Installation et gestion d'un cluster HPC

Logiciel Cluster | Philippe GRÉGOIRE

Installation d'un cluster

Processus d'installation d'un cluster

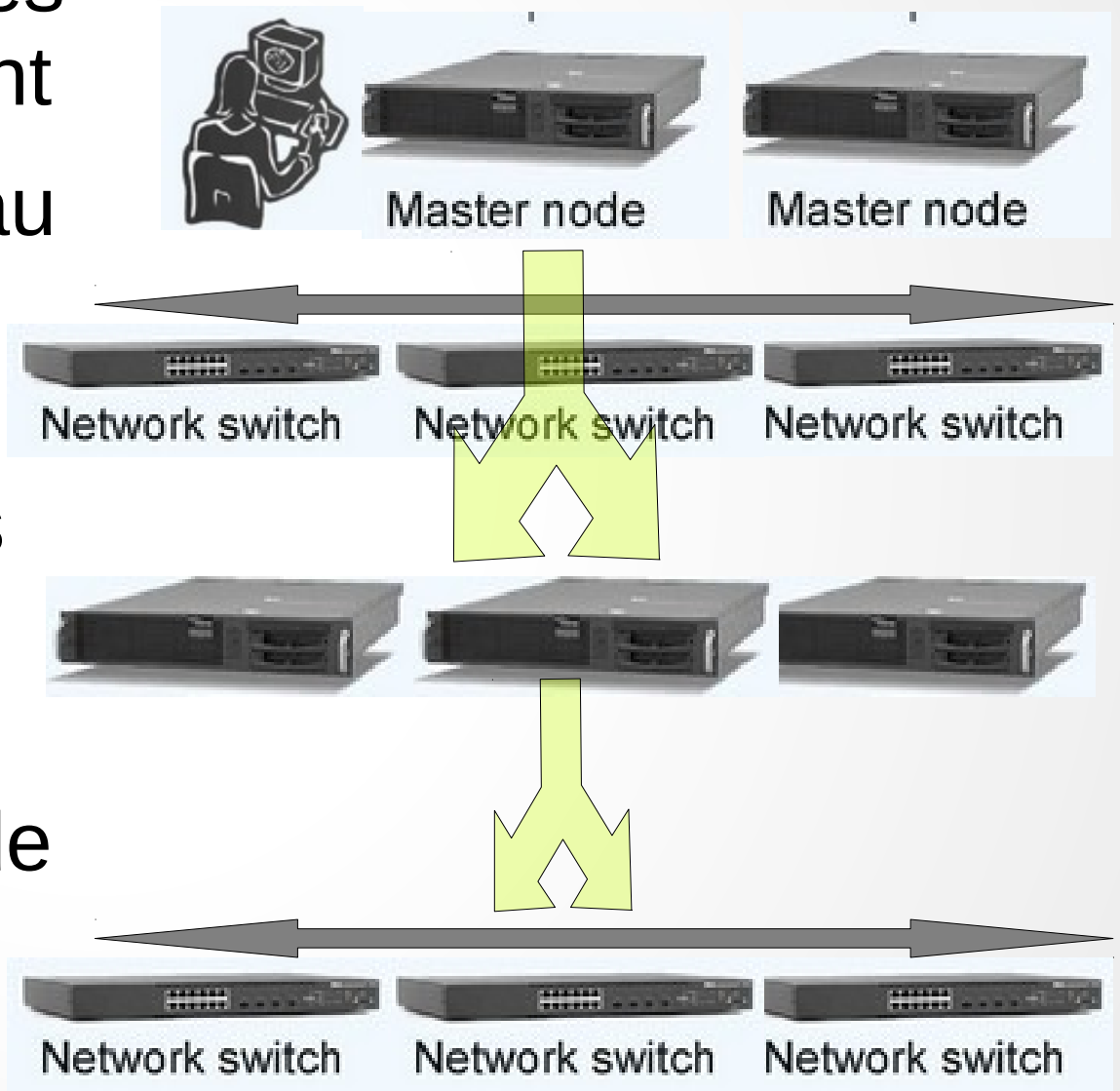
- Description précise de tous les composants matériels
- Assemblage en usine, validation par partie
  - Design par motif (template) reproduit N fois
- Plan d'implantation en salle machine
- Préparation de la salle machine
    - Renforcement des faux planchers
    - Refroidissement
    - Chemins de câbles réseaux
    - Chemins de câbles électriques

Installation matérielle

- Positionnement des racks en salle machine
- Raccordements électriques
- Raccordements hydrauliques
- Connexions réseau inter-racks
  - Réseau ethernet (management)
  - Réseau BackBone
  - Réseau de stockage
  - Réseau d'interconnexion
- Vérification du câblage

