LCG-France Tier-1 & AF

Réunion de Coordination

Fabio Hernandez
fabio@in2p3.fr

Lyon, 11 janvier 2008

dapnia
cea
saclay

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Table des Matières

- Réunions du GDB et du MB de janvier 2008
- Chantiers en cours
- Disponibilité, fiabilité, efficacité du site
- CCRC 08
- Thème du jour

GDBs & MBs

- Agendas:
  - GDB: http://indico.cern.ch/categoryDisplay.py?categId=3l181
  - MB: http://indico.cern.ch/categoryDisplay.py?categId=666
- Principaux sujets traités
  - CCRC 08
  - CPU benchmarking
  - Fichiers de faible taille
  - Déploiement de SRM v2.2

CPU Benchmarking

Standard benchmarks

CERN IT Department

- Compulsory:
  - SPECint_base 2000 (concurrent, transition only)
  - SPECint_base 2006 (concurrent)
  - SPECfp_base 2006 (concurrent, gcc 4)
- Optional, cross-checks only:
  - SPECint_rate 2006
  - All above as 64-bit applications
  - SPECint compiled with gcc 4
  - All above under SL5
- Notes:
  - "concurrent" – as many instances as there are cores run in parallel, results added over all cores
  - Applications are 32 bit unless noted otherwise

Source: Helge Meinhard, GDB 09/01/2008

Persistance des space tokens

Executive summary: dCache

- Keep T1D0 spaces relatively small
  - Use them only as buffers for writing into the storage system
- Keep most of the T1D0 disk space unassigned to any space tokens
  - Can be used for restoring large data sets concurrently
- Possibly configure paths to allow for the selection of specific pools when recalling files from tape
  - Depending on name space layout per experiment

Maarten Litmaath, GDB, 2008/01/09 12

Source: Maarten Litmaath, GDB 09/01/2008

Cette proposition, est-elle acceptable pour nous?

Modification de la publication des clusters et subclusters

Issues with Current Deployment

Enabling Grids for E-sciencE

- **Adding new hardware types or OSes requires sites:**
  - add complete new CE nodes.
  - unique queues to distinguish resources.
- **SubClusters can currently contain intersecting WN sets.**
  - This has always given gstat, the quarterly reports and now gridmap an impossible task when CPU counting.
- **More information is published than is needed.**
  - e.g CERN publishes 21 GlueClusters and 21 GlueSubClusters
  - Only one GlueCluster and 2 GlueSubClusters representing SL4-i686 and SL4-x86_64 is needed.

Steve Traylen, CERN, 9th January 2008 8

Short Term Fix at an lcg-CE Site

Enabling Grids for E-sciencE

1. **Split the CE node type into three (yaim) node types.**
   1. **CE-ClusterPublisher** - Publishes GlueCluster and GlueSubCluster
   2. **CE-GateKeeper** - Configuration of lcg-CE (or creamCE).
   3. **CE-CePublisher** - Publishes the GlueCE (and VOView) objects.
   - These may well run on the same physical node of course.
   - None of these components interact with one another.
   - Only a case of detangling their configuration.
     - One YAIM function - *config_gip_ce* - needs to be split.
2. **Extend the ClusterPublisher so GlueCEs can be joined to named GlueClusters.**
3. **Extend the ClusterPublisher to support multiple GlueSubClusters joined to GlueClusters.**
4. **Extend the ClusterPublisher to support multiple GlueClusters.**
   - Only needed for an lcg-CE world.

Steve Traylen, CERN, 9th January 2008 9

Source: Steve Traylen, GDB 09/01/2008

Tests d'interaction glexec et système de batch

Batch testing

- CERN LSF https://twiki.cern.ch/twiki/bin/view/FIOgroup/FsLSFGridglExec
- NIKHEF PBS
- CC-IN2P3 BQS
- LAL PBS
- CESGA SGE
- Others?

- CERN Tests are a good example. NIKHEF will publish a generic set soon.

