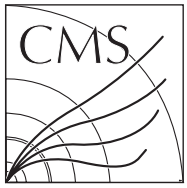
Available on CMS information server

CMS CR -2009/089



The Compact Muon Solenoid Experiment

Conference Report

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

13 May 2009

The commissioning of CMS sites: improving the site reliability

S. Belforte, I. Fisk, J. Flix, J. M. Hernández, J. Klem, J. Letts, N. Magini, P. Saiz, A. Sciabà

Abstract

The computing system of the CMS experiment works using distributed resources from more than 60 computing centres worldwide. These centres, located in Europe, America and Asia are interconnected by the Worldwide LHC Computing Grid. The operation of the system requires a stable and reliable behaviour of the underlying infrastructure. CMS has established a procedure to extensively test all relevant aspects of a Grid site, such as the ability to efficiently use their network to transfer data, the functionality of all the site services relevant for CMS and the capability to sustain the various CMS computing workflows at the required scale. This contribution describes in detail the procedure to rate CMS sites depending on their performance, including the complete automation of the program, the description of monitoring tools, and its impact in improving the overall reliability of the Grid from the point of view of the CMS computing system.

Presented at *CHEP09: International Conference On Computing In High Energy Physics And Nuclear Physics,21-27 Mar 2009,Prague,Czech Republic,15/05/2009*

The commissioning of CMS sites: improving the site reliability

S Belforte[1], I Fisk[2], J Flix[3,4], J M Hernández[4], J Klem[5], J Letts[6], N Magini[7,8], P Saiz[7], A Sciabà[7]

1. INFN, Sezione di Trieste, Trieste, Italy - 2. Fermi National Accelerator Laboratory, Batavia, USA - 3. Port d'Informació Científica, PIC (CIEMAT - IFAE - UAB), Bellaterra, Spain - 4. Centro de Investigaciones Energeticas Medioambientales y Tecnologicas, Madrid, Spain - 5. Helsinki Institute of Physics, Helsinki, Finland - 6. University of California San Diego, La Jolla, CA, USA - 7. CERN, Geneva, Switzerland - 8. INFN-CNAF, Italy

E-mail: jflix@pic.es, Andrea.Sciaba@cern.ch

Abstract. The computing system of the CMS experiment works using distributed resources from more than 60 computing centres worldwide. These centres, located in Europe, America and Asia are interconnected by the Worldwide LHC Computing Grid. The operation of the system requires a stable and reliable behaviour of the underlying infrastructure. CMS has established a procedure to extensively test all relevant aspects of a Grid site, such as the ability to efficiently use their network to transfer data, the functionality of all the site services relevant for CMS and the capability to sustain the various CMS computing workflows at the required scale. This contribution describes in detail the procedure to rate CMS sites depending on their performance, including the complete automation of the program, the description of monitoring tools, and its impact in improving the overall reliability of the Grid from the point of view of the CMS computing system.

1. Introduction

The Large Hadron Collider (LHC), located at CERN, will be operational in Autumn 2009 and will produce *p*-*p* collisions at centre-of-mass energy of 14 TeV, with a luminosity eventually two orders of magnitude larger than current hadron colliders. The Compact Muon Solenoid (CMS) [1] is one of the four detectors that will observe the collisions and it is foreseen to collect about 5 PB of data/year. The data processing requires exploiting computing and storage resources from several centres outside CERN. The resources are in fact coordinated by the Worldwide LHC Computing Grid (WLCG) [2], which at most sites exploits the computing infrastructure provided by other Grid projects, like EGEE, Open Science Grid and NorduGrid.

In the case of the CMS collaboration, around 60 sites from about 20 countries are involved. They are organized with a tiered structure, where different tier levels correspond to different functions. The Tier-0 site is CERN, and takes care of the prompt event reconstruction and detector calibration, the distribution of raw and processed data to external sites and the backup storage of the raw data. Seven Tier-1 sites run the subsequent reprocessing, including data skimming, keep an active copy of the raw data and store the Monte Carlo (MC) generated at

Tier-2 sites. Finally, around 50 Tier-2 sites get samples of the skimmed data for analysis and are used to run the MC simulation. A complete description of the CMS computing model and its services can be found elsewhere [3].

