



ECOLE
Centrale
Nantes

Solution Haute Disponibilité pour Linux

Nicolas Schmitz

Ecole Centrale de Nantes

Nicolas.Schmitz@ec-nantes.fr

Décembre 2005

Introduction

- La haute disponibilité c'est notamment :
 - Doubler au maximum le matériel
 - Mettre en place les mécanismes de bascule
- Ces mécanismes doivent prendre en compte les données.

Introduction

- Plusieurs solutions « classiques » :
 - Équipement SAN : SPOF, coût
 - Rsync en cron : non intégrité des données

Introduction

- Solution « Heartbeat + DRBD » :
 - Pas de Single Point Of Failure
 - Données parfaitement à jour sur les deux machines
 - Peu coûteuse.



ECOLE
Centrale
Nantes

Description générale

Description générale

- Deux serveurs identiques (disposant chacun de ressources en disque suffisantes pour assurer le service.)
- Un seul serveur actif :
 - Adresse IP du service
 - Système de fichiers monté
 - Services lancés

Description générale

- Système de « battement de coeur » entre les deux machines.
(logiciel Heartbeat)
- Plus de battement de coeur du maître : l'esclave prend la main
: adresse ip, filesystem, services

Description générale

- Le système de fichiers que l'on monte sur l'esclave doit contenir exactement les mêmes données que celui du maître au moment du crash.
- C'est DRBD qui nous le garantit avec l'implémentation d'un raid1 sur IP.

Schéma global, en temps normal :

