Dynamic Scheduling Strategy of HTCondor at IHEP

Shi, Jingyan (shijy@ihep.ac.cn)

On behalf of scheduling group of

Computing Center, IHEP

Content

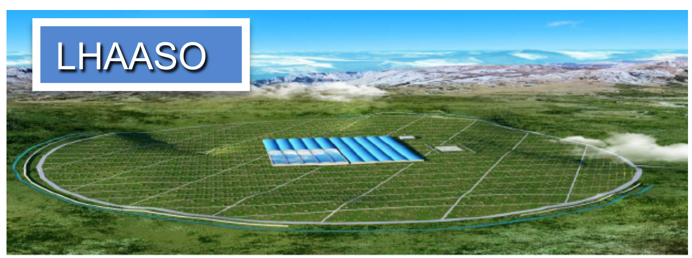
HEP Experiments at IHEP

IHEP: Institute of High Energy Physics

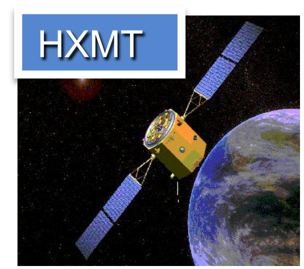
BESIII (Beijing Spectrometer III at BEPCII)
100TB raw data/year *10 years

DYB (Daya Bay Reactor Neutrino Experiment)
200TB/year* 9 years

JUNO (Jiangmen Underground Neutrino Observatory)
2PB/year*30 years



Large High Altitude Air Shower Observatory
1.2PB/year *10 years



Hard X-Ray Moderate Telescope

HTCondor Cluster Resource

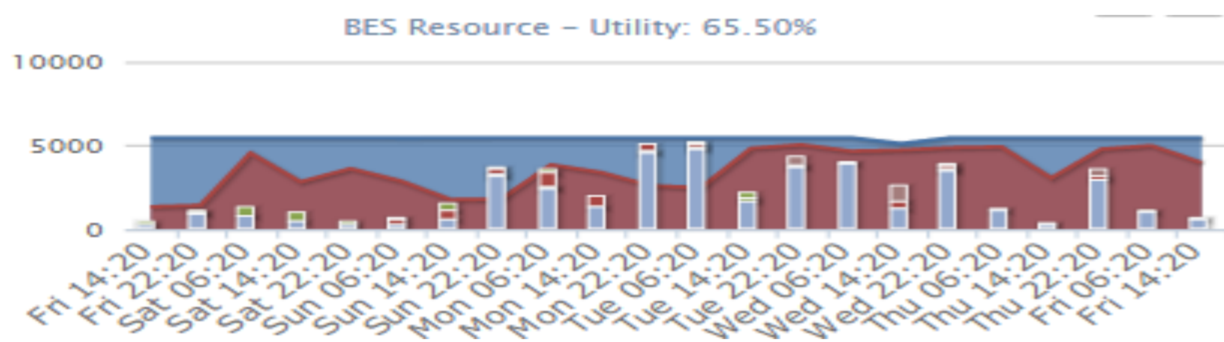
- Local cluster
  - ~10,500 CPU cores
  - Most are single core, series job slots
  - Managed by PBS for 10 years
    - Bottleneck: low scheduling performance
      - Large amount of jobs: 20,000 jobs in queues
      - Large scale: over 10,000 job slots
  - Migrated to HTCondor step by step with risk control
    - Jan, 2015 : ~ 1,100 CPU cores
    - May, 2016: ~ 3,500 CPU cores
    - Dec, 2016: ~ 11,000 CPU cores

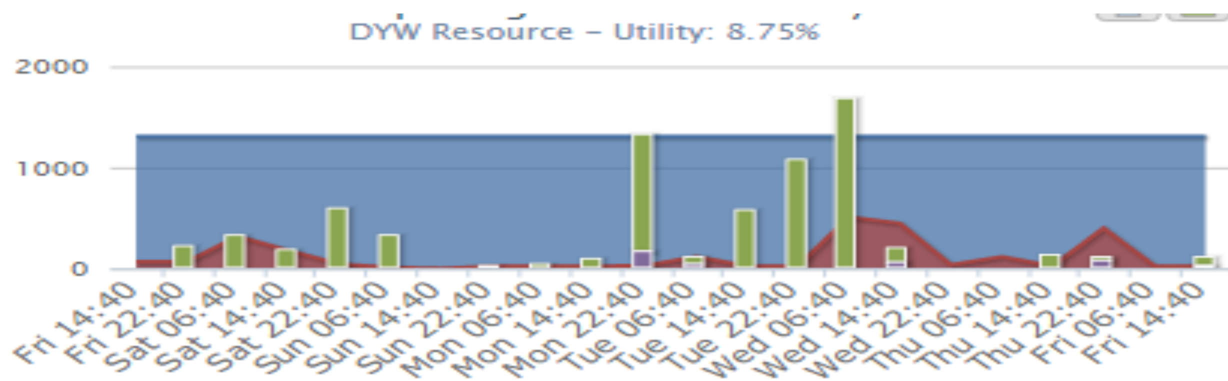
Content

Resource Managed by PBS Cluster

- Several experiments supported
  - BES, Daya Bay, Juno, Lhaaso, HXMT etc.
  - Resources are funded and dedicated for different experiments
  - No resource sharing among the experiments
  - 55 jobs queues with group permission limits configured
- Low resource utility
  - Coexistence of busy queues and free resources

