An Oracle Technical White Paper

May 2015

Microsoft Windows Server Multiprotocol Multipathing with the Oracle ZFS Storage Appliance

ORACLE

Table of Contents

Introduction

The Oracle ZFS Storage Appliance is formed by a combination of advanced hardware and software architecture with the purpose of offering a multiprotocol Storage Subsystem. The unified storage model enables you to simultaneously run a variety of application workloads while benefitting from advanced data services. The first-class performance characteristics of the Oracle ZFS Storage Appliance are illustrated by the results of industry standard benchmarks like SPC-1, SPC-2 and SPECsfs.

As a unified storage platform, the Oracle ZFS Storage Appliance offers the ability to access storage through multiple protocols, for both file and block.

This paper describes the ability to access a block storage volume through iSCSI and Fibre-Channel (FC) simultaneously in a Microsoft Windows Server 2008 R2 or later environment. This can be a very useful tool to provide a high-resilience architecture allowing high-bandwidth access through the relatively expensive (fiscally and administratively) FC connections falling back to the more traditional network-based iSCSI protocol to protect data availability.

Building High Availability into the Storage Network Infrastructure

High availability is or should be at the core of any storage network infrastructure design – whether it is FC or iSCSI. This design methodology should lead to redundant paths in the infrastructure design, which can effectively result in multiple mirrored sub-configurations. For example, in a standard resilient FC deployment, traditionally a dual fabric design is produced where the individual fabrics are duplicated.

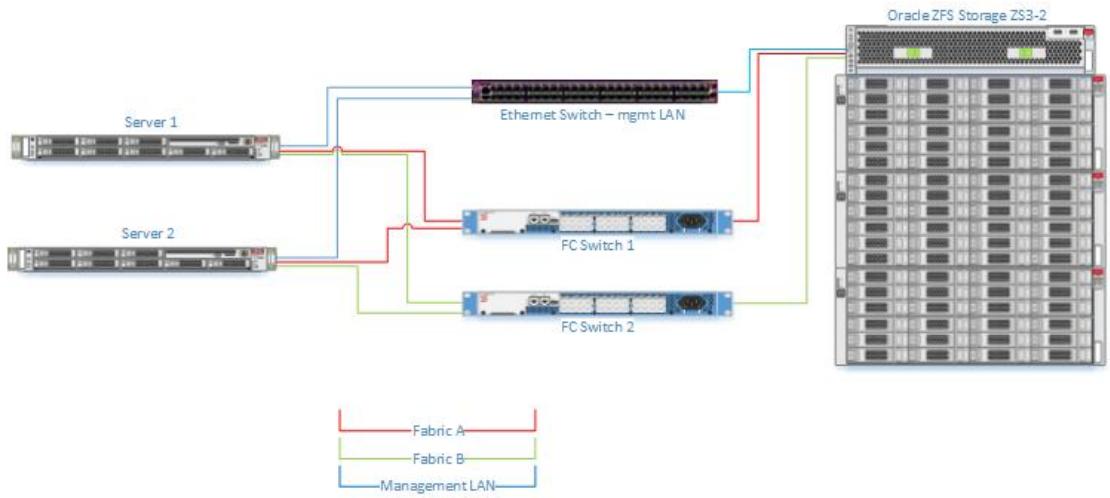

Figure 1. Example dual fabric FC SAN for high availability architecture

Similarly, any IP-based storage protocol should have a similar resilience built in with the server and storage straddling multiple subnets or VLANs.

Adding an Ethernet switch to provide TCP/IP connectivity between the servers and the Oracle ZFS Storage Appliance will allow the FC-presented volumes to also be presented through iSCSI to the server(s) when all the components are properly configured.

In the unlikely event of failure of all FC paths, the servers can be configured to use these additional iSCSI links. Given the unlikelihood of multiple failures causing loss of data access in a well designed SAN/power infrastructure, it may be desirable to deploy older, lower bandwidth network equipment for this task, but remember that these paths will carry all the traffic in the event of such a failure and bandwidth to data will be reduced to the level of the deployed network equipment for the period of the downtime.

If relatively low bandwidth equipment is deployed, the link(s) should be used in normal operation but only when absolutely necessary to preserve data access, albeit in a degraded fashion.

Updating the example shown in Figure 1, Figure 2 shows that the network infrastructure now includes an additional Gigabit Ethernet (GbE) switch to provide the network backup carrying iSCSI traffic.

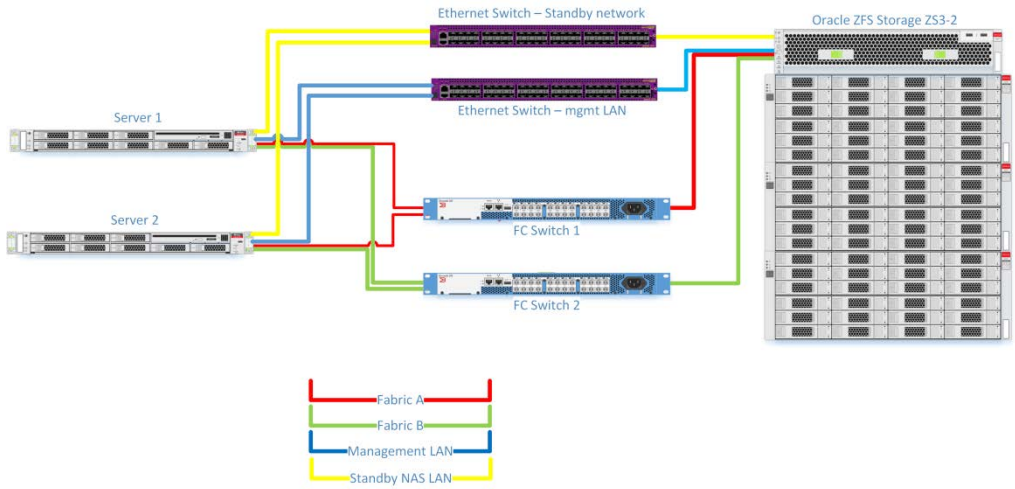

Figure 2. Dual fabric FC SAN with added GbE switch

Assumptions

For the purposes of this paper, it is assumed that the iSCSI Initiator service is not enabled but the Windows Multipathing (MPIO) service is (as would be the case where primary data access is over FC) and that the volumes being accessed are primarily presented over the FC interfaces and already configured using the Microsoft Device Specific Module (DSM).

The following sections describe how to enable the iSCSI Initiator service, assuming that the Windows Multipathing (MPIO) service is already enabled.

It is advisable to enable CHAP authentication but this is not covered in this document.

