

Workflow Advisor on The Grid

M. Ong, X. Ren, G. Allan, V. Kadiramanathan, HA Thompson and PJ Fleming

University of Sheffield,
Department of Automatic Control and Systems Engineering,
Mappin Street, Sheffield S1 3JD, United Kingdom.

Email: {M.Ong, X.Ren, J.Allan, Visakan, H.Thompson, P.Fleming}@sheffield.ac.uk

Abstract. There are an increasingly large number of web services available today. New services are constantly being introduced, each performing different tasks ranging from simple data transfer operations to highly complex engineering and mathematical processes. The existence of description languages, workflow tools, semantics, service publishing methods and Grid middleware have up to now enabled interoperability between resources available across the Grid environment. However, there is an emerging need for the ability to dynamically discover how available services, resources and data could be utilised not only to process a task, but to achieve the desired outcome in the most suitable manner. This paper investigates the emerging need for workflow advice to aid decision support on the Grid and proposes a workflow advice system that incorporates a reasoning system designed to perform such a task. Capturing domain knowledge and evaluation of workflow results is the key to creating suitable new workflows for new tasks. Thus the key challenges are to provide the best advice on constructing complex workflows using the available Grid-enabled resources to complete a desired task, and the underlying knowledge capture and knowledge-based search across complex workflow information to support that. The case for a Workflow Advisor system is explored within the context of advanced aero engine fault diagnostics in the UK e-Science Grid project, DAME. The use of Case-Based Reasoning (CBR) is explored as the underlying technology for intelligent information retrieval and matching. Open standards and interoperability are regarded as highly important issues in ensuring the design and implementation of a workflow advice system that will interact seamlessly across the heterogeneous Grid environment.

1. Introduction

There are an increasingly large number of web services available today. New services are constantly being introduced, each performing different tasks ranging from simple data transfer operations to highly complex engineering and mathematical processes. Each service may require a multitude of unique inputs and may produce outputs varying from simple images to complex data structures. While the Web Service Description Language (WSDL) has been introduced to define the interfaces to such services, workflow tools have also been constantly developed to manage semantics and interoperability between the services. In addition to this, web service publishing and service discovery methods are well underway in the development of Grid middleware which will enable users to dynamically locate and invoke the functions provided by any desired service.

However, there is an emerging need for the ability to dynamically discover how available services, resources and data can be utilised in a suitable manner to achieve a desired task.

It is clear that the way in which a user selects individual services and composes them into a string of processes within a workflow is very important in order to perform a task correctly. This can also affect the time taken to complete a job, and how well the available resources are utilised. A prime example is if a particular resource is scarce, expensive, time-consuming or not readily available, the user may choose to utilise another resource in a different manner to achieve similar end results. It is clear that there may exist more than one way to perform a task which may provide similar or different results. Here it would be ideal to be able to explore every possibility and compare all end results. However, given the time-constraints, human-resource factors and costs

involved in a real-world application, it would be highly advantageous to have an intelligent system that could operate at a higher level to provide the user with knowledge on how available services could be best used to execute a desired task beforehand.

evaluation of workflow results is the key to creating new workflows for new tasks that appear similar to previous tasks (Fig. 1). Thus the key challenges are to provide the best advice on constructing complex workflows using the available Grid-enabled resources to complete a desired task, and the underlying

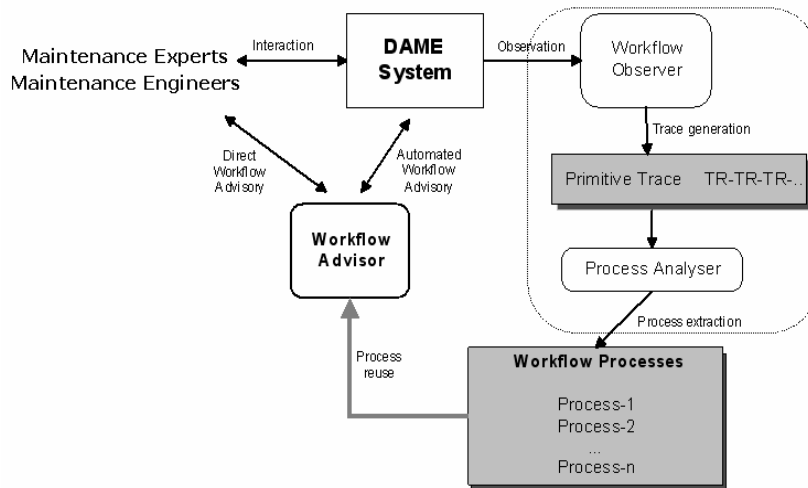


Fig. 1: General framework for a knowledge-based approach to the Workflow Advisor in DAME, leading from the observation level to the experience reuse level

