

Using Device-Mapper Multipath

5.1

Configuration and Administration

ISBN:

Publication date:

Using Device-Mapper Multipath

This book provides information on using the Device-Mapper Multipath feature of Red Hat Enterprise Linux 5.1

Using Device-Mapper Multipath: Configuration and Administration

Copyright © 2007 Red Hat, Inc.

Copyright © 2007 by Red Hat, Inc. This material may be distributed only subject to the terms and conditions set forth in the Open Publication License, V1.0 or later (the latest version is presently available at <http://www.opencontent.org/openpub/>).

Distribution of substantively modified versions of this document is prohibited without the explicit permission of the copyright holder.

Distribution of the work or derivative of the work in any standard (paper) book form for commercial purposes is prohibited unless prior permission is obtained from the copyright holder.

Red Hat and the Red Hat "Shadow Man" logo are registered trademarks of Red Hat, Inc. in the United States and other countries.

All other trademarks referenced herein are the property of their respective owners.

The GPG fingerprint of the security@redhat.com key is:

CA 20 86 86 2B D6 9D FC 65 F6 EC C4 21 91 80 CD DB 42 A6 0E

1801 Varsity Drive
Raleigh, NC 27606-2072
USA
Phone: +1 919 754 3700
Phone: 888 733 4281
Fax: +1 919 754 3701
PO Box 13588
Research Triangle Park, NC 27709
USA

Preface	vii
1. Audience	vii
2. Related Documentation	vii
3. Feedback	viii
4. Document Conventions	viii
1. Device Mapper Multipathing	1
1. Overview of DM-Multipath	1
2. Storage Array Support	4
3. DM-Multipath Components	5
4. DM-Multipath Setup Overview	5
2. Multipath Devices	7
1. Multipath Device Identifiers	7
2. Consistent Multipath Device Names in a Cluster	7
3. Multipath Device Attributes	8
4. Multipath Devices in Logical Volumes	8
3. Setting Up DM-Multipath	11
1. Setting Up DM-Multipath	11
2. Ignoring Local Disks when Generating Multipath Devices	12
3. Adding Devices to the Multipathing Database	15
4. The DM-Multipath Configuration File	17
1. Configuration File Overview	17
2. Configuration File Blacklist	18
2.1. Blacklisting by WWID	19
2.2. Blacklisting By Device Name	19
2.3. Blacklisting By Device Type	20
2.4. Blacklist Exceptions	20
3. Configuration File Defaults	21
4. Multipaths Device Configuration Attributes	24
5. Configuration File Devices	27
5. DM-Multipath Administration and Troubleshooting	31
1. Multipath Command Output	31
2. Multipath Queries with multipath Command	32
3. Multipath Command Options	33
4. Determining Device Mapper Entries with the dmsetup Command	33
5. Troubleshooting with the multipathd Interactive Console	34
Index	37

Preface

This book describes the Device Mapper Multipath (DM-Multipath) feature of Red Hat Enterprise Linux.

1. Audience

This book is intended to be used by system administrators managing systems running the Linux operating system. It requires familiarity with Red Hat Enterprise Linux.

2. Related Documentation

For more information about using Red Hat Enterprise Linux, refer to the following resources:

- Red Hat Enterprise Linux Installation Guide — Provides information regarding installation of Red Hat Enterprise Linux 5.
- Red Hat Enterprise Linux Deployment Guide — Provides information regarding the deployment, configuration and administration of Red Hat Enterprise Linux 5.

For more information about Red Hat Cluster Suite for Red Hat Enterprise Linux 5, refer to the following resources:

- Red Hat Cluster Suite Overview — Provides a high level overview of the Red Hat Cluster Suite.
- Configuring and Managing a Red Hat Cluster — Provides information about installing, configuring and managing Red Hat Cluster components.
- LVM Administrator's Guide: Configuration and Administration — Provides a description of the Logical Volume Manager (LVM), including information on running LVM in a clustered environment.
- Global File System: Configuration and Administration — Provides information about installing, configuring, and maintaining Red Hat GFS (Red Hat Global File System).
- Using GNBD with Global File System — Provides an overview on using Global Network Block Device (GNBD) with Red Hat GFS.

- Linux Virtual Server Administration — Provides information on configuring high-performance systems and services with the Linux Virtual Server (LVS).
- Red Hat Cluster Suite Release Notes — Provides information about the current release of Red Hat Cluster Suite.

Red Hat Cluster Suite documentation and other Red Hat documents are available in HTML, PDF, and RPM versions on the Red Hat Enterprise Linux Documentation CD and online at <http://www.redhat.com/docs/>.

3. Feedback

If you spot a typo, or if you have thought of a way to make this manual better, we would love to hear from you. Please submit a report in Bugzilla (<http://bugzilla.redhat.com/bugzilla/>) against the component rh-cs.

Be sure to mention the manual's identifier:

```
rh-MPIO(EN)-5.1 (2007-10-30T15:20)
```

By mentioning this manual's identifier, we know exactly which version of the guide you have.

If you have a suggestion for improving the documentation, try to be as specific as possible. If you have found an error, please include the section number and some of the surrounding text so we can find it easily.

4. Document Conventions

Certain words in this manual are represented in different fonts, styles, and weights. This highlighting indicates that the word is part of a specific category. The categories include the following:

Courier font

Courier font represents commands, file names and paths, and prompts .

When shown as below, it indicates computer output:

```
Desktop  about.html  logs  paulwesterberg.png
```


Mail backupfiles mail reports

bold Courier font

Bold Courier font represents text that you are to type, such as: **service jonas start**

If you have to run a command as root, the root prompt (#) precedes the command:

```
# gconftool-2
```

italic Courier font

Italic Courier font represents a variable, such as an installation directory:
install_dir/bin/

bold font

Bold font represents **application programs** and **text found on a graphical interface**.

When shown like this: **OK** , it indicates a button on a graphical application interface.

Additionally, the manual uses different strategies to draw your attention to pieces of information. In order of how critical the information is to you, these items are marked as follows:



Note

A note is typically information that you need to understand the behavior of the system.



Tip

A tip is typically an alternative way of performing a task.



Important

Important information is necessary, but possibly unexpected, such as a configuration change that will not persist after a reboot.



Caution

A caution indicates an act that would violate your support agreement, such as recompiling the kernel.



Warning

A warning indicates potential data loss, as may happen when tuning hardware for maximum performance.

Device Mapper Multipathing

Device Mapper Multipathing (DM-Multipath) allows you to configure multiple I/O paths between server nodes and storage arrays into a single device. These I/O paths are physical SAN connections that can include separate cables, switches, and controllers. Multipathing aggregates the I/O paths, creating a new device that consists of the aggregated paths.

