AFS Deployment and Activities of IHEP

Huang Qiulan

huangql@ihep.ac.cn

Computing Center, IHEP,CAS

2014-03-26

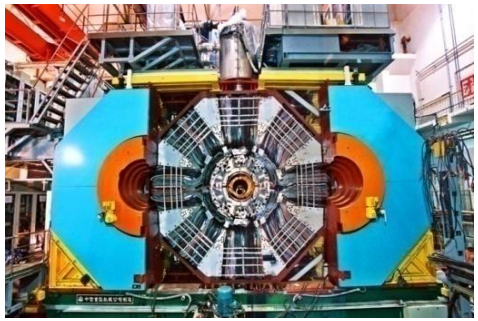
Outline

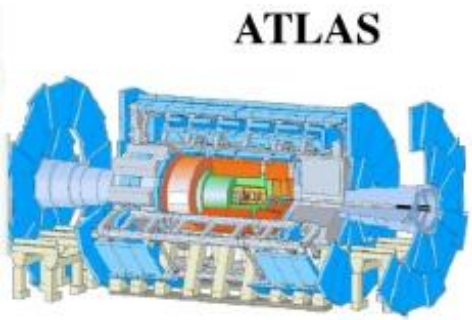
- **Overview of HEP computing in IHEP**
- **AFS status in IHEP**
  - Deployment, Performance and Issues
- **AFS activities in IHEP**
  - What we have done?
- **Summary**

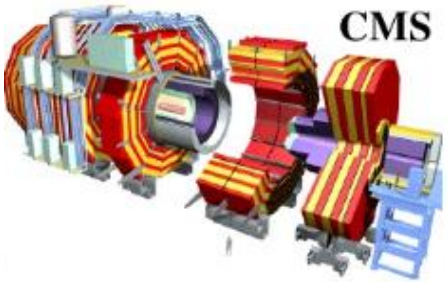
Overview of HEP computing in IHEP

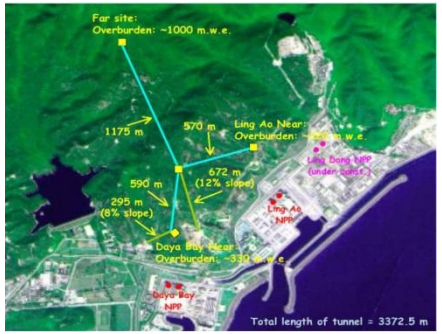
HEP computing in IHEP

- Support serveral experiments
  - BEPCII & BESIII
  - YBJ Cosmic Ray/Astrophysics in Tibet
  - DayaBay
  - CMS, ATLAS experiments on LHC
  - Future experiments
    - Lhaasso：1.2PB*10years
    - Jiangmen Undergroud Neutrino Observatory：1PB*10years









Computing center in IHEP

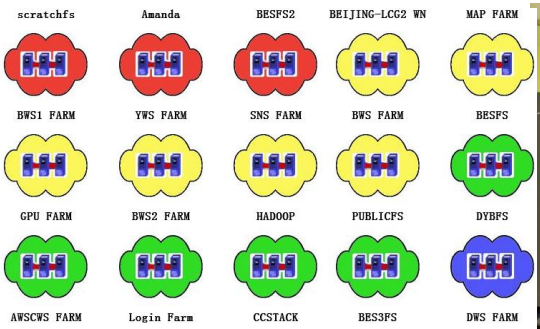
Power supply, cooling

主机组	主机状态汇总	服务状态汇总
备份服务器：姚秋玲郑伟 (Amanda-servers)	8 运行	18 正常
Atlas计算节点:负责人，系统组值班人员 (Atlas-servers)	28 运行	337 正常
BES计算节点:负责人，系统组值班人员 (BES-Servers)	444 运行	4849 正常 7 警报：7 未处理 28 严重：27 未处理 1 已确认
BIO计算结点负责人系统组值班人员 (Bio-servers)	17 运行	169 正常 1 未知：1 未处理
计算中心节点:负责人，系统组值班人员 (CC-Servers)	45 运行	495 正常
CMS节点:负责人，闻晓飞 (CMS-Servers)	38 运行	396 正常
DYB计算节点:负责人，系统组值班人员 (DYB-Servers)	36 运行	396 正常

Monitoring

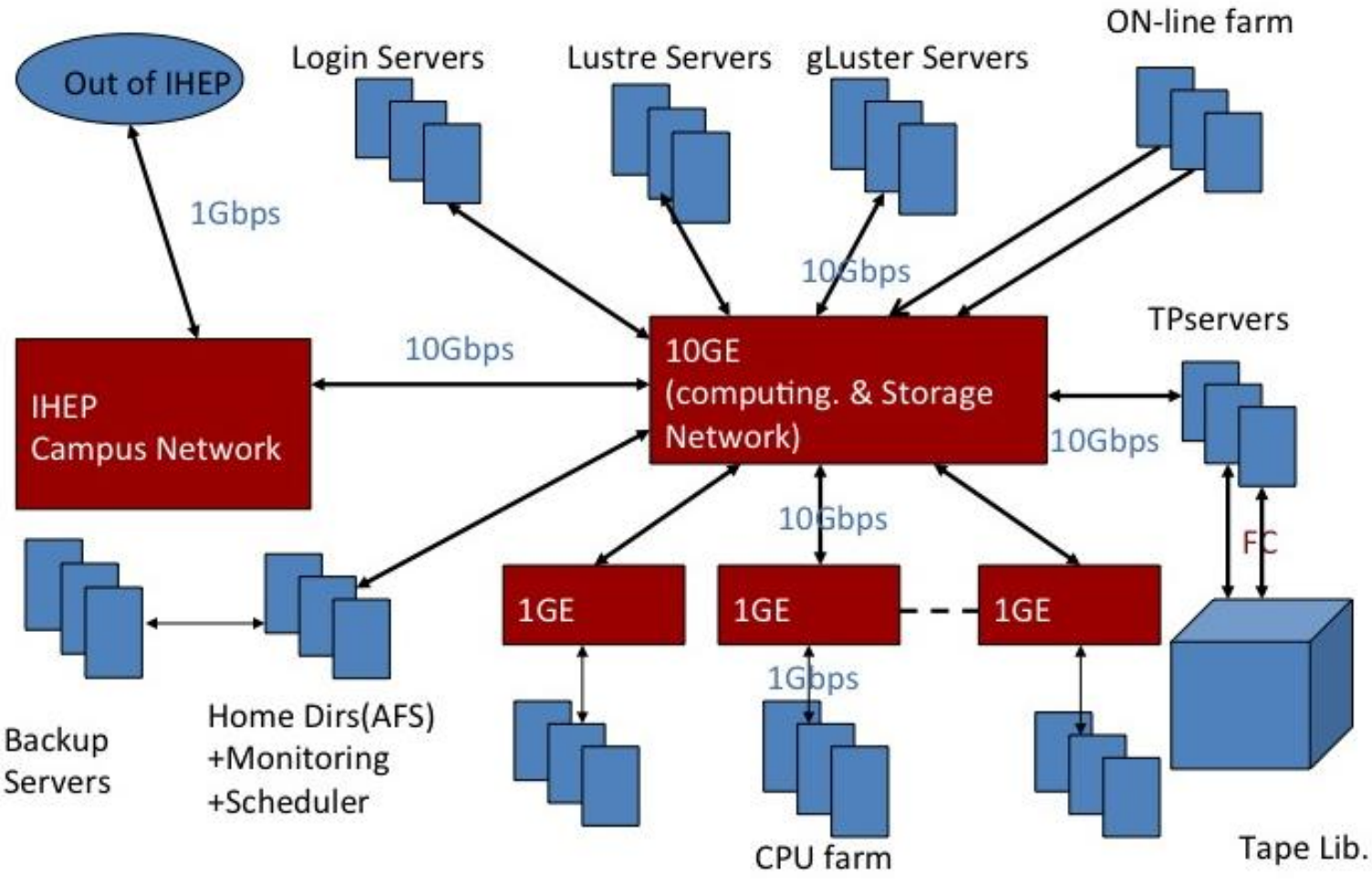
~10000 CPU cores

>4PB disk space(Lustre/Gluster)



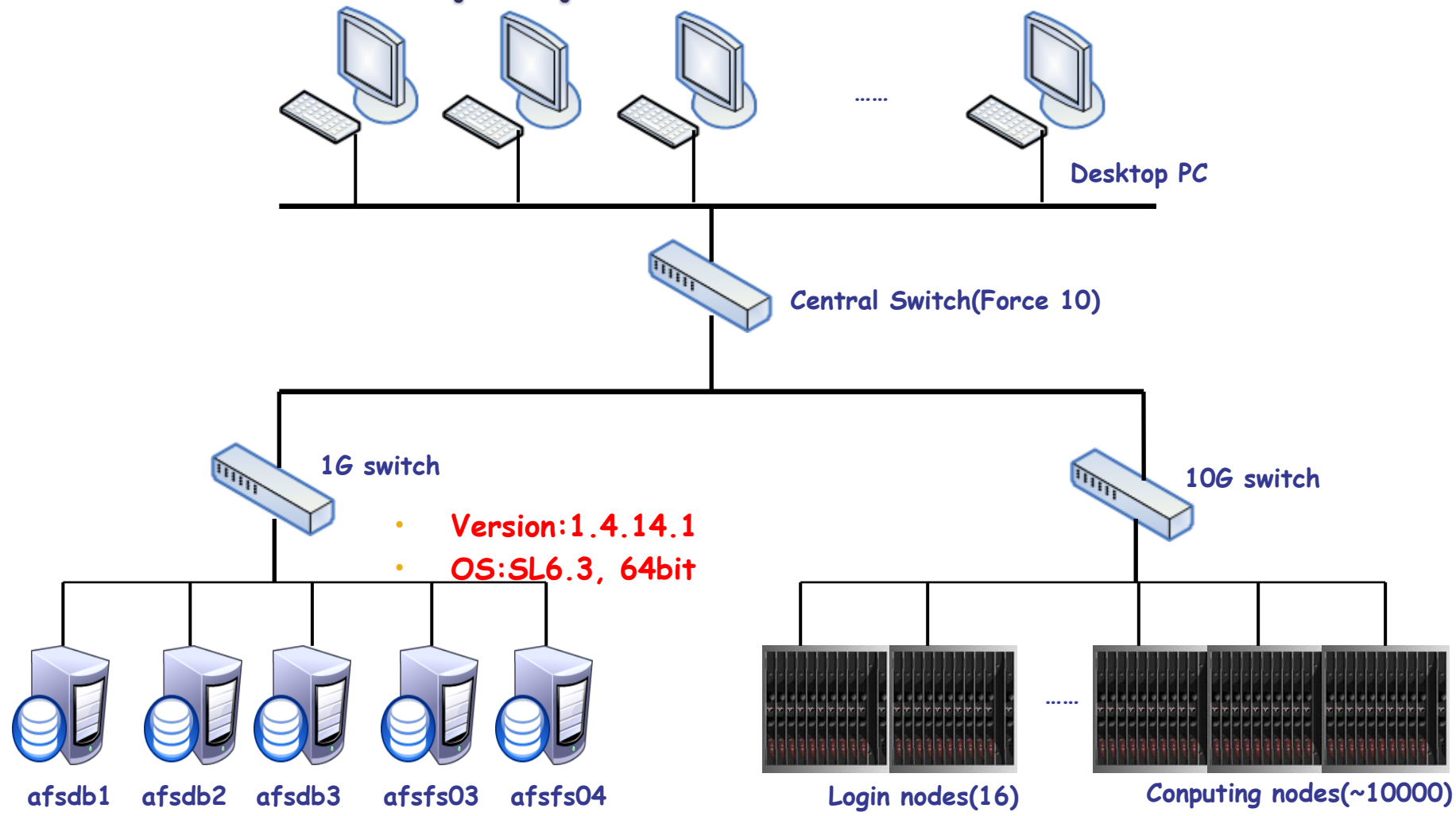
5PB tape library

Computing architecture



AFS status in IHEP

Deployment of AFS



1. Master database: afsdb1
2. Slave database:afsdb2,afsdb3
3. Fileserver:afsdb2,afsdb3,afsfs03,afsfs04
4. Total size: 7.8TB

1. AFS client installed in all login nodes
2. Tokens when login using PAM
3. UID and GID stored in /etc/passwd file, no password in /etc/shadow

1. AFS cache set 10GB
2. Jobs scheduled to computing nodes access software lib in AFS

Volume of AFS file system

- Obey the Rules of Mount Point Traversal

- User volume:500MB, home directory
- Soft volume
  - size is different, decieded by application requirements
  - Common software library like Boss,gcc,Gaudi,etc.

Status of AFS file system

- Capacity:7.8TB
- Provide Home Directory for all computing users(1384 users)
- Support thousands of concurrent access
- AFS authentication was integrated into some systems in computing environment
- Manage all users in AFS instead of LDAP

Problems

- I/O latency when high concurrent access
- Fine-granularity monitoring
  - Status of volumes(read/write)
  - capture who and where send requests to AFS in server side
  - Record client requests, filepath
- I/O performance
- Failed to get tokens when login(Kerberos 4)

AFS activities in IHEP

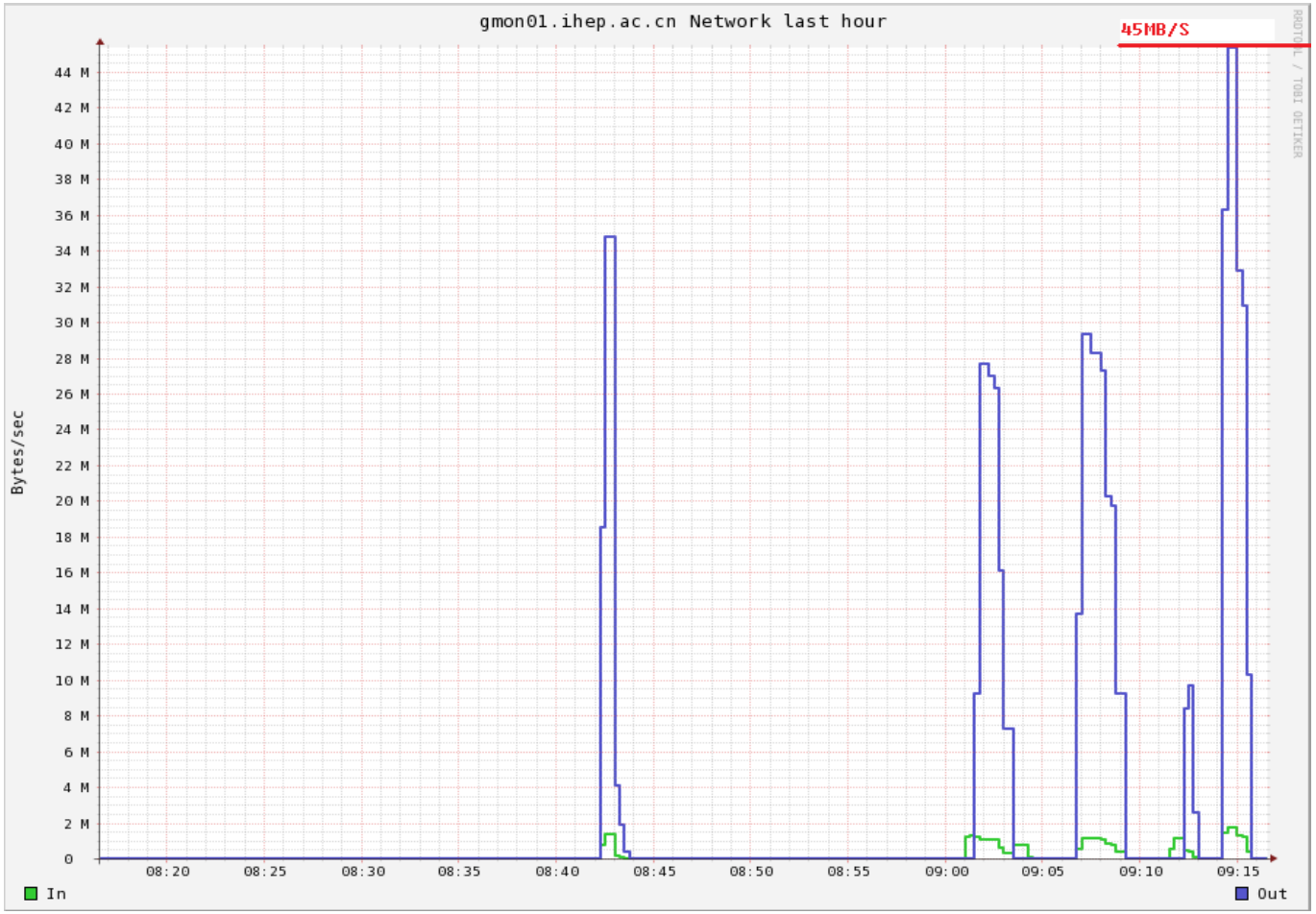
What we have done?

- Performance tuning
- Upgrading
  - AFS version:1.4.4→1.4.14.1
  - OS version:redhat AS 4(32bit)→SL6.3(64bit)
- Data backup
  - Backup user volumes and some soft volumes routinely using Amanda
- Unified authentication using AFS in computing environment
- Integration Torque and AFS
  - PAFSI（PBS and AFS Integration）
- Develop User Service System

Performance tuning

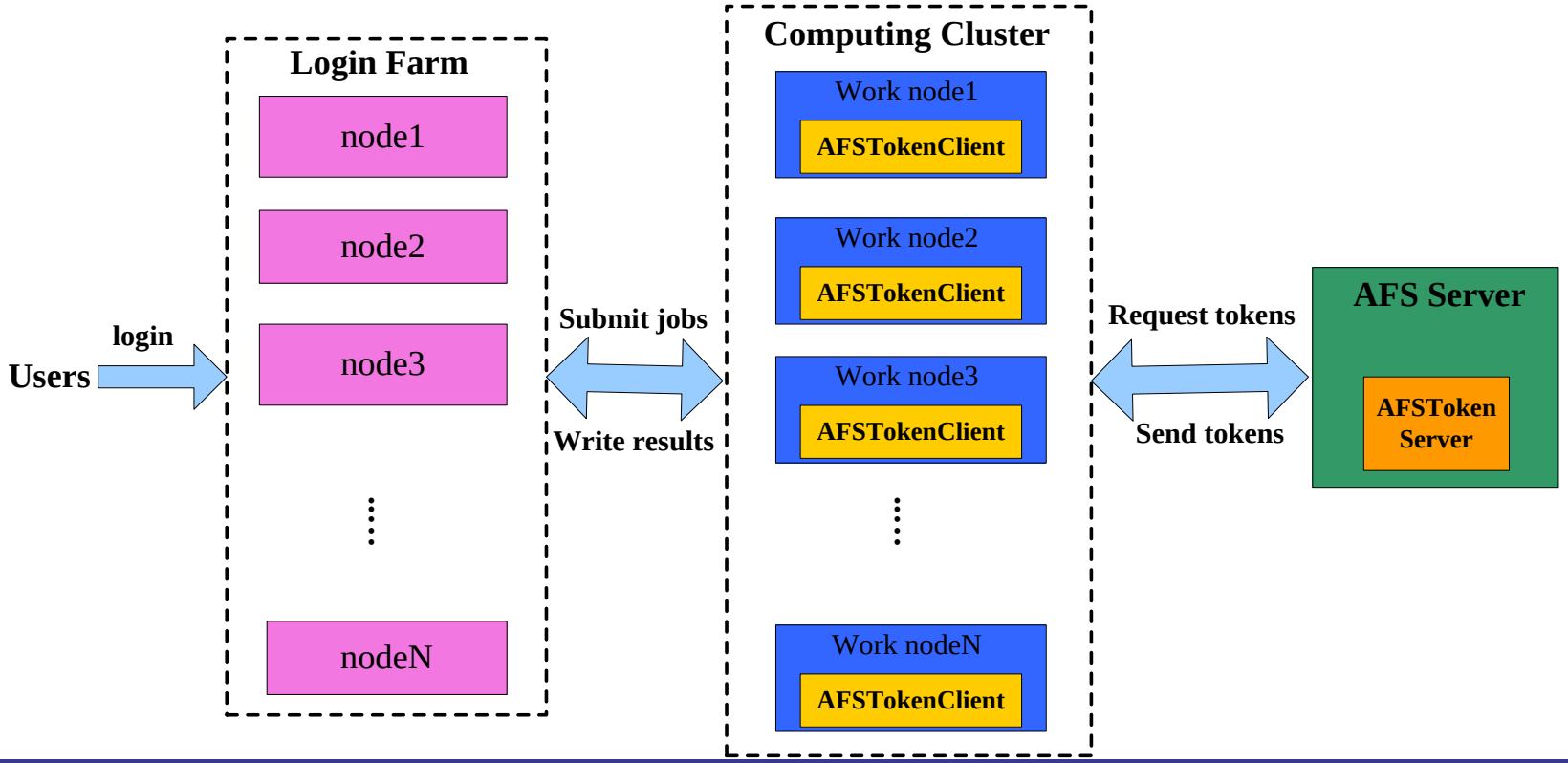
- Some configuration adjustments
  - AFS client
    - Cache size
    - chunk size
    - Files: the target number bof files in cache
    - Dcache :number of data cache entries
    - Stat: number of stat cache entries
  - AFS Server
    - Replication volume
- Latencies in AFS read access failures reduced to 0 when 3000 clients concurrent accessing
- I/O performance:45MB/s

I/O performance for each client



PAFSI: Integration Torque and AFS

- Integrate AFS and Torque, so that jobs can automatically access AFS areas
- No change to the AFS source just buliding some executables that use AFS API calls
- Some changes to Torque source



AFSTokenServer

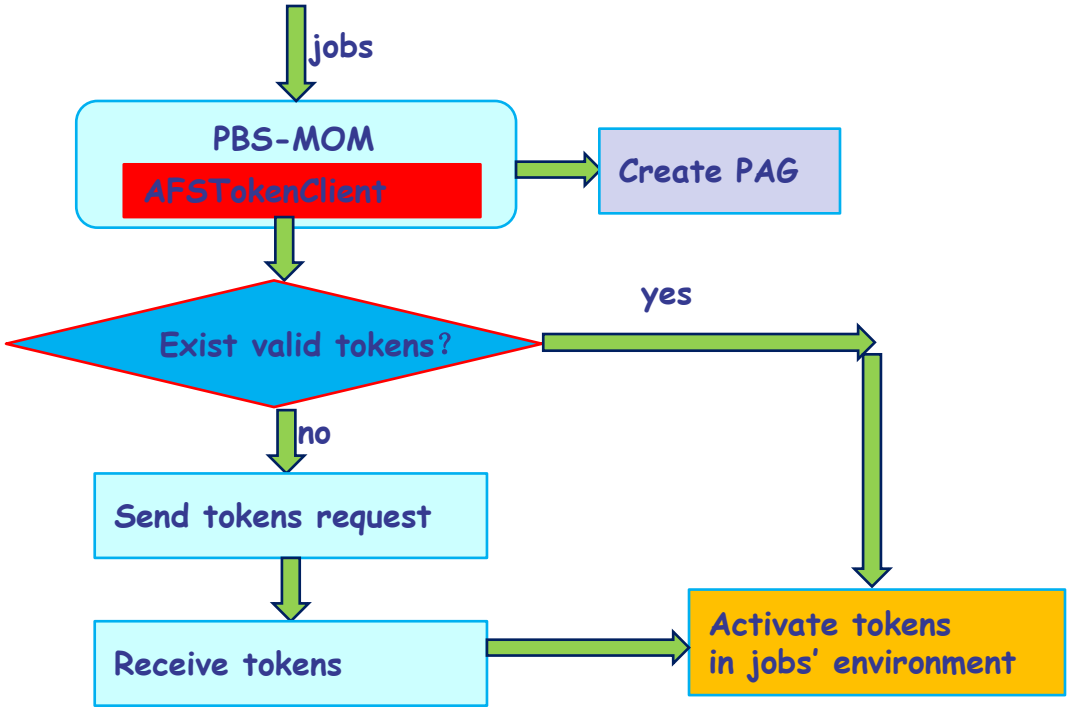
- No change to AFS source
- Coding some executables by calling AFS API in AFS-DEVEL
- Implemented:
  - forgeToken
  - activateToken
  - extendToken
- Forge tokens for jobs according to JOBID, JOBOWNER without any password
- Adopt ActiveMQ to manage communication between servers and clients, to send valid tokens to computing nodes

Directed by
Fabio Hernandez
(IN2P3)
fabio@in2p3.fr
2011

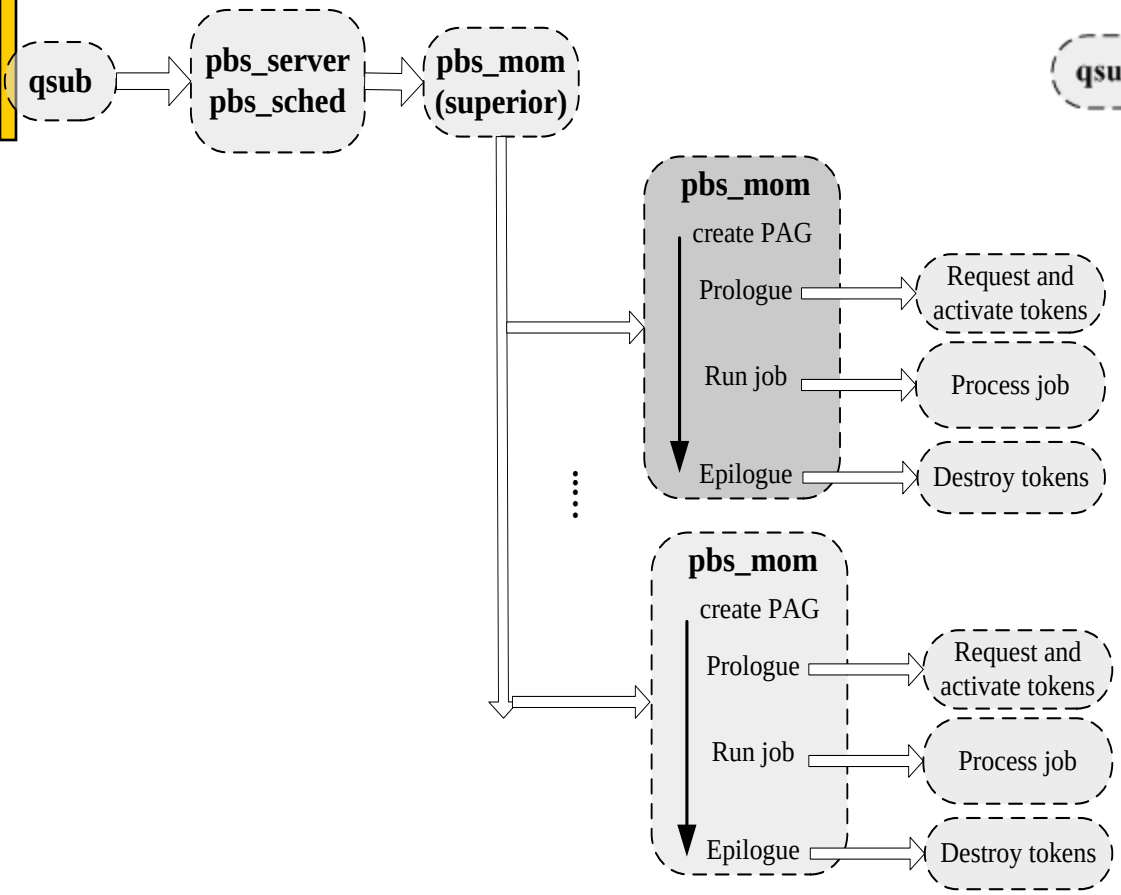
AFSTokenClient

- Change Torque PBS source: pbs-mom mudule
- Running on all computing nodes to be responsible for request, receive, save and activiate tokens

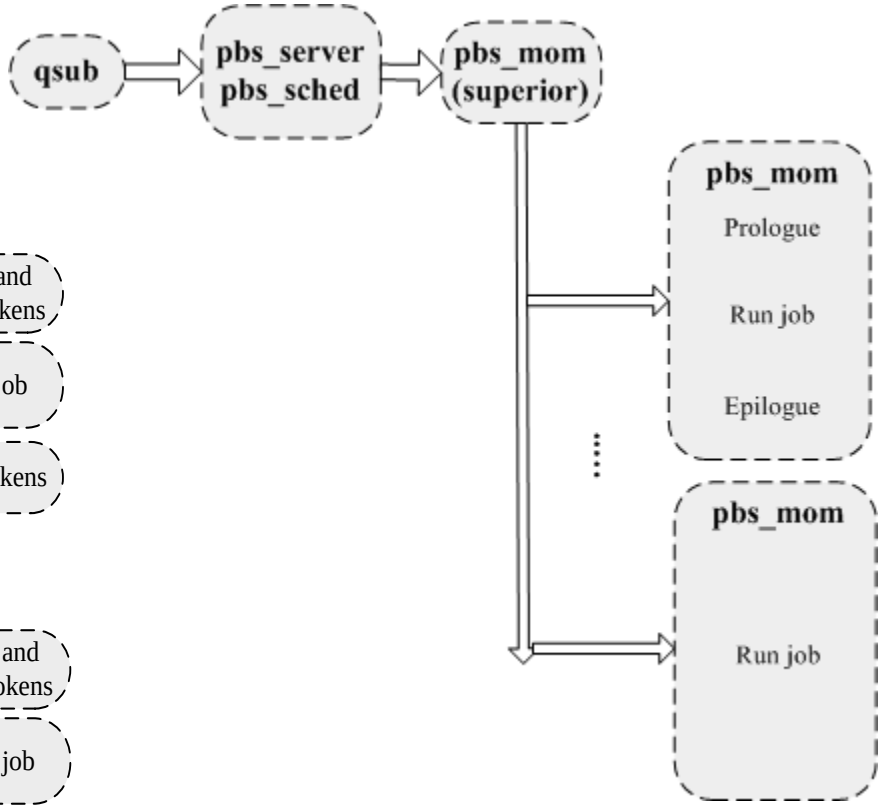
  - When new jobs coming, send tokens request, （ JOBID,JOBOWNER, HOSTNAME）
  - Receive and save tokens
  - Activate tokens in jobs environment



PAFSI VS Torque



Job process in PAFSI



Job process in Torque

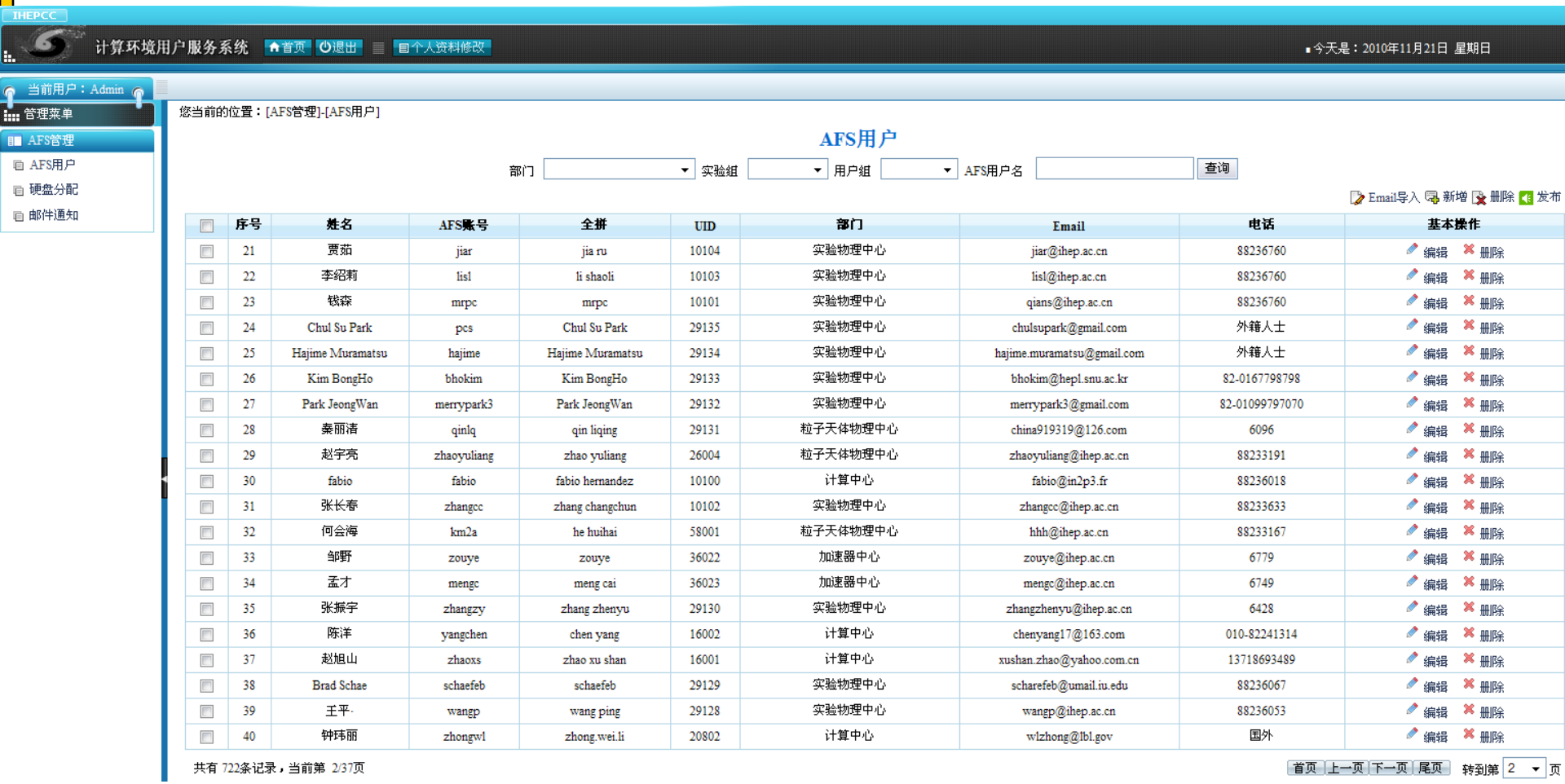
Status of PAFSI

- Components
  - AFS server:AFSTokenServer
  - ActiveMQ
  - Computing nodes (torque-IHEP-client)
- Deployed 84 computing nodes, 488 CPU cores
- AFSTokenServer is used to forge tokens for backup system
- Problems
  - ActiveMQ crashed when high concurrent access
  - AFSTokensServer not stable

Computing User Service System

- A Platform for user management
- Convenient
  - Administrator to Create/Edit/Search/Remove AFS users from WEB interface without complicate procedure.
- Information Integrity
  - All user information stored in DB
- Friendly
  - WEB interface is friendly and easy to operate
- The system is on line and more services being improved

To store all AFS users in DB



序号	姓名	AFS账号	全拼	UID	部门	Email	电话	基本操作
21	贾茹	jiar	jia ru	10104	实验物理中心	jiar@ihep.ac.cn	88236760	编辑 删除
22	李绍莉	lisl	li shaoli	10103	实验物理中心	lisl@ihep.ac.cn	88236760	编辑 删除
23	钱森	mrpc	mrpc	10101	实验物理中心	qians@ihep.ac.cn	88236760	编辑 删除
24	Chul Su Park	pcs	Chul Su Park	29135	实验物理中心	chulsupark@gmail.com	外籍人士	编辑 删除
25	Hajime Muramatsu	hajime	Hajime Muramatsu	29134	实验物理中心	hajime.muramatsu@gmail.com	外籍人士	编辑 删除
26	Kim BongHo	bhokim	Kim BongHo	29133	实验物理中心	bhokim@hepl.snu.ac.kr	82-0167798798	编辑 删除
27	Park JeongWan	merrypark3	Park JeongWan	29132	实验物理中心	merrypark3@gmail.com	82-01099797070	编辑 删除
28	秦丽清	qinlq	qin liqing	29131	粒子天体物理中心	china919319@126.com	6096	编辑 删除
29	赵宇亮	zhaoyuliang	zhao yuliang	26004	粒子天体物理中心	zhaoyuliang@ihep.ac.cn	88233191	编辑 删除
30	fabio	fabio	fabio hernandez	10100	计算中心	fabio@in2p3.fr	88236018	编辑 删除
31	张长春	zhangcc	zhang changchun	10102	实验物理中心	zhangcc@ihep.ac.cn	88233633	编辑 删除
32	何会海	km2a	he huihai	58001	粒子天体物理中心	hhh@ihep.ac.cn	88233167	编辑 删除
33	邹野	zouye	zouye	36022	加速器中心	zouye@ihep.ac.cn	6779	编辑 删除
34	孟才	mengc	meng cai	36023	加速器中心	mengc@ihep.ac.cn	6749	编辑 删除
35	张振宇	zhangzy	zhang zhenyu	29130	实验物理中心	zhangzhenyu@ihep.ac.cn	6428	编辑 删除
36	陈洋	yangchen	chen yang	16002	计算中心	chenyang17@163.com	010-82241314	编辑 删除
37	赵旭山	zhaoxs	zhao xu shan	16001	计算中心	xushan.zhao@yahoo.com.cn	13718693489	编辑 删除
38	Brad Schae	schaefeb	schaefeb	29129	实验物理中心	scharefeb@umail.iu.edu	88236067	编辑 删除
39	王平	wangp	wang ping	29128	实验物理中心	wangp@ihep.ac.cn	88236053	编辑 删除
40	钟玮丽	zhongwl	zhong.wei.li	20802	计算中心	wlzhong@lbl.gov	国外	编辑 删除

To create an AFS user

Summary

- Achievements
  - Latencies in AFS read access failures reduced
  - I/O performance:20MB/s→45MB/s
  - Design and implement PAFSI to integrate OpenAFS and Torque
  - Flexible Computing User Service System
  - Manage all computing users instead of LDAP
- Issues
  - Kerberos 4 authentication
  - Stability of PAFSI
  - I/O latency when high concurrent access
  - Fine-granularity monitoring

Thank you
Question?