

PROJET TUTEURÉ :
CONFIGURATION AUTOMATIQUE D'UN CLUSTER DE
CALCUL AVEC PUPPET

André DIMITRI
Quentin DEXHEIMER
Riwan BLONDE

23 mars 2012

Table des matières

1 Le Grid'5000	2
1.1 Présentation du Grid'5000	2
1.2 Les outils du Grid'5000	3
1.2.1 OAR2	3
1.2.2 Kadeploy3	3
1.2.3 Taktuk	4
1.2.4 KaVLAN	4
1.2.5 Les autres outils	5

Chapitre 1

Le Grid'5000

1.1 Présentation du Grid'5000

Le Grid'5000 est une grille informatique destinée à la recherche scientifique, la plateforme Grid'5000 fait partie de l'action de développement technique Aladin . Le plateforme a vu le jour en 2003 et a pour but de promouvoir la recherche sur les grilles informatiques en France. Le Grid'5000 est aujourd'hui composé de 1200 noeuds répartis sur 9 sites différents situés en France et au Luxembourg et interconnectés avec le réseau Réseau National de télécommunications pour la Technologie l'Enseignement et la Recherche (RENATER). L'objectif du Grid'5000 est de permettre aux scientifiques d'effectuer des expériences dans le domaine des systèmes informatiques et des réseaux distribués dans un environnement hétérogène aussi proche de la réalité que possible.

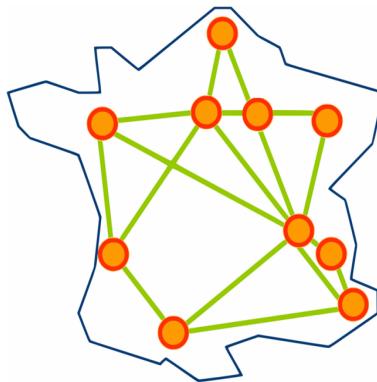


FIGURE 1.1 – Localisation des différents sites du Grid5000

Le Grid'5000 est donc composé de plusieurs sites distincts mais l'organisation au sein de chaque site est la même. Chaque site est composé d'un ou plusieurs clusters, c'est à dire un ensemble de machine homogène. Le site de Nancy par exemple héberge deux clusters nommés Griffon et Graphene. A l'intérieur de chaque cluster se trouvent des ordinateurs aussi appelés "nodes" ou "noeuds". Il existe deux types de nodes : les noeuds de services et les noeuds de travail, sur lesquels sont effectuées les expériences.

Les noeuds de services servent à l'administration des machines situées dans les cluster et à l'accès aux hôtes virtuels pour les administrateurs. Les machines de services sont commune aux différents clusters situées sur les sites. Certains noeuds de services appelés "frontends" sont utilisés par les utilisateurs pour accéder aux différents sites grâce au protocole ssh, la réservation de noeuds et le déploiement.

L'administration de la plateforme est centralisée et est assurée par 8 personnes parmi elle certains travaille à plein temps sur la plateforme alors que d'autres non.

Il existe donc une hiérarchie dans la plateforme Grid'5000. Au sommet de cette hiérarchie se trouve la plateforme en elle-même ensuite les différents sites géographique. Au niveau inférieur on trouve les différents services et clusters puis les noeuds. Ensuite on trouve les CPUS et enfin les cores.

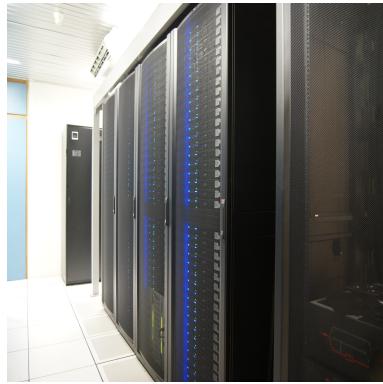


FIGURE 1.2 – Un des clusters de Nancy : Griffon

1.2 Les outils du Grid'5000

Le Grid'5000 est composé de plusieurs services : une partie ces services ont été développés uniquement pour cette plateforme comme par exemple Kadeploy3 qui a été développé par l'INRIA Nancy - Grand Est ; les autres sont des services standard déjà utilisés sur les systèmes Unix.

1.2.1 OAR2

OAR est un gestionnaire de ressources pour de grandes grilles informatique. Il est écrit en PERL et s'appuie sur une base SQL (PostgreSQL ou MySQL). OAR permet le déploiement, la réservation et le management des noeuds et des jobs¹.

Les principales fonctionnalités d'OAR sont les suivantes :

- Soumission : Le système décide du moment où votre travail commence afin d'optimiser l'ordonnancement global de la plateforme. S'il n'y a pas de noeud disponible, le travail est mis en attente. (correspond à oarsub-I ou oarsub scriptName syntaxes)
- Réservation préalable : On décide alors quand le travail doit commencer, les ressources fournies seront disponibles à la date spécifiée. Si les ressources demandées sont disponibles au moment du début du job la réservation est effectuée sinon la réservation échoue et il faut alors modifier les paramètres de la réservation. (correspond à oarsub-r date ou oarsub-R Date syntaxes scriptName)
- Mode interactif : On n'exécute pas de script lors de la soumission ou lors de la réservation mais on choisit de travailler de façon interactive. (correspond à oarsub-I pour la soumission ou oarsub-r jour ; oarsub-C jobid pour la réservation)
- Mode passif : On choisit d'exécuter directement un script sur les noeuds réservés. Il n'est alors pas obligatoire de se connecter sur les noeuds mais cela est toujours possible en utilisant oarsub-C jobid). (correspond à oarsub scriptName pour la soumission ou oarsub -r date scriptName pour la réservation)
- Type de job possible :
 - défaut : on utilise juste l'environnement par défaut des noeuds.
 - déploiement : on déploie un système d'exploitation définis lors de la réservation. Cette méthode utilise l'outils Kadeploy.
 - Il existe un autre type de job , destinée aux utilisateurs avancées qui utilise un mode d'utilisation différents.

