DEEP UNDERGROUND
NEUTRINO EXPERIMENT

DUNE Data Management Development at Edinburgh

Peter Clarke, Rob Currie, James Perry, **Wenlong Yuan**

THE UNIVERSITY *of* EDINBURGH

Data Management Development at Edinburgh

- The Data Management task includes all aspects of dealing with the data lifecycle from when it leaves the DAQ
- Edinburgh has been involved in development works to support Data Management, and some works on commissioning and validation as well:
  - Supporting centralised Data Management: e.g. DUNE RUCIO instance development and deployment
  - Monitoring and deployment of storage resources for UK/EU and all sites
  - Commissioning and validation: supporting/testing data access in various environments

RUCIO instance development and deployment

(James Perry)

- Rucio Lightweight Client
  - Removal of unnecessary dependencies on libraries required for uploading and downloading data, allowing lighter weight libraries to be used wherever possible. Optimising the use if the Rucio client as part of workflow and deploying a centralised installation of the client on a CVMFS
- Further objectstore work
  - Rucio's auditor must be modified to support objectstores, allowing Rucio's robust verification procedures to be applied to all of DUNE's storage elements.
- Integration of Rucio with other components of the DUNE data management system
  - Ensure that when new data is ingested, it is always declared both to Rucio and to the metadata catalogue, to avoid these becoming inconsistent. Rucio will need to be integrated with the DUNE Data.

RUCIO instance development and deployment (James Perry)

- Implementing Quality-of-Service Support
  - Rucio supports both tape and disk storage systems, and DUNE uses both. More sophisticated quality of service support in Rucio is required.
- Proximity mapping
  - Rucio uses a “proximity map” to determine how to optimise data movement. A new feature is needed to insert a custom proximity map that can handle not only distances between Rucio RSE's but distance from 3rd party sites that do not have an RSE.
- DUNE-specific Rucio test suite
  - Rucio test suites automatically run on the GitHub continuous integration platform whenever changes are made to the Rucio codebase, allowing for early detection of any regressions that might affect the experiment’s usage of Rucio.
- Protocol ordering
  - deletion of protocols preserves the ordering of the remaining protocols.

