

Installation d'un Cluster

Différentes approches

Romaric DAVID
david@unistra.fr

Pôle HPC, Direction Informatique
13/09/2011

changes
espiritualidad
insertion
perspectives
mutualisation
reussite
ouverture
fondation
CHEMISTRY
equation
biology
 $E = mc^2$
RECHERCHE
SYNERGIES
COMPETENCES
pi
TECHNOLOGY
doctorat
cosmopolite
ENSEIGNEMENT SUPÉRIEUR
biotechnologies
axiome
mécanique
management
capitale
droit
excellence
savoirs
wissenschaft
bibliothèques
médecine
tesis
théologie
gravitation
idéaux
connaissances
musica
langage
INTERNATIONAL
solution
HEURISTIQUE
partenariats
HISTOIRE
physique
mécanique quantique
insertion
PLURIDISCIPLINARITÉ
sciences
gravitation
humain
molécule
ambition
quantique
MASTER
cultures
NETWORK

- Pôle HPC = centre de calcul de l'Université de Strasbourg
- ▶ Support applicatif de haut niveau
 - ▶ Développement d'application et formation à destination des chercheurs
 - ▶ Veille technologique sur les outils logiciels et matériels



- ▶ Mise à disposition de ressources de calcul orientés HPC
 - Dans ce cadre, il nous est nécessaires de savoir administrer et gérer des systèmes d'exploitation
 - Besoin constant d'automatisation et de simplification
 - Nos méthodes et outils doivent permettre la gestion de plus de 100 serveurs

- ▶ A l'heure actuelle, nos ressources de calcul HPC sont financées directement par les labos de recherche
- ▶ Achats de machines **au coup par coup** (Asus, Dell, HP, IBM, Supermicro)
- ▶ Nous intégrons les machines dans notre salle machine
- ▶ Nous y installons notre liste de logiciels HPC



- ▶ Clusters : quels systèmes d'exploitation ?
 - Distributions, génériques, spécialisées
- ▶ Quelle base matérielle commune ?
 - Ce qu'il faut penser à demander
- ▶ Installation d'un cluster
 - Déploiement d'images (bloc, fichier)
 - Installations automatisées
 - Pré/Post Installs, rocks, xcat2
- ▶ Conclusion

- ▶ Clusters : quels systèmes d'exploitation ?
 - Distributions, génériques, spécialisées
- ▶ Quelle base matérielle commune ?
 - Ce qu'il faut penser à demander
- ▶ Installation d'un cluster
 - Déploiement d'images
 - Installations automatisées
 - Pré/Post Installs, rocks, xcat2
- ▶ Conclusion

► Pour construire un cluster prenez :

- Un serveur baptisé "maître" 
- Un ensemble de serveurs pour le calcul : "noeuds"



- Un réseau Ethernet pour les relier



- Un réseau Infiniband pour du HPC

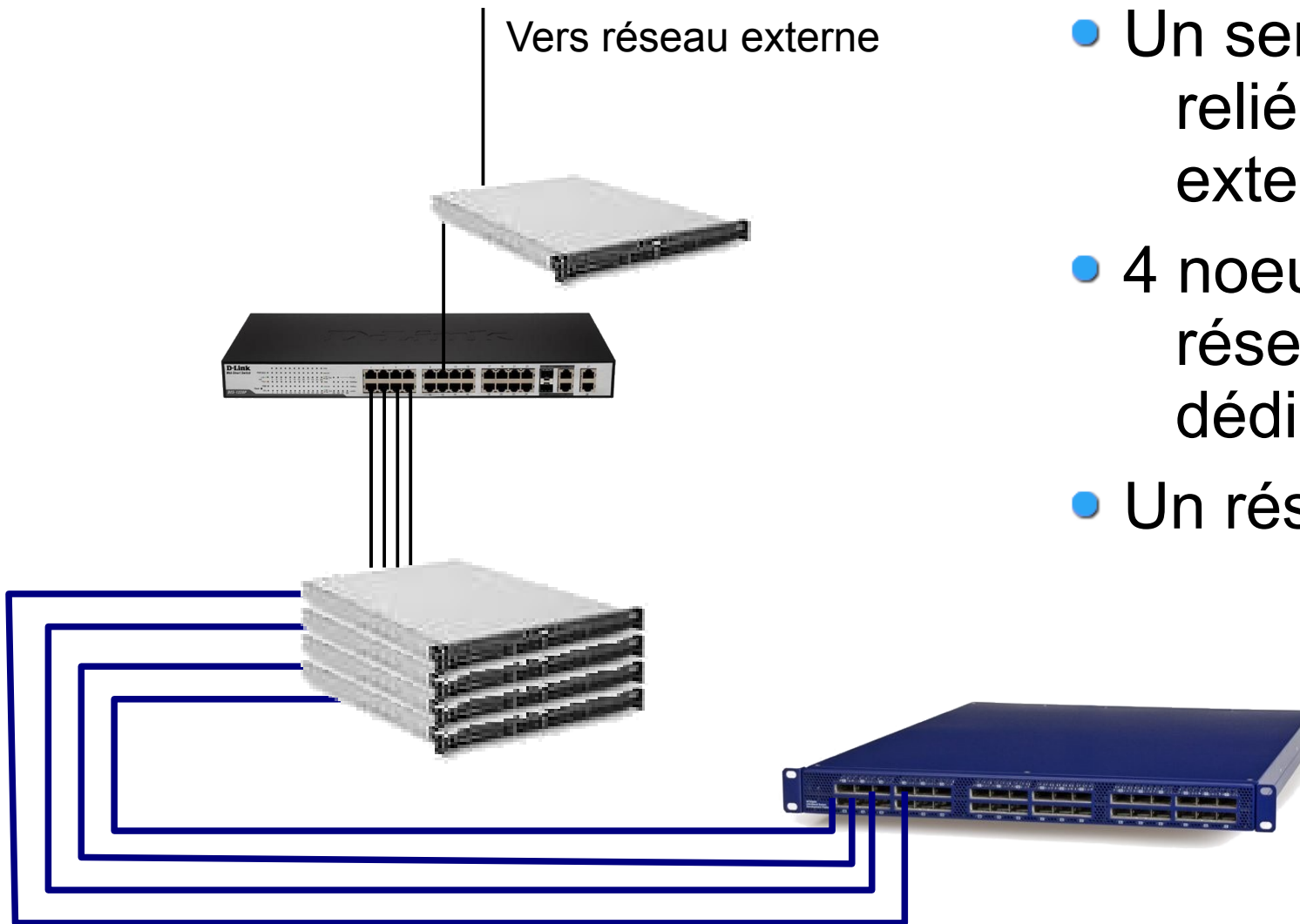


► Assemblez le tout

► Une fois assemblé

► Sur ce cluster

- Un serveur maître relié au réseau externe
- 4 noeuds sur un réseau isolé et dédié
- Un réseau IB



