Virtualisation système
&
Retour d'expérience(s) sur Xen

Franck.Elie@cnrs-orleans.fr

Cette présentation comporte des éléments issues des JoSy Virtualisation (28 septembre 2006) et documents trouvés sur le web

Plan

- Virtualisation
  - Qu'est-ce que la Virtualisation système
  - Avantages / inconvénients de la virtualisation
  - Techniques de virtualisation système
    - types de virtualisation
    - virtualisation et processeurs x86
    - Petit comparatif VMware / Xen ; benchmarks
- Retour d'expérience sur Xen
  - Présentation du système de virtualisation du LPCE
    - Matériel
    - Caractéristiques systèmes (stockage, réseau, etc.)
    - Utilisation
  - Problèmes rencontrés
  - Evolutions envisagées

Virtualisation

- Définition : ensemble de techniques matérielles et/ou logicielles destinées à faire fonctionner sur une seule machine **plusieurs systèmes d'exploitation et/ou applications**, séparément les uns des autres, comme s'ils fonctionnaient sur des **machines physiques distinctes**.

- Constat :
  - ↑ puissance du matériel informatique (pour faire tourner Vista ?)
  - ↓ taux d'utilisation des machines
  - ↓ nombre d'Administrateurs Système & Réseau (↑ nb de services)
  - ↓ budget (achat/recyclage, électricité, maintenance)
  - ↓ espace de travail (PC de bureautique + portable + machine de calcul)
  - ↑ prise en compte d'aspects liés à l'écologique (déchets) ou aux conditions de travail (nuisances dues au bruit, température)

Avantages de la virtualisation

- Consolider les serveurs tout en isolant les services :
  - meilleure exploitation des ressources matérielles
  - amélioration de la sécurité par isolation des services
  - amélioration de la flexibilité par séparation des services *
- Accroître la disponibilité des services (redondance, migration)
- Répondre rapidement aux nouveaux besoins (exemple : site web) *
- Faciliter les taches de tests, expérimentations
- Déléguer l'administration de certains serveurs
- Partager des ressources de calcul (pool de calculateurs) *
- Minimiser :
  - les coûts (maintenance, électricité, achat de matériel, etc.)
  - la quantité de matériel à maintenir et la place occupée *
  - les spécificités matérielles (les MV ne gèrent pas le matériel)

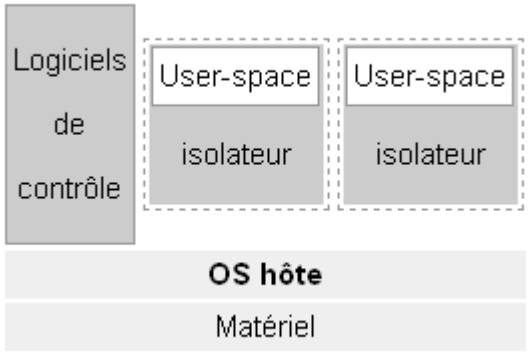
Inconvénients de la virtualisation

- Pertes plus importantes en cas de panne de la machine hôte (plusieurs services indisponibles)
- Complexité :
  - de mise en œuvre (évaluation des ressources matérielles, installation)
  - de gestion des MV (multiplication d'images systèmes, savoir gérer correctement les images systèmes, les mettre à jour)
- Performances inégales selon les technologies employées :
  - overhead variable selon les technologies et les systèmes clients
  - baisse des perf possible (nombre important d'I/O)
- Coût potentiellement élevé suivant le nombre de MV et options désirées :
  - nécessité d'un serveur hôte plus puissant
  - centralisation des MV ≡ techno DAS voire SAN
  - prix des technologies propriétaires (exemple : VMware infrastructure)

Techniques de virtualisation système
types de virtualisation 1/2

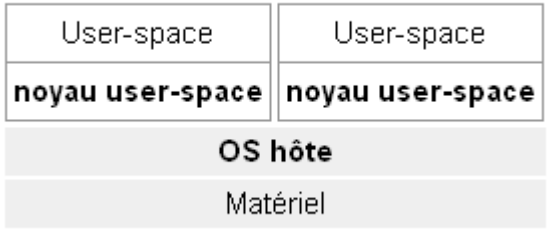
Isolateur :
• **couche logicielle** permettant d'isoler des applications dans des **contextes d'exécution** différents
• technique très performante (native), utilisé dans certains centres de calculs (Plateforme en Ligne Mathrice)
• les « machines virtuelles » sont issues du même OS
➔ exemples : *Linux-VServer* (isolation des processus en espace utilisateur), *chroot* (isolation par changement de racine)

Architecture d'un isolateur



Noyau en espace utilisateur :
• s'exécute comme une **application** « standard » dans l'**espace utilisateur** du système hôte (ce système hôte à lui-même un noyau qui s'exécute directement sur la machine matérielle en espace privilégié)
• technique peu performante car on empile deux noyaux, et donc plus adaptée à des développements
• les « machines virtuelles » sont issues du même OS
➔ exemple : *User Mode Linux (*noyau s'exécutant en user-space)

Architecture d'un noyau en user-space



(*Source : http://www.urec.cnrs.fr/IMG/pdf/asr.josy.virtualisation.perrot.pdf*)

Techniques de virtualisation système
types de virtualisation 2/2

Machine Virtuelle (superviseur) :

• **logiciel** (généralement assez complexe et lourd) qui s'exécute sous le contrôle d'un système hôte (host)

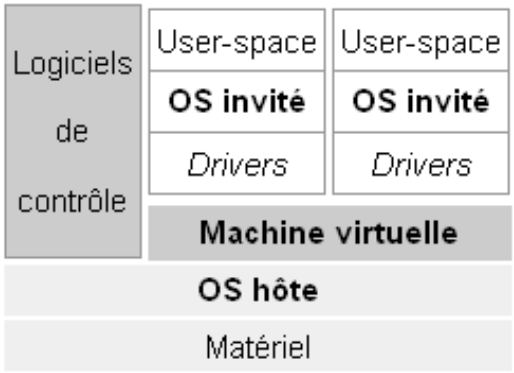
• le système hôte procure l'accès générique aux ressources physiques (disques, périphéries, connexions) : les OS invités croient être interfacés avec cette périphérie

• les OS hôtes et invités doivent être de même architecture matérielle (processeur en particulier), sauf...

