HEPIX 2015

Computing Resources Virtualization at IHEP

Tao Cui, Qiulan Huang
On behalf of IHEP Cloud Technology Group
2015-10-11

outline

- Requirements of Scientific computing
- IHEPCloud status
- Further work

Requirements of Scientific Computing

Big projects in next 7 years at IHEP

- **Computing request**
  - **16PB storage**
  - **40000 cores**

Project	Start time	Datas/year	total in next 7 years
BESIII	2014	1.5PB	10.5PB
Daya Bay	2014	0.14PB	1PB
JUNO	2018	1PB	2PB
LHAASO	2018	1.2PB	2.4PB
Total	15.9PB		

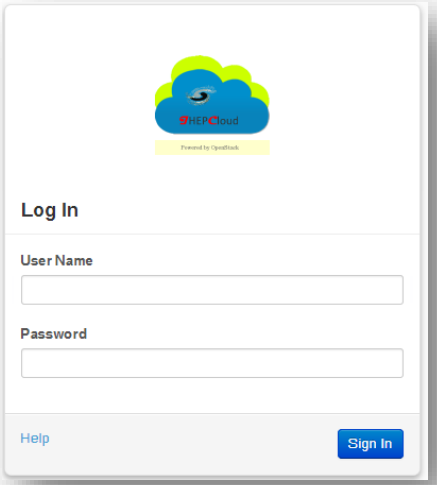
IHEPCloud Introduction

IHEPCloud

- A private IaaS platform based on **Openstack Icehouse**
- Deployed at 2014.11
- To service all IHEP users

Scenarios

- **User self-Service platform**
  - ✓ User register and destroy VM on-demand
- **Virtualized computing cluster**
  - ✓ Integrated with HTCondor/PBS
- **Distributed computing**
  - ✓ Integrated with Dirac
  - ✓ Start/Destroy VMs by calling API according to the requirements



IHEPCloud services

- Who can use?
  - any user who has IHEP email account
- How many resources for user?
  - By default, each user has 3 CPU cores and 15GB memory
- VM types
  - testing machine
    - full root privilege, no public storage
  - UI node
    - AFS authentication, No root privilege, public storage
    - No some limitations like memory, CPU time, process
- OS types
  - SL 55, SL 58, SL 65, SL 7 64 bits, SL 65 32 bits, Win7 64 bits
  - add new types depends on user requirement
- VM IP address
  - internal IP address (192.168.*.*) is allocated automatically
  - external IP address (202.122.35.*) need the approval of administrator

