The CERN Tape Archive (CTA) : Archival Storage for Scientific Computing

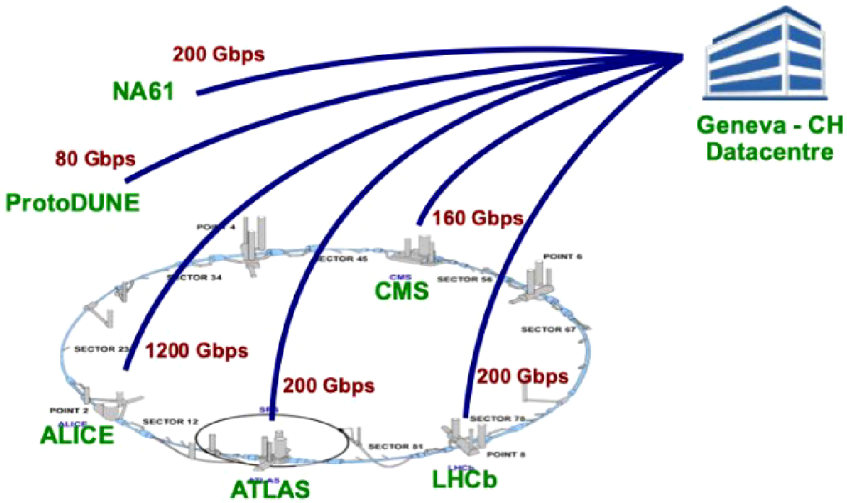
Dr. Michael Davis

CERN, IT Department, Storage Group

CERN/LHC Data Archival

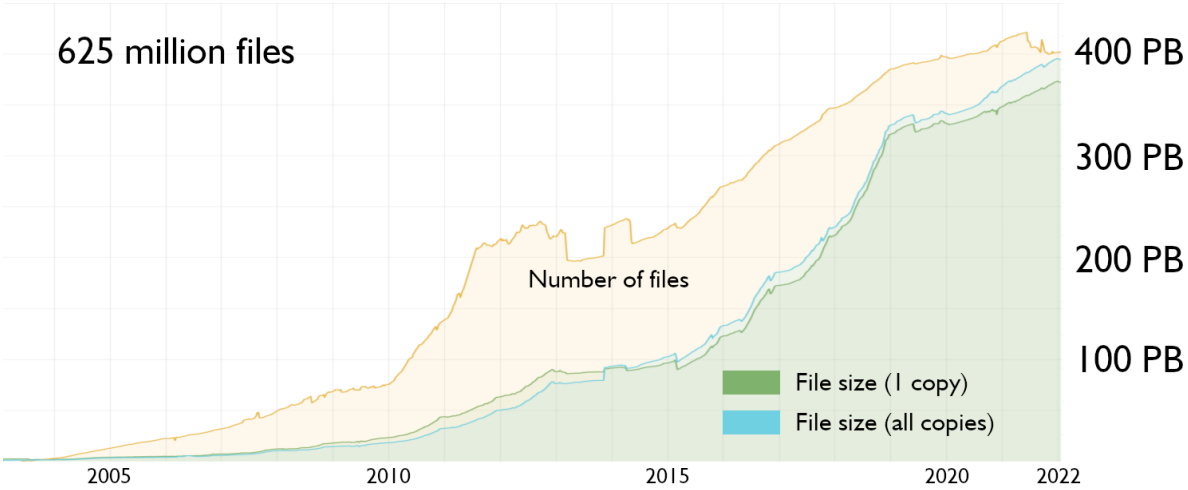
- All LHC physics data—400 PB—is stored on tape in the CERN Data Centre, Geneva (Tier–0)
- Additional copies of the data are distributed across 13 national computing centres (Tier–1)
- Tier–1s also provide tape archival storage

CERN Tier–0 Data Rates (2022–2025)

