Table of Contents

Week of 080818

Open Actions from last week:

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(Simone, Jean-Philippe, Harry, Luca, Ricardo);remote(Gonzalo, Derek, Daniele, Michael, Jeff).

elog review:

Experiments round table:

ATLAS (SC): Had 3 rather critical problems over the weekend. 1) Many Castor errors trying to export reprocessed data due to its being on tape not disk. These data were garbage collected before older data whereas we thought GC was a FIFO operation. We will follow up with CASTOR operations. Despite this all cosmics were exported (except to RAL) though at low efficiency and DAQ ran into CASTOR at 750 MB/s in 12 hours slots. 2) Affecting all of the Tier 1 is that during the weekend the data set type name embedded at the beginning of each file changed from data08_cos to data08_cosmag without advance warning to allow Tier 1 sites to switch to a new storage directory mapping. We will follow up this poor communication with the ATLAS management. 3) RAL had several problems over the weekend . On Saturday their SRM Oracle data base was giving errors then on Sunday they were not accepting FTS data from CERN though they were still exporting. We found the CERN channel set to 0% and it would have been good to tell us that had been done. Today I can see RAL is in an unscheduled downtime. D.Ross explained their CASTOR was down from 07.00 to 09.00 Saturday then Sunday they had first an LSF disk full (which stops staging) then a database disk full which lead to the unscheduled downtime. The expert called out set the CERN-ATLAS FTS channel share to zero as it was the only one active. H.Renshall said we should follow up how sites could indicate such a configuration change (e.g. site statusTwiki).

CMS (DB): CMS have started the CRUZET4 (cosmics but with magnet on) run with a couple of subdetectors in so far. They have a dataops shift in place concentrating on Tier 0 workflows.They have daily meetings at 16.00 and will use the CCRC08 elog for general (unstructured) observations and their cruzet3 elog for more detailed reports. H.Renshall invited them to make relevant observations to these minutes.

Sites round table: Jeff (NL-T1) reported they are at risk today and tommorow while they change network routers. Will be application transparent unless a cable switch over exceeds the tcp timeout.

Core services (CERN) report:

DB services (CERN) report: - The apply process at ATLAS OFFLINE was aborted on Friday afternoon when trying to replicate the statements in order to drop the tables from one schema. The problem is a known bug, reproduced on ATLAS after setting a new parallel Streams setup between the ONLINE and the OFFLINE databases to replicate the PVSS schemas. This bug is assigned to Oracle development but the progress is very slow. The workaround is to setup schema rules at the apply side. This change will be implemented this week.

- The corruption found on the atlas online server is not affecting services, but an intervention is scheduled on Wednesday from 14:00 till 14:45 to fix the issue via a switch to a new database using Oracle dataguard/standby technology.

Monitoring / dashboard report:

Release update:

AOB:

Tuesday:

Attendance: local(Andrea, Simone, Roberto, Patricia, Ricardo, Gavin, Luca, Harry, Julia, Nick, Jean-Philippe);remote(Jeremy, Michel, Ricardo).

elog review: Related to the CMS report

Experiments round table:

- CMS: 1) CRUZET-4 started, regarding the CMS summary of the shifts it has been found one run with no events. This issue, already observed in the past has still to be understood. 2) Transfers T0-T1 via Phedex working smoothly with a transfer rate over 600MB/s. 3) Regarding the CERN-FNAL transfers, the problems observed will tr to be solved using the CERN FTS server instead the server placed at the T1 4) Still discussing the list of T2 sites which will enter the next production.
- ATLAS: Follow up of the report presented yesterday. The problem observed with CASTOR@T0 during the weekend seems to be a combination of several issues: a bug included in the algorithm responsible of the file system. The problem has been solved and a new patch will be probably applied tomorrow (transparently to the experiment). In addition 100 TBs will be requested to the ATLAS pool as soon as Kors comes back the next week. Finally the setup of the garbage collection of the default pool will be simplified and changed basically following: 1st comes, 1st goes way. Few problems observed also today, all of them notified to the corresponding responsible: CASTOR@CNAF is showing 100% of failures, site in Michigan unreachable, and Oracle problems observed at RAL being checked by Jeremy.
- LHCB: Issues (most probably coming from local network interventions) observed at SARA have been reported via a GGUS ticket. Although the reason has not been totally clarified, the ticket has been closed as soon as the problem dissapeared. Several problems observed at PIC and GridKa and also some few italian sites while installing the new software in the corresponding software area. The experiment recommends the use of static accounts instead of pool accounts, but in any case this is still and open (and quite old and well known) issue. Finally the upgrade of Dirac2 to include the SRMv2 clients has been announced.
- ALICE: The experiment follows the MC production. There is an issue observed with the CREAM CE deployed at GridKa for Alice testings. Basically the std.err and the std.out files cannot be currently retrieved into the VOBOX. The issue will be followed with the developers and the site managers

together with the experiment software experts.

Sites round table: Oracle issue at RAL being followed by Jeremy. In addition Jeremy reported some DPM problem at some of the UK sites when they changed the size of their spacetokens. The bug # is 40273. Michel asked ATLAS about the elog entry reported during the weekend.

Core services (CERN) report: Nothing to report

DB services (CERN) report: stream replication for LHCb finished

Monitoring / dashboard report: Development performed for CMS to calculate the site availability is also in place for all VOs. Still the calculation is not included but it is possible already to visualize some results. VOs are required to check the links and provide feedback:

- ATLAS: http://dashb-sam-atlas.cern.ch/dashboard/request.py/latestresultsview ☒
- LHCb: http://dashb-sam-lhcb.cern.ch/dashboard/request.py/latestresultsview ☒
- ALICE: http://dashb-sam-alice.cern.ch/dashboard/request.py/latestresultsview ☒

Release update: gLite upgrade already available although still not included in the production repository. Those sites which have already migrated do not required any further operation. Those sites which still have not migrated should wait for the corresponding announcement. The reason is the Oracle testing which still has not finished. 2) Regarding the FTS service for SLC4, feedback from the experiments and sites is required. The issue will cbe discussed during the next weekly operations meeting.

AOB: Nothing reported

Wednesday

Attendance: local(DanieleB, Harry,Miguel Anjo, Roberto, Patricia,Sophie, Olof, Jean-Philippe,Julia,Simone);remote(Derek, Gonzalo,Jeremy,MichaelE,Jeff).

elog review: summary CMS

Experiments round table:

- CMS Tier-0 workflows: First runs yesterday (CRUZET-4 day-2) have been reported as successfully repacked and promptly reco'ed. Some later runs had missing files, guilty components have been identified, being fixed. Distributed Data Transfers: CRUZET-4 data transfers started. T0-T1: overall quality looked fine all through the day, except CNAF (note: scheduled downtime). T1-T1 trafic nothing of relevant.T1 processing: FNAL noted to be marked with a yellow warning on the SiteView: 50% of analysis jobs successful; failure comes primarily from oneuser trying to write to existing files in an SE, not a site issue. All other T1 sites marked as green in all CMS-specific tests. Julia: do you use site status board? Daniele: yes, extremely useful. Harry and Gonzalo: why do you re-route to FNAL-PIC. This is a Phedex feature that allows to choose the most convenient path and does not depend on the quality of thee transfers. Phedex updates constantly its view of various site-connections rating them on the basis of their performances and speed.
- ALICE: no huge activity. Production ramped down for preparing the first cycle of the next large MC production. Deployment of xrootd to all Tiers 2 (over DPM) that will be in the production mask as soon as they will restart. Conitnuing the all day testing activity.
- ATLAS: various problems reported last week are now fully under control. Simone reports that the week now of SD at RAL for ATLAS is fairly worrying and the fact that the window of this downtime is continuously increasing should perhaps require a postmortem analysis. Derek replies this is still under investigation, they cleaned up rogue entries but not still clues of the problem that might reside in the oracle backend. Jeremy asked whether the large number of aborted jobs at UK T2 will be

redistributed somewhere else being RAL not available Yes they will. There is also techincal discusison about how to re-associate a T2 to another T1. Simone presented also the next week plans. He's reporting that the FDR-2 phase-c will start mostl likely next Tuesday for the following three days. This is basically a full test that will also involve ATLAS T1s (the previous phase-b involved only the T0). This is not much data (no special requests for sites) but with the same priority of a data-taking like activity. Also shift procedures will be exercised.Tomorrow the announcement of 4 new dataset projects.

