Projet de demande de renouvellement de Plateforme Fédérative de Recherche

Titre du projet : MAGI - Plateforme de calcul et de traitement de données pour la recherche

Porteurs, rédacteurs du projet et contacts

M./M^{me}	Nom	Prénom	Laboratoire/Equipe /Composante	Téléphone/ email
M	CERIN	CHRISTOPHE	LIPN/AOC/GALILEE	4083/christophe.cerin@univ-paris13.fr
M	GRENECHE	NICOLAS	DSI	4035/nicolas.greneche@univ-paris13.fr

Relecteurs : Charles Desfrancois (LPL), Roland Chervet (Directeur Délégué SI USPC), Jérôme Le Rousseau (LAGA), Mohamed Boubekeur (LAGA), Philippe Djemia (LSPM), Emiliano Traversi (LIPN), Linda Chami (L2TI), Mustapha Lebbah (LIPN), Nadjib Achir (L2TI) et Jean-Luc Gaudiot (University of California, Irvine, professeur invité à Paris 13).

Type de demande : Renouvellement

Mots-clés : Infrastructure pour le calcul scientifique et la gestion de données massives.

Coordonnées de la plateforme fédérative (renouvellement).

Nom du responsable : Christophe Cérin

Localisation physique de la plateforme : DSI

Site Web de la plateforme : http://www.univ-paris13.fr/calcul/wiki/index.php?title=Accueil

Synthèse des demandes et des propositions

- Demande d'un budget annuel de 35k€ ;
- Propositions d'investir les domaines de la convergence des Systèmes et du Stewardship des données ;
- Au delà de la formation à l'utilisation de la plateforme, proposition de relance de la formation généraliste (écosystèmes du calcul scientifique, cloud, GPU, Big-data…) à destination en particulier des doctorants (périmètre USPC) ;
- Poursuite des journées scientifiques annuelles (Journées MAGI) ;
- Demande à toutes les SF de faire des appels à projets incluant explicitement un travail scientifique avec MAGI ; Les champs de la santé, des SHS etc. sont incités à faire de même ;

- Demande à la SF Math&Stic de co-financer du développement en rupture, au niveau système, pour la plateforme (afin d'accroître notre spécificité) ; mise en place d'un groupe de travail ;
- Demande de clarification des fonctions d'administrateur de la plateforme MAGI avec fiche de poste (voir Annexe);
- Propositions d'ouverture en direction des acteurs économiques (Kit MAGI), en particulier territoriaux, et de l'international.

Listes des Laboratoires/Équipes ou partenaires membres de la plateforme et noms des participants

La liste, la plus exhaustives possible, des projets qui ont bénéficié d'heures de calculs sur MAGI donne un aperçu de la diversité des champs scientifiques des utilisateurs. Les utilisateurs sont pour l'instant très majoritairement issus des Sciences mathématiques, de la physique / chimie et de l'informatique. En conséquence ce sont les laboratoires de Galilée qui donnent les utilisateurs de MAGI. Cependant nous verrons également, plus loin dans le texte, qu'il y a aussi des utilisateurs en dehors de Paris 13.

La liste des projets est : Énumération de marches dans le quart de plan et autres objets combinatoires, Simulations de la formation de microstructures et des interactions de dislocation dans les cristaux, Atomistic modelling of grain boundary migration and martensitic phase transition in pure titanium, Simulations numériques de problèmes d'écoulements de fluides et de plasmas, Testing of the performance of algorithms for mathematical optimization on large sets of instances, Classification taxonomique en métagénomique, Créqtion d'une bibliothèque de calcul de phénomènes streamer 2D-3D, Calcul de la configuration magnétique dans les nanostructures férromagnétique périodiques, Analyse de données temporelles multi-variées à large échelle, Calculs de Chimie Quantique moléculaire avec Gaussian, Calculs ab initio de propriétés structurales, optiques,magnétiques et élastiques de d'alliages métalliques, de semiconducteurs ou d'isolants, Grain boundary decohesion in presence of H with first principle DFT calculations, Molecular Dynamics simulations of solid-solid interface properties and migration, Réseaux de neurones & apprentissage à partir de modèles, Simulation des réseaux véhiculaires (VANETs & VDTNs) et des protocoles de routage véhiculaires, Numerical simulations of Spinor Bose–Einstein condensates, Algorithmes d'optimisation combinatoire, Calculs de propriétés et croissance de nanoparticules par dynamique moléculaire, Molecular Dynamics and Phase Field simulations of solid-solid interface migration, Calculs de Chimie Quantique moléculaire avec Gaussian, Distributed computation of high-dimensional heterogeneous agent economic models, Construction distribuée de Graphe d' Observation Symbolique, Experimental evaluations of data management/analytics algorithms in a distributed setting (using SPARK), Comparing and testing the performances of algorithms for combinatorial optimization problems, Mesoscale DD simulations of deformation of crystalline materials/ Molecular Dynamics simulations of solid-solid interface properties and migration, Simulations et calculs numériques sur les marches aléatoires sur les arbres aléatoires (programmes en C), Simulation intensive pour l'évaluation d'algorithme statistiques, Mesoscale simulations of dislocation microstructure near grain boundaries, Atomistic modelling of diffusion processes and mechanical properties, Testing and comparison of algorithm for mathematical optimization, Machine Learning - Social Network Analysis, Analyse de données séquençage haut débit et protéomique.

Nous donnons maintenant deux exemples de travaux scientifiques.

Philippe Djemia (LSPM) utilise le cluster MAGI pour des calculs de propriétés physiques (paramètre de maille, mécaniques, magnétiques et optiques) en matière condensée, soit

sur un nombre de cœurs limité mais avec beaucoup de mémoire ou bien pour des calculs distribués sur un grand nombre de cœurs (jusqu'à 1000 cpu soit 50% des ressources quand l'occupation le permet). C'est l'échelon indispensable avant d'accéder aux grandes infrastructures de calculs en France ou en Europe. Nous exploitons des programmes de calculs de dynamique moléculaire open source (LAMMPS) ou de calculs ab initio basés sur la théorie de fonctionnelle de la densité électronique, avec des licences recherche (VASP).

Lucas Letocart et Emiliano Traversi (LIPN), ont utilisé le cluster MAGI pour définir la première bibliothèque d'instances pour la programmation quadratique (QP), c'est-à-dire la famille des problèmes d'optimisation continus et d'entiers (mixtes) où la fonction objective et / ou les contraintes sont quadratiques. QP est une classe de problèmes très diverse, cette diversité se reflète dans la diversité des méthodes de solution QP, allant des approches entièrement combinatoires aux algorithmes entièrement continus, en passant par de nombreuses méthodes pour lesquelles les deux aspects sont fondamentaux. La bibliothèque finale comprend 453 instances, sélectionnées parmi un ensemble de plus de 9000 instances.

La liste, la plus exhaustive possible, des publications réalisées en partie grâce à MAGI est disponible en ligne à l'adresse suivante :

http://www.univ-paris13.fr/calcul/wiki/index.php?title=Misc:Publications

DOSSIER SCIENTIFIQUE

Rapport scientifique synthétique (4 pages max) en cas de demande de renouvellement.

La plateforme MAGI est l'infrastructure de calcul intensif de Paris 13, infrastructure mutualisée, qui offre depuis 2013 des moyens de calcul haute performance. Cet outil interdisciplinaire intéresse des chercheurs travaillant dans des domaines variés des sciences exactes et expérimentales parmi lesquelles la modélisation mathématique, la simulation numérique en Physique, Chimie et Biologie ou le développement d'outils de calcul ainsi que de traitement de données.

