



HAL
open science

Les grilles de calcul au service de la physique médicale

C. O. Thiam, D. Donnarieix, Vincent Breton, Z. El Bitar, L. Maigne

► **To cite this version:**

C. O. Thiam, D. Donnarieix, Vincent Breton, Z. El Bitar, L. Maigne. Les grilles de calcul au service de la physique médicale. 44èmes Journées Scientifiques de la Société Française de Physique Médicale (SFPM), Jun 2005, Avignon, France. pp.1-6. in2p3-00025135

HAL Id: in2p3-00025135

<http://hal.in2p3.fr/in2p3-00025135>

Submitted on 28 Nov 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LES GRILLES DE CALCUL AU SERVICE DE LA PHYSIQUE MEDICALE

Cheick O. THIAM (1), Denise DONNARIEIX (2), Vincent BRETON (1), Ziad EL BITAR (1), Lydia MAIGNE (1)

1 : Laboratoire de Physique Corpusculaire, 24 avenue des Landais, 63177 Aubière cedex

2 : Unité de physique médicale, département de radiothérapie curiethérapie du Centre Jean Perrin, 63000 Clermont-Ferrand.

Introduction :

Dans le cadre de ses activités de recherche, l'équipe Plate forme de Calcul pour les Sciences du Vivant (PCSV) du Laboratoire de Physique Corpusculaire de Clermont Ferrand est impliquée dans le déploiement des applications biomédicales sur des environnements de calcul informatique distribué. Etant donné la demande croissance des scientifiques en ressources informatiques (calcul et stockage), l'idée de grille de calcul est apparue, de manière à mutualiser des ressources réparties, afin de pouvoir accéder à de la puissance de calcul, à des données sécurisées géographiquement dispersées et de bénéficier d'une continuité de service pour les utilisateurs. Nous étudions donc les capacités de la grille informatique pour optimiser les temps de calcul des simulations de type Monte Carlo avec le code GATE qui fait office d'une application pilote pour l'infrastructure de grille.

GATE (Geant4 Application for Tomographic Emission) est une plateforme de simulation générique basée sur le code GEANT4 et élaborée pour répondre dans un premier temps aux besoins spécifiques de la communauté scientifique dans le domaine des applications SPECT/PET. Son champ d'application peut aussi s'étendre à des applications en radiothérapie et curiethérapie. Cependant, des résultats précis ne peuvent être obtenus avec une telle méthode que si nous générons beaucoup d'événements. Cela nécessite des heures de calcul d'où l'intérêt d'utiliser une infrastructure de grille informatique.

Dans un premier temps seront exposées les caractéristiques de l'infrastructure de la grille qui s'inscrit dans le cadre d'un projet européen EGEE. Ensuite nous nous intéresserons au déploiement de l'application GATE sur la grille EGEE et au portail web développé dans le cadre de cette activité. Ce portail intègre les fonctionnalités de l'intergiciel («middleware») EGEE utilisé pour accéder aux ressources de la grille afin de pouvoir faciliter l'utilisation de l'application GATE sur la grille de calcul. Une station de travail est en cours d'installation au centre anticancéreux Jean Perrin de Clermont Ferrand.

Matériel et méthodes :

Le projet EGEE (Enabling Grids for E-science in Europe) fournit une plate forme de production pour la recherche répartie sur plusieurs sites et impliquant plusieurs communautés d'utilisateurs organisés en groupes de travail. L'accès aux ressources de la grille nécessite donc pour l'utilisateur, d'être affilié à une organisation virtuelle (VO). Dans le cas de l'application GATE, il faut être membre de la VO "biomed". Un certificat d'autorisation, numérique et associé à des mots de

passé, est attribué à tous les utilisateurs afin de garantir un accès sécurisé aux ressources de la grille. L'utilisateur interagit avec la grille via une interface utilisateur (UI) lui permettant :

- d'interroger le système d'information pour connaître les ressources disponibles ;
- de manipuler des données : les déposer sur des unités de stockage (SE), les enregistrer sur des catalogues, les dupliquer...
- de soumettre des travaux (jobs).

