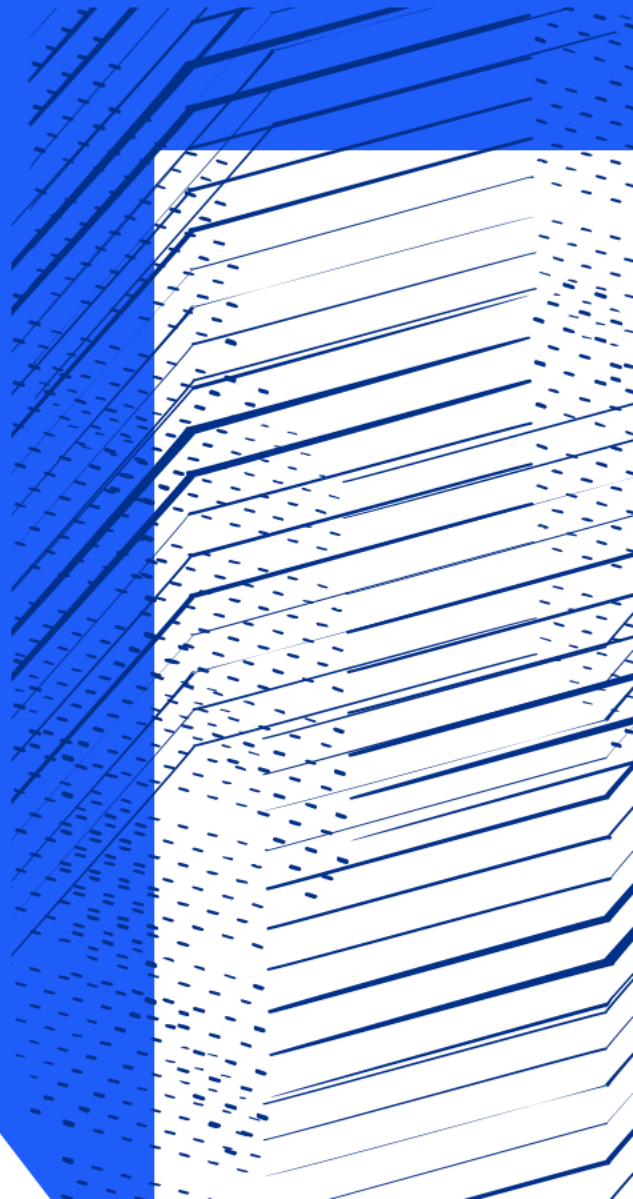
Science and Technology Facilities Council

