



www.cnrs.fr

Le CNRS et le calcul intensif

Avec l'aide de M. Daydé - Dr du COCIN



Le Comité d'Orientation pour le Calcul Intensif (COCIN)



Rôle et missions du COCIN

- Crée en Décembre 2010
- Réflexion collective sur les besoins, la structuration et les évolutions en calcul intensif au CNRS
- Prospective sur les besoins des différentes communautés, proposition de maintenance et de développement coordonné des moyens / ressources liées au calcul intensif, en particulier pour l'IDRIS.
- Dix personnalités scientifiques désignées par chacun des instituts du CNRS plus le Directeur de l'IDRIS.
- Le président et directeur désignés par le Président du CNRS

Composition / Fonctionnement



- Président : Ph. Baptiste (INS2I) parti au Ministère jusqu'en Avril 2013, Michel Bidoit nommé en Mai
- Dr : M. Daydé (INS2I)
- Membres :
 - S. Bosi (INSHS), D. Girou (IDRIS), Ph. Helluy (INSMI), L. Lellouch (INP), P.-E. Macchi (IN2P3), Th. Meinnel (INSB), C. Pouchan (INC), G. Bornette (INEE), D. Veynante (INSIS), JP Villette (INSU)
- Invités :
 - V. Breton (IDGC), O. Porte (DSI)
- Réunion tous les 1er mardi du mois

Quelques sujets abordés

- S'emparer des sujets du moment : IDRIS, Maison de la Simulation, coordination de l'informatique en appui à la recherche au CNRS avec la prise en charge du CCIS,
- Etat des lieux, cartographie, besoins, prospectives du calcul intensif au sein du CNRS (en cours)
- Réflexion autour de la pyramide du calcul en France : IDRIS, articulation Tiers 1 et Tiers 2 (méso-centres), ..
- Liens / complémentarités entre HPC / grilles / cloud
- Problématique du traitement des grands volumes de données
- Propositions pour une nouvelle stratégie du calcul au CNRS

Livre Blanc sur le Calcul Intensif au CNRS



- Lancé en Avril 2011 sur la base d'un questionnaire
- Contribution CNRS à la réflexion lancée par GENCI sur les besoins des communautés et sa stratégie pour 2012-2016 (renouvellement IDRIS en 2012)
- Panorama du calcul intensif dans les Instituts : pratiques, besoins, verrous, cartographie, aspects recherche, ...
- Certains des instituts (e.g. INC, INP) mènent régulièrement ce type de travail
- Objectif : travail finalisé fin 2012 avec pour objectif de mener des réactualisations régulières

http://www.cnrs.fr/ins2i/IMG/pdf/Livre_blan - dernière_version.pdf



COCIN@CD CNRS

Constats



- CNRS a en son sein tout l'écosystème du calcul intensif : PRACE, IDRIS, CC IN2P3, Institut des Grilles, recherche, interaction avec les méso-centres (Gdr Calcul, UMR), ... et toute la diversité des besoins / pratiques au CNRS (HPC, grille, cloud, ...) et caractère interdisciplinaire
- Favoriser les échanges entre communautés HPC / grilles / Cloud
- Synergie avec les acteurs du HPC (CEA, GENCI, INRIA, universités, ...)
- Promotion du calcul intensif à tous les niveaux centres nationaux et méso-centres
- Emergence rapide d'un fort besoin en matière de stockage / traitement sur les grands volumes de données

Calcul Intensif



- Plus possible de dissocier le calcul haute performance de l'analyse et valorisation des masses de données issues de :
 - simulations numériques, en climat, fluides turbulents (combustion, fusion, astrophysique) ... ;
 - grands instruments, i.e., LHC, ITER, LSST, LOFAR, plateformes génomiques ... ;
 - grands systèmes d'observation au sol, i.e., sismologie et géodésie (RESIF) ... et dans l'espace (Euclid, WFIRST, GAIA, imagerie et interférométrie) ...
- Compétitivité scientifique : adosser aux infrastructures et ressources informatiques un environnement d'expertise pluri et inter disciplinaire pour les valoriser et les exploiter (e.g. USA, Japon, Allemagne, UK)
- Calcul intensif pas uniquement guidé par les avancées technologiques

