

NeoKylin 3.2.1

DM 多路径配置 及管理



湖南麒麟信息工程技术有限公司

1. 设备映射器多路径	
1.1. DM-Multipath 总览	1
1.2. 存储阵列支持	3
1.3. DM-Multipath 组件	3
1.4. DM-Multipath 安装总览	3
2. 多路径设备	
2.1. 多路径设备识别器	5
2.2. 在群集中保持多路径设备名称一致	5
2.3. 多路径设备属性	5
2.4. 逻辑卷中的多路径设备	6
3. 安装 DM-Multipath	
3.1. 安装 DM-Multipath	7
3.2. 在生成多路径设备时忽略逻辑磁盘	8
3.3. 在多路径数据库中添加设备	9
4. DM-Multipath 配置文件	
4.1. 配置文件总览	11
4.2. 配置文件黑名单	11
4.2.1. 根据 WWID 将设备列入黑名单	12
4.2.2. 根据设备名称将设备列入黑名单	12
4.2.3. 根据设备类型将其加入黑名单	13
4.2.4. 黑名单之外的设备	13
4.3. 配置文件默认设置	13
4.4. 多路径设备配置属性	15
4.5. 配置文件设备	17
5. DM-Multipath 管理及故障排除	
5.1. The Multipath Daemon	21
5.2. Issues with Large Number of LUNs	21
5.3. Issues with queue_if_no_path feature	21
5.4. Multipath Command Output	22
5.5. Multipath Queries with multipath Command	22
5.6. Multipath Command Options	23
5.7. 使用 dmsetup 命令确定设备映射器条目	23
5.8. Troubleshooting with the multipathd Interactive Console	23

设备映射器多路径

多路径设备映射器（DM-Multipath）允许您将服务器节点和存储阵列间的多个 I/O 路径配置为一个单一设备。这些 I/O 路径是可包含独立电缆、交换机以及控制器的物理设备。多路径集合了 I/O 路径，并生成一个包整合路径的新设备。

1.1.DM-Multipath 总览

可使用 DM-Multipath 提供：

冗余

DM-Multipath 可在主动/被动配置中提供出错冗余。在主动/被动配置中，只有一半的路径在每次 I/O 时都使用。如果 I/O 路径的任意元素（电缆、交换机或者控制器）出现故障，就会将 DM-Multipath 切换到备用路径。

提高的性能

DM-Multipath 可配置为主动/主动模式，其中将 I/O 以调度算法方式分布到所有路径中。在有些配置中，DM-Multipath 可在 I/O 路径中检测负载并动态对负载进行再平衡。

图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.

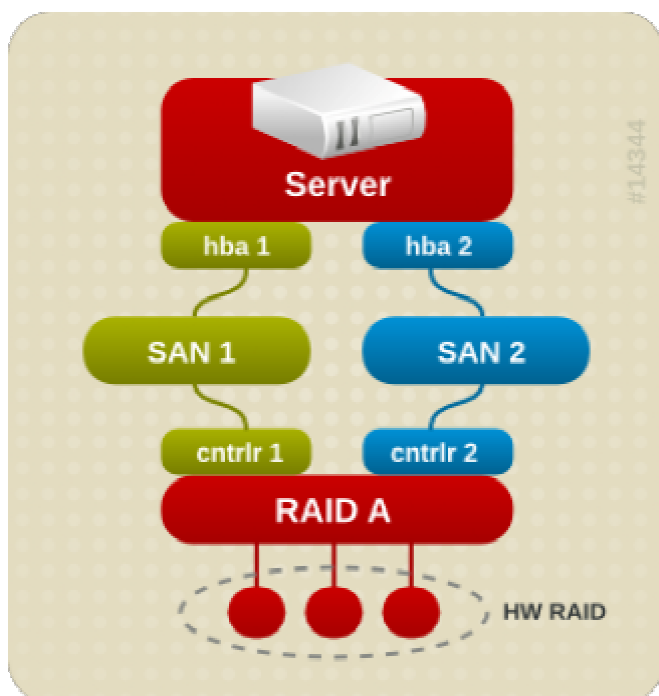


图1.1.Active/Passive Multipath Configuration with One RAID Device

在这个配置中，一个 I/O 路径通过 hba1、SAN1 以及控制器 1。第二个 I/O 路径通过 hba2、SAN2 以及控制器 2。在这个配置中有很多地方可能导致故障：

HBA 故障

FC 电缆故障

SAN 交换机故障

阵列控制器端口故障

如果配置了 DM-Multipath，在所有这些地方出现故障时都会使 DM-Multipath 切换到备用 I/O 路径。

图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.

