Scientific Workflows for Game Analytics

Apostolos Georgas
Hellenic Open University, Greece

Dimitris Kalles
Hellenic Open University, Greece

Vasileios A. Tatsis
Institute of Physical Chemistry, University of Muenster, Germany & Hellenic Open University, Greece

INTRODUCTION

Grid computing is about sharing the computing and storage elements of available heterogeneous computer systems to create a pool of resources that would match those of a powerful super computer (Foster, 2002). It is about providing the means to execute computational experiments or applications that, due to their complexity, often require a large amount of computer (storage and computing) resources.

It is the complexity of modern applications and experiments that lead to the development of scientific workflows. These are mostly used to describe, manage and share complex scientific analyses. Different types of tasks described by a workflow can be executed by employing local and/or remote Web services, scripts and sub-workflows (workflows used as fragments of larger ones). Most workflow systems offer a reasonably intuitive interface for setting-up, organizing, submitting and analyzing complex experiments; this interface usually combines a graphical metaphor to model the experiments and a formal language to describe that model. Workflow systems allow users to integrate external applications or services into workflows, deploy the workflows to a grid infrastructure, check for possible errors and finally, analyze the output of these calculations (Taylor, 2007), with relatively modest demands on a user's IT skills.

Bioinformatics has been one of the scientific fields that generated the demand for grid computing and scientific workflows. As an example, consider *in silico* studies of biomimetic enzymes, which employ protein structure databases and replica-exchange molecular dynamics algorithms; there has long been a huge demand for powerful computer resources to execute experiments and for sophisticated analysis tools to cope with the large amount of data generated (Dongarra & Rarhavan, 2000; Caragiannis 2013).

We have identified the domain of machine learning in games as a field where grid computing and workflows hold potential; therein, the distribution of computations that are required for the evolution of learned behaviors and strategies must be coupled with a well designed sequence of learning experiments and the accompanying data analysis (Kalles & Fykouras, 2010; Kalles & Kanellopoulos, 2001). We have thus oriented our contribution towards the design and implementation of scientific workflows to investigate game learning using a variety of state-of-the-art tools for workflows, both at the desktop ant at the grid. Based on this experience, we argue that these tools, especially the grid-based one, are of paramount importance to researchers who need to test complex hypotheses that demand excessive amounts of computations. We offer experimental evidence for this argument by showing how the setting we have used can be further exploited for

DOI: 10.4018/978-1-4666-5202-6.ch190

Copyright © 2014, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

designing optimized interaction sequences for machine learning of playing tactics in gaming.

The rest of this contribution is structured as follows. We first briefly describe scientific workflow software and then we review our initial implementation of experimental sessions in this software. Following that, we review recent developments on workflow systems for grid applications and we set out how to use these resources to analyze the behavior of game playing learning agents. The last section concludes our contribution and identifies promising and ongoing research directions.

A BRIEF BACKGROUND ON WORKFLOWS

Workflow Applications

The simplest computerized scientific workflows are scripts that use data, programs, and other inputs and produce outputs that might include visualizations and analytical results. These may be implemented in programs such as R or MATLAB, or using a scripting language such as Python or Perl.

More specialized scientific workflow systems, e.g. Discovery Net, Taverna (Hull *et al*., 2006; Oinn *et al*., 2006; Missier *et al*., 2010), BPEL (Tan, Missier, Madduri, & Foster, 2008) and Kepler (Altintas *et al*., 2004), provide a visual programming front-end enabling users to visually model their applications as a graph by connecting nodes. Each directed edge in the graph typically represents a connection from the output of one application to the input of another one.

Typically sitting on top of a middleware layer, scientific workflows are a means by which scientists can model, design, execute, debug, reconfigure and re-run their analysis and visualization pipelines.

Brief Review of Taverna and Kepler

Taverna and Kepler are two of the most widely used software platforms for designing, implementing and redistributing scientific workflows. Taverna is particularly popular in life sciences and in genomics research. It offers an environment to access Web services through a graphical user interface, purportedly with limited technical knowledge of Web services or programming (Oinn *et al*., 2006).

The primary aim of Taverna is to aid scientists to combine remote Web services to build scientific workflows (Tan *et al*., 2009; Damkliang, Tandayya, Phusantisampan, & Sangket, 2009), so a collection of components has been developed by suitably organizing available Web services (Oinn *et al*., 2004). Since Taverna works with components coming from different autonomous service providers, it has been designed to simply assume XML as a common data format. Additionally, it provides a set of generic component types for rapid integration or development of new components.

Kepler's (Altintas *et al*., 2004) conceptual predecessor is Ptolemy II, an actor-oriented modeling tool primarily targeted at embedded and real-time system design. Kepler's focus on data analysis and modeling makes it suitable for modeling processes in diverse scientific domains, such as physics, bioinformatics and information systems (Ludäscher *et al*., 2007; Wang, Crawl, & Altintas, 2009; Abramson, Bethwaite, Enticott, Garic, & Peachey, 2011). Kepler separates its execution engine from the workflow model; additionally,

Table 1. Comparison between Taverna and Kepler

	Taverna	**Kepler**
User interface	↑	↔
Execution speed	↑	↑
Syntactical features	↓	↑
Control behavior	↔	↑
Embedding	↔	↑
Access to Grid	↔	↓
Log creation	↔	↔
Extensibility	↔	↓
Support for Life Sciences	↑	↔
Availability (free)	↑	↑

Legend: ↑ Good, ↔ Fair, ↓ Poor.

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/scientific-workflows-for-game-analytics/107398

Related Content

Stock Market Prediction Using Elliot Wave Theory and Classification
Saeed Tabar, Sushil Sharmaand David Volkman (2021). *International Journal of Business Analytics (pp. 1-20).*
www.irma-international.org/article/stock-market-prediction-using-elliot-wave-theory-and-classification/269484

Applications of Social Graphs: Author, Email, Telephone Using Graph Mining Techniques
Bapuji Rao, Sasmita Mishraand Sarojananda Mishra (2018). *Improving E-Commerce Web Applications Through Business Intelligence Techniques (pp. 297-331).*
www.irma-international.org/chapter/applications-of-social-graphs/197199

A Mixture Price Trend Model for Long-Term Risk Management
Eric S. Fung, Wai-Ki Chingand Tak-Kuen Siu (2010). *Business Intelligence in Economic Forecasting: Technologies and Techniques (pp. 157-173).*
www.irma-international.org/chapter/mixture-price-trend-model-long/44254

Business Intelligence Enhances Strategic, Long-Range Planning in the Commercial Aerospace Industry
David Ellis (2011). *International Journal of Business Intelligence Research (pp. 17-28).*
www.irma-international.org/article/business-intelligence-enhances-strategic-long/60242

Reconnection of Wireless Sensor Network Partitions on Multi-Agent Platform
E. Anna Deviand J. Martin Leo Manickam (2019). *International Journal of Business Analytics (pp. 43-54).*
www.irma-international.org/article/reconnection-of-wireless-sensor-network-partitions-on-multi-agent-platform/218834