TRIUMF

Old WN mysteriaously dies but reboots ok.
FTS Oracle upgraded to 10.1.0.5.0

1 node offline for benchmarking.
*-TRIUMF FTS channels disabled since no disk space available

BNL

Problem:
A user generated significant load on our PNFS dCache core server. Other users experienced timeout errors.

Cause:
This user\\\"\"s program bug repetitively read the same files via dccp.

Severity:
PNFS server load was about 5 between 7:00PM to 10:00AM Friday morning.

Solution:

Kill the user\\\"\"s job.
DQ2 installation and upgrade. There is no data service during the upgrade.
Problem:
ERROR:GRAM Job submission failed because authentication with the remote server failed (error code 7)at 10:15:37AM

Cause:

A containercert.pem, which prima depended on was expired. The original containercert.pem should be deprecated and use the existing host certificate which is monitored by Nagios.

Severity:
Two Grid gatekeepers were out of server for two hours between 10:15 and noon.

Solution:
Our Grid administrator changed the prima configuration file to point to hostcert.pem and its key instead. So prima is now fixed.

Problem:

Severity:

Maintenance:
Two more GridFtp servers (dcdoor08 and dcdoor09) are installed and running. These two servers will share intensive data transfer load on each of GridFtp servers and reduce the number of timeout errors experienced by users and SAM probes. The network buffer settings are as following

# Set several kernel parameters to increase network buffers and improve performance:
net.core.wmem_max = 1048576
net.core.rmem_max = 1048576
net.core.wmem_default = 131072
net.core.rmem_default = 131072
net.ipv4.tcp_wmem = 4096 65536 1048576
net.ipv4.tcp_rmem = 8192 131072 1048576
#net.ipv4.tcp_mem = 98304 131072 196608

Problem:
Nagios reported errors on 4:40AM. ???ERROR:GRAM Job submission failed because authentication with the remote server failed (error code 7)???

TIME: Sun Jun 17 10:02:50 2007
PID: 12294 -- Failure: globus_gss_assist_gridmap() failed
authorization. gridmap.c:globus_gss_assist_map_and_authorize:1910:
Error invoking callout
globus_callout.c:globus_callout_handle_call_type:727:
The callout returned an error
prima_module.c:Globus Gridmap Callout:400:
Gridmap lookup failure: Identity Mapping Service could not be contacted

Cause: Note the \\\"host name mismatch in tcp_connect()\\\". For whatever reason, the prima modules on OSG gatekeeper is rejecting the alias (gums.racf.bnl.gov) when the actual canonical hostname is different from the alias, and the host certificates??? subject name contain the canonical host name.

Severity: The primary OSG gatekeeper is out of service between 4:00AM~10:AM until our system administrator did intervention.

Monday: June/11/2007

Problem:

One of GridFtp servers node (dcdoor02) crashed on 6:00AM Monday morning.

Cause: an IRQ kernel error.

Severity: the GridFtp server was off-line for four hours and 14% connectivity was lost due to this problem.

Solution:

Our system administrator updated the server since it hadn't been updated in many months. Then, after the update, we rebooted it a second time so it would be running the updated kernel.

Problem:
Two OSG gatekeepers were reported critical at 04:54:56 EDT 2007

Cause:
GUMS server went off-line (but log output had stopped. we restarted Tomcat and it appears to be functioning). The new host certificate generated for the GUMS server was missing an attribute needed for a server. The old one had

```
X509v3 extensions:
        Netscape Cert Type:
           SSL Client, SSL Server
```

while the new one had:

```
X509v3 extensions:
        Netscape Cert Type:
           SSL Client, S/MIME
```

Severity:

Two OSG gatekeepers were impacted for three hours before our administrator intervention.

Solution:

We disabled the GUMS on two OSG gatekeepers and used the static Grid map files to allow BNL CE to be accessible right away. In the mean time, we obtained a new host certificate with a proper server attribute for our GUMS server.

Tuesday: June/12/2007
Maintenance:
Kernel update was performed on all seven GridFtp server nodes one by one. Each GridFtp server has less than an hour downtime. The update is transparent to users. No need to make announcement.

Wednesday: June/13/2007

Problem:
A fraction of data transfers from BNL to other ATLAS Tier 1 sites failed with certificate mismatch errors.

Cause:
A fraction of our dCache read pool nodes have bad certificates that their DNs do not match with their hostnames. The error message is shown as follows:

06/07 12:01:04 Cell(SRM-dcsrm@srm-dcsrmDomain) : Authentication failed.
Caused by GSSException: Operation unauthorized (Mechanism level:
[JGLOBUS-56] Authorization failed. Expected
"/CN=host/acas0203.usatlas.bnl.gov" target but received
"/DC=org/DC=doegrids/OU=Services/CN=acas0399.usatlas.bnl.gov")

Severity: This problem affects the data transfer directly between the remote party and these affected dCache read pool nodes while the data transfer via (GridFtp servers) was not affected. BNL to other Tier 1 data transfer does not use GridFtp server nodes; all data transfer from BNL affected nodes to Tier 1 sites experienced transfer failures. USATLAS production is NOT affected by this problem.

Solution:

We replaced these bad certificates on Wednesday. We notified the ATLAS data operation team to confirm whether the lower performance problem with the data transfers from other Tier 1 sites to BNL. We will add scripts to validate the host certificates.

Thursday: June/14/2007
None.

DQ2 installation and upgrade. There is no data service during the upgrade.
DQ2 installation and upgrade. There is no data service during the upgrade.

CERN

Networking services
On Tuesday, the LANDB Oracle database was unavailable during one hour starting from 16:10 because of a cluster restart. This interrupted all the network operation including network/register which had to be blocked. From Wednesday 20:00 till Thursday 11:30, the Oracle application server stopped running normally. This interrupted all the network operation

since our internal tools rely on it. Although the public interfaces (network/register, mike, etc) were not affected, we received complaints because we were unable to performed database updates.
We have had three incidents in the last 2 weeks (see last C5 report for the first one) affecting seriously the maintenance of the network. The design of the complete system will be reevaluated in the next weeks in collaboration with DES

Fabric Services
- CASTORLHCB was successfully upgraded on Monday to CASTOR 2.1.3
- CASTORCMS is currently being upgraded
- Some SRM mapping problems have occurred during the week. Problem fixed. Extra measures and monitoring introduced.
- And unfortunate coincidence between SAM and SRM problems prolonged the apparent downtime of CERNPROD. The SRM issue was fixed 1h after the SAM tests stopped working. Since the SAM results were not being updated SRM appeared has down long after the problem had been fixed.
- Long tape queues for different reasons
-- Atlas using stagecdr...
-- Compass accessing same files through Castor-1 and Castor-2 disk caches (to be discussed Friday)
- The kernel upgrade (quarterly update) on SLC4/64 is going on with low priority (half way through now)
- An update of LSF 6.1 base and LSF CERN extensions rpms has been rolled out on lxbatch and CE nodes which addresses the following issues:
+ LSF pim daemon bug fix
(filling up /var with useless messages)
+ protection against hosts with exhausted memory:
when the difference between allocated address space and the limit is lower than 0.5GB, the host will refuse new jobs
+ bug fix for SGM pool accounts affecting the token grabbing and extension via gssapi
- problems on lxbatch and lxplus
+ many ATA kernel panics, possibly caused on some HW architectures by SMART
+ dead locks due to memory overcommitments: A parameter change to fix this problem was suggested in the last CCSR by the linux experts. This change was applied to lxplus, lxbatch and the grid CE cluster, and was meant to become effective with the next reboot, but in fact became active immediately.
It turned out that a large number of nodes were running with much more allocated memory (mostly by user processes) than physically present, and as a consequence batch nodes

as well as lxplus machines fell over. At the moment the
change became effective we lost about 1000 user jobs and 140
worker nodes this way. The problems are understood and under
control. The problem is being escalated to RedHat.
+ in the process of investigating the unexpected problems
caused by the parameter change described above, a bug in the
load balancing was found, affecting only SLC4 machines,
and in particular lxplus nodes. The consequence of this bug
was that the load balancing kept
redirecting users to lxplus nodes with exhausted memory.
This bug was fixed on the 19/6.
+ the SLS sensor which is monitoring lxplus has been updated
to test the offered nodes for this kind of problems
+ now the number of crashes on lxplus seems to have reduced a lot,
and with the bug fix in the load balancing in place the overall
availability of the cluster seems to improve.

EGEE CERN Regional Operations Centre (ROC):

---

* Registration/certification of new site UFRJ (Rio de Janeiro,
Brazil) started.
* We have got a number of open GGUS ticket concerning the site
YerPhI (AM): basically all tickets are to be attributed to
the poor network connection and to recently discovered
firewall issues in the network route to the Armenian site.
The network issues are under investigation by people in
CERN ROC and GD group.

EGEE Pre-Production Service Coordination:

---

* The SL4 UI was released to the PPS this week. Significant
issues have already been found regarding this new service
(see comments under the section \"CERN GRID Pre-Production
Site\")
* Setting up of the grid interoperability testbed (between the
EGEE grid and OSG grid) is almost complete. Verification
testing to start early next week.
* Testing of FTS 2.0 and SRM 2.2 continues, with the full
collaboration of the experiments.
* Procedure now in place to give rapid feedback to development
and testing teams from the testing of new middleware releases
in the PPS. This should lead to faster bug fixing and clearer
list of \"known issues\".

CERN GRID Pre-Production Site (CERN_PPS):

---

* Upgrade of the site to gLite 3.0.1 PPS-UPDATE 32 completed

* Upgrade to gLite 3.1.0 PPS-U01 (UI). During the installation we noticed that:
1) The directory $INSTALL_ROOT/etc was removed and its content put in $INSTALL_ROOT/external/etc. As a major consequence for the user the location of the environment script has changed.
2) The name of the environment script was changed from grid_env.(c)sh to grid-env.(c)sh,
Both these changes are likely, in our opinion, to create surprise and discomfort in the user community. Therefore we opened two critical bugs to track these issues.
- [10]https://savannah.cern.ch/bugs/index.php?27361
- [11]https://savannah.cern.ch/bugs/index.php?27362

CERN Grid Operations
---
* Sandbox partition full on one CMS WMS last week due to a bug in CRAB. The CRAB developers have been contacted.
* Preparation of new CDB templates containing a minimal set of packages to install on GD production nodes managed by quattor (SLC3 and SLC4 operating systems).
* Middleware upgrade (update 26 for gLite 3.0) for all the LCG RBs.

WLCG Transfer Service:
---
* Transfer ranging from 10 to 500 MB/s, averaging around 200MB/s per day.
* Mostly traffic from CMS.
* 0 tickets have been submitted to sites this week
4 tickets have been moved from last week
* Throughput plots:
[12]http://gridview.cern.ch/GRIDVIEW/

SAM (Service Availability Monitoring):
---
* 20 June ~20:00 - 21 June ~10:00 web services were down (tests not updated). Reason understood, temporary fix for the time being (permanent fix a little later)
* 15 June - 20 June: due to a GOCDB synchronization problem, downtimes were not updated correctly
* quattor component for tomcat released (needed by SAM servers)
* bdii2oracle (tool to get resources information from BDIIs) re-engineering: input modules ready *
* integrated WMS sensor being certified by TIC Team (Krakow)

Physics Database Services
Smooth upgrade to FTS2.0 in production. Defragmentation of FTS tables lead to recovery of 130GB of space and performance improvement.

PIC has joined the ATLAS Streams setup and the synchronization is going on.
SARA is frequently unavailable; they are trying to increase the space for database archive log files to leave sufficient headroom for the expected
workload from ATLAS and LHCb.

ARDA/EIS
AMGA:
Progress on getting AMGA to build in ETICS with a lot of help from Alberto Di Meglio so that it compiles now under ETICS.

ALICE:
up to 4000 concurrent jobs during last week (~2900 avg.)

ATLAS:
Migration to the new version of the Data Management system completed (v 0.2 --> 0.3)
Continuing working on central ATLAS services (reviewing the security part).
Production systems and Ganga have been migrated successfully within the agreed timeframe.
Job Priorities: very short term solution (rollback to previous stable scenario) for the problems observer by ATLAS has been agreed with SA1 and SA3; timescale is 2 weeks. A short term solution (introduction of appropriate tags in information system) also has been agreed, some remaining points will be clarified; timescale is 2 months.

CMS: testing the gLite WMS using CRAB, the CMS tool to send analysis jobs.

EIS:
Ongoing test activities with StoRM SRM. During last week two types of tests have been done: one consists job Grid jobs submitted to CNAF which access LHCb files on StoRM via a SRM2.2 client and analyze them with DaVinci. A second type consists of stress tests with up to 300 parallel requests to a new StoRM instance (~5 Hz). The goal is to continue stressing the system until saturation.

Geant4: running the validation of the 4th release candidate.
Helping the CERN theoretical group to run jobs on the Grid (running

QCD jobs).

Dashboard:
Support remote installation (NIKHEF).
LAL colleagues working in collecting BDII information to be fed in the dashboard.
Started the activity to monitor Condor-G jobs (jobs submitted bypassing the resource broker).
SAM results being stored in the CMS dashboard data base (under test)

Report for Tier1 GridKA (FZK):

[author : Jos van Wezel]

Almost all difficulties this (and last week) stem from stability problems on the CE\\\"\"s. More specific, the info provider system (gris) sometimes returns erroneous data (i.e. no data). Consequently the job requests fail. We are investigating and have setup more extensive monitoring of all relevant activity on the CE. However for the time being the situation remains unsatisfiable.

