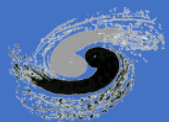# IHEP Site Report

Haibo Li (lihaibo@ihep.ac.cn)
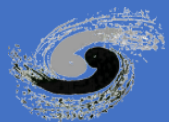
On behalf of Computing Center, IHEP

HEPiX Autumn 2020

Outline

- Brief Introduction
- Operating Status
  - Local Cluster
  - Grid Site
  - Network
- Activities in progress
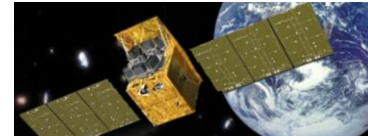 - HTCondor and Slurm Cluster
  - Storage
- Summary

Brief Introduction to IHEP
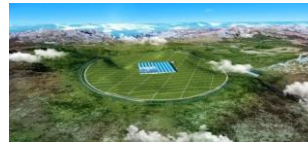

BESIII (Beijing Spectrometer III at BECPII)

JUNO (Jiangmeng Underground Neutrino Observatory)
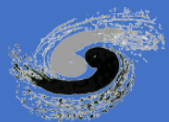

HXMT(Hard X-Ray Moderate Telescope)

China Spallation Neutron Source

LHAASO (Large High Altitude Air Shower Observatory)


HEPS (High Energy Photon Source)

Computing Resources

- 25,000 cpu cores, 180 GPU cards for more than 10 experiments
    - HTCondor cluster runs for HTC jobs
    - Slurm cluster runs for HPC jobs
    - WLCG tier 2 site

- About 30 PB storage
    - Lustre and EOS are two main file systems
    - Castor for tape storage

- Network
    - IP V4/ IP V6 dual stack
    - Ethernet(100Gb) / IB (100Gb) supported
    - LHCOne joint
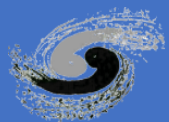
Outline

- Brief Introduction

- Operating Status
    - Local Cluster
    - Grid Site
    - Network

- Activities in progress
    - HTCondor and Slurm Cluster
    - Storage

- Summary
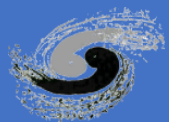
Updates to Infrastructure

- Foreman upgraded from 1.24 to 2.1

- Puppetserver and puppetca upgraded from puppet5 to puppet6
  - Puppet6 takes 5 times longer to compile directories than puppet5.
  - tuning:
    - Need more heap size than puppet5,currently set to 1.25G/instance heap size
   - Set 100M/instance codecache size

- Tuning Guide:
  - https://tickets.puppetlabs.com/browse/SERVER-2771

Puppetserver cpu usage last month
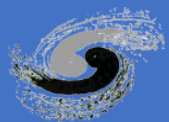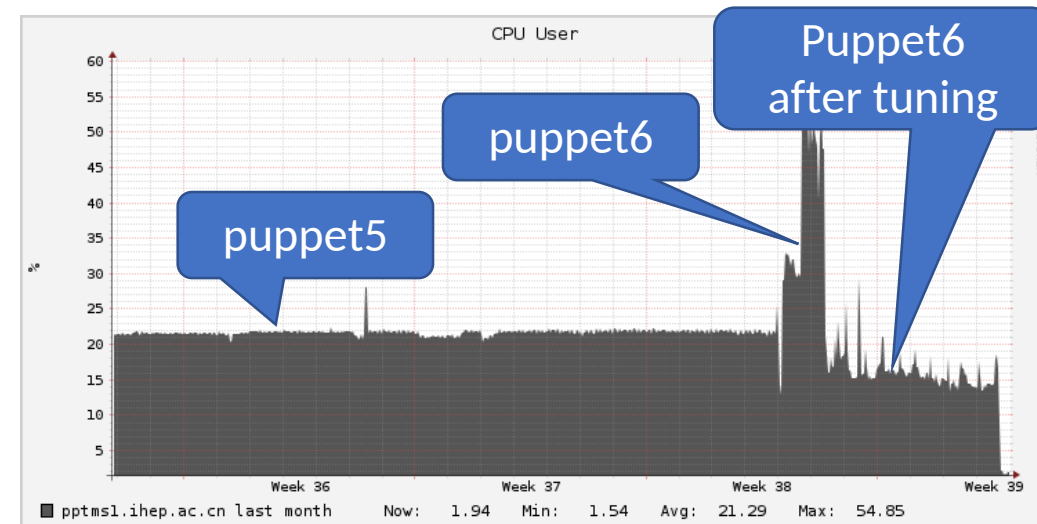
Centos 7 upgrade

