UNIVERSITI PUTRA MALAYSIA

A PLATFORM INDEPENDENT WEB-BASED DATA MANAGEMENT SYSTEM FOR RANDOM ACCESS FILE

KU DAY CHYI

FK 2002 66

A PLATFORM INDEPENDENT WEB-BASED DATA MANAGEMENT SYSTEM FOR RANDOM ACCESS FILE

By

KU DAY CHYI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirement for the Degree of Master of Science

July 2002

To God, my parents, brother and sisters

Abstract of the thesis submitted to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Master of Science

A PLATFORM INDEPENDENT WEB-BASED DATA MANAGEMENT SYSTEM FOR RANDOM ACCESS FILE

By

KU DAY CHYI

July 2002

Chairman : Associate Professor Borhanuddin Mohd. Ali, Ph.D.

Faculty : Engineering

With the advent of the Web, the Internet has evolved into a user-friendly medium capable of high speed, on demand information delivery. Putting data onto the World Wide Web in the form of a fully accessible, searchable database can open up a wide variety of possibilities for teaching, learning and research. There are many different types of web-based database management system (DBMS), e.g., Oracle, Informix/Illustra, IBM DB2, Object Design, ranging from small systems that run on personal computers to huge systems that run on mainframes. However, these systems have limitations such as being platform dependent, not portable and expensive.

This thesis describes the development of WebDB, a platform independent web-based data management system using Java servlets and random access files to address the problems. It is developed in order to provide the management functions to WebEd2000's database. WebEd2000 is a working prototype of Web-based distance learning system developed at the Broadband and ATM Research Group, Universiti Putra Malaysia (UPM). It enables delivering conventional lecture notes over the Web and providing various tools to help in managing and maintaining course materials in a server. The WebDB approach is for the ease of the centralized-management of database administrator over the WebEd2000 users and maintains the database. It also enables instructors to access to their database and update it when necessary. Instead of WebEd2000 database, the system allows its users to put another database on the server.

WebDB is mainly developed using the combination of Java servlets and JavaScript technologies. The server-side servlets are used to handle the requests from client and the responses from server. The random access file served as database repository in the database server where all the data is stored. The client-side JavaScript is used to enable DHTML features and perform less-security-concern processes in order to reduce the workload of the web-server.

Lastly, WebEd can be easily set up and deployed in any platform and web-servers with minimal modifications. Portability is achieved by utilizing Java technology for the system applications and random access file as the data repository.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia untuk memenuhi keperluan ijazah Master Sains

SISTEM PENGURUSAN MAKLUMAT YANG BEBAS-PELANTARAN BERASASKAN INTERNET UNTUK FAIL CAPAIAN RAWAK

Oleh

KU DAY CHYI

Julai 2002

Pengerusi : Profesor Madya Borhanuddin Mohd. Ali, Ph.D.

Fakulti : Kejuruteraan

Dengan kemunculan Web, Internet telah berubah menjadi sebuah perantaraan berupaya kelajuan tinggi, atas permintaan penghantaran maklumat yang ramah pengguna. Meletakkan data ke atas jaringan seantaro dunia yang berada di dalam bentuk mudah dicapai, pangkalan data yang membolehkan proses pencarian dapat membuka aneka jenis kemungkinan untuk pengajaran, pembelajaran dan penyelidikan. Terdapat pelbagai jenis sistem pengurusan pangkalan data (DBMS) berasaskan Internet seperti Oracle, Informix/Illustra, IBM DB2, Object Design, bermula daripada sistem kecil yang beroperasi di komputer peribadi ke sistem amat besar yang beroperasi di kerangka utama. Sistem-sistem tersebut walau bagaimanapun, mempunyai had seperti tidak bebas-pelantaran, tidak mudah alih dan berkos tinggi.

Tesis ini menerangkan pembangunan WebDB, sebuah sistem pengurusan maklumat berasas Internet yang tali bersandar pelantar yang ditulis dengan menggunakan Java servlets dan fail capaian rawak untuk mengatasi masalah ini. Ia dibina untuk membekalkan fungsi pengurusan ke pangkalan data WebEd 2000. WebEd 2000

sebenarnya adalah sambungan dari WebEd, iaitu satu prototaip sistem penbelajaran jarak jauh yang dibangunkan oleh Kumpulan Penyelidikan *Broadband* dan *ATM*, Universiti Putra Malaysia (UPM). Ia membolehkan penghantaran nota kuliah melalui Web dan membekalkan pelbagai alatan untuk membantu dalam mengurus dan memelihara kandungan kursus dalam sebuah pelayan. Kemunculan WebDB menyenangkan pengumpulan maklumat secara berpusat untuk membolehkan pengurus pangkalan data WebEd 2000 memelihara data and pengguna pangkalan dengan lebih mudah. Ia juga membolehkan pengajar mencapai ke pangkalan data mereka dan mengemaskinikannya bila diperlukan. Selain daripada pangkalan data WebEd 2000 yang sedia ada, sistem ini membenarkan pengguna-penggunanya meletakkan pangkalan data yang lain di atas pelayan.

WebDB dibina menggunakan gabungan teknologi Java Servlet dan JavaScript. Servlets yang beroperasi di sebelah pelayan digunakan untuk mengendalikan permintaan daripada pelanggan dan tindak balas daripada pelayan. Fail capaian rawak itu bertugas sebagai penyimpan fail-fail bagi pangkalan data di dalam pelayan. Sementara itu, JavaScript digunakan untuk menghasilkan kesan DHTML dan melaksanakan proses seperti mempamerkan antaramuka untuk input, dengan itu mengurangkan beban pelayan secara keseluruhannya.

Akhir kata, WebDB boleh beroperasi di atas semua pelantaran dengan sedikit modifikasi. Ini adalah disebabkan oleh penggunaan teknologi Java untuk aplikasi sistem dan fail capaian rawak untuk sistem penyimpanan data.

ACKNOWLEDGEMENTS

First of all, I would like to express my utmost thanks and gratitude to God for giving me the ability to finish my research project and this thesis successfully.

The author gratefully wish to express her profound appreciation and gratitude to her supervisor, Assoc. Prof. Dr. Borhanuddin Mohd. Ali, for his supervision, guidance, supporting, and constructive suggestion and comments throughout the duration of the project until it turns to real success.

The author also indebted to the members of her supervisory committee, Dr. V. Prakash and Dr. Abdul Rahman Ramli, for their affectionate guidance, prompt decision and valuable assistance during this period.

Appreciation also to the assistance rendered by the respective lecturers, staffs, technicians of the faculty of engineering for providing the facilities required for undertaking this project.

The author would like to thank her family for the encouragement and support without which is impossible for the success of this project, and my friends, especially Qussay A. Salih and Micheal Ng Chee Khun for offering their helps.

This thesis submitted to the Senate of Universiti Putra Malaysia has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisor Committee are as follows:

Borhanuddin Mohd. Ali, Ph.D.
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Veeraraghavan Prakash, Ph.D.
Faculty of Engineering
Universiti Putra Malaysia
(Member)

Abdul Rahman Ramli, Ph.D.
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Member)

AINI IDERIS, Ph.D.
Professor/Dean
School of Graduate studies
Universiti Putra Malaysia

Date:

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

LIST OF ABBREVIATIONS

API	-	Application Programmer Interface
ASCII	-	American Standard Code for Information Interchange
CGI	-	Common Gateway Interface
CORBA	-	Common Object Request Broker architecture
DBI	-	Database Interface Module for Pearl
DBMS	-	Database Management System
DFD	-	Data Flow Diagram
DHTML	-	Dynamic Hyper Text Markup Language
EJB	-	Enterprise JavaBeans
GUI	-	Graphic User Interface
HTML	-	Hyper Text Markup Language
HTTP	-	Hyper Text Transfer Protocol
IIS	-	Microsoft Internet Information Server
IP	-	Internet Protocol
ISAPI	-	Internet Server API
JDBC	-	Java Database Connectivity
JVM	-	Java Virtual Machine
MacOS	-	Macintosh Operating System
NOS	-	Network Operating System
NSAPI	-	Netscape Server API
ODBC	-	Object Database Connectivity
ODBMS	-	Object-oriented Database Management System

ODMG	-	Object Data Management Group
OMG	-	Object Management Group
OO	-	Object-oriented
OODB	-	Object-oriented database
PHP	-	Hyper Text Preprocessor
RAD	-	Rapid Application Development
RDBMS	-	Relational Database Management System
SQL	-	Sequence Query Language
UID	-	User Interaction Diagram
UML	-	Unified Modeling Language
URL	-	Uniform Resource Locator
WRBAPI	-	Web Request Broker API
WWW	-	World Wide Web

CHAPTER I

INTRODUCTION

Web-based Data Management System

The meteoric rise of the World Wide Web (WWW) is one of the defining technology-related events of the mid-1990s. The Web enables almost anyone, especially businesses, to broadcast timely information world wide at minimal cost. At the very least, the Web has augmented the marketing efforts of countless companies (Maurice, 1996*)*.

Databases are used in Web development in two ways; as interfaces to information repositories, which are manipulated using the browser as the client component of a traditional client-server system, and as the underlying system layer from which dynamically generated web-site content is composed (McGhan, 1998).

Today, Informix, Oracle and Sybase provide native connectivity to web-browser and ODBC gateway allows connections to other databases, including Microsoft databases. IBM is currently working to add native support for DB2 as well. Native database support automatically takes care of the vendor-specific query language, allowing developers to quickly develop applications that access databases in a generalized way (McGhan, 1998).

Web-based applications generally work across all computing platforms in an organization. either through a common web-browser client, or through Java applets (Matthew, 1998). Both of these approaches can be deployed locally at a site. or they can be made accessible over the Internet for the broader community to access. Finally, using the Web for data management affords the obvious advantages of easing the process of sharing and distributing data with collaborators, the general public, and archive centers.

The Challenges of Implementing a Web-based Data Management System

Because of the open nature of the Internet, security of Web-based DBMS is a serious issue. Extra steps must be taken to ensure that private information remains private. The level of security required for the application depends on the volatility of the information being passed between the Web servers to the browser.

Since the Web-based DBMS tends to centralize the processes and applications on servers, the risk that the entire DBMS will be completely down is significantly increased compared to the distributed DBMS.

The current commercial data management systems available in market are powerful in terms of applications and functions. However, there are several drawbacks faces by the users when implementing the systems. Some of the common problems faced are the lack support for extensibility of new formatted data types, difficult integration with existing data stores, slow in performance, expensive and not scalable (Dicken, 2002). Furthermore, the systems, though support for various platforms, are typically ported to specific platforms and shifted separately. This makes the system deployment somewhat inflexible (Yem and Ku, 2001).

Sybase provides separate, unintegrated servers for geospatial data and relational data. Other DBMS systems (e.g., Oracle, Informix/Illustra, IBM DB2, Object Design) provide "data blades" or Database Extenders, that is, well-supported libraries for some useful formatted data types like text, audio, images, video, spatial, time-series, HTML, Java, and other formatted data types. Some of these libraries are application-developer extensible but they are tied to particular DBMS products and not portable though it appears that they could be made portable (Craig and Paul, 1996). Other than that, the existing DBMS only allows user to define some simple compound data types and collections.

RDBMS mention above, though much more reliable than most user-written transaction processing code, are not nearly as reliable as a basic Unix system with a Web server pulling static files out a file system (Greenspun, 1998). Therefore, random access binary static file can be used as the database repository instead of the RDBMS to achieve the system reliability. Random access binary file provides an alternative to the database management system, which improves the reliability and independency of the system. It provides the portability, definition of new formatted data type by user, better performance, less complexity and easier integration with existing system and made web-accessible by web-server connectivity technologies.

The Challenges of Using Random Access Binary File as Data Repository

Besides database, one way to store permanent data is by using random access file. Random access binary file is a file that can be accessed randomly and stored in binary format. The binary file is computer-readable but not human-readable (Steve. 1999). Random access binary file is a simple and flexible file that stores data in sequence of bytes. All executable programs are stored in binary files, as are most numeric data files. A binary file format is defined in terms of named data segments and composites of data segments. Each named segment of the format is documented and described in terms of its size, which can be fixed or variable, and position relative to the beginning of the file or some other named data segment.

The compactness of the binary data product make it practical to represent the full detail available for data, which might be impractical in the limited ASCII format. On the contrary, the format will usually vary between data sets and require custom programming to extract the information. Naturally, sufficient documentation for the binary data product format is provided in the metadata (EMF Measurements Database, 1997).

An ASCII data file can take two to five times the disk space of a comparable binary data file. Not only is there less information in each byte, but extra bytes are needed for decimal points, delimiters, and end-of-line markers (Sgouros, 1999). In a text or ASCII file, each character requires 1 byte (8 bits) of storage (Singleton, 2001). By storing data files in binary format, disks space can be saved and applications can be accessed more efficiently.

The obvious inconvenience of binary files is that user cannot see the content in the file and so it is not always obvious how to read them (Singleton, 2001). The user has no way to read or write to the contents in the files hence will not be able to maintain the data efficiently. Redundant data will be kept in the file and occupied the disc memory of the server. It is however a wastage of resources and is not practical in long-term. Some form of file content or file description can overcome the problem. There are three possible ways of doing this (Singleton, 2001):

1. Write some character strings to the file giving the file description before writing the binary data. Those strings will be readable if you list the file. This method is messy.
2. Write an brief auxilliary file describing the format of the binary file and always keep the files together as a pair.
3. Use one of the available self-describing file format packages (netCDF is one such package). All file IO goes through a library and the package provides routines to inquire about the file content and format.

In this thesis, we developed a Web-based data management system called WebDB based upon random access binary file using Java and servlet technologies. It overcomes the limitations of random access binary file stated above with the second method suggested by Singleton.