Available on CMS information server

CMS CR -2009/087

The Compact Muon Solenoid Experiment

Conference Report

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

13 May 2009

Bringing the CMS distributed computing system into scalable operations

S Belforte, A Fanfani, I Fisk, J Flix, J M Hernández, T Kress, J Letts, N Magini, V Miccio and A Sciabà

Abstract

Establishing efficient and scalable operations of the CMS distributed computing system critically relies on the proper integration, commissioning and scale testing of the data and workload management tools, the various computing workflows and the underlying computing infrastructure, located at more than 50 computing centres worldwide and interconnected by the Worldwide LHC Computing Grid. Computing challenges periodically undertaken by CMS in the past years with increasing scale and complexity have revealed the need for a sustained effort on computing integration and commissioning activities. The Processing and Data Access (PADA) Task Force was established at the beginning of 2008 within the CMS Computing Program with the mandate of validating the infrastructure for organized processing and user analysis including the sites and the workload and data management tools, validating the distributed production system by performing functionality, reliability and scale tests, helping sites to commission, configure and optimize the networking and storage through scale testing data transfers and data processing, and improving the efficiency of accessing data across the CMS computing system from global transfers to local access. This contribution reports on the tools and procedures developed by CMS for computing commissioning and scale testing as well as the improvements accomplished towards efficient, reliable and scalable computing operations. The activities include the development and operation of load generators for job submission and data transfers with the aim of stressing the experiment and Grid data management and workload management systems, site commissioning procedures and tools to monitor and improve site availability and reliability, as well as activities targeted to the commissioning of the distributed production, user analysis and monitoring systems.

Presented at *Computing in High Energy and Nuclear Physics CHEP09,21 - 27 March 2009,Prague,Czech Republic,15/05/2009*

Bringing the CMS distributed computing system into scalable operations

S Belforte[1], A Fanfani[2], I Fisk[3], J Flix[4,5], J M Hernández[5], T Kress[6], J Letts[7], N Magini[8,9], V Miccio[10] and A Sciabà[8]

[1]INFN Trieste, [2]INFN Bologna, [3]FNAL, [4]PIC, [5]CIEMAT, [6]RWTH, [7]UCSD, [8]CERN, [9] INFN CNAF, [10]INFN Milano

E-mail: `jose.hernandez@ciemat.es`

Abstract. Establishing efficient and scalable operations of the CMS distributed computing system critically relies on the proper integration, commissioning and scale testing of the data and workload management tools, the various computing workflows and the underlying computing infrastructure, located at more than 50 computing centres worldwide and interconnected by the Worldwide LHC Computing Grid. Computing challenges periodically undertaken by CMS in the past years with increasing scale and complexity have revealed the need for a sustained effort on computing integration and commissioning activities. The Processing and Data Access (PADA) Task Force was established at the beginning of 2008 within the CMS Computing Program with the mandate of validating the infrastructure for organized processing and user analysis including the sites and the workload and data management tools, validating the distributed production system by performing functionality, reliability and scale tests, helping sites to commission, configure and optimize the networking and storage through scale testing data transfers and data processing, and improving the efficiency of accessing data across the CMS computing system from global transfers to local access. This contribution reports on the tools and procedures developed by CMS for computing commissioning and scale testing as well as the improvements accomplished towards efficient, reliable and scalable computing operations. The activities include the development and operation of load generators for job submission and data transfers with the aim of stressing the experiment and Grid data management and workload management systems, site commissioning procedures and tools to monitor and improve site availability and reliability, as well as activities targeted to the commissioning of the distributed production, user analysis and monitoring systems.

1. Introduction

Establishing efficient and scalable operations of the CMS distributed computing system requires proper integration, commissioning, scale testing and monitoring of the underlying computing infrastructure, the data and workload management tools and the computing workflows.

Computing challenges, undertaken periodically by CMS to test its computing model and the distributed computing system, have revealed the need for a sustained effort on computing integration and commissioning activities in CMS. In January 2008 CMS launched the Processing And Data Access task force (PADA) within the Computing Integration Program with the mandate of validating the computing infrastructure for data processing and analysis. This paper describes the activities, results and experience of the PADA task force.

