

# **WLCG workshop – 7<sup>th</sup> -9<sup>th</sup> July Barcelona**

**Draft notes: Jeremy Coles.**

## **1) WLCG STATUS**

**Introduction** – general introductions & logistics.

**PIC highlights** (Manuel Delfino)

<https://indico.cern.ch/event/305362/session/0/contribution/23/material/slides/0.pdf>

- Particle physics + Astrophysics + Cosmology
- dCache+Enstore. Open PBS/Torque/Maui.
- Proud of financial efficiency
- Focus has been on energy efficiency – recent containment of hot and cold air. Replaced CRACs by air-to-air heat exchangers. “Free” air cooling

Chris W observes: Notable that they support T2K - and mentioned it.  
Air conditioning - air-air cooling system payback time 5 years (approx) - even in Spanish heat. Replacing and moving condensers had a payback of about 3 years.

**Operations** – look back and review (Andrea Sciaba for team)

<https://indico.cern.ch/event/305362/session/0/contribution/1/material/slides/1.pdf>

- Reviewed background to ops coordination and recent successes and focus areas.
- Issues – site participation can be improved. Long tails in deployment campaigns. Uncertainty in middleware support. Lack of risk analysis. Manpower.
- Several WGs near completion. What else needs to be done?
- Priorities for the experiments?

... ATLAS – reduce needed operations effort. Fully enable usage of cloud resources.

-- shifting site admin from site to experiment. Virtual sites – what progress and milestones.

... CMS – provide solutions for availability monitoring, accounting, and utilization in cloud. Usage/load storage federations. All the usual...

WLCG – cloud, performance understanding, optimization, ‘virtual’ sites.

## **Tier-0 (Maite Barroso Lopez)**

<https://indico.cern.ch/event/305362/session/0/contribution/2/material/slides/1.pdf>

- Update on Wigner ( additional capacity mainly for OpenStack)
- Linux next version – adopt CentOS7. <http://seven.centos.org/>. Expect customized version in July/August.
- Networking. LHCONE expanding in Asia <https://indico.cern.ch/event/318813/>. CERN link moved to 30Gbps. Two Russian sites joined LHCOPN. IPv6 now deployed and 4 T1s have IPv6 LHCOPN connections.
- Moving from Havana-3 to Icehouse (OpenStack) – accounting with Celiometer.
- FTS2 stop on 1<sup>st</sup> August.
- Batch system migration selection:  
<https://indico.cern.ch/event/247864/session/5/contribution/22/material/slides/0.pdf> ... moving to HTCondor.
- New squid service to cover frontier as well as CVMFS.

## **MW readiness (Maarten Litmaath)**

<https://indico.cern.ch/event/305362/session/0/contribution/3/material/slides/1.pdf>

- Mandate of group to ensure infrastructure does not suffer with MW upgrade
- Reviewed meetings and defined workflow (including links with PTs and volunteer sites)
- MW officer more of a coordinator role.
- See slides for list of tasks.
- MW package reporter is work in progress – parses and selects WLCG MW rpms running at sites.

## **Multicore jobs (Alessandra Forti)**

<https://indico.cern.ch/event/305362/session/0/contribution/4/material/slides/1.pdf>

Objective: integrate single and multi-core jobs; avoid splitting resources; maximise CPU usage.

- Draining and job priorities considerations.
- Job running time – depends on work, input data, masking with pilot. Need steady flow of MC jobs. Entropy vs predictability. Draining rate has to be optimized.

- Without backfilling. Dynamic partitioning – moves nodes between single and m queues.
- Now combined submissions. (slide 24).
- Lots of useful details in slides (best to look at them directly).

### **Monitoring Evolution** (Pablo Saiz)

<https://indico.cern.ch/event/305362/session/0/contribution/5/material/slides/1.pdf>

- Past, present, future. Mon consolidation group.
- Present – in process of unifying, and improvement of deployment/operations. Eval. Other technologies. Agile infrastructure for dep. Decouple SAM. Common metric store (Nagios, hammer cloud, pledges, downtimes, SSB, rebus...). Can create derived metrics. Correlate things for Wigner analysis.
- Storage tech evaluation: ElasticSearch, hdfs, Impala, Riak
- Visualization: Ember and Angular
- Analytics: Storm, ESPER.
- CVMFS monitoring now has prototype <http://cvmfsmmon.cern.ch>. Package deployed on squid server.

ATLAS could publish on activity as well as services. E.g. you are available for production.

