CERN Tape Archive (CTA) : Experiences from the Migration of the LHC Experiments

Dr. Michael Davis
Project Leader
CERN Tape Archive

The archival storage solution from the CERN IT Storage Group

CERN
Tape Archive

The archival storage solution from the CERN IT Storage Group

CTA is the tape back-end to EOS

The archival storage solution from the CERN IT Storage Group

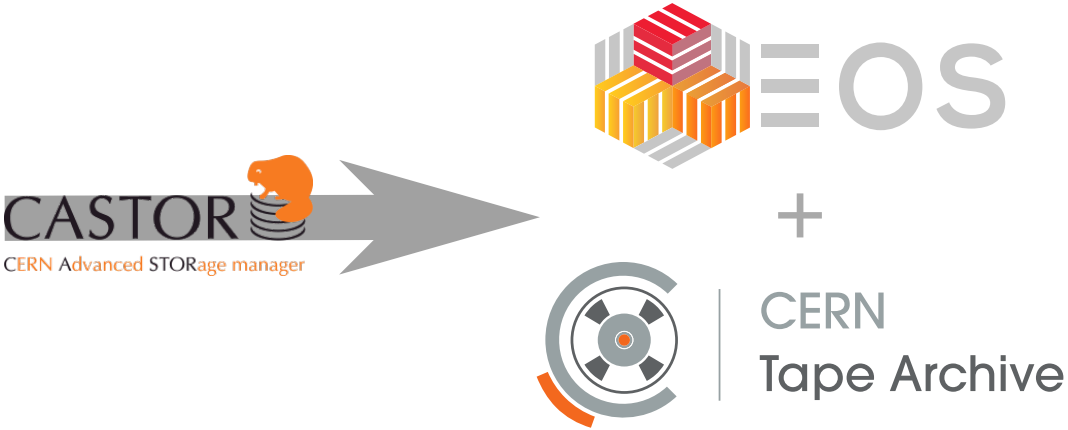
EOS +

CTA

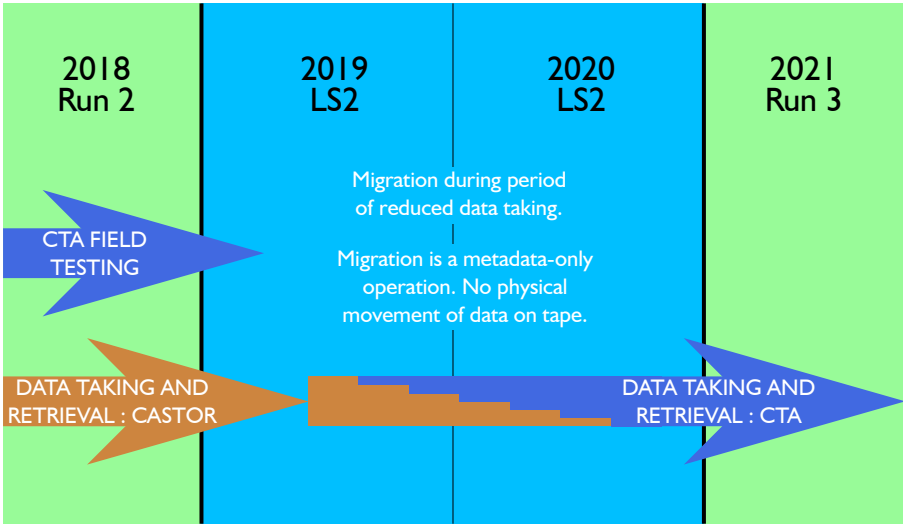
+

FTS

The four LHC experiments have been migrated from CASTOR to CTA…



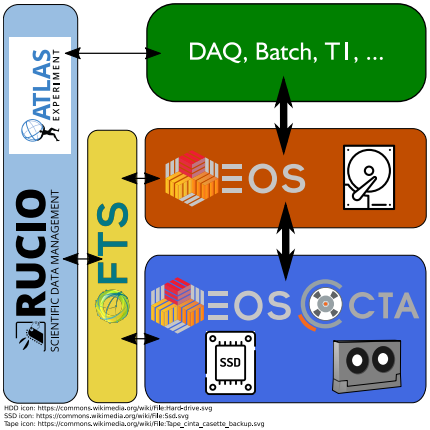
…on schedule (more or less)



