New solutions for large scale functional tests in the WLCG infrastructure with SAM/Nagios: the experiments experience

J Andreeva[1], P Dhara[2], A Di Girolamo[1], A Kakkar[3], M Litmaath[1], N Magini[1], G Negri[1], S Ramachandran[4], S Roiser[1], P Saiz[1], M D Saiz Santos[1], B Sarkar[5], J Schovancova[6], A Sciabà[1] and A Wakankar[3]

[1] European Organization for Nuclear Research, CH-1211 Genève 23, Switzerland
[2] Variable Energy Cyclotron Centre, 1/AF, Bidhan Nagar, Kolkata - 700 064, India
[3] Bhabha Atomic Research Centre, Trombay, Mumbai - 400 085, India
[4] Indira Gandhi Centre for Atomic Research, Kalpakkam, Tamilnadu - 603102, India
[5] Department of Atomic Energy, Anushakti Bhavan, C.S.M. Marg, Mumbai - 400 001, India
[6] Institute of Physics, Academy of Sciences of the Czech Republic, Na Slovance 2, CZ-18221 Prague 8, Czech Republic

E-mail: Andrea.Sciaba@cern.ch

Abstract. Since several years the LHC experiments rely on the WLCG Service Availability Monitoring framework (SAM) to run functional tests on their distributed computing systems. The SAM tests have become an essential tool to measure the reliability of the Grid infrastructure and to ensure reliable computing operations, both for the sites and the experiments. Recently the old SAM framework was replaced with a completely new system based on Nagios and ActiveMQ to better support the transition to EGI and to its more distributed infrastructure support model and to implement several scalability and functionality enhancements. This required all LHC experiments and the WLCG support teams to migrate their tests, to acquire expertise on the new system, to validate the new availability and reliability computations and to adopt new visualisation tools. In this contribution we describe in detail the current state of the art of functional testing in WLCG: how the experiments use the new SAM/Nagios framework, the advanced functionality made available by the new framework and the future developments that are foreseen, with a strong focus on the improvements in terms of stability and flexibility brought by the new system.

1. Introduction

The four main LHC experiments, ALICE, ATLAS, CMS and LHCb, rely on a vast infrastructure, operated by the Worldwide LHC Computing Grid project (WLCG) [1], for most of their offline activities. The WLCG is a federation of three Grid projects: EGI [2], OSG [3] and NorduGrid [4] and includes about 150 computing sites, organized in a Tier-0 site at CERN, 11 Tier-1 sites and approximately 140 Tier-2 sites.

To accomplish their physics analysis program, the experiments need high availability of the computing centres they use, both in terms of storage and of processing resources, and in particular the experiments need to know when a site is not correctly functioning, as this has an impact on their computing operations and on the users.