Science & Technology Facilities Council

Timescales

- glexec review – done
- Glexec testing - started
- Framework Review
  - Team formed.
  - Mandate defined (confirm with MB next week)
- Batch system testing
  - Started but not yet full coverage
- Glexec certification and deployment
  - Once testing
- LCAS/LCMAPS service
  - Still in development due now

Science & Technology Facilities Council

10

Source: John Gordon, GDB 09/01/2008

Performances I/O bande

FIO Analysis

CERN IT Department

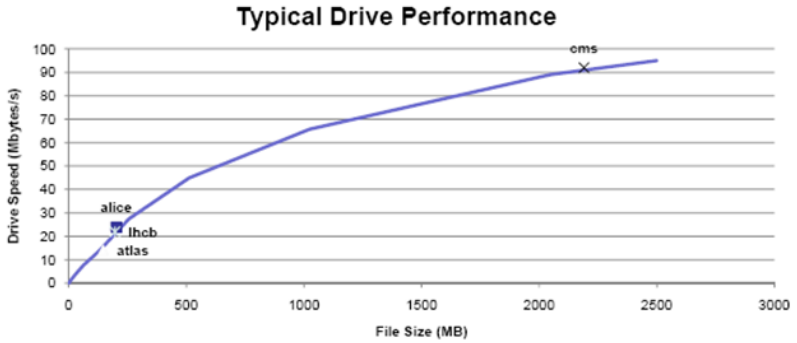
- Data collected during Nov/Dec 2007
  - Distribution of file sizes on tape
  - Tape mounts and performance
  - Production tapes only (no user tapes)
- Root causes identified
  - Small file sizes
  - Repeated mounting

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

3

File size and performance

CERN IT Department



Alice	Atlas	CMS	LHCb
200 MB	150 MB	2200 MB	200 MB

- Tape drives need to stream at high speeds to achieve reasonable performance.
- Per-file overheads from tape marks lead to low data rates for small files
- LHC tape infrastructure sizing was based on 1-2GB files.

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

4

Source: Tim Bell, MB 08/01/2008

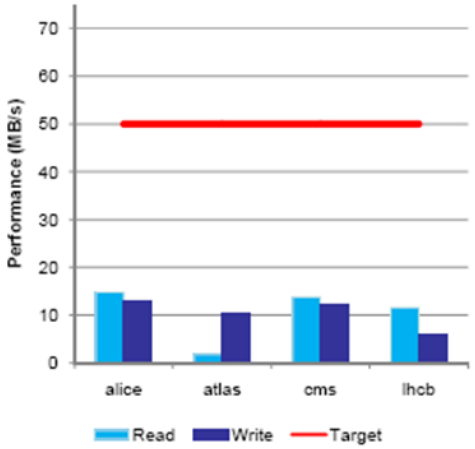
Performance I/O bande (suite)

FIO

Total performance to tape

CERN IT Department

- Planning was based on total performance of 50MB/s.



VO	File Size	Mounting Overhead
Alice	✗✗	✓
Atlas	✗✗	✗✗
CMS	✓✓	✗✗
LHCb	✗✗	✗

- Total performance is based on the sum of data transferred against the total time spent on drives (including mount unmount time).