NDGF

Network problems causing packet loss and low throughput between NORDUNet and GEANT since Wednesday last week identified to be a faulty DK GEANT router. Replaced.
Very low throughput between Slovenian and NDGF sites.
Throughput problems from Slovenia pinned down, an overloaded firewall replaced.

NIKHEF

We cleaned the group mappings in the DPM database. There were still old non-VOMS group entries corresponding to SGM and PRD groups, sometimes sharing the numeric group id with another group. Shared gid-s confused the new DPM dynamic info provider.

SARA

Problem: We have had problems with dcache pools running out of disk space on their root file systems. This problem was caused by idle gridftp doors generating lots messages stating that it has nothing to do. This generated huge log files.
Solution: Removed the gridftp logs and restarted the gridftp door and pools. In addition we have tightened the logrotate rules so that the dcache log files are not only rotated and compressed each day but also when they exceed a 2 GB limit.

RAL

Our OPN link to CERN went down on the 20th, this meant that the SEs and our FTS couldn\"t not be contacted and contact the SEs at CERN. Now that there are SAM tests for SEs, is the replication test in the SAM CE job still useful? Wouldn\"t it be more appropriate incorporated into the SE SAM tests?

One of our LCG RBs lcgrb01.gridpp.rl.ac.uk had a large number of idle jobs due to experiment submission scripts, this was preventing other jobs from completing successfully. Alice have agreed to stop using lcgrb01.gridpp.rl.ac.uk and we will be deploying a new RB for their use shortly.

PIC Tier-1

Date: 15/06/2007

Problem: Sporadic failures in the CE sft-lcg-rm test, due to timeouts or spurious errors in the srm-disk (dCache) service.

Severity: Medium. We observe spurious errors from dcache, which we believe are caused by the constant load on the system, coming mainly from CMS transfers.

Solution: We migrated to dcache-1.7 on 21-22 June and hope the new version will not present these spurious errors anymore.

--

Date: 16/06/2007 from 18h until aprox 24h.

Problem: Spurious errors in castorsrm SRM-tape service, with error message "Invalid CRL: The available CRL has expired".

Severity: Medium. The problem seems to appear in only one of the two castorsrm gridftp doors, and it was transient. Retrying transfer should eventually work.

Solution: We did not do anything, and the problem disappeared in few hours. Probably there was some problem in one of the runnings of the update CRLs in one of the gridftp servers, and the problem went away in the next running of the cron.

--

Date: 21/06/2007

Problem: Bad configuration in the STAR-CIEMAT channel. This channel, following request from CMS, is configured as 'srmcopy'. it seemed the requests to the srm server arrive with a not normalized URL. It looks like this is a bug of dCache.

Severity: Low. This is a configuration problem in an FTS channel which is currently being debugged/tested by CMS.

Solution: To solve it we have to change the surlnorm parameter in the fts, first manually and then using yaim. This corrects the problem.

--

Date: 22/06/2007

Problem: Problem with FTS in the morning. The agents were stopped and we were unable to restart them. Having a look at the logs it seemed a problem with a library that was missing. In contact with fts-support to try and solve it.

Severity: Medium. The FTS server at PIC is down, but the transfers from the experiments using fts.pic.es are stopped, due to the recent upgrade of dCache at PIC.

Solution: Problem still open. Working on it.

--

Date: From 22/06/2007 at 18:00h until 25/06/2007 at 11:00h

Problem: The SRM-disk was migrated to dcache-1.7 on 21-22 June. The service was open on 22 June at 18:00 and tested by CMS. File transfer by CMS was ok. On monday 25 we discover that lcg-utils commands to store/retrieve files from SRM-disk do not work because the information is not being correctly published.

Severity: Medium. The service restored ok, and proved to be working for FTS-driven transfers (as those used by CMS in the testing) but the lcg-utils commands were failing..

Solution: The information provider of SRM-disk service has been fixed on Monday 25 June morning. It was some missing configuration steps.

--
General Comments:

- We saw ATLAS request to unconfigure the VOViews. We started configuring all this stuff by hand, following ATLAS request. Then, some weeks ago, we re-configured with YAIM, as soon as the automatic configuration was available. Now we will wait until the new version of YAIM appears that unconfigures this automatically.

- The SRM-disk service was in scheduled downtime from 21/06 at 9:00 until 22/06 at 18:00 for upgrading dCache from 1.6 to 1.7 version. On Friday 22 June the service was

restablished, and CMS transferred files from CERN at a constant and high rate during the whole weekend.