gLite sur Grid'5000: vers une plate-forme d'expérimentation à taille réelle pour les grilles de production

Sébastien Badia et Lucas Nussbaum
{sebastien.badia,lucas.nussbaum}@loria.fr

ALGORILLE team, LORIA
INRIA Nancy - Grand Est & Nancy-Université

1 Overview

The Grid has become a huge and important instrument, playing a key role in the everyday work of many researchers. A large amount of software is being developed both to manage the grid infrastructure itself (gLite middleware), to facilitate the task of grid users (e.g workflow managers, pilot job managers, etc.), and to run the computations. That software must be designed to handle network and services outages in a highly distributed environment, while still providing the expected performance. It is inconvenient to test software using the production infrastructure, since (1) it might not exhibit the behaviour that is required to test extreme conditions (services are unlikely to crash as often as required when testing fault tolerance) ; (2) it might not be possible to replace key parts of the infrastructure without degrading the user experience ; (3) experiments are not easily reproduced in production conditions.

In this paper, we present our ongoing work on deploying the gLite middleware on the Grid'5000 testbed, a scientific instrument designed to support research on parallel, large-scale and distributed computing. Tools were written to automatize the deployment of the gLite middleware on several sites and clusters, resulting in a deployment of 5 sites and 8 clusters on 441 machines in less than an hour. This provides a solid basis for future experiments on the gLite middleware and on software that interact with the middleware.

2 Enjeux scientifiques

La grille EGI est devenue une immense et importante plate-forme, qui joue un rôle clé dans la travail quotidien de nombreux chercheurs. Un grand nombre de logiciels ont été développés pour gérer l'infrastructure elle-même (middleware gLite[1]), pour faciliter le travail des utilisateurs (moteurs de workflows comme MOTEUR [2], gestionnaires de jobs pilotes, etc.), et pour exécuter les traitements eux-mêmes. Ces logiciels doivent être conçus pour prendre en compte les pannes à différents niveaux (réseaux, services), dans une infrastructure largement distribuée, tout en fournissant les performances requises.

Il est en général difficile de tester ces logiciels en utilisant l'infrastructure de production. D'une part, il est peu probable que l'infrastructure de production fournisse le comportement souhaité pendant les tests : lors du test de la résistance aux pannes d'un logiciel, il serait peu confortable d'attendre une panne sur l'infrastructure de production pour vérifier son bon fonctionnement. D'autre part, il n'est pas forcément possible de remplacer un composant central de l'infrastructure pour en tester une version modifiée sans gêner les utilisateurs.

Dans ce travail, nous proposons d'utiliser la plate-forme Grid'5000 [3, 4], dédiée à la recherche sur les systèmes et le calcul parallèle, pour tester l'infrastructure logicielle des grilles de production.

3 Développements et outils

La plate-forme Grid'5000 est composée de 1700 machines (7000 coeurs) répartis dans 10 sites en France. Elle dispose de fonctionnalités de reconfiguration du matériel et du réseau : les utilisateurs peuvent, après avoir réservé des ressources, réinstaller les machines réservées avec le système requis pour leur expérience.

Cette fonctionnalité a été utilisée pour développer un ensemble de scripts permettant d'automatiser le déployement d'une infrastructure gLite sur Grid'5000, composée de :

1. une VO et son VOMS (*Virtual Organization Membership Service*), annuaire des utilisateurs ;
2. plusieurs sites, composés de :
   (a) un BDII (*Berkeley Database Information Index*), annuaire des ressources disponibles sur chaque site ;
   (b) un CE (*Computing Element*), service de soumission des tâches à un site de calcul donné, et un ou plusieurs clusters composés de noeuds de calcul (WN) ;
   (c) un BATCH, ordonnanceur ; dans le cadre de notre travail, le couple Torque/Maui a été utilisé ;
   (d) une UI (*User Interface*), interface d'accès pour les utilisateurs.

Le processus est composé de différentes étapes implémentées dans des scripts qu'il est possible d'enchainer automatiquement. D'abord, les machines réservées sont réinstallées avec une installation Scientific Linux 5.5 minimale. Puis le dépôt RPM de gLite (*glitesoft.cern.ch*) est ajouté, et les paquets nécessaires sont installés en fonction du rôle choisi pour la machine cible (VOMS, BDII, noeud de travail, ...). Enfin, l'ensemble des noeuds sont configurés.

Afin de générer les certificats utilisés pour identifier les machines et les utilisateurs, nos scripts créent une autorité de certification à l'aide des scripts *simpleCA* disponibles dans l'archive Globus [1]. Il est ainsi possible de générer et signer automatiquement les certificats nécessaires au déploiement de notre infrastructure.

