

ESCAPE DAC21

LSST data replication exercise results



Lionel Schwarz, Adrien Georget, Fabio Hernandez
(CC-IN2P3)

Reminder: purpose of this exercise

- Use Rucio to drive inter-site replication of one night's worth of LSST data
 - repeatedly over 5 consecutive days
- Realistic data selected for this exercise
 - 4,000 exposures, ~800 k files, ~15 TB
 - equivalent to 1 night of raw data in terms of volume and to 2 nights in terms of number of files
 - average file size: 18 MB (small by HEP standards)
 - time budget: less than 12 hours, equivalent to one night's data taking period

Principle of operation

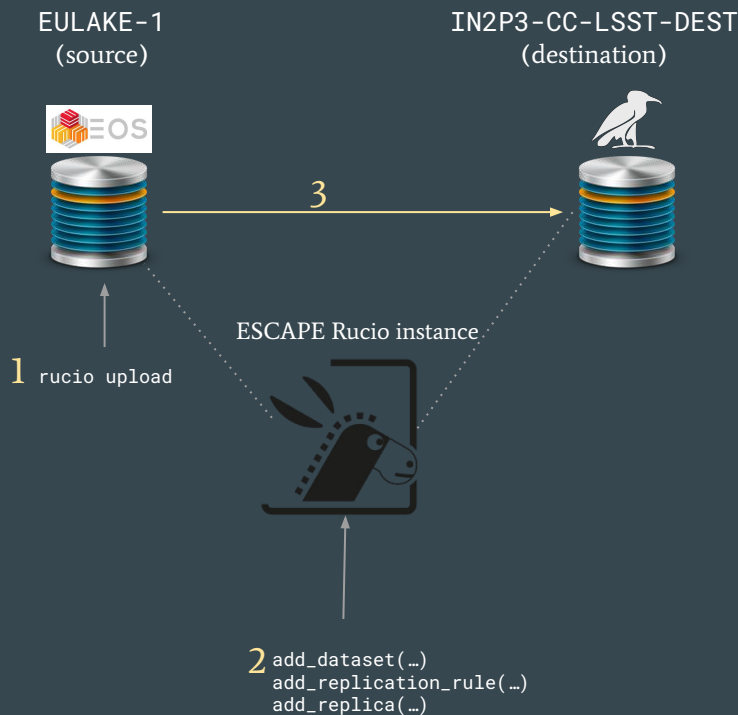
Step 1: prior to the start of the exercise, upload each exposure's files to the source RSE

```
rucio upload --rse EULAKE-1 \  
  --scope LSST_DAC21 --name myLFN \  
  --protocol davs /path/to/file.fits
```

Step 2: daily, for each exposure create a new Rucio dataset and a time-limited replication rule

```
# create a dataset and add a replication  
# rule with a lifetime of 16 hours  
DIDClient.add_dataset(...)  
RuleClient.add_replication_rule(...)  
  
# add each file in this exposure to the dataset  
ReplicaClient.add_replica(...)
```

Step 3: daily, let Rucio perform the replication



Warm up week (15 to 18 Nov.)

- Observed issues with `rucio` upload of 20 files (out of ~800k)
 - DID correctly created but no replica associated with it
- Rucio's `Reaper` configuration modified
 - 5k → 20k → 50k deletions per hour
- Observations at the end of warm up session
 - replication of half the exposure data in 13 hours: ~1 Gbps
 - suspiciously high transfer error rate of about 20%: automatically recovered by FTS via TPC push (instead of pull), not understood by then

Challenge week (22 to 26 Nov.)

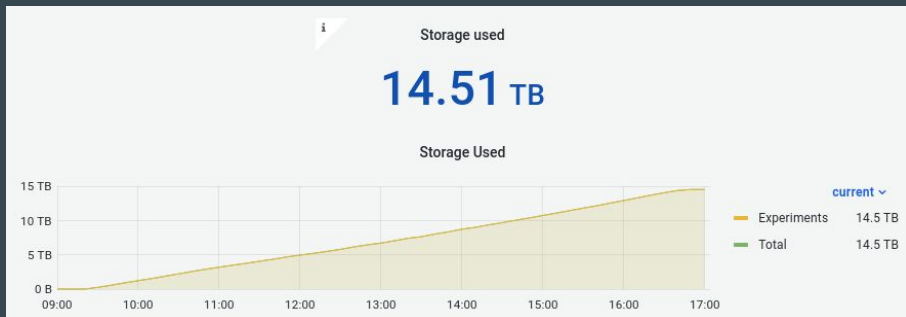
- Decision: perform the exercise by replicating only half of the exposures
- Day 1
 - 14.5 hours to replicate 2,000 exposures, average throughput 1.1 Gbps
 - expecting about 100 simultaneous transfers, observed 15 to 20
 - unexpected response HTTP 404 given by dCache to about 20% of gfal-copy requests
- Day 2
 - applied first dCache hotfix to fix HTTP 404 issue
 - increased throughput to 1.24 Gbps

Challenge week (cont.)

- Day 3:
 - 2 additional dCache hotfixes applied to improve throughput (details [here](#))
 - increased throughput to 4 Gbps
- Day 4:
 - replication of 3,000 exposures (567k files, 10 TB)
 - average throughput 3.6 Gbps
 - observed slight decrease in successful transfers by 14h-15h (FTS issue ?)

Challenge week (cont.)

- Day 5:
 - trigger replication of all 4,000 exposures (756k files, 14.5 TB)
 - average throughput 4Gbps
 - less than 8 hours to replicate without errors 🚀



Summary of results

- We reached our goal: to use Rucio to replicate one night's worth of raw LSST data in less than 12 hours, repeatedly over 5 days
 - we are very happy with the results obtained in this exercise and the lessons learned
- Detailed monitoring tools, both at CERN (Rucio & FTS activity) and local to the site (dCache activity) were instrumental
- To investigate
 - the number of observed simultaneous transfers (15 to 20) were significantly lower than the maximum configured for FTS (100)
 - could this be related to high churn rate induced by the small file sizes? (it takes less than 4s to transfer each file)

Acknowledgements

- Many thanks to the CERN team operating CRIC, Rucio, FTS and monitoring infrastructure for their support and reactivity
 - Riccardo, Rizart, Alba, the DepOps team and other people working behind the scene
- Special thanks to Paul and the dCache development team for diagnosing, debugging and providing several hot fixes in a extremely reactive way all over the week