }
```

2. Run `service multipathd restart`.

SETTING UP LINUX ISCSI INITIATORS

For the example used in this configuration, the following items should be noted:

- The Red Hat Linux initiator has an iSCSI card that uses IP addresses 192.168.10.119 and 192.168.11.119.
- The target is ION Accelerator in HA mode. Each node has a card installed and configured for iSCSI. The IP addresses are 192.168.10.8 and192.168.11.8 for the first node, and 192.168.10.9 and 192.168.11.9 for the second node.
- An ION Accelerator volume of 100GB is used, with LUNs already created and presented to the Linux initiator.


Each initiator iSCSI port should be in a separate subnet that is the same as that of the target iSCSI port.

Discovering and Accessing an ION Accelerator Volume

By running the `targets` command in the CLI, you can view the targets as seen from the HA side of an ION Accelerator system.

```
targets -dt --cluster
ionr8i48:
Id  |UUID                    |Protocol|Enabled|State
-----------------------------------------------------
eth4|iqn.2007-02.com.fusioni\|iSCSI   |true   |Online |o:sn.2m232406fv:eth4 |   |   |
eth5|iqn.2007-02.com.fusioni\|iSCSI   |true   |Online |o:sn.2m232406fv:eth5 |   |   |
ionr8i49:
Id  |UUID                    |Protocol|Enabled|State
-----------------------------------------------------
eth6|iqn.2007-02.com.fusioni\|iSCSI   |true   |Online |o:sn.2m232406fw:eth6 |   |   |
eth7|iqn.2007-02.com.fusioni\|iSCSI   |true   |Online |o:sn.2m232406fw:eth7 |   |   |
```

Discovering ION Target Portals from the RHEL Initiator

To see all the target portals that are discovered, you can run the iscsiadm command from the console:

```
# iscsiadm -m discovery -t st -p 192.168.10.8
192.168.10.8:3260,1 iqn.2007-02.com.fusionio:sn.2m232406fv:eth4
# iscsiadm -m discovery -t st -p 192.168.10.9
1  92.168.10.9:3260,1 iqn.2007-02.com.fusionio:sn.2m232406fw:eth6
# iscsiadm -m discovery -t st -p 192.168.11.8
192.168.11.8:3260,1 iqn.2007-02.com.fusionio:sn.2m232406fv:eth5
# iscsiadm -m discovery -t st -p 192.168.11.9
192.168.11.9:3260,1 iqn.2007-02.com.fusionio:sn.2m232406fw:eth7


# iscsiadm -m node -P1
Target: iqn.2007-02.com.fusionio:sn.2m232406fw:eth6
        Portal: 192.168.10.9:3260,1
                Iface Name: default
Target: iqn.2007-02.com.fusionio:sn.2m232406fw:eth7
        Portal: 192.168.11.9:3260,1
                Iface Name: default
Target: iqn.2007-02.com.fusionio:sn.2m232406fv:eth5
        Portal: 192.168.11.8:3260,1
                Iface Name: default
Target: iqn.2007-02.com.fusionio:sn.2m232406fv:eth4
        Portal: 192.168.10.8:3260,1
                Iface Name: default
```

The following example logs in to each target portal:

```
# iscsiadm -m node -T iqn.2007-02.com.fusionio:sn.2m232406fw:eth6 -p
192.168.10.9  -l
# iscsiadm -m node -T iqn.2007-02.com.fusionio:sn.2m232406fw:eth7 -p
192.168.11.9  -l
# iscsiadm -m node -T iqn.2007-02.com.fusionio:sn.2m232406fv:eth5 -p
192.168.11.8  -l
# iscsiadm -m node -T iqn.2007-02.com.fusionio:sn.2m232406fv:eth4 -p
192.168.10.8  -l
# iscsiadm -m session
tcp: [5] 192.168.10.9:3260,1 iqn.2007-02.com.fusionio:sn.2m232406fw:eth6
tcp: [6] 192.168.11.9:3260,1 iqn.2007-02.com.fusionio:sn.2m232406fw:eth7
tcp: [7] 192.168.11.8:3260,1 iqn.2007-02.com.fusionio:sn.2m232406fv:eth5
tcp: [8] 192.168.10.8:3260,1 iqn.2007-02.com.fusionio:sn.2m232406fv:eth4
```

Creating RHEL Initiators in ION Accelerator

1. From ION Accelerator (standalone or HA), create an initiator group in the CLI:

```
inigroup:create RHEL

Id  RHEL
Parent
Initiators  []
UUID  29f1007e-1751-11e3-8482-0015178fbc10
```

2. For each target, create an initiator. The examples below are for four iSCSI targets from ION Accelerator HA:

