

Table of Contents

Summary of GDB meeting, June 10, 2015 (CERN).....	1
Agenda.....	2
Introduction.....	3
GDB Future - I. Bird.....	5
LHCONE/OPN Update - E. Martelli.....	7
dCache Workshop and Storage Evolution - P. Millar.....	9
Large-scale MC simulation at Helix Nebula - D. Giordano.....	11
report - M. Alandes Pradillo.....	12
T0 Update - H. Meinhard.....	13
Belle2 - T. Hara.....	15
DPHEP Collaboration Workshop - J. Shiers.....	17
Cloud Traceability pre-GDB - I. Collier.....	18
HEPiX Update - H. Meinhard.....	19

Summary of GDB meeting, June 10, 2015 (CERN)

Agenda

<https://indico.cern.ch/event/319748/>

Introduction

Next GDBs

- October meeting: Grid day at HEPiX at BNL
- 2016 dates added to Indico: may be impacted by discussion about GDB future

Future GDB topics on the list:

- Pakiti
- machine/job features
- next generation benchmark
- security and AAI

Next Fall pre-GDBs: dates not yet settle

- multi-day workshops: HTCondor, clouds
- cloud and storage services
- IPv6
- No pre-GDB in October, September to be confirmed in July

ARGUS collaboration meeting last week

- OpenSSL increase of DH cipher keys to 1024 bits may be problematic for pepd
- work started on EL7 and Java 8 support: plan is to be ready with EL7 early Fall
- Issue of ARGUS becoming unresponsive only seen at CERN so far: other sites seeing a problem are asked to report to ARGUS (<https://github.com/argus-authz/>)
- Some effort available from Indigo Datacloud

Open actions

- machine/job features: need more volunteer sites to make progress
 - ◆ Please contact Stefan Roiser
- List of class 2 services: Jeff agreeing to start a list with ATLAS information, other VOs will have to complete with their info
 - ◆ Class 2 services are VO services like VOBOX or storage N2N plugins that are run inside the site and may have a potential security impact
- multi-core accounting: no major change since May, every region/NGI should look at mail from John after last GDB providing the situation per NGI
 - ◆ 4 visible missing "big" regions: China, Asia-Pacific, Italy, Spain. **Urgent action needed**
 - ◆ Others should look too...
 - ◆ Tickets will be open against NGI by John: starting today...
- list of storage protocols: started by W. Bhimji for ATLAS before leaving, other experiments must fill their part
 - ◆ <https://twiki.cern.ch/twiki/bin/view/LCG/WLCGDataAccessProtocolUse>
- HTCondor and BDII: new plugin for the CREAM CE, Share (GLUE2 version of VOVView) to be added to ARC but no plan for VOVViews
- WLCG workshop to be organised by operations' coordination: see OpsCoord report for details

Forthcoming meetings: fortnightly operations' coordination, HEPiX, EGI community forum, Supercomputing

Discussion

- Maria D: Having a testbed for ARGUS can be very beneficial.

- ◆ Michel: Not sure whether it will help in the case of the CERN issue, but in general of course a good idea.
- ◆ Maarten: together with IT-PES, have more concrete ideas now on how to move forward

GDB Future - I. Bird

Background: need to update technical discussions

- LHCC likely to request more formal plan for LHC upgrades, in particular wrt. to the lack of commonalities
- Had TEGs, computer model updates but one-off's
- With our experience, we need to be more flexible and more cost effective
- Need to create WGs and a forum to discuss technical strategy
 - ◆ Principle agreed in Management Board: small team to be set up in order to launch the effort

Current meetings/boards

- CB: meeting of every MoU party
- OB: a subset of CB discussing strategy but in fact mostly hearing report
- MB: policy and management
- Operations: not part of the MoU but created since then
- Architect's forum: stakeholders of AA
- GDB
 - ◆ Evolution from Deployment to Operations reflected by the creation of OpsCoord
 - ◆ Less policy issues (GDB was in charge of proposals)
- Missing a place to drive the technical strategy, except pre-GDB's