IHEPCloud Current status

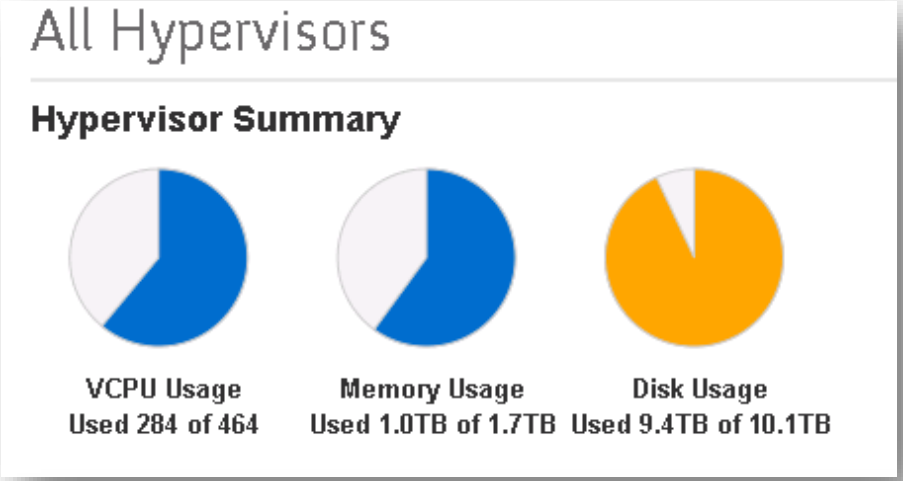
Resources

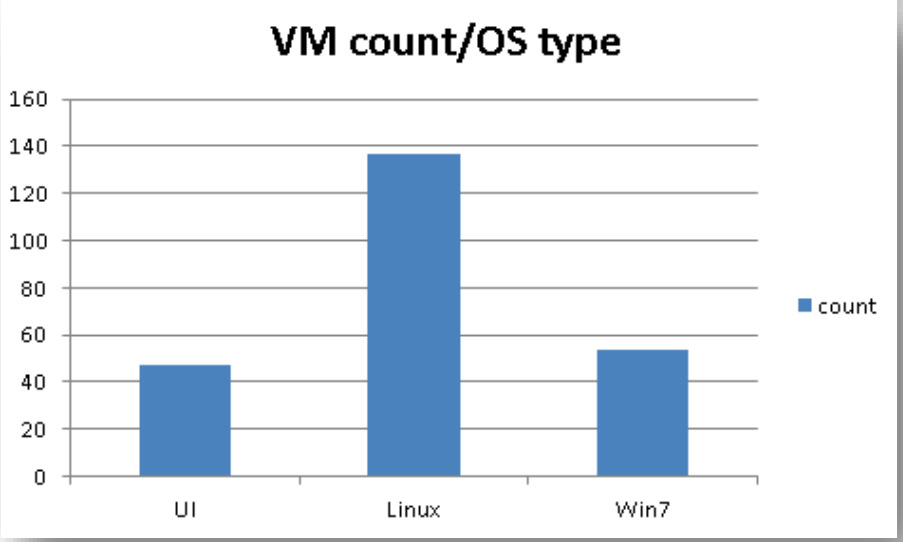
- 24 hypervisor
- 464 cores
- 1.7TB Ram
- 10.1TB Storage
- >1000 Users

