



LCG-France Tier-1 & AF

Réunion de Coordination

Fabio Hernandez
fabio@in2p3.fr

Lyon, 14 mai 2009



l r f u
cea
saclay

► Table des Matières



- Réunions GDB et MB de mai 2009
- Disponibilité, fiabilité, efficacité des sites
- Chantiers en cours
- Thème(s) du jour

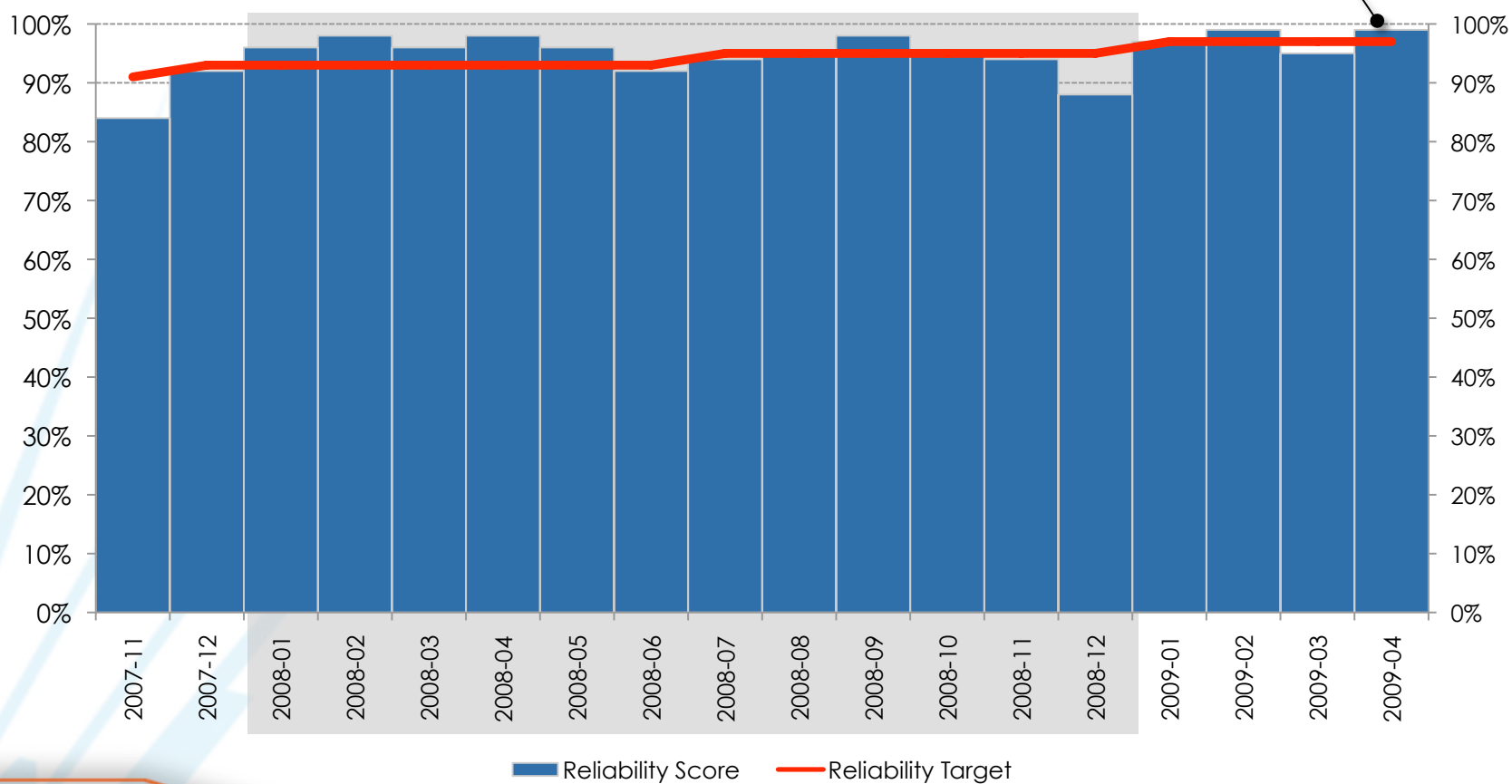
► GDBs & MBs



- Agendas:
 - GDB: <http://indico.cern.ch/categoryDisplay.py?categId=31181>
 - MB: <http://indico.cern.ch/categoryDisplay.py?categId=666>
- Principaux sujets traités
 - Préparation STEP'09
 - Scientific Linux 5

► Fiabilité du tier-1

LCG-France tier-1 - Reliability (VO OPS, last 18 months)

