

Atelier: Le RAID logiciel

A. Pré-requis

La commande **mdadm** gère un disque RAID (création, activation...)

Ses principales sous-commandes:

- create: Crée un disque RAID, met à jour les superblocs des disques spécifiés.
- assemble: Active un disque RAID déjà créé précédemment (on ne met pas à jour les superblocs)
- build: Crée un disque RAID sans superbloc (pour RAID-0 ou Linear)
- examine: Affiche le superbloc d'un disque
- add: Ajoute (à chaud) un disque RAID.
- remove: Retire (à chaud) un disque RAID.
- fail: Positionne un disque en erreur (simule le crash d'un disque).
- query: Examine un disque et indique s'il est membre d'un RAID.
- monitor: Surveille les disques, prévient l'administrateur si besoin (par e-mail...).
Activé par le RC mdmonitor.

Les fichiers:

- /proc/mdstat:** Etat du RAID (publié par le noyau).
- /etc/mdadm.conf:** Ce fichier n'est pas nécessaire mais il permet de garder une trace d'un disque RAID. L'activation d'un disque RAID est simplifiée si on utilise ce fichier.

Le superbloc:

Les disques faisant partie d'un RAID (sauf RAID-0 ou Linear) possèdent en tête un superbloc qui contient la configuration du RAID. Ces superblocs permettent également le démarrage à partir d'un disque RAID.

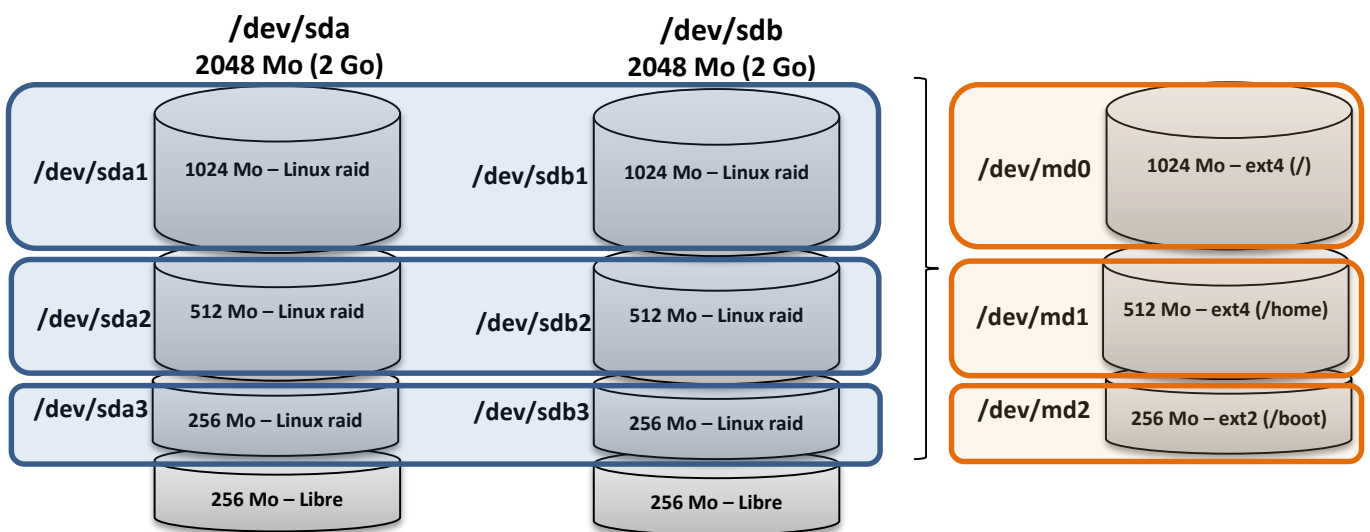
B. Installation et manipulation d'un système RAID1

Créez une machine virtuelle (VirtualBox) CentOS 6.x (64 bits) possédant les caractéristiques suivantes:

- 1 Go Ram, 4 processeurs
- Pas de carte son, pas d'usb
- Une carte réseau
- 2 disques SATA de 2048 Mo (2 Go) en taille fixe
- Possibilité de démarrer cette machine sur une image d'installation (.iso) CentOS 6.x minimale

Une fois configurée, le système devra pouvoir démarrer sur son système RAID opérationnel ou sur son RAID dégradé (bootage possible même si un des 2 disques est en panne).

Les disques durs respecteront le partitionnement et les RAID 1 suivants:



Il faudra d'abord créer les partitions RAID software puis créer leurs devices ...

1. Configuration sous VirtualBox

Dans VirtualBox:

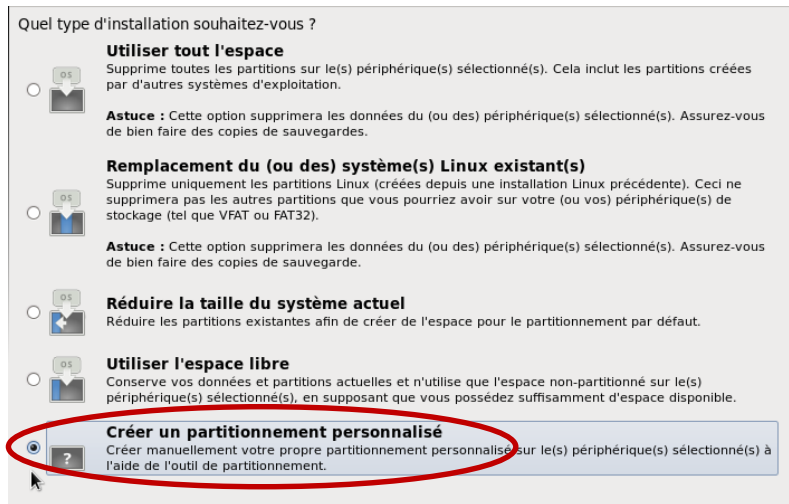
- Système → Processeur → 4
 - Carte mère → Mémoire vive → 1024 Mo
- Son → [] Activer le son
- USB → [] Activer le contrôleur USB
- Réseau → Carte1
 - [v] Activer la carte réseau
 - Avancé: Type de carte: Intel Pro/1000 MT Desktop (82540OEM)
- Stockage → Contrôleur SATA
 - HD1r1 (2 Go)
 - HD2r1 (2 Go)
 - Contrôleur IDE
 - CentOS-6.4...minimal.iso

2. Installation

a. Effectuez les 1^{ère} étapes de l'installation comme d'habitude...

