

Table of Contents

July 2016 GDB notes.....	1
Agenda.....	2
Introduction (Ian Collier).....	2
Security Operations Centre pre-GDB report (David Crooks).....	2
HNSciCloud update (Bob Jones).....	2
BDII and Information Systems (Maria Alandes Pradillo).....	3
Data matters in WLCG (Oliver Keeble).....	5
WLCG accounting task force update (Julia Andreeva).....	6
Accounting portal demonstration (Ivan Diaz).....	6
Discussion.....	6

July 2016 GDB notes

Agenda

<http://indico.cern.ch/event/394784/>

Introduction (Ian Collier)

<http://indico.cern.ch/event/394784/contributions/2225263/attachments/1308485/1957164/GDB-Introduction-20160713>

- Accounting portal check status + live demo
- August GDB cancelled
- may cancel October GDB
- People who use networks invited to LHCONe meeting
- DI4R including satellite meetings
- 20-22 September Helix nebula general assembly
- Also Supercomputing
- Moved Networking Pre GDB to December

Security Operations Centre pre-GDB report (David Crooks)

<http://indico.cern.ch/event/394784/contributions/2225261/attachments/1308495/1956805/SOCGDBSummary.pdf>

- Focussing on some specific areas
- Steady progress.

HNSciCloud update (Bob Jones)

http://indico.cern.ch/event/394784/contributions/2225260/attachments/1308473/1956765/BobJones_WLCG_GDB_13

- Lots of partners
- Potential users
- CERN active from start with commercial cloud services
- Microsoft Azure work published
- Deployment with T-Systems ongoing, Sept-Oct
- Hybrid model
- Procurement process not matched to commercial services
- PICSE came up with recommendations
- Pre-Commercial Procurement: 1- research organisations
- ESRF joined consortium
- >1.6M procurement funds
- Some manpower commitment
- Focused at infrastructure level
- Get fundamentals right
- Working with systems managers, IT staff rather than solely users
- WLCG, other physics, all sponsored.
- Needs <-> Provision use cases; risk assessed
- Results documented
- Payment models; which model is best for each use case
- Don't replace orchestration of experiments
- Running through to end of 2018
- Most Economically Advantageous Tender (not necessarily cheapest, best quality)

- Q: Are sure that someone will be able to live up to requirements?

- ◆ Reason for open market consultations. Companies rated themselves the difficulty. A number already have put in SAML based identity systems. They have to see there's a market.
- Downstream industries from some users
- Ian Neilson: What do you anticipate timescale
 - ◆ October/Nov framework, "We agree setup, 2016-2018"
 - ◆ Next stage have to sign another contract
 - ◆ Length of each one, 3 months design, more for prototyping...
- Ian N: Down the road?
 - ◆ What's good for community - looking to sign for annual period, a year-ish
- Ian N: Not a way of providing on-demand?
 - ◆ Thinking of framework agreement, actual sum you would pay would depend on consumption
- Q: if I compare current two procurements, difference in way experiments interacting. Train more going towards talking to central then fanning out
 - ◆ Not always a technical reason for that. What if failure of company. Who handles mess? Intermediate layer shields experiments. Contract management etc...
- Ian B: Really through batch system, through condor, so immune to what's underneath.
- Q: Operationally, day to day, I want to know that the cloud infrastructure, 50 nodes at place A are not working and want to disable
- Ian B: Yes - we'd want more or less transparent extension of T0.
- Maarten: Fed Ident? Companies would have to match each other because that's what this means. How much does this buy us since we'll need our own ident, voms etc.
 - ◆ Not going to create accounts for each user for each company. Doesn't mean will replace infrastructure or take away work of integration
- Ian B: Not see this being used in WLCG only. For WLCG overlay as currently looks. Also for individual user for other areas come with institutional id.

