La Physique des Particules

Catherine Biscarat (LPSC Grenoble)

Pour le groupe de travail « Calcul »

Le groupe de travail “Calcul”

Nicolas Arnaud (LAL) - Edouard Audit (Irfu - SAp) – Volker Beckmann (APC) – Dominique Boutigny (CC-IN2P3) – Vincent Breton (LPC Clermont et IdG) – Sacha Brun (Irfu - SAp) – Jaume Carbonell - Frédérique Chollet (LAPP) - Cristinel Diaconu (CPPM) – Pierre Girard (CC-IN2P3) – Gilbert Grosdidier (LAL) - Sébastien Incerti (CENBG) – Xavier Jeannin (CNRS – Renater) - Edith Knoops (CPPM) Giovanni Lamanna (LAPP) – Eric Lançon (Irfu – SPP) – Fairouz Malek (LPSC) – Jean-Pierre Meyer (Irfu – SPP)) – Thierry Ollivier (IPNL) – Yannick Patois (IPHC) – Olivier Pène - Roman Pöschl (LAL) – Ghita Rahal (CC-IN2P3) – Patrick Roudeau (LAL) Frédéric Schaer (Irfu – SPP) – Olivier Stezowski (IPNL)

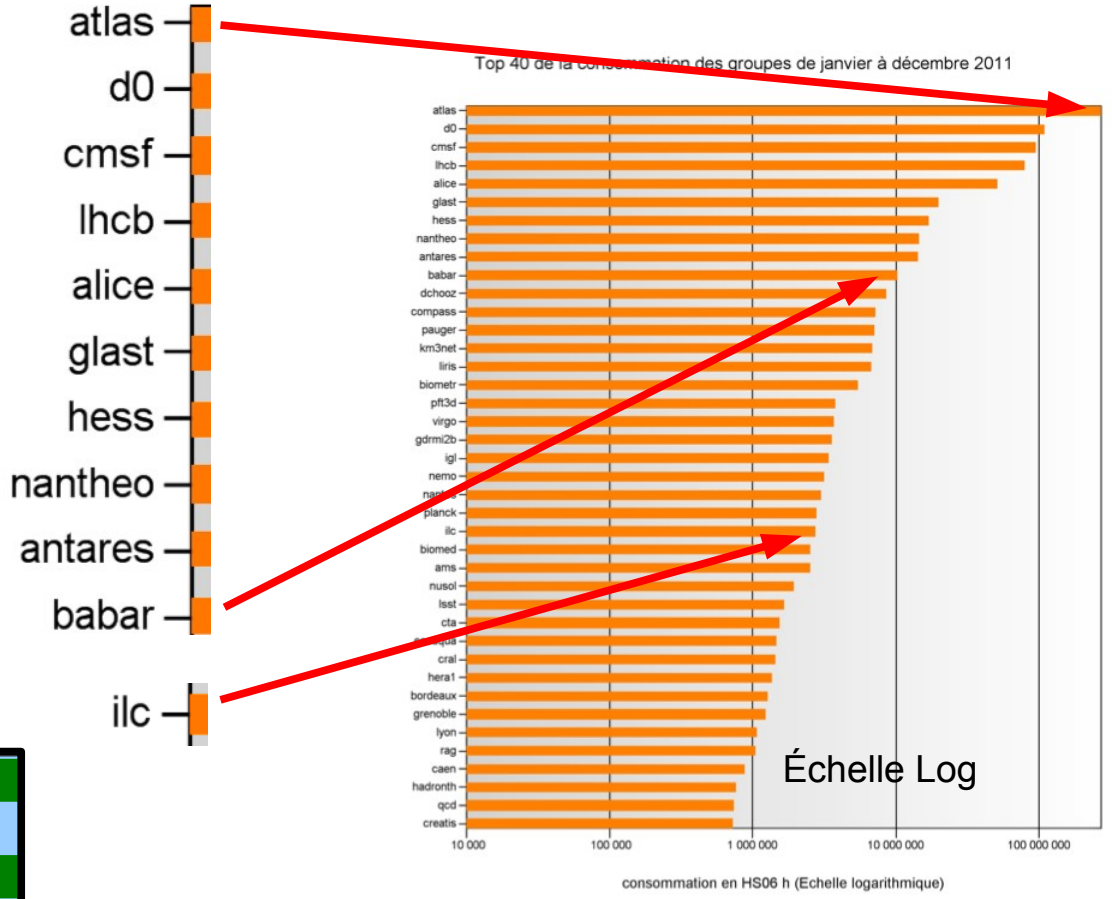
Remerciements à :

Claude Charlot (LLR), Sébastien Binet (LAL), Tibor Kurca (IPNL)

