Linux−RAID FAQ

Gregory Leblanc

Revision History

Revision v0.0.12	2003−03−05	Revised by: gml
Fleshed out questions that cover using mdadm, small formatting changes		
Revision v0.0.11	2003−01−08	Revised by: gml
Updated Archive Locations, information on when to patch, added a note about old patches being missing, removed question about the raidtools being dangerous (since they don't appear to be so labeled any longer).		
Revision v0.0.10	24 April 2001	Revised by: gml
Added a new section and question about benchmarking.		

This is a FAQ for the Linux−RAID mailing list, hosted on vger.kernel.org. vger.rutgers.edu is gone, so don't bother looking for it. It's intended as a supplement to the existing Linux−RAID HOWTO, to cover questions that keep occurring on the mailing list. PLEASE read this document before your post to the list.

Table of Contents

1. Linux RAID FAQ

1. General

1.1.1. Where can I find archives for the linux−raid mailing list?

The only archives left seem to be available at http://marc.theaimsgroup.com/?l=linux−raid&r=1&w=2

1.1.2. Where can I find the latest version of this FAQ?

The latest version of this FAQ will be available from the LDP website at http://www.tldp.org/FAQ/.

1.1.3. What sorts of things does this list cover?

Well, obviously this list covers RAID in relation to Linux. Most of the discussions are related to the raid code that's been built into the Linux kernel. There are also a few discussions on getting hardware based RAID controllers working using Linux as the operating system. Any and all of these discussions are valid for this list.

2. Kernel

1.2.1. I'm running *`[insert your linux distribution here]`*. Do I need to patch my kernel to make RAID work?

At this point, most major distributions are shipping with a 2.4 based kernel, which already includes the necessary patches. If your distribution is still using a 2.2.x kernel, upgrade!

If you download a 2.2.x kernel from ftp.kernel.org, then you will need to patch your kernel.

1.2.2. How can I tell if I need to patch my kernel?

That depends on which kernel series you're using. If you're using the 2.4.x kernels, then you've already got the latest RAID code that's available. If you're running 2.2.x, see the following instructions on how to find out.

The easiest way is to check what's in `/proc/mdstat`. Here's a sample from a 2.2.x kernel, *with* the RAID patches applied.

```
[gleblanc@grego1 gleblanc]$ cat /proc/mdstat
Personalities : [linear] [raid0] [raid1] [raid5] [translucent]
read_ahead not set
unused devices: <none>
```

If the contents of `/proc/mdstat` looks like the above, then you don't need to patch your kernel.

The "Personalities" line in your kernel may not look exactly like the above, if you have RAID compiled as modules. Most distributions will have RAID compiled as modules to save space on the boot diskette. If you're not using any RAID sets, then you will probably see a blank space at the end of the "Personalities" line, don't worry, that just means that the RAID modules aren't loaded yet.

Here's a sample from a 2.2.x kernel, *without* the RAID patches applied.

```
[root@serek ~]# cat /proc/mdstat
Personalities : [1 linear] [2 raid0]
read_ahead not set
md0 : inactive
md1 : inactive
md2 : inactive
md3 : inactive
```

If your `/proc/mdstat` looks like this one, then you need to patch your kernel.

1.2.3. Where can I get the latest RAID patches for my kernel?

The patches for the 2.2.x kernels up to, and including, 2.2.13 are available from ftp.kernel.org. Use the kernel patch that most closely matches your kernel revision. For example, the 2.2.11 patch can also be used on 2.2.12 and 2.2.13.

These patches are no longer available from this location! I haven't been able to find the new location for them, please email me if you know where they've gone.

The patches for 2.2.14 and later kernels are at http://people.redhat.com/mingo/raid−patches/. Use the right patch for your kernel, these patches haven't worked on other kernel revisions. Please use something like wget/curl/lftp to retrieve this patch, as it's easier on the server than using a client like Netscape. Downloading patches with Lynx has been unsuccessful for me; wget may be the easiest way.

1.2.4. How do I apply the patch to a kernel that I just downloaded from ftp.kernel.org?

First, unpack the kernel into some directory, generally people use `/usr/src/linux`. Change to this directory, and type **patch −p1 < /path/to/raid−version.patch**.

On my RedHat 6.2 system, I decompressed the 2.2.16 kernel into `/usr/src/linux−2.2.16`. From `/usr/src/linux−2.2.16`, I type in **patch −p1 < *`/home/gleblanc/raid−2.2.16−A0`***. Then I rebuild the kernel using **make menuconfig** and related builds.

1.2.5. What kind of drives can I use RAID with? Do only SCSI or IDE drives work? Do I need different patches for different kinds of drives?

