

# La virtualisation au Dpt. Informatique

Redondance, Archipel et  
délégation



# Le besoin

- Virtualisation en place: les invités en local
  - Utilisation de libvirt / KVM
  - Besoin de redondance
  - Ne pas tout réinstaller ...
- Services aux étudiants
  - Leur fournir des machines vierges
  - Gérer la sécurité entre machines
  - Besoin de pouvoir gérer des réseaux

# Les choix

- Libvirt/KVM
- Stockage partagé via le réseau
- Utilisation d'Archipel pour l'accès aux VM
  - surcouche libvirt
  - Client 100% web
  - Gestion des droits

# Les points faibles

- Stockage centralisé des disques des invités
  - Pb serveur de stockage = tout est arrêté
  - Panne de réseau pour accéder au serveur = idem
  - Stockage doit être extensible
- Besoin de migration pour la maintenance des hôtes

# Redondance du stockage

- Actif/Actif
  - Besoin d'un filesystem gérant les accès concurrents: GPFS, OCFS2, GFS2
  - Ou un filesystem distribué: GlusterFS, Lustre...
  - On a une répartition de charge
  - Complexe à mettre en œuvre
  - Ou stockage sur baie et système de réplication propriétaire

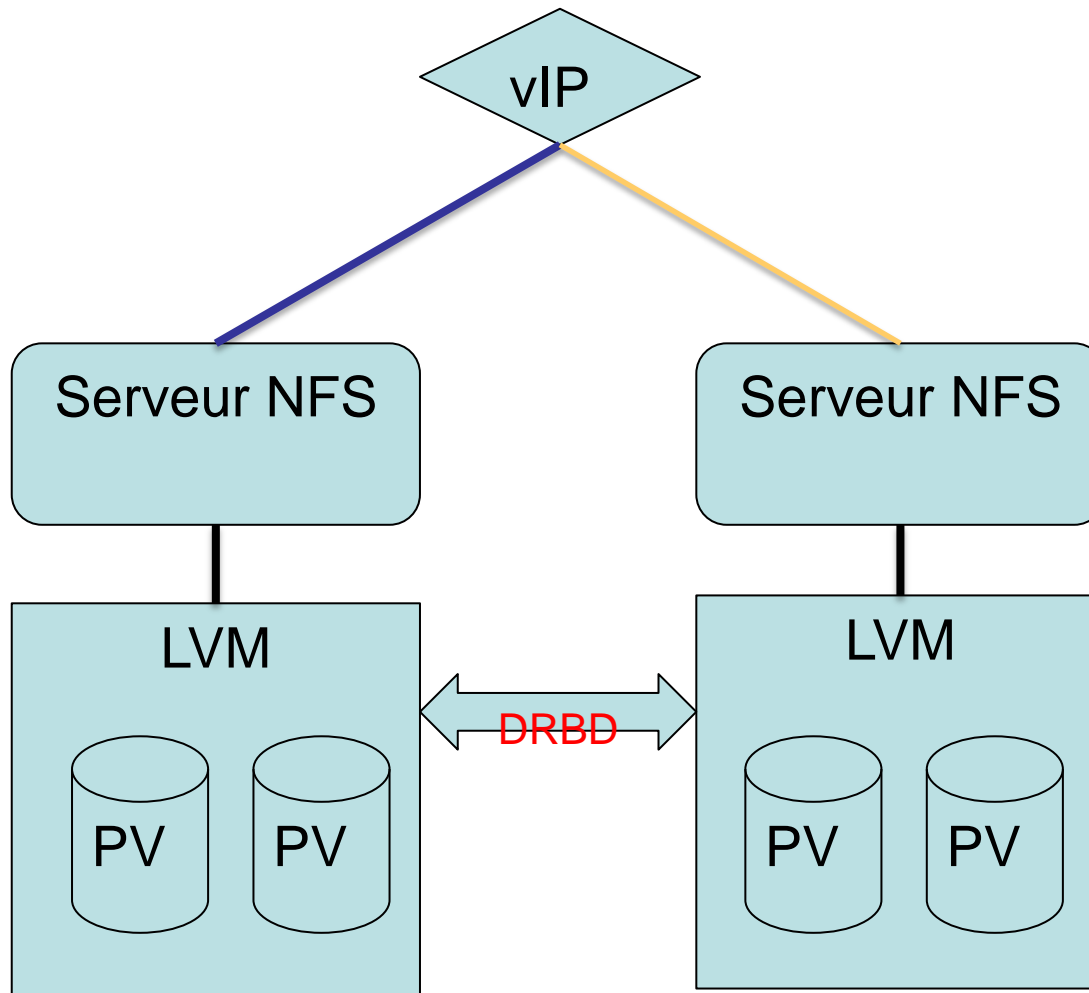
# Redondance du stockage

- Actif/Passif
  - On peut utiliser un filesystem classique
  - Facilité de mise en œuvre
  - Mais pas de répartition de charge

# Stockage: NFS + DRBD

- Choix de Actif/Passif
  - Utilisation de LVM
  - Puis filesystem ext3
- Réplication du stockage via DRBD
- Export via NFSv4
  - gestion des locks et reprise sur incident

# Schema stockage





# Stockage: LVM

```
pvcreate /dev/sdb1  
vgcreate storage /dev/sdb1  
lgcreate -L 500G storage
```

- Permet d'augmenter le partage plus tard:
  - LVM le permet
  - DRBD le permet
  - ext3 le permet
  - Transparent en NFSv4
  - Sans arrêt du service

# Stockage: DRBD

- Configuration (chaque serveur) `/etc/drbd.d/VMS.conf`:

```