... s'il s'agit d'un « émulateur »

➔ exemples : *Qemu* (émulateur de plateforme x86, PPC, Sparc), *Vmware player ou serveur* (plateforme x86), virtualBox

Architecture d'une machine virtuelle



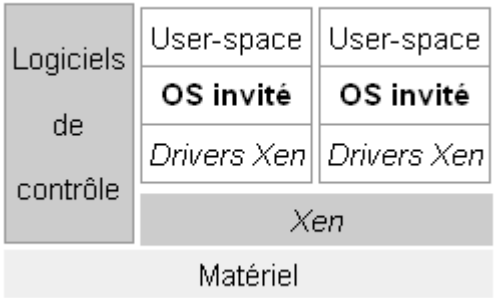
hyperviseur et Para-virtualiseur :

• L'hyperviseur est **un noyau hôte** restreint (allégé) et **optimisé** pour assurer l'exécution exclusive d'OS invités

• Les performances sont quasi natives, il s'agit de la technique de virtualisation logicielle la plus efficace

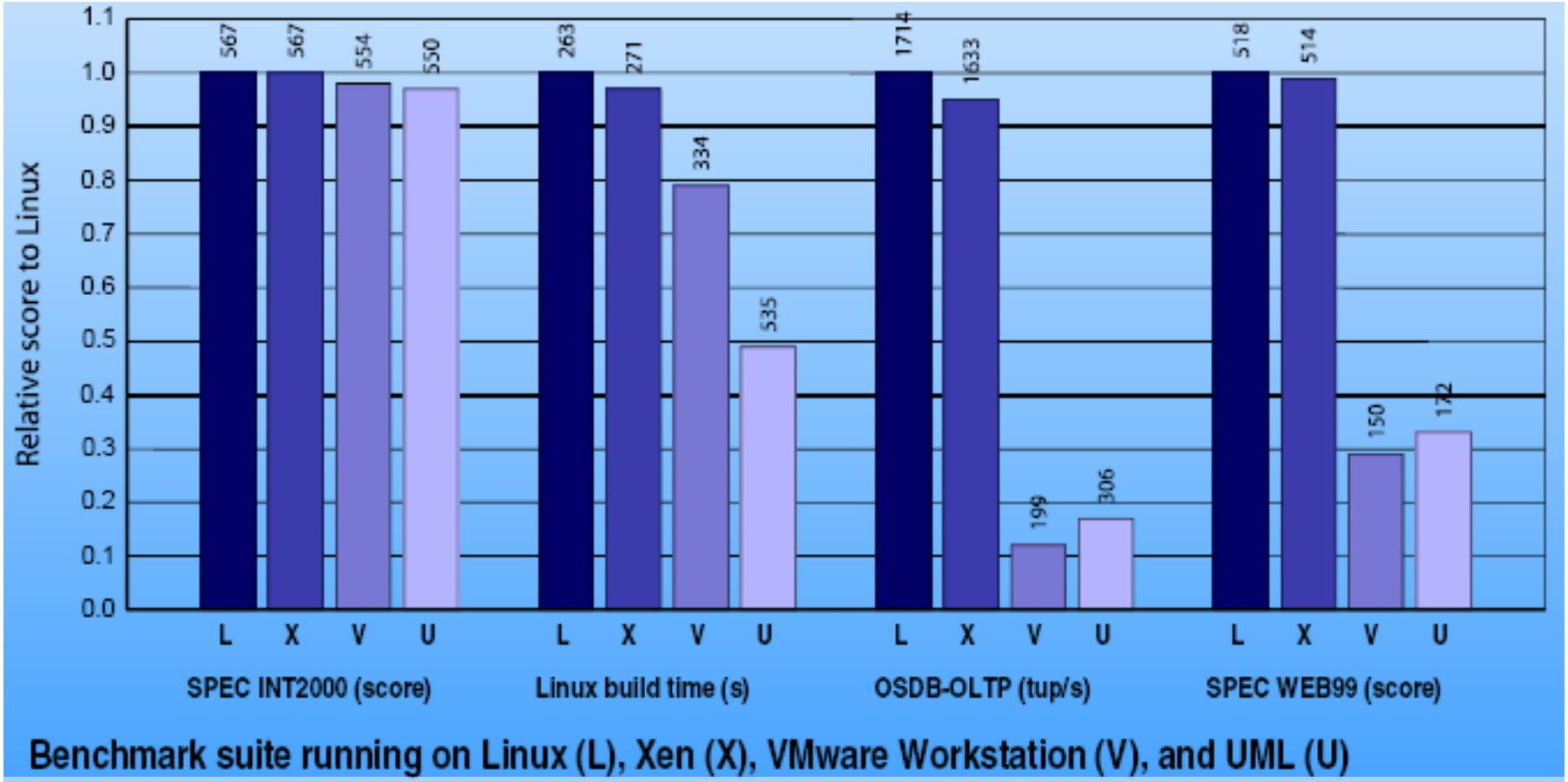
➔ exemples : *VMware ESX server, Xen*

Architecture Xen



(*Source : http://www.urec.cnrs.fr/IMG/pdf/asr.josy.virtualisation.perrot.pdf*)

Benchmarks



SPEC INT2000 (calcul intensif, peu d'I/O) ; OSDB-LLTP (transaction type SGBD) ; SPEC WEB99 (web dyn.)

