

DAC21 Debriefing Session - Q&A

Agenda: <https://indico.in2p3.fr/event/25767/>

Present: Gareth Hughes, Agustin Bruzzese, Alba Vendrell, Andrea Ceccanti, Diego Ciangottini, Elena Gazzarrini, Fabio Hernandez, Federica Agostini, Frederic Gillardo, Ghita Rahal, James Collinson, Jutta Schnabel, Maisam M. Dadkan, Marcelo Soares, Marek Szuba, Matthias Fuessling, Mieke Bouwhuis, Paul Millar, Riccardo di Maria, Rizart Dona, Ron Barnsley, Ron Trompert, Rosie Bolton, Xavier Espinal (*apologies if I missed some of you, please add your name*)

Apologies: Pandey, Yan Grange

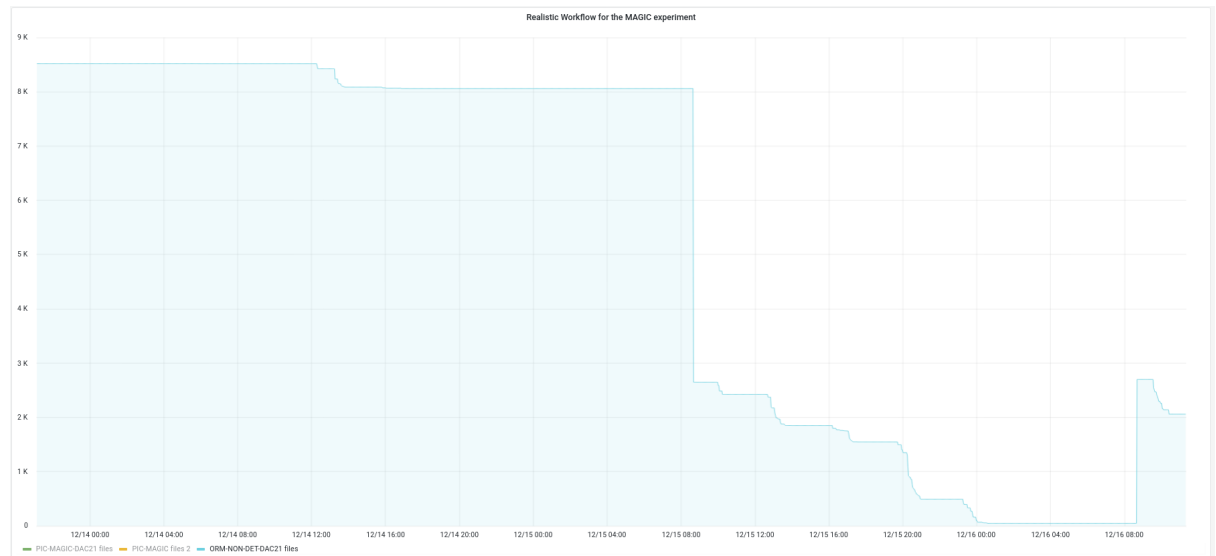
SKAO

- XE: Related to Rucio's conveyor daemon issues, would it be a good idea to setup alarms/healthchecks for all Rucio daemons in the several instances we run in the DL?
 - RMB A: We run something like this on our instance where we have a probe to periodically poll the time since the last daemon log entry and display it on a dashboard, this is how we knew it had fallen over. This could be implemented as a pod liveness probe also.
I think the recommended (by rucio dev) "fix" for this issue is to restart the daemons every day.
- XE: Based on the Long haul x-fers metrics, this is based on strict 10MB files, seems test data? Any plans to have simulated or realistic data in the next year?
 - RMB A: This is test data, yes, ranged between 50-5,000MB. We plan to probe this parameter space more in the coming year but not with simulated data.
- RdM: May you please expand (offline) on how DLaaS was not working? There was an issue since users filled the local storage instead of using SCRATCH...
 - JSC A: The 'upload' functionality (from UI) was not working during the DAC21 week - I think you or Rizart advised of this. Upload using the CLI worked fine, so it wasn't a blocker, but I just thought it was worth mentioning. Have a look at the linked demo (<https://drive.google.com/file/d/1nDDtbXipQeZBiDtN7GVp71NlfVqxYhga/view?usp=sharing>) starting around 7:00 to see what I mean.
 - Riccardo: small files should be uploaded using the web-friendly "Rucio Upload". Big files should be uploaded using "datalake upload" functionality available using the terminal. During DAC21, there was an issue since a user filled up the shared local storage space used by Jupyter. But that issue caused environments not even being loaded for users. Maybe this was just a misunderstanding? I believe other experiments successfully used both upload DLaaS functionalities. TO CHECK

MAGIC

- XE: Due to the situation in La Palma, this exercise ran totally unattended on the next to the telescope facility? How was this experience?
 JD: The setup in La Palma was precisely the one used for the telescope facility for regular transfers since we have dedicated machines for that purpose, the RSE is the same used for regular transfers. The current situation allowed us to at full capacity without any clash with other transferences
 AB: During 5 days, we ran through cronjobs a script that allowed the automatic transfer from the Palm to the PIC. During these days, there was moderate supervision on how data was continuously transferred. In general, there were no major problems. Apart from those mentioned above. Even so, we could say that our workflow ran unsupervised and successfully for 5 days (we are still running it and we have been running it for 10 days with no deletion or transfer problems)
- XE: Based on the resulted tests, plans to bring this setup into full production (not user if this is the case already...)?
 AB: Currently, the only drawback we have is the smooth migration from the old MAGIC transfer system to the rucio transfer system. For this, we have to keep feeding the database that allows the magic experimentalists to know the statuses for the MAGIC files. In short, once we apply this last step (feeding the old MAGIC database using rucio) we will be able to move the present workflow into production.
- XE: Would you recommend your current setup to other experiments with similar (remote facilities) needs?
 AB: In our experience, k8s is a great tool. Many applications, such as rucio, are greatly helped by it. So this architecture is recommended. However, we found it easier to connect rucio to some of the services already deployed on the PIC outside k8s, such as prometheus, elasticsearch and grafana. After this full design of the workflow is completed with the inclusion of the MAGIC data transfer database updates, we will define a full migration plan which also includes to make more robust the setup and deployment of RUCIO at PIC and also the Network configurations
- The source (observatory) RSE seems to show a number of files that is increasing over time. There are peaks when data is ingested and those peaks then go down again as Rucio replicas the data and the source files are deleted; however, the base-level seems to increase over time. Is this understood?
 AB: Thank you for this question. It's interesting, because we have found that the RUCIO base configuration was less optimal for transferring numerous files. As a consequence, the conveyor submitter daemon was slow to update the status of the files, and consequently the deletion of the files on the source RSE. To solve this problem, we changed the conveyor submitter configuration to transfer more than one file per fts's job using the --group-bulk flag. From then on, you can see a big change in the speed of file deletion
 [1] <https://github.com/rucio/rucio/blob/master/bin/rucio-conveyor-submitter#L99>

[2]



CTAO

- XE: How would you evaluate the effort vs. benefits of having a dedicated Rucio instance? Would you recommend this approach for experiments with a similar size, data volumes and distributed computing needs?
- XE: Interesting point: DIRAC capabilities regarding file catalogue management. This is something to follow-up. Could you enumerate some of the missing items in Rucio that you would inherit from DIRAC?

AB: at the rucio.cfg we are missing the following configuration:

```
[api]
```

```
endpoints =
```

```
accountlimits,accounts,auth,config,credentials,dids,dirac,export,heartbeats,id
entities,import,lifetime_exceptions,locks,meta,ping,redirect,replicas,requests,r
ses,rules,scopes,subscriptions
```

AB: at the apache configuration, we've observed that redirection of dirac library was missing.

```
WSGIScriptAlias /dirac
```

```
/usr/local/lib/python3.6/site-packages/rucio/web/rest/dirac.py process-group=rucio
application-group=rucio
```

- XE: CTA is a clear nexus between Data Lake, OSSR and DIRAC. Something to pursue in 2022 in view of further integration. The planned exercise for DAC21 is still relevant or you have something else in mind to tackle this in the next months?

GH:

The data lake DIRAC connection will be worked on in Jan/Feb 2022. There is the beginning of a plan with some initial questions regarding missing methods and data caching.

I have been working on launching DIRAC jobs from the ESAP, the main issue being authenticating. If this step is done by hand then I think everything else is relatively

LSST

- DID created but not the replica. May you please expand on this? By definition in Rucio, ALL DIDs are replicas, even if present in 1 copy. Are you referring to replication rule?
- XE: Was the reaper keeping up with 50k deletions per hour? What would be a limit?
 - Riccardo: we fine tuned for the requested rate. The actual limit should be identified yet. Following the fine tuning, the reaper kept up with 50k deletion/h.
- XE: Integrating US/SLAC in the pipeline?
 - FH: Already happening (!) uploading data now to SLAC and planning some tests in the next days.

KM3Net

- XE: Do you have a real need to update files (reuse same file names)? Would it be file versioning a solution? Usually storage systems do not love file updates.
 - Andrea: both tokens and x509 will need renovation procedures
 - Marek: Refresh tokens are normally fairly long-lived, though.
 - Andrea: yep, what I meant is that you do not just need to tokens (as an experiment) but also implement support for token renewal (i.e., handling the token refresh when needed)
- XE: X509 proxy renewal **needs follow-up**, is a standard procedure but needs to be done correctly. Point taken. Missing procedure and documentation?
- XE: Metadata browsing/filtering needs is a common demand, yes, need to follow up this.
 - RMB: This is ongoing. Metadata browsing using hardcoded keys has already been merged since 1.27 (<https://github.com/rucio/rucio/pull/4746>). The same functionality for custom metadata (json key/value) is in progress (<https://github.com/rucio/rucio/pull/5104>). I am also working on generic interfaces for external databases, e.g. mongo
- Paul: As a general comment: the IRODS-Rucio integration discussion might be of interest by others (at least, by me, Paul).

FAIR

CBM and PANDA

- I think checksum can be passed based on the filesystem info? In posterior TPC x-fers FTS verify checksums are honored on both sides. **To be followed up.**
 - <https://github.com/rucio/rucio/issues/5079>
- Rucio Domains usage (WAN vs LAN)

- XE: What about organising a dedicated session for long lived tokens/x509?
 - Andrea: interim solution in ESCAPE IAM in place
- Paul: Would be interesting to discuss further about data ingestion and how to use checksums that CBM / Panda are already calculating (Paul)

R3B

- XE: Which is the usage you would see the DLaaS fitting in the future in your community?
- XE: Token based auth on the overall infrastructure interoperability to be pursued. Point taken. **To be followed up.**
 - AC: Yep, especially since both X.509 and token-based auth worked fine for some use cases (e.g., CMS)

CMS

- XE: Could you expand a bit more the procedure used for the Data Discovery in Rucio?
- XEL What would be the use cases addressed by DASK service? is it targeted for the local user community or could it be seen as a more global analysis-facility “model”?
- XE: A WP2 wide exercise for embargo data might be a good idea, following the use case you exercised in DAC21 and based on the fact many experiments will be facing similar requirements. **To be followed-up.**
- XE: It is a good idea to explore ephemeral RSEs to be used as buffers/caches, for several reasons

ATLAS

Meeting with Task Leads to digest DAC21 debrief

Next WP2 meeting to be announced (~mid january) -> ideas and plans for 2022