2. Workflow Advise in DAME

The Rolls-Royce supported University Technology Centre (UTC) in the Department of Automatic Control and Systems Engineering at the University of Sheffield is currently engaged within the UK e-Science Grid project, DAME [1], contributing expertise in decision support, modelling, simulation and workflow advice. The case for Workflow advice is being explored within the context of aero engine fault diagnostics.

The work at the Sheffield UTC investigates the emerging need for workflow advice to aid decision support on the Grid and proposes a workflow advice system that incorporates a reasoning system designed to perform such a task [2]. Given that information is available from current and past workflows, it is highly feasible to build a repository of workflow history. Capturing domain knowledge and

knowledge capture and knowledge-based search across complex workflow information to support that.

Based on previous workflow information, an aircraft maintenance engineer may request for advice on which set of Grid-enabled tools and resources to use, and more importantly how to use them to analyse large datasets downloaded from aircraft engines. Upon analysis, further tools and services on the Grid will be used to identify any abnormal conditions and recommend a suitable maintenance action to rectify the abnormal condition based on the results of the previous steps. These tools are to be composed into a suitable workflow process and sent for execution across high performance Grid systems [3].

The task of providing workflow advice is a two step approach. The first step is the semantic service description approach involving the use of conceptual links between services and their properties [4]. This utilises available information about service profiles such as the preconditions, constraints and outputs of the service in order to assess the potential fit of each service to a desired task in the workflow. From that point on, it is feasible to automatically retrieve a service that matches the semantic description. External agents can use the result of such a reasoning

engine to select a service in line with their information processing goals. Here, the use of ontologies for describing the diagnostic process pipeline and related information items are very important. However, this approach is limited with respect to the appropriate selection of services suited for a specific task or with the appropriate configuration of service parameters. It is at this point that the second step, a knowledge-based approach, plays an important role.

A knowledge-based approach [5] to workflow advice can often succeed in situations where ontology-driven reasoning proves inadequate. For example, in the domain of modelling, simulation and engine performance optimisation there are over a hundred different optimisation methods, each of which is geared to solving a specific type of problem. Even with a single method, different configurations of control parameters may produce very different results. Knowledge about the correct method to choose in a particular situation as well as the appropriate configuration of method parameters is an important feature of expert-level performance and a vital ingredient of problem-solving success. Therefore such a system or systems with similar concepts require access to an exquisitely detailed representation of the knowledge contingencies relating problem characteristics and solution goals with the appropriate selection and configuration of available methods.

3. Case-Based Reasoning

In designing a prototype Workflow Advisor system, the use of Case-Based Reasoning (CBR) techniques has been explored as the underlying technology for intelligent information retrieval and matching. CBR is a knowledge-based, problem-solving paradigm that resolves new problems by adapting the solutions used to solve problems of a similar nature in the past [6] (see Fig. 2). A further advantage of this approach is that it allows consolidation of rule knowledge and provides a

reasoning engine that is capable of probabilistic-based matching. With CBR technology, development can take place in a dynamic and incremental fashion.

In contrast to conventional search engines, CBR systems contain a knowledge model of the application domain in which it operates on. It is therefore not universal but specifically designed for the domain. Hence, it is possible to develop intelligent search abilities, which even show reasonable results when given fuzzy or incomplete requests. Moreover, the results are ranked and complemented by variants and alternatives, thus, not only matches are given but information is valued with "more suitable" or "less suitable".

