Point technique léger sur l'activité de CiGri

Nicolas Capit, Laurent Desbat
TIMC-ID-IMAG, UMR CNRS 5525,
In3S, faculté de Médecine, UJF
38706 La Tronche France,
Nicolas.Capit@imag.fr, Laurent.Desbat@imag.fr

décembre 2003

1 Introduction

Le projet CIGRI a pour but de fournir aux utilisateurs la plus grande puissance de calcul disponible sur CIMENT pour des applications multi-paramétriques trivialement parallèlisables. Elles ont pour caractéristiques communes de mettre en oeuvre de très grands nombres de jobs et de nécessiter énormément de ressources CPU. Pour traiter ces applications réelles et dimensionnantes, nous utilisons les processeurs disponibles des clusters de calcul localisés dans les différents pôles de la communauté CIMENT. Le principe est donc de regrouper les clusters en une grille légère de calcul.

Les expériences des projets tels que Globus ou DataGrid nous ont montré qu'il est complexe de vouloir fédérer des sites très distants et possédant des politiques d'administration différentes. C'est pour cela que nous avons décidé de nous orienter vers un modèle de grille légère avec des clusters basés dans le bassin grenoblois (proximité spatiale des acteurs).

L'activité scientifique de recherche en informatique distribuée relative à la mise en place de la grille légère CiGri est essentiellement décrite dans le rapport du laboratoire ID ci-joint dans le cadre du projet "grappe de 200PC". L'utilisation de la grille légère sera plus précisément décrite l'an prochain. L'objet de cette courte note est de faire un point très pratique sur l'état de la grille CiGri et les fonctionnalités disponibles.

2 Point technique CiGri

Dans un premier temps nous avons mis au point une solution pour un utilisateur afin qu'il puisse se servir, pour son application, des processeurs disponibles parmi les clusters de calcul. Ce premier système, assez rustique, est basé sur des outils développés spécifiquement, permettant de gérer un

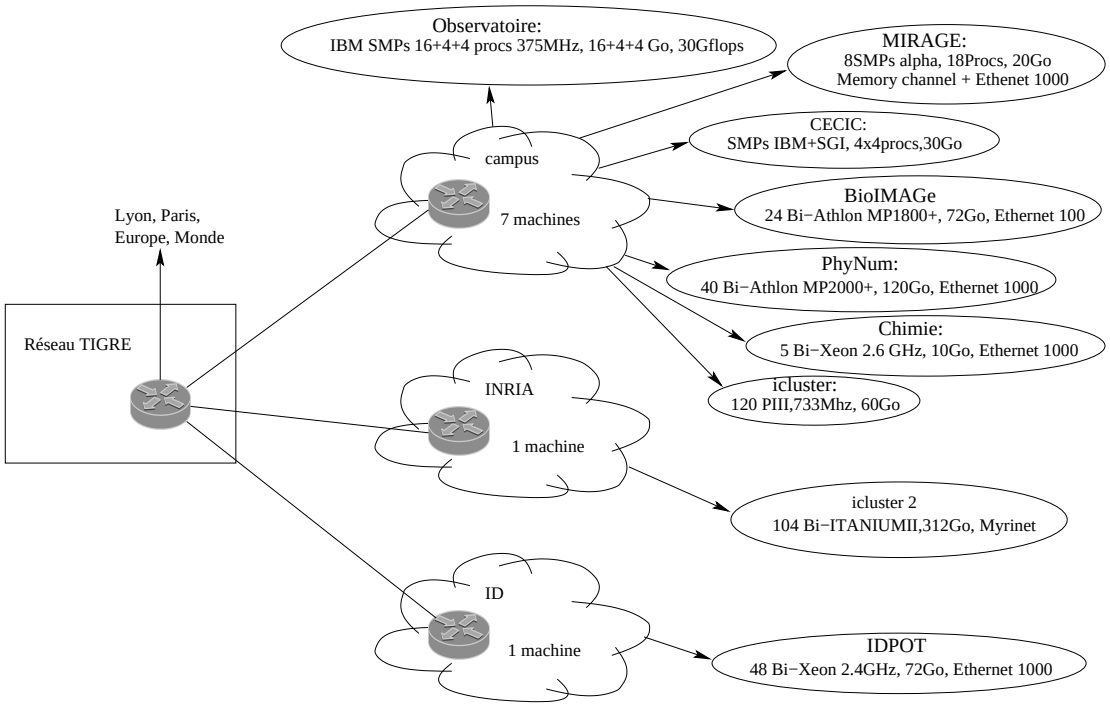

Figure 1: Répartition des clusters CIMENT

grand nombre de jobs et de scripts bash. Cela nous a permis d'identifier certaines contraintes telles que la gestion des erreurs ou la prise en compte d'un grand nombre de fichiers.