Une infrastructure de calcul, au sens de l'informatique, est un ensemble de processeurs (coeurs), de mémoire et de disques interconnectés par un réseau à haute performance. On parle aussi d'un cluster de calcul. Une infrastructure de calcul vient en appui à une démarche scientifique expérimentale. Elle permet souvent de valider des modèles mathématiques ou de trouver de nouveaux modèles en sciences des données par exemple. Une première forme de démarche scientifique avec un grand instrument numérique se résume principalement à modéliser la solution, à écrire un algorithme, à coder cet algorithme dans un langage de programmation, à construire des tests et un plan expérimental, à passer les tests sur l'infrastructure, à examiner la solution, à la critiquer pour éventuellement remettre en cause les hypothèses. Une deuxième forme de démarche scientifique s'illustre à travers la simulation de phénomènes physiques, chimiques ou biologique qui permet, par comparaison avec des résultats d'expériences réelles, de déduire des paramètres intrinsèques utiles pour comprendre et pouvoir transposer les résultats expérimentaux dans d'autres situations. Chaque discipline a ses algorithmes et ses méthodes qui lui sont propres. Certains noyaux de calcul sont utilisés dans plusieurs disciplines (mathématique, physique, informatique), d'autres sont spécifiques à la discipline.

Disposer d'une infrastructure sert à conforter et à développer les cultures de l'expérimentation à grande échelle. C'est un enjeu majeur pour celles et ceux qui ont un rapport fort aux machines dans leur démarche scientifique et pour produire de la Science de qualité. En effet les grands enjeux scientifiques contemporains considèrent des masses de données de plus importantes à fouiller, des maillages (discrétisation du phénomène

étudié) de plus en plus fins si bien qu'il est souvent nécessaire de montrer que le code développé passe à l'échelle c'est à dire que son comportement en temps décroît linéairement, à partir du moment ou l'on ajoute à l'infrastructure des ressources de calcul de manière linéaire. Ainsi on espère qu'en doublant les ressources, le temps d'exécution soit divisé par deux. Quand le chercheur s'intéresse à la qualité de la solution, avec une grande infrastructure il va pouvoir étudier cette qualité pour des petits et des grands jeux de données. Une infrastructure doit s'utiliser conformément à des procédures et des standards internationaux qui ne correspondent pas à une utilisation d'un ordinateur de bureau.

MAGI a été rénovée à partir de 2011 non seulement en terme de matériel (nœuds de calcul et de stockage, réseau) mais aussi en terme d'organisation de l'exploitation (soumission, réservation de travaux). Des comités sont en place. Un ingénieur de recherche a été recruté à Paris 13, en la personne de Nicolas Grenèche, pour installer et administrer la plateforme et peut intervenir ponctuellement sur du conseil et de la formation.

Les investissements effectués sur MAGI ont été co-financés jusqu'à présent par 1) USPC pour un montant de 400 k€ et 2) l'université Paris 13 pour un montant de 400k€. L'établissement, après un investissement initial de 200 k€ en 2011 par la Présidence de l'université a contribué à partir de co-financements issus du Conseil Scientifique (17k€ annuels) des chercheurs (90k€ au total), de la fédération Maths&Stic (ponctuel). Ces budgets ont permis de développer et d'assurer la jouvence du cluster MAGI. Les cofinancements apportés par les chercheurs de Paris 13 ont été les suivants :

Laboratoire	Contributeur	Date	Matériel	Montant
L2TI	Emmanuel Viennet	Nov 2014	Système de stockage Transtec	22335€
LSPM	Philippe Djemia	2014	Hardware	13k€
LSPM	Philippe Djemia	2014 et 2016	Logiciels de calcul scientifique à usage restreint : Vienna Ab initio software Package VASP (calculs ab initio)	4k€ 21k€
LIPN	Lucas Letocart	2015	Hardware	2.5k€
LIPN	Mustapha Lebbah	2013	Baie de calcul B500	16k€
Université	Christophe Cérin	2018	Serveur	10k€
LSPM	Oguz Umut Salman	2019	Hardware	4k€

Cette plateforme, localisée sur le campus de Villetaneuse dans les locaux de la DSI, comprend initialement deux parties : une pour le calcul sur mémoire distribuée l'autre pour le calcul en mémoire partagée. L'intégralité du cluster est reliée par un réseau rapide à faible latence de communication pour que les échanges entre nœuds soient les plus

performants possible. En 2018 un nouvel investissement a été réalisé sur un budget USPC de 32k€ pour une extension du stockage.

L'accès à la plateforme et le mode d'emploi s'effectue sur simple demande : http://www.univ-paris13.fr/calcul/wiki/index.php?title=Documentation:create_account. Son accès est limité aux chercheurs de Paris 13 et de SPC ou des chercheurs d'autres établissements ayant une collaboration en cours. La soumission de travaux est libre pour les chercheurs inscrits.

Les objectifs et réalisations sur la période 2014-2018

Le projet s'adresse à tous les laboratoires et au delà, pas seulement les laboratoires historiques. Il est important que l'ensemble des laboratoires de Paris 13 puissent être informés de ce qui a été fait et des projections. Nous visons à étendre à terme l'activité pour qu'elle concerne toutes les disciplines en capitalisant sur l'expérience acquise ces dernières années. Les infrastructures numériques pour la recherche sont faites pour tous et il faut aller au delà du « calculer plus ». Le conseil et la formation est importante comme l'ouverture nationale et internationale ainsi que la liaison avec les territoires.

Les chercheurs de Paris 13 ne disposaient pas, jusqu'en 2011, d'un outil informatique de calcul intensif. On trouvait une machine en jachère qui n'était pas mutualisée. Il y avait également une méconnaissance des meilleures pratiques quant à l'usage d'un tel instrument. Les principaux objectifs initiaux à atteindre ont donc été :

- Modifier le paysage de la recherche à Paris 13 tant au niveau méthodologique que thématique ; De nouvelles méthodologies scientifiques et approches de modélisation deviennent envisageables avec MAGI ;
- Toucher un public non spécialiste (experts dans les architectures, intergiciels, systèmes de la large échelle) dans l'usage des grandes infrastructures de calcul afin de réaliser leur montée en compétences ; Ne pas apparaître en concurrence avec certains équipements de calcul sur SPC mais complémentaire car nous touchons d'autres publics ; Faire en sorte que l'usage et les pratiques actuelles ne soient pas bouleversés par les nouvelles technologies mais intégrés ; Le cluster MAGI est en particulier complémentaire du cloud CUMULUS (http://cirrus.uspc.fr/cumulus/) de Paris 5.
- Accompagner le changement par la mise en place de formations à l'usage (pour mémo, l'ingénieur en charge de la plateforme assure des formations régulières à l'utilisation de la plate-forme[1] et une formation SPC pour les doctorants a été mise en place[2]
- Assurer de l'animation scientifique par l'organisation de journées scientifiques[3]

Comme démontré précédemment, les subventions de l'ex Conseil Scientifique (commission recherche) ont permis de financer du petit équipement, à co-financer les journées scientifiques, à co-financer des participations à des conférences spécialisées (NVIDIA submit, Super-Computing) et participer à sa jouvence.

