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

Dr. Muhammad Imran

National Centre for Physics, Islamabad, Pakistan

Agenda

- Overview of National Centre for Physics (NCP), and IT services

- Overview of WLCG T2_PK_NCP Site.

- Issues and Challenges in WLCG site.

- Overview of Local computing cluster.

- Future Directions

About NCP

- The National Centre for Physics, Pakistan has been established to promote research in Physics & applied disciplines in the country and the region.

- NCP has collaboration with many in international organizations including **CERN**, **SESAME**, **ICTP**, **TWAS**…

- Major Research Programs:

  - Experimental High Energy Physics, Theoretical and Plasma Physics, Nano Sciences and Catalysis, Laser Physics, vacuum Sciences and technology, Earthquake studies.

NCP IT Overview

NCP is maintaining a large IT infrastructure, which is mainly categorized into following areas:

CORE COMPUTING SERVICES

- Hosted a WLCG TIER-2 Site, comprising of **524 CPU Cores**, and **~540 TB of Disk Storage**

- Computing cluster of **96 CPU cores**, installed for local scientific community.

CORPORATE IT SERVICES

- Corporate Services, including Email, DNS, Public Web site, FTP , application databases etc... are hosted inside NCP data centre.

- All Of the Corporate Services are **Virtualized. ... 50 + VMs**

High Speed & fault tolerant network infrastructure is deployed to provide IT services

WLCG @ NCP

NCP-LCG2 site hosted at National Centre for Physics (NCP) , for WLCG CMS experiment.

- **Total Number of Physical CPUs = 106**

- **Total Number of Logical Cores = 524**

- **HEPSPEC06 = 6365**

- **KSI2K = 1591**

- **Storage Capacity = 561 TB (Raw)**
- **= 430 TB (Usable)**

NCP-LCG2 and TIER-2 Site Requirement

- Requirement of Tier-2 site Resources With respect to CMS Experiment

Resources	Nominal	Recommended	Tolerated	Installed@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 Gpbs	1 Gpbs

Hardware Specification

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
				524 CPUs

Storage Servers

Storage Server	Total Disks/Server	Servers Quantity	Raw Capacity
Transtec Lynx 4300	23 x 1TB	15	345 TB
Dell Power Edge T620 DELL EMC MD1200	12 X 4TB 10 x 6 TB	2 2	96TB 120 TB
			561 TB Total

Status of NCP-LCG2 site

Installed Resources from 2008-2018

Year	CPU	HEPSPEC06	Storage	Network Connectivity
Jan-08	14	67.2	3.2 TB	2 Mbps (Shared)
April-08	36	172.8	3.2 TB	2 Mbps (Shared)
Sep-08	74	355.2	3.2 TB	10 Mbps (dedicated)
Feb-10	160	1600	3.2 TB	10 Mbps (dedicated)
Jun-10	240	2400	69 TB	**155 Mbps (dedicated)**
Dec-10	524	**6365**	87TB	155 Mbps (dedicated)
Jun-11	524	6365	175TB	155 Mbps (dedicated)
May-12	524	6365	260TB	155 Mbps (dedicated)
Oct-14	524	6365	**330TB**	155 Mbps (dedicated)
April-2015----	524	6365	330TB	**1 Gbps (connectivity)**
Upto Mar 2018	524	6365	561 TB	**1 Gbps (connectivity)**

WLCG @ NCP

- Compute Elements (CEs)

  - 2 x CREAM-CE Hosts

    - pcncp04.ncp.edu.pk and pcncp05.ncp.edu.pk
    - Equipped with compute nodes cluster of ~ 360 cores
    - PBS batch system deployed

  - HTCONDOR-CE *(Latest Deployment*)

    - htcondor-ce.ncp.edu.pk
    - Equipped with compute nodes cluster of ~ 164 cores.
    - HT-condor batch system deployed on backend.
    - Connected on Dual stack IPv4/ipv6

- Storage Elements (CEs)

  - Disk Pool Manager (DPM) based storage is being used at NCP.
  - 12 DPM disk Nodes are aggregated to provide ~ 430 TB of storage capacity.
  - Connected on Dual stack IPv4/ipv6
  - Additional 100 TB from Ceph Cluster will be integrated soon…..

HT-condor CE deployment

- HT-Condor-CE and its all worker nodes are recently deployed on cloud

- Worker Nodes: 4-Core flavor VMs with 8GB RAM partitioned to 4 job slots.

- Batch System Resources: 164 cores, 328GB memory, ~20TB storage, 164 job slots.

- One-click new running(installed/configured) worker node (VM) spinning .

- In production at NCP, for about three months only.

