# Einrichten eines Software-RAID unter eisfair-2

Yves Schumann

08/2010

## Inhalt

Aufgabe	2
Ausgangssituation	2
Umsetzung (RAID-Setup):	3
Module laden	3
Platten partitionieren	3
Sektorgrössen anpassen (optional)	5
Reste eines eventuellen alten RAID-Verbundes löschen	6
RAID-Verbund anlegen	6
RAID Synchronisation	6
Dateisystem anlegen	7
RAID mounten	7
Umsetzung (Dauerhafte Integration)	8
mdadm-Konfiguration aktualisieren	8
fstab aktualisieren	9
GRUB-Konfiguration aktualisieren	10
Reboot & Check	11

## Aufgabe

Gegeben ist ein auf einer Platte installiertes System, in welches zwei weitere Platten als RAID-Verbund integriert werden sollen. In diesem Fall ist das ein stromsparendes Atom-Board, welches mit einem IDE und zwei SATA-Anschlüssen ausgestattet ist. Am IDE-Port befindet sich eine kleine 2,5" Notebook-Platte mit dem System, an den beiden SATA-Anschlüssen jeweils eine grosse Platte für den RAID-Verbund.

## Ausgangssituation

Es wurde eine Platte angeschlossen (hier eine 2,5" IDE-Platte) und das System komplett installiert. Dann wurden die beiden zusätzlichen Platten (hier 3,5" SATA-II) eingebaut. Damit ergibt sich bspw. folgendes Bild:

root@kessel:~# fdisk -1 Disk /dev/sda: 40.0 GB, 40007761920 bytes 255 heads, 63 sectors/track, 4864 cylinders Units = cylinders of 16065 \* 512 = 8225280 bytes Disk identifier: 0x0000000 Device Boot Start /dev/sda1 \* 1 /dev/sda2 10 End 9 267 Blocks Id System 72292 83 Linux 2072385 82 Linux swap / Solaris 268 4864 36925402+ 83 Linux /dev/sda3 Disk /dev/sdb: 2000.3 GB, 2000398934016 bytes 255 heads, 63 sectors/track, 243201 cylinders Units = cylinders of 16065 \* 512 = 8225280 bytes Disk identifier: 0x0000000 Disk /dev/sdb doesn't contain a valid partition table Disk /dev/sdc: 2000.3 GB, 2000398934016 bytes 255 heads, 63 sectors/track, 243201 cylinders Units = cylinders of 16065 \* 512 = 8225280 bytes Disk identifier: 0x0000000 Disk /dev/sdc doesn't contain a valid partition table root@kessel:~#

Die Platte sda ist die IDE-Systemplatte, sdb und sdc soll zum Software-RAID verbunden werden.

## **Umsetzung (RAID-Setup):**

## **Module laden**

Da das System ohne RAID aufgesetzt wurde, müssen zunächst die entsprechenden Module geladen werden:

```
root@kessel:~# modprobe raid1
```

Nun sollten in der Ausgabe von Ismod die folgenden beiden Einträge zu finden sein:

```
root@kessel:~# lsmod
...
raid1 25728 1
md_mod 81940 2 raid1
...
```

## **Platten partitionieren**

Nun wird auf den beiden SATA-Platten jeweils eine Partition angelegt und deren Typ gesetzt. Das wird mit fdisk erledigt. Folgende Schritte werden ausgeführt:

- Partition anlegen (hier eine primäre Partition voller Grösse der Platte)
- Typ der Partition auf "Linux raid auto" ändern
- Partitionstabelle schreiben

```
root@kessel:~# fdisk /dev/sdb
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF disklabel
Building a new DOS disklabel with disk identifier 0x4e29a19c.
Changes will remain in memory only, until you decide to write them.
After that, of course, the previous content won't be recoverable.
The number of cylinders for this disk is set to 243201.
There is nothing wrong with that, but this is larger than 1024,
and could in certain setups cause problems with:
1) software that runs at boot time (e.g., old versions of LILO)
2) booting and partitioning software from other OSs
   (e.g., DOS FDISK, OS/2 FDISK)
Warning: invalid flag 0x0000 of partition table 4 will be corrected by w(rite)
Command (m for help): p
Disk /dev/sdb: 2000.3 GB, 2000398934016 bytes
255 heads, 63 sectors/track, 243201 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Disk identifier: 0x4e29a19c
                  Start End Blocks Id System
   Device Boot
Command (m for help): n
Command action
  e extended
     primary partition (1-4)
   р
P
Partition number (1-4): 1
First cylinder (1-243201, default 1): <Return>
Using default value 1
Last cylinder or +size or +sizeM or +sizeK (1-243201, default 243201): <Return>
```

