WLCG Service Report

Olof.Barring@cern.ch

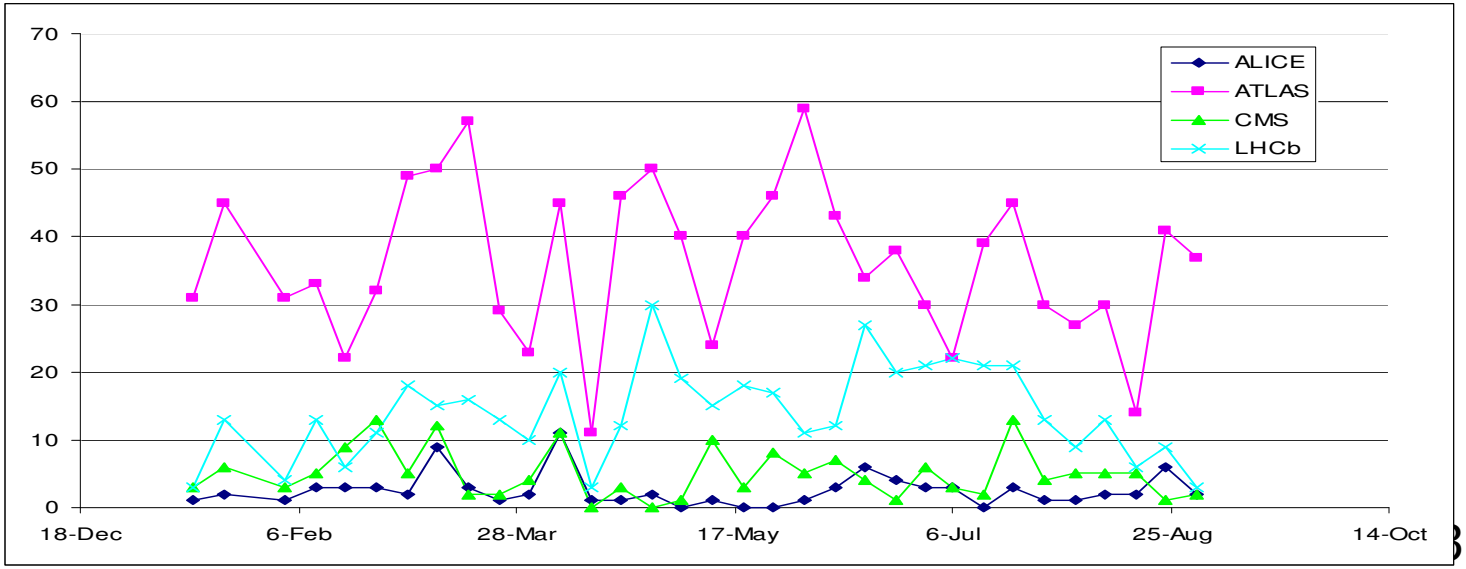
~~~

WLCG Management Board, 1st September 2009
~~~

