



LCG-France Tier-1 & AF

Réunion de Coordination

Fabio Hernandez

fabio@in2p3.fr

Lyon, 15 novembre 2007

dapnia

cea

saclay



CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

► Table des Matières



- Réunions du GDB et du MB d'octobre et novembre 2007
- Disponibilité, fiabilité, efficacité du site
- Chantiers en cours
- Thème du jour: stabilité



- Agendas:
 - **GDB:** <http://indico.cern.ch/categoryDisplay.py?categId=31181>
 - **MB:** <http://indico.cern.ch/categoryDisplay.py?categId=666>
- Principaux sujets
 - CCRC 08
 - Job priorities
 - Pilot jobs
 - Monitoring
 - Middleware sous SL4
 - Déploiement de SRM v2.2

- Combined Computing Readiness Challenge
 - *"A combined challenge by all experiments to demonstrate the readiness of the WLCG computing infrastructure before start of data taking at a scale comparable to the data taking in 2008."*
- Intervenants
 - 4 expériences, tous les sites et les coordinateurs du service WLCG
- Calendrier
 - **Février 2008: pré-challenge**
 - *But: tester la disponibilité de tous les composants nécessaires pour les expériences*
 - *Débit n'est pas le plus important*
 - **Mai 2008: challenge**
 - *But: vérifier le bon fonctionnement de l'ensemble des composants au niveau de performance et de stabilité nécessaires pour l'acquisition de données prévue en 2008*
 - *Entre 40% et 50% de données d'une année nominale*

- Rôles des intervenants
 - Voir [présentation de Jamie Shiers](#)
 - Expériences
 - *End-end experiment planning, online/ offline coordination, goals, metrics, requirements & publication of these*
 - Sites
 - *Provision of needed resources & services, response to problems, active participation in coordination meetings etc.*
 - WLCG service coordination
 - *Overall coordination, maintenance of commun schedule & results, interaction with service, summary of results to MB/GDB etc.*
 - *Provide technical expertise to help debug & resolve (complex) problems*

► Politiques de sécurité



- Grid Security Policy
 - <https://edms.cern.ch/document/428008>
- Grid Site Operations Policy
 - <https://edms.cern.ch/document/819783>
- VO Operations Policy
 - <https://edms.cern.ch/document/853968>
- Pilot jobs policy
 - <https://edms.cern.ch/document/855383>

Sommes-nous suffisamment informés des droits et des obligations de notre site convenus dans ces documents?

Jobs Pilotes



- Accord des sites sur la volonté de permettre les jobs pilotes multi-utilisateur avec changement d'utilisateur Unix pour l'exécution du sous-job
- Conditions
 - Changement d'utilisateur doit se faire via un mécanisme commun (aux expériences et aux sites)
 - *glexec* est le seul candidat
 - *Un audit sécurité de glexec est en cours*
 - *glexec* doit être intégré et testé avec succès avec les différents systèmes de batchs utilisés par les sites WLCG
 - *Y compris BQS: ces tests sont sous notre responsabilité*
 - LCAS/LCMAPS doit être disponible en version client-serveur

Job priorities



- Job priorities



Short Term Solution

- Make the VOViews mutually exclusive

- on a CE not more than one VOView can match a FQAN
- Introduce a DENY rule on the Information system:

```
GlueVOViewLocalID: atlas
GlueCEAccessControlBaseRule: DENY:/atlas/Role=production
GlueCEAccessControlBaseRule: VO:atlas
[...]
GlueVOViewLocalID: /atlas/Role=production
GlueCEAccessControlBaseRule: VOMS:/atlas/Role=production
```

- Required simple changes in the WMS

- Required careful publication of the rule on in the GIP

Quel est l'état
d'avancement
chez nous?

Middleware pour SL4



- gLite 3.1 construit pour SL4 32 bits
 - Etat d'avancement disponible
 - <https://twiki.cern.ch/twiki/bin/view/EGEE/Glite31NodeTracker>
 - Résumé
 - LCG-CE: disponible
 - CREAM-CE: objectif de certification janvier 2008
 - BDII: disponible
 - FTM: disponible
 - LFC/Oracle: intégration
 - FTS v2/Oracle: intégration
 - Objectif: déploiement en production aux tier-1s en janvier 2008
 - MON: configuration via YAIM pas encore disponible
 - VOMS: certification
 - VO Box: certification

Middleware pour SL4 (suite)



- gLite 3.1 construit pour SL4 64bits
 - Priorités: WN, DPM, UI
 - *WN pour SL4 64bits est en phase de tests*

