

电子资源备份、交换空间的建设

全成干 王 爽

(厦门大学图书馆 厦门 361005)

【摘要】 阐述厦门大学图书馆数字化建设过程中将 FTP 与 HTTP 服务相结合,提供面向大用户群的文档型数字资源备份、交换服务的可行性及其实现。

【关键词】 数据备份 交换 系统安全

【分类号】 G250.7

Digital Resource Backup and Exchange Service

Quan Chenggan Wang Shuang

(Library of Xiamen University, Xiamen 361005, China)

【Abstract】 Xiamen University Library provides a new type of service—Data Backup and Exchange Online (trial) based on File Transfer Protocol and Hypertext Transfer Protocol. The design and realization of this service is discussed in this paper.

【Keywords】 Data backup Exchange System security

厦门大学图书馆经过多年数字化建设,取得一定成果,使用户能较为方便地获取大量数字资源。在使用过程中,用户常常需要收集与其学科、专业或兴趣相关的资料作为参考和备用,这些资料与用户的论文、随笔及教案等通常都存放在个人电脑中,存在由于计算机硬件故障或病毒感染等意外事件而丢失资源的风险;此外,存放于个人电脑中的资料无法“移动”使用,虽然当前 CD、DVD 刻录机已经普及,但是使用光盘备份仍然存在相当多的不便,数据安全、资源交换与共享仍受到一定的限制。

厦门大学图书馆在数字化建设过程中除了进行资源建设,还进行相关的服务建设。针对用户遇到的问题,作者设想充分利用图书馆现有的 SAN 存储系统和磁带库备份系统,提出读者数据备份、交换空间建设的解决方案,进行服务建设的新尝试。

文件传输协议(File Transfer Protocol,FTP 协议)^[1]是当前 Internet 上最成熟、应用最广泛的用来传送文件的协议,系统的最初设计是基于 FTP Server 之上,但使用 FTP 客户程序来管理数据远不如 Web 页面方便,考虑到 HTTP 协议也具有文件上传/下载功能^[2],所以本系统设计中将 FTP 与 HTTP 服务有机结合,共同提供服务如图 1 所示。

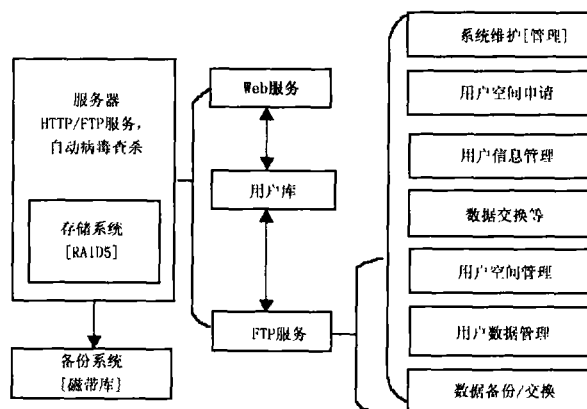


图 1 备份、交换系统结构示意图

1 技术设计和实现

(1) 硬件系统

包括服务器一台(提供 Web 服务和 FTP 服务),配置为双 Xeon PII400/2GB 内存;数据存放在 SAN 中(Dell/EMC FC4500);用户名、密码等用户信息数据由 SQL Server 2000 管理,存放于本馆专用数据库服务器中(可用 Access 数据库替代);磁带库备份系统。

(2) 软件系统

操作系统为 Windows Server 2003 Enterprise Edition; FTP 服务由 Serv-U 提供,Web 服务器采用系统自带的 IIS6。

备份、交换系统由厦门大学图书馆信息技术部自主开发,包括系统管理、用户管理、空间管理、数据管理和使用等四大

收稿日期: 2004-06-14

功能模块。

(3) 系统功能

系统提供 HTTP 和 FTP 两种上传途径。用户可将要备份的数据上传至系统内,系统每日进行病毒扫描和数据备份以保证数据安全。用户以 HTTP 和 FTP 方式下载数据,满足“移动”使用的需要。同时,用户只要设置受限密码并在用户空间下建立 Public 目录即可实现数据交换功能,例如教师可将用户名、受限密码告诉学生,学生即可通过 HTTP 方式获取教师存放在 Public 目录下的资料,也可将完成的作业等数据上传到 Public 目录下。

系统管理模块提供系统管理员进行用户管理和空间管理的操作接口,实现数据压缩、空间分配及用户信息维护等功能;用户管理模块实现新用户注册和已注册用户进行密码修改等操作;用户通过空间管理模块可以管理其使用空间、决定是否开放数据交换功能及进行备份文件的压缩操作以节约空间等;数据管理和使用则提供 HTTP 和 FTP 两种方式进行数据上传、下载、删除及目录操作等功能。特别要说明的是,使用受限密码只能通过 HTTP 方式访问 Public 目录。

2 关键问题的解决

(1) 用户管理问题

已有公开发表的论文详细探讨了 FTP 服务器的建设^[3,4],但所涉及的用户数均较少。对本系统而言,使用者至少有上千人,所以实现对数千个 FTP 用户进行管理是开发本系统要解决的第一个关键问题。作者认为使用数据库管理大量用户信息是当前最为合适的方式。

目前 Windows 平台下常用的 FTP 服务器主要包括 IIS 的 FTP 组件, Serv-U FTP Server、Ws-FTP Server、Xlight FTP Server 等,其中 Serv-U FTP Server 和 Xlight Ftp Server 均支持用户数据库管理^[1],但是在用户数的大情况下,自带的用户管理模块基本无法使用。作者通过开发 ASP 程序实现 Web 页面管理 Serv-U-FTP Server 用户。

(2) 用户密码问题

本系统采用 Serv-U Ftp Server 实现 FTP 功能。与 XLight FTP Server 不同, Serv-U Ftp Server 采用非标准的 md5 加密,对 Serv-U FTP Server 加密算法的研究和相关加密模块开发是本系统的另一个关键点。

(3) 安全性问题

由于系统是在 Windows 平台下开发,使用 IIS6 为主页服务器,如果将页面发布目录设置为只有“读取”权限,则无法保证用户通过 HTTP 协议下载其备份的动态页面程序(如 *.asp 等),若不进行限制,又存在系统被攻击的潜在威胁。解决安全问题是保障用户数据安全所要注重的第三个关键点。

(4) HTTP 上传文件大小限制问题

IIS6 对于 ASP 程序上传数据字节数的默认设置为不超过 200K,通过对 IIS6 默认设置进行调整方可实现大文件上传的需要。考虑到通过 HTTP 上传文件效率较低,本系统将 HTTP 上传大小限制设定为 5M。

3 功能扩展和探讨

作为数字化建设过程中的一种新服务,厦门大学图书馆读者电子资源备份、交换空间服务目前只面向教师开放。在试用期间,系统运行稳定、效果良好。在此基础上,作者还将逐步增加下列功能:

(1) 数据标引功能

电子资源备份、交换空间在提供数字资源备份的同时,为用户提供简单数据标引功能,从而实现对其收集的数据进行分类,提高数据的利用率。作者初步设计采用 DC 核心元数据规范开发数据标引模块。

(2) 全文检索功能

本系统主要是提供文献型数字资源的备份。通过引入全文搜索引擎,对用户的数据进行索引并提供全文检索功能,以方便用户对资源查找和使用。作者正在开发基于 Lucene 的全文检索功能。

参考文献:

- 1 File Transfer Protocol (FTP); Network Working Group, October 1985. <http://www.w3.org/Protocols/rfc959/Overview.html> (Accessed May 25, 2004)
- 2 Form-based File Upload in HTML, E. Nebel and L. Masinter, November 1995. <http://www.faqs.org/rfcs/rfc1867.html> (Accessed May 25, 2004)
- 3 张海波. 高校图书馆的 FTP 站点建设. 现代图书情报技术, 2004 (2): 75-77
- 4 马林山, 丁健. 图书馆 FTP 服务器的配置与维护. 合肥联合大学学报, 2002(4): 108-112
- 5 Jakarta Lucene: The Apache Jakarta Project, May 2004. <http://jakarta.apache.org/lucene/docs/> (Accessed May 25, 2004)

(作者 E-mail: ganr@xmu.edu.cn)

《现代图书情报技术》2004年(年刊)征订通知

《现代图书情报技术》2004年(年刊)已经出版。每册定价29元(含邮费)。内容涉及数字图书馆技术、图书馆自动化、信息检索技术、网络资源与建设、网络多媒体技术、图书情报技术工作交流等方面。如欲定购,请直接将款汇到:北京中关村北四环西路33号《现代图书情报技术》编辑部 邮编:100080,本刊编辑部收,款到后即将发票随刊一并寄出。印数有限,欲购从速!编辑部联系电话:(010)82624938。