(*Source : http://www.cl.cam.ac.uk/research/srg/netos/xen/performance.html*

Comparaison VMware / XEN

VmWare

- Hôtes supportés
  - Linux Mandrake, Red Hat, SuSe
  - Windows NT, 2000 server, XP
- Invités supportés
  - Windows, MS-DOS, Linux, Novell, Solaris, FreeBSD
- Pas de modification du code source des O.S. invités
- Dépendance commerciale
- Maturité
- Prix* (Stand. 2PE : 1884€ + 394€/an)

** prix à titre indicatif*

XEN

- Hôtes supportés
  - Linux Debian, Mandrake,Red Hat, SuSE, etc.
- Invités supportés
  - Mandrake, Red Hat, SuSE, Debian, FreeBSB, NetBSD
  - Windows (avec PE récents)
- Modification des sources (1.5%)
  - ➔ Meilleures performances
- Open Source
- Prix* :
  - Gratuit
  - Distribution payante via Xensource (avec IHM admin)
    (2PE : 750€ + 150€/an)

Architectures matérielles

- Questions préalables :
  - budget : aucun, petit ou gros ? global ou échelonnée ?
  - stockage des VM : local au(x) serveur(s) ou centralisé (SAN, DAS, iSCSI, NFS) ?
  - besoin CPU / VM : multiprocesseurs, BladeCenter ou pizza boxes ?
  - quantité en RAM et espace disque : avec ou sans interface X?
  - sauvegarde des VM : globale ou juste les données
- Choix CPU : Intel ou AMD ? Hypertransport ou Northbridge ?

Retour d’expérience sur XEN

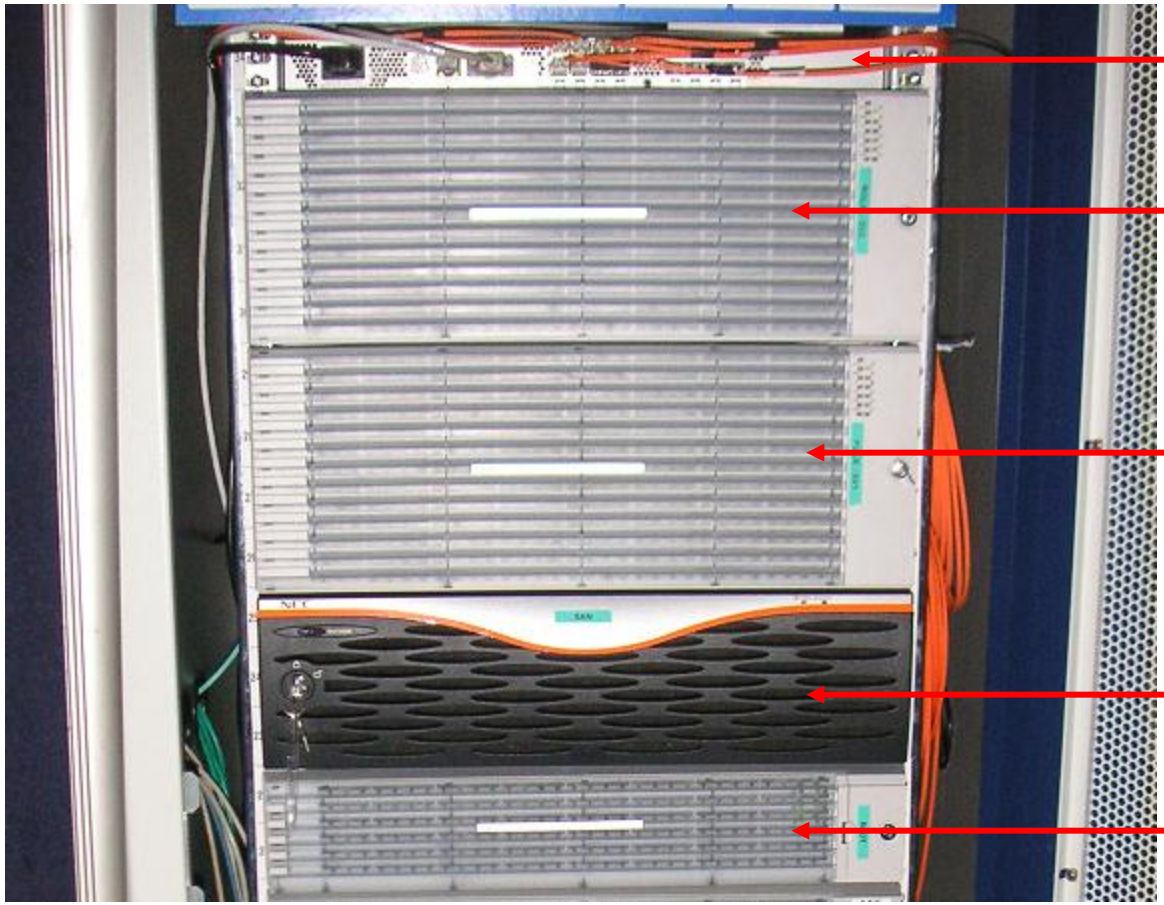
Spécification du besoin du LPCE

- Objectifs :
  - Ajouter des services ET diminuer le nombre de machines physiques à administrer (place, consommation électrique,etc.)
  - Améliorer la gestion des pannes matérielles
  - Partager les ressources de calcul (pool de calculateurs virtuels)

- Besoin :
  - Services : vingtaine (LDAP, SMTP(s), DNS, DHCP, CMS, etc.)
  - Plateformes de calcul et traitement intensif de données avec divers OS (Debian, Suse, Fedora) => administration déléguée

