Raid1-mini-Howto

Paolo Subiaco psubiaco@creasol.it - http://www.creasol.it

17 febbraio 2003

1 Introduzione

Scopo di questo documento è la descrizione sommaria dell'installazione di un sistema raid1 (mirroring) in Linux con kernel 2.4.9 o superiori).

Il sistema è stato provato con la distribuzione Mandrake 8.0, installata in un hdd con le seguenti partizioni:

- hda1 /boot 50MB
- hda5 /swap 250MB
- hda6 / 1.5 GB
- hda7 /usr ▷
- hda8 /var
- hda9 /home 12GB

Ovviamente non sono necessarie tutte queste partizioni, ma si può semplicemente optare per una piccola partizione di boot, una partizione di swap ed una di root.

Tutte le partizioni dovrebbero funzionare in raid: nel mio caso, essendo i due hard disk nello stesso canale IDE, ho evitato di utilizzare la partizione di swap in raid (la comunicazione sarebbe molto appesantita durante la scrittura) cosicché, nel caso di crash di un HDD, il computer si bloccherà per errore sull'uso delle due partizioni di swap (presenti in entrambi i dischi).

Il PC era dotato di un altro hdd, installato purtroppo nello stesso canale IDE (come hdb) mentre è preferibile installarlo in un altro canale (ad esempio come hdc o hdd).

Per il funzionamento sono richiesti

- raidtools 0.90 o superiori
- lilo 21.7 o superiori

2 Passaggio da un HDD al sistema raid1

Installato linux in un solo hdd, vorremo portare il nostro sistema ad utilizzare un raid1 composto di due hdd identici (ma potrebbero anche non esserlo: basta che le partizioni da unire in raid siano circa della stessa dimensione) di cui uno è quello in cui si è installato linux. La soluzione è possibile, ed ora vedremo i passi per attuarla.

2.1 Installazione del kernel

Installate il kernel 2.4.9 o superiori (magari funziona anche con kernel dalla versione 2.4.5 in poi o con quelli precedenti dopo una modifica al Makefile in driver/md).

Il kernel deve essere configurato per supportare il MULTI-DEVICE SUPPORT (RAID AND LVM), poi compilato ed installato in /boot; ovviamente modificare lilo.conf per caricare il nuovo kernel, e digitare *lilo* per aggiornare il MBR dell'hard disk hda. Quindi riavviare il computer con il nuovo kernel.

2.2 Partizionamento hard disk hdc

Si dovranno creare partizioni della stessa dimensione (circa) di quelle esistenti in hda.

Il tipo della partizione dovrà essere **fd** anziché 82 (swap) oppure 83 (linux native), in modo che il kernel, al bootstrap, riconosca le nuove partizioni come partizioni RAID.

Per il partizionamento si utilizzerà *fdisk /dev/hdc*, e si specificherà il comando *n* per aggiungere una nuova partizione, *t* per cambiarne il tipo, *p* per stampare la configurazione corrente.

2.3 Creazione del file /etc/raidtab

Si dovrà creare il file /etc/raidtab, inserendo tanti blocchi sottoriportati quante saranno le nostre partizioni raid.

```
#partizione /boot
raiddev /dev/md0
raid-level 1
nr-raid-disks 2
nr-spare-disks 0
chunk-size 64k
persistent-superblock 1
device /dev/hda1
failed-disk 0
device /dev/hdc1
raid-disk 1
```

Notare che abbiamo volutamente specificato che la partizione /dev/hda1 (attualmente utilizzata e montata come /boot) fa parte del raid, ma è inutilizzata: in questo modo sarà comunque possibile attivare il raid, che utilizzerà solo l'hard disk vergine (in questo caso la partizione /dev/hdc1).

