PUGNERE Denis
Institut de Physique Nucléaire de Lyon

Rapport de site Tier3 IN2P3-IPNL

Contributions de : Anne-Laure, Boris, Elvire, Stéphane, Tibor, Yoan

Rencontres LCG-France
14-16 décembre 2015 – CC-IN2P3 - Villeurbanne

Plan

1) **Présentation du laboratoire et du T3 IN2P3-IPNL**
2) **statistiques du T3 IN2P3-IPNL et activités des VOs**
3) **Problèmes et activités**
4) **Futur**

1

Présentation de l'IPNL

1.1

Présentation de l'IPNL

- **IPNL : CNRS / Université Claude-Bernard - Lyon 1 (UCBL) / Université Lyon**
  - **216 personnes (100 CNRS, 55 non CNRS)**
  - **Domaines de recherche :**
    - **PNHE - Physique Nucléaire et des Hautes Energies (80%)**
    - **PHY - Physique (10%)**
    - **SC - Chimie (5%)**
    - **SPU - Sciences de la Planète et Univers (5%)**

- **Groupes de recherche : CMS, Alice, ILC, AEgIS, Neutrinos, Matière Sombre, Matière Nucléaire, Théorie, Cosmologie observationnelle, EUCLID, R&D ebCMOS, CAS-PHABIO, IPM, ACE**

- **7 services techniques**

- **Service informatique : 17 personnes**

1.2 Activité du groupe CMS sur le T3_FR_IPNL

- **Activités principales :**
  - **Analyses de données**
  - **Production Monte Carlo privées**
  - **Crab & ganga**
  - **Activité computing réduit en 2015, la finalisation des analyses du RUN 1 a nécessité moins de calcul qu'en 2014,**

- **Disk: 900 TB for CMS**
  - **Host the Primary datasets for interesting the local analyses (where we can not afford to loose jobs)**
  - **Copie de certain MC de T1**
  - **Local rootuples**

- **Strong and essential support for the local analyses :**
  - **Rate of job success close to 100% when running crab in local.**
  - **Large priority: almost no time spent in queue**
  - **Good reactivity from the local IT team**

1.3 Activité Alice sur le T3 IPNL (1/2)

Types d'analyse de physique :

- **Simulations, reconstructions et analyse d'événements pour l'étude des performances physiques pour le projet d'upgrade MFT (utilisation distribuée dans l'année, pics de production possibles selon le calendrier d'activités établi par le CERN). Utilisation potentiellement lourde.**

- **Analyse des données (compression du format, extraction des informations physiques). Pics d'activités en cas de productions centrales du CERN, ou en préparation de conférences. Utilisation typiquement légère.**

- **Simulation et reconstruction d'événements en support de l'analyse des données. Utilisation typiquement lourde, mais distribuable sur de longues périodes.**

1.4 Activité Alice sur le T3 IPNL (2/2)

Intérêt pour Alice :

- **L'utilisation du Tier3 a donné et donne actuellement une possibilité unique d'accéder à d'importantes ressources de calcul en particulier dans deux cas :**

  - Simulations d'importants volumes d'événements, liées à l'étude des nouvelles méthodes de trajectographie pour le Muon Forward Tracker, nécessitant de versions de développement du code AliRoot pas encore disponibles sur la Grille.

  - Analyses des données ayant un accès très limité, voire interdit, aux ressources de calcul du CERN à cause de la priorité donnée par le CERN à d'autres analyses.

- **Les activités de traitement de données et simulations d'événements nécessaires pour l'analyse des dimuons de basses masses dans l'expérience ALICE, sujet de la thèse de doctorat de B. Teyssier, dépendent critiquement des ressources de calcul du Tier3 de Lyon.**

1.5 Présentation du T3 IN2P3-IPNL

- VO supportées : alice, cms, vo.agata.org, calice, dzero, vo.france-grilles.fr, vo.ipnl.in2p3.fr, vo.rhone-alpes.idgrilles.fr
- Ressources humaines : Guillaume Baulieu, Yoan Giraud, Tibor Kurca, Martin Mommey, Denis Pugnère (total 1,27 ETP)
- ~ 1700 slots (16k HEP-SPEC06) : Faireshare : 90% LCG (70% CMS, 30% Alice), 10 % autres vo
- Stockage :
  - 42TB XROOTD alice : rempli à 89%
  - ~900TB DPM cms : rempli à 68%
  - ~50 TB autres VOs
  - ~48TB espace POSIX partagé GPFS (en cours de renouvellement)
- Services : 1 CREAMCE, Torque, squid cvmfs, lfc, bdii, argus, head dpm, head xrootd, 2 vobox (alice + cms)
- Soumission jobs grille uniquement : Crab (cms), alien (alice), soumission directe, ganga, dirac
- Quelques UI pour le groupe CMS + quelques UI à usage général
- Configuration :
  - Quattor : T3
  - Puppet + foreman : Serveurs du laboratoire
- Monitoring environnement (systèmes, clims) : Nagios + centreon
- Monitoring T3 : Application Java (Jobs, fairshare, stockage, WN) + graphiques RRD
- Tour de garde monitoring hebdomadaire

