



CCRC'08 Weekly Update

Jamie.Shiers@cern.ch

~ ~ ~

WLCG Management Board, 27th May 2008

Introduction

- This is the last week of the May CCRC '08 exercise – and the one which should see most activity
 - **The Full Monty?**
- Generally, things continue to run rather smoothly with problems being addressed quickly
- This report focuses on the basic infrastructure and not on the experiments' (impressive) results
 - **The WLCG "Added Value"**
- These are covered e.g. in the experiments' wikis – see CMS example linked to MB agenda

Sites

- NIKHEF: increased installed disk by 120TB. Now have 88TB of space allocated to ATLAS. Problem with cooling last week meant power-off of WNs. DPM 1.6.7 problem (lifetime reset to 1.7 days after resize of space reservation) – upgraded to latest production release.
 - This was discussed prior to start of CCRC '08 – felt to be too tight (insufficient time for adequate testing...)
- RAL: closed for public holiday on Monday 26th (came up in relation to planned LCG OPN intervention). Problems will be dealt with by on-call system.
 - AFAIK, there were none...

Network (& Communications)

- LCG OPN: upgrade s/w of two CERN routers that connect to Tier1s – bug affecting routing for backup paths in some(?) cases.
Needs reboot, hence downtime of ~5' per router.
- Agreed to go ahead with upgrade (first one then the 2nd on Wednesday if no problems seen).
 - Upgrade of 2nd router scheduled for tomorrow
 - <http://it-support-servicestatus.web.cern.ch/it-support-servicestatus/>
- Problems seen with Alcatel phone conferencing system due to *“the SIP service crashed on one of the servers”*.
 - Hopefully this can be monitored / alarmed. It is not only extremely annoying but also highly disruptive to remote collaboration.

Core Services (1/2)

- Several problems and actions on Data services in relation with the ongoing CCRC tests
 - Castor and SRM services suffered from instabilities, which are actively being followed with the developers. Several of these issues are understood and fixed in software, and we have deployed a number of bug fix releases during this week Castorcms and Castoratlas now run the latest 2.1.7-7 release, addressing the slow stager_rm and putDone commands.
 - the Castor nameserver daemon has been upgraded to 2.1.7-7, to work around the overloaded service during backup runs of the databases
 - Castor SRM has been upgraded to 1.3-22, addressing some deadlock issues
All these upgrades were transparent.
- Long standing (since beginning of CCRC) CASTORCMS/t1transfer GC problem was finally understood by the castor development team yesterday:
 - the SRM v22 prepareToGet request was giving a too high weight for files already resident in the pool while files being recalled from tape (like in early May) or copied from other diskpools (like the last 2 weeks) were given a too low weight.
 - This caused a very high rate of internal traffic and it is quite impressive that the system managed to keep up given the load it generated on itself...

Core Services (2/2)

- Additional SRM problems were seen over the w/e. The atlas-operator-alarm@cern.ch e-mail list was used to report these problems but further clarification of the follow-up is required
 - The problem was solved, but it seems independently of this mail). (SMOD follow-up)
- We also used the emergency site contacts phone number (TRIUMF) which high-lighted two problems:
 - The printed version in B28 is out of date
 - The people in the TRIUMF control room did not seem to be aware of the procedure(s) for informing (Grid) contacts at the site

DB Services

- The capture process on the online database for the ATLAS Streams environment was aborted on Tuesday 20th around 02:23 due to a memory problem with the *logminer* component. The system was in sync already several minutes after restart.
- High-load has been seen from CMS dashboard application & traced to a specific query. Follow-up / tuning planned.
- ↳ **Various problems affecting SRM DB – high number of connections, threads, locking issues etc. (Still) being investigated...**
- 💣 **Some configuration / tuning being discussed – the implications and actions will be discussed with the experiments in the immediate future (aim is to initially ensure DB stays alive, possibly requiring different retry mechanisms at middleware and experiment-ware levels)**

Experiment Issues - LHCb

- LHCb have had a long standing issue at IN2P3 (a gsidcap door issue?) and RAL (where also jobs requiring to access files from WN crash, to be still understood)
- "File access issue(s)" continue at some sites
 - also starting to be a problem at NIKHEF, since yesterday
- Encouraging (first) results from the procedure of first downloading data into the WN and then accessing it locally.
- WN /tmp issue (FZK, NIKHEF) with files still used / required by running jobs being cleaned up "too enthusiastically"

Monitoring

- Gridmap moved onto new h/w – gets about 25K hits per day
- Agreed to do systematic follow-up on problems reported – were they picked up by monitoring?
- See next slides...

