Table of Contents

Week of 090720

WLCG Service Incidents, Interventions and Availability

VO Summaries of Site Availability				SIRs & Broadcasts	
ALICE⧉	ATLAS⧉	CMS⧉	LHCb⧉	WLCG Service Incident Reports	Broadcast archive⧉

GGUS section

- GGUS Escalation reports every Monday⧉ (used for WLCG Service Report to MB)
  - ♦ Full search: https://gus.fzk.de/ws/ticket_search.php⧉ and select "VO" - this will give all tickets, including team & alarm
- **LHCOPN** Tickets in GGUS: see https://gus.fzk.de/pages/all_lhcopn.php⧉ and change your selection criteria. Future actions are also listed.
  - ♦ If network group participation is necessary, please invite them in time.

- Procedure to become a LHC Experiment VO TEAM member⧉
- Other recent GGUS FAQs⧉

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			
CERN IT status board⧉	M/W PPSCoordinationWorkLog	WLCG Baseline Versions	Weekly joint operations meeting minutes

Additional Material:

STEP09	ATLAS	ATLAS logbook	CMS WLCG Blogs⧉

Monday:

Attendance: local(Simone, David, Eva, Julia, Gavin, Gang, Alessandro, Patricia, Harry, Jean-Philippe);remote(Xavier Mol, Angela, Michael, Ronald, Fabio, John Kelly, Fabio, Danielle).

Experiments round table:

- ATLAS - SC: Last Friday's problem at ASGC of "No default service classes defined" has been resolved - Castor had gone down following a power cut. Serial processing at ASGC still cannot start as stage pool cleaning is not working. There were a couple of problematic Tier 2 over the weekend. NDGF reported some lcorrupt files last week and these are now being recuperated from other sites.

- CMS reports - DB: Will give a full report today then return to incremental mode. 1) A Tier0 CAF user is getting error messages reading from CAF pools but this is thought to be a user environment problem. 2) There were 7 tickets reopened at ASGC - 4 have been understood and closed. Main issue left is that since some weeks important data are not being migrated to tape. Gang reported that their tape drives (currently 7) are busy reading CMS files from tape (1200 jobs in queue) but that they

should reprioritise the use perhaps dedicating 2 drives to migration. Also ASGC fails to install a new software release reporting a locked lcg-tags file. 3) The link from a Russion Tier 2 to ASGC is failing commissioning since 5-6 days - looks like the Russian site is overloaded by ATLAS jobs. 4) The GRIF (French) Tier 2 fails to install a new software release reporting a hanging installation job. 5) Phedex transfers FZK to TIFR (India) are failing and expiring. Not thought to be a Tier 1 issue. 6) Loadtest transfers from PIC to CALTECH have been failing - may be due to network overloads at Caltech as reported in another ticket. Probably already solved. 7) Two old tickets (from end June) will be escalated to experts - CRAB jobs failing in Estonia and Phedex transfers to IPHC (France) timing out.

- ALICE - PM: Have been testing a new version of the ALIEN module submitting jobs to the CERN gLite 3.2 WMS and due to a bug submitted huge numbers of jobs causing it to overload. It had to be drained over the weekend so not much production was done. Known problems have now been resolved and testing continues today. The WMS itself is working fine and does not show the scalability issues shown by previous versions.

- LHCb reports - 1 billion events MC production is now running its last 500 jobs. Some further signal production submitted and a long tail of reprocessing activity last week (500 jobs at NIKHEF)

Sites / Services round table:

* FZK AP: Had a disk problem with stale NFS handles that was quickly fixed but that they would like to make more automated.

* SARA RS: Have installed 12 new tape drives. Still needing some configuration/tuning but already being used for data migration.

* IN2P3 FH: Advance warning of downtime from 22-24 September where work on the electrical infrastructure will lead to significantly reduced batch capacity. Full service resumption on the morning of 25th and detailed planning will be available early in September.

- CERN Networking: There was a short (5 minutes) Geant network interruption between Paris and London on Saturday about 08.30 and there is no backup link so traffic stopped. No report from Geant yet.