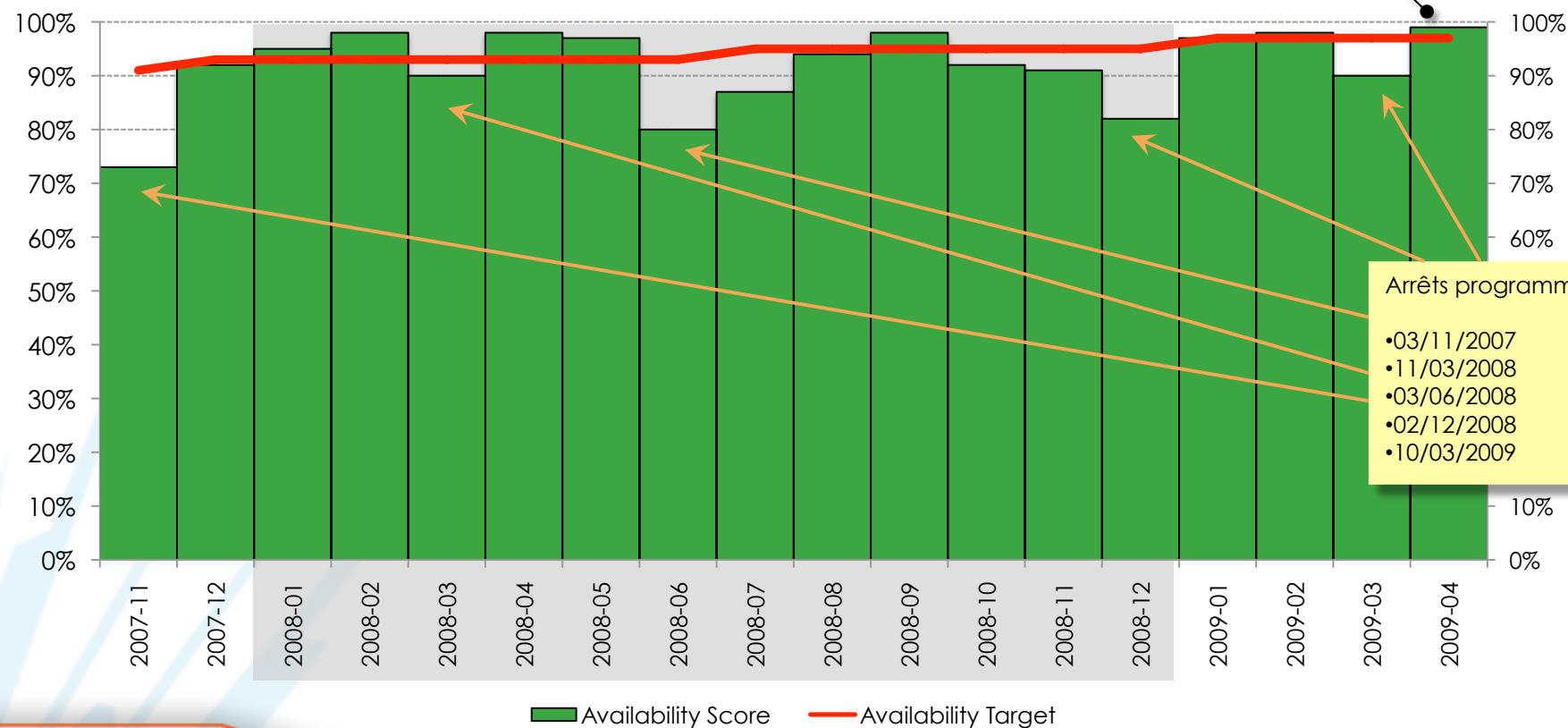


Source: [WLCG Reliability Reports](#)

► Disponibilité du tier-1



LCG-France tier-1 - Monthly availability score
(VO OPS, last 18 months)



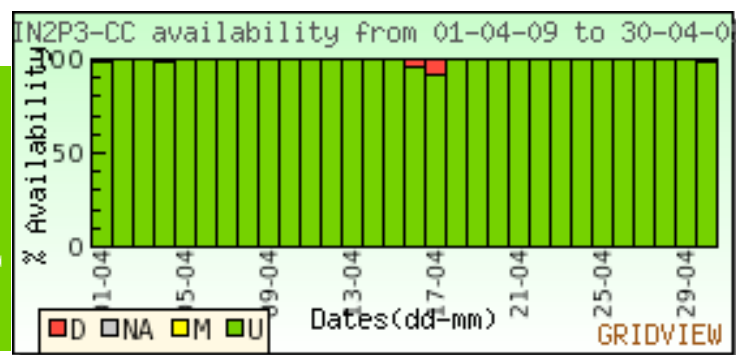
Source: [WLCG Reliability Reports](#)

► Disponibilité & Fiabilité tier-1

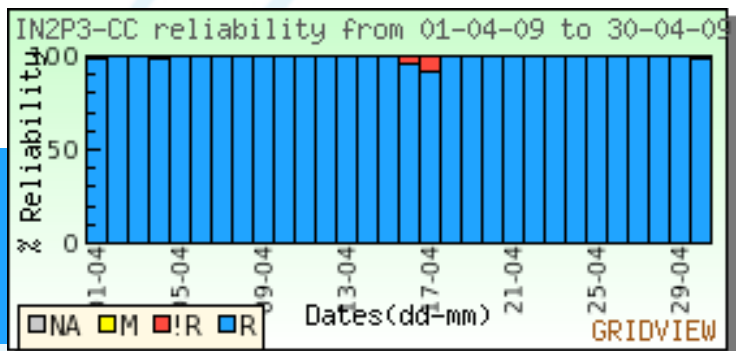


Avril 2009 (VO OPS)

Disponibilité



Fiabilité



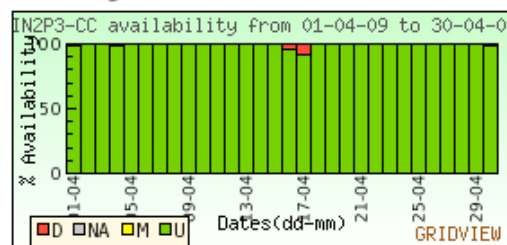
Source: <http://gridview.cern.ch>

► Disponibilité tier-1



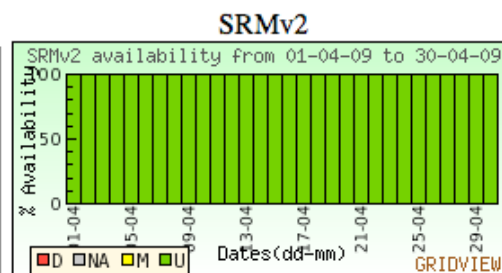
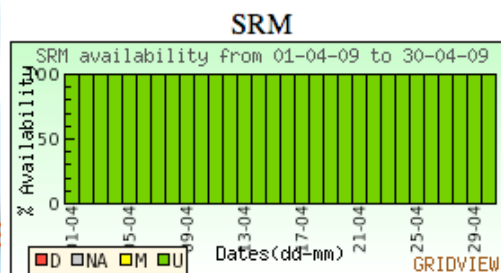
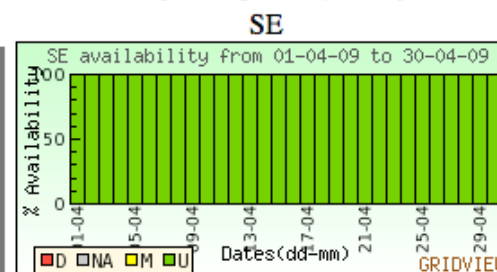
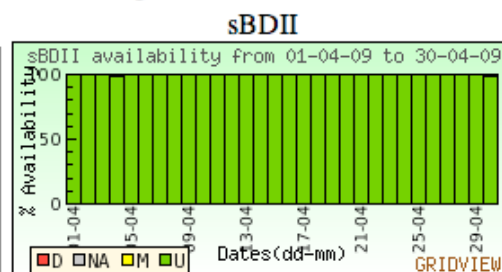
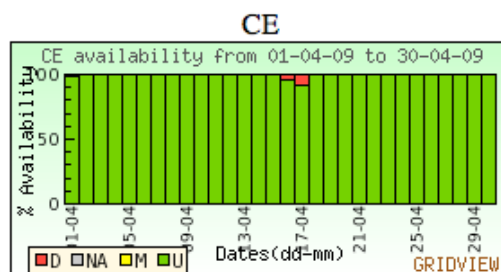
- Avril 2009