- Upgrade all node to Centos7 in August

- Update all puppet module related to centos7

- Issues with centos7:
  - Nic card sequence changed in every reboot
    - Set nic name in file /etc/udev/rules.d/60-net.rules
    - Set kernel parameter net.ifnames=1
    - Here is the puppet module(fixdnic):
    - Use Hiera to configure dual nic bonding

```
class clusters::fixednic {
  file {"/etc/udev/rules.d/60-net.rules":
    mode     =>  '0644',
    owner    =>  'root',
    group    =>  'root',
    content  =>   template("${module_name}/60-net.rules.erb"),
  }
  kernel_parameter { "net.ifnames":
    ensure  => present,
    value   => "1",
  }
  network::interface { 'eth0':
    enable_dhcp => true,
    hwaddr      => $mac,
    nm_name     => "eth0",
    defroute    => "yes",
  }
}
```
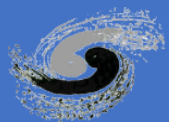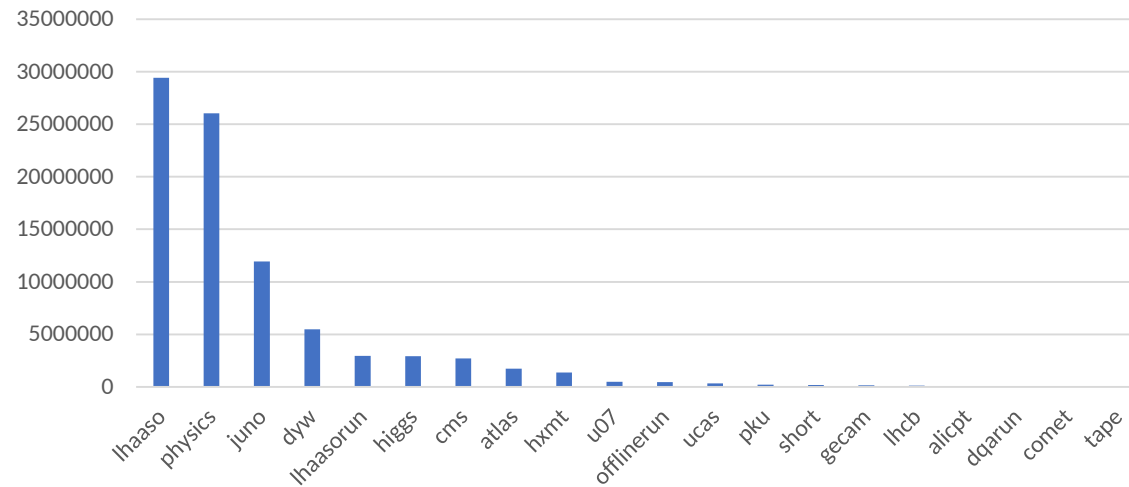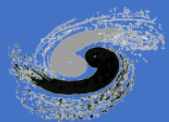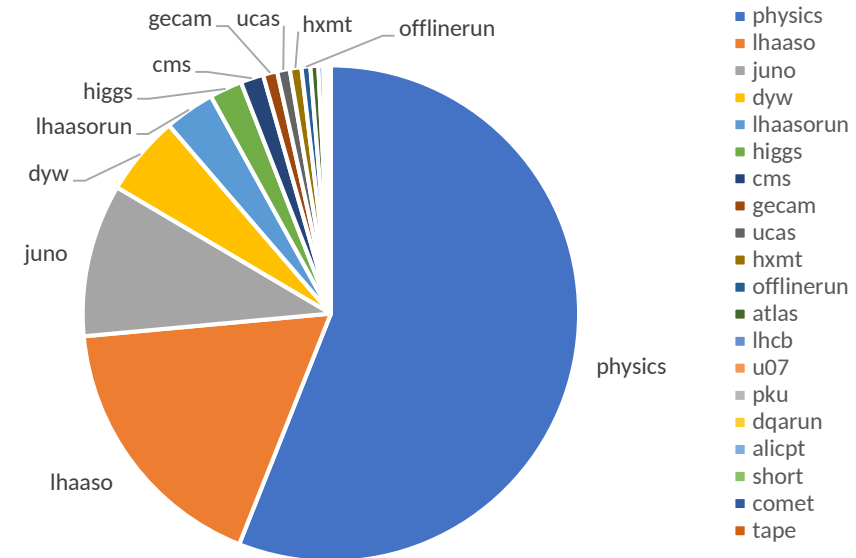
HTCondor Cluster Status

- Statistics (2019.10~2020.09)
  - Total job number: 86.5 millions
  - Total Walltime in hours: 141.8 millions

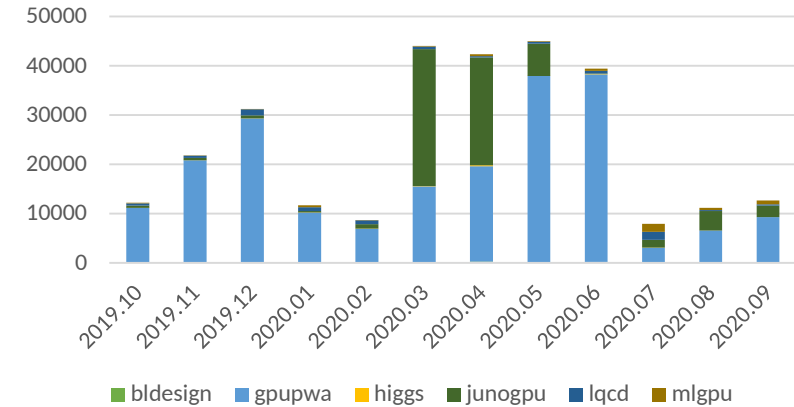Job Number by Group

Walltime Proportion by Group

Slurm Cluster

- **Cluster upgrade**
  - 2020.03 : 88 NVIDIA V100 GPU cards added
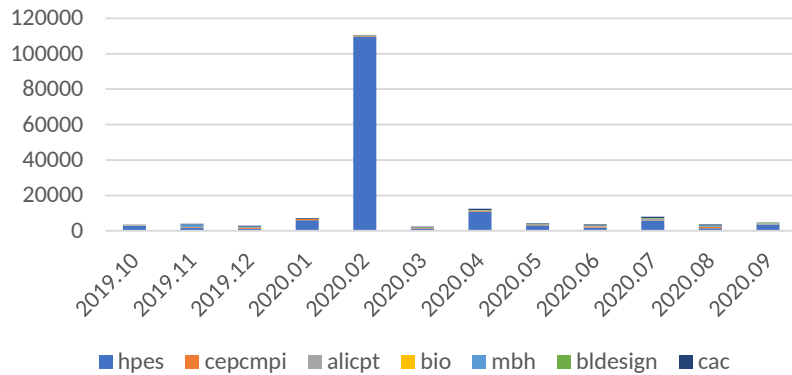 - 2020.08: Slurm upgraded to 19.05.6
- **Cluster jobs(2019.10~2020.09)**
  - CPU jobs: 167K, 23M CPU*Hours
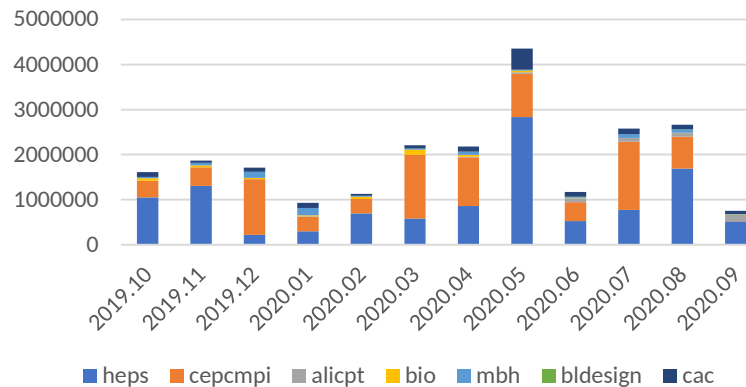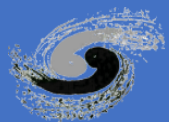 - GPU jobs: 288K, 591K GPU*Hours

