Rucio Community Workshop

Summary and highlights

Mario.Lassnig@cern.ch
on behalf of the Rucio team

https://indico.cern.ch/event/676472/

In autumn 2017…

- Three experiments were already using Rucio (ATLAS, AMS, Xenon1t)
- Some personal contacts to other interested communities have been established
- ATLAS S&C data strategy for the long-term was getting more concrete
- Significant contribution to the HSF community process for data management

→ Idea was born to do the "1st Rucio Community Workshop"

Workshop objectives

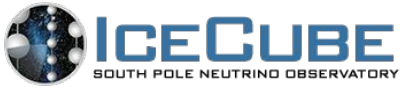
- Present Rucio to a larger audience beyond ATLAS
  - Design, concepts, operations, and demonstration of the system
  - And also the people behind it!
- Community-building and strengthening
  - For the experiments using Rucio
  - For the experiments which are evaluating Rucio or are interested to know more
  - And also gather information and feedback
- Discussing the future
  - The software development roadmap for the long-term
  - Development incubator
  - Identifying potential collaborations

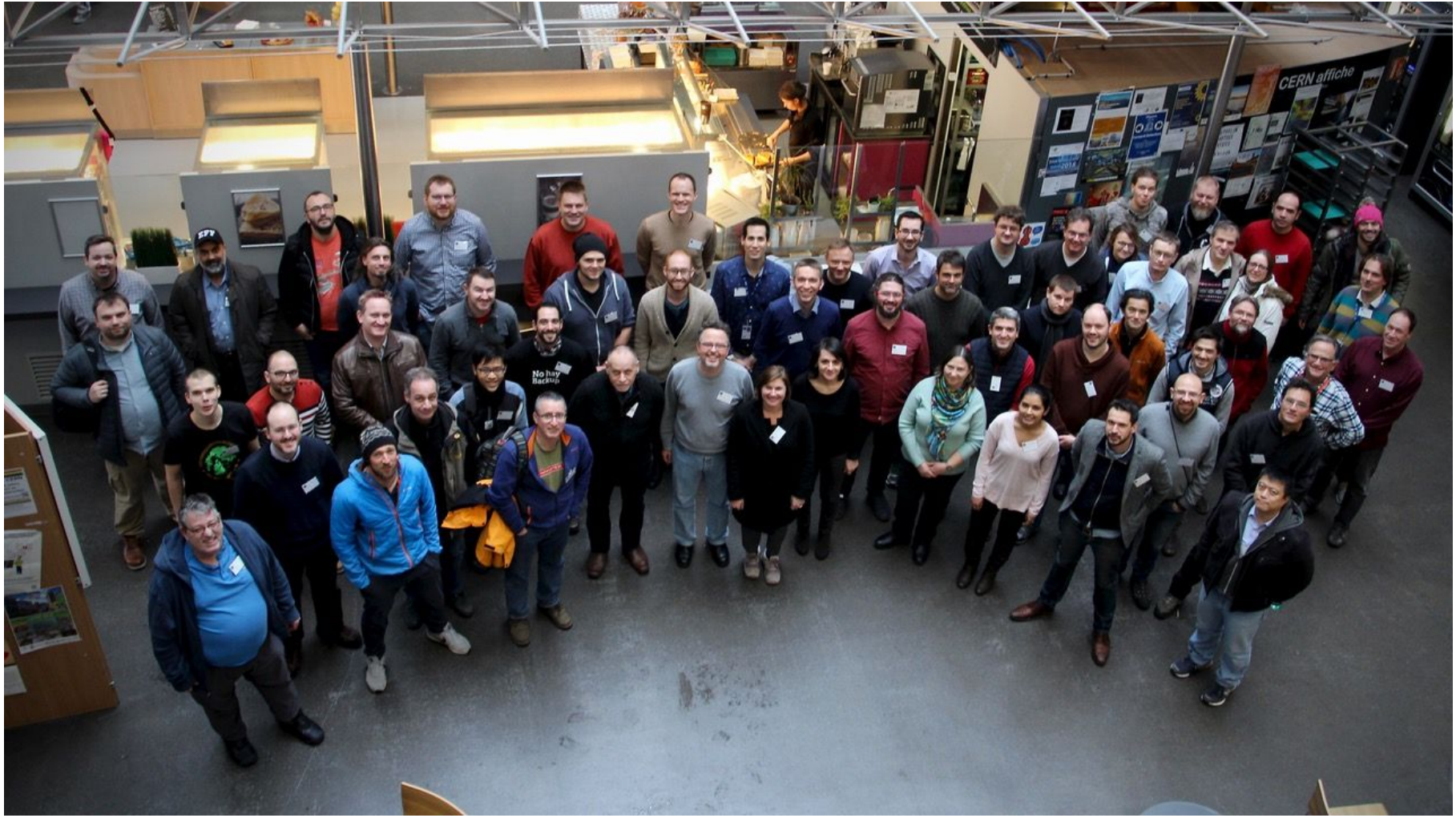
Overwhelming response

Google Cloud Platform

- 90+ persons registered
- Representing more than 14 communities

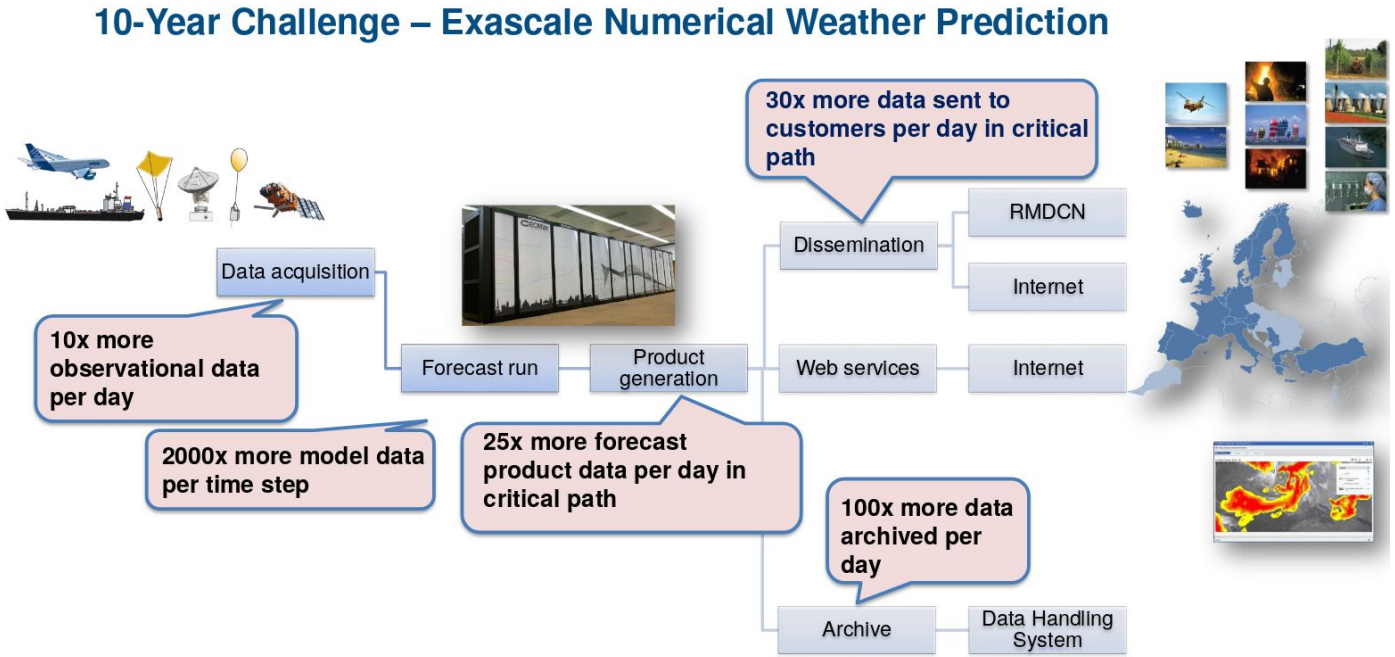
- This workshop marks the start of the process to evolve Rucio as a common distributed data management system for HEP and non-HEP




CERN affiche
No hay Backup

What is the value of a good weather forecast?