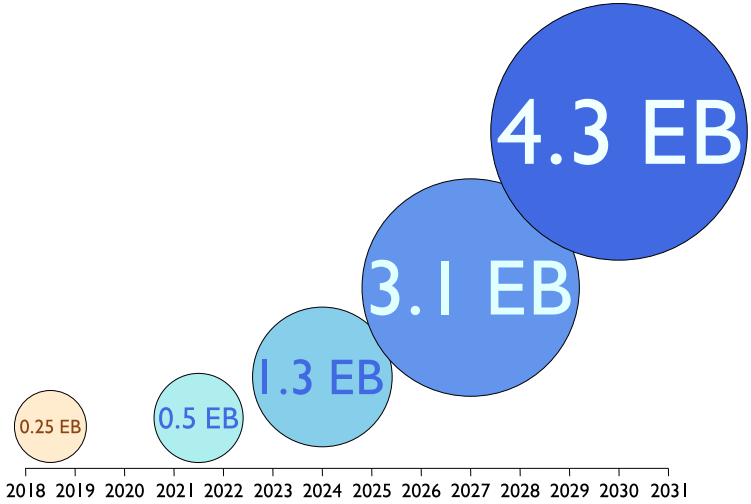


Custodial copy is archived on tape

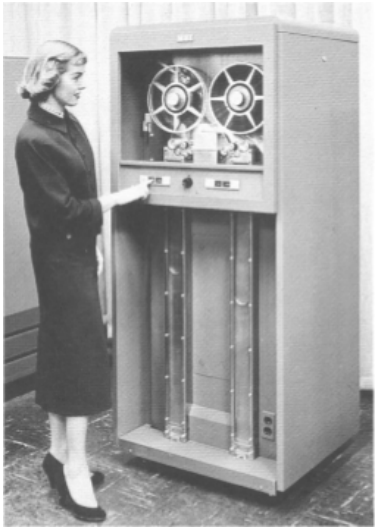
Physics Data on Tape



Predicted Tape Archival Storage Needs



Why use tape? Isn't it obsolete?



CERN Data Storage in the 1970s

140 MB 9-track tape (1974)

10 MB disk platter from CDC 7638 Disk Storage Subsystem (1974)

Tape and disk have evolved in step

IBM TS1160 Tape Drive

20 TB capacity

WD DC HC 530 Hard Drive

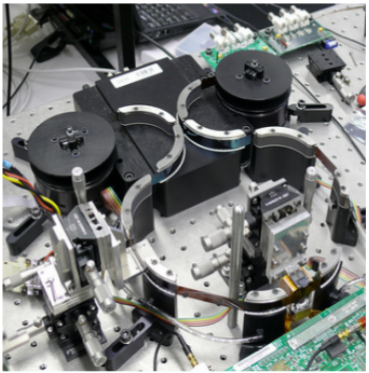
14 TB capacity

Comparing tape and disk

	Tape	Disk
Data transfer rate	400 Mb/s	200 Mb/s
Positioning type	**Fast Sequential Access**	**Fast Random Access**
Average positioning time	30 seconds (610 m @ 12 m/s)	5 milliseconds
Latency to first byte	A few minutes	5–10 milliseconds

Advantages of Tape : Reliability and Data Security

- Two heads are better than one : read after write verification
- No data loss if a drive fails
- Immutable files
- Air-gap security
- Long media lifetime (30+ years)



Advantages of Tape : Energy Efficiency

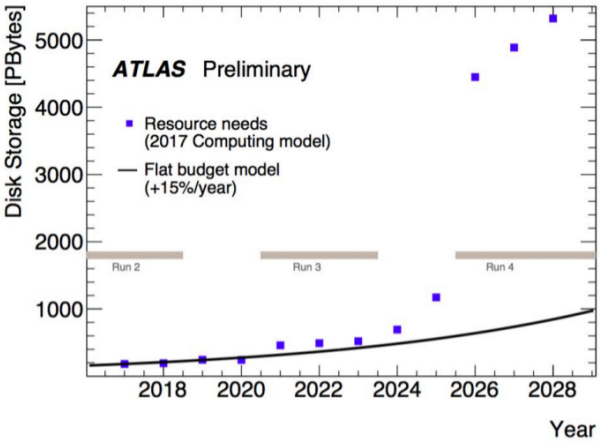
Hard disks are always on. They constantly consume power and generate heat.