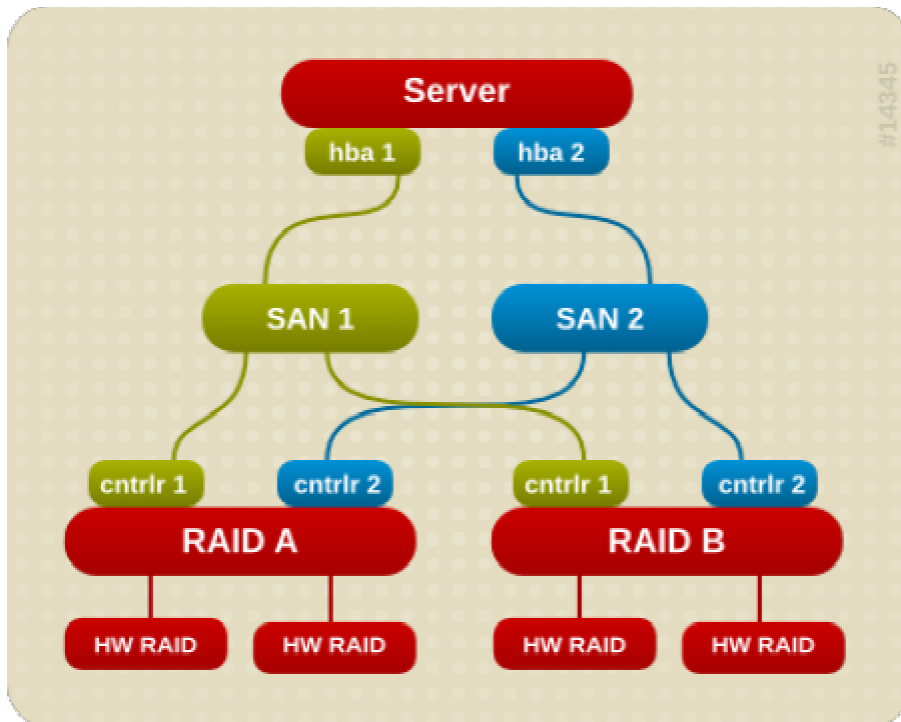


图1.2.Active/Passive Multipath Configuration with Two RAID Devices

In the example shown in 图1.2 “Active/Passive Multipath Configuration with Two RAID Devices,” there are two I/O paths to each RAID device (just as there are in the example shown in 图1.1 “Active/Passive Multipath Configuration with One 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.

图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 到控制器 1

hba1 到控制器 2

hba2 到控制器 1

hba2 到控制器 2

在这个配置中，可将 I/O 分布到那四条路径中。

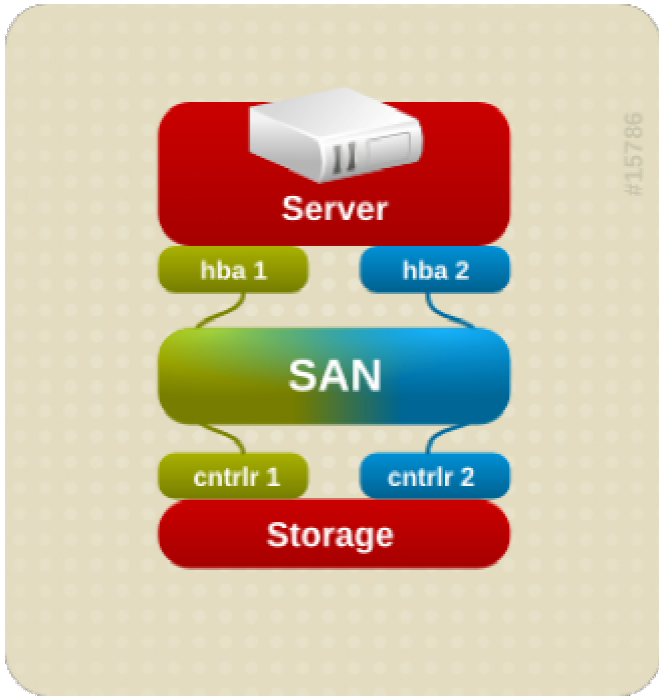


图1.3.Active/Active Multipath Configuration with One RAID Device

1.2.存储阵列支持

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 [第4章 DM-Multipath 配置文件](#).

有些存储阵列需要对 I/O 错误和路径切换的特殊处理。这些要求独立于硬件处理器内核模块。

1.3.DM-Multipath 组件

表1.1 “DM-Multipath 组件” describes the components of DM-Multipath.

表1.1.DM-Multipath 组件

组件	描述
dm-multipath kernel module	为路径和路径组群重新指定 I/O 并支持出错冗余。
multipath command	列出并配置多路径设备。通常从 <code>/etc/rc.sysinit</code> 开始，还可以使用 <code>udev</code> 程序随时添加块设备，或者可用 <code>initramfs</code> 文件系统运行。
multipathd daemon	监视路径，如果路径故障并返回，它可能会启动路径组群切换。可为多路径设备提供互动修改。对 <code>/etc/multipath.conf</code> 文件的任何修改都必须启动它。
kpartx command	为设备中的分区生成设备映射器设备。这个命令对带 DM-MP 的 DOS 分区是很必要的。 <code>kpartx</code> 在其自身软件包中就存在，但 <code>device-mapper-multipath</code> 软件包要依赖它。

1.4.DM-Multipath 安装总览

DM-Multipath 包含适用于通用多路径配置的内置默认设置。安装 DM-multipath 通常很简单。
以下是为您的系统配置 DM-multipath 的基本步骤：

1. 安装 `device-mapper-multipath` rpm。
2. 编辑 `multipath.conf` 配置文件：
 - 注释出默认黑名单
 - 根据需要修改现有默认设置
 - 保存配置文件
3. 启动多路径守护进程。
4. 使用 `multipath` 命令生成多路径设备。

Detailed setup instructions for several example multipath configurations are provided in see [第3章 安装 DM-Multipath](#).

