Profils des jobs LHC *(01-12-2014)*

Renaud Vernet
Journees LCG-Fr @ CEA Saclay

Introduction

- Le profil des jobs
  - Un des aspects d'un modèle de calcul
  - Comprendre les besoins
  - Situer l'utilisation des ressources vs SLA (MoU)
- Exercice non trivial
  - Monitoring VO vs monitorings site
  - Informations complémentaires, non exhaustives
  - Monitorings orientés 'global' plutot que 'par job'
  - Jobs pilotes
- Cette présentation
  - Basée sur observations sur quelques sites (surtout CC)
  - Donne quelques idées des grandeurs-type des profils de jobs
  - Montre quelles infos sont disponibles si besoin

Monitoring VO

- ALICE
  - MonaLisa *http://alimonitor.cern.ch*
  - Niveau de detail élevé, peu de vision par activité
- ATLAS et CMS
  - Dashboards
    - *http://dashb-atlas-job.cern.ch* et *http://dashb-cms-jobsmry.cern.ch/dashboard/request.py/dailysummary*
  - Oriente 'activités', vue globale
- LHCb
  - Dirac http://dashb-atlas-job.cern.ch
  - Controle d'accès

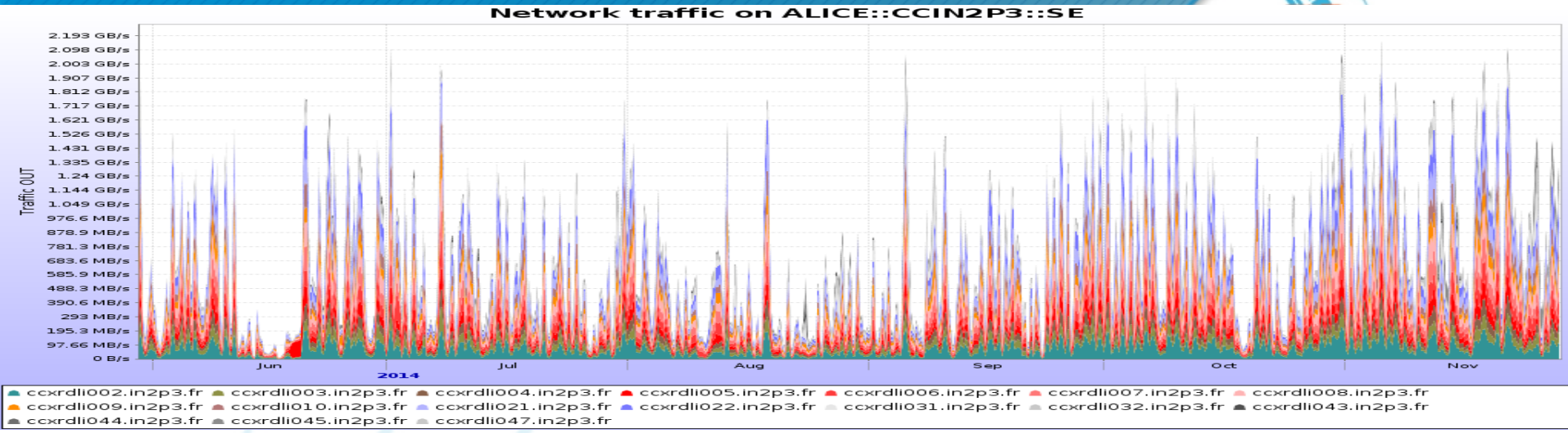
Monitoring sites

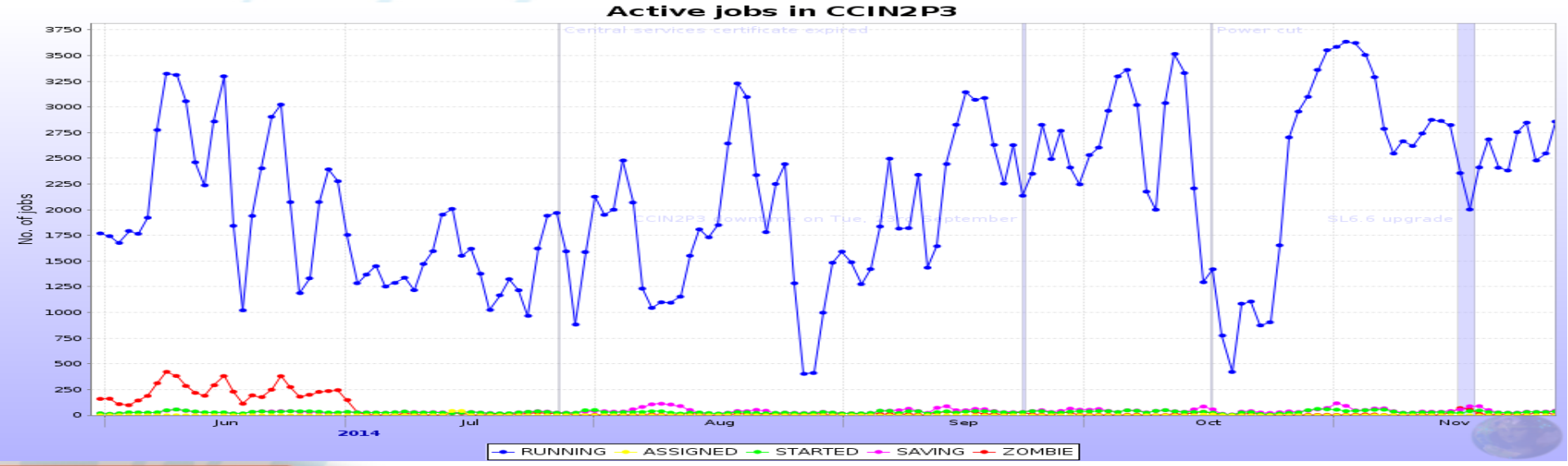
- T2/T3
  - Monitoring principalement système
    - Objectif : santé des machines et services
    - Vue globale (mémoire, trafic) par machine
    - Peu adapté pour faire du profiling des jobs
  - Monitoring batch
    - Nombre de jobs par groupe (running, pending...)
- T1
  - Mêmes outils que pour T2/T3
  - Monitoring spécifique *http://symod.in2p3.fr*
    - Notamment basé sur l'accounting GE

