

Etat des lieux des structures de type méso-centres en France (février 2008)

Rapport rédigé sur la base d'une enquête auprès des responsables de méso-centres.

Rédactrices : Violaine Louvet (Institut Camille Jourdan, Lyon) et Françoise Berthoud (LPMMC, Grenoble)
Relecteur : Pierre Valiron (LAOG, Grenoble)

Ce rapport a été rédigé en préalable à l'organisation d'une journée nationale dédiée aux méso-centres, le 13 février 2008 à Paris. L'offre des méso-centres, tant au niveau des moyens de calcul, qu'au niveau des services (formation, expertise, ...) est en pleine expansion et il semble important dans ce contexte d'informer, voire d'envisager une structuration de cette offre en terme de visibilité, d'interactions et d'échanges de compétences.

Les rapports Héon-Sartorius¹, du Groupe d'Initiative « Calcul Scientifique » de l'Académie des Sciences² et du HCST³ ont dressé un état des lieux des moyens de calcul français, en mentionnant de façon très succincte l'existence de centres de taille intermédiaire.

Les rapports ⁴ et ⁵ présentent quelques réflexions de prospective européenne.

Le présent document se concentre sur les pôles de calcul de taille intermédiaire en France.

La structure pyramidale de l'offre des moyens de calcul à destination de nos laboratoires de recherche, telle qu'elle est synthétisée par la réflexion de prospective européenne, présente les futurs moyens pétaflopiques européens à son sommet, les grands centres nationaux (IDRIS, CINES, CEA), Grands Instruments (CC-IN2P3) à son tronc, et les moyens de calcul de proximité (moyens de laboratoire et méso-centres) à sa base. Dans le paysage français, ces derniers constituent une « nébuleuse » dont l'importance en termes de puissance de calcul et surtout en termes de services de proximité (formation, expertise, aide, mise en relation des acteurs etc.) est (était ?) mal connue. A ce schéma pyramidal viennent s'ajouter les grilles dont il faut souligner qu'elles concernent en l'état actuel des choses une faible portion des communautés utilisatrices (hormis l'IN2P3). En effet, les codes qui peuvent tirer parti des grilles sont spécifiques : ce sont essentiellement des codes séquentiels multiparamétrique ou des codes qui manipulent de grandes masses de données. L'adaptation des codes existants à cette architecture nécessite un travail de portabilité non négligeable (avec une exception notable pour la communauté de physique des particules).

Concernant les méso-centres eux-mêmes, il nous apparaît utile ici d'en donner ici un essai de définition :

Un méso-centre pourrait être caractérisé par :

- **Un ensemble de moyens humains, de ressources matérielles et logicielles à destination d'une ou plusieurs communautés scientifiques, issus de plusieurs entités (EPST, Universités, Industriels) en général d'une même région, doté de sources de financement propres, destiné à fournir un environnement scientifique et technique propice au calcul haute performance.**
- **C'est une structure pilotée par un comité scientifique (plus ou moins structuré) et, en principe, évaluée régulièrement.**

¹ « La politique française dans le domaine du calcul scientifique », mars 2005,

<http://www.education.gouv.fr/cid2204/la-politique-francaise-dans-le-domaine-du-calcul-scientifique.html>

² http://www.academie-sciences.fr/comites/groupe_ics.htm, décembre 2006

³ <http://www.hcst.fr/articles.php?lng=fr&pg=278>, janvier 2008

⁴ <http://www.hpcineuropetaskforce.eu/>

⁵ <http://www.hpcineuropetaskforce.eu/deliverables>

Ces moyens de méso-informatique ont des particularités propres et présentent des intérêts multiples pour les communautés utilisatrices. Les paragraphes suivants apportent différents éclairages sur ces caractéristiques.

