

IHEP Computing Center Site Report

Gang Chen (Gang.Chen@ihep.ac.cn)

Computing Center

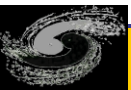
Institute of High Energy Physics



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



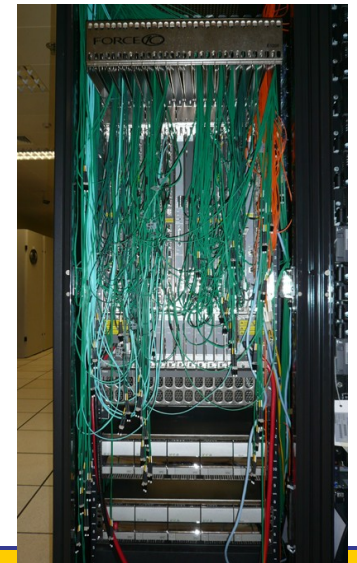
Computing Resources

- ~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



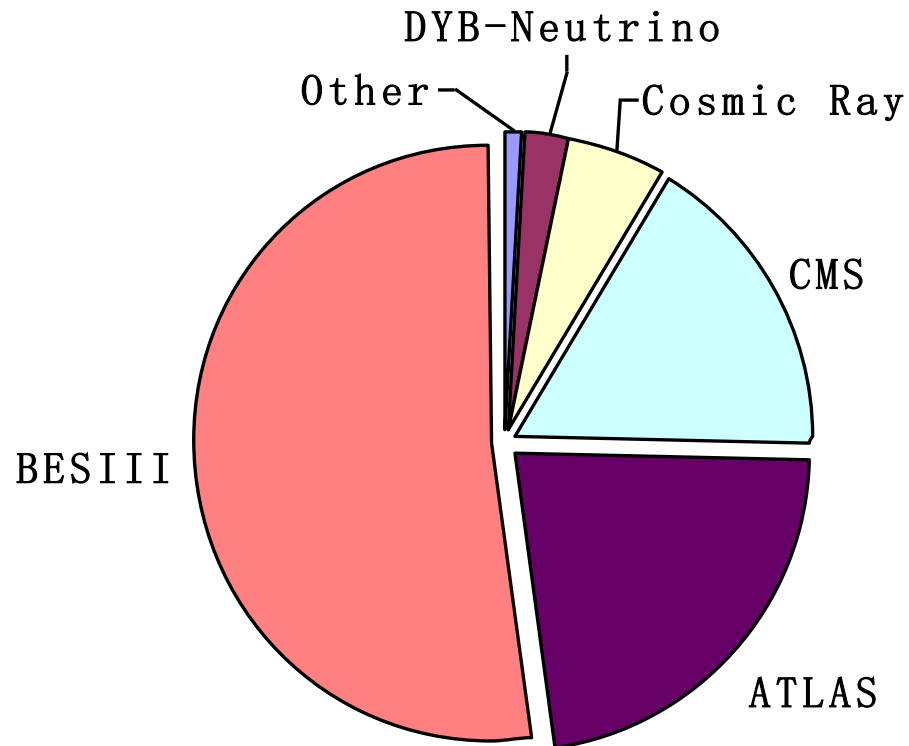
PC farm built
with blades

Force10 E1200
Central Switch



Resources used per VO

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



Storage Architecture

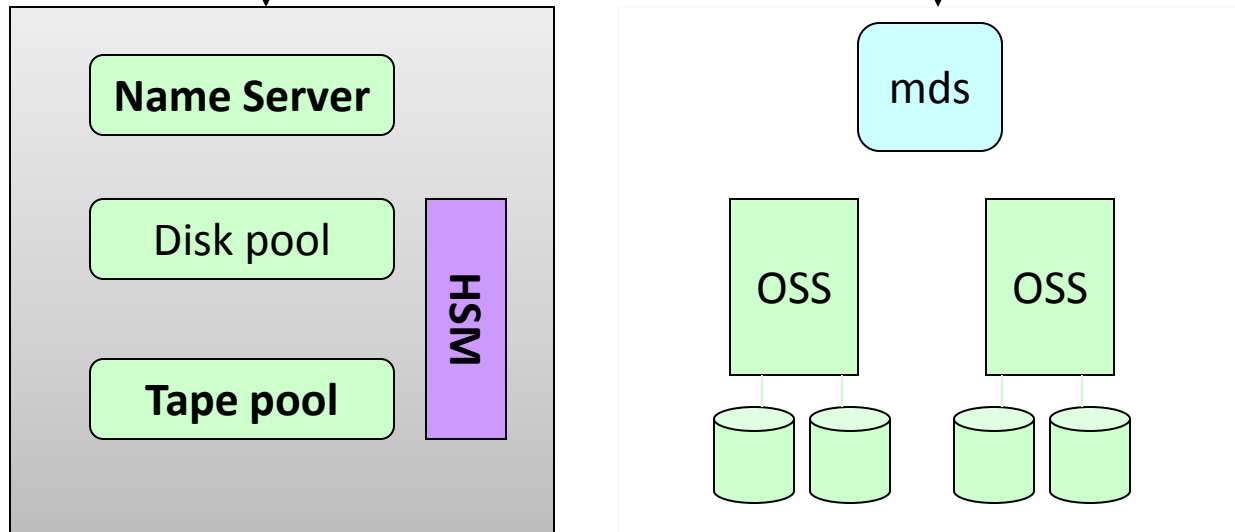
Computing nodes



Storage system

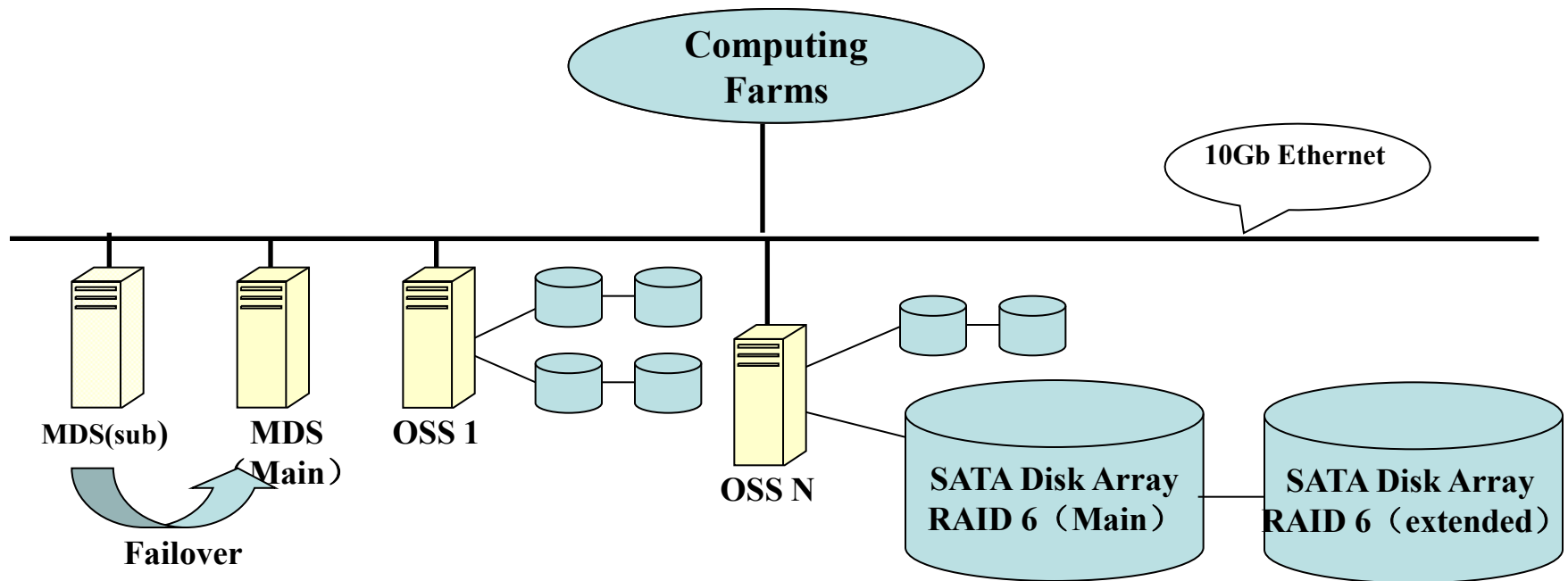