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



Score: 99%

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



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

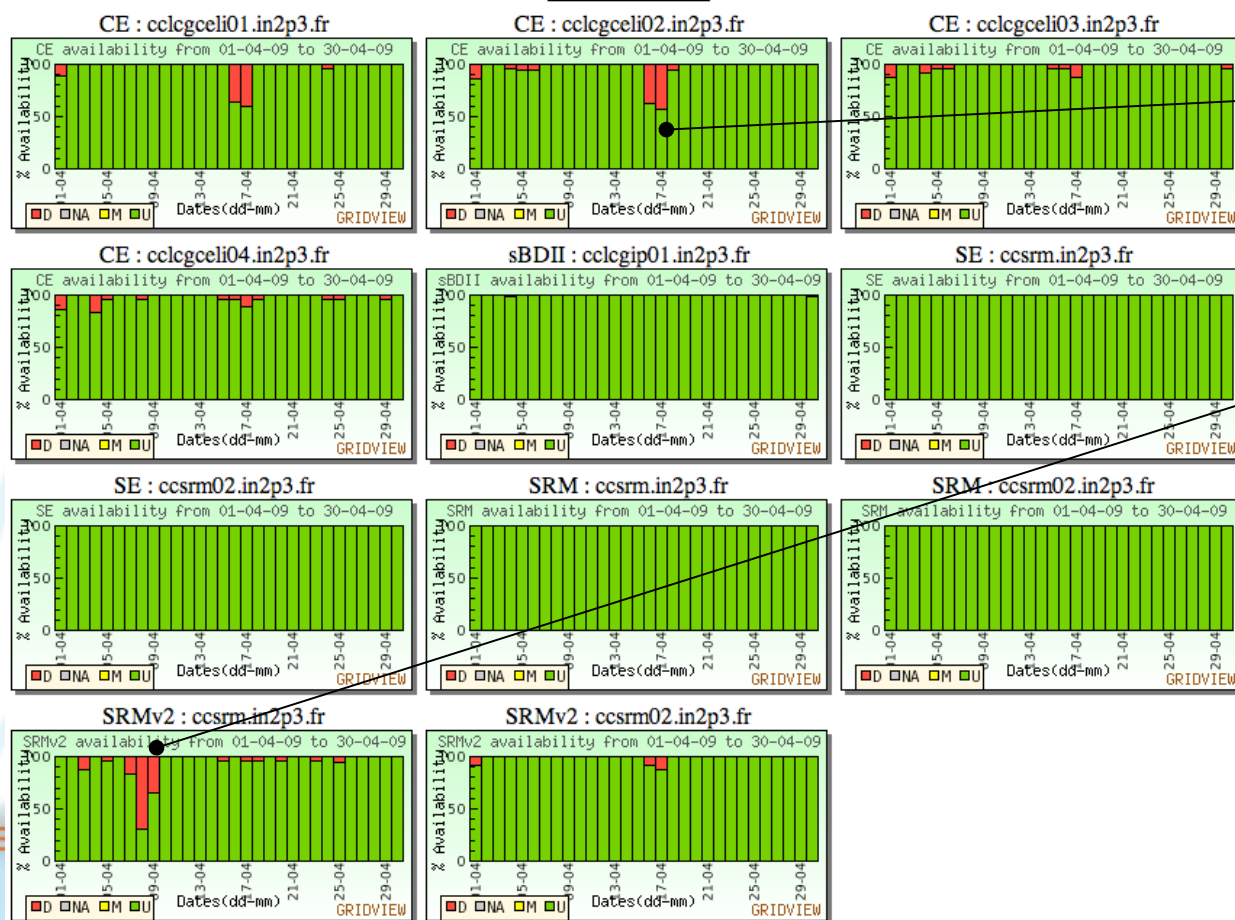
► Disponibilité tier-1 (suite)



• Avril 2009 (suite)

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

Site Services



???

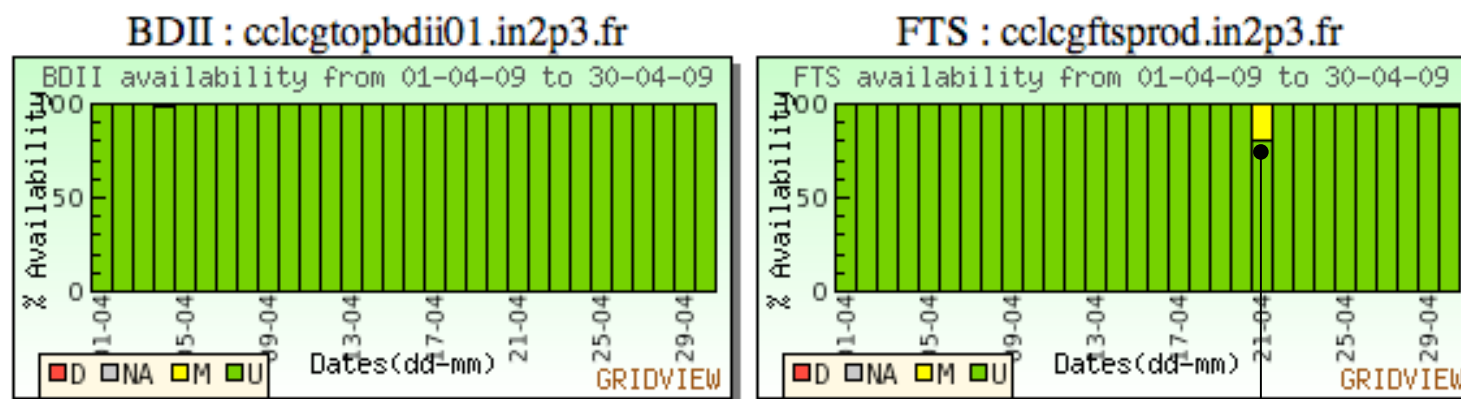
???