Figure 1 depicts the CMS computing model [1]. The CMS computing resources are structured in a tiered architecture with specific functionality at different levels. The Tier-0 centre is located at CERN. It archives on tape and promptly reconstructs the raw data coming from the detector at a rate of about 300 MB/s. In addition CERN hosts the CMS Analysis Facility (CAF). The CAF has access to the full raw dataset and is focused on latency-critical detector, trigger and calibration activities. It also provides some CMS central services like the storage of conditions data and calibrations. Reconstructed data at the Tier-0 together with the corresponding raw data are distributed to the next level in the tiered structure, a small number (7) of Tier-1 centres where organized mass data processing is performed. That includes calibration, re-processing, data skimming and other organized intensive analysis tasks. About 50k jobs/day are expected to run at the Tier-1 sites. The data coming from the Tier-0 (50-250 MB/s depending on the Tier-1 centre) are archived on tape as well as the simulated data produced at the Tier-2 centres at a moderate rate of about 1 TB/day/Tier-2. At the Tier-2 centres, more than 50 sites around the world, in addition to the production of simulated data, user analysis of data imported from Tier-1 centres takes place. About 200k jobs/day are expected to run at the Tier-2 sites. Data for analysis can be downloaded from any of the Tier-1 centres. These transfers are bursty, ranging from 50 to 500 MB/s.

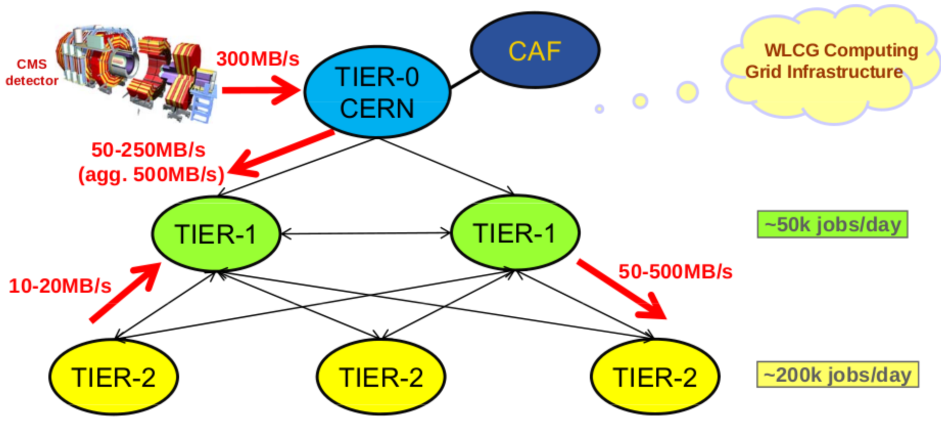

Figure 1. Schematic view of the CMS Computing Model.

2. The Processing And Data Access Task Force

CMS has undertaken periodic challenges of increasing scale and complexity to test its computing model and Grid systems. Performance values are measured, problems are identified and feedback into the design, integration and operation of the system is provided.

Lots of useful lessons have been learnt in the computing challenges. Sustainable operations are difficult to achieve. The computing system has to be robust enough to cope with the intrinsic unreliability of the distributed infrastructure. Continuous reliability and scaling tests running realistic workflows, at large scale on the computing resources using the production tools, are needed to ensure that the processing system scales adequately. Finding stable and efficient operating points takes quite some time and a well integrated monitoring system was considered essential. In general, the tests revealed the need to improve the reliability and performance of accessing and processing data.

The PADA processing task force was constituted with the following goals: validating the infrastructure for organized processing and user analysis including the sites and the workload and data management tools; validating the distributed production system by performing functionality, reliability and scale tests; helping sites to commission, configure and optimize the networking and storage through scale-testing data transfers and data processing; improving the efficiency of accessing data across the CMS computing system from global transfers to local access.

PADA was structured in four activities: the monitoring activity with the objective of evaluating, consolidating, integrating and extending the existing monitoring tools; the commissioning of the data transfers aiming at improving the quality and throughput of the transfers over the Wide Area Network (WAN); the site commissioning with the goal of improving the reliability and efficiency of the computing sites; and the commissioning of the distributed production and analysis systems to verify that data can be efficiently processed and analyzed over the Grid computing system. The results of the various PADA activities are reported in detail in these proceedings [2, 3, 4, 5, 6, 7, 8].

2.1. Monitoring integration

As the LHC approaches completion the need for clear and accurate monitoring tools for the CMS computing system becomes more pressing. It is important that different monitoring tools meet the requirements of all stakeholders and provide an integrated view of the computing operations. The PADA monitoring subgroup conducted a critical review of the CMS monitoring tools available to the various offline consumers (data and facilities operation, user support and detector commissioning). The requirements for the monitoring tools and missing functionality were identified. Recommendations on how to fill the gaps in the currently provided service were provided.

The PADA monitoring team led the adoption of the Site Status Board (SSB) [2] monitor by CMS. The SSB is a generic monitoring tool, developed by the ARDA Dashboard team, that displays the status and the history of sites according to some pre-defined metrics. The flexibility of the SSB makes it possible to use it for monitoring different activities according to specific metrics.

2.2. Commissioning data transfers

CMS needs to have working transfer links between the different tiers. Transfers from the Tier-0 to the Tier-1 centres are required to export the raw data collected from the detector and the products of the prompt reconstruction. After every reprocessing of the data, transfers between the Tier-1 sites take place to re-synchronize the Analysis Object Data (AOD) samples, the data format for high level analysis. The Tier-2 sites must be able to download selected data samples from any Tier-1 site and to upload the simulated data locally produced to the custodial Tier-1 centre.

