HAL open science

Calcul distribué pour l'imagerie médicale

Sorina Pop, Rafael Ferreira da Silva, Tristan Glatard

To cite this version:

Sorina Pop, Rafael Ferreira da Silva, Tristan Glatard. Calcul distribué pour l'imagerie médicale. JRES (Journées réseaux de l'enseignement et de la recherche) 2011, Renater, Nov 2011, Toulouse, France. hal-04804557

HAL Id: hal-04804557

https://hal.science/hal-04804557v1

Submitted on 26 Nov 2024

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.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Calcul distribué pour l'imagerie médicale

Sorina Pop
Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon ; Université Lyon 1
CREATIS - INSA LYON, Bâtiment Blaise Pascal, 7 Avenue Jean Capelle, 69621 Villeurbanne Cedex

Rafael Silva
Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon ; Université Lyon 1
CREATIS - INSA LYON, Bâtiment Blaise Pascal, 7 Avenue Jean Capelle, 69621 Villeurbanne Cedex

Tristan Glatard
Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon ; Université Lyon 1
CREATIS - INSA LYON, Bâtiment Blaise Pascal, 7 Avenue Jean Capelle, 69621 Villeurbanne Cedex

Résumé

Le traitement et la simulation d'images médicales peuvent être très demandeurs en puissance de calcul. Ceci est du à la fois à la complexité des algorithmes et aux volumes de données à traiter qui sont de plus en plus importants. Pour obtenir des résultats dans des délais raisonnables, ces applications peuvent être parallélisées et exécutées sur des architectures distribuées.

Le portage des applications sur des infrastructures distribuées est réalisé de manière générique en utilisant des workflows explicitant l'enchaînement des processus de calcul, ainsi que les flots de données. Chaque application est décrite par un workflow, qui est ensuite initialisé avec les entrées données par l'utilisateur pour chaque exécution de l'application. Les applications sont pré-compilées sur des machines compatibles avec les infrastructures utilisées et installées à la volée sur les nœuds de calcul. Pour la soumission et la gestion des tâches générées à partir des workflows, un système de tâches dites «pilotes» est utilisé. Ce système permet d'améliorer les performances en réduisant le temps d'attente des tâches et en améliorant la qualité de service sur des infrastructures distribuées et hétérogènes.

L'exécution et le suivi des ces applications sur des grilles de calcul sont facilités par une plateforme mise en place au laboratoire. La plateforme comporte un serveur pour la gestion des workflows et une interface web. A travers ce portail web, les utilisateurs ont la possibilité de s'authentifier, lancer l'exécution d'une application, la suivre ou bien accéder à l'historique et aux résultats des exécutions précédentes. Le transfert de données vers et depuis les éléments de stockage se fait de manière transparente.

Mots clefs

Calcul distribué, imagerie médicale, grille, workflow

1 Introduction

Le laboratoire Creatis (http://www.creatis.insa-lyon.fr) est impliqué depuis plusieurs années dans des activités de calcul distribué pour le traitement des images médicales, la modélisation et la simulation numérique dédiées à l'imagerie du vivant. Le calcul sur des infrastructures distribuées telles que les grappes et grilles de calcul est aujourd'hui devenu un outil quotidien pour la recherche en imagerie médicale à Creatis.

Plusieurs classes d'applications ayant de forts besoins de calcul peuvent être parallélisées afin de bénéficier des ressources des infrastructures distribuées. Une première possibilité de parallélisation concerne les applications de balayage de paramètres, c'est-à-dire l'exécution de plusieurs instances de la même application en faisant varier les paramètres d'entrée. Ceci est pertinent pour les chercheurs travaillant sur l'optimisation des paramètres utilisés par exemple dans l'initialisation des algorithmes de segmentation d'organes [1] ou dans le filtrage par décalage de moyenne (Mean-Shift) [2, Fig 5]. Une méthode comparable découpe les données d'entrée (souvent des images de grande taille) en plusieurs morceaux sur lesquels l'application est itérée parallèlement, avec une