2.4 Inizializzazione delle partizioni raid

Per ognuna delle partizioni raid che vorremo creare, bisognerà

- digitare *mkraid /dev/md0* (md1, md2, ...) per creare la partizione, che sarà visibile digitando *cat /proc/mdstat.*
- formattare le partizioni raid (in sostanza viene formattato solo l'hdd hdc perché hda è dichiarato failed, e questa opzione ci consente infatti di poter inizializzare le partizioni raid senza dover utilizzare un altro hdd supplementare) utilizzando il comando *mke2fs -j /dev/md0* (md1, md2, eccetera), nel caso si utilizzi il filesystem ext3.
- montare la partizione raid con il comando *mount /dev/md0 /mnt*
- copiare il filesystem dalle partizioni attualmente attive (hda1, ecc) alla nuova partizione raid, utilizzando il comando *tar cv /boot/ | tar x -C mnt/* in cui al posto di /boot bisognerà specificare la (o le) directory da inserire
  Un modo alternativo consiste nell'uso del comando *cp -axv /boot/* /mnt/*
  Nel caso si debba copiare tutto il contenuto del root filesystem (il quale sicuramente conterra' molte directory quali /bin /sbin /root eccetera, utilizzando il primo metodo basta digitare *cd /; for i in bin sbin root ... tmp; do tar cv $i |tar x -C mnt/; done* ; ricordarsi di creare le directory /mnt/* , /home, /var, /proc eccetera nel root filesystem!
- smontare la partizione raid, digitando *umount /mnt*

Nel caso di partizione swap, basterà applicare il punto 1 e poi digitare *mkswap /dev/md1* per la formattazione.

2.5 Test delle partizioni raid create

Il nuovo disco e' stato partizionato, e' stato creato il filesystem e vi sono stati inseriti tutti i file necessari. A questo punto e' possibile testare il funzionamento del raid per le partizioni diverse dalla root nel seguente modo:

- se ci si trova in modalita' multiutente, digitare *telinit 1* per cambiare runlevel andando ad operare in modalita' single user
- e' possibile a questo punto testare le partizioni var, usr, home, ... come ? Prendiamo ad esempio la partizione /var:

- modificare il file */etc/fstab* specificando che la partizione var, ad esempio, si trova in /dev/mdX anziche‘ in /dev/hdYz
- *umount /var*
- *mount var* ed a questo punto il raid corrispondente alla partizione var sara‘ disponibile, in modalita‘ degradata (perche‘ uno dei dischi e‘ dichiarato “failed-disk”.
- modificare /etc/raidtab dichiarando anche il primo disco come disco funzionante, quindi digitare *raidhotadd /dev/mdX /dev/hdYz* e monitorare il file /proc/mdstat per seguire come avviene la ricostruzione della partizione precedentemente marcata come danneggiata.

2.6 Modifica di fstab per montare le partizioni RAID

Modificare il file /etc/fstab specificando, in prima colonna, le nuove partizioni da utilizzare, ad esempio

```
/dev/md2 /       ext2  defaults  0 1
/dev/md0 /boot   ext2  defaults  0 2
/dev/md1 none    swap  sw        0 0
/dev/md3 /usr    ext2  defaults  0 2
```

eccetera.

A questo punto reboottare il PC per far caricare, al successivo riavvio, le nuove partizioni raid.

2.7 Attivazione del secondo hard disk di RAID

Riavviato il PC, digitando *mount* dovrete vedere che il sistema ora utilizza le nuove partizioni RAID: tuttavia, digitando *cat /proc/mdstat* , osserverete che solo un hard disk viene utilizzato per le partizioni di raid.

Affinché vengano utilizzati entrambi gli harddisk, bisognerà editare il file /etc/raidtab sostituendo tutte le linee **raid-failed 0** con **raid-disk 0** in modo da attivare anche il secondo hard disk (indispensabile per avere ridondanza).

Inoltre, dovranno essere modificate con *fdisk /dev/hda* il tipo di partizione, che dovrà essere impostata al valore **fd.**

2.8 Configurazione di lilo.conf

Bisognerà anche modificare /etc/lilo.conf onde specificare che la partizione di boot ora si trova in /dev/md0 (linea *boot=/dev/md0)* in modo che sia caricato il boot loader in entrambi gli hard disk. Alternativamente si potra’ specificare *boot=/dev/hdc* quindi digitare lilo per caricare il bootloader nell’hard disk hdc, poi modificare in *boot=/dev/hda* quindi ridigitare lilo affinche‘ sia caricato anche nel disco hda.

Inoltre si dovrà specificare che la partizione di root è /dev/md2, quindi digitare *lilo* per effettuare i cambiamenti.

Ricordatevi magari di aggiungere, se non presenti, le linee

password=PASSWORD
restricted

al fine di proteggere la macchina da eventuali malitenzionati che potrebbero al contrario guadagnarsi l’accesso di root da console, e quindi *chmod 600 /etc/lilo.conf.*

3 Test della nuova configurazione

Provate a riavviare, sperando che il tutto riparti.