Sur l'utilisation des moyens de calculs offerts par les mésocentres :

Certains types de besoins ne sont pas ou difficilement couverts par les centres nationaux. Il s'agit :

- Du développement, de la mise au point de codes, de tests, d'optimisation algorithmique, de débogage.
- De pré-études (jeux de données de tests).
- De certains types de calculs (pour des raisons liées soit à la nature des codes, soit à la durée de l'expérience, soit à l'inadaptation de la politique de gestion des files d'attente dans les centres nationaux, soit tout simplement à l'impossibilité d'accès aux moyens de calculs disponibles).
- De pré et post traitement des données / résultats de simulation numérique.
- De l'utilisation de codes commerciaux non disponibles dans les grands centres.

Sur les avantages de la proximité :

La création et le fonctionnement des méso-centres a un impact important sur :

- La structuration de l'offre de calcul haute performance au sein des universités.
- La qualité du support aux utilisateurs.
- Les liens avec les financeurs (les régions financent ces infrastructures dans près de 100% des cas)
- Les liens de proximité entre les acteurs dont le proche voisinage facilite les échanges :
 - Scientifiques,
 - Techniques,
 - Industriels (développement des relations avec le tissu industriel local).

Sur les aspects scientifiques :

Il est indéniable que l'existence des méso-centres a des conséquences au sein des communautés scientifiques impliquées. Ils constituent des lieux privilégiés d'échanges via des colloques ou journées scientifiques (par exemple au PSMN à Lyon, dans le cadre de CIMENT à Grenoble) et favorisent le développement de collaborations inter-disciplinaires (via le comité de pilotage par exemple).

Ils permettent dans certains cas des ouvertures vers des partenaires industriels (comme pour le CRIHAN à Rouen).

Enfin, la structuration des méso-centres autour d'un « *responsable scientifique* » (qui est souvent un comité de pilotage), ainsi que l'évaluation régulière de la structure elle-même (via le renouvellement des sources financières, ou par un conseil scientifique) garantissent une bonne utilisation des moyens pour des projets scientifiques sérieux. Le nombre de publications liées aux calculs réalisés dans l'ensemble de ces structures l'atteste, avec près de 800 publications par an.

De même, le nombre de laboratoires impliqués est un bon indicateur de la valeur scientifique intrinsèque des méso-centres. Environ 300 laboratoires utilisent ces moyens, soit en moyenne 14 laboratoires par méso-centre (avec des extrêmes de 2 à 30).

Sur les aspects compétences / expertises locales :

Un des aspects les plus positifs de l'existence des méso-centres concerne le développement de compétences et d'expertises locales, au service de la recherche, en contact direct avec ses acteurs.

Les équipes techniques partagent leurs savoir-faire sur des aspects assez variés autour de la gestion des méso-centres (appel d'offre, CCTP, administration des machines et logiciels, gestion des ressources).

L'identification de l'ensemble des méso-centres devrait permettre rapidement d'augmenter la quantité et la qualité de ces échanges.

Les différents acteurs (chercheurs, ingénieurs) ont de nombreuses occasions de communiquer sur les aspects davantage liés à l'informatique scientifique (algorithmes, méthodes numériques, bibliothèques scientifiques, programmation etc.).

Enfin, les expertises développées localement permettent d'offrir aux étudiants et chercheurs des formations adaptées. Ainsi la plupart des méso-centres proposent, par le biais des écoles doctorales, des formations au calcul haute performance ou à la parallélisation. L'ensemble de l'offre de formation des méso-centres (hors aspects grilles) est de l'ordre de 1500 jours*homme. Il ne fait aucun doute que ces efforts locaux constituent des leviers

importants pour permettre aux futurs utilisateurs des centres de calcul nationaux de porter leurs codes avec un meilleur rendement.