CERN - IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

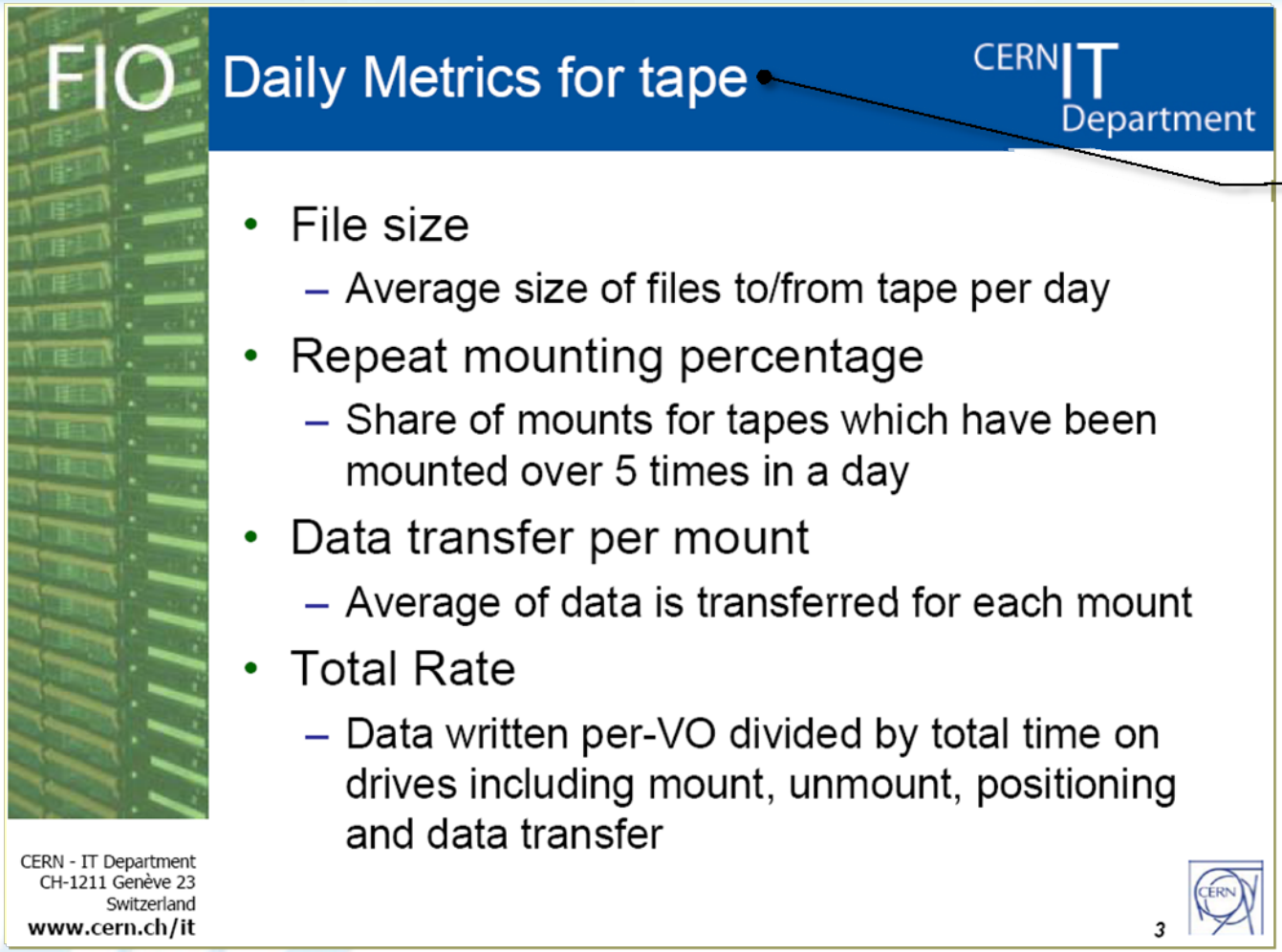
8

Source: Tim Bell, MB 08/01/2008

CCRC 08: rappel

- 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
  - 4 – 29 Février 2008: pré-challenge
    - *But: tester la disponibilité de tous les composants nécessaires pour les expériences*
    - *Débit n'est pas le plus important*
  - 5 – 30 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*

CCRC 08 (suite)



Avons-nous la possibilité de collecter systématiquement cette information?

Source: Tim Bell, CRRC 08 F2F meeting 10/01/2008

CCRC 08 (suite)

- Réunion au CERN le 10/01/2008
  - http://indico.cern.ch/conferenceDisplay.py?confId=24844
  - Avons-nous maintenant suffisamment d'information de la part des expériences pour configurer nos systèmes de stockage?
    - *dCache et HPSS*
  - Quels sont les éléments de notre site à monitorer pendant l'exercice?
    - *Les métriques ont-elles été identifiées?*