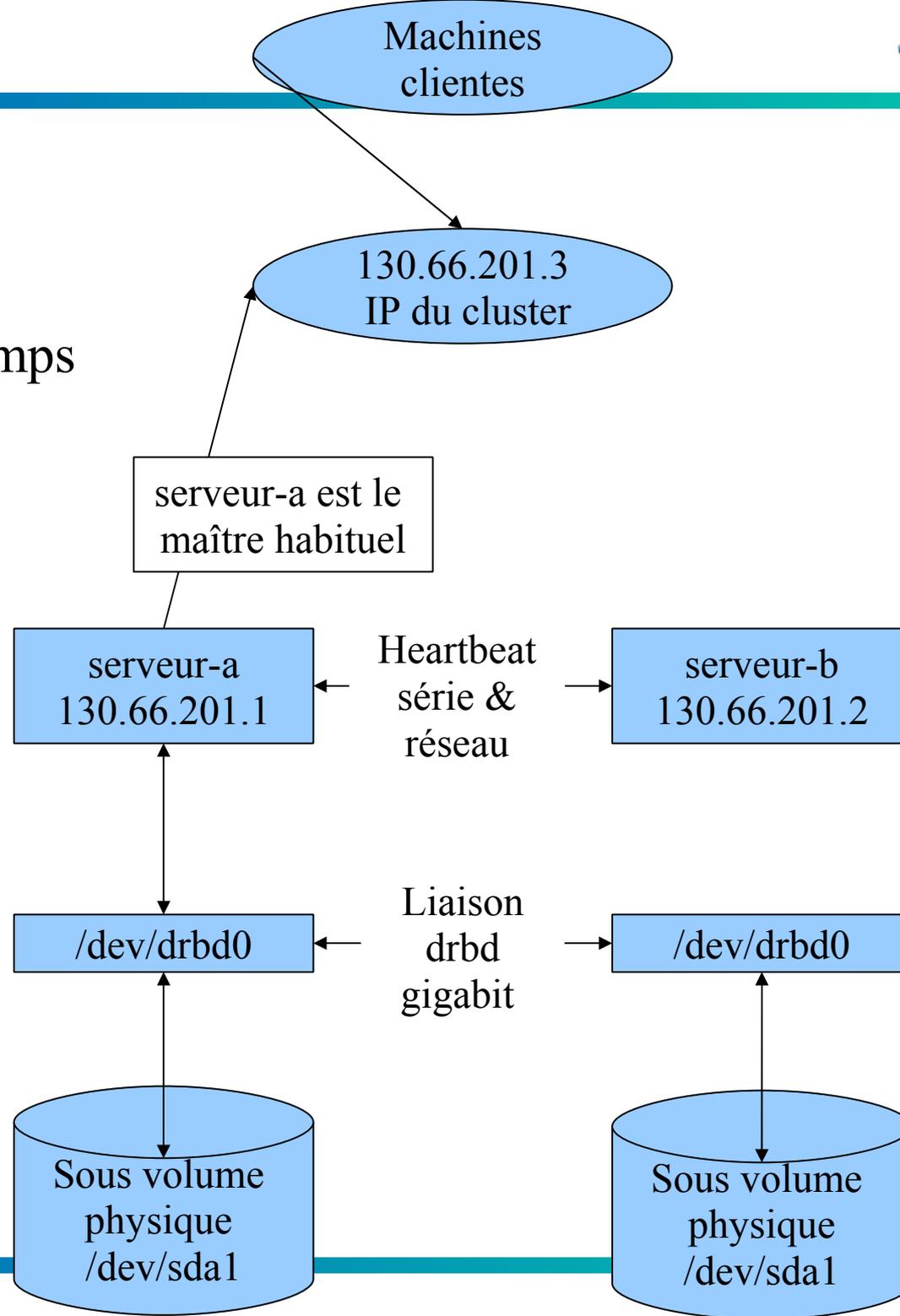


Schéma global, panne de
serveur-a, cluster dégradé :

