Operating the Worldwide LHC Computing Grid: current and future challenges

J Flix Molina[1], A Forti[2], M Girone[3] and A Sciabà[3]

[1] Port d'Informació Científica, E-08193 Bellaterra, Spain
[2] University of Manchester, Oxford Road, M13 9PL, Manchester, UK
[3] European Organization for Nuclear Research, CH-1211 Genève 23, Switzerland

E-mail: `Andrea.Sciaba@cern.ch`

Abstract. The Wordwide LHC Computing Grid project (WLCG) provides the computing and storage resources required by the LHC collaborations to store, process and analyse their data. It includes almost 200,000 CPU cores, 200 PB of disk storage and 200 PB of tape storage distributed among more than 150 sites. The WLCG operations team is responsible for several essential tasks, such as the coordination of testing and deployment of Grid middleware and services, communication with the experiments and the sites, followup and resolution of operational issues and medium/long term planning. In 2012 WLCG critically reviewed all operational procedures and restructured the organisation of the operations team as a more coherent effort in order to improve its efficiency. In this paper we describe how the new organisation works, its recent successes and the changes to be implemented during the long LHC shutdown in preparation for the LHC Run 2.

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 first LHC startup. Today it includes resources from more than 150 sites, jointly operated by WLCG and infrastructure projects like EGI [1], Open Science Grid [2] and NorduGrid [3].

The resources and services provided by sites vary according to their role in the computing model of each LHC experiment they support. The models are still largely hierarchical as by the initial design of the MONARC project [4]: typically, the Tier-0 at CERN is used for first pass reconstruction and long term raw data archiving; about a dozen Tier-1 sites for reconstruction, reprocessing and as secondary tape archive; more than a hundred Tier-2 sites for data analysis and simulation. Besides the compute and storage resources, there are several central services, hosted at the Tier-0 and/or the Tier-1 sites.

Historically, the central WLCG operational procedures were developed from practice during the many service and data challenges organised to test the readiness of the infrastructure. As a consequence, they evolved more as the coordination of the effort of many individuals than as the result of careful advance planning. This led to a successful and functionally efficient operation but at the expense of significant manpower cost. Therefore in 2011 WLCG decided to review every aspect of the project by identifying shortcomings and areas of improvement and to propose future strategies and new solutions.

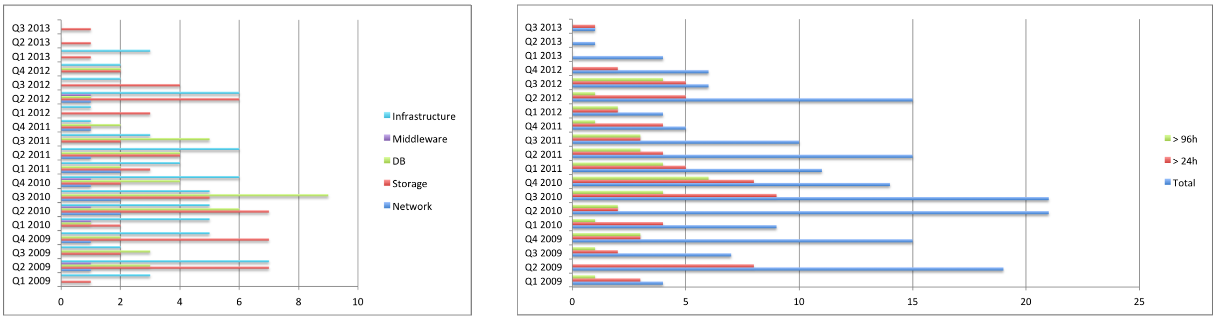

Figure 1. Service incidents by area (left) and by time to resolution (right).

The challenges to face were (and still are): to keep operating the system reliably enough with significantly less effort, and to evolve towards new, more open and flexible computing models.

