

Data storage accounting at RAL

Tuesday, 7 March 2017 14:20 (20 minutes)

Accounting for storage available and used is important to many communities using grids (and clouds). Several large grids came together to define the GLUE schema (Grid Laboratory Uniform Environment) in order to promote the interoperation and cross-grid use of the infrastructures. Ensuring consistency (across implementations and grids) and usefulness of the information published by these grids was a significant task for every storage system, and the consistency exercise had to be repeated for the second version of GLUE, GLUE2. Creating the schema alone is not sufficient, so collaborations defined their “interpretations” of the schemas and guidelines for publishing metadata.

The present paper describes the work involved in developing and deploying such an information system for the CASTOR and CEPH systems at the RAL Tier 1. In part 1 of the paper, we outline some of the important design decisions taken throughout the development process and focus on how we obtain the required information from several disparate parts of the systems, and describe the difficulties associated with accounting when files can be compressed on tape, have ephemeral copies on disk, filesystem vs object store, etc. We finish part 1 with an outlook towards dynamic (“cloudy”) use of storage resources. In part 2, we look at how to format the information to become standards-compliant and future-proof.

Primary author: Mr APPLEYARD, Rob (STFC)

Co-authors: Mrs PACKER, Alison (STFC); Mr CANNING, Bruno (STFC); Mr KETLEY, Cheney (STFC); Dr JENSEN, Jens (STFC); Dr GORDON, John (STFC)

Presenter: Mr APPLEYARD, Rob (STFC)

Session Classification: Data Management & Big Data

Track Classification: Data Management & Big Data