In July 2007 the Debugging Data Transfers (DDT) task force [3] was formed with the mandate of defining appropriate metrics, providing a procedure and the tools to test transfer links and assist sites in solving data transfer problems. A data transfer load generator was implemented in the CMS PhEDEx [9] data transfer and placement system in order to provide sustained data transfers between any two sites at a given rate. Transfer links were certified if they were capable of sustaining for 24 hours a transfer rate of 20 MB/s for Tier-0 to Tier-1, Tier-1 to Tier-1 and Tier-1 to Tier-2 links, and 5 MB/s for Tier-2 to Tier-1 links. This activity was extremely useful to uncover and fix several problems at the various layers of the data transfer system (PhEDEx, the File Transfer Service -FTS-, the Storage Resource Manager -SRM- layer, gridFTP, network, Storage Elements, tcp settings, etc).

In figure 2 the significant increase in production data transfer quality resulting from the DDT activity can be appreciated.

CMS has transferred over the WAN about 60 PBytes of data in the last two years. Currently 50-100 TB of data are transferred daily CMS-wide (see figure 3). Given that links are not

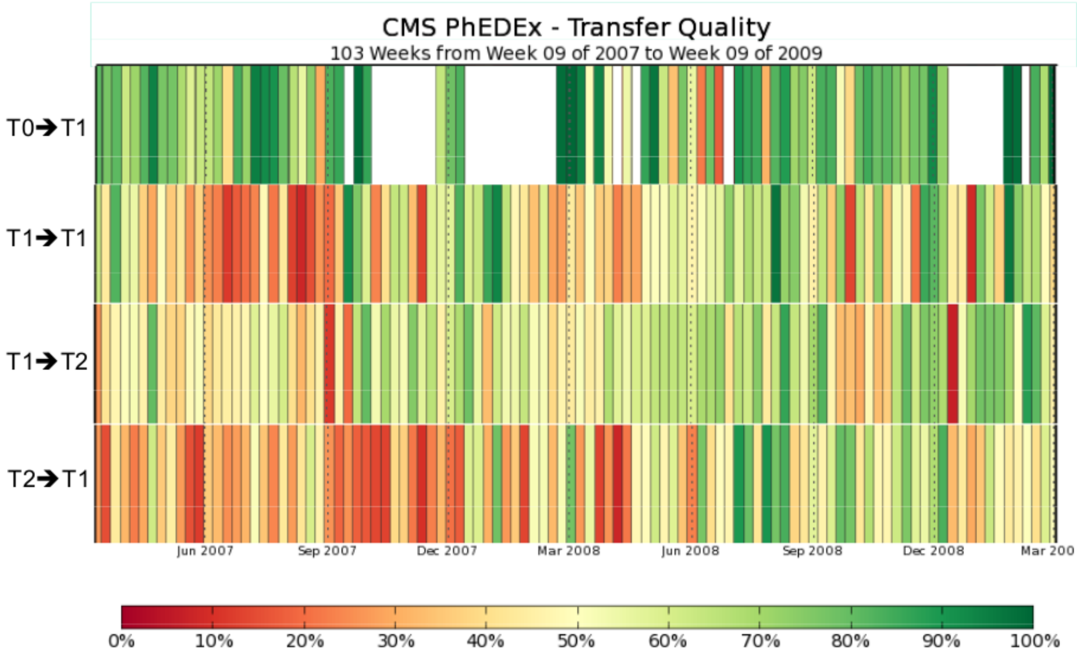

Figure 2. Production data transfer quality as a function of time for the last two years separated by transfers between different tier levels.

continuously utilized by production transfers, monitoring data transfers at low rate (0.5 MB/s) are conducted continuously in the certified links (about 500). This way transfer problems can be quickly identified before production transfers take place. This monitoring heartbeat results in a total traffic of about 30 TB/day. We are quite confident that the CMS data transfer system can cope with the needs of the experiment.

2.3. Site Commissioning

The CMS distributed computing requires a stable and reliable behavior of the underlying infrastructure at all times. This is difficult to achieve given the heterogeneity, the different amount of computing resources and support level at the sites. A framework to monitor the 'site readiness' for conducting the CMS workflows has been developed [4]. It monitors the reliability of the computing infrastructure and services at the sites by means of dedicated Site Availability Monitoring (SAM) tests that probe specific services, dedicated monitoring jobs that access data at the sites submitted periodically by a Job Robot, and by monitoring the quality of the data transfers in certified links.

SAM tests are high priority jobs, submitted every hour to the sites, that test the correct functioning of the Computing Elements, Storage Elements, the experiment software area, the experiment conditions cache, that data can be locally read/written from/to the local mass storage, etc. All those services are required to have a daily availability larger than 80%/90% for the Tier-2/Tier-1 sites respectively.

