Storage consistency

Ensuring the consistency between the Storage Elements (SE) and the catalog(s) is a long standing issue of grid computing [1]. Two types of inconsistencies exist:
- Files in the catalog(s) but not physically on the SE: Lost Files
- Files on the SE but not registered in the catalog(s): Dark Data

We describe here the work done in the Distributed Data Management system of ATLAS (Rucio [2]) to automatically detect these inconsistencies and fix them.

Principle

Automatic consistency check is based on dumps comparison
- Each site is supposed to provide storage elements dumps on a regular basis (monthly or quarterly)
- Rucio dumps of the file replicas supposedly located at the site are generated every day
- We compare the Storage Element dump with the Rucio dumps at -X days and +X days

- The files in the two Rucio dumps but not in the SE dumps are Lost files
- The files in the SE dump but in none of the Rucio dump are Dark Data

Automation

- A daemon called auditor is doing the comparison of the dumps:
  - It automatically checks the dumps uploaded on the Storage Element for each ATLAS site, gets their creation date and downloads them
  - It gets the Rucio dumps for +X days stored on an Hadoop cluster
  - It runs the comparison between the 2 dumps. The comparison times scales from a few seconds for the small sites to a few hours for the biggest one (70M files)

- Another daemon collects the list of Dark Data and automatically deletes them

- The files identified as LOST are reported to the clouds support/sites for investigation. If confirmed LOST, another daemon is taking care of them
  - If another replica exists, there are automatically recovered
  - If no other replica exists, the file is removed from the datasets it belongs to, and the owner of the dataset as well as the ATLAS Metadata Information system of ATLAS (AMI [3]) is notified

Results

- A campaign was conducted to ask all the sites supporting ATLAS (~120 sites) to provide regular dumps. About 70% is doing it regularly as of today

Plot showing the evolution of the space after identifying and cleaning the Dark Data. The grey part represents Dark Data, the orange one data that must be kept (primary), the green one cached data (secondary)

- The consistency service is now running in production for more than 6 months. It allowed to identify and clean ~4 PB of Dark Data (out of ~200 PB of data registered)

References


Contact

https://rucio-ui.cern.ch
rucio-dev@cern.ch