Grandeurs observées

- Trafic réseau
- Wallclock time
- Mémoire
- Efficacité CPU

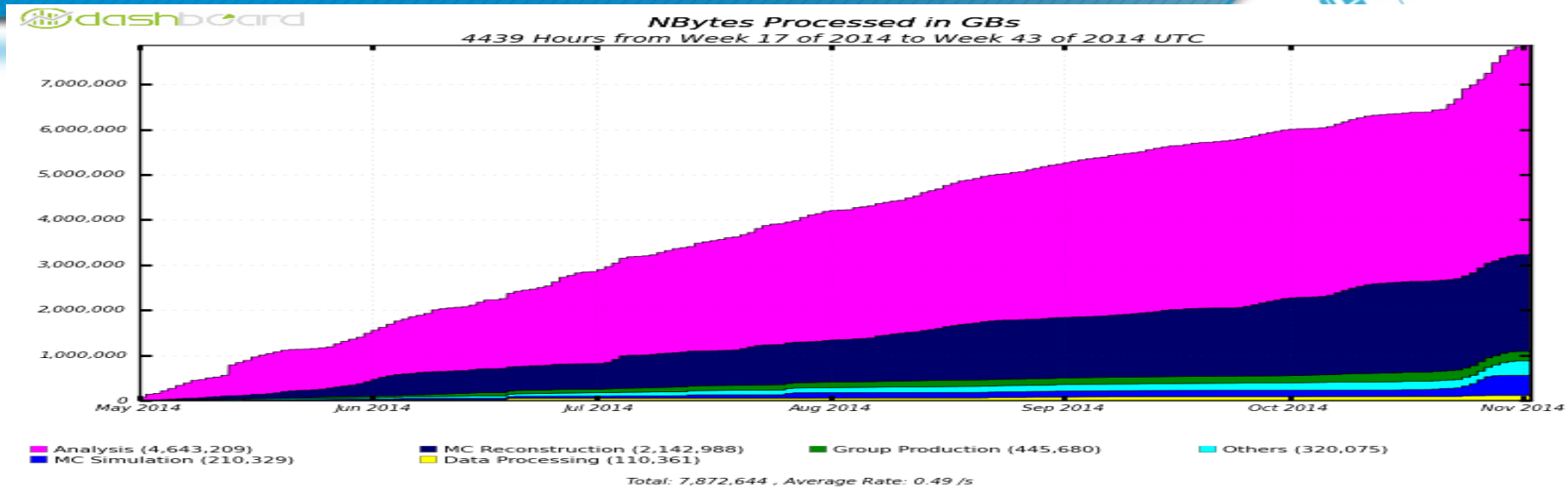
ALICE : trafic réseau

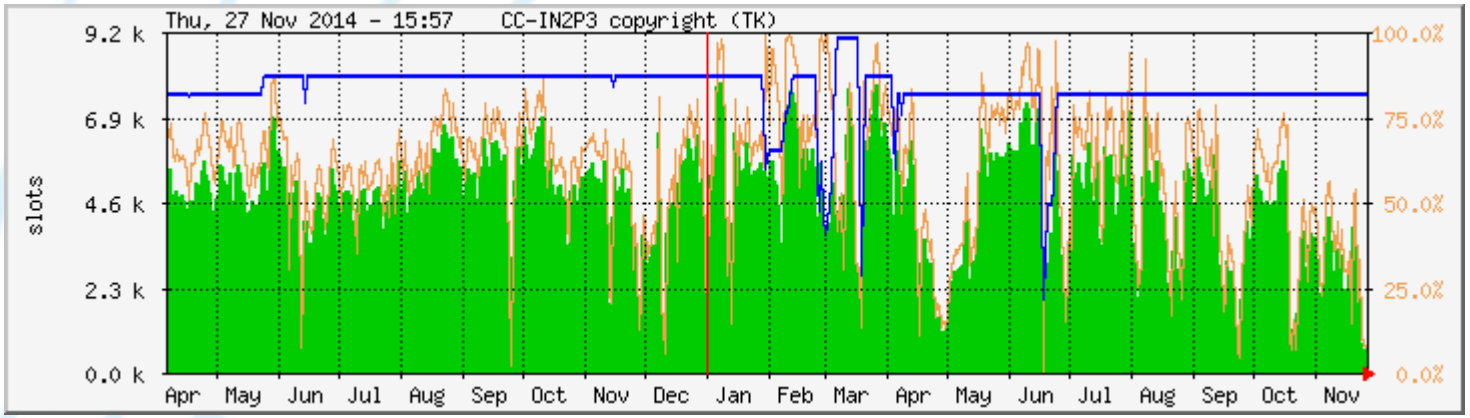




2120 jobs, 750 MB/s => ~0.4 MB/s/job