Les sites mettant à disposition des ressources possèdent un ou plusieurs des éléments suivants :

- UI (User Interface) : Interface utilisateur, machine depuis laquelle un utilisateur soumet un job.
- CE (Computing Element) : Élément de calcul comprenant un ou plusieurs nœuds de calcul (Worker Nodes WN) et un frontal qui assure la correspondance entre un utilisateur de la grille et un compte disponible sur un CE pour la VO de l'utilisateur (GateKeeper GK).
- SE (Storage Element) : Élément de stockage (serveur de disques)
- MON (Monitoring Node) : Nœud de contrôle, machine chargée de faire tourner les fonctionnalités du système d'information pour le site et le service d'optimisation des duplications de fichiers.
- RB (Resource Broker) : Courtier de ressources, machine qui recherche sur la grille l'élément de calcul le plus adéquat pour l'exécution du job en prenant en compte la proximité des données éventuellement requises.

Des services de maintenance des catalogues (localisations des fichiers originaux et leurs copies), d'information (R-GMA) (Etat de la grille, infrastructure et applications) et de gestion (événements relatifs à la vie d'un job) sont aussi disponibles.

Le but du système d'information de la grille est de permettre la gestion et la surveillance des ressources distribuées.

La plate forme de simulation Monte Carlo GATE est disponible sur plusieurs sites EGEE. Un ensemble de commandes développées dans le middleware EGEE permet de soumettre et de gérer des jobs sur la grille. L'utilisateur a la possibilité d'enregistrer et de gérer les données (Images médicales, logiciels...) et d'accéder aux unités de calculs à travers différents sites dans le monde. Il est possible d'évaluer les ressources de travail pour la VO à laquelle on appartient (temps CPU utilisation mémoire ...).

Nous étudions avec GATE, les distributions de doses simulées à partir des faisceaux de photons et d'électrons (radiothérapie) ou des sources radioactives (curiethérapie). Les simulations Monte Carlo utilisent des générateurs de nombres pseudo aléatoires de manière à reproduire la probabilité des interactions physiques dans la matière.

Plus les interactions dans la matière sont nombreuses, plus les simulations utilisent un nombre grandissant de nombres aléatoires. Le meilleur moyen pour réduire les temps de calcul incombant aux simulations Monte Carlo est de les subdiviser en un nombre de simulations générant moins de particules. A chacune de ces petites simulations est indexée une sous séquence de nombres pseudo aléatoires obtenue en « découpant » une très longue séquence de nombres, chaque sous séquence de nombre devant être indépendante.

Nous avons utilisé une méthode de segmentation pour découper la séquence de nombres aléatoires. A chaque sous séquence obtenue, correspond un « état » du générateur aléatoire stocké sous forme d'un fichier. Un job GATE générant donc des millions d'événements est parallélisé en plusieurs petites simulations (jobs) qui sont déployées sur la grille de calcul, chacune utilisant un « état » du générateur aléatoire.

La soumission d'un job se fait depuis une interface utilisateur (UI), en envoyant au courtier de ressources (RB) un script JDL (Job Description Language) décrivant les caractéristiques du job et ses pré requis sous forme d'attributs. Ces attributs décrivent les ressources nécessaires (CPU, mémoire, logiciel GATE...) et les fichiers nécessaires (GateMaterials.db : fichier contenant la description des matériaux utilisés dans la définition de la géométrie dans la simulation...).

Un ensemble de commandes développées dans le middleware EGEE permet à l'utilisateur de surveiller ses travaux sur la grille (soumission du job, vérification de l'état du job enregistrement et réplication de données récupération de résultats...). Une fois le job soumis, il passe par différents statuts (Submitted, Waiting, Ready, Scheduled, Running, Done).

Nous avons effectués des tests sur deux sites EGEE (CC-IN2P3 Lyon et LPC Clermont) sur lesquels l'application GATE est installée afin d'estimer les temps de calcul. Nous définissons les paramètres d'une simulation de traitement de radiothérapie générant plusieurs millions d'événements (20.000.000 particules). La simulation est divisée successivement en des séries de dix, vingt, cinquante et cent tâches qui sont ensuite déployées sur la grille de calcul. Nous avons évalués les temps relatifs à la vie de chaque tâche pour chaque série de test. Les résultats présentés ci-dessous représentent une moyenne des temps pour chaque cas.