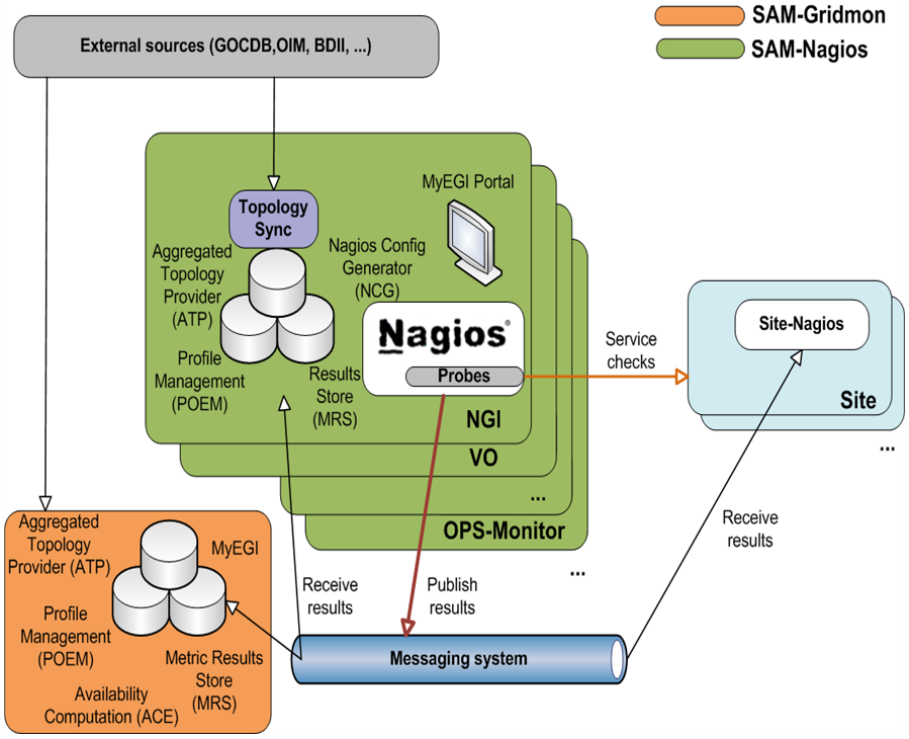

Figure 1. The new SAM architecture [5].

By agreeing to the WLCG Memorandum of Understanding, the sites commit to provide a certain level of availability and reliability. The availability of a site is the amount of time a site is successfully passing a certain set of "critical" tests divided by the duration of the total period, while the reliability is the availability of the site calculated only when it is not undergoing a scheduled intervention.

Monitoring of the status of the site services is very important and to accomplish this task a new SAM framework, more robust and flexible than the previous one, has been developed to address present and future requirements [5].

2. The SAM framework

The experiments migrated their tests from the old SAM system, based on a custom client/server system, to a new framework based on the Nagios monitoring system for the test submission. The new SAM architecture is represented in figure 1.

Submission of tests is accomplished via a *Nagios server* to remote hosts and services via the appropriate Grid interfaces. The test results are published into a messaging system (MSG) and consumed by the *metrics results store (MRS)*. Topology information from various sources is combined by the *aggregated tolopogy provider (ATP)*. The *availability computation engine (ACE)* uses the metric results, the topology and the profile definitions to calculate the site availability and reliability according to a well defined algorithm. All the stored information is made available via a programmatic interface.

The new SAM framework allows to define different sets of tests, aggregated in *profiles*, and compute the availability and reliability for each of these profiles, while the old SAM was letting the experiments choose only one list of critical tests on which the site availability and reliability were calculated. Each experiment has defined one of these profiles as the one to be reported to

the WLCG Management Board each month and compared to the MoU targets, but it may also use other profiles for its own internal computing operations. For example the ATLAS profile for the WLCG management contains just a few basic storage tests (copy from/to the Nagios server to/from a storage element), a job submission test and a local software installation test, while many more tests are submitted and included in other profiles.

The management of these profiles is now done centrally, but in few months from now, after the introduction of POEM (Profile Management), the experiments will be able to directly add or edit profiles by themselves.

The topology is now provided by each experiment through a *VO feed*. VO feeds are XML files published via the Web and used by the LHC VOs to define which are the services and the sites they use. At the same time, these feeds are used to define *service groups* relevant for the VO. Service groups can be used to query the SAM database and extract the corresponding list of services. Examples of service groups can be "clouds" (here a cloud means a Tier-1 site along with a group of dependent Tier-2/3 sites), level-N tiers or simply custom names for sites.

Presently it is possible to test and monitor only services defined in the GOCDB [6] or OIM [7] and endpoints that are properly publishing in the Information System the services they provide, but the feasibility of providing calculation of availability for meta-services is under discussion. This would address two important uses cases:

- a VO wants to calculate the site availability based on metrics that cannot be attached to particular services, for example the success rate of HammerCloud jobs [8] or any custom metric published on the Dashboard Site Status Board [9];
- a VO needs to test services that are not (yet) defined in GOCDB/OIM; examples are Frontier servers, Squid, CVMFS and Xrootd.

In terms of probes, some of them were rewritten while for others a wrapper was provided to be able to run old-style SAM tests in the new framework. In particular, the SRM tests were rewritten as a common shared test used now by ATLAS, CMS and LHCb; this probe consists of a set of functions to interact with the storage using the Grid File Access Library API, with a specific function for each experiment to get the configuration to be tested for each service endpoint and the correct logical-to-physical file name translation.

All the experiments use the standard Computing Element probes provided by SAM to check the job submission functionality via the EMI Workload Management System or directly via the CREAM computing element and to package and dispatch to the remote site the tests to be run on the worker nodes.