- CERN FTS: Some transfers have been observed to be going via version 1 endpoints by not using fully qualified surls. Details are being checked.

- CERN Databases: The CMS offline production database has been migrated to a RHEL5 64-bit platform (later news - hit performance issues and was rolled back).

- CERN CASTOR: Two switch modules in the CASTOR RAC infrastructure have to be changed. A test of the procedure was successful today so the intervention to be made tomorrow from 10.00-11.00 should be transparent but if not there is a risk of a 15 minute interruption.

AOB:

Tuesday:

Attendance: local(Gang, Graeme, Harry, Jean-Philippe, Simone, Miguel, Patricia, Ricardo, Gavin, MariaG);remote(Xavier (FZK), Tiju (RAL), Michael, Jeremy, Joel, Danielle).

Experiments round table:

- ATLAS - 1) Small issue with the ATLAS Panda to Oracle connections where we had two hanging connections on our side. Have now increased the logging on our side and also have worked out a procedure to kill such processes even from outside of CERN. 2) We want to retest STEP'09 reprocessing at ASGC, where we are waiting for confirmation, and at FZK. FZK then announced they are bringing in new tapes and doing some tape system reconfiguring so would like this to start Thursday of this week.

- CMS reports - 1) T0 CAF problem reported yesterday was a user problem. 2) ASGC software installation problem closed understood. 3) Link commissioning failure Russia to ASGC now passed to network experts. 4) 3 Tier 2 tickets closed - left open are failing test transfers CIEMAT to Brunel (London) which is waiting for site admin to come back from vacation and failing transfers T1 to IPHC (French T2). 5) The FZK to PSI (Switzerland) problem had been resolved the same day but the ticket was not closed so we need to check communications there.

- ALICE - ALIEN central servers are being relocated to new racking (within their building 12 centre) since 09.00 today so production has been halted. Scheduled to finish by 17.30.

- LHCb reports - Two GGUS tickets for stalled batch jobs at CERN (ticket 50406) and Nikhef (ticket 50408).

Sites / Services round table:

* BNL (ME): Some 2 PB of disk is being added requiring to reboot some 30 disk servers. Each takes less than 5 mins so they will be done today starting at about 14.00 UTC.

* CERN Databases (MG) : The CMS offline cluster migration to RHEL5 yesterday had to be rolled back adding an hour to the downtime. There is a known incompatibility between RHEL5 and QLogic switch firmware for which we have a workaround that does not appear to work well for databases under high load. Migration of the LCGR DB scheduled for tomorrow has been postponed. That of the LHCBR DB on the 5 August will also be reconsidered with a decision announced on Monday (Roberto Santinelli will coordinate for LHCb). Yesterday the ALICE offline cluster automatically evicted some failing disks and the DB had to be restarted today to rebalance the storage.

* CERN FTS (GM): Decided now to change FTS to use SRM 2.2 by default. Will be scheduled for next Thursday. Danielle questioned the FTS 2.2 release timeline. It is being tested in the PPS. ATLAS is testing the checksum changes and there are changes in how it interacts with SRM to test at some scale. Gavin will contact Nicolo Magini to involve CMS in the tests. Simone reported ATLAS have been testing it from CERN to TRIUMF since a few days at a steady 4 MB/sec and 30 files an hour with no problems seen.

* ASGC: Yesterday tried to clean a lot of data from the SAM disk pool so SAM CE tests recovered, at least temporarily. We have turned on garbage collection and plan to double the size of the pool.