1.6 Infrastructure générale du laboratoire

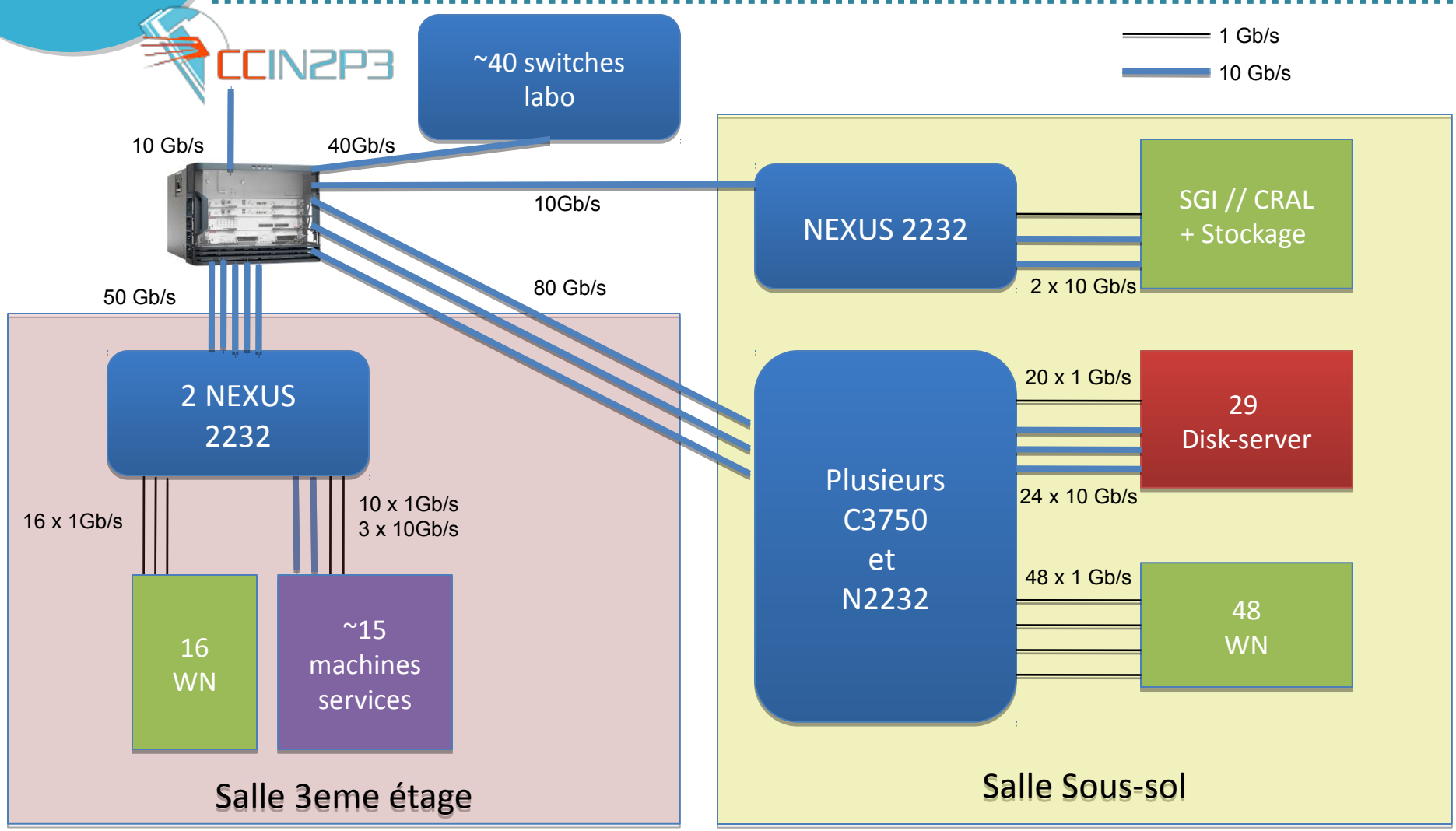
- 2 salles informatiques :
  - 3eme étage : 35 m^2, 8 racks (remplis à 70%), 100% ondulé, 2x30Kw froid
    - consommation actuelle ~40KW
    - Services et stockage du laboratoire
    - Services grille et quelques WN (~192 slots)
  - Sous-sol : 30m^2, 10 racks (remplis à 40%) + clims InRow APC,
    - Stockage sur courant ondulé,
    - Workers (~1500 slots) sur courant EDF
    - Cluster SGI // du CRAL : 32 workers + stockage
    - Quelques services labo
    - Quelques serveurs de groupe
    - consommation actuelle ~45KW
- Groupe froid 200KW frigo
- Onduleur 160KW
- Groupe électrogène 235KW

1.7 Infrastructure matérielle du T3

- Stockage grille :
  - 5 * X4500 (20TB : 4 x 1Gb/s)
    - 2 * xrootd : alice
    - 3 * DPM : vo.agata.org,calice,dzero, vo.france-grilles.fr, vo.ipnl.in2p3.fr, vo.rhone-alpes.idgrilles.fr
  - 9 * DELL r510 20TB (1 x 10Gb/s) : DPM cms
  - 8 * DELL r720xd 30TB (1 x 10Gb/s) : DPM cms
  - 7 * DELL r730xd 72TB (1 x 10Gb/s) : DPM cms

- Worker nodes : tous en 1Gb/s
  - 1 blade center : 16 bi E5540 @ 2.53GHz, 16c HT, 24Go RAM (16*16 coeurs, 12 slots / worker)
  - 5 C6100 : 20 bi E5645 @ 2.40GHz, 24c HT, 48Go RAM (20*24 coeurs, 24 slots / worker)
  - 1 C6220 : 4 bi E5-2670 @ 2,60GHz, 32c HT, 96Go RAM (4*32coeurs, 32 slots / worker)
  - 6 C6220 : 4 bi E5-2670v2 @ 2,60GHz, 40c HT, 128Go RAM (24*40 coeurs, 40 slots / worker)

1.8 Infrastructure réseau du T3



1.9

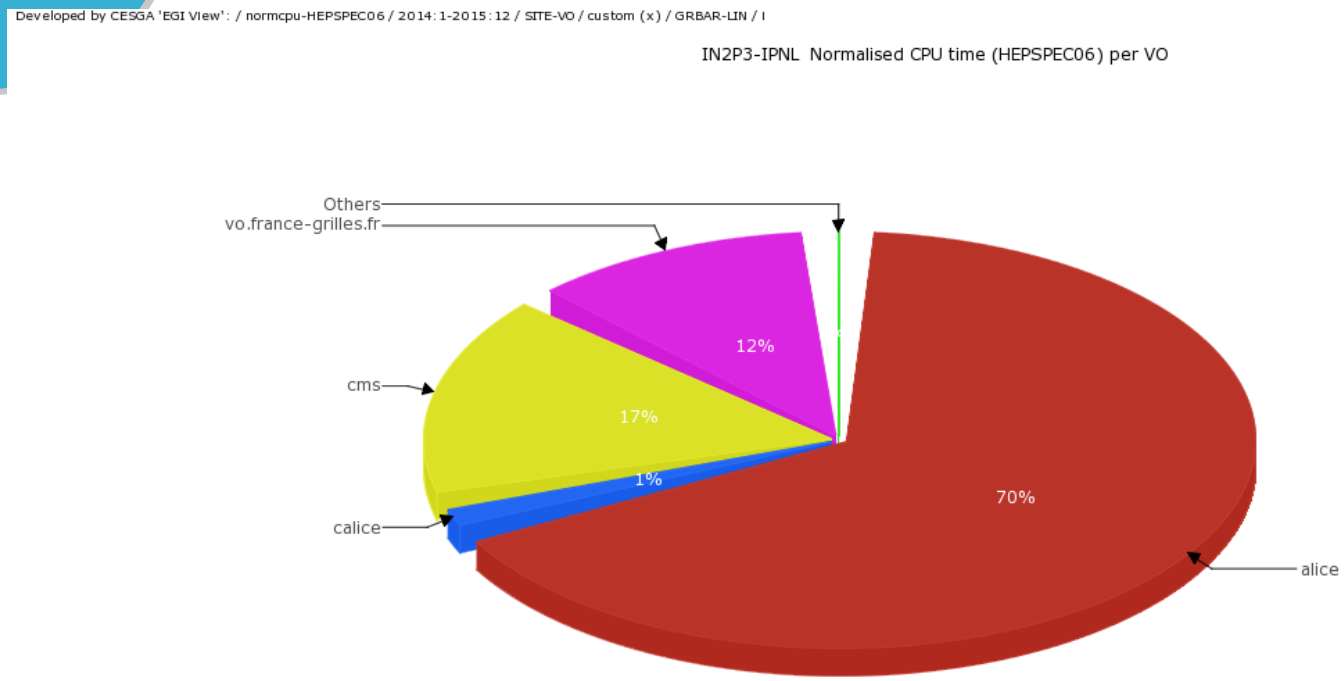
Évolutions depuis dernier rapport de site (Q2 2013 → Q4 2015)

- **Refonte complète du réseau backbone (Cisco Nexus 7004 + N*Nexus 2232)**
- **Extinction incendie salle sous-sol**
- **Sortie de production des DELL PE 1950, des Sun x4500**
- **Achats CPU (6 DELL C6220)**
- **Achats stockage : 7 DELL R730xd**
- **Migration de tout le stockage DPM CMS en SL6**
- **Installation de plusieurs UI à usage général,**
- **T3_FR_IPNL dans la fédération AAA de CMS**
- **Stockage GPFS : remplacement d'un cluster HA : 2 x Dell R610 + 1 MD3200 + 1 MD1200 par 2 x Dell R630 + 1 MD3460 (installation Q1 2016)**