Busy Queue and Free Resource at PBS



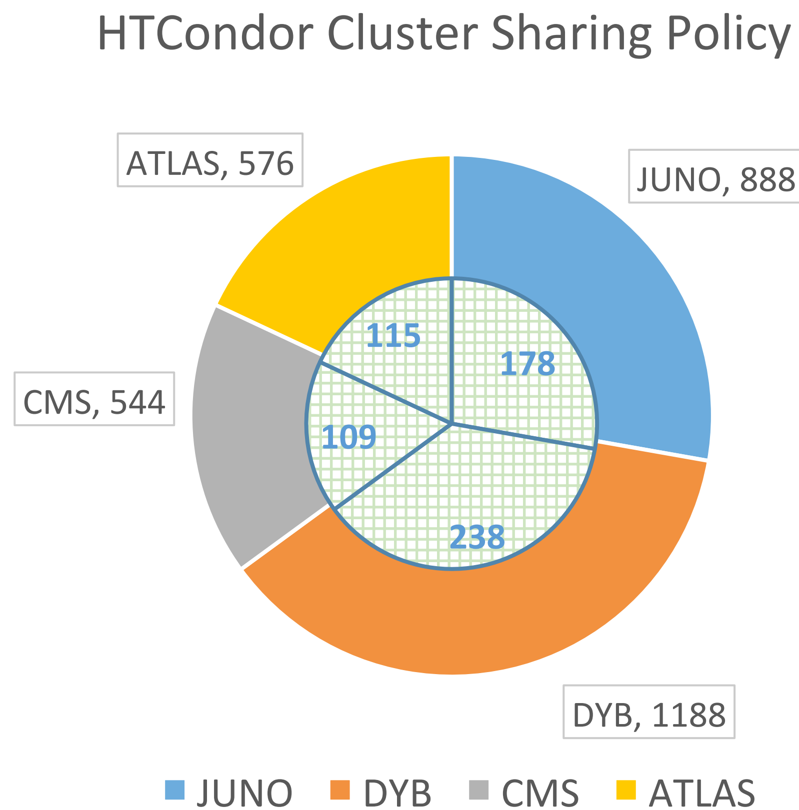


Basic Idea of the HTCondor Scheduling Strategy

- Resource sharing
  - Break the resource separation
  - Busy exp. can take more resources from that of the free exp.
- Fairness guarantee
  - Peak computing requirements from different exp. usually happened at different time periods
  - Jobs from free exp. have higher priority than the jobs from busy exp.
  - The more resources the exp. shares, the more its jobs can be scheduled