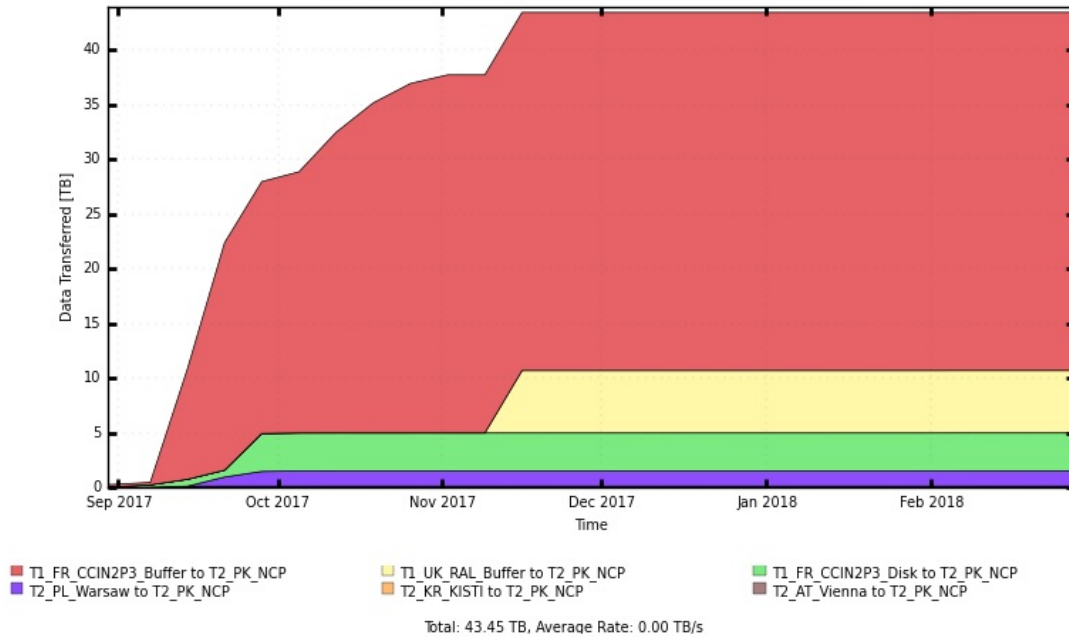- Jobs run: 8104 + remote.

Resource Utilization (2011-2018)

Year	KSI2K-Hours	No. of Jobs
2011	2,816,009	614,215
2012	3,791,319	630,629
2013	427,846	308,065
2014	609,034	165,002
2015	1,800,557	239,315
2016	1,474,279	347,339
2017	3,945,116	408,688

Jobs stats/ CPU Utilization 2017

CPU Utilization in (KHEPSPEC06)

