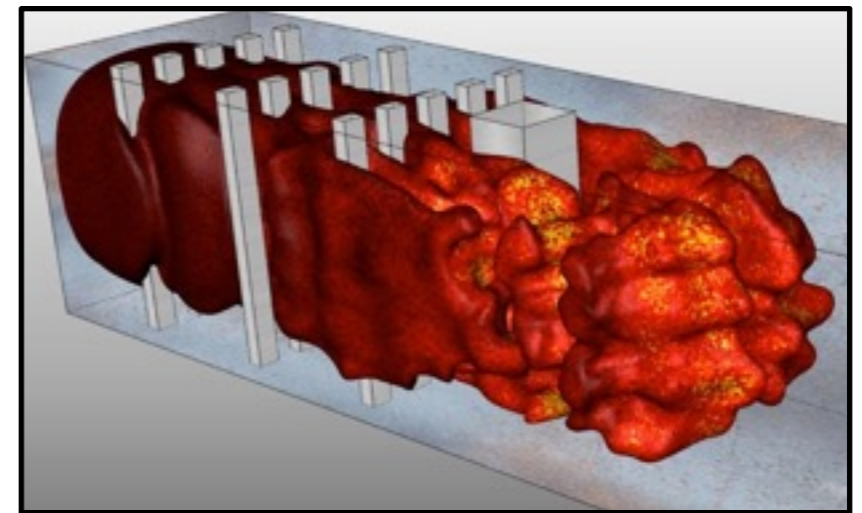
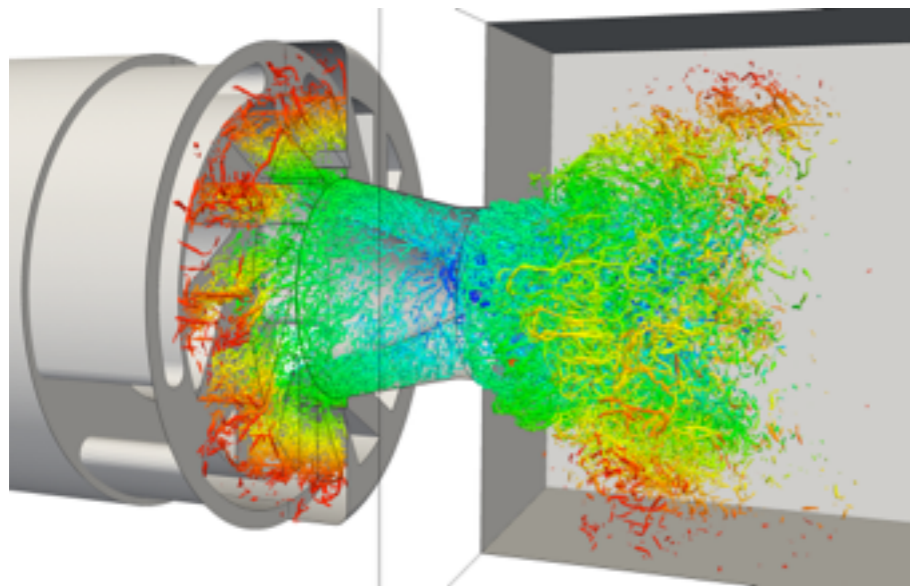


Mission Calcul - données (MiCaDo)

Denis Veynante

Président du comité directeur de la mission



Plan

➔ Contexte

- Importance croissante calcul / données
- Un paysage dispersé

➔ Volonté d'organisation au niveau de la Présidence

- Création du COCIN (2010)
- Mission Calcul - Données (2015)

➔ Le paysage, trois niveaux principaux

- *Centres nationaux (Tier1)*
- *Méso-centres (Tier2)*
- *Laboratoires (Tier3)*

➔ Agir principalement au niveau régional (*Tier2*)

- Actions structurantes et optimisation possibles
- Organisation pratique

➔ Insertion dans un paysage en évolution

- Infranum / Codornum, GENCI, PRACE, Renater

➔ Analyse scientifique des usages Tier1

- Constats
- Aller plus loin (Tier0, Tier2,...)

➔ Conclusions

Contexte

- ➔ **Développement modélisation / simulation**
- ➔ **Avancées scientifiques significatives**
 - *Mécanique des fluides / combustion*
 - *Climat (scénarii réchauffement climatique...)*
 - *Astrophysique*
 - *Physique*
 - *Chimie*
 - *...*
- ➔ **Explosion des besoins de puissance de calcul**
 - *Résolutions plus fines*
 - *Augmentation de la taille des problèmes*
 - *Simulations instationnaires sur des temps de plus en plus longs*
 - *Problèmes couplés*
 - *...*

Contexte

➔ Puissance de calcul

- *Machines massivement parallèles (dizaine de milliers de processeurs)*
- *Compatibilité modèles / codes / machines*

➔ Une large gamme de compétences

- *Mathématiques appliquées, méthodes numériques,...*
- *Algorithmique, informatique,...*
- *Modélisation physique,...*
- *Traitement de données,...*

➔ Des équipes multi-disciplinaires

- *Peu de codes partagés (production)*

➔ Des opportunités

- *Couplage de codes (exécution simultanée / échange de données)*
- *Problèmes multi-physiques*

Contexte

➔ **Explosion des volumes de données**

- *Issues des simulations*
- *Mais aussi :*
 - grands instruments (LHC, télescopes,...)
 - réseaux de capteurs
 - satellites d'observation
 - grandes bases de données
 - internet
 - ...

