Table of Contents

WLCG Operations Coordination Minutes, April 4, 2024

Highlights

- ATLAS proposal for downtime

- Preliminary agenda for OPS session during WLCG WS in May
    - Also see this discussion

- UMD-5 plans for EL9 products

Agenda

https://indico.cern.ch/event/1401087/⧉

Attendance

- local: Alexey (CRIC), Julia (WLCG), Maarten (ALICE + WLCG)
- remote: Andreu (ATLAS + PIC), Brian (RAL), Christoph (CMS), David C (Technion), David M (FNAL), Doug (BNL), Federico (LHCb), Giuseppe (CMS), Jyoti (RAL), Mario (ATLAS), Matt (Lancaster + GridPP), Ofer (BNL), Panos (WLCG), Petr (ATLAS + Prague), Priscilla (ATLAS), Rod (ATLAS + LRZ-LMU), Ron (NLT1), Sean (ZA_CHPC), Stephan (CMS), Thomas (DESY)
- apologies:

Operations News

- the next meeting is planned for May 2

Special topics

ATLAS proposal for downtime

please see the presentation⧉

Discussion

- Julia:
    - the described problem exists for all sites and experiments
    - is blacklisting functionality not sufficient to deal with it?
- Priscilla:
    - that functionality gets input from the GOCDB or manually
- Julia:
    - site admins may not know about high failure rates, while the ATLAS ops team do
    - site admins may assume all is fine when the SAM tests are OK
- Priscilla:
    - we discuss such matters in tickets and site admins thus know
    - we would like to have a policy to let them declare outages when a partial problem with a storage service gives us a much bigger problem

- ♦ if we just blacklist the site, that may not be appreciated
- Julia:
  - ♦ it appears we need to determine who puts the site in downtime
  - ♦ mind there is no downtime per VO in the GOCDB
  - ♦ ATLAS can do that in CRIC, though

- Stephan:
  - ♦ in CMS there is a 15-minute granularity of checks for source, destination or unspecified transfer problems
  - ♦ bad endpoints get removed and analyzed
  - ♦ a site can announce an outage and the experiment reacts
    - ◊ no ticket would be opened for that site
  - ♦ CMS do not need a 20% downtime to be declared: a site cannot know the impact, while the experiment does
  - ♦ on a daily basis it is decided which sites (not) to use for which workflows
- Rod:
  - ♦ mind it is not the site, but the service: just the ATLAS storage
  - ♦ if 20% is unavailable, an outage should be declared because the expected QoS is not being delivered
- Stephan:
  - ♦ CMS is informed of service outages, handled differently from failures:
    - ◊ scheduled work may be adjusted
    - ◊ tickets will be avoided
  - ♦ the experiment usually is faster than the site to determine the impact
  - ♦ sometimes the problem may rather be the network instead

- *Stephan then demonstrates the CMS test transfer matrix dashboard*

- Stephan:
  - ♦ Grafana does not give all the information
  - ♦ every 15 minutes the FTS logs are mined and the results are injected
  - ♦ from the matrix it is concluded which sources and destinations to remove
  - ♦ the lack of good error messages from `WebDAV` has made this harder, compared to how it worked for `GridFTP`
- Mario:
  - ♦ that is consistent with our proposal: a fixed algorithm, no discussion
  - ♦ it then is about how to implement it
- Stephan:
  - ♦ we do not want to rely on admins declaring downtimes
- Priscilla:
  - ♦ there are 2 options:
    1. the site does the GOCDB and we follow up
    2. we do it ourselves

- Rod:
  - ♦ for ATLAS it is a problem if a federated site has 1/3 of its storage down: what would be the threshold for CMS?
- Stephan:
  - ♦ if the site fails the SAM tests, it is taken out
  - ♦ if the majority of the storage is accessible, we do not need an outage
- Andreu:
  - ♦ if CMS stops using a site, does that impact its A/R?
- Stephan:
  - ♦ only the SAM tests determine the A/R: not perfect, but not a big problem either

- Thomas:
    - a site can declare a downtime, a VO can decide to use the site or not
    - might an intermediary state between "`at risk`" and "`outage`" be useful?
- Stephan:
    - on an at-risk downtime, we manually disable the site or not
- Julia:
    - the A/R is determined from basic tests
    - a VO needs to decide when to blacklist a site
- Federico:
    - there seem to be 2 problems:
        1. a federated site should declare downtimes properly
        2. should the use of resources by a VO be reflected back in the GOCDB? No!