Given the complexity of the infrastructure, it is important to measure its performance in a continuous way, in order to inform the sites of any problem CMS is encountering, or will encounter, when having activities at that site. The Grid projects operating the infrastructure have their own procedures to identify and correct problems, but these do not necessarily reflect the usage CMS does of the resources. For this reason, CMS has established a set of techniques and tools intended to provide a better picture of the site performance and reliability.

The following sections describe how this is performed: the procedure to test sites in an automatic and continuous way, the results obtained so far, and the implications for site usage.

2. Site evaluation techniques

CMS undertakes periodic computing challenges of increasing scale and complexity to test its computing model and the Grid infrastructure. Performance values are measured, problems are identified and feedback into the design, integration and operation is provided. As an example, during the 2008 Common Computing Readiness Challenge (CCRC08) [4] the CMS Tier-2 sites were tested using realistic analysis workflows, more importantly, having real users submitting a large number of analysis tasks. This allowed the collaboration to quantify the readiness of the Tier-2 sites and provided invaluable input to the operations of the computing centres.

However, the operation of the complete CMS computing system requires stable and reliable behavior of the underlying heterogeneous (in computing resources, regions and support) infrastructure at all times. The *Site Commissioning* [5] [6] activity is part of the CMS computing integration program and its mandate is to evaluate the readiness of every CMS site to execute the computing tasks assigned to it. This information is used by the sites to become aware of problems, and by CMS to plan accordingly the distribution of the workload such that temporarily unreliable sites are not used.

In order to accomplish that, custom tests are regularly run at each site, and they are conceived to check every possible functionality exploited by CMS, aiming at having the highest possible correlation between failures of these tests and of real CMS jobs. Sites must satisfy certain lower limits on the success rate of these tests to be considered reliable.

The following information is used to evaluate the readiness of a CMS site:

- The results of the CMS SAM tests, running on Grid site resources to check their functionality and the local CMS software and configuration;
- the success rate of the Job Robot, a load generator simulating user data analysis;
- the number and the quality of the data transfer links used in production;
- the downtimes scheduled by the site.

2.1. SAM site availability

In the WLCG, all Grid services are periodically tested using a framework called SAM (Service Availability Monitor) [7], which executes periodic tests on all the Grid services within the infrastructure. SAM provides one of the main sources of information for the Grid operations and is used to measure the availability of Grid services.

CMS has adopted SAM to run custom tests on the Computing Elements (CE) and Storage Resource Manager (SRM) instances at the sites. These tests allow to determine, among other things, that:

- It is possible to send and run jobs;
- the CMS software is correctly installed and configured;
- it is possible to access in a job local CMS data;
- it is possible to copy data in and out of the local storage.

These tests are listed in Table 1 and are run once per hour on all CE and SRM instances accessible by CMS. The CE tests are run on a worker node of the site batch system by submitting Grid jobs via the gLite WMS [8] from a central node; SRM tests are directly run from that node.

Test name	Test definition
Computing Element	
js/jsprod	Checks the job submission via Grid
basic	Checks the local CMS configuration
swinst	Checks the locally installed CMS software
mc	Checks the file transfer mechanism to the local SE
frontier	Checks access to the calibration data via the local cache
squid	Queries the local calibration server
analysis	Checks the accessibility of a local dataset
Storage Resource Manager	
get-pfn-from-tfc	Performs logical-to-physical file name translation
lcp-cp	Transfers files to and from the local SRM

Table 1. CMS critical tests for the CE and the SRM services.

A failure of any of these tests determines the "unavailability" of the service instance where the test ran. If all instances of a given service type in a site are unavailable, the service itself is considered unavailable. Finally, if either the CE or the SRM service is unavailable, the site itself is considered unavailable.

The site availability over a time interval is the fraction of that time interval when the site was available. An example of the daily evolution of this metric and ranking for all CMS sites tested, for a month period, is shown in Figure 1.