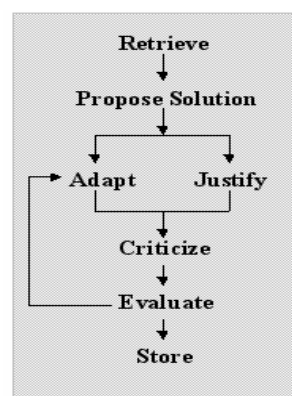


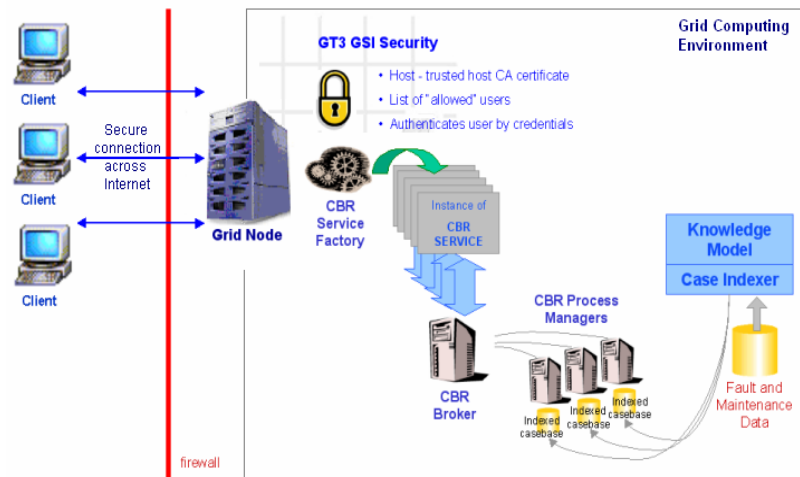
Fig. 2: Case-Based Reasoning

Essential to the CBR system is the casebase that represents a knowledge repository containing detailed workflows and descriptions containing domain knowledge gathered from the DAME system users such as engineers and engine experts performing complex engine diagnostics on The Grid. For a new problem-solving instance, little information is known initially but the advantage of CBR techniques is that a casebase of independent workflows and domain knowledge can be developed in a piecemeal manner and updated as and when knowledge about the behaviour of the problem is known. More importantly, the siting of the CBR system within a Grid environment also allows the integration of diagnostic knowledge from multiple information sources which can improve the accuracy and coverage of the CBR system. Useful workflow information previously available from separate

workflow management components, when brought together into a single knowledge-based system, provides for a powerful Workflow Advice tool.

The requirements for use of open workflow standards [7, 8], BPEL4WS [9], as well as Grid requirements for OGSA standards [10] and inter-operability are also to be taken into consideration. These are crucial in ensuring the

manually as well as via automatic workflow mechanisms. Previous DAME related work at the Sheffield UTC has resulted in a demonstrator implementation of a CBR Service [11] that is currently deployed on the Grid (see Fig. 3). The CBR Service was implemented to support engineers diagnosing an engine abnormality using the DAME system by providing best-practice advice based on a history of engine maintenance cases.



design of a workflow advice system that will interact seamlessly across the heterogeneous Grid environment.

Two levels of operation modes for Workflow Advisor have been identified as applicable within the context of the DAME system. The first level involves consulting the Workflow Advisor at the very start of a workflow process to obtain a suitable workflow. However, workflow processes can sometimes be long and complex requiring decisions to be made along the way depending on the output at each stage. Hence the second level of use involves the above with addition of re-consultation at each stage of a workflow process which may, taking into account results at each stage, yield higher accuracy of workflow advice.

It is highly feasible for a Workflow Advisor system itself to be implemented as a service on the Grid that can be consulted by users

Fig. 3: The CBR Service as deployed on the Grid, enabling users to search across an archive of case histories to support the problem-solving process. Grid service “factories” allow multiple instances of the CBR service to run in parallel on the Grid, each instance supporting an individual client request

4. Workflow Component Interaction in DAME

The Workflow Manager defines, creates and manages the execution of the activities within workflows through the use of software, running on one or more workflow engines, which is able to interpret the process definition, interact with Grid Services and invoke the use of available tools and applications. This process can be performed manually, interactively or automatically. The Workflow Manager also manages the metadata about each Grid Service invocation that it subsequently stores within the DAME provenance database.

