

Activités LHC au CC-IN2P3

Renaud Vernet

Subatech , September 19th 2012



Plan



- Quelques nouvelles du LHC
- Les ressources du T1 français et leur utilisation
- LAF
- Nouvelles du site



Nouvelles du LHC



- Véritable succès du LHC
- La Grille de calcul en est un des fondements !
- Bonne collaboration entre VO et sites
- Transmission du savoir essentiel
 - Des physiciens vers les sites
 - Et vice versa !
- Nouvelles technologies à suivre
 - Cloud, multi-core, federation...
- Les 4 expériences préparent leur upgrade pour l'après-LSD
 - → nouvelles conditions, ce sera un défi pour le computing aussi !

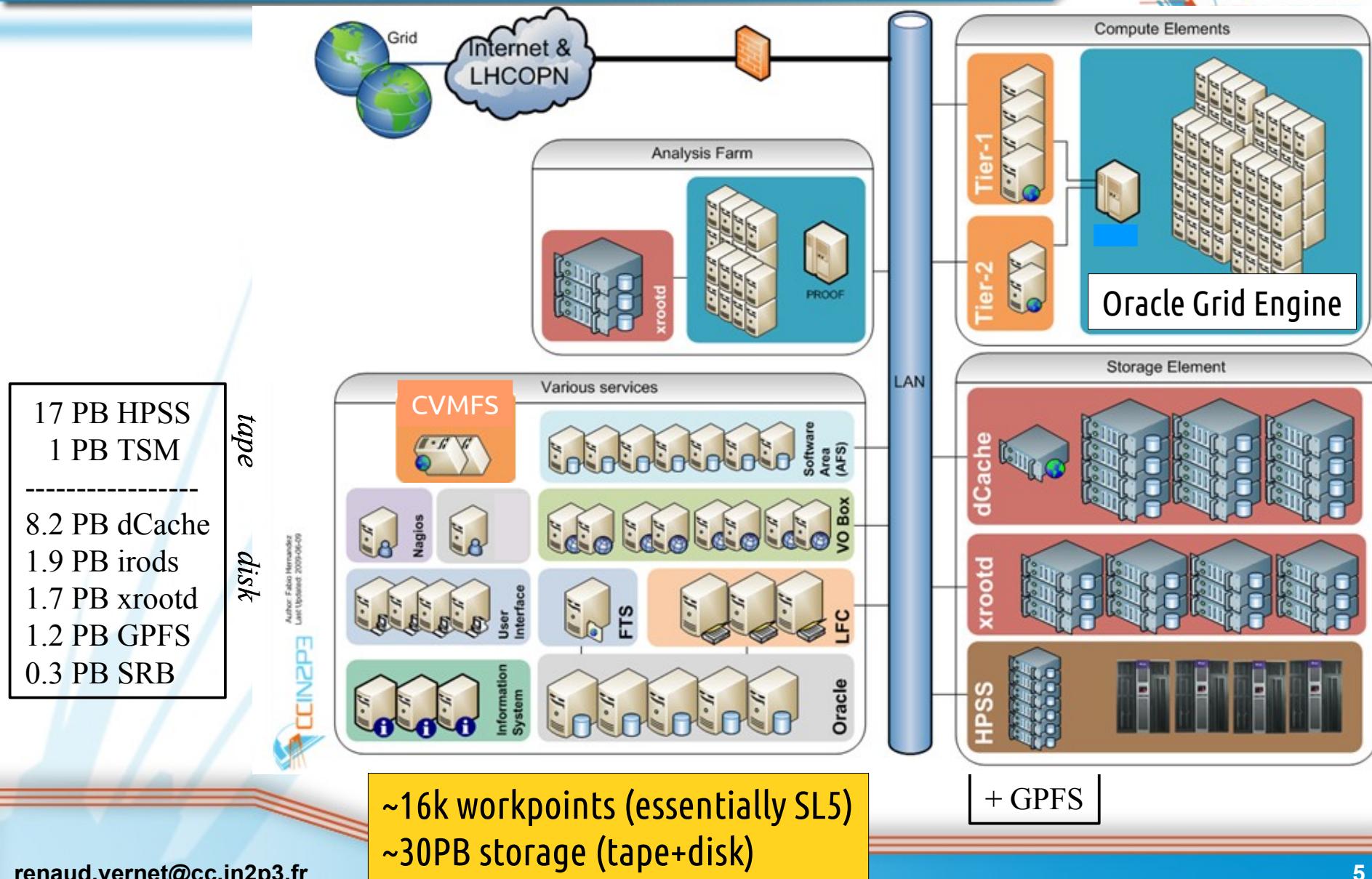
Nouvelles du LHC (2)



■ Extension du run 2012

- p-p jusqu'à mi-décembre
 - Accumuler davantage de statistique avec $\sqrt{s}=8\text{TeV}$ pour améliorer études en cours
 - Dernier run avant l'augmentation energie faisceau à $\sim 14\text{ TeV}$ apres shutdown
- p-Pb en janvier-février
 - physique des ions lourds
- Conséquences sur ressources
 - Mais pas de mise à jour des pledges prévue
 - Demandes des VO difficiles (impossibles) à satisfaire
- « best effort » de la part des sites et des VO

Ressource du CC-IN2P3



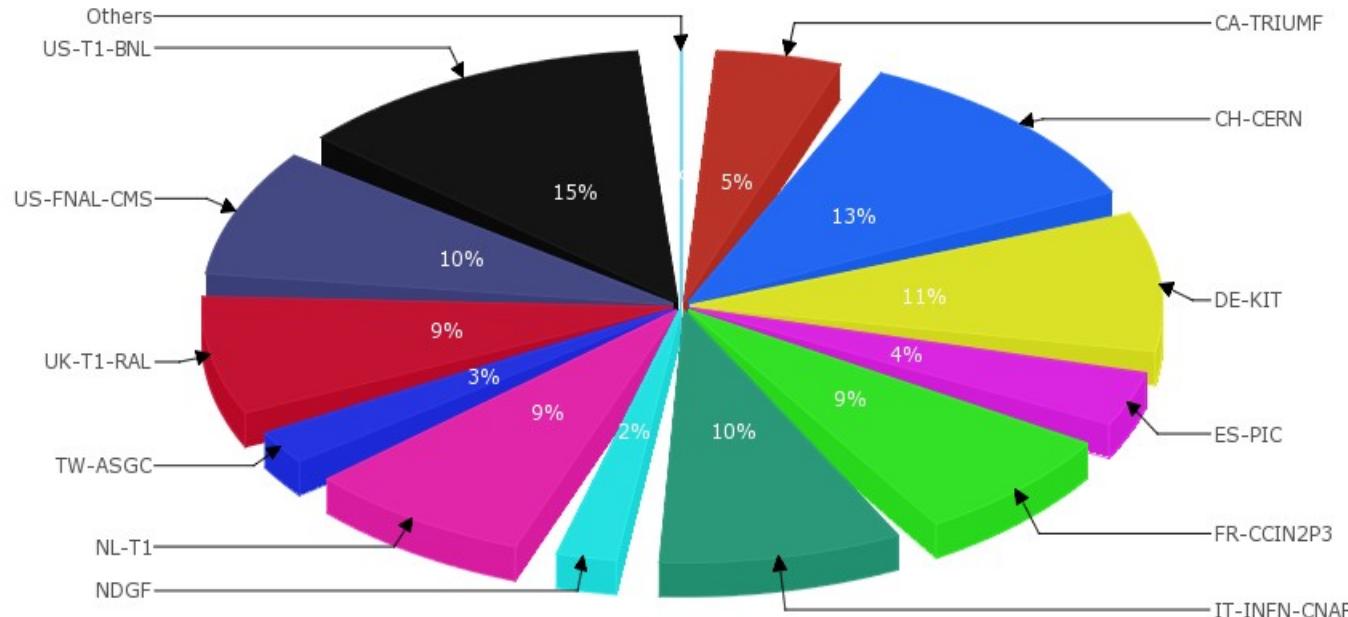
Contributions CPU des T1 (*depuis un an*)



Developed by CESGA 'EGI View': / normcpu-HEPSPEC06 / 2011:9-2012:8 / TIER1-VO / lhc (x) / GRBAR-LIN / i