2.2. Job Robot

Another complementary testing method consists of regularly submitting jobs similar to real analysis jobs. The difference with respect to the SAM tests is the fact that the statistics are much higher, ∼600 jobs/(site×day), the fact that the accessed data can be spread on several disks, as it sizes 0.5 TB, and a higher load on the site storage system. A tool called Job Robot was developed to implement such automatic job submission system using CRAB [9], the CMS analysis job submission tool.

At regular time intervals, a new analysis task is created for each site, to be run on a specific dataset. The task is then split into several jobs, which are submitted as a collection to the gLite WMS. Each job performs a trivial data analysis on a fraction of the dataset. All submitted jobs are classified as successful, as failed at the application level or as aborted at the Grid level.

The Job Robot daily statistics are used to measure the success rate for each site. The usual usage of the Job Robot is in monitoring mode, i.e. with a load that continuosly ocuppy some slots at the sites. However, the Job Robot can be tuned to saturate all available CMS slots at the sites, and then compare the results to the resource pledges to CMS; this stress mode is useful also to uncover possible bottlenecks or scaling problems in the site services, although this has not been tested yet.

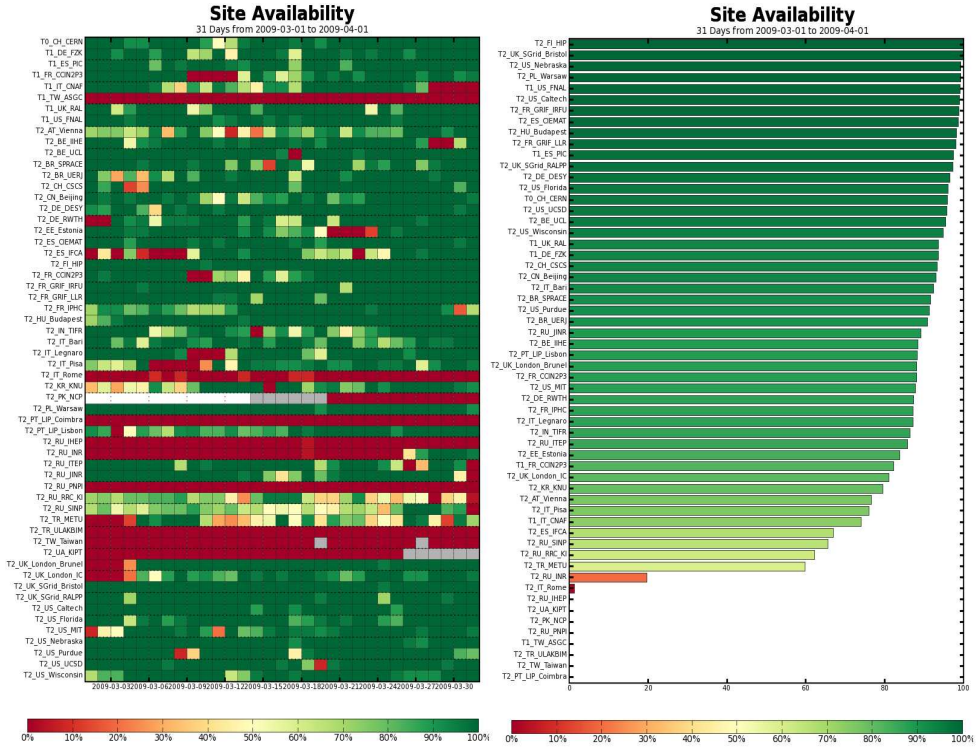


Figure 1. Both history (left) and ranking (right) SAM site availability results of March 2009 for all CMS sites.

Figure 2 shows the Job Robot success rate for all jobs submitted in monitoring mode on a month, both history and ranking views.

2.3. Data transfer links

A site needs to have sufficient data transfer connections to other sites in order to perform CMS workflows. In the CMS computing model the reconstructed data is distributed between Tier-1 sites; for analysis based on reconstructed data, transfers from all Tier-1 to all Tier-2 sites are required. Transfers between Tier-1 sites at very high rate will be needed every time there is a global replication of reduced data, result of different reprocessing passes ocurring at the Tier-1 sites. For uploading MC data produced at Tier-2 sites, typically the regional Tier-1 is used but transfers to other Tier-1 sites are also required. Hence, data is massively moved in the organized system for further processing and the data placement needs to work efficiently.

In 2007, a Debugging Data Transfers (DDT) task force was created to design and enforce a procedure to debug problematic links [10][11]. A clear certification procedure was set, using a traffic generator to test the quality of a link and considering a link to be commissioned when it demonstrates:

- for links with source at Tier-0 or Tier-1 sites, 20 MB/s averaged over 24 hours;
- for links with source at Tier-2 sites, 5 MB/s averaged over 24 hours.

Only links which are commissioned are used to move data for production usage. The DDT activities have been extremely useful and the data transfers have increased in number and improved in quality since the DDT program started.

Figure 2. Both history (left) and ranking (right) Job Robot success rates of March 2009 for all CMS sites.

At end of March 2009, a total of 523 links had achieved commissioning:

- 56/56 Tier-(0,1)↔Tier-1 links (100%);
- 300/352 Tier-1→Tier-2 links (85%);
- 140/352 Tier-2→Tier-1 links (40%);
- 27 Tier-2→Tier-2 links (not reflected in computing model).

It is worth to grossly mention that missing commissioned links are typically the ones not used for production (for example, some non-regional Tier-2→Tier-1 links). They are not failing the commissioning process; they are simply not being used at all.

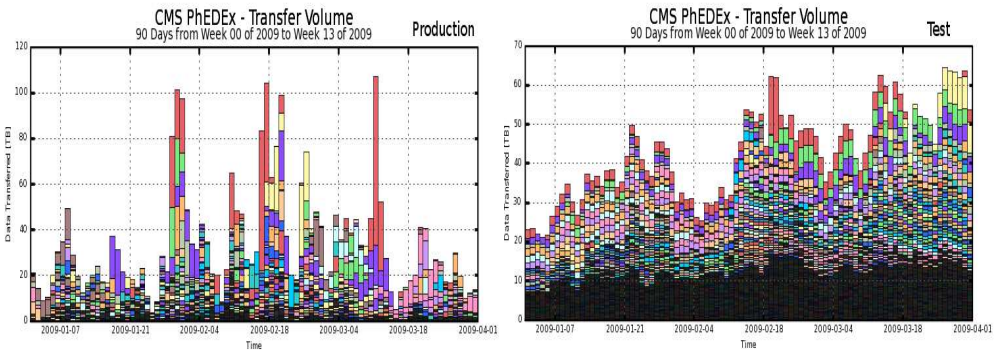


Figure 3. Production (left) and test (right) daily data transfer volumes, for the first 2009 quarter.

Additionally to DDT, a procedure has been set up to continuously monitor the data transfer quality on all active links by using both production data transfers, when available, and a constant low rate data transfer running at 0.5 MB/s on each link. Figure 3 shows the daily amount of transferred data on all active links for both production and test transfers, for the first 2009 quarter. This measurement allows to detect systematic problems, not only at the network level, but also in the data transfer services and the storage infrastructure. This information may also be used to decommission data transfer links which have a very poor quality since a very long time. Figure 4 shows the transfer quality for all transfers from Tier-1 to Tier-2 sites ocurring in the first 2009 quarter. Ocassionally, some periods in which Tier-1 sites had export problems to the Tier-2 sites are clearly seen.

Figure 4. Daily average quality for Tier-1→Tier-2 transfers, for the first 2009 quarter.

2.4. Scheduled downtimes

When evaluating the reliability of a site, one needs to properly trace the scheduled downtimes of Grid services used by CMS. These downtimes are published in the SAM database and can concern individual service instances as well as the whole site. In a given day, a site is in a scheduled downtime from the point of view of CMS if either:

- The whole site had a scheduled downtime during the day;
- the CE, or all the SRM instances used by CMS at the site, had a scheduled downtime during the day.

