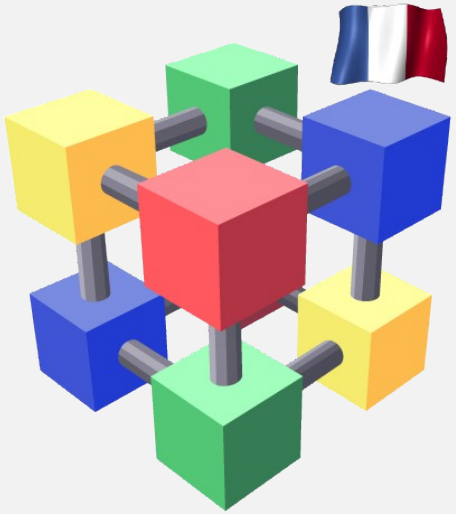
Evolution du modèle de calcul des expériences LHC

Catherine Biscarat, LPSC/IN2P3/CNRS



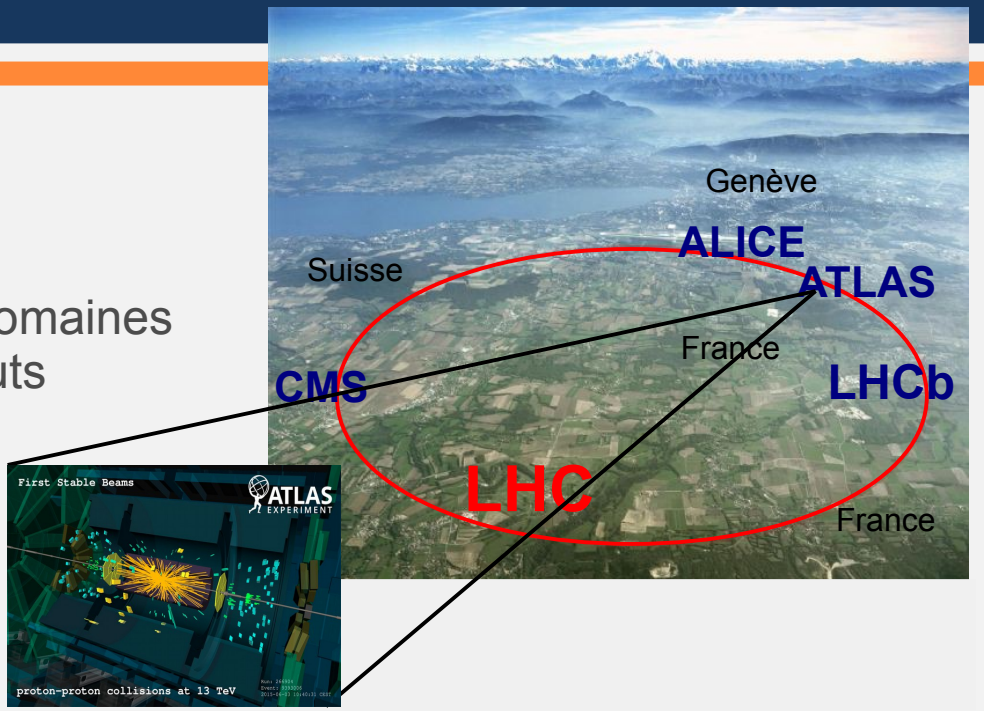
LCG France

10èmes Journées Informatique de l'IN2P3-IRFU
26-29 septembre 2016, Le Grand Lioran

Préambule

Large Hadron Collider

- Nouvelle génération de collisionneur
  - Haute énergie, haute intensité
- S'inscrit dans les axes principaux des domaines de recherche scientifiques de nos instituts
  - Origine de la masse
  - Confinement des quarks
- Equipé de quatre détecteurs
- Opérations : 2010 - 2035



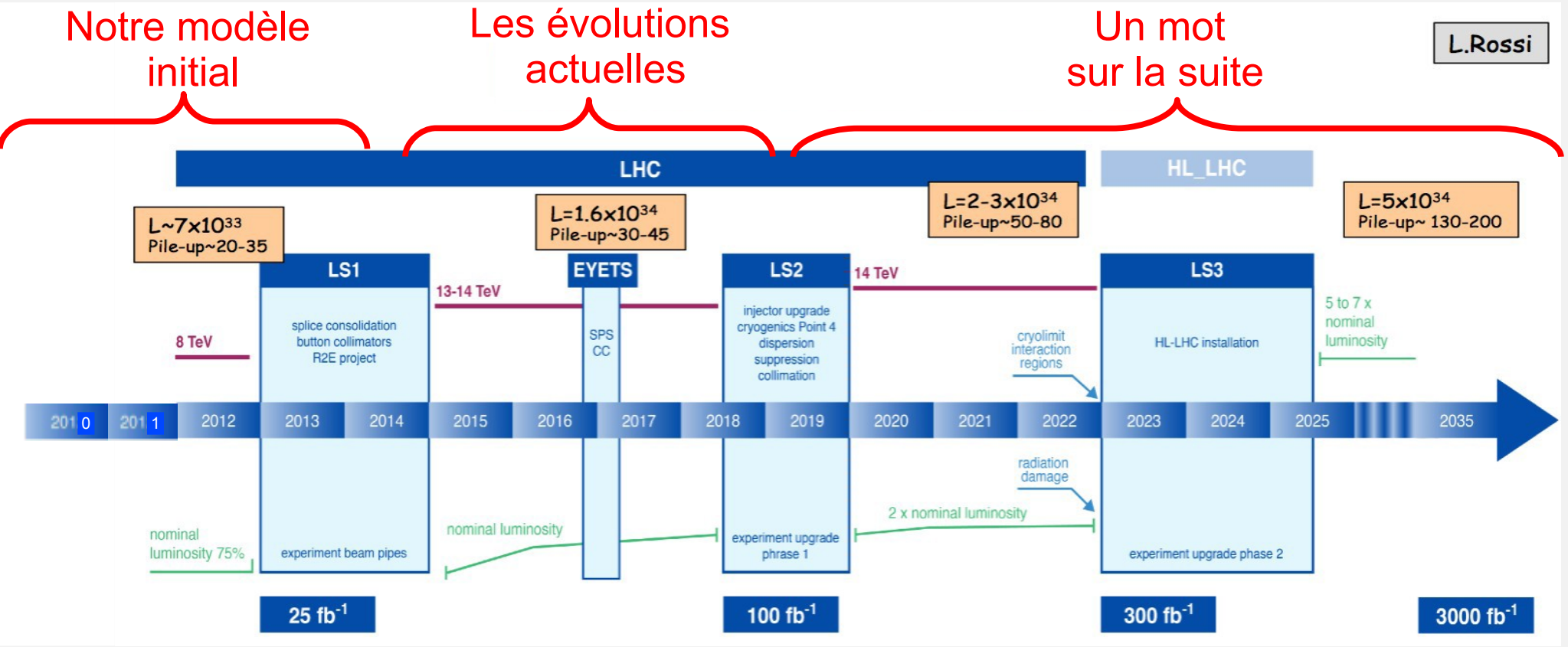
Modèle de calcul

- “Classique” pour la physique sur collisionneurs
- Petits événements indépendants -> traitement séquentiel
- Recherche de signaux rares -> grande statistique

