ATLAS Input to Storage

8 June 2021

David Cameron (University of Oslo), David South (DESY)

Key requirements and concerns

- Changes in workflows and effects of larger files (100GB or more)
  - Heavier checksum calculation, RAID/EC recovery of lost data
  - Consequences for tapes (reliability in reading/writing, gaps in the tapes)
  - Copying to/from worker nodes (but today we already have up to 100GB input for 8 core jobs)
  - Heavier use of (zip) archives and access through all protocols (also remotely)
- Storage location relative to processing
  - Storage local to CPU mostly holds today and probably in the future at least for large sites
  - Smaller/opportunistic CPU resources may not have local storage and can take advantage of caching technologies
  - Distributed storage (e.g. NDGF-T1) is used mostly transparently already but with significant differences in workflow (push-model for jobs, ARC data staging and caching)
    - Not necessarily reduced operational cost for site or experiments
- QoS and breaking the disk/tape paradigm
  - requires coordinated QoS knowledge in Rucio, FTS and storage
  - Up to now all data is treated as precious - can the experiments dynamically trade reliability for cost?
    - i.e. pay for CPU to recreate lost data on cheaper more unreliable storage

Key requirements and concerns

- Object stores
  - Rucio provides some support for native object store access, but outside of R&D projects this is not used
  - ATLAS prefers filesystems on top of any OS
  - Performance benefits of OS only applicable in certain use cases (e.g. future analysis facilities?)
- Tapes
  - Fundamental to the storage cost model but is it a risk to rely on a possibly soon obsolete technology?
  - With heavier reliance on reading from tape even in run 3 several developments are key
    - SRM must be replaced with a unified API across the different storage implementations
    - More intelligence is needed in both writing and reading to optimise throughput -> tighter collaboration between experiments and sites
- Human resources
  - Six storage technologies in the document essentially all doing the same thing
    - A risk that any changes required in functionality or interfaces need to be duplicated six times
      - For example transition to WebDAV as 3rd party transfer protocol, QoS standardisation
    - On the other hand a motivation to not do expt/HEP-specific things, to allow "standard" storage to fit in
  - Long term sustainability - experiments in general have no direct involvement in storage software