Elastic Computing Resource Management Based on HTCondor

Qiulan Huang

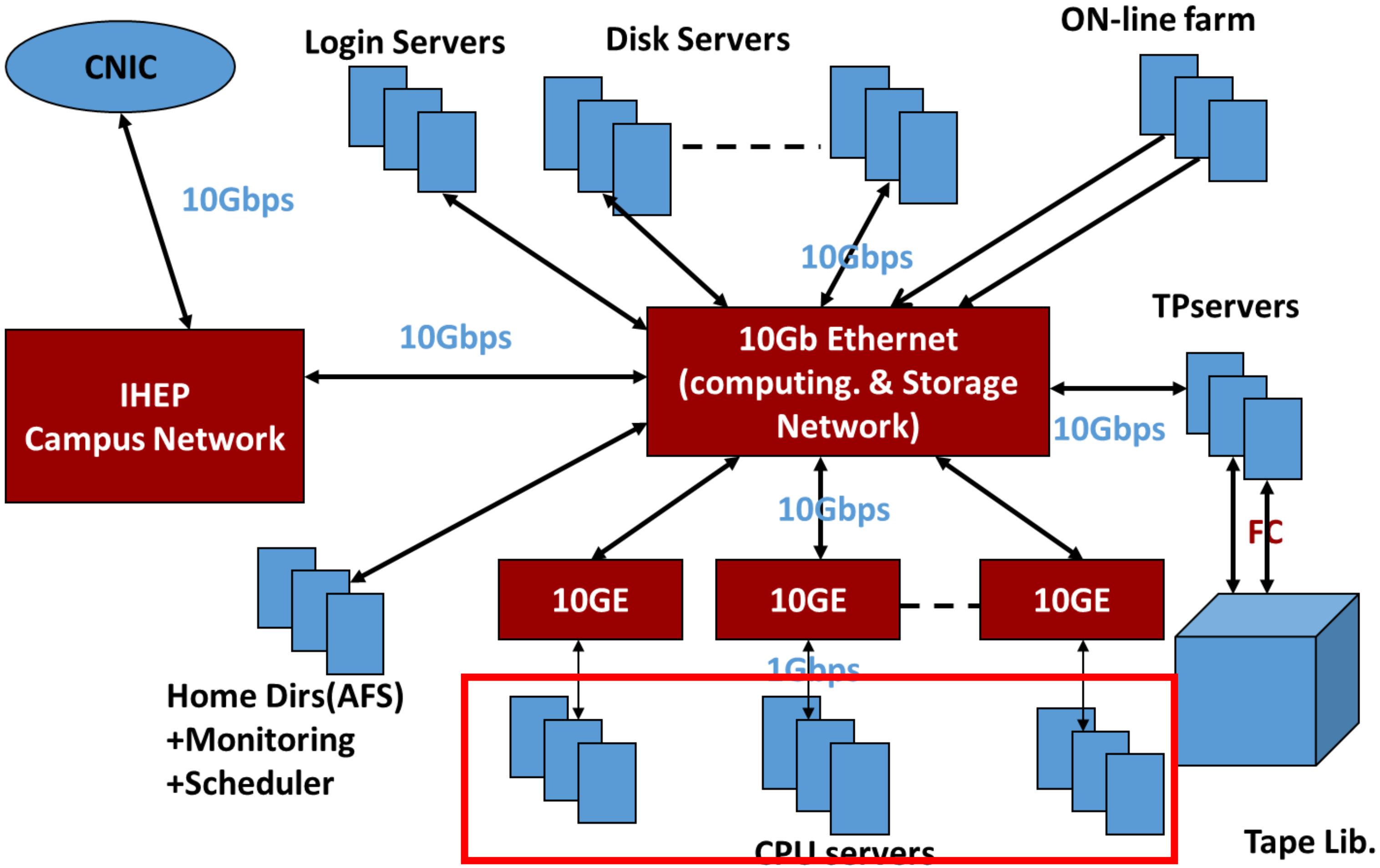
On behalf of IHEP Computer Center

CHEP2016 San Francisco,2016-10-12

Outline

- Background
- VCondor: Elastic Computing Resource Management System Based on HTCondor
- Running Status and Performance Analysis in IHEPCloud
- Summary

