La virtualisation au Dpt. Informatique

Redondance, Archipel et délégation

Robert Schuman
UNIVERSITÉ DE STRASBOURG

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'acces 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 arreté
  - 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

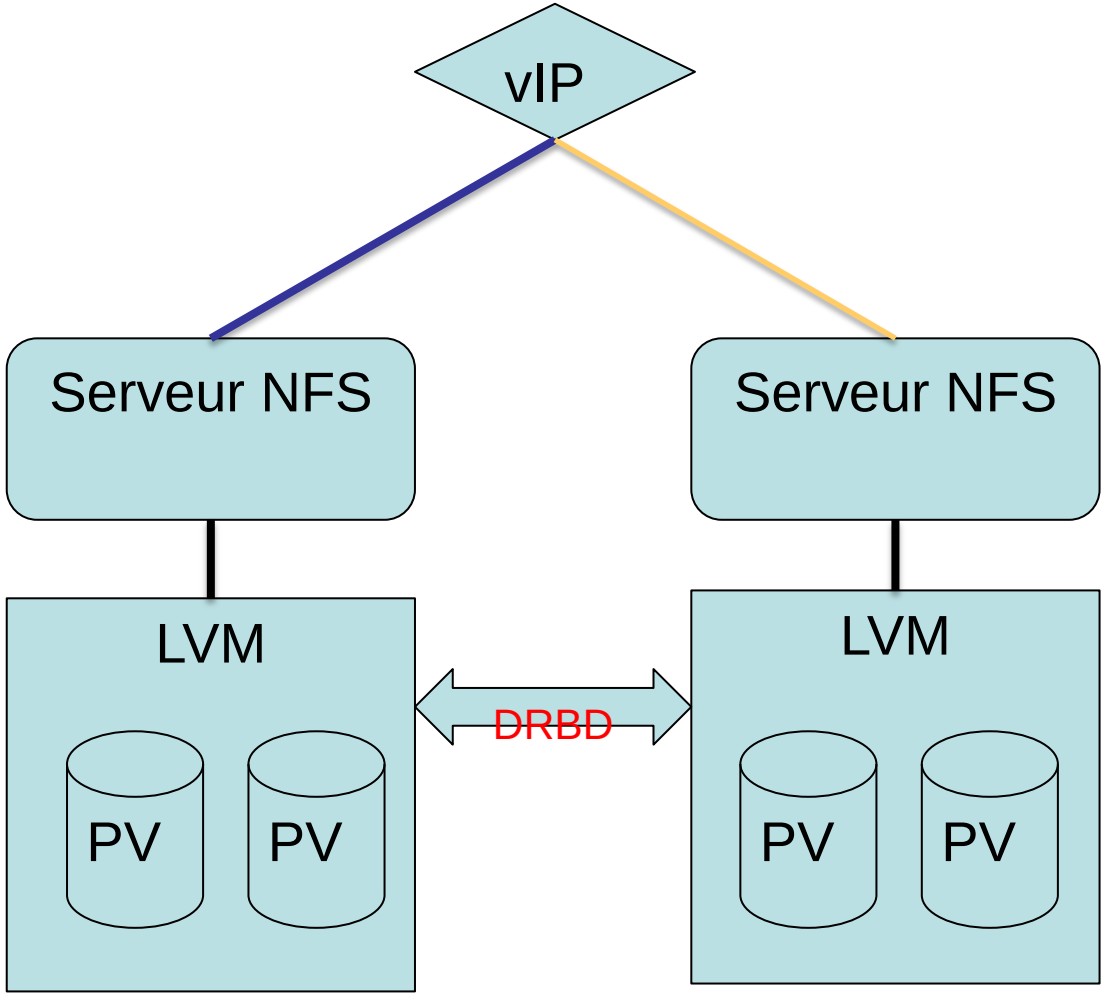
Redondance du stockage

- 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



Robert Schuman
UNIVERSITÉ DE STRASBOURG

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-lx1 {
              address 130.79.80.54:7790;
      }
      on eb-lx2 {
              address 130.79.80.59:7790;
      }
}
```

- Démarrage sur le serveur primaire

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

Robert Schuman
UNIVERSITÉ DE STRASBOURG

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)
```

Robert Schuman
UNIVERSITÉ DE STRASBOURG

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
```

Robert Schuman
UNIVERSITÉ DE STRASBOURG

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
```

Robert Schuman
UNIVERSITÉ DE STRASBOURG

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)

