

# Calcul, algorithmes et données pour les expériences du LHC

Catherine Biscarat, L2IT

*catherine.biscarat@l2it.in2p3.fr*

Journée IN2P3 du bâtiment 3R1, université Paul Sabatier

L2IT

# Un mot sur l'oratrice

- Jusqu'en 2011 : expérimentaliste en physique des particules, sur collisionneur

Détecteur aux EU, à Chicago, collisionneur ppbar Tevatron à  $\sqrt{s} = 2$  TeV

- calorimétrie, recherche SUSY
- production d'événements simulés

Détecteur au CERN, à Genève, collisionneur pp LHC à  $\sqrt{s} = 14$  TeV

- calorimétrie, recherche de nouvelles particules
- responsable des activités pour ATLAS dans un centre de calcul majeur

- Depuis 2011 : ingénieur en informatique à l'IN2P3 (LPSC), un des instituts du CNRS
  - grilles de calcul, infrastructures
  - centre secondaire pour le LHC
  - portage d'applications sur les grilles



- 2014-2019 : responsable technique du projet LCG-France
  - fournir les ressources de calcul et de stockage aux expériences du LHC

# L'appareillage

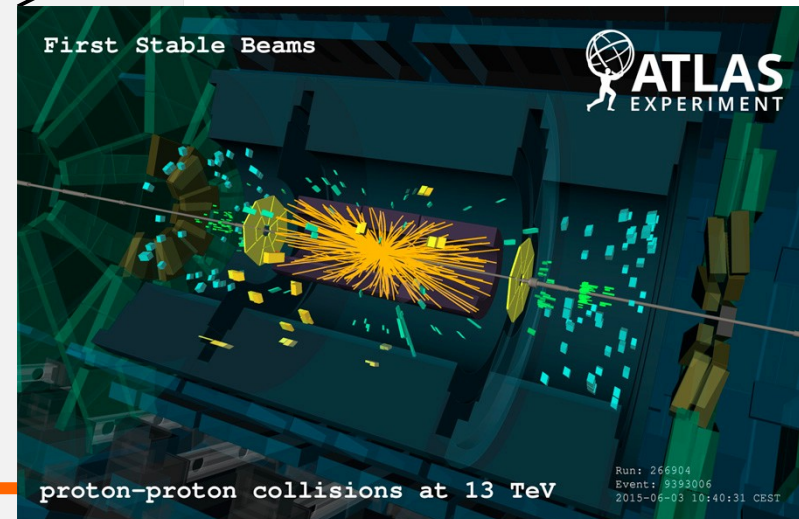
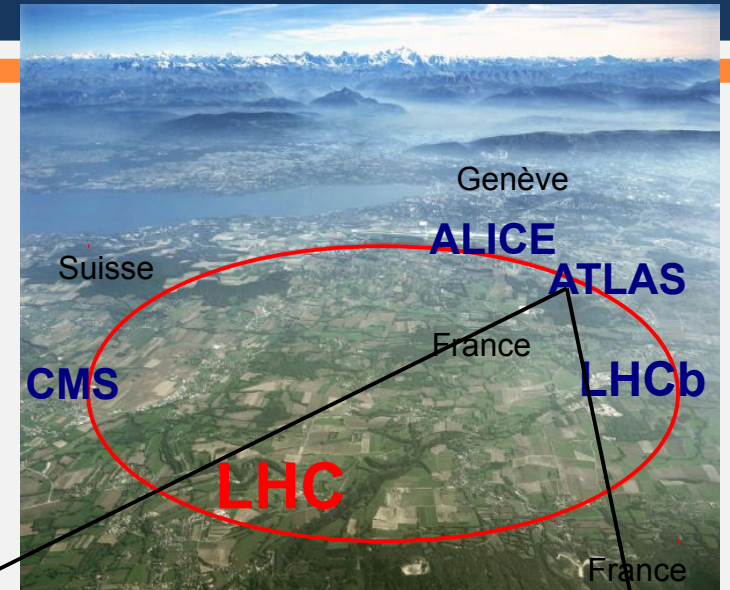
## Large Hadron Collider

- 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
- Des algorithmes très simples et intuitifs
- Petits événements **indépendants**  
→ traitement **séquentiel**