Using default value 243201 Command (m for help): p Disk /dev/sdb: 2000.3 GB, 2000398934016 bytes 255 heads, 63 sectors/track, 243201 cylinders Units = cylinders of 16065 \* 512 = 8225280 bytes Disk identifier: 0x4e29a19c Start End 211 1 243201 1953512001 Device Boot Blocks Id System /dev/sdb1 83 Linux Command (m for help): t Selected partition 1 Hex code (type L to list codes): L 1e Hidden W95 FAT1 80 Old Minix be Solaris boot 0 Empty 24 NEC DOS 1 FAT12 81 Minix / old Lin bf Solaris 39 Plan 9 82 Linux swap / So c1 DRDOS/sec (FAT-2 XENIX root 3 XENIX usr 3c PartitionMagic 83 Linux c4 DRDOS/sec (FAT-40 Venix 80286 41 PPC PReP Boot 84 OS/2 hidden C: c6 DRDOS/sec (FAT-85 Linux extended c7 Syrinx 4 FAT16 <32M 5 Extended 42 SFS 86 NTFS volume set da Non-FS data 6 FAT16 x 4d QNX4.x 87 NTFS volume set db CP/M / CTOS / . 7 HPFS/NTFS 8AIX4eQNX4.x 2nd part 88Linux plaintext deDell Utility9AIX bootable4fQNX4.x 3rd part 8eLinux LVMdfBootItaOS/2 Boot Manag 50OnTrack DM93Amoebae1DOS accessbW95 FAT3251OnTrack DM6 Aux 94Amoeba BBTe3DOS R/O c W95 FAT32 51 OnTrack DM6 Aux 94 Amoeba BBT e3 DOS R/O c W95 FAT32 (LBA) 52 CP/M 9f BSD/OS e4 SpeedStor e W95 FAT16 (LBA) 53 OnTrack DM6 Aux a0 IBM Thinkpad hi eb BeOS fs f W95 Ext'd (LBA) 54 OnTrackDM6 a5 FreeBSD ee EFI GPT 55 EZ-Drive 56 Golden Bow a6 OpenBSD a7 NeXTSTEP ef EFI (FAT-12/16/ f0 Linux/PA-RISC b 10 OPUS Induced FAIL256Golden Bowa7NeXTSTEPf0Linux/PA-I12Compaq diagnost 5cPriam Ediska8Darwin UFSf1SpeedStor14Hidden FAT16 <3 61</td>SpeedStora9NetBSDf4SpeedStor16Hidden FAT1663GNUL HURD on Sure 11T f2 DOS secondary 17Hidden HPFS/NTF 64Novell Netware b7BSDI fsfdLinux raid auto18AST SmartSleep 65Novell Netware b8BSDI swapfeLANstep1bHidden W95 FAT3 70DiskSecure Mult bbBoot Wizard hid ffBBT 1c Hidden W95 FAT3 75 PC/IX Hex code (type L to list codes): fd Changed system type of partition 1 to fd (Linux raid autodetect) Command (m for help): p Disk /dev/sdb: 2000.3 GB, 2000398934016 bytes 255 heads, 63 sectors/track, 243201 cylinders Units = cylinders of 16065 \* 512 = 8225280 bytes Disk identifier: 0x4e29a19c Device Boot End Blocks Id System Start /dev/sdb1 243201 1953512001 fd Linux raid autodetect 1 Command (m for help): w The partition table has been altered! Calling ioctl() to re-read partition table. Syncing disks. root@kessel:~#

Selbiges muss nun auch für /dev/sdc durchgespielt werden.

