An efficient, modular and simple tape archiving solution for LHC Run-3

V Bahyl, G Cancio, E Cano, D Kruse, J Leduc and S Murray – CERN/IT
V Kotlyar – IHEP Russia

The tape solution for LHC Run-3 physics

Integrate tape with the EOS disk system

Preemptive tape drive scheduler

Clean separation between disk and tape

Motivations for the CERN Tape Archive (CTA)

Archive 100 petabytes per year

LHC Run 2 currently archives approximately 40 petabytes per year.
LHC Run 3 is predicted to more than double this demand.

Add tape to the de facto disk storage system for physics - EOS

CTA will enable EOS to directly archive and retrieve files to and from tape.

Avoid functional duplication through a clean, consolidated separation between disk and tape

EOS will focus on providing high-performance disk storage, end-user access [protocols/tools] and the namespace. CTA will focus on providing the tape back-end.

Easy migration of data from CASTOR

Today CASTOR stores the custodial copy of all LEP and LHC physics data. This data needs to be safely, painlessly and efficiently migrated.

Achieve full speed data transfers, full time

Preemptive drive scheduling will keep tape drives fully occupied.

Provide an additional disaster recovery solution

Tape storage is a safety net. CTA should not only store the contents of physics files it should also permit the storage and update of disk file metadata for the purpose of disaster recovery.

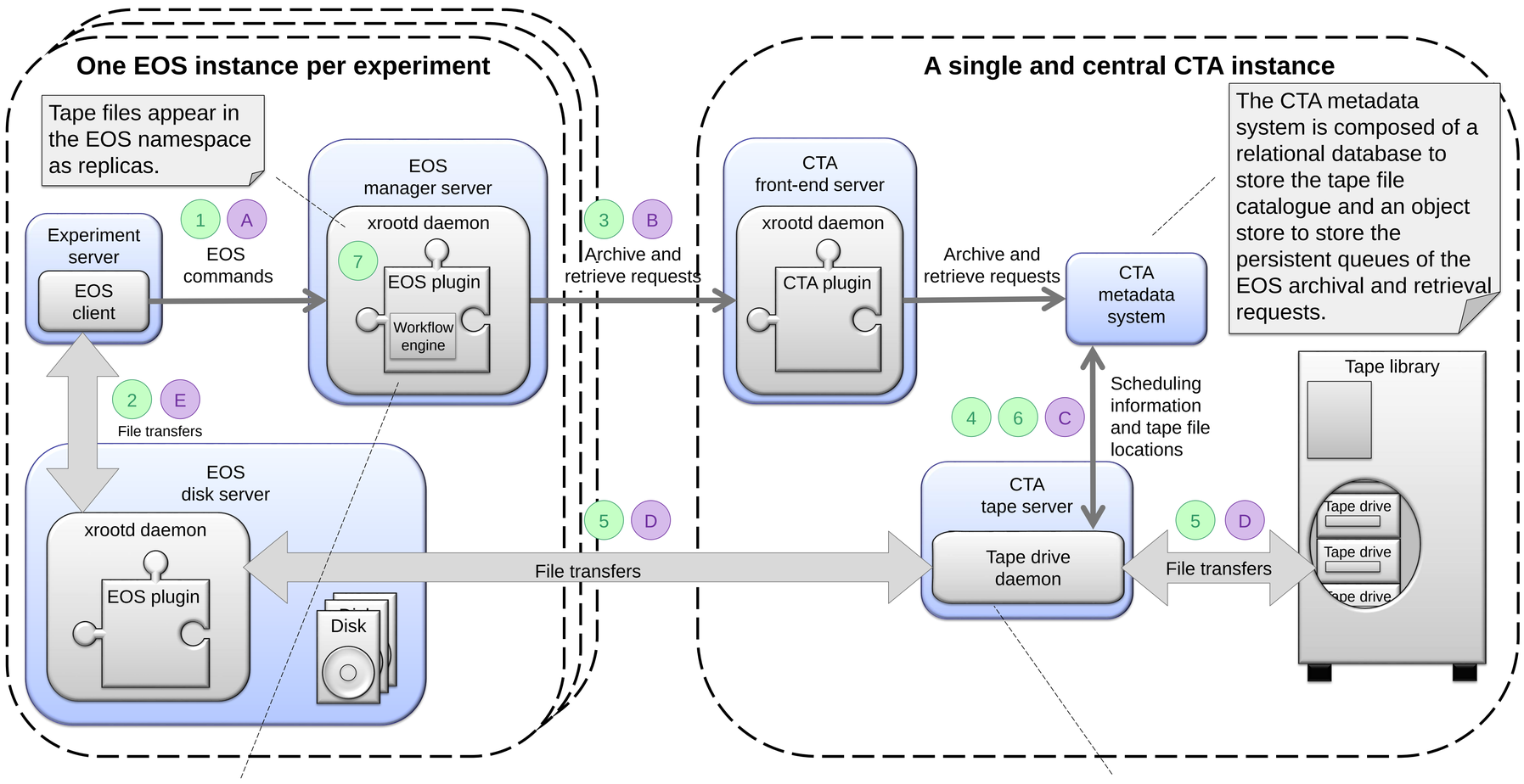
Ready for new performance technologies

CTA will actively add new and proven tape features to its tape server as they become available.

Ready for the new logging infrastructure

CTA will actively take advantage of technologies outside of the tape server such as providing tailored log messages to facilitate new log analysis technologies and associated alarming systems.

The architecture of the CERN Tape Archive



How a file is archived

1. Experiment creates an EOS directory configured to use an EOS workflow that automatically archives files to tape.
2. Experiment transfers a file into the newly created directory.
3. On the close of the file the EOS workflow engine queues an archive request with the CTA front end.
4. Tape server pulls the archive request from the CTA metadata system.
5. Tape server copies the file from disk to tape.
6. Tape server stores the tape-file location in the CTA metadata system.
7. EOS or user deletes the copy of the file on disk.

How a file is retrieved

A. Experiment requests a file from EOS.
B. EOS workflow engine queues a retrieve request with the CTA front-end.
C. Tape server pulls the retrieve request from the CTA metadata system together with the location of the tape file.
D. Tape server copies the file from tape to disk.
E. Experiment reads the file from the EOS disk server.

The EOS workflow engine can be configured to provide:
- D1T0 – Disk only files.
- D1T1 – Files replicated on both disk and tape.
- D0T1 – Tape files cached on disk.
- Blocking or “bring-online” tape file retrievals.

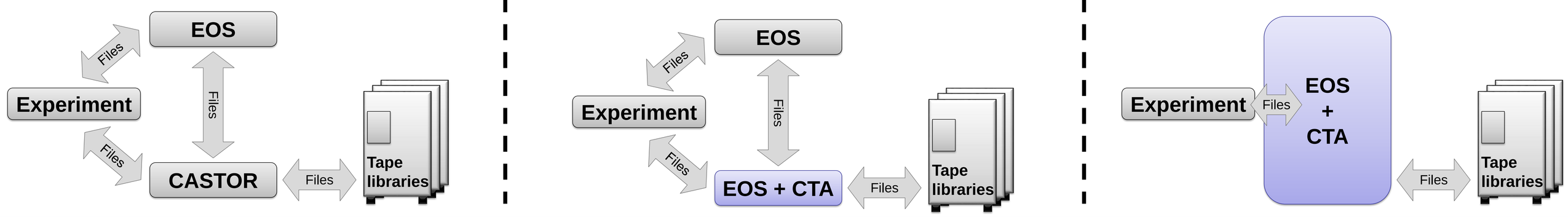
The EOS garbage collector can be configured to delete disk files that have been safely archived to tape.

The tape drive daemon from CASTOR will be augmented with a new preemptive tape drive scheduler that provides:
- Full speed data transfers, full time:
  - CASTOR achieved full tape drive speed.
  - New scheduler will run tape drives full speed, full time by keeping otherwise idle drives occupied with the preemptible background tasks of tape verification and repack.
- Fair load balancing across tape drives through the late binding of drives and tapes.

Easy migration from CASTOR

- The CASTOR namespace will be migrated to EOS and the CTA tape file catalogue.
- The CASTOR tape format will be reused.
- Only metadata will need to be migrated.
- No files will need to be copied between tapes.

Migrating from CASTOR



CTA Milestones

Q4 2016	Bare functionality working with physical tape drives and libraries.
Q2 2017	Optimized for production. Ready for redundant use cases such as additional backups of filer data (AFS/NFS) and additional copies of LEP data.
Q2 2018	Tape media repack functionality. Ready to migrate small VOs such as non-LHC experiments from CASTOR to EOS + CTA.
Q4 2018	Full, robust, scalable system. Ready to migrate large virtual organizations such as LHC experiments from CASTOR to EOS + CTA.

Contact: Steven.Murray@cern.ch