

# 그리드 정보검색 시스템을 위한 OGSA-DAI 기반 확장된 Collection Manager 서비스 설계

김혁호\*, 김양우\*  
\*동국대학교 정보통신공학과  
e-mail : {hulegea, ywkim}@dongguk.edu

## Design of Advanced Collection Manager Service for Grid-IR System Based on OGSA-DAI component

Hyukho Kim\*, Yangwoo Kim\*  
\*Dept. of Information and Communication Engineering, Dongguk University

### Abstract

The interest in the access and integration of distributed massive data resources has increased recently. This paper presents the Advanced Collection Manager(CM) service with OGSA-DAI component which can access and integrate the distributed data resources. The Advanced CM service supports the data resource of various types. And it can provide the query, updating, transforming and delivering data via cooperating with other services in Grid Information Retrieval(Grid-IR or GIR) System. As a result, it can access and manage the data resource more flexible and efficient.

### 1. Introduction

The interest in the access and integration of distributed massive data resources has increased recently. And there need the treatment methods to satisfy this requirements efficiently. Data Grid[1, 2] provides API and lots of services which can satisfy those requirements. OGSA-DAI(Open Grid Service Architecture – Data Access and Integration) component[3] is the middleware for Data Grid. It provides services of various types and API which can access and integrate the distributed data. This paper presents Advanced Collection Manager (CM) Service architecture. This CM service manages the collected data resources in Grid-IR system[4, 5]. We proposed to enlarge the existed CM service model with OGSA-DAI component. Therefore it can access and manage the distributed data resources efficiently.

The rest of this paper is organized as follows. In Section 2, we briefly explain the related work. Section 3 presents the Grid Information Retrieval System and Grid-IR Services. And then we explain the overall architecture of Advanced CM Service. Finally, we present our conclusion and future works in Section 4.

### 2. Related work

#### 2.1 Globus Toolkit

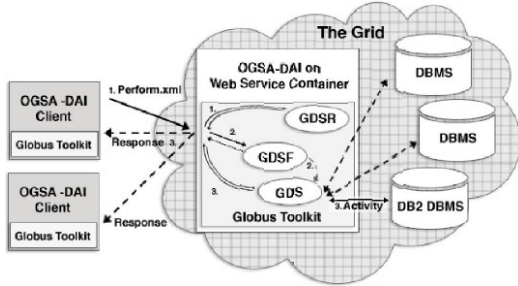
The Globus Toolkit[6] is a fundamental enabling technology for the Grid, letting people to share computing power, databases, and other tools securely online across corporate, institutional, and geographic boundaries without sacrificing local autonomy. The toolkit includes software services and libraries for resource monitoring, discovery, and management, plus security and file management.

The toolkit includes software for security, information

infrastructure, resource management, data management, communication, fault detection, and portability. It is packaged as a set of components that can be used either independently or together to develop applications. Every organization has unique modes of operation, and collaboration between multiple organizations is hindered by incompatibility of resources such as data archives, computers, and networks. The Globus Toolkit was conceived to remove obstacles that prevent seamless collaboration. Its core services, interfaces and protocols allow users to access remote resources as if they were located within their own machine room while simultaneously preserving local control over who can use resources and when.

#### 2.2 OGSA-DAI

OGSA-DAI[7, 8, 9] is one kind of integrated middleware, which realize the data access and the integration in Grid environment. OGSA-DAI is constructed into a tool box with many expansion points, so as to let the developers expand the ability to meet the specific requirements. The middleware of OGSA-DAI permits the same data resources to carry on the synthesis in a compatible OGSA system structure. The Grid service in the OGSA-DAI provides the basic operation. The standard module is used for accessing and controlling data information and the resources. Also OGSA-DAI provides expansibility mechanism, thus it can increase more activities which the users define and are executable in the OGSA-DAI. The implementation of OGSA-DAI is shown in Figure 1.



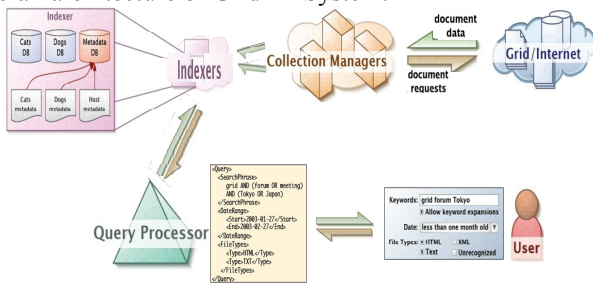
(Figure 1) Implementation Process of OGSA-DAI

At present the middleware of OGSA-DAI is in research phase and development continuously. The main flaws of the existing middleware of OGSADAI are: Lacking the database automatic registration service; the capability to tolerate a bad database connection; the data integration of the isomerism database is not good.

### 3. System Design

#### 3.1 Grid Information Retrieval System

GIR applies the tools of grid computing to IR to provide a common infrastructure for distributed IR. It also brings the capabilities of IR to grid computing. The basic idea of GIR is to define an IR system in terms of three functional components, implemented as grid services: Collection Manager service (CM), Indexing/Searching service (IS), and Query Processing service (QP). These services are autonomous, and being grid services, they are distributed. Since they can be created dynamically and in any combination of Virtual Organizations, they can be used to create new IR systems or link existing ones together in an interoperable network of IR services. Figure 6 shows the overall architecture of Grid-IR System.



(Figure 2) Grid-IR Architecture

The Collection Manager (CM) is concerned with collecting and managing source documents intended to be indexed, searched, and retrieved via one or more Grid-IR services. It does this by retrieving specified documents from various locations (remote or local), preprocessing and managing them in a local store, and providing them to clients according to specified rules. “Client” here normally refers to the Index/Search service, which uses CM’s as the sources of documents that will be indexed and searched.