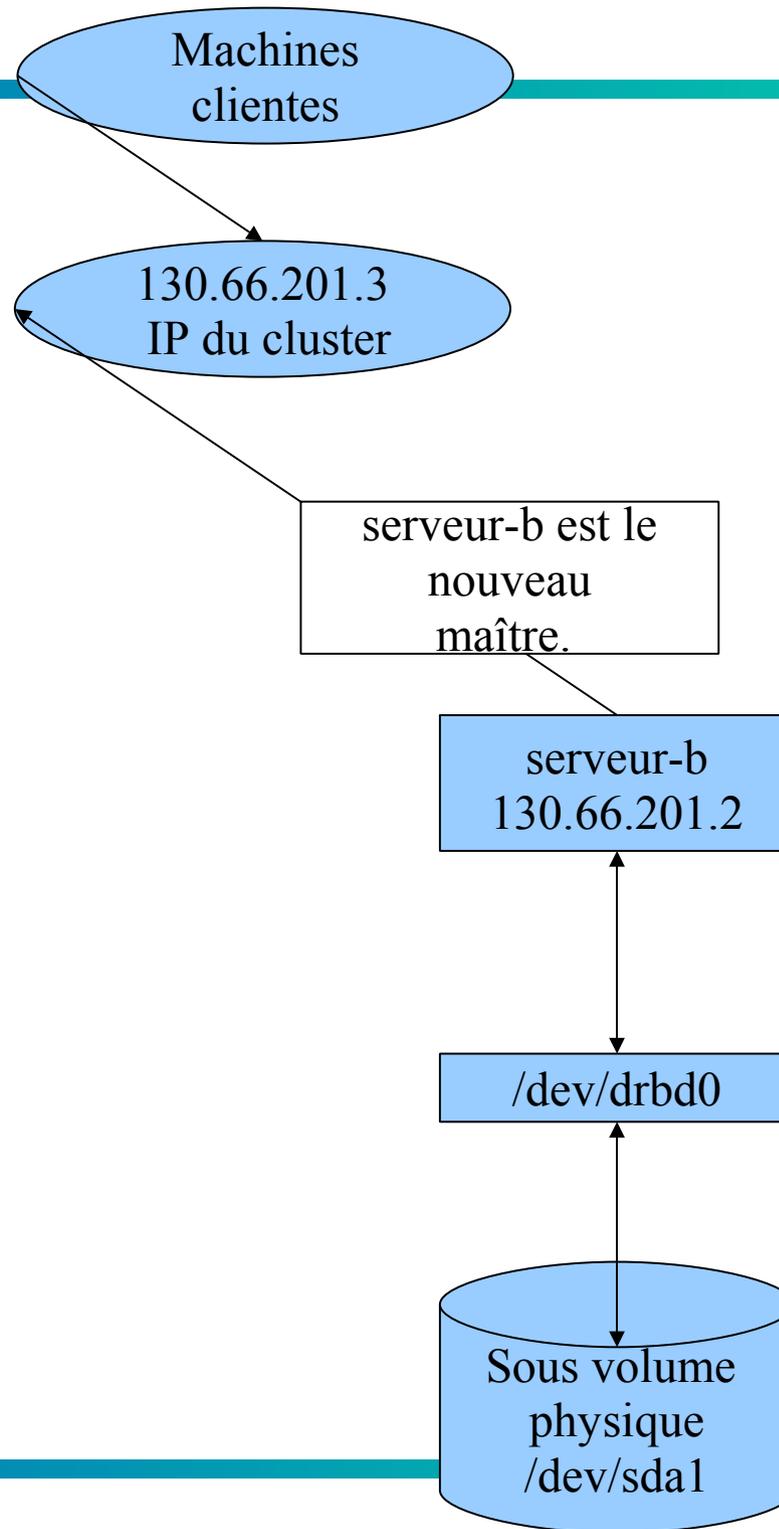
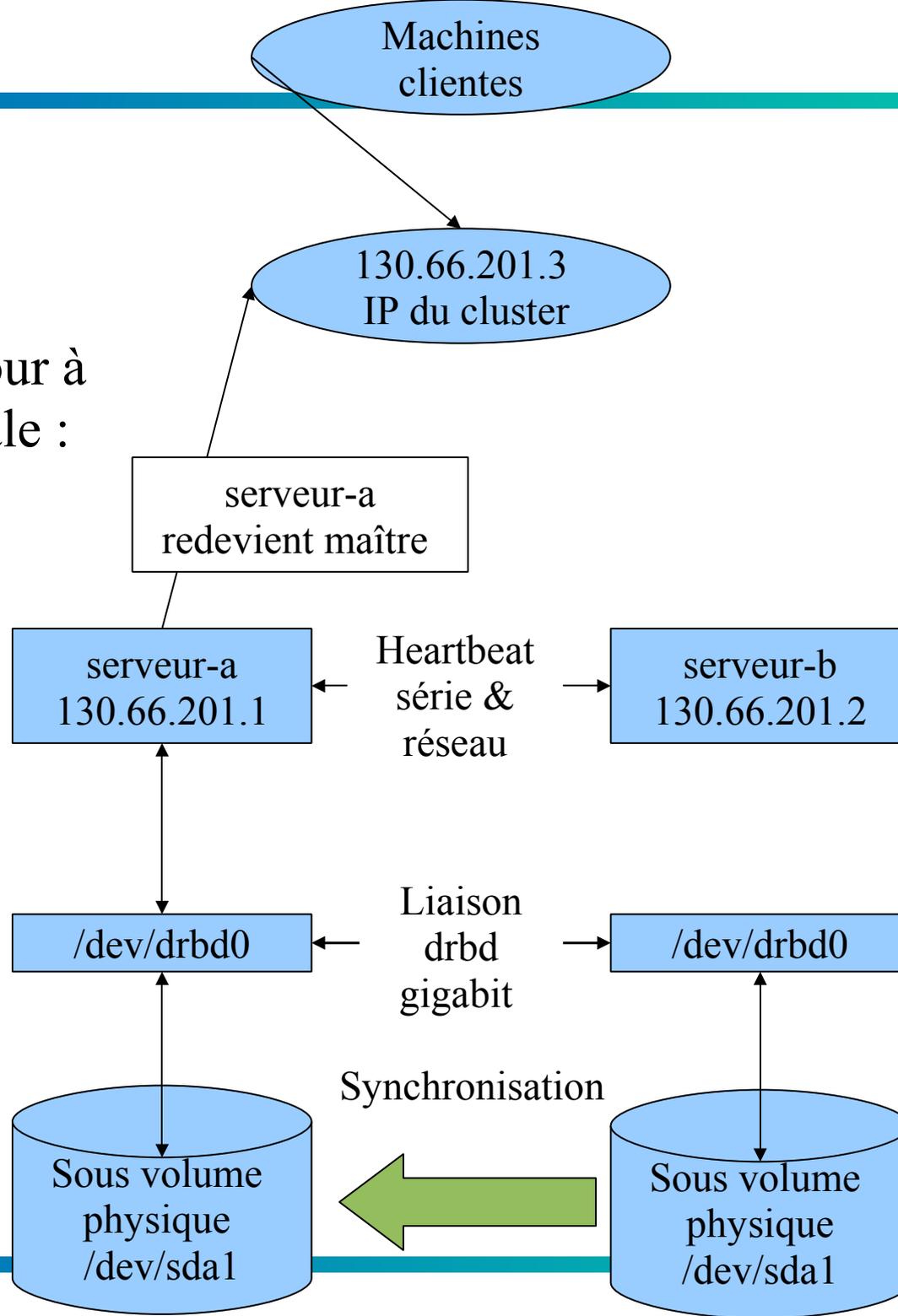