BDII and Information Systems (Maria Alandes Pradillo)

http://indico.cern.ch/event/394784/contributions/2225259/attachments/1308430/1956701/GDB_july_2016.pdf

- Representation of IS today, tools interacting with it
- LHC dependencies, help us to understand complexities, define doesn't match consumption
- All depend on BDII, Top-BDII, rebus
- Only ALICE use dynamic attributes
- BDII pros and cons
- Cons: info quality, do not validate, effort to do post validation, in the end VOs don't trust
- Proliferation of home made IS
- Discussing: CRIC system [Computing Resource Information Catalogue]
- Central CRIC + expt CRICs
- API to applications like monitoring to query CRIC etc.
- Experiment specific CRICs (ALICE/LHCb), very basic topology information like site names
- Consider whether we stop relying on BDII
- Stopping dependencies on BDII discussed
- Need to document plans to stop dependencies and where to get information
- Until CRIC is there, not going to change anything

- Known issues (capacities in REBUS)
- Waiting for new system in place making sure info reliable...
- Extra slides on info sources, etc...

- Michel: Compared to previous plan, is it the same idea?
 - ◆ CRIC using some concepts from AGIS, new thing, incorporated CMS needs/systems. AGIS devs + new devs. AGIS code base as starting point

- Maarten: Advent of CRIC would be good thing, first time nice consistent overview of what WLCG is. Always been missing. Good thing, proven AGIS technology + CMS requirements, plus discussions in TF. Generic, flexible tool. Hopefully not controversial. Making ourselves independent, we don't have to worry if the BDII data is of bad quality, nice evolution of technology.

- John Gordon: Based on Open Source?
 - ◆ Python, Django

- John Gordon: Other people use it?
 - ◆ The idea is to make something quite generic able to be adopted by other communities

- Maarten: beware of potential scalability concerns; with our use case in mind, not going to be queried 100 times a second. Specific to ATLAS, made more generic, now in principle capable of seeing variety, be careful what you'd want to use it for
- John G: Make it safe against 100/second?
- Maarten: Can always DoS, we have other ways of dealing with that
- Gavin McCance: Is it designed to scale out?

- Jeff: Three things in BDII don't want to replicate by hand
- 1 CE drain state? yes/no
 - ◆ Maarten: only ALICE use it

- 2 CE machine, list of endpoints, not in GOCDB.
 - ◆ Need to understand if this goes to GOCDB or to CRIC

- 3 ACLs for queues.
- Jeff: Connector BDII -> GOCDB
 - ◆ Maria: Need to understand use cases,

- Maarten: EGI use cases, not about dropping BDII, make us independent of it.
- Jeff: Fine if independent of BDII as long as this doesn't increase load on site personnel
 - ◆ Simpler systems are intention, less work for sites.

- Maarten: Important input (work for sites). What is in BDII is largely ignored. Idea was system fully dynamic. If today a queue has a different name, the original grid paradigm was: auto discovery -> auto use. However, a site may have multiple queues, not every queue suitable for every flow, so anyway have to discover which queue to use, by talking to the site admin. Existing practice, try to cast into sustainable official position now.

- Maria: How to proceed?

- Q: Effort coming from to sustain this? Is this going to require new component at sites?
 - ◆ Intended to be central system, Central CRICs etc.
 - ◆ Don't see sites have to install anything, maybe too early to say but that's the plan.

- Ian C: all dynamic info, not GOCDB, rely on telephone and email?

- ◆ Maria: Which info? Essentially all info under consideration is static, not dynamic
- Ian C: If the only info sources are static?
 - ◆ only ALICE want dynamic, happy to continue using the BDII for that
- Ian C: Queue names, ACLs, people ignore it so we rely on current situation
- Michel: Tend to agree with Ian. Use of HTCCondor bypasses dynamic info: will it also become a requirement for ALICE?
- Maarten: Don't spend too much time on ALICE
- Michel: Have backup plans for dynamic? are these putting requirements on sites?
- Maarten: No backup plans for dynamic information, only talk about static information. New queue name, new VO ACL, usually require discussion with site + VO, been for 10 years. Obviously could try to implement fully dynamic system. Have tried to make reliable but fundamental battle. Even dynamic info is highly questionable much of time for many sites. In TF shall we continue on this path trying to make system reliable, or consider alternatives?
- Michel: One info not mentioned, DIRAC for submitting jobs, how many jobs in queue?
 - ◆ Not through BDII, query the CE directly
- Maarten: Main point was introduction of CRIC, can only bring benefits, finally have right technologies, frame of mind, positive thing. Can we depend less on other things. Don't forget this, discussion is difficult, that's why we have a TF, have been having these discussions for many months. This is what we came up with that has most traction, maybe not perfect but in complex system
- Maria: ask for input on whether this is appropriate, info sources, getting data into CRIC, put in practice. Need to know if WLCG is fine with this, can keep discussing but need to move forward, this is a proposal. Need green light.
- Julia Andreeva: ATLAS: rely mostly on AGIS, biggest info, if they succeed?
- Michel: In ATLAS data AGIS collects include dynamic data. Can also in CRIC add things.
 - ◆ Maria: Slide 4, ATLAS doesn't use dynamic info in BDII
- Peter Solagna: Would encourage fact that WLCG specific info, want to migrate BDII -> something else, encourage have them in a general purpose tool. Endpoints CEs, queues, in BDII, needed to submit jobs, if moved away from BDII if effort to move these, should move them to GOCDB
 - ◆ Maria: That's the idea. Did exercise with one UK site put minimum set of info in GOCDB, quite easy to add this today.
- Ian Bird: For approval, needs to be put before Management Board (next week), need to continue with TF, edge cases, discussing dynamic and static queue name not dynamic even if changed every day. A lot of info never used, need something similar. One slide requesting approval with design of service etc.