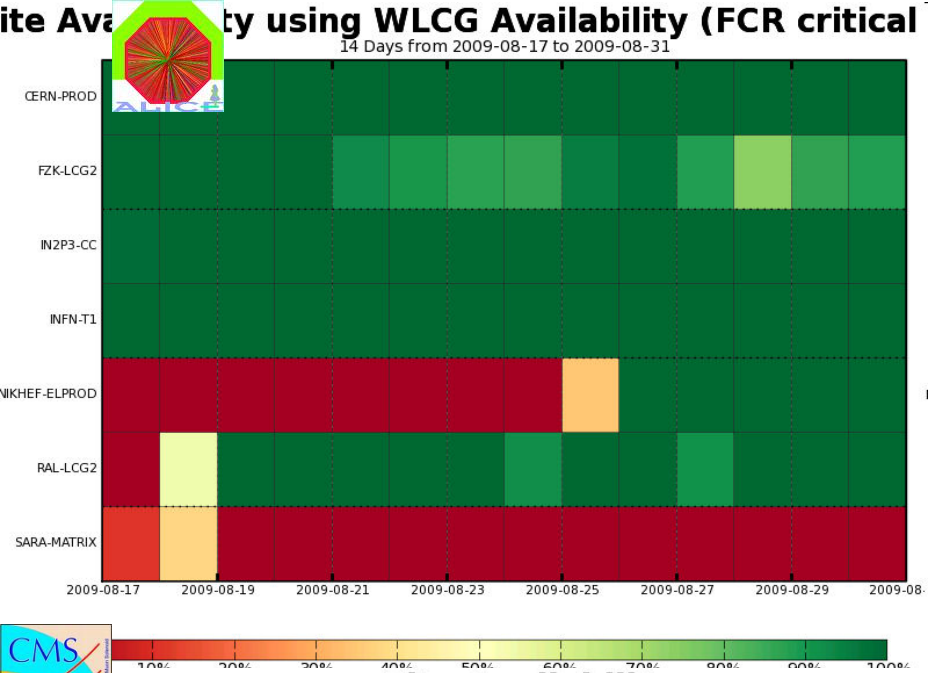
Introduction

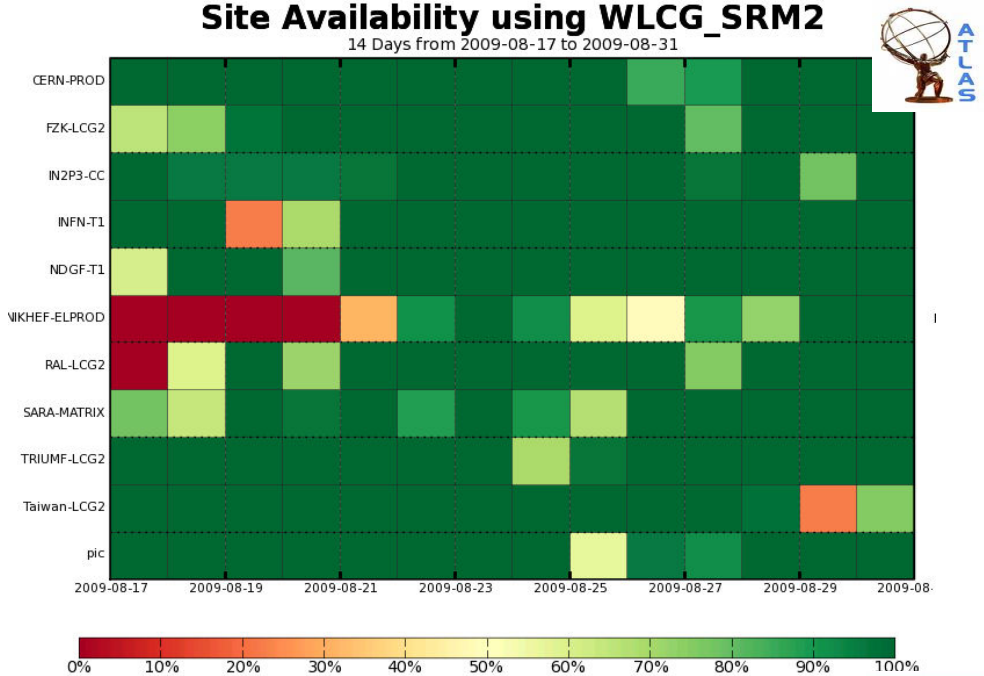
- Covers the two weeks 17th to 29th August
- Attendance
  - First week affected by holiday period
  - Significant ramp up with high attendance in second week
- Main events:
  - BNL site name change: BNL-LCG2 → BNL-ATLAS
  - Kernel patch for published local exploits
- Two alarm tickets:
  - 51206 ATLAS → BNL: TEST ALARM Ticket after name change
  - 51172 CMS → CERN: All dedicated CMS LSF queues at CERN disabled with no clear pre-warning
- Incidents leading to service incident reports
  - 26/8/2009 CERN: Closing of public and production queues for Emergency batch reboot to apply kernel upgrade

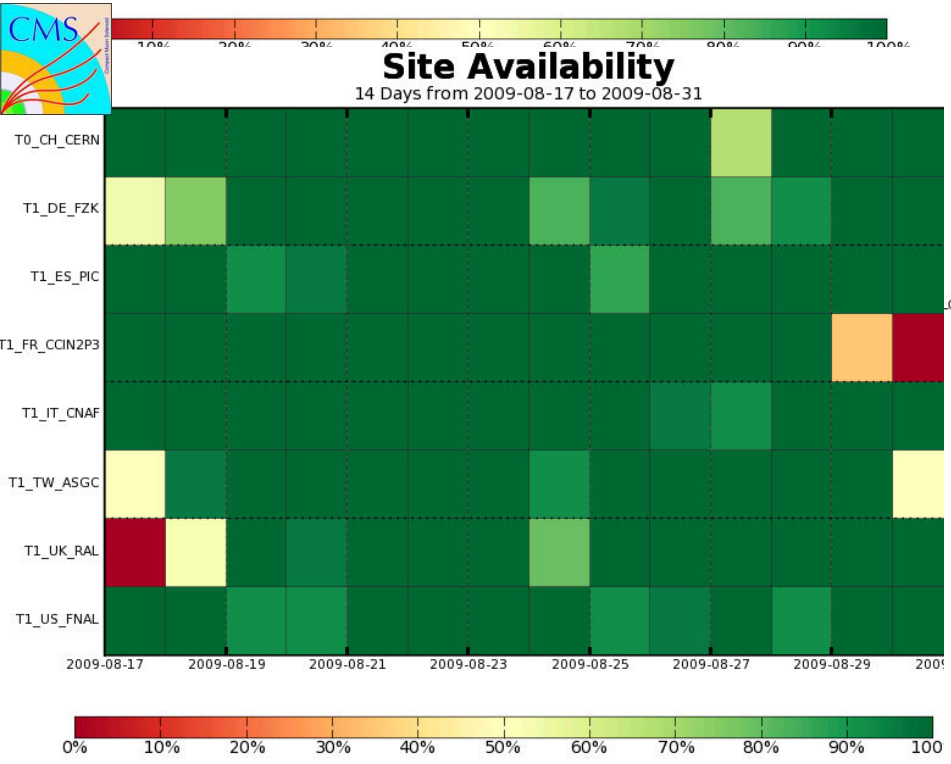
GGUS summary (2 weeks)

