Comments on VO SAM Results – March 2009

ALICE

SAM Report for ALICE (March 2009). In general March has been a good period for all T1 sites, only some issues to comment:

FR-CCIN2P3

01/03: downtime: due to problems on the local batch system as it can be seen in SAM. The problem in fact was coming from February and already explained at that time.

25-03: downtime: again the same issues observed at the beginning of the month with the local batch system

SARA

04/03: downtime: listmatch problems associated to the local bat h system which was suffering of some issues. I attached you the error message which can also be applicable to the errors observed in France (FR-CCIN2P3).

22/02: unknown: It is difficult to figure out the problem because SAM is not giving any report for that day. So I went to check the status for LHCb and ATLAS and in fact both experiments are reporting similar issues that day. It seems therefore that the local queue was unavailable during that period, although the site did not announced any downtime at that moment.

```
Event: Transfer
- Arrived                    =    Wed Mar  4 02:15:32 2009 CET
- Dest host                  =    unavailable
- Dest instance              =    /var/glite/logmonitor/CondorG.log/CondorG.1236107731.log
- Dest jobid                 =    unavailable
- Destination                =    LRMS
- Host                       =    wms209.cern.ch
- Reason                     =    22 the job manager failed to create an internal script argument
file
- Result                     =    FAIL
- Source                     =    LogMonitor
- Src instance               =    unique
- Timestamp                  =    Wed Mar  4 02:15:32 2009 CET
- User                       =    /DC=ch/DC=cern/OU=Organic
Units/OU=Users/CN=pmendez/CN=477458/CN=Patricia Mendez Lorenzo/CN=proxy/CN=proxy
        ---
Event: Done
- Arrived                    =    Wed Mar  4 02:15:39 2009 CET
- Exit code                  =    1
- Host                       =    wms209.cern.ch
- Reason                     =    Job got an error while in the CondorG queue.
- Source                     =    LogMonitor
- Src instance               =    unique
- Status code                =    FAILED
- Timestamp                  =    Wed Mar  4 02:15:39 2009 CET
- User                       =    /DC=ch/DC=cern/OU=Organic
Units/OU=Users/CN=pmendez/CN=477458/CN=Patricia Mendez Lorenzo/CN=proxy/CN=proxy
        ---
```

ATLAS

CERN 95% OK

CA 97% OK

DE OK, but SAM test problems: storage endpoint changed, updateded after few days. Site was ok from the point of view of the availability

ES 100% OK

FR 98% OK

INFN 99% OK (availability ~75%, few days of downtime)

NDGF 60%: problems in the SRM tape endpoints: intermittent timeouts. Problem solved mid April

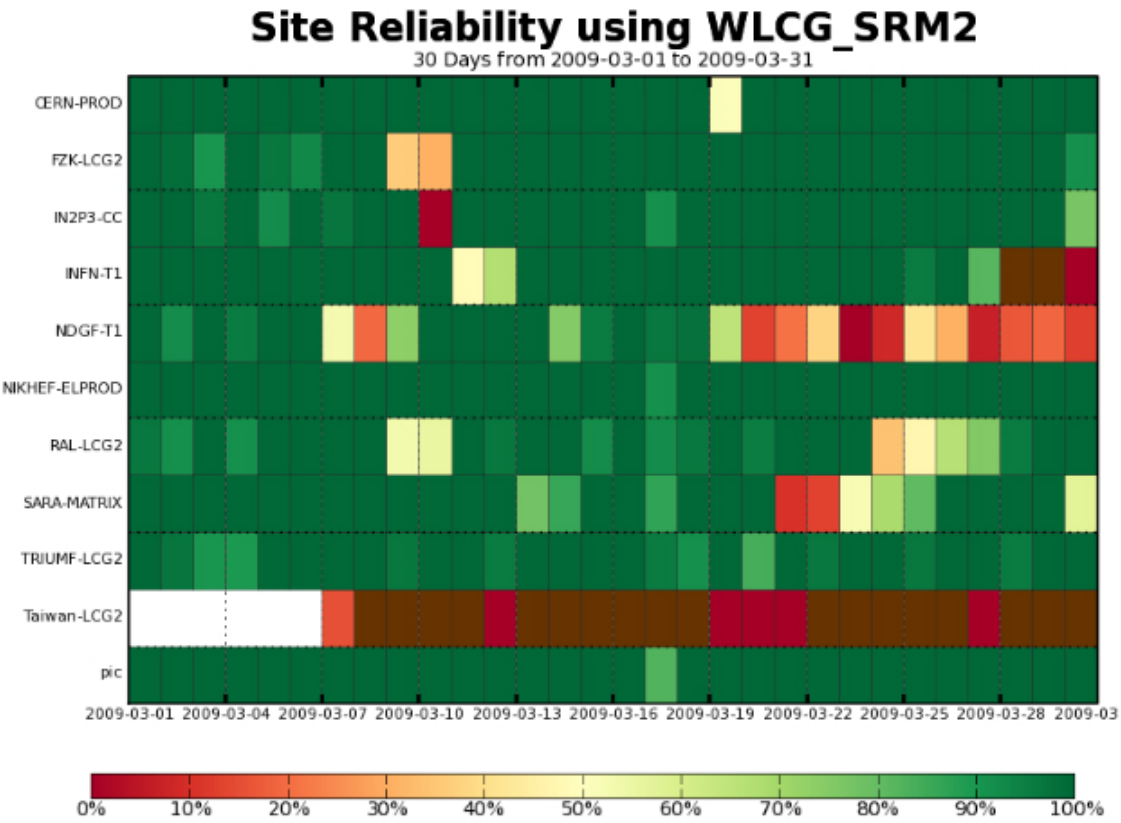
NL 89%: few CE job submission problems

TW ...

UK 93% OK

US still not measurable for naming convention mismatch

Attached here few plots that we can get from the sam atlas dashboard **http://dashb-atlas-sam/dashboard/request.py/historicalsmryview**



CMS

CERN-PROD

Between 10/3 and 12/3, the SRMv2 test trying to copy a file from a UI to srm-cms.cern.ch failed ~50% of times with the error:

[SE][Mkdir] httpg://srm-cms.cern.ch:8443/srm/managerv2: CGSI-gSOAP: Could not open connection ! lcg_cp: Communication error on send

Strangely enough, there is no mention of it in the WLCG reports or in the GOCDB.

FZK

SRMv2 instabilities on the 10/3 and 17/3. They might have been related to a very high load caused by massive pool-to-pool transfers and long staging queues due to a metadata problem in dCache.

PIC

OK.

IN2P3-CC

The CEs were not working from 10/3 to 13/3. A power outage was scheduled for the 10-11/3, but apparently the CEs did not fully recover for some time after the end of the scheduled downtime.

INFN-T1

OK

ASGC

Down due to the fire.

RAL

On 9-10/3, lcg-cp errors on SRMv2. The error logging of lcg-cp is insufficient to determine the cause, but they might be related to an unscheduled downtime affecting the site DNS.

FNAL

OK.

LHCb

General remarks

There is a long gray period from 11th to 16th of March commonly to all T1's. This is because SAM suite was not submitting because the WMS were sick. (Remedy ticket open, patch applied).

In the rest of the failures the CE sensor was the one failing (not SE issues). This is true for all T1s. Site by site:

CERN

CERN was failing some tests the 13th and 14th and 19th-20th.

A close investigation on historical tests show that the js (job submission test) was failing there because the user proxy expired after 12 hours. Most likely SAM jobs were queued in the queue and did not manage to start running in duly time before the proxy expired

RAL

1st of March all day off. User proxy expired after 12 hours. Site busy.

11-16 gray zone because of CERN WMS. We had a failure the 16 with the error: X509 proxy not found (on the WMS). It managed to work few hours later.

The 24th and 25th no jobs submitted, RAL was in Downtime.