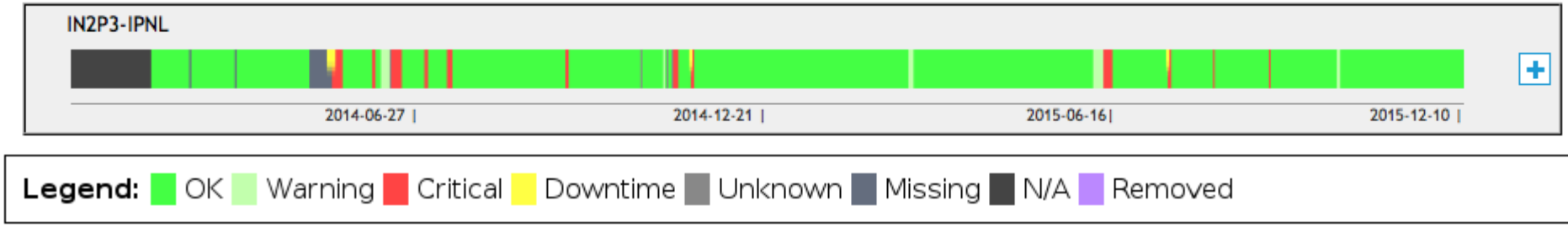
2

Statistiques du T3 IN2P3-IPNL et activité des VOs

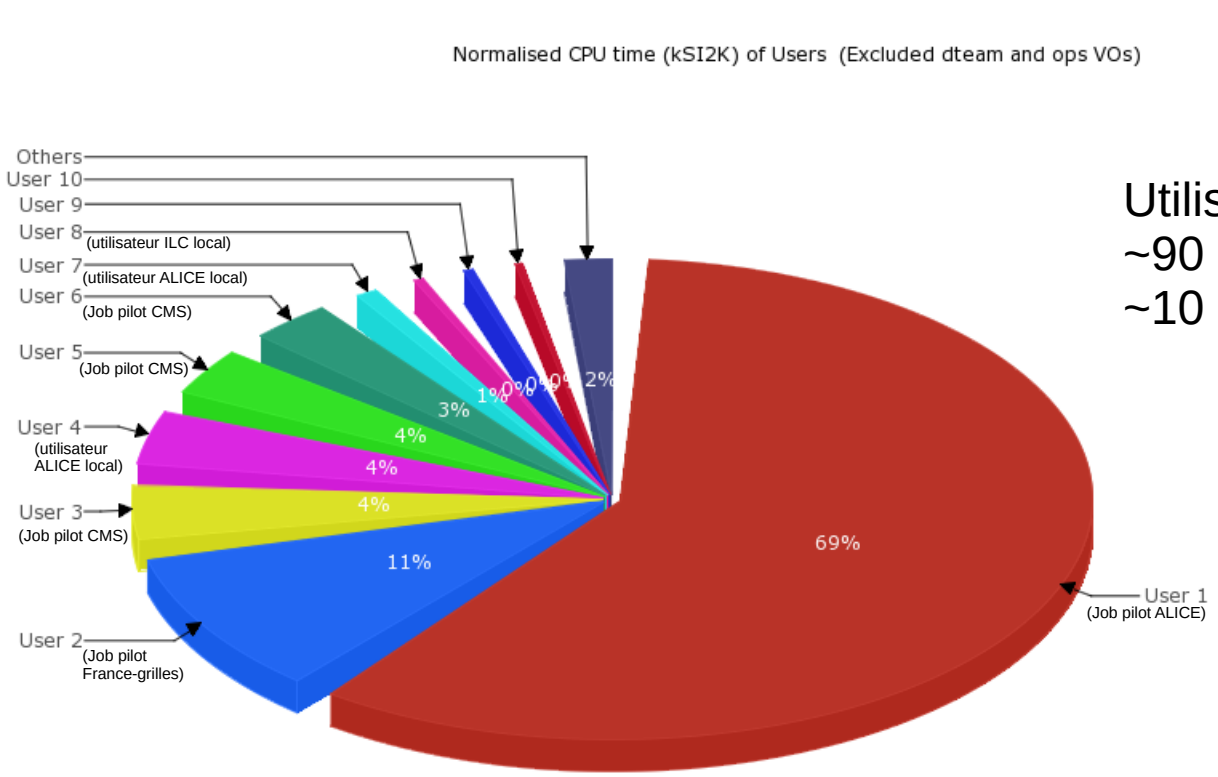
2.1 Statistiques 01/2014 -> 12/2015



EGI Accounting portal : http://accounting.egi.eu/egi.php?ExecutingSite=IN2P3-IPNL



