Scientific and Administrative Report

Manuel Caeiro-Rodríguez

March 21, 2008

1 Summary

This is the report of my stay at Rennes (France) as a CoreGRID PostDoc since December 2007 to March 2008.

The work I have performed during these four mounths here in Rennes has been focused on Goals 1 and 2 of the initial research plan. These goals are related with the identification of scenarios requiring dynamic composition and execution of workflows, and the identification of main issues to support decentralized autonomous enactment of workflows, respectively. Eventually, these goals have been focused specifically in the support of dynamicity in scientific workflows.

In relation with the activities and outcomes proposed I have already performed the three first ones and I have developed an initial proposal towards the fourth one. Activity A, related with the chemical paradigm, was performed by reading the available literature and conversations with partners. Activity B and C, related with the search of dynamicity scenarios and study of existing solutions, have been performed through a extensive and intensive review of the literature. A initial proposal has to be depicted in order to attain Activity D, related with the development of an HOCL scientific workflow solution.

2 Publications

Two documents have been created during this four month period.

1. A report entitled as *"Dynamicity in Scientific Workflows"*.

Abstract

Dynamicity is a recurrent topic in traditional workflow systems. The need and feasibility to perform changes in workflow process instances while they are being executed has been a main (and to a long extend yet unsolved) challenge. More recently, the scientific workflow domain has also paid attention to this topic and some of the current scientific workflow management systems give a certain support for dynamism. In general, there is a common agreement that dynamicity is an intrinsic requirement for scientific workflows, but the understanding about the real needs and functionalities to be provided is confuse. This report is mainly focused on contributing to enhance such an understanding by analysing dynamicity scenarios, requirements and proposals in scientific workflows. First, five general

scenarios involving different dynamicity needs are described introducing concrete examples. Then, these scenarios are used to identify a set of dynamicity requirements for scientific workflows support. Finally, a review of current well-known scientific workflow execution systems is presented, focusing on their proposals to support dynamicity.

2. A paper has been accepted for presentation in the Mynisimposia "Workflow Abstractions for Parallel and Distributed Computing in e-Science" in the PARA 2008 conference (http://para08.idi.ntnu.no/index.php?page=minisymposia#Workflow-eScience) entitled *A Proposal to Support the Execution of Scientific Workflows based on a Higher Order Chemical Language*.

Abstract

Scientific workflows are being developed to support experimentation in an increased number of fields (e.g., astronomy, biology, nuclear physics, climate, geology). Such workflows use to involve the management of large amounts of data and the performance of computational-intensive tasks, requiring the use of large amounts of resources. During the last years, several execution systems have been developed. These systems perform the distribution of workflows' tasks and data in resources, control and manage their operation in accordance with the prescriptions. For them, a main point is to obtain a good performance from the numerous resources involved through appropriate parallelization, trying to get as less idle resources as possible. In addition, these execution systems need to take care of several important requirements such as dynamic changes, fault-tolerance, provenance, steering, etc. This paper will introduce an initial proposal for a scientific workflow execution system based on a /Higher Order Chemical Language /(HOCL) approach. This language provides a natural parallelization basics that enables to conceive a kind of execution system different from the traditional ones. Instead the common /push-oriented/ behavior, where the central execution system distributes the tasks and data among the resources, the solution proposed follows a /pull-oriented /mode, requiring a more active role for the resources. In this way, the solution is very flexible and scalable, enabling the achievement of the numerous requirements involved in the execution of scientific workflows.

3 Travels

I have performed three travels during my stay in Rennes:

1. To Budapest, Hungary (27/01/2008 to 30/01/2008). A meeting with Zsolt Nemeth and Thierry Priol to review the work performed and take decisions about the next steps.

2. To Sophia Antipolis, France (14/02/2008 to 16/02/2002). A meeting with Johan Montagnat about scientific workflows and dynamicity needs.

3. To Madrid, Spain (20/02/2008 to 22/02/2002). To receive the award of the Spanish Chapter of the IEEE Education Society to the best 2007 PhD thesis.