## Sektorgrössen anpassen (optional)

Wenn die verwendeten Platten Sektorgrössen von 4k unterstützen, sollten diese auch verwendet werden. Dafür ist eine Verschiebung des Startpunktes der Daten auf einen durch acht teilbaren Wert notwendig. Das wird ebenfalls mit fdisk durch die folgenden Schritte erreicht:

- Expertenmodus aktivieren
- Beginn der Daten in der Partition auf durch 8 teilbaren Wert setzen (hier vom Default-Wert 63 auf 64)
- Partitionstabelle schreiben

```
root@kessel:~# fdisk /dev/sdb
The number of cylinders for this disk is set to 243201.
There is nothing wrong with that, but this is larger than 1024,
and could in certain setups cause problems with:
1) software that runs at boot time (e.g., old versions of LILO)
2) booting and partitioning software from other OSs
   (e.g., DOS FDISK, OS/2 FDISK)
Command (m for help): p
Disk /dev/sdc: 2000.3 GB, 2000398934016 bytes
255 heads, 63 sectors/track, 243201 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes
Disk identifier: 0x1361d2f0
   Device BootStartEndBlocksev/sdc112432011953512001
                                                   Id System
fd Linux raid autodetect
                                            Blocks
/dev/sdc1
Command (m for help): x
Expert command (m for help): m
Command action
   b move beginning of data in a partition
   c change number of cylinders
   d print the raw data in the partition table
   e list extended partitions
     fix partition order
   f
      create an IRIX (SGI) partition table
   q
      change number of heads
   h
   i change the disk identifier
   m print this menu
   p print the partition table
     quit without saving changes
   q
   r
      return to main menu
     change number of sectors/track
   S
     verify the partition table
   V
   w write table to disk and exit
Expert command (m for help): b
Partition number (1-4): 1
New beginning of data (63-3907024064, default 63): 64
Expert command (m for help): w
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
root@kessel:~#
```

Selbiges muss nun auch für /dev/sdc durchgespielt werden.

## Reste eines eventuellen alten RAID-Verbundes löschen

Falls die Platten bereits in einem RAID-Verbund verwendet worden sind, muss sichergestellt werden, dass keine "Reste" der alten RAID-Konfiguration vorhanden sind:

```
root@kessel:~# mdadm --zero-superblock /dev/sdb1
mdadm: Unrecognised md component device - /dev/sdb1
root@kessel:~# mdadm --zero-superblock /dev/sdc1
mdadm: Unrecognised md component device - /dev/sdc1
root@kessel:~#
```

Wenn keine Reste eines früheren RAID-Verbundes vorhanden waren, wird die oben zu sehende Meldung ausgegeben. Das ist an dieser Stelle kein Problem und bei neuen Platten normal.

#### **RAID-Verbund anlegen**

Die SATA-Platten sind nun vorbereitet, um sie zu einem RAID-Verbund zusammenzuführen. Somit wird im nächsten Schritt das RAID mit mdadm wie folgt angelegt:

```
root@kessel:~# mdadm --create /dev/md0 --level=1 --raid-disks=2 /dev/sdb1 /dev/sdc1
mdadm: excess address on MAIL line: mdadm/mail_to - ignored
mdadm: excess address on MAIL line: doesnt exist - ignored
mdadm: array /dev/md0 started.
root@kessel:~#
```

## **RAID Synchronisation**

Jetzt beginnt die Synchronisation der RAID-Disks, welche man bis zum Ende abwarten sollte. Das wird je nach Grösse der Disks mehrere Stunden dauern. Den aktuellen Status der Synchronisation kann man wie folgt anzeigen lassen:

Beim Aufruf mit 'watch cat /proc/mdstat' erhält man eine alle zwei Sekunden aktualisierte Ausgabe und kann den Sync-Prozess somit nebenher laufen lassen und im Auge behalten:

Zurück auf die Konsole gelangt man mit Ctrl+C.

