

# Table of Contents

<b>Summary of GDB meeting, May 8, 2013 (CERN)</b> .....	<b>1</b>
<b>Agenda</b> .....	<b>2</b>
<b>Welcome - M. Jouvin</b> .....	<b>3</b>
<b>LHCONE/OPN Update and Plans - E. Martelli</b> .....	<b>4</b>
<b>Migration to Grid Engine - M. Alef</b> .....	<b>5</b>
<b>MW Development and Provisioning after EML</b> .....	<b>6</b>
MEDIA Initiative - A. De Meglio.....	6
EGI after EMI - P. Solagna.....	6
Discussion.....	7
<b>Actions in Progress</b> .....	<b>8</b>
Ops Coordination Report - M. Girone.....	8
SL Migration Plans - A. Forti.....	8
Xrootd Monitoring Infrastructure - D. Giordano.....	9
DPM Community Status - Oliver Keeble.....	10
<b>Computing Model Updates</b> .....	<b>11</b>
Introduction - I. Bird.....	11
Technology trends - B. Panzer.....	11
Experiment Computing Models - M. Cattaneo.....	12
Application SW - F. Carminati.....	13
Computing Services - I. Bird.....	13

# Summary of GDB meeting, May 8, 2013 (CERN)

# Agenda

<https://indico.cern.ch/conferenceDisplay.py?confId=197803>

# Welcome - M. Jouvin

Upcoming meetings: in June we need to decide if there is to be another outside CERN GDB.

- Michel finds people in favour but the meetings have low attendance and especially less people from CERN.

News on actions:

- Message that new APEL publishers released in EMI-3 will be available in EMI-2.
- WN in CVMFS. Have any experiments used it? Do we close the action?
- Deployment of xrootd and WebDav access at all ATLAS DPMs. Any update? (WebDav is to allow access for the new naming convention. This is not federated WebDav)
  - ◆ Simone: not yet started, an action for the coming months, no follow-up needed by GDB or Ops Coord at this point
- Are jobs with high memory requirements still an issue?

HEPiX news: a puppet Working Group had been started led by Ben Jones (CERN) and Yves Kempf (DESY).

- More detailed report at next GDB

Ian Bird: after the end of EMI and EGI-Inspire SA2, CERN IT/GD and IT/ES groups have merged under leadership of Markus

- Jamie Shiers left the service coordination to lead the data preservation effort
- Maria Girone becomes the new service coordinator
- **Jamie deserves a great deal of thanks for the immense work he did in building the WLCG service**
- **Others agreed that Jamie's contribution and focus on such things as service challenges had a huge impact and the collaboration is very grateful.**

# LHCONE/OPN Update and Plans - E. Martelli

LHCONE: private routed network, currently based on L3VPN

- 44 sites, 17 countries, 15 NRENs
- Main operators: GEANT, CERN, Internet2, ESnet and CANARIE

LHCONE R&D activity: P2P service

- On-demand P2P link established when 2 sites want to exchange data
- Not needed today but may be an attractive technology when/if network becomes a bottleneck
- Potential benefit: guaranteed bandwidth, deterministic
- But a lot of challenges including:
  - ◆ Need for a API used by SW to provision the circuits
  - ◆ Challenge of multi-domain provisioning
  - ◆ Not covering the last mile that can be also an issue and degrade the guaranteed bandwidth
  - ◆ Routing over these circuits
  - ◆ Billing: are users ready to pay for this service? How to do the billing?
- Workshop last week decided to create a testbed based on static P2P links between volunteer sites (T2s) to compare performance with the currently existing infrastructure
  - ◆ Static link used to avoid change in SW

LHCONE also exploring SDN

- First use case: WAN load balancing over the 6 transatlantic links using OpenFlow
  - ◆ Goal: optimize the use of these expensive resources
  - ◆ Main challenge is coordination between multiple domains
- Is the SDN coordinated with the Network Working Group?
  - ◆ EM: Not at the moment as the activity is recent.

Simone: can P2P service help solve the poor connection of a few sites?

- Edoardo: not really... first must join LHCONE! Which is open to everybody...
  - ◆ It's an expensive service: it is for sites with a good connectivity already that want to have a better one.
- Ian: you need to convince people to improve their local/national connectivity. We have been successful in a few cases.