Middleware pour SL4

- gLite 3.1 pour SL4 32 bits & 64 bits

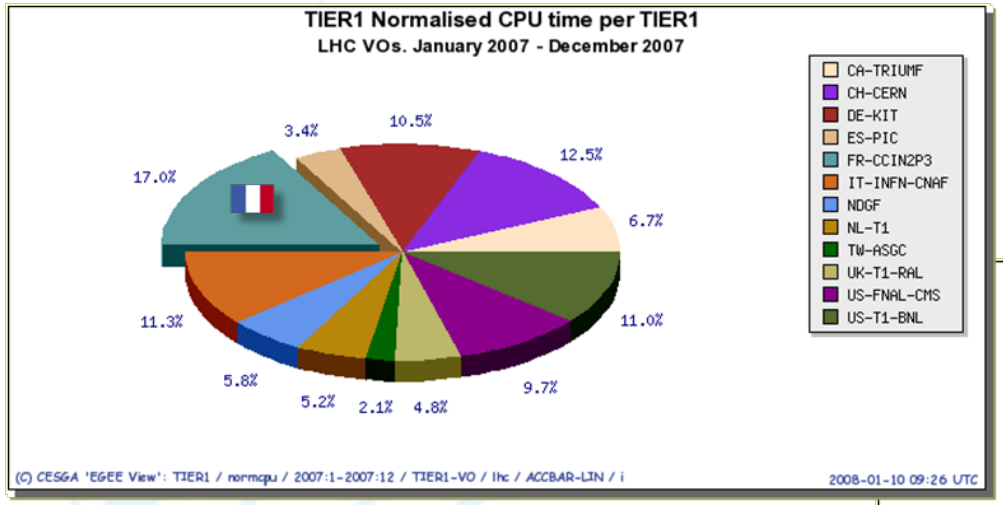
Le passage en SL4 32bits des nos CEs est donc possible

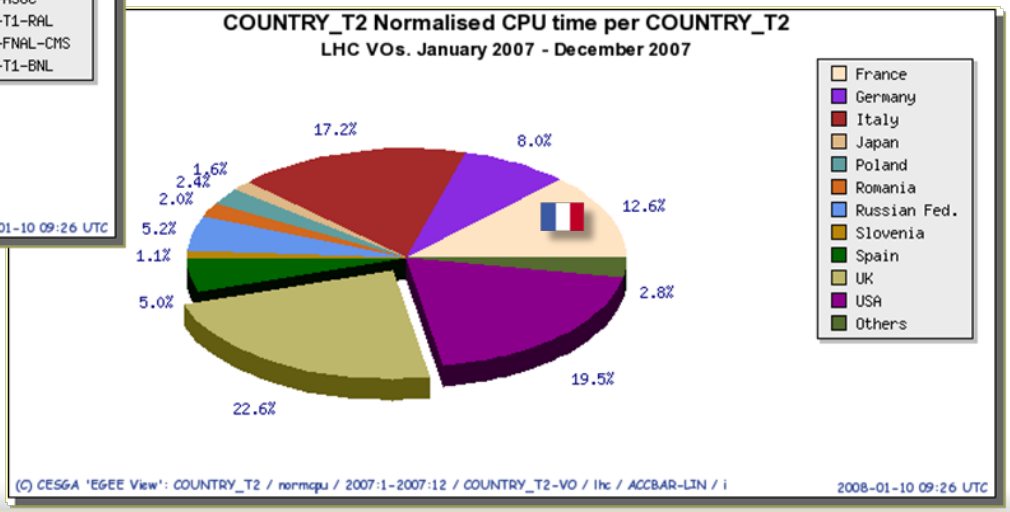
Status of individual services scheduled for gLite 3.1 on SL4

service	SL(C)4/i386	SL(C)4/x86_64	Comments
glite-WN	Released	Certification	
glite-UI	Released	Build	
glite-AMGA_postgres	PPS	Integration	
glite-BDII	Released	Configuration	
lcg-CE	Released	Build	
glite-CREAM	Integration	Build	
glite-FTA_oracle	Integration	Build	
glite-FTS_oracle	Integration	Integration	
glite-FTM	Released	Integration	
glite-LB	Certification	Build	WMS/LB is being installed as an 'experimental service'
glite-LFC_mysql	Released	Certification	
glite-LFC_oracle	Released	Certification	
glite-MON	Certification	Build	
glite-PX	PPS	Integration	
glite-SE_classic	Certification	Build	
glite-SE_dcache_*	PPS	Integration	
glite-SE_dpm_disk	Released	Certification	
glite-SE_dpm_mysql	Released	Certification	
glite-SE_dpm_oracle	Integration	Integration	
glite-TORQUE_utils	Released	Build	
glite-TORQUE_client	Released	Build	
glite-TORQUE_server	Released	Build	
glite-VOMS_oracle	Released	Integration	
glite-VOMS_mysql	Released	Integration	
glite-VOBOX	PPS	Integration	
glite-WMS	Certification	Build	WMS/LB is being installed as an 'experimental service'