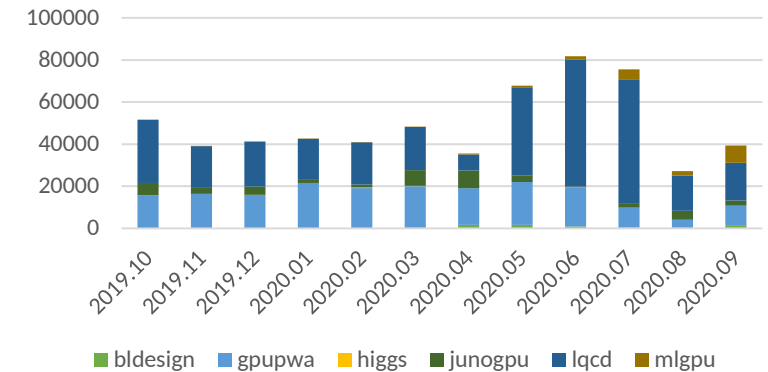Number of GPU Jobs

Number of CPU Jobs

CPU * Hours of Jobs

GPU * Hours of Jobs

Storage Status
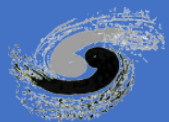
- **Space capacity**
  - Lustre :20 PB total, 9 PB used, 6 PB will be added soon
  - EOS: 8 PB total, 6.5 PB used, 5 PB will be added soon
- **Performance -- Aggregate bandwidth**
  - Read :42.35 GB/s peak, 14.3 GB/s average
  - Write :5.56 GB/s peak, 1 GB/s average
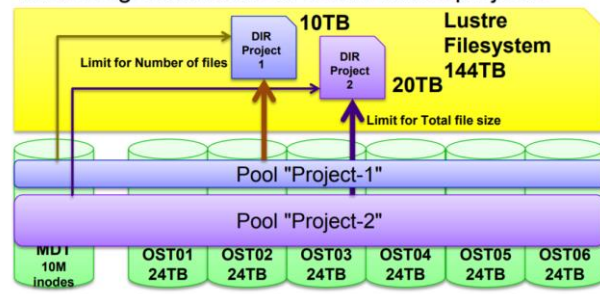- **Availability Time**
  - >99%

Lustre updates

- All the server and clients upgraded to Lustre 2.12.5, the lastest LTS version
  - More stable, better support with recent Linux kernel and hardware
  - Enabled a new feature as project quota, statistics of /besfs/groups will be faster
- With project quota function, system administrator can set a shared project id for a set of directories which is related to a specific project
  - Lustre records space and inode usages during file I/O
  - Project id will be inherited when a sub directory is created
  - By command chattr , directories can be added and removed from a project freely
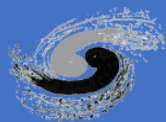 - With new directories add/remove, project quota will change simultaneously

EOS updates

- All the server and clients upgraded to EOS 4.7.7 and QuarkDB 0.4.2
  - Boot time reduced from 5 minutes to seconds

- Use new architecture and layout
  - Using 25gb Network and JBOD array (5u 84 disks)
  - 8PB raw capacity, launched in November 2019
  - Configure with two replicas


2019.07

2019.12

Upgrades of Tape Storage

- Tape Upgraded from LTO 4 to LTO 7
  - >11PB data is stored on tapes
  - Raw data is stored in two copies on tapes
  - To support the annual PB growth of data storage requirements
- Features and Benefits of LTO 7 Tape
  - **Higher Storage Capacity** - 6TB native and up to 15TB Compressed
  - **Better Performance** - Expect data transfer rates of up to 700 MB/sec

TAPE	Native Capacity	Compressed capacity	Native Data Rate	Compressed Date Rate
LTO-4	800GB	1.6TB	120MB/s	240MB/s
LTO-7	6TB	15TB	300MB/s	750MB/s
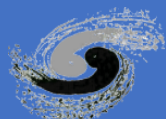
BEIJING-LCG2 Tier2 Resources