```
initiator:create -a RHEL iqn.1994-
05.com.redhat:595862111da4#192.168.10.8 rhel_1

id  rhel_1
UUID  iqn.1994-05.com.redhat:595862111da4#192.168.10.8
Protocol  iSCSI
Discovered  false
Initiator Group  29f1007e-1751-11e3-8482-0015178fbc10

initiator:create -a RHEL iqn.1994-
05.com.redhat:595862111da4#192.168.11.8 rhel_2

id  rhel_2
UUID  iqn.1994-05.com.redhat:595862111da4#192.168.11.8
Protocol  iSCSI
Discovered  false
Initiator Group  29f1007e-1751-11e3-8482-0015178fbc10

initiator:create -a RHEL iqn.1994-
05.com.redhat:595862111da4#192.168.10.9 rhel_3

id  rhel_3
UUID  iqn.1994-05.com.redhat:595862111da4#192.168.10.9
Protocol  iSCSI
Discovered  false
Initiator Group  29f1007e-1751-11e3-8482-0015178fbc10

initiator:create -a RHEL iqn.1994-
05.com.redhat:595862111da4#192.168.11.9 rhel_4

id  rhel_4
UUID  iqn.1994-05.com.redhat:595862111da4#192.168.11.9
Protocol  iSCSI
Discovered  false
```

```
      Initiator Group  29f1007e-1751-11e3-8482-0015178fbc10
```

The Fusion ION LUN is now discovered with a MPIO multi-path configuration disk device, with two active and two enabled paths:

```
# multipath -ll
mpathag (26538623635336336) dm-3 FUSIONIO,ION LUN
size=93G features='3 queue_if_no_path pg_init_retries 50' hwhandler='1 alua'
wp=rw
|-+- policy='queue-length 0' prio=130 status=active
| |- 11:0:0:0 sdd 8:48 active ready running
| `- 12:0:0:0 sde 8:64 active ready running
`-+- policy='queue-length 0' prio=1 status=enabled
  |- 9:0:0:0  sdb 8:16 active ready running
  `- 10:0:0:0 sdc 8:32 active ready running
```

Performance Tuning and Filesystems for Linux Initiators

The `iontuner` utility does performance tuning for Linux-based initiators used with ION Accelerator. It is available for download at http://www.fusionio.com/files/ion-optimization-scripts.

Here are a few of its features:

- Balances an initiator's interface IRQ load between CPUs and NUMA nodes
- Tunes block device parameters for ION block devices
- Warns the user of power settings that may negatively impact performance
- Supports RHEL6.x/OL6.x and SLES11
- Has been tested on the Intel Nehalem / Westmere and Intel Sandy Bridge / Ivy Bridge architectures

INSTALLATION

To install `iontuner.rpm` , run the following commands:

```
rpm -ivh iontuner*.rpm

reboot
```

When upgrading from releases prior to 1.1.0, remove the old package and install the new one, rather than upgrading (use `rpm -e, rpm -i` rather than `rpm –U`).

PERFORMANCE EXPECTATIONS

`iontuner` improves performance only when the bottleneck is initiator storage performance. Configurations that are application-limited, fabric-limited, or target-limited will see little benefit from `iontuner`.

`iontuner` is primarily aimed at improving performance on NUMA initiators, but it can improve performance on any Linux initiator. The impact of `iontuner` will scale with the number of NUMA nodes in the initiator system.

For best use with NUMA initiators, each storage volume should be exported only to initiator ports on one of the initiator's NUMA nodes. When multiple volumes are used, access should be distributed between the NUMA nodes such that ports on all NUMA nodes are in use, but a single volume is accessed only through one NUMA node. This reduces inter-node data traffic and improves the effectiveness of any storage tuning.

MOUNTING A FILESYSTEM ON A LUN DURING LINUX BOOT TIME

In order to mount a filesystem on top of an ION Accelerator LUN, during Linux system boot time, follow these steps:

1. Determine the iSCSI ID of the ION Accelerator LUN. This is the number shown between parentheses in the output of `multipath -l` that starts with the digit "2" (indicated by the arrows below). For example:

```
# multipath -l
→ 26164613638323832 dm-0 FUSIONIO,ION LUN
size=34G features='3 queue_if_no_path pg_init_retries 50'
hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=0 status=active
`- 17:0:0:0 sdc 8:32 active undef running
→ 23630313437393135 dm-1 FUSIONIO,ION LUN
size=31G features='3 queue_if_no_path pg_init_retries 50'
hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=0 status=active
`- 17:0:0:1 sdd 8:48 active undef running
```

2. Add an entry in `/etc/fstab` for that LUN and use the `noauto` option so the entry is skipped during boot time. For example:

```
/dev/disk/by-id/dm-uuid-mpath-23165353937623137 /mnt auto noauto
0 0
```

3. Add a script in `/etc/init.d` that waits until `multipathd` has detected the LUN and then mounts the filesystem. See *Script for Mounting the Filesystem* below.

4. Make the system run during startup and shutdown. For example:

```
chkconfig mount-ion-filesystems on
```

Script for Mounting the Filesystem

```
#!/bin/sh
### BEGIN INIT INFO
# Provides: mount-ion-filesystems
# Required-Start: multipathd
# Required-Stop: multipathd
# Default-Start: 2 3 4 5
# Default-Stop: 0 1 6
# Description: Mount ION File Systems
### END INIT INFO
### BEGIN CHKCONFIG INFO
# chkconfig: 2345 13 87
# description: Mount ION File Systems
### END CHKCONFIG INFO

timeout=120

usage() {
echo
echo "Usage: `basename $0` {start|stop|restart|status}"
echo
return 2
}

is_ion_lun() {
[ "${1#/dev/disk/by-id/dm-uuid-mpath-}" != "$1" ]
}

luns_exist() {
grep -v '^#' /etc/fstab |
while read dev mountpoint vfstype options freq passno; do
is_ion_lun "$dev" && [ ! -e "$dev" ] && return 1
done
return 0
}

mount_luns() {
grep -v '^#' /etc/fstab |
while read dev mountpoint vfstype options freq passno; do
is_ion_lun "$dev" && fsck -n "$dev" && mount "$dev"
done
}

unmount_luns() {
```

```
grep -v '^#' /etc/fstab |
while read dev mountpoint vfstype options freq passno; do
is_ion_lun "$dev" && umount "$dev"
done
}

start() {
for i in $(seq $timeout); do
luns_exist && break
sleep 1
done
mount_luns
}

stop() {
unmount_luns
}

case "$1" in
start) start;;
stop) stop;;
restart) stop; start;;
status) ;;
*) usage;;
esac
```

Troubleshooting Linux Issues

LOGIN FAILURE WITH RHEL INITIATORS

Problem

When using RHEL's `qlogic qla2xxx` driver, the initiator may fail to log in to the target. When this happens, disc failure messages will appear in `/var/log/messages`.

Solution

Use the `qla2xxx` module parameter on the initiator side:

```
modprobe qla2xxx ql2xasynclogin=0
```

To have this setting used automatically, create the `/etc/modprobe.d/qla2xxx.conf` file and add the following line:

```
options qla2xxx ql2xmaxqdepth=2
```

If `qla2xxx` is loaded during boot time, then run `mkinitrd` (or `dracut -f` on RHEL 6.x) to pick up the changes.

ORACLE LINUX 6.3 ISSUES

Problem

When running Oracle Linux 6.3, OpenSM and srptools may encounter problems. The following error may occur:

```
# srp_daemon -oaev >/dev/null
26/07/13 12:49:01 : umad_open_port failed for device mlx4_0 port 1
```

The reason is that neither the OpenSM package nor the `srptools` package works on an unmodified Oracle Linux 6.3 system.

Solution

To resolve this issue,

1. Add the following code to the end of the `/lib/udev/rules.d/50-udev-default.rules` file:

```
# InfiniBand
KERNEL=="umad*", NAME="infiniband/%k"
KERNEL=="issm*", NAME="infiniband/%k"
KERNEL=="uverbs*", NAME="infiniband/%k"
KERNEL=="ucm*", NAME="infiniband/%k"
KERNEL=="uat", NAME="infiniband/%k"
KERNEL=="ucma", NAME="infiniband/%k"
KERNEL=="rdma_cm", NAME="infiniband/%k"
```

2. Run the following command to restart `rdma`:

```
# /etc/init.d/rdma restart
```

3. Wait for a few seconds after `rdma` has restarted, and then run the following commands to verify that the changes have solved the issue:

```
# srp_daemon -oac
# opensm
```

   OpenSM will now start as expected, as will SRPTools.

LINUX CLEANUP AFTER LUN REMOVAL

Linux does not automatically clean up after ION Accelerator LUN removals. For RHEL and OL, the required steps for LUN removal are described here:

https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/removing_devices.html

These instructions also work for SLES.

For all distros, *before* running `multipath –f`, run the following command:

```
dmsetup message mpathXYZ 0 "fail_if_no_path"
```

where `mpathXYZ` is the name of the device.

SOLVING NETWORK MANAGER AND UDEV ISSUES WITH RHEL 6.4

Problem

When using the Mellanox ConnectX®-3 driver, `mlx4_en`, in RHEL, the Network Manager and `udev` tools will incorrectly reset and set up the network devices. This results in the network devices being renamed or incorrectly shut down, and possibly incapable of receiving frames.

Solution

Network Manager must be disabled, and one of the `udev` rules must be modified, as explained below.

1. Follow the steps in this Red Hat link to stop the Network Manager service: https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux_OpenStack_Platform/3/html/Installation_and_Configuration_Guide/Disabling_Network_Manager.html)

2. Ensure that the Network Manager service is stopped by using the `service` command:

   ```
   service NetworkManager stop
   ```

4. Ensure that the Network Manager service is disabled at startup by using the `chkconfig` command:

   ```
   chkconfig NetworkManager off
   ```

5. Open each interface configuration file on the system in a text editor. Interface configuration files are found in the `/etc/sysconfig/network-scripts/` directory. They have names of the form "`ifcfg-X`", where X is replaced by the name of the interface. Valid interface names include `eth0`, `p1p5`, and `em1`.

6. In each configuration file, ensure that the `NM_CONTROLLED` configuration key is set to `no` and the `ON_BOOT` configuration key is set to `yes`.

   ```
   NM_CONTROLLED=no
   ONBOOT=yes
   ```

5. Ensure that the network service is started by using the `service` command:

   ```
   service network start
   ```

6. Ensure that the network service is enabled at startup by using the `chkconfig` command:

   ```
   chkconfig network on
   ```

7. To prevent `udev` from creating duplicate `eth-X` entries in `/etc/udev/rules.d/70-persistent-net.rules`, modify the following `udev` rule:

In:
`/lib/udev/rules.d/75-persistent-net-generator.rules`

by adding this line:

```
DRIVERS=="mlx4_core", ENV{MATCHDEVID}="", ENV{MATCHIFTYPE}=""
```

When `udev` is updated, this line must be re-added.

HANDLING CONFIGURATION CHANGES

After volumes have been added, deleted, or resized, it is important to rescan LUNs on each affected initiator system. The most convenient way to rescan LUNs is by running the `rescan-scsi-bus.sh` script. For example:

```
# rescan-scsi-bus.sh --forcerescan --hosts=34
Scanning SCSI subsystem for new devices
 and remove devices that have disappeared
Scanning host 34 for SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning for device 34 0 0 0 ...
OLD: Host: scsi34 Channel: 00 Id: 00 Lun: 00
      Vendor: FUSIONIO Model: ION LUN          Rev: 3243
      Type:   Direct-Access                    ANSI SCSI revision: 05
0 new device(s) found.
0 device(s) removed.
```

Failure to rescan LUNs after having reduced the size of a volume may lead to data loss.

Informing the Multipath Daemon of Resized Devices

After a multipath device is resized on Linux and a rescan is done, the multipath daemon needs to be informed of the change. To do this, run the following commands:

```
# multipathd -k
multipathd> resize
multipathd> map <your_multipath_device>
multipathd> exit
```

Tips for Resizing Devices with Multipathing

To resize a device, you should first resize it on the target/ION Accelerator side, and then on the initiator side.

For RHEL 6.X and OEL 6.X, follow the steps out lined in the Red Hat Storage guide:

https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/online-iscsi-resizing.html

Resolving Path Failures during Multipathing Scans

If a path fails while `multipathd` is scanning the path UID, then `multipathd` won't recognize that path until it is rescanned explicitly. To verify whether `multipathd` failed to add one or more paths, run the following command:

```
# echo show paths | multipathd -k | grep faulty
388:0:0:10 sdlt  68:432  1   undef  faulty running orphan
388:0:0:11 sdlu  68:448  1   undef  faulty running orphan
```

You can also force `multipathd` to reconsider faulty paths, as follows:

```
# echo reconfigure | multipathd -k
```

Part Four: Application Tuning

Best Practices for Oracle Database

This section outlines best practices for using ION Accelerator with Oracle Database and Oracle ASM.

ORACLE DATABASE IMPLEMENTATION BEST PRACTICES

- After configuring ION Accelerator, export the ION Accelerator volumes to the Oracle Database nodes (standalone or RAC).
- Install Oracle with the "database binary" (software-only) option.
- Run the Oracle synthetic benchmark tool Orion to measure IOPS and bandwidth on ION volumes with various workloads (OLTP, DSS, RAID). You can also use the FIO tool to benchmark these performance categories.
- Make sure the Orion/FIO performance stats come closer to the theoretically possible performance of the hardware setup (HBAs on the target, initiator, and ioDrive/ioScale).
- Install Oracle Grid Infrastructure. Configure the database and run the "read only" `ioCalibration` tool to determine the IOPS, MBPS, and latency of the setup.
- Record your performance stats. These will serve as the baseline for the setup if application performance issues are noted during workload tests.
- Configure NUMA and HugePages for Oracle.

ASM BEST PRACTICES

- ASM limits the LUN size to 2TB or less, so ION Accelerator volumes should be created no larger than that.
- An ASM diskgroup can be created using one or multiple ION Accelerator volumes. For better performance it's recommended to create fewer ASM diskgroups on more ION Accelerator volumes.

- Use the 512B sector size when exporting volumes to initiators. If you need to export LUNs with a 4KB sector size, use Oracle or third-party-provided ASMLIB support packages. ASMLIB support is now available for RHEL 6.x. For details, visit this link:

  http://www.oracle.com/technetwork/server-storage/linux/asmlib/rhel6-1940776.html

  You may also need to use the `_disk_sector_size_override` parameter to overcome an Oracle ASM diskgroup creation error.

- ASM distributes data on all the ASM disks in the disk group. A diskgroup can be created with an option for NORMAL REDUNDANCY (two-way mirroring), HIGH REDUNDANCY (three-way mirroring) or EXTERNAL REDUNDANCY (no mirroring). For ION Accelerator, no significant performance differences have been observed between these ASM mirroring options.

- It is generally advisable to use the default 1MB allocation unit (AU) size. For larger (3TB+) databases, the AU size can be increased to 4M or 8M.

- If more than 2 IONs used (without HA options), then by carefully creating ASM diskgroup mirroring options can avoid outages due to the rare scenario of ION node/component failures. For example, if you have 3 set of ION volumes from 3 different ION nodes and all part of a HIGH redundancy mirroring option, then losing 2 IONs completely would still make the application function without an issue (assuming the free space is enough for the database transactions).

EXPORTING LUNS TO INITIATORS

In RAC, it's recommended to have separate, small ION Accelerator volumes for Oracle cluster voting disks, with Normal or High redundancy ASM mirroring. Optionally, versions 11gR2 and 12c support voting disks as part of regular ASM diskgroups. Careful planning is needed here to avoid Oracle RAC node evictions.

For Oracle RAC configuration, it's recommended to have one initiator group for all participating Oracle initiator nodes. This will greatly decrease the time spent in LUN-mapping.

If you are planning to assign LUNs to various initiators, you should create the initiator groups individually.

