

Infrastructures de calcul et de données : les méso-centres, la vision du CNRS



www.cnrs.fr

D. Veynante

**Président du Comité directeur de la mission
« Calcul Données » au CNRS**

M. Daydé

**Directeur du Comité d'Orientation pour le
Calcul Intensif au CNRS (COCIN)**

Big Data & HPC : des enjeux stratégiques



- **Modélisation et simulation** : 3^{ème} pilier de la science (après théorie et expérimentation)
- **Exploitation des données** (« Big Data ») maintenant considérée comme le 4^{ème} pilier
- **Au cœur des grandes avancées de la recherche scientifique** :
 - Génome humain, boson de Higgs, évolution du climat, risques naturels, pollution atmosphérique, environnement...
- **Nombreux autres défis scientifiques** :
 - Structure de l'univers, astrophysique, neuroscience, combustion, sismologie, climat, biologie et recherche médicale, matériaux,
- **Enjeu stratégique de compétitivité et d'attractivité internationale** : multiples champs disciplinaires ; importantes retombées socio-économiques



HPC : la course aux performances



**USA, EU, Chine, Japon et Russie
systèmes exascales vers 2020**

1. **Mflops** : 10^6 opérations flottantes / s
(Intel 8086/8087, Motorola 68000) fin
des années 80,
2. **Gflops** : 10^9 flops,
 - **Tflops** : 10^{12} flops,
 - **Pflops** : 10^{15} flops,
1. **Exaflops** : 10^{18} flops



Tianhe-2

Actuellement top500 :

1. **National Super Computer Center, Guangzhou, China** Tianhe-2 (MilkyWay-2) – 33,86 Pflops avec 3,120,000 cœurs, 17,8 MW, peak 54,9 Pflops, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P
2. **Titan, Cray XK7, DOE Oak Ridge** : 17,59 Pflops avec 560,640 cœurs, 8,209 MW, peak 27,112 Pflops, Opteron + NVIDIA
3. **Sequoia, BlueGene/Q, DOE, LLNL** : 17,173 Pflops avec 1,572,864 cœurs, 7,890 MW, peak 20,132 Pflops
4. **“K Computer” au Japon**, 10.51 avec 705,024 SPARC64 cœurs, 12,7 MW, peak 11,8 Pflops

1ère machine européenne : CSCS en Suisse 7^{ème}

1ères machines françaises : TOTAL (33^{ème}) – OCCIGEN (CINES) 44^{ème}.

Mais USA > 50 % des systèmes installés

Objectifs



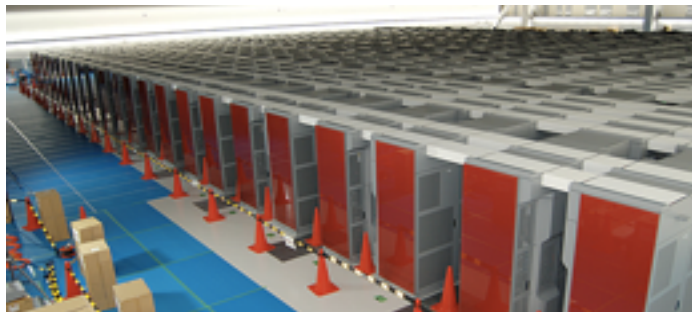
Objectifs aux USA :

- 100 PFlops 2016-2017
- 1 Exaflops en 2018-2020

Contraintes :

- Utiliser des technologies sur étagère ou au moins viables commercialement
- < 20 MW
- Calculateurs suffisamment généralistes

K Computer,
Riken, Kobe,
Japan



- Cons. actuelle : 130 W / chip
- Bande passante mémoire augmente moins vite due fréquence CPU
- #coeurs : 10^4 en 2000 => proche de 10^7 en 2012
- #sockets : 10^3 vers 2005 => 10^5 en 2012
- 100 millions de threads sur un système exaflops
- Avec la technologie actuelle : **475 MW** pour un LINPACK à l'exaflops mais uniquement 2% énergie pour le calcul :
 - 50% accès stockage données et déplacement des données
 - 50% : cache, mémoire virtuelle, ...

