Evaluating the CTA Storage System for Small Files

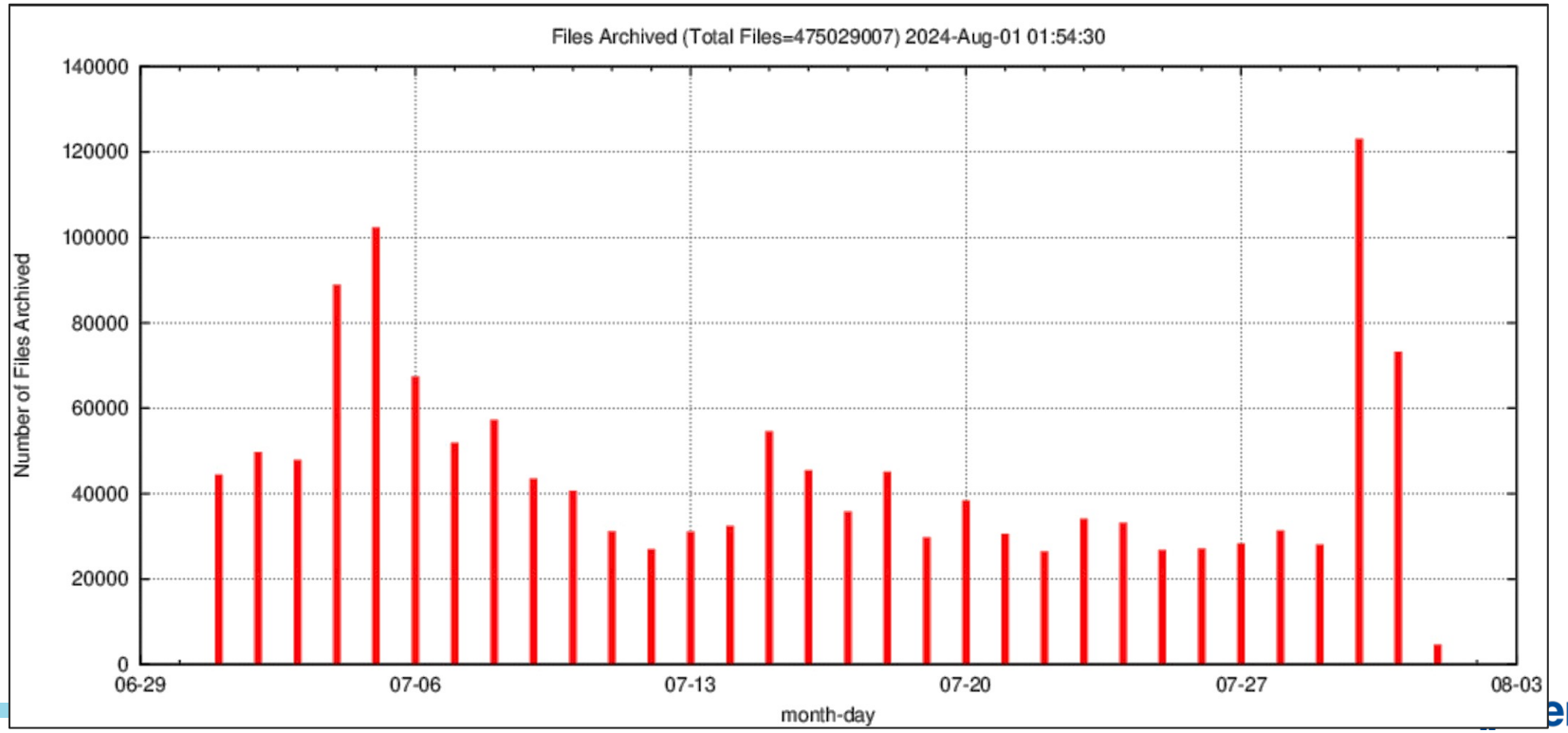
Tammy Walton on behalf of the CTA Project