Windows Server iSCSI Initiator Service

The first step is to ensure that the Windows Server iSCSI Initiator service has been enabled, which can be done using the following steps:

Windows Environment Details

1. Log on to the Windows Server using an administrative account.
2. Navigate to the Control Panel.
3. Open the iSCSI control as shown in Figure 3.

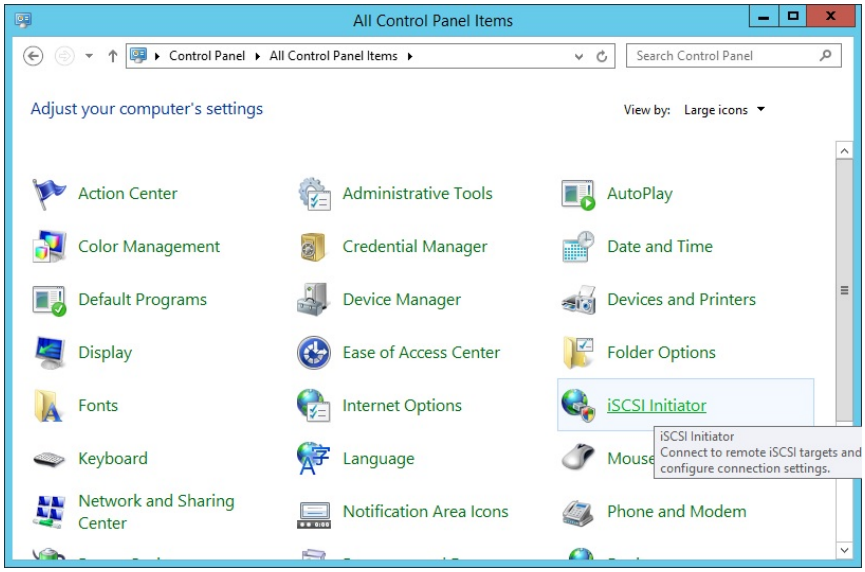

Figure 3. iSCSI control panel in Windows Server

4. If the iSCSI service has not been previously used, the following dialog box is shown. To maintain the iSCSI connection as the fallback, click the "**Yes**" button to ensure that the iSCSI initiator service is automatically started at every boot.

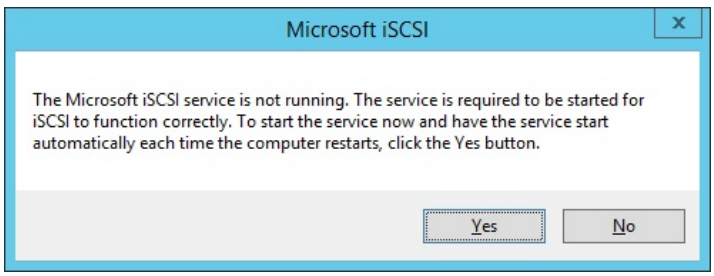

Figure 4. iSCSI autostart dialog window

The iSCSI Control panel will then be displayed as shown in Figure 5.

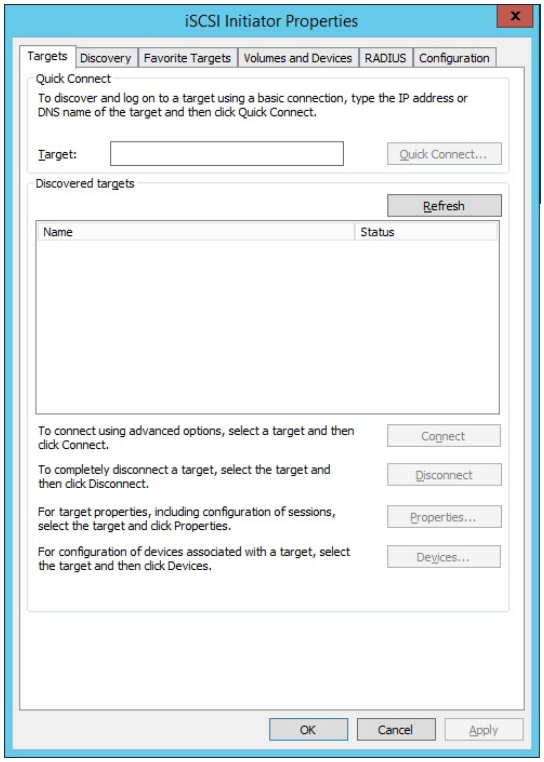

Figure 5. iSCSI Initiator Properties window

5. Select the **Configuration** tab. The Windows server IQN is shown as the Initiator Name. This IQN is used to assign exclusive access for a particular server or servers to a volume. It is analogous to LUN masking in FC SANs. Copy this IQN in order to configure the Oracle ZFS Storage Appliance initiator.

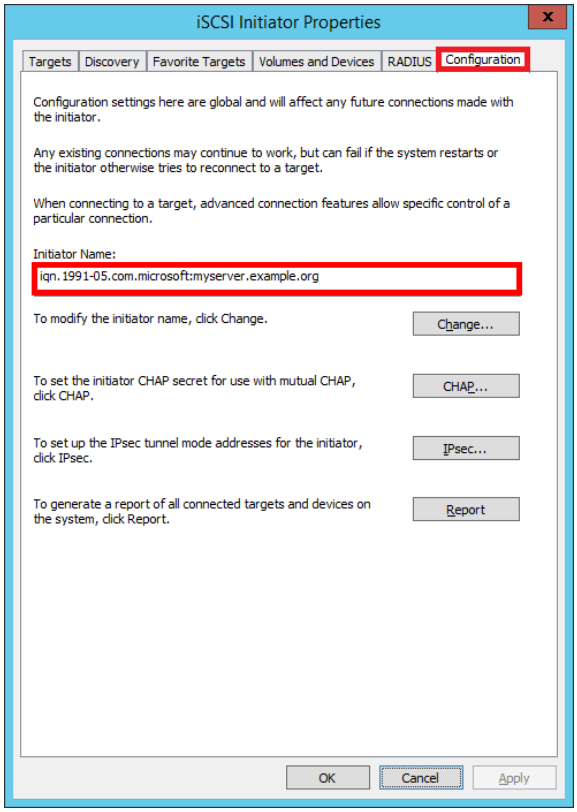

Figure 6. iSCSI initiator qualified name details

With this information in hand, you can now configure the Oracle ZFS Storage Appliance.

