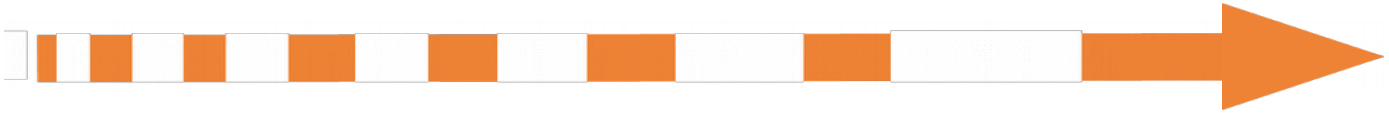
DOMA-QoS



Paul Millar

paul.millar@desy.de

eXtreme DataCloud

Data Management for extreme scale computing

DOMA-ACCESS meeting

Tuesday 20th November 2018

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

DOMA-QoS: why?

- Anticipating a fixed budget for RUN-4.
  - Likely not have as much storage capacity as we would like
- We want to make optimal use of the available budget
  - Allow optimal use of deployed storage media (RAID, n copies, JBOD, erasure coding, …)
  - Allow optimal choice of media (enterprise HDD, cheap consumer HDD, SSD, …)
- Rephrase this as minimising the cost per file
  - Think of different storage configurations has having different costs
  - Which storage option provides the cheap cost, while providing the expected behaviour characteristics (i.e., the required QoS)

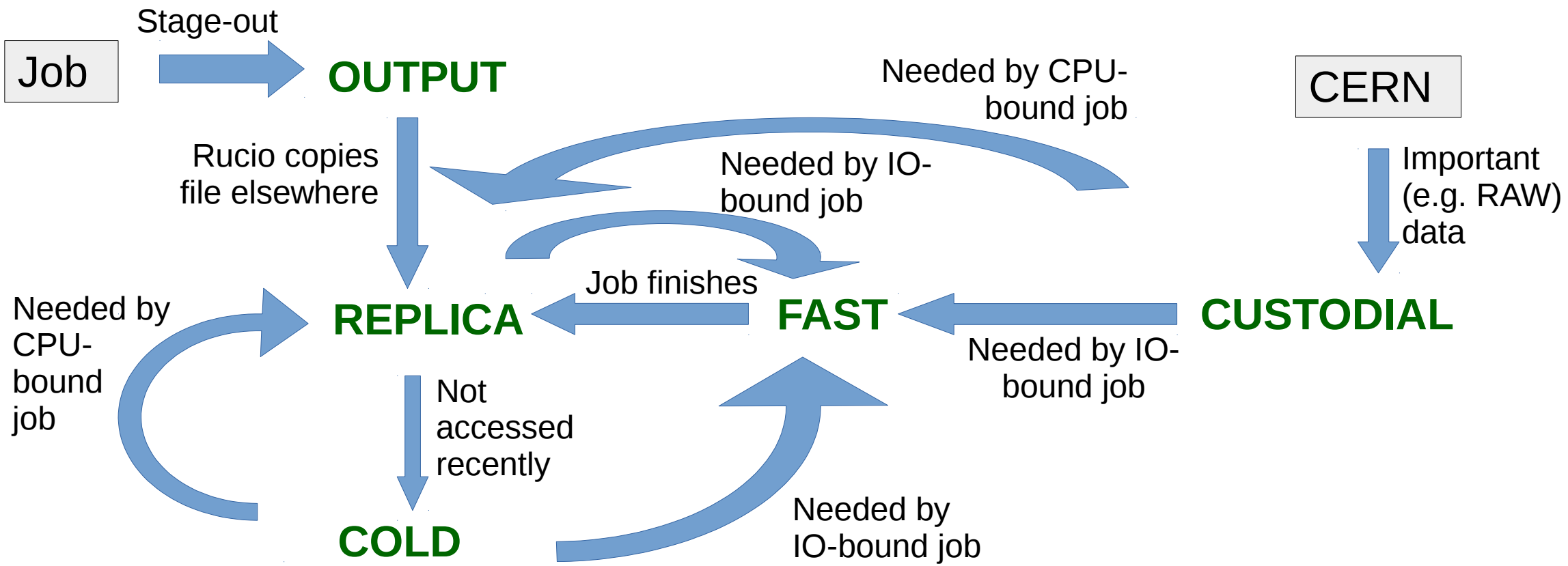
DOMA-QoS: strawman model

- The strawman model exists to explain QoS concepts.
  - This is not (necessarilly) what we will end up with!
  - The real QoS model requires a collaboration with the VOs
- The model takes the current storage QoS (DISK & TAPE) and expands them in a simple fashion.
  - There are other possible QoS not covered here – remember, this is meant as a pedagogic aid.

DOMA-QoS: strawman model

- DISK → OUTPUT, REPLICA
  - **OUTPUT** storing only existing copy of data
  - **REPLICA** storing one copy of data
- TAPE → CUSTODIAL, COLD
  - **CUSTODIAL** storing data that must not be lost.
  - **COLD** storage data that is 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: examples

- Example storage QoS:
  - Enterprise HDD as RAID: **OUTPUT**, **REPLICA**, **COLD**
  - Consumer HDD as JBOD: **REPLICA**
  - (public) cloud storage: **COLD**
  - SSD as JBOD: **FAST**
  - Enterprise HDD as RAID, with multiple replicas existing on separate 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.
  - Would like some notion of “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.

DOMA-QoS and DOMA-ACCESS

- What can we learn about how jobs “use” storage?
  - QoS distinction makes sense if jobs are also somehow distinct
  - For example IO-bound vs CPU-bound
- Do we include caching as a QoS attribute?
  - It’s not guaranteed to read a file through a cache
  - It does (potentially) bring benefits.
  - Is this a question for scheduling
- Do we need to have geographic aware QoS?
  - For example: two copies within the data lake, but don’t care where.
  - Do we have a concept of “pinning” data to a geographic location (for jobs). Is this a QoS operation?