Au dernier recensement il y a eu environ 100 publications en liaison avec l'usage de MAGI. Les personnes et la liste est disponible en ligne et l'effet structurant est donc important : http://www.univ-paris13.fr/calcul/wiki/index.php?title=Misc:Publications

[1] http://www.univ-paris13.fr/calcul/wiki/index.php?title=Accueil#Formation
[2] http://lipn.univ-paris13.fr/~cerin/ClusterGridCloud.html
[3] voir les éditions sur : https://lipn.univ-paris13.fr/~cerin/VILLETANEUSE2018.html

L'insertion dans le paysage – régional, national, international – de la PF.

MAGI se positionne comme une infrastructure de calcul de taille régionale, à l'image par exemple du cluster ROMEO de Reims. Il a pour vocation de préparer éventuellement le passage des codes sur les grandes infrastructures nationales. Les outils et les modes d'utilisation de MAGI sont désormais semblables à ce que l'on trouve dans les grands centres nationaux.

MAGI a donc une offre de services qui repose sur du matériel informatique (processeurs, réseaux, stockage) semblable à ce que l'on trouve dans une infrastructure régionale ou nationale, mais en nombre moindre. MAGI dispose à ce jour d'environ 2000 coeurs (organes) de calcul alors qu'une machine nationale en possède de l'ordre de 80000 pour la machine CEA/TGCC-GENCI. Il se distingue du matériel disponible dans un laboratoire qui, généralement, est de performance moindre que celui de MAGI. La distinction doit continuer à exister afin que les laboratoires puissent mener leurs politiques propres en matière de matériel informatique, dans le respect de la mutualisation. La mutualisation consiste en particulier à faire en sorte qu'un achat profite au plus grand nombre au lieu de finir sur l'étagère d'un chercheur. C'est un effort de tous les jours. Cela correspond par ailleurs aux volontés de nos tutelles, par exemple au CNRS.

Enfin, il convient de noter que MAGI accueille des expériences réalisées par des enseignants chercheurs en dehors des frontières nationales. Les utilisateurs extérieurs à Paris 13 sont en collaboration avec des chercheurs de Paris 13. Pour être un peu exhaustif, à ce jour, il y a 336 chercheurs, d'origines diverses sur MAGI, avec 221 comptes de Paris 13. Les principaux laboratoires utilisateurs de MAGI sont :

- **Université Paris 13 :** LIPN UMR 7030, LAGA UMR 7539, LSPM UPR 3407, LPL MR 7338, L2TI EA 3043, SMBH,
- **Extérieurs à Paris 13 :** Université Paris 7, IPGP, Université Paris 5 : MAP5, LMO, INSERM, INSA Lyon, Faculty of Mechanical Engineering (Prague), ENS Cachan, Université d'Orléans, Université de Yaoundé (Cameroun), Université de Xi'an Jiaotong (Chine), Université de Sétif (Algérie), Université de Picardie, Université de Nice, Université de Milan (Italie), Université de Grenoble, Supcom (Tunisie), Science Po, Polytechnique, Orange, ONERA, Normalesup, LPL, Institut Jacques Monod, Institute of Metal Research (Chine), INRIA, Fédération de Recherche en Mathématiques de Paris Centre, Fealinx, ESPCI, Centre de Recherches Interdisciplinaires, Bioaxial (Hôpital Cochin).

MAGI fait partie intégrante de CIRRUS qui est un projet ayant pour objet de proposer une infrastructure numérique mutualisée portée politiquement et financée par USPC dans le cadre de l'IDEX 2012. Cette plateforme est ouverte aux chercheurs, de tous domaines, depuis fin 2016. CIRRUS a été présenté au COmité du Calcul, numérique et INtensif du CNRS (https://lipn.univ-paris13.fr/~cerin/COCIN05072016.pdf). Le comité rassemble un représentant par institut, auquel est rajouté des représentants des grands équipements nationaux. Les coûts d'acquisition totaux (incluant la co-founding) sont de 1 900 000 € (900 000 € financé par USPC dans le cadre de l'IDEX). Trois ingénieurs informaticiens sont en poste pour son administration.

CIRRUS comprend environ 5000 coeurs (dont 2000 dans MAGI de Paris 13), 2,5 Peta-Octets (2.10^{15} octets) de stockage et une provision de 500 machines virtuelles.

Son originalité tient au fait que la plateforme CIRRUS est née de la mutualisation de trois plateformes déjà existantes dans les établissements. A ce jour, et à compter de 2016 ce

sont plus de 950 chercheurs qui utilisent CIRRUS qui ont produit plus de 250 publications identifiées utilisant au moins l'une des trois plateformes.

Concernant le positionnement avec USPC, MAGI s'est "intégré" dans la plateforme CIRRUS (http://cirrus.uspc.fr) composée d'une infrastructure de cloud (Paris René Descartes), d'un cluster de calcul à l'IPGP et de MAGI à Paris 13. L'idée initiale a été de se doter d'un centre de calcul et d'expérimentations informatiques avec pour ambition de renforcer le positionnement de la France dans le TOP500 (classement des supercalculateurs mondiaux). Un tel centre permettrait à terme de décupler les interactions à l'international, de crédibiliser les projets déposés au niveau national et européens par les chercheurs de SPC. CIRRUS permettrait aussi de développer des relations encore plus fortes avec les entreprises de nos territoires.

Outre le domaine du Calcul Haute Performance, cette infrastructure de production et d'expérimentation mutualisée (CIRRUS) offre la capacité de traiter des problématiques émergentes, comme le « Big-Data » et la visualisation avancée des données. Un objectif a été d'offrir une plateforme de virtualisation et de stockage, un peu à l'image de ce que l'on trouve dans les plateformes de Cloud Computing des GAFA (Google, Amazon, Facebook, Apple), via le cloud CUMULUS. Il s'agissait de mettre à disposition des chercheurs de tous les domaines, une grappe de serveurs puissants adaptés à leurs besoins en haute disponibilité et de fournir des capacités de stockage sécurisé de leurs données de recherche.

Le troisième objectif sur USPC a été de créer un réseau de partage de pratiques numériques dans les communautés scientifiques. Il est évident que SPC a la nécessité de se positionner par rapport au GAFA et leurs services. SPC n'a peut être pas vocation à se marier avec l'un ou l'autre mais de faire en sorte que chacun puisse se faire sa propre idée, et donc de choisir en conscience le type de services, en fonction de ce qu'il souhaite faire sur le plan scientifique. Il s'agit donc de se former aux nouvelles techniques et aux nouveaux usages. CIRRUS a pour rôle d'aider à se former, à apprendre en faisant. Les formations mises en place sont détaillées plus bas dans la section Formation.