Configuring iSCSI on the Oracle ZFS Storage Appliance

The following outlined steps are basic instructions for configuring iSCSI on the Oracle ZFS Storage Appliance. A more comprehensive tutorial on Oracle ZFS Storage Appliance configuration for iSCSI (including CHAP authentication) is available at

http://www.oracle.com/technetwork/articles/servers-storage-admin/o11-129-iscsi-luns-windows-522255.html

1. Log on to the Oracle ZFS Storage Appliance in an Internet browser as a user with the following permissions assigned:

- stmf Configure
- nas.*.*.*: changeAccessProps, changeProtocolProps, changeGeneralProps

2. Enter an address in the URL field of a Web browser that includes the IP address or hostname of the Oracle ZFS Storage Appliance:

```
https://<ip-address or hostname>:215
```

The login dialog window shown in Figure 7 is displayed.

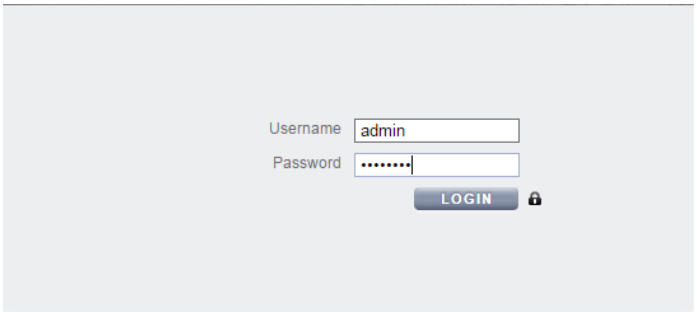

Figure 7. Oracle ZFS Storage Appliance login

3. Enter a **Username** and **Password** and click **LOGIN**.

Defining an iSCSI Target Group

A target group is created on the Oracle ZFS Storage Appliance to define the ports and the protocol by which the LUN to be presented to the Windows Server is accessed.

To define an iSCSI target group on the Oracle ZFS Storage Appliance, complete these steps:

1. Click **Configuration > SAN** to display the Storage Area Network (SAN) screen as shown in Figure 8.
2. Click **iSCSI** on the right hand side of the panel and **Targets** on the left hand side of the panel as shown in Figure 8.

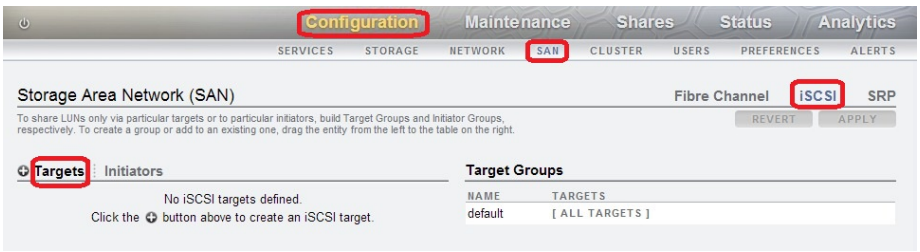

Figure 8. Configuring an iSCSI target in the Oracle ZFS Storage Appliance BUI

3. Click the **+** icon at the left of **Targets** to display the Create iSCSI Target dialog shown in Figure 9.
4. Fill in the displayed fields:
   - Enter an **Alias** for the target and select the **Initiator authentication mode**.
   - Additionally, choose the network interfaces over which iSCSI LUNs will be presented.

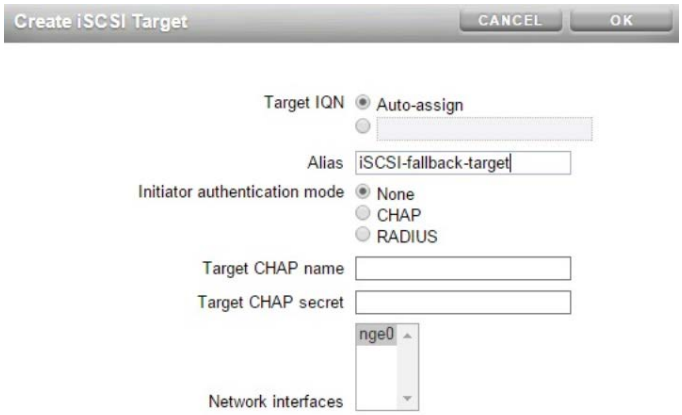

Figure 9. Create iSCSI Target dialog window

5. Click **OK** to confirm.

4. To create an iSCSI target group that includes the iSCSI target just defined, place the cursor over the new **iSCSI-Targets** entry in the left panel. The Move icon ✥ appears to the left of the entry as shown in Figure 10.

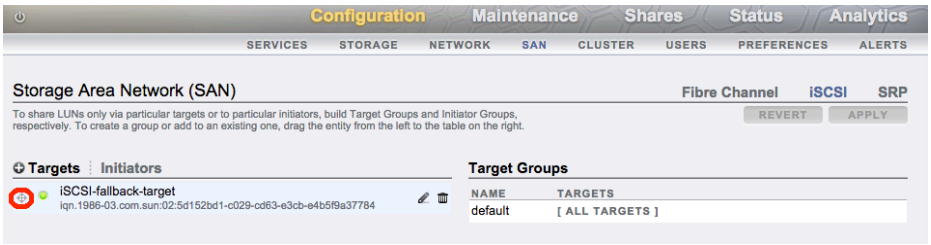

Figure 10. Selecting the iSCSI target entry

5. Click the ✥ icon and drag it to the iSCSI Target Groups panel on the right. A new entry (highlighted in yellow) appears at the bottom of the iSCSI Target Groups column as shown in Figure 11.

Figure 11. Creating an iSCSI target group

6. Move the cursor over the new target group and release the mouse button. A new iSCSI target group is created with a name `targets-`*n*, where *n* is an integer, as shown in Figure 12.

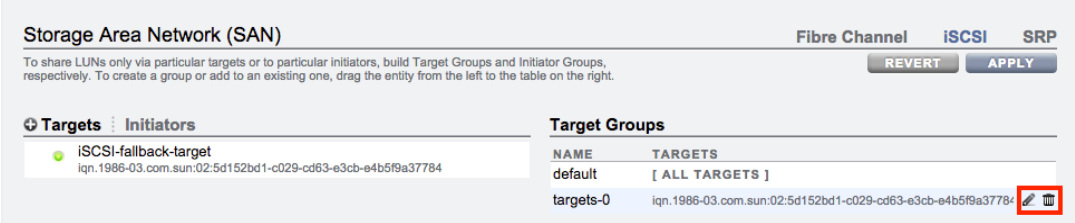

Figure 12. Selecting the new target group for editing

7. Click the edit icon (✎) to display the dialog seen in the preceding figure.
8. In the **Name** field, replace the default name with the name to be used for the iSCSI target group and click **OK**. For this example, the name `iscsi-FallbackGroup` is used.