Après cette première expérience nous avons décidé de développer un mécanisme simple, modulaire, s'appuyant sur une base de données et exploitant les gestionnaires batch des clusters (PBS, OAR). Nous sommes actuellement dans une phase de tests de notre outil "cigri" de soumission d'un grand nombre de jobs sur la grille légère de calcul CIMENT. Ces essais se font en collaboration avec des utilisateurs sur des applications réelles et de grande dimmension. Elles nécessitent de très grands temps CPU (quelques dizaines à quelques centaines de milliers d'heures CPU sur des processeurs modernes) et mettent en oeuvre de nombreux jobs (plusieurs dizaines de milliers à plusieurs centaines de milliers) identiques sur des jeux de paramètres différents. CiGri offre à l'utilisateur une interface lui permettant soummettre ses campagnes de job. Pour cela, il doit remplir un fichier de description de sa campagne de jobs selon un jdl "job description langage" que nous avons mis au point.

Exemple de la simulation d'echo radar, campagne.jdl :

```
DEFAULT{
    paramFile = echos.param ;
}
tomte.ujf-grenoble.fr{
    execFile = /home/nis/lpg-grid/runTime.sh ;
}
bioimage.imag.fr{
    execFile = /home/lpg-grid/runTime.sh ;
}
```

Commentaires : Donc l'application "runTime.sh" va pouvoir être lancée sur deux clusters (tomte et bioimage) et le fichier de paramètres sera echos.param. L'utilisateur soumets ses jobs avec la commande "gridsub campagne.jdl" qui lui retourne un "campagneID".

CiGri offre à l'utilisateur des outils de suivi de ses jobs. L'utilisateur peut savoir à tous moments l'état de ses jobs : "terminés", "en attente" ou "en cours d'exécution", la liste des clusters sur lesquels il calcule, des statistiques de charge de ces clusters.