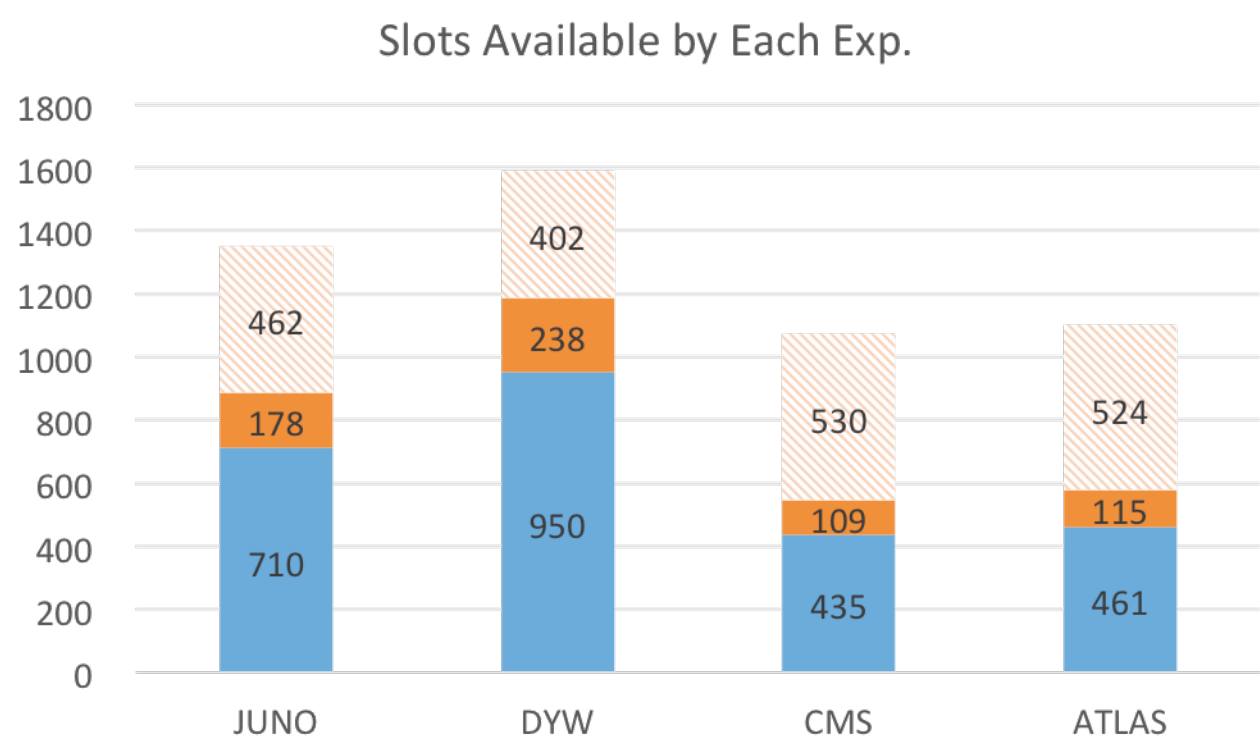
Resource Sharing at HTCondor

- Based on job slots (mainly CPU cores)
- As first step, part of resources are contributed to be shared
- Some dedicated resources are kept by experiments own
  - Only run jobs from owner exp.
- Sharing resource pool
  - Sharing resources contributed by all experiments
  - Sharing slots can be dispatched to all jobs
  - At least 20% slots of each exp. are shared
  - encourage exp. to share more resources



The dedicated and sharing slots of different groups

Resource Sharing with HTCondor



The dedicated, sharing and max allocable slots for each exp.
(In May, 2016)

Fairness and priority

- Scheduling preference
  - Jobs are preferred to run at dedicated slots owned by exp.
  - The shared slots are kept for busy experiments
- Group quota
  - Define linux group for each exp.
  - The initial group quota is set to the amount of real resources from exp.
  - Group quota can be exceeded if there are free slots in the sharing pool
- Group priority and User priority
  - Group priority is correlated to the group quota and the group slots occupancy
  - User priority is effective within the same group users