多路径设备

没有 DM-Multipath，从服务器节点到存储控制器的所有路径都将被系统视为独立的设备，即便连接相同服务器节点到相同存储控制器的 I/O 路径也不例外。DM-Multipath 提供了逻辑管理 I/O 路径的方法，即在基础设备的顶部生成单一多路径设备。

2.1.多路径设备识别器

每个多路径设备都有一个全球识别号（WWID），它是一个全球唯一的无法更改的号码。默认情况下，多路径设备的名称将设为它的 WWID。另外，您还可以在多路径配置文件中设置 `user_friendly_names` 选项，该选项可将别名设为格式为 `mpathn` 的节点唯一名称。

例如：当在有两个端口的存储控制器中通过单一不分区 FC 切换附加带两个 HBA 的节点时，可看到四个设备：`/dev/sda`、`/dev/sdb`、`dev/sdc` 和 `/dev/sdd`。DM-多路径生成一个有唯一 WWID 的单一设备，该设备可将 I/O 根据多路径配置重新路由到那四个基础设备。当将 `user_friendly_names` 配置设定为 `yes`，会将该多路径设备的名称设定为 `mpathn`。

当在 DM-多路径中添加新设备时，这些新设备会位于 `/dev` 目录的三个不同位置：`/dev/mapper/`
`mpathn`、`/dev/mpath/mpathn` 和 `/dev/dm-n`。

`/dev/mapper` 中的设备是在引导过程中生成的。可使用这些设备访问多路径设备，例如在生成逻辑卷时。

提供 `/dev/mpath` 中的设备是为了方便，这样可在一个目录中看到所有多路径设备。这些设备是由 `udev` 设备管理器生成的，且在系统需要访问它们时不一定能启动。请不要使用这些设备生成逻辑卷或者文件系统。

所有 `/dev/dm-n` 格式的设备都只能是作为内部使用，且应该永远不要使用。

For information on the multipath configuration defaults, including the `user_friendly_names` configuration option, see [第4.3节 "多路径设备配置属性"](#)

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 [第4.4节 "多路径设备配置属性"](#)

2.2.在群集中保持多路径设备名称一致

当将 `user_friendly_names` 配置选项设为 `yes` 时，该多路径设备的名称对于节点来说是唯一的，但不保证对使用多路径设备的所有节点都一致。如果您使用 LVM 生成逻辑设备，这不会给您造成任何困难。

。但如果您需要将您的多路径设备名称在群集中的每个节点上都保持一致，请执行以下步骤之一：

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 [第4.4节 "多路径设备配置属性"](#).

如果您想要系统定义的用户友好名称在群集中的所有节点间保持一致，请在一台机器中设置所有多路径设备。然后将那台机器中的 `/var/lib/multipath/bindings` 文件复制到群集中其它所有机器中。

2.3.多路径设备属性

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 [第4.4节 "多路径设备配置属性"](#)

2.4.逻辑卷中的多路径设备

生成多路径设备后，多路径设备的名称可以与您生成一个 LVM 物理卷时使用的物理设备名称相同。例如：如果多路径名称为 `/dev/mapper/mpath0`，以下命令可将 `/dev/mapper/mpath0` 标记为一个物理卷。

```
pvccreate /dev/mapper/mpath0
```

您可以如在使用其它 LVM 物理设备一样在创建 LVM 卷组时使用得到的 LVM 物理卷。当您创建一个使用主动/被动多路径阵列的 LVM 逻辑卷作为基础物理设备时，您应该在 `lvm.conf` 中添加过滤器以便排除位于多路径设备下的磁盘。这是因为如果阵列在接收 I/O 时自动将主动路径改为被动路径，不过滤掉这些设备，无论何时 LVM 扫描被动路径时多路径将执行出错冗余并进行出错切换。对于需要使用命令激活被动路径的主动/被动阵列，LVM 会在发生这种情况时给出警告信息。

要过滤多路径配置文件（`lvm.conf`）中的所有 SCSI 设备，请在该文件的 `devices` 部分添加以下过滤器。

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


安装 DM-Multipath

本章提供了配置 DM-Multipath 的分步示例，其中包括以下步骤：

基本 DM-Multipath 安装

忽略本地磁盘

在配置文件中添加更多设备

3.1. 安装 DM-Multipath

在您的系统中安装 DM-Multipath 前，请确定您的系统已经进行了升级，其中包括对 `device-mapper-multipath` 软件包的升级。

使用以下步骤为基本出错冗余配置安装 DM-Multipath。

1. 通过注释出 `/etc/multipath.conf` 文件的开始几行来编辑该文件。配置文件的这一部分在开始的时候是将所有设备都列入黑名单的。您必须将其注释出来以便启用多路径。

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

注释出这些行以后，这部分的形式如下。

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

2. DM-Multipath 的默认设置已经被编译到系统中，且无需在 `/etc/multipath.conf` 文件中特别指出。

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 [第4章 DM-Multipath 配置文件](#).

