1. sudo apt-get install mdadm
2. /proc/mdstat ： 当前md(软RAID)的状态信息

/etc/mdadm.conf ： mdadm的配置文件

1. mdadm管理软raid陈列

dadm程序是一个独立的程序，能完成所有的软件raid管理功能，主要有7种使用模式：

1. Create

使用空闲的设备创建一个新的阵列，每个设备具有元数据块

1. Assemble

将原来属于一个阵列的每个块设备组装为阵列

1. Build

创建或组装不需要元数据的阵列，每个设备没有元数据块

1. Manage

管理已经存储阵列中的设备，比如增加热备磁盘或者设置某个磁盘失效，然后从阵列中删除这个磁盘

1. Misc

报告或者修改阵列中相关设备的信息，比如查询阵列或者设备的状态信息

1. Grow

改变阵列中每个设备被使用的容量或阵列中的设备的数目

1. Monitor

监控一个或多个阵列，上报指定的事件

1. 创建过程：
2. 查看系统下面的硬盘编号 # fdisk -l
3. 用mdadm命令将sda和sdb创建为raid1。

# mdadm --create --auto=yes /dev/md0 --level=1 --raid-devices=2 /dev/sda /dev/sdb

以上命令表示：创建raid1 ；虚拟的raid设备为md0 ；用于raid的设备为2块。

1. 创建完成以后，用如下命令查看raid的情况：

# mdadm --detail /dev/md0

1. 创建完成以后，可以对虚拟的raid设备进行格式化等操作，
2. 跟操作一个硬盘的效果是一样的。格式化的命令为：

# mkfs.ext4 /dev/md0

ext4是debian6新支持的文件格式，好处比ext3多很多。

1. 格式化完成以后，开始mount。

# mkdir /data

# mount /dev/md0 /data

查看/data，以确定挂载是否成功；也可以创建文件或文件夹；删除文件或文件夹来测试对raid的操作是否正常。

1. 修改fstab，将raid挂载到/data命令添加到开机自启动。

 # vim /etc/fstab

1. 配置完成后，重启系统，检测raid是否正常自动挂载。

**如果以后有个硬盘坏了，怎么办？**

步骤是:先把坏的硬盘拔下来，拔下来以后，在raid设置里面,把这个坏的挪走。然后插上新的，再在raid设置里面把新硬盘加入到raid就可以了。(举例更换sdb)步骤如下:

 1.  移除坏的硬盘

# mdadm --manage /dev/md0  --remove /dev/sdb

2.  插入新硬盘后,加入到整列

# mdadm --manage /dev/md0 --add /dev/sdb

 3.  查看raid组建情况

 # mdadm --detail /dev/md0

等一段时间后,raid将自动恢复。

现在讲下如果硬盘不够用了，如何新加入硬盘即，对raid扩容的方法。扩容的步骤如下，先插入到主板上，把新硬盘加入到raid整列中。命令如下：

1.  加入新硬盘到整列。

# mdadm --manage /dev/md0 --add /dev/sde

2.  设定raid可用硬盘数量。

# mdadm --grow /dev/md0 --raid-devices 4

3.  查看raid情况,发现已经开始同步了，估计要很久的时间。

# mdadm --detail /dev/md0

1. mdadm软RAID的删除方法和注意事项:

删除整个RAID：

mdadm /dev/md0 --fail /dev/sdb --remove /dev/sdb  
mdadm /dev/md0 --fail /dev/sdc --remove /dev/sdc  
mdadm /dev/md0 --fail /dev/sdc --remove /dev/sdd  
mdadm --stop /dev/md0  
mdadm --remove /dev/md0

mdadm --misc --zero-superblock /dev/sdb  
mdadm --misc --zero-superblock /dev/sdc  
mdadm --misc --zero-superblock /dev/sdd  
即： 先删除RAID中的所有设备，然后停止该RAID即可