In Nagios the frequency of the tests, the test timeouts and the number of retries in case of failure can be customized by each experiment; for example ATLAS decided to run the SRM tests every 30 minutes and CMS has longer timeouts on the worker node and for the job submission tests.

A new functionality of the system is the possibility to directly publish test results in the WLCG messaging system MSG using Apache ActiveMQ [10], allowing the experiments to report in SAM metrics that are not run inside the framework like the HammerCloud tests.

The experiments migrated to the new SAM in February 2012, after a period of three months in which the results were carefully compared and the few differences explained satisfactorily.

3. Experiment tests and configuration

Each experiment runs several tests, some of which are standard generic probes and others are developed by the experiments and testing functionality specific to them. In this section we describe the tests run by the experiments and specific configuration choices.

Table 1. SAM tests run by ALICE.

Test name	Functionality	Standard
org.sam.CREAMCE-DirectJobSubmit	Direct job submission to CE	yes
org.sam.CREAMCE-JobSubmit	Indirect job submission to CE	yes
org.sam.WN-SoftVer	Deployed middleware version	yes
org.sam.WN-sft-vo-swdir	Accessibility of software area	yes

3.1. ALICE

Since many years the ALICE experiment has been using its own integrated monitoring system based on MonALISA [11] and covering the needs of ALICE operations to a large extent. The SAM framework is being used for tests that are complementary and whose results are presented in a standard way, thereby allowing for comparison with results of tests run by the other experiments and the ones run by EGI. The SAM tests for ALICE currently include CREAM CE job submissions, both direct, as used for normal ALICE jobs, and indirect ones via an EMI WMS. The latter path is currently required in SAM to allow additional tests to be run on a worker node at the target site. The set of worker node tests for ALICE include only a trivial middleware version check and a test of the availability of the software area dedicated to the experiment. The set of tests are summarized in table 1; it can be seen that only tests included in the standard SAM probes are used.

The storage services for ALICE are not being tested through SAM because their type, XRootD, is not yet supported by the SAM framework. That is expected to change in the second half of 2012 and would allow the corresponding MonALISA test results to be forwarded into SAM and thus provide a more complete picture of how each of the tested sites is doing.

A final service to be tested would be the gLite VOBOX, deployed on the majority of ALICE sites and critical for their operations. A set of VOBOX tests were used in the old SAM framework, but will need to be reconsidered and made more robust before they can be included in the current framework. The ALICE VO feed provides the selection of all service endpoints to be tested, taken from the AliEn LDAP server [12].

3.2. ATLAS

ATLAS uses SAM to monitor the availability and reliability of the sites since 2007, running both standard and custom probes. For what concerns the processing resources (Computing Elements) the jobs are submitted using standard probes via the EMI WMS to each CE. This allows the availability to be measured for each individual CE, which currently is not possible via the pilot-based workload management system used for production and analysis (PanDA) [13]. There is an ongoing development to make those tests more realistic: pilot jobs landing on worker nodes will run the ATLAS tests and directly publish their results to SAM to allow for the full chain to be measured as seen from the ATLAS experiment and bypassing Nagios and the EMI WMS. The full list of the ATLAS SAM tests is in table 2.

For the SRM services ATLAS uses the standard test developed in collaboration between the SAM team and the experiments, now shared between ATLAS, CMS and LHCb. The *GetATLASInfo* test is the only storage test specific to ATLAS and it relies on the ATLAS Grid Information System to know which space tokens on each SRM should be tested. The SRM tests are running every 30 minutes, and, in case of failures, each test is tried two times to avoid that glitches would affect the availability calculation.

ATLAS is feeding SAM also with tests run outside the Nagios framework and which publish

Table 2. SAM tests run by ATLAS.