Resources usage

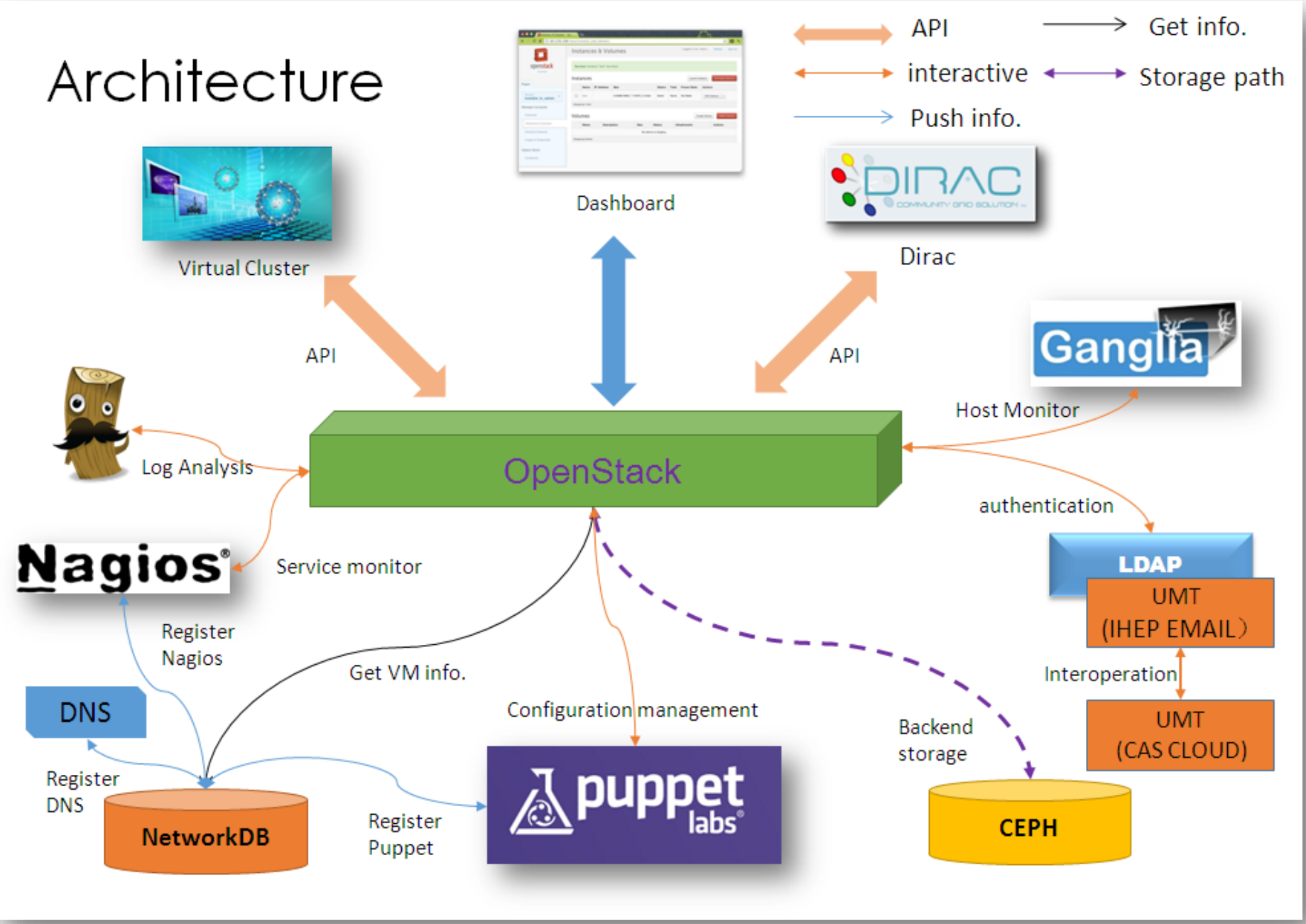
- Cores 57%
- MEM 58%
- Storage 88%

VM CPU real usage about 5-10%



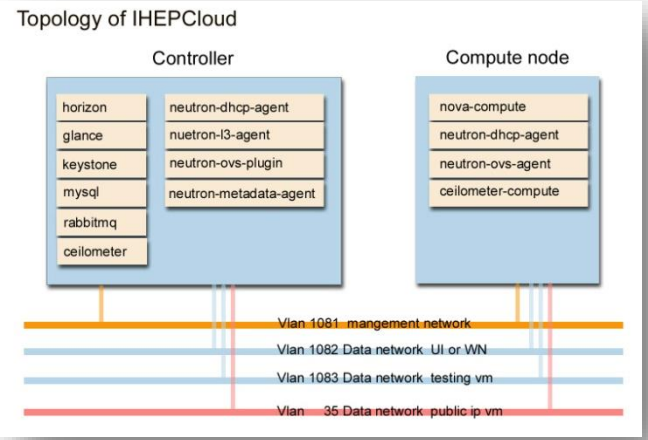


Architecture



IHEPCloud Components

- VM management: Openstack
  - open source cloud management system
  - most popular
- Configuration management tool: Puppet
  - create VM image
  - manage applications in VM
  - keep the consistency of VM and computing environment
- Authentication
  - IHEP Email account and password
  - AFS authentication for UI node
- Network management
  - centric NETDB record MAC, IP,
  - hostname, user, …
  - each VM has a hostname, *.v.ihep.ac.cn
  - network traffic accounting
- External storage
  - Currently, images and instances are stored in local disk
  - evaluating CEPH to support GLANCE, NOVA and Cinder



NETDB

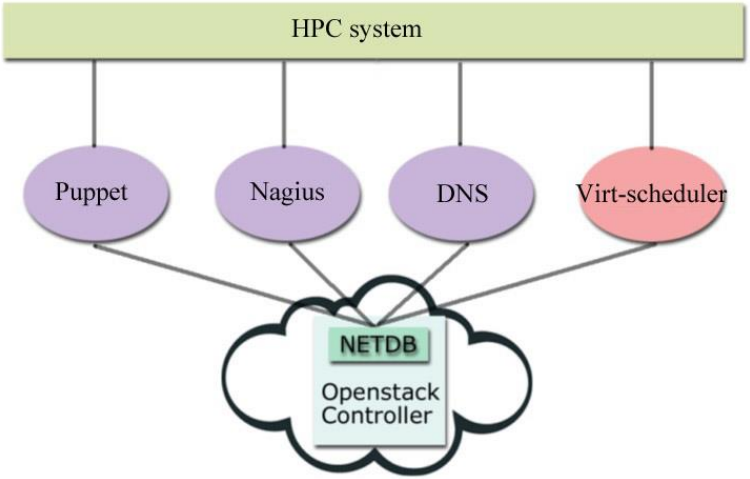
NETDB -- Environment configuration

VM Domain

vm + ip segment address + host address + .v.ihep.ac.cn

vm083011.v.ihep.ac.cn 192.168.83.11

- Scan DB to get any new VM information(ip, mac) and generate the VM domain for puppet, Nagios and DNS server
- Provide Query API for other modules