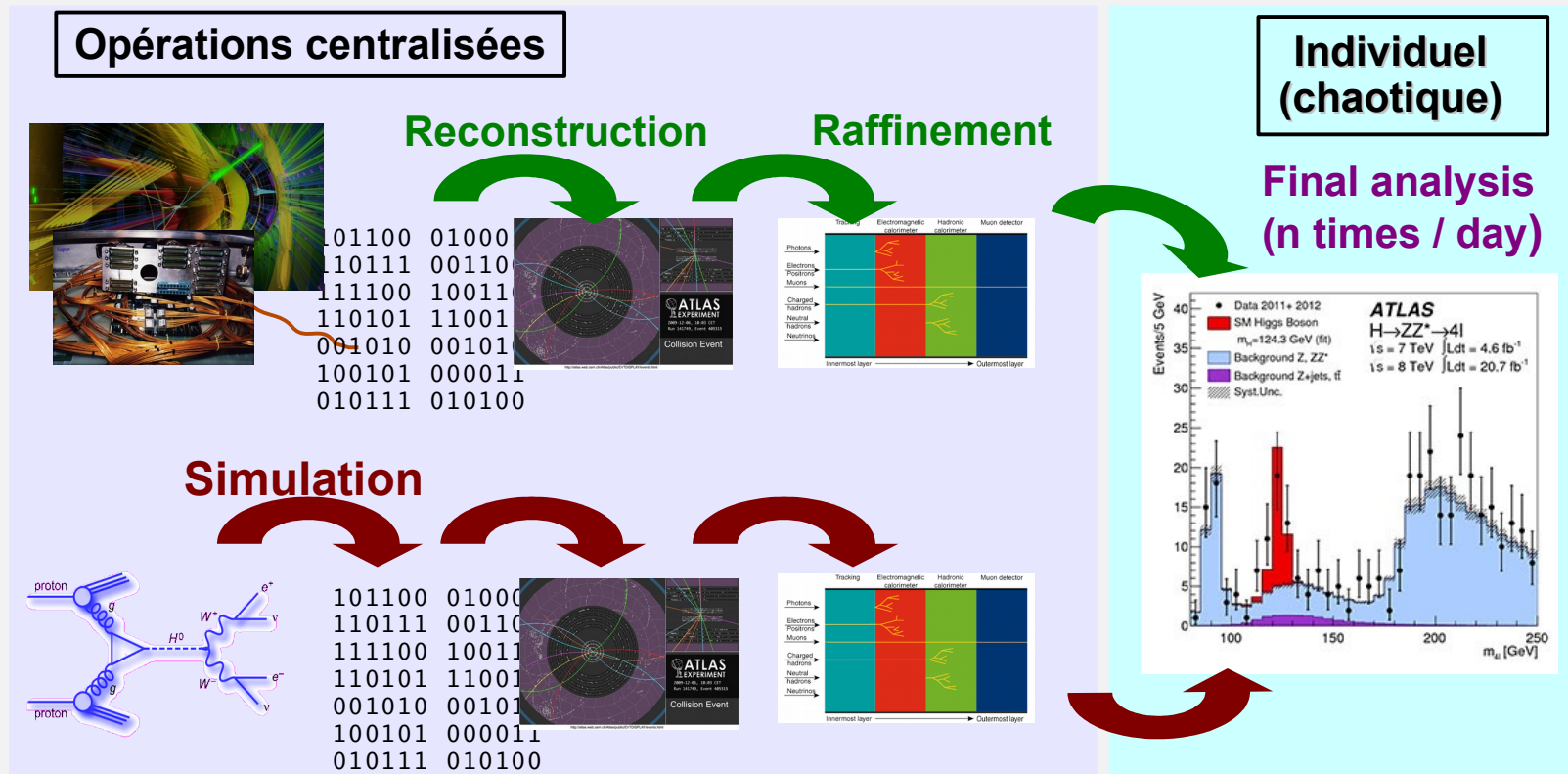
Un événement  
dans ATLAS



L2IT

# Organisation du calcul

Préparation des données pour les analyses finales (objets de physique).  
Données réelles et simulées.  
Réduction de la taille des événements.



# Un nouvel ordre de grandeur

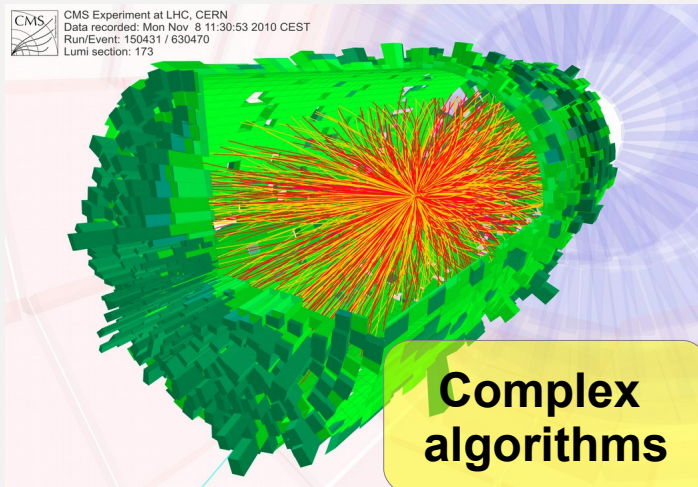
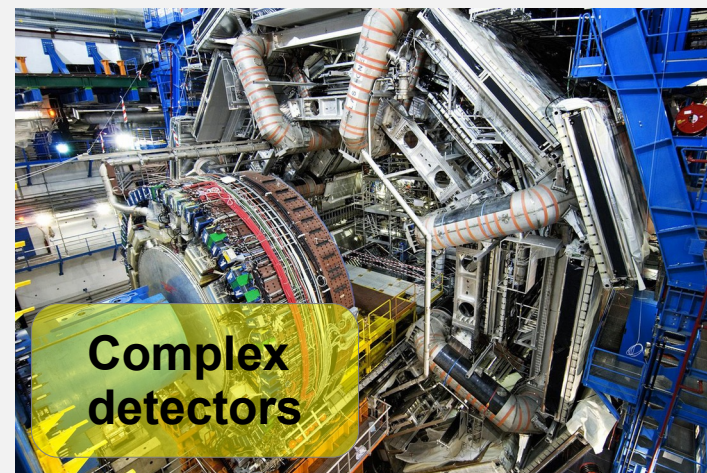
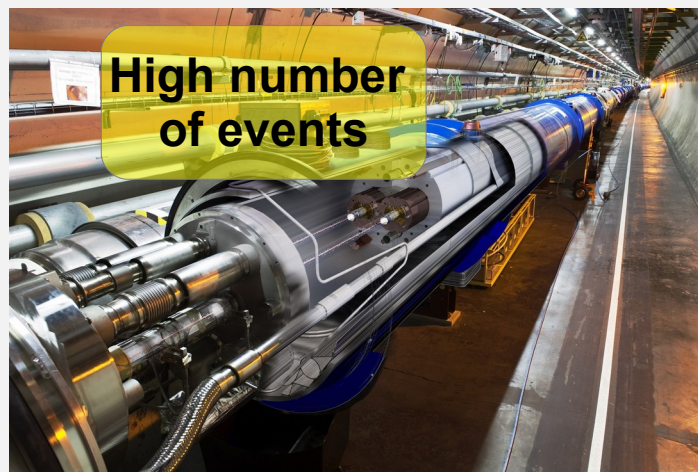
## Contraintes :

Large volume de données  
Ressources (CPU+stockage)  
Milliers utilisateurs finaux

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

Un centre de calcul seul ne pouvait pas satisfaire les besoins à venir.