The Job Robot load generator is a tool for automatic job preparation, submission, collection and evaluation. It submits 600 jobs/day to each CMS site resulting in a total of 30k jobs/day. These jobs are simple but realistic jobs that access a large dataset locally stored at the sites. The JobRobot supports two modes of operation: the monitoring mode where a constant rate of jobs are submitted to sites for monitoring purposes, and the stress mode in which sites are

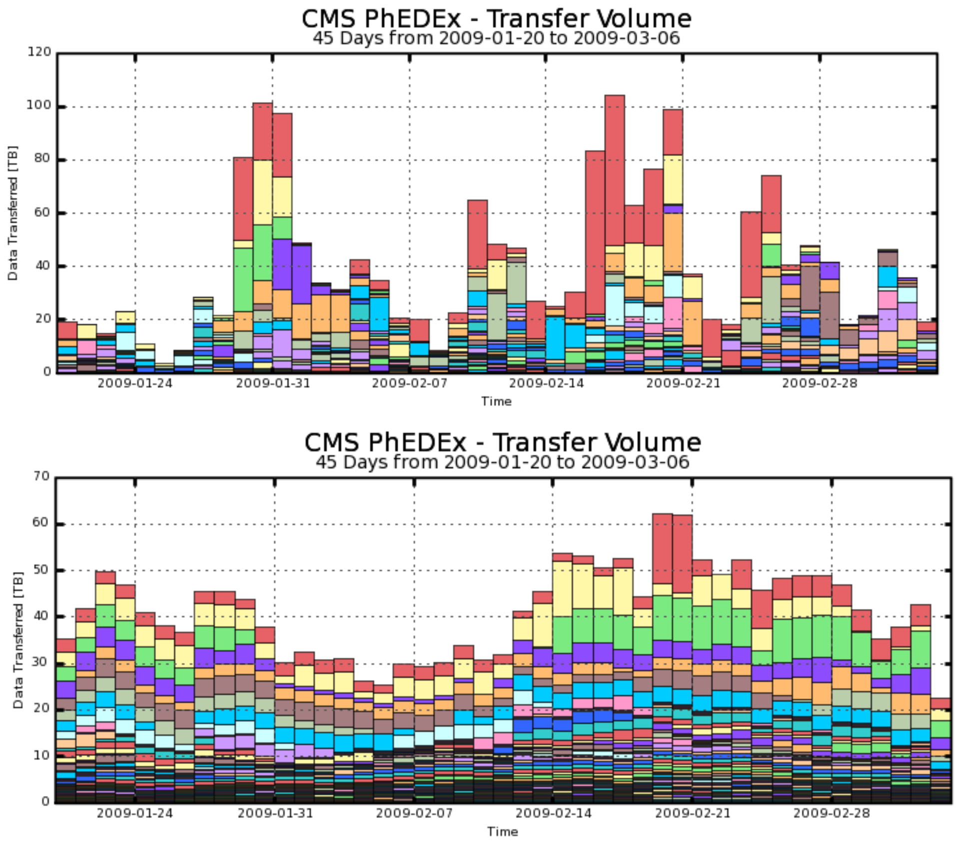

Figure 3. Daily volume of data transferred CMS-wide, during a period of 45 days, including production (top) and monitoring (bottom) transfers.

filled up with jobs to stress the data serving system. For the site readiness metrics, a Job Robot success rate higher than 80%/90% is required for the Tier-2/Tier-1 sites respectively.

Data transfers are also taken into account in the site readiness metrics. Sites are required to have a minimum number of DDT-certified links. For Tier-1 sites, a minimum of 20 download links to Tier-2s and at least 4 links from/to other Tier-1's are required. Tier-2 sites must have at least 4/2 certified links from/to Tier-1 sites. In addition, a minimum number of links with a transfer quality above certain threshold is required. Production and monitoring transfers are counted. A transfer quality above 50% for at least 50% of the certified links is required. These modest requirements are enough to detect systematic transfer problems at a site since typically when a site develops transfers problems they reflect in a poor transfer quality in most of the links.

The site readiness monitoring information is collected and displayed in the Site Status Board. In figure 4 a snapshot of the SSB for the site commissioning metrics is shown. All metrics are combined into a single daily 'site readiness status'. A site is *Ready* if the SAM availability and the JobRobot success rate are above the thresholds quoted above, and the data transfer metrics are satisfied. If any of the metrics is not fulfilled during a day, the site goes to the *Warning* status. The site can be in such status at most for two days in the last seven days. If the problem is not fixed, the site degrades to the *Not-Ready* status. It can quickly recover after two days of fulfilling the metrics. Scheduled downtimes (and week-ends for Tier-2 sites) are taken into account in the evaluation of the metrics. Figure 5 shows as an example for a site the values of the daily metrics as well as the site readiness status for a period of three weeks.