Propose to refresh the GDB

- Good time to do it: Michel is well over its term (2 years) and is ready to step down as soon as a replacement has been decided
 - ◆ A good target could be the election of a new chair in September
 - ◆ MoU says that the chairman must come from a site and not from CERN
- discuss and agree directions and details of implementation and deployment
- Sponsor technical developments
- Avoid duplication with other meetings
 - ◆ Drop summary reports from OpsCoord: OpsCoord can raise specific issues when needed
- Useful as a regular meeting of the community: need to keep it

WLCG workshops: useful and opportunity for wider discussion, must be different from GDB and OpsCoord

- Currently 9-monthly: 6-monthly? What is the role of the GDB in this case?
- Need to have time to write summary/position papers during workshops
- Should be agenda driven
- Make sure OSG and EGI are represented

Discussion

- Jeremy Coles: sounds very much like pre-GDBs which have been difficult to organise.
 - ◆ Ian B: interest in GDB has faded away, room was fully packed in the past
 - ◆ Ian B: optimistic that with the right preparation technical topics of relevance to WLCG would attract a sufficient number of people
- Eric Lancon: Support Ian's proposal, experiments need a clear strategy to be defined for the next years. But proposal is addressing only part of the problem body needed that drives the technical strategy.
 - ◆ Ian B: Agree, it's this body that sets the topics for GDB.
 - ◆ Eric: avoid duplications, make clear choices. Also need site validation because of subsequent implications.

- Ian Collier: at its best, GDB does set technical directions (as it happened with the cloud traceability working group for example), but this needs to be reinforced.
- Jeff Templon: Support Ian Collier's statement; will need to carefully consider the relationship of the steering group and the GDB chair. Would propose that the GDB is chaired by the steering group.
 - ◆ Alessandra Forti: support.
- Michel: would have been very happy to run GDB in a more collegiate fashion; not much input received from the MB. One option is not to have a permanent chair, but to rotate among the members of the steering group.
 - ◆ Ian: a lot of scope for useful work.
 - ◆ Michel: "topical WGs" are important (this is where we attract other people) for make the GDB lively. Not sure we can have a permanent body in charge of strategic directions: probably need a cycle where we have TEG-like process producing new ideas, some time to digest them, a new TEG-cyle... We are probably at the time to do it: we "digested" the previous TEG ideas... Also need to clarify relationship of GDB and steering group to the Management Board.
- Michel, Jeff and others: doubt that MB can drive this process. Too technical, not the right membership.
- Oxana Smirnova: Glue2, Glexec are already examples of topics discussed by a few experts that need broad discussions.
 - ◆ Ian B: Exactly this is the reason to have GDB
- Jeremy Coles: How to proceed?
 - ◆ Ian B: I will get a few people together and start. Those people should be from experiments. Pre GDB discussions are very good, we need to put that into a framework, priorities etc. First start with a short strategy document setting directions: could be based on the computing model update.
- Jeff: suggest a model based around how it went with wLCG accounting. Management decided "we need accounting", working it out was done in the GDB, lots of fierce discussions on what would be collected, how interpreted, how implemented, limitations, requirements etc.
 - ◆ Ian B: That was the way when significant progress was made.
- Tendency is that the steering group should own the GDB

LHCONE/OPN Update - E. Martelli

Meeting last week in Berkeley

LHCOPN

- New sites with 10G connectivity: KISTI, TRIUMF
- T1-T1 links are being shut down and moved to LHCONE
 - ◆ Still need to test this backup link at several sites
 - ◆ Proposal from FNAL to remove it from LHCOPN: not yet agreed, some sites with a better connectivity to LHCOPN
- IPv6 activated at CNAF

LHCONE

- CERN has 100G to ESnet and GEANT
- 5 US T2 now connected to ESnet
- Polish NREN will soon join LHCONE
- JANET and Brazil joined LHCONE
- Slow progress in Asia
- IPv6 being deployed: already 6 sites connected
- AUP was finalized at February meeting
 - ◆ Existing sites will have to acknowledge it: an email to be sent soon
- Belle2 and Auger asked to join LHCONE
 - ◆ Most sites are already connected to LHCONE or are WLCG sites

perfSonar

- LHCONE MadDash is getting better
- New service based on datastore + MadDash + OMD being deployed: new perfSonar version expected after the summer

P2P service is still a prototype: circuit awareness demonstrated with PhEDEx

- No guaranteed bandwidth yet
- No scalable L3 routing solution through P2P circuits identified yet

