User Data on the Grid

Dario Barberis

CERN & Genoa University

User space on the Grid (1)

- Many discussions took place recently on this matter. Two extremes:
  - "Traditional" computing model:
    - All ATLAS VO members can write to all user space (ATLASUSERDISK token) wherever this token is defined
      - Usually Tier-2s
    - All this space is defined as "scratch" with a retention time of a week to a month
      - This retention time gives users time to decide if they wish to keep the data longer in their private site, promote them to group level and move them to a group area, or delete them
    - Problem: how to guarantee a minimum retention time? The SE may be flooded by other users that force all existing data out
    - Also, how to synchronize the SE with the catalogue(s)? How would the garbage collector work in practice? Would it be a central or local operation?
  - Recent proposal:
    - Allocate all ATLASUSERDISK space to (national or local) groups
      - Pro: no garbage collection, but if the SE gets full, people are readily findable
      - Contra: it breaks down the concept of the Grid and of equal access for all ATLAS members
      - Contra: it forces all analysis jobs to write their output to the SE where the user is authorized to write (lots of unnecessary data movement in real time)

User space on the Grid (2)

- My proposal:
  - Define the minimal need for user-managed (home) space on the Grid and allocate it at Tier-2s
    - Assuming everyone has at least one "friendly" Tier-2
      - Some ATLAS-wide agreement may be needed to make sure that this is actually true
    - There is additional user space at Tier-3s, which are not under central control
  - Set up all the rest of disk space at Tier-2s as a scratch area with garbage collection
    - Set up a garbage collector that cleans (for example):
      - All non-catalogued files older than a few days
      - All catalogued files older than a month (cleaning the catalogue too, of course)
        - Possibly sending a warning a week in advance
  - Set up distributed analysis tools so that they follow one of these work models:
    1) Jobs write the output to the local SE; either an automatic data transfer is triggered at the end of the job, or it is then the user's responsibility to recover the data or let it fade away
    2) Jobs write the output to the default SE of the user; if that fails for whichever reason, write the output to the local SE or have a list of failover SEs
  - Almost all tools to implement this proposal exist
    - Some may need some further development and tuning to make sure that we are not going to lose data