Sur les aspects souplesse d'utilisation et les liens avec les grands centres nationaux :

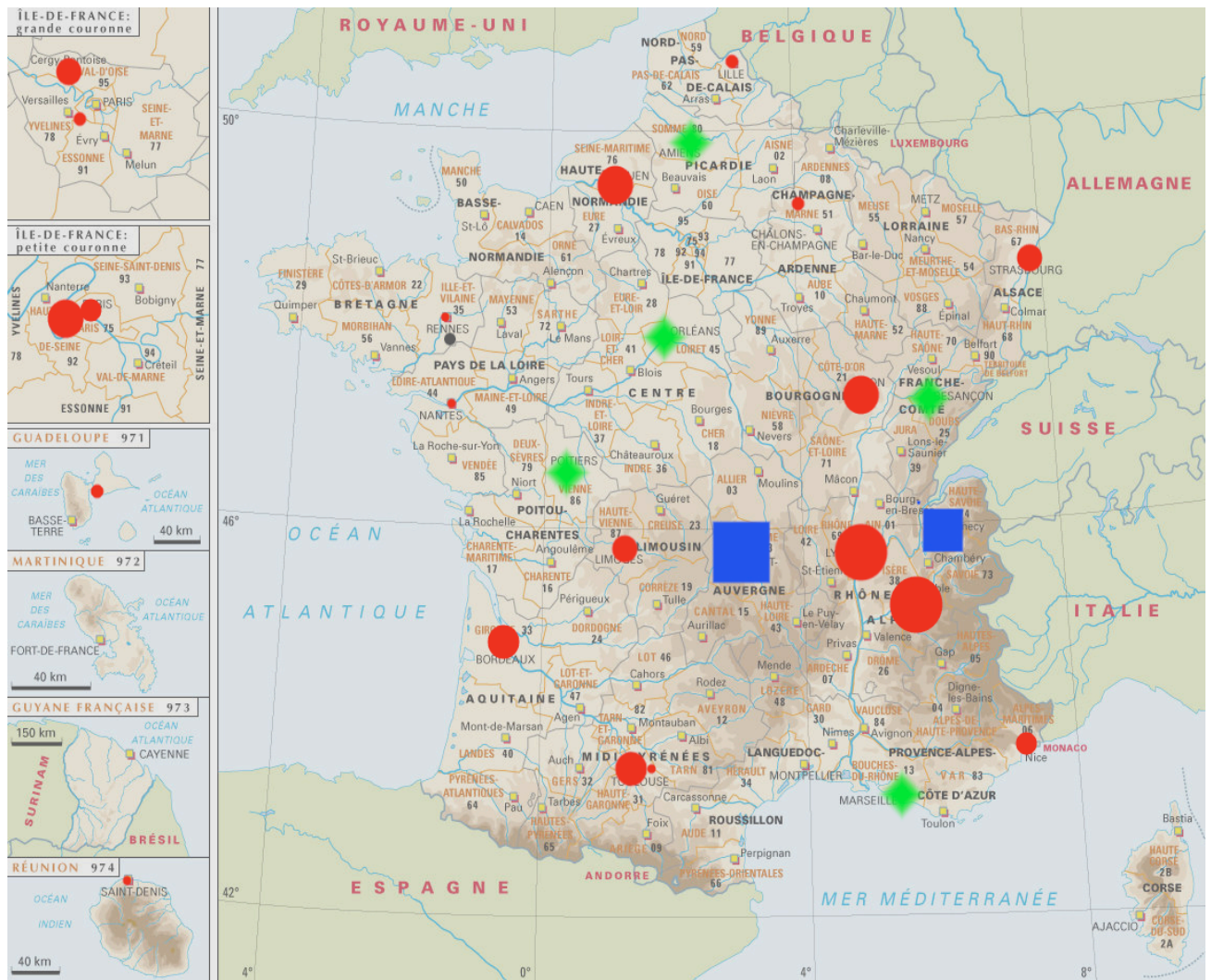
L'adéquation des choix architecturaux et de la configuration des gestionnaires de ressources aux besoins locaux permet l'exécution de jobs qui n'auraient pas pu tirer bénéfice des centres nationaux (en particulier du fait de la « lourdeur » de la procédure pour des besoins ponctuels et de certaines contraintes d'exploitation, notamment pour les jobs très longs).