Next meeting end of October in Amsterdam

Discussion

- Ian B: nice that Auger, BelleII joined LHCONE but other experiments are asking/will ask. What is the process to have other experiments join the network ?
 - ◆ Edoardo: no formal process exists. LHCONE currently reluctant to open to much, to be able to keep control. Current ones were not a problem as most sites involved were already part of LHCONE.
 - ◆ Michel/Ian: LHCONE should discuss and propose a process as it should be prepared to received other similar request with most of the sites already part of LHCONE. In the meantime, requests will be directed to Tony and Edoardo.
- Eric: is there any global/central statistics for LHCONE usage ?
 - ◆ Edoardo: see LHCONE twiki giving operator monitoring pages. There is no central collector for LHCONE, difficult to make it: too many risks of double counting. Usually operators show their statistics at the LHCONE meetings. Each network has its own statistics.
 - ◆ Eric: is there no way to calculate the total volume transmitted through LHCONE ?

- ◆ Michel: it has always been like that, even before LHCONE. Total volume is estimated through statistics collected at the storage/transfer application level, e.g. FTS, xrootd
- Simone: how are connected russian sites to other T2s in Europe?
 - ◆ Edoardo: due to unpaid bills, cannot use GEANT network but connected directly to NORDUnet and ESnet. Connectivity to other European sites is going through the GP network: connectivity exists but bandwidth is more limited.

dCache Workshop and Storage Evolution - P. Millar

dCache workshop: also presentations from CERN collaborators and industries

- 35 people, 13 countries

Funding: LSDMA (German project) and INDIGO Datacloud

- DESY WP4 lead in INDIGO Datacloud
- dCache focus: Data QoS and data lifecycle management
 - ◆ QoS: generalization of SRM storage class
 - ◆ lifecycle: changing between 1 QoS to another one

SRM

- very much supported by dCache
- An interface, not a functionality: plan to provide other interfaces (e.g. CDMI/RESTful) to the same functionality
 - ◆ Waiting for input from experiments in shaping such an interface
 - ◆ Will be delivered as part of INDIGO Datacloud

WLCG http deployment WG

- dCache an active participant since the beginning
- dCache supports the http dynamic federation: plan to use it within INDIGO Datacloud
 - ◆ Installing FTS and WebFTS at DESY to get more experience

Ceph integration

- Currently, each dCache pool only sees its own private data
- Working on allowing several pools to share a common storage repository to provide the associated benefits: scaling, redundancy...
 - ◆ Require a major evolution of pool: currently designed to own their storage, concurrent operation raises a lot of tricky issues like file deletion...

NFS 4.1

- Survey of 21 sites: 1/2 running it in production or pre-production
 - ◆ None using it for WN access yet: more for T3 access or non WLCG use cases
- Recent perf study by KIT showing a 35% improvement over dcap
- A couple of sites moving forward

Federated AAI

- Activity at the intersection of LSDMA, EGI FedCloud, AARC and INDIGO Datacloud

Many other topics about future directions including storage evolution, dCache integration with DDN

- Look at Indico

Discussion

- David Crooks: quality of service and data lifecycle: does that include that data moves back and forward if the status had changed?
 - ◆ Paul: yes. So you can change your mind and switch status to precious. Can mark it that there will be multiple copies available. With data lifecycle can also specify a period to recycle space on the fast media.

Large-scale MC simulation at Helix Nebula - D. Giordano

Nov. 2014: CERN price enquiry for 2000 VMs during 45 days, resulted in a bid with ATOS

Production phase during March: ATLAS simulation

- 200 KVM HVs with 16 cores each: started with 2000 VMs, reached 3000
- Each VM running a HTCondor startd
- Each VM benchmarked at startup
- Key role of VM monitoring + accounting
 - ◆ Ganglia with 15s resolution
- VM provisioned through SlipStream
 - ◆ Status of VMs retrieved through SlipStream

Main results

- 93% CPU efficiency: jobs of ~9h
- 97% of WC time used by successful jobs
- 7th contribution to ATLAS simulation in March
- Showed some problems in provisioning and stuck deployment when orchestrated by SlipStream: more efficient to manage VM provisioning directly from CERN resource managers
 - ◆ Reached 93% of provisioned resources used
 - ◆ Some issues related to misconfiguration with iSCSI backend

Consumer-side accounting

- Required to validate RP invoices and assess effective efficiency of the resources provisioned
- CERN Ganglia used as the reference

VM benchmarks done with 100 single muon events simulation: ~2 min

- Spread of 15 to 20%
- A few less performing VMs detected by the benchmark: consistent with the time to execute the actual payload
 - ◆ This benchmark seems a prompt and effective solution to identify VM with poor performances: restart them rather than use them?