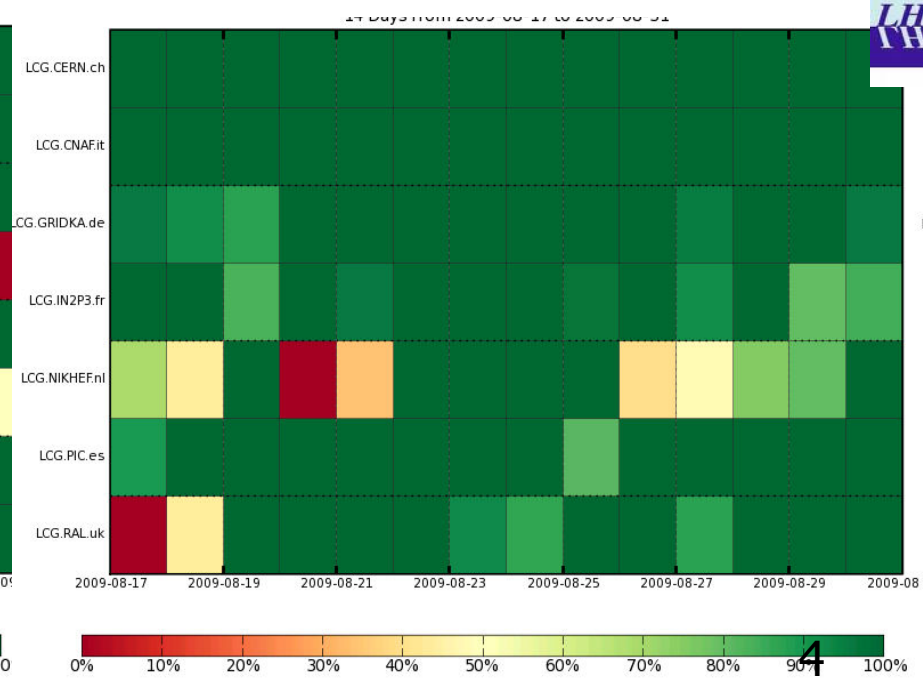
VO	User	Team	Alarm	Total
ALICE	8	0	0	8
ATLAS	21	56	1	78
CMS	2	0	1	3
LHCb	0	12	0	12
Totals	31	68	2	101




ite Ava ty using WLCG Availability (FCR critical
14 Days from 2009-08-17 to 2009-08-31
CERN-PROD
FZK-LCG2
IN2P3-CC
INFN-T1
NIKHEF-ELPROD
RAL-LCG2
SARA-MATRIX
2009-08-17
2009-08-19
2009-08-21
2009-08-23
2009-08-25
2009-08-27
2009-08-29


Site Availability using WLCG_SRM2
14 Days from 2009-08-17 to 2009-08-31
CERN-PROD
FZK-LCG2
IN2P3-CC
INFN-T1
NDGF-T1
NIKHEF-ELPROD
RAL-LCG2
SARA-MATRIX
TRIUMF-LCG2
Taiwan-LCG2
pic
2009-08-17
2009-08-19
2009-08-21
2009-08-23
2009-08-25
2009-08-27
2009-08-29
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%


Site Availability
14 Days from 2009-08-17 to 2009-08-31
T0_CH_CERN
T1_DE_FZK
T1_ES_PIC
T1_FR_CCIN2P3
T1_IT_CNAF
T1_TW_ASGC
T1_UK_RAL
T1_US_FNAL
2009-08-17
2009-08-19
2009-08-21
2009-08-23
2009-08-25
2009-08-27
2009-08-29
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100


LCG.CERN.ch
LCG.CNAF.it
LCG.GRIDKA.de
LCG.IN2P3.fr
LCG.NIKHEF.nl
LCG.PIC.es
LCG.RAL.uk
2009-08-17
2009-08-19
2009-08-21
2009-08-23
2009-08-25
2009-08-27
2009-08-29
0%
10%
20%
30%
40%
50%
60%
70%
80%
100%

Experiment Availabilities Reports

NIKHEF completed move ~24/8

SARA: the ALICE ongoing downtime is due to a failing ALICE VObox there. GGUS ticket 51238 submitted on the 31st .

RAL was back in full prod on Tuesday 18th after the air conditioning stoppages due to water chiller failures that had started on 12th (see Harry's report at the last MB)

Incidents: closing batch queues at CERN

https://twiki.cern.ch/twiki/bin/view/FIOgroup/PostMortem20090826

On the advice of the CERN security team [illegible] to two severe kernel vulnerabilities (one with no [illegible]ces had to be reinstalled with a new kernel and re[illegible]ally drained first.

Wide announcement: 'CERN is vulnerable'

Communication race condition where the subtle change 'public' → 'all' wasn't noticed by MoD

Wednsday

- 16.00 - Long running public batch queues set Inact[illegible]
- 16.50 - Mail sent to info-experiments describing thi[illegible] day.
- 18.40 - All batch queues set Inactive.
- 18.49 - Mail sent for SB to Computer Operations de[illegible]
- 20.11 - MOD sees original info-expeiments mail and moves the 18.49 update [saying it is all batch queus] to the usual place on SSB, but links from it the original mail sent at 16.50 saying it is only the public batch queues
- 21.38 - CMS "T0 operations list" mail describing the problem of the Inactive T0 queue and complaining about the confused messages ("public" queus vs. "all" queues)
- 22.26 - Update to computer operations on SSB saying that public and grid queues are now open with limited capacity.
- 22.38 - GGUS alarm ticket from CMS arrives detailing the issue with Inactive CMS T0 queue
- 23.06 - Operator responds to alarm ticket, asking for clarification
- 23.30 - CMS T0 hosts rebooted and queues reopened.

Thursday

- 09.45 - Remaining batch jobs killed. Service rebooted.
- 12.00 - Reboots ongoing (30% already available). All queues reactivated.
- 16.00 - All machines rebooted and in production (except for a few outliers).

Closing batch queues at CERN cont…

- To drain or not?
  - Only ~650 jobs out of 17,000 were lost
  - But, closed queues == lost processing time
- Some VOs prefer us to keep queues open and not drain the nodes while rolling out the new kernel but rather kill the running jobs
  - Potentially a large number of failed jobs
- The CERN batch configuration is highly customized with >70 dedicated queues
  - For a simplified configuration with only a (few) large resources it would have been possible to handled the rollout almost transparently without closing the queue
  - During the emergency rollout there were no time for negotiating how to handle each dedicated queues
- We are considering different options for how to better handle similar situations in the future
  - Will make a proposal to the VOs later this month

BNL site name change

- Site name change in BDII on 25/8 at 15:00 CEST (09:00 EDS):
  - BNL-LCG2 → BNL-ATLAS
- A few minutes (~15') later ATLAS Tier-1 sites could start updating their FTS channel definitions following a procedure provided by Gavin
  - The update of the channel definitions had to be synchronized with the BDII updates.
    - RAL and TRUIMF have disabled daily BDII updates. **The update is still pending…**
  - Some sites were not actively confirming the update though successful.
- BNL planned the operation and executed it across the grid with the assistance of the Tier-0 FTS team
- **The operation was transparent for ATLAS and transfers did not suffer.**
- The site name was also done for GGUS and a TEST ALARM ticket was injected by ATLAS in order to confirm that the workflow was working.

Miscellaneous Reports

- ALICE VOBox issue at CERN
  - Problem had been reported from several sites where ALICE VOBoxes couldn't connect to the Alien DB at CERN. The port, 8084, had been closed recently and must now be re-opened.
- RAL lost one disk server when recovering from the cooling problems
  - 99k files out of 4M in MCDISK space token
- CMS tape migrations stalled at ASGC
  - Ongoing for two weeks. Local CASTOR staff are investigating
  - Help from CASTOR teams @ CERN may be required. If so, the castor-operation-external@cern.ch mailing list should be used
- OPN problems at PIC
  - connectivity between PIC and CNAF didn't work over backup route
- Early announcement for downtime registration not possible anymore?
  - Reported by PIC who wanted to register a several hours downtime a week ahead
  - The option to send an advance announcement seems to have disappeared?

Summary

- BNL site name change successful
- NIKHEF move completed. Back in prod
- CMS alarm ticket for lost processing time when closing dedicated queues at CERN for emergency kernel rollout
  - Procedure and communication need to be improved