

An e-Learning Platform Based on Grid Architecture

CHAO-TUNG YANG AND HSIN-CHUAN HO
High-Performance Computing Laboratory
Department of Computer Science and Information Engineering
Tunghai University
Taichung, 407 Taiwan
E-mail: ctyang@thu.edu.tw

The overall popularity of the Internet has helped e-learning become a hot method for learning in recent years. Over the Internet, learners can freely absorb new knowledge without restrictions on time or place. Many companies have adopted e-learning to train their employees. An e-learning system can make an enterprise more competitive by increasing the knowledge of its employees. E-learning has been shown to have impressive potential in e-commerce. At present, most e-learning environment architectures use single computers or servers as their structural foundations. As soon as their work loads increase, their software and hardware must be updated or renewed. This is a big burden on organizations that lack sufficient funds. Thus, in this study we employ a kind of Grid Computing technology, called the "Data Grid" to integrate idle computer resources in enterprises into e-learning platforms, thus eliminating the need to purchase costly high-level servers and other equipment.

Keywords: e-learning, grid computing, data grid, partner schools, grid portal

1. INTRODUCTION

E-learning is overtaking conventional classroom teaching methods. Via this approach, knowledge can be more flexibly propagated and absorbed irrespective of the teaching personnel and learners. Many schools and enterprises are adopting e-learning as their principal instruction or training method [1].

An e-learning Platform generally requires a high-capacity storage device, such as a Disk Array, Network Attached Storage (NAS), or Storage Area Network (SAN), to store teaching content. Schools with insufficient budgets often cannot afford this expensive equipment. However, upon examining their current equipment, they may discover many idle computer resources that can be utilized. For example, a personal computer with a 20GB HDD bought two years ago, if used solely for online browsing and word processing, requires only about 6GB, 30%, of its total HDD capacity; the remaining 14GB will be idle. Therefore, we can connect computers used by an organization for teaching or administration together and combine their idle storage space by means of Grid Computing technology. The storage space thus obtained will be equal to that of an expensive storage device.

This study focused on connecting idle storage devices on campuses using Data Grid techniques to serve as substitutes for expensive storage devices, such as Disk Arrays or

NAS, and as e-learning platform storage devices. Campuses that have insufficient budgets can services and obtain enormous teaching resources. With continuing participation by educational organizations, the expected accomplishments achieved and experience gained through this research can serve as a reference for agencies seeking to develop e-learning and can help them save time and money in developing similar systems.

2. E-LEARNING TECHNOLOGIES

Developments in communication and networking technology in recent years have led to gradual improvement of network bandwidth and quality, and have made real-time transmission of high-quality video and audio a reality. Multimedia transmission and related network application technologies have helped fuel the development and popularization of such e-learning techniques as Distance Education, Video Conferencing, and Video on Demand.

Distance Education is an essential link in the application of e-learning. Whether is used in academic courses on campuses or in training courses within enterprises, Distance Education is a very effective learning method. With conventional teaching methods, learners must meet at the same time and location. When enterprises have numerous learners [2], exorbitant expenditures of money and time are often required. The advantage of Distance Education is that it can solve the problem of geographical location and enable students at remote sites to feel as if they are attending classes in a conventional classroom. Furthermore, it can conserve money and eliminate the time and energy consumed in traveling between classes.

Although Distance Education has many advantages, its biggest drawback is the required investment in equipment. In order to help more enterprises and schools apply Distance Education, this study adopted Grid Computing technologies to build an e-learning platform that features expandability and resource-sharing capability. Current Distance Education Systems for e-learning can roughly be classified into the three types described below.

2.1 Multicast

This type enables teachers and students to be at different locations. Through the use of networking technology, classroom video/audio and multimedia teaching materials can be transmitted in real-time to remote-site classrooms. It also allows two-way real-time communication between teachers and students in remote-site classrooms.

2.2 Virtual Classroom

This type uses a management system to simulate the experience of attending class in a classroom (including teacher lectures, written examinations, specific assignments, question-and-answer sessions, student questions, oral quizzes, etc). Teachers and students can be linked to the management system at any time via the Internet.

2.3 Video on Demand

This type uses a technology called VOD (Video on Demand). Students obtain teaching-learning materials via the Internet using computers or television sets combined with set-top boxes. This enables the distance learning process to proceed in accordance with each student's personal learning speeds by controlling the broadcasting process.

The teaching mode studied here combines Multicast and VOD teaching methods. Real-time video teaching is performed at fixed times, with the rest of the time reserved for web teaching. For instance, video teaching courses are recorded as video stream files after every class and broadcast online for the convenience of learners who cannot watch them at the regular times.

3. GRID INFRASTRUCTURE

Grid infrastructures support the sharing and coordinated use of resources in dynamic global heterogeneous distributed environments. This includes resources that can manage computers, data, telecommunication, network facilities, and software applications provided by different organizations [3]. A Grid is a collection of distributed computing resources available over a local- or wide-area network that appears to an end user or application as one large virtual computing system.

3.1 The Concept of Grid Computing

Distance Education or e-learning can only be put into practice through the application of computer technology. However, the elements of traditional computing, including software, hardware, and the Internet, suffer from heavy workloads, leading to slower computing speeds. In order to solve this problem, it is necessary to upgrade or renew software and hardware, which is often very difficult for institutions that have limited budgets.

