Status of IHEP Site

Jingyan Shi , shijy@ihep.ac.cn

Computing Center, IHEP

2016 Spring HEPiX Workshop

Outline

- New Resources
- Current Status
- Problem We Met
- Next Plan

New Resources Added Last Half Year

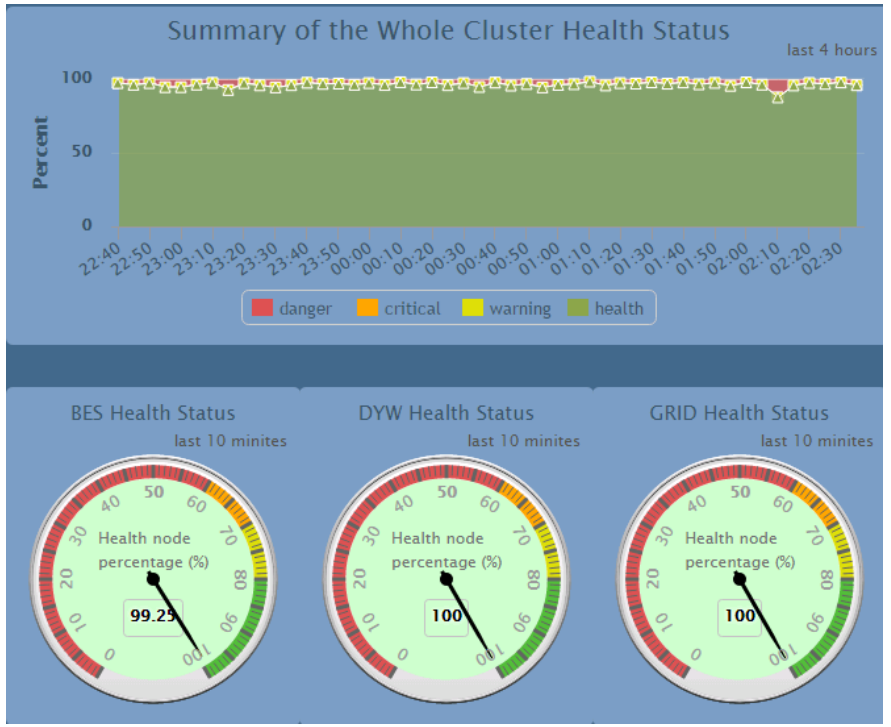
- Computing Resources
  - **86** new blade servers have been added to the local cluster
    - Lenovo Flex System x240 M5
    - CPU E5-2680 v3
    - Total CPU cores is **2064**
  - 368 slow CPU cores have been retired
  - Computing power of local cluster has increased by **32%**
    - HEPSPEC06 -- Before: **173099** Now: **229047**
- Storage
  - **1PB** storage added
    - DELL MD 3860F: 4T * 60 / Disk Array
  - **740** TB 4 year old disk arrays have been retired

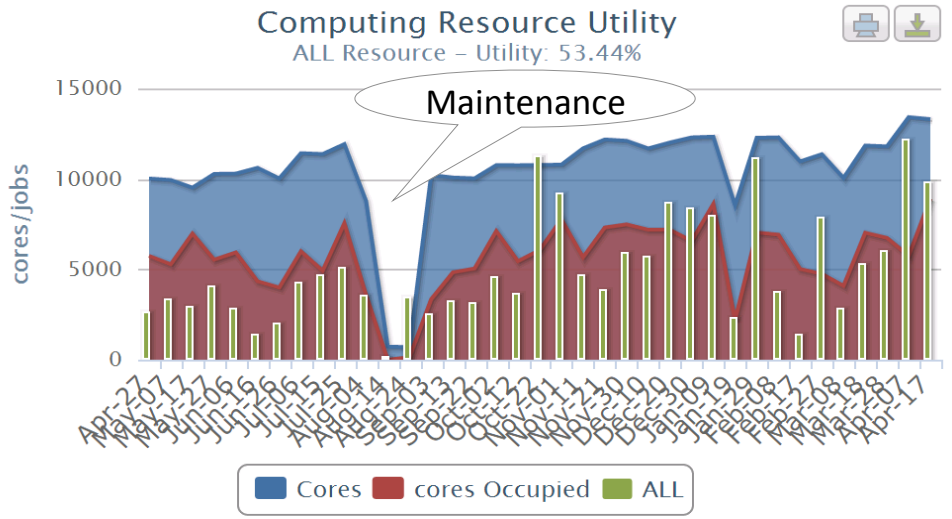
Local Cluster (1)