2.5. The Site Status Board

The *Site Status Board* (SSB) is a synoptic view of the status of all CMS computing sites [12]. It is designed to allow users to correlate the output of their workflows with known problems at sites, and to provide experts with a single entry point to the full suite of CMS monitoring tools. The provided information is often changed as the understanding of what is most relevant for making a good diagnosis of problems improves.

The SSB is a flexible presentation layer above a dynamic framework where information is stored in the columns of a database table, having the site name as key. These columns are filled by processes collecting data from the internal CMS dashboard database [13], the WLCG information system and ASCII files on the web. New columns can be freely added and defined

via a web interface and grouped in "views", or collections of columns. The time history of any column can be graphically displayed or retrieved in XML format.

Site Readiness plots and results are kept and accessed from the SSB.

3. Site Readiness evaluation

3.1. Site Readiness criteria

The quantities defined above must satisfy some constraints to consider the site as ready. Ideally, these constraints should be defined in such a way as to *a)* allow temporary glitches, *b)* enforce a reasonable level of reliability over a period of time and *c)* allow sites to quickly recover their ready status when problems are solved. In addition to that, downtimes due to scheduled maintenance and failures during weekends (for Tier-2 sites) should not be negatively considered in the site evaluation.

Currently, each day a site is evaluated as good (tagged as 'O', Ok) if the conditions in Table 2 are satisfied, and as bad if at least one metric is not satisfied (tagged as 'E', Error). Additionally, the scheduled downtimes are as well accounted (tagged as 'SD'). In order to take into account the stability of a site, a readiness daily status of a site is evaluated using the history of the last 7 days overall daily metrics (ignoring downtimes) and it is expressed by a flag with four possible values: *Ready (R)*, *Warning (W)*, *Not-Ready (W)* and *Scheduled-Downtime (SD)*, which means respectively that the site is fully usable, that it is usable but suffering from temporary problems, that the site is unusable and that the site is under a maintenance period.

The transition rules between these states are shown in Table 3. Weekend daily failures do not negatively count for Tier-2 sites in the evaluation of the Site Readiness daily status. If the site is in maintenance, its daily readiness status is 'SD', regarless the last 7 days history of daily states. Note that the intermediate warning state gives sites reasonable time to recover.

Tier-1 sites	**Tier-2 sites**
daily SAM availability $\geq 90\%$	daily SAM availability $\geq 80\%$
daily Job Robot efficiency $\geq 90\%$	daily Job Robot efficiency $\geq 80\%$
commissioned link from Tier-0	commissioned links to Tier-1s$\geq$2
commissioned links to Tier-2s$\geq$20	commissioned links from Tier-1s$\geq$4
commissioned links from/to other Tier-1s$\geq$4	

Table 2. Site Commissioning daily metrics required for Tier-1 and Tier-2 CMS sites.

from/to	**NR**	**R**	**W**
NR	-	O for last 2 days	-
R	E for >2 days over last 7	-	E but not E for >2 days over last 7
W	E for >2 days over last 7	O and not E for >2 days over last 7	-

Table 3. Site Readiness daily status: transition rules.

3.2. Site Readiness results

The Site Commissioning program is active since October 2008. Apart from consulting the results and status in the SSB, each site is provided with an overall picture of its last 15 days Site Readiness status and daily metric status, as well as the independent daily measurements. Figure 5 shows an example of such view for a Tier-2 at the end of March 2009.

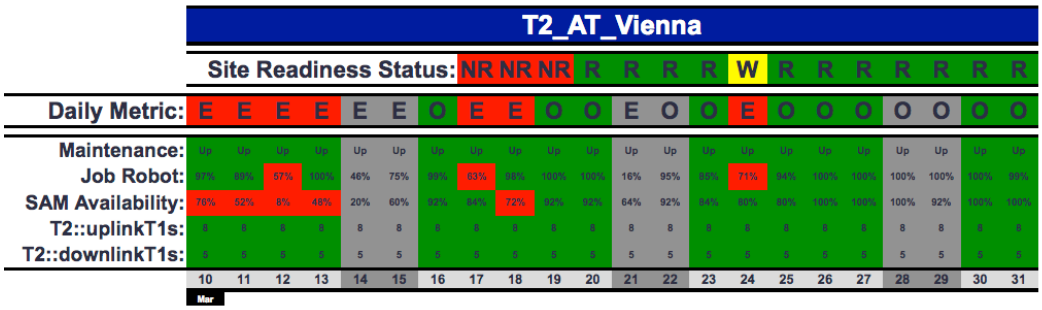


Figure 5. Overall picture on Site Readiness results for a CMS Tier-2 site, at the end of March 2009.

