



Contribution ID: 140

Type: **Poster**

## Mucura: your personal file repository in the cloud

*Tuesday 22 May 2012 13:30 (4h 45m)*

By aggregating the storage capacity of hundreds of sites around the world, distributed data-processing platforms such as the LHC computing grid offer solutions for transporting, storing and processing massive amounts of experimental data, addressing the requirements of virtual organizations as a whole. However, from our perspective, individual workflows require a higher level of flexibility, ease of use and extensibility, which are not yet fully satisfied by the deployed storage systems.

In this contribution we report on our experience building Mucura, a prototype of a software system for building cloud-based file repositories of extensible capacity. Intended for individual scientists, the system allows you to store, retrieve, organize and share your remote files from your personal computer, by using both command line and graphical user interfaces.

Designed with usability, scalability and operability in mind, it exposes web-based standard APIs for storing and retrieving files and is compatible with the authentication mechanisms used by the existing grid computing platforms. At the core of the system there are components for managing file metadata and for secure storage of the files' contents, both implemented on top of highly available, distributed, persistent, scalable key-value stores. A front-end component is responsible for user authentication and authorization and for handling requests from clients performing operations on the stored files.

We will present the selected open-source implementations for each component of the system and the integration work we have performed. In particular, we will present the rationale and findings of our exploration of key-value data stores as the central component of the system, as opposed to the usage of traditional networked file systems. We will also describe the pros and cons of our choices from the perspectives of both the end-user and the operator of the service. Finally, we will report on the feedback received from the early users and from the operators of the service.

This work is inspired not only by the increasing number of commercial services available nowadays to individuals for their personal storage needs (backup, file sharing, synchronization, ...) such as Amazon S3, Dropbox, SugarSync, bitcasa, etc., but also by several efforts in the same area in the academic and research worlds (NASA, SDSC, etc.). We are persuaded that the level of flexibility offered to individuals by this kind of systems adds value to the day-to-day work of scientists.

**Primary author:** Mr HERNANDEZ, Fabio (IN2P3/CNRS Computing Centre & IHEP Computing Centre)

**Co-authors:** DU, Ran (Chinese Academy of Sciences (CN)); Dr WU, Wenjing (Institute of High Energy Physics, Chinese Academy of Sciences (CN))

**Presenter:** Mr HERNANDEZ, Fabio (IN2P3/CNRS Computing Centre & IHEP Computing Centre)

**Session Classification:** Poster Session

**Track Classification:** Distributed Processing and Analysis on Grids and Clouds (track 3)