Regular CERN-SixSq-ATOS meetings

- Comprehensive notes of issues and actions, in particular about problems with SlipStream orchestration

Beneficial experience of managing VMs in a commercial cloud: good input for upcoming European procurement

Next

- Other VOs
- Go beyond simulation and cover analysis.

report - M. Alandes Pradillo

MW

- New GFAL version fixing FTS3 bringonline daemon crash
- Issues with last Globus library in EPEL: new behaviour to check if a host cert is legitimate, potentially breaking services using aliases
 - ◆ Interim release restoring the previous default behaviour

Experiments: high activity due to preparation of Run2

- See slides for details on recent issues
- CMS Global xrootd redirector (@CERN [↗](#)) added to WLCG critical service list

RFC proxies

- Readiness of WLCG infra will be tested soon through SAM preprod instance
- SAM proxy renewal requiring an easy fix

http deployment

- Agreement on goals and work to do

MW Readiness

- MS Readiness app available: should replace the manually maintained baseline list

Information System

- 14h incident with OSG BDII caused problem to access resources because caching was configured to 12h
- OSG is in the process of deprecating and removing OSG BDII in coordination with USCMS and USATLAS

New WLCG Operations Portal being developed: will centralize the Ops information for everybody

- Gathering useful information for sites
- Will soon advertize the first version to get feedback

Next WLCG workshop end of January or beginning of February

- New format with less topics and more in-depth discussions
- Main topics to be decided in the next weeks: feedback welcome

Discussion

- Maarten: about OSG problem with BDII, in the past we had a caching period of 4 days in the BDII: personally checked that, including the SAM BDII. Was it dropped ? Else why the problem?
 - ◆ To be checked/followed up offline

T0 Update - H. Meinhard

Cloud: more than 4700 servers

- Moving to CERN CentOS 7 for HV: proved to be advantageous

Database: piquet reestablished as during Run 1

- No critical service incidents
- Golden Gate migration completed
- DB on demand: bought Appdynamics for monitoring by users to help debugging tricky problems

Storage

- HW consolidation for both CASTOR and EOS
- CASTOR config simplified: 2-3 PB disk pool per experiment
- EOS: 140 PB, 50/50 Meyrin-Wigner
- ATLAS and CMS now EOS-centric: raw data sent to EOS and then archived into CASTOR
- CERNBOX open to all CERN users: 1 TB/user, 1 Mfiles/user

CERN mobile phone numbers are changing end of June: +41 75 411 instead of +41 76 487

- Short number (from CERN) unchanged

Grid services

- ATLAS and CMS reported availability for CERN as low as 90%: studied, understood and contained
 - ◆ Overload of various services, in particular batch system
- Efficiency studies: a dedicate person will be hired to work on this topic/project
- LFC to be closed later this month
- HTCondor migration started: ramp-up over 2015
 - ◆ Small prototype already run successfully during a few months
 - ◆ BDII integration and accounting done
 - ◆ Second phase: local job submission requiring Kerberos tickets
 - ◆ Aim: stop LSF at the end of Run 2

LXPLUS: performance improvements using hypervisors with SSD caches

- Got many complaints about performance problems

Infrastructure

- Version control: GitLab established, users encouraged to move
 - ◆ For projects for which GitHub hosting is not appropriate
 - ◆ Want to stop the current Git service
 - ◆ Will have to discuss end of service for SVN: nothing set yet, not in the short term...
- Savannah -> JIRA migration completed
- Volunteer computing: very significant uptake by ATLAS, significant activities by others
 - ◆ BOINC funding is a concern: NSF stopped its funding, discussion about turning it into a community project
- Twiki: considering moving to FOSwiki, compatible with additional features
- Quattor closed at the end of last year but some hosts are left unmanaged
 - ◆ May require some special actions from users to keep them up-to-date with security fixes

Discussion

- Jeff: why CERN CentOS rather than CentOS? Was hoping the end of CERN specificities...
 - ◆ Helge: bottom line is binary compatibility between ALL EL flavours. CERN had to make its own version due to some specific requirement like AFS. Has no impact on users (or other sites).