- ▶ Le système d'exploitation (GNU)/Linux est largement utilisé dans le monde du calcul scientifique
- ▶ Une grande part de ce cours y fait référence
- ▶ Le système GNU/Linux est :
 - gratuit
 - multi-utilisateurs : plusieurs utilisateurs peuvent accéder simultanément au système
 - temps partagé : De manière cyclique, les programmes accèdent au processeur qui les fait progresser. Le reste du temps, ils sont inactifs.

- ▶ Le nom Linux vient de la contraction du nom du premier développeur du noyau, Linus Torvald, et de Unix
- ▶ Une **distribution** Linux est un ensemble cohérent noyau + programmes
- ▶ Une distribution comporte ses propres :
 - mécanismes d'installation
 - mécanismes de mise à disposition des programmes :
le **packaging**
 - variations d'arborescence
 - fichiers de configuration

- ▶ Il existe un grand nombre de distributions Linux, certaines dérivant les unes des autres :
http://fr.wikipedia.org/wiki/Distribution_Linux
- ▶ Voici les principales

Nom	Année d'Origine	Paquets	Usages	Orientation	Distributions dérivées
Debian	1993	apt	Tous	Logiciel Libre	Ubuntu
RedHat	1994	rpm	Serveurs	Entreprise. Payante	Centos, Fedora, Scientific Linux
Slackware	1993	pkgtool	Tous		
Suse	1994	rpm	Tous	Entreprise. Payante	OpenSuse

- ▶ Comment choisir sa distribution ?
 - De la même manière que l'on choisit un OS : en fonction des applications qui y tournent
 - Par exemple : vérifier les applications packagées pour une distribution particulière
 - Version des outils (en particulier développement) disponibles
- ▶ Certaines applications commerciales ne bénéficient d'un support que sous RedHat
 - Fonctionneront sous distributions dérivées de RedHat, mais en perdant le support et au prix d'un éventuel travail d'intégration

Exemple : configuration d'une interface réseau

► Distribution Ubuntu 9.10

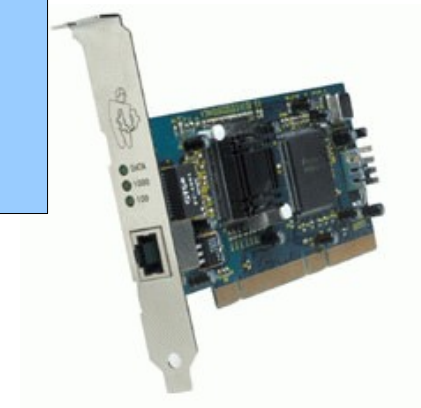
```
# The primary network interface
auto eth0
iface eth0 inet dhcp
```

Fichier /etc/network/interfaces

► Distribution Centos 5.3

```
DEVICE="eth0"
ONBOOT="yes"
BOOTPROTO="dhcp"
```

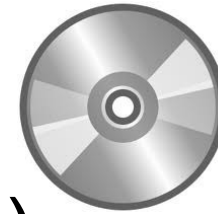
Fichier /etc/sysconfig/network-scripts/ifcfg-eth0



- ▶ Par leur structure interne, les distributions Linux diffèrent entre elles par :
 - l'organisation des répertoires
 - leurs outils de configuration
 - leur système de packaging
 - Pour autant, elles offrent plus ou moins les mêmes fonctionnalités
- ▶ Certaines applications ne sont pas packagées pour toutes les distributions
 - Cela constitue un critère de choix prépondérant

- ▶ Un **dépôt logiciel** désigne un emplacement en réseau ou en local (DVD, Disque dur) où des paquetages logiciels sont disponibles
 - En réseau : utile pour les mises à jour (sécurité) des logiciels
 - En local : souvent utilisé lors de l'installation
- ▶ Les dépôts logiciels en réseau:
 - Peuvent être indépendants des éditeurs de la distribution (par exemple, Personal Package Archives d'Ubuntu, karan.org pour Centos)
 - Ajoutent quantité de logiciels à la distribution de base

- ▶ Petit rappel du processus d'installation d'une distribution sur un ordinateur
- ▶ Ingrédients : un support d'installation :
- ▶ Un ordinateur (ici un serveur de calcul)

