IHEP Computing Center Site Report

Gang Chen (Gang.Chen@ihep.ac.cn)
Computing Center
Institute of High Energy Physics

HEPiX 2011 Spring Meeting

CC-IHEP at a Glance

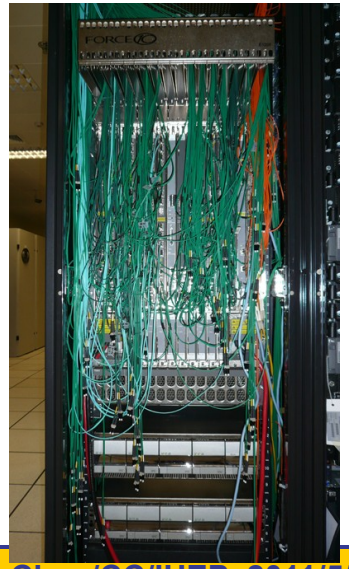
- The Computing Center was created in 1980's
  - Provided computing service to BES, the experiment on BEPC
- Rebuilt in 2005 for the new projects:
  - BES-III on BEPC-II
  - Tier-2's for ATLAS, CMS
  - Cosmic ray experiments
- 35 FTEs, half of them for computing facility

Computing Resources

- ~6600 CPU-cores
  - SL5.5 (64 bit) for WLCG
  - SL4.5 (32 bit) for BES-III, Migrating to SL5.5
  - Toque: torque-server-2.4
  - Maui: maui-server-3.2.6
- Blade system, IBM/HP/Dell
  - Blade links with GigE/IB
  - Chassis links to central switch with 10GigE