1. Overview of DM-Multipath

DM-Multipath can be used to provide:

- Redundancy

DM-Multipath can provide failover in an active/passive configuration. In an active/passive configuration, only half the paths are used at any time for I/O. If any element of an I/O path (the cable, switch, or controller) fails, DM-Multipath switches to an alternate path.

- Improved Performance

DM-Multipath can be configured in active/active mode, where I/O is spread over the paths in a round-robin fashion. In some configurations, DM-Multipath can detect loading on the I/O paths and dynamically re-balance the load.

Figure 1.1, “Active/Passive Multipath Configuration with One RAID Device” shows an active/passive configuration with two I/O paths from the server to a RAID device. There are 2 HBAs on the server, 2 SAN switches, and 2 RAID controllers.

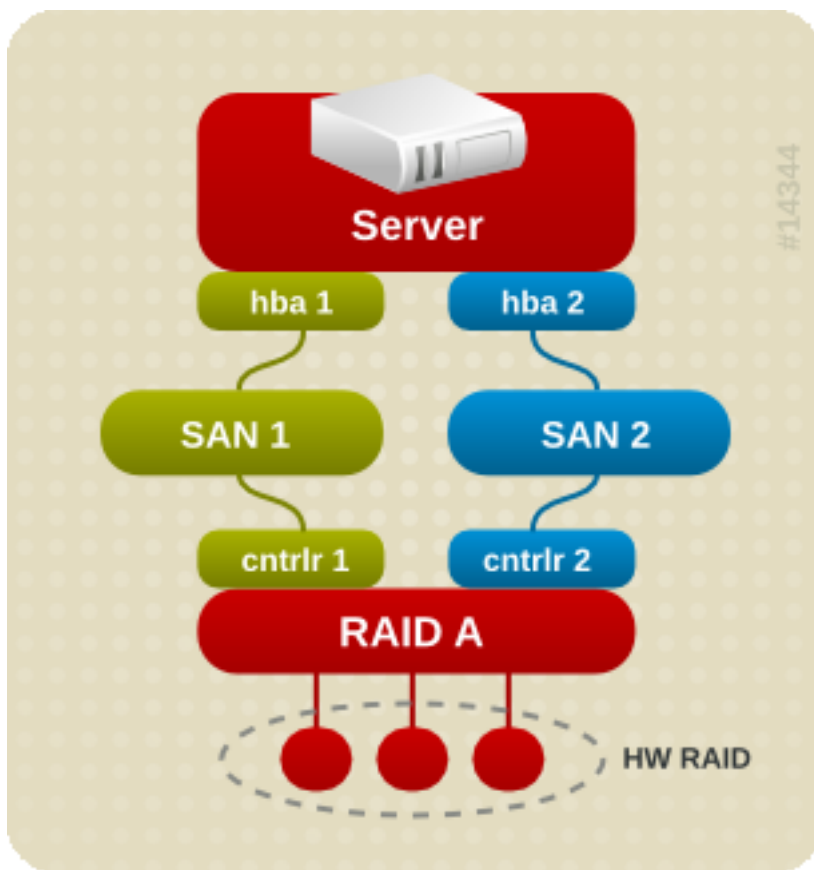


Figure 1.1. Active/Passive Multipath Configuration with One RAID Device

In this configuration, there is one I/O path that goes through hba1, SAN1, and controller 1 and a second I/O path that goes through hba2, SAN2, and controller2. There are many points of possible failure in this configuration:

- HBA failure
- FC cable failure
- SAN switch failure
- Array controller port failure

With DM-Multipath configured, a failure at any of these points will cause DM-Multipath to switch to the alternate I/O path.

Figure 1.2, “Active/Passive Multipath Configuration with Two RAID Devices” shows a

more complex active/passive configuration with 2 HBAs on the server, 2 SAN switches, and 2 RAID devices with 2 RAID controllers each.

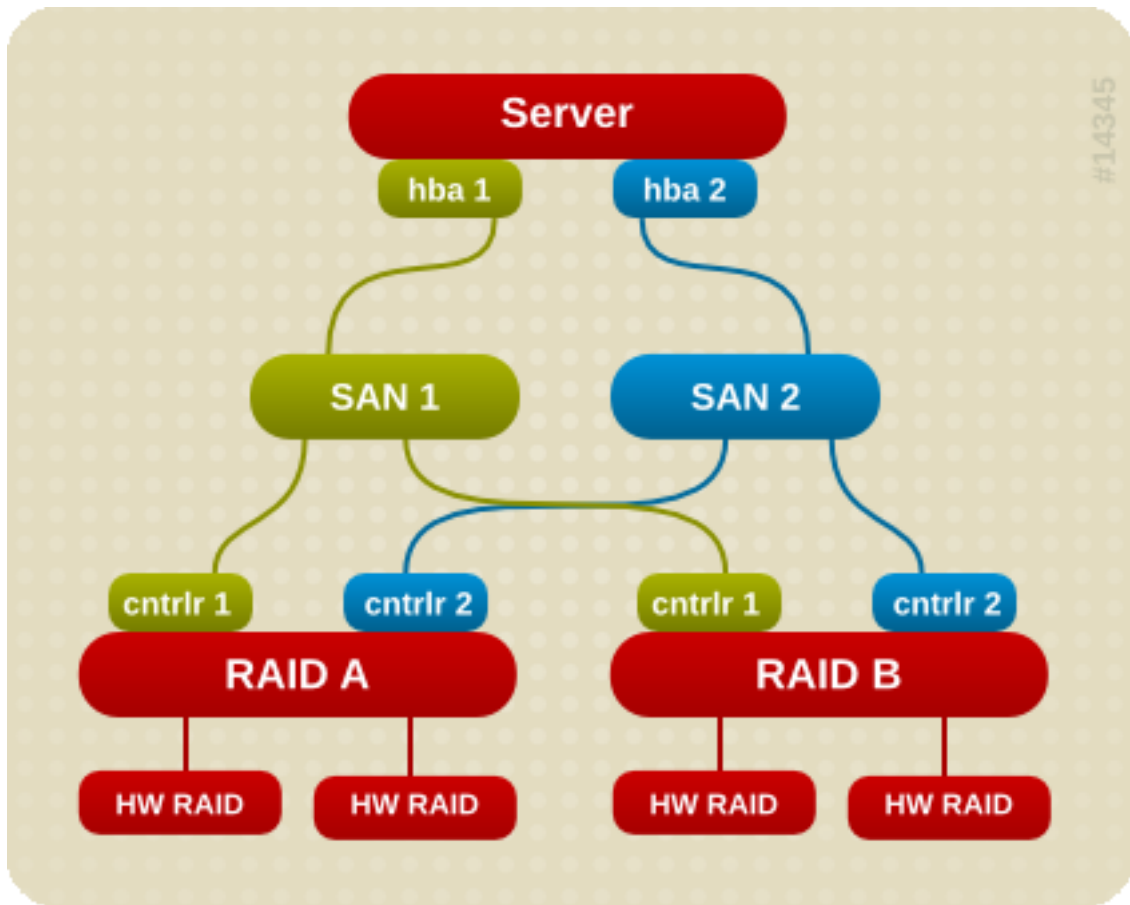


Figure 1.2. Active/Passive Multipath Configuration with Two RAID Devices

As in the example shown in [Figure 1.1, “Active/Passive Multipath Configuration with One RAID Device”](#), there are two I/O paths to each RAID device. With DM-Multipath configured, a failure at any of the points of the I/O path to either of the RAID devices will cause DM-Multipath to switch to the alternate I/O path for that device.