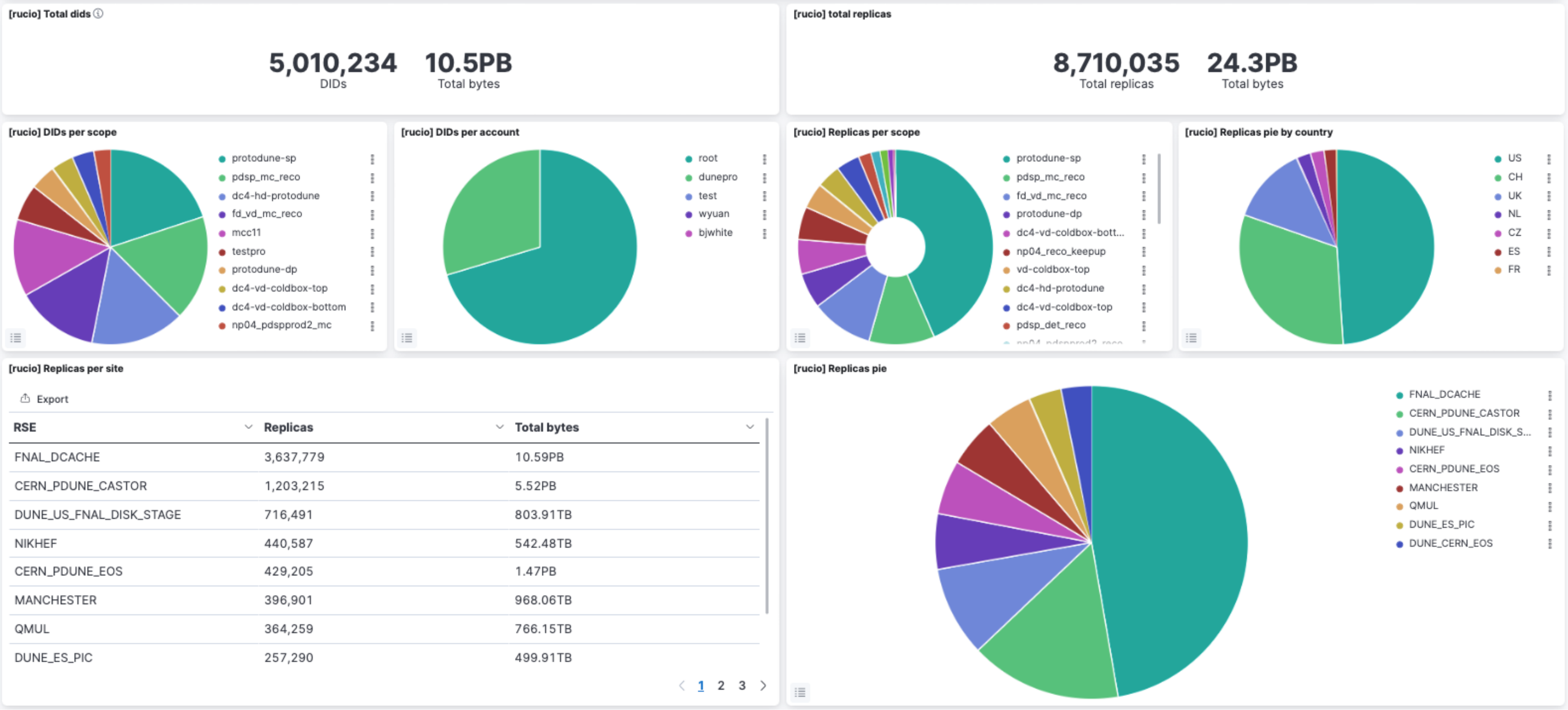
DEEP UNDERGROUND NEUTRINO EXPERIMENT

Development of monitoring system

(Wenlong Yuan)

- Rucio monitoring
  - In a good shape at the moment, useful in the weekly DUNE computing Ops meeting, monthly DUNE UK computing meeting
  - Expected to see more parameters/details, e.g. the files heatmaps in the monitoring to find out which files are popular and where is the closest file replica locations.
  - Plan to study how to use AI/ML technology to analysis tons of monitoring data source
- Developing and combining the DUNE data management monitoring system
  - DUNE has various data management monitoring sources, e.g. Rucio, data pipeline monitoring, job status, storage usage, etc.
  - These will be harmonised and combined in one place, plan to use monitoring facility at Edinburgh, interact with Messaging Queue at FNAL
  - Will further develop a new DUNE data management transfer monitoring system based on FTS monitoring, a new workflow job status monitoring system as well

Rucio monitoring at Edinburgh



[rucio] Total dids

5,010,234 DIDs

10.5PB Total bytes

[rucio] total replicas

8,710,035 Total replicas

24.3PB Total bytes

[rucio] DIDs per scope

[rucio] DIDs per account

[rucio] Replicas per scope

[rucio] Replicas pie by country

[rucio] Replicas per site

Export

RSE	Replicas	Total bytes
FNAL_DCACHE	3,637,779	10.59PB
CERN_PDUNE_CASTOR	1,203,215	5.52PB
DUNE_US_FNAL_DISK_STAGE	716,491	803.91TB
NIKHEF	440,587	542.48TB
CERN_PDUNE_EOS	429,205	1.47PB
MANCHESTER	396,901	968.06TB
QMUL	364,259	766.15TB
DUNE_ES_PIC	257,290	499.91TB

‹ 1 2 3 ›

[rucio] Replicas pie

https://dune.monitoring.edi.scotgrid.ac.uk/app/dashboards#/view/7eb1cea0-ca5e-11ea-b9a5-15b75a959b33

Commissioning and validation

(Wenlong Yuan)

- Supporting/testing data access in various environments
- Data Challenges
  - We need to be centrally involved in the DUNE participation in all data challenges
- RSE commissioning:
  - When new storage resources from various faculties join DUNE, full testing and commissioning is needed. We've commissioned and bring lots of disks online, e.g. IN2P3, RAL-PPD, QMUL, SURFSara, etc, and tapes from CERN and RAL.
  - The commissioning includes VO authentication test, direct, multi-hop, TPC transfers tests
  - Will continue the RSE commissioning work as more and more European and worldwide (UK, NL, FR, CERN, India, etc.) tape and disk sites will join DUNE

Other Data management stuff

(Wenlong Yuan)

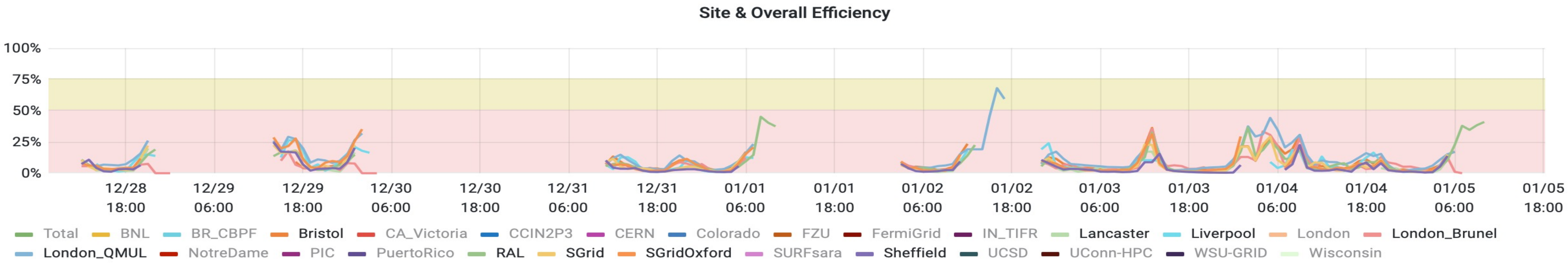
- XCache and Virtual Placement (VP) for DUNE
  - Edinburgh is the first XCache and Virtual Placement (VP) test site for DUNE
  - XCache has been adopted by various VOs, e.g. ATLAS, as well as the ATLAS Virtual Placement
  - DUNE and LBNC would like XCache and VP work for DUNE and for many small sites. Edinburgh will be a centre for development and commissioning
- Token migration:
  - The OSG has formally dropped x509 based authentication in favour of “Tokens”, whilst the WLCG (Europe) has a much longer timescale (at least until after LHC Run-3) for migration to Tokens. We must mitigate this by hybrid operation for the rest of the period.
- DAQ interface and data format evolution in HDF5
  - Raw data files from the DAQ are written in Hierarchical Data Format (HDF5) format. We plan to be working with the DAQ group to interface HDF5 to the computing chain. This will be a continuous process as the data format will continue to evolve after ProtoDUNE.

Other Data management stuff - StashCache

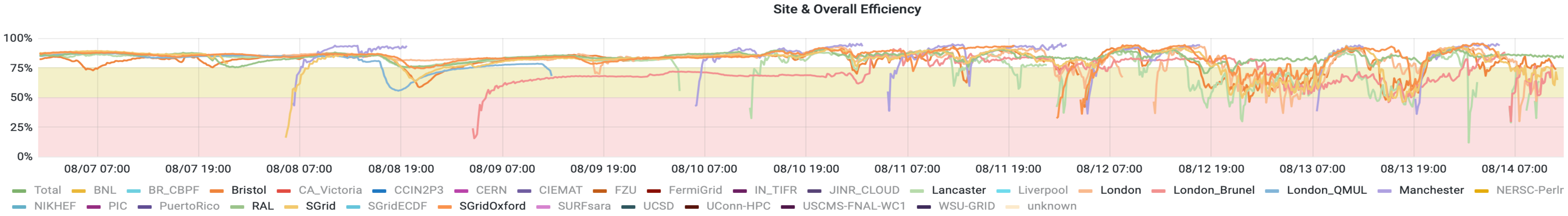
DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

(Rob Currie, Wenlong Yuan)

- The Edinburgh StashCache is running smoothly since April 2022, solved the low efficiency problem when processing “Flux files”
- The Edinburgh StashCache has not only been of great benefit to DUNE but also supports all US OSG VOs running in the UK, except LIGO
- A talk about this on CernVM Workshop



Dec 21’ - Jan 22’



Aug 22’

THE UNIVERSITY of EDINBURGH

Summary

- Working to support DUNE Data Management activities within the UK
- Contributed to different areas of data management development, commissioning and validation, within the experimental workflow
- Will contribute to essential, but unforeseeable, data management and related tasks that arise over the project