**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2015 Т. 7 № 3 С. 455−462**

Ки&М

УДК: 004.75

JINR TIER-1-level computing system for the CMS experiment at LHC: status and perspectives

N. S. Astakhov, A. S. Baginyan, S. D. Belov, A. G. Dolbilov, A. O. Golunov, I. N. Gorbunov, N. I. Gromova, I. A. Kashunin, V. V. Korenkov [a], V. V. Mitsyn, S. V. Shmatov, T. A. Strizh, E. A. Tikhonenko, V. V. Trofimov, N. N. Voitishin, V. E. Zhiltsov

Joint institute for nuclear researches, Laboratory of Information Technologies,
Joliot-Curie, 6, Moscow reg., Dubna, 141980, Russia

E-mail: [a] korenkov@jinr.ru

Received October 28, 2014

The Compact Muon Solenoid (CMS) is a high-performance general-purpose detector at the Large Hadron Collider (LHC) at CERN. A distributed data analysis system for processing and further analysis of CMS experimental data has been developed and this model foresees the obligatory usage of modern grid-technologies. The CMS Computing Model makes use of the hierarchy of computing centers (Tiers). The Joint Institute for Nuclear Research (JINR) takes an active part in the CMS experiment. In order to provide a proper computing infrastructure for the CMS experiment at JINR and for Russian institutes collaborating in CMS, Tier-1 center for the CMS experiment is constructing at JINR. The main tasks and services of the CMS Tier-1 at JINR are described. The status and perspectives of the Tier1 center for the CMS experiment at JINR are presented.

Keywords: grid computing, CMS experiment, CMS Tiers

The work is carried out within the federal program “Research and development on the priority directions of the development of the scientific — technological complex in Russia for 2007–2013” (contract №07.524.12.4008).

© 2014 Николай Степанович Астахов, Андрей Сергеевич Багинян, Сергей Дмитриевич Белов, Андрей Геннадьевич Долбилов,
Алексей Олегович Голунов, Илья Николаевич Горбунов, Наталья Ивановна Громова, Иван Андреевич Кашунин,
Владимир Васильевич Кореньков, Валерий Валентинович Мицын, Сергей Владимирович Шматов,
Татьяна Александровна Стриж, Елена Александровна Тихоненко, Владимир Валентинович Трофимов,
Николай Николаевич Войтишин, Виктор Евгеньевич Жильцов

**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2015 Т. 7 № 3 С. 455–462**

Ки&М

Статус и перспективы вычислительного центра ОИЯИ 1-го уровня (TIER-1) для эксперимента CMS на большом адронном коллайдере

Н. С. Астахов, А. С. Багинян, С. Д. Белов, А. Г. Долбилов, А. О. Голунов, И. Н. Горбунов, Н. И. Громова, И. А. Кашунин, В. В. Кореньков, В. В. Мицын, С. В. Шматов, Т. А. Стриж, Е. А. Тихоненко, В. В. Трофимов, Н. Н. Войтишин, В. Е. Жильцов

*Лаборатория информационных технологий, Объединенный институт ядерных исследований
Россия, 141980, г. Дубна, ул. Жолио-Кюри, д. 6*

Компактный мюонный соленоид (CMS) — высокоточная детекторная установка на Большом адронном коллайдере (LHC) в ЦЕРН. Для осуществления обработки и анализа данных в CMS была разработана система распределенного анализа данных, предполагающая обязательное использование современных грид-технологий. Модель компьютинга для CMS — иерархическая (в смысле создания вычислительных центров разного уровня). Объединенный институт ядерных исследований (ОИЯИ) принимает активное участие в эксперименте CMS. В ОИЯИ создается центр 1-го уровня (Tier1) для CMS с целью обеспечения необходимой компьютерной инфраструктурой ОИЯИ и российских институтов, участвующих в эксперименте CMS. В работе описаны основные задачи и сервисы центра Tier1 для CMS в ОИЯИ и представлены статус и перспективы его развития.

Ключевые слова: грид компьютинг, эксперимент CMS, центры CMS (Tiers)

Citation: *Computer Research and Modeling*, 2015, vol. 7, no. 3, pp. 455–462.