 **64 bit - dual arch install**
Enabling Grids for E-science

- In order to properly support **32bit** applications on **x86_64**, we will ship a number of **32bit** libraries with the **64bit WN**.
- For dual arch install, the following constraints apply
 - rpms must be at the same version
 - rpms must be installed simultaneously
 - shared files must be identical, EXCEPT;
 - libs must be relocated properly -> /lib64
 - executables will have the 64 bit version chosen
- Our management scripts need to be updated to accommodate packages which must be installed 32/64
- Dual architecture meta-rpms are proving problematic
 - Will have to move to 'groups'
 - #yum install glite-WN -> #yum groupinstall glite-WN

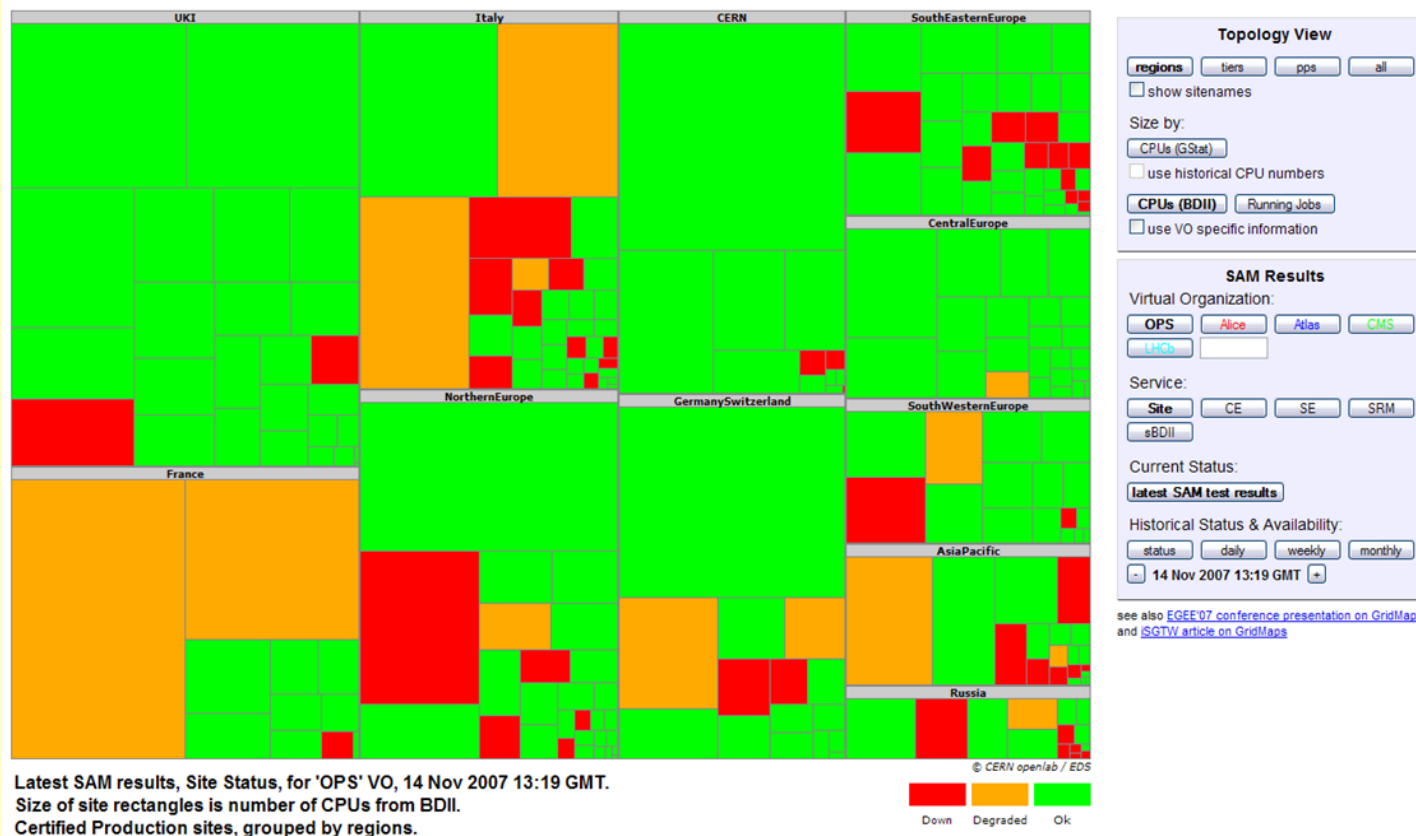
Est-ce suffisant pour nous permettre d'installer nos WN en SL4 64bits?

