Towards the R&D proposal

Plenary Session: Performance and efficiency of large data storage

Fabric Infrastructure

- What we want to avoid
  - Tape-related issues
    - Market trends SSD/HDD/Tape
    - Is it technically possible to completely abandon tapes?
    - Investigate solutions for long-term storage
    - Establish SuperB-related metrics for the “HSM caches”
- What we absolutely want
  - Investigate usage of WN disks (performance and applicability to apps/frameworks eg Hadoop/Proof/caching a-la-STAR)
  - Investigate simulation-based approach to define the minimum requirements for storage sites

User functionalities

- What we want to avoid
  - Unneeded complexity, unneeded “software layers”
  - Heavy “personal” setup phase
  - Differences between local run and batch/grid run.
  - Difficulties in analysis developing/debugging and final run
  - Heavy dependencies on central services to access local (and remote) data
- What we absolutely want
  - Interactivity + usage of personal resources for personal analysis tasks

Authorization/Authentication

- Define simplicity oriented requirements for
  - Authorization/Authentication
  - Access policies (eg r/o, n/a, groups, directories)
- Evaluate/design secure gateways/proxies to allow controlled access to/from computing farms

Integration with CM

- Investigate on dependencies between storage/data access tech and computing model E.g.
  - "prepare" requests (called pre-staging in the case of tapes)
  - Requests in advance of data chunks to allow WAN data access
  - Investigate common (readable) logical FN definitions
  - Investigate Extended Attrs and possibility to do searches (currently unavailable)

  - File/Replica location. Investigate on storage systems able to do it natively.
  - How many sites? Sites' roles? QoS differentiation and relationship with data type (eg RAW, user, etc)
  - Formulate simplicity-oriented requirements for data movements
    - Eg. BaBar-like datamovers? Native third party copies?
  - Metadata? Investigate how to avoid storage-dependent metadata (e.g. a file's location) in the Bookkeeping DB