- LHCb: CCRC-like activities ongoing through DIRAC3 together with normal prodution still through Dirac2. Problems with gfal_ls at CNAF due to CASTOR in downtime (GGUS 39890) and accessing shared area (because some shortage of the GPFS disk servers). Problem accessing at RAL CondDB from Brunel application (correlated with the ORACLE problem there), problem with libdcap at PIC (GGUS 39898) , fixed as soon as LHCb started using the local dcap installation. Gonzalo reports that the dcache clients available in the WN are there since three months from now. NL-T1, GRIDKA and IN2P3 are running smoothly. Remarkable that at Lyon (with DIRAC2) LHCb managed to run a job accessing all input files via xrootd. Plans to use xrootd both at IN2P3 and SARA. General remark: we experienced at sevrral sites failures accessing through Persistency the ConditionDB information with jobs crashing. This is under investigation to exclude it is a problem server side. Jeff comments: his understanding is that xrootd had no real authorization except for the ALICE model. So if both LHCb and ALICE want to use xrootd, LHCb and ALICE will likely need to agree that it is ok to see each other's data or alternatively to ask to setup a a dedicated xrootd, which probably means a dedicated dcache, which we might choose not to do (it would cost money -- another set of machines,and more manpower as dcache is not so stable). Hence it is a good idea to ask SARA what is possible. Gonzalo also comments about the need to have a sensor for the CondDB service.

Sites round table:

JT: bad idea for the experiments to hardcode the name of the CE. NIKHEF is going to replace its historical hostname tbn20.nikhef.nl. The Computing Element service at host tbn20.nikhef.nl (grid site NIKHEF-ELPROD) will be stopped next week. On Friday August 22 at 10:00, tbn20 will be removed from the information system. Then, new jobs cannot be submitted to tbn20 anymore, but already submitted jobs can still complete. On Tuesday August 26 at 10:00, the CE service at tbn20 will be stopped. A replacing CE service is already operational at host gazon.nikhef.nl.Soon an additional CE will be installed on host trekker.nikhef.nl. Rely on the BDII is always a better idea...

Derek: There is a problem with the ATLAS instance of Castor at RAL. The result is corrupted entries for Castor requests, not files, in the Castor database. This has only been observed for ATLAS but we cannot tell if this is due to the instance or ATLAS's load or use patterns.

Diagnosis of the problem is continuing but we have no expectation of a quick resolution as we will need to cleanup the database once the problem is fixed. Thus we are announcing a downtime for ATLAS Castor from now until noon BST on Wednesday 27th August. This status will be reviewed in the afternoon of Tuesday 26th. When the service is working again we will involve ATLAS in extensive testing before we go back into production.

We understand that this is a very serious situation that has a large impact on ATLAS. However, we believe it is better to be realistic about the situation (serious and unknown problem) rather than promise a solution within a day or two.

We are very sorry for the disruption that this will cause to ATLAS operations and users and will provide further updates as soon as we have significant progress to report.

Core services (CERN) report:

Nothing

DB services (CERN) report:

The scheduled intervention on the ATLAS ONLINE database supposed to finish before three o'clock didn't yet because some small problems. The intervention consists in fixing a corruption on the storage side by moving the database to a new storage.

Monitoring / dashboard report:

Nothing

Release update:

Nothing

AOB:
Nothing

Thursday

Attendance: local(Julia, Simone, Ricardo, Roberto, Jean-Philippe, Harry, Daniele, Steve, Luca, Miguel);remote(Derek, Michael, Shaun, Gonzalo,Jeff).

elog review:

Experiments round table:

CMS (DB): Last 24 hours has not been so good with serious CRUZET4 issues in storage management at the pit. They are planning to roll back today to the software version used in CRUZET3 in order to continue. In data distribution CNAF had a scheduled downtime but are still not back 100%. They have made progress in exercising Tier 0 to Tier 2 transfers.

ATLAS (SC): Over this weekend they will use the cosmics run to observe if the recent CASTOR 2.1.7 patch 16 has fixed the garbage collection problem (of young files) seen last week. Next week they will perform the FDR2-c run, exercising all components at full nominal rates, from Tuesday evening till Friday morning. Also a new type of DPD will be generated. They also have transfer problems to CNAF which came back recently at 100% for only 30 minutes then went back to failing. They are going to perform some specific low level tests with RAL to help debug their ongoing CASTOR problem.

LHCb: 1) Some LHCB specific SAM CE tests were systematically failing everywhere because of a bug in one of the modules for installing software. These tests have been temporary set "not critical" until a fix will be in place and fully tested. 2) RAL Castor problem (it seems to be the same problem affecting ATLAS). 3) CNAF shared area got full yesterday. This morning CNAF admins freed 2TB. The last days problems with GPFS volume serving the shared area seem to have severely affected the ratio CPU/wall clock (as reported yesterday). CNAF CASTOR: the same error reported in the past against CASTOR is still there even after the downtime: Error: Too many threads busy with Castor at the moment Unknown error . 4) SARA problem accessing data via gsidcap (problem failed to create control line(#39953). 5) GRDIKA: The 'Error reading token data: Success' when issuing gfal_ls problem has appeared today at GridKa and seems to be persistent.

Sites round table:

BNL: Michael asked if ATLAS were resuming throughput tests. Simone said not as they prefer to not be distracted from preparing for the FDR-2c run.

RAL (email from DR): The problem reported yesterday, which seemed limited to the Atlas Castor instance, has now also occurred on the LHCb instance:

Last night the castor instance for lhcb at RAL began experiencing the same problems as previously seen on the atlas instance. This involved violation of a primary key constraint on uniqueness in Oracle. At about 02:30 the instance was shutdown to avoid any additional logical corruption in the database. Investigation showed that the error first began to appear at 10:56 on 20 Aug.

Investigations are still continuing, but the cause is still unclear. We have not been able to reproduce the error on the Atlas Stager through internal load testing, therefore we intend to work with Atlas and open the Atlas stager to Atlas functional testing to see if that will cause the problem to reoccur.

The LHCb and Atlas instances share the same Oracle RAC, so we have not ruled out an issue on the RAC.

NL-T1: Jeff reported the LHCb gsidcap problem is not due to their recent network changes. The gsidcap service had been restarted and it still did not work but mysteriously restarted later.

Core services (CERN) report:

DB services (CERN) report: The corruption found on the atlas online server and reported in the C5 of last week has not affected DB services and data availability. However the corruption has caused a serious risk of further spreading and affecting services in case of subsequent HW failure. Therefore over the weekend a standby database has been set up with Oracle Dataguard technology to further protect the database in case of data loss. An intervention has been performed on Wednesday afternoon when the standby database has been activated for production of Atlas online. This has fixed the corruption issue. The intervention was scheduled for 45 minutes but has finally lasted 1.5h because of some unforeseen issues with archivelogs movement and streams setup. There are strong indications that the 'silent corruption' was generated by a faulty drive, this is currently under investigation by IT-FIO. We have also added additional monitoring at the DB level to allow early detection of corruption issues.

Monitoring / dashboard report:

Release update:

AOB:

Friday

Attendance: local(Harry, Daniele, Julia, Jean-Philippe, Luca, Nick, Ricardo, Patricia, Alessandro, Roberto);remote(Derek, Michael, Gonzalo).

elog review:

Experiments round table:

CMS (DB): Tier-0 workflows: Online ran with CMSSW_2_0_10 for hrs to get rid of a problem with lumi-sections nbs. No changes in any of the subsequent workflows (prompt reco is still going with CMSSW_2_1_4, with only two failures seen). The problem with the inability to properly log data through the Storage Manager was diagnosed to the Global Trigger crate (which when rebooted sent proper lumi section numbers). We are now rolling forward to CMSSW_2_1_4 online again to return to our original program. --- Suffered of a backlog of DBS registrations, this cleared late in the Fermilab morning, it leftus with ~7 hrs of DBS registrations backlog, already cleared out. Now back into catching & registering files at real time like we should.

Distributed Data Transfers: Following CERN-T1 transfers. CERN-IN2P3 shows some "file exists" error, but few active transfers. RAL, PIC, and ASGC are all receiving CRUZET-4 samples frequently from FNAL: some number of the files had staging problems at CERN so PhEDEx rerouted from FNAL, which apparently has lower cost. Some evidence also of CERN FTS server heavy load reported by shifters: the FNAL transfers in /Debug use the FTS server at CERN and the quality is low, the /Prod link uses the FNAL FTS and the quality is good. The errors on /Debug appear to be FTS issues of loosing the transfer.

Tier-2 Processing: Production continues. Generally the job exit codes looks fine. There are a higher number of non-exit code zeros at the T2_FR.* The predominant code is 60314 (problem in SRM stage-out). To be understood if this is a workflow or site issue, if the latter then we will open GGUS appropriately (we now track internally via Savannah and CMS contacts).

ATLAS (AdG): Last night the production dashboard was not showing any ATLAS activity. This may be a problem in the http service of the dashboard or in the dq2 related services (a restart of a VObox brought the monitoring back). FDR2-c starts on Tuesday and the Tier-1 coordinators have been primed. We expect to handle 38 TB of raw data, 13-16 TB of ESD and a few TB of AOD and DPD. The directory label will be fdr08_run2.

ALICE (PM): No problems in production. A new version of Aliroot is in preparation. They are contacting Tier 2 sites to make sure their data storage will be ready to be used as soon as collisions start.

LHCb (RS): 1) GridKA gfal_ls issue (https://gus.fzk.de/pages/ticket_details.php?ticket=39995⧉) seems to be related to a conflict of version between clients used by LHCb (1.10.15) and the server running there. Should we expect to have backwards incompatibilities for a minor release?

2) at CNAF opened an ALARM ticket yesterday about the problem of CASTOR (not yet fixed despite the downtime of the days before!) and the ticket has not been updated since then. I'm wondering whether the ticketing alarm system works as it is supposed to do otherwise no other plausible and acceptable explanation for this abnormal delay in dealing such top priority tickets. Discovered that the GPFS server serving the shared area there (critical service for all VOs) also serves other disk services. It has to be understood that the "shared Area" is a crucial service and agreed QoS must be provided.

3) NL-T1 several problems discovered and reported at about the same that (accordingly SAM at least) : 1. file access problem https://gus.fzk.de/ws/ticket_info.php?ticket=39953⧉ 2. tURL ONLINE but not tURL available problem https://gus.fzk.de/pages/ticket_details.php?ticket=40001⧉ and 3. timeout problem uploading files (from SAM, https://gus.fzk.de/pages/ticket_details.php?ticket=39996⧉)

They all seem to be due to a SRM process not running (and restarted). SARA people are working very hard to sort all their instability problems in the Storage area (and we appreciate that) by improving their monitoring systems and understanding the cause why daemons are dying but LHCb needs to anticipate that the observed QoS there in the last months is far to be optimal.

Sites round table:

BNL (ME): Can Alessandro explain the recent high rate of transfers CERN to BNL failing with file not found (at CERN). The reply was they do not yet have a clear answer but suspect a tape migration issue. It was not a large number of files and investigations continue.

RAL (DR) : No progress yet on reproducing the CASTOR request corruption problem. Monday is a UK bank holiday so they will be absent from the daily meeting. On Tuesday they will upgrade the LFC instances for LHCb and ATLAS.

NL-T1 (by email from JT): 1) our old trusty CE 'tbn20.nikhef.nl' has been pulled from the info system. the DNS alias 'ce03.nikhef.nl' which had pointed to tbn20 now points to 'gazon.nikhef.nl'.

2) an additional CE has been installed on host 'trekker.nikhef.nl'

3) we are aware of the gsidcap problems at SARA. As dCache itself prints no error messages associated with this failure mode, we are having a hard time understanding what is actually going wrong. SARA folk are improving the monitoring to try and catch the problem earlier.

PIC (GM): Last night a disk filled up about 23.00 stopping the ATLAS 3D service. It was fixed this morning by restarting the complete Oracle instance.

Core services (CERN) report: A network switch problem at 09.59 caused intermittent router problems for the next hour which lead to breaks in access to disk storage.

DB services (CERN) report:

- an HW problem affected lhcb online from Thursday morning for about 24h. Database node number 1 went down and 1 path to the storage was lost. Oracle cluster has been put back to full production on Friday afternoon after a reboot of the cluster nodes. An issue with DHCP configuration affecting database service for lhcb online was also discovered and fixed together with lhcb sysadmins

- streams replication between Atlas online and offline DBs stopped on Friday afternoon. The cause was investigated and found to be due to a user-created table which was missing the primary key.

Monitoring / dashboard report:

Release update:

AOB:

This topic: LCG > WLCGDailyMeetingsWeek080818
Topic revision: r14 - 2010-06-11 - PeterJones