Quality of Service (QoS)

for cost-effective storage and improved performance

Martin Barisits, Xavier Espinal, Patrick Fuhrmann, Edward Karavakis, Oliver Keeble,
Mario Lassnig, Paul Millar, Markus Schulz, Vincent Garonne, Eric Lancon, Alexei Klimentov, Xin Zhao

Outline

- From motivation to objective
- The QoS site survey
- QoS examples from the experiments
- Roadmap for site-level and grid-level activities

Motivation

- Quality of Service
  - A quantitative measure of performance characteristics
  - Intended to be associated with a cost and a workflow
  - "Unreliable and cheap", "Fast and expensive", …
- QoS is asking questions such as
  - Are there places in experiment work-flows where it makes sense to trade performance/reliability for increased storage capacity?
  - Are there places in experiment work-flows where a small amount of higher performance storage would yield significant benefits?
- QoS our umbrella term for optimally mapping our dataflows onto our storage under a *throughput*, *latency*, *resiliency*, *space*, and *cost* constraint

The QoS working group activities

- Site survey
  - Understand the current and potential QoS landscape
- Experiment contact
  - Map workflows onto QoS
- White paper
  - A short reference on status and opportunities for cost savings through QoS in WLCG
- Liaise with
  - WLCG Working Groups, Storage Development, Experiment R&D efforts
- Get involved
  - *https://twiki.cern.ch/twiki/bin/view/LCG/QoS*
  - *https://e-groups.cern.ch/e-groups/EgroupsSubscription.do?egroupName=wlcg-doma-qos*

Site survey

- Understand how embracing diversity in storage can save money
- Capture relevant ongoing efforts and directions relating to cost-saving and QoS

- Around 80 sites responded across the whole Tier spectrum
- Raw results available at
  - *https://twiki.cern.ch/twiki/bin/view/LCG/QoSSurveyAnswers*
- Our attempt at an analysis
  - *https://twiki.cern.ch/twiki/bin/view/LCG/QoSSurveyConclusions*
  - This contains much more detail than presented here

Site survey — Summary of a summary 1/2

- Underlying media
  - Majority of sites report using enterprise media
  - No obvious request for consumer-grade drives
- RAID6 with 12-16 disks represents over 2/3 of our sites
  - This does not give much margin for further cost savings
  - All the rest do JBOD, where redundancy comes from Ceph, EOS, HDFS and GPFS
  - Abandoning redundancy would give ~15% increase in space — when would it be worth it?
- Few surprises in storage systems
  - Almost always is a shared POSIX filesystem required at the site
  - Saving effort due to consolidation of different solutions required at each site
- Practically all T1s share multiple experiments, half of T2/3s are shared
  - LHC experiments typically on separate systems

Site survey — Summary of a summary 2/2

- Effort estimation is inconclusive, except when reading between the lines
  - No clear mapping of human efficiency onto system cost (T1: 2.5 FTE, T2: 0.6 FTE)
  - Majority of cost does not come from storage operation
- The vast majority of T2s are neither planning/wanting to move to storage-less setups
  - Local storage is needed
  - This would cut off independent lines of funding
  - Diverging opinions regarding caching esp. between T1s (helpful) and T2s (not helpful)
  - At the same time, T1s express uncertainty about extra load from T2s
- Few signals from sites regarding storage cost-saving R&D
  - Mostly towards JBOD+Ceph, but concerns about breaching MoU
- Clear interest for better integration with experiment frameworks
  - Even sites seem to take a more user-centric view — granularity of workflows/dataflows

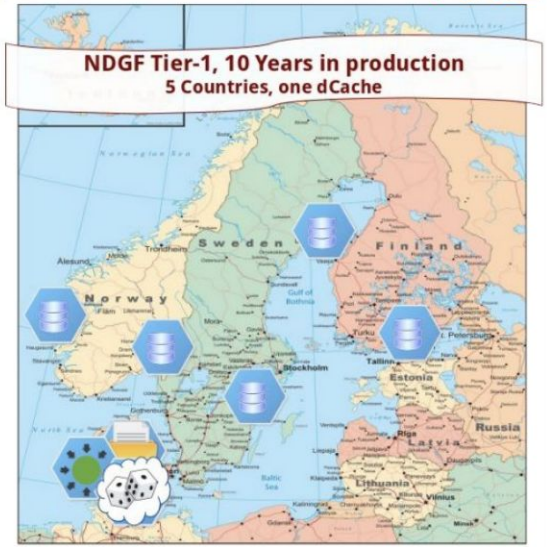
The NDGF Example

dCache.org powered

ARC

Multi-Site deployment

- Distribute data over multiple locations
- Multiple administrative domains
- Use available resources