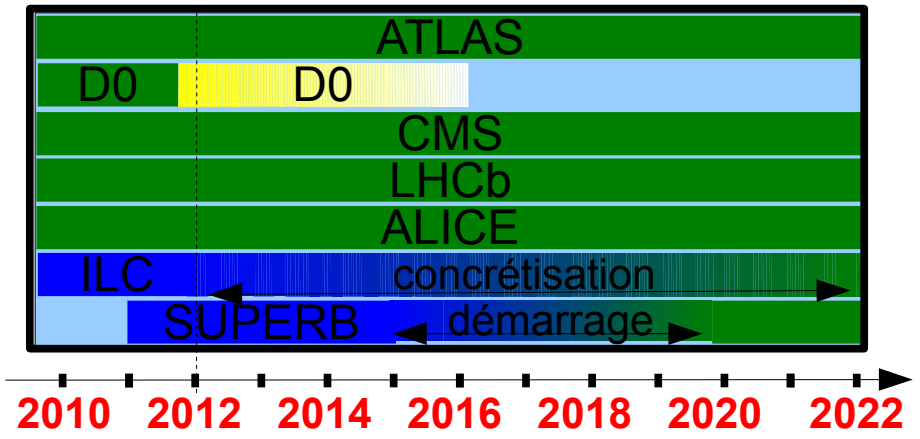
Les consommateurs de ressources

Une vue actuelle :

Classement des groupes par consommation de CPU en 2011 au CCIN2P3



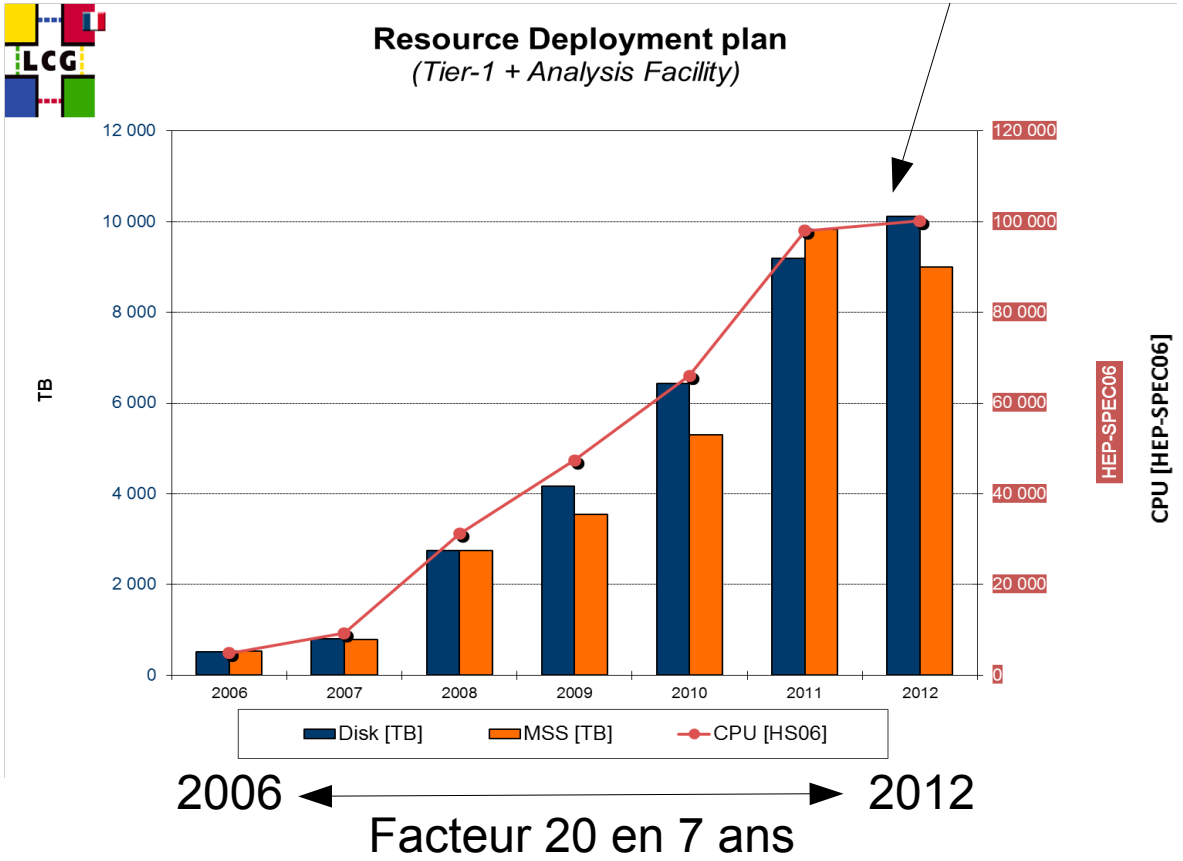
Les calendriers :