EOS+CTA Deployment at CERN Tier–0

- 4 instances for the LHC experiments
  - ATLAS
  - ALICE
  - CMS
  - LHCb
- 1 instance for PUBLIC
  - Active non-LHC experiments
    AMS, CAST, CLIC, Compass, Dune, NA61/SHINE, NA62, nTOF, …
  - Data Preservation
    BaBar, LEP-era experiments (*Aleph*, *Chorus*, *Delphi*, *Harp*, …), others
  - Certain backup use cases
  - User directories

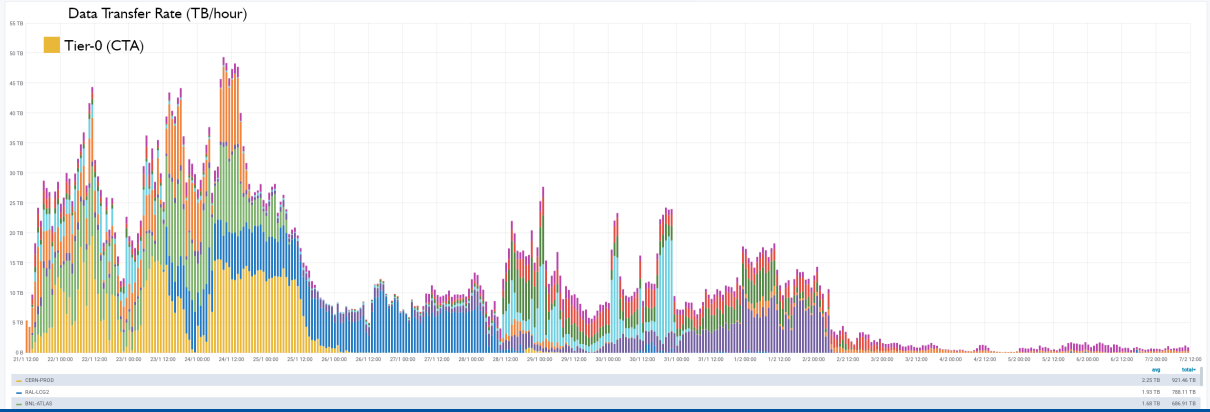
ATLAS Migration (Jan–Jun 2020)



- Rucio + FTS + EOS + CTA
- **EOS ATLAS**: "Big EOS". Spinning disks. Storage accounted to pledge. Data on disk is long-lived.
- **EOSCTA ATLAS**: "Little EOS". SSDs. Small, fast buffer with no contention. Data is short-lived.