CIRRUS ne doit pas seulement être un système de production pour les initiés de telle ou telle discipline. CIRRUS doit contribuer à une montée en compétence de tous, autour des écosystèmes du numérique d'aujourd'hui (IA, bigdata, machine learning, high performance computing, Internet of Object). Cela inclut aussi les publics SHS, langues, médecine etc. Une étude sur les processus métiers des chercheurs sur SPC a été conduite sur la fin 2015 afin de comprendre les pratiques actuelles et de mieux cibler les attentes et les outils utilisés. Le rapport d'enquête et les résultats de cette étude sont disponibles en ligne à l'adresse https://lipn.univ-paris13.fr/~cerin/Rapport_Enquete.pdf

Le mode de gouvernance de MAGI

La plateforme MAGI est organisée avec un comité technique, dont la dernière réunion s'est déroulée le 29 mars 2019. Ce comité est constitué de représentants désignés par les laboratoires. Il est co-animé par le responsable de la plate-forme et l'ingénieur. Il s'agit de discuter de technique, des modes d'accès etc. Sa fréquence est de l'ordre d'une réunion par an. Par exemple il a été décidé d'être souple et de faire simple dans la demande de création de comptes afin d'attirer le plus d'utilisateurs possibles. Les grands centres nationaux ont des procédures d'attribution des heures de calcul beaucoup plus formalisées que celles mises en place. L'ingénieur inspecte les usages et un mail est envoyé si une

consommation élevée de ressources est observée. Tous les problèmes se sont réglés à l'amiable, par une bonne compréhension de tous des pratiques de chacun.

La plateforme MAGI est également présente, via son responsable, dans le comité stratégique de la plateforme CIRRUS d'USPC et le comité technique. En effet, comme nous l'avons vu, des financements très importants de MAGI émanent d'USPC jusqu'à ce jour. Il a donc été important d'être actif au niveau d'USPC. Nous pouvons affirmer que Paris 13 a joué un rôle très important, avec un fort impact, dans l'offre de services numériques pour la recherche d'USPC.

Projet scientifique pour la période 2019-2023 (8 pages max)

Le projet scientifique et ses effets structurants seront explicités.

Contexte scientifique

Objet du projet et problématique scientifique générale

Le projet vise d'une part à conforter l'acquis des dernières années en matière de modélisation et de simulation numérique et d'autre part à s'ouvrir aux problématiques de données massives, en continuant à investir dans du matériel expérimental pour soutenir les pratiques scientifiques. Pour MAGI, il s'agit de trouver un équilibre entre équipements pour le calcul et équipements pour le stockage et le traitement des données, sachant que pour l'instant les efforts se sont portés sur le calcul (2000 coeurs de calcul). Or les infrastructures sont amenées à devenir «) généralistes ». Le document intitulé *Livre blanc sur les données au CNRS -- Etat des lieux et pratiques,* par la mission Calcul et Donnée (http://www.cocin.cnrs.fr/IMG/pdf/livre_blanc_donne_es_2018.pdf), donne un état des lieux précis des attentes, compétences des Instituts vis à vis de la donnée en janvier 2018. Il donne également un état des lieux des grands centres de calcul nationaux et des unités mixtes de services (UMS). Pour reprendre la terminologie du livre blanc, le projet MAGI, *d'un point de vue des infrastructures*, concerne les défis suivants (en reprenant le texte exact du rapport) :

- Convergence HPC et HDA : au-delà de la diversité des applications scientifiques exploitant les données, de nombreuses communautés scientifiques combinent et orchestrent aujourd'hui des applications d'analyse de pointe (HDA, pour High-end Data Analysis) et de calcul haute performance (HPC, pour High-Performance Computing) ;
- Stewardship des données : un autre défi critique est l'organisation des communautés (c.-à-d. stewardship) et des ressources autour des données, en phase avec leurs pratiques de recherche.

L'objet du projet est la consolidation de l'infrastructure MAGI actuelle au regard de ces deux défis. Il s'agit d'un projet d'équipement, avec l'ingénierie qui va autour, en s'appuyant sur l'expertise des enseignants-chercheurs de Paris 13 en matière d'exploitation des grandes infrastructures pour la recherche et l'innovation en rupture. En un mot, il s'agira d'être pro-actif vis à vis de la convergence des infrastructures tout en veillant à satisfaire le plus grand nombre possible d'utilisateurs et d'usages.

Périmètre des équipes/laboratoires et partenaires impliqués

Au titre d'un équipement numérique mutualisé, toutes les équipes et laboratoires de Paris 13 peuvent se sentir concernés par le projet. Par expérience nous avons des laboratoires "leaders", à savoir le LSPM, le LPL, le LAGA, le LIPN, le L2TI qui consomment à ce jour 90% des heures de calcul de MAGI. Cependant, à l'image de l'ouverture de CIRRUS vers toutes les communautés scientifiques d'USPC et du fait que Paris 13 est une université

pluri-disciplinaires, nous devons veiller à accueillir les nouveaux venus, indépendamment de la discipline. C'est le défi de Stewardship des données.

Nous pensons qu'il y a encore des communautés de Paris 13 qui pourraient rentrer dans l'initiative MAGI. Les communautés, par exemple en SHS, éthologie, gestion etc. ont peut être des besoins autres que le HPC pur, plutôt orientés vers l'accès à des environnements de travail de type notebook, voire à accéder à leurs logiciels habituels mais au travers d'un cloud. C'est tout du moins ce que nous avons constaté dans l'expérience CIRRUS, pour le public de l'imagerie du vivant par exemple. Il s'agit pour MAGI, sur la prochaine période, de continuer à promouvoir la plateforme et d'en faciliter l'accès et l'utilisation au plus grand nombre.

Contexte scientifique local et positionnement régional/national/international

Pour l'instant, MAGI fait partie de la fédération CIRRUS (http://cirrus.uspc.fr) et les forces ont été déployées exclusivement sur le périmètre de la COMUE. Cependant, certaines personnes du comité technique suivent et participent à des réunions techniques nationales concernant les grands instruments, par exemple les journées SUCCES ou celles de GRID'5000.

Appartenance à un réseau scientifique national ou international (préciser le ou lesquels)

Pour l'instant il n'y a pas eu de contractualisation avec un réseau scientifique national. Il est sans doute encore nécessaire que Paris 13 fasse ses preuves et communique sur ses savoirs faire et réalisations. Le problème de la notoriété, puisque MAGI est jeune, se pose.

Projet détaillé

Objectifs scientifiques précis

Le projet INRIA IPL-bigdata (https://project.inria.fr/hpcbigdata/) pour la période 2018-2022 pose les enjeux scientifiques de la convergence. Le HPC et Big Data ont évolué avec leurs propres infrastructures (superordinateurs et cloud respectivement), leurs applications (simulations scientifiques et analyses de données respectivement) et leurs outils logiciels (intergiciels MPI et OpenMP d'un coté et Map / Reduce ou Deep Learning de l'autre). Cependant, l'analyse Big Data nécessite de plus en plus de calculs (grâce à un apprentissage approfondi), tandis que la gestion des données devient une préoccupation majeure pour le calcul scientifique. D'autres projets de recherche à l'INRIA, dont Kerdata par exemple (https://team.inria.fr/kerdata/phd-position-hpc-bigdata-convergence/) permettent d'envisager des solutions techniques pour MAGI. MAGI est une infrastructure de production et, à ce titre, a l'impérieuse nécessité de répondre à des utilisateurs ayant des besoins immédiats. Dans le dossier d'information intitulé Big data en santé (https://www.inserm.fr/information-en-sante/dossiers-information/big-data-en-sante),
L'INSERM relève également que, concernant l'Internet des Objets, ceux-ci génèrent également de très nombreuses données transmissibles et partageables : appareils mesurant le nombre de pas, la fréquence cardiaque, la glycémie, la pression artérielle etc. Ces données sont le plus souvent stockées et gérées par des géants d'internet ou GAFAM : Google, Apple, Facebook, Amazon et Microsoft. De manière générale, nous pouvons affirmer que les GAFAM préparent des interfaces de programmation (API) capables, de manière unifiée, de capter tous les marchés, dont ceux de la santé. Chez Amazon par exemple, concernant les objets connectés, donc une partie de la santé, le lecteur peut consulter la page https://aws.amazon.com/fr/iot/ qui permettra de mesurer le très grand nombre de services déjà déployables.

