



LHCb: L'analyse aux T1s

Luisa Arrabito - CC IN2P3



IPNL Journées Grille France, 14-16/09/2009



Rappel du Computing Model de LHCb (1/2)

- Les Raw data sont transférées en temps réel au Tier-0
 - Enregistrement dans le Grid File Catalog (LFC)
 - Provenance des Raw data dans une base de données (Bookkeeping)
 - Deux copies sont distribuées dans les Tier1s
 - Taux: ~2000 événements/s (35 Ko/év) → 70 Mo/s
- Tout le traitement des données jusqu'à la production des μ DST ou des NTuples est distribué
 - Il n'est pas possible d'effectuer la première phase de reconstruction entièrement au Tier0
 - ↳ Le Tier0 est considéré comme distribué
 - ↳ La première phase de reconstruction a lieu dans les T1s, ainsi que le re-processing
 - L'analyse a lieu dans les T1s
- Une partie de l'analyse n'est pas associée aux données
 - Extraction des paramètres physiques sur la violation de CP (toy-MC, procédures de fit...)
 - Utilisation de ressources distribuées (T1s/T2s)



Rappel du Computing Model (2/2)

- Les Tier2s contribuent à:
 - Simulation MC
- Les Tier1s contribuent à:
 - Reconstruction
 - Stripping
 - Analyse
 - Stockage sur le MSS (Raw, rDST, DST, données MC)
 - Stockage sur disque (DST)
- Les jobs s'exécutent là où les données sont stockées
- Les jobs d'analyse utilisent des datasets de dimensions réduites:
 - Plusieurs canaux d'intérêt
 - Peu d'événements dans chaque canal (entre 10^2 et 10^6 événements/an)
 - Le physicien analyse 10^7 événements au maximum
 - Le dataset final peut être traité sur un laptop (quelques centaines de Go)

Les 6 Tier1s:

- CNAF (IT, Bologna)
- GridKa (DE, Karlsruhe)
- IN2P3 (FR, Lyon)
- NIKHEF (NL, Amsterdam)
- PIC (ES, Barcelona)
- RAL (UK, Didcot)



L'analyse dans LHCb

LHCb @ LCG-France

- Pourquoi l'analyse aux Tier1s?
 - L'emplacement des fichiers est crucial
 - Le stockage et l'accès aux fichiers doivent être fiables, robustes, performants et scalables
 - Les T1s sont en principe moins sujets à des arrêts et il est plus facile de les contrôler centralement
 - Les besoins de CPU sont assez modestes
- Le concept de la Grille pour LHCb
 - Un ensemble de ressources travaillant d'une façon collaborative
 - Les ressources sont à disposition de toute la collaboration dans son ensemble
 - La reconnaissance des contributions est indépendante du type de jobs qui tournent dans un site
 - ↳ Il n'y a pas de tâches plus ou moins nobles. Toutes les tâches sont nécessaires pour la réussite de l'expérience
 - ↳ Les ressources ne sont pas rendues disponibles en fonction des appartenances nationales
 - La disponibilité des ressources est un point crucial