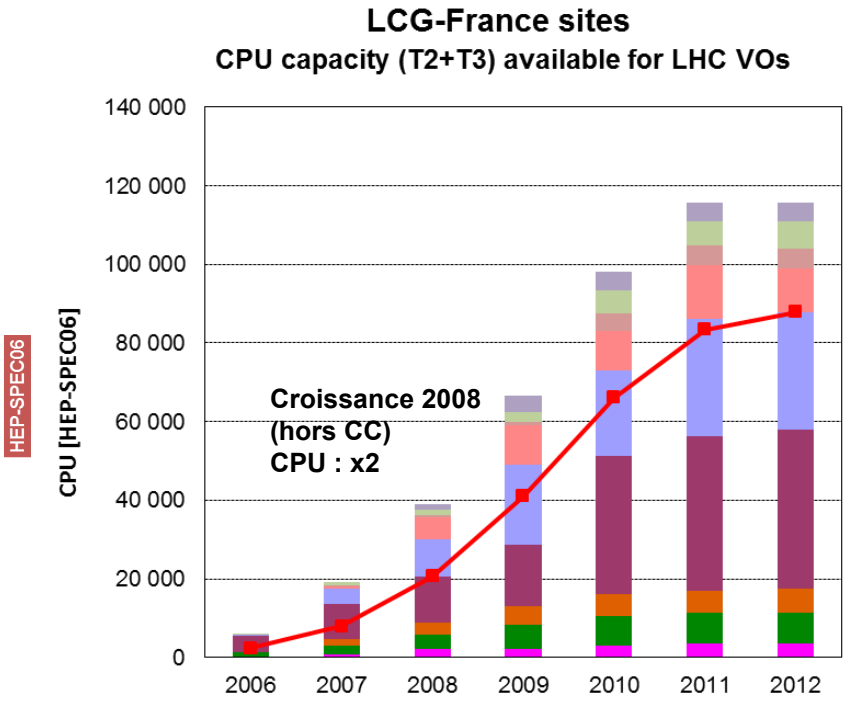


Les ressources allouées au LHC en France

- Contribution de la France à l'effort mondial : **10%**
  - Clé de répartition budgétaire : 15% (ALICE, LHCb), 45% (ATLAS), 25% (CMS)

Restriction budgétaire





Les ressources à venir LHC

Expression des besoins des expériences : 2 ans à l'avance

- Pas de plan à long terme (modèles changent, luminosité)

2014 sera annoncée à la fin du mois

2012 requêtes

Disk (PB)	T0	CAF	T1s	T2s
Required	7.6	0.24	7.0	12.4
Pledged	8.1		7.22	9.11 (12.9)
Difference	6%		3%	-36%

CPU (KHEP06)	T0	CAF	T1s	T2s
Required	90.0	35.0	95.0	207
Pledged	90		95	115 (194)
Difference	0%		0%	-80%

Chaque T1

+50% (accumulation des données)

2013 requêtes

Disk (PB)	T0	CAF	T1s	T2s
Required	13.2	0.24	10.9	19.4
Pledged				
Difference				

CPU (KHEP06)	T0	CAF	T1s	T2s
Required	90.0	35.0	95.0	194.8
Pledged				
Difference				

[Ian Bird, Overview Board WLCG, 9th mars 2012]

Ressources dans les sites (calcul grossier) :

- Budget constant + loi de Moore → augmentation des ressources de **30% / an**
  - On compte 10 % / an lors des arrêts du LHC (2013, 2017)

ILC

Activités de calcul actuelles :

- Analyse données faisceaux tests
- Simulations détaillées

- *Technical Design Report* ILC
- *Conceptual Design Report* CLIC

Requêtes au CCIN2P3

Année, scénario	Stockage [TB]	CPU [kHS06]
2011	50	2 000
2012	60	12 000
2013	20	2 000
2014	Idem 2013	
≥ 2014, ILC concretisé	3-4 x 2011	
≥ 2014, ILC retardé	Idem 2013	

Modèle de calcul :

- **Distribué** (grille, VO ILC ; ouvert au *cloud* si cette technologie prédomine)
- Le **CCIN2P3** sera sollicité pour jouer un rôle majeur (25% cette année)
- **2014-1016 : années charnières pour organiser le calcul**

Requêtes :

- modestes vs LHC (2012 : 3% de la réquête ATLAS au CC)
- Mais besoin de services stables
- Développer la **structure en France** et l'intégrer dans une **structure mondiale**

SuperB

Très grandes quantités de données (ab^{-1})

- Comparable à **une expérience LHC** au moment de SLHC

Modèle de calcul :

- Idem Babar (Données brutes : 2 copies)
- sur la **grille** (architecture LCG)
- *TDR* computing prévu en 2013
  - Avec **MAJ** du modèle/estimation

Estimation à long terme des besoins

- Souhaité : **rôle majeur du CCIN2P3** (gd succès avec Babar)

Préconisé au CCIN2P3

Année	Données Brutes [PB]	Stock. bande [PB]	Stock. disque [PB]	CPU [kHEP06]
1 (mise en place)	0	0	0	0
2 (montée 1/2)	10	0.5	1	45
3 (montée 2/2)	40	2	3	175
4 (nominale 1/5)	80	4	5	360
5 (nominale 2/5)	120	7	7	550
6 (nominale 3/5)	160	9	8	750
7 (nominale 4/5)	200	12	9	940
8 (nominale 5/5)	240	14	11	1130

1 copie des données

Participation de 10% au calcul et stockage

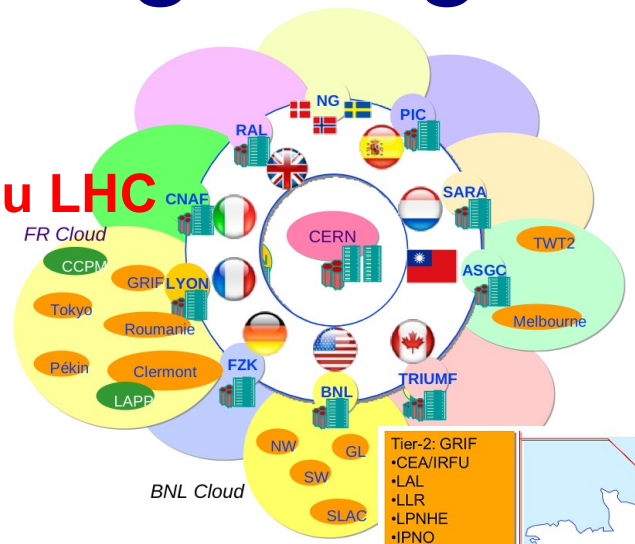
Futur :

