

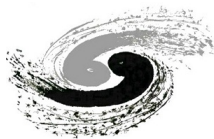
Data and Software Storage in LHAASO

2019 LHAASO 1st Conference

Yujiang Bi

on behalf of **Computing Center, IHEP, CAS**

2019.04 Nanjing



Outline



- 1 Disk Storage - EOS
- 2 Tape Storage - Castor
- 3 Software Repositories
- 4 Remote Access - XCache
- 5 Summary

Overview



- User Directories:
 - [/afs/ihep.ac.cn/users/u/user](#), default 500 MB.
- Data Storage
 - [/eos](#): primary storage for LHAASO
 - [/scratchfs](#): storing personal temp files.
 - [/workfs](#):
 - temp storage for personal files and softwares.
 - **do not** submit jobs under [/workfs](#).
- Software Repositories
 - [AFS](#)
 - [CVMFS](#)
- Tape Backup
 - LHAASO: **1 PB** in total, 5 drivers and 1700 slots.

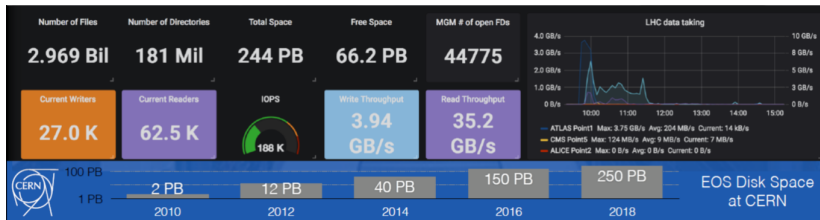
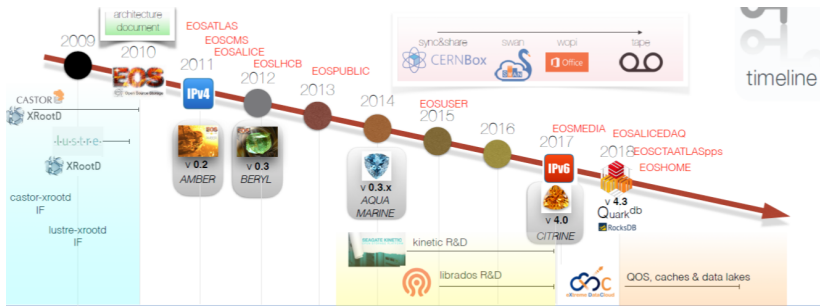
What is EOS?



- Distributed file system developed by CERN since 2010.
- In-memory MD server, supporting erasure code, disk group...
- Main system for experiments and users data, up to 250PB.
- Many EOS sites at tens of HEP labs in US, Russia, Australia.



EOS Current Status



EOS at IHEP



EOS Status at IHEP

- First EOS for LHAASO at Beijing online from 2016/02.
- Four instances: **LHASSO**, **HXMT**, **IHEPBox**, **Daocheng**.

EOS Capacities

Instance	Capacity	FSTs	Mount point	Usage
LHAASO beijing	2.39PB	40	/eos/lhaaso	Offline analysis
HXMT	350TB	7	/eos/hxmt	HXMT Exp.
LHAASO daocheng	152TB	14	/eos/daocheng	Online analysis
IHEPBox	158TB	4	-	IHEPBox storage



EOS at IHEP

EOS Problems.

- Outdated: v0.3, not supported anymore..
- Mounting via FUSE, and performance issues...
- In-memory NS and hard to scaling up.

EOS Upgrade

- Upgraded EOS LHAASO to v4.4.23, up to date.
- Mounting EOS with fusex and fuse both.
- Is eos **faster** with FUSEX mounting.

EOS Plans

- Using QuarkDB for NS, easy to scaling up.
- FUSEX on login nodes and fuse on computing nodes.
- Cancel fuse and use xrootd for computing jobs.

EOS Usage



FUSE/FUSEX Mode

- Accessing files like local files
- EOS commands like **cp**, **mv**, **ls** et. al.

```
$ cp /eos/user/b/bijj/testfile .  
$ eos cp ~/localfile /eos/user/b/bijj/  
$ eos mv /eos/user/testfile /eos/user/testfile2
```

XRootD Protocol

- No local mount endpoint
- Using **xrootd protocol** to access files.

```
$ xrdfs root://eos-server cat /eos/user/testfile  
$ xrscp xrootd://eos-server:1094//eos/user/testfile .  
$ xrscp local_file xrootd://eos-server:1094//eos/user/  
$ xrscp root://proxy//root://eos-server:1094//eos/user/file .
```




XRootD Protocol

Why XRootD Protocol?

- Far smaller failure rate than FUSE, more suitable for analysis jobs.
 - Input/Output error caused by fuse mounting failure.
 - Fuse access affected if node iowait is high.
- XRootD dominates and increasingly popular.
 - LHC experiments like Alice, CMS, ATLAS.
 - EOS has native support for XRootD protocol.
 - Tape storage CTA from CERN will support XRootD.
 - Transferring via a XRootD proxy to speed up access.

Software Preparations

- Geant4, KM2A, COSIKA, BOSS supporting or after modification.
- MDS addr <root://eos01.ihep.ac.cn> as the file path prefix.
- Using `TFile::Open(File)` instead of `TFile(File)` or `new TFile(File)`.

Data Storage Location



User directories

Directory	Space/Files Quota	Quota Usage
/afs/ihep.ac.cn/users/username	500MB/none	fs listquota \$HOME
/workfs/ybj/username	5GB/50,000	lfs quota -u username /workfs
/scratchfs/ybj/username	500GB/none	lfs quota -u username /scratchfs
/eos/user/a-z/username	1TB/250,000	eos quota /eos/user/a-z/username

Data directories

Directory	Space/Files Quota	Usage	Owner
/eos/lhaaso/raw	none/none	original data	Computing Center
/eos/lhaaso/spt	none/none	single particle data for each exp.	Computing Center
/eos/lhaaso/decode	none/none	decode data	Assigned by LHAASO
/eos/lhaaso/rec	none/none	reconstructing data	Assigned by LHAASO
/eos/lhaaso/cal	none/none	calibration parameter data	Assigned by Each Exp.
/eos/lhaaso/monitor	none/none	Monitoring data for data quality, moon shadow and Crab	Assigned by LHAASO

Account Applying



- Applying an AFS account if not. Click [here](#) to apply.
- Need to be approved by the supervisor or director.

AFS Account Application

Name *	<input type="text"/>	Full Name *	<input type="text"/>
Sex	<input checked="" type="radio"/> male <input type="radio"/> female	Phone *	<input type="text"/>
Email *	<input type="text"/>	Project group	<input type="text"/>
Shell Type *	<input checked="" type="radio"/> bash <input type="radio"/> tcsh <input type="radio"/> csh		
Application *	<input type="text"/> Read Me	User Group *	<input type="text"/>

contact person (supervisor/primary investigator)info

Name *	<input type="text"/>	Email *	<input type="text"/>
Phone	<input type="text"/>		

Remark

Security Code SOBVD8

If you can not see the security code, please use Internet Explorer(IE). [Help](#)

Castor



Castor Service

- A hierarchical storage system based on Castor 1 by CERN.
- Stores files accessed infrequently and kept for a long time.
- A file system integrating disks and tapes, a unified namespace.
- Transfer data between disks and tapes by file access frequencies.

Status for LHAASO

- 2 IBM tape libraries.
- **1700** tape slots and 5 tape drivers.
- **1 PB** in total and more in the future.





Castor Usage

Castor Usage

- Ready and will be used for backup late this year.
- DO NOT upload small files like logs and scripts.($< 80\text{M}$)
- Or pack small files up and upload the tarballs.

Examples

1. List files

```
$ nsls /castor/ihep.ac.cn/lhaaso
```

2. Download directories or files

```
$ rfcpx /castor/ihep.ac.cn/lhaaso/raw/ed/2019/09/ /tmp
```

3. Upload directories or files

```
$ rfcpx /etc/hosts /castor/ihep.ac.cn/lhaaso/test/
```

4. Delete files on castor

```
$ rfrm [-r] /castor/ihep.ac.cn/lhaaso/test/testfile
```

Plans

- New uploading/downloading system and will be online later..



AFS Service

Use Scenario

- Managing accounts for all users of IHEP and collaborators.
- User HOME directory: [/afs/ihep.ac.cn/users/u/user](#).
- Software repository: [BESIII](#), [LHAASO](#), [Dayabay](#), [JUNO](#), [LQCD...](#)

Policy

- Default quota for user home: **500MB** (applying more in need)
- Support replication for software repos, and default is **1**.
- Publish updates if software repos has multi-replicas.

LHAASO

- Software Volume
 - [/afs/ihep.ac.cn/soft/YBJ](#), **50 GB**, 3 replica.
 - [/afs/ihep.ac.cn/soft/LHAASO](#), **500 GB**, 1 replica.
- Readable for all users, modifiable only to software administrators.

CVMFS Service



CERNVM-FS

- Read-only FS and supporting fuse mode.
- Caching files to local disks to speed up access.
- Replacing [AFS](#) for hosting frequently used softwares.

LHAASO

- Software Volume: **`/cvmfs/lhaaso.ihep.ac.cn`**
- Release controlled by software repos administrators.
- Release procedure seeing [User Manual](#).

Advices

- Putting frequently used scripts/softwares in **CVMFS** but **AFS**.
- **DO NOT** storing big files larger than **1 GB** if possible in CVMFS.

XCache



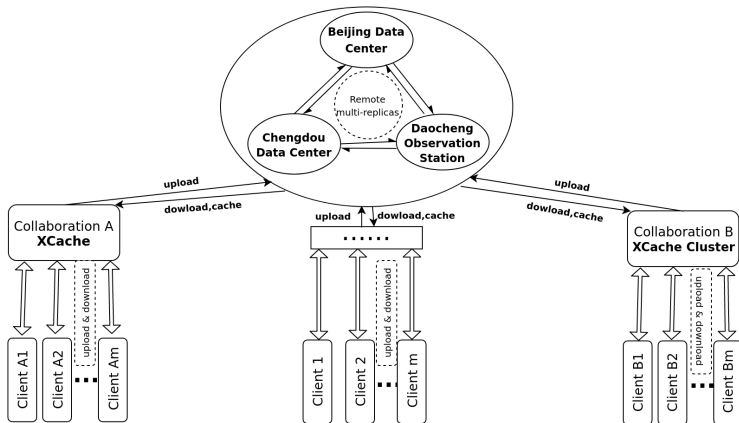
Problems of Remote Access

- Many collaborations distributed worldwide.
- Large latency (> 100ms) for long distance accessing
- Limited network bandwidth.
- Waste of network for repetitive downloading files.
- One Possible solution - XCache.

XCache - XRootD Proxy Storage Service.

- Optimizing network latency and saving bandwidth.
- Two mode:
 - Pre-fetch and cache complete files.
 - Download file blocks on-demand
- XCache Cluster for service increasing and load balancing.

XRootD Proxy Cache



- XCache deployment seeing at [Proxy Storage Services](#).
- Technical support provided by [IHEP Computing Center](#).



Summary

Current Status

- User Home in AFS: [/afs/ihep.ac.cn/users/](https://afs.ihep.ac.cn/users/)
- Data Storage:
 - EOS Beijing and Daocheng for storing experiment data
 - Castor system for data backups.
- AFS & CVMFS serve as softwares storage.
 - Two volumes on AFS: [/afs/ihep.ac.cn/soft/ybj,lhaaso](https://afs.ihep.ac.cn/soft/ybj,lhaaso)
 - CVMFS volume: [/cvmfs/lhaaso.ihep.ac.cn](https://cvmfs/lhaaso.ihep.ac.cn)
 - Try to using CVMFS to storing softs.
- Data Remote Access
 - Using XCache sites to speed up file accessing remotely.

Overlook

- Scaling up EOS and Castor Capacities for LHAASO.
- New Castor uploading/downloading system online.
- Deploying XCache sites for collaborations.

Thanks



Q & A