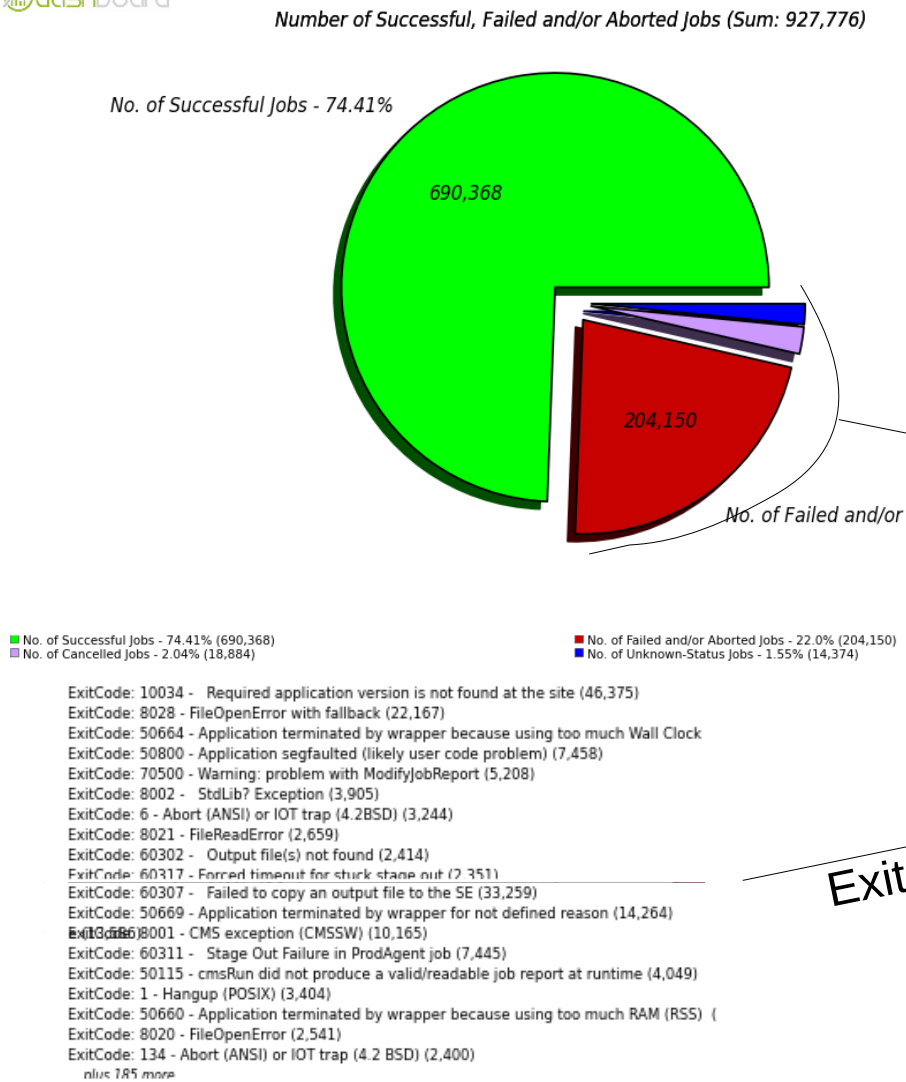
2.2 Statistiques 01/2014 -> 12/2015



Utilisation du T3 :
~90 % externe
~10 % interne

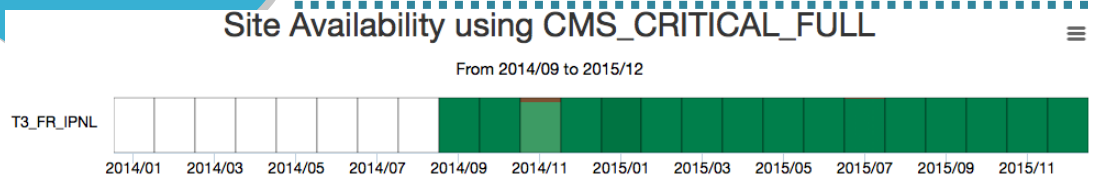
EGI Accounting portal : http://accounting.egi.eu/egi.php?ExecutingSite=IN2P3-IPNL

Statistiques CMS : 01/2014 -> 12/2015



Dashboard CMS :
http://dashboard.cern.ch/cms/

2.4 Statistiques CMS : 01/2014 -> 12/2015

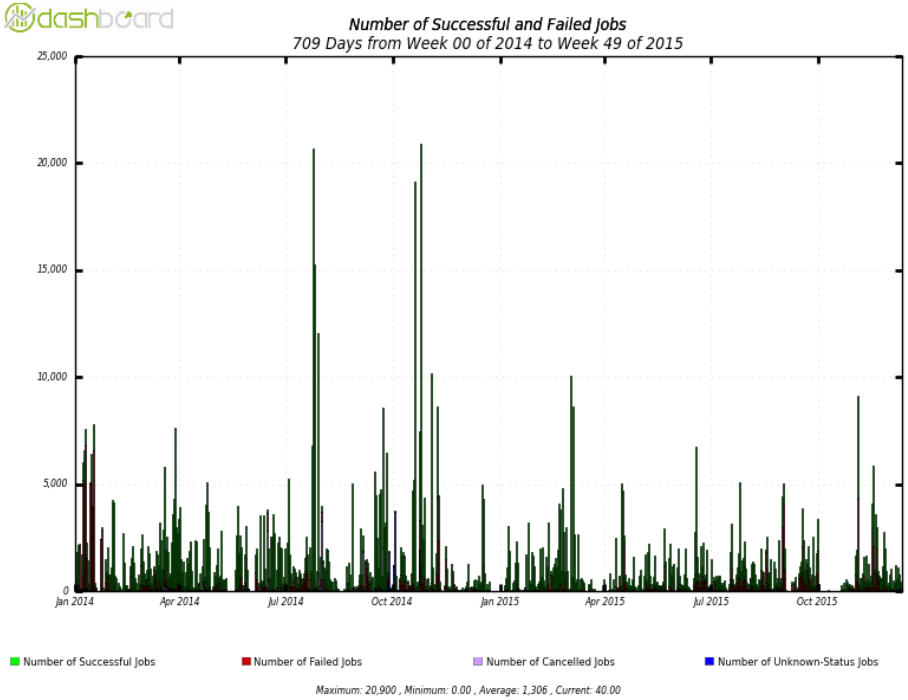


CMS_CRITICAL_FULL Algorithm =

(OSG-CE + CREAM-CE + ARC-CE) *
(all SRMv2 + all OSG-SRMv2)

Dashboard CMS

http://wlcg-sam-cms.cern.ch/templates/ember/#/historicalsmry



Dashboard CMS : http://dashb-cms-job-dev.cern.ch

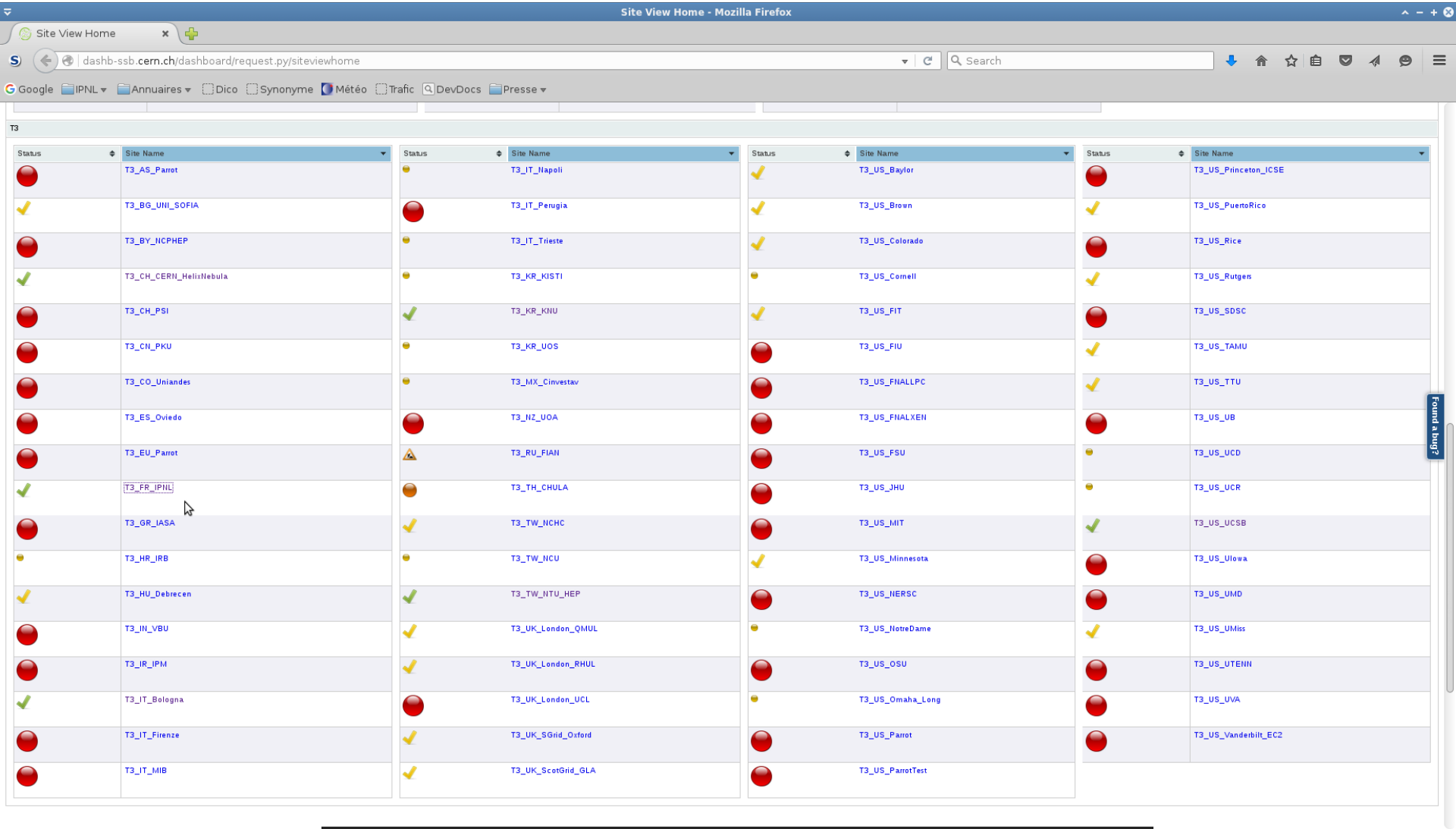
From Node ↓ To Node →	T3_FR_IPNL
T1_DE_KIT_Buffer	
T1_ES_PIC_Buffer	
T1_FR_CCIN2P3_Buffer	
T1_IT_CNAF_Buffer	
T1_IT_CNAF_Disk	
T1_UK_RAL_Buffer	
T1_UK_RAL_Disk	
T1_US_FNAL_Buffer	
T2_CH_CERN	
T2_DE_DESY	
T2_FR_CCIN2P3	
T2_FR_IPHC	
T2_IT_Legnaro	
T2_UK_SGrid_RALPP	
T3_US_FNALLPC	

