Dr. Mahmoud Mohamed El-Gayyar

PhD Thesis Abstract

Grids and service-oriented technologies are emerging as dominant approaches for distributed systems. With the evolution of these technologies, scientific workflows have been introduced as a tool for scientists to assemble highly specialized applications, and to exchange large heterogeneous datasets in order to automate and accelerate the accomplishment of complex scientific tasks. Several Scientific Workflow Management Systems (*SWfMS*) have already been designed to support the specification, execution, and monitoring of scientific workflows. Meanwhile, they still face key challenges from two different perspectives: system usability and system efficiency.

From the system usability perspective, current *SWfMS* are not designed to be simple enough for scientists who have quite limited IT knowledge. What's more, there is no easy mechanism by which scientists can share and re-use scientific experiments that have already been designed and proved by others. From the perspective of system efficiency, existing *SWfMS* are coordinating and executing workflows in a centralized fashion using a single scheduler and / or
a workflow enactor. This creates a single point of failure, forms a scalability bottleneck, and enforces centralized fault handling. In addition, they don't consider load balancing while mapping abstract jobs onto several computational nodes. Another important challenge exists due to the common nature of scientific workflow applications that need to exchange a huge amount of data during the execution process. Some available *SWfMS* use a mediator-based approach for data transfer where data must be transferred first to a centralized data manager, which is completely inefficient. Other *SWfMS* apply a peer-to-peer approach via data references. Even this approach is not sufficient for scientific workflows as a single complex scientific activity can produce an extensive amount of data.

In this thesis, we introduce SWIMS (Scientific Workflow Integration and Management System) framework. It employs the Web Services technology to originate a distributed management system for data-intensive scientific workflows. The purpose of SWIMS is to overcome the previously mentioned challenges through a set of salient features: i) Support for distributed execution and management of workflows, ii) diminution of communication traffic, iii) support for smart re-run, iv) distributed fault handling and load balancing, v) ease of use, and vi) extensive sharing of scientific workflows. We discuss the motivation, design, and implementation of the SWIMS framework. Then, we evaluate it through the Montage application from the astronomy domain.

Master Thesis Abstract

Recently dominant application field of Web Services is the integration of various applications towards a coherent, workflow-based application. Unlike ”traditional” business processes, composite services delivered through the Internet have to cope with a highly dynamic environment, where new services become available on a daily basis . In addition, the availability of many service providers increases the competition and forces companies to provide better services to satisfy different customers’ needs. These two characteristics of SOA environments impose demanding requirements on a framework that allows normal end-users to adapt easily their business processes in order to cope with dynamic changes in their environment. However, current web services composition engines like ActiveBPEL provide only facilities to create static workflow compositions. Once a workflow is deployed and started, the constructs that represent the workflow cannot be altered subsequently.

In fact, only administrator could re-define a workflow at any time. We claim, however, that such an admin-based deployment is not sufficient. For future application scenarios, we see a need for adapting workflow compositions in a personalized manner even at runtime of the deployed workflow. According to this idea, regular end-users should be able to adapt a workflow for accommodating new requirements.

Based on our review for the component-based tailoring methods, we propose the multi-tier architecture TailorBPEL, based on the ActiveBPEL engine that supports the tailoring of personalized BPEL-based workflow compositions. We believe that such architecture is significant for having highly flexible and adaptable service-oriented architectures that are considered for the adoption to user-centric application scenarios.

Publications

M. El-Gayyar, Y. Leng and A. B. Cremers. "*Distributed Management of Scientific Workflows in SWIMS*". In Proceedings of the 9th International Symposium on Distributed Computing and Applications to Business Engineering and Science (DCABES), pages 327 –331, August 2010.

Scientific workflows are emerging as a dominant approach for scientists to assemble highly-specialized applications, and to exchange large heterogeneous datasets to automate the accomplishment of complex scientific tasks. Several Scientific Workflow Management Systems (SWfMS) have already been designed so as to support the execution, and monitoring of scientific workflows. Even though, there are still some additional requirements and challenges must be met in order to provide a fully distributed and efficient SWfMS. SWIMS (Scientific Workflow Management and Integration System) environment has been developed trying to examine the nature of these challenges and to accommodate the missing requirements. In this paper we are going to highlight these requirements and show how the workflow management in SWIMS fulfills them.

Y. Leng, M. El-Gayyar and A. B. Cremers. "*Semantics Enhanced Composition Planner for Distributed Resources*". In Proceedings of the 9th International Symposium on Distributed Computing and Applications to Business Engineering and Science (DCABES), pages 61 –65, August 2010.

The integration of distributed data resources is a pervasive challenge facing almost all modern application fields, especially for large-scale scientific applications. In those applications, data resources produced independently by multiple domain scientists need to be shared to accomplish a special goal. Typically, such data resources have high heterogeneity, which makes the integration process more difficult. Current Web Services have emerged as a paradigm for managing complex distributed applications. The lack of machine readability in representation prevents Web Services from supporting the effective integration of heterogeneous data resources. Our work tries to enhance web services technology with semantic annotations and ontologies. Such semantic technologies can be used to support semi-automatic heterogeneous data conversion and integration. This paper presents the comparison of some existing systems and proposes a three-layer architecture of Semantically Enriched Integration System (SEIS), which uses ontologies for correlating different data resources available as SAWSDLs and builds OGSA-DAI workflows for reducing heterogeneities.

I. Markovic, S. Jain, M. El-Gayyar, A. B. Cremers, and N. Stojanovic. "Modeling and Enforcement of Business Policies on Process Models with Maestro", Proceedings of the Demo Session at the 6th European Semantic Web Conference (ESWC), Greece, June 2009.

Business policies and rules govern and guide the business processes of an organization. Enterprises usually document their business policies and rules in natural language. This makes the procedure of determining which business policies and rules apply to a certain process and the validation of their adherence within this process very costly and cumbersome. In order to support explicit specification and automated validation of business policies and rules against business processes, in this work we have devised an approach supported by Semantic Web technologies. The approach uses the notion of anti-patterns of process models to capture policies and rules as constraints on process models. Using our approach, we can formally capture and match the context in which a policy/rule applies and validate it against a process model in an automated way. The approach was designed based on a described set of requirements and prototypically implemented within a process modeling tool. An example scenario is provided to illustrate the benefits of the approach.