Thus, we employ a technique called "Grid Computing" to solve the above problem. Though the use of a proper operating system and software, Grid Computing can connect scattered computer resources via Internet links. These resources include computer facilities, storage media, and various input and output devices.

As PCs or other facilities are added to the grid, computers already on the grid can enlist their help in computing work. This involves the use of local computer resources as well as powerful virtual resources in remote computers, including computing and storage capacity, application programs, databases, and I/O devices spread over the entire grid. Thus, Grid Computing can be likened to a virtual supercomputer [4].

If, for example, a host needs to finish a task in one hour, the Grid Computing system can search for other idle computers and distribute the task among them, allocating available resources as required. This efficient use of computer resources may allow the task to be completed in less than ten minutes. The computers employed can be in different organizations or locations, even in different countries. They merely share computing resources with one another via the Internet [5].

Nowadays, many educational institutions cannot renew their computer software and

hardware facilities due to a lack of funds. However, if they examine the current facilities and equipment on their campuses, they may find many idle computers that can be used. For example, if a PC containing a Pentium III CPU, 128MB of RAM, and a 20GB HDD purchased a couple of years ago is only used for browsing online or word processing, it surely has sufficient computing capability and storage space. But if it is used as a server, it will be quite inadequate. We can connect PCs like this, which are spread over a campus and used for teaching or administration purposes, and obtain a combined computing capacity may be equal to that of an expensive high-speed server. Furthermore, institutions that cannot upgrade their computer software and hardware facilities can share other institutions' resources using Grid Computing techniques.

Although many institutions have adopted e-learning, the software and hardware facilities they use vary greatly. This causes major difficulties in sharing teaching resources. Web service techniques enable the integration of different information systems within grids and solve this problem. Therefore, the idea behind the present study is to integrate Grid Computing with web services to build an e-learning platform [6]. Fig. 1 illustrates the grid computing concept.

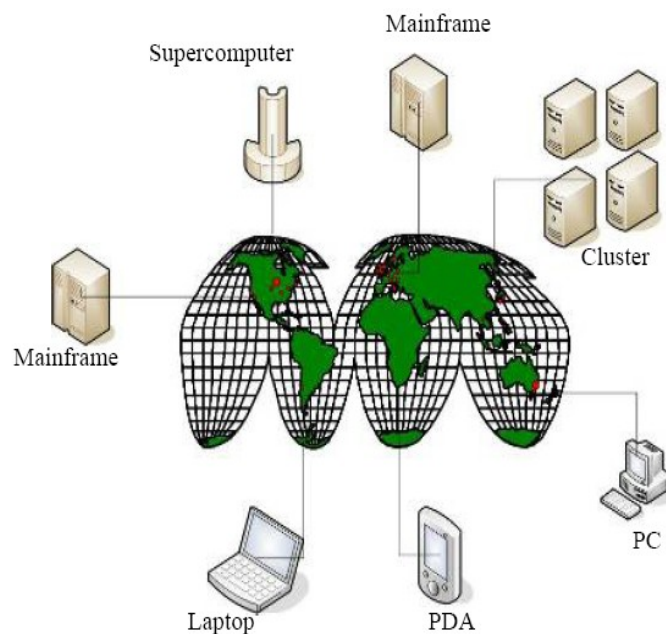


Fig. 1. Grid computing infrastructure.

3.2 Globus Toolkit

The Grid represents common properties that have very large, distributed dynamics that cross the boundaries of human organizations [7]. Due to these characteristics, its architecture must be very complex and its relationships like a maze. The vital task in realizing a Grid is establishing a common open standard. Its core is based on an open set

of standards and protocols – e.g., the Open Grid Services Architecture (OGSA) – that enable communication across heterogeneous, geographically dispersed environments. With Grid Computing, organizations can optimize computing and data resources, pool them for large-capacity workloads, share them across networks, and enable collaboration [8]. Some protocols and applications have been proposed and implemented in actual works, such as the Globus Toolkit [9].

The Globus Toolkit is an open-architecture, open-source suite of services and software libraries that support Grids and Grid applications. It provides software tools that enable coupling with computers, databases, and instruments. With Globus, jobs can be run on two or more high-performance machines at the same time, even though these machines might be located far apart and owned by different organizations. The toolkit includes software for security, information infrastructure, resource management, data management, communication, fault detection, and portability.

The Globus toolkit components most relevant to OGSA are the Grid Resource Allocation and Management (GRAM) protocol, the Meta Directory Service (MDS-2), and the Grid Security Infrastructure (GSI). These components provide the following essential elements of a service-oriented architecture:

1. GRAM protocol: this component provides for secure, reliable, service creation and management of arbitrary computations.
2. MDS-2: this component provides a uniform framework for information discovery through soft-state registration, data modeling, and a local registry.
3. GSI protocol: this component supports single sign-on, authentication, communication protection, and certification mapping.

3.3 Data Grid

Some newly developed scientific fields, such as high-energy physics and computational genomics, demand access to large amounts of data, a requirement that leads to such considerations as data file management, replicated file management and transfer, and distributed-data access management. The Data Grid infrastructure is designed to integrate data storage devices and data management services in a grid environment. Data grids consist of scattered computing and storage resources, which, though located in different countries, remain accessible to users [10].

