ADVANCED TOPICS IN SCIENCE AND TECHNOLOGY IN CHINA

Zhaohui Wu
Huajun Chen

Semantic Grid

Model, Methodology, and Applications

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

Springer

Zhao-hui Wu
Hua-jun Chen

Semantic Grid: Model, Methodology, and Applications

With 64 figures

ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

Springer

图书在版编目 (CIP) 数据

语义网格:模型、方法和运用= Semantic Grid: Model, Methodology, and Applications: 英文 / 吴朝晖,陈华钧著. —杭州:浙江大学出版社,2008.5
ISBN 978-7-308-05830-8

Ⅰ.语… Ⅱ.①吴…②陈… Ⅲ.语义网格-研究-英文 Ⅳ.TP18

中国版本图书馆 CIP 数据核字 (2008) 第 048967 号

语义网格:模型、方法和运用

吴朝晖　陈华钧　著

封面设计　Joe Piliero
责任编辑　傅　强　尤建忠
出版发行　浙江大学出版社
网址:http://www.zjupress.com
Springer-Verlag GmbH
网址:http://www.springer.com
排　　版　浙江大学出版社电脑排版中心
印　　刷　杭州富春印务有限公司
开　　本　787mm×960mm　1/16
印　　张　15.25
字　　数　549 千
版 印 次　2008 年 5 月第 1 版　2008 年 5 月第 1 次印刷
书　　号　ISBN 978-7-308-05830-8 (浙江大学出版社)
ISBN 978-3-540-79453-0（Springer-Verlag GmbH）
定　　价　90.00 元

浙江大学出版社发行部邮购电话 (0571)88072522

ADVANCED TOPICS
IN SCIENCE AND TECHNOLOGY IN CHINA

ADVANCED TOPICS IN SCIENCE AND TECHNOLOGY IN CHINA

Zhejiang University is one of the leading universities in China. In Advanced Topics in Science and Technology in China, Zhejiang University Press and Springer jointly publish monographs by Chinese scholars and professors, as well as invited authors and editors from abroad who are outstanding experts and scholars in their fields. This series will be of interest to researchers, lecturers, and graduate students alike.

Advanced Topics in Science and Technology in China aims to present the latest and most cutting-edge theories, techniques, and methodologies in various research areas in China. It covers all disciplines in the fields of natural science and technology, including but not limited to, computer science, materials science, life sciences, engineering, environmental sciences, mathematics, and physics.

Preface

The Internet has been an indispensable means of communication in our daily life. We rely upon it to communicate with others, search information to procure a solution to a knotty problem, book tickets to arrange our trips, look for business opportunities, entertain ourselves, and so forth. Without the Internet our life would have been a largely different one. However, has the Internet reached its full potential? Can it change our life more than we have seen? What will the Internet and the Web look like 20 years later?

A plurality of researchers from different areas have been working on these issues for a long time. At the fore, two distinguished and influential ones are Grid Computing and the Semantic Web.

The term the Grid was first used around 1990 as a metaphor for making the use of computer power as easy as the electric power Grid. It was originally coined as a new paradigm for solving computation-intensive problems by taking advantage of the computation power of idling computers, which could be a super computer or just a desktop computer. Grids were then described as well-organized virtual systems that may span across many organizational boundaries. Grid applications feature in the capability of the dynamic formation of cross-institutional Virtual Organizations in an *ad hoc* way to enable coordinated resource sharing and problem-solving across multiple administrative domains on the Internet.

The term Semantic Web was coined around 1998 by the web inventor Tim Berners Lee. It aims at leading the Web to its full potential by making its content machine-understandable. It draws on the standardization effort of a formal representation framework and advanced web languages such as RDF/OWL that can be used to enrich web resources with semantic descriptions and describe complex semantic relations among them. The semantic theory underlying this formal representation framework provides a formal account of meaning in which the logical relationship of web resources, which can be a webpage, a database record, a program, a web service, and so forth, can be explicitly described and specified without loss of the original meaning. This makes the web smarter and more intelligent, thereby enabling more sensible searches, far more accurate information retrieval, and seamless information integration.

Basically, Grids are concerned with the design and development of the architecture of the future Internet. Their ultimate goal is to provide a flexible, adaptive, manageable, service-oriented architecture for future Internet-based applications. Meanwhile, the Semantic Web looks at the semantic heterogeneity issue that has hindered almost all integration systems in the perspective of information representation. Although these two technologies offer different solutions, they actually complement each other. The concept of the Semantic Grid was then brought up by many researchers with the intention of combining them together to address many difficult problems that cannot be resolved by only one of them.

Commonly, the *Semantic Grid* refers to a branch subject under the umbrella of *Grid Computing* in which computing resources and services are described in a meaningful way that can be discovered, aggregated, joined up more readily and automatically. The description typically draws upon the technology of the Semantic Web, such as the Resource Description Framework (RDF) and Web Ontology Language (OWL). It is stated that the semantic grid, as a combination of the technologies from both grid computing and the semantic web, would provide a promising alternative for developing a future interconnection environment, particularly geared to enable highly selective resource sharing, very sensible knowledge discovery and collective intelligence.

As a synthesis of many different technologies, the Semantic Grid has a broad spectrum of topics. The book attempts to give a comprehensive introduction of these topics, including knowledge representation and semantic description for semantic grid applications, semantic-based data integration, grid service management and process orchestration, trust management in grids, ontology management for problem solving in the semantic grid, and integrative knowledge discovery based on the integration capability of the semantic grid.

How to use this book. This book can be a reference book for researchers in Internet-related technologies. Generally, the topics in Chapters 2-4 are introduced in a fundamental way, while the ones in Chapters 5-8 are presented from an applied perspective. Moreover, Chapters 9,10 are devoted to the experience of applying specific semantic grid technology to two typical application domains: the life science domain and an intelligent transportation system. Specifically, this book is organized in the following structure:

- Chapter 2 describes the relationship between knowledge representation and the semantic grid. The semantic web languages largely draw upon the fruits of the long-standing research on knowledge representation in the area of artificial intelligence. A good information representational framework is vital for a smarter and more intelligent grid system. For example, grid resource discovery relies upon a better description of the resources and the relationships between them. Rules are useful for specifying mappings, coordination policy, security settings, transaction configurations, and trust dependencies for grid applications.

- Chapter 3 describes typical issues such as sub-ontology management for a problem solving in the Semantic Grid. With the Semantic Grid as the problem solving environment, we will face many unexpected problems as in traditional problem solving. The problems to be solved are often complex and refer to large-scale domain knowledge from crossover disciplines. This chapter focuses on how to manage and reuse ontology that embodies domain knowledge based on the infrastructure of the Semantic Grid.
- Chapter 4 mentions an important issue: trust management in the Semantic Grid. Enabling trust to ensure more effective and efficient interaction is at the heart of the Semantic Grid vision. This chapter presents an integrated computational trust model based on statistical decision theory and Bayesian sequential analysis. The model helps users to select an appropriate service provider within the Semantic Grid environment.
- Chapter 5 introduces specific technology that can be used for semantic data integration in the Semantic Grid, with particular emphasis on integrating relational databases with semantic web ontologies. Integrating legacy relational databases is important for both Grid and Semantic Web applications. However, experience in building such applications has revealed a gap between semantic web languages and the relational data model. This chapter presents an intelligent framework with a formal mapping system to bridge the gap, and studies the problem of reasoning and query answering using the semantic view of the mapping system.
- Chapter 6 provides a comprehensive introduction to service management in the Semantic Grid including service description, service orchestration, service discovery, service composition, and so on. How to collaborate, cooperate and co-experiment conveniently and efficiently in the grid environment has become a hot topic in the research and application of the grid, and service flow management will be the key technology in solving the problem.
- Chapter 7 proposes the general ideas and the preliminary implementation of knowledge discovery in the Semantic Grid, with the emphasis on mining based semantic integration. The Semantic Grid provides a new computational environment, and also a new architecture for data mining. The dynamic extension of the algorithm, the transparent integration of data, and the circular refinement of knowledge, are main characteristics of knowledge discovery using such architecture, as high-level services of the Semantic Grid, data mining and knowledge discovery greatly enhance the effectiveness of the Semantic Grid.
- Chapter 8 presents a semantic grid platform called DartGrid. The Semantic Grid combines many technologies coming from the Grids, the Web