► Disponibilité tier-1 (suite)



- Avril 2009 (suite)

Central Services

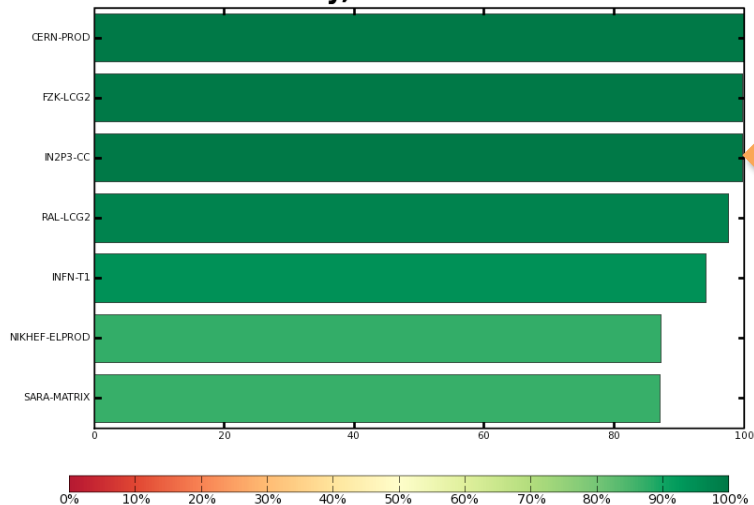


Shutdown du service
pour mise en
production de FTS v2.1
sous SL4 64bits

► Disponibilités tier-1s + CERN

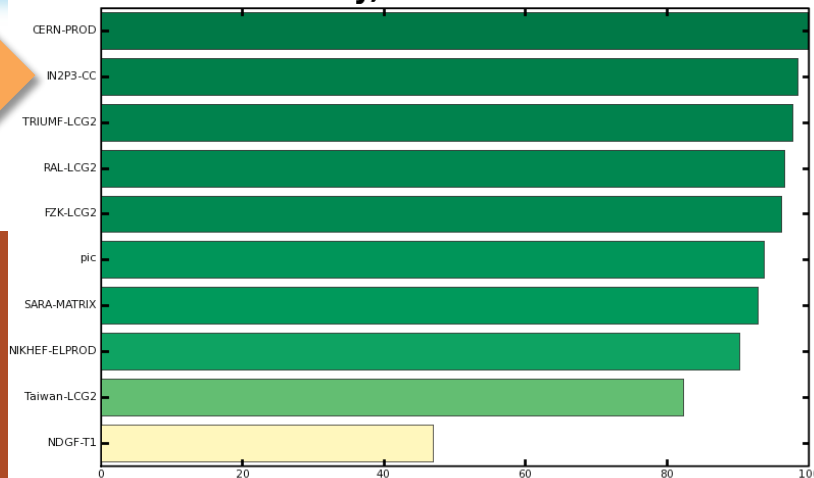


Site Availability, 2009-04-01 - 2009-04-30



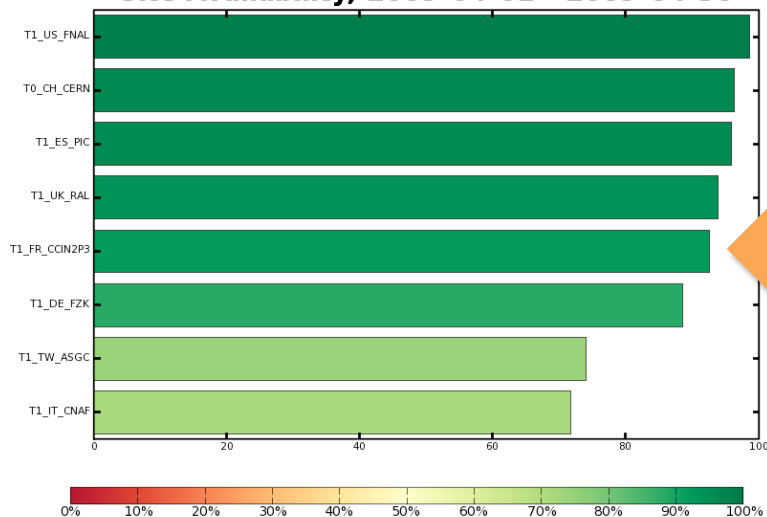
ALICE

Site Availability, 2009-04-01 - 2009-04-30



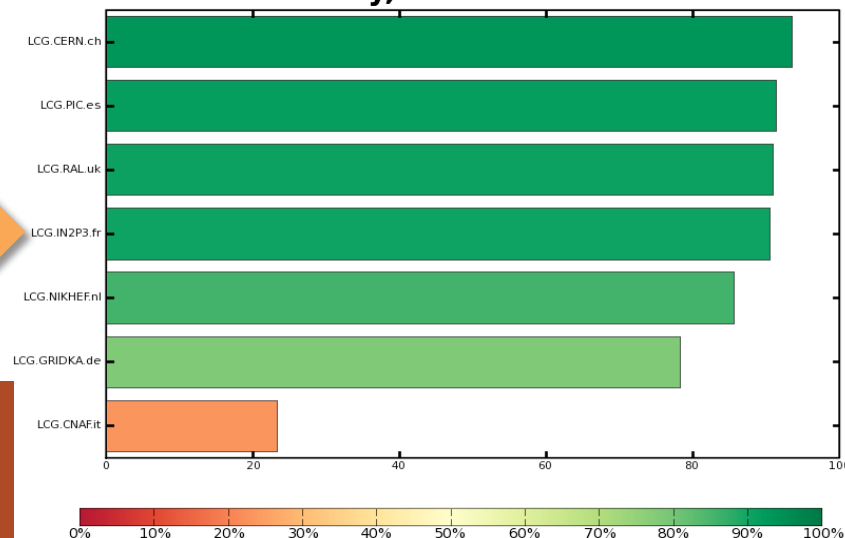
ATLAS

Site Availability, 2009-04-01 - 2009-04-30



CMS

Site Availability, 2009-04-01 - 2009-04-30

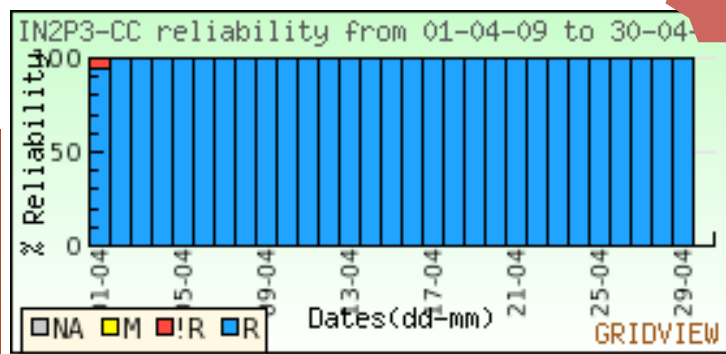


LHCb

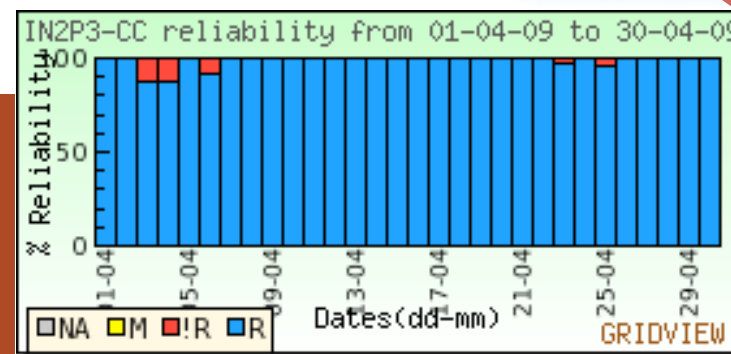
▶ Fiabilité vue par les VOs: Avril 2009



