



www.cnrs.fr

Ecosystème du HPC en France, éléments de la stratégie du CNRS

M. Daydé

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

Délégué Scientifique CNRS / INS2I

Rôle et missions du COCIN



- Créé 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 + Directeur de l'IDRIS + 4 ingénieurs experts
- Le président et directeur désignés par le Président du CNRS

Missions et production du COCIN



Créé en Décembre 2010 : réflexion collective sur les besoins, la structuration et les évolutions de l'écosystème du calcul intensif et des données scientifiques au CNRS

- **Composition : 1 représentant / institut + DU IDRIS + 4 experts**

Michel Bidoit (INS2I) : Président

D. Girou (IDRIS)

S. Lamarre (INEE)

JC. Michalski (INSB)

S. Bosi (INSHS)

F. Godefert (INSIS)

L. Lellouch (INP)

C. Pouchan (INC)

M. Daydé (INS2I) : Directeur

Ph. Helluy (INSMI)

P.-E. Macchi (IN2P3)

JP Vilotte (INSU)

Experts : D. Bascle (INC), F. Berthou (INP), M. Libes (INSU), V. Miele (INEE)

Invités : V. Breton (IDGC), O. Porte (DSI), JP Proux (GENCI)

- **Production :**

- *Livre Blanc sur le Calcul Intensif au CNRS fin 2012*

- *Propositions pour une nouvelle stratégie du calcul et des données au CNRS (Décembre 2013)*

- *Livre blanc sur l'Informatique en appui à la recherche (Mars 2014)*

- *Evolution des coûts des infrastructures pour le calcul intensif et le traitement des données à grande échelle (étude lancée 2014)*

Données scientifiques & HPC : des enjeux stratégiques



- Modélisation et simulation : 3^{ème} piliers de la science après la théorie et l'expérimentation
 - L'exploitation des données (« Big Data ») est maintenant considérée comme considérée comme le 4^{ème} plier de la science
- 
- Au cœur des grandes avancées de la recherche scientifique:
 - Génome humain, découverte potentielle du boson de Higgs, évolution du climat, risques naturels, pollution atmosphérique, environnement...
 - De 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



US, EU, Chine, Japon et Russie ont tous annoncé qu'ils auraient des systèmes exascales vers 2020

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



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 Pétaflop/s sur le Linpack benchmark avec 705,024 SPARC64 cœurs, 12,7 MW, peak 11,8 Pflops

1ère machine européenne CSCS en Suisse 6ème mais USA > 50% des systèmes installés

Tianhe-2

Objectifs



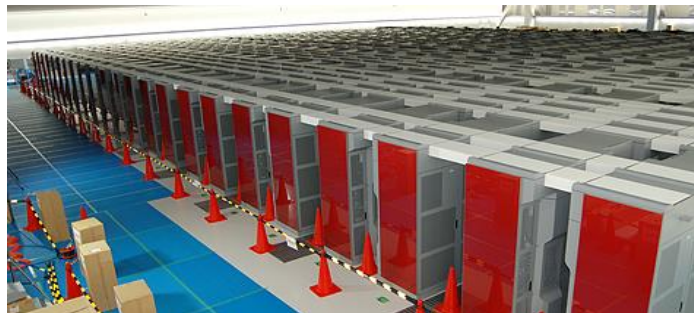
Objectifs aux USA :

- 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 que 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, ...

Compute & Data intensive



Large Synoptic Survey Telescope (LSST):
Chili, $30 \cdot 10^{12}$ octets d'images / jour (*télescope*
2.5m, Apache Point Observatory, New Mexico)



Large Hadron Collider (LHC): 60 To /
jour soit 15 Po / an

- 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 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

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

What do we mean by Big ?



From Mokrane Bouzeghoub (CNRS / INS2I)



Data inflation

Unit	Size	What it means
Bit (b)	1 or 0	Short for "binary digit", after the binary code (1 or 0) computers use to store and process data
Byte (B)	8 bits	Enough information to create an English letter or number in computer code. It is the basic unit of computing
Kilobyte (KB)	1,000, or 2^{10} , bytes	From "thousand" in Greek. One page of typed text is 2KB
Megabyte (MB)	1,000KB; 2^{20} bytes	From "large" in Greek. The complete works of Shakespeare total 5MB. A typical pop song is about 4MB
Gigabyte (GB)	1,000MB; 2^{30} bytes	From "giant" in Greek. A two-hour film can be compressed into 1-2GB
Terabyte (TB)	1,000GB; 2^{40} bytes	From "monster" in Greek. All the catalogued books in America's Library of Congress total 15TB
Petabyte (PB)	1,000TB; 2^{50} bytes	All letters delivered by America's postal service this year will amount to around 5PB. Google processes around 1PB every hour
Exabyte (EB)	1,000PB; 2^{60} bytes	Equivalent to 10 billion copies of <i>The Economist</i>
Zettabyte (ZB)	1,000EB; 2^{70} bytes	The total amount of information in existence this year is forecast to be around 1.2ZB
Yottabyte (YB)	1,000ZB; 2^{80} bytes	Currently too big to imagine