Monitoring



- Niveau grid

GridMap Prototype – Visualizing the "State" of the Grid



Source: <http://gridmap.cern.ch>

► Monitoring (suite)

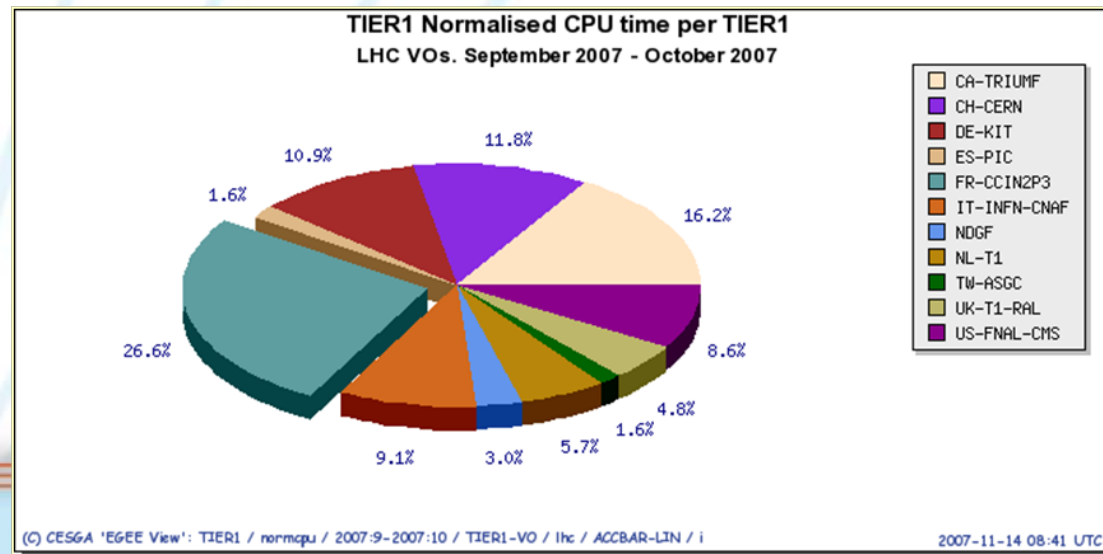


- Niveau site
 - Outil pour le monitoring au niveau site basé sur Nagios bien avancé
 - *Suivons-nous ce travail? Pourrions-nous en bénéficier?*

Accounting CPU



- Modification planifiée de la configuration de BQS pour utiliser les unités SI2000 selon le benchmark convenu par l'ensemble des sites
 - Détails: <http://indico.in2p3.fr/conferenceDisplay.py?confId=431>
- Contribution du tier-1 en temps CPU
 - Septembre – octobre 2007



Source: [EGEE Accounting Portal](#)

Accounting CPU (suite)