Software RAID works with any block device in the Linux kernel. This includes IDE and SCSI drives, as well as most harware RAID controllers. There are no different patches for IDE drives vs. SCSI drives.

3. RAID tools

1.3.1. What tools are available for dealing with my Linux Software RAID arrays?

There are currently two sets of tools available. Both sets work quite well, and have essentially the same functionalty. I recommend the newer set of tools, because they're much easier to use, but I'll mention where to get the older tools as well.

The new set of tools is called mdadm. It doesn't have much of a homepage, but you can download tarballs and RPMs from http://www.cse.unsw.edu.au/~neilb/source/mdadm/. I suggest that anyone who isn't already familar with the 'raidtools' package use these (and in fact, I suggest that folks who already know the raidtools package switch over to these).

The older set of tools is called raidtools. They're available from http://people.redhat.com/mingo/raidtools/. I believe there are other locations available, since Red Hat Linux is shipping based on a tarball numbered 1.00.3, which I can't find online. If anybody knows where these are, please let me know.

4. Disk Failures and Recovery

1.4.1. How can I tell if one of the disks in my RAID array has failed?

A couple of things should indicate when a disk has failed. There should be quite a few messages in `/var/log/messages` indicating errors accessing that device, which should be a good indication that something is wrong.

You should also notice that your `/proc/mdstat` looks different. Here's a snip from a good /proc/mdstat

```
[gleblanc@grego1 gleblanc]$ cat /proc/mdstat
Personalities : [linear] [raid0] [raid1] [raid5] [translucent]
read_ahead not set
md0 : active raid1 sdb5[0] sda5[1] 32000 blocks [2/2] [UU]
unused devices: <none>
```

And here's one from a `/proc/mdstat` where one of the RAID sets has a missing disk.

```
[gleblanc@grego1 gleblanc]$ cat /proc/mdstat
Personalities : [linear] [raid0] [raid1] [raid5] [translucent]
read_ahead not set
md0 : active raid1 sdb5[0] sda5[1] 32000 blocks [2/1] [U_]
unused devices: <none>
```

I don't know if `/proc/mdstat` will reflect the status of a HOT SPARE. If you have set one up, you should be watching `/var/log/messages` for any disk failures. I'd like to get some logs of a disk failure, and `/proc/mdstat` from a system with a hot spare.

1.4.2. So my RAID set is missing a disk, what do I do now?

Software−RAID generally doesn't mark a disk as bad unless it is, so you probably need a new disk. Most decent quality disks have a 3 year warranty, but some exceptional (and expensive) SCSI hard drives may have wararnties as long as 5 years, or even longer. More and more hard drive vendors are giving a 1 year warranty on their "consumer" drives. I suggest avoiding any drive with a 1 year warranty if at all possible. Try to have the manufacturer replace the failed disk if it's still under warranty.

When you get the new disk, power down the system, and install it, then partition the drive so that it has partitions the size of your missing RAID partitions. Once you have the partitions set up properly, just run **mdadm −−add *`/dev/md0`* *`/dev/hdc1`***, where *`/dev/md0`* is the RAID array you're adding the partition to, and *`/dev/hdc1`* is the partition that you're trying to add. Reconstruction should start immediately.

If you would prefer to use the RAIDtools suite, you can use the command **raidhotadd** to put the new disk into the array and begin reconstruction. See Chapter 6 of the Software RAID HOWTO for more information.

1.4.3. dmesg shows "md: serializing resync, *`md4`* has overlapping physical units with *`md5`*" (where md4 and md5 are two of your software RAID devices). What does this mean?

In that message "physical units" refers to disks, and not to blocks on the disks. Since there is more than one RAID array that needs resyncing on one of the disks in use for your RAID arrays, the RAID code is going to sync md4 first, and md5 second, to avoid excessive seeks (also called thrashing), which would drastically slow the resync process.

5. Benchmarking

1.5.1. How should I benchmark my RAID devices? Are there any tools that work particularly well?

There are really a few options for benchmarking your RAID array, depending on what you're looking to test. RAID offers the greatest speed increases when there are multiple threads reading from the same RAID volume.

One tool specificly designed to test and show off these performance gains is tiobench. It uses multiple read and write threads on the disk, and has some pretty good reporting.

Another good tool to use is bonnie++. It seems to be more targeted at benchmarking single drives that at RAID, but still provides useful information.

One tool *NOT* to use is hdparm. It does not give useful performance numbers for any drives that I've heard about, and has been known to give some incredibly off−the−wall numbers as well. If you want to do *real* benchmarking, use one of the tools listed above.