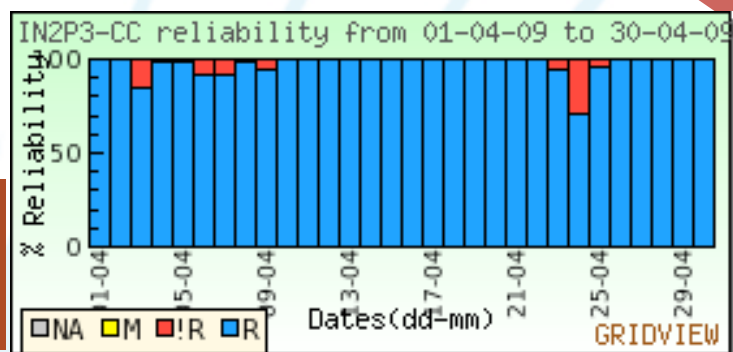
ALICE



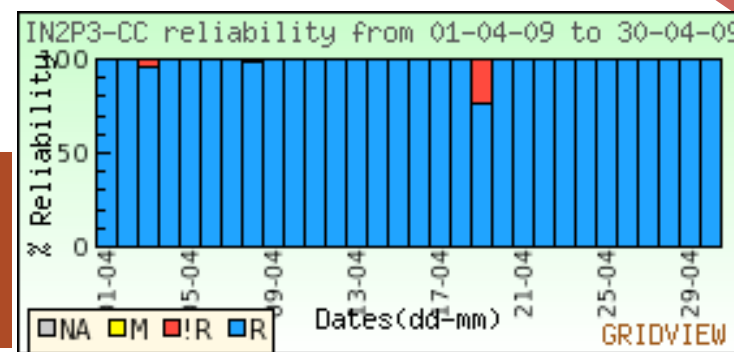
ATLAS



CMS



LHCb



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

▶ EGEE France – disponibilité & fiabilité



Avril 09

Region	Avail- ability	Reli- ability
France	97 %	98 %
SouthWesternEurope	95 %	97 %
UKI	94 %	96 %
NorthernEurope	94 %	96 %
SouthEasternEurope	93 %	94 %
GermanySwitzerland	93 %	93 %
CentralEurope	93 %	93 %
CERN	88 %	88 %
AsiaPacific	80 %	82 %
Italy	78 %	88 %
Russia	60 %	68 %

Source: <https://edms.cern.ch/document/963325>

EGEE France – disponibilité & fiabilité



Avril 09

Ces chiffres paraissent suspects pour certains sites (Phy.CPU = Log.CPU)

Region Site
France (France)