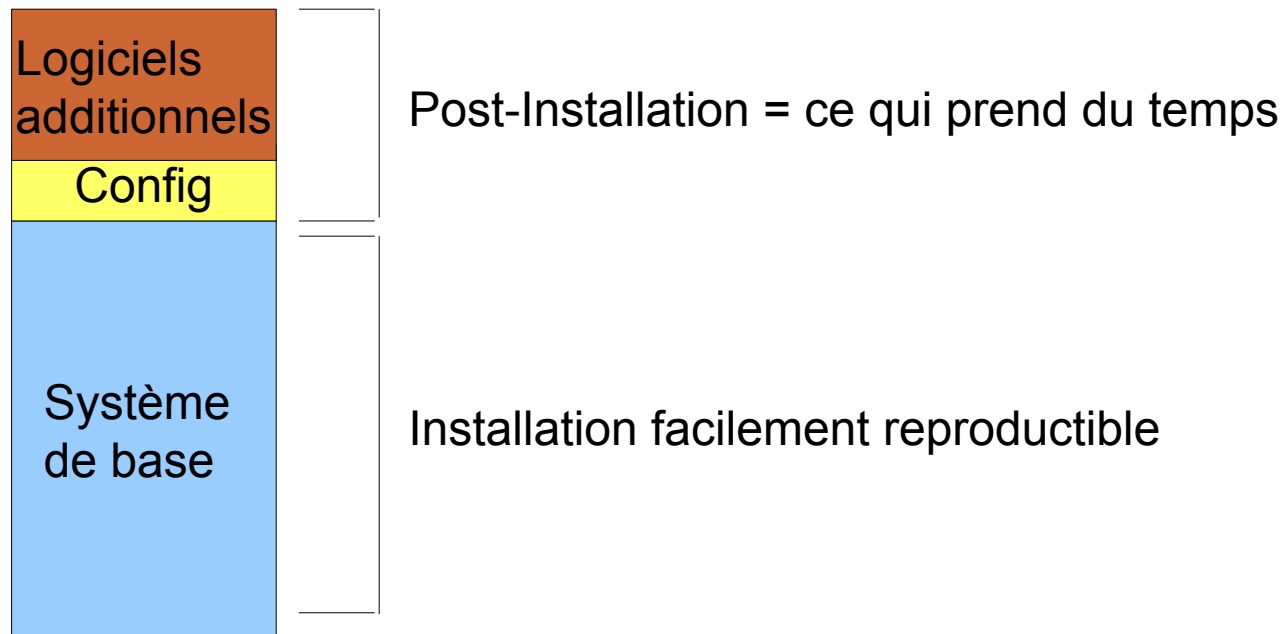


- ▶ Un peu de temps...
- ▶ Comment fonctionne l'installation ?

- ▶ Amorçage (boot) sur le support d'installation
- ▶ Lancement d'un système dédié à l'installation qui :
 - **Partitionne** le disque dur. Pour le calcul, on peut prévoir une partition de *scratch*, contenant les données temporaires des codes de calcul
 - Installe les **paquets** sur le disque dur
 - Installe le noyau et configure le démarrage de l'OS
- ▶ Redémarrage du système ainsi installé et boot sur le disque dur
- ▶ Pré-requis : le 1er support de boot recherché par l'ordinateur ne doit pas être le disque dur (en pratique ce sera souvent le dernier)

- ▶ Une série de questions sont posées pendant l'installation
 - Ces questions peuvent provenir des scripts d'installation présents dans les paquets
- ▶ Des étapes de post-installation, de configuration spécifique sont indispensables.
 - Configuration avancée non prévue dans le process d'installation des paquetages
 - Certains logiciels ne sont pas packagés
- ▶ Des mécanismes d'automatisation sont propres à chaque distribution (kickstart sous Redhat, preseeds sous Debian)

- ▶ Un système installé et opérationnel = le système de base + la configuration



- ▶ Tous les mécanismes d'installation automatique prévoient des mécanismes de post-installation *scriptés* et spécifiques

- ▶ Sur une machine virtuelle équipée de 2 Interfaces réseau et de 40 GO de disque dur, nous installons un système Linux dédié aux clusters
- ▶ Nous installons un serveur qui aura un rôle central dans le cluster (et qui resservira dans la suite)
- ▶ Remarquez les phase de partitionnement du disque et les questions posées.

- ▶ Nous avons vu le principe générique permettant d'installer une distribution Linux sur une machine
- ▶ Nous allons maintenant mettre en pratique sur un cluster :
 - Quelles spécificités ?
 - Quels logiciels sont utilisés ?

- ▶ Un cluster sous Linux est un ensemble de serveurs se répartissant des tâches de calcul
 - Un de ces serveurs joue un rôle particulier (accès aux fichiers, lancement des tâches de calcul) : il est appelé noeud maître.
 - Les autres sont munis du même modèle de système répliqué n fois : les noeuds de calcul
- ▶ Des services différents tournent sur maître et noeuds. Par exemple :
 - Agent de monitoring sur les noeuds de calcul
 - Serveur d'annuaire sur le noeud maître
- ▶ En général, les noeuds de calcul sont administrés à partir du noeud maître

- ▶ Quelle est la différence entre un ensemble de serveurs et un cluster ?
 - La configuration du serveur maître
 - La configuration des noeuds de calcul
- ▶ Un cluster, c'est un peu de Linux et beaucoup de configuration mais...
 - Des configurations toujours identiques
 - Quelques astuces pour se simplifier la vie
 - Les astuces principales suivent...

- ▶ Il faut centraliser la configuration
- ▶ Pour cela, on fait appel en particulier à quelques services utiles, sur le mode client-serveur
 - Dynamic Host Configuration Protocol (DHCP) :
 - Attribution automatique d'adresse IP, en fonction de l'adresse physique unique identifiant les carte réseaux, **donc les machines**
 - Fournit également des paramètres de configuration liés à l'interface réseau et au boot réseau
 - Partage de fichiers :
 - Nous utilisons beaucoup NFS, Network File System
 - Il existe d'autres protocoles

- ▶ Pour répondre aux besoins des utilisateurs, il est parfois nécessaire de :
 - faire cohabiter plusieurs logiciels offrant les mêmes fonctionnalités (compilateurs commerciaux, libres)
 - plusieurs versions des mêmes logiciels
- ▶ Pour cela, on utilise la commande *module* qui manipule l'*environnement* des utilisateurs
 - <http://modules.sourceforge.net>

- ▶ L'administration des noeuds de calcul se fait à partir d'un logiciel spécifique intégrant un ensemble de logiciels de plus bas niveau
- ▶ On y retrouve les fonctionnalités suivantes :
 - installation et mise à jour automatisée des OS
 - exécution de commandes à distance
 - inventaire des machines
 - allumage/extinction électrique des noeuds à distance
 - intégration de gestionnaires de batch

- ▶ Clusters : quels systèmes d'exploitation ?
 - Distributions, génériques, spécialisées
- ▶ Quelle base matérielle commune ?
 - Ce qu'il faut penser à demander
- ▶ Installation d'un cluster
 - Déploiement d'images (bloc, fichier)
 - Installations automatisées
 - Pré/Post Installs, rocks, xcat2
- ▶ Conclusion

- ▶ Dans un cluster il y a un certain nombre de machines
- ▶ Ces machines ne sont pas toujours à côté de notre bureau
- ▶ Comment pouvons-nous simplifier des tâches répétitives comme :
 - Nous déplacer dans une salle pour allumer une machine ?
 - Brancher un clavier ou un écran sur une machine ?
- ▶ Réponse : les constructeurs de matériel ont (un peu..) pensé à nous !