European infrastructure for HPC

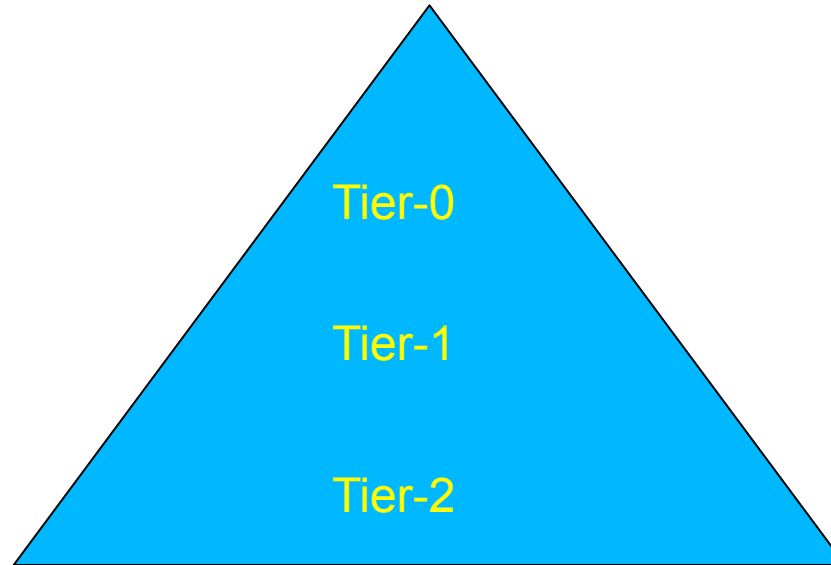


2013
Total capacity

2 Pflops
France /15
Pflops EU

1,6 Pflops

0,5 Pflops



>2017
Expected total capacity

50 Pflops France /200
Pflops EU

20 Pflops

5 Pflops

- **Tier-0** : European centres for Pflops, 6 computers in 4 countries (Germany, France with Curie at TGCC / CEA, Spain, Italy) + *Suisse (PRACE2)*
- **Tier-1** : 3 national centres (CINES, IDRIS CNRS, TGCC CEA) currently 1 Pflops cumulated, more than 2,4 Pflops in 2015
- **Tier-2** : Regional / University centres

GENCI is in charge of investments in the national computing centres and the french representative in PRACE

Besoins immenses autour des infrastructures de données



- **Communautés scientifiques avec des besoins / compétences bien établies :**
 - Sciences de l'univers (OSU, observatoires virtuels,)
 - Physique des hautes énergies (grille WLCG)
 - Biologie (RENABI, France Génomique, IFB)
 - ...
- **Reste des besoins immenses plus ou moins émergents**
Besoin structuration nationale et locale



Demandes IPSL pour CMIP6 2015-2018 (JL Dufresne) : 100 millions d'heures pendant 3 ans (12,000 cœurs par an), 25 Po stockage

Big Data: which specificities ?

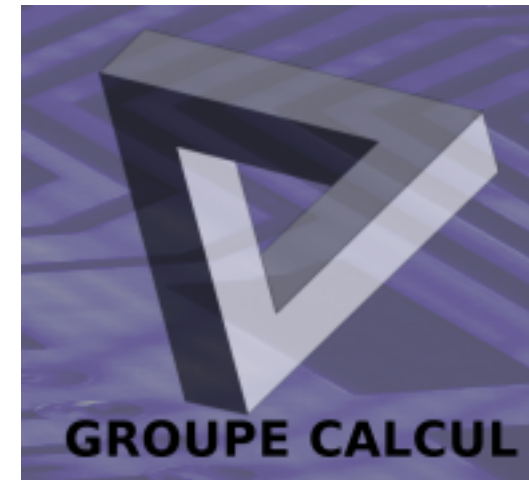


- **Volume:**
 - > PetaBytes which is a challenge for storage architectures
- **Variety:**
 - Diversity of contents, formats and data, structured, unstructured
- **Velocity:**
 - A challenge for networks
 - New models for processing streams of data
- **Veracity (quality / truthfulness)**
 - Data often souvent “write-once, read-many”(WORM), sometime costly to acquire
 - Hopefully, analysis can often be highly parallelized
- **Value of data**