Traditional Architecture of HEP Computing Cluster



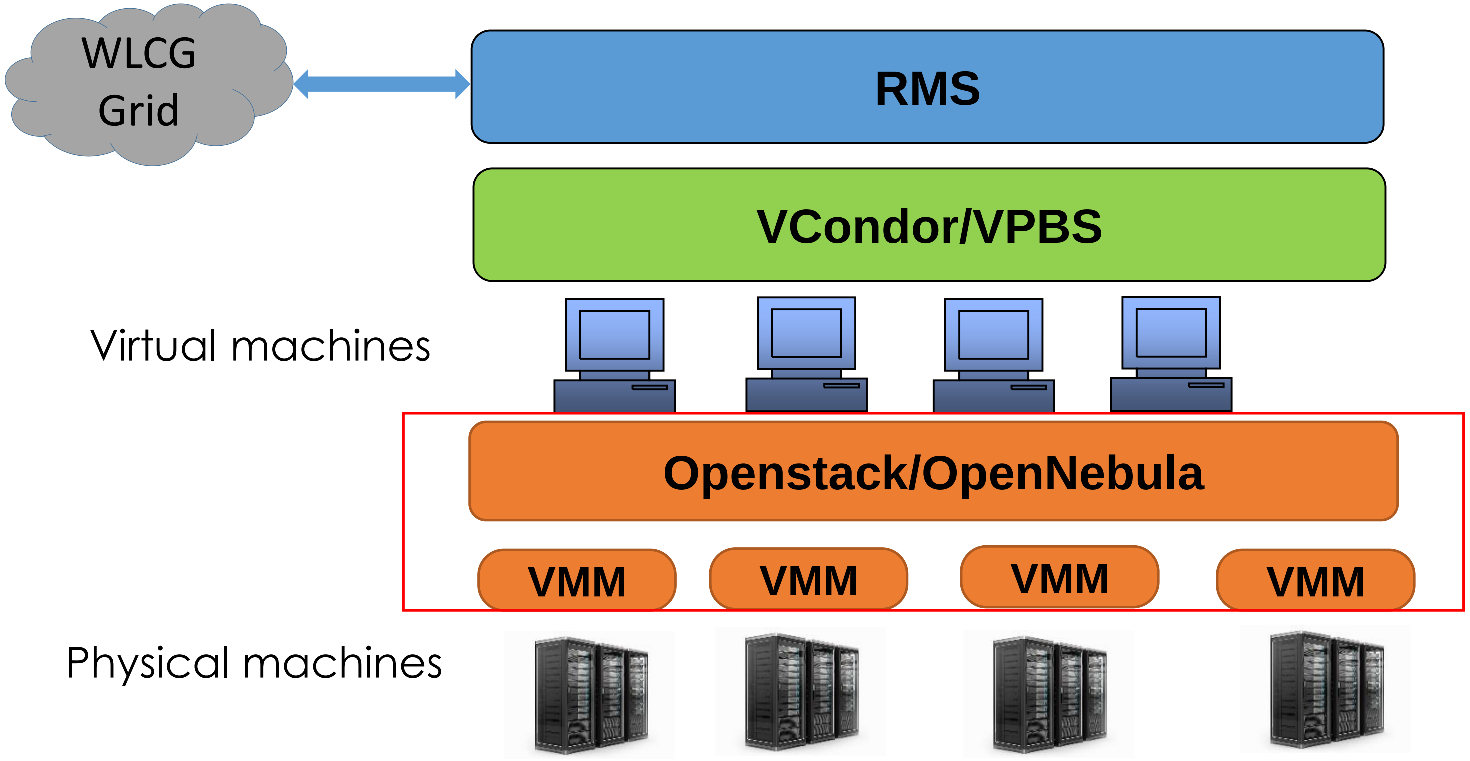
- Cluster Resource Provided by Independent HEP Application
- Multiple Independent Application vs Multiple Independent Computing Queue