Schéma global, retour à
une situation normale :





ÉCOLE
Centrale
Nantes

Heartbeat

Heartbeat

- Heartbeat : le gestionnaire du cluster : c'est lui qui décide des actions nécessaires.
- Il gère une adresse ip virtuelle qu'il bascule entre les machines avec la technique du « gratuitous arp ».

Heartbeat

- Les deux noeuds échangent en permanence des battements de coeur.
- Toutes les décisions sont prises à partir de ces battements de coeur.
- => il faut fiabiliser cet échange en le faisant passer par plusieurs liens physiques (par exemple lien série + lien réseau).

Heartbeat

- Configuration : 3 fichiers de conf :
 - /etc/ha.d/ha.cf : configuration locale du noeud
 - /etc/ha.d/authkeys : méthode d'authentification des battements de coeur.
 - /etc/ha.d/haresources : configuration générale des ressources (ip, fs, services...) du cluster

Heartbeat

- **/etc/ha.d/haresources** : les ressources (ip, fs, services...) à rendre hautement disponibles.
- Une ligne par groupe de ressources.
- Cette ligne commence par le nom du « maître habituel » suivi des appels aux scripts dans /etc/ha.d/ressource.d ou /etc/init.d

Heartbeat

- Ces scripts seront appelés :
 - dans l'ordre où ils sont énumérés et avec l'argument « start » sur le noeud qui prend la main.
 - dans l'ordre inverse avec l'argument « stop » sur le noeud qui rend la main.

Heartbeat

- /etc/ha.d/haresources :

```
serveur-a drbddisk ::user Filesystem ::/dev/drbd0 ::/user ::ext3  
IPaddr2 ::130.66.201.3/24/eth0:0 cyrus mysql-debian MailTo::root
```

(malgré les apparences, ces éléments sont sur une seule ligne)

Heartbeat - Conclusion

- C'est simple et rapide à configurer.
- On peut l'utiliser sans DRBD :
 - serveur ftp en lecture seule, webmail
 - dns, replica ldap...
 - proxy imap, relai smtps

Heartbeat - Conclusion

- Version 2 en aout 2005, apporte en plus :
 - Gestion de plus de deux noeuds
 - Système de contrôle du fonctionnement des services



ECOLE
Centrale
Nantes

DRBD

DRBD

- DRBD se place entre le sous volume physique et le filesystem.
- Il présente un « block device » (/dev/drbd0) sur lequel on va pouvoir créer un filesystem dédié aux services à clusteriser
- Toutes les écritures sur ce filesystem seront répliquées sur le deuxième noeud du cluster

