DataDirect™
NETWORKS

Storage Fusion Architecture

Multipath

(v1.8)

User Guide

1. OVERVIEW

This document provides instructions on how to install DDN's multipath RPM and manage Linux multipathing on DDN's family of Disk Arrays. The ddn_mpath_RHEL5_SLES10-1.1-0.x86_64.rpm provides validated and optimized configuration settings to achieve load balancing, path failover, and controller failover of DDN disk arrays in supported Linux environments. DDN's multipath ddn_mpath_RHEL5_SLES10-1.1-0.x86_64.rpm is an extension to the Linux multipath tools package: device mapper multipathing, *dm-multipath*.c

2. DDN MULTIPATH RPM VERSIONS

Use the table below to determine which DDN multipath RPM package required for each DDN array type and supported Linux Distribution
The figure below shows the current compatibility matrix with DDN Products, Linux Distribution, and DDN multipath RPM.

Products Supported in all versions:
EF 3015, S2A 6620, S2A 9900, SFA 10000

Linux Distribution version	DDN Linux Multipath rpms required
Red Hat Enterprise Linux 5.4	ddn_mpath_RHEL5_SLES10-1.1-0.x86_64.rpm
Red Hat Enterprise Linux 5.5	ddn_mpath_RHEL5_SLES10-1.1-0.x86_64.rpm
Red Hat Enterprise Linux 5.6	ddn_mpath_RHEL5_SLES10-1.1-0.x86_64.rpm
Red Hat Enterprise Linux 6.0	ddn_mpath_RHEL6-1.1-0.el6.x86_64.rpm
SUSE LINUX Enterprise Server 10 SP2	ddn_mpath_RHEL5_SLES10-1.1-0.x86_64.rpm
SUSE Linux Enterprise Server 11	multipath-tools-0.4.8-40.21.1ddn.x86_64.rpm kpartx-0.4.8-40.21.1ddn.x86_64.rpm ddn_mpath_SLES11-1.1-0.x86_64.rpm

3. INSTALLATION

3.1 RPM INSTALLATION

Determine the Linux distribution on the host and cross-reference with the version listed in the table in section 2.0. Apply the RPM installation steps for the specific Linux version installed on the Linux host.

3.1.1 DETERMINING THE LINUX DISTRIBUTION AND VERSION

Determing the distribution source can be done by running the Linux shell command: cat /etc/issue

For example:

```
#cat /etc/issue
Welcome to SUSE Linux Enterprise Server 11 SP1  (x86_64) - Kernel \r (\l).
```

For Redhat Linux distributions, the specific version can be determined by running the Linux shell command: cat /etc/redhat-release

Example for RHEL 6.0:

```
cat /etc/redhat-release
Red Hat Enterprise Linux Server release 6.0 (Santiago)
```

For SUSE Linux distribution, the specific version can be determined by running the Linux shell command: cat /etc/suse-release

Example for SLES 11:

```
cat /etc/suse-release
SUSE Linux Enterprise Server 11 (x86_64)
VERSION = 11
PATCHLEVEL = 0
```

3.1.2 RPM INSTALLATION ON RHEL 5.4 – RHEL 5.6 AND SLES 10 SP2

For Red Hat Enterprise Linux 5, update 4,5 and 6, and for SuSE Enterprise Linux Server 10 SP2, run the command shown in the example below:

```
# rpm -ivh ddn_mpath_RHEL5_SLES10-1.1-0.x86_64.rpm
Preparing...               ########################################### [100%]
  1:ddn_mpath_RHEL5_SLES10 ########################################### [100%]

The DDN mpath config file has been installed as /etc/multipath.conf.ddn.
Rename this file to /etc/multipath.conf if the file does not already exist.
Manually merge the DDN supplied file with /etc/multipath.conf if one is
already in use. Please read the DDN Multipath Manual for more information.
```

3.1.3 RPM INSTALLATION ON RHEL 6.0

For Red Hat Enterprise Linux 6.0, run the command shown in the example below:

```
# rpm -ivh ddn_mpath_RHEL6-1.1-0.el6.x86_64.rpm
Preparing...               ########################################### [100%]
  1:ddn_mpath_RHEL6        ########################################### [100%]

The DDN mpath config file has been installed as /etc/multipath.conf.ddn.
Rename this file to /etc/multipath.conf if the file does not already exist.
Manually merge the DDN supplied file with /etc/multpath.conf if one is
already in use. Please read the DDN Multipath Manual for more information.
```

3.1.4 RPM INSTALLATION ON SLES 11

For SLES 11, install the RPMs using the commands shown in the example below:

Step 1:

```
# rpm -Uvh kpartx-0.4.8-40.21.1ddn.x86_64.rpm multipath-tools-0.4.8-
40.21.1ddn.x86_64.rpm
Preparing...               ########################################### [100%]
  1:kpartx                 ########################################### [ 50%]
  2:multipath-tools        ########################################### [100%]
   Scanning scripts ...
   cResolve dependencies ...
```

Step 2:

```
# rpm -ivh ddn_mpath_SLES11-1.1-0.x86_64.rpm
Preparing...                ########################################### [100%]
  1:ddn_mpath_SLES11        ########################################### [100%]

The DDN mpath config file has been installed as /etc/multipath.conf.ddn.
Rename this file to /etc/multipath.conf if the file does not already exist.
Manually merge the DDN supplied file with /etc/multipath.conf if one is
already in use. Please read the DDN Multipath Manual for more information.
```

3.2 MERGING THE /ETC/MULTIPATH.CONF.DDN FILE

After installing the DDN multipath rpm, a file name /etc/multipath.conf.ddn will be installed in the hosts /etc/ directory.