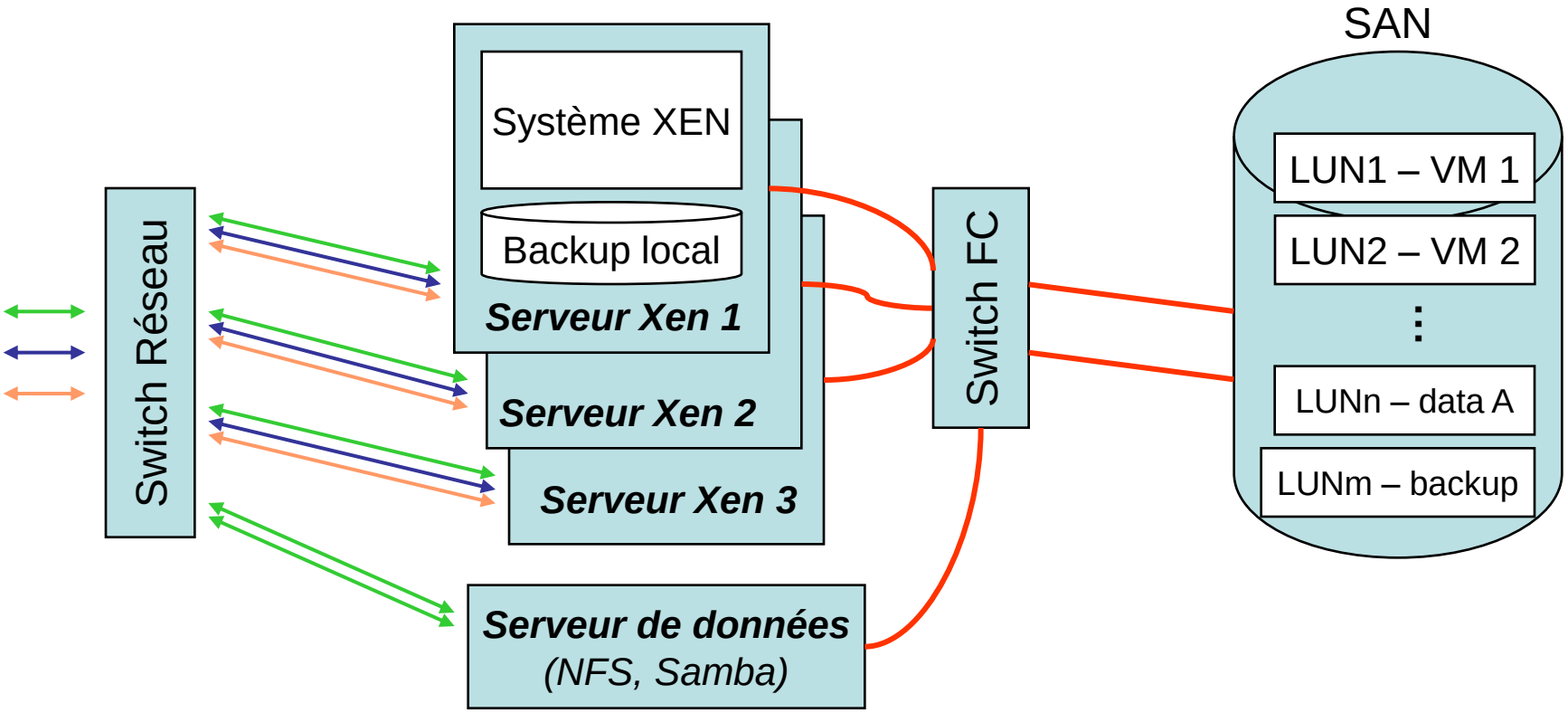
    ➔ séparation des machines de calcul et du stockage

Serveurs de virtualisation du LPCE



Organisation système

- Serveurs Xen avec backup des MV en local lors de l'arrêt système (coupures de courant)
- SAN : 1 MV ⇔ 1 LUN (2Go par défaut) ; 2 contrôleurs FC (redondance)
- 3 NIC / serveurs Xen (3 VLANs) ; 2 NIC pour le serveur de données