hardware



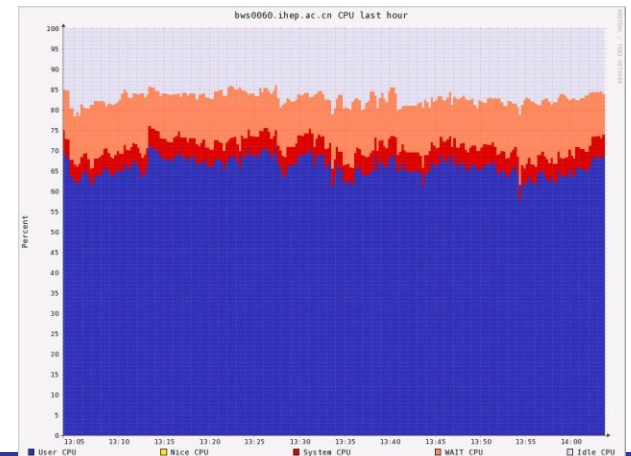
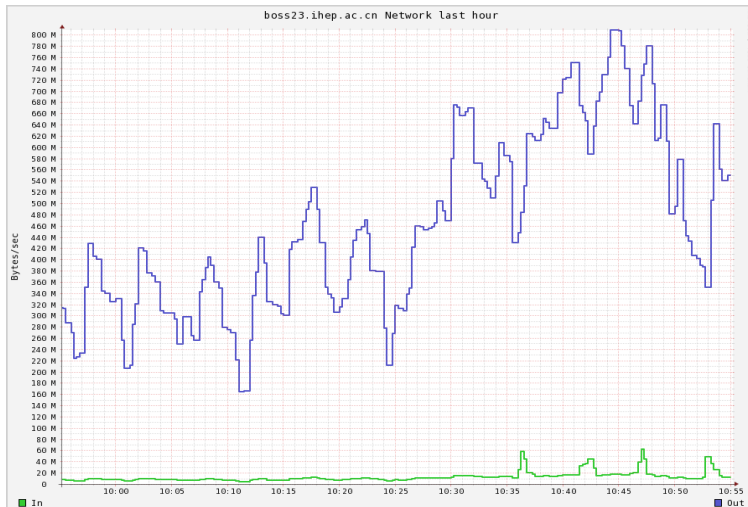
Lustre System

- 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)



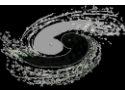
Lustre Performance

- Peak throughput of data analysis: 800MB/s per I/O server.
- Total throughput $\sim 25\text{GB/s}$



