International Workshop on Applications of Workflows in Computational Science (AWCS 08)

Adam Belloum[1], Zhiming Zhao[1], and Marian Bubak[1,2]

[1] Informatics Institute, University of Amsterdam, Amsterdam, the Netherlands
{adam,zhiming,bubak}@science.uva.nl
[2] Institute of Computer Science AGH, Krakow, Poland
{bubak}@agh.edu.pl
http://staff.science.uva.nl/~zhiming/workshop/awcs/2008/

Abstract. The goal of the Workshop on Applications of Workflows in Computational Sciences (WACS) is to provide a forum for sharing knowledge and experience on developing workflow applications, and to highlight important requirements for developing workflow systems. This short paper gives an overview on scientific workflow management systems and their application in e-Science, and introduces the topics of the Workshop. Several research focuses on utilizing scientific workflow in enhancing e-Science applications and on developing workflow management systems will be enumerated.

Keywords: scientific workflow, e-Science, application integration, data management, resource discovery, frameworks, security.

1 Aims and Scope

In many scientific domains, workflows become an important mechanism to prototype and perform complex experiments and to achieve scientific discoveries. A workflow system allows scientists to wrap and integrate legacy applications via an intuitive interface, to schedule the computing of different tasks at an abstract level, and to automate the processes in data processing. In this way, domain scientists can effectively utilize available resources and focus on the logic of the experiments instead of low level technical details.

During the passed years, a large number of workflow systems have been developed, e.g., DAGMan[1] and Pegasus[2] focus on managing massive computing tasks for processing distributed data, Taverna[3] integrates distributed web services based components, and Kepler[4] provides flexible GUI to prototype workflows and execute them. Decoupling workflow engine from the GUI and implementing the engine as a service allow better interoperability among workflow management systems, a good example for this interoperability has been achieved by invoking the WS-VLAM engine[5] from the Taverna workbench as a sub-workflow in a Taverna workflow.

Most of the SWMS are historically driven by applications in specific domains, e.g., bio informatics, high energy physics, and astronomical observations. Investigating the common characteristics in these domain specific systems and implement them as part of a generic framework emerges as an important need for e-Science applications. A number of research projects such MyExperiment[6] in the UK and VL-e in the Netherlands[7] aim to realize a Grid enabled generic framework where scientists from different domains can share their knowledge and resources, and perform domain specific research. Sharing knowledge and resource requires more interoperability among the major workflow management systems. More sophisticated solution are needed to achieve a seamless integration among workflow, approaches like the workflow bus[8] developed in the context of VL-e present a potential solution to the interoperability problem. Different requirements for supporting domain specific applications are important driving force for the development of workflow systems. It is therefore an important motivation for us organize such workflow to collect use cases from different application domains and understand specific requirements.

An overview of available scientific workflow systems and their application in e-Science is given in[9], while the challenges in this area are examined in[10]. Recently, workflow systems are build exploiting achievements in semantics and ontologies[11], [12], [13]. The goal of the Workshop on applications of workflows in computational sciences (WACS) is to provide a forum for sharing knowledge and experience on developing workflow applications, and highlight important requirements for system development.

2   Workshop Content

The Workshop addresses a number of issues related to the development of workflow applications.

Data management. Jablonski et al. present an infrastructure called DaltOn for managing the complexity in processing and integrating scientific data. DaltOn provides a data centered vision, namely data logistics, on modeling processes of collecting, storing, transporting and processing data in scientific experiments. The DaltOn aims at decoupling the data processing issues, e.g., syntactic and semantic conversions, from the actual workflow steps, and promoting the reuse of experiment logics.

Resource discovery. Huang discusses a resource discovery in distributed environment. A P2P environment called Virgo is used as the test bed. A multi-level virtual group architecture is used to manage the complexity of the system.

Security. Yau, et al. discuss the security and privacy issues scientific workflows. A trusted computing based scheme is proposed for selecting trusted resource providers, for protecting confidentiality and integrity of job information, and for auditing data for process provenance.

Usability. Buckingham et al. present a web accessible scientific workflow system called GPFlow. A model for collection processing based on key aggregation and

slicing which guarantee the processing integrity and facilitates automatic association of inputs. The GPFlow workflow defined on a single value may be lifted to operate on a collection of values with no change required to the workflow. Combing this workflow framework with Grid infrastructure will be the next step to achieve.

Frameworks. Jakimovski et al. present new framework for the Gridification of Genetic Algorithms. The framework enables easy implementation of Genetic Algorithms and also enables researchers easy and stable usage of the Grid for their deployment. The design of the framework was based on principles that make it very open and extensible.

Wibisono et al. describe ongoing efforts for designing a framework for performing parameter sweep experiments. The paper discusses the requirements gathered from use cases in various scientific domains indicate that interactivity is needed but not fully supported by most of existing frameworks designed to support parameter sweep applications. Preliminary design of a framework that would support interactivity is presented.

Application use cases. Fernandez-Quiruelas et al. describe how a well known climate model (CAM) can take advantage of the Grid computing power. In this work usability and robustness are the major requirements because the potential end-users have little background in computer systems and should not be bothered with the complexity of the underlying infrastructure. This requires involve managing a complex workflow involving long-term jobs and data management in a user-transparent way.

3   Summary and Outlooks

Several related workshops have been organized during passed years, e.g., WSES[1], SWF[2], and SWBES[3] The development of workflow management systems faces two important challenging issues. On one hand, the domain specific experiments from different applications require customized solutions in workflows for particular problems; on the other hand, to enable knowledge transfer and information sharing between different domains, a generic workflow solution is also demanded. A successful workflow system no only means it has mature conceptual design and engineering but more importantly it can be effectively enhance real life applications. The usability of a workflow system is essential to introduce the system to different domains scientists: not only suitable interface for composing and executing workflow, but also a set of user oriented tools for viewing, moving and processing data and for the provenance of the workflow and reproducing the

---

[1] International Workshop on Workflow Systems in e-Science,
 `http://staff.science.uva.nl/~zhiming/workshop/wses/`
[2] International Workshop on Scientific WorkFlows,
 `http://www.cs.wayne.edu/~shiyong/swf/`
[3] International Workshop on Scientific Workflow and Business workflow standards in e-Science, `http://staff.science.uva.nl/~adam/workshops/e-science2007/cfp-swbes-2007.htm`

results. Sharing the knowledge in meaningful workflows is becoming be an important requirement for e-Science framework. It is necessary to integrate different technologies, e.g., semantic based annotation and searching, and collaborative working facilitates, with scientific workflow system.

Acknowledgement

We would like to express our great appreciation for all PC members, referees and colleagues who have supported AWCS 08.

References

1. DAGMan. Directed acyclic graph manager (2005),
   http://www.cs.wisc.edu/condor/dagman/
2. Gil, Y., Deelman, E., Blythe, J., Kesselman, C., Tangmunarunkit, H.: Artificial intelligence and grids: Workflow planning and beyond. IEEE Intelligent Systems 19(1), 26–33 (2004)
3. Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M.R., Wipat, A., Li, P.: Taverna: A tool for the composition and enactment of bioinformatics workflows. Bioinformatics Journal, online (June 16, 2004)
4. Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludäscher, B., Mock, S.: Kepler: An extensible system for design and execution of scientific workflows. In: SSDBM, pp. 423–424 (2004)
5. Wibisono, A., Belloum, A., Inda, M., Roos, M., Breit, T., Hetrzberger, L.O., Korkhov, V., Vasunin, D.: Vlam-g: Interactive dataflow driven engine for grid-enabled resources. Scientific Programming 15(3), 173–188 (2007)
6. Goble, C.A., De Roure, D.C.: myexperiment: social networking for workflow-using e-scientists. In: WORKS 2007: Proceedings of the 2nd workshop on Workflows in support of large-scale science, pp. 1–2. ACM, New York (2007)
7. VL-e. Virtual laboratory for e-science (2005), http://www.vl-e.nl/
8. Zhao, Z., Booms, S., Belloum, A., de Laat, C., Hertzberger, B.: Vle-wfbus: a scientific workflow bus for multi e-science domains. In: Proceedings of the 2nd IEEE International conference on e-Science and Grid computing, Amsterdam, the Netherlands, December 4- 6, 2006, pp. 11–19. IEEE Computer Society Press, Los Alamitos (2006)
9. Taylor, I., Deelman, E., Gannon, D., Shields, M. (eds.): Workflows for e-Science. Springer, Heidelberg (2007)
10. Gil, Y., et al.: Examining the Challenges of Scientific Workflows. IEEE Computer 40(12), 24–32 (2007)
11. Bubak, M., Unger, S. (eds): K-WfGrid - The Knowledge-based Workflow System for Grid Applications. In: Proceedings of CGW 2006, vol. II, http://www.cyfronet.krakow.pl/cgw06/, ISBN 978-83-915141-8-4
12. Lee, S., Wang, T.D., Hashmi, N., Cummings, M.P.: Bio-STEER: A Semantic Web Workflow Tool for Grid Computing in the Life Sciences. Future Generation Computer Systems 23(3), 497–509 (2007)
13. Truong, H.-L., Dustdar, S., Fahringer, T.: Performance Metrics and Ontologies for Grid Workflows. Future Generation Computer Systems, vol 23(6), 760–772 (2007)