The provenance database within the DAME project focuses on not only providing an audit trail for

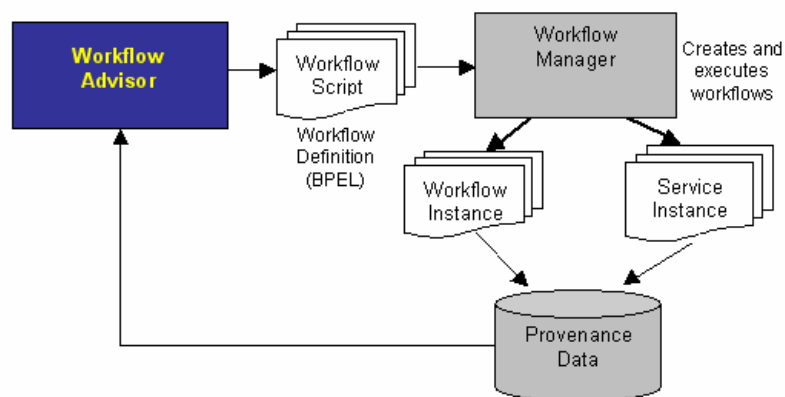
accountability purposes but also to facilitate the re-running of workflow definitions on new problem-solving instances and the creation of new workflows based on past history [12].

Information available within the archive of workflows stored in the provenance database is used by the Workflow Advisor to form a CBR casebase. Each case in the casebase represents a diagnostic process from start to finish. Each case can possess one or more workflows associated with it. Each workflow can in turn possess one or more resource service invocation associated with it. This provides all the relevant historical information regarding the outcome of previous workflow executions.

Figure 4 illustrates the interaction between the

5. Conclusion

There is an emerging requirement for Workflow Advise in the increasingly complex Grid environment. The key challenge is to provide best advice on constructing complex workflows using the available Grid-enabled resources to achieve a desired task, and the underlying knowledge-based techniques to support that. Further work is currently in progress to develop and implement a prototype Workflow Advisor system [2] to be deployed as a service on the Grid. This is to be made available as an integrated feature within the DAME demonstrator system deployed on the White Rose Grid. This particular application of the Workflow Advisor will focus on detailed workflow scenarios derived from formal use cases in DAME [13].



Workflow Advisor with the Workflow Manager and Provenance Data. Workflow Advisor searches and retrieves matching workflows from a repository of captured workflows contained in the Provenance Data. From the workflow advisor result set, suitable workflows are selected and adapted within a workflow script for execution by the Workflow Manager. In processing a new workflow, new workflow instances and new service instances are created. These instances are further captured and stored as new workflow processes within the Provenance Data, effectively expanding the workflow repository with new information for future use.

Fig. 4: The interaction between the Workflow Advisor component, workflow management and provenance data

Acknowledgements

The authors gratefully acknowledge financial support of the Engineering and Physical Research Council in the UK under the DAME project Grant Number GR/R67668/01 and input from engineers at Rolls-Royce and Data Systems & Solutions.

References

1. Distributed Aircraft Maintenance Environment (DAME) project; www.cs.york.ac.uk/dame
2. M. Ong *et al.* Initial Requirements for DAME Workflow Advisor. DAME Technical Report No. DAME/Sheffield/TR/03.004, April 2004.
3. I. Foster, C. Kesselman. The GRID 2. Morgan Kaufman, 2003.
4. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. In *Scientific American*, May 2001.
5. J. Blythe, E. Deelman, Y. Gil, C. Kesselman. Transparent Grid Computing: A Knowledge Based Approach. In *Innovative Applications of Artificial Intelligence Conference (IAAI)*, 2003.
6. J. Kolodner. Case-Based Reasoning. *Morgan Kaufman*, 1993.
7. Workflow management Coalition (WfMC) www.wfmc.org
8. L. Fischer, Workflow Management Coalition (WfMC). The Workflow Handbook 2003.
9. BEA Systems, IBM, Microsoft, SAP AG and Siebel Systems: Business Process Execution Language for Web Services (BPEL4WS) version 1.1, <http://www-106.ibm.com/developerworks/library/ws-bpel/> (2003)
10. I. Foster, C. Kesselman, J. Nick, and S. Tuecke. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, *Open Grid Service Infrastructure WG, Global Grid Forum*, 22nd June 2002.
11. M. Ong, X. Ren, G. Allan, V. Kadiramanathan, H. Thompson, P. Fleming. Decision Support System on The Grid. In proceedings of the *Int'l Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES)*, September 2004.
12. R. Bagshaw *et al.* Initial Requirements for DAME Data Provenance. *DAME Technical Report No. DAME/Leeds/TR/04/002*, March 2004.
13. M. Fletcher *et al.* DAME Mid Term Demonstration Detailed Workflows. *DAME Technical Report No. DAME/York/TR/03.004*. April 2003.