If the host has no devices managed under dm-multipath and the current file /etc/multipath.conf file is not actively used, then it is possible to backup the existing /etc/multipath file and replace it with the/etc/multipath.conf.ddn file using the following steps:

```
mv -f/etc/multipath.conf /etc/multipath.config.original

cp -f /etc/multipath.conf.ddn /etc/multipath.conf
```

Be sure to edit the {blacklist} section to ignore the devices that should not be managed or queried by dm-multipath. The administrator must analyze system SCSI device targets and update blacklist appropriately.

Example 1: the configuration below, when merged into, will cause the /dev/sda and all its partitions to be ignored by dm-multipath:

```
blacklist {
      devnode "^sda[0-9]*$"
}
```

Blacklist and blacklist_exception example to where multipathd will only attempt to manage devices with a vendor inquiry string of “DDN”

```
blacklist {
      device {
            vendor=”*”
      }
}

blacklist_exception {
      device {
            vendor=”DDN”
      }
}
```

4. ARCHITECT AND CONFIGURE YOUR DDN STORAGE ARRAY

Determine the LUN mapping configuration required for each host and architect the storage layout and mapping required. Storage architecture and host connectivity implementation details are beyond the scope of this guide. Each application requirements will demand different implementations. This document focuses on concepts required to achieve the primary goals of linux multipathing:

- Target device path failover redundancy
- Storage controller failover redundancy
- Preferred target device path optimization
- Target device path load balancing across multiple preferred paths

4.1 OVERVIEW OF LUN PRESENTATION PROCESS ON S2A 9900

Refer to the S2A 9900 User Guide for full descriptions of the commands and options available.

A simple example below with four LUNS is provided for illustration purposes. Note that LUN ownership is optimally balanced across the two controllers. The same principles are identical for Fibre Channel and Infiniband S2A 99000 models:

```
                         Logical Unit Status

                                     Capacity  Block
 LUN  Label     Owner  Status        (Mbytes)  Size  Tiers Tier list
-----------------------------------------------------------------------------
  1 lun_1         1    Ready         3815470   512     1    1
  2 lun_2         2    Ready         3815470   512     1    2
  3 lun_3         1    Ready         3815470   512     1    3
  4 lun_4         2    Ready         3815470   512     1    4
```

On S2A 9900 Linux multipath configurations, preferred access to LUNs will be automatically be assigned to host paths serviced by the owning controller.

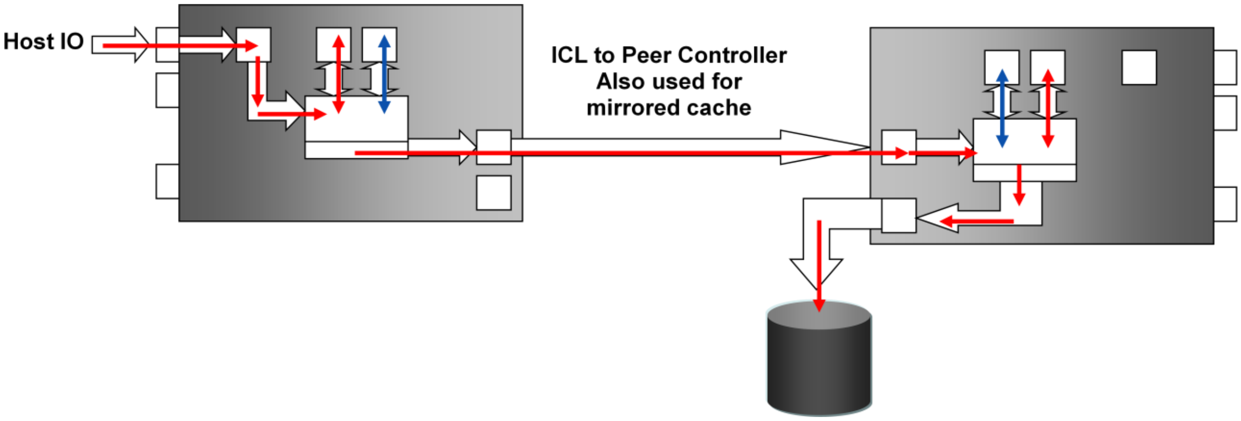
To implement S2A 9900 controller failover on Linux, it is important to zone S2A 9900 LUNs from ports both controllers.

On S2A 9900, use the “user” and/or “zoning” directOS commands to achieve a port zoning configuration which provides access to the LUN from ports of both controllers. In the zoning example below, all four LUNs are presented on all 8 ports. Highlighted in yellow are the internal LUNs owned by the respective controllers, illustrating which S2A 9900 host ports will provide preferred access to each LUN.

```
                                    LUN Zoning
 Port    World Wide Name    (External LUN, Internal LUN)
----------------------------------------------------------------------------
  1      21000001FF0207EE    000,001      001,002      002,003      003,004
  2      22000001FF0207EE    000,001      001,002      002,003      003,004
  3      23000001FF0207EE    000,001      001,002      002,003      003,004
  4      24000001FF0207EE    000,001      001,002      002,003      003,004


                                    LUN Zoning
 Port    World Wide Name    (External LUN, Internal LUN)
----------------------------------------------------------------------------
  1      25000001FF0207EE    000,001      001,002      002,003      003,004
  2      26000001FF0207EE    000,001      001,002      002,003      003,004
  3      27000001FF0207EE    000,001      001,002      002,003      003,004
  4      28000001FF0207EE    000,001      001,002      002,003      003,004
```