In this paper we summarise the findings and the conclusions of the WLCG technical evolution group on operations and tools; then we describe how the final recommendations were implemented, mainly by the creation of a WLCG operations coordination group; we give an overview of the activities and the achievements of the group and the associated task forces; and finally we concentrate on the changes to be implemented during the current LHC shutdown and beyond.

2. Review of WLCG operations

The process of reviewing the status of the WLCG operations took place between October 2011 and April 2012 and was summarised in a document [5]. A full summary of the review is outside the scope of this contribution: we will instead concentrate on the most important operational aspects that were found in need of improvement. The following sub-areas were identified:

- monitoring and site quality metrics
- operational procedures and support tools
- management of application software
- middleware validation and deployment.

Most monitoring issues are now dealt by a separate project and will not be covered here.

Operational procedures were found to be more than adequate to ensure reasonably smooth operations, as it can be concluded by the number and resolution time of serious incidents (figure 1). Still, the occurrence of incidents was not decreasing, implying that a limit was reached that could not be overcome without changes. Incidents were discussed on a daily basis at the WLCG operations meeting, with a special focus on Tier-1 sites and experiment services. Service incident reports were produced for particularly serious incidents. Ticketing systems as GGUS and Savannah were used to keep track of problems and discuss the solution. A fortnighly Tier-1 service coordination meeting was held to discuss issues, past and future service changes and experiment plans with a special focus on Tier-1 sites. The weakest points were considered to be:

- the lack of an effective communication with Tier-2 sites
- the lack of a central WLCG operations team taking care of driving all required actions
- the lack of a committee, including experiment and site representatives, entitled to take operative decisions concerning WLCG operations (but leaving strategic choices to the WLCG management board).

Table 1. Recommendations for improving the WLCG operations.

Description
Establish a core team for coordinating WLCG operations
Expand the scope of existing meetings to fully involve Tier-2 sites
Adopt CVMFS to distribute experiment software and middleware at sites
Simplify the middleware stack and improve documentation and procedures
Strengthen the participation of sites and experiments to the commissioning

Concerning the management of experiment software, it was observed that there was a multiplicity of software installation and runtime configuration tools, some of which causing serious problems at some sites (for example when causing intensive I/O on a shared filesystem). The recent experience of ATLAS and LHCb with CVMFS was considered encouraging enough to propose it as a replacement for previous systems, due to a much better scalability and robustness with respect to shared filesystems like NFS.

Another area of improvement was identified in the impact of middleware on operations. Software which is poorly documented, producing insufficient logging information, difficult to configure and inherently fragile generates a huge operational load. Therefore, a consolidation of the middleware stack (both in terms of architecture and code), better packaging and support for popular configuration tools (e.g. Puppet) and further investment in good documentation and procedures were seen as highly desirable.

Finally, middleware validation and rollout were considered. Much of the process was taken care of by infrastructure projects (EGI, OSG, NorduGrid), while WLCG's role is much more focused towards its applications, whereby the experiments themselves tested their frameworks with new versions of some clients or services. Given the cost of rolling back to previous versions in case of incompatible versions being deployed, it was recommended to expand the test infrastructure and provide proper incentive to participating sites. Middleware deployment would also take advantage from consolidating the several repositories available to sites (EMI, UMD, EPEL, etc.) and from providing some sort of "WLCG distribution" of the middleware.

Table 1 summarises the main recommendations of the review (amended from those related to monitoring).

3. The operations coordination working group

In order to proceed with the implementation of the recommendations described in the previous section, in autumn 2012 the WLCG Service Operations, Coordination and Commissioning Team was established. This group acts as the core operations and deployment coordination team and manages ongoing operational issues as well as new deployments, in synergy with the EGI and OSG operations teams. Its goals are specifically:

- discuss the plans and the needs of the experiments for their computing operations
- define a work plan and the actions to be pursued
- form task forces for specific issues, with members from experiments, service developers and operators
- ensure an effective communication among experiments, sites, infrastructure projects via their representatives in the group, mailing lists, web pages and meetings.

