Rucio and the ESCAPE community

Forum discussion

Martin Barisits

on behalf of the Rucio team

RUCIO
SCIENTIFIC DATA MANAGEMENT

Token Authentication

- OAuth tokens one of many ways to authenticate Rucio
  - Functionality exists in Rucio since many years, but it is largely a 1:1 replacement to X509 Proxies
- Rucio token evolution is coming
  - C.f. Rucio token discussion in March
- Capabilities tokens will enable Rucio to enforce data embargos
  - New permission model strongly connected to tokens
  - Design is complex and community-specific though
  - **Input from the community will be key!**
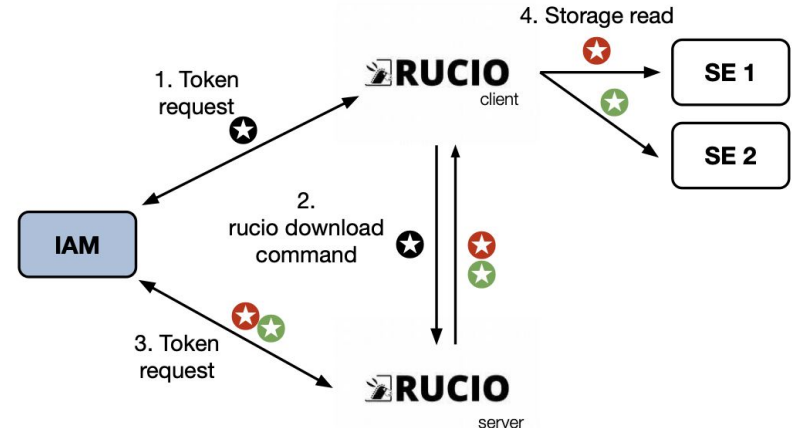
Metadata

- Historically metadata functionality was very limited in Rucio, due to the limited needs in ATLAS
- It is clear that most communities want an integrated metadata solution with their data catalog
  - ESCAPE developments (CERN & SKAO) introduced metadata plugins in Rucio
    - Abstract the metadata queries in Rucio from the actual metadata storage backend (Relational databases, JSON stores, MongoDB, …)
    - Strongly expanded query capabilities with a new filter engine
    - Enable communities to pick the ideal backend technology to store their metadata OR even write plugins to connect to their proprietary metadata stores
- Not the end of the journey
  - **We need to understand what metadata capabilities are missing, but also where we conceptually draw the line of supported metadata functionality**

Deployment and documentation

- Better documentation → New markdown based repository
  - Documentation is slowly improving but we need dedicated effort to make substantial improvements
  - Long-term contributors are focused on improving the software for their community, new contributors are focused on getting the system to run → Nobody does documentation
- Make running and deploying the system simpler
  - Rucio is a very complex system able to manage advanced workflows at large scale
  - Lots of complexity
    - Partly coming from Rucio itself
    - Partly coming from the larger infrastructure: K8s, Databases, Certificates, Storage, …
  - It will probably never be trivial, but we still have a long way to go to make it "easy"
  - Documentation is key, good demonstrators might go a long way as well
  - **What can we do to make the deployment easier?**

Collaboration

- Rucio is a community project
  - By using Rucio you get the benefit of 100k+ accumulated person-hours invested in the project
  - But the project can only be sustained by getting contributions back
    - Wide variety of participation from testing to writing documentation to software development
    - We are somewhat struggling with that
      - Influx of communities put a much larger support strain on experts
- It is (and will be) key to identify commonly needed functionality and find the best ways to achieve these
  - Form small special interest groups to tackle the problem
  - Identify resources (Persons, budget, funding calls, …)
  - Currently preparing the creation of an open steering board
    - Informal; Goal is to understand community priorities, identify common goals, collaboration and funding possibilities
- **What can we do to better utilize and engage the community?**

Thank you all for the inspiring years of collaboration