- SuperB ouvert aux progrès dans les modèles des expériences LHC
- SuperB est engagé dans R&D (parallélisme et IO stockage)

Le LHC computing : la grille

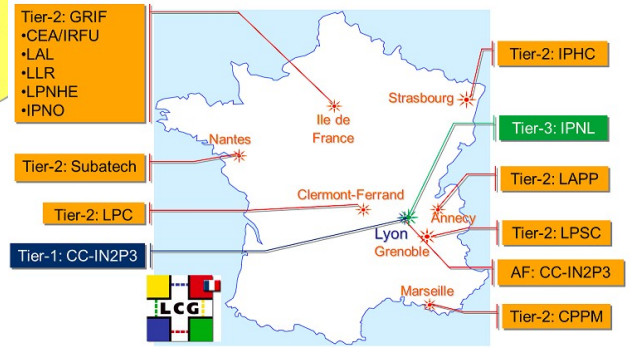
La grille WLCG en bref

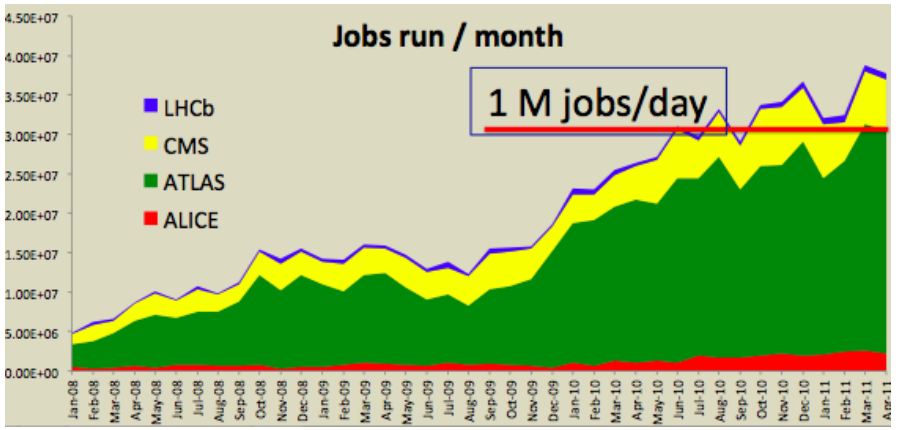
- **La solution au défit informatique du LHC**
- Des centres de calculs distribués
- Un intergiciel commun, des VO

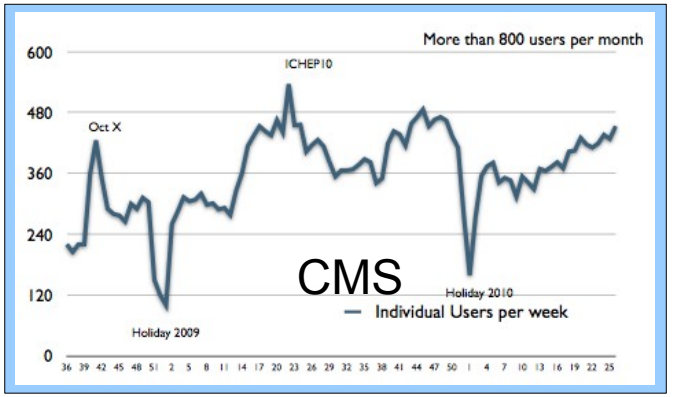


Depuis le démarrage du LHC

- La grille **marche vraiment**
  - Production MC, traitement des données
- De nombreux utilisateurs pour l'analyse finale :
  - ~800 (ATLAS, CMS); ~250 (LHCb, ALICE)







Leçons apprises LHC

La grille est complexe

- Beaucoup de travail sous-jacent (sites)
- Beaucoup d'effort pour cacher la complexité aux utilisateurs

Modèle hiérarchique des débuts (2002)

- Jobs envoyés aux données
- Données résidentes et pré-placées MAIS ne tient pas compte de la popularité
- Inquiétude vis-à-vis du **réseau** à l'époque MAIS réseau très stable
- Services distribués → nouvelle tendance à centraliser (simplification)