Existing Trouble in Traditional HEP Cluster Computing

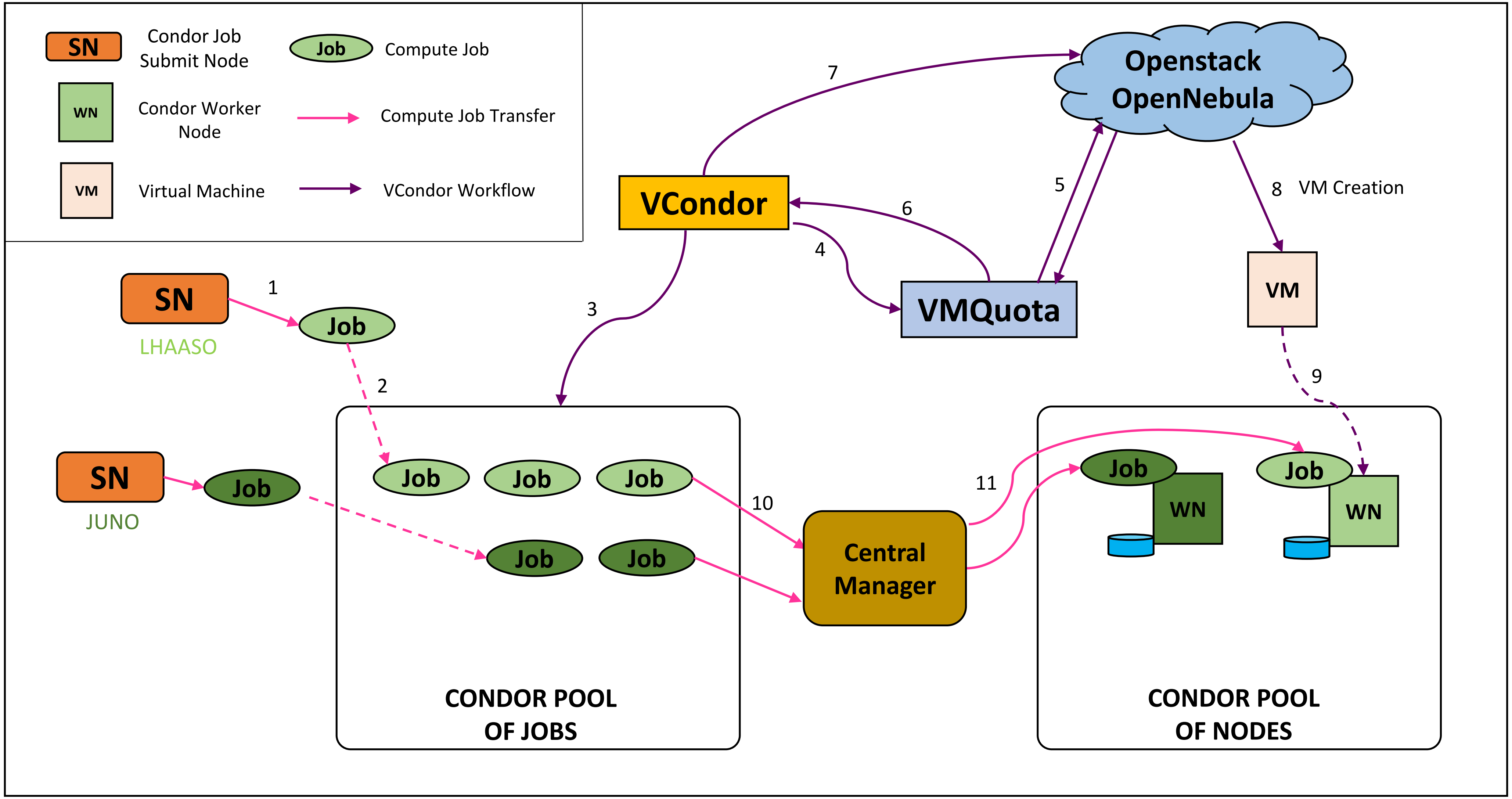
- Computing queues don’t share resource
  - Different queues have different peak times of resource usage
  - Massive jobs with a little resource or vice versa
  - The overall resource utilization is low
- Virtual Computing Cluster
  - Elastic computing resource management
  - Elastic resource pool, scale up and down dynamically according to HEP Application jobs to improve resource utilization