在您系统配置文件的初始默认部分，多路径设备名称的格式为 `mpathn`；如果没有这个设置，多路径设备的名称将会是该设备 WWID 别名。

3. 保存配置文件并退出编辑器。
4. 执行以下命令：

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

`multipath -v2` 命令输出多路径设备的路径来显示哪些设备是多路径设备。如果没有输出任何结果，请确定正确调试了所有 SAN 连接并使系统多路径。

For further information on the `multipath` command output, see [第5.4节 “Multipath Command Output”](#).

5. 执行以下命令以确定在引导时启动了多路径守护进程：

```
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 [第4章 安装DM-Multipath 配置文件](#).

3.2.在生成多路径设备时忽略逻辑磁盘

有些机器在其内部磁盘中有本地 SCSI 卡。不建议在这些设备中使用 DM-Multipath。以下步骤显示如何在配置多路径时修改多路径配置文件来忽略逻辑磁盘。

1. 确定哪些磁盘是内部磁盘并将其列入黑名单。

在这个示例中，`/dev/sda` 是内部磁盘。请注意：因为在默认多路径配置文件中已经对其进行了配置，所以执行 `multipath -v2` 会在多路径映射中显示本地磁盘 `/dev/sda`。

For further information on the multipath command output, see [第5.4节 “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. 要防止设备映射器在其多路径映射中映射 `/dev/sda`，请编辑 `/etc/multipath.conf` 文件的 `blacklist` 部分，使其包括该设备。尽管您可以使用 `devnode` 类型将 `sda` 设备列入黑名单，但那不是一个安全的步骤，因为在重启是无法保证 `/dev/sda` 是同一个名称。要将单独的设备列入黑名单，您可以使用那个

设备的 WWID 将其列入黑名单。

请注意：在 `multipath -v2` 命令的输出中，`/dev/sda` 设备的 WWID 是 `SIBM-ESXSST336732LC____F3ET0EP0Q000072428BX1`。要将这个设备列入黑名单，请在 `/etc/multipath.conf` 文件中添加以下内容。

```
blacklist {
    wwid SIBM-ESXSST336732LC____F3ET0EP0Q000072428BX1
}
```

3. 在您更新 `/etc/multipath.conf` 文件后，您必须手动让 `multipathd` 守护进程重新载入该文件。以下命令可重新载入更新的 `/etc/multipath.conf` 文件。

```
service multipathd reload
```

4. 运行以下命令：

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

本地磁盘或者不应再列入新的多路径映射的磁盘如下。

```
[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
```

3.3. 在多路径数据库中添加设备

默认情况下，DM-Multipath 包括对大多数支持 DM-Multipath 的通用存储阵列的支持。默认配置值请参考 `multipath.conf.defaults` 文件，其中包括支持的设备。

如果您需要添加一个在默认情况下不支持的存储设备作为已知多路径设备，请编辑 `/etc/multipath.conf` 文件并插入正确的设备信息。

例如：要添加有关 HP Open-V 系列，条目示例如下：

```
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 [第4.5节 “配置文件设备”](#)

DM-Multipath 配置文件

默认情况下，DM-Multipath 为大多数多路径的使用提供配置值。另外，DM-Multipath 包括对支持 DM-Multipath 大多数通用存储阵列的支持。默认配置值和支持的设备请参考 `/usr/share/doc/device-mapper-multipath-0.4.7/multipath.conf.defaults` 文件。

您可以通过编辑 `/etc/multipath.conf` 配置文件为 DM-Multipath 覆盖默认配置值。如果需要，您还可以在配置文件中添加默认不支持的存储阵列。本章提供了解析和修改 `multipath.conf` 文件的信息，它包括以下部分：

配置文件总览

配置文件 `blacklist`

配置文件 `defaults`

配置文件 `multipaths`

配置文件 `devices`

在多路径配置文件中，您只需要指定需要的配置，或者您想要修改的在 `multipath.conf.defaults` 文件中指定的默认值。如果文件中有和您的环境不相关的部分或者您不需要覆盖默认值，您可以将其注释出来，因为它们在文件的起始部分。

配置文件允许常规表达式描述语法。

配置文件的注释版本请参考 `/usr/share/doc/device-mapper-multipathd-0.4.7/multipath.conf.annotated`。