2012-09-17 02:50

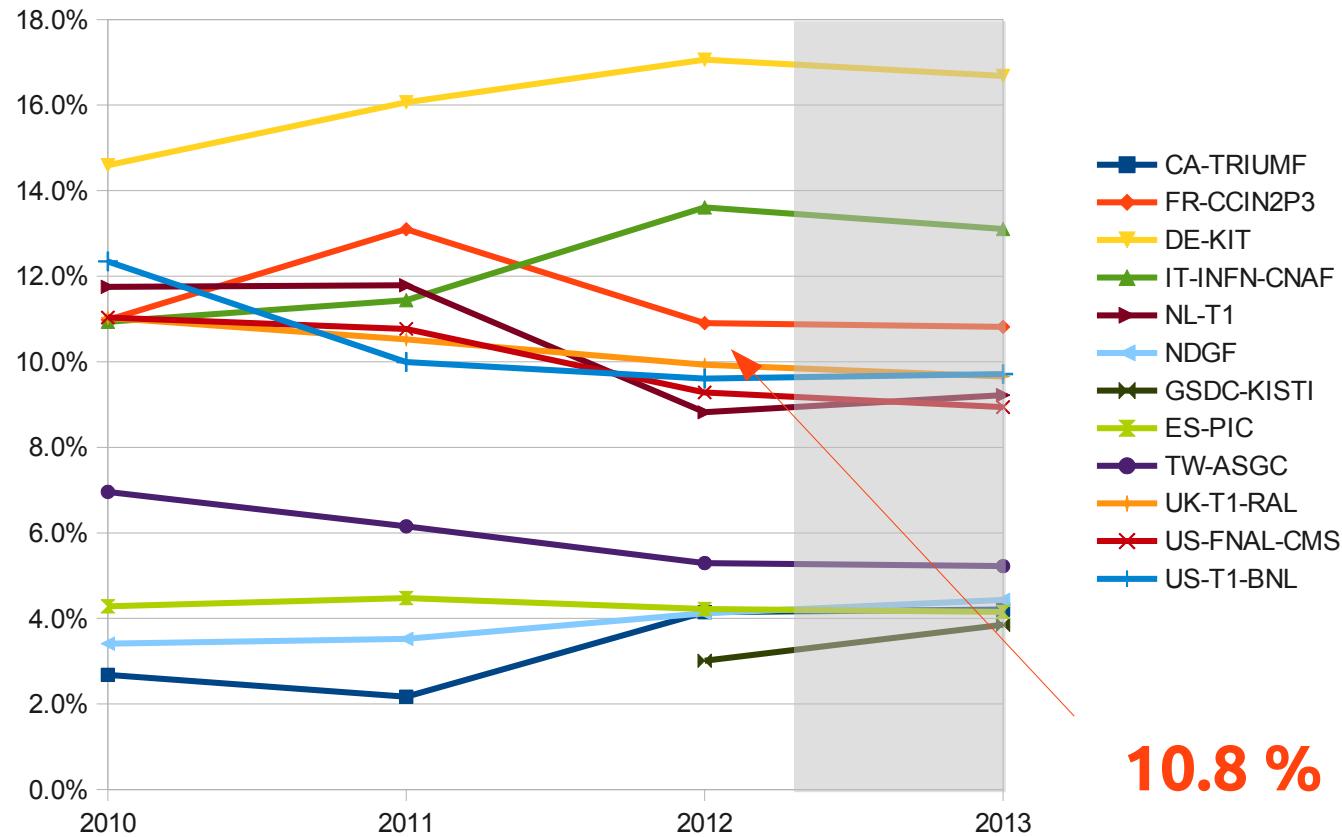
TIER1 Normalised CPU time (HEPSPEC06) per TIER1



Pledges T1 : CPU



T1 shares (CPU)

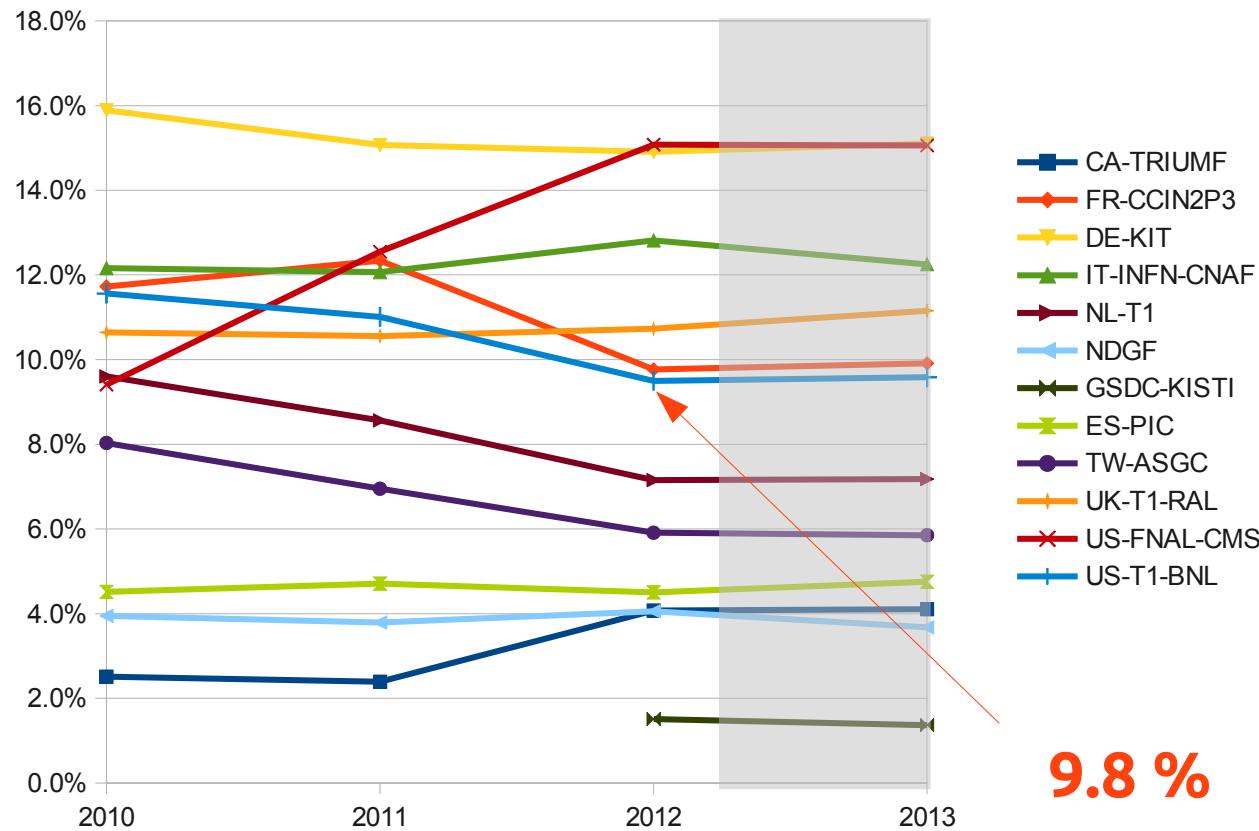


10.8 %

Pledges T1 : disque



T1 shares (Disk)

