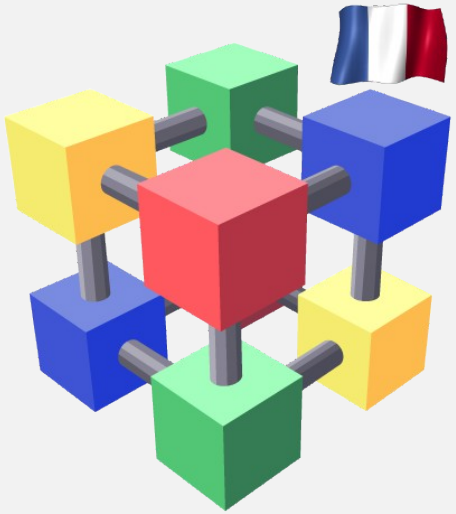
Vers le Run 4 du LHC

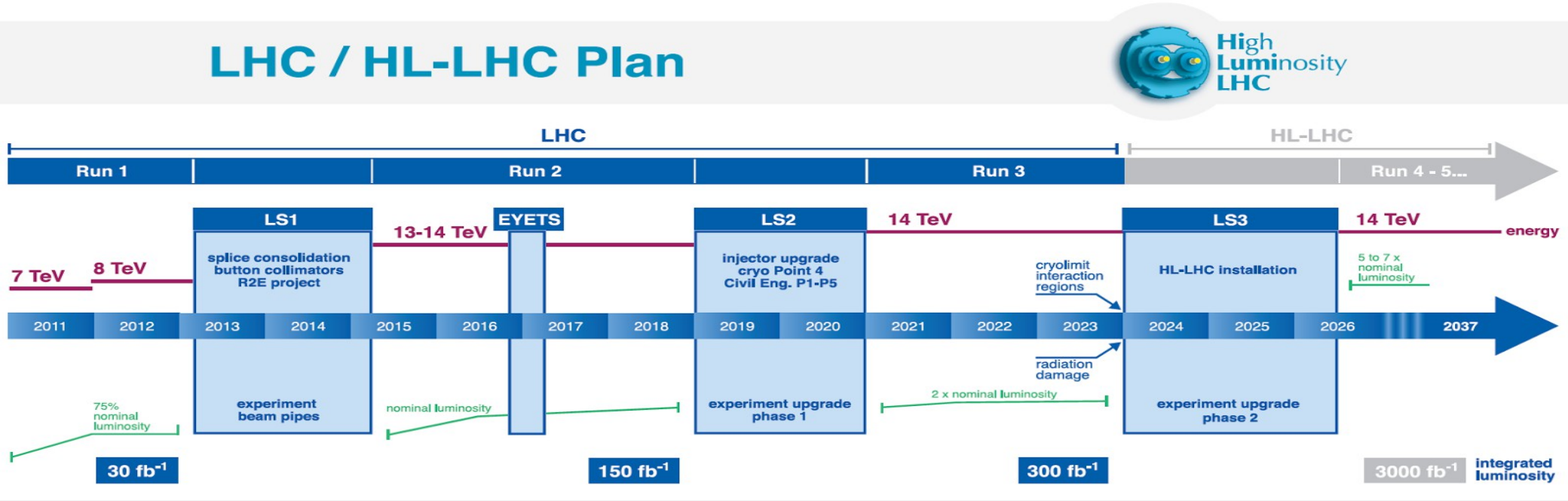
C. Biscarat & M. Jouvin
Avec l'aide de L. Duflot, E. Fede, A Sartirana, R. Vernet