[Figure 1.3, “Active/Active Multipath Configuration with One RAID Device”](#) shows an active/active configuration with 2 HBAs on the server, 1 SAN switch, and 2 RAID controllers. There are four I/O paths from the server to a storage device:

- hba1 to controller1
- hba1 to controller2

- hba2 to controller1
- hba2 to controller2

In this configuration, I/O can be spread among those four paths.

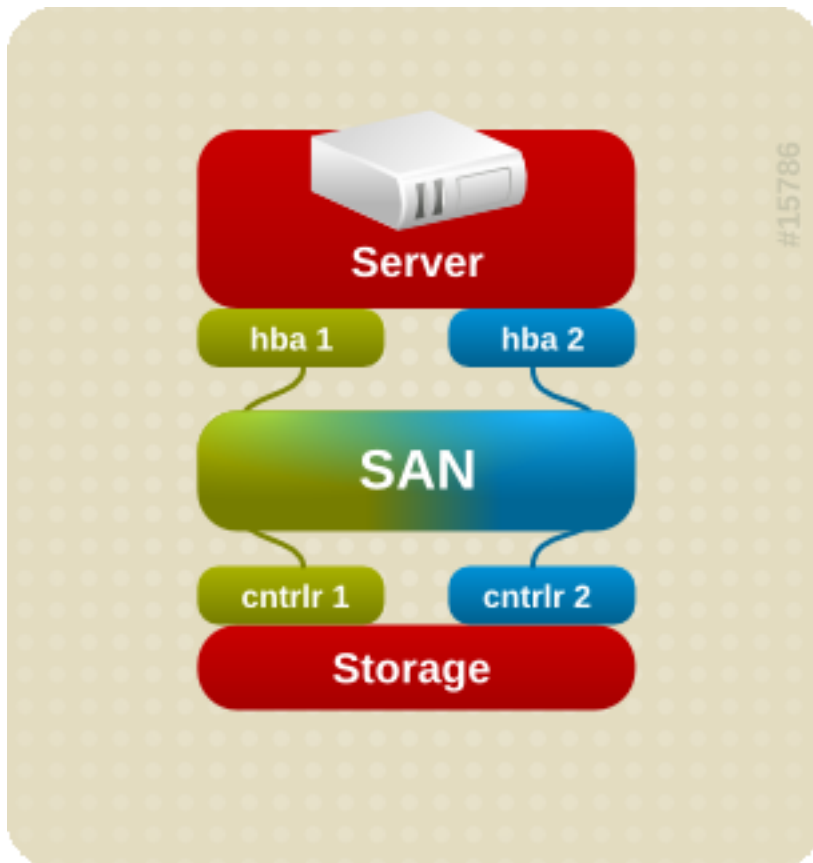


Figure 1.3. Active/Active Multipath Configuration with One RAID Device

2. Storage Array Support

By default, DM-Multipath includes support for the most common storage arrays that support DM-Multipath. The supported devices can be found in the `multipath.conf.defaults` file. If your storage array supports DM-Multipath and is not configured by default in this file, you may need to add them to the DM-Multipath configuration file, `multipath.conf`. For information on the DM-Multipath configuration file, see [Chapter 4, The DM-Multipath Configuration File](#).

Some storage arrays require special handling of I/O errors and path switching.

These require separate hardware handler kernel modules.

3. DM-Multipath Components

Table 1.1, “DM-Multipath Components”. describes the components of DM-Multipath.

Component	Description
dm-multipath kernel module	Reroutes I/O and supports failover for paths and path groups.
multipath command	Lists and configures multipath devices. Normally started up with <code>/etc/rc.sysinit</code> , it can also be started up by a <code>udev</code> program whenever a block device is added or it can be run by the <code>initramfs</code> file system.
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 <code>/etc/multipath.conf</code> file.
kpartx command	Creates device mapper devices for the partitions on a device. It is necessary to use this command for DOS-based partitions with DM-MP. The <code>kpartx</code> is provided in its own package, but the <code>device-mapper-multipath</code> package depends on it.

Table 1.1. DM-Multipath Components

4. DM-Multipath Setup Overview

DM-Multipath includes compiled-in default settings that are suitable for common multipath configurations. Setting up DM-multipath is often a simple procedure.

The basic procedure for configuring your system with DM-Multipath is as follows:

1. Install `device-mapper-multipath` rpm.
2. Edit the `multipath.conf` configuration file:
 - comment out the default blacklist
 - change any of the existing defaults as needed

- save the configuration file
3. Start the multipath daemons.
 4. Create the multipath device with the multipath command.

Detailed setup instructions for several example multipath configurations are provided in see [Chapter 3, Setting Up DM-Multipath](#).

Multipath Devices

Without DM-Multipath, each path from a server node to a storage controller is treated by the system as a separate device, even when the I/O path connects the same server node to the same storage controller. DM-Multipath provides a way of organizing the I/O paths logically, by creating a single multipath device on top of the underlying devices.

1. Multipath Device Identifiers

Each multipath device has a World Wide Identifier (WWID), which is guaranteed to be globally unique and unchanging. By default, the name of a multipath device is set to its WWID. Alternately, you can set the `user_friendly_names` option in the multipath configuration file, which sets the alias to a node-unique name of the form `mpathn`.

For example, a node with two HBAs attached to a storage controller with two ports via a single unzoned FC switch sees four devices: `/dev/sda`, `/dev/sdb`, `dev/sdc`, and `/dev/sdd`. DM-Multipath creates a single device with a unique WWID that reroutes I/O to those four underlying devices according to the multipath configuration. When the `user_friendly_names` configuration option is set to `yes`, the name of the multipath device is set to `/dev/mpath/mpathn`.

For information on the multipath configuration defaults, including the `user_friendly_names` configuration option, see see [Section 3, “Configuration File Defaults”](#).

You can also set the name of a multipath device to a name of your choosing by using the `alias` option in the `multipaths` section of the multipath configuration file. For information on the `multipaths` section of the multipath configuration file, see see [Section 4, “Multipaths Device Configuration Attributes”](#).