### **Database Evolution at CERN** (Eva Dafonte Perez)

<https://indico.cern.ch/event/305362/session/0/contribution/6/material/slides/1.pdf>

- Overview and evolution of service
- 100 databases. 500TB data files.
- DB clusters based on RAC.
- Plan to deploy Oracle Golden Gate at CERN and T1s.
- within CERN offer DataBase on Demand service.
- Looking at Hadoop (MapReduce) + Impala (SQL engine on top of HDFS)

There are some WLCG Collaboration Board slides. Worth looking through the report from Ian Bird if interested in budgets and RRB outputs (<https://indico.cern.ch/event/305362/session/4/contribution/25/material/slides/1.pdf>). Michel gave a GDB overview (<https://indico.cern.ch/event/305362/session/4/contribution/26/material/slides/0.pdf>).

## WLCG medium-term Evolution

### Cloud (Laurence Field)

<https://indico.cern.ch/event/305362/session/2/contribution/7/material/slides/1.pdf>

- Image management – balance pre/post-instantiation operations
- CERNVM – Resource management – Start, stop, status
  - ➔ Capacity Management (care about this more than ‘resource’).
- Tooling – VAC, v-cycle, Cloud schedule, glideinWMS, ALiEC2, BOINC.
- Monitoring – using Ganglia.
- Accounting/commercial – Helix Nebula > billing > consumer side accounting.
- Trust WLCG sites but check efficiency of resource usage.
- Framework & supporting services – PJ, support (squid, ganglia...)...who manages.
- EGI Federated Cloud (description)
- State of adoption – draft for August across experiments. Moving towards production.
- Questions ... regarding name CernVM and accounting.

### WLCG Network Outlook (Tony Cass for Edoardo Martelli)

<https://indico.cern.ch/event/305362/session/2/contribution/8/material/slides/1.pdf>

- CERN b. LS2. 40G server connections. Top rack -> 100G. 100G links to T1s.

Address allocations.

WLCG – b. LS2. LHCONE to more remote sites. Major T1/2s -> 100G.

>LS2 Possible use of SDN. Use of IPv6.

Upcoming 400G for datacenter interconnections. 1Tbps after 2020.

### IPv6 (Dave Kelsey)

<https://indico.cern.ch/event/305362/session/2/contribution/9/material/slides/1.pdf>

IPv6 growth – rapid. MS Azure – Ipv4 now in Brazil.

2/3 response to survey. About half of WLCG sites have no plan. T2 there are 5 sites having issues.

Testbed – data transfers.

Software readiness survey. <http://hepiv6.web.cern.ch/wlcg-applications>

Maddash now used for the dashboard.

Next steps – join tesbed; coordinate dual-stack; test IPv6 only WNs.

### **Network monitoring and metrics (Marian Babik)**

<https://indico.cern.ch/event/305362/session/2/contribution/12/material/slides/0.pdf>

Status – perfSONAR. Only 8 sites missing. Not all sites/links are monitored.

Needed metrics.

Network and transfer metrics WG: mandate – metrics coverage; understanding; tools (efficiency & optimization)

Monitoring and network control.

Find & localize the problems.

Advanced Network Services for Experiments (ANSE)

Q: Are sites going to see experiment optimization decisions – somehow they need to be alerted if they are not performing. Is this metric in scope?

### **Data Storage Evolution in Run2 (Wahid Bhimji)**

<https://indico.cern.ch/event/305362/session/2/contribution/20/material/slides/1.pdf>

Storage interfaces.

SRM. “All WLCG experiments will allow non-SRM disk-only resources by or during Run-2”.

Xrootd data federations: monitoring and zoom (interest for site)

Site tools for xrootd throttling.

HTTP/WebDav: Not as much http monitoring.

Benchmarking & I/O: number of root I/o developments.

The rest:

http>SPDY>http2 and noSQL>NewSQL

Relaxing requirements.

## **New processor evolution evaluation** (Luca Dellagnello)

<https://indico.cern.ch/event/305362/session/2/contribution/22/material/slides/0.pdf>

- Test with Avoton CPU.

CNAF – high kWh cost.

Intel Atom Pro C2000 [ark.intel.com](http://ark.intel.com)

HP Moonshot 1500 chassis. M350 server cartridges.

Avoton vs Opteron. 3 times performance but 0.5 vs 4.3 HS06/W.

ALICE – 61% slower. ATLAS 3x slower. LHCb 63-39% slower.

Pros: native x86\_64 code. Cons: More hardware.

Possibility to save 25% in TCO. Dependent on PUE of centre.

