

Technology Investigations Update

Brian Bockelman
December 17, 2014

Where is Brian Today?

- It's been a travel-heavy month-and-a-half. Many apologies for un- and under-answered emails.
- Last week was very interesting:
 - Helped lead the “European HTCondor Site Admins” meeting at CERN.
 - 49 registered participants; about 40 in-room and 10 via videoconference. About 85% WLCG-related participants.
 - 22 different sites were represented.
 - Part of a sea change within HEP/WLCG; as the free version of PBS becomes less-maintained, many sites are looking for alternate batch systems.
 - Within big sites- about 3 Tier-1s and CERN are considering switching.
 - View this as helping spread the “HTC philosophy”.

HTCondor-CE & Info Services

- Over the past 2 months, two new releases of the info services made. These cover our base requirements.
 - The service is deployed at GOC, but I am still helping shake out a few bugs.
 - I expect this service will be useful for gWMS / APF / AGIS developers in the January / February timeframe.
- SAMv3 can discover and submit to HTCondor-CE sites. At least two CMS T2s have removed GRAM!
 - Since October, several sites have progressed in their deploy (more CEs, glideins go to HTCondor-CE by default, RSV monitoring, SAMv3 etc). *However* only one additional WLCG Tier-2 has started the transition since then.
 - All CMS T1/T2 sites have an install now; we are behind in supporting the ATLAS deploy.

Scalability Tests

- Scale test activity wraps up at the end of calendar year to release resources (hardware and personnel) to other tasks.
- Highlights of the scale test activity:
 - Achieved goal for idle jobs in single schedd (goal was 200K; hit 500K).
 - Achieved goal for overall number of running jobs (goal was 200K; hit 200K).
 - Achieved goal for running jobs in single schedd (goal was 40K; hit 100K!).
 - Did not achieve the goal for running jobs in single HTCondor-CE (goal was 20K; hit 12K).
 - I consider this acceptable because the limit was the batch system behind the HTCondor-CE, not the CE itself. Largest production GRAM CE in OSG is currently 10K.
 - Many recovery modes successfully (restart collector, restart schedd, etc).
 - Made significant progress on reducing number of TCP connections required per glidein.
 - Result was several development requests for glideinWMS.
- I hope to have another, shorter round of scale tests in approximately 6 months. Goals will be to validate subsequent HTCondor and glideinWMS improvements as well as retry the HTCondor-CE goal.

OASIS Software & Service

- BNL is running an instance of OASIS v2 locally for USATLAS Tier-3 sites.
 - Progress toward deployment at GOC is on hold as developer is gone for the holidays.
- External repos are now live! Initial users are IceCube, LSST, and Fermilab VOs.
 - Required changes at the Stratum-0, OIM registration, and Stratum-1s.
Thanks to all involved!
 - You give a mouse a cookie... Fermilab has asked about whether a VO could have multiple repositories. Informally discussing this (not clear if it's needed); if it becomes a request, I'll put it in the stakeholder request system.
 - As part of the Fermilab SCD reorganization, the responsibility for the FNAL CVMFS hosts (repositories and Stratum-1) is transitioning to Tony Tiradani.

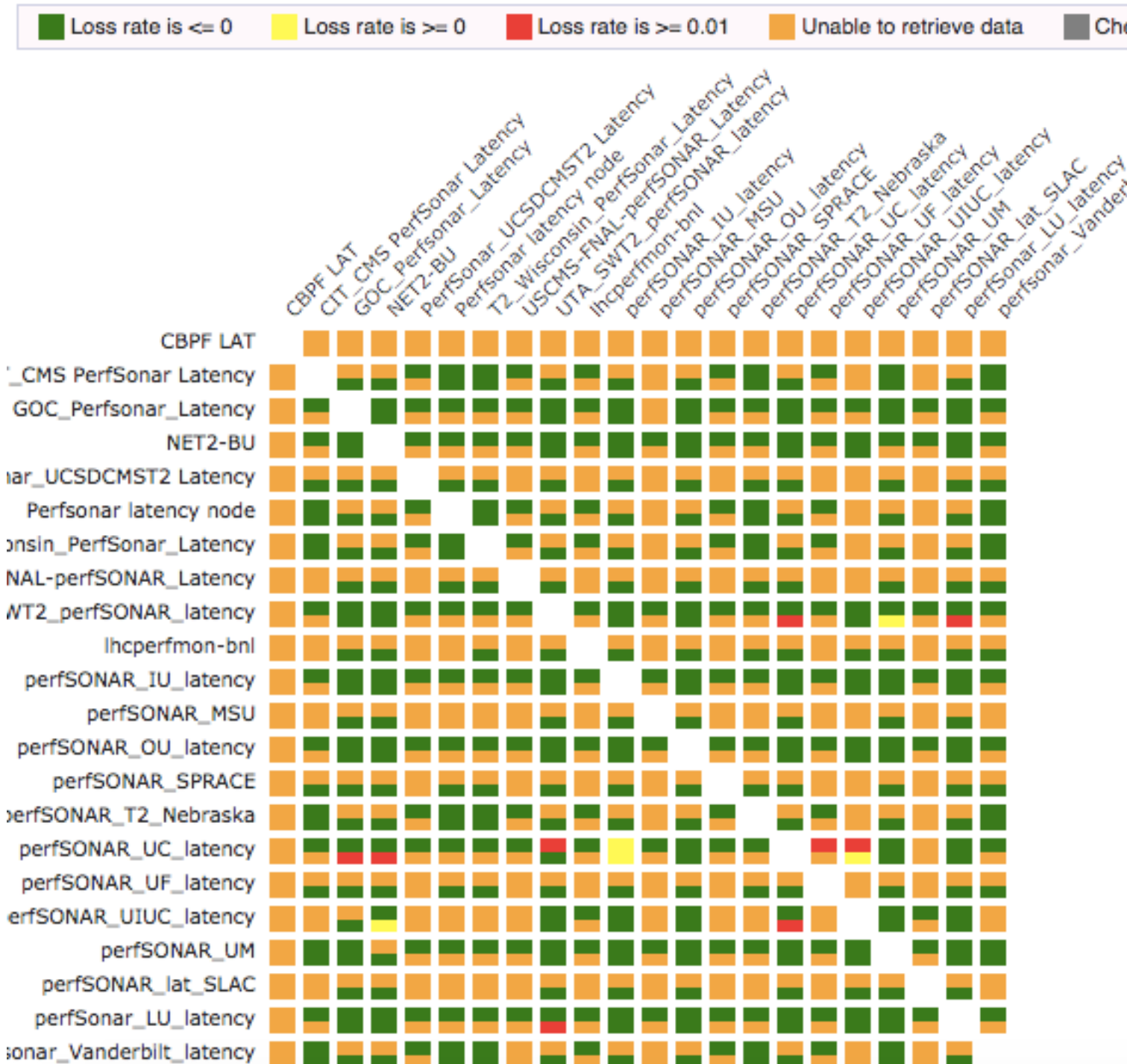
OASIS & External Repos

- We've had a long-standing goal of assisting VOs who want to access repositories signed by the egi.eu key.
 - **Issue #1:** We need a technical mechanism to revoke access to these repositories during security events.
 - We believe we have a solution for this, based on the repository blanking process in place for osg.org repos.
 - Unfortunately, the ITB hardware needed for testing this process has been in use for OASIS v2. While the OASISv2 dev is gone on holidays, we will “steal” this back for ITB testing.
 - **Issue #2:** We need an improved technical mechanism to manage OASIS configuration.
 - For example, we don't control the egi.eu keys, meaning frequent updates on their side would result in frequent RPM release **on every worker node** on ours!
 - Agreement at Blueprint:
 - (Near-term): Ship egi.eu repos once issue #1 is resolved.
 - (Short-term): Switch to a non-RPM-based automatic update mechanism for CVMFS configs.
 - (Long-term): Investigate the possibility of “config-free” CVMFS; pilots would bring along their own configuration.
- I predict issue #1 will be solved by next update and we will be ready to ship the RPMs outlined in #2.

Network Monitoring

- Centralized network repository is progressing nicely but slowly.
 - Recall this is fed by a RSV probe.
- Matrix to right is generated solely from the central repo. Now have several weeks of data in the test instance.
- Not without problems - a few performance issues solved, a few to go. A few times have had to go back to ESN Net devs (Andy Lake) for help.
- Tomorrow is the final ITB update, then will cut a production version.
- After the “first release”, plan on whittling down unmet requirements.
 - Let’s not snatch defeat from the jaws of victory - I’m concerned we’re putting too many requirements ahead of “Version 1.0”.

OSG Grid Operations Center Test Mesh Config - GOC Latency Tests



Authorization

- A secondary meeting last week at CERN was about authorization futures in EGI.
 - The current software, Argus, mostly works but key components are unsupported.
 - I attended to educate the room about how things work in OSG with GUMS (both in terms of support and technical functionality).
- Conclusion of workshop: Community of devs would provide best-effort support of Argus components. There was a strong desire to use this as a base for future work.
- **My** observations:
 - There's still a strong emphasis on new features over solving operational problems - even when there's no effort available to do either!
 - Comparatively, OSG has made extraordinary progress in simplifying the authorization model. "Sites trust VOs, VOs manage users."
 - This allows sites to operate production-quality legacy infrastructure independent of innovations done by sites.
 - This concept appears to be a non-starter in EGI. Accordingly, the existing EGI authorization deploys - which are much more complex than GUMS - are not complex enough as VOs start considering user-friendly models like federated identity.
 - OSG is headed in the other direction - i.e., looking to change the default documentation from GUMS to edg-mkgridmap.
 - I personally had not previously appreciated how much the two grids had deviated on the subject (and I'm very happy to be on the "OSG side" of things).

Blueprints

- Beyond the discussion of CVMFS, the most recent Blueprint tried to tackle discussion of “what is a site?” “What services should be on a CE?” “What is the minimal effort required to participate in the OSG?”
- There’s some loose organization appearing (sorted in increasing effort and increasing usability for science):
 - “SSH-endpoint”: Site provides a SSH-login where we can submit jobs to a batch system.
 - “Minimal CE”: Site provides a CE where we can submit jobs to a batch system.
 - “Full CE”: Site additionally provides CVMFS on worker nodes.
 - “Complex site”: Site additionally provides some sort of storage service.
- I think we’ll have many good future discussions on this! I’m sure even my labels are controversial.

Goals for next time

- **Stakeholder requests:** Finish of REQUEST-66 and REQUEST-67 in early January at OSG council meeting.
- **User Data Management:** Looking to organize a phone call in mid-January on this topic.
 - In particular, Derek and I are preparing a demo of one approach for the user support team.
- **HTCondor-CE:** 10 sites in the info services; half of CMS Tier-2s turn off GRAM; add another ATLAS Tier-2 CE.