In this study, we adopted the Globus Toolkit as the Data Grid infrastructure. The Globus Toolkit provides solutions for such requirements as security, resource management, data management, and information services. Many research projects, such as GriPhyN [11], PPDG [12], and EU DataGrid [13], are based on the Globus Toolkit.

The Globus Data Grid comes in two layers. On the Low Level are Data Grid Core Services, and on the upper layer are High Level Components [14]. Fig. 2 shows the Data Grid Architecture.

The storage system is a basic data grid component. It defines and covers all storage technologies capable of adding, deleting, reading, writing and operating file instances [15], such as HPSS, and DPSS (Distributed Parallel Storage System). Data access services are set up to access, manage, and transfer data in the storage system [16].

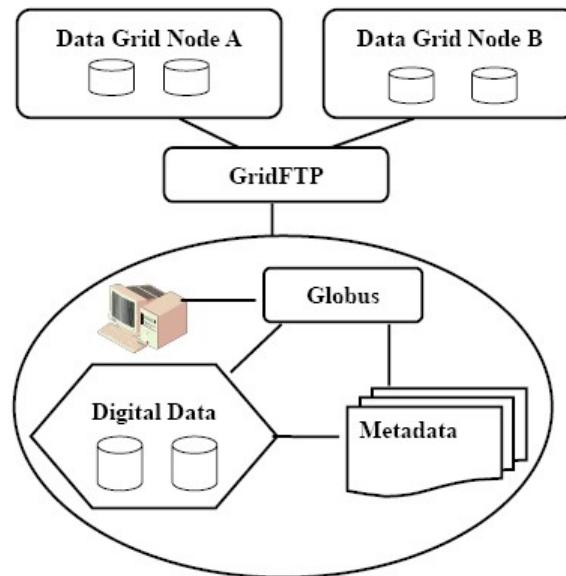


Fig. 2. Data grid architecture.

Metadata service is a way to manage and access Metadata containing Data Grid information. Application Metadata include information describing files and data environments. Replica Metadata are used to manage the replication of data objects and provide a way to map logical files to physical files.

Resource Management is responsible for the storage system, networks and other data grid resources needed to assure end-to-end efficiency, technical assessment of efficiency testing, and crucial resources. The Grid Security Infrastructure (GSI) [17] provides environment authorization and certification mechanisms for large numbers of users.

Underlying the Data Grid framework is a key element known as replica management [18], which is important for successful processing of large amounts of data by the Data Grid. It mainly decides when and where to set up replicas and provides information about replica locations, including these key functions:

1. Registration: this function adds new files to the Replica Catalog and shows where the new files may be accessed by users.
2. Creation and deletion: these functions create and delete registered replicas.
3. Publishing: this function reproduces unpublished files in destination storage systems for publication.
4. Copying: this function reproduces files among storage systems.
5. Query: this function checks physical locations in storage systems where replicas are stored.

Data Grid applications produce huge amounts of data, and managing replicated data in a Data Grid is a major problem. The Replica Location Service (RLS) provides a

mechanism for registering and discovering replica data on the grid. The RLS architecture contains the following five elements:

1. Local Replica Catalogs (LRCs) that contain mappings from logical file names (LFNs) to Physical file names (PFNs);
2. Replica Location Indexes (RLIs) that aggregate state information about one or more LRCs and support the discovery of replicas at multiple sites using soft state protocols, such as LDAP, to maintain RLI states;
3. optional compression of soft state updates to reduce communication, CPU, and storage overheads; and
4. management of RLS membership for locating LRCs and RLIs.

3.4 Grid Portal

Grid portals allow communication between grids and the outside world, and are always huge, complex frameworks. The main portals are the Application Portal (AP) and the User Portal (UP). AP portals [19] allow specific grid operations for specific applications, such as the Astrophysics Simulation Collaboratory (ASC) portal [20] and the Diesel Combustion Collaboratory (DSC) portal [21]. User portals provide special services to specific members of the public and researchers, such as the HotPage Portal user portal [22], the Gateway project [23], and UNICORE [24].

From a user standpoint, the primary requirements of a Grid portal system include [25] the following:

1. **Security:** Users visit portals using web browsers and are authenticated by means of user IDs and passwords. While better authentication technologies exist, this one is demanded by users. More secure systems, such as smart cards, are possible, but unlikely to be deployed anytime soon.
2. **Remote access:** Tools for accessing file metadata directories and remote file archives are a central portal requirement. Simple Grid FTP tools are essential, but many files are likely to be managed by virtual data systems in which data is cataloged, curretted, and staged by back-end grid services.
3. **Remote execution:** The ability to submit jobs to the Grid for execution and monitoring is a standard portal requirement. Users allocating specific resources want to be able to see the job queues on those resources and consult scheduling assistants. They also need to be able to read logs to keep track of job execution and know when operations fail.
4. **User information services:** Access to directories and index tools is an essential portal function. All users should have private, persistent stores of references to important information they have stored on the Grid.
5. **Application interfaces:** The key to scientific portals is being able to hide Grid details behind useful application interfaces. Users need to be able to launch, configure, and control remote applications in the same way they use desktop applications.
6. **Access to collaboration:** All Grid organizations must permit resource sharing. This includes real-time and asynchronous collaboration.