Figure 6 shows the monthly averaged SAM availability/reliability and Job Robot results for CMS Tier-1 and Tier-2 sites, from November 2008 to March 2009. Ignoring the days in which a site has declared a maintenance, the CMS SAM reliability (as similar as evaluated by WLCG SAM tests) is obtained. Since January 2009, the maintenances per service are better traced (from SAM DB). Being generic, WLCG SAM tests show slightly better results (∼95% values for Tier-1s) than CMS SAM tests, which probe a bunch of VO-speficic functionalities, and the difference is expected. Note the improvement of CMS SAM availibilities/reliabilities: in average, Tier-2 sites are above the Site Commissioning metric (80%), while Tier-1 sites do (90%) from March 2009. Job Robot success rate results shows that sites are slightly improving as well, with both Tier-1 and Tier-2 sites above the metrics (90% and 80%, respectively), in average.

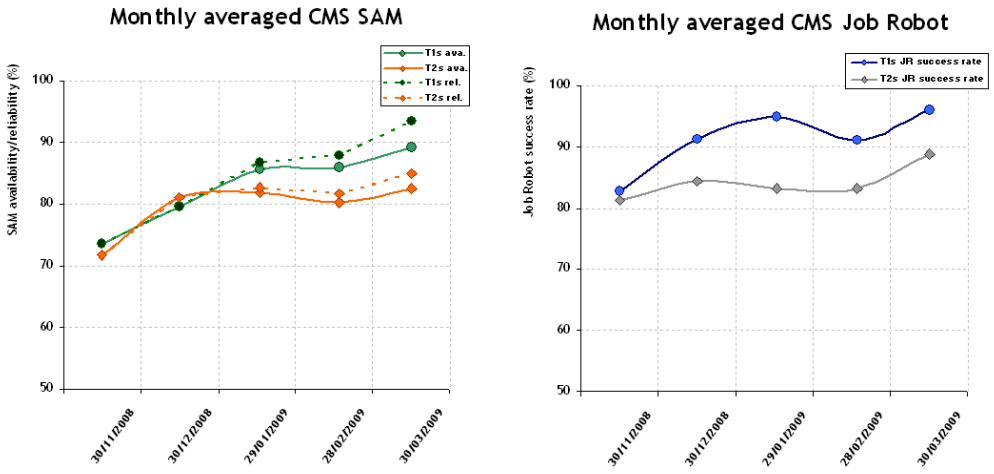


Figure 6. Monthly averaged SAM and Job Robot results for CMS Tier-1 and Tier-2 sites.

Figure 7 shows an historical view of the number of Tier-2 sites in each readiness status, since the Site Commissioning activities were set. A trend towards increasing numbers of good sites is evident, from 15 sites at the beginning of October 2008 to an average of 35 end March 2009. In 24th February 2009, a more restrictive metric was included in the commissioned links, and some Tier-2 sites were negatively affected. They rapidly commissioned the missing data transfer links and became 'ready' again. Though the new policy is more restrictive, only 30 commissioned data transfer links, affecting only to Tier-2 sites, are missing the certification to achieve the Site Readiness goal on commissioned transfer links for all the current CMS sites.

