WLCG Service Report

Jean-Philippe.Baud@cern.ch

~~~

WLCG Management Board, 27th October 2009
~~~

Introduction

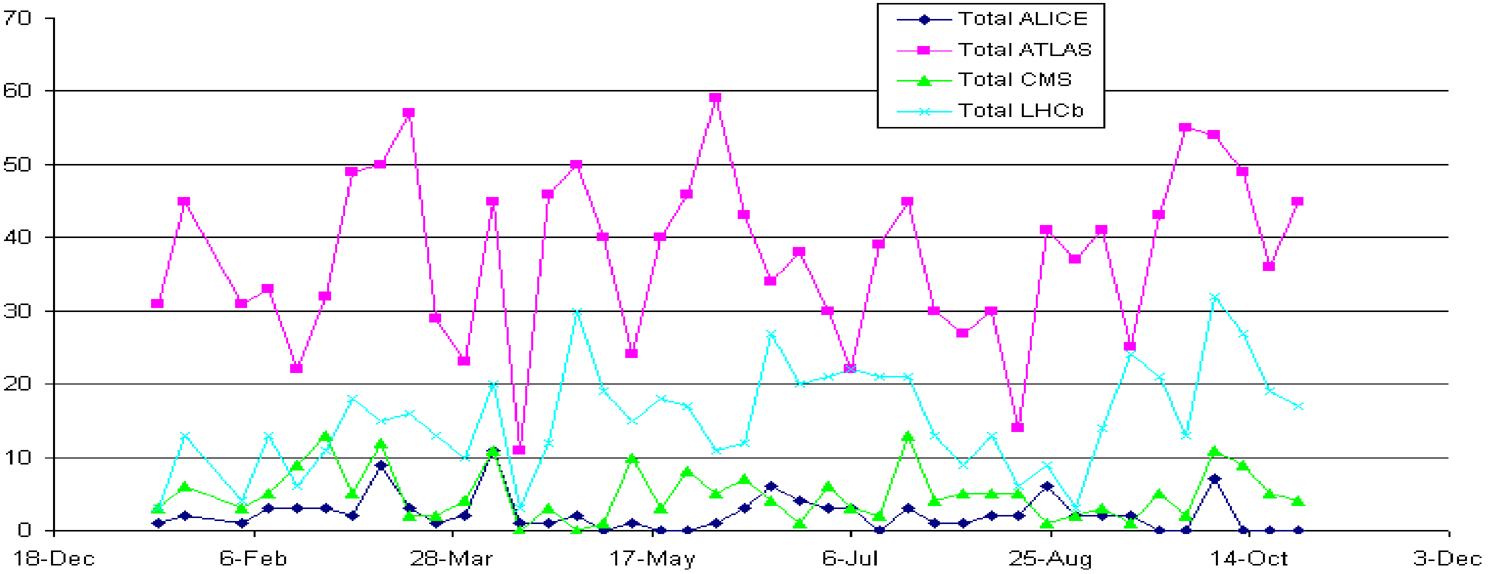
- Covers the weeks 12th to 25th October.
- Mixture of problems
- Incidents leading to (eventual) service incident reports
  - RAL disk subsystem failures taking down FTS, LFC and CASTOR from 4th to 9th and eventually led to the loss of 10 days data (200000 files) at RAL.
  - ASGC ATLAS conditions database still not synchronized.
  - ASGC CASTOR DB corrupted (21th October, not recovered yet)

Meeting Attendance Summary

Site	M	T	W	T	F
CERN	Y	Y	Y	Y	Y
ASGC	Y	Y	Y	Y	Y
BNL	Y	Y	Y	Y	Y
CNAF				Y	
FNAL					
FZK	Y	Y	Y	Y	
IN2P3		Y			
NDGF					
NL-T1	Y	Y	Y	Y	
PIC	Y	Y			Y
RAL	Y	Y	Y	Y	Y
TRIUMF					

GGUS summary (2 weeks)

VO	User	Team	Alarm	Total
ALICE	0	0	0	0
ATLAS	13	68	0	81
CMS	8	0	1	9
LHCb	1	33	2	36
Totals	22	101	3	126