Fluctuations importantes

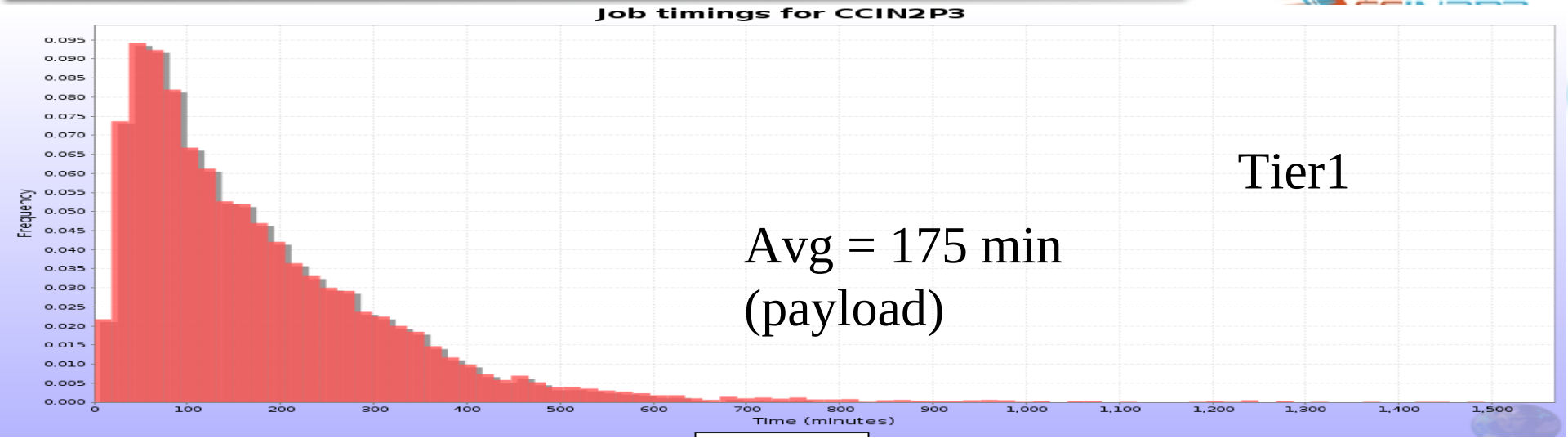
ATLAS : trafic réseau

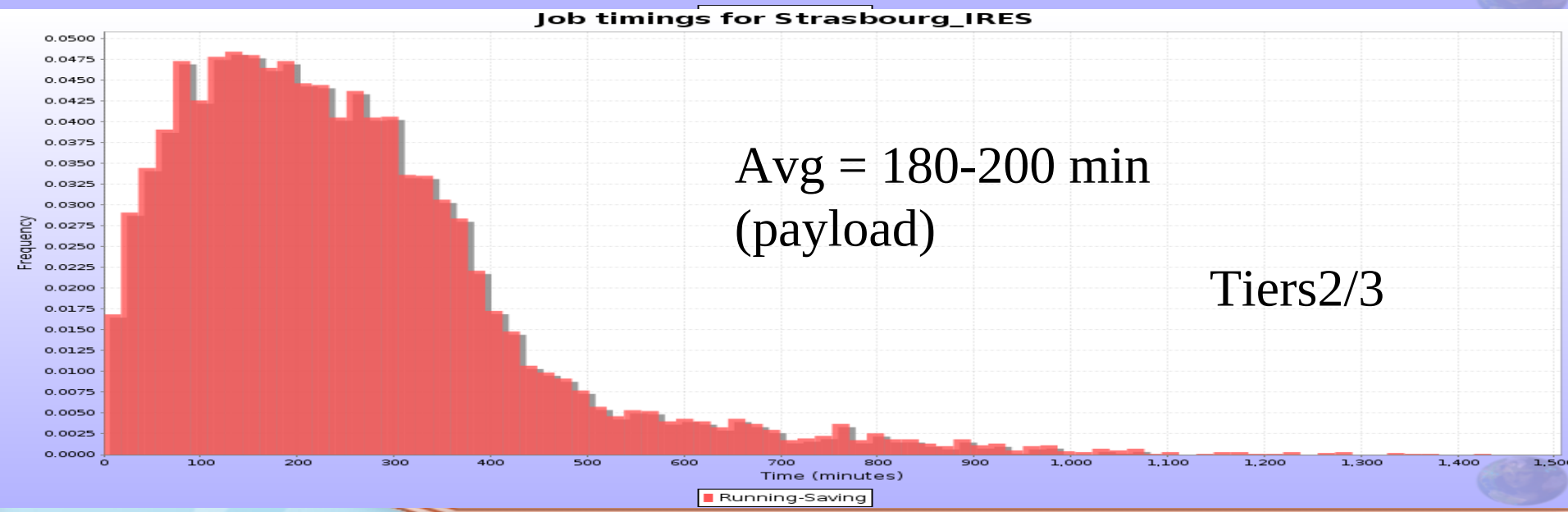




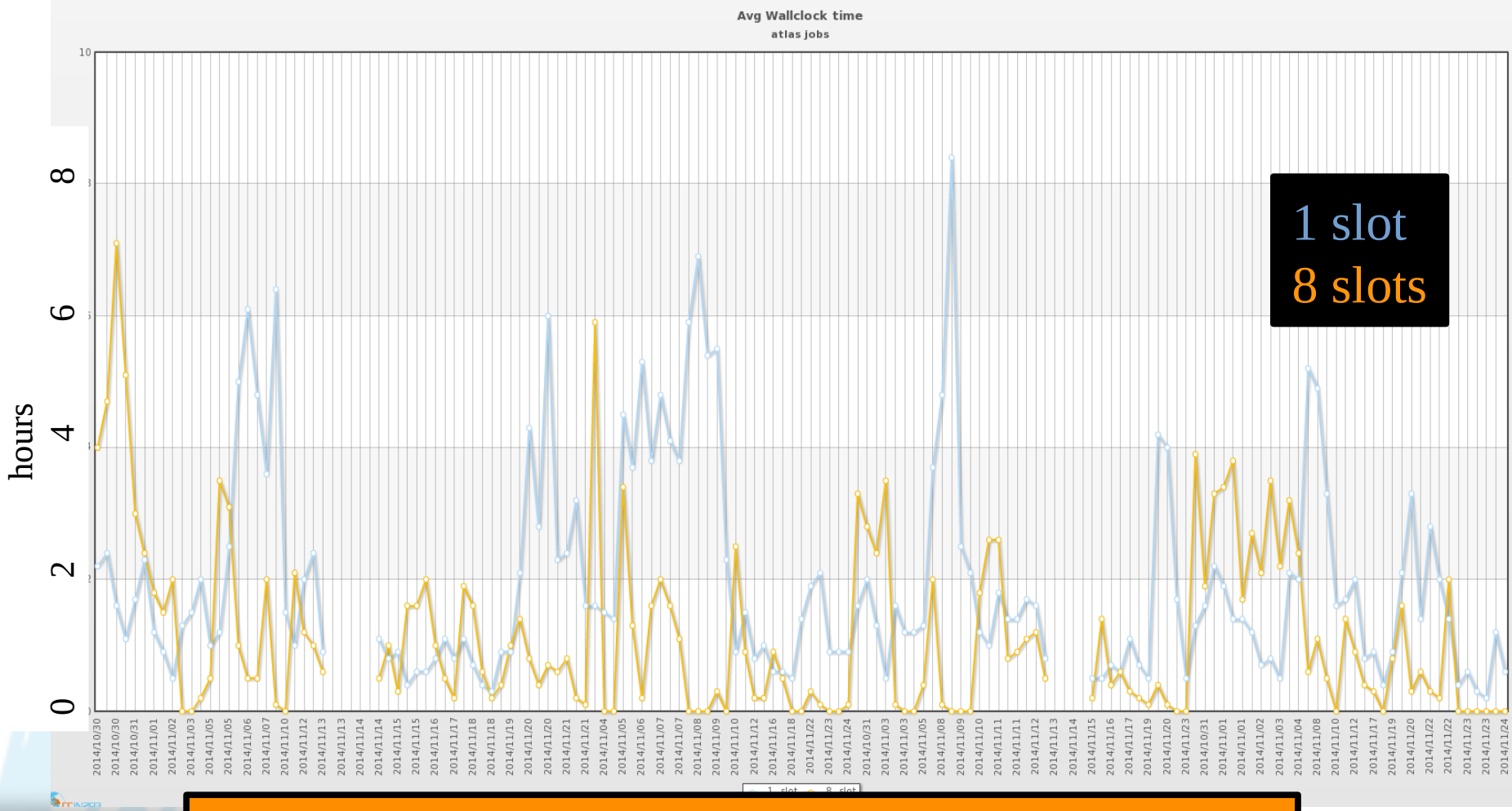
4500 jobs, 7 PB sur 6 mois => ~0.1 MB/s/job