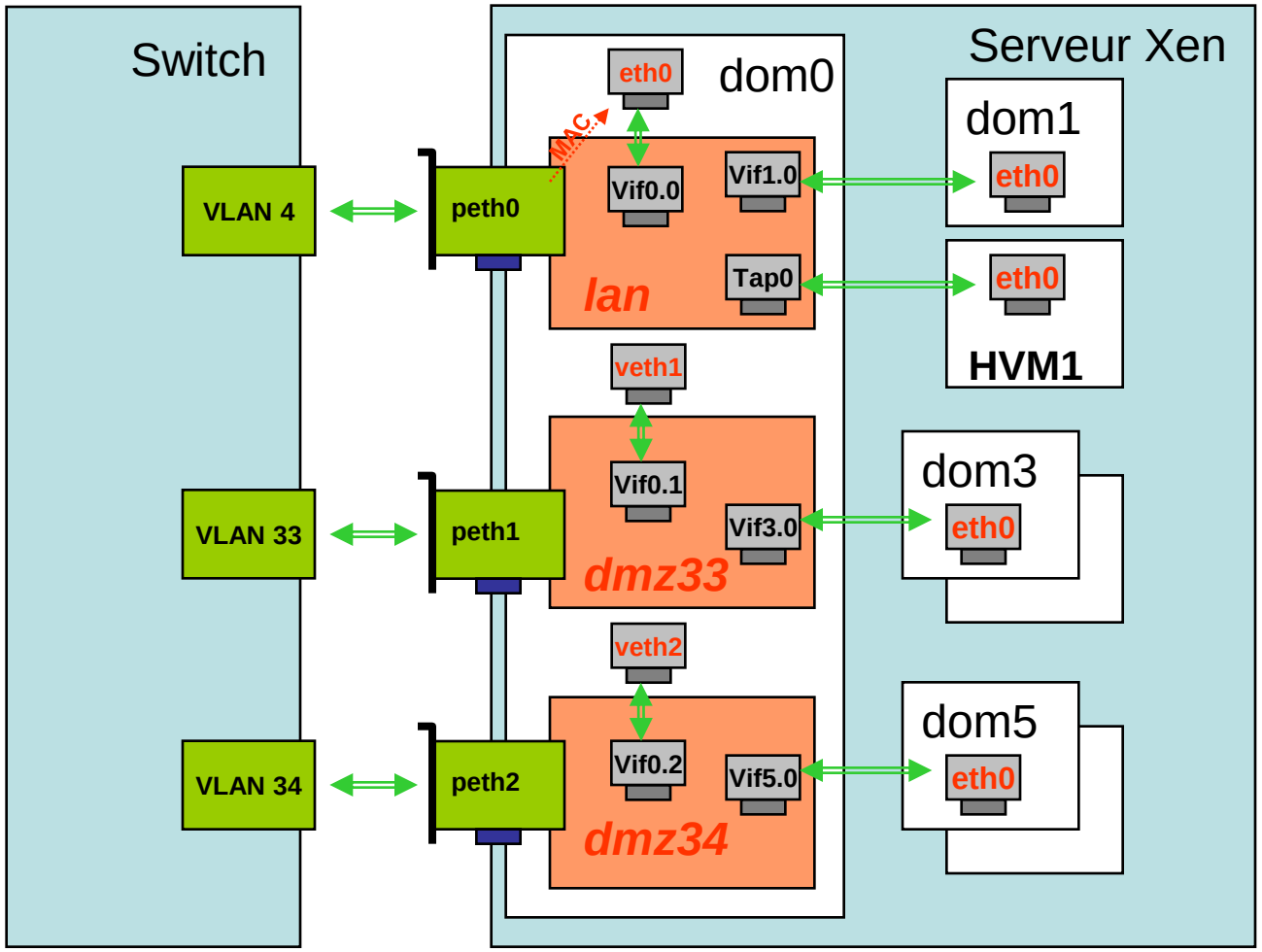
Matériel de virtualisation

- 3 Serveurs XEN :
  - 2 NEC 140 Rh-4 (4*MP 7110M) avec 8Go de RAM, 1 carte FC
  - 1 IBM xseries 235 avec 1Go de RAM, disque en RAID5, 1 carte FC
  - Système GNU Linux/Debian 4.0 (Etch) avec :
    - package Xen 3.0.3 (mise à jour de sécurité)
    - Multipath (SAN avec 2 contrôleurs FC)
    - LVM pour gestion locale du dom0 (pas les domU)
    - Gestion réseau en mode bridge avec 3 NIC dans 3 VLAN différents
  - dom0 en adressage privé avec filtrage via le parfeu + KVM IP

- 1 système de stockage SAN Nec S1500 (3.3To) avec
  - 15 disques 300Go 10KTm FC (13 en RAID 6 et 2 en spares)
  - 2 contrôleurs FC

- 1 switch FC 8 ports (Brocade 200E)