Catégories de calcul

- Reconstruction des événements
- Analyses organisées
- Simulation des événements
- Analyse individuelles

Ce dont nous allons parler



Pile-up : complexité des événements
Lumi. intégrée : nb d'événements produits

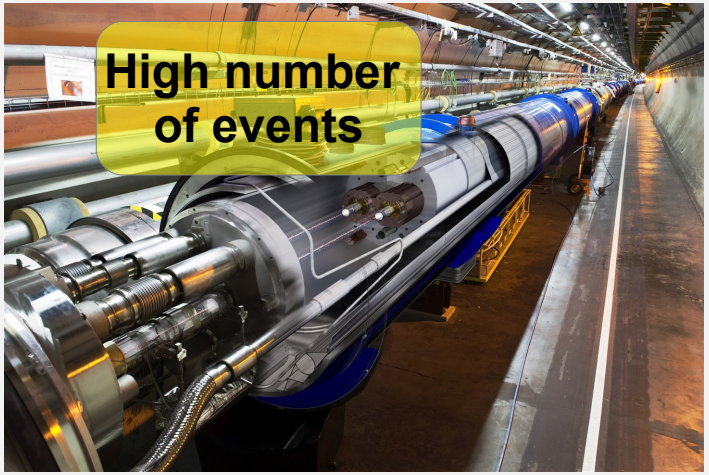
Le fruit du travail d'un nombre incalculable de personnes

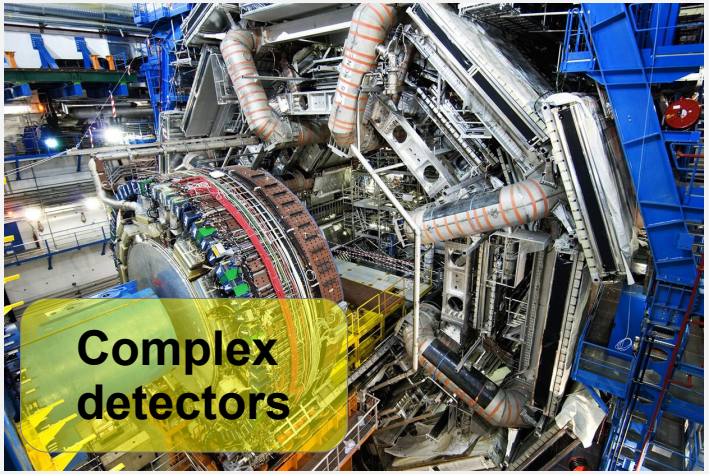
Notre modèle initial

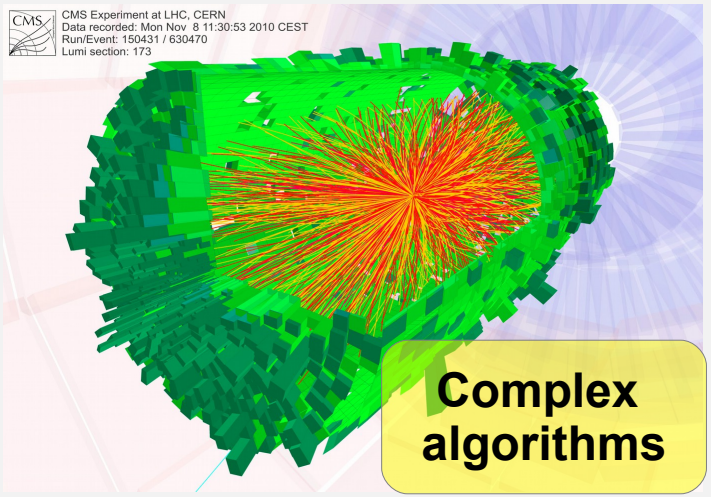
Un nouvel ordre de grandeur

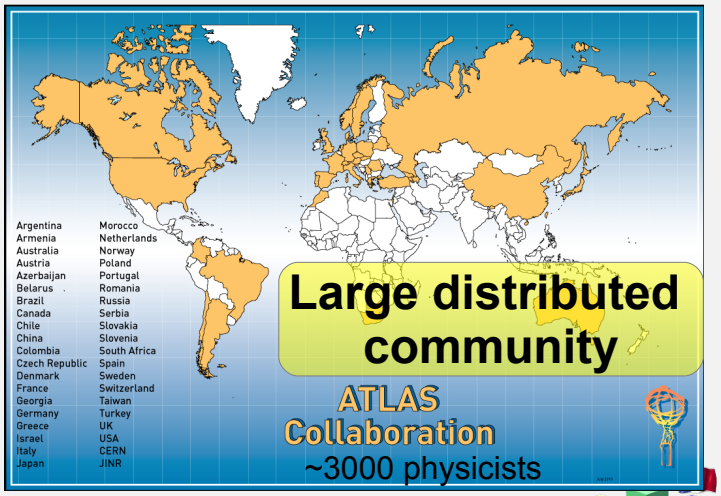
Contraintes :
Large volume de données
Ressources (CPU+stockage)
Milliers utilisateurs finaux
Archivage à long terme

Collisions : O(1) PB de données /an /détecteur
Avec les dérivés : ~15 PB / an de données LHC