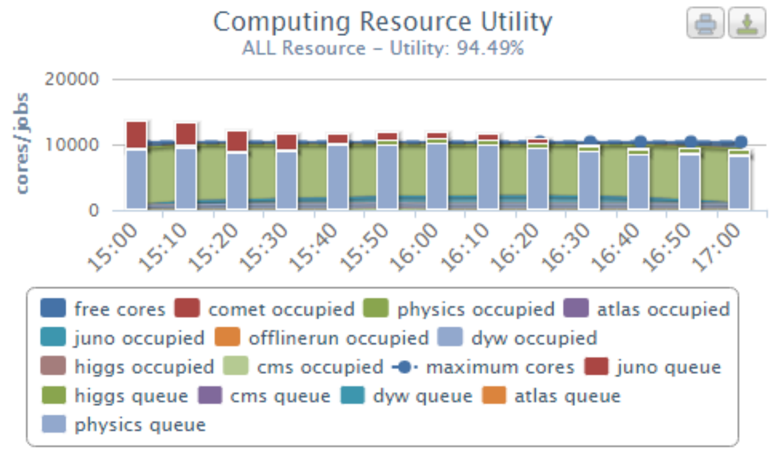
Content

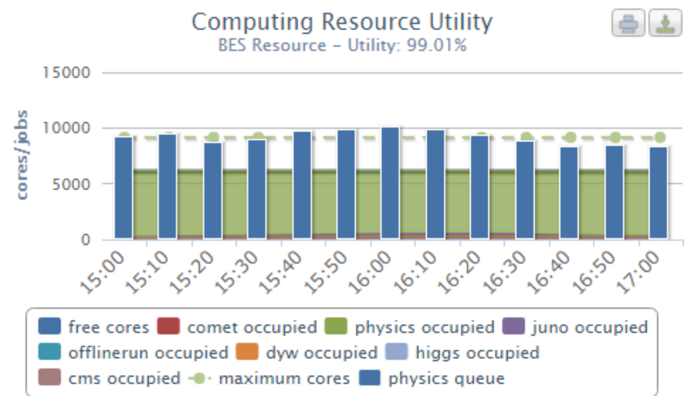
Current HTCondor Status

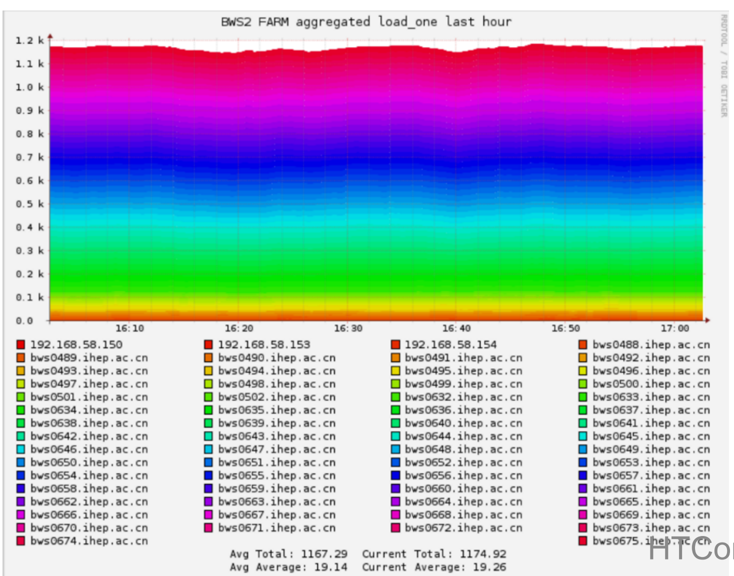
- Architecture
  - 28 submitting nodes
  - 2 scheduler machine (local cluster, virtual cluster)
  - 2 central manager (local cluster, virtual cluster)
  - ~ 10,000 physical CPU cores + an elastic number of virtual slots
- Jobs
  - Avg 100,000 jobs/day;
  - 60,000 jobs in queue at peak time
  - Serial single-core jobs

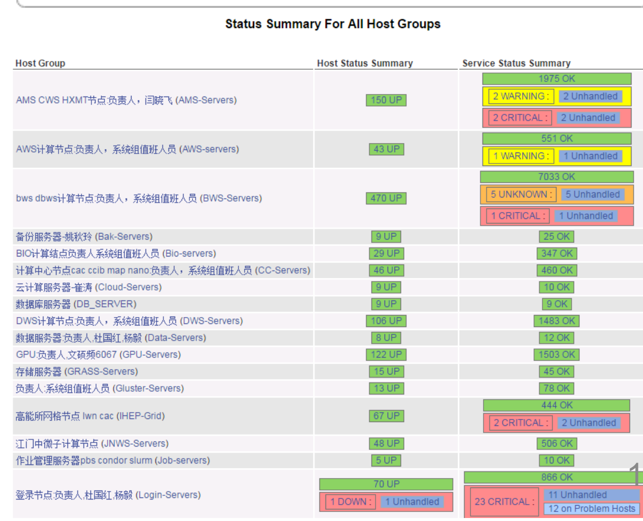
Job Monitoring

- Queuing and running statistics
  - The overall clusters
  - Each group/experiment
- The dedicated and sharing resource statistics
- Nagios and Ganglia





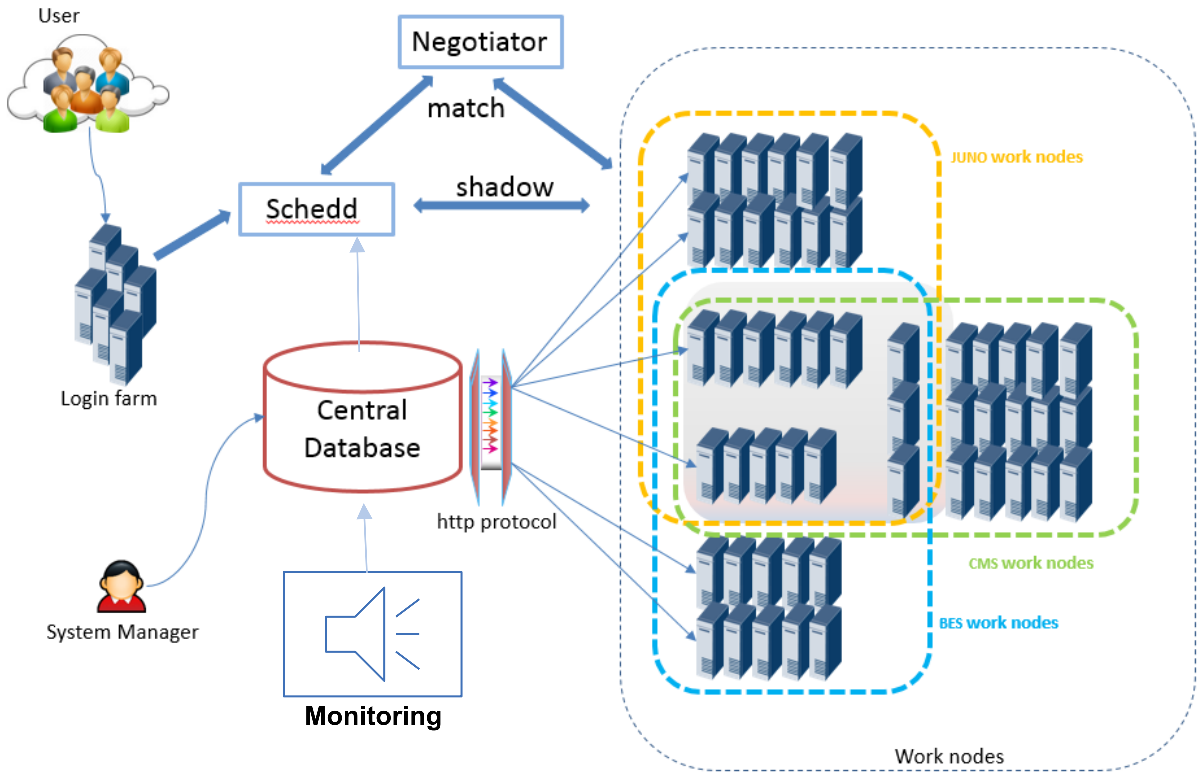



Status Summary For All Host Groups

Host Group	Host Status Summary	Service Status Summary
AMS CWS HXMT节点:负责人，闫晓飞 (AMS-Servers)	150 UP	1975 OK; 2 WARNING : 2 Unhandled; 2 CRITICAL : 2 Unhandled
AWS计算节点:负责人，系统组值班人员 (AWS-servers)	43 UP	551 OK; 1 WARNING : 1 Unhandled
bws dbws计算节点:负责人，系统组值班人员 (BWS-Servers)	470 UP	7033 OK; 5 UNKNOWN : 5 Unhandled; 1 CRITICAL : 1 Unhandled
备份服务器-姚秋玲 (Bak-Servers)	9 UP	25 OK
BIO计算结点负责人系统组值班人员 (Bio-servers)	29 UP	347 OK
计算中心节点cac ccib map nano:负责人，系统组值班人员 (CC-Servers)	46 UP	460 OK
云计算服务器-崔涛 (Cloud-Servers)	9 UP	10 OK
数据库服务器 (DB_SERVER)	9 UP	9 OK
DWS计算节点:负责人，系统组值班人员 (DWS-Servers)	106 UP	1483 OK
数据服务器:负责人,杜国红,杨毅 (Data-Servers)	8 UP	12 OK
GPU:负责人,文硕频6067 (GPU-Servers)	122 UP	1503 OK
存储服务器 (GRASS-Servers)	15 UP	45 OK
负责人:系统组值班人员 (Gluster-Servers)	13 UP	78 OK
高能所网格节点 lwn cac (IHEP-Grid)	67 UP	444 OK; 2 CRITICAL : 2 Unhandled
江门中微子计算节点 (JNWS-Servers)	48 UP	506 OK
作业管理服务器pbs condor slurm (Job-servers)	5 UP	10 OK
登录节点:负责人,杜国红,杨毅 (Login-Servers)	70 UP; 1 DOWN : 1 Unhandled	866 OK; 23 CRITICAL : 11 Unhandled, 12 on Problem Hosts

