

LBNE Data Handling

Qizhong Li

Fermilab

Jan. 31, 2014

LBNE Software Tool Workshop

Outlines

- Introduction
- Tape Storage:
 - SAM
 - Enstore
- Disk Storage:
 - BlueArc
 - dCache
- Summary and Plans

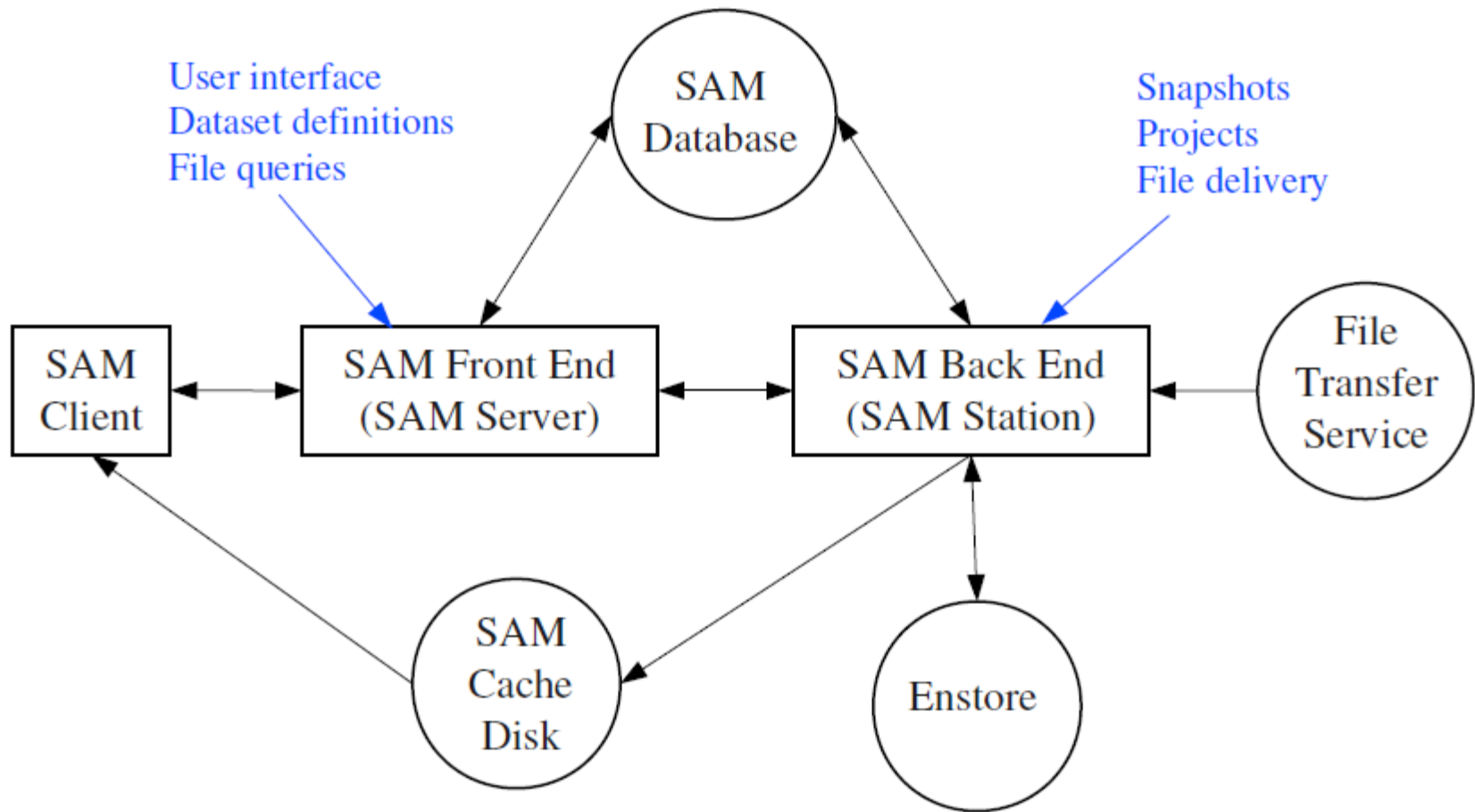
Data Handling Overview

- SAM
 - File catalog (database) consisting of file metadata and locations.
 - Organizes file delivery to worker nodes (via sam station).
 - Provides user interface to tape system (enstore).
- File transfer service
 - Provides interface for uploading files to Enstore and updating SAM database.
- Enstore
 - Fermilab tape mass storage system.
- dCache
 - Provides cache disk for staging tape-resident files.
- FNAL SCD provides excellent support teams.

SAM

- SAM serves as file catalog.
 - SAM database holds information about data files (metadata).
- Users interact with SAM http server (samweb).
 - SAM commands are sent to sam server via http port.
- SAM handles staging of tape files to cache disk as necessary.
- SAM knows how to intelligently schedule file delivery to many worker nodes simultaneously.
- SAM is being used by almost all intensity frontier experiments.
 - Also used by Tevatron collider experiments.
- SAM supporting team is very strong, a lot of experience with handling user problems.

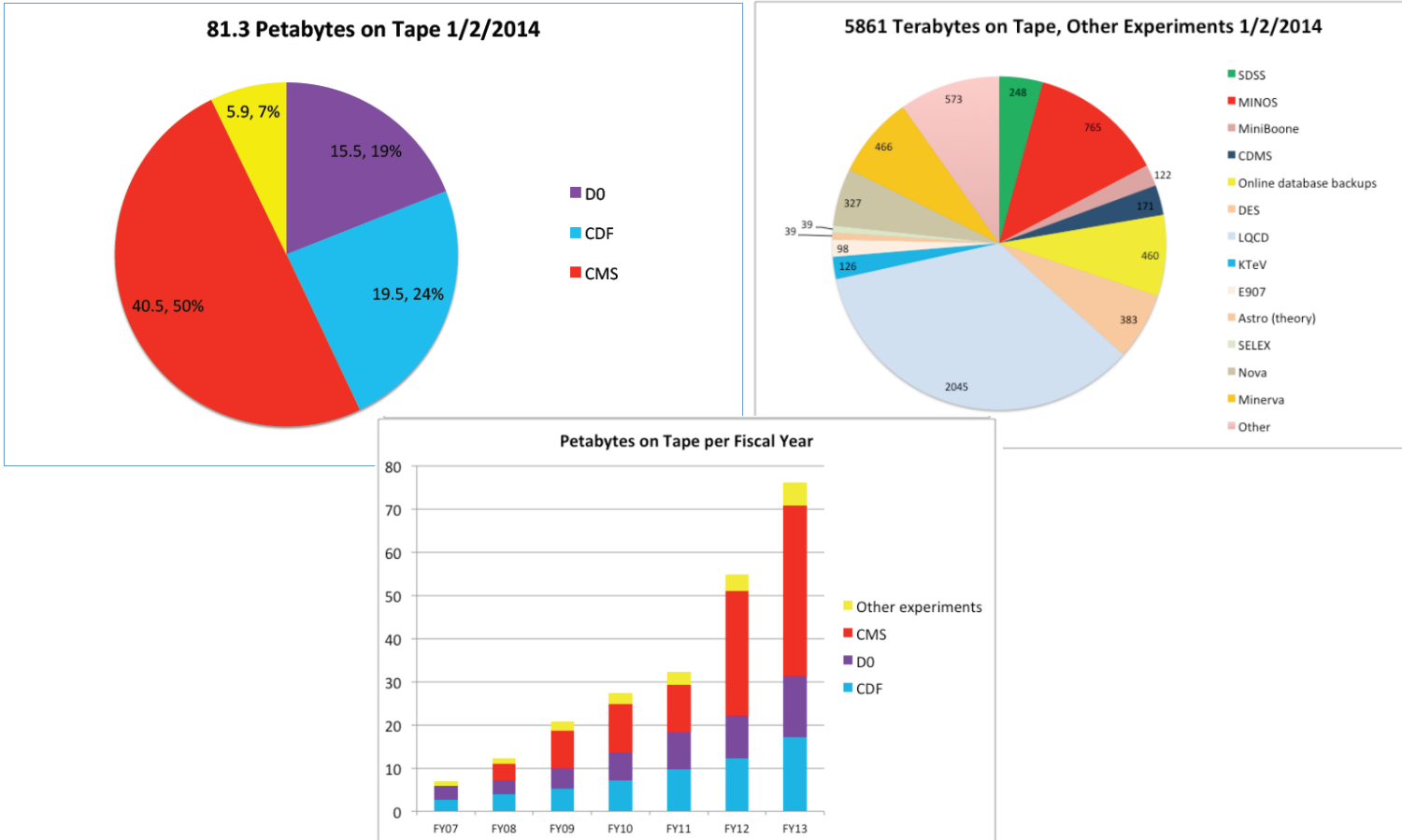
How SAM Works



Enstore

- Enstore is the mass storage system developed and implemented at Fermilab as the primary data store for scientific data sets.
- It provides access to data on tape to/from a user's machine on-site over the local area network, or over the wide area network through the dCache disk caching system.
- It is a very reliable tape storage system with a strong support team
 - on-call 24x7 and
 - the tape libraries and tape drives have 24x7 support from the vendor.

Enstore Tape Usage



All FNAL experiments store data in Enstore.
 Total more than 80PB data on tape, as of Jan. 2014.

Progress on lbne Data Handling

- LBNE metadata is defined
 - (see DocDB **LBNE-doc-8093**)
- SAM and Enstore library for lbne are all setup
- More than 3000 MC files are stored into SAM:
 - 35t single particle MC
 - 10kt single particle MC
 - GenieCry MC
 - Babu's single particle MC
- All these files can be read back from SAM.
- Recently setup dCache as lbne SAM cache.
- We are working on creating a scratch area on dCache for users.

To Get Files from SAM

- Dataset definitions:
 - Users can make dataset definition by your own requirements (need to use metadata parameters)
 - Or to use existing dataset definitions
 - For example. I defined the following dataset definitions for general user to use:

genieCRY_20130423_anuflux_nuebarflux_nuebar_g280_Ar40

genieCRY_20130423_anuflux_nueflux_nue_g280_Ar40

genieCRY_20130423_anuflux_numubarflux_nuebar_g280_Ar40

genieCRY_20130423_anuflux_numubarflux_numubar_g280_Ar40

...

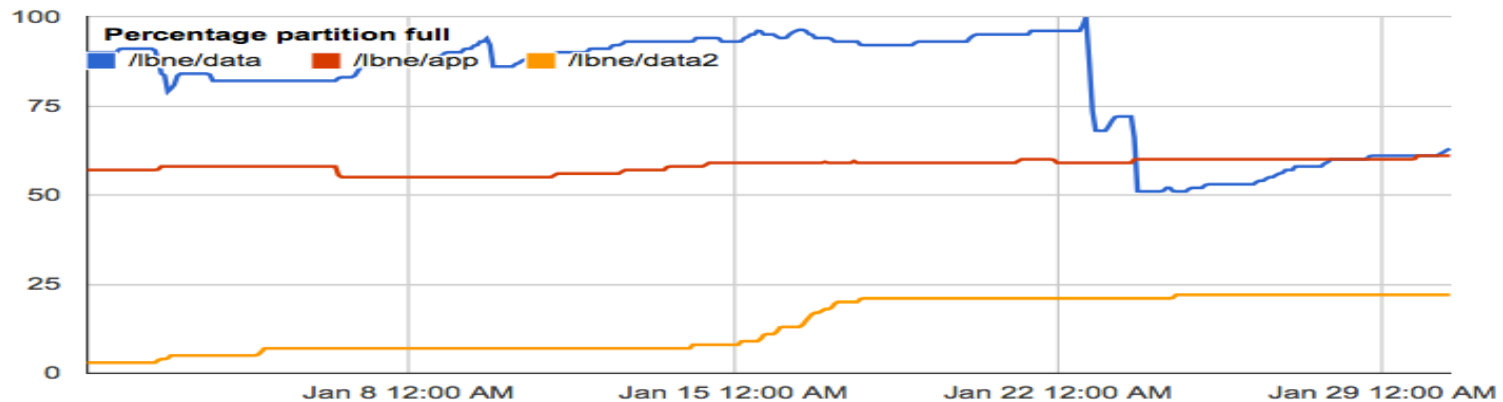
BlueArc Disks

- Lbne has two BlueArc data disks:
 - /lbne/data (30TB)
 - /lbne/data2 (20TB, new)
- Currently most people use /lbne/data, which was above 90% full in the last two months.
 - It was 100% full, 10 days ago.
 - General e-mails were sent asking people to clean up. Didn't have much effect.
 - By contacting individual users (many thanks to Tom!) made big difference.
- Users are encouraged to use /lbne/data2 (the new disk).

BlueArc Disks

- BlueArc disk space usage in January:

BlueArc for lbne



- Several experiments share a BlueArc controller.
 - If one experiment's BlueArc area is full, may hang other experiment's BlueArc access.

Batch and BlueArc

- We found some users are directly accessing BlueArc disk from batch jobs. This is not good practice.
- Batch jobs should write output to local worker scratch disk, and copy to BlueArc at end of job using “ifdh cp” command.
- Same for input. Batch jobs that read files from BlueArc should copy the entire file from BlueArc to local worker scratch at beginning of job using “ifdh cp”.

dCache Disks

- Fermilab has newly setup a very large dCache area for Intensity Frontier experiments
 - The new dCache has ~4 PB disk space
 - dCache is mainly used as SAM cache area, shared by IF experiments
 - There will be also a public scratch area on dCache, which can be used by users as a disk storage area.
 - Scratch area is also shared by IF experiments
 - We are working on setup `/pnfs/lbne/scratch`.

dCache Scratch Area

- From SLF6 machines, to access the scratch area is very easy:
 - > `ls /pnfs/lbne/scratch`
- From SLF5 machines, users need to access it using `ifdh cp` command (through `gridftp`, `xrootd`,...)
 - > `ifdh cp /pnfs/lbne/scratch/username/file ...`
- The scratch area, by definition, is temporary. It will delete “old files” when it is full.
 - But 4 PB will take long time to fill.
 - So far IF experiments data are not very large.
 - Therefore the scratch area can help us to relieve pressure on BlueArc disks.

Summary and Plans

- Data handling for lbne has been progressing well.
- More than 3000 MC files are stored into SAM.
 - They are on tape, and can be fetched to disk.
- Files in SAM can be read back.
- lbne dCache area is recently created as SAM cache.
- New BlueArc disk, /lbne/data2, (20TB) is available.
 - Please clean up non-needed files from /lbne/data
 - Should not let BlueArc disk usage exceed 90%
- Work in progress on setting up the scratch area on the very large dCache disks for users to do analysis.
- Working on generating metadata from the software program automatically (Herbert Greenlee).
- Several institutions are interested in storing 35t data files into their institution storage.