

Data Management in EGEE

Thursday, March 26, 2009 2:40 PM (20 minutes)

Data management is one of the cornerstones in the distributed production computing environment that the EGEE project aims to provide for a e-Science infrastructure.

We have designed and implemented a set of services and client components, addressing the diverse requirements of all user communities. LHC experiments as main users will generate and distribute approximately 15 PB of data per year worldwide using this infrastructure. Another key user community, biomedical projects, have strict security requirements with less emphasis on the volume of data.

We maintain three service groups for grid data management: The Disk Pool Manager (DPM) Storage Element (with more than 100 instances deployed world-wide), the LCG File Catalogue (LFC) and the File Transfer Service (FTS) which sustains an aggregated transfer rate of 1.5GB/sec. They are complemented by individual client components and also tools which help coordinating more complex uses cases with multiple services (GFAL-client, lcg_util, hydra-cli).

In this paper we show how these services, keeping clean and standard interfaces among each other, can work together to cover the data flow and how they can be used as individual components to cover diverse requirements. We will also describe areas that we consider for further improvements, both for performance and functionality.

Primary author: FROHNER, Ákos (CERN)

Co-authors: SMITH, David (CERN); GROSDIDIER, Gilbert (LAL/IN2P3/CNRS); BAUD, Jean-Philippe (CERN); TEDESCO, Paolo (CERN); GARCIA RIOJA, Rosa Maria (CERN); MOLLON, Rémi (CERN)

Presenter: FROHNER, Ákos (CERN)

Session Classification: Grid Middleware and Networking Technologies

Track Classification: Grid Middleware and Networking Technologies