

Raphaël Bolze

**Dossier de candidature
au concours
N° 43/05
CR2 CNRS**

Sommaire

Curriculum vitæ	2
Activités de recherche	3
Publications	3
Résumé de thèse	4
Relectures	5
Co-encadrement	5
Développement Logiciel	5
Exposés et communications	6
Projet de recherche	7
Intitulé	7
Contexte	7
Axes de recherche	7
Intégration à l'IGBMC	9
Collaborations et ouvertures	10
Activités d'enseignements	11
Documents	13
Rapport de soutenance	14
Rapports de thèse	16
Attestation de thèse	20
Attestation de monitorat	21
Lettres de recommandation	22

Raphaël Bolze

Laboratoire de l'Informatique du Parallélisme
École Normale Supérieure de Lyon
46 allée d'Italie 69364 Lyon Cedex 07, France
Nationalité : Française

+33 4 37 28 76 45
raphael.bolze@ens-lyon.fr
<http://graal.ens-lyon.fr/~rbolze>

CURSUS

- fev-2009 - : **Postdoc** Information Sciences Institute - University of Southern California (ISI/USC)
Equipe : Advanced System Division
Chef d'équipe : Ewa Deelman
Projets : Peegasus Workflow Management System, Ocean Modeling
- 2005 - 2008 : **Doctorat en Informatique**
Titre : *Analyse et déploiement de solutions algorithmiques et logicielles pour des applications bioinformatiques à grande échelle sur la grille*
Title : *Analysis and deployment of algorithm and software solutions for large scale bioinformatic applications into computing grid*
Directeur de thèse : Frédéric Desprez (Directeur de Recherche INRIA)
Financement : Bourse de doctorat pour l'ingénieur BDI-CNRS
Laboratoire de l'Informatique du Parallélisme (LIP)
École Normale Supérieure de Lyon
Date de soutenance : 31/10/2008.
- 2005 - 2008 : **Moniteur**
École Normale Supérieure de Lyon
- Fév. 2005 - Oct. 2005 : **Ingénieur Décrypton CNRS**
Laboratoire de l'Informatique du Parallélisme (LIP)
École Normale Supérieure de Lyon
- Fév. 2004 - Fév. 2005 : **Ingénieur Expert INRIA**
Laboratoire de l'Informatique du Parallélisme (LIP)
École Normale Supérieure de Lyon
- 2002 - 2003 : **DEA en Informatique**
l'institut de Recherche Informatique de Nantes (IRIN)
- été 2002 : **Certification Cisco CCNA**
Université de l'Assumption Bangkok - Thaïlande.
- 2000 - 2003 : **Diplôme d'ingénieur**
Département : Systèmes Informatiques : Logiciels et Réseaux.
Option : Réseaux
École Polytechnique de l'Université de Nantes
- 1997 - 2000 : **Classes préparatoires**
Option : Physique, Chimie
Lycée Camille Vernet - Valence

PUBLICATIONS

Mes travaux ont donné lieu à 3 publications dans des journaux internationaux, 3 dans des conférences internationales, et 1 dans une conférence nationale. Ces publications sont disponibles sous la forme de rapports de recherches de l'INRIA et/ou du LIP.

Journal international (avec comité de lecture)

- [1] A. Amar, R. Bolze, Y. Caniou, E. Caron, B. Depardon, J.-S. Gay, G. Le Mahec, and D. Loureiro. Tunable scheduling in a GridRPC framework. *Concurrency & Computation : Practice & Experience*, 2008. To appear.
- [2] R. Bolze, F. Cappello, E. Caron, M. Daydé, F. Desprez, E. Jeannot, Y. Jégou, S. Lanteri, J. Leduc, N. Melab, G. Mornet, R. Namyst, P. Primet, B. Quetier, O. Richard, E.-G. Talbi, and T. Irena. Grid'5000 : a large scale and highly reconfigurable experimental grid testbed. *International Journal of High Performance Computing Applications*, 20(4) :481–494, Nov. 2006.
- [3] N. Garnier, A. Friedrich, R. Bolze, E. Bettler, L. Moulinier, C. Geourjon, J. D. Thompson, G. Deleage, and O. Poch. MAGOS : multiple alignment and modelling server. *Bioinformatics*, 22(17) :2164–2165, 2006.

Conférences Internationales (avec comité de lecture)

- [4] A. Amar, R. Bolze, A. Bouteiller, P. K. Chouhan, A. Chis, Y. Caniou, E. Caron, H. Dail, B. Depardon, F. Desprez, J.-S. Gay, G. Le Mahec, and A. Su. Diet : New developments and recent results. In L. et al. (Eds.), editor, *CoreGRID Workshop on Grid Middleware (in conjunction with EuroPar2006)*, number 4375 in LNCS, pages 150–170, Dresden, Germany, August 28-29 2006. Springer.
- [5] V. Bertis, R. Bolze, F. Desprez, and K. Reed. Large scale execution of a bioinformatic application on a volunteer grid. In *Workshop on Parallel and Distributed Scientific and Engineering Computing (PDSEC), in conjunction with IPDPS*, April 2008.
- [6] R. Bolze, E. Caron, F. Desprez, G. Hoesch, and C. Pontvieux. A monitoring and visualization tool and its application for a network enabled server platform. In M. Gavrilova, editor, *Computational Science and Its Applications - ICCSA 2006*, volume 3984 of LNCS, pages 202–213, Glasgow, UK., May 8-11 2006. Springer.

Conférences nationales

- [7] P. d'Anfray, R. Bolze, and F. Desprez. Le programme decrypthon. In *Journée Réseaux, Strasbourg*, 2007.

Cette thèse présente un ensemble d'objectifs scientifiques visant à l'exploitation efficace des ressources des grilles informatiques. En particulier, nous avons étudié le déploiement d'applications sur des infrastructures hétérogènes et nous avons développé des techniques d'ordonnancement pour graphe de tâches. Nous nous sommes intéressés à plusieurs types de grilles aux caractéristiques différentes allant de la grille d'internaute World Community Grid à la grille expérimentale Grid'5000. La thèse a été réalisée dans le cadre du programme Décryphon (projet tripartite entre l'AFM, le CNRS et IBM) qui vise à accélérer la recherche dans les domaines qui intéressent l'AFM en donnant accès aux technologies de grilles.

Dans un premier temps, à partir de l'analyse des besoins des scientifiques et de leurs applications, nous avons conçu et maintenu la grille Décryphon. Les ressources de cette grille sont supportées par les centres de calcul des cinq universités partenaires (Bordeaux I, Lille I, ENS-Lyon, Pierre et Marie Curie Paris VI et Orsay), ainsi que le réseau RENATER (Réseau National de Télécommunications pour l'Enseignement et la Recherche), sur lequel est connecté l'ensemble des machines. Le Centre de Ressources Informatiques de Haute Normandie (CRIHAN) participe également au programme, il héberge les données volumineuses des projets scientifiques. La grille universitaire Décryphon s'articule autour de plusieurs éléments séparés : l'intergiciel de grille DIET, un portail web DIET_Webboard et des gestionnaires locaux de ressources (OAR, Loadleveler) propres à chaque centre de calcul. Le portail Web s'appuie sur l'intergiciel DIET pour soumettre les calculs sur les serveurs adaptés aux besoins de chaque application. Chaque site définit une politique indépendante d'utilisation des ressources. Les machines dédiées Décryphon s'intègrent dans le parc des centres universitaires, de cette manière l'ensemble des machines d'un centre de calcul sont partagées par les utilisateurs locaux et les scientifiques des projets Décryphon.

Nous nous sommes intéressés, en outre, au projet "Help Cure Muscular Dystrophy", un des projets sélectionnés par le programme Décryphon. Nous avons montré comment deux grilles, l'une dédiée et l'autre volatile, peuvent être utilisées de manière complémentaire pour préparer des calculs qui auraient demandé plus de 8000 ans sur une seule machine. À partir de la notion de *processeurs virtuels à plein temps* et de processeurs dédiés de référence nous avons donné un facteur de conversion entre les deux grilles. Ce facteur nous a permis de proposer une estimation des besoins de calcul pour la deuxième phase du projet.

Pour terminer, nous avons développé une nouvelle fonctionnalité à l'intergiciel DIET, le rendant capable de gérer l'exécution de tâches ayant des dépendances au travers d'un composant nommé MA_DAG. Nous nous sommes intéressés à développer des algorithmes prenant en compte plusieurs applications qui demandent l'accès aux mêmes ressources en même temps. Ces heuristiques permettent la prise en compte d'une priorité inter-applications et intra-applications, et peuvent rendre l'exécution concurrente de celles-ci plus équitable. Nous avons validé ces heuristiques avec des applications issues des projets du programme Décryphon.