Configuration réseau

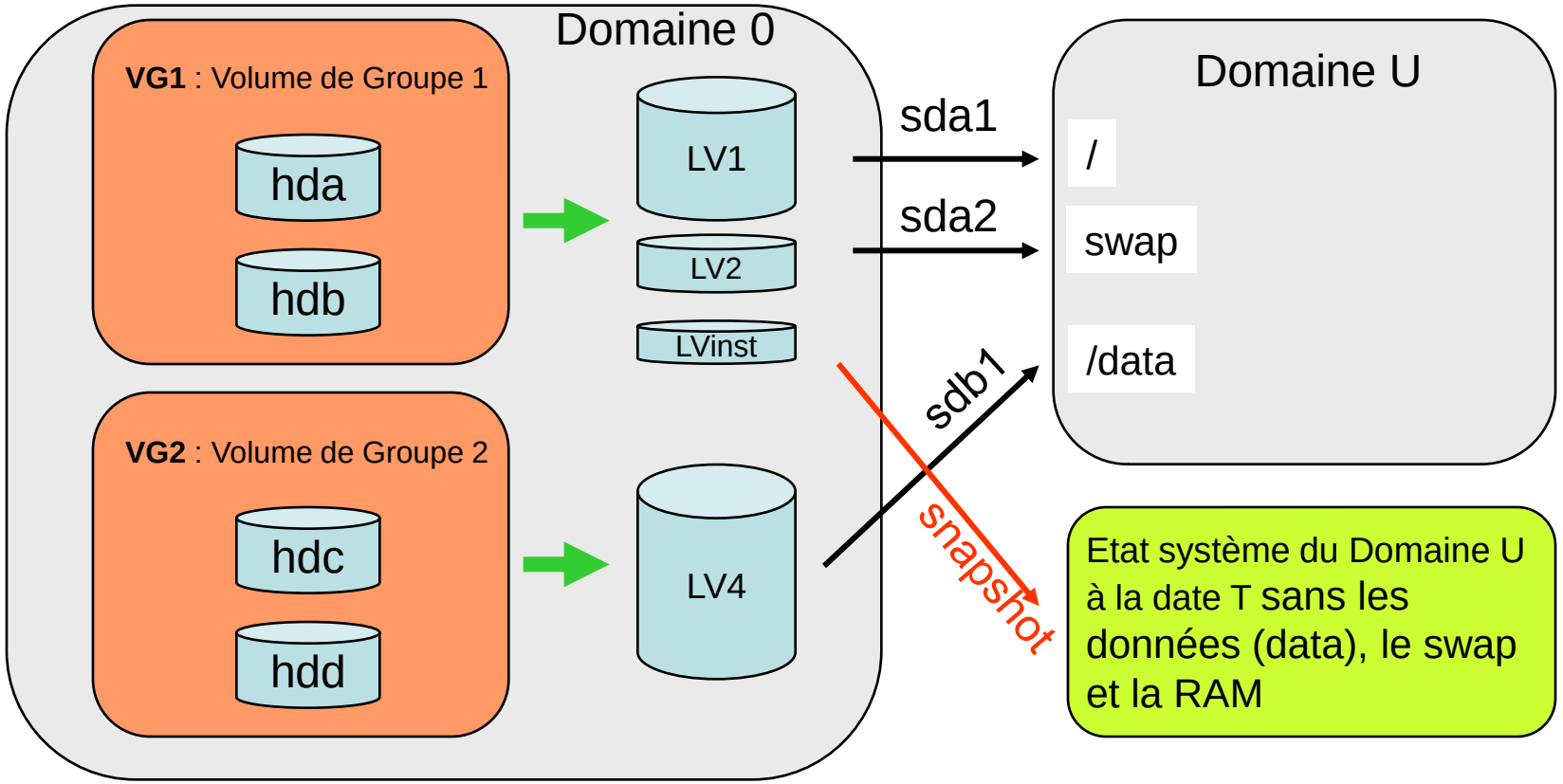


Stockage des machines virtuelles

- 3 types de stockage :
  - fichiers locaux : perf moins bonnes, pas de migration à chaud
  - partition du SAN (LUN) : gestion des noms de partition ⇔ VM
  - LVM : attention à la gestion de l'accès multiple (mode cluster)
    - dans le dom0 : 1 volume logique du dom0 ⇔ 1 partition du domU
    - dans les domU : 1 VG du domU ⇔ plusieurs VL du dom0

- Choix 1 MV par LUN car :
  - Renommage des LUNs avec multipath (/dev/mapper/VM_LDAP)
  - Possibilité d'augmenter la taille des LUNs (RAID6)
  - Limitation des risques relatifs au LVM en mode cluster

XEN et LVM
utilisation classique (1/4)

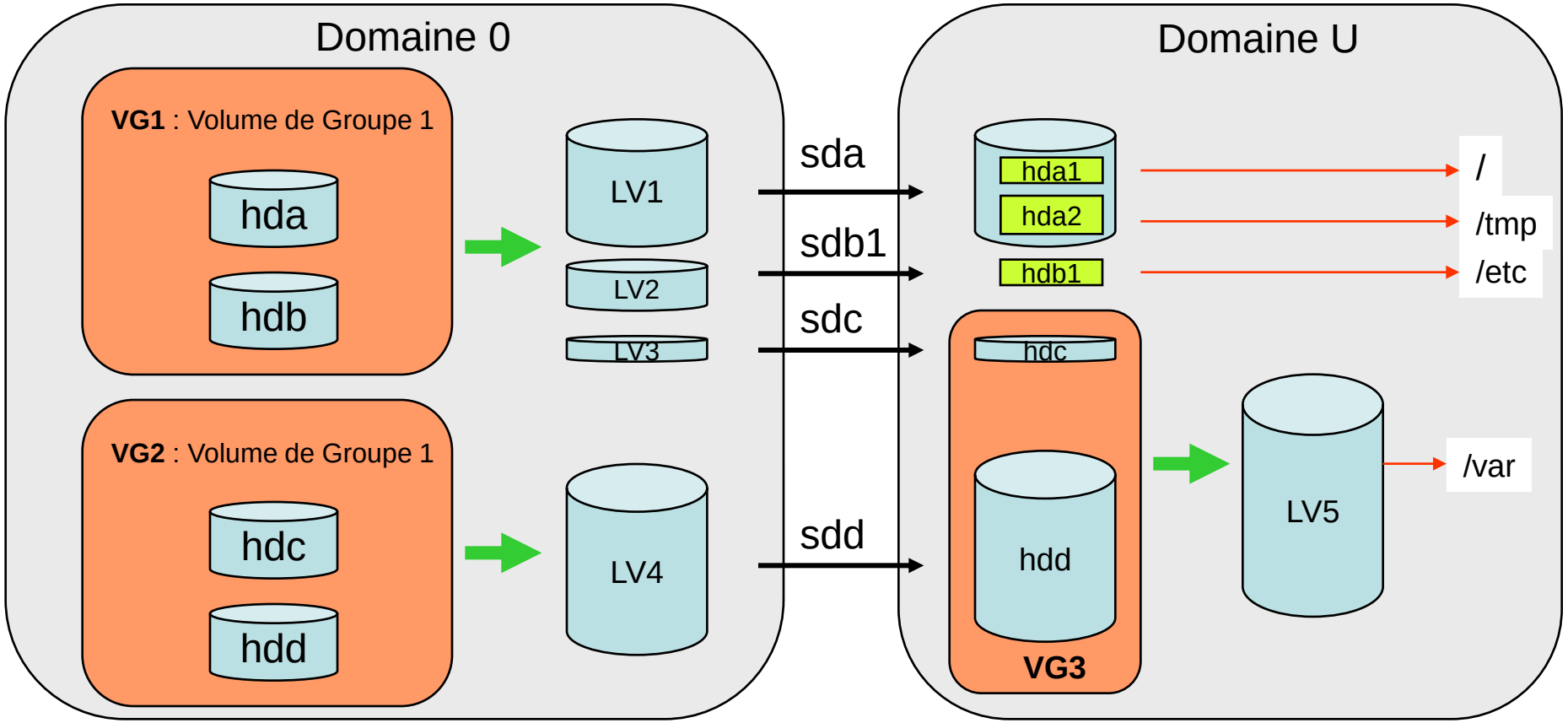


/dev/hda /dev/hdb
/dev/hdc /dev/hdd

/dev/mapper/VG1-VL1
/dev/mapper/VG1-VL2
/dev/mapper/VG2-VL4

disk = ['*phy:/mapper/VG1-VL1,sda1,w',*
/mapper/VG1-VL2,sda2,w',
/mapper/VG2-VL4,sdb1,w']

