

# Support d'Infiniband sur Grid5000

*Modules Infiniband*

*Tests de performance*

*Enregistrement d'un environnement compatible*

---

Licence ASRALL Nancy-Charlemagne

*(Sébastien BADIA <seb@sebian.fr> <badia.seb@gmail.com> - Projet tutoré 2009-2010)*

24 mars 2010

## Table des matières

<b>1</b>	<b>Déploiement</b>	<b>2</b>
1.1	Upgrade Debian Lenny vers Testing . . . . .	2
1.2	Erreur de udev avec le noyau . . . . .	2
1.3	Drivers Infiniband . . . . .	2
1.4	Personnalisation . . . . .	3
<b>2</b>	<b>Tests de reconnaissance</b>	<b>3</b>
2.1	Ping Over Ib . . . . .	3
2.2	RDMA . . . . .	3
2.2.1	Perftests et Ibutils . . . . .	3

# 1 Déploiement

## 1.1 Upgrade Debian Lenny vers Testing

L'upgrade du système n'est pas obligatoire pour installer les drivers. Cependant certains outils ne sont pas disponibles dans Lenny, ou sont trop anciens...

L'upgrade est une mise à jour classique, pour ce tuto nous utiliserons un environnement *lenny-x64-base*, disponible dans la base des environnements de Grid5000.

Passage en *testing* :

Commencer par remplacer les occurrences de **lenny** dans le fichier */etc/apt/sources.list* par **testing**

```
1 node: apt-get update && apt-get dist-upgrade
2 node: apt-get install firmware-bnx2 firmware-linux-nonfree firmware-linux firmware-
    linux-free
```

**Note :** Vous pouvez répondre par défaut à toutes les questions lors du processus de Màj, l'installation de "firmware-linux-nonfree" règle des soucis de modules avec le passage au noyau 2.6.32

## 1.2 Erreur de udev avec le noyau

Si lors du **dist-upgrade**, cette erreur survient :

```
1 Errors were encountered while processing:
2 /var/cache/apt/archives/udev_150-2_amd64.deb
```

Vous pouvez alors exécuter cette commande :

```
1 node: dpkg --configure -a
```

Cela permet de configurer tous les paquets restés en attente suite à l'erreur de udev.

Afin de rebooter sur le noyau 2.6.32 et continuer la mise à jour, il enregistrer son environnement (tgz-g5k), puis préciser les nouveaux noyaux et initrd installés (2.6.32-trunk-amd64), dans le *.env*, ou le *.dsc*. Redéployer avec le nouveau *.env*, ou le *.dsc* et relancer la commande de **dist-upgrade** pour terminer les mises à jour (Cette opération de redéploiement est apparenté à un reboot sur un pc classique).

On peut maintenant enregistrer son environnement <sup>1</sup> sur le *frontend*, afin de conserver un système upgradé propre, le *.env* n'a plus besoin d'être modifié.

## 1.3 Drivers Infiniband

Infiniband est entièrement intégré et utilisable sur les noyaux récent de linux, il vous faut donc juste activer certains modules.

Deux options s'offrent alors à vous, soit vous activer le module avec un *modprobe* après chaque démarrage du système.

Soit vous ajouter ces modules dans le fichier */etc/modules*, ceux-ci seront ainsi chargé au démarrage du système.

- mlx4\_ib Module supportant la plus part des cartes Mellanox Infiniband
- rdma\_ucm
- ib\_umad
- ib\_uverbs
- ib\_ipoib Module pour le support de ip over infinband
- ib\_srp
- bnx2

```
1 node: modprobe mlx4_ib && modprobe rdma_ucm && modprobe ib_umad
2 node: modprobe ib_uverbs && modprobe ib_ipoib && modprobe ib_srp && modprobe bnx2
```

1. Réf : [https://www.grid5000.fr/mediawiki/index.php/Deploy\\_environment](https://www.grid5000.fr/mediawiki/index.php/Deploy_environment) (section **Create a new environment from a customized environment**)

## 1.4 Personalisation

Changer le mot de passe root (default = grid5000) :

```
1 node: passwd
```

Créer votre utilisateur <sup>2</sup> (exécuter la commande `id` sur un *frontend* afin de récupérer les informations à fournir ; Le nom du groupe correspond théoriquement à la ville d'inscription) :

```
1 node: addgroup --gid <GID> <NOM_GROUPE>
2 node: adduser --uid <UID> --ingroup <NOM_GROUPE> <LOGIN>
```

Enregistrer l'environnement <sup>3</sup>.

## 2 Tests de reconnaissance

```
1 node: ibv_devinfo
2 node: ibstat
```

### 2.1 Ping Over Ib

Communication entre deux machines (Ping Over Ib)

Sur la première machine on va déterminer le guid de la carte pour que la deuxième machine puisse "ping" :

```
1 node: ibstat -p
```

Sur la seconde :

```
1 node: ibping -G <guid-hote-vise>
```

### 2.2 RDMA

Remote Direct Memory Access sur Infiniband (Bande Passante)

Sur la première machine :

```
1 node: ib_rdma_bw
```

Sur la seconde :

```
1 node: ib_rdma_bw <Node1>
```

#### 2.2.1 Perftests et Ibutils

Les Paquets Perftest et Ibutils, apportent un panel de commandes de tests supplémentaires, les classiques test de ping-pong mais aussi bien d'autres.

2. Réf : [https://www.grid5000.fr/mediawiki/index.php/Deploy\\_environment](https://www.grid5000.fr/mediawiki/index.php/Deploy_environment) (section **Customization**)

3. Réf : [https://www.grid5000.fr/mediawiki/index.php/Deploy\\_environment](https://www.grid5000.fr/mediawiki/index.php/Deploy_environment) (section **Create a new environment from a customized environment**)