PhEDEx – CMS Data Transfers

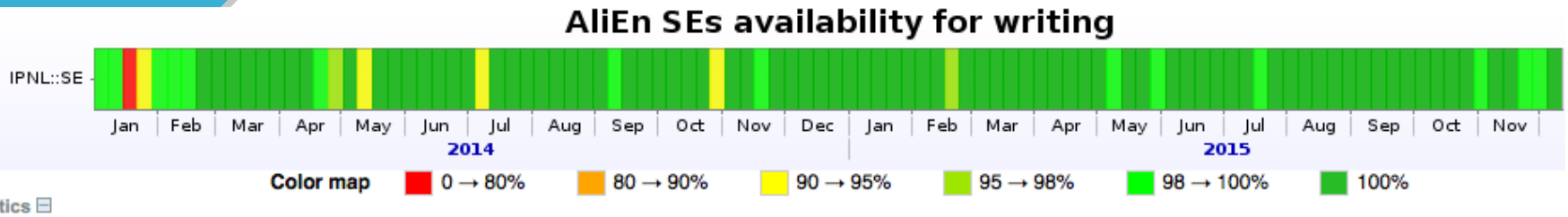
https://cmsweb-testbed.cern.ch/phedex

2.5 Snapshot du statut des T3 CMS au 10/12/2015

Dashboard CMS :
http://dashb-ssb.cern.ch/dashboard/request.py/siteviewhome



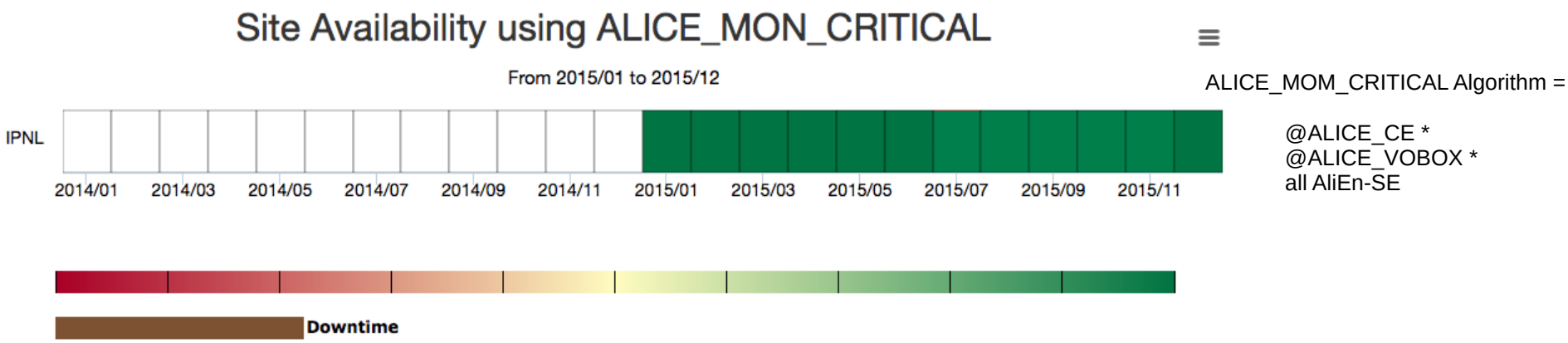
2.6 Statistiques Alice 12/2014 -> 12/2015



Statistics

Link name	Data		Individual results of writing tests			Overall
	Starts	Ends	Successful	Failed	Success ratio	Availability
IPNL::SE	01 Jan 2014 16:14	11 Dec 2015 14:13	8429	67	99.21%	99.28%