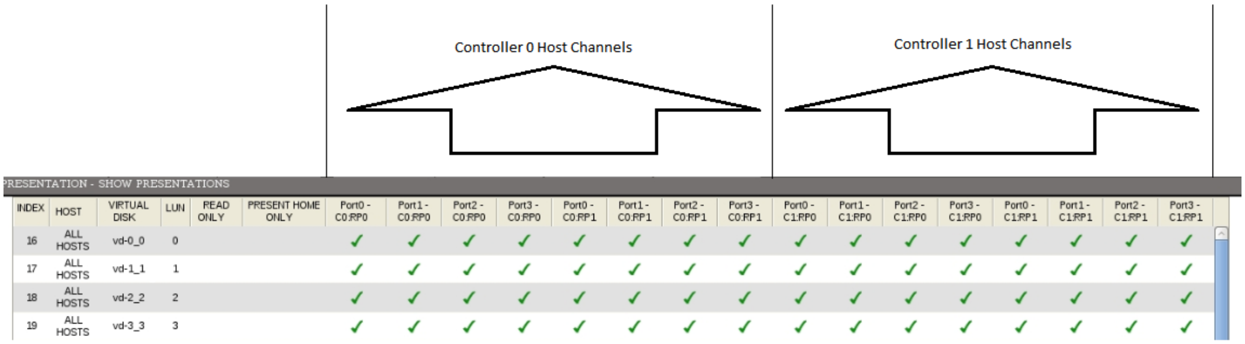
4.1.1 LUN PRESENTATION PROCESS ON S2A 6620 AND SFA 10000

S2A 6620’s architecture includes two controllers with one RAID Processor on each S2A 6620 controller. The architecture is asymmetrical active/active – the preferred home controller, which owns the VDs storage pool, performs all I/O to a VD storage pool member disks.
Host I/O performed on the host channel of the non-preferred controller are forwarded to the VDs home controller, which processes the disk operations and returns data to the peer controller via the inter-controller link connecting the two controllers.

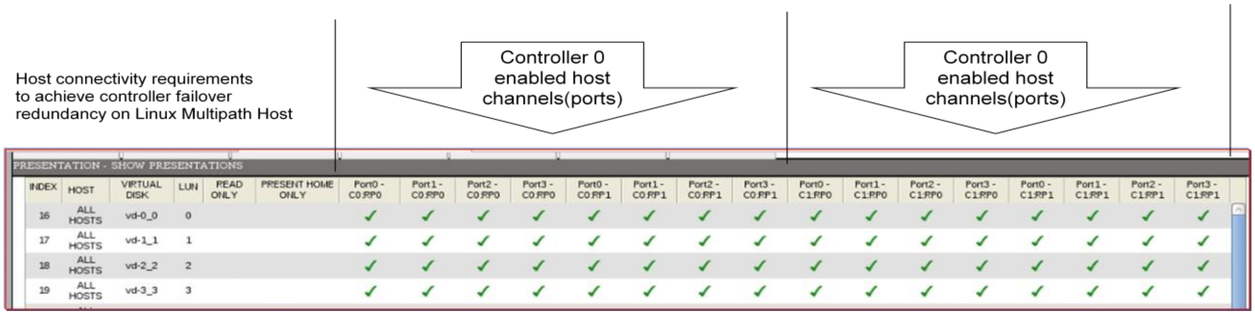


In the SFAOS, host channels are referenced by Controller, Raid Processor, and Port values, and in the example below VDs 0 to 3 are presented to all hosts through all host channels.
S2A 6620 models have 4 Fibre Channel Ports.

The WebUI illustrates enabled host channels on which the presentations will be available:


Controller 0 Host Channels
Controller 1 Host Channels
PRESENTATION - SHOW PRESENTATIONS

INDEX	HOST	VIRTUAL DISK	LUN	READ ONLY	PRESENT HOME ONLY	Port0 - C0:RP0	Port1 - C0:RP0	Port2 - C0:RP0	Port3 - C0:RP0	Port0 - C0:RP1	Port1 - C0:RP1	Port2 - C0:RP1	Port3 - C0:RP1	Port0 - C1:RP0	Port1 - C1:RP0	Port2 - C1:RP0	Port3 - C1:RP0	Port0 - C1:RP1	Port1 - C1:RP1	Port2 - C1:RP1	Port3 - C1:RP1
16	ALL HOSTS	vd-0_0	0			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
17	ALL HOSTS	vd-1_1	1			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
18	ALL HOSTS	vd-2_2	2			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
19	ALL HOSTS	vd-3_3	3			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓



Host connectivity requirements to achieve controller failover redundancy on Linux Multipath Host

Controller 0 enabled host channels(ports)
Controller 0 enabled host channels(ports)
PRESENTATION - SHOW PRESENTATIONS