Run jobs on Virtual Computing Cluster

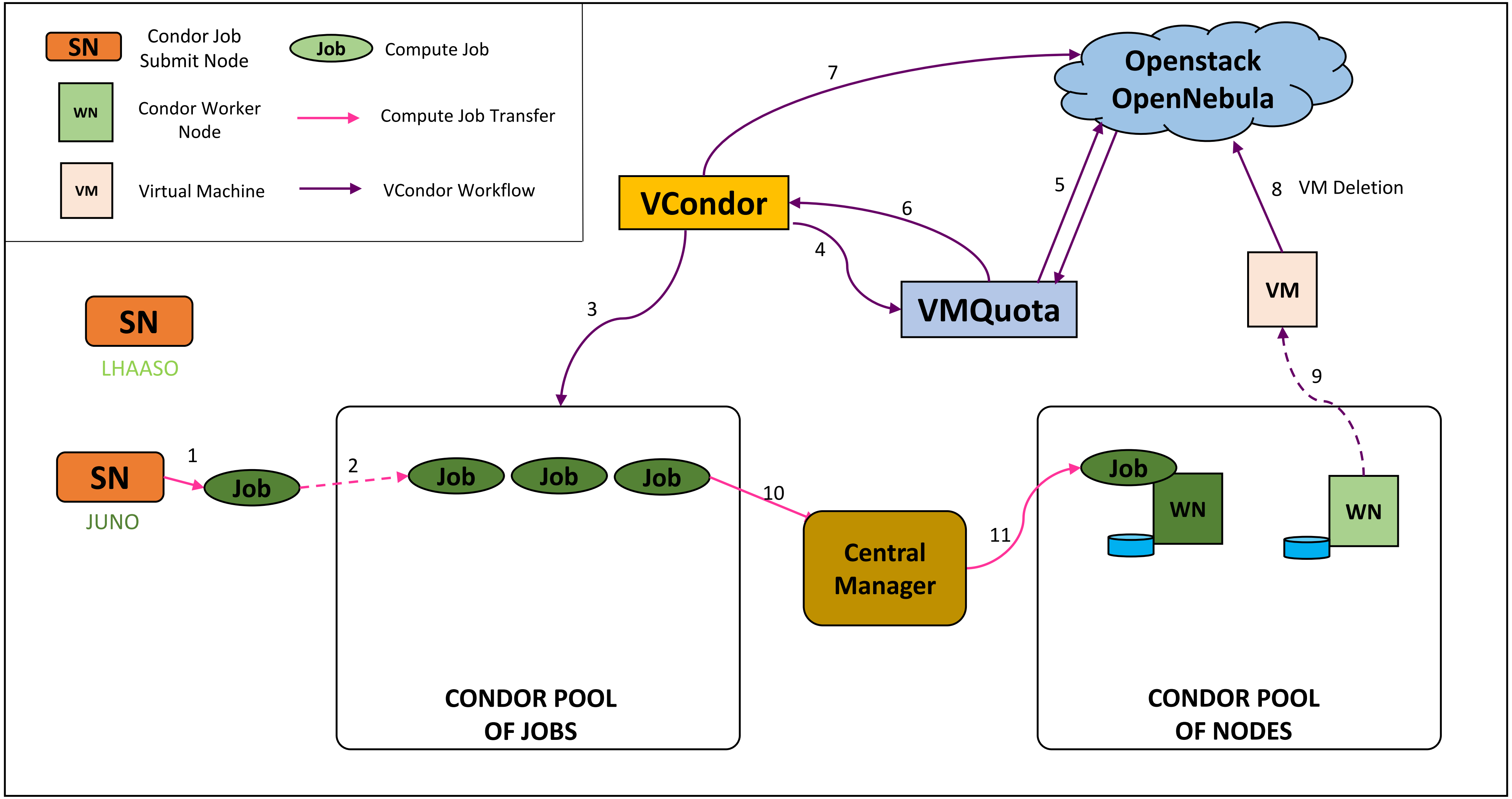
- A virtual layer is constructed between the physical machines and the RMS (resource management system such as HTCondor or PBS)
- In the case of free resources in the pool, a job queue can use more than it owns