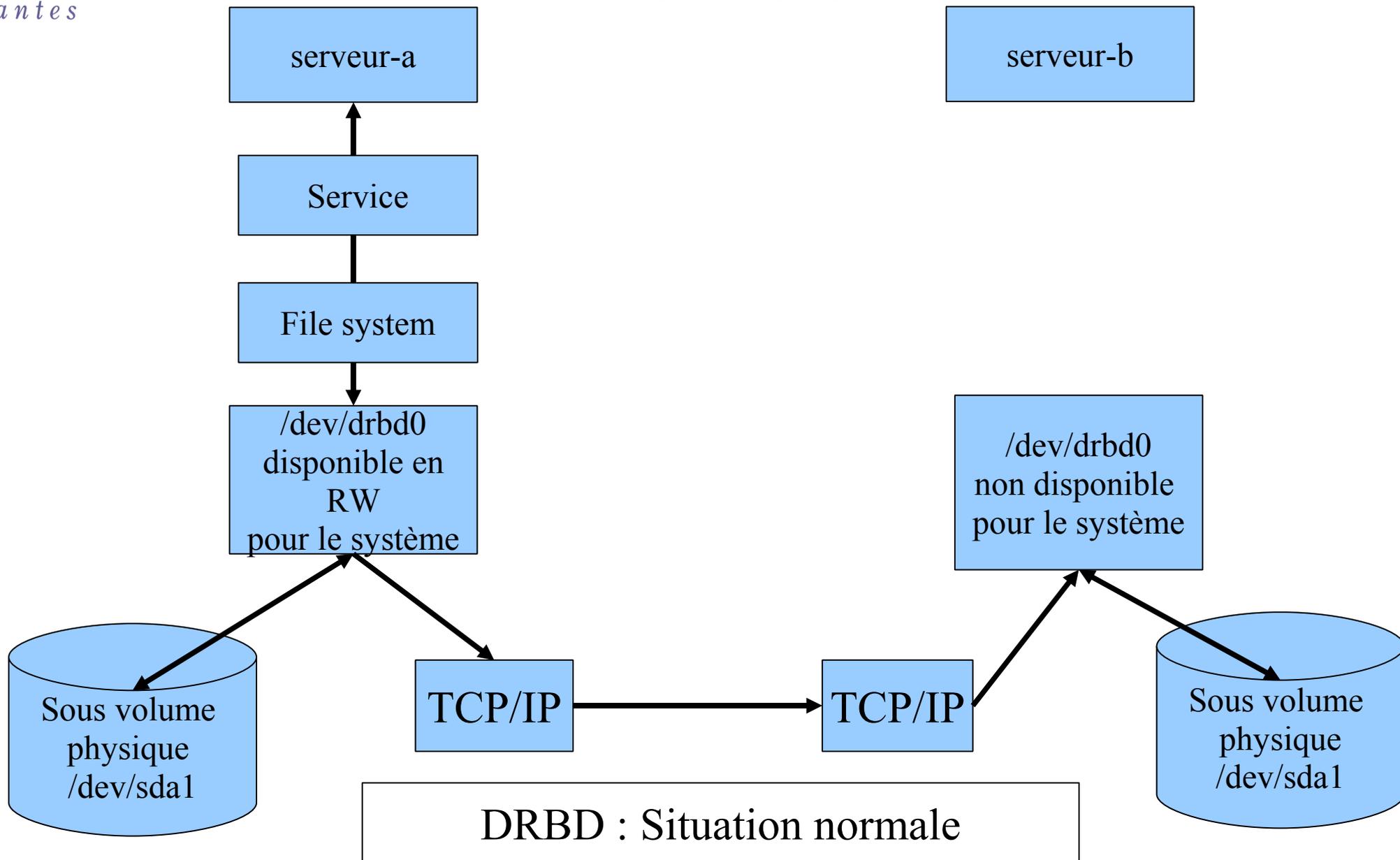
DRBD

- On choisit quand on envoie l'acquiescement d'écriture :
 - En protocole C, l'acquiescement n'est envoyé que lorsque les données ont été écrites sur les deux noeuds.