Number of Jobs
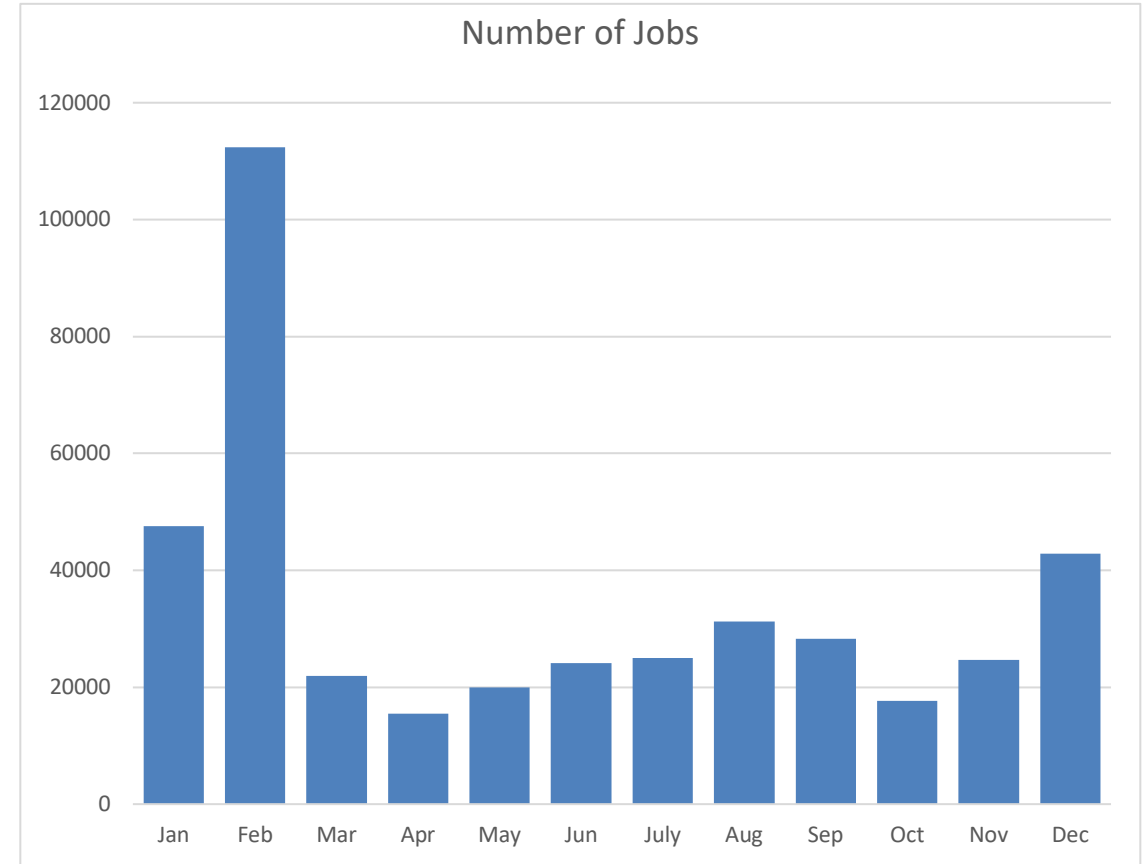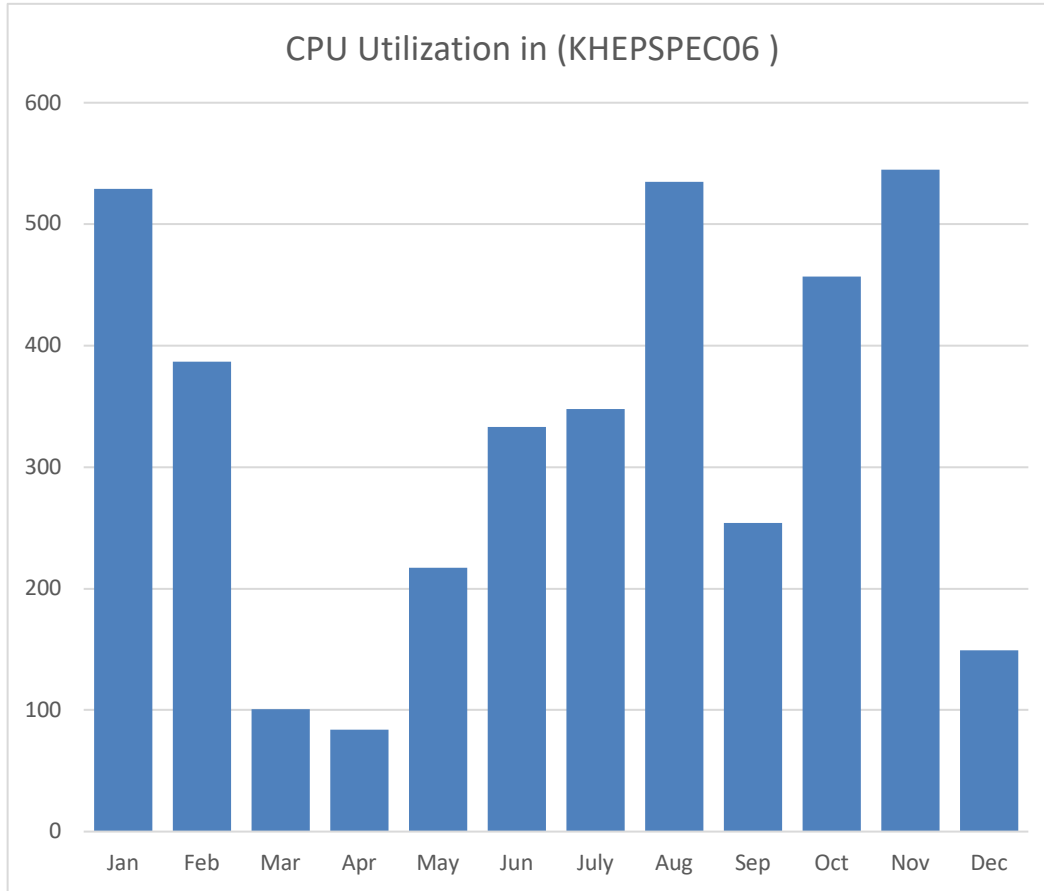
NCP Node Data Transfer stats (Last 6 months)