FIFE Meeting

August 2, 2024

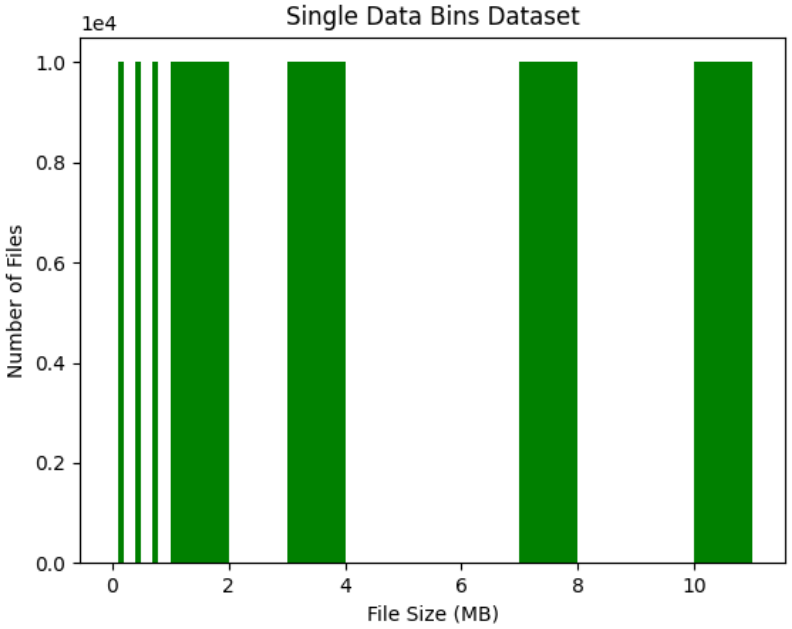
Objective

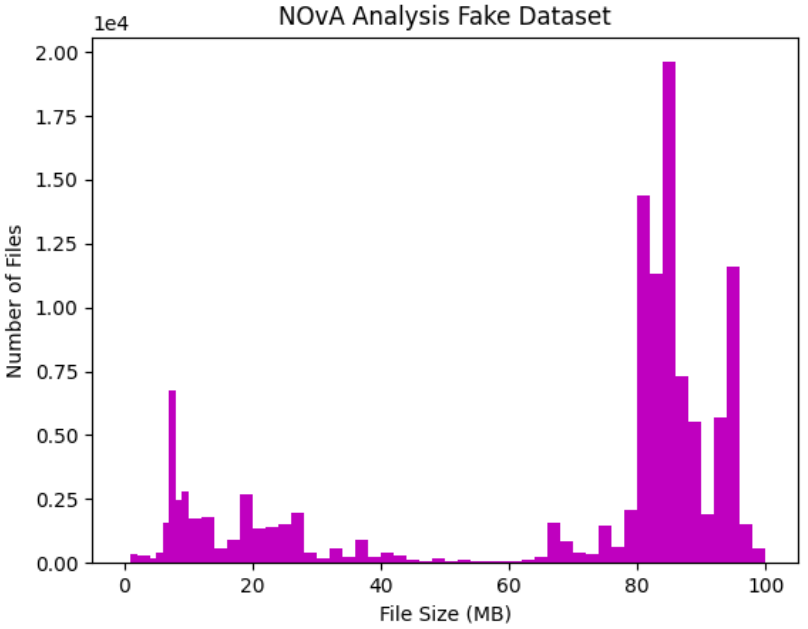
- In 2025, the CERN Tape Archive (CTA) storage system will replace ENSTORE
- CTA does not support dedicated utilities to handle small files (< 250 MB)
- Experiments are actively archiving small files to tape
- The objective is to understand CTA performance for writing and reading data to and from tape



Technical Details

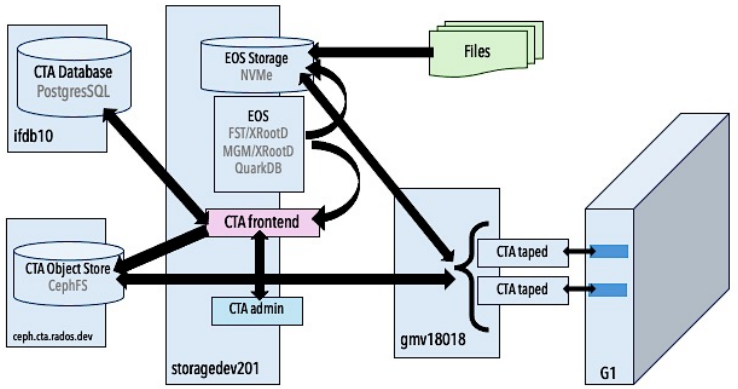
- NOvA produces files with sizes ranging from kilobytes to gigabytes
  - This talk focuses on kilobytes to 100 megabytes files
- Two types of small file datasets are evaluated
  - Single-binned dataset (233 GB with 70,048 files)
    - Composed of files with sizes ranging from 100 KB to 10 MB
  - NOvA freight train dataset (8.1 TB with 123,746 files)
    - Composed of files with sizes ranging from 100 KB to 100 MB and