Grid portal data must be described using XML. Although grid computing integrates computing resources inside grid organizations, these resources may be located on different kinds of platform. Thus, basic information transmission must rely on XML Web Services. In fact, the Globus Toolkit 3.0 already has a standard Web Services structure called the Open Grid Services Architecture (OGSA) [26, 27], which is referred to as “Grid Service.”

All Grid portals provide grid users with a web-based interface for using grid resources. Successful manipulation of the grid portal framework by means of Web Mining can bring about more efficient resource access.

Our Data Grid Portal utilizes the NSF Middleware Initiative’s (NMI’s) OGCE Portal [28] as the Grid Portal Site for the overall Grid organization. This portal technology provides access to Grid technologies through sharable and reusable components for web-based access to scientific- and business-oriented applications.

4. PLATFORM ARCHITECTURE

4.1 Hardware and Software Configuration

High-end servers and other supercomputers remain expensive, while PC prices continue to fall. Thus, the productivity gains the former provide are costly. But similar computing tasks can be performed by far less costly PC Cluster systems, and the difference in efficiency is not enormous. Therefore, substituting a PC Cluster system for a high-end server is cost-effective.

Table 1. Hardware configuration of grid-nodes.

Node 1*				
No	Domain Name	Host Name	Processor (CPU = 4)	Memory
1	beta1.hpc.csie.thu.edu.tw	beta1*	Intel Celeron 1.7GHz	512MB
2	beta2.hpc.csie.thu.edu.tw	beta2	Intel Celeron 1.7GHz	256MB
3	beta3.hpc.csie.thu.edu.tw	beta3	Intel Celeron 1.7GHz	256MB
4	beta4.hpc.csie.thu.edu.tw	beta4	Intel Celeron 1.7GHz	256MB
Node 2				
No	Domain Name	Host Name	Processor (CPU = 8)	Memory
1	gamma1.hpc.csie.thu.edu.tw	gamma1*	Pentium III × 2	512MB
2	gamma2.hpc.csie.thu.edu.tw	gamma2	Pentium III × 2	512MB
3	gamma3.hpc.csie.thu.edu.tw	gamma3	Pentium III × 2	512MB
4	gamma4.hpc.csie.thu.edu.tw	gamma4	Pentium III × 2	512MB
Node 3				
No	Domain Name	Host Name	Processor (CPU = 4)	Memory
1	lz01.lzsh.tcc.edu.tw	lz01	Pentium IV × 1	256MB
2	lz02.lzsh.tcc.edu.tw	lz02	Pentium IV × 1	256MB
3	lz03.lzsh.tcc.edu.tw	lz03	Pentium IV × 1	256MB
4	lz04.lzsh.tcc.edu.tw	lz04	Pentium IV × 1	256MB

* Master node of the cluster; the others are slaves.

Our Grid architecture, called a grid-cluster, is implemented on top of the Globus Toolkit. It consists of three Clusters that form a multiple-cluster environment. Clusters 1 and 2 each contain four PCs, one of which is a master node, while the rest are slaves. Cluster 3 also consists of four PCs, with a master node and three slaves, as shown in Table 1. The nodes are interconnected via a 10/100 Fast Ethernet Card connected to a Switch HUB. At present, the Education Grid includes three database nodes, with two nodes located at Tunghai University and another node at Lizen High School.

The operating system is RedHat Linux release 9, and for the Data Grid application, we use the Globus Toolkit 3.0.

4.2 The Front End – Grid Portal

Using Access Grid technologies, schools can integrate training courses and materials within the Grid environment, which provides greater teaching flexibility.

It is first necessary to set up a Portal Web Site in the Grid system. This provides services to Grid members and acts as a teaching platform for other institutions via the Internet. Organizations or individuals who have not joined the Grid will be restricted in their access to some resources.

Our Education Grid utilizes the NMI's OGCE Portal [33] as the network Grid Portal Site. The OGCE Portal includes the following functions:

1. posting on the Discussions board and communicating with friends in Chat rooms, and providing users with the latest grid-related technological updates in the News section;
2. monitoring Grid resources, such as node operating conditions;
3. submitting tasks for Grid operations;
4. transmitting data within the Grid via GridFTP; and
5. managing the Grid CA via the Proxy Manager.

The OGCE Portal framework provides a general portal architecture that supports virtual organizations composed of scientists and project developers, and also provides the API for the development of reusable, modular components that can be used to access the services being developed within the Grid organization.

4.3 The Back End – Data Grid Contents

The experimental location of this research is Lizen High School in Taichung, Taiwan. To establish this platform, we will design a system in accordance with the architecture shown in Fig. 3.

CAI software additions, along with growing numbers of users and increasing amounts of teaching materials will continuously increase the server load. Therefore, the storage device may need to be updated or replaced after a certain period of time. To overcome this problem, we will work on connecting idle storage devices within the institution using Data Grid techniques in place of expensive storage devices, such as a Disk Array or NAS, in serving as an e-learning platform storage device.

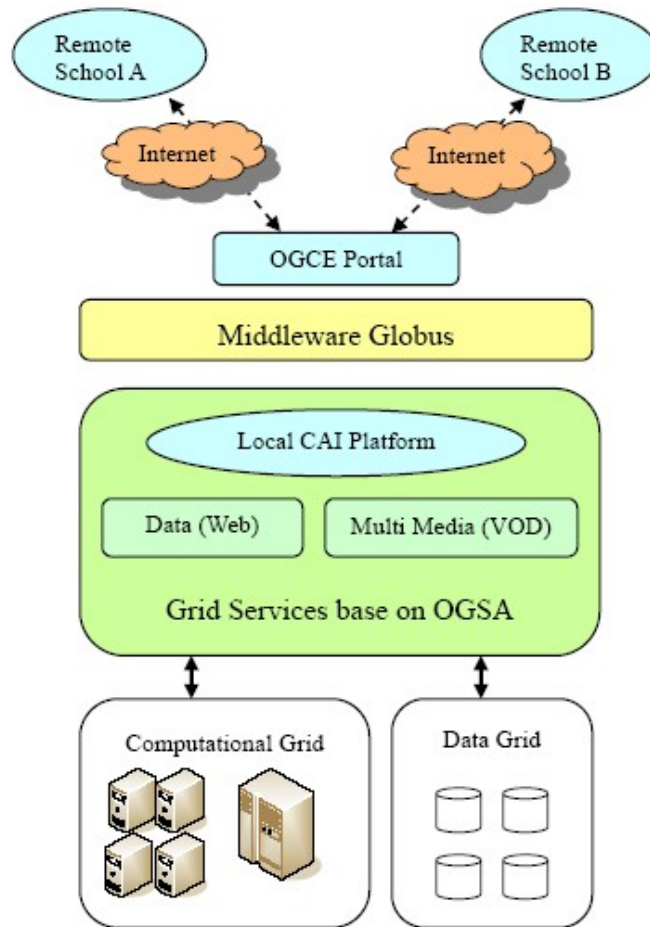


Fig. 3. E-learning system framework.

We will use a PC Cluster and LAN to simulate a data server and application server as the front end of the basic computational capability and as a data source for e-learning once materials have been acquired. Current plans include integrating the English CAI training software and database, and teaching resources. There will also be a VOD system providing multimedia streaming service connections via the Web that can be integrated with the CAI software.

4.4 Data Grid Implementation

Our Education Grid utilizes a Data Grid structure as the implementation technology. Within the Data Grid, the RLS [29] plays a vital role, providing a comparison feature between logical and physical file names. Data Grid users also have a convenient interface for saving and loading resources, which alleviates worries about where and how user resources are stored.

At the boundary of the Globus structure, the RLS Server is divided into two parts; one part consists of Local Replica Catalogs, which allow the LFN to communicate with the PFN in the physical storage system, and the other part is the RLI, which allows the LFN to communicate with the LRC. Although the LRC and the RLI have different functions, both can be installed on the same server. If they are installed on different servers, the LDAP is utilized as the communication protocol. To further streamline the system structure, we will install the LRC and the RLI together, and designate them as the most important parts within the Data Grid system.

Fig. 4 shows the complete portal structure we intend to create. The bottom-most level is the physical database. We will use the MySQL Server for the database system. At present, the Education Grid includes three database nodes, two of which are located at Tunghai University, and the other at Lizen High School.

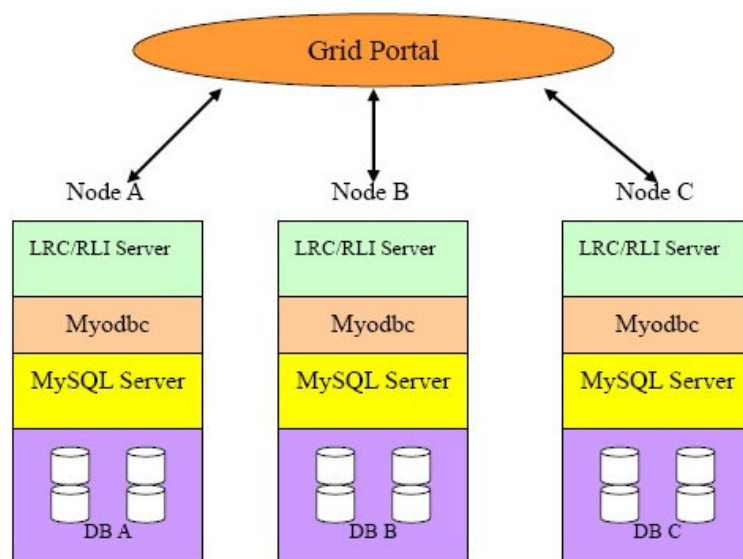


Fig. 4. E-learning portal framework.

Proceeding to the upper level, the database is connected to the Myodbc and ODBC, both of which provide interface connection to the LRC/RLI Servers. The GridFTP Protocol is used as the transmission device for copying and moving files on the Data Grid. The GridFTP is highly efficient and allows speedy, safe, dependable file transfers.

Control of the LRC/RLI Server requires another RLS client program, so we install the RLS client along with the Grid Portal, thus enabling network users to access resources on the grid through the Grid Portal.

4.5 SCO Repository Based on Data Grid

An e-learning Platform needs a Learning Management System (LMS) to store and manage its teaching content. The LMS plays two important roles: delivering the desired

courseware when and as needed, and tracking learners' responses and reactions. However, every LMS platform runs its own learning materials, which cannot be exchanged with those of other LMSs. This means e-learning resources are not available for sharing, which does not make the system very versatile. To deal with this problem, the U.S. government launched the Advanced Distributed Learning Initiative (ADL) program in late 1997 [30]. The result was an interactive technical standard known as SCORM (Sharable Course Object Reference Model). SCORM aims to establish a mechanism for repeated use and sharing of courseware as a way to reduce the time and cost of developing courseware and to make courseware reusable and acceptable to different LMSs.

The SCORM standard is divided into 2 parts: the Content Aggregation Model (CAM) and Run-time Environment (RTE) [31]. CAM-produced courseware is based on the principles of reusability, interoperability, and shareability. CAM includes three major modules: **content model**, **metadata**, and **content packaging**. Courseware elements are defined as content objects in the **content model** and must be properly arranged to make a reusable course, also known as an **SCO** (Sharable Content Object). **SCO** elements, such as html files, graphic files, and multimedia files, are known as "assets." **Metadata** files describe courseware information using XML. The description of courseware and elements made by **metadata** enables further management of course resources. **Content packaging** uses the Manifest XML files, denominated as `imsmanifest.xml`, to arrange and package SCOs in a course framework.

In conventional e-learning systems, each learning platform and courseware has a different method of communication since all systems have different programs and their API functions are defined differently. The SCORM RTE is designed to overcome this communication problem. We view the RTE as a simple Client-Server relationship in which the client end is the browser and the API Adapter provided by the LMS plays a key role in communication between LMS and courseware by taking responsibility for relay and reception of information between the Client and Server.

Some LMSs will be used as Data Grid nodes when implementing a complete Data Grid platform in which each node is provided with an interface for linkage between the Grid interface and the LMS. Upon completion, the Data Grid framework will enable sharing of SCOs between nodes.

Although an SCO meets the SCORM standards and can be run in every LMS, it may still be inconvenient for sharing among multiple LMSs because of the lack of a fast, safe, and secure mechanism. In Fig. 5, we show an SCO-sharing mechanism based on a Data Grid platform. Each SCO Repository in the LMS is linked through Globus middleware, and each and every LMS node can share SCOs with other LMSs within the Data Grid structure. As the figure shows, Globus is in the middle of the communication channel between local LMSs and the Grid Organization, and communication between nodes is conducted via the Grid Portal, which is the interface between grid nodes.

Figs. 6 and 7 display screenshots of the Educational Grid's Portal Site. We plan to place multimedia educational materials with different contents on each node. For instance, Node A could be used for math materials, Node B for English materials, and Node C for biology materials. File backup and querying can be achieved through the aforementioned RLS Server. We will thus be able to truly realize the establishment of an educational platform.

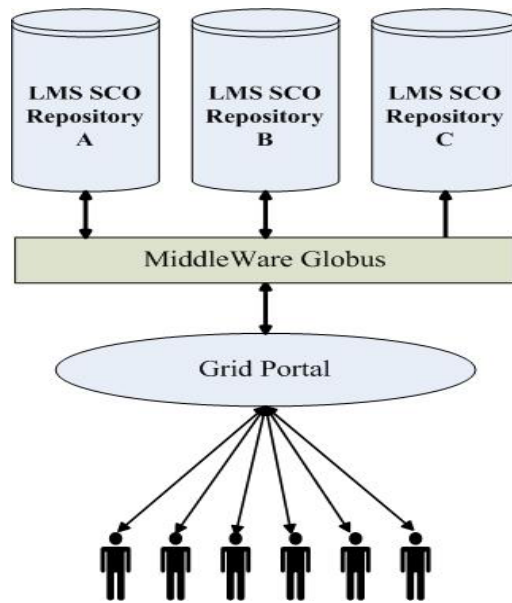


Fig. 5. SCO data grid platform.



Fig. 6. E-learning portal screenshot-1.

Name	Size	Time
../		
cog-1.1		23:28 Jun 10
cog-1.1-bin.tar.gz	2739777	2003 Jul 15
examples		04:04 May 26
globus_simple_ca_ffa40f5d_setup-0.13.tar.gz	149893	17:04 Jun 10
gram_job_mgr_11168.log	395334	09:22 Jun 11
gram_job_mgr_11441.log	384631	09:22 Jun 11
gram_job_mgr_12500.log	268050	19:18 Apr 30

Fig. 7. E-learning portal screenshot-2.

5. PARTNER SCHOOL ARCHITECTURE

The system established for this research can be applied not only in our school but also in other schools via the Internet to share teaching materials. We propose a Grid virtual organization based on the concept of “Partner Schools” that can be either “Resource receivers” sharing existing resources on the grid or “Resource providers” contributing resources. Partner schools can contribute various teaching resources according to their professional specialties. For instance, School A may contribute mathematics resources, School B literature resources, and so on. In this way, each partner school can concentrate on developing resources for its own professional specialty while also contributing professional teaching resources to other schools. Expenditures for teaching cost investments can thus be avoided. Fig. 8 shows a partner academy framework.

6. CONCLUSIONS AND FUTURE WORKS

This research is aimed at applying Grid Computing technologies to integrate idle computer resources in schools to reduce costs and make efficient use of applications and resources. If it is widely applied, huge savings in software and hardware equipment purchasing expenditures can be realized, and the goals of complementation and sharing of resources among schools achieved. Moreover, schools with limited budgets can also obtain better services and vast teaching resources using Grid technology. With further

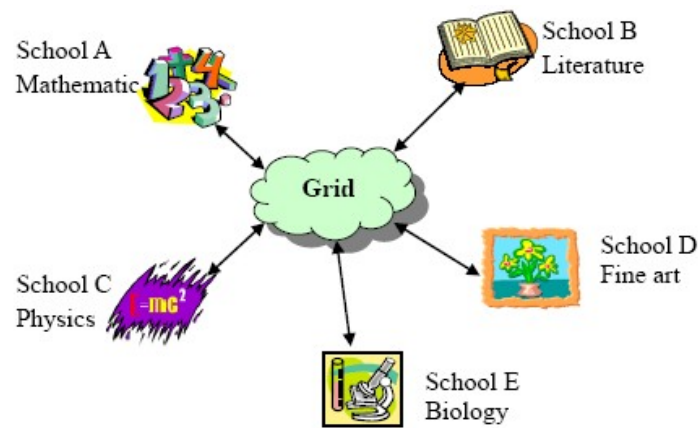


Fig. 8. Partner academy framework.

linking of partner schools, we can also expect the resulting experience to provide a reference for schools to assist them in developing similar e-learning environments and reducing development times and costs.

Besides e-learning platforms, there are many Data Grid technology applications. The following topics can be further explored in the future.

6.1 Global Digital Library Grid

Data Grids could be used to link Digital Collection institutions, such as libraries, museums, and archives, to form an enormous virtual museum [37], which we call a Digital Library Grid. Through an internal file management mechanism, a Digital Library Grid could allow users to share digital resources, conduct catalog management, and replica transfer management, and enjoy efficient access to files. Once scattered resources are linked together, all that end users would have to do is go to a single access point to search for files scattered all over the world. With replica selection, the most efficient site would be located automatically, and a copy of the file would then appear at the local site.

With the Data Grid architecture, information on each Local Grid should be unique and not duplicated. For example, Library A could house fine arts and Library B literature. This would result in reducing resource investments. However, since the libraries would be connected over the Internet, it is possible that in the course of data transmissions, abnormalities could occur due to unstable connection quality. In order to reduce this possibility and increase data retrieval efficiency, we should separate Data Grid storage spaces as Proxies [32].

6.2 Access Grid and Data Grid

The Access Grid is a technology developed by the U.S. Argonne National Laboratory. It enables many people to process video and audio interactions, and exchanges opinions. At present, it is used on many sites that need video for training, teaching, conferences and seminars [33]. The Access Grid application software is called the Access

Grid Toolkit and is freeware anyone can download from the Access Grid Web site. The Web site also provides complete technical archives to give necessary assistance to users. As of September 2003, the latest version of Access Grid Toolkit was Version 2.1 [34].

In Taiwan, the most well-known example of Access Grid application was its use in the fight against SARS during 2003 [35]. With assistance from three U.S. Argonne National Laboratory researchers, some Taiwan medical institutes successfully installed an Access Grid, thus enabling Taiwanese doctors to discuss SARS patients' conditions with specialists in other medical institutes around the world and to obtain more objective and precise judgments to cope with the spread of the SARS virus.

Before the creation of the Access Grid Toolkit Version 2.0, the Access Grid was only a kind of video conference system. With Version 2.0, the Access Grid has been tightly combined with Grid technology. All the functions of Grid technology, such as authentication, resource allocation, and remote data access and fault detection, are standard Access Grid functions. Therefore, the Grid Middleware Globus Toolkit must be installed before installing the Access Grid Toolkit 2.0.

Audio and video materials recorded by an Access Grid are usually so huge that they require high-capacity storage devices, such as a Disk Array, NAS, or SAN. These devices are very expensive, however, Data Grid technology can be substituted for them.

REFERENCES

1. S. Tucker, A. Pigou, and T. D. Zaugg, "e-learning – making it happen now," in *Proceedings of 30th Annual ACM SIGUCCS Conference on User Services*, 2002, pp. 292-293.
2. Distance Education and the WWW, <http://tecfa.unige.ch/edu-comp/edu-ws94/contrib/peraya.fm.html>.
3. F. Berman, G. Fox, and T. Hey, *Grid Computing-Making the Global Infrastructure a Reality*, John Wiley and Sons Ltd., New York, 2002, pp. 9-49.
4. A. Reinefeld and F. Schintke, "Concepts and technologies for a worldwide grid infrastructure," in *Proceedings of 8th International Euro-Par Conference on Parallel Processing*, LNCS 2400, 2002, pp. 62-71.
5. "What is grid computing?" http://www-1.ibm.com/grid/about_grid/what_is.shtml.
6. C. O'Hagan, "The integration of television and the Internet," in *Proceedings of IEEE International Conference on Advanced Learning Technologies (ICALT '01)*, 2001, pp. 475-477.
7. A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, and S. Tuecke, "The data grid: towards an architecture for the distributed management and analysis of large scientific datasets," *Journal of Network and Computer Applications*, Vol. 23, 2001, pp. 187-200.
8. G. Fox, "Grid computing environments," *Computing in Science and Engineering*, Vol. 5, 2003, pp. 68-72.
9. Introduction to Grids and the Globus Toolkit, <http://www.globus.org/training/grids-and-globus-toolkit/>.
10. W. Johnston, "Computational and data Grids in large-scale science and engineering," *Future Generation Computer Systems*, Vol. 18, 2002, pp. 1085-1100.