Accounting CPU

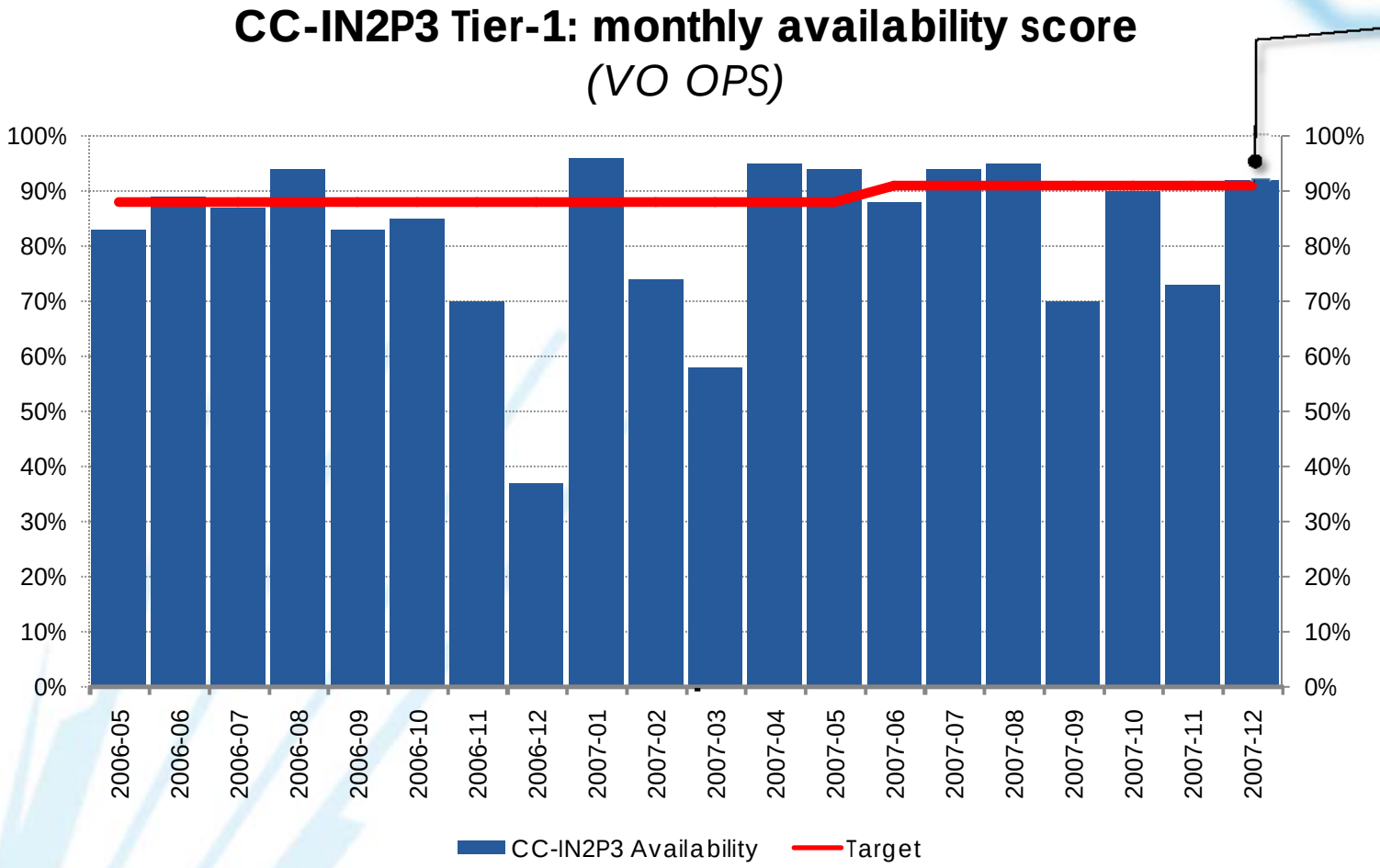
- Contribution du **tier-1** et des **tier-2s français** en temps CPU
  - Janvier-décembre 2007





Source: [EGEE Accounting Portal]

