Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
terabit

Adopting RUCIO for the Italian datalake

National experience and perspectives

A. Troja, M. Biasotto, D. Ciangottini, M. Delli Veneri, F. Fanzago, A. Italiano, N. Marcelli, L. Morganti, A. Rendina, M. Sgaravatto, D. Spiga, B. Spisso, S. Stalio, M. Verlato

30/09/2024

7° Rucio Community Workshop

INFN Computing infrastructure

The National Institute for Nuclear Physics (INFN) is the Italian research agency dedicated to the study of the fundamental physics.

Distributed heterogeneous computing infrastructure: resources are both integrated in WLCG and exposed through Cloud interfaces.

Resources are about:

- 200k CPU cores;
- 250PB disk storage;
- 200PB tape storage;
- More than 100 international scientific collaborations served.

The TeRABIT project

The Terabit network for Research and Academic Big data in Italy (TeRABIT) envisions the creation of a a distributed, hyper-connected, hybrid HPC-Cloud environment that offers services designed to meet the evolving needs of research and innovation.

The environment will federate and strengthen the three existing research infrastructures GARR-T, PRACE-Italy and HPC-BD-AI (HPC-Big Data-Artificial Intelligence), leveraging their existing of connections to other national and European research infrastructures and data spaces through the GÉANT backbone.

Main objectives:

- Enable widespread data transfer, up to Terabits per second, and services on a national scale in Italy, connected to Europe;
- Innovate the central HPC node of PRACE-Italy;
- Innovate the HPC services offered to researchers, beyond the centralized calculation model, adding distributed "HPC-Bubbles"

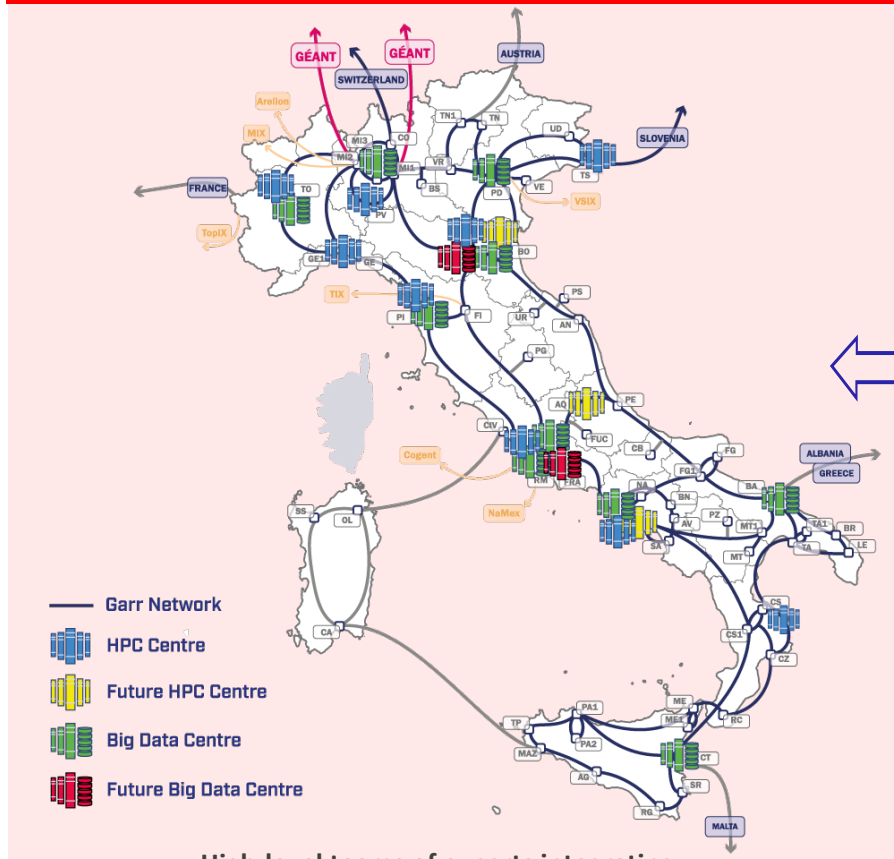
ICSC

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

The Italian Research Center in High Performance Computing, Big Data and Quantum computing conducts R&D at a national and international level.