	Phy. CPU	Log. CPU	KSI2K	Avail- ability	Reli- ability	Availability History		
						Jan-09	Feb-09	Mar-09
AUVERGRID	42	42	75	98 %	98 %	98 %	100 %	98 %
CGG-LCG2	80	80	49	91 %	91 %	46 %	73 %	96 %
ESRF	16	16	43	86 %	86 %	83 %	89 %	98 %
GRIF	3,338	2,180	3,908	100 %	100 %	99 %	100 %	92 %
IBCP-GBIO	10	10	5	55 %	97 %	60 %	67 %	94 %
IN2P3-CC	1,074	4,296	3,832	99 %	99 %	97 %	98 %	90 %
IN2P3-CC-T2	1,074	4,296	3,832	99 %	99 %	96 %	97 %	89 %
IN2P3-CPPM	358	358	537	98 %	99 %	99 %	97 %	95 %
IN2P3-IPNL	452	440	656	98 %	99 %	98 %	96 %	98 %
IN2P3-IRES	664	628	1,526	84 %	84 %	95 %	96 %	93 %
IN2P3-LAPP	512	512	1,133	96 %	98 %	94 %	100 %	98 %
IN2P3-LPC	448	448	802	84 %	99 %	94 %	93 %	99 %
IN2P3-LPSC	120	112	43	71 %	96 %	97 %	99 %	94 %
IN2P3-SUBATECH	275	380	803	97 %	97 %	99 %	96 %	99 %
IPSL-IPGP-LCG2	34	34	41	96 %	96 %	94 %	100 %	96 %

Source: <https://edms.cern.ch/document/963325>

LHCb Analysis Tests

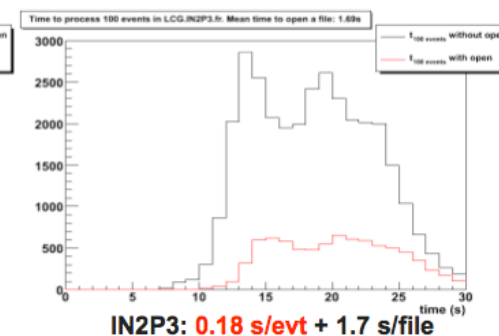
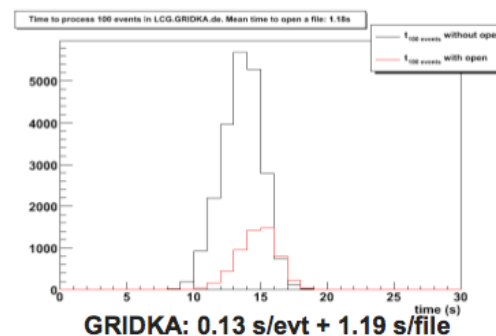
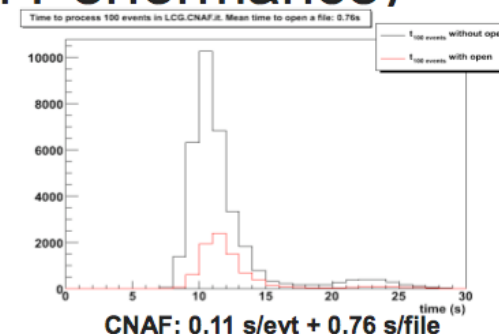


Running the Test

- Each submission consist of:
 - ~ 600 jobs, reading 100 different files with 500 events per file.
 - Average file size 200 MB
 - 1 Job = 100 Files = 50.000 events ~ 20 GB ~ 1-2 hours~ 2-3 MB/s
- InputData:
 - User submits list of LFNs
 - DIRAC checks LFC for replicas and check availability at the sites.
 - Job receives SURLs.
 - SURLs are converted to tURLs by asking to SRM (getturls)
- Application:
 - Root base application open the files in sequence, reads and analyzes all events directly from the SE (without copy to WN).

Results (Detail Performance)

- Distribution of Wall Clock time to process 100 consecutive events.
- Distribution of Wall Clock time to process 100 consecutive events including a new file opening.
- The difference of the means represents the file opening time.



Source: Ricardo Graciani, [GDB, 13/05/2009](#)

► LHCb Analysis Tests (suite)

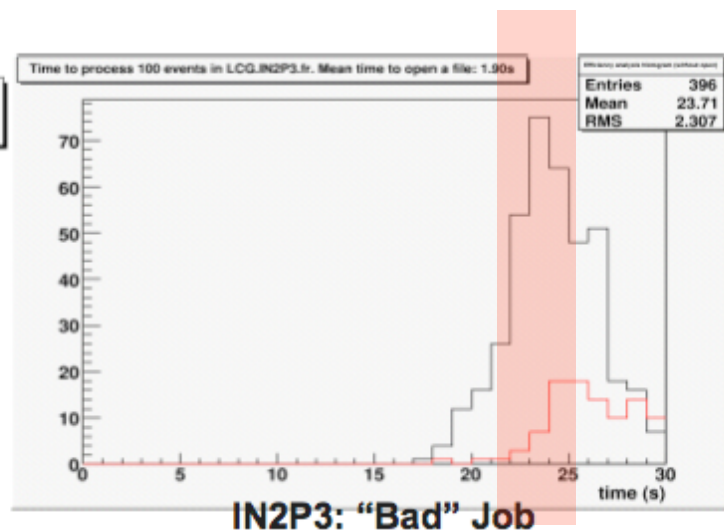
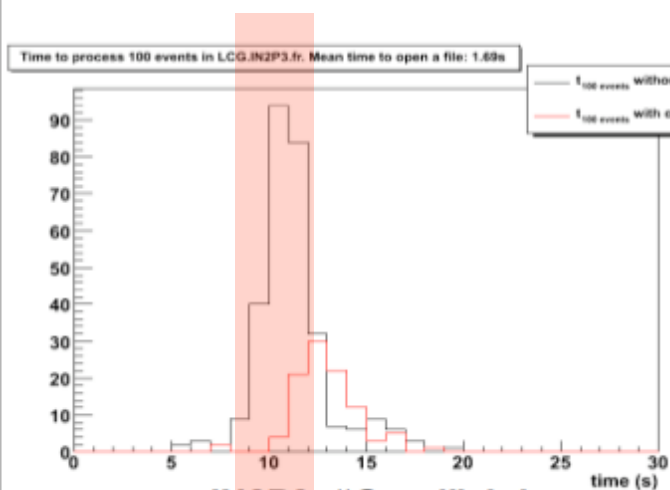


IN2P3 observations

- Poor CPU Efficiency at IN2P3 ($< 50\%$)
- Large variation in time to process events
- Seems to be Job to Job variations
- **Needs further investigations**

Comportements différents des jobs.
Pistes à explorer, d'après un expert dCache:

- dCache debug = 2 sur certains worker nodes
- Nombre maximum de fichiers ouverts par pool (MaxMovers?)