http://alimonitor.cern.ch/status/index.jsp

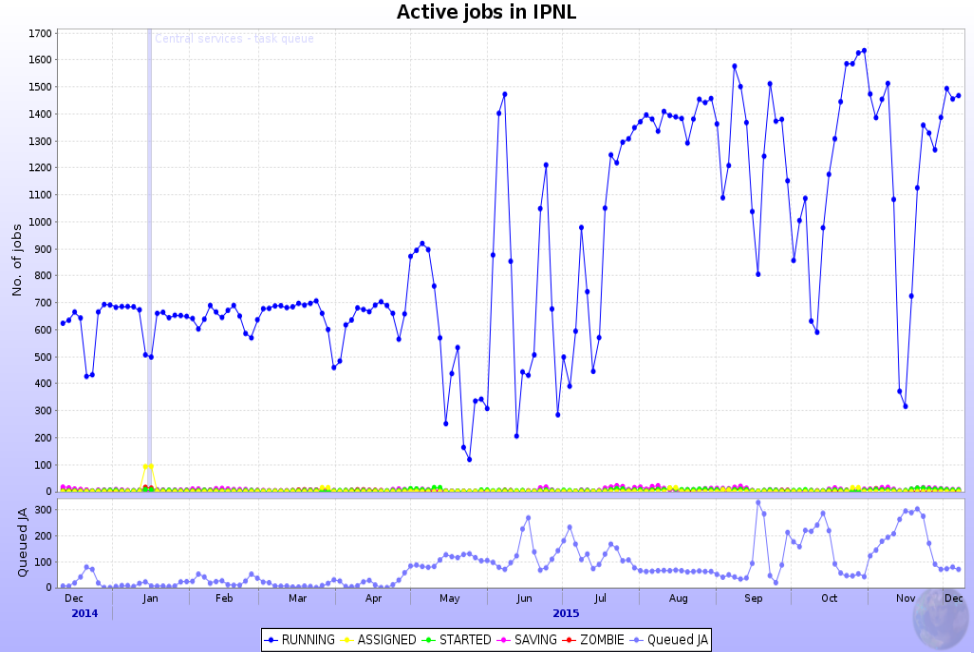


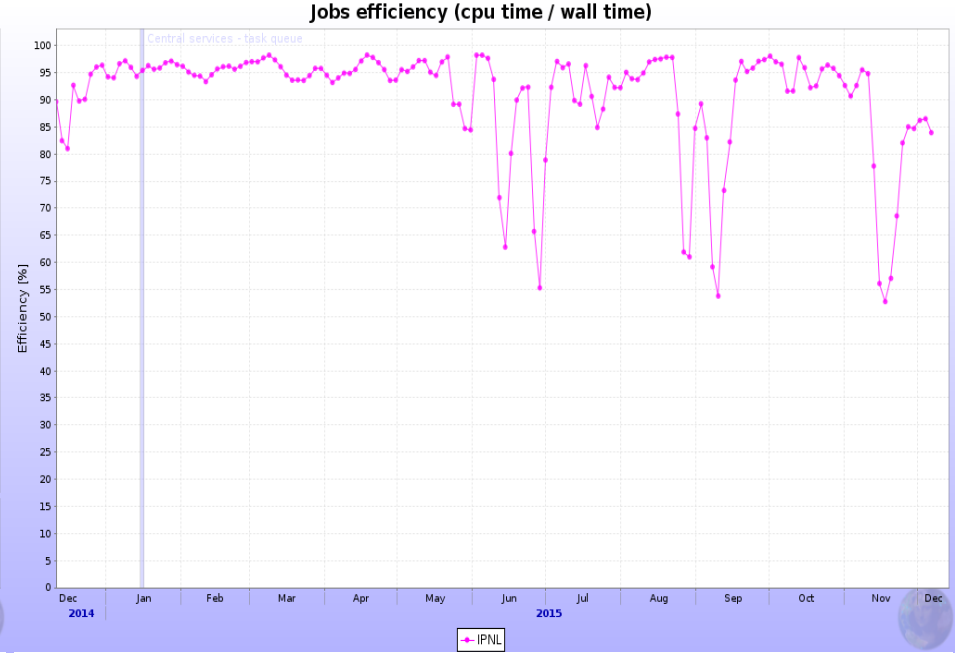
ALICE_MOM_CRITICAL Algorithm =

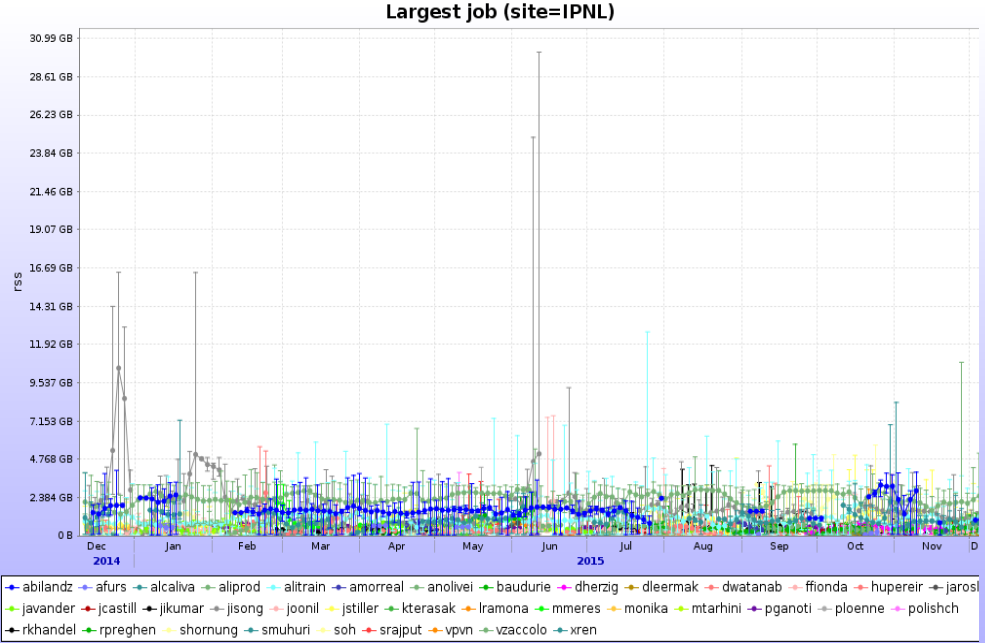
@ALICE_CE *
@ALICE_VOBOX *
all AliEn-SE

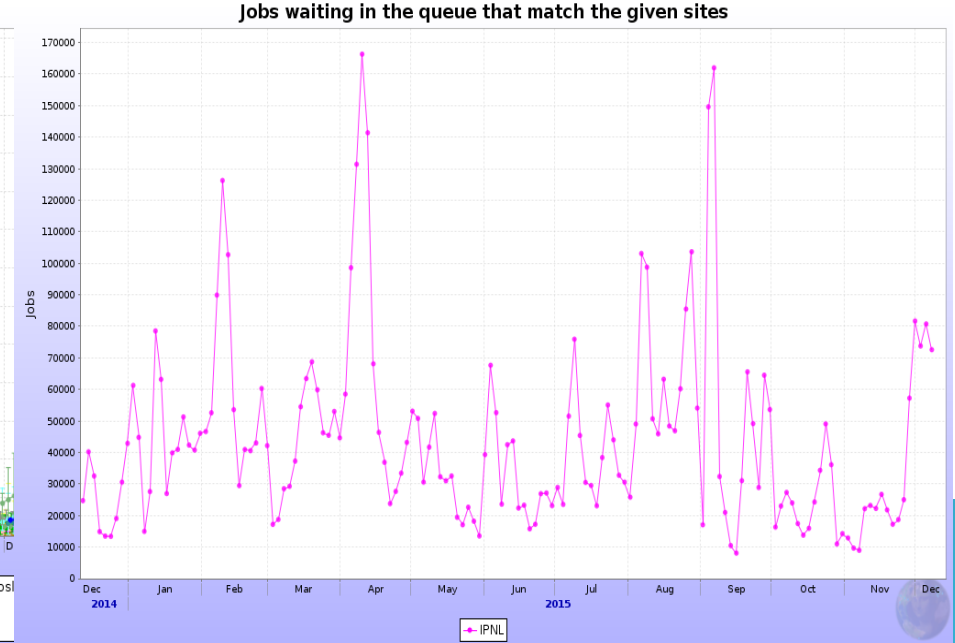
http://wlcg-sam-alice.cern.ch/templates/ember/#/historicalsmry

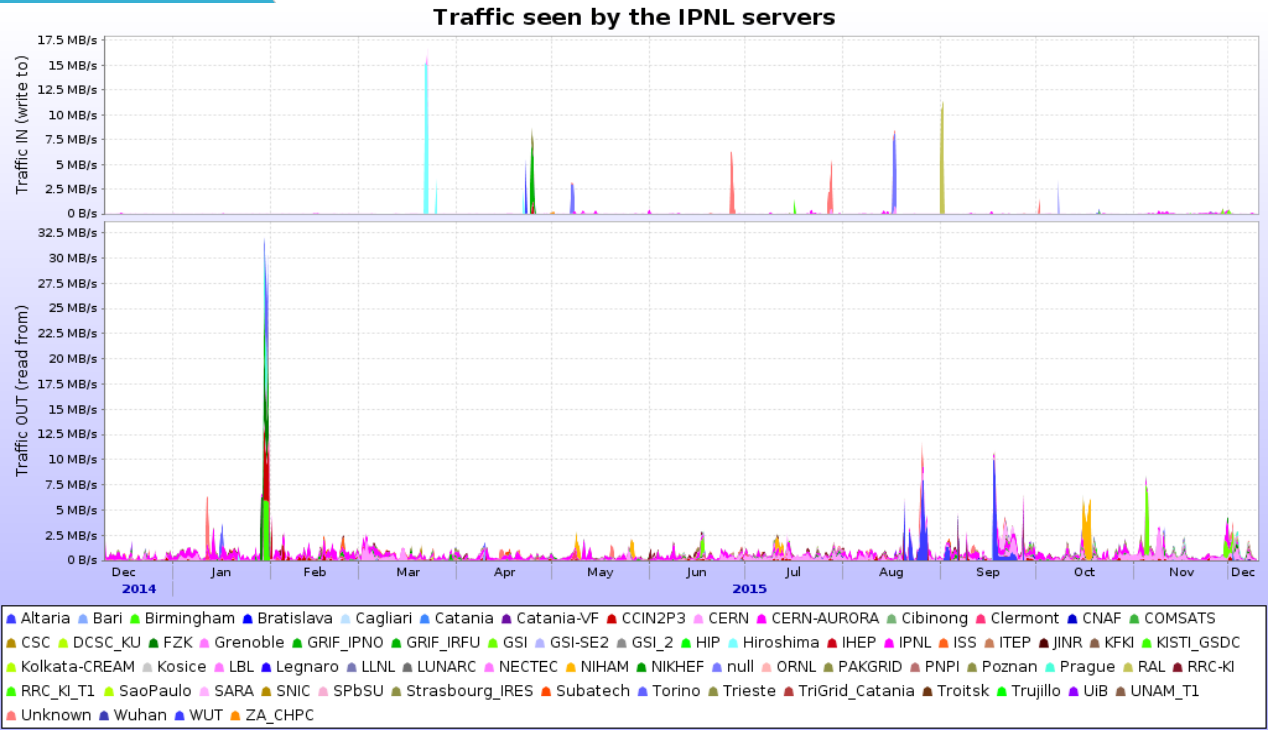
2.7 Statistiques Alice 01/2015 -> 12/2015