## **EXPERIMENT SESSION**

### **ALICE** (Maarten Litmaath)

- 8GB/s run2
- s/w and process improvements.
- LHCONe developments. Good progress IPv6.
- User analysis declined, organized trains increased (same input data).
- Popularity service.
- 13 working groups looking at various run3 areas.
- DATA MANAGEMENT ... reading remotely 0.7MB/s vs 1.7MB/s
- Infrastructure monitoring
- Will ask sites to move to xrootd 4 when ready. Recommend EOS.
- CVMFS monitoring:
- Clouds: [cernvm.online.cern.ch](http://cernvm.online.cern.ch). CernVM Elastic Clusters. Running HTCondor job scheduler. HLT may go with OpenStack.
- RUN3: about 1.1TB/s detector readout
- Reducing the complexity is the key. Want to replace the global file catalog.

### **ATLAS** (Simone Campana)

- LANDSCAPE Pile up 25-> 40. R3 -> 70. R4 -> 150.
- Up to 1.5M jobs waiting in March. Many Run1 data based papers still to be published.
- LIMITATIONS. Partitioning (user vs central production), memory increase, current DM & prod systems (new uses and tech). Data formats. Optimisation needed everywhere with -10% manpower. Run4 has factor 15 increase in processing time. Flexibility – some T2s>T1s in resource. Less and larger sites would be better!

- RESPONSES R2
- Simulation is CPU intensive. Mix full and fast. 64-bit 25-30% performance but memory req means going to AthenaMP. 2014 – production goes MC whereas analysis single core. xAOD read by root and Athena. Use trains (combined performance). Dec 3-5 Jamboree  
<https://indico.cern.ch/event/276502>
- Distributed computing: JEDI babysit user analysis jobs. Panda manage the execution.
- ProdSys2: imp. Data loss recovery, transient data handling, dynamic dimensioning (scout jobs)
- Rucio enables use of diff protocols. WebDAV can be used for log files, download to laptop, ...
- Site “useability” metric

### **CMS** (Christoph Wissing)

- multi-core - if thread safe lose efficiency improvements with number
- PromptRECO
- Use of HLT. “inter-fill” usage mandatory in Run2.
- Opportunistic resources
- Proposed mini-AOD.
- Bigger changes in Run4 than 3.

### **LHCb** (Stefan Roiser)

- Deferred trigger. IN Run2 split.
- Automated online calibration
- LHCb upgrade x5 in luminosity
- TWO... http is planned. LFC (replica and bookkeeping) to DFC.
- New pilot 2.0. Only CVMFS for sw deployment.
- How to find out about new resources (VM) at sites
- If marking downtime in GOCDB make sure service type correct.
- 2<sup>nd</sup> Sept CVMFS -> v2.

## **EU-T0** (Giovanni Lamanna)

Data research and innovation hub.

Hub built on computing centres which participate in WLCG. CERN and EIROforum members published a vision for the evolution of European e-Infra.

Rationale: Bring research communities closer. Share ideas/standards and avoid repetition. Ecosystem.

H2020 calls. "Datacloud" and "Data backbone, ZEPHYR".

Called T0 because of history. Dealing with raw data from many domains.

## **OSG** (Lothar Bauerdick)

In numbers. 700M hours HTC per year. 26.6 FTE.

Campus grids.

Flexible provisioning – raises new questions (usage, security....) Forum for discussion. Dropbox.

## **EGI Future Activities** (Peter Solgna)

Core services – continue supported: ops, security, portal....

EGI-ENGAGE – H2020 EINFRA-1 (6).

e-Infrastructure commons. AAI, Permanent-ID, Market place.

Ops: extend cloud activities. Cloud PaaS and IaaS.

Knowledge commons; Open data – collaborate with RDA.

## **WLCG Future** (Ian Bird)

- 10yrs from HLHC... took 10 yrs to get computing in shape for Run1. 400PB/yr by 2023. Need changes like LHCb moving reco to trigger.
- Must optimize. Reduce ops costs.
- Q: what more can we do to optimize?
- HEP SW FOUNDATION: not alone. 12 white papers received.



- CLOUDS: Strategy (cloud interfaces, pilot factories, Ces)? Needs discussion.
- OPS COSTS: reduce MW. (Volunteer support).
- EC: ICT-8 in November. Common procurement technical spec.
- Input to BD in EU
- Discuss with other communities at local institute level

Need a coherent view of what you want to do not just go for money.

WLCG – function for the community. Different as to whether infrastructure can be used more widely. e-I. Part of larger entity.

Questions: CPU per byte. Is future x86 64?