## **Dateisystem anlegen**

Nachdem die Synchronisation abgeschlossen worden ist, wird das Dateisystem angelegt. Dieser Vorgang wird einige Minuten dauern:

```
root@kessel:~# mkfs.ext3 /dev/md0
mke2fs 1.40.8 (13-Mar-2008)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
122101760 inodes, 488377984 blocks
24418899 blocks (5.00%) reserved for the super user
First data block=0
Maximum filesystem blocks=0
14905 block groups
32768 blocks per group, 32768 fragments per group
8192 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632, 2654208,
        4096000, 7962624, 11239424, 20480000, 23887872, 71663616, 78675968,
        102400000, 214990848
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
This filesystem will be automatically checked every 24 mounts or
180 days, whichever comes first. Use tune2fs -c or -i to override.
root@kessel:~#
```

#### **RAID mounten**

Damit ist das RAID eingerichtet und kann gemountet werden. Hier in diesem HowTo soll der RAID-Verbund nach /data/backup gemountet werden:

```
root@kessel:~# mkdir /data/backup
root@kessel:~# ls -l /data/
total 8
drwxr-xr-x 2 root root 4096 Aug 28 16:45 backup
drwxr-xr-x 2 root root 4096 Jun 14 15:22 packages
root@kessel:~# mount /dev/md0 /data/backup
root@kessel:~#
```

Nun ist der RAID-Verbund dem System bis zum nächsten Reboot bekannt:

root@kesse	el:~# mount		#(Ausgabe für das HowTo formatiert)	
/dev/sda3	on /	type ext3	<pre>(rw,errors=remount-ro,acl,user_xattr)</pre>	
proc	on /proc	type proc	(rw,noexec,nosuid,nodev)	
/sys	on /sys	type sysfs	(rw,noexec,nosuid,nodev)	
varrun	on /var/run	type tmpfs	(rw,noexec,nosuid,nodev,mode=0755)	
udev	on /dev	type tmpfs	(rw,mode=0755)	
devshm	on /dev/shm	type tmpfs	(rw)	
devpts	on /dev/pts	type devpts	(rw,gid=5,mode=620)	
/dev/sda1	on /boot	type ext3	(rw,errors=remount-ro,acl,user_xattr)	
/dev/md0	on /data/backup	type ext3	(rw)	
root@kessel:~#				
root@kessel:~# <b>df -h</b>				

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/sda3	35G	408M	33G	2%	/
varrun	1010M	36K	1010M	18	/var/run
udev	1010M	40K	1010M	18	/dev
devshm	1010M	0	1010M	0 %	/dev/shm
/dev/sda1	69M	14M	51M	22%	/boot
/dev/md0	1.9T	196M	1.8T	18	/data/backup
root@kessel:~#					

## **Umsetzung (Dauerhafte Integration)**

Natürlich soll der RAID-Verbund auch nach einem Reboot zur Verfügung stehen. Dafür sind die folgenden Schritte notwendig.

## mdadm-Konfiguration aktualisieren

Zunächst muss die mdadm-Konfigurationsdatei aktualisiert werden:

```
root@kessel:~# mdadm --examine --scan >> /etc/mdadm/mdadm.conf
mdadm: excess address on MAIL line: mdadm/mail_to - ignored
mdadm: excess address on MAIL line: doesnt exist - ignored
root@kessel:~#
```

Die Konfiguration sollte nun in etwa wie folgt aussehen:

```
root@kessel:~# cat /etc/mdadm/mdadm.conf
# mdadm.conf
#
# Please refer to mdadm.conf(5) for information about this file.
#
# by default, scan all partitions (/proc/partitions) for MD superblocks.
# alternatively, specify devices to scan, using wildcards if desired.
DEVICE partitions
# auto-create devices with Debian standard permissions
CREATE owner=root group=disk mode=0660 auto=yes
# automatically tag new arrays as belonging to the local system
HOMEHOST <system>
# instruct the monitoring daemon where to send mail alerts
MAILADDR 10 mdadm/mail_to doesn't exist
# This file was auto-generated on Mon, 14 Jun 2010 13:02:32 +0000
# by mkconf $Id$
ARRAY /dev/md0 level=raid1 num-devices=2 UUID=5048f0d3:e561afa8:36b8364f:ac71c0cc
root@kessel:~#
```

Nachdem die Array-Zeile eingefügt wurde, müssen noch die folgenden Anpassungen in der mdadm.config vorgenommen werden:

1. Eintragung der zu verwendenden Platten unter DEVICE:

## Eintrag alt:

DEVICE partitions

### Eintrag neu:

DEVICE /dev/sdb[0-9] /dev/sdc[0-9]

### 2. Erweitern der ARRAY-Zeile um die metadata-Angabe nach folgendem Schema

#### Eintrag alt:

ARRAY /dev/md0 level=raid1 num-devices=2 UUID=8f00...