Thursday Keynote — ECMWF



Thursday morning session — Presenting Rucio

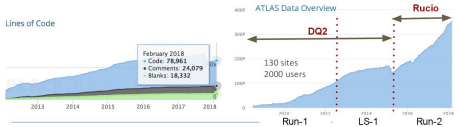
- Prologue: A history lesson
- Chapter 1 — Namespace
- Chapter 2 — Storage
- Chapter 3 — Replica management
- Chapter 4 — Basic usage
- Chapter 5 — Advanced usage
- Chapter 6 — Experiment perspective

Integration with Workload Management

- Rucio is integrated with ATLAS Workload Management System (PanDA/Prodsys)
- Rucio takes care of
  - Transferring of data (input/output) for the jobs
  - Staging data from TAPE
  - Notifying Panda when transfer of dataset is done
  - Notifying prodsys in case of lost or corrupted files
- Rucio also reads the PanDA tables to identify the datasets currently used by the jobs and can make extra copies to improve the brokering
- Rucio can possibly be used by any other Workflow Management system

Rucio Development & Commissioning

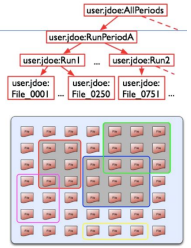
- Long initial process:
  - 2012: User surveys, technical studies & design phase ~1 year
  - 2012-2014: Initial development ~2 years
  - 2015: Commissioning & gradual migration from predecessor system DQ2 ~1 year



Data Hierarchy

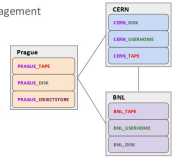
- At the heart of everything is a file*
- Files are grouped into datasets
- Datasets are grouped into containers
  - Datasets only hold files
- Containers are grouped into containers
  - containers only hold datasets or containers
- Collections can be organised freely
  - Files can be in multiple datasets
  - datasets can be in multiple containers
  - containers can be in multiple containers

* sub-file support being explored



RSE grouping and metadata

- The set of all RSEs form the topology for data management
- Fixed set of necessary metadata
  - `name, type, availability, ...`
- To give users flexibility each RSE can host an arbitrary amount of custom metadata
- Can be set and redefined at runtime
- Can be used to build replication rules (cf. Ch 3)
- Examples
  - `is_tape=True`
  - `user_software_available=True`
  - `min_free_space=10000`
  - `region=Europe`

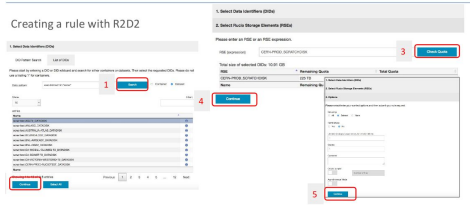


Replication rules II

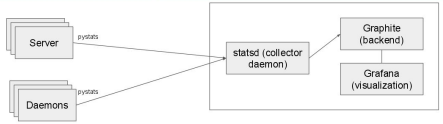
- Required parameters for a rule are: number of copies, did, RSE Expression
  - E.g. 3 copies of data:dataset1 on `(country=de|country=fr)&type=disk`
- Rules get enforced continuously, but are not re-interpreted
  - Order of creation matters:
    - 1 copy on RSEa; 1 copy on RSEa|RSEb → 1 physical replica on RSEa
    - 1 copy on RSEa|RSEb; 1 copy on RSEa → 1 physical replica each on RSEa and RSEb
- Rules can have a lifetime, after which the replicas become eligible for deletion
- Grouping options (for rules on dataset/container) for rules
  - DATASET (When rule is on a container): Each dataset is distributed to the same RSE
  - NONE: All files of the dataset/container are distributed randomly
  - ALL: All files of the dataset/container are distributed to the same RSE
- Different notification modes to notify creator as well as external applications