- Support BESIII, Da-ya Bay, JUNO, astrophysics experiments ……
- Computing
  - ~**13.5K** CPU cores, **300** GPU cards
  - Mainly managed by Torque/Maui
  - 1/10 has been migrated to HTCondor
  - HTCondor will replace Torque/Maui this year
- Storage
  - **5** PB of tape library.
  - **5.7** PB of Lustre. Another 2 PB will be added this year
  - **734** TB of gLuster with replica feature
  - **1.2** PB of other disk spaces
  - EOS is being evaluated, and a possible solution in the future

Local Cluster (2)

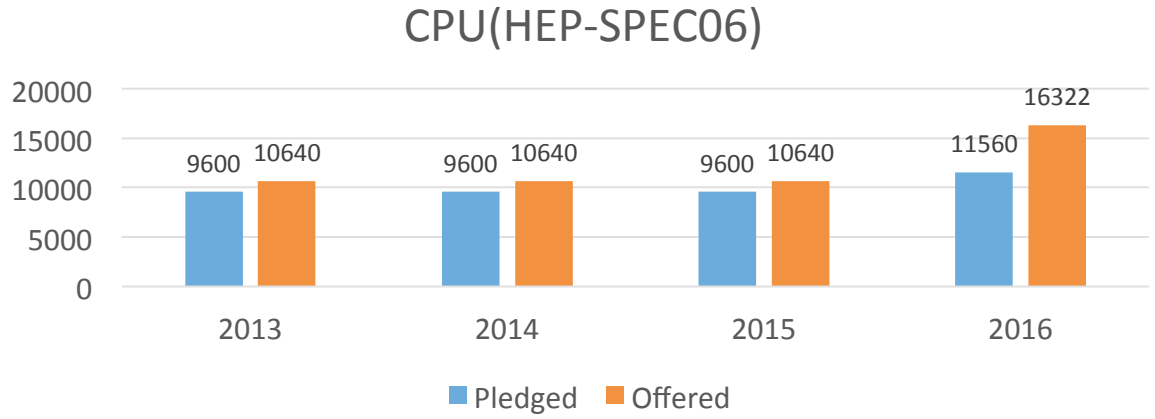
- Automatic deployment: Puppet + Foreman
- Monitor:
  - Icinga
  - Ganglia
  - Flume+Elastic Search
- Resource utilization rate: ~53%

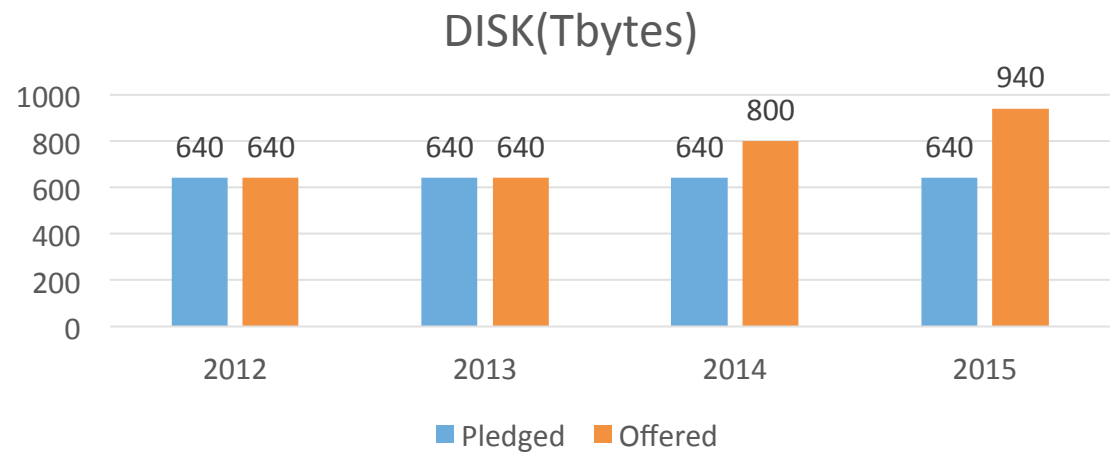




Grid Tier 2 Site (1)