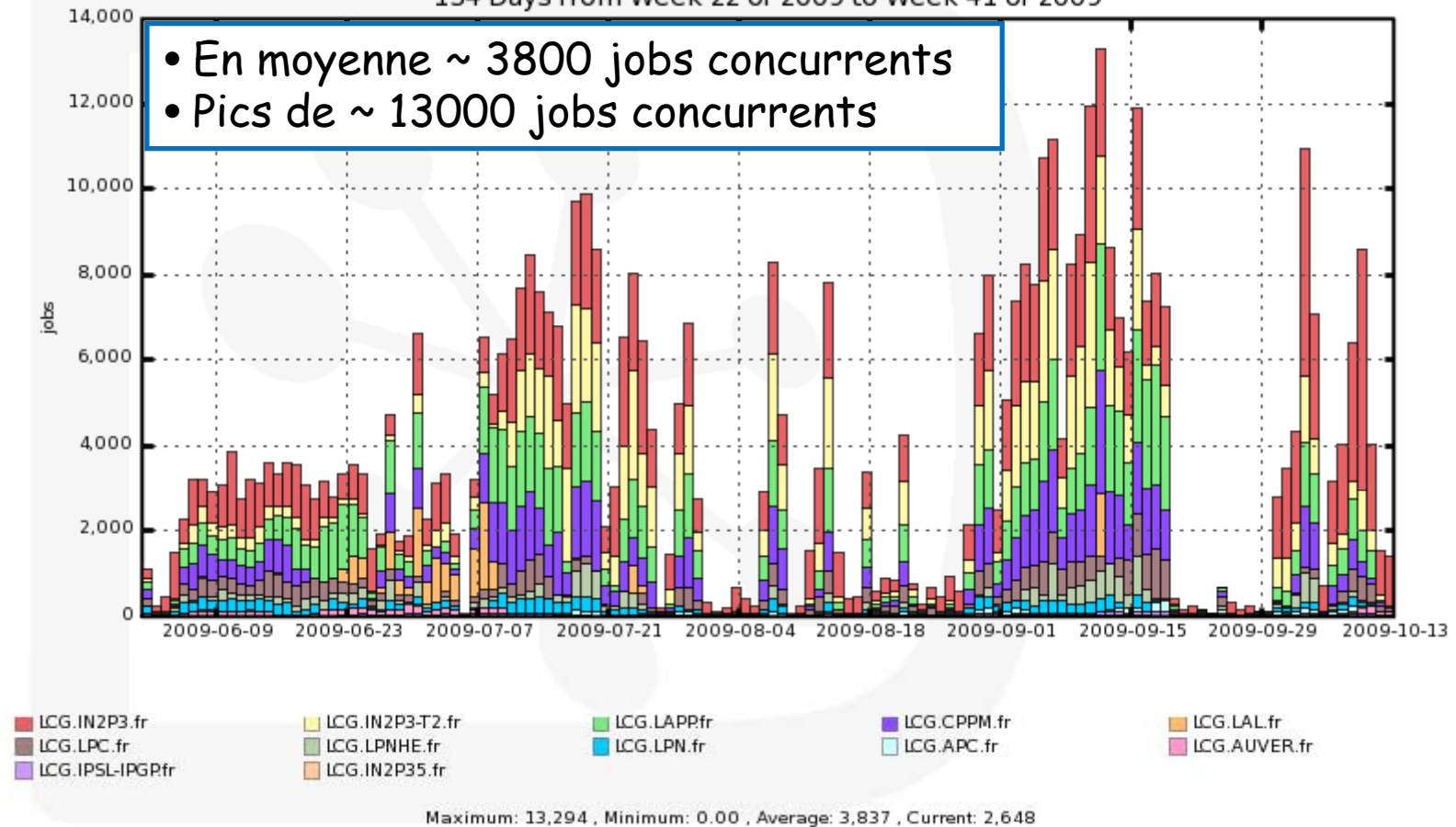
- L'analyse aux Tier2s
 - Les ressources dédiées à l'analyse doivent être des ressources supplémentaires
 - ↳ Il y a juste assez de ressources pour la simulation
 - Le stockage doit être un Grid-SE
 - Le CE est banni centralement en cas de problèmes (comme pour les Tier1s)
- L'analyse aux Tier3s
 - Les ressources sont réservées aux utilisateurs locaux (pas d'accès au travers des CEs, queues de batch locales)
 - Le stockage peut être ou non un Grid-SE