CTA status at RAL

George Patargias on behalf of the Antares team

CTA 2024 Workshop

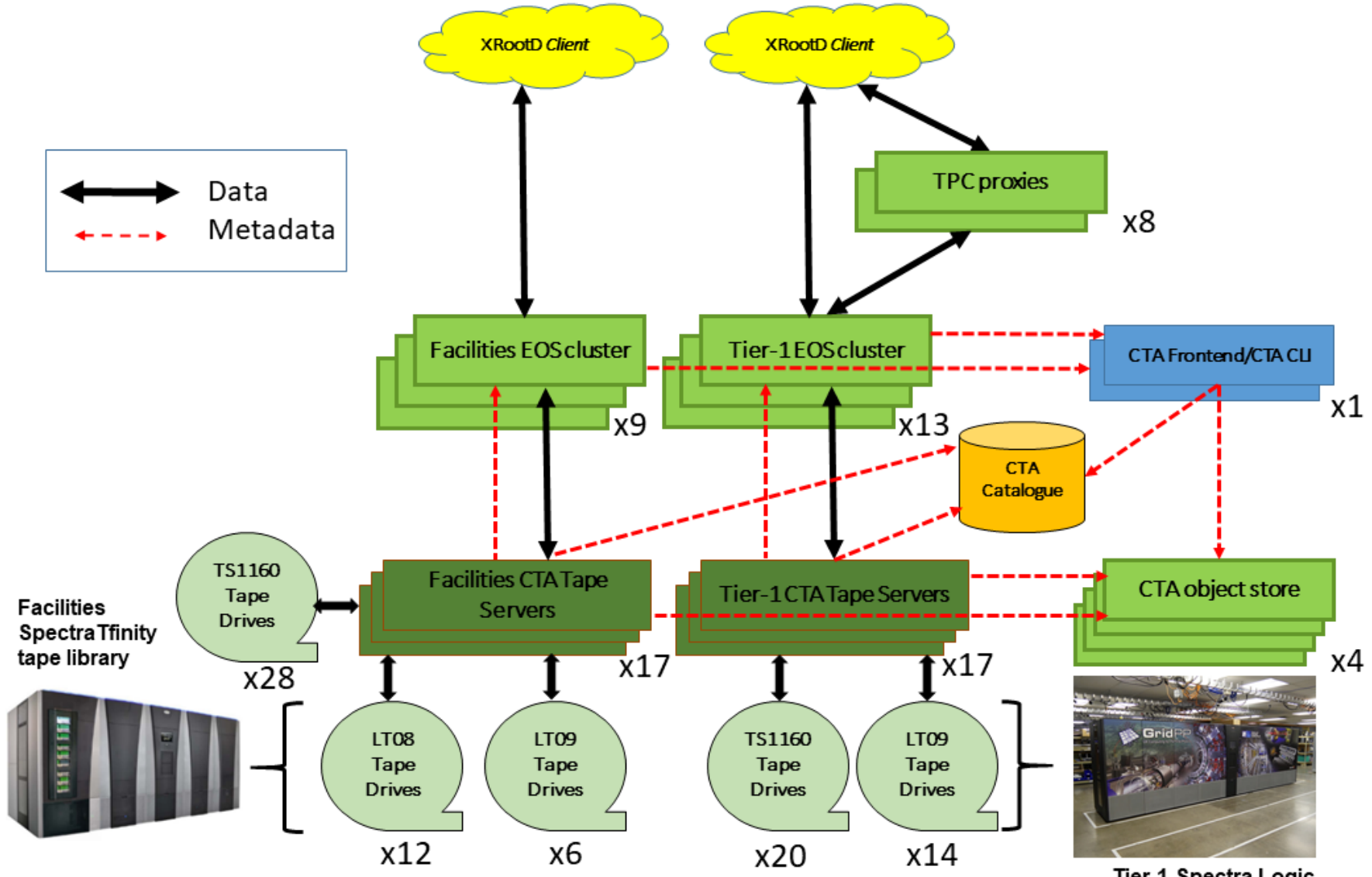


Introduction

- **Antares: Tape Archive service at RAL Tier-1 managing LHC and local Facilities data**

- **Antares team: George Patargias, Tom Byrne, Maha Agilandamurthy, Alison Packer, Tim Folkes**

Antares: Current setup



Science and Technology Facilities Council

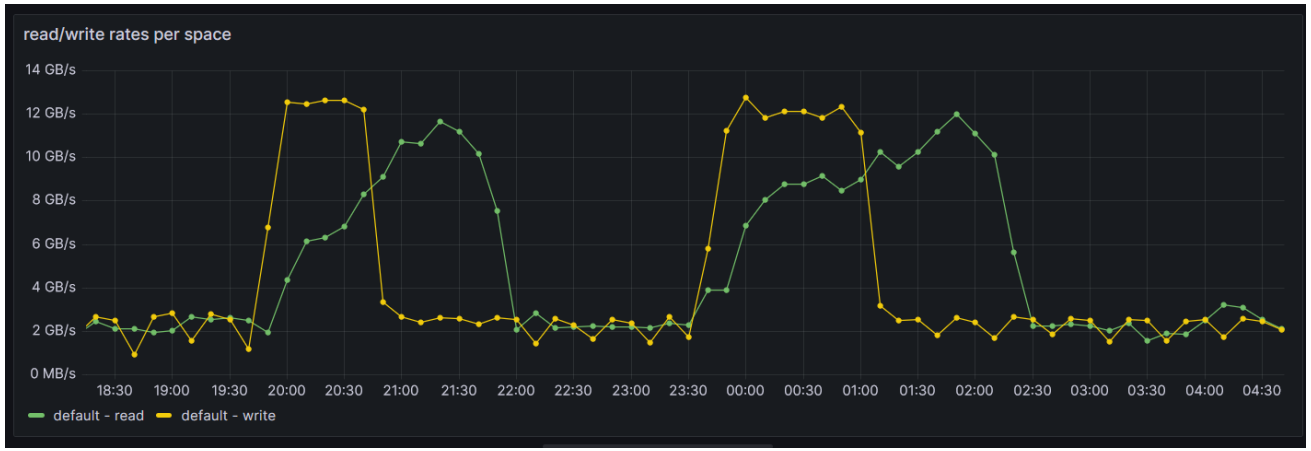
GridPP UK Computing for Particle Physics