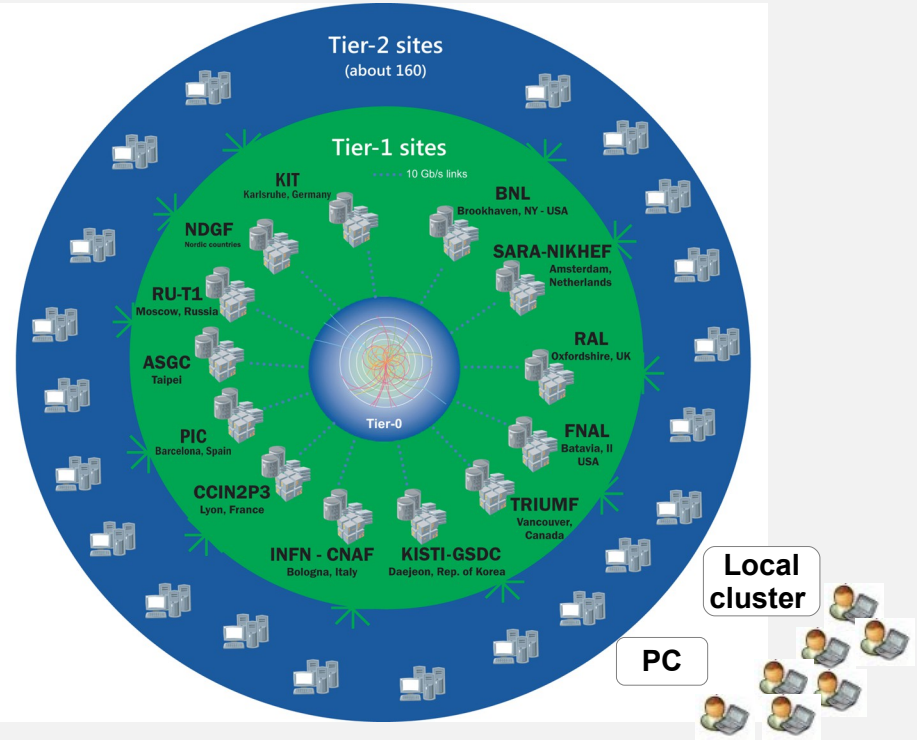




Modèle MONARC

Premier modèle pour l'informatique au LHC (1999)

- Modèle en étoile, hiérarchique, distribué
- Focus sur le contrôle du réseau (1Gb/s attendu)
- Préplacement des données et réplication



Tier-0 (CERN):

- Raw data storage
- Calibration
- Initial reconstruction
- Data distribution to T1

Tier-1:

- Long term archiving
- Subsequent reco passes
- Large scale organised analysis

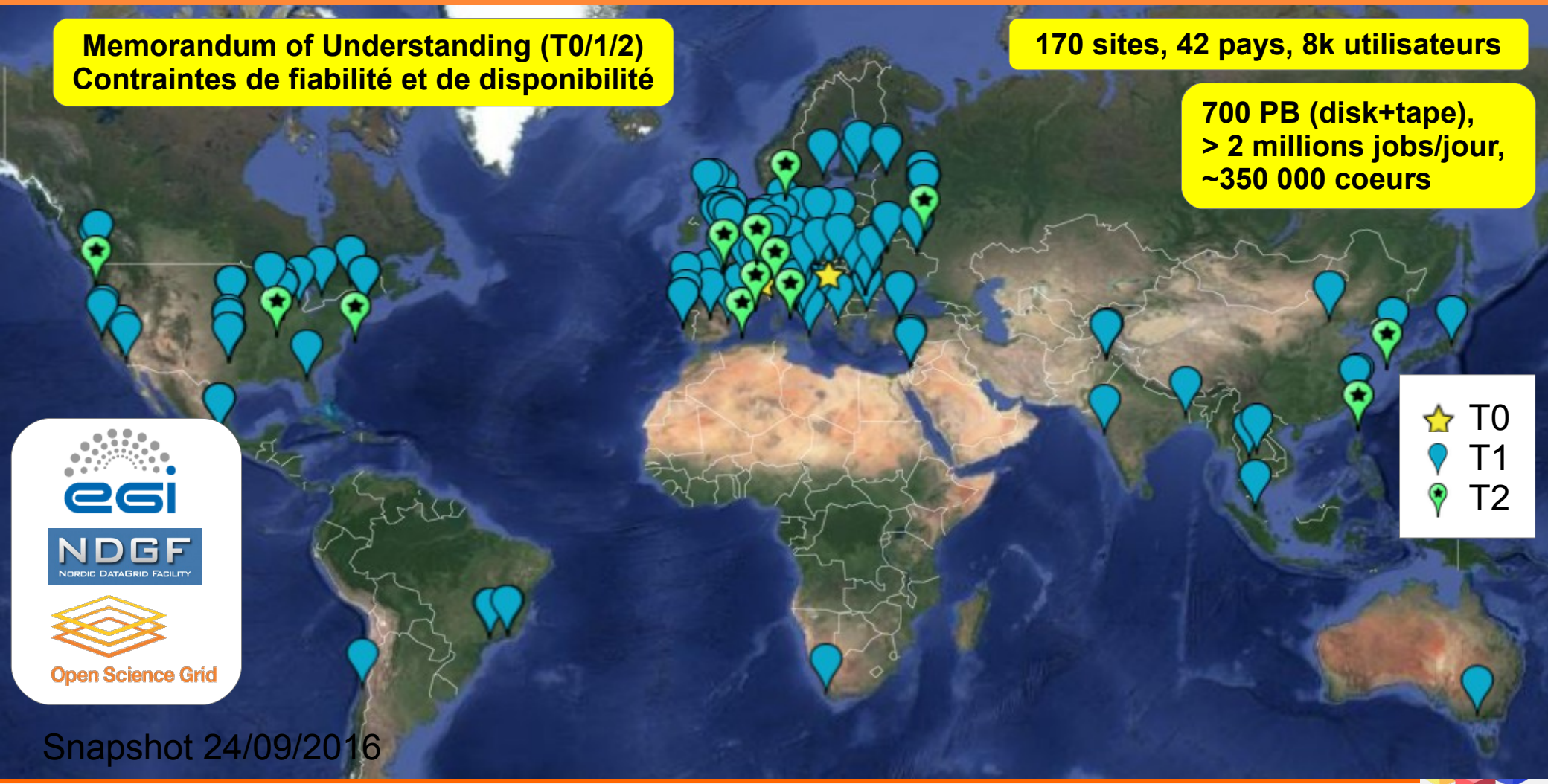
Tier-2:

- Simulation
- End user analysis

In addition (end user analysis):