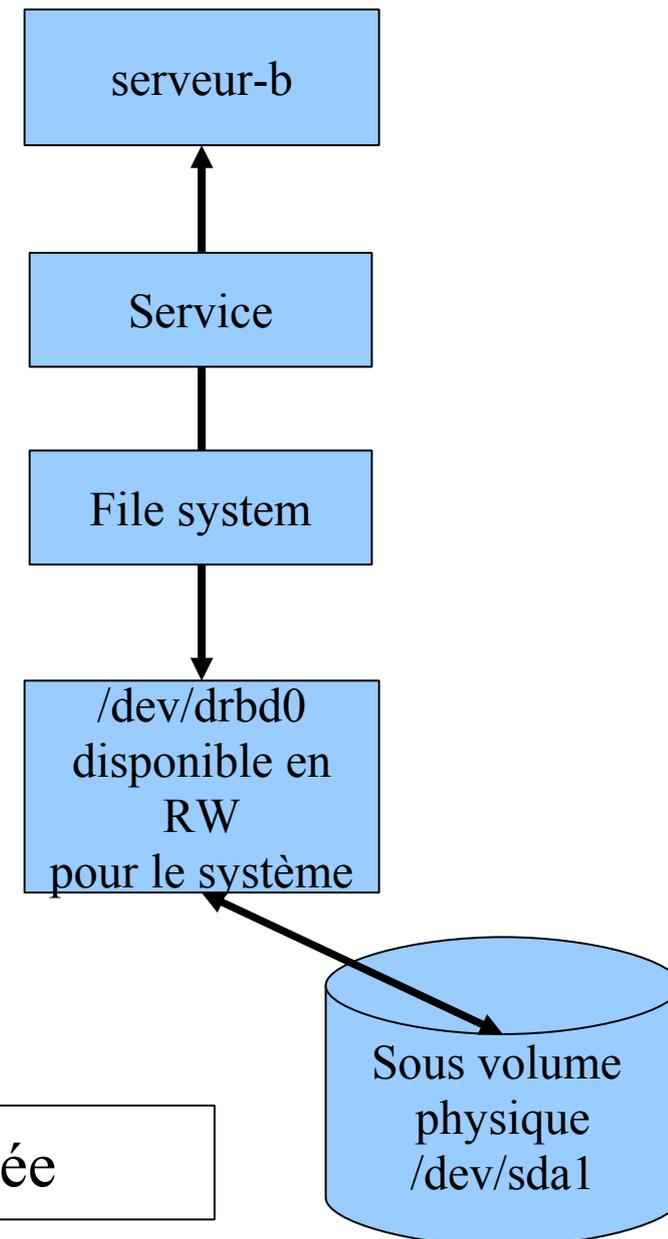
DRBD

- En cas de crash du maître, l'esclave dispose donc de toutes les données à jour, il peut rendre le service.
- Au retour du maître, synchronisation esclave => maître.
- Cette synchro est intelligente.

