RAID Logiciel

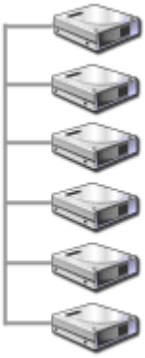
Ce tutoriel développe la mise en place des différents types de RAID Logiciel sous Linux avec Mdadm ou Raidtools.

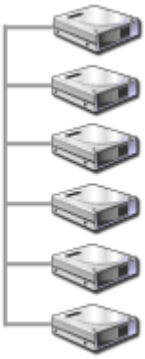
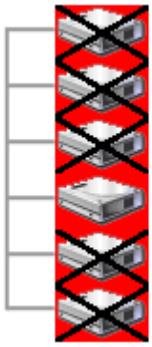
Présentation des types de RAID Logiciel

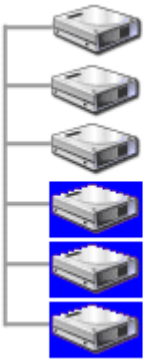
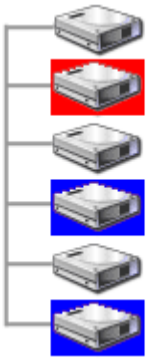
Nous allons parcourir les RAID suivants :

- Le mode Linéaire
- Le RAID 0
- Le RAID 1
- Le RAID 0+1
- Le RAID 10 (1+0)
- Le RAID 4
- Le RAID 5

Légende des représentations de Disques					
En Service	Mort	Données Perdues	Parité ou Redondance	Parité distribuée	Spare (rechange)
					

Le mode Linéaire (linear)	
Additionne les capacités de stockage des disques afin d'obtenir une seule entité de stockage. 2 Disques au minimum.	
✔ Utilisation des disques de taille différentes ou non. ✔ 100% de l'espace disque disponible. ✔ La perte d'un disque n'entraîne pas la perte **de toutes** les données.	✗ Pas d'amélioration des performances en lecture/écriture. ✗ En cas de perte d'un disque, les données présentes sur celui-ci sont perdues. ✗ Pas de disque de Spare.
	

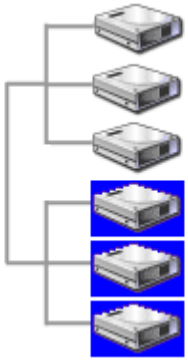
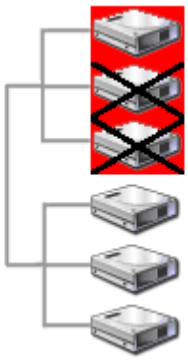
Le RAID 0 (striping)	
Découpe les données en blocs et les disperse linéairement. 2 Disques au minimum.	
✔ Utilisation de disques de taille différentes ou non. ✔ 100% de l'espace disque est disponible. ✔ Amélioration des performances en lecture/écriture.	✗ En cas de perte d'un disque, **toutes les données de l'ensemble des disques composant le RAID sont perdues.** ✗ Pas de disque de Spare.
	

Le RAID 1 (mirroring)	
Duplique les données des disques (redondance). 2 Disques au minimum.	
✔ La perte d'un disque n'entraîne pas la perte des données. ✔ Amélioration des performances en lecture. ✔ Pas de dégradation des performances en lecture lors de la reconstruction des données après la perte d'un disque. ✔ Possibilité d'utiliser un disque de Spare.	✗ Dégradation des performances en écriture. ✗ 50% de l'espace disque est disponible. ✗ Consomation en CPU.
	

Le RAID 0+1

Création de 2 stripes (RAID 0) indépendants de même capacité, associés en mirroring (RAID 1).

4 Disques au minimum.

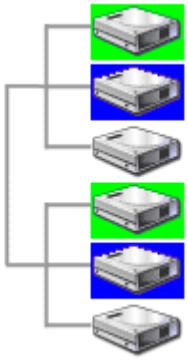
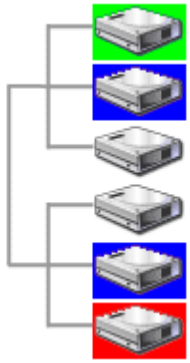
✔ La perte d'un disque n'entraîne pas la perte des données. ✔ Amélioration des performances en lecture (RAID 1) et en écriture (RAID 0). ✔ Pas de dégradation des performances en lecture lors de la reconstruction des données après la perte d'un disque.	✗ Immobilise plusieurs disques. ✗ 50% de l'espace disque est disponible. ✗ Consomation en CPU.
	