Download (226.85 TB)
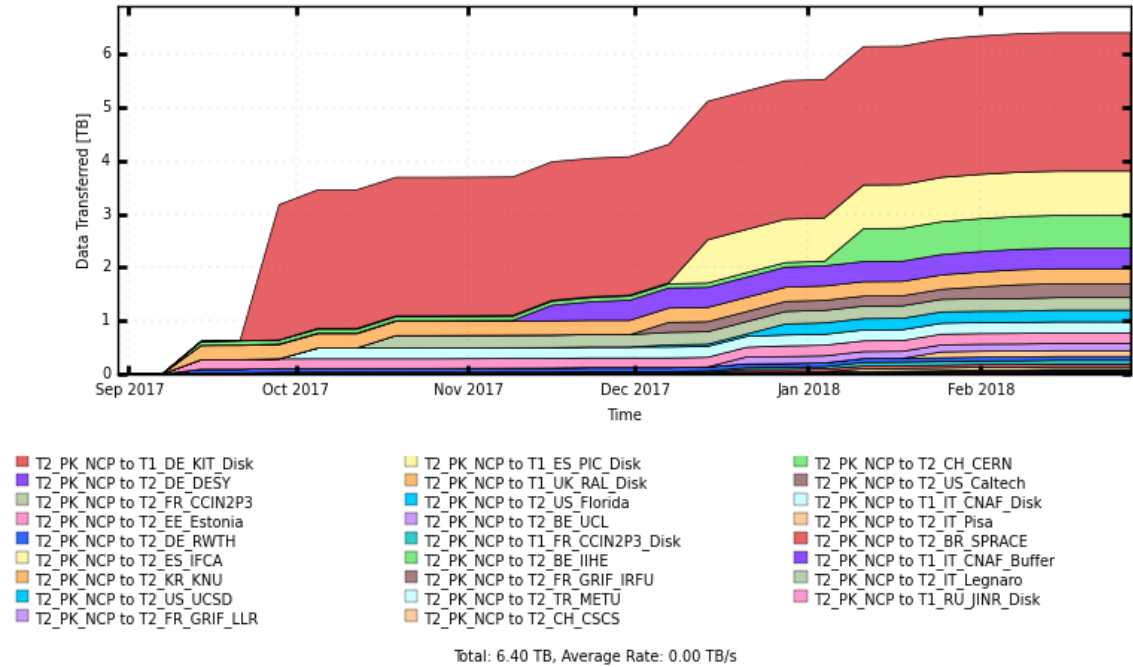
Upload (164.89 TB)

NCP_LCG2 Site Availability and Reliability

AsiaPacific	2018-02		2018-01		2017-12	
	Availability	Reliability	Availability	Reliability	Availability	Reliability
Australia-ATLAS (Australia)	99.9	99.9	100	100	100	100
HK-HKU-CC-01 (Hong Kong)	100	100	99.97	99.97	99.88	99.88
IN-DAE-VECC-02 (India)	99.75	99.75	99.74	99.74	99.21	99.21
INDIACMS-TIFR (India)	99.93	99.93	96.89	99.3	65.6	65.6
IR-IPM-HEP (Iran)	99.94	99.94	99.7	99.7	99.27	99.27
JP-KEK-CRC-02 (Japan)	100	100	96.76	100	99.99	99.99
KR-KISTI-GSDC-01 (South Korea)	100	100	100	100	99.93	99.93
KR-KISTI-GSDC-02 (South Korea)	99.73	99.73	76.28	78.82	96.73	98.33
KR-KNU-T3 (South Korea)	100	100	99.81	99.96	98.69	98.69
LCG-KNU (South Korea)	100	100	99.58	99.94	98.63	98.63
NCP-LCG2 (Pakistan)	99.18	99.18	97.81	97.81	99.69	99.69
PK-CIIT (Pakistan)	0	0	0	0	0	0
T2-TH-SUT (Thailand)	92.18	92.18	98.4	98.4	87.59	87.59
TOKYO-LCG2 (Japan)	99.37	99.37	95.24	96.12	100	100
TW-FTT (Taiwan)	100	100	99.98	99.98	96.44	100
TW-NCHC (Taiwan)	65.09	65.09	0	0	0	0
TW-NCUHEP (Taiwan)	100	100	90.73	90.73	53.03	53.03
TW-NTU-HEP (Taiwan)	77.85	77.85	0	0	0	0
Taiwan-LCG2 (Taiwan)	77.77	77.77	65.94	65.94	90.63	93.66
BEIJING-LCG2 (China)	98.35	100	95.32	99.78	96.36	99.62
HK-LCG2 (China)	97.88	99.97	100	100	99.55	99.55

PerfSonar Node @NCP

- **NCP has recently deployed a Perfsonar node at NCP, to troubleshoot network throughput issues between NCP and WLCG T1/T2 sites.**

  - **http://ps.ncp.edu.pk**
  - **Configured on dual stack i.e. IPv6/IPv4.**

- **Registered in GOCDB.**
- **Visible in WLCG/OSG perfsonar dashboard**
 CMS bandwidth MaDdash

