FAKULTÄT FÜR INFORMATIK

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Representation and Visualization of load consumption in D2WORM work units.

Rajendra Kharbuja

FAKULTÄT FÜR INFORMATIK

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Representation and Visualization of load consumption in D2WORM work units.

Repräsentation und Visualisierung von Lastverbrauch in D2WORM Arbeitseinheiten.

Author:	Rajendra Kharbuja
Supervisor:	Prof. Dr. rer. pol. Hans-Arno Jacobsen
Advisor:	Martin Jergler
Submission Date:	

I confirm that this master's thesis in informatics is my own work and I have documented all sources and material used.

Munich, Rajendra Kharbuja

Acknowledgments

Abstract

Contents

1 Introduction

1.1 Motivation

Business enterprises in current world have complex requirements and expectations in terms of time, speed as well as agility. Increased level of competition in the global market has been one of the key influencing factors for business to be cost effective at runtime, as well as to be open to any future changes. In order to make it possible, enterprises should be agile as much as possible supporting the ability to enhance the business processes along with the underlying information technology.[SGH95] This insists on the technology infrasturcture capable of incorporating information among various heterogenous systems. [Alo+np]

Scientific research and experiments are no different than business processes in a way that they require sophisticated infrastructure as well as high processing cabability. Scientific research are no longer limited to single domain or geographical area. Moreover, they need to utilize fresh data as well as services from various interdisciplinary scientific groups with the intension to improve speed, efficiency and accuracy. It clearly suggests that an infrastructure capable of handling large intensive data accross network of scalable systems in a reliable and fast delivery, is required.[Lud+05]

The valid requirements discussed in previous segments are fulfilled by workflow management system (WMS). It is a well-build technology to automate, monitor and controll business processes. It provides the required infrastructure to model the business processes or workflows, generate workflow implementation from the model and monitor the workflow providing required inputs to optimize the business process. It separates the business workflow logic from the core logic and thus provides robust system to define the business process.[SGH95; SSnp]

Thus, workflow can be considered as the basic element of any WMS. Depending upon, where wokflow are processesed, a WMS can be either centralized or distributed. In centralized WMS, the workflows are modeled as well as executed at central location however in case of distributed or decentralized WMS, the workflows are distributed across various nodes and executed.[LHL00] Moreover, the scientific as well as business

processes consists of collection of loosely coupled heterogenous computing devices distributed across network spaning to heterogenous domains with different administration control and scalability as well as different performance requirement. [SLBnp; Alo+np]

A centralized approach for WMS cannot be an appropriate choice for such cases but the central workflow controller can be bottleneck and complicated to amend as well as understand. [Dog+np]Instead, a decentralized workflow management can be the optimum solution. [Alo+np]

Additionally, depending upon the way workflows are modeled and processed, WMS can be either activity-centric, document-centric or data-centric. Activity-centric WMS represents workflow as graph of activities, nodes depicting tasks whereas edges modeling the flow sequence. A slightly improved approach than activity-centric is document-centric approch where workflow are modeled using documents. Both these approaches give high priority to the process-flow rather than the base data model. The static representation of the process-flow makes it hard to change and the complexity of the model increases with the number of tasks. On the other hand, data-centric WMS uses global data model using application level data and process status information, to visualize the workflow. This kind of modeling using business artifacts provides loose coupling as well as different level of abstraction among agents, which eventually contribute to simplicity and adaptablity.[JSJnp; CHnp]

Taking into considerations all different scenerios and their comparative study as discussed in earlier segments, Distributed Data-centric Workflow On-line Resource Management System (D2WORM) has been designed and implemented. D2WORM has a global data model to represent the process state and application data. The overall workflow logic is handled by a coordinated agents or workflow units, each responsible for a part of the complete workflow logic and also updating a subset of global data model. Furthermore, the co-ordination among the workflow units is handled by publish-subscribe based network, contributing loose coupling and flexibility.[JSJnp]

One of the major goals of distributed systems (with D2WORM being no exception) is to provide high performance.[KCH08] However, there are many challeges to acheive that, heterogenous nodes, unsymmetrical load and network connection being few of them. The overall performance can depend upon various elements such as individual component properties, system definition, underlying environment, etc. Load distribution is one of the major problems as well as contributing element to achieve high performance. A high degree of performance can be achieved if load can be distributed

proportionally to the units based on some criteria (defined by following section) so that the overall response time of the system is increased.[CMGnp; SPV11]

1.2 Problem Statement

Acronyms

D2WORM Distributed Data-centric Workflow On-line Resource Management.

WMS Workflow Management System.

List of Figures

Bibliography

[Alo+np] G. Alonso, D. Agrawal, A. E. Abbadi, and C. Mohan. *Functionality and Limitations of Current Workow Management Systems*. Tech. rep. IBM Almaden Research Center and Institute of Information Systems, np.

[CHnp] D. Cohn and R. Hull. *Business Artifacts: A Data-centric Approach to Modeling Business Operations and Processes*. Tech. rep. IBM T.J. Watson Research Laboratory, np.

[CMGnp] S. Chakraborty, S. Majumder, and D. Goswami. *A Model for Load Balancing in Distributed System using -Congestion Game*. Tech. rep. Indian Institute of Technology, np.

[Dog+np] A. Dogac, E. Gokkoca, S. Arpinar, P. Koksal, I. Cingil, B. Arpinar, N. Tatbul, P. Karagoz, U. Halici, and M. Altinel. *Design and Implementation of a Distributed Workow Management System METUFlow*. Tech. rep. Middle East Technical University, np.

[JSJnp] M. Jergler, M. Sadoghi, and H.-A. Jacobsen. *D2WORM: A Management Infrastructure for Distributed Data-centric Workflows*. Tech. rep. Technische Universitaet Munich, np.

[KCH08] I. Kuz, M. M. T. Chakravarty, and G. Heiser. *COMP9243 — Week 1*. Tech. rep. The University of New South Wales, 2008.

[LHL00] S. I. Lee, D. Han, and D. Lee. *A Pattern for Managing Distributed Workflows*. Tech. rep. Information and Communications University, 2000.

[Lud+05] B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. A. Lee, J. Tao, and Y. Zhao. *Scientific Workflow Management and the Kepler System*. Tech. rep. 2005.

[SGH95] A. SHETH, D. GEORGAKOPOULOS, and M. HORNICK. *An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure*. Tech. rep. Kluwer Academic Publishers, 1995.

[SLBnp] J. Schneider, B. Linnert, and L.-O. Burchard. *Distributed Workflow Management for Large-Scale Grid Environments*. Tech. rep. Technische Universitaet Berlin, np.

[SPV11] M. van Steen, G. Pierre, and S. Voulgaris. *Challenges in very large distributed systems*. Tech. rep. np, 2011.

[SSnp] K. P. Stoilova and T. A. Stoilov. *Evolution of the workflow management systems*. Tech. rep. Institute of Computer and Communication Systems, np.