IPv6: CERN expects IPv4 shortage to happen in 2014

- WLCG community urged to consider using IPv6 and help with the testing
- Testbed activities driven by HEPiX WG

# Migration to Grid Engine - M. Alef

GridKA was using PBSpro: many stability problems, problematic fairshares

Evaluated several alternatives, in particular Torque/MAUI and Grid Engine (GE)

- Torque/MAUI proved to be less stable than PBSpro
- Several testing stages since 2010, migration undertaken in 2012 to Univa GE

Current configuration

- Single server, no failover
- Flat files, no DB
- Certificate Security Protocol enabled
- Fairshare based on wall-clock time
  - ◆ historical data can be mixed with other parameters to decide the actual job priority

Lesson learnt

- Steep learning curve at the beginning
  - ◆ Documentation not always very clear
- Very good support from Univa (based in Europe)

Luca: any figure about the scalability limit?

- Manfred: successful tests done with 100K slots and 1M of jobs

# MW Development and Provisioning after EMI

## MEDIA Initiative - A. De Meglio

EMI legacy: EMI-3, released in March 2013

- 61 products, including new ones: STS, EMIR, Hydra
- "Supported until April 2014", "bug fixes until April 2015"
- EMI-2 status: "bug fixes until April 2014"

Current situation: Product Teams (PTs) are now fully responsible for their SW build process

- EMI-2 built using standard tools (mock, pbuilder) but ETICS still needed for some parts
- EMI-3 relies only on standard tools

PTs now fully responsible for SW releases and announcements

- EMI EMT officially stopped but mailing list will continue as an announcement list for PTs
  - ◆ Monthly summary of the announcements will be done but not technical coordination
- EMI EMT responsibility transferred to EGI URT for the product teams who agree to join UMD
  - ◆ Cristina will co-chair EGI URT during a transition period

PTs now fully responsible for their product support

- GGUS still used as user support tracking system
- No monitoring of any SLA compliance
- Each product has a lifetime defined at <http://www.eu-emi.eu/retirement-calendar>
- Some products with unclear support
  - ◆ WN, UI and YAIM-core: discussion ongoing with INFN were responsible for them
  - ◆ Hydra, STS: a commercial company may handle the support

MEDIA: is a lightweight collaboration that has been launched

- Collaboration between interested PTs
  - ◆ All PTs from EMI confirmed their participation
  - ◆ Not restricted to former EMI PTs
- 3 main objectives: forum for technical discussions and follow-up of actions decided, coordination for future funding proposals
- First co-chairs: Balazs Konya (LU) and P. Fuhrmann
- Idea validated with all main EMI stakeholders: agreement as long as this is lightweight and voluntary

## EGI after EMI - P. Solagna

During EMI and IGE, no direct (formal) contact between EGI and PTs: EGI TCB in contact with EMI and IGE

With the end of the projects, EGI needs to deal with a larger number of technology providers (PTs)

- New board created: URT (UMD Release Team)
- URT will take over part of the release coordination work previously done by EMI
- Membership: PTs (or group of PT), UMD provisioning team
- Lightweight coordination: mainly wiki pages plus bi-weekly meetings
  - ◆ Will track only works from PTs relevant to UMD

- ◆ PTs are invited to subscribe

#### URT mandate

- Discuss the UMD roadmap and updates schedule
- Report the status of products to the UMD provisioning pipeline
- Agree on Quality Criteria
- Gather development plans from PTs and anticipate changes that potentially affect other PTs
- Track the status of critical bugs/requirements
  - ◆ Eg. SHA-2 support
- Evolve UMD repository structure

#### TCB evolution

- New representatives from main MW providers
- MEDIA representatives
- Mandate: EGI technical roadmap/strategy, SLAs, high-level cross-product requirements

#### 3d level support in GGUS

- Support units being reorganized, with new contacts, to match PT structure after EMI
- New process for handling unresponsive tickets
- Minimum level of commitment requested for each PT, even if best effort

## Discussion

Ian: WLCG would like to deal directly with PTs, not with an integrated distribution. Not sure how EGI vision matches the WLCG needs (described in a document discussed over the last 6 months). Would like to see products in EPEL and MEDIA as the coordination.