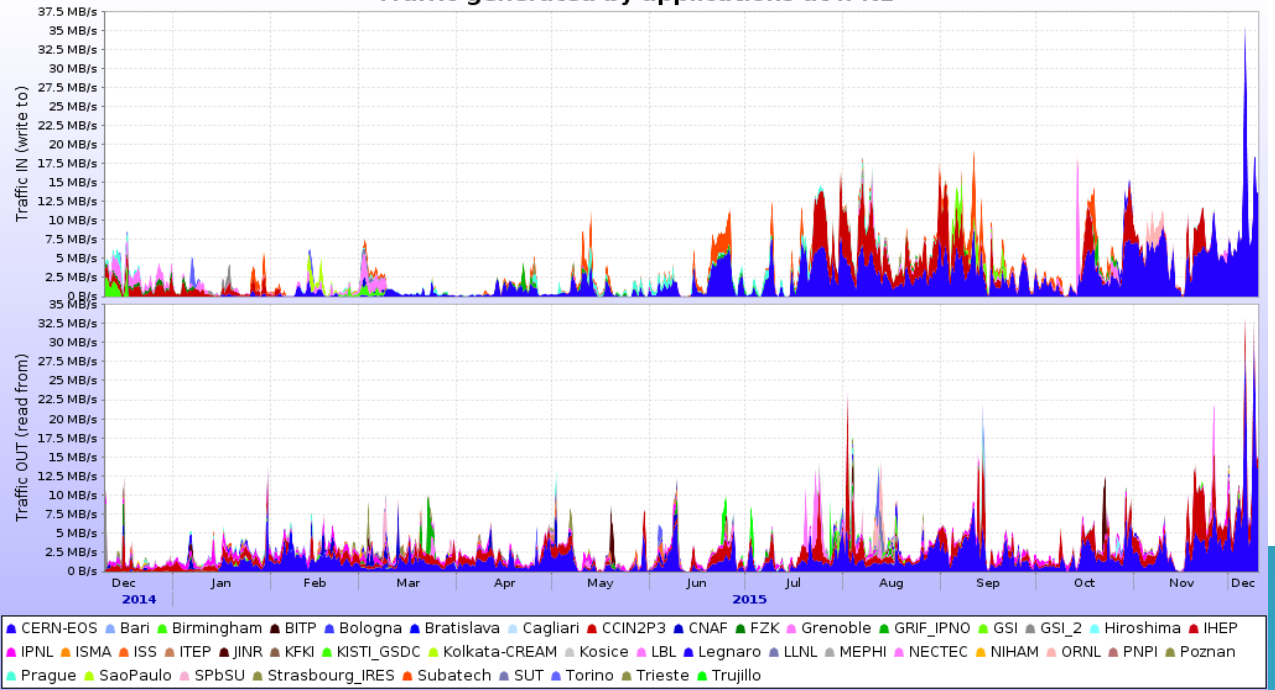












Clients / SE
(trafic que voit notre SE)

SE / client
(accès WAN de nos jobs)

2.8 Particularité d'être T3 vs T2

- **ALICE**
  - **+ Pas de REBUS à remplir**
  - **= Point de vue admin de site : Aucune différence**
  - **Point de vue utilisateur : ?**
- **CMS**
  - **Point de vue admin de site :**
    - **+ Pas de REBUS à remplir**
    - **= Même contrainte morale de disponibilité / réactivité**
  - **Point de vue utilisateur :**
    - **- Tâches de services des membres CMS non reconnues**
    - **+ Plus de ressources disponibles car non réservées par la VO**
  - **À venir : « Production » and « Transitional » federations**

3

Problèmes et activités

3.1 Problèmes et activités

Quelques exemples de problèmes de site rencontrés :

- publication dans BDII (pas de jobs alice depuis la VOBOX)
- publication dans BDII (jobs CMS depuis la job factory)
- bursts de jobs Alice > 2.5Go de RAM RSS
- saturation de la base MySQL du CREAMCE (→ purge)

Problèmes remontés par les utilisateurs :

- CMS :
- problèmes de compatibilité entre CRAB3 et CMSSW
- versions gfal* nécessaires > aux versions baseline

- ALICE : difficulté de faire passer les jobs dans les trains alice dans un temps raisonnable

3.2 Priorité des jobs

Problèmes historiques : priorité des jobs de nos utilisateurs passant par les job-pilots des VOs (alice : alien, cms : crab) : pas possible de leur donner une meilleure priorité par rapport aux autres.

Solution proposée dans CMS :

Les « Site-customized glideins » : Les sites avec un CE peuvent contrôler/prioritiser les accès pour les utilisateurs locaux arrivant depuis les jobs pilots CMS

CMS peut envoyer des job pilots avec un rôle VOMS spécifique pour une liste d'utilisateurs pré-déterminée

