WLCG DOMA kick-off follow up

Maria Girone, Simone Campana
CERN

From Napoli: Next steps

- Create a Data Organization Management Access evolution project:
  - keep track of developments and advancements in all DOMA areas
  - provide a forum to discuss ideas
  - foster interoperability of solutions
  - an umbrella for stakeholders, national initiatives, EU projects, already existing working groups
- Stakeholders: experiments, middleware developers and storage providers, facilities
- Kick-off meeting held at CERN on June 4th and 5th https://indico.cern.ch/event/729930/

News from the LHCC

- The LHCC will start a review process of the WLCG strategy for HL-LHC
- The timescale for the review is early next year. This is the starting point for a several years process
- There will be regular checkpoints of the progress after that
- DOMA is a key aspect of the strategy and therefore will be a key area of the review process
- The progress we make in the next six months will be extremely valuable as input to the review
- We should be prepared to report regularly our progress

Facilities and Storage Consolidation

Different scenarios depending on the regions and funding model

- Some regions want to consolidate the storage in a geographically distributed instance
- Some centers will invest less in storage and more in CPU capacity
- Some centers will continue running large storage services and absorb some use cases from smaller facilities

Fewer storage endpoint and a large number of processing facilities, including HPCs and Clouds

Generally the experiments see the reduction of the number of storage endpoints as the right direction

- **Need to avoid however loss of efficiency and hardware**

DOMA activities on Storage services

- Define and implement Quality of Service in WLCG storage services
  Allows optimizing hardware cost and offers added value. Facilitates the integration of cloud storage
  - Define a common vocabulary and the correct granularity
  - Understand transition between QoS classes and QoS lifecycles
  - **Leverage existing initiatives like XDC**
- Share experience with different hardware resiliency solutions
  - RAID vs ERASURE in the LAN and WAN
- Implement a set of standardized tools for storage management
  Dealing e.g. with data loss, data unavailability, space reporting
  - Expose the information in a consumable way as first step
  - Evaluate the possibility of an automated system
- GDB and HEPIX are some obvious places for discussing progress

DOMA activities on Protocols

- SRM-free data management
  - No SRM needed for disk access
  - Currently developed tape solutions will not provide SRM (RAL, CTA@CERN)
- Xrootd as primary option for gridFTP replacement and data access.
  - Adopted by all experiments for data access
  - Used by Alice for 3rd party (gridFTP replacement) since ling time protocol for access/transfer today
- HTTP data access is an important solution for non HEP applications
- Proposal: WLCG operations takes the lead of protocols commissioning/decommissioning. Presents a plan, organizes the work.

DOMA activities on Data Access

Data Locality is not guaranteed: facilities w/o storage and facilities with consolidated storage but w/o local data

- Client level caching
  - TTreeCache for remote I/O
  - xCache on the WN for asynchronous pre-fetch
- Site Level Caching
  - xCache as site level cache
  - ARC CE + ARC cache (shared filesystem on the WN)
- **Areas of study:**
  - **Providing hints about file re-use to the cache**
  - **Pre-filling site caches**

- We need to leverage on the ongoing studies and organize future discussions on end-to-end performance

Caching and Latency-Hiding is a key element of the strategy at many centers, including CERN

Workflows and DOMA

Integrate the concepts of QoS and Caching into the workflows to optimize cost

- Integrate workflows with QoS
  - E.g. Tape Carousels. The archival Storage WG ("best practices for tape system usage and identify and report metrics") could be broadened to follow this
- Caching strategies
  - Understand access patterns. Which formats and which information is accessed?
  - Evaluate different caching options depending on the data and the workflow
- High I/O facilities
  - We might need to restrict high I/O workflows in dedicated facilities
  - Understand how this fits in the current WLCG facilities

Work tightly integrated with the WLCG performance and cost model WG

Next discussions

- AAI and DOMA
  - Discuss the evolution of AAI in the DOMA context
  - Federated identities, Auth Tokens and the HTTP transfer ecosystem prototype
- Networks
  - Now that we have a clearer idea of how facilities and DOMA will look like we should underatnd the implication from/to networks
- Other communities outside HEP
  - Understand their plans and how they fit into WLCG DOMA
- Planning for regular monthly meetings, possibly in conjunction with pre-GDB slots