étape de fusion à la fin des calculs. Cette méthode est utilisée à Creatis par exemple pour le simulateur FIELD-II [4, 5] (http://server.electro.dtu.dk/personal/jaj/field/) et pour le calcul des cartographies paramétriques de perfusion hépatique [3, Fig 6]. Dans le cas des simulateurs Monte-Carlo tels que GATE [6, Fig 7] et Sorteo, la parallélisation est faite en divisant le nombre d'événements en plusieurs taches indépendantes. Chaque tâche reçoit ainsi l'intégralité des données mais ne contribue que partiellement à la simulation [7].

2 Présentation générale

2.1 Description de la plateforme

La plate-forme web développée et utilisée à Creatis permet d'accéder à des applications d'imagerie médicale pour la simulation en imagerie et radio-/hadronthérapie ainsi que pour des études paramétriques. Cette plate-forme est supportée par les projets ANR «Virtual Imaging Platform» (VIP) et «hGATE».

L'intégration (le portage) des applications est réalisée de manière générique en utilisant des workflows explicitant l'enchainement des processus de calcul, ainsi que les flots de données. Ces workflows sont gérés par le moteur de workflow MOTEUR [8] et sont écrits dans le langage GWENDIA [9]. Chaque application est décrite par un workflow, qui est ensuite initialisé avec les entrées données par l'utilisateur pour chaque exécution de l'application. Les applications sont pré-compilées sur des machines compatibles avec les infrastructures utilisées et installées à la volée sur les nœuds de calcul.

La plateforme comporte un serveur pour la gestion des workflows et un portail web comme interface pour les utilisateurs. Elle offre principalement les fonctionnalités suivantes :

- Définition de classes d'applications, groupes d'utilisateurs et droits associés.
- Transfert et réplication de fichiers et répertoires vers et depuis le catalogue de fichiers LFC ; gestion d'espaces privés et partagés.
- Lancement, exécution et suivi de workflows sur les ressources de l'organisation virtuelle (VO) «biomed»[1] de la grille de calcul EGI et sur des grappes locales.
- Visualisation d'objets 3D et statistiques d'utilisation

Cette plate-forme est disponible à l'adresse http://vip.creatis.insa-lyon.fr [10].

Pour la soumission et la gestion des tâches générées à partir des workflows, nous utilisons le système de tâches « pilotes » DiRAC [11]. Ce système permet d'améliorer les performances en réduisant le temps d'attente des tâches et en améliorant la qualité de service sur des infrastructures distribuées et hétérogènes.

La plateforme compte aujourd'hui une dizaine d'applications et une vingtaine d'utilisateurs.

2.2 Amélioration des performances

Les ressources à disposition de la VO biomed, qui comporte plus d'une centaine de sites de calculs, sont fortement hétérogènes. Cette hétérogénéité matérielle et logicielle, ainsi que l'absence d'une qualité de service (QoS) assurée, complexifient significativement l'ordonnancement des tâches. Typiquement, le makespan (le temps d'exécution d'une application perçu par l'utilisateur) est déterminé par la date de fin de la dernière tâche, qui peut pénaliser considérablement les performances si elle est exécutée sur une machine plus lente que les autres ou si elle doit être re-soumise suite à un échec.

La Figure 1 illustre le fait que les dernières tâches d'une simulation sont responsables d'une fraction significative du makespan.

Pour remédier à ce problème, nous développons des méthodes d'équilibrage de charge propres aux systèmes distribués à très grande échelle. La Figure 2 présente un résultat obtenu grâce à l'algorithme présenté en [7].

[1] à laquelle contribuent une quinzaine de sites français

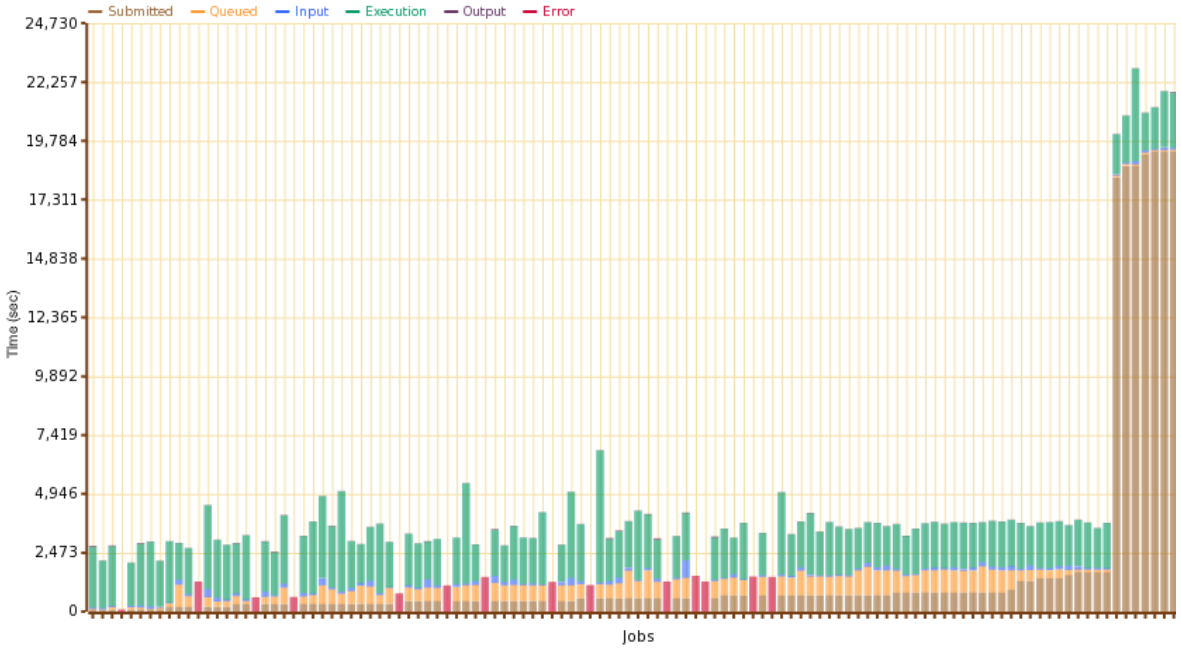

Figure 1: Parallélisation statique d'une simulation GATE. Les tâches les plus longues (échouées et re-soumises) pénalisent l'ensemble de la simulation.