2. Consistent Multipath Device Names in a Cluster

When the `user_friendly_names` configuration option is set to `yes`, the name of the multipath device is unique to a node, but it is not guaranteed to be the same on all nodes using the multipath device. This should not cause any difficulties if you use LVM to create logical devices from the multipath device, but if you require that your multipath device names be consistent in every node in the cluster you perform one

of the following procedures:

- Use the `alias` option in the `multipaths` section of the multipath configuration file to set the name of the multipath device. The alias for the multipath device is consistent across all the nodes in a cluster. For information on the `multipaths` section of the multipath configuration file, see [Section 4, “Multipaths Device Configuration Attributes”](#).
- If you want the system-defined user-friendly names to be consistent across all nodes in the cluster, set up all of the multipath devices on one machine. Then copy the `/var/lib/multipath/bindings` file from that machine to all the other machines in the cluster.

3. Multipath Device Attributes

In addition to the `user_friendly_names` and `alias` options, a multipath device has numerous attributes. You can modify these attributes for a specific multipath device by creating an entry for that device in the `multipaths` section of the multipath configuration file. For information on the `multipaths` section of the multipath configuration file, see [Section 4, “Multipaths Device Configuration Attributes”](#).

4. Multipath Devices in Logical Volumes

After creating multipath devices, you can use the multipath device names just as you would use a physical device name when creating an LVM physical volume. For example, if `/dev/mpath/mpath0` is the name of a multipath device, the following command will mark `/dev/mpath/mpath0` as a physical volume.

```
pvccreate /dev/mpath/mpath0
```

You can use the resulting LVM physical device when you create an LVM volume group just as you would use any other LVM physical device.

When you create an LVM logical volume that uses active/passive multipath arrays as the underlying physical devices, you should include filters in the `lvm.conf` to exclude the disks that underlie the multipath devices. This is because if the array automatically changes the active path to the passive path when it receives I/O, multipath will failover and failback whenever LVM scans the passive path if these devices are not filtered. For active/passive arrays that require a command to make

the passive path active, LVM prints a warning message when this occurs.

To filter all SCSI devices in the multipath configuration file (`lvm.conf`), include the following filter in the devices section of the file.

```
filter = [ "r/disk/", "r/sd.*/", "a./*" ]
```


Setting Up DM-Multipath

This chapter provides step-by-step example procedures for configuring DM-Multipath. It includes the following procedures:

- Basic DM-Multipath setup
- Ignoring local disks
- Adding more devices to the configuration file

1. Setting Up DM-Multipath

Before setting up DM-Multipath on your system, ensure that your system has been updated and includes the device-mapper-multipath package.

Use the following procedure to set up DM-Multipath for a basic failover configuration.

1. Edit the `/etc/multipath.conf` file by commenting out the following lines at the top of the file. This section of the configuration file, in its initial state, blacklists all devices. You must comment it out to enable multipathing.

```
blacklist {
    devnode "*"
}
```

After commenting out those lines, this section appears as follows.

```
# blacklist {
#     devnode "*"
# }
```

2. The default settings for DM-Multipath are compiled in to the system and do not need to be explicitly set in the `/etc/multipath.conf` file.

The default value of `path_grouping_policy` is set to `failover`, so in this example you

do not need to change the default value. For information on changing the values in the configuration file to something other than the defaults, see [Chapter 4, The DM-Multipath Configuration File](#).

The initial defaults section of the configuration file configures your system that the names of the multipath devices are of the form `mpathn`; without this setting, the names of the multipath devices would be aliased to the WWID of the device.

3. Save the configuration file and exit the editor.

4. Execute the following commands:

```
modprobe dm-multipath
service multipathd start
multipath -v2
```

The `multipath -v2` command prints out multipathed paths that show which devices are multipathed. If the command does not print anything out, ensure that all SAN connections are set up properly and the system is multipathed.

For further information on the multipath command output, see [Section 1, "Multipath Command Output"](#).

5. Execute the following command to ensure sure that the multipath daemon starts on bootup:

```
chkconfig multipathd on
```

Since the value of `user_friendly_name` is set to `yes` in the configuration file the multipath devices will be created as `/dev/mapper/mpathn`. For information on setting the name of the device to an alias of your choosing, see [Chapter 4, The DM-Multipath Configuration File](#).

2. Ignoring Local Disks when Generating Multipath Devices

Some machines have local SCSI cards for their internal disks. DM-Multipath is not

recommended for these devices. The following procedure shows how to modify the multipath configuration file to ignore the local disks when configuring multipath.

1. Determine which disks are the internal disks and mark them as the ones to blacklist.

In this example, `/dev/sda` is the internal disk. Note that as originally configured in the default multipath configuration file, executing the `multipath -v2` shows the local disk, `/dev/sda`, in the multipath map.

For further information on the multipath command output, see [Section 1, "Multipath Command Output"](#).

```
[root@rh4cluster1 ~]# multipath -v2
create: SIBM-ESXSST336732LC___F3ET0EP0Q000072428BX1
[size=33 GB][features="0"][hwhandler="0"]
\_ round-robin 0
  \_ 0:0:0:0 sda 8:0  [-----]

device-mapper ioctl cmd 9 failed: Invalid argument
device-mapper ioctl cmd 14 failed: No such device or address
create: 3600a0b80001327d80000006d43621677
[size=12 GB][features="0"][hwhandler="0"]
\_ round-robin 0
  \_ 2:0:0:0 sdb 8:16
  \_ 3:0:0:0 sdf 8:80

create: 3600a0b80001327510000009a436215ec
[size=12 GB][features="0"][hwhandler="0"]
\_ round-robin 0
  \_ 2:0:0:1 sdc 8:32
  \_ 3:0:0:1 sdg 8:96

create: 3600a0b80001327d800000070436216b3
[size=12 GB][features="0"][hwhandler="0"]
\_ round-robin 0
  \_ 2:0:0:2 sdd 8:48
  \_ 3:0:0:2 sdh 8:112

create: 3600a0b80001327510000009b4362163e
[size=12 GB][features="0"][hwhandler="0"]
\_ round-robin 0
```

```
\_ 2:0:0:3 sde 8:64  
\_ 3:0:0:3 sdi 8:128
```

2. In order to prevent the device mapper from mapping `/dev/sda` in its multipath maps, edit the blacklist section of the `/etc/multipath.conf` file to include this device. Although you could blacklist the `sda` device using a devnode type, that would not be safe procedure since `/dev/sda` is not guaranteed to be the same on reboot. To blacklist individual devices, you can blacklist using the WWID of that device.

