Table of Contents

Week of 090105

WLCG Baseline Versions

WLCG Service Incident Reports

- This section lists WLCG Service Incident Reports from the previous weeks (new or updated only).
- Status can be received, due or open

Site	Date	Duration	Service	Impact	Report	Assigned to	Status

GGUS Team / Alarm Tickets during last week

- Tickets: see https://gus.fzk.de/ws/ticket_search.php⧉ and select "VO" - this will give all tickets, including team & alarm - or use the following link: * In the GGUS Escalation reports every Monday⧉
- Probably due to Year End holidays the escalation reports were unavailable today so here is the manual selection of GGUS tickets per experiment for the 19/12/2008-05/01/2009 period:
  - Alice: 0 tickets
  - Atlas: 30 tickets
  - CMS: 4 tickets
  - LHCb: 1 ticket

Daily WLCG Operations Call details

To join the call, at 15.00 CE(S)T Monday to Friday inclusive (in CERN 513 R-068) do one of the following:

1. Dial +41227676000 (Main) and enter access code 0119168, or
2. To have the system call you, click here⧉

General Information

- Web archive: https://mmm.cern.ch/public/archive-list/w/wlcg-operations⧉
- CERN IT status board: http://it-support-servicestatus.web.cern.ch/it-support-servicestatus/⧉
- elogs: https://prod-grid-logger.cern.ch/elog/CCRC'08+Logbook/⧉

See the weekly joint operations meeting minutes

Additional Material:

Monday:

Attendance:
local(Olof,Eva,Harry,Nick,Jean-Philippe,Julia,simone,Andre,MariaDZ);remote(Fabio,Gareth,Jeff).

elog review:

Experiments round table:

ALICE (reportedby HR): Over the xmas break the ALICE WMS at CERN became unable to schedule jobs sufficiently fast (both gLite 3.0 and 3.1 but with different symptoms) thus losing production time.

ATLAS (SC): Positive news. The only activity during the break was a major reprocessing of the most valuable part (0.5 TB out of 2 TB) of last Autumn's cosmic data. 8 Tier 1 were validated for this (not ASGC and NL-T1) plus 5 US Tier 2 sites and CERN and the production ran smoothly. Merging of AODs and DPDs is complete and distribution (from the merge sites) has started. Merging of ESDs is ongoing. JT asked how to identify merge jobs the answer being that they are far more i/o intensive than reprocessing jobs.

CMS (AS): Nothing special to report,

Sites round table:

IN2P3 (FH): They switched to using srmv2 and found some cms jobs putting a high load on on their dcache srm servers by issuing srmLs commands on subdirectories and therefore switched back to srmv1 (which has no srmLs command). Andrea followed this up and reported: I understand that the problem you mentioned at the daily WLCG operations meeting was the one experienced at IN2P3 before December 20th, when your SRMv2 was overloaded by many "srmLs" calls coming from CMS jobs. As Farida knows very well, as it was discussed extensively in the CMS Hypernews, the cause was that the srmls command was used to check if the directory where to stage out the files existed, but without the -recursive_depth=0 option. That was equivalent to an "ls" in UNIX. Now the ProdAgent code has been changed to use that option, which is equivalent to a "ls -d" and is much lighter on the SRM server. As soon as CMS starts using the ProdAgent with the updated code, it should be safe again to use SRMv2.

RAL (GS): There will be an outage of the castoratlas instance tomorrow to reconfigure it in preparation for the next ATLAS 10-million files test which should start next week.

ASGC (by email from JS): earlier this morning, one of the raid subsystem serving oracle rac showing overheat alarm and cause instability of i/o access error intermittently. OPS help resetting the controller and front end cluster service able to resume normally but the database serving CASTOR and LFC (they're accessing same partition map to same LUN from the controller) can't resume normally.

The error code 'ORA-00600: internal error code, arguments: [3020]' remain the same observed earlier but this mainly because the DB service shutdown unexpectedly when three of the cluster nodes offline in the same time.