Méso-centre

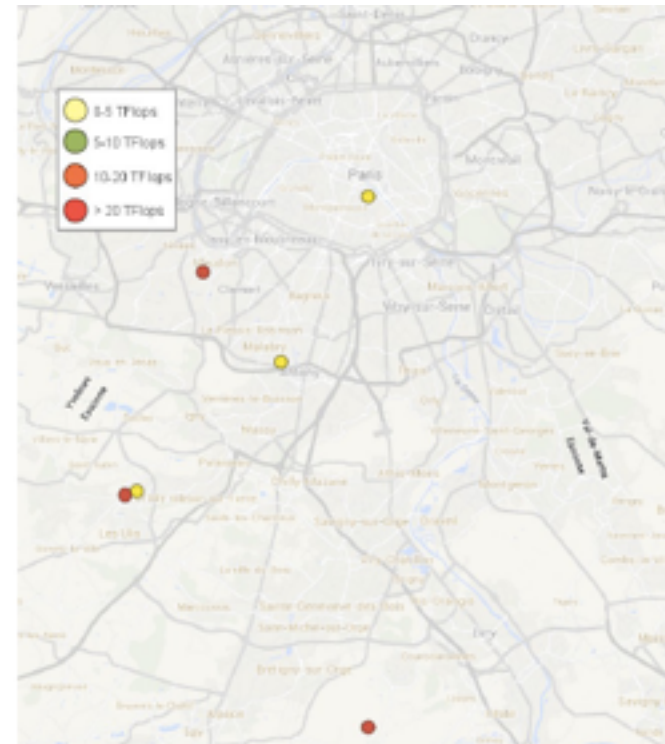
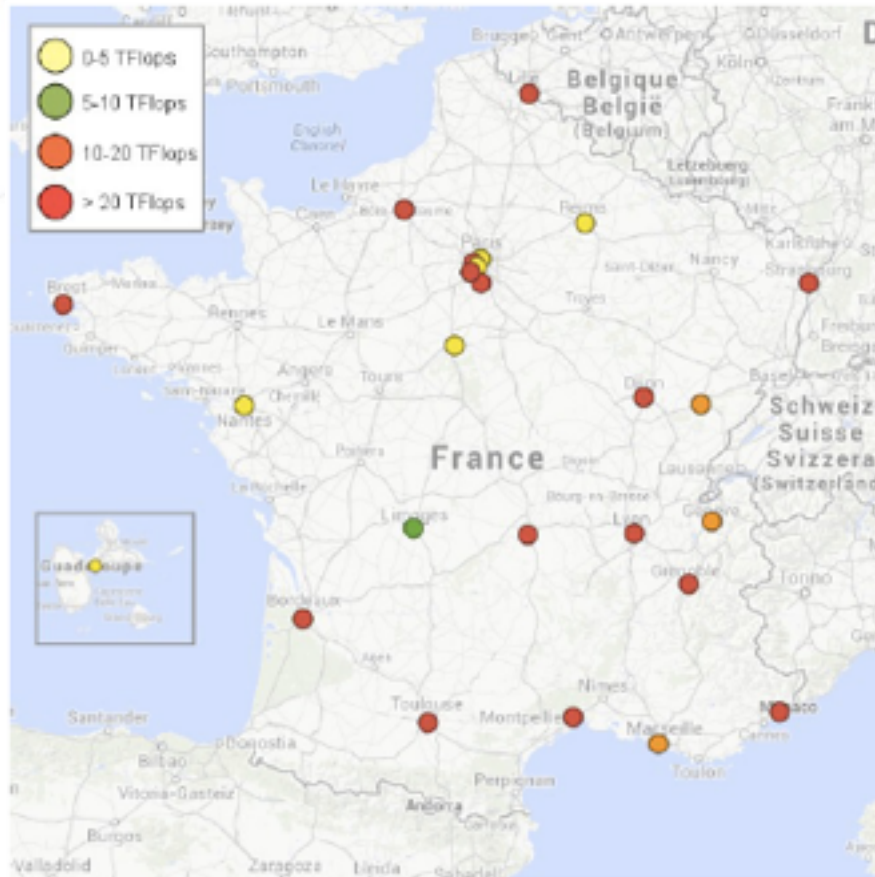
(définition du Groupe Calcul)
<http://calcul.math.cnrs.fr>



