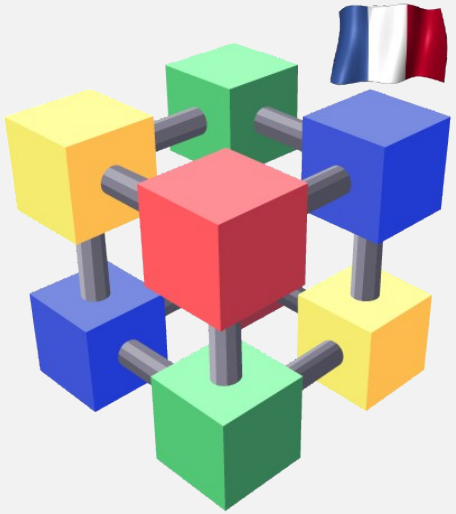
# Evolution du modèle de calcul des expériences LHC

Catherine Biscarat, LPSC/IN2P3/CNRS



Journées SUCCES 2015, 5-6 novembre, IPGP

# En guise d'introduction

- Présentation préparée avec l'aide de la communauté LCG-France
- Le fruit du travail d'un nombre de personnes incalculable

**Ce dont nous allons parler**

- Notre modèle original
- Les évolutions actuelles
- Un mot sur la suite

# Le contexte

# Contexte scientifique

Physique des particules

Physique hadronique
et physique nucléaire

## Les recherches du CEA
## en sciences de la matière et de l'Univers

**Les axes de recherche**

- Théorie et simulation
- Nanosciences
- Matériaux
- Physique nucléaire
- Physique des particules
- Astrophysique
- Instrumentation pour les grandes expériences de physique

cnrs dépasser les frontières

Rechercher ok

Le CNRS | Annuaires | Mots-Clefs CNRS | Autres sites

Institut national de **physique nucléaire** et de **physique des particules**

Centre national de la recherche scientifique

Présentation de l'Institut
Structures de recherche
Conseil scientifique
Infos aux laboratoires
Vie de la recherche
Carrières et emplois
Physique subatomique pour tous
English version

Accueil > Présentation de l'Institut > Politique scientifique

Statuts et missions
Axes stratégiques
Organisation et management
Questions scientifiques majeures

### Questions scientifiques majeures

Les problématiques scientifiques traitées à l'IN2P3 peuvent être résumées sous la forme des quelques questions scientifiques majeures.

**Y a-t-il une équation ultime des lois de la physique ?**

- D'où vient la masse des particules et donc de toute la matière dont nous sommes faits ?
- Quelle est la physique qui sous-tend la structure du Modèle standard composé de trois familles de particules élémentaires et de trois interactions ?
- Quelle est la nature et quelle est la masse du neutrino, cette particule insaisissable, au rôle encore inconnu dans la structure et l'évolution de l'Univers, et pourtant très répandue dans l'Univers ?
- Où est passée l'antimatière qui était présente au tout début de l'Univers ?

**Quelle structure pour la matière nucléaire ?**

- Comment les quarks sont-ils confinés dans les noyaux des atomes ?
- Comment se comporte la matière nucléaire aux confins de la stabilité ?

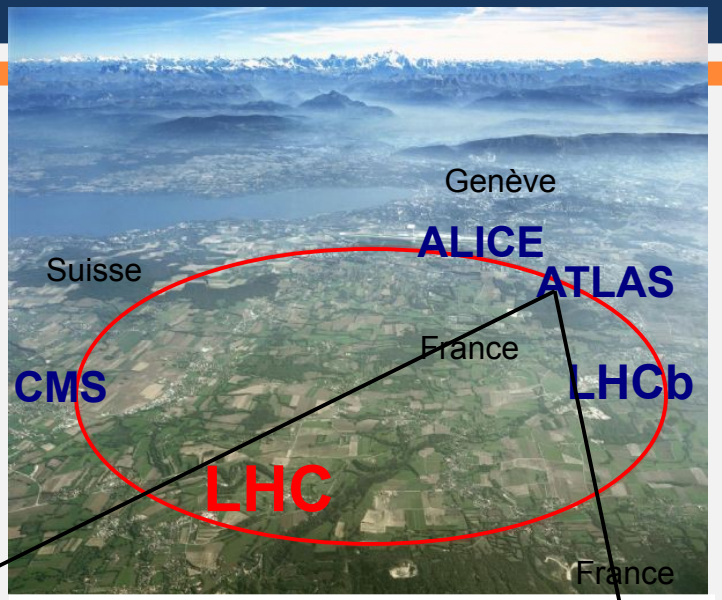
**De quoi est fait l'Univers et comment se comporte-t-il ?**

- Comment l'Univers s'est-il comporté dans le passé (quête des origines) ?
- Comment les éléments lourds se sont-ils formés dans l'Univers ?
- Qu'est-ce que la matière noire et l'énergie noire, cette part essentielle mais invisible de notre Univers ?
- D'où viennent les rayons cosmiques et quels sont leurs mécanismes de production et d'accélération ?

# L'appareillage

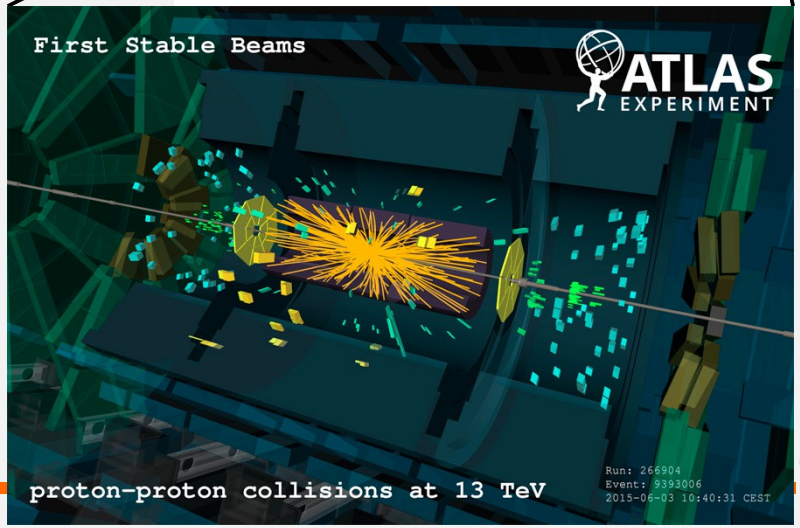
## Large Hadron Collider

- Nouvelle génération de collisionneur
  - Haute énergie, haute intensité
- Equipé de quatre détecteurs
- Début des opérations en 2010
- Détection de signaux rares  grande statistique




## Modèles de calcul

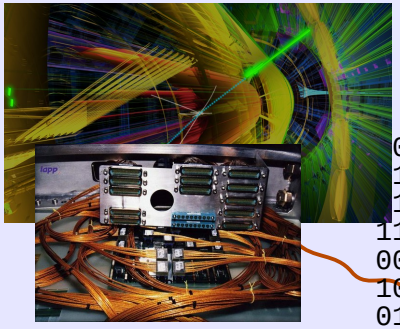
- “Classique” pour la physique sur collisionneurs
- Petits événements indépendants
  traitement séquentiel




Un événement dans ATLAS

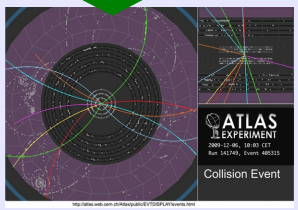
# Organisation du calcul

## Opérations centralisées

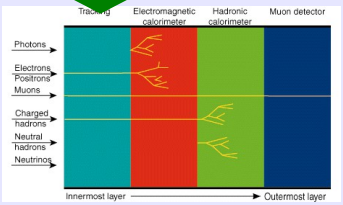


01100 010001
10111 001100
11100 100110
110101 110011
001010 001010
100101 000011
010111 010100

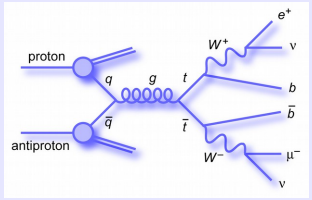
### Reconstruction



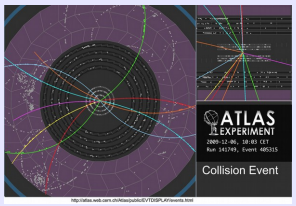

### Analyse

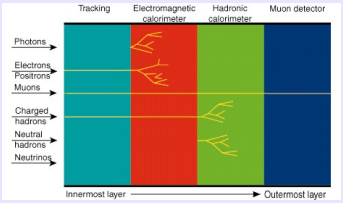



### Simulation



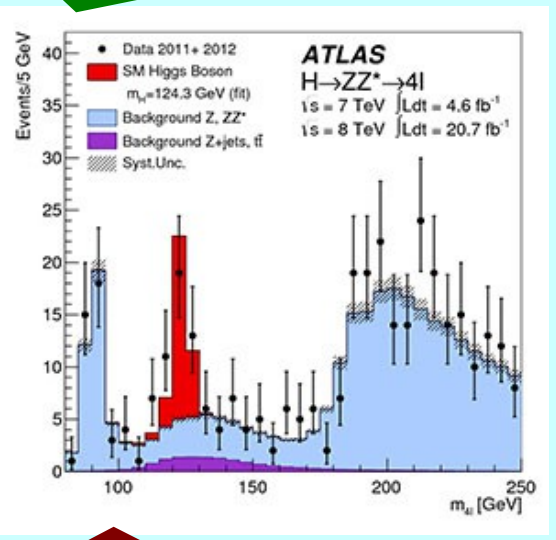

101100 010001
110111 001100
111100 100110
110101 110011
001010 001010
100101 000011
010111 010100




## Individuel (chaotique)

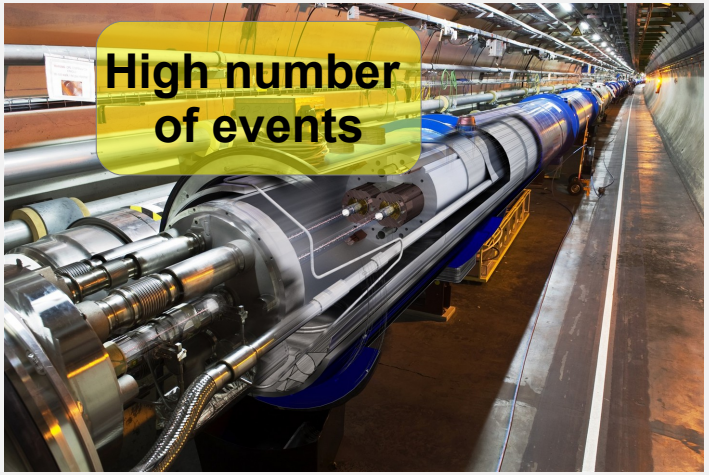
### Final analysis (n times / day)

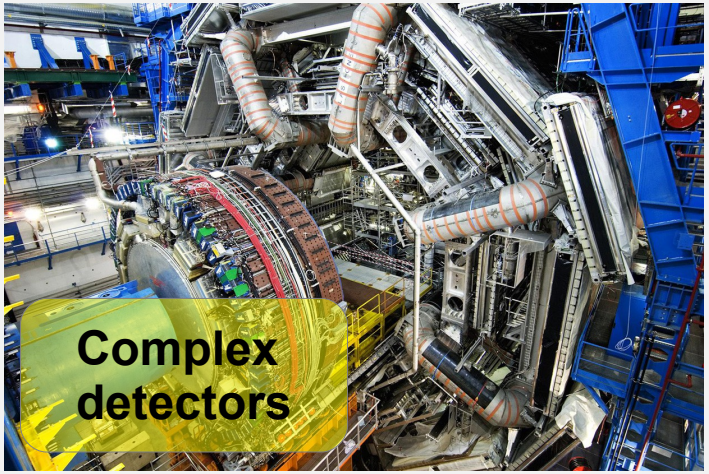


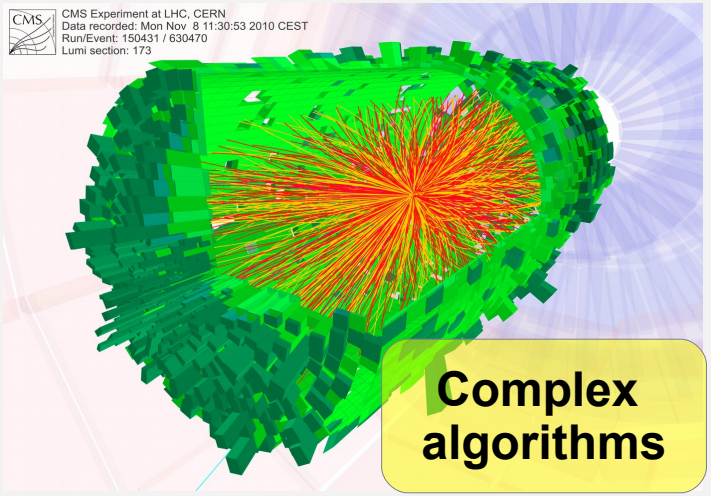
# Un nouvel ordre de grandeur

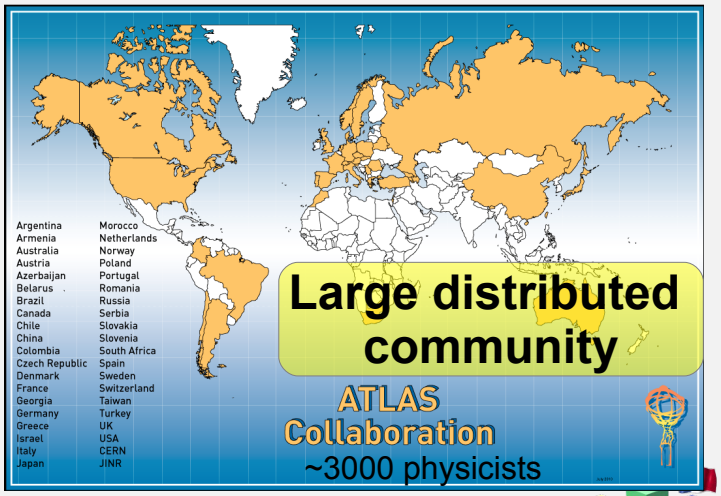
Contraintes :
Large volume de données
Ressources (CPU+stockage)
Milliers utilisateurs finaux
Archivage à long terme

