

Storage Evolution

Simone Campana - CERN

Introduction

- This short presentation is meant to kick-start a discussion
- Such discussion will likely take several months
- It will happen at many levels (technical, funding, strategic,..)
- It will involve different bodies (GDB, OB/CB, ...)
- Today we discuss it in the context of WLCG operations and facilities. You have to start from somewhere ..

Storage Evolution

We would like to understand how national efforts see storage evolving over the next 5 years

- Motivations and constraints:
 - Budget pressure increasing for HL-LHC (and before that ...): higher lumi, flat funding at best, declining density laws
 - Recent technologies studies show opportunities for efficiency increase (caches, QoS)
 - Site dependencies and constraints (including non WLCG ones) to be considered evaluating cost effective solutions

- Storage Consolidation for T2 federations. What are the plans at national level? Options could be:
 1. Continue with the current scenario, growing up the size of storage of each center with flat budget
 2. Consolidate storage in fewer centers, supporting the whole federation. Some centers will run CPU only (+ caches?)
 3. Expose fewer storage instances but distributed across the federation. E.g. NDGF
 - Of course you can (and likely will) have mixed models

General implications

- Scenario (1) is the simplest on paper but leaves us with the same issues we have today
 - Offers little/no opportunity to reduce operational cost
 - Requires centers with little expertise or effort to run a complicated service (reliability)
 - Leaves the experiments to deal with a multitude of storage endpoints
- Scenarios (2) and (3) intend to reduce the storage operation cost, while presenting challenges
 - break explicitly data/CPU locality
 - require large data centers to take more load (operations and hardware)

Implications for DOMA

- The ACCESS working group is looking at defining a content delivery network model. This strongly depends on how storage evolves
 - Do we need caches, where, which kind, which problem should they solve?
- Implications for networks
 - Capacity at the level of LAN, NREN, global
 - Needs for dynamic network shaping?
- Implications for storage development projects
 - Improvements to support distributed instances
 - Developments driven by the QoS working group
 - Particularly relevant for DPM, where there is little funded effort today to support several of the discussed evolutions in DOMA

Inquiry

- As part of this process, WLCG Operations Coordination would send out a set of questions to sites/federations
- Not concretely discussed, but questions might be along those lines:
 - How much effort is required to support storage at the site?
What is storage capacity?
 - Which storage middleware implementation the site is using?
Is there a plan to migrate to other implementation in the coming years?
 - Would the site be interested or is the site being pushed to follow one of the scenarios described at slide 3? If yes, which one?
 - Would the site be interested to take part in the prototyping/testing activity?

Time to start this discussion

Your ideas, thoughts?