Comparison of Manual vs. Automatic CPU Accounting

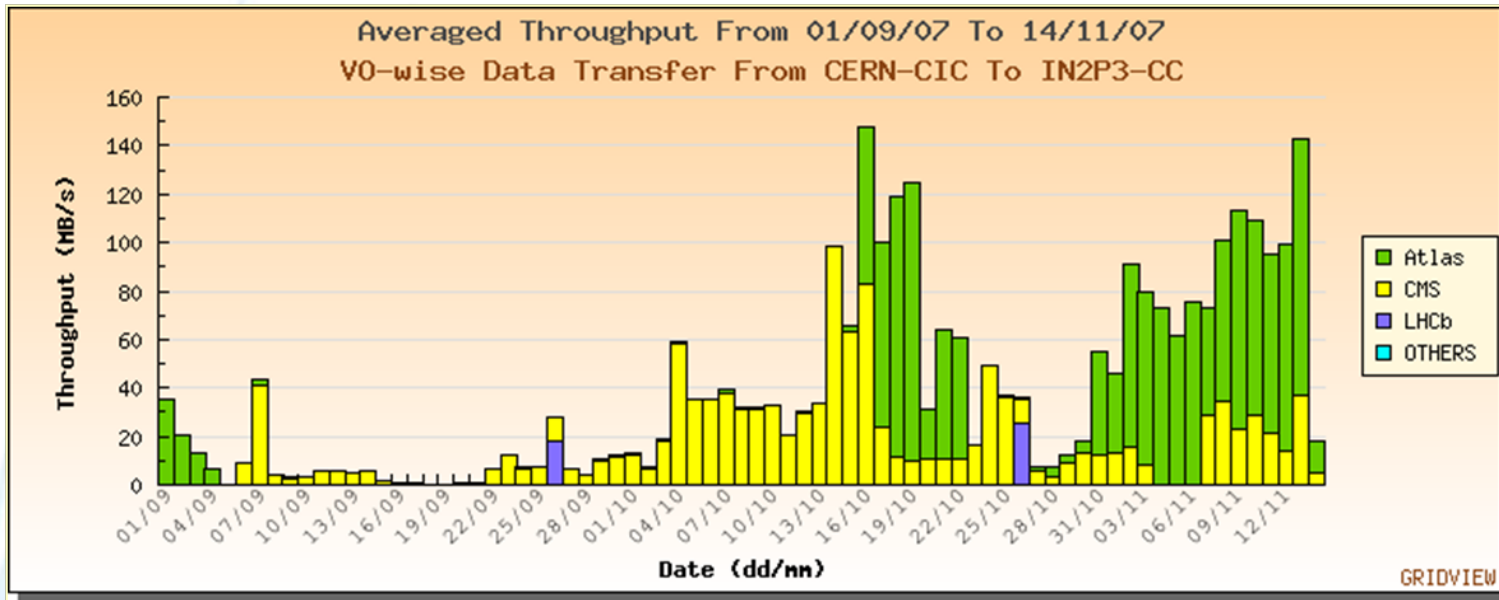
CPU Time in kH 2000														
		2007-01	2007-02	2007-03	2007-04	2007-05	2007-06	2007-07	2007-08	2007-09	2007-10	2007-11	2007-12	Aggregate
Alice	Manual	18 021	32 655	38 886	33 109	28 028	124 870	152 597	151 668	787 939	838 464	0	0	2 206 236
	Automatic	17 230	32 563	37 785	32 464	28 022	124 263	144 440	136 578	787 707	600 370			1 941 422
	Δ (automatic vs. manual)	-4,4%	-0,3%	-2,8%	-1,9%	-0,0%	-0,5%	-5,3%	-9,9%	-0,0%	-28,4%			-12%
Atlas	Manual	190 681	135 599	85 813	202 934	209 299	175 242	219 484	168 601	111 264	341 600	0	0	1 840 518
	Automatic	190 681	135 600	84 025	202 933	209 289	169 443	214 889	165 066	111 264	237 627			1 720 817
	Δ (automatic vs. manual)	-0,0%	+0,0%	-2,1%	-0,0%	-0,0%	-3,3%	-2,1%	-2,1%	-0,0%	-30,4%			-7%
CMS	Manual	83 135	118 364	86 624	167 267	66 313	66 101	232 528	334 313	471 174	362 673	0	0	1 988 491
	Automatic	83 113	118 364	86 612	167 262	66 313	59 065	233 100	329 708	471 146	259 102			1 873 785
	Δ (automatic vs. manual)	-0,0%	+0,0%	-0,0%	-0,0%	-0,0%	-10,6%	+0,2%	-1,4%	-0,0%	-28,6%			-6%
LHCb	Manual	159 138	138 702	29 857	24 647	20 308	50 348	135 945	68 870	16 774	40 381	0	0	684 969
	Automatic	159 086	138 722	29 402	24 638	20 308	50 332	126 999	68 866	16 774	23 942			659 069
	Δ (automatic vs. manual)	-0,0%	+0,0%	-1,5%	-0,0%	-0,0%	-0,0%	-6,6%	-0,0%	+0,0%	-40,7%			-4%

Ecart importants observés en octobre

Transfert de données



- Débit moyen journalier CERN→CCIN2P3
 - Sep-nov 2007



► Déploiement de SRM v2.2



- Tests en cours par les expériences
 - SRM v2.2 implémenté par une instance de dCache 1.8 dédiée



LHCb SRM v2.2 testing experience

- The test:
 - Pre-staged 12K files (about 2.2TB) in CASTOR@CERN dedicated pools
 - Performed FTS transfers through DIRAC DMS between CASTOR@CERN (srmv1) to IN2P3, SARA, CASTOR@CNAF, and StoRM@CNAF (all running srmv2).
 - FTS channels overloaded with 3 jobs; each jobs performing replications of 100 files.
- Satisfactory results with an average 1-2% failure rate. Nominal transfer rate for LHCb reached everywhere (40MB/sec).

WLCG GDB
CERN, 7 November 2007

3

Source: Flavia Dono, [GDB 07/11/2007](#)

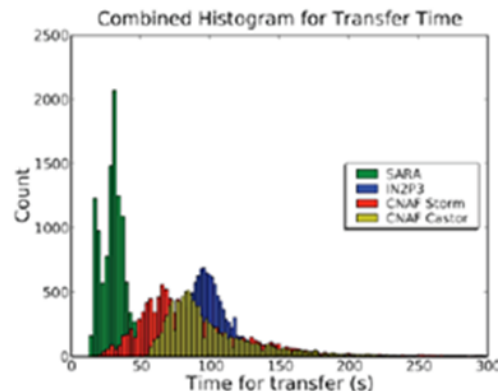
► Déploiement de SRM v2.2 (suite)



• Tests par les expériences (suite)



LHCb SRM v2.2 testing experience



• No errors observed at IN2P3

- SARA quite fast
- Need to understand how to tune the installations for better performance
- Check Andrew's slides for details

- Accepted proposal: LHCb will perform stress tests to understand the response of the systems.
- They will coordinate with sites and developers for assistance
- Ongoing production activities should not be disrupted. This might imply configuration changes: new FTS channels ? Storage resources ?