Ces travaux ont nécessité un développement logiciel important, d'une part sur les applications du Décryphon elles-mêmes et sur leur portage afin de rendre transparente leur utilisation sur la grille Décryphon, mais aussi au niveau de l'intergiciel DIET et de son écosystème : DIET_Webboard, VizDIET, GoDIET, LogService, MA_DAG, *etc.* Les résultats présentés ont été obtenus et validés sur trois grilles différentes : la grille universitaire du Décryphon, la grille d'internautes (World Community Grid) et la grille expérimentale Grid'5000.

Jury de thèse :

Monsieur Vincent	BRETON	DR CNRS - Rapporteur
Monsieur Frédéric	DESPREZ	DR INRIA - Directeur de thèse
Monsieur Jacques	DELPLANCQ	Délégué du PDG IBM France
Monsieur Pierre	MANNEBACK	Professeur - Rapporteur
Monsieur Thierry	PRIOL	DR INRIA - Président du jury
Monsieur Thierry	TOURSEL	AFM

Par ailleurs, ma thèse est disponible via le serveur des thèses en ligne TEL

– <http://tel.archives-ouvertes.fr/tel-00344249/fr/>

RELECTURES

J'ai effectué des relectures scientifiques pour différentes conférences internationales :

- Super Computing (SC 2006) ;
- Grid-Enabling Legacy Applications and Supporting End Users Workshop (GELA 2006) ;
- International Meeting High Performance Computing for Computational Science (VECPAR 2006).
- Algorithms and Architectures for Parallel Processing (ICA3PP 2007)
- Advancing Science and Society through Computation (ICCS 2007)
- International Symposium on High-Performance Distributed Computing (HPDC 2007)
- Leading Supercomputing Conference (ISC 2007)
- Cluster Computing and the Grid (CCGrid 2009)

CO-ENCADREMENT

Au cours de mes activités de recherche, j'ai eu l'occasion d'effectuer plusieurs co-encadrements de stagiaires et d'ingénieurs recrutés pour les projets de recherche auxquels j'ai participé.

- Encadrement d'un ingénieur Décryphon, Nicolas Bard. Développement et maintenance de la grille Décryphon, co-encadrements Eddy Caron, Frédéric Desprez ;
- Encadrement d'un ingénieur pour le projet ANR « MDCA » GWENDIA, Benjamin Isnard. Gestion et développement d'heuristiques d'ordonnancement de *workflows*, co-encadrements Eddy Caron, Frédéric Desprez ;
- Stage de 2 mois niveau BTS, Marc Boury. Découverte et développement de VizDIET, co-encadrement Eddy Caron ;
- Encadrement de trois élèves niveau L3 de l'ENS-Lyon, Pierre Rannou , Loic Magnin, Sylvain Collange. Projet de programmation orienté objet sur VizDIET et GoDIET.

DÉVELOPPEMENT LOGICIEL

Mes activités de recherche m'ont amené à développer en équipe et à maintenir plusieurs logiciels distribués publiquement :

- Intergiciel DIET :
 - intégration du LogService ;
 - maintenance et stabilisation du code ;
 - gestion de l'ordonnancement multiple de workflows ;<http://graal.ens-lyon.fr/DIET/index.html> ;
- VizDIET : logiciel de visualisation de l'activité d'une plate-forme DIET, <http://graal.ens-lyon.fr/DIET/vizdiet.html> ;
- GoDIET : logiciel de déploiement d'une plate-forme DIET, <http://graal.ens-lyon.fr/DIET/godiet.html> ;
- LogService : logiciel de monitoring d'une plate-forme distribuée, <http://graal.ens-lyon.fr/DIET/logservice.html> ;
- DIET_Webboard : portail web de gestion d'une plate-forme grille gérée par l'intergiciel DIET, <http://graal.ens-lyon.fr/DIET/dietwebboard.html> ;

Voici une liste de quelques exposés que j'ai eu l'occasion de donner au cours de ma thèse. Il convient d'ajouter les exposés liés aux publications dans les conférences internationales (cf. [liste publications](#)) et les séminaires internes au sein des groupes de travail de mon équipe GRAAL. La plupart des communications ont été faites en langue anglaise.

- Open Grid Forum GGF16 Athènes 2006 - Exposé : « Proposal for data management in DIET » ;
- École Grid'5000 : retour d'expériences 2006 : « Épopée d'expériences DIET sur Grid'5000 » ;
- Workshop Workflow Optimisation in Distributed Environments Octobre 2006 - « Workflow management within DIET applied to multi-application scheduling ».
- Réunion ALPAGE 2007 : « Ordonnancement de Workflows » ;
- Exposés séminaire d'équipe RAINBOW (I3S Nice) 2007 - « Multi-Workflow scheduling », « Exécution à grande échelle d'une application bioinformatique sur une grille d'internautes ».
- Workshop invité : Global and Peer-to-Peer Computing (GP2PC08) - « Large Scale Execution of a Bioinformatic Application on a Volunteer Grid »
- Workshop invité : XtremWeb Users Group Workshop (XW08) - « From Dedicated Grid to Volunteer Grid : Large Scale Execution of a Bioinformatic Application » .
- Réunion "BIOinformatique à GRAnde échelle - BIOGRALE 2008" - « La grille Décryphon ».

PROJET DE RECHERCHE

Intitulé :

Changement de dimension : outillage, caractérisation et optimisation des grilles informatiques pour une science du vivant à grande échelle

CONTEXTE

Le traitement parallèle des calculs et des données devient indispensable pour de nombreux domaines scientifiques. Hier, réservé aux applications gourmandes en temps de calcul issues notamment de la simulation en mécanique ou en météorologie, il a rejoint maintenant d'autres applications aux besoins différents, comme par exemple l'imagerie médicale, la modélisation « in silico » d'interactions médicamenteuses ou même la recherche de documents sur Internet. Ces applications nécessitent de nouvelles approches axées sur la manipulation de grandes quantités de données et la représentation de processus d'analyses complexes.

Du côté des architectures matérielles, nous sommes passés des gros calculateurs propriétaires comme les machines Cray, NEC, IBM, *etc*, à des grappes montées à partir d'ordinateurs personnels (PC) interconnectés par des réseaux rapides. Nous avons vu ensuite ces grappes s'agréger en grappes de grappes puis, depuis le début des années 2000, une mise en commun globale des moyens de calcul et de stockage, que ce soit avec des machines énormes de centres de calcul ou des PCs de bureau. Après des débuts chez les universitaires, cet agrégat de ressources appelé « grille informatique » trouve maintenant une place grandissante aussi bien dans les entreprises soucieuses de rationaliser et de rentabiliser leurs investissements informatiques que dans les laboratoires de recherche. Par ailleurs, la démocratisation des ordinateurs personnels et des connexions Internet chez les particuliers a fait pénétrer les technologies de grilles jusque dans les foyers. Ainsi, plusieurs dizaines de projets de calcul sont supportés par des internautes volontaires désireux de contribuer à l'avancer des connaissances.

Les capacités de calcul de ces plates-formes diffèrent énormément, à la fois en terme de nombre d'opérations par seconde (nous dépassons le Teraflops) mais aussi en fonction de leur disponibilité ou de leur capacité à échanger des données. Les scientifiques d'aujourd'hui, qu'ils soient issus de laboratoires de recherche ou de grandes entreprises privées, ont accès à des moyens informatiques hétéroclites : ressources dédiées et réservées (type Cloud computing), ressources partagées et hébergées dans des centres de calcul (EGEE, Décryphon, *etc*), et ressources volatiles (grille d'internautes). Ils leur est donc indispensable d'avoir des outils d'exploitation efficaces de ces infrastructures de grilles qui prennent en compte leurs usages. Le problème commun à toutes ces architectures n'est donc pas le matériel mais plutôt le logiciel permettant d'exploiter cette puissance (du système à l'algorithmique).

Un domaine d'application important des grilles concerne les sciences du vivant et la bioinformatique : le domaine des sciences où la biologie, l'informatique et les mathématiques se rejoignent pour former une seule et unique discipline. Au centre de nombreuses recherches dans le monde, que ce soit dans la recherche académique que dans l'industrie de la santé, ce domaine, à l'impact sociétal important, continue à se développer de manière très importante. Les caractéristiques de ces applications diffèrent de celles du calcul numérique « classique » qui s'exécute actuellement en masse sur les grilles de production du monde entier. Elles posent donc de nouveaux problèmes de recherche informatique importants auxquels je m'intéresse en mettant au centre de mes préoccupations les besoins des scientifiques utilisateurs.

