

Optimising costs in WLCG operations

María Alandes Pradillo¹, Maria Dimou¹, Josep Flix^{2,3}, Alessandra Forti⁴, Andrea Sciabà¹

¹ European Organization for Nuclear Research, CH-1211 Genève 23, Switzerland

² Port d'Informació Científica, E-08193 Bellaterra, Spain

³ Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, CIEMAT, Madrid, Spain

⁴ University of Manchester, Oxford Road, M13 9PL, Manchester, UK

E-mail: Andrea.Sciaba@cern.ch

Abstract. The Worldwide LHC Computing Grid project (WLCG) provides the computing and storage resources required by the LHC collaborations to store, process and analyse the 50 Petabytes of data annually generated by the LHC. The WLCG operations are coordinated by a distributed team of managers and experts and performed by people at all participating sites and from all the experiments. Several improvements in the WLCG infrastructure have been implemented during the first long LHC shutdown to prepare for the increasing needs of the experiments during Run2 and beyond. However, constraints in funding will affect not only the computing resources but also the available effort for operations. This paper presents the results of a detailed investigation on the allocation of the effort in the different areas of WLCG operations, identifies the most important sources of inefficiency and proposes viable strategies for optimising the operational cost, taking into account the current trends in the evolution of the computing infrastructure and the computing models of the experiments.

1. Introduction

The Worldwide LHC Computing Grid (WLCG) is a project established in 2001 to provide the software and the resources needed by the LHC experiments for their computing activities. After a long development and commissioning phase, it finally went into operation in 2008, in time for the initial LHC startup. Today it includes resources from about 170 sites, jointly operated by WLCG and infrastructure projects like EGI [1], Open Science Grid [2] and NorduGrid [3].

Historically, the central WLCG operations evolved from experience during the many service and data challenges organised to test the readiness of the infrastructure, which led to a successful and functionally efficient operation but at the expense of significant manpower cost. A review conducted in 2012 [4] led to several improvements, mainly via the creation of an operations coordination team, which implemented all the recommendations from the review and accomplished considerable work via task forces and working groups dealing with specific operational issues.

Recently, the increasing pressure from budgetary constraints has led the WLCG community to pursue new ways to provide computing and storage resources, aiming at reducing the operational effort as much as possible. The WLCG operations coordination group collected a wide spectrum of information from site administrators to determine which areas are the most effort-consuming and what aspects of operations should be most improved. This contribution presents the results



of this investigation and defines a set of strategies that should be pursued in the near future to achieve the desired cost reduction.

2. Current trends in computing operations

During the first long LHC shutdown the experiments have revised their computing models in several aspects. Data management and access is much more flexible than during Run1: unused data is more aggressively removed from disk, and data is more often accessed remotely via storage federations, decreasing the need for data replication. Access to opportunistic resources has become common, often by means of cloud computing techniques. Experience with cloud computing is quickly increasing, although it is not yet mature enough for large scale adoption in WLCG.

There is also an ongoing shift of complexity from sites to experiments: for example, the batch system configuration is simplified thanks to passing more job parameters at submission time; VOMS roles at sites are reduced as priorities are managed in the experiment job queues; data access protocols are consolidated (POSIX, xrootd and HTTP/WebDAV).

Infrastructure monitoring is improving, thanks to the WLCG monitoring consolidation project [5] and new tools and services for network monitoring [6]; however, it remains difficult to estimate the deployed capacity very precisely.

On the other hand, experiments have to spend too much time solving operational problems with small sites, and from their point of view it would be better to have “mature” sites acting as single entry points to regions, or even to have smaller sites federate their resources to reduce the operational overhead.

3. Site survey

There is a common perception that WLCG operations are still a too heavy process: Grid sites are complex to set up and maintain, implementing changes is often a long process and very little time is left available to participate in activities of common interest (task forces, testing new technologies, etc.), combined with a shrinking availability of manpower.