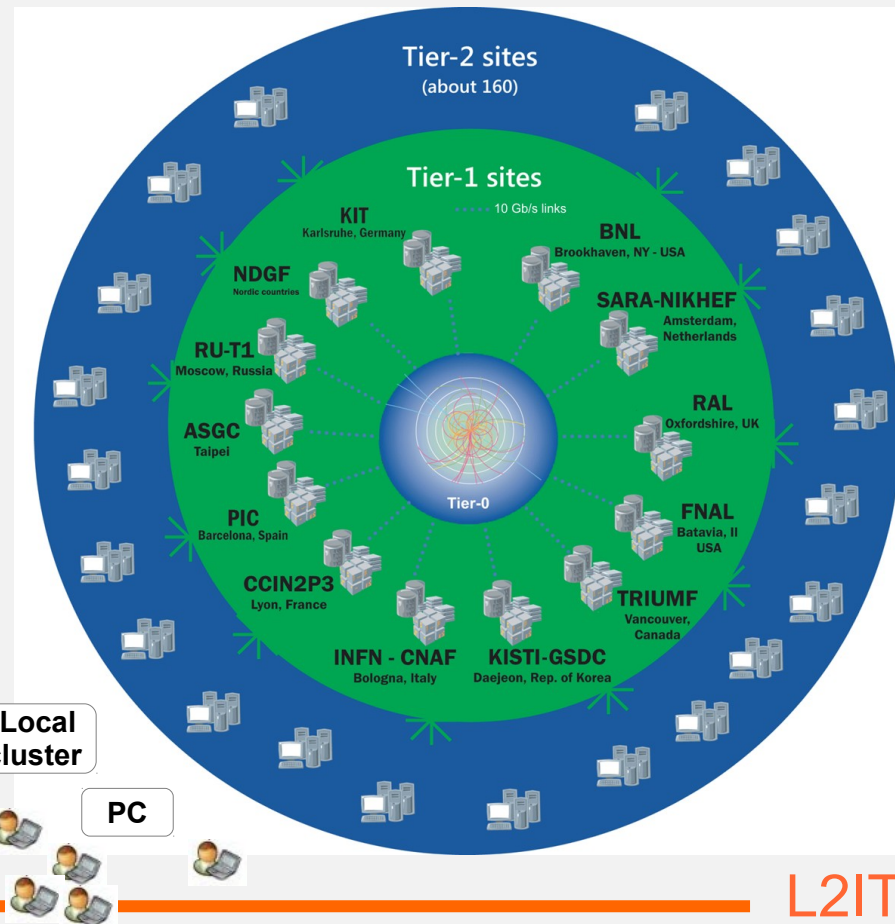
Des sites existaient déjà de part le monde.



# Innovation : la grille de calcul

## Décision de construire une grille de calcul pour le LHC (1999)

- Fournir aux utilisateurs un accès **transparent** à des ressources distribuées
  - Par analogie avec « electric power grid »
  - Les pionniers : HEP et biologie
- Chaque site garde le **contrôle** des accès à son site (droits des utilisateurs) et les technologies sous-jacents (protocoles standards)
  - développement de **middleware**
- Modèle distribué, hiérarchique (Tiers LCG)
  - Evénements **indépendants** !
- Focus sur le contrôle du réseau (1Gb/s attendu)
  - Début des réseaux étendus à haut débit



