Networking and Computing at IHEP/China

Fazhi QI

CC, IHEP, CAS, Beijing

AFAD2018, Daejeon, Korea

2018-01-29

Contents

- HEP experiments in China
- Networking and Computing environment
  - Network
  - Local Cluster
  - WLCG Tier 2 in Beijing
  - BESIII grid computing
  - Cloud computing
  - High performance computing
- Summary

Experiments at IHEP

BESIII (Beijing Spectrometer III at BEPCII)

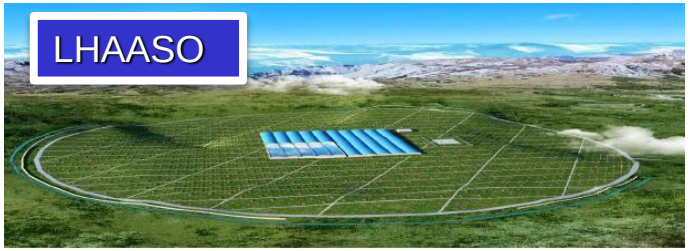
DYB (Daya Bay Reactor Neutrino Experiment)

JUNO (Jiangmen Underground Neutrino Observatory)

CMS

SNS 中国散裂中子源 China Spallation Neutron Source

ATLAS EXPERIMENT



Large High Altitude Air Shower Observatory

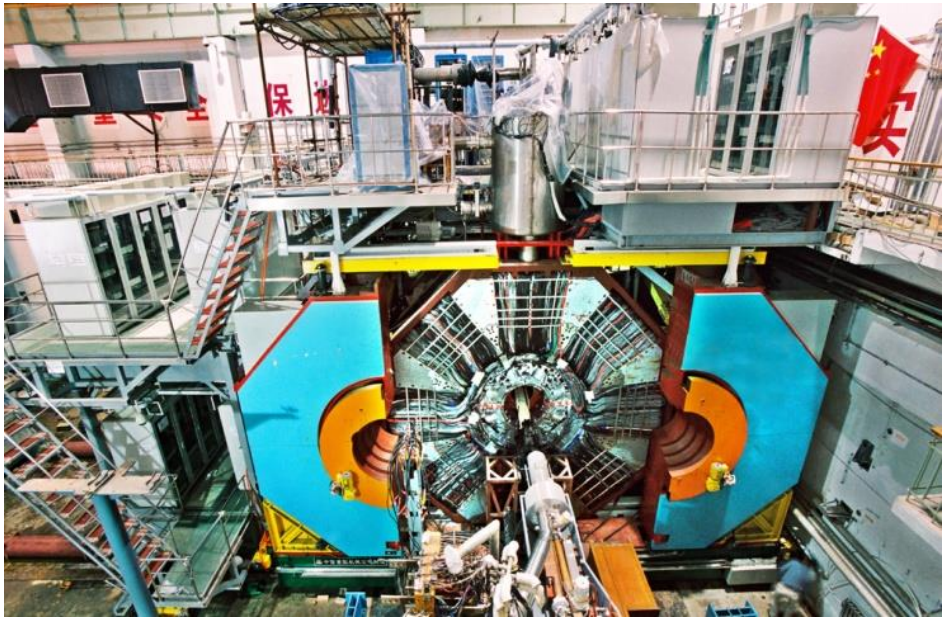
Hard X-Ray Moderate Telescope

BEPCII/BESIII

BEPC II: Beijing Electron-Positron Collider II

BES III: BEijing Spectrometer II, general-purpose detector on BEPC II

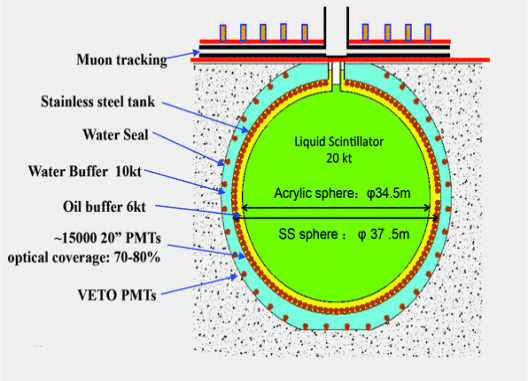
- Studying tau-charm physics
- Upgrade: BEPCII/BESIII, operational in 2008
- 2.0 ~ 4.6 GeV/C
- (3~10)$\times 10^{32}$ $cm^{-2}s^{-1}$
- Produce ~100 TB/year raw data
- ~ 5000 CPU cores for data process and physics analysis

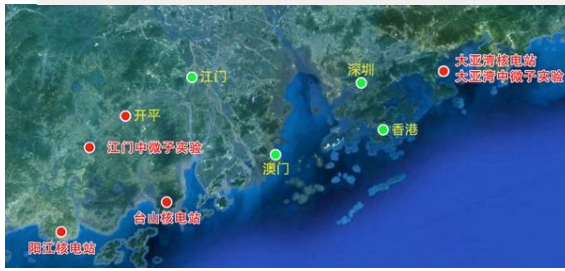


Neutrino experiments

- Daya Bay Reactor Neutrino Experiment
  - To measure the mixing angle θ_{13}
  - 300 collaborators from 38 institutions
  - Produces ~200 TB/year (2011-2018)

- JUNO - Jiangmen Underground Neutrino Observatory
  - Start to build in 2014, operational in 2020
  - 20 kt LS detector, 3% energy resolution
  - To determine the neutrino mass hierarchy using reactor antineutrino oscillations
  - Estimated to produce 2 PB data/year for 20 years





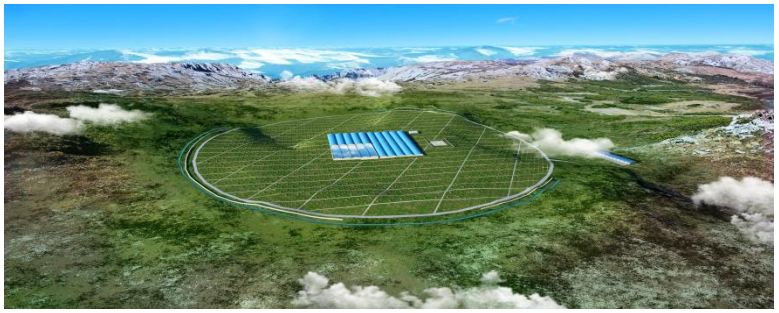
CSNS

- China Spallation Neutron source
- Will on production from Mar. 2018
- Requirements
  - Data Transfer
  - Metadata Catalogue
  - Data Process and Analysis

- Phase One: 3 instruments
  - 1000 cores(cloud) + 2000 cores(HPC)
- Phase Two: 20 instruments
  - The amount of resources currently required * 7

LHAASO

- Large High Altitude Air Shower Observatory, located on the border of Sichuan and Yunnan Province
- Multipurpose project with a complex detector array for high energy gamma ray and cosmic ray detection
- Expected to be operational in 2019
- ~1.2 PB data/year * 10 years
- On-site storage and computing resources. Data will be filtered and compressed and transferred back to IHEP.