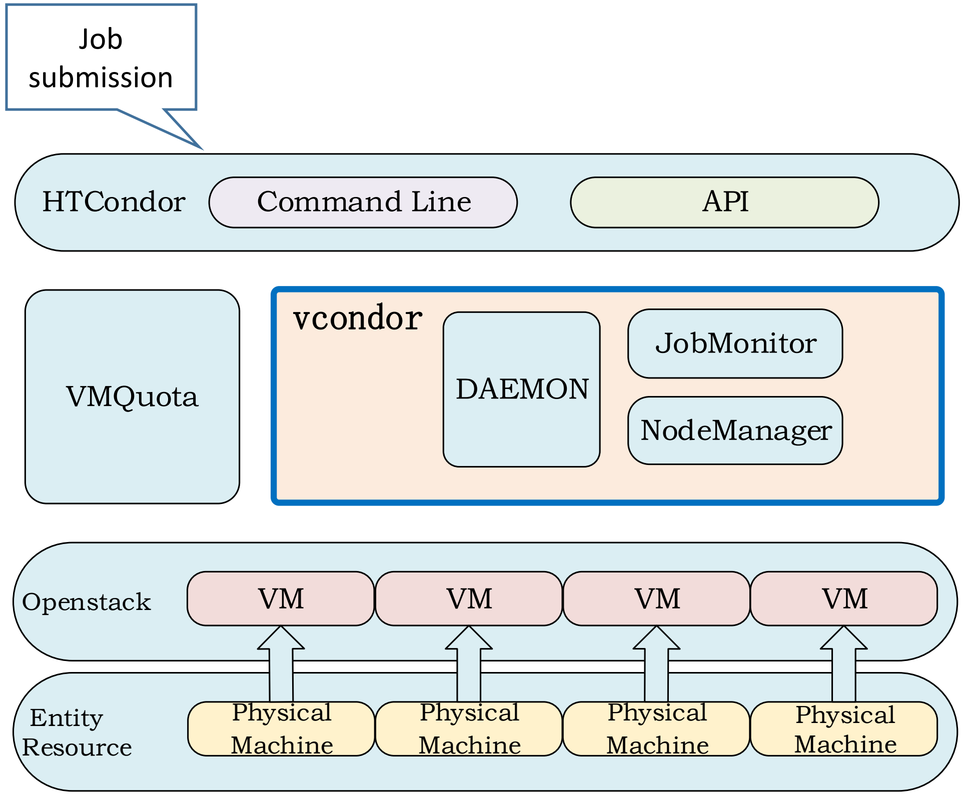
Resource pool expansion



Resource pool shrink



VCondor's Components



https://github.com/hep-gnu/VCondor.git

(1)vcondor:

I. JobMonitor: query and record job information and queue length changes

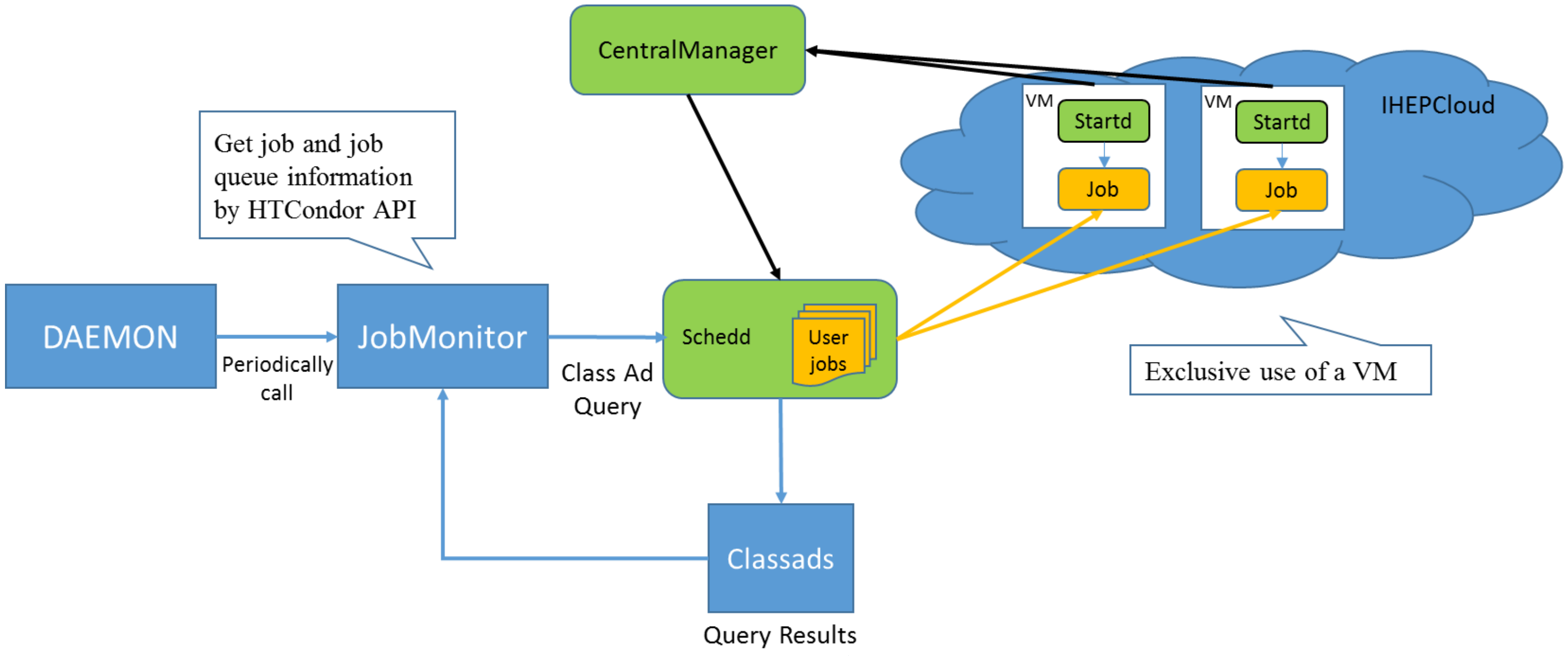
II. NodeManager: use openstack api to create and destroy virtual machines

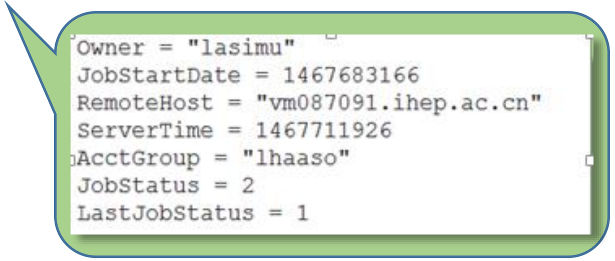
III. DAEMON: Main module, periodically executed

(2)VMQuota:

Computing resource share management system

(1) JobMonitor: Analyze resource requirements





```
Owner = "lasimu"
JobStartDate = 1467683166
RemoteHost = "vm087091.ihep.ac.cn"
ServerTime = 1467711926
AcctGroup = "lhaaso"
JobStatus = 2
LastJobStatus = 1
```

IHEPCloud:

- Virtual computing cluster
- 28 computing nodes,672 CPU cores
- HTCondor version:8.2.5
- Provide support for IHEP experiment such as LHAASO, JUNO, BES, CEPC accelerator design