1. un job correspond à une tâche affectée à une réservation

1.2.2 Kadeploy3

Kadeploy est un outil qui permet le déploiement des différents systèmes d'exploitation sur les noeuds de la plateforme. Il permet aussi de configurer les noeuds, de les cloner et de les manager. Il permet le déploiement de système Linux, BSD, Windows et Solaris.

1.2.3 Taktuk

Taktuk est un outil complémentaire à OAR2. Taktuk permet l'exécution de commandes à distance sur un grand nombre de noeuds hétérogènes. Il met en place un liaison entre la "frontends" et les noeuds concernés par la commande et s'adapte à l'environnement de la machine (les performances , la charges , le réseau ...).

Exemple :

```
1 taktuk -l root -s -m $puppet_master broadcast exec [ apt-get -q -y install
puppet facter puppetmaster ]
```

L'option -l permet d'utiliser le compte root pour l'installation.

L'option -m permet de déployer sur une seule machine alors que l'option -f permet de déployer sur toute les machines contenu dans le fichier.

Cette commande permet donc de déployer les paquets puppet, facter et puppetmaster sur la machine puppet_master.

1.2.4 KaVLAN

KaVLAN est un outil qui a pour but de permettre la mise en place d'un VLAN sur des noeuds du Grid5000. Il permet la mise en place de plusieurs type de VLAN : local, "router" et global. KaVLAN peut être utilisé en complément de Kadeploy et de OAR pour certains types d'expérimentation.

- Un KaVLAN local est un VLAN complètement isolé du reste de la plateforme Grid5000. Il est alors obligatoire d'utiliser une gateway pour accéder aux noeuds se trouvant à l'intérieur de VLAN.
- Un KaVLAN "router" permet l'accès à tous les noeuds du VLAN depuis le reste du Grid5000 sans utiliser de gateway.
- Un KaVLAN global est un VLAN qui est disponible sur tous les sites du Grid5000. Un routeur est alors configuré sur le site où le VLAN a été configuré.

Exemple :

```
1 oarsub -I -t deploy -l {"type='kavlan-local'}/vlan=1+/nodes=$nbr_nodes ,
walltime=$tmps -n "$nom"
```

La commande suivante permet de réserver un nombre donné de noeuds en utilisant un VLAN local. Kavlan dispose d'autres commande utile comme :

- La commande qui permet d'avoir la liste des machines qui se trouvent à l'intérieur d'un VLAN pour un job donné est :

```
1 kavlan -V -j JOBID
```

- Celle qui permet d'activer le dhcp interne à KaVLAN est :

```
1 kavlan -e
```

- La commande qui permet de désactiver le dhcp est :

```
1 kavlan -d
```

1.2.5 Les autres outils

Le Grid5000 utilise pour son fonctionnement d'autres outils :

- une base de donnée MySQL . Cette base de donnée va être utilisée par Kadeploy et OAR.
- un serveur DNS (Domaine Name Server) qui permet de faire la correspondance entre les adresses ip et les nom de domaine.
- un serveur DHCP (Dynamic Host Configuration Protocol) qui permet la configuration automatique des paramètres IP des machines.
- un serveur NFS (Network File System) pour permettre aux utilisateurs de stocker des informations dans leur /home sur les différents sites.
- des outils de supervision :
 - Nagios : surveille les hôtes et services spécifiés, alertant lorsque les systèmes vont mal et quand ils vont mieux.
 - Ganglia : permet de superviser des clusters et des grilles informatiques.
 - Munin : présente ses résultats sous forme de graphiques disponibles via une interface web.
 - Cacti : mesure les performances réseau et serveur.
- un serveur weathermap : qui permet la visualisation du réseau sous forme d'une carte.
- un serveur syslog : qui permet les journaux d'évènements.
- un annuaire LDAP (Lightweight Directory Access Protocol) : permet le stockage d'informations et de données.
- un serveur web Apache avec un proxy.
- un serveur Squid : un proxy/reverse proxy
- un serveur NAT (Network Address Translation) : permet de faire correspondre une seule adresse externe publique visible sur Internet à toutes les adresses d'un réseau privé
- un serveur mail.
- un dépôts de paquets et de logiciels.
- API Rest g5k : utiliser par KaVLAN, Kadeploy et UMS (User Management System). Elle permet d'utiliser toute les fonctions du Grid'5000 et de les automatiser.
- une infrastructure de virtualisation XEN comportant de 2 à 5 dom0 et environ 30 domU avec un service par domU.