Test name	Functionality	Standard
org.sam.(CREAM)CE-JobSubmit	Job submission to CE	yes
org.sam.glexec.(CREAM)CE-JobSubmit	Job submission to CE for gLExec test	yes
org.atlas.WN-swtag	ATLAS local software installation	no
org.atlas.WN-swspace	ATLAS software area	no
org.atlas.WN-LocalFileAccess	Local file access	no
org.atlas.WN-gangarobot_wms	HammerCloud test via EMI WMS	no
org.atlas.WN-gangarobot_panda	HammerCloud test via PanDA	no
org.atlas.WN-FrontierSquid	Local Squid server	no
org.sam.glexec.WN-gLExec	gLExec identity change	no
org.atlas.GetATLASInfo	Get the list of SRM space tokens	no
org.atlas.SRM-VOPut	SRM copy from client to SE	yes
org.atlas.SRM-VOGet	SRM copy from SE to client	yes
org.atlas.SRM-VODel	SRM deletion from SE	yes
org.atlas.SRM-VOGetTURLs	SRM getTURL method	yes

their results directly in the MSG: a notable example is represented by the HammerCloud test results, published in a test called *WN-gangarobot*, which is sent both via the EMI WMS and via the PanDA framework.

The ATLAS VO feed is generated dynamically from the ATLAS Grid Information System which queries the GOCDB, OIM, the WLCG information system and the Atlas Grid Information System (AGIS) [14] to generate the proper topology.

3.3. CMS

CMS uses SAM in production since 2007 to run both standard and custom tests for the computing element (CE) and the storage (SRM). Standard tests are used for job submission and for SRM (in the new version shared with ATLAS and LHCb); only the SRM test that translates a logical file name into a physical file name is specific to CMS, as it relies on the CMS trivial file catalogue (a collection of translation rules based on pattern matching having the role of local file catalogue). Currently the job submission proceeds via a dedicated EMI WMS service, but it is planned to write a probe using Condor glidein submission, as this is today the mainstream submission method in CMS. Table 3 lists all tests run by CMS.

CMS uses two profiles for the availability calculation. The first contains only the job submission and the SRM tests listed in table 3 and it is used by WLCG to calculate the official WLCG availability and reliability. The second corresponds to the availability used in CMS as one of the quality metrics to assess if the site was performing well in a given period of time; as such, it is an input to the CMS Site Readiness algorithm, combining different metrics into a single estimator [15].

The usage of two different profiles is motivated by the fact that WLCG should not consider any of the CMS-specific CE tests as critical, in order to factor out failures not caused by problems in the site infrastructure. However, the resulting availability is not sufficiently realistic as indicator of the site usability and a plan to separate tests of site functionality contained in CMS-specific tests is already foreseen to improve the significance of the WLCG calculation.

An interesting feature of the CMS tests is that some of them use a production VOMS role to ensure that they better simulate the site behaviour for production jobs; as an undesirable side

Table 3. SAM tests run by CMS.

Test name	Functionality	Standard
org.sam.(CREAM)CE-JobSubmit	Job submission to CE	yes
org.sam.glexec.(CREAM)CE-JobSubmit	Job submission to CE for gLExec test	yes
org.cms.WN-basic	CMS local site configuration	no
org.cms.WN-swinst	CMS local software installation	no
org.cms.WN-mc	Local file stageout	no
org.cms.WN-analysis	Reading local data	no
org.cms.WN-frontier	Reading calibration data from Frontier	no
org.cms.WN-squid	Local Squid server	no
org.cms.glexec.WN-gLExec	gLExec identity change	no
org.cms.SRM-GetPFNFromTFC	Convert logical to physical file name	no
org.cms.SRM-VOPut	SRM copy from client to SE	yes
org.cms.SRM-VOGet	SRM copy from SE to client	yes
org.cms.SRM-VODel	SRM deletion from SE	yes
org.cms.SRM-VOGetTURLs	SRM getTURL method	yes

effect, though, this sometimes causes the test results to expire after the 24 hours time period imposed by ACE, when the site is very busy with production jobs, causing it to be considered unavailable until the tests are run again. These occurrences need to be treated on a case-by-case basis with an *a posteriori* recalculation of the CMS site readiness estimator. An improvement with respect to the default Nagios configuration was obtained by increasing the timeout on job submission tests from 11.5 hours to 23.5 hours.

3.4. LHCb

LHCb makes extensive use of SAM. The test results are used as additional information for the shift crew with respect to the existing monitoring in the LHCb Dirac framework [16]. LHCb Dirac is the framework used by LHCb for interacting with the distributed computing infrastructure for workload management and data management. The tests run by LHCb are described in table 4.

LHCb uses SAM/Nagios to test computing elements, storage elements and LFC instances. The Nagios testing is driven by the so-called *topology.xml*, a file describing the infrastructure services used by the VO. LHCb is producing this file from the LHCb Dirac Configuration Service automatically every hour. The Configuration Service is updated by LHCb Dirac Agents that use the WLCG Information System for service discovery and included new service instances as soon as they appear.