AXES DE RECHERCHE

Une partie des challenges ouverts par la bioinformatique à grande échelle (génomique, protéomique, biologie moléculaire, *etc*) doivent maintenant être résolus par l'informatique répartie et plus particulièrement les grilles de calcul. Pour permettre l'exploitation efficace d'infrastructures distribuées et hétérogènes, je propose un projet scientifique qui se décompose suivant trois axes de recherche cohérents et complémentaires, qui s'appuiera sur les interactions avec les scientifiques consommateurs de moyens informatiques à

savoir : (i) un formalisme simple, accessible et compréhensible par l'utilisateur non expert, permettant l'expression d'applications gridifiées sous forme d'enchaînement de programmes ; (ii) des métriques d'analyse de performances des différentes grilles (dédiée, partagée et volatile) accessible aux scientifiques ; et (iii) des stratégies d'exploitation inter-grilles efficaces prenant en compte les spécificités et les atouts de chacune d'elles.

Mon projet de recherche s'appuie sur mon expérience dans des projets d'envergure internationale (Décrypthon, World Community Grid, ANR GWENDIA) et sur les aspects innovants des grilles informatiques pour la recherche informatique appliquée au sciences du vivant.

Dans mon première axe de recherche, je m'intéresse à l'exploitation efficace des infrastructures distribuées pour le traitement de grandes quantités de données grâce à des outils permettant l'expression des flots de traitements (workflow), la gestion des données scientifiques distribuées, et l'allocation des calculs sur des infrastructures hétérogènes de type grille. Au cours de ma thèse, je me suis intéressé, en particulier, aux stratégies d'ordonnancement des flots de travaux, modélisés sous forme de DAG (*Directed Acyclic Graph*), dans un contexte où plusieurs utilisateurs accèdent aux mêmes ressources d'une grille. À cette occasion, j'ai développé un environnement et des fonctionnalités intégrant les préoccupations de la communauté qui cherche à traiter leurs données scientifiques volumineuses par des procédures d'analyses complexes conceptualisées sous forme d'un enchaînement de tâches. Le déploiement des applications est assuré sur une plate-forme grille et les utilisateurs composent les programmes « gridifiés » de manière simple. Dans les approches d'autres moteurs d'exécution de workflows (Pegasus/DagMan, Triana, MOTEUR, *etc*) l'enchaînement des tâches (*workflow fonctionnel*) est séparé de la description des données. Chacun de ces outils propose des langages de description des applications qui leur sont propres (Scufl, DAX, BPEL, Swift, *etc*). Je propose de travailler au développement d'un formalisme commun et d'outils permettant de prendre en compte séparément le *workflow fonctionnel* et les données appliquées à celui-ci. Ce formalisme permettra une description des traitements utilisateurs indépendantes de l'infrastructure qui exécutera l'application et des stratégies d'ordonnancement de DAG élaborées tirant parti des spécificités de l'infrastructure cible. L'enjeu est de fournir à l'utilisateur non spécialiste un outil simple et adapté à l'exploitation d'environnements massivement parallèles comme les grilles. Ce travail poursuit l'effort initié dans l'ANR MDCA GWENDIA, aussi je propose de l'intégrer et de le développer dans un groupe de recherche de standardisation au sein de l'*Open Grid Forum* dans le but de produire un standard pour l'expression des workflows scientifiques.

Dans un second axe de recherche, je me focalise sur l'analyse et la prise en compte de plusieurs architectures de grilles aux caractéristiques et comportements différents (grille d'internautes, grille dédiée, grille partagée). Ce deuxième sujet de travail étudie la comparaison des grilles de calcul et des différentes architectures suivant plusieurs critères (échelle, coût, consommation électrique, disponibilité, performance, fiabilité, qualité de service *etc*). Au cours de ma thèse, je me suis attaché à produire une analyse comparative entre une infrastructure dédiée et la grille de volontaires World Community Grid à partir des besoins de l'application d'ancrage moléculaire MAXDo du projet « Help Cure Dystrophy Musclar ». À cette occasion, j'ai établi un rapport de performance entre des ressources dédiées et des ressources volatiles d'internaute. Il me semble important d'observer dans quelle mesure le facteur de comparaison évoluera avec le renouvellement des machines des volontaires. Je pense déjà à baser mon analyse sur les points/crédits accordés aux volontaires lorsqu'ils contribuent aux projets du World Community Grid. En outre, dans un contexte où la préoccupation d'économie d'énergie devient de plus en plus importante pour les particuliers et pour les centres de calcul, je pourrais établir les profils de consommation électrique des machines lorsqu'elles travaillent sur les calculs scientifiques. Ainsi, je serai capable de répondre à la question suivante : quelle est la proportion d'énergie consommée par une machine restant allumée mais inactive par rapport à une machine calculant sur des projets scientifiques. De surcroît, il serait intéressant d'établir d'autres points de comparaison avec d'autres travaux semblables menés sur des grilles de production telles que les *data challenge* de l'initiative WISDOM sur la grille européenne EGEE. Ainsi, nous aurions une analyse comparative complète des infrastructures de grille accessibles aux scientifiques : ressources dédiées et réservées (Cloud computing type Amazon EC3/S3, Grid'5000), ressources partagées et hébergées dans des centres de calcul (EGEE, Décrypthon, *etc*) et ressources volatiles (World Community Grid, grille d'internautes).

Cet axe de recherche soulève des problématiques d'actualité au moment où les mastodontes du monde informatique (Amazon, Google, HP, IBM, Microsoft, Sun Microsystems, *etc*) prennent place et travaillent à fournir une offre commerciale basée sur la virtualisation des ressources (« cloud computing ») répondant aux contraintes de confidentialité et de qualité de services. Aussi, il est important de pouvoir fournir des outils et une analyse comparative entre les différents moyens de calcul mis à la disposition des scientifiques pour leur permettre de les utiliser au mieux.

Dans un dernier axe de recherche, je propose d'étendre mes travaux de thèse sur les stratégies d'ordonnancement multi-utilisateurs et multi-workflows. Il reste encore des champs d'heuristiques à explorer. Je me suis principalement intéressé aux techniques d'ordonnancement dites *list scheduling* qui fixent une priorité aux tâches d'un DAG pour ensuite les affecter à une ressource disponible suivant un critère de choix. Il existe d'autres stratégies applicables, comme la duplication ou le regroupement de tâches que j'ai détaillé dans mon état de l'art. De plus, les critères de choix auxquels j'ai réfléchi, privilégiaient la minimisation du temps de fin de l'application ou l'équité d'accès aux ressources entre les utilisateurs, aussi, nous pourrions incorporer des critères issus de l'étude des différents types de grilles, à savoir (coût, fiabilité, rapport qualité/prix, énergie consommée, disponibilité, *etc*). De cette manière, je travaillerai à la prise en compte de l'accès et des spécificités de plusieurs infrastructures de grilles (dédiée, partagée ou volatile) dans des heuristiques d'ordonnancement qui conjugueront à la fois les aspects multi-utilisateurs et multi-grilles. De cette façon, je proposerai de nouvelles stratégies d'ordonnancement qui exploiteront ces facteurs dans le but d'optimiser suivant les besoins de l'utilisateur l'exécution de son application.

Enfin, il est important de souligner que mes perspectives de recherche mettent au centre de mes préoccupations les besoins des utilisateurs. Ainsi, à l'image de mon travail autour du programme Décryphon et de mon post-doc, il sera nécessaire de disséminer et de promouvoir les connaissances autour des technologies de grilles auprès des scientifiques non spécialistes. Cette étape est très enrichissante humainement et intellectuellement. Elle permet de saisir les points de vue et le vocabulaire des domaines scientifiques qui appréhendent l'informatique comme un outil de travail sans chercher forcément à optimiser leurs programmes pour des traitements parallèles efficaces. Le rapprochement entre l'informatique du parallélisme des grilles et les autres sciences explore de nouveaux champs de connaissances pluridisciplinaires. Cette diffusion du savoir passera par une rationalisation des standards (groupes de réflexion et de standardisation de l'OGF), par une simplification et une meilleure compréhension des outils informatiques disponibles et enfin par une optimisation des applications

Ce projet de recherche apparaît prometteur et porteur d'écho dans un monde où les sciences du vivant et la bioinformatique sont un enjeu crucial. De plus, le projet que je propose est à l'image de ma thèse et de mon post-doc, ils se retrouvent à la croisée de plusieurs domaines de recherche, ils s'intègrent dans un contexte pluridisciplinaire dans le but de produire de la valeur ajoutée scientifique et technique inter-domaine. Par ailleurs, il entre en synergie avec la création de l'Institut des grilles par le CNRS qui vise à fédérer les efforts dans les infrastructures et à faciliter l'accès aux utilisateurs.