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



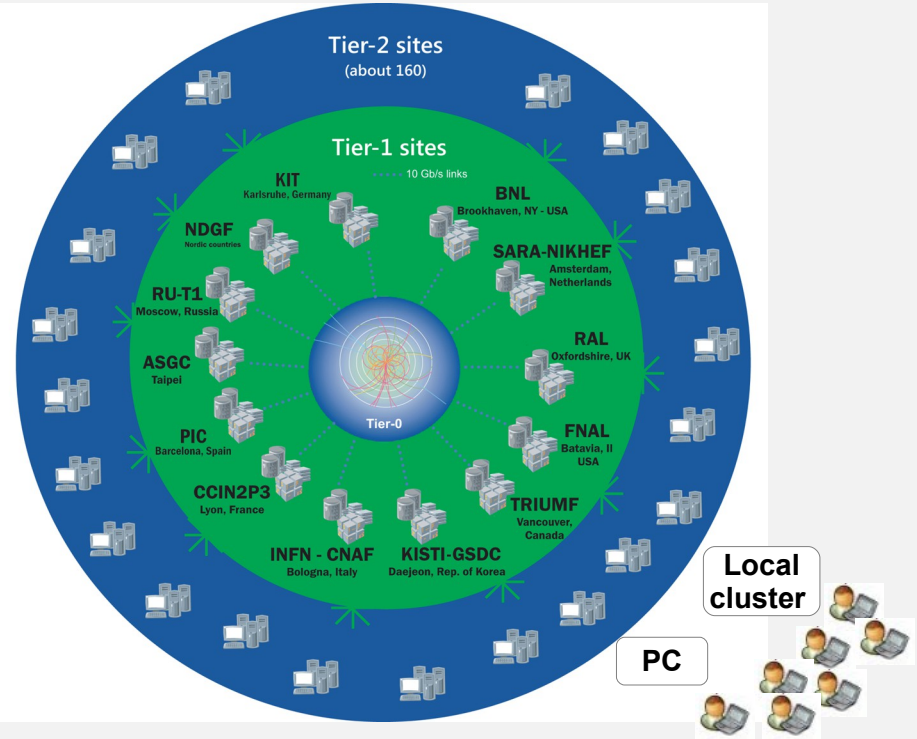



# Hiérarchie des sites / modèle MONARC

**Premier modèle pour l'informatique au LHC (1999)**

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




**Tier-0 (CERN):**

- Raw data storage
- Calibration
- Initial reco
- 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

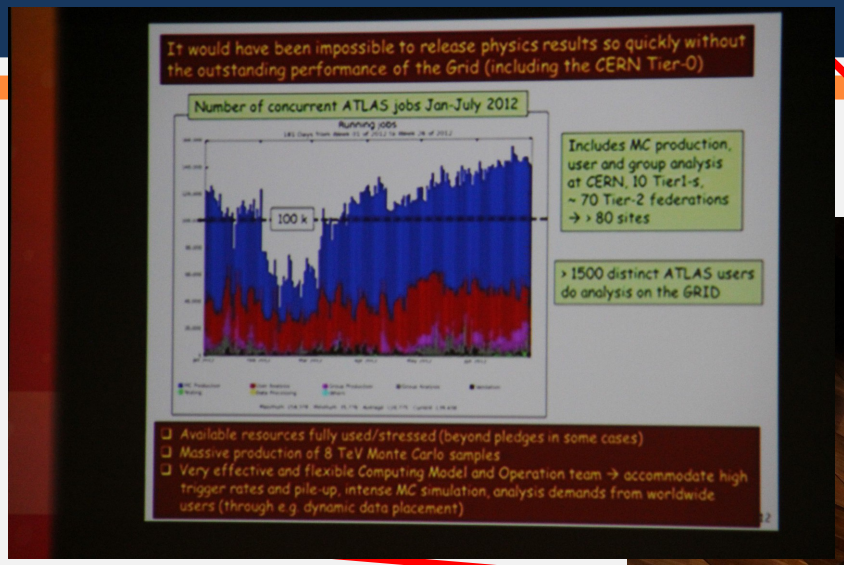
**Memorandum of Understanding (T0/1/2)**
**Contraintes de fiabilité et de disponibilité**

**170 sites, 40 pays, 8k utilisateurs**

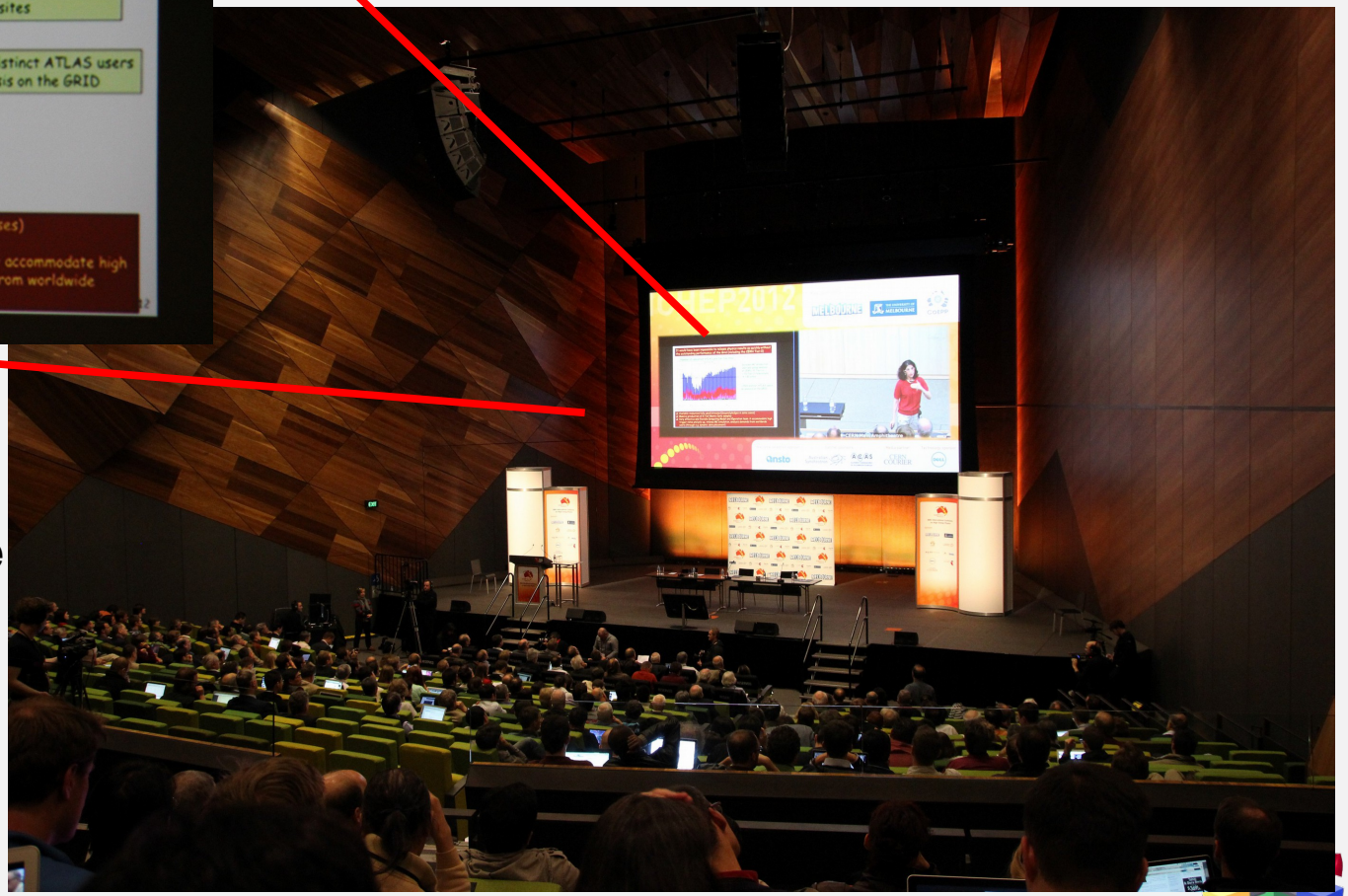
**500 PB,**
**> 2 millions jobs/jour,**
**~350 000 coeurs**


Snapshot 30/10/15

# « computing enables physics »




Photography: C. Biscarat



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

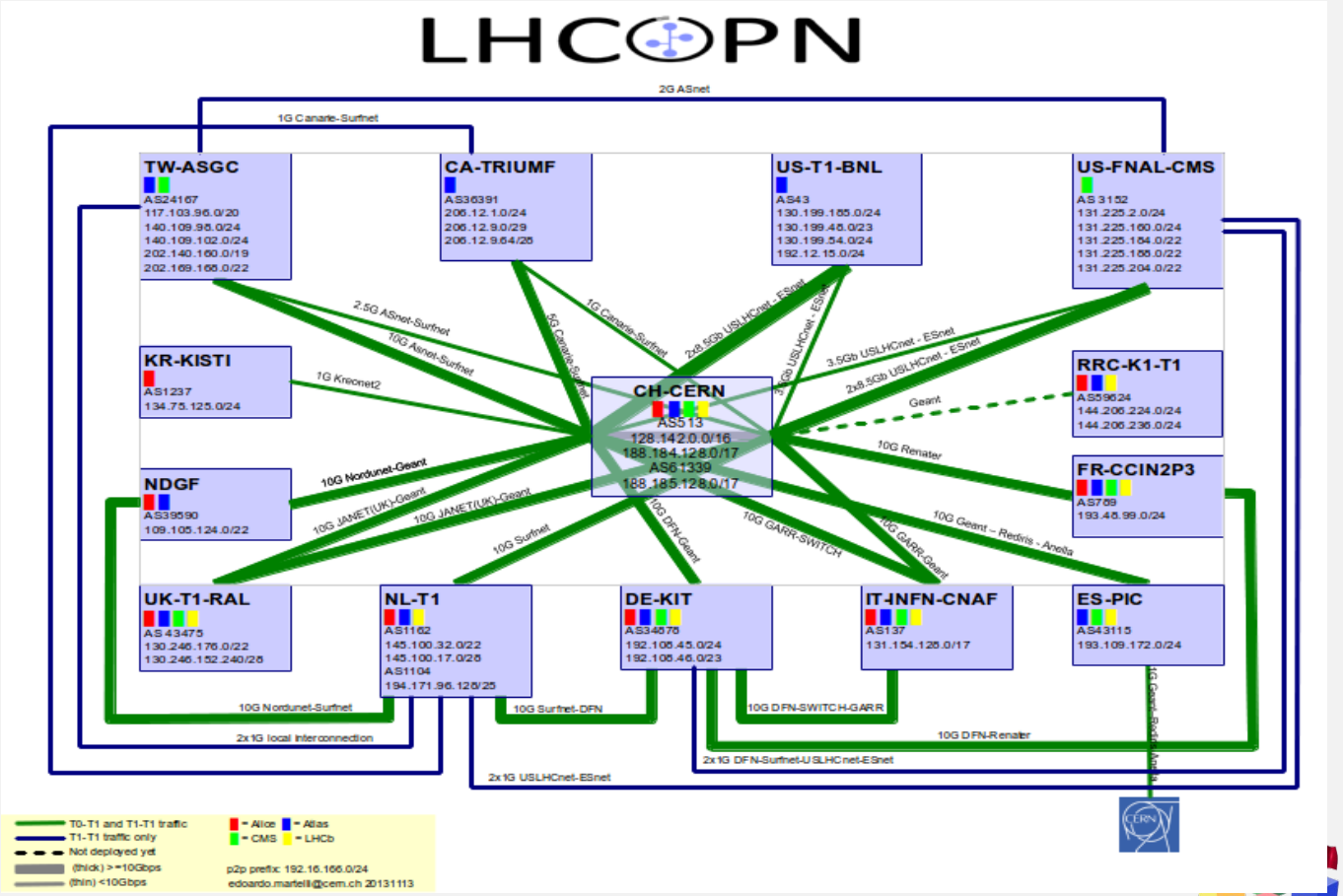
CERN seminar, July 4th 2012,
retransmitted at ICHEP (Melbourne)