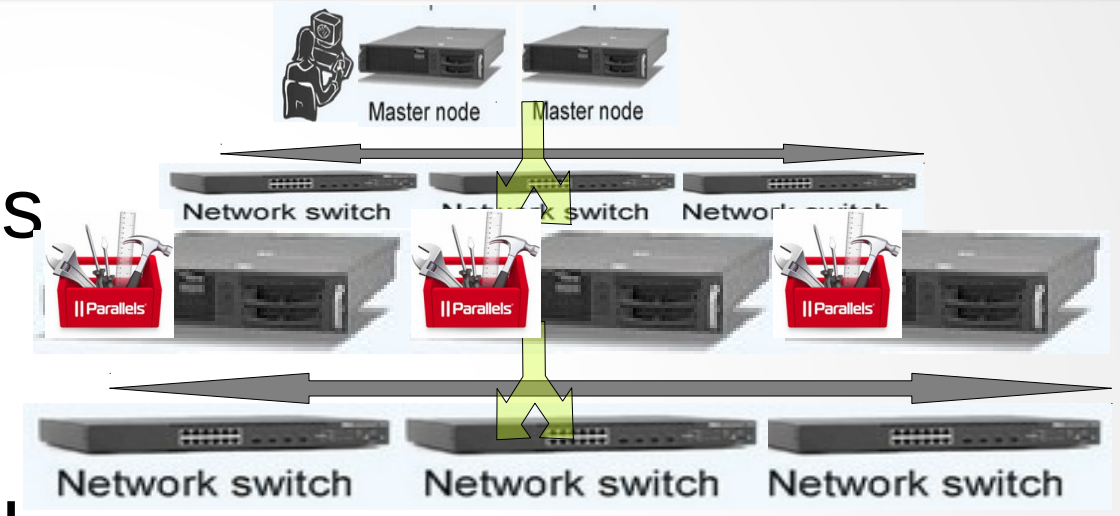
Processus d'installation d'un cluster

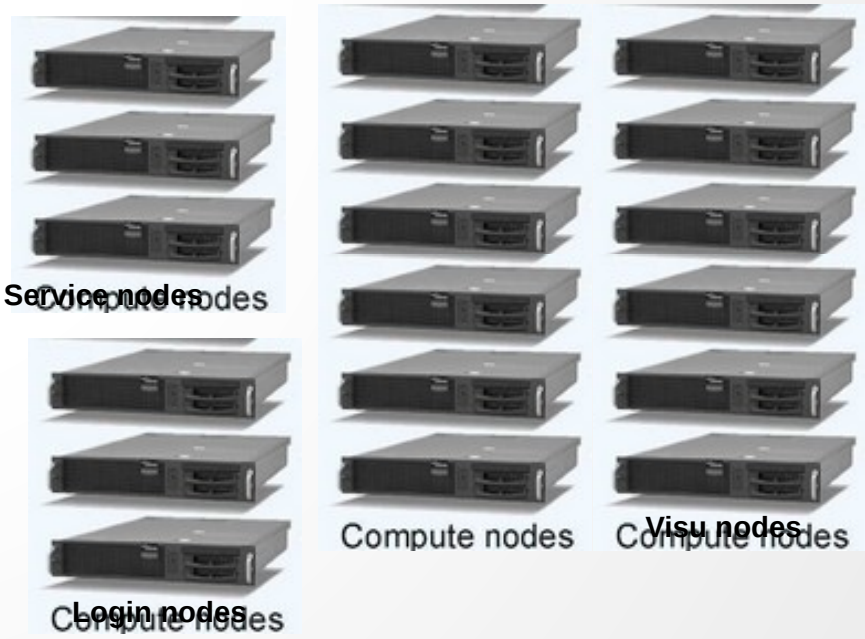
- Installation logicielle des nœuds de management
- Configuration du réseau principal de management
- Installation des nœuds de service
- Configuration des réseaux secondaires de management



Processus d'installation d'un cluster

- Mise en place des outils de d'installation des nœuds de calcul
- Déploiement des nœuds de service
- Déploiement des nœuds de calcul, visu, login,
- Configuration des nœuds de service, calcul, visu, login.



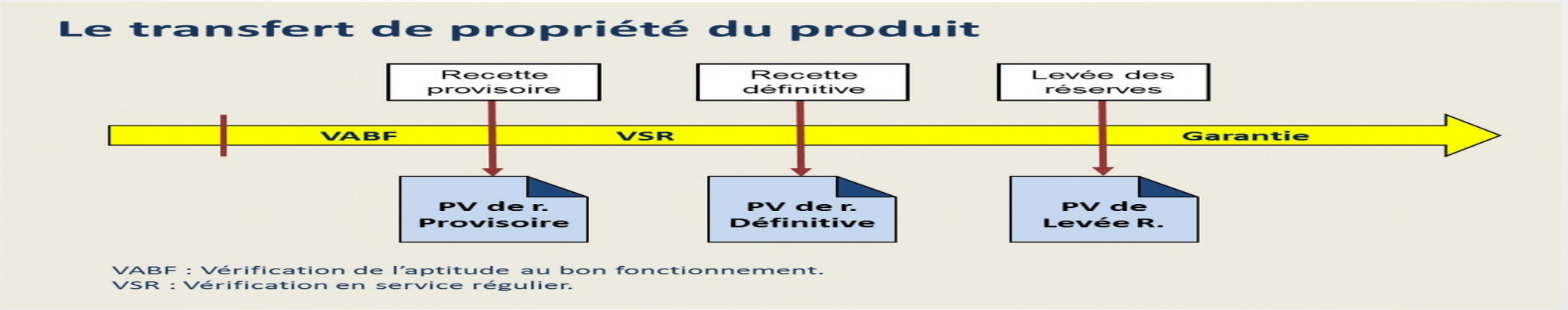


Processus d'installation d'un cluster

- Validation des performances :
  - Unitaires : performance de chaque nœud
  - Au niveau d'un châssis (niveau L1)
  - D'un sous-ensemble de châssis (niveau L2)
  - De la machine complète (L3 ou +)
- Recette (ou tests d'acceptation)
  - Tests de performance : Flops, Go/sec, temps d’exécution
  - Tests de robustesse (taux de réussite)
  - Vérification d'aptitude au bon fonctionnement (VABF)

Processus d'installation d'un cluster

- Mise en service opérationnelle
  - Configuration pour intégration dans le centre de calcul
  - Démarre une période de vérification de service régulier (VSR) en environnement de production
  - Premiers utilisateurs, comparaisons des résultats/performances
  - Grands challenges
  - Mise en production



Outils d'installation d'un cluster

- Au début, à l'installation du cluster (quelques mois)
- Plus tard, lors des extensions matérielles
- Complexité de l'architecture matérielle :
  - Des dizaines à des centaines de racks
  - Des milliers de nœuds
  - Des milliers de câbles

Outils d'installation d'un cluster

- Référentiel indispensable
  - Quel modèle de représentation ?
- Génération automatique de certaines configurations
  - Switchs Ethernet des réseaux de management
  - Contrôle des équipements
- Vérification de la cohérence du câblage avec sa représentation
- Intérêt des motifs de configuration matériels et logiciels

Outils d'installation matérielle

- Des composants de base de plus en plus complexes
  - Commutateurs réseaux (switchs), cartes réseaux,
  - Cartes mères, contrôleurs disques
  - Systèmes d’alimentation (PSU),
  - Systèmes de refroidissement
  - Intègrent des firmwares (micro-code ou micro-logiciel) qui permet de le faire fonctionner, de le contrôler, de l'interroger.

