Data Management in HEP
What next?

Predrag Buncic

ALICE Input

- With growing number of Grid sites and exploding data volume we must take steps to manage the complexity
  - 100 Storage Elements in the ALICE grid, 10^9 files in the existing File Catalog, 10^6 new files per day during data taking
  - We need to move from 100 SEs and Grid sites to 10 Cloud like regions in order to
    - simplify job scheduling
    - delegate the responsibility for custodial data storage to individual clouds
- This approach calls for a high level data federation
  - We recognize a need for a common scalable file namespace and agreed upon mechanism to keep such namespace up to date
- We think that we should stick to not more than 2 protocols for data access
  - Avoid gfal like approach
  - Use xrootd as primary (HEP specific) and HTTP as commonly available (failover) protocol
- EOS catalog on global scale for us would be a preferable starting point for such global namespace

Reducing complexity is the key

for illustration only

- Virtually joining together the sites based on proximity (latency) and network capacity into Regional Data Clouds
- Each cloud/region provides reliable data management and sufficient processing capability
  - Dealing with handful of clouds/regions instead of the individual sites