Nous reprenons maintenant les deux objectifs généraux présentés au préalable et nous détaillons ce qu'il convient de faire à partir de technologies existantes, éprouvées et déjà en production :

- Convergence HPC et HDA : sur le plan technique, la convergence doit prendre la forme de l'intégration à la plateforme actuelle de processeurs capables à la fois de faire du calcul HPC mais aussi de l'apprentissage profond. Pour cela nous pouvons ajouter, à côté des processeurs généralistes actuels, des accélérateurs tels que nous les trouvons chez NVIDIA (coeurs CUDA et coeurs Tensor au sein d'un même système). Il convient de veiller que la gamme choisie corresponde à une gamme pour un cluster et non pas pour une machine de bureau (à cause de la distinction entre la politique de l'établissement et la politique des laboratoires). Ce type de technologie va par ailleurs introduire de l'hétérogénéité ce qui va compliquer les questions de métrologie de la plate-forme (remontée des informations relatives à la consommation des ressources afin de qualifier les usages), mais ce problème de métrologie sera traité dans le projet. Ensuite il conviendra d'étudier et de solutionner la question du rapprochement entre les données sur disques et les calculs dans les processeurs. Cela passe par des systèmes de fichiers parallèles et distribués performants permettant de coupler de manière efficace les dispositifs matériels. Il existe des solutions purement logicielles et Open source, qui seront privilégiées dans un premier temps, ou bien des solutions matérielles qui seront déployées ultérieurement en fonction des budgets. La priorité est donnée, en début de projet, à des investissements en coeurs de calcul et en coeurs destinés à de l'apprentissage profond. L'infrastructure mise en place devra également pouvoir ingérer les données de l'Internet des Objets (http://www.iot-planet.org/). En effet, plusieurs expériences sont en cours au L2TI qui a reçu du financement, notamment BQR pour ces expériences.
- Stewardship des données : pour continuer à intégrer tous les publics de la recherche de Paris 13, il convient d'avoir une offre matérielle et logicielle qui dépasse les problématiques de convergence HPC et HDA. Nous avons pu observer sur SPC qu'une solution de type cloud (CUMULUS localisé à Paris 5), permet au chercheur de déployer des machines virtuelles comme on peut le faire chez les GAFA. Le cloud étant un système centralisé, la gestion des traitements et l'analyse des données se trouve facilité. Un exemple phare d'utilisation du cloud CUMULUS est le projet SPC Imageries du vivant (http://idv.mi.parisdescartes.fr/) dont Paris 13 est partie prenante et moteur. L'exploitation de CUMULUS se fait via un référent technique qui hérite d'un virtual data-center dans lequel il peut gérer ses propres machines, ouvrir ses propres services réseau, sans le contrôle de l'administrateur de CUMULUS. La responsabilité est déléguée. Nous continuerons d'une part à promouvoir CUMULUS comme système numérique généraliste et nous mettrons en place également une politique cloud propre à Paris 13, axée sur les domaines scientifiques historiques, à savoir le HPC et les big-data. Nous envisagerons plus précisément comment coupler des ressources localisées à Paris 13 et des ressources hébergées dans des clouds publics. Il s'agit de pouvoir prendre des ressources en dehors de MAGI, voire d'échanger de la ressource avec partenaires à l'extérieur de MAGI. Bien évidemment, l'utilisateur de MAGI ne doit pas être impacté au niveau de sa méthodologie de travail. Une preuve de concept (POC) est actuellement menée sur le sujet, sur la base de la technologie OpenStack. Ce POC n'est pas encore ouvert à la production, par définition.

Pour conclure, plutôt que de convergence, peut être faudrait-il parler de faire entrer en résonance les éco-systèmes du HPC et des big-data à la fois sur le plan du matériel mais aussi sur le plan des méthodes d'exploitation et de travail des chercheurs. Pour Paris 13, cela devrait prendre la forme de projets scientifiques transversaux, pluri-partenaires

(Médecine et SHS compris) avec comme objectif de produire des résultats scientifiques en s'appuyant sur MAGI. Il n'y a actuellement pas d'appels à projets internes et d'offres claires, mettant MAGI au coeur. Il est suggéré à la commission recherche de l'université de se saisir de la question. Le comité technique de MAGI est d'ores et déjà prêt à prendre toute sa part dans la réflexion à mener.

De plus le comité de pilotage de MAGI suggère aux SF d'effectuer un fléchage explicite de leurs appels à propositions vers des expériences avec MAGI. Afin de continuer à effectuer une montée en compétences de tous vis à vis du numérique pour la recherche, il suggère également aux autres composantes de l'université de suivre cette proposition (un accompagnement technique peut se mettre en place spécifiquement).

Le comité de pilotage demande également à la SF Math&Stic d'étudier la possibilité d'un groupe de travail dédié aux aspects Système de l'infrastructure. Il s'agirait aussi de co-financer du développement en rupture, propre à MAGI. La stratégie vise à démontrer une spécificité, des savoirs faire, propre à MAGI afin de nous distinguer dans CIRRUS en particulier. Il serait avantageux de mener ce travail avec des partenaires internationaux.

Description de la plateforme

L'administration de la plateforme est assuré par Nicolas Grenèche. Outre différentes missions pour le compte de l'université, le comité de pilotage demande une clarification et sa fiche de poste relative à l'ensemble des services qu'il rend à la communauté. Une proposition de fiche de poste est donnée en annexe.

D'un point de vue technique, MAGI comporte 2000 coeurs : 50 nœuds équipés de 40 coeurs et de 64 Go de RAM chacun, 3 nœuds orientés mémoire partagée possédant chacun 512 Go de RAM équipés de respectivement: 16, 40 et 80 coeurs, un nœud GPU équipé de 4 cartes NVIDIA K40M. L'intégralité du cluster est connecté en Infiniband qui est un réseau rapide pour le calcul haute performance.

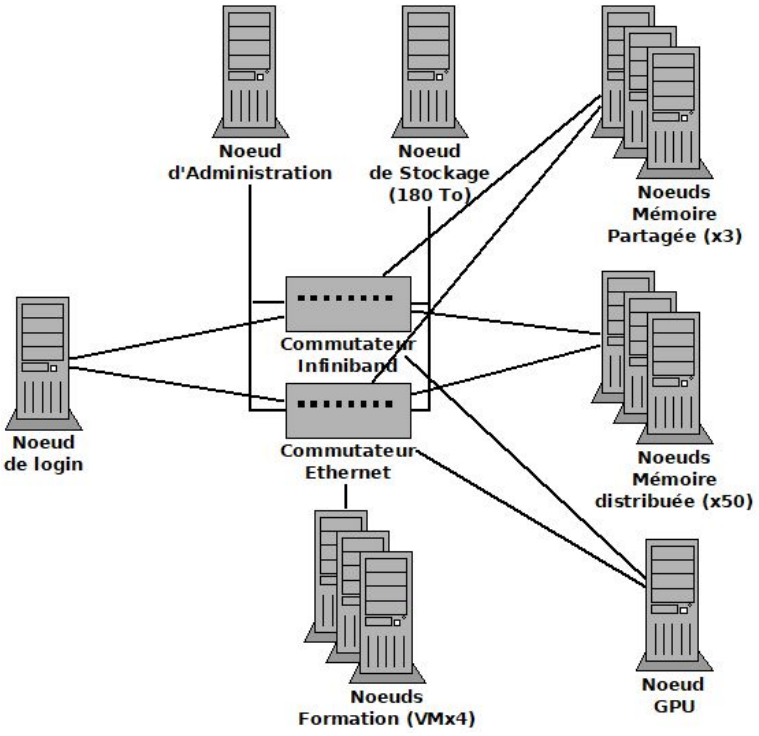


Figure 1 : Architecture du cluster MAGI

La Figure 1 détaille l'architecture du cluster MAGI avec le nœud de stockage, le nœud d'administration et le nœud de login qui constitue le point d'entrée. Enfin il y a des nœuds pour la formation à MAGI. Ils sont distincts des nœuds de calcul, qui servent à la production, et qui apparaissent sur la partie droite de la Figure.