Outils d'installation matérielle

- Des composants nécessitant des
  - outils pour mettre à jour ces firmwares
  - outils pour les contrôler
- Pour un système complexe, nécessité de valider ces composants ensemble
  - Notion d'état technique
- Outil de gestion des firmwares à l'échelle du cluster :
  - Mise à jour des composants en parallèle
  - Vérification de la cohérence firmware du cluster

Firmware

- Nécessaire au fonctionnement du matériel
- Occupe peu de place en mémoire
- Embarqué sur le matériel
- Difficile à mettre à jour
- Nécessite souvent un arrêt électrique
- Une mise à jour peut-être fatale !

Maintien en conditions opérationnelles

MCO

- Maintien en condition opérationnelles
  - Disponibilité, sécurité, fiabilité des calculateurs
- Garantir que la dégradation des moyens de calcul n’entraîne pas des conditions de travail inacceptables.
  - Critères contractuels

MCO

- Mettre en œuvre, valider et suivre les procédures :
  - Conduite de l'exploitation
  - Gestion des alarmes
  - Consignes de sécurité
- Détecter et suivre les incidents liés aux dysfonctionnements techniques
- Garantir la disponibilité des services
- Suivre l'évolution des consommations de ressources
  - Espace disque
  - Espace bandes robotique

Maintenance

Maintenance

- Maintenance matérielle :
  - Certaines opérations ne peuvent se faire qu'en arrêtant le serveur
  - Mise à jour de firmwares
  - Changement de cartes
  - Remplacement de pièces

Maintenance

- Maintenance logicielle:
  - Les logiciels ont toujours des bugs
  - Attente de nouvelles fonctionnalités
  - Correction de failles de sécurité
  - Raisons de pérennité de support
  - Compatibilité avec d'autres composants

Maintenance

- Avantages et Inconvénients
  - Un système à jour est plus sur
  - Un système à jour est plus fiable
  - Un système à jour a des nouveaux bugs
    - Les applications peuvent ne plus fonctionner
  - Les maintenances coûtent du temps.
  - Nécessité d'avoir un système de tests.

Maintenance

- Un grand nombre de composants matériels :
  - Barrettes mémoire
  - Disques
  - Processeurs
  - Câbles réseau
- Le temps moyen entre 2 pannes est très court à l'échelle du cluster
  - MTBF Mean Time Between Failure
  - Outils de MCO

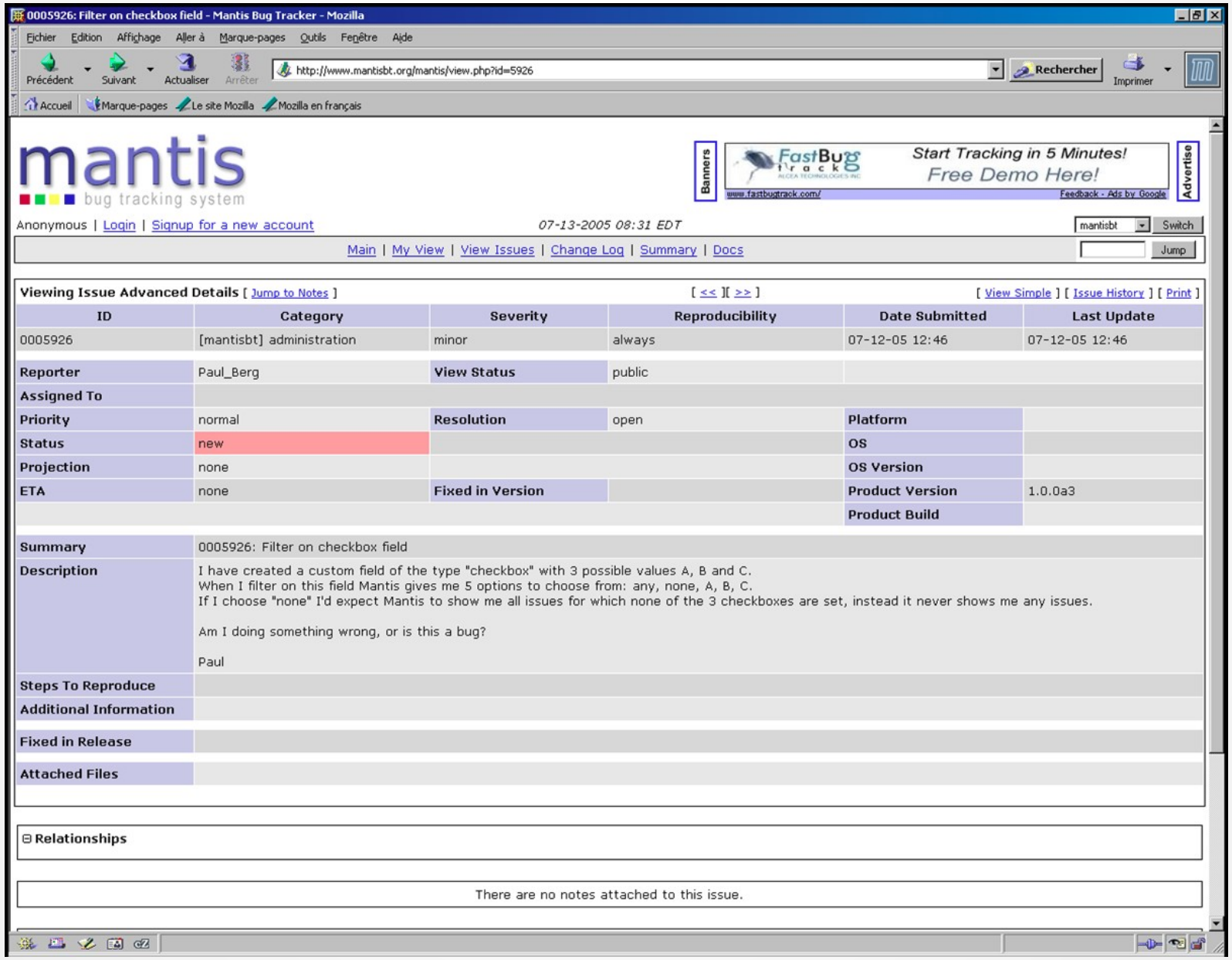
Maintenance

- Nécessite d'un processus de maintenance continue
  - Maintenance corrective
    - ensemble des tâches effectuées après la détection d'une panne et destinée à remettre un système dans l'état dans lequel il peut accomplir une fonction requise
    - Actions curatives ou palliatives (durée limitée)
    - Ex :
      - Remplacement d'un contrôleur disque
      - Remplacement d'une carte dans un commutateur réseau
      - Mise à jour d'un paquet pour corriger une faille de sécurité
      - Mise en place d'un patch live logiciel (systemtap)