Disk and tape distributed over many sites
but appear as one entity

ARC cache for computing element caching
Shared file system on the worker node

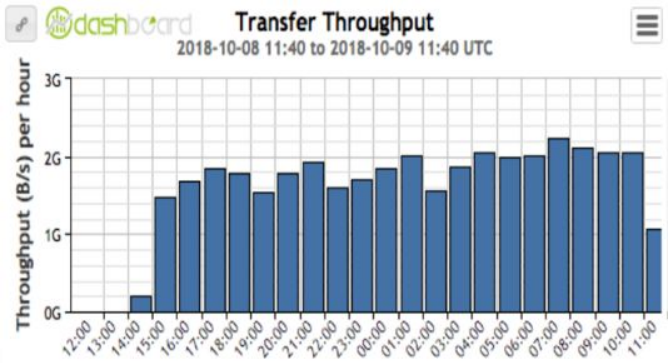
ATLAS Data access numbers

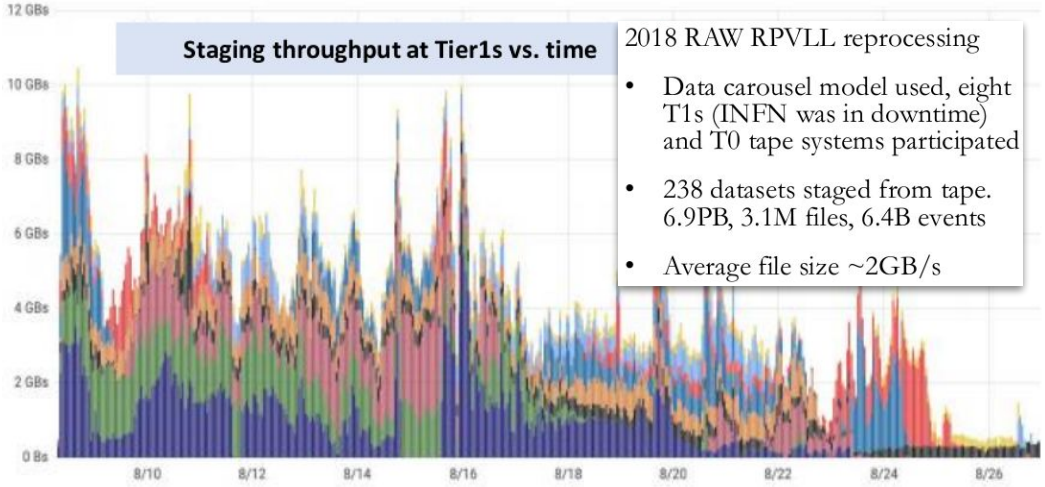
Production & Analysis	6 PB/day of input data
	2M jobs/day
Transfers	1 PB/day
	2.5 M transfers/day
User download	500 TB/day

Cache space is not reported to the experiment

Data Carousel

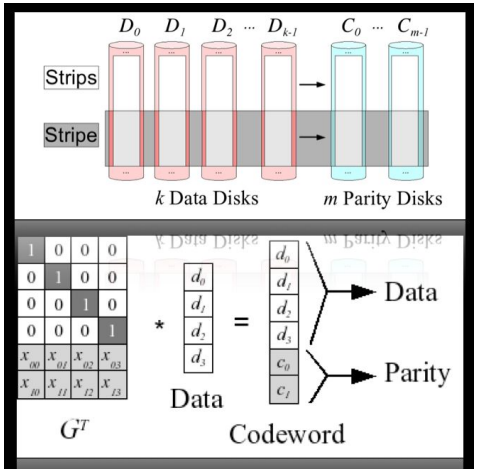
- ATLAS started the Data Carousel R&D in mid 2018 — client-based view of QoS
  - Study the feasibility to get inputs from tape directly for various ATLAS workflows, including derivation production and RAW reprocessing
  - Phases: (1) Tape Sites Evaluation, (2) ProdSys2/Rucio/Facilities integration, (3) Production
- Investigate smart staging submission
- Investigate bulk writing methods





EOS Erasure coding

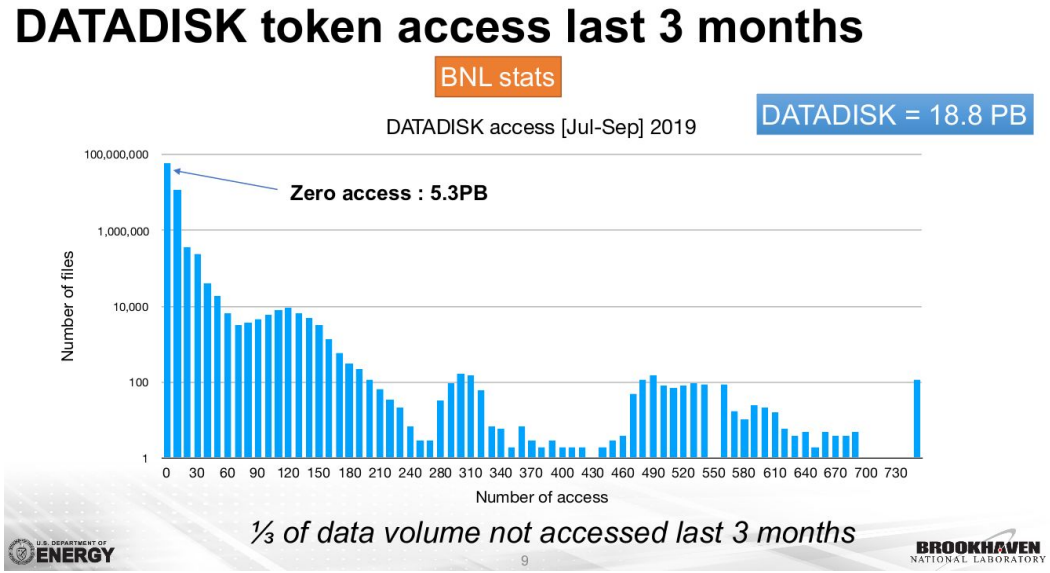
- Uses Reed-Solomon algorithm from JERASURE
- EOS Erasure Encoding is file based
- CERN
  - Wigner decommissioning allows us to move from double replica to erasure coding
  - Offers significant potential for cost saving
  - Exact policy used is likely to depend on file size and use cases
- Demonstrated in the ALICE DAQ instance of EOS
  - RS(12,10)
  - Encoding of 2.4 (4.8 RAW) PB freed 2 PB of storage in the pool
- https://indico.cern.ch/event/775181/contributions/3292106/



Andreas-Joachim Peters

Multilayer Automatic Storage (MAS)

- R&D Initiative at BNL
- Created a 3rd QoS class
- Not all data are equal
  LOGs are not RAW
- Move under some conditions some data from disk to tape and pretend they are on disk
- Large dCache buffer size in front of HPSS: 2.5 PB (arbitrary size) from retired equipment

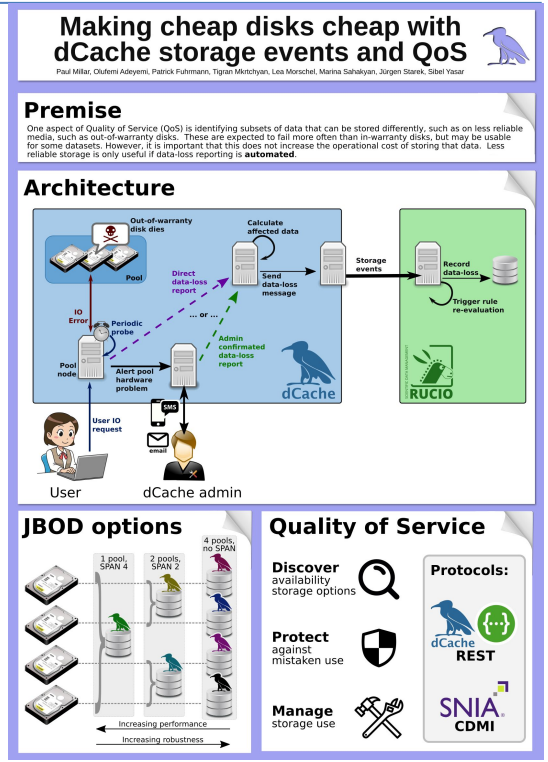


- https://indico.cern.ch/event/823340/#21-bnl-rd-on-multilayer-automa

Eric Lancon

dCache low-reliability QoS

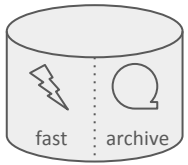
- Adding support for new QoS class: low-reliability
- Based on end-of-warranty media
  - “Free” extra storage, but not expected to be reliable.
  - Failed drives are not replaced: “you get what you get”.
  - Anticipate using JBOD: drive failure ⇒ data lost.
  - New drives added as media falls out of warranty period.
- Low expectation: not suitable for all data
  - Examples: COLD data, staged data speculatively kept on disk, …
- Storage events used to avoid operational overhead
  - Allow external agents (e.g., Rucio) to learn of data loss.
  - Recovery options: copying, restaging, rerunning analysis, …
- Data-loss events may be useful for other QoS classes



https://indi.to/Z58b7

Rucio

- Replica management in Rucio is based on replication rules
  - Put `2` copies of `file.001` on RSEs matching expression `country=uk&type=disk`
  - Rule engine matches the rule against existing replicas (rules); only requests transfers if necessary
- QoS for experiments thus means expressing QoS on data, not on storage
  - `2` copy of `file.001` on RSEs matching `country=uk` with qos `latency<50` for `30` days, after that change qos to `latency<1000` for `∞` days.
- Rucio will support RSEs with multiple storage classes/zones
- Need to agree on standard how to communicate QoS needs to storage
- Can storage independently move a file into different QoS zone?
  - This implies a de-facto ranking of QoS classes/attributes
  - Ranking not obvious, as QoS classes may not “outrank” each other in all attributes



Site-level activities

- Procurement, densification and media
  - Purchasing strategy, server density and overheads, SMR, SSDs
  - Networking & implications of the data lake model on origin storage
  - This is trying to provide the current QoS, unchanged, at lower price
- Software Defined Storage (SDS)
  - Configurable storage characteristics
    - Replication, erasure coding, media transitions
    - Future of RAID-6 and higher capacity disks
    - Pure JBOD operation
  - Introducing SDS into the stack
    - Use of Ceph, HDFS etc behind existing grid storage systems
    - Direct use of cluster FS tech by experiments (e.g. CephFS)
  - Identification of which QoS classes can be mapped onto which WLCG workflows

Site-level actions

- Invite sites to begin a classification of their current offerings
  - Do this after the first set of recommended classes has been produced
  - Sites should also add any they are interested in providing
- Invite sites to report on current relevant directions
  - Media diversity
  - Redundancy layer
- Attempt the "JBOD experiment" in conjunction with an experiment
  - Remove all redundancy and work on handling data loss gracefully
  - Collaborate with the Access&Caching Working Group
- Understand where this should be progressed
  - Many points are under the experiment radar and more aligned with forums like HEPiX

Grid-level activities

- WLCG QoS classes
  - Definition of the most useful set of QoS classes beyond "disk" and "tape"
  - Tagging and brokering
  - MoU and pledges
- Client-driven QoS
  - Interfaces, clients, orchestration
  - Including bring-online
- Data Lifecycle

Grid-level actions

- Organise a dedicated consultation meeting with each of ALICE, CMS and LHCb
  - Identification of QoS classes useful to the experiment
  - Define reference use cases for these classes
  - Plan for how QoS class can be introduced, integrated, tested, and exploited

- Solicit sites to provide new experimental storage areas with novel QoS features
  - Trigger exploration of what adaptations are needed over the stack
  - Experiment/Site combinations could be encouraged to report

Summary and roadmap

- The WG is part of the drive to solve WLCG's cost problem
  - Introduces a new consideration - QoS
  - Attempt to discover, promote, trigger and report all such activities in the project
- Our current activities
  - Ongoing consultations with WLCG experiments
  - Delivering a white paper
- Next steps
  - Dedicated Pre-GDB meeting
  - Site-level activities and actions
  - Grid-level activities and actions