WLCG Service Report

Olof.Barring@cern.ch

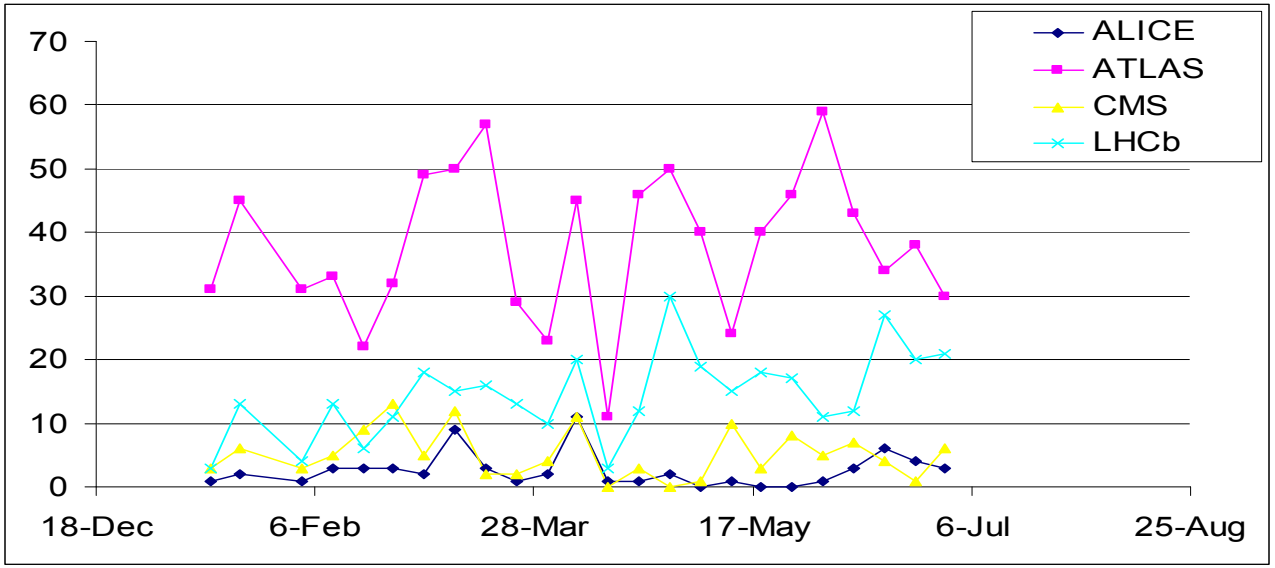
~~~

WLCG Management Board, 30th June 2009
~~~