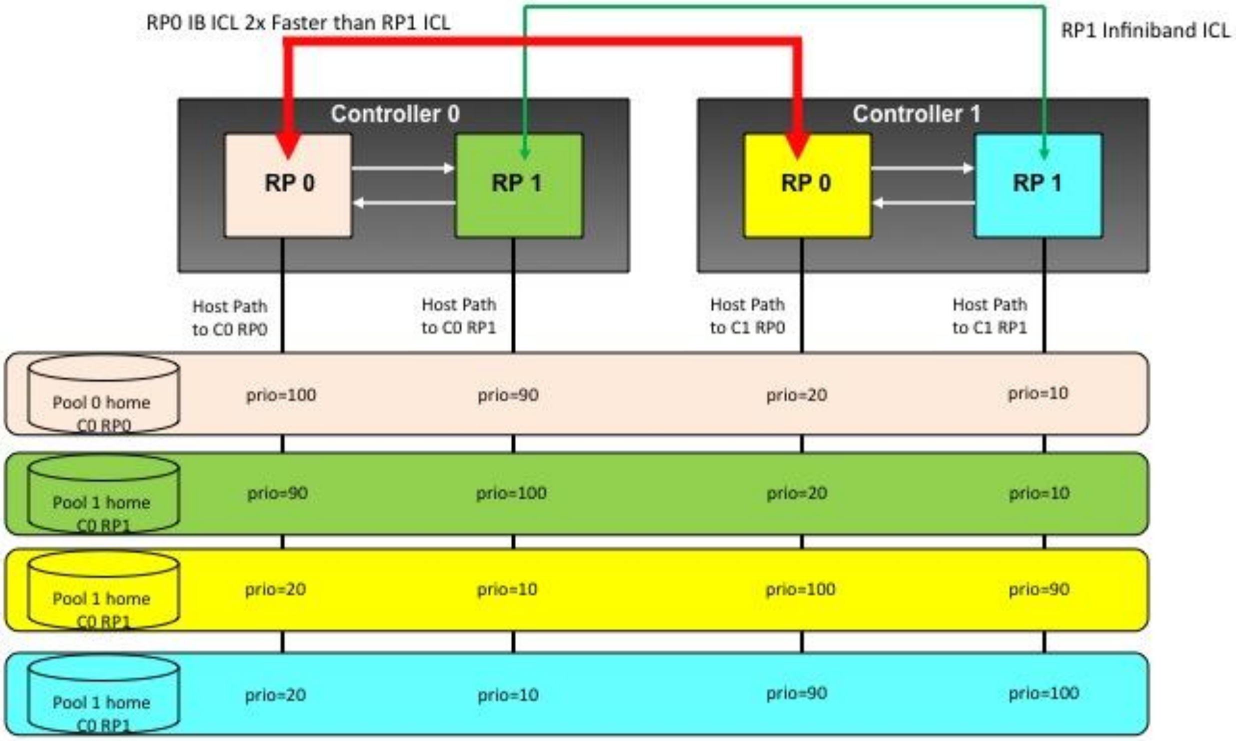
INDEX	HOST	VIRTUAL DISK	LUN	READ ONLY	PRESENT HOME ONLY	Port0 - C0:RP0	Port1 - C0:RP0	Port2 - C0:RP0	Port3 - C0:RP0	Port0 - C0:RP1	Port1 - C0:RP1	Port2 - C0:RP1	Port3 - C0:RP1	Port0 - C1:RP0	Port1 - C1:RP0	Port2 - C1:RP0	Port3 - C1:RP0	Port0 - C1:RP1	Port1 - C1:RP1	Port2 - C1:RP1	Port3 - C1:RP1
16	ALL HOSTS	vd-0_0	0			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
17	ALL HOSTS	vd-1_1	1			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
18	ALL HOSTS	vd-2_2	2			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
19	ALL HOSTS	vd-3_3	3			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

5. PERFORMANCE CONSIDERATIONS FOR SFA 10000

To achieve optimal performance on SFA 10000 appliances, it is important to make note that the pool home attribute is assigned to raid processors (RP's), not to controllers as in the case of S2A 6620 and S2A 9900. This design feature requires special consideration when optimizing multipath performance.

The example below illustrates pool priority values reported by the SFA10k prioritizer provided as part of the DDN linux multipath RPM. In this example, four pools are shown with color-coding that matches the pool ownership defined in the example.
The most optimal path will always be a host port belonging to the home RP for the VD and underlying pool of any given presented LUN. The second most optimal path will be the peer RP in the same controller as the home RP. The third most optimal choice is always RP0 of the peer controller. The least optimal RP is always RP1 of the peer controller. The DDN linux multipath SFA10k prioritizer determines the topology of each path and reports a priority value for each path with values of 100,90,20 and 10 reflecting the four RP priorities.

SFA10K_PRIO_ALUA RETURN VALUES



	Host Path to C0 RP0	Host Path to C0 RP1	Host Path to C1 RP0	Host Path to C1 RP1
Pool 0 home C0 RP0	prio=100	prio=90	prio=20	prio=10
Pool 1 home C0 RP1	prio=90	prio=100	prio=20	prio=10
Pool 1 home C0 RP1	prio=20	prio=10	prio=100	prio=90
Pool 1 home C0 RP1	prio=20	prio=10	prio=90	prio=100

6. ACTIVATING MULTIPATHD

multipathd is the daemon that will monitor device paths at intervals configured using the device “polling_interval” directive in the /etc/multipath.conf file. Note that device path failover is controlled by the “no_path_retry” directive in the event of a primary path failure. This component is described in detail in later sections of this document. The multipathd path checking service should be enabled on the server using the chkconfig program as shown below:

```
chkconfig multipathd on
```

6.1 VERIFYING THAT MULTIPATHD SERVICE IS CONFIGURED TO AUTO-START

The “--list” option to chkconfig will display the auto-start setting for each Linux Run Level:

```
chkconfig multipathd --list

Sample Output:
multipathd      0:off   1:off   2:on    3:on    4:on    5:on    6:off
```

6.2 STARTING THE MULTIPATHD SERVICE

Reboot the host or run the service command to start the multipathd daemon:

```
service multipathd start
```

7. INTRODUCTION TO DM-MULTIPATH

DDN’s multipath ddn_mpath_RHEL5_SLES10-0.4-7.x86_64.rpm is an extension to the linux multipath tools package: device mapper multipathing, *dm-multipath*. *dm-multipath* provides a means for accessing a device with multiple paths to that device in Linux. The device mapper kernel module creates a single SCSI block device for every LUN probed by Linux at boot time (or manually – see later in this document). This device(s) for each LUN can be found in /dev/mpath as well as /dev/mapper directories in Linux.