  - **Few PERN sites are also Added monitoring dashboard, for network problem troubleshooting.**

  - **Deployment of this service, facilitated a lot in identification of low network throughput, reasons, and bottlenecked links.**

NCP in WLCG/OSG Production perfSONAR Dashboard
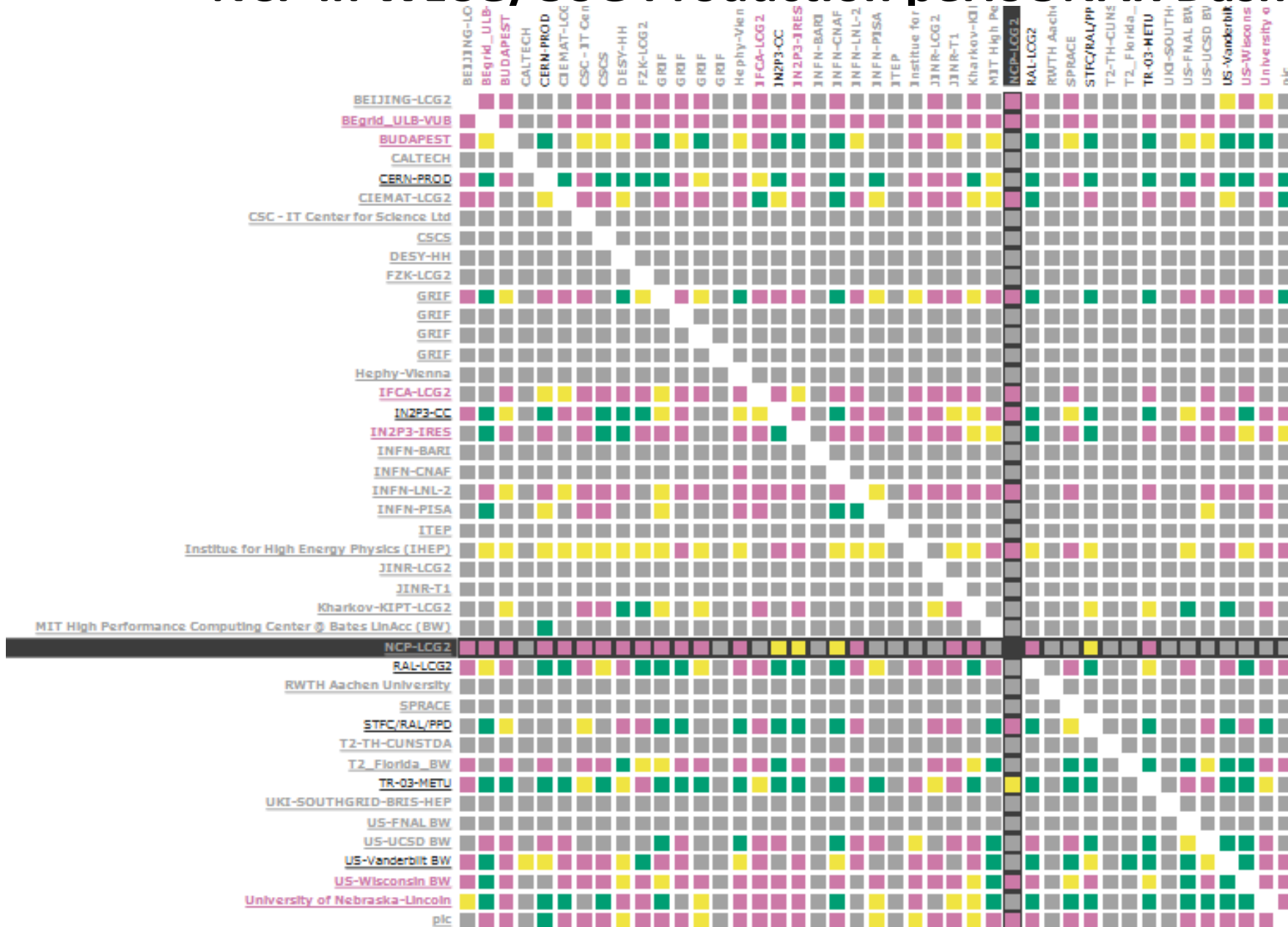
Cloud Deployment @ NCP

- NCP is running a Openstack based Private Cloud, and all of the hardware resources are accumulated in this cloud.

- Three Projects are running under this private cloud.
  - **WLCG Project**
    - » Almost 70 % of the total resources are reserved for NCP WLCG T2 site.
  - **Local Batch system (HTC)**
    - » 15-20 % of Resources are reserved for HT-condor based HTC cluster.
  - **Local HPC Cluster**
    - » 15-20 % of compute Resources are reserved for MPI based HPC cluster, which can scale in and out.