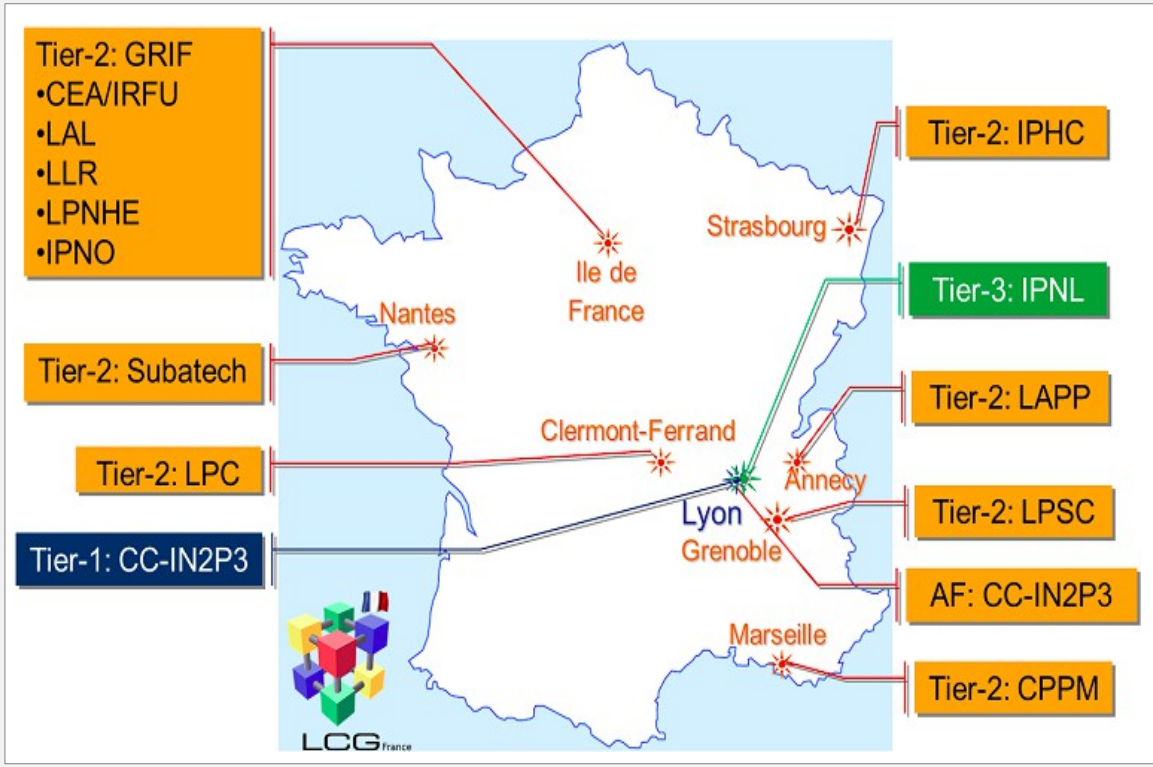
- Tier-3
- Local clusters

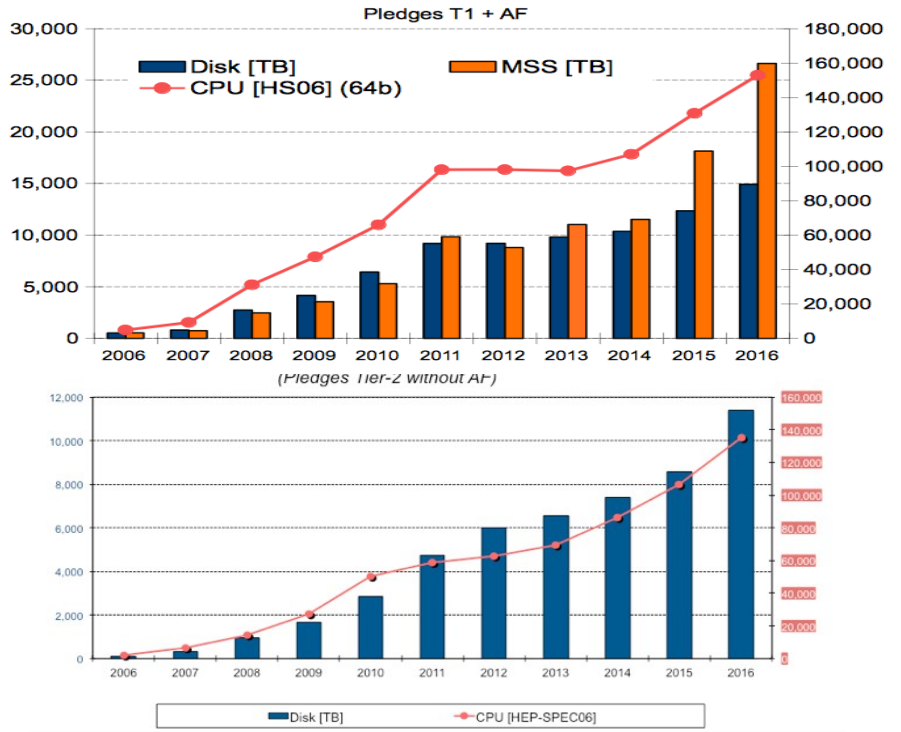
WLCG – ordres de grandeurs



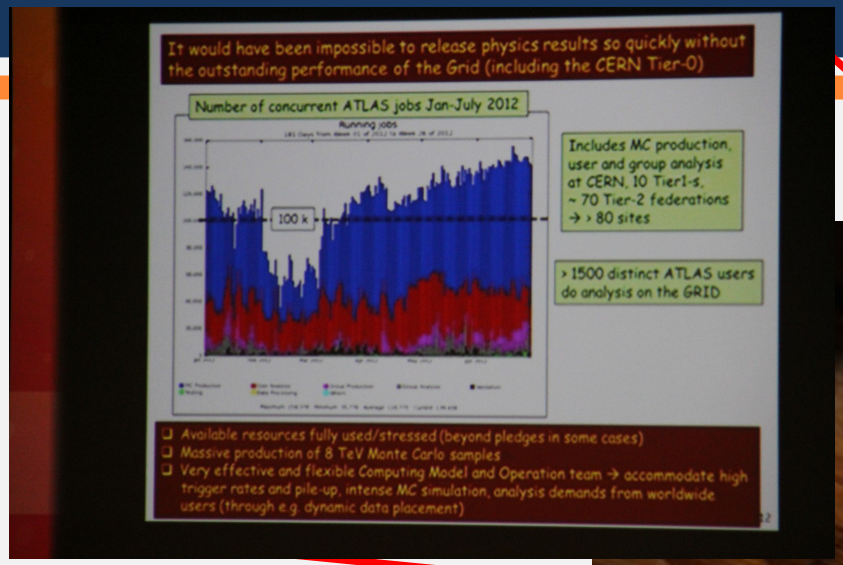
Les sites en France

- Organisés avec les expériences dans le projet « LCG-France » - CNRS/IN2P3 et CEA/IRFU
- Fournir ~10% des ressources informatiques mondiales aux expériences LHC (MoU, T1+T2)
  - En 2016 au T1 : 27 PB bande, 14 PB disk, 140 kHS06
  - Les T2 offrent environ les même capacités disk et CPU.