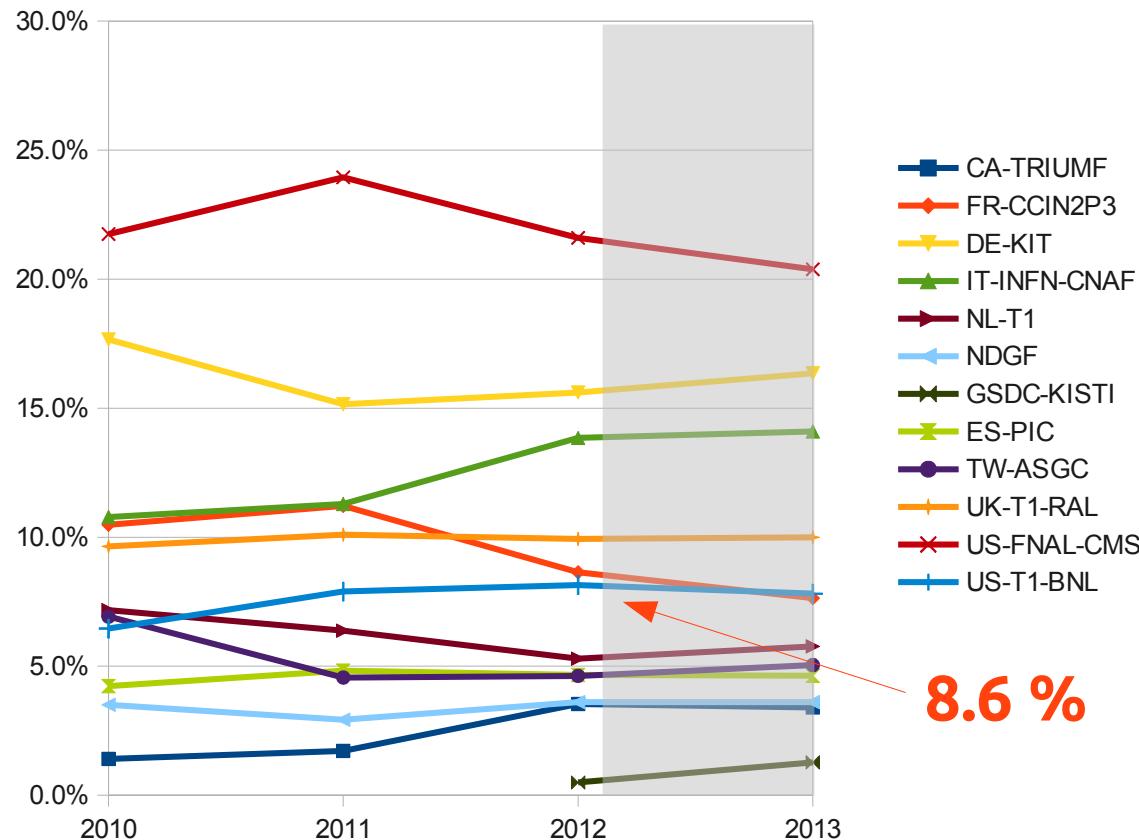


9.8 %

Pledges T1 : bande

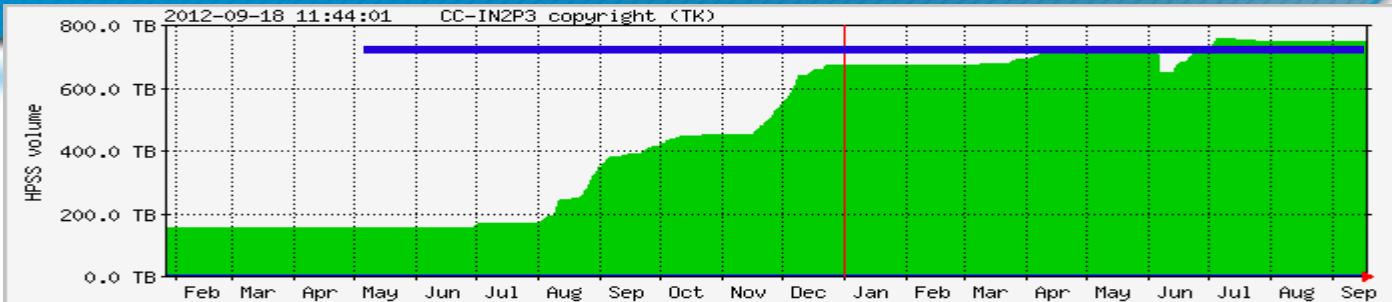


T1 shares (Tape)

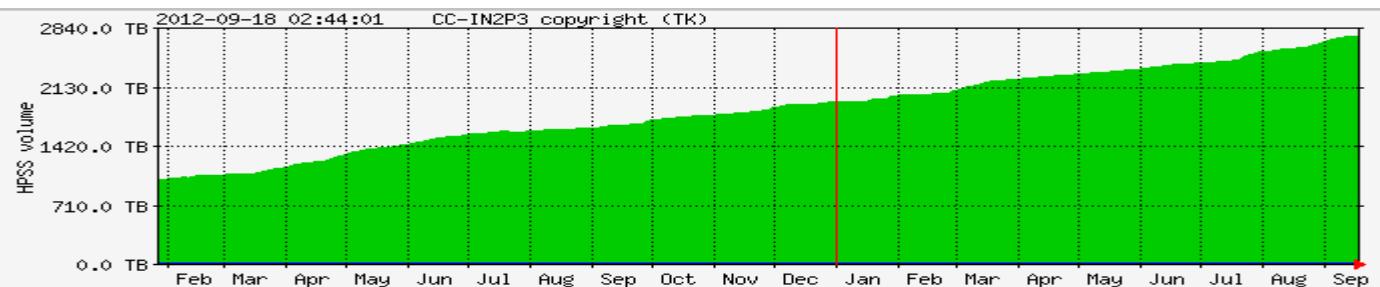


Le CC-IN2P3 a perdu de son importance dans les T1 l'annee dernière
Previsions peu optimistes pour les années à venir

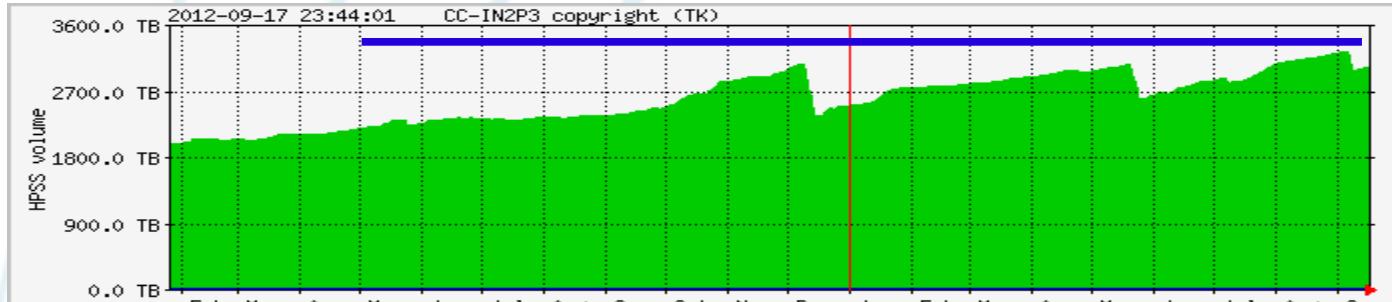
Remplissage des bandes



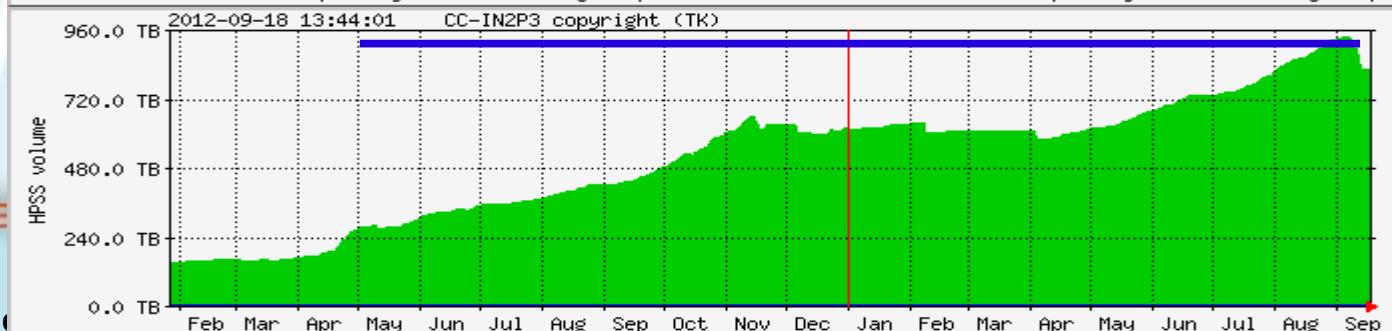
ALICE



ATLAS



CMS



LHCb

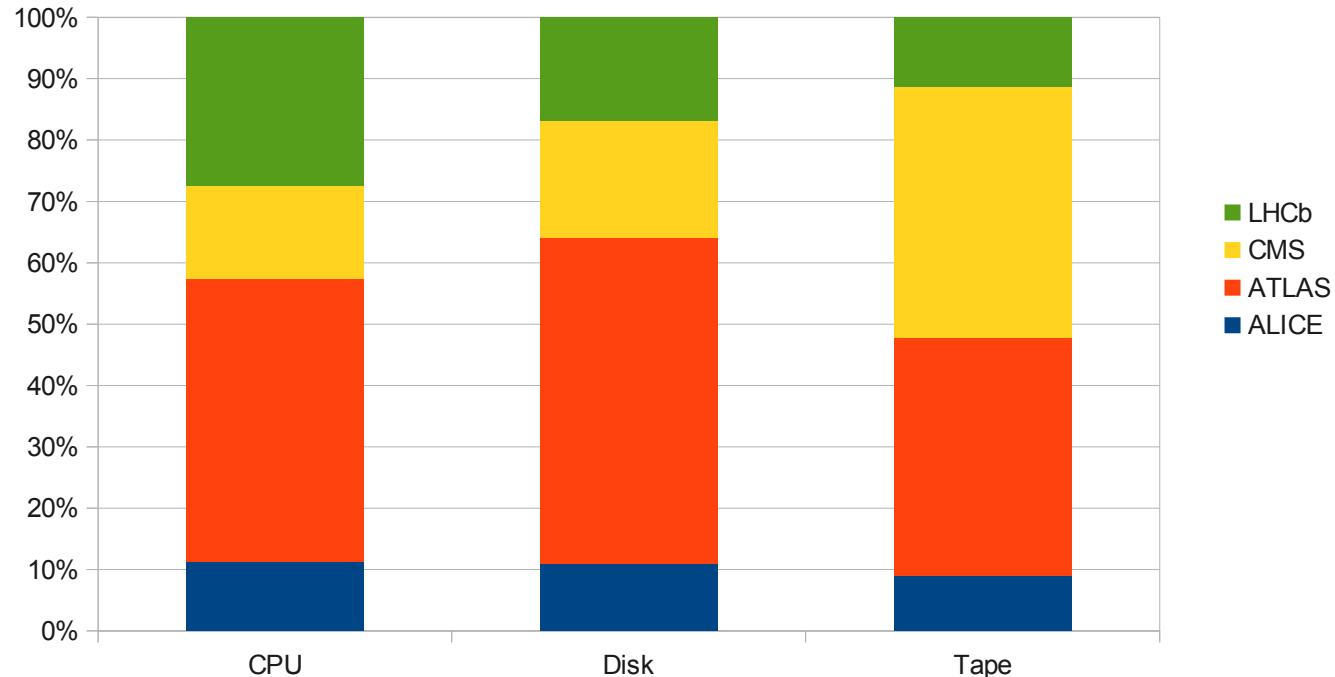
Repartition des ressources par VO au CC



2012

Resource sharing at CCIN2P3

Tier1 pledges

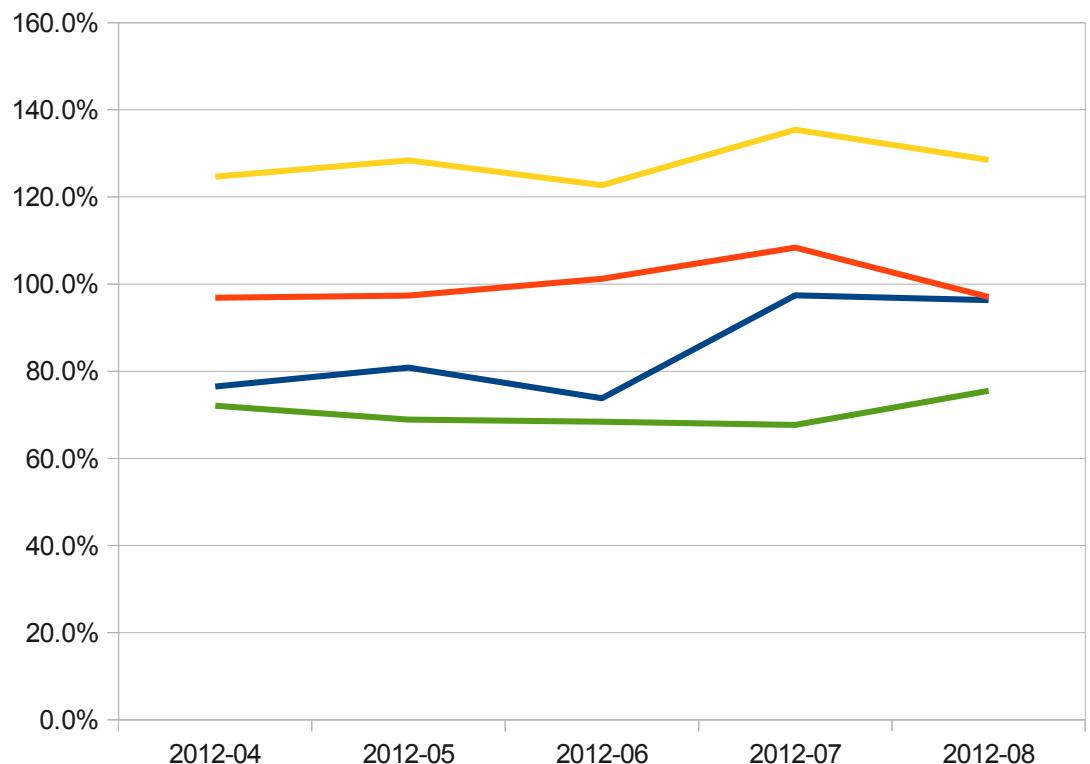


Utilisation des ressources CPU



Wall / pledges

T1+T2



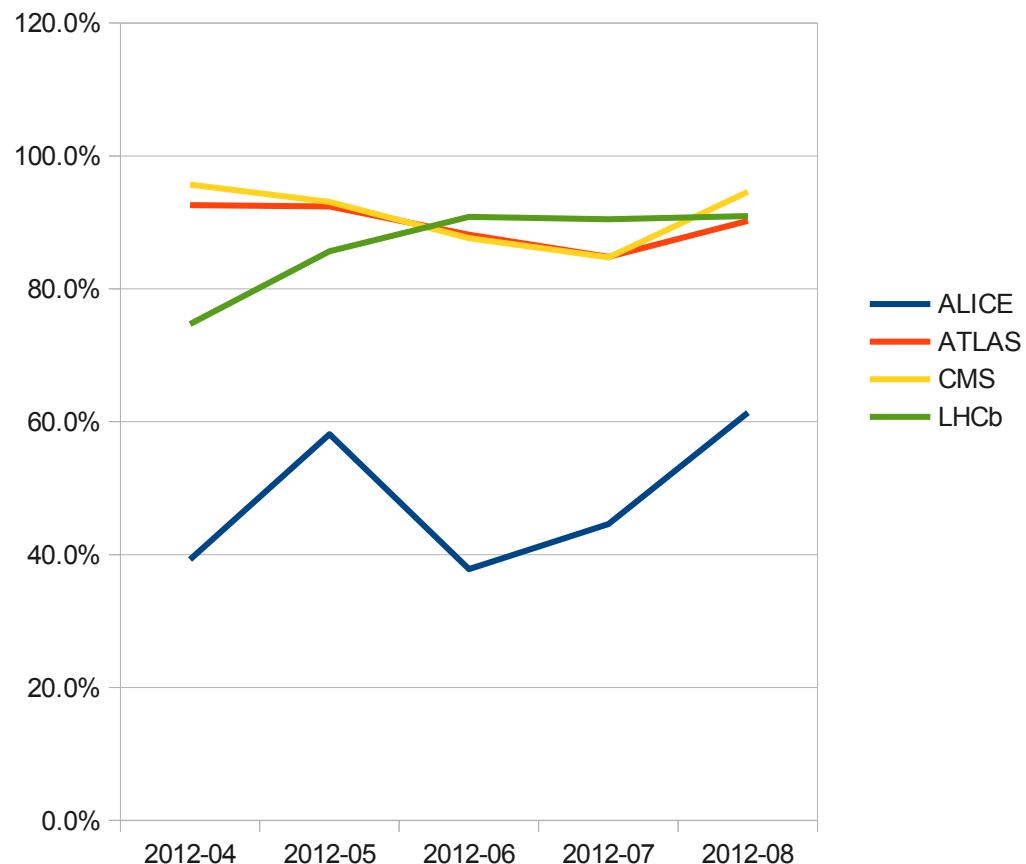
- ALICE : nous devons ameliorer
- CMS : tres opportuniste
- ATLAS : OK
- LCHb : plusieurs problemes en cours

Efficacite CPU recente



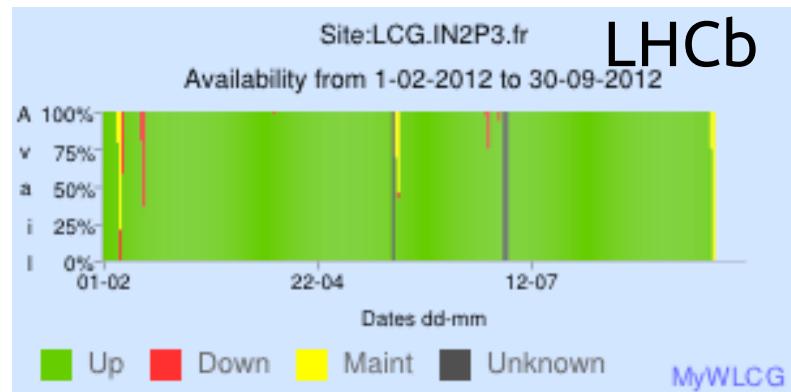
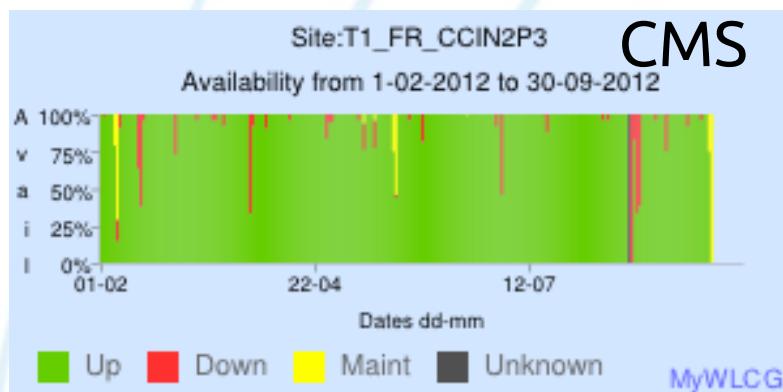
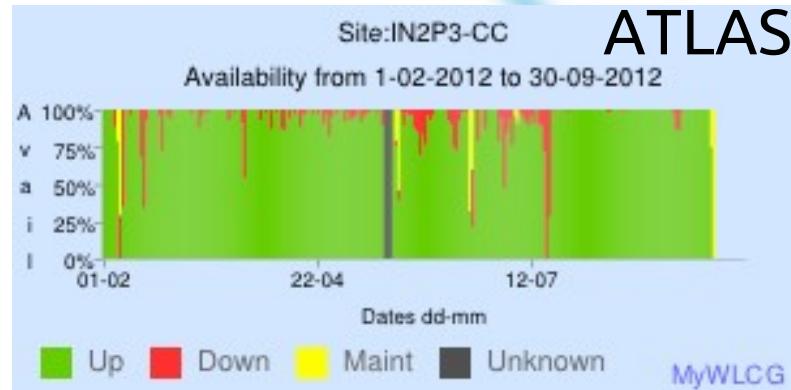
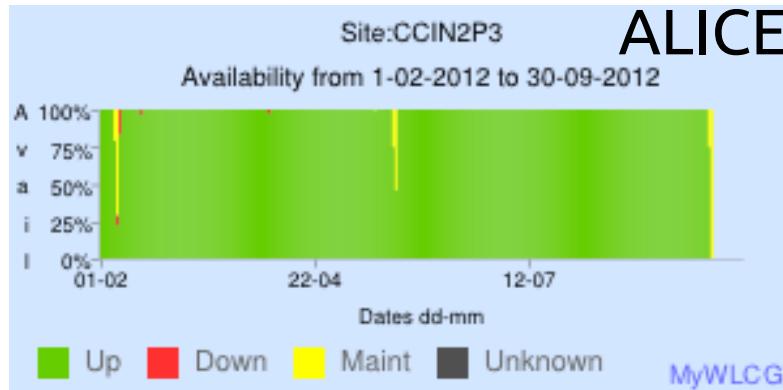
CPU efficiency

T1+T2



- ALICE : faible efficacite CPU des jobs
Non-simulation
Importante proportion analyse
Acces donnees a distance

Disponibilité du site depuis janvier





Nouvelles du staff LCG au CC



- Pierre Girard nous a quittes pour d'autres horizons
 - Tache de « representant T1 » assuree au CC par moi-même depuis juin dernier
- Nouvelle grid admin: Vanessa Hamar
 - Reprend activite systeme de Pierre
 - Complete l'équipe avec Ch. Eloto et J. Garnier
- Support des VO LHC
 - ATLAS → Ghita Rahal, Emmanouil Vamvakopoulos
 - CMS → Sébastien Gadrat
 - LHCb → Aresh Vedaee
 - ALICE → R.V.
- Ce sont les supports dedies
 - → ils connaissent mieux que personne les activités de la VO au CC
 - → ils doivent être vos interlocuteurs privilégiés

Migration middleware



- <https://twiki.cern.ch/twiki/bin/view/LCG/WLCGBaselineVersions>
 - Evoluera prochainement !
- gLite → EMI1
 - Migration a pris plusieurs mois (debute en mai)
 - Faite pour la plupart des services WLCG recommandes
- A prevoir a court terme
 - EMI1 → EMI2 et gLite → EMI2
- Une campagne de validation des services EMI2 est en preparation par WLCG operations coordination
 - Sites invites a participer
 - Nouvelles sur migration seront donnees

Service	Version	Availability	Notes
ARC CE	1.1.0	EMI / UMD / EPEL / Debian Wheezy / NorduGrid	Release notes
ARC Infosys	1.1.0	EMI / UMD / EPEL / Debian Wheezy / NorduGrid	Release notes
Argus	1.4.0	EMI / UMD / SL5	Release notes
BDII_site	1.0.1	EMI / UMD / SL5	Release notes
BDII_top	1.0.1	EMI / UMD / SL5	Release notes
CASTOR SRM	2.11-0	CASTOR Savannah release area	Release notes
CASTOR	2.1.12-4	CASTOR Savannah release area	Release notes
	2.1.11-9		Release notes
CREAM	1.13.3	EMI / UMD / SL5	Release notes
CernVM-FS	2.0.13	CernVM-FS release	Release Notes
dCache	1.9.12	dCache.ORG	Recommended releases are defined here
FTS_oracle	2.2.8	EMI / SL5	Release notes
Frontier	3.24 ?		Installation notes
Frontier/Squid Launchpad	2.7.STABLE9-3.7 ?		Deployment notes
GLEXEC_wn	3.2.5-1	gLite 3.2 / SL5	Release notes
LB	3.1.0	EMI / SL5	Release notes
LFC_mysql	1.8.0-1	gLite 3.2 / SL5	Release notes , adds banning facility
LFC_oracle	1.8.2-2	gLite 3.2 / SL5	Release notes
SE_dpm_disk	1.8.2-3	gLite 3.2 / SL5 EMI / UMD / SL5	Release notes Release notes
SE_dpm_mysql	1.8.2-3	gLite 3.2 / SL5 EMI / UMD / SL5	Release notes Release notes
StoRM	1.8.2	EMI / UMD / SL5	Release notes
UI	3.2.11-1	gLite 3.2 / SL5	Release notes
VOBOX	3.2.15-1	gLite 3.2 / SL5	Release notes
WMS	3.3.5	EMI / UMD / SL5	Release notes Note: check its known Issues
WN	3.2.12-1	gLite 3.2 / SL5	Release notes

Tape



- LHCb et ALICE pleins, CMS a 90 %
- L'extension du run 2012 n'arrange pas les choses
- Situation blocante
- La priorité est donnée aux bandes pour le budget LCG des années à venir
- Migration vers bandes T10KC (5 TB/bande) en cours
 - Renouvellement des bandes plus anciennes
- L'équipe HPSS a pu recycler une partie des bandes !
- → 1 PB de tapes supplémentaires disponibles très prochainement pour VO LHC
 - Une partie l'est d'ores-et-déjà
 - Répartition à voir, probablement au moins 100 TB par VO dispo. très rapidement

Major incidents



- Corrupted files on dCache-LHCb (2012/06)
 - Mismatch observed between some LFC and dCache file checksums (0.03 % files)
 - A working group has been setup to understand the problem
 - The problem was not understood
 - Other LHCb sites were informed and some were much worse
 - CCIN2P3 has asked LHCb to activate checksum verification during transfers
- CVMFS Squid (2012/07/03)
 - Server failure, no failover → production down for ATLAS and LHCb
- Plus others concerns
 - Network performance, lost files dCache etc...