The composition of the group consists of four chairs, a secretary, representatives for the four LHC experiments, for regions of Tier-2 sites, for the Tier-1 sites and for the infastructure projects (EGI and OSG). The group meets every two weeks to discuss the status of the ongoing activities and any issue of relevance for operations. Once every three months, a planning meeting is organised, to discuss the experiment plans on such timescale, approve the creation of new task forces or declare them completed.

Given the scarcity of manpower, the group is largely based on the contribution from volunteers, most of which from sites other than CERN. This model also has the advantage of ensuring a closer involvement of external sites in WLCG.

3.1. Task forces

CVMFS deployment. CVMFS is a caching, http based read-only filesystem optimised for delivering experiment software to (virtual) machines [6]. Originally developed as part of the CernVM project, it gained popularity in being used in physical batch nodes to serve the LHC experiment software, since the caches can persist longer on physical machines. This implementation removes the need for local software install jobs and shared software areas at every WLCG site. At the sites, the CVMFS system scales very easily with additional squid caches. Although the load on the squid servers is small, the sites typically have, at least, a pair of them for load balancing and redundancy.

CVMFS tests for serving experiment software started in 2010 and it was progressively adopted by ATLAS, LHCb and CMS. Early in 2013 it was decided that CVMFS should be the preferred method for serving the experiment software at all WLCG sites. CERN hosts a central repository where the experiments place the software releases to be distributed (the so-called Stratum 0 repository). The RAL Tier-1 hosts a replica of the repository (Stratum 1 repository).

The aim of the CVMFS task force is to help in the full deployment of CVMFS for software and other data at all sites. The coordination, progress reports and issues discovered by the task force are regularly discussed. At the moment of this writing, most of the ATLAS, CMS and LHCb sites are running CVMFS and ALICE is close to its completion as well.

perfSONAR deployment. Data distribution and access models of the LHC experiments have been evolving from the original hierarchical MONARC model to a full mesh model. In addition, Tier-2 sites are acquiring a greater role, as they become more reliable and better connected, which made storage federations a viable solution. The proof is the increased throughput of the past few years.

While in the hierarchical model the LHCOPN network connecting the Tier-0 and the Tier-1 sites and its monitoring were the key components, in the mesh model there are no boundaries and the network becomes important at every level. Hence the need for a WCLG-wide, experiment independent network monitoring system capable of identifying problems on the network paths between sites, finding (to the extent possible) when and where they occur, alerting in case of significant changes, setting expectations about what is possible and expected and providing network metrics to existing and future services.

The perfSONAR-PS tookit [7] was chosen to monitor lower level networking (bandwidth, latency, packet loss, routing) and present them via a dedicated dashboard. The results should be combined with the FTS transfers monitoring to give a complete picture of the network status at all layers.

The role of the task force is to help sites install perfSONAR, simplify the installation and the amount of manual intervention, define the different types of tests and work on the monitoring dashboards and metrics. The most challenging part of setting up an instance is to add sites to the tests; centrally configured mesh tests were introduced so that experiments can setup different tests according to their policies and requirements without needing the intervention of the system administrators.

Journal of Physics: Conference Series **513** (2014) 062044 doi:10.1088/1742-6596/513/6/062044

SHA-2 deployment. The usage of the SHA-1 hash algorithm, currently the default in the WLCG community, is increasingly seen as dangerous due to known weaknesses which might soon pave the road to attacks. As a consequence, all the Grid projects participating in WLCG and the International Grid Trust Federation (IGTF) agreed to move to the SHA-2 algorithm as default for new certificates by December 2013. This requires that all Grid middleware and services used by experiments are tested to work with SHA-2 certificates. On top of the certification done by middleware developers, EGI, NorduGrid and OSG, the task force coordinates testing within the experiments of their own services and in some cases of the full workflow and data management chains.

gLExec deployment. The pilot job philosophy of serially pulling workflows from a remote location at execution time ensures flexibility in scheduling and executing workflows, but it has serious security issues related to traceability and user banning.