resource VMS {  
    device /dev/drbd1;  
    disk /dev/storage/lvol0;  
    meta-disk internal;  
    on eb-1x1 {  
        address 130.79.80.54:7790;  
    }  
    on eb-1x2 {  
        address 130.79.80.59:7790;  
    }  
}
```

- Démarrage sur le serveur primaire

```
drbdadm up VMS  
drbdsetup /dev/drbd1 syncer -r 200M  
drbdadm primary VMS
```

# Stockage: NFS

- Baisser le délai de bascule:

```
echo 10 > /proc/fs/nfsd/nfsv4leasetime  
NEED_IDMAPD=yes # /etc/default/nfs-common
```

- Mettre les metadonnées nfs sur DRBD:

```
mv /var/lib/nfs /local/.private/ # sur actif  
ln -s /local/.private/nfs /var/lib/nfs #sur actif et passif
```

- Partager le stockage:

```
/local/nfs4  
192.168.0.0/24(rw,async,fsid=0,crossmnt,no_subtree_check,no_root_squash)
```

# Heartbeat + Corosync

- Sert à vérifier les services actifs et les bascules
- Veille par multicast, et authentification par clef commune (corosync-keygen)
- Configuration: modif dans `/etc/corosync/corosync.conf`

```
interface {  
    ringnumber: 0  
    bindnetaddr: 130.79.80.64  
    mcastaddr: 224.0.0.199  
    mcastport: 1974  
}
```
- A dupliquer:  
`/etc/corosync/authkey`  
`/etc/corosync/corosync.conf`

# Heartbeat + Corosync

- Configurer les services: `crm configure edit`
- Les primitives

```
primitive p_drbd_VMS \  
  ocf:linbit:drbd \  
  params drbd_resource="VMS" \  
  op monitor interval="15s" role="Master" \  
  op monitor interval="30s" role="Slave "
```

```
primitive p_fs_vms \  
  ocf:heartbeat:Filesystem \  
  params device=/dev/drbd/by-res/VMS \  
  directory=/local \  
  fstype=ext3 \  
  op monitor interval="30s"
```

```
primitive p_nfs_service lsb:nfs-kernel-server op monitor interval="10"
```

```
primitive p_ip_nfs \  
  ocf:heartbeat:IPaddr2 \  
  params ip=130.79.80.64 \  
  cidr_netmask=24 \  
  op monitor interval="30s"
```