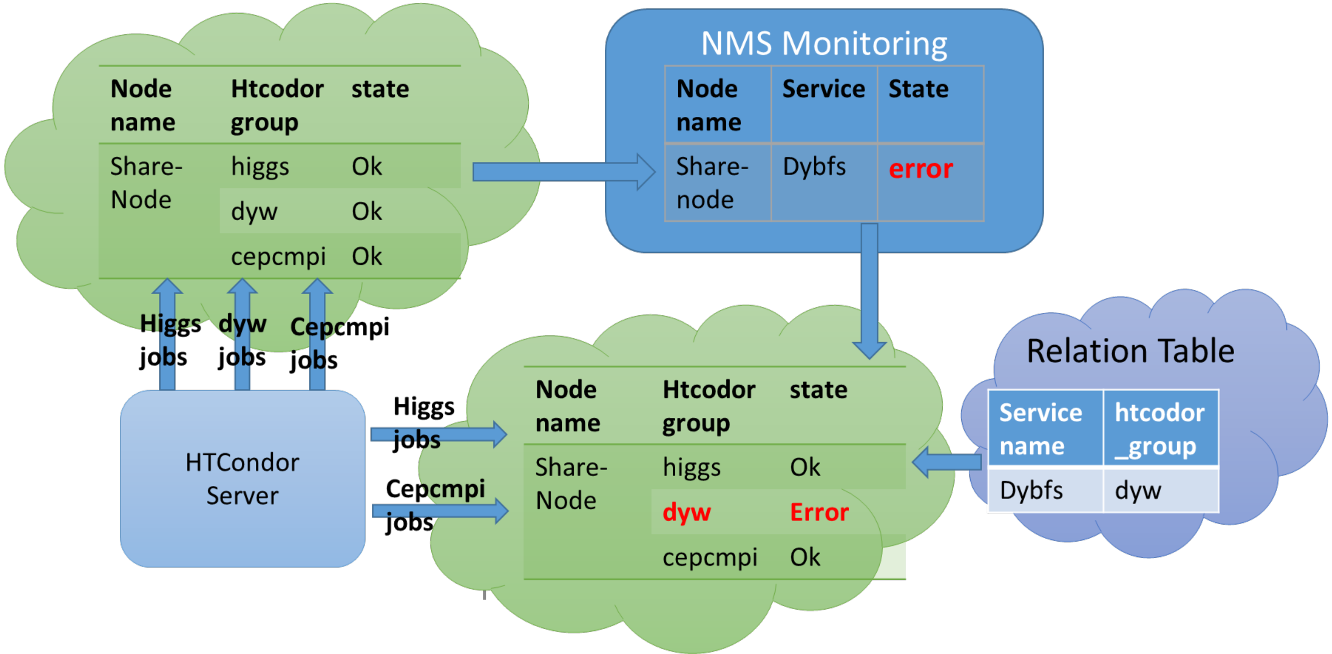
Central Controller

- Control of groups, users and resources
  - All information is collected into the Central Database
  - Necessary information is published to the relative services



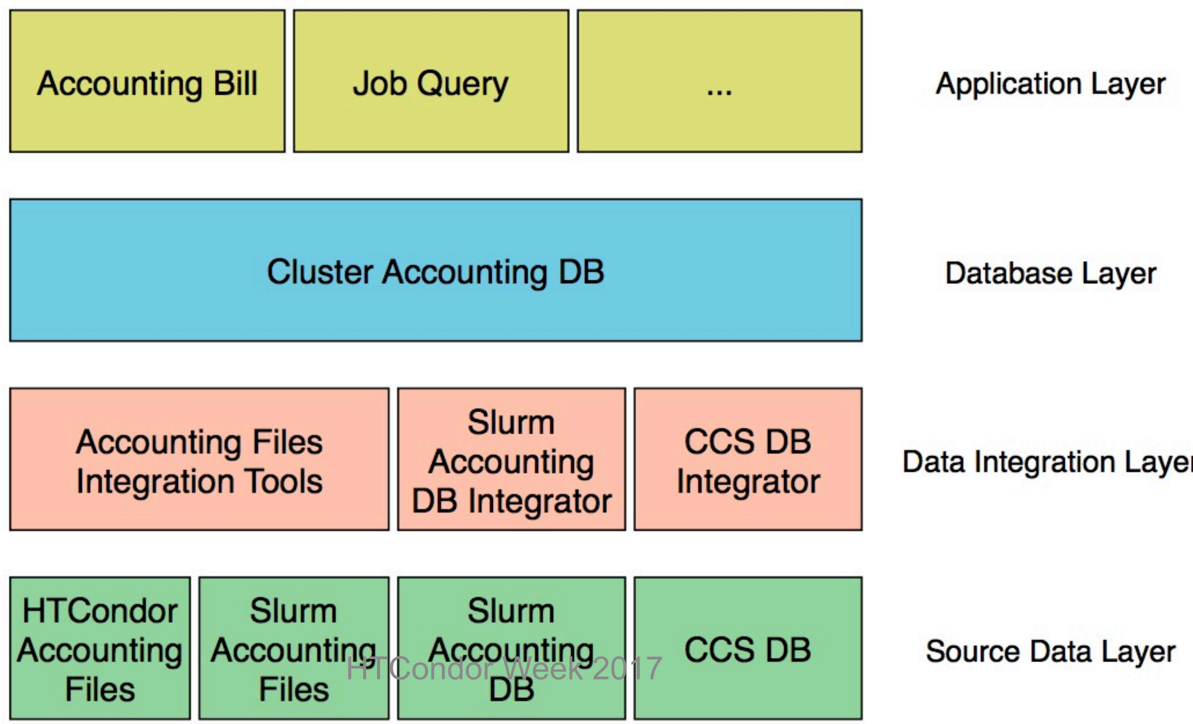
Error Detection and Recovery

- Workers' health status are collected and report to Central Controller
- Workers' attributes are self-updated automatically through the information published by Central Database
- No job will be scheduled to error worker