▶ SL5 – Portage du code des expériences

3 – Status and Plans

▶ ATLAS

- The goal is to have a full native build by next week (deployed on the GRID by ~August 2009)
- Will continue to produce binaries for SL4 and SL5 concurrently
- Inclined to retain the existing default (SLC4/gcc34/32-bit) as primary platform until after the 2009–2010 physics run

▶ CMS

- Finished native port 64 bit/SL5. Everything builds, basic tests seem okay, will proceed to full validation.
- Rather switch from SL4 to SL5 binaries in one go at their convenience

▶ LHCb

- Port not yet finalized (not yet a complete working release)
- Plan is to use native SL5/gcc43/64bit for the real data and the corresponding MC productions

▶ ALICE

No problem. They would like to see the transition to 64 bits/SL5 to happen as soon as possible

pere.mato@cern.ch May 12, 09

7

Etat d'avancement de compilation native sur SL5 avec compilateur gcc-4.3 (pas d'utilisation du compilateur par défaut sur SL5 qui est gcc-4.1)

Activation de SELinux a un impact sur certains librairies, y compris Oracle.

Passage en production n'est pas possible pour toutes les expériences avant septembre 2009.

Des WNs sous SL5 pour la validation restent nécessaires, avec un CE spécifique.

Source: Pere Mato, [GDB, 13/05/2009](#)



VMEM in 64bit Architecture

- We observe that any process running in 64bit has a VMEM footprint of ~50MB (c.f. 5MB for a 32bit process)

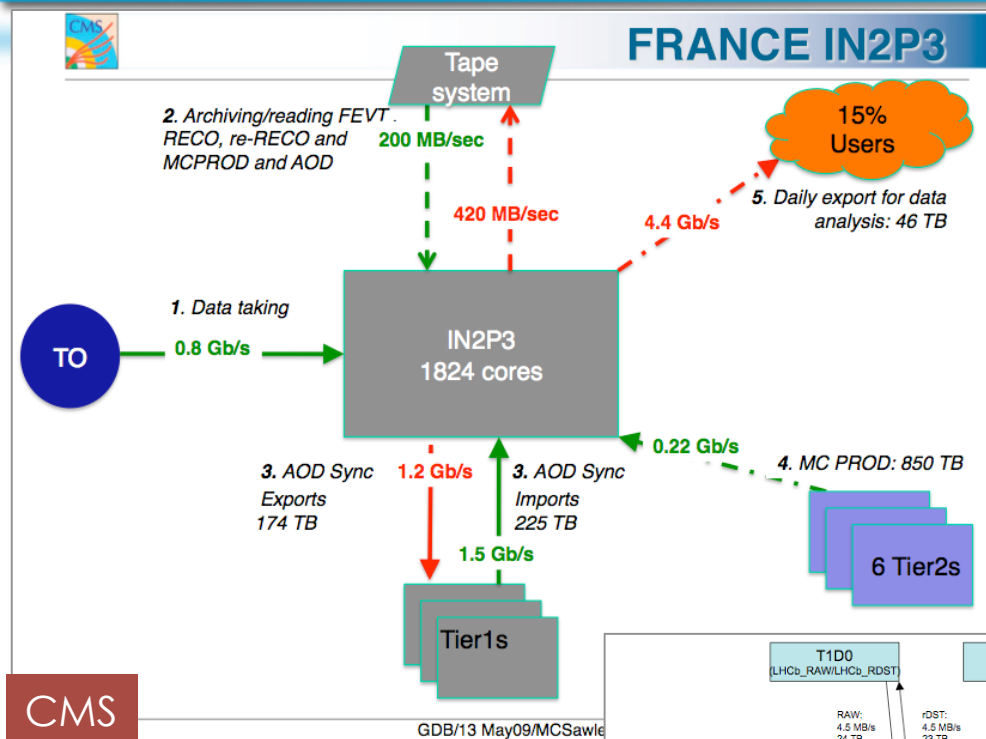
USER	PID	%CPU	%MEM	VSZ	RSS	TTY	STAT	START	TIME	COMMAND
gla012	1741	0.1	0.0	58892	2612	pts/13	S	12:49	0:00	python yawn.py
gla012	1745	0.0	0.0	5816	2176	pts/13	S	12:49	0:00	python32 yawn.py

- This means that running 'hello world' on the grid on a 64bit machine has a VMEM footprint of ~450MB (c.f. 32bit footprint of ~60MB)
 - In both cases the RSS is ~22MB
- Which means that killing grid jobs based on VMEM consumption is probably not a good idea...
 - N.B. also killing jobs based on RSS doesn't work as the kernel actively tries to keep pages in memory
 - And it seems that, e.g., torque kills based on the memory consumption of the process tree, not of the payload