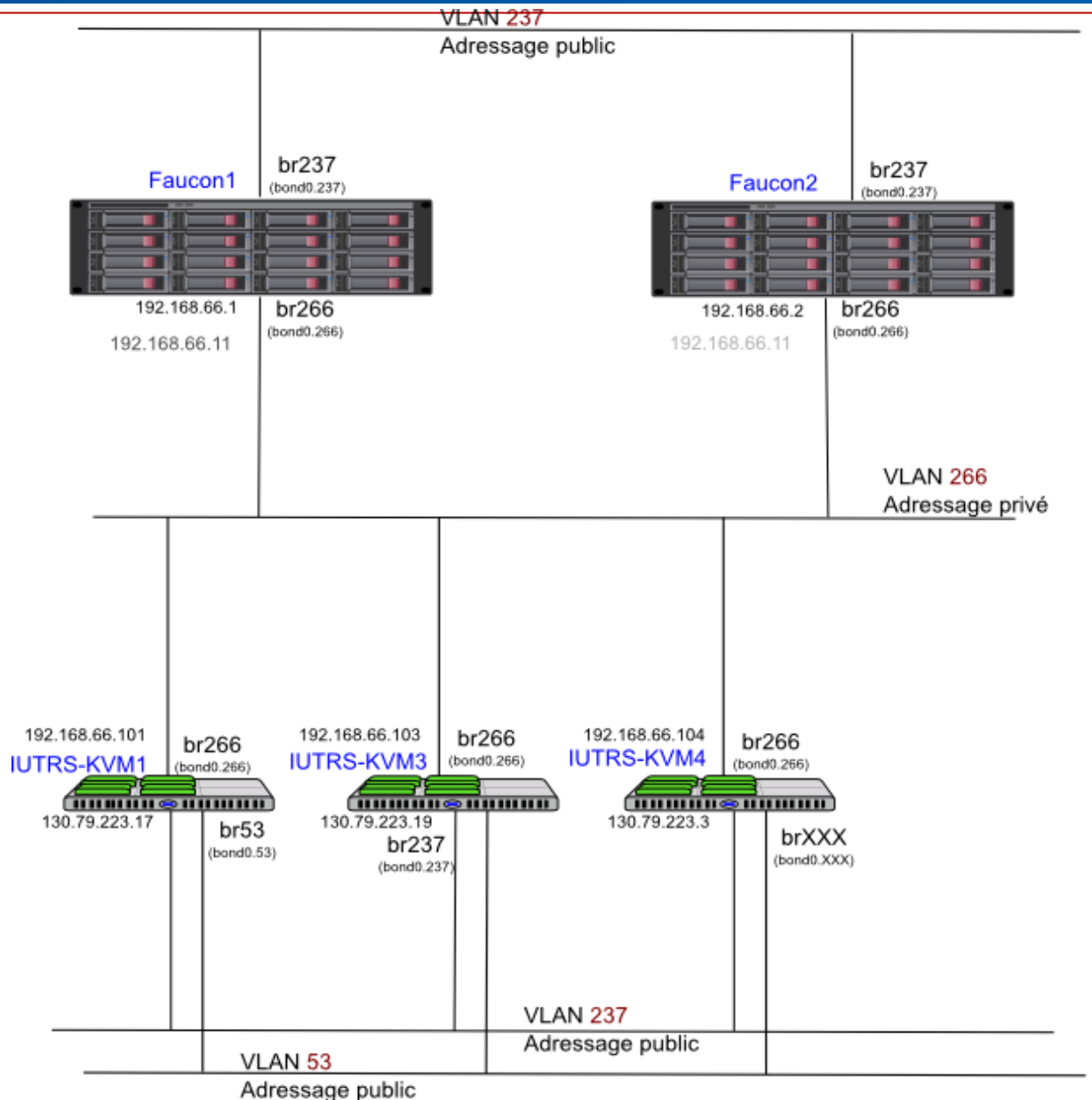
```
#sur chaque serveur
pvadd /dev/sdc1
vgextend storage /dev/sdc1
lvextend -L +500G /dev/storage/lvol0

#Sur le serveur actif
drbdadm  resize VMS
cat /proc/drbd # attendre la fin

ext3resize /dev/drbd1
```

Robert Schuman
UNIVERSITÉ DE STRASBOURG

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



Robert Schuman
UNIVERSITÉ DE STRASBOURG

Archipel

- 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.