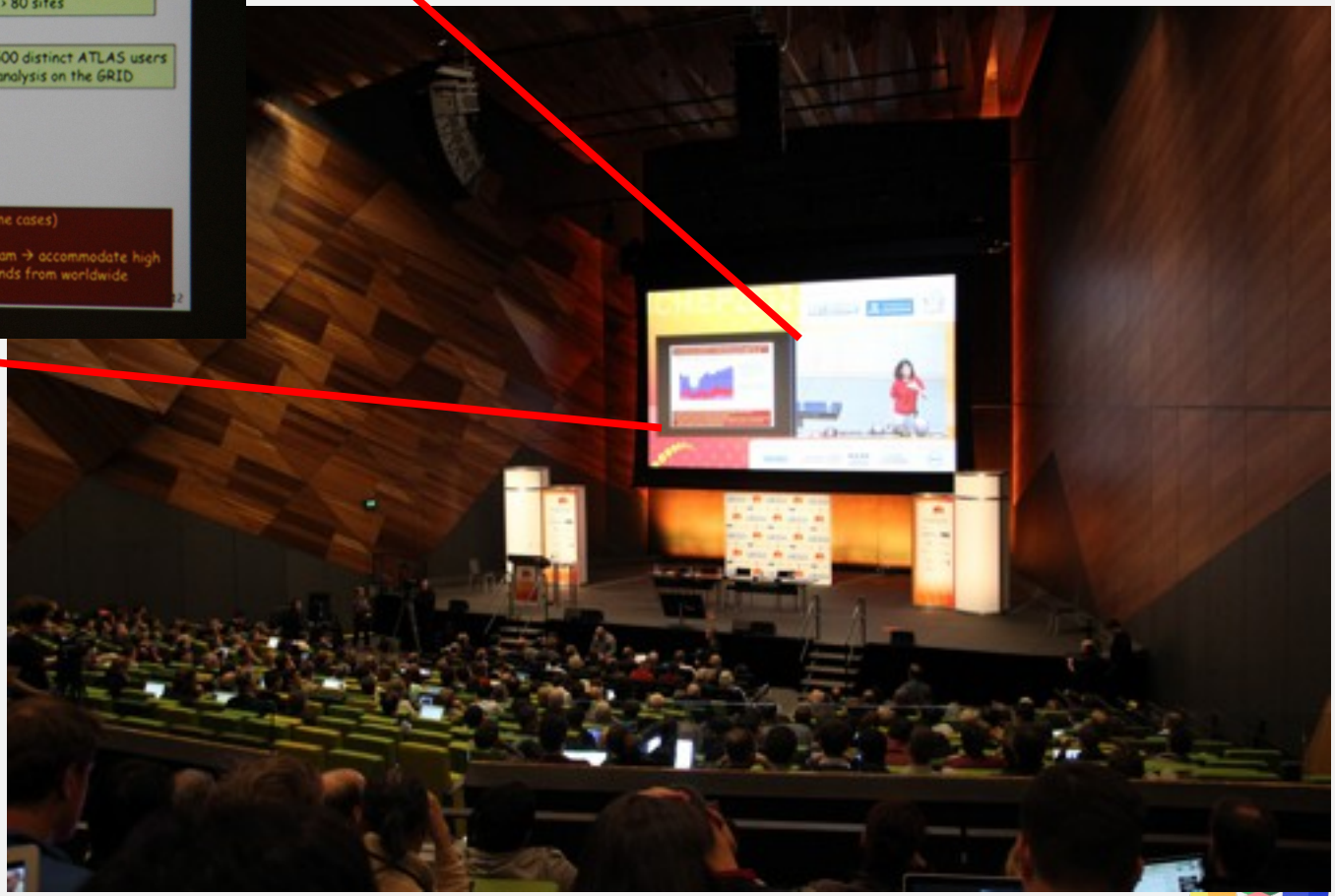




« computing enables physics »



Photography: C. Biscarat



Annonce de la découverte du maillon manquant de notre Modèle Standard, le boson de Higgs

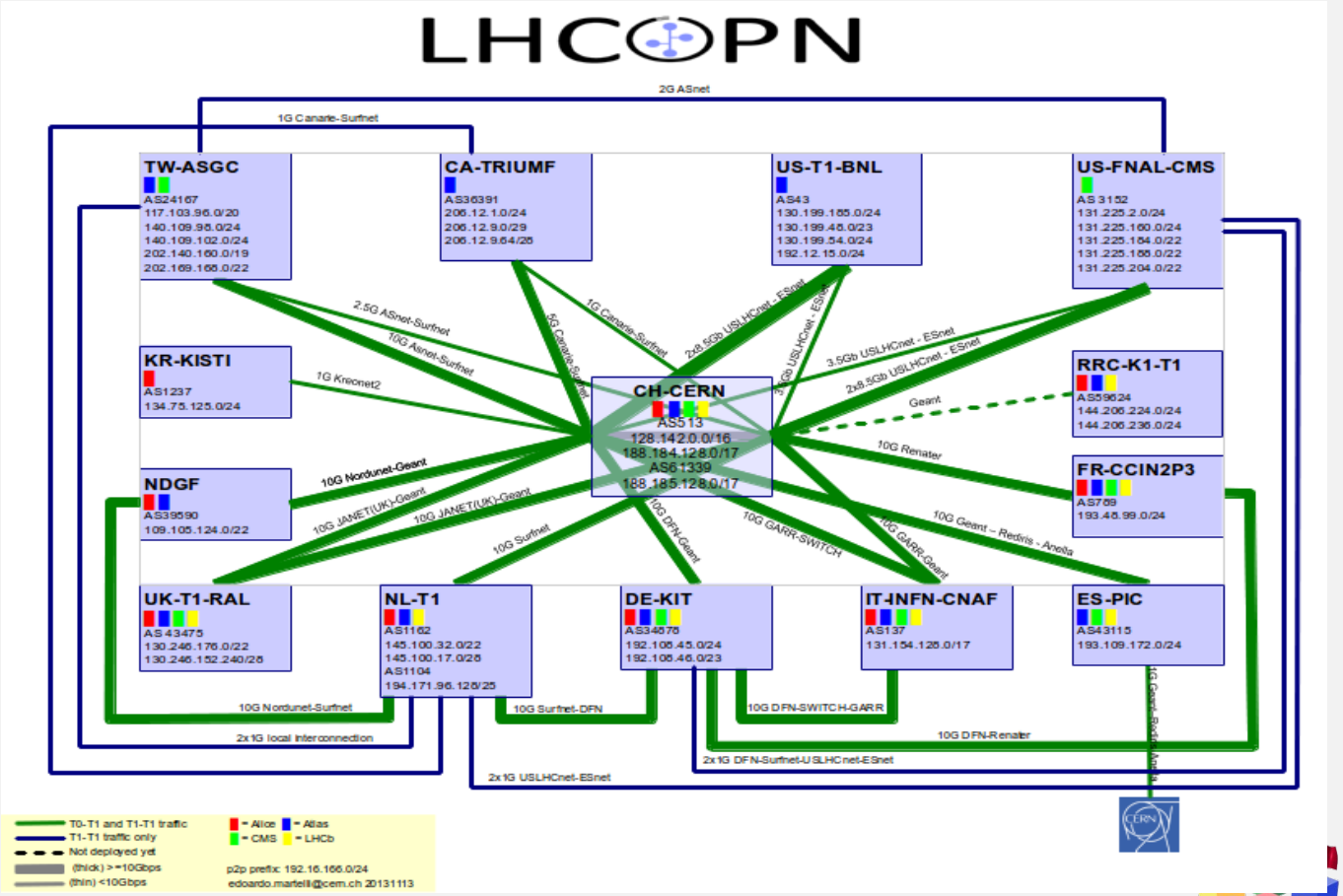
CERN seminar, July 4th 2012, retransmitted at ICHEP (Melbourne)