Site Name	SiteComm JR	Commissioned Links (expand this column)	Site availability	SiteReadiness Status	Maintenance (expand this column)	Good links
T0_CH_CERN	98%(699)	n/a	100%	n/a	GOCDB	n/a
T1_CH_CERN	n/a	combined	100%	n/a	n/a	yes
T1_DE_FZK	91%(594)	combined	100%	R	GOCDB	yes
T1_ES_PIC	100%(600)	combined	100%	R	GOCDB	yes
T1_FR_CCIN2P3	99%(500)	combined	80%	W	GOCDB	no
T1_IT_CNAF	100%(600)	combined	100%	R	GOCDB	yes
T1_TW_ASGC	n/a	combined	0%	NR	GOCDB	n/a
T1_UK_RAL	100%(600)	combined	100%	R	GOCDB	yes
T1_US_FNAL	100%(600)	combined	100%	R	OIM	yes
T2_AT_Vienna	99%(600)	combined	64%	NR	GOCDB	yes
T2_BE_IIHE	100%(600)	combined	100%	R	GOCDB	no
T2_BE_UCL	100%(600)	combined	100%	R	GOCDB	yes
T2_BR_SPRACE	87%(797)	combined	76%	R	OIM	no
T2_BR_UERJ	100%(475)	combined	100%	R	OIM	no
T2_CH_CAF	n/a	combined	n/a	R	n/a	n/a
T2_CH_CSCS	100%(600)	combined	100%	R	GOCDB	yes
T2_CN_Beijing	99%(700)	combined	100%	R	GOCDB	yes
T2_DE_DESY	99%(600)	combined	92%	R	GOCDB	yes
T2_DE_RWTH	n/a	combined	0%	W	GOCDB	no
T2_EE_Estonia	100%(600)	combined	100%	R	GOCDB	yes
T2_ES_CIEMAT	98%(600)	combined	100%	R	GOCDB	yes
T2_ES_IFCA	18%(200)	combined	0%	W	GOCDB	no
T2_FI_HIP	n/a	combined	100%	R	GOCDB	yes
T2_FR_CCIN2P3	n/a	n/a	80%	R	GOCDB	n/a
T2_FR_GRIF_IRFU	99%(600)	combined	100%	R	GOCDB	yes

Figure 4. A snapshot of the Site Status Board.

T2_ES_CIEMAT

	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Site Readiness Status:								R	R	R	R	R	W	R	R	R	R	R	R	R	R	R
Daily Metric:	E	E	E	O	O	O	E	O	E	O	O	O	E	O	O	O	O	O	O	O	O	O
Maintenance:	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up	Up
Job Robot:	100%	100%	99%	99%	100%	80%	75%	100%	100%	98%	100%	100%	100%	100%	100%	82%	100%	100%	100%	99%	100%	100%
SAM Availability:	52%	52%	52%	96%	100%	100%	96%	96%	44%	96%	88%	100%	76%	92%	100%	100%	100%	100%	100%	100%	100%	100%
T2::uplinkT1s:	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	8	8	8	8	8	8	8	8
T2::downlinkT1s:	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	5	5	5	5	5	5	5	5

Feb

Figure 5. Example of Site status metrics values and status for a site during a three-week period.

The site status history during the last two weeks is used to assess the stability and reliability of a site. Ranking plots like that in figure 6 are considered by data operations to select sites Ready-for-CMS to run CMS workflows. Tier-1/Tier-2 sites not in Ready status for more than 10%/20% of the time respectively during the last 15 days are considered not usable.

The site readiness program has already produced quite positive effects in site reliability. In the accumulated data over the past six months a clear trend of increasing number of Ready-for-

