Re: Problem with Raid Array persistence across reboots.

Source: http://linux.derkeiler.com/Mailing−Lists/Debian/2006−08/msg02572.html

- *From*: Miles Fidelman <mfidelman@xxxxxxxxxxxxxxx>
- *Date*: Fri, 25 Aug 2006 10:57:24 −0400

I recently set up a Sarge box, with raid and LVM. It took a while, and there where a bunch of nits that had to be worked through, but now it's been running stably for a month or so, and everthing comes back back after reboots.

So, in case it helps:

− I found the guidance at http://www.planamente.ch/emidio/pages/linux_howto_root_lvm_raid.php
to be very helpful

− here's a copy of the install report I filed afterwards, which includes a blow−by−blow − including comments on what didn't work from the recipe above

Package: installation−reports

Boot method: netinstall CD
Image version:
3.1r2, downloaded July 2006 (not sure of the exact date)
downloaded from
http://cdimage.debian.org/debian−cd/3.1_r2/i386/iso−cd/debian−31r2−i386−netinst.iso

Date: 7/20−24/06

Machine: Supermicro P8SCT 1U rack−mount server
Processor: P4 processor 640 3.2GHZ 2MB CACHE L775 CPU (i686 architecture)
Memory: 4G (3G usable, the other 1G overlaps system resources)
Partitions: <df −Tl will do; the raw partition table is preferred>

server1:~# fdisk −l

Disk /dev/sda: 400.0 GB, 400088457216 bytes
255 heads, 63 sectors/track, 48641 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

Device Boot Start End Blocks Id System
/dev/sda1 * 1 122 979933+ fd Linux raid

autodetect
/dev/sda2 123 450 2634660 82 Linux swap / Solaris
/dev/sda3 451 12608 97659135 fd Linux raid
autodetect
/dev/sda4 12609 36923 195310237+ 5 Extended
/dev/sda5 12609 36923 195310206 fd Linux raid
autodetect

Disk /dev/sdb: 400.0 GB, 400088457216 bytes
255 heads, 63 sectors/track, 48641 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

Device Boot Start End Blocks Id System
/dev/sdb1 * 1 122 979933+ fd Linux raid
autodetect
/dev/sdb2 123 450 2634660 fd Linux raid
autodetect
/dev/sdb3 451 12608 97659135 fd Linux raid
autodetect
/dev/sdb4 12609 36923 195310237+ 5 Extended
/dev/sdb5 12609 36923 195310206 fd Linux raid
autodetect

Disk /dev/sdc: 400.0 GB, 400088457216 bytes
255 heads, 63 sectors/track, 48641 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

Device Boot Start End Blocks Id System
/dev/sdc1 * 1 122 979933+ fd Linux raid
autodetect
/dev/sdc2 123 450 2634660 82 Linux swap / Solaris
/dev/sdc3 451 12608 97659135 fd Linux raid
autodetect
/dev/sdc4 12609 36923 195310237+ 5 Extended
/dev/sdc5 12609 36923 195310206 fd Linux raid
autodetect

Disk /dev/sdd: 400.0 GB, 400088457216 bytes
255 heads, 63 sectors/track, 48641 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

Device Boot Start End Blocks Id System
/dev/sdd1 * 1 122 979933+ fd Linux raid
autodetect
/dev/sdd2 123 450 2634660 fd Linux raid
autodetect
/dev/sdd3 451 12608 97659135 fd Linux raid
autodetect
/dev/sdd4 12609 36923 195310237+ 5 Extended
/dev/sdd5 12609 36923 195310206 fd Linux raid
autodetect

raid set up as follows:
md3 : active raid1 sdb5[0] sdc5[2] sdd5[1]
195310080 blocks [2/2] [UU]

md0 : active raid1 sda1[0] sdb1[2] sdc1[1]
979840 blocks [2/2] [UU]

md2 : active raid1 sdb3[2] sda3[0] sdc3[1]
97659008 blocks [2/2] [UU]

md1 : active raid1 sda2[0] sdb2[2] sdc2[1]
2634560 blocks [2/2] [UU]

/dev/md0 mounted as /boot (ext3)
/dev/md1 allocated for swap
/dev/md2 allocated as LVM pv /dev/rootvolume
/dev/md3 allocated as LVM pv /dev/backupvolume