iut Robert Schuman
UNIVERSITÉ DE STRASBOURG

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

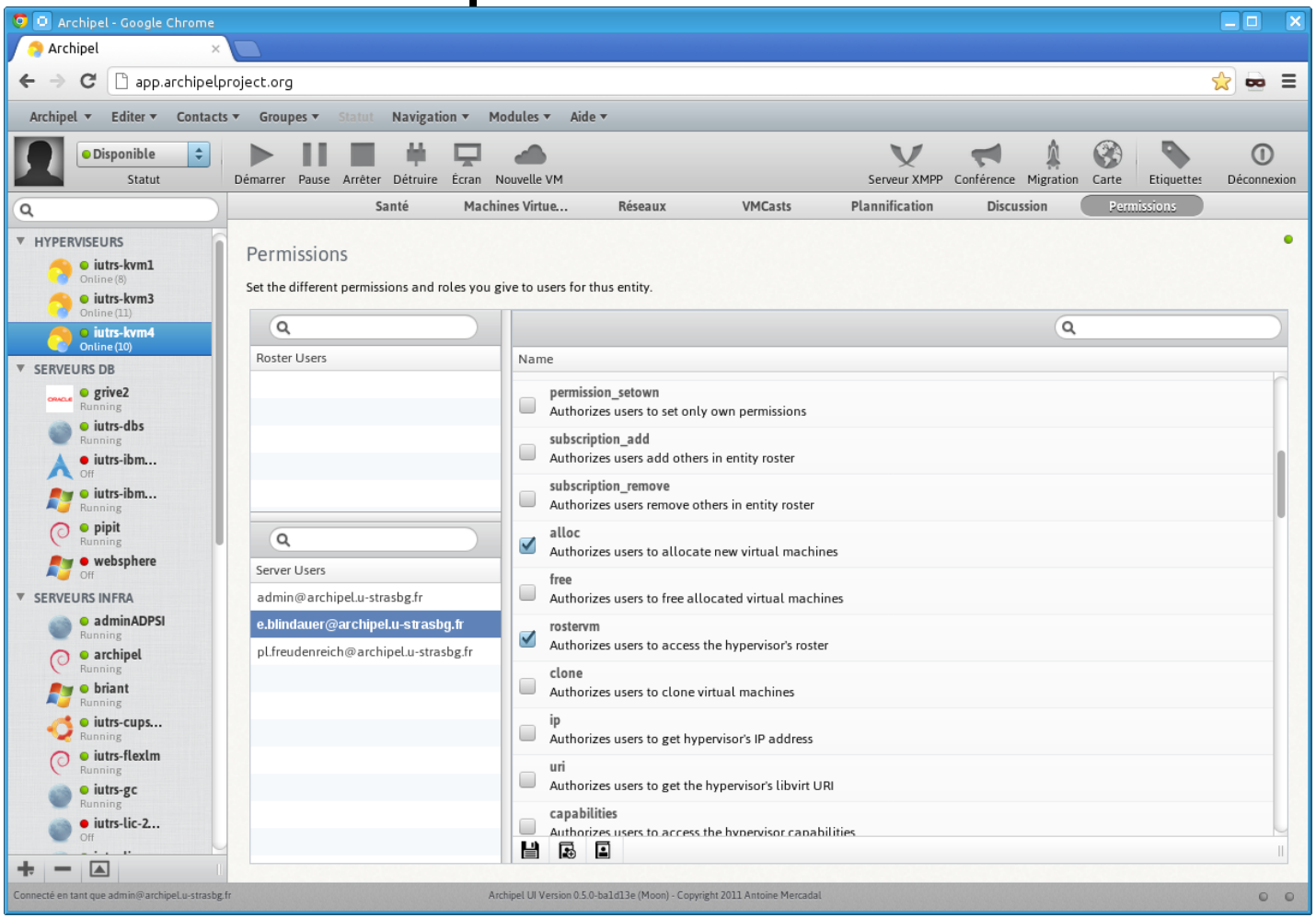
Robert Schuman
UNIVERSITÉ DE STRASBOURG

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

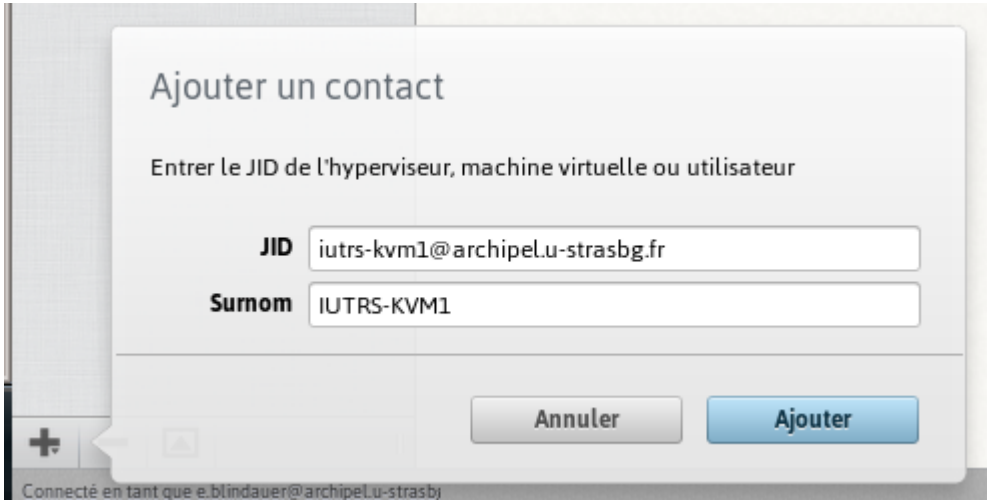
