

Managing scientific data with Rucio

INTRODUCTION:

- 1) Major Atmospheric Gamma IMaging Cherenkov (MAGIC) telescopes:
 - 1) 210240 files
 - 2) 126 Tb data

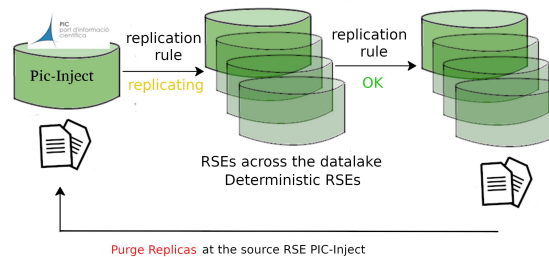
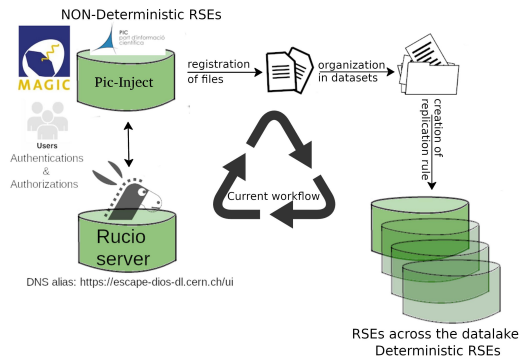
- 2) Data transfers from the Observatory site to PIC:
 - 1) Does not allow for fully automatic replication.
 - 2) Limited amount of space

- 3) Rucio as a global data management system:
 - 1) Provides scientific collaborations with the functionality to organize, manage, and access their data at scale.
 - 2) The data can be distributed across heterogeneous data centres at widely distributed locations

PIC data injector

AIMS :

- 1) The main aim is to develop solutions to handle the large data sets produced by Gamma ray telescopes :
- 2) Files streamed from the telescope to the Data Lake for permanent storage and access.



File creation, discovery, replication and deletion cycle

PROGRESS :

- 1) Data is injected from the storage system located at the source (ORM La Palma in the future) :
 - a) Currently testing a non-deterministic RSE configuration at PIC that allows to register files with their original path in the detector. Mimicking the future config of an RSE at ORM.

Table 1. Organization of the Magic namespace at RUCIO

Files -> LFN 20200114_M1_05088328.001_P_CrabNebula-W0.40+035.raw.gz

DID -> (scope:name) = MAGIC_PIC_BRUZZESE:20200114_M1_05088328.001_P_CrabNebula-W0.40+035.raw.gz

Container 3 -> Data Type = RAW_M1

Container 2 -> Source = CrabNebula

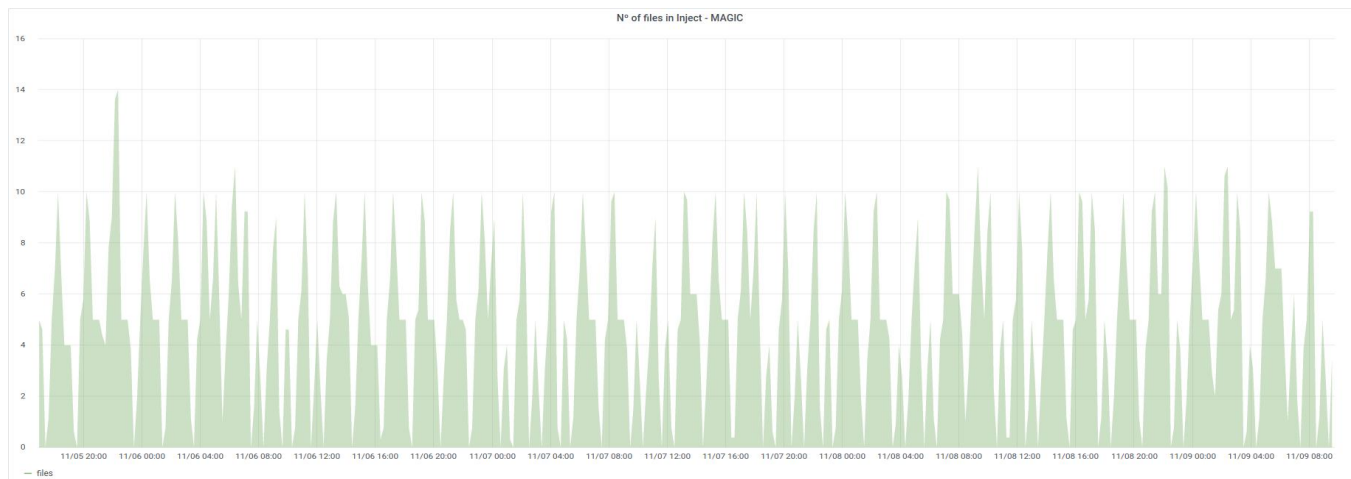
Container 1 -> Date = 20200114

Datasets -> Runs = 05088328

File creation, discovery, replication and deletion cycle

PROGRESS :

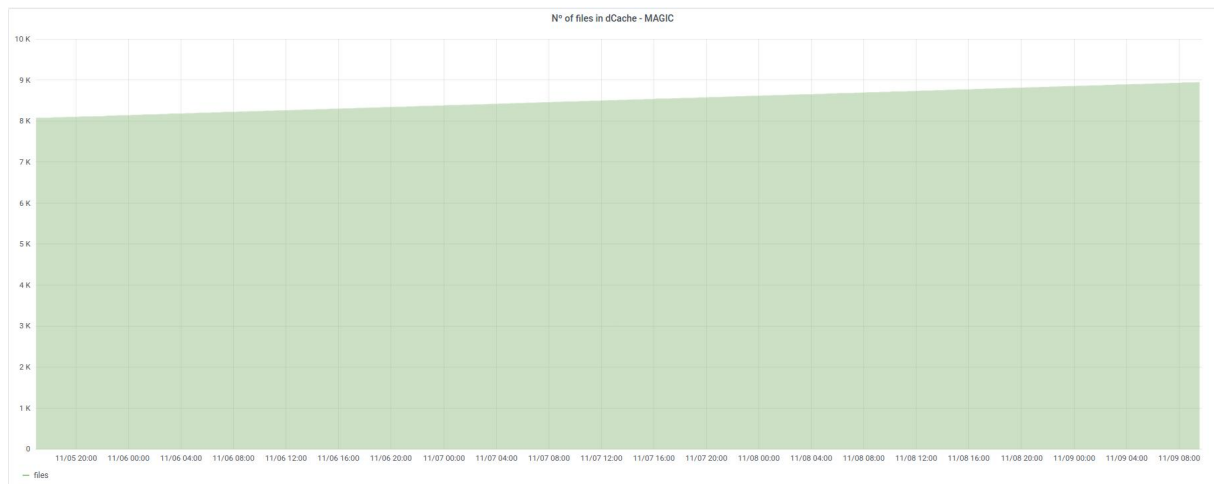
- 1) Data is injected from the storage system located at the source (ORM La Palma in the future) :
 - a) Currently testing a non-deterministic RSE configuration at PIC that allows to register files with their original path in the detector. Mimicking the future config of an RSE at ORM.



Continuous storage for MAGIC experiments

PROGRESS :

- 1) Data is By means of the RUCIO API, we automate the orchestration of the data: using replication rules :
 - a) Files are replicated through different research facilities among the Data Lake.
 - b) Once replicated, source files are removed.



Current Problems

Rucio UI

Monitoring

Data Transfers (R2D2)

Admin

pattern OR name OR rule id

Search

Using account: bruzzese

You are here: Rucio Rule Definition Droid - List Rules

Rucio Version (WebUI / Server): 1.23.11 / 1.23.11

Rules

New request

Account

RSE

State

Activity

Interval

days

Apply

Show

100

entries

Search:

Showing 1 to 7 of 7 entries

Previous

1

Next

Select all

Delete rule(s)

download as JSON

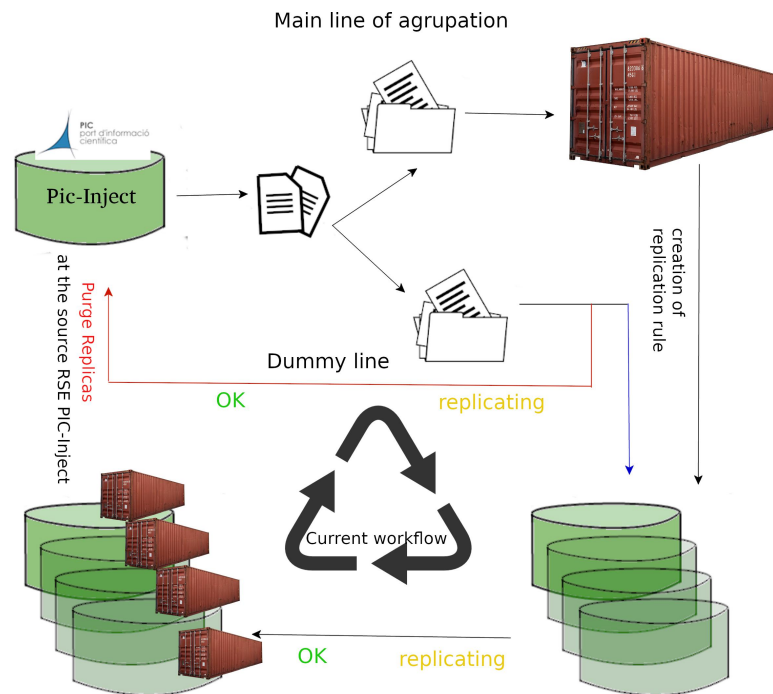
© 2012-2019 European Organisation for Nuclear Research (CERN)

Tue Dec 08 2020 20:56:57 GMT+0100 (hora estándar de Europa central)

Orchestration of rule creation and automatic removal

Conclusions:

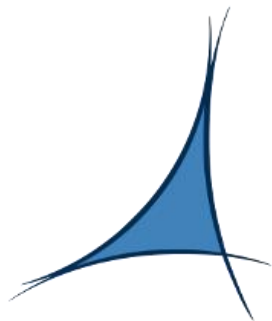
1. Develop an automatic workflow for the distribution of scientific data through the use of RUCIO
2. Setup a RSE in La Palma for the ingest operation
3. We want to start working with real data from MAGIC experiments. To do this, having a private data section in the ESCAPE project would be ideal
4. In addition, we would like to start distributing some data from the LST experiments



Thanks for listening

Questions?

Thank you all for your tremendous support.



PIC
port d'informació
científica

SI Figures