Ricordate eventualmente che in fase di boot si possono specificare a lilo diversi parametri: ad esempio digitando LILO: *linux single* verrà avviato linux con runlevel 1, ovvero in modalità single user senza caricare tutti i servizi normalmente presenti.

Provate poi a staccare un harddisk, poi l’altro, e vedere se il sistema funziona.

4 Recupero di una partizione

Può accadere che, dopo mesi di utilizzo, al boot il sistema rilevi una partizione in RAID non utilizzabile: nel syslog troverete alcuni messaggi di errore del tipo

```
Dec  3 08:30:57 linux kernel: md: superblock update time inconsistency   -- using the mos
Dec  3 08:30:57 linux kernel: md: freshest
  ide/host0/bus1/target1/lun0/part5
Dec  3 08:30:57 linux kernel: md0: kicking faulty   ide/host0/bus0/target0/lun0/part5!
Dec  3 08:30:57 linux kernel: md:   unbind<ide/host0/bus0/target0/lun0/part5,1>
Dec  3 08:30:57 linux kernel: md:   export_rdev(ide/host0/bus0/target0/lun0/part5)
```

Si noti che è stato rilevato un problema nel device *ide/host0/bus0/target0/lun0/part5* ovvero /dev/hda5: infatti digitando *cat /proc/mdstat* viene segnalato l'utilizzo di un solo disco del RAID1:

```
md0 : active raid1 ide/host0/bus1/target1/lun0/part5[1]
      1799168 blocks [2/1] [_U]
```

A questo punto possiamo montare la partizione corrotta con il comando *mount /dev/hda5 /mnt/hd* dove hda5 è la partizione non funzionante e /mnt/hd e‘ una directory che ho creato nella quale montare il filesystem danneggiato.

Se il filesystem danneggiato è comunque "montabile", cancellare l'eventuale file .fsck* presente, quindi smontare il filesystem con *cd /; umount /mnt/hd* , e provare a reinserire la partizione nel raid con il comando *raidhotadd /dev/md0 /dev/hda5*.
A questo punto controllare con *cat /proc/mdstat* il risultato dell'operazione... se tutto va bene verrà ricostruita la partizione appena aggiunta e vedrete

```
md0 : active raid1 ide/host0/bus0/target0/lun0/part5[2] ide/host0/bus1/target1/lun0/part5
      1799168 blocks [2/1] [_U]
      [===========>.........]  recovery = 57.0% (1027456/1799168)
      finish=0.5min speed=21834K/sec
```

ed al termine dell'operazione, se conclusa con successo, vedrete

```
md0 : active raid1 ide/host0/bus0/target0/lun0/part5[0] ide/host0/bus1/target1/lun0/part5
      1799168 blocks [2/2] [UU]
```

In caso contrario suppongo dovrete riformattare la partizione (con *mke2fs /dev/hda5* o qualcosa del genere) oppure sostituire l'hard disk.

5 Monitoraggio stato del raid

Per verificare il corretto funzionamento del raid e' possibile schedulare nel crontab, ad esempio ogni 5 minuti, un complicatissimo script in grado di notificare in email l'amministratore riguardo eventuali guasti nei dischi.

Lo script che propongo e' il seguente:

```
#!/bin/bash
# /usr/local/sbin/raid-control.sh
# Controlla lo stato del RAID
RAIDSTRING=‘grep blocks /proc/mdstat |grep -v UU‘
if [ -z "$RAIDSTRING" ]; then
    cat /proc/mdstat |mail -s "raid problem at ‘hostname‘" psubiaco@creasol.it
fi
```

Per inserire lo script nel crontab basta digitare, da root, *crontab -e* quindi inserire la linea
*/5 * * * * /usr/local/sbin/raid-control.sh

6 Credits

Nel caso vogliate segnalare degli errori o suggerimenti, potete farlo all'indirizzo email psubiaco@creasol.it.
Per la stesura di questo documento ho fatto riferimento ad i seguenti testi:

- *Configuring Linux Software RAID 1* di Warwick Duncan warwick@clug.org.za e Craig Balfour craig@clug.org.za
- *The Software-RAID HOWTO* di Jakob OEstergaard jakob@ostenfeld.dk