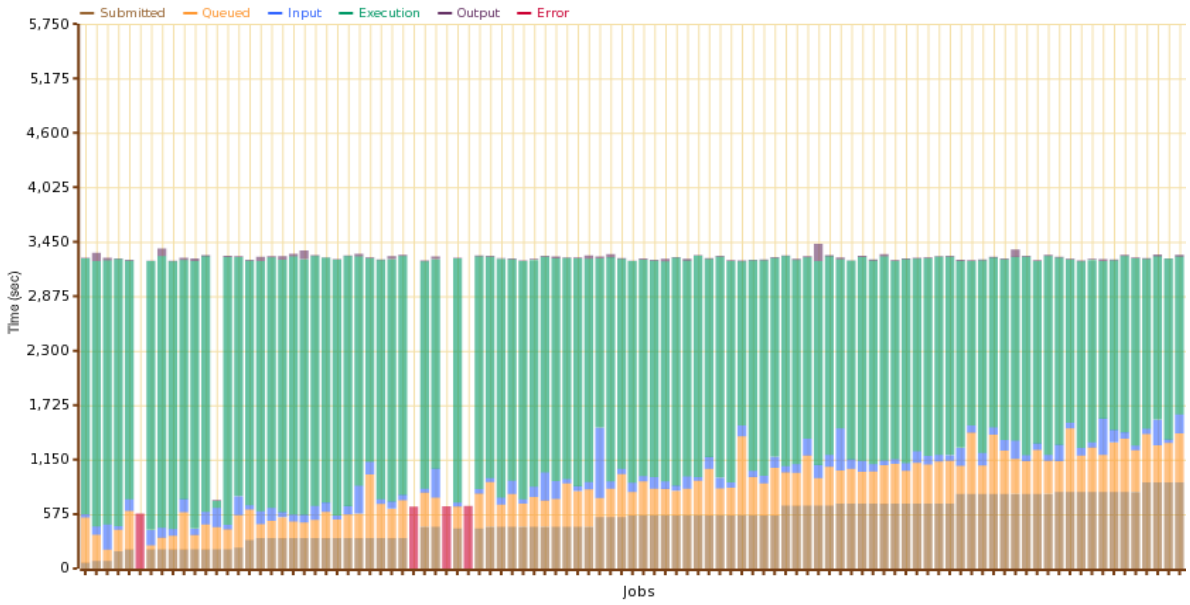

Figure 2: Parallélisation dynamique d'une simulation Gate. Les temps de terminaison des tâches sont sensiblement identiques et les erreurs n'ont pas besoin de re-soumission.

3 Résultats

Depuis janvier 2011, la plateforme a été utilisée par une vingtaine d'utilisateurs pour une dizaine d'applications portées sur la grille et intégrées dans le portail. Les applications les plus utilisées ont été Gate, le calcul de cartographies, les simulateurs FIELD II et Simri, ainsi que le filtrage MeanShift. Depuis le 1er janvier 2011, ces applications ont produit des résultats[2] correspondant à plus de 10 ans de calcul réparti sur les ressources de la VO biomed. Quelques statistiques d'utilisation sont présentées dans le Tableau 1, ainsi que les figures 3 et 4.

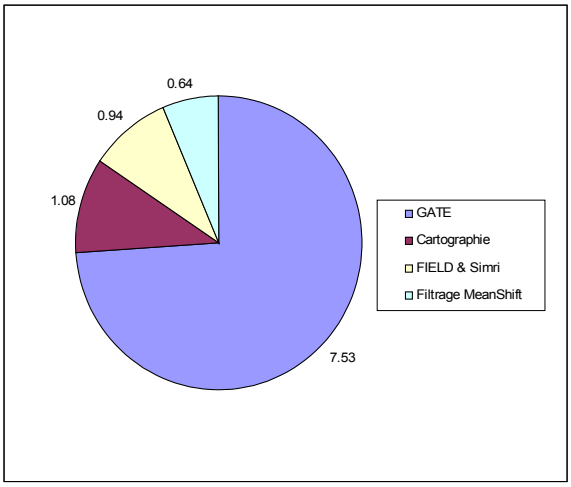

Figure 3: Années CPU consommées par les applications les plus utilisées dans la plateforme depuis le 01/01/2011

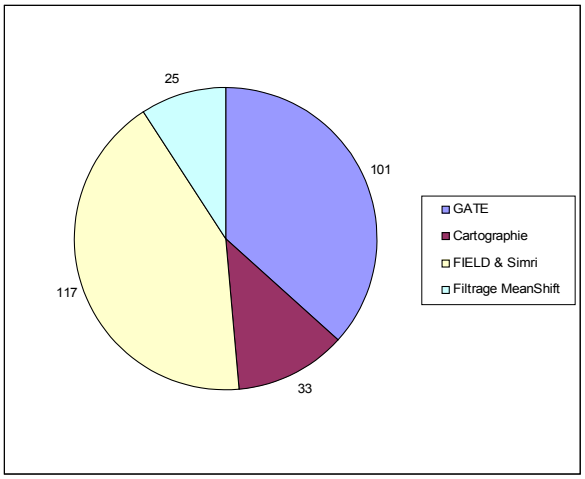

Figure 4: Nombre d'exécutions terminées avec succès pour les applications les plus utilisées dans la plateforme depuis le 01/01/2011

Applications	Années CPU	Nb de workflows
GATE	7.53	101
Cartographie	1.08	33
FIELD & Simri	0.94	117
Filtrage MeanShift	0.64	25
TOTAL	**10.19**	**276**

Tableau 1: Activité des principales applications portées dans la plateforme depuis janvier 2011

Les figures ci-dessous illustreront quelques images obtenues avec ces 5 applications.

