



Contribution ID: 170

Type: Poster

Grid-enabling a Problem Solving Environment: Implementation and Everyday Use

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).

We have conveniently integrated Grid access capabilities within Jylab, a Problem Solving Environment (PSE) we designed. This PSE already provides interactive access to a suite of Java libraries targeting the applied scientist in general however also serving his need to compute with data and peer machines available over the Internet. Jylab can be dynamically installed at Grid nodes, thus being the execution environment for the scenario we tested (e.g. Web search engine functionality over the Grid)

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.

Grid-enabling and Grid-ifying (dynamic installation) of our PSE are mature enough for submitting jobs to Grid nodes (used mainly HellasGrid infrastructure). The applications tested are coded in Python (with the extra benefit of calling into Java libraries). E.g. we have suitably packaged Nutch search engine (<http://lucene.apache.org/nutch/>) with the scripts needed to crawl and index predefined small sets of URLs down to some linking depth from within Grid nodes; all indices remotely constructed are currently downloaded to the submission machine -not necessarily the User Interface machine, ssh tunnels can be transparently employed- and merged for subsequent visualization using Jung library (<http://jung.sourceforge.net/>) or queries using ready-to-use, servlet-based Web interface, all without leaving the comfort of the PSE environment. Grid interaction scripts are light-weight Python scripts wrapping the CLI (Command Line Interface) of standard available Grid toolkits (mostly gLite).

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

We consider the simplification of access to Grid resources for the applied scientist as a priority task. By Grid-enabling our PSE and also providing for its dynamic installation at Grid nodes we also propose a simple example of the implementation path that could be taken by more established PSEs (like Matlab, Scilab, etc) for actually exposing the Grid to applied scientists lacking the skills of a computer engineer, however in need of aggregate computational resources for their applications. In particular our Grid applications are taken from the field of Internet algorithmics - computing with data and metadata (e.g. link structure) - collected directly from the Web. We particularly follow the interesting pattern of exploiting the network bandwidth available

to Grid nodes for collecting this info in a distributed manner (by interactively submitting a set of Grid jobs from within our PSE) and then merging it for further processing (e.g. with numerical linear algebra kernels)

Authors: Prof. GALLOPOULOS, Efstratios (University of Patras); KOLLIAS, Giorgos (University of Patras); GEORGIOU, Konstantinos (University of Patras)

Presenter: KOLLIAS, Giorgos (University of Patras)

Track Classification: Demo and Poster session