The Index/Search interface (IS) conducts jobs which are assemble or indexing of documents from the variety of existing sources. The IS also creates and manages data structures needed to provide searching capabilities. (These data structures are normally referred to as an “index.”) The

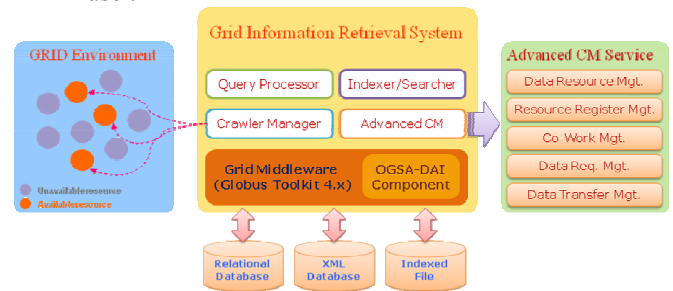
IS also exposes a “search interface,” which is a set of functions that allow searching of the indexed document collection.

The Query Processor (QP), the third architectural component of Grid-IR, is responsible for managing queries and result sets (i.e. response sets from searching operations). As mentioned earlier, QP provides a search interface identical to that of IS. In other words, for purposes of searching, QP is a “virtual IS.” It has no indexing capability of its own (and provides no indexing interface) but serves as a search only gateway to one or more IS’s.

#### 3.2 Advanced CM Service Architecture

The Grid-IR system architecture is shown in Figure 3. As I mentioned above, Grid-IR system consists of Collection Manager service, Indexer/Searcher service and Query Processor service. In this paper, we designed the Advanced CM service with OGSA-DAI component. In this paper, the proposed Advanced CM service has the function following:

- Data Resource Manager manages the existing data resources. It is in charge of create, update and delete of data resources.
- Data Resource Register Manager manages the registration of the new data resource actually. It makes the space for new data resource and manages the information about the new data resource.
- Co-Work Manager is in charge of co-operation with other services in Grid-IR system. It sends a message to IS service to make the new index file for the new data resource. And it sends a work to Crawler Manager for collecting the data.
- Data Request Manager is in charge of the request classification according to work types. And it notify to the suitable and available service and resources.
- Data Transfer Manager is in charge of the data transmission. When the user searches the specific data using the QP service, it transfers the result data to the user.



(Figure 3) Advanced CM Service Architecture

The proposed Advanced CM service can co-operate with other services in Grid-IR system because it is implemented as a Grid service[10]. And it can access and integrate the distributed data using OGSA-DAI component. Therefore it can support the data resource of various types and manage the distributed data resource more flexible and efficiently.

### 4. Conclusion and Future Works

In this paper, we designed the advanced collection manager service model in Grid Information Retrieval(Grid-IR, GIR) System with OGSA-DAI component which can

access and integrate the distributed data. The proposed advanced CM service has the following features: Firstly, Advanced CM service can co-operate with other service in Grid-IR system because it is a grid service. Secondly, it can create, update and delete the data resource with OGSA-DAI component, so it can manage the data resource more flexible and efficient.

We have to implement the advanced CM service and need to port it into Grid-IR System for co-operation with other services. Also we plan an experiment in comparison with the existed CM service for system stability and excellence, so that we verify the practical proposed service's efficiency.

### References

- [1] I. Foster, C. Kesselman, S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", International Journal of High Performance Computing Applications, 15(3):200-222, 2001.
- [2] I. Foster, C. Kesselman, J. Nick, S. Tuecke. "The Physiology of the Grid-An Open Grid Services Architecture for Distributed Systems Integration", Global Grid Forum, June 22, 2002.
- [3] OGSA-DAI. Available from URL: <http://www.ogsadai.org/>.
- [4] N. Nassar, K. Gamiel, and J. Morris, "Grid Information Retrieval Architecture", [http://www.gir-wg.org/papers/Grid\\_Information\\_Retrieval\\_Architecture.pdf](http://www.gir-wg.org/papers/Grid_Information_Retrieval_Architecture.pdf), September 2003.
- [5] Yangwoo Kim, Seungha Lee, Pillwoo Lee and Newby G. B. "Grid Information Retrieval Management System for Dynamically Reconfigurable Virtual Organization", Grid and Cooperate Computing 2006(GCC 2006) , Fifth International Conference, Oct. 2006.
- [6] Globus Toolkit Primer, See [http://www.globus.org/toolkit/docs/4.2/4.2.0/key/GT4\\_Primer\\_0.6.pdf](http://www.globus.org/toolkit/docs/4.2/4.2.0/key/GT4_Primer_0.6.pdf).
- [7] Xiufen Fu, Haishui Xu, Wenguo Hou, Yangseng Lu, and Boxing Chen, "Research of the Access and Integration of Grid Database", Proc. of the 10<sup>th</sup> International Conference on Computer Supported Cooperative Work in Design, 2006.
- [8] Database Access and Integration Services Working Group, DAIS-WG. Available from URL: [http://www.ogf.org/gf/group\\_info/view.php?group=dais-wg](http://www.ogf.org/gf/group_info/view.php?group=dais-wg).
- [9] N. Hardman, A. Borley, J. Magowan, OGSA-DAI: A look under the hood: Part 1: Architecture and database access, IBM United Kingdom Limited, January 2004.
- [10] I. Foster, C. Kesselman, J. Nick, S. Tuecke. "Grid Services for Distributed System Integration", Computer, 35(6): 37-46, 2002.