OPS already perform recovery procedures and escalation (over 12Hr from the first event observed in monitoring interface) have been proceed to expert. we hope the service able to resume shortly.

SD have been add for two services (srm and srm2 of CASTOR) to avoid confusion from dashboard. keep you posted if having any update.

Services round table:

Databases (ED): Since 1 January there is a problem with CMS streams replication due to their running big transactions which are exhausing the memory of the DB servers. Replication to ASGC is down with a lock corruption.

FIO (US): seeing a big increase in the load on the CERN myproxy servers due to ALICE renewing submit server proxies on each job submission. They will change this to a 10 minute interval.

VOMS (MDZ): 2 minor VOMS problems over the break - one on Dec 23 fixed in a few hours and one yesterday transparent to users.

AOB:

Monday: 2

Tuesday:

Attendance: local(Eva, Harry, Jan, Jean-Philippe, Simone, Olof, Ulrich, Julia, MariaDZ, Nick, Kors);remote(Jeff, Michael, Gareth, Brian).
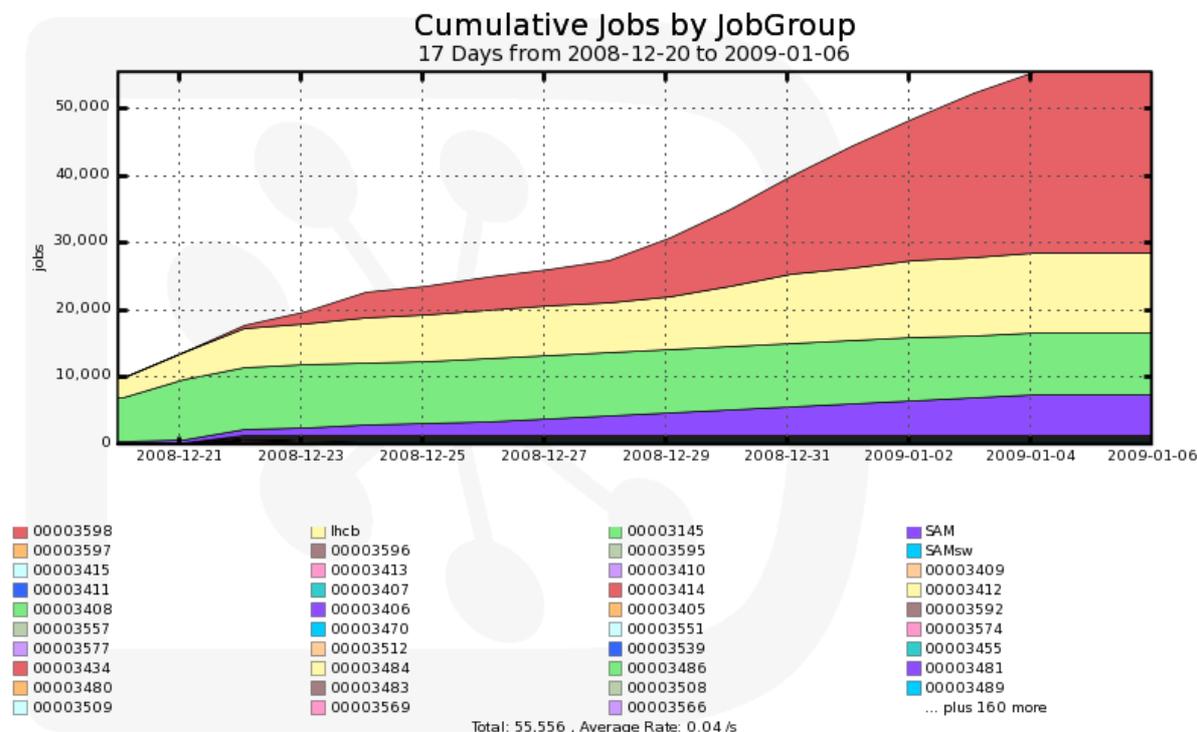
elog review:

Experiments round table:

ALICE (by email from L.Betev): 1. Despite the partial overload of the WMS-es at CERN, the production was very successful with ~4K jobs in parallel over the 3 weeks of Christmas holidays. 2. Maarten Litmaath (always online) warned us of the overload status and provided quick info on the situation with the WMS-es at CERN. With Patricia Mendez we were able to reconfigure the sites to use less loaded WMS, thus the loss of production time was very small. 3. Thanks to the expert information from both Patricia and Maarten, we were also able to identify couple of improvement points for the submission framework, which will be implemented and deployed shortly. 4. One of the biggest issues is that most of the sites rely on WMS at CERN, thus making it a critical point for the ALICE GRID operation. Further post-mortem will be performed in the coming week.
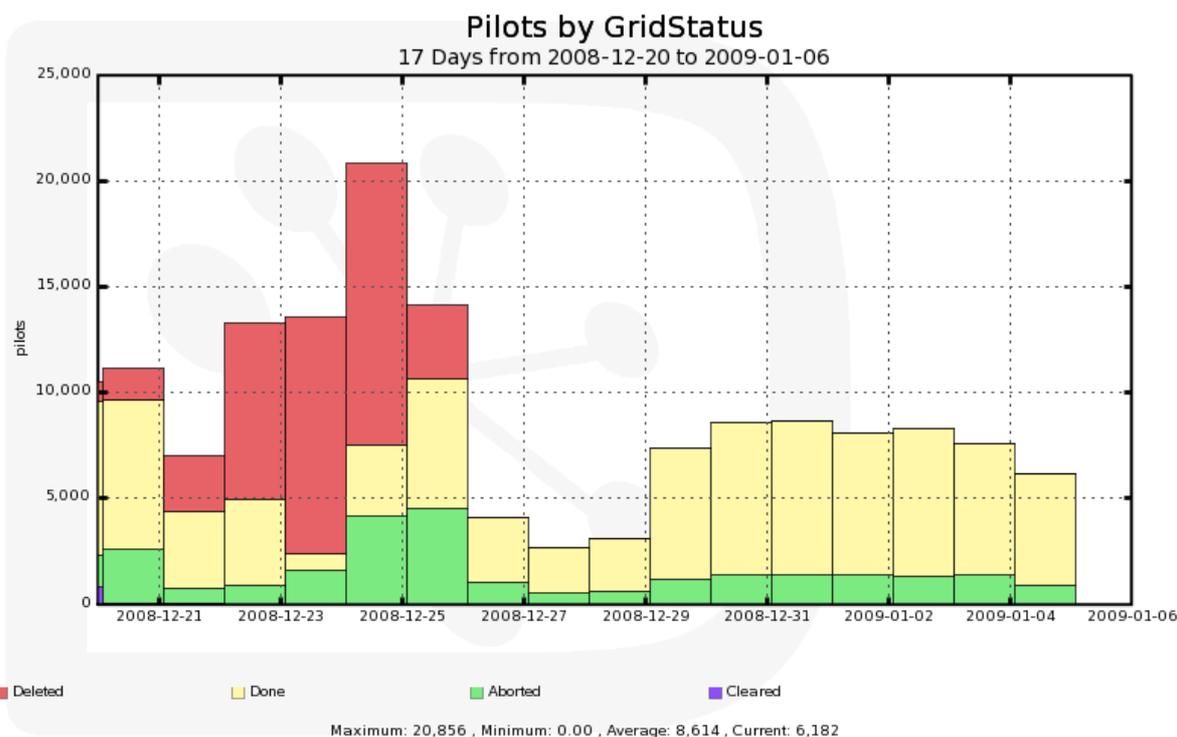
LHCb:

The plot in attachment summarizes the activities during the Christmas period:

- Red and green represent the dummy productions ~35k (most marked as failed, see
http://lblogbook.cern.ch/Operations/1114)
- Yellow shows user jobs ~12k
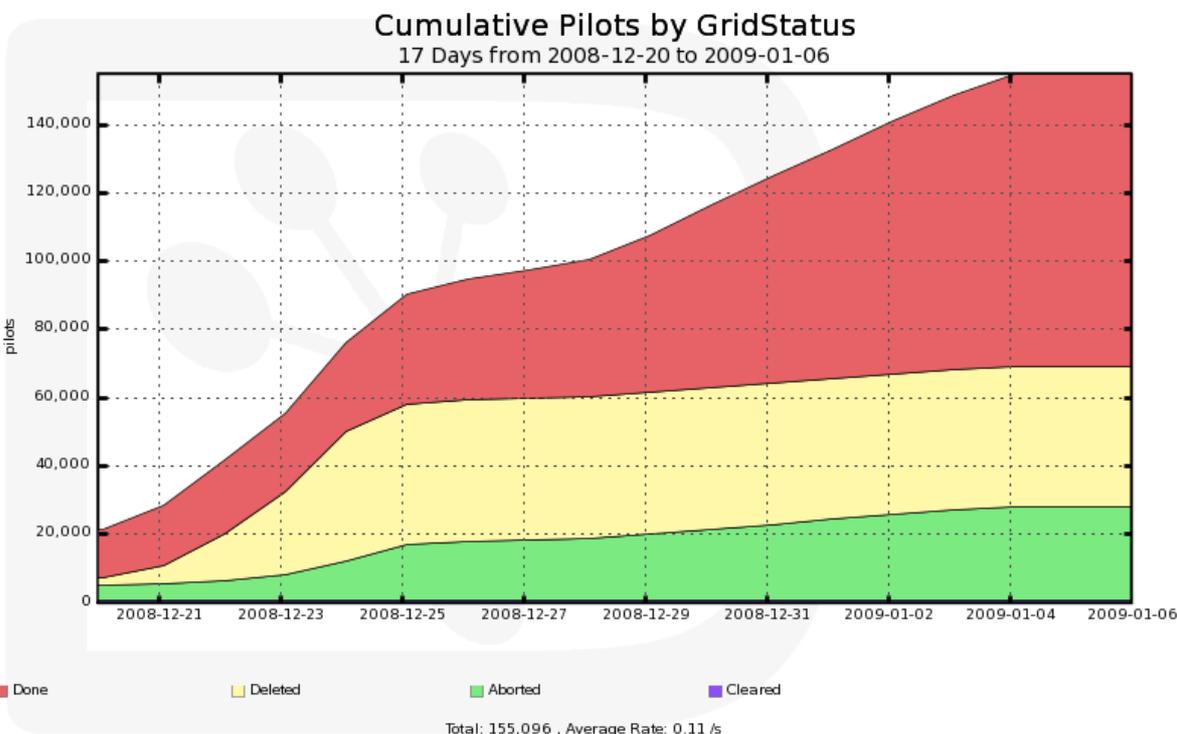- Purple are SAM jobs ~5k.

The dummy productions were being extended as necessary during the holidays so a relatively small number of jobs were executed. Jobs that managed to run show a reasonable ratio CPU/wall clock then the real problem experienced can be traced down into a problem getting pilots running. This is more evident in the following plot that shows the status of submitted pilots over the Xmas period and gives an indication of the throughput of the activities.

Pilots by GridStatus
17 Days from 2008-12-20 to 2009-01-06

Maximum: 20,856 , Minimum: 0.00 , Average: 8,614 , Current: 6,182

With the central task queues saturated we observed a small throughput (with ~8WMSes in use). The cause of large numbers in the 'Deleted' status is when the status of pilots is not determined after several days e.g. if an WMS becomes not available. Only after the 29th of December a steady ramp up of Done pilots (see last plot) shows then a more stable submission. From a look on the DIRAC logs there seemed to be a large number of list-match failures. It would be nice to know whether there were known problems (BDII, Network) that could explain instabilities on the RB side during the first week of the Xmas break time (21-29 of December) .