Outil de suivi d'incidents (Mantis)

- Conserver l'historique des incidents
- Connaître l'impact sur la machine
- Corrélation avec des événements
  - mises à jour logicielles, nouveaux codes
- Description précise de l'incident
  - Messages d'erreurs, répertoire des fichiers de traces et de dumps.
- Conserver l'historique des actions
  - Description des actions effectuées
    - Analyse
    - Rapport d'anomalies/incidents vers le constructeur (référence ticket)
    - Contournement
  - Dates des différentes étapes (contractuelles) (schéma?)

Mantis


0005926: Filter on checkbox field - Mantis Bug Tracker - Mozilla
Fichier Edition Affichage Aller à Marque-pages Outils Fenêtre Aide
Précédent Suivant Actualiser Arrêter
http://www.mantisbt.org/mantis/view.php?id=5926
Rechercher
Imprimer
Accueil Marque-pages Le site Mozilla Mozilla en français
mantis
bug tracking system
Banners
FastBug track
Start Tracking in 5 Minutes!
Free Demo Here!
www.fastbugtrack.com/
Feedback - Ads by Google
Advertise
Anonymous | Login | Signup for a new account
07-13-2005 08:31 EDT
mantisbt
Switch
Jump
Main | My View | View Issues | Change Log | Summary | Docs
Viewing Issue Advanced Details [Jump to Notes]
[<<][>>]
[View Simple] [Issue History] [Print]
ID
Category
Severity
Reproducibility
Date Submitted
Last Update
0005926
[mantisbt] administration
minor
always
07-12-05 12:46
07-12-05 12:46
Reporter
Paul_Berg
View Status
public
Assigned To
Priority
normal
Resolution
open
Platform
Status
new
OS
Projection
none
OS Version
ETA
none
Fixed in Version
Product Version
1.0.0a3
Product Build
Summary
0005926: Filter on checkbox field
Description
I have created a custom field of the type "checkbox" with 3 possible values A, B and C.
When I filter on this field Mantis gives me 5 options to choose from: any, none, A, B, C.
If I choose "none" I'd expect Mantis to show me all issues for which none of the 3 checkboxes are set, instead it never shows me any issues.
Am I doing something wrong, or is this a bug?
Paul
Steps To Reproduce
Additional Information
Fixed in Release
Attached Files
Relationships
There are no notes attached to this issue.

Mantis

Anonymous | Login | Signup for a new account *2009-02-01 03:44 EST*

Main | My View | View Issues | Change Log | Roadmap | Docs | Wiki | ManTweet | Repositories Jump

Unassigned [^] (1 - 10 / 2006)

0010092	Logging of user action (login, download, ...) feature - 2009-01-30 09:55
0009394	MantisConnect Webservice crashes when trying to get issue information api soap - 2009-01-30 07:43
0010094	EMail generation if not assigned bugtracker - 2009-01-30 07:02
0010091	CHANGE STATUS TO: Assigned vs ASSIGN TO ?? customization - 2009-01-29 23:37
0004640	New custom field types: "Version", "User" custom fields - 2009-01-29 22:52
0003790	Additional Custom-Field-Type "users" custom fields - 2009-01-29 22:49
0010065	Custom field "date" not saved silently custom fields - 2009-01-29 08:22
0006418	Change Resolve status and Assignment at the same time feature - 2009-01-29 05:41
0007104	Auto assign on resolve bugtracker - 2009-01-29 05:39
0007150	Automatic reassignment on status transition feature - 2009-01-29 05:39

Resolved [^] (1 - 10 / 254)

~~0010093~~	(phpmailer) class.phpmailer.php => private $smtp is accessed by mantis email - 2009-01-31 14:37
~~0010056~~	Category national character upgrade - 2009-01-30 14:30
~~0009455~~	Database configuration does not correctly deal with array administration - 2009-01-30 11:18
~~0010089~~	Updating version field does not trigger history note administration - 2009-01-28 03:55
~~0008628~~	The "Show Content" link on attached tx file doesn't show text webpage - 2009-01-28 03:47
~~0007790~~	Links protected by brackets are not processed properly bugtracker - 2009-01-27 12:27
~~0009216~~	Problems with Russian language localization - 2009-01-27 11:40
~~0007144~~	Unable to set realname because of existing unsername administration - 2009-01-25 14:21
~~0010078~~	logging into the site with out entering login details security - 2009-01-24 12:44
~~0010073~~	Form Validate in JavaScript javascript - 2009-01-22 14:43

Outils de suivi d'interventions

- Outil de suivi d'interventions matérielles
  - Propre au constructeur
  - Adapté à son organisation interne
    - Références des pièces détachées (FRU : Field Replaceable Unit)
- Outil d'inventaires matériels et logiciels (firmware)
  - Nécessaire pour suivre les mouvements de matériel
    - À l'intérieur du cluster
    - Entre le site et le constructeur

Outils de travail collaboratif

- Outils de planification
  - Coordination des équipes
  - Durée des interventions
- Wiki
  - Partage de connaissances
  - Procédures de maintenance
  - Procédures de mises à jour, d'installation
- Messagerie instantanée
  - Échange rapide des informations,
  - Synchronisation des intervenants