# Le réseau – socle de notre modèle

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

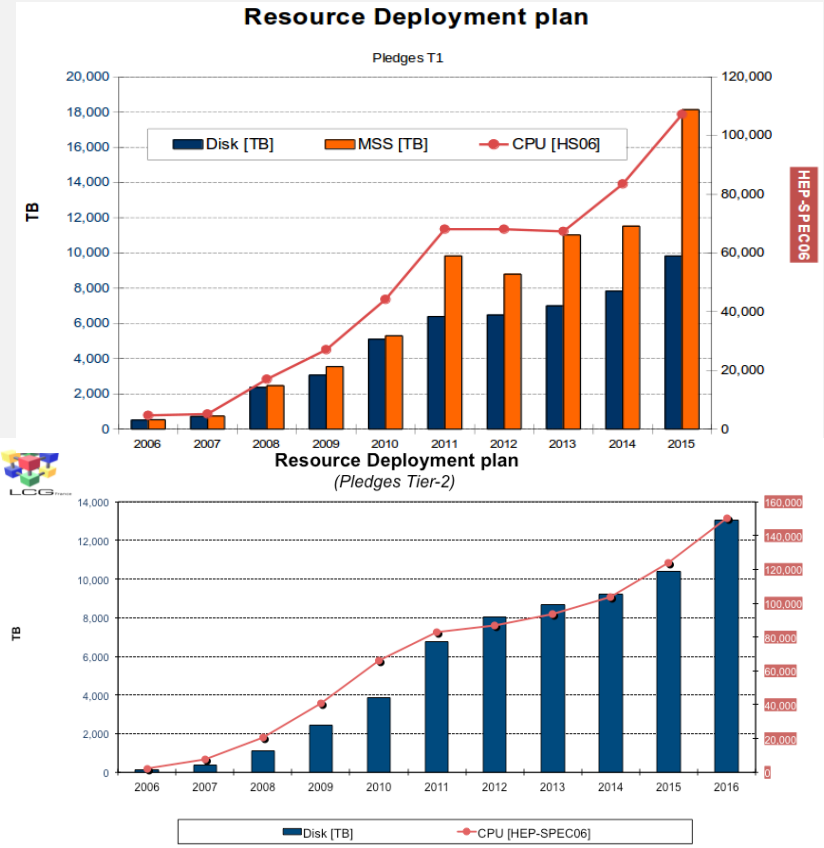
- Optical private network
- Liens dédiés et redondants
- T0-T1 et  T1-T1

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

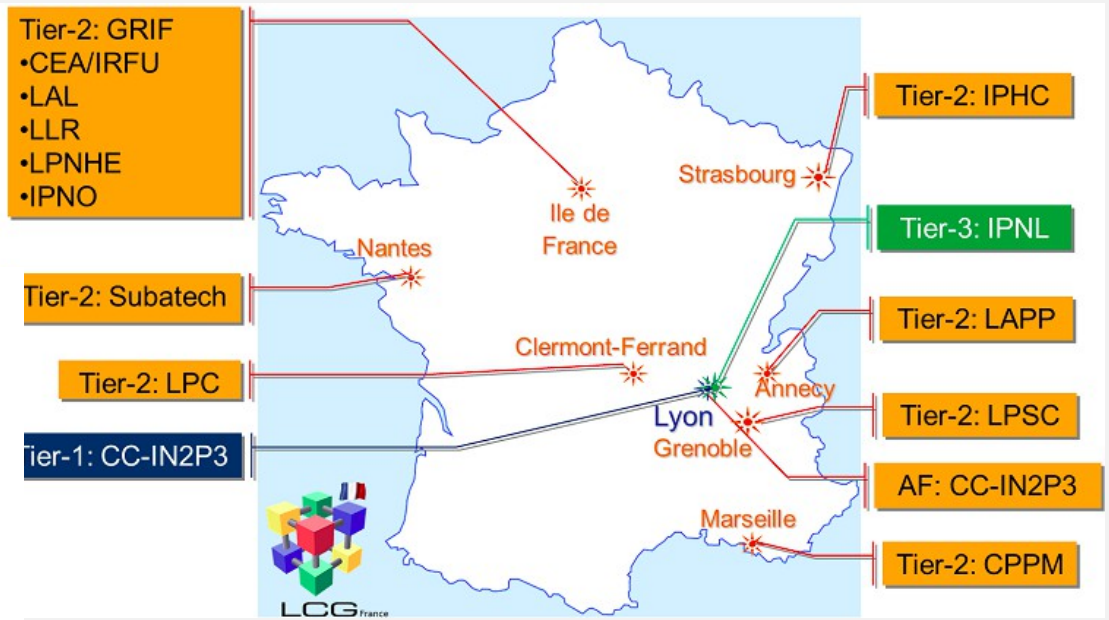


# Les sites en France

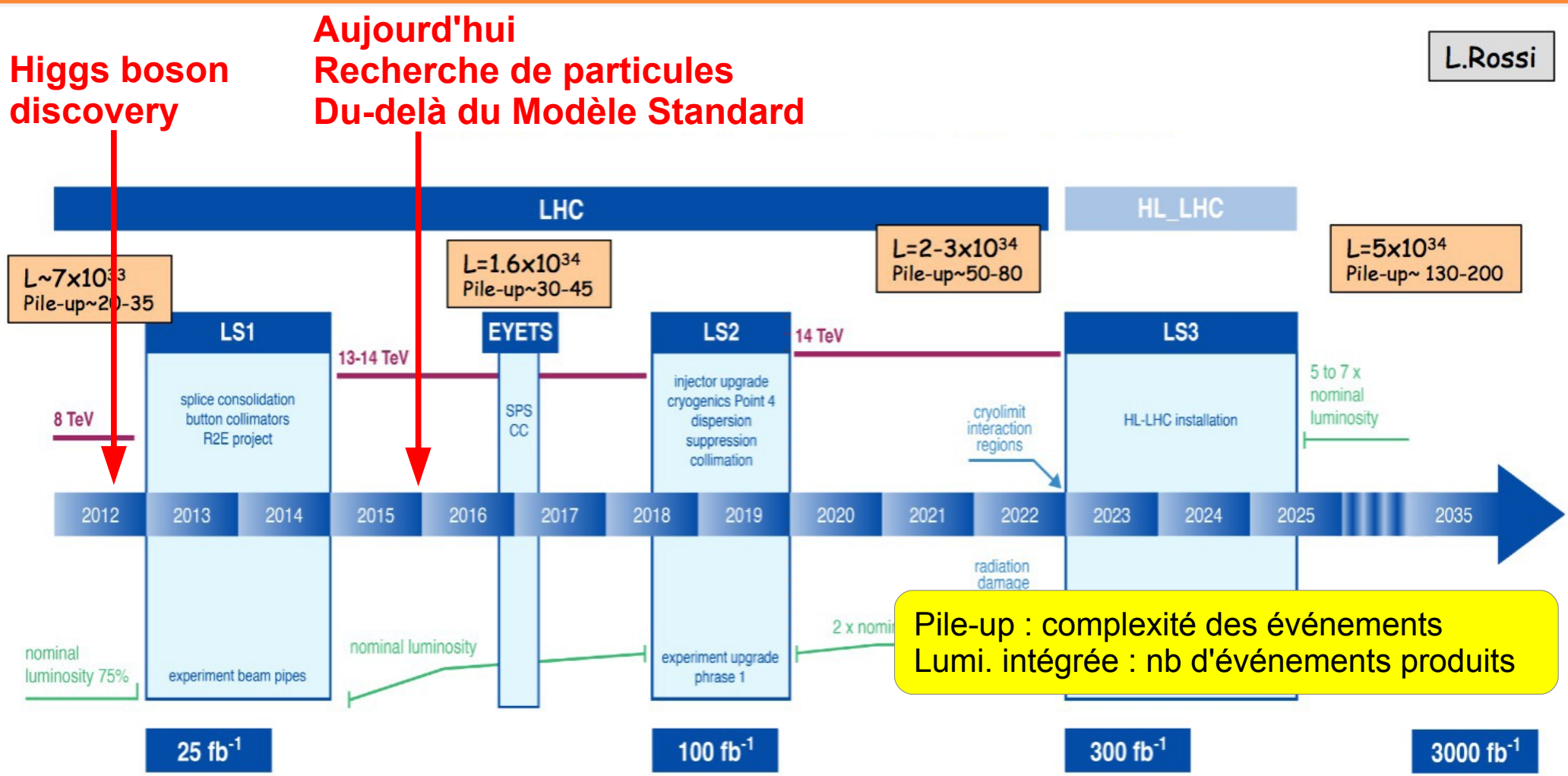
- Organisés avec les expériences dans le projet « LCG-France » - CNRS/IN2P3 et CEA/IRFU
- Fournir ~10% des ressources informatiques mondiales aux expériences LHC (MoU, T1+T2)




En 2015 au T1 : 18 PB bande, 10 PB disk, 110 kHS06
Les T2 doublent les ressouces disk et CPU.