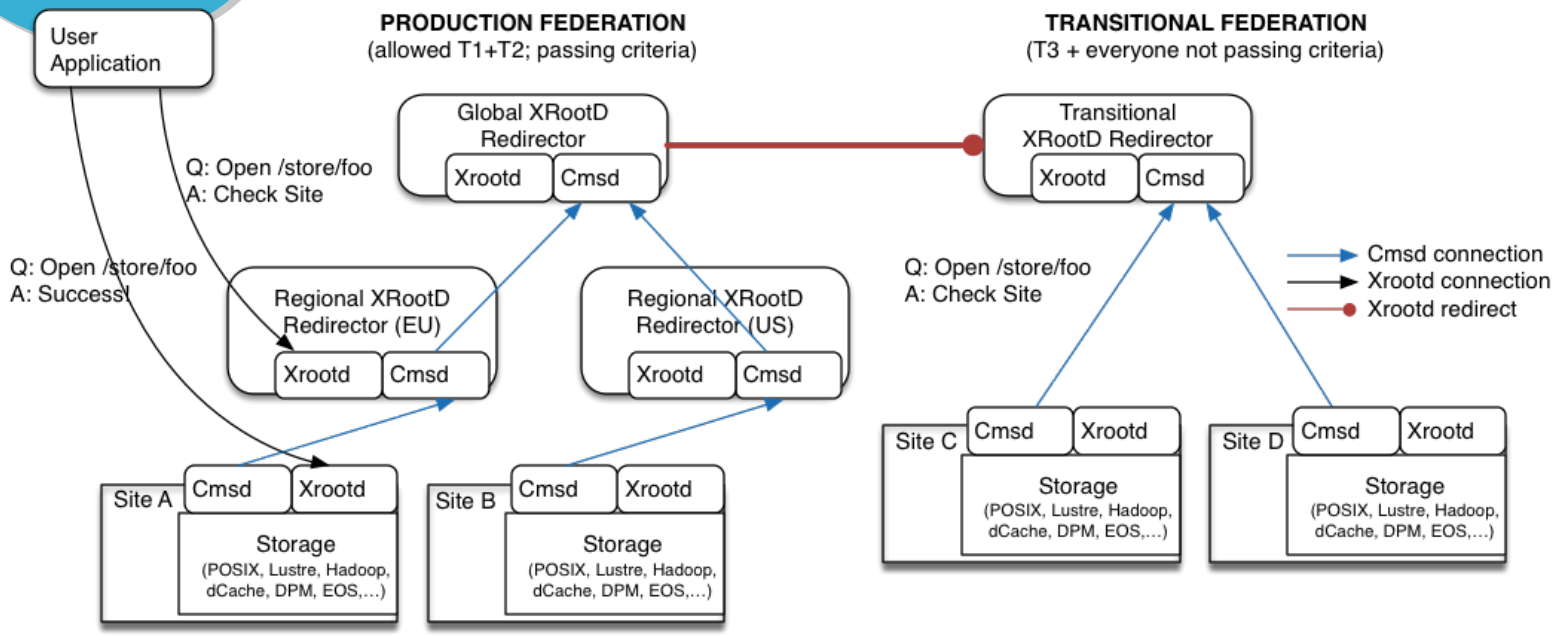
Comment ?

- **ajouter un fichier GlideinConfig/local-users.txt (avec CERN logins) dans SITECONF**
- **CRAB3 lit ce fichier et va créer un jobad approprié**
- **fronted va soummetre pilot joba avec le rôle "local" (/cms/local/Role=pilot)**
- **le site mappe cet attribut sur compte, exemple cmslocal (juste utilisateurs local)**
- **ajout du compte cmslocal dans la configuration glexec et argus**
- **dans le submit_filter de torque, reconnaissance de ce rôle et affectation dans un accounting group « ipnl »**
  --> cmslocal fait partie de "accounting" group ipnl

En production dans le sites : T3_FR_IPNL, T2_ES_CIEMAT, T2_US_Nebraska, T3_US_TAMU, T3_IT_Bologna
En test dans : T2_DE_DESY

4 Le futur

4.1 CMS « Production » and « Transitional » federations

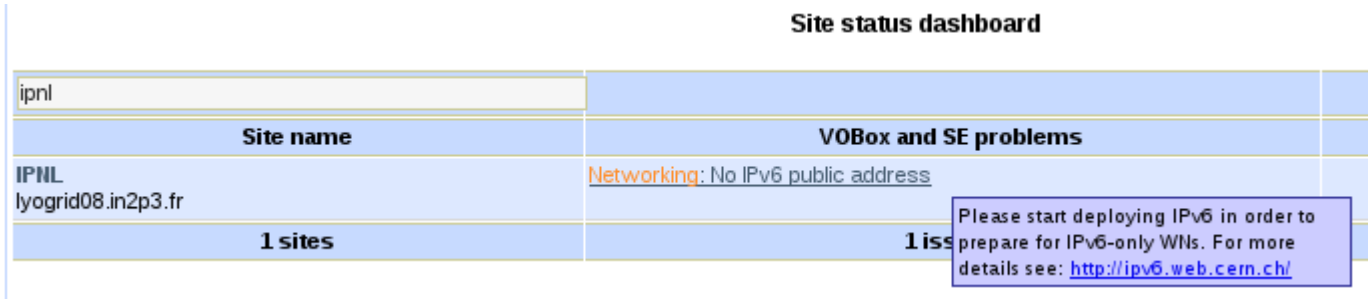


cf : https://twiki.cern.ch/twiki/bin/view/Main/RedirectorsSubscription

- **Classement des sites dans l'une des 2 fédérations : Production federation (PF) ou Transitional federation (TF)**
- **Les jobs exécutés dans sites de la PF accèderont aux données dans les deux fédérations.**
- **Les jobs exécutés dans sites de la TF n'accèderont qu'aux données de la TF**

4.2 Le futur

- **Alice :**



- **CMS : Production federation ou Transitional federation ?**

Fin

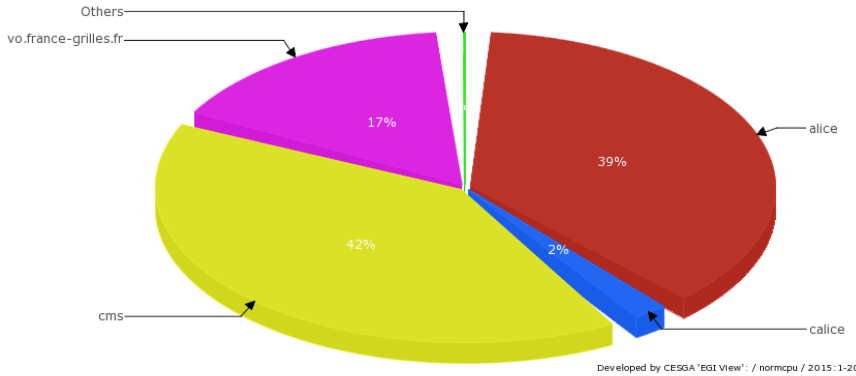
Merci pour votre attention

4.2 Backup slide : détail accounting cms

Developed by CESGA 'EGI View': / normcpu / 2014:1-2014:12 / SITE-VO / all (x) / GRBAR-LIN / l

2015-12-14 10:33

IN2P3-IPNL Normalised CPU time (kSI2K) per VO

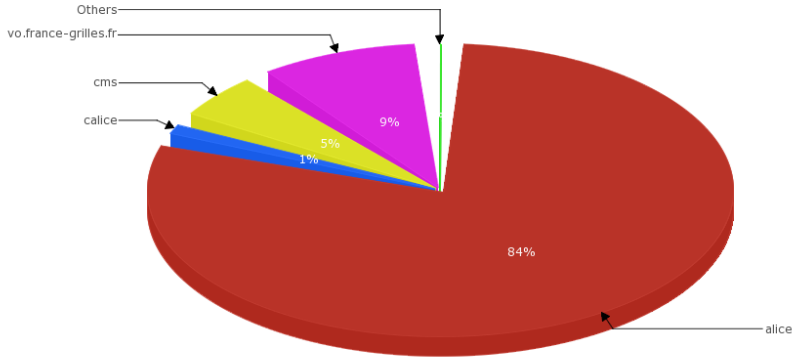


Accounting CMS 2014

Developed by CESGA 'EGI View': / normcpu / 2015:1-2015:12 / SITE-VO / all (x) / GRBAR-LIN / l

2015-12-14 10:33

IN2P3-IPNL Normalised CPU time (kSI2K) per VO



Accounting CMS 2015