- CPU: 888 Cores
  - Intel E2680V3: 696 Cores
  - Intel X5650 192 Cores
  - Batch: Torque/Maui
- Storage: **940** TB
  - DPM: Total 400TB
    - 4TB X 24slots With Raid 6
    - 5 Array boxes
  - dCache:Total:540TB
    - 4TB X 24slots With Raid 6. 8 Array box.
    - 3TB X 24slots With Raid 6. 1 Array box
- Network
  - 5 Gbps link to Europe visa Orient-plus and 10 Gbps to US

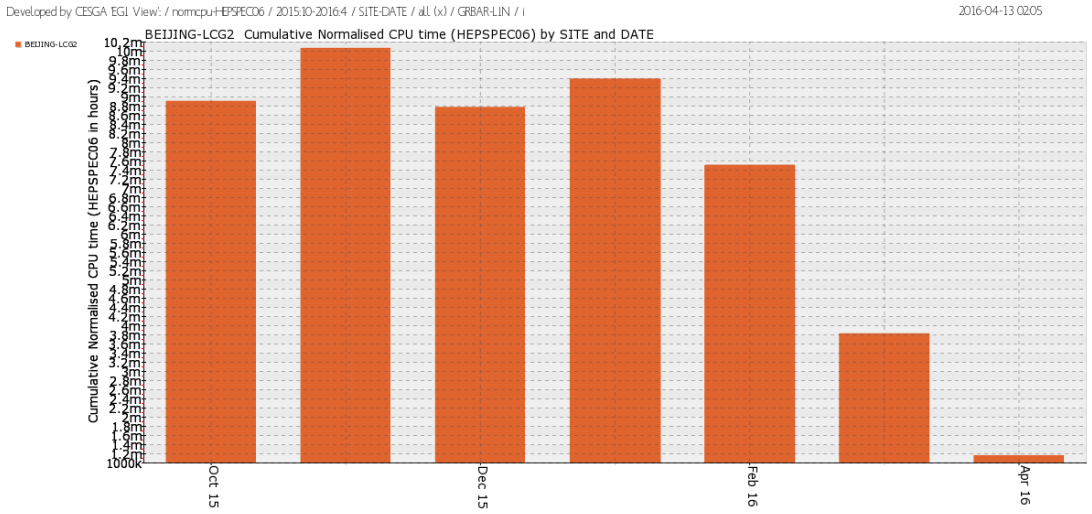




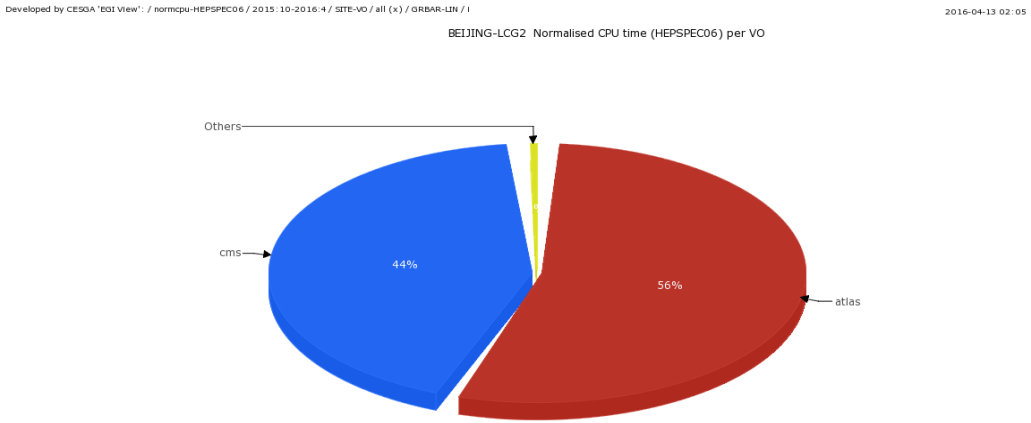
Grid Tier 2 Site (1)

- Site Statistics last half year

CPU Time (HEPSPEC06 in hours)



CPU Time (HEPSPEC06) per VO

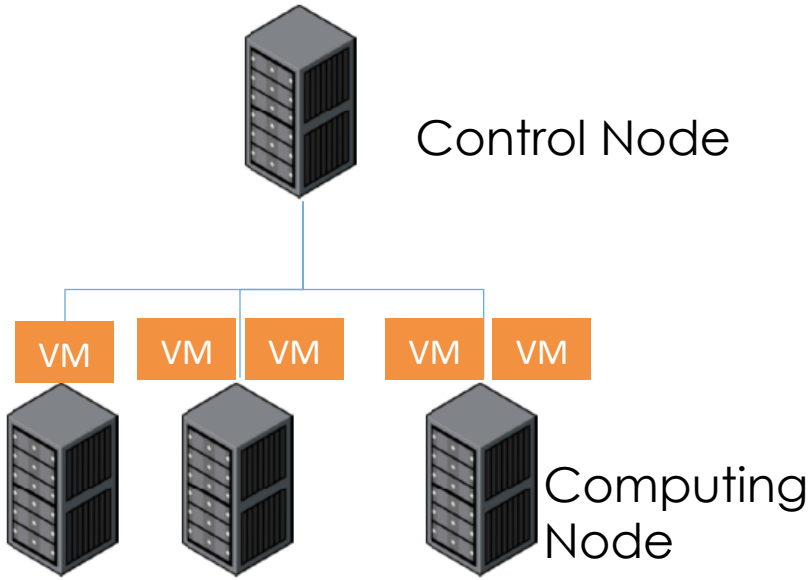


Grid Tier 2 Site Trouble Shooting

- 2016.01: cvmfs hung on some new node.(over monitored)
  - Increase interval of Nagios monitor, and stop reload cvmfs client
- 2016.02: iowait very high: 35% of cpu tps:290.
  - Lots of pilot jobs running on one node.
  - change the share policy between production and pilot jobs

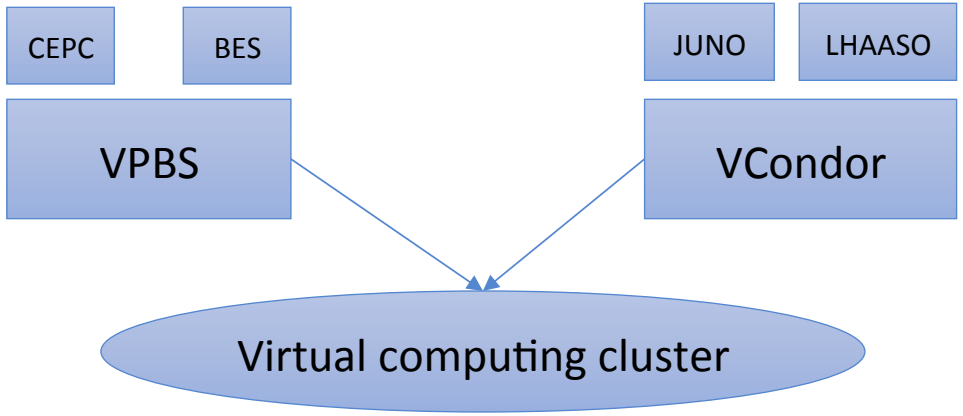
IHEP Cloud

- IHEP Private Cloud
  - Aims to fit peak computing requirement and promote resource utilization
  - Provides virtual machine on demand of real computing requirement
- Resource: 30 physical machines - 720 CPU cores
  - Established in 2015
  - Based on Openstack
    - Upgrade from JUNO to KILO
- More detail in Haibo's talk on Thursday

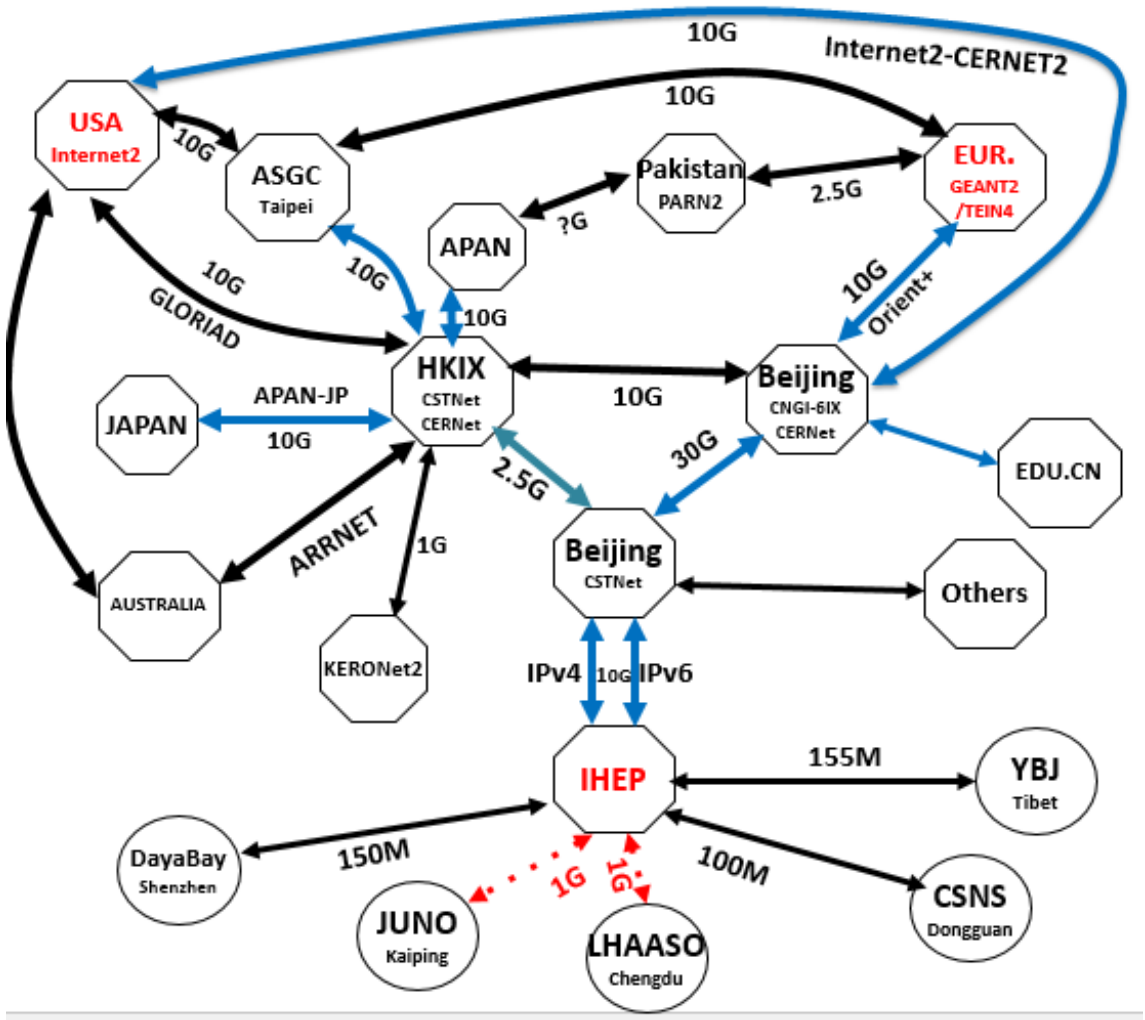


IHEP Cloud Status

- Support 3 experiments
- Running status:
  - VPBS CEPC: ~1,500 jobs, 12,000 CPU hours/week.
  - Vcondor LHAASO: ~1,700 jobs, 23,000 CPU hours/week
  - Vcondor JUNO: ~45,545 jobs(mainly short job, average 152s), 2,000 CPU hours/week
- More experiments will be supported