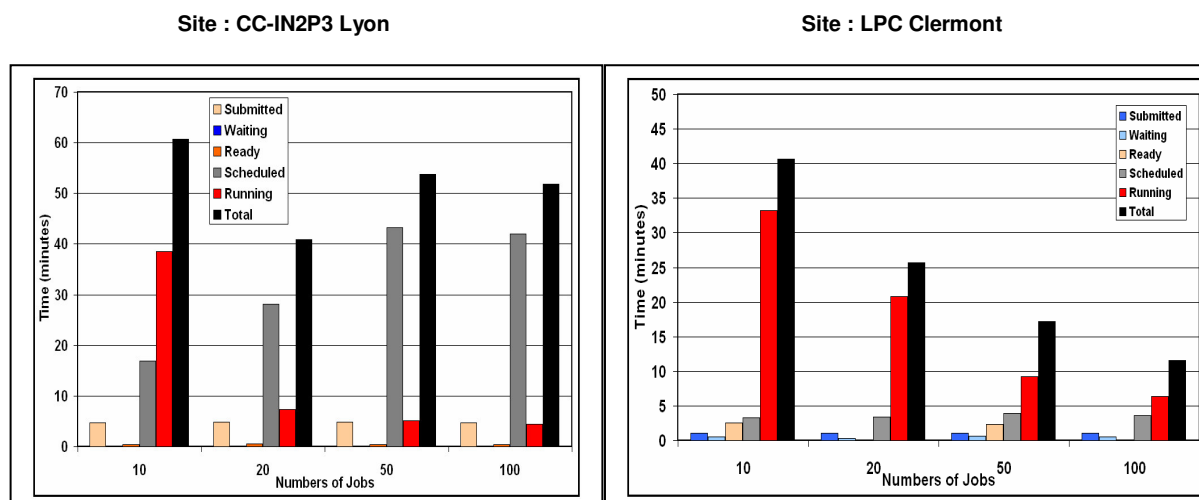


Figure 1 : Comparaison des temps pour les différents statuts des jobs

Les chiffres représentés sur la figure 1 décrivent un état du job au cours de sa vie sur la grille.

Une fois que la requête de soumission est adressée au courtier de ressources (RB), la tâche est dans l'état "submitted"

L'état "waiting" : la tâche est en attente pendant que le RB interroge le système d'information pour trouver l'élément de calcul le plus approprié pour accueillir la tâche.

L'état "Ready" : la tâche est préparée pour être soumise au "Computing Element" (CE).

L'état "Scheduled" : la tâche est en queue de batch sur le CE.

L'état "Running" : la tâche tourne.

L'état "Done" : l'exécution sur le CE est terminée et s'est bien déroulée; le résultat peut être rapatrié à l'utilisateur.

Le temps correspondant à l'état "Scheduled" peut être important (figure1), cela dépend directement de la charge du CE et de la performance des machines qui constituent les nœuds. Les autres temps sont négligeables par rapport au temps total que passe le job sur la grille.

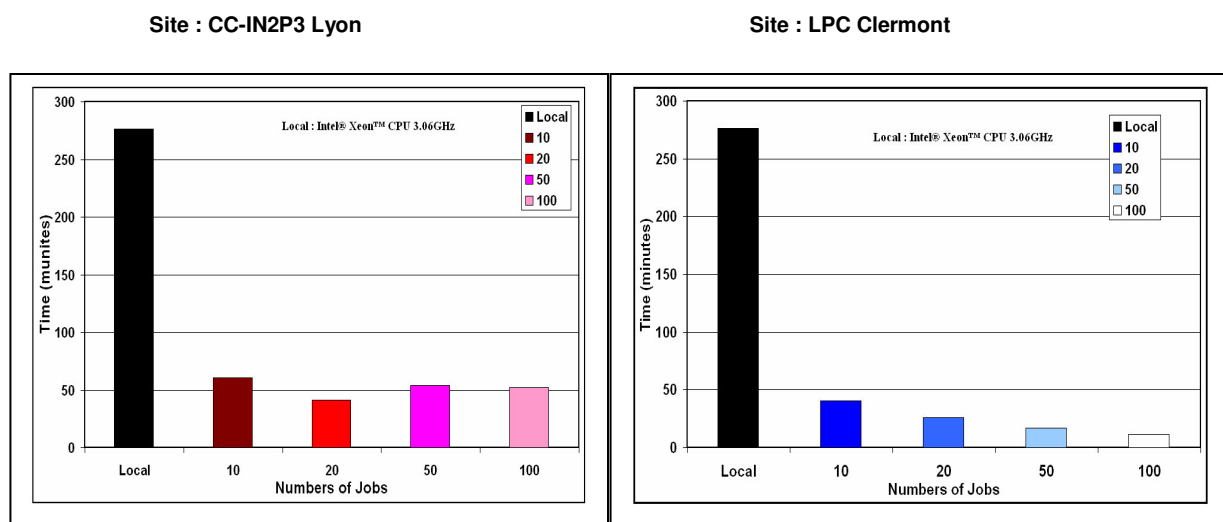


Figure 2 : comparaison temps total LPC / CC-IN2P3 par rapport au temps en local

La figure 2 représente une comparaison du temps d'exécution sur une machine locale de la simulation subdivisée et des jobs déployées sur la grille par rapport aux temps totaux pour chaque série de test.