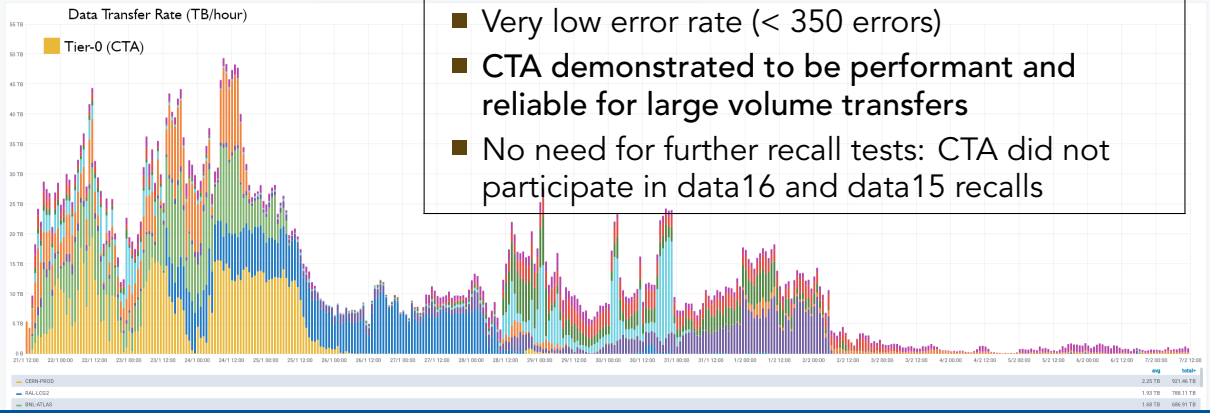
ATLAS Reprocessing Campaign

January: Recall of Data18
February: Recall of Data17



ATLAS Reprocessing Campaign

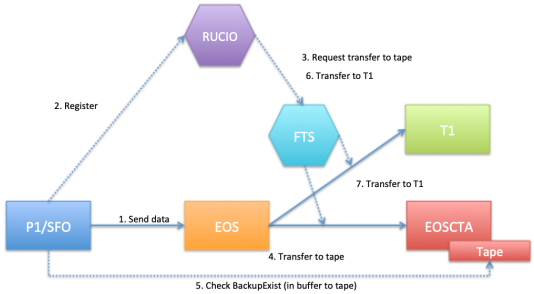
- 5 PB recalled @ 4–5 GB/s
- Very efficient tape recall (80% tape drive efficiency for Enterprise drives)
- Very low error rate (< 350 errors)
- **CTA demonstrated to be performant and reliable for large volume transfers**
- No need for further recall tests: CTA did not participate in data16 and data15 recalls

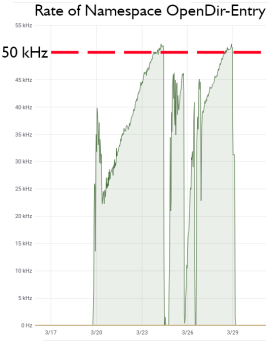


Recommended Access Order (RAO)

- Enterprise drives (with RAO) read speed during data17 recall 250–280 MB/s
- LTO-8 drives (without RAO) read speed during data17 recall 70–110 MB/s
- Software RAO for LTO has been implemented in CTA and deployed in production

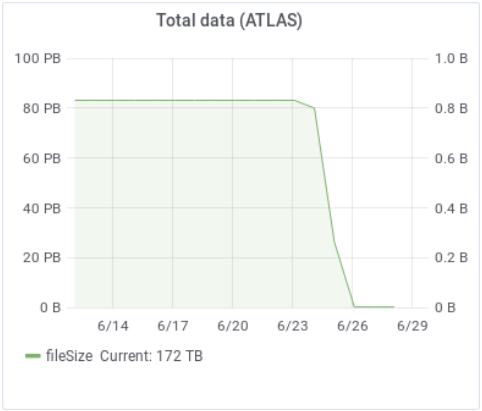
Mar/Apr : TDAQ/SFO Integration Tests



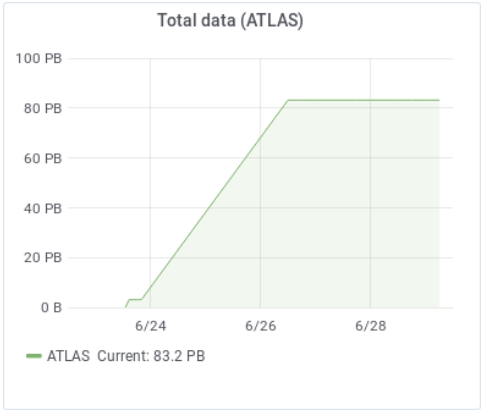


- "File safely on tape" check caused a high rate of metadata queries to EOS
- Problem has been fixed using new FTS Archive Monitoring feature (see HEPiX 2020 presentation, FTS: Towards tokens, QoS, archive monitoring and beyond)

23–25 June 2020 : ATLAS Migration

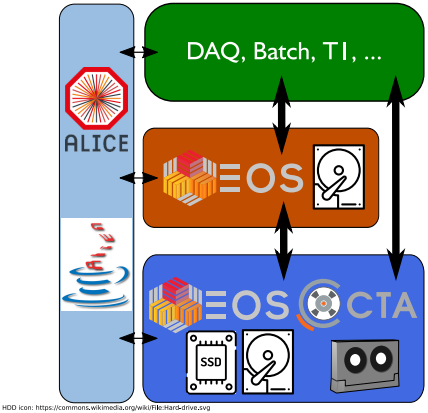


CASTOR



CTA

ALICE Migration (Jul-Oct 2020)