ALICE : job wallclock distribution



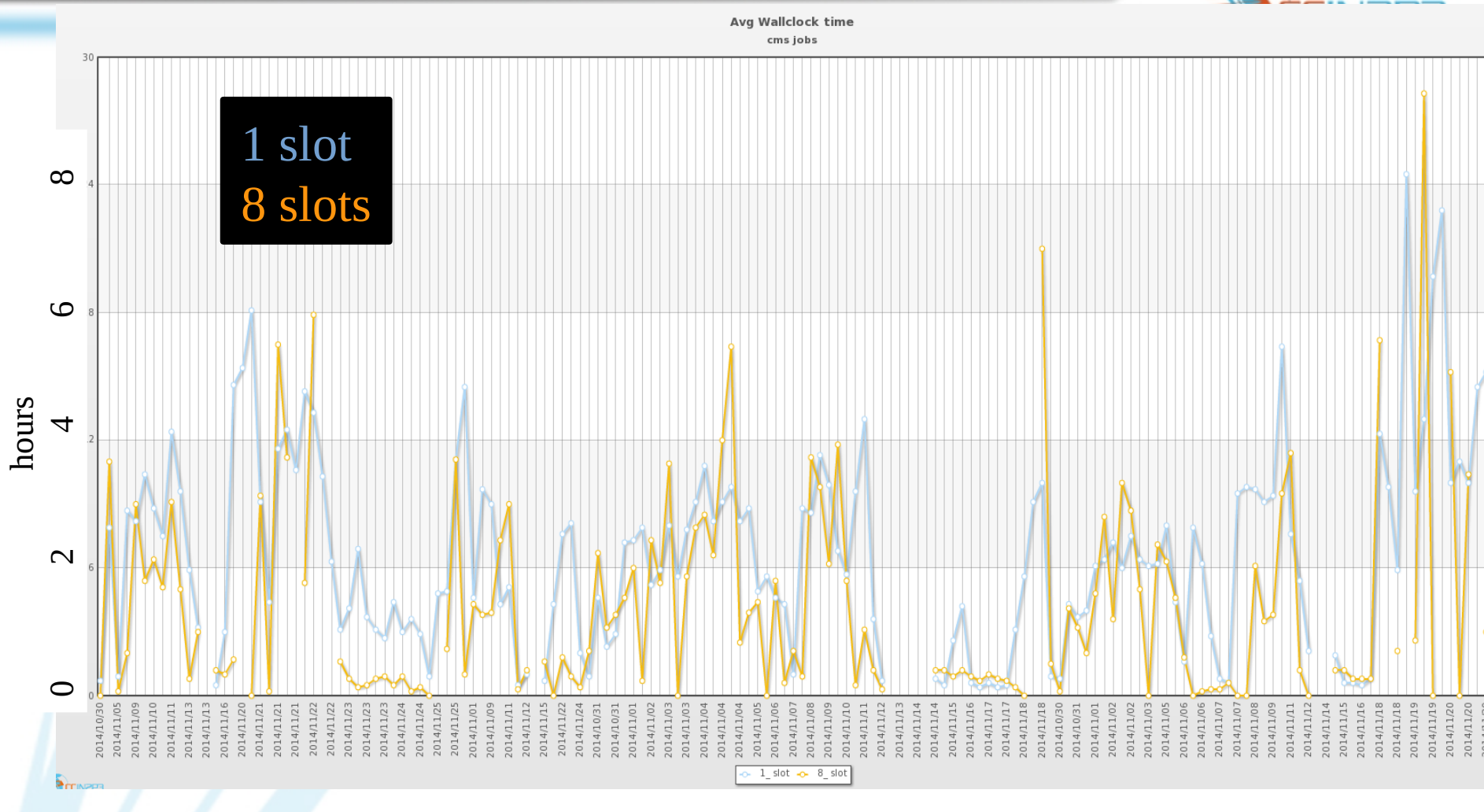


ATLAS : wallclock (pilotes)



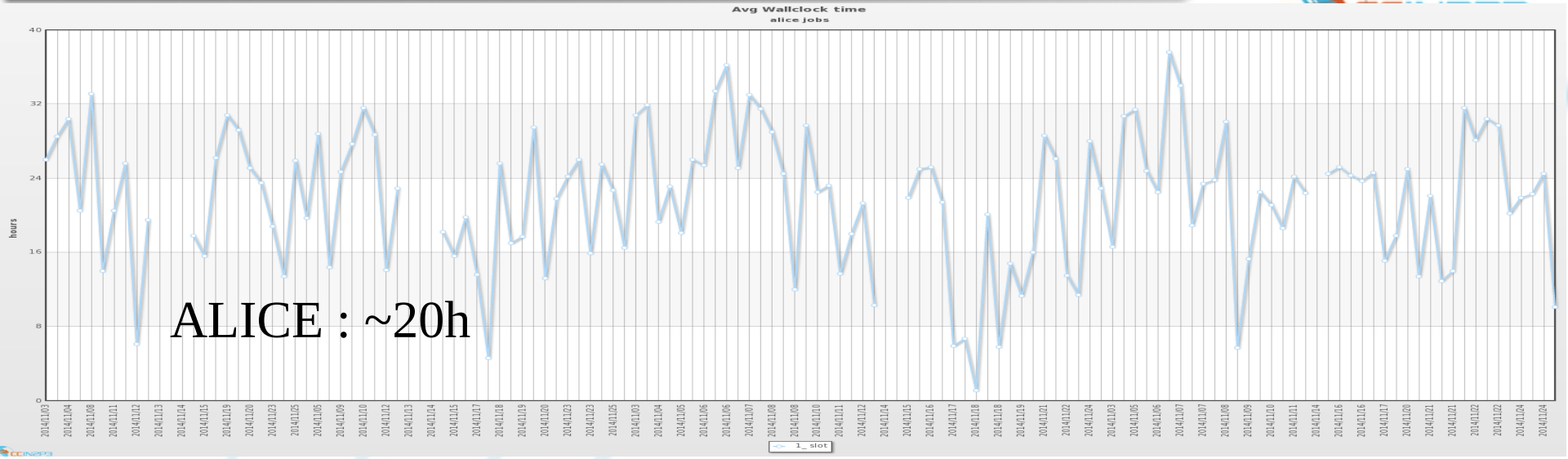
Sensiblement même comportement (sequ. vs mc)
~2h / job

CMS : wallclock (pilotes)