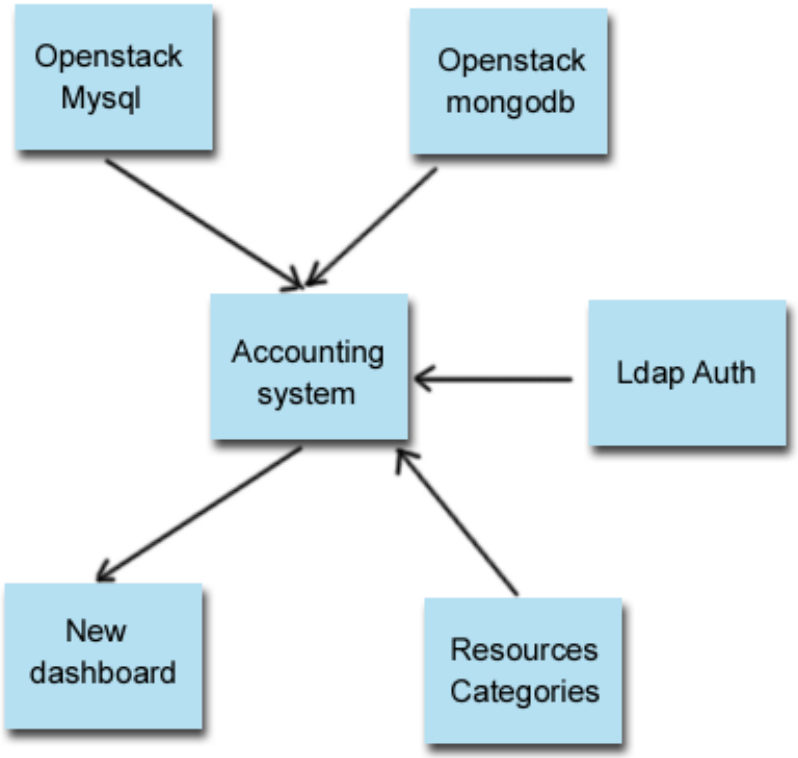
NETDB: dynamical database and an interface of controller

Details of IHEPCloud

- Network design
  - neutron + OVS, Vlan mode, Only one NIC
- Unified Authentication
  - Keystone+Ldap
  - IHEP mail account/password
- Simplify VM customization
  - Simplify the user selection, Conceal some labels and add VM type selection
  - Network type selection be included in VM type selection
- User Access control
  - Not allow non-privilege users to configure network and upload VM images
  - Shared network and images (created by Privileged users) to all users
  - Quota: One tenant one user