Convention d'utilisation du matériel entre les partenaires

MAGI faisant partie de la fédération CIRRUS (Paris 13, 5, IPGP), il n'y a pas de convention d'utilisation du matériel entre les partenaires. La convention d'utilisation est implicite.

Pilotage, animation et organisation

La nomination du directeur/directrice de la structure se fait sur proposition du comité de pilotage suite à une élection à la majorité des voix. Si le besoin se fait sentir, il sera proposé un directeur/directrice adjoint(e). Dans ce cas, la parité devra être respectée.

Modalité de nomination d'un.une ingénieur (IR/IE) chargé.e du pilotage opérationnel de la plateforme

Voir fiche de poste en Annexe

Fonctionnement et composition de l'instance de pilotage

Le pilotage de la structure continuera de s'appuyer sur un comité pilotage (COPI). Ce dernier sera composé d'un représentant par laboratoire de Paris 13, désigné par le laboratoire, du directeur-trice de la PF, du/la ingénieur(es) de la PF et du Directeur délégué SI d'USPC. Les fonctions que devront assurer les membres du COPI, outre la présence au(x) réunion(s) pléinière(s), seront d'animer la réflexion en matière d'achat, de formation, de valorisation, de préparation des journées scientifiques, de relations avec les autres PF et SF de Paris 13, de l'insertion dans les éco-systèmes nationaux et internationaux, de rédaction de compte rendus annuels, de veille sur les Appels à projets etc. pour ne citer que quelques tâches. Le cas échéant, la PF MAGI fera appel à des experts scientifiques et techniques extérieurs.

Processus de prise de décision (ex budgétaire) et personnes impliquées, le rôle et la fonction des participants.

Jusqu'à présent la prise de décision s'est toujours faite sur le mode du consensus dans le COPI, y compris en cas d'utilisation « outrancière » des ressources de calcul. Nous proposons de continuer de cette manière.

Principaux éléments de la convention de fonctionnement entre les participants (financements, modalités d'accès, tarification, recours, etc.).

Il n'est pas prévu de convention de fonctionnement entre les participants. La question pourrait se poser dans le cas de l'accueil d'industriels (voir le point « Les évolutions à plus long terme »).

Modalités d'accueil de nouveaux membres, Modalités d'ouverture de la PF à des partenaires extérieurs

L'accès à MAGI est possible à des extérieurs de Paris 13 à partir du moment où il y a une collaboration avec Paris 13 ou SPC. Dans ce cas, un référent Paris 13 / SPC est demandé à la création du compte utilisateur.

Budget

Il n'est pas courant dans les plateformes académiques de taille régionale d'avoir un modèle économique qui fait payer les utilisateurs. Il est plutôt d'usage de s'appuyer essentiellement sur les aspects recherche, innovation et visibilité pour assurer les renouvellements. Les aspects de services aux entreprises (collaboration et prestation) servant aussi pour la partie innovation et visibilité. Le « coût horaire » qui pourrait être facturé à un industriel, s'il devait faire des calculs sur la plateforme sera à étudier au cas par cas en fonction des besoins[4]. En effet l'industriel pourrait demander un confinement de son travail ce qui nécessite là aussi des technologies dédiées. Dans notre cas, nous proposons de continuer selon les principes suivants :

1. La DSI héberge la machine (électricité et refroidissement) et en assure l'administration. L'université de Paris 13, via sa politique de PF, contribue significativement pour couvrir ces coûts de fonctionnement et de jouvence ;
2. Les enseignants chercheurs et leurs laboratoires sont invités à contribuer au fonctionnement en mentionnant un budget, systématiquement, dans leurs réponses aux appels à projets qu'ils déposent ;
3. Le comité de pilotage de MAGI répond aux appels à projets infrastructures qui peuvent émaner de l'Université de Paris (Idex), du CNRS, de la Région Ile-de-France etc. afin de coordonner une réponse qui satisfasse le plus grand nombre d'utilisateurs de MAGI. Il s'agit ici de déterminer les co-financements pour le renouvellement de PF MAGI.

La demande budgétaire à la CR - BRED Paris 13 est de 35k€ annuel.

- Ventilation du budget en fonctionnement, équipement, jouvence du matériel, maintenance du matériel. Pas de frais de personnel.

BUDGET PREVISIONNEL						
Catégorie	**2019**	**2020**	**2021**	**2022**	**2023**	**Total**
Fonctionnement						
Soutien aux évènements professionnels internationaux (participations à SuperComputing, NVIDIA submit etc.)	4k€	4k€	4k€	4k€	4k€	**20k€**
Soutien à la valorisation et au transfert (développement logiciel POC, études préalables etc.)	3k€	3k€	3k€	3k€	3k€	**15k€**
Soutien aux actions de rayonnement (visites de travail auprès de	3k€	3k€	3k€	3k€	3k€	**15k€**

[4] Techniquement, cela nécessite l'installation et la configuration d'outils de monitoring permettant de capter le montant des ressources utilisées, performant pour des services classiques comme pour des micros services.

centres/collègues liés aux infrastructures)						
Journée scientifique annuelle	2k€	2k€	2k€	2k€	2k€	**10k€**
Communication et vulgarisation	1k€	1k€	1k€	1k€	1k€	**5k€**
Sous-total Fonctionnement	**13k€**	**13k€**	**13k€**	**13k€**	**13k€**	**65k€**
Equipement						
Appels à projets « Gros équipement » (investissement)	75k€	75k€	75k€	75k€	75k€	**375k€**
Contributions des enseignants/chercheurs à la jouvence	15k€	15k€	15k€	15k€	15k€	**75k€**
Complément jouvence	20k€	20k€	20k€	20k€	20k€	**100k€**
Sous-total Equipement	**110k€**	**110k€**	**110k€**	**110k€**	**110k€**	**550k€**
Total	**123k€**	**123k€**	**123k€**	**123k€**	**123k€**	**615k€**

Le plan de financement sur cinq an est indiqué dans le tableau suivant. Au niveau des co-financements nous avons réparti sur 5 ans ce qui correspondrait à un « Gros équipement » (pour l'instant nous ne savons pas à quel moment déclencher la demande de financement).