VLDB

Big Data

XLDB

Very Big Data

Massive Data

Data Deluge

The prefixes are set by an intergovernmental group, the International Bureau of Weights and Measures. Yotta and Zetta were added in 1991; terms for larger amounts have yet to be established.
Source: *The Economist*

European infrastructure for HPC

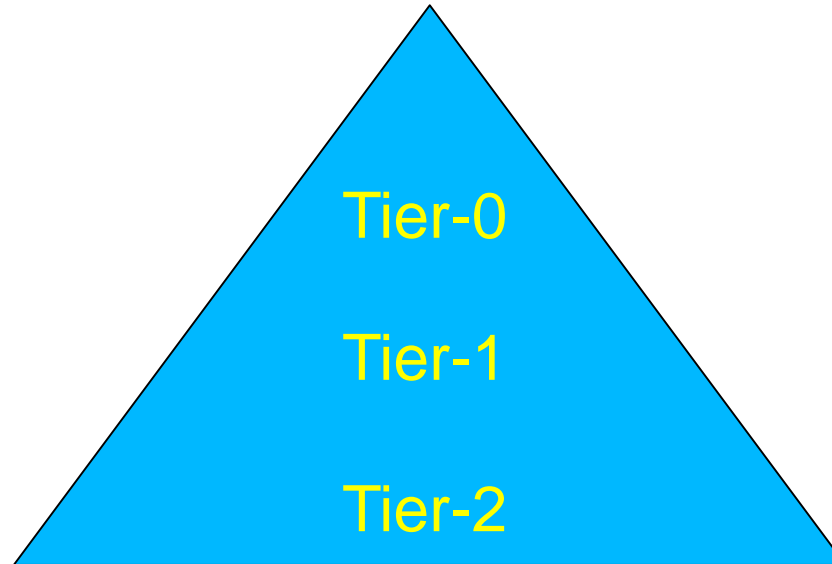


2013
Total capacity

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)
- 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 in addition to be the french representative in PRACE

National supercomputing center from CNRS:

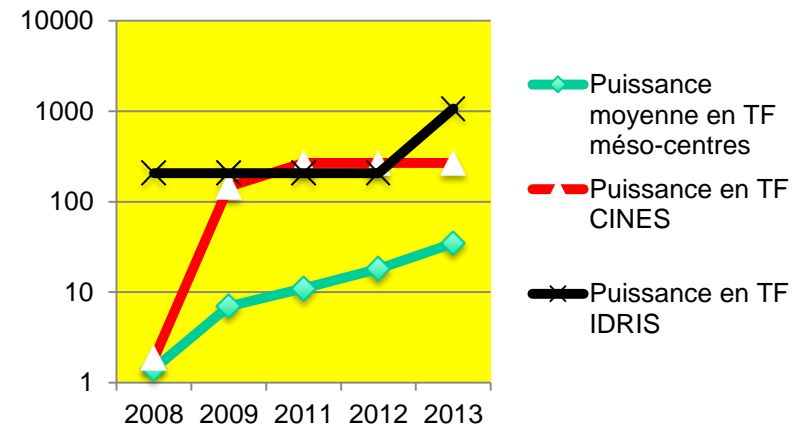
- Host and operates GENCI computers
- Around 1000 utilisateurs, 370 scientific projects
- Two new computers installed in 2012:
 - Turing: IBM Bluegene/Q, 836 Tflop/s with 65 536 cores (proc. PowerPC A2)
 - Ada: IBM 3750, 336 large SMP nodes SMP, 233 Tflop/s with 10,752 cores (proc. Intel Sandy Bridge)



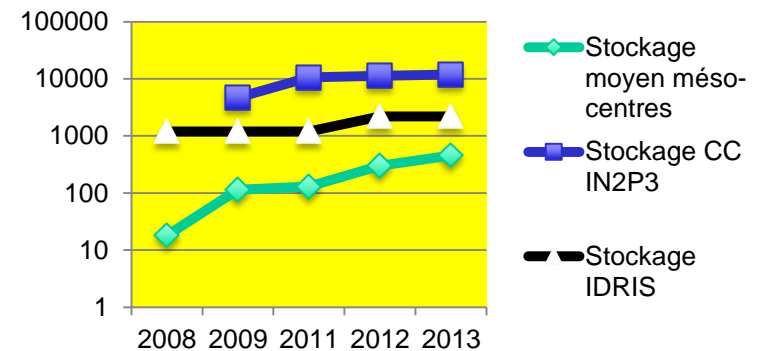
Evolution des Tiers-1 et Tiers-2 2009-2014 (données Groupe Calcul et EcoInfo)