XEN et LVM
avec LVM dans les dom0 et domU (2/4)

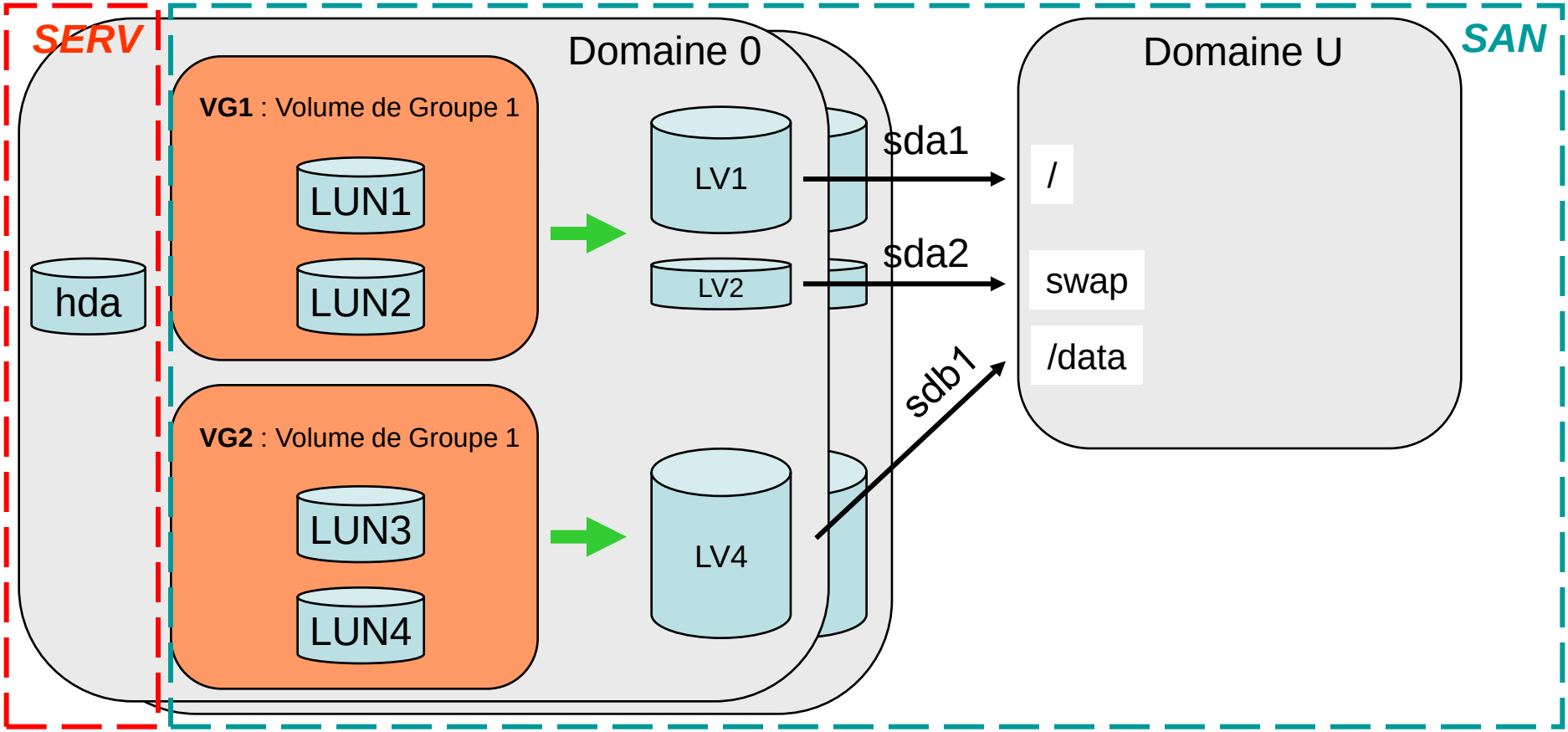


/dev/hda /dev/hdb
/dev/hdc /dev/hdd

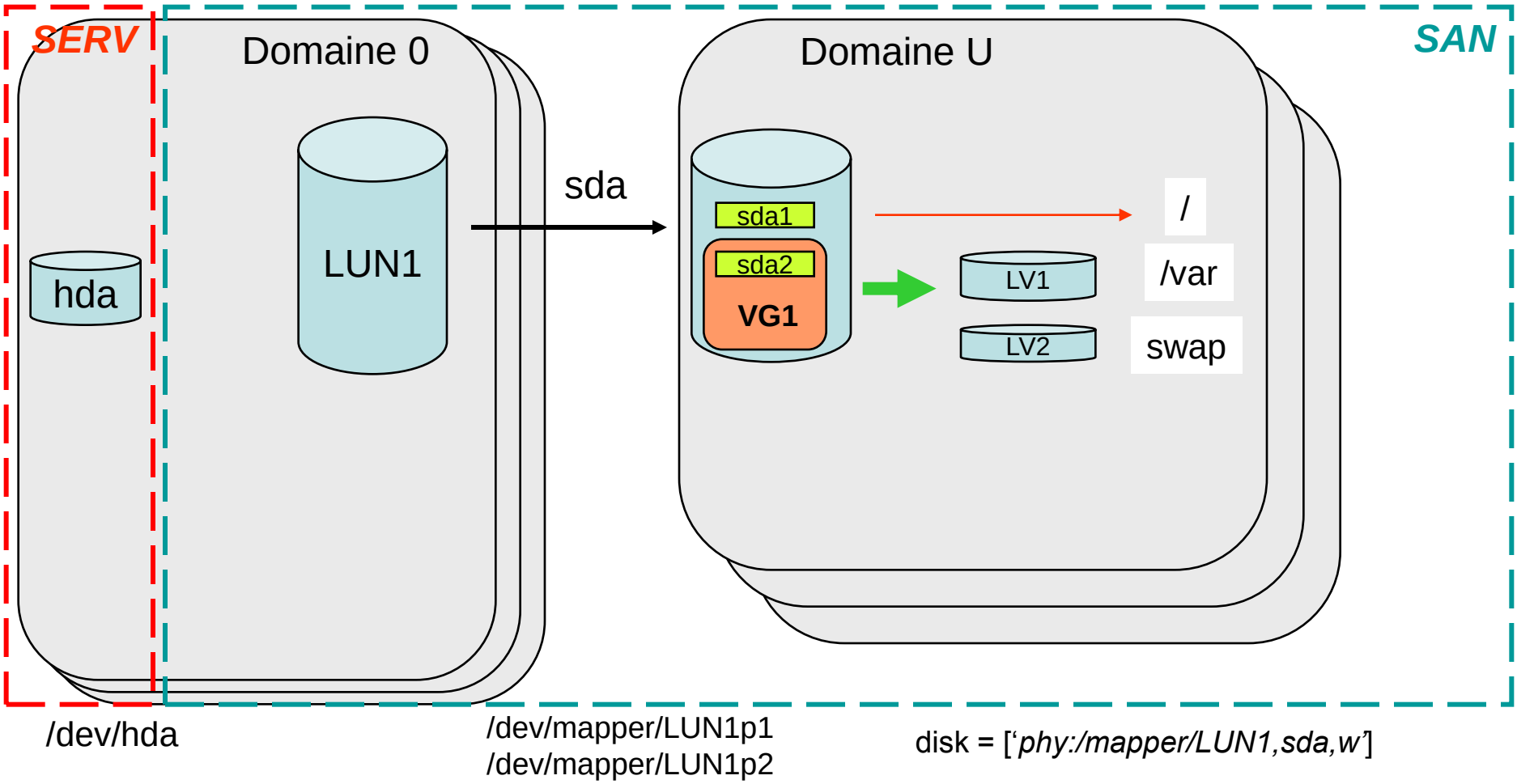
/dev/mapper/VG1-VL1
/dev/mapper/VG1-VL2
/dev/mapper/VG1-VL3
/dev/mapper/VG2-VL4

disk = [‘*phy:/mapper/VG1-VL1,sda,w’,*
/mapper/VG1-VL2,sdb1,r’,
/mapper/VG1-VL3,sdc,w’,
/mapper/VG2-VL4,sdd,w’]

XEN, LVM et SAN
avec LVM géré par dom0 (3/4)



XEN, LVM et SAN
avec LVM géré par domU (4/4)



Administration des VM (1/2)

- **Accès aux machines virtuelles :**
  - Mode console :
    - Xen console domU
    - connexion ssh (avec éventuellement un renvoi X11)
  - Mode graphique :
    - Vnc,
    - Xdmcp ou connexion bureau distance (Windows)
  - Montage du système de fichier (MV arrêtée)

- **Sauvegardes** :
  - Sauvegarde via la MV des données (mysqldump)
  - Sauvegarde via arkeia (client dans les VM)
  - Copie (dd) des VM arrêtées dans un fichier stocké dans le SAN (montage NFS) et sur bande

Administration des VM (2/2)

- **Nouvelles VM** :
  - *debootstrap* (debian)
  - duplication (VM template : CMS avec joomla)
  - conversion VMware en mode HVM (qemu) puis xenification
- **Modification de l'attribution des ressources** :
  - Taille de la RAM
  - Nombre et choix des CPU
- **Changement d'hôte** : ➔ Migration 'live'

Fonctionnalités non testées

- IHM d'administration (xenman, enomalisme,etc.)
- Le scheduling
- Compilation de Xen
- Approche NFS
- Fonctionnalités des version 3.0.4 et 3.1
  - HVM save/restore/migrate
  - Amélioration du support SMP & ACPI

Problèmes rencontrés

- Délégation de l'administration (root sinon rien)
  - ➔ relance d'une VM à partir d'une sauvegarde