New feature in Dashboard

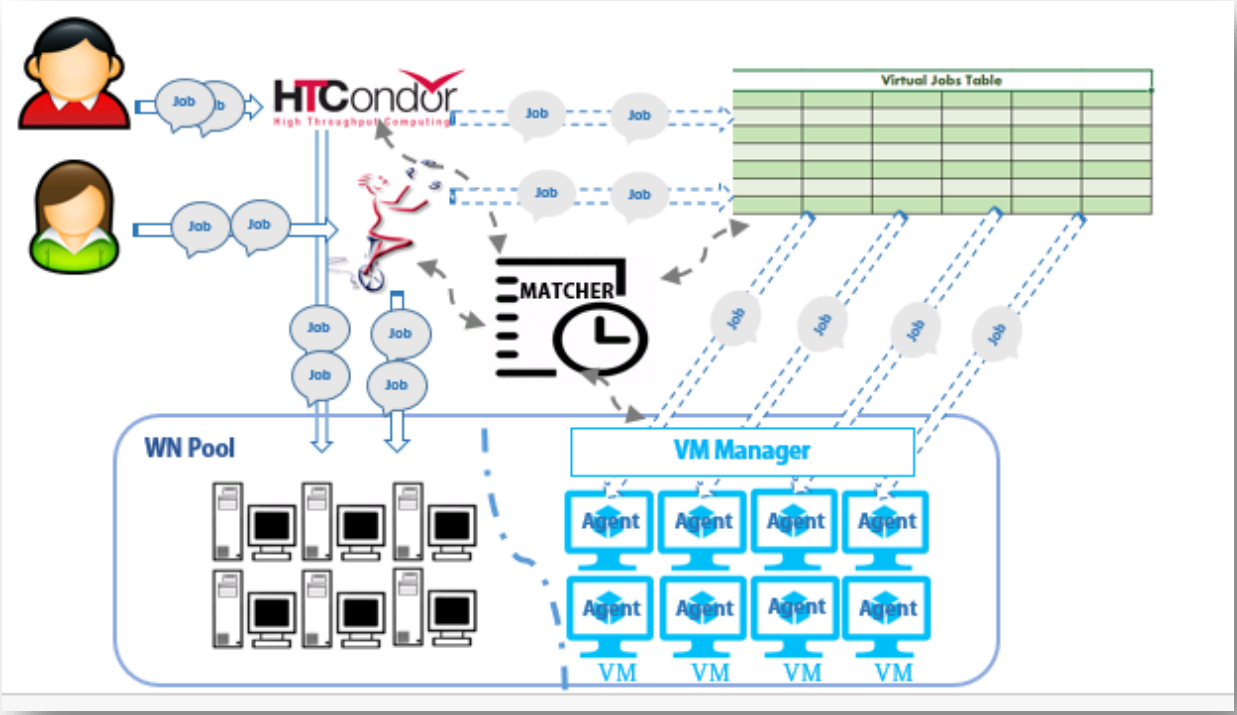
- **Introduce bonus**
  - Periodically scan Mysql database to find VM information
  - Periodically scan mongodb to find VM real usage information
  - Record VM resources usage, owner and running time
  - Estimate the resource contribution of each experiment



- **Reword rules**
  - How much resources contribution per experiment?
  - Bonus can be consumed when using others' resource
  - Bonus can be reworded if you shared resources for others

Virtualized Computing Cluster

- KVM+OpenStack+batch system: Integrated Openstack and HTCondor/PBS to provide computing service based on IHEPCloud
- Virtual cluster and physical cluster work together
- Resources for physical machine and virtual machine can be switched dynamically