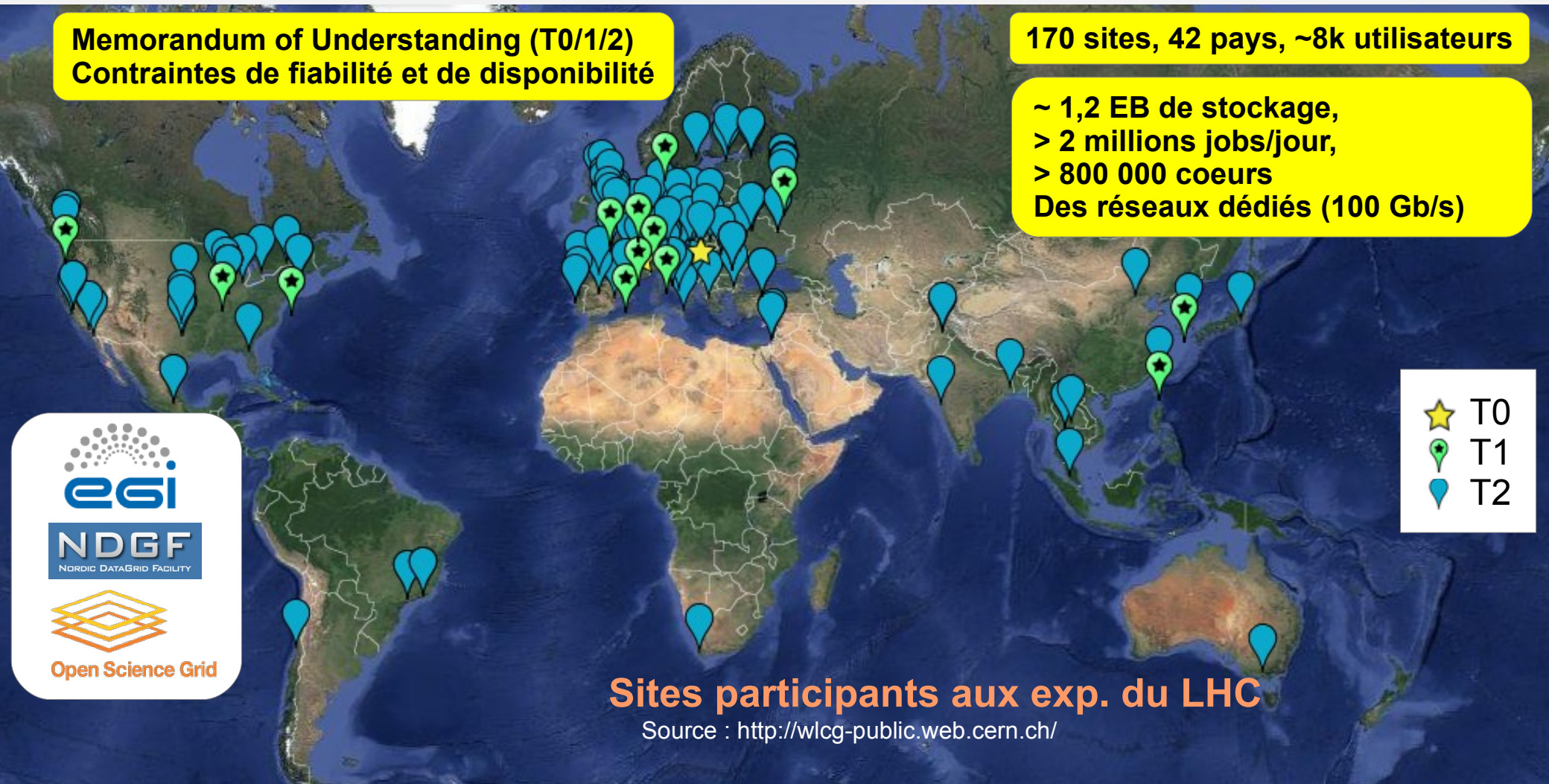


# La grille pour le LHC (WLCG)

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

170 sites, 42 pays, ~8k utilisateurs

~ 1,2 EB de stockage,  
> 2 millions jobs/jour,  
> 800 000 coeurs  
Des réseaux dédiés (100 Gb/s)



