ATLAS ‘Data Carousel’ R&D

- Objective: use tape as input for I/O intensive workflows
  - Define realistic expectations with the current tape setup, evaluate possible evolutions
    - working with WLCG archival WG
- Complex topic, touching several areas:
  - ATLAS DDM (Rucio) --- improve tape usage, e.g. bulk mode, queue limits
  - FTS --- optimize scheduling of transfers involving tape endpoints
  - SE endpoints (dCache, Storm etc) --- any bottlenecks and possible improvements on interfacing with respective tape systems
  - Storing data to tape --- improve the way in how data is stored to tape, so to optimize their recall speed
    - Implication on the planning of what we want to retrive in optimal way.
  - In general: expectation management
  - ……

ATLAS ‘Data Carousel’ R&D: strategy

- Several phases of this R&D project
  - First phase: understand tape systems performance at the various sites
  - Second phase: address the issues of data retrieval/exchange between tape/disk
    - Deeper integration between two main distributed computing components, data management system and workflow management system
  - Third phase: exercise various workflows using tape at scale
- Starting point
  - Iterative test production campaigns --- main driver, to reveal bottlenecks of the current system and measure improvements of evolving system
  - Start with derivation workflow
    - Demanding workflow, in both data intensity and time to completion
  - Start with BNL --- run sample derivation tasks, examine tape performance, profile possible improvements.
    - and replicate the study at various Tier1s
- At every step, define metrics for success : turnaround time, failure rate, tails management, etc.