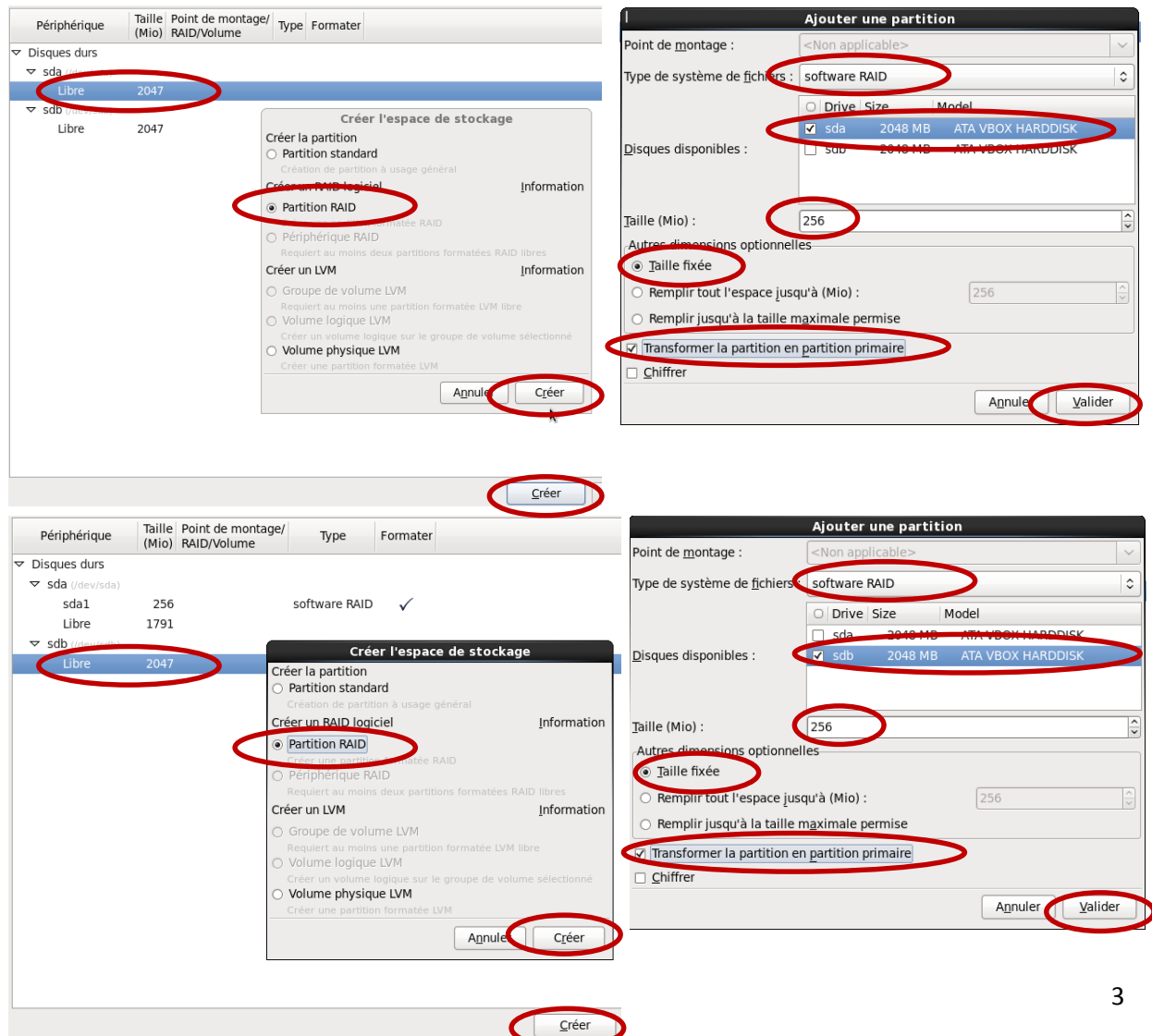
...

b. Respectez le partitionnement suivant



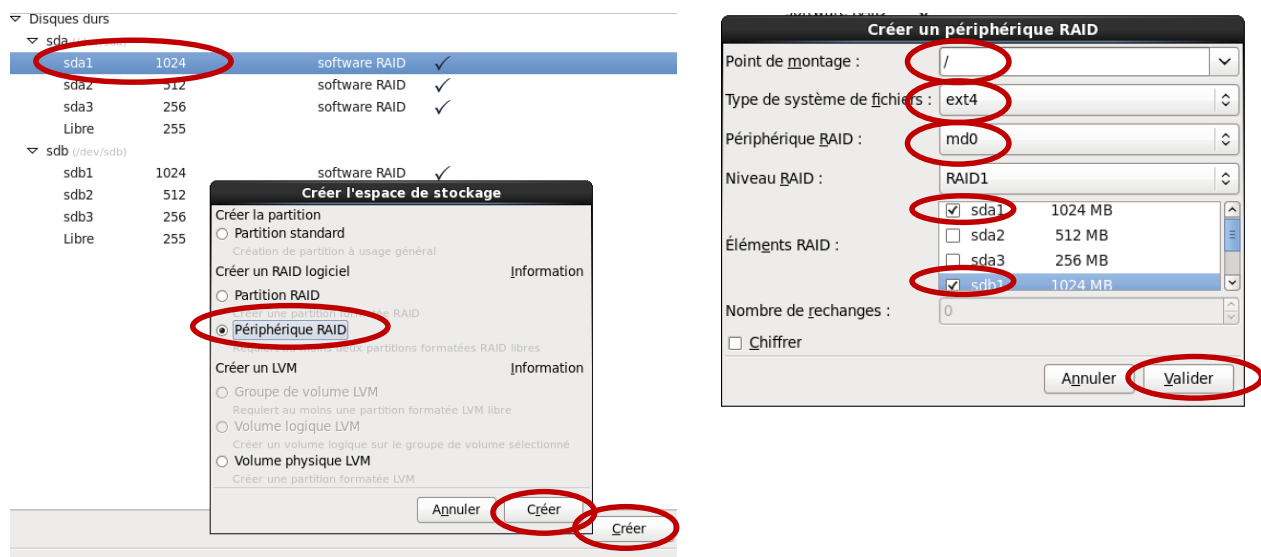
i. Création des partitions RAID software