- Par GUI ou par CLI directement



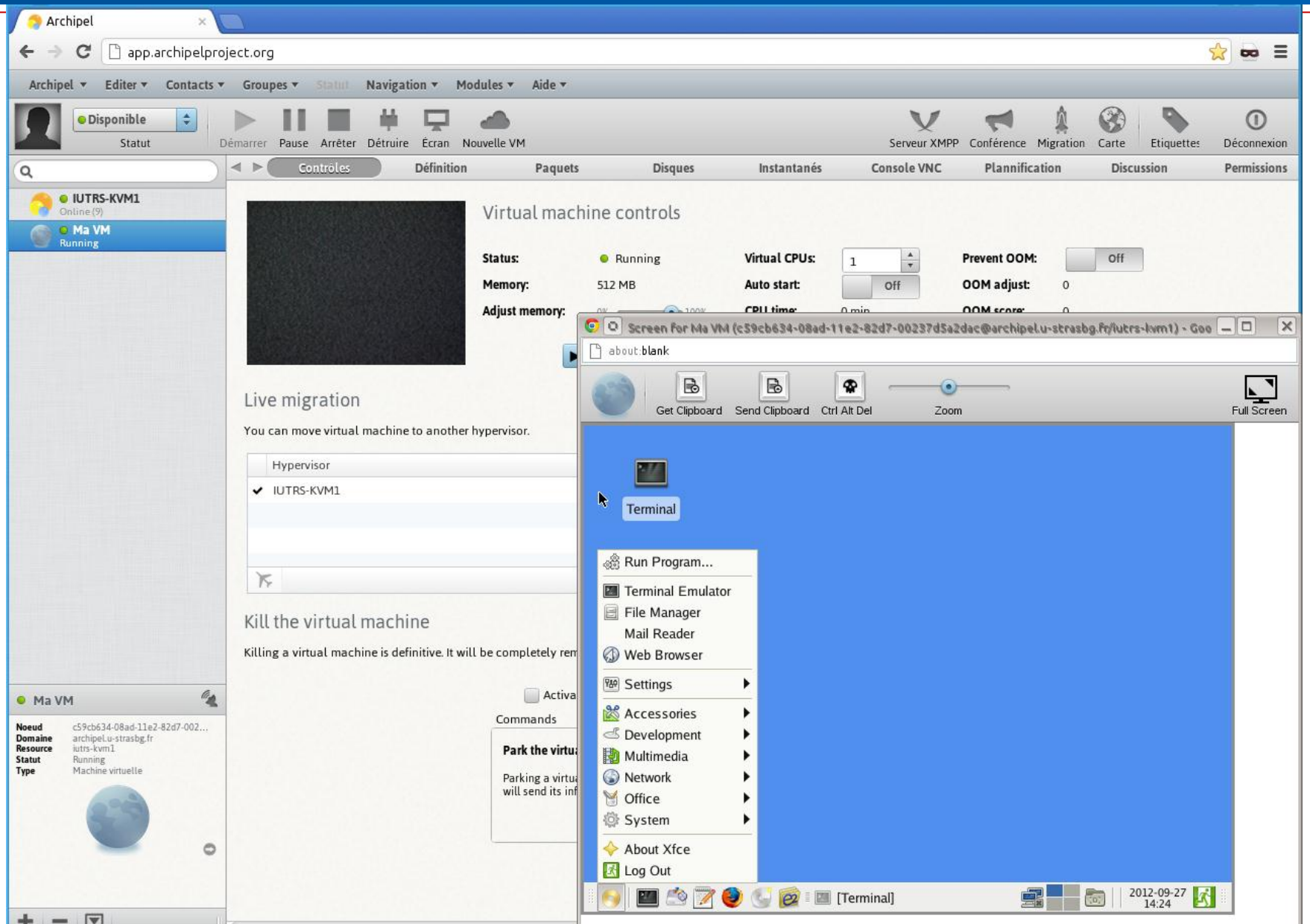
Robert Schuman
UNIVERSITÉ DE STRASBOURG

Action de l'utilisateur

- Ajouter l'hyperviseur auquel il a accès



Vue utilisateur



Robert Schuman
UNIVERSITÉ DE STRASBOURG

Conclusion

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