7.1 HOW IT WORKS

Linux dm-multipath queries each Linux SCSI disk devices and determines which disk devices are duplicate paths to disk targets on a Linux host computer.

Each path is a physical connection (Fibre Channel or Infiniband for DDN storage) between the initiator (the server) and a specific LUN on the target (data storage) device. Paths to the same target are assembled into priority groups. Only one of these priority groups will be used at a time for I/O to the device. The priority group that is being utilized for IO is labeled “*active”*.
A component to dm-multipath is used to determine which path to use for the next IO.
This component is called the *Path Selector*.
If an I/O fails on the selected active path, that path will be disabled and the I/O is retried down a different path within the same group of paths called a *Priority Group*. There can be more than one path in this priority group, and each path is weighted for a priority level called a *Path Group Priority*. The highest priority level in the *Group* determines the primary path to use to access the device. If every path in the path group fails, then a different priority group will chosen and enabled to continue IO to the target device.

7.2 DM-MULTIPATH COMPONENTS

- The **dm-multipath kernel module** –
  Provides control over paths and priorities.

- The **multipath daemon (multipathd)** –
  Used by the Linux kernel to monitor and control the multipath paths.

- The **multipath command** –
  Utilized by the user to manipulate (view, flush cached entries…) multipath devices.

- The **/etc/multipath.conf** file –
  The configuration file read by multipathd to describe the behaviors and attributes of multipath devices.

The reader is encouraged to reference the man pages for these: **multipath, multipathd, multipath.conf, kpartx, dmsetup, mpath_prio_alua**

7.3 DDN MULTIPATH RPM CONTENT

- Enhancements to the /etc/multipath.conf, that defines the correct multipath settings for a DDN disk arrays

- An optimized prioritizer for the SFA 10000 capable of optimizing path priorities for all four of the SFA 10000's Raid Processors (RP's) by priority group.

7.4 LINUX SCSI DEVICE ENUMERATION

Linux enumerates SCSI devices in order of Host:BUS:ID:LUN

```
# lsscsi -g
[0:0:0:0]    disk    SEAGATE  ST373207LC       D703  /dev/sda    /dev/sg0
[0:0:6:0]    process PE/PV    1x2 SCSI BP      1.0   -           /dev/sg1
[1:0:0:0]    disk    DDN      S2A 6620         1.03  /dev/sdd    /dev/sg3
[1:0:0:1]    disk    DDN      S2A 6620         1.03  /dev/sde    /dev/sg5
[1:0:0:2]    disk    DDN      S2A 6620         1.03  /dev/sdg    /dev/sg7
[2:0:0:0]    disk    DDN      S2A 6620         1.03  /dev/sdb    /dev/sg2
[2:0:0:1]    disk    DDN      S2A 6620         1.03  /dev/sdc    /dev/sg4
[2:0:0:2]    disk    DDN      S2A 6620         1.03  /dev/sdf    /dev/sg6
```

The *lsscsi* utility is part of some Linux distributions, but is also available for free download. It displays SCSI path information, Vendor and Product inquiry strings, block devices, and associated SCSI generic devices in a nice easy to read output.

7.5 FUNCTIONING MULTIPATH OUTPUT (ON RHEL 5 AND SLES 10)

```
# multipath -ll

mpath19 (360001ff07211600000000002688e10002) dm-2 DDN,S2A 6620
[size=21T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=50][active]
 \_ 1:0:0:2 sdg 8:96  [active][ready]
\_ round-robin 0 [prio=10][enabled]
 \_ 2:0:0:2 sdf 8:80  [active][ready]
```

```
mpath19 (360001ff07211600000000002688e10002) dm-2 DDN,S2A 6620
   ^                    ^                    ^    ^     ^
   |                    |                    |    |     |___ Product
   |                    |                    |    |_______ Vendor
   |                    |                    |__________ Sysfs Name
   |                    |_____________________________ WWID of the Device
   |_________________________________________________ User Defined Alias Name
```

- **Product** and **Vendor** are returned from the SCSI inquiry string
- The **sysfs** name is the device mapper SCSI block device name
- **WWID** is the unique identifier for the multipath device, which includes OEM vendor strings, owning controller MAC, and LUN id's
- The **alias** name is optional and can be defined in /etc/multipath.conf in conjunction with enabling user_friendly_names

```
# multipath -ll

mpath19 (360001ff07211600000000002688e10002) dm-2 DDN,S2A 6620
[size=21T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=50][active]
 \_ 1:0:0:2 sdg 8:96  [active][ready]
\_ round-robin 0 [prio=10][enabled]
 \_ 2:0:0:2 sdf 8:80  [active][ready]
```

```
 [size=21T][features=1 queue_if_no_path][hwhandler=0][rw]

      ^                       ^              ^        ^
      |                       |              |        |___ Device Permissions
      |                       |              |___________ Hardware Handler
      |                       |_________________________ Supported Features
      |________________________________________________ Size of the DM Device
```

