IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

**ITJ2912**

# Information Management for Computational Grids

Wei Jie, Institute of High Performance Computing, Singapore
Tianyi Zang, Institute of High Performance Computing, Singapore
Terence Hung, Institute of High Performance Computing, Singapore
Stephen J. Turner, Nanyang Technological University, Singapore
Wentong Cai, Nanyang Technological University, Singapore

## ABSTRACT

*Information Service is a key component of a Grid environment and crucial to the operation of a computational Grid. This paper presents an Information Service for a computational Grid Virtual Organization (VO). The main functionality of this Information Service is the provision of information essential for applications running on a computational Grid. This Information Service is a hierarchical structure that consists of VO layer, site layer, and resource layer. We propose different models of information data organization for Information Service, and experiments were conducted to evaluate the performance of these models. Based on the experimental results, we further introduce the data organization model for our Information Service. The implementation of the Information Service is based on the Globus Toolkit and complies with the OGSI (Open Grid Services Infrastructure) specifications. In this paper, some application experiences of using the Information Service are also presented.*

*Keywords:* *computational Grid; Grid computing; Grid Service; Information Service; meta-data*

## INTRODUCTION

A Grid (Foster, 1999) is concerned with coordinated resource sharing and problem solving in dynamic, multi-institutional Virtual Organizations (VO; Foster, 2001). The Information Service (Plale, 2002) is a key component in such a Grid environment. It provides fundamental mechanisms for information discovering and monitoring and serves as an underlying service for other Grid services, such as meta-scheduler, execution management service, and performance diagnosis. However, information management in a Grid environment is a challenging issue, because the information belonging to different organizations is characterized as diverse, dynamic, heterogeneous, and geographically distributed. The autonomy of resource owners needs to be honored, along with their local resource management and usage policies.

The Open Grid Services Architecture (OGSA) (Foster, 2002) was presented to address the challenges in a dynamic, heterogeneous, and geographical Grid environment, and it provides a universal resolution for the Grid Information Service. The OGSA is a formulation of Grid services as a special subset of Web Services, using standard Web service definitions and protocols (Curbera, 2002). By providing a set of operations from which all Grid services are implemented, the OGSA allows consistent resource access across multiple heterogeneous platforms with local or remote location transparency. It also allows the composition of services to form more sophisticated services, regardless of how the services are implemented, and supports integration with various underlying native computing platforms. The OGSI (Open Grid Services Infrastructure) (Tuecke, 2003) is a formal and technical specification of the concepts described in the OGSA, including Grid services.

The Globus Toolkit 3 (GT3, 2004), a Reference Implementation of the OGSI specifications, provides MDS-3 (MDS-3, 2004), which is an OGSI-compliant framework for information management for a Grid. MDS-3 provides interfaces for information operations like accessing, aggregating, generating, and querying of service data. MDS-3 also provides a standard mechanism for registration, polling, and notification/subscription of service data. But basically, MDS-3 only presents a framework for information management within an OGSI-compliant environment. It does not address information management and organization in one or multiple Virtual Organizations. In addition, very few results have been published that quantitatively study the performance of an MDS-3 based or an OGSI-compliant Information Service. Other related works on Grid Information Service includes GMA (Tierney, 2002), Hawkeye (Hawkeye, 2004) and MCS (Singh, 2003). GMA (Grid Monitoring Architecture) is a producer-consumer model for monitoring resource information in a Grid. Hawkeye uses the monitoring agents for automatic resource monitoring and problem detection within a distributed system. MCS (Metadata Catalog Service) introduces the metadata mechanism for information and data management in a data Grid environment.

In this paper, we present a VO-oriented Information Service within the OGSA framework for computational Grids (Fox, 2001). First of all, the information model for a computational Grid is discussed. Then, the architecture of this Information Service is described. We proposed two models of data organization for Information Service. To evaluate the performance of these models, we designed and conducted a set of experiments. The analysis of results is given. Based on the experimental results, we further introduce the mixed model of data organization for our Information Service. In addition, the application experiences of using our Information Service are presented. Finally, we conclude the paper and give our future research directions.

## INFORMATION MODEL FOR A COMPUTATIONAL GRID

The first issue to construct an information service is its information model. The information model defines the types of data an information service should provide. As an underlying service, our Information Service intends to provide required information to support applications' execution in a computational Grid. We specify four types of information that are vital in a computational Grid environment (see Figure 1):

- **Resources Information.** An end user or a job submission service needs to know what resources are available in a Grid. In order to achieve this goal, the knowledge of all kinds of computational resources involved in a Grid environment should be provided. Some typical information of computational resources includes machine architecture, CPU number and speed, memory and disk capacity, operating system, and so forth. Increasingly, we also need to capture and provide information, such as available software services, license constraints, and so forth.
- **Job Status.** Job running status is the key information that should be provided by an

Efficiently Compositing and Optimizing the Quality of Heterogeneous Services

Hanhua Chen and Hai Jin (2014). *International Journal of Web Services Research (pp. 76-95).*

www.igi-global.com/article/efficiently-compositing-and-optimizing-the-quality-of-heterogeneous-services/122816?camid=4v1a