Automatisation

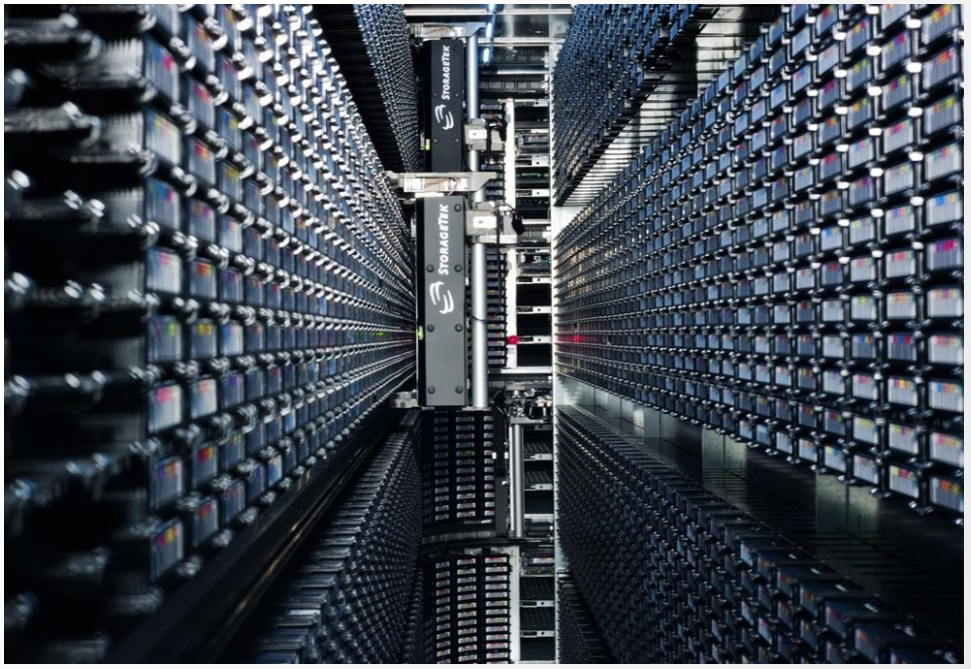
Résilience dès la conception

- Redondance électrique et refroidissement
- Redondance en interne des nœuds
  - Disque en raid
  - Parité mémoire
- Redondance entre nœuds
  - Haute disponibilité
- Redondance des services
- Redondance dans les réseaux
  - Plusieurs chemins

Automatisation des processus

- Interface métiers utilisateurs
  - Automatisation du workflow utilisateur
- Ordonnanceur (Schéduler)
  - Automatisation de l'ordonnancement des travaux
- Résilience des applications
  - Utilisation d'intergiciels (middleware) tolérants aux pannes
    - Mécanisme de retry dans les couches logicielles
  - Mécanisme de Checkpoint/Restart
    - Applicatif : interne, déclenché par l'application
    - Externe : fourni par l'OS
    - Permet la reprise d'une simulation à un moment particulier

Evolution du centre de calcul depuis 1970



Automatisation de la maintenance

- Automatiser certaines taches de maintenance
  - Remise en production des nœuds
  - Application de patches de sécurité
  - Réparation des pannes simples connues
- Réduire le nombre d’arrêt global du cluster
- Réduire le temps des maintenances

Contraintes d'un robot de maintenance

- Sécurisé
  - Impossible d'attaquer le cluster
- Modulaire
  - Interfaçable avec d'autres composants logiciels
- Scalable
  - A l’échelle des clusters
- Fiable
  - Doit améliorer pas dégrader l'état des machines
  - Peut intervenir sur une alerte avant l'astreinte

Robot NetFlix : Winston

- https://qconsf.com/sf2016/system/files/presentation-slides/qconsf_2016_presentation.pdf

- Système d'automatisation
  - Réagit à des événements
  - Élimination des faux positifs
  - Récupération des informations de diagnostics
  - Exécution de procédures de réparation

Facebook AutoRemediation FBAR

- https://www.usenix.org/conference/srecon16/program/presentation/komorn
- Système de surveillance
- Framework de réparation
- Systèmes de maintenance proactive
  - Planifie les actions en fonction des services hébergés dans les racks

Outils Admin

Ergonomie des outils

- Un incident peut impacter un grand nombre de nœuds
  - Panic dans les couches kernel du système de fichiers Lustre
  - Panne IB → Plantage MPI dans une application
  - Débordement mémoire (Out of Memory OOM Killer)
- L'unité de panne / travail est le nodeset
  - Ensemble de nœuds : nodes[1000-1500]
  - Au bout de quelques semaines, le nodeset peut être plus morcellé

Ergonomie des outils

- Nécessite d'adapter les outils
  - À la taille de la machine
  - À la taille des équipes
    - Ne suit pas la taille des machines
- Efficacité des outils importante pendant
  - Phase de mise au point de la machine
  - Maintenance
- Interface graphique (GUI) peu adaptée
- Interface commande (CLI) plus efficace

Ergonomie des outils

- Les outils conçus pour les clusters
- Nodeset en arguments
- Présentation synthétique des résultats
  - Agrégation des résultats identiques
  - Nodesets en sortie
- Passage à l'échelle (scalabilité)
  - Parallélisation de l’exécution
  - Contrôle du degré de parallélisation (fan-out)
- Résilience aux pannes de nœuds
- Sécurité

Ergonomie des outils

- ipmitool
  - Logiciel opensource pour gérer des serveur
  - Permet l'accès /controle distant d'un serveur via l'interface réseau de sa BMC
- ipmifree
  - Logiciel opensource pour gérer des serveur
  - Permet l'accès /contrôle distant d'un serveur via l'interface réseau de sa BMC
  - Supporte les nodesets

Ergonomie des outils : un nœud

- Exemple ipmitool sur un nœud

  # Ipmitool -I lanplus -H node1200-bmc chassis power off

- Exemple freeipmi sur un nœud

  # Ipmipower -D lan_2_0 -h node1200-bmc --off

Ergonomie des outils : un nœud

- Exemple ipmitool sur un nœud

```
# ipmitool -I lanplus -H node1200-bmc chassis power off
```