LHCb would be interested in feeding information to the SAM/Nagios framework from LHCb Dirac, which is also monitoring the Grid infrastructure. This could provide a more complete and detailed level of monitoring information than what is currently been done via Nagios and would better reflect the usage of the Grid by the VO.

4. Visualisation

The Experiment Dashboard Framework [17] was used to develop the *Site Usability Monitor (SUM)* to visualize test results, availabilities and reliabilities of sites and site services. The SUM UI is used by all the LHC experiments.

Table 4. SAM tests run by LHCb.

Test name	Functionality	Standard
org.sam.CREAMCE-DirectJobSubmit	Direct job submission to CE	yes
org.sam.(CREAM)CE-JobSubmit	Indirect job submission to CE	yes
org.sam.glexec.(CREAM)CE-JobSubmit	Job submission to CE for gLExec test	yes
org.lhcb.WN-sft-brokerinfo	Get CE name from gLite BrokerInfo	no
org.lhcb.WN-sft-csh	C-shell installation	no
org.lhcb.WN-sft-lcg-rm-gfal	Setting of LCG_GFAL_INFOSYS	no
org.lhcb.WN-sft-vo-swdir	LHCb local software installation	no
org.lhcb.WN-sft-voms	VOMS client	no
org.lhcb.SRM-GetLHCInfo	Get information from site storage	no
org.lhcb.SRM-VOPut	SRM copy from client to SE	yes
org.lhcb.SRM-VOGet	SRM copy from SE to client	yes
org.lhcb.SRM-VODel	SRM deletion from SE	yes
org.lhcb.SRM-VOLs	SRM ls method on file	yes
org.lhcb.SRM-VOLsDir	SRM ls method on directory	yes
org.lhcb.LFC-Ping	Check that LFC is alive	no
org.lhcb.LFC-Read	Read file in LFC	no
org.lhcb.LFC-Readdir	Read directory in LFC	no
org.lhcb.LFC-Replicate	Replicate/read file in master/slave LFC	no

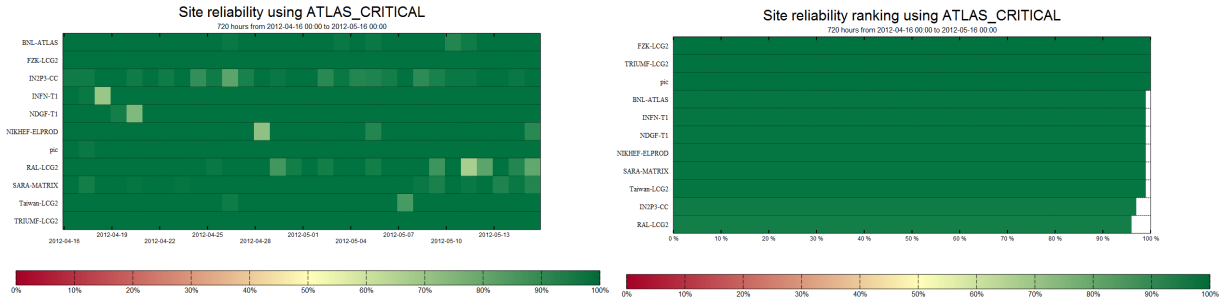

Figure 2. (left) A site reliability historical plot; (right) a site reliability ranking plot.

The SUM web interface provides two main views: one showing the *latest results*, that is the current status of all the tested services and endpoints, and the *historical view*, showing the historical trends for the availability and reliability of sites and services and for the test results for individual service instances. In figure 2 examples of a quality plot and a ranking plot of monthly reliability are shown. The history views are linked together in a chain with different granularities of the status information. In this way, users can start from the site availability and drill down through site services and test results all the way to the detailed log files per test result.

The visualisation of status quantities follows a simple colour scheme which enables the majority of users to immediately spot ongoing issues with services, and to determine their duration retrospectively. The target group of users contains the experiment computing operations experts, shifters, and site administrators, who can easily identify and address or

escalate issues with services at the sites.