Evolutions passées des modèles

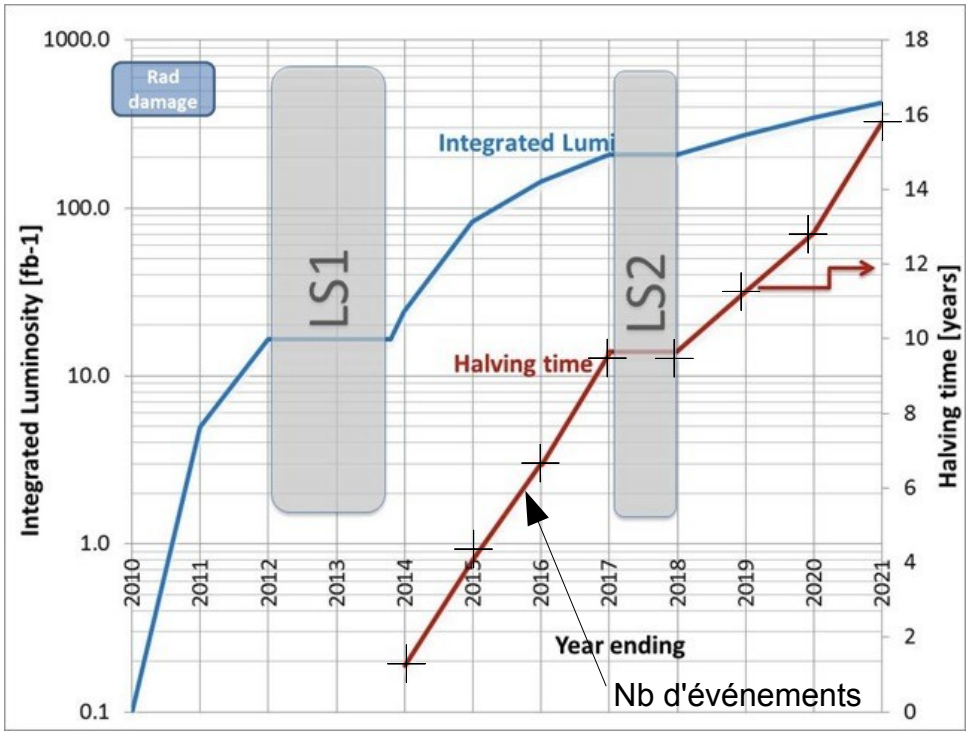
- Placement **dynamique** des données
- Distribution des **software** sur les sites (CVMFS)

Un déluge de données

Halving time : date où il y a un doublement des données

Doublement du dataset chaque année

- Stockage, accès et traitement des données
- Traitement des données
- **Passage à l'échelle** avec les technologies actuelles ?



– Problème qui n'est plus propre à notre communauté (sciences, e-commerce, réseaux sociaux, ...)

Travail en cours :

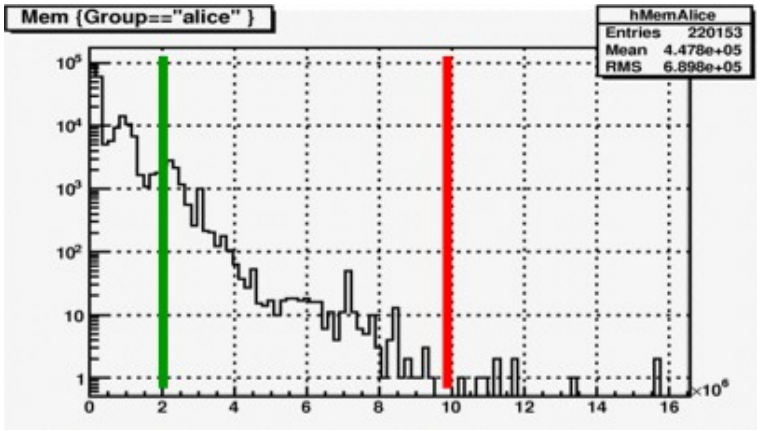
- Fédération de **stockage** (un point d'entrée unique, xRootd)
  - Condition d'un bon réseau (LHCOne)
- Développements dans les **bases de données** (rapidité d'accès)
  - Investigations de produits issus du web (Hadoop)
- **Passage au cloud** (« self-service », élasticité)

cf. Dominique Boutigny

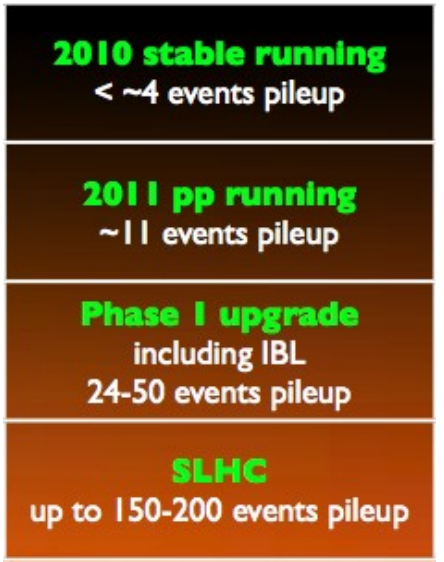
La consommation de mémoire

- Cahier des charges pour les sites : **2 GB/ core de mémoire**
- Flot d'informations à traiter
  - Eg : ATLAS MC event avec trigger et « truth » dépasse 3GB
    → demande aux sites d'augmenter les limites des queues
- Le pile-up en 2012 devrait excéder les prévisions initiales
  - Eg : CMS pourraient être limité en mémoire pour le *reprocessing* en 2012
    → augmentation de la mémoire par core
- **Pas en adéquation** avec les machines modernes