DRBD

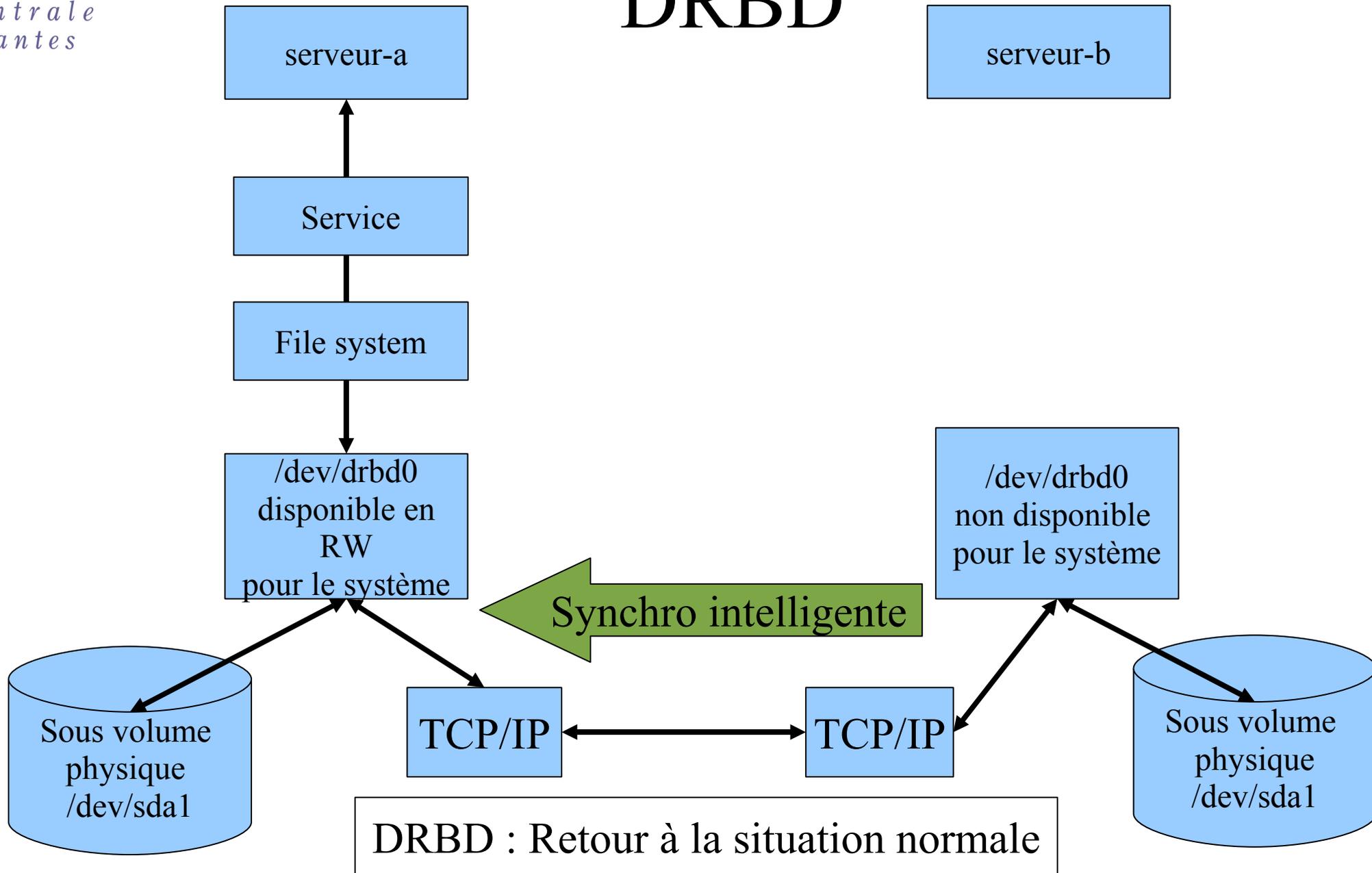


DRBD



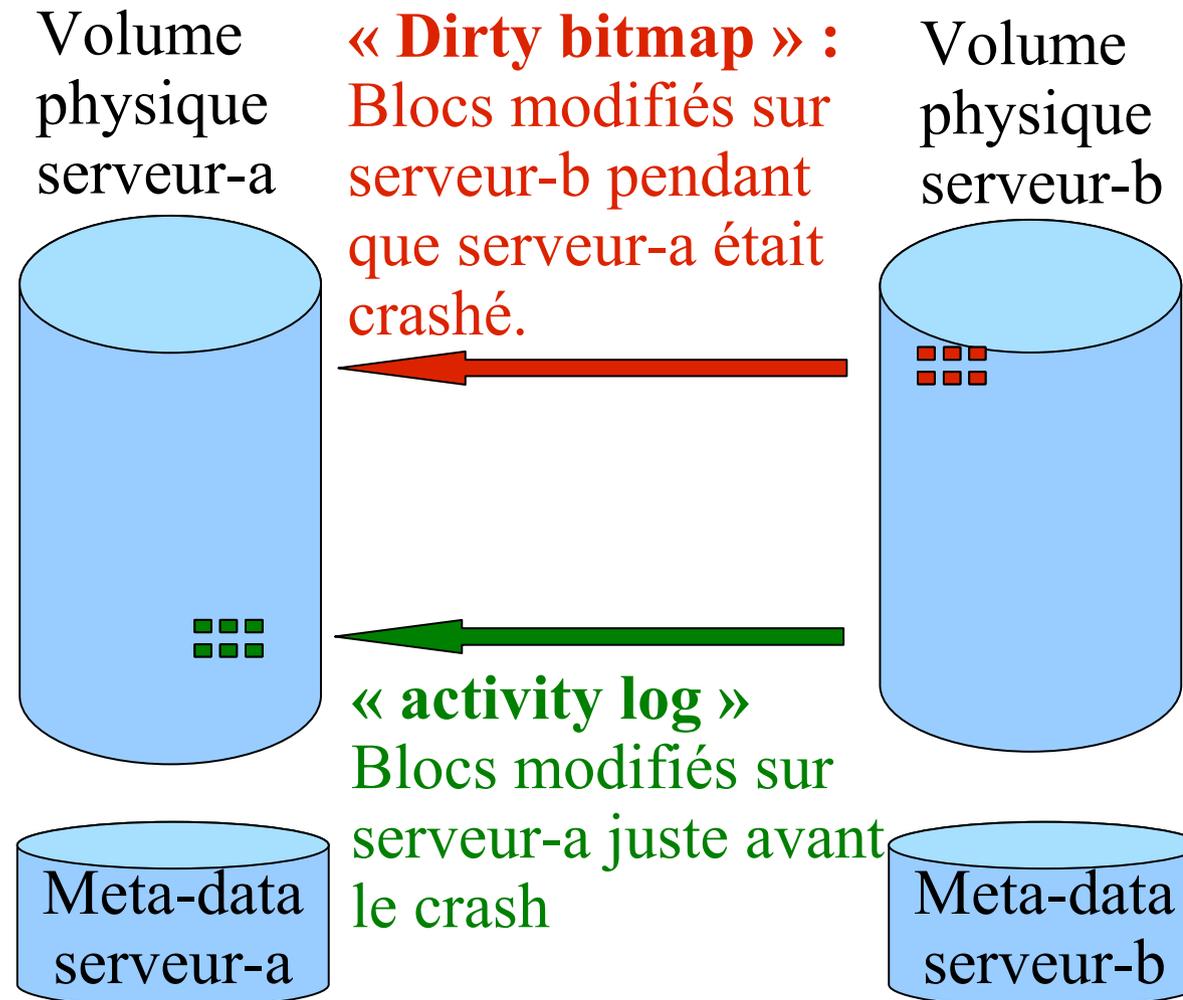
DRBD : Situation dégradée

DRBD



DRBD

Retour de serveur-a : synchro intelligente



DRBD

- Installation :
 - Une partie noyau : patch direct du noyau ou module
 - Outils de gestion : package.

DRBD

- Deux points de configuration qui influent sur les performances :
 - al-extents
 - meta-disk

DRBD

- Fichier de configuration `/etc/drbd.conf` identique sur les deux noeuds, ce qui facilite l'administration

DRBD

- Sur le maître :

```
mkfs.ext3 /dev/drbd0
```

- Mieux vaut choisir un filesystem journalisé

DRBD - Conclusion

- Grande robustesse du système.
- Impact sur les performances léger.
- Dans l'avenir, écriture simultanée sur les deux noeuds, et dépassement des 4Tera.



Intégration Heartbeat et DRBD

Intégration Heartbeat et DRBD

- Heartbeat dirige DRBD :

```
serveur-a drbddisk ::user Filesystem ::/dev/drbd0 ::/user ::ext3  
IPaddr2 ::130.66.201.3/24/eth0:0 cyrus mysql-debian MailTo::root
```

Intégration des services

- Ajout d'un service « par dessus » :
 - Script implémentant la commande « status »
 - Pointer vers le bon filesystem
 - Gestion du service dans /etc/ha.d/haresources
 - S'assurer que le service redémarre bien après un crash



ECOLE
Centrale
Nantes

Conseils - Astuces

Conseils - Astuces

- Gérer correctement les pannes de courant :
 - Onduleur avec remontée de statut.
 - Ordre d'extinction / rallumage à gérer

Conseils - Astuces

- Gestion des fichiers de conf :
 - Scp avec authentification par clef.
 - Autres outils : csync2, cfengine...
 - Placer les fichiers de conf sur la partition drbd

Conseils - Astuces

- Tester, tester, et tester encore avant mise en exploitation.
 - Une semaine/homme
 - Machine à état

Conseils - Astuces

- Support :
 - Listes de discussions de bon niveau
 - Support commercial directement par la SSII autrichienne qui maintient DRBD
 - Support commercial en France par SSLL

Conseils - Astuces

- Cluster actif/actif :
 - Deux partitions, deux ip
 - Répartition de serveurs virtuels
 - DRBD08 + GFS
 - Sans DRBD : Round Robin DNS et deux ip gérées par Heartbeat.



ECOLE
Centrale
Nantes

Conclusion

Conclusion

Les moins :

- Gestion du statut manquante

Les plus :

- Liberté de rendre hautement disponible n'importe quel service
- Solution fiable.
- Matériel standard donc peu coûteux.
- 100% libre