Belle2 - T. Hara

Compute requirements similar to ATLAS or CMS

Belle2 production infrastructure built around DIRAC

- Will use the same grid infrastructures as WLCG: EGI, ARC (SIGNET), OSG
 - ◆ DIRAC currently has an issue with last version of ARC CE client
 - ◆ OSG CE through WMS: DIRAC plugin for OSG CE being developed by ILC and will be available soon
- Also clouds through VMDirc or standalone clusters through SSH tunnels
- Current resources are 78% from EGI, 19% from clouds, 3% from standalone clusters (mainly Japanese universities)

Data distribution workflow quite similar to WLCG workflow

- Using FTS3: servers at KEK and PNNL

Metadata/catalog: AMGA + LFC

Current status: running 18K simultaneous jobs for MC production

- Only 60% of the resources used pledged to Belle2
- 15 regions, 31 sites
- 95% of resources SL6.x
 - ◆ Main SL5.x are current KEK CC and Japanese universities
 - ◆ Current plan to replace the computing systems at KEK this summer may be delayed to summer 2016 for financial reasons: discussing upgrade of current resources
- No DB service at KEK CC: have to be managed by Belle2 directly
- CVMFS Belle2 repository: Stratum0 at CERN
 - ◆ Also using grid.cern.ch for getting certificates
- Using GOCDB, GGUS and Dashboard (from CERN)
 - ◆ Redmine for non WLCG sites
 - ◆ Dashboards in particular for transfers
- A few shutdown expected at KEK CC: need to see how to make services redundant

Monitoring developments

- Periodic job submission tests
- Pilot activity: automatic error diagnosis
- Also retrieve information from DIRAC
- In the event of a site problem, currently manually open a ticket: considering moving to automatic ticket submission

Software framework used by jobs: Basf2 (Belle2 Analysis SW Framework)

- Working on multicore support
 - ◆ Multicore queue available at most sites as they have it in place for WLCG
- Memory footprint (RSS): 1 GB/core for single core jobs
 - ◆ Tests showing that with 8-core jobs, can reduce mem footprint (PSS) to 2 GB (instead of 8)
 - ◆ Most sites implementing mem limit in batch systems currently doing it with RSS: to be seen if PSS can be used

Collaboration with WLCG

- Resources/infrastructures: many sites in common, using the same network and MW
- Many useful tools in WLCG for Belle2, including operation tools like dashboard
- Would like to learn from WLCG experience

Discussion

- Massimo: have you studied the CPU efficiency of N single job cores versus multicore with N slots?
 - ◆ Takanori: No, will try to do it.
- Jeremy: Why using LFC rather than DIRAC File Catalog (DFC)?
 - ◆ Belle2 believes that LFC and AMGA are good product, that already demonstrated working at a scale larger than Belle2. No need for new features. Not much manpower, so difficult to change direction today, despite this remains an option for the future.
 - ◆ Michel: would be good to have exchange with ATLAS an LHCb to understand why they moved away from LFC and if it may apply to Belle2. As using DIRAC, DFC would have seem a more natural choice than LFC...
 - ◆ Markus: will have the same scalability problems with DFC and LFC because they both use a relational db. EOS-like memory-based namespace is much more scalable but not yet production ready. And moving from one catalog to the other is very difficult if a rename operation is involved.
- Michel: other possible topics for collaboration between Belle2 and WLCG
 - ◆ memory limits: see discussions at last GDBs. Current conclusion seems that moving to cgroups is probably the best way to do it properly (rather than letting the batch scheduler implement the limitation).
 - ◆ monitoring infrastructure: looks pretty similar to what SAM does. May be good if you could get in touch with SAM people (handled offline)
- Michel: what's the reason for using GGUS only for WLCG sites?
 - ◆ Maria: to use GGUS, sites must be registered into GOCDB. This is where site contacts used for notifications are taken from. May be a requirement not acceptable for some sites.
 - ◆ Maarten: we may need to revisit this at some point. Not only Belle2 has a problem, some WLCG T3 also cannot use GGUS because of the requirement to be a production site.
- Ueda: what is the exact status of CVMFS grid.cern.ch/grid-security/certificates ?
 - ◆ Several: it is production and if there is a problem, open ticket. This will allow to check that the support is properly in place.