WLCG GDB
CERN, 7 November 2007

4

Source: Flavia Dono, [GDB 07/11/2007](#)

► Déploiement de SRM v2.2 (suite)



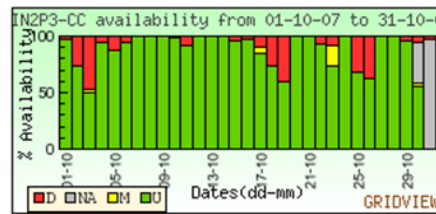
- Passage en production de dCache v1.8.x planifié pour le 26 et 27 novembre
 - NDGF, FZK déjà fait
 - SARA: planifié pour le 19 nov.

Disponibilité du site (1/4)



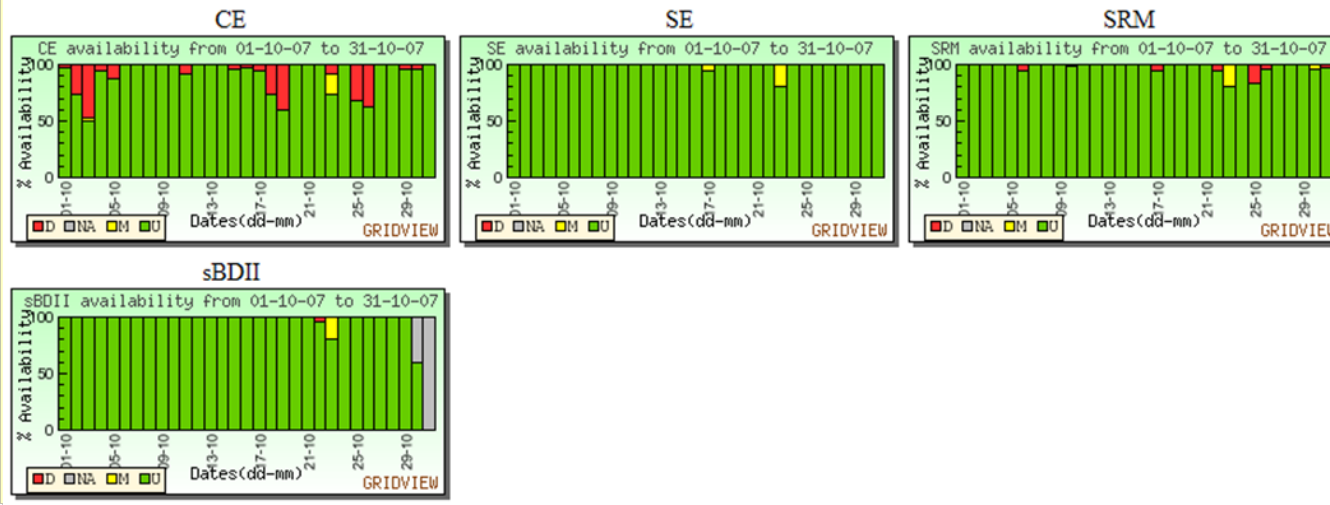
- Octobre 2007

Overall Service Availability for Site:IN2P3-CC VO:OPS (Daily Report)



Score: 90%

Individual Service Availability for site:IN2P3-CC VO:OPS (Daily Report)