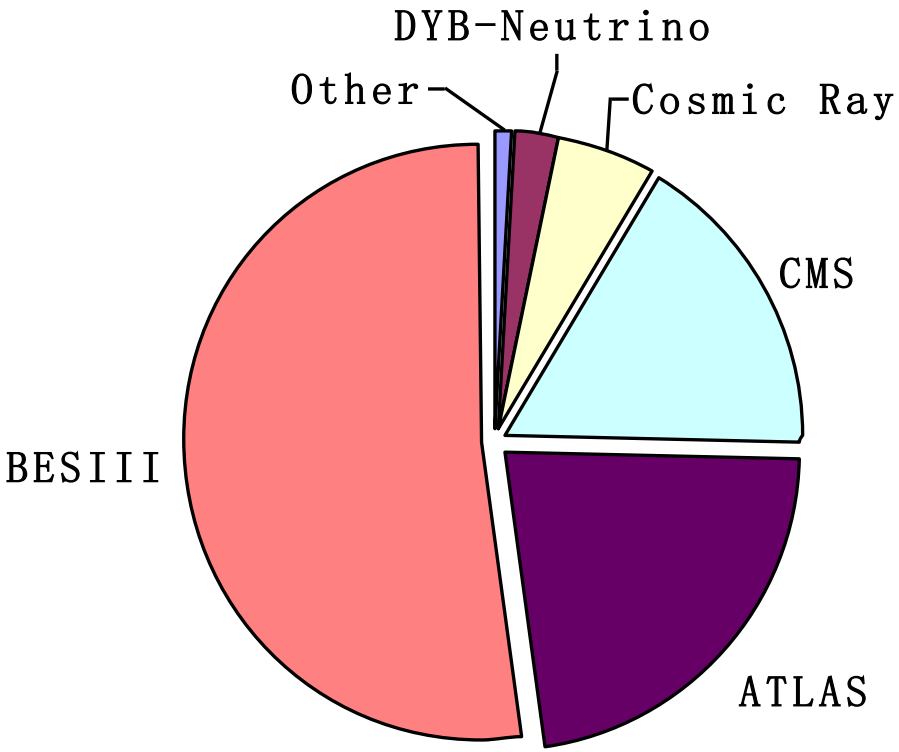
PC farm built with blades

Force10 E1200 Central Switch

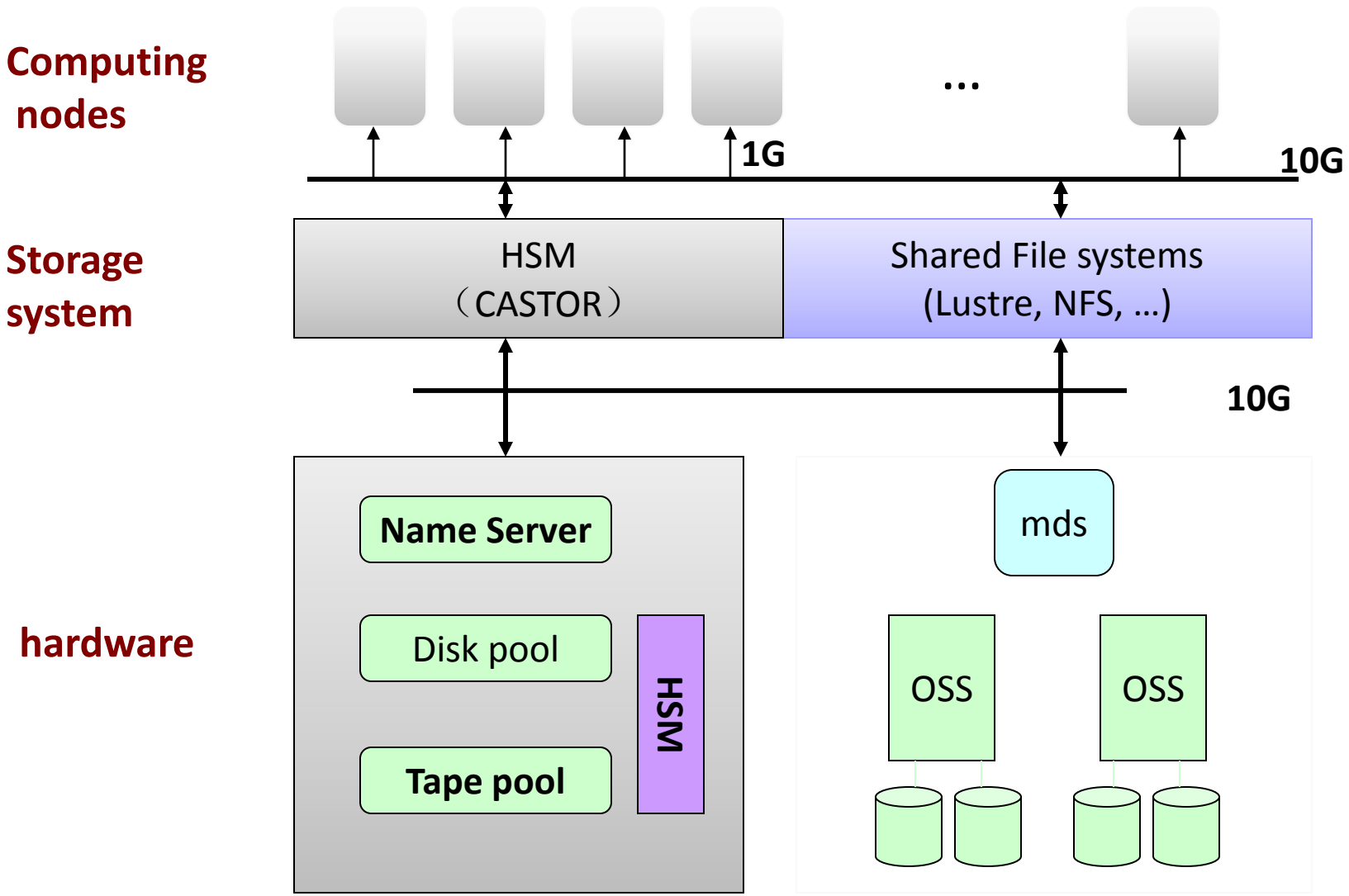


Resources used per VO

CPU hours
