# ATLAS DQ2 to Rucio renaming infrastructure

**C. Serfon[1], M. Barisits[1,2], T. Beermann[1], V. Garonne[1], L. Goossens[1], M. Lassnig[1], A. Molfetas[1,3], A. Nairz[1], G. Stewart[1], R. Vigne[1] on behalf of the ATLAS Collaboration**

[1] CERN PH-ADP-CO/DDM, 1211 Geneva, Switzerland

[2] University of Innsbruck, Innsbruck, Austria

[3] University of Melbourne, Melbourne, Australia

## 1. Introduction

Rucio [1] is the new evolution of the ATLAS Data Management system that will replace the current one called DQ2 [2]. This new system has many improvements, but it breaks the compatibility with DQ2. One of the biggest changes is the fact that no external replica catalog like the LCG File Catalog (LFC) is being used in Rucio: the physical path of a file can be simply extracted from the Logical File Name via a deterministic function. Therefore, all files replicas produced by ATLAS have to be renamed to follow a new naming convention. It represents around 300M files split between ∼120 sites with 6 different storage technologies. An infrastructure to perform this renaming has been developed and is presented here.

## 2. New naming convention

In DQ2 there is no deterministic way to get a Physical File Name (PFN) from a Logical File Name (LFN), therefore an external catalog (LCG File Catalog) was needed to do the mapping.

The new naming convention (also known as Rucio naming convention) is based on a deterministic function that allows to transform every LFN into a PFN. The path can be directly obtained from the LFN via a deterministic function.

Deterministic function : For `LFN=scope:filename` (scope is a way to partition the namespace), the PFN will be `<prefix>/scope/L1/L2/filename` where L1/L2 are the 2 first bytes of MD5(LFN), e.g. :

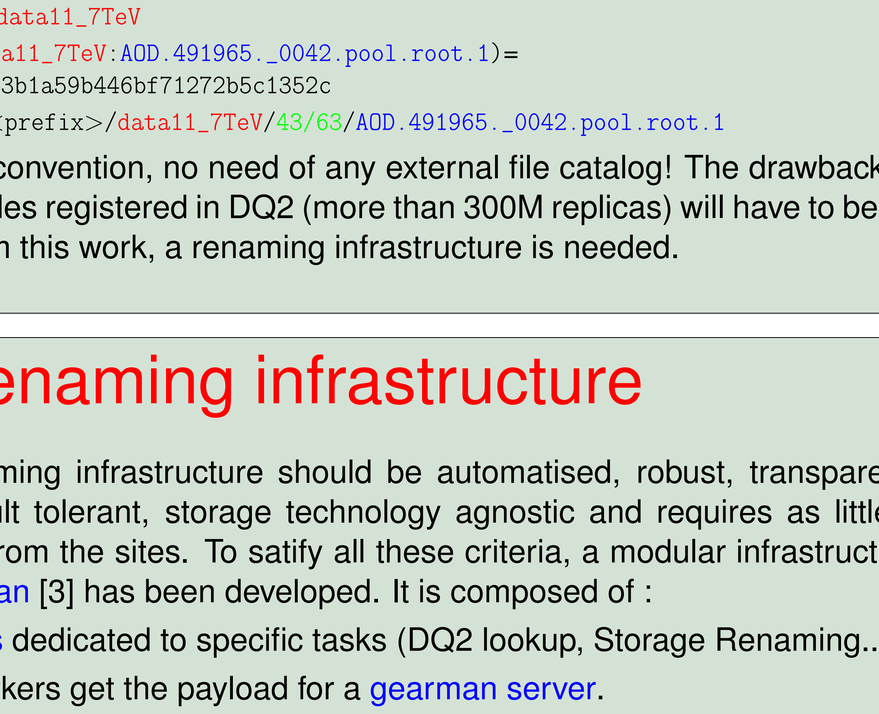
- For file `user.jdoe:004406.EXT0._00011.root` :
  - scope=`user.jdoe`
  - `md5(user.jdoe:004406.EXT0._00011.root)= 35be9fb53d01500d33011414abccde53`
  - PFN : `<prefix>/user/jdoe/35/be/004406.EXT0._00011.root`
- For file `data11_7TeV:AOD.491965._0042.pool.root.1` :
  - scope=`data11_7TeV`
  - `md5(data11_7TeV:AOD.491965._0042.pool.root.1)= 436335c43b1a59b446bf71272b5c1352c`
  - PFN : `<prefix>/data11_7TeV/43/63/AOD.491965._0042.pool.root.1`

With this convention, no need of any external file catalog! The drawback is that all physical files registered in DQ2 (more than 300M replicas) will have to be renamed. To perform this work, a renaming infrastructure is needed.

## 3. Renaming infrastructure

The renaming infrastructure should be automatised, robust, transparent for the users, fault tolerant, storage technology agnostic and requires as little work as possible from the sites. To satify all these criteria, a modular infrastructure based on gearman [3] has been developed. It is composed of :

- Workers dedicated to specific tasks (DQ2 lookup, Storage Renaming...)
- The workers get the payload for a gearman server.
- The gearman server is under the supervision of a process called the supervisor that feeds it with list of datasets.




The renaming is done in 6 steps :

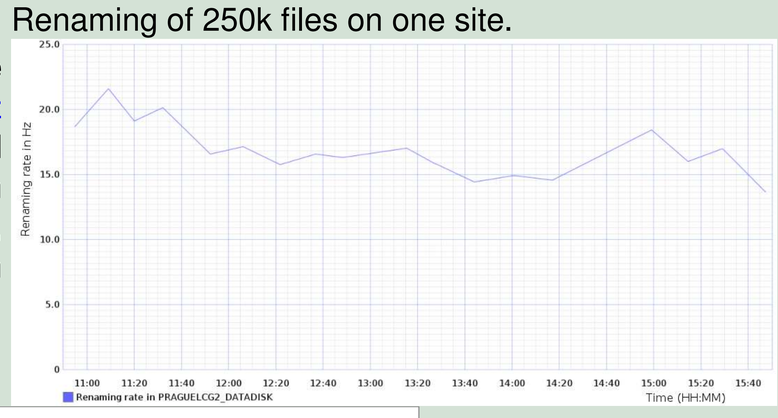
1. The supervisor queries the DQ2 catalog that stores the dataset replicas to get a list of datasets on a given site. The operation consists in a single query, and don't need to be parallelized between multiple workers.
2. The supervisor sends the list of datasets to the DQ2 lookup workers that query the DQ2 catalog that stores the dataset content. A query is performed for each dataset and therefore takes advantage of working on multiple workers. The result returned is a list of logical files identified by a GUID (Globally Unique Identifier).
3. The list of GUIDs is then used as input by the LFC lookup agents that identify the replicas (URI) associated to the GUID in the target site.
4. The list of replicas extracted by the LFC lookup agents is used as input by the LFC addreplica agents, which for each replica that follows the old naming convention will register in the LFC a new replica with the new convention. If some replicas already follow the new convention, they are skipped.
5. The list of files with the old and new convention is used as input by the Storage Rename agent, that interacts with the Storage Element. It renames the files with the old convention on the Storage Element to the new convention using the WebDAV [4] protocol.
6. The last part of the rename is to delete the old replicas from the LFC. This is done by the Cleaner agents.

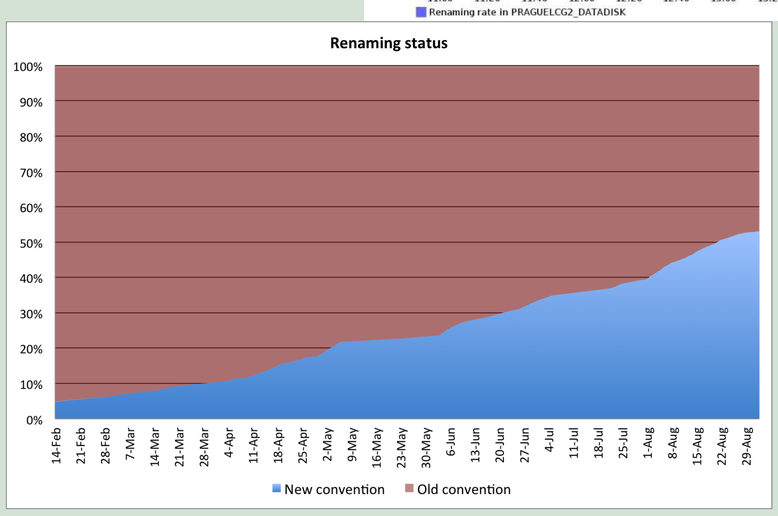
## 4. Results

The performance of the system is monitored via Graphite [5], a tool to store time series and visualise them in real-time.

- Renaming rate on one site of at least 10 Hz (joint physical and logical operations) can easily be achieved, even for sites with large RTT.

Renaming of 250k files on one site.




- Since the renaming can be run in parallel on multiple sites, 2-3M files can be renamed each day.
- As of today already more than 50% of the files (∼200M files) follow the new convention.

## 5. Conclusion

The renaming infrastructure developed to prepare the transition to Rucio has already successfully renamed hundreds millions of files without disrupting the ongoing computing activities, and it should be possible to finish the renaming campaign in time for Rucio deployment.

## References

[1] V. Garonne et al., "Rucio - The next generation of large scale distributed system for ATLAS Data Management.", see talk on Tuesday afternoon.

[2] V. Garonne et al., "The ATLAS Distributed Data Management project: Past and Future", *Journal of Physics: Conference Series*, **396** (032045), IOP 2012.

[3] Gearman: `http://gearman.org/`

[4] The IETF Trust, "HTTP Extensions for Web Distributed Authoring and Versioning (WebDAV)", *RFC 4918*, 2007.

[5] Graphite: `http://graphite.wikidot.com/`