Pour la partition RAID software qui devra accueillir /boot



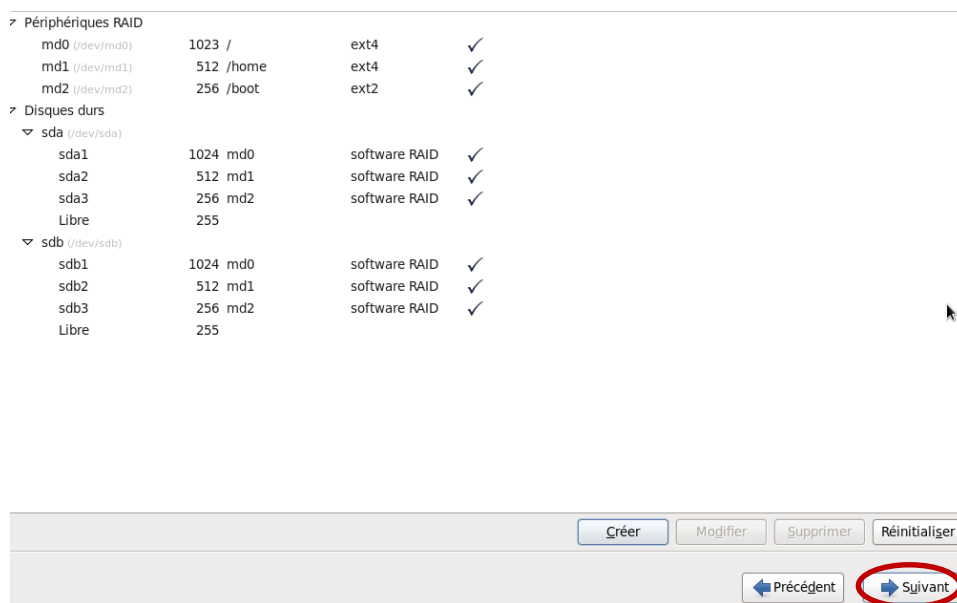
Procédez de façon similaire pour les partitions RAID software qui devront accueillir / et /home...

ii. Création des devices RAID à partir de ces partitions



Procédez de façon similaire pour les devices RAID qui gèreront /home et /boot...

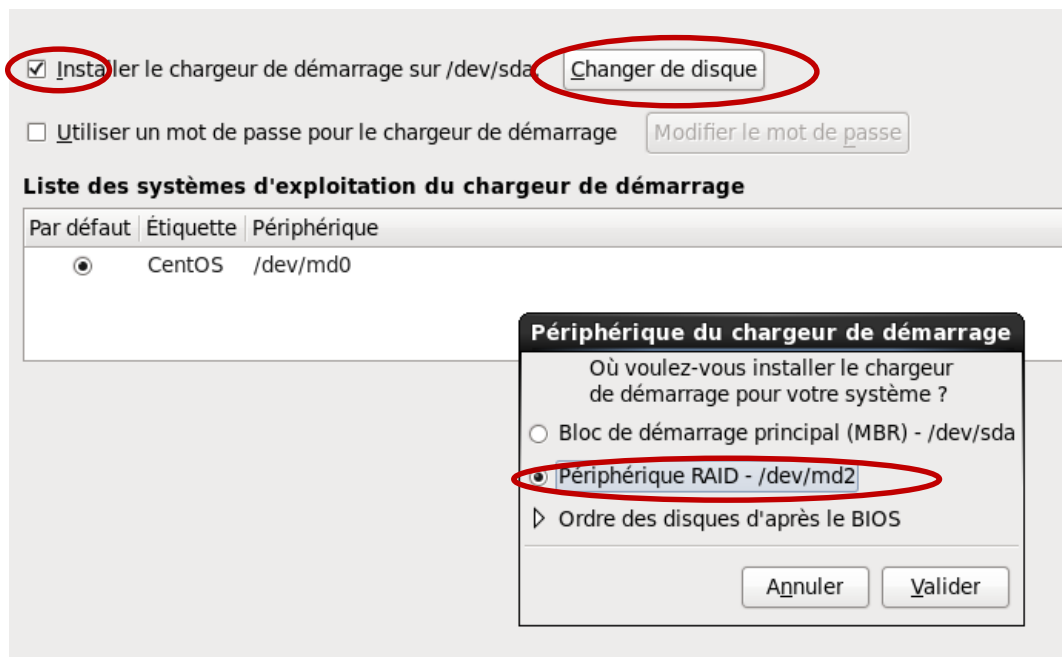
Fenêtre de fin de configuration du RAID



Une fenêtre signale qu'on n'a pas de swap... On continue quand même.

On "formate" et on "écrit les modifications sur disque"

On installe Grub sur le RAID pour pouvoir démarrer en RAID dégradé.



On installe Grub sur le RAID pour pouvoir démarrer en RAID dégradé.

Puis on clique "Suivant"

L'installation des paquets commence...

En fin d'installation, on redémarre...

Il y a beaucoup de chance que le système freeze au redémarrage.

Cela est dû au fait que l'installation correcte de Grub sur un périphérique RAID avec l'assistant d'installation est "boggée".

Comment résoudre ce problème ?

1. Refaire une installation complète (⊗) mais "Installer le chargeur de démarrage sur /dev/sda"
2. Redémarrer avec l'image .iso en mode "rescue" et réaliser les actions suivantes:

```
# chroot /mnt/sysimage
# grub → On réinstalle l'appel à Grub correctement sur
... les MBR de sda et sdb et Grub dans /boot de sda et sdb
grub> device (hd0) /dev/sda → Association des noms Grub aux noms Linux
grub> device (hd1) /dev/sdb

grub> root (hd0,2) → Racine sur laquelle seront lus les fichiers de Grub
root (hd0,2)

grub> setup (hd0) → Restauration de Grub dans le MBR de hd0
setup (hd0) (hd0,2) → restauration dans le boot secteur de
/boot)
Checking if "/boot/grub/stage1" exist... no
Checking if "/boot/stage1" exist... yes
Checking if "/boot/stage2" exist... yes
... succeeded
Done

grub> root (hd1,2)
root (hd1,2)
Filesystem type is ext2fs, partition type 0x fd
```

```
grub> setup (hd1)
setup (hd1)
Checking if "/boot/grub/stage1" exist... no
Checking if "/boot/stage1" exist... yes
Checking if "/boot/stage2" exist... yes
... succeeded
Done

grub> quit
# exit
+ Désactivation de SELINUX
# halt
```

On éjecte l'image .iso du CD virtuel et on démarre la MV...

3. Vérifiez que le monitoring RAID tourne

```
#cat /proc/mdstat
```

4. Vérifiez que les types de RAID logiciels sont acceptés par le noyau

```
[root@raid ~]# cat /proc/mdstat
Personalities : [raid1]
md2 : active raid1 sda3[0] sdb3[1]
      262080 blocks super 1.0 [2/2] [UU]

md1 : active raid1 sda2[0] sdb2[1]
      523968 blocks super 1.1 [2/2] [UU]

md0 : active raid1 sda1[0] sdb1[1]
      1048000 blocks super 1.1 [2/2] [UU]
      bitmap: 0/1 pages [0KB], 65536KB chunk
```

5. Lister les matrices RAID

```
#mdadm --detail --scan
```

6. Afficher des informations sur un disque RAID (composition, niveau ...)

```
#mdadm --detail /dev/md0 | more
```

7. Afficher des informations sur un disque RAID (fait-il partie d'un RAID...)

```
#lsblk
```

8. Avoir un aperçu rapide de l'état de nos ensembles RAID.

```
#cat /proc/mdstat
```

9. Utiliser les grappes RAIDs

- a. Utiliser la grappe RAID1:0 (md0)

- b. Utiliser la grappe RAID1:1 (md1)

- c. Utiliser la grappe RAID1:2 (md2)

```
[root@raid ~]# lsblk
NAME        MAJ:MIN RM   SIZE RO TYPE MOUNTPOINT
sr0          11:0    1  1024M  0 rom
sda           8:0     0    2G  0 disk
├─sda1        8:1     0     1G  0 part
│ └─md0       9:0     0 1023,4M  0 raid1 /
├─sda2        8:2     0    512M  0 part
│ └─md1       9:1     0  511,7M  0 raid1 /home
├─sda3        8:3     0    256M  0 part
│ └─md2       9:2     0    256M  0 raid1 /boot
└─sdb         8:16    0     2G  0 disk
   ├─sdb1      8:17    0     1G  0 part
   │ └─md0       9:0     0 1023,4M  0 raid1 /
   ├─sdb2      8:18    0    512M  0 part
   │ └─md1       9:1     0  511,7M  0 raid1 /home
   └─sdb3      8:19    0    256M  0 part
      └─md2       9:2     0    256M  0 raid1 /boot
```

10. On teste si le système démarre avec seulement un disque (démarrage RAID dégradé)

- a. Redémarrage sur le 1^{er} disque uniquement

On arrête le système et on détache le 2^{ème} disque dans VirtualBox

On redémarre.

Ca marche (excepté un certain temps de latence au 1^{er} boot... et un message "device node not found")

Avoir un aperçu rapide de l'état de nos ensembles RAID.

```
#cat /proc/mdstat
```

```
[root@raid ~]# cat /proc/mdstat
Personalities : [raid1]
md2 : active raid1 sda3[0]
      262080 blocks super 1.0 [2/1] [U_]

md1 : active raid1 sda2[0]
      523968 blocks super 1.1 [2/1] [U_]

md0 : active raid1 sda1[0]
      1048000 blocks super 1.1 [2/1] [U_]
      bitmap: 1/1 pages [4KB], 65536KB chunk

unused devices: <none>
```

- b. Redémarrage sur le 2^{ème} disque uniquement

On arrête le système et on rattache le 2^{ème} disque et on détache le 1^{er} dans VirtualBox

On redémarre.

Ca marche (excepté un certain temps de latence... et un message "device node not found")

Avoir un aperçu rapide de l'état de nos ensembles RAID.

```
#cat /proc/mdstat
```

11. Remplacement du disque défectueux et reconstruction du RAID

- a. On affiche le partitionnement du disque opérationnel

```
#fdisk -l | grep sda
```

- b. On se procure un 2^{ème} disque identique

On arrête le système et on crée un nouveau disque dans VirtualBox de 2048 Mio.

On redémarre le système.

- c. On partitionne ce disque à l'identique

```
#fdisk /dev/sdb
```

Un partitionnement à l'identique sera difficile vu que les partitions du 1^{er} ne se terminent pas sur une frontière de cylindre.

```
>n      >p      >1      >1      >131
>n      >p      >2      >132     >197
>n      >p      >3      >198     >231     >w
```

- d. On lance un affichage continu de l'état des grappes sur un 2^{ème} terminal.

```
#watch cat /proc/mdstat
```

- e. On reconstitue les grappes RAID 1 et on constate sur le 2^{ème} terminal

```
#mdadm --manage /dev/md0 --add /dev/sdb1
```

```
#mdadm --manage /dev/md1 --add /dev/sdb2
```

```
#mdadm --manage /dev/md2 --add /dev/sdb3
```

- f. En fin de reconstitution, on vérifie l'état de nos ensembles RAID.

```
#cat /proc/mdstat
```

- g. On resynchronise toutes les partitions sur leurs grappes respectives

```
#mdadm --assemble --scan
```

```
#mdadm --detail --scan --verbose > /etc/mdadm.conf
```

- h. On vérifie si nos manipulations de l'étape 9 sont toujours présentes (et donc reconstituées)

12. On n'oublie pas le chargeur de démarrage pour rendre à nouveau le système bootable en mode dégradé

```
#grub
```

```
>device (hd0) /dev/sda
```

```
>device (hd1) /dev/sdb
```

```
>root (hd0,2)
```

```
>setup (hd0)
```

```
root (hd1,2)
```

```
>setup(hd1)
```

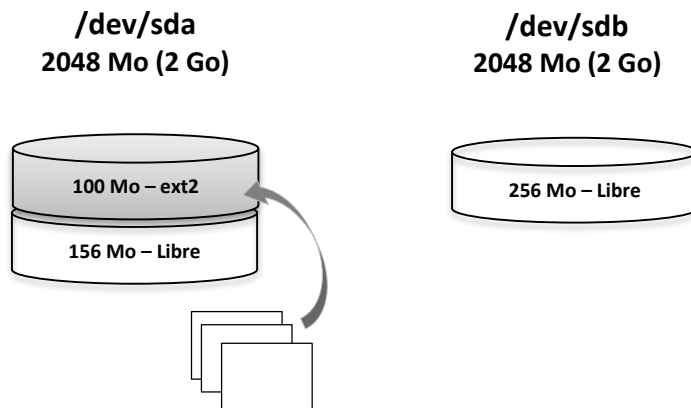
```
>quit
```

```
#reboot
```

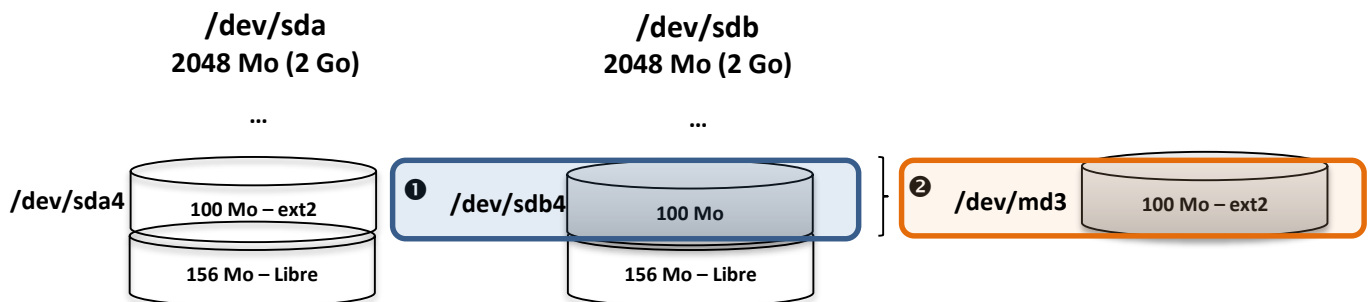
C. Transformation d'un disque en disque RAID1 logiciel sans perte de données

Principe:

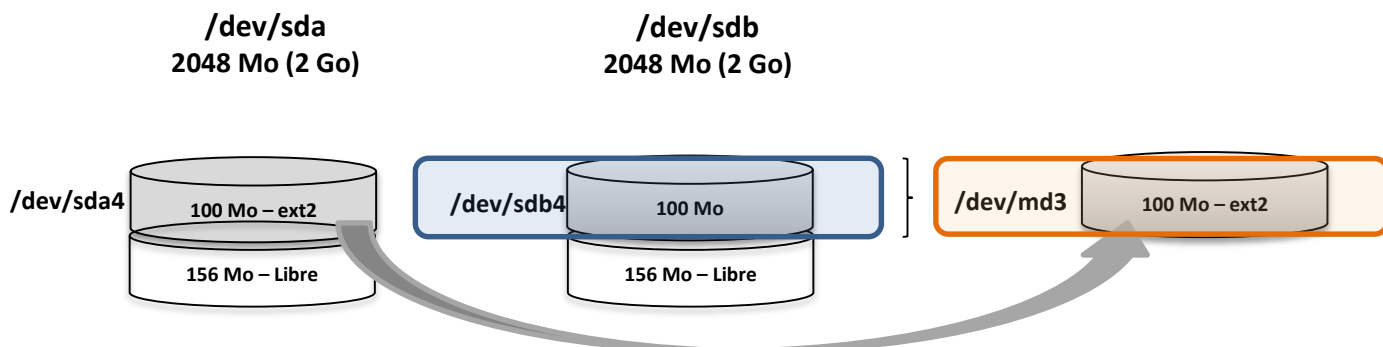
- Créer une partition standard Linux de 100 Mo sur `/dev/sda4`, la formater et y installer des fichiers.



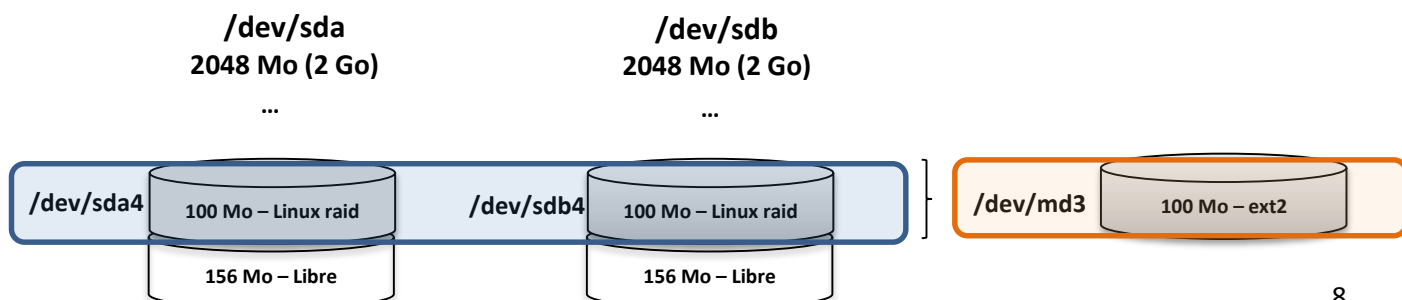
- Créer une partition standard Linux de 100 Mo sur `/dev/sdb4`, l'intégrer dans un disque RAID1 et formater le nouveau RAID1 en ext2.



- Copier les données du disque à encapsuler vers la grappe RAID1



- Ajouter le disque d'origine au RAID1



Mise en oeuvre:

1. Créer une partition standard Linux de 100 Mo sur /dev/sda4, la formater et y installer des fichiers.

```
#fdisk /dev/sda >n >p >(enter) >+100M >w #reboot
```

Attention: L'idéal est de commencer cette partition au début d'un nouveau cylindre.

```
#mkfs -t ext2 -q /dev/sda4 #mount -t ext2 /dev/sda4 /mnt #cp -r /usr/share/doc/* /mnt
```

2. Créer une partition standard Linux de 100 Mo sur /dev/sdb4, l'intégrer dans un disque RAID1 et formater le nouveau RAID1 en ext2.

Attention: La taille de cette partition doit impérativement être supérieure ou égale à celle de sda4.

L'idéal est de commencer cette partition au début d'un nouveau cylindre.

- a. Création de la partition /dev/sdb4 de taille strictement identique à sda4.

```
#fdisk /dev/sdb >n >p >(enter) >+100M >w #reboot
```

- b. Création de la grappe RAID1 mais uniquement avec cette partition.

```
#mdadm --create /dev/md3 --level=1 --raid-devices=2 missing /dev/sdb4 -y
```

- c. Formatage de la grappe.

```
#mkfs -t ext2 -q /dev/md3
```

3. Copie des données du disque à encapsuler vers la grappe RAID1.

```
#mkdir /mnt2 #mount -t ext2 /dev/md3 /mnt2 #mount -t ext2 /dev/sda4 /mnt #cp -dpRx /mnt/* /mnt2/
```

4. Ajout du disque d'origine au RAID1

```
#mdadm --manage /dev/md3 --add /dev/sda4
```

Dans la construction du RAID1, le disque d'origine va être écrasé. Ce n'est pas grave, il va recevoir la copie des données qui sont une copie de ses propres données.

5. Accéder maintenant aux données grâce à notre disque en miroir

```
#mount /dev/md3 /mnt
```

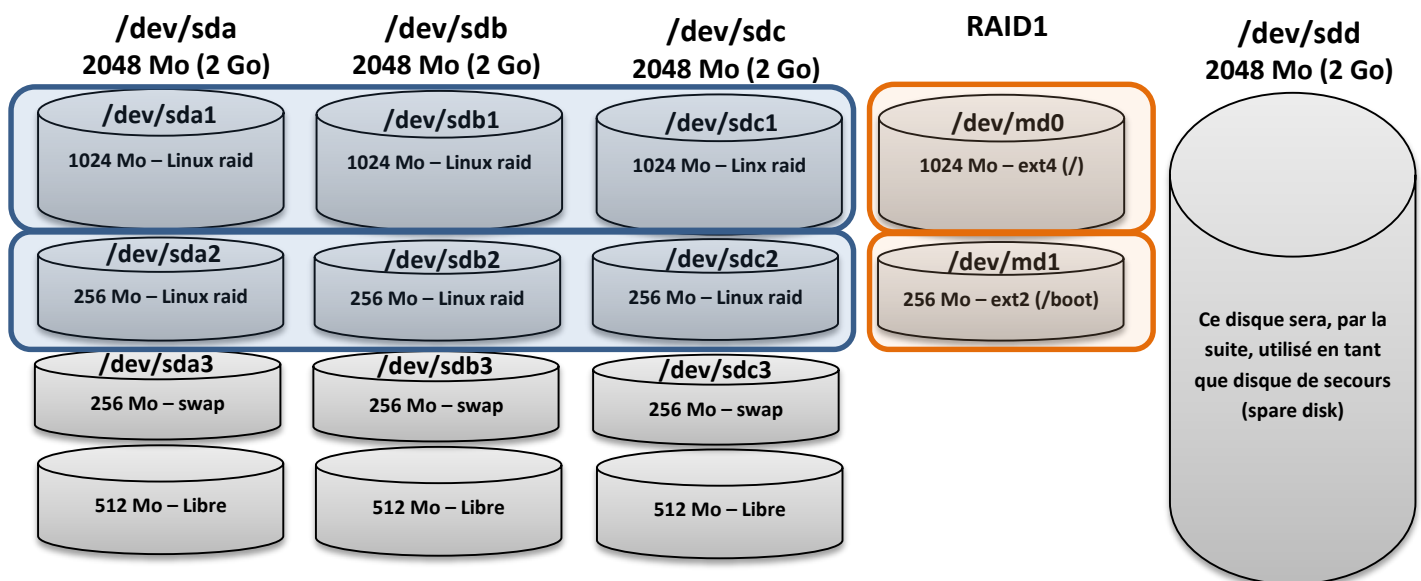
D. Installation et manipulation d'un système RAID5

Créez une machine virtuelle (VirtualBox) CentOS 6.x (64 bits) possédant les caractéristiques suivantes:

- 1 Go Ram, 4 processeurs
- pas de carte son, pas d'usb
- une carte réseau
- 4 disques SATA de 2048 Mo (2 Go) en taille fixe
- Possibilité de démarrer cette machine sur une image d'installation (.iso) CentOS 6.x minimale

Une fois configurée, le système devra pouvoir démarrer sur son système RAID1 opérationnel ou sur son RAID1 dégradé (bootage possible même si un des 3 disques est en panne).

Les disques durs respecteront le partitionnement et les RAID 1 suivants:



1. Configuration sous VirtualBox

Dans VirtualBox: *Système* → *Processeur* → 4
→ *Carte mère* → *Mémoire vive* → 1024 Mo
Son → [] *Activer le son*
USB → [] *Activer le contrôleur USB*
Réseau → *Carte1*
→ [v] *Activer la carte réseau*
→ *Avancé: Type de carte: Intel Pro/1000 MT Desktop (82540OEM)*
Stockage → *Contrôleur SATA*
HD1r5 (2 Go)
HD2r5 (2 Go)
HD3r5 (2 Go)
HD4r5 (2 Go)
→ *Contrôleur IDE*
CentOS-6.4...minimal.iso

2. Installation

- a. Effectuez les 1^{ère} étapes de l'installation comme d'habitude et choisissez un partitionnement personnalisé ...

...

- b. Respectez le partitionnement suivant

- i. Création des partitions RAID software

Pour la partition RAID software qui devra accueillir /boot:

Procédez de la même façon que pour le point "A. Installation et manipulation d'un système RAID1" mais pour les partitions sda1, sdb1, sdc1 des disques sda, sdb et sdc...

Procédez de façon similaire pour les partitions RAID software qui devront accueillir /.

- ii. Création des devices RAID1 à partir de ces partitions

Pour le périphérique RAID1 (/dev/md0) qui devra gérer le RAID software sur /:

Procédez de la même façon que pour le point "A. Installation et manipulation d'un système RAID1" mais pour les 3 partitions sda1, sdb1 et sdc1....

Procédez de façon similaire pour le périphérique RAID1 (/dev/md1) qui devra gérer le RAID software sur /boot.

- iii. Création des 3 zones de swap

Créez la zone de swap standard sur chacun des 3 disques...

Fenêtre de fin de configuration du RAID

Périphérique	Taille (Mio)	Point de montage/ RAID/Volume	Type	Formater
▼ Périphériques RAID				
md0 (/dev/md0)	1023	/	ext4	✓
md1 (/dev/md1)	256	/boot	ext2	✓
▼ Disques durs				
▼ sda (/dev/sda)				
sda1	1024	md0	software RAID	✓
sda2	256	md1	software RAID	✓
sda3	256		swap	✓
Libre	511			
▼ sdb (/dev/sdb)				
sdb1	1024	md0	software RAID	✓
sdb2	256	md1	software RAID	✓
sdb3	256		swap	✓
Libre	511			
▼ sdc (/dev/sdc)				
sdc1	1024	md0	software RAID	✓
sdc2	256	md1	software RAID	✓
sdc3	256		swap	✓
Libre	511			
▼ sdd (/dev/sdd)				
Libre	2047			

On "formate" et on "écrit les modifications sur disque"

On installe Grub sur le RAID pour pouvoir démarrer en RAID dégradé.

Procédez de la même façon que pour le point "A. Installation et manipulation d'un système RAID1" mais l'installation se fait sur /dev/md1.

On installe Grub sur le RAID pour pouvoir démarrer en RAID dégradé.

L'installation des paquets commence...

En fin d'installation, on redémarre...

On clôture l'installation correcte de Grub.

Redémarrer avec l'image .iso en mode "rescue" et réaliser les actions suivantes:

```
# chroot /mnt/sysimage
# grub          →      On réinstalle l'appel à Grub correctement sur
...              les MBR de sda, sdb et sdc et Grub dans /boot de sda,
                  sdb et sdc.

grub> device (hd0) /dev/sda
grub> device (hd1) /dev/sdb
grub> device (hd2) /dev/sdc

grub> root (hd0,1)
root (hd0,1)
Filesystem type is ext2fs, partition type 0xfd

grub> setup (hd0)
setup (hd0)
Checking if "/boot/grub/stage1" exist... no
Checking if "/boot/stage1" exist... yes
Checking if "/boot/stage2" exist... yes
... succeeded
Done

grub> root (hd1,1)
root (hd1,2)
Filesystem type is ext2fs, partition type 0xfd

grub> setup (hd1)
setup (hd1)
Checking if "/boot/grub/stage1" exist... no
Checking if "/boot/stage1" exist... yes
Checking if "/boot/stage2" exist... yes
... succeeded
Done

grub> root (hd2,1)
root (hd1,2)
Filesystem type is ext2fs, partition type 0xfd

grub> setup (hd2)
setup (hd1)
Checking if "/boot/grub/stage1" exist... no
Checking if "/boot/stage1" exist... yes
Checking if "/boot/stage2" exist... yes
... succeeded
Done

grub> quit
# exit      + 'disabler' SELINUX
# halt
```

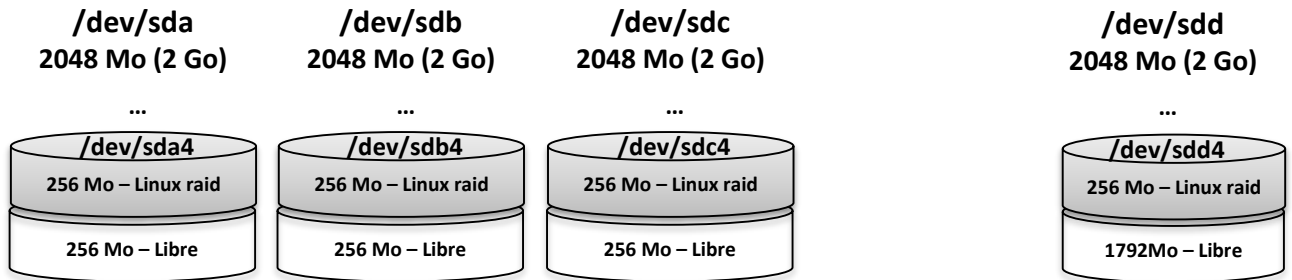
On éjecte l'image .iso du CD virtuel et on démarre la MV.

*Un ré-étiquetage SELINUX a lieu puis ... (pas si SELINUX a été désactivé !!!)
elle rebootte correctement ☺*

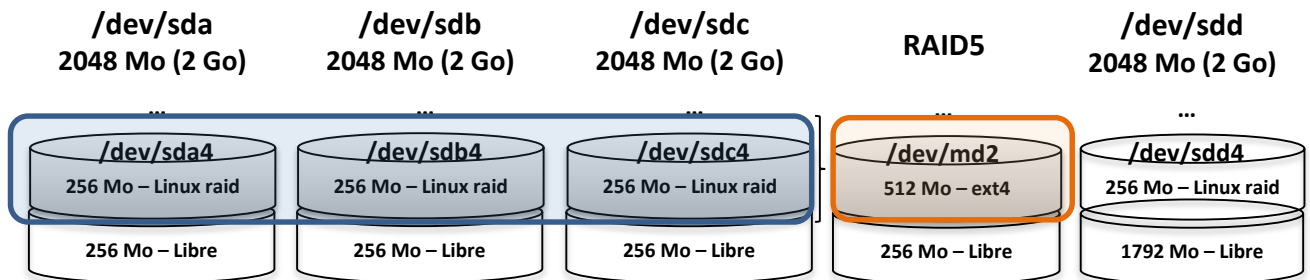
- Lister les matrices RAID
#mdadm --detail --scan
#cat /proc/mdstat
- Afficher des informations sur un disque RAID (composition, niveau ...)
#mdadm --detail /dev/md0 | more
- Installation d'un disque RAID5 et de son disque de secours

Principe:

- Créer 4 partitions Linux raid (flag fd) de 256 Mo sur /dev/sda4, /dev/sdb4, /dev/sdc4 et /dev/sdd4.



- Pour être propre, supprimer les signatures éventuelles sur chacune d'elles.
- Créer le disque RAID5, le formater et l'alimenter de données.



- Si besoin en est, mettre à jour le fichier /etc/fstab.
- Clôturer la configuration du RAID5 avant un redémarrage.
- Redémarrage et tests.

Mise en œuvre:

- Créer 4 partitions Linux raid (flag fd) de 256 Mo sur /dev/sda4, /dev/sdb4, /dev/sdc4 et /dev/sdd4

Attention: L'idéal est de commencer ces partitions au début d'un nouveau cylindre.

N'oubliez pas d'y coller le flag de code hexadécimal 'fd'.

Ces partitions doivent avoir exactement la même taille ...

→ Même si sdd ne contient encore aucune partition, vous pouvez créer une partition sdd4 commençant au cylindre 197 et se terminant au cylindre 230...

```
#fdisk /dev/sd* >n >p >4 >196 >+256M >t >fd >w
```

- Pour être propre, supprimer les signatures éventuelles sur chacune d'elles.

- c. Créer le disque RAID5, le formater et l'alimenter de données.

```
#mdadm --create /dev/md2 --level=5 --raid-devices=3 /dev/sda4 /dev/sdb4 /dev/sdc4 --spare-devices=1 /dev/sdd4
```

- d. Si besoin en est, mettre à jour le fichier /etc/fstab.

- e. Clôturer la configuration du RAID5 avant un redémarrage.

```
#mdadm --assemble --scan
```

```
#mdadm --detail --scan --verbose > /etc/mdadm.conf
```

- f. Redémarrage et tests.

```
#reboot
```

6. Simulation de panne

- a. Sur la grappe RAID5 avec le disque de secours

- i. On provoque une panne sur sdc4

```
#mdadm --manage --set-faulty /dev/md2 /dev/sdc4
```

- ii. On vérifie si la reconstruction sur le disque de secours a bien lieu (durant ce temps le système reste opérationnel).

```
#cat /proc/mdstat
```

- iii. On retire le disque en panne.

```
VIRTUALBOX
```

- iv. On ajoute un disque neuf

```
#fdisk /dev/sdc >n >p >4 >196 >+256M >t >fd >w  
#mdadm --manage /dev/md2 --add /dev/sdc4
```

- v. Clôturer la configuration du RAID5 avant un redémarrage.

```
#mdadm --assemble --scan
```

```
#mdadm --detail --scan --verbose > /etc/mdadm.conf
```

- vi. Redémarrage et tests.

```
#reboot
```

- b. Sur la grappe RAID5

- i. On regarde l'image de la matrice

```
#cat /proc/mdstat
```

- ii. On enlève le disque de secours

```
#mdadm --manage /dev/md2 --remove /dev/sdc4
```

- iii. On vérifie.

```
#cat /proc/mdstat
```

- iv. On provoque une panne sur sda4.

```
#mdadm --manage --fail /dev/md2 /dev/sda4
```

- v. On vérifie qu'il n'y a pas de reconstitution (durant ce temps le système reste opérationnel).

```
#cat /proc/mdstat
```

- vi. On retire le disque en panne.

```
#mdadm --manage /dev/md2 --remove /dev/sda4
```

- vii. On ajoute un disque neuf

VIRTUAL BOX

```
#fdisk /dev/sda >n >p >4 >196 >+256M >t >fd >w
```

```
#mdadm --manage /dev/md2 --add /dev/sda4
```

- viii. Clôturer la configuration du RAID5 avant un redémarrage.

```
#mdadm --assemble --scan
```

```
#mdadm --detail --scan --verbose > /etc/mdadm.conf
```

- ix. Redémarrage et tests.

```
#reboot
```

7. Arrêt – redémarrage et suppression d'un RAID.

- a. Arrêt du RAID5 de la grappe md2.

```
#umount /dev/md2
```

```
#mdadm --stop /dev/md2
```

```
#cat /proc/mdstat
```

- b. Redémarrage de ce RAID5.

```
#mdadm --assemble --scan
```

```
#mdadm --query /dev/md2
```

```
#mount /dev/md2 /home
```

- c. Arrêter le RAID5 et le supprimer.

```
#umount /dev/md2
```

```
#mdadm --stop /dev/md2
```

```
#mdadm --zero-superblock /dev/sd[a-d]4
```

PRINCIPALES COMMANDES MANIPULEES LORS DE CET ATELIER

Principales commandes:

mdadm, fdisk, tar, wipefs

Principaux fichiers

/proc/mdstat, /etc/mdadm.conf