The **features** value determines what to do if the path has failed. It is recommended to have ***no_path_retry*** defined in multipath.conf

```
# multipath -ll

mpath19 (360001ff07211600000000002688e10002) dm-2 DDN,S2A 6620
[size=21T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=50][active]
 \_ 1:0:0:2 sdg 8:96  [active][ready]
\_ round-robin 0 [prio=10][enabled]
 \_ 2:0:0:2 sdf 8:80  [active][ready]
```

```
Path Group 1
\_ round-robin 0 [prio=50][active]

 ^       ^           ^        ^
 |       |           |        |_________________ Path Group State
 |       |           |_________________________ Path Group Priority
 |       |____________________________________ Path Selector and Repeat Count
 |___________________________________________ Path Group Level
```

A **Path** is the connection from the server (initiator – HBA/HCA) to a specific LUN (target device). The server will create a device path for each available path to the target device if the device (LUN) is mapped/masked in such a way that it is available across multiple server initiator ports (for redundancy, HA, and performance reasons). Thus the server will have multiple paths to the same target device in this situation.

```
# multipath -ll

mpath19 (360001ff0721160000000002688e10002) dm-2 DDN,S2A 6620
[size=21T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=50][active]
 \_ 1:0:0:2 sdg 8:96  [active][ready]
\_ round-robin 0 [prio=10][enabled]
 \_ 2:0:0:2 sdf 8:80  [active][ready]
```

The previously mentioned *Paths* are organized into **Path Groups**. Only one path group can be active at any time. The **Path Selector** determines which path in the path group will be used to handle the next IO. This IO will only go down the active path.

Below is an example of a configuration with 8 paths to the same LUN, with 2 path groups, 4 paths per path group.

```
[fc_host]# multipath -ll
ddn03_lun076 (360001ff010308e504c000100001d1bf1) dm-75 DDN,S2A 9900
[size=15T][features=0][hwhandler=0][rw]
\_ round-robin 0 [prio=200][active]
 \_ 18:0:2:75  sduv  67:624  [active][ready]
 \_ 19:0:2:75  sduw  67:640  [active][ready]
 \_ 19:0:3:75  sdacf 135:624 [active][ready]
 \_ 18:0:3:75  sdadj 65:848  [active][ready]
\_ round-robin 0 [prio=40][enabled]
 \_ 19:0:0:75  sdev  129:112 [active][ready]
 \_ 18:0:0:75  sdew  129:128 [active][ready]
 \_ 19:0:1:75  sdml  69:464  [active][ready]
 \_ 18:0:1:75  sdne  71:256  [active][ready]
```

All paths have a specific **Priority**. The **Priority Callout** function within multipath is defined in multipath.conf and determines the priority for all paths. The *group_by_prio* path grouping policy provides the path priority and is used to group paths together and determine the priority value within the path selector.

The **Path Group State** displays the current status of the path to the path group. Each path within the path group may show one of a few different status states. The “active” state means the path is the optimal path and is capable of handling IO. The “enabled” state means the path is capable of handling IO, but is not the optimal path to use. The “disabled” state infers that no path is available to the active path group, and IO will go down another path group if it is in the ready state.

The **Path Group Priority** is a weighted value. The multipath daemon will use the highest path group priority value to determine the active path group.

The **Path Selector** is a component of multipath that chooses which path to take for the next IO. The “round-robin” algorithm is recommended for load balancing IO across available paths, and this is configured in the multipath.conf file using the “path_selector” directive. The “active-passive” algorithm could also be used.

```
# multipath -ll

mpath19 (360001ff0721160000000002688e10002) dm-2 DDN,S2A 6620
[size=21T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=50][active]
 \_ 1:0:0:2 sdg 8:96  [active][ready]
\_ round-robin 0 [prio=10][enabled]
 \_ 2:0:0:2 sdf 8:80  [active][ready]
```

```
First Path on Path Group 1 (could be more than one)
\_ 1:0:0:2 sdg 8:96  [active][ready]

      ^     ^    ^      ^       ^
      |     |    |      |       |________ Physical Path State
      |     |    |      |______________ DM Path State
      |     |    |____________________ Device Major/Minor Numbers
      |     |________________________ Block Device Name
      |_____________________________ SCSI Path Info, host:channel:scsi id:LUN
```

The **Path State** refers to the physical state of a path. There are currently 4 states in which a path can be. The *"ready"* state means the path is available to handle IO requests. A *"faulty"* state means the path is currently down and cannot handle IO requests. A *"shaky"* state means the path is available, but for some reason is not capable of handling IO requests. The "*ghost*" state is a passive path in an "active-passive" arrangement.

The device mapper path state is the multipath kernel module's state of a path. An "*active"* status means that the last IO requested through this path completed without incidence. A *"failed"* status means that the last IO requested down the path did not complete and failed.

The device **Major/Minor** numbers are assigned by the Linux kernel and are used to read/write to the device file itself. The major number refers to the device driver type and the minor number refers to the unique device within the device driver type.

Block Device names are used to address the Major/Minor devices. The sdg (in the example above) refers to a SCSI device driver type (sd) and the enumerated, unique device of this type (g).

SCSI Path Information displays the specific attributes of the path to the device. There are four comma separated attributes; host, channel, SCSI id, and LUN id. The host attribute refers to the initiator (Fibre Channel or Infiniband) port within the server. The channel and SCSI id attributes are set within the HBA/HCA interface. The LUN id displays the LUN number of the SCSI device from the presented target device.