Sur les aspects optimisation des moyens financiers :

Dans le contexte actuel de la multiplicité des ressources financières (plan méso du COMI en 1999, appel d'offre méso du ministère en 2003, financement dans le cadre des CPER -contrats projet état région- ou financement région, financement type "mi-lourds" du CNRS Région, PPF université, Europe, etc.), leur mutualisation pour l'acquisition et le fonctionnement des méso-centres optimise bon nombre de critères :

- moyens matériels, logiciels (jetons de licences),
- humains,
- hébergement (bâtiments, énergie),
- réseaux,
- « visibilité » par rapport aux fournisseurs qui permettent de bénéficier d'un effet volume sur les prix.

Les paragraphes suivants rendent compte de l'état des lieux de la méso-informatique en date de février 2008. Il faut noter que de nombreux projets sont en gestation, comme l'illustre la carte de France de ces structures. Les emplacements bleus désignent les méso-centres faisant partie de la grille EGEE. Le centre de calcul de l'IN2P3 n'est pas représenté sur la carte, car il est trop anachronique, en termes de moyens et en termes d'utilisation, par rapport aux autres structures. Les moyens de calcul dédiés à la grille de recherche GRID 5000 ne figurent pas sur cette carte.



 Mésocentres, échelle proportionnelle à la puissance crête

 Mésocentres participant à la grille EGEE, échelle proportionnelle à la puissance crête, différente de l'échelle précédente

 Mésocentres en projet, en cours de création

Le tableau ci-dessous regroupe l'ensemble des données en notre possession, fournies par les responsables des différents centres (classés par ordre alphabétique des noms de région). Un descriptif plus complet des mésocentres est disponible à partir du site web du groupe calcul dans la rubrique méso-centre.

Nom, localisation	Contact	web	Formations (HPC)	Personnel ETP	TFlops théo.
CECPV (Strasbourg, Alsace)	Éric Sonnendrücker	http://www-cecpv.u-strasbg.fr/	oui	2,1	1,2
M3PEC (Bordeaux, Aquitaine)	Samir MATAR	http://www.m3pec.u-bordeaux1.fr/	oui	2	1,5
AuverGrid (Clermont-Ferrand, Auvergne)	Vincent Breton	http://www.auvergrid.fr/	oui (grille EGEE)	4	5
CRI, Centre de Calcul (Dijon, Bourgogne)	Olivier Politano	https://haydn2005.u-bourgogne.fr/CRI-CCUB/	oui	2,5	1,7
GenOuest (Rennes, Bretagne)	Jacques Nicolas	http://genoweb.univ-rennes1.fr/Serveur-GPO/	oui	3+5CDD	0,4
ROMEO II (Reims, Champagne Ardennes)	Michaël Krajecki	http://www.romeo2.fr/	oui	1,5	0.61
CCUR (Saint Denis de la Réunion, DOM)	Delphine Ramalingom	http://www.univ-reunion.fr/ccur/	oui	1	0.22
C3I (Point à Pitre, DOM)	Pascal Poulet	http://www.univ-ag.fr/c3i/	oui	1	0,64
CRIHAN (Rouen, Haute Normandie)	Daniel Puechberty	http://www.crihan.fr/	oui	2	1,77
CEMAG (Ile de France)	Michel Pérault	Site web en intranet	non	0,25	1,740
Métacentre de Calcul et de Données (Paris, Ile de France)	Jean-Pierre Vilotte	http://www.ipgp.jussieu.fr/rech/scp/	oui	1,5	1
CRI (Orsay, Ile de France)	Laurence Puel	http://www.cri.u-psud.fr/machine/index.html http://www.u-cergy.fr/sir/	oui	1,5	0,65
SIR – UCP (Cergy Pontoise, Ile de France)	Yann COSTES		oui	2	1,275
CALI (Limoge, Limousin)	Jean Pierre Lainé	http://www.unilim.fr/sci/article106.html	oui	2	1,2
CALMIP (Toulouse, Midi-Pyrénées)	Dominique Astruc	http://www.calmip.cict.fr/	oui	2	1,5
GenoToul (Toulouse, Midi-Pyrénées)	Christine Gaspin	http://bioinfo.genopole-toulouse.prd.fr/main/	oui	3,4	0.403
Calcul Intensif USTL (Lille, Nord-Pas de Calais)	Didier Dangoisse	http://ustl1.univ-lille1.fr/projetUstl/universite/sces_communs/criservice/calcul_intensif/informations.htm	oui	1	0.6
CC IPL (Nantes, Pays de la Loire)	Gael Choblet	http://www.cnrs-imn.fr/CCIPL/	oui	1,5	0.147
CRIMSON (Nice, Provence Alpes Côte d'Azur)	Helene Politano	https://crimson.oca.eu/rubrique1.html	non	3	0,96
FLCHP (Lyon, Rhône Alpes)	Marc BUFFAT	http://www.flchp.univ-lyon1.fr/	non	3	3,137
CIMENT (Grenoble, Rhône Alpes)	Laurent Desbat	https://ciment.ujf-grenoble.fr/	oui	4	2,9
MUST (Savoie, Rhône Alpes)	Bernard Caron		Oui (grille EGEE)	3	2.9
CC IN2P3 (Lyon, Rhône Alpes)	Dominique Boutigny	http://cc.in2p3.fr/	Oui (grille EGEE)	68	35