LCG France

Journées LCG-France, 24 novembre 2017,
Clermont-Ferrand

Les challenges

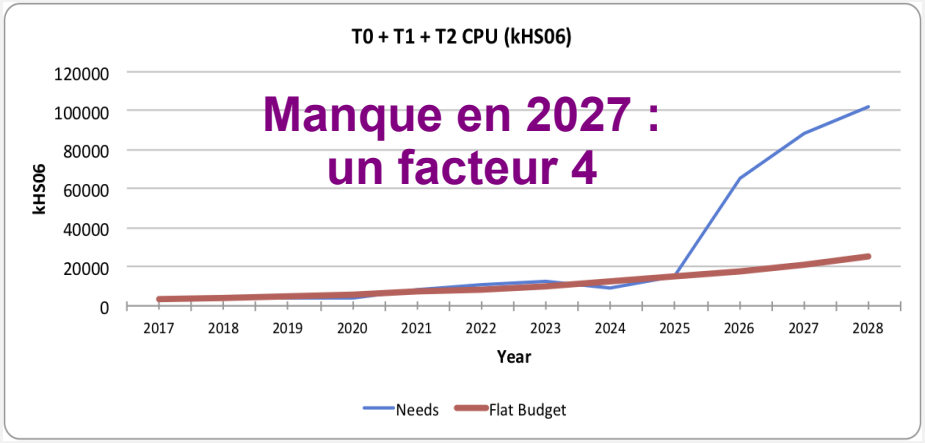


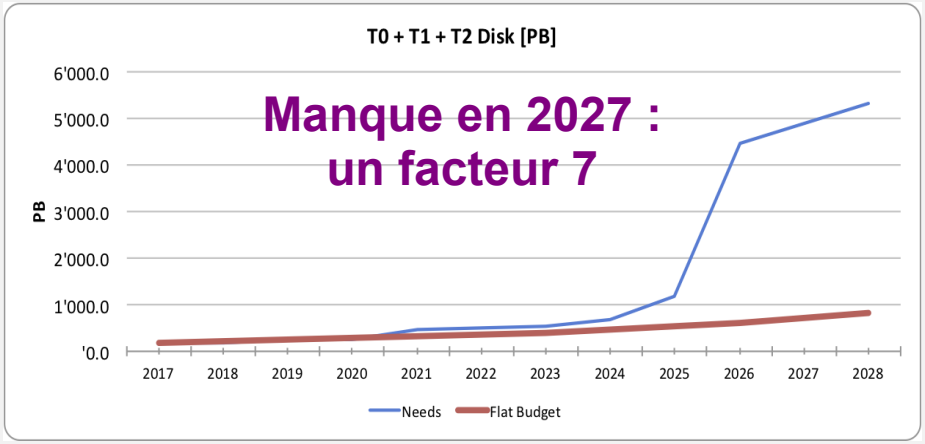
- Ailleurs en HEP : FNAL muon et neutrinos (DUNE), FAIR (GSI Darmstardt), Belle-2 …
- Autres programmes avec des technologies proches : SKA, CTA

En ressources

- Upgrades des triggers de ATLAS et CMS - > nombre d'événements enregistrés x 5 à 10
  - cf. présentations de Luc et Andréa (hier)
- Faisceau plus intense - > plus de collisions par événements x 60 à 200

Exemple de ATLAS: besoins en ressource avec une extrapolation du modèle de calcul Run2 en prenant en compte des projections de décroissance des coûts – comparaison avec une croissance à budget plat





- Challenges techniques : s'adapter aux nouvelles architectures, inclure le HPC
- Challenges humains : ré-écriture des codes, convergence entre les expériences
- Au-delà du LHC : DUNE, Belle-2 et SKA (avec des jeux de données comparables)

La première brique

HSF web site

- **HSF Community white paper**
  - Réunir toute la communauté des Hautes Energies (+ que LHC)
  - Chalenges sur horizon à 10 ans
  - SW et computing
  - Planifier et s'entendre sur les actions en termes de R&D

- **Second draft du CWP**
  - Disponible pour commentaires (dead-line : 1 décembre) - > lien
  - Pour appuyer le contenu du CWP, chacun est invité à le signer - > lien
  - Cf. courriel de Michel (éditeur) sur les listes LCG-France

- **Un effort communautaire**
  - Une large participation avec des workshops dédiés (San Diego 110 participants janvier 2017 ; Annecy, LAPP 90 participants juin 2017)

Le processus jusqu'au TDR HL-LHC (Run 4)

- Le LHCC demande un document à WLCG sur sa stratégie pour le Run 4
  - Points de R& D
  - En vue de la préparation du TDR (2020)

- Document stratégique issu du CWP
  - Avec une vue WLCG
  - Les points clés pour le HL-LHC
  - Avec des milestones de R&D
  - Une stratégie pour les coûts

- Document dû en février 2018
  - Sera revu en MB et en OB

Le document stratégique - Plus en détails

Document outline

- Introduction
  - Update of the resource requirements estimates
- Themes that address cost and performance
  - Next slide
- R&D programme & timescale

Réf. Ian Bird, MB

Key strategies to manage cost

1. Software performance and efficiency
   - Fabric performance & efficiency, including data access
   - This cuts across all applications
2. Minimise data volumes
   - HLT processing, selection, nano-AODs
   - Reduce replication
   - Remote (bulk) processing without need to ship data
   - Cloud analysis
3. Reduce processing costs
   - Fast simulation will be a huge benefit
   - Improved/new algorithms (ML/DL is seen as important technology)
4. Minimise operations costs
   - Consolidation or federation at National level
   - Common solutions for infrastructure & software stacks
   - Move common services (data, workflow) into infrastructure layers
5. Sustainability
   - Common solutions and joint efforts
   - Common infrastructure – across HEP and with other related sciences
   - Authentication/authorization infrastructure – which direction?
   - Data preservation and reproducibility