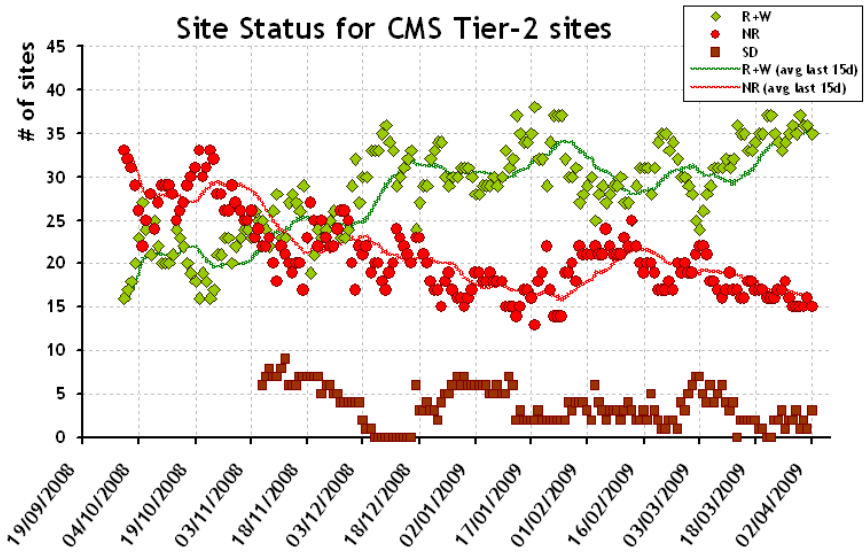


Figure 7. Evolution of Site Readiness status for CMS Tier-2 sites, since Site Commissioning started.

By means of this readiness evaluation, sites have an easy way to know if CMS is seeing problems at them. The program provides all kind of monitoring plots, XML feeds, a Nagios-plugin [14], and automatic alerts to stable sites prior being *Not-Ready* via Savannah ticketing system. Site performances according to the Site Readiness criteria are periodically reviewed in the CMS collaboration on a weekly basis.

3.3. Site Readiness usage

CMS plans to use the measured Site Readiness as a guide to mark sites as good or bad for running computing activities, like data reprocessing, MC simulation and analysis. For example, at Tier-1s, trends will be used to determine data placement and long term job routing, as the typical lenght of a workflow is of the order of days and the readiness of the site is a useful indicator to schedule the activities.

To help the guide, the fraction of time a site has been stable and reliable is daily estimated based on the last 15-days history of Site Readiness daily status. Figure 8 shows the ranking plots based on this *Site Readiness* for Tier-1 and Tier-2 sites, on 31st March 2009. The goal is to chose those sites for operations above a certain threshold on *Site Readiness* (placed at 90% for Tier-1 and 80% for Tier-2 sites), as well as the scheduled downtime information at any time.

4. Current activities and deployments

Data transfer quality is planned for inclusion in the Site Readiness criteria. It can also be used to detect links to be removed from production when they suffer from a persistent low transfer quality in during more than one month (they are approximately 1% of the total). Link removals from production will happen monthly, and sites will need to recommission them according to the usual DDT criteria.

A task force team has been created to determine usual failures, help sites to improve, give feedback for robustness of CMS tools/services, and to pursue the increase of reliability of CMS sites. Moreover, it is expected the evaluation of the intrinsic levels of inefficiencies per service.

Other activities include running the Job Robot in stress mode, interfacing the Site Readiness

