Fermilab Rucio Operations for ICARUS, DUNE, and the Rubin Observatory

Brandon White, Dennis Lee, Yuyi Guo

Rucio Workshop

November 2022

The Scope of Fermilab Operational Involvement

- About Me
    - Distributed Data Development @ FNAL
    - Rucio Technical PoC @ FNAL (except for CMS)
    - Vera C. Rubin Observatory Data Curation Group Lead
    - Assisted by Dennis Lee and Yuyi Guo

- Fermilab maintains 6 Rucio deployments
    - 3 VOs
    - Hosted at two different physical sites
    - Two disparate cluster deployment platforms
    - OKD is the open-source upstream of Redhat Openshift and the method of deployment used at FNAL
    - Kubernetes is used at SLAC, with some caveats
    - Multi-VO plans to migrate all experiments eventually
    - Databases at Fermilab are supported by a centrally managed DB service
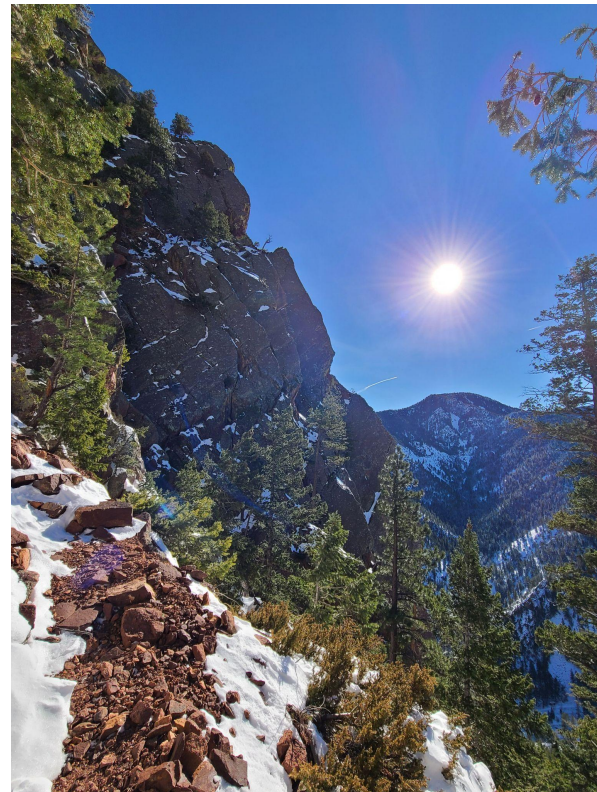
🎳 **Fermilab**

ICARUS Rucio Deployments

- Experiment at Fermilab searching for sterile neutrinos
- Fermilab supports a 1.26 deployment for ICARUS
  - Will be upgraded to 1.29 once the FNAL OKD4 cluster comes online, scheduled for after the DUNE upgrade
  - Limited issues with this deployment, has run fairly stably on OKD3 since 2020
  - Stability is usually limited by usual service management issues (e.g. certificate renewal)

- The ICARUS use case is fairly simple
  - Deterministic RSEs, no custom data placement
  - Primarily used for data movement from FNAL->INFN
  - For the moment replica cataloging functionality for ICARUS is provided by the legacy Sequential Access via Metadata service

Fermilab
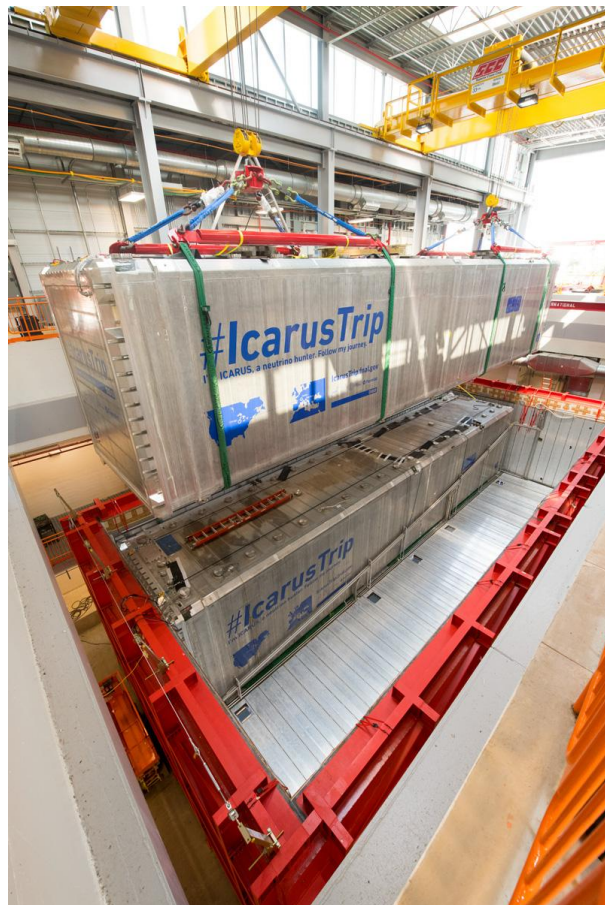
DUNE Rucio Deployments

- DUNE has used a Rucio deployment on OKD3 (frozen at 1.26) for production operations since late 2019
  - Uses a custom image build/deployment framework to allow for customization such as policy packages
  - Versus K8s: are Routes vs. Ingresses, custom ServiceAccount permissions, monitoring config
  - Deterministic RSEs. No custom data placement

- Running a development deployment of 1.29
  - Deployed on Fermilab's development OKD4 cluster
  - Uses the DUNE policy package for custom data placement at tape sites
 - Production OKD4 cluster is newly online!

DUNE Considerations

- Custom Policy Package (James Perry)
  - Uses DUNE's MetaCat software (Igor Mandrichenko) to query metadata for dynamic creation of PFNs
  - Enforces the constraint that any file declared to Rucio must have first been declared to MetaCat

- Most 1.26->1.29 upgrade pains were due to our custom deployment process or accounting for changes in Helm charts. Smooth, if time consuming.
- Support for HTTPS+SRM needs to be enabled to allow for delivery of tape-resident data from Fermilab
- Token support, to be discussed further by other DUNE reps

🟦 **Fermilab**
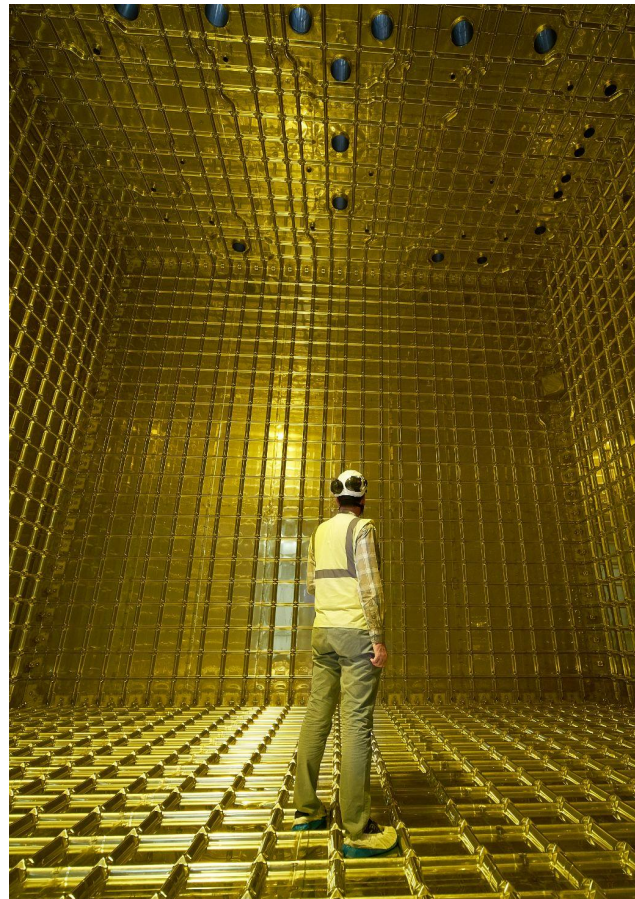
Data Challenge Retrospective

- Data Challenge Issues
  - Initial configuration issues caused instability
  - Caused by custom server configuration
  - Some remediations to policy package

- Rucio Replicated a 500TB total dataset
  - Full copies at FNAL/BNL, split copies in UK/EU
  - Demonstrated speeds of 3.5GB/s from CERN, average ~2GB/s
  - There is still a need to reclaim the ~2.5PB of data challenge data spread around the DUNE RSEs

🔷 **Fermilab**
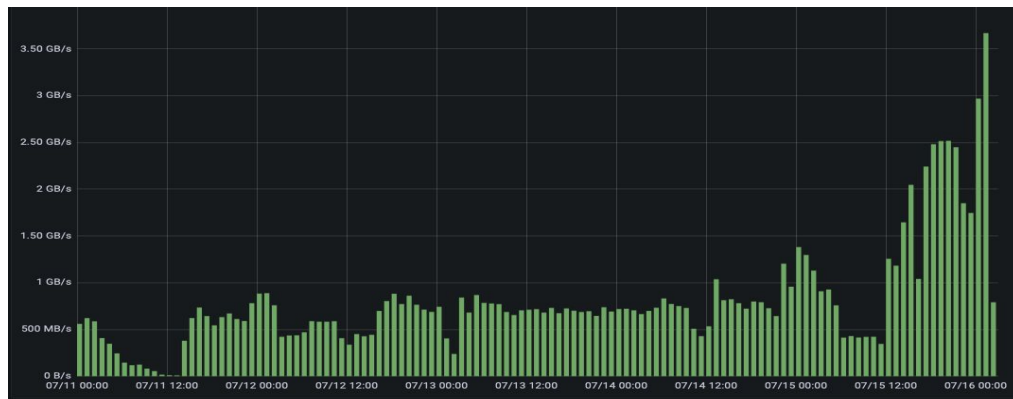
Rubin Observatory Rucio Deployments

- Rubin has had an operational 1.26 Docker-based deployment since the start of 2022
- Kubernetes deployment of 1.29 at SLAC: Dev Late Sept.
- Production deployment to K8s at the end of October

- SLAC is using vCluster from Loft to provision virtual clusters, enabling a multitenant institutional cluster
  - Will utilize separate vClusters on the same backend for development and production
  - Using a Kustomize-based deployment framework
  - Deterministic RSEs. No custom data placement. Limited policies are already in the Rucio core as LSST
- Database at SLAC is deployed in same vCluster with Cloud-Native PostgreSQL, simplifying scaling greatly

🟰 **Fermilab**

Rubin Considerations

- Major Data Facilities: USDF, UKDF, FrDF (CCIN2P3-Lyon)
- Rucio is not being utilized for summit->SLAC transfers of raw camera data
    - Latency is too high due to the periodicity of Rucio
- Many small files
    - Ensure sufficient transfer throughput given the overhead of many small transfers
    - Order of magnitude more files than HEP experiments
    - Storage sites will be using object stores
    - Construct an accurate prediction of total Rucio database size over the lifetime of the experiment
- Steve Pietrowicz is modifying Hermes into Hermes-Rubin
    - Will deliver messages to both ActiveMQ and Kafka
    - Kafka will function as the link between Rucio and the Data Butler (metadata catalog) via a Butler Ingest Service

🟥 **Fermilab**

Operational Lessons Learned

- Continued contributions to documentation
    - Environment variables
    - Multi-VO operations
    - Use of lesser known daemons

- A persistent Slack history would be helpful

- Kustomize can really speed up the deployment process

- Fermilab has to policy package operational support to contribute

🐾 Fermilab

Questions?