
Track 4 Highlights – Data Organisation, Management and Access

Alessandra Forti, Brian Bockelman, Xavier Espinal, [Tigran Mkrtchyan](#)

Thank You!

- 71 abstracts, 43 oral
- Excellent content & peaceful discussions
- Almost always full room
- Almost all speaker perfect in time

Main Directions

- Storage
- Experiment
- Middleware

Storage System Evolution

- Updates for scale-out
- Geo-distribution
- Reliability
- Token-based authorization
- New Tape system and replacements

Common Problems, Common Solutions

- LHC experiments do not dominate in data volume
- FTS is de-facto THE data transfer service
- Rucio is Breaking ATLAS Boundaries
 - missing functionality contributed by community

“Data Lakes”

- Various type of distributed deployments
 - DPM, dCache, Dynafed, EOS, xCache
- Aim to optimize storage usage
 - HW
 - Manpower
- R&D activities to prof the economical benefits
 - Result at next CHEP (or earlier)?

Latency hiding

- CPU and storage not always co-located
- Latency hiding makes remote access affordable
- xCache is de-facto caching technology around
- Remote access to Conditions Database subject to latency hiding

3rd Party copy (the problem)

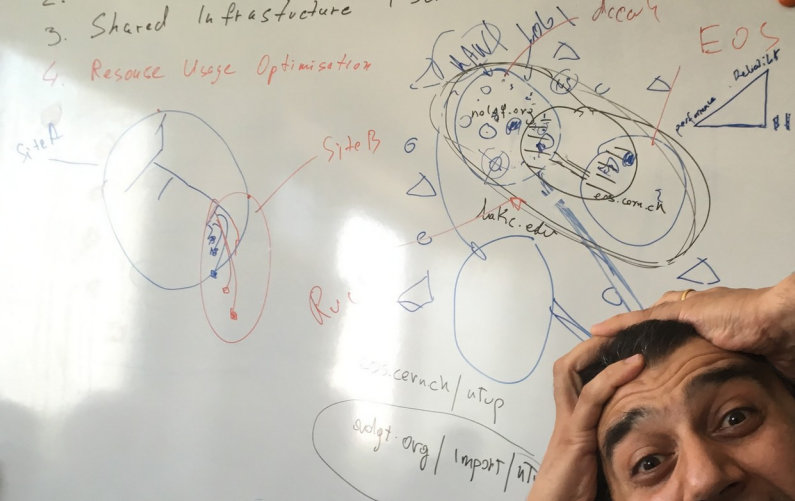
- GridFTP (Globus Toolkit) is not supported anymore (Dec 2018)
- Working alternatives are required at all sites ...
 - DPM, StoRM, dCache, XrootD, EOS, ECHO, Dynafed
 - FTS, Rucio, gfal ...

3rd Party copy (starting point)

- Two protocol is considered
 - Xroot
 - HTTP
- All storage provides must provide a test end-points
- First production deployments should be available by the end of 2019!

0. Evolution Project (Not A Revolution!)

- 1. Reduce Cost: local + global
- 2. Scale-out:
- 3. Shared Infrastructure (Science edu operative)
- 4. Resource Usage Optimisation



Data Organisation Managa

Cost: local site adm
 QoO: Ops team large
 Site mostly UQ

RAIN-X → L&O

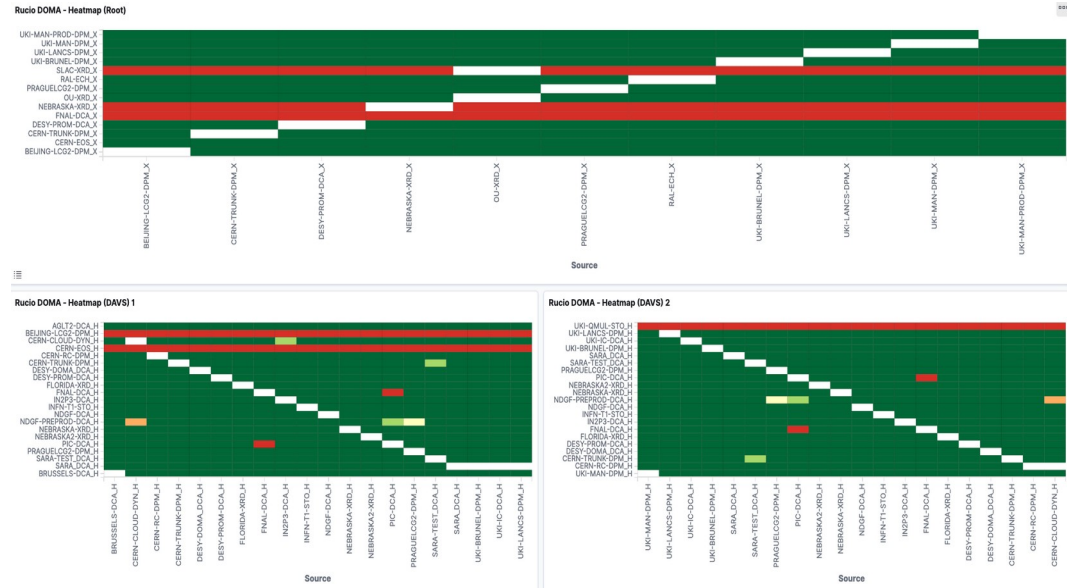
N-Rep

QoS
 QoS
 QoS
 QoS



3rd Party copy (today)

- All storage flavors
- 36 sites in testbeds
- Multiple test-runs per day



Do we use storage efficient?

- 90% of data deleted in 80 days
- 90% of RAW data re-read in the same day
- 90% of accessed < 15 times

Summary²

- Storage systems becomes smarter, bigger ... and simpler
- Commonality in experiments allow us to share existing solutions
- Lot of effort goes into understanding of data usage
- Latency hiding solutions make remote data access more affordable
- And yes, we need that network works!

Let **thanks** all speaker again!