#### Eintrag neu:

ARRAY /dev/md0 level=raid1 num-devices=2 metadata=00.90 UUID=8f00...

3. Prüfen und ggf. korrigieren der eingetragenen UUID.

In der Ausgabe von 'blkid' ist die UUID der beiden Platten im RAID-Verbund zu sehen. Diese ist für beide Platten identisch und muss auch in der mdadm.config stehen. Achtung, die Schreibweise unterscheidet sich geringfügig, bei der Ausgabe von blkid werden die UUIDs mit "-" getrennt, in der mdadm.config mit ":" und noch dazu an anderer Stelle. Hier ein korrektes Beispiel:

```
root@kessel:~# blkid
...
/dev/sdb1: UUID="f728db1b-06e6-0380-36b8-364fac71c0cc" TYPE="linux_raid_member"
/dev/sdc1: UUID="f728db1b-06e6-0380-36b8-364fac71c0cc" TYPE="linux_raid_member"
...
root@kessel:~# cat /etc/mdadm/mdadm.conf | grep ARRAY
ARRAY /dev/md0 level=raid1 num-devices=2 metadata=00.90
UUID=f728db1b:06e60380:36b8364f:ac71c0cc
root@kessel:~#
```

## fstab aktualisieren

Nun muss der RAID-Verbund in die fstab eingetragen werden. Bei einer Standard-Installation von eisfair-2 ohne separate Datenpartition sieht die fstab bspw. wie folgt aus:

root@kessel:~#	cat /etc/fstab	# (Au	nsgabe für das HowTo formatiert)		
proc	/proc	proc	defaults	0	0
UUID=41b0	/	ext3	<pre>defaults,errors=remount-ro,acl,user_xattr</pre>	0	1
UUID=c99e	/boot	ext3	<pre>defaults,errors=remount-ro,acl,user_xattr</pre>	0	1
UUID=a101	none	swap	SW	0	0
/dev/cdrom	/media/cdrom	udf,iso9660	user,noauto,exec,utf8	0	0
/dev/fd0 root@kessel:~#	/media/floppy	auto	rw,user,noauto,exec,utf8	0	0

Um nun den RAID-Verbund in die fstab einzutragen, wird die UUID benötigt. Diese wird entweder wieder mit 'blkid' oder mit 'vol\_id' wie folgt ermittelt:

```
root@kessel:~# vol_id -u /dev/md0
d2dcblcd-2ba4-4f89-8bee-8c44213b2848
root@kessel:~#
```

Mit dieser UUID wir nun der entsprechende Eintrag in die fstab eingefügt:

root@kessel:~# vi /etc/fstab

und eine Zeile nach folgendem Schema eintragen:

```
UUID=<hier-UUID-eintragen> /data/backup ext3 defaults,errors=remount-ro,acl,user_xattr 0 2
```

## **GRUB-Konfiguration aktualisieren**

Im letzten Schritt muss in der Grub-Konfig das Laden der RAID-Module aktiviert werden. Dafür die Datei /boot/grub/menu.lst im Editor geöffnet:

root@kessel:~# vi /boot/grub/menu.lst

und die Kernel-Zeile um die Angabe "raid=1" ergänzt werden:

Eintrag alt:

kernel /boot/vmlinuz root=UUID=41b0... ro quiet console=tty7 kmapger

## Eintrag neu:

kernel /boot/vmlinuz root=UUID=41b0... ro quiet console=tty7 kmapger raid=1

## **Reboot & Check**

An dieser Stelle ist das Setup des RAID-Verbund und dessen dauerhafte Integration abgeschlossen. Das sollte nun durch einen Reboot überprüft werden:

root@kessel:~# reboot ... root@kessel:~# df -h Filesystem Size Used Avail Use% Mounted on /dev/sda3 35G 408M 33G 2% / varrun 1010M 32K 1010M 1% /var/run udev 1010M 48K 1010M 1% /dev devshm 1010M 0 1010M 0% /dev/shm /dev/sda1 69M 14M 51M 22% /boot /dev/md0 1.9T 196M 1.8T 1% /data/backup root@kessel:~#