11. GriPhyN – Grid Physics Network, <http://www.griphyn.org/index.php>.
12. Particle Physics Data Grid (PPDG) Project at SLAC, <http://www.slac.stanford.edu/xorg/ngi/ppdg/ppdg-slac.html>.
13. The Data Grid Project, <http://eu-datagrid.web.cern.ch/eu-datagrid/>.
14. S. Vazhkudai, S. Tuecke, and I. Foster, “Replica selection in the globus data grid,” in *Proceedings of 1st IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2001, pp. 106-113.
15. Getting Started with the Globus Replica Catalog, <http://www.globus.org/datagrid/deliverables/replicaGettingStarted.pdf>.
16. S. Malaika, A. Eisenberg, and J. Melton, “Standards for databases on the grid,” *ACM SIGMOD Record*, Vol. 32, 2003, pp. 92-100.
17. Globus: Grid Security Infrastructure (GSI), <http://www.globus.org/security/>.
18. Globus: Replica Management, <http://www.globus.org/datagrid/replica-management.htm>.
19. M. Thomas, M. Dahan, K. Mueller, S. Mock, C. Mills, and R. Regno, “Application portals: practice and experience,” *Journal of Concurrency and Computation: Practice and Experience*, Vol. 14, 2002, pp. 1427-1443.
20. ASC Portal Users Guide, <https://www.ascportal.org/portal/usersguide.html>.
21. Diesel Combustion Collaboratory, <http://doecollaboratory.pnl.gov/research/diesel/homepage.html>.
22. NPACI HotPage Portal, https://hotpage.npaci.edu/status/hotpage_main_in.html.
23. M. E. Pierce, C. H. Youn, and G. Fox, “The gateway computational Web portal,” *Journal of Concurrency and Computation: Practice and Experience*, Vol. 14, 2002, pp. 1411-1426.
24. Uniform Interface to Computing Resources, <http://www.fz-juelich.de/unicore/>.
25. J. Novotny, “The grid portal development kit,” *Journal of Concurrency and Computation: Practice and Experience*, Vol. 14, 2002, pp. 1129-1144.
26. Towards Open Grid Services Architecture, <http://www-fp.globus.org/ogsa>.
27. G. Xue, M. J. Fairman, and S. J. Cox, “Exploiting web technologies for grid architecture,” in *Proceedings of 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID 2002)*, 2002, pp. 256-257.
28. NMI Portal Installation Instructions, <http://www.ogce.org/install.php>.
29. S. Vazhkudai, S. Tuecke, and I. Foster, “Replica selection in the Globus Data Grid,” in *Proceedings of 1st IEEE/ACM Cluster Computing and the Grid*, 2001, pp. 106-113.
30. Advanced Distributed Learning – Home, <http://www.adlnet.org/>.
31. H. T. Li, C. H. Lin, Y. C. Chang, and J. T. D. Yang, “On the distributed management of SCORM-compliant course contents,” in *Proceedings of IEEE International Conference on e-Technology, e-Commerce and e-Service*, 2004, pp. 538-539.
32. W. Hoschek, J. Jaen-Martinez, A. Samar, H. Stockinger, and K. Stockinger, “Data management in an international data grid project,” *Lecture Notes In Computer Science*, Vol. 1971, 2000, pp. 77-90.
33. R. Stevens, M. E. Papka, and T. Disz, “Prototyping the workspaces of the future,” *IEEE Internet Computing*, Vol. 7, 2003, pp. 51-58.
34. Access Grid, <http://www.accessgrid.org/>.
35. Sars Hope, <http://www.sarshope.org/>.



Chao-Tung Yang (楊朝棟) was born on November 9, 1968 in I-Lan, Taiwan, R.O.C. He received a B.S. degree in Computer Science and Information Engineering from Tunghai University, Taichung, Taiwan in 1990, and an M.S. degree in Computer and Information Science from National Chiao Tung University, Hsinchu, Taiwan in 1992. He received a Ph.D. degree in Computer and Information Science from National Chiao Tung University in July 1996. He earned first-class honors in the National Higher Examination for the Information Processing field in 1994, and won the 1996 Acer Dragon Award for outstanding Ph.D. Dissertation. He has worked as an associate researcher for ground operations in the ROCSAT Ground System Section (RGS) of the National Space Program Office (NSPO) in Hsinchu's Science-based Industrial Park since 1996. In Aug. 2001, he joined the faculty of the Department of Computer Science and Information Engineering at Tunghai University, where he is currently an Associate Professor. His present research interests are in parallel and cluster computing, grid computing, and internet-based applications.



Hsin-Chuan Ho (何信權) was born on September 16, 1963 in Taichung, Taiwan, R.O.C. He received a B.S. degree from the Department of International Trade at Tunghai University, Taichung, Taiwan, in June 1986, and an M.S. degree in 2004 from the Department of Computer Science and Information Engineering at Tunghai University. Since 2001, he has worked as a computer teacher at Lizen High School in Taichung, Taiwan. His research interests include grid computing, e-learning, and data mining.