- 25 universities
- 12 Research institutes
- 14 Strategic private companies

0 SUPERCOMPUTING CLOUD INFRASTRUCTURE



High-level teams of experts integrating the Spokes working groups (mixed cross-sectional teams)

EDUCATION & TRAINING, ENTREPRENEURSHIP, KNOWLEDGE TRANSFER, POLICY, OUTREACH

1 FUTURE HPC & BIG DATA

2 FUNDAMENTAL RESEARCH & SPACE ECONOMY

3 ASTROPHYSICS & COSMOS OBSERVATIONS

4 EARTH & CLIMATE

5 ENVIRONMENT & NATURAL DISASTERS

6 MULTISCALE MODELING & ENGINEERING APPLICATIONS

7 MATERIALS & MOLECULAR SCIENCES

8 IN-SILICO MEDICINE & OMICS DATA

9 DIGITAL SOCIETY & SMART CITIES

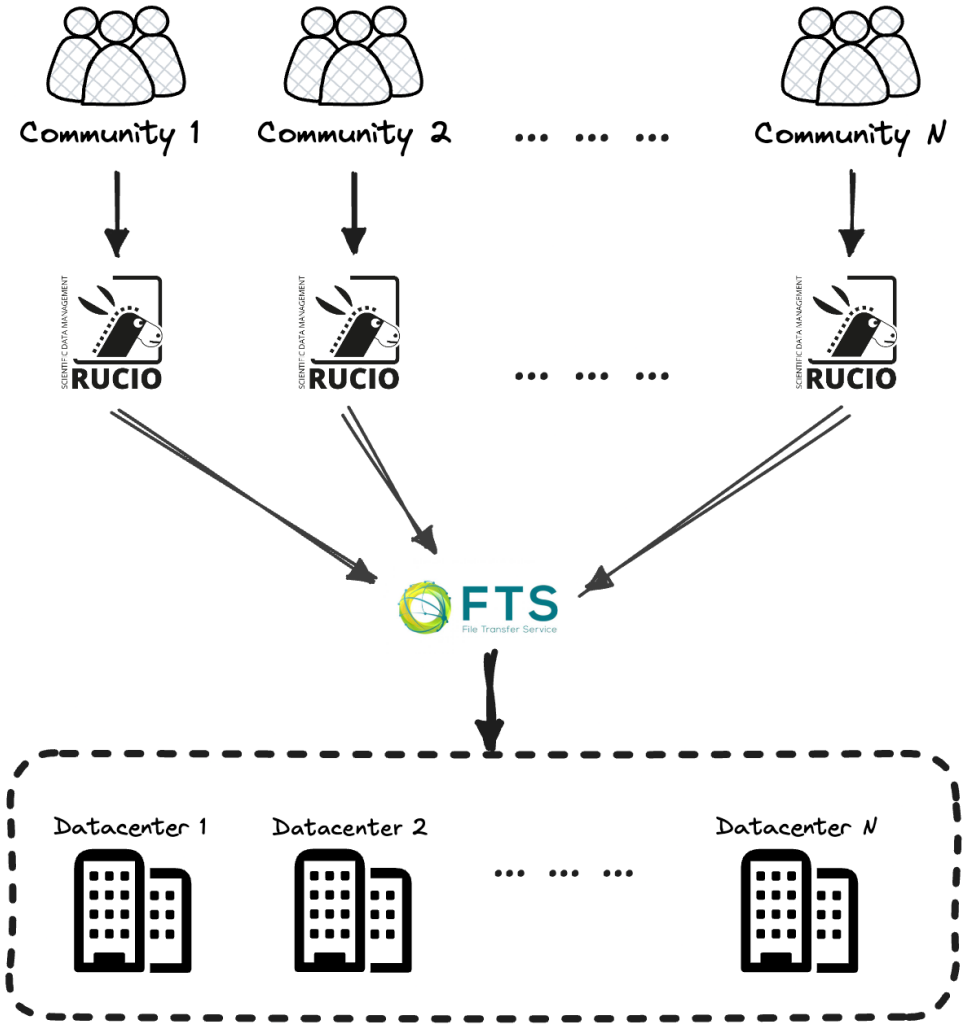
10 QUANTUM COMPUTING

Data-centric model and the National datalake

Each INFN computing centre hosts a storage system, peculiar to that site, including both Grid and Cloud backend. The **need to federate these heterogeneous storages** brings to the creation of a National datalake where users are allowed to access and manage data from any federated node. The underlying infrastructure must be transparent to them.

Each community would interact with its Rucio instance, which in turn relies on a single FTS server centrally managed. TPC are performed among the federated datacenter.

Federation will include INFN facilities as well as external ones, just like CINECA Prace-Italy HPC.



Our experience with Rucio deployment

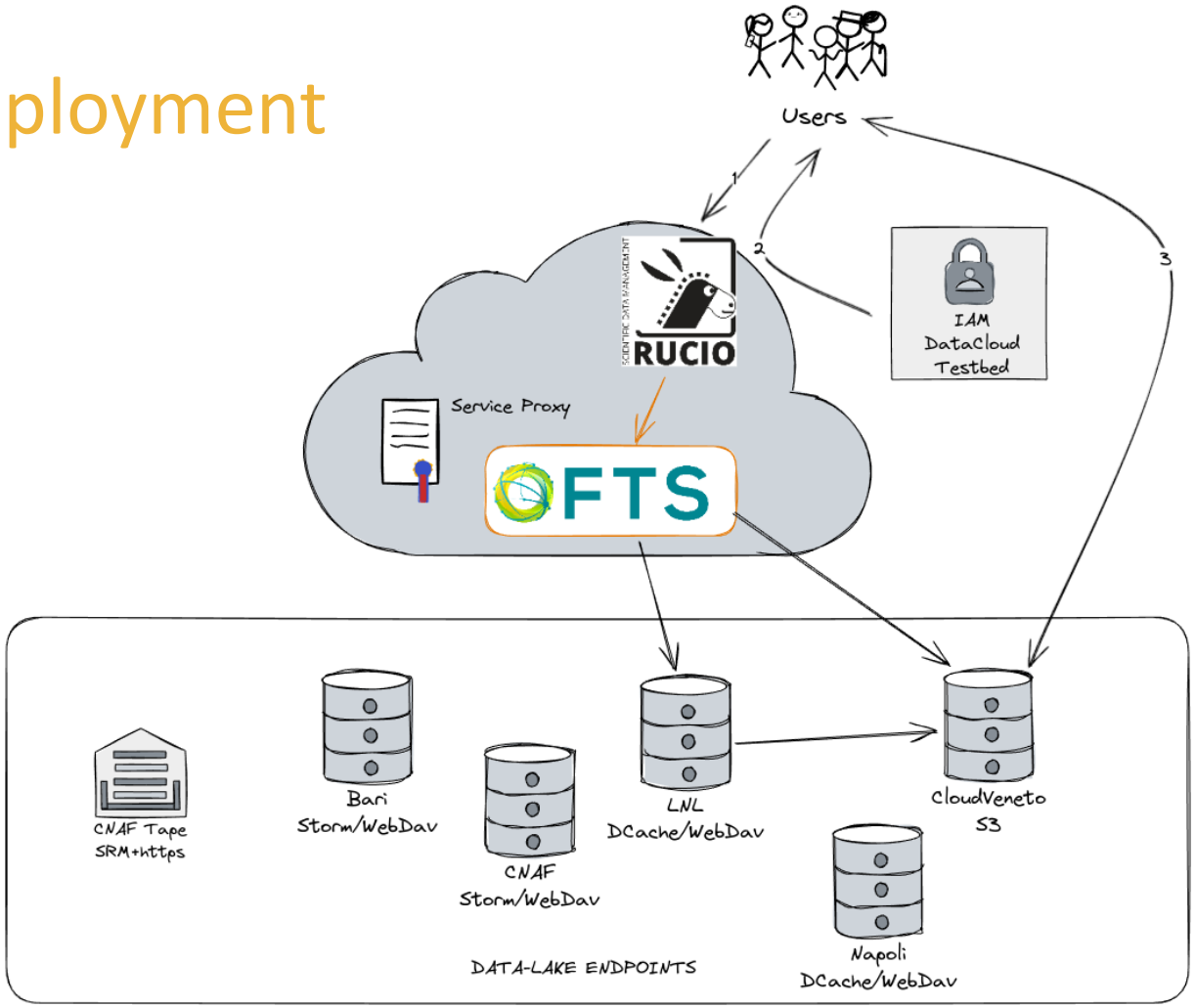
In order to test the federation of geographically distributed, etherogeneous INFN storages, we deployed a **data management testbed**, on which we tested and validated several configurations.

We deployed Rucio on a K8s cluster, hosted in INFN Cloud machines. Same goes for the FTS server. Users interact with Rucio via JWT token, while x.509 VOMS proxies are used by FTS for the TPCs

INDIGO IAM is used to AuthN/Z, and the Rucio database is hosted by a high-availability PostgreSQL infrastructure.

RSEs abstract six INFN storage systems, with different QoS (disk and tape), implementation and protocols.

We use Flux to manage the infrastructure.



Our experience with scientific communities

We're already supporting scientific communities that decided to adopt Rucio (and also act as beta testers):

- Cygno is an experiment on dark matter detectors. Data produced are stored in the INFN Cloud Backbone, analyzed and then moved to tape storage (and deleted from disks when processed). We completed the implementation of the the Rucio infrastructure early this year;
- DarkSide: research on direct evidence of dark matter from 2013 to 2020. They are developing the Rucio infrastructure by themselves, following the documentation we wrote;
- DAMPE: research on indirect detection of dark matter signature. We're setting up an infrastructure that allows interaction between Tier-1 storage at INFN and data center located in China;
- Euclid Italy: The INFN contribution in the Euclid space mission is remarkable, but it lacks of a way to manage data. The process of developing a Rucio instance started with a kick-off meeting early September.

AuthZ model

We tested several authorization models on the data, in order to fulfill the use cases requirements:

- **Group-based policies**: group A users won't be allowed to access data owned by group B, while it can perform any action on group A data (even deletion);
- **Scope-based policies**: scopes are defined within the IAM, and passed to Rucio through the token. Scopes give user permission to certain action on certain area in the storages. As an example, we implemented *user isolation* (e.g., each user can read everything but only write in his/her own designated area);
- **RSE-based policies**: users can perform actions only on a specific type RSE and not on others, for example only privileged users can trigger TPC from/to a TAPE RSE. This is useful to avoid misuses of the tape library, which would impact also other user communities.

User registration

Assigning scim:read scope to the RUCIO operator (as required for the integration with IAM) is not compliant with INFN policies unless this operator has been nominated to manage personal data (and this is not always the case)

We therefore **developed a user registration layer**, put in front of the data management infrastructure.

A user that wants to create an account in Rucio, must authenticate through OAuth2 on the WebUI developed specifically for the user registration. Upon accepting the terms, a Rucio account is created, along with:

- User IAM identity attached;
- Quota assigned;
- User-owned scope created.

We think this feature can be **useful for the general Rucio audience**. We're available to talk about the details.

S3 integration

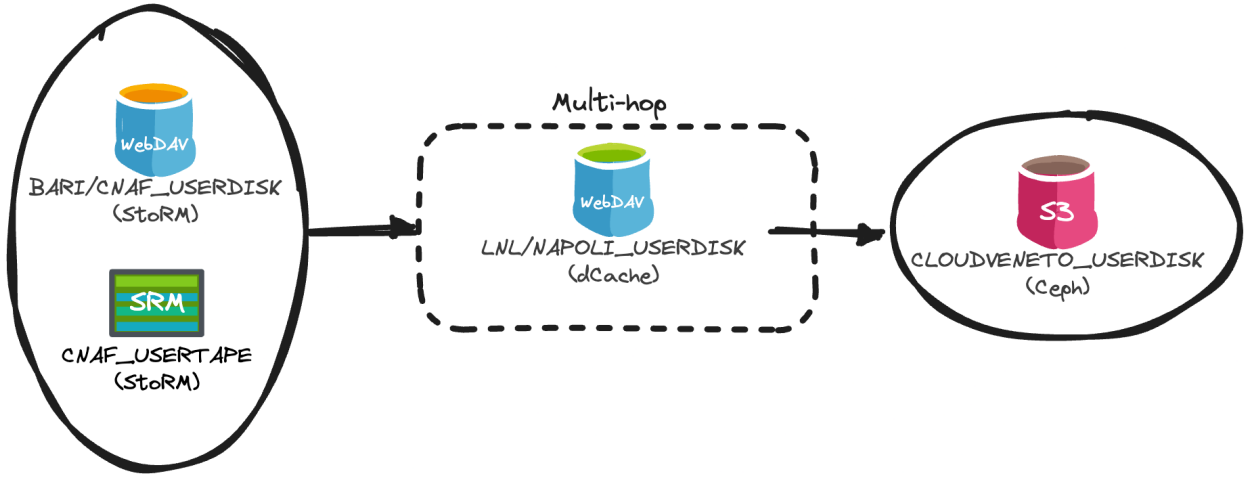
Besides "Grid" storage systems we need to integrate also some S3 endpoints (ceph-rgw and minio)

S3 endpoints must be **manually registered in both RUCIO and FTS** (registering also the S3 credentials that allow read/write access to the relevant buckets)

- As far as we understand this is needed even if the S3 endpoint is configured to authenticate with tokens
- Exposing these credentials to the RUCIO and FTS instances is not very "clean"
- Is this (and more in general the S3 integration model) going to change/evolve ?

To implement TPCs between S3 endpoints **we rely on multi-hop**:

Implemented by deploying an RSE supporting both push and pull mode (dCache webdav in our deployments)
The same approach is used also for transfers from STORM based RSEs to S3 RSEs (since STORM doesn't support the push mode)



Issues at deployment

Thanks to the chat in Mattermost and the help of the developers, we were able to face the issues that arose during development. These include:

- Rucio-server container doesn't include mail templates;
- Reaper daemon doesn't work unless we import "threading" and "concurrent.futures.thread";
- Reaper daemon Helm chart overwrites the certificates in "/etc/grid-security/certificates" with the ones in secret "rucio-ca-bundle-reaper";
- Hermes daemon doesn't allow the use of smtp port and authentication.

The solution we implemented was to create new containers starting from official ones, with all the necessary patches. (of course, it is a sub-optimal solution).

We may push these code developments upstream if there is some interest (and we don't want to diverge from the official code base).

Could make sense for us to create a PR? Will you review these items?

External Metadata catalog

One of the requirements we got from our early adopters is about the support for **the integration of domain specific metadata catalog**:

- by definition are external to the Rucio system.

We might use the RUCIO DB (i.e. extending it) however also following what already done at SKA we understand that using an external DB properly linked (i.e. via rucio.dids) to the RUCIO DB is a better choice.

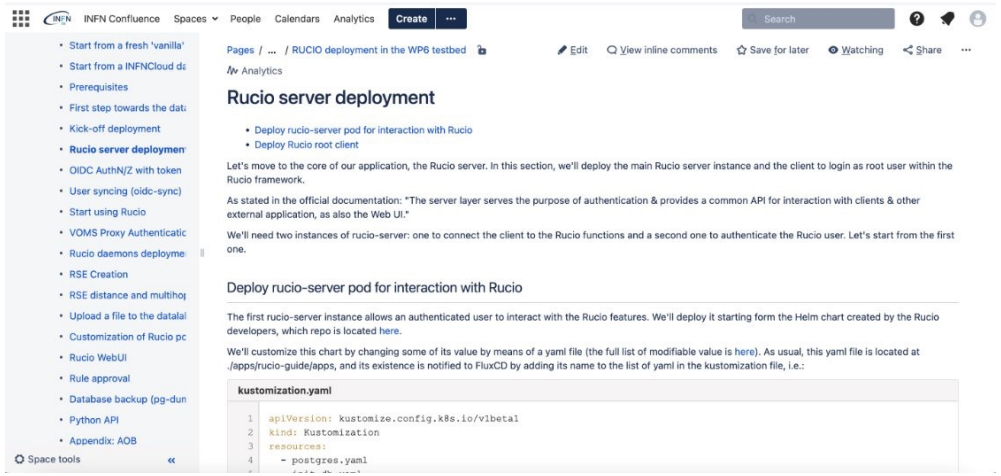
Few questions:

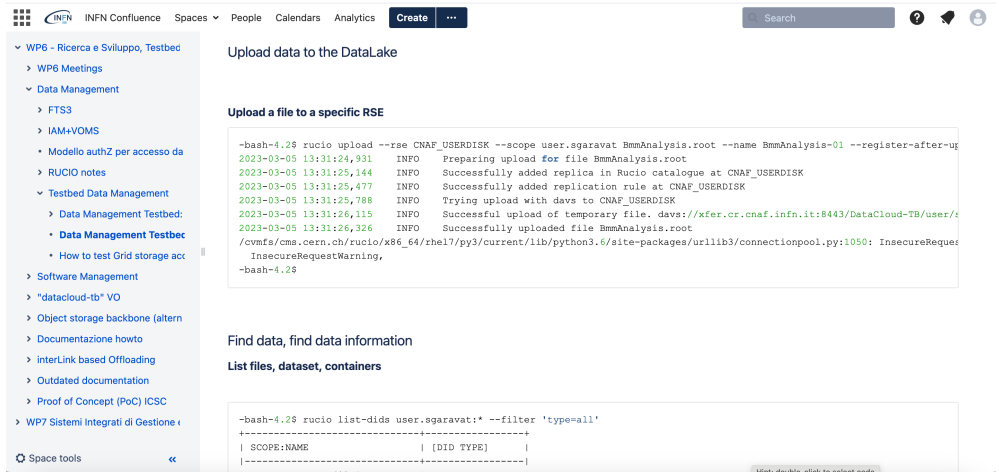
- Is this following the right integration model?
- Is already planned to be embedded in Rucio?
- Any forum where discussions related to this topic happen?

Documentation

We have produced documentation for both Rucio users and operators on INFN Cloud, especially tackling the gaps in the official documentation (for example rule approval and Hermes configuration.)

This documentation **allows any user to replicate our DM system** anytime is required, without major effort.



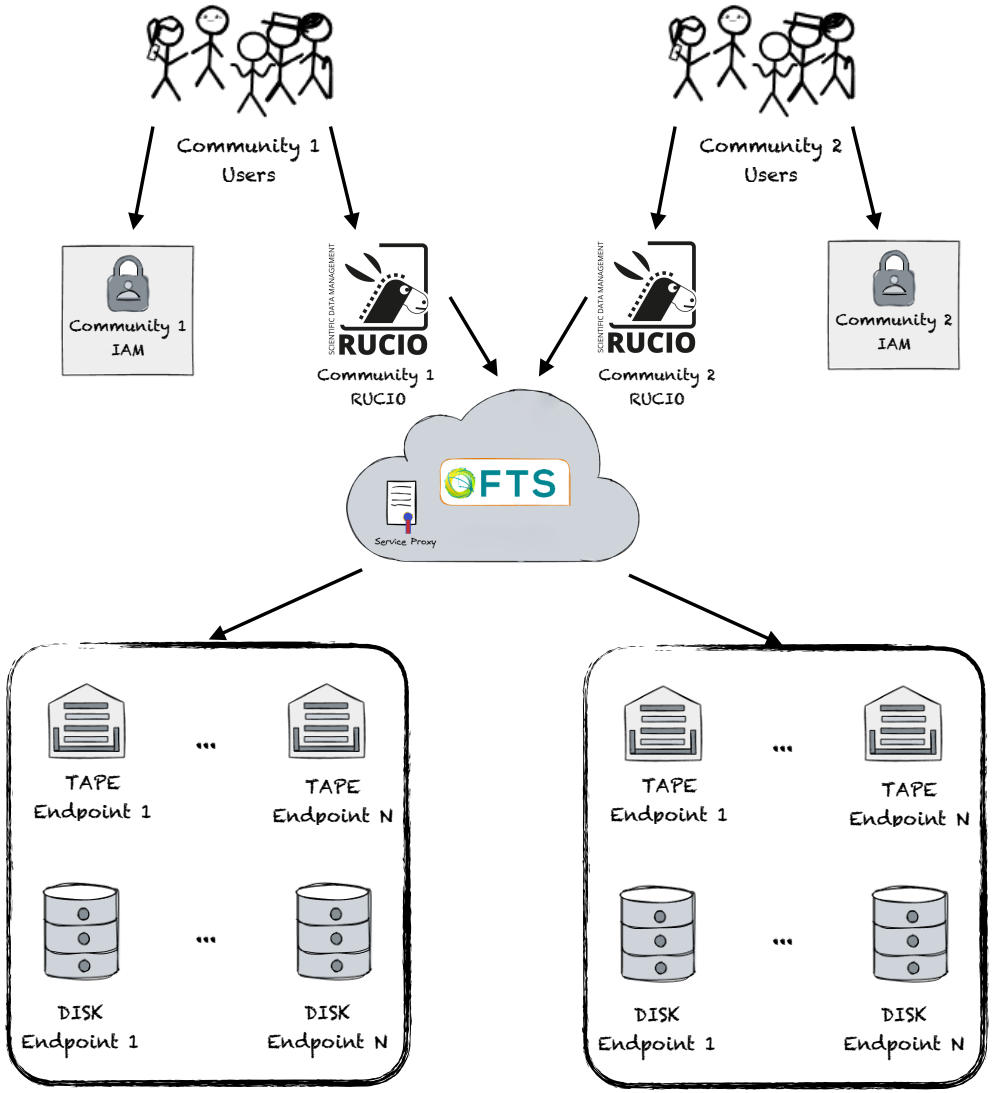


Rucio as a Service for communities

We aim to add Rucio as a a service in the portfolio of INFN Cloud applications, supporting scientific communities with infrastructure and know-how.

We want to **simplify the deployment** of a community RUCIO istance, e.g. Rucio as a service among the services offered by the INFN Cloud dashboard with the following characteristics:

- Rucio community-specific, managed by them;
- INDIGO IAM community-specific, managing users and policies;
- Centrally managed database in high-availability;
- Centrally managed multi-VO FTS;
- Federation of pledged storages.



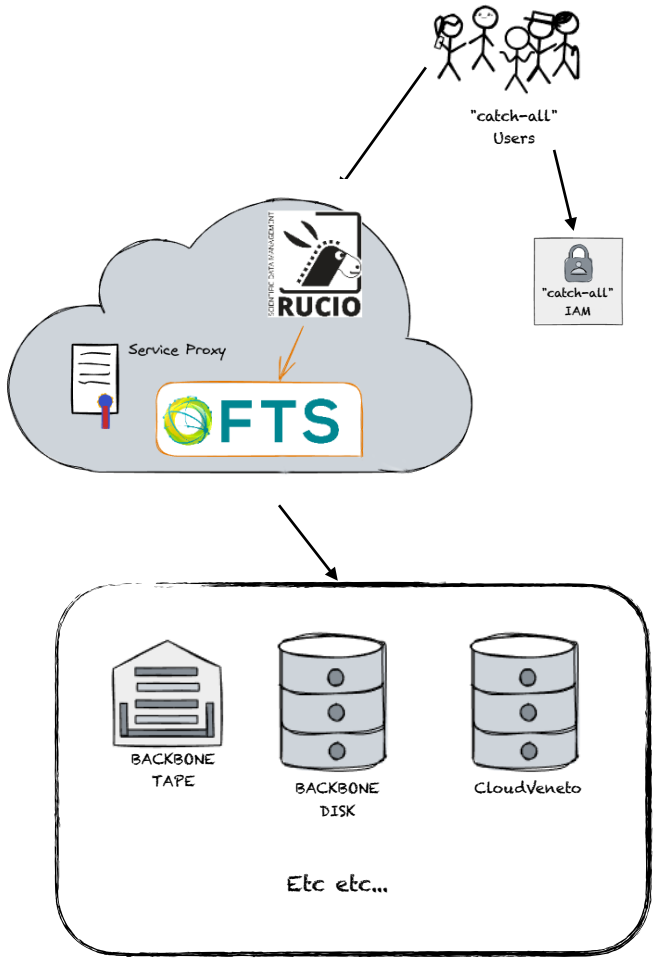
Centrally managed Rucio for single users

We also plan to have a «central» Rucio instance **to serve also single users**, who can take advantage of a federation of storages created specifically for this purpose.

A single user will be able to use Rucio to manage «personal» data, with a single interface in front of several storage systems → **makes it simple.**

- Rucio centrally managed;
- INDIGO IAM centrally managed;
- Centrally managed database in high-availability;
- Centrally managed multi-VO FTS;
- Federation of dedicated INFN storages.

Are there already similar experiences?



Summary and final remarks

- Our national computing infrastructure is further **consolidating and improving**;
- We recognize that "Data Management" has a central role and **we choosed Rucio as a middleware** to abstract the logical layer from the physical one;
- We’re moving towards a model where Rucio is **within reach of the single** user, not only communities;

From the Proof of Concept phase, we want to move to the production phase. Some of the next steps:

- Transfers FTS with token: is it production ready?
- Replace rucio-ui with rucio-webui when the on-going refactoring is done;
- Rucio JupyterLab interface;
- Monitoring: suggestions on tools and technologies that could be used are welcome.

In order to avoid the maintenance of custom Rucio setup, we are **willing to contribute upstream** with possible contributions (i.e. via PR), following the Rucio project's best practices.

Thank you!

antonino.troja@pd.infn.it