Figure 13. Renaming the iSCSI target group

9. Click **APPLY**. The changes are shown in the iSCSI Targets panel on the left.

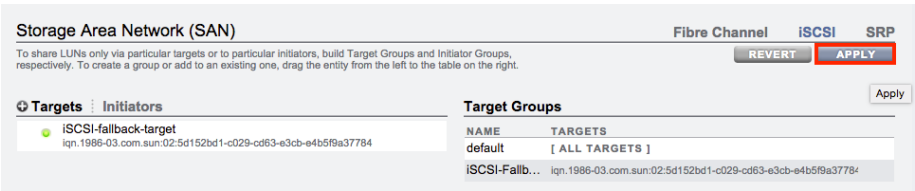

Figure 14. Viewing the new iSCSI target group

Defining an iSCSI Initiator

An iSCSI initiator is defined to restrict which servers have access to a particular volume. If more than one host can write to a given volume concurrently, inconsistency in file system caching between hosts can cause corruption in the on-disk image, unless there is a multi-concurrency-aware file system built on the volume. Typically, a single initiator is given access to a volume, unless a specialized cluster file system is being used.

The iSCSI initiator serves to define the "host" from the point of view of the Oracle ZFS Storage Appliance. To identify the Windows Server to the Oracle ZFS Storage Appliance, the Windows Server iSCSI initiator IQN must be registered with the Oracle ZFS Storage Appliance by completing the following steps:

1. Click **Configuration >SAN** to display the Storage Area Network (SAN) screen shown in the following figure.
2. Click the **iSCSI** tab at the right and then select **Initiators** at the top of the left panel.

Figure 15. Selecting the iSCSI Initiators panel

3. Click the + icon to the left of **Initiators** to display the Identify iSCSI Initiator dialog window seen in the following figure.
4. Enter the **Initiator IQN** for the Windows server (as determined in the previous section). Enter a more meaningful symbolic name as the **Alias**.

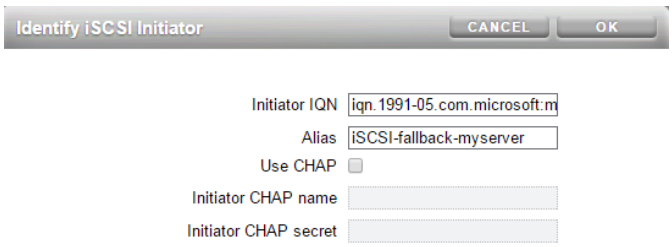

Figure 16. Configuring a new iSCSI initiator

5. Click **OK**.

Defining an iSCSI Initiator Group

Related iSCSI initiators can be combined into logical groups to allow single commands to be executed on multiple iSCSI initiators; for instance, assigning LUN access to all iSCSI initiators in a group with one command. For this example, the iSCSI initiator group will contain one initiator, but in a cluster, where multiple servers are treated as a single logical entity, the initiator group may contain multiple initiators.

To create an iSCSI initiator group, complete these steps:

1. Select **Configuration > SAN** to display the Storage Area Network (SAN) screen.

   Select the **iSCSI** tab at the right and then click **Initiators** at the top of the title panel.

   Place the cursor over the entry for the iSCSI initiator created in the previous section. The Move icon appears to the left of the entry as shown in Figure 17.

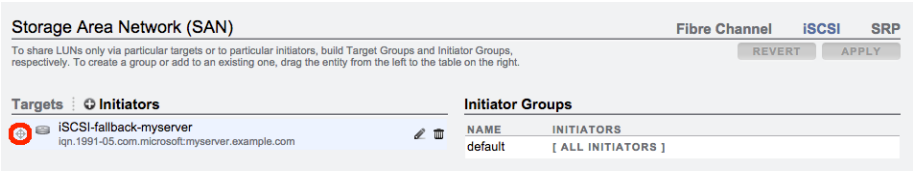

Figure 17. Displaying the Move icon for a new iSCSI initiator

2. Click the ✥ icon and drag it to the Initiator Groups panel on the right. A new entry (highlighted in yellow) appears at the bottom of the Initiators Groups panel as shown in Figure 18.

   Move the cursor over the new entry box and release the mouse button. A new iSCSI initiator group is created with a name `initiators-`*n*, where *n* is an integer, as shown in Figure 19.

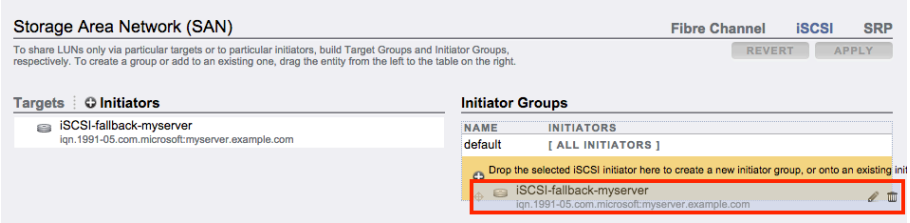

Figure 18. Creating an iSCSI initiator group

3. Move the cursor over the entry for the new initiator group. Two icons appear to the right of the target group box as shown in Figure 19.

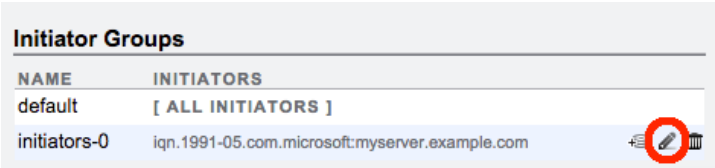

Figure 19. Selecting the iSCSI initiator group

4. Click the edit icon (✎) to display the following dialog window.

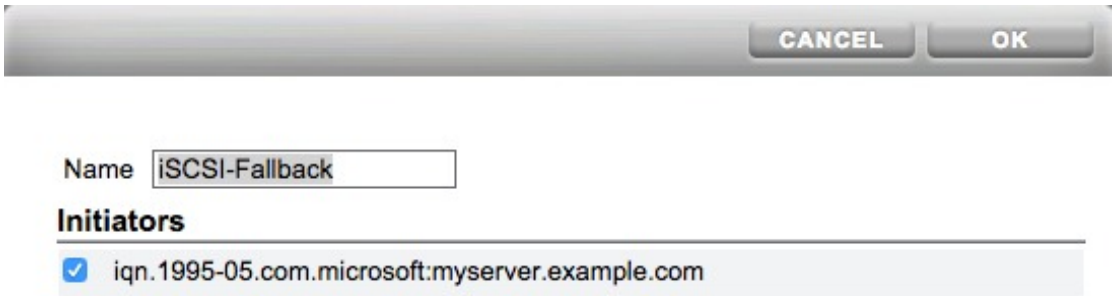

Figure 20. Replacing the default name with a new name for the iSCSI initiator group

   In the **Name** field, replace the default name with the name to be used for the iSCSI initiator group and click **OK**. For this example, the name "iSCSI-Fallback" may be appropriate here.

