



Contribution ID: 21

Type: **not specified**

A Data Engine for Grid Science Gateways enabling easy transfer and data sharing

Thursday, 29 March 2012 16:40 (25 minutes)

Conclusions

The component will be deployed and tested in some SGs (e.g. DECIDE portal) before to be officially released. Future developments could allow direct data transfers from the user to the Grid storage elements and viceversa, i.e. without the Science Gateway “in the middle”, when the storage element interface will official support web protocols. This will widen the scope of the data engine to files of any size.

Overview (For the conference guide)

Grid infrastructures allow users to access and use computational and storage facilities distributed in different locations all around the world.

However, a real exploitation of such platforms by large communities is still not happening mainly because of the complex architecture of the Grid Security Infrastructure and of the command line based interface that turn out to be quite cumbersome for the vast majority non IT-expert users.

In this work we show a new SG component providing users with the ability to move data to storage elements and share them in an easy and clever way.

Impact

Science Gateways (SG) are recently emerging as high-level web environments that ease Grid access and use. However, although SGs simplify job management on the Grid, providing a web interface by which a job can be submitted with a few “clicks”, their capabilities to do an effective and intuitive data management are still at an early stage. Actually, Grid storage elements use dedicated protocols not supported by common desktop applications such as web browsers. This makes a smooth integration of data management services in a Science Gateway quite difficult.

Using the component described in this work is possible to build a new generation of SGs providing the full set of operations available on Grid, including the data movement which is

Description of the Work

The new SG component operates between the Grid storage and the user, providing a file-system-like view that implies a two-steps transfer through the Science Gateway.

Data are transferred in an easy way from users to the SG by means of a standard protocol, such as HTTPS, and then they are asynchronously moved to the Grid storage elements. Files are not immediately removed but they are kept in a cache in order to optimise the bandwidth by reducing multiple transfers of the same data.

File information, such as logical names, is managed by the data engine avoiding the use of Grid file catalogues and/or metadata servers. This is stored in a local DB for a more efficient management. Moreover, the DB is used to identify files ownership on the SG. In fact, from the middleware side the only owner of the files is the SG as a consequence of the shared robot-certificate proxy generated by the portal to perform Grid transactions. To avoid unwanted sharing of files, an additional access privileges flag is kept by the data engine. Users can access their own files only, unless they get that privilege from others. In this way, it is possible to share data among (sub-)groups of people inside the SG.

From a technical point of view, the data engine has been implemented as a pure Java library developed for the integration in a JSR-286-standard portlet. The library uses the jSAGA implementation of the SAGA standard to communicate with Grid services in a middleware-independent way. The web-based file-system-like user interface has been based on elFinder.

Primary authors: Dr FARGETTA, Marco (Consorzio COMETA); Mr ROTONDO, Riccardo (GARR); Prof. BARBERA, Roberto (INFN, Consorzio COMETA and University of Catania - Physics Department)

Presenter: Mr ROTONDO, Riccardo (GARR)

Session Classification: Portals & Gateways

Track Classification: Software services for users and communities