Deux difficultés importantes ont été rencontrées lors de la mise au point des scripts :

– S'il existe une documentation abondante sur la configuration de gLite, les différents documents sont de qualité inégale, parfois dépassés ou incomplets. Il est dommage qu'il n'y ait pas plus d'effort de centralisation de la documentation, ce qui améliorerait sa qualité.

1. http ://www.globus.org/toolkit/docs/4.0/admin/docbook/ch07.html

– Le processus d'installation de gLite est prévu pour être réalisé à la main. L'automatiser, et en particulier automatiser la gestion des certificats, a été loin d'être évident.

Les scripts utilisés sont disponibles et documentés sur `https://github.com/sbadia/gdeploy/`.

4 Résultats scientifiques

Avec les scripts développés, nous avons réalisé le déploiement de gLite sur 441 machines de Grid'5000, réparties dans 5 sites et 8 clusters. Le déploiement de l'environnement Scientific Linux sur l'ensemble des machines avec Kadeploy a pris 10 minutes, et la configuration de l'ensemble des noeuds gLite a pris 40 minutes.

5 Perspectives

Ce travail vise à démontrer la faisabilité de l'utilisation de Grid'5000 pour tester ou évaluer des logiciels de l'écosystème des grilles de production.

Nos scripts peuvent être assez largement améliorés. Une première étape est de permettre de configurer une plate-forme composée d'un nombre de VOs, de sites et de clusters arbitraires afin de pouvoir se placer dans des configurations expérimentales plus variées et intéressantes.

Concernant la durée de déploiement, nous explorons actuellement l'utilisation d'un moteur de workflow pour orchestrer les différentes étapes du déploiement, ce qui permettra d'extraire le parallélisme implicite entre les différentes étapes et devrait significativement réduire la durée du déploiement. D'autres améliorations sont également prévues, comme l'ajout d'un cache pour les paquets RPM de Scientific Linux ce qui réduira la charge sur les serveurs *proxy* HTTP de Grid'5000, et la modification de l'image Scientific Linux afin qu'elle puisse fonctionner sur l'ensemble des clusters de Grid'5000.

De plus, l'infrastructure actuellement déployée est relativement simple. Nous ne déployons pour l'instant pas de services de gestion de stockage ni de services annexes (Monitor, par exemple). La gestion des VO et des utilisateurs est également assez simpliste.

Ce travail n'est pas une fin en soi. Nous cherchons des utilisateurs ou développeurs de logiciels pour la grille ayant des besoins en expérimentation de leurs logiciels. En particulier, nous envisageons les cas d'utilisation suivants :

– *Expériences sur des évolutions de composants du middleware gLite*. Ces expériences pourraient être réalisées à grande échelle, sans affecter les utilisateurs de la grille ;
– *Expériences sur des outils interagissant avec le middleware gLite*. Grâce à l'utilisation d'une infrastructure distincte, il sera possible de simuler des pannes de services, d'injecter de la charge applicative, et de soumettre un grand nombre de jobs factices sans perturber le travail des autres utilisateurs.

6 Documentations utilisées

- http://glite.cern.ch/admin_documentation
- http://www.gridpp.ac.uk/wiki/Main_Page
- https://twiki.cern.ch/twiki/bin/view/LCG/WebHome
- http://www.globus.org/toolkit/docs/4.1/admin/docbook/gtadmin-simpleca.html
- http://www.nordugrid.org/papers.html

Remerciements

Ce travail a été en partie financé par le thème EDGE du CPER MISN de la Région Lorraine et par l'appel *Interfaces Recherche en grilles - Grilles de production* de l'Institut des Grilles du CNRS et de l'action INRIA Aladdin-G5K.

Références

[1] gLite - lightweight middleware for grid computing. http://glite.cern.ch/.

[2] Tristan Glatard, Johan Montagnat, Diane Lingrand, and Xavier Pennec. Flexible and efficient workflow deployement of data-intensive applications on grids with MOTEUR. *International Journal of High Performance Computing Applications*, 22(3) :347–360, August 2008.

[3] Franck Cappello, Frédéric Desprez, Michel Dayde, Emmanuel Jeannot, Yvon Jégou, Stephane Lanteri, Nouredine Melab, Raymond Namyst, Pascale Primet, Olivier Richard, Eddy Caron, Julien Leduc, and Guillaume Mornet. Grid'5000 : a large scale, reconfigurable, controlable and monitorable Grid platform. In *6th IEEE/ACM International Workshop on Grid Computing - GRID 2005*, Seattle, USA, November 2005. Grid 2005 held in conjunction with SC'05.

[4] Grid'5000 website. https://www.grid5000.fr/.