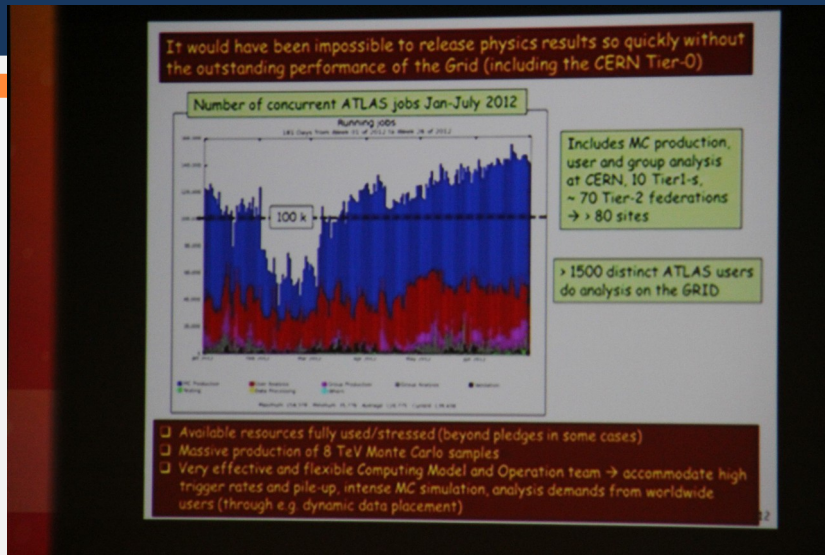
- ★ T0
- 📍 T1
- 📍 T2



Sites participants aux exp. du LHC

Source : <http://wlcg-public.web.cern.ch/>

# « computing enables physics »



Photography: C. Biscarat

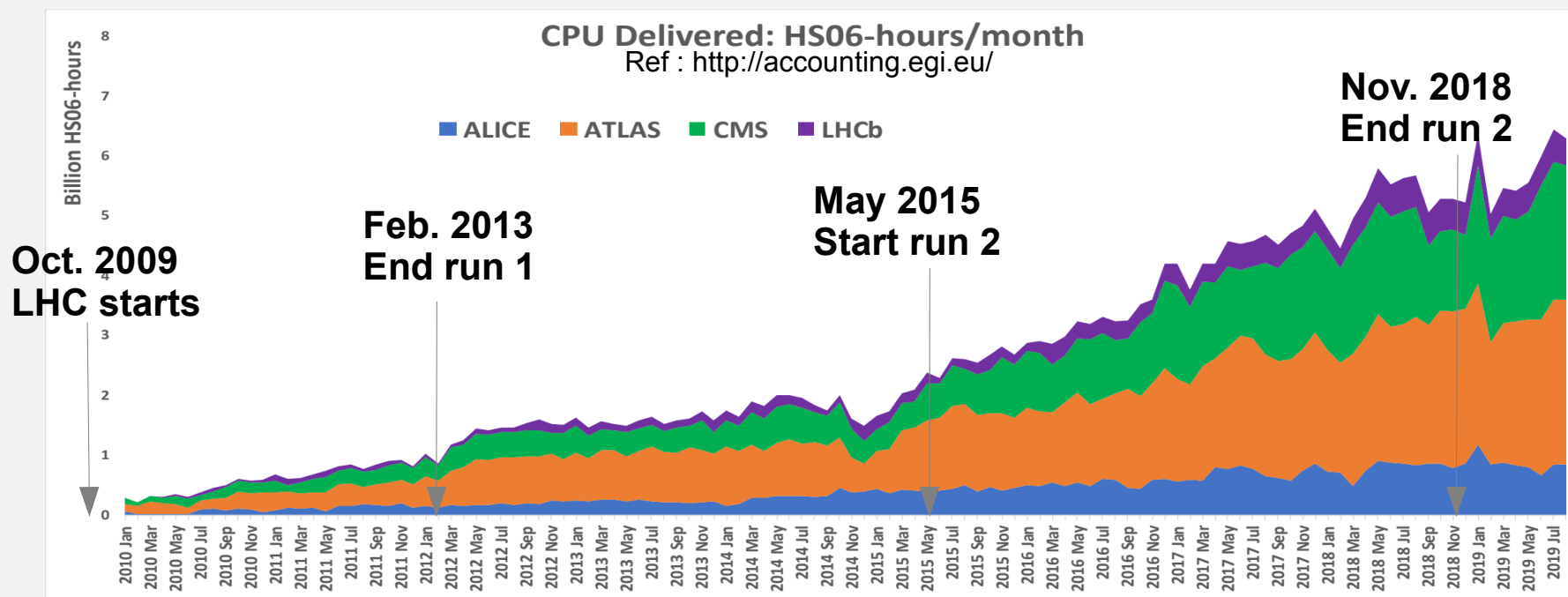


Announce de la découverte du maillon manquant de notre Modèle Standard, le boson de Higgs.

CERN seminar, July 4<sup>th</sup> 2012, retransmitted at ICHEP (Melbourne)



# Flash sur les activités de WLCG

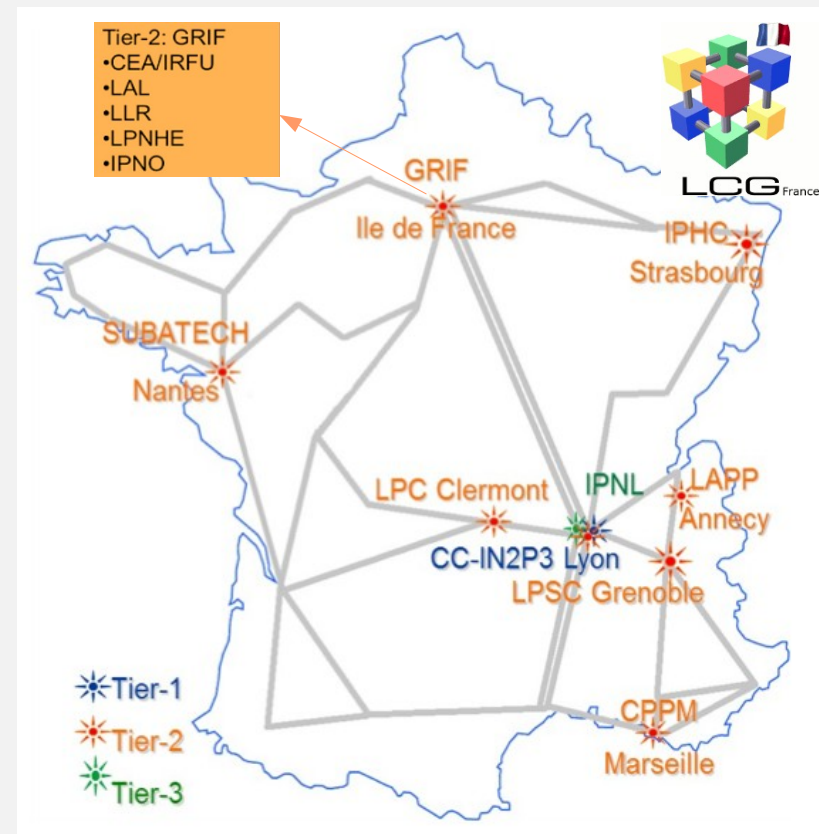


~900 000 coeurs en continu (production centralisée et analyses individuelles)  
Transferts réguliers de > 100 PB/mois.