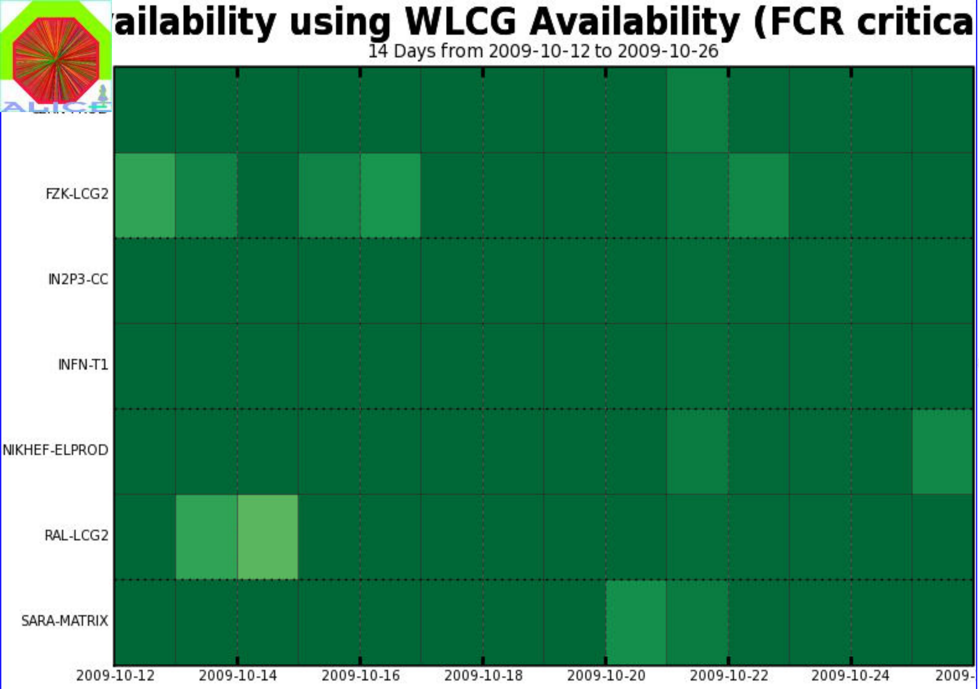


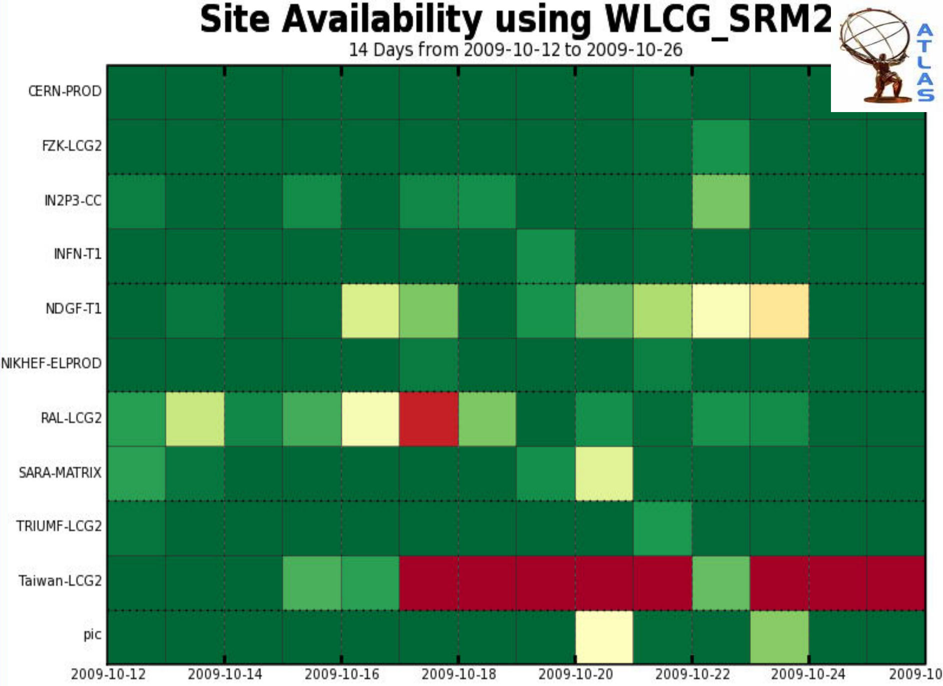
Alarm tickets

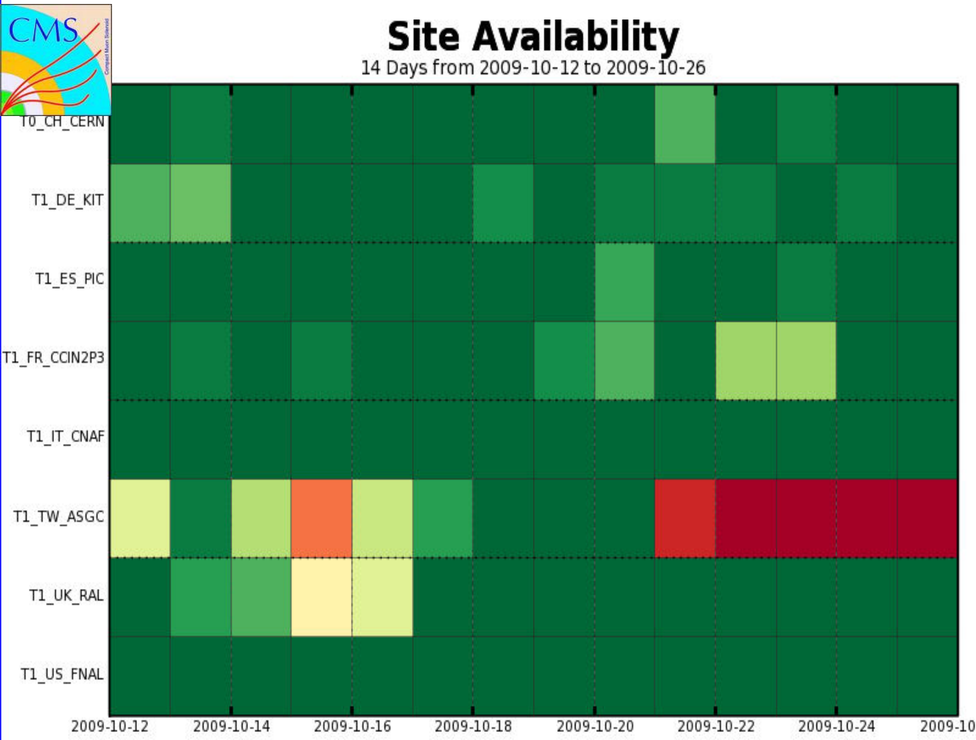
- There were 3 alarm tickets in the week starting 12th October
  - CERN CASTOR stager stuck reported by LHCb
  - CERN CASTOR Name Server problem reported by CMS
  - CERN CASTOR Name Server problem reported by LHCb

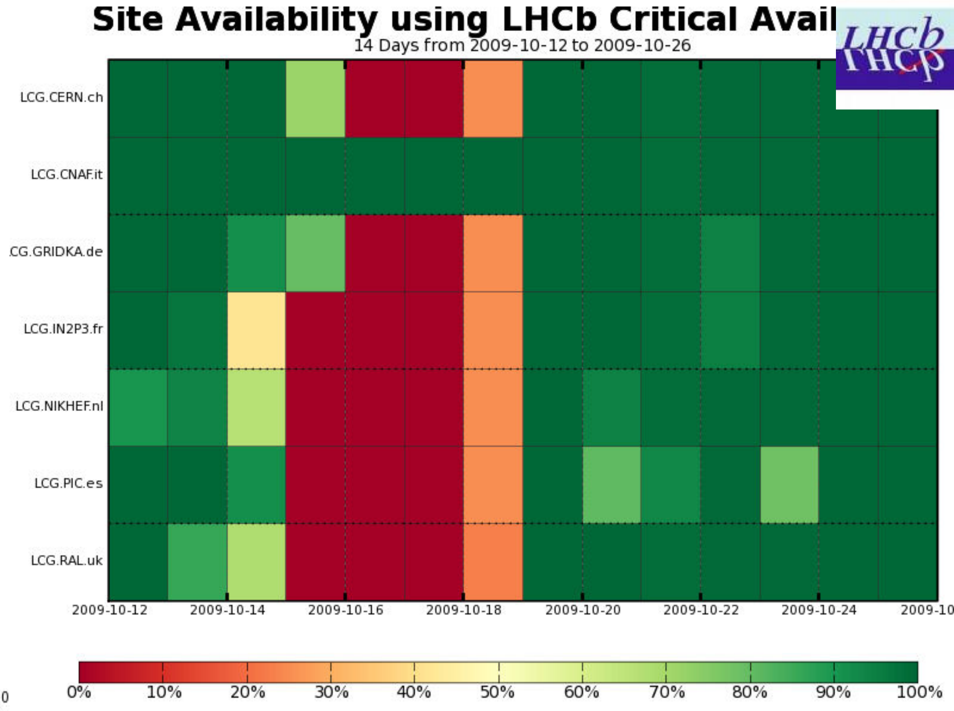
GGUS tickets and OSG

- The OIM view provided to GGUS should list only the ‘resource group’ name (BNL_ATLAS) with valid contact-email and emergency-email addresses


ailability using WLCG Availability (FCR critica
14 Days from 2009-10-12 to 2009-10-26
FZK-LCG2
IN2P3-CC
INFN-T1
NIKHEF-ELPROD
RAL-LCG2
SARA-MATRIX
2009-10-12
2009-10-14
2009-10-16
2009-10-18
2009-10-20
2009-10-22
2009-10-24
2009-