- CPU: 3048 cores
  - Intel Golden 6140 2160 Cores
  - Intel E5-2680V3: 696 Cores
  - Intel X5650   192 Cores
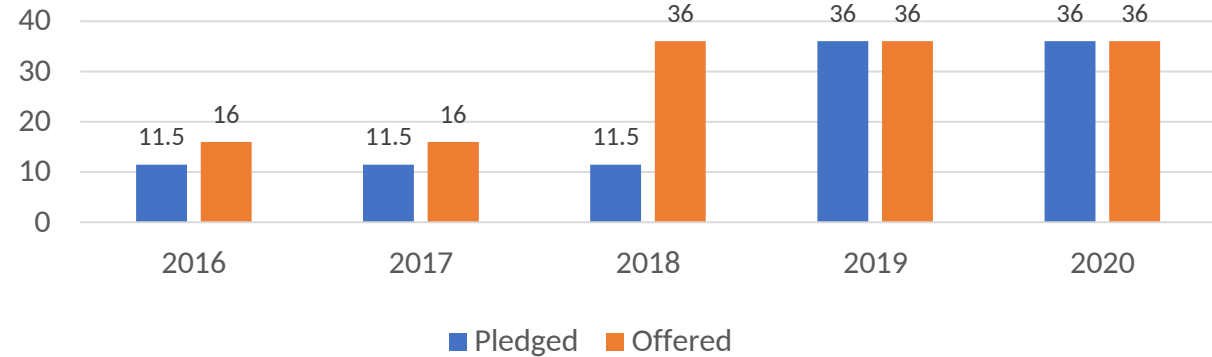- Batch: HT-condor
- VO: ATLAS, CMS, LHCb, BelleII, CEPC

- DPM: 1055TB
  - 4TB * 24slots with Raid 6, 5 Array boxes
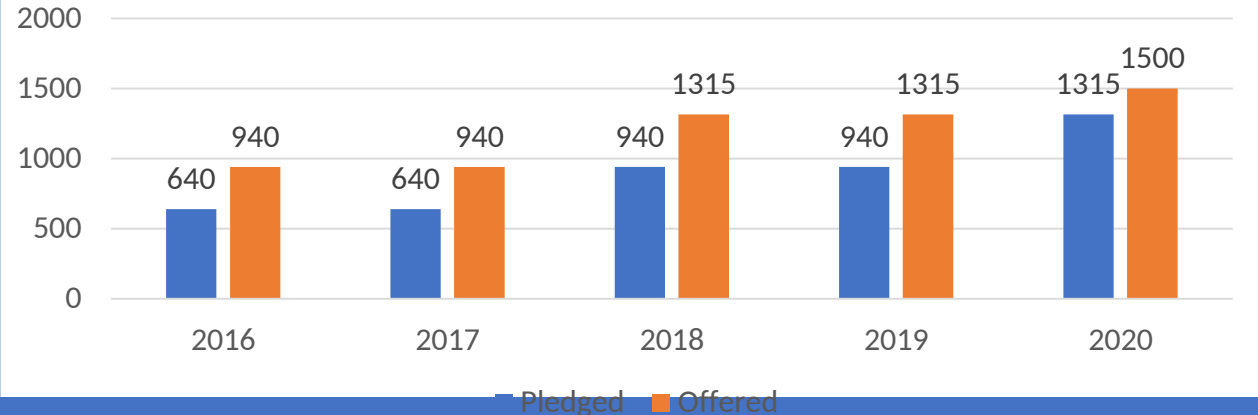 - DELL MD3860 8TB*60 slots
  - DELL ME4084 10TB*42 slots
- dCache: 540TB
  - 4TB * 24slots with Raid 6, 6 Array boxes
  - 3TB * 24slots with Raid 6, 2 Array boxes

CPU(HEP-SPEC06X1000)

Year	Pledged	Offered
2016	11.5	16
2017	11.5	16
2018	11.5	36
2019	36	36
2020	36	36

DISK(Tbytes)

| Year	Pledged	Offered
2016	640	940
2017	640	940
2018	940	1315
2019	940	1315
2020	1315	1500

BEIJING-LCG2 Tier2 Operations

- DPM Upgraded from 1.12 to 1.13, will be upgraded to 1.14 in October

- Moving to HTCondor-CE&HTCondor
  - Replace Cream-CE by HTCondor-CE, PBS by HTCondor
  - Provide a unified computing endpoint for ATLAS, CMS, LHCb, Belle
  - Allocate resources to the proper VO by accounting_group
  - Current status
    - Already in production for ATLAS and CMS
    - Under testing for LHCb and Belle
    - Serving for JUNO and CEPC which are not WLCG sites
 - Considering high availability, we're planning to add an extra CE
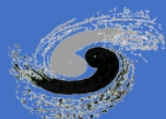
Network architecture updates