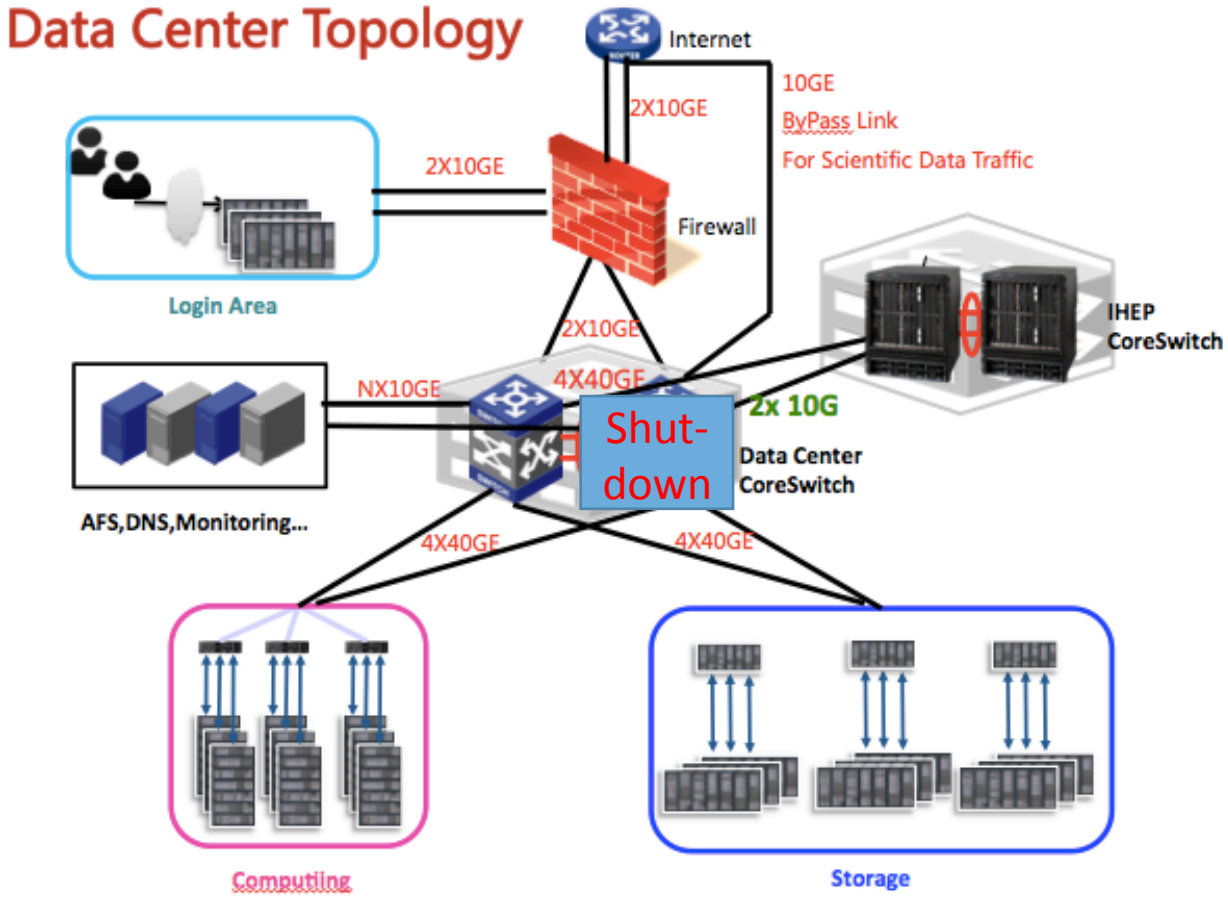
Internet and Domestic Network



- IHEP-EUR.: 10Gbps
- IHEP-USA: 10Gbps
- IHEP-Asia.: 2.5Gbps
- IHEP-Univ.: 10Gbps