Disponibilité du site

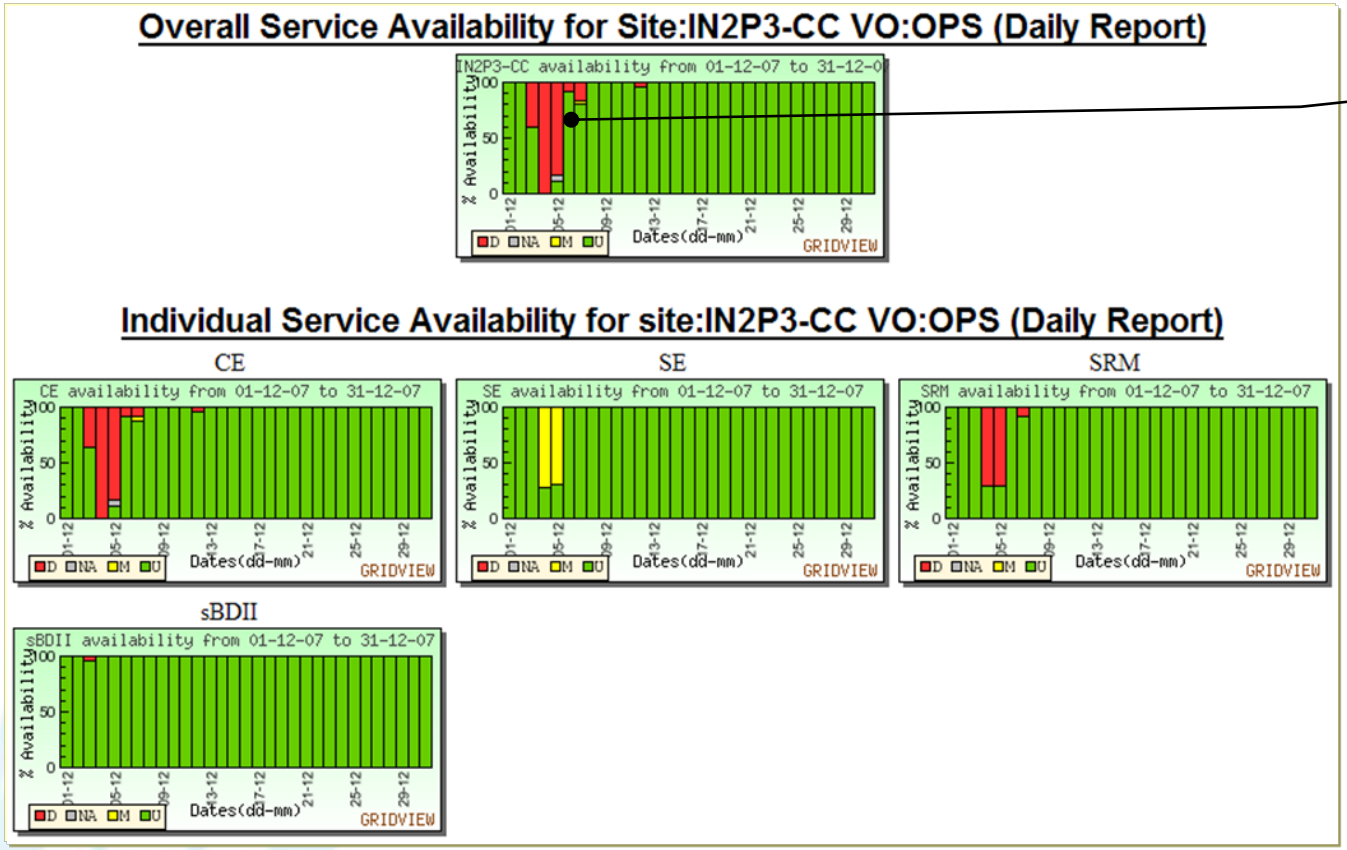


En décembre 2007, l'objectif de disponibilité a été atteint

Disponibilité du site (1/3)