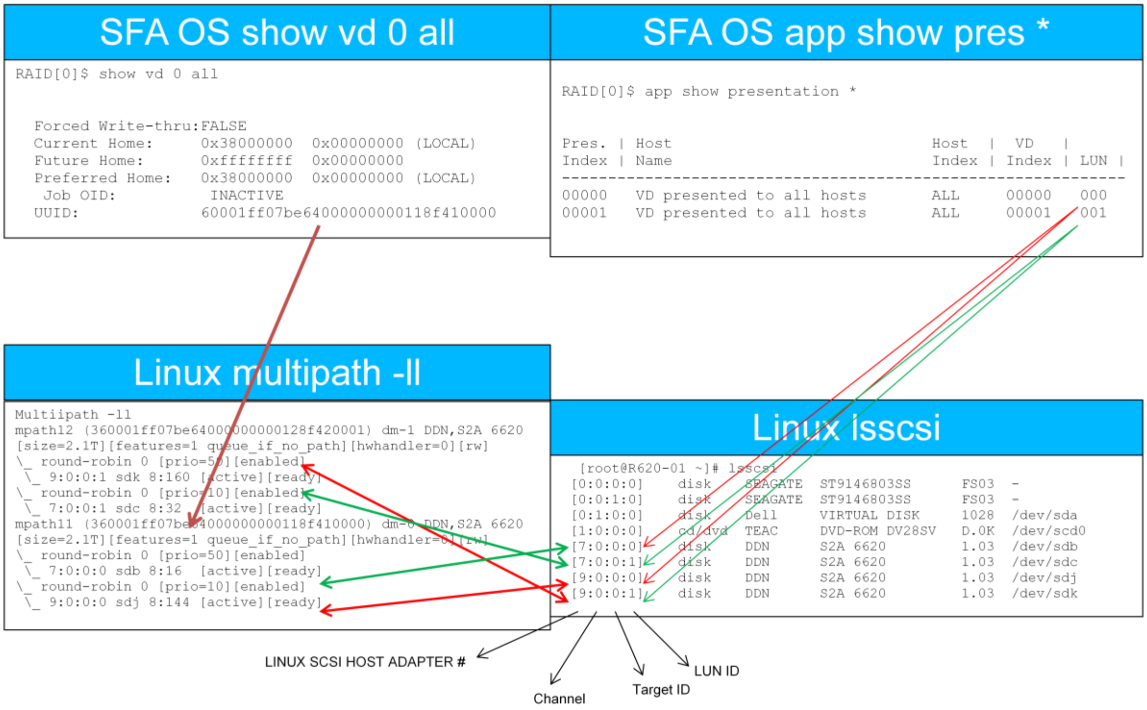
```
# multipath -ll

mpath19 (360001ff07211600000000002688e10002) dm-2 DDN,S2A 6620
[size=21T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=50][active]
 \_ 1:0:0:2 sdg 8:96  [active][ready]
\_ round-robin 0 [prio=10][enabled]
 \_ 2:0:0:2 sdf 8:80  [active][ready]
```

```
Path Group 2
\_ round-robin 0 [prio=10][enabled]
 \_ 2:0:0:2 sdf 8:80  [active][ready]
```

The example above shows a second path group to the same device (referenced in Path Group 1) within the *Priority Group*. This path group has a lower priority (prio=10) than path group 1 (prio=50). The path group state of this path group is “enabled”, meaning it is ready to service IO, but is not the optimal path to send IO requests through. The physical path state is “ready” to accept IO requests and the device mapper path state is “active”, which infers the path was last tested successfully.

8. S2A 6620 EXAMPLE

The figure below identifies the relationships between linux SCSI devices, S2A 6620 virtual disks, presentations, and dm-multipath devices.



```
RAID[0]$ show vd 0 all


  Forced Write-thru:FALSE
  Current Home:     0x38000000  0x00000000 (LOCAL)
  Future Home:      0xffffffff  0x00000000
  Preferred Home:   0x38000000  0x00000000 (LOCAL)
   Job OID:          INACTIVE
  UUID:             60001ff07be6400000000118f410000
```

```
RAID[0]$ app show presentation *


Pres. | Host                          Host  |  VD   |
Index | Name                          Index | Index | LUN |
--------------------------------------------------------------
00000   VD presented to all hosts     ALL     00000   000
00001   VD presented to all hosts     ALL     00001   001
```

```
Multiipath -ll
mpath12 (360001ff07be6400000000128f420001) dm-1 DDN,S2A 6620
[size=2.1T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=50][enabled]
 \_ 9:0:0:1 sdk 8:160 [active][ready]
\_ round-robin 0 [prio=10][enabled]
 \_ 7:0:0:1 sdc 8:32  [active][ready]
mpath11 (360001ff07be6400000000118f410000) dm-0 DDN,S2A 6620
[size=2.1T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=50][enabled]
 \_ 7:0:0:0 sdb 8:16  [active][ready]
\_ round-robin 0 [prio=10][enabled]
 \_ 9:0:0:0 sdj 8:144 [active][ready]
```

```
 [root@R620-01 ~]# lsscsi
[0:0:0:0]    disk    SEAGATE  ST9146803SS      FS03  -
[0:0:1:0]    disk    SEAGATE  ST9146803SS      FS03  -
[0:1:0:0]    disk    Dell     VIRTUAL DISK     1028  /dev/sda
[1:0:0:0]    cd/dvd  TEAC     DVD-ROM DV28SV   D.0K  /dev/scd0
[7:0:0:0]    disk    DDN      S2A 6620         1.03  /dev/sdb
[7:0:0:1]    disk    DDN      S2A 6620         1.03  /dev/sdc
[9:0:0:0]    disk    DDN      S2A 6620         1.03  /dev/sdj
[9:0:0:1]    disk    DDN      S2A 6620         1.03  /dev/sdk
```