- **Ensemble de moyens humains, matériels et logiciels**
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.
- **Structure pilotée par un comité scientifique**
(plus ou moins structuré) et, en principe, évaluée régulièrement.

Les méso-centres en France

Rapport annuel 2013 sur les structures de type méso-centre en France, Groupe Calcul



34 méso-centres en 2013

En moyenne :

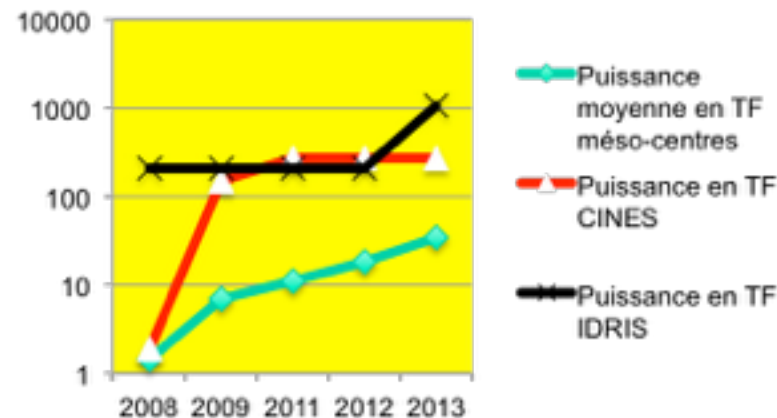
- Performance : 34,8 TéraFlops
- Stockage disque : 12 PétaOctets
- ETPT : 2,5

Evolution des Tiers-1 et Tiers-2

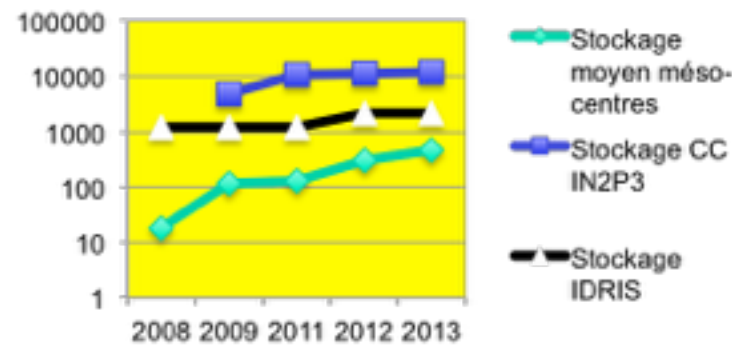
2009-2014 (données Groupe Calcul et EcoInfo)



- Performances de calcul multipliées par **30**,
- Volumes de stockage multipliés par **25**,
- Facture électrique augmentation d'un facteur allant de **1,25** au CC IN2P3 à **1,5** à l'IDRIS,
- **Avec un coût d'investissement stable** (qui n'a pas toujours permis de satisfaire la demande des utilisateurs) **et un volume de maintenance informatique et d'ETP décroissant.**
- **ETPT** entre $\frac{1}{2}$ et $\frac{3}{4}$ des dépenses de fonctionnement
- Estimation consommation électrique : cœur Intel environ 21W et 1 Teraoctets 18W



Evolution des performances en TFlops (TF) de deux Tiers-1 (CINES et IDRIS) et en moyenne sur l'ensemble des méso-centres



Evolution des volumes de stockage sur disques en Teraoctets (TB) sur le CC IN2P3 et IDRIS et en moyenne sur l'ensemble des méso-centres

Equip@meso

(Projet EQUIPEX 2010)