© 2014 Николай Степанович Астахов, Андрей Сергеевич Багинян, Сергей Дмитриевич Белов, Андрей Геннадьевич Долбилов, Алексей Олегович Голунов, Илья Николаевич Горбунов, Наталья Ивановна Громова, Иван Андреевич Кашунин, Владимир Васильевич Кореньков, Валерий Валентинович Мицын, Сергей Владимирович Шматов, Татьяна Александровна Стриж, Елена Александровна Тихоненко, Владимир Валентинович Трофимов, Николай Николаевич Войтишин, Виктор Евгеньевич Жильцов

Introduction

The 6 million billion proton-proton collisions were produced by the Large Hadron Collider (LHC) [The Large Hadron Collider] at CERN in its first physics Run (2010–2012). Around 5 billion of these collisions were recorded in real time by the ATLAS and CMS experiments each for further processing, reconstruction of physics objects and physics analysis. Including simulation events, all in all, the LHC experiments have generated during the LHC Run 1 about 200 PB of data.

Data storage, processing and analysis of such a huge amount of data have been completed in the framework of the distributed computing infrastructure within the Worldwide LHC Computing Grid (WLCG) Project [LHC Computing…, 2005]. The WLCG computing model joints the three-level recourse centers (tiers) and originally assumed hierarchical structure according to their functionality. Then data distribution and replication was optimized by allowing transfers between any two centres. Now WLCG is formed by more than 170 centres spread around the world among them the Tier-0 center in CERN and thirteen Tier-1 centers. 2 million jobs run every day in this global infrastructure [The Worldwide LHC Computing Grid].

CMS Tier1 center at JINR (Dubna)

For CMS Tier1 centers are in Germany, United Kingdom, USA (FNAL), Italy, France, Spain, Taipei and JINR (Dubna). Starting 2011 the WLCG Tier-1 site is under development in the Russian Federation for all four LHC Experiments [CMS Dashboard; Korenkov 2013]. The special Federal Target Programme Project is aimed to construction of a Tier-1 computer-based system in National Research Center "Kurchatov Institute" and JINR for processing experimental data received from LHC and provision of grid services for a subsequent analysis of the data at the distributed centers of the LHC computing grid. It is shared that the National Research Center "Kurchatov Institute" is responsible for support of ALICE, ATLAS, and LHCb experiments, while the JINR provides Tier-1 services for the CMS experiment. In 2012 the WLCG Overview Board approved the plan of creating a Tier1-level center in Russia.

In agreement with the CMS Computing model [Grandi, Stickland, Taylor, 2005], the JINR Tier-1 site will provide acceptance of an agreed share of raw data and Monte Carlo data and provision of access to the stored data by other CMS Tier-2/Tier-3 sites of the WLCG, will serve FTS-channels for Russian and Dubna Member States Tier-2 storage elements including monitoring of data transfers.

The Tier1 CMS infrastructure at JINR consists of the following elements (services) (Figure 1): Data storage subsystem, Computing system (Computing Elements), Data transfer subsystem (FTS), Management of data transfer and data storage (CMS VOBOX), Load distribution system, and CMS Tier-1 network infrastructure.

Since October 2013 the JINR Tier-1 supports CMS as the tape-less Tier-1 with 1200 cores (17K HS06) and 450 TB disk-only dCache storage system. The prototype of mass-storage system constitutes 130 TB dCache pools and 72 TB tapes. All required grid services were installed and successfully validated and tested for high memory (6GB) jobs, in particular, File Transfer Service FTS 2.2.8, CMS Data Transfers service PhEDEx 4.1.2 for disk-only dCache and MSS, authorization and authentication Argus service, 2x Frontier Squids (access to the calibration data via the local cache), site BDII (Berkeley DB Information Index) and top-level BDII, User Interface service UI, Credential Management Service MyProxy, Workload Management System WMS, Logging and Bookkeeping service LB, LCG File Catalog — LFC for internal testing.

Since the LHC Run-2 start-up in line with the WLCG and LHC Experiments requirements, the JINR has to provide a support of a number of the main Tier-1 services for the CMS experiment: user-visible services (Data Archiving Service, Disk Storage Services, Data Access Services, Reconstruction Services, Analysis Services, User Services) and specialized system-level services (Mass storage system, Site security, Prioritization and accounting, Database Services).

The network bandwidth as part of LHCOPN for Tier-0-Tier-1 and Tier-1-Tier-1 connections was about 2 Gbps for 2012 and now is 10 Gbps. The JINR link to public network with a bandwidth of 20 Gbps is used to connect the Tier-1 with all other Tier-2/Tier-3 sites.