➔ **Non sans difficultés :**

- *Stockage (coûts, protection, conservation, sauvegardes,...)*
- *Manipulation / transferts*
- *Pré- et post- traitements*
- *Valorisation des données*
- *...*

Contexte : un paysage dispersé

➔ **Une “maturité” très différente selon les disciplines**

➔ **Une “discipline” transverse**

- *Par nature très inter-disciplinaire*
- *Pas de réelle “appropriation” par chaque communauté...*
 - qui compte un peu sur les autres pour défendre le calcul !
- *Pas de réelle communauté “calcul intensif” (≠ “computational science”)*
 - Difficultés à tenir un discours commun cohérent

➔ **Une multitude d’acteurs :**

- *Internationaux (PRACE,...), Nationaux (GENCI), Régionaux (méso-centres)*
- *Locaux (laboratoires)*

➔ **Une tendance à la dispersion des moyens :**

- *Machines locales “à soi” plutôt que mutualisation*
- *Foisonnement d’initiatives locales plus ou moins coordonnées*
- *Situation peu optimale en période de restrictions budgétaires*
- *Moyens humains ????*



**Volonté d’organisation / de coordination
de la Présidence du CNRS**

CNRS : COCIN

➔ **COCIN : Comité d'Orientation pour le Calcul INTensif**

- *Créé par A. Fuchs en décembre 2010*
- *Présidé par P. Baptiste puis M. Bidoit, animé par M. Daydé*
- *1 représentant par Institut + DSI + IDRIS + CC IN2P3*
- *invités : GENCI + quelques experts*

➔ **Mission de réflexion**

- *Besoins / organisation / évolution du calcul intensif au CNRS*
- *Au delà : stockage, informatique scientifique,...*

➔ **Une note de réflexion stratégique**

- *Présentée en CD le 20 février 2013*
- *Note du Président au COCIN du 8 avril 2013*

➔ **Recommandations**

- *Mettre l'accent sur la science (≠ course à la puissance)*
- *Renforcer les synergies entre des composantes parfois dispersées*
- *Intégrer les enjeux pluri- et inter- disciplinaires*
- *Mettre en place une "mission" pour le calcul intensif rattachée à la Direction Générale ("opérationnaliser le COCIN")*

Volonté d'organisation de la présidence du CNRS

➔ **Novembre 2015 : Mission Calcul - Données (MiCaDo)**

- *Définition et mise en oeuvre d'une politique globale et cohérente du CNRS (HPC, grilles, "cloud", infrastructures de données massives,...).*
- *Coordonner avec les instituts le pilotage des ressources dédiées IDRIS, CC-IN2P3 ; Maison de la Simulation...*
- *Aider les DSR à construire avec nos partenaires une stratégie conjointe en particulier avec méso-centres et centres de compétences mutualisés.*
- *Représenter le CNRS dans les instances du domaine (GENCI, Renater, Infranum,...).*
- *S'appuyer sur le COCIN, comité de pilotage de la mission.*

Même volonté au niveau de l'Etat (INFRANUM / CODORNUM)

Le paysage : trois niveaux principaux

➔ Moyens nationaux (*Tier1*)

- *Pour le CNRS :*
 - IDRIS (UPS, Orsay)
 - CC-IN2P3 (USR, Lyon)



➔ Moyens régionaux (*Tier2*) :

- *Méso-centres :*
 - Promotion du calcul intensif
 - Répondre à des besoins puissance de calcul / stockage
 - Etape préparatoire à l'utilisation des centres nationaux ou européens
- *Centres de compétences (maisons de la simulation...)*
 - Pas directement opérateurs de moyens de calcul
 - Souvent adossés à un méso-centre

➔ Moyens locaux des laboratoires (*Tier3*)

- *Nécessaires*
- *Tendance à la dispersion, pas toujours optimale*

Le paysage : trois niveaux principaux

➔ Moyens nationaux (*Tier1*)

- *Pour le CNRS :*
 - IDRIS (UPS, Orsay)
 - CC-IN2P3 (USR, Lyon)

- Périmètre, missions et moyens bien identifiés
- Ressources ?