Statistiques depuis le 1er juin 2009

Jobs par site en France

134 Days from Week 22 of 2009 to Week 41 of 2009



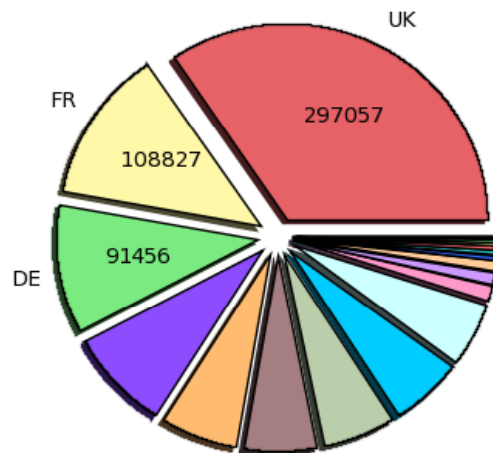
LHCb @ LCG-France





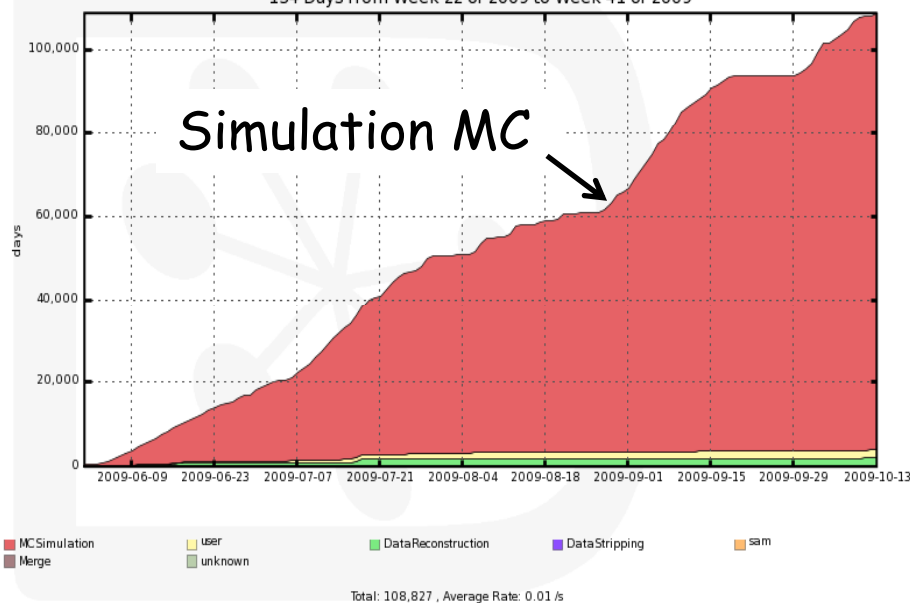
Statistiques depuis le 1er juin 2009

CPU days used by Country (Sum: 860836)

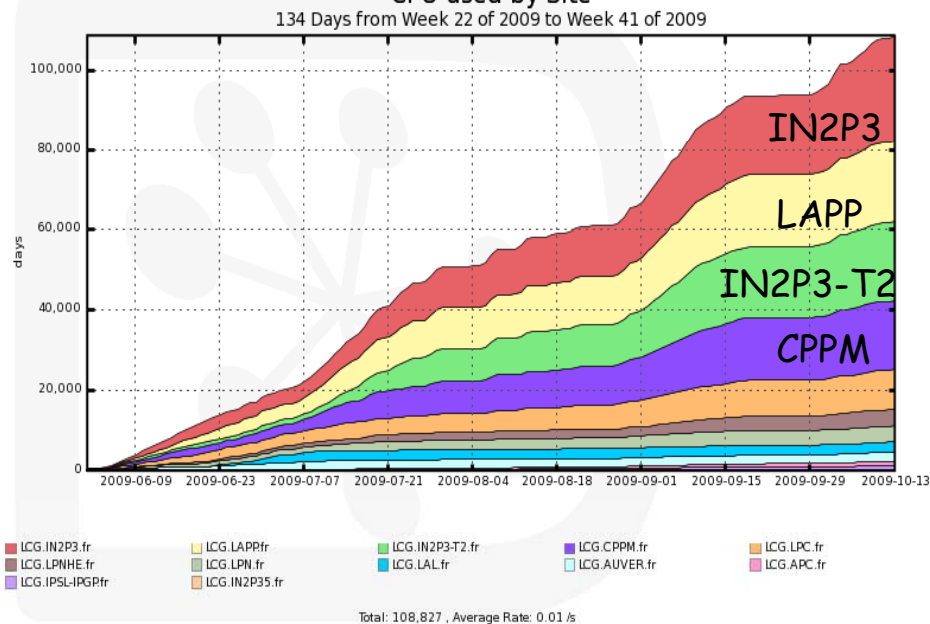


- 22 pays au total
- France : 13% du CPU utilisé

CPU used by JobType
134 Days from Week 22 of 2009 to Week 41 of 2009



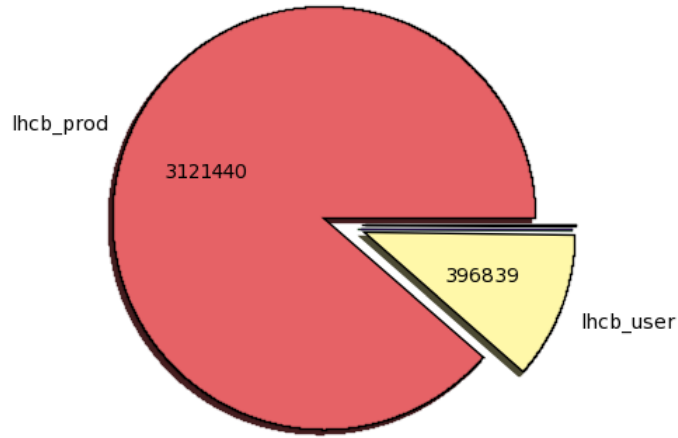
CPU used by Site
134 Days from Week 22 of 2009 to Week 41 of 2009