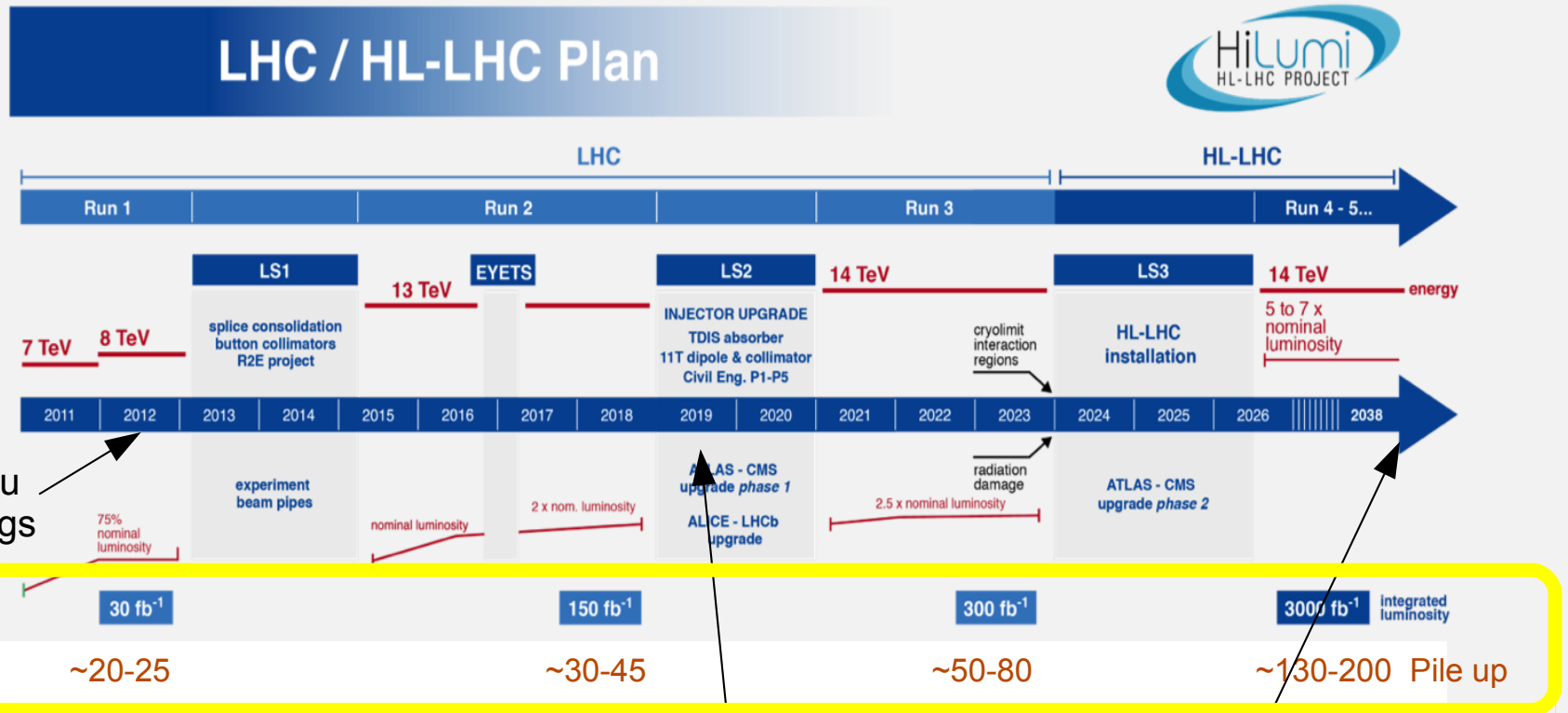
Améliorations continues dans le *middleware*, les *workflows*, les algorithmes

# 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)
- Représente la France auprès de WLCG.
- Accords site/WLCG pour assurer une haute disponibilité (7/24) :
  - > 99% pour les Tier-1
  - > 95% pour les Tier-2
- Le Tier-1 : le Centre de Calcul de l'IN2P3 (Lyon)
  - ~60 ingénieurs
  - 2x850 m<sup>2</sup> de salles machines
  - 4 expériences LHC (2/3 du centre)
  - 10% de WLCG Tier-1
  - 80 PB, 40k coeurs
- Les Tiers-2 doublent les capacités (CPU, disque).