Le RAID 10 (1+0)

Création de 2 mirroring (RAID 1) indépendants de même capacité, associés en striping (RAID 0).

4 Disques au minimum.

Précision : Vous n'êtes pas obligé d'utiliser un disque Spare pour du RAID 10.

✔ La perte d'un disque n'entraîne pas la perte des données. ✔ Amélioration des performances en lecture (RAID 1) et en écriture (RAID 0).	✗ Immobilise plusieurs disques. ✗ de 50 à 33% de l'espace disque disponible. ✗ Consomation en CPU. ✗ Dégradation des performances en lecture lors de la reconstruction des données après la perte d'un disque.
	

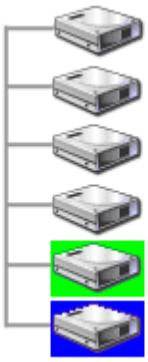
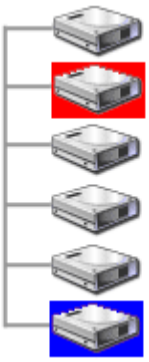
Le RAID 4

Immobilise 1 disque entier pour le stockage des informations de parité.

Équivaut à un RAID 0 mais avec la possibilité de reconstruire les données après la perte d'un disque.

3 Disques au minimum.

Précision : Vous n'êtes pas obligé d'utiliser un disque Spare pour du RAID 4.

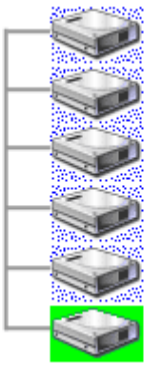
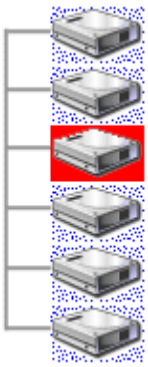
✔ La perte d'un disque n'entraîne pas la perte des données. ✔ Amélioration des performances en lecture.	✗ Dégradation des performances en écriture à cause du disque de parité. ✗ L'espace disque disponible est réduit à N-1. ✗ Dégradation des performances en lecture lors de la reconstruction des données après la perte d'un disque.
	

Le RAID 5

Ressemble à du RAID 4 mais en mieux, car il ne répartit pas les données de parité dans 1 seul disque mais dans chaque disques.

3 Disques au minimum mais 5 recommandés.

Précision : Vous n'êtes pas obligé d'utiliser un disque Spare pour du RAID 5.

✔ La perte d'un disque n'entraîne pas la perte des données. ✔ Amélioration des performances en lecture.	✗ Légère dégradation des performances en écriture à cause de la répartition de la parité. ✗ L'espace disque disponible est réduit à N-1 à cause de la répartition de la parité. ✗ Légère dégradation des performances en lecture lors de la reconstruction des données après la perte d'un disque.
	

Configuration de Mdadm

Installation du package :

Mandrake	debian
mdadm-1.9.0-3mdk.i586.rpm	mdadm_1.9.0-4_i386.deb

MDadm utilise un fichier de configuration pour activer les périphériques RAID au boot de votre machine. Toutes les commandes se font avec le binaire **/sbin/mdadm**

Mise en place d'un RAID 5 avec 1 disque de spare

Pour créer et activer la 1er fois son lecteur logique /dev/md0 contenant les 6 disques SCSI /dev/sdb / dev/sdc /dev/sdd /dev/sde /dev/sdf /dev/sdg :

```
[root@pc user]# mdadm --create /dev/md0 --level=5 --raid-devices=5 --
spare-devices=1 /dev/sd[bcdefg]
 mdadm: array /dev/md0 started.
```

Pour visualiser tous les MD (multi disks) actifs :

```
[root@pc user]# mdadm --detail --scan
 ARRAY /dev/md0 level=raid5 num-devices=5
UUID=9f37ae88:a3b50329:ce24f75c:b18f1a2a
   devices=/dev/scsi/host0/bus0/target1/lun0/disc,/dev/scsi/host0/
bus0/target2/lun0/disc,/dev/scsi/host0/bus0/target3/lun0/disc,/dev/
scsi/host0/bus0/target4/lun0/disc,/dev/scsi/host0/bus0/target5/lun0/
disc,/dev/scsi/host0/bus0/target6/lun0/disc
```