# Heartbeat + corosync

- Les types, groupes et ordres

```
ms ms_drbd_VMS p_drbd_VMS \  
  meta master-max="1" \  
  master-node-max="1" \  
  clone-max="2" \  
  clone-node-max="1" \  
  notify="true" .
```

```
order o_drbd_before_fs inf: ms_drbd_VMS:promote p_fs_vms:start
```

```
group g p_fs_vms p_ip_nfs p_nfs_service
```

```
order o_nfsservice_after_fs inf: p_fs_vms:start p_nfs_service:start
```

```
colocation c_drbd_and_mount inf: ms_drbd_VMS:Master g
```

# Etat corosync

```
root@faucon1:~# crm_mon --one-shot
```

```
=====
```

```
Stack: openais
```

```
Current DC: faucon2 - partition with quorum
```

```
Version: 1.0.9-74392a28b7f31d7ddc86689598bd23114f58978b
```

```
2 Nodes configured, 2 expected votes
```

```
2 Resources configured.
```

```
=====
```

```
Online: [ faucon2 faucon1 ]
```

```
Resource Group: g
```

```
  p_fs_vms      (ocf::heartbeat:Filesystem):      Started faucon1
```

```
  p_ip_nfs      (ocf::heartbeat:IPaddr2):                Started faucon1
```

```
  p_nfs_service (lsb:nfs-kernel-server):                Started faucon1
```

```
Master/Slave Set: ms_drbd_VMS
```

```
  Masters: [ faucon1 ]
```

```
  Slaves:  [ faucon2 ]
```

# Augmenter le stockage

- (Plugger des disques -> /dev/sdc1)

```
#sur chaque serveur
```

```
pvadd /dev/sdc1
```

```
vgextend storage /dev/sdc1
```

```
lvextend -L +500G /dev/storage/lvo10
```

```
#Sur le serveur actif
```

```
drbdadm resize VMS
```

```
cat /proc/drbd # attendre la fin
```