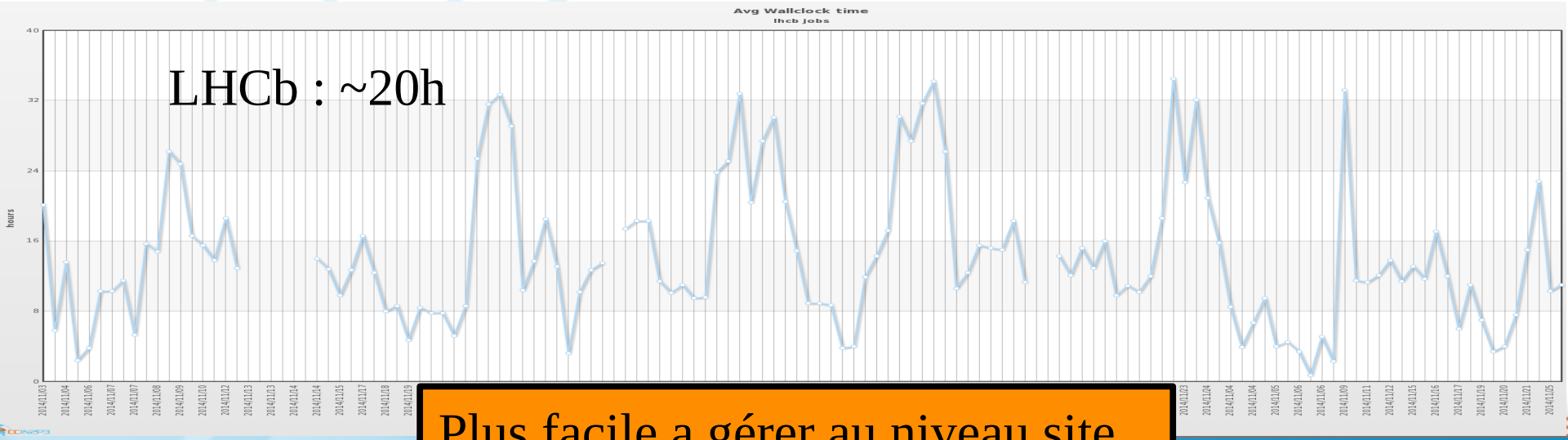


Sensiblement même comportement (sequ. vs mc)
~2h / job

ALICE, LHCB : wallclock (pilotes)



