Rucio Community Workshop 2019

Notes

Session 1

- Keynote 1: Concepts and Architectures for the Next Generation Academic Networks
    - Quite many points to be tracked and discussed over future Rucio dev meetings. E.g. how the application layer can propagate info to the network one
- DUNE
    - Metadata: what is missing?
        - Datasets defined using metadata → They do complicated queries
            - Interested in relations between files → more than simple key/value searches
        - 20-30 metadata items per file
        - At what level is the metadata? Only file?
            - Also on object/observation level?
        - To be discussed also with other experiments like ATLAS. Flexible metadata recording and querying might be interesting
    - PostgreSQL
        - FNAL fully focusing on postgres
            - No scalability issues
        - Backend for Rucio?
        - Everyone interesting in PostgreSQL!
    - Staging area
        - Usage of tapes a challenge for the feature
        - Should work together on this!

Session 2

- ATLAS
    - No questions
    - Open points: what can we do together? Monitoring, accounting, lifetime of files, operational intelligence (automation of operations)
        - not only a common operational effort, but a common operational strategy e.g., "we would like to have this" —> "you can do it like that"
- CMS :
    - Dry-run of rules would be a beneficial!
    - Validation of migration to TAPE :
        - Concerns that the source replica is not deleted before being migrated to TAPE
    - DDM candidates other than Rucio ?
        - Home grown Dynamo
    - Requirement on ACLs ?
        - No requirement

    - Integration DDM/WFSM
- BELLE2 :
    - Many technical questions that will be addressed offline or during the Friday afternoon session
    - Naming scheme :
        - Propose naming scheme is the same as CMS except the fact that the filename are unique for CMS
    - ATLAS has expertise in migration from LFC to Rucio. Can provide expertise and help
    - Shutdown like LS2 ?
        - No
        - Migrate in-between during short stops?
    - Postgres (10) doesn't allow partitioning
        - (well, yes, but only range and list partitioning, not dynamic enough)
        - 11 will
- FTS :
    - Possibility to have FTS instances talking to each other?
        - One FTS instance for all of CERN?

Session 3

- XenoN1T/XenoNnT
    - Will you keep XenoN1T installation?
    - Recommendation for new Rucio users?
        - Use the Rucio API as often as possible!
        - Documentation improved a lot!
    - Python 3.6?
        - Right now Rucio not tested against 3.6, only 3.5 - but very similar
        - Gfal on Python 3 not tested yet, but is available
- IceCube
    - Why adler32?
        -
    - Archive Containers
    - Better documentation!
        - :-)
        - Should be improved!
        - Communities can help a lot in improving the documentation!
        - 150 pages of "installing and running rucio" journal from LIGO (cannot share publicly)
        - Better demo needed
- CTA (Telescope, not tape ;-)

    - Policy: Protect data for at least one year?
        - ivoa compliance?
        - http://www.ivoa.net/documents/PR/ResMetadata/RM-20040323.html#rm
- Rucio Database @ ATLAS
    - Lots of expertise need to run Rucio DB?
        - Possible to run Postgres without all this DBA expertise?
    - Popularity of different partitions?
        - Some extremely popular, some are not
    - Cache fusion
    - PostgreSQL
        - Big database run in postgres (dCache, big web projects etc)
            - Confidence that big Rucio instances should also be able to run on Postgres (but with tunings…)
- Rucio for an SKA Regional center
    - Generic metadata support is important to have.
    - Alternative method of authentication (WebUI)
    - Lightweight CLI
    - Permissions/ACLs
    - Policies
        - Reading too much from tape etc.

Session 4

- NSLS-2
    - Integration of globus online if FTS not suitable?
        - working GO transfertool implementation?
    - Advice from Rucio community about “data layout”
- XDC
    - Storage events
- CTA (The tape, not the telescope)
- LCLS-II

Discussion Day 1

Topics to discuss:

1. Drop of Python 2.6 support for Rucio Clients
    a. Confirmed
2. Release model
    a. Do we need LTS releases?
        i. request from UK: security LTS release (once per year) would be useful maybe last patch release from previous feature release cycle
        ii. try not to overcomplicate

    b. Any particular observations from current upgrade procedures?
        i. validation of releases beyond "alpha"/"beta" testing
3. Support model
    a. (how-to/tutorial?) contribute to documentation!!
    b. mission of cern is also to work with other communities

Session 5

- Keynote II: The Nordic e-Infrastructure Collaboration (NeIC)
    - new project cycle >2020
    - Collaboration also possible with non-european countries? E.g. Canada?
        - Yes, very much so, if the objectives align
- EGI
    - Multi VO instance?
        - Would be desirable; DIRAC is run in multi-vo way for EGI
    - ONEDATA support → More applicable for FTS?
        - Being looked into via XDC
    - OIDC support
        - Will be addressed in Rucio via XDC
- LSST
    - Feature requests: register existing files, customized folder/file structure, AAI, switch for protocol src-dst mapping, customised transfertool-"loader" (local transfertool that does something a-la atlas LSM but for async)
    - Mapping between source/destination protocol
        - Some pairs are blocked, even if FTS could stream it, or even if e.g.dCache could do http to gsiftp
    - Globusonline needed for NERSC?
        - no, can use gridftp (client-side loadbalancing across NERSC DTNs)
        - problem with GO is hardware-token, needs person every two days pushing a button, cannot be automated
    - will give recommendation to NCSA
        - Dynafed
    -

Session 6

- RAL/IRIS
    - Integration with Dirac. 1 FTE for 2 years
    - Multi-VO support expected Sept. 2019
    - Sharing monitoring (+monitoring interpretation/knowledge)
- ARC
    -

- Eiscat3D
    - data embargoes becoming important!
    - What tools for analysis ?
        - Currently matlab. For Eiscat3D, not possible. Not known yet

Discussion Day 2

Topics to discuss:

1.