9. MULTIPATH -LL OUTPUT WITH SLES 11 AND RHEL 6

Under RHEL 6 and SLES 11, the version of multipath tools now displays the same information with very slight changes to the formatting, as show in the example below:

RHEL 6.0:

```
360001ff0802bd0000000004a8f950003 dm-12 DDN,SFA 10000
size=7.1T features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=100 status=active
| `- 6:0:0:3  sdaf 65:240 active ready running
`-+- policy='round-robin 0' prio=10 status=enabled
  `- 5:0:0:3  sdq  65:0   active ready running
```

SLES 11:

```
360001ff0802bd0000000004a8f950003 dm-9 DDN,SFA 10000
size=7.1T features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='round-robin 0' prio=100 status=active
| `- 6:0:0:3  sdaf 65:240 active ready running
`-+- policy='round-robin 0' prio=10 status=enabled
  `- 5:0:0:3  sdq  65:0   active ready running
```

10. KNOWN ISSUES

10.1 RHEL5X/SLES10X

10.1.1 INFINIBAND SCSI DEVICES THAT DISAPPEAR AND COME BACK ARE NOT ADDED TO MULTIPATH DEVICE MAPS (INFINIBAND ONLY)

Description:
IBSRP changes the host number of the SRP host thereby causing the sysfs path of the device to change. For example LUN 13 = host:bus:target:lun = 6:0:0:13 before the paths fail and then LUN 13 = 7:0:0:13 after the paths come back. Multipath caches the sysfs entries for devices to prevent repeated path lookups. When a device disappears and then comes back with a different host, multipath tries to look up the device using cached (and now incorrect) path. Hence multipath fails to add the device-to-device map.

Solution (RHEL5x):
Install updated multipath provided by RedHat. Multipath >= 0.4.7-46 fixes this problem.

Workaround (SLES10x):
Manually trigger multipath to reload dev maps by either issuing multipath -r or multipath <device-WWN>.

10.2 SLES 10X

10.2.1 CERTAIN MULTIPATH COMMANDS COMPLAIN ABOUT DEPRECATED PRIO_CALLOUT

Description:
Certain versions of multipath print the following warning:
Using deprecated prio_callout '/sbin/mpath_prio_sfa10k /dev/%n' (controller setting)
Example:

```
360001ff0775480000000000db8e400019 dm-9 DDN,S2A 6620
[size=16G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=100][enabled]
\_ 6:0:0:25  sdae 65:224 [active][ready]
sdaj: Using deprecated prio_callout '/sbin/mpath_prio_alua /dev/%n' (controller setting)
      Please fixup /etc/multipath.conf
```

Workaround:
Ignore the warnings.

10.3 VARIOUS

10.3.1 DEFAULT FOR PG_PRIO_CALC HAS CHANGED BETWEEN MULTIPATH VERSION 0.4.8 AND 0.4.9

Description:
The default for pg_prio_calc method has changed from “sum” in 0.4.8 to “avg” in 0.4.9.

With 0.4.8 (pg_prio_calc = “sum”) the priority of the path group is set to the sum of all path weights in the group. Thus in the priority group below the priority is 600, which is the sum of the six available paths which each have an individual path priority of 100.

```
# multipath -ll

360001ff080329000000002df8af00004 dm-8 SGI,DD6A-IS16K-10000
[size=1.0T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=600][enabled]
 \_ 3:0:8:4   sdcg 69:64   [active][ready]
 \_ 3:0:12:4  sddu 71:192  [active][ready]
 \_ 3:0:14:4  sdeo 129:0   [active][ready]
 \_ 4:0:8:4   sdik 135:64  [active][ready]
 \_ 4:0:12:4  sdjy 65:448  [active][ready]
 \_ 4:0:14:4  sdks 67:256  [active][ready]
```

With 0.4.9 (pg_prio_calc = “avg”) the priority is set to the average of the available paths in the path group. Thus in the example below the priority is set to 100 for the priority group, which is the average of the six available paths which each have an individual path priority of 100.

```
# multipath -ll

360001ff080329000000002df8af00004 dm-8 SGI,DD6A-IS16K-10000
[size=1.0T][features=1 queue_if_no_path][hwhandler=0][rw]
\_ round-robin 0 [prio=100][enabled]
 \_ 3:0:8:4   sdcg 69:64   [active][ready]
 \_ 3:0:12:4  sddu 71:192  [active][ready]
 \_ 3:0:14:4  sdeo 129:0   [active][ready]
 \_ 4:0:8:4   sdik 135:64  [active][ready]
 \_ 4:0:12:4  sdjy 65:448  [active][ready]
 \_ 4:0:14:4  sdks 67:256  [active][ready]
```

Workaround:
Edit the /etc/multipath.conf file to add the pg_prio_calc option. The addition must be made to the “default” section of the multipath.conf file, not in a device section.

Example:

```
defaults {
        udev_dir                /dev
        user_friendly_names     no
        pg_prio_calc            avg
}
```

Warning: This setting may impact the default behavior of other vendor storage devices.