1. 各种模式的使用：

**Assemble模式**

用法：mdadm –assemble md-devices options-and-component-devices…

例子：#mdadm –assemble /dev/md0 /dev/sda1 /dev/sdb1

说明：把sda1和sdb1重组成/dev/md0。

用法：mdadm –assemble –scan md-devices-and-options…

例子：#mdadm –assemble –scan /dev/md0

说明：从配置文件读出设备列表，根据超级块中的信息，重组/dev/md0。

用法：mdadm –assemble –scan options…

例子：#mdadm –assemble –scan –uuid=xxxxxxx

说明：从配置文件读出设备列表，根据超级块中的uuid信息，重组uuid是xxxxxxx的RAID。

假如#mdadm –assemble –scan命令后面没有设备列表，mdadm会读取配置文件中所列的RAID信息，并尝试重组。假如系统中没有/etc/mdadm.conf配置文件，则会从/proc/partitions中读取设备列表。因此使用mdadm时，必须保证mdadm中的状态信息(/proc/mdstat)和/etc/mdadm.conf配置文件一致,否则重启操作系统后，会出现问题。

## Create模式

用法：mdadm –create md-device –chunk=X –level=Y –raid-devices=Z devices

例子：# mdadm –create /dev/md0 –chunk=64 –level=0 –raid-devices=2 /dev/sda1 /dev/sdb1

说明：使用sda1和sdb1创建RAID0，条带大小是64KB。

例子：#mdadm –create /dev/md1 –chunk=64 –level=1 –raid-devices=1 /dev/sdc1 missing

说明：创建一个降级的RAID1，同样可以使用missing创建降级的RAID4/5/6。

## Misc模式

用法：mdadm options… devices..

例子：#mdadm –detail –test /dev/md0

说明：这条命令的返回值：0代表md0正常；1代表md0至少有一个failed的组件设备；2代表md0有多个failed组件设备，这个md0已经不能使用，即失效(md0是raid1、raid5、raid6、raid10时)；4代表获取md0设备信息错误。

## Monitor模式 (只监控raid1/5/6/10，不监控raid0)

用法: mdadm –monitor options… devices..

说明：mdadm除了报告事件以外，mdadm还可以把一个RAID中的热备盘移动到另一个没有热备盘的RAID中，前提条件是这些RAID都属于同一个spare-group(RAID的spare-group可以在配置文件里设置)。

说明：当命令中有设备列表时，mdadm只会监控这些设备。当没有设备列表时，配置文件中的所有RAID都会被监控。当使用–scan选项时，/proc/mdstat中的设备也会被监控。

说明：传给program的三个参数是事件名、涉及到的md device名、涉及到的其他设备(比如组件设备失效)。

说明：监控的事件有

    DeviceDisappeared 当RAID0和linear中某个设备失效时，就会出现RAID消失。

    RebuildStarted    重建RAID

    RebuildNN    重建百分比，NN代表20,40,60,80

    RebuildFinished 重建结束

    Fail RAID中某个活动组件设备失效

    FailSpare RAID中某个热备盘失效

    SpareActive RAID中热备盘启用，用于重建RAID

    NewArray 在/proc/mdstat中监控到有新的RAID被创建

    DegradedArray RAID降级

    MoveSpare 热备盘从一个RAID中移动到另外一个RAID中，前提是这两个RAID属于同个spare-group

    SparesMissing 发现RAID中的热备盘数比配置文件中的少

    TestMessage 测试

说明：只有Fail、FailSpare、DegradedArray、SparesMissing、TestMessage事件才会触发发送Email。

## Grow模式

说明：能改变RAID1、5、6中的”size”属性。

说明：能改变RAID1、5中的”raid-disks”属性。

说明：增加移除RAID中的write-intent bitmap。

1. 命令概要： **mdadm** [模式选项] [RAID设备名] [子选项…] [组件设备名…]
2. 无特定模式

|  |  |
| --- | --- |
| 子选项 | 备注 |
| --verbose | 显示更详细的信息用于 –detail –scan 或者 –examine –scan |
| --force | 某些选项强制执行 |
| --config= | 指定配置文件，默认是”/etc/mdadm.conf”或者是”/etc/mdadm/mdadm.conf”，假如配置文件名是 “partitions”，则mdadm会读取/proc/partitions的设备用于scan。假如配置文件名是”none”，则mdadm会认为配置文件是空的。 |
| --scan | 从配置文件或者/proc/mdstat中扫描信息。 |
| --metadata= | 定义组件设备上超级块的类型。对于–create，默认是0.90。0,0.90 ： 限制一个RAID中的设备数为28个，限制组件设备大小为2TB1,1.0,1.1,1.2 ：不同的子版本号标识在不同的地方存储超级块。1.0在设备的结尾，1.1在设备的开头，1.2在设备的4K处。 |
| --homehost= | 在创建一个RAID时，homehost名会记录在超级块中。在1.X超级块中，它是RAID名字的前缀。0.90超级块中，homehost名的的SHA1值会保存在UUID的后半部分。当使用Auto-Assemble时，只有相同homehost名的RAID才会被组建。 |