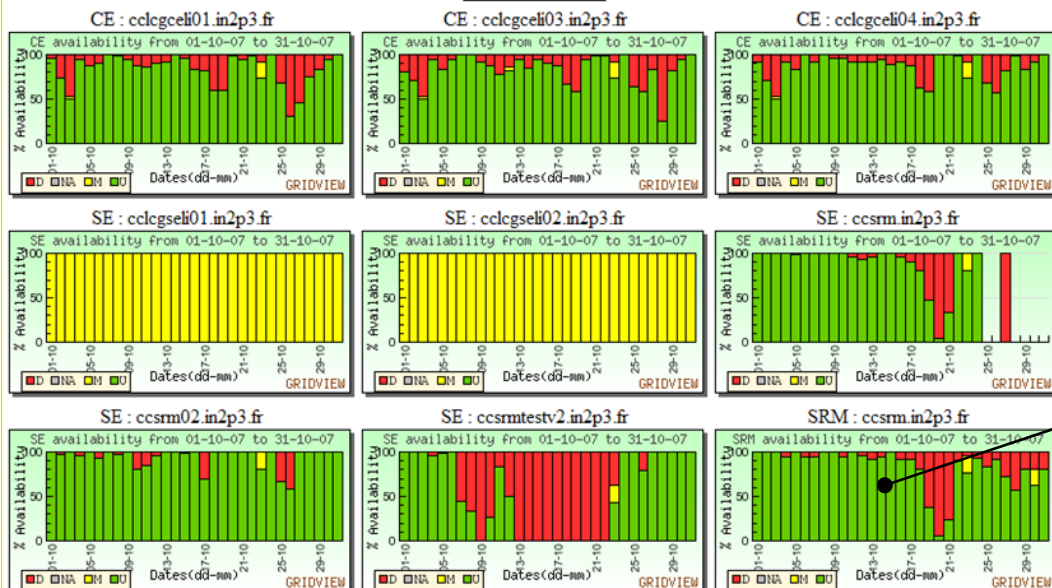
Disponibilité du site (2/4)



● Octobre 2007 (suite)

Service Instance Availability for site:IN2P3-CC VO:OPS (Daily Report)

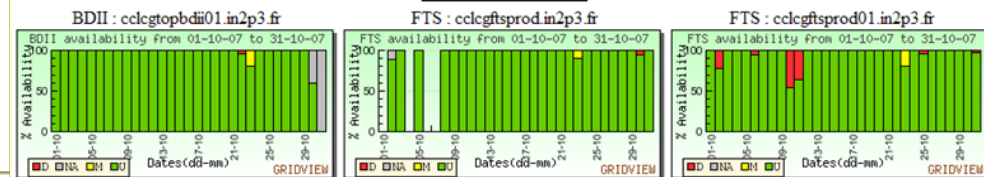
Site Services



Instabilité
simultanée de tous
les CEs.

Est-ce l'instabilité
du service SRM qui
affecte tous les
CEs?

Central Services



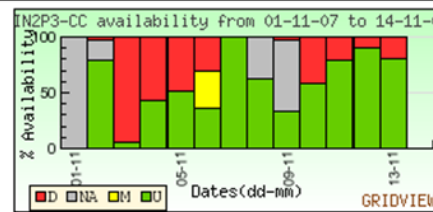
Source: <http://gridview.cern.ch>

Disponibilité du site (3/4)



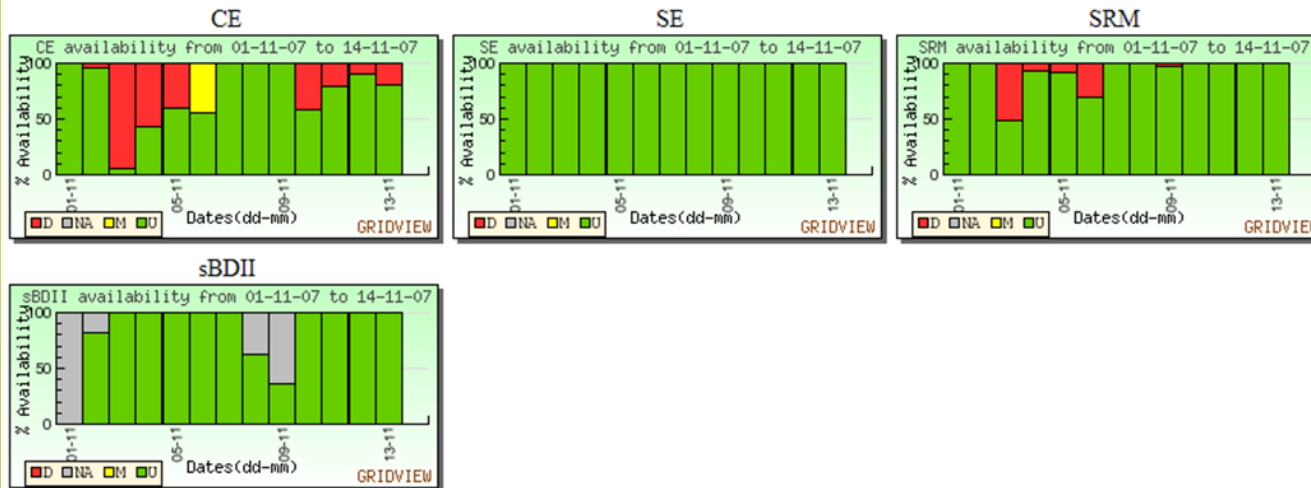
- Novembre 2007

Overall Service Availability for Site:IN2P3-CC VO:OPS (Daily Report)



