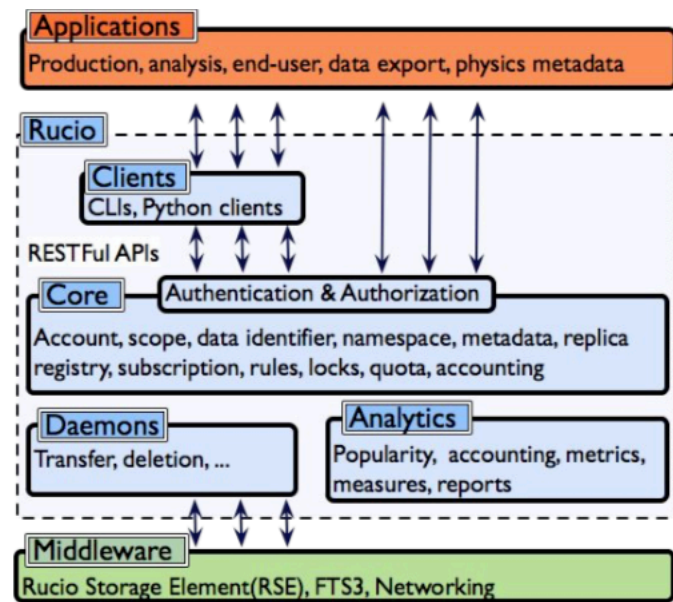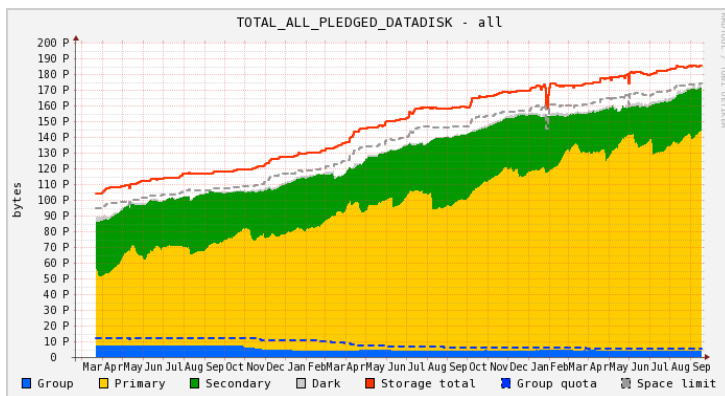# Rucio

Alastair Dewhurst

What is Rucio?

- Rucio provides a complete and generic scientific data management service

    - Made by ATLAS with more than 10 years of operational experience in large-scale data management.

    - CMS are moving to Rucio during the long shutdown.

    - DUNE and LIGO also using it.

Advantages

- Users

  - Policy based data management.

  - Hardened against transient failures.

- Sites

  - Supports different site setups and protocols.

  - Provides an abstraction layer between users and storage.

- Both

  - Large user community.

  - Active development team that are leading data management evolution in the WLCG.

Tier-1 Strategy?

- We have perceived a need for complex data management and movement capabilities.

- Having reviewed the landscape I believe Rucio is best position to fulfil these future needs.

- From experience we believe that the best way to run this as a central service.

- The Tier-1 has decided to run a Rucio instance.

  - Is developing Multi-VO capabilities.

Rucio setup at RAL
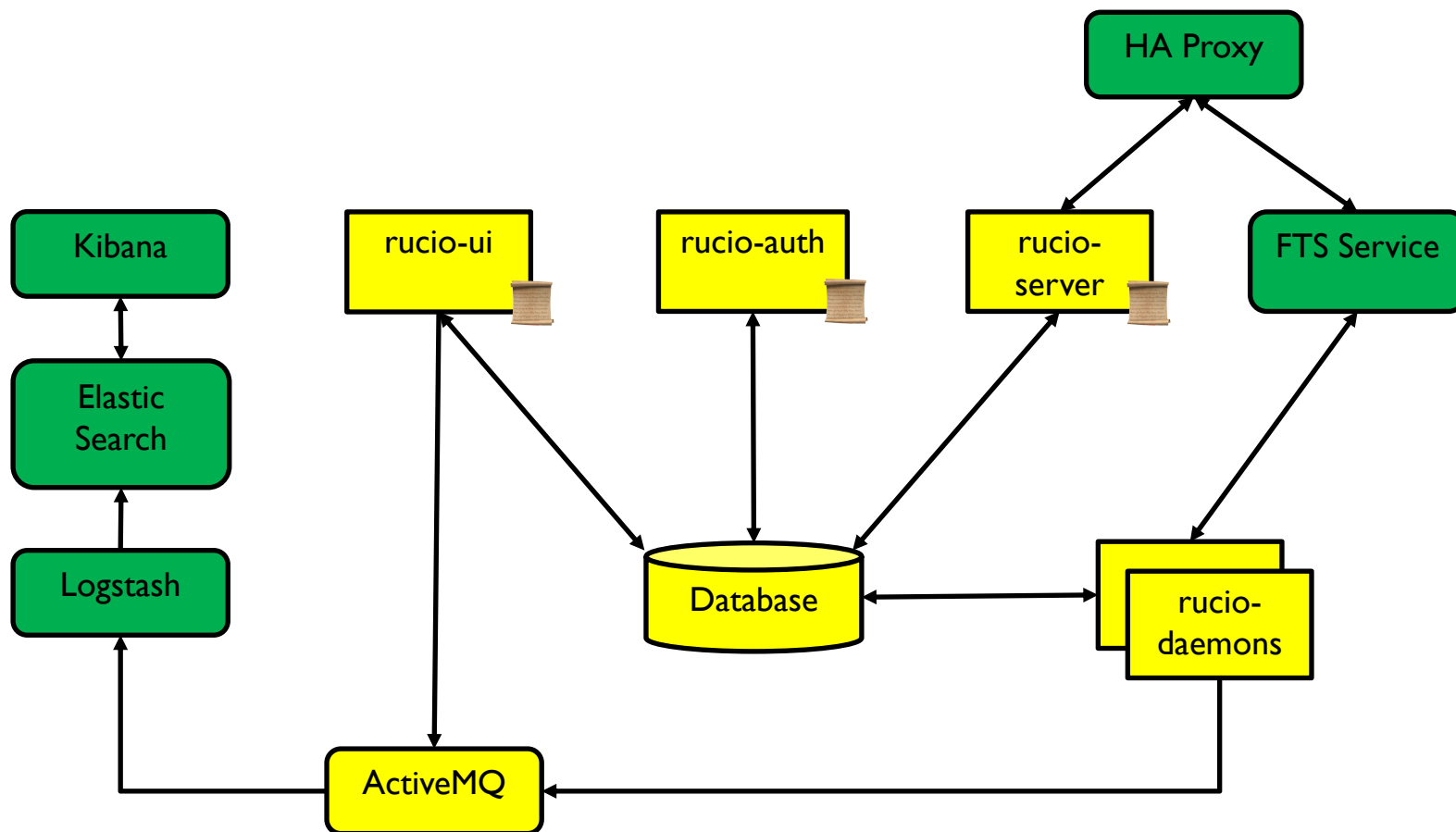
Current status

- Working instance for AENEAS project.

- Continuing to work on improving monitoring.

  - Transfers logs being sent to Elastic Search

- Operational cost ~0.1FTE

  - Security patching has been performed twice.

  - Two (minor) Rucio upgrades performed.

Rucio Monitoring

- ~80 Rucio metrics are being pushed into Elastic Search

- Some basic dashboard are available internally.

  - Transfer matrix for AENEAS sites.

- Still a way to go before we consider it production ready.

  - Also need to make plots public.
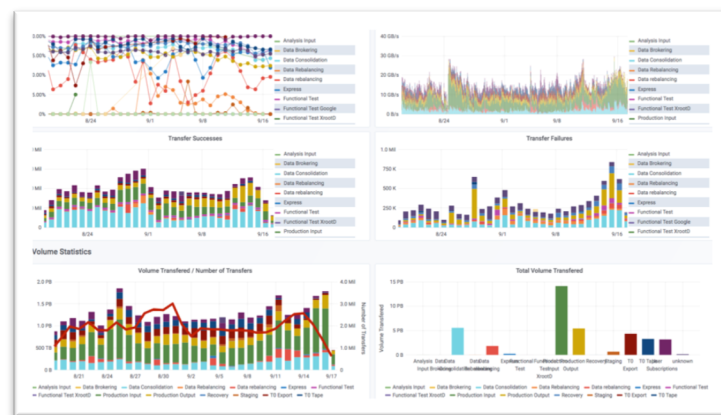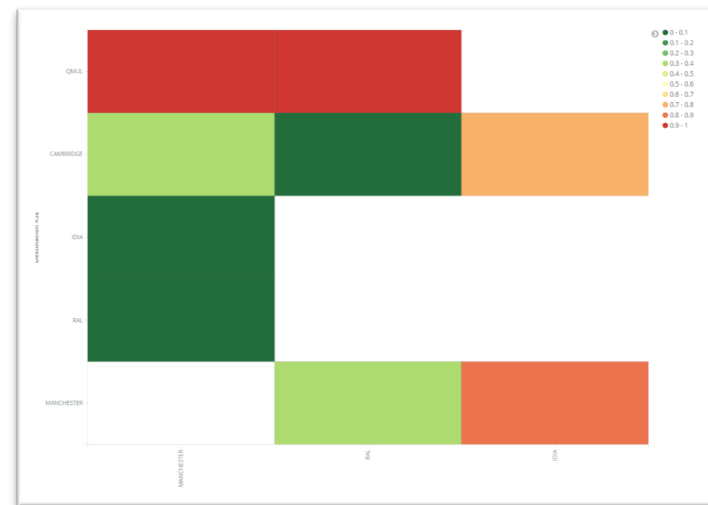
Multi-VO Rucio

- We want to add a VO field in Rucio, to allow a single instance to support multiple experiments.

- Andrew Lister (RAL) started work on 6th March to develop multi-VO support for Rucio.

  - Supervised by Ian Johnson.

  - Has spent time at CERN sorting out database changes.

- On course for completion in September 2019.

  - Previous talk by Daniela should have covered their work plan for integration with DIRAC.

[1] https://docs.google.com/document/d/1RzrEhRiTCzpylf3Fpp6SLGc5svfJOLjiI0dCo-zUPos/edit#heading=h.3tdjqhs1g071
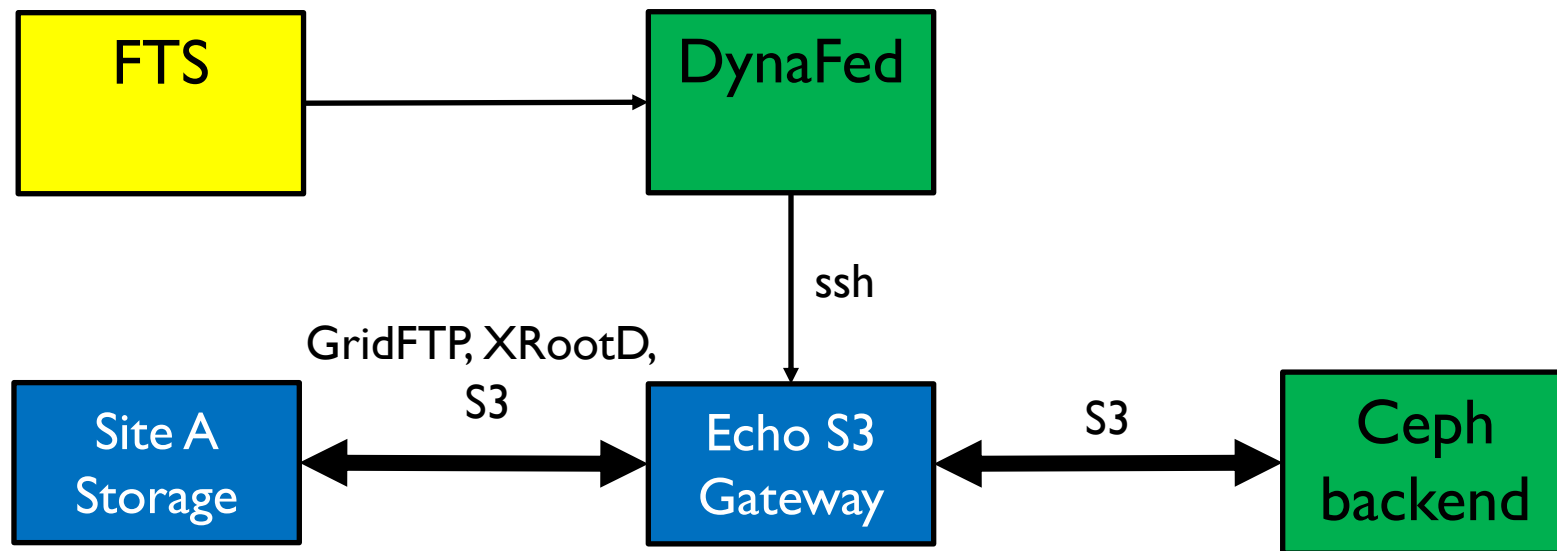
Object Stores

- James Perry (ECDF) is working on supporting object stores in Rucio (for DUNE).

    - Working on Rucio talking to S3/Swift directly.

- RAL are working on DynaFed integration with S3.

- Further joint digital asset request from RAL and ECDF to continue to improve:

    - Object Store support

    - Monitoring

DynaFed integration

```
┌──────────┐                    ┌──────────┐
│   FTS    │ ─────────────────▶ │ DynaFed  │
└──────────┘                    └──────────┘
                                      │ ssh
    GridFTP, XRootD,                  ▼
          S3
┌──────────┐      ┌──────────┐      ┌──────────┐
│ Site A   │ ◀──▶ │ Echo S3  │ ◀─S3─▶│  Ceph   │
│ Storage  │      │ Gateway  │      │ backend  │
└──────────┘      └──────────┘      └──────────┘
```

- Using DynaFed to make a cloud endpoint look like a grid endpoint.

- Authentication and Authorisation via Grid-mapfile plugin.

- Multi-part uploads were fixed in December 2018.

Managing the Service?

- The Rucio instance at RAL is built on the STFC Cloud.

  - With the exception of the backend database.

  - Relatively straight forward to give external collaborators access.

- How should administrative access be distributed?

AENEAS work

- We are in discussion with Manchester about getting a CDT student to work on AENEAS.

- Plan to test transfers via DTNs from remote locations (e.g. Australia).

- Should we use this opportunity to integrate more UK sites in the Rucio instance?

Conclusions

- Operational Rucio instance for AENEAS for almost a year.

- Digital asset on course for delivery in September 2019.

- Building Rucio expertise across GridPP Collaboration.