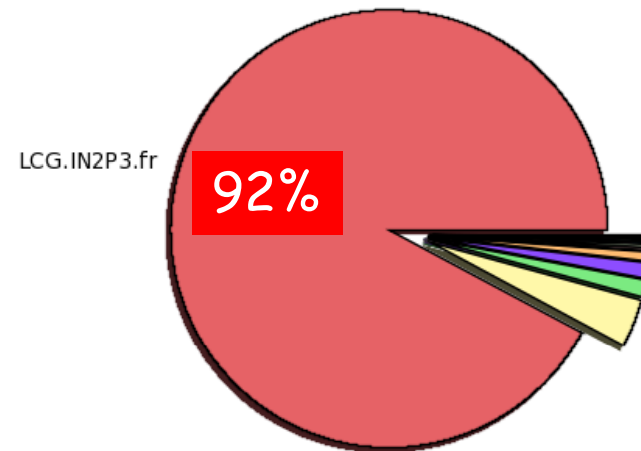
Statistiques depuis le 1er juin 2009

Total Number of Jobs by UserGroup (Sum: 3523493)



- Plus de 3.5 millions de jobs au total
- 11% sont des "vrais" jobs d'analyse

Jobs d'analyse en France



- 92% des jobs d'analyse au T1

LHCb @ LCC France

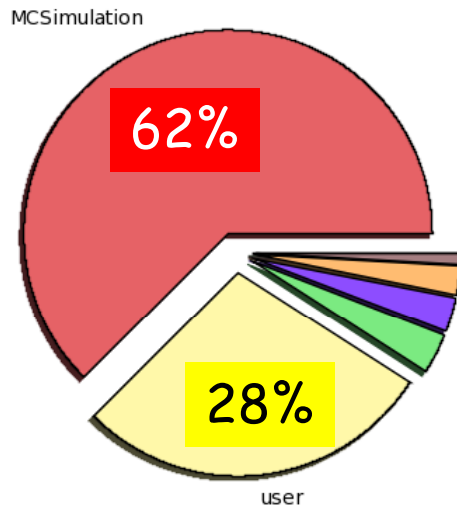


LCG.IN2P3.fr (47014)	LCG.IN2P3-T2.fr (1845)	LCG.LPC.fr (635)	LCG.LAPPfr (614)	LCG.LAL.fr (141)
LCG.LPN.fr (123)	LCG.APC.fr (94)	LCG.LPNHE.fr (82)	LCG.IN2P35.fr (16)	LCG.IPSL-IPGPr (25)



Statistiques depuis le 1er juin 2009

Total Number of Jobs by JobType (Sum: 165746)

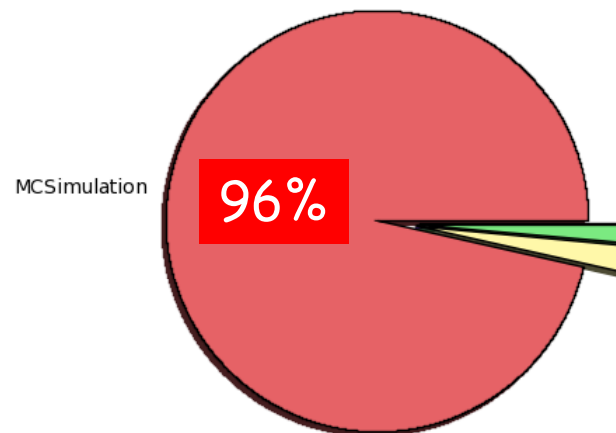


■ MCSimulation (103663)
 ■ user (47014)
 ■ Merge (5264)
■ DataReconstruction (3736)
 ■ sam (1415)
 ■ unknown (52)

Au T1 CC-IN2P3:

- 62% simulation
- 28% analyse
- 3% stripping
- 3% merging
- 2% reconstruction

Total Number of Jobs by JobType (Sum: 89265)



■ MCSimulation (86146)
 ■ user (1845)
 ■ sam (1272)
 ■ unknown (3)

Au T2 CC-IN2P3:

- 96% simulation
- 2% analyse

LHCb @ LHC





- Les besoins de l'analyse concernent essentiellement un accès aux fichiers fiable et performant (rôle majeur des T1s)
- Des tests de performance d'accès aux fichiers ont déjà permis de réajuster la configuration du système de stockage dans les différents sites
- Ce point nécessite un monitoring et des réajustements constants
- Les besoins de la simulation sont liés essentiellement aux ressources CPU (rôle majeur des T2s)