Because LUN mappings include the target/initiator port addresses, it's strongly recommended to tag the cables and ports. This will help in reconnecting to the same ports if the direct-connected cables needed to be removed or exchanged.

ASM STORAGE CONFIGURATIONS

The configurations described below are variations that may serve your particular storage needs, using Oracle ASM and ION Accelerator.

HA Mode, One Diskgroup

One large ASM diskgroup can be created with External Redundancy, which means no mirroring is done by ASM. Here is the basic procedure:

1. Place all ION Accelerator storage from both the HA nodes into one ASM diskgroup called DATA.
2. Have Oracle write the OCR and Voting Disk files to DATA, which gets passed through to the underlying LUNs on ION Accelerator.
3. The HA implementation of ION Accelerator takes care of creating the redundant copies.

Standalone Mode

In standalone mode, Oracle ASM can be used with Normal Redundancy. The procedure would use the first two steps in *HA Mode, One Diskgroup* above.

A variation is to use ASM High Redundancy for the OCR and Voting files, and Normal Redundancy for the database files. To do this,

1. Create a 1GB partition on *n* LUNs from each ION Accelerator node (where *n* is 6 divided by the number of nodes).
2. Put those 6 partitions into an ASM diskgroup named "CRS" with High Redundancy.
3. Put all of the remaining storage into the DATA diskgroup with Normal Redundancy.

Standalone Mode: Storing OCR and Voting Files on ION Accelerator

When using ASM (not ION Accelerator HA) for redundancy, there is no requirement to store the OCR and Voting files in ASM. During Oracle Clusterware installation, the checkbox for this can be left clear so ASM won't be used for these two types of files.

The basic procedure is to create a 1GB partition on the first *n* LUNs from each ION until there are six total partitions (two LUNs from each of three ION Accelerator nodes, or one LUN from each of six nodes, etc.).

During installation, Oracle Clusterware prompts for the 6 devices on which Oracle will store the OCR and Voting Disk files. There are 6 OCR files and 5 Voting Disk files (the Voting Files are small extents stored in the headers of the OCR files, not separate files themselves).

It is very important to note that this count of files (6 OCR and 5 Voting) means one of the LUNs will not have a Voting File. It is critical that no single ION Accelerator node have a majority of the Voting Disk files, because if a majority of them go offline, Oracle aborts the entire cluster.

For example, in a two-node configuration there would be three voting files on the first node and two on the second node. Taking the second node offline for maintenance is not a problem, because the majority of Voting Files are still online. However, taking the first node offline causes the entire cluster to abort. The solution calls for a third storage node, which can be a third ION Accelerator node or a legacy SAN. With three storage nodes, two voting files are placed on the first, two on the second, and one on the third. Now, no single node contains a majority of voting files, so the cluster is not aborted.

Best Practices for MS SQL Server

This section outlines best practices for using ION Accelerator with MS SQL Server.

MS SQL SERVER IMPLEMENTATION BEST PRACTICES

- Configure ION Accelerator and export its volumes to the SQL DB nodes (standalone or cluster)
- Run the SQLIO synthetic benchmark tool to measure the IOPS and bandwidth for volumes, with various workloads.
- Make sure the SQLIO performance stats come closer to the theoretically possible performance of the hardware setup (HBAs on the target, initiator, and ioDrive/ioScale).
- Record your performance stats. These will serve as the baseline for the setup if application performance issues are noted during workload tests.

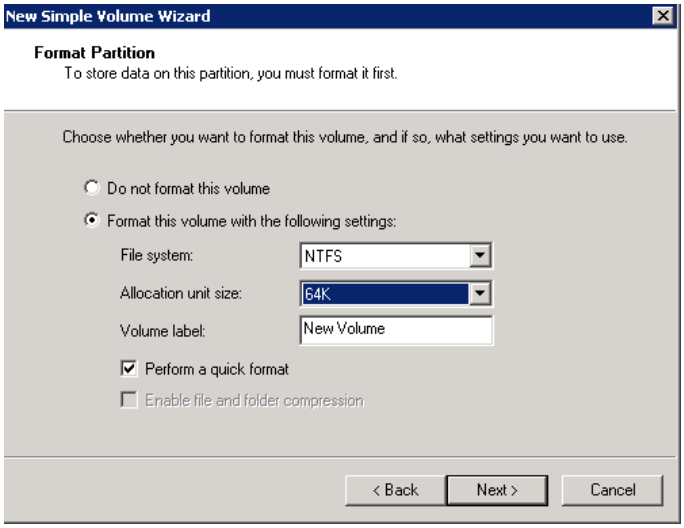
SQL SERVER MEMORY ALLOCATION

- Leave enough memory for the OS processes. As a best practice, deduct 1GB for every 4 cores and allocate the rest to the "Max server memory (MB)" parameter.
- Use the following dynamic management views to troubleshoot memory issues that may occur:
  - `sys.dm_os_memory_brokers` provides information about memory allocations using the internal SQL Server memory manager. The information provided can be useful in determining very large memory consumers.
  - `sys.dm_os_memory_nodes` and `sys.dm_os_memory_node_access_stats` provide summary information of the memory allocations per memory node and node access statistics, grouped by type of page. This information can be used to quickly obtain summary memory usage, instead of running `DBCC MEMORYSTATUS` .
  - `sys.dm_os_nodes` provides information about CPU node configuration for SQL Server. This DMV also reflects software NUMA (soft-NUMA) configuration.
  - `sys.dm_os_sys_memory` returns the system memory information. The "Available physical memory is low" value in the `system_memory_state_desc` column is a sign

of external memory pressure that requires further analysis.

OTHER BEST PRACTICES

- When formatting the partition that will be used for SQL server data files, you should use a 64 KB allocation unit size for data, logs and the TempDB database, as shown below:

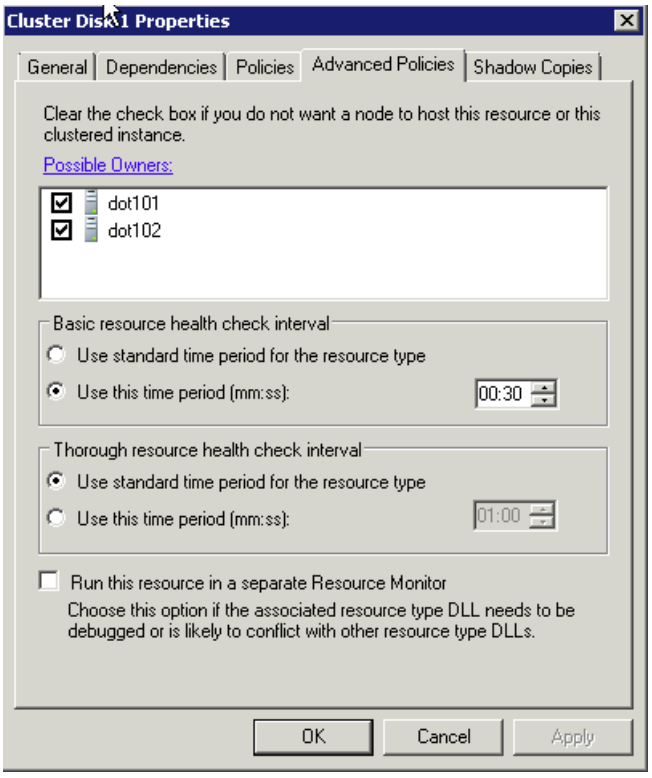


- Lock pages in memory. To reduce SQL server paging, grant the service account "Lock pages in Memory" privileges through the Windows Group Policy editor, for both 32- and 64-bit servers.
- Adjust the "Degree of Parallelism" option to the number of cores in a single NUMA node, which is eight or fewer. Test the workload with various degrees of parallelism to arrive at the best combination.
- The default setting for Number of Worker Threads works well for most configurations.
- To ensure efficient operation, create multiple TempDB files (1 per physical CPU core). Pre-size the TempDB files and create them in equal sizes.
- For ease of maintenance, create more than one log file (there is no performance gain in configuring more files).

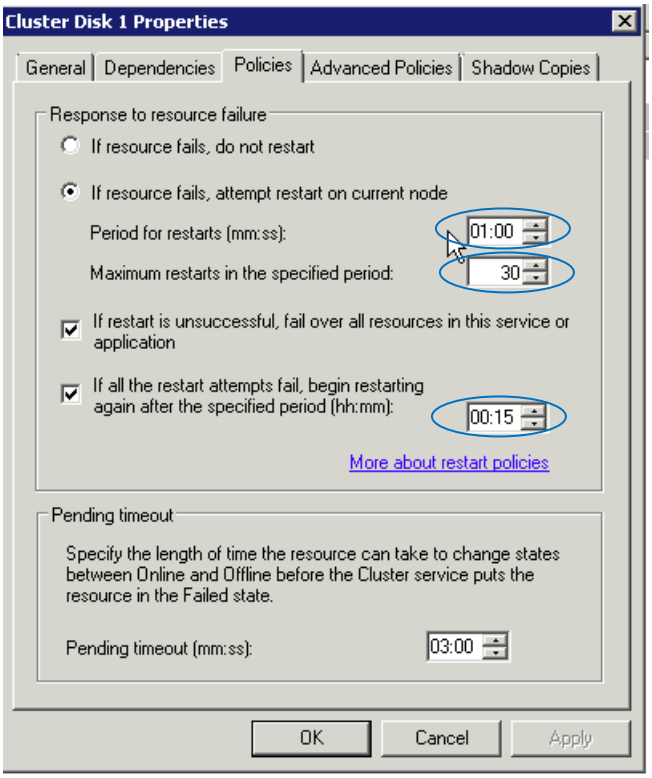
Configuring HA with Windows Server Failover Cluster (WSFC)

If you are deploying Windows Server 2008 with ION Accelerator appliances, follow the steps below to configure your storage to use Windows Server Failover Cluster. For MPIO settings, refer to *Configuring Multipathing* earlier in this guide.

1. In WSFC, navigate to **Storage**.
2. Click the **Advanced Policies** tab.
3. As shown below, set "Use this time period" to 30 seconds.



4. Click the **Policies** tab.

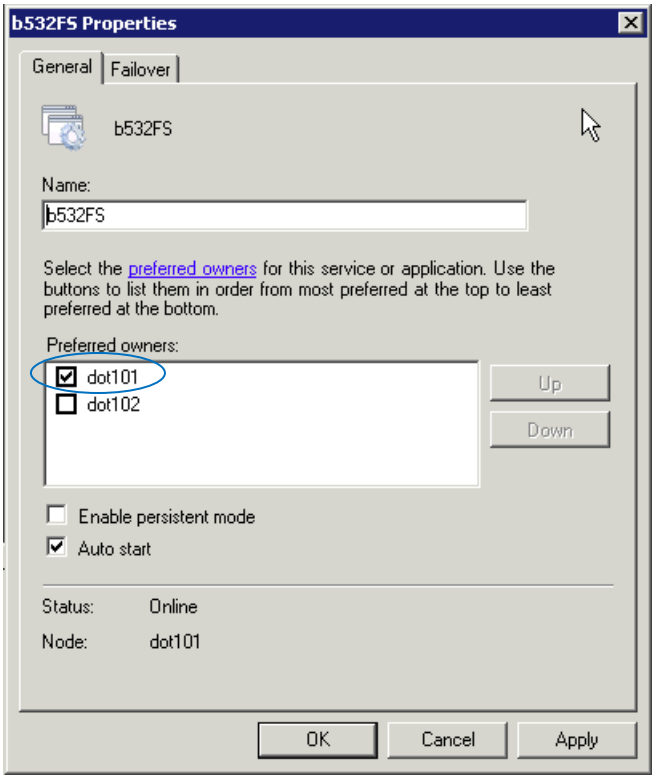


5. Set "Period for restarts" to 1:00.
6. Set "Maximum restarts in the specified period" to 30.
7. Set "If all restart attempts fail, begin restarting again after the specified period" to 15 minutes.

CONFIGURING AUTO-FAILBACK SETTINGS

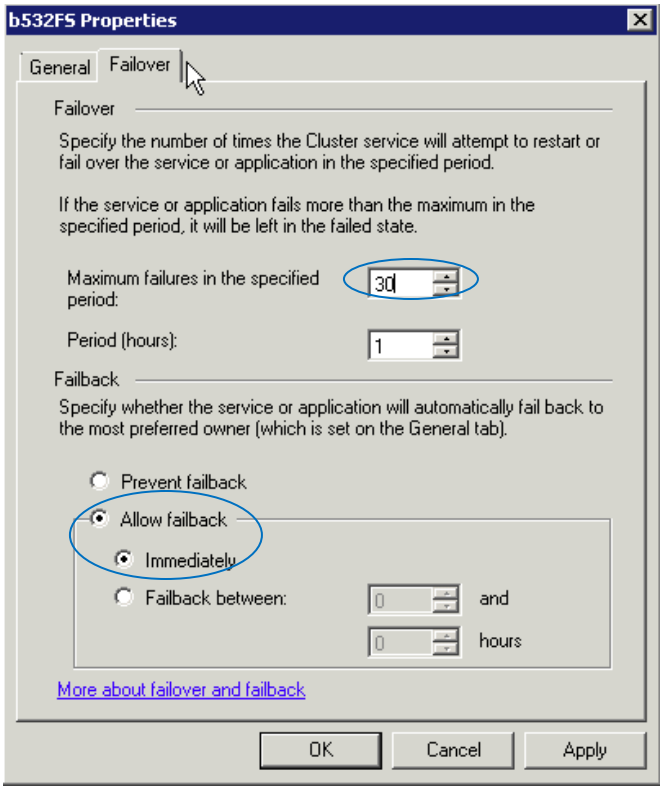
To specify auto-failback settings on a WSFC node for any service or application,

1. Verify that the preferred owner is selected in the General tab. In the example below, the preferred owner is "dot101".