5. Click **APPLY** on the SAN configuration screen to confirm all the modifications, as shown in the following figure.

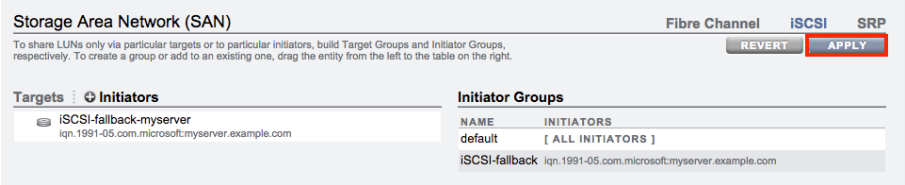

Figure 21. Completing the iSCSI initiator configuration

Windows Server Multipathing

The next configuration steps will create additional paths to the Windows Server connectivity.

Prior to taking any administrative actions, verify the number of paths that currently exist. Do this from the Windows Server Control Panel MPIO Properties panel under '**Configuration Snapshot**' as shown in the following figure.

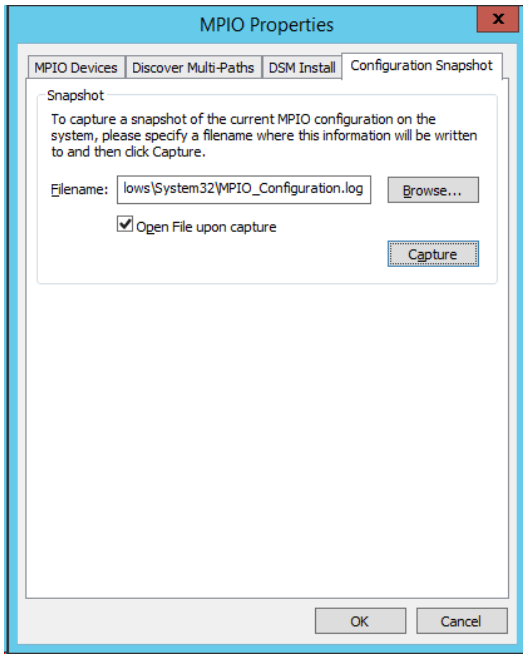

Figure 22. Windows Server MPIO Properties

The following example is a configuration snapshot showing one volume connected with FC using the default Microsoft DSM, which is appropriate for the Oracle ZFS Storage Appliance.

```
Registered DSMs: 1
==================
+-------------------------------|--------------------|----|----|----|---|-----+
|DSM Name                       |      Version       |PRP | RC | RI |PVP| PVE |
|-------------------------------|--------------------|----|----|----|---|-----|
|Microsoft DSM                  |006.0003.09600.17476|0020|0003|0001|030|False|
+-------------------------------|--------------------|----|----|----|---|-----+

Microsoft DSM
=============
MPIO Disk0: 02 Paths, Round Robin, Symmetric Access
        SN: 600144F0AE323ECA000054DE03C30004
        Supported Load Balance Policies: FOO RR RRWS LQD WP LB
```

```
        Path ID          State              SCSI Address      Weight
        ---------------------------------------------------------------------------
        0000000077030000 Active/Optimized   003|000|000|000   0
        * TPG_State: Active/Optimized  , TPG_Id: 0, TP_Id: 2
        Adapter: Emulex LightPulse HBA - Storport Mini...  (B|D|F: 019|000|001)
        Controller: 46616B65436F6E74726F6C6C6572 (State: Active)

        0000000077020000 Active/Optimized   002|000|000|000   0
        * TPG_State: Active/Optimized  , TPG_Id: 0, TP_Id: 1
        Adapter: Emulex LightPulse HBA - Storport Mini...  (B|D|F: 019|000|000)
        Controller: 46616B65436F6E74726F6C6C6572 (State: Active))

MSDSM-wide default load balance policy: N\A
No target-level default load balance policies have been set.
===============================================================================
```

As reflected in the example, there are two paths to the FC volume – one for each fabric. Both are active and online and carry equal weight, as the standard path selection policy is '`round-robin`.'

This can be confirmed by accessing the volume on the Oracle ZFS Storage Appliance BUI under **Shares > LUNs** and verifying that the GUID, highlighted in following figure example from the BUI, matches the Serial Number (SN) in the output of the MPIO configuration . The corresponding volume details are shown in Figure 23.

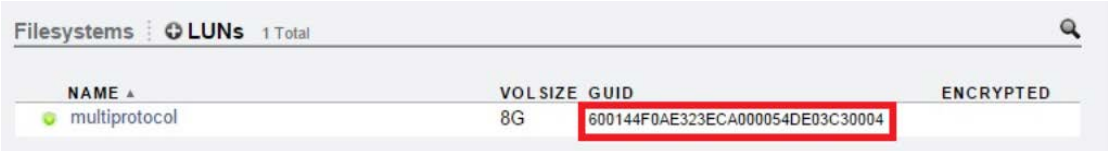

Figure 23. LUN details displayed in the Oracle ZFS Storage Appliance BUI

Presenting the Volume to Both iSCSI and FC Initiators

At this point, the volume is already being presented to FC initiators, so the next step is to present the volume to the appropriate iSCSI initiator(s). Use the following steps to change the configuration of the volume access:

1. From the Oracle ZFS Storage Appliance BUI, select **Shares** > **LUNs** (and optionally, the relevant project from the Projects panel on the left of the BUI).

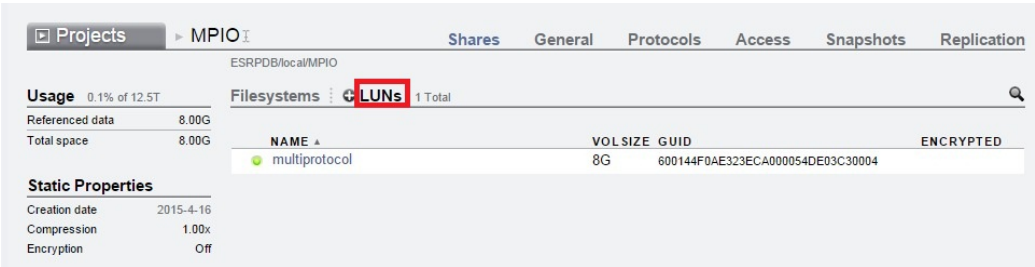

Figure 24. Selecting the volume to be modified

2. Verify the volume to be modified from the GUID information shown on the BUI and the Windows Server MPIO report shown previously. In this example, the displayed volume is

the correct one, as the GUID (600144F0AE323ECA000054DE03C30004) matches the LUN serial number.