In 2015 the CMS Tier-1 site in JINR will provide computing facilities about 10% of the total existing CMS Tier-1 resources (excluding CERN).

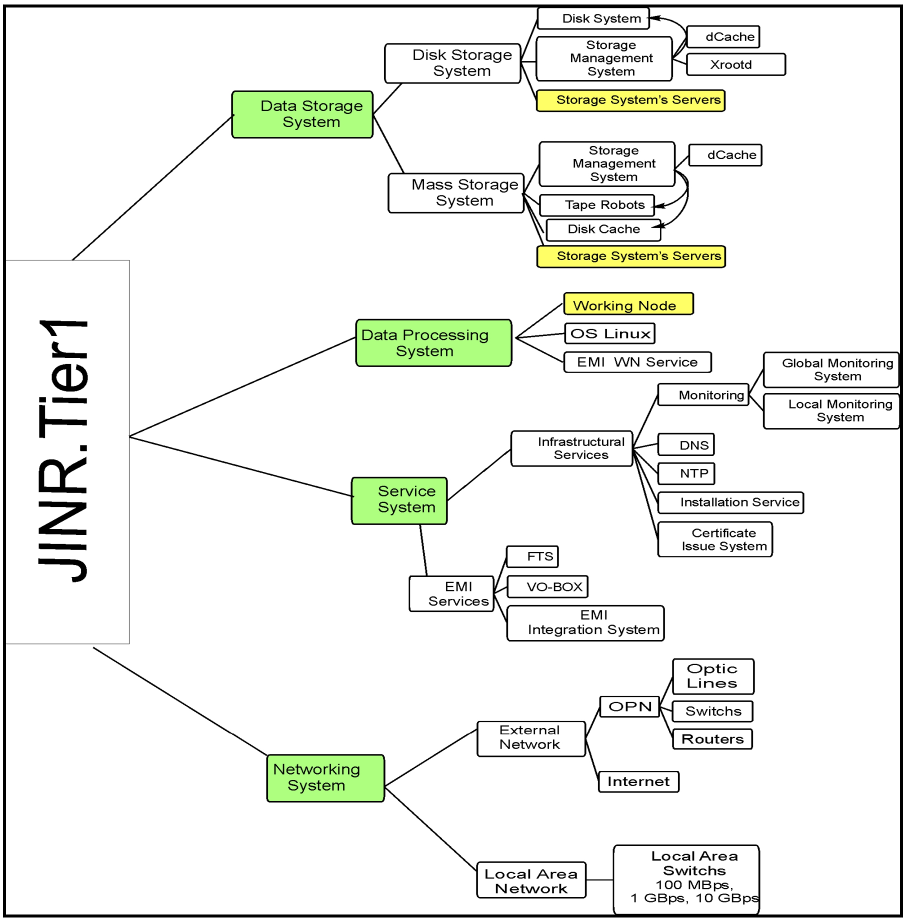


Fig. 1. JINR Tier1 infrastructure scheme

Structure of JINR CMS Tier1 local network

Figure 2 presents the Tier1 network topology at JINR. To realize it, we have to commutate 160 disk servers, 25 blade servers and 60 infrastructure servers.

For the networks with a star topology, each network host is connected to a central node with a point-to-point connection. All traffic passes through the central node. An advantage of the star topology is simplicity of supplementing additional nodes, while its primary disadvantage is that the hub represents a single point of failure. The type of the network topology in which some of the nodes of the network are connected to more than one other node in the network makes it possible to take advantage of some of the redundancy that is provided by a physical connected mesh topology. A fully connected network is a communication network in which each of the nodes is connected with one another. In graph theory it is known as a complete graph [Education-portal].

Network designers implement mesh topology and Spanning Tree Protocol (STP) on switches in order to prevent loops in the network, i.e. to use STP in situations where you want redundant links, but not loops. A failure of primary links activates the backup links so that users can continue using the network. Without STP on switches, such a failure can result in a loop. STP defines a tree that spans all the switches in an extended network. STP forces certain redundant data paths into a blocked state and leaves other paths in a forwarding state. If a link in the forwarding state becomes unavailable, STP reroutes data paths through the activation of the standby path.