2. In the Failover section of the tab, set the maximum number of failures as 30 in 1 hour.

3. In the Failback section, select "Allow failback: and "Immediately".



Windows Clustering Hot Fixes

Several hot fixes for Windows Clustering are available from Microsoft Support. These may be useful in preventing or resolving issues with Windows clusters used in an ION Accelerator system.

Hot Fix Reference	Description
http://support.microsoft.com/kb/2718576/en-us	This fixes an issue in which an MPIO disk is removed unexpectedly when a `PR_IN` command fails. This issue occurs when the Failover Clustering feature is configured in a Windows Server 2008 R2 environment.
http://support.microsoft.com/kb/2733575/en-us	SAN targets should reply to `PR_IN` commands within 3 seconds on Windows Server 2008 and Server 2008 R2.
http://support.microsoft.com/kb/2522766	The MPIO driver fails over all paths incorrectly when a transient, single failure occurs in Windows Server 2008 or in Windows Server 2008 R2.
http://support.microsoft.com/kb/2550886	A transient communication failure causes a Windows Server 2008 R2 failover cluster to stop working.

Appendixes

Appendix A: Dell PowerEdge R720 Configuration

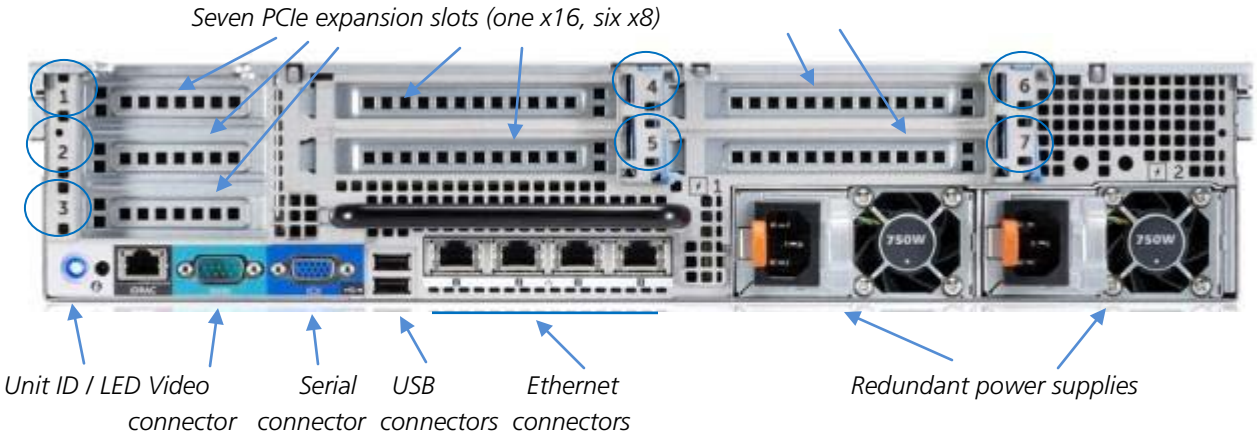
The pictures below show various views of the ION Accelerator appliance as a Dell PowerEdge R720 Server (2U, two-socket).

ION ACCELERATOR APPLIANCE FRONT PANEL, DELL R720

Front panel: Dell PowerEdge R720 server, with bezel

(Without bezel) *LCD control panel and video connector*

REAR PANEL

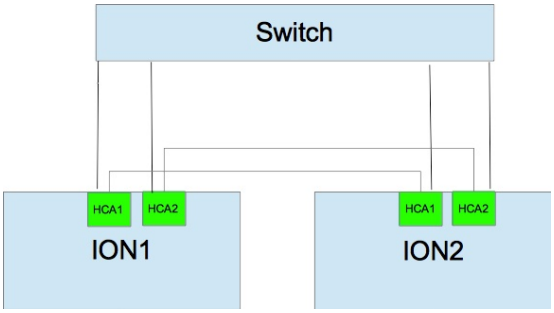


Low-Profile Slot Assignments

	Non-HA, FC	Non-HA, iSCSI	HA, FC	HA, iSCSI
Slot 1	empty	empty	Cluster Interconnect (both Eth ports)	Left port = iSCSI; Right port = Cluster Interconnect (Eth)
Slot 2	FC	iSCSI (Eth)	FC (both ports)	iSCSI (both ports)
Slot 3	FC	iSCSI (Eth)	FC (both ports)	Left port = iSCSI; Right port = Cluster Interconnect (Eth)

Split-Function Ports for iSCSI – Dell R720

For increased redundancy/high-availability, when using two Mellanox ConnectX-3 dual-port adapters for both connectivity between HA IONs and fabric connectivity to the initiator nodes (as when running iSCSI), it is best to split the functionality between the cards. This allows each card to be used for both a redundant connection to the partner node and to the fabric. That way, if one adapter fails, and both its ports are lost, the system can continue to run. The mode will be partially degraded, but not system-failed, with the remaining cluster interconnect and iSCSI ports on the other adapter.



It is important to use the correct ports for fabric and ION interconnectivity in this case. The diagram above, along with the rear-view picture of the ION chassis shown previously, describe the ports to be used to connect IONs, and the ports to be used when connecting to the fabric.

iSCSI Port Assignments

- PCIe Slot 1: left port = iSCSI, right port = cluster interconnect (Eth)
- PCIe Slot 2: both ports = iSCSI
- PCIe Slot 3: left port = iSCSI, right port = cluster interconnect (Eth)

ConnectX®-3 iSCSI card

Appendix B: About the ION Accelerator Appliance Guides

The *ION Accelerator Appliance Configuration Guide* provides an introduction to the ION Accelerator software, as well information on First Boot setup, host multipathing, and application tuning.

Other ION Accelerator guides include:

- *ION Accelerator GUI Guide* – explains how to use the ION Accelerator GUI to administer shared PCIe flash storage
- *ION Accelerator CLI Reference Guide* – an equivalent to the GUI guide, using a command-line interface for tasks