Table of Contents

Summary of GDB meeting, September 12, 2012

Agenda

https://indico.cern.ch/conferenceDisplay.py?confId=155072

Introduction

October pre-GDB and GDB will take place in Annecy (France)

- Registration required at least 1 week in advance to organize cathering
  - Registration is free
  - Separate registration for both events: see Indico
- Accomodation and transportation information will be added by end of this week (Sept. 16)

October pre-GDB will be on storage issues

- Co-organized by all storage-related WGs

Follow-up of actions without a specific presentation/discussion today

- WN environment variable
  - Proposal adding variables related to multicore support now available at https://twiki.cern.ch/twiki/bin/view/LCG/WMTEGEnvironmentVariables
  - Ulrich started to work on LSF implementation
  - Action on Nikhef people : look at proposal and give feedback, agree in group, implement
  - Be careful to make sure it doesn't conflict with GLUE2 multicore work.
  - Simone points out that we need an agreement: twiki plus what all sites deploy should agree. Meeting and alignment is needed. Open question who owns the action to coordinate the meeting, alignment, agreement, implementation? Ops coord team?
- Storage Accounting (StAR)
  - Changes requested by WLCG (see June GDB) accepted by EMI and OGF
  - dCache started to work on implementation, DPM and StoRM have not yet started
  - Non EMI storage providers: need to engage with them
  - APEL has a test consumer ready

Topics for future GDBs, forthcoming related meetings: see slides

Market Trends

Processor market : server is only a very small niche, with a very low growth rate

- 50% of server market consumed by Microsoft, Amazon, Google
- Highest growth rate market: smartphone (60% this year), tablet (270% this year)
- More smartphones than PCs sold this year
    - ♦ Consequence: pushing mobile processors, the low end of the processor market

Many core: several architecture candidates but many promising projects didn't materialize in production or where canceled.

Processor complexity: growing number of transistors but 50+% for L1+L2++L3 caches in the server processors

- Intel/AMD: 100M transistors to compare with ARM: 6M transistors...

Server price : small improvement per year expected (~15-20%)

- Processor price (Euro/HS06) doesn't reduce over the years for a given generation
    - ♦ Euro/HS06 reduction only by going to newer generation
- Memory price very volatile
- Increasing market concentration for memory and disks: 2 or 3 companie sharing the market for every technology

Memory: pb to improve bandwith and consumption

Hard disk: at the end of PMR technology, future is HAMR

- Will bring a factor 10 in density
- But it is a very expensive technology: huge investments needed, will impact price reduction
- 50% of disk consumption from clouds
- Prices not yet back to their level before the Thailand crisis

SSD: factor 10 in price compared to HDD, no convergence in a foreseeable feature

Future storage technologies: many innovative technologies but no market relevance in the 3-5 years

- Memristor (not yet a prototype), PCRAM...

No breaking technology expected before 5 years: too difficult to compete with huge investment made to incrementally inmprove current technologies

GPU: no price reduction expected.

Tape: 90% of market is LTO

HEP is not benifiting from most innovations in processors because of its SW: most of the performance gain is coming from vector instructions that we don't use

- Manpower required to improve our applications
    - ♦ Several investigations showed that ~10% of performance improvement (10% of HW less) can result from SW improvement

- Markus: suppose we can improve 10% in speed -> 10% less hardware purchase... this is quite a bit of money allowing to fund many people. May be worth the investment!

Operations Coordination Team

Emphasis is on coordination: the team will not do operations itself.

Would like participation from many people with a low level of (committed) effort (10-20%)

- Evolving service is a challenging activity
- Decommissioning obsolete services also require some coordination effort
- Several initial (short term) task forces: CVMFS, glexec, perfSONAR; Squid monitoring, FTS3 deployment, SHA-2
    - ♦ To be decided at kick-off meeting
- Mix of core members and targeted experts from sites/region, experiments and services
    - ♦ Serveral signed up already

Interaction planned with other WGs and daily OPS.

Meetings planned

- Quarterly operation meetings with experiments: discuss needs, task forces...
- Fornightly meetings

Kickoff meeting planned on Monday the 24th

Maria calls for participation in both the core ops coord team, and for members of the dedicated task forces.

- GDB members are expected to forward this call

I/O Classification and Benchmarking WG - D. Duellmann

3 major players trying to improve some aspect of I/O: experiments, sites, storage providers

- May converge but also may just interfere
- Benchmarking is important to assess the reality

WG motivations

- WLCG aggregate I/O access patterns difficult to reproduce
- No agreed common metrics
- Several benchmarks not able to run in multi-client configuration

WG goals: help to optimize sites, exp frameworks and storage implementations for a given I/O access pattern

- Review market for agreed key metrics, existing tools
- Survey existing monitoring information at experiments and sites wrt to the metrics

Non goals

- Compare different storage implementations against each other: done by HEPiX WG
- Experiment framework and ROOT I/O layer optimisation: already taking place as part of ROOT I/O workshops (2/year)
    - Collaboration with this forum is essential
    - A key forum for promoting commonalities between experiments

A question example: what balance between WN storage and storage cluster

- Copy-local: all random I/O take place on the WN
    - Storage systems optimized for put.get
    - easier integration with cloud storage
    - IOPS perf essential on WN
- LAN access: mainly random I/O backend perf matters

Logs/monitoring data are a good source of information to mine to understand I/O access pattern.

