

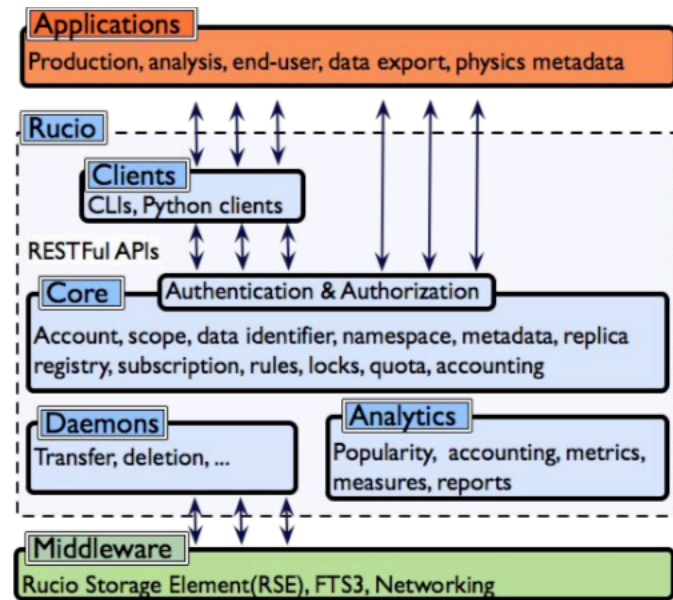
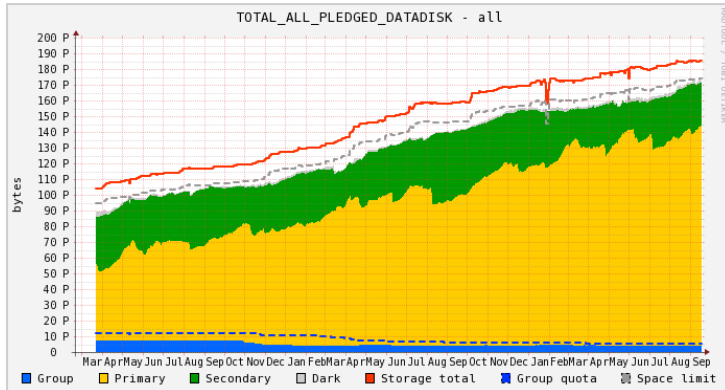
Multi-VO Rucio

Alastair Dewhurst

What is Rucio?

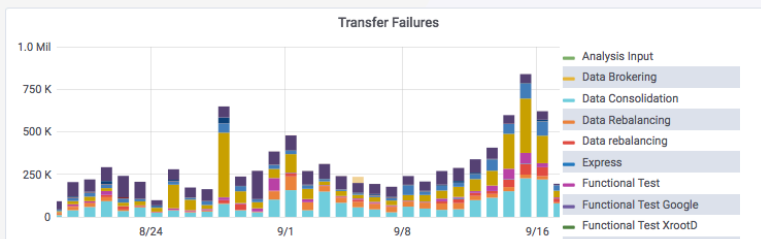
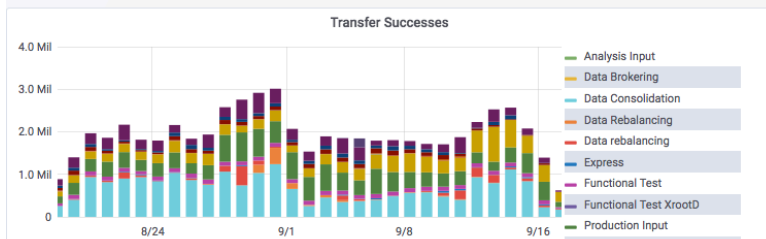
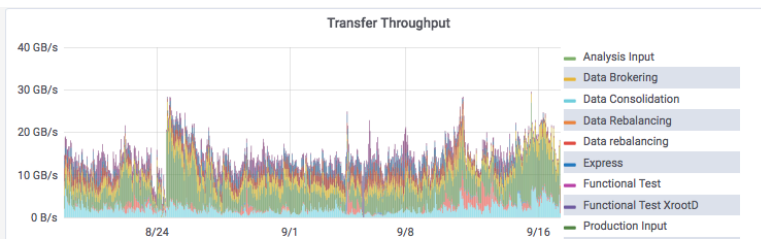
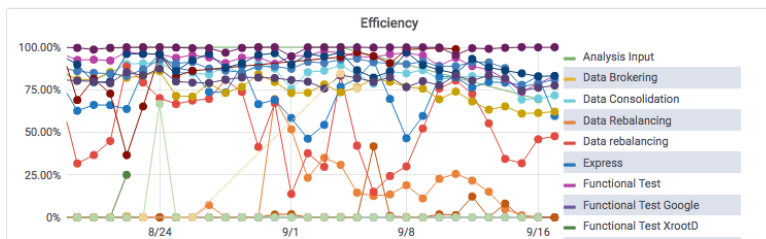
- Rucio provides a complete and generic scientific data management service
- Made by ATLAS with more than 10 years of operational experience in large-scale data management.

ATLAS Store around 1 billion files which take up 400PB of data on ~120 sites.

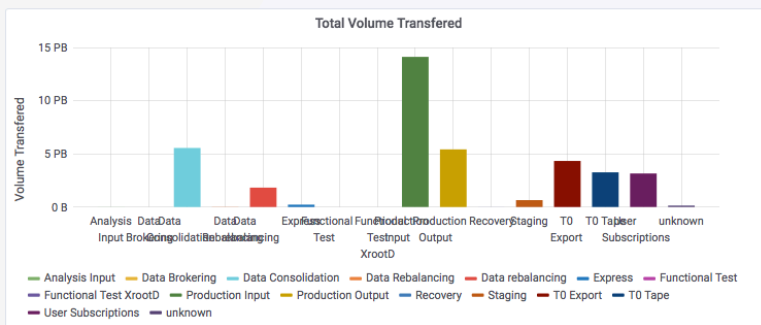
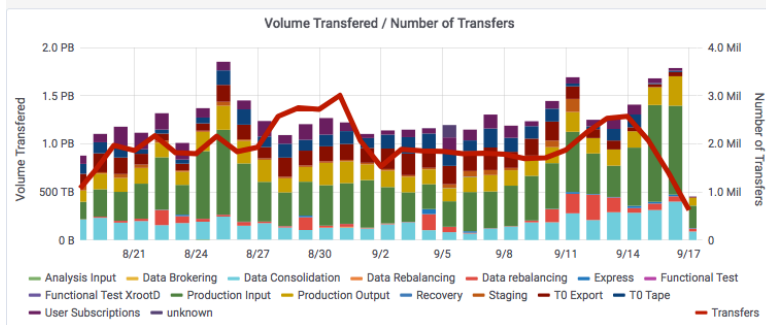


Usage

ATLAS are accessing / transferring 2 - 3 million files a day. This corresponds to 1 - 1.5 PB of data with an aggregated total throughput of 10 - 20 GB/s.
 Note: Rucio is handling 250 thousand failures a day too!



Volume Statistics



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.



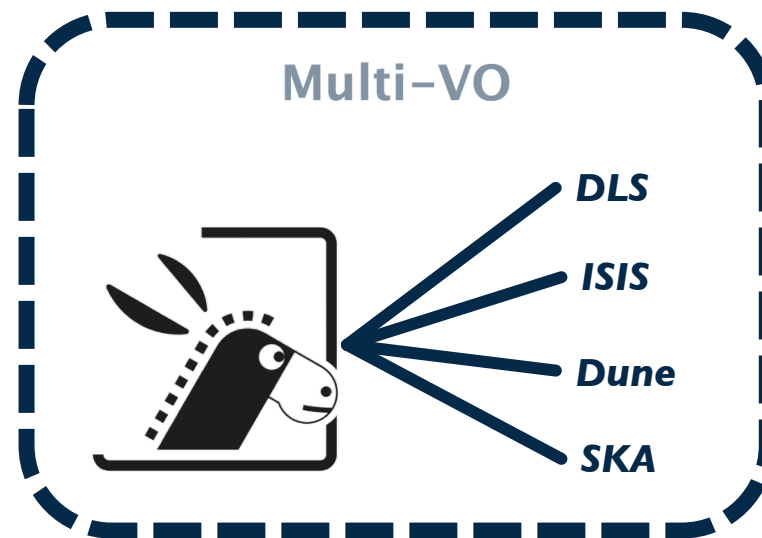
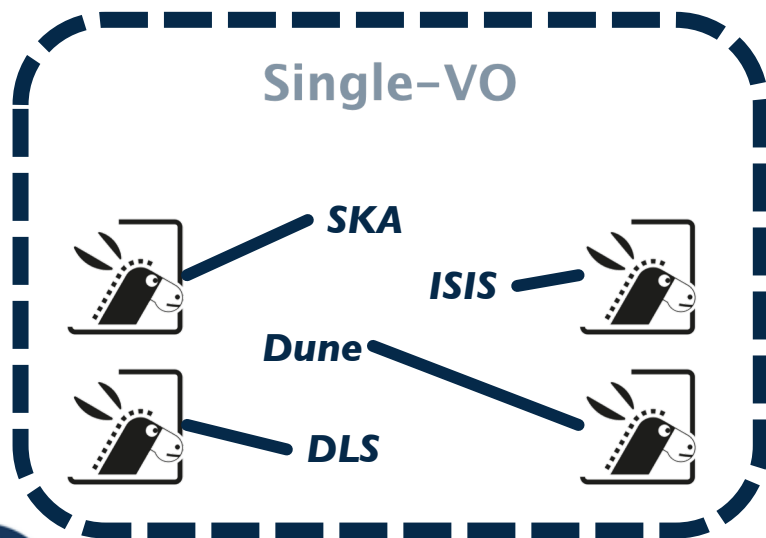
Use Cases

- Types of users:
 - Small < 1PB of data, < 1 million files
 - Medium, < 10PB of data, < 10 million files
- In the long term, we expect to be supporting lots (10+) small users and a handful of medium users.
- We could create Rucio instances in Kubernetes for each experiment.
 - Significant overhead per Experiment.
 - Long term support is more complicated.



Multi-VO Rucio

- We want to add a VO field in Rucio, to allow a single instance to support multiple experiments.
- 12 Month IRIS Digital Asset:
 - 3 months to setup a production quality instance
 - 9 months to develop Multi-VO support.

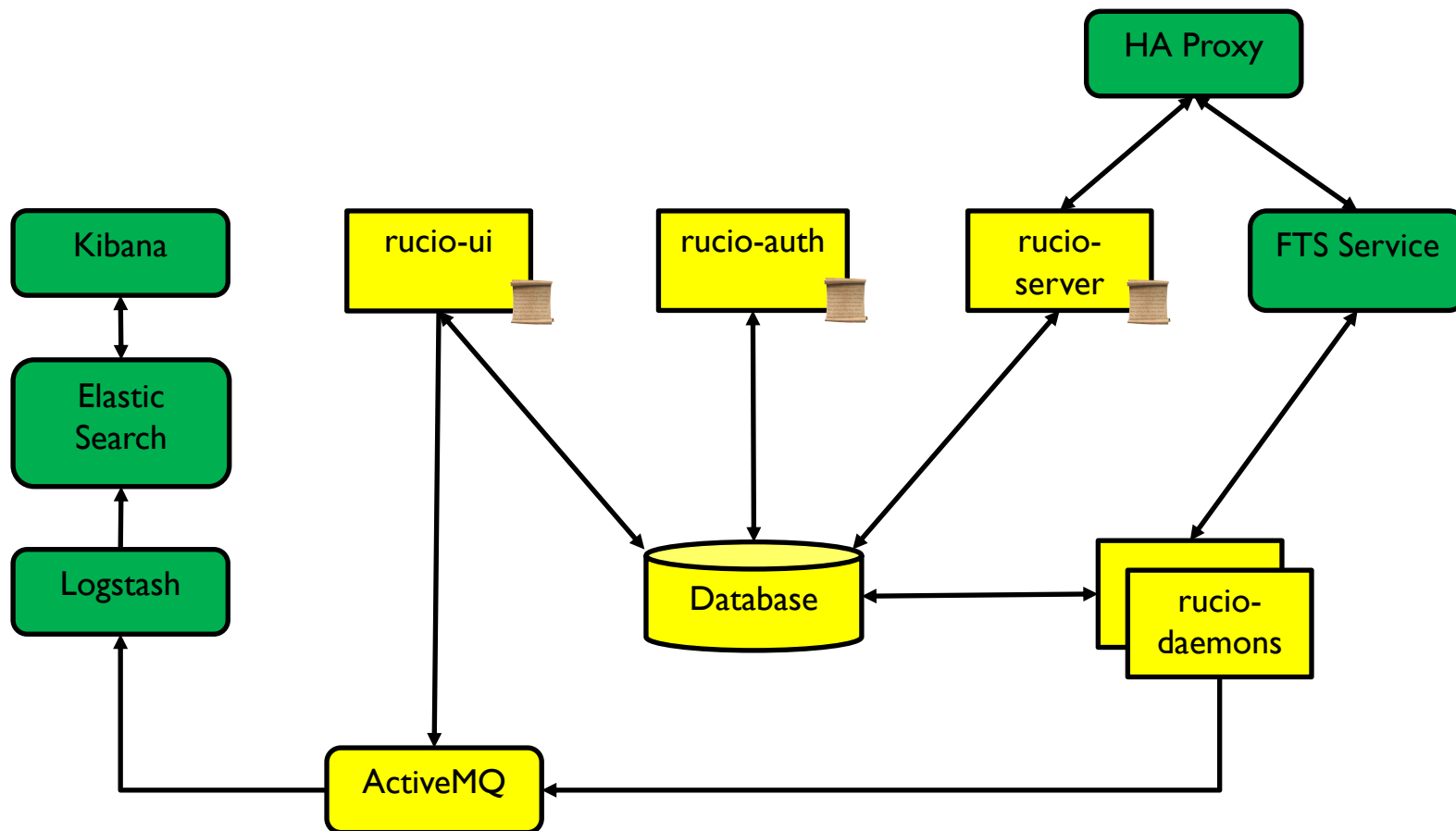


Current Status

- We want to add a VO field in Rucio, to allow a single instance to support multiple experiments.
- Andrew Lister worked between March and September 2019 to develop multi-VO support for Rucio.
- Extremely productive development period achieved around 90% of the tasks we needed in 6 months, when we budgeted 9 months.
- Eli Chadwick took over in September 2019.
- His work is currently on hold while he is on a 3 month placement at ILL.
- Full multi-VO support will be in the next Rucio release 1.22 expected in February 2020.



Rucio setup at RAL



Plans

- Long term strategy for Tier-I to provide Multi-VO Rucio for non-LHC VOs.
- Replacing legacy services like LFC.
- Working with ECDF on Object store support and monitoring. (See their talk later).
- Collaborating with Imperial to integrate multi-VO DIRAC and Rucio.
- Will be applying for EGI funds to provide multi-VO instance for EGI.



Rucio demo

- ~~This afternoon Ian Johnson will present a short Rucio demonstration.~~
- Ian is unfortunately ill today, I will see what I can manage.
- This will:
 - Configure a sites storage to be a Rucio Storage Element (RSE).
 - Create a dataset and upload files to it.
 - Create a rule to trigger a transfer of this dataset.
- We want to find out about IRIS site storage.
 - Will any further development work be required?



Conclusions

- Operational Rucio instance for AENEAS for almost 2 years.
- Digital asset on course for delivery in March 2020.
- Building Rucio expertise across GridPP Collaboration.