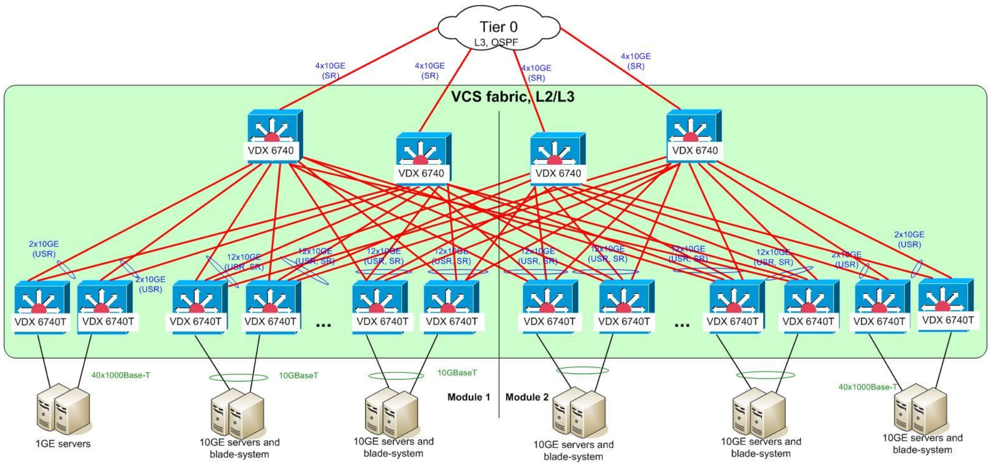


Fig. 2. Network topology of JINR CMS Tier 1 center

Recently the IS-IS protocol for network calculating has been applied. For this protocol Dijkstra's algorithm from graph theory is used. It compares and calculates the shortest path through all nodes in the network. It is constructing a shortest-path tree from the first vertex to every other vertex in the graph. On its basis a modern protocol Transparent Interconnection of Lots of Links (TRILL) was developed [Transparent Interconnection of…].

A new Layer 2 routing protocol, Transparent Interconnection of Lots of Links (TRILL), offers many advantages over STP. While STP maintains a single path and blocks all redundant paths, TRILL provides Layer 2 multipath and multi-hop routing. The proposed TRILL protocol enhances Layer 2 routing by introducing multipath and multihop routing capabilities. The new capabilities represent significant improvements over Spanning Tree Protocol (STP). TRILL provides Layer 2 multiple paths by splitting the traffic load among several paths. TRILL also is faster at self-healing during a network failure. While one link gets unavailable others continue transfer traffic. The maximum reduction in a network bandwidth does not exceed 50 % and taking into account our design in the network Tier 1 at JINR it is not more than 25 %. Accordingly, all the nodes continue to operation, only their bandwidth decreases.

In conclusion, it is worth noting, that the topology Tier 1 at JINR will be based on protocol TRILL. This protocol will help network designers to create a coherent Virtual Cluster Switching (VCS) fabric with distributed switches and allow creating highly reliable, mobile and multi-port systems.

Monitoring and statistics

The various metrics (example for JINR Tier-1 is given in Fig. 3) are based on the result of common WLCG tests and CMS specific tests, in particular NAGIOUS tests are applied (Fig. 4). These tests are used to establish site availability and readiness for the CMS collaboration usage. Test results are summarized in the status summary table at Site Status Board (SSB) monitoring at the CMS Dashboard [CMS Dashboard, URL].

The JINR CMS Tier1 site shows good results in the monitoring rank of CMS Tier1 sites on availability and readiness (see Fig.5).

The JINR Tier sites is enabled to process more than 230 000 jobs per months (Fig. 6, left), i. e. about 6 % of all CMS jobs, with very high efficiency (~ 90%) (Fig. 6, right). The utilization metrics shown the efficiency of usage of job slots for all CMS Tier-1 sites are given in Figure 7.

Summary

Current status and activities on creation of CMS Tier1 center at JINR were reported at NEC'2013 [Korenkov, 2013] and GRID'2012 conferences and are published in [Korenkov et al., 2012; Astakhov et al., 2013; Astakhov et al., 2013/04]. In February 2015 the JINR CMS Tier1 resources will be increased to the level that was outlined in JINR's rollout plan: CPU 2400 cores (28800 HEP-Spec06), 2.4 PB disks, and 5.0 PB tapes. It is planned the JINR CMS Tier-1 site will be included in the WLCG infrastructure as a Tier-1 production-level member with resources indicated above from February 2015 for WLCG use by the CMS Collaboration.

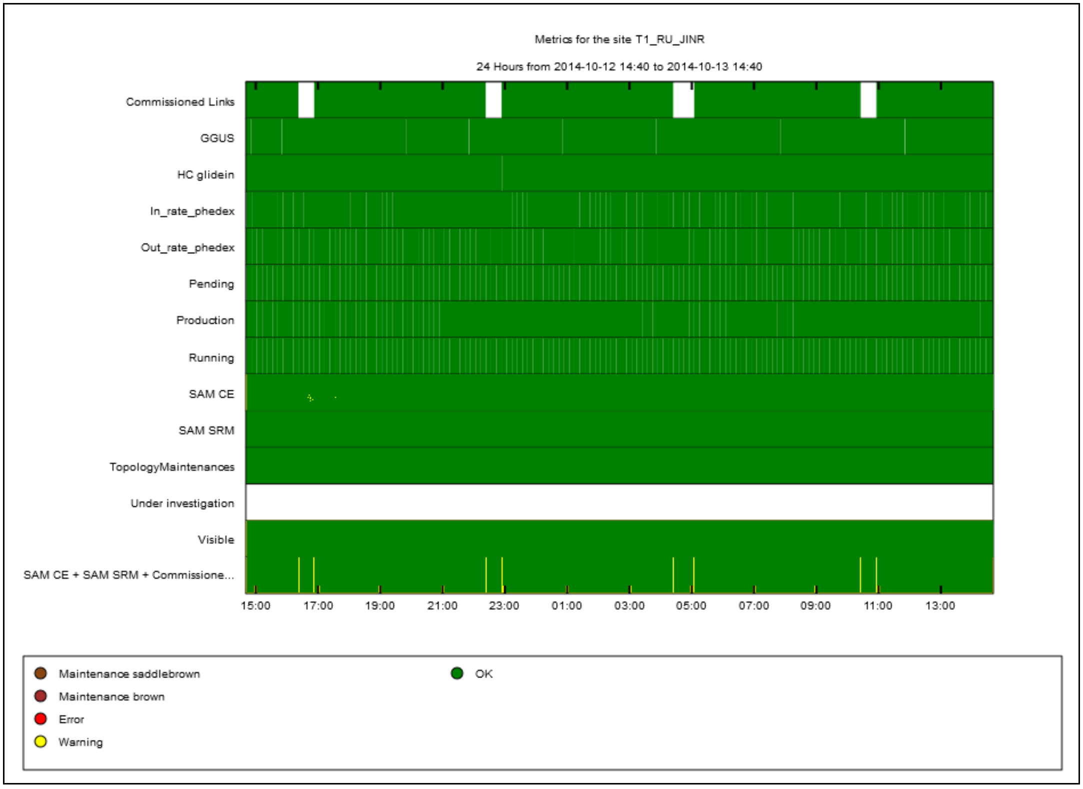
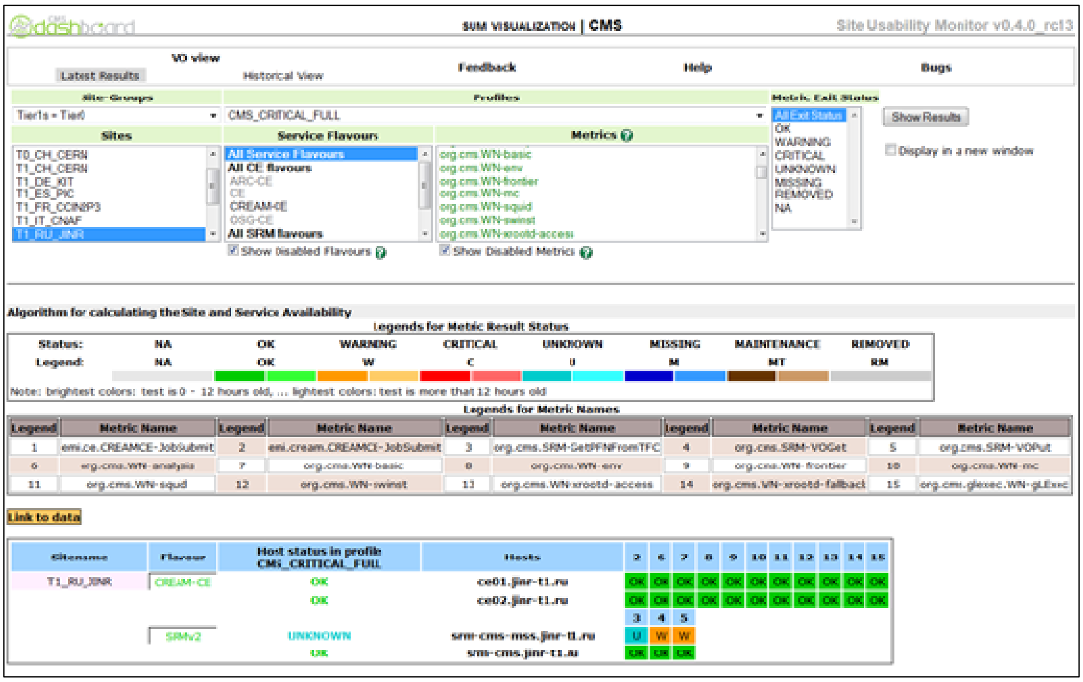


Fig. 3. Metrics for the JINR Tier-1 site



Site-Groups: Tier1s + Tier0

Profiles: CMS_CRITICAL_FULL

Metric Exit Status: All Exit Status, OK, WARNING, CRITICAL, UNKNOWN, MISSING, REMOVED, NA

☐ Display in a new window

Sites: T0_CH_CERN, T1_CH_CERN, T1_DE_KIT, T1_ES_PIC, T1_FR_CCIN2P3, T1_IT_CNAF, T1_RU_JINR

Service Flavours: All Service Flavours, All CE flavours, ARC-CE, CE, CREAM-CE, OSG-CE, All SRM flavours

☑ Show Disabled Flavours

Metrics: org.cms.WN-basic, org.cms.WN-env, org.cms.WN-frontier, org.cms.WN-mc, org.cms.WN-squid, org.cms.WN-swinst, org.cms.WN-xrootd-access

☑ Show Disabled Metrics

Algorithm for calculating the Site and Service Availability

Legends for Metric Result Status

Status:	NA	OK	WARNING	CRITICAL	UNKNOWN	MISSING	MAINTENANCE	REMOVED
Legend:	NA	OK	W	C	U	M	MT	RM

Note: brightest colors: test is 0 - 12 hours old, ... lightest colors: test is more that 12 hours old

Legends for Metric Names

Legend	Metric Name	Legend	Metric Name	Legend	Metric Name	Legend	Metric Name	Legend	Metric Name
1	emi.ce.CREAMCE-JobSubmit	2	emi.cream.CREAMCE-JobSubmit	3	org.cms.SRM-GetPFNFromTFC	4	org.cms.SRM-VOGet	5	org.cms.SRM-VOPut
6	org.cms.WN-analysis	7	org.cms.WN-basic	8	org.cms.WN-env	9	org.cms.WN-frontier	10	org.cms.WN-mc
11	org.cms.WN-squid	12	org.cms.WN-swinst	13	org.cms.WN-xrootd-access	14	org.cms.WN-xrootd-fallback	15	org.cms.glexec.WN-gLExec

Sitename	Flavour	Host status in profile CMS_CRITICAL_FULL	Hosts	2	6	7	8	9	10	11	12	13	14	15
T1_RU_JINR	CREAM-CE	OK	ce01.jinr-t1.ru	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
		OK	ce02.jinr-t1.ru	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK
				3	4	5								
	SRMv2	UNKNOWN	srm-cms-mss.jinr-t1.ru	U	W	W								
		OK	srm-cms.jinr-t1.ru	OK	OK	OK								

Fig. 4. The JINR Site usability based on the NAGIOS test

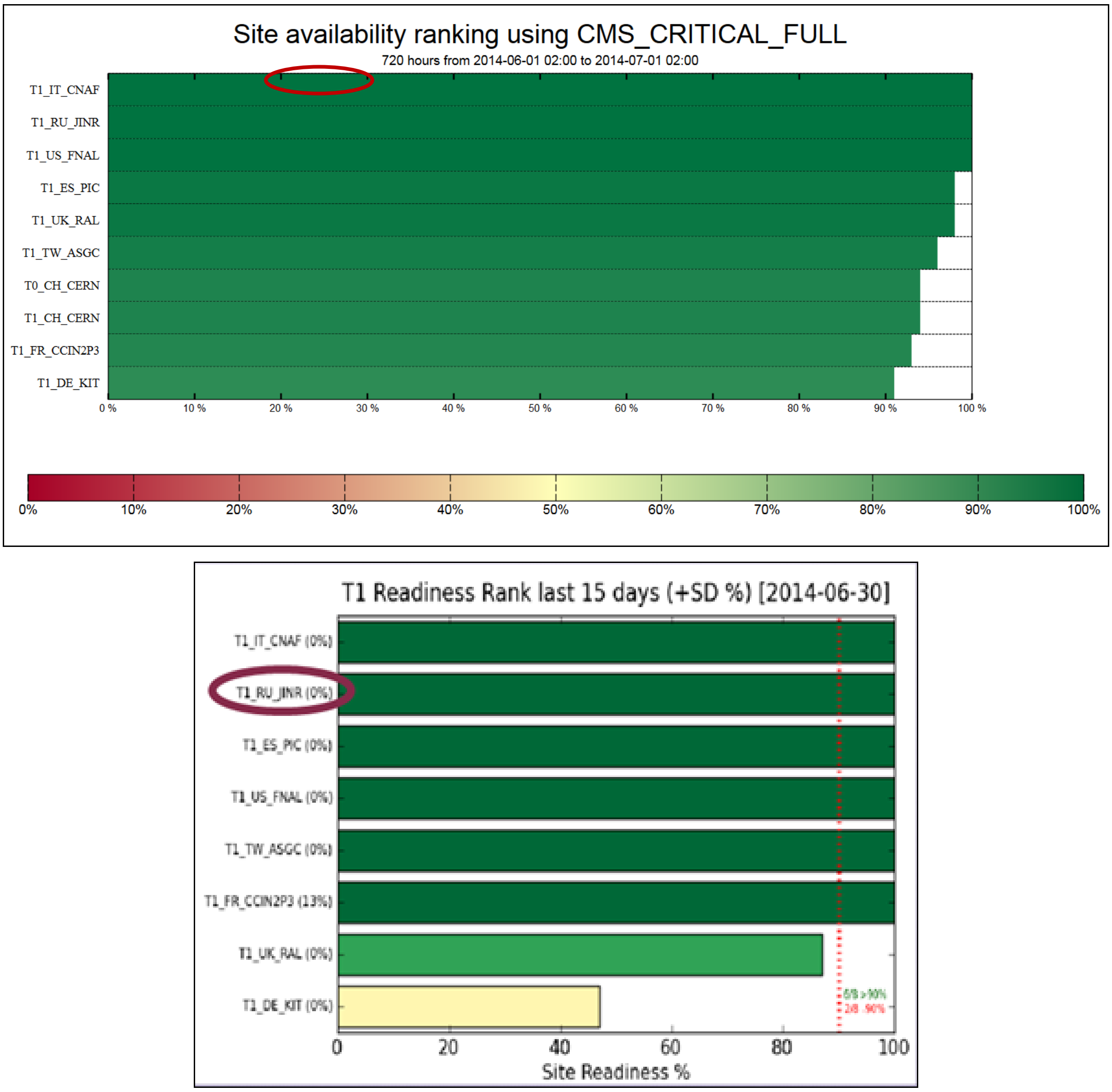
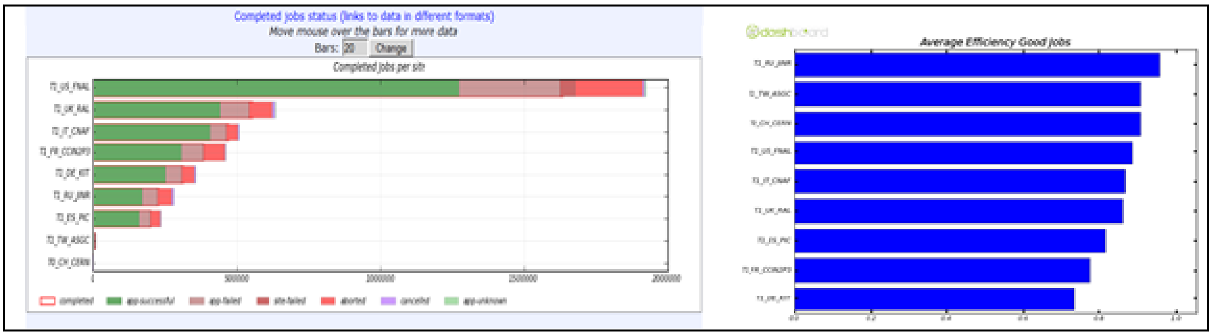


Fig. 5. Availability (top) and readiness(bottom) of CMS Tier-1 Centers


Completed jobs status (links to data in different formats)

Move mouse over the bars for more data

Bars: 20 Change

Fig. 6. Complited jobs per site for one month (left) and Average CPU efficiency for Good Jobs (right)

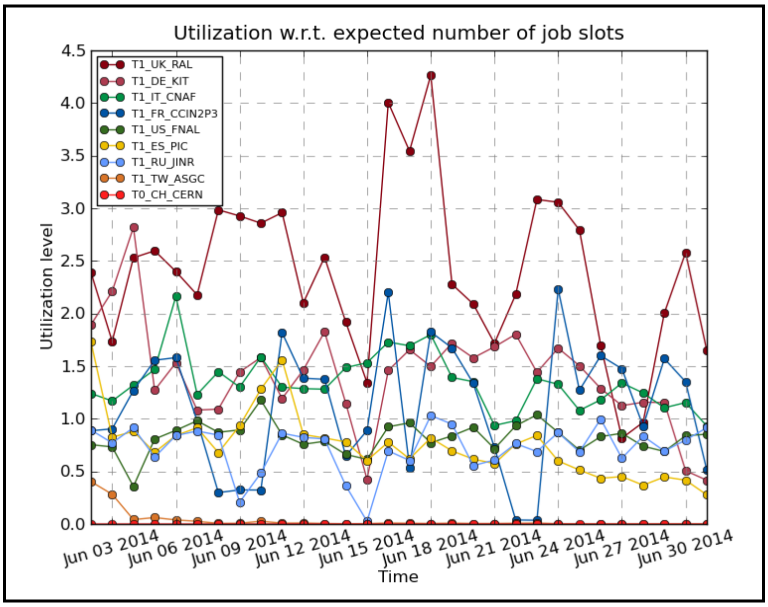


Fig. 7. CMS T1 Site Activity Summary (utilization level)

References

Astakhov N. S., Belov S. D., Dmitrienko P. V., Dolbilov A. G., Gorbunov I. N., Korenkov V. V., Mitsyn V. V., Shmatov S. V., Strizh T. A., Tikhonenko E. A., Trofimov V. V., Zhiltsov V. E. CMS Tier-1 at JINR, in NEC’2013 Proceedings, Dubna, 2013, pp.19–23.

Astakhov N. S., Belov S. D., Gorbunov I. N., Dmitrienko P. V., Dolbilov A. G., Zhiltsov V. E., Korenkov V. V., Mitsyn V. V., Strizh T. A., Tikhonenko E. A., Trofimov V. V., Shmatov S. V. The Tier-1-level computing system of data processing for the CMS experiment at the Large Hardon Collider. 15 p., “Information Technologies and Computation Systems”, 2013/04, pp. 27–36 (in Russian).

CMS Dashboard. http://dashb-ssb.cern.ch/dashboard/request.py/siteviewhome

Education-portal. http://education-portal.com/academy/lesson/how-star-topology-connects-computer-networks-in-organizations.html#lesson

Grandi C., Stickland D., Taylor L. CMS NOTE 2004-031 (2004), CERN LHCC 2004-035/G-083; CMS Computing Technical Design Report, CERN-LHCC-2005-023 and CMS TDR 7, 20 June 2005.

Korenkov V.V. CMS Tier 1 at JINR. // XXIV International Symposium on Nuclear Electronics & Computing, NEC2013. 2013. http://nec2013.jinr.ru/files/13/Korenkov.ppt

Korenkov V. V., Astakhov N. S., Belov S. D., Dolbilov A. G., Zhiltsov V. E., Mitsyn V. V., Strizh T. A., Tikhonenko E. A., Trofimov V. V., Shmatov S. B. Creation at JINR of the data processing automated system of the TIER-1 level of the experiment CMS LHC. // Proceedings of the 5th Inter. Conf. "Distributed Computing and Grid-technologies in Science and Education", ISBN-5-9530-0345-2, Dubna, 2012, pp. 254-265 (in Russian).

LHC Computing Grid Technical Design Report. CERN-LHCC-2005-024, 2005; Worldwide LHC Computing Grid (WLCG), http://lcg.web.cern.ch/LCG/public/default.htm

The Large Hadron Collider. http://home.web.cern.ch/topics/large-hadron-collider

The Worldwide LHC Computing Grid. http://wlcg-public.web.cern.ch/

Transparent Interconnection of Lots of Links. http://www.ipinfusion.com/products/zebos/protocols/data-center-ethernet/TRILL