2)、--create

|  |  |
| --- | --- |
| 子选项 | 备注 |
| --raid-devices= | 指定一个RAID中active devices的数目。 |

3）--build

|  |  |
| --- | --- |
| 子选项 | 备注 |
| devices= | 指定一个RAID中active devices的数目。 |

4）、--grow

|  |  |
| --- | --- |
| 子选项 | 备注 |
| --spare | 指定创建一个RAID时spare devices中的数目 |
| -devices= |  |
| --size= | 在RAID1/4/5/6中每个设备所能利用的数据容量。这个值必须是 chunk size的整数倍。而且必须留128K的设备空间用于RAID超级块。假如没有指定，则默认使用最小设备空间。该选项能用–grow 进行扩容。 |
| --chunk= | 条带大小 |
| --rounding= | 在linear array中的rounding factor，等于条带大小 |
| --level | 设置RAID级别，RAID级别有（有些是同义词，比如raid5和5）：Linear,raid0,0,stripe,raid1,1,mirror,raid4,4,raid5,5,raid6,6,raid10,10,multipath,mp,faulty。–build只支持linear,stripe,raid0,0,raid1,multipath,mp,faulty。–grow不支持改变RAID级别。 |
| --layout= | 设置RAID5、RAID10数据布局类型，控制faulty级别的failure的模式。 |
| --bitmap= | 这个选项对性能可能有影响，具体查看《mdadm手册翻译》设置一个文件用于保存write-intent位图。当文件名是”internal”时，位图复制保存在RAID组件设备（一个RAID的所有组件设备）的超级块中。当–grow，文件名是”none”时，位图会被移除。 |
| --bitmap- | 这个选项对性能可能有影响，具体查看《mdadm手册翻译》 |
| chunk= | 设置位图中每位所映射块的大小。  当使用”internal”位图时，映射块的大小是自动设置的（根据超级块中可用空间）。 |
| --write-mostly | –build、–create、–add后的设备都被标记上”wirte-mostly”。这个选项只对RAID1有效，–build、–create、–add后的设备都被标记上”wirte-mostly”。这个选项只对RAID1有效，即”md”driver会避免从RAID1的所有设备读取数据。假如镜像的速度很慢，这是非常有用的。 |
| --write-behind= | 该选项只对RAID1有效。这个选项会设置最大的写队列深度，默认值是256。使用write-behind的前提是先设置write-intent bitmap，先设置设备为write-mostly。 |
| --assume-clean | 告诉mdadm这个array已经clean。当array从一个严重的故障中恢复时，这个选项会保证没有数据会被覆盖。当创建RAID1和RAID10时，这个选项也能避免初始化同步。但是使用该选项必须要很谨慎。 |
| --backup-file= | 当使–grow为RAID5增加组件设备数时，该文件保存关键数据。（该文件不能在该RAID5上，以免发生死锁。） |
| --name | 给一个RAID设置名字，只在1.X超级块中有用，它是简单的字符串，用于assembling时识别RAID组件设备。 |
| --run | 强制激活RAID。（当一个RAID中的某些组件设备被其他RAID或者文件系统占用时）。使用这个选项，设备上有旧的元数据信息的提示会被忽略。 |
| --force | mdadm无条件接受指定的参数。  使用该选项可以只使用一个设备创建RAID；  创建RAID5时使用该选项，使得RAID5初始化不使用recovery模式，而是校验同步所有组件设备上的数据（比recovery模式要慢）。详情请见命令举例中的创建RAID |
| --auto= | 创建md设备文件。选项是{no,yes,md,mdp,part,p}{NN}。默认是yes。  “yes”要求RAID设备名是标准格式的，然后设备文件类型和minor号会自动确定。 比如RAID设备名是”/dev/mdx” 查看/proc/partitions可以看到mdx的major号是9，minor号是x。  当使用”md”时，RAID设备名可以是非标准格式，比如”/dev/md/zhu”，然后创建两个设备文件/dev/md/zhu 还有 /dev/mdx，并给这两个设备文件分配相同的major号和minor号（也就是这两个设备文件指向同一个设备）。分配minor号的方法是：分配一个没有使用过的minor号，这个minor号就是/dev/mdx中的数字x。查看/proc/partitions和/proc/mdstat，RAID设备名还是/dev/mdx。  当使用”mdp,p,part”时，RAID设备名可以是非标准格式，比如”/dev/md/zhu”，除了创建设备文件/dev/md/zhu 还有 /dev/mdx外，还会创建 /dev/md/zhup1, /dev/md/zhup2, /dev/md/zhup3, /dev/md/zhup4，这些是分区设备文件。 |
| --symlink=no | 默认下–auto会创建/dev/md/zhu的软连接 /dev/md\_zhu。假如使用该选项，则不会创建软连接。 |

5）、--assemble

|  |  |
| --- | --- |
| 子选项 | 备注 |
| --uuid= | 重组RAID，要求组件设备的uuid相同 |
| --super-minor= | minor号会保存在每个组件设备中的超级块中，可以根据这个重组RAID。 |
| --name= | 根据RAID名重组RAID。 |
| --force | 即使一些超级块上的信息过时了，也可以强制重组。 |
| --run | 即使RAID中的组件设备不完整（例如原来创建4块盘的RAID5，现在只发现3块成员盘），RAID也被重组，并启动。  （假如不用–run，RAID只被重组，但是不启动） |
| --no-degraded | 和–scan选项一起使用。  禁止RAID中的组件设备不完整时启动RAID，知道RAID中的组件完整。 |
| --auto | 如–create中的 –auto |
| --bitmap | 指定bitmap文件（当RAID创建时所指定的bitmap文件），假如RAID使用internal类型的bitmap，则不需指定。 |
| --backup-file= | 当增加RAID5的组件设备数，指定backup-file文件。在RAID5重构过程中，假如系统当机，backup-file文件会保存关键数据，使得重启系统之后，重构可以继续进行。假如没有指定backup-file，mdadm会使用热备盘上的空间作为备份空间。 |
| --update= | 更新RAID中每个组件设备的超级块信息。选项有sparc2.2、summaries、uuid、name、homehost、resync、byteorder、super-minor |
| --auto-update-homehost | 只对auto assembly情况下有用。 |

6）、Manage模式

|  |  |
| --- | --- |
| 子选项 | 备注 |
| --add | 给RAID在线添加设备(可用于添加热备盘) |
| --re-add | 给RAID重新添加一个以前被移除的设备。  假如一个RAID使用write-intent bitmap时，它的一个设备被移除后又被重新添加，bitmap可以避免完全重建，而是只更新那些设备被移除后已经被更新过的块数据。 |
| --reomve | 移除设备，只能移除failed(失效)和spare(热备)设备。(因此假如要移除RAID5中的一个活动设备，需要先使用–fail选项使该设备的状态变成failed，然后才能移除。)  该选项后面跟的是设备名(比如是 /dev/sdc)，也可以是**failed**和**detached**关键字。Failed使得所以失效的部件被移除，detached使得所以被拔出的硬盘被移除。 |
| --fail | 使RAID中某个设备变成failed状态。  该选项后面跟的是设备名(比如是 /dev/sdc)，也可以是**detached**关键字。 |

7）、Misc模式

|  |  |
| --- | --- |
| 子选项 | 备注 |
| --query | 查询一个RAID或者一个RAID组件设备的信息 |
| --detail | 查询一个RAID的详细信息 |
| --examine | 查询组件设备上的超级块信息 |
| --sparc2.2 | 用于修正超级块信息，详情请见用户手册。 |
| --examnie-bitmap | 查看bitmap文件中的信息 |
| --run | 启动不完整的RAID(比如本来是有4块盘的RAID5,现在3块盘也可以启动)。 |
| --stop | 禁止RAID活动，释放所有资源。但是RAID中组件设备上的超级块信息还在。还可以重新组建和激活RAID。 |
| --readonly | 使RAID只能只读 |
| --readwrite | 使RAID能读写 |
| --zero-superblock | 假如一个组件设备包含有效的超级块信息，那么这个超级块会被写0覆盖。假如使–force选项，则不管超级块中是否有信息，都会被写0覆盖。 |
| --test | 假如–detail一起使用，则mdadm的返回值是RAID的状态值。  0 代表正常  1 代表降级，即至少有一块成员盘失效  2 代表有多快成员盘失效，整个RAID也失效了(对于RAID1/5/6/10都适用)。  4 读取raid信息失败 |