(2)NodeManager: VM creation and deletion

- ***VM Class***

  —Store vm properties, such as hostname, job queue name, idle or busy

- ***Icluster Abstract Class***

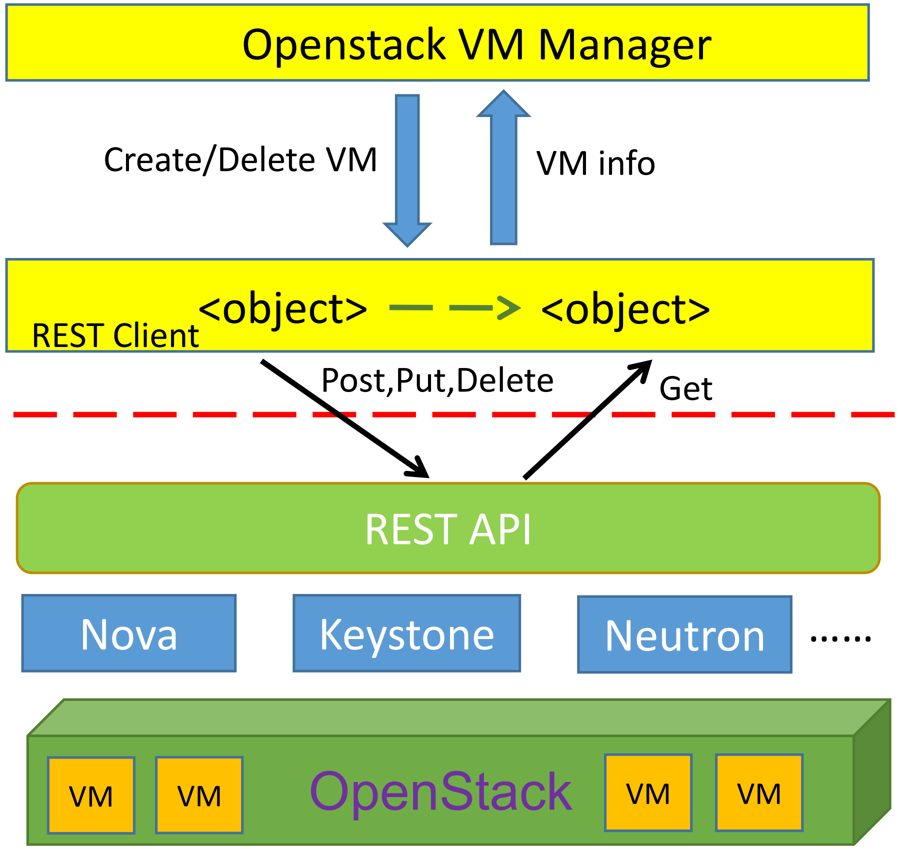
  —Implementation of the interface support a variety of cloud computing platforms, like Openstack, OpenNebula, AWS EC2

- ***Openstack Cluster Class***

  —Communicate with the controller Nova through a REST client to create or destroy VMs on Openstack

- ***ResourcePool Class***

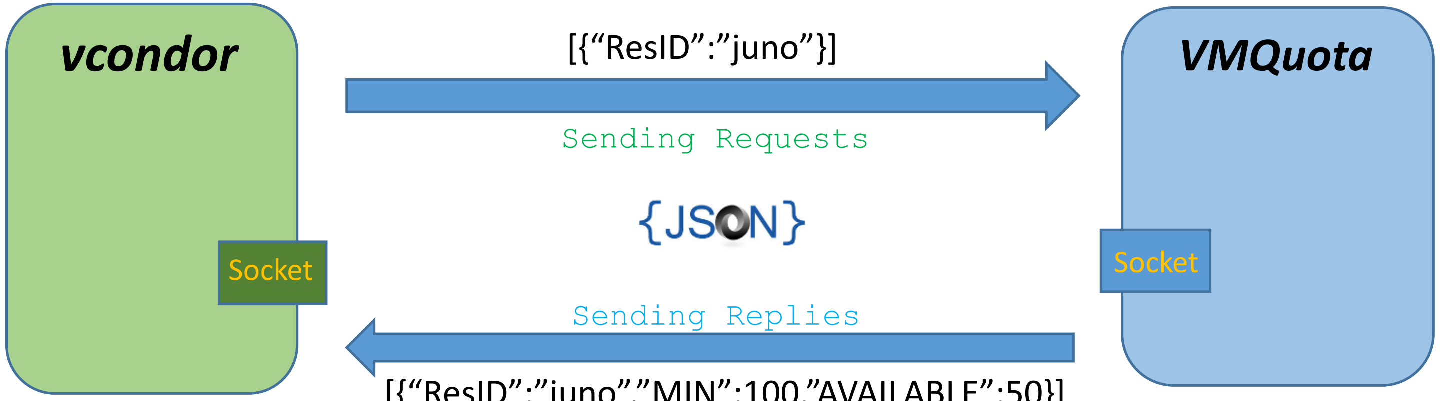
  —Provide support for resource management across multiple cloud platforms