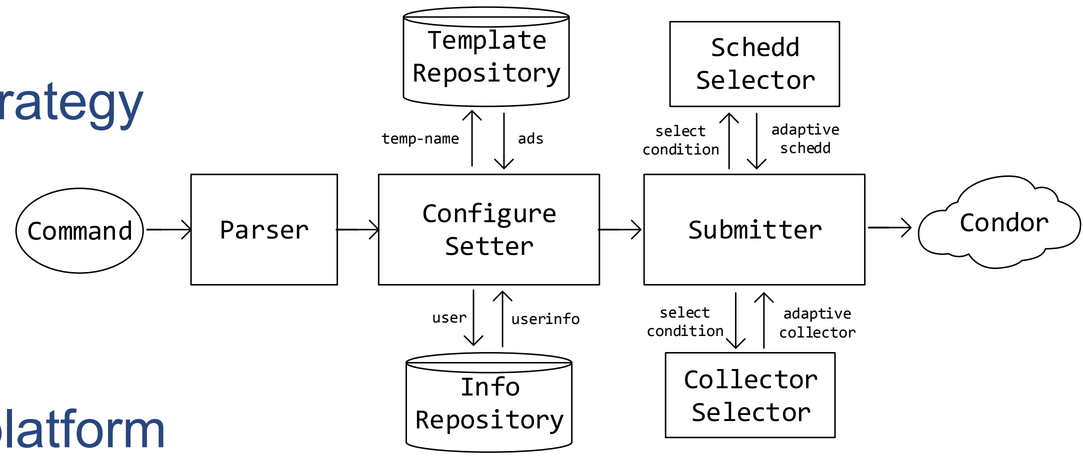
Global Accounting

- Detailed accounting to each group and each user
  - Accounting the contribution to other exp.
  - Accounting the extra resources occupied from other exp.
- Weighting slots with slow/fast CPU, Memory, Disk, etc.

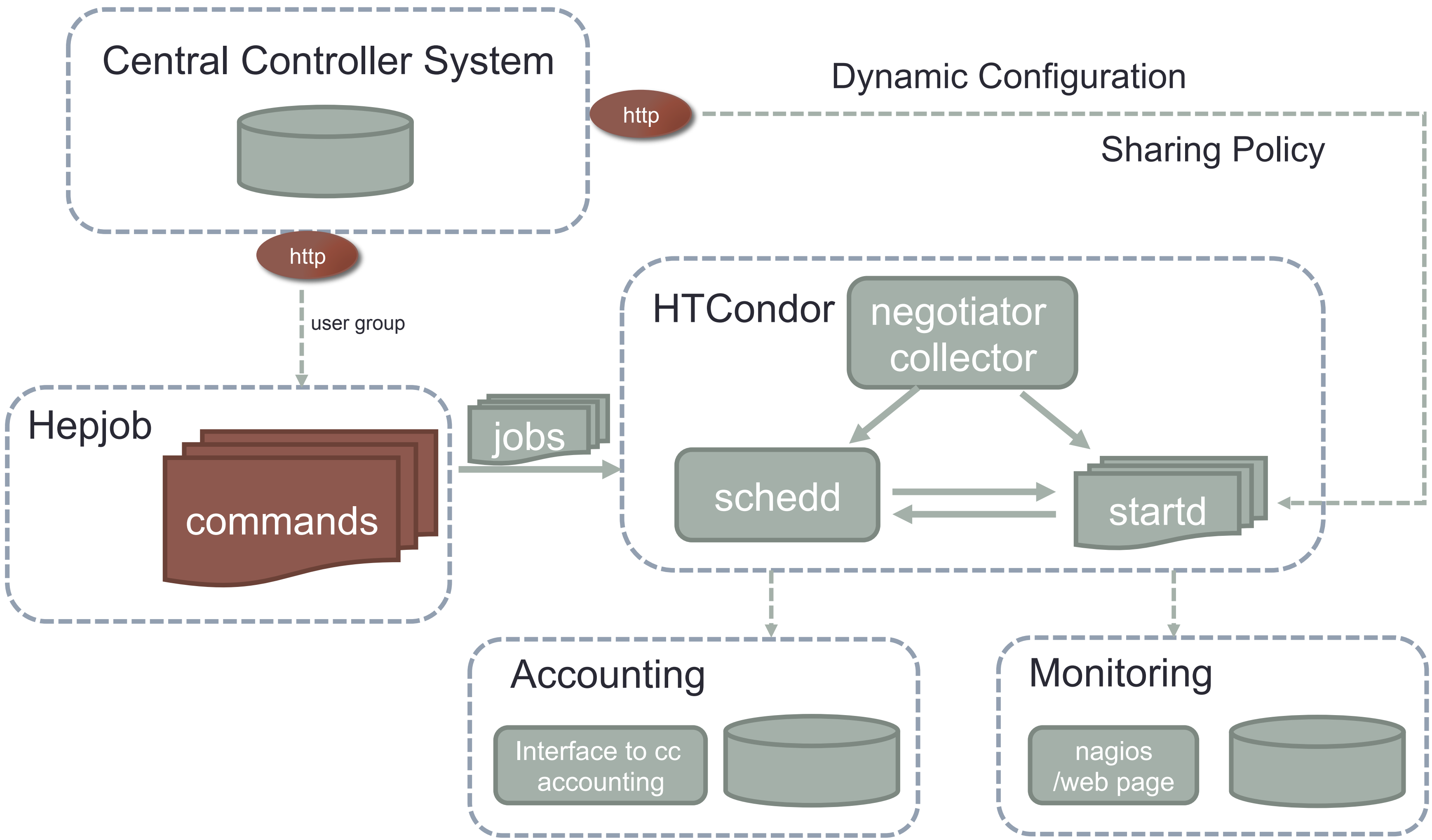


The Toolkit for user: hep_job

- Motivation
  - Smooth migration from PBS to HTCondor for users
  - Simplify users' work
  - Help to achieve our scheduling strategy
- Implementation
  - Base on python API of HTCondor
  - Integrated with IHEP computing platform
    - Server name, group name
  - Several Jobs template according the experiments requirements