DE-KIT

All OK apart the period WMS was not available. I do not understand with Gridview report gray bins the 30 and 31 having valid test submitted and run through GridKA

INFN-T1

fine apart the 11-16 gray zone. See RAL.

IN2p3-CC

only occasional failures with the CE sensor test js (job submission) with Broker Helper error (listr match problem). The site was most likely wrongly publishing after the scheduled intervention they had until the 11th of March. Then the reliability, taking into account the last valid test (failed) was compromised seriously at IN2p3 in the periodo w/o data available.

PIC

between the 5th and 6th some failures managing to get jobs running (user proxy expired - within the 12 hours validity). The same the 17th and 18th. List match problems the 19th of March.

SARA

CE tests failing again. (SARA is the NL-T1 site serving the Storage and other Grid services) while NIKHEF is the once serving CE (that has been demonstrated perfect BTW). We had problems in submitting jobs to ce-gina ce-lisa with error: "Unspecified Grid Manager Error" that usually means a wrong configuration of the local LRMS. I insist however to merge the two sites in just one conceptually coherent site: NL-T1 and consider CE from NIKHEF and SE from SARA. This problem with the CE at SARA indeed should not affect the overall site reliability for the NL-T1.

RAL

For RAL the quick summary is:

1) The main dips in availability almost all correlate with times of entries (scheduled and unscheduled) in the GOC DB. (see 3. Below)

2) Atlas see quite a lot of cases where one or two SAM tests fail in a day leading to high but not 100% availability.

3) LHCb have a couple of significant cases where they were the only one of the four VOs failing tests, but we had not logged any significant problem.

4) A power failure at RAL hit the whole service on 24/3 and the effects lasted until 25th. This is visible in tests of all VO's.

OPS

```
7th. Single or double failure across all CEs early hours of morning
(Saturday). Looks like problem at CERN end.
```

ALICE

```
5/6th: Upgrade of Castor 'Gen' instance on 5th, but found configuration
problems afterwards (on 6th)
```

ATLAS

9/10th DNS problems.

26th: CE SAM tests failing. Checks showed the CEs to be working fine, but it took some time for things to settle down after the Tier1 was off air for over 24 hours. Quite a lot of the time for these failures

Gridview reports a "n/a".

Other Days: Lots of days with only ninety-something percent success. Usually failures of the SRMv2 test. There are a variety of reasons SRM timeouts usually due to Castor and BigID problems in Oracle.

CMS

3rd. Castor upgrade (scheduled).

4th. Problems in castor (LSF) for CMS following upgrade.

9/10th DNS problems.

23rd Two nodes in Oracle RAC failed. Failover didn't work.

LHCb

1st. Don't know cause of this. (Was confined to LHCb V0.The same CEs as failed for LHCb were passing SAM tests for other VOs.)

2nd. Castor upgrade (scheduled).

12 - 15th. Lots of n/a. Some overlap with same seen at other Tier1s for this time.

16th Most of the 16th shows as 'n/a' in GridView. After this the same CEs as failed for LHCb were passing SAM tests for other VOs.

NL-T1

ALICE

plots seem to be based only on SARA, Nikhef CE not included.

ATLAS and LHCb

the problems around the 10th thru 13th is due to the "hot file issue", ATLAS maxing out internal network bandwidth causing various random things to time out.
the problems around the 22nd : the system was working, just very slow. Problem caused by Life Sciences Grid user accessing the SRM in an "unusually effective fashion".

NDGF

The NDGF ATLAS tests appear to be wrong: only one CE was having problems (a scheduled maintenance gone wrong), but the rest were mostly OK. Probably some issue with AND of SAM tests, we'll investigate. Other glitches were caused by various maintenance downtimes announced on a too short notice.

Fron A.Di Salvo: https://gus.fzk.de/ws/ticket_info.php?ticket=47922

NDGF-T1 ATLAS SAM tests failing: copyback from TAPE problems since 15th of March