Contribution ID: **127**   Type: **Oral Presentation**

Job Submission Tool, web interface and WebDAV data management

Monday, 11 April 2011 16:30 (30 minutes)

Overview

JST (Job Submission Tool) allows the exploitation of the Grid computing power to many research communities with highly compute intensive applications. The tool helps subdividings large applications in single independent tasks and execute them on the grid nodes in a optimized time. Furthermore with the implementation of 'ad hoc' graphic interface, users have the opportunity to use the grid without knowing all the technological details and without taking care of its complex authentication methods. So far, several bioinformatics applications have been successfully executed onto the grid using JST allowing the users to reach important goals on their researches.

Impact

Researchers often are not familiar with X509 certificates and grid technology: JST has proved since few years the capability to solve the problem of submitting huge challenges on the grid with a little grid competence and small human effort.

The first goal achieved by JST was to split application born to be executed on single machines in order to reach the same results executing single independent jobs in a parallelized way.

The JST team ported over the grid about 20 different bioinformatics application submitting around half a million of jobs for about 50 years of CPU.

After the release of the graphic interface more than 50 challenges have been submitted to grid through the web. A total of 359 days of CPU usage on the grid worker nodes has been reached for the completion of all the tasks. The submission with the GUI has produced 19656 jobs correctly executed and 583 failed jobs. Considering other similar tools that allows to use the grid, we have tried to completely mask the underneath technology trying to provide to the final users the chance to use a classic a simple portal that do not take care of the complex steps they should face using directly the grid. Furthermore we have overcome the limit that a classic portal could give for the input files dimensions setting up the new webdav server for JST that offer a new way to manage very big files. We decided to use WebDAV as it is a standard and well supported protocol, indeed each operative system as its own client (this is true for Linux, MacOS and Windows).

Description of the work

Several bioinformatics scientific questions often require months or even years to be solved using the computational power provided by a single cluster. In such cases it is important to exploit the grid in order to reduce the execution to few days or hours. The collaboration between bioinformatics researchers and grid expert has been found to be very productive: the grid expert takes care of splitting the application into independent tasks, filling the central DB with the independent task, submitting the tasks to the grid and retrieving the output. In other words the final user does not have to deal with all the grid technicalities which are taken by

the grid expert. Several important challenges have been successful executed whit this approach.

We have tried to go one step behind, trying to enable final users to use the grid on their own. For this purpose we have developed a tool, JST, which uses a web graphic interface, written using PhP, Javascript and XSLT, where a user can authenticate using username and password. Using the same graphic interface the user uploads the input files. Since it was found not feasible to upload big input files (up to 10GB) using the http protocol, a new mechanism based on the "webdav" protocol has been implemented to upload input files on a file server in an efficient way and with an automatic registration of the files on the grid Storage Elements. The user then customizes the application (changing the configuration parameters) and specify the name of the output files. After the submission of the web form, JST executes all the step required for the submission of the application to the grid (slicing of the problem, user authentication by means of a robot certificate, submission of the jobs to the grid, retrieving of the output files). At the end, JST notifies to the user the link of the output files by sending a mail.

Conclusions

The JST has shown a good reliability and efficiency and few improvements are needed. The development team is just working on the integration of the interface in other portals. Since JST has been developed inside the BioinfoGRID and LIBI project, so far all the work planned has been focused on the bioinformatics applications, but the generalization of the tool allows us to open our technology to other communities.

Primary author: Dr DONVITO, Giacinto (INFN)

Co-authors: Prof. MAGGI, Giorgio Pietro (INFN); CUSCELA, Guido (INFN)

Presenter: Dr DONVITO, Giacinto (INFN)

Session Classification: User Environments

Track Classification: User Environments - Portals