PLAN DE FINANCEMENT PREVISIONNEL						
Catégorie	**2019**	**2020**	**2021**	**2022**	**2023**	**Total**
Ressources propres						
Contributions des Enseignants/chercheurs	15k€	15k€	15k€	15k€	15k€	**75k€**
Sous-total ressources propres	15k€	15k€	15k€	15k€	15k€	**75k€**
Co-financements à demander						
SESAME et structure similaire -- « Gros équipement »	70k€	70k€	70k€	70k€	70k€	**350k€**
Organismes et institutions nationaux -- « Petit	3k€	3k€	3k€	3k€	3k€	**15k€**

équipement »						
Sous-total cofinancement demandé	73k€	73k€	73k€	73k€	73k€	**365k€**
Aide demandée						
CR – BRED UP13	35k€	35k€	35k€	35k€	35k€	**175k€**
Total	**123k€**	**123k€**	**123k€**	**123k€**	**123k€**	**615k€**
Pourcentage d'implication CR - BRED UP 13	28,8%	28,8%	28,8%	28,8%	28,8%	**28,5%**

La gestion et mutualisation des moyens financiers.

Le centre de responsabilité financier sera localisé au BRED.

Respect des règles d'hygiène & sécurité, de la protection et sécurité des données, de l'archivage des données, de la valorisation des résultats obtenus.

La PF MAGI est localisée à la DSI, sur le Campus de Villetaneuse. Les accès à la salle machine sont protégés. Les utilisateurs de MAGI n'ont pas d'accès physiques à la salle informatique de la DSI. Concernant la protection des données personnelles, MAGI ne stocke pas d'information autre que les informations d'authentification (identifiant, mot de passe). MAGI est donc conforme à L'application et le respect du RGPD (Règlement Général de Protection des Données) en vigueur depuis mai 2018. A l'heure actuelle, MAGI ne fait pas de redondance de données à l'extérieur de la salle informatique de la DSI. **Aucun accès super-utilisateur n'est donné. Autrement dit, un enseignant chercheur ne peut pas installer un nouveau logiciel sur MAGI dans une partition du système.** Ce n'est pas une machine de bureau mais un système de production mutualisé.

La valeur ajoutée éventuelle pour la formation/interaction recherche/formation

Pour l'instant, la formation concerne l'utilisation de la plateforme MAGI par les enseignants/chercheurs. A ce jour il y a eu une dizaine de formation collective à l'utilisation de MAGI. De manière générale, l'activité de l'administrateur de MAGI est divisée comme suit et cela permet de mesurer les interactions entre la formation au sens large et l'aide en appui à la recherche :

- Administration de la machine : 50%
- Assistance et formation aux utilisateurs : 30%
- Assistance pour passer les codes sur centre nationaux : 5%
- Déploiement de nouveaux nœuds : 5%
- Écriture de codes ou de scripts pour les usages peu courants : 8%
- Veille : 2%

Du côté de la formation, il y a eu, pendant deux ans, une inscription d'un cours de 32h au catalogue des formations de SPC pour les doctorants, qui allait un peu plus loin que

l'utilisation de MAGI (https://lipn.univ-paris13.fr/~cerin/ClusterGridCloud.html). Il conviendrait de relancer cette initiative. Pour l'instant MAGI a été conçue pour la recherche exclusivement, comme système de production. Faire de la formation (Master par exemple), nécessite des modifications architecturales. Pour l'instant nous avons une demande du LEEC (C. Feron) pour de la formation en liaison avec leurs processus métiers.

Les évolutions et les développements éventuels à plus long terme.

La question d'un comité Stratégique / Scientifique de la PF MAGI peut se poser. Nous pouvons faire ici plusieurs observations. Un comité stratégique pour CIRRUS (fédération MAGI+S-CAPAG+CUMULUS) a été mis en place sur la COMUE. Ce comité n'a pas (encore) repris ses activités après la perte de l'IDEX. Pour Paris 13, faut-il s'appuyer exclusivement sur ce comité stratégique ou en monter un autre ? La question mérite une réflexion approfondie. Ensuite, nous estimons que la PF MAGI n'a pas encore fait le pleinven terme d'utilisateurs. Elle devrait par exemple s'ouvrir plus en direction des publics de SHS et de médecine. Nous proposons d'abord de capter et d'évangéliser d'autres publics potentiels pouvant tirer partie de MAGI avant de mettre en place un comité Stratégique. Pour l'instant, c'est le COPI qui joue le rôle de comité stratégique. Nous pouvons aussi faire remarquer que nous n'avons pas encore atteint une taille suffisante pour nous organiser comme une Unité Mixte de Service comme par exemple GRICAD (https://gricad.univ-grenoble-alpes.fr/a-propos) à Grenoble. Cela peut constituer un objectif à long terme, mais pas une organisation à implémenter immédiatement.

Ensuite plusieurs PF actuelles couvrent les champs disciplinaires de l'**Electrical Engineering** (au sens de l'**Institute of Electrical and Electronics Engineers** ou **IEEE** - http://www.ieee.org) et les focus de ces PFs vont des matériaux aux systèmes en passant par la physique et le composant. Voici quelques mots clefs associés : photonique, LiFi, quantique, IoT, Transmission Haut débit, Data, Clouds. Nous pourrions travailler ensemble, accompagné par des services de l'universités, aux questions de valorisation et de recherche partenariale avec les PFs. Nous proposons de considérer dès à présent ces questions afin de proposer un plan d'action d'ici à 12 mois. Parmi ces questions, il y aurait celle d'une structure d'accompagnement des acteurs du territoire vis-à-vis de leurs besoins Ressources Humaines, Recherche et Développement/Valorisation, et pour ce qui nous concerne en rapport avec les PFs.

En effet, pour l'instant, la PF MAGI est vue uniquement comme une plateforme de service (de calcul) pour des enseignants/chercheurs de Paris 13. Elle est donc centrée sur l'usager universitaire de Paris 13. Il existe d'autres organisations ou les usagers, mais aussi les citoyens, habitants, sont considérés comme des acteurs clés des processus de recherche et d'innovation. Cette approche s'éloigne de l'innovation centrée sur l'usager pour relever du projet porté par l'usager. Quels seraient les avantages de transformer le périmètre d'intervention actuel des PFs afin d'aller vers un schéma intégrant plus d'acteurs que seulement les enseignants chercheurs. Cela ne permettrait-il pas de mieux faire dialoguer Science et Société ? Nous pourrions par exemple imaginer offrir un "Kit MAGI" aux entrepreneurs de l'incubateur ou refacturer du temps d'utilisation de MAGI dans le cadre de conventions de partenariat Université-Entreprise(s).

Analyse SWOT dans la perspective du projet scientifique à 5 ans (une demi page)

Identification des forces

Les années de 2013 à 2018 ont clairement été l'occasion de structurer à la fois l'offre mais aussi un peu la demande en matière d'infrastructure numérique pour la recherche. De l'animation scientifique a pu se mettre en place soit sous la forme de formations à

l'utilisation, soit sous la forme de journées afin de croiser les expériences. Cela n'existait pas avant 2013 et les enseignants chercheurs étaient vraiment isolés, travaillant avec des méthodes qui n'étaient pas alignées sur des standards du domaine. Un indicateur du succès est le nombre de publications (100) qui ont pu bénéficier de l'apport de MAGI. Nous pouvons donc estimer qu'un palier a été franchi permettant au plus grand nombre (au sens des sciences dures) de s'approprier les outils. C'est une première étape. Les référents et compétences sont maintenant identifiés, conduisant à l'émergence d'un réseau d'experts internes à Paris 13.