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

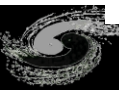
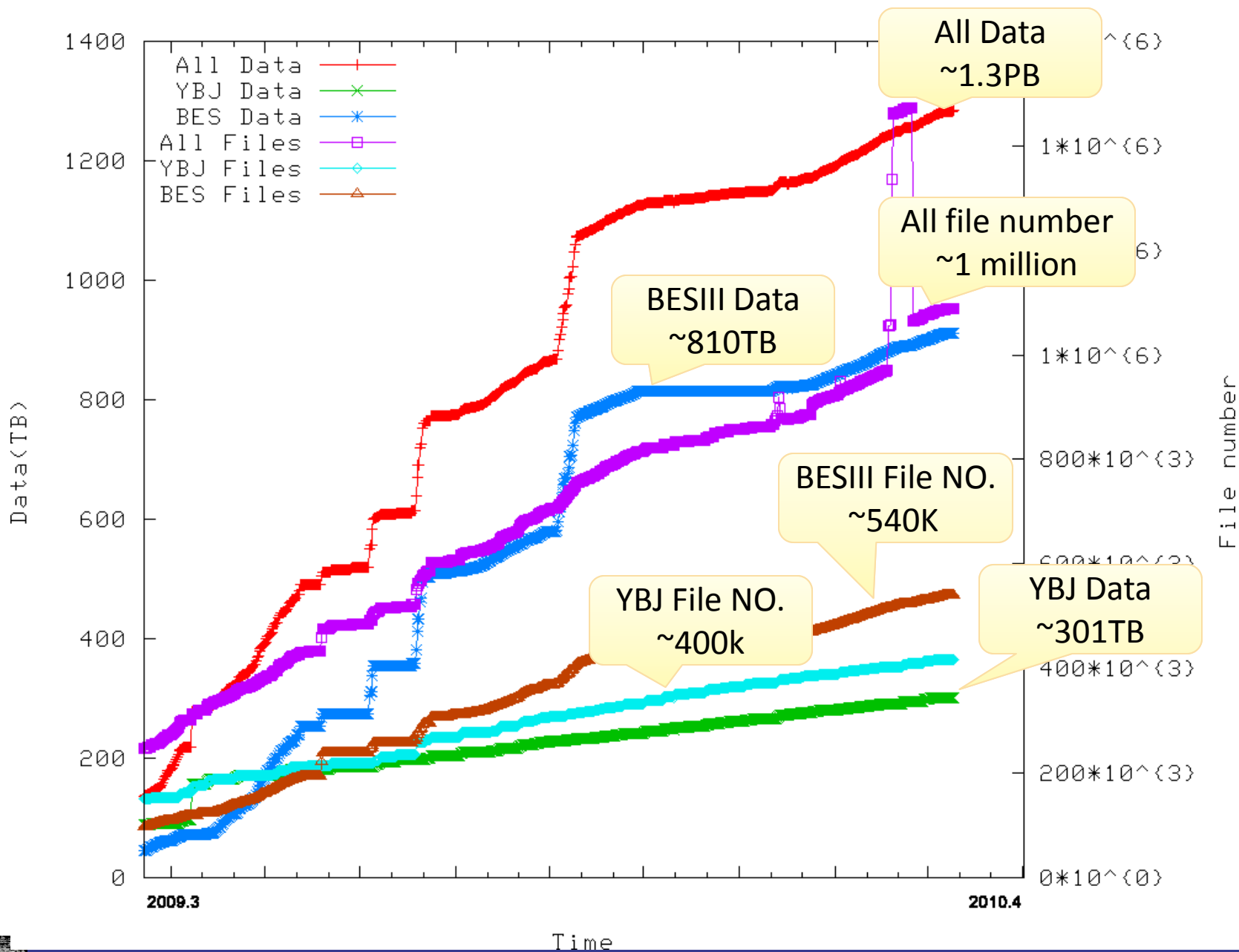
- **Software**

- 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

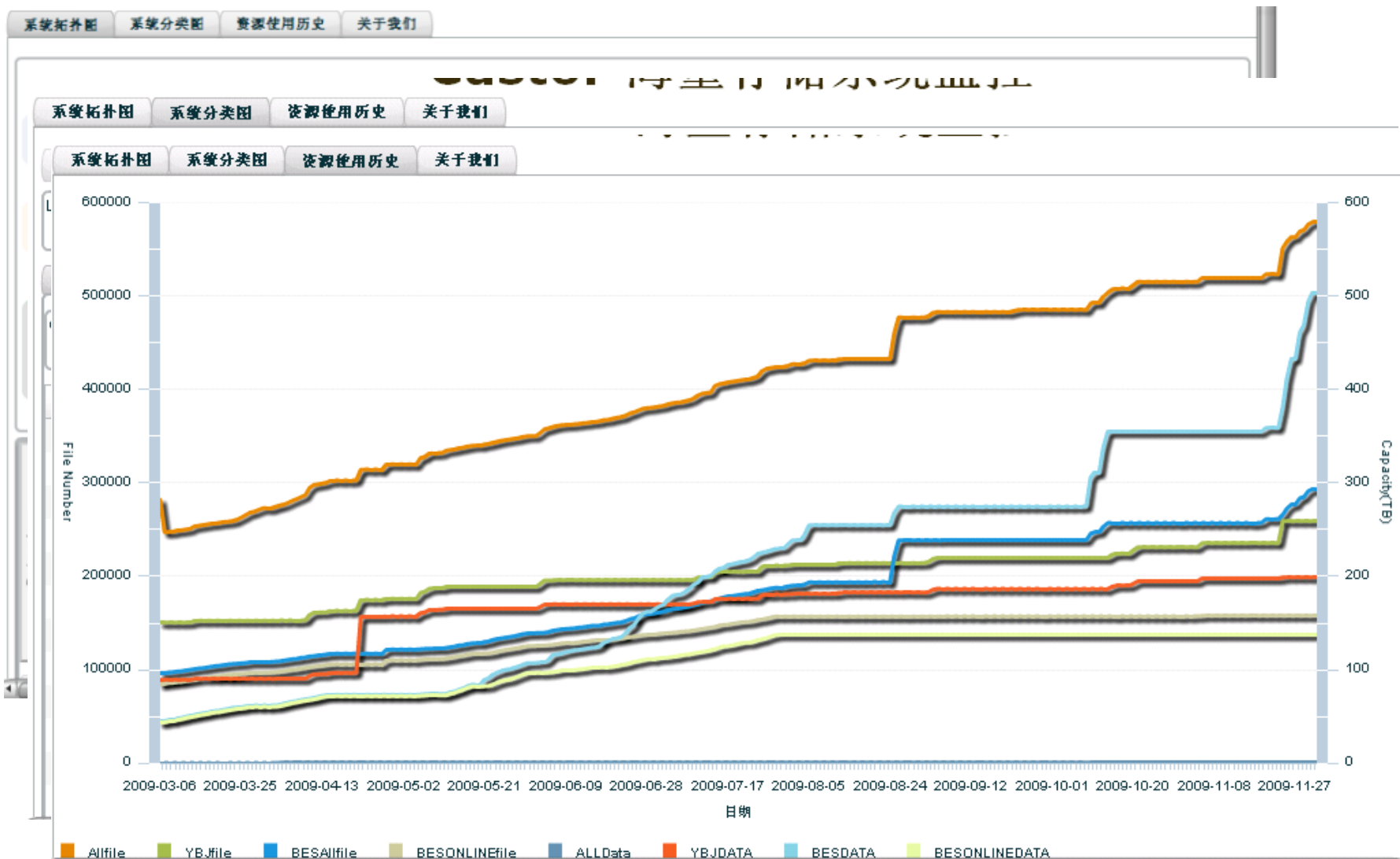
- **Network**

- 10Gbps link between disk servers and tape servers



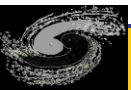


Realtime Monitoring of Castor



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



[首 页](#) |
 [文件预留](#) |
 [文件查询](#) |
 [文件修改](#) |
 [文件删除](#)

Loading data..

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

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



联系我们 | 帮助

[首 页](#) |
 [文件预留](#) |
 [文件查询](#) |
 [文件修改](#) |
 [文件删除](#)

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

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	run1704.raw	0		编辑 删除
3	raw	run1705.raw	0		编辑 删除
4	raw	run1706.raw	0		编辑 删除
5	raw	run1707.raw	0		编辑 删除
6	raw	run1708.raw	0		编辑 删除
7	raw	run1709.raw	0		编辑 删除
8	raw	run1710.raw	0		编辑 删除

需要预留的文件列表如下:

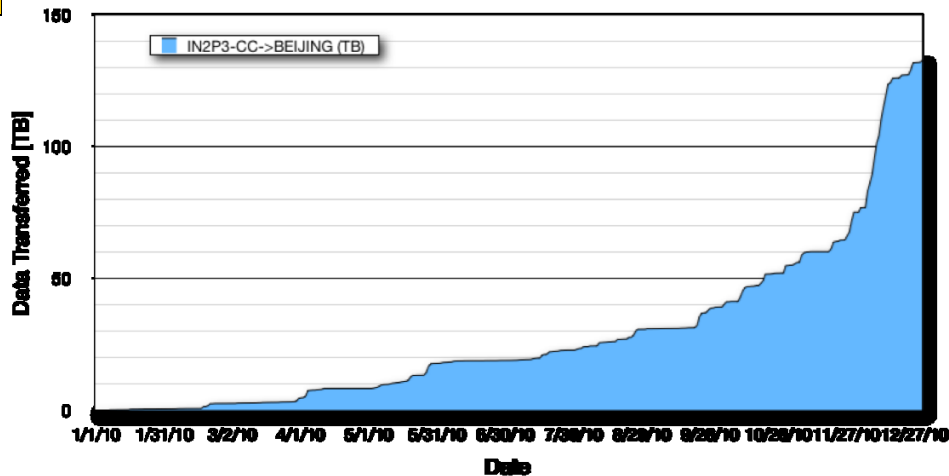
预留文件名	预留天数
/castor/ihp.ac.cn/bes/bes2/raw/17605.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1704.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1705.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1706.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1707.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1708.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1709.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1710.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1711.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1712.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1713.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1714.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1715.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1716.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1717.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1718.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1719.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1720.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1721.raw	30 天*
/castor/ihp.ac.cn/bes/bes2/raw/run1722.raw	30 天*

共有 20 条记录, 当前第 1/2 页

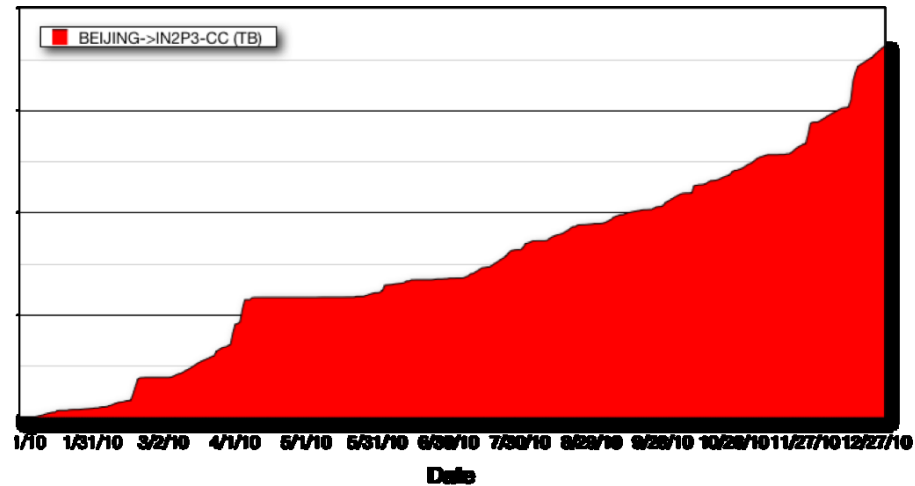
[确定](#) [重置](#)

[首页](#) [上一页](#) [下一页](#) [尾页](#) [转到第 页](#)

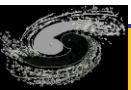
ATLAS Data transfer between Lyon and Beijing



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

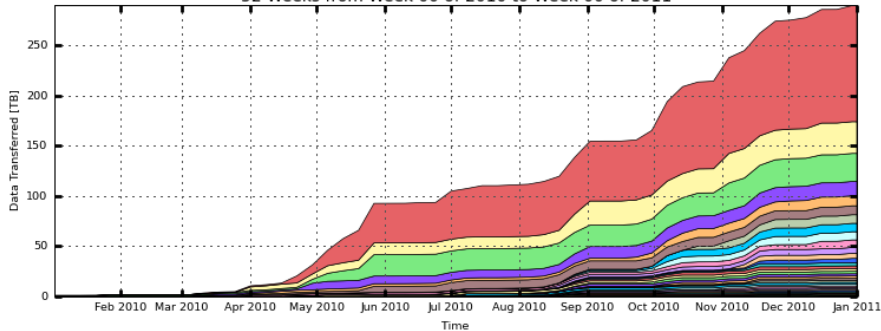


> 35 TB of data transferred from Lyon to Beijing in 2010



CMS Data transfer from/to Beijing

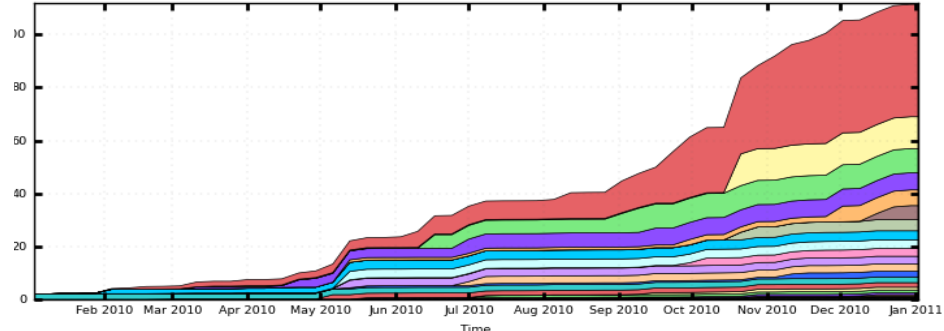
CMS PhEEx - Cumulative Transfer Volume
52 Weeks from Week 00 of 2010 to Week 00 of 2011



- T1_US_FNAL_Buffer to T2_CN_Beijing
- T1_UK_RAL_Buffer to T2_CN_Beijing
- T3_US_FNALLPC to T2_CN_Beijing
- T1_TW_ASGC_Buffer to T2_CN_Beijing
- T2_ES_IFCA to T2_CN_Beijing
- T1_ES_PIC_Buffer to T2_CN_Beijing
- T2_UK_London_Brunel to T2_CN_Beijing
- T2_US_Florida to T2_CN_Beijing
- T2_DE_RWTH to T2_CN_Beijing
- T2_FI_HIP to T2_CN_Beijing
- T1_IT_CNAF_Buffer to T2_CN_Beijing
- T1_FR_CCIN2P3_Buffer to T2_CN_Beijing
- T2_UK_London_IC to T2_CN_Beijing
- T2_US_Wisconsin to T2_CN_Beijing
- T2_US_UCSD to T2_CN_Beijing
- T2_US_Nebraska to T2_CN_Beijing
- T2_UK_London_Brunel to T2_CN_Beijing
- T2_US_Caltech to T2_CN_Beijing
- T2_ES_CIEMAT to T2_CN_Beijing
- T2_FR_GRIF_IRFU to T2_CN_Beijing
- T2_FR_IPHC to T2_CN_Beijing
- T1_DE_KIT_Buffer to T2_CN_Beijing
- T2_CH_CSCS to T2_CN_Beijing
- T2_US_MIT to T2_CN_Beijing
- T2_TW_Taiwan to T2_CN_Beijing
- T2_IT_Legnaro to T2_CN_Beijing
- T2_IT_Pisa to T2_CN_Beijing
- T2_US_Caltech to T2_CN_Beijing
- T2_FR_GRIF_LL2 to T2_CN_Beijing
- T2_RU_JINR to T2_CN_Beijing
- T2_BR_UERJ to T2_CN_Beijing
- ... plus 8 more

Total: 289.28 TB, Average Rate: 0.00 TB/s

CMS PhEEx - Cumulative Transfer Volume
52 Weeks from Week 01 of 2010 to Week 01 of 2011



- ejijing to T1_FR_CCIN2P3_Buffer
- ejijing to T2_CH_CSCS
- ejijing to T2_UK_London_IC
- ejijing to T2_US_MIT
- ejijing to T1_DE_KIT_Buffer
- ejijing to T2_ES_IFCA
- ejijing to T2_US_Nebraska
- ejijing to T2_IT_Bari
- ejijing to T2_ES_CIEMAT
- ejijing to T2_IT_Rome
- T2_CN_Beijing to T3_US_FNALLPC
- T2_CN_Beijing to T2_TW_Taiwan
- T2_CN_Beijing to T2_FI_HIP
- T2_CN_Beijing to T2_US_Caltech
- T2_CN_Beijing to T1_ES_PIC_Buffer
- T2_CN_Beijing to T2_US_Purdue
- T2_CN_Beijing to T2_EE_Estonia
- T2_CN_Beijing to T2_IT_Pisa
- T2_CN_Beijing to T2_BR_UERJ
- T2_CN_Beijing to T2_BR_SPRACE
- T2_CN_Beijing to T2_DE_DESY
- T2_CN_Beijing to T1_UK_RAL_Buffer
- T2_CN_Beijing to T2_US_UCSD
- T2_CN_Beijing to T1_US_FNAL_Buffer
- T2_CN_Beijing to T2_US_Florida
- T2_CN_Beijing to T2_US_Wisconsin
- T2_CN_Beijing to T2_AT_Vienna
- T2_CN_Beijing to T3_US_Omaha
- T2_CN_Beijing to T2_RU_JINR
- ... plus 5 more

Total: 111.41 TB, Average Rate: 0.00 TB/s

~290 TB transferred from elsewhere to Beijing in 2010

~110 TB transferred from Beijing elsewhere in 2010

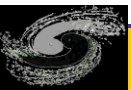
Cooling System

- 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