CMS

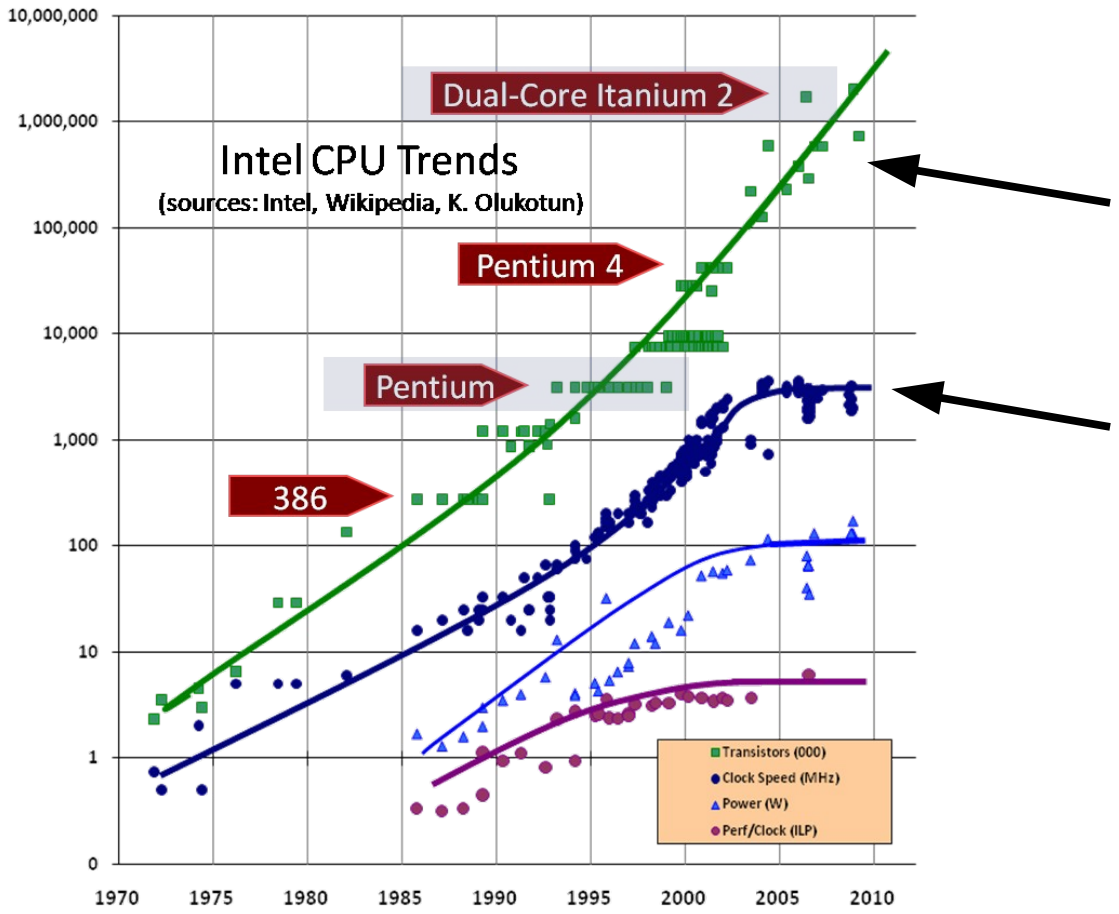


ALICE



2005 : « The free lunch is over »

- Les performances des CPUs conventionnelles **ne progressent plus**
- Si les besoins en HEP augmentent, il faut **s'adapter aux nouvelles architectures** : multi-cores, many-cores, GPU (console de jeux)



Multi-cores actuelles
- Limitation mémoire et IO par coeur

Fréquence d'horloge atteint un plateau

Solutions

Thread = fil d'exécution

- La technique HEP **ne marche plus**
  - 1 processus et un *thread* sur 1 *core* ; n cores indépendants

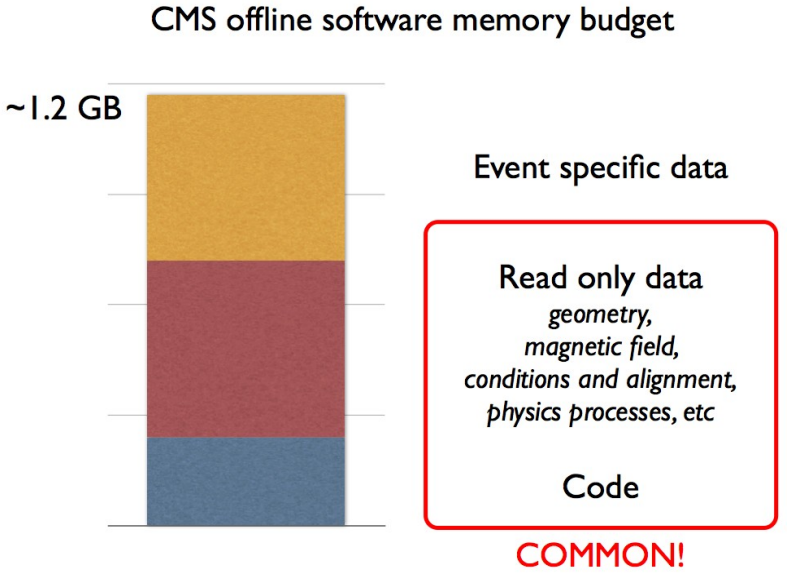
- Obligation de **paralléliser**
  - Les événements (relativement facile)
  - Les algorithmes (difficile et peu de gain en mémoire actuellement)