Cumulative Pilots by GridStatus
17 Days from 2008-12-20 to 2009-01-06

Total: 155,096 , Average Rate: 0.11 /s

ATLAS (SC): Functional tests are running smoothly with one communications problem between FZK and NDGF (put down to networking as their links to other sites are working). Both sites have been contacted by the ATLAS expert on call. There will be a past mortem analysis of the reprocessing exercise tomorrow. Jeff then queried the information that SARA had not been validated to participate in the reprocessing. Simone

Tuesday: 4

reported the site had failed validation tests before xmas possibly due to pilot jobs not coping with space tokens at two different subsites. To be confirmed at the post mortem analysis.

Sites round table:

RAL: Brian reported they had made a configuration change to one of the disk pools serving reprocessing jobs which should improve their efficiency. Some jobs were timing out waiting to stage data in from tape due to a too tight garbage collection strategy. Gareth reported the castoratlas changes reported on yesterday had been completed with a slight overrun and they were just waiting now to flip to a new version of srm.

NL-T1 (JT): There will be router changes at SARA on 12 January causing several short network interrupts.

Services round table:

DataBases (ED): The problem of CMS streams replication reported yesterday has now been fixed.

AOB:

Wednesday

Attendance: local(Harry, Simone, Ulrich, Roberto, Gavin, Julia, Nick, Jan, Sophie, MariaDZ);remote(Gonzalo, Michael, Jeff).

elog review:

Experiments round table:

ATLAS (SC): 1) There was a round table discussion of the Xmas reprocessing this morning and a full analysis will be given to the Atlas Distributed computing operations meeting tomorrow. The sites which failed validation for this were ASGC and FZK (not SARA as thought yesterday). Validation requires sites to pass the full processing chain including reproducing at the byte level the results of standard processing jobs. ASGC failed on data access and FZK was giving wrong results on the standard jobs, sometimes by a lot. There is no final conclusion on this yet. Four sites, SARA, IN2P3, RAL and PIC are still finishing off reprocessing job queues. Gonzalo will follow up why a backlog developed at PIC. 2) A CERN Remedy ticket has been issued for some files which are in the default pool when they should be in atldata and from which they can neither be copied nor deleted (they appear with 'invalid status'). JvE will follow this up. 3) Brian Davies of RAL has requested ATLAS to increase the rate of functional tests at RAL. This will probably be done for the start of the next weekly round of FT tests.

LHCb (RS): Dummy Monte Carlo running continues. From 21-29 December their main concern was low throughput of pilot jobs submission, thought to be due to WMS problems, then somehow things started to work smoothly again. Jeff asked if LHCb had stopped using gsidcap at SARA ? The reply was that LHCb are evaluating xrootd but still using gsidcap at dcache sites (like SARA).

Sites round table:

RAL (by email from Gareth): Owing to a clash of meetings both today and tomorrow I will not be able to phone into the daily WLCG meetings. (It is possible someone else will phone today.) We have scheduled an "At Risk" for the RAL Tier1 site over the coming weekend. Maintenance will be carried out on the electrical supply to the building. In order to reduce load during this work we will run at significantly reduced batch capacity for the weekend. Furthermore, in order to drain out the batch nodes in time the queues on many worker nodes will be closed during today.

NL-T1 (JT): 1) The December MB asked T1 sites to take a critical look at VO SAM tests and he has been putting in GGUS tickets for observed problems e.g. an ATLAS CE test ran 25 times since January 1 of which 28% have failed. Simone agreed to look at this case (later reported to be at https://gus.fzk.de/ws/ticket_info.php?ticket=45062 ☞). 2) Jeff reported progress on 3D database planning. Tenders are about to go out so deployment should be expected in about 3 months time.

BNL (ME): Have just taken delivery of a significant set of Sun Thumper disk arrays (an increment of just over a petabyte) which they are now working hard to install. This will allow to convert some exiting 650 TB of non-space token managed disk space into space token managed disk space. Internal tools will be used to move data from the non-space token areas.

Services round table:

CERN Lxbatch (US): Have started to drain 2 CE to migrate them to SLC5 submission. There are currently 58 SLC5 worker node of 9 job slots each. They will be published as being production SLC5 services.