Python 64bits
VSZ = 58 MB

Python 32bits
VSZ = 5 MB

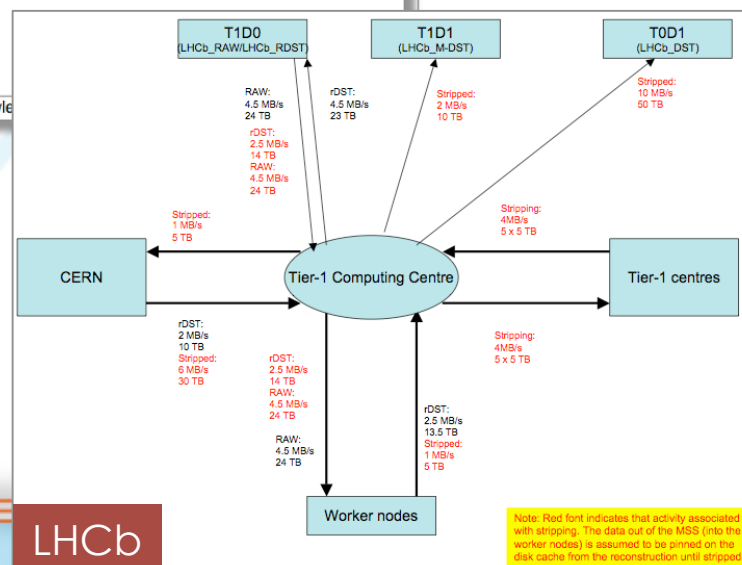
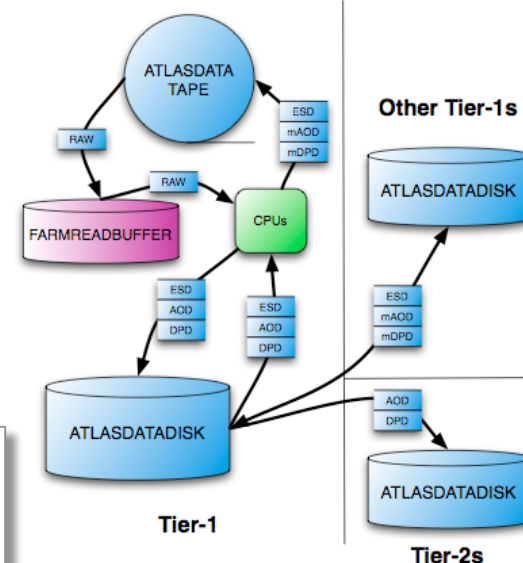
Workflows & débits I/O



ATLAS

TI Reprocessing Workflow

- Here mAOD/mDPD means merged files



Beaucoup d'information dans ces présentations.

A étudier en détail et fournir un retour.

▶ Tests glxexec/SCAS



Status of the pilot

■ Sites

- FZK: Running well
- Lancaster:
 - SCAS plug-in for 64 bit not building (ETICS issue). Will use pool account plug-in for now. Should be ready this week.
- IN2P3: running well

■ Experiments

- LHCb have started testing and are planning to increase this.
- ATLAS: Work on the framework is almost done. Expect to be sending jobs to prod in a few days.

■ Issues:

- Currently there is an incompatibility with CREAM (fix almost ready)



WLCG Grid Deployment Board, CERN 13 May 2009



Cette affirmation ne correspond pas du tout aux résultats des tests:

- Impossibilité d'installer glxexec partagé
- glxexec détruit l'environnement du job lancé par le pilote

Détails Pierre Girard

► XROOT comme protocole d'accès aux données



CERN IT Department
CH-1211 Genève 23
Switzerland
Dirk.Duellmann@cern.ch

XROOT Motivation & Schedules

CERN IT
Department

- Several different goals
- XROOT I/O server for CASTOR
 - Motivation:
 - more scalable I/O server for analysis
 - using XROOT server framework to implement a de-coupled CASTOR disk pool management (w/o LSF scheduling)
 - Time-scale: now (2.1.8)
 - aim at LSF free disk pools with next CASTOR release 2.1.9
- XROOT as potential common protocol
 - Motivation:
 - fewer file access protocols to support by users / providers
 - no intrinsic performance enhancements expected
 - Time-scale: unlikely that consolidation could finish before 2009/10 run
 - but planning / evaluation should start

2



Tuesday, 12 May 2009

► Chantiers récents et en cours



- Migration vers HPSS v6.2
 - Interruption de service: Juin 1-4
- Pas de migration vers dCache/Chimera avant l'été
- Plusieurs tests et évolutions de la chaîne de soumission et de gestion des jobs
 - Voir présentation de Pierre aujourd'hui
- Problème d'installation et de réplication du logiciel ATLAS sur AFS
 - Réunion programmée le 29/05/2009

► Chantiers récents et en cours (suite)



- Ménage dans les espaces de stockage (dCache et HPSS) des expériences LHC
- dCache
 - Poursuite de l'introduction de nouveaux serveurs de disques Thors et campagne de sortie de production progressive des premiers Thumpers (acquis en 2006)
- VO boxes
 - Toutes mise à jour et en production, à l'exception de une VO box ATLAS

▶ RAPPEL: axes de travail prioritaires



- Tier-1
 - Améliorer l'efficacité de la copie bande → disque, via l'ordonnancement des requêtes passées à HPSS par dCache
- Séparation claire des données tier-1 et tier-2
 - Portes d'accès différentes pour chacun des sites, sans pour autant dupliquer les données
- Déploiement d'un prototype de ferme d'analyse interactif basée sur PROOF
- Finaliser le mécanisme d'installation du logiciel des expériences dans la zone AFS
 - Utilisation de la réplication de volumes AFS pour l'équilibrage de charge
 - Documentation et généralisation pour les 4 expériences LHC
- Monitoring !!!

► Événements à venir



- Réunion semestrielle des sites LCG-France
 - Annecy, 18-19 mai 2009
 - <http://indico.in2p3.fr/conferenceDisplay.py?confId=1660>
- Hepix Spring 2009
 - Umeå (Suède), 25-29 mai 2009
 - <http://www.hpc2n.umu.se/events/workshops/09/hepix/>
- Workshop on adapting applications and computing services to multi-core and virtualization
 - CERN, Juin 24-26
 - <http://indico.cern.ch/conferenceDisplay.py?confId=56353>

▶ Aujourd'hui et à venir



- Aujourd'hui
 - Evolutions chaîne de gestion de jobs
 - Etat d'avancement de l'interface CREAM pour BQS
 - SYMOD: plate-forme générique de monitoring
- Prochaine réunion
 - Jeudi 11 juin, 13h30-15h30, salle 202

► Prochaines réunions en 2009



- Juin 11
- Juillet 9
- Septembre 10
- Octobre 15
- Novembre 12
- Décembre 10
- Toutes les réunions
 - <http://indico.in2p3.fr/categoryDisplay.py?categId=102>

► Questions/Commentaires