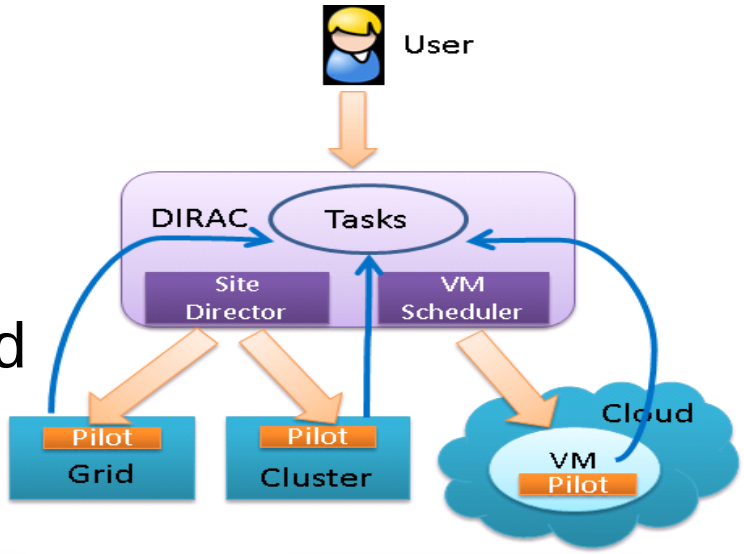


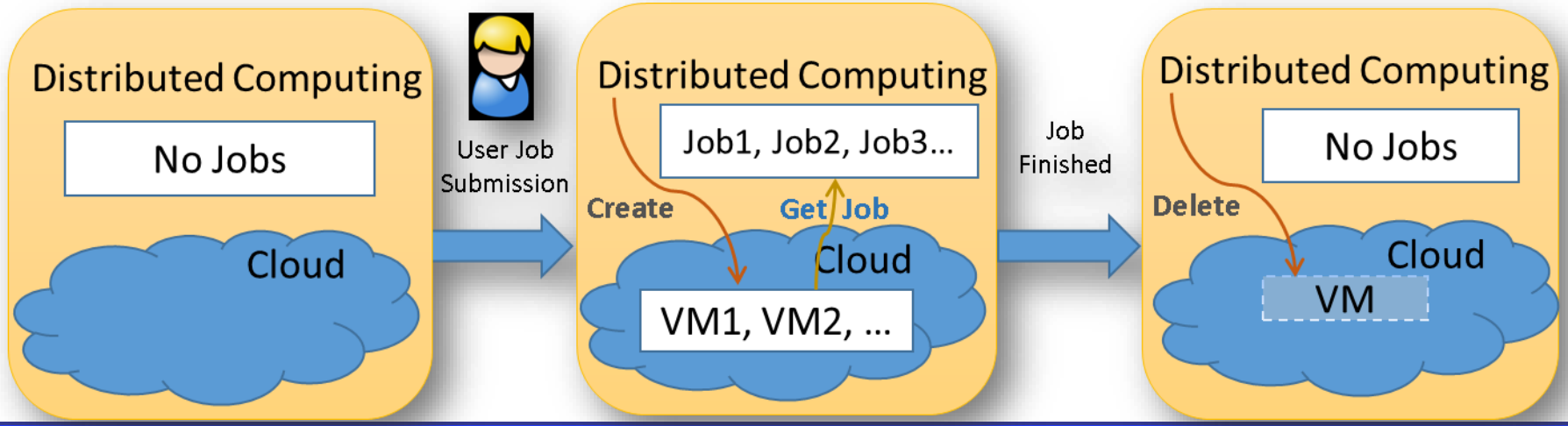
Current status

- Completed the testbed
- Submitted hundreds of test jobs
- Next steps:
  - Need more scale testing
  - Applied to other experiments like JUNO,DAYABAY,YBJ and so on.
  - Provide online service in this year.

Distributed computing system

- Distributed computing has integrated cloud resources based on pilot schema, implementing dynamic scheduling
- Cloud resources can be shrunk and extended dynamically according to job requirements