WEB UI

Creating a rule with R2D2



Internal Monitoring Architecture

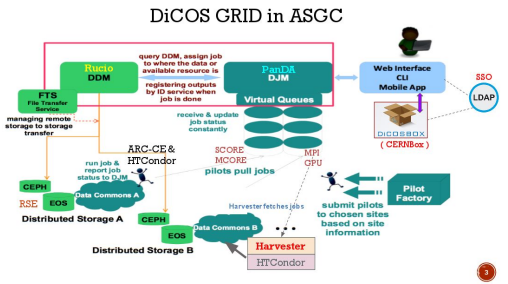


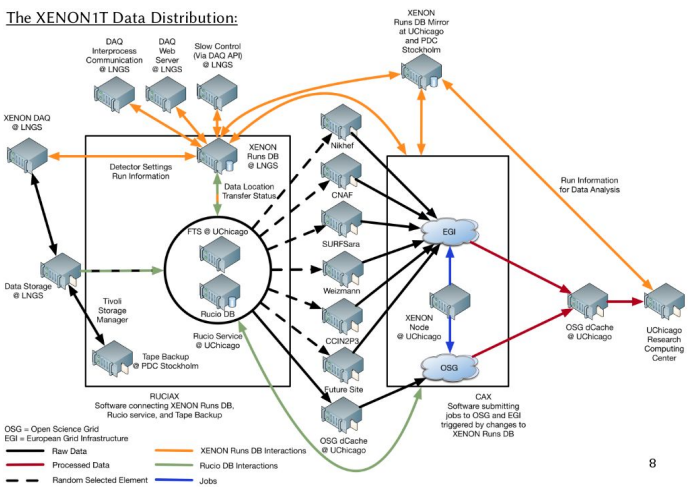
- Easy to set up
  - Docker images available both for Graphite and Grafana
  - Otherwise a manual install is also easily possible
  - Then only need to configure Rucio to point to the Graphite host and data will automatically start to flow in
  - Dashboard templates for Grafana can be provided

Thursday afternoon — First community session

- Communities already using Rucio
- ATLAS, ASGC/AMS, Xenon1t

- Main things discussed
  - Deployment and operational experiences
  - Integration experience with existing systems
  - Split between WFMS and DDM — which drives which
  - Python 3 support
  - Flexible permissions and policies



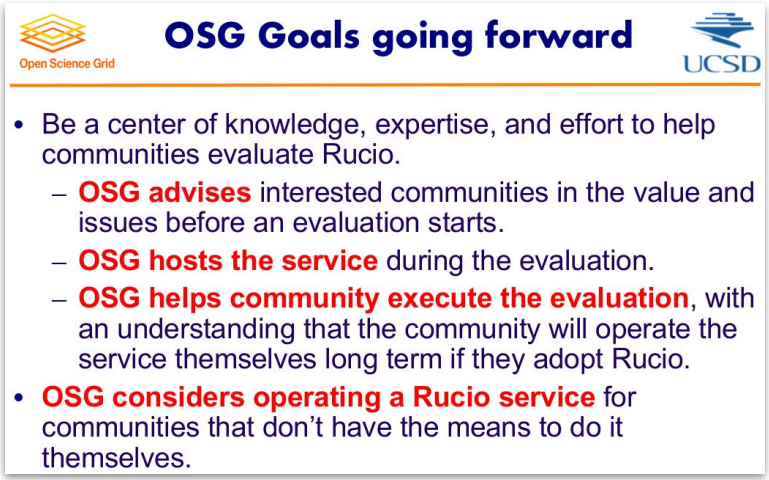


Thursday afternoon — Second community session

- Testbeds: CMS, LIGO, IceCube
  - Combined effort from OSG, FermiLab, and UChicago, with support from the Rucio team
- OSG considering hosting "Rucio as a Service" for OSG experiments
  - Without taking ownership of experiment's data(!)
- NA62, EISCAT_3D presented their data needs

Summary

- LIGO GW archival data: frame files, various data reduction levels
- LIGO Data Replicator (LDR): mature/stable replication tech.
- Rucio -> *potential* drop-in replacement with long-term support, large dev- & user-communities
- LIGO rucio evaluation: early days but progressing well, broad interest for future observing runs
  - Replication rules, subscriptions & integration with broader HTC community particularly appealing
- Short/medium-term needs: archival frame data management & replication
- Potential longer-term interest in post-processed data products:
  - Short-time Fourier transforms, newer `cleaned' data, analysis output (e.g., posterior samples)



Thursday afternoon — Community session 2/3

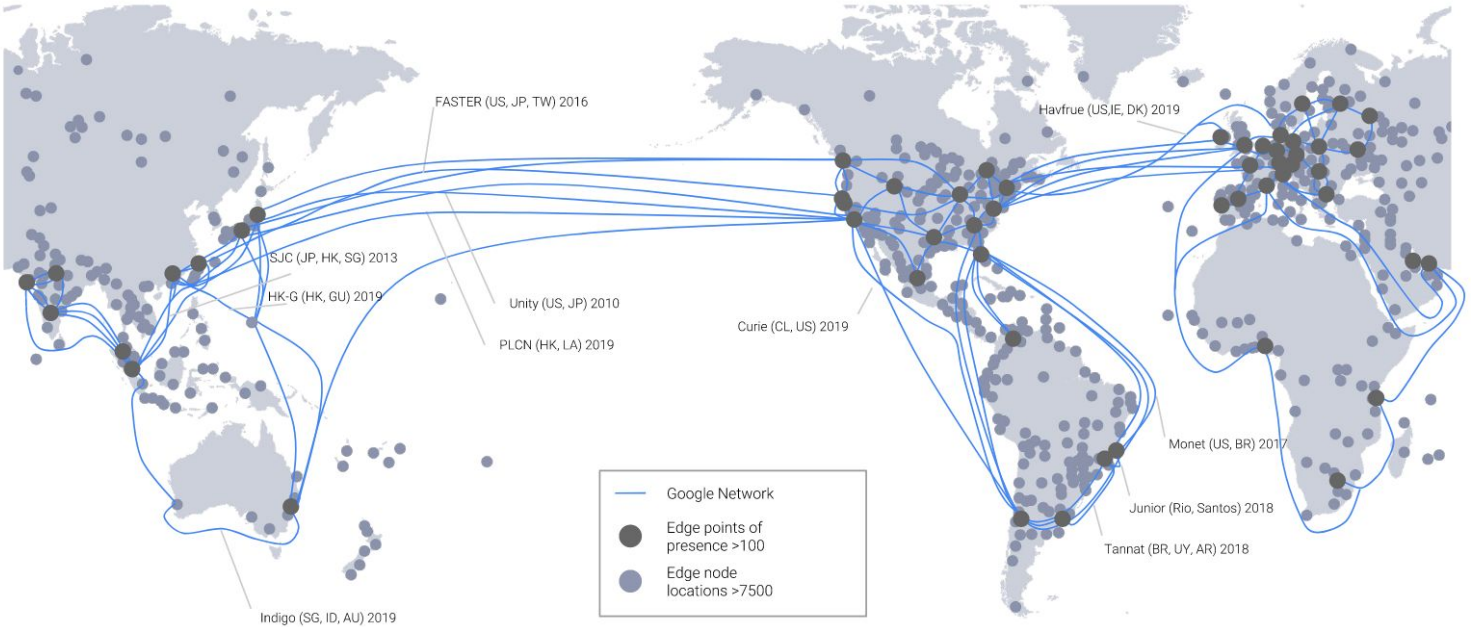
Conclusions

- Rucio meets CMS's immediate scalability needs and is a good enough fit to our existing data model
- Rucio developers have been very accommodating and encouraging
  - It is a concern for CMS if the effort continues to be owned by ATLAS
  - Community project would be ideal
- U Chicago system has been instrumental in helping with a quick start
- We still have some milestones to meet to show that CMS should adopt Rucio, but we are all optimistic they can be met before the summer review
  - We now know we *could* adopt rucio
  - Transition would take place during 2019-2020 LHC shutdown
  - Still need to map out exactly how this would happen
    - ★ Need to explore how to run both in parallel

Friday Keynote — Google

Google Network

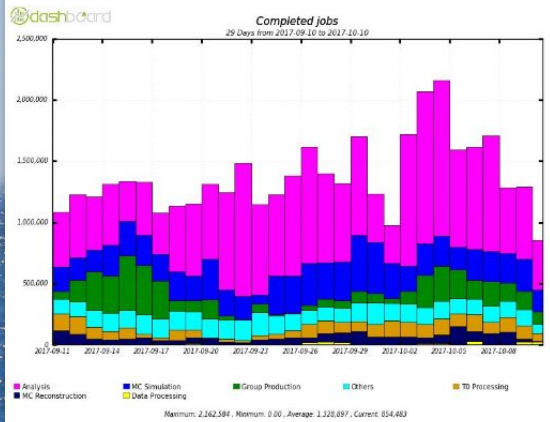
The largest cloud network, comprised of 100+ points of presence



Keynote — Google

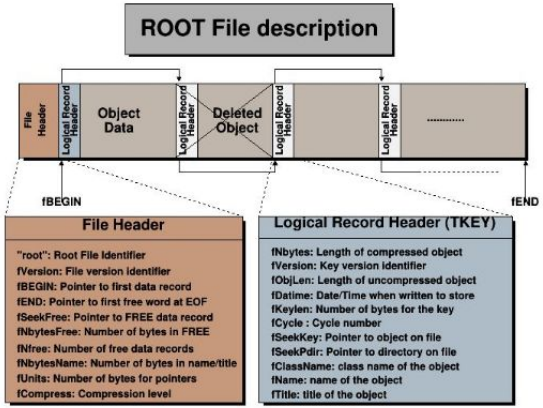
Keynote — Google

User Analysis



Data Analysis, Replication and Placement

Data Streaming



Google Cloud

Confidential & Proprietary

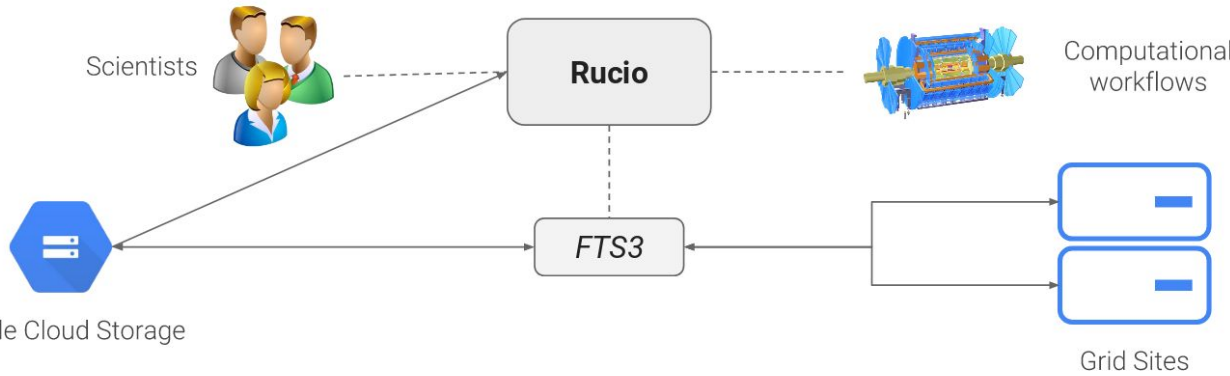
Indigo (SG, ID, AU) 2019

Edge node locations >7500

Keynote — Google

Google Cloud Storage and Rucio integration

- Rucio supports S3-style buckets out of the box
- Functionality was added for orchestrated file transfers
- ATLAS can now use Rucio replication rules to move data in and out of GCS
- Fully integrated with ATLAS computational workflows

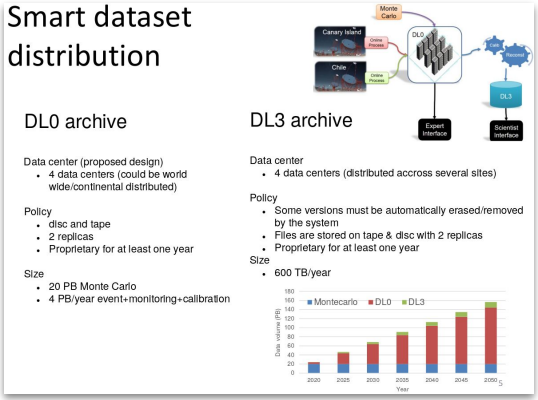


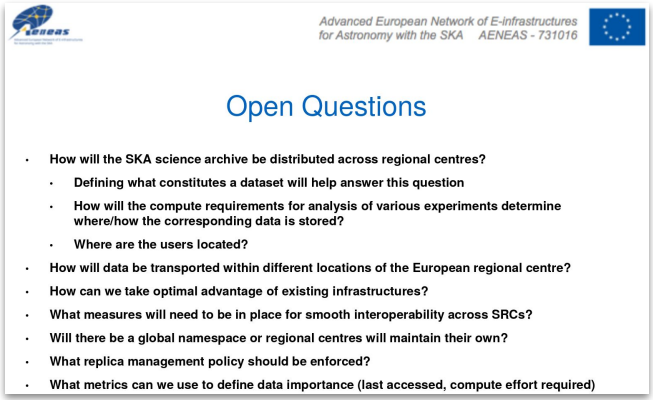
Google

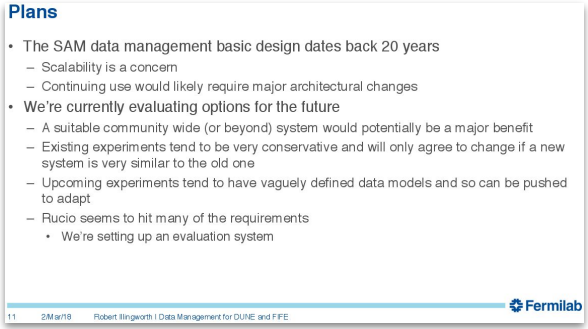
Confidential + Proprietary

Friday morning — Third community session

- ARC Cache support already integrated in Rucio
- COMPASS wants to start evaluating Rucio
- CTA needs smart data distribution using OAIS flexible metadata
- SKA needs a 300+PB/year scalable data management system
- DUNE and FIFE (Frontier Experiments at FermiLab) are looking for SAM replacement

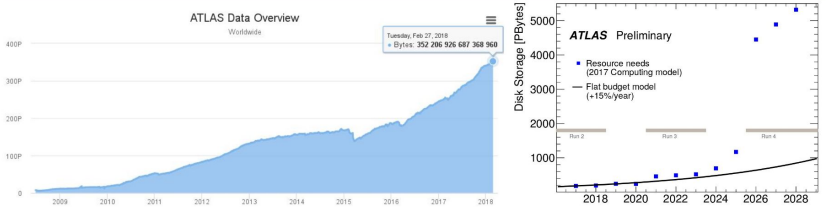




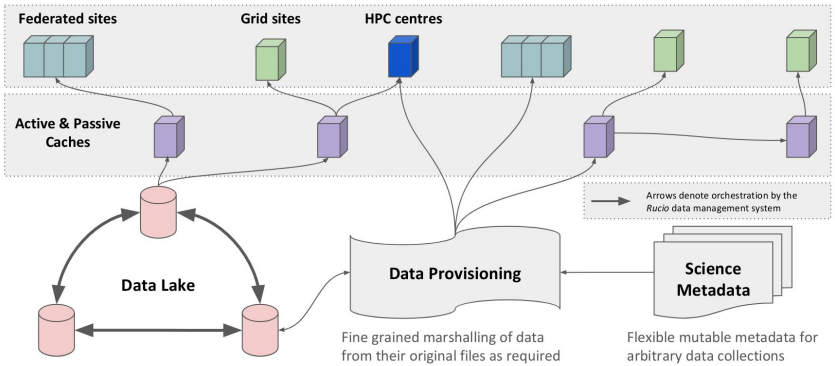


Friday morning — Discussion session

Long-term development strategy



- **Major challenges that will drive HEP data management R&D in the next 5-10 years**
  - Factor 10+ increase in storage requirements (>5 Exabytes)
  - Factor 100+ increase in number of objects to manage (>100 billion)
- **Long-term data management R&D strategy**
  - Explore the usage of large reliable data centres ("lakes") with wide area cache control
  - Explore sub-file data object processing to better utilise heterogeneous compute and associated networks
  - Stronger integration of software (file formats, new IO patterns, interactive analysis) with distributed computing



- Scalability vs Python
- Subfile support and data formats
- Flexible metadata support
- Database support
  - PostgreSQL, non-relational
- Authentication & Authorisation
  - EduGAIN, SciTokens, …
- Workflow management support
  - DIRAC, HTCondor, Pegasus
- More transfertools
  - GlobusOnline, WebDAV, HTCondorFile
- Deployment models
  - Dockers, Multi-VO Service, …
- Dependent services
  - ActiveMQ, FTS, …

Friday afternoon — Technical & demo session

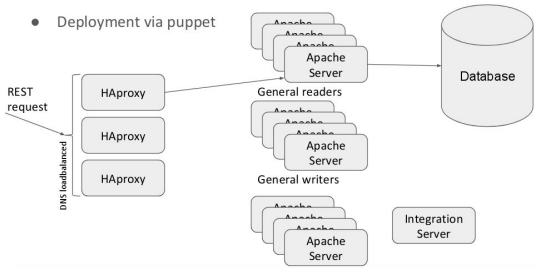
How to contribute 1/3

- Join our slack channel to interact with the other developers
- Fork the repository on GitHub
  - Read / Write issues
  - Review / Comment Pull Requests
  - Submit Pull Requests
- Read the documentation on readthedocs
- Download the packages from PyPi
- Download docker images on docker hub
  - Server, daemon, UI, clients, dev
  - Pushed automatically on new release

2018-03-01

ATLAS deployment

- Deployment via puppet



2018-03-01

https://rucio-blueprint.readthedocs.io/en/latest/rucio_demo.html

Testing dataset upload, creation of dataset and rules

There are no datasets created yet. To generate datasets and copy them to one of the RSEs, you can use a daemon called automatix:

```
[root@3a6d4527e1f6 rucio]# /usr/bin/rucio-automatix --run-once --input-file /opt/rucio/etc/automa
...
2018-02-19 13:47:07,532 277      DEBUG    https://localhost:443 "POST /dids/tests/test.24659.automa
2018-02-19 13:47:07,533 277      INFO     Thread [1/1] : Upload operation for tests:test.24659.auto
2018-02-19 13:47:07,534 277      INFO     Thread [1/1] : Run with once mode. Exiting
2018-02-19 13:47:07,541 277      INFO     Thread [1/1] : Graceful stop requested
2018-02-19 13:47:07,541 277      INFO     Thread [1/1] : Graceful stop done
```

The daemon has created and uploaded a new dataset in the tests scope. One can list all the DIDs in this scope:

```
[root@3a6d4527e1f6 rucio]# rucio list-dids tests:*
+-------------------------------------------------------+--------------+
| SCOPE:NAME                                            | [DID TYPE]   |
|-------------------------------------------------------+--------------|
| tests:AOD.a9753781316c4b2f8bd88c60e9dd3570            | FILE         |
| tests:AOD.fc50eb5e2b194991980f8218bf62108             | FILE         |
| tests:test.24659.automatix_stream.recon.AOD.917       | DATASET      |
+-------------------------------------------------------+--------------+
```

And one can list the content of the dataset:

```
[root@3a6d4527e1f6 rucio]# rucio list-files tests:test.24659.automatix_stream.recon.AOD.917
+-------------------------------------------+--------------------------------------+----------+
| SCOPE:NAME                                | GUID                                 | ADLER32
|-------------------------------------------+--------------------------------------+----------+
| tests:AOD.a9753781316c4b2f8bd88c60e9dd3570 | 32C89A5A-F0BD-43F6-A958-099C46954C7F | ad:480900d
| tests:AOD.fc50eb5e2b194991980f8218bf62108 | 1937B5B8-BFE3-4AE0-B5CE-28AFE964F5F8 | ad:32bf834
+-------------------------------------------+--------------------------------------+----------+
Total files : 2
```

Next steps

- Continue to support already invested communities
  - Set up collaboration plans for Rucio feature developments which are not directly needed by ATLAS
  - e.g., DIRAC and HTCondor integration, flexible metadata support, EduGAIN authorisation, GlobusOnline, component-wise deployments, …
- Engage more closely with the newly interested communities
  - Deploy, test, and monitor; Integrate existing systems; Register existing data into Rucio
  - News from SKA European Regional Center: will set up Rucio testbed (likely RAL)
- Start going into details for the upcoming "infrastructure sharing" requirements
  - Data Lakes, Networks, Cloud, Capability Authentication
- Think about Rucio Community Workshop #2!
  - Eg., a coding camp to follow up on technical discussions

More information

Website	http://rucio.cern.ch
Documentation	https://rucio.readthedocs.io
Repository	https://github.com/rucio/
Continuous Integration	https://travis-ci.org/rucio/
Images	https://hub.docker.com/r/rucio/
Online support	https://rucio.slack.com/
Developer contact	rucio-dev@cern.ch

Backup

Workshop dinner

Thanks from the Rucio team!


Cédric


Joaquín


Mario


Martin


Nicolò


Stefan


Thomas


Tobias


Tomáš


Vincent