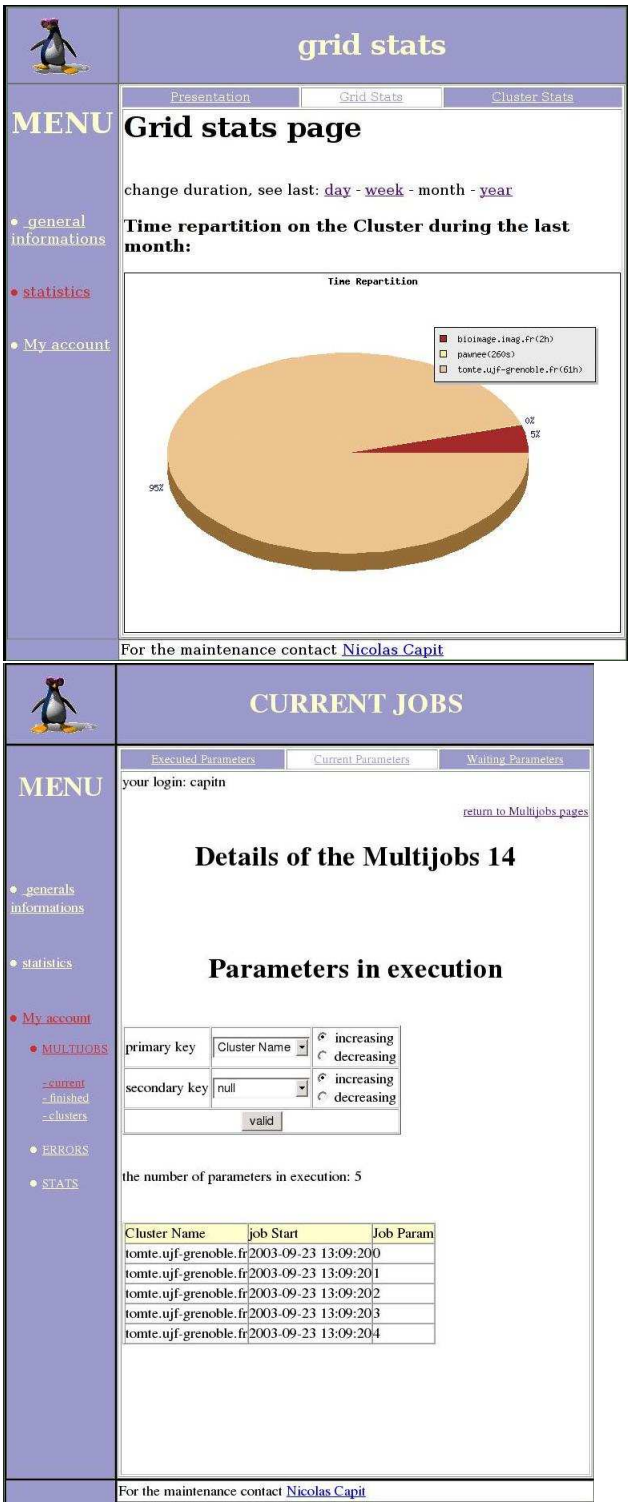

Figure 2: Illustration de l'interface web

L'utilisateur peut à tout moment stopper sa campagne de job avec la commande "griddel campagneID". Pour relancer une campagne qui a été stoppée, il suffit de mettre à jour la liste des paramètres. Actuellement, ceci est laisser à la charge de l'utilisateur.

Différentes stratégies d'ordonnancement des tâches sur la grille légère sont actuellement en cours d'expérimentation. Cette activité est décrite dans le rapport ci-joint du laboratoire ID dans le cadre du projet de grappes de 200 PC. Nous nous efforçons de rendre ces expérimentations les moins intrusives possibles sur les grappes de calcul pour ne pas perturber leur exploitation. Ceci permet l'utilisation des clusters ainsi que celle de la grille légère sur des calculs réels conjointement à l'expérimentation de l'efficacité de la distribution des calculs.

3 Le projet OAR

Ce projet a débuté en début d'année 2003 au sein du laboratoire ID-IMAG. C'est une expérience autour d'une nouvelle conception d'un gestionnaire de travaux pour grappe. En effet les codes de gestion de batch actuels (et accessibles) sont complexes et ne permettent pas facilement d'implémenter de nouvelles fonctionnalités. Notre gestionnaire de batch doit permettre aux utilisateurs de réserver de la puissance de calcul sur une ferme d'ordinateurs. Très rapidement les problématiques de "scheduling" apparaissent et doivent être abordés. OAR est entièrement modulaire et permet d'implémenter des algorithmes d'ordonnancement de façon simplifiée. Ainsi, dans l'environnement CIMENT, les spécialistes de l'ordonnancement peuvent tester leurs différents algorithmes de placement sur de "réelles" machines avec de "réels" utilisateurs. Grâce à OAR, il a été possible d'implémenter des fonctionnalités que nous ne retrouvions pas dans les autres logiciels. Par exemple, la notion de tâches "best effort" a été implémentée. Il s'agit de tâches non prioritaires dont l'execution ne doit pas perturber pas les autres (elles sont automatiquement tuées si un job d'un autre type veut de la puissance). Cela permet d'utiliser au mieux la puissance de la grappe pour ces tâches, sans déranger les utilisateurs ni la politique d'administration définies sur un cluster pour les autres types de tâche. Bien entendu, le type de tâche “best effort” est utilisé dans le logiciel CIGRI pour que les jobs grilles soient les moins intrusifs possibles sur les clusters des pôles CIMENT.

A ce jour OAR est installé sur 4 grappes différentes : - tomte (CECIC) - IDPOT (ID-IMAG) - BioImage (TIMC-IMAG) - Icluster2

et bientôt PhyNum...

OAR est dans un état fonctionnel et nous sommes en train de réaliser un travail sur sa documentation et sur sa diffusion auprès des utilisateurs.

Il est possible de trouver plus d'informations sur le site : http://oar.imag.fr

Voici quelques exemples des commandes de base:

```
# oarsub -l nodes=5 job.sh
```

Permet de lancer la tâche "job.sh" avec 5 noeuds réservés.

```
# oardel 2341
```

Permet de supprimer le job 2341.

```
# oarstat
```

Permet d'obtenir des informations sur les jobs qui s'exécutent sur le cluster.

4 Résultats préliminaires

OAR et CiGri ont été déployés sur respectivement 4 et 3 clusters de CIMENT. Ces logiciels sont utilisés par des utilisateurs sur des machines en exploitation. Inversement, les recherches en informatique distribuées bénéficient d'un cadre réel de validation.

Dans le mode actuel de la grille légère CiGri, environ 100000 tâches, chacune de 5 à 15 minutes de CPU, ont été traitées (simulation d'échos radar de Jean-François Nouvel). De grandes campagnes de jobs, sur cette problèmatique, sur la chimie quantique et sur l'imagerie médicale, devraitent très rapidement démarrer et seront décrites dans le rapport d'activité de CiGri en octobre 2004.