Note that in the output to the `multipath -vs` command, the WWID of the `/dev/sda` device is `SIBM-ESXSST336732LC___F3ET0EP0Q000072428BX1`. To blacklist this device, include the following in the `/etc/multipath.conf` file.

```
blacklist {  
    wwid SIBM-ESXSST336732LC___F3ET0EP0Q000072428BX1  
}
```

3. Run the following commands:

```
multipath -F  
multipath -v2
```

The local disk or disks should no longer be listed in the new multipath maps, as shown in the following example.

```
[root@rh4cluster1 ~]# multipath -F  
[root@rh4cluster1 ~]# multipath -v2  
create: 3600a0b80001327d80000006d43621677  
[size=12 GB][features="0"][hwhandler="0"]  
\_ round-robin 0  
  \_ 2:0:0:0 sdb 8:16  
  \_ 3:0:0:0 sdf 8:80  
  
create: 3600a0b80001327510000009a436215ec  
[size=12 GB][features="0"][hwhandler="0"]  
\_ round-robin 0  
  \_ 2:0:0:1 sdc 8:32  
  \_ 3:0:0:1 sdg 8:96
```



```
create: 3600a0b80001327d800000070436216b3
[size=12 GB][features="0"][hwhandler="0"]
\_ round-robin 0
  \_ 2:0:0:2 sdd 8:48
  \_ 3:0:0:2 sdh 8:112

create: 3600a0b80001327510000009b4362163e
[size=12 GB][features="0"][hwhandler="0"]
\_ round-robin 0
  \_ 2:0:0:3 sde 8:64
  \_ 3:0:0:3 sdi 8:128
```

3. Adding Devices to the Multipathing Database

By default, DM-Multipath includes support for the most common storage arrays that support DM-Multipath. The default configuration values, including supported devices, can be found in the `multipath.conf.defaults` file.

If you need to add a storage device that is not supported by default as a known multipath device, edit the `/etc/multipath.conf` file and insert the appropriate device information.

For example, to add information about the HP Open-V series the entry looks like this:

```
devices {
  device {
    vendor "HP"
    product "OPEN-V."
    getuid_callout "/sbin/scsi_id -g -u -p0x80 -s /block/%n"
  }
}
```

For more information on the devices section of the configuration file, see [Section 5, “Configuration File Devices”](#).

The DM-Multipath Configuration File

By default, DM-Multipath provides configuration values for the most common uses of multipathing. In addition, DM-Multipath includes support for the most common storage arrays that support DM-Multipath. The default configuration values and the supported devices can be found in the

`/usr/share/doc/device-mapper-multipath-0.4.7/multipath.conf.defaults` file.

You can override the default configuration values for DM-Multipath by editing the `/etc/multipath.conf` configuration file. If necessary, you can also add a storage array that is not supported by default to the configuration file. This chapter provides information on parsing and modifying the `multipath.conf` file. It contains sections on the following topics:

- Configuration file overview
- Configuration file blacklist
- Configuration file defaults
- Configuration file multipaths
- Configuration file devices

In the `multipath` configuration file, you need to specify only the sections that you need for your configuration, or that you wish to change from the default values specified in the `multipath.conf.defaults` file. If there are sections of the file that are not relevant to your environment or for which you do not need to override the default values, you can leave them commented out, as they are in the initial file.

The configuration file allows regular expression description syntax.

An annotated version of the configuration file can be found in

`/usr/share/doc/device-mapper-multipathd-0.4.7/multipath.conf.annotated`.

1. Configuration File Overview

The `multipath` configuration file is divided into the following sections:

blacklist

Listing of specific devices that will not be considered for multipath. By default all devices are blacklisted. Usually the default blacklist section is commented out.

blacklist_exceptions

Listing of multipath candidates that would otherwise be blacklisted according to the parameters of the blacklist section.

defaults

General default settings for DM-Multipath.

multipaths

Settings for the characteristics of individual multipath devices. These values overwrite what is specified in the defaults and devices sections of the configuration file.

devices

Settings for the individual storage controllers. These values overwrite what is specified in the defaults section of the configuration file. If you are using a storage array that is not supported by default, you may need to create a devices subsection for your array.

When the system determines the attributes of a multipath device, first it checks the multipath settings, then the per devices settings, then the multipath system defaults.

2. Configuration File Blacklist

The blacklist section of the multipath configuration file specifies the devices that will not be used when the system configures multipath devices. Devices that are blacklisted will not be grouped into a multipath device.

By default, all devices are blacklisted, since the following lines appear in the initial configuration file.

```
blacklist {
    devnode "*"
}
```

To enable multipathing on all of the devices that are supported by default, comment

out those lines, as described in [Section 1, “Setting Up DM-Multipath”](#).

After commenting out the universal blacklist, you can specify general device types and individual devices to blacklist. You can blacklist devices according to the following criteria:

- By WWID, as described in [Section 2.1, “Blacklisting by WWID”](#)
- By device name, as described in [Section 2.2, “Blacklisting By Device Name”](#)
- By device type, as described in [Section 2.3, “Blacklisting By Device Type”](#)

By default, a variety of device types are blacklisted, even after you comment out the initial blacklist section of the configuration file. For information, see [Section 2.2, “Blacklisting By Device Name”](#).

2.1. Blacklisting by WWID

You can specify individual devices to blacklist by their World-Wide IDentification with a `wwid` entry in the blacklist section of the configuration file.

The following example shows the lines in the configuration file that would blacklist a device with a WWID of 26353900f02796769.

```
blacklist {
    wwid 26353900f02796769
}
```

2.2. Blacklisting By Device Name

You can blacklist device types by device name so that they will not be grouped into a multipath device by specifying a `devnode` entry in the blacklist section of the configuration file.

The following example shows the lines in the configuration file that would blacklist all SCSI devices, since it blacklists are `sd*` devices.

```
blacklist {
    devnode "^sd[a-z]"
}
```

You can use a devnode entry in the blacklist section of the configuration file to specify individual devices to blacklist rather than all devices of specific type; this is not recommended, however. Unless it is statically mapped by udev rules, there is no guarantee that a specific device will have the same name on reboot. For example, a device name could change from `/dev/sda` to `/dev/sdb` on reboot.

By default, the following devnode entries are compiled in the default blacklist; the devices that these entries blacklist do not generally support DM-Multipath. To enable multipathing on any of these devices, you would need to specify them in the `blacklist_exceptions` section of the configuration file, as described in [Section 2.4, “Blacklist Exceptions”](#).

```
blacklist {
    devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*"
    devnode "^hd[a-z]"
}
```

2.3. Blacklisting By Device Type

You can specify specific device types in the blacklist section of the configuration file with a device section. The following example blacklists all IBM DS4200 device and all HP devices.

```
blacklist {
    device {
        vendor "IBM"
        product "3S42"    #DS4200 Product 10
    }
    device {
        vendor "HP"
        product "*"
    }
}
```

2.4. Blacklist Exceptions

You can use the `blacklist_exceptions` section of the configuration file to enable

multipathing on devices that have been blacklisted by default. For example, by default, devices with vendor = "IBM" and product = "S/390.*" are blacklisted since usually DASD devices are used for local storage. If you do need to enable multipathing on these devices, uncomment the following lines in the configuration file:

```
blacklist_exceptions {
    device {
        vendor "IBM"
        product "S/390.*"
    }
}
```

When specifying devices in the `blacklist_exceptions` section of the configuration file, you must specify the exceptions in the same way they were specified in the blacklist. For example, a WWID exception will not apply to devices specified by a devnode blacklist entry, even if the blacklisted device is associated with that WWID. Similarly, devnode exceptions apply only to devnode entries, and device exceptions apply only to device entries.

