CTA Status and Plans

CERN update for pre-GDB Tape Evolution

Richard Bachmann, on behalf of the CTA team

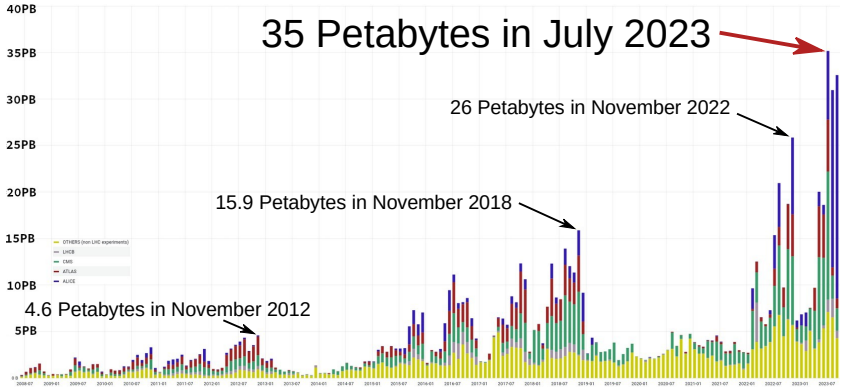
What is the CERN Tape Archive (CTA)?

CTA:

- Archive for physics data
- FOSS
- Pure tape system, using EOS as back-end for disk
- Provisioned capacity: ~840PB

New CERN Record

Data archived to tape storage each month since 2008

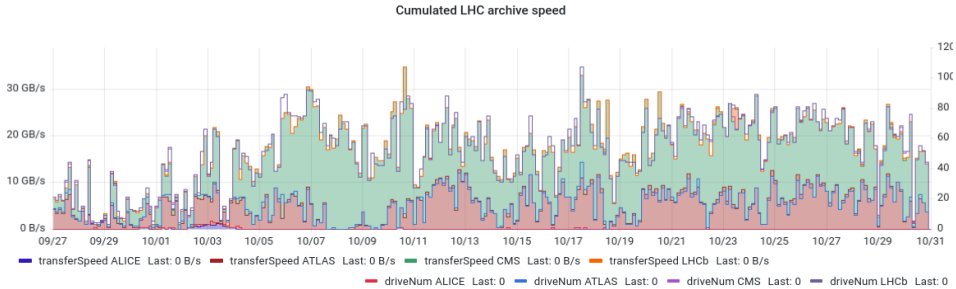


Data on Tape 2004–2023

CERN
Tape Archive

- +205 PB since start of Run–3
- +125 PB in last 12 months
- Does not include 2023 Heavy Ion run

2023 Data Taking —Heavy Ion



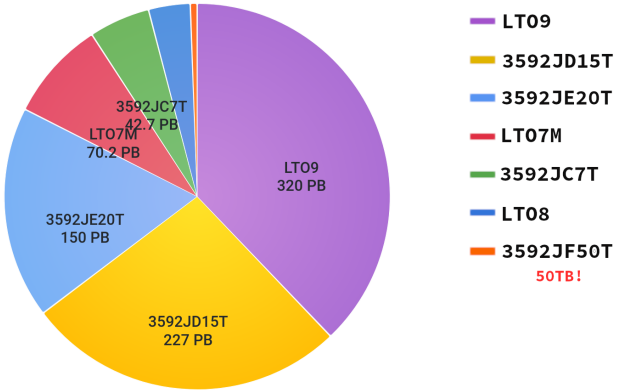
Hardware Inventory

6 Libraries:

- 4x IBM TS4500
- 2x Spectra Logic TFinity

184 Drives:

- 1:1 drive to *Tape Server* mapping
- *New:* IBM TS1170



Operations Achievements 2023

HTTP REST API rollout: Done

- Allows HTTP transfers for tape, including *disk eviction* and *checkOnTape*
- Deployed in Q1 2023 for all interested parties: ATLAS, LHCb

EOS 5 deployment:

- Complete for non-LHC experiments
- The rest will follow now that Heavy Ion is complete
- CTA 5 + XRootD 5 deployment after that

CTA Features and code

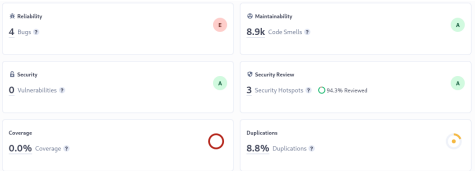
Public releases

- Introduce Physical Library concept
- Introduction of dedicated Repack VO
  - Separate tape-lifecycle operations from VO resource quotas
- New Tape State Machine


```
{
  "name": "IBMLIB4",
  "manufacturer": "IBM",
  "model": "TS4500",
  "type": "3592",
  "guiUrl": "http://ibmlib4.cern.ch/",
  "webcamUrl": "http://ibmlib4.cern.ch/",
  "location": "513-S-034",
  "nbPhysicalCartridgeSlots": "16938",
  "nbAvailableCartridgeSlots": "16938",
  "nbPhysicalDriveSlots": "48",
  "comment": "",
  "creationLog": {
    "username": "jleduc",
    "host": "ctaproductionfrontend02",
    "time": "1691655768"
  },
  "lastModificationLog": {
    "username": "jleduc",
    "host": "ctaproductionfrontend02",
    "time": "1691655768"
  }
}
```

Sonarcloud

- Supplements Cppcheck



Also happening in CTA-land —cback

- Modern tape-based backup orchestrator using CTA
- CERN-developed, Restic-based, open source
- In use by CERNBox, EOSMEDIA, CephFS, Ceph-smb, Oracle backup to S3 commercial cloud (experimental)
- Full presentation

Also happening in CTA-land — CTA Operations Utilities

- Automation + monitoring tools and library for CTA operators
- FOSS
- Repository

```
ctautils    <-- Misc. utilities
tapeadmin   <-- Tape interaction utilities
atresys     <-- Repack automation tools
ctaopsadmin <-- CTA admin operation tools
poolsupply  <-- Tape pool media refill automation
tapeverify  <-- Data health check automation
ctaopseos   <-- Tools for interacting with metadata found in EOS
```

CTA Community

Online community forum:
https://cta-community.web.cern.ch/

EOS workshop April 2023

1. New CTA tape lifecycle
2. Disk File Metadata for Tape Files - Migrating, Restoring, Replicating
3. CTA Challenges and Roadmap

CTA at CHEP2023

1. CHEP2023 HTTP REST
2. The CTA Run3 production experience
3. CTA Beyond CERN
4. Evolution of the CERN Backup system

Future plans

Short term

Next public release:

- Focus on consolidation and bugfixes until end-of-year
- Next public feature release Q1 2024

gRPC

- Code in process of being refactored to accommodate the gRPC-frontend
- Target: By end-of-year

Medium term —OS and Metadata

Alma9

- Code: Replacement of CC7 dependencies with Alma9 ones
- Migration of CERN infra in Q1 2024, before LHC activity resumes

Data Co-location / Archive Metadata

- Supply additional metadata with archive requests to improve tape utilization
- Will be a focal point during 2024 data taking
- To be discussed at the DataChallenge 24 Workshop

Medium term —Repack and Monitoring

Repack

- Look into moving repack to a separate objectstore

Monitoring

- OpenTelemetry support
  - Plan: Create prototype for the scheduler first
- Add support for log output in JSON format (decouple from rsyslog)

Tape server

- Better multi-drive support: Keep going while >0 drives are alive

Long term

Scheduler

- Looking to move from Ceph objectstore to PostgreSQL DB
- Will first be rolled out for repack workflows, then for user requests later
- We will keep the present solution for the rest of Run3
- Hope: Prototype by end of 2024

Catalogue database support

- CTA supports both Oracle and PostgreSQL, Oracle in use at CERN
- The choice of DB will in the coming future be done using the plugin system

Summary and Q&A

- CTA at CERN met expectations for 2023 and the Heavy Ion Run
- Deployment of HTTP REST API + EOS5 went/going well
- New in code: Physical Libraries and Repack improvements
- Up next:
  1. gRPC
  2. Alma9
  3. Archive metadata
  4. More Repack and Monitoring
  5. Scheduler separation and migration to PostgreSQL


CERN

CTA Service :Storage Needs 2023–2024

	ALICE	ATLAS	CMS	LHCb	Total
2022 (reported)	61.4	102	140	29.8	333.2 PB
2022 (CTA)	63.7	110	154	32.3	360 PB
2022 (pledged)	95	120	155	81	451 PB
2023 pledge	131	174	228	91	624 PB
2024 request	181	207	320	117	825 PB

- Plus 140 PB for Small and Medium Experiments (SMEs), growing at $\approx$ 20 PB/year