- Interconnection isolation between IHEP network and remote site network
 - Layer 2 interconnection -> Layer 3 interconnection
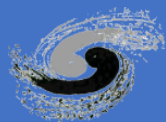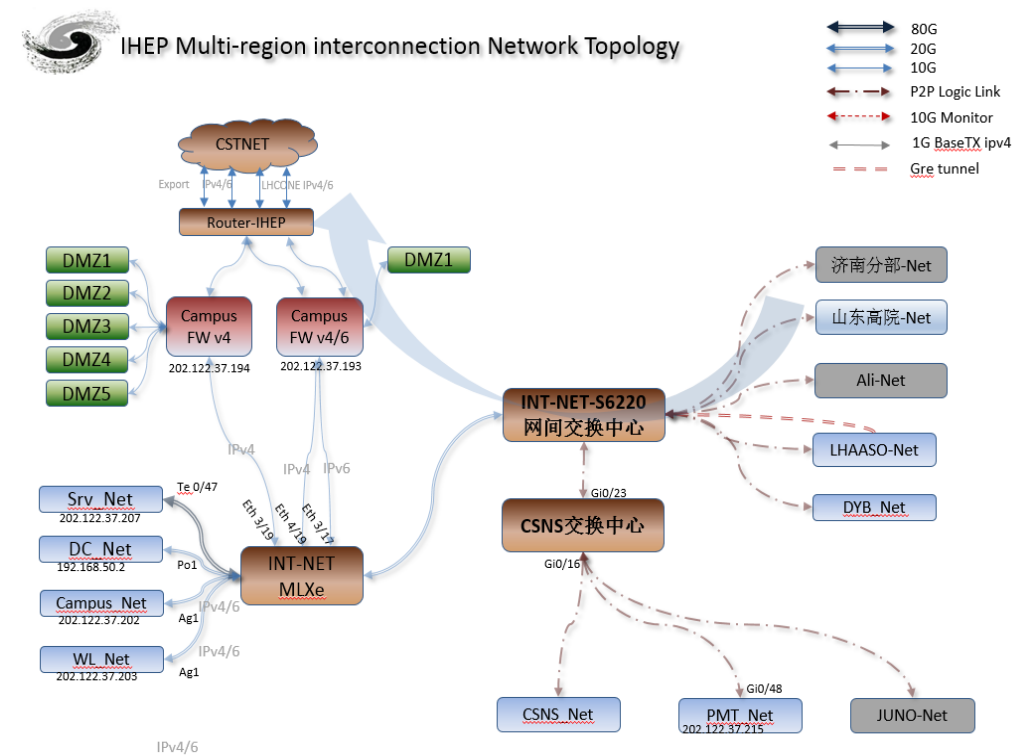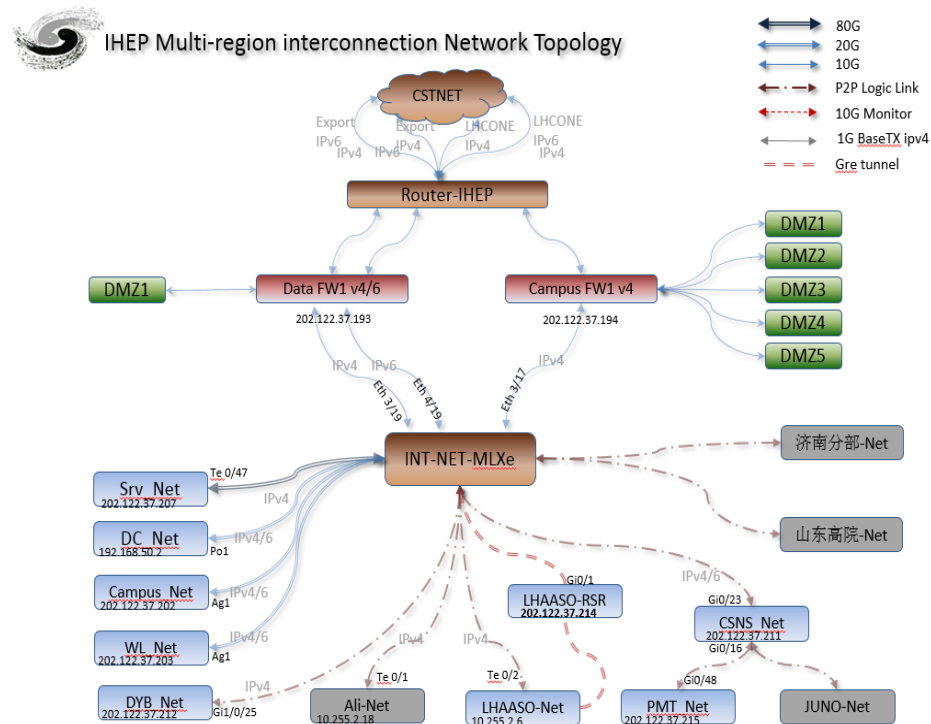
Cyber Security

- **Updates in the annual summer maintenance**
  - The firewall configuration was adjusted according to the core network topology updates
  - Starting daily black IP/URLs blocking for malicious out connection
  - Starting the daily email notice for users whose PC are infected by virus

- **Security Operation Center**
  - Three MISP (Malware Information Sharing Platform) instances for BESIII/DYB/CSNS was established.
  - Bro and web logs are send to Kafka and then to ES
 - 6 distributed data collectors of NGSOC are deployed
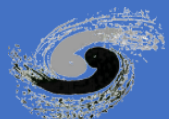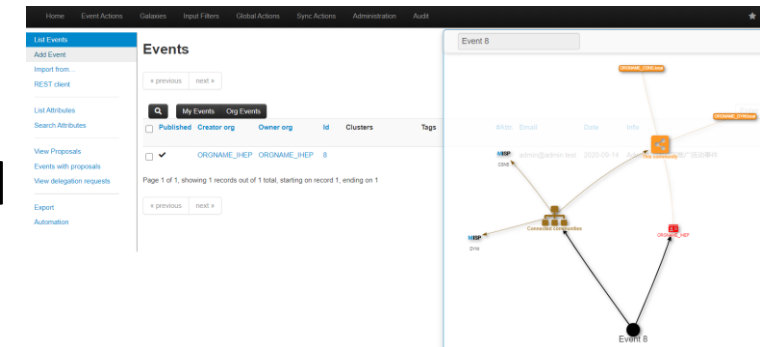
Outline

- Brief Introduction

- Operating Status
  - Local Cluster
  - Grid Site
  - Network

- Activities in progress
  - HTCondor and Slurm Cluster
  - Storage

- Summary
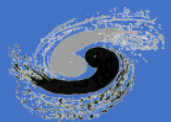
Activities in progress– HTCondor cluster

- **Multiple Schedds**

- **One schedd performs slow**
  - The new experiment LHAASO brought 100,000 jobs into Schedd on some weekends

Pressure of Single Schedd (this July)

- **Binding the specific groups to each Schedd**
  - Integrated multiple schedds via HepJob (a frontend tool of schedulers)
  - A user's operation will be mapping to the corresponding Schedd as the user group, before running hep_sub, hep_q, …