INTÉGRATION À L'IGBMC

Mon intégration dans l'« Institut de Génétique et de Biologie Moléculaire et Cellulaire » apportera à la fois une expertise dans les technologies de grille permettant le traitement informatique en masse des données biologiques et une mixité des connaissances au sien d'une équipe dynamique à la pointe de son domaine. Ma venue au sein de l'IGBMC et plus particulièrement dans l'équipe dirigée par Olivier Poch apparaît évidente de part les thématiques de recherche que je vise et les compétences que j'ai développé à la fois dans les sciences de la vie et l'informatique distribuée. Il est particulièrement intéressant et motivant pour ma carrière de chercheur que j'intègre ce laboratoire de haut niveau et l'équipe d'Olivier Poch dont les préoccupations tournées vers la bioinformatique à grande échelle trouvent une extension naturelle et nécessaires vers les grilles de calcul. Par ailleurs, les applications biologiques développées dans l'équipe seront des cas concrets d'utilisation qui me permettront de poursuivre les objectifs scientifiques que j'ai

fixés. D'autre part, mes projets de recherche sur les workflows et l'analyse des différents types de grilles (Cloud Computing, grille volatile et grille partagée) apporteront une dimension supplémentaire à l'IGBMC tout en me laissant le champs libre pour développer ma recherche. J'ai déjà eu des interactions marquées avec Olivier Poch et son équipe qui ont donné lieu à une publication commune. La conjonction de ces facteurs me laissent à penser que l'environnement de travail de cette équipe stimulera ma créativité.

COLLABORATIONS ET OUVERTURES

Ma candidature au CNRS dans l'équipe d'Olivier viendra enrichir le réseau de collaboration de l'IGBMC des liens privilégiés que j'ai sus nouer au cours de ma jeune carrière de chercheur. D'une part, ma formation initiale au Laboratoire de l'informatique du Parallélisme (LIP) de l'ENS-Lyon, d'autre part ma participation déterminante au projet Décryphon apportera les interactions et les échanges scientifiques nécessaires au brassage d'idées et à la pluri-disciplinarité. Il est d'ors et déjà établi que je continuerai à avoir des liens avec le programme Décryphon et avec les équipes scientifiques qui souhaitent capitaliser nos échanges. De même, mon post-doc aux États-Unis dans l'équipe d'Ewa Deelman du laboratoire « Information Sciences Institute » de l'université de Californie du Sud (USC) apporte une stature international à mon profile. De surcroît, j'ai su interagir et développer un réseau de connaissance au sein de l'entreprise IBM aussi bien en France (programme Décryphon) qu'aux État-Unis au travers du World Community Grid qui laisse présager de collaborations futures. Enfin, mes capacités d'intégration et ma curiosité scientifique me permettront de développer de nouvelles interactions profitables pour l'IGBMC et le CNRS.

ACTIVITÉS D'ENSEIGNEMENTS

Pendant la période 2005-2008, j'ai été moniteur de l'École Normale Supérieure de Lyon. À ce titre, j'ai effectué mon service annuel de 64 heures (équivalent TD) dans le Département Informatique (DI) de l'ENS-Lyon. Les enseignements se répartissent de la manière suivante :

2007-2008 (64 h)

- **Algorithmique des Réseaux et Télécommunication**, enseignant Eric Fleury, Niveau Master 1 (32 h) : Pour la deuxième année consécutive, j'ai assuré les TD et TP pour cette matière. L'enseignant de cette matière étant nouveau dans le corps des enseignants de l'ENS, j'ai été en charge de faire le lien entre les deux années d'enseignements afin d'assurer une cohérence dans la formation. A cette occasion la majeure partie des TD/TP ont été réécrits et adaptés au contenu du cours du nouvel enseignant.
- **Architecture, Systèmes et Réseaux**, enseignant Eric Fleury, Niveau Licence 3 (32 h) : Ce cours a pour objectif de présenter aux étudiants les problématiques de conception des systèmes d'exploitation. La gestion des ressources systèmes du point de vue du système d'exploitation sera abordée : gestion de processus, gestion de la mémoire primaire, systèmes de fichiers et gestion de la mémoire secondaire, protection ... Une seconde partie du cours s'intéressera aux aspects réseaux inhérents aux systèmes d'exploitations. J'ai assuré les TD/TP de cette matière.

2006-2007 (64 h)

- **Tuteur du projet Compilation**, enseignant Daniel Hirschkof, Niveau Licence 3. (équivalent 30h) : Ce projet s'est étalé sur un semestre. Les élèves avaient pour but de réaliser un compilateur : À partir d'un langage initial (Cool, Lucid synchrone, MetaOCaml, miniCaml) faire l'analyse lexicale et syntaxique pour ensuite construire l'arbre de syntaxe abstrait, et finir par générer du code compilé. J'étais présent aux cotés de l'enseignant afin d'aider les élèves en cas de difficultés. À la fin du projet, j'ai été en charge de l'évaluation de la partie code pour les groupes d'élèves ayant choisi le langage JAVA.
- **Initiation à la programmation java et à la programmation orientée objet**. Niveau Licence 3 (2h) : Au début du projet de compilation, j'ai donné un cours magistral en classe entière afin de donner les notions de bases en JAVA.
- **Algorithmique des Réseaux et Télécommunication**, enseignant Anne Benoit, Niveau Master 1 (32 h) : ce cours s'intéresse aux techniques algorithmiques liées aux nouvelles technologies et plates formes réseaux. Nous avons abordé notamment : Routage dans les réseaux, Réseaux pair à pair, Réseaux ad-hoc et sans fil, Internet et le Web. J'ai assuré les TD et TP en collaboration avec l'enseignant et un autre moniteur. Nous avons repris les TD/TP existants et en avons créé de nouveaux. De plus, lorsque la correction des exercices n'était pas faite, nous avons produit une correction permettant de pérenniser la base d'exercices. Enfin, j'ai établi le sujet du partiel de cette matière.

2005-2006 (62 h)

- **Responsable de la gestion des ordinateurs portables** prêtés aux élèves du Département Mathématiques et informatiques de l'ENS de Lyon. (équivalent 32 h) : mon travail consistait à installer les machines (double boot : windows, linux) et à aider les élèves en cas de problème d'utilisation de ces machines. Dans ce cadre, j'ai mis en place une gestion des ordinateurs permettant de réinstaller une machine à la demande, et ainsi revenir à un état fonctionnel, tout en conservant les données des utilisateurs.
- **Systèmes Dynamiques Discrets**, enseignant Jacques Demongeot, Niveau Master 1 (30 h) : Ce cours est constitué de deux parties. Dans la première nous avons étudié d'un point de vue théorique le comportement de certains systèmes dynamiques classiques (déterministes et stochastiques). Nous montrons dans une deuxième partie comment de tels systèmes sont utilisés en pratique pour modéliser des situations issues de domaines scientifiques aussi variés que les sciences humaines ou la biologie.

Période	Intitulé	horaires, niveau
2007-2008	Algorithmique des Réseaux et Télécommunications	32h, M1
	Architecture, Systèmes et Réseaux	32h, L3
2006-2007	Co-gestion du projet compilation	32h, L3
	Algorithmique des Réseaux et Télécommunications	32h, M1
2005-2007	Systèmes dynamiques discrètes	32h, L3
	Gestion des ordinateurs portables	32h, L3

TAB. 1 – Récapitulatif Monitorat, Département Informatique, ENS-Lyon.

Nous avons étudié en détail des exemples issus de ces domaines : diffusion de la criminalité pour les sciences humaines ; système de contrôle génétique de la morphogénèse pour la biologie. J’ai assuré les TD de ce cours à partir des sujets fournis par l’enseignant.