- Peter: not incompatible with EGI vision/structure. PTs will remain independent but EGI will try to take care of dependencies between products. They will release in EPEL and UMD will pull from it. MEDIA is more about technical evolution than release process.
- Maarten: the baseline from UMD can be helpful (safe baseline). EGI have been taking responsibility for the non-US infrastructure. There are many things that we would have to do for ourselves. Would not help sites to have conflicting messages – one from WLCG and another from EGI. Need to collaborate where we can. Do not want to return to big-bang releases but we need to be careful of disruption caused by conflict between packages or timelines.

Intense discussion on EGI proposal vs. WLCG needed flexibility as far as SW provisioning is concerned

- WLCG insists that it may need to deal directly with PTs and pick-up version not yet released in UMD
- EGI agrees this is not a problem as long as the version goes through a minimum certification
  - ◆ Always the case with WLCG
  - ◆ Worked well in the past, should work even better with the lighter coordination proposed

Review where we are in a few months...



# Actions in Progress

## Ops Coordination Report - M. Girone

CVMFS: target date (April 13) met

- ALICE beginning evaluation
- Very good site participation in evaluating new versions

SHA-2 good progress: ready to be tested, hope to be able to have full support in production during autumn

glxexec: now integrated into DIRAC, sorting out a few new problems at sites

- No scalability problem seen so far
- **Deployment is ready for pushing. TF recommends to MB to aim for 90% adoption by end of 2013.**

perfSONAR: v3.3 RC3 being tested and a few issues found

- Sites should wait before installing it
- New modular dashboard: an alpha release available, very promising

FTS3: successful tests; still waiting for a deployment roadmap

http proxy discovery: new TF started, led by D. Dykstra

MW

- All sites still running EMI-1 service must upgrade them ASAP
- EMI-2 VOBOX now available
- New WLCG repository created for things that cannot be hosted somewhere else

Conclusion

- Good progress reported by several TFs, eg. CVMFS deployment done in 6 months
- Good collaboration with OSG and EGI

## SL Migration Plans - A. Forti

TF membership: sites, experiments, EGI and IT/ES

- Good representation
- A mailing list + a Twiki page (LCG/SL6Migration)

Timeline proposed

- Before June 1st, sites supporting several VOs including ATLAS encouraged to test without upgrading
- Before end of October: upgrade bulk of WLCG resources to SL6 (5 months)
- After: discuss move to SL6/EMI-3

Recommendations

- Gradual migration rather than big-bang

- Do not mix SL5 and SL6 resources in the same queues
- LHCb wants SL6 queues to be published into BDII

HEP\_OSlibs for SL6 available and documented

- SL5 version still available but unmaintained
- Sites encouraged to install it on WNs

New WLCG repository created, in particular for HEP\_OSlibs

- <http://linuxsoft.cern.ch/wlwg>
- Sites encouraged to enable it
- Xrootd plugins will be soon added

Migration status

- LXPLUS alias migrated to SL6 last Monday
  - ◆ VMs
  - ◆ lxbatch currently has 10% of resources migrated to SL6: migration will be done inline with lxplus usage growth
  - ◆ Details of CERN plans can be found at <http://itsb.web.cern.ch/service-change/lxplus-alias-migration-slc6/06-05-2013>.
- T1s: most have already complete testing or are doing it, several already planned migration
- T2s: 81 over 131 forwarded their plans, many awaiting the green light from experiments (particularly ATLAS)

### Need more country T2 representatives in the TF

I. Fisk: CMS encourage sites to migrate their WNs to SL6 asap.

- No problem to run SL5 binaries

S. Campana: would prefer big-bang migration at each site (all site resources migrated at once)! Progressive migration requires creation of additional queues, a time-consuming operation for just a migration.

- ATLAS would prefer the risk of temporary problems or downtimes
- Avoid migrating all sites at the same time...
- MJ: for small sites it is okay to go ahead right now, for larger sites it is best to coordinate with central experiment teams to decide the most appropriate timing and approach.