Members: experts from all 3 actors (site, experiments, storage/ROOT)

- Call for participation: people interested should contact Dirk.
    - GDB members are expected to forward this call
    - Jeff: Nikhef student is doing a systematic comparison of analysis on grid vs. cloud, may have some information coming out of that which will be helpful

Kickoff planned next week.

Action Follow-Up

CVMFS Deployment

No major figure changes due to the summer

- A couple of additional sites for LHCb
- CMS is pushing adoption: 1 T1 migrated

Issues identified at last GDB preventing adoption

- WN disk space requirement
- Inability to install SW on WN (OS support)

New capabilities in development CVMFS branch (2.1.x): both client-side and server-side

- NFS support
- Shared cache
- MacOS X support
- Unpinning of local catalogs during catalog updates
    - Has been causing problems with long running jobs keeping metadata unnecessarily pinned and filling up the cache
- Server can now store uid/gid and implement standard POSIX checks

Testing of new capabilities welcome: new stable version expected by the end of the year

- Not stable yet

A CVMFS task force will be created in the Operations Coordination to help with deployment

- Sites and VOs representatives
- Kickoff meetings during the Operations Coordination kickoff

Maarten expresses its concern about sites who cannot access any shared resources from WN and will not benefit from NFS export feature.

- Addressed only by ALICE approach based on bittorrent
- Short discussion saying it is a corner issue that should not prevent deployment of new features

SHA-2 Proxies

Agreement at IGTF last monday that:

- No SHA-2 cert before August 1, 2013
- Aim to have the production infrastructure ready by that time
    - EMI-1 end of life is April 30 + 3 month grace period
- Monthly tracking of progress and blocker if any: SHA-2 introduction date to be delayed further if needed
    - If there is no sign of SHA-1 breaking (concrete) risk

Which SHA-2 variant is expected to be supported?

- See last IGTF agenda : SHA-2 versions 256 & 512 Maarten's SHA2 talk

All EMI-2 productions should support RFC proxies

- WMS not yet available
- Little uptake so far: a few bugs were found and fixed
- SHA-2 supported by every service except dCache?

Updated milestones

- Deployment of SW supporting RFC proxies by beginning of Spring
- During spring switch to RFC proxies
- Uggrade dCache and Bestman at the end of Spring

Plan B status

- Plan B presented as July GDB: setup of a catch all CA
- Currently put on hold: not very attractive, will require a significant effort
- Concentrating on meeting the new deadline for SHA-2
    - Don't rely on further postponement of the deadline: will really depend on concrete possibility that SHA-1 may be broken

glexec

3 more CEs supporting it compared to July

- Core sites supporting it are stable but others are not
- ASGC : a few problematic WNs for ATLAS and CMS
- BNL : not yet setup

Need to follow ARGUS deployment in addition to glexec

- For central banning

glexec priority will be ramped up in ops coord team

Multi-core jobs in IS

Proposal converging and documented: see slides.

Open issues

- Single or multiple value for RequestableCores
- RequestableCors: A numerical value with the number of cores or just a "multicore" capability with the number discovered by the job
    - Need better understanding of the use cases

Need site people (call for participation) with LRMS scheduling experience and multicore to look at Maria's presentation on multicore and make comments.

EMI-2 Transition

Proposed plan and methodology

For EMI-1, recommendations given on a baseline services page

- Distinguishing between services with or without long-term state information (eg. SE vs. CE)
- WN client validation started but didn't end with a clear status

Relying on EGI staged rollout: overall a success

- A few exception where severe problems were discovered by real users (eg. CREAM CE for ALICE)

Current result is an infrastructure running mainly obsolete, unsupported SW

- Despite a few sites running EMI-1 services and FTS 2.2.8 entirely EMI
- gLite 3.2 is in the process of being retired: even people with knowledge on how to build a gLite release have left

Obsolete MW must be upgraded to EMI: v2 better than v1

- EGI will follow-up this through NGIs
- WLCG: balance between progress and risk may be different
    - ♦ Clients can be relocated to application area

WLCG validation needs to start

- EGI staged rollout/validation doesn't cover everything
- No perfect moment: LS1 will not be quieter, don't wait for it
- Better to be ready to upgrade before the end of the EMI project

Worker nodes: Some willing sites agreed to install EMI-2 WN (SL5/64-bit)

- SL6 will follow
- DESY, CERN, Rutherford, Brunel, INFN-Napoli, INFN-CNAF
- Best effort: difficult to converge in a short time
- Done using the EMI test repository: allow quicker fixes in case of problems

Need coordination: need a more formal collaboration to increase efficiency

- A first step is to create a wiki page to track the status of testing/validation