Service and the Semantic Web. Organic integration of these technologies that are actually complement each other can result in competent implementation for both Grid and Semantic Web applications. This chapter presents a semantic grid implementation, called DartGrid, which is made up of several components that are intended to support data integration and service management in Grids.
- Chapter 9 introduces the application of the specific technology of the Semantic Grid in building an e-Science environment for the Traditional Chinese Medicine community from the perspectives of knowledge engineering, data integration, and knowledge discovery.
- Chapter 10 introduces the attempted application in an intelligent transportation system, the goal of which is to build an integrated intelligent transportation information and service platform, to integrate traffic data resources and cooperate with existing ITS subsystems and services.

The book is the result of several years of study, research and development of the faculties, PhD candidates and many others affiliated to the CCNT Lab of Zhejiang University. We would like to give particular thanks to Yuxin Mao, Xiaoqing Zheng, Shuiguang Deng, Yi Feng, Yu Zhang, Chunyin Zhou, Tong Yu, Wei Shi, Guozhou Zheng, Jian Wu who have devoted their energy and enthusiasm to the book and relevant projects.

In addition, the work in this book was mainly sponsored by the China 973 project of the Semantic Grid initiative (NO. 2003CB317006), the National Science Fund for Distinguished Young Scholars of China NSF Program (NO. NSFC60533040), the Program for New Century Excellent Talents in University of the Ministry of Education of China (NO. NCET-04-0545). The work was also partially supported by the National Program for Modern Service Industry (2006BAH02401), the Program for Changjiang Scholar(IRT0652) and the NSFC Program under Grant NO. NSFC60503018 and NSFC60603025.

The Semantic Grid is still an undergoing area of rapid development. Although this book cannot give a complete account of all issues and topics, we hope it can shed some light on the most important aspects relevant to the future Internet and can be valuable for those who are interested in the future development of the amazing Internet technology.

Zhaohui Wu
Zhejiang University
October 2007

Contents

1 Introduction

Abstract: The Semantic Grid commonly refers to a branch subject under the umbrella of Grid Computing. In a typical semantic grid application, the network resources and services are described in a meaningful way so that the resources can be discovered, aggregated, joined up more readily and automatically. The description typically draws upon the technology from the Semantic Web, such as the Resource Description Framework (RDF) and Web Ontology Language (OWL). In this chapter an overview to the background of the Semantic Grid is given.

1.1 Background

The concept of the Semantic Grid was derived from Grid Computing and the Semantic Web. This section offers the background introduction on the basic concepts and histories of these two technologies.

1.1.1 Grid Computing

Grid computing (Foster I, 2002a) was originally coined as a new paradigm for solving computation-intensive problems by taking advantage of the computation power of idling computers, which can be a super computer or just a desktop computer. The Grids were then described as a well-organized virtual cluster that may span many organizational boundaries. Grid applications feature in the capability for the dynamic formation of cross-institutional *Virtual Organizations* (VO) in an *ad hoc* way to enable coordinated resource sharing and problem-solving across multiple administrative domains. The features make themselves distinguishable from traditional computer clusters or distributed computing that are often confined to a local domain and within only one organization.

1.1.1.1 Basic Concepts

In the beginning, the Grids were particularly advocated as a brand-new solution for solving those "big-science" problems such as those in life science, physics, climate modeling, financial computing and others. At the present time, as the concept of the Grid grows broader, Grids are also considered as a means of offering information as a utility service such as a computational center or a data center for commercial purposes, with those clients paying as they use it, just like other traditional utilities (electricity, water, etc.).

As Grids have kept gaining considerable momentum since their emergence, a wide variety of definitions about Grid Computing are proposed by different communities. The rather authoritative definition is from Ian Foster, who identifies the real and specific problem that underlies the Grid concept as coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations, and defined the concept of Grid in his article "What Is the Grid? A Three Point Checklist"(Foster I, 2002b). The three points of this checklist are: a) Computing resources are not administered centrally, b) Open standards are used, c) Non-trivial quality of service is achieved.

Another popular definition for the Grid concept comes from IBM who defines Grid Computing as "the ability, using a set of open standards and protocols, to gain access to applications and data, processing power, storage capacity and a vast array of other computing resources over the Internet. A Grid is a type of parallel and distributed system that enables the sharing, selection, and aggregation of resources distributed across 'multiple' administrative domains based on their (resources) availability, capacity, performance, cost and users' quality-of-service requirements." [1]

From a general perspective, Grid computing involves sharing and managing heterogeneous resources with different hardware architecture, software platforms, and network protocols across geographical locations and administrative domains over a network, and emphasizing an open-standard-based approach. Functionally, Grids can be classified into several types:

- **Computational Grid** that focuses primarily on the sharing of computational power such as a CPU cycle with computationally-intensive operations.
- **Data Grid** emphasizing the controlled sharing and management of large amounts of distributed data, usually at a tera-scale.
- **Device Grid** that offers the service of controlling the equipment such as a telescope remotely and analyzing the data produced.

Additionally, more Grid-inspired concepts have been proposed, such as the Access Grid intended to support group-to-group video-based interactions across the Grid, the Semantic Grid as a synthesis of grid technology and the Semantic Web technology.

---

[1] http://www-306.ibm.com/software/globalization/terminology/gh.jsp

1.1.1.2 Brief History

The term Grid Computing sprang up in the early 1990s as a metaphor for making the utilization of computer power as easy as that of an electric power grid. The basic ideas and concepts were firstly brought together and further formalized by Ian Foster, Carl Kesselman and Steve Tuecke (Foster I, Kesselman C, Tuecke S, 2001).

The early Grid development was mostly propelled by the massive computational requirements of scientific applications. During that time the most prominent development was the Globus Toolkit (Foster I, 2006)[2], led by Ian Foster. As the concept of the Grid evolves, the open source Globus Toolkit has included software services and libraries for not only CPU cycle sharing and disk storage management, but also many other non-high-performance-computational-oriented components, such as those for data integration, resource monitoring, security provisioning, agreement negotiation, notification mechanisms, and dynamic service aggregation. It provides a wide variety of choices of the basic building blocks for developing various grid application developments and a real-time running environment. Similar toolkits and development endeavors such as the Unicore[3], CGSP [4] emerged quickly and world-wide.

With the rapid swelling of the Grid community and because of the well-recognized fact that the success of Grid technology heavily relies upon well-defined standards, the Global Grid Forum (GGF) was established in 2001 as a standing global organization to standardize Grid specifications and solutions. Since then the GGF has produced numerous standards, specification documents and best practices. The first fruitful outcome of the GGF is the Open Grid Service Architecture and Open Grid Service Infrastructure (OGSA/OGSI)(Foster I, Kesselman C, Tuecke S, 2002). The goal of OGSA is to standardize all of the most basic services that one can find in typical grid applications, such as those for job management, resource management, security services, etc., while the OGSI gives more concrete specifications with respect to how Grid Services can be implemented and function coordinately as a virtual organization.

Another important event that happened in grid history was the convergence of core grid standards and web service standards. WSRF (the Web Service Resource Framework) (Humphrey M, Wasson G, et al, 2005), as a successor of OGSI, was first announced in 2004 by many key players from both the grid computing community and the web service community. WSRF, as a result of the merging of the Grid and Web Services, provides the standards and a set of operations that make web services stateful and supports the fine-grained control and management of the resource status, which was

[2]Globus Toolkit: http://www.globus.org/

[3]Unicore: http://www.unicore.org/

[4]China Grid Supporting Platform: http://www.chinagrid.edu.cn/cgsp/