# Temps forts du LHC



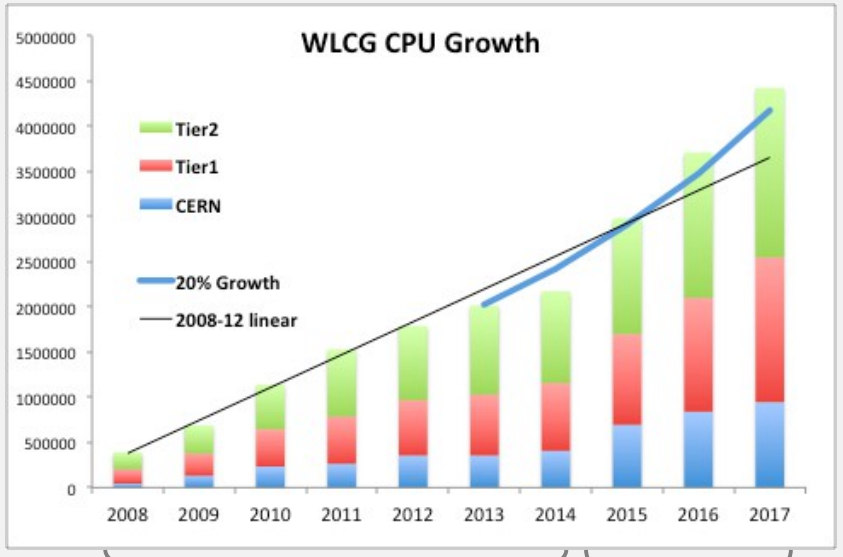

Pile-up : complexité des événements
Lumi. intégrée : nb d'événements produits

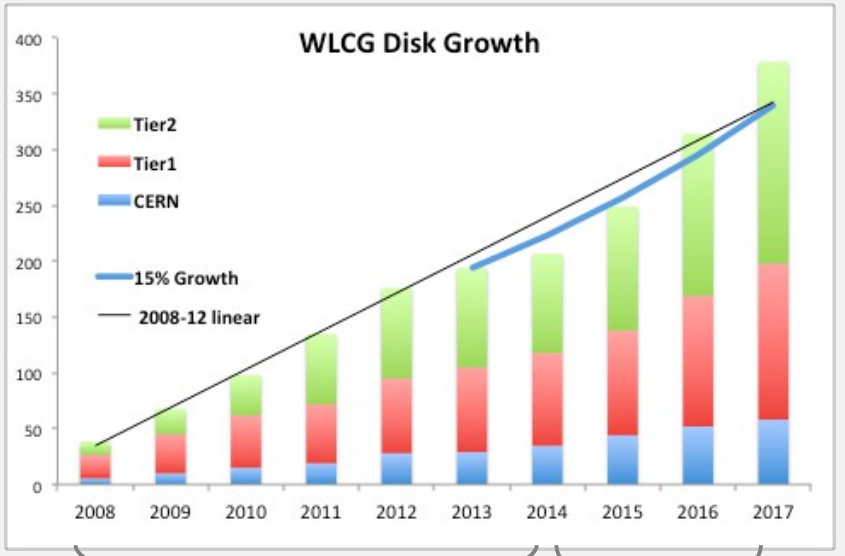
# Les évolutions aujourd'hui

# 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



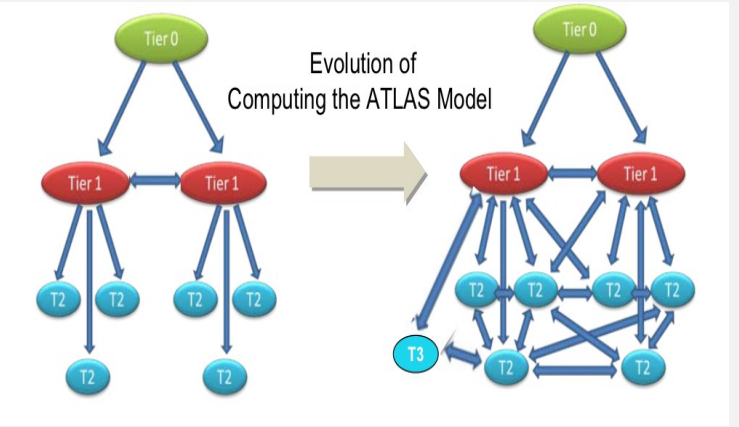



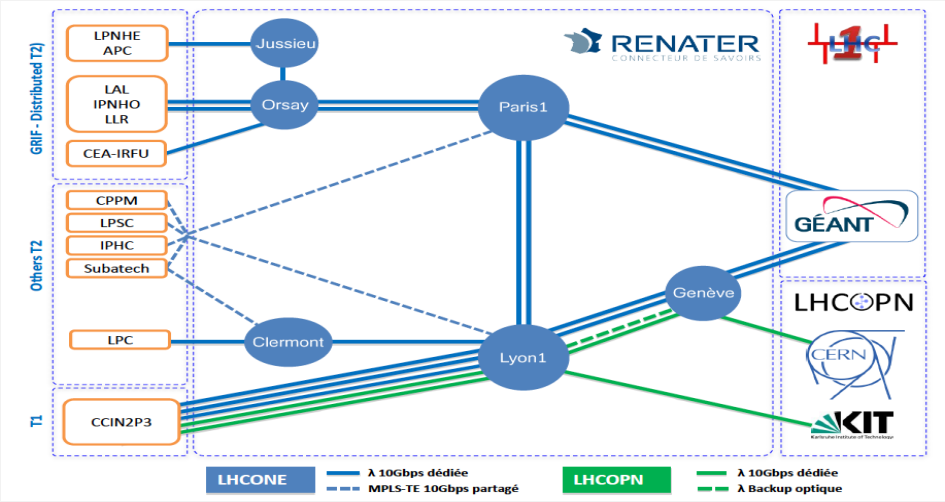

Courbes : croissance annuelle estimée à budget constant (CPU : 20% ; Disk : 15%)

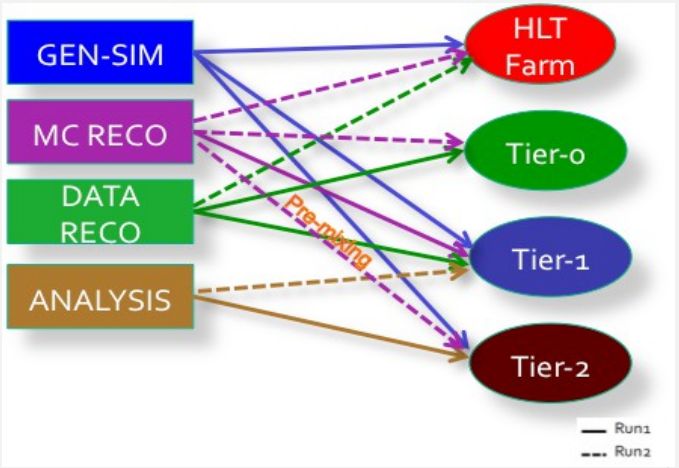
# Un modèle plus souple

## Bénéfices du réseau

- Le modèle MONARC (réplication et pré-placement des données) 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) pour tous les sites – 70 sites connectés



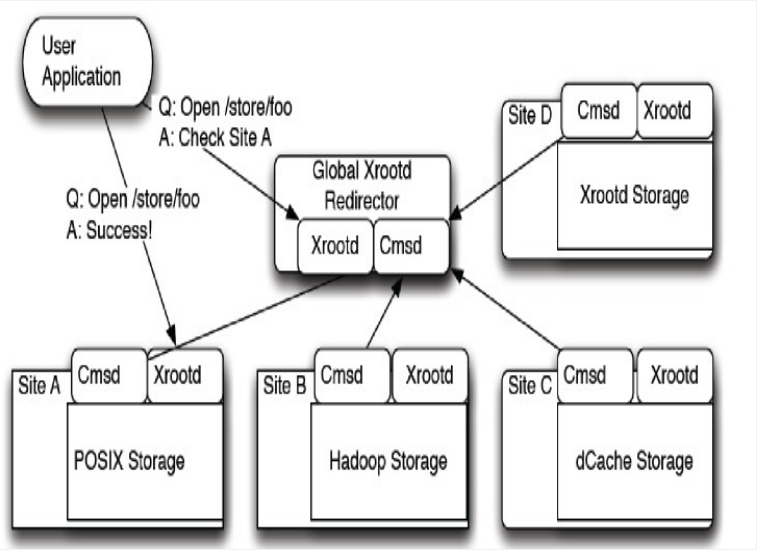
# Gestion et accès aux données

## Gestion dynamique des données

- Examen de la « popularité des données »
- création et effacement dynamique

## Accès transparents aux données

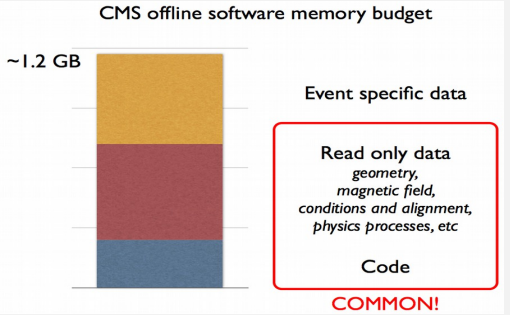
- Fédération de données
- En production routinière :  fall-back
  - récupération des données si l'accès local échoue.
- Exploratoire : accès distant aux données



# Les flots de traitement

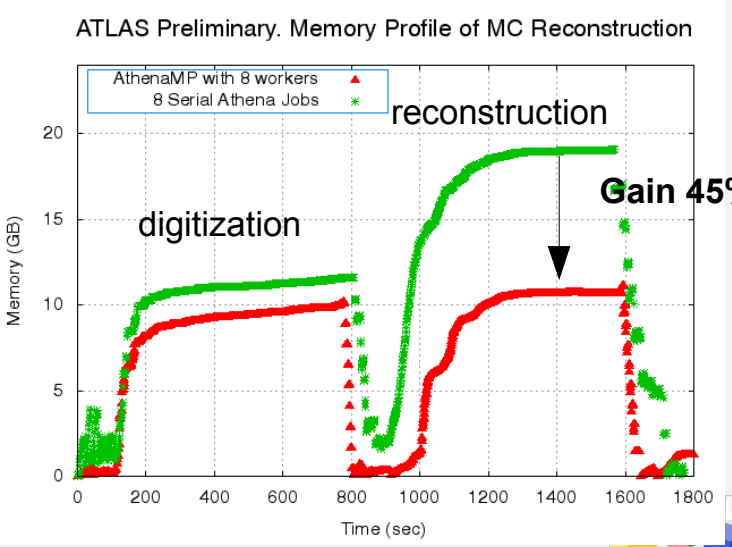
## Des améliorations continues

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



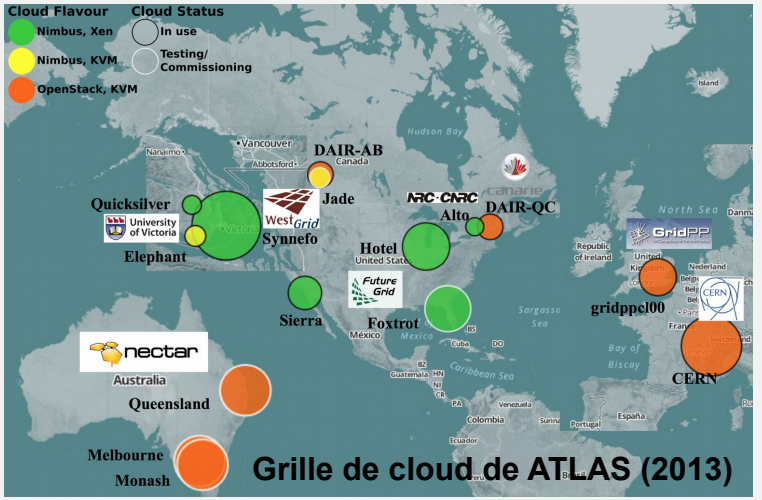

## Utilisation optimale du parc de CPU

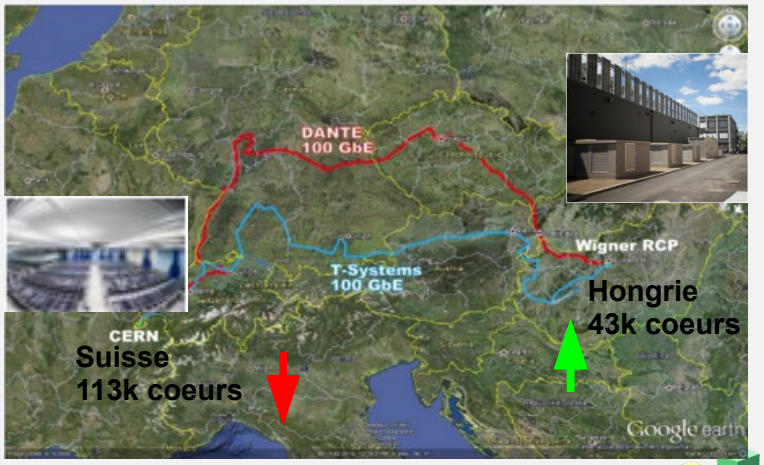
- Constructeurs : moins de mémoire par cœur
  - Augmentation des besoins avec la luminosité
  - Parallélisation des codes (nouveau en HEP)
    - fork des événements



# Des sites naissent « clouds »

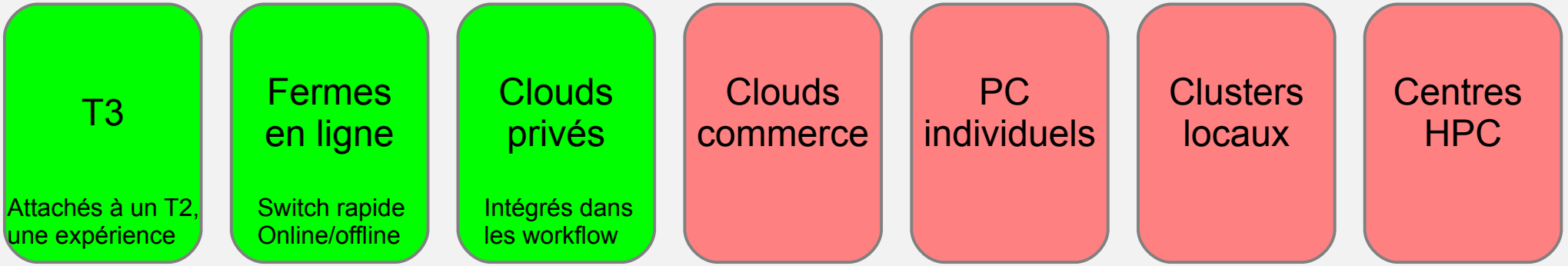
- Intégrer aux workflow des expériences
- Deux modèles principaux :
  - La VO instancie les VM (image de la grille - pilotes)
  - Le site crée lui-même les VM des expériences (modèle Vacuum) qui demandent elles-même leurs tâches
- Le Tier-0 s'est doté d'une annexe
  - Les ressources sont orchestrées dans un cloud privé (extension dynamique du Tier-0 historique au CERN)
  - Aujourd'hui : 4600 HV, 125 000 coeurs



# Une avalanche de données

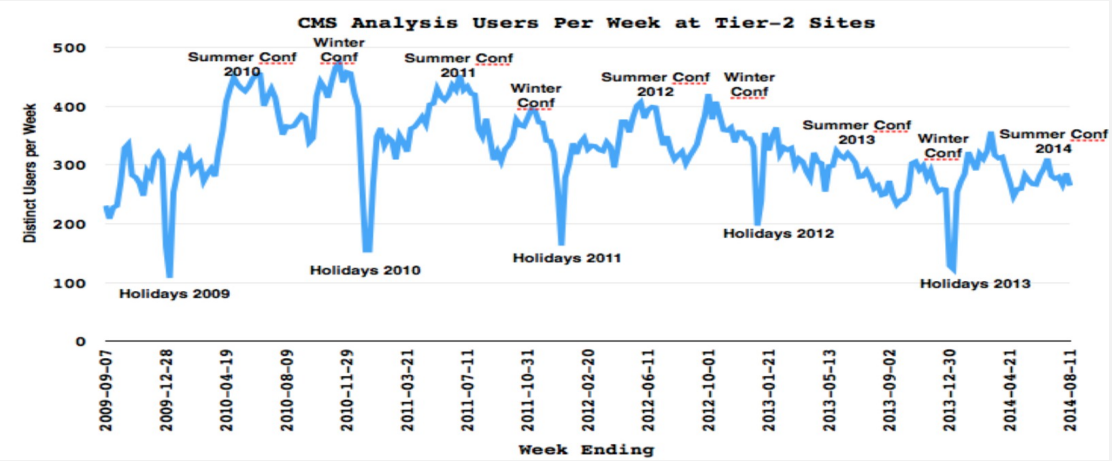
- Intégration de plateformes de **ressources « opportunistes »** (hors des « MoU »)
  - Le stockage n'est pas opportuniste




- Tâches préférées : génération et simulation (CPU intensif, pas de connexion aux DB)
- Portabilité : utilisation de la virtualisation (cernVM)
- Optimisation : tâches qui peuvent être interrompues (Event Service – ATLAS)
- Services : simplification, utilisation de standards

# Les *clouds* commerciaux

- L'activité est variable dans le temps




## Elasticité

- Caractéristique des *clouds* commerciaux - « pay as you go »
- Comprendre et maitriser les coûts
  - Les données, le réseau, les services
- Des expériences avec Google, HELIXNebula, Amazon Web Service

- En Europe, HNSciCloud (démarrage 2016)
  - Prototype de *cloud* public/privé
  - Comprendre les coûts et l'utilisation possible par les expériences LHC et d'autres communautés

# Volunteer computing

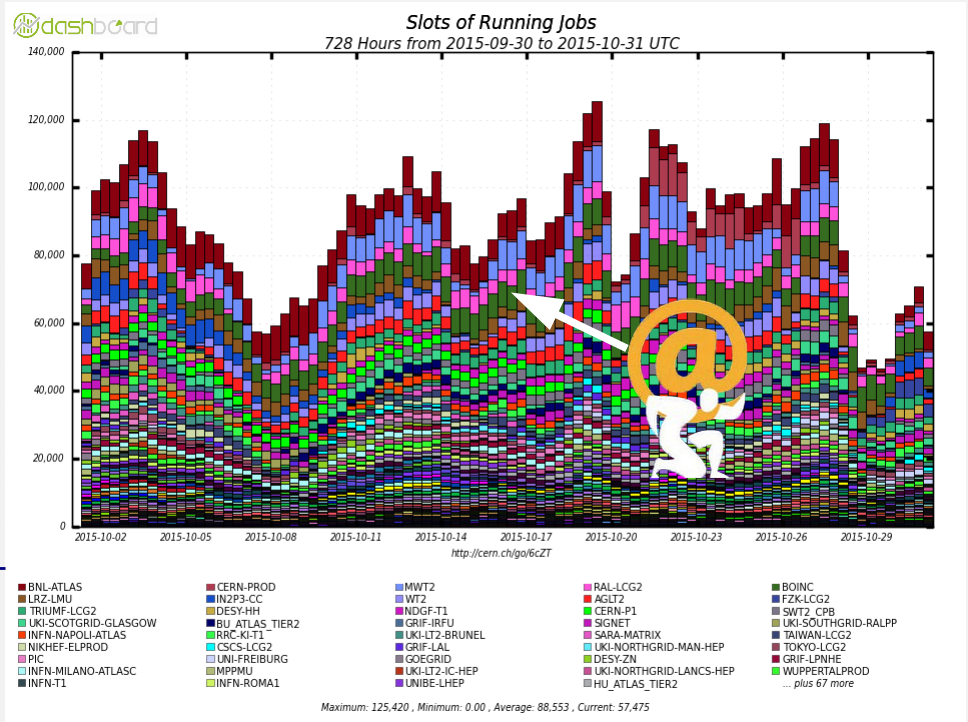
## BOINC et les projets @home

- Sur le modèle de SETI@home
  - Inclut la virtualisation
- Des volontaires installent BOINC sur leur PC personnel et donnent des cycles de calcul aux projets de leur choix

- La première tentative : ATLAS@home
  - Une énorme source de CPU

- Une solution pour intégrer les clusters hors grille

http://atlasathome.cern.ch/




Un des “sites” majeurs pour la simulation
7000 tâches en moyenne en octobre 2015

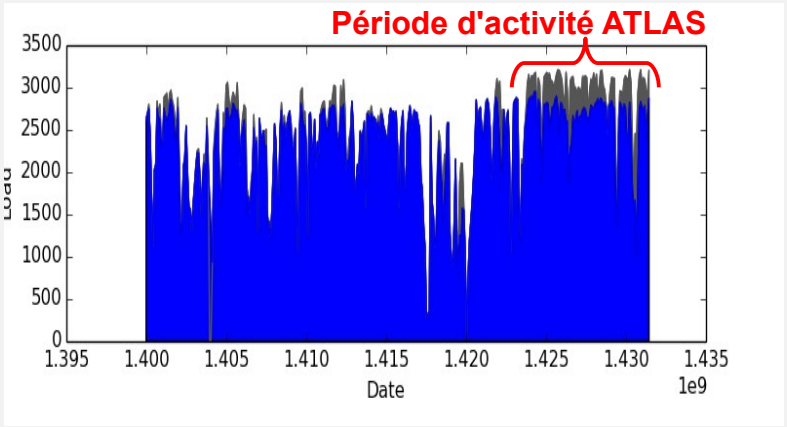
# Les centres HPC

## HTC

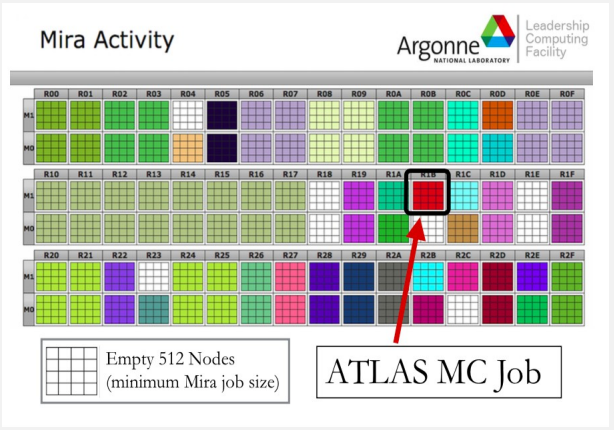
- En HEP : traitement d'événements séquentiels (Hight Througput Computing)
- Sans besoin des spécificités du HPC

## HPC

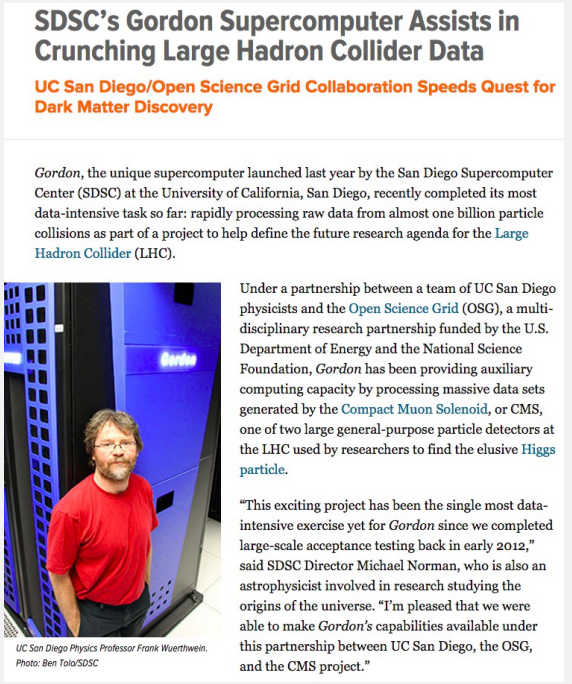
- Des collaborations avec de grands centres HPC (allocation par grant)
- Des initiatives locales/régionales (cycles vides)




Génération DIPHOX, FROGGY (CIGRI), LPSC/ATLAS




Génération ALPGEN, MPI, ATLAS



SDSC's Gordon Supercomputer Assists in Crunching Large Hadron Collider Data

UC San Diego/Open Science Grid Collaboration Speeds Quest for Dark Matter Discovery

*Gordon*, the unique supercomputer launched last year by the San Diego Supercomputer Center (SDSC) at the University of California, San Diego, recently completed its most data-intensive task so far: rapidly processing raw data from almost one billion particle collisions as part of a project to help define the future research agenda for the Large Hadron Collider (LHC).

UC San Diego Physics Professor Frank Wuerthwein. Photo: Ben Tolo/SDSC

Under a partnership between a team of UC San Diego physicists and the Open Science Grid (OSG), a multi-disciplinary research partnership funded by the U.S. Department of Energy and the National Science Foundation, *Gordon* has been providing auxiliary computing capacity by processing massive data sets generated by the Compact Muon Solenoid, or CMS, one of two large general-purpose particle detectors at the LHC used by researchers to find the elusive Higgs particle.

"This exciting project has been the single most data-intensive exercise yet for *Gordon* since we completed large-scale acceptance testing back in early 2012," said SDSC Director Michael Norman, who is also an astrophysicist involved in research studying the origins of the universe. "I'm pleased that we were able to make *Gordon's* capabilities available under this partnership between UC San Diego, the OSG, and the CMS project."

Re-processing de CMS

# Pour finir

# En guise de conclusion

- 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

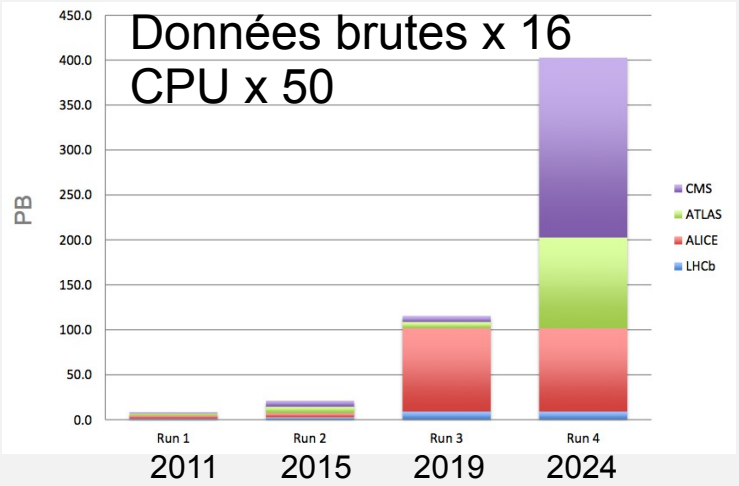
# Horizon à 10 ans

## Run 3 et 4 – où comment faire encore plus ?

- Les besoins sont gigantesques
- Dépasse une extrapolation simple à budget plat

- Il faut repenser les modèles
- S'adapter aux nouvelles architectures et revoir les logiciels

- WLCG Technical Forum
- HEP Software Fundation
  - http://hepsoftwarefoundation.org/