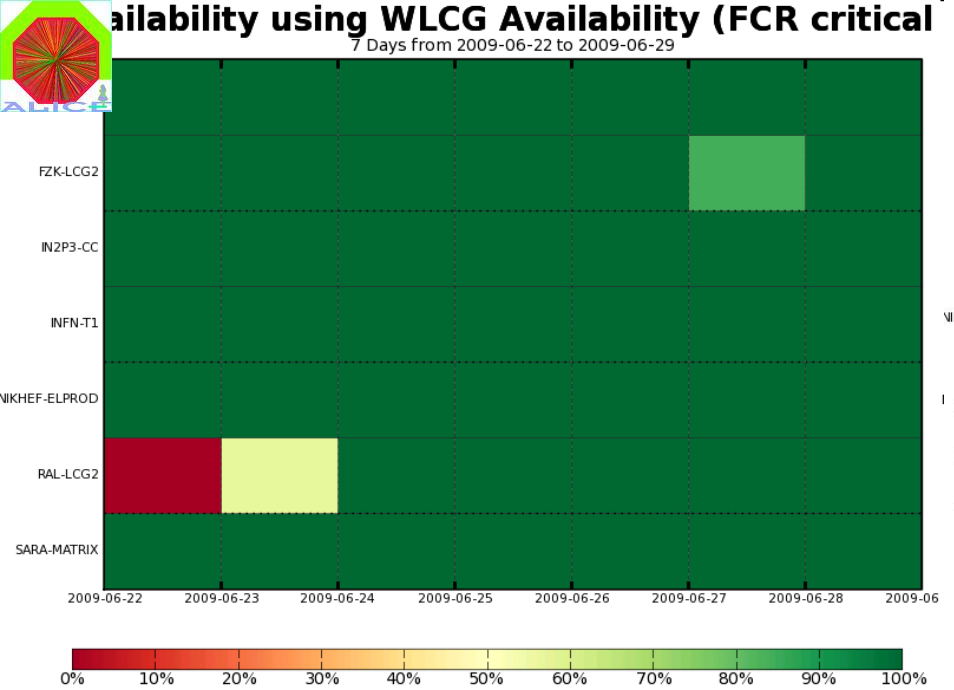
Introduction

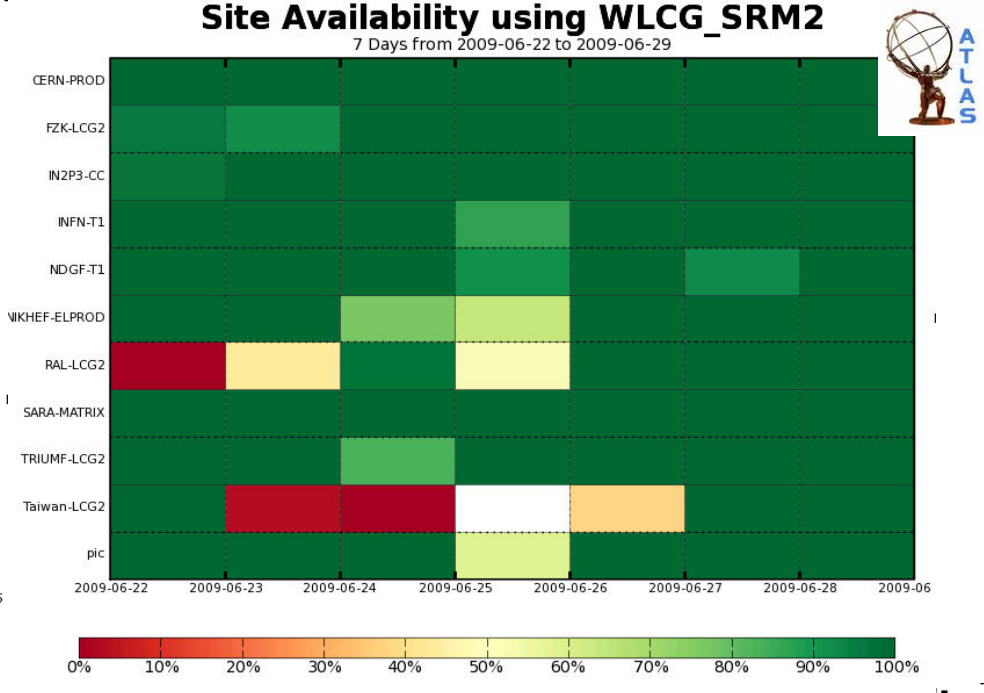
- Quiet week
- No alarm tickets
- Instabilities in VO specific frameworks
  - ATLAS post-mortem: Degraded PanDA service, impact on other offline DB services on ATLR
  - LHCb reported some issues with DIRAC
- RAL scheduled downtime for move to new Data Centre
- ASGC not stable with several unscheduled downs
- Several CE problems @CERN

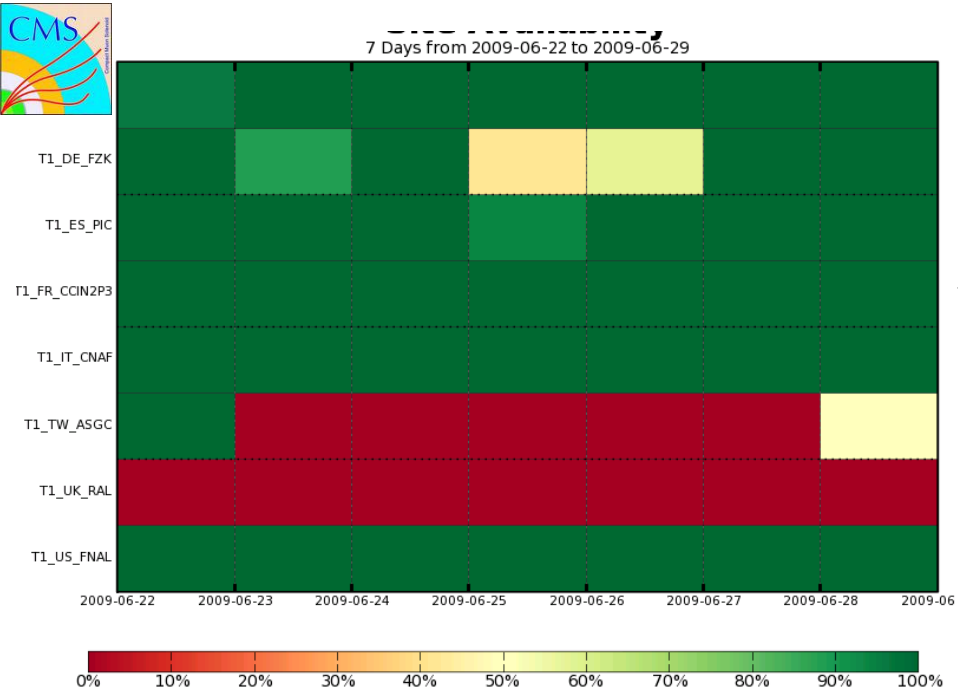
GGUS summary

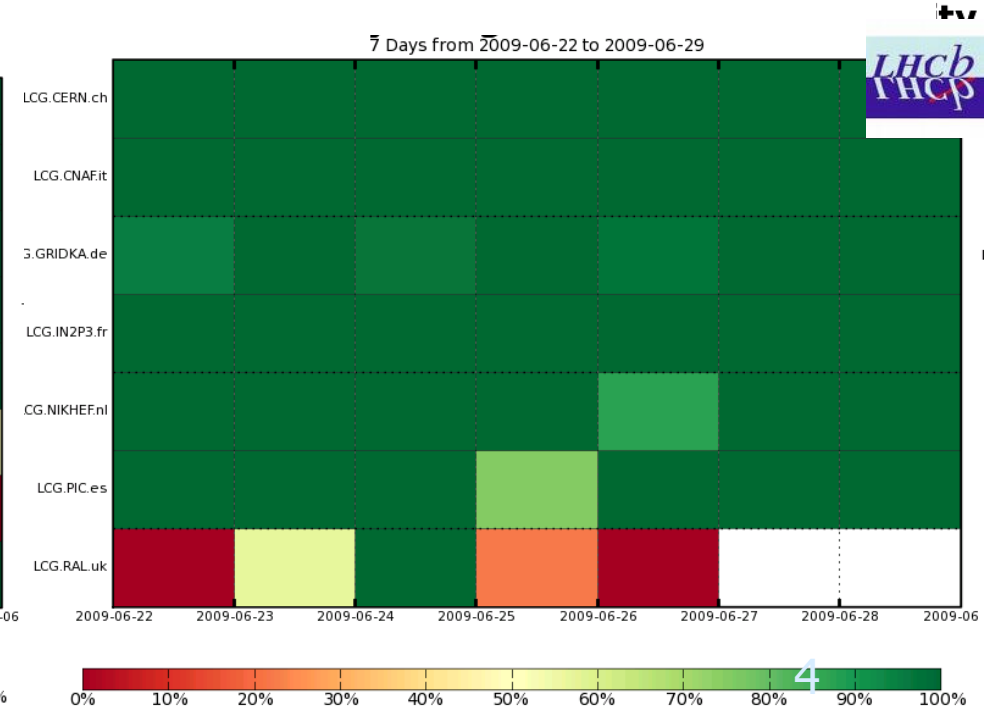
VO	User	Team	Alarm	Total
ALICE	3	0	0	3
ATLAS	5	25	0	30
CMS	6	0	0	6
LHCb	1	20	0	21
Totals	15	45	0	60




ailability using WLCG Availability (FCR critical
7 Days from 2009-06-22 to 2009-06-29
FZK-LCG2
IN2P3-CC
INFN-T1
NIKHEF-ELPROD
RAL-LCG2
SARA-MATRIX
2009-06-22
2009-06-23
2009-06-24
2009-06-25
2009-06-26
2009-06-27
2009-06-28
2009-06
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%


Site Availability using WLCG_SRM2
7 Days from 2009-06-22 to 2009-06-29
CERN-PROD
FZK-LCG2
IN2P3-CC
INFN-T1
NDGF-T1
NIKHEF-ELPROD
RAL-LCG2
SARA-MATRIX
TRIUMF-LCG2
Taiwan-LCG2
pic
2009-06-22
2009-06-23
2009-06-24
2009-06-25
2009-06-26
2009-06-27
2009-06-28
2009-06
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%


7 Days from 2009-06-22 to 2009-06-29
T1_DE_FZK
T1_ES_PIC
T1_FR_CCIN2P3
T1_IT_CNAF
T1_TW_ASGC
T1_UK_RAL
T1_US_FNAL
2009-06-22
2009-06-23
2009-06-24
2009-06-25
2009-06-26
2009-06-27
2009-06-28
2009-06
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%


7 Days from 2009-06-22 to 2009-06-29
LCG.CERN.ch
LCG.CNAF.it
G.GRIDKA.de
LCG.IN2P3.fr
CG.NIKHEF.nl
LCG.PIC.es
LCG.RAL.uk
2009-06-22
2009-06-23
2009-06-24
2009-06-25
2009-06-26
2009-06-27
2009-06-28
2009-06
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

Degraded PanDA service, impact on other offline DB services on ATLR (1/3)

Incident Post Mortem for ATLAS PanDA Monitor, June 23

Overview

- ATLAS PanDA monitor opens many connections to the ATLR Oracle database where PanDA system state is maintained.
- At the moment these connections are not pooled, but development of pooling code is being undertaken to improve response times and reduce load on the database. To take proper advantage of this code the http daemon running the service needs to be run in multi-threaded mode (a.k.a. MPM-worker).
- The panda monitor service runs on 3 machines, voatlas19,20,21.

Degraded PanDA service, impact on other offline DB services on ATLR (2/3)

Incident Timeline

Prequel

- Connection pooling code had been developed and tested on a pre-production monitor and was believed to be robust.

June 22

- **1500 CEST** The intent to upgrade to the pooled connection service was announced to developers with instructions not to update code. Locks were placed on key files in SVN.
- **1530** Connection pooling code for the panda monitor was deployed on voatlas21. The httpd service was switched from MPM-prefork to MPM-worker to support this change.
- **1600** A problem with a different part of the monitor code, the logger, which uses the connection pool, was identified. This caused DB connections to not be released when clients tried to use the http logging service.
  - The total number of connections used by voatlas21 is well beneath the connection limit for panda on ATLR, so only this machine was affected with no knock-on effects to the rest of panda or other ATLR hosted services at this point.
- **1630** The logger was disabled on voatlas21, allowing the monitor service to continue in a slightly degraded form.
- **1730** Initial attempts to debug the logger service were not able to identify the problem. As the degradation was minor, and the pooling code on voatlas21 otherwise worked well, an announcement was made to developers to not make any monitor code changes while investigations continued.
- **2330** To fix an unrelated problem a developer updated a piece of code on voatlas19. *This update put the pooling code in place on this machine but without enabling MPM-worker.* The configuration of MPM-prefork, combined with the logger bug, caused the consumption of large numbers of DB connections.
- **June 23**
- **0030** The same code change was applied to voatlas20, resulting in two monitor machines swallowing almost all DB connections for the panda service on ATLR.
  - Severe degradation of the panda service is reported by ATLAS shifters.
- **0116** One node in ATLR is rebooted by CERN IT.
- **0303** Another node in ATLR is rebooted by CERN IT.
- **0730** Initial attempts are made to rollback monitor machines. Hampered by lack of understanding about code updates on voatlas19,20.
- **0930** Full rollback done, restoration of service.

Degraded PanDA service, impact on other offline DB services on ATLR (3/3)

Conclusions

- A review of panda monitor code development and deployment is being undertaken to ensure no repeat of the above incident occurs. Expansion of the panda monitor test suite to encompass use of the logger service is being done.

DIRAC issues reported by LHCb

- All MC physics production has been temporary halted because of inconsistencies in DIRAC bookkeeping
  - Reported fixed on Friday
- Issue with DIRAC task queue scheduling.
  - Causing some destructive interference between the 1 billion MC production and other (user) jobs in the system
  - Still an issue?

RAL scheduled downtime for DC move

- An interesting example for how to cope with long extended downs of a Tier-1
  - ATLAS elected Glasgow as stand-in UK Tier-1 taking over the main transfer and data distribution role while RAL was being down.
  - Limits the impact on the Cosmic data taking to ~24 hours down of the LFC @ RAL
- Despite presumably hectic activity with equipment movements, RAL continued to attend the daily conf call ☺
  - Move seems to progress according to the original schedule
  - CASTOR and Batch down until next Monday (6/7).
  - Planning and detailed progress reported at http://www.gridpp.rl.ac.uk/blog/2009/06/23/r89-migration-tuesday-23rd-june/

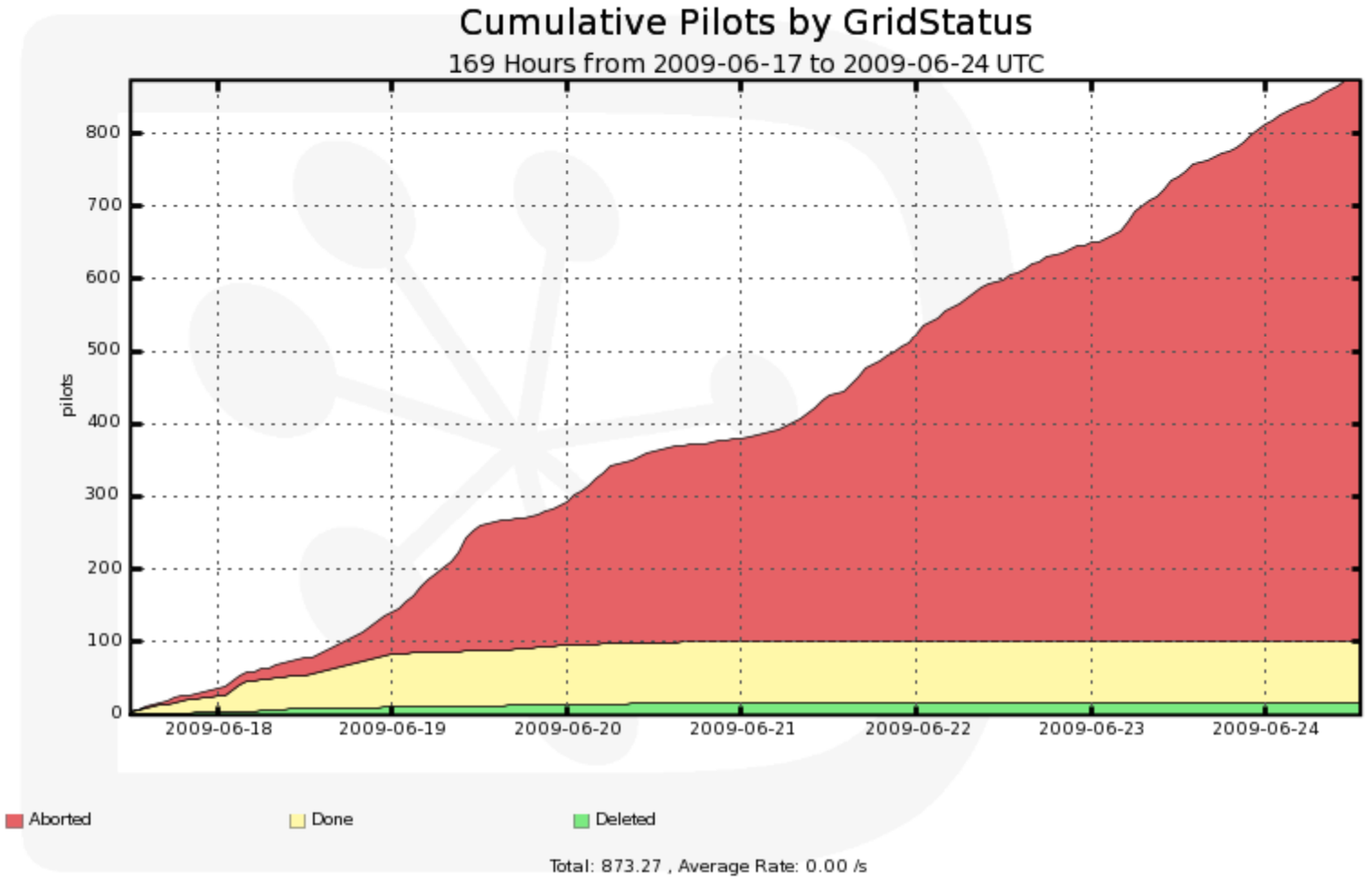
ASGC instabilities

- Site services are not yet stable and fully useable from VO's perspective
- Split ASGC from ATLAS streams replication because they were down and we were running out of space
  - Resumed within recovery window
- ATLAS reports several unscheduled or extended downtimes → ASGC was taken out from data distributions at several occasions
- CMS decided to give this week as grace period for ASGC to become fully functional
  - No new tickets and opened tickets put on hold
  - Resume on Monday 6/7

CE instabilities @ CERN

- Several problems with CREAM CE reported by ALICE
  - configuration problems and bugs found (https://savannah.cern.ch/bugs/?48144, https://savannah.cern.ch/bugs/?52392)
  - … service back but underlying cause(s) not yet fully understood
- Also a service problem with some of the production LCG CEs
  - Hardware problems on ce124. It had been put in scheduled down in GOCDB but unfortunately the service had not been stopped and this caused 'blackhole' for LHCb
  - A known bug (globus_gma process pile-up, https://savannah.cern.ch/bugs/?48588) affected two of the production CEs when coming back after the power-cut the week before

CERN ce124 causing blackhole effects for LHCb pilot submission



Summary

- No new serious site issues but ASGC still suffers from instabilities
- RAL long downtime for DC move is progressing to plans. ATLAS copes by diverting Cosmic data taking to Glasgow
- Issues with VO frameworks
  - ATLAS: degraded PanDA service, impact on other offline DB services on ATLR
  - LHCb: issues with DIRAC
- Bad week for the CE services at CERN