8）、–monitor

|  |  |
| --- | --- |
| 子选项 | 备注 |
| --mail | 设置警报邮件 |
| --program | 当监测到一个事件发生时，关于该事件的信息会作为参数被发给该程序 |
| --syslog | 所有事件都会通过syslog报告 |
| --delay | Mdadm会隔多少秒轮询各个RAID，默认是60秒 |
| --daemonise | Mdadm会创建一个子进行作为后台监控程序。该子进程的PID号会输入到stdout上。 |
| --pid-file | 当–daemonise一起使用时，子进程的PID号会写入该文件中 |
| --oneshot | 只会检测RAID一次。它会检测NewArray、DegradedArray、SparesMissing事件。  在cron脚本中运行”mdadm –monitor –scan –oneshot”会定期报告。 |

1. 选项：

-A, --assemble：加入一个以前定义的阵列

-B, --build：创建一个没有超级块的阵列(Build a legacy array without superblocks.)

 -C, --create：创建一个新的阵列

-F, --follow, --monitor：选择监控(Monitor)模式 -G, --grow：改变激活阵列的大小或形态

-I, --incremental：添加一个单独的设备到合适的阵列，并可能启动阵列

 --auto-detect：请求内核启动任何自动检测到的阵列

-h, --help：帮助信息，用在以上选项后，则显示该选项信息

 --help-options：显示更详细的帮助

-V, --version：打印mdadm的版本信息

 -v, --verbose：显示细节

-b, --brief：较少的细节。用于--detail 和 --examine 选项

-Q, --query：查看一个device，判断它为一个 md device 或是 一个 md 阵列的一部分-D, --detail：打印一个或多个 md device 的详细信息

-E, --examine：打印 device 上的 md superblock 的内容

-c, --config= ：指定配置文件，缺省为 /etc/mdadm.conf

-s, --scan：扫描配置文件或 /proc/mdstat以搜寻丢失的信息。

1. 创建新的阵列，mdadm使用：

--create(或其缩写-C) 参数来创建新的陈列并且将一些重要阵列的标识信息作为元数据可写在每一个底层设备的指定区间

--level(或者其缩写-l) 表示阵列的RAID级别

--chunk(或者其缩写-c) 表示每个条带单元的大小，以KB为单位，默认为64KB，条带单元的大小配置对不同负载下的阵列读写性能有很大影响

--raid-devices(或者其缩写-n) 表示阵列中活跃的设备个数

--spare-devices(或者其缩写-x) 表示阵列中热备盘的个数，一旦阵列中的某个磁盘失效，MD内核驱动程序自动用将热备磁盘加入到阵列，然后重构丢失磁盘上的数据到热备磁盘上

1. 创建事例：

1）创建一个RAID 0设备：

mdadm --create /dev/md0 --level=0 --chunk=32 --raid-devices=3 /dev/sdb1 /dev/sdc1 /dev/sdd1

2）创建一个raid 1设备：

mdadm --create /dev/md0 --level=1 --chunk=128 --raid-devices=2 --spare-devices=1 /dev/sdb1 /dev/sdc1 /dev/sdd1

3）创建一个RAID5设备：

mdadm --create /dev/md0 --level=5 --raid-devices=5 /dev/sd[c-g]1 --spare-devices=1 /dev/sdb1

4）创建一个RAID 10设备：

mdadm -C /dev/md0 -l10 -n6 /dev/sd[b-g] -x1 /dev/sdh

5）创建一个RAID1 0设备：

mdadm -C /dev/md0 -l1 -n2 /dev/sdb /dev/sdc

mdadm -C /dev/md1 -l1 -n2 /dev/sdd /dev/sde

mdadm -C /dev/md2 -l1 -n2 /dev/sdf /dev/sdg

mdadm -C /dev/md3 -l0 -n3 /dev/md0 /dev/md1 /dev/md2

1. 其它事项：