From 01-01-2010 to 31-12-2010

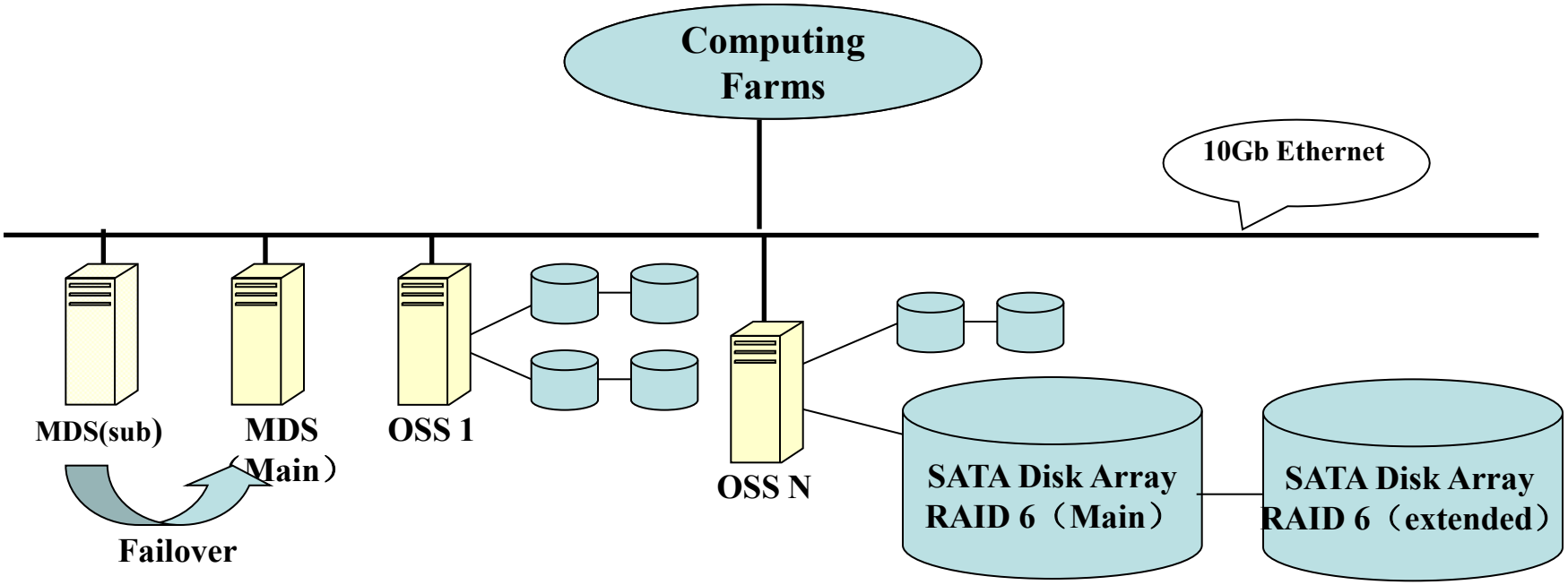


Storage Architecture



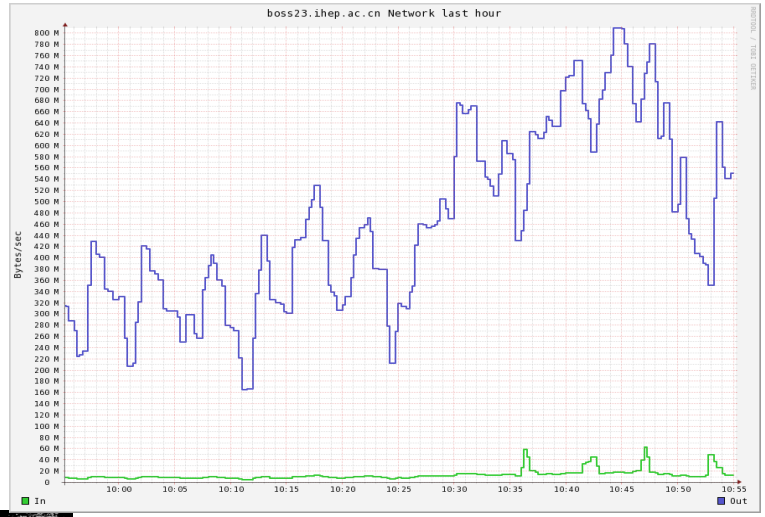
Lustre System

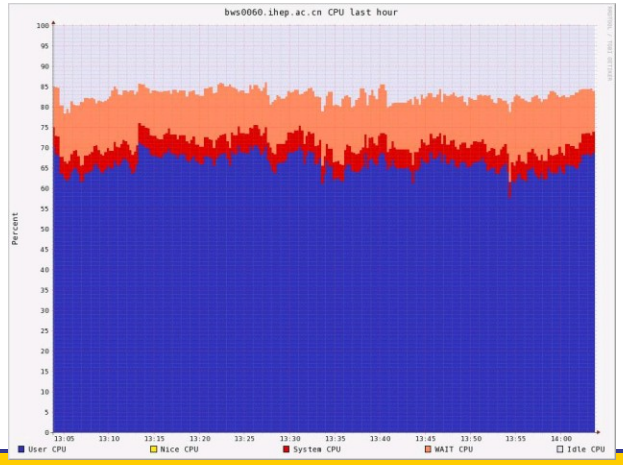
- Version:1.8.1.1
- 32 I/O servers, each attached with 4 SATA Disk Arrays
- Storage capacity: 1.7 PB
- Name Space: 3 mount points (for different experiments)



Lustre Performance

- Peak throughput of data analysis: 800MB/s per I/O server.
- Total throughput ~25GB/s



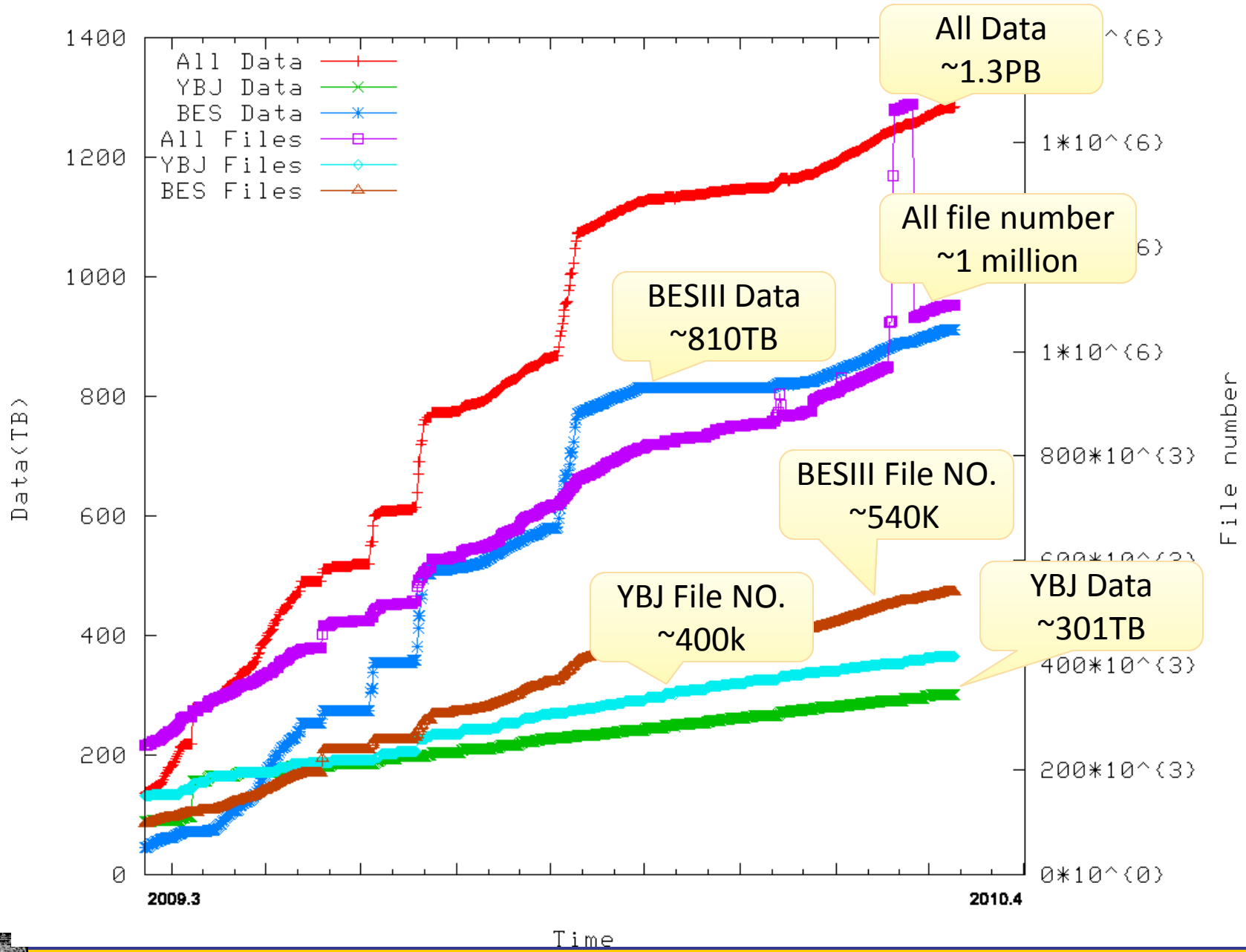


Lustre Lessons

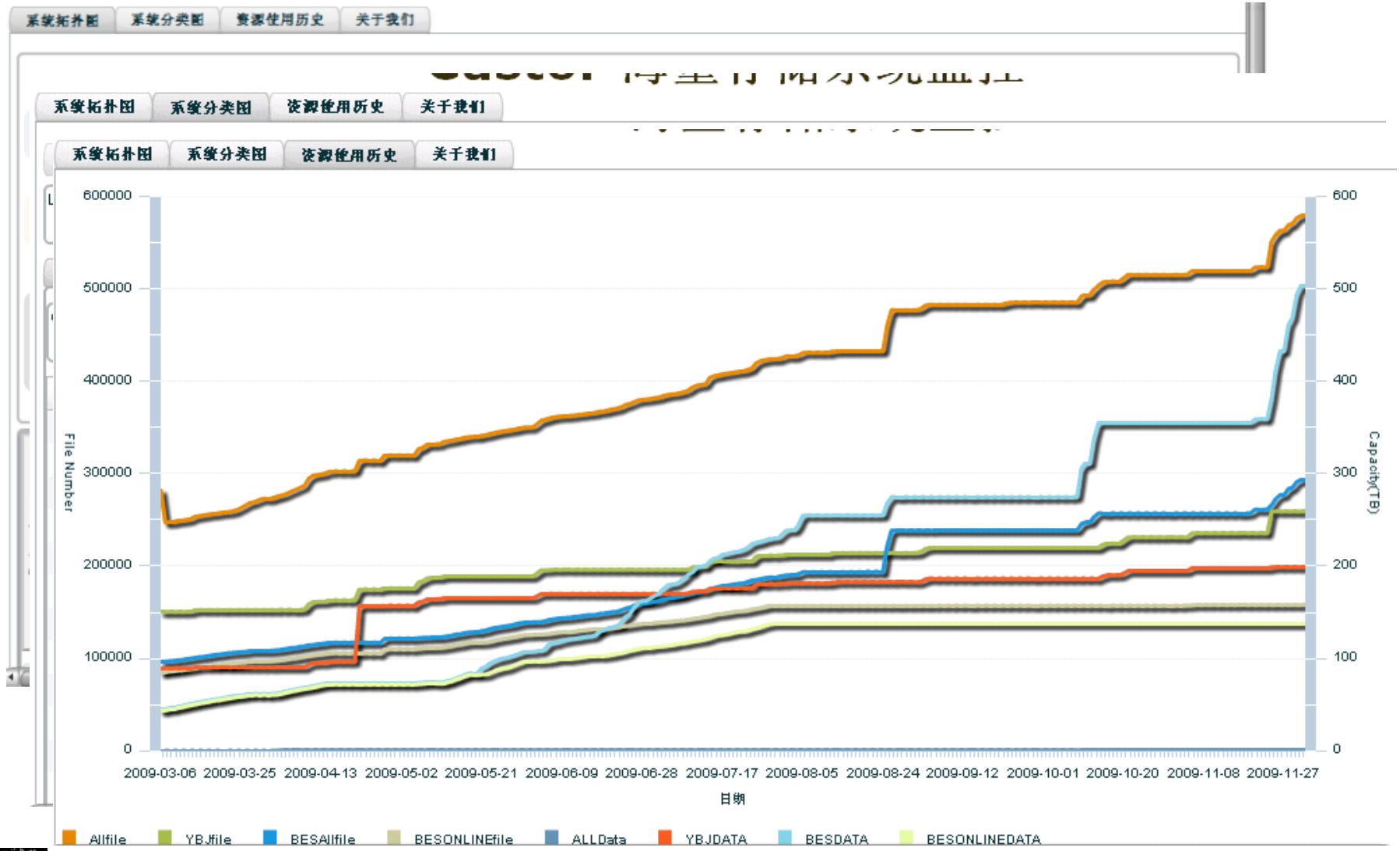
- Low-Memory runs out may cause the system crash
  - Move to 64-bit OS
  - Optimize the patterns of read/write
- Security and user-based ACL
  - recompilation of source code is needed to add certain modules

HSM Deployment

- Hardware
  - Two IBM 3584 tape libraries
  - ~5800 slots, with 26 LTO-4 tape drivers
  - 10 tape servers and 10 disk servers with 200TB disk pool
- Software
  - Customized version based on CASTOR 1.7.1.5
  - Support the new types of hardware
  - Optimize the performance of tape read and write operation
  - Stager was re-written
- Network
  - 10Gbps link between disk servers and tape servers


All Data
YBJ Data
BES Data
All Files
YBJ Files
BES Files
Data(TB)
File number
Time
2009.3
2010.4
All Data ~1.3PB
All file number ~1 million
BESIII Data ~810TB
BESIII File NO. ~540K
YBJ File NO. ~400k
YBJ Data ~301TB

Realtime Monitoring of Castor



File Reservation for Castor

- The File Reservation component is a add-on component for Castor 1.7.
- Developed to prevent the reserved files from migrating to tape when disk usage is over certain level.
- Provides a command line Interface and a web Interface. Through these two Interfaces, user can:
  - Browse mass storage name space with a directory tree
  - Make file-based ,dataset-based and tape-based reservation
  - Browse, modify and delete reservation.

File Reservation System for Castor

Castor 文件预留系统

首 页 | 文件预留 | 文件查询 | 文件修改 | 文件删除

Loading data..
- castor
  - ihep.ac.cn
    - bes
      - bes2
        - raw
      - besTest
      - offline
      - run
      - userTest
    - public
    - user
    - ybj

你当前的位置：/castor/ihep.ac.cn/bes/bes2/raw-文件列表

Castor 文件预留系统

联系我们 | 帮助

首 页 | 文件预留 | 文件查询 | 文件修改 | 文件删除

castor
- ihep.ac.cn
  - bes
    - bes2
      - raw
    - besTest
    - offline
    - run
    - userTest
  - public
  - user
  - ybj

你当前的位置：/castor/ihep.ac.cn/bes/bes2/raw-文件列表

当前页 全选 新增

文件名： 确定 过滤条件： 确定

序号	类型	文件名	预留天数	修改时间	基本操作
1	raw	17605.raw	0		编辑 删除
2	raw	run11704.raw	0		编辑 删除
3	raw	run11705.raw	0		编辑 删除
4	raw	run11706.raw	0		编辑 删除
5	raw	run11707.raw	0		编辑 删除
6	raw	run11708.raw	0		编辑 删除
7	raw	run11709.raw	0		编辑 删除
8	raw	run11710.raw	0		编辑 删除

共有 109

需要预留的文件列表如下：

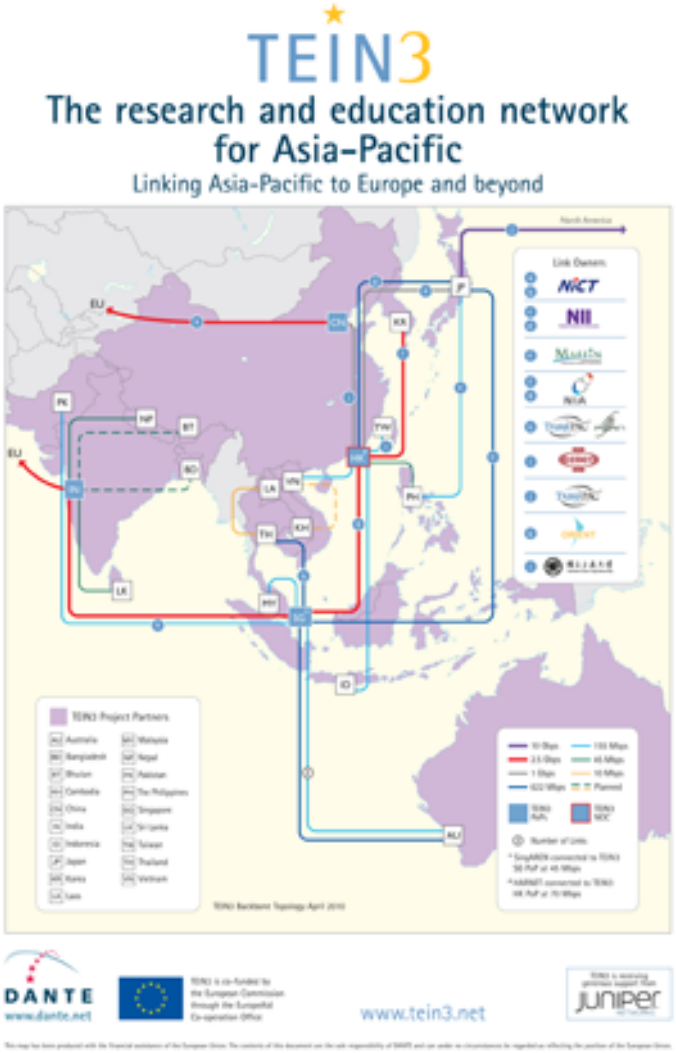
预留文件名	预留天数
/castor/ihep.ac.cn/bes/bes2/raw/17605.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11704.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11705.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11706.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11707.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11708.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11709.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11710.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11711.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11712.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11713.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11714.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11715.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11716.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11717.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11718.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11719.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11720.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11721.raw	30 天*
/castor/ihep.ac.cn/bes/bes2/raw/run11722.raw	30 天*

共有 20 条记录，当前第 1/2页 首页 上一页 下一页 尾页 转到第 页

确 定 重 置

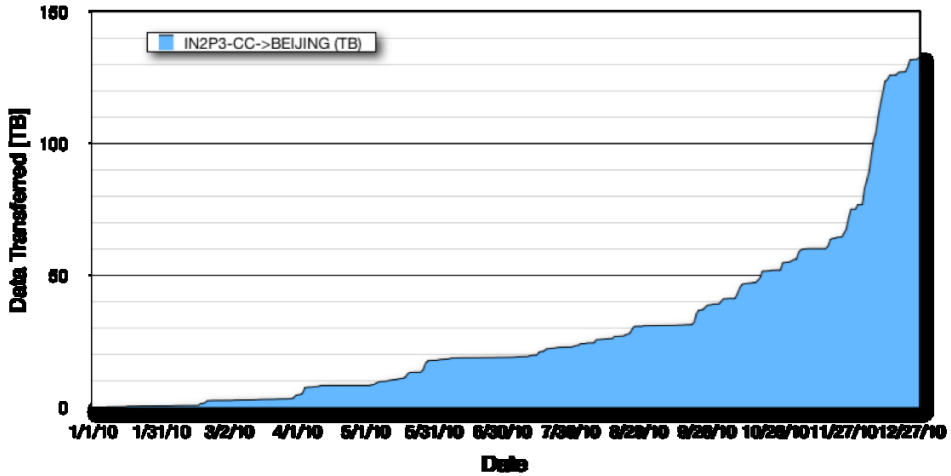
Global Networking

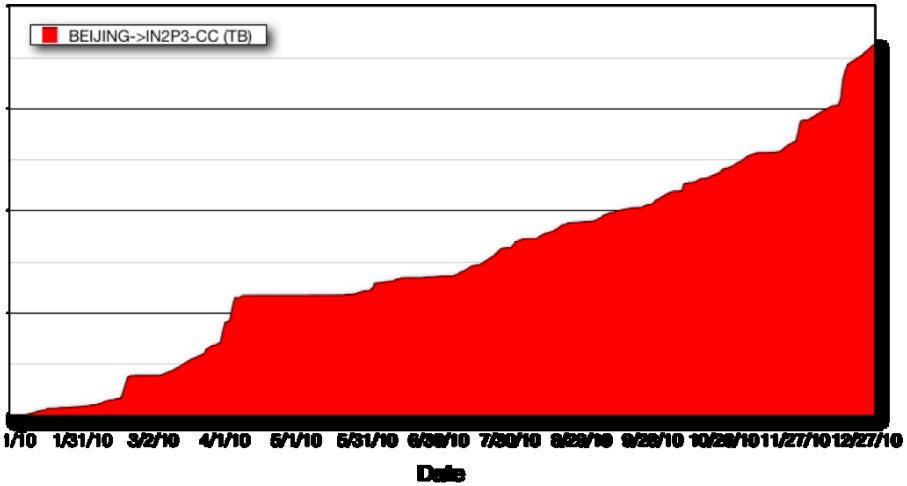
Via ORIENT/TEIN3 to Europe



Via Gloriad to US

ATLAS Data transfer between Lyon and Beijing

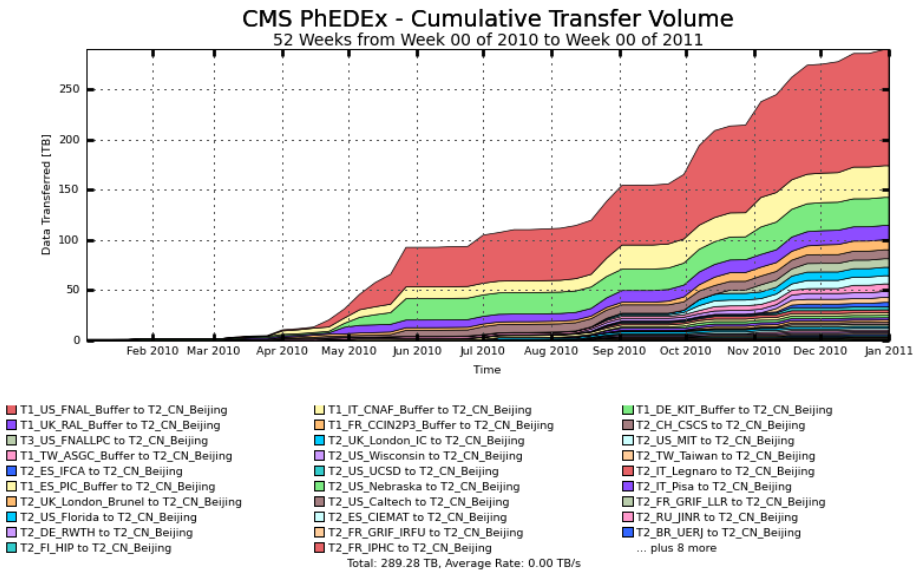


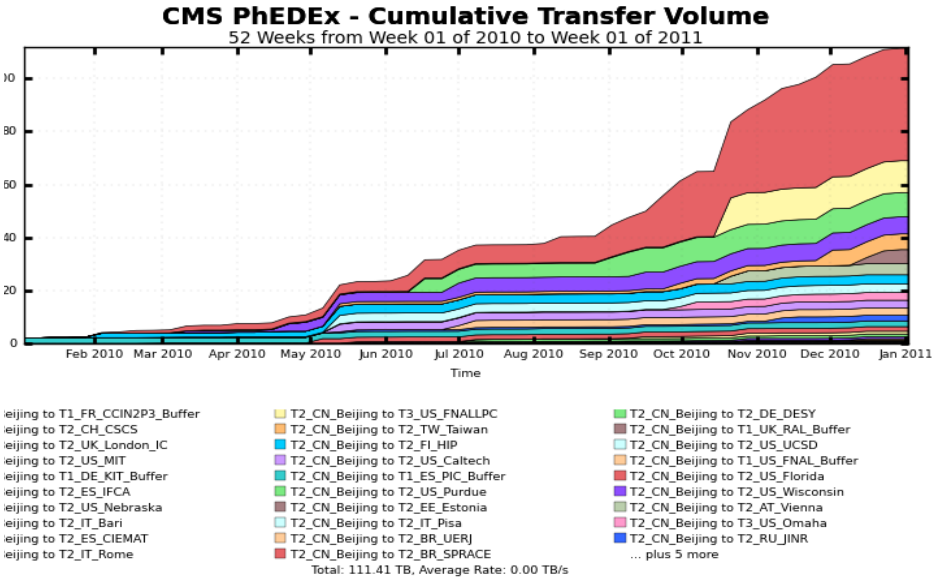


> 130 TB of data transferred from Lyon to Beijing in 2010

> 35 TB of data transferred from Lyon to Beijing in 2010

CMS Data transfer from/to Beijing





~290 TB transferred from elsewhere to Beijing in 2010

~110 TB transferred from Beijing elsewhere in 2010

Cooling System

- Air Cooling system reached 70% of capacity
- Cool air partition was built in 2009 and 2010
- Water cooling is being discussed

Conclusion

- CPU farms work fine, but must migrate the 32-bit system to 64-bit as soon as possible.
- Lustre is the major storage system at IHEP with acceptable performance but also some trivial problems.
- Resources, CPU and storage, increase much faster than what we expected, which cause problems: system stability, batch system scalability, cooling, etc.

Thank you