- Exemple freeipmi sur un nœud

```
# ipmipower -D lan_2_0 -h node1200-bmc --off
```

Ergonomie des outils : nodeset simple

- Exemple ipmitool sur un nodeset simple

```
# for ((n=1000 ; n<=1200 ; n++))
do
   ipmitool -I lanplus -H node${n}-bmc chassis power off
done
```

- Exemple freeipmi sur un nodeset simple

```
# ipmipower -D lan_2_0 -h node[1000-1200]-bmc --off --fanout=64
--consolidate-output
```

Ergonomie des outils : nodeset complexe

- Exemple ipmitool sur un nodeset complexe

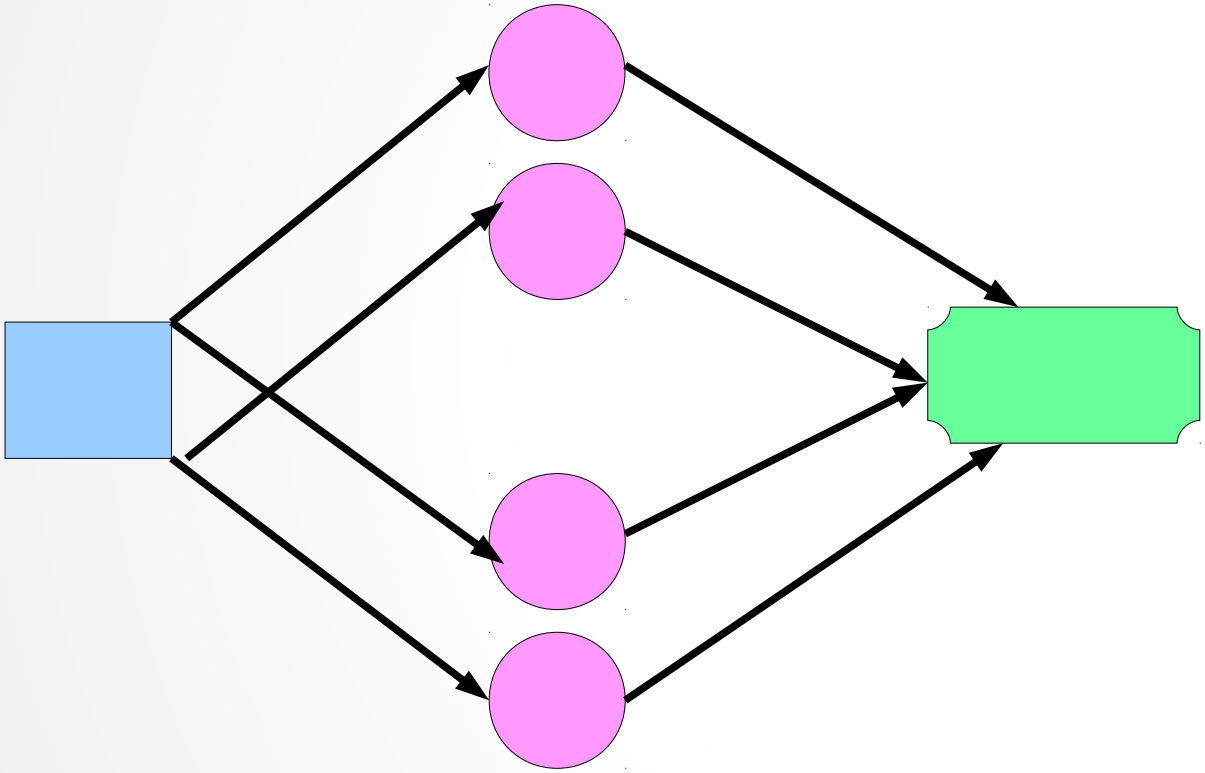
```
    # ??? découper le nodeset en nodeset simple ???
```

- Exemple freeipmi sur un nodeset complexe

```
   # ipmipower -D lan_2_0 \
      -h node[1000-1200,1206,1213,1301,1329,1455-1623,
       1867-1923,2311,2672,3450-3459,4061]-bmc \
       –off  --fanout=64  --consolidate-output
```

Conception des outils //

- Un élément important le fan-out



ClusterShell

- Bibliothèque Python pour exécuter des commandes en parallèle

- Améliorer l'administration des grands systèmes
  - Efficacité et scalabilité
  - Mutualisation des techniques
  - Facilité d'utilisation

- Cibles
  - Clusters « top 10 »
  - Clusters d'entreprise
  - Ferme de serveurs
  - Parc de stations, etc.

ClusterShell

- Points forts du langage Python
  - Programmation orientée objet
  - Lisibilité/réutilisabilité/maintenance du code : *code clarity*
  - Langage mature et robuste
  - *Python Standard Library*
  - Portabilité
  - Evolution (*Python C/C++ bindings*)

- Points faibles / contraintes
  - Vitesse d'exécution de Python ?

ClusterShell

- Points forts en pratique pour l'administrateur système
  - Maîtrise rapide du langage
    - Cycles de développement rapides
    - Concepts de bonne programmation (code design) dans des outils systèmes!
  - Support des exceptions et traceback
  - Nombreux modules dans la *Python Standard Library*
    - regexp, réseau, base de données, algorithmes…

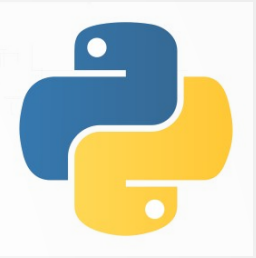
ClusterShell

- **Bibliothèque** Python pour exécuter des commandes en parallèle

- Améliorer l'administration des grands systèmes
  - Efficacité et scalabilité
  - Mutualisation des techniques
  - Facilité d'utilisation

- Cibles
  - Clusters « top 10 »
  - Clusters d'entreprise
  - Ferme de serveurs
  - Parc de stations, etc.

ClusterShell : Python ?