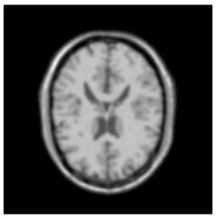

Figure 5: simulation du cerveau avec le simulateur IRM Simri

[2]Résultats produits par et pour les utilisateurs finaux. Les expériences réalisées par les administrateurs de la plateforme ne sont pas comptés dans ces valeurs

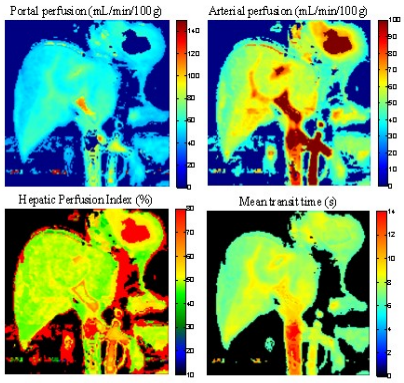

Figure 8: étude paramétrique de la perfusion hépatique [3].

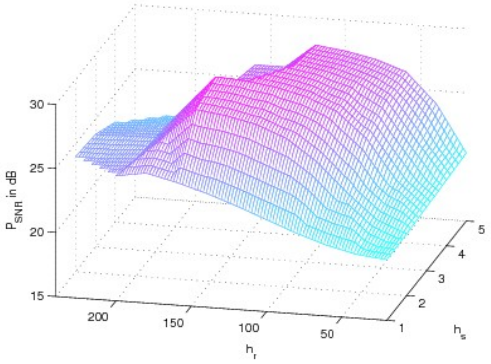

Figure 6: optimisation des paramètres d'échelle du filtrage par décalage de moyenne (Mean-Shift) [2].

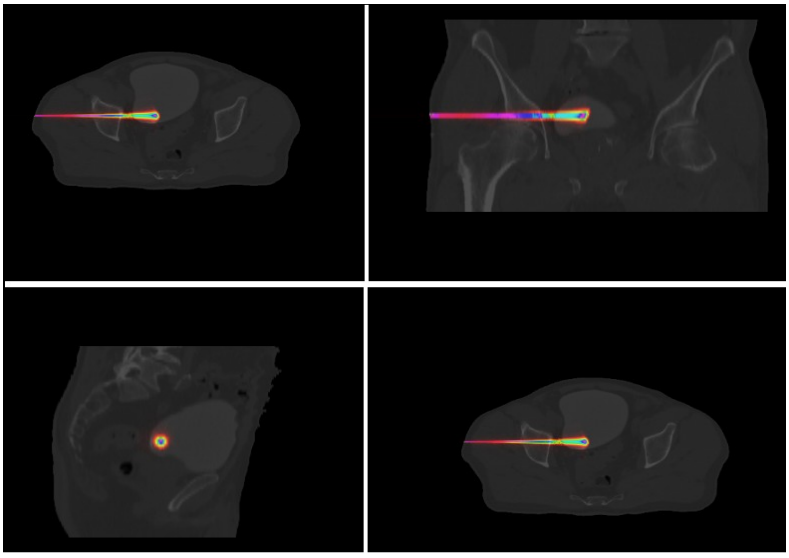

Figure 7: simulation avec le logiciel GATE.

Figure 9: Simulation échocardiographique obtenue avec le logiciel FIELD-II.

4 Bibliographie

[1] Ketan Maheshwari, Tristan Glatard, Joel Schaerer, Bertrand Delhay, Sorina Camarasu-Pop, Patrick Clarysse, Johan Montagnat. “Towards Production-level Cardiac Image Analysis with Grids”, HealthGrid'09, Berlin, 2009.