HEPS

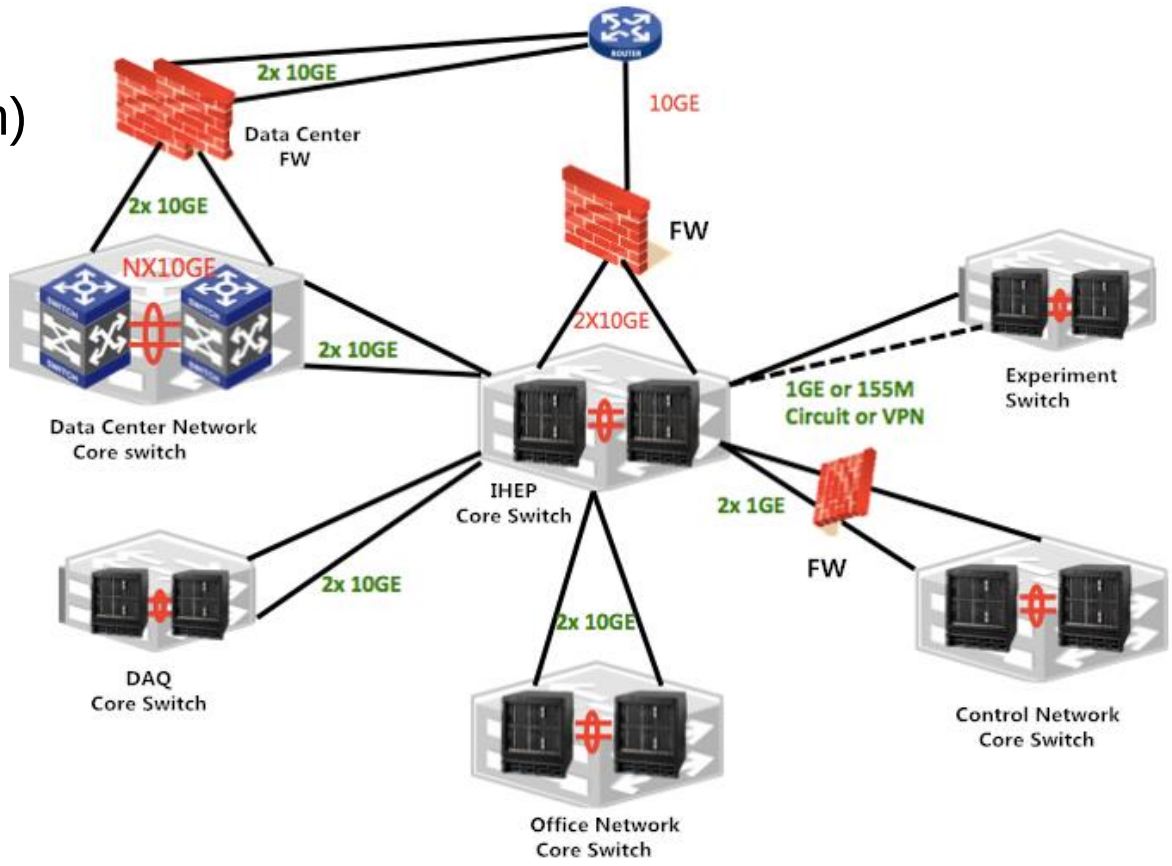
- High Energy Photon Source
- Locates in Huairou/Beijing
- Will start to build from 2018 to 2024 with 14 beam lines
- Data: 180TB/day, 60PB/year
- New computing resources and technologies should be considered for HEPS

Contents

- HEP experiments in China
- Networking and Computing environment
  - Network
  - Local Cluster
  - WLCG Tier 2 in Beijing
  - BESIII grid computing
  - Cloud computing
  - High performance computing
- Summary

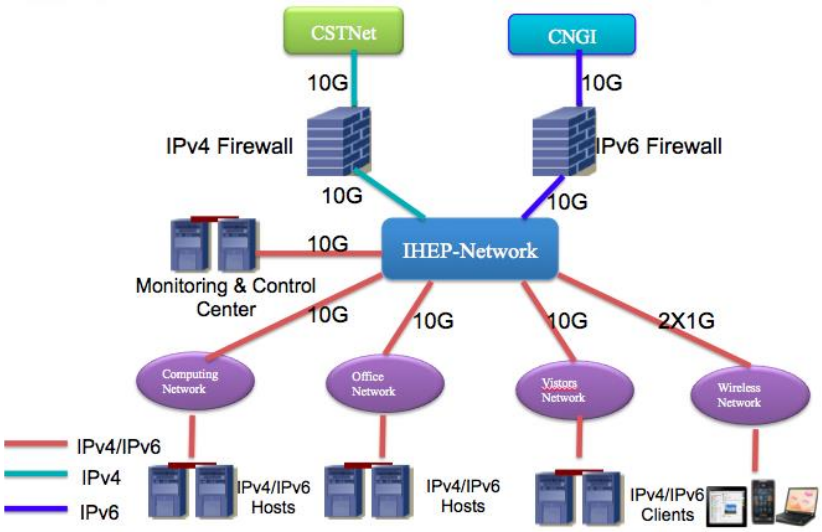
Internal Network in IHEP

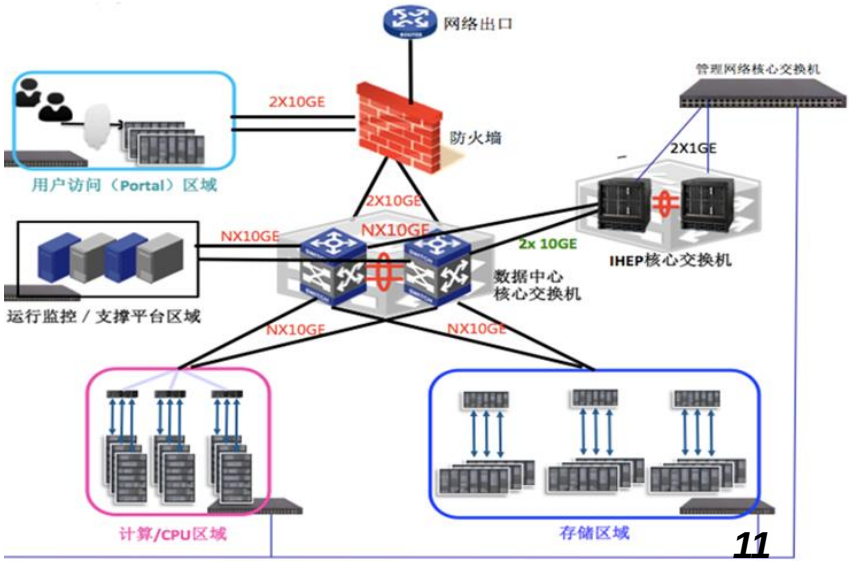
- Multi-Campus
  - Beijing
  - DayaBay (Shen Zhen)
  - CSNS(Dong Guan)
  - Ybj (Tibet)
  - LHAASO(Cheng Du)
  - JUNO(Kai Ping)
  - HEPS (Huai Rou)
  - ……



Network Structure

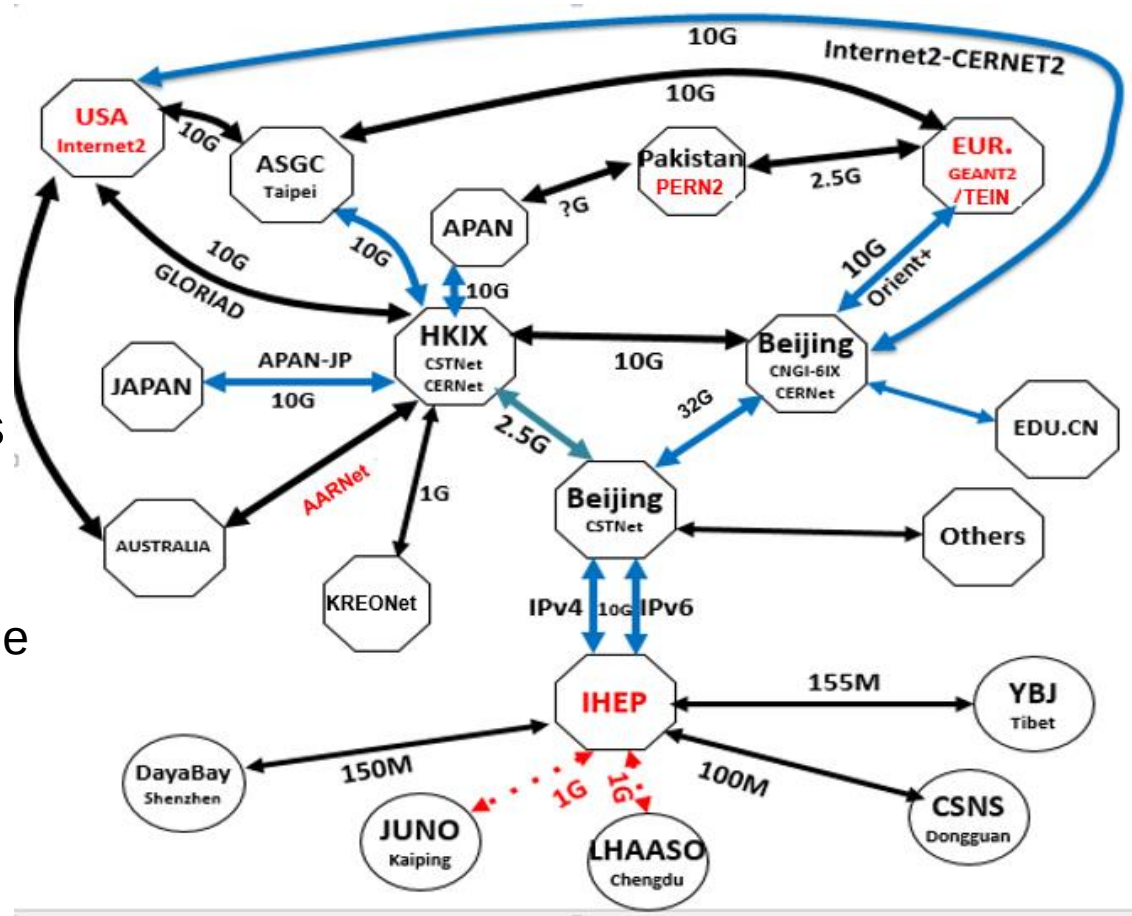
- For office users
  - The largest campus network and bandwidth among all CAS institutes
    - 10G backbone
    - IPv4/IPv6 dual-stack
    - Wireless covered
  - Email/web/ services
  - >5000 end users
- For the data center at computing center
  - 160 Gbps (4X40Gbps) for 2-layer switches
  - 2X10 Gbps for storage nodes