3. Select the volume from the BUI by moving the mouse onto the appropriate volume line so that the pencil icon (✎) appears as shown in Figure 25. Click on the pencil icon.

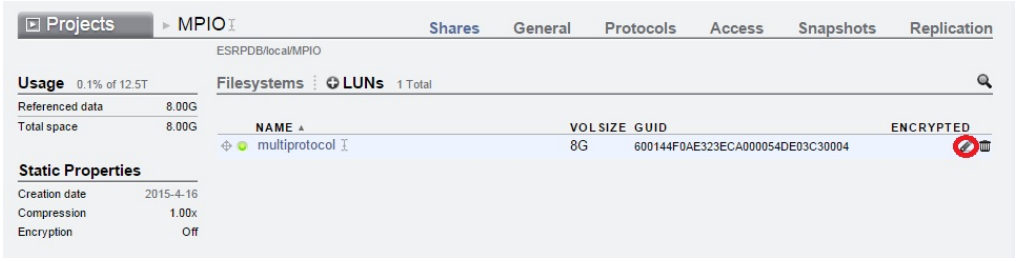

Figure 25. Volume edit icon

4. Select the Protocols tab and click on the **Initiator group: LU number** ✎ "**Edit**" link as shown.

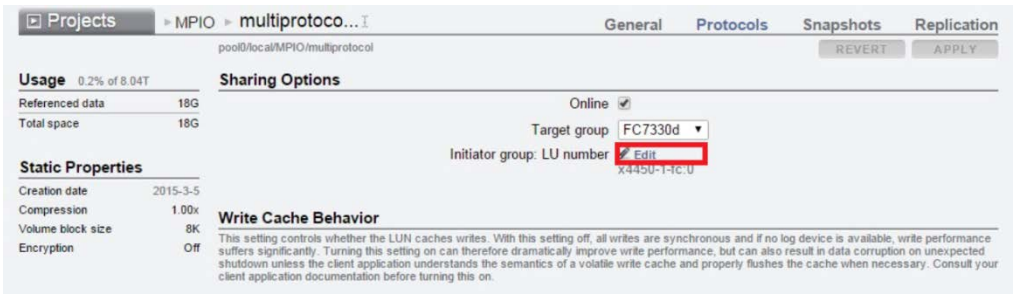

Figure 26. Edit Initiator groups

5. The "Edit Initiator Group(s)" dialog window will be shown as in Figure 27.

Figure 27. Edit Initiator Group(s) dialog

As seen in the example, the volume is currently assigned as LUN 0 to the Initiator Group x4450-1-fc. In order to allow access to the volume by the iSCSI initiator, the volume needs to be assigned a LUN number. Unless there is a particular reason to alter the LUN number, it is a good practice to let the Oracle ZFS Storage Appliance choose the appropriate LUN number by leaving the Policy at **Auto-assign**. Select the checkbox to the left of the appropriate initiator group.

Click **OK** to continue.

6. Next, the volume has to be presented to all the target groups as it is not possible to select multiple, particular target groups. This means that while the volume is potentially present over all the Oracle ZFS Storage Appliance configured target interfaces, it will only be accessible by the initiator groups nominated in the previous step. For added security, iSCSI access should be configured with CHAP authentication to ensure a further level of security over the volume access.

Figure 28. Selecting all target groups for presentation

Click **Apply** to continue.

The FC and iSCSI initiators associated with the Windows Server are allowed access to the volume and, while the volume is discoverable on all the assigned and configured interfaces, access is only granted to these initiators explicitly nominated.

Configuring iSCSI Multipathing

It may be necessary to explicitly enable iSCSI Multipathing in the Windows Server environment. This option only becomes available after at least one iSCSI volume is accessible to the Windows Server.

To enable iSCSI MPIO, follow these steps:

1. Select **Control Panel > MPIO**. If the text `MSFT20005iSCSI BusType_0x9` is present, iSCSI Multipathing is already configured and the rest of the steps in this section can be skipped.
2. Select the Discover Multi-Paths tab as shown in the following figure. Enable the checkbox for '**Add support for iSCSI devices**' and select the **Add** button.

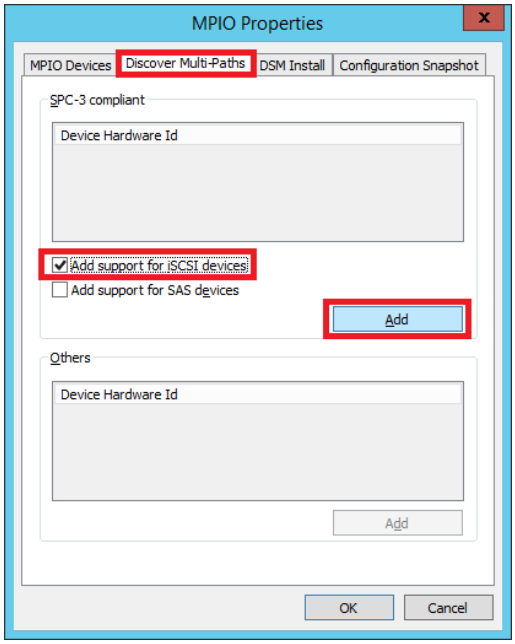

Figure 29. Enabling iSCSI Multipathing in the Windows Server interface

3. It may now be necessary to reboot the server to complete the operation. If this is the case, a "**Reboot Required**" dialog box will open. Click **Yes** if it is appropriate to reboot the server at this time.

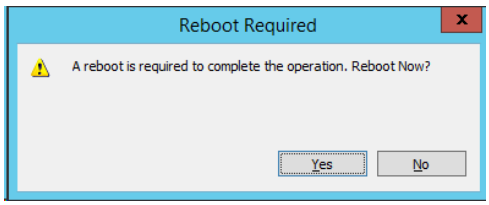

Figure 30. Reboot Required display

NOTE: If the iSCSI selection line from figure 28 is greyed out and cannot be selected, it may be necessary to open either a PowerShell or CMD session and run the following command:

```
C:\>mpclaim -r -i -d "MSFT2005iSCSIBusType_0x9"
```

This may cause the server to reboot automatically to allow the MPIO framework to claim control of the iSCSI volumes.

7. Once the server has rebooted, log in as an administrator and open **Control Panel** > **MPIO** once again.

   The list of supported devices should now include the string `MSFT2005iSCSIBusType_0x9` as shown in Figure 31.