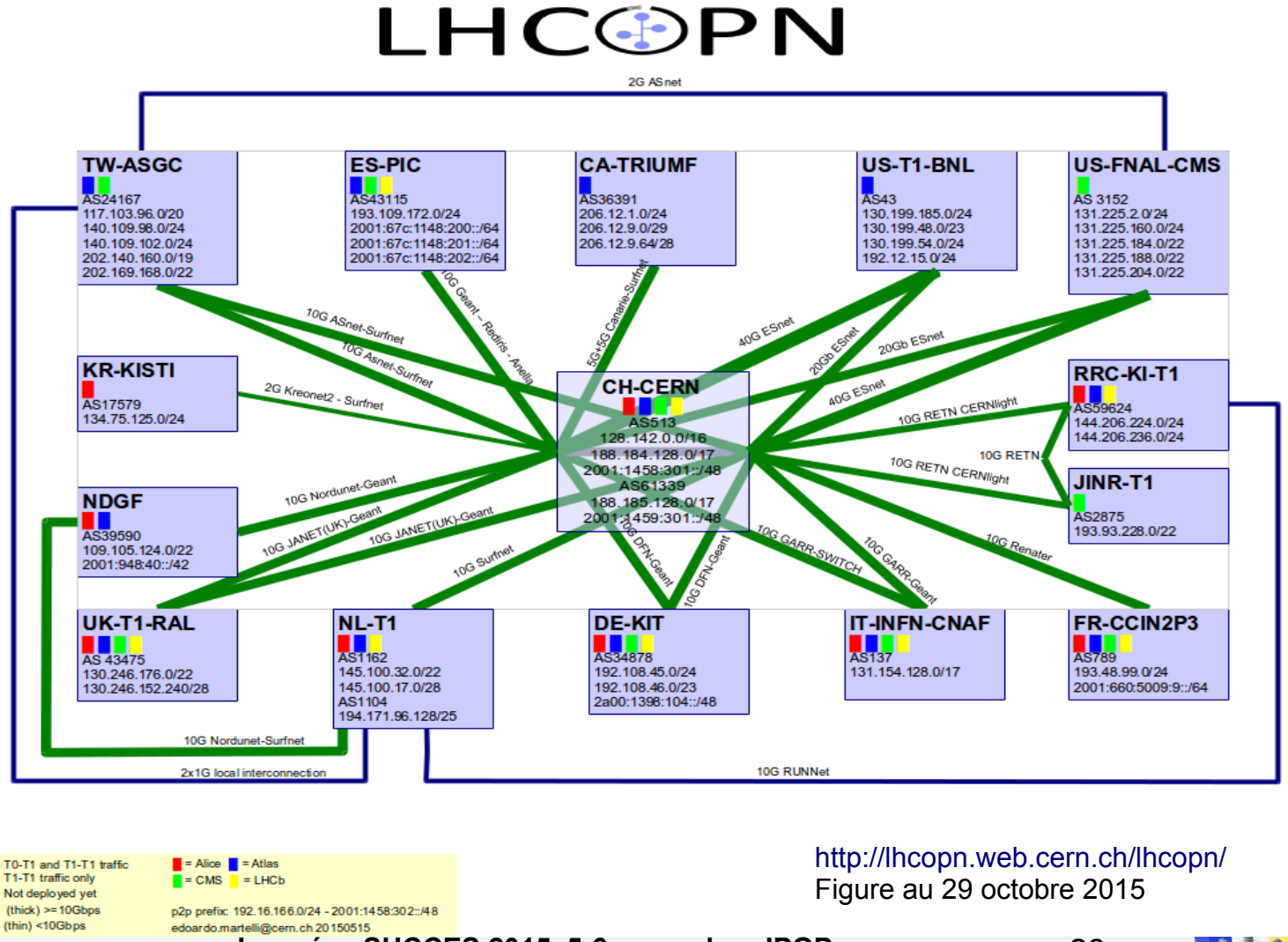

Taille des lots de données brutes

# Diapositives supplémentaires

# Le réseau – socle de notre modèle

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

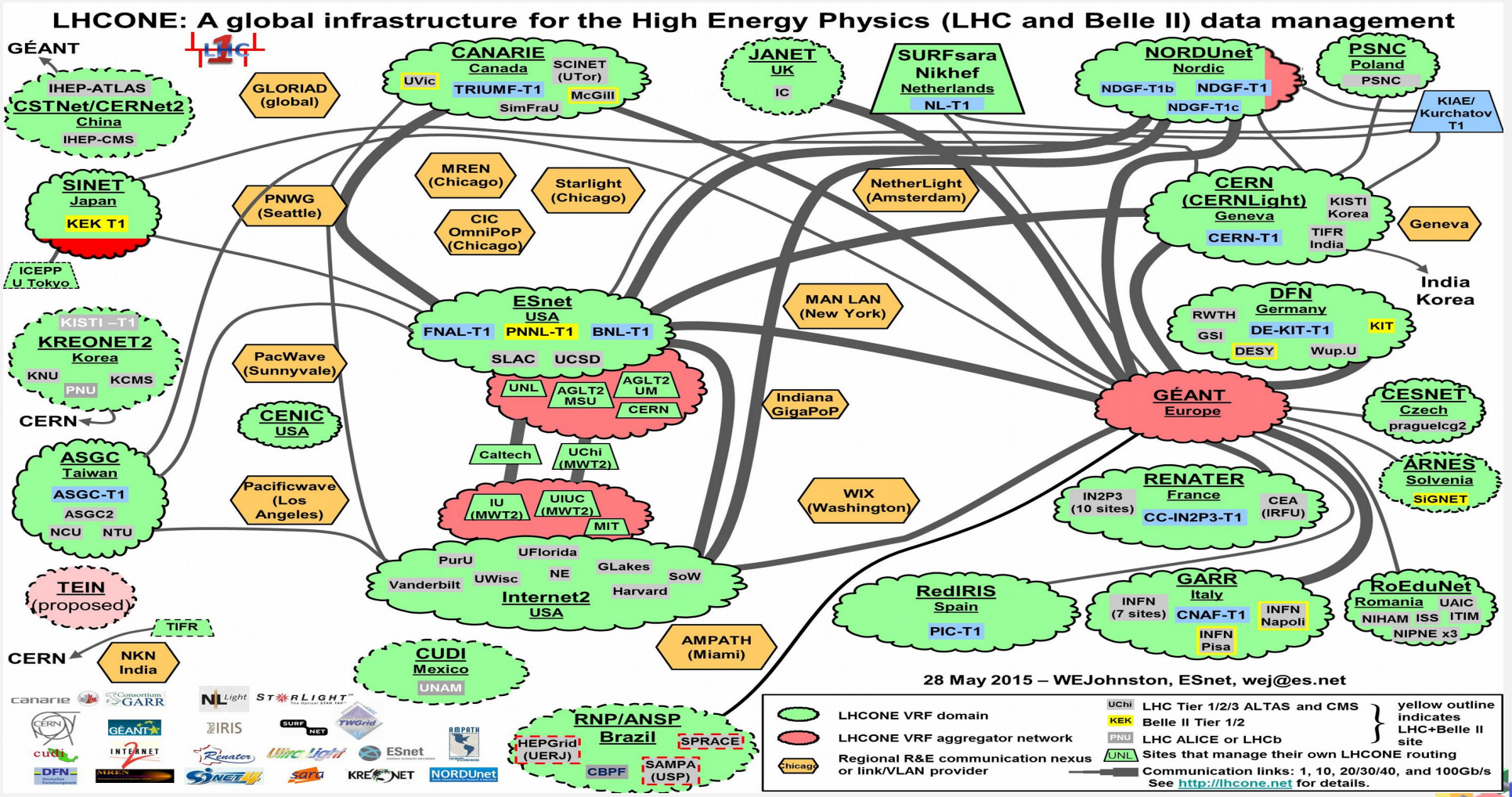
NB : certains liens T1-T1 ont été mis sur un autre infrastructure LHCONE




http://lhcopn.web.cern.ch/lhcopn/
Figure au 29 octobre 2015

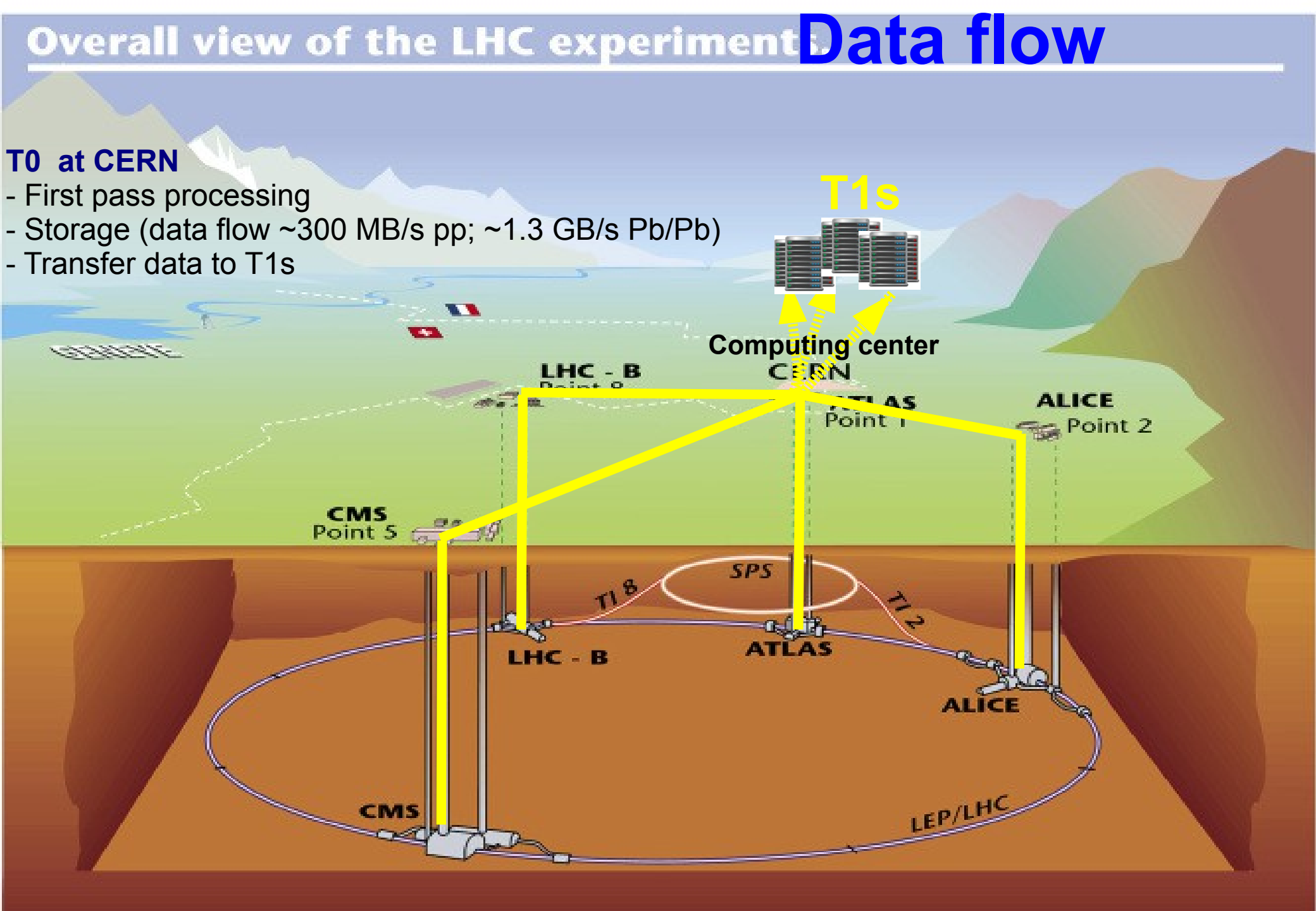
# Evolutions du réseau - LHC-ONE

http://lhcone.web.cern.ch/

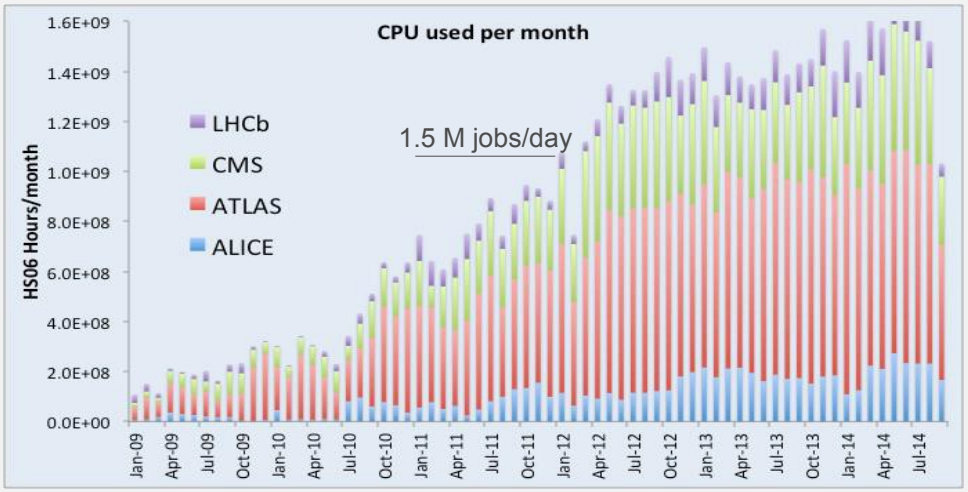


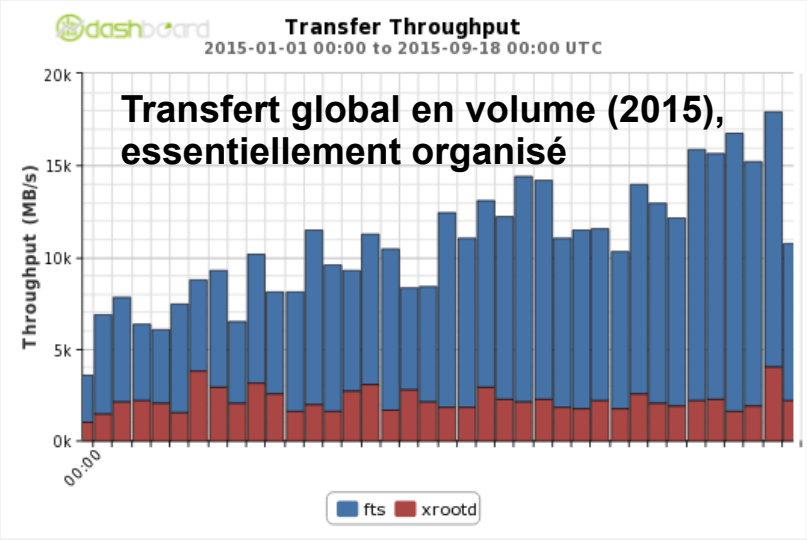
# Overall view of the LHC experiments. Data flow

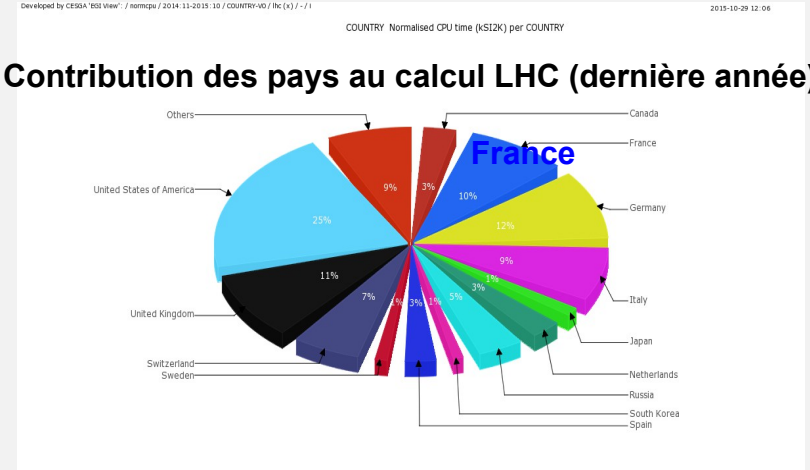
**T0 at CERN**
- First pass processing
- Storage (data flow ~300 MB/s pp; ~1.3 GB/s Pb/Pb)
- Transfer data to T1s



# Clin d'oeil sur l'utilisation des ressources

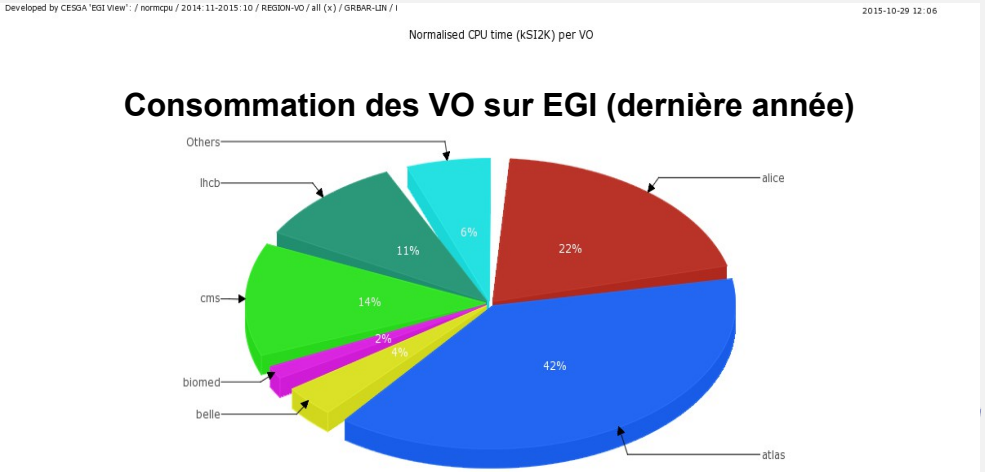


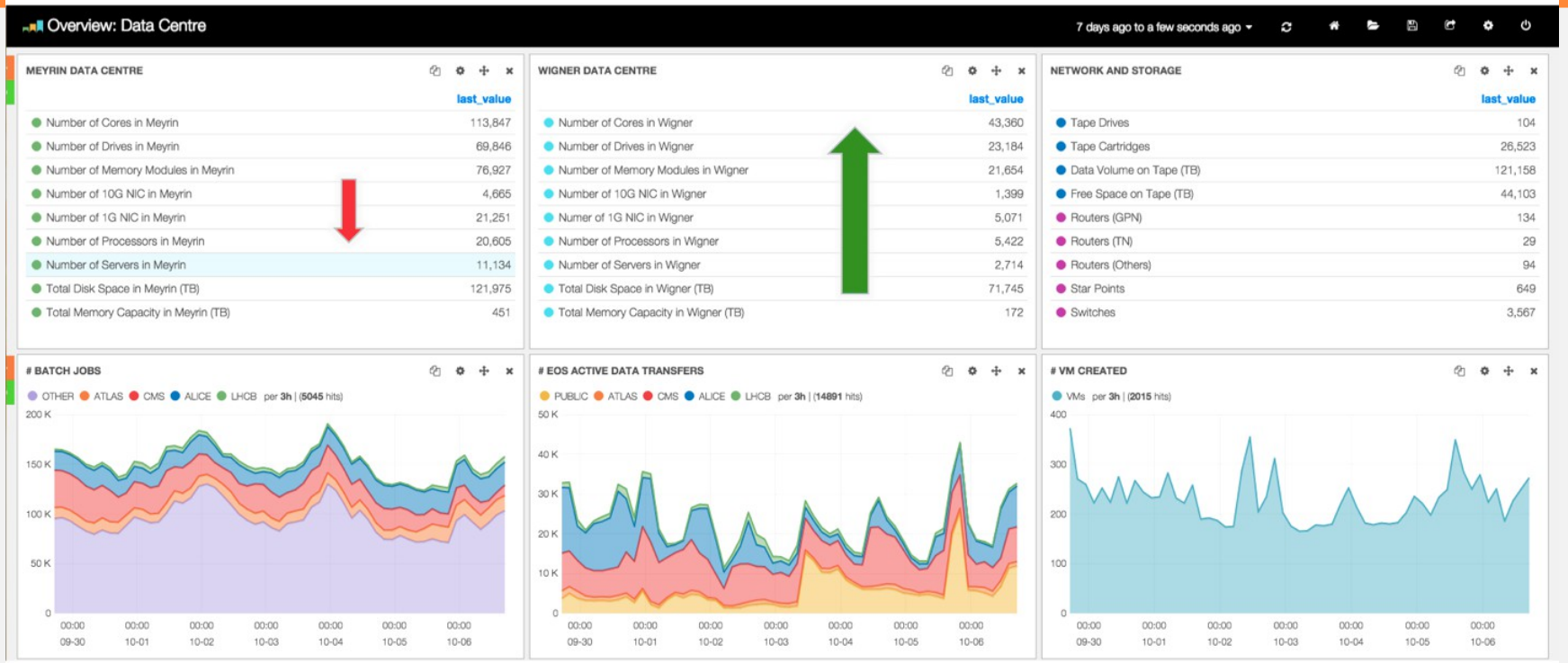

**Contribution des pays au calcul LHC (dernière année)**



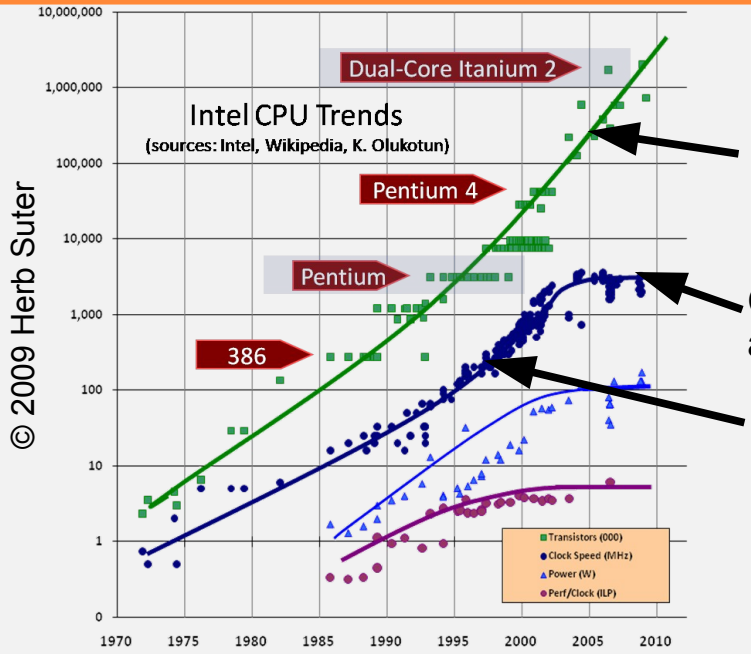

**Consommation des VO sur EGI (dernière année)**

# The CERN DCs in Numbers




More details at https://meter.cern.ch

# The free lunch is over




Ajout de transistors dans les CPU
Memory per core per IO is limited

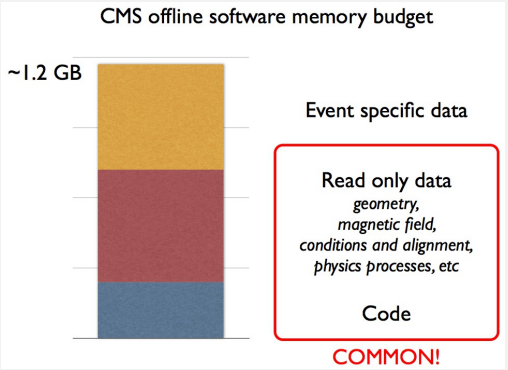
CPU clock reaches a plateau

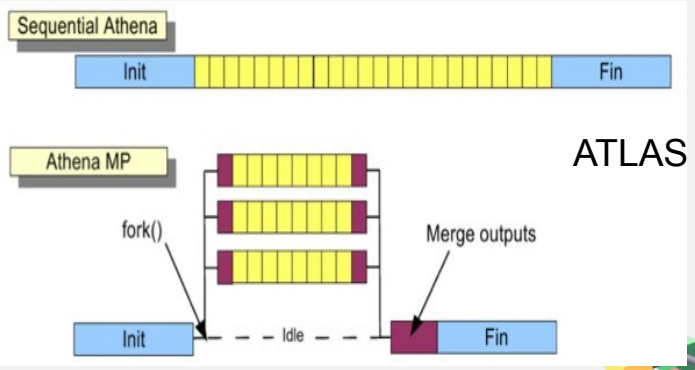
Fréquence x 2 chaque année

- En HEP : petits événements indépendants
- Augmentation de la luminosité, pile-up
- → augmentation de la mémoire

**Besoin d'introduire du parallélisme**

- Fork des événements
- Mise en commun d'une grande partie de la mémoire

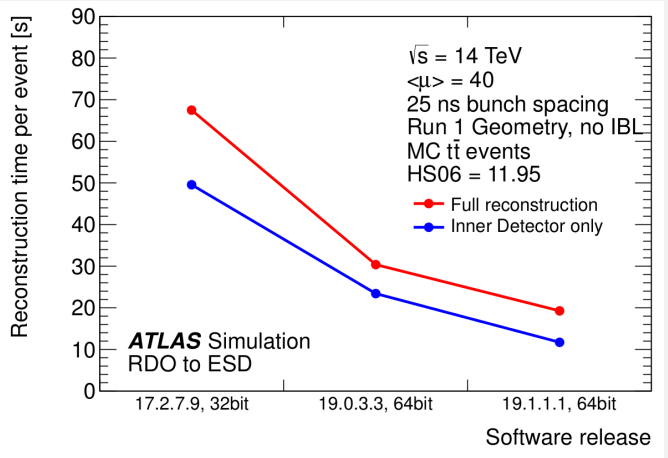


# Le défi logiciel

## Optimisation du software

- Exemple ATLAS : reconstruction 3 x plus rapide



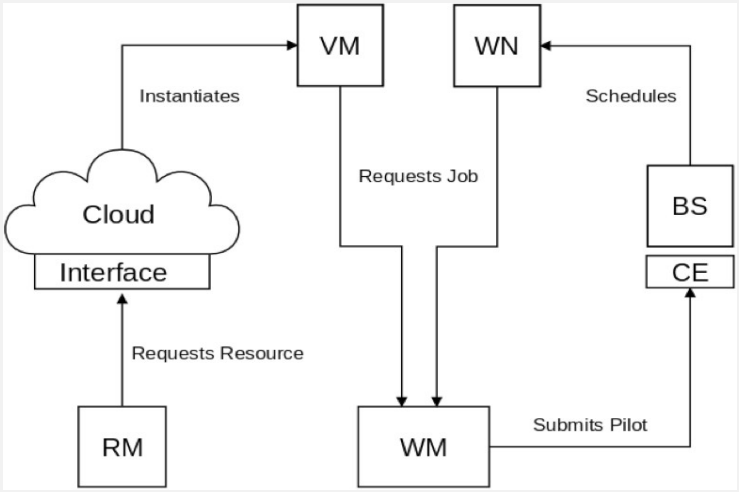

## Aller plus loin : un challenge pour le HEP

- Exploiter efficacement les nouvelles architectures (CPU modernes, GPU, ...)
- Parallélisation à tous les niveaux (algorithmes), revoir la structure des données, …
- Effort colossal de ré-écriture, nouvelles compétences à acquérir