Autres responsabilités et activités liées à l’enseignement

- **Membre du Jury d’admission pour les auditeurs libres de l’ENS** : J’ai participé au jury d’admission des élèves désirant entrer en tant qu’auditeur libre à l’ENS en L3. Dans ce jury, j’avais pour rôle de donner mon avis sur les élèves et de répondre aux questions des candidats sur les conditions de vie au sein de l’école.
- **Membre du jury de soutenance de stage Niveau Licence 3** : J’étais chargé de la lecture et l’appréciation d’une partie des rapports de fin de stage des étudiants.
- En outre, lors de ma formation d’ingénieur et de mon DEA à l’École Polytechnique de l’Université de Nantes, j’ai eu l’occasion d’assurer au cours de l’année 2003 une trentaine d’heures de formation au langage JAVA. Cette formation a été dispensée dans le cadre de la formation « DU Chef de projets logiciels et réseaux ».

Dans la suite de ce dossier, vous trouverez les documents suivants :

- [Rapport de soutenance](#)
- Rapports de thèse
 - [Rapport de Pierre Manneback - Professeur](#)
 - [Rapport de Vincent Breton - DR CNRS](#)
- [Attestation de thèse](#)
- [Attestation de monitorat](#)
- Lettres de recommandation
 - [Lettre de Frédéric Desprez](#) - Directeur de thèse ;
 - [Lettre de Dino Moras](#) - Directeur de l'IGBMC ;
 - [Lettre de Jean Lemerle](#) - Membre Comité Directeur Décrypthon ;
 - [Lettre de Thierry Tournel](#) - Chef de projet AFM ;
 - [Lettre de Bill Bovermann](#) - Project Executive for the IBM Corporation ;
 - [Lettre d'Alessandra Carbone](#) - Professeur Université Pierre et Marie Curie et chef d'équipe « Analytical Genomics » ;
 - [Lettre de Christophe Pouzat](#) - CR1 CNRS et chef de l'équipe « Statistiques Computationnelles pour l'Analyse de Données Neurophysiologiques » ;

RAPPORT DU PRÉSIDENT DE JURY
APRÈS LA SOUTENANCE ET SIGNATURE
DES MEMBRES DU JURY

Thèse de Raphaël BOLZE

Raphaël BOLZE est l'auteur d'un travail remarquable sur l'analyse et le déploiement d'applications bio-informatiques au sein d'infrastructures de grille. Ce travail a été réalisé dans un contexte pluridisciplinaire avec une collaboration étroite avec des chercheurs d'autres disciplines, notamment des biologistes et des bioinformaticiens. Il s'inscrit également dans le cadre d'une collaboration entre l'Association Française contre le Myopalmes et son projet phase le Décrypton, la société IBM et le CNRS. Les contributions scientifiques de Raphaël Bolze concernent, entre autres, l'étude et l'expérimentation d'heuristiques pour l'ordonnement simulé de plusieurs workflows en provenance d'applications bioinformatiques.

Le jury a souligné que Raphaël Bolze a joué un rôle pivot dans le projet Décrypton qui, grâce à sa grande connaissance des technologies grilles, a permis à ce projet de prendre des décisions sur le choix des technologies à employer qui se sont révélées d'une grande justesse a posteriori. Ce travail a été accompagné d'un développement logiciel conséquent de l'intergiciel de grille DIET qui a permis de passer du stade de prototype de recherche à une solution dont la robustesse a autorisé son déploiement dans un contexte de production. Il est à noter que le travail de Raphaël Bolze a joué un rôle fondamental dans la création d'une start-up pour commercialiser

l'intégral DiET.

Le document de thèse, de taille conséquente, est de très bonne qualité et agréable à lire ce qui démontre les capacités de Raphaël Balze à bien expliquer l'ensemble de ses contributions. Le jury a apprécié un exposé clair et bien structuré qui a présenté les principales contributions apportées par Raphaël Balze dans le cadre de son travail de thèse. Cette présentation a révélé de très bonnes aptitudes pédagogiques, la grande expertise et connaissance du domaine, et la capacité à bien synthétiser le résultat de ses travaux. Enfin le jury tient à souligner la pertinence des réponses aux questions qui lui ont été posées ce qui montre une très grande maîtrise du sujet.

En conclusion, le jury décerne le titre de Docteur de l'Université de Lyon, Ecole Normale Supérieure de Lyon, spécialité Informatique à Raphaël Balze, avec la mention Très Honorable.

PRIOZ

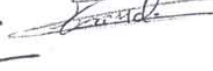
T. TOUSSAET

J. DELPLANCK

P. MANVERCH

F. DESPREZ











RAPPORT

sur le manuscrit de thèse présenté par Monsieur Raphaël BOLZE

Analyse et déploiement de solutions algorithmiques et logicielles pour des applications bioinformatiques à grande échelle sur grille

**Prof. Pierre Manneback
Faculté Polytechnique de Mons, Belgique
10 octobre 2008**

Le travail de Raphaël BOLZE présente des expérimentations d'utilisation de grilles informatiques pour des applications bioinformatiques aux besoins de calcul considérables. Les grilles exploitées sont essentiellement les grilles d'internautes, réputées instables et volatiles. Les intergiciels (*middleware*) doivent être adaptés à ces infrastructures, tout comme les classes d'applications qui peuvent y tourner efficacement. Bon nombre d'applications bioinformatiques se prêtent bien à ces architectures, qui ont l'avantage de pouvoir fédérer de très nombreuses ressources informatiques à un coût raisonnable.

Le premier chapitre du travail reprend de façon claire les définitions des différents types de grilles informatiques, ainsi que quelques exemples connus de grilles exploitées. En ce qui concerne les ressources d'une grille informatique, l'auteur cite les ressources de communication, de calcul, de stockage, ainsi que les ressources humaines. Si on avait voulu être plus global, on aurait pu également citer les ressources d'entrées/sorties (capteurs, actuateurs, etc), voire les ressources énergétiques.

Le chapitre 2 décrit le programme Décryphon et les projets scientifiques associés. Le Décryphon est un programme de recherche coordonné et financé par l'Association Française contre les Myopathies (AFM), le CNRS et IBM. Il vise à soutenir des projets exigeant des moyens de calcul ou de stockage considérables, projets traitant de recherche en génomique et protéomique. 8 projets sélectionnés durant le travail de thèse sont décrits de manière synthétique mais accessible pour les non-spécialistes.

Au chapitre 3, l'auteur commence par décrire la procédure d'évaluation des intergiciels pour la grille du Décryphon. La méthodologie d'évaluation est détaillée et rigoureuse. Il ne manque pratiquement aucun critère, la difficulté étant peut-être d'en avoir trop pour décider, une sous- ou une sur-pondération de certains critères pouvant biaiser la décision. Certains sous-critères sont questionnables : par exemple, le nombre maximal de tâches contrôlables en parallèle peut être dépendant de la taille de ces tâches, le critère synchrone/asynchrone n'est pas clair.

Le processus d'utilisation de la grille Décryphon I, exploitée avec l'intergiciel GridMP, est ensuite décrit. Une section motive en 2007 le passage à l'utilisation de l'intergiciel DIET, comme solution de logiciel libre, avec les avantages d'adaptabilité possible. L'architecture de DIET (*Distributed Interactive Engineering Toolbox*, développé à l'ENS Lyon) est ainsi présentée de manière claire et pédagogique. L'adaptation de DIET à la grille

Décryphon s'est déroulée avec succès, et dans un temps minimal. Un portail web DIET (*DIET_Webboard*) a été développé, dont les fonctionnalités évoluées, notamment de déploiement et de gestions des utilisateurs, facilitent considérablement le travail d'administration de la grille. Cet outil est générique et donc utilisable par d'autres grilles que celle du Décryphon.

Le chapitre 4 présente les expériences menées de portage d'applications bioinformatiques sur grille. Dans un premier temps, des expériences ont été conduites sur l'utilisation de Grid'5000 avec DIET. L'objectif initial était de mesurer le nombre de traitement de tâches par seconde. Ces tâches étaient régulières (multiplication matricielle 10x10). Les résultats de scalabilité sont satisfaisants jusqu'à plus de 1000 processeurs. L'auteur a amélioré les fonctionnalités de déploiement de DIET sur grille, en proposant un nouvel outil de génération de configuration (*XmlGoDIETGenerator*). De même, un outil de monitoring (*LogService*) adossé à DIET a été conçu et développé. Il s'agit d'un outil très précieux pour le développement d'une application sur grille, car les performances dépendent d'un nombre important de paramètres, qu'il est difficile d'évaluer sans monitoring. L'auteur a procédé à une modélisation rigoureuse du nombre de messages générés par le service de monitoring. Il a pu mesurer que ce service n'entravait pas les performances intrinsèques de DIET.