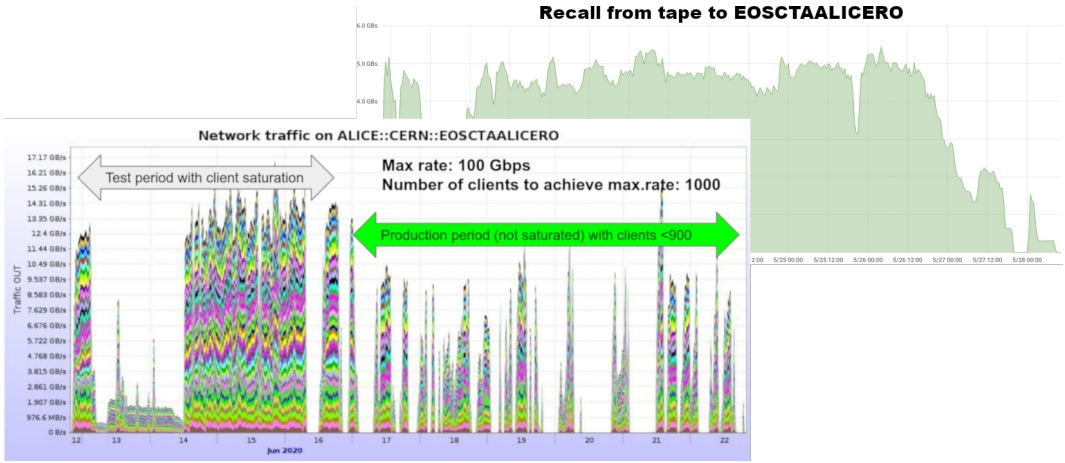


HDD icon: https://commons.wikimedia.org/wiki/File:Hard-drive.svg
SSD icon: https://commons.wikimedia.org/wiki/File:Ssd.svg
Tape icon: https://commons.wikimedia.org/wiki/File:Tape_cinta_casette_backup.svg

- JAlien + EOS + CTA
- Dual-space buffer for recalls :
  - Small tape buffer (SSD)
  - 5 PB recall-only cache, accounted to ALICE pledge (HDD)

ALICE Tests

- Testing of ALICE-specific configuration with 5 PB disk buffer for recalls



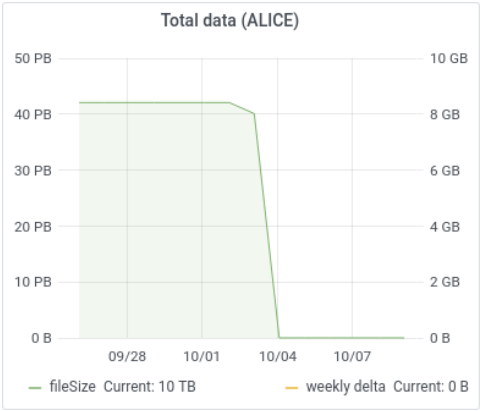
ALICE Tests : Findings

- 5 PB disk cache requires spinning disks
- Tape recall to HDDs less efficient than recall to SSDs due to contention on write
- Tape drives starving and dismounting tapes

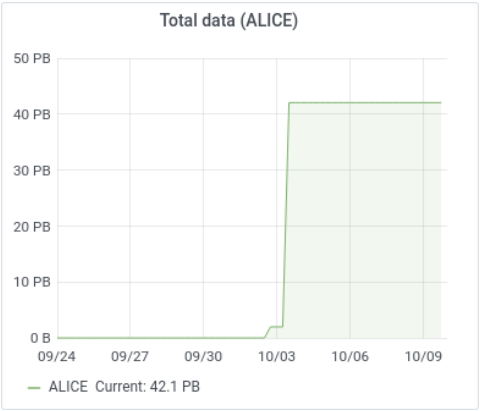
Solution

- Fast recall from tape to SSD, EOS conversion gradually moves files from SSD space to spinner space
- Disk files in the cache are single replica