- Enveloppe globale : 10,5 millions € (2011-2019)
- 10 mésocentres partenaires – 5 mésocentres adhérents
- Coordonné par GENCI
- **Objectifs :**
 - Développer équipements et interactions entre centres régionaux de calcul
 - Soutenir localement l'Initiative HPC-PME
lancée fin 2010 par Bpifrance, GENCI et Inria pour faciliter et encourager l'accès des PME à la simulation numérique et au calcul intensif
- **Résultats :**
 - Doublement de la puissance de calcul
800 Tflops supplémentaires sur des configurations très diverses
 - Formation et animation scientifique pilotée par la Maison de la Simulation et en collaboration avec le Groupe Calcul du CNRS
Point fort du projet - Plusieurs formations nationales et locales par an
 - Journée thématique annuelle

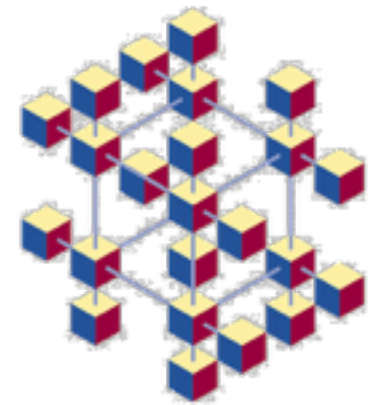
Partenaires Equip@meso



Exemple : méso-centre Toulouse



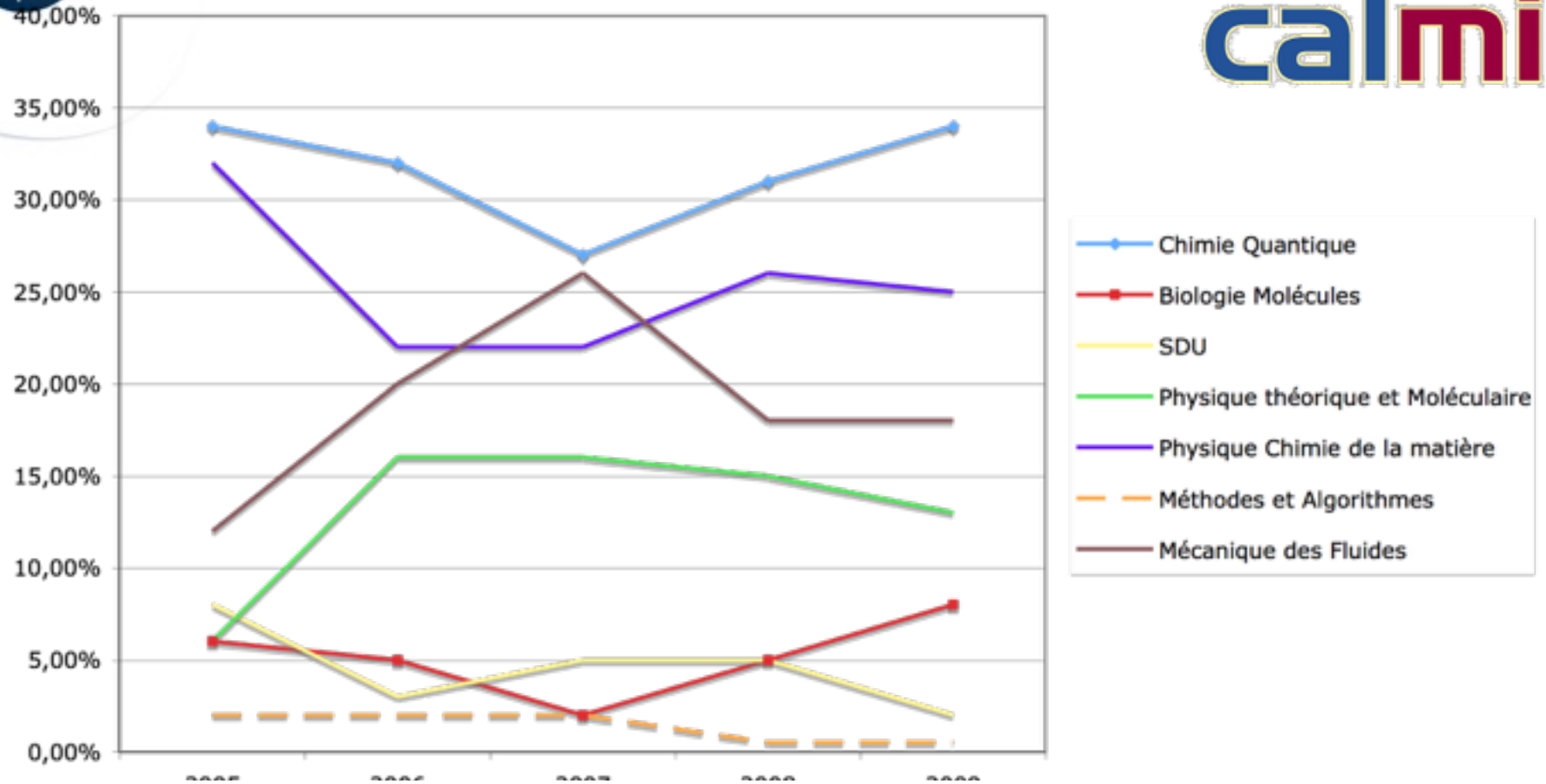
- Téraflopp Crête : 274 TFlops (Linpack 255 TF)
- 612 nœuds
- Processeurs Intel(r) IVYBRIDGE 2,8 Ghz 10-cores
- 12240 cœurs
- RAM par nœud 64 Go
- RAM totale 39 To
- Capacité disques initiale 1 Po extension (fin 2015) de 7 Po pour répondre à la convergence des besoins calcul intensif et traitement massif de données