Le réseau – socle de notre modèle

"The Network infrastructure is the most reliable service we have"
Ian Bird, WLCG project leader

- Optical private network
- Liens dédiés et redondants
- T0-T1 et T1-T1

http://lhcopn.web.cern.ch/lhcopn/
Figure du printemps 2015

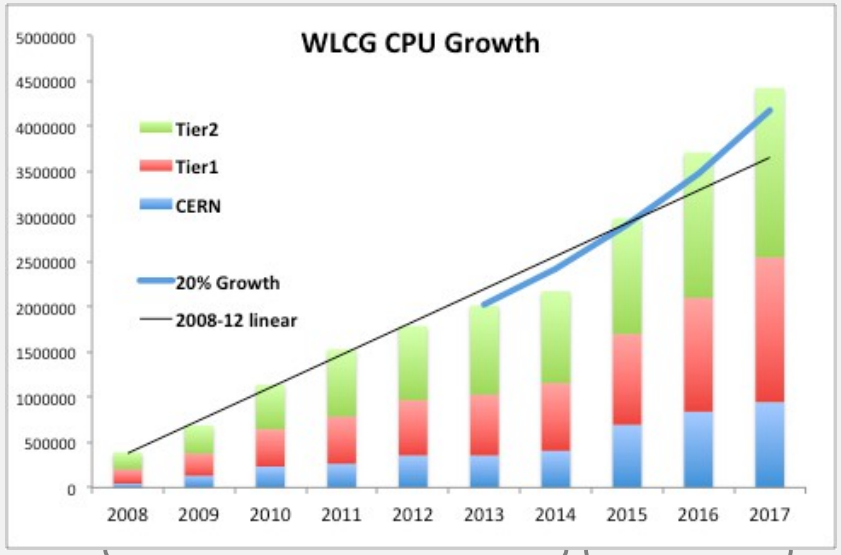


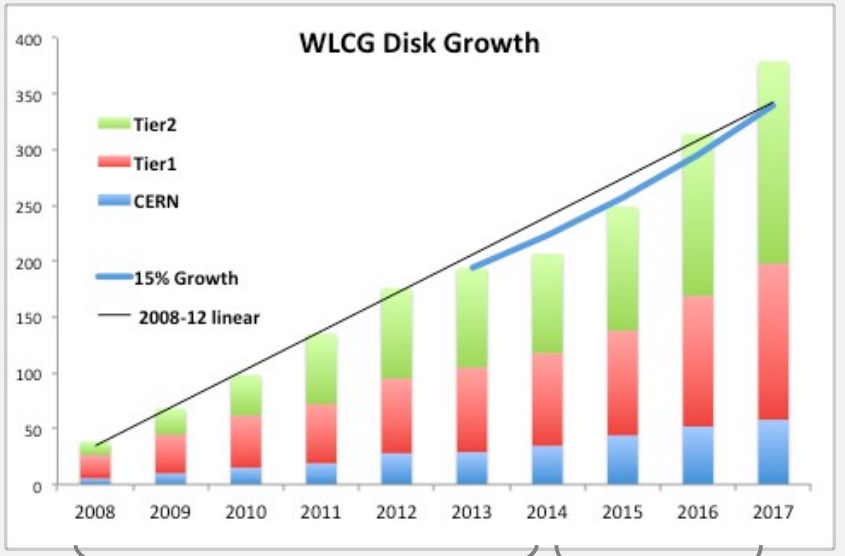
Les évolutions aujourd'hui

Evolution des besoins des expériences

Run 1 → run 2

Source : CERN-LHCC-2014-04
Document édité par WLCG (2014)
http://cds.cern.ch/record/1695401

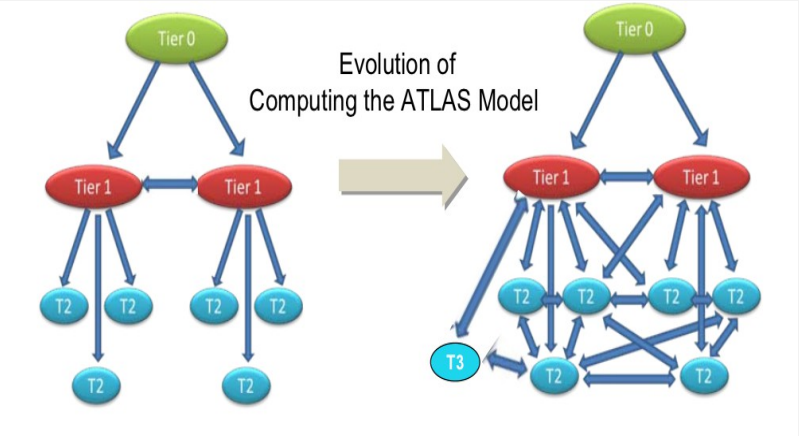


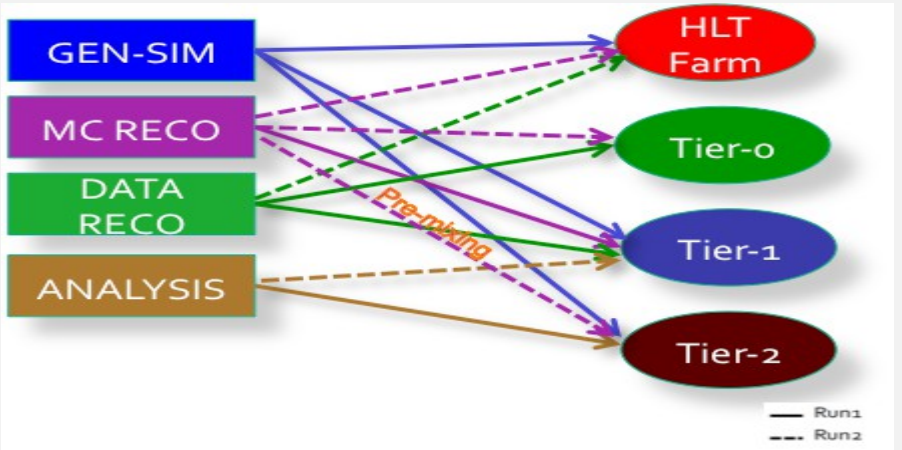


Courbes : croissance annuelle estimée à budget constant (CPU : 20% ; Disk : 15%)