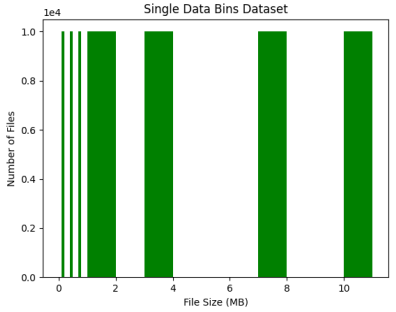
Technical Details

- The evaluation is performed on a development CTA storage system deploying LTO-8 tapes
  - LTO-8 baseline data transfer rate is 360 MB/s
  - The tape's size is 9 TB
  - 4 tapes are active
- EOS delegates the disk functionalities
- Fake files are generated via "dd" and stored on dCache scratch
- CTA tape daemons send out commands (archive, recall, cancel,..) when
  - 100 MB of data are stored on the EOS disk
  - 5 elapsed minutes after the first request is received from the CTA object store
- Analysis is performed via a fermicloud SL7 VM

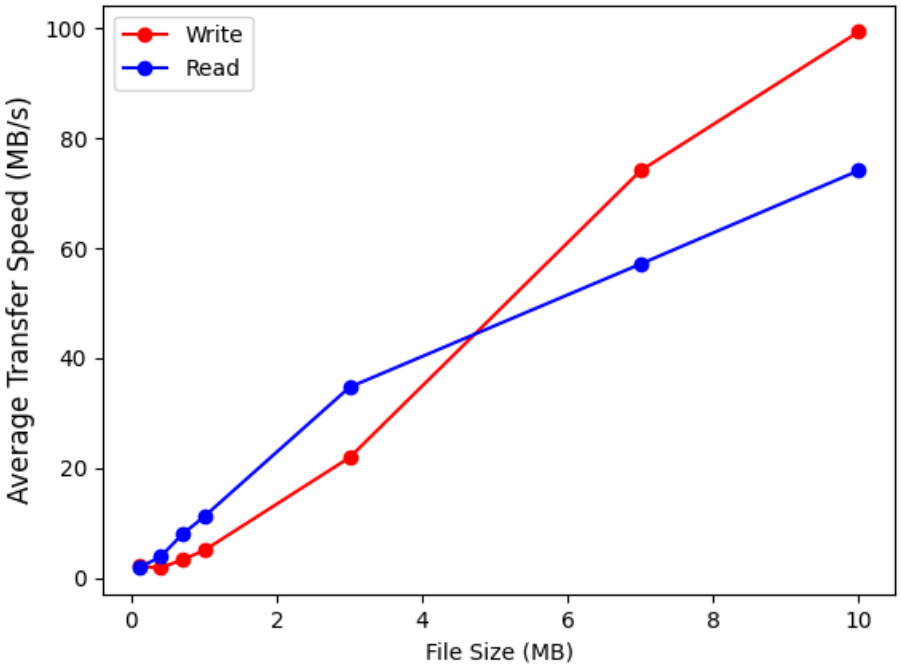


ilab

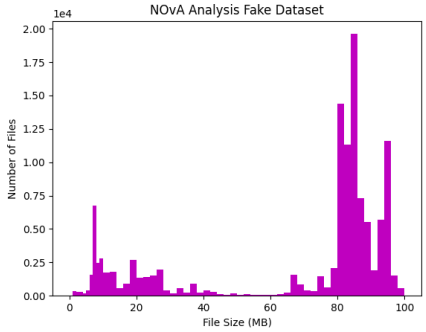
Single-Binned Analysis



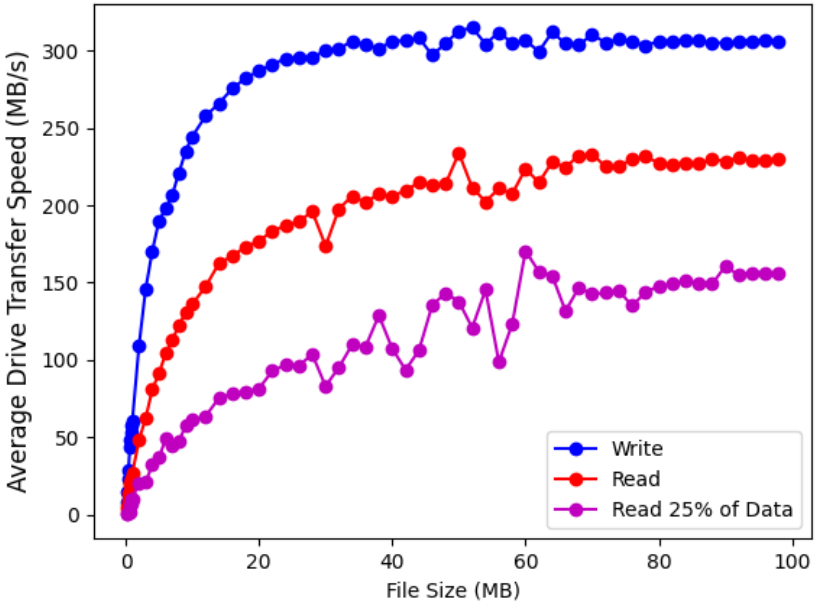
- The dataset consists of files with the following sizes:
  - 100 KB, 400 KB, 700 KB, 1 MB, 3 MB, 7 MB, 10 MB
- An ensemble of same size files is archived or retrieved to or from tape
- The tape drives are disabled while data are transferred to EOS disk or data are prepared for stagging
- After the ensemble of same size files are ready for archival or retrieval, the tape drives are returned to active
- CTA issues commands to archive or retrieve the data
- The results show the CTA backend performance
  - Archival data transfer rate reaches 28% of the potential
  - Retrieval reaches about 20%
  - The tape system malfunctions for the 400 KB



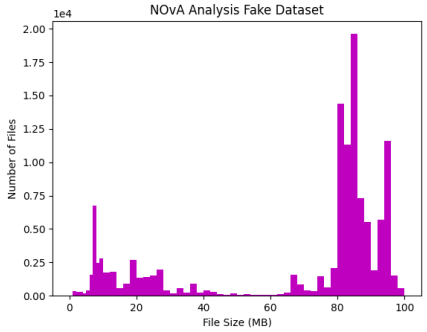
NOvA Freight Train Analysis



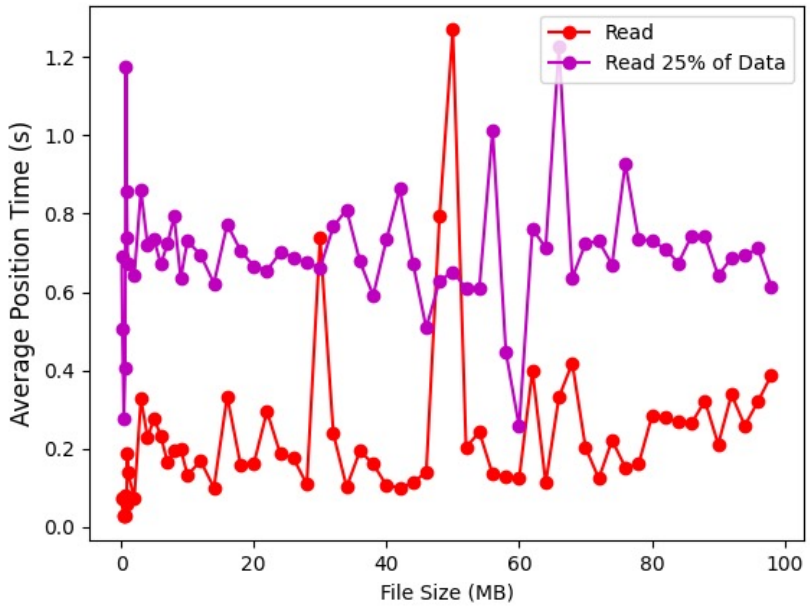
- The dataset consists of small files that are ordered in an unstructured format
- The tape system is operated as nominal
- The files are singlehandedly archived or retrieved to or from tape
- Two types of retrieval tests are performed:
  - All files in a directory are prepared for stagging
  - A random 25% of the files per directory are prepared for stagging
- The results present the performance of the EOSCTA deployment
  - CTA is 80% efficient for writing files with sizes > 20 MB
  - Performs 63% of the potential for reading files with sizes > 40 MB
  - Read performance depends on the requested data and data structure



NOvA Freight Train Analysis



- The dataset consists of small files that are ordered in an unstructured format
- The tape system is operated as nominal
- The files are singlehandedly archived or retrieved to or from tape
- Two types of retrieval tests are performed:
  - All files in a directory are prepared for stagging
  - A random 25% of the files per directory are prepared for stagging
- The results present the performance of the EOSCTA deployment
  - CTA is 80% efficient for writing files with sizes > 20 MB
  - Performs 63% of the potential for reading files with sizes > 40 MB
  - Read performance depends on the requested data and data structure



Summary

- CTA is evaluated for small files deployed on LTO-8 tapes
- The conservative results show CTA performs inefficient for files with sizes < 50 MB
- The next studies will evaluate the dCache-CTA performance for small files that are archived or retrieved to or from LTO-9 tapes

- Experiments should avoid writing small files
- Fermilab dCache-CTA tape storage system will support small files that are archived to ENSTORE

- For more information, see the *preliminary* technical document that is attached to the meeting page