4.1. 配置文件总览

多路径配置文件可分为以下几个部分：

blacklist 多路径不会考虑类出具体设备。默认情况下所有设备都会列在黑名单中。通常黑名单部分是被注释出的。

blacklist_exceptions

列出了根据黑名单部分的参数可能会被列到黑名单中的多路径候选人。

defaults

DM-Multipath 的通用默认设置。

multipaths

各个独立多路径设备的特性设置。这些数值覆盖了在配置文件的 `defaults` 和 `devices` 部分中指定的数值。

devices

各个存储控制器的设置。这些数值覆盖了在配置文件的 `defaults` 部分指定的数值。如果您要使用不是默认支持的存储阵列，您可能需要为您的阵列创建 `devices` 子部分。

当系统决定多路径设备的属性时，首先它会检查多路径设置，然后是每个设备的设置，然后才是多路径系统默认设置。

4.2.配置文件黑名单

多路径配置文件的 `blacklist` 部分指定在系统配置多路径设备时不能使用的设备。黑名单中的设备将无法分组到多路径设备中。

默认情况下，所有设备都列在黑名单中，因为下面这些行出现在配置文件开始的部分。

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

To enable multipathing on all of the devices that are supported by default, comment out those lines, as described in [第3.1节 "安装DM-Multipath."](#)

注释出通用黑名单后，您可以在黑名单中指定一般设备类型和个体设备。您可以根据以下标准将设备列入黑名单：

By WWID, as described in [第4.2.1节 "根据 WWID 将设备列入黑名单"](#)

By device name, as described in [第4.2.2节 "根据设备名称将设备列入黑名单"](#)

By device type, as described in [第4.2.3节 "根据设备类型将其加入黑名单"](#)

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 [第4.2.2节 "根据设备名称将设备列入黑名单."](#)

4.2.1.根据 WWID 将设备列入黑名单

您可根据全球识别号将单独的设备列入黑名单，即在配置文件的 `blacklist` 部分加入 `wwid` 条目。

以下示例显示了在配置文件中可将 WWID 为 26353900f02796769 的设备列入黑名单的行。

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

4.2.2.根据设备名称将设备列入黑名单 您可以根据设备名称将设备类型列入黑名单，以便在配置文件 `blacklist` 部分的 `devnode` 条目中指定不要将它们分组到多路径设备中。

以下示例显示了配置文件中可将所有 SCSI 设备列入黑名单的行，因为它的黑名单中有 `sd*` 设备。

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

您可以使用配置文件 `blacklist` 部分的 `devnode` 条目指定要列入黑名单的每个设备，而不是指定特定类型的所有设备，我们不推荐使用后一种情况。除非根据 `udev` 规则的静态映射，我们无法保证在重启后每个设备的名称是一样的。例如：重启后，某个设备的名称可能从 `/dev/sda` 变为 `/dev/sdb`。

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 [第4.2.4节 黑名单之外的设备](#)

```

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

```

4.2.3.根据设备类型将其加入黑名单

您可在配置文件 `blacklist` 部分使用 `device` 部分指定具体设备类型。以下示例将所有 IBM DS4200 设备以及所有 HP 设备都列入了黑名单。

```

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

```

4.2.4.黑名单之外的设备

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.*"
    }
}

```

当在配置文件的 `blacklist_exceptions` 指定设备时，您必须以指定黑名单的相同方法来指定例外情况。例如：在 `devnode` 黑名单条目中指定的设备无法使用 `WWID` 将其指定为例外情况，即使列入黑名单的设备和该 `WWID` 关联也不行。同样，`devnode` 例外也只适用于 `devnode` 条目，而 `device` 例外只适用于 `device` 条目。

4.3.配置文件默认设置

`/etc/multipath.conf` 配置文件包括 `defaults` 部分，在该部分中将 `user_friendly_names` 参数设为 `yes`，如下。

```

defaults {
    user_friendly_names yes
}

```

这可覆盖 `user_friendly_names` 参数的默认值。

该配置文件包括配置默认模板。这部分要被注释出来，如下。

```

#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
# max_fds 8192
# rr_weight priorities
# failback immediate
# no_path_retry fail
# user_friendly_names yes
#}
```

要覆盖任意配置参数的默认值，您可将这个模板中相关的行复制到 `defaults` 部分并取消其注释。例如：要覆盖 `path_grouping_policy` 参数以使用 `multibus` 覆盖默认的 `failover`，请将模板中正确的行复制到配置文件的 `defaults` 部分并取消对它的注释，如下。

```
defaults {
    User_friendly_names yes
    path_grouping_policy multibus
}
```

表4.1 “多路径配置默认设置”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.

表4.1.多路径配置默认设置

属性	描述
<code>udev_dir</code>	指定生成 <code>udev</code> 设备节点的目录，默认值为 <code>/udev</code> 。
<code>polling_interval</code>	以秒数指定检查两个路径间的时间间隔，默认值为 <code>5</code> 。
<code>selector</code>	指定用来确定下一个 I/O 操作所使用的路径的默认算法，默认值为 <code>round-robin 0</code> 。
<code>path_grouping_policy</code>	指定用于未指定路径的默认路径分组策略，可能的值包括： <code>failover</code> = 每个优先组群有一个路径 <code>multibus</code> = 所有有效路径在一个优先组群中 <code>group_by_serial</code> = 每个检测到的系列号有一个优先组群 <code>group_by_prio</code> = 每个路径优先值有一个优先组群 <code>group_by_node_name</code> = 每个目标节点名有一个优先组群 默认值为 <code>failover</code> 。
<code>getuid_callout</code>	Specifies the default program and arguments to call out to obtain a unique path identifier. An absolute path is required. 默认值为 <code>/sbin/scsi_id -g -u -s</code> 。
<code>prio_callout</code>	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
<code>path_checker</code>	指定用来确定路径状态的默认方法。可能的值有 <code>readsector0</code> 、 <code>tur</code> 、 <code>emc_clariion</code> 、 <code>hp_sw</code> 和 <code>directio</code> 。默认值为 <code>readsector0</code> 。
<code>rr_min_io</code>	指定在当前路径组群中切换到下一个路径前指向一个路径的 I/O 请求数目。默认值为 <code>1000</code> 。
<code>max_fds</code>	（RHEL 5.2 以及之后的版本）会为 <code>multipathd</code> 进程设定打开文件描述符的最大值。在 RHEL 5.3 中，这个选项允许值 <code>max</code> ，它可将

打开文件描述符的数目设置为系统可承受的最大值。

属性	描述
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>如果将其设为 uniform，那么所有路径加权是相等的。默认值为 uniform。</p>
failback	<p>指定路径组群出错切换。</p> <p>数值 0 或者 immediate 指定一旦存在优先权高于现有路径组群的路径组群，那么系统将切换到那个路径组群。</p> <p>大于 0 的数字值指定推迟出错切换，以秒表示。</p> <p>manual 值表明只能在操作员介入时才会发生出错切换。</p> <p>默认值为 manual。</p>
no_path_retry	<p>这个属性的数字值指定系统在禁用队列前应该试图使用失败的路径的次数。</p> <p>fail 值表明立即失败，无需排队。</p> <p>queue 值表明在路径修复前不应停止排队。</p> <p>默认至为（空白）。</p>
flush_on_last_del	<p>（在 RHEL 5.3 以及之后的版本中）如果将其设定为 yes，multipathd 守护进程将在删除了最后一个到某个设备的路径时 禁用查询。默认值为 no。</p>
queue_without_daemon	<p>（在 RHEL 5.3 以及之后的版本中）如果将其设定为 no，multipathd 守护进程将在关闭后禁用对所有设备的查询。默认 值为 yes。</p>
user_friendly_names	<p>如果将其设为 yes，即该系统应该使用文件 bindings 来将一个持久且唯一的别名分配给多路径，格式为 mpathn。bindings 文件的默认位置为 /var/lib/multipath/bindings，但使用 bindings_file 选项可改变此位置。如果将其设为 no，即该系统应该使用 WWID 作为多路径的别名。在这两种情况下，您在这里指定的数值将被您在配置文件 multipaths 部分指定的具体设备别名覆盖。默认值为 no。</p>
bindings_file	<p>（在 RHEL 5.2 以及之后的版本中）bindings 文件的位置与 user_friend_names 选项一同使用。默认值为 /var/lib/multipath/bindings。</p>
mode	<p>（在 RHEL 5.3 以及之后的版本中）多路径设备节点使用的 mode 是八进制的。默认值由具体进程决定。</p>
uid	<p>（在 RHEL 5.3 以及之后的版本中）多路径设备节点使用的用户 ID。您必须使用数字用户 ID。默认值由具体进程决定。</p>
gid	<p>（在 RHEL 5.3 以及之后的版本中）多路径设备节点使用的组群 ID。您必须使用数字组群 ID。默认值由具体进程决定。</p>

4.4.多路径设备配置属性

表4.2 “多路径属性”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.

表4.2.多路径属性

属性	描述
wwid	指定 multipath 属性采用的多路径设备的 WWID。alias 指定 multipath 属性采用的多路径符号名称。
path_grouping_policy	指定用于未指定路径的默认路径分组策略，可能的值包括： failover = 每个优先组群有一个路径 multibus = 所有有效路径在一个优先组群中 group_by_serial = 每个检测到的系列号有一个优先组群 group_by_prio = 每个路径优先值有一个优先组群 group_by_node_name = 每个目标节点名有一个优先组群
path_selector	指定用来确定下一个 I/O 操作所使用路径的默认算法。
failback	指定路径组群出错切换。 数值 0 或者 immediate 指定一旦存在优先权高于现有路径组群的路径组群，那么系统将切换到那个路径组群。 大于 0 的数字值指定推迟出错切换，以秒表示。 manual 值表明只能在操作员介入时才会发生出错切换。
rr_weight	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. 如果将其设为 uniform，那么所有路径的加权都相等。
no_path_retry	这个属性的数字值指定系统在禁用队列前应该试图使用失败的路径的次数。 fail 值说明立即失败，无需排队。 queue 值表明在路径修复前不应停止排队。
flush_on_last_del	（在 RHEL 5.3 以及之后的版本中）如果将其设定为 yes，multipathd 守护进程将在删除了最后一个到某个设备的路径时 禁用查询。默认值为 no。
rr_min_io	指定在当前路径组群中切换到下一个路径前指向一个路径的 I/O 请求数目。
mode	（在 RHEL 5.3 以及之后的版本中）多路径设备节点使用的 mode 是八进制的。默认值由具体进程决定。
uid	（在 RHEL 5.3 以及之后的版本中）多路径设备节点使用的用户 ID。您必须使用数字用户 ID。默认值由具体进程决定。
gid	（在 RHEL 5.3 以及之后的版本中）多路径设备节点使用的组群 ID。您必须使用数字组群 ID。默认值由具体进程决定。

以下示例显示在配置文件中为两个特定多路径设备指定的多路径属性。第一个设备的 WWID 为 3600508b4000156d70001200000b0000，符号名称为 yellow。

示例中第二个多路径设备的 WWID 为 1DEC____321816758474，符号名称为 red。在这个示例中，rr_weight 属性被设为 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
    }
}
```