Dans le cadre de cette application, nous obtenons un résultat très significatif en parallélisant une simulation GATE sur la grille informatique.

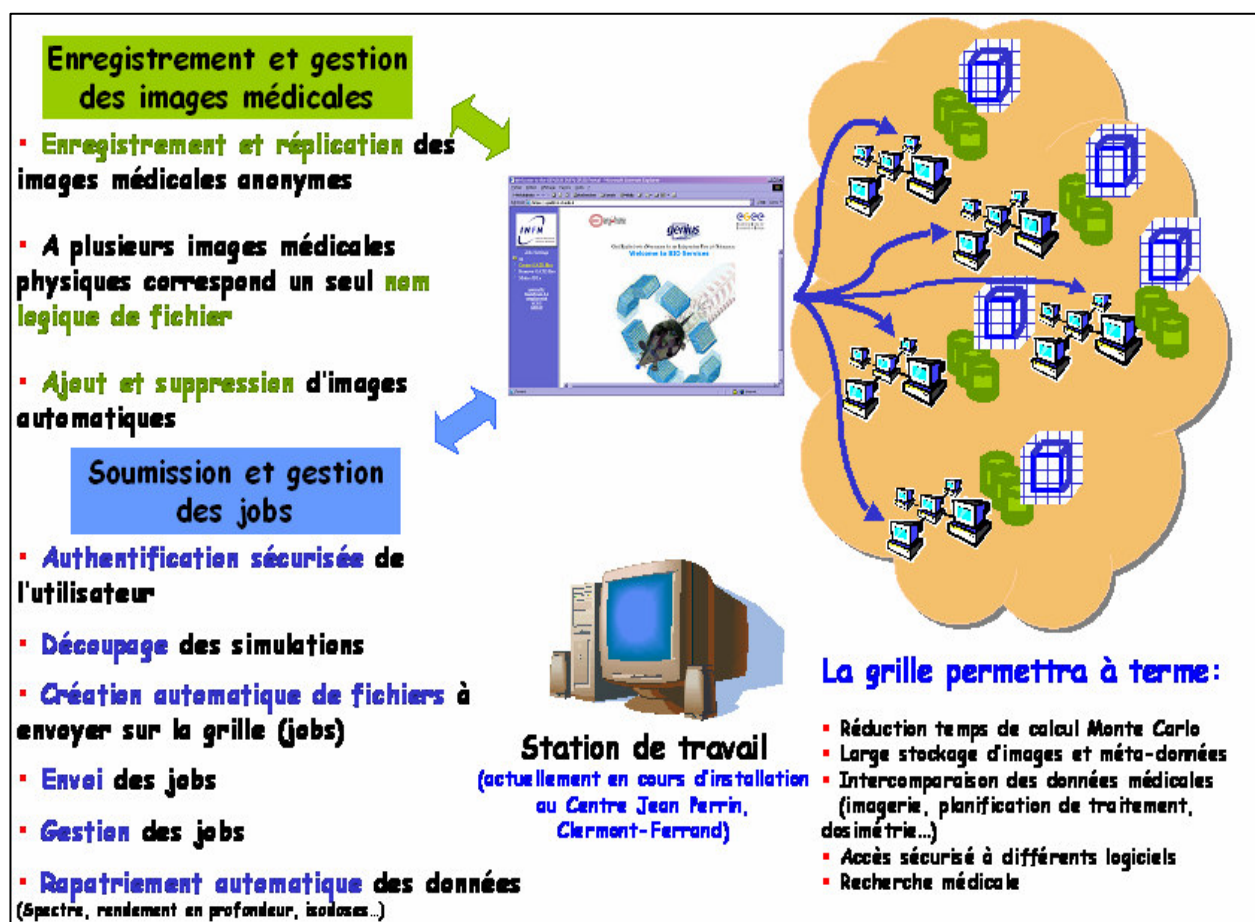
Nous avons un gain maximum de 24 pour une division en 100 jobs sur le site EGEE LPC de Clermont, par contre sur le site CC-IN2p3 Lyon le gain maximum est de 7 pour une répartition en 20 Jobs.

Ces chiffres montrent que la division de la simulation en un très grand nombre de petites simulations n'est pas forcément la meilleure solution pour paralléliser un job GATE. Il est nécessaire de tenir compte de plusieurs facteurs notamment le nombre de ressources disponibles pour l'application, la charge des machines et leurs performances.

L'application GATE étant un outil destiné à la physique médicale, un portail internet a été développé afin de pouvoir faciliter l'utilisation des ressources de la grille informatique. Le portail GENIUS intègre toutes les fonctionnalités de la grille de calcul de telle sorte que les lignes de commandes EGEE sont donc élaborées de façon automatique.

Ce portail internet permet donc à un physicien médical de pouvoir utiliser l'application GATE dans les planifications de traitement en curiethérapie et radiothérapie, de soumettre des simulations Monte Carlo sur les ressources de la grille avec les temps de calcul suffisamment réduits. Ce portail est sécurisé et nécessite un certificat d'autorisation.

Les Fonctionnalités du portail GENIUS :



Pour un utilisateur, il est donc important que l'accès aux données médicales stockées dans des ressources de stockage de la grille via le portail soit rapide (par rapport à la durée de calcul d'une simulation). Ces données (images médicales binaires) peuvent être utilisées dans les simulations GATE pour des calculs dosimétriques par exemple. Cependant, elles doivent être anonymes avant l'enregistrement sur les ressources de la grille. A ces images médicales doivent correspondre des bases de données relatives à un patient.

Afin d'être sûr que les données médicales soient accessibles pour les calculs, leur répliquation sur différents éléments de stockage (SE) devrait être possible. Les services de grille donnent la possibilité de gérer ces informations de façon libre et

transparentes. Les opérations de manipulation des données et des catalogues sur la grille sont assurées par le système « Replica Manager » qui intègre tous les outils permettant de gérer les données sur la grille, et assure un point d'entrée unique pour les autres utilisateurs et services de la grille.

Les fichiers déposés sont identifiés par un nom logique de fichier LFN (Logical File Name), alias d'un GUID (Grid Unique IDentifier), défini par l'utilisateur.

Conclusion :

Les simulations Monte Carlo effectuées avec le logiciel GATE donne des résultats très concluants. Sa facilité d'utilisation et son support basé sur le code GEANT4 en font un outil convivial et performant.

Les travaux effectués de manière à paralléliser le code de simulation réduisent considérablement le temps de calcul pour une simulation donnée. Les chiffres obtenus dans les tests de temps sont très bons sachant que les clusters utilisés pour ces tests sont très chargés. Le seul problème quand on découpe trop les jobs est que les temps de soumission des jobs et de rapatriement en local des résultats deviennent important d'où la nécessité de paralléliser la soumission également.

Le portail web offre une possibilité d'utiliser les ressources de la grille de façon simple et transparente notamment dans un milieu médical.

Références :

[1]: **GATE: a simulation toolkit for PET and SPECT**

Jan S., Santin G., Strul D., Staelens S. and al, *Phys. Med. Biol.*, 49 (2004) 4543-4561.

[2]: **Parallelization of Monte Carlo Simulations and Submission to a Grid Environment**

L.Maigne, D. Hill, P. Calvat, V. Breton, R. Reuillon, D.Lazaro, Y. Legré, D. Donnarieix,, *Parallel Processing Letters journal*, Vol. 14, No. 2 (June 2004), p 177-196.

[3] **GENIUS: a simple and easy way to access computational and data grids**

A. Adronico, R. Barbera, A. Falzone, P. Kunszt, G. Lo Re, A. Pulvirenti, A. Rodolico, *Future Generation Computer Systems*, volume 19, Issue 6, pages 805-813, 2003.

[4]: **L'Equipe PCSV Clermont Ferrand**

<http://clrpcsv.in2p3.fr>

[5]: **OpenGATE collaboration**

<http://www-iphe.unil.ch/~PET/research/gate/>

[6]: **Le projet EGEE**

<http://public.eu-egee.org/>

[7]: **Le portail web GENIUS**

<https://genius.ct.infn.it>