- **Other updates:**
  - Dynamic job walltime control: test, short, default, mid, long
 - Multiple Singularity image (SL5, SL6, SL7), decided by the parameter '-os' of hep_sub
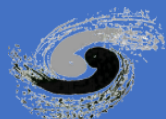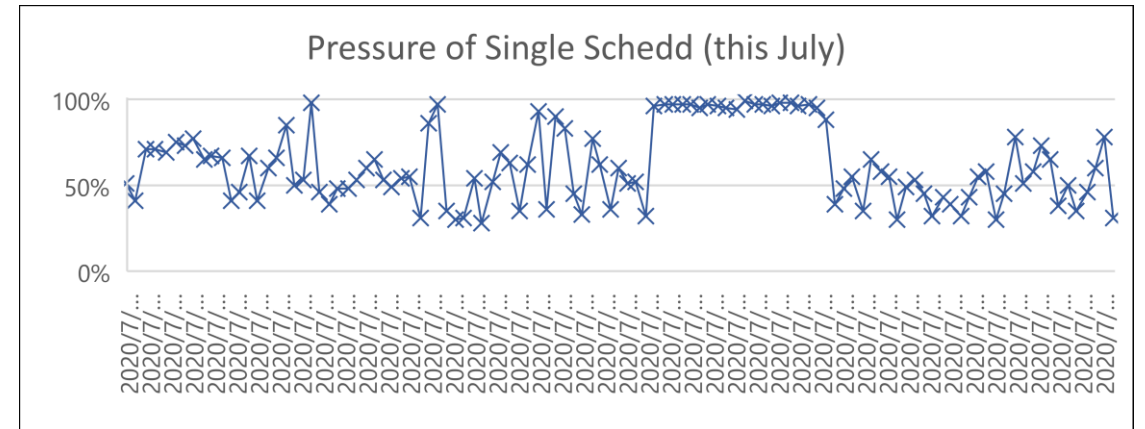
HPL testing

- Target：practical vs. theoretical performance of GPU cards
- Testbed
  - HPE Apollo6500 Gen10: 384GB DDR4 / 8 * NVIDIA V100 / 4 * Mellonox IB 100Gbps
  - Benchmark tools are provided by HPE & NVIDIA
- Results
  - Performance ratio ~70%

Practical and Theoretical Performance Ratio

	1 node	2 nodes	4 nodes	6 nodes
	72	73.8	67.1	69.8
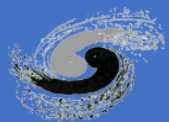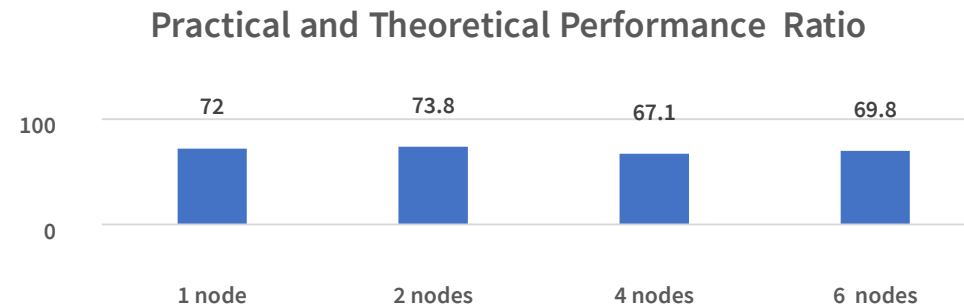
AFS issue

- **Worked as home directory and software publish**
  - 3 replicas of software volume

- **Two cases of AFS clients blocked causing jobs failure**
  - 1) Accessed some V* files cached in local cache directory blocked
    - Locate the files and rewrite on server side and release again
  - 2) RX_CALL_TIMEOUT（-3) error. The /afs mount instances is totally blocked
    - Nothing can help except restart the client nodes to recovery

- **It prefer to happen in the compute blades when there are large IO jobs especially the network throughput is up to or over the export bandwidth(10Gbps/s)**

- **Discussed and diagnostic with AFS administrator (Jan Iven) at CERN, We all think it's a bug and we did some work**
  - Upgrade the fileserver configuration to have more callbacks("-cb 4000000") and switch to the dafileserver
  - Downgrade the AFS client version to 1.6.22.3-1
  - To reduce the AFS access, most experiments software published in AFS has been migrated to CVMFS

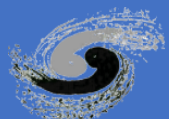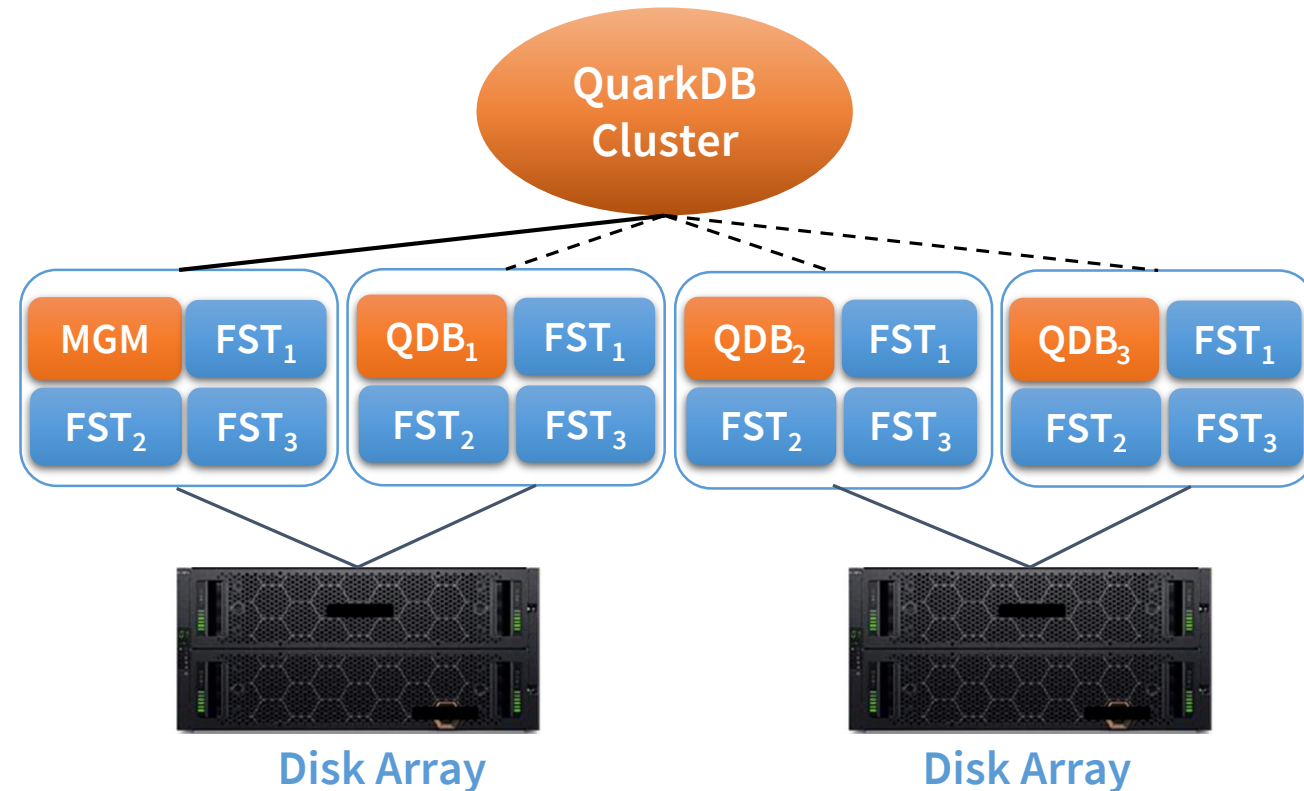- **After these activities, the case 1) disappear, but RX_CALL_TIMEOUT error exist**

Peak: 2.74GB/s
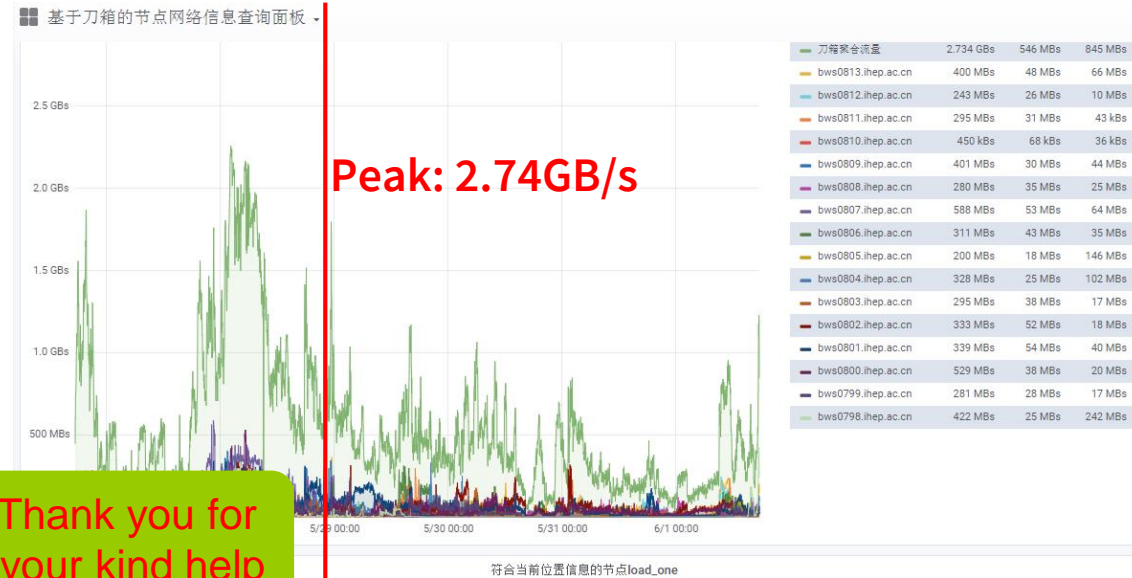
Thank you for your kind help

