SUPPORTING WORKFLOW-LEVEL PARAMETER STUDY APPLICATIONS BY THE P-GRADE GRID PORTAL*  

Peter Kacsuk and Zoltan Farkas and Gergely Sipos and Gabor Hermann  
MTA SZTAKI, 1111 Kende utca 13  
Budapest, Hungary  
{kacsuk,zfarkas,sipos,gabor.hermann}@sztaki.hu  

Tamas Kiss  
University of Westminster  
Cavendish School of Informatics  
115 New Cavendish Street London W1W 6UW  
T.Kiss@westminster.ac.uk  

Abstract  
Workflow applications are frequently used in many production Grids. There is a natural need to run the same workflow with many different parameter sets. P-GRADE portal has been providing a high-level, graphical workflow development and execution environment for various Grids (EGEE, UK NGS, GIN VO, OSG, TeraGrid, etc.) built on second and third generation Grid technologies (GT2, LCG-2, GT4, gLite). Feedback from the user communities of the portal showed that parameter study support is highly needed and hence the recent release (2.5) of the portal supports the workflow-level parameter study applications. The current paper describes the semantics and implementation principles of managing and executing workflows as parameter studies. Special emphasis is on the generation of the parameter input files, concurrent management of large number of files and jobs as well as collecting the large number of results generated at the PS execution time.  

Keywords: Portal, Parameter Study, Workflow, Execution  

*This research work is carried out under the FP6 Network of Excellence CoreGRID funded by the European Commission (Contract IST-2002-004265) and under the SEE-GRID-2 project funded by the European Commission (Contract number: 031775).
1. Introduction

One of the most promising utilizations of Grid resources comes to life with parameter study (or “parameter sweep”) applications where the same application should be executed with a large set of input parameters. Such parameter study applications are easy to run in the Grid since executions started with different parameters are completely independent. Indeed, there are several projects [1],[2] that demonstrated that parameter study applications are easily manageable in the Grid. However, most of these projects tackled only single job based applications. The real challenge comes when complex applications consisting of large number of jobs connected into a workflow should be executed with many different parameter sets. There have been only two projects that tried to combine parameter studies with workflow-level support in the Grid. ILab [3],[4] enables the user to create a special parameter study oriented workflow. With the help of a sophisticated GUI, the user can explicitly define statically how to distribute and replicate the parameter files in the Grid and how many independent jobs are to be launched for each segment of the data files. The SEGL [5] approach puts much more emphasis on exploring the dynamic nature of the Grid. They also provide a GUI to define the workflows and to hide the low level details of the underlying Grid. The SEGL workflow provides tools for several levels of parameterization, repeated processing, handling conclusions and branches during the processing as well as synchronization of parallel branches and processes. The problem with this GUI is that it might be too sophisticated, requiring very large skill from the application developer.

Although our approach to support workflow-level parameter study applications in the Grid has many similarities with these two projects, there are significant differences, too. Our main goals are as follows:

1. Keep both the workflow GUI and the parameter study support concept as simple as possible. This enables the fast learning and easy usage.
2. Enable run any existing workflow with different parameter sets without modifying the structure of the workflow.
3. Provide an easy-to-use way of generating the various input parameter files needed for the PS execution.
4. Manage the execution of the workflows on as many Grid resources as possible. Enable the collection of Grid resources from several Grids.
5. Enable the access of the workflow-oriented GUI and the available Grids via a Grid portal without installing any software on the user’s machine.
6. Provide an easy-to-use way of collecting and processing the large number of result files generated during the PS execution.