CEGA: A Workflow PSE for Computational Applications¹

Yoonhee Kim

Dept. of Computer Science, Sookmyung Women's University, Korea
yulan@sookmyung.ac.kr

Abstract. The importance of problem solving environment (PSE)s over Grid has been emphasized owing to the heterogeneity and the large volume of resources involved and the complexity of computational applications. This paper proposes a Grid-enabled PSE called as Computing Environment for Grid Applications (CEGA) and discusses how it is evolving to develop a computational application in a style of workflow model and incorporates Grid computing services to extend its range of services and handle information for development, deployment, execution and maintenance for an application as well as an application requirement itself. The paper describes the architecture of CEGA and its implementation for development, execution and visualization of an application.

1 Introduction

The concept of Grid computing has been investigated and developed to enlarge the concept of distributed computing environment to create infrastructure that enables integrated services for resource scheduling, data delivery, authentication, delegation, information service, management and other related issues [1]. As the Grid provides integrated infrastructure for solving problems, interfacing services such as web portal to access Grid services, PSEs (Problem Solving Environments) have been developed to improve the collaboration among Grid services and reduce significantly the time and effort required to develop, run, and experiment with large scale Grid applications.

However, most PSEs to support parallel and distributed computing focus on providing environments for successful execution of applications and providing reasonable resource scheduling schemes. There have been several application-specific tools and PSEs to utilize Grid environment efficiently. ASC Grid Portal [2] is a PSE for large-scale simulation in astrophysics. Hotpage [3] is another PSE targeted toward high performance computing applications. Cactus [4] provides a problem solving environment for developing large-scale distributed scientific applications. GrADS [5] is a toolkit to help users to build applications over heterogeneous resources with ease of use. Similarly, UNICORE [6] provides graphical user interface to access heterogeneous resources uniformly. However, an effort on generalizing a PSE to support development and execution of applications (i.e. applications in workflow management), has been not fully investigated.

¹ This Research was supported by the Sookmyung Women's University Research Grants 2003.

A Computing Environment for a Grid Application (CEGA) has been developed to provide a computing environment for a computational application in a workflow model. The CEGA provides an efficient Graphical User Interface (GUI) approach for developing, running, evaluating and visualizing large-scale parallel and distributed applications that utilize computing resources connected by local and/or wide area network. To support Grid services through CEGA, the server creates application configuration using Resource Specification Language (RSL), which runs over the Globus toolkit [7].

The organization of the remaining sections of the paper is as follows. I present an overview of the CEGA architecture in Section 2. In Section 3 I describe the major functionality in workflow editor for development and visualization in detail. The conclusion is followed in Section 5.

2 The Architecture of CEGA

The CEGA is a workflow based problem solving environment for grid computational applications. It provides transparent computing and communication services for large scale parallel and distributed applications. The architecture of CEGA consists of CEGA portal services, CEGA system serves and Globus (see Fig.1) Globus is the Grid middleware toolkit to launch jobs and collects their results over Grid resources. CEGA Portal Service help users to generate a workflow of an application, based on the application logic, specify the input/output and runtime requirements to provide various execution environments without having big efforts for changing their experiment options. In addition, the status of job executions is reported back to a user using visualization tools, which are attached to the editor. To interface between the Portal and Globus, System Services collect and handle data from the Editor and Globus services, steer the workflow over Globus. In addition, it monitors the status of job execution. That is, CEGA provides two important services over Grid services: 1) Problem Solving Environment: to assist in the development of large scale parallel and distributed applications; and 2) Evaluation Tool: to analyze the performance of parallel and distributed applications under different resource environments.

3 CEGA Services

3.1 CEGA Editing Service

Graphical user interface helps users to develop, modify and execute an application easily over Grid resources. CEGA editor provides workflow patterns which includes various job patterns and linkages. User can use customized workflow patterns which a user can directly modify on; as well as built-in patterns. As Object List Tree is added in the left side of the window, new object is easily added to the graph with just drag & drop scheme. For editing service, open, save, new, execute and monitor interface is developed on the top menu. Once an application is developed within the editor, the information is passed to the Workflow Engine for workflow management to proceed application execution.

3.2 CEGA Execution Service

An application is defined as a set of jobs; which are connected to one another based on their job dependency. The execution of jobs can be executed either in parallel or sequence based on the dependency. When Job Control service collects application workflow information from the editor, it controls the order of execution of the jobs based on the dependency over Grid environment as the results of analysis of the workflow information. The service generates a set of activities in Resource Specification language (RSL), one of which can run over Globus in parallel. The deployment of the data in appropriate locations among the execution of activities is done by the control of the service based on the workflow. The system service checks the job dependency on the graph and divides multiple phases, which means making execution groups of jobs in parallel. Each group is mapped to grid resources by means of job scheduling service in Globus. Beside of job steering, it provides cooperative administration for dynamic Globus environment such as administrating Grid nodes and coordinating other Grid services, GRAM, MDS, and FTP as examples. The monitoring service also provides runtime monitoring and collects runtime log data in XML including starting and ending time, and total execution time. Whenever a user wants to monitor the runtime data, the Engine provides them to a visualization tool.

Fig. 1. The Architecture of CEGA

3.3 CEGA Monitoring and Visualization Service

The Workflow Engine controls and monitors the application execution based on the application workflow. A user is informed as the Engine provides the status data to the Editor in real time. The monitoring information is classified as two types: one is parsing status caused by syntax error before execution and the other is execution status. The information of monitoring syntax error and job execution result is displayed on the bottom window in text. At the same time, the color of each job model shows a

current status of job execution in 6 states: ready, active, done, failed, suspended, and pending. And when the execution is successfully done, the result of the application is stored in the location requested at the beginning of the application development. The visualization of the result of the application is dependent of the characteristics of the application. CEGA can basically provide the result in text mode and choices of some plugged in graphic tools (See Fig.1).

Currently, CEGA has been developed for a potential energy surface calculation in molecular dynamics simulation, which requires high performance parallel computing. It also provide interface to customize experimental options easily for diverse execution conditions. It reduces deployment cost for experimentation.

4 Conclusions

The paper describes the architecture of CEGA to provide a parallel and distributed programming environments; it provides an easy-to-use graphical user interface that allows users to develop, run and visualize parallel and distributed applications running on heterogeneous computing resources connected by networks. To support Grid services through CEGA the server creates application configurations using Resource Specification Language (RSL), which runs over the Globus toolkit. This paper shows the functionality of the workflow editor to develop an application and generate application configuration over Grid environment. As it is adapt a workflow model, it can be easily extended to other application domain when diverse workflow patterns are added. Currently, the extended workflow models are under development.

References

1. Foster, C. Kesselman, "The Grid:Blueprint for a New Computing Infrastructure," Morgan-Kaufmann, 1998.
2. Astrophysics Simulation Collabratory: ASC Grid Portal, <http://www.ascportal.org>
3. HotPage, <http://hotpage.npaci.edu>
4. Cactus Code, <http://www.cactuscode.org>
5. GrADS Project, <http://nhse2.cs.rice.edu/grads/index.html>
6. Romberg,M., "The UNICORE Architecture Seamless Access to Distributed Resources," High Performance Distributed Computing, 1999
7. Globus Project, <http://www.globus.org>
8. J. Novotny, "The Grid Portal Development Kit," Concurrency: Practice and Experience, Vol.00, pp1-7, 2000