Situation ALICE



- **Disques pleins, bandes pleines**
 - Recu une grosse partie des donnees 2011 sur bande
 - Moins de place pour 2012
 - Une partie des jobs recuperent les donnees a distance
- **Grosse consommation memoire**
 - Une grosse partie des jobs frisent les 4 GB de VMEM
- **Efficacite**
 - Les CPU ne sont pas pleinement utilises
 - Surbooking difficile (memoire)



Situation ATLAS



- Problèmes transferts réseau avec certains sites
 - Une instance perfSonar a été déployée
 - Performances altérées par mauvaise combinaison java+sunOS → résolu
- Espace disponible sur bande (~400TB)
- Configuration des queues pour jobs multi-core en cours
 - Activité croissante à prévoir à court terme

Situation CMS



- Bandes quasi pleines
 - → Menage a prevoir, utilisation « locale » reste marginale (2%)
- Installation cache CVMFS (arret d'hier)
 - Plus de dependance AFS pour VO LHC au CC
- Cf talk de Sébastien



Situation LHCb



- Pilotes
- Metrique d'utilisation memoire
 - VMEM elevee mais reelle consommation RAM faible
 - LHCb demande davantage de SWAP
 - En cours : mise en place de machines virtuelles
- Cache CVMFS
 - Redimensionnement (arrêt d'hier) :
 - 10 GB software
 - 1.5 GB cond-db
 - Demande d'un cache plus gros (16 GB)
 - → nous répondrons à cette demande en fonction de la taille des machines
- Sous-utilisation de la ferme par LHCb
 - Beaucoup de problèmes en cours qui font que LHCb a du mal à utiliser les ressources du CC



LAF



En production depuis 1 an

Interactive
machine
(ccage)



16 Dell PEM610
* 16 cores
2.53GHz

**LAF
nodes**



master



workers



10 Gb/s



LAF storage (xrootd)

3 Dell EqualLogic PS 6010xv
16 blades x 600GB SAS 15krpm
RAID5



20 TB storage

Local disks
(system + proof daemon)

dCache

SPS

2*1 Gb/s



LAF storage (xrootd)