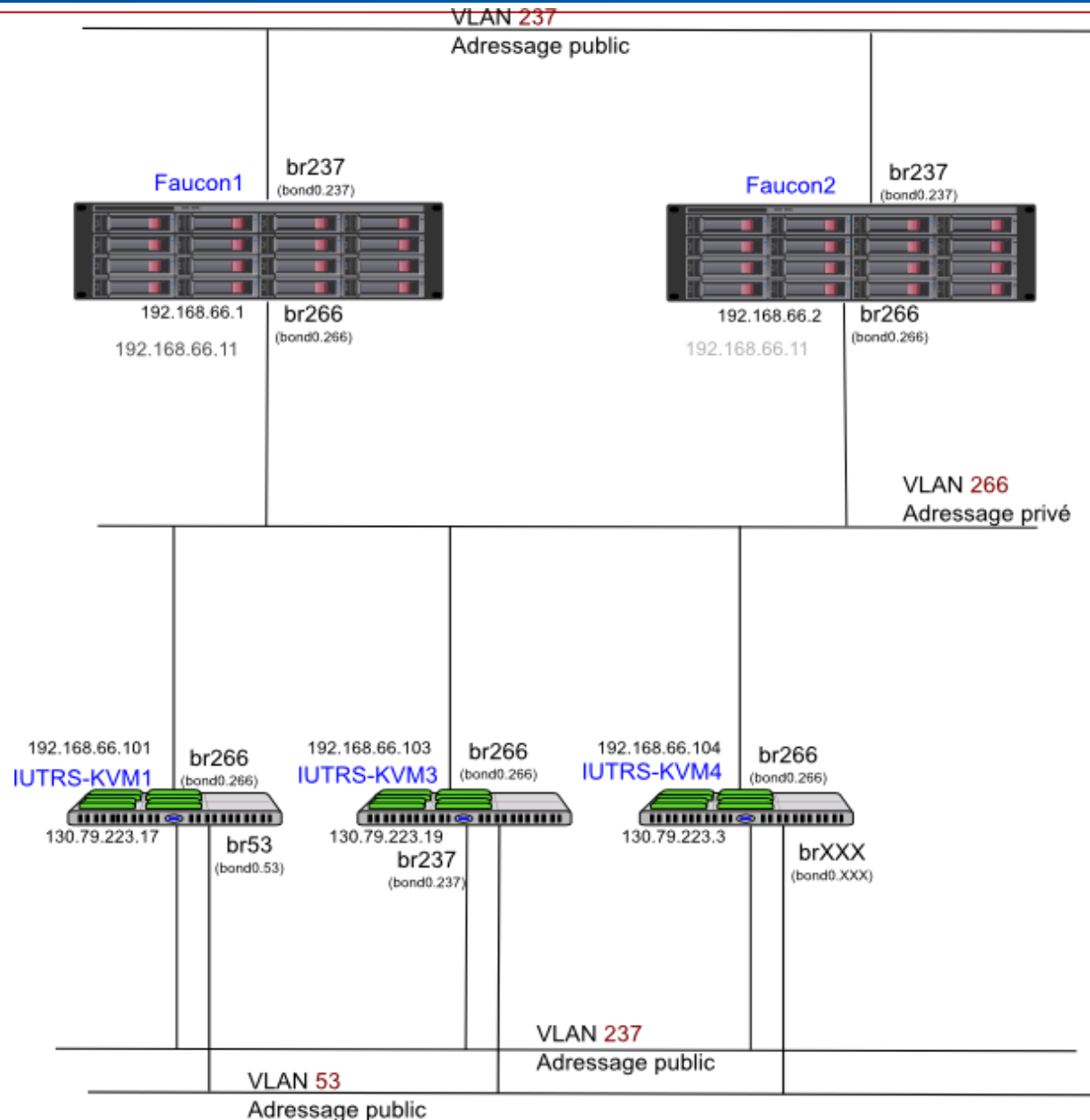
```
ext3resize /dev/drbd1
```



# Redondance: Réseau (stockage et hôtes)

- Double attachement réseau des serveurs
- Utilisation de bonding et de tagging
  - Bonding: redondance des liens
  - Tagging: utilisation de plusieurs VLAN
  - Attention au bonding:
    - Fault tolerance vs. Load balancing
    - Niveau de la panne réseau: port / switch / lien

# En production



- Serveur Jabber pour messages entre
  - Entité gérant hyperviseur
  - Entité gérant une VM
  - Entité interface client
- Version ejabberd 2.1.8 nécessaire ☹️
- Client lourd Javascript: (pas de web)
  - Chrome, safari (webkit), Firefox.

# Installation ejabberd

- Suivre la doc dans le wiki du projet
- Utiliser le fichier de conf fourni
- Tenter une connexion avec un client jabber.

# Installation Agent

- En charge de gérer les entités Hôte et VMs
- Agent en python
- Configuration `/etc/archipel/archipel.conf`
- Se charge de créer les comptes jabber
- Expose les propriétés de libvirt
- Stockage dans `/vm/drives/<UUID>/`

# Installation Client

- Rien à installer.
- Charge le javascript depuis un stockage
- Par exemple: <http://app.archipelproject.org/>
- Communication directe entre le navigateur et le serveur jabber.

# Intégration LDAP

- Autorisation du S2S sur le serveur ejabberd
- Installation serveur ejabberd classique avec config vers son LDAP
  - Réutiliser l'authentification pour les utilisateurs

# BOSH: Serveur à utiliser



**Archipel**  
XMPP based orchestrator

**Logon**

Jabber ID

Password

BOSH service

Remember  Off



# Scenario d'utilisation

- Droit de création de VM sur un hyperviseur
- Mise à disposition de switchs virtuels
- Gestion autonome par l'étudiant de sa VM:
  - Configuration (Disques, RAM ...)
  - Réseaux
  - Installation OS