Service updates

- May 2023: Enable Tape REST API (ATLAS, LHCb)
- June 2023: Completed the migration for CASTOR Facilities instance
- October 2023: Replace CTA Frontend (VMware with physical)
- November 2023: 6 x LTO9 drives were added to the Facilities tape library
- December 2023: Moved from 3-node Oracle RAC to 2-node RAC
- January 2024: Upgrade to EOS 5.1.28 and CTA 4.10.0-2

Antares performance

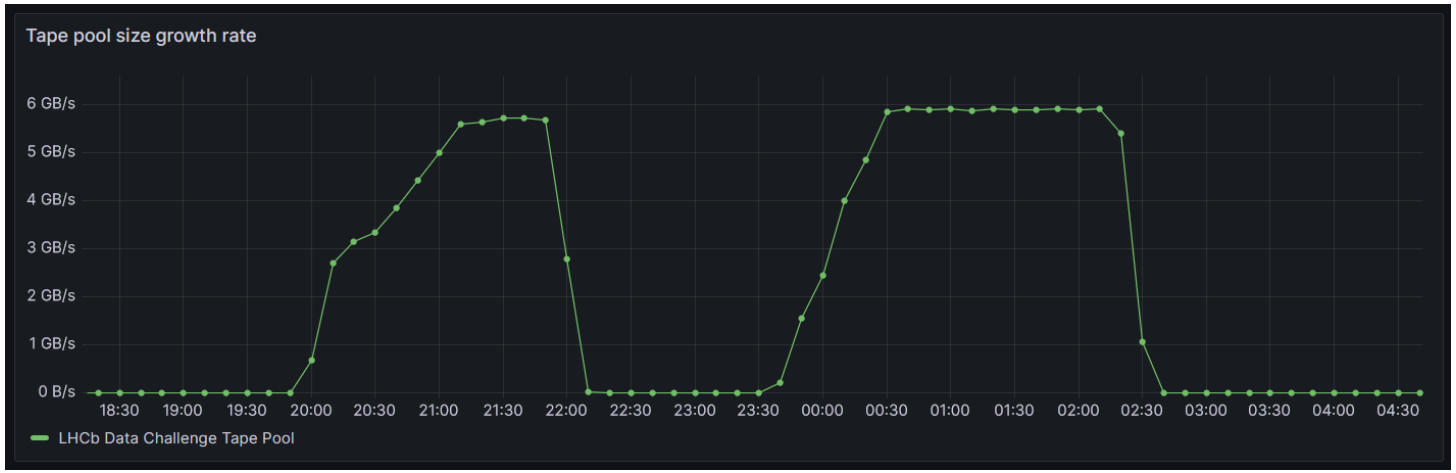
- Data challenge 2024: Testing WLCG infrastructure at 25% of HL LHC rates
- Although not all VOs were testing rates to tape, LHCb tested rates to Tier-1 tape via Tier-1 disk



- Rates into to the Antares disk buffer from Echo peaked at ~12 GB/s (6 EOS nodes in the default space)

Antares performance

- LHCb tape pool growth rate (rate of data going to tape)
- LHCb were using 15 drives for writing



- The LHCb target writing rates were comfortably met by Antares without any tuning or additional resources

CASTOR Facilities migration

- The second and last CASTOR to CTA migration
- Three groups of data (“VOs”) migrated: CEDA/JASMIN, Diamond
  - ~13.3 million files (~102PB)
- More complex than the Tier-1 data migration
  1. Integrate Facilities client code with EOS-CTA:
     - RFIO → XRootD
     - Implement calls to XRootD API to stage, stat and evict files
  2. Remove FileID clashes with some Tier-1 files (Tom’s talk at EOS2023 https://indico.cern.ch/event/1227241/contributions/5335998/)
  3. Test large file ingest and recall – work with users to determine an optimum file size for their use case

Operational issues: Tape libraries

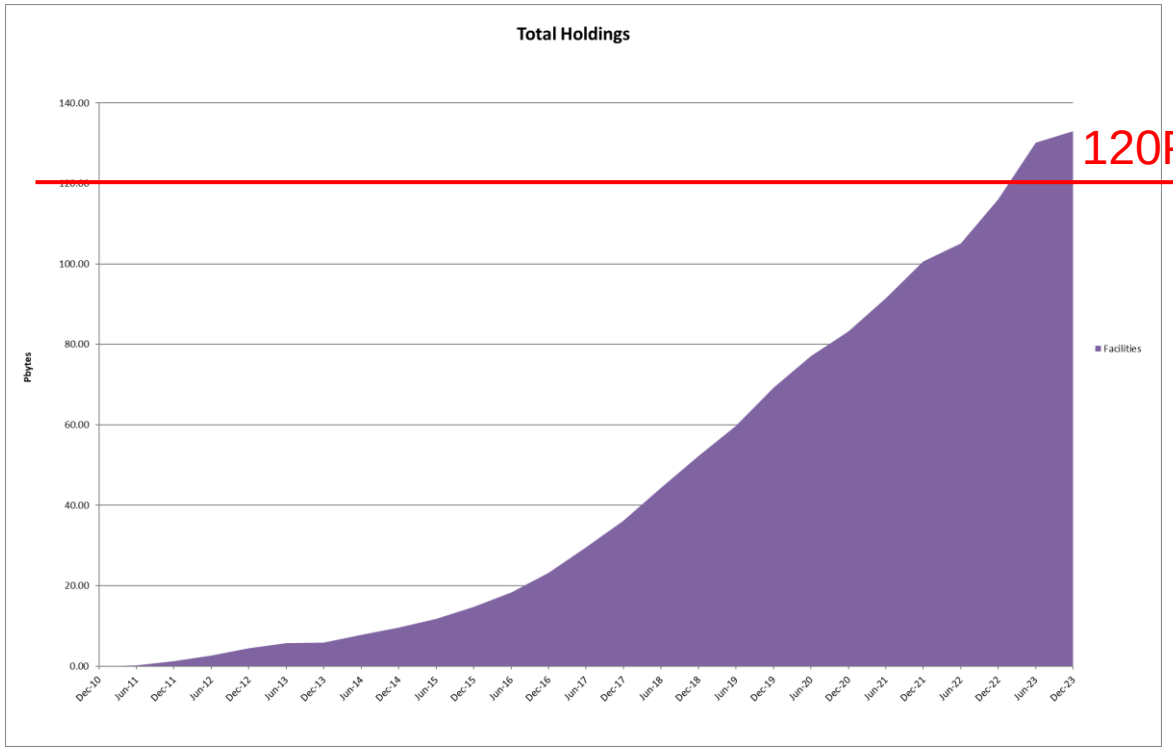
- Asterix (Tier1) has had 4 support cases opened against it in the last year while Obelix (Facilities) 21 support cases opened
- Configured both libraries to use two RIMs concurrently. Asterix worked as expected first time, Obelix had issues due to
  - RIM unit in one frame failing even after three replacements
  - RIM failure caused RCM restart which leaves tapes stuck in drives
  - Paths to logical libraries presented on FC in a different order on each RIM
- Movers failing due to mechanical problems and issues with firmware leaving terapacks stuck in mover.
- Upgraded Obelix to 12.8.08.03 two weeks ago. Already had a new case due to mover not being able to put terapacks away in some slots

New Use Case: EPAC | Extreme Photonics Applications Centre

- EPAC is a new user facility at STFC, due to come online in 2025.
  - EPAC will produce laser pulses with intensities up to 1 Petawatt at 10 Hertz.
- Plasma accelerators will produce multi-GeV electron beams and spatially coherent x-ray and gamma-ray beams for cutting-edge science.
- Expected data rate 500MB/s – 5GB/s
  - Individual images will be of the order of 50MB
  - The plan is for data to be archived on Antares

Facilities File aggregation

- Our facilities scientific data is written to our tape archive via a file aggregation layer
  - The system packs files into ~5GB *aggregates*, organised by time/user group.
- Over 100PB of data stored on tape via this system.
- When this was developed in 2010, aggregation was vital for achieving acceptable archive and recall speeds for small files
  - There has been a number of key changes in the underlying tape system in the intervening period
    - Virtual file markers
    - T10K -> TS/LTO
    - CASTOR -> CTA



Total data holding from the facilities experiments as of Feb 2024

Evaluation of file archival requirements for EPAC

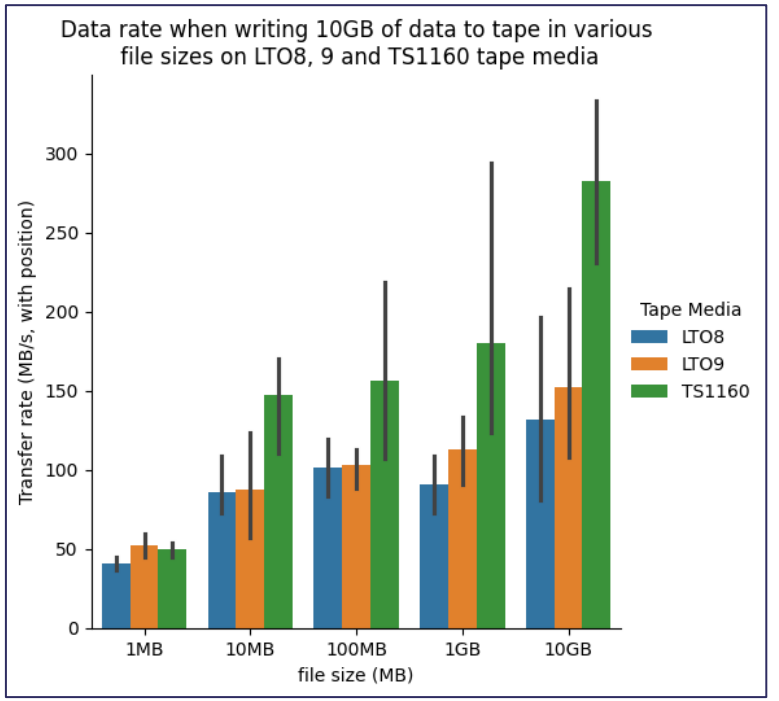
- It's worth considering if EPAC need a file aggregation layer for their archival data.
  - Simplifying the data pipeline wherever possible is a good thing
- We have done some analysis of small file archival and retrieval performance on our CTA instance
  - Tests involved 10GB of data in a variety of file sizes chosen based on expected EPAC data
  - Evaluating data rates and overall tape session time in each scenario on the available media types
  - Recall file order randomised to simulate real world recalls

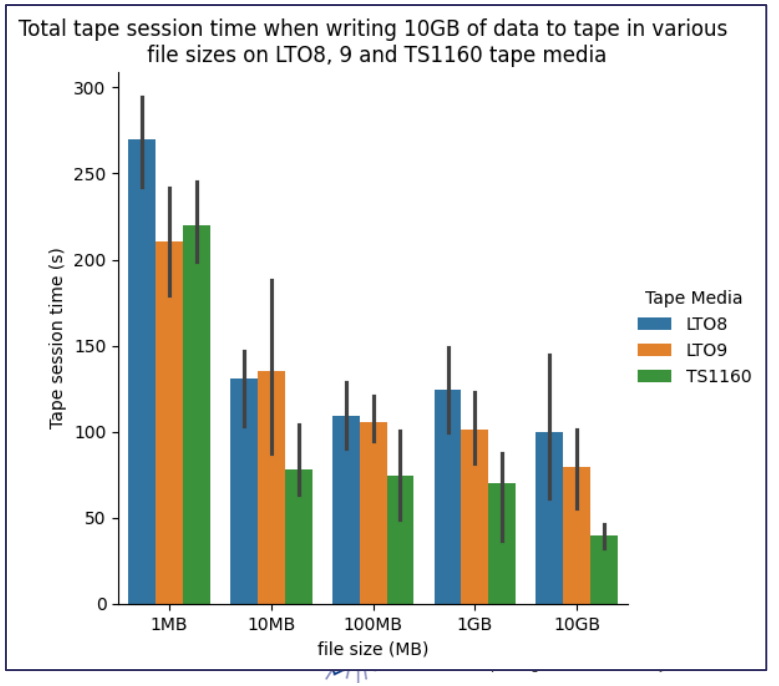
Number of files	File size
1	10GB
10	1GB
100	100MB
1000	10MB
10,000	1MB

Test scenarios – equal data volumes in each scenario allows easy comparison

First impressions - Archival

- In general – writing rates reasonably consistent over most sizes tested
- The move to virtual file markers in the early 2010s was likely the main reason for the reasonable small file archive performance
- The sharp drop at 1MB file sizes is probably due to the spacing of tape file markers: every 1000 files in our config
  - All other scenarios are not writing any file markers before the end.
  - Possible tuning to do here, can this be safely raised?

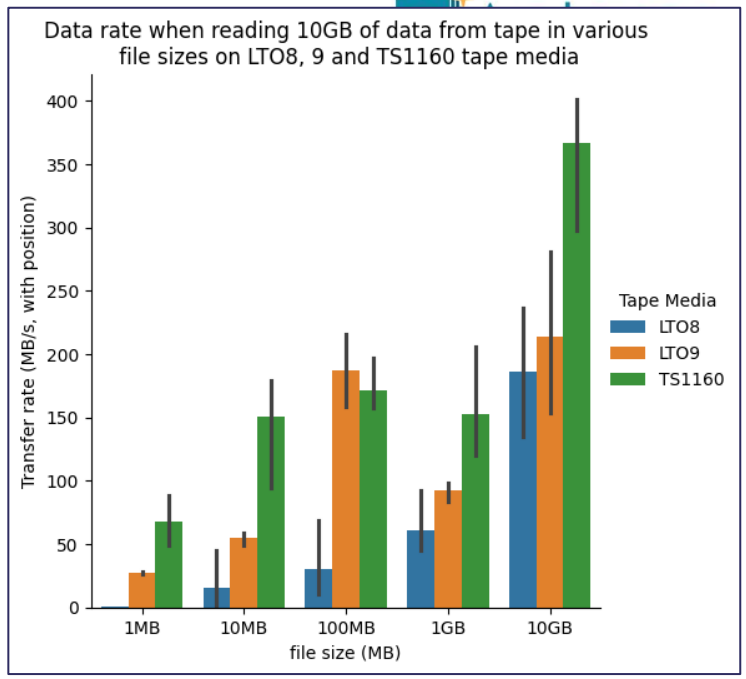


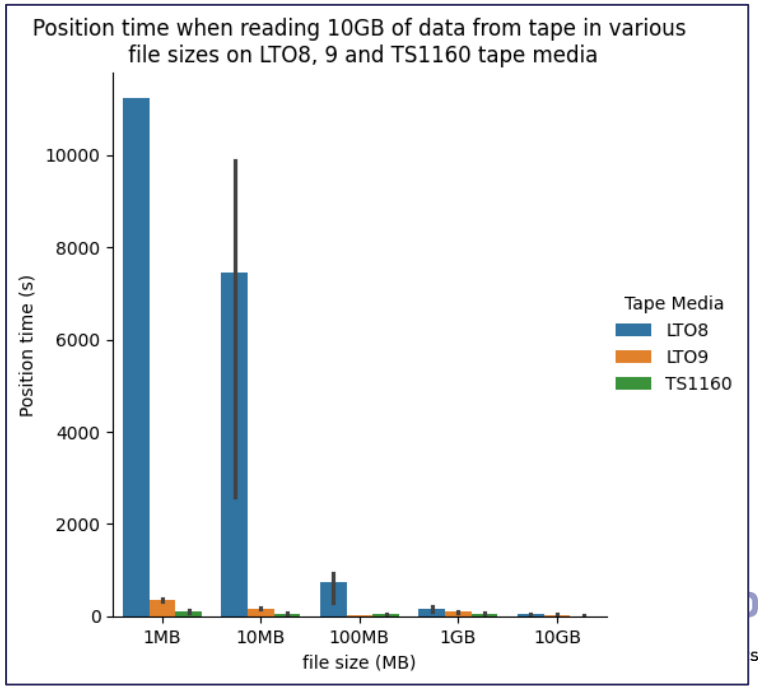


First impressions - Retrieval

- As expected, retrieval rates drop as the files get smaller, particularly on LTO8 media
  - For LTO8, session time was dominated by positioning. RAOLTO algorithm apparently enabled, but possibly misconfigured.

- LTO9 performance not on par with TS for files 10MB and under, but still much better than LTO8





EPAC archival summary

- **CTA small file performance is generally encouraging**
  - **Being able to write data directly to CTA would simplify things hugely.**
  - **Further analysis of expected file sizes needed to ensure we're not missing something.**

- **This testing was very limited in scope, we may be missing other issues**

Future plans

- **Upgrade to CTA 5**
- **Upgrade to Rocky9**
- **Enable XRootD token support**