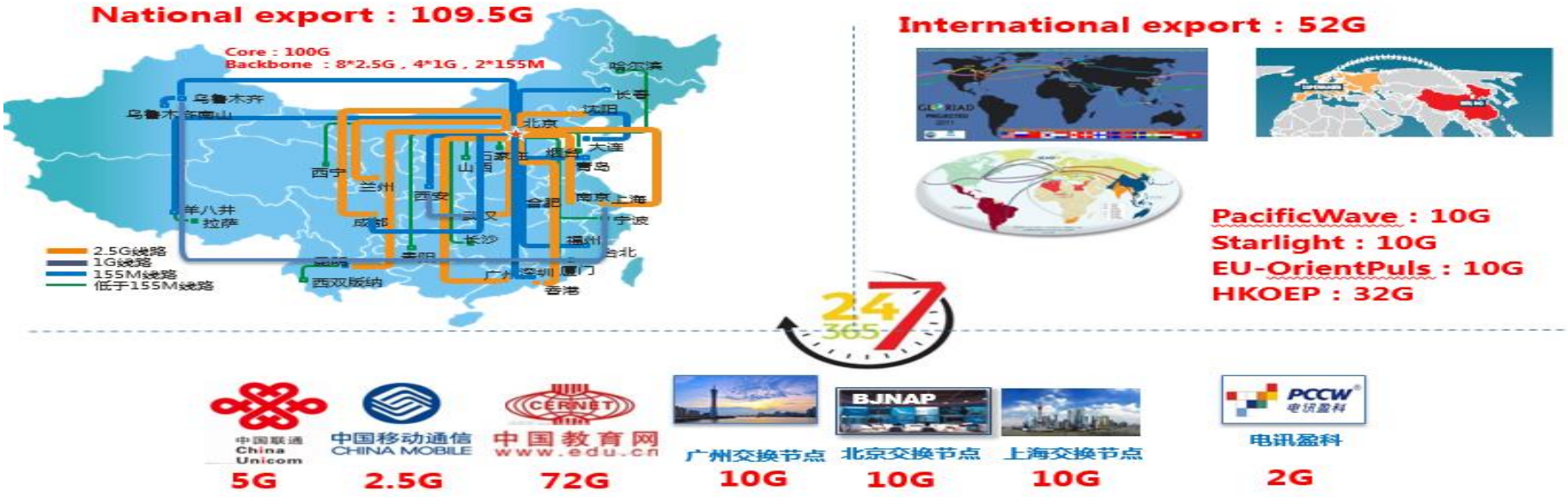
International and domestic links

- Dedicated Links for three other IHEP sites (other three in the future)
  - Shenzhen (Dayabay)
  - Dongguan (CSNS)
  - Tibet (YBJ/ARGO)
  - Kaiping (JUNO)
  - Chengdu (LHAASO)
  - Huairou(HEPS)
- Good Internet connections
  - IHEP-Europe: 10 Gbps
  - IHEP-USA: 10 Gbps
  - ~4 PB/year data exchange
- Network for R&E in China
  - CSTNet
  - CERNet/CERNet2



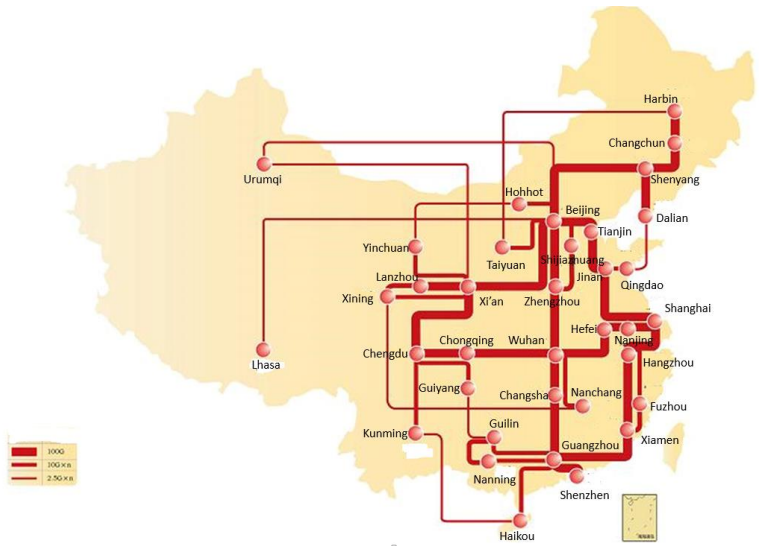
R & E Networking in China(1)

- CSTNet
  - China Science and Technology Network
  - Comprises more than 100 research institutes, three universities
  - Has one core center and 12 regional branch centers
  - Backbone bandwidth is 10Gpbs for domestic
  - Shares the 10Gbps link between Beijing and Europe (ORIENT+) for the China-Europe traffic with CERNET
  - Also the Gloriad member (from Hong Kong to USA,10Gbps)



R & E Networking in China(2)

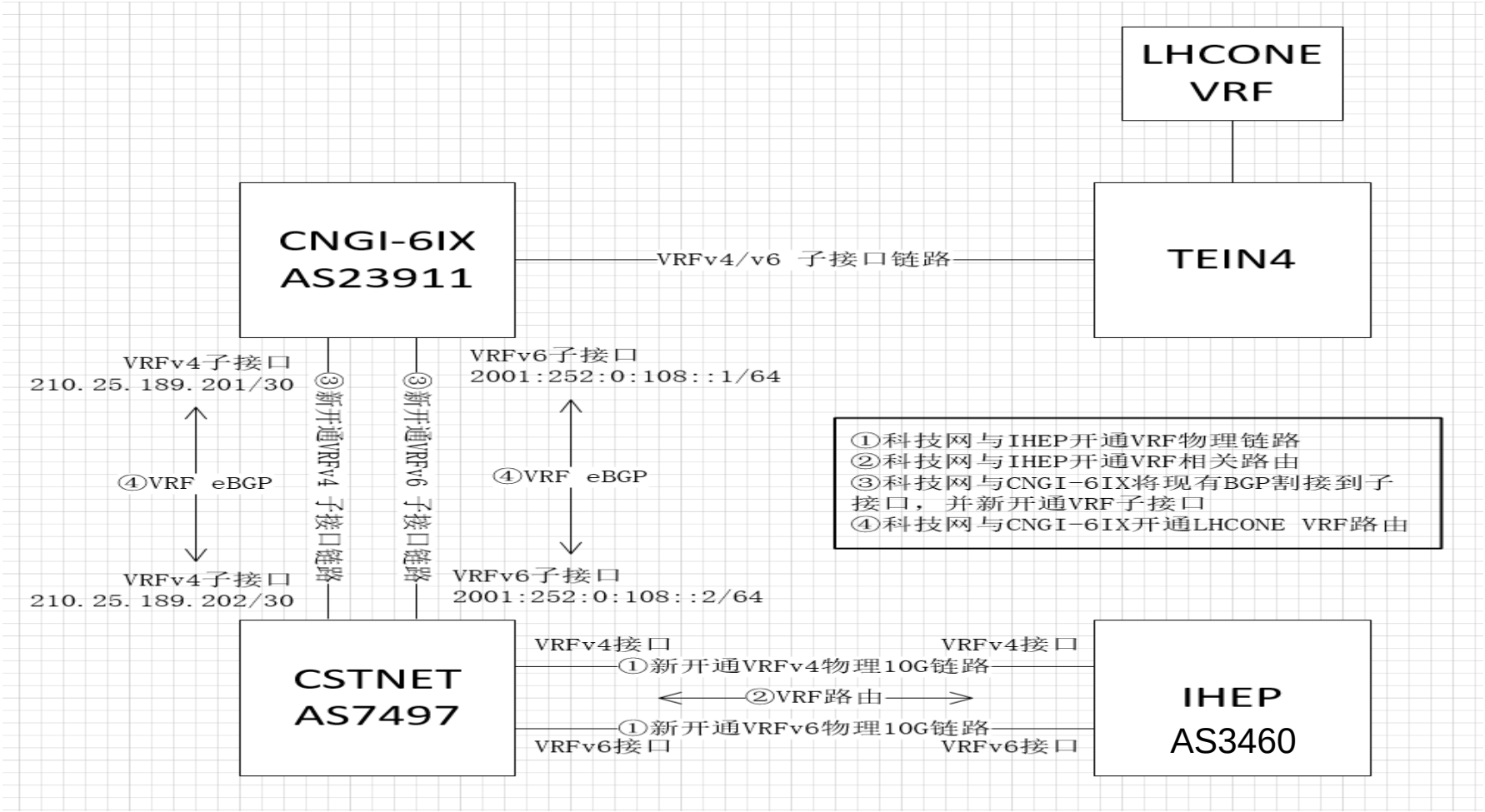
- CETNet & CERNet2
  - China Education and Research Network / 2nd generation
  - 38 PoPs in 36 cities and over 2600 universities/institutes/organizations connected
  - Backbone bandwidth is 10~100Gbps
  - Two international network exchange centers: Beijing and Hong Kong
  - The Beijing NOC locates in Tsinghua University
  - Good connections to North America, Europe and Asia-Pacific