# Et après ? Horizon à 10 ans



Découverte du boson de Higgs

Integrated Lumi.: nb of events  
Pile-up : event complexity

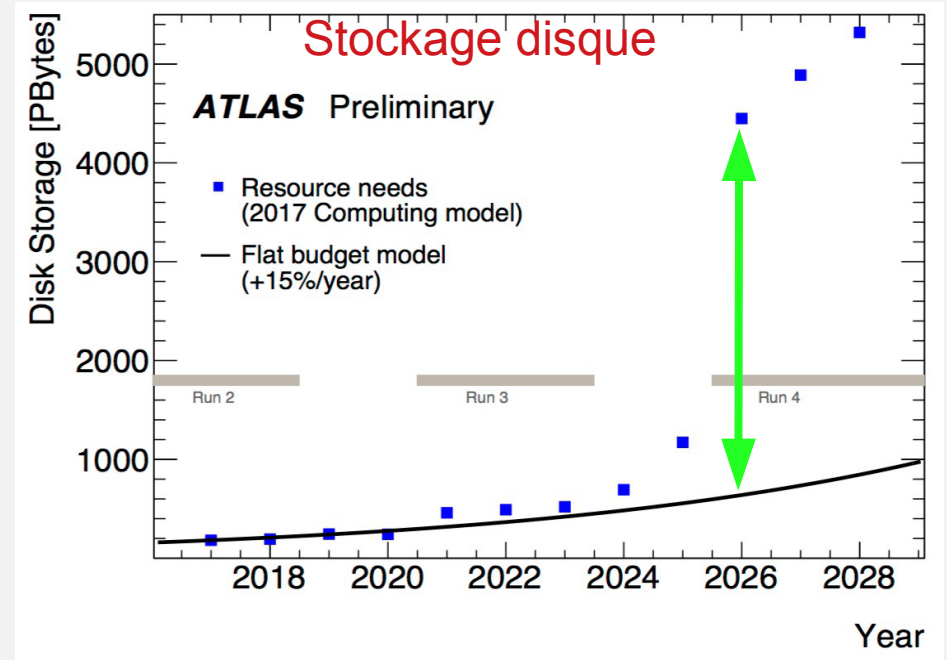
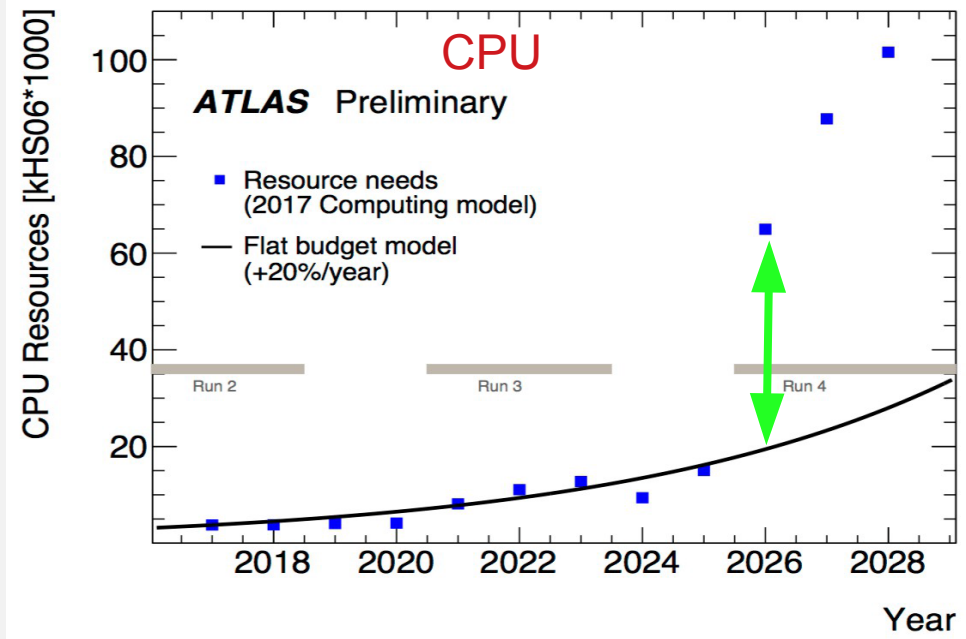
Aujourd'hui : 90 fb<sup>-1</sup>

Fin du LHC ~ 2040 : 3000 fb<sup>-1</sup>

30 fois plus de données !  
Des événements complexes

L2IT

# En terme de ressources



- Si nous gardons notre modèle de computing actuel, il **manquera un facteur important** dans les ressources de calcul et de stockage → **changements importants**
  - Algorithmes, architectures, infrastructures, gestion des données, analyses ...
- Au HL-LHC (2026) : **la richesse du programme de physique sera limitée** par la performance des algorithmes et les ressources.

# Le « community white paper »

## *A Roadmap for HEP Software and Computing R&D for the 2020s*

- D'autres expériences avec des besoins importants démarreront sur la même échelle de temps → **collaboration** → Community White Paper [arxiv.org 1712.06982]
- La résolution de notre problème viendra des algorithmes, pas des évolutions du matériel.
- Le machine learning est un des grands axes identifiés pour améliorer notre consommation de ressources.
  - Rapidité des algorithmes (traces)
  - Classification des événements
  - Détection d'anomalie en cours d'opérations
- Une nouvelle thématique dans notre conférence d'informatique la plus prestigieuse (CHEP)

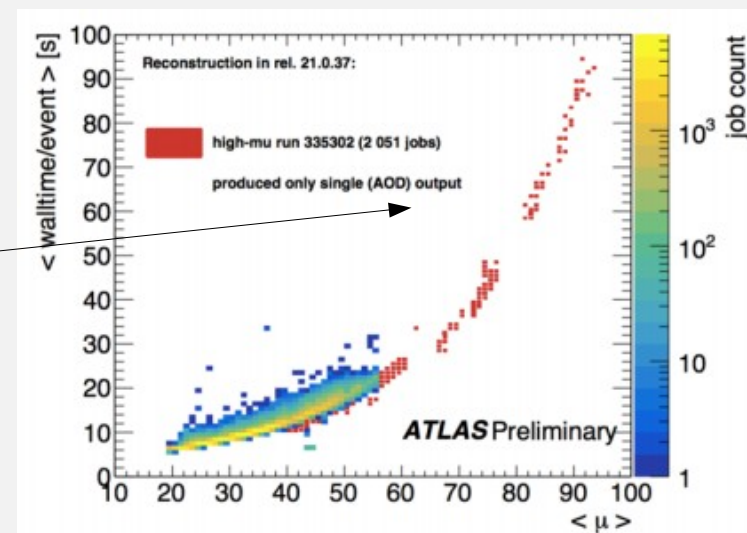
### Contents

1	Introduction	2
2	Software and Computing Challenges	5
3	Programme of Work	11
3.1	Physics Generators	11
3.2	Detector Simulation	15
3.3	Software Trigger and Event Reconstruction	23
3.4	Data Analysis and Interpretation	27
3.5	Machine Learning	32
3.6	Data Organisation, Management and Access	36
3.7	Facilities and Distributed Computing	41
3.8	Data-Flow Processing Framework	45
3.9	Conditions Data	48
3.10	Visualisation	51
3.11	Software Development, Deployment, Validation and Verification	53
3.12	Data and Software Preservation	58
3.13	Security	61
4	Training and Careers	65
4.1	Training Challenges	66
4.2	Possible Directions for Training	67
4.3	Career Support and Recognition	68
5	Conclusions	69
	Appendix A List of Workshops	72
	Appendix B Glossary	74
	References	80