4.5.配置文件设备

表4.3 “设备属性”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.

默认情况下，在多路径配置中包括很多支持多路径的设备。这些设备支持的值默认列在 multipath.conf.defaults 文件中。您可以不需要修改这些设备，但如果修改了这些值，您可以通过在配置文件中为覆盖那些值的设备添加条目来覆盖默认值。您可以从 multipath.conf.defaults 文件中为设备复制设备配置默认设置并覆盖那些您想要修改的值。

要在配置文件的这个部分添加没有默认自动配置的设备，您需要设定 vendor 和 product 参数。您可在 /sys/block/device_name/device/vendor 和 /sys/block/device_name/device/model 文件中找到这些值，其中 device_name 是要进行多路径操作的设备，示例如下：

```
[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 表4.3 “设备属性”.

如果设备是主动/被动模式，但它会自动将路径切换到被动路径，那么您需要将检查器改为一个不需向路径发送 I/O 来测试其是否工作（否则，您的设备将一直进行出错冗余）的功能。这几乎意味着您将 path_checker 设为 tur。这对所有支持 Test Unit Ready 命令的 SCSI 设备都适用。

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.

表4.3.设备属性

属性	描述
vendor	指定 device 采用的存储设备的零售商名称，例如 COMPAQ。
product	指定 device 属性使用的存储设备产品名，比如 HSV110 (C) COMPAQ。
path_grouping_policy	指定用于未指定路径的默认路径分组策略，可能的值包括： failover = 每个优先组群有一个路径 multibus = 所有有效路径在一个优先组群中 group_by_serial = 每个检测到的系列号有一个优先组群 group_by_prio = 每个路径优先值有一个优先组群 group_by_node_name = 每个目标节点名有一个优先组群
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	指定用来确定路径状态的默认方法。可能的值包括 readsector0 和 tur、emc_clariion、hp_sw 以及 directio。
path_selector	指定用来确定下一个 I/O 操作所使用路径的默认算法。
features	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. For information on issues that may arise when using this feature, see 第5.3节 “Issues with queue_if_no_path feature.”
hardware_handler	指定将在切换路径组群或者处理 I/O 错误时用来执行硬件具体动作的模块。可能的值包括 0、1 emc 和 1 rdac。默认值为 0。
rr_weight	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. 如果将其设为 uniform，那么所有路径加权是相等的。默认值为 uniform。
no_path_retry	这个属性的数字值指定系统在禁用队列前应该试图使用失败的路径的次数。 fail 值表明立即失败，无需排队。 queue 值表明在路径修复前不应停止排队。 默认至为（空白）。
failback	指定路径组群出错切换。 数值 0 或者 immediate 指定一旦存在优先权高于现有路径组群的路径组群，那么系统将切换到那个路径组群。 大于 0 的数字值指定推迟出错切换，以秒表示。 manual 值表明只能在操作员介入时才会发生出错切换。
rr_min_io	指定在当前路径组群中切换到下一个路径前指向一个路径的 I/O 请求数目。默认值为 1000。

<code>flush_on_last_del</code>	(在 RHEL 5.3 以及之后的版本中) 如果将其设定为 <code>yes</code> , <code>multipathd</code> 守护进程将在删除了最后一个到某个设备的路径时 禁用查询。默认值为 <code>no</code> 。
<code>product_blacklist</code>	根据产品指定用来将设备列入黑名单的常规表达式。

以下示例显示了多路径配置文件的 `device` 条目

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

DM-Multipath 管理及故障排除

本章提供了在运行的系统中管理 DM-Multipath 的信息，其中包括以下部分：

The Multipath Daemon

Issues with Large Number of LUNs

Issues with `queue_if_no_path` feature

Multipath Command Output

Multipath Queries with `multipath` Command

Multipath Command Options

用 `dmsetup` 命令进行多路径查询

Troubleshooting with the `multipathd` Interactive Console

5.1.The Multipath Daemon

If you find you have trouble implementing a multipath configuration, you should ensure that the multipath daemon is running, as described in [第3章 安装DM-Multipath](#). The `multipathd` daemon must be running in order to use multipathed devices.

5.2.Issues with Large Number of LUNs

当在某个节点中添加大量 LUN 时，使用多路径设备可明显延长 `udev` 设备过滤器为其生成设备节点所用的时间。如果您遇到过这个问题，您可在 `/etc/udev/rules.d/40-multipath.rules` 文件中删除以下行解决这个问题：