Les expériences ont été ensuite conduites pour une application d'amarrage moléculaire (MAXDo), provenant du projet HCMD (*Help Cure Muscular Dystrophy*) du Décryphon. Il s'agit de trouver la meilleure façon d'apparier deux protéines pour former un complexe protéique. Cette application nécessite un nombre considérable de calculs, étant donné le nombre de configurations possibles. Grid5000 a été utilisée pour estimer le temps de calcul total nécessaire pour l'amarrage de 168 protéines, soit près de 15 siècles sur un processeur standard 2GHz. L'application a alors été portée sur la grille d'internautes de la *World Community Grid*. L'ensemble des calculs a duré 26 semaines. L'auteur relate avec précision le déploiement, le déroulement et les résultats de l'expérience. Le concept intéressant de *processeur virtuel équivalent temps plein* est proposé. Il donne une borne minimale de processeurs dont on a eu besoin durant une période donnée pour effectuer un calcul donné. Cet indicateur représente fidèlement l'activité d'une application sur la grille d'internautes.

Enfin, le chapitre 5 présente le problème très étudié d'ordonnement de graphes de tâches sur grille. Il rappelle de manière pédagogique les différents types de workflow et les algorithmes d'ordonnement classiques y associés. Il donne un rapide aperçu des principaux gestionnaires d'exécution de workflow. Il s'intéresse plus particulièrement aux environnements multi-utilisateurs et multi-applications. Dans cette optique, il propose différentes variantes de l'algorithme HEFT (*Heterogeneous Earliest Finish Time*), différant suivant le type de sélection de tâches multi-DAG (un graphe orienté acyclique par application).

Les algorithmes d'ordonnement sont implantés sous DIET, puis testés et validés. L'approche originale est d'avoir travaillé avec des multi-DAG.

Conclusion

Le manuscrit reflète un travail important de compréhension, maîtrise et implantation d'outils intergiciels et d'applications bioinformatiques sur grille. Il comprend certaines contributions significatives, comme la conception des outils d'administration et de monitoring de DIET, l'implantation d'une application d'amarrage moléculaire sur une grille d'internautes et la proposition d'algorithmes d'ordonnement multi-DAG. Certaines ont été publiées dans des journaux ou conférences internationales. Pour ces raisons, j'é mets un avis favorable à la soutenance de la thèse de Raphaël Bolze en vue de l'obtention du titre de Docteur de l'Ecole Normale Supérieure de Lyon.

Rapport rédigé à Mons (Belgique) le 10 octobre 2008



Pierre MANNEBACK
Professeur



Dr Vincent Breton
Directeur de recherches au CNRS
Laboratoire de Physique Corpusculaire de Clermont-Ferrand
Campus des Cézeaux
63177 Aubière Cedex
France

Aubière, le 10 Octobre 2008

Rapport sur la thèse de Mr Raphaël Bolze

Etabli par Vincent Breton, au lu de son manuscrit

La présentation de l'ensemble de la thèse de Mr Raphaël Bolze intitulée « Analyse et déploiement de solutions algorithmiques et logicielles pour des applications bioinformatiques à grande échelle sur la grille » est très agréable. Le manuscrit est rédigé de façon très satisfaisante.

La thèse de Monsieur Raphaël Bolze porte sur les enjeux de l'utilisation des grilles pour la bioinformatique. Le plan de la thèse est clair : après une introduction générale, le chapitre 1 propose une présentation générale des grilles informatiques. Il commence par une définition des concepts principaux illustrés sur trois grilles actuelles : EGEE, Grid'5000 et World Community Grid. Le choix de ces projets est judicieux car ces grilles sont emblématiques des infrastructures distribuées déployées aujourd'hui dans le monde. Le chapitre dans son ensemble constitue une très bonne synthèse de l'état de l'art.

Le chapitre 2 introduit l'organisation et les enjeux scientifiques de la grille Décryphon à travers une description résumée des projets qui ont fait l'objet ou sont encore en cours de déploiement. Le travail de synthèse est rendu difficile par la variété des problématiques biologiques abordées mais le chapitre permet aux lecteurs de se faire une idée assez claire des enjeux scientifiques des différents projets et de l'importance des besoins de calculs qu'ils induisent.

Le chapitre 3 nous place au cœur de la problématique du déploiement d'une grille de production. En 2004, peu d'intergiciels étaient suffisamment robustes pour gérer efficacement plusieurs centaines de processeurs distribués géographiquement. Le choix initial d'une solution propriétaire proposée par la société United Devices s'inscrit dans ce contexte. Avec les progrès significatifs accomplis par les groupes académiques dans le développement de solutions logiciels libres, le choix de DIET à partir de 2007 apparaît excellent. Le chapitre fournit un ensemble d'informations pertinentes sur les critères d'évaluation des intergiciels suivant lesquels il serait intéressant d'analyser les solutions actuelles. Il décrit aussi en détails les fonctionnalités et le fonctionnement de DIET.

Le chapitre 4 intitulé « expériences dimensionnantes sur grilles informatiques » propose une description et une analyse du déploiement de DIET sur l'infrastructure Grid'5000 et du déploiement de calculs d'ancrage moléculaire sur World Community Grid, la grille volontaire d'IBM. La comparaison des performances de la grille dédiée Grid'5000 et de la grille volontaire WCG est très originale et particulièrement intéressante. Elle met en évidence les atouts d'une grille volontaire comme WCG, notamment en termes de volumes de ressources, mais aussi les facteurs d'inefficacité qui réduisent son impact par rapport à une grille dédiée comme Grid'5000. Cette comparaison abondamment documentée met finalement bien en évidence la complémentarité entre ces infrastructures.

Le chapitre 5 présente les enjeux de l'ordonnancement de graphes de tâches sur les grilles de calcul. Il introduit plusieurs heuristiques pour la gestion des « workflows » dont plusieurs ont été implémentées dans l'intergiciel DIET et testées avec succès sur trois applications scientifiques du Décryphon.

Au final, il apparaît clairement que Mr Raphaël Bolze a effectué un ensemble de propositions pertinentes et adaptées à la problématique du traitement des données biologiques dans des environnements de grille. Le manuscrit présente un travail que j'ai beaucoup apprécié et qui constitue, sans aucun doute, une contribution tout à fait significative. Il faut souligner en particulier la démarche très originale de coupler différentes infrastructures de grille (Grid'5000, World Community Grid et Décryphon) pour l'étude d'une thématique scientifique, en l'occurrence le docking moléculaire. Comme le montre clairement le manuscrit, la grille Décryphon constitue aujourd'hui un environnement privilégié pour les déploiements de calculs à grande échelle de la communauté française de bioinformatique, environnement dont la gestion est assurée par un intergiciel issu directement de la recherche en informatique française. Il est important de souligner que Raphaël Bolze a contribué significativement à enrichir les fonctionnalités de DIET au cours de sa thèse.

Je conclurai donc en donnant un avis très favorable pour que ce manuscrit soit présenté et défendu en vue de l'obtention par Mr Raphaël Bolze du grade de docteur.

Fait à Clermont-Ferrand, le 10 Octobre 2008



Vincent Breton
Directeur de Recherches au CNRS

ATTESTATION

Le directeur de l'école normale supérieure de Lyon, soussigné, certifie que

*Monsieur Raphaël BOLZE, né(e) le 28/08/79 à VALENCE,
inscrit(e) à l'école normale supérieure de Lyon, sous le numéro d'étudiant 147853,
a soutenu, le 31/10/08, conformément aux règlements en vigueur, la thèse intitulée
"Analyse et déploiement de solutions algorithmiques et logicielles pour des applications bioinformatiques à grande
échelle sur la grille."
devant le jury composé de :*

Président

Monsieur Thierry PRIOL

Membres

Monsieur Vincent BRETON

Monsieur Jacques DELPLANCQ

Monsieur Frédéric DESPREZ

Monsieur Pierre MANNEBACK

Monsieur Thierry TOURSEL

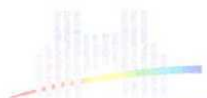
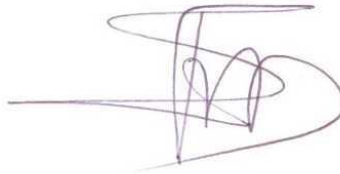
*A l'issue de cette soutenance, le jury a accordé à Monsieur Raphaël BOLZE, le grade de Docteur
de l'Université de Lyon délivré par l'Ecole Normale Supérieure de Lyon, spécialité Informatique, avec la
mention "très honorable",
pour en jouir avec les droits et prérogatives qui y sont attachés par les lois, décrets et règlements.*