Si on considère la puissance crête théorique totale de l'ensemble des méso-centres, excepté ceux appartenant à la grille EGEE, on arrive à environ 24 TFlops⁶ (La puissance crête pour l'ensemble des méso-centres participant à EGEE est de l'ordre de 40 TFlops). Par comparaison avec les centres nationaux, les calculateurs du CINES ont une puissance crête théorique de 1.85 TFlops, et ceux actuels de l'IDRIS, 6.55 TFlops (hors vectoriel).

La puissance brute déployée dans les structures de méso-informatique est donc loin d'être négligeable puisque pour l'instant elle est supérieure à celle proposée dans les deux principaux centres nationaux purement académiques. Par ailleurs, en raison de la complémentarité d'objectifs entre les moyens nationaux et les méso-centres, les solutions techniques généralement mises en œuvre dans les méso-centres sont légères et donc relativement peu coûteuses, plus proches des architectures de grilles que des architectures massivement parallèles. Cette légèreté a permis en général des cycles de jouvence plus rapides que pour les supercalculateurs

⁶ Calcul basé sur les 22 méso-centres répertoriés en février 2008 ; Il s'agit donc d'une valeur minimale ; les méso-centres MUST et AuverGrid sont intégrés ici à la hauteur du % d'utilisation hors EGEE.

nationaux, mais l'on peut espérer que la création de l'agence GENCI améliore également à l'avenir la jouvence de ces supercalculateurs. Des machines plus spécialisées sont présentes dans certains méso centres, dédiées à des travaux ou des logiciels spécifiques ou pour des jobs très longs, mais ne sont pas majoritaires en nombre de processeurs.

Il faut cependant noter que la nouvelle acquisition de l'IDRIS augmentera la puissance crête de ce centre à 207 TFlops (hors vectoriel) et celle du CINES à 50 TFlops.

Par ailleurs, il est intéressant de noter que dans la dernière liste du top 500 (en date de novembre 2007), les 10 premiers ont une puissance crête théorique entre 100 et 600 TFlops, et les 10 derniers entre 15 et 25 Tflops.

En conclusions, L'enquête réalisée dans le cadre des ces journées a permis de recenser plus de vingt méso-centres et de réaliser l'importance des moyens disponibles, tant en terme de puissance installée, que du point de vue de la formation et de la structuration des communautés scientifiques utilisatrices.