In order to quantify the impact of these issues and to collect enough information from the WLCG sites on all the relevant aspects of site and central WLCG operations, a survey was launched in November 2014, covering the following areas:

- *Site organisation*: number of supported VOs
- *Effort by service*: estimation (in FTE) of the effort spent in the operations of services and other WLCG-related tasks
- *Service upgrades and changes*: feedback on tools and procedures for middleware upgrades
- *Communication*: feedback on communication between WLCG operations and WLCG sites, on communication channels and meetings, on GGUS [7] and task forces and working groups
- *Monitoring*: feedback on site functional tests and other monitoring sources
- *Service administration*: feedback on access to documentation, deployment, upgrades and reconfiguration, troubleshooting and support.

The survey was the first done by WLCG at this level and saw the participation of 102 sites, several of which gave detailed and constructive feedback. In a few cases, further interaction was needed to clarify the meaning of some questions. Answers were not anonymous but the identity of the submitters was strictly confidential. The survey was also a valuable opportunity to enhance the interaction between sites and WLCG operations coordination and to set realistic expectations on what can be achieved.

4. WLCG Operational effort estimates and site survey results

This section offers a summary of the main results of the survey and other information collected, while a fully detailed analysis of the survey results is available online [8].

4.1. Operations coordination

It is agreed by the experiments that having coordinated projects of common tools and common deployments of new services reduces the overall effort and completion time. This is one of the primary missions of the WLCG operations coordination team and is done in its many active task forces and working groups, mostly based on voluntary effort. A general meeting is held twice per month and two shorter operations meetings are organised each week. The project coordination itself is carried out by about 1.5 FTE, while around 10 FTE are active in task forces and working groups.

4.2. Site organisation

In terms of supported virtual organisations, 38% of the Tier-1 sites support only one LHC VO, and 31% all four of them. Conversely, 44% of the Tier-2 sites support one LHC VO and 19% claim to support all of them. The average number of non-LHC VOs is similar for both Tier-1 and Tier-2 sites and is around 7 (with a maximum of 40).

4.3. Effort by service

Sites were asked to provide estimates of the operational effort spent in services and areas, expressed in *full time equivalent (FTE)*, defined as the number of hours spent on the task in a year, divided by 1,600 hours; the results are subject to large uncertainties, also due to misinterpretation, which was apparent in a few cases (*a posteriori* corrections were applied). Hence, one should be careful in drawing strong conclusions from the numbers presented.

Figure 1 shows the FTE effort for Tier-0/Tier-1 sites and Tier-2 sites for the different asked areas. Not unexpectedly, the most time-intensive service is related to storage. The Tier-0 and Tier-1 sites provide experiment and/or middleware services, as well as some significant development effort. For all sites there is a large effort in operating core infrastructure services, such as networking, monitoring, batch systems and computing elements. Experiment contact people use a non-negligible fraction of the effort, though not all experiments need them at all of their sites. Participation in task forces and meetings is very limited. The (large) impact of other WLCG-related tasks at Tier-1 sites can be split into two groups: (a) tasks inherent to running a large computing centre, and (b) support tasks for the local or the entire community. Many Tier-2 sites listed APEL in 'other Grid services' as being a service/tool which requires some attention.

The average number of FTEs to operate Tier-0/Tier-1s is 12.3, while it is 2.8 for the Tier-2 sites. Larger Tier-1 sites need a higher number of FTE, although this trend is not seen for Tier-2 sites, where the total FTE effort is not correlated with the number of supported LHC VOs. Surprisingly, the support via tickets for WLCG does not scale with the total FTE effort, though this is a really difficult quantity to estimate.

4.4. Service upgrades and changes

Sites agree that the frequency of middleware upgrades is manageable, and that the WLCG support during the process is satisfactory. The most used repositories to find packages are EPEL [9] and EMI [10] repositories. Normally, it is easy to find the right documentation, though in some cases it might be incomplete or not very detailed, which imposes some difficulties for a fraction of the sites at the time of performing a service upgrade.

