UDMGrid: A Grid Application for University Digital Museums [1]

Xiaowu Chen, Zhi Xu, Zhangsheng Pan,
Xixi Luo, Hongchang Lin, Yingchun Huang, Haifeng Ou

The Key Laboratory of Virtual Reality Technology, Ministry of Education,
School of Computer Science and Eng., Beihang University, Beijing 100083, P.R. China
{chen, xuzhi}@vrlab.buaa.edu.cn

Abstract. Because the eighteen online university digital museums of China confront a problem that the multi-discipline resources at digital museums are isolated and dispersed without sufficient interconnection, ChinaGrid (China Education & Research Grid) supports a project named UDMGrid (University Digital Museum Grid), which studies a grid application for University Digital Museums using grid technology, especially information grid technology. According to the analysis on the problem, UDMGrid focuses on resource sharing, information management and information service. This paper presents research work on UDMGrid's framework, metadata, replica management, application server, etc.

1 Introduction

Eighteen featured university museums have already been digitized, which mainly relate to Geology & Geography, Archaeology, Humanities & Civilization, and Aeronautics & Astronautics. These digital museums play an important role in the fields of education, scientific research, as well as specimen collection, preservation, exhibition, and intercommunication.

But, these digital museums dispersed on different nodes in CERNET (China Education and Research Network) have several shortcomings: first, the present digital museums lack sufficient interconnection, and only connected by web links. This brings problems on management such as information redundancy and storage resources waste. And, the stored information is isolated in the so-called information island, which makes it difficult to intercommunicate and share resources among museums; second, these digital museums are unable to filter, classify, and process the stored raw information to fit for different kinds of requirements of visitors, and all the visitors having different knowledge background now access information through the same channel [3].

The information grid technology gives a better solution to these digital museums, since information grid is good at resource sharing, information

[1] This paper is supported by China Education and Research Grid (ChinaGrid) under the grant No.CG2003-GA004, and National 973 Program of China under the grant No.2002CB312105.

management, and information service. In an information grid, the raw data from various sources is standardized, classified, and organized to be information; then, the grid digs out the information's internal relations; finally, uniform and transparent information service is presented to users [4][5]. The recent hot spots of information grid research focus on the grid architecture, information representation and organization, information interconnection and uniformity, security, robust, etc. [6]The San Diego Supercomputer Centre (SDSC) at UCSD is one of the original research institutes, which pays attention to the information grid and relationship among data, information and knowledge. SDSC has devoted to the abstraction, organization and management of data, and advised to manage them by layers.UK e-Science concerns three levels including computational and data grid, information grid, knowledge grid. [8] It discusses the case of automating the process from raw data to information, and further to knowledge, and it plans to create new types of digital libraries for scientific data.[7] The e-Science proposes to integrate distributed data, computing and storage resources, and implement further data mining and knowledge discovery in grid environment. IBM has been actively involved in providing access to heterogeneous files, databases, and storage systems, and it concentrates on sharing of data for processing and large-scale collaboration.[9] IBM supports virtualizing data across diverse formats to solve the challenge of accessing data stored in different format, supports using of Storage Access Network (SAN) technology to solve the poor storage resource utilization, supports developing a replication solution to solve the problem of having to move a large volume of data across network to facilitate remote processing. Institute of Computing Technology (ICT), Chinese Academy of Sciences, emphasizes the information management and processing in the information grid. ICT has developed some applications based on the Vega Grid, including the Vega Information Grid (Vega-IG)[10][4] and Railway Information System [11]. The UK-wide digital museums have been built since 1999. Recently, UK planned to integrate the collections of most famous museums by Grid, so as to offer an interconnected and interactive environment for learning [12]. For example, visitors can receive the information service filtered according to their requirements, such as age and interests.

So, it is necessary for ChinaGrid (China Education & Research Grid) to research on the UDMGrid [13], whose purpose is to integrate the enormous dispersed resources of various digital museums, to share the resources effectively and eliminate the information island, to filter and classify the collection information, and to provide appropriate information service to users according to their knowledge levels and motivation through unified grid portal. The UDMGrid will promote the education and scientific research in universities as well as in public, and also improve the public infrastructure facilities development in the Information Age

The rest of this paper is organized as follows. Section 2 describes UDMGrid's system framework, metadata, replica management, and A. Server. In section 3, we present the description and analysis on a prototype of UDMGrid. Finally, section 4 draws its conclusion and outlines future work.

2 Design of UDMGrid

Using information grid technology, the UDMGrid is designed to overcome the existent shortcomings of present digital museums.

The UDMGrid is composed of three parts, as illustrated in Fig.1. The Grid Supporting Middleware offers basic grid services, such as authentication, authorization, file transfer, index service, etc. Bolstered by Grid Supporting Middleware, the Grid Application Platform is developed and provides high level services to Grid Application. Using services through their APIs, Grid Applications can easily be developed and deployed oriented to user's requirements.

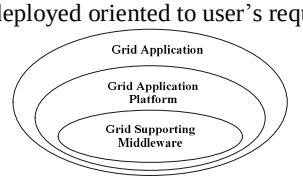

Fig. 1. the relationship of Grid Supporting Middleware, Grid Application Platform and Grid Application

Some developed grid software can be applied in UDMGrid as the Grid Supporting Middleware, such as Globus Toolkit 3.0[14], Unicore [15], VegaGos [16] and WebSASE4G [3].

2.1 System Framework

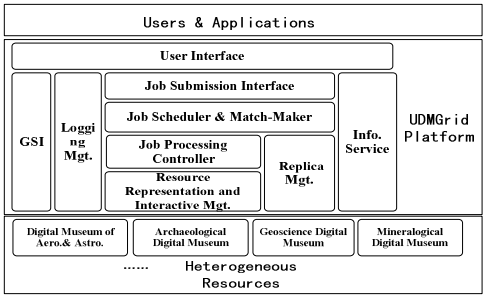

Fig. 2. the framework and function modules of UDMGrid

As shown in Fig. 2, the function modules can be divided into four parts: User Interface, Job Management, Job Monitor, and Resource Management.

User Interface is the entrance for users to grid, which provides service environment for job submission and job information acquirement. Before submitting the job, user interface will pre-standardize user's request to job descriptions following specified criterions.

Job Management takes charge of job processing control and management in the workflow [17], including three modules: Job Submission Interface, Job Scheduler & Match-maker, and Job Processing Controller.

Job Monitor, composed of Grid Security Infrastructure (GSI) [18] and Logging Management [19], monitors the status of UDMGrid system and the procedure in job processing. It provides information feedback when requested by other modules.

Resource Management comprises Information Service module and Replica Management module. The functions of Resource Management part include providing resource information for match-making, metadata description and organization, replica management, etc. It is a research emphasis in UDMGrid.

2.2 Metadata

It is required to describe and standardize the mass and various data present in grid using metadata. The metadata in UDMGrid is designed according to the characters of the users, the grid it self, and the different service and various collections in digital museums, such as mass digital specimen information.

The metadata in the UDMGrid mainly includes: System Status & Configuration metadata, Replica metadata, Job metadata, Resource metadata, User metadata, which mainly service for system monitoring, replica management, job monitoring, resource accessing, and grid security. The detailed classification of the metadata in UDMGrid is depicted as following:

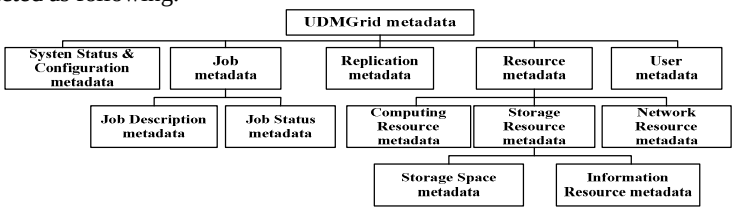

Fig. 3. the metadata in UDMGrid

System Status & Configuration metadata, stored in the Logging Management module, describes the status and configuration information of the grid system. It services for the system monitoring and provides information for cooperation of modules.

Replica metadata expresses the attributes and status of the replica. It provides the mapping information between logical files and replica files.

Job metadata describes the jobs in grid, including Job Describing metadata and Job Status metadata. The Job Describing metadata describes the content of submitted jobs, and provides information for the match-making between resources and job. Its design is suitable for data operation and accords with the characters of job in UDMGrid, such as time-sensitive and interactive.

Resource metadata describes the related information about computing resource, storage resource, and network resource. Because the computing resource in UDMGrid

is generally represented as A.Server, the Computing Resource metadata is mainly designed according to the attributes of A.Server (the details is presented in part 2.4). The Storage metadata includes Storage Space metadata and Information Resource metadata. The Storage Space metadata describes the information of storage resource which can provide storage service, and the Information Resource metadata describes the attributes and some content information of the information resource. The Network resource metadata describes the status and parameters of networking.

User metadata describes information concerning users and resource providers. This information of users and providers is required in authorization, authentication, accounting, etc.

Among the metadata listed above, the Information Resource Metadata is one of the most important types in UDMGrid, since how to describe the numerous and various collections in these university digital museums is very challenging. The typical collections in university digital museums are images, text, video, etc. Take image for example, a picture about P-61 airplane in Digital Museum of Aero. & Astro. can be described as following:

Table 1. a example of Information Resource Metadata in UDMGrid

title	P-61
type	Picture
format	Jpeg
coverage	Aviation
identifier	http://digitalmuseum.buaa.edu.cn/picture/craft/black.jpg
description	P-61 is the first battle plant equipped with radar
rank	0
…	…

2.3 Replica Management

Replica Management provides transparent access to distributed data and improves the efficiency of data accessing and the quality of information service. In the UDMGrid, the Replica is managed in several domains, each domain comprise five modules: Replica Catalog, Replica Manager, Replica Catalog Query, Replica Selector and Consistency Manager (see Fig. 4).

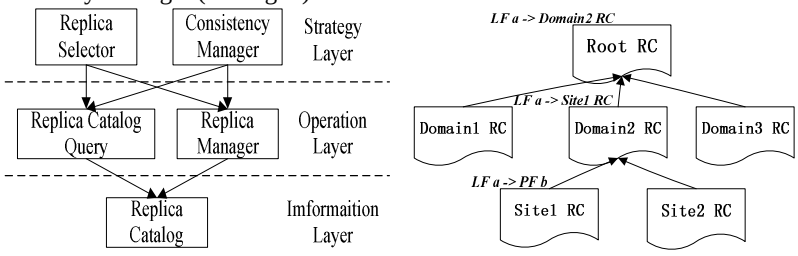

Fig. 3. modules of replica management **Fig. 4.** hierarchical Replica catalog

The five modules are explained as following:

Replica Catalog (RC) stores the mapping information from logical file name (LFN) to physical file names (PFN). As illustrated in Fig. 5, the Replica metadata is stored in Site RC; the upper Domain RC contains the mappings from LFNs to the Site RCs; the Root RC contains mapping from LFNs to Domain RCs. For example, if a mapping from LF a to PF b is stored at Site1 RC, the Domain2 RC will keep a mapping record form LF a to Site1 RC and the Root RC will keep a mapping form LF a to Domain2 RC[20]

Replica Manager executes actual creation and deletion of replica within its domain and updates the RC.

Replica Catalog Query provides query service and a unified view of the replicas in grid.

Replica Selector receives request for replicas from its domain and searches the RC for a proper replica according to the request. Considering the result of searching, the Replica Selector will decide to select a proper replica or to create a new replica. In different storage conditions, different strategies will be used to make the decision. If create, it will inform the Replica Manager to apply the actual actions. In addition, Replica Selector maintains a table that records the accessing frequencies to logical files in this domain.

Consistency Manager detects the change of master, and informs Replica Manager to maintain the corresponding replicas. In the UDMGrid, only the master can be modified or deleted. If modified, the master will be republished and all the replicas related to it will be deleted.

2.4 Application Server

Application Server (A.Server) is a kind of deployed computing resource which provides specialized and persistent service in the UDMGrid. This specialized service is pre-designed and deployed according to the requirements of users and applications.

In UDMGrid, A.Server gets raw data about the collections from dispersed university digital museums, explores data's internal and external relationships, and processes it to information according to user's requirements. Take information classifying for example: a middle school teacher wants to query some information about "china's aeronautics history". The serving A.Server will firstly find all kinds of correlative data, such as images, video, files, etc. in the university digital museums, and filter useful information according to the job description and user's information, such as "teacher". In the end, the result will be sent back by time order.

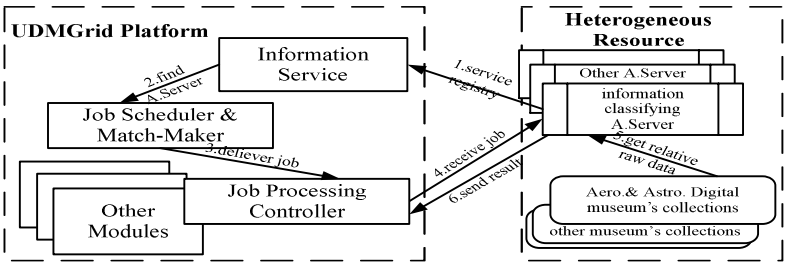

Fig. 5. A.Server in the UDMGrid

Fig.6 illustrates the procedure of information classifying service. Firstly, A.Server providing such service must exist and has registered its service as Grid Service in the Information Service module. When a job arrived, the Job Scheduler & Match-Maker module will seek proper A.Server in the Information Service module, and deliver the job to the chosen A.Server, here is information classifying A.Server. While processing the job, A.Server may need to get a lot raw data about different digital museums' collections. The results will return to Job Processing Control module.

3 Prototype of UDMGrid

In this prototype, we propose to integrate the information resource of several university digital museums, and provide integrative and intelligent information searching, filtering and classifying service through grid Portal. The university digital museums involved include the Digital Museum of Aeronautics and Astronautics (BUAA)[21], the Archaeological Digital Museum (SDU)[22], the Geoscience Digital Museum (NJU)[23], and the Mineralogical Digital Museum(KUST)[24], etc.

The services and resources in each university digital museum compose one site, and the Fig.7 shows the architecture of site BUAA. The Collection provide service gives access to digital museum's collections as raw data.

and these Collection provide services may be acquired in several different sites.

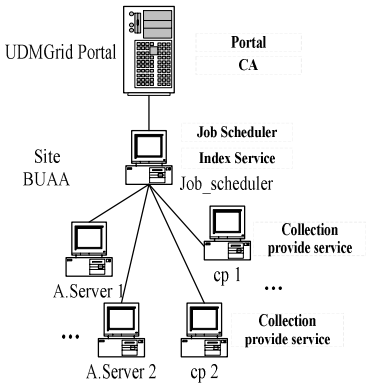

Fig.7. the prototype in site BUAA

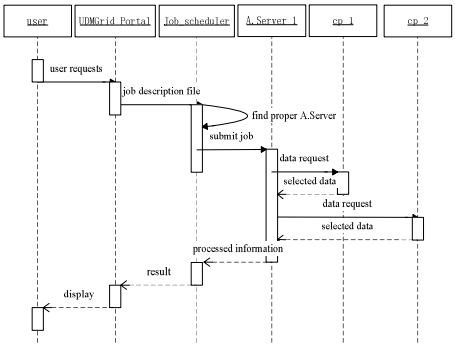

Fig. 8. Interaction in the BUAA site

The interactive process is presented in Fig.8. A user submits his request to UDMGrid from Portal. Then the portal formats user's requests to job description file and delivers it to Job_scheduler. Then the Job_scheduler searches the Index Service for proper A.Server and Collection provide services which provide materials for processing in A.Server. And the Job_scheduler creates a job according to the job description file. Then this job is submitted to the chosen A.Server. While processing, the A.Server needs to get raw data from Collection provide services as materials, and these raw data may be acquired from different sites, for example, the A.Server located in site BUAA may need resource from site NJU. After processing, the result will be send back to Job_scheduler and finally displayed to user.

4 Conclusion and Future Work

We present an information grid infrastructure for UDMGrid, and our research work concentrates on system framework, metadata, replica catalog, A.Server, etc. Its purpose is to solve the information island problem of established university digital museums, provide easy resource sharing, and offer intelligent information filtering and classifying service. As an important part of ChinaGrid Project, the UDMGrid will be the next generation information service infrastructure for museums and play an important role in the fields of learning, teaching, and public education.

The research is just at the beginning and needs continued efforts on the key technologies, such as architecture, metadata description, replica management, resource information service, A.Server, and the implementation

References

[1] The digital museum research group at China Univ. of Geosciences, Hankou Campus, The methods and steps of building university digital museum;

[2] University Digital Museums, http://www.edu.cn/20020118/3018035.shtml;

[3] ChinaGrid project--BeiHang University Research Group, Grid Supporting Platform based on Web Services and its Typical Application, ChinaGrid Newsletter No.1;

[4] Xu Zhi-Wei, Li Xiao-Lin, and You Gan-Mei, Architectural Study of the Vega Information Grid, Journal of Computer Research and Development;

[5] Keith G Jeffer, EU GRIDs Activity: Information Issues;

[6] Reagan W. Moore, Digital Libraries, Data Grids, and Persistent Archives, San Diego Supercomputer Center(SDSC);

[7] National e-Science Centre, http://www.nesc.ac.uk/;

[8] Tony Hey and Anne Trefethen, the Data Deluge: An e-Science Perspective;

[9] The information grid, Rob Vrablik, Melissa Hyatt, IBM developerWorks;

[10] Xu Zhi-Wei and Li Wei, Research on Vega Grid Architecture, Journal of Computer Research and Development;

[11] Railway information grid Research group at Institute of Computing Technology, CAS, Information Grid and Railway Information Grid platform;

[12] UK-wide digital museum-National Grid for Learning (NGfL), http://www.ngfl.gov.uk;

[13] ChinaGrid Project, Ministry of Education, http://www.cergrid.cn/;

[14] Globus project, http://www.globus.org/;

[15] UNICORE forum, http://www.unicore.org/;

[16] Vega Grid project，http://vega.ict.ac.cn/;

[17] Massimo Sgaravatto, INFN-EDG, WP1: Workload Management (DataGrid-01-D1.4-0127-1_0);

[18] Grid Security Infrastructure, http://www-unix.globus.org/security/;

[19] Francesco Giacomini, INFN-EDG, definition of architecture, technical plan and evaluation criteria for scheduling, resource management, security and job description (DataGrid-01-D1.2-0112-0-3);

[20] Ben Segal, CERN-EDG, Data Management (WP2) Architecture Report;

[21] The Digital Museum of Aeronautics and Astronautics (BeiHang University) http://digitalmuseum.buaa.edu.cn/;

[22] The Archaeological Digital Museum (Shandong University) http://museum.sdu.edu.cn/index/index.asp;

[23] The Geoscience Digital Museum (Nanjing University) http://202.119.49.29/museum/default.htm;

[24] The Mineralogical Digital Museum (Kunming Univ. of Sci. & Technol.) http://www.kmust.edu.cn/dm/index.htm.