ATLAS Top 5 Worries

Dario Barberis

(CERN & Genoa University/INFN)

ATLAS Top 5 Worries

1) Castor @ CERN/Tier-0

2) Integration of DDM/FTS/SRM/GridFTP/LFC etc

3) (Lack of) SRM 2.2

4) Data Storage Management Tools

5) Stability of the Information System

1 - Castor @ CERN/Tier-0

- We see Tier-0 as an extension of the online computing farm
  - Therefore it must be reliable and shielded from possible interference from users' activities on the CAF
- The recent misadventures with Tier-0 internal tests and data export to the Tier-1s prompted multiple actions (Task Force):
  - Short term:
    - Implement improvements already in the pipeline
    - Split the ATLAS Tier-0 Castor instance from the ATLAS CAF instance
  - Longer term:
    - Re-think (parts of) the architecture…
- We are still unhappy at the current status, but we see that the problems are now addressed properly
  - The slope is positive…
- Nevertheless this remains our top worry.

2 - DDM/FTS/SRM/GridFTP/LFC etc.

- The worry is about the integration of all these components
  - Each software development project overruns by a factor π.
  - Each software integration project overruns by a factor π^2.
- Data distribution tests show that rates are not limited by hardware
  - But rather by configuration parameters in the many software layers
- Work in progress to understand this complex system and optimise these parameters is encouraging, but this is not really the task we are paid for!
  - In this multi-parameter space, it may take for ever to find the right configuration for each site pair

3 - (Lack of) SRM 2.2

- More than one year passed since the first SRM 2.2 discussion in Mumbai
  - The first implementation was promised for LAST September
- We rely on several new features in SRM 2.2 to build a robust data management system
  - In the meantime, we can survive, as long as:
    - We do not try to exercise reprocessing (with prestaging from tape)
    - Most data access is for production activities (few analysis users)
- SRM 2.2 was the highest priority until 2 months ago…
  - But it has been superseded in this role by the Castor problems!

4 - Data Storage Management Tools

- We need tools for data storage management:
  - Including quotas, accounting, tools for moving and deleting files
- Some accounting tools are in preparation
  - EGEE work on storage accounting
- Some other tools are part of the SRM 2.2 specs
  - File management tools
- Generally work in this area is much less advanced than in the job management area
  - Albeit equally (or perhaps even more) necessary
- As always we have interim solutions
  - Such as asking sites to provide separate disk pools for production and user data
    - Where a real quota system, coupled to VOMS groups/roles, would be needed on each site

5 - Stability of the Information System

- The BDII system has overgrown its original purpose
  - It stores information of very different type and scope:
    - Static info on site services setup, CEs, SEs etc.
    - Dynamic info on site availability, running jobs, free storage space etc.
  - This info is refreshed often but sometimes the refresh fails and a site disappears altogether
    - For a short time, but during this time the static info becomes unavailable too
- Each experiment has in addition its own view of the Grid
  - In ATLAS, the TiersOfATLAS file contains the Tier-1–Tier-2 associations, the LFC and FTS information and other useful stuff for ProdSys and DDM
    - This file has also overgrown its original purpose!
- We need a more structured approach to publishing this type of (useful and very much used) information