Nous ne voyons que `/dev/md0` pour le moment

Pour visualiser en détail notre /dev/md0 :

```
[root@pc user]# mdadm --detail /dev/md0
/dev/md0:
        Version : 00.90.00
  Creation Time : Mon Mar 15 21:28:08 2004
     Raid Level : raid5
     Array Size : 16776960 (15.100 GiB 17.18 GB)
    Device Size : 4194240 (3.100 GiB 4.29 GB)
   Raid Devices : 5
  Total Devices : 7
Preferred Minor : 0
    Persistence : Superblock is persistent

    Update Time : Mon Mar 15 21:33:14 2004
          State : dirty, no-errors
 Active Devices : 5
Working Devices : 6
 Failed Devices : 1
  Spare Devices : 1

         Layout : left-symmetric
     Chunk Size : 64K

     Number   Major   Minor   RaidDevice State
       0       8       16       0       active sync   /
```

```
dev/scsi/host0/bus0/target1/lun0/disc
       1       8       32       1       active sync   /
dev/scsi/host0/bus0/target2/lun0/disc
       2       8       48       2       active sync   /
dev/scsi/host0/bus0/target3/lun0/disc
       3       8       64       3       active sync   /
dev/scsi/host0/bus0/target4/lun0/disc
       4       8       80       4       active sync   /
dev/scsi/host0/bus0/target5/lun0/disc
       5       8       96       5       spare   /
dev/scsi/host0/bus0/target6/lun0/disc
           UUID : 9f37ae88:a3b50329:ce24f75c:b18f1a2a
         Events : 0.4
```

Pour arrêter le lecteur logique /dev/md0 :

```
[root@pc user]# mdadm --stop /dev/md0
```

Pour activer le lecteur logique /dev/md0 :

```
[root@pc user]# mdadm --assemble /dev/md0 /dev/sd[bcdefg]
 mdadm: /dev/md0 has been started with 5 drives and 1 spare.
```

Pour formater le lecteur logique /dev/md0 en ext3 par exemple :

```
[root@pc user]# mke2fs -j /dev/md0
mke2fs 1.32 (09-Nov-2002)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
2097152 inodes, 4194240 blocks
209712 blocks (5.00%) reserved for the super user
First data block=0
128 block groups
32768 blocks per group, 32768 fragments per group
16384 inodes per group
Superblock backups stored on blocks:
       32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632,
2654208, 4096000

Writing inode tables: done
Creating journal (8192 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 38 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.
```

Il ne reste plus qu'a monter le lecteur logique /dev/md0 dans /mnt/stockage :

```
[root@pc user]# mkdir /mnt/stockage
[root@pc user]# mount /dev/md0 /mnt/stockage
[root@pc user]# mount
...
/dev/md0 on /mnt/stockage type ext3 (rw)
```

Pour surveiller le raid si un disque fail, un mail est envoyé au root :

```
[root@pc user]# mdadm --monitor --mail=root@alex.fr --delay=120 --
daemonise /dev/md0
```

Pour mettre en panne le disque /dev/sdd par exemple :

```
[root@pc user]# mdadm /dev/md0 --fail /dev/sdd
mdadm: set /dev/sdd faulty in /dev/md0
```

Pour supprimer un disque en panne :

```
[root@pc user]# mdadm /dev/md0 --remove /dev/sdd
mdadm: hot removed /dev/sdd
```

Pour ajouter un disque de spare en RAID 1,4,5 :

```
[root@pc user]# mdadm /dev/md0 --add /dev/sdh
mdadm: hot added /dev/sdh
```

Mise en place d'un RAID 1 + 0 avec 1 disque de spare

Consiste à créer 2 lecteurs logique /dev/md2 et /dev/md3 en RAID 1.
Puis assembler ces 2 lecteurs logique en RAID 0 sur /dev/md4

Pour créer et activer le lecteur logique /dev/md2 contenant les 3 disques /dev/sdb /dev/sdc / dev/sdd :

```
[root@pc user]# mdadm --create /dev/md2 -l 1 -n 2 -x 1 /dev/sd[bcd]
 mdadm: array /dev/md2 started.
```

