



Contribution ID: 77

Type: **Oral contribution**

The EGRID facility

Wednesday 1 March 2006 15:00 (15 minutes)

The EGRID project aims at implementing a national Italian facility for processing economic and financial data using computational grid technology. As such, it acts as the underlying fabric on top of which partner projects, more strictly focused on research in itself, develop end-user applications.

The first version of the EGRID infrastructure has been in operation since October 2004. It is based on European Data-Grid (EDG) and the Large Hadron Collider Computing Grid (LCG) middleware, and it is hosted as an independent Virtual Organization (VO) within INFN's grid.IT. Several temporary workarounds were implemented mainly to tackle privacy and security issues on data management: in these last few months the infrastructure was fully re-designed to better address them. The redesigned infrastructure makes use of several new tools: some are part of EDG/LCG/EGEE middleware, while some others were developed independently within EGRID. Moreover the EGRID project joined recently EGEE as pilot application in the field of finance, which means that the EGRID VO will be soon recognized on the full EGEE computational grid; this may impose some compatibility constraints because of the afore mentioned additions we make, which we will handle when the time comes.

The new infrastructure will be composed of various architectural layers that will take care of different aspects.

Security issue has been handled at the low middleware level that manages data: an implementation of the SRM (Storage Resource Manager) protocol is being completed where novel ideas have been applied, thereby breaking free from the limitations of current approaches. Indeed, the SRM standard is becoming widely used as a storage access interface and, hopefully, it will soon be available on the full EGEE infrastructure. The EGRID technical staff has an on-going long time collaboration with INFN/CNAF on the StoRM SRM server, with the intention to use this software for providing the kind of fine grained access control that the project demands. What StoRM does is to add appropriate permissions (using POSIX ACLs) to a file being requested by a user, and to remove them when the client is done with the file. Since permissions are granted on-the-fly, grid users can be mapped into pool accounts, and no special permission sets need to be enforced prior to grid usage. An important role is given to a secure web service (ECAR) built by EGRID to act as a bridge between the (resource-level) StoRM SRM server, and the (grid-level) central LFC logical filename catalog from EGEE that replaces the old RLS of EDG. The LFC natively implements POSIX-like ACLs on the logical file names; the StoRM server can thus read (via ECAR) the ACLs on the logical filename corresponding to a given physical file and grant or deny access to the local files, depending on the permissions on the LFC. This provides users with a consistent view of the files in grid storage.

At a higher level, in order to make even more transparent the usage of data in the grid, we also developed ELFI that allows grid resources to be accessed through the usual POSIX I/O interface. Since ELFI is a FUSE file-system implementation, grid resources are seen through a local mount-point so all the existing tools for

managing the file-system automatically apply: the classical command line, any graphical user interface such as Konqueror, etc. Programs too will only have to be interfaced with POSIX, thereby aiding in grid prototyping/porting of applications.

ELFI will be installed on all WN of the farm, so applications will no longer need to explicitly run file transfer commands but simply access them directly as though they were local. Moreover, ELFI will be able to fully communicate with StoRM, and it will be installed in the host where the portal resides thereby easing portal integration of SRM resources.

The new EGRID infrastructure can be accessed via a web portal, one of the most effective ways to provide an easy-to-use interface to a larger community of users: the portal will become the main interface for naive users.

The EGRID portal that is currently under development is based on P-grade, and inherits all the features already available there: still some parts must be enhanced to comply with our requirements. The P-grade technology was chosen because it seemed sufficiently sophisticated and mature to meet our needs.

However there are still missing functionalities important to EGRID. We are currently collaborating with the P-grade team in order to develop and integrate what we need:

Improved proxy management

Currently private key of the user must go through the portal, and then into the MyProxy server; we feel that for EGRID it should instead be uploaded directly from the client machine without passing through the server: this is needed to decrease security risks. To accomplish it we implemented a Java WebStart application which carries out the direct uploading. The application is seamlessly integrated into P-grade, through the standard "upload" button of the "certificates" portlet.

Data management portlet that uses ELFI

Currently P-grade does not support the SRM protocol and does not support browsing of files present in the machine hosting the portal itself. Since ELFI is our choice for accessing grid disk resources in general, including those managed through StoRM, a specific Portlet was written to browse and manipulate the file system present in the portal server itself. In fact ELFI allows grid resources to be seen as a local mount point as already mentioned it becomes easier to modify the portal for local operations rather than for some other grid service.

The Portlet allows manual transfer of files between different directories of the portal host, but since some of these directories are ELFI mount points then automatically a grid operation takes place behind the scenes. So what happens is a file movement between the portal server, remote storage and computing elements.

File management and job submission interaction

A new file management mechanism is needed besides that currently supporting "local" and "remote" files: similarly to the previous point what is required is "local on the portal server", since the portal host will have ELFI mount points allowing different grid resources to be seen as local to the portal host. In this way the workflow manager will be able to read/write input and output data through the SRM protocol.

Moreover, EGRID also needs a special version of job submission closely related to workflow jobs: what we call swarm jobs. These jobs are such that the application remains the same while the input data changes parametrically over several possible values; then a final job collects all results and makes some aggregate computation on them. At the moment the specification of each input parameter is done manually: an automatic mechanism is required.

Authors: TERPIN, Alessio (Egrid project ICTP); LETO, Angelo (Egrid project ICTP); MESSINA, Antonio (Egrid project ICTP); CORSO, Ezio (Egrid project ICTP); MURRI, Riccardo (Egrid project ICTP); DI MEO, Riccardo (Egrid project ICTP); Dr COZZINI, Stefano (CNR-INFN Democritos and ICTP)

Presenter: Dr COZZINI, Stefano (CNR-INFN Democritos and ICTP)

Session Classification: 1d: Computational Chemistry - Lattice QCD - Finance

Track Classification: Computational Chemistry - Lattice QCD - Finance