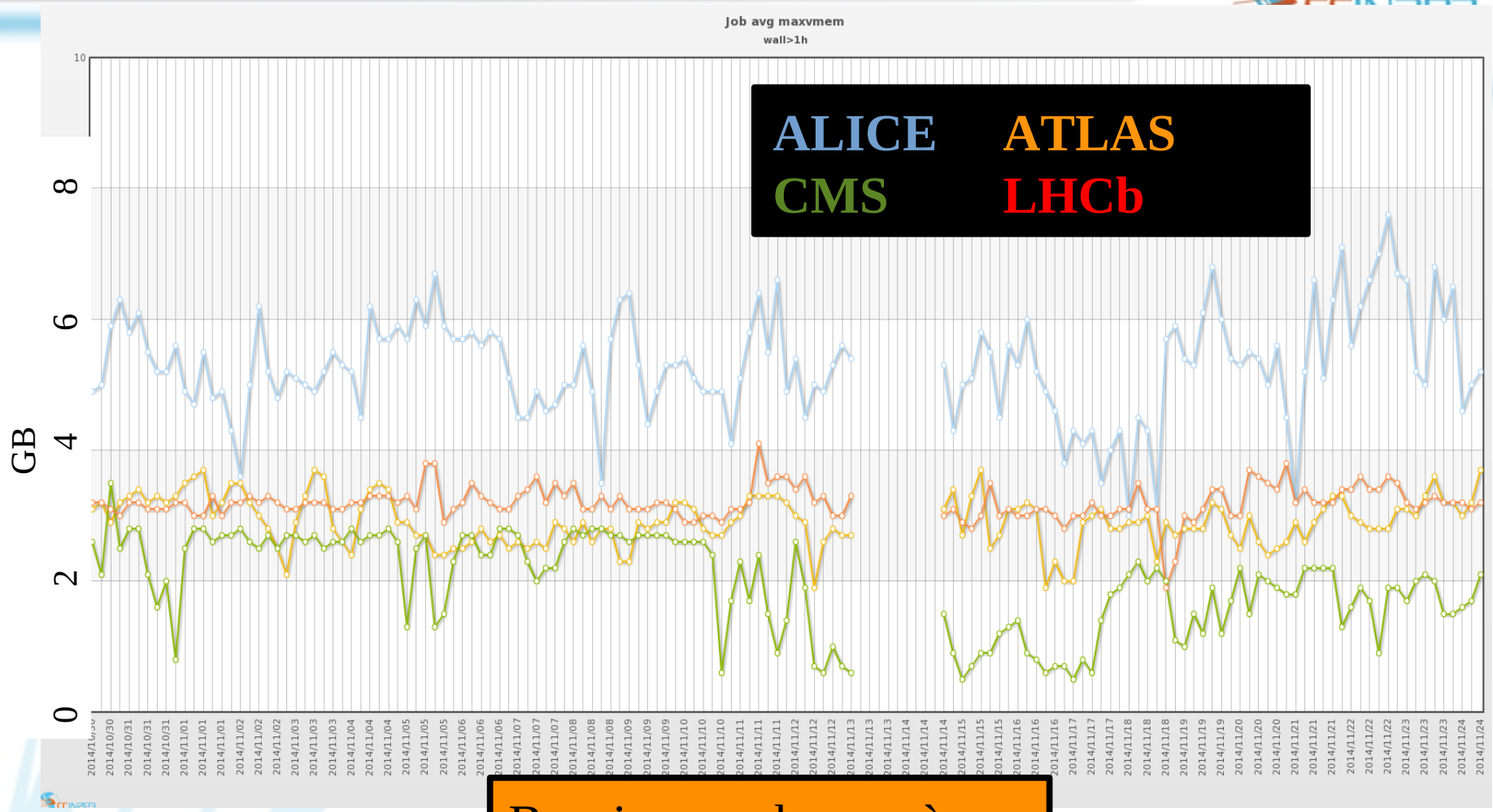
ALICE : ~20h



LHCb : ~20h

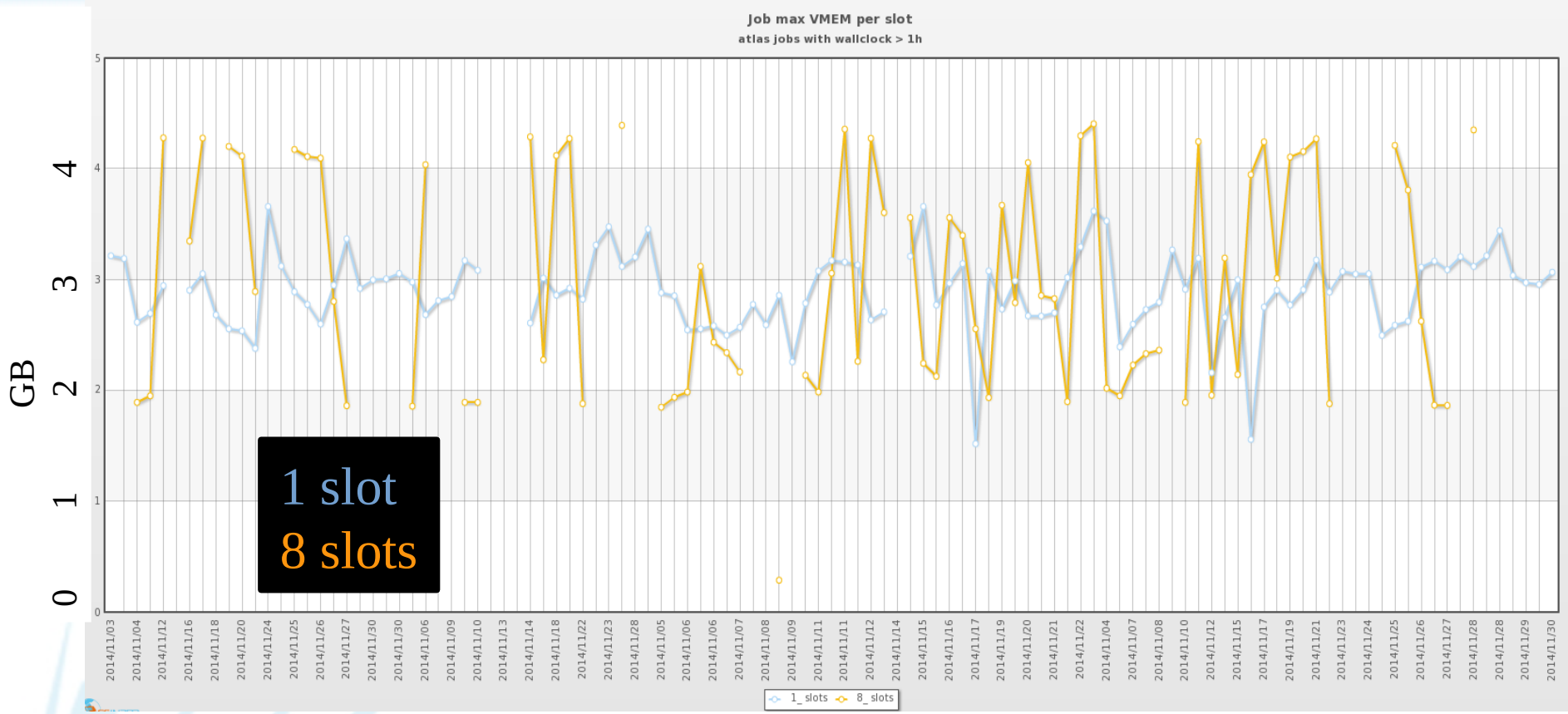
Plus facile a gérer au niveau site

max(VMEM) per slot



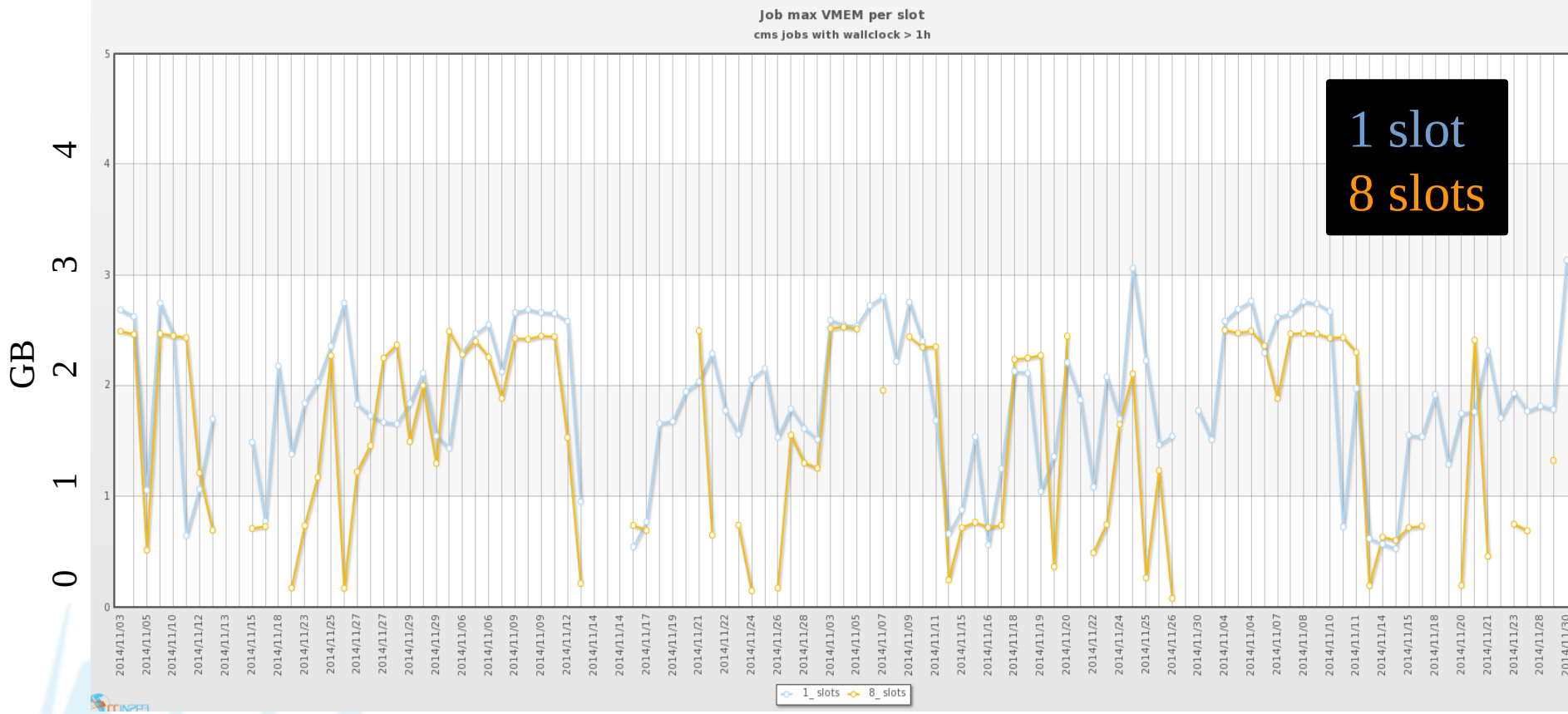
Besoins peu homogène

ATLAS : max(VMEM) par slot