Identification des faiblesses

Paris 13 n'est rentrée dans une démarche active vis à vis de l'utilisation des grandes infrastructures que depuis peu d'années, disons 8 ans. Les forces ont été concentrées sur le secteur du calcul haute performance (HPC) parce qu'il y avait des utilisateurs avec déjà une expérience sur les clusters. Ces utilisateurs avancés tirent leurs expériences de leur travaux de thèse, de leurs travaux collaboratifs avec d'autres chercheurs dans le monde, de voyages d'étude et de participation à des conférences dédiées aux machines ou faisant une part importante à l'expérimentation sur machines. Ce sont donc des sources extérieures de connaissance qui sont venues enrichir l'écosystème de Paris 13. Cet écosystème du HPC doit s'ouvrir à d'autres problématiques parce qu'une certaine convergence des infrastructures, tant sur le plan matériel que logiciel s'opère actuellement. Cette convergence ne s'opère pas avec le HPC au centre. Il conviendrait peut être d'avoir une vision beaucoup plus généraliste, par exemple dans nos choix technologiques, plutôt qu'une vision qui resterait sur le champs du HPC.

Identification des opportunités

MAGI ne couvre pas actuellement tous les besoins des disciplines de Paris 13. Comme grand instrument numérique pour la recherche nous pouvons penser que les secteurs de la médecine et de SHS pourraient être intéressés pour bénéficier de services innovants allant bien au delà d'un lieu de stockage de fichiers. Il faut certes avoir des données, mais ensuite pouvoir les analyser. Il est donc nécessaire de rapprocher les données des calculs ce qui est précisément un des objectifs techniques de MAGI. D'ores et déjà, le cloud CUMULUS, localisé à Paris Descartes, de la fédération CIRRUS d'USPC, permet d'embrayer avec des outils de traitement massif et surtout facilement utilisables par le non-expert. L'enjeu est ici de faire en sorte que l'accès soit aussi simple qu'une utilisation d'un service Google Docs par exemple. Dans ce cadre l'informatique et les sciences du numérique en général doivent plus que jamais prendre en compte les usages et le profil des utilisateurs.

Identification des menaces

Une première menace pour Paris 13 est de se voir décrochée par rapport à la maîtrise des grands instruments numériques pour la recherche. Sur la période récente nous avons vu le renforcement des GPU (processeurs graphiques) dans les systèmes de calcul, et celui des processeurs spécialisés pour traiter par exemple des problèmes d'Intelligence Artificielle. Nous parlons alors d'accélérateurs. Des fusions industrielles s'opèrent comme NVIDIA qui absorbe Mellanox pour 6.9 milliard de dollars (printemps 2019) ou encore Intel qui rachète Altera pour 16 milliard de dollars en 2015. Ces réorganisations conduisent à des innovations logicielles et matérielles dont on ne pourra plus se passer.

En parallèle, les GAFA, mais aussi IBM, ALIBABA etc. sont en train de proposer des écosystèmes de programmation unifiés afin de capter tous les marchés (IA, machine learning, high performance computing, Internet of Object, big-data...), ceux qui concernent directement les citoyens. Il ne s'agit plus ici de capter les marchés de la Big Science mais aussi les marchés pour les Big Populations. Ici le chercheur risque de se voir enfermé, capté, par des outils ou systèmes beaucoup trop vastes pour qu'il/elle en mesure seul les tenants et les aboutissants. Le chercheur devra se positionner vis à vis d'une attitude de formation tout au long de sa vie professionnelle et rechercher des compétences là où elles se trouvent (travail pluri-disciplinaire).

Une deuxième menace pour Paris 13, pour être encore plus précis dans l'argumentaire précédent, serait de se voir imposer des éléments méthodologiques la rendant dépendant fortement d'une seule manière de conduire et de faire de la recherche. Une menace de cette nature pourrait être celle qui consisterait à se voir imposer une manière de faire de la Science reproductible. Amazon par exemple propose dès à présent des moyens pour empaqueter son travail logiciel, le déposer sur une place de marché afin que d'autres puisse se l'approprier. Par exemple, les évaluateurs de nos travaux de recherche ont d'ores et déjà, avec Amazon, les moyens de rejouer les expériences que nous proposons dans nos articles, de changer les paramètres, de vérifier que nos modèles sont reproductibles.

Il est donc important à plus d'un titre d'avoir une attitude active en matière d'infrastructure pour la Recherche en créant les conditions favorables pour un dialogue entre les problématiques scientifiques et les problématiques d'ingénierie. Ce travail doit s'accomplir à la bonne échelle et Paris 13 ne peut pas le conduire seul. C'est pourquoi SPC par exemple s'est doté dès l'origine, pour les plateformes, d'un comité stratégique et d'un comité technique. En sommeil depuis 18 mois, ces comités doivent être impérativement relancés, avec l'implication particulière de Paris 13, avec des missions d'études, de prospectives, en se reposant sur les ingénieurs recrutés ou à venir. Cela vient en complément des missions premières des laboratoires. Sur le sujet de la recherche reproductible, l'équipe d'Inria Learning Lab propose la deuxième session du MOOC "Recherche reproductible : principes méthodologiques pour une science transparente". Christophe Pouzat (CNRS/MAP5, université Paris-Descartes), Arnaud Legrand (CNRS, équipe Inria Polaris, université de Grenoble) et Konrad Hinsen (CNRS, Centre de Biophysique Moléculaire à Orléans, Synchrotron SOLEIL) y posent les bases d'une recherche reproductible et expliquent l'utilisation de bons outils.

Une autre menace, très terre à terre mais réelle, serait que certaines entités ne jouent plus le jeu de la mutualisation, préférant revenir à une politique d'entité. La question peut s'exprimer simplement par "entité X, acceptez-vous qu'au dessus d'un seuil de 4k€ HT, qu'un équipement demandé par un de vos membres soit mutualisé dans MAGI ?". C'est une façon simple de mettre à contribution les laboratoires, sans passer par une taxe systématique, et de satisfaire le bonheur du plus grand nombre de personnes à Paris 13. La question mérite d'être posée à chaque entité.

CV court du ou des porteur(s) de projet (une demi page)

Christophe Cérin has been a professor of computer science at the University of Paris 13, France since 2005. In 2015, he was involved with an infrastructure project related to big data and high performance computing for e-sciences for USPC (Université Sorbonne Paris Cité). At Paris13, he chairs the board for the cluster computing facility available to all campus scientists and he chaired (2011-2016) the 'Expert Committee' in charge of recruiting and mentoring full time junior and senior professors in computer science. His current industrial experience includes serving as local chair for the Wolphin project (Alterway, Gandhi, Objectif Libre and Paris 6). His recent industrial experience was with the Wendelin and Resilience projects related to Cloud and Big-Data. He is also active with RSU (Rèves de

Scènes Urbaines - industrial demonstrator for sustainable cities), with Qarnot Computing and with Umanis. His research focuses on High Performance Computing, including Grid and Cloud Computing and he develops middleware, algorithms, tools and methods for distributed systems. Find more information on http://lipn.univ-paris13.fr/~cerin/

Transmission du dossier (format pdf) par le.la porteur /porteuse de projet après avis des conseils de laboratoires impliqués et de la commission recherche ou scientifique de ou des composante(s) impliquée(s) à commissionrecherche@univ-paris13.fr et vpcr@univ-paris13.fr.

ANNEXE : fiche de poste actuelle de Nicolas Grenèche