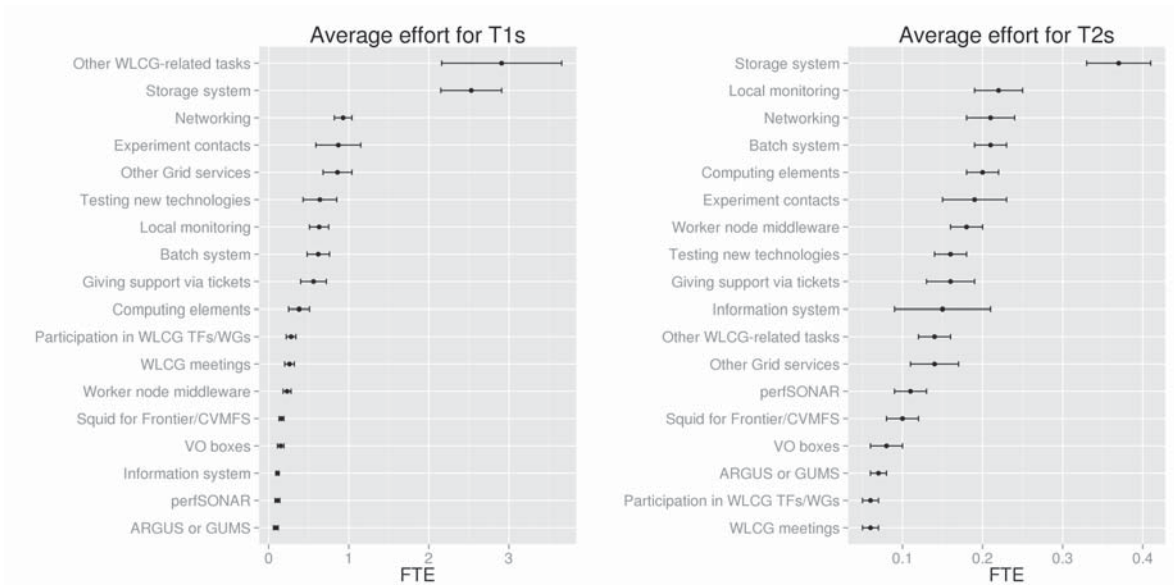


Figure 1. Site operations effort, per category, for Tier-0/Tier-1 (left) and Tier-2 (right) sites.

4.5. Communication

The feedback on several communication channels in WLCG shows to be positive. Figure 2 shows an example of the answers collected from the sites. In general, there is a good communication between central operations and Tier-1 sites, and Tier-2 sites are regularly reading relevant minutes from the WLCG operations coordination meetings. Task forces and operations meetings are considered useful, with adequate meeting frequencies. Many Tier-2 sites expressed interest in participating in task forces; however, the lack of manpower is limiting the participation. GGUS is a very appreciated tool and GGUS tickets and WLCG broadcasts are judged equally important when communicating (and tracking) requests to sites. Overall, the communication channels seem to be sufficient, with some room for improvement (discussed in the next section). Sites expressed their concerns about having to use experiment web pages with access restrictions. Those pages cannot be accessed by site administrators and they are not indexed by search engines. This imposes difficulties for sites in getting some messages from the experiments. Sites expressed interest in having dedicated sessions and forums in which they can better communicate, discuss and benefit from site experiences.

4.6. Monitoring

Monitoring and functional tests are essential to operate WLCG efficiently and provide a high quality service. The sites acknowledge the work done to consolidate WLCG monitoring tools, offering better views and faster interfaces. False positive/negatives are sometimes occurring and take some effort to identify. The sites would like more work in isolating 'site problems' from those which are related to the experiment chain. Local monitoring is essential to catch problems before experiments realise. From the survey, the lack of awareness of the SAM Nagios plugin was clear, despite it being available and ready to use and a very useful tool for sites [11].