- Décembre 2007

Score: 92%

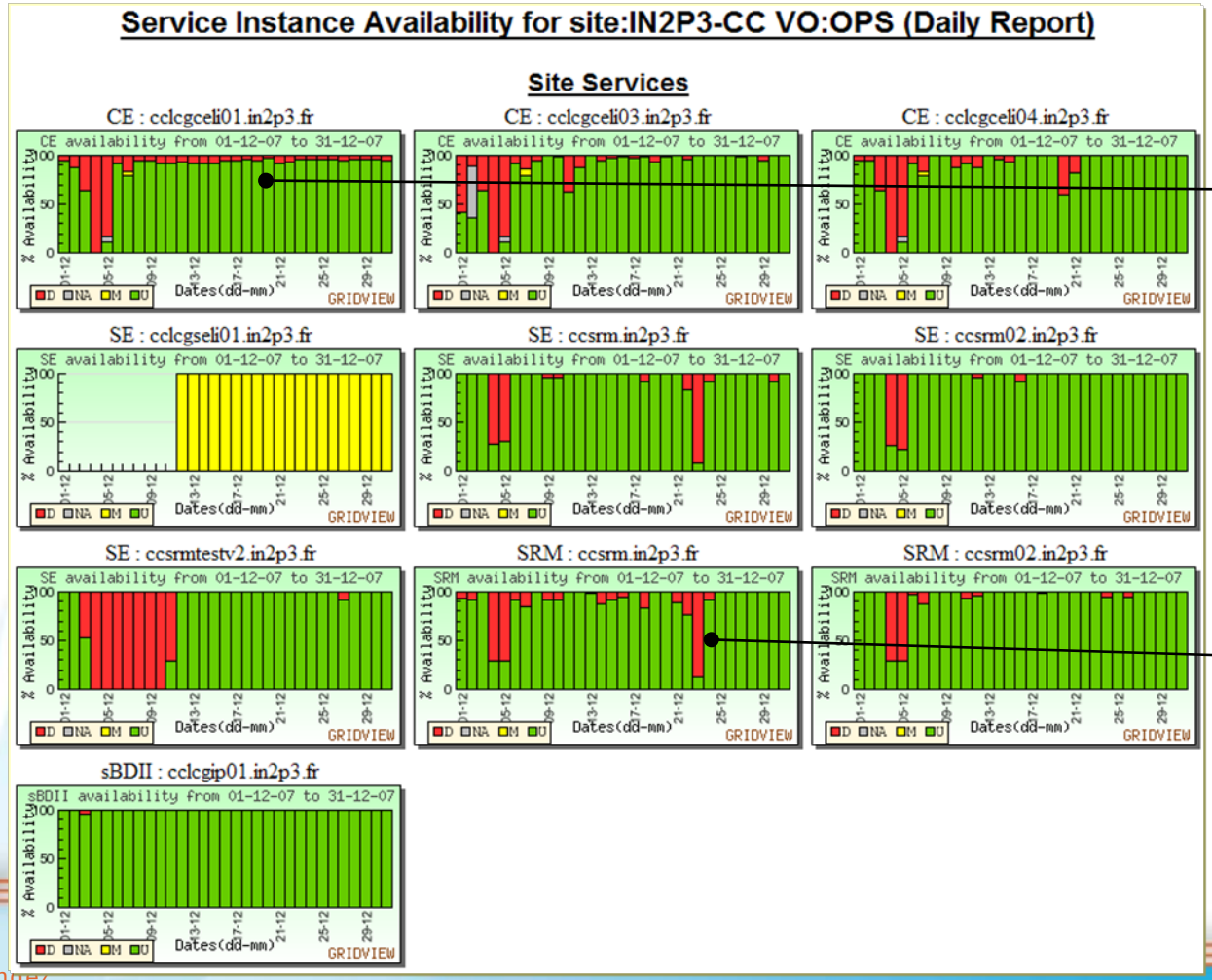


Arrêt programmé du 04/12/2007 (downtime enregistré du 03/12 13h au 05/12 18h)

Source: http://gridview.cern.ch

Disponibilité du site (2/3)

- Décembre 2007 (suite)