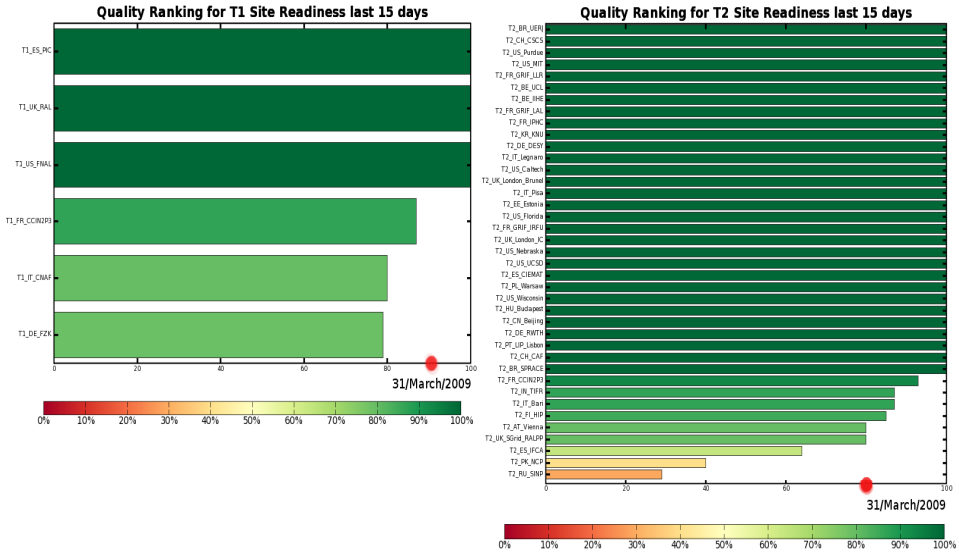


Figure 8. Site Readiness ranking for Tier-1 (left) and Tier-2 (right) CMS sites on 31st March 2009.

results to production and analysis tools, or adapting the tools for computing shifts to facilitate the identication of unreliable sites during CMS operations on shifts.

5. Conclusions

Site Commissioning activities are crucial for bringing the CMS distributed computing system into stable and reliable operations. We have continuously monitored Grid and CMS services at sites for about 6 months. All the available information is now condensed in a single estimator, whose value takes in account also the stability of the site. This helps production and users to select reliable CMS sites. A positive trend in reliability for Tier-1 and Tier-2 sites has been already observed. More metrics will be included, hoping to provide the CMS sites with sufficient feedback and monitoring results so they can make their sites more reliable to CMS, and for CMS to use them in the most efficient way.

References

[1] CMS Collaboration, *The Compact Muon Solenoid Technical Proposal*. CERN/LHCC 94-38, 1994.
[2] Worldwide LHC Computing Grid. *Technical Design Report*. E-ref: http://lcg.web.cern.ch/LCG/TDR/LCG_TDR_v1_04.pdf
[3] CMS Collaboration, "CMS Computing Project: Technical design report". *CERN-LHCC 2005-023*, 2005.
[4] J.D. Shiers et al., "The (WLCG) Common Computing Readiness Challenge(s)", *CCRC08, contribution N29-2, session Grid Computing*. Nuclear Science Symposium, IEEE (Dresden), October 2008.
[5] J. Flix et al., "The commissioning of CMS computing centres in the worldwide LHC computing Grid", *contribution N29-5, session Grid Computing*. Nuclear Science Symposium, IEEE (Dresden), October 2008.
[6] A complete description of the program in https://twiki.cern.ch/twiki/bin/view/CMS/PADASiteCommissioning
[7] A. Duarte, P. Nyczyk, A. Retico, D. Vicinanza, "Monitoring the EGEE/WLCG Grid Services", *J. Phys.: Conf. Ser. 119 052014*, 2008.
[8] P. Andreetto et al., "The gLite workload management system", *J. Phys.: Conf. Ser. 119 062007*, 2008.
[9] D. Spiga et al., "The CMS Remote Analysis Builder (CRAB)", *LNCS vol. 4873*, pp. 580-586, 2007.
[10] G. Bagliesi et al., "The CMS Data Transfer Test Environment in Preparation for LHC Data Taking", *contribution N67-2, session Applied Computing Techniques*. Nuclear Science Symposium, IEEE (Dresden), October 2008.
[11] J. Letts et al., "Debugging Data Transfers in CMS", *CHEP09*, Prague, Czech Republic, March 2009.
[12] R. Rocha et al., "Experiment Dashboard for Monitoring of the Computing Activities of the LHC Experiments on the Grid", *contribution N29-4, session Grid Computing*. Nuclear Science Symposium, IEEE (Dresden), October 2008.
[13] J. Andreeva et al., "Dashoard for the LHC Experiments", Proceedings of International Conference on Computing in High Energy and Nuclear Physics (CHEP 07), *J.Phys.Conf.Ser.119:062008*, 2008.
[14] Nagios: *Industry Standard in Enterprise System, Network, and Application Monitoring*. (www.nagios.org).