- Priscilla:
    - in the CMS approach, decisions are made by the VO, not the GOCDB
    - "`at risk`" downtimes currently are not used in ATLAS
    - when a site is blacklisted, is it notified?
- Stephan:
    - on blacklisting, a ticket is opened and some CMS teams may be informed

- Thomas:
    - if a RAID controller breaks on a Friday, its replacement will only arrive the next Tuesday and some pools are affected: should an outage be declared?
- Rod:
    - if at least 20% is affected
- Thomas:
    - mind that new pool hosts attract lots of new files, which may lead to lots of jobs wanting just those files in the future
- Rod:
    - we propose a simple 20% threshold irrespective of use cases
- Stephan:
    - CMS would appreciate an at-risk downtime in such cases

- Maarten:
    - we seem to be converging on the experiments controlling such matters?
    - in ALICE there is hardly any reliance on downtime information: everything is automatic, there are no operations shifters
    - when things keep failing, an expert will take a look eventually
    - downtime announcements are appreciated to help:
        - ◊ the understanding of failures
        - ◊ the planning of big operations

- Alexey:
    - ATLAS already has a tool for automatic blacklisting, with various input sources:
        - ◊ downtimes, quota issues, and more
    - it could be extended with transfer matrix information
- Priscilla:
    - quite some work may be needed
    - we will report on these matters within ATLAS and follow up from there

Preliminary agenda for OPS session during WLCG WS in May

please see the document☐ attached to the agenda

Discussion

- Julia:
  - ♦ we already decided **not** to include a discussion on GPU resources, as they are not for the short term and there still is a lot to be done

- Federico:
  - ♦ what about ARM? which sites are interested?
- Julia:
  - ♦ we can look into that indeed
- Maarten:
  - ♦ that topic could even feature in a plenary session

Middleware News

- Useful Links
  - ♦ WLCGBaselineTable
- Baselines/News
  - ♦ The status of the new UMD service has been presented in the EGI OMB meeting⬠ on March 21.
    - ◊ UMD-4 for CentOS 7 should again be updatable as of the 2nd week of April
    - ◊ UMD-5 for EL9 *as designed* should become available by the 2nd half of May
      - · A temporary, less sophisticated version will be looked into!

Tier 0 News

Tier 1 Feedback

Tier 2 Feedback

Experiments Reports

ALICE

- Activity levels fluctuated between very high and low (recent days)
- Raw data replication to CERN CTA and the T1 sites has finished
- No **major** problems
  - ♦ Upgrade of `myproxy.cern.ch` to EL9 is paused
    - ◊ One of the 2 hosts behind the alias was upgraded to EL9
    - ◊ ALICE VOboxes at many sites could not be served by that host
    - ◊ Almost all are OK now after upgrades of `myproxy` and `Globus` packages
    - ◊ Two VOboxes failed due to obsolete syntax in their host certificate DNs
      - · The `'host/'` convention appears to be unsupported on EL9
      - · One case was fixed by a new certificate
    - ◊ A few tickets are still open

ATLAS

- Generally smooth running
  - ♦ Good job mix with lots of Full Sim
  - ♦ 720k average slots to 900k peak slots in the last 30 days

- ♦ 3PB average daily transferred in the last 30 days
- ♦ Large scale 60 PB data consolidation to tape (average rate of 1PB/day)
- Issue with Vega HPC site
  - ♦ Space issue at the arc-ce level
  - ♦ The info that the job ended in the arc-ce did not go back to Panda.
  - ♦ For some days mid march vega reported running o(1 million) jobs but this was not real as no job was running in Vega.
  - ♦ Problem corrected
  - ♦ Vega monitoring repainted 20:00 13.03   20:00 19.03
- Modified curl 8.7.1 without line length limitation on EL9 on ALRB
  - ♦ We noticed the issue as sometimes requests with very large payload were failing on production when trying to bump the cURL version.
  - ♦ Custom wrapping
  - ♦ Needed for EVNTMerge at TOKYO
  - ♦ https://github.com/curl/curl/issues/11431
  - ♦ Any way to properly package this for EL9?
- Automatic blacklisting of sites (mentioned in dedicated report)

Discussion

- Maarten:
  - ♦ while `curl` has not been patched upstream and released in (RH)EL, ATLAS will need to keep bringing its own version through CVMFS

CMS

- overall smooth running, no major issues
  - ♦ CPU usage between 340k and 570k cores
  - ♦ reduced analysis activity with production/analysis split of about 5:1
  - ♦ Run 3 production followed by Run 2 ultra-legacy Monte Carlo main activities
- still waiting on python3 version/port of HammerCloud
- three remaining DPM sites need to migrate their storage
- first two tape endpoints migrated to REST interface
- token migration progressing steadily
  - ♦ two-third of the dCache sites ready (plus all StoRM and all but one native xrootd sites)
  - ♦ waiting on EOS (scitoken library) to be fully token-ready
- we will ask sites to remove x86-64-v1 microarchitecture worker nodes with the end of SL7, i.e. June 30, for CMS

Discussion

- Maarten:
  - ♦ how much is CMS affected by the lack of Python-3 support in HC?
- Stephan:
  - ♦ we cannot use Run-3 CMSSW releases in HC tests
    - ◊ cannot test new features
    - ◊ AAA tests are also impacted
- Julia:
  - ♦ we will ask the HC service manager to give an update in the next meeting

- Stephan:
  - ♦ stopping `x86_64-v1` WNs: could that be discussed at the workshop?
- Maarten:

- ◆ ALICE jobs match CPU features to workloads, but it would be better if very old HW were no longer proposed by sites
  - ◆ it would allow SW builds to take more advantage of modern CPU features
- Alexey:
  - ◆ the ATLAS pilot can exclude `v1`
- Julia:
  - ◆ this looks a good topic for the workshop indeed

LHCb

- smooth running
- Data Challenge:
  - ◆ more tests with INFN-CNAF
  - ◆ "DC" with Beijing proto-Tier1:
    - ◊ Staging: target was 0.63GB/s, and 1.12GB/s was achieved. Efficiency was very good as well.
- ETF tests updated in pre-production, still fixing tests for a couple of sites before going into production (same as previous month)
- Moved to EL9-compatible Apptainer image, and discovered a few sites with pre-`x86_64-v2`

Task Forces and Working Groups

Accounting TF

- NTR

Migration from DPM to the alternative solutions

- 34 out of 42 WLCG sites either migrated or decommissioned their storage. 7 are in progress.
- 3 which have been migrated still need to enable SRR
- 22 sites selected dCache, 9 EOS, 4 xrootd/CEPHFS

Information System Evolution TF

- NTR

IPv6 Validation and Deployment TF

Detailed status here.

Monitoring

Network Throughput WG

WG for Transition to Tokens and Globus Retirement

- the current state of affairs was presented in the GDB⬀ on March 27

Action list

Creation date	Description	Responsible	Status	Comments

Specific actions for experiments

Creation date	Description	Affected VO	Affected TF/WG	Deadline	Completion	Comments

Specific actions for sites

Creation date	Description	Affected VO	Affected TF/WG	Deadline	Completion	Comments

AOB

- Next meeting will take place on the 2nd of May

This topic: LCG > WLCGOpsMinutes240404
Topic revision: r15 - 2024-04-09 - MaartenLitmaath