WLCG 'Weekly' Service Report

Jamie.Shiers@cern.ch

~ ~ ~

WLCG Management Board, 8th July 2008

Introduction

- This 'weekly' report covers two weeks, as last week's MB was cancelled
  - Last week:
    - Monday: EGI_DS workshop
    - Tuesday: LHCC "mini-review" of LCG
  - This week:
    - Tuesday: MB F2F
    - Wednesday: GDB, C-RSG
    - Friday: OB
- Notes from the daily meetings can be found from:
  - https://twiki.cern.ch/twiki/bin/view/LCG/WLCGOperationsWeb
    - https://twiki.cern.ch/twiki/bin/view/LCG/WLCGOperationsMeetings
- How many people read these daily? Regularly? Sometimes? Never?
- (Some additional info from CERN C5 reports & other sources)

Power & Cooling

- IN2P3 – weekend of 21/22 June suffered a serious problem with A/C & had to stop ~300 WNs.
    - Not sure what follow-up was – PM?
    - **[ROC France]:** Air conditioning trouble at IN2P3-CC due to excessive heat. **(operations minutes of June 30 – somewhat terse?)**

> Simone mailed Luca directly at 17:00 hrs on Sunday after CNAF had been down for 40 hours and nobody reacted to the alarms we raised. Luca answered around 20:00 hrs. Network problems reported and problems with StoRM and LFC. At the time of this eLog entry (23:00) CNAF still dead.

- CERN – scheduled power test
    - Check that problems in previous (unscheduled) cut were fixed. Power tests will continue – quarterly basis?

- N.B. post-mortems should be 'push' – not 'pull'!

Monitoring

- CERN availability at zero from Friday 20th – traced to limit (~100) in # tuples published from gstat to SAM. Caused by increase in # CEs at CERN (hardware 'rotation')
- CMS dashboard load: index on query suspected for performance bottleneck. Seems better, but more follow-up needed...
- GridView throughput display collapsed on 26/06. Restart of TomCat cured it (monitor the monitoring?)
- Check of which sites have installed FTM – follow-up at weekly joint ops meeting **[more later]**
- Network problem affecting all BDIIs hit SAM monitoring on Wednesday 2nd July. Will availability be automatically corrected? John Shade to follow-up
- Request from CMS to use WMS to submit CMS SAM tests – need to understand user mapping for this
- Problem with SAM Thursday night which causes many alarms to be sent to RAL (also CERN). Called RAL staff out by SMS. And the SAM staff???

Jumping to the conclusions

- The main monitoring sources for the challenge were experiment specific monitoring tools.

  For activities at CERN (Tier0, CAF) Lemon was widely used.

  SAM and SLS were used by all experiments for monitoring of the status of the services and sites

  In general worked quite well and provided enough information to follow the challenge, to see whether the targets are met, to spot the problem rather quickly

- In most cases the problems were triggered by people on shifts using the monitoring UIs, alarms are not yet common practice.

  We do not yet have a straight forward way to show what is going on in the experiments for people external to the VO and even for users inside the VO (non-experts).

For performance measurements except Lemon for CERN related activities and T0-T1 transfer display in GridView, nothing else was provided to show the combined picture of experiments metrics sharing the same resources. [FTM deployment]

Sites are still a bit disoriented. They do not have clear idea how to to understand their own role/performance and whether they are serving the VOs well

Work is ongoing to address the last points

Status of FTM deployment at Tier1s

- Which LCG tier-1 sites have successfully deployed FTM?
  For those tier-1 sites which have not deployed FTM, when is this planned to take place?
  The reason the experiments want this is because the FTM publishes transfer logs to GridView (thanks Steve ;o)