# Un exemple concret : le tracking



- La reconnaissance de trajectoire domine le temps CPU de la reconstruction d'un événement.
- Augmentation dramatique du temps CPU avec le nombre d'interactions.
- Augmentation dramatique du nombre d'interactions :
  - Run 1 (2012) → Run 2 (2015) → HL-LHC (2027)
  - $\langle \mu \rangle \sim 20$  →  $\langle \mu \rangle \sim 30$  →  $\langle \mu \rangle \sim 200$



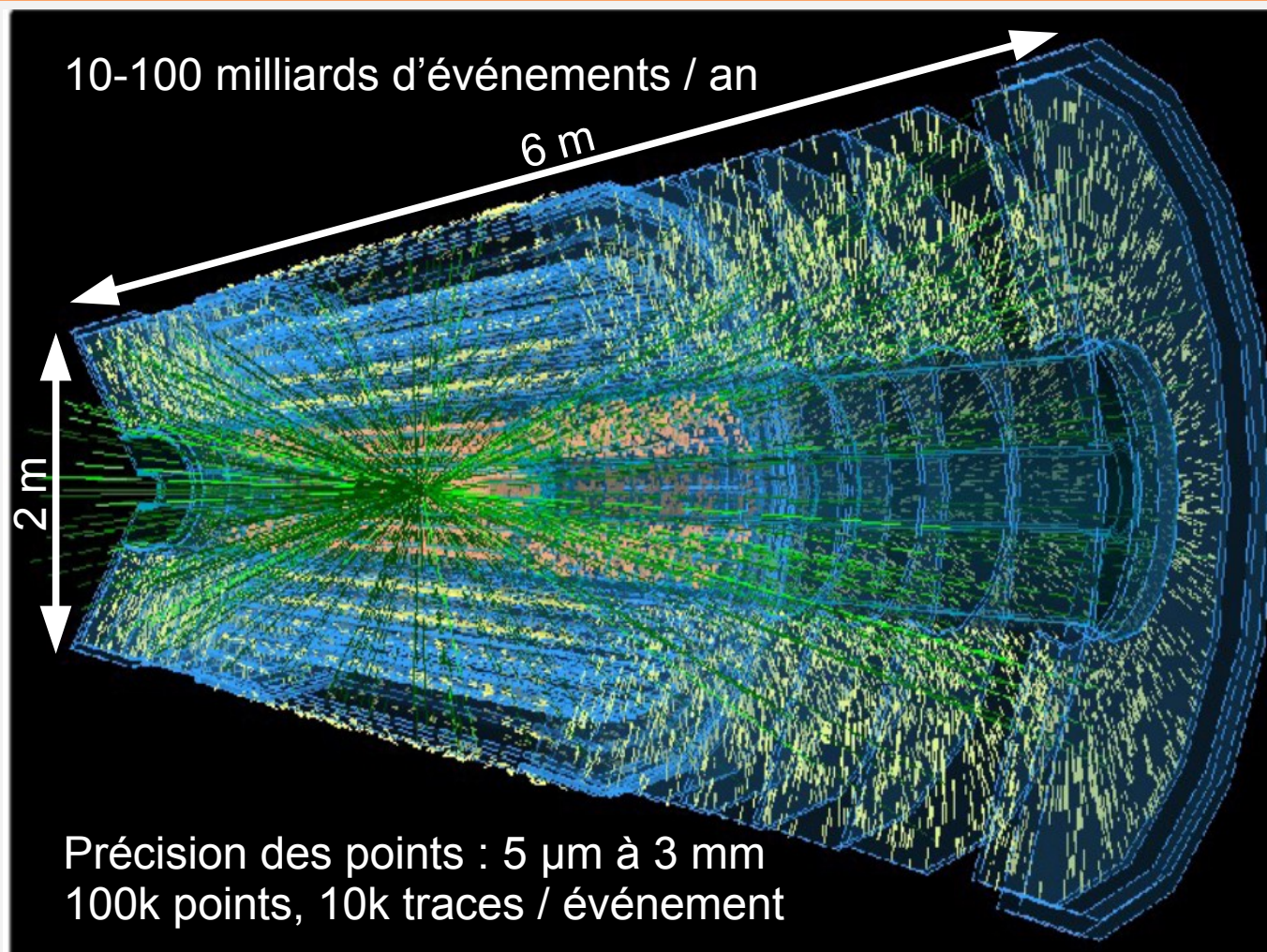
# Un exemple concret : le tracking

Notre communauté a commencé à s'organiser (IN2P3, ATLAS, «HEP» ...).

Nous cherchons à bâtir des solutions avec les experts de ce domaine (Computer scientists).

## Le *TrackML challenge*

- Un lot de données d'entraînement (vérité) et un lots de points → traces
- Des algorithmes 10 fois plus rapides



# En guise de conclusion

- La naissance du LHC s'est accompagnée par une rupture dans notre modèle de calcul.
- Nous avons alors contribué à construire la grille de calcul. Celle-ci s'est avérée un succès et a parfaitement rempli son rôle en couvrant nos besoins.
- La phase de haute luminosité du LHC s'accompagne d'un nouveau défi pour l'informatique.
- Celui-ci pourra être relevé seulement à travers une **révision profonde de nos algorithmes**.
- Le L2IT a pour ambition de contribuer activement à cette aventure et nous cherchons à bâtir des collaborations avec les experts du domaine sur Toulouse.
- Plus que le LHC, les techniques de machine learning bénéficieront à d'autres domaines de l'IN2P3 et aussi à des expériences qui auraient les mêmes défis à relever.