/dev/mapper/rootvolume−rootlv on / type ext3 (rw,errors=remount−ro)
/dev/mapper/backupvolume−backuplv on /backup type ext3 (rw)

Output of lspci and lspci −n:

server1:/etc# lspci
0000:00:00.0 Host bridge: Intel Corp. Server Memory Controller Hub (rev 05)
0000:00:01.0 PCI bridge: Intel Corp. Server Memory Controller Hub PCI
Express Port (rev 05)
0000:00:02.0 VGA compatible controller: Intel Corp. Graphics Controller
(rev 05)
0000:00:1c.0 PCI bridge: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6 Family)
PCI Express Port 1 (rev 03)
0000:00:1c.1 PCI bridge: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6 Family)
PCI Express Port 2 (rev 03)
0000:00:1d.0 USB Controller: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6
Family) USB UHCI #1 (rev 03)
0000:00:1d.1 USB Controller: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6
Family) USB UHCI #2 (rev 03)
0000:00:1d.2 USB Controller: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6
Family) USB UHCI #3 (rev 03)
0000:00:1d.3 USB Controller: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6
Family) USB UHCI #4 (rev 03)
0000:00:1d.7 USB Controller: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6
Family) USB2 EHCI Controller (rev 03)
0000:00:1e.0 PCI bridge: Intel Corp. 82801 PCI Bridge (rev d3)
0000:00:1f.0 ISA bridge: Intel Corp. 82801FB/FR (ICH6/ICH6R) LPC
Interface Bridge (rev 03)
0000:00:1f.1 IDE interface: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6
Family) IDE Controller (rev 03)
0000:00:1f.2 IDE interface: Intel Corp. 82801FR/FRW (ICH6R/ICH6RW) SATA
Controller (rev 03)

0000:00:1f.3 SMBus: Intel Corp. 82801FB/FBM/FR/FW/FRW (ICH6 Family)
SMBus Controller (rev 03)
0000:01:00.0 PCI bridge: Intel Corp. PCI Bridge Hub (rev 09)
0000:01:00.1 PIC: Intel Corp. PCI Bridge Hub I/OxAPIC Interrupt
Controller A (rev 09)
0000:03:00.0 Ethernet controller: Broadcom Corporation NetXtreme BCM5721
Gigabit Ethernet PCI Express (rev 11)
0000:04:00.0 Ethernet controller: Broadcom Corporation NetXtreme BCM5721
Gigabit Ethernet PCI Express (rev 11)

server1:/etc# lspci −n
0000:00:00.0 0600: 8086:2588 (rev 05)
0000:00:01.0 0604: 8086:2589 (rev 05)
0000:00:02.0 0300: 8086:258a (rev 05)
0000:00:1c.0 0604: 8086:2660 (rev 03)
0000:00:1c.1 0604: 8086:2662 (rev 03)
0000:00:1d.0 0c03: 8086:2658 (rev 03)
0000:00:1d.1 0c03: 8086:2659 (rev 03)
0000:00:1d.2 0c03: 8086:265a (rev 03)
0000:00:1d.3 0c03: 8086:265b (rev 03)
0000:00:1d.7 0c03: 8086:265c (rev 03)
0000:00:1e.0 0604: 8086:244e (rev d3)
0000:00:1f.0 0601: 8086:2640 (rev 03)
0000:00:1f.1 0101: 8086:266f (rev 03)
0000:00:1f.2 0101: 8086:2652 (rev 03)
0000:00:1f.3 0c05: 8086:266a (rev 03)
0000:01:00.0 0604: 8086:032c (rev 09)
0000:01:00.1 0800: 8086:0326 (rev 09)
0000:03:00.0 0200: 14e4:1659 (rev 11)
0000:04:00.0 0200: 14e4:1659 (rev 11)

Base System Installation Checklist:
[O] = OK, [E] = Error (please elaborate below), [ ] = didn't try it

Initial boot worked: [O]
Configure network HW: [O]
Config network: [O]
Detect CD: [O]
Load installer modules: [O]
Detect hard drives: [E]
Partition hard drives: [E]
Create file systems: [O]
Mount partitions: [O]
Install base system: [E]
Install boot loader: [E]
Reboot: [E]

Comments/Problems:

Detect Hard Drives:

− default kernel (2.4) couldn't detect drives
− needed to run linux26/expert26 to detect SATA drives − then worked ok
− could use a little better documentation

Partition Hard Drives:
− this was VERY time consuming and error prone
− I was using the combination of md and lvm, which is very poorly documented, and doesn't work completely (at least for sarge)
− things I discovered along the way:
−−−− after specifying a RAID set, things work badly unless you wait for the disks to sync (alt−f2, watch and wait, for each raid set − wait a long time)
−−−− after creating raids and trying to write them, I got errors along the lines of note: "unable to re−read the partition table, you should reboot" − turns out that it's ok to ignore these, but you have to go into and out of the partition editor a couple of times to move on to the next step of configuring LVM
−−−− create file systems worked after figuring out the md and lvm setups
− the documentation needs a lot of work for this. some suggestions:

−− sections 6.3.2.2 and .3 should be reversed (setting up RAIDs needs to happen before setting up logical volumes!)
−− need a lot more discussion about complex setups (lvm over raid, and so forth) − I found http://www.planamente.ch/emidio/pages/linux_howto_root_lvm_raid.php to be the best resource − suggest either adding a link, or incorporating text from it

Install boot loader and reboot:
− neither lilo nor grub set up properly after the disk partioning
− got errors about not being able to install (tried a lot of different locations)
− had to use the install disk as a rescue disk to boot and figure out the proper paths to the kernel and initrd
− it seemed to get grub into the MBR on one hard drive, so I was able to reboot − which got me into grub
− manually booted via grub
− manually installed grub into MBRs of other disks
− had to set up menu.lst manually
− then everything pretty much worked

A few more nits that I had to clean up manually:

− the −386 kernel only recognized 1G of 3G: I had to install a −686 version of the kernel and reboot to see all my memory (might want to add more kernel choices to the installer and/or put some notes in sec. 3.6.4 of the installation manual regarding what to do if you don't see all your memory − all that's mentioned now is to try using the mem= kernel option, which doesn't help)

− dmesg contains about 165 lines of

devfs_mk_dir: invalid argument.<4>devfs_mk_dev: could not append to
parent for /disc
which turns out to be a harmless, and wrong error (apparantly, there's a
bug in the devfs code such that it's printing this error instead of
a "device not found" error as it cycles through all the possible /md
devices − and nobody is planning on fixing this, since devfs is going
away) − it might be worth a note in the documentation somewhere, though,
that this can be safely ignored

− the installer did not quite get swapping or the fstab right, I had to
> mkswap/dev/md1
> swapon

and add the following to fstab:
/dev/md0 /boot ext3 defaults 0 0
/dev/md1 none swap sw 0 0

and there seems to be some kind of subtle bug or interactions in the
raid code:

I have spare devices configured for each raid, but they didn't show up
on initial boot. As best as I can reconstruct from the boot logs:

/dev/md0 − which mounts as boot, tried to bind it's spare device, and
then kicked it out as being out of date. I did a manual
mdadm /dev/md0 −a /dev/sdb1
and it mounted cleanly, and then automounted on subsequent reboots.

/dev/md2 − which mounts as / − would not mount it's spare device at all,
and there's no indication in the boot log that it even tries. I finally
gave up and put a crude work−around, by executing "mdadm /dev/md2 −a
/dev/sdb3" in a locally defined init file (executed via rcS.d).

One last suggestion: it might be worth enabling bootlog to be on by
default − it would make debugging a lot easier, rather than first having
to figure out that it exists and then how to turn it on.

Overall: not a complete snap, but given that mine is a somewhat
complicated setup, and I'm coming at this from a Solaris and Red Hat
background, I'd consider this a reasonably smooth setup.

Now I'm off to load and configure application software and migrate apps
from my old Red Hat box (and not completely happy that Debian uses Exim
instead of Sendmail by default, sigh...).

−−
To UNSUBSCRIBE, email to debian−user−REQUEST@xxxxxxxxxxxxxxx with a subject of "unsubscribe".
Trouble? Contact listmaster@xxxxxxxxxxxxxxx