Score jusqu'au 13 nov:
77%

Individual Service Availability for site:IN2P3-CC VO:OPS (Daily Report)

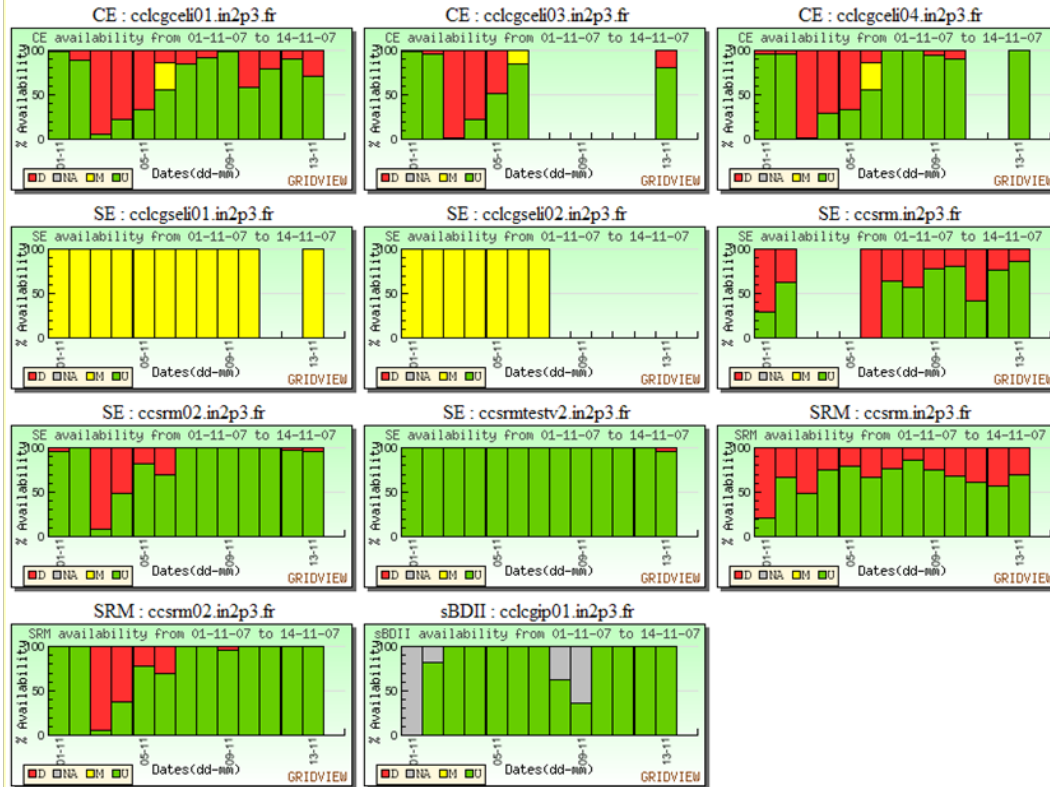


Disponibilité du site (4/4)



Service Instance Availability for site:IN2P3-CC VO:OPS (Daily Report)

Site Services

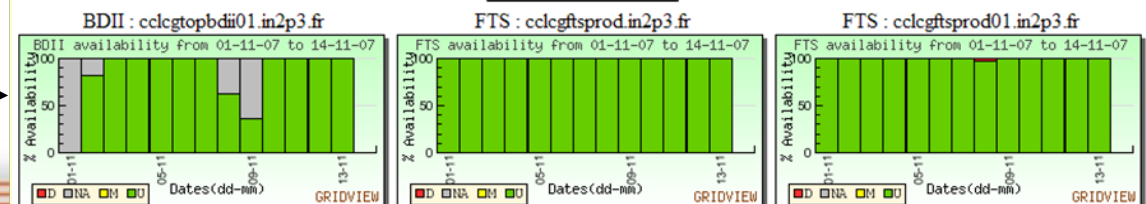


Instabilité
simultanée de tous
les CEs.