(3)VMQuota: Resource share management

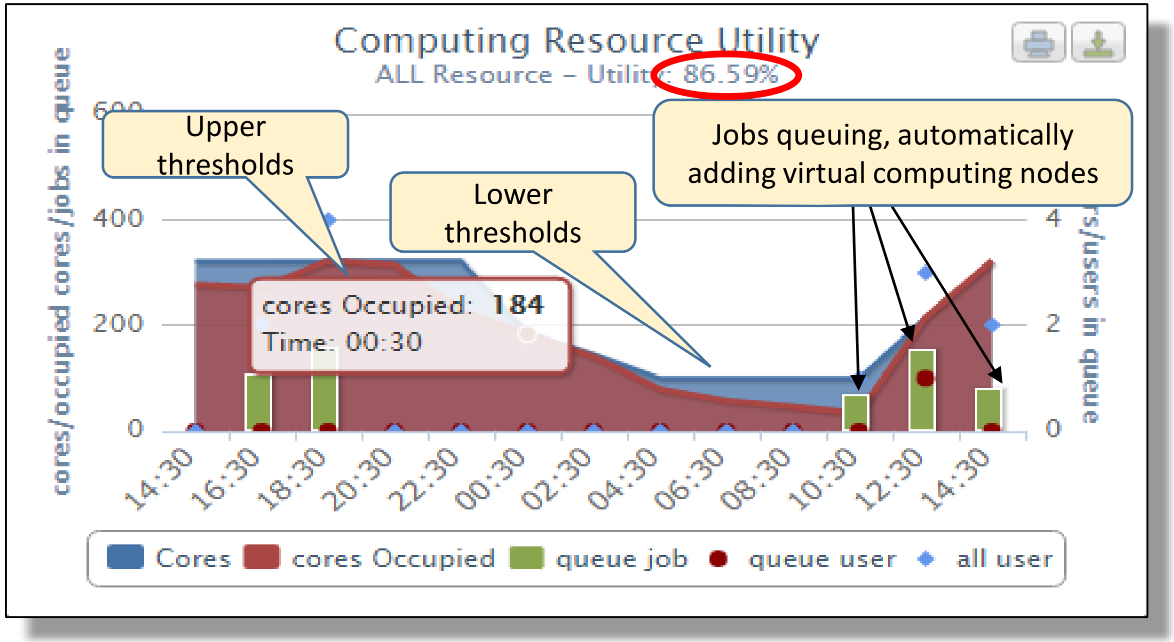
- Set up virtual computing queues for each HEP application, like LHAASO and JUNO
- Manage the upper thresholds, lower thresholds, and resource reservation time for each queue

Queue	Lower thresholds	Upper thresholds	Available resource	Resource reservation time(seconds)
LHAASO	100	400	200	600
JUNO	100	300	200	600



VMQuota

Dynamic Scheduling Effect



Upper thresholds

Lower thresholds

Jobs queuing, automatically adding virtual computing nodes

LHAASO Resource Pool:

Automatically Scale up and down on demand

VCondor:How to use

- Download VCondor from https://github.com/hep-gnu/VCondor.git
- Make sure HTCondor and Openstack or OpenNebula are well configured
- Setup a VM Image with Condor installed
- Setup a VM Template with Image in the above
- The VCondor configuration file allows you to configure most of its functionality, open it up to get a usable installation
- Start VCondor and submit jobs, resource pool scale up and down dynamically

Summary

- VCondor enables elastic resource management
  - ➢ Has begun trial operation in IHEPCloud
  - ➢ Current experiment： LHAASO, JUNO
  - ➢ Provide support for HEP application resource sharing plan
- Future
  - ➢ Achieve preemptive scheduling based on HTCondor
  - ➢ Improve scheduling algorithm further：

Welcome to Contact us !
E-mail: chyd@ihep.ac.cn lihaibo@ihep.ac.cn
chengzj@ihep.ac.cn