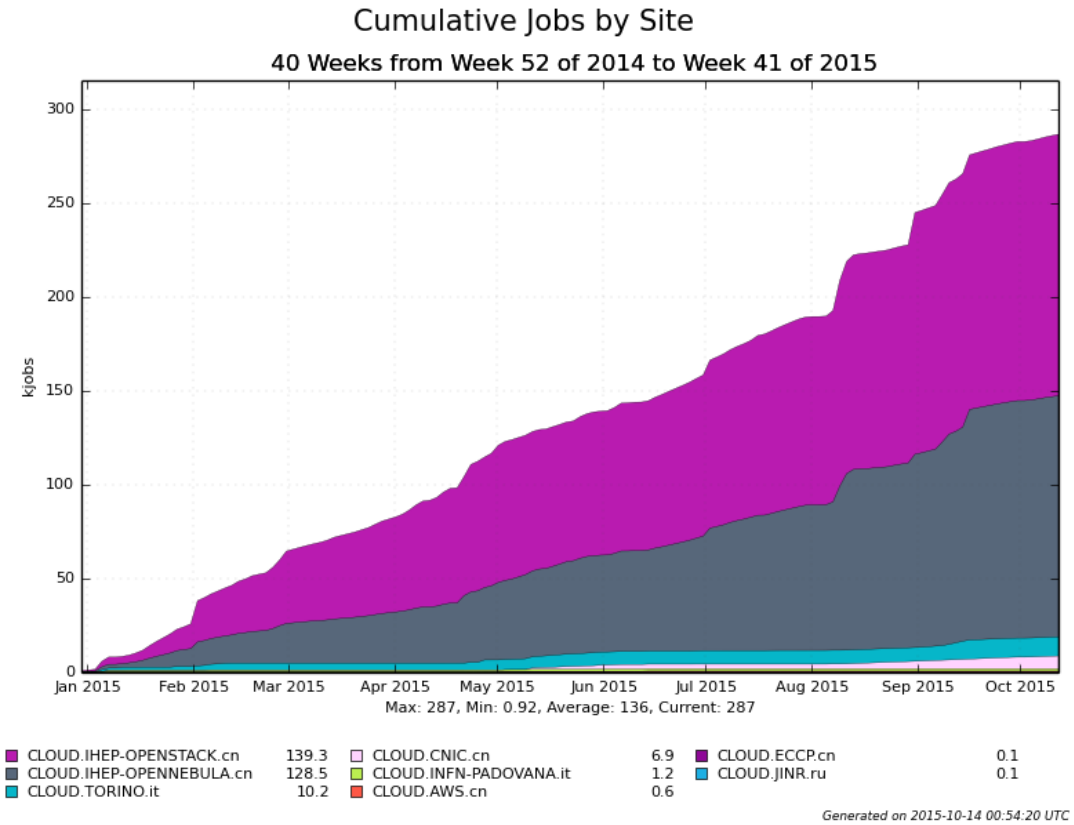
Cloud sites

- 6 cloud sites from Torino, JINR, CNIC, INFN and IHEP have been set up and connected to distributed computing system
  - About 455 CPU cores, 608GB Memory, 10.65TB disk

Site	CloudType	vCPU	Memory GB	Storage TB
Cloud.IHEP-Openstack.cn	Openstack	120	144	2.5
Cloud.IHEP.OpenNebular.cn	OpenNebular	120	144	2.5
Cloud.Torino.it	OpenNebular	60	60	5.1
Cloud.JINR.ru	OpenNebular	5	10	0.3
Cloud.CNIC.cn	Openstack	50	50	0.05
Cloud.INFI-PADOVANA.it	Openstack	100	200	0.2

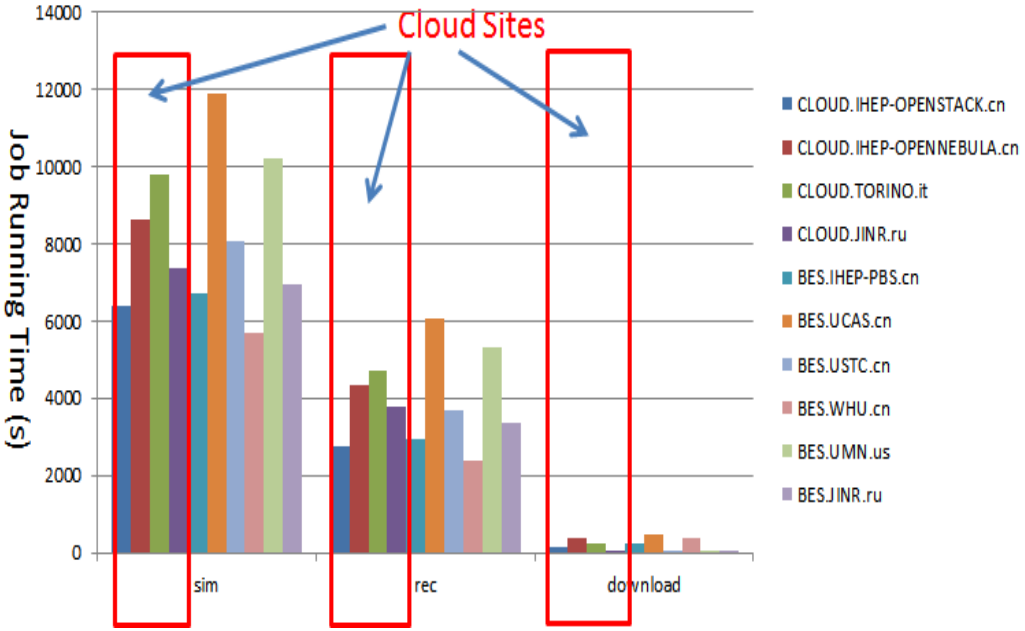
Cloud in production

- Cloud is put into production since end of 2014
- In 2015 until now, about 200K jobs have been done



Performance and Physics validation

- Performance tests has shown that running time in the cloud sites are comparable with other production sites
  - Simulation, Reconstruction, Download random trigger data
- Physics validation has proved that physics results are highly consistent between clusters and cloud sites



Future work

- Upgrade IHEPCloud to Kilo version
- Optimize the monitoring of virtual machine
- Introduce rewarding bonus to each experiment based on the resource contribution
- Push Virtualized computing cluster on line in this year
- Consider to introduce CELL in Cloud deployment to manage the distributed computing resource

Thanks & question?