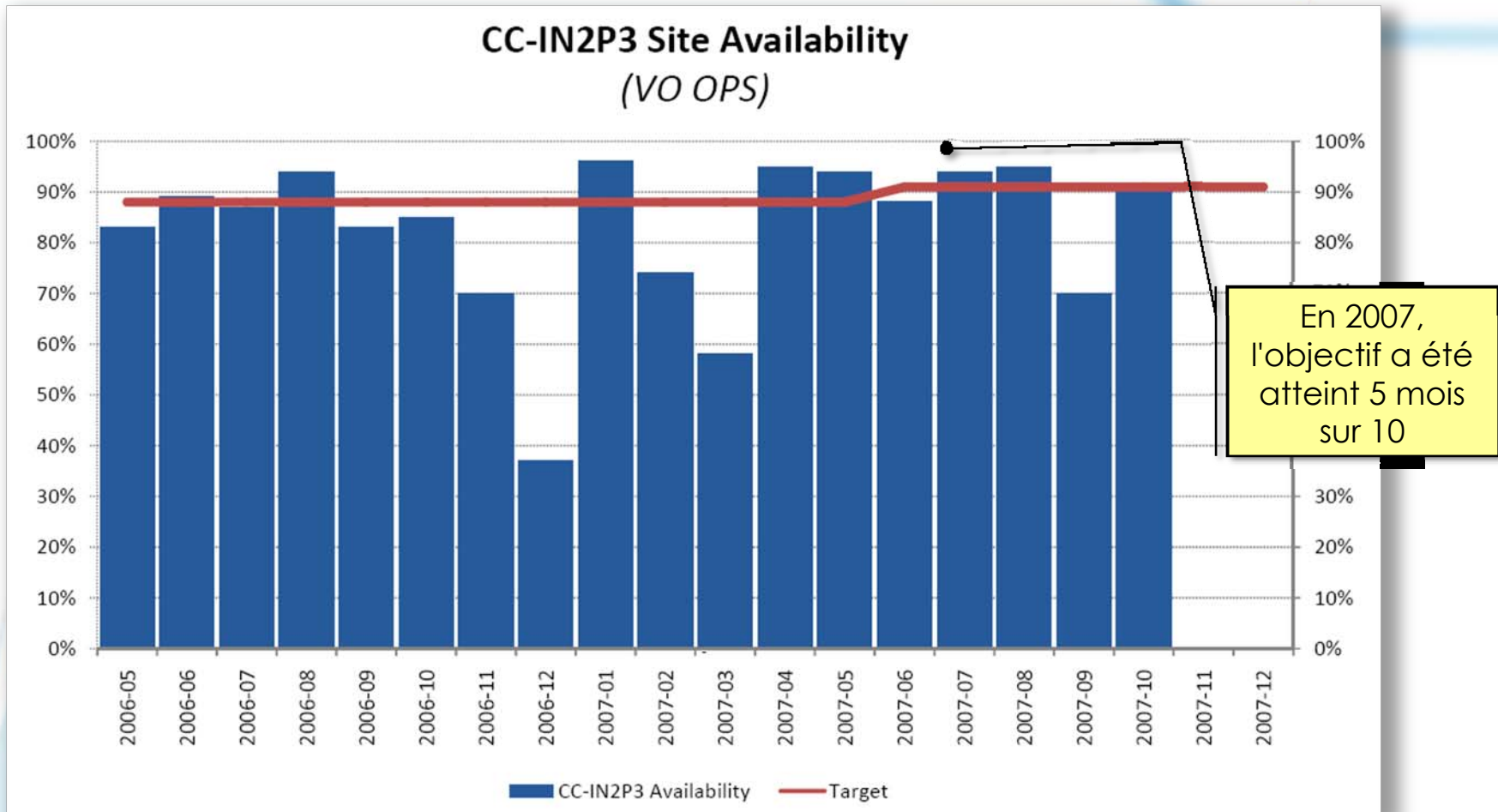
Plate-forme
matérielle
insuffisante?

Système
d'information

Central Services



Disponibilité du site: historique



- Quelques idées à explorer
 - Configuration de la chaîne de soumission et d'exécution afin que les jobs (courts) envoyés via le framework SAM soient mise en exécution "immédiate"
 - *Pour OPS mais aussi pour les autres VOs*
 - *Le résultat des sondes SAM pour OPS est utilisé par les expériences pour "blacklister" les sites où la sonde échoue*
 - *Moyen d'éviter l'expiration du proxy*
 - Modification de la fermeture du site grille lors d'un arrêt programmé prolongé
 - *Est-il possible de permettre la réception et l'exécution des jobs courts jusqu'au dernier moment?*

► Chantiers ouverts ou à ouvrir



- Validation le middleware sur SL4 64bits
 - En particulier pour les worker nodes
 - Implémenter la répartition de ressources pour le tier-1 et tier-2 en accord avec nos engagements
- Finaliser la configuration de tous les services grid avec le niveau de **redondance** nécessaire
- Accounting
 - Séparation tier-1, tier-2 et AF pour l'accounting interne
- Outils et procédures d'exploitation
 - Monitoring
 - Interface d'exploitation des services grid
 - Formalisation des procédures
- CREAM-compatible BQS-backed computing element

- Stabilité!!!
- Prochaine réunion
 - 13 décembre 2007
- Agendas de toutes les réunions
 - <http://indico.in2p3.fr/categoryDisplay.py?categId=102>

► Questions/Commentaires