* CNAF - by email from Luca: On Sunday 12 July at 01:13 am the ATLAS LFC standby database in Roma has bocome unreachable because of a storage problem. Moreover, at CNAF, on Sunday afternoon, a not well understood problem has caused the loss of connectvity to the storage area network from several Oracle clusters among wich there was the ATLAS LFC one. Due to this connectivity problem, several clusters have been automatically rebooted, after the reboot, the connection between the LFC front-end and the back-end has been automatically restored, but unfortunately the software wasn't functional. On Monday the 13th, the database was in hang with an error ORA-29702 (error in cluster group service operation). We found a lot of connections (order of 100) on the database, while the usual number is 40. The investigation of this problem is difficult because in the LFC front-end logs there is an hole between July 12 at 22:41 and July 13 at 10:19, probably due to the fact that the lfcdaemon was in hang. As the database in Roma was unavailable, the failover didn't succeed. The service has been restored in the evening on Monday the 13th, in both CNAF and Roma sites.

Tuesday: 3

AOB:

Wednesday

Attendance: local(Oliver, Harry, Graeme, Miguel, Jean-Philippe, Maria, Simone);remote(Tiju, Angela).

Experiments round table:

- ATLAS - GS: 1) The ASGC stage buffer has been cleared so reprocessing tests with pre-staging recall from tape have been launched. CPU activity will peak after about 8 hours. ASGC confirm that CMS is also busy running stripping work so this will be a useful multi-experiment test of sharing batch and data access. 2) ATLAS have been running a large scale analysis stress test in the UK for the last 24 hours. The Glasgow T2 had upgraded their dpm to version 1.7.2-4 and during these 24 hours under high load conditions their srm version 2.2 has crashed 7 times - not seen under low load conditions. J-P volunteered to look at a dump.

- CMS reports - apologies from Danielle who is unable to attend today.

- ALICE - After the relocation of the alice central services nodes the production has continued since this morning. They are not seeing any remarkable issue in any of the alice sites. For the coming two weeks ALICE will connect to this meeting only when they have a specific issue.

- LHCb reports - busy in another meeting.

Sites / Services round table:

* BNL (by email from ME): Yesterday I mentioned that we will have to reboot ~30 servers to connect the ~2PB of new storage. Turned out that this was not necessary since Solaris (as others do) allow to perform a "live attachment". Our experts found that this was working well and almost all of the 50 storage arrays are now visible behind the existing storage servers. The storage management group will be installing/configuring dCache pools on the new arrays and we expect to have them all available by the end of today.

* CERN Databases (MG): 1) The ATLAS DB at IN2P3 is down for a 12 hour intervention to fix a block corruption (announced yesterday). Also they have a problem that they cannot restart the capture process for replication of the ATLAS AMI database from IN2P3 to CERN. 2) The restart of the ALICE online to rebalance the disks after yesterdays failure did not work as planned and a complete recovery from disk is now being performed. Should be back in 1 more hour. 3) All upgrades to RHEL5 64-bit (would have been LHCb next week) have now been put on hold following yesterdays problems after the CMS upgrade and rollback.

* CERN CASTOR (MCS): Tomorrow there will be a transparent rolling upgrade to the CASTOR nameserver from 10.00 to 12.00. AOB:

Thursday

Attendance: local(Graeme, Alessandro, Simone, Gang, Julia, Jean Philippe, Michele, Gavin, Maria G, Steve, Dirk);remote(Xavier+Angela/FZK, Ono/SARA, Daniele/CMS, Brian+John/RAL, Luca/CNAF).

Experiments round table:

- ATLAS -ATLAS - (Simone) : At ~11:00 ATLAS had access problems with STORM at CNAF, a GGUS ticket has been created, the problem disappeared between 13:00 and 14:00 but the ticket has not been updated yet. Graeme: pseudo reconstruction at ASGC in progress. after an initial phase without tape activity, now 1000 files have been obtained from tape in 6h (not fast, but ok). On the

other hand back to again experiencing problem with balancing between ATLAS T1 and T2 resources (most of jobs running are in t2 jobs, more than configured limit to T2 slots). All jobs (t1+t2) are mapped for to same user which could be part of the problem. ASGC is working on this.

- CMS reports - (Daniele) Daniele gave an overview of a larger number of open tickets at T1,T2 and T3 which is well summarised on the CMS twiki (link above). During the discussion Angela/FZK suggested to review the number of parallel transfers at FZK. Luca/CNAF pointed out that iperf tests are also being done between CNAF and CALTECH.