AOB:

Thursday

Attendance: local(Harry, MariaDZ, Simone, Andrea, Nick, Roberto, MariaG, Gavin, Ulrich, Jean-Phillipe, Julia, Sophie, Jan);remote().

elog review:

Experiments round table:

CMS (AS): Since today some 2% of jobs are being scheduled on the FZK cream-CE, where they fail, so they must again be publishing this CE as available for CMS. A GGUS ticket has been raised.

LHCb (RS): Next Monday Dirac2 will be finally phased out so LHCb will no longer neeed an RB service (the CERN stoppage is scheduled for 26 Jan).There were problems with CERN and Tier 1 WMS yesterday indicative of a network problem.

Sites round table:

ASGC (by email from JS on their recent problems): The first Oracle cluster recover since '01-05-2009 21:40', and SRM database resume after '01-06-2009 03:20'. The CASTOR database restart with full functionality since couple of hours after RAC block corruption recovered. T0-T1 transfer able to resume after SRM service restart. Other affected services are LFC and FTS as cluster all base on same back-end raid subsystem.

ASGC LFC service should be back online since '01-06-2009 6AM' after fixing the back-end oracle database. Together with file transfer agent the other front end applications base on same database backend, that job submission able to resume after channel/vo agents restart.

Service availability of all critical core services base on Oracle should be back online after around '01-06-2009 6AM'.

Services round table:

Lxbatch (US): during the last two days some CERN lxbatch user jobs got killed automatically because of memory problems on worker nodes. The problem has been traced down to a badly behaving user process. The normal protections we have in place to avoid such situation did not catch this specific case. The offending processes have been terminated now, and we are adding additional protections at the queue level to reduce the

impact of such incidents. However, these changes only become effective for newly submitted jobs. For CERN users it can help if you specify job requirements concerning memory usage and reservations at submission time, using the -R option of bsub. A post-mortem is available at https://twiki.cern.ch/twiki/bin/view/FIOgroup/PostMortem20090601

ATLAS have been sending jobs to CEs that are being prepared for SLC5 so must not be checking their availability. Simone confirmed that two of the three job pilot factories use hard-coded CE names.

CASTOR (JvE): Name server problems last night resulted in a ATLAS team ticket. A weak spot in the software has been reported to the developers.

AOB: MariaDZ requested that ATLAS send her some GGUS alarm ticket use cases so the GGUS developers can prepare templates for common cases (LHCb have already sent some).

Friday

Attendance: local(Jan, Ulrich, Jean-Phillipe, Roberto, Patricia, Harry, Miguel, Sophie, Ricardo,);remote(Conf call failed to start).

elog review:

Experiments round table:

ALICE (PM): WMS backlog problems seen over Xmas are still happening and not understood. There has been a new version of ALIEN but the job submission part has not changed. A possibilty is because of overloading the myproxy server - Alice will reduce how they do this. Could also be related to VO-box overloads or some bad CEs that are provoking job resubmission (one was detected in Russia) though ALICE has requested zero resubmission be configured for them. Planning is to soon have 3 large WMS for ALICE at CERN (like the current WMS204). They also rely on WMS at NL-T1, RAL and FZK with a smaller one in Russia serving only the Russian federation.

LHCb (RS): A ticket was opened yesterday after a failure to register an LFC file that is to be used for next weeks MC production runs. No errors were returned or logged and it is still under investigation.

ATLAS (SC): Outstanding issue is transfer failures in both directions between FZK and NDGF. Thought to be networks and will need in-depth follow up. Planning is now for the rerun of the 10 million files test to start next Wednesday and run for 2 weeks. If successful all ATLAS VO-boxes will be upgraded. The data samples are already in place.

Sites round table:

Services round table:

Lxbatch (US): ALICE have been successfully using the CERN SLC5 worker nodes. LHCb hope to be ready for tests in a few weeks. Two SLC5 CE, CE128 and CE129, are available for such tests.

AOB:

This topic: LCG > WLCGDailyMeetingsWeek090105
Topic revision: r11 - 2009-01-10 - HarryRenshall

or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback