WLCG Tier-2 site at NCP, Status Update and Future Direction

Saqib Haleem,[*] **Fawad Saeed**,[†] **Adeel-ur-Rehman**,[‡] **Muhammad Imran**[§¶]

National Centre for Physics, Islamabad, Pakistan.
E-mails: saqib.haleem@ncp.edu.pk, fawad.saeed@ncp.edu.pk,
adeel.rehman@ncp.edu.pk, muhammad.imran@ncp.edu.pk

The National Centre for Physics (NCP) in Pakistan maintains a computing infrastructure for the scientific community. A major portion of the computing and storage resources are reserved for the CMS experiment through the WLCG infrastructure, and a small portion of the computing resources are reserved for other non experimental high-energy physics (EHEP) scientific experiments. For efficient utilization of resources, many scientific organizations have migrated their resources into infrastructure-as-a-service (IaaS) facilities. The NCP has also taken such an initiative last year, and has migrated most of their resources into an IaaS facility. An HT-condor based batch system has been deployed for the local experimental high energy physics community to allow them to perform their analysis task. Recently we deployed an HT-Condor compute element (CE) as a gateway for the CMS jobs. On the network side, our Tier-2 site is completely accessible and operational on IPv6. Moreover, we recently deployed a Perfsonar node to actively monitor the throughput and latency issues between NCP and other WLCG sites. This paper discusses the status of NCP Tier-2 site, its current challenges and future directions.

---

[*]Saqib works as the Manager IT at NCP.

[†]Fawad works as the Grid Computing Manager at NCP.

[‡]Adeel works as the Manager Software Development at NCP.

[§]Speaker.

[¶]Dr. Imran works as a Senior Scientific Officer at NCP.

1. Introduction

The National Centre for Physics has been established to promote research in physics and applied disciplines in Pakistan [15]. NCP will evolve into a centre of excellence in the field of physics and related emerging disciplines such as nanoscience, vacuum science and earthquake studies. NCP follows the research directions of the Abdus Salam International Centre for Theoretical Physics (AS-ICTP) in Trieste, Italy. The mission of the NCP is to break the isolation of Pakistani researchers from the international scientific community and to establish links between industry and research institutes for the development of society in Pakistan.

NCP started its scientific activities on January 27, 1999 and obtained its charter from the government of Pakistan in April 2004. Now it is an autonomous body funded by the government of Pakistan. The Centre was initially located at Quaid-i-Azam University. Since July 2006, NCP has its own building. Two advisory committees, namely the International Advisory Scientific Council (IASC) and the National Advisory Scientific Program Council (NASPC) guide the scientific and academic proceedings of the NCP.

The main objective of the Centre is to raise the standard of physics at par with international norms of productivity and originality, and to act as an entity for acquisition, generation, transmission and dissemination of knowledge from the frontiers of physics for universities and Research and Development (R&D) organizations covering the broad areas of physics. NCP offers opportunity to work with scientists who are invited to lecture and also to participate in the Centre's research activities on a regular basis.

The experimental high energy physics (EHEP) group of NCP is actively involved in various collaborative research projects, particularly related to the Large Hadron Collider (LHC) at CERN, Switzerland [21]. Members of the EHEP group are contributing to the analysis of LHC data. The large amount of data produced at CERN has been managed by NCP through the Worldwide LHC Computing Grid (WLCG).

Pakistan has two operational grid computing nodes, of which only NCP operates the CMS Tier-2 site. It is registered with the name "NCP-LCG2" and connected to the WLCG infrastructure for the CMS experiment at CERN [7]. In this paper, we discuss the computing infrastructure of NCP for the scientific community. A major portion of the computing and storage resources are reserved for the CMS experiment in the context of the WLCG infrastructure, and a small portion of the computing resources are reserved for other scientific experiments. For efficient utilization of resources, many scientific organizations have migrated their resources to an IaaS facility. NCP has also taken such an initiative last year, and migrated most of their resources to an in-house IaaS facility for scientific use. An HT-condor [9] based batch system has been deployed for the local experimental high energy physics community to support them in performing their analysis task. Recently we deployed an HT-Condor compute element (CE) as a gateway for the CMS jobs. On the network side, our Tier-2 site is completely accessible and operational on IPv6. Moreover, we recently deployed a Perfsonar node [16] to actively monitor the throughput and latency issues between the NCP Tier-2 centre and other WLCG sites. This paper discusses the status of the NCP Tier-2 site, its current challenges, and future directions.

2. Computing Infrastructure at NCP

NCP is maintaining an IT infrastructure, which is mainly categorized into two areas: Core Computing Services (CCS) and Corporate IT Services (CISs). In CCS, NCP hosts a WLCG TIER-2 site that comprises 524 CPU cores and 540 TB of disk storage. Furthermore, a computing cluster of 96 CPU cores is installed for the local scientific community. In CISs we have provided several services such as e-mail, DNS, a public web site, FTP, and application databases, which are hosted inside the NCP data centre. All of these services are virtualized, collectively comprising more than 50 virtual machines (VMs). For facilitating dynamic management of the deployments underlying the CCS, an Openstack-based IaaS cluster has been configured.

2.1 WLCG at NCP

The NCP-LCG2 site is registered and connected to the WLCG infrastructure for the CMS experiment at CERN [7] since 2005. At that time, the NCP grid computing site was the first of its kind in South-Asia and fifth in Asia. A photographic representation illustrating its current setup is shown in Figure 1.

Figure 1: NCP Computing Infrastructure at a glance.

2.1.1 NCP-LCG2 and TIER-2 Site Requirements

The WLCG infrastructure as well as the CMS experiment specify requirements regarding the minimum capacity of resources a site should make available for its smooth functioning in terms of CPU, storage and network bandwidth [6]. A picture of installed resources at NCP compared to the officially recommended figures are presented in Table 1.

Table 1: NCP-LCG2 and TIER-2 Site Requirements [19]

Resources	Nominal	Recommended	Tolerated	Installed at NCP
CPU	10.9 kHS06	5 kHS06	4 kHS06	6.3 kHS06
Disk	810 TB	400 TB	300 TB	430 TB
Network	10 Gbps	1 Gbps	1 Gbps	1 Gbps

2.1.2 Hardware Specification

The NCP computing servers come from multiple vendors and comprise various models. The distribution of those are shown in Table 2. Similarly, storage systems used to store and analyze

Table 2: Computing Servers

Hardware	No. of Sockets	Cores	Quantity	Total Cores
Sun Fire X4150(Intel(R) Xeon(R) CPU X5460 @ 3.16GHz)	02	04	28	224
Dell Power Edge R610 (Intel (R) Xeon(R) CPU X5670 @ 2.93GHz)	02	06	25	300

WLCG data, multiple vendors as well as models distribution is presented in Table 3.

Table 3: Storage Servers

Storage Server	Total disks per server	Server Quantity	Raw Capacity
Transtec Lynx 4300	23 x 1 TB	15	345 TB
Dell Power Edge T620	12 X 4 TB	02	96 TB
DELL EMC MD1200	10 x 6 TB X 4 TB	02	120 TB

2.2 Status of NCP-LCG2 Site

The historical escalation of NCP installed resources during the past decade is depicted in Table 4. Also, the summary of NCP resource utilization statistics in the past few years in terms of kHS06-hours as well as number of jobs is expressed in Table 5 below. The resource utilization statistics have also been shown in terms of kHS06 as well as number of jobs both for year 2017 in Figure 2a and 2b respectively. It can be noticed in Figure 2b that the number of jobs increased in the month of February to more than 100k. This is because of the testing activities performed by our local user community.

Another significant metric of a grid site is measured in terms of amount of data downloaded as well uploaded from/to other WLCG sites. The corresponding data transfer statistics is given for data downloaded and uploaded during a period of last 6 months in Figures 3a and 3b respectively.

Apart from that, a critical performance oriented metrics of a grid site in WLCG is measured in terms of its availability (i.e. how much a site was active in a given time) and reliability (i.e. how much efficient a site is in processing the data in a given time). The availability / reliability statistics of the NCP grid site for the last 8 months are illustrated in Table 6.

2.3 Compute Elements (CEs)

The T2_PK_NCP is running three compute elements (CEs). The CREAM-CEs are connected to a portable batch system (PBS). Recently a new HT-condor CE has been installed, which is connected to an HT-Condor batch farm. The HT-condor CE and its compute farm have been installed on the virtual machines (VMs) in a local OpenStack-based IaaS cloud. In total, 524 CPU cores are

Table 4: Installed resources from 2008-2018

Year	CPU	HS06	Storage	Network Connectivity
Jan-08	14	67.2	3.2 TB	2 Mbps (Shared)
Apr-08	36	172.8	3.2 TB	2 Mbps (Shared)
Sept-08	74	355.2	3.2 TB	10 Mbps (Shared)
Feb-10	160	1600	3.2 TB	10 Mbps (Shared)
Jun-10	240	2400	69 TB	155 Mbps (Shared)
Dec-10	524	6365	87 TB	155 Mbps (Shared)
Jun-11	524	6365	175 TB	155 Mbps (Shared)
May-12	524	6365	260 TB	155 Mbps (Shared)
Oct-14	524	6365	330 TB	155 Mbps (Shared)
Apr-15	524	6365	330 TB	1 Gbps (Shared)
Mar-18	524	6365	561 TB	1 Gbps (Shared)

Table 5: NCP Resource utilization [17]

Year	kHS06-hours	No. of Jobs
2011	11,264	614,215
2012	15,165	630,629
2013	1,711	308,065
2014	3,654	165,002
2015	7,515	239,315
2016	5,897	347,339
2017	3,945	408,688

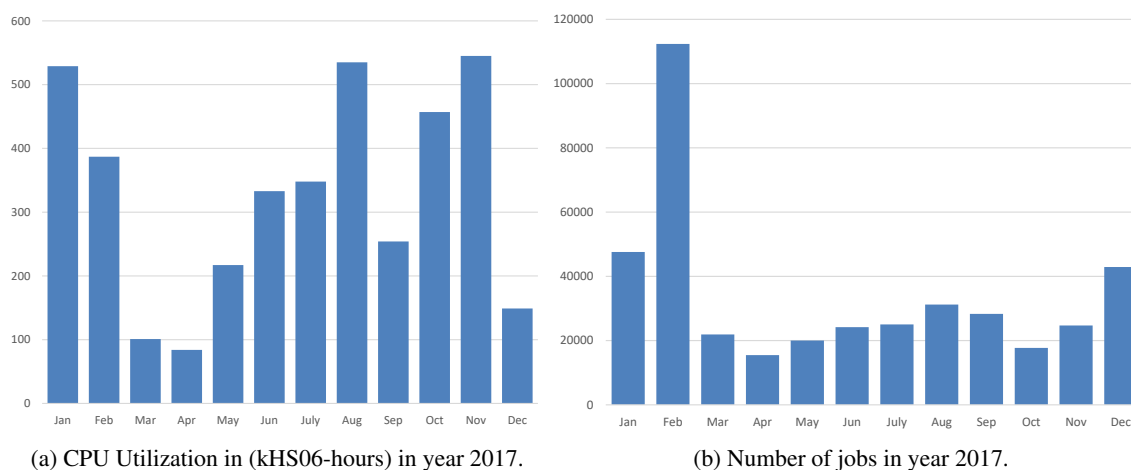(a) CPU Utilization in (kHS06-hours) in year 2017.

(b) Number of jobs in year 2017.

Figure 2: CPU Utilization of grid computing resources and the number of jobs run in the grid computing facility in year 2017.

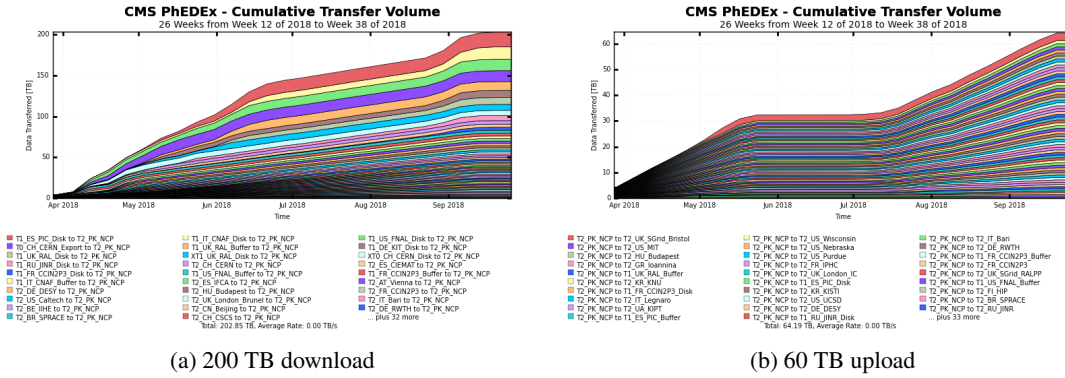(a) 200 TB download

(b) 60 TB upload

Figure 3: NCP site data transfer stats (last 6 months).

Table 6: NCP grid site availability and reliability [4]

Year-month	Availability	Reliability
2018-01	97.81	97.81
2018-02	99.18	99.18
2018-03	97.42	97.42
2018-04	99.80	99.80
2018-05	99.90	99.90
2018-06	99.59	99.59
2018-07	100	100
2018-08	99.94	99.94

provided for CMS jobs. Each compute VM comprises 8 cores with 16 GB of memory. All of the compute nodes support single core jobs. As CMS is moving towards multi-core jobs. Therefore HT-condor based compute nodes are now configured to also support multi-core jobs.

2.4 Grid Storage

NCP is using Disk Pool Manager (DPM) [10] based storage technology in its grid site. 12 DPM disks nodes are connected along a 'head' node to provide approximately 400 TB of storage capacity. All of the disk nodes are connected with a dual-stack network interface i.e. IPv6 and IPv4. In order to bring flexibility in the existing architecture and to make the grid storage capacity more useful, NCP is adopting software defined storage model. A Ceph-based distributed storage solution has been deployed which offers storage services in a flexible and reliable way. [5].

2.5 The PerfSonar node

Last year NCP deployed a PerfSonar node [16] for monitoring and troubleshooting throughput for issues encountered in the WLCG site commissioning process. It is configured with a dual-stacked network i.e. IPv4 and IPv6, both of which are publicly accessible.

The NCP PerfSonar node has been added on the WLCG/OSG Production PerfSonar Dashboard in order to determine the throughput rate between between NCP and other WLCG sites. The

NCP in WLCG/OSG Production perfSONAR Dashboard

http://psmad.grid.iu.edu/maddash-webui/index.cgi?dashboard=WLCG%20CMS%20Bandwidth%20Mesh%20Config
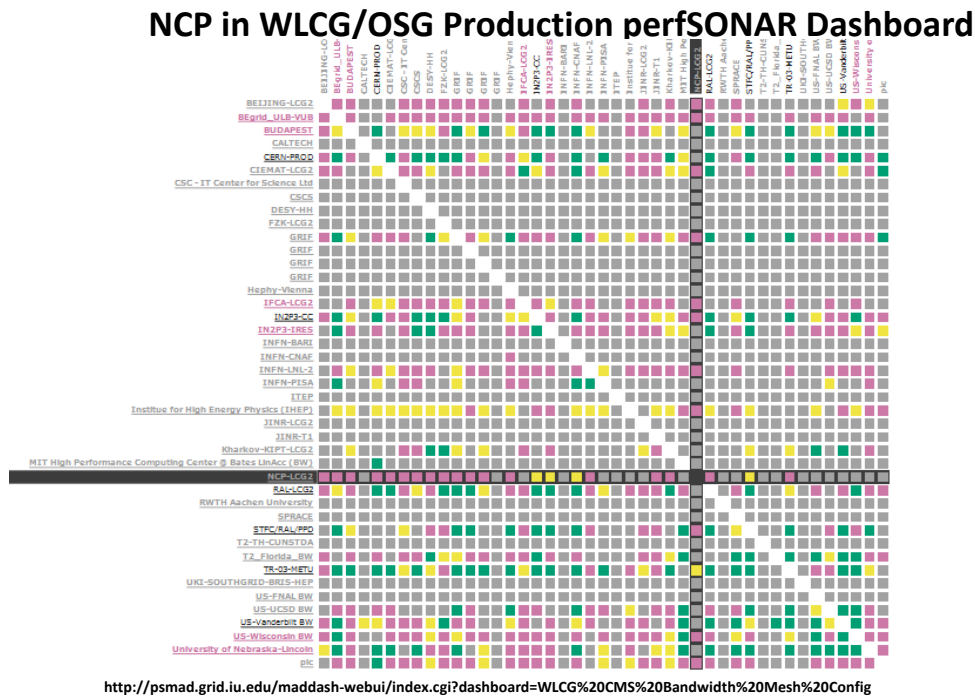
Figure 4: NCP in WLCG/OSG Production perfSONAR Dashboard.

data in Figure 4 shows that the NCP site is getting low throughput compared to most of the other WLCG sites and hence it faces problems in its link commissioning process. The reasons for low throughput are discussed in section 5.

3. Cloud Deployment at NCP

NCP is running Openstack-based private IaaS facility. There are three main projects utilizing this private cloud: the NCP Tier-2 site, the Rocks based high-performance computing (HPC) cluster [18], and the HT-Condor batch system. Almost 70 % of the total resources are reserved for the NCP WLCG Tier-2 site in WLCG infrastructure. Besides this service, 15-20 % of the resources are reserved for the HT-Condor based HTC cluster and the remaining 15-20 % of the compute resources are reserved for the MPI based HPC cluster [14], both of which can scale down or out.

3.1 High Throughout Computing Cluster for EHEP

An HT-Condor based compute cluster has been deployed for experimental high energy physics group. Currently, 96 CPU cores are reserved for the local CMS analysis, which can be scaled up according to workload. The local batch system supports the CMS software environment [8]. The storage capacity of 100 TB has been reserved for this cluster.

3.2 High Performance Computing (HPC) Cluster

The HPC cluster at NCP is based on Linux and supports the use of MPI, and uses Rocks for systems configuration and management. There are a total of 96 CPU cores reserved for this
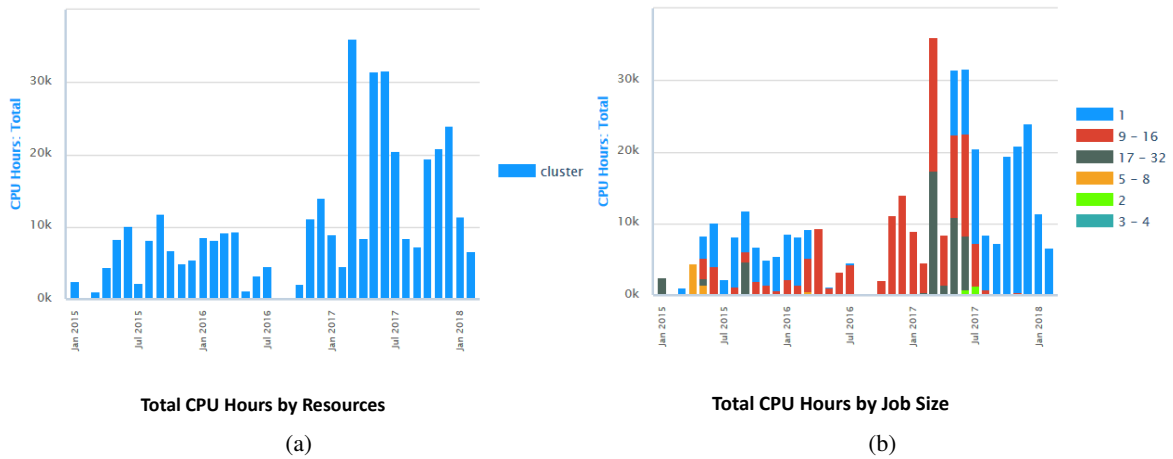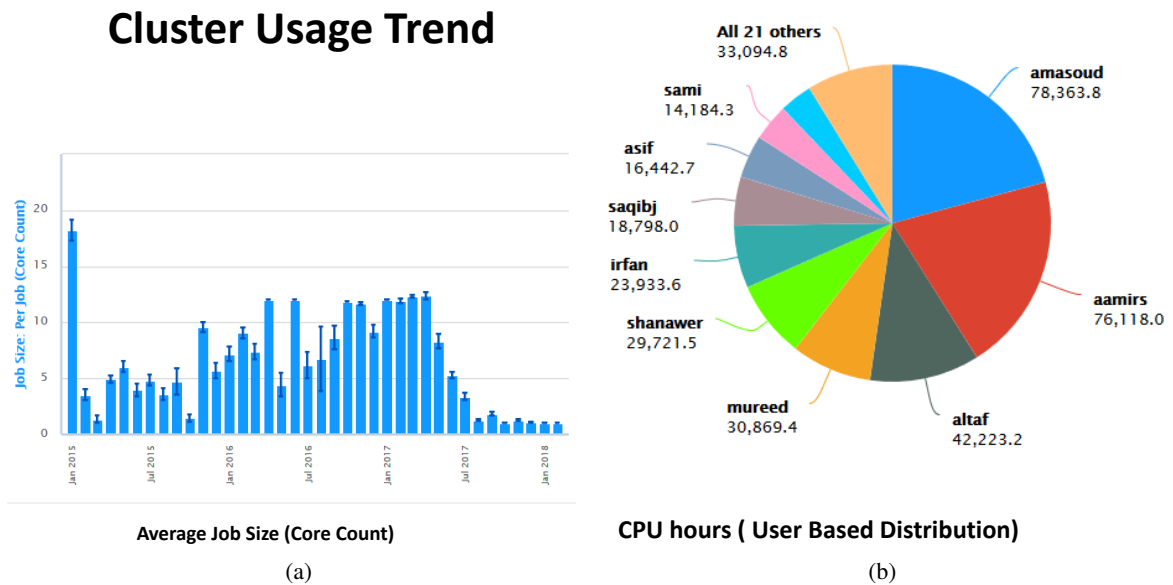
Figure 5: Cluster Usage Trend.

Figure 6: Cluster Usage Trend.

HPC cluster. This cluster is being used by researchers working in different disciplines across Pakistan. The user base of this facility comprises students, researchers and faculty members of different universities and research institutes. Some relevant utilization statistics of the HPC cluster are shown in Figure 5 and Figure 6.

4. Challenges in T2_PK_NCP

There are a couple of problems that we are facing currently. The first problem is related to the commissioning of the dedicated wide-area network link with peer sites in WLCG. NCP's WLCG Tier-2 site is connected with other grid sites through the TEIN network [3] with 1 Gbps shared con-

nectivity. However, currently NCP is facing a problem in its link commissioning, due to low network throughput and other issues between T1 sites and NCP. For example, fiber ring outages, bandwidth bottleneck issue in Pakistan's NREN due to increased load, and maintenance/developmental works. In order to solve this issue, our service provider is implementing a couple of solutions such as bifurcation of the fiber optic ring in metro-area, to increase fault tolerance and acquisition of a new transmission wavelength for increased capacity. Apart from local problems, there are certain non-optimized routing issues between NCP and US/Europe T1 sites, which are being actively addressed.

Another issue is related to accounting and monitoring for the HT-Condor CE. Accounting data of newly deployed HT-condor CE is not being published on EGI accounting portal [12]. The CREAM-CE uses the APEL packages (APEL parser and APEL publisher) for publishing accounting data to the EGI accounting portal [1]. But APEL parser does not officially support HT-Condor. However, some work has been done by a couple of EU sites towards adding that support [2]. This issue will be addressed in future work.

5. Conclusion

In this paper, we presented a brief overview of IT infrastructure and updates on technologies that are being used at NCP. NCP maintains a computing infrastructure for scientific community. A major portion of computing and storage resources are reserved for the CMS experiment by means of the WLCG infrastructure with a smaller portion of the computing resources reserved for other non experimental high-energy physics scientific experiments. For efficient utilization of resources, many scientific organizations have migrated their resources to be part of an IaaS cloud infrastructure. NCP has also taken such an initiative last year, and has migrated most of its resources into a private cloud for scientific use. An HT-Condor based batch system has been deployed for the local EHEP community, to perform their analysis task. Recently we deployed an HT-Condor Compute element (CE) as a gateway for CMS jobs. On the network side, our Tier-2 site is completely accessible and operational on IPv6. Moreover, we recently deployed a Perfsonar node to actively monitor the throughput and latency issues between NCP and other sites. This paper discusses status of NCP Tier-2 site, its current challenges, and future directions.

