2008

Accelerating time to scientific discovery with a grid-enhanced microsoft project

Gregor von Laszewski

Leor Dilmanian

Accelerating Time to Scientific Discovery with a Grid-Enhanced Microsoft Project

Gregor von Laszewski and Leor E. Dilmanian
Rochester Institute of Technology, Rochester, NY 14623, Email: laszewski@gmail.com

Abstract

The composition, execution, and monitoring of challenging scientific applications is often a complex affair. To cope with the issue of workflow management, several tools and frameworks have been designed and put into use. However, the entry barrier to using these tools productively is high, and may hinder the progress of many scientists, or non-experts, that develop workflows infrequently. As part of our Cyberaide framework we enable workflow definition, execution and monitoring through the Microsoft Project software package. The motivation for this choice is that many scientists are already familiar with Microsoft Project, a project management software package that is perceived to be user friendly. Through our tool we have the ability to seamlessly access Grids, such as the NSF sponsored TeraGrid. Cyberaide abstractions have also the potential to allow integration with other resources, including Microsoft HPC clusters. We test our hypothesis of usability while evaluating the tool as part of several graduate level courses taught in the field of Grid and Cloud computing.

1 Introduction

The workflow paradigm has proven itself to be an effective paradigm in response to challenges imposed by many scientific problems. To address these challenges, several different frameworks have been developed [1] to provide usable tools which simplify the use of executing distributed applications on a computational Grids. As part of the scientific communities, the most demanding computational problems require large-scale resources and orchestration of activity. To date, numerous Grid workflow management systems have been created. Unfortunately, there is no one-size-fit all workflow management tool or solution. The multitude of solutions is based on a number of challenges that we outline next:

- *Changing standards to integrate Grid concepts:* Although organizations such as the Workflow Management Coalition has already defined and implemented a standard set of tools, documents, and reference models, the standards are being redefined to address a whole set of issues unique to the Grid, including the need for a "service-based architecture intended for activities ranging from supporting large, distributed virtual organizations for e-science to autonomic and on-demand computing for commercial enterprises [2]."
- *Complexity and unfamiliarity of the user interface:* Although graphical user interfaces in general can be used to simplify interaction with a scientific workflow composition system, they require the learning of yet another tool that may in practice be used infrequently. The learning curve for a basic workflow remains high.
- *Integration of middleware is limited:* In many cases, the workflow tools developed by industry and also the scientific community are limited and may not provide or allow the integration of other frameworks and middleware. Although Grid protocol standards have been defined, additional frameworks to enable an advanced cyberinfrastructure which might not have Grid services installed need to be considered.
- *Work modalities of the exiting community:* In some communities, existing tools already cover portions of the complex workflow process, but are not integrated with Grid or Cloud solutions and frameworks.

As a direct result, we are in search for a multitude of interfaces that support use cases driven by community needs. However, the resulting tool must be easy to learn, adopt, expandable, and integrate with high performance computing resources as part of deployed Grids and Clouds.

To identify such a solution, *commodity* tools that are being used by the community to support increase scientific productivity need to be identified. One such tool is *Microsoft Project* [3]. It is often used by scientists to plan, coordinate, and execute various activities; track and display progress, and build reports on project related activities with colleagues and sponsors.

The following questions are used to guide our research effort:

- Is Microsoft Project feasible to support the scientific workflow that is required as part of Grid and Cloud computing activities?
- Which types of scientific activities which can be supported using a Microsoft Project based workflow tool?
- Which types of middleware solutions that can be integrated into Microsoft Project?
- What are the advantages and limitations of using such an approach?

2 Use Cases

As obvious from the functionality of Microsoft Project, the use cases can include the execution of dependent jobs with resource constraints. Such constraints are typical for unique and large-scale instruments that need to be shared amongst a group of scientists. As such resources are scarce. Microsoft Project allows a framework for sharing and collaboration. Priorities can be used to order the tasks and to create a schedule that is suitable to address not only multiple tasks by a single scientist but by an entire research group. Such functionality can be made available through both the Microsoft Project client and server products.