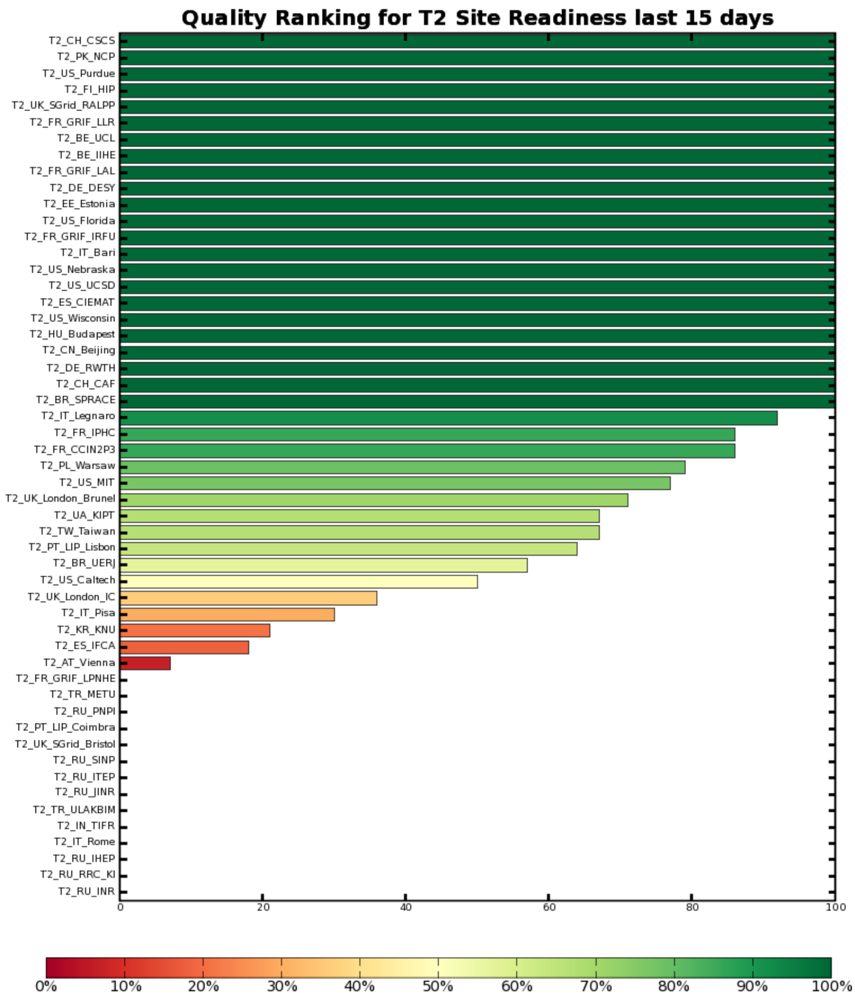

Figure 6. Site status ranking with the percentage of the time in Ready status considering a two-weeks period

CMS sites is observed. In April 2009 about 70% of the sites lie in this category, representing more than 90% of the resources available to CMS. The site readiness machinery is very useful for continuously monitoring the Grid and CMS services at the sites and helps production and users to select reliable sites. There is however still room for improving the site reliability. Efforts are now focused in determining the usual failures, help sites to improve and provide feedback for robustness of the CMS tools and services. The site commissioning and readiness monitoring has become a regular activity inside the computing facilities operation area.

2.4. Commissioning distributed production and analysis

Another important activity of the PADA task force is the systematic integration and validation of the CMS workload management tools, ProdAgent [10] for organized data processing [5, 6], and CRAB [11]/CRAB-server [12] for distributed analysis [7]. The tools have been scale-tested with various Grid Workload Management Systems (WMS), the gLite WMS, Condor-G and glidein WMS, a pilot-job-based approach in use in CMS.

Functionality, reliability and scaling tests have been conducted as well as the integration of new components of the production and analysis systems. Scaling tests have shown that each instance of ProdAgent or CRAB-server can sustain a rate of 30-50k jobs per day. Any number of instances can be run in parallel in order to scale the system up to the expected number of 250k jobs/day CMS will be running.

The integration, testing and commissioning of the data and workload management tools has become a regular activity within CMS computing. A formal validation procedure is in place that allows a better understanding of the functionality and performance of the tools before they go into operations. The result is an improvement in production efficiency.

CMS has recently started to run at the Tier-1 sites the so-called backfill jobs whenever there is no real production activity. These jobs are realistic data processing jobs which continuously test at scale the Tier-1 sites. It is certainly an important diagnostic tool for the operations team and the sites.

Figure 7 shows the number of jobs executed by CMS as a function of time during the last year classified according to the various activities. About 70k jobs are routinely run every day (25k analysis jobs, 25k JobRobot jobs and 20k production jobs). A volume of 150k jobs/day has been reached during computing challenge periods. We expect no problems in scaling up the WMS system with the current production and analysis WMS tools.