The existing hardware is becoming obsolete and therefore procurement of new compute and storage servers is in progress at NCP. For example, servers containing the 10th generation Intel Xeon Silver 4116 series of processors will soon be purchased. Furthermore, 50 TB of additional storage will be added to the setup shortly. We are also looking forward to deploy an ELK platform [13] for log management and visualization of data from cloud and grid nodes. Furthermore, Ceph-based storage cluster deployment is in progress which will provide flexible storage services, and will be available soon in the production environment. Other options for compute resource provisioning to the CMS experiment will be evaluated soon such as dynamic on demand VM provisioning using VAC/Vcycle [20], and DoDAS [11].

References

[1] APEL. Website: `https://apel.github.io/`. last accessed: 25.09.2018.

[2] APEL Parser to HT-Condor. Website: `https://github.com/apel/apel/pull/131`. last accessed: 25.09.2018.

[3] Asia - The Trans-Eurasia Information Network (TEIN). Website: `https://ec.europa.eu/europeaid/regions/asia/tein-3_en`. last accessed: 25.09.2018.

[4] Availabilities Reliabilities Report. Website: `http://argo.egi.eu/lavoisier/site_ar?site=NCP-LCG2,%20&cr=1&start_date=2018-01-01&end_date=2018-08-31&granularity=MONTHLY&report=Critical`. last accessed: 25.09.2018.

[5] Ceph storage. Website: `https://ceph.com/`. last accessed: 25.09.2018.

[6] CMS Resource Requirements. Website: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SiteOperationProcedures/`. last accessed: 25.09.2018.

[7] CMS SiteDB. Website: `https://cmsweb.cern.ch/sitedb/prod/sites`. last accessed: 25.09.2018.

[8] CMSSW Application Framework. Website: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework`. last accessed: 25.09.2018.

[9] Computing with HTCondor. Website: `https://research.cs.wisc.edu/htcondor/`. last accessed: 25.09.2018.

[10] DPM - Disk Pool Manager. Website: `http://lcgdm.web.cern.ch/dpm`. last accessed: 25.09.2018.

[11] Dynamic On Demand Analysis Service: DODAS. Website: `https://dodas-ts.github.io/dodas-doc/`. last accessed: 25.09.2018.

[12] EGI Accounting Portal. Website: `https://accounting.egi.eu/`. last accessed: 25.09.2018.

[13] Elastic Search, Logstash, kibana. Website: `https://www.elastic.co/elk-stack`. last accessed: 25.09.2018.

[14] Message Passing Interface (MPI). Website: `https://hpcc.usc.edu/support/documentation/message-passing-interface/`. last accessed: 25.09.2018.

[15] National centre for physics. Website: `http://www.ncp.edu.pk`. last accessed: 05.04.2018.

[16] Perfsonar. Website: `https://www.perfsonar.net/`. last accessed: 25.09.2018.

[17] Resource Utilization in EGI Portal. Website: `http://accounting-devel.egi.eu/egi.php?ExecutingSite=NCP-LCG2&query=normcpu-HEPSPEC06&startYear=2017&startMonth=12&endYear=2017&endMonth=12&yrange=SITE&xrange=VO&groupVO=lhc&chart=GRBAR&scale=LIN&localJobs=onlygridjobs`. last accessed: 25.09.2018.

[18] Rocks Cluster. Website: `http://www.rocksclusters.org/`. last accessed: 25.09.2018.

[19] Tier 2 Resources Requirement for CMS sites. Website: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SiteOperationProcedures`. last accessed: 25.09.2018.

[20] Vcycle: VM lifecycle management. Website: `https://www.gridpp.ac.uk/vcycle/`. last accessed: 25.09.2018.

[21] Worldwide LHC Computing Grid. Website: `http://wlcg.web.cern.ch`. last accessed: 25.09.2018.

PoS(ISGC 2018 & FCDD)015