



## Evaluation of CTA for use at Fermilab

Ren Bauer

EOS Workshop 2022

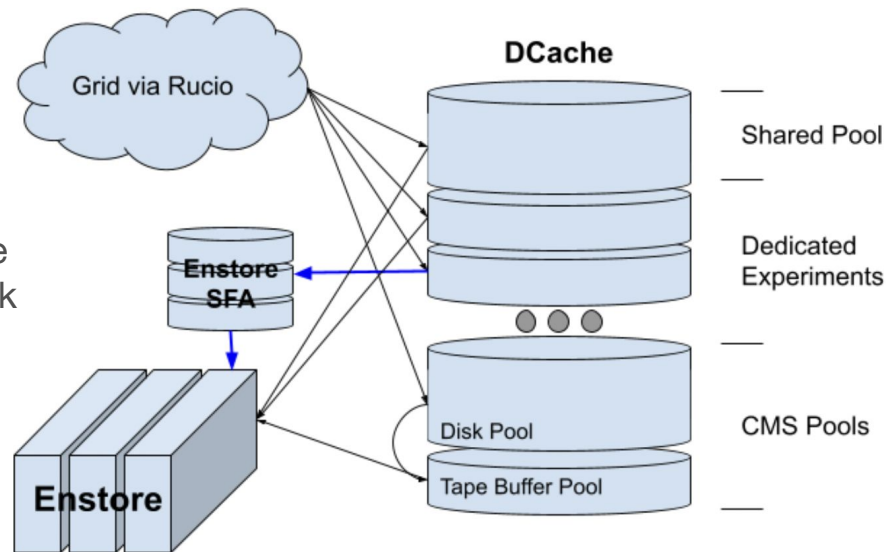
09 March 2022

# State of Fermilab Tape Storage

- Tape Storage on Enstore
- DCache provides Disk storage and buffer space

For CMS in particular, there are disk storage and tape buffer pools, similar to the two disk and buffer EOS instances in CERN's CTA deployment

- Enstore's Small File Aggregation (SFA) provides capability to stage small files on disk until they can be packaged into a file large enough for tape storage
- All services run on bare metal hardware, no virtualization or release automation



Data ingresses to Fermi from the grid via Rucio, and goes to DCache where, depending on pool, it can take a variety of paths to Enstore

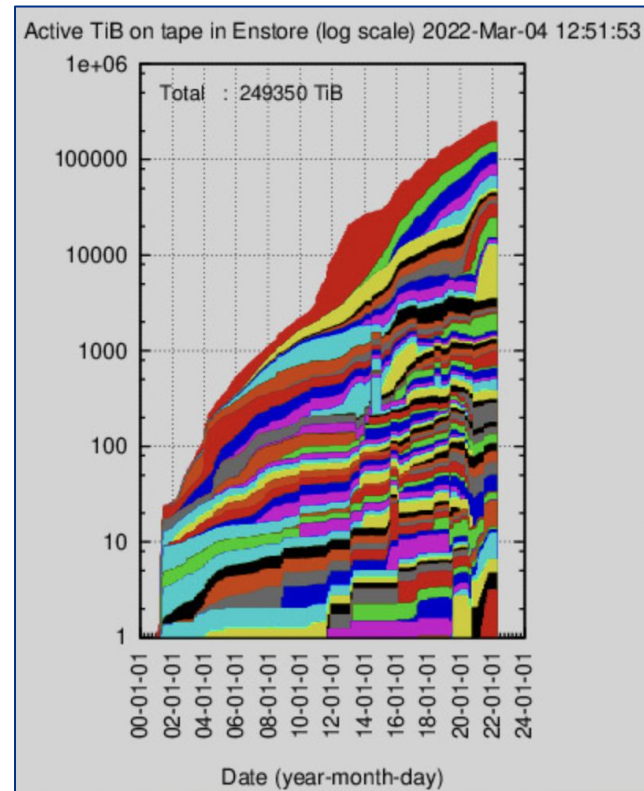
# Fermilab Tape Storage Statistics

## Enstore

- Three IBM TS4500 libraries and one Spectra TFinity
- 300+ Petabytes stored data
- 500+ Terabytes R/W per day

## SFA

- Small File Aggregation service, utilized by public experiments to package tape data
- Processes up to 5+ TiB per day



**Bytes store over time, all Enstore**  
**Colors indicate storage groups, which roughly correlate to experiments, but CMS groups dominate**

# CTA Evaluation at Fermilab

## Key Evaluation Points

- Integration with Fermilab virtualization infrastructure
- Compatibility with CPIO tape format
- Integration with DCache disk storage
- Replacement for SFA functionality

# CTA Evaluation at Fermilab

## Key Evaluation Points

- **Integration with Fermilab virtualization infrastructure**
- **Compatibility with CPIO tape format**
- Integration with DCache disk storage
- Replacement for SFA functionality

# CTA Evaluation at Fermilab

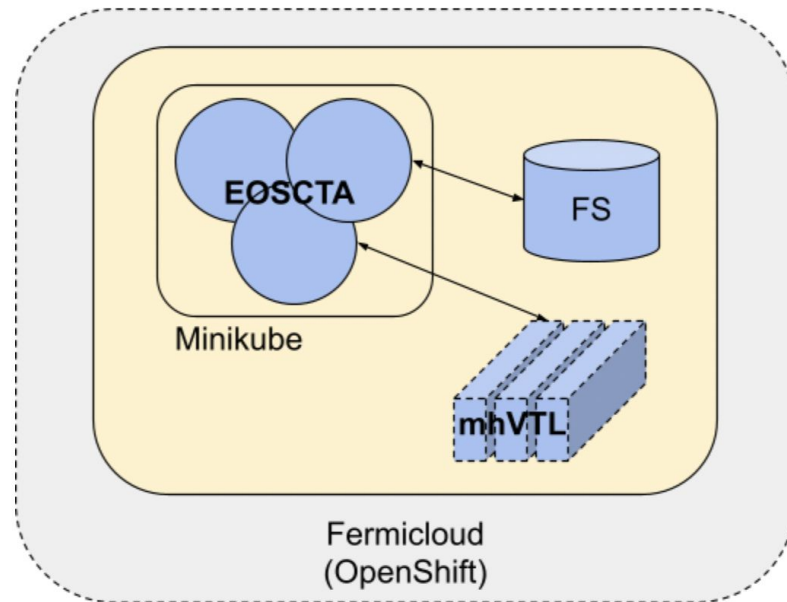
## Virtualization Infrastructure Integration: Current

- Currently have EOSCTA running on a single-VM Kubernetes cluster

Use CTA Continuous Integration build to deploy to Minikube on a Fermicloud OpenStack VM

ObjectStore data in mounted FS Volume

Tape data read/written to mhVTL tape virtualization library

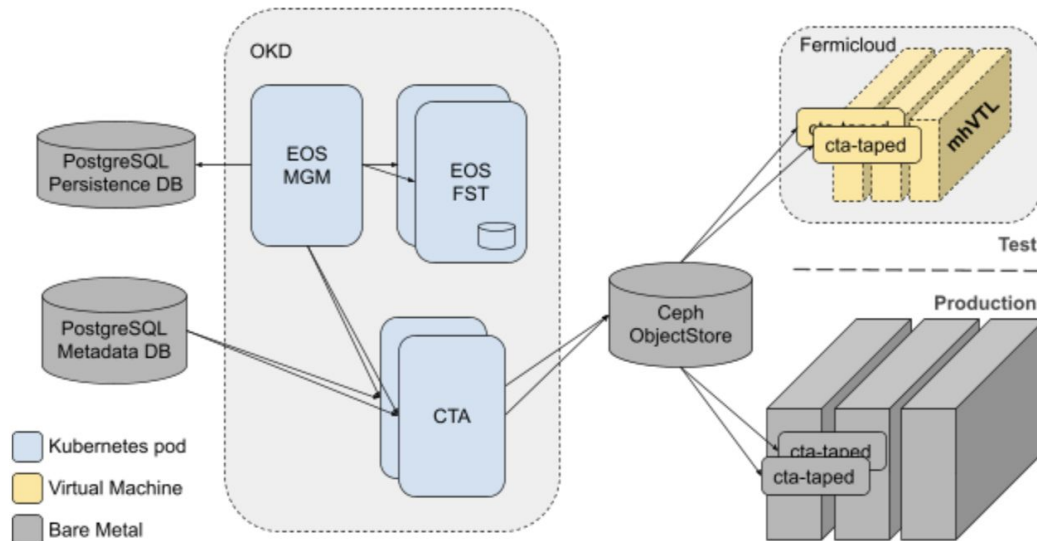


Existing functional EOSCTA deployments at Fermilab

# CTA Evaluation at Fermilab

## Virtualization Infrastructure Integration: Proposed

- EOSCTA on OKD
- Postgres for Metadata and EOS Persistence DBs
- Ceph cluster for ObjectStore
- Tape daemons deployed on:
  - Fermicloud with mhVTL backend
  - Bare metal machines with direct tape library interface



Proposed EOSCTA deployment at Fermilab, with virtual and physical tape backends

# CTA Evaluation at Fermilab

## Compatibility with CPIO Tape File Format

- Data at Fermilab is largely stored in CPIO tape format

Enstore also supports CERN tape format used in CTA

This format is used by CMS and for large files (8GB+)

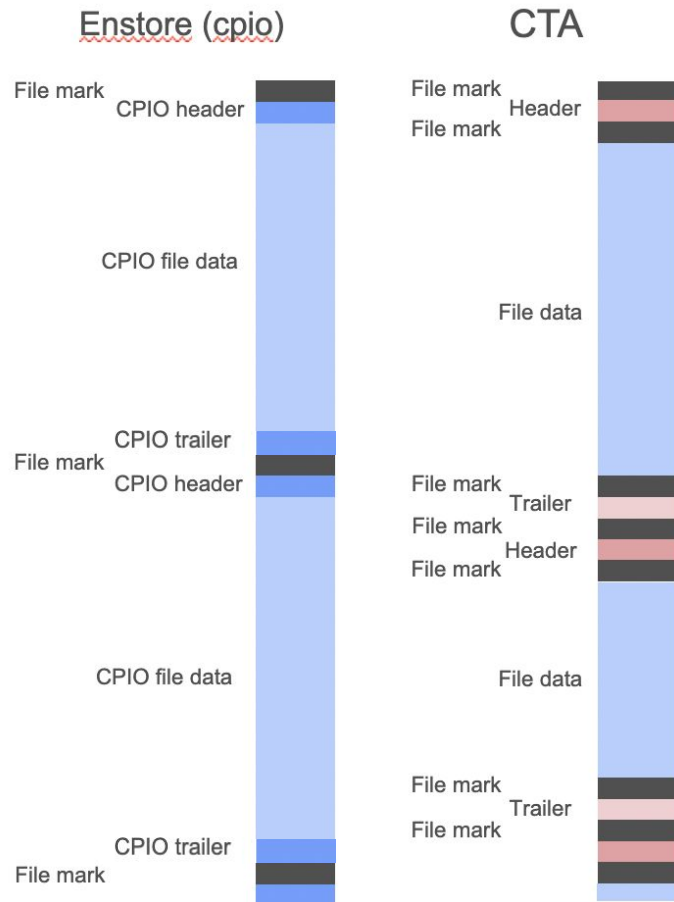
We and our colleagues at PIC find CPIO much more performant in Enstore

However, CTA currently has no support for CPIO format

Much of our dev effort focused on implementing this support

Planning to implement read, not necessarily write, functionality for this format

Contact: ewv@fnal.gov





# CTA Evaluation at Fermilab

## CPIO Compatibility Implementation: Progress

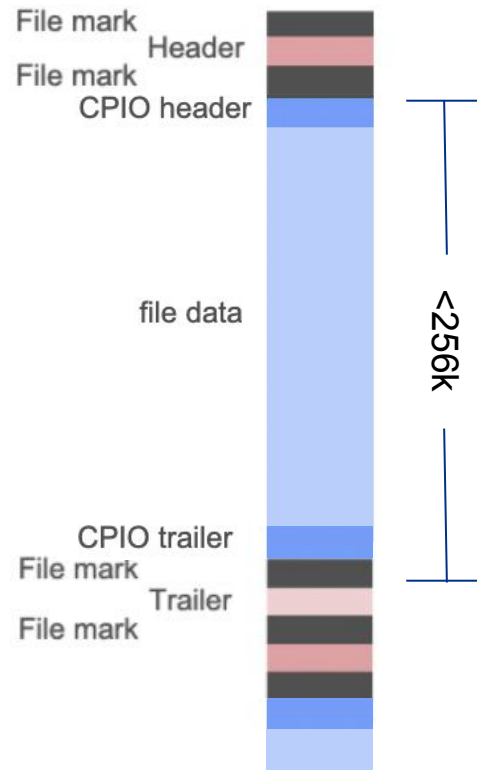
- We've modified CTA (CASTOR) code to:

Wrap short (< 256 kB) data files with CPIO header and trailer, in order to match how they are written by Enstore

For files which fit into a single CASTOR read buffer (256kb), unwraps CPIO header and read file correctly

Will handle files spanning multiple buffers using CPIO metadata on file contents

Not thoroughly tested, as we don't yet have a method to create such files written with the CPIO wrapper



# CTA Evaluation at Fermilab

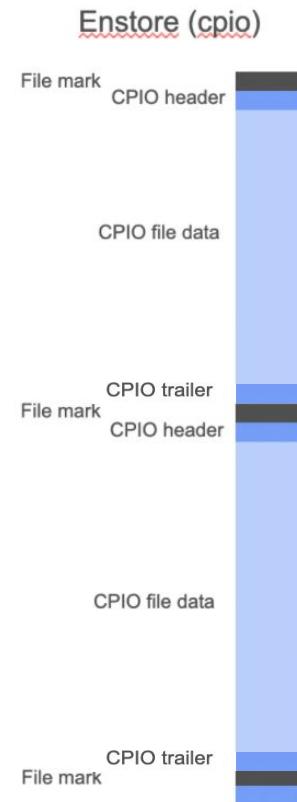
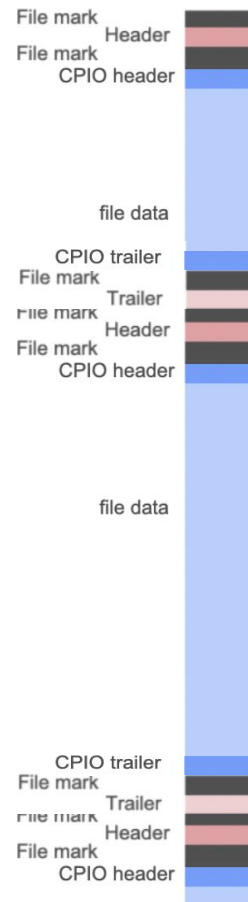
## CPIO Compatibility Implementation: Next steps

- Create virtual tape with CERN-style tape headers and multiple CPIO-wrapped data files

Includes importing tape details to CTA metadata database

- Thoroughly test read code, ensuring it can handle CPIO-wrapped files of all lengths
- Modify code to no longer expect CERN-style tape headers
- Import an Enstore-written tape into mhVTL, and test read capability

This tape will have CPIO wrapped files, and no CERN-style tape headers or file headers / trailers



# CTA Evaluation at Fermilab

## CPIO Compatibility Implementation: Collaboration

- Folks at DESY are also working on multi-format compatibility in CTA

Interested in supporting OSM format tapes

(OSM is identical to CPIO for sizes under 2GB)

Looks like they're a little closer than we are with Enstore

We'll be working with them to converge on a solution



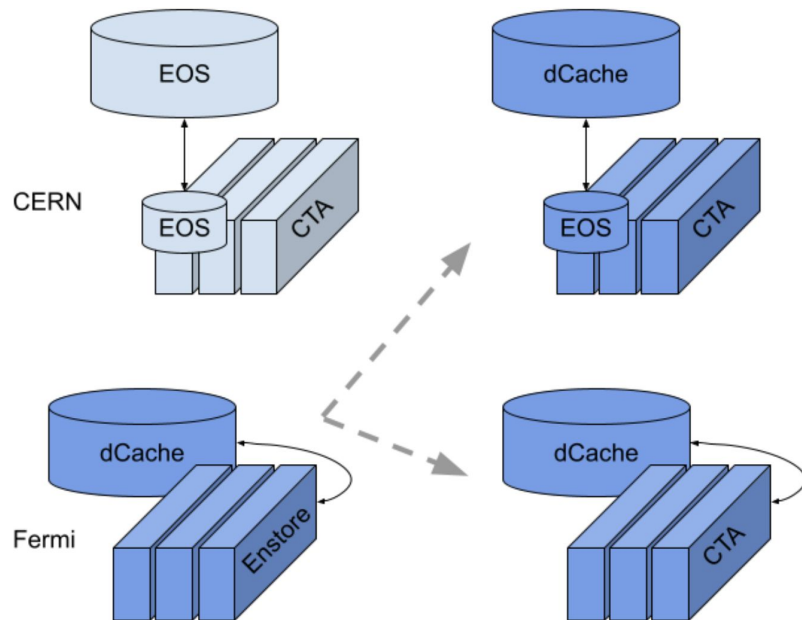
# CTA Evaluation at Fermilab

## Integration with dCache disk storage

- Fermilab employs dCache for all disk storage
- Open to replacing EOS entirely in the EOSCTA system, or replacing large disk pool and maintaining small EOS deployment as part of CTA system

A dCache-only system likely simplifies deployment

- Dedicated session on this integration led by Tigran Mkrtchyan of DESY coming up



# CTA Evaluation at Fermilab

## Replacement for SFA functionality

- SFA is an Enstore feature that allows staging of small files on disk, until enough are aggregated to package into a single file to be written to tape

Over 60 TB of SFA-packaged data written to tape in the last month

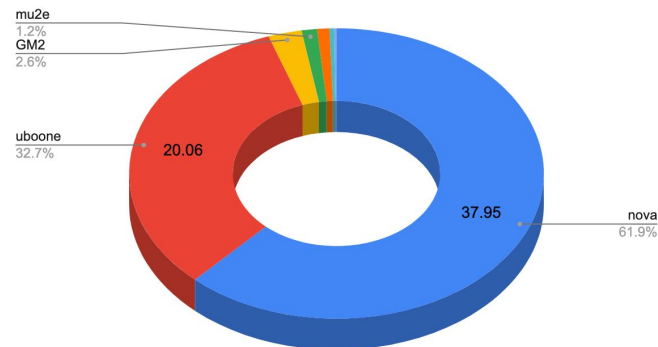
Estimated over 90 PB of SFA archived data currently in Enstore

No roadmap for replicating this functionality in CTA

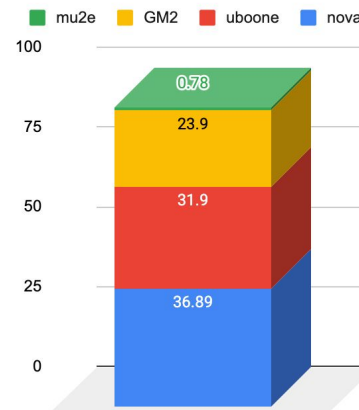
- Open to other methods of small file handling

SFA TB Written by Storage Group (1 Month)

Total: 61.27 TB



Total SFA Storage (PB)



# CTA Evaluation at Fermilab

## Key Evaluation Points

- Integration with Fermilab virtualization infrastructure
  - Test instance running CI build in Kubernetes
  - Evaluating feasibility of running on OKD cluster
- Compatibility with CPIO tape format
  - Making progress on CPIO Read functionality
  - Hope to work with folks at DESY to enable multi-format
- Integration with DCache disk storage
  - Open to options, following DESY progress
- Replacement for SFA functionality
  - Important feature set for our users, no roadmap

# CTA Evaluation at Fermilab

## Thanks to Contributors

- Eric Vaandering - ewv@fnal.gov  
CPIO Compatibility Implementation
- Robert Illingworth - illingwo@fnal.gov  
Project Oversight and Tape Format Diagrams
- Bo Jayatilaka - boj@fnal.gov  
Project Oversight
- Rafael Rocha, Jeff Derbyshire, Dmitry Litvintsev  
Enstore Stats and Figures