LHCONE Status/Progress in China

- Network services provides in China
  - Supports from CERNET/CERNER2 and CSTNet are very important and necessary
  - Approved by CERNet and CSTNet in 2017
- LHCONE community
  - 2017.2: Taipei ; ASGC (Hsin-Yen Chen), IHEP(Gang Chen, Fazhi Qi); technology discussion
  - 2017.3 : CERN; LHCONE community (Edoardo Martelli),IHEP(Gang Chen, Fazhi Qi) ; technology discussion
  - 2017.4 : Information exchange among IHEP(Gang Chen, Weidong Li, Fazhi Qi) , LHCONE community (Edoardo Martelli) and CERNet(Xing Li, Zhonghui Li); technology discussion
- Progress
  - The VRF peer to TEIN finished by CERNET2 in Sep. 2017
  - 2017.12:Start to deploy the network infrastructure for LHCONE in CERNet-CSTNet-IHEP
  - Two dedicated links (10Gbps ipv4 + 10Gbps IPv6) have been set up for LHCONE between IHEP and CSTNet
  - Start to do some tests for LHCONE
  - And will setup VRF peer to Internet2,ASGC in Hongkong(HKIX).....

LHCONE Deployment in CERNET-CSTNET-IHEP

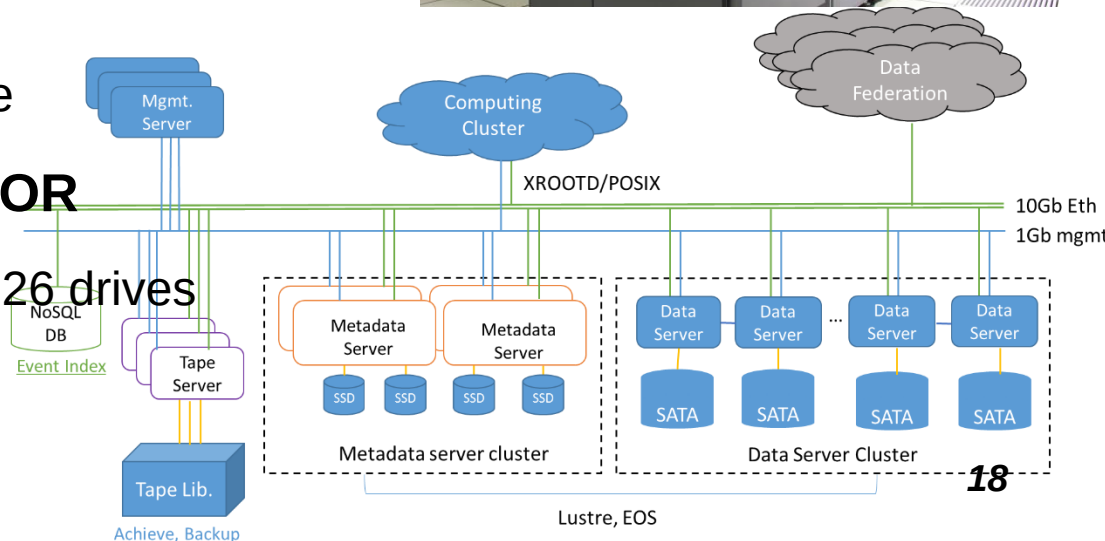


Computing resources in IHEP

- Clusters in Beijing
  - ~13,500 CPU cores
  - 8 GPU cards
  - Scheduler:
    - PBS-2.5.5 with Maui-3.3.1
    - HTCondor 8.2.5
- Clusters in Dongguan(CSNS)
  - ~3,000 CPU cores
  - Scheduler:
    - PBS-2.5.5 with Maui-3.3.1
    - Slurm
- Grid site (WLCG)
  - 1,200 CPU Cores
  - CreamCE (PBS-2.5.5 with Maui-3.3.4)
- The BESIII DIRAC-based distributed computing system
  - ~ 2,000 CPU cores
- IHEPCloud based on Openstack
  - ~ 1000 CPU cores

Storage

- **Lustre as main disk storage**
  - Capacity: 10 PB storage
- **Gluster system**
  - 734TB +350TB storage with replica feature
- **EOS system**
  - Capacity: 1.8 PB storage
- **DPM & dCache**
  - 940TB, With SRM interface
- **HSM, with modified CASTOR**
  - 2 tape libraries + 2 robots, 26 drives
  - Capacity: 5 PB



LCG Tier-2 Site at IHEP

- **Have been running for more than 10 years**
  - MoU signed with CERN in 2006
  - Mainly supports ATLAS and CMS
- **Resources**
  - 1200 cpu cores, 980TB storage
  - 3-5 PB data exchange each year
- **Use atlas@home to exploit extra CPU**
  - BOINC acts as a second scheduler, pulling in more jobs (with lowest priority) to increase the CPU utilization of every single work node.
  - CPU utilization can reach 90% on all work nodes, and extra of 23% CPU can be exploited, in the period of 3 weeks, from IHEP Tier-2 site while the site is fully loaded (wall time utilization > 86%)

BEIJING_LCG2 daily utilization in 3 weeks										
	#cores	HS06	Walltime (days)	CPUtime (days)	HS06 (days)	Job CPU eff (%)	Wall Util (%)	CPU Util (%)	HS06 Util (%)	Extra Cores
Grid	420	7560	359	283	5591	78.83	85.48	67.38	73.96	
BOINC	420	7560	367	97	1730	26.43	87.38	23.10	22.88	144.0
Total	420	7560	726	380	7321		172.86	90.48	96.84	
Note: site has 420 dedicated cores for ATLAS, numbers are the average in a consective 3 weeks (starting 2017-08-25), BOINNC processes **0.03 M events/day**										

BESIII Grid Computing

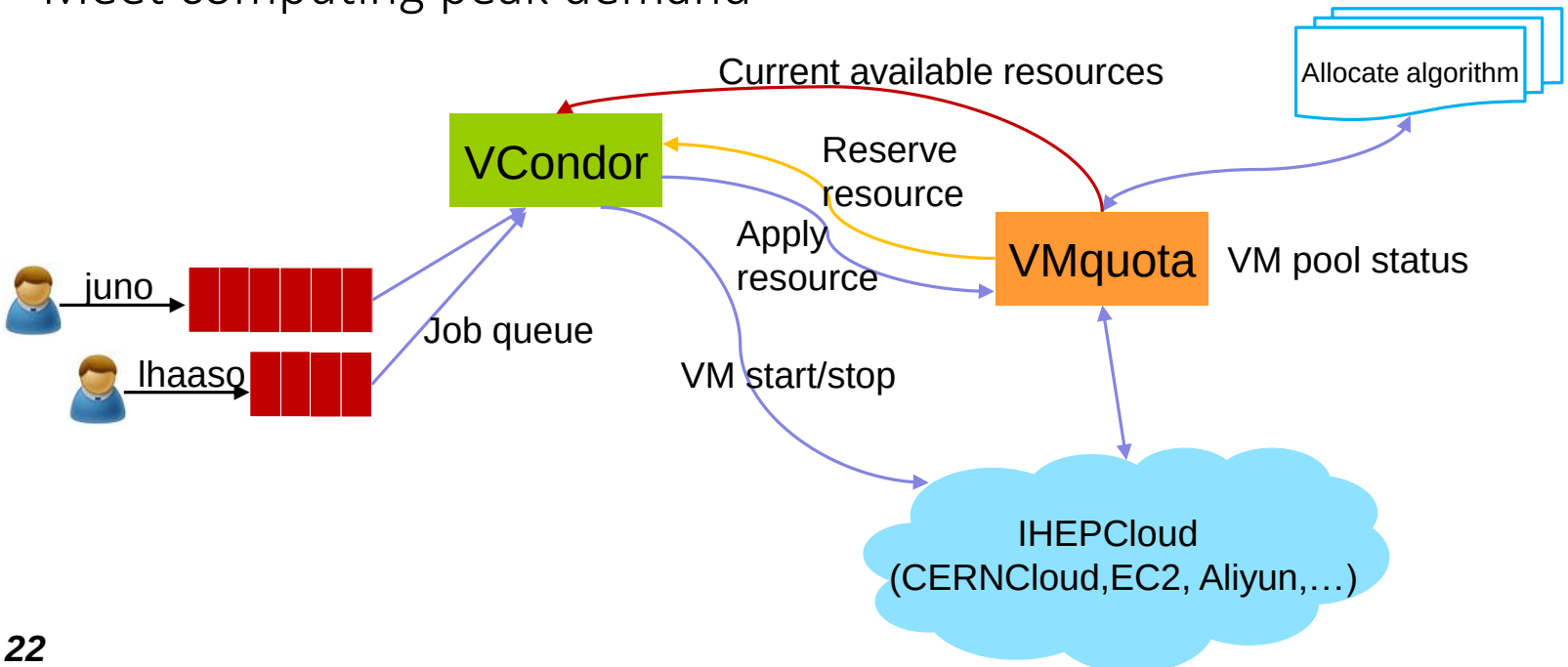
- IHEP as central site
  - Raw data processing, bulk reconstruction, analysis etc
- Remote sites for peak needs
  - MC production, analysis
- Resources
  - 14 sites from USA, Italy, Russia, China universities
  - About 2000 cores CPU resources, 500 TB storage have been integrated