Sun Fire X4500

17 TB

storage

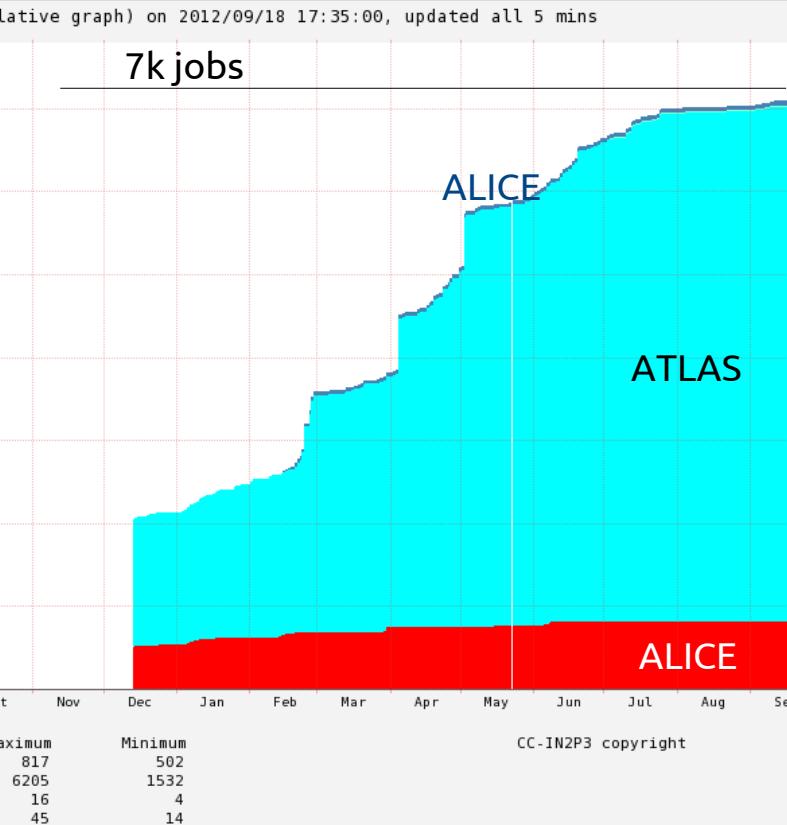
sun
microsystems

Stockage xrootd pour staging des
fichiers sur GRID

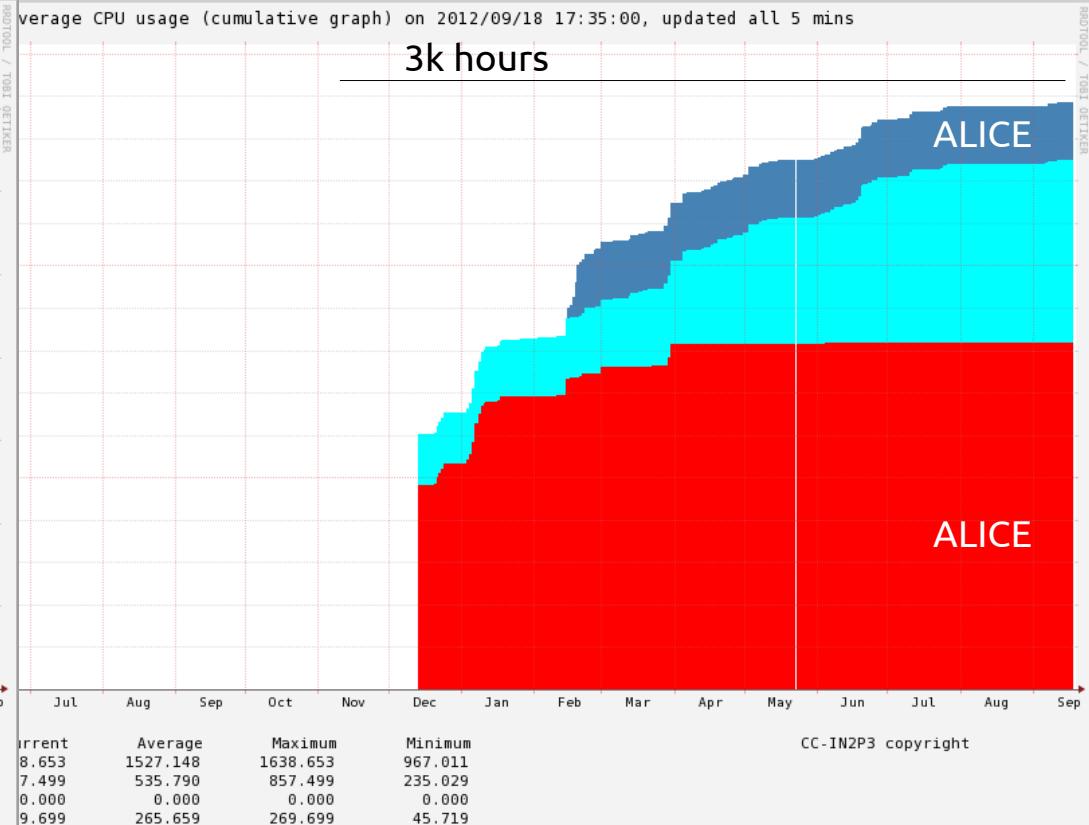
Statistique utilisation LAF



jobs



CPU time



Une certaine activité...
Mais très faible taux d'occupation des workers



- 2 separate instances are running

- Standard PROOF

- 8 slaves (128) workpoints)
 - 1 storage server 17 TB

- ALICE Analysis Facility

- 5 slaves (80 workpoints)
 - 1 storage server 19 TB

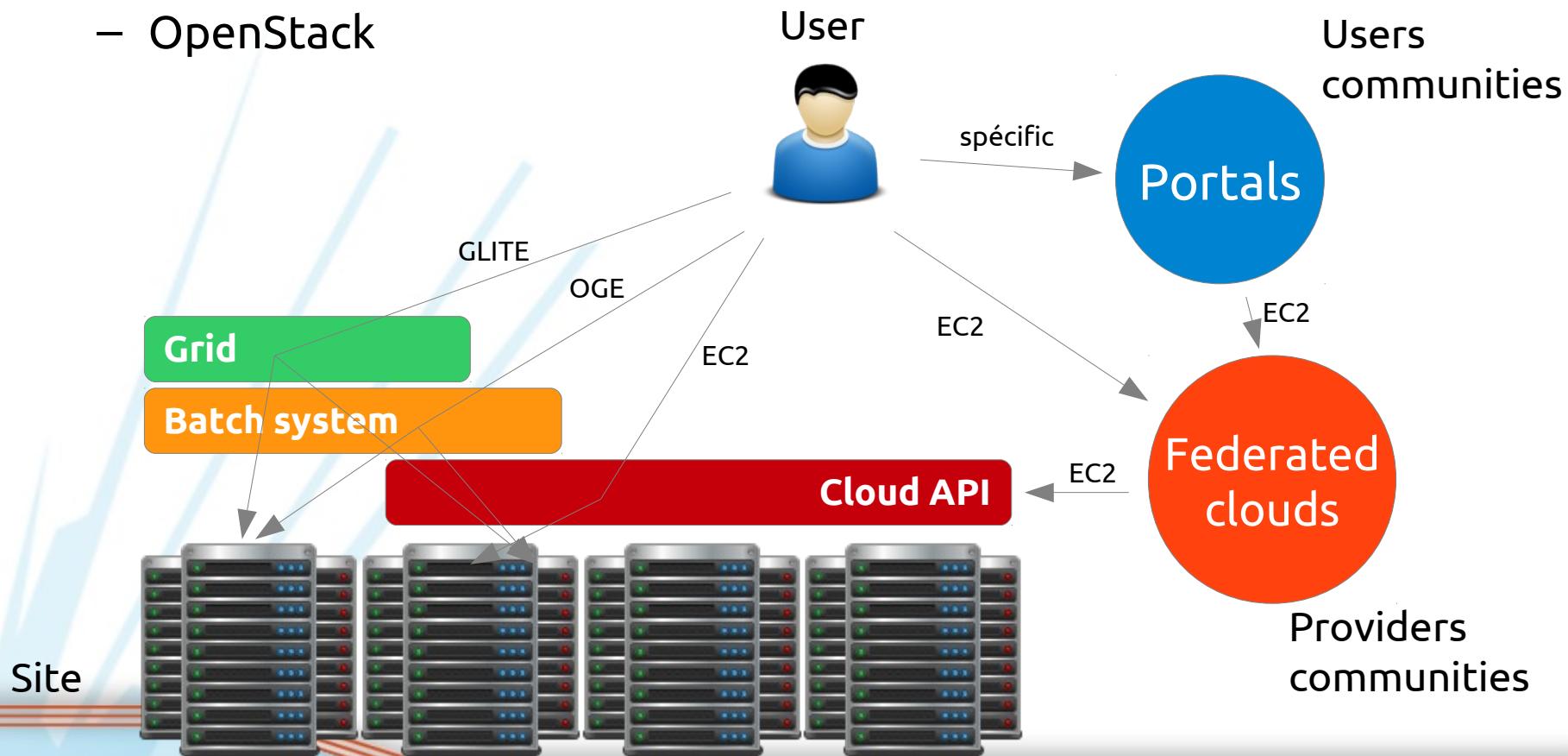
- Activity monitored by MySQL db
- None of the instances is being actively used
 - Connections from time to time, mostly ATLAS
- Very little demand
 - Performance 20 times lower than expected
 - Account at CCIN2P3 required for its use → discourage ALICE people



Cloud



- Mise en place d'une infrastructure de Cloud au CCIN2P3
 - Pool de machines dédiées
 - OpenStack



Testbed infrastructure



- 16 Dell Poweredge C6100 hosts
- On each node
 - 2 Xeon 24cores X5675 @ 3.07GHz
 - 96GB RAM
 - 2TB raid 10 local storage (4 SAS 7.2 krpm)
- Catalog of images (1.4 TB)
- Public IPv4 subnet → full VM network isolation
- Powered by Openstack
- Available interfaces :
 - EC2 and Nova API at the moment
 - Ongoing work on OCCI

Activite LHC sur cloud



- Rien pour l'instant, mais...
- Demarrage recente d'un groupe de travail pour LHCb
 - Mesure de l'utilisation memoire difficile sur grosses machines
 - Solution ? → fournir des VM pour etudier le comportement des jobs
 - Offline LHCb interesse par cette solution
- A terme
 - Utiliser future (?) fonctionnalite OpenStack de Dirac pour soumission cloud-like
 - Experience interessante pour VO et site

Perspectives



- Ameliorer notre monitoring
- Etre impliques dans le massivement parallele
- Developper notre cloud
- Penser a la federation
- Etre impliques dans l'operation WLCG
- Migration EMI2

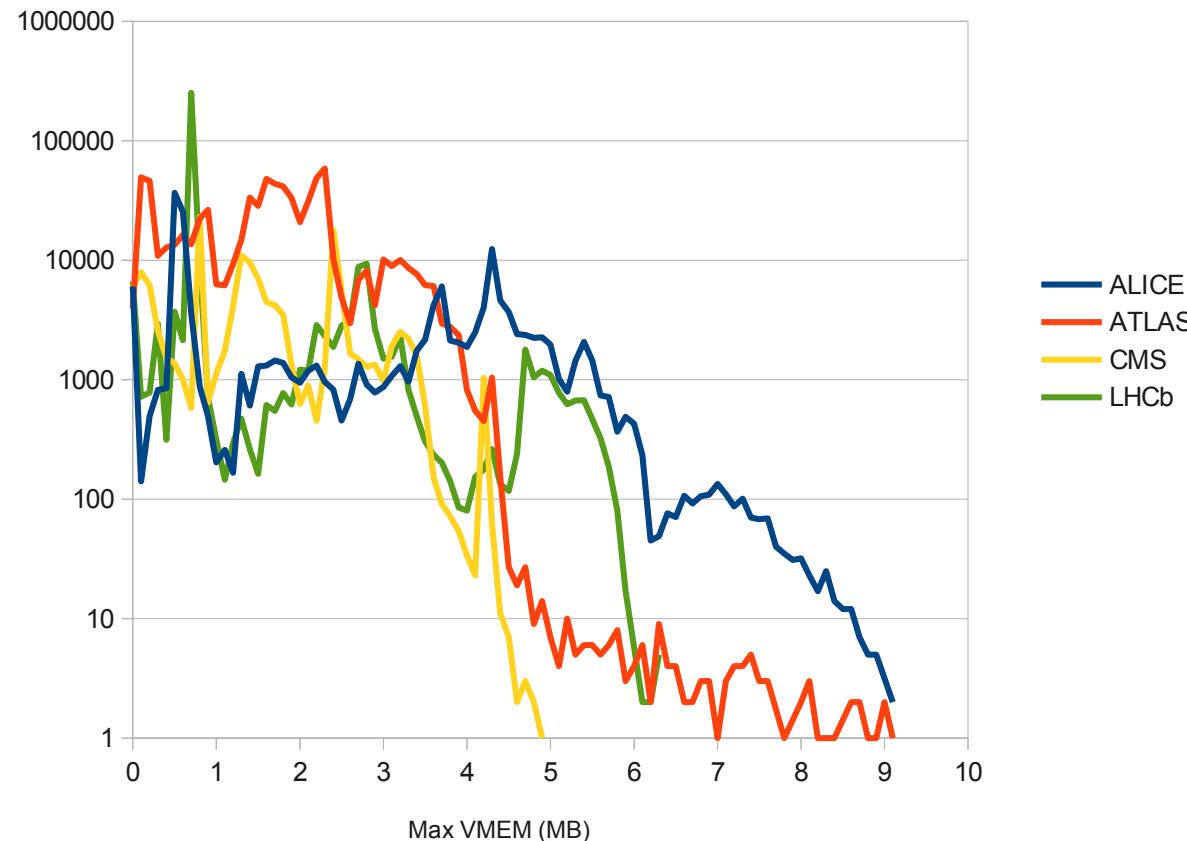
- Questions en tete (HEPiX)
 - Signal envoyé au job pour dépassement limite (mémoire, wall...)
 - Métrique utilisation mémoire

