Contribution ID: **134**   Type: **oral presentation**

Comparative evaluation of tools providing access to different types of data resources exposed on the Grid

Friday 11 May 2007 09:20 (20 minutes)

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

The problem of managing and accessing huge datasets distributed across multiple sites
and stored into heterogeneous databases is common to several research areas.
We report on the comparative evaluation of three tools to access different types of
data resources exposed onto Grids: G-DSE
(http://wwwas.oats.inaf.it/grid/index.php?option=com_content&task=view&id=70&Itemid=88),
GRelC (http://grelc.unile.it)
and OGSA-DAI (http://www.ogsadai.org.uk/).

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

A test bed, which includes several sites in the Italian production Grid
infrastructure, Grid.it, has been set up. In each site the instances of the G-DSE,
GRelC and OGSA-DAI servers have been installed together with several instance of
their respective clients.
A set of databases provided by the bioinformatics and astrophysical communities were
exposed over the grid, some of them located on the same LAN as the access tool,
others are accessed remotely.
The test plan spans from very simple queries, use cases provided by the
bioinformatics and astrophysical communities, up to stress tests with multiple
queries submitted simultaneously from different sites.

With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possible, point out the experience limitations (both in terms of existing services or missing functionality)

All the three tools relay on the basic authentication mechanism provided by GSI to
grant the authorization for accessing the Grid exposed data.
However the Virtual Organization Membership Service, presently employed on the EGEE
infrastructure, provides detailed information on the user's relationship with the
Virtual Organization and could allow conditional access depending from the user
group, role and capability.
An evolution of the tools to become VOMS compliant is highly desirable.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

The Grid infrastructure has proven to be capable of providing storage and computing resources for the computational needs of the modern research. However eScience projects need also a way to access widespread databases within a computational grid environment, through a set of secure, interoperable and efficient data grid services. The evaluation test, reported here, addresses the needs of the bioinformatics community engaged, through the BioinfoGRID (http://www.bioinfogrid.eu/) and the LIBI (http://www.libi.it/) projects, in the adoption of a grid infrastructure layer at the base of their research activities and of the Astrophysical community of the INAF (Istituto Nazionale di Astrofisica) (http://www.inaf.it/) interested to access data in astronomical databases from the GRID,

The access to data from the Grid is also a crucial problem for the adoption of the grid technology to provide services in public administration (EGG project).

Primary author: MAGGI, Giorgio (INFN)

Co-authors: NEGRO, Alessandro (Università Lecce); GISEL, Andreas (CNR-ITB); GHISELLI, Antonia (INFN); PIERRO, Antonio (INFN); VUERLI, Claudio (INAF); AIFTIMIEI, Cristina (INFN); GIORGIO, Emidio (INFN); MANNA, F. (INAF); PASIAN, Fabio (INAF); DEL FREO, Federico (EGG project); DONVITO, Giacinto (INFN); ALOISIO, Giovanni (Università Lecce); TAFFONI, Giuliano (INAF); LA ROCCA, Giuseppe (INFN); ATUL, Jain (INFN+Politecnico Bari); CAROTA, Luciana (INFN); VERLATO, Marco (INFN); CAFARO, Massimo (Università Lecce); MAZZU-CATO, Mirco (INFN); BARBERA, Roberto (INFN+Università Catania); VADACCA, Salvatore (Università Lecce); FIORE, Sandro (Università Lecce); BARISANI, andrea (INAF)

Presenters: DONVITO, Giacinto (INFN); MAGGI, Giorgio (INFN)

Session Classification: Data Management

Track Classification: Data Management