Bénéfices du réseau - Un modèle plus souple

- Le modèle MONARC est relaxé
- Focus sur les capacités et les ressources des sites plus que sur leur rôle stricte
  - La hiérarchie T0/T1/T2 s'efface
- LHCONE : réseau privé mondial du LHC (et Belle 2, …), complémente LHCOPN pour tous les sites – la moitié des sites sont connectés



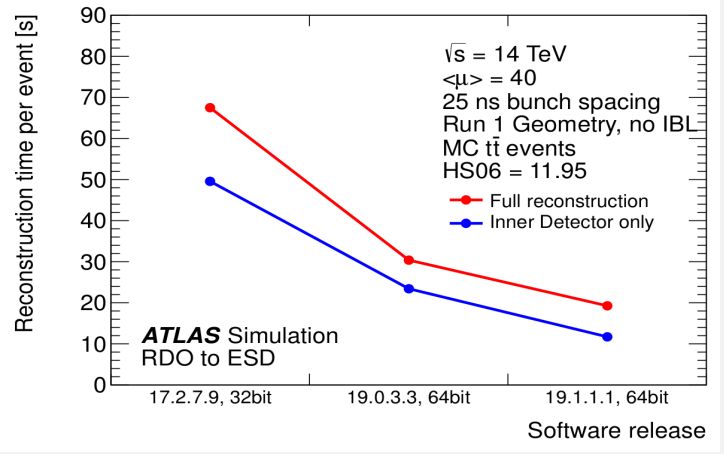


- Tout nouveau modèle de ATLAS : « Noyaux et satellites »
  - Choix dynamique (performances réseau) des centres distributeurs de données

Les flots de traitement

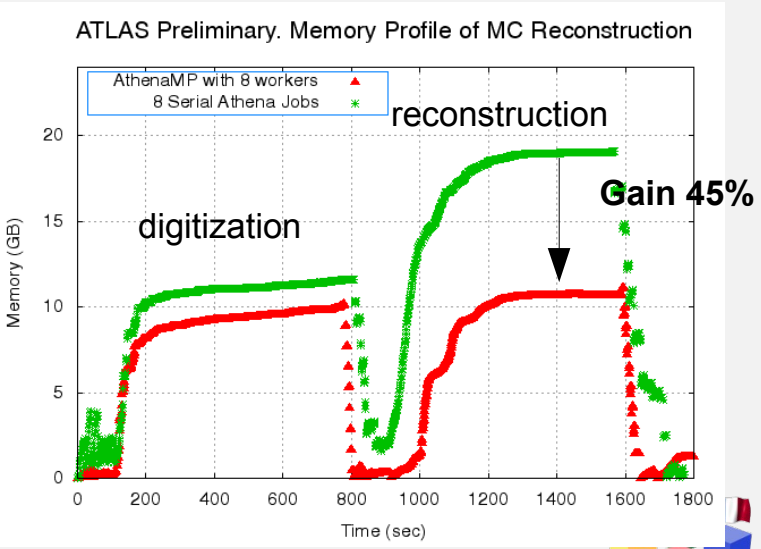
Des améliorations continues

- Optimisation du software
- Moins de passes de re-reconstruction
- Analyses « organisées »
- Formats d'analyse réduits



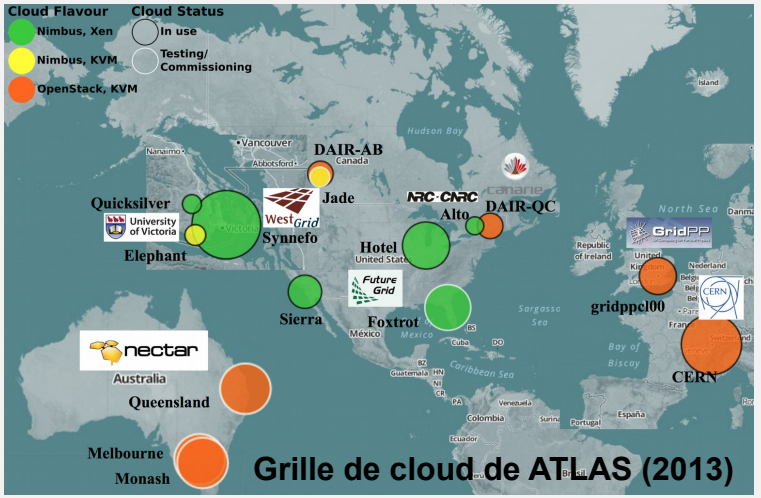
Utilisation optimale du parc de CPU

- Constructeurs : moins de mémoire par cœur
- Augmentation des besoins avec la luminosité
- Parallélisation des codes (nouveau en HEP)
  - fork des événements (ATLAS)
  - au niveau des algorithmes (CMS)



Des sites naissent « clouds »

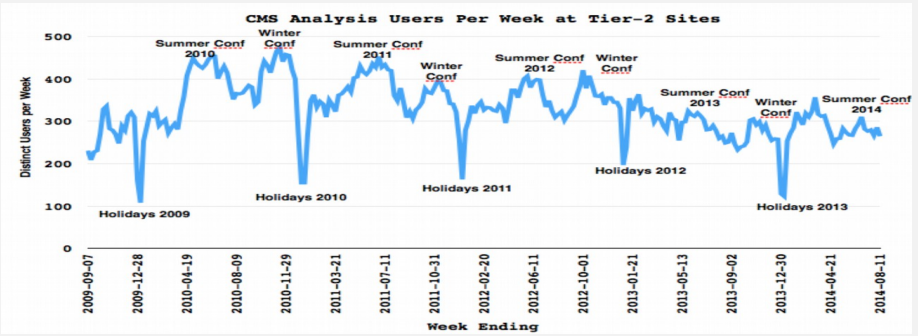
- Intégrés aux *workflow* des expériences
  - Over-head < 5% (CERN)
  - Les containers peuvent améliorer ce chiffre
- Le Tier-0 s'est doté d'une annexe
  - Les ressources sont orchestrées dans un cloud privé (extension dynamique du Tier-0 historique au CERN)
  - Il y a un an : 4600 HV, 125 000 cœurs
- Fermes de déclenchement en ligne
  - Dotées d'une couche cloud pour un switch rapide online/offline entre les remplissages du LHC



Les ressources opportunistes

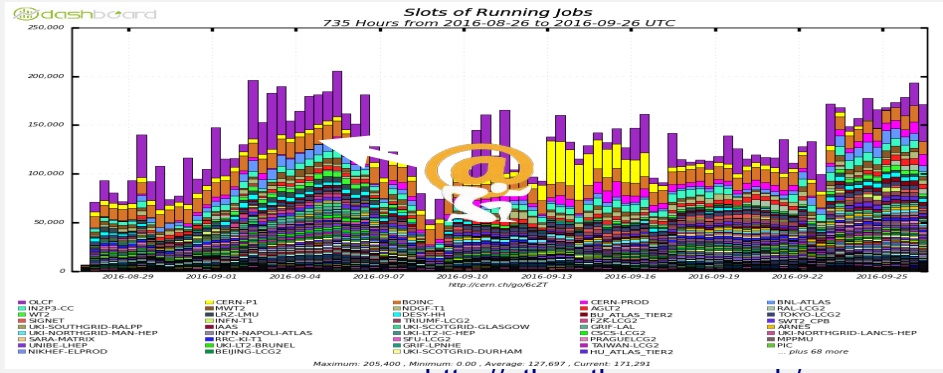
Les clouds commerciaux - Elasticité

- Déjà de l'expérience : Google, HELIXNebula, Amazon Web Service
- En Europe, HNSciCloud (démarrage 2016)
  - Prototype de *cloud* public/privé
  - Comprendre et maitriser les coûts



BOINC et les projets @home

- Sur le modèle de SETI@home
- La première tentative : ATLAS@home



http://atlasathome.cern.ch/

Les centres HPC

- HEP : sans besoin des spécificités du HPC
- Des collaborations avec de grands centres HPC
- Des initiatives locales/régionales (cycles vides)

- Tâches préférées : génération et simulation (CPU intensif, sans connexion aux DB)
- Optimisation : tâches qui peuvent être interrompues (Event Service – ATLAS)

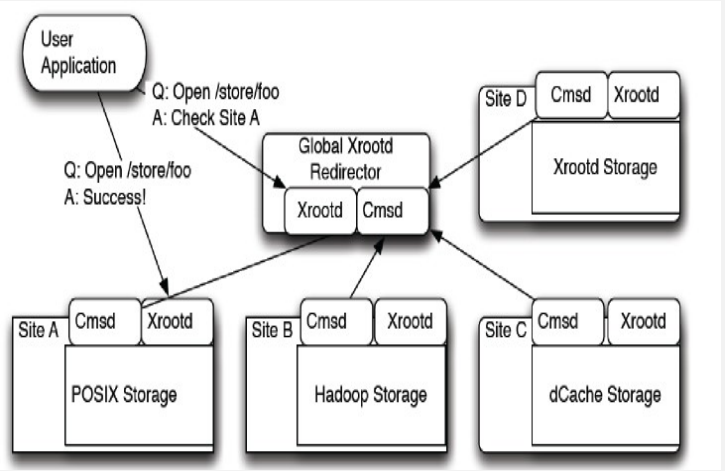
Le stockage n'est pas opportuniste

Gestion dynamique des données

- Examen de la « popularité des données »
- Effacement dynamique

Accès transparents aux données

- Fédération de données
- En production routinière : fall-back
  - récupération des données si l'accès local échoue.
- Exploratoire : accès distant aux données



Rôle des « petits » sites

- Le stockage représente le plus d'effort (sites et expériences)
- Petits == peu de support dans le site
- Diminuer le nombre de points d'entrées de stockage
  - Fédération de sites
  - Des sites complètement « CPU »
    - ATLAS recommande une coupure à 400 TB
    - Cache avec ou sans catalogue, disk-less
- Simplifier les opérations du stockage
  - Diminution du nombre de protocoles

Pour finir

Où nous en sommes

- Les données du LHC sont traitées sur la grille avec succès
  - Un socle maitrisé, solide et fiable
  - Découverte du boson de Higgs en 2012 !

- Les évolutions au Run 2 se basent sur l'expérience gagné au Run 1
  - Utilisation plus efficace et agile des ressources (réseau)

- Intégration de nouvelles technologies et de ressources supplémentaires

- Un autre ingrédient : l'organisation
  - les opérations, l'implication des sites

Horizon à 10 ans

Où comment faire encore mieux ?

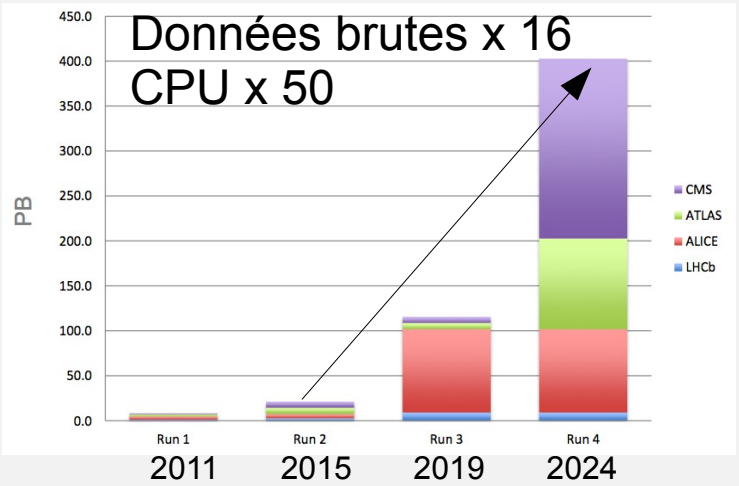
- De gigantesques besoins
- Dépasse une extrapolation simple à budget plat
  - Déjà le cas pour 2017

Run 3 : upgrade de ALICE et LHCb

- Calibration et reconstruction « en ligne »
- Ne plus faire de tri des données (event index LHCb)

Run 4 : upgrade de ATLAS et CMS

- 0.5 à 1 exabytes de données brutes / an
- Nous devons gagner un facteur 5-10
  - Modèles des expériences - quantité de données, algorithmes, calibration en ligne
  - Software – nouvelles architectures, mémoire --- HEP SW Fundation
  - Infrastructure – cpu/disk/tape/network, cloud privés, DC virtuel
- Repenser les modèles, «être inventifs »
- Etre en ligne avec les expériences « non-HEP»



Taille des lots de données brutes
ECFA High Lumi. LHC experiments workshop (2013)