



Contribution ID: 113

Type: **not specified**

Exploiting grid resources for data simulation by using a general-purpose framework

Thursday, 29 March 2012 14:40 (25 minutes)

Description of the Work

We have developed a general purpose framework for executing data simulations in a distributed environment, which is effective in exploiting multi flavour Grid resources. Its design is based on a minimal set of standard Grid services and is capable to fit the requirements of many different Virtual Organizations (VO). At the moment, a centralized site provides the EGI services, a web portal and the metadata management tools.

The web portal allows the definition of VO-specific tasks, the so called 'session'. The job runtime environment, the job executable parameters, the Input and Output datasets, the enabling of distributed resources, for instance, can be customized to the VO needs through a session web interface. The framework can thus easily manage several sessions and VOs. A relational database system is used to store the session- related metadata; it acts as backend for all the sub-services and communicates via RESTful protocol with running jobs on remote sites.

The suite makes use of EGI grid services such as the WMS as job brokering service and Grid flavours interoperability element, the VOMS as authentication and accounting system, the LCG File Catalog and Utils, GANGA as job management system and SRM as protocol for data access.

The web portal provides a user-interface for job preparation and submission, for database interactions and monitor functionalities; it automatically configures to adhere to the VO-specific session customization. Jobs are grouped into requests by specifying the set of job parameters and submissions are performed either automatically to all the available sites by request or by a manual fine-grain selection of the job parameters and/or sites.

The job workflow is handled by a python script, which retrieves and executes the VO specific application, communicates status information to the bookkeeping database and transfers output files to a request-specific data repository. Location info are registered into the LFC service.

Conclusions

The suite has proven to be reliable and efficient although it still requires a careful initial configuration. At the moment, some VO specific constraints should be configured in the Web-interface; although this is a minor effort to cope with, in order to provide an agnostic tool, a completely user-configurable interface is under development.

Our suite can be seen as a light-weight general-experiment framework which focuses on basic functionalities, designed specifically for organizations that cannot afford the use of the more specialized HEP frameworks but that still require an easy-to-use interface to the Grid. Customization of the bookkeeping database, web portal, job executable and site requirements are the key points to achieve this goal as well as a small installation and configuration footprint.

Impact

The suite we developed ensures both a high degree of customization and usability and high performances in job submission and status retrieval.

The web portal has been developed in PHP with Smarty as the template engine to separate the logic layer from the presentation one and makes extensive use of Javascript for AJAX functionality. It is strictly bounded to the bookkeeping database in order to allow the job definition, submission and monitoring. It presents several sections, depending on how many VO-specific sessions have been defined in the configuration phase. It is important to stress that their content is dynamically generated from the bookkeeping database schema and state in order to include the session-specific fields.

The job submission scheduling is managed by the web portal itself taking into account the VO-specific load of remote sites and the status of the previous submission. The upcoming integration with the Nagios service availability monitor will strength the scheduling algorithm.

The authentication and authorization layer is based on GSI. It makes use of myproxy server to realize a VOMS proxy certificate authentication that allows job submission through the web portal and assures secure communication between jobs and the database. Web portal submission triggers the launch of a GANGA session, which in turn submits a python parametric-script to the remote sites. The latter is the effective job that will run on the remote site WNs.

The monitor functionalities are mainly based on the bookkeeping database through the information sent by jobs. This allows a real-time monitoring of job status and activity. A closer integration with the Grid LB will provide an additional source of information, especially in case of hanging jobs.

Our suite has been successfully tested with two superbvo.org sessions for Monte Carlo simulation. Tests involved more than 20 remote sites in Europe, USA and Canada, and three different Grid middlewares.

Overview (For the conference guide)

We present a general-purpose software framework, which allows different multi-disciplinary communities to take advantage of a distributed computational infrastructure. It has been designed specifically for organizations that cannot afford the adoption costs of more specialized and complex frameworks, but that still require a user-friendly, standard and highly customizable access to the Grid.

Our framework heavily relies on a bookkeeping database, storing both application-specific and infrastructure metadata, which is tightly coupled with a web portal. The first makes available to the users information on the execution status of jobs and their specific meaning and parameters, and contributes in orchestrating the submission mechanism. The latter provides job submission management, bookkeeping database interactions and monitoring functionality.

Primary authors: Dr ARMANDO, Fella (INFN Sezione di Pisa); Prof. ELEONORA, Luppi (University of Ferrara and INFN); Mr MATTEO, Manzali (University of Ferrara and INFN); Dr LUCA, Tomassetti (University of Ferrara and INFN); Mr ENRICO, Vianello (University of Ferrara and INFN)

Presenter: Dr LUCA, Tomassetti (University of Ferrara and INFN)

Session Classification: Portals & Gateways

Track Classification: Software services for users and communities