

Title:

How to access huge amounts of data on the grid?

Author:

Paul Millar, DESY

Abstract:

Providing a grid-wide storage service for huge amounts of data is challenging.

Work on analysing an experiment is traditionally centred at one point: where the experiment is physically located. The recent trend in the scientific research is that the research community working to analyse results from an experiment is increasingly widely geographically distributed.

For large experiments, including those associated with the LHC facility at CERN, the required resources now exceeds what an individual institute can provide. To analyse the experimental results, the research community must combine the resources of many institutes to form a single virtual resource. The challenge is to provide this virtual resource to the end-users scientist without them being aware on which institute the resources they are utilising reside.

These constraints are true for all grid resources (computational, storage, identity management, group membership, ...); however, it is a particularly tough challenge for storage. Storage involves shipping data between sites: high bandwidth usage that must be coordinated.

This talk will discuss how these problems are solved within the WLCG grid alliance. It will focus on the High-Energy Physics LHC usage and describe how the different software components fit together. It will also include some discuss on current trends in storage and some upcoming storage technologies, such as cloud computing.