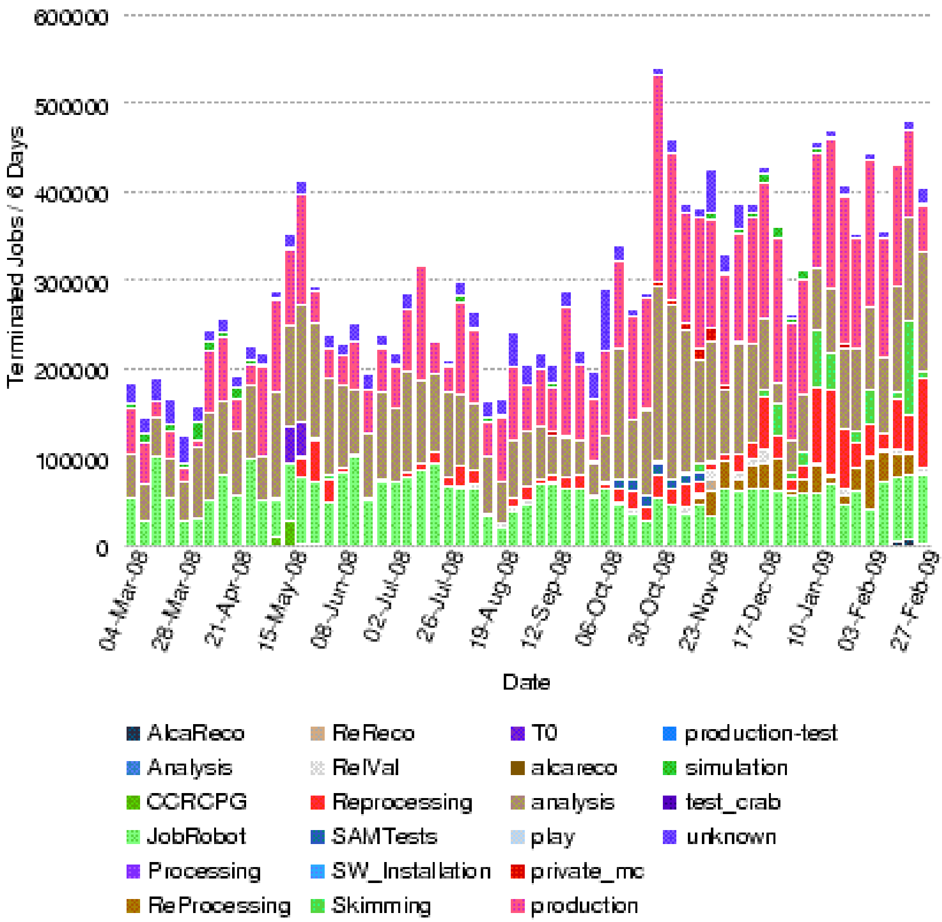

Figure 7. Distribution of jobs for one-year period classified according to the various activities.

Figure 8 shows the Grid and application job success rates for monitoring (JobRobot) jobs, production jobs and analysis jobs. Application inefficiencies include also site-related problems. Monitoring jobs have a global success rate (taking into account Grid and application/site efficiencies) of 85%. The submission environment and the job configuration are very well controlled in this kind of jobs, thus the higher efficiency compared to other types of jobs. The global efficiency for production jobs is 80%. Failed jobs are automatically resubmitted so that after few resubmissions almost all production jobs get done. However even small inefficiencies produce a significant operational overhead since all data have to be processed. Analysis jobs have only about 60% global efficiency. Failures are dominated by application crashes and inefficient remote stage out of the output products. CMS is developing a system for asynchronously transferring the output files into the final destination using a temporary buffer at the local site where the job was executed.