Conséquences



- Evolution technologique + consommation énergétique croissante + évolution requise des codes + compétences + support + impact sur l'organisation de la recherche => ***augmentation des coûts***
- Calcul intensif: pas uniquement problème de ressources mais un ***changement de paradigme*** dans la recherche scientifique
- Nouvelle ***approche holistique*** autour des défis scientifiques:
 - Inter/pluridisciplinarité et plus d'interactions entre informatique, mathématiques et autres disciplines,
 - Nouvelles méthodes et algorithmes: défis logiciels
 - Infrastructures de calcul et de données *en synergie* avec celles des grands instruments, plateformes expérimentales et des systèmes d'observation

Ecosystème riche au sein du CNRS



- Centres de calcul d'envergure nationale :
 - **CC IN2P3** Tier-1 WLCG : traitement et analyse des données issues des grands instruments en physique des hautes énergies (CERN), ouverture vers biologie, Astronomie & Astrophysique (données du LSST). Compétence / expertise reconnues en technologies Grilles et Cloud
 - **IDRIS** Tier-1 GENCI: HPC et analyse des masses de données produites par les simulations. Expertise, qualité de support aux projets pluridisciplinaires et formation, uniques et reconnues internationalement.
- Institut des Grilles et du Cloud (coordonne France Grilles), Maison de la Simulation, Mission pour l'interdisciplinarité (MASTODONS), lien avec les méso-centres ...
- Autres éléments d'envergure nationale du CNRS : Observatoires des Sciences de l'Univers (OSU), les grandes plateformes génomiques ...
- Besoin d'une d'une stratégie globale autour de ces initiatives

Vision du calcul intensif au CNRS



- Calcul intensif: un *grand instrument scientifique pluridisciplinaire*, catalyseur de nouvelles connaissances scientifiques
- Exploitation et valorisation scientifique dépendent de la capacité à:
 - l'insérer au sein des pôles de recherche et d'expertise associant recherche informatique, analyse numérique aux autres disciplines autour de grands enjeux scientifiques ;
 - évoluer en phase avec les pratiques de la recherche ;
 - Promouvoir l'appropriation du calcul intensif par les communautés encore trop peu utilisatrices des moyens nationaux et européens de calcul intensif, alors qu'elles pourraient en tirer profit pour aborder des sujets ambitieux nouveaux, au meilleur niveau mondial

Avec une stratégie globale



- Structurante ambitieuse, holistique et incitative.
- Meilleure coordination les initiatives locales
- *Qui a un coût !!*
- Un plan financier programmé dans le temps à l'échelle des enjeux scientifiques et des efforts de structuration nécessaires
- Spécificité pluri/interdisciplinaire du calcul intensif: modèle de financement différent des grands instruments disciplinaires
- Aussi bien en interne CNRS que au niveau national

Recommandations



- Dépasser modèle actuel de pyramide de ressources calcul et données
- Structurer les besoins en calcul, traitement de données, ... autour de grands défis scientifiques: en leur associant infrastructures de calcul et de données, des équipes de recherche inter/pluridisciplinaires.
 - Intégrés aux grands instruments, plateformes expérimentales et systèmes d'observation
- Reconnaissance des chercheurs et ingénieurs, à l'interface domaine disciplinaires / maths / informatique /

Financements côté TGIR CNRS



- Budget TGIR 2012 environ 140 M€
- Budget calcul de l'ordre de 20 M€ pour IDRIS, GENCI, CC IN2P3, RENATER, IDGC / France GRILLES
- *Est-ce suffisant eu égard aux enjeux scientifiques ?*
- *Situation assez similaire au niveau national*
- Personnel: ~ 30-40% du budget d'un centre
- Engagements nationaux et internationaux pour des centres d'envergure comme le CC IN2P3 ou l'IDRIS
- Aspects données / stockage pas encore suffisamment pris en compte (e.g. Tiers-1)
- Problématique du financement de l'évolution des infrastructures des centres

Conclusion



- Aller vers une stratégie holistique du calcul intensif au CNRS incitative, structurante et dimensionnante
 - Renforcer synergies entre les diverses initiatives au sein du CNRS mais plus largement au niveau national
 - Prendre en compte les enjeux pluri / interdisciplinaires des défis scientifiques du calcul intensif et du *data-intensive*
 - Développer des synergies avec grands instruments, plateformes expérimentales nationales, systèmes d'observation et infrastructures nationales de données (e.g. Observatoires des Sciences de l'Univers, ...).
 - Réfléchir à l'émergence de pôles agrégant infrastructures calcul et données, équipes de recherche autour de grands défis scientifiques, s'appuyant sur le contexte local