ESCAPE WP2 Fortnightly 4th April 2019

Indico: https://indico.in2p3.fr/event/19024/

Participants: Simone Campana, Yan Grange, Paul Millar, Xavier Espinal, Tigran, Aristides, Ghita Rahal, Martin Barisits, Ron Trompert, Fabio Hernandez, Manuel, Paul Kemp, Eric Fede, Lars Fischer, Mario Lassnig, Frederic Gillardo,

Meeting notes

News

- The proposal for the first F2F meeting of WP2 is to held it on Amsterdam 1st-3rd of July (Mon 13:00 to Wed 16:00 , 2 and ½ days)
- Next presentations:
    - (TODAY) Storage QoS and data lifecycle orchestration (Paul Millar - DESY)
    - Ideas on federated storage infrastructures (Kilian - GSI)
    - Integration with compute: caching and content delivery (Xavier Espinal– CERN, Diego Ciangottini - INFN)
    - Hammercloud as commissioning framework (Jarka/Aris – CERN)
    - Network related aspect of WP2 (SKAO + CERN)
    - AAI (Andrea Ceccanti – INFN)

QoS and ideas for task 2.2

(https://indico.in2p3.fr/event/19024/contributions/70958/attachments/52906/68603/2019-04-04_ESCAPE_QoS.pdf)

- Why QoS?
    - No single storage technology can match desired behaviour, hence QoS is needed to build hybrid solutions. Example: cheap storage that is both robust (“tape”-like), and fast (SSD-like) doesn’t simply exist.

- Different media options have different characteristics. Different combinations of media: RAID, RAIN, JBOD, Erasure coding
- These also have different costs: cost in terms of raw capacity and cost in terms of money/budget-usage
- This is all very complicated. It is better to describe expectations, rather than dictate how storage operates.
    - Experiments decide what they really need and sites aim to provide what is desired – at a minimum cost.
    - This could work fine, provided everyone is honest.
- Available QoS at a site level: A site provides finite choices, not arbitrary selection
- QoS options at a site: a site may provide a single QoS, a site could provide multiple storage system each with a single QoS or a single storage system with multiple QoS.
    - Example: SCRATCH (latency), ARCHIVAL(durable) and FAST (latency and bandwidth)
- WLCG has a long tradition of working with QoS: they were called tape and disk until now.
- DOMA is the WLCG project looking at the Data Lake model. DOMA is split in three groups: Protocols, QoS and Access.
- DOMA-QoS is asking two main rhetorical 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?
- Paul presented a strawman model of possible data workflows with some simple 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
- DOMA-QoS: current activity
    - Engage with experiments to explore adapting workflows to include QoS concepts
    - Engage with sites to learn what technologies are currently available

    - Coordinate our activities within the wider community: other DOMA activities, WLCG workgroups, and (potentially) further afield.
- Take-away messages:
    - QoS is motivated by: Saving money and building something “better” than any one site can provide.
    - QoS is an abstraction of storage.
    - QoS is an experiment driven activity: It only makes sense if integrated into experiment work-flows (this is HARD)

Q (Ghita): Ensure experiment ideas?
A (Paul): Representative from experiments in WLCG DOMA working groups. Could follow the same approach.
A (Simone): Not discussed yet, but should proceed as we do n WLCG. Will be an iterative process. Ideally we need to have some understanding for the F2F meeting on QoS ideas from experiments and sites according to experiment workflows. Need named representatives from experiments to follow up our activities.
A (Ghita): ESCAPE could also provide consolidated feedback from other sciences to WLCG.
A (Simone): Suggest to Yan and Ghita to take this onboard and follow it up with CTA and LSST.

Q (Manuel): problem I see is that many of the projects will not produce data until 5+ years. Propose to work with some existing more mature projects that we can take as example and map it to reality.
A (Simone): Agree. Example is ASTRON and LOFAR, we can use this data.
Q (Manuel): PIC is also involved in MAGIC since long and new telescope with some new and readable data will be starting on fall. Ramping up operations during the execution of ESCAPE project. DES is also mature and have data but possibly data is on US (to be seen).
NdT: Manuel will be the contact for MAGIC.

Q (Paul): Need to iterate on this before F2F meeting.
A (Simone): Yes. As starting point the whitepaper being produced in WLCG for QoS should be distributed in WP2.

Q (Fabio): Vocabulary matters. Output labels should be called precious or unique (if we mean raw data). Characteristics of some data depends on experiments view but sites serving different experiments would need to use the same vocabulary for all of them.
A (Paul): Labels matters. And labels come from experiments and they might come with different names for the same “definition”. Sites would need to map these experiment labels to the storage QoS.

Q (Simone): One QoS might be that we accept to lose files (ie. MC files, loosing 1% might be cheaper to reproduce…) Need to deal with operational aspects as losing files is a pain to handle. Need to automate as much as possible. For instance for ATLAS there is a handshaking between the experiment and the site, eventually the site report which files are lost and then the info is injected into Rucio to cure the problem if possible. Need to think about the orchestration of data loss handling (ie. notifications). I suggest to discuss this in Amsterdam to become part of our plan although is not in the proposal.

Round table: NTA

Date for next meeting:
Thursday 18th of April at 10:00 (TBC due to eastern holidays, will circulate a doodle poll)

Xavier