The pilot job has no knowledge of the payload it will eventually execute when it starts running. A site wishing to pinpoint the source of a suspicious workflow must currently contact an experiment representative. In addition a site does not have the ability to block individuals or groups inside a virtual organisation. Less important to the individual sites, but important for the security of the experiments and the WLCG is the lack of reasonable sandboxing between the pilot credentials and the user workflows. While unlikely, it is potentially possible to take control of a pilot from the downloaded workflow. The two issues are both addressed by the consistent use of gLExec [8], which authenticates and logs the change of workflows with a user credential and changes the ownership of the user environment to separate it from the pilot environment.

After many years of deployment WLCG is close to complete the gLExec service. Functional tests monitoring the health of the service are in production. The LHC experiments are either capable of using gLExec now or are actively developing support for it in their frameworks. The current schedule from the Operations Coordination Team is to monitor glexec as a critical test by the end of 2013, and the experiments should use the following year to finalise the adoption.

Tracking tools evolution. As the day-to-day communication in WLCG operations happens mostly via ticketing systems, it is important to make sure that they fulfill the needs of the project. Therefore, this task force is also formed by developers and service managers for GGUS, Service Now, Savannah and JIRA. A large part of the activity is devoted to the final decommissioning of Savannah, to be replaced by GGUS or the JIRA service at CERN, and to the addition of new GGUS features and policies in GGUS, typically in agreement with EGI.

Squid monitoring. Squid caches [9] are widely used by the LHC experiments on WLCG sites. These caches allow for remote access of detector conditions data via Frontier [10], used by ATLAS and CMS. Additionally they allow serving collaboration software via CVMFS, as described earlier. As a result, squid cache deployment across WLCG has rapidly increased, with hundreds of squid servers now in production.

While some effort had been made to share monitoring tools among experiments, this happened on an ad-hoc basis, and it was found that some sites had duplicated monitoring, while others had none whatsoever for their cache service. Therefore a task force was established to decide how to improve and better deploy squid monitoring within WLCG common operations and produce an architecture for a common squid monitoring system configuration. The task force concluded, providing a list of recommendations and the subsequent work necessary to implement them. For example, the setup of a web page for monitoring of WLCG squid servers, the registration of squid servers in GOCDB and OIM and a setup of functional tests for squid services.

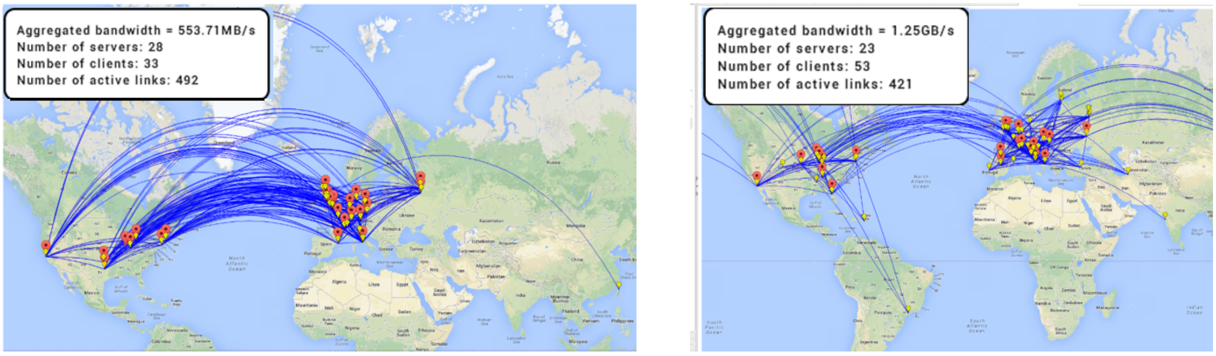

Figure 2. xrootd monitoring for FAX (left) and for AAA (right).