Data matters in WLCG (Oliver Keeble)

<http://indico.cern.ch/event/394784/contributions/2225217/attachments/1308070/1957089/GDBJul16v1.pdf>

- Ian B: in Lisbon we decided a forum is needed for data management aspects
 - ◆ a GDB WG is a good fit
 - ◆ its terms of reference need to be agreed
- Ian C: the upcoming WLCG workshop has a good slot for it

- Markus: mind that storage and data access are different
 - ◆ both need to be covered
 - ◆ "Data Management" could imply that
- Jamie: outside WLCG this would be called a Data Management **Planning** Group
- Alessandra: Data Management Coordination?
- Renaud: can site issues be discussed there? E.g. unused data?
 - ◆ A: yes

WLCG accounting task force update (Julia Andreeva)

<http://indico.cern.ch/event/394784/contributions/2225219/attachments/1308692/1957157/WLCGAccountingGDB.13.0>

- John:
 - ◆ inconsistencies across views, if any, are bugs to be fixed
 - ◆ WLCG defines the WLCG views
 - ◆ it should (also) be documented what is not included
- Julia:
 - ◆ we will only validate the WLCG view
 - ◆ non-pledged resources should also be viewable there
- John: the TF should propose a cut-down WLCG view and invite comments

Accounting portal demonstration (Ivan Diaz)

Discussion

- John:
 - ◆ examples of issues that have been fixed:
 - ◇ the problem for DESY
 - ◇ CPU time appearing to be larger than wall-clock time
 - ◆ overall good progress!
- Ian C: are all things on track?
- Ian B: the main worry was that numbers could not be trusted
 - ◆ that looks under control now
- Ian C: is the interface OK?
- Ian B:
 - ◆ not sure if the reports should pull in the pledges?
 - ◆ should the reports be there at all?
- John:
 - ◆ the portal reports can also be yearly, quarterly, etc.
 - ◆ REBUS does not offer such flexibility
- Ian B: we should be careful not to mix functionalities
- Ian B:
 - ◆ could the interface allow a selection of filters to be built up?
 - ◆ and then save them for future use?
 - ◆ in any case the interface should show which ones are applied
 - ◆ can we e.g. have an "all WLCG" selection?
- John:
 - ◆ we will add the country view to the WLCG set
 - ◆ the "*advanced options*" allow fine-tuning the selections

- John: we were asked some questions about desirable APEL enhancements
 - ◆ Q: can batch system scaling be undone?
 - ◆ A: that would be a big work, more for the Run-3 time scale
 - ◆ Q: can we have daily granularity?
 - ◆ A: that would mean a **lot** more data to process
 - ◇ plus changes in the client
 - ◇ we rather would see more sites publishing summaries instead
 - ◇ i.e. less work for APEL
 - Julia: you have the raw data, i.e. any desired granularity?
 - John: monthly summaries already can take ~9 days to compute!
 - Ian C: can you move to NoSQL for such computations?
 - Adrian: yes, we will already investigate that for EGI later this year
-
- Ian C: we are where we wanted to be at this time!

This topic: LCG > GDBMeetingNotes20160713

Topic revision: r6 - 2016-07-27 - unknown



Copyright &© 2008-2021 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