- 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

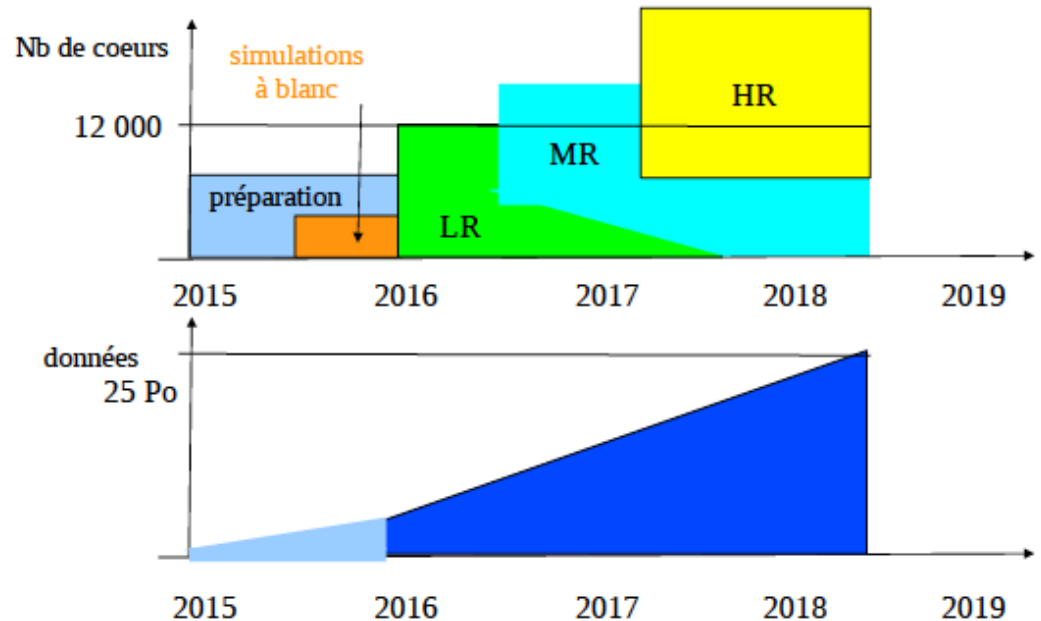


Evolution des volumes de stockage sur disques en TOctets (TB) sur le CC IN2P3 et IDRIS et en moyenne sur l'ensemble des méso-centres

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 et un besoin de structuration nationale et au niveau des sites



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

Conséquences : coûts croissants



- Calcul intensif :
 - Coordination au moins au sommet de la pyramide du calcul (Tiers-0 / Tiers-1 / Tiers-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 !!!

Maîtrise des coûts



- Autres pistes permettant de maîtriser les coûts suggérées par le Groupe **EcoInfo** :
 - **Maîtriser les coûts électriques** en installant une métrologie systématique
 - **Eviter les installations surdimensionnées** et opter pour de la modularité
 - Innover en terme de systèmes de froid
 - Contraindre à l'achat de matériel **efficace énergétiquement**, ce qui implique pour avoir des gains significatifs **mutualisation des infrastructures en particulier au niveau des Tiers-3** : e.g. économie de 60% sur le coût électrique avec un hébergement plus efficace (réduction du PUE de 2 à 1,3) qui induit économie de **500 K€ / an** pour les 2000 serveurs de calcul et de données achetés sur les marchés CNRS entre 2003 et 2013.
- Etude mutualisation du stockage à l'Université Grenoble-I :
 - Recensement de 44 locaux climatisés
 - Coût fluides estimé à 700 K€ (1/3 consommation énergétique de l'établissement)
 - Besoins de stockage croissant de 1,2 Po en 2012 à 3,4 Po en 2018 (sans compter le stockage nécessaire au calcul intensif).

Ecosystème riche au sein du CNRS



- 2,500 personnes directement concernées par le HPC au CNRS comme outil ou objet de recherche
- Données au cœur des pratiques de plusieurs instituts : IN2P3, INSB, INEE, INSU, INS2I, INSMI, besoins émergents à INSHS (humanités, sciences sociales, économie,...)
- Au CNRS centres / initiatives d'envergure nationale :
 - **CC IN2P3** Tier-1 WLCG : données en physique des hautes énergies (CERN), ouverture vers biologie, astronomie & astrophysique (données du LSST).
Compétence / expertise Grilles et Cloud
 - **IDRIS** Tier-1 GENCI: HPC et données produites par les simulations. Expertise / support HPC aux projets pluridisciplinaires et formation
 - **Institut des Grilles et du Cloud** (France Grilles), **Maison de la Simulation**, Mission pour **l'interdisciplinarité** (MASTODONS), lien avec **les méso-centres**, observatoires des Sciences de l'Univers (**OSU**), plateformes génomiques ...

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 eut 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

Vision du calcul intensif au CNRS



- Calcul intensif et valorisation des données : *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 (et de l'analyse de données) 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 des 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

Éléments stratégiques



- 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 de Recherche Scientifique et Technologique en Calcul et Données (RSTCD) agrégeant infrastructures calcul et données, équipes de recherche autour de grands défis scientifiques, **s'appuyant sur le contexte local**
- Reconnaissance des chercheurs et ingénieurs, à l'interface domaine disciplinaires / maths / informatique /

Conclusion



- Calcul intensif / données : *grand instrument scientifique pluridisciplinaire*, catalyseur de nouvelles connaissances scientifiques
- Besoins calcul / stockage **en forte croissance** d'où l'**explosion des demandes au niveau du 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**
- *Stratégie du CNRS : coordination / rationalisation des investissements aux niveaux site / national autour de défis scientifiques avec l'ensemble des acteurs concernés*