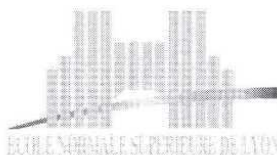
Fait à Lyon le 19/11/2008, pour servir et valoir ce que de droit

Le directeur de l'École normale supérieure de Lyon

*N.B. : Le Conseil Scientifique de l'ENS Lyon a
décidé, le 06 mai 1997, de ne plus décerner la mention
"très honorable avec les félicitations du jury".*

Jacques SAMARUT





ÉCOLE NORMALE SUPÉRIEURE DE LYON

46 allée d'Italie 69364 LYON Cédex 07 - FRANCE

Département Informatique

Tél : (+33) 04 72 72 86 52 - Fax : (+33) 04 72 72 86 93

E-mail : departement.info@ens-lyon.fr

Web : <http://www.ens-lyon.fr/DI/>

ATTESTATION DE MONITORAT

Je soussigné Monsieur **Yves ROBERT**, Professeur, Responsable du Département Informatique à l'École Normale Supérieure de Lyon, certifie que :

Monsieur Raphaël BOLZE né le 28 août 1979 à VALENCE (26 – Drôme)

a exercé la fonction de **Moniteur** au sein de l'École Normale Supérieure de Lyon :

ANNÉES	Intitulés des enseignements et nature (TD, TP)	Volume horaire
2005-2006	• Responsable de la gestion des ordinateurs portables du Département Informatique	32
	• Systèmes Dynamiques Discrets (TD)	32
2006-2007	• Tuteur du projet Compilation et Initiation JAVA	32
	• Algorithmique des Réseaux et Télécommunication (TD)	32
2007-2008	• Algorithmique des Réseaux et Télécommunication (TD)	32
	• Architectures Systèmes et Réseaux (TD)	32
Répartition du service équivalent TD		192

Pour servir et valoir ce que de droit.

Fait à Lyon le 2 décembre 2008






Laboratoire de l'Informatique du Parallélisme
Unité mixte CNRS - ENS Lyon - UCB Lyon n°5668 associée à l'INRIA

Frédéric Desprez, Directeur de Recherche Inria
Laboratoire LIP, Equipe CNRS-ENSL-UCB-INRIA GRAAL,
École normale supérieure de Lyon,
46 Allée d'Italie, 69364 Lyon Cedex 07 France
Email : Frederic.Desprez@ens-lyon.fr

Lyon, le 02/01/09

Objet : Lettre de recommandation pour Raphaël Bolze

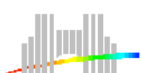
Messieurs,

Après avoir été recruté comme ingénieur expert dans mon équipe dans le cadre du projet Décryphon (AFM-CNRS-IBM), Raphaël Bolze s'est décidé à faire une thèse et j'ai obtenu pour lui une bourse de thèse BDI au CNRS. Les travaux de Raphaël ont été sur trois points. Il y a eu une partie importante autour du choix du logiciel, de la conception de l'architecture et mise en place de la grille de production. Même si il s'agit d'une partie plus logicielle, elle a nécessité l'acquisition de compétences techniques sur l'existant et sur le déploiement d'une grille de production pour la bioinformatique. Il a également participé à des développements autour du logiciel DIET développé dans l'équipe GRAAL, notamment autour du monitoring de la plateforme et des ordonnanceurs modifiables (avec validation sur Grid'5000). Ensuite, il a étudié le déploiement d'une application sur une grille d'internautes avec l'équipe du World Community Grid. Il a comparé les résultats de la « gridification » de cette application avec une étude théorique pour une grille de production. Il a enfin étudié des heuristiques d'ordonnement de workflows multiples sur grilles de serveurs avec notamment étude de l'équité entre les applications. Les résultats ont été validés à la fois théoriquement et expérimentalement sur Grid'5000. En parallèle avec ces travaux, il a géré le projet Décryphon, assisté le portage d'application (dont une a donné lieu à une publication dans la revue Bioinformatics).

Les travaux de Raphaël autour des grilles de production et des grilles d'internautes ont été largement validés théoriquement, expérimentalement et enfin publiés. Le projet Décryphon est un succès et il nous a permis de valider le logiciel DIET pour lequel Raphaël a participé au développement. Raphaël partira en postdoc au USA à Los Angeles, dans l'équipe d'Ewa Deelman à partir de Mars 2009. Il possède des qualités de chercheur indéniables et une capacité à s'intéresser à des applications d'autres sciences, notamment dans les sciences de la vie. Je soutiens donc très vivement l'obtention d'un poste de chargé de recherche au CNRS pour Raphaël Bolze.

Frédéric Desprez

Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France
Tél. (+33) 4 72 72 80 37 Fax (+33) 4 72 72 88 06, Adresse électronique : lip@ens-lyon.fr



Illkirch, le 19 Décembre 2008

A qui de droit,

Par la présente je signifie mon accord enthousiaste à la venue du Dr. Raphaël Bolze à l'IGBMC. Il intégrera et renforcera le Laboratoire de Bioinformatique et Génomique Intégratives dirigé par le Dr. Olivier Poch. Cette équipe de renommée internationale est impliquée dans le développement d'outils et de stratégies d'analyse des données biomédicales provenant notamment, des maladies génétiques et des cancers. En renforçant l'axe de recherche sur les technologies des grilles, la venue de Mr Bolze constituera un atout supplémentaire dans la volonté de l'institut de développer de nouvelles approches informatiques pour générer et exploiter les données à haut débit de la biologie moderne et contribuera à l'essor de la biologie Intégrative. Cet axe de recherche majeur est soutenu par les investissements du CPER dans le cadre de la construction prochaine du Centre de Biologie Intégrative. Le Dr. Bolze sera accueilli dans les locaux existants et bénéficiera des ressources de calcul de l'IGBMC.

Recevez, Madame/Monsieur, mes respectueuses salutations.



Dr. Dino Moras

Directeur : Dino Moras
IGBMC UMR7104 – U596
1 rue Laurent Fries
BP 10142
F - 67404 ILLKIRCH Cédex
Tél : + 33 (03) 88 65 32 20
Fax : + 33 (03) 88 65 32 01
igbmc@igbmc.u-strasbg.fr
www.igbmc.u-strasbg.fr

PJ: organigramme de l'équipe

Organigramme

Laboratoire de Bioinformatique et Génomique Intégratives

8 statutaires

O. POCH	DR2 CNRS
J. THOMPSON	CR1 CNRS
O. LECOMPTE	MC1 ULP
N. WICKER	MC1 ULP
R. RIPP	IR1 CNRS
W. RAFFELSBERGER	IR2 CNRS
L. MOULINIER	IE CNRS
L. POIDEVIN	IE ULP

13 doctorants et post-doctorants

L.P. ALBOU	Doctorant
G. ALTOBELLI	Post-doctorant
R. ANIBA	Doctorant
Y.N. ANNO	Doctorant
F. BEDEZ	Doctorant
Y. BRELIVET	Post-doctorant
A. FRIEDRICH	Post-doctorant
N. GAGNIERE	Doctorant
D. KIEFFER	Doctorant
N.H. NGUYEN	Post-doctorant
E. PERRODOU	Post-doctorant
F. PROSDOCIMI	Post-doctorant
V. RUANO-RUBIO	Post-doctorant

Le 20 décembre 2008

Recommandation de Monsieur Raphaël Bolze

Le comité directeur du programme « Décryphon », crée en partenariat entre le CNRS, l'AFM et IBM, a eu très vite à s'assurer la collaboration de personnel qualifié pour mettre en place une grille de serveurs universitaires. Raphaël Bolze a été alors recruté comme ingénieur pendant quelques mois avant d'obtenir une BDI CNRS pour préparer une thèse de doctorat.

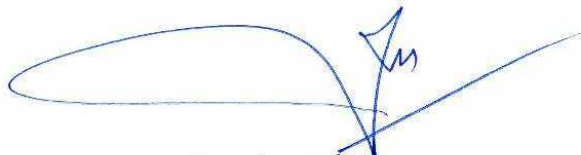
La contribution de R. Bolze a été déterminante pour le fonctionnement de « Décryphon »; Ainsi le projet a pu évoluer en fonction de l'avancement de ses travaux de thèse. D'autre part les projets de recherche retenus dans « Décryphon » lui ont permis de valider ses résultats.