Problems We Met

- Lustre mds hung unexpectedly
  - Some measures have been taken
    - Check high density blade servers
    - Shutdown one core switch
      - will add network cards to remain the 4X40Gbps bandwidth
    - Reduced unnecessary mds loads
    - Identify lustre bug
  - Error frequency and impacts have been decreased for two months
- NIC of gLuster servers was unstable
  - replaced these network adapters
  - **running well** now

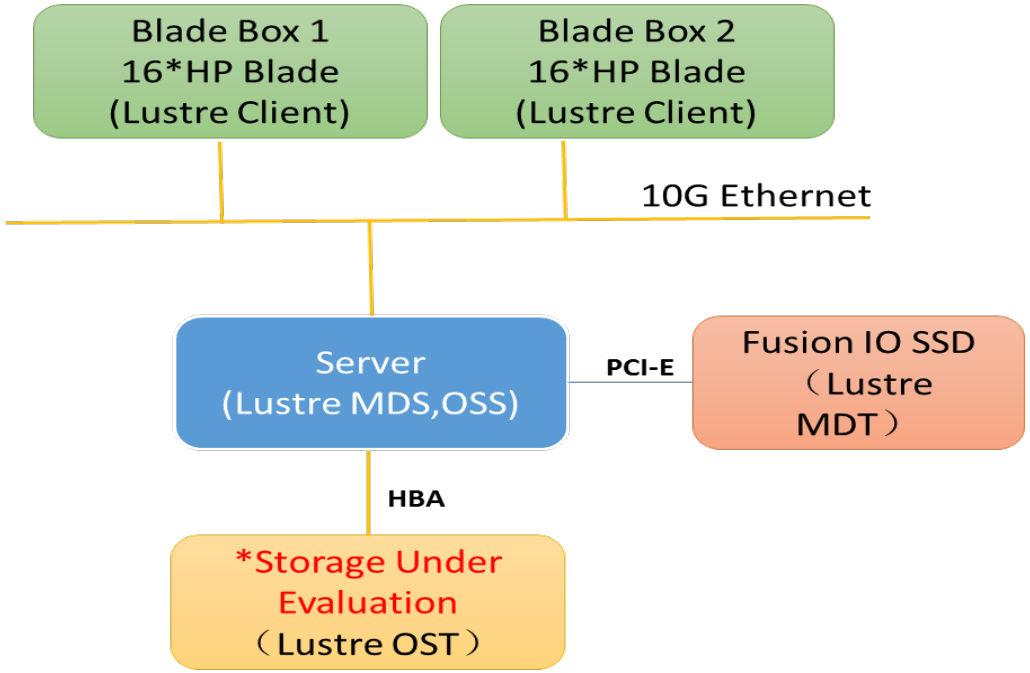


Next Plan

- New storage purchase
- HPC plan

Storage Hardware Evaluation(1)

- Basically, we are happy with the DELL storage hardware.
- According to the requirement of the government procurement, **at least 3** suppliers should be invited to bid for future storage purchasing .
- A uniformed testbed has been setup.
- Products included in the evaluation:
  - MicroSAN MS3300
  - Toyou ISum780
  - Huawei Oceanstor



Storage Hardware Evaluation(2)

- I/O performance is **not the only** metric in our consideration.

Metrics	Evaluation Method
IO throughput	IOzone Cluster Test on the testing bed
Job Efficiency	Realistic BES job efficiency on testing bed
Reliability	Job failure rate during disk unplugging test
Availability	Efficiency decrease and rebuild time after disk unplugging test
Durability	Historical disk failure records in Computing Center's database
manageability	Configuration/Monitoring/Alerting functions provided by each product
Space density	Capacity(TB)/U
Service response time	Experience of system administrators

- Dell and Huawei are under our consideration
- The bid will be started soon

HPC Cluster Plan

- Funds for HPC cluster will be given next year
- A new heterogeneous hardware platform
  - CPU, Intel Xeon Phi, GPU
- parallel programming supports
  - MPI, OpenMP, CUDA, OpenCL …
- potential usage cases
  - simulation, partial wave analysis, bio-medical study, QCD …
- Evaluation is underway.
  - Network Architecture & technologies
  - InfiniBand network for HPC testbed will be setup soon
  - Slurm is under estimated

Thank you!

Question?