Table of Contents

Summary of September GDB, September 12, 2018

Agenda

Agenda⬀

Introduction - I. Collier

presentation⬀

GDB in March 2019 will be at CERN

GDB in April 2019 will be colocated with ISGC in Taiwan

Additional meeting: LHCONE/LHCOPN meeting on the 30-31 of October 2019 at Fermilab

HEPIX-WLCG-HSF Technology working group will be launched on Monday 24th. Contact Helge or Bernd to join the mailing list.

HSF Workshop

- The joint WLCG-HSF workshop in Naples worked very well.
- As stated at July GDB, the next workshop will be in North America on the 18th-22nd of March 2019. (Problem is over constrained - other weeks not feasible.)
- Jefferson Lab (VA) and BNL/Stony Brook are being considered, no decision made yet.

Registration and accommodation would be cheaper in JLAB.

Comments: some preferences expressed for JLAB.

Comment: There two other broad community events in the following weeks (HEPiX, ISGC, etc.) participations may be difficult.

GDPR discussion

David Foster: Activities at CERN on GDPR

CERN, being an intergovernmental organisation, is not bound, but it has obligation concerning the GDPR because of the many collaborations with EU.

Main activities at CERN:

- Established a centre of competence, to provide advices to the organization. It has written an Operational Circular, which will come into effect very soon, that gives a code of conduct to CERN staff. The OC requires service owners to register their services in the catalogue; the manager will be responsible of the protection of the personal data collected by the service.
- The responsible of the service has to give information on how the data is treated and retained.

Privacy notices for all the services will be produced.

  - Training: for all the responsible, to help them produce the necessary documents

Main problem to solve are accountability and transparency, i.e. how to make sure individuals can know how and what of their personal information are treated.

Q: How GDPR applies to WLCG?

A: Each single entity is legally responsible for what they do. There's no need to panic: in case of problems, the regulator in a country, before sending fines, will tell the subject to stop the wrongdoing.

Q: What about the data stored in Indico (participants, authors of presentations...)?

A: It could be claimed that keeping that info forever is for the public interest. And users should be made aware of it and be given the possibility to be removed. Anyway the Public Interest needs to be justified.

Q: GGUS decided to play safe to anonymise all the ticket older than 18 months. Wasn't too much?

A: if they determined they have no ongoing purpose of the data, they did it right, because they are liable.

Dave Kelsey

presentation

AARC has produced guidelines on how to implementing GDPR (link on slide 5)

WLCG has drafted a risk assessment for DPIA. It will be presented to the WLCG MB.

Julia Andreeva

Survey of WLCG services that may need a privacy notice.

A twiki page has been created: https://twiki.cern.ch/twiki/bin/view/LCG/GDPRandWLCG

Work in progress. At the moment there is a list of services which do not expose personal data.

There are also contact people for each experiment that are responsible for updating the data on the twiki.

Comment: putting up a privacy notice is not enough for those services that expose too much info to too many people. Works need to be done to reduce the exposure.

Comment: work is not finished, November may be the good time to review it.

Q: Is a task force needed to help with the effort?

A: for now it's not needed. (But very likely will be later.)

IPv6 Update - A. Sciaba

presentation

Short-term milestones: T2 storage dual-stacked by the end of run 2 (end of 2018)

Changes since March

- ~ +20% in IPv6 to/from CERN (not included LHCOPN/ONE)
- +25% IPv6 endpoints in WLCG

FTS transfers

- CERN dual-stack but not BNL and FNAL (no information yet)
- IPv6 GridFTP between half of the sites (see slides)

T2 status

- 40% of T2 sites have storage on dual stack (55% for CMS, 33% ATLAS)
- Sites asked about their plans throug a GGUS ticket: most sites responsive, very few who declared they will not meet the deadline
    - ♦ Some sites depend on campus infrastructure being IPv6-ready
- USATLAS and USCMS now tracked but directly by experiments
- Leading regions: NDGF, Spain, France, USCMS

Conclusion: no major concern about meeting the deadline for a wide deployment of IPv6 at the end of the year

- Acceleration expected at T2 sites as we approach the deadline

HTCondor Workshop @RAL🔗 - H. Meinhard

presentation🔗

Complementary to yearly HTCondor weeks in Madison

- Last one took place last week @RAL🔗: 54 participants (9 from STFC), 28 affilitations, good mix of experienced and new attendants

38 contributions: 2/3 from HTCondor team (agenda: https://indico.cern.ch/event/733513/🔗)

- Hot topic: accounting
    - ♦ What is needed
    - ♦ What can be provided by HTCondor (not specific to APEL)?
- Very positive feedback from attendees, including HTCondor team: in particular more discussions than during the HTCondor week

Next workshop: JRC of EU at Ispra (Lago Maggiore, Italy)

- Tentative date: September 24-27 2019
- Follow-up discussion on accounting

Also a short Asian HTCondor workshop planned the days before ISGC

Container WG - G.

presentation🔗

4 meetings since January: focus on Singularity

- Main goal: simple command line to run a job in a containerised environment for payload separation

Singularity packaging in EPEL

- Former maintainer not very responsive and adding a lot of unreferenced patches: WLCG had to maintain it own version
- Brian and Dave took over: now following the upstream version with some patches still in PR added when required by WLCG

Bind mount and overlay: not really practical, requiring privileged Singularity on EL7

- "Underlay" proposed by Vincent and implemented by Dave to allow bind-mount in arbitrary locations with an unprivileged Singularity: part of the EPEL version but not yet merged upstream
- No expectation to have the underlay feature in upstream 2.6 as the effort upstream is concentrated on v3, a Golang rewrite, that will include this feature

OCI container technologies: a possible, more mainstream, alternative to Singularity in the feature but not possible until we can get rid of EL6

Image and image distribution: several open questions

- Can we converge on a common base image from which the experiment can build and distribute it in CVMFS
- Can we agree on a common CVMFS repo? Is it desirable?

SW: Estimates of Potential Efficiency Gains - M. Schulz

presentation⬚

Several kind of improvements identified from code profiles and compilation strategies (vectorisation, feedback...) to linker/loader

- Also take into account impact of evolution at the infrastructure level (centralisation, reduction of storage replica...)
- Attempt to classify the effort required for each kind of changes

Figures are best-cases: should be taken with precaution

Storage

- Reduction of sites with disk storage, moving most T2s as cache sites (with no stateful service): potential of saving 40% (~40 FTEs) of the effort currently required to operate WLCG storage
- Reduction of the redundancy at sites (removal of RAID, EC...): probably a marginal impact on job failure rate, based on CERN experience (disk error rate is very low) and may allow to save 30-50% disk space
- Reduction of number of copies at experiment level, based on popularity. Potentially complex if managing a larger hierarchy, probably 15% saving.

Compiler/Code/Build

- Potential for 15-20% CPU improvement: difficulty is that HEP code is very sensitive to compiler switches and require a lot of physics validation
  - ♦ Should try to regain control on our code to avoid the very heavy physics validation that is required today
- Memory management is another area identified with profiling where gains are possible

- Study of HW counters for several applications shows that HEP code is executing 0.8-1.5 instruction per clock cycle compared to 4 for HPC: a potential 100% gain possible at the price of a rewrite of the code
    - ◆ ALICE demonstrated it within its HLT where the GPU-friendly code brought important improvements on CPUs too
- Does not take into account new algorithms, new smaller data formats...

Requirements: work with other communities to learn from them, increase cooperation between engineers and physicists

- Important role for the IT Performance group to liaise with experiments and ensure that cross-experiment communication takes place
- Join WLCG/HSF workshop demonstrated to be a good place for ensuring the information flows between the people in charge of the infrastructure and the DW developers in the experiments

-- IanCollier - 2018-09-13

This topic: LCG > GDBMeetingNotes20180912
Topic revision: r3 - 2018-09-25 - AndreaSciaba