Post-EMI SW lifecycle still in discussion. General ideas:

- As much independance between PTs as possible
- "You need it, you provide it": this will be the new philosophy with no supporting project
    - ♦ +: development will be made by people who need/use it

DPM: migration from gLite 3.2 to EMI-2

Already 25% of sites running EMI-1 DPM

- Was 5% 2 months ago

- Good example that upgrade of service with state data can be migrated in a non disruptive way

Most new features only in EMI

Several package name changes to be Fedora/EPEL compliant

- Both EPEL 5 and 6 are available
- Only missing package: Oracle-related ones and YAIM-related ones
  - ♦ Will never make it
- EPEL and EMI packaging are the same and provide the same versions: can be installed from EPEL
  - ♦ Keep EMI repository for YAIM

Both reinstallation and upgrade are possible

- Reinstallation always simpler when possible
- No need to upgrade all nodes at the same time: can mix gLite/EMI head nodes and disk servers

Several sites already installed on SL6

Discussion

Need to start evaluation asap to better discussion migration timeline

- Need to extend evaluation to all services, not just clients (WN/UI)
- Maarten: WN validation expected to be completed soon (end of the month?)

Lots of discussion about the WLCG, EMI, EGI recommendations / deadlines for upgrading from glite 3.1 / 3.2 and EMI-1 .... really hard to make sense of it.

- bottom line is that we (WLCG) need to be proactive and to start validation then share information about whatever explorations we do in moving services, provide feedback and help to other sites.
- Jeff's suggestion : start posting to ROLLOUT again ?? 

Actions

- Twiki page tracking validation work for EACH service: list of sites involved, pending issues, recipes...

perfSONAR

Number of instances doubled in last 6 months.

PerfSONAR dashboard is a key piece of the infrastructure * Several new communities created

A wiki collecting deployement best-practicies and issues * http://www.usatlas.bnl.gov/twiki/bin/view/Projects/LHCperfSONAR * Everybody can contribute * Shared with developers: expect many fixes to be incorporated in future releases

Managing site configurations is currently a bit painful: need a configuration update at every site when a new site is added to get it tested * Looking at implementation of centralized VO-configuration, maintained by the VO and used directly/transpartently by sites

Actions

- Circulate the main entry page for sites wanting to configure perfSONAR
    - ♦ Ensure it is VO independent and even LHCONE independent
- Prepare a survey/questionnaire for sites who installed perfSONAR and those who didn't to better understand issues that could prevent wider development
    - ♦ distribute through GDB members
    - ♦ Review at a later GDB

/Nebula Results

40K CPUxday run during the PoC

Basically successful, a few issues identified.

TechArch group working to federate providers

- Common API
- Single sign-on
- Image and data marketplace
- ...

Next steps

- lightweight federation, common API
- image management, brokerage

Dan's remark: HEP is not very much interested in TechArch proposed features, except the common API * Lack of common API between cloud provides is very painful: need a specific implementation for every provider * Other topics discusses by TechArch are driven by other communities, like EMBL, that want to use clouds as their data repository, with the ability to allocate computing resources on demand to process these data: not matching the HEP vision

Storage Interface WG

Motivation for the WG related to the future of SRM

- Increasing storage not manageable by SRM
  - Other well known issues like inconsistencies between implementations and performance
- WG approach: maintain SRM at archive sites, look at possible alternatives for disk-based storage
  - Do not invent a "new SRM"
  - Evaluate possible alternative as they emerged: in particular ensure they can be supported by FTS and lcg_utils to allow interoperability and smooth migration
- Clearly not a short-term WG...
  - Report at least every 3 months on progress at GDB
  - In 12 months, propose a roadmap

Experiments, middleware experts and sites should agree on alternatives to be considered for testing and deployment, targeting not the full SRM functionality but the subset determined by its actual usage in non-archive sites. * Small working group reporting to GDB for wider discussion

- No intent to exclude anybody
- Representatives from sites, experiments and storage providers * Ability to integrate cloud-based storage must be taken into account

Wrap-UP

Members welcome for operations coordination and IO Classification/benchmark WG

Operations coordination kickoff on Sept. 24th

- Several task forces planned, related to GDB actions

SHA-2 proxy and EMI-2 transition tightly linked

- Need to start asap evaluation: well underway for WNs but need to cover all services
- A central wiki page will be created and advertized

glexec deployment: follow-up ARGUS deployment in conjunction

- Central banning is as important as identity switch

Storage: a lot of activties started by different focused WG

- Interactions, pre-GDB

perfSONAR

- Good progress in deployment
- Need to better understand issues that may prevent wider deployment

Need to review MW clients in AA in the future in the light of EMI migration

- Update on the process from getting software (clients) from the MW or infra projects into the stuff the experiments distribute in their area on the WN
- How does this work, how has it changed, is it still appropriate in the "new era".

This topic: LCG > GDBMeetingNotes20120912
Topic revision: r1 - 2012-09-14 - MichelJouvin