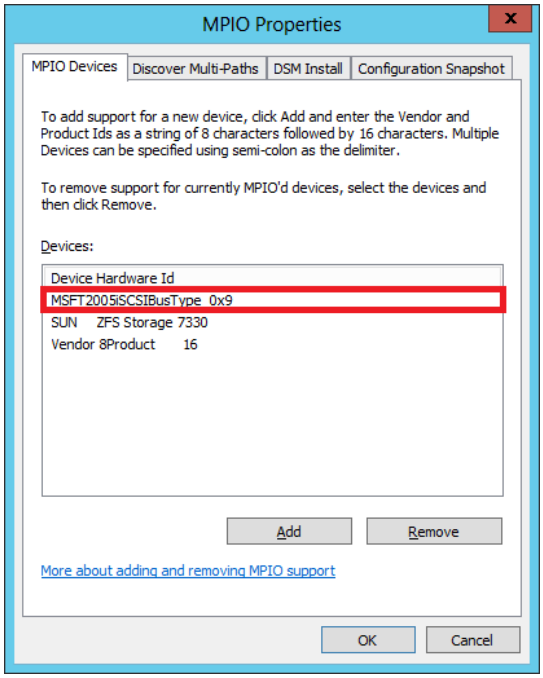

Figure 31. iSCSI MPIO configuration

Verifying iSCSI Multipathing

The volume presented to both iSCSI and FC target groups on the Oracle ZFS Storage Appliance should now be accessible by the existing FC connections and also the iSCSI adapter. To verify, rerun the **Configuration Snapshot** from the **MPIO** control panel.

```
Registered DSMs: 1
================
+-------------------------------|-------------------|----|----|----|---|-----+
|DSM Name                       |      Version      |PRP | RC | RI |PVP| PVE |
|-------------------------------|-------------------|----|----|----|---|-----|
|Microsoft DSM                  |006.0002.09200.16384|0020|0003|0001|030|False|
+-------------------------------|-------------------|----|----|----|---|-----+


Microsoft DSM
=============
MPIO Disk0: 03 Paths, Round Robin, Symmetric Access
        SN: 60144F0AE323ECA0054DE3C304
        Supported Load Balance Policies: FOO RR RRWS LQD WP LB
    Path ID          State              SCSI Address      Weight
    ---------------------------------------------------------------------------
    0000000077040001 Active/Optimized   004|000|001|000   0
      * TPG_State: Active/Optimized  , TPG_Id: 0, TP_Id: 13
        Adapter: Microsoft iSCSI Initiator...              (B|D|F: 000|000|000)
        Controller: 46616B65436F6E74726F6C6C6572 (State: Active)

    0000000077030000 Active/Optimized   003|000|000|000   0
      * TPG_State: Active/Optimized  , TPG_Id: 0, TP_Id: 2
```

```
        Adapter: Emulex LightPulse HBA - Storport Mini...  (B|D|F: 019|000|001)
        Controller: 46616B65436F6E74726F6C6C6572 (State: Active)

    0000000077020000 Active/Optimized   002|000|000|000   0
      * TPG_State: Active/Optimized  , TPG_Id: 0, TP_Id: 1
        Adapter: Emulex LightPulse HBA - Storport Mini...  (B|D|F: 019|000|000)
        Controller: 46616B65436F6E74726F6C6C6572 (State: Active)
MSDSM-wide default load balance policy: N\A
No target-level default load balance policies have been set.
============================================================================
```

Comparing the latest MPIO configuration snapshot with the earlier one, the difference is that the Microsoft iSCSI Initiator controlled volumes are now forming part of the multipath set.

Most probably, access to the volumes through the iSCSI route is only desirable in the case in which both FC connections are inoperable. In Microsoft MPIO, this setup is facilitated by assigning each path to storage a priority value or weight or by marking the iSCSI path as a standby path – the method used depends on which allocation strategy is being used.

From the Oracle ZFS Storage Appliance advanced analytics screen, you can verify the access through iSCSI or FC paths. In the example shown in Figure 31, the LUN is being accessed reasonably symmetrically through both iSCSI and FC.

Figure 32. Verifying access through iSCSI and FC paths

The MPIO settings for the Windows Server can be changed by selecting the Properties dialog from the list of volumes in an explorer window.

Navigate to a display of the hardware properties, as seen in figure 32, by selecting the Hardware tab, then the "disk drive" SUN ZFS Storage, and click Properties.

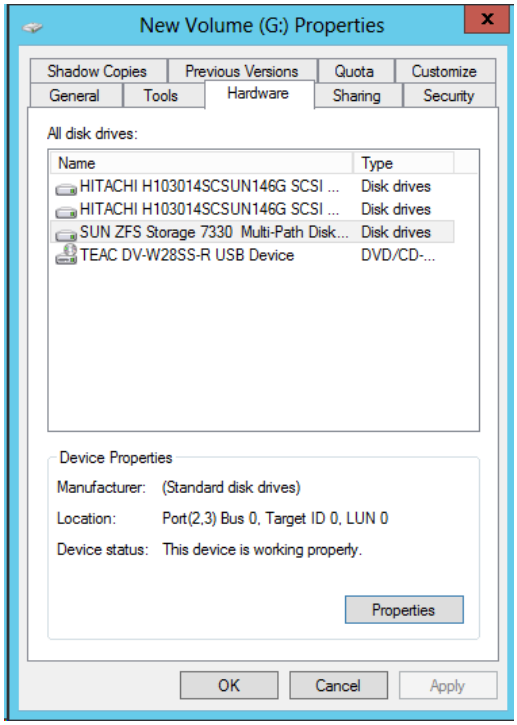

Figure 33. Volume Properties in the Windows Server MPIO interface

The following window showing the properties for the selected hardware contains the tab for MPIO, where the MPIO policy can be set.

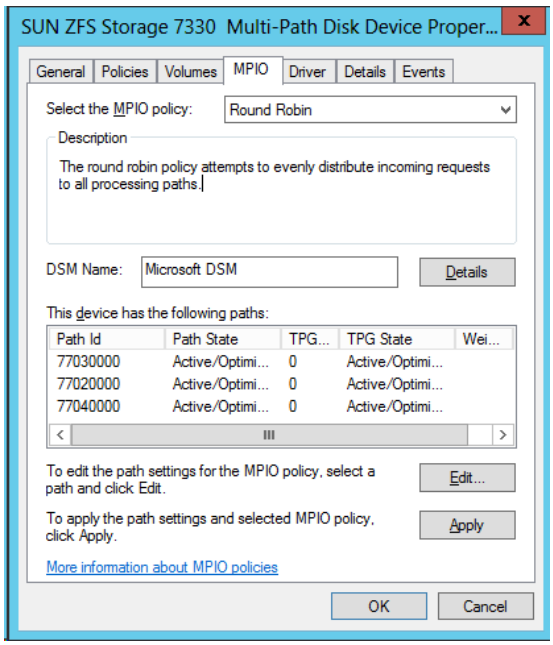

Figure 34. Hardware Properties MPIO

There are multiple solutions for configuring the iSCSI path to provide the necessary standby. The following subsections list some of those options.

Failover

When the MPIO policy for a LUN is set to Failover, no load balancing is performed. This setting is not recommended, as only one path can be set as active/optimized and all other paths are active/unoptimized, standby or unavailable. When the active/optimized path fails, the other paths are then tried sequentially until an available path is found.

Round Robin

Round Robin policy instructs Windows to use each Active/Optimized path in turn. This is the default setting and may not be optimal, depending on the application, as I/O requests are evenly distributed without any recognition of unbalanced requests.

This policy does not allow any paths to be manually configured as Standby. Do not choose Round Robin if the LUN will be configured with both FC and iSCSI access.

Round Robin with Subset

Round Robin with Subset is similar to Round Robin but does allow particular paths to be set as Standby. To set the iSCSI path to be the standby path, select the MPIO Path Details from the Hardware Properties dialog.

The Path State drop-down should be modified to show Standby as shown in Figure 35.

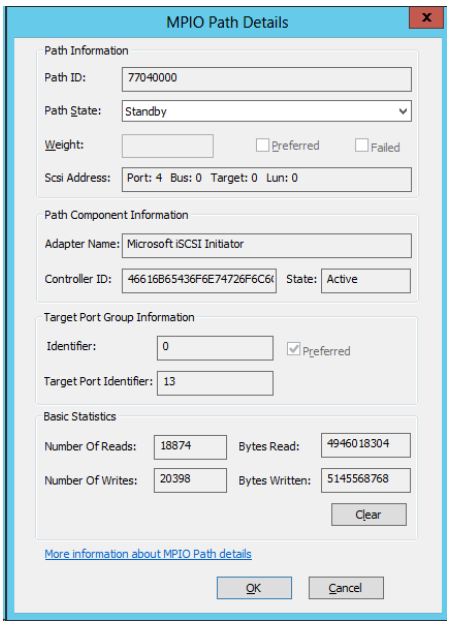

Figure 35. MPIO path details

If the FC paths fail, the iSCSI path becomes the Active one and I/O resumes using only the iSCSI path.

When the FC paths are returned to service, failback automatically stops I/O on the iSCSI path and continues on the FC ones.

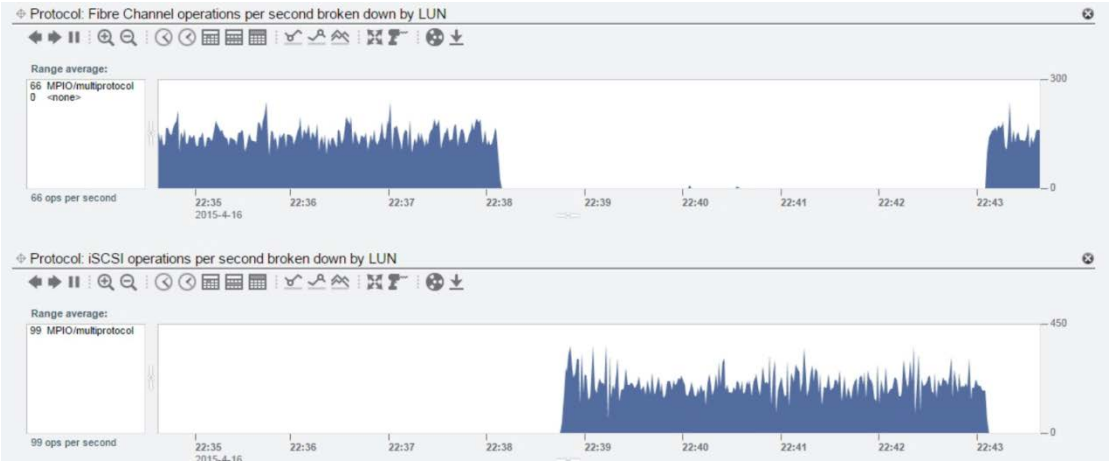

Figure 36. Round Robin with subset path failure and recovery

Least Queue Depth

Least Queue Depth is analogous to Round Robin but the I/O request is assigned to the path with the smallest queue of outstanding requests. It is not possible to assign a path as Standby with Least Queue Depth or to assign a preference of path choice, so this policy should not be chosen for implementing the multiprotocol multipathing. Since iSCSI is usually higher latency and lower bandwidth than FC, choosing to use both types would present performance challenges. Typical operation would use the higher bandwidth interface exclusively until a failure occurs in all the paths provided by that interface.

Weighted Paths

In Weighted Paths, a number is assigned to each path that represents the priority by which the path will be chosen relative to the other assigned weights. The higher the number, the lower the priority of that path. In practice, the favored FC paths should be assigned a very low weight and the iSCSI path a relatively high number.

For the purposes of the examples, the low weight was set at 0 and the high one at 100. These designations result in all the I/O being allocated to the FC channels when they are in service.

As shown in the following two figures, path failure detection and retries can take approximately 30 seconds (depending on the DSM timer settings).

Figure 37. Weighted Path failover and recovery

Conclusion

The unified approach to storage provided by the Oracle ZFS Storage Appliance family can be used to provide the framework for high resilience at low cost and management overhead. By simultaneously presenting LUNs for access by multiple protocols, the complete dependence on a single protocol is removed, allowing for a higher level of resilience to be designed into architecture deployments.

By using relatively low-cost networking infrastructure components, high availability is an affordable option when deployed with Fibre-Channel dual-fabric configurations.

References

See the following resources for additional information relating to the products covered in this document.

- Oracle ZFS Storage Appliance White Papers and Subject-Specific Resources
  http://www.oracle.com/technetwork/server-storage/sun-unified-storage/documentation/index.html
- Oracle ZFS Storage Appliance Product Information
  https://www.oracle.com/storage/nas/index.html
- Oracle ZFS Storage Appliance Documentation Library, including Installation, Analytics, Customer Service, and Administration guides:
  http://www.oracle.com/technetwork/documentation/oracle-unified-ss-193371.html
- The *Oracle ZFS Storage Appliance Administration Guide* is also available through the Oracle ZFS Storage Appliance help context.
  The Help function in Oracle ZFS Storage Appliance can be accessed through the browser user interface.

Microsoft Windows Server Multiprotocol Multipathing with the Oracle ZFS Storage Appliance
May 2015 Version 1.0

Author: Andrew Ness

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com

Oracle is committed to developing practices and products that help protect the environment

Hardware and Software, Engineered to Work Together