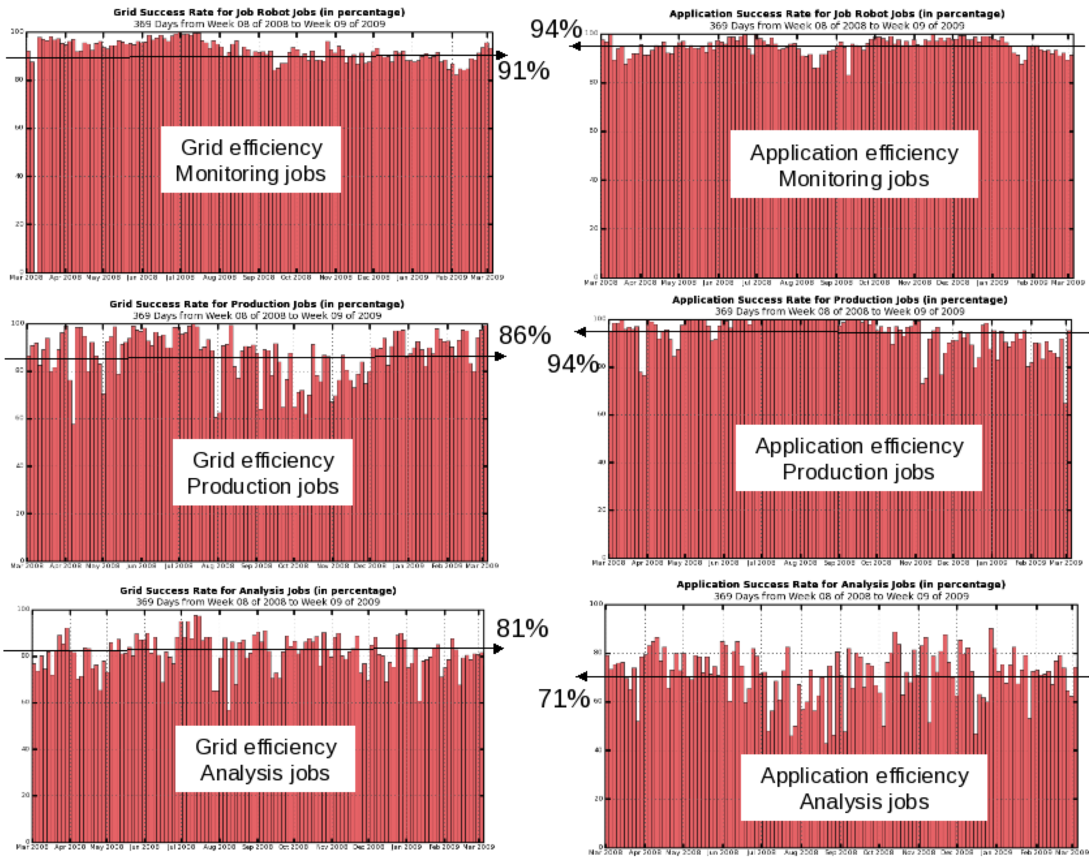

Figure 8. Job success rates for the last year for CMS jobs. Monitoring jobs correspond to the upper plots, production jobs the middle plots and analysis jobs the bottom plots. Grid efficiencies are displayed in the left hand side plots and application efficiencies on the right hand side plots.

3. Other integration activities

Apart from the PADA task force, several other initiatives have been taken directed towards the integration and commissioning of the CMS distributed computing system.

The Analysis Support Task Force (ASTF) was formed [8] with the aim of improving the reliability of analysis jobs. Its mandate was to perform a thorough analysis of the failures, evaluate the available monitoring tools, assess the user support techniques, and collect direct feedback from users about the distributed analysis system. In this respect a comprehensive survey was conducted. The task force will transition into a stable group supporting analysis operations. The ASTF concluded that most of the technical requirements to do effective analysis operations exist today or are under development. It made recommendations for what the Analysis Operations team should look like in future. Metrics to understand the system as a whole and identify problems should be developed; data placement, consistency and quality checks of analysis datasets should be actively managed; reference sites to facilitate debugging problems should be established; the coordination with the rest of the computing groups should be enhanced and the operation of the distributed analysis machinery (CRAB servers, WMS servers, etc) should be strengthened.

Given the limited manpower available for computing development, integration and operation, integration campaigns were regularly conducted in order to focus the effort into specific development targets driven by an imminent operational need. A campaign identifies an issue that needs development effort, integration tests and operational commissioning to bring some piece of the system up to a usable level, and directly applies manpower for a defined time from all areas to address the issues. Starting a campaign is basically a commitment from all sides involved to work upon the identified issue to some satisfactory resolution. Successful campaigns rolling out the data consistency tools, some of the components of the production/analysis systems, etc, were carried out. The campaigns have been a useful instrument to make progress in a labour-limited environment.

4. Summary

Integration and commissioning activities have been crucial in CMS for bringing the distributed computing system into stable and scalable operations. Computing challenges, task forces, integration campaigns and end-to-end tests have very well served this purpose. Load generator tools for data transfers and job submission have been very useful in testing the system at large scale.

The CMS data and job management system are scaling well. Routine large scale data transfers and job submissions are performed efficiently. There is a continuous improvement in reliability and robustness of the sites and the tools.

Production and analysis tools, operations and the sites have greatly benefited from the computing integration and commissioning program.

References

[1] C.Grandi, D.Stickland, L.Taylor et al. *The CMS Computing Model* CERN-LHCC-2004-035/G-083 (2004).
[2] P. Saiz et al. *Generic monitoring solution for LHC site commissioning activity and LHC computing shifts.* These proceedings.
[3] J. Letts et al. *Debugging Data Transfers in CMS.* These proceedings.
[4] J. Flix et al. *The commissioning of CMS sites: improving site reliability.* These proceedings.
[5] G. Codispoti et al. *Use of the gLite WMS in CMS for production and analysis.* These proceedings.
[6] S. Padhi et al. *Use of glide-ins in CMS for production and analysis.* These proceedings.
[7] A. Fanfani et al. *Commissioning Distributed Analysis at the CMS Tier-2 Centers.* These proceedings.
[8] J. Letts et all. *CMS Analysis Operations.* These proceedings.
[9] R. Egeland. T. Wildish and S. Metson,*Data transfer infrastructure for CMS data taking.* PoS (ACAT08) 033.
[10] D. Evans et al., *CMS MC Production System Development and Design.* Proceedings of Computing in High Energy Physics (CHEP07). Victoria, Canada (2007).
[11] M. Corvo et al., *CRAB, a tool to enable CMS Distributed Analysis.* Proceedings of Computing in High Energy Physics (CHEP06), Mumbai, India (2006).
[12] D. Spiga et al., *Automatization of user analysis workflow in CMS.* These proceedings.