Case 2 appears(/afs blocked) after the Peak
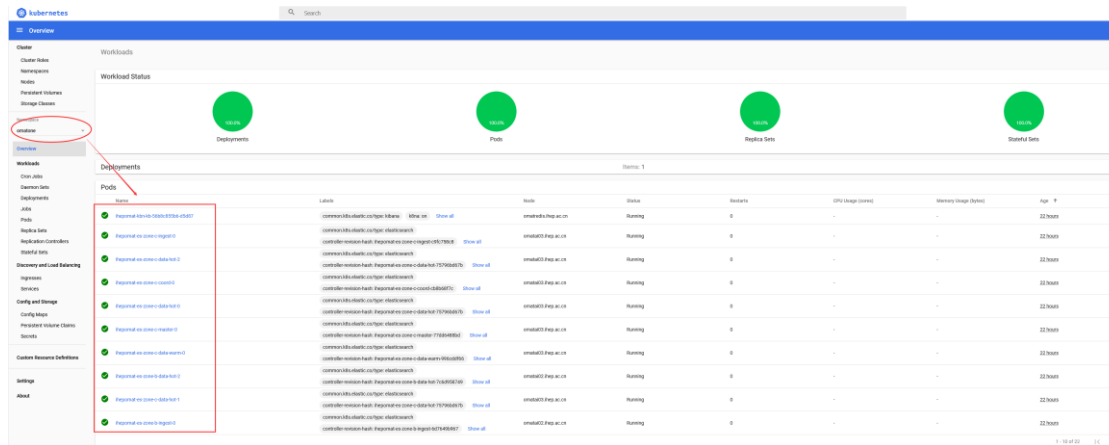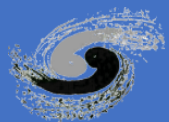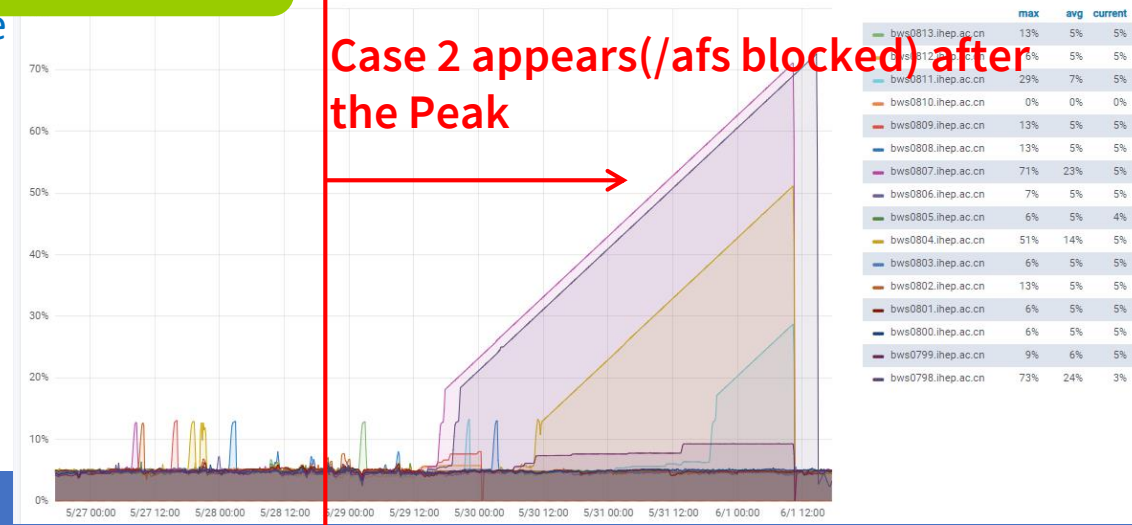
JUNO EOS testbed

- **JUNO EOS testbed Status**
  - Use QuarkDB as namespace
  - Hardware Setup
    - 4 nodes, 2 HDD arrays, and 84 HDDs.
  - Software Setup
    - EOS : 4.7.7
    - QuarkDB: 0.4.2
  - EOS Setup
    - 1 MGM/FST node: junoeos01
    - 3 FST/QuarkDB nodes: junoeos[02-04]
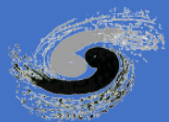   - 3 FSTs on each FST node
  - Capacity: 900 TB, 150 TB used.

QuarkDB Cluster

| MGM | FST_1 | | QDB$_1$ | FST_1 | | QDB$_2$ | FST_1 | | QDB$_3$ | FST_1 |
| FST_2 | FST_3 | | FST_2 | FST_3 | | FST_2 | FST_3 | | FST_2 | FST_3 |

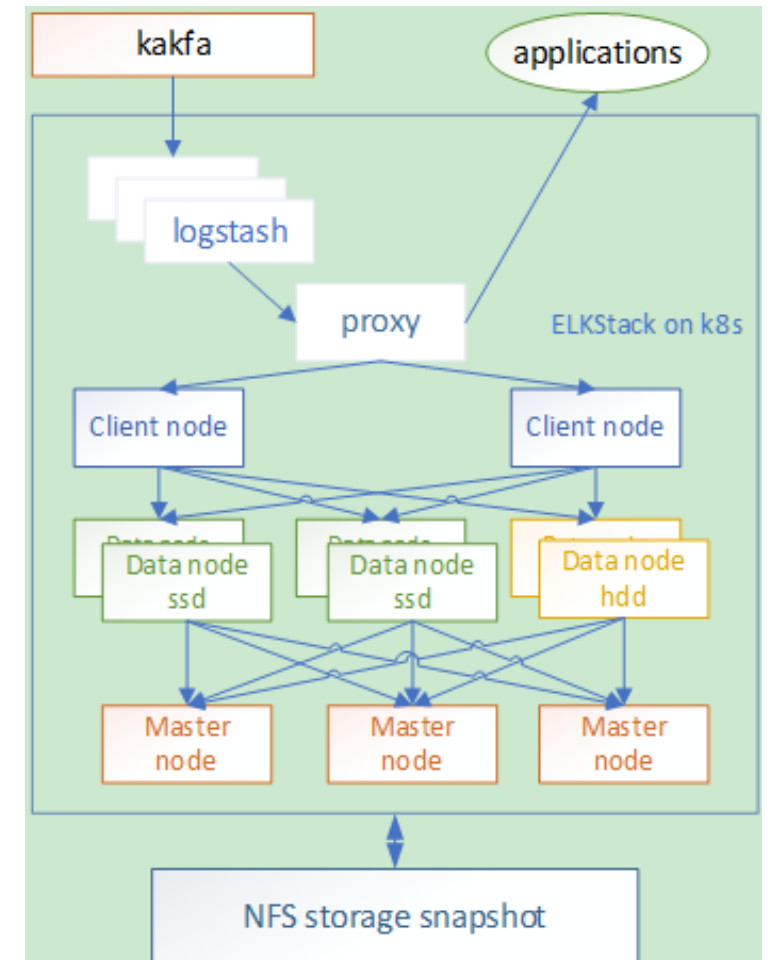**Disk Array**   **Disk Array**

Monitor and Analysis

- A new framework based on k8S clustering has been deployed
  - more stable data query
  - faster monitoring data indexing
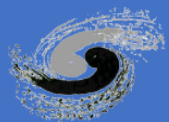 - Better hardware resources efficiency

Summary

- Both computing and storage scale expanded

- Software upgrades has been done and the IHEP site keeps running smoothly

- Taking efforts to meet the requirements from the experiments
  - Container job
  - EOS SE

Thank you!

Question?