Baisse de consommation par multicore... pas évidente

CMS: max(VMEM) par slot



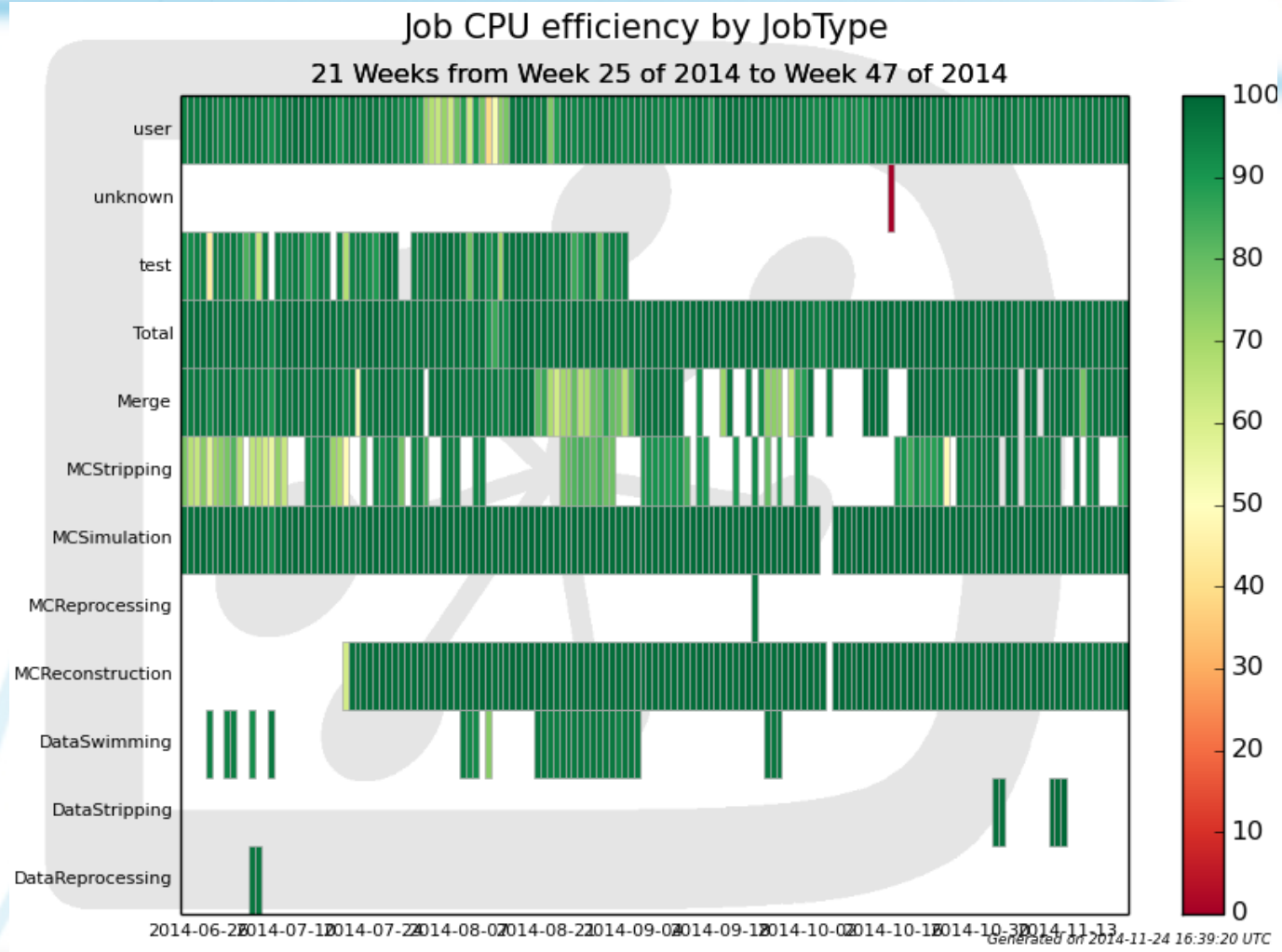
Baisse de consommation par multicore... légère ?

Mémoire

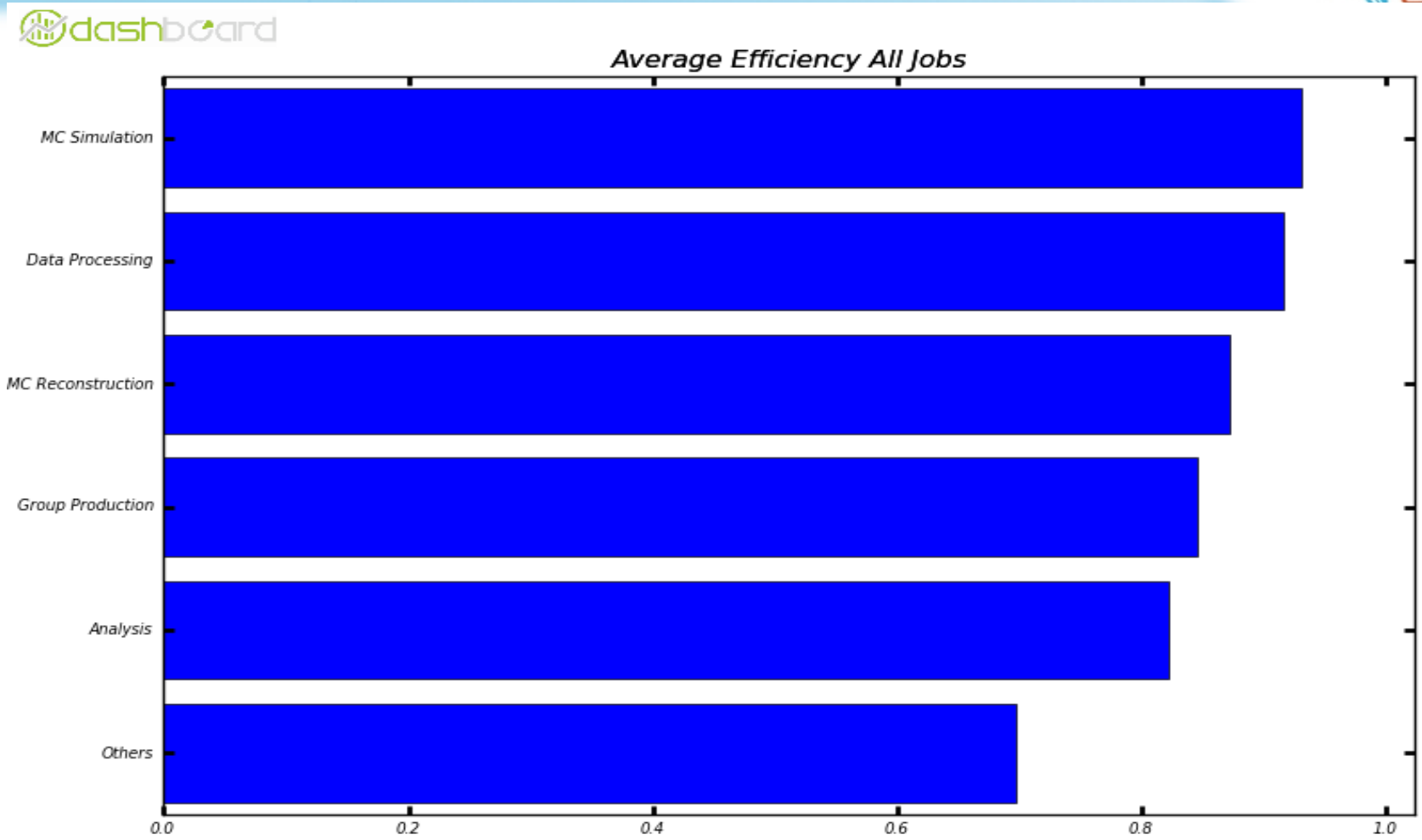
- Profils précédents valables au CCin2p3 avec VMEM

