

New results on a comparative evaluation of software providing access to different relational databases interfaced to the grid

Wednesday, February 13, 2008 3:00 PM (20 minutes)

A way to access widespread databases within a computational grid environment, through a set of secure, interoperable and efficient data grid services is very common in eScience projects.

Use cases in the bioinformatics and astrophysical communities span from very simple queries up to really stressing ones.

A stress tests has been set up exploiting the EGEE grid infrastructure by submitting jobs and monitoring them by means of the Job Submission Tool(webcms.ba.infn.it/cmssoftware/index.html/index.php/Main/JobSubmissionTool).

In this configuration it is possible to easily reach the server software and hardware limits and to test the software in a real environment, taking into account the different latency between the server and all the clients, to put in evidence the efficiency of each software tool in delivering the output.

1. Short overview

The problem of managing and accessing huge datasets distributed across multiple sites and stored interfaced into heterogeneous databases is common to several research areas.

We report on the comparative evaluation of four tools to access different types of data resources exposed onto Grids: G-DSE

(www.was.oats.inaf.it/grid/index.php?option=com_frontpage&Itemid=1),

GRelC (www.spaci.it/content.php?loc=projects&pg=prj.php&cat=gm&id=3), OGSA-DAI (www.ogsadai.org.uk/) and AMGA (amga.web.cern.ch/amg)

3. Impact

The evaluation test, reported here, addresses the needs of the bioinformatics community engaged, in the BioinfoGRID (www.bioinfoGRID.eu/) and the LIBI (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) (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).

These software could be integrated on the gLite grid infrastructure in order to add the possibility to access Relational Databases

4. Conclusions / Future plans

Each of the four tested software, shows some specific strength that can be helpful in some particular application environment. We will show this characteristics for each software and the final results obtained running the client in a widely distributed environment. We will highlight also the capability of each software to be integrated on the gLite infrastructure and on the work on-going in this field

Provide a set of generic keywords that define your contribution (e.g. Data Management, Workflows, High Energy Physics)

Data Management, Astrophysics, Bioinformatic, Relational Databases

Primary authors: Dr DONVITO, Giacinto (INFN-Bari); Prof. MAGGI, Giorgio (INFN-Politecnico)

Co-authors: NEGRO, Alessandro (Università Lecce); BARISANI, Andrea (INAF); GISEL, Andreas (CNR-ITB); VUERLI, Claudio (INAF); AIFTIMIEL, Cristina (INFN); MANNA, F. (INAF); PASIAN, Fabio (INAF); ALOISIO, Giovanni (Università Lecce); TAFFONI, Giuliano (INAF); ATUL, Jain (INFN+Politecnico Bari); CAROTA, Luciana (INFN); CAFARO, Massimo (Università Lecce); BARBERA, Roberto (INFN+Università Catania); VADACCA, Salvatore (Università Lecce); FIORE, Sandro (Università Lecce); CALANDRUCCHI, Tony (INFN)

Presenter: Dr DONVITO, Giacinto (INFN-Bari)

Session Classification: Data Management

Track Classification: Existing or Prospective Grid Services