## Xrootd Monitoring Infrastructure - D. Giordano

2 monitoring streams from Xrootd (using UDP)

- Summaries: aggregated by MonALISA
- Detailed (and f-stream): requires the use of GLED collector for aggregation
  - ◆ Collector is publishing into ActiveMQ: information processed by dashboard, popularity calculation, WLCG transfer dashboard...
  - ◆ Both for EOS and federation activities
  - ◆ Require the use of new Xrootd plugins that will be distributed in the new WLCG repository: dCache-Xrootd monitoring, VOMS-Xrootd

Open issues

- Concern about user privacy if user information is stored in the monitoring DB (and only there)?
- Detailed user-specification published for all users irrelevant of the VO to which they belong
  - ◆ Is it enough to filter at the collector level?
  - ◆ Is it necessary to do it at the site-level (more work needed)?

Xrootd services must be published into GOCDB

- Sites must declare downtimes when relevant

Xrootd federation services are being instrumented as all WLCG services.

## **DPM Community Status - Oliver Keeble**

Presentation postponed as we were running late: an update talk will be given at the June GDB.

The main points:

- The collaboration started on 2nd May.
- Includes contributions from CERN, Czech Rep., France, Italy, Japan, Taiwan and UK.
- A collaboration agreement has been agreed and the allocation of tasks (development, testing, support ) started.
- Support for non-HEP communities will be best effort.

# Computing Model Updates

## Introduction - I. Bird

Request from LHCC in December after the apparent significant increase of resources requests for Run 2:

- Describe changes since original computing TDR (2005)
- Emphasize what is being done to adapt to new technologies
- Produce 1 document rather than 5 (TDRs): common format, tables...
- Updated computing models should cover LS1 to LS2

Timescales

- To prepare for 2015, need a good draft to be discussed at the autumn RRB (October)
- Require a draft to be available at the end of the summer to be first discussed at the LHCC meeting in September.

This document will be an opportunity to:

- Describe significant changes and improvements already made
- Insist on commonalities between experiments in recent evolutions and plans
- Describe WLCG needs for the next 5 years
- Review our organisation
- Raise issues/concerns like staffing issues and missing skills

A draft TOC is available: see slides

Challenges

- Need to make best use of resources
- Major investment needed in SW but missing skills, tools...

Resource Needs and Evolution

- Basic assumptions on running conditions, desired physics and impact on resources
- Take flat funding as the most optimistic funding hypothesis: much less than the 20%/year growth for CPU (15% for disk) that we were used to until now (until 2015).

Collaboration Organisation and Management: should we need to be more inclusive and open to other HEP experiments?

- Also need to be collaborative with other communities (funding and efficiency)

## Technology trends - B. Panzer

See slides.

Tape: increase by steps every 2-3 year

- Currently: 2.5-5 TB/tape
- Next cycle: 6-8 TB/tape

HDD: declining market

- notebook, tablets, smartphones using flash memory
- Cloud storage leading to consolidation (less disks used compared to individual disks)

HDD current technology reaching its limit: new technologies are just emerging and are very expensive

- Price: x2 between commodity and enterprise markets
- Expect a slowdown in the space/price ratio
- New technologies will mean more TB per spindle without much improvement in sequential/random access: impact on global performance?

Network: HEP IP is 1/1000 of global IP traffic, so negligible

Processors/servers: tables and smartphones are the drivers

- 2013: 1st year number of tablets is exceeding number of laptops sold, smartphones even more and larger than standard phones
- HEP buying mainly from the server market: dominated by a limited number of very big consumers, dominated by INTEL as a provider (96% of the market, almost no competition).
- Emergence of the Microserver market, based on ARM and ATOM: competition between them is hard, no clear winner, picture changing frequently
- Moore's law still applies for the main cores
  - ◆ Specific (co-)processors are something different

With flat budgets and increasing computing requests, the only solution is to increase efficiency

- Combination of new HW combinations and SW improvements
  - ◆ HW only will not provide the 10x improvement we are looking for, coding efficiency has a major role to play
- Must take into account the side effect on global efficiency of higher processor usage factor: need to adopt a holistic view instead of concentrating on just one element
  - ◆ Impact on I/O and networks
  - ◆ High end GPUs are a niche market: desktop market decrease impacts negatively the availability and price of discrete graphic cards
  - ◆ Look at market sites programming etc. as a system

Lesson to be learnt from previous attempt to understand impact of technology evolution on our computing: look at new technology, evaluate them but avoid to concentrate on only one thing.