- ▶ Certaines caractéristiques matérielles des serveurs sont très pratique pour faciliter l'intégration dans un cluster et simplifier l'administration :
 - Présence d'une carte d'administration - Baseband Management Controller : BMC
 - Pour les serveurs rackables, report des connectivités USB/Ecran en façade
- ▶ Certaines sont optionnelles et peuvent constituer des éléments de coût superflus :
 - Double-alimentation
 - Lecteurs de DVD, de cartes mémoires

- ▶ Une BMC est un **dispositif matériel** (*ordinateur dans l'ordinateur*) qui permet de contrôler les fonctions de base de la carte mère :
 - Allumage/extinction électrique hors OS
 - Reset de la machine
 - Identification visuelle de la machine
 - Redirection de l'affichage texte ou graphique (Bios puis système)
- ▶ La BMC dispose d'une adresse IP qui lui est propre et propose en général :
 - Une interface en ligne de commande
 - Une interface Web
 - Une interface standardisée appelée IPMI

- ▶ Pour accéder en réseau à la BMC, on peut choisir :
 - d'une interface réseau dédiée si présente sur le serveur
 - Performant
 - Complique le câblage : 2 câbles Ethernet par serveur
 - une interface réseau partagée avec celles du serveur
 - Câblage allégé
 - Bien concevoir son réseau : se rappeler que 2 adresses IP cohabiteront sur le même port de switch
- ▶ Le standard logiciel est IPMI (Intelligent Platform Management Interface)
 - En local sur le noeud ou dialogue client/serveur

- ▶ Exemples de commandes Ipmi :
 - Redémarrage d'un serveur :

```
ipmitool -I lan -H nom_machine -U user  
-P motdepasse power cycle
```
 - Accès à la console texte (Serial On Lan)

```
ipmitool -I lanplus -H nom_machine -U  
user -P motdepasse sol activate
```
 - Obtention de la température d'un serveur

```
ipmitool -I lan -H nom_machine -U user  
-P motdepasse sdr type Temperature
```
- ▶ Il s'agit de commandes de base à intégrer dans une suite logicielle plus vaste

- ▶ La réponse aux commandes Ipmi dépend du serveur.
- ▶ **Serveur HP**

CPU0 Diode	20h ok 3.0 51 degrees C
CPU1 Diode	22h ok 3.1 36.50 degrees C
Power Ambient	24h ok 0.0 17.50 degrees C
CPU0 Prochot	07h ok 3.0 Limit Not Exceeded
CPU1 Prochot	08h ok 3.1 Limit Not Exceeded

- ▶ **Serveur SuperMicro + GPU**

GPU1 Temp	18h ok 7.1 41 degrees C
GPU2 Temp	19h ok 7.1 41 degrees C
System Temp	03h ok 7.1 29 degrees C
P1-DIMM1A Temp	60h ok 7.1 40 degrees C
P1-DIMM1B Temp	61h ns 7.1 No Reading

- ▶ Exemple d'interface Web fournie par une carte d'administration à distance (Asus)

The screenshot shows the ASUS iKVM remote administration interface. The browser window is Mozilla Firefox, displaying the URL `https://192.168.80.69/index.html`. The page has a blue header with the ASUS logo and the text 'ASMB4-iKVM'. Below the header is a navigation menu with the following items: System Information, **Server Health**, Configuration, Remote Control, Maintenance, Languages, and HELP. The main content area is titled 'Server Health' and includes a sub-section 'Sensor Readings'. The 'Sensor Readings' section contains a table with the following data:

Name	Status	Reading
CPU1 Temperature	Normal	17 degrees C
CPU2 Temperature	Normal	16 degrees C

- ▶ Autre solution pour accéder aux machines sans se déplacer : KVM (Keyboard Video Mouse)
- ▶ Permet de regrouper les branchements de claviers, écrans, souris... de plusieurs ordinateurs sur un seul dispositif physique
- ▶ Prévoir un KVM sont accessible en réseau



Elements de coût et de choix

- ▶ KVM : 100 € par serveur. Limite dans le nombre de machins pouvant être connectées
- ▶ Carte IPMI additionnelle si le serveur n'est pas équipé : 75 €
- ▶ En raison du coût très peu élevé des solutions à base d'IPMI, elles sont recommandées
- ▶ De plus, IPMI est utilisé dans les suites logicielles de gestion de clusters

Démonstration (en réseau) !

Live web imaging unleashed.



- ▶ Clusters : quels systèmes d'exploitation ?
 - Distributions, génériques, spécialisées
- ▶ Quelle base matérielle commune ?
 - Ce qu'il faut penser à demander
- ▶ Installation d'un cluster
 - Déploiement d'images
 - Installations automatisées
 - Pré/Post Installs, rocks, xcat2
- ▶ Conclusion

- ▶ Sur un cluster comprenant de nombreux noeuds, le but est de maintenir parfaitement **cohérents et identiques** tous les systèmes :
 - paquets installés identiques
 - configurations similaires
- ▶ But : éviter un dysfonctionnement d'application lié à une différence système
- ▶ Sans outil spécifique, avec un peu de rigueur, on pourrait y arriver

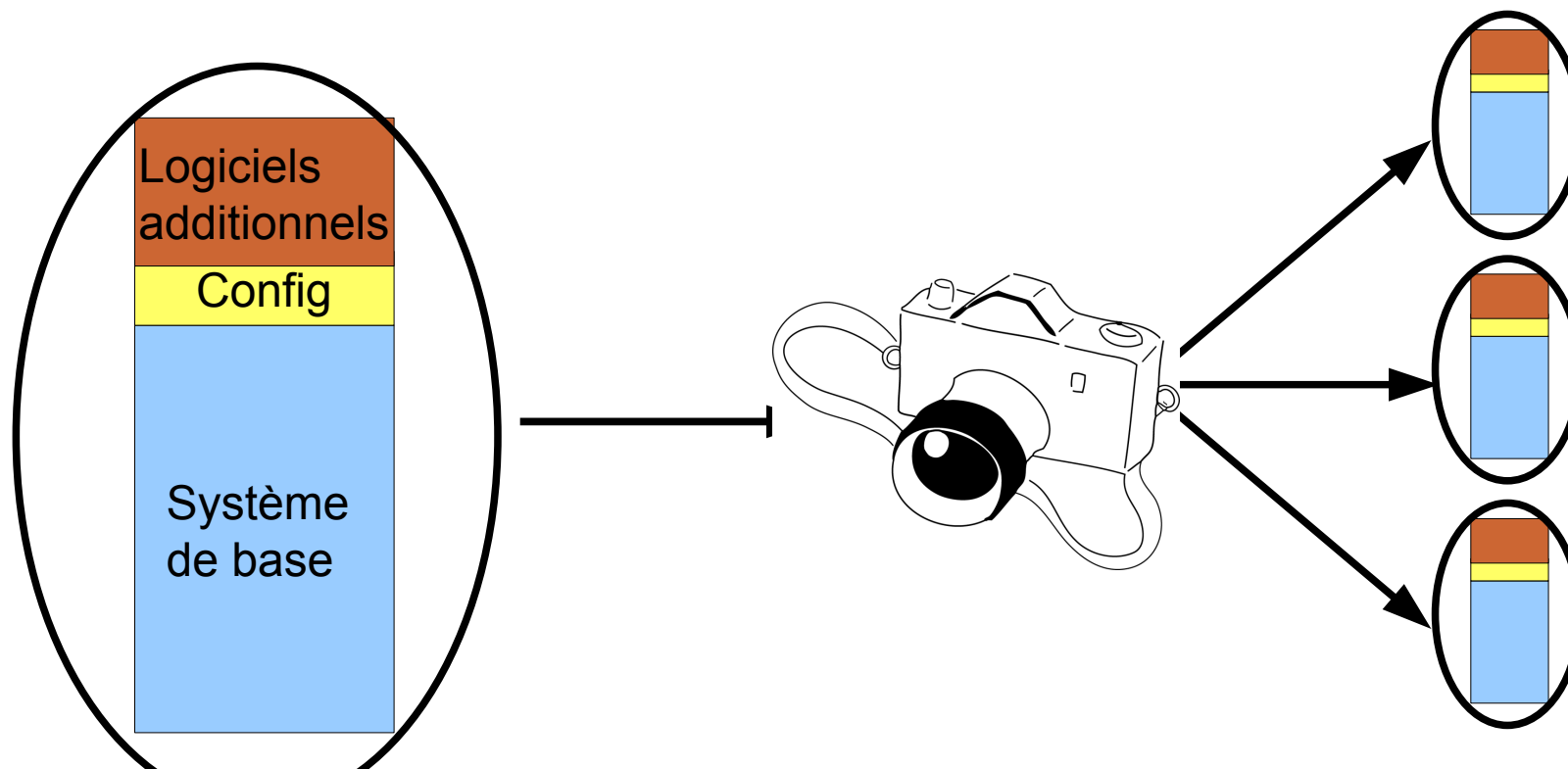
- ▶ Par exemple, une fois les systèmes installés, il suffit de passer les mêmes commandes sur toutes les machines :
 - <http://sourceforge.net/projects/clusterSSH/> : cluster ssh
 - scripts maison
- ▶ Il s'agit d'une tâche répétitive, raison pour laquelle des outils ont été proposés pour l'automatiser
- ▶ Voici leur fonctionnement et les grandes étapes à suivre

- ▶ Lors de l'installation d'un cluster, le principe est le suivant :
- ▶ Installation du noeud maître. Configuration du réseau sur ce noeud (2 cartes réseau) :
 - Une interface reliée au réseau du laboratoire, de l'université : interface vers l'*extérieur*
 - Une interface réseau dédiée aux noeuds du clusters : interface *interne*
- ▶ Installation du premier noeud de calcul sur le réseau interne.
- ▶ Duplication du système sur les autres noeuds

- ▶ La duplication du système sur les autres noeuds se réalise en deux étapes :
 - Construction d'un modèle à partir du système du premier noeud de calcul
 - Recopie automatisée de ce système sur les autres noeuds
- ▶ Le serveur frontal réalise ces deux étapes. Il existe deux familles de méthodes
 - Clonage
 - Replay

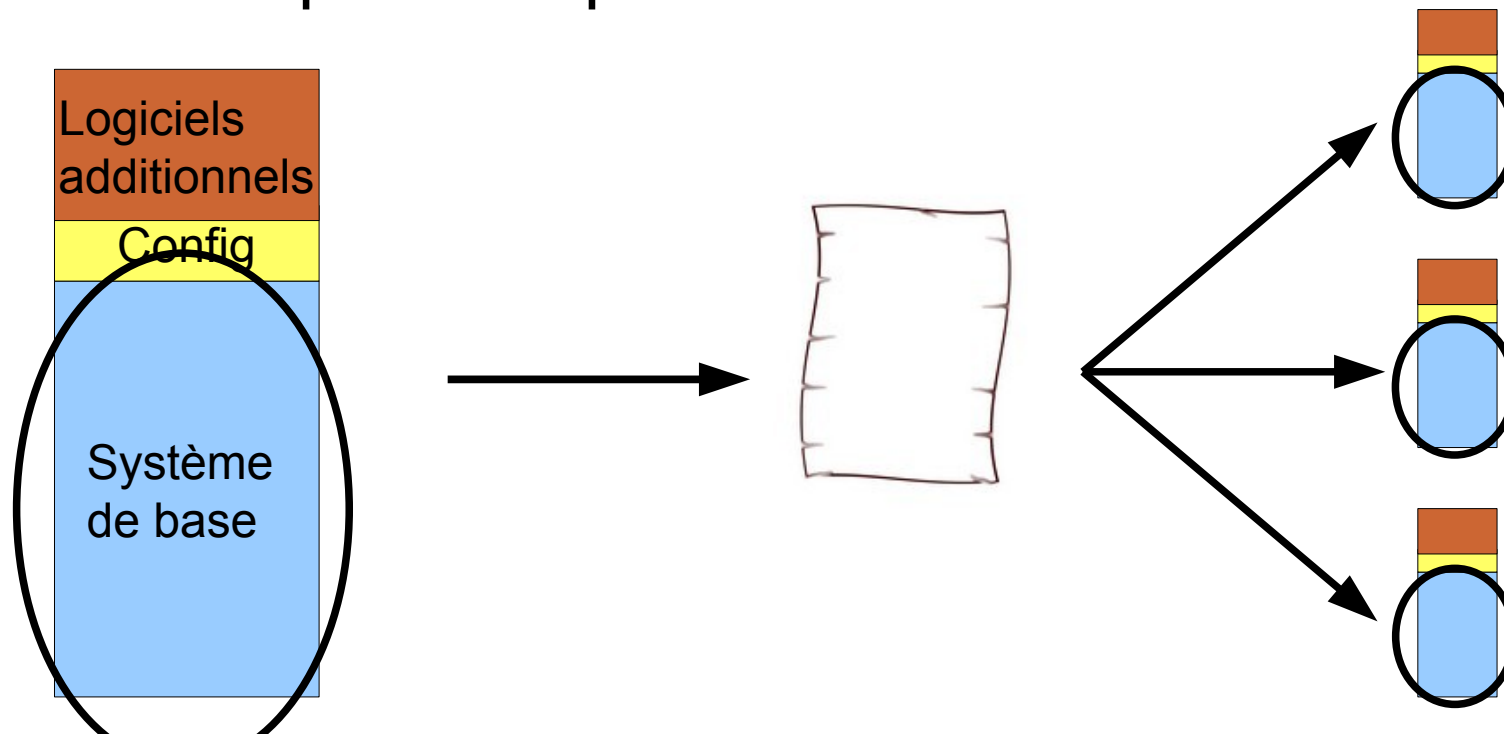
► Méthode par clonage :

- On reprend le système *finalisé* sur le noeud de *référence* = *image*
- Par des mécanismes propres au système de clonage, cette image est *dupliquée*



► Méthode par replay :

- On reprend la liste des paquets installés et les réponses aux questions posées lors de l'installation : *fichier de réponses*
- L'installation est *rejouée* : l'installeur du système utilise les réponses qui lui sont fournies automatiquement



- ▶ Après clonage, chaque noeud de calcul doit pouvoir être identifiée de manière unique sur le réseau sans conflit.
 - Risque d'incohérence si le système cloné est mis en route tel quel
 - C'est pourquoi la configuration réseau doit être reprise par les outils de clonage
- ▶ De manière générale, conserver le minimum d'informations de configuration sur le noeud
 - Configuration réseau par DHCP
 - Centraliser les comptes utilisateurs en créant un annuaire LDAP sur le noeud maître.

- ▶ http://wiki.systemimager.org/index.php/Main_Page
OU
<http://www.projet-plume.org/fr/fiche/systemimager>
- ▶ SystemImager est un ensemble d'outils permettant d'automatiser l'installation **et la mise à jour** d'un parc de machines tournant sous systèmes Linux.
- ▶ Les installations d'images se font automatiquement via le réseau avec les protocoles pxe et tftp ou en créant un CD d'amorçage
- ▶ Sur le site de l'Université de Strasbourg, utilisé en production depuis 2005

- ▶ Principe de systemimager : déployer un système Linux peut se résumer à :
 - copier l'ensemble des fichiers (mode fichier)
 - à effectuer quelques étapes préparatoires
 - il suffit de scripter le tout...
- ▶ **Systemimager fournit des scripts pour déployer des systèmes**
- ▶ Les outils de systemimager automatisent ces étapes en ajoutant une gestion de l'amorçage réseau.

- ▶ 1ère étape : installation du système sur le noeud de référence
 - Configuration soignée des logiciels
 - Installation des logiciels spécifiques au cluster
 - Installation des utilitaires systemimager permettant de transférer les fichiers vers le serveur d'installation
 - Préparation du noeud au clonage
- ▶ 2ème étape : installer un serveur de clonage
 - DHCP, TFTP (Trivial File Transfer Protocol)
 - Utilitaires systemimager
- ▶ 3ème étape
 - Télécharger l'ensemble des fichiers du système du noeud de référence sur le serveur

- ▶ 4ème étape : installation d'un noeud
 - Démarrage du noeud en réseau ou sur un CD.
 - Le noeud télécharge un mini-noyau
 - Ce mini-noyau exécute un script compliqué (produit automatiquement par systemimager) qui :
 - Partitionne les disques
 - Télécharge le système à cloner sur le noeud
 - Effectue une configuration système (principalement réseau) sur le système téléchargé
 - Redémarre le noeud à la fin de l'installation
- ▶ De notre expérience, systemimager fonctionne bien pour le clonage de systèmes de type Redhat (Centos)

- ▶ Systemimager permet également la mise à jour des systèmes sans ré-installation
- ▶ Cette mise à jour se fait en deux étapes :
 - Mise à jour du noeud de référence :
 - Installation d'un nouveau paquet
 - Modification d'une configuration
 - Report de cette mise à jour sur le serveur
 - Seul le différentiel est transféré (pas de mécanisme d'historique)
 - Sur chacun des noeuds de calcul, transfert des mises à jour de l'image depuis le serveur
 - Seul le différentiel est transmis

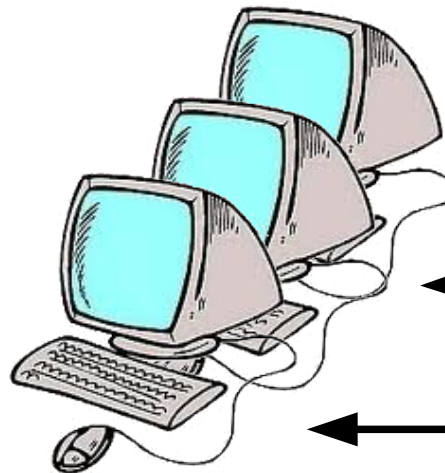
► Workflow dans systemimager



1 : Installation soignée
(ou mise à jour)



2 : Aspiration des fichiers
sur le serveur



3 : Installation ou mise
à jour des autres noeuds

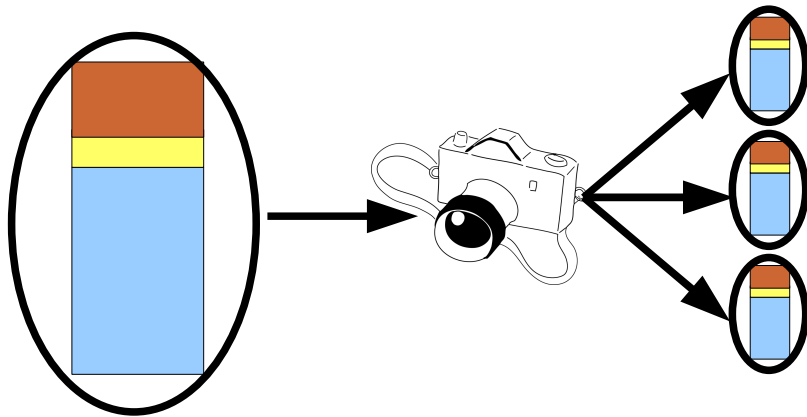
► Points forts de l'outil

- La mise à jour *en ligne* des systèmes est possible
- Prend en charge des configurations de clusters hétérogènes :
 - scripts d'installation adaptés à chaque génération de noeud
 - configurations différenciées suivant les noeuds

► Limitations de systemimager :

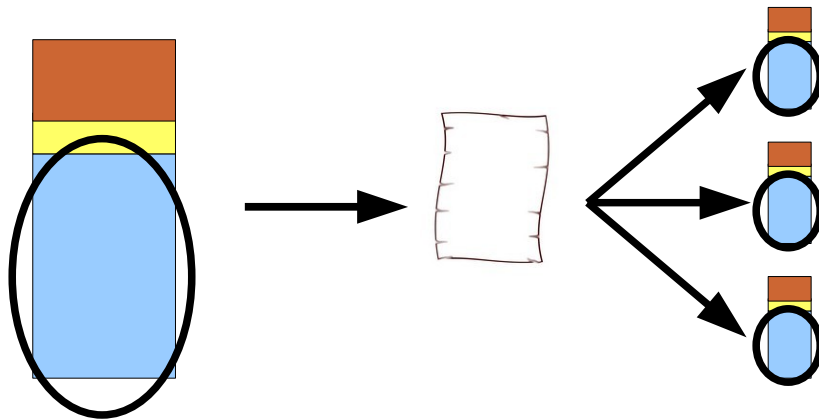
- L'outil n'est pas intégré (pas d'IPMI, gestion de parc)
- Les scripts d'autoinstallation sont complexes
- La mise à jour en ligne s'appuie sur des *listes d'exclusions* lourdes à gérer

- ▶ Le principe de Systemimager est de greffer une suite de clonage sur un système existant, construit *à la main*



- ▶ D'autres systèmes Linux sont déjà conçus pour tourner sur cluster et intègrent des mécanismes de mise en cluster

- ▶ Rocks est une distribution Linux s'appuyant sur Centos (donc sur RedHat) permettant d'automatiser l'installation de systèmes en cluster



- ▶ Configuration maître / noeuds intégrée à l'installation
- ▶ <http://www.rocksclusters.org>

- ▶ Rocks utilise des ensembles de logiciels Linux re-packagés en *rolls*
- ▶ Les rolls contiennent des logiciels usuels dans les clusters, **prêts à être déployés** et préconfigurés pour un schéma en cluster : un noeud maître, des noeuds de calcul.
- ▶ rolls de base indispensables :
 - Kernel/Boot Roll
 - base Roll
 - OS Roll

Welcome to Rocks



Selected Rolls

No rolls have been selected.

If you have CD/DVD-based rolls (that is, ISO images that have been burned onto CDs or a DVD), then click the *CD/DVD-based Roll* button. The media tray will eject. Then, place your first roll disk in the tray and click *Continue*. Repeat this process for each roll disk.

If you are performing a network-based installation (also known as a *central* installation), then input the name of your roll server into the *Hostname of Roll Server* field and then click the *Download* button. This will query the roll server and all the rolls that the roll server has available will be displayed. Click the *selected* checkbox for each roll you will to install from the roll server.

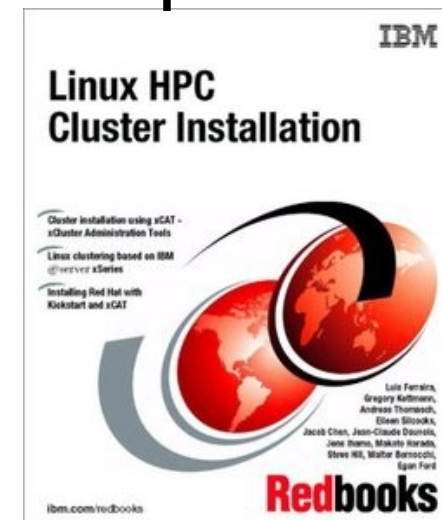
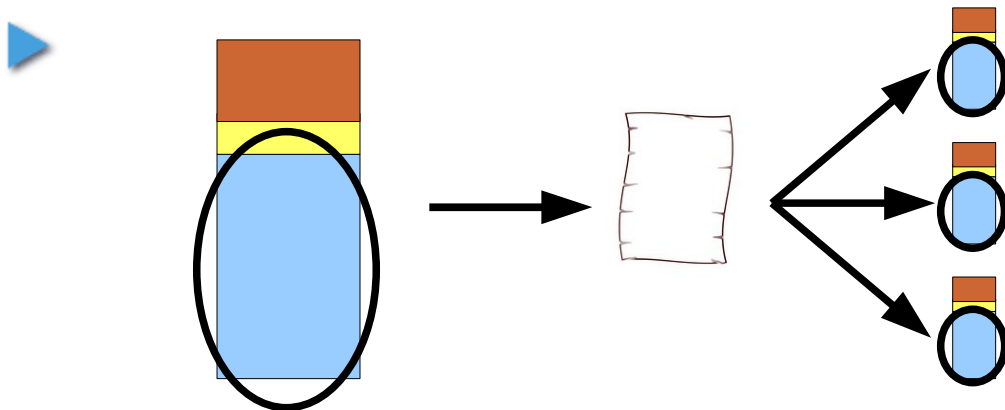
When you have completed your roll selections, click the *Next* button to proceed to cluster input screens (e.g., IP address selection, root password setup, etc.).

Selected	Roll Name	Version	Arch
<input type="checkbox"/>	area51	5.4.3	x86_64
<input type="checkbox"/>	base	5.4.3	x86_64
<input type="checkbox"/>	bio	5.4.3	x86_64
<input type="checkbox"/>	condor	5.4.3	x86_64
<input type="checkbox"/>	ganglia	5.4.3	x86_64
<input type="checkbox"/>	hpc	5.4.3	x86_64
<input type="checkbox"/>	java	5.4.3	x86_64
<input type="checkbox"/>	kernel	5.4.3	x86_64
<input type="checkbox"/>	os	5.4.3	x86_64
<input type="checkbox"/>	perl	5.4.3	x86_64
<input type="checkbox"/>	python	5.4.3	x86_64
<input type="checkbox"/>	sge	5.4.3	x86_64
<input type="checkbox"/>	web-server	5.4.3	x86_64
<input type="checkbox"/>	xen	5.4.3	x86_64

- ▶ Quelques rolls fournis :
 - SGE, Torque : gestion de batch
 - HPC : MPI, benchmarks
 - Bio : Bioinformatique (HMMER, NCBI BLAS, MpiBLAST, biopython, ClustalW, MrBayes)
 - Ganglia : monitoring
 - Cuda (programmation GPU)

- ▶ Il existe une version de Rocks commerciale :
Rocks+HPC
<http://www.stackiq.com/cluster-computing/>
- ▶ Propose du support
- ▶ Fournit des rolls de logiciels commerciaux
- ▶ Prix de base annoncé : 149\$ / noeud

- ▶ xcat2 est une suite logicielle gratuite de IBM permettant le déploiement d'un parc hétérogène de serveurs
- ▶ xcat : *Extreme* cluster administration toolkit
- ▶ xcat2 prend en charge plusieurs distributions Linux : familles RedHat, Debian
- ▶ Intègre le boot réseau et la gestion via Ipmi



- ▶ xcat2 est écrit en Perl
- ▶ xcat2 utilise une base de données de configuration du Cluster qui comprend un certain nombre de *tables*
 - switches
 - machines
 - groupes de machines
- ▶ Un grand ensemble de commandes texte permet de modifier les configurations
- ▶ xcat2 utilise des versions spécifiques de certains services : dhcp, résolution de nom, tftp

► Workflow dans xcat2



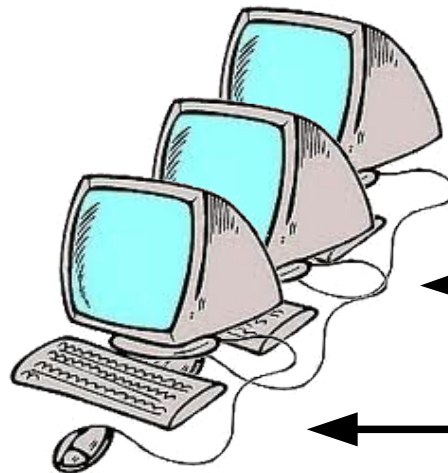
1 : Installation d'un noeud de calcul



2 : Transfert de la liste de paquets sur le serveur



3 : Creation d'un dépôt d'Installation sur le serveur



4 : Déploiement des autres noeuds

- ▶ xcat2 permet également d'écrire un ensemble de scripts de post-installation
 - paramétrables par groupe de machine
 - spécifiques à une tâche : plusieurs scripts seront exécutés sur les machine
- ▶ Pour la gestion courante du cluster, un mécanisme d'exécutions de commandes en parallèle est disponible (y compris ACL)
- ▶ Limites de xcat2 :
 - Le fonctionnement en mode cluster (montage du partage NFS, batch) est à configurer.
 - Compétences *systeme* nécessaire

► Nous avons vu :

- Une solution de clonage de système : systemimager
- une distribution Linux destinée à être installée sur un cluster : Rocks
- un outil d'administration intégré, xcat2

Outil	Mode cluster	IPMI	Mise à jour online	Portail Web
System imager	À paramétrer	Non	Oui	Non
Rocks	Pré-configuré	Non	Non	Non
xcat2	À paramétrer	Oui	Non	En cours

- ▶ Il existe un grand nombre d'outils permettant de simplifier et de centraliser l'administration système
 - puppet (<http://www.puppetlabs.com/>)
 - oscar
- ▶ Choisir un outil **dédié aux clusters** permet de s'affranchir d'actions de configuration complexes.
- ▶ On trouve également des outils commerciaux :
 - Bright Cluster Manager
 - HP Cluster Management Utility

- ▶ À partir du serveur maître installé hier, nous allons installer un cluster avec Rocks
- ▶ Rappel : installation du serveur maître
 - Celui-ci dispose de 2 interfaces réseau
 - La première sur le réseau interne
 - La deuxième reliée à Internet
 - Nous l'avons installé à partir de l'image d'un CD
- ▶ Installation d'un noeud de calcul
 - Celui-ci dispose d'une seule interface réseau
 - Nous l'installons à partir du réseau
 - Nous préparons l'opération sur le maître
 - Il ne reste qu'à allumer le noeud

- ▶ Clusters : quels systèmes d'exploitation ?
 - Distributions, génériques, spécialisées
- ▶ Quelle base matérielle commune ?
 - Ce qu'il faut penser à demander
- ▶ Installation d'un cluster
 - Déploiement d'images
 - Installations automatisées
 - Pré/Post Installs, rocks, xcat2
- ▶ Conclusion

- ▶ Le système Linux est parfaitement adapté aux clusters
 - Choix restants : Type de distribution
 - Expérience subjective : bonne satisfaction avec dérivées de RedHat
- ▶ Il existe de nombreuses solutions de mise en cluster
 - Expérience subjective : satisfaction avec Systemimager
 - Rocks semble parfaitement adapté à une mise en cluster rapide
- ▶ À vous d'expérimenter...