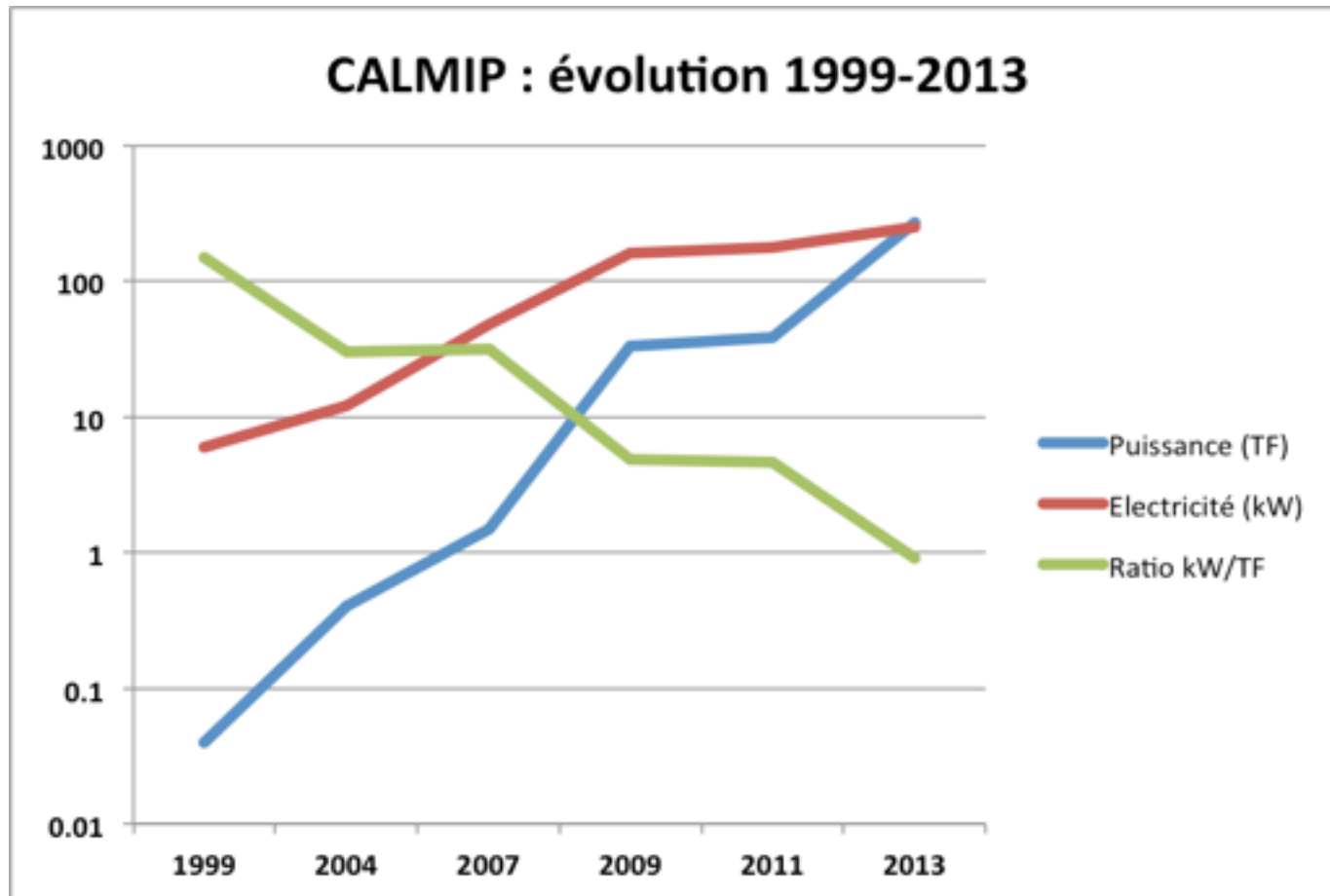
calmip



Evolution de la part des thématiques scientifiques au cours de la période 2005-2009



Evolution de la puissance et de la consommation électrique de CALMIP entre 1999 et 2013

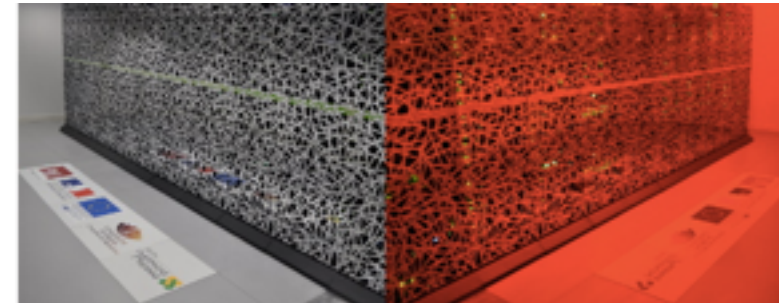


Exemple 2 : Romeo Reims



Cluster de calcul Romeo

- Installation : novembre 2013
- Puissance : 254,9 Tflops
- Consommation : 81,41 kW
- **130 noeuds de calcul hybrides Bullx:**
 - Deux accélérateurs NVIDIA TESLA K20X
 - Deux processeurs Intel Ivy Bridge @2,6 GHz, soit 16 coeurs de calcul par nœud
 - 32 Go de DDR3
 - Disque dur local
- **Réseau d'interconnexion**
 - Infiniband QDR à 40 Gbits par seconde
- **Données**
 - 57 To pour les données utilisateurs
 - 88 To de scratch global `/scratch_p` via un système de fichier Lustre
 - 100 To d'espace de sauvegarde



Calcul + données : coûts croissants à maîtriser



➤ Calcul intensif :

- Coordination au moins au sommet de la pyramide du calcul (Tier-0 / Tier-1 / Tier-2, EQUIP@MESO,...)
- Evolution technologique + consommation énergétique croissante + adaptation des codes + compétences + support

➤ Données :

- Explosion des besoins et des demandes non coordonnées (CPER) même si certaines communautés sont structurées (e.g. physique des hautes énergies, sciences de l'univers, bio, ...)
- Conforter compétences + support

➤ Impact sur l'organisation de la recherche



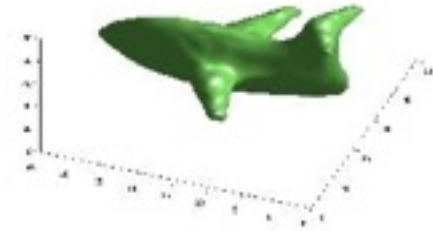
Rationaliser déploiement des infrastructures / coordonner les demandes
Stratégie nationale / de site autour de défis scientifiques et maîtrise des coûts !!!