- Points forts du langage Python
  - Programmation orientée objet
  - Lisibilité/réutilisabilité/maintenance du code : *code clarity*
  - Langage mature et robuste
  - *Python Standard Library*
  - Portabilité
  - Evolution (*Python C/C++ bindings*)

- Points faibles / contraintes
  - Vitesse d'exécution de Python ?
    - Rapidité = code « Pythonic »

ClusterShell : CLI

- Pour manipuler des listes de machines et *ranges* de noeuds.

```
$ nodeset --count mars,venus,lune
3

$ nodeset --expand azur[1-10]
azur1 azur2 azur3 azur4 ... azur10

$ nodeset --fold azur5 azur7 azur6 azur9
azur[5-7,9]
```

- Exemple

```
$ for i in $(nodeset --expand azur[1-10])
do
   echo $i | ...
done
```

ClusterShell : CLI

- Opérations ensemblistes sur ces listes :
  - Union

```
$ nodeset -f azur[1-10] azur[8-12]
azur[1-12]
```

  - Exclusion

```
$ nodeset -f azur[1-10] -x azur[8-12]
azur[1-7]
```

  - Intersection

```
$ nodeset -f azur[1-10] -i azur[8-12]
azur[8-10]
```

  - XOR

```
$ nodeset -f azur[1-10] -X azur[8-12]
azur[1-7,11-12]
```

ClusterShell : CLI

- Stepping
  - Déclarer des *ranges* avec un pas :

```
$ azur[1-10/3] <=> azur[1,4,7,10]
```

  - Détection automatique d'un pas : *autostepping*

```
$ nodeset -f azur1 azur[5,9] azur13 azur17
azur[1,5,9,13,17]
$ nodeset -f --autostep=3 azur1 azur[5,9] azur13 azur17
azur[1-17/4]
```

  - Exemples :

```
$ nodeset -e azur[0-10/2 «  tous les noeuds pairs
$ # some specific files
$ nodeset -e /dev/mapper/vol[1-9/3]  -> 1,4,7
$ nodeset -e /dev/mapper/vol[2-9/3]  -> 2,5,8
$ nodeset -e /dev/mapper/vol[3-9/3]  -> 3,6,9
```

ClusterShell : CLI

- Splitting

```
$ nodeset --split=3 -f azur[1-10]
azur[1-4]
azur[5-7]
azur[8-10]
```

- Séparateur
  - Lors d'un affichage *expand* (-e), le séparateur peut être spécifié via -S

```
$ nodeset -S "\n" -e azur[1-3]
azur1
azur2
azur3
```

ClusterShell : CLI

- Rangeset
  - Manipulation de liste d'entier avec les mêmes fonctionnalités
  - Avec l'option -R

```
$ nodeset -R -e 1,5-7,10
1 5 6 7 10
$ nodeset -R -f 1-10 -i 8-12 -X 5
5,8-10
```

  - Exemple : Qui répond au ping ?

```
# ping -b  192.168.1.255 -c 5 | awk '/icmp_seq/ { print $4 }'  |   \
     sed -e 's/192.168.1.1//' -e 's/://'  |   nodeset -R -f
01-13
```

ClusterShell : CLI

- Il est possible de déclarer des formules, une arithmétique
- Un caractère par opération :
  - Union « , »
  - Intersection « & »
  - Exclusion « ! » (sauf … , not ...)
  - XOR « ^ »
- La priorité se fait de gauche à droite
- Exemples :

```
azur[1-10],azur[5-12]
  azur[1-12]
azur[1-10]!azur[8-12],azur9
  azur[1-7,9]
```

ClusterShell : Groupes

- Un groupe représente une liste de machines. Il se préfixe d'un @

```
$ nodeset -f @foo
azur[1-12]
```

- Il peut être utilisé à la place de n'importe quel nom

```
$ nodeset -f @foo,mars,lune,@bar,azur[1-12]
```

- La liste des groupes peut être obtenue facilement :

```
$ nodeset -l
@foo
@bar
@baz
$ nodeset -ll
@foo azur[1-6]
@bar azur[7-9]
@baz azur[10-12]
```

ClusterShell : Groupes

- Les groupes s'obtiennent à partir de *sources*
- Une source est définie à partir de 4 commandes externes
  - map : Convertit un nom de groupe en liste de éléments
  - list : Liste tous les noms de groupes disponibles
  - all : Renvoie une liste d'éléments
  - reverse : Convertit une liste d'éléments en nom de groupe
- Seul *map* est obligatoire.
- Les définitions de groupes sont récursives.
- Les sources se définissent dans */etc/clustershell/groups.conf*

```
[main]
default: getent

[getent]
map:  awk -F: -v grp=$GROUP '($1 == grp) {print $4}' /etc/group
list: awk -F: -v grp=$GROUP '($1 == grp) {print $1}' /etc/group
```

ClusterShell : Groupes

- Fichier de groupe personnalisé simple avec Awk.

```
$ cat /etc/clustershell/mygroups
foo: azur[1-5]
bar: azur[6-8]
qux: azur[9-12]
```

- Map : nom vers noeuds

```
$ awk -F: '/^$GROUP:/ {print $2}' /etc/clustershell/mygroups
```

- List : liste des groupes

```
$ awk -F: '/^$GROUP:/ {print $1}' /etc/clustershell/mygroups
```

- All: un groupe qui représente tous les noeuds
  - C'est une convention. Vous pouvez y mettre la valeur que vous souhaitez.
  - Les groupes sont récursifs : un simple « echo @foo » suffit.
  - S'il n'est pas définit et que map et list le sont, l'union de tous les groupes est utilisés.
  - Accessible via l'option -a : nodeset -f -a

ClusterShell : Groupes

- Reverse
  - Permet de convertir une liste de machines en groupes
  - Seulement si le groupe est complet
  - Utilise en priorité les plus grands groupes
  - Utilise le call *reverse* si définit.
    - S'il est manquant, utilise *list* et *map*
- Exemple :

```
$ cat /etc/clustershell/mygroups
foo: azur[1-5]
bar: azur[6-8]
qux: azur[9-12]
$ nodeset -r azur[1-12]
@foo,@bar,@qux
```