S'appuyant sur l'interlogiciel DIET R. Bolze a mis au point le déploiement d'une grille de production pour la bioinformatique ce qui a permis aux 3 partenaires de « Décryphon », en se passant d'une société prestataire de service au demeurant peu satisfaisante, de consacrer tous les moyens disponibles aux laboratoires. Ses travaux ont aussi été d'un grand intérêt lorsqu'il s'est agit d'effectuer des calculs sur plusieurs milliers de protéines et qu'il a été fait appel à une grille d'internautes (WCG). La comparaison des résultats de « gridification » avec une étude théorique a permis au comité directeur de « Décryphon » de conforter son choix sur le type de grille utilisée pour le projet de recherche soutenu par le programme.

Au cours de ses travaux de thèse R. Bolze a su constamment faire l'aller et retour entre recherche et application en évitant que les laboratoires de biologie responsables des projets soutenus par « Décryphon » l'entraînent vers un travail essentiellement de service.

Raphaël Bolze a les qualités du jeune chercheur que j'aurais aimé recruter lorsque je présidait l'université Pierre et Marie Curie.

Je donne un avis très favorable à sa demande.



Jean Lemerle

Membre du comité directeur de « Décryphon »
Ancien directeur des relations avec l'enseignement supérieur au CNRS
Ancien président de l'université Pierre et Marie Curie

Evry, le 05 janvier 2009

Thierry Tourse
Direction Scientifique
Nos réf : TT /DS 001/01/09
Tel : 01 69 13 22 23
Fax : 01 69 13 22 22
ttoursel@afm.genethon.fr

Objet : recommandation pour la qualification de Mr. Raphaël Bolze

La collaboration avec Raphaël Bolze a débutée en 2005 avec le démarrage du programme Decryphon, programme de grille bioinformatique initié par l'Association Française contre les Myopathies, IBM et le CNRS.

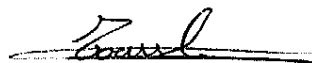
La problématique de ses travaux a consisté dans un premier temps à rechercher et mettre en œuvre une solution d'intergiciels pour la gestion de la grille Decryphon. Raphaël a assuré grâce à ses compétences un déploiement rapide des solutions retenues ce qui a permis le démarrage rapide du projet.

Raphaël a ensuite eu la lourde tâche de gridifier les différentes applications issues des projets scientifiques sélectionnés. Il a su démontrer une grande capacité d'adaptation et de compréhension des différents projets. Il a ainsi effectué des propositions pertinentes pour traiter efficacement l'ensemble des données biologiques de ces différents projets.

J'ai apprécié lors de notre collaboration la capacité d'adaptation de Raphaël, sa disponibilité et ses initiatives dans la mise en œuvre de ce programme. Il a, en outre, grandement contribué au succès de la grille Decryphon, qui est actuellement l'une des grilles de production pour la communauté française de bioinformatique.

Raphaël Bolze possède toutes les qualités requises pour obtenir un poste de chargé de recherche et c'est en toute confiance que je vous recommande sa candidature.

Avec mes meilleures salutations.



Thierry TOURSEL
Chef de Projet Recherche

Paris, December 29, 2008

To whom it may concern.

Dear Colleagues,

It is a pleasure for me to write a recommendation letter for Dr. Raphael Bolze to report on his competence and involvement on a project that I am leading as a principal investigator. Help Cure Muscular Dystrophy is a project supported by AFM/IBM/CNRS, which answered to the Decryphon call in 2006. It is a large scale docking project aimed to the detection of protein partners that integrates structural data with evolutionary signals used to predict potential docking interfaces. The project was first supposed to run on the French University Grid dedicated to the Decryphon Project where we could run a very first full cross-docking analysis on 6 pairs of proteins. Right from the beginning Raphael had to communicate and interact with computer scientists, physical-chemists and biologists to understand the problem, and translate our program into the world of parallelism (this amounted into an adaptation of the algorithm, an accurate definition of subunits and a scheduling of work units). He did it very well. This first analysis was crucial to us for the development of numerical criteria to select true protein partners. I wish to say here that the correct and efficient running of the parallel version of the program was crucial for us.

Soon we realized, with the help of Raphael, that the huge computational time demanded by our project (we wanted to realize a full cross-docking on a set of about 100 proteins, followed by a targeted cross-docking on a set of 4000 proteins) could not be handled by the University Grid and a potential solution could be to use the much more powerful World Community Grid offered by IBM in Seattle. The project was then considered by the WCG Committee and selected as one of the 6 international projects dedicated to health to run on WCG. In this occasion, Raphael helped to estimate the computer time of the project, he was our link with the WCG team, and defined the parallelisation of the workflow units.

The first phase of the project run successfully for 7 months on WCG, and in march 2009 we shall launch the second phase. For the second phase, Raphael adapted our new version of the docking program (integrating a prediction of the binding site to dock) for a parallel computation and delivered it to WCG.

I know that, even if Raphael will soon leave for a postdoc, he wishes to follow the project until its end. This will shows once more his alive interest for his work and for science.

Sincerely yours,


Alessandra Carbone

Professor, Department of Computer Science
Université Pierre et Marie Curie
and

Director of the laboratory "Génomique des Microorganismes", FRE 3214 CNRS-UPMC



December 29, 2008

1200 Fifth Avenue
Seattle, WA 98101

To Whom It May Concern:

I am very pleased to provide this letter for a very high recommendation for Dr. Raphael Bolze for the position of Assistant Researcher (CR2).

I am a Project Executive for the IBM Corporation. In this role I manage numerous Information Technology projects, one of which is called World Community Grid. World Community Grid is an IBM philanthropic initiative that provides qualified researchers with free access to large amounts of computer power.

I have known Raphaël since 2006, when the "Help Cure Muscular Dystrophy" (HCMD) project was accepted to run on World Community Grid. The HCMD project proposed using the power of World Community Grid for the detection of protein-protein interactions. HCMD accomplished this by setting the goal of screening a database containing thousands of proteins for functional sites involved in binding to other proteins targets.

Raphael was the Senior Architect for the HCMD project. In this role he provided leadership for all computer science aspects of the project, acted as the primary liaison person between HCMD scientific team and the World Community Grid technical team, and provided the expertise required to model the behavior of the scientific program. The value that Raphael provided became evident in the successful accomplishment of the HCMD project. In addition, Raphael conducted some very interesting work in the area of studying several volunteer computing platforms. This additional work was very interesting and valuable to World Community Grid and was subsequently published with collaboration from two people on the World Community Grid team.

In my opinion Dr. Raphaël Bolze has demonstrated not only superior organizational and communication skills, he has also shown the scientific expertise and creativity needed to become a leading researcher. He accomplished all of this while maintaining a friendly, collaborative, and professional demeanor.

In summary, I found Raphael a pleasure to work with and have no reservations in recommending Dr. Bolze for the position of Assistant Researcher.

Please contact me if you have any questions.

Sincerely,

A handwritten signature in black ink that reads "William Bovermann". The signature is fluid and cursive, with a long horizontal line extending from the end of the name.

William Bovermann, Project Executive
IBM Corporation
bbover@us.ibm.com
206-295-7949

Christophe Pouzat
Laboratoire de Physiologie Cérébrale
Université Paris-Descartes & CNRS
Responsable de l'équipe:
Statistiques computationnelles pour l'analyse des données neurophysiologiques
45, rue des Saints-Pères
75006 Paris
tel: 01 42 86 38 28
mail: christophe.pouzat@gmail.com

Le 3 Janvier 2009

Madame, Monsieur,

J'ai rencontré Raphaël Bolze alors qu'il travaillait comme responsable de la grille de calculs Decryphon. Mon projet de recherche venait d'être sélectionné par le comité scientifique du projet Decryphon (composé de membres de l'AFM, d'IBM et du CNRS). Raphaël était alors mon interlocuteur principal puisqu'il était chargé de (m'aider à) porter mon projet sur la grille. Ceci impliquait une compilation du logiciel utilisé: R (<http://www.r-project.org>) sur les machines de la grille puis la mise en oeuvre d'une interface web permettant de lancer les calculs "en batch", de visualiser leur progression et de télécharger les résultats. Les nombreuses interactions que j'ai eu avec Raphaël durant cette période m'ont permis d'apprécier pleinement les qualités de rigueur, de précision et de rapidité de son travail. Il a de plus fait preuve d'une réelle capacité d'écoute et s'est toujours montré à la fois patient et très pédagogue. Je ne peux donc que le recommander chaudement comme collaborateur et espérer que nos routes se croiseront à nouveau si sa candidature au poste de CR2 CNRS se trouve couronnée de succès.



Christophe Pouzat, CRI CNRS