Site Availability using WLCG_SRM2
14 Days from 2009-10-12 to 2009-10-26
CERN-PROD
FZK-LCG2
IN2P3-CC
INFN-T1
NDGF-T1
NIKHEF-ELPROD
RAL-LCG2
SARA-MATRIX
TRIUMF-LCG2
Taiwan-LCG2
pic
2009-10-12
2009-10-14
2009-10-16
2009-10-18
2009-10-20
2009-10-22
2009-10-24
2009-10
ATLAS


Site Availability
14 Days from 2009-10-12 to 2009-10-26
T0_CH_CERN
T1_DE_KIT
T1_ES_PIC
T1_FR_CCIN2P3
T1_IT_CNAF
T1_TW_ASGC
T1_UK_RAL
T1_US_FNAL
2009-10-12
2009-10-14
2009-10-16
2009-10-18
2009-10-20
2009-10-22
2009-10-24
2009-10
CMS


Site Availability using LHCb Critical Avail
14 Days from 2009-10-12 to 2009-10-26
LCG.CERN.ch
LCG.CNAF.it
CG.GRIDKA.de
LCG.IN2P3.fr
LCG.NIKHEF.nl
LCG.PIC.es
LCG.RAL.uk
2009-10-12
2009-10-14
2009-10-16
2009-10-18
2009-10-20
2009-10-22
2009-10-24
2009-10
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
LHCb

RAL Disk failures 1/2

ATLAS, CMS and LHCb SAM tests saw the RAL LFC, FTS and CASTOR downtimes (4 to 7 October for LFC and FTS and up to 9 October for CASTOR) due to failing disk sub-systems. ALICE only test their VOboxes and saw an SL4 to SL5 migration interrupt.

RAL CASTOR runs on a SAN with disk systems containing primary and mirrored databases. Hardware faults on mirror since 10 September also hit primary on 4 October and CASTOR went down.

Decision was to revert to older hardware then revalidate the failing systems. Suspicion early on was temperature problems.

8 October CASTOR being restored without loss for ALICE and CMS and losing a few hours transactions for ATLAS and LHCb – estimated at 10000 files. List of lost files being prepared for experiment decision.

9 October CASTOR restored – experiments to recover lost files or to clean catalogues. Vendor working with RAL to understand root cause of failures.

14 October Discovered problem with database used following the restore. Resulted in loss of around last ten days data added to Castor.

The database restore had been OK. The problem arose when Oracle opened the database and picked up the 'wrong' disk array.

21 October List of lost files (200000 for Atlas) produced and LFC cleanup started.

Actually only one dataset did not have another copy available at another site.

RAL Disk failures 2/2

SIRs:

Hardware failures and loss of service.

http://www.gridpp.ac.uk/wiki/RAL_Tier1_Incident_20091004

Loss of data following restoration of services:

http://www.gridpp.ac.uk/wiki/RAL_Tier1_Incident_20091009

ASGC DB Problems

- 2 major DB problems
  - Atlas Condition DB:
    - has not been available for more than 4 weeks now
    - CERN DM group recommends to perform a complete re-instantiation using transportable tablespaces. BNL will be the source.
    - Synchronization should happen tomorrow 28th October (09:00 CET).
  - CASTOR DB:
    - Has not been available for almost a week
    - All recovery attempts failed
    - Should the DB be reset?
    - A phone conference will take place tomorrow 28th October (09:15 CET)

Miscellaneous Reports

- CASTOR Name Server problem at CERN due to new CASTOR release (2.1.9).
- 180 files lost at NLT1 (tape destroyed).
- Problem at CNAF installing CMS SW release (not understood).
- Instability of Atlas Condition Database at BNL due to high load from Tier2s; solved by increasing memory.
- Problems with new SRM release at CERN (needed a rollback to the previous version).
- Problems with new BDII release at CERN (needed a rollback to the previous version).
- SRM problem at BNL last Friday due to a Java exception.

Summary/Conclusions

- Very long standing problems at ASGC (CASTOR and Condition Database).
- Serious disk hardware failures at RAL: 200000 files lost.
- A number of sites, including ASGC, have been unable to recover production databases from backups / recovery areas with major downtimes occurring as a result. A coordinated DB recovery validation exercise will take place the 26th November: RAL and ASGC are especially encouraged to participate.