- Expensive to run
- CERN Tier–0 Data Centre is at the limit of how much power and cooling it can deliver (3.5 MW)
- Disk storage requires power and cooling

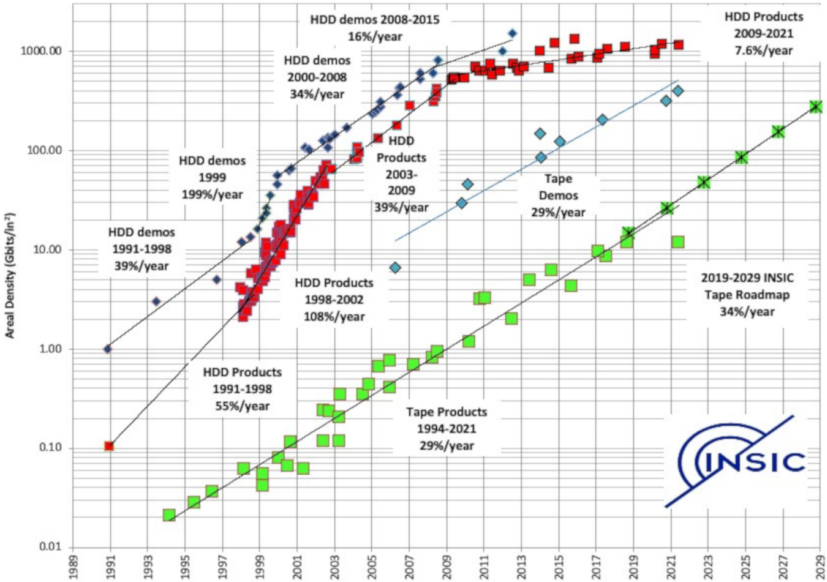
Tape cartridges don't consume any power when they are not mounted in a drive.

- Tape capacity can be increased without requiring additional power

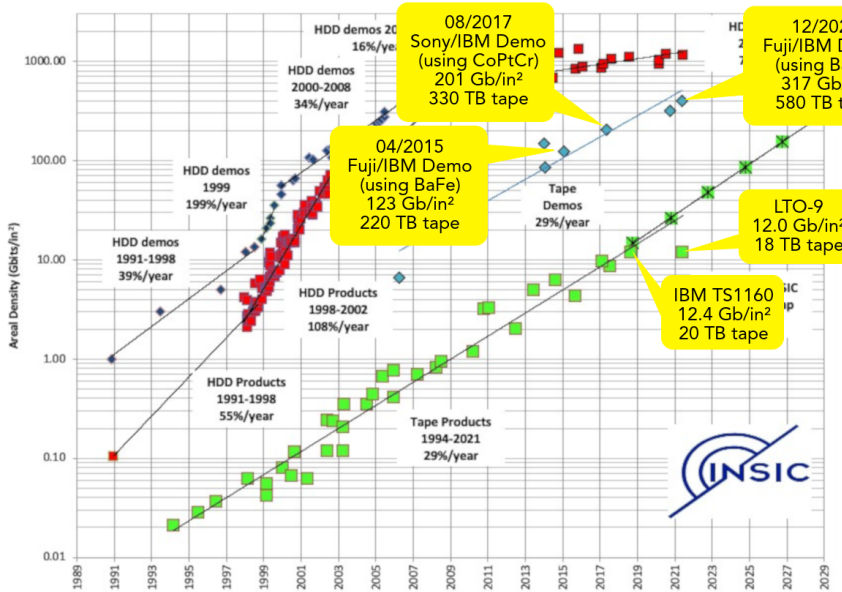
Advantages of Tape : Cost!



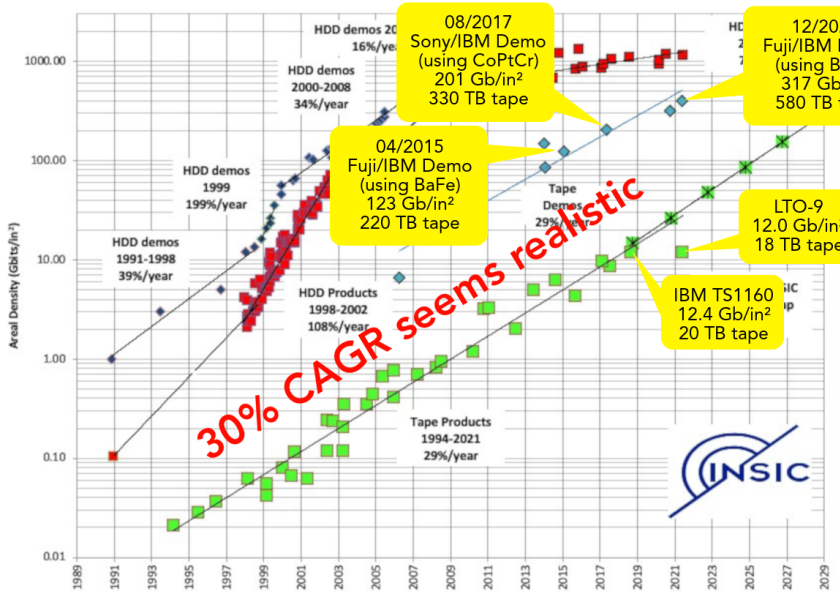
- Storage needs are increasing exponentially
- Budget is not increasing to match needs
- Tape storage is 3–5× cheaper than disk storage



Magnetic Media : Evolution of Areal Density

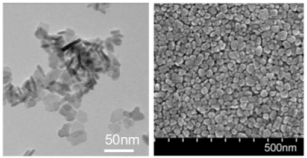


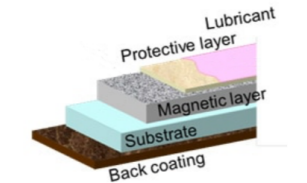
Magnetic Media : Evolution of Areal Density

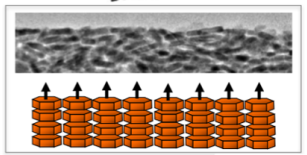


Magnetic Media : Evolution of Areal Density

The Outlook for Tape Technology







New Advanced Materials

- Very fine magnetic particles
- Smooth surfaces with low friction
- 3D stacking of magnetic particles

Disk technologies are pushing the limits of storage density. Tapes have plenty of room to improve capacity.

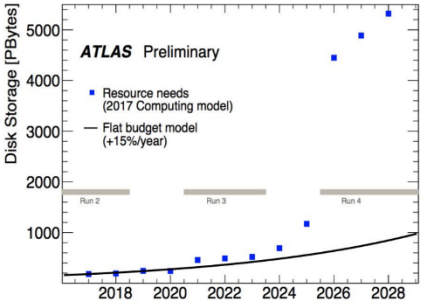
- **The cost advantages of tape will increase over time**

High-Luminosity LHC (2029–32)

Hi-Lumi LHC : Data Carousel
Store data for **online analysis** on tape

Data storage challenge of HL-LHC :

- ➔ ‘Opportunistic storage’ basically doesn’t exist
- ➔ Format size reduction and data compression are both long-term goals, require significant efforts from the software and distributed computing teams
- ➔ Tape storage is 3~5 times cheaper than disk storage, increasing tape usage is a natural way to cut into the gap of storage shortage for HL-LHC



‘Data Carousel’ R&D → to study the feasibility to use tape as the input to various I/O intensive workflows.

The Archival Storage Solution from CERN IT Storage Group

EOS +

CTA

+

FTS

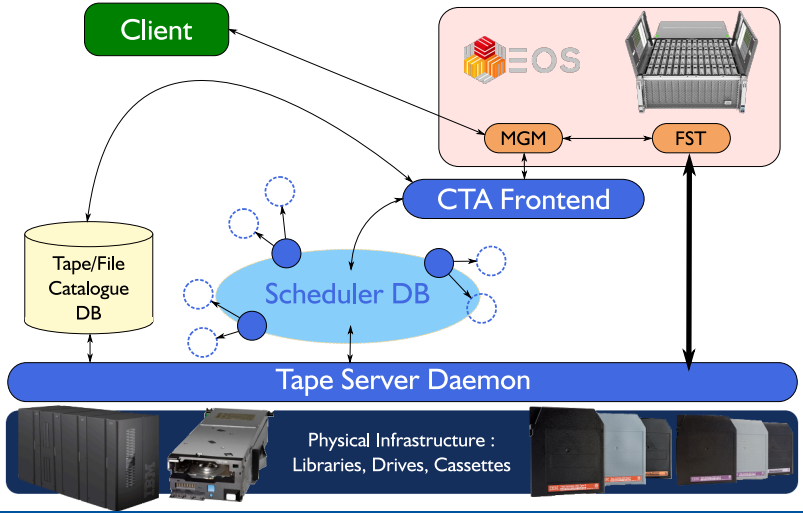
EOS+CTA : “Best of Both Worlds”

- CTA is the tape back-end to EOS (CERN’s high-performance disk system)
- Interface, file operations and disk pool management provided by EOS
- Scheduling and tape operations provided by CTA

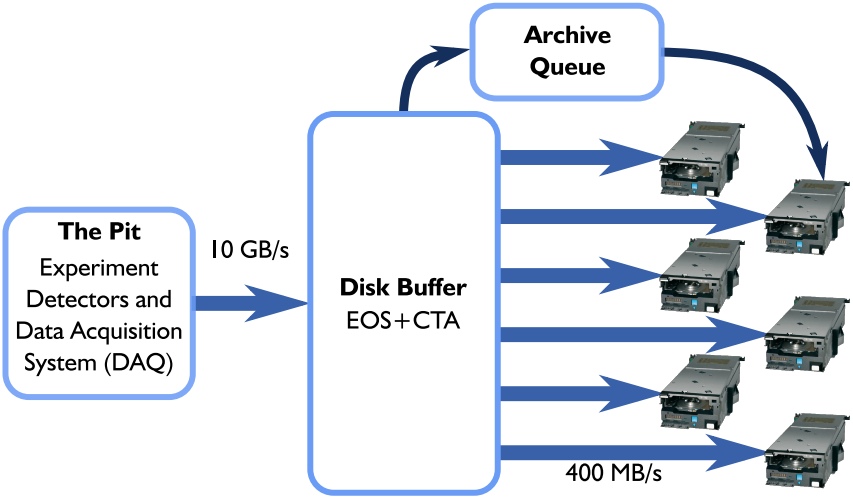
EOS+CTA : “Best of Both Worlds”

Function	Provided by
File Metadata Operations	EOS (MGM/XRootD)
Namespace	EOS (QuarkDB)
Disk Buffer for Staging	EOS (FST)
Tape File Metadata Ops	CTA (Frontend)
Archive/Recall Requests	CTA (Scheduler DB)
Tape File Catalogue	CTA (Catalogue DB)
Tape Operations (libraries, drives, cassettes)	CTA (Tape Server)

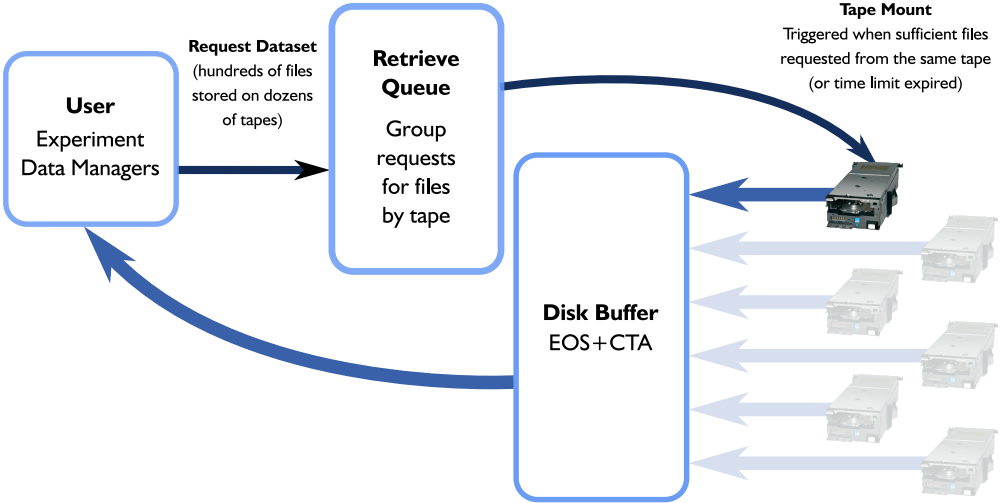
EOS+CTA Architecture



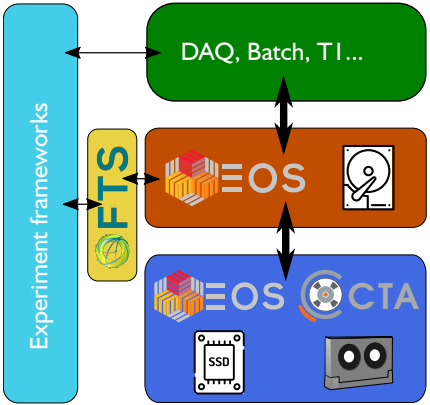
Physics Data Flow : Archival



Physics Data Flow : Retrieval



EOS for Analysis vs. EOS for Staging



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

"Big EOS" (Analysis)

- Tens of PB of storage for physics jobs and staging to Tier–1s.
- File replicas have a long lifetime.
- Spinning disks.

"Little EOS" (Staging)

- Small buffer for copying files to/from tape.
- File replicas have a very short lifetime. Deleted as soon as tape copy exists (archival) or copied to "Big EOS" (retrieval).
- SSDs: reduce contention and give the best price/performance ratio.

Complementary roles of Disk and Tape

	Disk	Tape
Use Case	**Online** : Data Analysis	**Offline** : Long-Term Archival
Drives	110 000 disks (85% HDD/15% SSD)	145 tape drives (in 5 libraries)
Media	—	32 000 tape cartridges
Capacity	Nominal capacity 490 PB	Capacity 460 PB but can be easily extended
	Effective capacity is lower due to redundancy	Currently ≈395 PB on tape (625 million files)

Beyond CERN : CTA at other sites

- Wide diversity of tape software deployed at Tier–1 sites, consolidation likely in coming years
- Increasing license fees for commercial alternatives/ risk of lock-in
- EOS+CTA is Free and Open Source Software (GPLv3)
  https://cta.web.cern.ch
- Active and growing CTA Community
  https://cta-community.web.cern.ch/

Summary

- Tape is the best currently-available technology for archival storage, in terms of reliability, stability over long periods of time and cost
- Power and cooling constraints make it difficult to add more disk storage capacity; tape storage capacity can easily be increased
- Cost benefits over disk storage are significant and look set to increase over the next decade
- Storage needs are growing but budgets are flat
  - The CERN physics archive is 400 PB but will soon grow to 1 EB
  - Data retrievals already exceed 1 EB/year
  - The storage demands of HL-LHC will mean more data on tape and new tape workflows
- CERN is investing in tape as its primary archival storage medium for LHC Run–3 and Run–4


CERN