Non-exhaustive list

Les groupes de travail du CWP

13 groupes de travail – focus sur la R&D pour le Run 4

2 Software and Computing Challenges **6**

3 Programme of Work **11**

3.1 Physics Generators 12

3.2 Detector Simulation 16

3.3 Software Trigger and Event Reconstruction 23

3.4 Data Analysis and Interpretation 28

3.5 Machine Learning 31

3.6 Data Organisation, Management and Access 37

3.7 Facilities and Distributed Computing 42

3.8 Data-Flow Processing Framework 46

3.9 Conditions Data 49

3.10 Visualisation 52

3.11 Software Development, Deployment, Validation and Verification 55

3.12 Data and Software Preservation 59

3.13 Security 62

4 Training and Careers **67**

Lien : page des groupes de travail et leurs docs

CWP – Facilities and Distributed computing

- Enjeux : lots de données gigantesques
  - des infrastructures partagées avec d'autres exp.
  - diversité des ressources de computing (HPC, cloud, commercial, architectures)
  - consolidation du stockage
- Utilisation du network intensif (“moins cher”), mais partagé et plus intégré aux applications
- Evolution de notre système : oui mais dans le principe d'une optimisation des coûts (HW, RH)

Programme de R & D

- Comprendre les coûts pour tout changement - > “System Perfomance & Cost modeling WG”
- Comprendre le modèle des “data lakes” (gain sur le stockage), dans quelle mesure on gagne et comment ce modèle s'inscrit dans les modèles régionaux distribués
- Converger entre les expériences sur les modèles de data management pour accommoder
  - Sites disk cache
  - Processing par petites parties (on s'éloigne des datasets)
  - Intégration de ressources hétérogènes (HPC, cloud)

cf. talk de Michel hier

- Optimiser le scheduling
- Evaluer l'apport d'Analysis Facility (interactive)

Systems Performance and Cost Modeling WG

- Motivation : prédire l'impact global d'un changement
  - Modélisation l'infrastructure, workload & workflow en vue d'optimiser les coûts pour un objectif de performance
- Vue globale de notre système
  - En intégrant diversités de sites, de pays et de modèles organisationnels (financements)
  - Manpower
- WLCG working group (long terme) :
  - très fortement couplé avec HSF et HEPIX
  - focus sur WLCG
  - sera d'intérêt pour les autres exp. (CTA, SKA)

Mandat

- Caractériser le système avec des métriques (workload < - > infrastructure)
- Décortiquer des workloads de référence (implication des expériences)
- Etablir des modèles (et vérifier) leur pouvoir de prédiction (ref. workloads vs changements)
- Etablir une méthodologie pour prédire les variations de coûts (local)

Systems Performance and Cost Modeling WG

- Cela démarre (2 réunions) dont établissement du mandat
- 1ète étape :
  - Printemps 2018 / Naples : métriques et modèle simplifié, un workload
- Long terme avec un processus de raffinement

Members wlcg-SystemPerformanceModeling@cern.ch

- Andrew Sansum
- Biscarat(AT)in2p3.fr
- Daniele Bonacorsi
- Eric Fede
- Frank Würthwein
- Michel Jouvin
- Martin Gasthuber
- Michele Michelotto
- Renaud Vernet
- Sartiran(AT)llr.in2p3.fr
- Erik Mattias Wadenstein
- Gareth Roy
- Yves Kemp
- Concezio Bozzi
- Helge Meinhard
- Markus Schulz
- Andrea Sciaba
- Oxana Smirnova
- Andrea Valassi
- Torre Wenaus
- Davide Costanzo
- Tommaso Boccali
- Alessandra Forti
- Jose Flix Molina
- Johannes Elmsheuser
- Jan Iven
- Andrey Kiryanov
- Domenico Giordano
- David Lange
- Graeme Stewart
- Raul Cardoso Lopes

En savoir plus :

- Indico
- wiki

- Expériences, sites, HSF, HEPIX
- FR : Andrea, Catherine, Eric, Michel (HSF), Renaud

Conclusions et prochain rendez-vous

- Le CWP identifie les enjeux et donne des pistes pour relever le défi du Run 4
  - Focus plus large que WLCG
  - R&D
  - Optimisation globale (coûts)

- Opportunité
  - De valoriser notre expertise
  - De réfléchir notre modèle de fonctionnement

- Ce document est important pour WLCG
  - Document stratégique pour le LHCC
  - HL-LHC TDR 2020

- WLCG workshop conjoint avec HSF – 26-29 mars 2018 Naples
  - Organiser la R&D (workload management, data lakes) ; que HL-LHC