High Throughout Computing Cluster for EHEP

- HT-condor based compute cluster has been deployed for Experimental High Energy Physics Group.

- Currently 96 CPU Cores, are reserved for local CMS analysis, which can be scaled up according to workload.

- Local batch system fully supports CMSSW environment, and all necessary packages.

- 100 TB of storage is reserved for data storage.

High Performance Computing (HPC) Cluster

- Local Linux/MPI cluster based on **Rocks Cluster Distribution**.

- 96 CPU cores are reserved for non grid computing.

- 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.

Non GRID HPC Research Areas

Some of the scientific areas where researchers are benefitting from this facility are as follows:

- Computation Fluid Dynamics (CFD)
- Molecular Biology
- Bio Chemistry
- Condensed Matter Physics
- Space Physics
- Weather Forecasting
- Density Functional Theory (DFT)
- Ion channeling
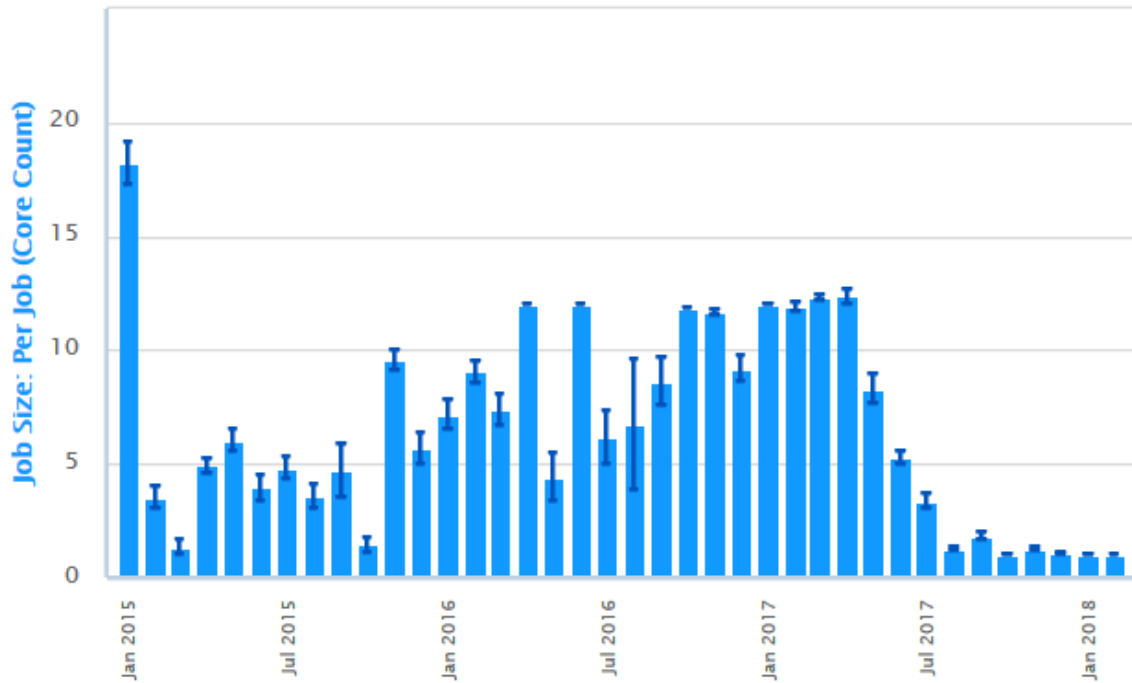- Multi-Particle Interaction
- Earthquake studies.