Vérification de /dev/md2

```
[root@pc user]# mdadm --detail /dev/md2
/dev/md2:
        Version : 00.90.00
  Creation Time : Wed Mar 17 21:32:54 2004
     Raid Level : raid1
     Array Size : 4194240 (3.100 GiB 4.29 GB)
    Device Size : 4194240 (3.100 GiB 4.29 GB)
   Raid Devices : 2
  Total Devices : 3
Preferred Minor : 2
    Persistence : Superblock is persistent

    Update Time : Wed Mar 17 21:32:54 2004
          State : dirty, no-errors
 Active Devices : 2
Working Devices : 3
 Failed Devices : 0
  Spare Devices : 1


    Number   Major   Minor   RaidDevice State
       0       8       16        0      active sync   /
dev/scsi/host0/bus0/target1/lun0/disc
       1       8       32        1      active sync   /
dev/scsi/host0/bus0/target2/lun0/disc
       2       8       48        2      spare   /
dev/scsi/host0/bus0/target3/lun0/disc
           UUID : f9234199:ca00d95a:8d3e868b:e4be3b88
         Events : 0.1
```

Pour créer et activer le lecteur logique /dev/md3 contenant les 3 disques /dev/sde /dev/sdf /dev/sdg :

```
[root@pc user]# mdadm --create /dev/md3 -l 1 -n 2 -x 1 /dev/sd[efg]
 mdadm: array /dev/md3 started.
```

Vérification de /dev/md3

```
[root@pc user]# mdadm --detail /dev/md3
/dev/md3:
        Version : 00.90.00
  Creation Time : Wed Mar 17 21:33:04 2004
     Raid Level : raid1
     Array Size : 4194240 (3.100 GiB 4.29 GB)
    Device Size : 4194240 (3.100 GiB 4.29 GB)
   Raid Devices : 2
  Total Devices : 3
Preferred Minor : 3
    Persistence : Superblock is persistent
```

```
    Update Time : Wed Mar 17 21:33:04 2004
          State : dirty, no-errors
 Active Devices : 2
Working Devices : 3
 Failed Devices : 0
  Spare Devices : 1


    Number   Major   Minor   RaidDevice State
       0       8       64       0       active sync   /
dev/scsi/host0/bus0/target4/lun0/disc
       1       8       80       1       active sync   /
dev/scsi/host0/bus0/target5/lun0/disc
       2       8       96       2       spare   /
dev/scsi/host0/bus0/target6/lun0/disc
           UUID : be07cc8d:4e8b5aaa:43e96495:5d45d401
         Events : 0.1
```

Pour créer et activer le lecteur logique /dev/md4 contenant les 2 lecteurs logique /dev/md2 et /dev/md3 :

```
[root@pc user]# mdadm --create /dev/md4 -l 0 -n 2 /dev/md[23]
 mdadm: array /dev/md4 started.
```

Vérification de /dev/md4

```
[root@pc user]# mdadm --detail /dev/md4
/dev/md4:
        Version : 00.90.00
  Creation Time : Wed Mar 17 21:29:13 2004
     Raid Level : raid0
     Array Size : 8388352 (7.100 GiB 8.59 GB)
   Raid Devices : 2
  Total Devices : 2
Preferred Minor : 4
    Persistence : Superblock is persistent

    Update Time : Wed Mar 17 21:29:13 2004
          State : dirty, no-errors
 Active Devices : 2
Working Devices : 2
 Failed Devices : 0
  Spare Devices : 0

     Chunk Size : 64K

    Number   Major   Minor   RaidDevice State
       0       9       2       0       active sync   /dev/md/2
       1       9       3       1       active sync   /dev/md/3
           UUID : 6827a760:8cc69ddd:8531f69d:5ba54c7e
         Events : 0.1
```

Pour activer les périphériques RAID au boot de votre machine

Quand votre RAID est actif, passez la commande suivante :

Mandrake	debian

`[root@pc user]# echo "DEVICE partitions" >> /etc/mdadm.conf`
 	
`[root@pc user]# mdadm --detail --scan >> /etc/mdadm.conf`
	`pc:~# mdadm --detail --scan >> /etc/mdadm/mdadm.conf`

Aperçu du fichier mdadm.conf pour un RAID 1 :

```
DEVICE partitions
ARRAY /dev/md1 level=raid1 num-devices=2
UUID=4c45dba8:3e094c34:5d50ebcd:60d50d67
   devices=/dev/hda8,/dev/hda9
```

Configuration de Raidtools

Mandrake

Installation du package **raidtools-0.90-13.rpm**

Le développement du logiciel Raidtools étant stoppé, vous trouverez des packages pour Mandrake. Pour Debian Sarge il faut utiliser mdadm.

