



Contribution ID: 156

Type: **Oral**

## **Job submission tool: bioinformatic application porting on the gLite grid Infrastructure.**

*Tuesday, 13 April 2010 11:30 (15 minutes)*

The Job Submitting Tool provides a solution for the submission of a large number of jobs to the grid in an unattended way. Indeed the tool is able to manage the grid submission, bookkeeping and resubmission of failed jobs.

It also allows the monitor in real time of the status of each job using the same framework. In this work we will introduce some key new features and application that we have added to this tool. In the work several already executed challenges will be reported together with a logical description

### **Detailed analysis**

In this work we will present the status of development of the Job Submission Tool both in terms of new functionalities and new applications ported within this framework.

The main new feature added are: automatic distribution of input data using standard gLite Data Management Tools, the possibility to exploit the dependency between tasks belonging to the same run, the possibility to create on the fly the needed task from both existing portals or application running over the grid (without any human intervention).

The Job Submission Tool is an easy and highly customizable framework that speeds up the process of porting new application to the grid glite infrastructure.

The applications submitted using this framework could be executed both using robot certificate or user certificates.

In the work we will present also the portal built on top of the framework in order to allow final users to submit large collection of jobs by filling up a small number of web form.

We will also describe the auto-adaptive submission algorithm that could guarantee to distribute the load of the calculation over all the farm available within the grid infrastructure.

### **Conclusions and Future Work**

This framework for submitting control and managing the jobs over the glite grid infrastructure could sensibly reduce the man power needed to run large challenges. With the new features it allows the final user to deal easily with large input files and a huge collection of output file in an unattended way. It was mainly tested with bioinformatics applications but it surely could work with other applications coming from different science, thanks to its flexibility and to the possibility to be customized to meet the end users needs both for serial and for workflow driven applications.

### **Impact**

Using this tool the end user could easily create a collection made up of a large number of jobs, where the complete run could take months to be executed also in a large grid infrastructure like EGEE.

In this case one needs to solve problems like: detecting application failures, resubmit the failed jobs, collect the output back. Moreover in some bioinformatics applications the input files are fairly large, and this could

easily become the main bottleneck if those files are available on one or a few sites. In this case it could be difficult to know where the files should be replicated as it may depend on where the CPU are available. In those cases it could be interesting to have a procedure to move the data where the CPU are free at a given time, and for the framework do this automatically without human intervention, in order to run on several farms, using data locally available.

Moreover as the number of the jobs needed to finish the whole run could easily be several thousands it is useful to be sure that the submission procedure does not overload a single or few sites and that the jobs are spread among the largest number of available sites.

## **Keywords**

Job Submission, Workflow, gLite, bioinformatics, Data Management

## **URL for further information**

<http://webcms.ba.infn.it/cms-software/index.html/index.php/Main/JobSubmissionTool>

**Primary authors:** DONVITO, Giacinto (INFN-Bari); MAGGI, Giorgio Pietro (INFN and Politecnico di Bari); CUSCELA, Guido (INFN-Bari)

**Presenter:** DONVITO, Giacinto (INFN-Bari)

**Session Classification:** Bioinformatics

**Track Classification:** Software services exploiting and/or extending grid middleware (gLite, ARC, UNICORE etc)