- I. Bird: one the lesson learnt is that we should have invested much more in SW that we did. **We have to do it now.**

## Experiment Computing Models - M. Cattaneo

Important to emphasize what has been done since initial TDR (2005) to improve efficiency of LHC experiment computing

- Sometimes a factor of 10 has been achieved for some applications (eg. CMS reconstruction)
- Computing has not been a limiting factor during Run 1

Document overview for Computing model chapters

- Will cover data flows, event rates, non-event data (e.g. DBs), software evolution, data preservation .
- Need to document that many optimisations have been done.

Ian: already very substantial input from 3 of the experiments for this chapter.

## Application SW - F. Carminati

Massive parallelism is here: we got away many times but now we probably can't anymore

Different dimensions to performance

- Multi-socket/core improves memory footprint but not throughput
- Only the micro-parallelism (instruction level //, instruction pipeline, vectors) can allow to gain a significant performance factor
  - ◆ According to OpenLab survey of industry best-practices, a factor of 10 could be gained...

For post LS1, must work on scenario like pileup=140, bunch spacing=25ns

Work to be done is very significant but a few first successful experiences already exist

- GEANT4 MT: providing a significant improvement in performance with a low penalty in a 1 thread context, first example of moving a large code base
- Several interesting results with GPU usage: NA62 L0 trigger, CMS tracking
- GaudiHive very promising

ROOT6 will be a major step forward in this direction.

Need to evolve the current Concurrency Forum to a HEP SW Collaboration

- Attract new skills/expertise
- Attract new people from the community and acknowledge them for their contribution
- Explore possibility of Horizon 2020 project on these challenges

TechPark idea: build upon the OpenLab success to provide HW and SW resources with the right connection to company engineers

- A resource for the SW collaboration driven by technology rather than demand
- Exact relationship with OpenLab still being discussed: for a company to join OpenLab it is quite a heavy process, and may be too heavy to start hence new approach being considered.

## Computing Services - I. Bird

Contents will come mostly from the work done by TEGs and related work since the Amsterdam meeting in 2010

- Describe commonalities
- Describe and justify differences between experiments

Review Tier definitions based on required functions by each experiments

- Be more flexible in the mapping of functions to sites: give "credit" for providing various services
- How to make better use of opportunistic resources
- Define the different level of services expected: not the same for all functions and for all the sites

Workflow management

- Use of pilots, needs for interfaces at sites
- Strategy for future: CE vs. cloud, glexec problem

#### Data management

- Clarify strategy for tape, disks, federations, role of data popularity
- Access control and security must be clarified
- Storage solutions/interfaces: role of HEP specific solutions vs. standard solutions
- Storage abstraction (caching) for opportunistic resource use

Also database needs, distributed services...

#### Operations and infrastructure services

- Describe coordination, tools and monitoring

Security development, in particular federated identities: are we advanced enough to mention it?

This chapter should focus on the strategy statement rather than detailed technical discussions.

Frederico: Don't forget data preservation!

- Ian: will be added, based on DPHEP proposal led by Jamie

#### Discussion about Tiers:

- MD: Is there work on-going to develop the definitions different from the hierarchy of Tiers?
- IB: Need to be pragmatic but also consider implications of changes for example prestige attached to historical contributions. Newer Tier-1s may contribute a subset of the functions.
- Roger Jones: I would be wary of this because this would lead to a renegotiation of the MoU.
- IB/General: We have sites coming scalability and reliability may be an assessment criteria. Lack of resources is an issue and some may be excluded by a naming convention if we do not address this problem. Notion of credit associated with services may be developed. If it means you have to completely redevelop your computing model then that may be an answer guiding what we can do in this area. This may include opportunistic use of other large resource offerings (via cloud or whatever mechanism) where certain other functions are excluded such as large connected storage. Beyond 2015 we may wish to revisit some of these wider questions and this would not be material for the present document, but it is a discussion worth having.

We'll come up with a further draft of this document through various computing groups in the next month(s). A solid draft of the whole document must be ready for the end of August.

- IB will write the summaries on all aspects of this chapter, based on input received from TEG, WGs and others.

---

This topic: LCG > GDBMeetingNotes20130508

Topic revision: r4 - 2013-05-27 - MichelJouvin



Copyright &© 2008-2019 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback