

IceCube Long Term Archive

Patrick Meade
28 Feb 2019

IceCube Laboratory (ICL)

I'm sorry Dave, I'm afraid I can't do that...



IceCube Neutrino Observatory

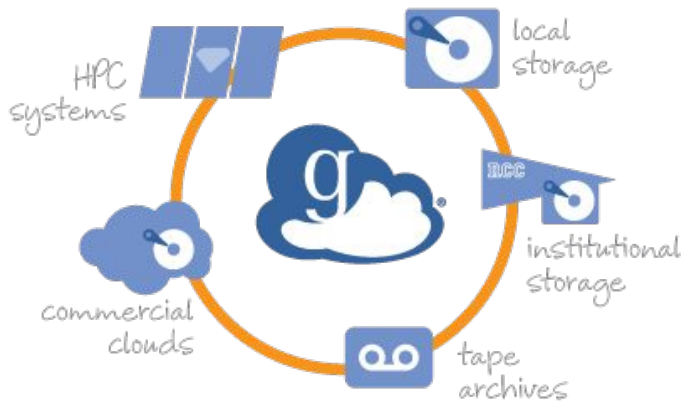
- Located at the Geographic South Pole
- ~5400 Digital Optical Modules (DOMs) in the ice (1-2 km deep)
- Raw data is ~1 TB/day, sent via cargo ship; 1 shipment/year
- Filtered data is ~80 GB/day, sent via satellite; daily transfer
- South Pole Station is open from October through January

IceCube Neutrino Observatory

- Detector was completed at the end of 2010
- Since completion, IceCube has generated...
 - 3 PB of Raw data
 - 200 TB of Filtered data
 - 2 PB of Level 2 and Level 3 analysis data
- Backup of Filtered, L2, and L3 data to DESY
- Backup of Raw data to NERSC

Problem

- JADE Long Term Archive (LTA) software was written to use Globus Online...



- Globus Online is going to a commercial, closed source model.

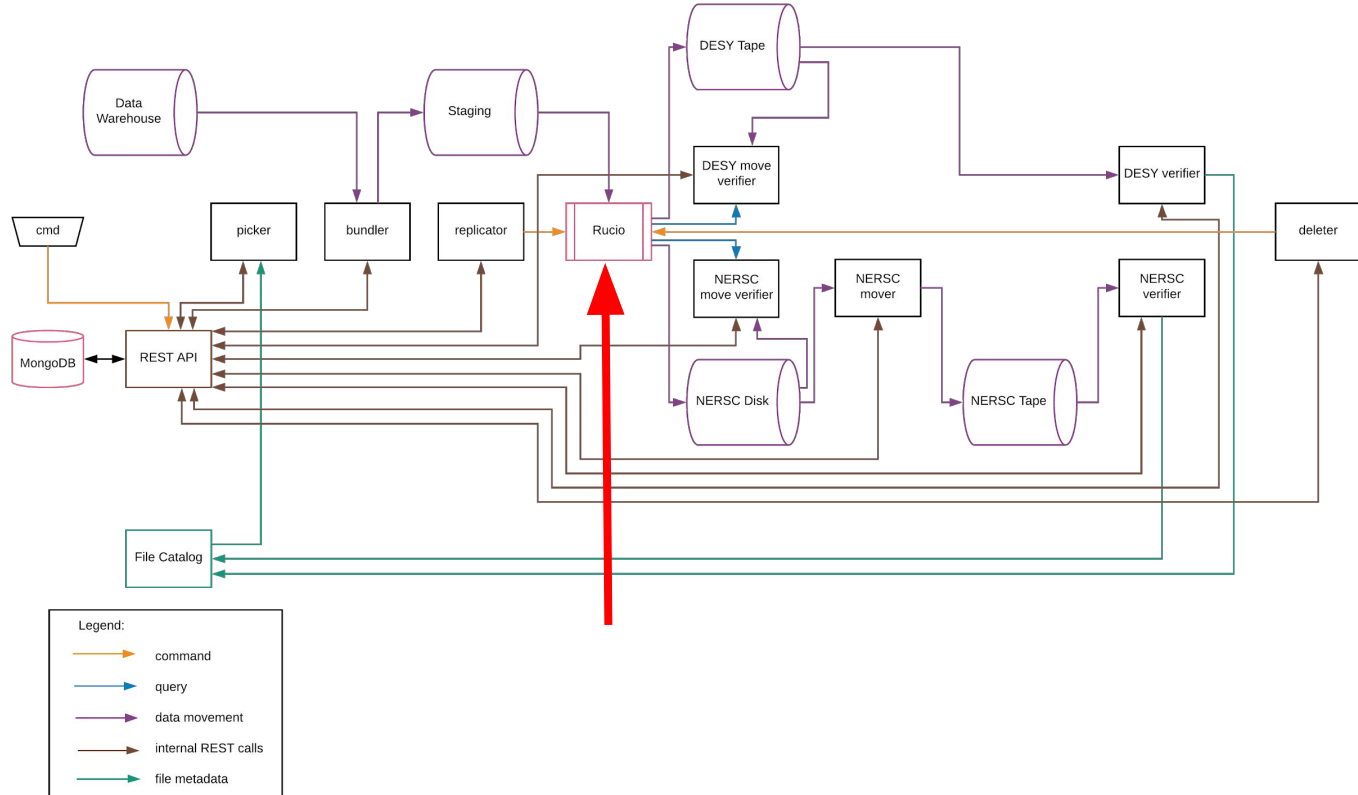
Lemons into Lemonade

If Globus is going away ... why not write new LTA software in Python?



<https://github.com/WIPACrepo/lta>

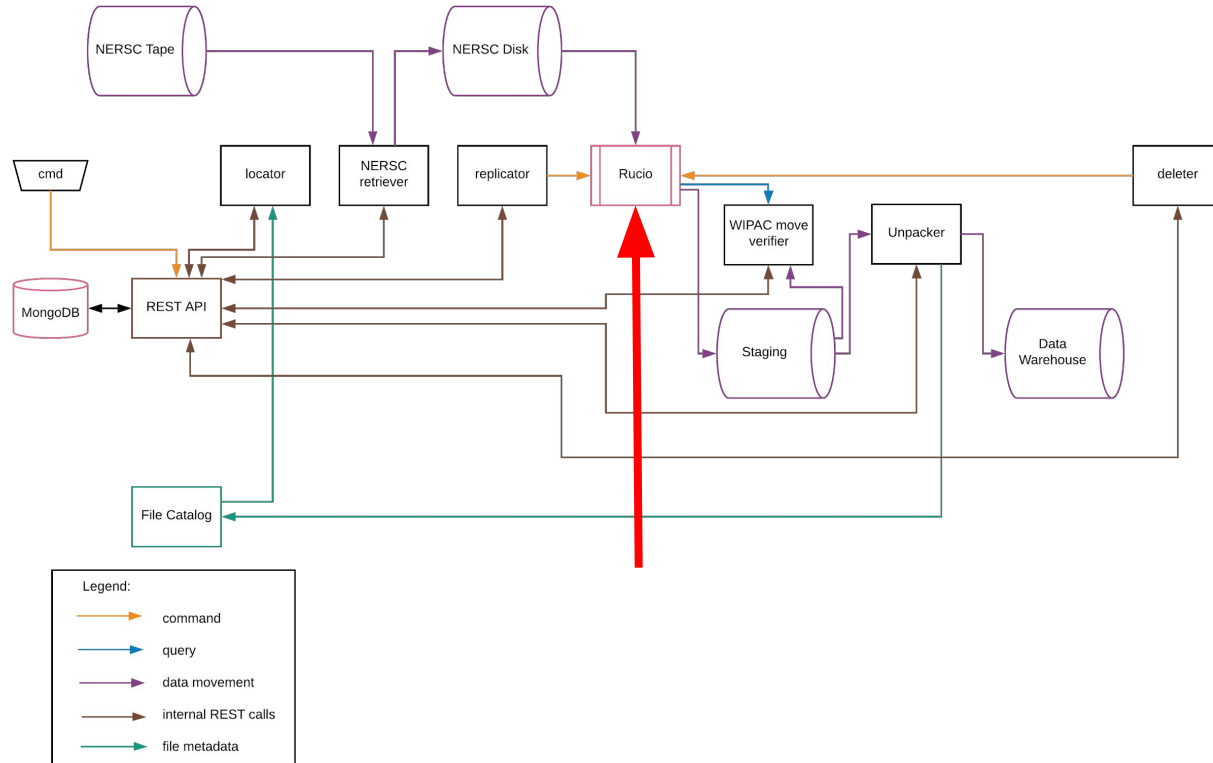
UW Madison -> {DESY, NERSC}



Long Term Archive - Archive

- Picker identifies Files to be sent to an archival destination
- Bundler consumes Files to be made into Bundles (ZIP64)
- Replicator prompts Rucio to copy Bundles to archival destination
- Move Verifier verifies the Bundle was copied to the destination
- Verifier verifies the archive can be read back from the destination

NERSC -> UW Madison



Long Term Archive - Restore

- Locator identifies Bundles to be brought back to the Data Warehouse
- Replicator prompts Rucio to copy from archive to UW-Madison
- Move Verifier verifies that the Bundle was copied to UW-Madison
- Unpacker decompresses the Bundle and stores Files in the Data Warehouse
- Deleter deletes the Bundle after the Files are verified in the Data Warehouse

Rucio Question 1

- Why Adler32 for a checksum?
 - It is fast, but not particularly strong.
- SHA512 usually has hardware support
- BLAKE2 is a fast secure hipster hash if SHA-2 is too mainstream

Rucio Question 2

- Consider Support for Archival Containers?
- Rucio Data Model
 - Files
 - Datasets
 - Containers
- Would be nice if Rucio could consider an archive file a container
 - (Un)Compressed Tarball
 - ZIP Archive
- If within scope, can IceCube help with this?

Rucio Question 3

- Any chance for Better Documentation?
 - Full Tutorial on Rucio setup; including all major daemons
 - Setup including multiple RSEs and FTSs
- IceCube is setting up its own Rucio instance for LTA
- An LDAP/Oath2 example might be:
<https://www.ory.sh/docs/hydra/5min-tutorial>
- Can IceCube help with this?

Thank You for Your Kind Attention ^_^



Patrick Meade - pmeade@icecube.wisc.edu