Mission Calcul – Données (MiCaDo)



- **Nomination d'un chargé de mission auprès de la Présidence du CNRS (mars 2015) : Denis Veynante**
- **Contexte :**
 - Développement modélisation / simulation
 - Avancées scientifiques significatives
 - Explosion des besoins de puissance de calcul
 - Explosion des volumes de données
- **« Discipline » transverse**

Par nature très inter-disciplinaire, Pas de réelle "appropriation" par chaque communauté, "Maturité" très différente selon les disciplines
- **Création de la mission calcul-données (décembre 2015)**
 - **En cours de mise en place**



Mission Calcul – Données (MiCaDo)



- *Définition et mise en oeuvre d'une politique globale et cohérente du CNRS sur les objets à coût significatif (y compris en ETPT)*
- *Pilotage ou participation au pilotage des ressources dédiées*
 - IDRIS, CC-IN2P3
 - France-Grilles
 - Maison de la Simulation
 - ...
- *Cohérence de la politique du CNRS avec les “objets” nationaux (GENCI, Renater, Infranum,...) ou internationaux (PRACE,...)*
- *Aider les DSR à construire une stratégie conjointe avec nos partenaires*
- **Comité de Pilotage : COCIN**
(Comité d'Orientation pour le Calcul INTensif)



Le paysage national du calcul

➔ Moyens nationaux (Tiers-1)

• Pour le CNRS :

- IDRIS (UPS, Orsay)
- CC-IN2P3 (USR, Lyon)

- Périmètre, missions et moyens bien identifiés
- « Provenance » des ressources

➔ Moyens régionaux (Tiers-2)

Action structurante possible

• Méso-centres :

- Promotion du calcul intensif
- Réponse à des besoins puissance de calcul / stockage

• Centres de compétences (maison de la simulation,...)

- Pas directement opérateurs de moyens de calcul
- Souvent adossés à un méso-centre

CNRS pas toujours présent dès l'origine (CPER,...) mais sollicité après coup :

- Structures pérennes (USR, UMS)
- Frais de fonctionnement
- Personnel

Éléments clés de la stratégie du CNRS



- **Ne plus dissocier HPC de l'analyse et valorisation des masses de données** issues des simulations numériques (climat, fluides turbulents,...), grands instruments (, LHC, ITER, LSST, LOFAR, plateformes génomiques ...) et grands systèmes d'observation au sol (i.e., sismologie et géodésie : RESIF) et dans l'espace (Euclid, WFIRST, GAIA, imagerie et interférométrie)...
- *Calcul + data intensif pas uniquement problème de ressources mais un **changement de paradigme** dans la recherche scientifique :*
 - Plus d'inter/pluridisciplinarité (informatique, maths et autres disciplines),
 - Vision holistique des Infrastructures calcul / données / grands instruments / plateformes expérimentales / systèmes d'observation
- **Rationaliser le déploiement des infrastructures, coordonner les demandes**
- *En s'appuyant stratégie nationale / de site autour de défis scientifiques et maîtrise des coûts !!!*

Conclusion



- Calcul intensif / données : *grand instrument scientifique pluridisciplinaire*, catalyseur de nouvelles connaissances scientifiques
- Besoins calcul / stockage toujours **en forte croissance** (explosion des demandes CPER)
- **Facteur majeur** de la dérive des coûts informatiques : **foisonnement d'infrastructures de calcul et de données au niveau local** (i.e. Tiers-3) aggravé par **l'augmentation des demandes non-coordonnées + morcellement** et de **désorganisation des infrastructures de données**
- Le niveau régional dans les infrastructures calcul / données crucial
- Réflexions INFRANUM au MESR : aller vers 4 centres nationaux calcul / données et 13 datacenters régionaux

Stratégie du CNRS : coordination / rationalisation des investissements aux niveaux site / national autour de défis scientifiques avec l'ensemble des acteurs concernés