  - ➔ Administration centralisée des serveurs XEN via une IHM
- Temps de vidage du cache lors de la sauvegardes des domU avec dd
  - Mise en évidence par un calcul md5sum en boucle après un shutdown du domU
  - Cache système + cache Xen ?
- Gestion délicate du LVM + serveurs multiples + SAN

Evolutions envisagées

- CLVM pour des domU dans LVM => sécurité et snapshot
- GFS pour système de fichier partagé
- Heartbeat pour le serveur de données (serveur NFS virtuel en spare)
- Poursuite de la politique de séparation des données et calcul ; mise en commun des demandes de moyen
  - ➔ projet d'ajout de 10To au SAN et raccordement de 3 bi-proc quad-core
- Upgrade de XEN en version 3.1 (attente Lenny « Stable » ?)
  - gestion multiprocesseur des HVM
  - Sauvegardes d'états
- Exploiter les linux-Vserver pour les machines dédiées au calcul

Conclusions sur XEN

- Avantages :
  - Performances (para-virtualisation)
  - Flexibilité (XEN n'est pas un soft, mais une fonctionnalité système)
  - Gratuité : possibilité de déployer à l'infini
  - *Stimulant sur le plan intellectuel …*

- Inconvénients :
  - Pas de support technique (à moins d'utiliser la solution xensource)
  - Nécessite des connaissances UNIX (lignes de commandes)

➔ Solution actuellement la mieux adaptée à mon besoin

Quelques URLs

- http://www.urec.cnrs.fr/article350.html
- http://xen.xensource.com/ (Xen Opensource)
- http://www.cl.cam.ac.uk/research/srg/netos/xen/performance.html
- http://pub.leblond-fr.info/spip.php?article78
- http://grid.ncsa.uiuc.edu/ggf12-sec-wkshp/panel4/hand.ppt (XEN)
- http://en.wikipedia.org/wiki/Comparison_of_virtual_machines