Raidtools utilise le fichier de configuration **/etc/raidtab** pour fonctionner.

Mise en place d'un RAID 5 avec 1 disque de spare

Créez le fichier /etc/raidtab :

```
# Exemple RAID 5 + 1 Spare
#
# Nom du device utilisé
raiddev             /dev/md0

# Niveau du RAID (linear|raid0|raid1|raid4|raid5|...)
raid-level          5

# Nombre de disques de données composant ce RAID
nr-raid-disks       5

# Nombre de disques Spare (remplacement) composant ce RAID
nr-spare-disks      1

# Indique la taille des segments virtuels des données
# sur le device. Doit être une puissance de 2 et faire au
```

```
# minimum 4ko. Ne pas utiliser en mode linear.
chunk-size          64

# Active l'auto détection de la configuration au démarrage du système.
persistent-superblock   1

# Défini le type d'algorithme pour la répartition des données
# de parités. Valable en RAID 4 et 5 uniquement.
parity-algorithm    left-symmetric

# Liste des devices (disques ou partitions)
device          /dev/sdb
raid-disk       0

device          /dev/sdc
raid-disk       1

device          /dev/sdd
raid-disk       2

device          /dev/sde
raid-disk       3

device          /dev/sdf
raid-disk       4

# Disque de Spare
device          /dev/sdg
spare-disk      0
```

Pour créer et activer le lecteur logique /dev/md0 contenant les 6 disques SCSI /dev/sdb /dev/sdc / dev/sdd /dev/sde /dev/sdf /dev/sdg :

```
[root@pc user]# mkraid /dev/md0
```

ou si vous voulez détruire l'ancienne configuration

```
[root@pc user]# mkraid -R /dev/md0
DESTROYING the contents of /dev/md0 in 5 seconds, Ctrl-C if unsure!
handling MD device /dev/md0
analyzing super-block
disk 0: /dev/sdb, 4194304kB, raid superblock at 4194240kB
disk 1: /dev/sdc, 4194304kB, raid superblock at 4194240kB
disk 2: /dev/sdd, 4194304kB, raid superblock at 4194240kB
disk 3: /dev/sde, 4194304kB, raid superblock at 4194240kB
disk 4: /dev/sdf, 4194304kB, raid superblock at 4194240kB
disk 5: /dev/sdg, 4194304kB, raid superblock at 4194240kB
```

Pour formater le lecteur logique /dev/md0 en ext3 par exemple :

```
[root@pc user]# mke2fs -j /dev/md0
mke2fs 1.32 (09-Nov-2002)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
2097152 inodes, 4194240 blocks
209712 blocks (5.00%) reserved for the super user
First data block=0
```

```
128 block groups
32768 blocks per group, 32768 fragments per group
16384 inodes per group
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632,
2654208, 4096000

Writing inode tables: done
Creating journal (8192 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.
```

Monter le lecteur logique /dev/md0 dans /mnt/stockage :

```
[root@pc user]# mount /dev/md0 /mnt/stockage/
[root@pc user]# mount
...
/dev/md0 on /mnt/stockage type ext3 (rw)
```

Démonter le lecteur logique /dev/md0 pour l'arrêter :

```
[root@pc user]# umount /dev/md0 /mnt/stockage/
[root@pc user]# raidstop /dev/md0
```

Redémarrer le lecteur logique /dev/md0 :

```
[root@pc user]# raidstart /dev/md0
```

Pour mettre en panne le disque /dev/sdc par exemple :

```
[root@pc user]# raidsetfaulty /dev/md0 /dev/sdc
```

Pour supprimer un disque en panne :

```
[root@pc user]# raidhotremove /dev/md0 /dev/sdc
```

Pour ajouter un disque de spare en RAID 1,4,5 :

```
[root@pc user]# raidhotadd /dev/md0 /dev/sdh
```

Pour visualiser les informations sur /dev/md0 :

```
[root@pc user]# cat /proc/mdstat
Personalities : [raid5]
read_ahead 1024 sectors
md0 : active raid5 scsi/host0/bus0/target2/lun0/disc[6]
scsi/host0/bus0/target6/lun0/disc[1] scsi/host0/bus0/target5/lun0/disc[4]
scsi/host0/bus0/target4/lun0/disc[3] scsi/host0/bus0/target3/lun0/disc[2]
scsi/host0/bus0/target1/lun0/disc[0]
      16776960 blocks level 5, 64k chunk, algorithm 2 [5/5] [UUUUU]
unused devices: <none>
```