Complements

Memory usage

Job max VMEM distribution

sub-title

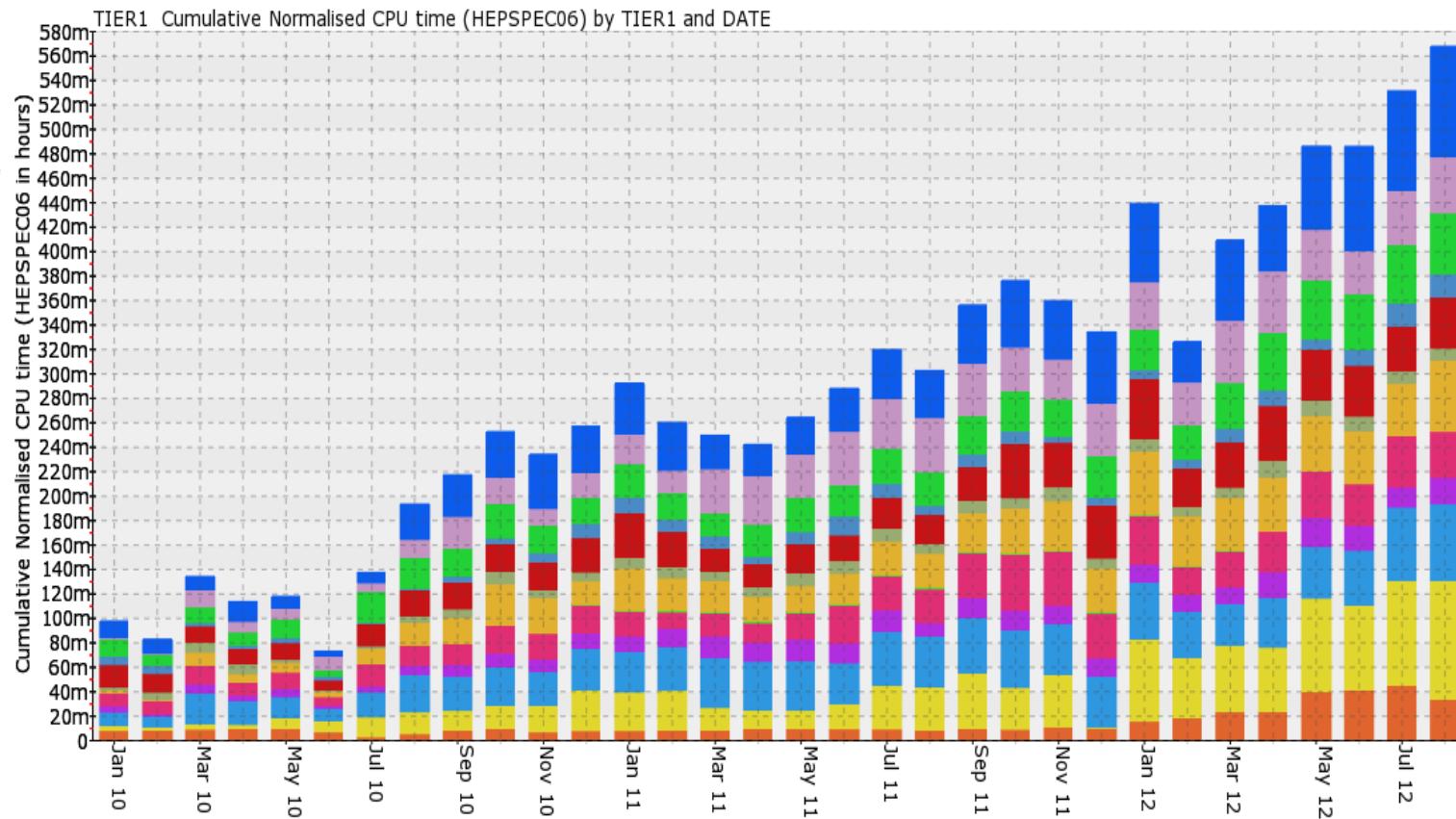


CPU consumption vs time vs T1

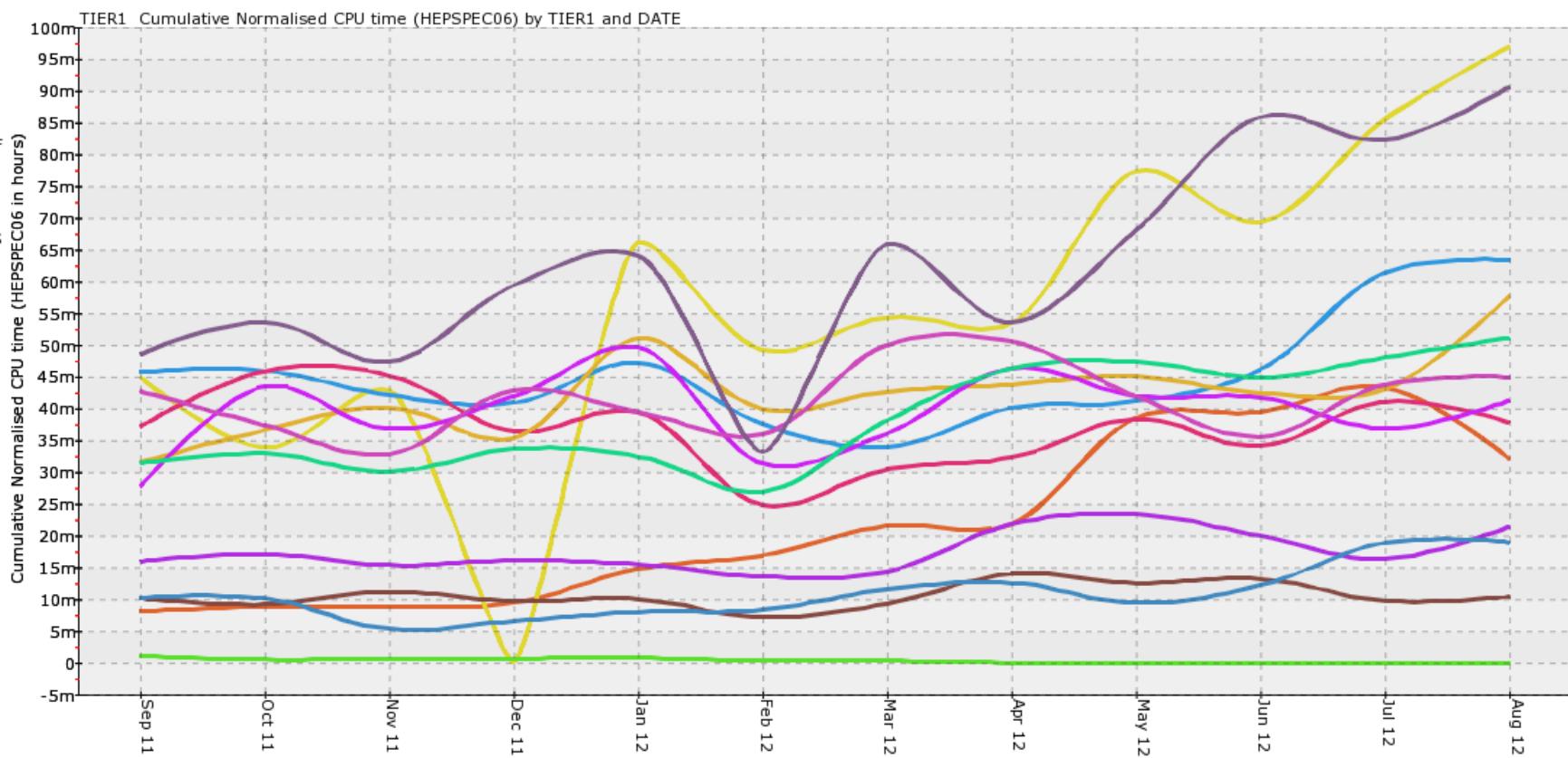


Developed by CESGA EGL View: / normcpu-HEPSPEC06 / 2010-1-2012-8 / TIER1-DATE / lhc (x) / ACCBAR-LIN / i

2012-09-17 0250



CCIN2P3



Ressources T2



T2 shares at CC-IN2P3 (2012)