ClusterShell : Groupes

- nombre illimité de sources :

```
$ nodeset –groupsources # give me all sources
```

- Toujours une source par défaut

```
$ nodeset -f @foo # group foo from default source
```

- Sinon <nom de la source:groupe>

```
$ nodeset -f @local:foo # group foo from source local
```

- Les sources peuvent être mélangées

```
$ nodeset -f @local:foo&@other:bar,@last:baz
```

- La source par défaut peut être surchargée

```
$ nodeset -s av -l
```

ClusterShell : Groupes

- Documentation
  - Man pages

```
$ man 1 nodeset
$ man 5 groups.conf
```

  - Site Web
    - http://clustershell.sourceforge.net
    - https://clustershell.readthedocs.io/en/latest
    - http://sourceforge.net/apps/trac/clustershell/wiki/NodeGroups

ClusterShell : Clush

- Exécute en parallèle des commandes sur des machines distantes

```
$ clush -w azur[1-4] echo ok
azur2: ok
azur1: ok
azur4: ok
azur3: ok
```

- Copie des fichiers depuis et vers des machines distantes

```
$ clush -w azur[1-4] --copy /etc/passwd
```

- Configuration :
  - Globale : `/etc/clustershell/clush.conf`
  - Par utilisateur (ignore le fichier global) : `~/.clush.conf`

ClusterShell : Clush

Selection des machines

- -w: machines à utiliser
- -x: machines à exclure

```
$ clush -w azur[1-12] -x azur[8-12] echo ok
```

- -a : Utilise le call *all* du *groupsource*

```
$ clush -a echo ok
$ clush -a -x @test echo ok
```

- Les *extended patterns* sont supportés

```
$ clush -w azur[1-12]&foo[4-5] echo ok
```

ClusterShell : Clush

Selection des machines

- Donc les groupes aussi

```
$ clush -w @av:prod&@compute echo ok
```

- Options pour les groupes
  - Pas de « @ »
  - *extended patterns* supportés.
  - Groupes à utiliser : -g
  - Groupes à exclure : -X
  - Exemples

```
$ clush -g prod,test!hardware ...
$ clush -g chassis[15-16] -X test ...
```

ClusterShell : Clush

- Contrôle de la charge : fanout
  - Contrôle le nombre d'exécutions concurrentes

```
$ clush -f 128 -w azur[1-2000] ...
```

  - Régler sa valeur par défaut en fonction des performances de la machine

```
$ time clush -f xxx -w azur[1-2000]
```

- Timeouts
  - Connect timeout : Temps pour établir la connexion SSH : -t

```
$ clush -w ... -t 5 echo ok
```

  - Command timeout : Temps global d'exécution d'une commande, incluant le temps de connexion.

```
$ clush -w  ... -u 5 echo ok
```

- Passage de paramètre à la commande SSH

```
$ clush -w ... -o '-oPort=2022' echo ok
```

ClusterShell : Clush

- Affichage
  - Affiche en couleur par défaut, mais désactivable

```
$ clush -b –color=never ….
```

  - Regroupement par affichage identique : -b

```
$ clush -b -w azur[1-4] echo ok
---------------
azur[1,3]
---------------
ok
azur2: ssh: connect to host azur2 port 22: No route to host
azur4: ssh: connect to host azur4 port 22: No route to host
clush: azur[2,4]: exited with exit code 255
```

  - Fusion des sorties standards et d'erreurs : -B

ClusterShell : clush

- Affichage triée par noeud: -L

```
$ clush -L -w azur[1-4] echo ok
azur1: ok
azur2: ok
azur3: ok
azur4: ok
```

- Regroupement ligne par ligne : -bL / -BL

```
$ clush -bL -w azur[1-4] cat /etc/motd
azur[1-4]: Welcome on CEA computing complex
azur2: This is curie machin
azur3: This is titane machine
azur4: This is platine machine
```

ClusterShell : clush

- Différence

```
$ clush --diff -w azur[1-4] echo ok
azur1: ok
azur2: ok
azur3: ok
azur4: ok
```

ClusterShell : clush

- Copie de fichiers ou de répertoires
  - Local vers distant

```
$ clush -w azur[1-12] -c /etc/passwd ...
$ clush -w azur[1-12] -c /etc/passwd --dest
/etc/passwd.new
```

  - Distant vers local
    - Chaque version du fichier distant est copié localement et préfixé du nom de la machine d'origine :

```
$ clush -w azur[1-12] --rcopy /etc/motd
$ ls -m /etc/motd.*
/etc/motd.azur1, /etc/motd.azur2, /etc/motd.azur3, ...
```

  - Il est possible de changer la destination :

```
$ clush -w azur[1-12] --rcopy /etc/motd --dest /tmp
```

ClusterShell : clush

- Mode interactive
  - Si aucune commande n'est précisée, *clush* se lance en mode interactif
  - Chaque ligne de commande est exécutée sur tous les noeuds spécifiés

```
$ clush -w azur[1-3]
Enter 'quit' to leave this interactive mode
Working with nodes : azur[1-3]
clush> echo ok
azur2: ok
azur1: ok
azur3: ok
clush> ...
```

  - Il est possible sans quitter le mode,
    - de modifier le mode d'affichage : « = »
    - de modifier les noeuds courants

```
clush> +azur5
clush> -azur3
```

ClusterShell : clush

- Documentation
  - Man pages

```
$ man 1 clush
$ man 5 clush.conf
```

Outils Infra

PowerMan

- Logiciel de contrôle de PDU (Power Distribution Units)
- Conçu pour un environnement cluster
- Extensible par un système de greffons (pluggins)
- Supporte SNMP, IPMI, et autres
- Permet d'unifier la vue et le contrôle des alimentations

ConMan

- Logiciel de gestion de consoles
- Conçu pour un environnement cluster
- Maintient une connexion persistance avec les consoles des équipements
- Supporte différents types de connexions
- Permet d'enregistrer tous les messages envoyés sur une console dans un fichier
- Permet la connexion interactive à une console

Questions ?