Mise en place d'un RAID 1 + 0 avec 1 disque de spare

Créez un nouveau fichier /etc/raidtab :

```
# Exemple RAID 1 + 0 avec Spare
#
# RAID 1
# Nom du device utilisé pour le 1er mirror
raiddev               /dev/md0
raid-level            1
# 2 disques de données
nr-raid-disks         2
# 1 disque de Spare
nr-spare-disks        1
chunk-size            16
persistent-superblock 1

# Liste des devices pour le 1er mirror
device                /dev/sdb
raid-disk             0
device                /dev/sdc
raid-disk             1
# Le disque de Spare du 1er mirror
device                /dev/sdd
spare-disk            0

#
# Nom du device utilisé pour le 2em mirror
#
raiddev               /dev/md1
raid-level            1
# 2 disques de données
nr-raid-disks         2
# 1 disque de Spare
nr-spare-disks        1
chunk-size            16
persistent-superblock 1

# Liste des devices pour le 2em mirror
device                /dev/sde
raid-disk             0
device                /dev/sdf
raid-disk             1
# Le disque de Spare du 2em mirror
device                /dev/sdg
spare-disk            0

#
# RAID 0
# Nom du device utilisé pour le striping
#
raiddev               /dev/md2
raid-level            0
# Stripe de nos 2 devices RAID 1
nr-raid-disks         2
nr-spare-disks        0
persistent-superblock 1
```

```
chunk-size              16

# Liste des 2 devices RAID 1 précédantes
device                  /dev/md0
raid-disk               0
device                  /dev/md1
raid-disk               1
```

Consiste à créer 2 lecteurs logique /dev/md0 et /dev/md1 en RAID 1.
Puis assembler ces 2 lecteurs logique en RAID 0 sur /dev/md2

Pour créer et activer le lecteur logique /dev/md0 contenant les 3 disques /dev/sdb /dev/sdc / dev/sdd :

```
[root@pc user]# mkraid -R /dev/md0
DESTROYING the contents of /dev/md0 in 5 seconds, Ctrl-C if unsure!
handling MD device /dev/md0
analyzing super-block
disk 0: /dev/sdb, 4194304kB, raid superblock at 4194240kB
disk 1: /dev/sdc, 4194304kB, raid superblock at 4194240kB
disk 2: /dev/sdd, 4194304kB, raid superblock at 4194240kB
```

Pour créer et activer le lecteur logique /dev/md1 contenant les 3 disques /dev/sde /dev/sdf /dev/sdg :

```
[root@pc user]# mkraid -R /dev/md1
DESTROYING the contents of /dev/md1 in 5 seconds, Ctrl-C if unsure!
handling MD device /dev/md1
analyzing super-block
disk 0: /dev/sde, 4194304kB, raid superblock at 4194240kB
disk 1: /dev/sdf, 4194304kB, raid superblock at 4194240kB
disk 2: /dev/sdg, 4194304kB, raid superblock at 4194240kB
```

Pour créer et activer le lecteur logique /dev/md2 contenant les 2 lecteurs logique /dev/md0 et / dev/md1 :

```
[root@pc user]# mkraid -R /dev/md2
DESTROYING the contents of /dev/md2 in 5 seconds, Ctrl-C if unsure!
handling MD device /dev/md2
analyzing super-block
disk 0: /dev/md0, 4194240kB, raid superblock at 4194176kB
disk 1: /dev/md1, 4194240kB, raid superblock at 4194176kB
```

Pour formater en ext3 le lecteur logique /dev/md2 qui est constitué des 2 RAID 1 :

```
[root@pc user]# mke2fs -j /dev/md2
mke2fs 1.32 (09-Nov-2002)
Filesystem label=
OS type: Linux
Block size=4096 (log=2)
Fragment size=4096 (log=2)
1048576 inodes, 2097088 blocks
104854 blocks (5.00%) reserved for the super user
First data block=0
64 block groups
32768 blocks per group, 32768 fragments per group
16384 inodes per group
Superblock backups stored on blocks:
       32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632

Writing inode tables: done
Creating journal (8192 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.
```

Sources :
http://linux.cudeso.be/misc.php
http://www.rezalfr.org/

Document mis à jour : 10/10/05