Optimising costs in WLCG operations

M. Alandes, M. Dimou, J. Flix, A. Forti, A. Sciabà

on behalf of the WLCG Operations Team
Outline

• Overview of WLCG operations
• Evaluation of the operational costs
  – WLCG site survey
  – Areas for improvement
• Communication effectiveness
• Site administration
• Possible optimisation approaches
• Conclusions
The Worldwide LHC Computing Grid project ([http://cern.ch/lcg](http://cern.ch/lcg)) supports the distributed computing for the LHC experiments and provides the resources to store, distribute and analyse the LHC data.

– Established in 2001, stable operations in 2010

<table>
<thead>
<tr>
<th>Tier</th>
<th>CPU (kHS06)</th>
<th>Disk (PB)</th>
<th>Tape (PB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Tier-0</td>
<td>687</td>
<td>49</td>
<td>95</td>
</tr>
<tr>
<td>14 Tier-1</td>
<td>1,055</td>
<td>95</td>
<td>179</td>
</tr>
<tr>
<td>160 Tier-2</td>
<td>1,354</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3,096</td>
<td>250</td>
<td>274</td>
</tr>
</tbody>
</table>
WLCG operations are the result of 10 years of evolution

- From data and service challenges to data taking
- Relatively small central team
- Strongly coupled to experiment computing operations
- Lots of voluntary contributions from sites and experiments
  - Site managers provide the bulk of the effort
  - Heavily rely on OSG, EGI, NorduGrid, GridPP, etc. and on external tools (GGUS, GOCDB, OIM, ops portals, etc.)

Already reviewed in 2011 in the Technical Evolution Group on operations and tools

- Identified weak points and recommended actions
- Led to the creation of the WLCG operations coordination team in 2012
  - Implemented all TEG recommendations
WLCG operations coordination: tasks and effort

- **Stakeholders**
  - LHC experiments, Tier-0/1/2’s, Grid projects

- **Tasks**
  - Operate the WLCG sites
  - Foster WLCG site operations stability
  - Recommend middleware versions and advise on service upgrades
  - Coordinate deployment campaigns
  - Find solutions to specific technical problems
  - Define new procedures, work plans, actions
  - Facilitate communication between experiments and sites and with the management

- **Effort**
  - Distributed coordination team
    - Stakeholder representatives
    - Task forces and working groups
    - Contributions from T0, T1’s and T2’s
  - Site operations
    - The site administrators running the day-to-day service management
Cost optimisation

• WLCG operations are perceived as still too heavy on manpower
  – Grid sites are complex to setup and maintain, require expert administrators to be stable
  – Implementing changes on the infrastructure is a long and sometimes painful process
  – Difficult to balance the limited available effort between “routine” operations and e.g. participation to task forces and testing new technologies
  – Shrinking budgets exacerbate the problems

• Mission
  – Determine if and how the effort spent on operations can be reduced

• Method
  – Start by asking the sites to provide information via a general survey
    • FTE estimates by service and activity
    • Evaluation of current tools and procedures
    • Evaluation of service administration aspects
    • Explicit suggestions and new ideas
  – ≈ 100 sites answered, many giving valuable feedback
Effort for central operations

- **Meetings**
  - One 1½ h fortnightly coordination meeting, ≈20 participants
  - Two 20’ operations meetings per week, ≈15 participants

- **Coordinators:** ≈1.5 FTE

- **Task forces and working groups**
  - Manpower is difficult to quantify
    - O(10) FTE in total
• Measured by a recent survey on WLCG sites, by category
  – Tier-0, Tier-1’s: 12.3 FTE on average
  – Tier-2’s: 2.8 FTE on average

• Other Grid services: APEL, VOMS, MyProxy, CVMFS, FTS, LFC, Frontier, Xrootd redirectors, ...
• Other WLCG-related tasks: virtualisation, security, databases, OS, hardware provisioning, IPv6, regional operations and support, planning, development, testing, validation, training, documentation, outreach, volunteer computing, ...
The most time-intensive service is **storage**

Followed by the other **core infrastructure services**: networking, monitoring, batch and computing elements

- Other grid services amount to a small fraction
- APEL dominates the “other Grid services” category

Experiment contacts use a non-negligible fraction of the effort

- Not all experiments have them at all their sites

Participation in task forces and meetings uses a very small fraction of resources

The (large) impact of “other WLCG-related tasks” at Tier-1 sites can be split into two groups

- Tasks inherent to running a large computing centre
- Support tasks for the local or the entire community
Site effort: how to improve

• **Reduce effort on core infrastructure services**
  – Mostly inherent to running a storage/compute facility
  – Some room for improvement certainly exists
    • More details from the service administration
  – Considerable gain if e.g. running a batch-less farm or “simpler” storage, it may be adequate for small sites

• **Reduce the need for experiment contacts**
  – Without increasing the effort in the experiments
  – Expected gain: minor but not negligible

• **Reduce the number of services**
  – There is no “low hanging fruit”: no services are at the same time effort-heavy and of limited use
  – Still, worth pursuing in specific scenarios
Communication: site feedback

• Work well
  – GGUS very appreciated
  – Task forces are and operations meetings are considered useful
  – Communication between central operations and Tier-1 sites
  – Reasonable information sharing among sites
  – The operations coordination meeting is followed by most Tier-1’s and the minutes are usually being read also by most Tier-2’s
    • Timezones are an obstacle to participation
  – The “3 o’clock” biweekly meeting is ≈OK with respect to frequency and usefulness

• Need improvement
  – High-level view of the WLCG plans
  – Communication and tracking of requests to sites and their deadlines
  – Consolidate information in a single entry point for documentation, recipes, procedures, known issues

<table>
<thead>
<tr>
<th>Channel</th>
<th>Most important</th>
<th>Less important</th>
<th>Least important</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGUS tickets</td>
<td>54%</td>
<td>33%</td>
<td>13%</td>
</tr>
<tr>
<td>WLCG broadcasts</td>
<td>51%</td>
<td>35%</td>
<td>14%</td>
</tr>
<tr>
<td>Operations meetings</td>
<td>4%</td>
<td>29%</td>
<td>67%</td>
</tr>
</tbody>
</table>

GGUS tickets and WLCG broadcasts are judged equally important when communicating (and tracking) requests to sites
Communication rating plots

- Is communication between site and WLCG ops effective?
- Is information sharing across WLCG sites effective?
- Do you follow the WLCG ops coord meeting?
- Do you read the minutes of the WLCG ops coord meeting?
- Are WLCG task forces useful for you?
- Are you satisfied with GGUS?
- Is the 3 o'clock meeting useful?
- Is the frequency of the 3 o'clock meeting adequate?
Service administration: plots

How easy are these aspects of the operation of the batch system?

How easy are these aspects of the operation of the WN middleware?

How easy are these aspects of the operation of the storage system?

How easy are these aspects of the operation of the computing elements?

How easy are these aspects of the operation of the local monitoring?

How easy are these aspects of the operation of the squid servers?

How easy are these aspects of the operation of the ARGUS/GUMS servers?

How easy are these aspects of the operation of the information system?

How easy are these aspects of the operation of the VO boxes?
Service administration: site feedback

• Identified areas to be improved
  – Documentation: generally OK but dispersed
  – Reconfiguration: not an issue
  – Upgrades: not an issue
  – Support from developers: not an issue, except for batch systems
  – Deployment: difficult for storage and computing elements
  – Troubleshooting: often difficult

• SEs and CEs are overall the most difficult services to operate
  – Squid and the local monitoring are the least difficult
  – Authentication servers could be better
Service administration: how to improve

• The largest gain would come from making troubleshooting easier for CEs, SEs and batch
  – Improve logging, documentation, etc.

• Other recommendations
  – Migrate to “better” implementations
  – Simpler batch system and storage configurations
  – Maintain an official, up-to-date documentation portal and build a knowledge base (searchable mailing lists, tickets, web forums, etc.)
  – Create e-groups to share information on site-specific services or issues
Other approaches

• WLCG has a large number of small sites supporting few VOs, with very limited dedicated effort

• Recent experience shows that also bringing up a new Tier-2 (however small) is quite a strenuous effort
  – Several services to deploy
  – Steep learning curve

• Still, it is essential that adding new resources (both opportunistic and pledged) becomes easier
  – But there is work that scales with the number of sites, so very small sites are not desirable

• Need a model for a “simple” but fully functional Tier-2
  – And an even simpler model for a “bare-bones” site
    • E.g. no analysis, single VO, etc. Can be applied to opportunistic resources
Ideas for a “simple” site

- One (few) computing element(s)
  - Or none at all for a “bare-bones” site
- A well supported and documented batch system with a simple configuration
  - Or no batch system at all
- A storage element which is easy to operate and supporting xrootd, HTTP/WebDAV (SRM not required)
  - Or no storage element at all?
- All client software from CVMFS
- ARGUS/GUMS + gLExec required
  - This can be relaxed if a suitable alternative is available (e.g. traceability guaranteed by the experiment)
- perfSONAR desirable (but needs to be very simple to configure)
- Virtualisation of worker and/or service nodes desirable
- Local VOboxes not needed or remotely managed by experiments
Conclusions

• Central operations largely based on voluntary effort
  – Mainly via task forces and working groups
• A first cost reduction analysis has been attempted based on results from a survey on WLCG sites
• Reduction in site effort constrained by infrastructure services
  – Still some margin exists for reduction
  – Existing expertise is an extremely valuable resource
• Overall communication level good but not yet optimal
• Service administration should be improved in deployment and troubleshooting
• Need to design a model for smaller sites, simple but functional
  – Or better two models (one Grid-like, one Cloud-like)
  – Good for pledged AND opportunistic resources
  – Simpler and/or fewer services