  Responses:
  ASGC: Already deployed and operational.
  BNL: Already deployed and operational.
  CNAF: Installed last week but still being tested.
  DE-KIT (FZK/GridKa): Already deployed and operational.
  IN2P3-CC: Not yet installed. Hope to have it in place during July.
  NDGF: Not installed. Will take at least 3 weeks if needed.
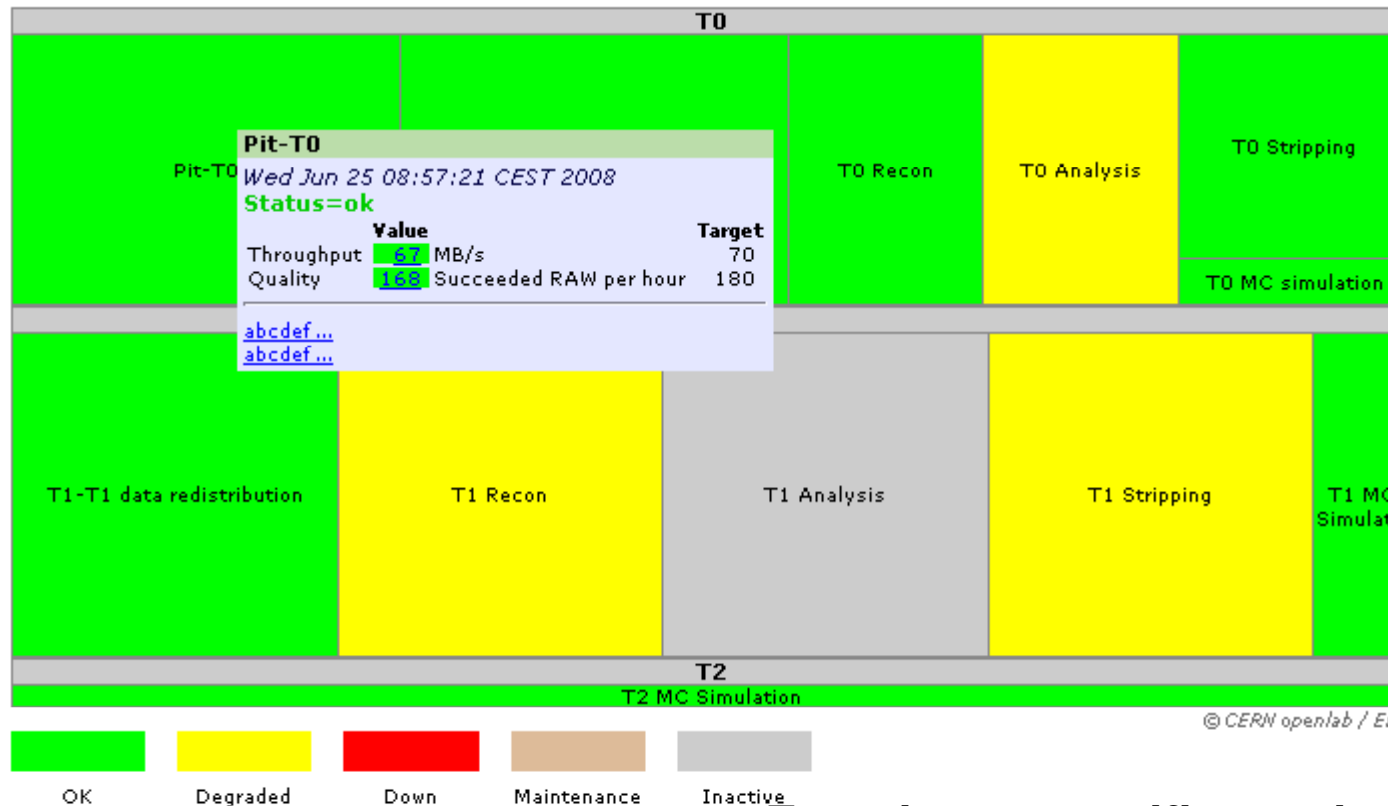 PIC: A test instance is being deployed now and is planned to be in production by mid July
  RAL: Already deployed and operational.
  SARA: Intend to install FTM early in July.
  TRIUMF: Already deployed and operational.

LHCb ActivityMap Test Page

AIM:

1. Exp... vis... per...
2. Pas... and...
3. Hid... spe... rel...
4. Pro... Sit...
5. Def...
6. Ext... (DI...
7. A p... sta...

Experiment specific monitoring page

javascript:_GM("http://lhcbweb.pic.es/DIRAC/systems/accountingPlots/dataOperation")

Data Management

- Long-standing problem (LHCb) with gsidcap at IN2P3 (multiple connections). Fixed by dcache patch level 8?
  - More in slide notes...
- Encouraging tests from xrootd (LHCb) at SARA
- Gfal 1.10.14 and thread safety? TBC
- LFC (RAL) – crashing – recommend to install 1.6.10-6 which fixes this problem. Push out info through Joint Operations meeting(s)
- LFC (LHCb at CERN) – wrong (20) # threads – reset to 60
- New version of CASTOR + CASTOR-SRM. Requires an intrusive upgrade (to be scheduled)
- New space tokens for ATLAS (ATLASSPECIALDISK) and LHCb
- Further clarification of ATLAS space token permissions for DPM needed – requires DPM development?
  - See slide notes for follow-up...
- ATLAS problems with reprocessing at dcache(?) sites – needs follow-up to locate problem (dcache, gfal, ATLAS etc.)

