Data Injection: LOFAR

Author: Yan Grange

1.First part of the exercises for the ESCAPE-DIOS project: Data Injection

commands used:

a. Building Singularity container from Docker hub container (requires Singularity 3.5+):
   i. singularity build --fakeroot rucio-py3.simg docker://projectescape/rucio-client:py3
b. Preparing environment for the container execution:
   i. mkdir -p ~/.rucio
   ii. voms-proxy-init --valid 12:00 --vomses /home/ygrange/vomses --vomsdir /home/ygrange/vomses/ --voms escape:/escape/lofar
   iii. export RUCIO_CFG_AUTH_TYPE=x509_proxy
   iv. export RUCIO_CFG_ACCOUNT=grange
   v. export RUCIO_CFG_CLIENT_X509_PROXY=/tmp/x509up_u38360
   vi. singularity run -B ${HOME}/.rucio/:/opt/rucio/etc rucio-py3.simg

Following commands are all executed inside the Singularity container:

c. Creating scope:
   i. rucio-admin scope add --account grange --scope LOFAR_ASTRON_GRANGE
d. Upload a file:
   i. rucio upload --scope LOFAR_ASTRON_GRANGE --rse GSI-ROOT L557208_SB091_uv.MS_3f73d3c6.tar
e. Add replication rule:
   i. rucio add-rule LOFAR_ASTRON_GRANGE:L557208_SB091_uv.MS_3f73d3c6.tar 1 SARA-DCACHE
f. Check rule status:
   i. rucio list-rules LOFAR_ASTRON_GRANGE:L557208_SB091_uv.MS_3f73d3c6.tar
g. Download the file:
   i. rucio download LOFAR_ASTRON_GRANGE:L557208_SB091_uv.MS_3f73d3c6.tar

Output of failures:

Singularity> rucio upload --scope LOFAR ASTRON GRANGE --rse CNAF-STORM
L557208_SB091_uv.MS_3f73d3c6.tar

2020-09-29 18:18:43,920   INFO   Preparing upload for file
L557208_SB091_uv.MS_3f73d3c6.tar
2020-09-29 18:18:44,243   INFO   Successfully added replica in Rucio catalogue at
CNAF-STORM
2020-09-29 18:18:44,424   INFO   Successfully added replication rule at CNAF-STORM
2020-09-29 18:18:45,071   INFO   Trying upload with davs to CNAF-STORM
2020-09-29 18:18:45,541   WARNING   Upload attempt failed
2020-09-29 18:18:45,542   INFO   Exception: The requested service is not available at the
moment.
Details: An unknown exception occurred.
Details: Result Authentication error, reached maximum number of attempts after 1 attempts
2020-09-29 18:18:45,667   ERROR   Failed to upload file
L557208_SB091_uv.MS_3f73d3c6.tar
2020-09-29 18:18:45,667   ERROR   None of the given files have been uploaded.

Singularity> rucio upload --scope LOFAR_ASTRON_GRANGE --rse LAPP-DPM
L557208_SB091_uv.MS_3f73d3c6.tar

2020-09-29 18:19:06,706   INFO   Preparing upload for file
L557208_SB091_uv.MS_3f73d3c6.tar
2020-09-29 18:19:06,935   INFO   File DID already exists
2020-09-29 18:19:07,169   INFO   Successfully added replica in Rucio catalogue at LAPP-
DPM
2020-09-29 18:19:07,570   ERROR   The requested service is not available at the
moment.
Details: An unknown exception occurred.
Details: Failed to stat file (Permission denied)

Singularity> rucio upload --scope lofar_astron_grange --rse CNAF-STORM
L557208_SB091_uv.MS_3f73d3c6.tar

2020-09-29 17:09:23,865   INFO   Preparing upload for file LOFAR_dataset
2020-09-29 17:09:24,025   ERROR   Scope does not exist.
Details: no scopes found for account ID 'grange'

Feedback:
One issue I notice is that sometimes error messages are not really clear to me. But in this
specific case that was not really the case.

As a future exercise I would like to investigate the hierarchical relation of files to datasets
and containers. In our case one observation consists of many files. For imaging, one can
combine a target and a calibrator, and multiple calibrated images can then be combined to
a further data product. Provenance and data hierarchy of those can nicely been done in
datasets and containers.

Also I would like to experiment with non-deterministic file systems for our data.
Also an exercise of listing all files in a scope/container/dataset using the CLI would be useful.

A more advanced use case, that I had at hand when importing existing LOFAR data into the data lake, is to import data from an external gridftp location, which is fairly straightforward using the python bindings of the Rucio client.