Principal Researcher & email	Farrukh Nadeem. abdullatif@kau.edu.sa
Co-researchers & emails	Thomas Fahringer
Affiliation	National University of Computers and Emerging Sciences, Lahore, Pakistan; University of Innsbruck, Austria.
Research Title	Using Templates to Predict Execution Time of Scientific Workflow Applications in the Grid
Research Topic	Performance Modeling of Scientific Workflow Applications
Publisher	IEEE Computer Society
Publishing Year	2009
ISBN	978-0-7695-3622-4
Key Words	Scientific workflow applications, Performance modeling, Grid
Journal Name, or (Conference + place and date being held)	CCGRID '09 Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, May 18-May 21, Shanghai, China.
Volume No.+ Issue No. and the Number of Pages in case it has been published in a scientific journal	pages = 316-323
Research Abstract	Workflow execution time predictions for Grid infrastructures is of critical importance for optimized workflow executions, advance reservations of resources, and overhead analysis. Predicting workflow execution time is complex due to multeity of workflow structures, involvement of several Grid resources in workflow execution, complex dependencies of workflow activities and dynamic behavior of the Grid. In this paper we present an online workflow execution time prediction system exploiting similarity templates. The workflows are characterized considering the attributes describing their performance at different Grid infrastructural levels. A “supervised exhaustive search” is employed to find suitable templates. We also make a provision of including expert user knowledge about the workflow performance in the procession of our methods. Results for three real world applications are presented to show the effectiveness of our approach.

Using Templates to Predict Execution Time of Scientific Workflow Applications in the Grid

Farrukh Nadeem, Thomas Fahringer
Institute of Computer Science, University of Innsbruck, Austria
farrukh@dps.uibk.ac.at

Abstract

Workflow execution time predictions for Grid infrastructures is of critical importance for optimized workflow executions, advance reservations of resources, and overhead analysis. Predicting workflow execution time is complex due to multeity of workflow structures, involvement of several Grid resources in workflow execution, complex dependencies of workflow activities and dynamic behavior of the Grid. In this paper we present an online workflow execution time prediction system exploiting similarity templates. The workflows are characterized considering the attributes describing their performance at different Grid infrastructural levels. A "supervised exhaustive search" is employed to find suitable templates. We also make a provision of including expert user knowledge about the workflow performance in the procession of our methods. Results for three real world applications are presented to show the effectiveness of our approach.

1 Introduction

Workflow applications from scientific and business domains typically consist of several different tasks (activities) with (complex) execution dependencies among them. The execution of such workflows in an automatic fashion, usually accomplished through Grid application development and runtime environment like ASKALON [3], requires to make several dynamic decisions (like resource selection, scheduling workflow activities etc.) regarding their optimized executions. While the performance can be gained by intelligently planning the execution of workflow activities in the Grid, there are also several overheads associated with them [12], which degrades the performance. An online workflow performance prediction service provides a decisive base for such optimization decisions.

Predicting the execution time of a workflow application is a complex and has been ignored so far due to involvement of several Grid resources and their inherent architectural and functional heterogeneity. In addition, a large variety of structures of the workflows and complex control flow and dataflow dependencies among their activities [13], external load and dynamic behavior of the Grid make the problem even more challenging, and debar the classical prediction methods at outset.

In this paper, we present an online workflow application execution time prediction system exploiting similarity templates. A *Template* refers to a set of selected workflow attributes. Workflow attributes correspond to the features describing workflow execution. We parameterize workflow application execution in terms of its attributes defining its structure and execution (Section 3), and use similarity templates to deliver the execution time predictions (Section 4). A comprehensive search method *Supervised Exhaustive Search* (Section 4.1) has been employed to define and search suitable templates. Supervised exhaustive search iteratively generates templates by selecting different workflow attributes and selects the best of them. The selection of workflow attributes in the templates is supported by the supervision of expert user through providing probabilities for the selection of different attributes. We define relationships among different workflow attributes and exploit these relationships during the search for suitable templates. Results are presented (Section 5) from a prototype development of our system, which has been integrated as a Grid service in ASKALON. Please note that in the course of this paper, by *workflow performance prediction* we mean execution time prediction of the entire workflow.

2 Scientific Workflow Applications

The workflow is a common way to model compound applications as a graph of activities, dependencies among them. Dependencies among the activities define their execution order and the dataflow from one activity to another. An activity in a workflow may be an executable to be executed or a service to be invoked.Usually, a Directed Acyclic Graph (DAG) is used to represent a simple workflow, however, additional constructs are used to show loops and con-