Projects using synchrotrons, flow-cytometers, and electron microscopes as resources can benefit greatly from such a tool. Even without specifying resource constraints as part of the workflow, it allows scientists to organize the experiments and enables a rigorous planning of the often complex calculations while being able to integrate them in the overall planning of an experiment. Furthermore, it can be used to guide resource allocations needed to invoke scientific applications required as part of emergency management situations, such as intrusion detection of contaminants in the water management system of a city.

Other use cases are based on the actual graphical user interfaces supporting workflow designs. Our framework will allow the import and export of workflows to be executed with other tools and workflow frameworks common in the Grid community.

3 Implementation

As part of our framework to simplify the use of advanced cyberinfrastructure called *Cyberaide* we developed a tool that uses Microsoft Project and is able to steer computations on the Grid. It is extensible as it provides an abstraction layer for the execution of tasks on computational resources, which requires activities such as fault management and data transfer. Figure 1 shows a screenshot of our tool in action while composing and executing jobs on a variety of resources on the TeraGrid.

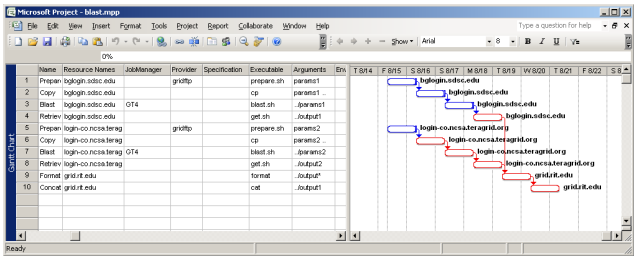

Figure 1: Using Microsoft Project to compose and execute grid workflows on the Teragrid.

Microsoft Project provides the appropriate views for the display of resource related information. Such views include resource graphs, calendars, charts, and usage diagrams. Using both standard and custom fields associated with tasks and resources, the software package can be used to store and monitor various quality of service parameters, allowing built in data analysis to make performance estimation. Resource discovery can be accomplished using built in support from active directory, the address book, or a Microsoft Project server. Programmatic manipulation of the Microsoft Project 2007 client in the C# console application was made possible through the Microsoft Project Primary Interoperability Assembly. Besides the manipulation of the object through a graphical user interface, our implementation also contains a command line console allowing integration of scripts and ad-hoc computational steering. Events generated from the Grid middleware are forwarded to the Microsoft Project and allow for dynamic monitoring.

4 Status

Our Cyberaide tool demonstrates a feasible workflow framework that uses Microsoft Project as a composition and execution tool to perform jobs on a Grid such as the TeraGrid. Our current research efforts includes the creation of abstractions for Cloud services such as Amazons Cloud Services and the integration of a workflow enactment engine that takes advantage of the failure handling, check-pointing, dynamic and distributed workflow features. In addition, we are integrating our Grid and Cloud shell that is under active development as part of the Cyberaide project.

5 Evaluation

The availability of additional Microsoft tools can support planning, specification, execution, publication, sharing not only of the workflows, but be helpful for the entire scientific discovery process. This can be of immense value to many scientists. Microsoft Project provides for many users the necessary abstraction. For them Microsoft Project provides a tool of choice, as our framework has demonstrated. Simplicity and ubiquity are the essential factors for the user adoption.

Acknowledgment

Work conducted by Gregor von Laszewski is supported (in part) by NSF CMMI 0540076 and NSF SDCI NMI 0721656.

References

[1] I. Taylor, E. Deelman, D. Gannon, and M. Shields, Eds., *Workflows for e-Science: Scientific Workflows for Grids*. Springer, 2007, iSBN: 978-1-84628-519-6.

[2] G. Fox and D. Gannon, "Workflow in Grid Systems," *Concurrency and Computation: Practice and Experience*, vol. 18, no. 10, pp. 1009–1019, 2006.

[3] *MSDN - Project 2007*, Microsoft. [Online]. Available: http://msdn.microsoft.com/en-us/library/bb258902.aspx