  Et les conclusions subséquentes aussi
- Profils plutôt plats pour ATLAS et CMS
  - Faibles fluctuations
  - ALICE et LHCb plus erratiques
- Consommation ALICE élevée
  - Améliorations récentes
    - Reste un problème potentiel dans petits sites (?)
    - Etude avec mémoire résidente pourrait être intéressante
    - Autre grandeur pertinente ? (PSS…)

LHCb : efficacité CPU par activité

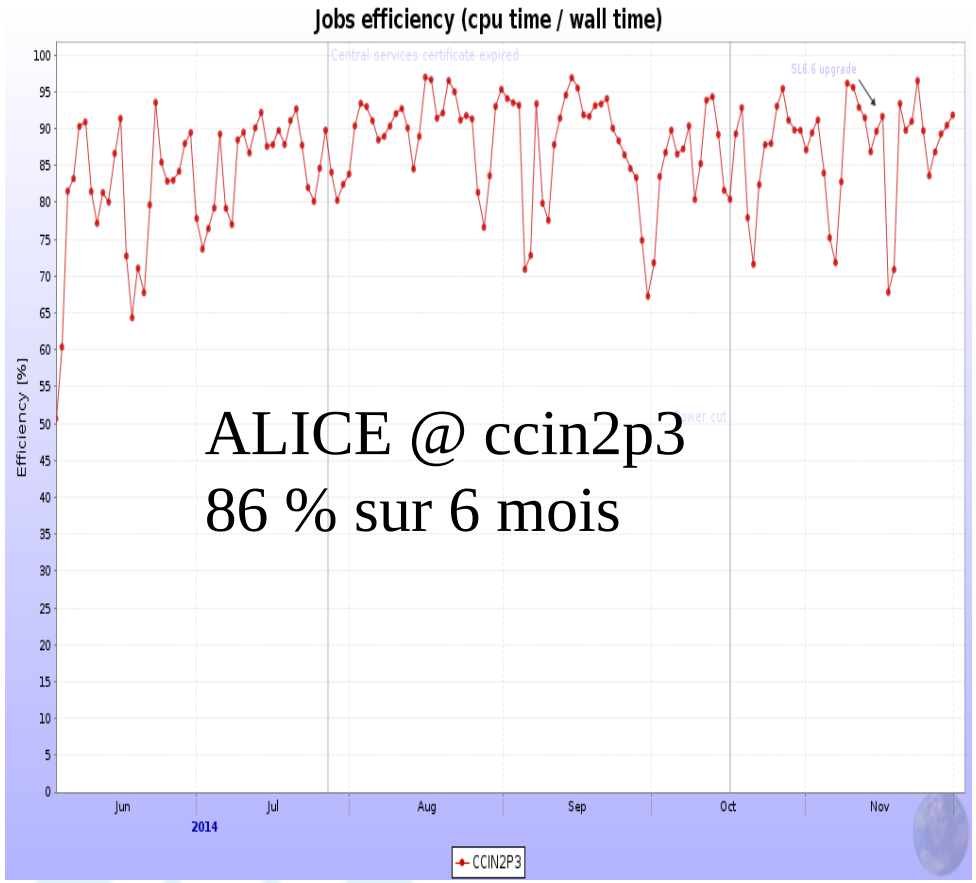


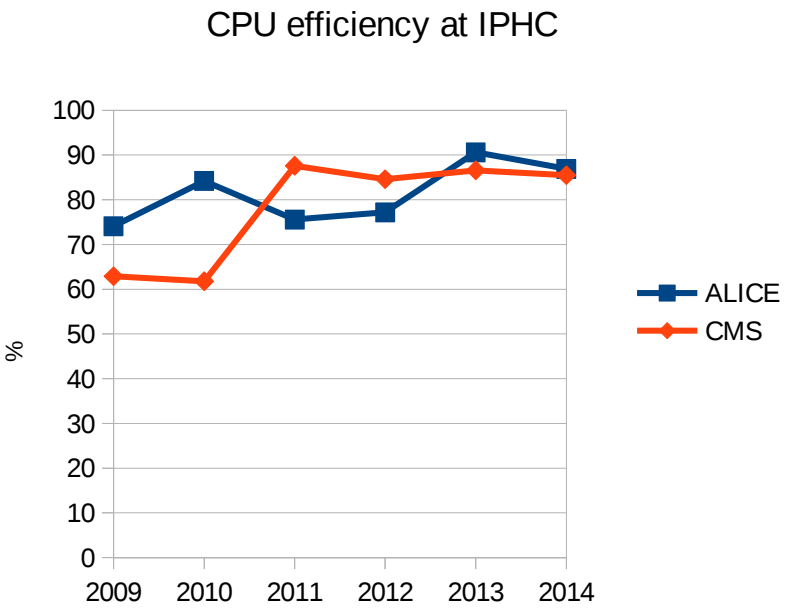
ATLAS : efficacité CPU par activité



Activites principales : eff_cpu > 80 %

ALICE


Jobs efficiency (cpu time / wall time)
Central services certificate expired
SL6.6 upgrade
Efficiency [%]
Jun
Jul
Aug
Sep
Oct
Nov
2014
CCIN2P3
ALICE @ ccin2p3
86 % sur 6 mois


CPU efficiency at IPHC
%
2009
2010
2011
2012
2013
2014
ALICE
CMS

Conclusions

- Exercice de profiling des jobs pas facile
  - Accès au au détail parfois compliqué
  - Ce serait bien si les expériences pouvaient fournir ce genre d'estimations aux sites pour années à venir (besoins réseau notamment...)
- Dasboards CERN plutôt orientés 'expérience'
  - Quelque intérêt pour sites cependant
- Déclenchement discussion chez CMS
  - ajouter données dashbd pour meilleur profiling...
- Jobs ALICE et LHCb plus longs
  - Mieux pour le batch system
- Pas de différence notable entre jobs sequentiels et jobs multi-core *(dans les grandeurs considerées ici du moins)*
- Mémoire toujours un problème potentiel avec ALICE
  - en amelioration

Credits

- Merci a Christine, Edith, Denis, Jerome, Jean-Michel, Jean-Claude, Eric, Sebastien, Aresh, Emmanouil pour leurs infos :-)