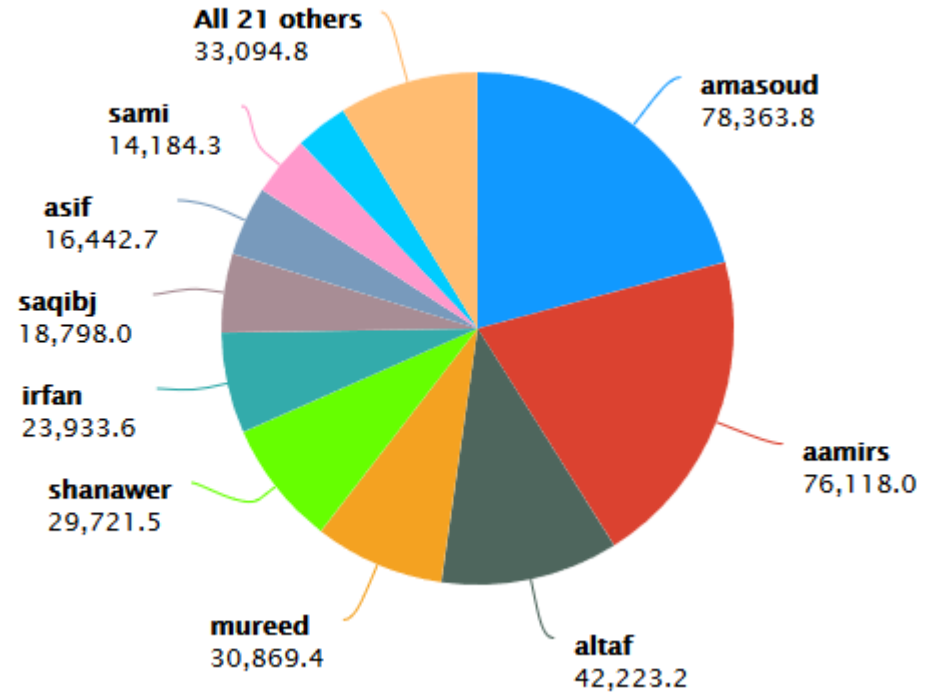# Cluster Usage Trend
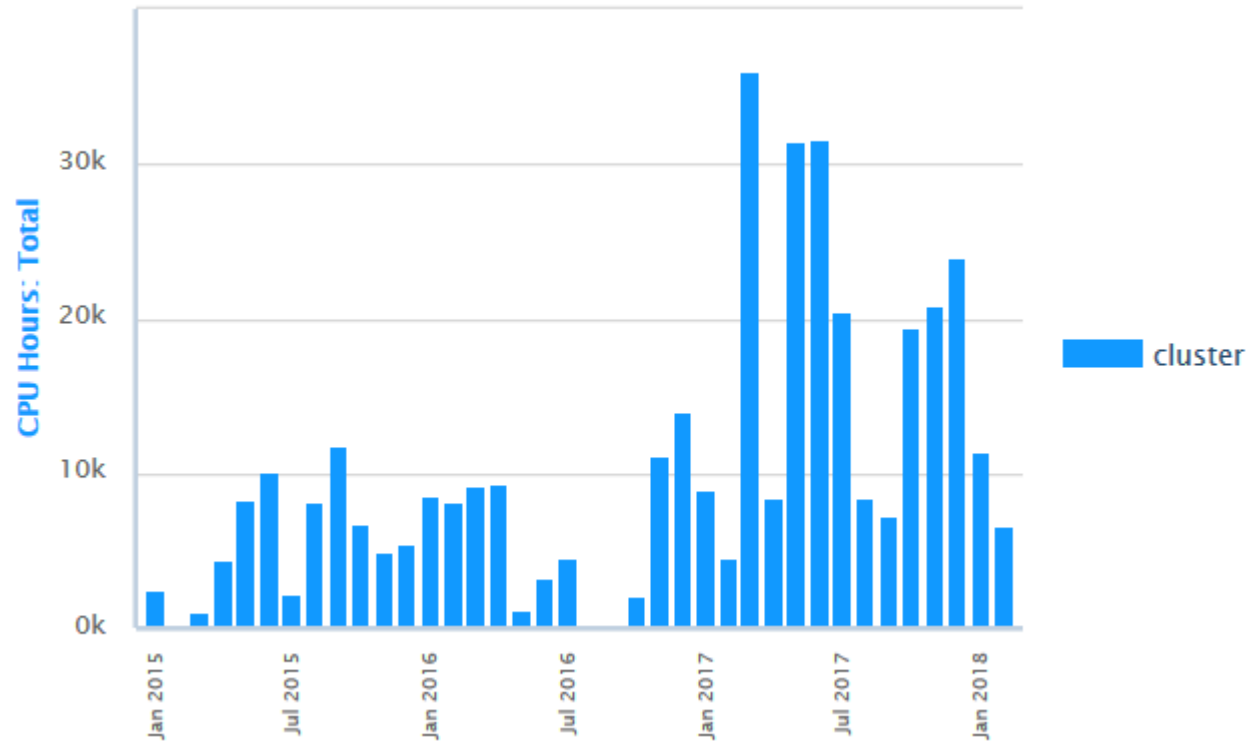
Total CPU Hours by Resources
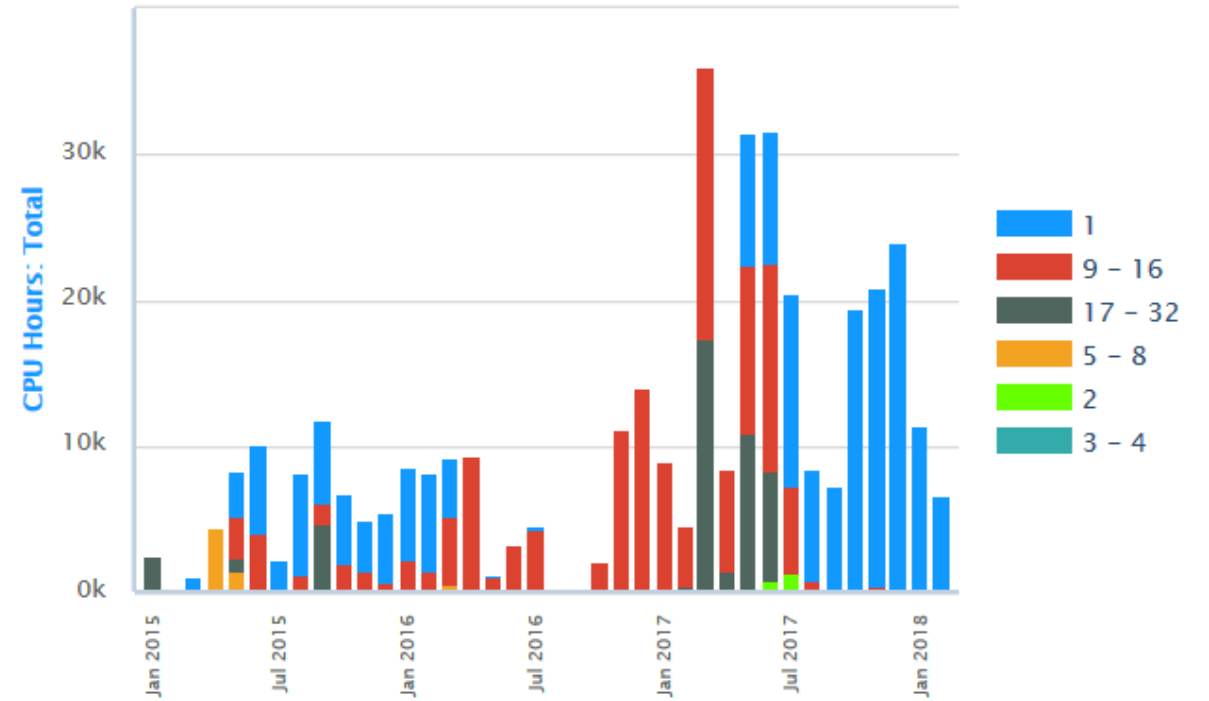
Total CPU Hours by Job Size