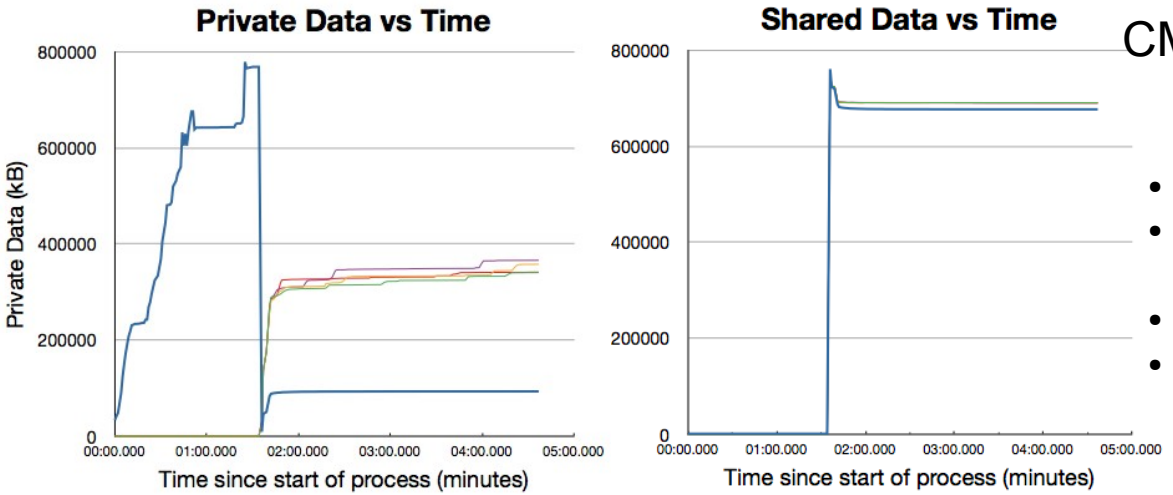
Court terme : multi-cores

Le Fork() : partage la mémoire entre parent et enfants

- fourni par Unix
- facile à adapter

- **Toutes les expériences** l'exploitent.
- Queues de **batch de tests** en place.





CMS, résultat 4 CPU, 8 core/CPU

- Shared memory / child : ~ 700 MB
- Private memory / child : ~ 375 MB

- Tot. mem. used by 32 jobs indiv. : **34 GB**
- Tot. mem. used by 32 child : **13 GB**

Gain facteur 2.5

Moyen-Long terme

Multi-cores : parallélisation des algorithmes (*multi-threading*)

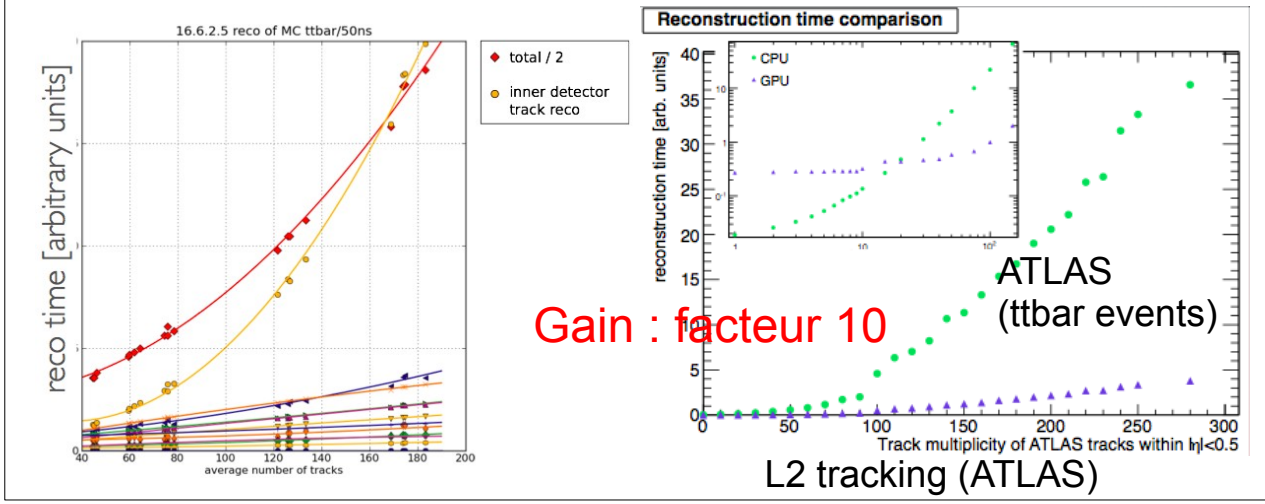
- Taille des événements
- **Ré-écriture du software** → effort global des **collaborations**
- Échelle de temps : **2014-2015**

GPUs

Many-cores

Cf. Dominique Boutigny

GPU appliqué à la recherche de “seeds”



