



Parametric study workflow support by P-GRADE portal and MOTEUR workflow enactor

Wednesday 1 March 2006 18:30 (20 minutes)

1. Composing and executing data-intensive workflows on the EGEE infrastructure

Grid computing is naturally very well suited for handling data-intensive applications involving the analysis of huge amounts of data. In many scientific areas the need for composing complex applications on grids from basic processing components has emerged. The classical task-based job description approach is providing a mean of depicting such applications but it becomes very tedious when trying to express complex application logics and large input data sets. Indeed, a different task needs to be described for each component and each input to consider. Higher level interfaces for easing the migration of applications to grid infrastructures are drastically needed. To ease the migration to grids of such complex and data intensive applications we are proposing a powerful tool which:

- Simplifies the application logic description through a graphical and intuitive editor.
- Enables the seamless integration of data intensive application running on different grid infrastructures.
- Permit try-and-retry experiments design and tuning through a flexible description and execution environment.
- Eases legacy code migration.
- Provides high level monitoring and trace analysis capabilities.

This tool is based on the integration of the PGRADE grid portal [1] and the MOTEUR workflow execution engine [2].

1. MOTEUR workflow execution engine

The service-based paradigm, plebiscited in the grid community, is elegantly enabling the composition of different application components through a common invocation interface. In addition, the service-based approach nicely decouples the description of processing logic (represented by services) and data to be processed (given as input parameters to these services). This is particularly important for describing the application logic independently from the experimental setting (the data to process).

MOTEUR is a service-based workflow enactor developed to efficiently process application workflows by exploiting the parallelism inherent to grid infrastructures. It is taking as input the application workflow description (expressed in Scuf language from the MyGrid project [3]) and the data sets to process. MOTEUR is orchestrating the execution of the application workflow by invoking asynchronously applications services. It takes care of processing dependencies and preserves the causality of computation on a highly distributed and heterogeneous environment.

Very complex data processing patterns may be described in a very compact way. In particular, the dot product (pairwise data composition) and cross product (all-to-all data composition) patterns from the Scuf language are very efficiently reducing

complex data-intensive application graphs into much simpler ones. They significantly enlarge the expressiveness of the workflow language.

In addition, MOTEUR enables all level of parallelism that can be exploited in a data-intensive workflow: workflow parallelism (inherent to the workflow topology), data parallelism (different input data can be processed independently in parallel), and services parallelism (different services processing different data are independent and can be executed in parallel). To our knowledge, MOTEUR is the first service-based workflow enactor implementing all these optimizations.

1. The PGRADE portal GUI

During the last few years the P-GRADE portal has been chosen as the official portal by several Globus and LCG-2 middleware based Grid projects around Europe. In its original concept the P-GRADE Portal supported the development and execution of job-oriented workflows by the Condor DAGMan workflow manager. While DAGMan is a robust scheduler to submit jobs and to transfer input-output files among grid resources, it uses a quite simple scheduling algorithm, it is not able to invoke Web/Grid services and it cannot exploit every possible level of application parallelism (e.g. pipelining).

To overcome these difficulties the P-GRADE portal has been integrated with the MOTEUR workflow manager. On top of that the P-GRADE Portal has been equipped with a universal interface by which it can be easily connected to other types of workflow engines. As a result every EGEE user community with its own application-specific scheduler can use the P-GRADE Portal to manage the execution of domain-specific programs on the connected Grids or VOs.

Based on the DAGMan and MOTEUR workflow managers the P-GRADE Portal supports the development and execution of stand-alone applications, parameter study applications and workflows composed from normal and/or parameter study components. These applications can be executed in LCG-2, Web services or Globus-based grids. During the execution the portal automatically selects the most appropriate plugged-in workflow manager to perform the scheduled submission of jobs, service invocation requests or data transfer processes.

The presentation introduces the capabilities of the MOTEUR-enabled P-GRADE Portal and the way in which the EGEE bioscience community is using it to solve a medical image processing problem. The community is going to develop a workflow of parameter study components that is capable to perform large number of operations on a huge set of medical images. The different components of the workflow represent Web services and are described by graphical notations. The MOTEUR workflow manager is responsible for the pipelined invocation of these Web services driven by the medical images and the different control input parameters.

[1] PGRADE portal, <http://www.lpds.sztaki.hu/pgportal>

[2] MOTEUR, http://www.i3s.unice.fr/_glatard/software.html

[3] UK eScience MyGrid project, <http://www.mygrid.org>

Summary

This presentation addresses two special new features of the P-GRADE portal:

1. Enabling the efficient parallel Grid execution of parametric study type of applications both at job level and at workflow level. More than that it enables the creation and execution of workflows where certain components of the workflow are parametric study applications themselves.
2. Enabling the plug-in of any user community specific workflow enactors.

The talk will present how the MOTEUR enactor plug-in is handled and used in the portal in order to support the above mentioned parametric study workflow execution.

Author: Mr SIPOS, Gergely (MTA SZTAKI)

Co-authors: Dr MONTAGNAT, Johan (CNRS, I3S laboratory); Prof. KACSUK, Peter (MTA SZTAKI); Mr GLATARD, Tristan (CNRS, I3S laboratory); Mr FARKAS, Zoltan (MTA SZTAKI)

Presenter: Mr SIPOS, Gergely (MTA SZTAKI)

Session Classification: Poster and Demo session + cocktail

Track Classification: Poster session