Another visualisation interface, MyWLCG, is developed and maintained by the SAM team and will be evaluated by the LHC experiments, as a consolidation of the development effort is clearly desirable [18].

5. Future work and conclusions

The migration from the old to the new SAM framework was a lengthy process, executed in various steps: the migration from the SAM client to Nagios for the test submission, the migration from the old SAM database to the new infrastructure based on ACE, ATP and MRS and the migration to the SUM visualisation. The migration turned out to be fairly smooth and the new SAM now is completely integrated in the experiment monitoring systems. However, several improvements are desirable in order to achieve a better match with the experiment needs:

- a better separation between site-specific functionality and experiment-specific functionality; this would allow achieving a much clearer understanding of which problems should be solved by the site and which problems by the experiment;
- the creation of job submission probes using the same job submission mechanisms as used in the experiment; this is particularly important for those experiments that are not using, or plan to stop using, the EMI WMS service;
- the addition of tests for new functionality (the gLExec identity change mechanism is a recent example);
- an even greater flexibility in SAM, to allow "custom" services to be tested (either abstract, if they correspond to a set of functionalities that cannot be attached to real services, or real, if they are services not registered in OIM or GOCDB).

On a much shorter time scale, the imminent introduction of POEM for profile creation and editing will make the SAM configuration much easier and allow separating identical tests run with different VOMS proxy attributes.

The SAM framework is confirmed to be an essential tool for reliable computing operations in WLCG, not only for the infrastructure itself, but also for the experiments. The transition to mainstream technologies like Nagios and ActiveMQ had a visible impact on stability and flexibility, from a user point of view. The main challenge will be to further expand its functionality in order to match the evolving requirements of WLCG and experiments in the years to come.

6. Acknowledgements

We would like to thank the SAM team, in particular Pedro Andrade, Marian Babik, David Collados, Paloma Fuente Fernández, Emir Imamagic, Wojciech Lapka, Konstantin Skaburskas and Jacobo Tarragón for the excellent support we received and our many fruitful interactions.

References

[1] Worldwide LHC Computing Grid `http://wlcg.web.cern.ch/`
[2] European Grid Infrastructure `http://www.egi.eu/`
[3] Open Science Grid `http://www.opensciencegrid.org/`
[4] NorduGrid `http://www.nordugrid.org/`
[5] Collados D, Shade J, Traylen S and Imamagic E 2010 *J. Phys.: Conf. Ser.* **219** 062008
[6] Mathieu G *et al* 2010 *J. Phys.: Conf. Ser.* **219** 062021
[7] OSG Information Management System `http://oim.grid.iu.edu/oim/home`
[8] Van Der Ster D C *et al* 2012 Experience in Grid site testing for ATLAS, CMS and LHCb with HammerCloud *J. Phys.: Conf. Ser.* (not yet published)
[9] Tuckett D *et al* 2012 Collaborative development. Case study of the development of flexible monitoring applications *J. Phys.: Conf. Ser.* (not yet published)

[10] Cons L and Paladin M 2012 The WLCG Messaging Service and its Future *J. Phys.: Conf. Ser.* (not yet published)
[11] Legrand I *et al* 2009 *Comp. Phys. Comm.* **180** 12 2401-98
[12] Saiz P *et al* 2012 AliEn: ALICE Environment on the GRID *J. Phys.: Conf. Ser.* (not yet published)
[13] Elmsheuser J *et al* 2010 *J. Phys.: Conf. Ser.* **219** 072002
[14] Anisenkov A, Di Girolamo A, Klimentov A and Senchenko A 2012 AGIS: The ATLAS Grid Information System *J. Phys.: Conf. Ser.* (not yet published)
[15] Flix J, Hernández J and Sciabà A 2011 Monitoring the Readiness and Utilization of the Distributed CMS Computing Facilities *J. Phys.: Conf. Ser.* **331** 072020
[16] Tsaregorodtsev A 2010 *J. Phys.: Conf. Ser.* **219** 062029
[17] Saiz P *et al* 2012 Experiment Dashboard - a generic, scalable solution for monitoring of the LHC computing activities, distributed sites and services *J. Phys.: Conf. Ser.* (not yet published)
[18] Lapka W *et al* 2012 Distributed monitoring infrastructure for Worldwide LHC Computing Grid *J. Phys.: Conf. Ser.* (not yet published)