```
KERNEL!="dm-[0-9]*", ACTION=="add", PROGRAM=="bin/bash -c '/sbin/lsmode | /bin/grep ^dm_multipath'", RUN+="/sbin/multipath -v0 %M:%m"
```

这行会在每次向该节点中添加块设备时让 `udev` 设备管理器运行 `multipath`。虽然删除了这一行，`multipathd` 守护进程仍将自动生成多路径设备，在引导过程中仍将为带多路径根文件系统的节点调用 `multipath`。唯一的改变就是没有运行 `multipathd` 守护进程不再自动生成多路径设备，对大多数多路径用户来说应该不是个问题。

5.3.Issues with queue_if_no_path feature

If features "1 queue_if_no_path" is specified in the /etc/multipath.conf file, then any process that issues I/O will hang until one or more paths are restored. To avoid this, set the no_path_retry N parameter in the /etc/multipath.conf file (where N is the number of times the system should retry a path). When you do, remove the features "1 queue_if_no_path" option from the /etc/multipath.conf as well.

If you need to use the features "1 queue_if_no_path" option and you experience the issue noted here, use the dmsetup command to edit the policy at runtime for a particular LUN (that is, for which all the paths are unavailable). For example, if you want to change the policy on the multipath device mpath2 from "queue_if_no_path" to "fail_if_no_path", execute the following command.

```
dmsetup message mpath2 0 "fail_if_no_path"
```

请注意：您必须指定 mpathn 别名而不是该路径。

5.4.Multipath Command Output

当您生成、修改或者列出多路径设备时，您会获得一个当前设备组成的打印输出，格式如下。

对于每个多路径设备：

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

对与每个路径组群：

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

对于每个路径：

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

例如，多路径命令的输出可能是如下形式：

```
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]
```

如果该路径已经启用并准备好进行 I/O，那么路径的状态就是 ready 或者 active。路径的状态由 multipathd 守护进程根据在 /etc/multipath.conf 文件中定义的通信间隔进行周期性更新。

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.



注意

当生成或者修改一个多路径设备时，路径组群状态和 dm 状态是未知的。同时，特性也不一定正确。当列出多路径设备时，路径组群优先权是未知的。

5.5.Multipath Queries with multipath Command

您可以使用 `multipath` 命令的 `-l` 和 `-ll` 选项来显示现有多路径配置。`-l` 选项显示从 `sysfs` 以及设备映射器中收集的多路径拓补信息。`-ll` 选项显示 `-l` 显示的所有信息以及该系统额外可用的其它组件。

当显示多路径配置时，您可以使用 `multipath` 命令的 `-v` 选项指定三种详细等级。指定为 `-v0` 时没有输出。指定为 `-v1` 则只输出生成或者更新的路径名称。指定 `-v2` 将输出所有检测到的路径、多路径和设备映射。

以下示例显示了 `multipath -l` 命令的输出。

```
# 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]
```

5.6.Multipath Command Options

表5.1 有用的multipath 命令选项” describes some options of the `multipath` command that you may find useful.

表5.1.有用的multipath 命令选项

属性	描述
<code>-l</code>	显示来自 <code>sysfs</code> 和设备映射器的当前多路径配置。
<code>-ll</code>	显示来自 <code>sysfs</code> 、设备映射器以及系统中的其它可用组件的当前多路径配置。
<code>-f device</code>	删除命名的多路径设备。
<code>-F</code>	删除所有多路径设备。

5.7.使用 dmsetup 命令确定设备映射器条目

您可以使用 `dmsetup` 找出哪个设备映射器条目与多路径的设备映射。

以下命令显示所有设备映射器设备及其主、副号码。副号码确定 `dm` 设备的名称。例如：副号码 3 与多路径的设备 `/dev/dm-3` 对映。

```
# 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)
```

5.8.Troubleshooting with the multipathd Interactive Console multipathd

`-k` 命令是一个 `multipathd` 守护进程的互动界面。执行此命令将进入互动的多路径控制台。执行此命令后，您可以输入 `help` 来获得可用的命令列表。您可以输入互动命令，或者您还可以输入 `CTRL-D` 退出。

`multipathd` 互动控制台可用来在您的系统有问题时进行故障排除。例如：以下命令会在退出控制台前显示多路径配置，其中包括默认配置。

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

以下命令确定多路径已经识别了所有对 `multipath.conf` 的修改。

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

使用以下命令组确定路径检查器工作正常。

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