[2] T. Li, S. Camarasu-Pop, T. Glatard, T. Grenier, and H. Benoit-Cattin, "Optimization of Mean-Shift scale parameters on the EGEE grid", Studies in health technology and informatics, Proceedings of Healthgrid 2010, vol. 159, pp. 203-214, 2010.

[3] B. Leporq, S. Camarasu, F. Pilleul, and O. Beuf, "3D-liver quantitative perfusion mapping using EGEE grid with MR-DCE imaging and MS-325 blood pool contrast agent", ISMRM, Montreal, Canada, pp. 2046, May, 2011.

[4] J.A. Jensen, « Field: A Program for Simulating Ultrasound Systems », Paper presented at the 10th Nordic-Baltic Conference on Biomedical Imaging Published in Medical & Biological Engineering & Computing, pp. 351-353, Volume 34, Supplement 1, Part 1, 1996.

[5] J.A. Jensen and N. B. Svendsen, « Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers », IEEE Trans. Ultrason., Ferroelec., Freq. Contr., 39, pp. 262-267, 1992.

[6] S. Jan, G. Santin, D. Strul, S. Staelens, K. Assié, D. Autret, S. Avner, R. Barbier, M. Bardiès, P.M. Bloomfield, D. Brasse, V. Breton, P. Bruyndonckx, I. Buvat, A.F. Chatziioannou, Y. Choi, Y. H. Chung, C. Comtat, D. Donnarieix, L. Ferrer, S.J. Glick, C.J. Groiselle, D. Guez, P.F. Honore, S. Kerhoas-Cavata, A.S. Kirov, V. Kohli, M. Koole, M. Krieguer, D.J. van der Laan, F. Lamare, G. Largeron, C. Lartizien, D. Lazaro, M.C. Maas, L. Maigne, F. Mayet, F. Melot, C. Merheb, E. Pennacchio, J. Perez, U. Pietrzyk, F. R. Rannou, M. Rey, D.R. Schaart, C.R. Schmidtlein, L. Simon, T.Y. Song, J.M. Vieira, D. Visvikis, R. Van de Walle, E. Wieërs, C. Morel. “GATE: a simulation toolkit for PET and SPECT.”, Phys Med Biol, Vol. 49, 2004

[7] Sorina Camarasu-Pop, Tristan Glatard, Jakub Moscicki, Hugues Benoit-Cattin, David Sarrut. “Dynamic partitioning of GATE Monte-Carlo simulations on EGEE.”, Journal of Grid Computing (JGC), Vol. 8 241-259, 2010

[8] Tristan Glatard, Johan Montagnat, Diane Lingrand, Xavier Pennec. “Flexible and Efficient Workflow Deployment of Data-Intensive Applications On Grids With MOTEUR.” International Journal of High Performance Computing Applications (IJHPCA). vol.22 no.3 347-360, 2008

[9] Johan Montagnat, Benjamin Isnard, Tristan Glatard, Ketan Maheshwari, Mireille Blay-Fornarino. “A data-driven workflow language for grids based on array programming principles.” Proceedings of the Workshop on Workflows in Support of Large-Scale Science (WORKS'09), pages 1–10, ACM, Portland, USA, 2009

[10] R. Ferreira da Silva, S. Camarasu-Pop, B. Grenier, V. Hamar, D. Manset, J. Montagnat, J. Revillard, J. R. Balderrama, A. Tsaregorodtsev, and T. Glatard, "Multi-infrastructure workflow execution for medical simulation in the Virtual Imaging Platform", HealthGrid 2011, Bristol, UK, 2011.

[11] A Tsaregorodtsev, N Brook, A Casajus Ramo, Ph Charpentier, J Closier, G Cowan, R Graciani Diaz, E Lanciotti, Z Mathe, R Nandakumar, S Paterson, V Romanovsky, R Santinelli, M Sapunov, A C Smith, M Seco Miguelez, A Zhelezov. “DIRAC3 - the new generation of the LHCb grid software.” Journal of Physics: Conference Series, 2010.