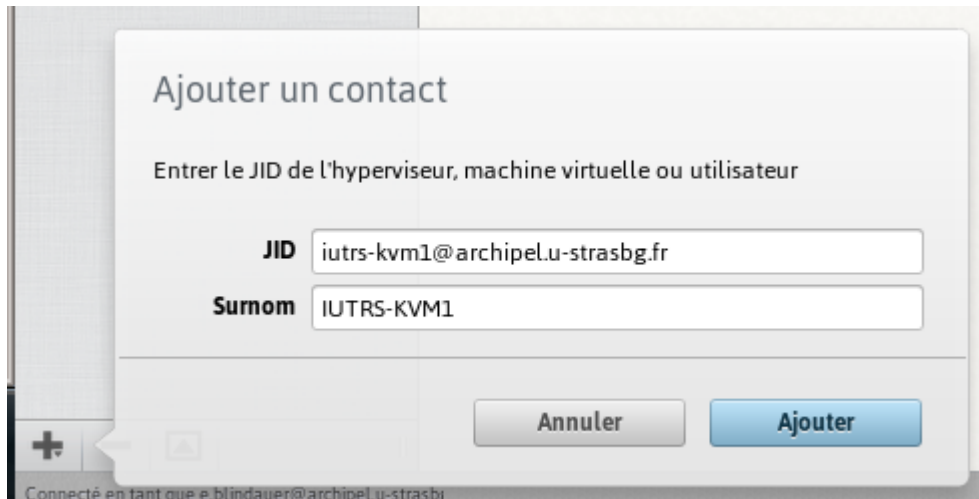
# Permissions

- Par GUI ou par CLI directement

The screenshot displays the Archipel web interface in a Google Chrome browser. The address bar shows 'app.archipelproject.org'. The interface includes a top navigation bar with menus for 'Archipel', 'Editer', 'Contacts', 'Groupes', 'Statut', 'Navigation', 'Modules', and 'Aide'. Below this is a toolbar with various icons for server management and communication. The main content area is titled 'Permissions' and contains the instruction: 'Set the different permissions and roles you give to users for this entity.' The interface is divided into two main sections: 'Roster Users' and 'Server Users'. The 'Server Users' section lists several users, with 'e.blindauer@archipel.u-strasbg.fr' selected. To the right, a list of permissions is shown with checkboxes: 'permission\_setown', 'subscription\_add', 'subscription\_remove', 'alloc' (checked), 'free', 'rostervm' (checked), 'clone', 'ip', 'uri', and 'capabilities'. The left sidebar shows a tree view of the system components, including 'HYPERVISEURS', 'SERVEURS DB', and 'SERVEURS INFRA'. The bottom status bar indicates the user is connected as 'admin@archipel.u-strasbg.fr' and provides version information: 'Archipel UI Version 0.5.0-ba1d13e (Moon) - Copyright 2011 Antoine Mercadal'.

# Action de l'utilisateur

- Ajouter l'hyperviseur auquel il a accès



Ajouter un contact

Entrer le JID de l'hyperviseur, machine virtuelle ou utilisateur

**JID**

**Sumom**

Connecté en tant que e.blindauer@archipel.u-strasbg.fr

# Vue utilisateur

The screenshot displays the Archipel web interface in a browser window. The address bar shows `app.archipelproject.org`. The interface includes a navigation menu with options like "Archipel", "Editer", "Contacts", "Groupes", "Statut", "Navigation", "Modules", and "Aide". A toolbar contains icons for "Statut", "Démarrer", "Pause", "Arrêter", "Détruire", "Ecran", and "Nouvelle VM".

The main content area is titled "Virtual machine controls" and shows the status of a virtual machine named "Ma VM". The status is "Running". Other details include "Memory: 512 MB", "Virtual CPUs: 1", "Auto start: Off", "Prevent OOM: Off", "OOM adjust: 0", and "CPU time: 0 min".

Below the controls, there is a "Live migration" section with the text "You can move virtual machine to another hypervisor." and a list of hypervisors where "IUTRS-KVM1" is selected. A "Kill the virtual machine" section is also visible, with the text "Killing a virtual machine is definitive. It will be completely removed".

Overlaid on the interface is a terminal window titled "Screen for Ma VM (c59cb634-08ad-11e2-82d7-00237d5a2dac@archipel.u-strasbg.fr/iutrs-kvm1) - Goo". The terminal shows a desktop environment with a blue background and a "Terminal" window icon. A context menu is open over the terminal icon, listing options such as "Run Program...", "Terminal Emulator", "File Manager", "Mail Reader", "Web Browser", "Settings", "Accessories", "Development", "Multimedia", "Network", "Office", "System", "About Xfce", and "Log Out". The system tray at the bottom of the terminal window shows the date and time as "2012-09-27 14:24".

# Conclusion

- Redondance stockage assuré
- Facilité de mise en œuvre
- Reprise de l'existant
- Archipel: Permet délégation sur des VM
- Archipel: encore des bugs