4.7. Service administration

Feedback on tools and procedures for managing middleware upgrades was also obtained from the sites. Tasks such as service reconfigurations and upgrades do not take much operational

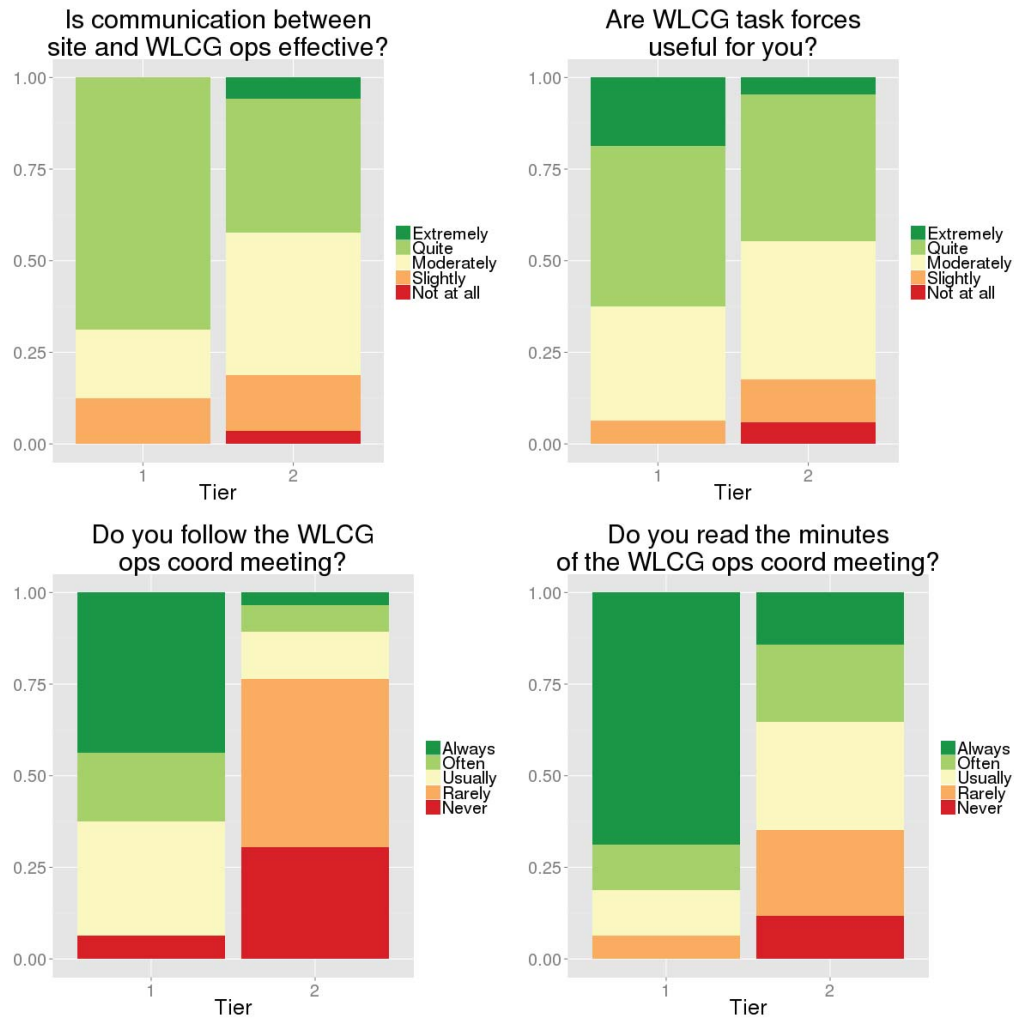


Figure 2. Communication rating plots for some of the survey questions to Tier-0/Tier-1 and Tier-2 sites.

time. The tasks that take most of the operational effort are service deployment (from scratch), in particular for storage and computing elements, and troubleshooting. There is a perception that the number of services operated should be reduced. Additionally, hardware provisioning in many cases involves tendering processes which can be very time-consuming.