3. Configuration File Defaults

The `/etc/multipath.conf` configuration file includes a defaults section that sets the `user_friendly_names` parameter to yes, as follows.

```
defaults {
    user_friendly_names yes
}
```

This overwrites the default value of the `user_friendly_names` parameter.

The configuration file includes a template of configuration defaults. This section is commented out, as follows.

```
#defaults {
#    udev_dir          /dev
#    polling_interval  10
#    selector          "round-robin 0"
#    path_grouping_policy multibus
```

```
# getuid_callout    "/sbin/scsi_id -g -u -s /block/%n"
# prio_callout     /bin/true
# path_checker     readsector0
# rr_min_io        100
# rr_weight        priorities
# failback         immediate
# no_path_retry    fail
# user_friendly_name  yes
# }
```

To overwrite the default value for any of the configuration parameters, you can copy the relevant line from this template into the defaults section and uncomment it. For example, to overwrite the `path_grouping_policy` parameter so that it is `multibus` rather than the default value of `failover`, copy the appropriate line from the template to the initial defaults section of the configuration file, and uncomment it, as follows.

```
defaults {
    user_friendly_names  yes
    path_grouping_policy multibus
}
```

[Table 4.1, “Multipath Configuration Defaults”](#) describes the attributes that are set in the defaults section of the `multipath.conf` configuration file. These values are used by DM-Multipath unless they are overwritten by the attributes specified in the `devices` and `multipaths` sections of the `multipath.conf` file.

Attribute	Description
<code>udev_dir</code>	Specifies the directory where <code>udev</code> device nodes are created. The default value is <code>/udev</code> .
<code>polling_interval</code>	Specifies the interval between two path checks in seconds. The default value is 5.
<code>selector</code>	Specifies the default algorithm to use in determining what path to use for the next I/O operation. The default value is <code>round-robin 0</code> .
<code>path_grouping_policy</code>	Specifies the default path grouping policy to apply to unspecified multipaths. Possible values include: <code>failover = 1 path per priority group</code>

Attribute	Description
	<p>multibus = all valid paths in 1 priority group group_by_serial = 1 priority group per detected serial number group_by_prio = 1 priority group per path priority value group_by_node_name = 1 priority group per target node name The default value is failover.</p>
getuid_callout	<p>Specifies the default program and arguments to call out to obtain a unique path identifier. An absolute path is required. The default value is /sbin/scsi_id -g -u -s.</p>
prio_callout	<p>Specifies the the default program and arguments to call out to obtain a path priority value. For example, the ALUA bits in SPC-3 provide an exploitable prio value for example. "none" is a valid value. The default value is no callout, indicating all paths are equal</p>
path_checker	<p>Specifies the default method used to determine the state of the paths. Possible values include: readsector0, tur, emc_clariion, hp_sw, and directio. The default value is readsector0.</p>
rr_min_io	<p>Specifies the number of I/O requests to route to a path before switching to the next path in the current path group. The default value is 1000.</p>
rr_weight	<p>If set to priorities, then instead of sending rr_min_io requests to a path before calling selector to choose the next path, the number of requests to send is determined by rr_min_io times the path's priority, as determined by the prio_callout program. Currently, there are priority callouts only for devices that use the group_by_prio path grouping policy, which means that all the paths in a path group will always have the same priority. If set to uniform, all path weights are equal. The default value is uniform.</p>
failback	<p>Specifies path group failback.</p>

Attribute	Description
	<p>A value of 0 or immediate specifies that as soon as there is a path group with a higher priority than the current path group the system switches to that path group.</p> <p>A numeric value greater than zero specifies deferred failback, expressed in seconds.</p> <p>A value of manual specifies that failback can happen only with operator intervention.</p> <p>The default value is immediate.</p>
no_path_retry	<p>A numeric value for this attribute specifies the number of times the system should attempt to use a failed path before disabling queuing.</p> <p>A value of fail indicates immediate failure, without queuing.</p> <p>A value of queue indicates that queuing should not stop until the path is fixed.</p> <p>The default value is (null).</p>
user_friendly_names	<p>If set to yes, specifies that the system should use the bindings file <code>/var/lib/multipath/bindings</code> to assign a persistent and unique alias to the multipath, in the form of <code>mpathn</code>. If set to no, specifies that the system should use the WWID as the alias for the multipath. In either case, what is specified here will be overridden by any device-specific aliases you specify in the <code>multipaths</code> section of the configuration file. The default value is no.</p>

Table 4.1. Multipath Configuration Defaults

4. Multipaths Device Configuration Attributes

Table 4.2, “Multipath Attributes” shows the attributes that you can set in the `multipaths` section of the `multipath.conf` configuration file for each specific multipath device. These attributes apply only to the one specified multipath. These defaults are used by DM-Multipath and override attributes set in the `defaults` and `devices` sections of the `multipath.conf` file.

Attribute	Description
wwid	Specifies the WWID of the multipath device to which the multipath attributes apply.
alias	Specifies the symbolic name for the multipath device to which the multipath attributes apply.
path_grouping_policy	<p>Specifies the default path grouping policy to apply to unspecified multipaths. Possible values include:</p> <ul style="list-style-type: none"> failover = 1 path per priority group multibus = all valid paths in 1 priority group group_by_serial = 1 priority group per detected serial number group_by_prio = 1 priority group per path priority value group_by_node_name = 1 priority group per target node name
path_selector	Specifies the default algorithm to use in determining what path to use for the next I/O operation.
failback	<p>Specifies path group failback.</p> <p>A value of 0 or immediate specifies that as soon as there is a path group with a higher priority than the current path group the system switches to that path group.</p> <p>A numeric value greater than zero specifies deferred failback, expressed in seconds.</p> <p>A value of manual specifies that failback can happen only with operator intervention.</p>
rr_weight	<p>If set to priorities, then instead of sending rr_min_io requests to a path before calling selector to choose the next path, the number of requests to send is determined by rr_min_io times the path's priority, as determined by the prio_callout program. Currently, there are priority callouts only for devices that use the group_by_prio path grouping policy, which means that all the paths in a path group will always have the same priority.</p> <p>If set to uniform, all path weights are equal.</p>
no_path_retry	

Attribute	Description
	<p>A numeric value for this attribute specifies the number of times the system should attempt to use a failed path before disabling queueing.</p> <p>A value of fail indicates immediate failure, without queueing.</p> <p>A value of queue indicates that queueing should not stop until the path is fixed.</p>
rr_min_io	Specifies the number of I/O requests to route to a path before switching to the next path in the current path group.

Table 4.2. Multipath Attributes

The following example shows multipath attributes specified in the configuration file for two specific multipath devices. The first device has a WWID of 3600508b4000156d70001200000b0000 and a symbolic name of alias.

The second multipath device in the example has a WWID of 1DEC____321816758474 and a symbolic name of red. In this example, the rr_weight attributes is set to priorities.

```
multipaths {
  multipath {
    wwid          3600508b4000156d70001200000b0000
    alias         yellow
    path_grouping_policy multibus
    path_checker  readsector0
    path_selector "round-robin 0"
    failback      manual
    rr_weight     priorities
    no_path_retry 5
  }
  multipath {
    wwid          1DEC____321816758474
    alias         red
    rr_weight     priorities
  }
}
```

5. Configuration File Devices

Table 4.3, “Device Attributes” shows the attributes that you can set for each individual storage device in the devices section of the multipath.conf configuration file. These attributes are used by DM-Multipath unless they are overwritten by the attributes specified in the multipaths section of the multipath.conf file for paths that contain the device. These attributes override the attributes set in the defaults section of the multipath.conf file.

Many devices that support multipathing are included by default in a multipath configuration. The values for the devices that are supported by default are listed in the multipath.conf.defaults file. You probably will not need to modify the values for these devices, but if you do you can overwrite the default values by including an entry in the the configuration file for the device that overwrites those values. You can copy the device configuration defaults from the multipath.conf.defaults file for the device and override the values that you want to change.

To add a device to this section of the configuration file that is not configured automatically by default, you need to set the vendor and product parameters. You can find these values by looking at `/sys/block/device_name>/device/vendor` and `/sys/block/device_name>/device/model` where `device_name` is the device to be multipathed, as in the following example:

```
[root@cypher-06 ~]# cat /sys/block/sda/device/vendor
WINSYS
[root@cypher-06 ~]# cat /sys/block/sda/device/model
SF2372
```

The additional parameters to specify depend on your specific device. If the device is active/active, you will usually not need to set additional parameters. You may want to set `path_grouping_policy` to `multibus`. Other parameters you may need to set are `no_path_retry` and `rr_min_io`, as described in Table 4.3, “Device Attributes”.

If the device is active/passive, but it automatically switches paths with I/O to the passive path, you need to change the checker function to one that does not send IO/ to the path to test if it is working (otherwise, your device will keep failing over). This almost always means that you set the `path_checker` to `tur`; this works for all SCSI devices that support the Test Unit Ready command, which most do.

If the device needs a special command to switch paths, then configuring this device for multipath requires a hardware handler kernel module. The current hardware handlers are emc and "rdac. If these are not sufficient for your device, you may not be able to configure the device for multipath.

Attribute	Description
vendor	Specifies the vendor name of the storage device to which the device attributes apply, for example COMPAQ .
product	Specifies the product name of the storage device to which the device attributes apply, for example HSV110 (C)COMPAQ .
product_blacklist	Specifies a regular expression used to blacklist devices by product.
path_grouping_policy	Specifies the default path grouping policy to apply to unspecified multipaths. Possible values include: failover = 1 path per priority group multibus = all valid paths in 1 priority group group_by_serial = 1 priority group per detected serial number group_by_prio = 1 priority group per path priority value group_by_node_name = 1 priority group per target node name
getuid_callout	Specifies the default program and arguments to call out to obtain a unique path identifier. An absolute path is required.
prio_callout	Specifies the the default program and arguments to call out to obtain a path weight. Weights are summed for each path group to determine the next path group to use in case of failue. "none" is a valid value.
path_checker	Specifies the default method used to determine the state of the paths. Possible values include readsector0 and tur, emc_clariion, hp_sw, and directio.
path_selector	Specifies the default algorithm to use in determining what path to use for the next I/O operation.
failback	Specifies path group failback.

Attribute	Description
	<p>A value of 0 or immediate specifies that as soon as there is a path group with a higher priority than the current path group the system switches to that path group.</p> <p>A numeric value greater than zero specifies deferred failback, expressed in seconds.</p> <p>A value of manual specifies that failback can happen only with operator intervention.</p>
features	<p>The extra features of multipath devices. The only existing feature is queue_if_no_path, which is the same as setting no_path_retry to queue.</p>
hardware_handler	<p>Specifies a module that will be used to perform hardware specific actions when switching path groups or handling I/O errors. Possible values include 0, 1 emc, and 1 rdac. The default value is 0.</p>
rr_min_io	<p>Specifies the number of I/O requests to route to a path before switching to the next path in the current path group. The default value is 1000.</p>
rr_weight	<p>If set to priorities, then instead of sending rr_min_io requests to a path before calling selector to choose the next path, the number of requests to send is determined by rr_min_io times the path's priority, as determined by the prio_callout program. Currently, there are priority callouts only for devices that use the group_by_prio path grouping policy, which means that all the paths in a path group will always have the same priority.</p> <p>If set to uniform, all path weights are equal. The default value is uniform.</p>
no_path_retry	<p>A numeric value for this attribute specifies the number of times the system should attempt to use a failed path before disabling queuing.</p> <p>A value of fail indicates immediate failure, without queuing.</p> <p>A value of queue indicates that queuing should not stop until the path is fixed.</p>

Attribute	Description
	The default value is (null).

Table 4.3. Device Attributes

The following example shows a device entry in the multipath configuration file.

```
# }
# device {
#     vendor          "COMPAQ "
#     product         "MSA1000  "
#     path_grouping_policy multibus
#     path_checker    tur
#     rr_weight       priorities
# }
#}
```


DM-Multipath Administration and Troubleshooting

This chapter will provide information on administering DM-Multipath on a running system. It includes sections on the following topics:

- Multipath Command Output
- Multipath Queries with multipath Command
- Multipath Command Options
- Multipath Queries with dmsetup Command
- Troubleshooting with the multipathd Interactive Console

1. Multipath Command Output

When you create, modify, or list a multipath device, you get a printout of the current device setup. The format is as follows.

For each multipath device:

```
action_if_any: alias (wwid_if_different_from_alias) [size][features][hardware_handler]
```

For each path group:

```
\_ scheduling_policy [path_group_priority_if_known] [path_group_status_if_known]
```

For each path:

```
\_ host:channel:id:lun devnode major:minor [path_status] [dm_status_if_known]
```

For example, the output of a multipath command might appear as follows:

```
mpath1 (3600d0230003228bc000339414edb8101) [size=10
GB][features="0"][hwhandler="0"]
\_ round-robin 0 [prio=1][active]
\_ 2:0:0:6 sdb 8:16 [active][ready]
\_ round-robin 0 [prio=1][enabled]
\_ 3:0:0:6 sdc 8:64 [active][ready]
```

If the path is up and ready for I/O, the status of the path is ready or active. If the path is down, the status is faulty or failed. The path status is updated periodically by the multipathd daemon based on the polling interval defined in the `/etc/multipath.conf` file.

The dm status is similar to the path status, but from the kernel's point of view. The dm status has two states: failed, which is analogous to faulty, and active which covers all other path states. Occasionally, the path state and the dm state of a device will temporarily not agree.



Note

When a multipath device is being created or modified, the path group status and the dm status are not known. Also, the features are not always correct. When a multipath device is being listed, the path group priority is not known.

2. Multipath Queries with multipath Command

You can use the `-l` and `-ll` options of the `multipath` command to display the current multipath configuration. The `-l` option displays multipath topology gathered from information in `sysfs` and the device mapper. The `-ll` option displays the information the `-l` displays in addition to all other available components of the system.

When displaying the multipath configuration, there are three verbosity levels you can specify with the `-v` option of the `multipath` command. Specifying `-v0` yields no output. Specifying `-v1` outputs the created or updated multipath names only, which you can then feed to other tools such as `kpartx`. Specifying `-v2` prints all detected paths, multipaths, and device maps.

The following example shows the output of a `multipath -l` command.

```
# multipath -l
```

```

mpath1 (3600d0230003228bc000339414edb8101)
[size=10 GB][features="0"][hwhandler="0"]
  \_ round-robin 0 [prio=1][active]
  \_ 2:0:0:6 sdb 8:16 [active][ready]
  \_ round-robin 0 [prio=1][enabled]
  \_ 3:0:0:6 sdc 8:64 [active][ready]
    
```

3. Multipath Command Options

Table 5.1, “Useful multipath Command Options” describes some options of the multipath command that you may find useful.

Option	Description
-l	Display the current multipath configuration gathered from sysfs and the device mapper.
-ll	Display the current multipath configuration gathered from sysfs, the device mapper, and all other available components on the system.
-f device	Remove the named multipath device.
-F	Remove all multipath devices.

Table 5.1. Useful multipath Command Options

4. Determining Device Mapper Entries with the dmsetup Command

You can use the dmsetup command to find out which device mapper entries match the multipathed devices.

The following command displays all the device mapper devices and their major and minor numbers. The minor numbers determine the name of the dm device. For example, a minor number of 3 corresponds to the multipathed device /dev/dm-3.

```

# dmsetup ls
mpath2 (253, 4)
mpath4p1 (253, 12)
mpath5p1 (253, 11)
mpath1 (253, 3)
    
```

```
mpath6p1    (253, 14)
mpath7p1    (253, 13)
mpath0      (253, 2)
mpath7      (253, 9)
mpath6      (253, 8)
VolGroup00-LogVol01  (253, 1)
mpath5      (253, 7)
VolGroup00-LogVol00  (253, 0)
mpath4      (253, 6)
mpath1p1    (253, 10)
mpath3      (253, 5)
```

5. Troubleshooting with the multipathd Interactive Console

The `multipathd -k` command is an interactive interface to the `multipathd` daemon. Entering this command brings up an interactive multipath console. After entering this command, you can enter `help` to get a list of available commands, you can enter an interactive command, or you can enter `CTRL-D` to quit.

The `multipathd` interactive console can be used to troubleshoot problems you may be having with your system. For example, the following command sequence displays the multipath configuration, including the defaults, before exiting the console.

```
# multipathd -k
> > show config
> > CTRL-D
```

The following command sequence ensures that multipath has picked up any changes to the `multipath.conf`,

```
# multipathd -k
> > reconfigure
> > CTRL-D
```

Use the following command sequence to ensure that the path checker is working properly.

```
# multipathd -k  
> > show paths  
> > CTRL-D
```

Index

Symbols

/etc/multipath.conf package, 11

A

active/active configuration

definition, 1

illustration, 3

active/passive configuration

definition, 1

illustration, 1

alias parameter, 24

configuration file, 7

B

blacklist

configuration file, 18

default devices, 20

device name, 19

device type, 20

in configuration file, 11

WWID, 19

blacklist_exceptions section

multipath.conf file, 20

C

chkconfig command, 12

configuration file

alias parameter, 24

blacklist, 11, 18

failback parameter, 22, 24, 28

features parameter, 28

getuid_callout parameter, 22, 28

hardware_handler parameter, 28

no_path_retry parameter, 22, 24, 28

overview, 17

path_checker parameter, 22, 28

path_grouping_policy parameter, 22,
24, 28

path_selector parameter, 24, 28

polling-interval parameter, 22

prio_callout parameter, 22, 28

product parameter, 28

product_blacklist parameter, 28

rr_min_io parameter, 22, 24

rr_weight parameter, 22, 24, 28

selector parameter, 22

udev_dir parameter, 22

user_friendly_names parameter, 21,
22

vendor parameter, 28

wwid parameter, 24

configuring

DM-Multipath, 11

D

defaults section

multipath.conf file, 21

device name, 7

device-mapper-multipath package, 11

devices

adding, 15, 27

devices section

multipath.conf file, 27

DM-Multipath

and LVM, 8, 8

components, 5

configuration file, 17

configuring, 11

definition, 1

device name, 7

devices, 7

failover, 1

overview, 1

redundancy, 1

setup, 11

setup, overview, 5

dm-multipath kernel module , 5
dmsetup command, determining device mapper entries, 33

F

failback parameter, 22, 24, 28
failover, 1
features parameter, 28
feedback, viii, viii

G

getuid_callout parameter, 22, 28

H

hardware_handler parameter, 28

K

kpartx command , 5

L

local disks, ignoring, 12
LVM physical volumes
 multipath devices, 8
lvm.conf file , 8

M

modprobe command, 12
multipath command , 5, 12
 options, 33
 output, 31
 queries, 32
multipath devices, 7
 logical volumes, 8
 LVM physical volumes, 8
multipath.conf file, 4, 17
 blacklist_exceptions section, 20
 defaults section, 21
 devices section, 27
 multipaths section, 24
multipath.conf.annotated file, 17

multipath.conf.defaults file, 4, 17

multipathd

 command, 34
 interactive console, 34

multipathd daemon , 5

multipathd start command, 12

multipaths section

 multipath.conf file, 24

N

no_path_retry parameter, 22, 24, 28

P

path_checker parameter, 22, 28
path_grouping_policy parameter, 22, 24, 28
path_selector parameter, 24, 28
polling_interval parameter, 22
prio_callout parameter, 22, 28
product parameter, 28
product_blacklist parameter, 28

R

rr_min_io parameter, 22, 24
rr_weight parameter, 22, 24, 28

S

selector parameter, 22
setup
 DM-Multipath, 11
storage array support, 4
storage arrays
 adding, 15, 27

U

udev_dir parameter, 22
user_friendly_names parameter , 7, 21, 22

V

vendor parameter, 28

W

World Wide Identifier (WWID), 7

wwid parameter, 24