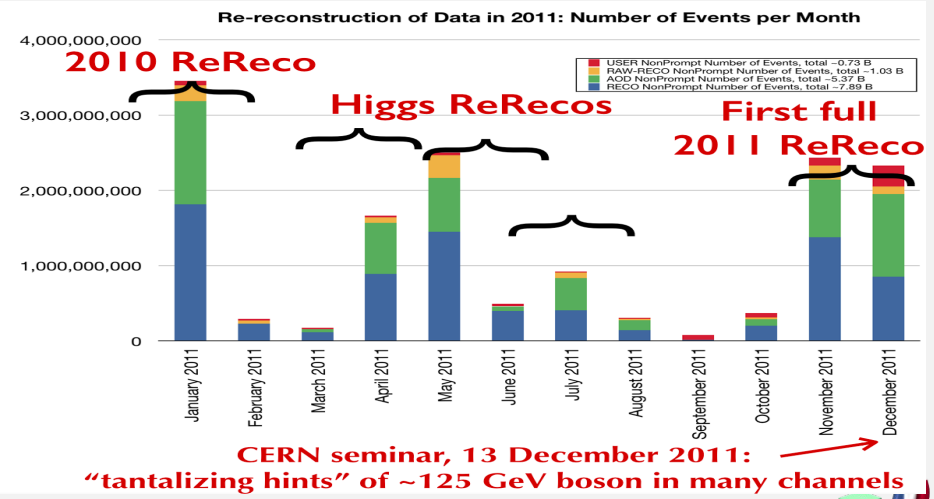
## *HEP software foundation*

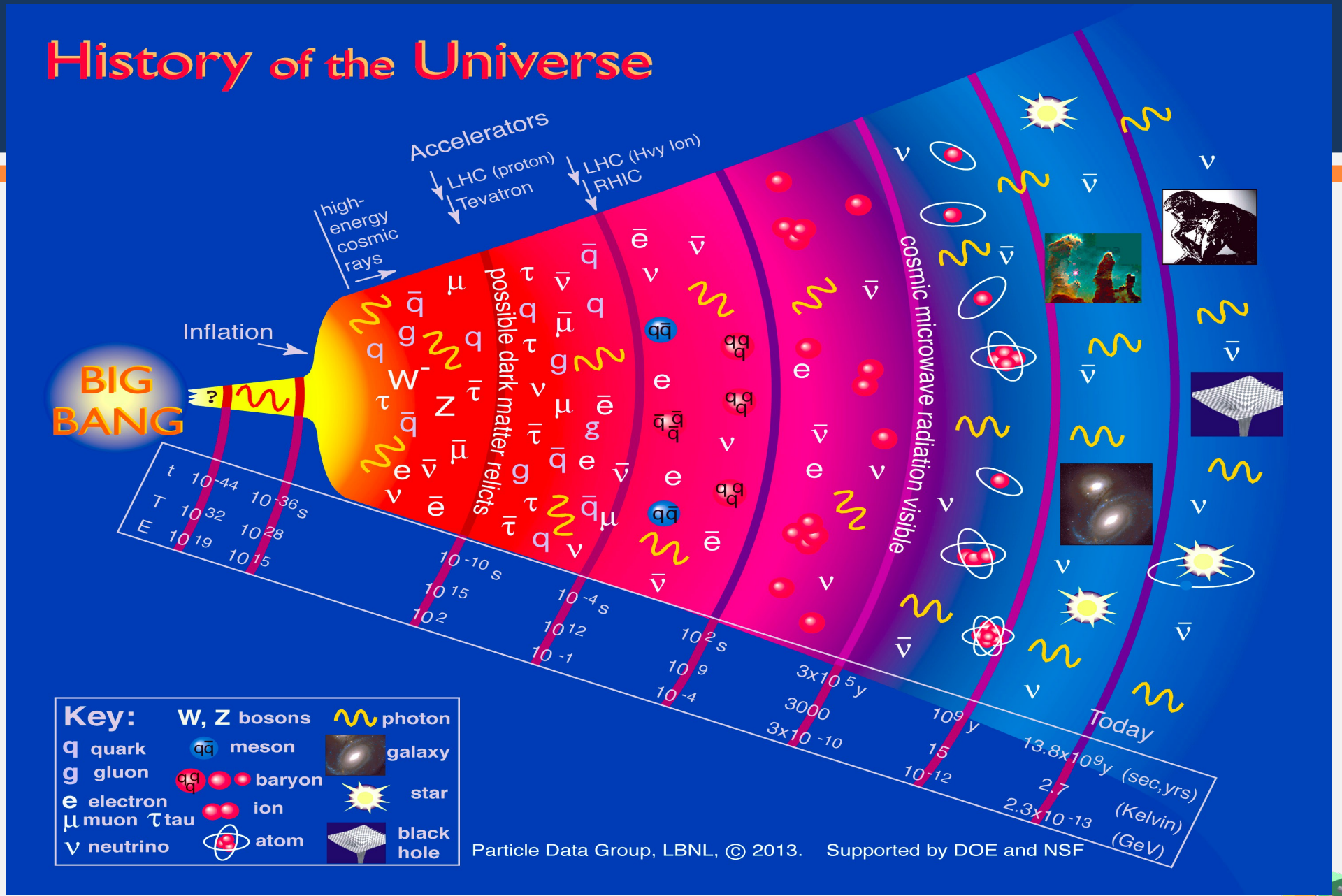
- Première initiative pour structurer les efforts
- Identification des thématiques où contribuer
- Promotion des projets et solutions communs
- Plateforme pour une collaboration entre disciplines

http://hepsoftwarefoundation.org/


VM
WN
Instantiates
Schedules
Requests Job
Cloud
Interface
BS
CE
Requests Resource
RM
WM
Submits Pilot


RECO
promptly
MiniAOD
Analysis1 tuple
Analysis2 tuple
if needed?
Analysis5 tuple
AOD
when new high level calibrations
or recipes become available


Re-reconstruction of Data in 2011: Number of Events per Month
4,000,000,000
3,000,000,000
2,000,000,000
1,000,000,000
0
2010 ReReco
Higgs ReRecos
First full
2011 ReReco
January 2011
February 2011
March 2011
April 2011
May 2011
June 2011
July 2011
August 2011
September 2011
October 2011
November 2011
December 2011
CERN seminar, 13 December 2011:
"tantalizing hints" of ~125 GeV boson in many channels


History of the Universe
Accelerators
LHC (proton)
Tevatron
LHC (Hvy Ion)
RHIC
high-energy cosmic rays
Inflation
BIG BANG
possible dark matter relicts
cosmic microwave radiation visible
Today
Key:
W, Z bosons
photon
q quark
qq meson
galaxy
g gluon
qqq baryon
star
e electron
ion
μ muon τ tau
atom
black hole
ν neutrino
Particle Data Group, LBNL, © 2013. Supported by DOE and NSF

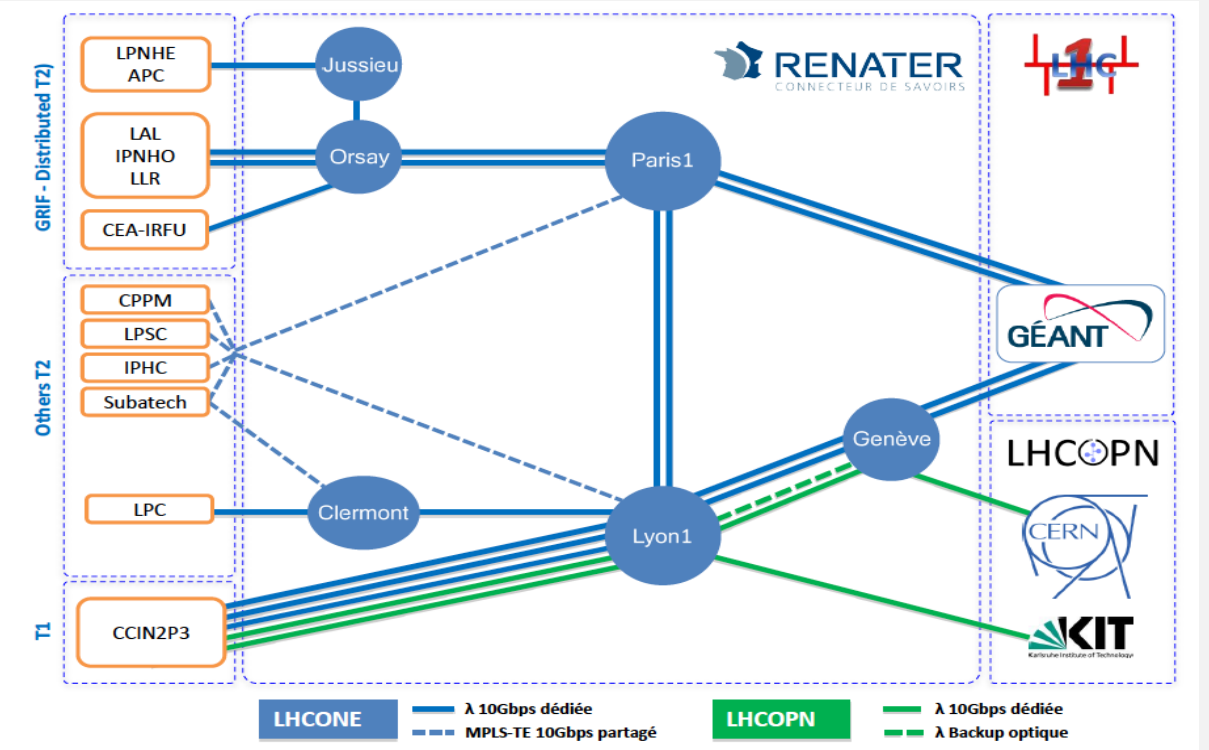
# Evolutions du réseau - LHCONE

Objectif : fournir mondialement une collection de points d'accès pour la connexion des T1/T2/T3 à un réseau privé du LHC (aussi Belle 2).

Complémente LHCOPN.

Aujourd'hui sont connectés :

- tous les sites en France
- ~70 sites dans le monde (dont 40 en EU)

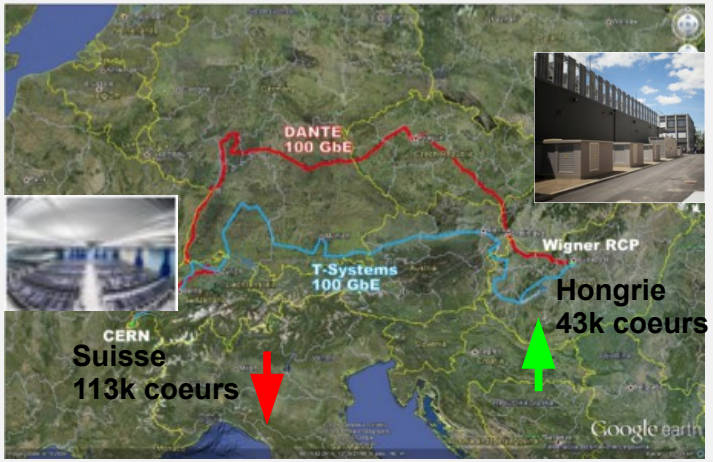



Source : N. Garnier https://indico.in2p3.fr/event/11617/

# Evolution du Tier-0

## Une nouvelle annexe

- Le centre de calcul du CERN (Meyrin) a atteint sa capacité en terme de capacité électrique
- Une nouvelle annexe a été créée à Budapest - Wigner
- Elle est opérée comme une partie du DC de Meyrin
  - Jusqu'à 2,7 MW
  - Connecté à 2 x 100 Gb/s (22ms de latence)
  - Actuellement en production




## Des ressources orchestrées dans un *cloud* privé

- Flexibilité (notamment services)
- Croissance de la taille du Tier-0 à RH constant
  - Aujourd'hui : 4600 HV, 125k coeurs
- Extension dynamique à Wigner
  - Et aussi sur *clouds* commerciaux + HLT