Monitoring / Logging / Reporting Follow-up

Issue	Comments / Actions
How to see T1-T1 transfers?	Sites to install FTM – CCRC'08 baseline components updated.
Failures accessing SRM_ATLAS	Detected by users → ticket
ATLAS online-offline streams	☺ Picked up by alarms – see example mail
NIKHEF worker nodes	Problem with EGEE broadcast – should have been picked up on ATLAS dashboard.
BNL low incoming data rate	Detected by “manual” monitoring (aka eye-balling)
Garbage collection on (CMS) t1transfer pool	Detected by users – seen by PhEDEx.
NIKHEF dCache upgrade to p3 patch level – space token definitions not visible	All FTS transfers stopped – manual monitoring
srmServer 'stuck' (monitoring also stuck ☹)	Actuator to “re-incarnate” as stop-gap

Streams Monitoring

De : Streams Monitor <LCG3D.Monitor@cern.ch> Répondre à : <Pdb.Service@cern.ch> Date : Sun, 11 May 2008 12:51:31 +0200 À : <PDB.Service@cern.ch>, <Gancho.Dimitrov@cern.ch>, <florbela.tique.aires.viegas@cern.ch>

Objet : CERN-PROD : Process error report STRMADMIN_APPLY_ONLT@ATLR.CERN.CH

Streams Monitor Error Report

Report date: 2008-05-11 12:51:31

Affected Site: CERN-PROD

Affected Database: ATLR.CERN.CH

Process Name: STRMADMIN_APPLY_ONLT

Error Time: 11-05-2008 12:51:21

Error Message: ORA-26714: User error encountered while applying

Current process status: ABORTED

See also: oms3d.cern.ch:4889/streams/streams

Jobs (not) running at NIKHEF

- ... this is clear from the ATLAS production monitoring interface which is used by the production shifters:
 - <http://dashb-atlas-prodsys-test.cern.ch/dashboard/request.py/overview?grouping=site&start-date=2008-05-14%2000:00:00&end-date=2008-05-15%2023:59:59&grouping=grid>
- For example the 14th and 15th of May, there is `no jobs in NIKHEF and the site is shown as 'no activity', though it is not turned to red, since the site is red when the failure rate is high.
 - The few jobs which were submitted all failed:
 - <http://dashb-atlas-prodsys-test.cern.ch/dashboard/request.py/errors-detail?&status=failure&site=NIKHEF-ELPROD&end-date=2008-05-15%2023:59:59&start-date=2008-05-14%2000:00:00&maxN=4&grouping=task>
- But since ATLAS is using the panda pilot system, if sanity check of pilot are not ok , the site might be not turned to red since the pilot just does not pull real jobs, so there is no evidence of failure of the application.
- I'll ask Benjamin, whether there is any kind of alarm for the shifters when the site does not process jobs , or they just do not care and redistribute the load among other sites.

Experiments

- As this report is based on the minutes of the daily con-call, it tends to focus on (operational) issues – rather than on performance metrics
- Grosso modo* the work on is being carried out
- Whereas, during the day, the people can be prepared and working during data-taking?
 - Re-processing is scheduled
- Add at least a few dashboards(?) – even if Tier0 processing largely de-couples accelerator operation from first pass processing & data export, is it *entirely* irrelevant?

```
111 CERN SL
LHC Run 0000 updated: 27-05-08 08:24:04
COOLDOWN: INJECTION NOMINAL
T(average)/K:
SECTOR arc 1ss dfb
5-6 : 1.86 3.47 4.13

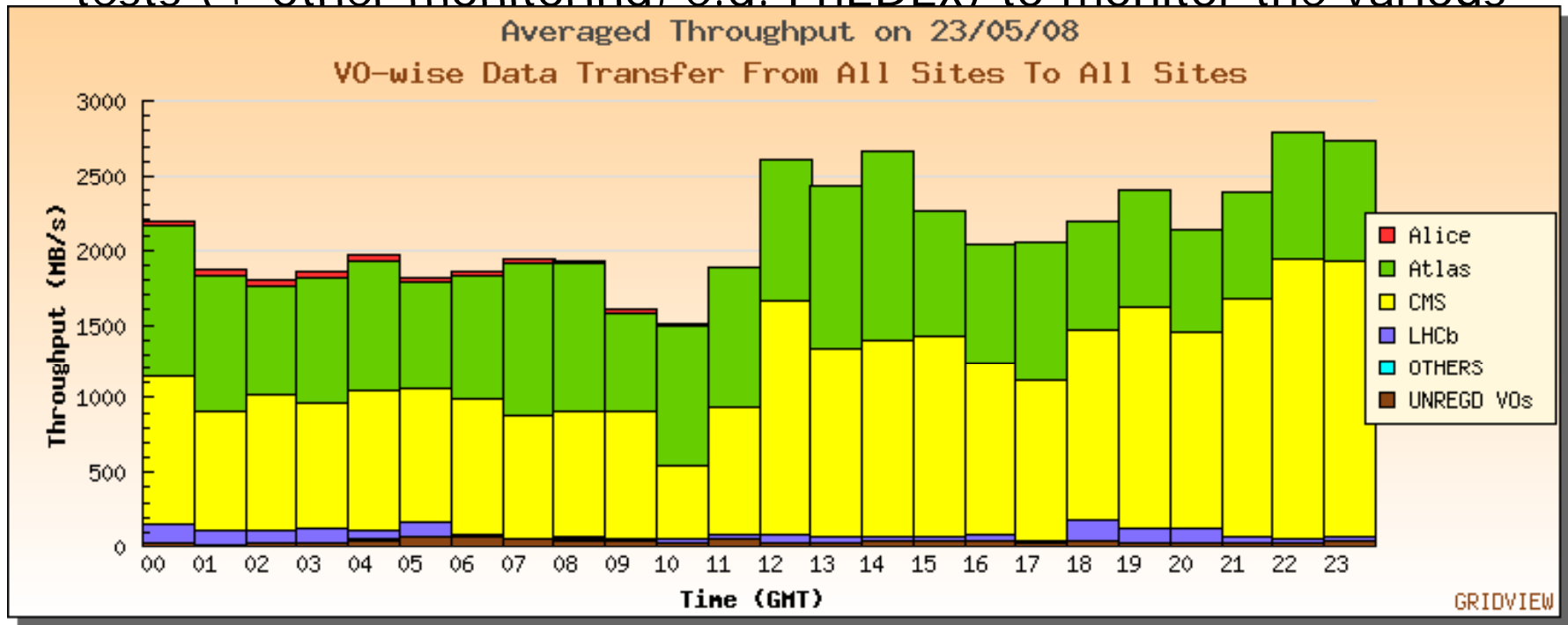
LOCATION CRYO-START CRYO-MAINTAIN
arc56 OK OK
msr5 OK OK
msl6 OK OK
itr5 OK OK

Comments 26-05-08 01:17 :
Sector 56: powering tests ongoing...
Sector 78: powering test starting...

LHC OPERATION in CCC : 77600, 70480
```

CCRC '08 – How Does it Work?

- Experiment “shiffters” use Dashboards, experiment-specific SAM-tests (+ other monitoring, e.g. PhEDEX) to monitor the various



the above → fix [+ MB report]

- With time, increase automation, decrease eye-balling

CCRC '08 – Areas of Opportunity

- Tier2s: MC well run in, distributed analysis still to be scaled up to (much) larger numbers of users
- Tier1s: data transfers (T0-T1, T1-T1, T1-T2, T2-T1) now well debugged and working sufficiently well (most of the time...); reprocessing still needs to be fully demonstrated for ATLAS (includes conditions!!!)
- **Tier0: data / storage management services still suffer from load / stability problems. These will have to be carefully watched during initial data taking. Only a very few (configuration?) changes are now feasible before data taking...**

Summary

- CCRC '08 post-mortem: 12/13 June; LHCC “mini-review” 1 July
- ☺ **Overall, CCRC '08 – both February and May runs – have been largely successful in their goals**
- ↳ **Must assume readiness for data taking in July – according to LHC commissioning schedule**
- This only leaves June for some final – **minor** – configuration changes (which by definition will not have been tested in CCRC '08...)
- **The number of such changes must be kept to the absolute minimum – integrated over all services – and must be fully motivated!**
- We will – inevitably – have to live with some limitations and “features”
- These should be well documented – together with any work-arounds