Gain : facteur 10

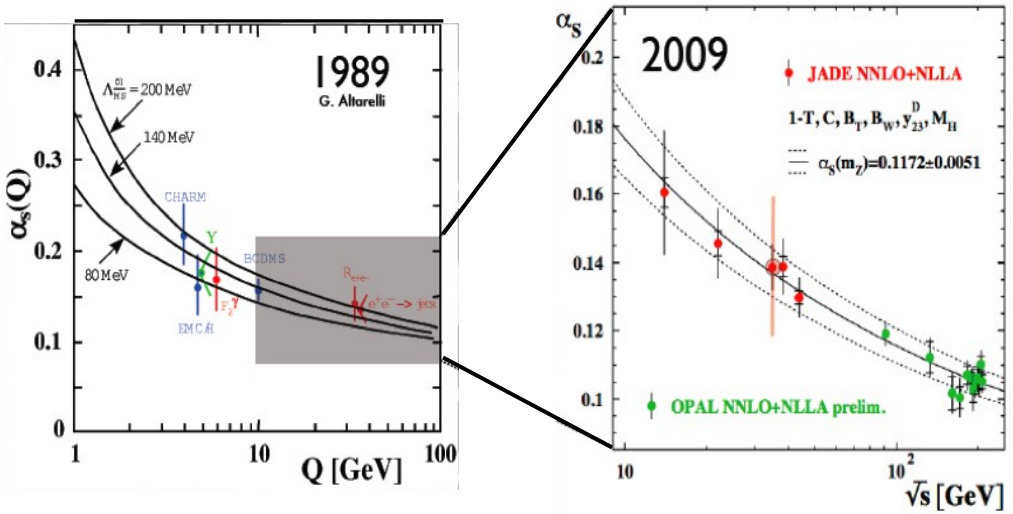
L2 tracking (ATLAS)

Prédire à long terme sur les technologies est difficile.

Mais les besoins sont immenses.

DPHEP: International Study Group on Data Preservation (ICFA Panel)

Study Group for Data Preservation and Long Term Analysis in High Energy Physics



Buts initiaux (2009)

- Confronter les modèles de données
- Clarifier les concepts, mettre en place un langage commun
- Investiguer les **aspects techniques**
- Comparer et créer des liens avec les autres champs
  - astrophysique, sciences de la vie, librairies

Achèvements et plans

- Première **publication** du concept: *arXiv:0912.0255*
- 5 w-shops: DESY, SLAC, CERN, KEK, FNAL
- 2 **projets** dédiés
  - SLAC/Babar and DESY/HERA
- 2012**:**
  - **Status Report Document**: mai 2012
  - Session à CHEP2012, w-shop automne Europe
  - **Consolider la collaboration internationale**

En guise de conclusion

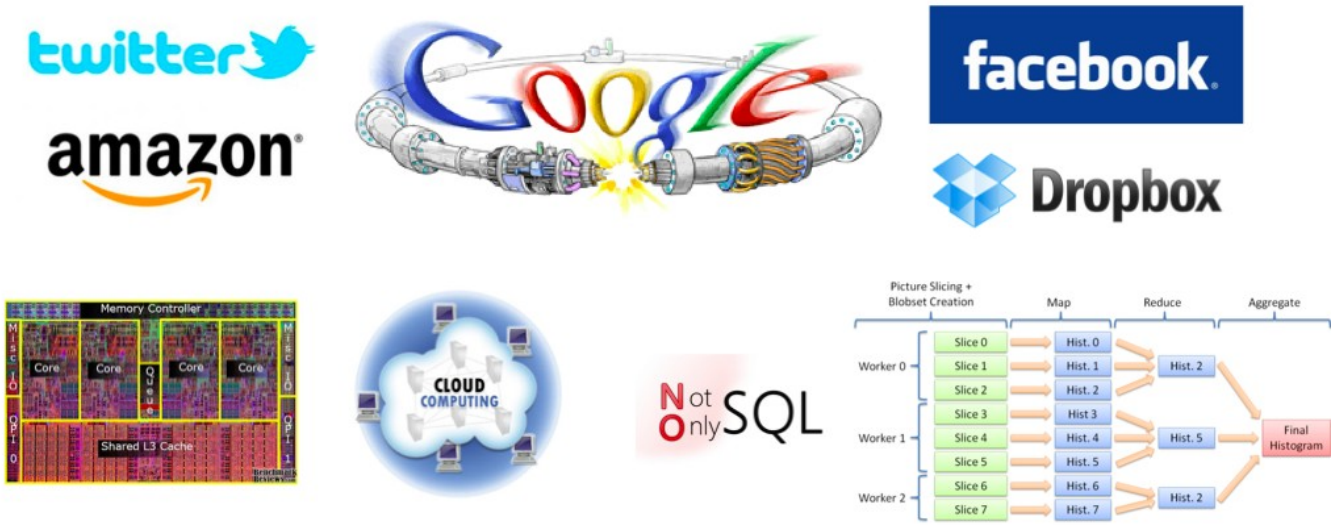
La grille est à peine née : elle a montré ses limites

- Nous sommes à **un tournant**
- Difficile de faire des projections

Sur les 10 prochaines années :

- tout une **richesse d'expériences** à soutenir dont 6 projets phares
- la physique a **besoin de plus en plus de calcul**

Pendant que nous développions le grille, **d'autres idées sont nées**



Les risques

Nos besoins sont en très forte croissance

- LHC : soutenir l'augmentation du taux de pile-up, luminosité, ecm
- Multiplication des projets très gros consommateurs de ressources informatiques
- Il faut budgétiser l'informatique en amont d'un projet.

« *End of the free lunch* », la loi de Moore ne marche plus

- L'industrie a des parades mais notre software doit suivre
- Il faut préparer les futurs achats et investissements

Passage vers le cloud

- Centres de calcul de la communauté ↔ cloud
- Le stockage est le problème délicat.

Et tout ce qui n'est pas spécifique à l'informatique :

- **Budget plat au mieux**
- **Perte d'expertise par mouvements de personnels**
- **Gestion par projets de courte durée (EGI et budgets)**