- ALICE - no report

- LHCb reports - no report

Sites / Services round table:

IN2P3 (by email): This mail to let you know that the DBATL database was put back into production yesterday night (around 11 PM). Unfortunately I clicked on the delete link instead of the update one. (may be a confirm dialog should be useful to avoid such errors (certainly due to the late hour 😵)). The problem we hit was that a corrupt block in the sysaux tablespace made the all 4 instances to coredump with the following errors :

ORA-00607: Internal error occurred while making a change to a data block

ORA-00600: internal error code, arguments: [kddummy_blkchk], [3], [19539], [18007], [], [], [], []

Thanks to the Oracle support we were able to mark the block as corrupted and then to drop it. At this time, there is no clear explanation nor clue to find a reason to this corruption (which was logical not physical). All we know is that this corrupted block was Block Type = Pagetable segment header block, and the HWM was wrong (HWM block 50 beyond extent boundary 8).

- MariaG/CERN: in addition also AMI replcation from IN2P3 is still down
- Ono/SARA: problem with black hole in one cluster due to ldap problem after reboot - now fixed
- John/Brian: RAL - postponed split of LFC (ATLAS instance) due to staff availability
- Gang/ASCG: ATLAS and CMS tests (800 CMS jobs and 1000 ATLAS jobs) show that fair share works between the VOs. Since ASGC applied the ORACLE patch no further big id problems have been detected.
- Luca/CNAF: will also provide information about the reasons for the storm access problems during the scheduled "at risk" period
- MariaG: ALICE online problem required a full recovery, now fully operational again. Problem has been tracked down to series of ASM bugs, which are already documented in the Oracle bug tracking system.

AOB:

- Graeme: DPM bug under very high load reported by Glasgow has been fixed (affects 1.7.0 and above) by the DPM developers. A patch has been provided for DPM 1.7.2. Simone: why did this not happen during STEP? Graeme: seems to happen only under very high load.

Friday

Attendance: local(Eva, Harry, Julia, Simone, Alessandro);remote(Jeremy/Gridpp, Riccardo/INFN-T1,John+Tiju/RAL,Michael/BNL,Danielle/CMS).

Experiments round table:

- ATLAS - Problems accessing INFN software area - solved a few hours ago. See their site report. Main issues were at ASGC where pseudo reprocessing tests are running routed to their T1 together with MonteCarlo routed to their T2. Their T1 and T2 endpoints point to the same batch resources using the same userids and the long MonteCarlo jobs were grabbing all the job slots and holding up the shorter production jobs which also risk to have their pre-staged data purged before they start. ASGC attempted to improve this but killed all the MC jobs unannounced which confused the shifters. ATLAS resubmitted but the problem was still there. ATLAS then decided to treat ASGC as a single site with a single grid queue where PANDA decides the next work package to schedule and this is working much better having run 4000 reprocessing jobs since. There are still delays accessing data from tape even though we heard CMS skimming jobs have finished with some jobs now having waited 6 hours for tape data. Tests will continue over the weekend.

- CMS reports - There is a new VObox problem at CERN affecting Phedex - details on Monday. CAF data access timeouts at CERN due to a server that was down most of the day - waiting for confirmation from the user before closing. Other ticket reports are given in detail in the linked report.

- ALICE -

- LHCb reports -

Sites / Services round table:

* FZK (by email): unable to attend in person today but nothing to report.

* CNAF: One of the shared software area servers (see the ATLAS report) failed and the automatic failover was prevented by a second hardware failure in a switch port. Batch jobs were stopped while repairs were done and were resumed a few hours ago.

* CERN Databases: Replication of the ATLAS AMI database from IN2P3 to CERN has resumed. The problem was with a rule set used by the capture process.

AOB:

-- JamieShiers - 16 Jul 2009

This topic: LCG > WLCGDailyMeetingsWeek090720
Topic revision: r10 - 2009-07-24 - HarryRenshall