

Preface

Dynamic computational workflows: Discovery, optimisation and scheduling

David Walker^a and Ewa Deelman^b

^a*School of Computer Science, Cardiff University, Cardiff, CF24 3AA, United Kingdom*

^b*USC Information Sciences Institute, Marina del Rey, CA 90292, USA*

The Grid computing community is converging on a service-oriented architecture in which applications are composed from geographically-distributed, interacting web services, and are expressed in a workflow description language, usually based on XML. Such workflows are viewed as offering a useful representation of service-based applications or applications composed of standalone components that are to be run in a distributed environment. A workflow can be conveniently displayed as a directed graph in which nodes represent services and edges represent the flow of data and/or control between them. In some sense workflows represent “how the Grid is programmed”, and consequently much recent research has focussed on software and tools that support the use of workflows for grid computing. The semantic description of services, components and workflows, service and resource discovery and selection, and the scheduling of workflows present significant research challenges in making workflows of practical benefit to grid users. The solution of these challenges, in turn, depends on languages for workflow representation, enactment engines for workflow execution, and ontologies.

This special issue arose mainly out of two international workshops:

1. The Workshop on Workflow Optimization in Distributed Environments, held 19–20 October 2006, at the e-Science Institute, Edinburgh, UK.
2. The Second Workshop on Workflows in Support of Large-Scale Science, held 25 June 2007, in Monterey Bay, California, USA.

In addition to selected papers from these two workshops, the special issue also contains papers submitted in response to an open call for papers. The resulting six papers published here provide a snapshot of the broad range of workflow-related issues currently being investigated by researchers.

Two of the papers discuss mature workflow management systems, and both emphasize realtime techniques for supporting the efficient execution of scientific workflows in grid environments. The first of these, by Prodan, describes workflow support in the ASKALON environment. The paper presents an abstract hierarchical model for workflows, and then goes on to show how a performance prediction service and a scheduling service can interact with each other, and with a workflow enactment engine, to provide for the efficient, scaleable execution of large-scale scientific workflows. The second paper, by McGough et al., focuses on the techniques, such as workflow restructuring, used by the GridCC realtime workflow management system to ensure that pre-defined quality-of-service constraints are met when executing workflows in a grid environment. The paper by Walker et al. is also concerned with the optimization of workflow execution. This paper describes how a generic proxy service is used to discover and select from among a set of semantically equivalent service instances, based on near-realtime performance data. Experimental results are presented showing the effect of dynamic service discovery and selection on workflow makespan, together with a best-case statistical analysis. Whereas the papers discussed above main-

ly consider the optimization of workflows with respect to execution time, the paper by Singh et al. addresses the issue of scheduling large-scale, data-intensive workflows that require large amounts of storage. The paper describes how such workflows can be restructured to reduce their data footprint, and illustrates this with some realtime distributed scientific applications. In the paper by Al-Shakarchi et al., a music recommendation system is described that makes use of cooper-

ative caching on a peer-to-peer (P2P) network. This requires the distribution of Triana workflows across the P2P network, with the results being stored in a distributed database and used to help users in discovering and selecting music. The final paper by Bahsi, Ceyhan and Kosar reviews the most widely-used workflow management systems, paying particular attention to the support they provide for conditional execution in workflows.