IHEPCloud: a Private IaaS platform

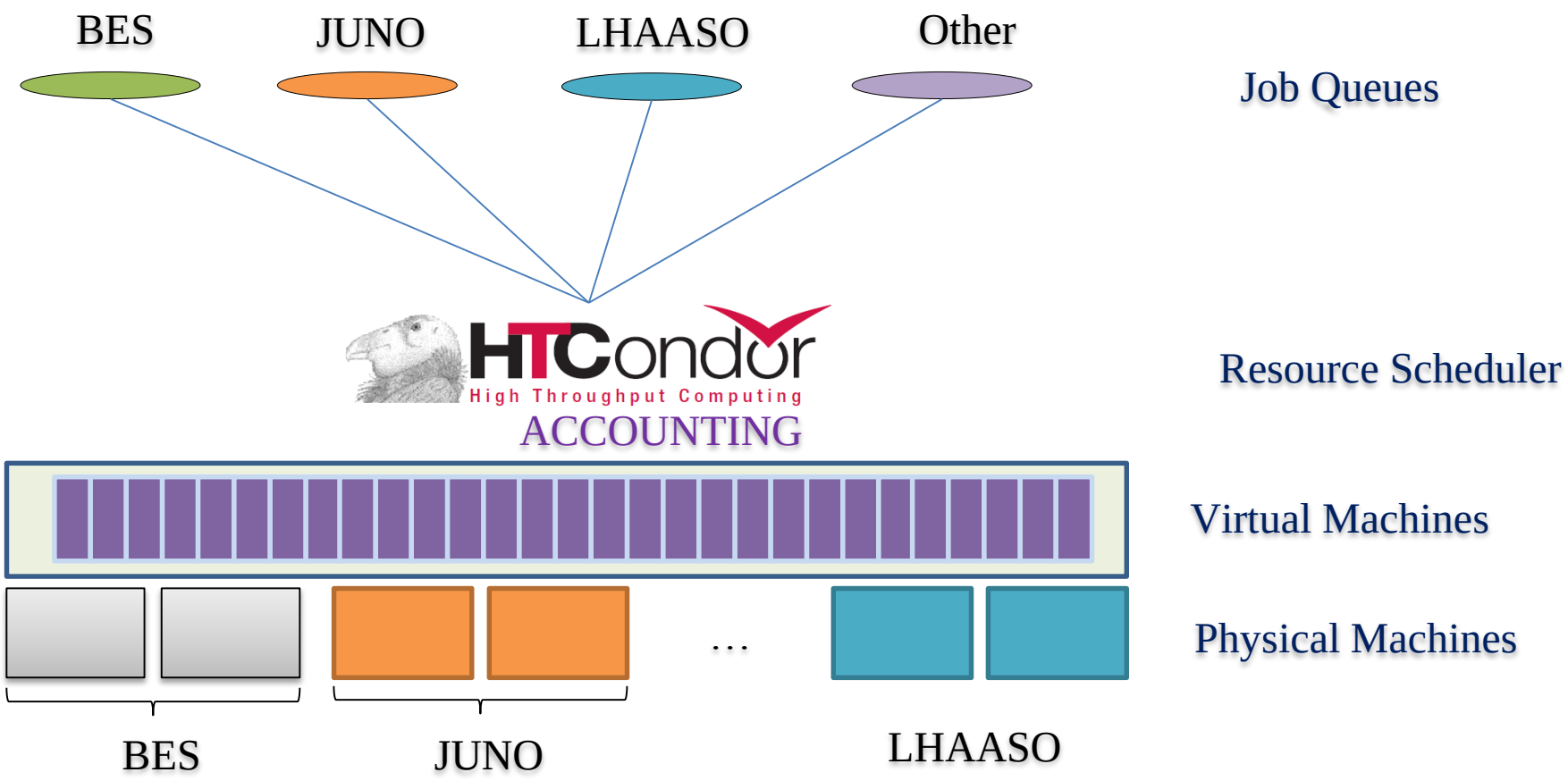
- Launched in May 2014
- Three use cases
  - User self-service virtual machine platform (IaaS)
    - User register and destroy VM on-demand
  - Virtual Computing Cluster
    - Combined with physical queue, jobs will be allocated to virtual queue automatically when physical one is busy.
  - Distributed computing system
    - Working as a cloud site: Dirac call cloud interface to start or stop virtual work nodes

Virtual computing cluster

- ∽1000 physical cpu cores
  - provide virtual machine job slot on demand
  - Allocate resources dynamically to improve resource utilization
- Implement resource integration and share different experimental / Organizational Computing Resources
- Meet computing peak demand