Cluster Usage Trend

Average Job Size (Core Count)

CPU hours (User Based Distribution)

Challenges in T2_PK_NCP

- **WLCG link Commissioning problems:**

  - NCP is connected with European and US sites through TEIN network with 1 Gbps of connectivity. However, Currently NCP is facing a problem in its Link commissioning, due to low network throughput and other issues between T1 sites and NCP.

  Like:

  - Fiber ring outages, bandwidth bottleneck Issue in Pakistan's NREN due to increased load, and maintenance/developmental works.
    - Bifurcation of fiber optic ring, to increase fault tolerance…..in progress.
    - Acquisition of new transmission Lambda, for increased capacity….in progress.
  - Downlink from US T1 (FNAL) to NCP , is not following R&D TEIN route…. Hence creating bandwidth bottleneck.
    - Issue is already forwarded to network team of fermilab, and being addressed.
  - Certain European T1 sites (like GRIDKA & PIC…) are showing low throughput, with NCP.
    - Issues already raised to wlcg-network-throughput@cern.ch mailing list, and being addresses by relevant persons.

Challenges in T2_PK_NCP

- **Accounting and Monitoring (HTCONDOR-CE)**
  - Accounting data of New Deployed HT-condor CE, is not being published on EGI accounting portal.

  - CREAM-CE uses APEL packages for publishing accounting data on EGI accounting portal. But APEL does not yet officially support HT Condor.

  - Currently there is no tool developed to show nice graphical view of jobs detail like job owners, CPU cycles, time, other resources utilizations.

Future Direction

- Existing hardware is becoming obsolete and therefore Procurement new of **Compute Servers** and **Storage** is in progress….

  - 10th Generation latest Intel Xeon Silver 4116 series of processors.
  - 50 TB of additional storage is being procured.

- **ELK platform** deployment is in progress for log management and visualization of data from cloud and grid nodes.

- **Ceph based** storage cluster deployment in progress due its flexibility, and will be available soon in production environment….

- Other Options for Compute resource provisioning to CMS experiment will be evaluated soon like Dynamic on Demand VM provisioning, VAC/Vcycle, and DoDAS
  - https://www.gridpp.ac.uk/vcycle/
  - https://twiki.cern.ch/twiki/bin/view/CMSPublic/DynamicOnDemandAS

Thankyou For Your Attention