FTS-3 integration and deployment. The File Transfer Service (FTS) is the service responsible for distributing the majority of LHC data across the WLCG infrastructure [11]. Several years of operational experience have shown the importance of reliability, robustness and performance in data transfers. The latest version of FTS addresses most of the problems and limitations found in FTS-2. Among other features, it provides a simple interface for transfer job submission and status retrieval, advanced monitoring capabilities, support for multiple data access and transfer protocols and for more database backends and a simpler configuration. The FTS-3 service has already undergone extensive pre-production validation and a pilot service has been setup.

FTS-3 is scheduled to become the new data movement service for the WLCG infrastructure. Currently, FTS-3 pilot instances are installed at CERN, RAL, ASGC, BNL and PIC for functional testing (the CERN is using Oracle as backend and the rest is using MySQL). High volume production transfers have been already performed on the pilot service. The FTS-3 task force helps with integrating and validating the results of the pilot tests and with debugging of the software, by bringing together the service developers, the sites involved on the tests and the WLCG experiments. Under the task force, demos on how to install and operate the service have been provided, as well as many discussions on the transfer tests, aimed to progressively ramp up the load to achieve as realistic conditions as possible, achieving breakthrough performances. The goal is the delivery of a production-like service, which could be easily deployed in production in WLCG, better performant, and without any disruption of the transfer service.

xrootd deployment. The main goal of the xrootd deployment task force is to provide a common coordination body to support the ATLAS and CMS experiments in the deployment of their xrootd based infrastructures. These projects, the Federated ATLAS xrootd (FAX) for ATLAS [12] and Any Data Anytime Anywhere (AAA) for CMS [13] are internally driven by the experiments, but they share common needs and to a large extent also common deployment issues. The task force targets to identify these commonalities and to act as a catalyser of shared solutions on deployment issues. Another area of common interests and work is monitoring: the task force is responsible for collecting monitoring requirements and liaises to the monitoring development efforts [14] for a common layout, as shown in figure 2.

Scientific Linux 6 migration. In the past, software upgrades for WLCG were rarely coordinated among sites and experiments and this made the process somewhat chaotic. Therefore for the upgrade to SL6 it was decided to have it driven by a dedicated task force, including Grid experts and software librarians from the experiments, sites representatives, software producers and people from the CERN IT department. The task force had three stages:

Journal of Physics: Conference Series **513** (2014) 062044 doi:10.1088/1742-6596/513/6/062044

(i) set a timeline agreed by all experiments and a set of procedures sites should follow to upgrade;
(ii) make sure experiments were ready;
(iii) follow and coordinate the sites upgrade tracking the status on a web page.

The first stage also defined the experiment strategies (rolling vs. fixed upgrades, mixed SL5/SL6 queues, time constraints, etc.). Sites where encouraged to start testing but upgrade only between June 1 and October 31 2013, after which only a small number of sites should be left to upgrade. Communication with sites was done following existing experiment procedures. It was decided to also start a WLCG repository where to put *HEP_OSlibs*, a metapackage used to install the experiment software dependencies, and to collect other Grid software that was not easily available before. Until June it was also ensured that the experiments were ready, that their procedures worked and to address any issues.

The last stage started in June and it is still ongoing. At the moment of writing 10 over 16 current of future Tier-1 sites and 62% of the Tier-2 sites have completed the migration, and the expectation for the end of October is respectively 15/16 and 77%. This means that since June an average of one site every 1.3 days has moved to SL6 with no major disruption in the experiments operations. The majority of the remaining sites are likely to move in November with not much delay.

Machine/Job features. The job environment typical of WLCG resources offers very little information about the capabilities of the processing nodes and this can lead to a suboptimal utilisation of the resources. For this reason a mechanism has been defined to provide both static (number of cores, CPU power, local disk space, etc.) and dynamic (residual run time or CPU time for the job, virtual machine residual lifetime, etc.) information to running jobs via the machine environment. This involves agreeing on a uniform interface, writing plugins for the most common batch systems, validating the information via monitoring probles and coordinating the deployment on a wide scale. Currently, the task force is well into the implementation phase despite having started only very recently.