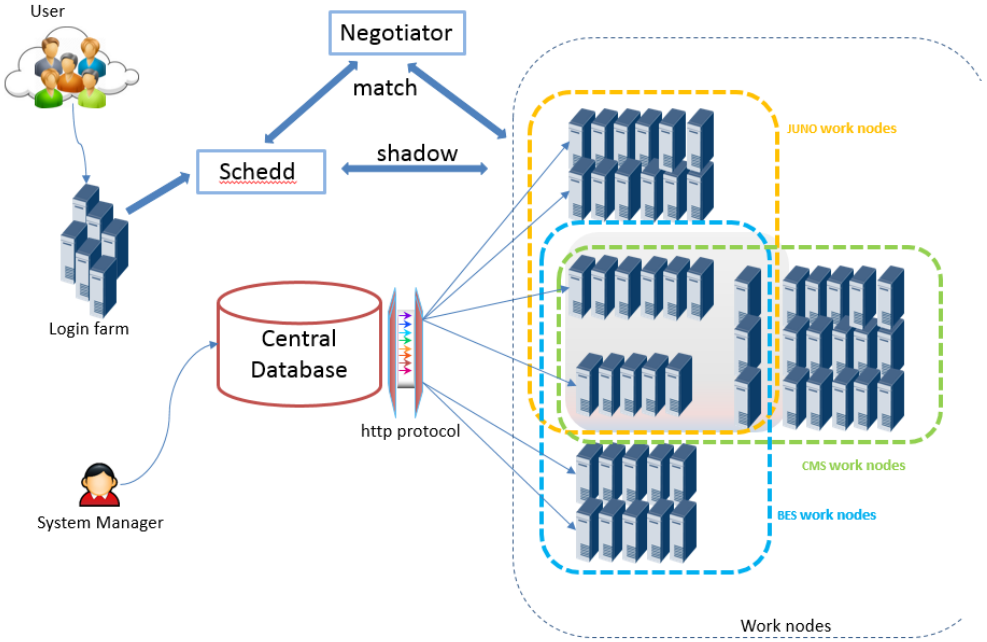


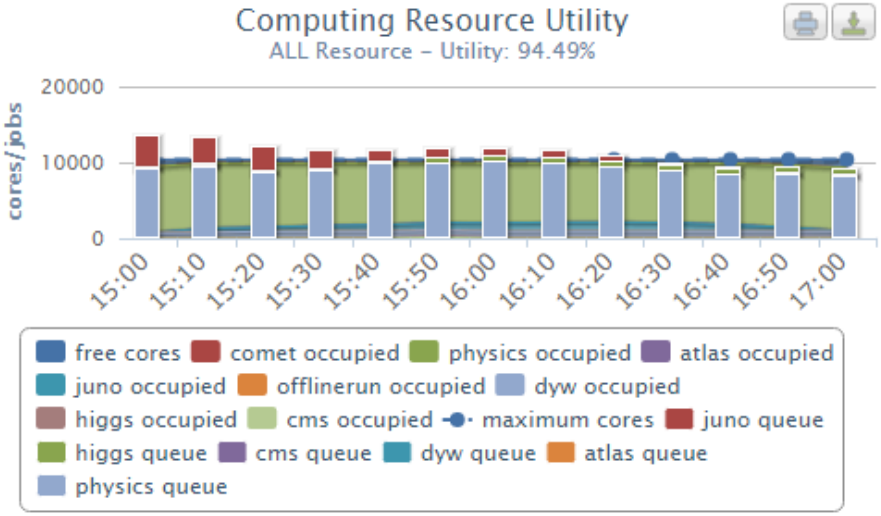
Future setup



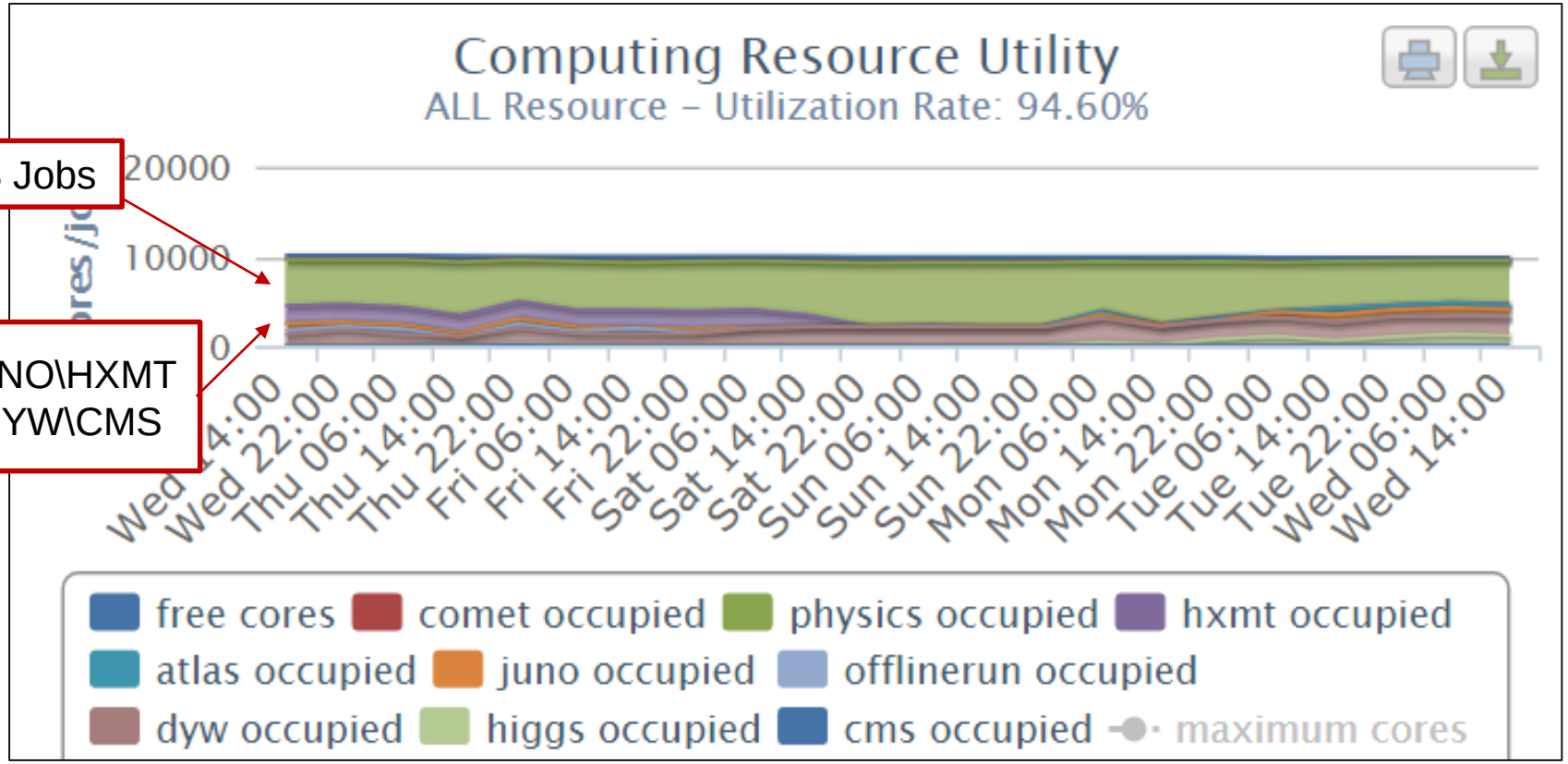
Sharing Policy for HTC

- Sharing Policy
  - Computing nodes are shared for all experiments
    - Less queuing time
  - Quota surplus configured
    - Fair share between experiments guaranteed
  - Error detection and recovery automatically
- Job slots utilization increased dramatically
  - From 60% to 90% each year



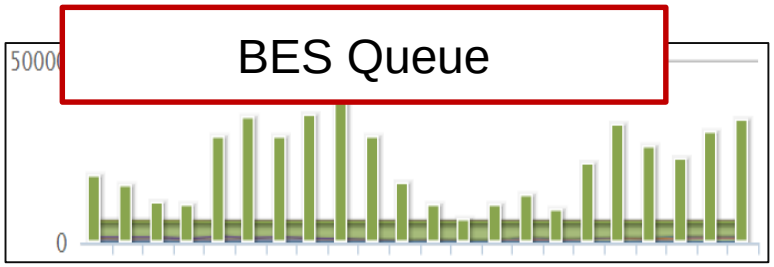


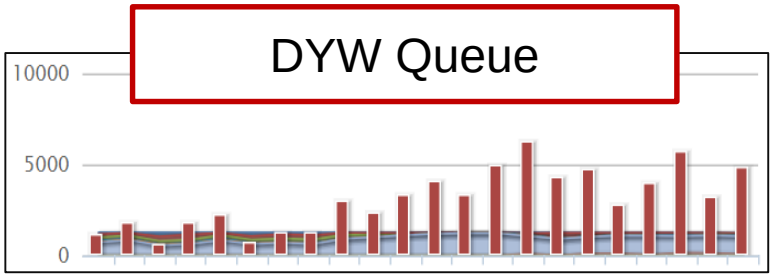
Computing resources Utility



BES Jobs

JUNO\HXMT
\DYW\CMS





High Performance Computing

- Needs from experiments and theoretical calculation
  - BESIII partial wave analysis
  - Geant4 detector simulation (CPU time and memory consuming)
  - Simulation and modelling for accelerator design
  - Lattice QCD calculation
- A HPC cluster at IHEP is being planned in 2017 and will be deployed in 2018
  - NVIDIA Tesla GPUs
  - Xeon Phi coprocessors
  - Interconnected by the InfiniBand network
  - 2P Flops capability

Summary

- China will be a member of LHCONE in 2018 !
- About 20000 CPU cores and 15PB storage resources for HEP experiments at IHEP
- Grid and cloud computing technologies were adopted to support various types of HEP experiments in China.
- Sharing policies among different experiments for CPU resources improve the resources utility
- A HPC platform will be deployed in 2018 at IHEP

Thank You !

謝謝