➔ Moyens régionaux (*Tier2*) :

- *Méso-centres :*
 - Promotion du calcul intensif
 - Répondre à des besoins puissance de calcul / stockage
 - Etape préparatoire à l'utilisation des centres nationaux ou européens
- *Centres de compétences (maisons de la simulation...)*
 - Pas directement opérateurs de moyens de calcul
 - Souvent adossés à un méso-centre

Action structurante possible

CNRS

- Généralement pas ou peu actif à l'origine (Régions, CPER,...)
- Sollicité après coup :
 - structures opérationnelles (UMS, USR)
 - participation aux frais de fonctionnement
 - **personnel permanent pérenne**

Structuration et optimisation des moyens au niveau régional

➔ Identification d'un certain nombre de centres régionaux

- *Politique de sites du CNRS (Directoire, CD, DASTR, DSR, DR,...)*
- *Labellisation ministère (CODORNUM / INFRANUM)*

➔ Attribution de moyens financiers et humains

- *Sous condition de rationalisation, optimisation et mutualisation*
- *Plus généralement, soutien à des actions transverses "calcul-données"*

➔ Exemples

▸ *Aujourd'hui*

- CALMIP (mésocentre Toulouse, UMS INSU)
- Maison de la Simulation (Orsay, USR INS2I)
- FLMSN (Lyon, FR 3403, INSMI) - *évolutions ?*
- GRICAD (Grenoble, UMS INSMI), créée au 1/01/2016

▸ *Demands (plus ou moins avancées)*

- ISCD (Sorbonne Université), PSL, Nice, Orléans, Lille, Bordeaux,...



Gestion de structures transverses aux instituts du CNRS

Gestion des structures calcul / données

Création d'une structure légère

- *Affectation de moyens financiers et humains à la mission*
- *Gestion administrative déléguée aux instituts*

Commentaires :

- *Initialisation :*
 - Affectation à la mission des moyens actuels des centres nationaux, de la participation aux méso-centres et structures type MdS (UMS, USR)...
 - Cas particulier du CC-IN2P3 (*affectation partielle à la mission*)
- *Moyens supplémentaires pour initier / soutenir des actions incitatives*
 - ETPT : 5 en 2016 et 2017 (*niveau IE et IR*)
- *Avantages :*
 - La mission reste une structure légère
 - Dialogue avec les instituts
 - *attribution de moyens à des unités gérées par ceux-ci*
 - *incitation à des actions structurantes mutualisées (moyens supplémentaires)*

Rôle stratégique du COCIN

➔ Comité de pilotage de la mission

- *Evaluation des demandes*
- *Suivi des unités et de toutes les questions relevant de la mission*
- *Transparence de l'affectation des moyens*
(décisions finales relevant du comité directeur de la mission)

➔ Implication accrue de ses membres

- *Chargés de mission de la mission*
- *Renforcement des liens avec les instituts*
- *Réalisation d'un état des lieux détaillé*
 - Utilisation effective des moyens de calcul (Tier0, 1 et 2 par les laboratoires)
 - Moyens existants dans les laboratoires (Tier3)
 - Moyens humains affectés par le CNRS (calcul, données)

Evolution du paysage

➔ **INFRANUM / CODORNUM** (*réflexions interministérielles*)

- Structuration
 - *Labellisation de quatre datacentres nationaux*
 - *Processus de labellisation de datacentres régionaux (13 ?)*
 - *“Cloud” - liens avec European Open Science Cloud (EOSC)*
 - *IR FR-T2 ?*

➔ **GENCI** (*Tier0 / Tier1 français*)

- Evolution des missions (stockage)
- Financement des frais de fonctionnement des Tier1 ?
- Entrée du CC-IN2P3 dans le périmètre (partie non IN2P3) - Offre ?
- Coordination Tier1 / Tier2 ???

➔ **PRACE** (*Europe*)

- PRACE 1 terminé
- Démarrage PRACE 2 pour 3 ans (call 14 - mars 2017)

➔ **RENATER**

- Plan stratégique 2017 - 2020
 - *Réseaux de collecte*
 - *Services*
 - *“Cloud” / Datacenters*

Conclusions

➔ Ne plus dissocier HPC et analyse des données

- Données issues de calculs de plus en plus volumineuses
- Traitements des données massives de plus en plus lourds

➔ Rationaliser le déploiement des infrastructures

- Autour de défis scientifiques
- Coordination et rationalisation des investissements (*maîtrise des coûts*)
- Stratégies nationale et de sites
- Séparer “Béton” et “science”

➔ Etat des lieux - prospectives

- *Connaissance du paysage*
 - *Enquête sur les moyens de calcul utilisés par les laboratoires...*
 - *Utilisation scientifique des moyens de calcul (PRACE, Tier1, Tier2,...)*
- *Comité du Besoin en Calcul (DGRI)*
 - *Identifier les besoins indépendamment des “outils” (Tier1, Tier2,...)*

➔ Quelques points complémentaires :

- *Attribution des moyens / mutualisation*
 - *Promouvoir et soutenir la mise en commun de codes, outils,...*
 - *Optimisation des moyens, notamment ETPT*
- *Valorisation des bases de données (issues de calculs ou expériences)*
 - *Mise à disposition (cf Centre de Données astronomiques de Strasbourg)*
 - *Optimisation*
 - *Notoriété*