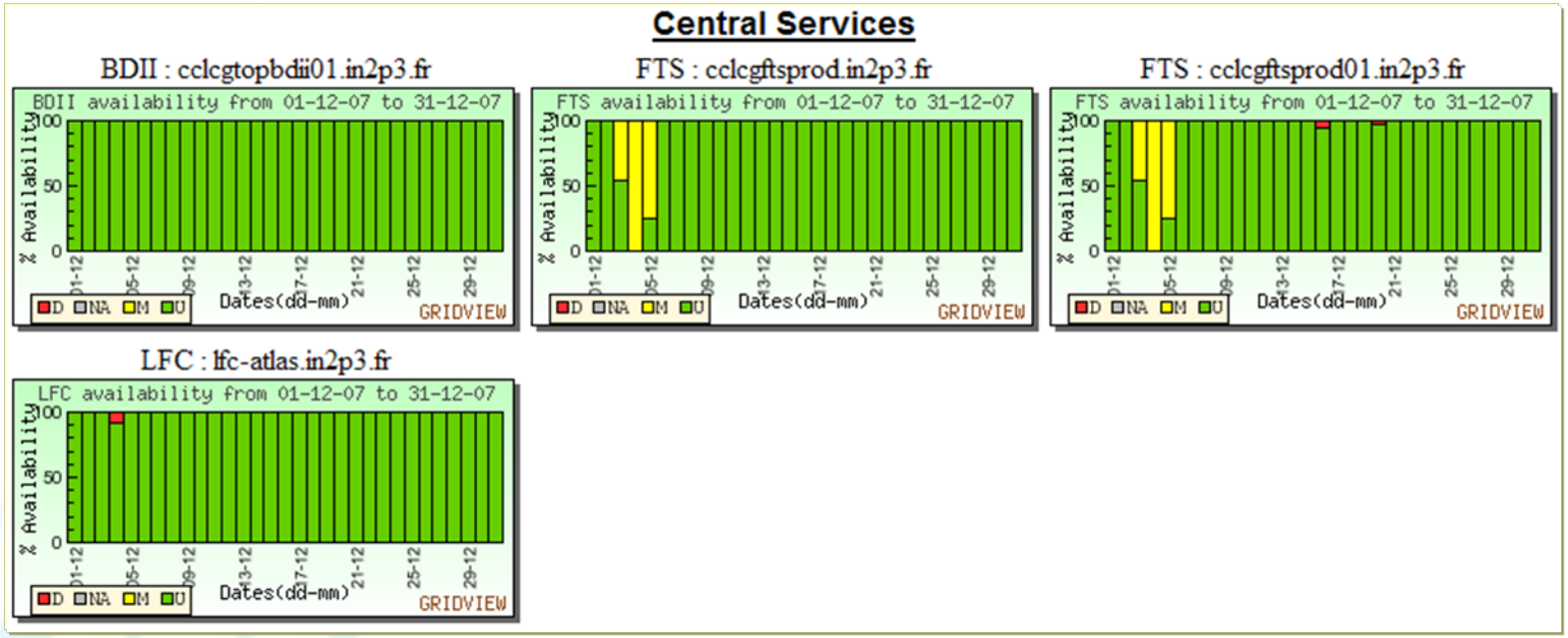
Ce CE s'est montré moins stable que les autres. Des raisons particulières?

La raison de cette indisponibilité est-elle comprise?

Source: http://gridview.cern.ch

Disponibilité du site (3/3)

- Décembre 2007 (suite)



VO Boxes: rappel

- Travail systématique nécessaire pour identifier
  - Configuration requise de la machine
    - *Mémoire, partitionnement du disque*
  - Niveau de redondance nécessaire
  - Service système à démarrer
    - *gsiSSH, PNFS, …*
  - Services spécifiques VO à démarrer
  - Tâches périodiques
    - *Rotation des logs, backups, …*
  - Monitoring
    - *Système et services VO*
  - Alertes
- Documentation du processus

Revue des chantiers en cours

- Intégration
  - Tests de LCG-CE sous SL4
  - Développement de l'interface BQS pour CREAM-CE
  - Tests d'impact du changement d'utilisateur UNIX (par glexec) pour un job BQS
- Exploitation
  - Consolidation des services grid: LFC, FTS, CE, système d'information
  - Intégration des VO boxes à l'exploitation standard
  - Procédures d'exploitation services grid
  - Déploiement de LFC-RO pour LHCb
  - Plate-forme de monitoring basée sur NAGIOS
  - Séparation des la consommation CPU tier-1 et tier-2 dans la comptabilité interne
  - Arrêt de FTS v1.5

Aujourd'hui

- Consolidation
  - Système d'information
  - CEs
- Monitoring basé sur NAGIOS
- Prochaine réunion
  - Jeudi 7 février 14h salle 202
- Agendas de toutes les réunions
  - http://indico.in2p3.fr/categoryDisplay.py?categId=102

Questions/Commentaires