In general, the support from the product developers is seen as very satisfactory, except for services whose support has ended, or products with poor documentation or poor error reporting. Discontinued products are seen as problematic, as they do not allow to offer the most up-to-date functionality and can become insecure. The most evident case as of today is the Maui cluster scheduler, which is widely used in WLCG, but whose support is almost non-existent since some years. Worries on future support for other critical components were also expressed.

5. Optimisation strategies

The general picture resulting from the survey shows that a cost reduction for operations could be achieved by acting on the following areas:

- reduce the effort needed for the existing services
- improve the communication processes
- reduce the number of services needed at a site.

In this section we describe several improvements that are expected to be effective and could be implemented in a short time scale.

5.1. Reducing effort for existing services

As we have seen in the previous section, the services that require more manpower are the core services of any computing centre: the network, the storage, the monitoring and the batch system. Among the other Grid services, computing elements are the heaviest, while the rest require individually only a small amount of manpower, although the sheer number of services makes it rather heavy to operate a Grid site.

Without reducing the set of services, the only improvement can come from reducing the effort per service. It was seen from the survey that troubleshooting is the most difficult task across the board and first deployment is hard for storage and computing services (which affects the site commissioning process). This leads to the following actions to be considered:

- improve the documentation and the logging of services;
- create a searchable WLCG documentation portal to host all the information (procedures, installation and deployment how-to's, monitoring pages, etc.) needed by WLCG sites and link to relevant web sites from local WLCG communities;
- in case the "LCG rollout" mailing list is not sufficient, create dedicated and searchable mailing lists or web forums, with the goal of establishing a community support;
- ensure that all WLCG services have a clear support channel (even if it is community support).

5.2. Improving communication

Another aspect that should be improved is the support role of WLCG operations with respect to sites, as it is not always clear what is requested of them, what are the recommended technical solutions and what is the general direction. The following changes should therefore be considered:

- clearly define the short and medium term planning, a timeline and a list of changes to be implemented, in particular by sites;
- give clear recommendations about middleware choices and other technical matters, explaining pros and cons of alternative solutions;
- extend the use of the WLCG repository, for example to distribute tested virtual machine images made to be compatible with any cloud infrastructure, or distribute middleware using container technologies.

5.3. Reducing the services

The number of services required per WLCG site has increased over the years, which produces a significant load particularly on small sites with limited manpower. This is partly due to the fact that many sites also belong to other infrastructure projects with different requirements, or have to support non-LHC virtual organisations or local user communities. Moreover, some of these services have very little or no application outside the WLCG context, which means that the know-how does not exist outside the community and they have a limited appeal for the average site administrator.

The consequence is that WLCG should strive to either find simpler alternatives, with reduced but sufficient functionality for the WLCG scope, or make some services optional or altogether unnecessary.

This approach will be described in more detail in the next section.

6. A new site model

One of the future challenges will be to find additional resources or to keep existing ones with less manpower by simplifying operations at WLCG sites. The LHC experiments have already started to extend resource provisioning to other types of resources (e.g. HPC sites and opportunistic computing); however these are not pledged resources and thus cannot form the core of the computing resources.

Existing WLCG sites have a broad spectrum of sizes and manpower; the number of services required per WLCG site has increased over the years, which produces a significant load. It is clear that a simplification would benefit everyone in general and smaller sites in particular.

Experiments would prefer to consolidate the resources in fewer bigger sites, but this is in contrast with the need to increase the resources, finding them anywhere they are offered. The Grid model has worked so far because the load was highly distributed and every site could contribute.

For existing Grid sites the simplification is not trivial, as there are not many services which require manpower that can be easily eliminated maintaining a fully functional site. As seen in figure 1, the services requiring most manpower are those at the core of a site. As mentioned before, the experiments have started to differentiate the usage of resources by running different applications on different types of resources. This increased flexibility and the evolution of technology such as cloud resource provisioning, makes it possible to introduce different types of sites other than classical Grid sites, by consolidating more manpower-demanding services at bigger sites.