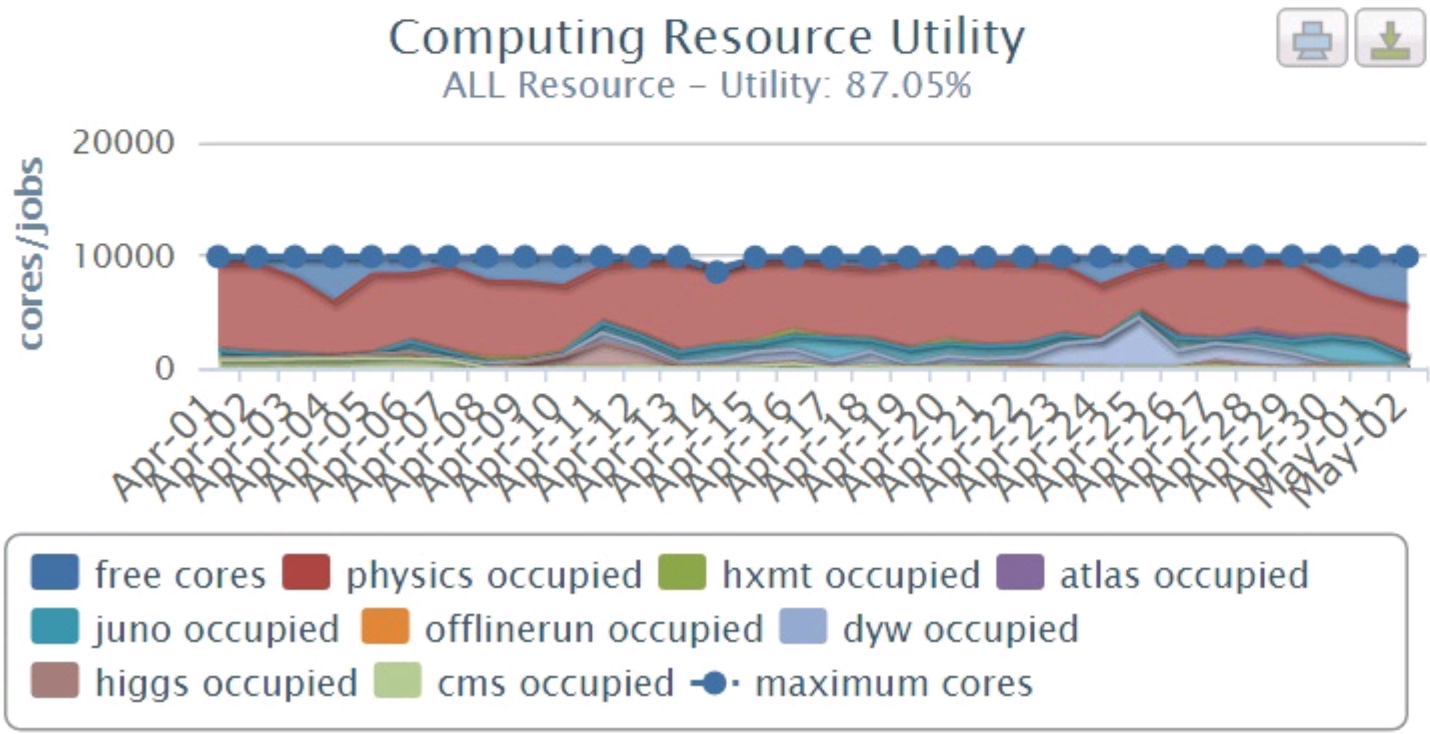


Put all together



Resource Utility Improvement

The HTCondor overall resource utility last month: ~87%



- The typical resource utility without resource sharing: 50% - 60%
- There is a significant improvement with the resource sharing policy

Content

1 IHEP Cluster Introduction

2 Scheduling Strategy to HTCondor

3 Implementation and Deployment

4 **Summary and Future Plan**

Summary and Future work

- Summary
  - The resource utility is significantly improved by the resource sharing strategy
  - Central controller helps to provide stable computing service

- Future work
  - Resource sharing ratio will be tuned dynamically according to the overloads of each group
    - The integration of Job Monitoring and Central Controller
  - Fine grain accounting system need to be developed

Thank you !

Question?