4. Future challenges

The WLCG operations coordination group is charged with implementing the changes that the WLCG management with the experiments agree to pursue. Of course, this applies in particular to the evolution of the computing models which will have as goal a much more efficient usage of the resources and a reduction of the operational costs. Several activities in this direction have started or are expected to start very soon.

In the timescale of the Long Shutdown 1 (which will last until February 2015) the main focus will be on the decommissioning of FTS-2, the expansion of the storage federations and the implementation of dynamic data placement techniques (currently used only by ATLAS); this will allow for a much more efficient data access and transfer. In particular, the two main storage federations, AAA for CMS and FAX for ATLAS have several commonalities and would be best coordinated in a WLCG context.

On the computing side, experiments are increasingly committed to integrating opportunistic resources (trigger farms, supercomputers, commercial clouds, etc.) in their frameworks. Currently, as these resources are by definition not part of WLCG, the operations coordination group is not playing a large role in these developments. However, there is reason to think that in the near future also WLCG sites will choose to adopt cloud interfaces, in which case the group will certainly need to be involved in the transition. On a shorter timescale the deployment of multicore queues will be addressed, also profiting from the work of the machine/job features task force.

Other evolution processes are taking place and will eventually impact WLCG operations. Work has already started to assess the IPv6 readiness of the middleware and the experiment applications in collaboration with the HEPiX IPv6 working group [15]. The significant reduction of EU funding for Grid projects is also posing a challenge and in particular it will require WLCG to develop a sustainable process to validate (and to some extent distribute) the needed middleware. Last, but not least, new hardware technologies and computing architectures are being seriously evaluated and their likely integration in the production infrastructure will need to be properly coordinated.

Acknowledgements
We would like to thank all members of the WLCG operations coordination and commissioning team for their dedication and their continued effort.

References

[1] European Grid Infrastructure `http://www.egi.eu/`
[2] Open Science Grid `http://www.opensciencegrid.org/`
[3] NorduGrid `http://www.nordugrid.org/`
[4] The MONARC project `http://monarc.web.cern.ch/MONARC/`
[5] Girone M and Templon J 2012 Final report of the operations & tools technical evolution group `https://espace.cern.ch/WLCG-document-repository/`
[6] Blomer J, Aguado-Sánchez C, Buncic P and Harutyunyan A 2011 *J. Phys: Conf. Series* **331**
[7] Tierney B *et al* 2009 perfSONAR: Instantiating a Global Network Measurement Framework *4th Workshop on Real Overlays and Distributed Systems (ROADS'09)* `http://roads.mytestbed.net/`
[8] Groep D, Koeroo O and Venekamp G 2008 *J. Phys.: Conf. Ser.* **119** 062032
[9] `http://squid-cache.org/`
[10] Blumenfeld B, Dykstra D, Lueking L and Wicklund E *et al* 2007 *J. Phys.: Conf. Ser.* **119** 072007
[11] Salichos M, Keeble O, Alvarez Ayllon A and Kamil Simon M 2014 FTS3 – Robust, simplified and high-performance data movement service for WLCG *J. Phys.: Conf. Ser.*
[12] Gardner R W and Vukotic I 2014 Data Federation Strategies for ATLAS Using XRootD *J. Phys.: Conf. Ser.*
[13] Bloom K 2014 CMS Use of a Data Federation *J. Phys.: Conf. Ser.*
[14] Beche A *et al* 2014 Monitoring of large-scale federated data storage: XRootD and beyond *J. Phys.: Conf. Ser.*
[15] Kelsey D *et al* 2014 WLCG and IPv6 - the HEPiX IPv6 working group *J. Phys.: Conf. Ser.*