DPHEP Collaboration Workshop - J. Shiers

Goals

- Establish motivation for long-term preservation: what are the common set of use cases like those agreed between the 4 LHC experiments
- Site/experiment roundtable to capture the HEP-wide situation
- A longer document, like a blueprint update, is expected by next workshop

Move to Open Science happening externally but relevant and matching what happens inside experiments

Use cases for "all HEP"

- Agreement on bit preservation and preservation of data, software and know-how
 - ◆ But still lacking a clear policy to avoid being dependent on particular person motivations
- Not a consensus of sharing data and SW with larger scientific communities: many experiments sharing only with their own community and no possibility to change it in the short term
- Open access to a reduced dataset

Knowledge capture "beyond the grave": no one thinks it's doable today but LHC and FCC with their long life time will have to face the succession problem

Joint projects

- Bit preservation: HEPiX
 - ◆ Copying data onto new media and transfer errors well under control
 - ◆ LEP era data is ~400 TB: 20 tapes today, probably 2 or 3 in the future. Negligible costs to have a few copies outside CERN.
- Virtualization: CernVM + CVMFS
 - ◆ Bootloader technology with everything in CVMFS improved a lot the situation
 - ◆ Put in CVMFS old experiment like Jade or legacy SW like Cernlib (and its doc!)
- Analysis capture: experiment work
- Open access, open data, open science

Data preservation impact on physics output still to be understood/evaluated

- But the cost is not the main issue... main issue is probably experiment manpower
- It is never too early to think about DP... but this advice is rarely followed...

Remarkable progress with DPHEP in the last 3 years (since blueprint has been published)

Discussion

- Markus: about open science, astro community do that, can we learn from them ?
 - ◆ Jamie: open science want paper and data, not reasonable. But this is not a well defined concept. Should be more active in discussion around it.
- Maarten: still lot of work for experiments to be able to reproduce analysis and results several years after
 - ◆ Jamie: even with virtualisation the environment will still be different, we have to be very careful and take part in ongoing activities

Cloud Traceability pre-GDB - I. Collier

A bit less attended that the previous one in February

Main area of investigations from the Feb. meeting

- Externally observable behaviours, like net flows
 - ◆ Survey of technologies done: see slides from the meeting
 - ◆ Several tools surveyed in particular nfsen, CERNSOC/OpenSOC. CERNSOC has similar features to OpenSOC with a few variations in the implementation. OpenSOC + BRO IDS look as the main trend at commercial providers.
 - ◆ Several tools to generate net flows: softflows, fprobe
 - ◆ Share "intelligence sources"?
- Logging inside VM but unfortunately no report as the meeting yesterday
- Better tools for managing large volume of logs
 - ◆ ELK adopted by several sites, particularly in UK
 - ◆ RAL visited IBM research and are discussing the possibility to work with them around their cloud analytics tools based on machine learning techniques: we can offer a diverse environment.
- Deferred deletion and quarantine of VMs
 - ◆ Currently only in StratusLab
 - ◆ RAL managed to implement it with OpenNebula for HTCondor managed VMs
 - ◆ OpenStack: some people interested, concrete implementation not yet clear, CERN scale (300-400 VM/hour) is a challenge

Security Service Challenges: S. Gabriel organizing it for EGI Fed Clouds, will see how it can be applied to WLCG instances

- May be a problem for Vac which has no way to inject jobs directly into it: has to go through an experiment machinery

Discussion

- Maarten: ELK and Flume both used will there be a winner, or are the use cases sufficiently different?
 - ◆ Romain: strong convergence on OpenSOC, we would be foolish to ignore this. But this is not one tool: in fact integrating many tools. At the higher level, its better to keep what is in OpenSoc, like Flume. But at lower levels can replace components by something fitting better with the local site infrastructure. This is what was done at CERN with CERNSOC.

HEPiX Update - H. Meinhard

Postponed to July as the meeting was running late.

This topic: LCG > GDBMeetingNotes20150610

Topic revision: r5 - 2015-06-15 - MaartenLitmaath



Copyright &© 2008-2022 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.
or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback