

# Rucio Storage Dumps

mario.lassnig@cern.ch

# Overview

---

- Objectives
  - Verify the contents of storage with Rucio database
  - Make it automatic where possible
  - Make it easy for site-admins when manual verification is needed
- Solution (for now)
  - Provide periodically updated dumps of content
  - HTTP access, tab-separated, (compression TBD)
- ETA
  - End of week

# Data volume

---

- Data Identifiers (DIDs)
  - Files 550'000'000
  - Datasets 6'800'000
  - Containers 4'500'000
- Rules 7'500'000
- Replicas 690'000'000
- Locks 650'000'000
- RSEs 700

Sqoop from Oracle to Hadoop: ~4 hours

Volume: ~16GB bz2 compressed

# Contents - RSEs

---

Description of the Rucio Storage Elements.  
Needed to resolve RSE IDs to RSE Names in  
other dumps.

**rse\_id:** <uuid>

**rse:** CERN-PROD\_DATADISK

**rse\_type:** DISK

**deterministic:** 1

**volatile:** 0

# Contents - Dataset Locks

---

Holds the dataset replicas requested via rules.

**scope:** cmccond

**name:** cmccond.000001.conditions.recon.pool.v0000

**rule\_id:** <uuid>

**rse\_id:** <uuid>

**account:** root

**state:** 0

# Contents - DIDs

---

Holds the definition of containers/datasets/files.

**scope:** data08\_rof

**name:** data08\_rof.00060688.calibration\_calibration.daq.RAW.o3

**account:** root

**did\_type:** D

**hidden:** ....

**is\_open, complete, obsolete, bytes, length, events, project, datatype, run\_number, stream\_name, prod\_step, version, task\_id, panda\_id, campaign, lumiblocknr, provenance, phys\_group, transient**

# Contents - Replicas

---

Holds the location of files.

**scope:** data08\_idcomm

**name:** daq.ATLAS.0096542.physics.RPCwBeam.LB0013.SFO-2.\_0014.data

**rse\_id:** <uuid>

**bytes:** 1509885748

**state:** A

**lock\_cnt:** 1

**adler32:** ce3acd0d

**created\_at:** ....

**accessed\_at:** ....

**path:** data08\_idcomm/RAW/data08\_idcomm.00096542.physics\_RPCwBeam.daq.RAW.o4\_T1227757005/daq.ATLAS.0096542.physics.RPCwBeam.LB0013.SFO-2.\_0014.data

# How to get them?

---

Proposed URLs (not ready yet):

<https://rucio-hadoop.cern.ch/dump/2014-12-03/rses.bz2>

<https://rucio-hadoop.cern.ch/dump/2014-12-03/dslocks.bz2>

<https://rucio-hadoop.cern.ch/dump/2014-12-03/dids.bz2>

<https://rucio-hadoop.cern.ch/dump/2014-12-03/replicas.bz2>

Will notify you once live at the end of week.

Current time estimate: Daily 05:00 CERN time



# Additional dumps

---

We will also provide smaller, RSE-split, replica dumps:

**scope, name, SURL, Adler32, bytes**

Proposed URL:

[https://rucio-hadoop.cern.ch/dump/2014-12-03/replicas/CERN-PROD\\_DATADISK.bz2](https://rucio-hadoop.cern.ch/dump/2014-12-03/replicas/CERN-PROD_DATADISK.bz2)

We can easily prepare more merged/mapped/split dumps out of these, so you don't have to.

So please go through available attributes, select what you like, and we will prepare them for you.