DOMA-QoS

eXtreme DataCloud

Data Management for extreme scale computing

Paul Millar

paul.millar@desy.de

GDB meeting

Wednesday 16th January 2019

eXtreme DataCloud is co-funded by the Horizon2020 Framework Program – Grant Agreement 777367

QoS: two rhetorical questions

- QoS is asking two questions:
  - 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?

    (Note that these questions are strongly experiment focused: this effort will only be successful with strong input from experiments.)
- Assuming the answer to these questions is “yes” then how do we achieve these trade-offs?

QoS: background

- HEP has a long tradition of handling storage QoS:

  We have stored data on tape as reliable and cheap media, and recall data back to disk when needed.

- This has served us well, but the terms DISK and TAPE are increasingly problematic:
  - DISK: NVMe → SSD → HDD (SAS/SATA/Shingled/Commodity…)
  - TAPE: magnetic, optical, highly-redundant geographically distibuted disks
- Also want to understand whether there is redundancy
  - R`AID, plain disks (with multiple copies or erasure coding) – is this needed?
- Better to describe storage by **expectation**, rather than media:
  - Support adding new technologies.
  - Allows sites to innovate

DOMA-QoS: our motivation

“Given the expected **flat budget** for High-Lumi / RUN 4, create a mechanism to allow a **diversity** where **sites** can offer specific QoS options through innovative solutions that **save cost**. Through this **competition**, drive down the total cost of storage, while allowing **experiments** to optimise their **storage usage**.”

from DOMA-QoS Mandate

DOMA-QoS: our motivation

eXtreme DataCloud

“Given the expected **flat budget** for High-Lumi / RUN 4, create a mechanism to allow a **diversity** where **sites** can offer specific QoS options through innovative solutions that **save cost**. Through this **competition**, drive down the total cost of storage, while allowing **experiments** to optimise their **storage usage**.”

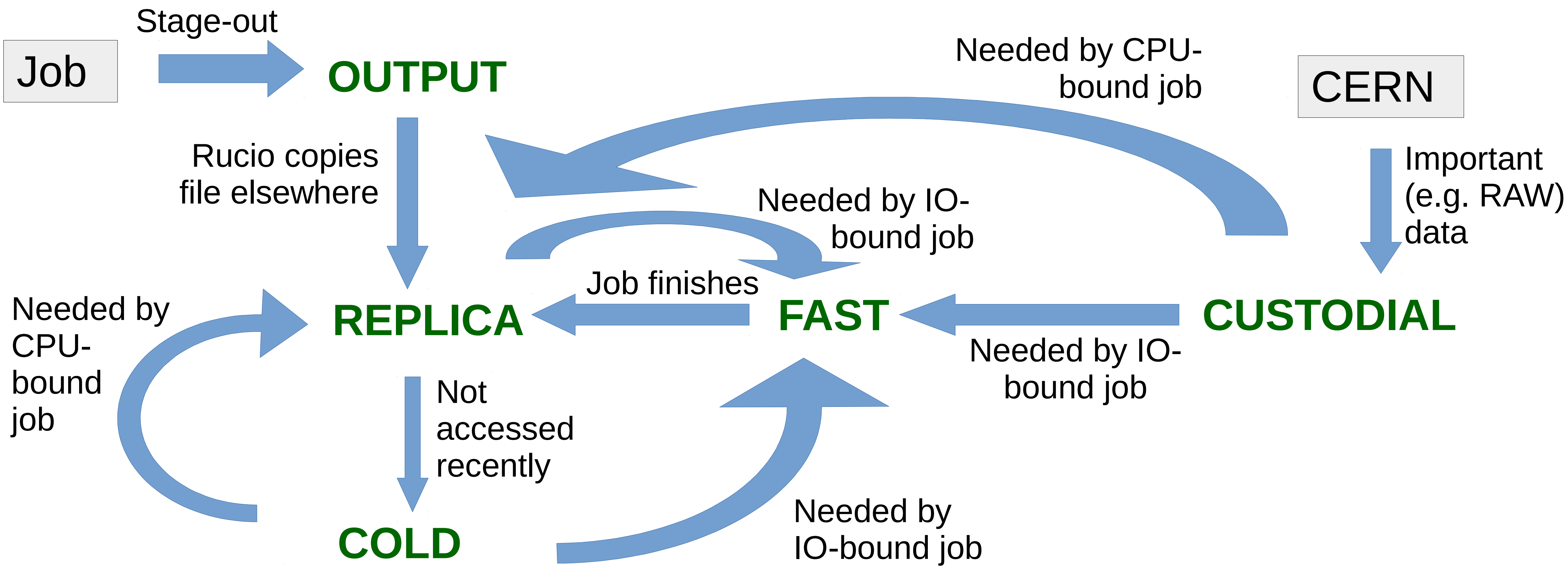
SAVE MONEY

from DOMA-QoS Mandate

DOMA-QoS: strawman model

- DISK → OUTPUT, REPLICA
  - **OUTPUT** storing only existing copy of data
  - **REPLICA** data also exists elsewhere (data loss more acceptable)
- TAPE → CUSTODIAL, COLD
  - **CUSTODIAL** storing data that must not be lost.
  - **COLD** data that is only used in bursts, and currently not being used.
- DISK → {OUTPUT/REPLICA}, FAST
  - **OUTPUT/REPLICA** input data for non-IO bound (analysis) jobs
  - **FAST** input data for IO bound jobs.

DOMA-QoS: strawman model



DOMA-QoS: strawman examples

- Example storage QoS:
  - Enterprise HDD as RAID: **OUTPUT**, **REPLICA**, **COLD**
  - Consumer HDD as JBOD: **REPLICA**
  - (public) cloud storage: **COLD**
  - SSD as JBOD: **FAST**
  - Internal replicas existing on multiple server nodes: **FAST**
- Same site could have multiple QoS that have required QoS label
  - For example, enterprise RAID and consumer JBOD both have **REPLICA** label.
  - Use “cost” to drive decision: cheaper to store data on JBOD than RAID.
- Different sites could implement QoS using different technologies
  - As above, would like “cost” to drive decision.

Current activity

- Engage with **experiments** to explore adapting workflows to include QoS concepts,
- Engage with **sites** to learn what technologies are currently available, and from their experiences of technologies that are currently not available to experiments,
- **Coordinate** our activities within the wider community: other DOMA activities, WLCG workgroups, and (potentially) further afield.

Engaging with experiments: ATLAS

- Very enthusiatic participation
  - Our “QoS minion” (Mario Lassnig) is very active within ATLAS and QoS.
- Current focus in ATLAS is on a data carousel prototype
  - Needs a small development effort and changes in workflow to support this
- Will use experience from data carousel to drive further QoS changes
  - There is a clear plan on how to move forward: run a full derivation from tape, using “manual” QoS, and instrument what we gain from it.
- ATLAS is looking at a adopting some common WLCG technology for QoS
- ATLAS will also consider networking at part of QoS
  - Networking QoS is non-trivial (see NOTED project),
  - Although we currently consider network out-of-scope for DOMA-QoS, likely be some connection between Storage QoS and networking QoS.

Engaging with experiments: CMS

- Internal discussion have started
- Many ideas CMS presented resonate with concepts within DOMA-QoS
  - Looks promising: we seem to be on the right track.
  - Informal communication suggests QoS would support some existing ad-hoc workflows.
- CMS haven’t identified an official “QoS minion”
  - Informal feedback received suggests CMS management appreciate QoS as an interesting and potentially useful technology.

Engaging with experiments: ...

- We currently have no formal participation from LHCb or Alice
  - Participation welcome!
- We are also investigating other ways of interacting
  - For example, preparing a white paper providing concrete ideas, allowing VOs to comment.

Engaging with Sites: site survey

- Aim to learn more about storage at WLCG sites, and learn from their experiences.
- Survey is now in “Release Candidate” stage.
- Use CERN and DESY as “guinea pig” sites: both to test the questions and provide sample responses.
- Aware of a somewhat overlapping survey by the Cost Modelling WG
  - The CM WG survey targets only Tier-1 sites and has a different focus
  - Include text in DOMA-QoS survey, saying sites may use their CM WG survey answers, where appropriate.
- Plan to send out the survey in January.

Summary

- QoS can save us money.

  i.e., Increase capacity for fixed budget

- Our job is to understand if we can use it

  i.e., our job is NOT to impose it!

- Get involved: **sites are very welcome**!

  https://twiki.cern.ch/twiki/bin/view/LCG/QoS

  Egroup: **WLCG-DOMA-QoS**

  https://e-groups.cern.ch/e-groups/EgroupsSubscription.do?egroupName=wlcg-doma-qos

Thanks for listening!