2–4 Oct 2020 : ALICE Migration

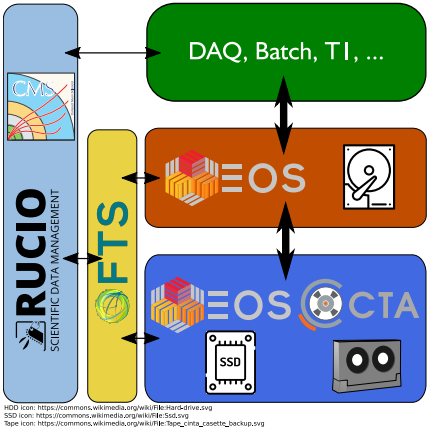


CASTOR



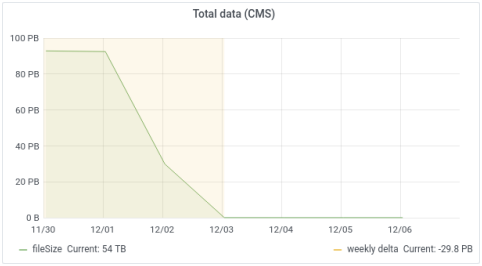
CTA

CMS Migration (Oct–Dec 2020)

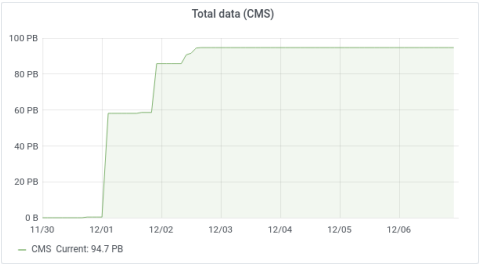


- Similar setup to ATLAS : Rucio + FTS + EOS + CTA
- Dual migration : Phedex to Rucio and CASTOR to EOS+CTA
- Integration with FTS Archive Monitoring feature

1–3 Dec 2020 : CMS Migration

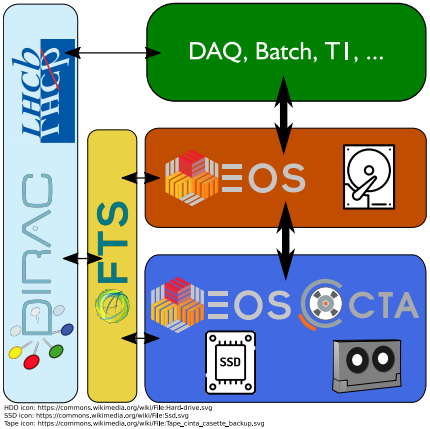


CASTOR



CTA

LHCb Migration (Jan–Mar 2021)

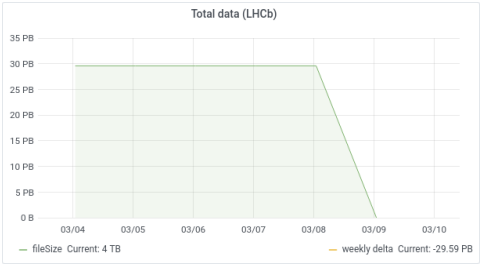


- Similar setup to ATLAS and CMS : Dirac + FTS + EOSCTA
- Additional requirement to make Third Party Copy (TPC) transfers between T0 (EOS+CTA) and T1 SEs

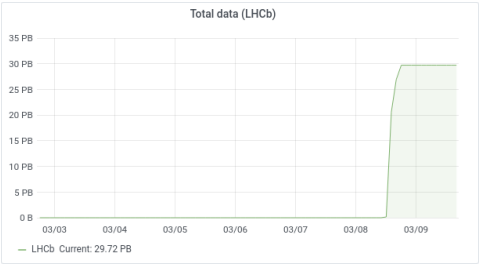
LHCb Migration (Jan–Mar 2021)

- Integration with Dirac+FTS straightforward
- Transfers between T0 and T1
  - CTA does not support SRM
  - XRootD TPC with delegation of credentials: required software upgrades at the T1s **DONE**
  - StoRM does not support XRootD TPC, but does support HTTP TPC
  - HTTP lacks an equivalent to SRM BringOnline/XRootD Prepare
  - Short-term solution: FTS stages file with XRootD Prepare, then transfers with HTTP TPC **DONE**
  - Long-term solution: Add staging functionality to HTTP based on the dCache REST interface **IN PROGRESS**

8–9 Mar 2021 : LHCb Migration



CASTOR



CTA

Migration to CTA: Summary

- ATLAS, ALICE, CMS and LHCb have all been migrated from CASTOR to CTA
- Each experiment presented its own unique use cases and problems which had to be solved
- Additional tests and data challenges will be carried out during 2021 to ensure readiness for Run–3
- The respective four CASTOR instances are in the process of being decommissioned

CTA Website : Publications, Presentations, Documentation and Source Code


CERN