Other Services

- Continued upgrade of CERN Oracle services to 10.2.0.4. Tier1 sites in the process of upgrading (majority done...)
- Break(s) in CNAF streams synchronization due to site downtime(s)
- Some experiment scripts pointing to old AFS version of UI. /current is recommended.
- New kernel in PPS – experiments encouraged to test!
- CMS 'rogue user' submitting ~10k jobs overloaded batch system – various actions to 'fix' and isolate experiments. Should there not have been more discussion on Tier1 workflows???
- NIKHEF – cluster maintenance Friday 27th June. # cores available will temporarily reduced
- BNL – on-going problem with Michigan calibration site (invalid space token).
- SL4 FTS testing on PPS. Aim to move to production (CERN) before pp collisions (end July? Early August?)

Data Services - Detail

- We have received the new Castor 2.1.7-10 software, and will now upgrade our Castor pre-production service.

- As there are many changes in this release, the deployment will take some time.

- In addition, the first release candidate of the Castor/SRM 2.7 has been installed on our pre-production SRM service.

- We hope to receive the first official release by the middle of next week, and we are starting to plan the upgrades of the production endpoints.

- As both these upgrades will be intrusive, we will try to schedule them at the same time, to limit the number of service interruptions to one per LHC experiment.

DB Services - Detail

- During last week we had 2 reboots of 2 different nodes of the CMSR database (no service unavailability, although database connections were lost on those nodes and applications running there needed to reconnect):

- On Sunday 29th June around 23:00 the node number 3 was evicted from the cluster and rebooted. After the reboot Oracle processes started normally and continued working without any further problems

- On Tuesday 1st July around 11:50 the node number 2 was evicted from the cluster and rebooted.

- In both cases the symptoms were very similar but the root cause is still not clear. We are in the process of deploying some diagnostics tools that should help to discover the source of the problem.

- There was a general power cut at CNAF during the weekend. On Monday, the Streams setups were split in order to avoid the impact on the main flow replication. As soon as CNAF confirms that the situation is stable, we will merge the configurations.

- NDGF will upgrade their databases to Oracle 10.2.0.4 on Tuesday 1st July.
- PIC has scheduled an intervention to upgrade all the 3d databases 8th July.
- IN2P3 has received new hardware and plans to migrate the ATLAS and LHCb databases to Linux in the next weeks.

ATLAS

- Discussions – at least partially prompted to site outages – on "cloud vulnerability" – particularly in the light of recent events at CNAF – and ways of affordably / manageably protecting against this
- Updates for space token requirements for Tier2s (see ATLAS ops meeting)
- Running FTS transfers using SLC4 server in PPS
- Pre-staging / recall from tape: 5 sites (2 castor, 3 dcache) – major recall problems
- Effort starting to understand these issues within ATLAS OPS

CMS

- RAL job limits
- CMS mask T1 sites so that no analysis jobs are scheduled there
- Preparing for restart of next cosmics run – CRUZET 3
- Some 100 TB of data – custodial site will be IN2P3. CERN CAF and FNAL will get full copy, CNAF, PIC and Taiwan get partial copy.
- Request to use WMS for CMS SAM tests

LHCb

- Delayed roll-out of interim DIRAC3.

- Debugging of file access problems at IN2P3

- Testing of generic pilots & capability of DIRAC3 to run multiple payloads in July

- Additional space tokens at CERN for MC simulation

ALICE

- Redeployment of Alien at all sites. Changes in 'LCG module' to submit jobs to RB/WMS sites in homogeneous manner.

- (Seriously?) affected by one of myproxy nodes being down (proxy renewals not working and proxies corrupted – all VO boxes down.

- Need to follow-up why:
    a) Service is supposed to be H/A
    b) Why ALICE is particularly exposed in this case

- And here it is...
    https://twiki.cern.ch/twiki/bin/view/FIOgroup/ScLCGPxOperations

Summary

- A lot of activity – less structure than in May, but this is likely to be the trend...

- Will not be easy to provide continuity / coverage over the summer months, but it is presumably preferably that people are away July / August than later...

- **Please automatically provide post-mortems as / when appropriate!**
  - IMHO the CERN-IT-FIO examples show the value of these!

- "Post-mortem" events carried over into next QR? OB?