The model that is emerging as most viable is that of having smaller sites running only virtualised worker nodes that pull payloads directly from the experiment submission system rather than the experiments remotely submitting pilot jobs to such sites [12]. The WNs themselves can generate the virtual machines using images provided by the experiments according to a simplified concept of fair share based only on CPU usage configured via a simple configuration file. This would remove the need to have a batch system at sites that support a limited number of experiments and do not support other VOs, in which case a batch system, with its more complex fair share implementation, is required. This model can also be integrated in existing cloud provisioning systems such as Openstack (that can be run at other sites) or in existing batch systems which could run in this way normal jobs and cloud jobs using the same fair share system. The model works also with types of virtualisation other than virtual images, such as Linux containers [13].

Smaller sites could also benefit from either being associated to bigger sites for the storage, or from the distributed storage model the experiments developed to access data from remote wherever it is, hence removing also the need for local storage. However this would put a much bigger strain on those sites offering storage and needs to be carefully evaluated, in particular to run analysis applications but also for some of the more heavy-weight production applications that require large amounts of input data.

So far a limited number of nodes have been used for a feasibility study for three of the four LHC experiments and how to fully integrate this in the experiments workflow and what applications can be run at what sites is still not completely understood. For example this new model will require a complete revision of what site availability means and how to monitor it, or the security implications of running in this way have not been fully evaluated.

7. Conclusions

Current constraints on budget and restricted manpower in many WLCG sites are strongly motivating the WLCG operations team to find ways of optimising their current processes and reducing the overall operational effort. The sites survey has now given the WLCG operations team the necessary input to start applying optimisation measures. In the survey, sites have pointed out which areas that consume a high part of their available effort could be targeted for improvement. The WLCG operations team has now the opportunity to make an impact based on the received feedback. Areas where improvements could be carried out include the reduction and simplification of services deployed at sites. The WLCG operations team should follow up with middleware developers in evaluating whether current maintenance could be made simpler and also should try to understand the feasibility of getting rid of some of the existing services. Sites have also expressed the need for improving the quality and availability of the documentation, and also for a more accessible way to exchange information within the community. WLCG operations should work on an operations portal centralising useful documentation, providing a single entry point for middleware documentation, operations announcements, future plans and site documentation, making it easier for system administrators to share knowledge and access relevant information. The WLCG operations team should also embrace common solutions that will minimise efforts at all levels, as well as encourage the adoption of industry standards, making sure the best technology is recommended in each situation. The WLCG operations team will have to define and implement a roadmap describing all these measures in detail in order to optimise the WLCG operational cost.

Acknowledgments

We would like to thank all members of the WLCG operations coordination and commissioning team for their dedication and their continued effort, the WLCG site managers who contributed to this work by means of their feedback, and Catherine Noble for her invaluable technical support in the preparation of the survey.

References

- [1] European Grid Infrastructure <http://www.egi.eu/>
- [2] Open Science Grid <http://www.opensciencegrid.org/>
- [3] NorduGrid <http://www.nordugrid.org/>
- [4] Flix J, Forti A, Girone M and Sciabà A 2014 *J. Phys.: Conf. Series* **513** 062044
- [5] Saiz P *et al* 2015 WLCG Monitoring, consolidation and further evolution, *J. Phys.: Conf. Series*
- [6] Campana S *et al* 2014 *J. Phys.: Conf. Series* **513** 062008
McKee S, Babik M *et al* 2015 Integrating network and transfer metrics to optimize transfer efficiency and experiment workflows, *J. Phys.: Conf. Series*
- [7] <https://ggus.eu/>
- [8] <https://twiki.cern.ch/twiki/bin/view/LCG/WLCGSiteSurvey>
- [9] <https://fedoraproject.org/wiki/EPEL>
- [10] <http://emisofit.web.cern.ch/emisofit/>
- [11] <https://github.com/jcasals/nagios-plugins-lcgsam>
- [12] McNab A *et al* 2014 *J. Phys.: Conf. Series* **513** 032065
- [13] <http://www.docker.com/>