



项目编号:

2015032

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称: 大数据环境下云存储在高校图书馆数字资源
长期保存中的应用研究

项目关键词: 大数据 云存储 数字资源

项目单位(盖章): 塔里木大学图书馆

通信地址: 新疆阿拉尔市塔里木大学图书馆 843300

项目主持人: 席亚军

联系电话: 13565142267

电子邮件: smallx1002@sina.com

提交日期: 2016年4月20日

项目结题验收单

1 专家验收表（项目单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究		
主持人	席亚军	职务/职称	馆员
所在单位	（加盖公章）塔里木大学图书馆		
专家意见	<p>根据 CALIS 农学中心项目验收的要求，专家组对《大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究》进行了验收。专家组认真听取了课题组汇报，审阅了相关材料，经讨论后形成如下验收意见：</p> <p>1、该课题于 2015 年 5 月立项，课题组成员能够按照 CALIS 农学中心的要求，前期基础性工作扎实，研究大纲条理清晰，有一定的创新。</p> <p>2、课题组成员克服了经费和设备不足的困难，合理分工，充分利用图书馆工作的优势，广泛查阅资料，对图书馆数字资源的概念、特点和价值有了深入的理解和认知。在此基础上，通过对数字资源各种存储方式的对比分析，得出了云存储策略是未来高校图书馆数字资源存储的发展趋势，对构建塔里木大学数字图书馆云平台，完成云存储策略知识普及和人员培训，注重技术的更新和设备的维护，保证平台的正常运行奠定了坚实的基础。</p> <p>课题研究期间共举办相关知识培训 3 次；举办技能考核 1 次，发放云存储知识普及试卷 89 份，收回有效试卷 80 份；课题组成员对各自的研究形成了阶段性成果，共撰写并发表论文 7 篇；获得软件著作权 1 项。</p> <p>3、该课题组研究人员工作认真，阶段研究成果丰硕，以严谨的学术研究完成了课题研究，研究成果具有较高的学术水平，取得了预期效果。</p> <p>经评议，专家组同意项目通过验收。</p>		
专家签字	曹子元	周春宏	苏峰 黄慎
职务/职称	副研究员	副研究员	研究员 研究员

目 录

(1) 研究背景.....	1
(2) 研究目的和意义.....	2
(3) 研究现状.....	2
(4) 研究方法及内容.....	3
(5) 研究思路及创新点.....	4
(6) 研究计划及步骤.....	4
(7) 结论与建议.....	5
(8) 项目成果... ..	16
(9) 参考文献.....	17
(10) 已发表 7 篇文章附录.....	18
(11) 软件著作权复印件.....	40
(12) 课题组成员培训演示照片.....	41
(13) 已发表 7 篇文章录用通知书.....	43
(14)数字图书馆云存储系统 V1.0 设计说明书.....	43
(15)数字图书馆云存储系统 V1.0 程序关键代码.....	63

大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究

关键词：大数据 云存储 数字资源

绪 论

1 研究背景

美国国会研究协会对 113 个成员图书馆调查结果显示，1999 年数字资源占资源总购置经费的 12.9%，到 2007 年其所占比率高达 46.55%，而到 2009 年这个数字攀升至 56.33%，最高甚至达到 85.4%，我国的情况也与此类似，2010 年高校图书馆用于采购纸质文献的馆均经费约是采购电子资源经费的 1.7 倍，低于 2009 年的 2.16 倍，可见图书馆的数字资源在逐年增加。而作为数字资源重要组成部分之一的网络信息资源数量更是大得惊人。一组名为“互联网上一天”的统计图在各大科技网站上流传，它告诉我们：一天中，约有 2940 亿封电子邮件发出，200 万篇博客在网上发布，两亿五千张照片上传到脸谱网，86.4 万小时的视频上传到 Youtube，1.87 亿个小时的音乐在 Pandora 上播放，这意味着，互联网一天之内产生的信息总量，可以装满 1.68 亿张 DVD 光盘。尽管这些数字看起来庞大，却也仅只是数字资源总量的冰山一角。

数字资源的获取，使用以及共享，相较于传统文献资源有着无可比拟的优势，使其在人们的生活、学习、工作中扮演起越来越重要的角色。然而，正如刘家真教授所说数字技术在给人类带来空前机遇的同时，也正在编织一个数据陷阱。人们发现，数字资源的长期保存显得非常棘手。首先，数字资源是以数字代码形式存储在一定的物理载体之上的，主要以光、电、磁为主。这些材料的载体对存储环境的要求相当高，高温，潮湿，磁场影响都会造成信息的丢失，并且这些载体的寿命都远远短于传统载体纸张成百上千年的寿命，普遍认为，磁性载体保存时间为 10 年左右，光盘保存时间为 10-20 年。其次，数字资源的读取依赖于计算机软件，当技术不断发展，操作系统、软件升级不再支持传统的文件格式，将导致数字资源面临无法利用的尴尬局面。2013 年 12 月，微软宣布自 2014 年 4 月起将停止更新修复 Windows XP 系统，据微软官方所说，XP 系统用户的电脑将极有可能遭遇病毒风暴和恶意软件，保存在这些电脑中的数字文件必然会受到安全威胁。再次，现如今，数字资源已经达到一个非常庞大的量，且仍在持续发展，这就使得保存数字资源的费用无法预估，且数字资源的保存费用不仅包括保存传统纸质文献所需的物理空间和环境控制费用，还涉及到为保证数字资源再现的其他费用，如技术更新方面、数字迁移方面所需要的费用。此外，数字资源的长期保存还存在缺乏公共标准以及法律方面的难题。

2011 年 6 月 7 日，举世瞩目的苹果全球开发者大会 WWDC2011 召开，乔布斯在发布会上发布了本次大会的重点产品 iCloud，宣布苹果将提供云存储服务。而早在此前，云存储就已经出现在人们的视野中，那么云存储到底是什么？要了解云存储，也许先要明白云计算。早在 1961 年，享有“人工智能之父”的美国计算机科学家 John McCarthy 就提出云计算的概念，他认为，计算能力可以像水、电和气一样的公共事业被用户所使用，而在长达 40 年的时间里，这一思想都只停留在梦想阶段。随着分布式计算，并行计算，虚拟化技术的发展，成熟与融合，云计算这一理念终于被人们所了解。美国国家标准和技术研究院(National Institute of Standards and Technology, NIST)将云计算定义为“一种通过网络连接 IT

资源(如服务器、存储、应用和服务等)的应用模式,组成一个共享资源池,向用户提供按需服务,同时能够实现资源的快速部署。”总的来说,云计算实际上是对 IT 基础资源在部署、管理和使用上的一种理念的革新和技术的再创造。那么,云存储到底是什么?事实上,到目前为止,云存储并没有权威的定义,但可以明确的是,云存储不仅是存储技术或设备,更是一种服务。云存储的定义应该由两部分构成:第一,在面向用户的服务形态方面,它是提供按需服务的应用模式,用户可以通过网络连接云端存储资源,实现用户数据在云端随时随地的存储。第二,在云存储服务构建方面,它是通过分布式、虚拟化、智能配置等技术,实现海量、可弹性扩展、低成本、低能耗的共享存储资源。

我国的数字资源长期保存研究始于上世纪 90 年代末期,相较于国外发达国家起步较晚。比较有影响的项目有中国国家图书馆建立的网络信息资源保存试验项目 WICP-project,清华大学参与的数字资源长期保存系统的研究与开发项目,中科院国家科学图书馆进行的数字资源持久保存国际合作实验项目 CSDL 项目,数字化科技信息资源长期保存体系与政策机制 NSDL 项目,以及北京大学主持的 WED 信息博物馆项目,中国高等教育文献保障系统 CALIS 的建立,中国知网的数字资源保存体系的建立与完善等。

2 研究目的和意义

数字资源是承载当今文明的载体,开展数字资源长期保存活动就是保护人类文化遗产,为今后社会的发展提供智力支持。

英国国家图书馆将数字资源长期保存定义为对数字资源的连续获取,或至少这些数字资源无限期地包含着这些信息。可见数字资源长期保存工作不仅包括长期存储与长期获取,还应保证数字信息的完整性,而可靠的存储环境是实现这一目的的保障。由前文可知,当下数字资源长期保存活动遭遇瓶颈,数字资源数量庞大且仍处于不断增长中,而传统的存储系统容量有限,且扩容困难。且数字资源长期保存任务繁重,任何一个机构都无法独立完成,传统存储环境限制了机构间协同合作。此外,由于无法预测数字资源的容量,机构在开展保存活动的过程中,要么购置过多的存储设备造成资源闲置,要么购置过少的设备导致存储空间不足,进而又要为升级设备以扩展存储空间支付额外的费用,无形中增加了数字资源长期保存的成本。而开展数字资源长期保存活动的机构多为图书馆、研究所等非盈利性机构,经费有限。总的来说,可靠的存储环境需满足:支持海量存储和动态扩展,支持跨地域、跨机构的协同合作,有效地降低人力,设备等方面的成本。云存储拥有海量的存储空间,采用并行扩展的架构,具有良好的可扩展性能,且成本低廉。在理论上完全符合数字资源长期保存的需求,故而,将云存储引入到数字资源长期保存研究领域具有一定的实践意义。

3 研究现状

2006 年,“云”的先驱者 Amazon 推出弹性计算云服务(Elastic Compute Cloud, EC2)拉开了轰轰烈烈的“云”序幕。紧接着各大设备商和服务商也纷纷推出了类似的计算能力服务,开始了在“云”市场的角逐。2007 年,Google 与 IBM 带领“云”进驻美国大学校园,包括卡内基梅隆大学、麻省理工大学、斯坦福大学、马里兰大学等,进而掀起了“云”研究的热潮。关于“云”的研究正如火如荼地开展着,国外一些研究机构也展开了关于云存储与数字资源长期保存的研究。

我国关于数字资源长期保存的研究始于 20 世纪 90 年代末,国内学者从政策、法律法规、管理、技术等多种视角对数字资源长期保存进行了广泛研究。云技术的兴起,为数字存储提供一种新的可能,也为数字工作研究者带来新的研究方向,国内研究者迅速抓住机遇,开展起研究。笔者以“云存储”为主题检索了 CNKI 数据库,共获取检索结果 1730 条,为了确保文献的查全率与查准率,笔者再分别结合“数字资源”、“数字文献”、“数字档案”、“电子文

件”、“信息资源”等主题进行检索，经过筛选、剔除不相关文献以及重复文献后，共获取论文 92 篇，其中学位论文 13 篇，期刊论文 79 篇。

国内外关于云存储与数字资源长期保存的研究正如火如荼地展开着，均取得了一些研究成果，相比较而言，我国研究者对云存储与数字资源长期保存的研究与国外发达国家存在不小的差距，主要体现在：

(1) 从研究主体来看：国外研究机构如 Fedora Commons, DSpace Foundation 以及各大高校都广泛开展合作，不仅研究机构，云服务提供商如 Amazon, Google 等也广泛参与。而我国关注云存储与数字资源长期保存的研究者虽然不算少数，但大多集中于高校，图书馆等学术性机构，而像云存储服务提供商，因特网服务提供商，数据库生产商等机构的介入较少，而云存储与数字资源长期保存涉及到云存储服务水平、网络带宽、知识产权等方方面面，需要机构间的协同合作，且该论题涉及计算机，信息管理等多学科方面的知识，因此跨学科合作也相当必要。此外，研究者大多单兵作战，缺乏与同行间的学术对话，这也导致论文重复现象严重。

(2) 从研究内容来看：国外研究者从论题各个方面展开研究，切入视角有宏观也有微观，研究对象也较为丰富。我国的研究则并不深入，大多停留于表面，仅对云存储的概念，特点，框架，优势以及存在的问题进行复述。研究视角也较为单一，大多是从宏观上分析云存储如何应用于数字资源长期保存，而涉及到相关细微问题，通常仅简单介绍一遍，缺乏细致的挖掘与深入的探讨，难以掌握云存储与数字资源长期保存的实质，因而解决问题的对策也是如出一辙，缺乏创新性。

(3) 从研究方法来看：学理研究与实证研究相辅相成，缺一不可，国外研究者深谙此道，开展了多个云存储与数字资源长期保存的相关项目。而我国研究者大多从理论层面下手，关于云存储与数字资源长期保存的实证研究少之又少，这就导致学理研究的科学性，可行性，可信性都大打折扣。

这种研究局面的形成与云存储技术的新颖性，行业发展模式不成熟，以及我国云基础设施建设不完善有关。针对这种情况，我国研究者应广泛展开跨学科，跨机构、跨国界的交流合作，深化对云存储技术的认识。此外，还应积极实践，真正做到学术、商业等的整合，并最终实现数字资源长期保存的目标。

4 研究方法及内容

4.1 研究方法

(1) 文献分析法。查找与阅读国内外相关文献，通过前人的研究，为自己对云存储、对数字资源长期保存以及云存储应用与数字资源长期保存科学认知和深入研究奠定一定的理论基础。

(2) 比较分析法。对云存储与传统存储方式进行比较，进而从技术、成本、利用方面，对云存储满足数字资源长期保存的要求进行论证。对比云存储与传统存储方法的不同之处，结合高校图书馆数字资源的价值、特点构建塔里木大学图书馆数字资源云存储方案。

4.2 研究内容

4.2.1 高校图书馆数字资源概述

- (1) 高校图书馆数字资源的概念
- (2) 高校图书馆数字资源的特点
- (3) 高校图书馆数字资源的价值

4.2.2 高校图书馆数字资源长期保存的云存储选择分析

- (1) 高校图书馆长期保存面临的挑战

- (2) 云存储的原理及其优势
- (3) 云存储是高校图书馆数字资源长期保存的最佳选择

4.2.3 高校图书馆数字资源长期保存的云策略

- (1) 普及知识, 培训人员
- (2) 协调优化
- (3) 撰写与云存储相关的管理标准草稿
- (4) 注重技术的更新与设备维护

4.2.4 构建塔里木大学数字图书馆云平台

5 研究思路及创新点

5.1 研究思路

首先, 通过国内外关于云存储的研究, 明晰云存储的基本概念, 了解云存储的关键技术, 总结其技术优势, 并通过与传统存储策略的对比分析, 得出云存储在成本以及利用方面的优势, 为研究云存储在数字资源长期保存中的应用奠定基础。

其次, 深入挖掘塔里木大学图书馆存储主体数字资源的价值, 分析该存储主体数字资源的特点, 结合云存储的优势, 为塔里木大学图书馆构建合适的云存储平台提出建议。

5.2 创新点:

- (1) 调查与分析当前大数据环境下, 高校图书馆数字资源的概念、特点和价值。
- (2) 根据调查结果分析在大数据环境下, 高校图书馆长期保存面临的挑战, 并且根据分析云存储的原理和优势, 得出云存储是高校图书馆数字资源长期保存的最佳选择。
- (3) 提出了高校图书馆数字资源长期保存的云策略, 给出一系列的实施方案, 最终构建塔里木大学数字图书馆云平台。

6 研究计划及步骤

6.1 研究计划

(1) 文献调研。本课题充分利用各类中英文数据库, 广泛查阅、筛选、归纳资料。2015年5月至6月。

(2) 形成思路。课题组成员探讨提炼大纲, 探讨内容包括: 高校图书馆数字资源的概念、特点和价值; 云存储的原理及其优势。2015年7月至8月。

(3) 对比分析。通过对比分析, 得出结论云存储是高校图书馆数字资源长期保存的最佳选择。2015年9月至10月。

(4) 制定云存储策略。高校图书馆数字资源长期保存的云策略包括: 普及知识, 培训人员; 协调优化; 撰写与云存储相关的管理标准草稿; 完成技术的更新与设备维护。2015年11月至12月。

(5) 搭建平台。构建塔里木大学数字图书馆云平台。2016年1月至3月。

6.2 本课题研究计划分为四个阶段:

第一阶段: 2015年5月-6月, 查阅资料、筛选文献, 进行架构设计, 制定研究大纲。

第二阶段: 2015年7月-12月, 根据研究大纲和计划, 课题组成员以实际情况安排研究任务撰写相关研究论文。

第三阶段: 2016年1月-3月, 构建塔里木大学数字图书馆云平台。

第四阶段: 2016年4月-5月, 完成研究报告《大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究》, 字数2万字左右。

7 结论与建议

数字资源是现代文明的载体。为了确保人类文化的传承与发展，必须通过技术、经济、政策等多种手段，科学有效地开展数字资源长期保存活动。数字资源是信息技术发展的产物，在考虑对数字资源进行长期有效保存的时候，首先要突破长期保存的技术难题。云存储是IT技术的最新成果，它支持海量存储和动态扩展，实现了协同保存、按需使用并且成本低廉，因此一经提出就受到国内外数字资源长期保存研究领域学者的关注。

将云存储应用于图书馆数字资源长期保存毕竟是一个全新的课题，在本课题中，笔者仅仅从理论层次对该课题进行了初步探索，如何将科学选择图书馆数字资源与云存储系统进行无缝结合、云存储的安全保障等都是图书馆数字资源长期保存中需要关注的问题。这些问题笔者将在今后的研究中继续探索。

7.1 数字图书馆面临的挑战

7.1.1 图书馆数据存储现状

目前，图书馆采用的主要存储技术包括单一磁盘存储数据、磁带备份和 RAID(Redundant Arrays of Independent Disks)磁盘阵列存储数据以及 DAS(Direct Attached Storage)直接附加存储、NAS(Network Attached Storage)网络附加存储和 SAS(Storage Area Network)存储区域网络等网络存储系统。

(1)磁盘阵列存储(RAID)

磁盘阵列是由很多价格较便宜的磁盘，组合成一个容量巨大的磁盘组，利用个别磁盘提供数据所产生加成效果提升整个磁盘系统效能。

(2)直接附加存储(DAS)

直接附加存储是指将存储设备通过SCSI接口或光纤通道直接连接到一台计算机上^[3]。

(3)网络附加存储(NAS)

网络附加存储是将存储设备通过标准的网络拓扑结构（例如以太网），连接到一群计算机上。NAS是部件级的存储方法，它的重点在于帮助工作组和部门级机构解决迅速增加存储容量的需求。

(4)存储区域网络(SAN)

采用光纤通道（Fiber Channel，简称FC）技术，通过光纤通道交换机连接存储阵列和服务器主机，建立专用于数据存储的区域网络。SAN经过十多年历史的发展，已经相当成熟，成为业界的事实标准。

几种数据存储方式优缺点如表 1 存储方式比较所示。

表 1 存储方式比较

	RAID	DAS	NAS	SAN
存储性能	一般	较好	较好	优
实施难度	容易	容易	较难	难
数据共享	无	无	较好	优
扩展性	一般	难	较好	优
可靠性	差	差	一般	优
安全性	差	差	较好	优
安装维护成本	较低	一般	高	昂贵

SAN 经过十多年历史的发展，已经相当成熟，成为业界的事实标准。它具有传输速度快，扩展能力强，共享能力高以及可靠性、安全性好的优点，但是 SAN 技术比较复杂，实

施难度较大，安装以及后期升级、维护费用高，大部分图书馆特别是图书馆难以承受。目前，图书馆急需一种存储能力强，高度数据共享，可靠性和安全性优越，同时易于实施，价格低廉的数据存储方式。

7.1.2 图书馆数字资源存储存在的问题

目前，图书馆数字资源在存储过程中主要面临以下问题：数字资源高速增长，存储技术标准不统一以及设备兼容性差等。

(1) 数字资源高速增长

图书馆数字资源内容涵盖文本、图片、音频、视频、网络资源等。以国家图书馆为例，2012年《国家图书馆数字资源与服务概览》中指出，截止2012年国家图书馆主要包括电子图书百万余种，电子期刊、电子报纸约数十万种，学位论文约四百万篇，音频资料约101.6万首，视频资料约8.9万小时，且内容还在不断更新。近几年国家图书馆数字资源增长如图1数字资源高速增长所示。

(2) 存储技术标准不统一

各种存储技术没有统一的技术标准是目前数字资源存储所面临的重要问题。存储系统的建设没有统一的规范，使得系统从最初的立项到结果的验收都无据可循，从而无法保证建设效果，更加无法应对未来数字资源的高速增长，最终造成了资源的大量浪费。

(3) 设备兼容性差

企业为了打到市场垄断的效果，将设备的兼容性降低。图书馆为了满足自身需求，丧失选择权被迫投入巨额的购买成套的设备，且后期维护艰难。

再加上图书馆为了达到数字化的要求，在没有清楚了解自身需求、对系统运行缺乏理性分析的基础上，盲目追求主流技术，造成所建存储系统投入大量资金却无法自身需求的现象。

国家图书馆数字资源高速增长图



图1 数字资源高速增长

7.2 云存储技术在图书馆应用的优势

7.2.1 图书馆数字资源存储需求分析

随着数字资源以爆炸性增长，图书馆在不断扩大数字资源存储容量的同时，也在考虑如何节约有限的资金投入、高度数据共享，可靠性和安全性优越等方面的问题。

(1)经济适用是图书馆建设存储系统时考虑的重要因素。作为资金有限的国家单位，图书馆需要根据自身发展和资金现状选择性价比较高的存储方案，切不可盲目追求投入昂贵技术先进，却暂时无法应用的设备。

(2)数据高度共享是图书馆存储系统建设的重要条件。图书馆资源数字化是图书馆发展的必然趋势,这有利于提供图书馆资源的使用效率,提高数据的共享性,也有利于资源的保护和利用。

(3)安全可靠是数字图书馆服务的保证。为了避免不可抗力和人为因素对存储系统的损害,图书馆要求建设存储系统的时候,必须要有安全的存储备份方案。

7.2.2 云存储的技术特点

与传统的存储技术相比,云存储技术具有绝对优势,二者对比如表 2 云存储与传统存储对比所示。

表 2 云存储与传统存储对比

比较项	云存储	传统存储
架构	不仅是一种架构,更是一种服务。底层采用分布式架构和虚拟化技术,易于扩展,单点失效不影响整体服务。	针对某种特殊应用而采用的专用、特定的硬件组件构成的架构
服务模式	按需使用,按使用计费,服务提供商可迅速交付和响应	用户通过整机购买或租赁获取存储容量
容量	支持 PB 级以上无限扩展	针对某个特定的应用存储,由使用需求方决定容量,难于扩展。
数据管理	不仅提供传统访问方式,而且提供海量数据的管理和对外的公众服务支撑,同时采用保护数据安全的策略,用户可灵活配置。	用户数据管理员可见,信息不够安全。通常使用 RAID 提供数据保护,用户无法灵活配置个性化存储策略和保护策略。

结合图书馆数字资源存储需求分析结果,云存储技术在图书馆数字资源存储中应用的优势如下。

(1) 节约成本

云存储技术的应用可以为图书馆数字化过程中从实施到后期的维护节约成本。目前的存储技术需要图书馆必须购买昂贵的软件和硬件,才能将数字资源放在服务器上。此外,图书馆还要聘请专业的 IT 人员,管理这些硬件和软件的维护工作,并且还要更新这些设备和软件。云存储技术在很大程度上解决了这些问题。

(2) 更好的数据备份并可以异地处理日常数据

云存储中硬盘坏掉,数据会自动迁移到别的硬盘,大大减轻了管理人员的工作负担。对云存储来说,再多的存储服务器,在管理人员眼中也只是一台存储器,每台存储服务器的使用状况,通过一个统一管理界面监控,使得维护变得简单和易操作。

(3) 更多的访问和数据共享

图书馆的借阅者不再需要通过本地网络来访问图书馆相关的数字资源,云存储可以让借阅者在任何地方,任何时间访问他们需要的数据。图书馆也不需要花费上千万美元来打造最新技术和最新应用来创造最好的系统,这就为图书馆数字资源最大程度的数据共享提供了可能。

以上特点为云存储技术在图书馆数字化中的应用铺平道路。除了以上特点,云存储还具有超强的计算能力、永远以客户为核心、高度的安全性、较优的性价比和节能环保等优势。

7.3 云存储技术在图书馆管理系统中的应用模式

数字图书馆的云存储系统是基于高性能网络传输技术,通过集群技术、分布式文件系统和网络计算技术等,根据客户服务需求将计算资源与已存储的数据资源按需分配。图书馆的数据资源云存储系统是一个由高性能网络传输设备、超大容量数据存储设备、超高性能服务

器、特色服务应用软件和形式多样客户端等多个部分组成的复杂系统。通过虚拟化技术对云图书馆存储和数据操作进行管理，系统形成了以存储设备为核心，通过软件为客户提供数据存储和业务访问功能。数字图书馆云存储系统的结构模型如图 2 所示。

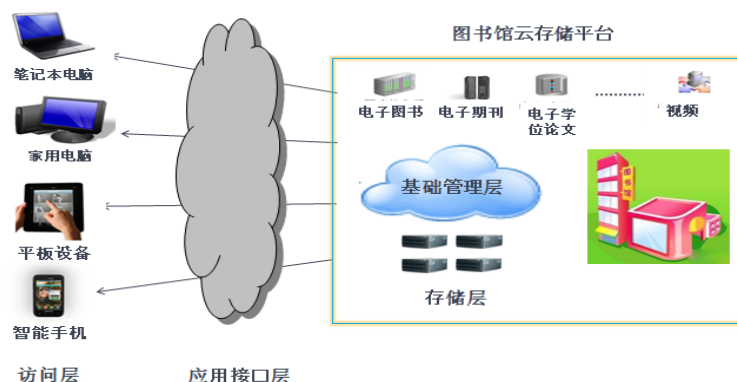


图 2 数字图书馆云存储系统结构模型

7.3.1 存储层

存储层是数字图书馆云存储系统的基础。图书馆将所有数字资源通过存储设备统一存放于存储层中，分布于世界不同地域的存储设备再通过各种网络设备连接在一起，由统一的存储设备管理系统管理，并利用虚拟化技术忽略掉硬件之间的差异，对外提供服务，实现海量数据的统一管理。存储设备包括：光纤通道存储设备、IP 存储设备和 DAS 存储设备等。存储层的功能结构图如图 3 所示。

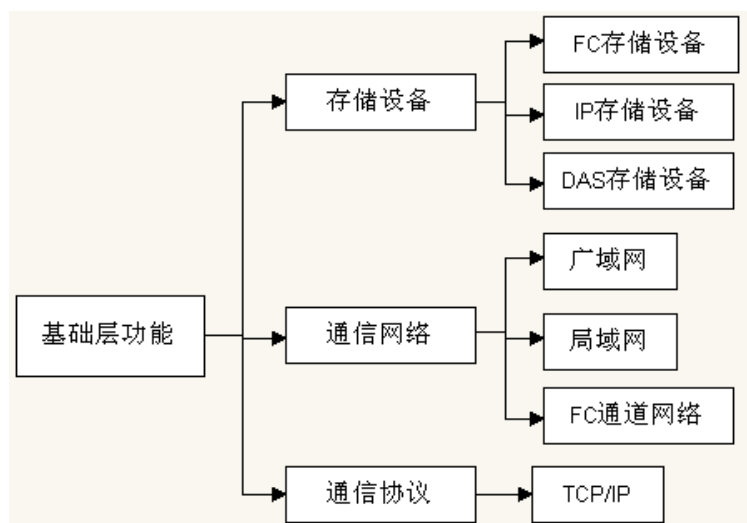


图 3 存储层的功能结构图

7.3.2 基础管理层

基础管理层是云存储的核心部分，也是最难以实现的部分。该层通过集群技术、分布式文件系统和网格计算技术等相关技术，实现系统中多个存储设备间的协同工作，并对外提供统一的服务标准，为存储系统提供更高效、更强大、更优越的数据访问性能打下基础。此层还具有内容分发系统、数据加密技术、数据备份和容灾技术和措施等保证了存储系统中的数据的安全性。基础管理层的功能结构图如图 4 所示。

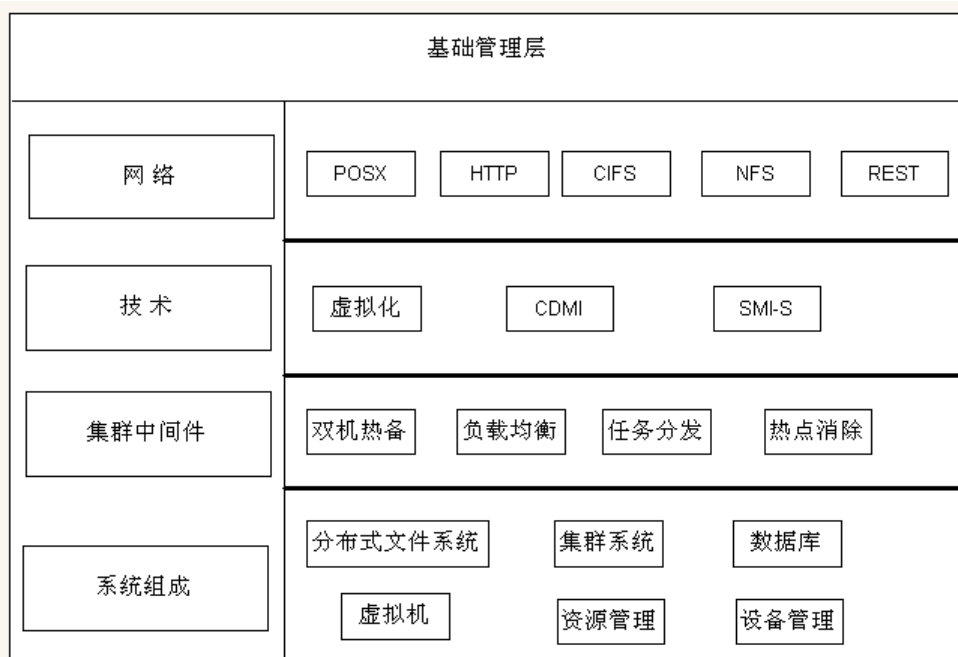


图4 基础管理层的功能结构图

7.3.3 应用接口层

应用接口层是云存储系统中最灵活多变的部分。不同的云存储运营单位可以根据实际业务类型，开发不同的应用服务接口，为不同的图书馆需求提供不同的应用服务。该层满足了云图书馆基于网络的跨平台的二次应用程序开发的要求，为图书馆建立特色平台，如：视频监控应用平台、IPTV 和视频点播应用平台、网络硬盘引用平台和远程数据备份应用平台等，提供了技术支持。应用接口层的功能结构图如图 5 所示。

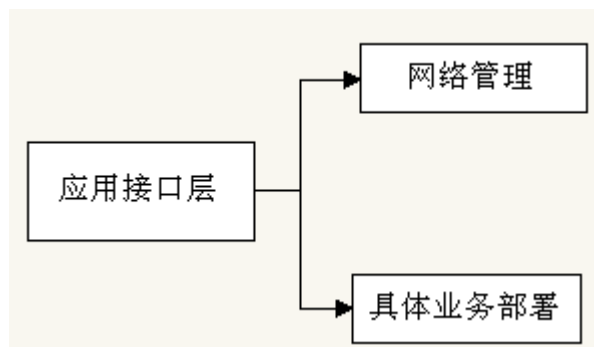


图5 应用接口层的功能结构图

7.3.4 访问层

任何一个授权用户都可以通过标准的公用应用接口来登录数字图书馆云存储系统，享受相关服务。此时的数字图书馆馆员以及数字读者不再需要了解底层的硬件设备类型，也不需要知道数字资源存放的物理位置，只需要在获得授权后，直接调用标准的 API，享受相关服务。此服务的使用不受读者所使用的终端设备限制。访问层的功能结构图如图 6 所示。

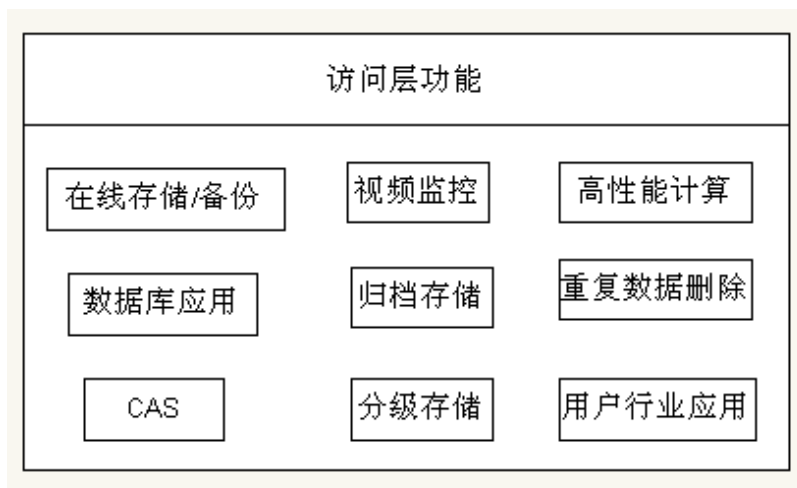


图6 访问层的功能结构图

7.4 系统设计

7.4.1 图书馆数字资源存储需求分析

随着数字资源以爆炸性增长，图书馆在不断扩大数字资源存储容量的同时，也在考虑如何节约有限的资金投入、高度数据共享，可靠性和安全性优越等方面的问题。

(1) 经济适用是图书馆建设存储系统时考虑的重要因素。作为资金有限的国家单位，图书馆需要根据自身发展和资金现状选择性价比比较高的存储方案，切不可盲目追求投入昂贵技术先进，却暂时无法应用的设备。

(2) 数据高度共享是图书馆存储系统建设的重要条件。图书馆资源数字化是图书馆发展的必然趋势，这有利于提供图书馆资源的使用效率，提高数据的共享性，也有利于资源的保护和利用。

(3) 安全可靠是数字图书馆服务的保证。为了避免不可抗力和人为因素对存储系统的损害，图书馆要求建设存储系统的时候，必须要有安全的存储备份方案。

7.4.2 系统架构设计

数字图书馆云存储系统以图书馆海量数字资源存储和共享为研究对象，以图书馆—读者为基本模式，从图书馆馆员、管理员和读者的不同角度设计该系统。系统包括数字资源存储模块、数据管理模块和索引检索模块三部分，其系统模块图如图7。从图中可以看出，图书馆馆员将纸质版图书、期刊、论文、报纸等纸质资源，以及光盘等数字资源，通过传统手段录入到计算机中，形成数字资源。针对电子数据文件大小异构的特点，图书馆馆员通过数据存储模型，将视频、音频等大数据文件利用 HDFS 分布式文件系统存放；将图片、文档等较小的文件利用 HBase 直接存放。管理员利用数据管理模块，对数据文件进行上传，下载和删除等操作。做图书馆的服务对象，读者可以在 PC 机、手机、平板电脑等不同平台，利用系统提供的各种查询方式进行资源检索。

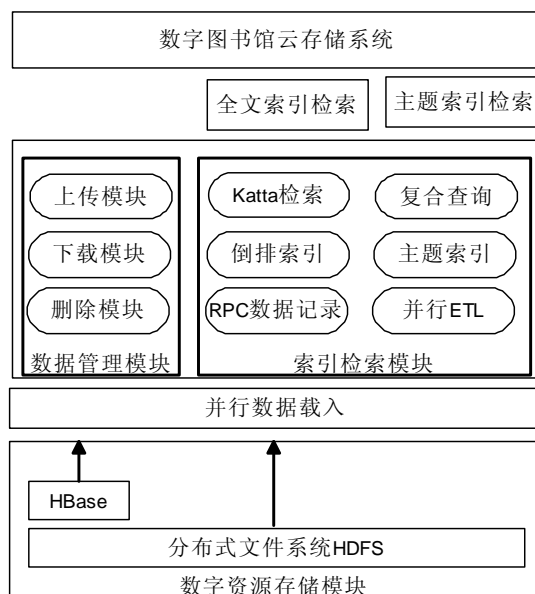


图 7 系统模块图

7.4.3 系统部署

本系统硬件节点部署示意图如图 8 所示，分为三层：数据存储层、应用服务层和用户层。数据存储层由若干不同类型的存储节点或者服务节点组成，用于存储图书馆数字资源。应用服务层主要由具有不同功能的服务器组成，其中一部分服务器用于运行 NameNode、SecondNameNode、Hmaster 和 JobTracker 服务；一部分用于单独运行 ZooKeeper 服务；一部分用于运行 Katta 分布式索引检索服务；一部分用于运行 DataNode、TaskTracker 和 HregionServer 服务。用户层中的读者可以通过不同形式的平台使用该存储系统。

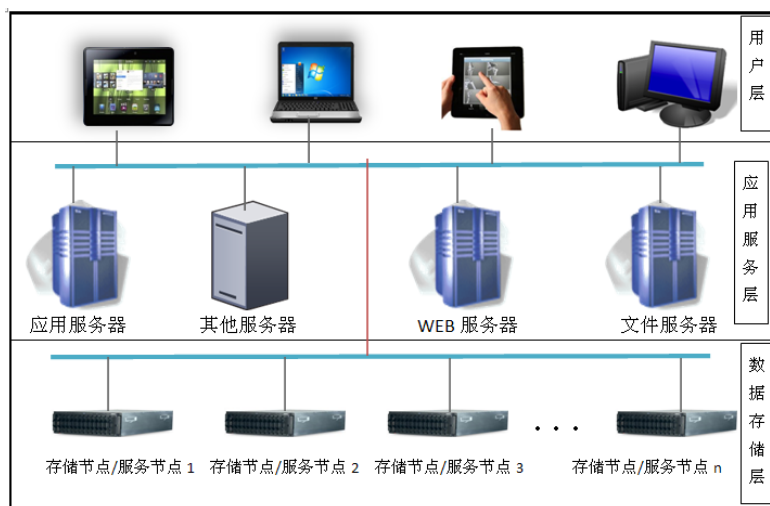


图 8 系统部署示意图

7.5 数据处理流程设计

数字图书馆云存储系统的数据流程示意图如图 9 所示。系统对外提供图书馆馆员使用的管理接口和普通读者使用的查询接口。图书馆馆员借助于图形管理界面，实现数字资源的管理（上传、下载、删除等）操作。数据管理过程中，数据信息通过网络平台在当地计算机和存储集群之间的传递。数字资源按照类别，分别存储于 HBase 或直接存放于 HDFS。读者获

得授权后，可以浏览当前数字资源，通过系统利用 HBase 自身的 RowKey 索引，进行精确查询、模糊查询和复合查询。

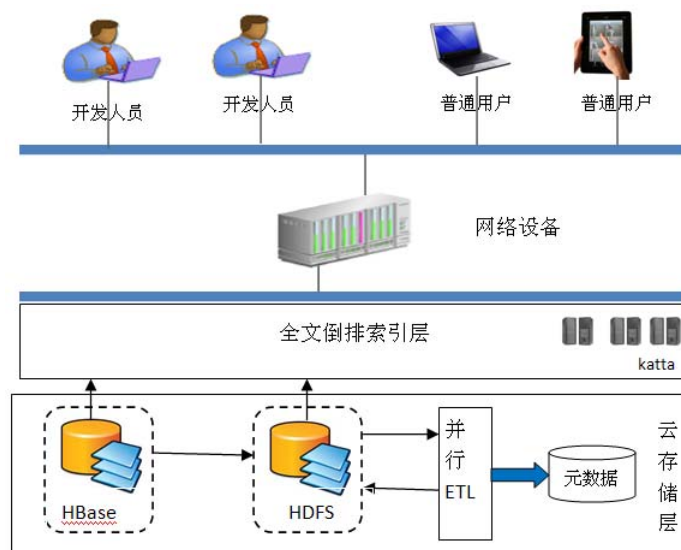


图 9 数据流程示意图

7.5.1 文件上传

数字资源的上传主要是对 HDFS 分布式系统和 HBase 数据库的操作。数字图书馆云存储系统根据数据文件大小进行了分类存储，对不同数据文件的上传进行了不同设计。图书馆馆员上传文件之前，先调用 HBase 数据库，获取 TableName 表的 Scanner，检查该文件是否存在。如果文件已经存在，则提示文件已经存在，是否需要更新。如果文件没有存在，则按照数据文件大小进行分类存储。

(1) 较小文件上传

较小的数据文件包括：文本、图片和网络资源，直接存储在 HBase 数据库中。文本文件要进行 RPC 通信，建立倒排索引，监控更新消息，记录对文本表的更新操作。图片上传不需要相应的通信。

(2) 较大文件上传

较大的数据文件包括：音频、视频、高清扫描图片等，需要借助分布式文件系统 HDFS 存储，同时在 HBase 数据库中有相应的地址映射表。在地址映射表 MyData 中查重，若存在则更新，否则添加映射关系。

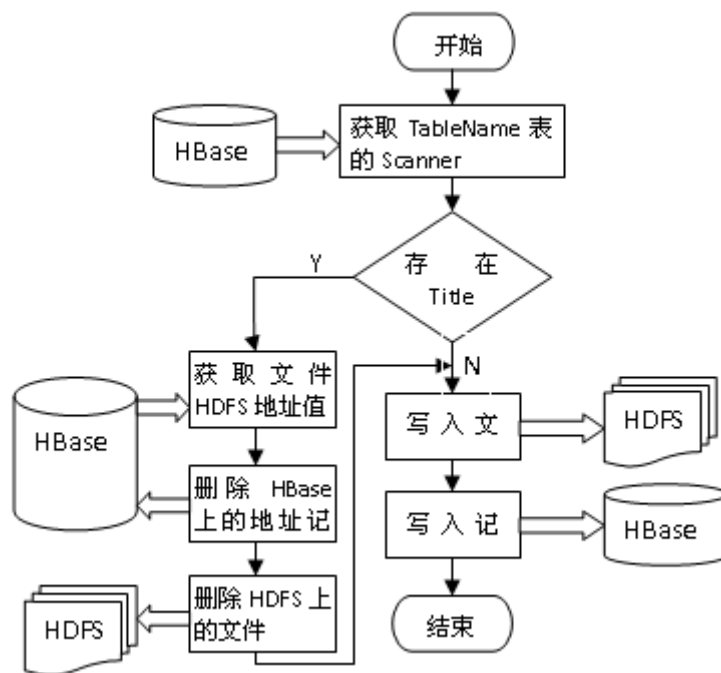


图 10 文件上传流程

7.5.2 文件下载

文件的下载模块也是根据文件大小采用不同的方式进行下载的。对于相对较小的数据文件直接获取数据库 HBase 数据表 Infor 中的 content 字段中的内容，而对于相对较大的数据文件需要获取 Infor 中的 content 字段中的值后，从分布式系统 HDFS 中下载文件内容。

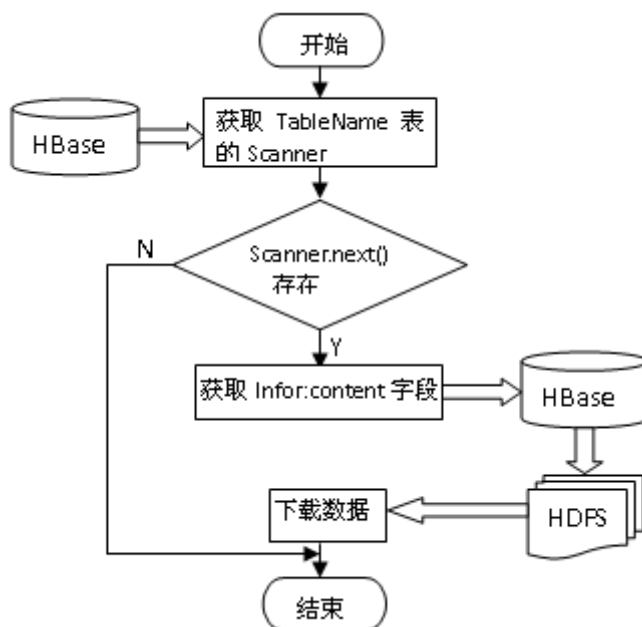


图 11 文件下载流程

7.5.3 索引模块

针对文本文件内容的检索相对比较复杂，本系统通过分布式检索工具 Katta，利用分布式倒排索引 Lucene，将索引进行分片存储加以实现，检索流程图如图 12 所示。客户端首先

向检索服务器群组中的分布式查询服务器发送查询请求,由分布式查询服务器根据检索服务器集群中设备的当前状态将查询任务发送,然后检索服务器将检索结果返回给分布式查询服务器,用户最终得到查询结果,完成查询任务。

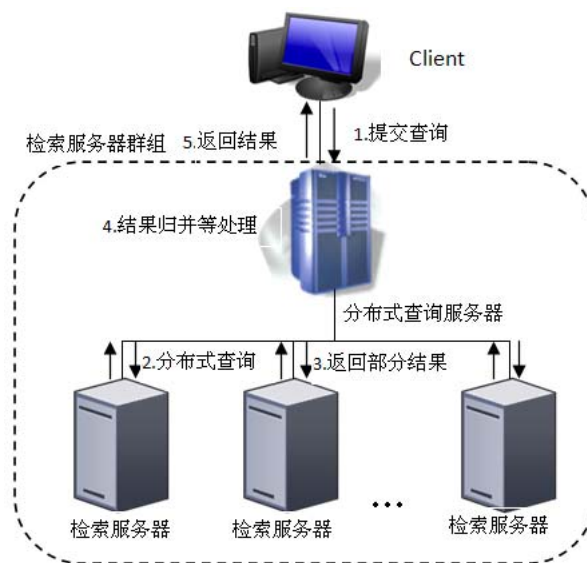


图 12 基于 Katta 的检索流程图

7.6 主要关键技术

7.6.1 Lucene 索引结构

传统的索引方式^[6]是以文档的ID为关键字,表中记录文档中每个字的位置信息,查找时扫描表中每个文档中字的信息直到找出所有包含查询关键字的文档。这种组织方法在建立索引的时候结构比较简单,建立比较方便且易于维护。但是在查询的时候需对所有的文档进行扫描以确保没有遗漏,这样就使得检索时间大大延长,检索效率低下。倒排索引以字或词为关键字进行索引,表中关键字所对应的记录表项记录了出现这个字或词的所有文档,一个表项就是一个字表段,它记录该文档的ID和字符在该文档中出现的位置情况。由于每个字或词对应的文档数量在动态变化,所以倒排表的建立和维护都较为复杂,但是在查询的时候由于可以一次得到查询关键字所对应的所有文档,所以效率高。在全文检索中,能够达到检索的快速响应。文档文件不同于其他文件,需要对其全文的内容进行检索。为了提高效率,数字图书馆云存储系统对文档文件的检索采用倒排索引。

7.6.2 行列式存储模式

Facebook 曾在 2010 ICDE (IEEE International Conference on Data Engineering) 会议上介绍了数据仓库 Hive。Hive 存储海量数据在 Hadoop 系统中,提供了一套类数据库的数据存储和处理机制。它采用类 SQL 语言对数据进行自动化管理和处理,经过语句解析和转换,最终生成基于 Hadoop 的 MapReduce 任务,通过执行这些任务完成数据处理。数字图书馆云存储系统中较大文件的上传存储便利用了数据仓库 Hive,即借助分布式文件系统 HDFS 存储,同时在 HBase 数据库中有相应的地址映射表。在地址映射表 MyData 中查重,若存在则更新,否则添加映射关系。Hive 数据仓库的系统结构如图 13 所示。

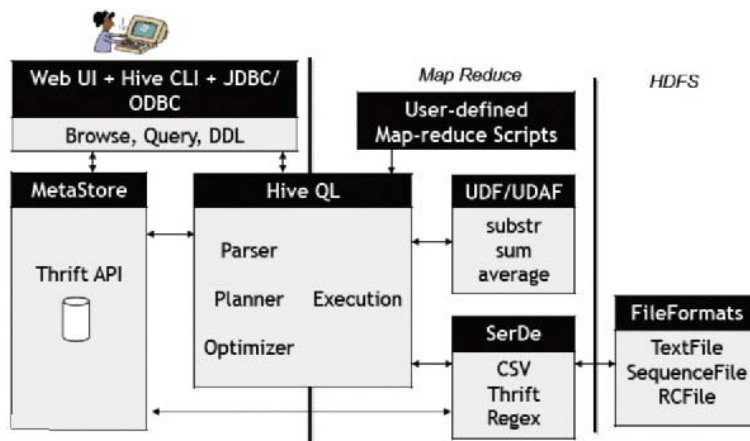


图 13 Hive 数据仓库的系统结构

7.7 系统实现

数字图书馆云存储系统考虑到多用户性和可扩展性，系统采用 C/S 的体系结构；云存储系统为了给公众提供一个查询的平台，采用网站形式。整个系统以 JSP 为开发平台，采用面向对象的开发思想。在编写系统公共类的基础上实现相应功能。系统界面图如图 14 所示。



图 14 系统界面示例图

数字图书馆云存储系统采用 RedHat Enterprise Linux5 操作系统，使用 13 台不同型号服务器组建 Hadoop 集群，使用 Hadoop, HBase, Zookeeper, Lucene, Katta 等多种开发工具。该系统已在新疆塔里木大学图书馆应用，满足了图书馆数字资源存储形式多样化、总量迅猛增长、具有动态可扩展性以及高度共享性的要求，取得了良好的经济效益和社会效益。

7.8 结束

云存储技术为构建现代图书馆存储系统提供了新的思维方式。构建数字图书馆云存储平台，可以解决许多长期困扰图书馆数字资源存储管理和数据共享中存在的问题，通过云存储可提升图书馆各类数据资源的利用率，提高图书馆的服务水平。云存储作为一种新技术，在实际的应用过程可能会面临很多问题，在以后的研究过程，将以数字读者需求为基本点，运用成熟的技术，集合多方面的力量来完善数字图书馆云存储系统。

8 项目成果（发表的文章、开发的软件、取得的实践效果等）

8.1 发表的文章：

序号	论文题目	期刊名称	刊期	作者	发表或录用情况
1	大数据时代下云技术在图书馆数据存储中的应用	《农业图书情报学刊》	2015. 10	席亚军	已发表
2	数字图书馆云存储系统的设计与实现	《河南图书馆学刊》	2015. 10	席亚军	已发表
3	高校图书馆数字资源建设现状及发展趋势研究	《河南图书馆学刊》	2016. 02	席亚军	已发表
4	基于 RFID 技术的图书馆编目系统建设探索	《农业网络信息》	2016. 02	席亚军	已发表
5	新媒体背景下西部高校图书馆信息服务研究	《农业网络信息》	2016. 03	席亚军	已发表
6	边疆高校图书馆图书借阅量下降分析及对策-以塔里木大学图书馆为例	《江苏科技信息》	2016. 05	席亚军	已发表
7	基于 TPI 系统的塔里木大学图书馆特色数据库建设实践	《农业网络信息》	2016. 04	席亚军	已发表

8.2 软件著作权：

序号	软件名称	登记号	开发完成时间	著作权人	登记时间
1	数字图书馆云存储系统 V1.0	2015SR147383	2015.06.02	塔里木大学：席亚军	2015.7.30

9 参考文献:

- [1]马晓亭. 大学图书馆数字资源采购评估指标体系研究[J]. 现代情报, 2010, 10: 92-94.
- [2]吴高, 韦楠华. 我国图书馆数字资源采购问题与对策研究[J]. 图书馆建设, 2012, 12: 20-24.
- [3]刘家真. 跨越数字困境的第一步[J]. 情报学报, 2002, 04: 458-464.
- [4]周建军, 李肖军. 数字资源长期保存的技术障碍及对策[J]. 兰台世界, 2009, 16: 4-5.
- [5]徐义全. 电子文件的特性与长期保存[J]. 档案学研究, 2000, 01: 53-57.
- [6]张继平主编. 云存储解析[M]. 北京:人民邮电出版社, 2013: 12.
- [7]李素梅. 基于云计算的高校图书馆信息资源建设[J]. 河南图书馆学刊, 2011, 31(3): 151-153.
- [8]陈臣. 数字图书馆云存储系统安全架构与安全策略研究[J]. 现代情报, 2011(9): 44-47.
- [9]刘鹏. 中国云存储发展报告[M]. 北京:电子工业出版社, 2013: 100-126.
- [10]杨信廷. 蔬菜安全生产管理及质量追溯系统设计与实现[J]. 农业工程学报, 2008, 24(3): 162-166.
- [11]吴海华. 基于云计算数字图书馆的应用实践研究[J]. 河南图书馆学刊, 2013, 33(4): 74-75.

大数据时代下云技术在图书馆数据存储中的应用

席亚军

(塔里木大学图书馆, 新疆 阿拉尔 843300)

摘要: 随着信息技术的快速发展, 图书馆目前的存储方式已不能很好地满足数字资源存储形式多样化、总量迅猛增长、安全性要求越来越高、具有动态可扩展性以及高度共享性。分析了大数据时代数字图书馆存储面临的挑战, 介绍了云存储在图书馆数字化建设上的优势, 提出了大数据时代下云存储技术在图书馆管理系统中的应用模式。该模式解决了图书馆数字资源海量存储需求的问题。

关键词: 大数据; 云存储技术; 数字资源

中图分类号: G250.76; G255.76

文献标识码: A

文章编号: 1002-1248 (2015) 11-0005-04

The Application of Cloud Storage in Library at the big data age

XI Ya-jun

(Tarim University, Xinjiang Alar 843300, China)

Abstract: With the rapid development of information technology, the present storage methods of library can't settled for the diversification of stored form, the rapid growth of storage gross, the growing request of security, the high sharability and dynamic extensibility. Cloud storage is a network system, which includes a large number of various types of storage devices. These devices are set up by the software to work together to provide data storage and business functions. This paper analyzed the challenges of digital library storage at the big data age and the advantages of cloud storage, and proposed an application mode of cloud storage technology in the library management system, which solved the problem of mass storage in library.

Keywords: Big data; Cloud storage technology; Digital resource

1 引言

大数据时代下, 信息量呈现了海量增长的趋势。作为信息的主要表现形式, 数字资源在图书馆中的存储呈现了规模不断扩大的特点, 自 20 世纪 90 年代以来, 投入到数字资源建设上的专项经费逐年上涨, 且增幅较大; 形势多样化, 目前图书馆数字资源内容涵盖文本、图片、音频、视频和网络资源等; 比重不断增加, 数字资源已经成为图书馆主要的资源类型, 且国外已经出现了“无书图书馆”^[1]。如何充分利用现有的技术, 建设安全、可靠、经济有效地图书馆数字存储系统, 成为图书馆发展所面临的新问题。

云存储是一个以数据存储和管理为核心的云计算系统。该系统是一种将分布式计算、网格计算、并行

计算以及互联网技术结合起来的新的 IT 资源提供模式。它可将动态、可伸缩的 IT 资源以服务的方式通过互联网提供给用户^[2]。云存储技术的应用可以改变现有图书馆的服务模式, 使图书馆将海量、繁杂的数字信息资源存储于云服务器中, 所有的运算均在服务器终端完成。海量的数字信息资源连接在一起, 为数字资源的共知、共建、共享提供新途径, 真正实现数字资源的集约化。

针对目前图书馆数据存储的方式、缺点及数据发展的趋势, 笔者研究了其未来数据存储所面临的挑战, 提出了大数据时代下云存储技术在图书馆管理系统中的应用模式。

收稿日期: 2015-06-01

基金项目: 2015 年 CALIS 农学中心研究项目“大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究”(项目编号: 2015032)

作者简介: 席亚军 (1981-), 男, 馆员, 学士, 研究方向: 数字图书馆。

2 数字图书馆面临的挑战

2.1 图书馆数据存储现状

目前,图书馆采用的主要存储技术包括单一磁盘存储数据、磁带备份和 RAID (Redundant Arrays of Independent Disks) 磁盘阵列存储数据以及 DAS (Direct Attached Storage) 直接附加存储、NAS (Network Attached Storage) 网络附加存储和 SAS (Storage Area Network) 存储区域网络等网络存储系统。

(1) 磁盘阵列存储(RAID)

磁盘阵列是由很多价格较便宜的磁盘组合成一个容量巨大的磁盘组,利用个别磁盘提供数据所产生加成效果提升整个磁盘系统效能^[3]。

(2) 直接附加存储(DAS)

直接附加存储是指将存储设备通过 SCSI 接口或光纤通道直接连接到一台计算机上^[3]。

(3) 网络附加存储(NAS)

网络附加存储是将存储设备通过标准的网络拓扑结构(例如以太网),连接到一群计算机上。NAS 是部件级的存储方法,它的重点在于帮助工作组和部门级机构解决迅速增加存储容量的需求^[3]。

(4) 存储区域网络(SAN)

采用光纤通道(Fiber Channel,简称 FC)技术,通过光纤通道交换机连接存储阵列和服务器主机,建立专用于数据存储的区域网络。SAN 经过十多年历史的发展,已经相当成熟,成为业界的事实标准^[3]。

几种数据存储方式优缺点如表 1 存储方式比较所示。

表 1 存储方式比较

	RAID	DAS	NAS	SAN
存储性能	一般	较好	较好	优
实施难度	容易	容易	较难	难
数据共享	无	无	较好	优
扩展性	一般	难	较好	优
可靠性	差	差	一般	优
安全性	差	差	较好	优
安装维护成本	较低	一般	高	昂贵

SAN 经过十多年历史的发展,已经相当成熟,成为业界的事实标准。它具有传输速度快,扩展能力强,共享能力高以及可靠性、安全性好的优点,但是 SAN 技术比较复杂,实施难度较大,安装以及后期升级、维护费用高,大部分图书馆难以承受。目前,图书馆急需一种存储能力强,高度数据共享,可靠性和安全性优越,同时易于实施,价格低廉的数据存储方式。

2.2 图书馆数字资源存储存在的问题

目前,图书馆数字资源在存储过程中主要面临以下问题:数字资源高速增长,存储技术标准不统一以及设备兼容性差等。

2.2.1 数字资源高速增长

图书馆数字资源内容涵盖文本、图片、音频、视频、网络资源等。以国家图书馆为例,2012 年《国家图书馆数字资源与服务概览》中指出,截止 2012 年国家图书馆主要包括电子图书百万余种,电子期刊、电子报纸约数十万种,学位论文约四百万篇,音频资料约 101.6 万首,视频资料约 8.9 万小时,且内容还在不断更新。近几年国家图书馆数字资源高速增长,如图 1 所示。

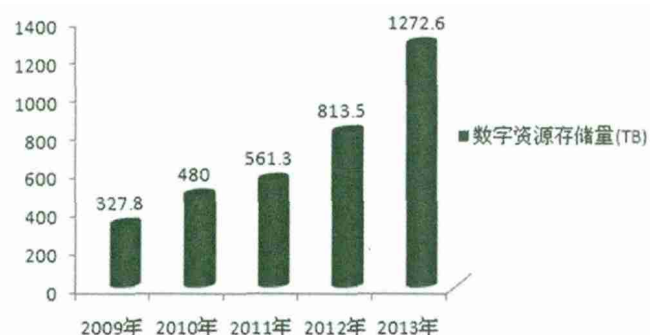


图 1 数字资源高速增长

2.2.2 存储技术标准不统一

各种存储技术没有统一的技术标准是目前数字资源存储所面临的重要问题。存储系统的建设没有统一的规范^[4],使得系统从最初的立项到结果的验收都无据可循,从而无法保证建设效果,更加无法应对未来数字资源的高速增长,最终造成了资源的大量浪费。

2.2.3 设备兼容性差

企业为了达到市场垄断的效果,将设备的兼容性降低^[4]。图书馆为了满足自身需求,丧失选择权,被迫投入巨额的成本购买成套的设备,且后期维护艰难。再加上图书馆为了达到数字化的要求,在没有清楚了解自身需求、对系统运行缺乏理性分析的基础上,盲目追求主流技术,造成所建存储系统投入大量资金却无法自身需求的现象。

3 云存储在图书馆应用的优势

3.1 图书馆数字资源存储需求分析

随着数字资源的爆炸性增长,图书馆在不断扩大数字资源存储容量的同时,也在考虑如何节约有限的资金投入、高度数据共享,可靠性和安全性优越等方面的问题。

(1) 经济适用是图书馆建设存储系统时考虑的重

要因素。作为资金有限的国家单位,图书馆需要根据自身发展和资金现状选择性价比较高的存储方案,切不可盲目追求投入昂贵的先进技术,却暂时无法应用的设备。

(2) 数据高度共享是图书馆存储系统建设的重要条件。图书馆资源数字化是图书馆发展的必然趋势,这有利于提供图书馆资源的使用效率,提高数据的共享性,也有利于资源的保护和利用。

(3) 安全可靠是数字图书馆服务的保证。为了避免不可抗力和人为因素对存储系统的损害,图书馆要求建设存储系统的时候,必须要有安全的存储备份方案。

3.2 云存储的技术特点

与传统的存储技术相比^[5],云存储技术具有绝对优势,二者对比如表 2,云存储与传统存储对比所示。

3.2.1 节约成本

云存储技术的应用可以为图书馆数字化过程中从实施到后期的维护节约成本。目前的存储技术需要图书馆必须购买昂贵的软件和硬件,才能将数字资源放在服务器上。此外,图书馆还要聘请专业的 IT 人员,进行硬件和软件的维护工作,还要更新这些设备和软件。云存储技术在很大程度上解决了这些问题。

3.2.2 更好地数据备份并可以异地处理日常数据

云存储中硬盘坏掉,数据会自动迁移到别的硬盘,大大减轻了管理人员的工作负担。对云存储来说,再多的存储服务器,在管理人员眼中也只是一台存储器,每台存储服务器的使用状况,通过一个统一管理界面监控,使得维护变得简单和易操作。

3.2.3 更多地访问和数据共享

图书馆的借阅者不再需要通过本地网络来访问图书馆相关的数字资源,云存储可以让借阅者在任何地方、任何时间访问他们需要的数据。图书馆也不需要花费上千万美元来打造最新技术和最新应用来创造最好的系统,这就为图书馆数字资源最大程度的数据共享提供了可能。

以上特点为云存储技术在图书馆数字化中的应用

铺平了道路。除了以上特点,云存储还具有超强的计算能力、永远以客户为核心、高度的安全性、较优的性价比和节能环保等优势^[6]。

4 云存储技术在图书馆管理系统中的应用模式

数字图书馆的云存储系统是基于高性能网络传输技术,通过集群技术、分布式文件系统和网络计算技术等,根据客户服务需求将计算资源与已存储的数据资源按需分配。图书馆的数据资源云存储系统是一个由高性能网络传输设备、超大容量数据存储设备、超高性能服务器、特色服务应用软件和形式多样客户端等多个部分组成的复杂系统。通过虚拟化技术对云图书馆存储和数据操作进行管理,系统形成了以存储设备为核心,通过软件为客户提供数据存储和业务访问功能。数字图书馆云存储系统的结构模型如图 2 所示。

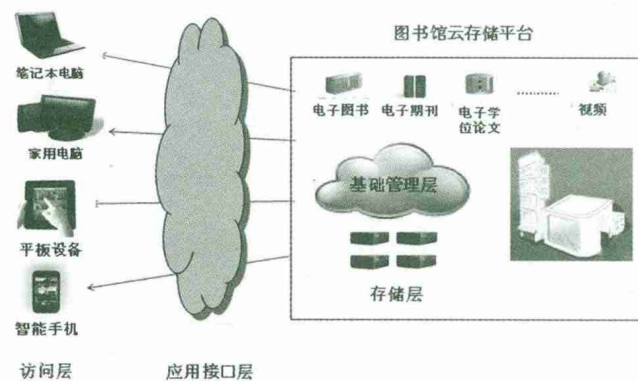


图 2 数字图书馆云存储系统结构模型

4.1 存储层

存储层是数字图书馆云存储系统的基础。图书馆将所有数字资源通过存储设备统一存放于存储层中,分布于世界不同地域的存储设备再通过各种网络设备连接在一起,由统一的存储设备管理系统管理,并利用虚拟化技术忽略掉硬件之间的差异,对外提供服务,实现海量数据的统一管理。存储设备包括:光纤通道存储设备、IP 存储设备和 DAS 存储设备等。存储层的功能结构如图 3 所示。

表 2 云存储与传统存储对比

比较项	云存储	传统存储
架构	不仅是一种架构,更是一种服务。底层采用分布式架构和虚拟化技术,易于扩展,单点失效不影响整体服务	针对某种特殊应用而采用的专用、特定的硬件组件构成的架构
服务模式	按需使用,按使用计费,服务提供商可迅速交付和响应	用户通过整机购买或租赁获取存储容量
容量	支持 PB 级以上无限扩展	针对某个特定的应用存储,由使用需求方决定容量,难于扩展
数据管理	不仅提供传统访问方式,而且提供海量数据的管理和对外的公众服务支撑,同时采用保护数据安全的策略,用户可灵活配置	用户数据管理员可见,信息不够安全。通常使用 RAID 提供数据保护,用户无法灵活配置个性化存储策略和保护策略



图 3 存储层的功能结构图

4.2 基础管理层

基础管理层是云存储的核心部分，也是最难以实现的部分。该层通过集群技术、分布式文件系统和网格计算技术等相关技术，实现系统中多个存储设备间的协同工作，并对外提供统一的服务标准，为存储系统提供更高效、更强大、更优越的数据访问性能打下基础。此层还具有内容分发系统、数据加密技术、数据备份和容灾技术和措施等，保证了存储系统中数据的安全性。基础管理层的功能结构如图 4 所示。

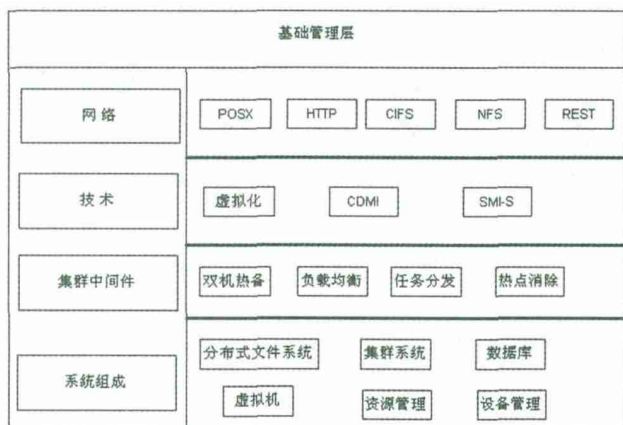


图 4 基础管理层的功能结构图

4.3 应用接口层

应用接口层是云存储系统中最灵活多变的部分。不同的云存储运营单位可以根据实际业务类型，开发不同的应用服务接口，为不同的图书馆需求提供不同的应用服务。该层满足了云图书馆基于网络跨平台的二次应用程序开发的要求，为图书馆建立特色平台，如：视频监控应用平台、IPTV 和视频点播应用平台、网络硬盘引用平台和远程数据备份应用平台等，提供了技术支持。应用接口层的功能结构如图 5 所示。

4.4 访问层

任何一个授权用户都可以通过标准的公用应用接口来登录数字图书馆云存储系统，享受相关服务。此

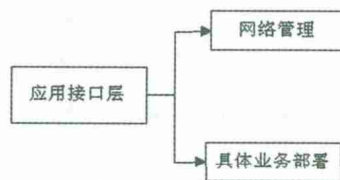


图 5 应用接口层的功能结构图

时的数字图书馆馆员以及数字读者不再需要了解底层的硬件设备类型，也不需知道数字资源存放的物理位置，只需要在获得授权后，直接调用标准的 API，享受相关服务。此服务的使用不受读者所使用的终端设备限制。访问层的功能结构如图 6 所示。

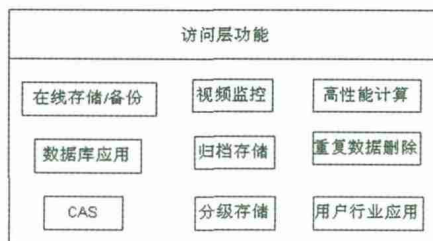


图 6 访问层的功能结构图

5 结束语

云存储技术为构建现代图书馆存储系统提供了新的思维方式。构建数字图书馆云存储平台，可以解决许多长期困扰图书馆数字资源存储管理和数据共享中存在的问题，通过云存储可提升图书馆各类数据资源的利用率，提高图书馆的服务水平。云存储作为一种新技术，在实际的应用过程中可能会面临很多问题，在以后的研究过程，将以数字读者需求为基本点，运用成熟的技术，集合多方面的力量来完善数字图书馆云存储系统。

参考文献：

- [1] 吴国华.云存储技术在数字图书馆建设中的应用探讨[J].龙岩学院学报,2011,(2):130-133.
- [2] 张继平.云存储解析[M].北京:人民邮电出版社,2013.
- [3] 龙斌.网络存储技术在数字图书馆系统中的应用分析[J].农业图书情报学刊,2010,(9):54-55,62.
- [4] 常青.基于网络存储架构的数字图书馆建设研究[J].农业图书情报学刊,2008,(11):44-47.
- [5] 陈臣.数字图书馆云存储系统安全架构与安全策略研究[J].现代情报,2011,(9):160-164.
- [6] 徐超.大数据时代下云存储技术在消防信息化系统中应用[J].信息通信,2015,(3):118.

数字图书馆云存储系统的设计与实现^{*}

席亚军

(塔里木大学图书馆 新疆 阿拉尔 843300)

关键词: 数字图书馆; 云存储系统; 分布式系统

摘要: 文章以图书馆数字资源存储为研究对象, 设计了系统主要功能实现的流程, 探讨了关键技术, 综合采用了网络技术、云存储技术、Hadoop 分布式文件系统(HDFS)、HBase 技术、Lucene 倒排索引技术等, 最终以 JSP 为开发平台, 构建了一个以实现数字资源长期有效存储为目的的数字图书馆云存储系统。

中图分类号: G250.76 文献标识码: A 文章编号: 1003-1588(2015)10-0121-03

自从 Google 发布云计算的概念以来, 很多公司以及组织对其进行了研究, 同时也产生了一大批产品, 包括 Google 的大数据表、MapReduce、谷歌文件系统、ChubbyLock、Windows Azure、雅虎的集群计算系统, 以及亚马孙的 EC2^[1]。其中 Apache 基金会的 Hadoop^[2] 使用最为广泛。使用 Hadoop 用户可以在完全透明的情况下使用集群, 进行读写和计算。做一个免费并且开源的计算框架, 国内外使用 Hadoop 的公司都比较多。雅虎^[3] 拥有 25 000 个节点, 主要用于支持网页搜索和广告系统, 成为全球最大的 Hadoop 集群。国内百度、淘宝、网易、华为、中国移动等都是 Hadoop 集群的使用用户, 其中淘宝拥有 1 700 个服务器节点, 属于较大的 Hadoop 集群。

塔里木大学图书馆作为大量数字资源的拥有者, 为了改变现有存储系统的缺陷, 以海量数据资源的存储为研究对象, 以 JSP 为开发平台, 利用 HBase、Lucene、Katta 等多种技术, 搭建 Hadoop 集群, 开发数字图书馆云存储系统, 实现对数字资源的存储形式多样化、安全有效、具有动态可扩展性以及高度共享性的存储。

1 系统设计

1.1 图书馆数字资源存储需求分析

随着数字资源以爆炸性增长, 图书馆在不断扩大数字资源存储容量的同时, 也在考虑如何节约有限的资金投入、高度数据共享、可靠性和安全性等方面的问题^[4]。

1.1.1 经济适用是图书馆建设存储系统时需要考

虑的重要因素。作为资金有限的公用事业单位——图书馆, 需要根据自身发展和资金现状选择性价比较高的存储方案, 切不可盲目追求投入昂贵、技术先进, 却暂时无法应用的设备。

1.1.2 数据高度共享是图书馆存储系统建设的重要条件。图书馆资源数字化是图书馆发展的必然趋势, 他有利于提供图书馆资源的使用效率, 提高数据的共享性, 也有利于资源的保护和利用。

1.1.3 安全可靠是数字图书馆服务的保证。为了避免不可抗拒和人为因素对存储系统的损害, 要求图书馆在建设存储系统时, 必须要有安全的存储备份方案。

1.2 系统架构设计

数字图书馆云存储系统以图书馆海量数字资源存储和共享为研究对象, 以图书馆——读者为基本模式, 从图书馆馆员、管理员和读者的不同角度设计该系统。系统包括数字资源存储模块、数据管理模块和索引检索模块三部分, 其系统模块图如图 1。从图 1 可以看出, 图书馆员将纸质版图书、期刊、论文、报纸等纸质资源, 以及光盘等数字资源, 通过传统手段录入到计算机中, 形成数字资源。针对电子数据文件大小异构的特点, 图书馆员通过数据存储模型, 将视频、音频等大数据文件利用 HDFS 分布式文件系统存放; 将图片、文档等较小的文件利用 HBase 直接存放。管理员利用数据管理模块, 对数据文件进行上传、下载和删除等。读者可以在 PC 机、手机、平板电脑等不同平台, 利用系统提供的各种查询方式进行资源检索。

收稿日期: 2015-09-04

作者简介: 席亚军(1981—) 塔里木大学图书馆馆员。

* 本文系 CALIS 农学中心研究项目“大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究”的阶段性成果之一, 项目编号: 2015032。

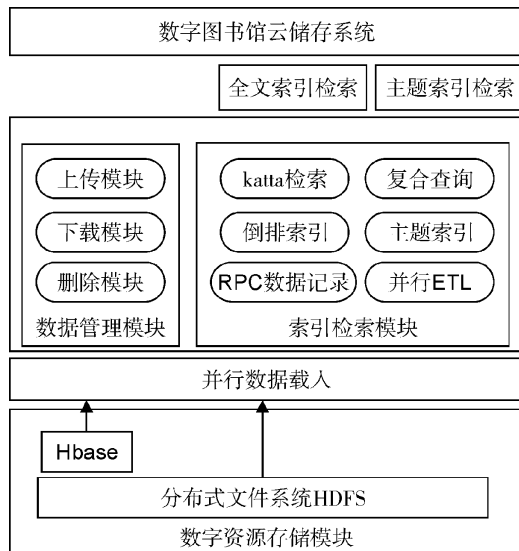


图1 系统模块图

1.3 系统部署

本系统硬件节点部署示意图如图2所示,系统分为三层:数据存储层、应用服务层和用户层。数据存储层由若干不同类型的存储节点或者服务节点组成,用于存储图书馆数字资源^[5]。应用服务层主要由具有不同功能的服务器组成,其中一部分服务器用于运行NameNode、SecondNameNode、Hmaster和JobTracker服务;一部分用于单独运行ZooKeeper服务;一部分用于运行Katta分布式索引检索服务;一部分用于运行DataNode、TaskTracker和HregionServer服务^[6]。用户层中的读者可以通过不同形式的平台使用该存储系统。

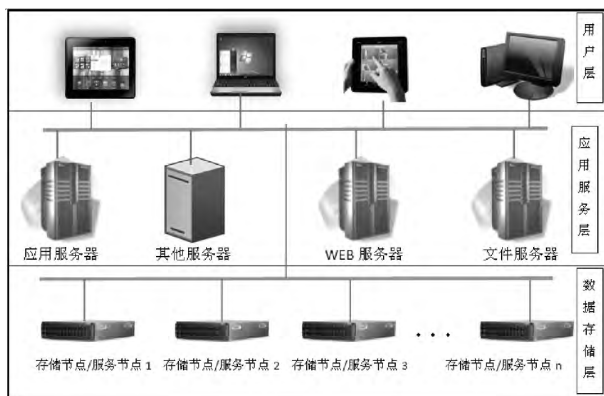


图2 系统部署示意图

2 数据处理流程设计

数字图书馆云存储系统的数据流程示意图如图3所示。系统对外提供图书馆员使用的管理接口和普通读者使用的查询接口。图书馆员借助于图形管理界面,实现数字资源的管理(上传、下载、删除等)操作。数据管理过程中,数据信息通过网络平台在当地计算机和存储集群之间的传递。数字资源按照类别,分别存储于HBase或直接存放于HDFS。读者

获得授权后,可以浏览当前数字资源,通过系统利用HBase自身的RowKey索引,进行精确查询、模糊查询和复合查询。

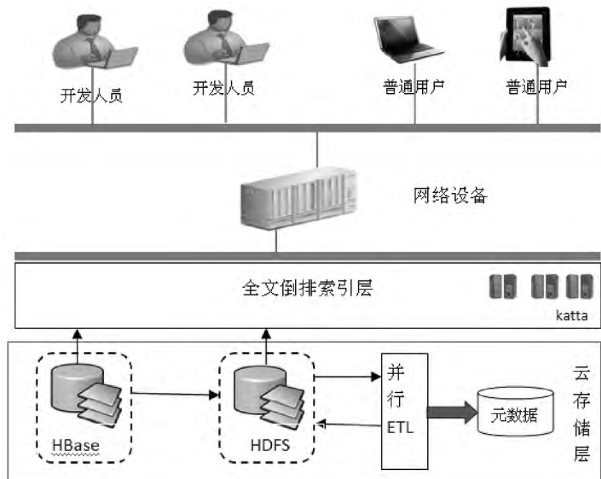


图3 数据流程示意图

2.1 文件上传

数字资源的上传主要是对HDFS分布式系统和HBase数据库的操作。数字图书馆云存储系统根据数据文件大小进行了分类存储,对不同数据文件的上传进行了不同设计。图书馆员上传文件之前,先调用HBase数据库,获取TableName表的Scanner,检查该文件是否存在。如果文件已经存在,则提示文件已经存在,是否需要更新。如果文件没有存在,则按照数据文件大小进行分类存储。

2.1.1 较小文件上传。较小的数据文件包括文本、图片和网络资源,直接存储在HBase数据库中。文本文件要进行RPC通信,建立倒排索引,监控更新消息,记录对文本表的更新操作。

2.1.2 较大文件上传。较大的数据文件包括音频、视频、高清扫描图片等,需要借助分布式文件系统HDFS存储,同时在HBase数据库中有相应的地址映射表。在地址映射表MyData中查重,若存在则更新,否则添加映射关系。

2.2 文件下载

文件的下载模块也是根据文件大小采用不同的方式进行下载的。对于相对较小的数据文件直接获取数据库HBase数据表Infor中的content字段中的内容,而对于相对较大的数据文件需要获取Infor中的content字段中的值后,从分布式系统HDFS中下载文件内容。

2.3 索引模块

针对文本文件内容的检索相对比较复杂,本系统通过分布式检索工具Katta,利用分布式倒排索引Lucene,将索引进行分片存储加以实现,检索流程图如图4所示。客户端首先向检索服务器群组中的分布式查询服务器发送查询请求,由分布式查询服务器根据检索服务器集群中设备的当前状态将查询任务发送,然



后检索服务器将检索结果返回给分布式查询服务器, 用户最终得到查询结果, 完成查询任务。

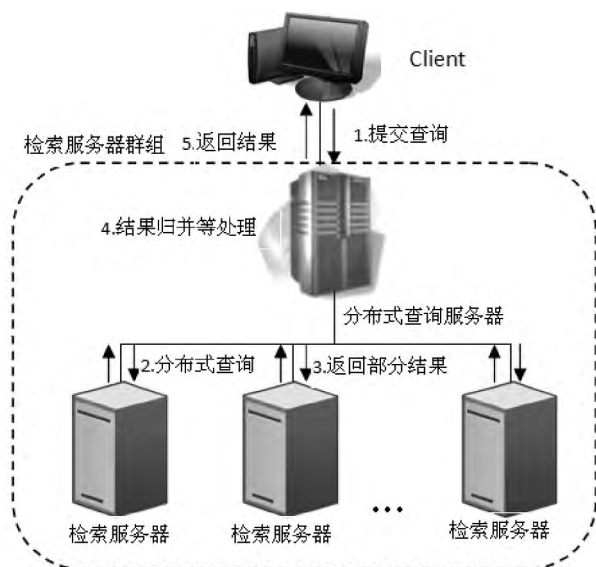


图4 基于 Katta 的检索流程图

3 主要关键技术

3.1 Lucene 索引结构

传统的索引方式是以文档的 ID 为关键字, 表中记录文档中每个字的位置信息, 在查找时, 扫描表中每个文档中字的信息直到找出所有包含查询关键字的文档。这种组织方法在建立索引的时候结构比较简单, 建立比较方便且易于维护。但是在查询的时候需对所有的文档进行扫描以确保没有遗漏, 这样就使得检索时间大大延长, 检索效率低下。倒排索引以字或词为关键字进行索引, 表中关键字所对应的记录表项记录了出现这个字或词的所有文档, 一个表项就是一个字表段, 他记录该文档的 ID 和字符在该文档中出现的位置情况。由于每个字或词对应的文档数量在动态变化, 所以倒排表的建立和维护都较为复杂, 但在查询的时候由于可以一次得到查询关键字所对应的所有文档, 所以效率高。在全文检索中, 能够达到检索的快速响应。文档文件不同于其他文件, 需要对其全文的内容进行检索。为了提高效率, 数字图书馆云存储系统对文档文件的检索采用倒排索引。

3.2 行列式存储模式

Facebook 曾在 2010 ICDE (IEEE International Conference on Data Engineering) 会议上介绍了数据仓库 Hive^[7]。Hive 存储海量数据在 Hadoop 系统中, 提供了一套类数据库的数据存储和处理机制。他采用类 SQL 语言对数据进行自动化管理和处理, 经过语句解析和转换, 最终生成基于 Hadoop 的 MapReduce 任务, 通过执行这些任务完成数据处理^[8]。数字图书馆云存储系统中较大文件的上传存储便利用了数据仓库 Hive, 即借助分布式文件系统 HDFS 存

储, 同时在 HBase 数据库中有相应的地址映射表。在地址映射表 MyData 中查重, 若存在则更新, 否则添加映射关系。

4 系统实现

数字图书馆云存储系统考虑到多用户性和可扩展性, 系统采用 C/S 的体系结构; 云存储系统为了给公众提供一个查询的平台, 采用网站形式。整个系统以 JSP 为开发平台, 采用面向对象的开发思想。在编写系统公共类的基础上实现相应功能。

数字图书馆云存储系统采用 RedHat Enterprise Linux5 操作系统, 使用 13 台不同型号服务器组建 Hadoop 集群, 使用 Hadoop, HBase, Zookeeper, Lucene, Katta 等多种开发工具。该系统已在新疆塔里木大学图书馆应用, 满足了图书馆数字资源存储形式多样化、总量迅猛增长、具有动态可扩展性, 以及高度共享性的要求, 取得了良好的经济效益和社会效益。

5 结论

笔者利用云存储平台的分布式存储机制, 设计云存储环境下的数字图书馆云存储系统。该系统主要是针对图书馆数字资源存储过程中产生海量数据进行存储和共享, 并针对数字资源存储的特点, 进行相应的操作处理。

参考文献:

- [1] Amazon E. C. Amazon elastic compute cloud (Amazon EC2) [J]. Amazon Elastic Compute Cloud(Amazon EC2), 2010(5): 18-23.
- [2] Fay Chang, Jeffrey Dean. Gruber Bigtable: A Distributed Storage for Structured Data [C]. OSDI 2006: 47-50.
- [3] Shvachko K, Kuang H, Radia S. The hadoop distributed file system [C]. Mass Storage System and Technologies (MSST), 2010 IEEE 26th Symposium on IEEE, 2010: 1-10.
- [4] 李素梅. 基于云计算的高校图书馆信息资源建设 [J]. 河南图书馆学刊, 2011(3): 151-153.
- [5] 陈臣. 数字图书馆云存储系统安全架构与安全策略研究 [J]. 现代情报, 2011(9): 44-47.
- [6] 刘鹏. 中国云存储发展报告 [M]. 北京: 电子工业出版社, 2013: 100-126.
- [7] 杨信廷. 蔬菜安全生产管理及质量追溯系统设计与实现 [J]. 农业工程学报, 2008(3): 162-166.
- [8] 吴海华. 基于云计算数字图书馆的应用实践研究 [J]. 河南图书馆学刊, 2013(4): 74-75.

(编校: 马怀云)

高校图书馆数字资源建设现状及发展趋势研究

席亚军

(塔里木大学 新疆 阿拉尔 843300)

关键词: 数字资源建设; 高校图书馆; 发展趋势; 共建共享

摘要: 文章在论述高校数字资源概念和优势的基础上, 阐述了高校数字资源建设现状, 分析了高校数字资源建设过程中存在的问题, 最终探讨了高校数字资源建设发展趋势。

中图分类号: G251

文献标识码: A

文章编号: 1003 - 1588(2016)02 - 0041 - 03

数字资源具有形式多样、易保存、检索便捷等特点。高校图书馆依据数字资源的特点, 将其作为重点发展方向, 而数字化技术、网络技术和计算机技术的发展为其提供了技术保证。数字资源的建设改变了传统图书馆工作效率低下、形态单一以及社会地位低的现状, 提高了高校图书馆整体的服务水平, 为图书馆的高速发展创造了新的机遇。目前, 高校图书馆数字资源建设过程中存在的一些问题已经阻碍了高校图书馆的进一步发展。因此, 寻找高校图书馆数字资源建设发展的趋势, 成为当前高校图书馆亟须解决的首要问题。

1 图书馆数字资源的概念和优势

数字资源是指把文字、声音、图像、动画等形式多样的电子数据存储在非纸质载体中, 通过计算机技术、通信技术及多媒体技术进行发布、存取和利用的信息资源。数字资源主要包括数据库、电子期刊、电子图书、网页和多媒体资料等^[1]。图书馆根据自身发展需求, 通过购买电子资源数据库, 建设特色数据资源、学科导航库, 建设光盘资源管理系统以及采集网络数据资源等手段, 形成高校各自的图书馆数据资源系统。

数字资源可以利用通信网络进行传输。因此, 数据资源具有查询检索方便、复制传播迅速的特点, 同时其安全性面临考验。数据资源的传输还要依赖于通信网络软硬件设备的匹配和兼容性, 使其传输变得复杂。综上所述, 数据资源具有共享性、依赖性、复杂性、动态多变性、跨时空性、可检索性及可分离性等特点^[2]。

2 高校图书馆数字资源建设现状

为了较全面客观地反映我国高校图书馆数字资源建设现状, 笔者从全国 2 246 所高校中随机选取 20 所, 其中重点建设大学 10 所, 普通二本院校 7 所, 普通高职院校 3 所, 对其数字资源的数量、种类、类型等方面进行系统全面的调查, 并对调查数据进

行统计分析。调查方法主要采用从这些高校数字图书馆的信息门户网站入手, 辅以电话和走访等方式。调查时间自 2015 年 5 月至 2015 年 8 月。调查的数据资源类型为所购数据库、自建特色数据库、随书光盘管理系统等。

2.1 高校图书馆拥有数据库总量情况

通过调查 20 所大学图书馆的数据库统计情况见表 1。这些数据库包括购进的中外文数据库、自建特色数据库等各类数据库总量。将表 1 中的数据进行分类汇总, 数据库数量分布情况见图 1。

表 1 高校图书馆拥有的数据库总量

序号	单位名称	数量	序号	单位名称	数量
1	清华大学	533	11	兰州大学	91
2	北京大学	294	12	北京信息科技大学	78
3	复旦大学	261	13	塔里木大学	49
4	浙江大学	251	14	华北电力大学	46
5	四川大学	235	15	河北工程大学	40
6	武汉大学	177	16	太原师范学院	26
7	昆明理工大学	157	17	西藏农牧学院	18
8	东北大学	135	18	天津职业大学	14
9	天津大学	126	19	承德高等专科学校	13
10	上海师范大学	98	20	辽宁交通高等专科学校	5

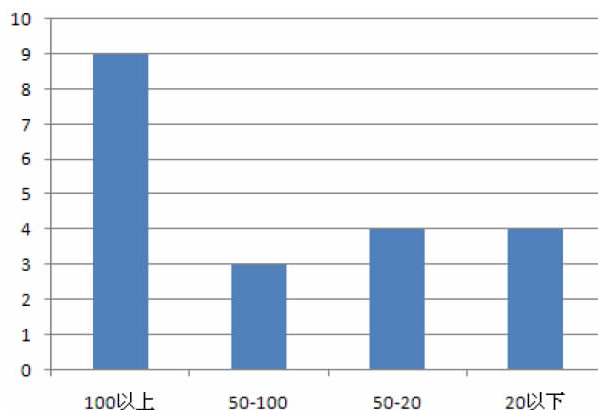


图 1 数据库数量分布图

收稿日期: 2016 - 01 - 09

作者简介: 席亚军(1981—) 塔里木大学馆员。

如图 1 所示,本调查随机抽选的 20 所高校中有 9 所高校拥有 100 个以上数字资源数据库,3 所拥有 50 个以上资源数据库,4 所拥有 20 个以上数据库,另有 4 所还未达到 20 个。其中重点院校占有绝对优势,拥有 100 个以上数据库,绝大多数普通二本院校拥有 20 个以上数据库,而高职高专院校由于受资金、人才、设备、管理和思想理念等多方面的制约,数据库建设滞后。

2.2 购买中文数据库情况

依据实际调查结果如表 2 所示,表明 20 所高校图书馆都不同程度地引进了中文数据库,但在数量上还存在很大的差距。

从表 2 可知,清华大学所购买的中文数据库数量具有绝对优势,而第二名复旦大学仅是清华大学的一半。清华大学所拥有的数量是最后一名辽宁交通高等专科学校所拥有数量的近 40 倍。1 所高校所购买的中文数据库超过 100 个,8 所高校在 50 个到 100 个之间,5 所高校在 20 个到 50 个之间,6 所高校未达到 20 个。

表 2 高校图书馆购买中文数据库数量汇总表

序号	单位名称	数量	序号	单位名称	数量
1	清华大学	195	11	兰州大学	55
2	复旦大学	86	12	塔里木大学	22
3	北京大学	60	13	华北电力大学	6
4	昆明理工大学	34	14	河北工程大学	12
5	武汉大学	92	15	北京信息科技大学	72
6	四川大学	63	16	西藏农牧学院	49
7	浙江大学	12	17	太原师范学院	29
8	天津大学	56	18	承德高等专科学校	26
9	上海师范大学	18	19	天津职业大学	17
10	东北大学	46	20	辽宁交通高等专科学校	4

根据中文数据库被引进的情况,对各高校图书馆所购买的中文数据库进行分类统计,其中购买次数达到 10 次以上的中文数据库如表 3 所示。

表 3 购买次数达 10 次的中文数据库

序号	数据库名称	所购买图书馆数量
1	读秀知识库	19
2	超星数字图书馆	18
3	万方数据资源系统	18
4	中国知网(CNKI)	16
5	维普数据资源系统	15
6	新东方多媒体学习库	12

读秀知识库集文献检索、试读、传递为一体,拥有强大的资源优势、文献的深度检索、全文显示、统一快捷的整合服务和高效快捷的文献传递,被绝大多数的高校图书馆所购买。各高校均引进了综合类中文科技期刊类数据库,例如万方和中国知网,但专业数据库的引入较少。

2.3 购买外文数据库情况

实际调查结果如表 4 所示,除了承德高等专科学校未购买外文数据库,其余 19 所高校图书馆都不同程度地引进了外文数据库,但在数量上存在很大的差距。

表 4 高校图书馆购买外文数据库数量汇总表

序号	单位名称	数量	序号	单位名称	数量
1	清华大学	327	11	上海师范大学	49
2	北京大学	197	12	兰州大学	48
3	浙江大学	191	13	塔里木大学	20
4	四川大学	164	14	华北电力大学	20
5	复旦大学	162	15	河北工程大学	16
6	武汉大学	107	16	太原师范学院	7
7	东北大学	86	17	天津职业大学	6
8	昆明理工大学	85	18	西藏农牧学院	1
9	天津大学	69	19	辽宁交通高等专科学校	1
10	北京信息科技大学	56	20	承德高等专科学校	0

由表 4 可知,各高校图书馆在购买外文数据库上的态度差异较大,自清华大学至上海师范大学,都在大力购买外文数据库,其购买外文数据库的数量超过所购买中文数据库的数量。自兰州大学至承德高等专科学校,其外文数据库的购买量少于中文数据库。

根据外文数据库被引进的情况,统计汇总被高校图书馆购买次数达 10 次以上的外文数据库如表 5 所示。

表 5 购买次数达 10 次的外文数据库

序号	数据库名称	所购买图书馆数量
1	Springer Link	15
2	ProQuest 检索平台	13
3	EBSCO Host	12
4	Association for Computing Machinery (ACM)	12
5	American Chemical Society(ACS)	11

由表 5 可知,有 5 个外文数据库被 15 所高校同时购买,被购买次数最多的外文数据库为 Springer Link。各高校所购买的外文数据库中除了综合性的科技期刊数据库,还有专业数据库。数据库种类分布的专业也较为广泛。

2.4 自建特色数据库情况

20 所高校中,有 14 所高校根据自身的特点建立了特色数据库,还有 6 所高校未建立。

表 6 高校图书馆自建特色数据库数量汇总表

序号	单位名称	数量	序号	单位名称	数量
1	清华大学	11	11	天津大学	2
2	北京大学	11	12	河北工程大学	2
3	四川大学	11	13	天津职业大学	2
4	兰州大学	9	14	承德高等专科学校	1
5	复旦大学	7	15	昆明理工大学	0
6	武汉大学	7	16	上海师范大学	0
7	太原师范学院	7	17	塔里木大学	0
8	浙江大学	4	18	华北电力大学	0
9	北京信息科技大学	4	19	西藏农牧学院	0
10	东北大学	3	20	辽宁交通高等专科学校	0

由表 6 可知,清华大学、北京大学和四川大学分别建立了 11 个特色数据库。其中北京大学根据自己的优势建立了北大机构知识库、北大博文、北大讲座、北大名师、北京历史地理、燕大论文、古文献资源库和民国旧报刊,另外还以北大名人所捐赠的书刊建立了 3 个图书馆。

3 高校图书馆数字资源建设过程中存在的问题

3.1 高校图书馆数字资源建设发展不均衡

依据 20 所高校图书馆数据资源库调查的结果显示,清华大学拥有各类数据库 533 个,而辽宁交通高等专科学校仅拥有 5 个数据库。绝大部分重点高校所拥有的数据库的数量都多于普通二本院校,而大部分普通二本院校拥有数据库的数量多于高职高专院校。数字资源建设需要充裕的资金支持、高素质的技术管理人才和先进的数字图书馆建设理念等必要条件,高职高专院校由于受资金、人才、设备、管理等多方面的因素制约,数字资源建设长期处于较低水平。

3.2 数据库的重复建设问题突出

一方面,高校图书馆建设过程中为了能够涵盖更多的数据资源,会花费更多的资金购买更多的数据库资源。数据库服务商为了能够提供更多的服务,会将自身数据库的数据资源数量涵盖更多的内容。因此,各高校所购买的数据库存在重复建设的问题。对比两个目前高校购买较多的数据库:中国知网(CNKI)和万方数据资源系统^[3],中国知网和万方数据资源系统均提供学术期刊、学位论文、会议论文、专利文献、科技成果和中外标准等服务,所提供的数据有一部分内容完全一致。另一方面,调查结果显示各高校图书馆所购买的数据库有部分数据库是相同的,20 所高校中,读秀知识库就被 19 所高校购买。

3.3 数字资源需要整合

根据调查结果,每所高校图书馆至少拥有 5 个以上数据库。不同的数据库来自不同的供应商,彼此之间没有任何联系。读者在检索数据资源时,通常在一个数据库内输入关键字进行检索,然后再打开另外的数据库重新输入关键词检索。对于拥有几百个数据库的高校图书馆,读者往往要不厌其烦地不断重复操作,这样不但耗费大量的时间和精力,且效率低下,也给读者带来了重复操作的烦恼。数据整合是由简单的链接整合、导航整合向跨数据库检索不断发展,将在一定程度上提高检索效率,节约时间和精力。

3.4 标准不统一,操作复杂

在我国数据资源建设过程中,国家或者行业协会尚未对数据资源建设制定标准和规范。各数据库服务商根据自己的建设要求,各自为政,各行其是,自由发展。读者在使用过程中发现系统多种多样、操作困难、数据兼容性和互操作性差以及原始数据处理不一等问题,这些问题限制了数据库的共享,影响了其作用的发挥。

3.5 特色数据库建设缺少方向

目前,特色数据库建设中低水平重复建设、缺乏特色、更新缓慢等现象比较普遍。调查结果显示,部分高校图书馆将学生学位论文和本校教师发表的学术论文作为特色数据库,20 所高校中有 6 所高校未建立特色数据库。作为高校图书馆,应根据各自馆藏特色划定范围,找准各自定位,做到有所为有所不为,这样才能建设出既符合高校的办学特点,又具有可持续发展潜力的特色数据库。

4 高校图书馆数字资源建设发展趋势

对于投入资金有限的高校图书馆数字资源建设来说,未来高校图书馆将遵循资源建设标准化、规范化、坚持现实可能与可持续发展、资源建设共建共享以及特色化的原则,彻底改变目前建设中存在的问题。

4.1 高校数字资源共建,达到资源共享

高校图书馆所购买数据库的资金主要来源于国家教育部所划拨的教育专项资金,各高校用有限的资金重复购买相同的数据库,造成了资源的浪费。实现高校图书馆之间的数字资源共享是高校图书馆发展的必然趋势。各高校图书馆应在管理体制和资源配置方式上进行相应改革,变单独建设为群体共建,变独立运营为全局开放,避免重复建设,浪费资源。资金雄厚、数字资源建设较完善的高校更应该转变观念,积极加入数字资源共建的群体中来。通过高校之间通力协作,制定统一标准和规划,最终完成高校数字资源共建共享的目标。

4.2 建设特色数据库,突出各自特色

特色数据库是指图书馆在充分利用自己的馆藏特色或本区域特色文化资源基础上建立起来的一种具有本馆特色的可供共享的文献信息资源库^[4]。目前,各高校图书馆应充分利用计算机技术、通信技术和网络技术,将在长期社会分工实践中积累和形成的、具有本馆特色的馆藏资源数字化,将本区域特色文化资源数字化,为科学技术和经济建设发展战略信息需求提供决策依据。特色数据库建设是高校图书馆馆藏资源得以生存和发展的基础,也是图书馆数字化的核心内容。

4.3 数字资源深度整合,提高检索效率

数字资源整合是一种数字资源优化组合的存在状态,是依据一定的需求,对各个相对独立的资源系统中的数据对象、功能结构以及互动关系的融合、聚类和重组,重新结合为一个新的有机整体^[5]。数字资源能否被高效利用和吸收取决于其整合程度。数据资源整合方式有基于传统书目管理的 OPAC 系统、基于跨库检索的资源整合、基于资源导航的资源整合以及基于超级链接的资源整合等多种模式。

参考文献:

- [1] 黄海清. 高校图书馆数字资源建设的现状及对策初探[J]. 中国管理信息化, 2015(15): 199-201.
- [2] 丛全滋,王学军,耿晓宁. 图书馆数字资源建设的问题与对策[J]. 图书馆学刊, 2014(11): 44-48.
- [3] 孔毅. 万方数据资源系统与中国知网数据库对比分析[J]. 情报理论与实践, 2010(2): 342-347.
- [4] 秦波,邹丽红. 浅谈地方高校图书馆数字资源建设的构想[J]. 现代情报, 2008(3): 99-101.
- [5] 周建清. 试论图书馆数字资源整合[J]. 现代情报, 2007(3): 88-89, 92.

(编校: 崔 萌)

基于 RFID 技术的图书馆编目系统建设探索

席亚军

(塔里木大学图书馆, 新疆 阿拉尔 843300)

摘要: 在应用了 RFID 技术之后, 图书馆的管理系统融合了新的 RFID 系统, 传统的图书流通和管理业务流程等技术体系也产生了极大的变革。本文主要是针对基于 RFID 技术的图书馆编目系统的探索。

关键词: RFID; 编目系统; 图书馆

中图分类号: G250

文献标识码: A

文章编码: 1672-6251(2016)02-0074-02

Exploration on the Construction of Library Cataloging System Based on RFID Technology

XI Yajun

(Library of Tarim University, Xinjiang Alaer 843300)

Abstract: With the application of RFID technology, library management system has integrated the new RFID system architecture, and the technology system of traditional library circulation and management has also changed greatly. This paper focused on the exploration of library cataloging system construction based on RFID technology.

Key words: RFID; cataloging system; library

进入 21 世纪, 方便快捷的信息传播是信息时代的特征之一。图书管理是 RFID 技术应用的一个重要方面, 图书馆是图书管理需求最为集中的应用场所。自 1999 年开始, 图书馆行业开始逐渐应用 RFID 技术来管理图书和读者。到 2006 年, 国内也开始使用 RFID 技术设计图书馆功能。RFID 技术使得图书借还更贴近现代化的技术, 方便读者对图书的利用, 适合现在快节奏的生活, 更方便了图书馆员对图书的管理, 减少了很多从采购到图书借还到图书盘点的繁琐工作, 释放了劳动力, 提高了效率^[1]。

1 RFID 概述

RFID (Radio Frequency Identification) 称为无线射频识别技术, 国际图联大会上信息技术组提出 RFID 技术与图书馆的关系后, RFID 技术在图书馆应用至今已有很长的历史。1998 年, 新加坡 Bukit Batok 社区图书馆首先采用了 RFID 系统, 之后, 世界各国和地区的图书馆开始大规模实施图书馆 RFID 系统。据 Checkpoint 统计, 2005 年全球约有 440 家图书馆采用了 RFID 技术, 2007 年为 2 000 家^[2], 并且以每年 30% 的速度增加^[3]。在应用数量上, 美国处于世界领先

地位, 其次是日本和英国, 此外澳大利亚、新加坡、荷兰、墨西哥等也有相当大数量的应用案例。

在我国, 图书馆 RFID 技术的应用比国外晚。2007 年, 集美大学首家采用 RFID 图书管理系统; 7 月, 深圳图书馆实施了 RFID 图书管理系统。随后汕头大学图书馆、上海长宁区图书馆、国家图书馆、陕西省图书馆等也相继实施, 他们的成功实践为 RFID 的应用产生了良好的示范效应。据统计, 截止到 2015 年底, 国内应用 RFID 系统或已招标的图书馆数量已超过百余家。以新疆为例, 应用 RFID 系统的有石河子大学图书馆、新疆大学图书馆。从国家范围来看, 全国图书馆总数约 1 万家, 但 RFID 应用比例还很小, 未来有很大的发展空间。

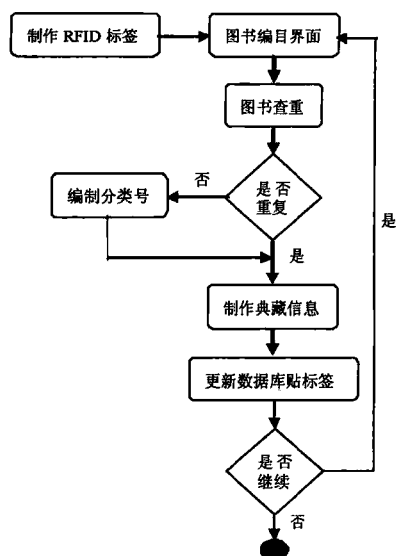
2 基于 RFID 技术的编目系统结构及编目流程

基于 RFID 技术的图书馆编目系统涉及多个相互关联的子系统, 各个子系统之间相互协调完成数据的交换与认证。编目的工作主要是完成图书信息制作。图书编目主要工作是 RFID 标签制作和分类号编制。图书编目流程如附图所示。

基金项目: 2015 年 CALIS 农学中心研究项目“大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究”(编号: 2015032)。

作者简介: 席亚军(1981-), 男, 本科, 馆员, 研究方向: 数字图书馆。

收稿日期: 2016-02-24



附图 编目流程

2.1 图书编目流程

进行图书编目时,使用书名或ISBN号检索到图书记录,使用手持设备激活RFID标签,形成图书RFID编号作为识别码,系统自动生成财产号,再编制分类号,最终生成完整的图书典藏信息。

2.2 RFID标签制作流程

操作员激活图书中RFID标签。激活RFID标签之前人工要粘贴RFID标签。若系统扫描不到RFID标签信息,则剔出图书进入下一本图书制作环节;也可以人工粘贴RFID标签后,重新扫描图书中的RFID标签信息。

2.3 分类号编制流程

分类号编制的过程是操作员制作图书分类号并打印书标的过程。如果在图书馆藏数据中检索到数据,则直接生成图书财产号;否则将由人工根据《中国图书分类法》编制新的分类号,再生成图书财产号。一本书需要至少一个书标提供给读者查找图书时使用,图标可以当时打印,也可以做好一批图书后统一打印。图书制作完成后,还需人工将书标粘贴到书脊上。

3 RFID在图书馆应用中的优势

3.1 简化借还书流程,提高流通效率

条码的借还书流程仍然需要人工打开图书扉页并找到条码位置后才能扫描条码。这样的操作流程繁琐、效率很低。同时,由于条码容易磨损或脱落,更影响借还书的效率,同时也会影响读者对图书馆的满意程度。RFID可以将多本书随意放置在读写器读写范围,一次性实现借阅或归还操作,是条码操作效率的几倍,效率大大提高。

3.2 大幅降低图书盘点和查找工作量

依靠人工的图书盘点工作,特别是人工对书架图书的盘点工作量大而且效率很低。图书管理员盘点书架图书要凭自身的记忆对图书进行分类放置和记录,费时劳神又很难达到目的。引入先进的RFID图书馆盘点工具和方法,可以实现图书盘点的自动化。利用RFID非接触、远距离、快速读取多个标签的特点,结合盘点车和层架标签,可以很方便地实现书籍盘点、顺架、查错架、缺架清查工作;同时上架时还可根据书库图形化路线指示正确位置摆放馆藏。

3.3 提高图书馆工作人员的工作满意度

图书馆对RFID系统的应用,减少了流通工作量,剩余的流通工作也是配合RFID技术的自动化工作,解放的流通馆员,可以从事其它的高级咨询工作。

3.4 提高读者满意度

利用RFID技术,读者不用排队等候,一次性借阅图书,图书馆的开放时间更长以及RFID技术带来的全新感受。

4 结束语

RFID技术引入图书馆领域,并非简单照搬,也非为图书馆行业的特殊需要而量身制作,它存在诸多局限和空白。因此,在积极引进、应用国外RFID技术的同时,应结合国内图书馆工作实际,以先进的文献服务理念 and 馆藏管理模式为先导,打破原有的传统束缚,大胆地进行探索和改革,使RFID技术成为图书馆提高服务水平、实行馆藏文献管理智能化的有效手段,进而解决全行业性难题,使图书馆读者和社会、国家普遍受益^[9]。基于RFID技术的图书馆编目系统建设是一项系统工程,不仅建设周期长,而且涉及到人、财、物等诸多方面的管理,需在建设前期做好充分的调研工作,既要当前技术发展和应用现状有深入了解,又要充分考虑本馆自身建设的需求,避免盲目建设。

参考文献

- [1] 李佳.基于RFID技术的昆明学院图书馆信息管理系统分析与设计[D].昆明:云南大学,2015.
- [2] 马瑞,洪光宗.以开放的心态迎接新的信息技术[J].中国图书馆学报,2010,(3):86-89.
- [3] 周文豪.数字图书馆以及RFID在世界图书馆中的应用现状[J].中国电子商情,2008,(3):66-70.
- [4] 姜爱蓉.图书馆新技术趋势与应对[C].2010高校图书馆论坛,2010.

新媒体背景下西部高校图书馆信息服务研究

席亚军

(塔里木大学图书馆, 新疆 阿拉尔 843300)

摘要: 文章概述了移动互联网高速发展背景下国内西部高校图书馆新媒体服务的现状, 并指出存在的问题, 提出今后发展的对策。

关键词: 西部高校图书馆; 新媒体; 移动图书馆

中图分类号: G250

文献标识码: A

文章编码: 1672-6251(2016)03-0047-02

Research on the Information Service of West-China University Libraries under New Media Environment

XI Yajun

(Library of Tarim University, Xinjiang Alaer 843300)

Abstract: The paper summarized the current status of new media service of domestic West-China university libraries in the context of rapid development of mobile network, and pointed out the existing problems, and proposed the solutions for the further development.

Key words: West-China university library; new media; mobile library

据《中国新媒体发展报告 2015》统计显示, 2015 年我国互联网网民将接近 7 亿人, 手机网民超过 6 亿人, 同时, 手机网民规模首次超过台式机网民。网民中利用手机上网的比率达 85.8%。这些数据表明, 网民正大规模向移动互联网平台迁徙, 促使新媒体的移动化转型。

我国已有多家图书馆通过积极探索和努力, 在向全媒体形态转型方面取得了较好的成果。上海图书馆于 2005 年启动了手机图书馆服务, 是国内手机服务的最早实践者; 2010 年推出 24 小时自助图书馆。杭州数字图书馆开通了“文澜在线”, 实现了基于互联网的多种技术应用, 包括在线展览、虚拟现实等服务, 实现了所有杭州地区的 IP 地址用户通过网站平台免费获取各种资源。同时开通了手机服务, 注册用户都可通过手机平台使用数字图书馆。杭州有线电视用户还可以 24 小时点击杭州数字图书馆的电视服务。

1 西部图书馆现状分析

基于国家对西部对口支援的大好时机, 国家大力

扶持西部各项建设, 特别是对教育的重视和投入都大大加强, 近些年来西部高校图书馆取得了跨越性发展。但由于地处西部欠发达地区, 起步较晚、发展缓慢, 现在正处于由数字图书馆向智能化图书馆的转变的初级阶段。

据《中国统计年鉴》(2003 年) 统计, 全国共有普通高等学校 13% 所, 西部地区 12 省(市) 共有普通高等学校 331 所, 占全国的 23.7%。而且各地区高校布局和发展都不平衡。多年来, 尽管西部地区高校图书馆, 尤其是一些重点高校为西部的经济建设做出了很大贡献, 但由于西部地区集中了我国大部分的贫困人口, 造成了西部经济、交通、通讯相对落后, 人均国内生产总值只有全国水平的 1/2 左右。绝大部分高校图书馆的藏书、自动化、信息化、数字化建设方面仍处于落后状态, 整体水平比起东部地区还存在一定的差距^[1]。

2 西部高校图书馆新媒体服务实践

移动技术是促使高等教育快速达到目标的一种让

基金项目: 2015 年 CALIS 农学中心研究项目“大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究”(编号: 2015032)。

作者简介: 席亚军(1981-), 男, 本科, 馆员, 研究方向: 数字图书馆。

收稿日期: 2016-03-04

人信服的、新的手段,移动设备解放了学习工具对学习者的限制,实现了无处不在的学习,新的移动计算科学即将引入要教育之中。在一定的时期内,移动技术也是西部高校图书馆发展的重点。

2.1 移动图书馆

高校移动图书馆主要为高校的老师、学生提供便捷的、随时随地的信息服务,极大地方便了老师和学生的教学科研学习活动。随着移动图书馆的蓬勃兴起,众多高校也开始开发建设移动图书馆,探索着为用户提供移动服务。但东西部高校移动图书馆的发展有一定差距,今后一段时间移动图书馆是西部高校发展的重点。

移动图书馆界面一般包括资源导航模块、我的中心模块和信息发布模块。其中,资源导航模块包含热门书推荐功能、图书分类导航功能、期刊分类导航功能;我的中心模块可通过移动图书馆平台与图书馆集成管理系统的对接,提供借阅证挂失、馆藏查询、预约借书、个人借阅历史查询、图书续借、咨询、移动图书馆检索历史记录、浏览历史记录等个性化自助服务;信息发布模块可以自主完成移动图书馆提供的新闻、图书馆通告、新书推荐、借书到期提醒等信息订阅服务。

2.2 电视图书馆

电视图书馆依托丰富的馆藏文献资源和数字化成果,电视图书馆基于开放互联网络、多媒体、通讯等多种技术于一体,向读者推出了“文献检索课”互联网电视服务。学生在宿舍中通过互联网,即可以观看图书馆的精品课程、特别活动、图书推荐、典籍鉴赏、名著赏析等内容,以视频点播方式为用户提供新型电视图书馆体验。

2.3 触摸屏服务

读者无需键盘和鼠标,只需手指轻触液晶显示屏

的图形、文字或按钮,便可实现触摸屏系统内数字资源的浏览与互动。触摸屏系统是高校图书馆向读者提供的查阅数字资源的新的方式,其方便、快捷、人性化的设计受到了读者的欢迎。该系统依托高校图书馆的馆藏资源和各项服务,为读者提供了内容丰富的资源展示。数字资源触摸体验系统展示的资源已经包含了馆藏资源、电子报刊、在线讲座、在线展览、服务介绍及政府信息公开等丰富的内容。

触摸屏系统为读者查阅馆藏资源提供了新的途径。触摸屏系统内展示的馆藏资源包含了古籍、书刊、老照片等内容。通过触摸屏系统访问馆藏资源,既实现了对馆藏资源的保护,也实现了文明的传承,方便了读者使用。电子报刊是触摸体验系统内更新速度最快的资源之一。通过触摸体验系统查阅电子报刊,读者可以方便快捷地了解国内外的时事动态。读者通过点击、滑动等简单操作,便可实现电子报刊的选择、放大、缩小、翻页、拖动、查看往期电子报刊等操作。

3 结束语

西部高校图书馆意识到新媒体的价值,主动应用新媒体技术,开展了丰富多彩的活动,并在实践中总结了经验,拓宽了思路,凝练出行之有效的服务模式,勾勒出一套操作性强的技术路线、如何利用和驾驭好新媒体服务已成为图书馆行业热议的问题,加强新媒体服务的研究、并使这些研究成果真正应用于图书馆,是业内学者和图书馆员的共同使命^[2]。

参考文献

- [1] 米玲.西部高校图书馆为地方经济建设开展信息服务探索[J].重庆工商大学学报(自然科学版),2011,(1):99-102.
- [2] 黄国凡,肖铮.图书馆新媒体服务的实践与思考[J].情报资料工作,2014,(6):85-88.

边疆高校图书馆图书借阅量下降分析及对策

—以塔里木大学图书馆为例

席亚军

(塔里木大学图书馆, 新疆 阿拉尔 843300)

摘要: 文章针对高校图书馆纸质图书借阅量下降这一现象,以塔里木大学图书馆 2013-2015 年图书借阅数据进行统计分析,得出近三年来各类纸质图书的借阅量的变化情况。在分析图书馆借阅量下降原因的基础上,提出相应的建议和对策。

关键词: 高校图书馆; 图书借阅量; 纸质图书

中图分类号: G251

文献标识码: A

文章编号:

Study on the causes of and countermeasures for the downward trend of books' circulation rate of university library

—A case study of library of Tarim University

XI Yajun

(Library of Tarim University, Xinjiang Alaer 843300)

Abstract: On the condition of the decline of paper books' loan amount, the loan amount changing situation of all kinds of Chinese books are made based on the statistical analysis of loan data of Chinese book from 2013 to 2015 in the library of Tarim University. on the basis of analyzing the reasons of the decline, it finally puts forward some proposals and countermeasures.

Key words: university library; books' circulation rate; paper books

塔里木大学图书馆在文献资源建设中始终把农业、生物科学、水利、建筑、计算机、经济管理、法律、文学等学科文献作为收藏重点,注重对重点学科文献及新疆地方特色文献的采集,同时兼顾其他学科文献的系统收集,图书馆纸质文献以每年 5 万册增量稳步增长,截止 2014 年底,馆藏纸质文献达 120 余万册,电子图书 80 万种,其中,中文电子图书 75 万种,外文电子图书 5 万种;每年订购中外文报刊约 800 种,现已基本形成了以新疆地方馆藏文献为特色,其中农学、生物科学、新疆地方文献已达到或接近研究级馆藏要求。文章以塔里木大学图书馆近几年纸质图书流通借阅数据为依据,对借阅情况作实证研究。通过研究,了解读者对文献资源借阅量的动态变化,分析影响借阅量的原因,有针对性地提出了改善的对策和建议,以期为学生提供有针对性的、更加优质的服务,充分发挥图书馆在高校教育中的职能和作用^[1]。

1 纸质图书借阅量数据统计与分析

1.1 统计方法

本文以塔里木大学图书馆为调查对象。调查时间为近几年读者图书借阅量以及近一年图书借阅学科分布统计,以《中国图书馆分类法》为依据,把每年图书借阅总量进行对比,并对图书借阅数据进行了调查与分析,找出存在的问题,为促进纸质资源的借阅与利用提供理论依据。

1.2 塔里木大学图书馆纸质图书借阅现状

根据近 3 年图书借阅量统计数据,我馆纸质文献以每年 5 万册增量稳步增长,图书借阅量却呈现下降趋势,其中,2013 年度纸质图借阅量为 90876 册,2014 年度纸质图借阅量为 82630 册,2015 年度纸质图借阅量为 68006 册,图书借阅量下降幅度很明显,尤其是 2015 年度下降趋势更为明显,其中,2013-2014 年度图书借阅量下降为 9%,2014-2015 年度图书借阅量下降 17.69%。截止 2015 年年底,与 2014 年同期相比,借阅册数少了 14624 册,

2014 年较 2013 年同期相比, 借阅册数少了 8246 册, 总之, 图书借阅量是逐年下降趋势, 但是与之相反的, 我校招生每年都有所增加, 这就给我们图书馆急需解决的问题。

1.3 读者对各类图书的借阅情况统计与分析

我馆自动化系统使用的是汇文图书管理系统, 读者借阅身份包括本专科生、研究生、学校教职工和校外读者, 所用数据均为真实数据, 能较真实地反映我校读者对流通纸质图书的实际借阅情况。笔者对我馆 2014 年-2015 年近两年图书借还分类综合统计, 统计结果如下表 1。

表 1 塔里木大学图书馆 2014 年-2015 年图书借阅分类综合统计结果

类别	类号	2014 年	2015 年	类别	类号	2014 年	2015 年
马列类	A	135	86	自然科学	N	103	98
哲学类	B	7706	6124	数学	O	4296	3215
社科总论	C	1634	1340	天文地球	P	218	153
政治法律	D	2626	2210	生物科学	Q	1496	1302
军事类	E	200	183	医药、卫生	R	1200	738
经济	F	4127	3355	农业科学	S	2627	1989
文化教育	G	1896	1683	工业技术	T	10801	8378
语言、文字	H	10115	11769	交通运输	U	224	237
文学	I	24041	19730	航空、航天	V	14	12
艺术	J	1465	1740	环境科学	X	75	91
历史、地理	K	5372	4745	综合图书	Z	149	158

根据统计结果显示, A、E、N、P 类图书借阅量整体呈下降趋势, 年图书借阅总量不高, B、C、D、F、G、I、K、O、Q、R、S、T、V 类图书借阅量整体呈下降趋势, 其中, T 工业技术类图书下降比较明显, 由 2014 年 13.07% 下降到 2015 年 12.12%; I 文学类图书借阅总量整体持平, 下降幅度不是很明显, 由 2014 年 29.09% 下降到 29.01%。H 类语言文学类图书借阅总量有所上升, 但上升幅度不是很明显, 由 2014 年 14.24% 上升到 2015 年 14.87%。从上述结果分析来看, 爱好文学书的读者在塔里木大学占了相当大的比例, I 文学类图书成了多数读者的首选。H 语言文学和 T 工业技术类图书借阅量总体也很高, 主要是英语、计算机的普及程度越来越高, 尤其是近些年来, 大学生就业形势严峻, 企业在选拔人才时对外语、计算机操作应用能力都有一定的要求, 如英语四六级证、计算机等级证书。国家教委也明文规定大学生必须参加英语、计算机等级考试, 迫使每一位大学生都必须学习英语、计算机, 因而这两类图书需求较强。上述分析结果得出, 纸质图书的流通率整体呈下降趋势, 我馆还要有针对性对馆藏进行调整。

随着互联网 APP 的迅速发展, 阅读形式呈现多元化, 网络阅读、手机阅读、浏览器下载阅读等各种阅读模式, 以及电子书兴起使图书馆纸质图书借阅率下降的主因。在“互联网+”、背景下, 移动图书馆、手机图书馆、触摸屏服务、互联网电视等新媒体越来越完善, 让读者更加便捷和随心所欲。同时遇到需要解决的实际问题或写论文时不再单纯依赖纸质书籍, 而是把眼光投向读者网络和搜索引擎, 其中, 我馆图书馆也购买了电子资源供读者使用, 超星电子书和维斯博文外文电子书图书总量为 80 万种, 中文电子书 75 万种, 外文电子书 5 万种。其实, 不论是纸质阅读还是数字阅读, 人们对阅读的需求还是存在的, 只是方式的改变。阅读方式的多元化是导致纸质图书借阅量下降的主要因素。

2 新环境下提高图书馆图书借阅量的对策

越来越多的电子书、电子期刊、电子出版物的人量涌入, 改变着图书馆的文献资源结构, 也改变了读者的阅读方式, 迫使图书馆改变观念和工作方式, 创新适应新环境下的工作模式。

2. 1 根据学科专业建设和读者阅读倾向, 制定合理的采购计划

图书采购人员积极调查研究学校专业设置及重点学科建设情况, 详细了解我校所设专业学科性质、培养口标、学科建设发展规划等。分析研究读者需求, 掌握读者的阅读倾向和图书需求的规律及特点, 并据此进行综合分析研究, 制定合理的采购计划, 使藏书体系不断丰富和完善。

2. 2 调整书库布局, 优化借阅环境

首先, 在书库正中的书架上增设经典书架, 将经典图书和流通率极高的各类图书组织到这几个最显眼的书架上, 方便读者借阅;其次, 可将读者借阅率较高的图书依次从中间几排的架位开始上架;最后, 再将读者借阅较少的基础类藏书放到书库靠后的书架上^[2]。

2. 3 延长图书馆开放时间

塔里木大学图书馆现在每周开放时间 50 小时左右, 开放时间很短, 远远达不到《普通高等学校图书馆规程》所要求的每周开放 90 小时的要求, 延长时间后, 从早上 10:00 到晚上 21:00, 每周开放时间达到 77 小时, 最大限度的满足读者阅读需要, 从一定程度上能提高图书借阅量。

3 结束

总之, 图书馆要根据图书馆藏书特点和读者需要, 充分利用图书馆各类工具, 充分利用现代化信息技术手段, “以人为本”, 为科研、教学等方面更好的服务。以便能够最大限度地发挥图书馆的馆藏效益, 能最大限度地促进学校教学、科研及学生的学习^[3]。

参考文献

[1]白海燕. 高校图书馆读者阅读倾向分析与对策研究[J]. 陕西教育(高教版),2014,03:62-63.

[2]梁秀玉. 纸质图书借阅率下降分析及应对策略[J]. 内蒙古科技与经济,2011,16:128-129.

[3]李秀凤,曹健. 图书借阅统计分析[J]. 农业图书情报学刊,2008,09:83-84.

作者简介: 席亚军 (1981.10-), 男, 河北平泉县, 馆员, 学士, 研究方向: 数字图书馆。
电话: 13565142267

基金项目：2015 年 CALIS 农学中心研究项目“大数据环境下云存储在高校图书馆数字资源长期保存中的应用研究”（编号:2015032）。

作者简介：席亚军（1981-），男，本科，馆员，研究方向：数字图书馆。

收稿日期：2016-04-07

邮寄地址：843300 新疆阿拉尔市塔里木大学教工区 30 号楼 1 单元 301 室 席亚军（2 本）

手机号码:13565142267

E-mail: smallx1002@sina.com QQ:187170403

基于 TPI 系统的塔里木大学图书馆特色数据库建设实践

席亚军

（塔里木大学图书馆，新疆 阿拉尔 843300）

摘要：文章概述了 TPI 系统技术及功能，阐述了利用 TPI 建立塔里木大学图书馆特色库的方法及步骤。

关键词：TPI；特色数据库；数字图书馆；

中图分类号：G250.74

文献标识码：A

文章编号：

based on TPI systems Library Database Construction Practice Tarim University

XI Yajun

(Library of Tarim University, Xinjiang Alaer 843300)

Abstract: The article provides an overview of TPI system technology and functions. The method of using TPI established Tarim University Library Features library methods and procedures.

Key words: TPI; Special Date-base; Digital Library

1 引言

随着图书馆智能化的日益发展，特色数据库已经成为现代图书馆数字化建设的重要内容。每所高校都有自己的办学重点，高校图书馆也会在相应的学科方向上形成较为丰富的文献资源。高校图书馆也会根据自身的特点和资源收藏文献，形成较为丰富的馆藏文献，那么，怎样更好地发挥馆藏优势，为读者、教学和科研提供更好的服务；怎样整合现有资源，开发新资源，提高资源利用率；又怎样将自己拥有的具有特色的信息资源数字化，放到网上进行共享，这些都是目前我们面临的最迫切的问题。其可行的办法就是走特色化办馆之路，在文献资源建设与用户服务上充分挖掘自身的优势，来达到融合学院、教学及科研的目的。特色数据库是指针对用户的信息需求，对某一学科或某一专题有利用价值的信息进行收集、分析、评价、处理、储存，并按照一定标准和规范将特色资源数字化，以满足用户个性化需求的信息资源库^[1-2]。

2 TPI 系统简介

TPI 是清华同方股份有限公司开发的一个以管理内容为核心的综合数字图书馆解决方案，它是以数字化资源、分布式管理和智能化服务为基本要素而提出的解决方案，它适用于数字图书馆、各大档案馆等领域。清华同方 TPI6.0 专业数据库制作管理系统是同方知网（北京）技术有限公司自主研发的数字化图书馆系统，它是一套基于网络平台上用于知识仓库创建、生产、管理、维护和发布的工具软件系统，提供全文检索、电子书制作、文档统一转换、

元数据标引、文档分类管理、内容动态发布等功能，实现对文本、图像、音频、视频等各种类型资源的管理，可广泛应用于数字图书馆特色库建设、数字档案馆建设和各种类型的文档管理系统。使用它可以快捷、方便地掌握知识仓库建库、管理以及网上发布技术，建设自己的知识仓库和知识网站，将知识消息资源的建设与交流推广到全社会。TPI6.0系统的核心是KBASE检索服务器，它实现了信息的组织、存储与检索。电子图书制作、元数据加工工具系统实现了原始文献资料的数字化和组织；内容发布、检索网关和TPI检索服务器一起实现了资源发布。

3 塔里木大学图书馆特色数据库建设

目前大部分图书馆都有自己特色数据库，如何将本馆拥有的具有特色的文献信息资源数字化，并放在网上供读者查阅，是目前各图书馆迫切解决的问题。2015年塔里木大学图书馆购买清华同方数字图书馆TPI6.0系统，作为我馆建设数字图书馆特色数据库的应用建设平台，进行了该馆特色馆藏的建设工作。在实际工作中取得了一定的成绩。塔里木大学图书馆西域研究文献库藏书2万多册，现在不对外借阅，只限参考阅读和复印，考虑到上述情况，我馆要进行特色数据建设，以解决目前图书馆西域研究文献库图书不能外借问题。目前，塔里木大学图书馆西域研究文献库重点是学科新疆地方特色文献，包含新疆地名志、新疆人物志、新疆博物馆志、新疆印象等内容。

3.1 电子书加工流程

通过TPI电子书加工工具可以把书籍、报刊、公文等各类纸张文档，通过书页扫描、书页图像处理、目录树编改、书页压缩等过程简单快速地形成电子书。也可以把已有各种电子文件，通过电子书加工软件打印驱动转化为电子书。其中包括书页图像扫描及管理、书页图像处理、书页图像识别导入及编改、目录加工处理、书页图像及目录合并形成电子书。工作流程，如图1

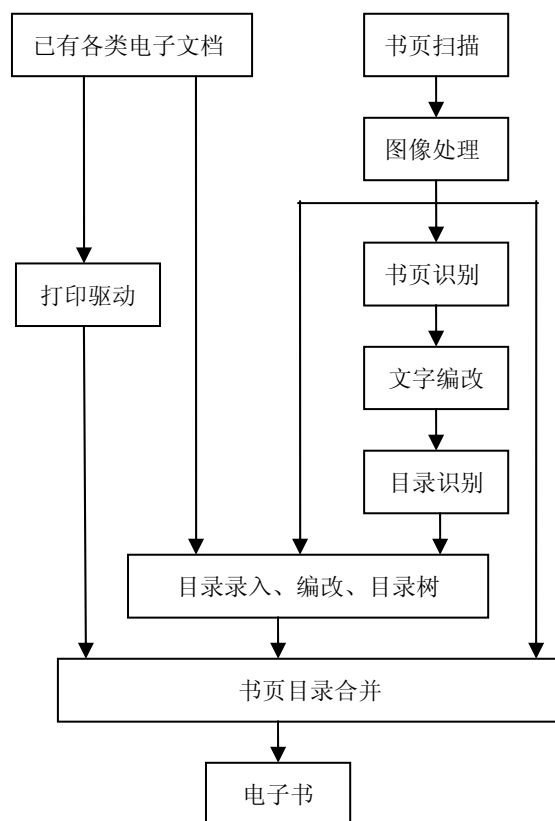


图1 电子书加工流程

3.2 扫描图书

书页扫描是指对要进行加工成电子书的书籍、报刊、公文等各类纸质文档通过扫描仪进行扫描,使它变成可以利用计算机处理的图像或文字的电子文档。书页扫描模块利用 TWAIN 接口驱动扫描仪对书页进行扫描。它包括选择扫描仪、扫描新页、扫描前页、扫描后页、覆盖页等功能。书页扫描时各种参数的选择对制作电子书有一定的影响,其中以压缩率、分辨率、扫描模式影响最重要,压缩率是影响图像文件大小的主要参数,分辨率是影响图像清晰度和扫描速度的主要参数,如果综合考虑扫描得到的图片质量、文件大小、工作效率等因素,一般来说,按照以下方法设置扫描仪,黑白文字页:300DPI,线条图(二值)扫描模式;灰度图像页:200DPI,256 灰度扫描模式;彩色图像页:150DPI,RGB 真彩色扫描模式;以上扫描仪设置满足了阅读清晰度,又使文件大小适宜,达到了比较满意的效果了。

3.3 图像处理

从扫描仪获得数字图像后,由于存在图像倾斜、噪声、污迹、装订孔、文字内容版心不正等问题,所获得数字图像并不理想。如果直接生成电子书,影响读者的阅读。同时,大量的污迹也会降低 OCR 图像识别的识别率,增加编改的工作量。因而有必要对书页图像进行图像编辑加工。电子书加工软件包含书页加工过程中所必需的图像处理功能,包括书页高速成批扫描、自动倾斜校正、去噪、剪裁、翻转、灰度图像页自动搜索、灰度区自动标记、灰度图像页亮度及对比度的调整,还增加了自动目录树识别功能,目录树节点图像 OCR 识别录入等方式,加工书籍导航树的效率成倍提高。

3.4 生成电子书

TPI 生成电子书模块对书页扫描图像进行压缩、打包处理,形成电子书。电子书加工软件还能利用打印驱动把 WORD、PDF、HTML、PS、S2、S72、PS2、PSD、TXT 等已有的各种电子文件转换为浏览器可识别的 NH 文件。如果有目录树,在生成电子书的过程中,自动把目录树并入打包文件中,形成带目录树的电子书。电子书支持将“书页列表区(书页树)”中的所有图书的图像页压缩、打包生成各自的电子书。自动背景二值化,灰度图叠加显示,生成最优图像版电子书。背景二值图采用 JBIG 压缩,前景灰度图采用 JPEG 压缩,使生成的电子书文件大小最小。

3.5 页面发布

TPI 数据库发布提供了多种风格的模板,同时也可以自建发布模板。发布时,首先选择要发布的数据库,选择“数据库发布菜单”,配置数据库版权信息;然后选定适合内容发布的模板,定制要发布的检索字段,作为发行数据库的检索字段、选择要发布的概览字段和细览字段,配置检索选项,完成发布设置。特色数据库发布的页面设计与 CNKI 页面设计符合用户的检索习惯,但也可以根据所建特色数据库的内涵,突出特色,建立个性化页面。在发布前设定检查项,可以在数据入库发布前发现问题,适时修改。TPI 在线浏览技术使读者在阅读电子书时,可以选择下载阅读和在线阅读两种方式。TPI 数据库发布与检索系统可实现各种检索服务及全文检索。该系统能为用户提供新闻、网络导航、学位论文、教师参考书、博硕士论文提交等内容发布形式^[4]。

数据库发布及检索是 TPI6.0 的主要功能,新增加了新闻发布、跨库检索、数据库内容展示等功能。基于浏览器的后台管理,更加方便使用。每种模板发布时都可以选取检索字段,概览字段以及细览字段,概览和细览字段还可以为每个字段及字段内容选择颜色、大小、字段、粗体、斜体等属性。TPI 系统提供 6 种数据库发布类型,分别为检索字段发布、概览字段发布、细览字段发布、排序字段发布、关联数据库发布、关联字段发布,可适应不同用户的检索习惯^[3]。我馆选择了 CNKI 格式作为发布模式。发布时,只要在“数据库管理栏”里选择要发布的数据库,设置好概览字段和细览字段等项目,就可以进行数据库的 Web 发布。

4 结束

TPI 系统已经是现代图书馆数字化建设的重要内容,尤其是特色数据库的建设,随着高

校图书馆服务的转型,建设自己的特色数据库是高校图书馆的首要选择,这样会更好的为地方服务,也更好的将当地的文化遗产下来。总之,通过构建特色资源数据库,可以探索出一条由传统图书馆向数字图书馆转化的道路。塔里木大学图书馆在建设数字化图书馆过程中,始终把建设特色数据库作为重要的基础工作,利用 TPI6.0 系统工具将特色文献资源数字化并纳入计算机网络中共享,可使更多的读者了解本馆的情况,提高馆藏文献资料的利用率,发挥自己的特点和优势,为社会提供更多的信息服务,同时提高图书馆的知名度。

参考文献

- [1]吴南. TPI 系统在黑龙江科技学院图书馆特色数据库建设中的应用[J]. 科技情报开发与经济, 2008, 17: 65-66.
- [2]陈春娥. 基于 TPI 系统的集美图书馆特色数据库建设实践[J]. 情报探索, 2011, 01: 79-81.
- [3]徐良. 利用 TPI 系统加工电子书的方法[J]. 图书馆学刊, 2006, 01: 119-121.
- [4]吴涛, 李锋. 基于 TPI 的特色数据库建设实践[J]. 现代情报, 2005, 07: 165-168.

附录

附录一：软件著作权扫描件

附录二：数字图书馆云存储系统机房搭建环境及培训演示照片

附录三：论文录用通知《大数据时代下云技术在图书馆数据存储中的应用》

附录四：论文录用通知《数字图书馆云存储系统的设计与实现》

附录五：论文录用通知《高校图书馆数字资源建设现状及发展趋势研究》

附录六：论文录用通知《基于 RFID 技术的图书馆编目系统建设探索》

附录七：论文录用通知《新媒体背景下西部高校图书馆信息服务研究》

附录八：论文录用通知《边疆高校图书馆图书借阅量下降分析及对策
-以塔里木大学图书馆为例》

附录九：论文录用通知《基于 TPI 系统的塔里木大学图书馆特色数据库建设实践》

附录十：数字图书馆云存储系统设计说明书

附录十一：塔里木大学数字图书馆云存储系统程序部分关键代码

附录一

软件著作权 1 项，名称：数字图书馆云存储系统 V1.0，登记号：2015SR147383



附录二



课题主持人介绍数字图书馆云存储系统使用说明



课题组成员进行数字图书馆云存储系统搭建演示



课题组成员进行数字图书馆云存储系统培训

附录三

《农业图书情报学刊》(月刊)稿件录用通知

席亚军作者:

您好!

您的文稿:《大数据时代下云技术在图书馆数据存储中的应用》稿件编号:15-0870;字数:____11103____(按版面折合),经审查研究,拟全文刊用(编辑可作文字修改),将在2015年第____10或11____期刊登。

请缴纳:发表费(人民币:90元/千字):人民币____990____元。

审稿费(50元/篇):人民币__50__元。

加急费:人民币____元。

合计:人民币__1040__元。

请将上述费用于2015年__6__月__25__日前汇至编辑部,两种汇款方式如下:

1. 邮寄(针对个人汇款):100081北京海淀区中关村南大街12号《农业图书情报学刊》编辑部收。汇款时请在收款人处写清:“《农业图书情报学刊》编辑部收”字样。请务必在留言栏注明:“《学刊》版面费,稿件编号15-0870,要(不要)发票”字样及第一作者姓名。对开具发票的“付款单位(抬头)”必须注明同时汇款人(作者)的收件地址需清楚标明。

2. 信汇(只限通过单位向银行转账汇款之用):

账号:11050601040009874

开户银行:中国农业银行北京北下关支行

户名:中国农业科学院农业信息研究所

信汇时请务必在用途栏注明:“《学刊》版面费,稿件编号15-0870”及第一作者姓名。

查询汇款,请明示汇款日期,请向邮局索要汇款收据,以备之需。同时录用两篇文章以上的作者,如刊期不能在同一期上,需向编辑部声明。

稿件一经录用,不予更改。汇款超期,发表刊期顺延。超期2个月,文稿将从系统删除,但录用仍有效。谢谢合作!

查询稿件请明示稿件编号及录用刊期。另外,赠送作者的样刊为挂号邮寄,需要快递的作者请及时与编辑部联系并汇款。

编辑部地址:100081北京海淀区中关村南大街12号中国农业科学院农业信息研究所

《农业图书情报学刊》编辑部

联系电话:010-82109667

电子信箱:xuekan@caas.cn

联系人:孔晔晗

如有异议,请速与编辑部联系。谢谢支持与信任!

《农业图书情报学刊》编辑部;2015-06-15(已盖章备案);(此件请保留至文稿发表)

附录四

《河南图书馆学刊》用稿通知

作者 你好：

你的文章经专家认真审阅被我刊采用，拟安排 2015 年择期刊发，请勿他投，请按以下要求修改文章，[修改后的文章请发往 993489580@qq.com](mailto:993489580@qq.com) 邮箱中，请在电子邮件发送主题中注明“作者姓名”及“已按模版修改稿”字样，论文中请附上作者（第二及第三作者亦需）信息（出生年、工作单位、现从事工作、职称、职务等），作者详细联系方式（手机、详细的邮寄地址）。

文章修改整体要求：

1. 论文题目控制在 20 字之内（不含副标题）。
2. 请作者对文章进行认真校对，字数控制在 4000-6000 字，因差错（包括错别字、标点符号错误）过多导致文章延期发表，由作者承担责任。
3. 文中参考文献请遵照国家标准（《文后参考文献著录规则》（GB/T 7714-2005））标注，将文中注释号变为红色右上标。

4. 表格为田字格模式

一、字体字号

1. 摘要、关键词、中图分类号、文献标识码、文章编号：五号黑体。
2. 正文：五号宋体。
3. 文尾参考文献：五号黑体。

二、文中标题级别排序顶满格

1. 一级标题用：五号黑体（标题后无句号，不接排文字）。
2. 二级标题用：五号宋体（标题后无句号，不接排文字）。
3. 三级标题用：五号宋体（标题后无句号，不接排文字）。
4. 各级标题序号均使用阿拉伯数字（如一级标题：1 2 3；二级标题 1.1 1.2；三级标题 1.1.1 1.1.2）

三、参考文献

参考文献必须使用方括号，字体大小与正文同，标以红色，以便于检查核对。

论文编审费：1000 元

邮局汇款到：《河南图书馆学刊》编辑部 徐黎娟收

地 址： 郑州市嵩山南路 150 号 河南省图书馆 邮编：450052

联系电话： 0371-67181468

联 系 人： 徐黎娟 严真

汇款单上请注明：[论文标题](#) [手机号](#) [作者详细地址](#)

编审费只能通过邮局汇款，不支持银行汇款，请保留邮局汇款单。

文章刊出后会给每位作者以快递到付的形式邮寄一本，如需平邮请注明。

《河南图书馆学刊》编辑部

2015 年 7 月 14 日

附录五

《河南图书馆学刊》用稿通知

作者 你好：

你的文章经专家认真审阅被我刊采用，拟安排 2016 年择期刊发，请勿他投，请按以下要求修改文章，**修改后的文章请发往 2078743893@qq.com** 邮箱中，请在电子邮件发送主题中注明“作者姓名”及“已按模版修改稿”字样，论文中请附上作者（第二及第三作者亦需）信息（**出生年、工作单位、职称**），作者联系方式（**手机、详细的邮寄地址**）。

文章修改整体要求：

1. 论文题目控制在 20 字之内（不含副标题）。
2. 请作者对文章进行认真校对，字数控制在 4000-6000 字，因差错（包括错别字、标点符号错误）过多导致文章延期发表，由作者承担责任。
3. 文中参考文献请遵照国家标准（《文后参考文献著录规则》（GB/T 7714-2005））标注，将文中注释号变为**红色右上标**。

4. 表格为田字格模式

一、字体字号

1. 摘要、关键词、中图分类号、文献标识码：五号黑体。
2. 正文：五号宋体。
3. 文尾参考文献：五号黑体。

二、文中标题级别排序顶满格

1. 一级标题用：五号黑体（标题后**无**句号，**不**接排文字）。
2. 二级标题用：五号宋体（标题后**无**句号，**不**接排文字）。
3. 三级标题用：五号宋体（标题后**无**句号，**不**接排文字）。
4. 各级标题序号均使用**阿拉伯数字**（如一级标题：1 2 3；二级标题 1.1 1.2；三级标题 1.1.1 1.1.2）

三、参考文献

参考文献必须使用方括号，字体大小与正文同，标以**红色**，以便于检查核对。

论文编审费：1000 元

邮局汇款到：《河南图书馆学刊》编辑部 周雪芹收

地 址：郑州市嵩山南路 150 号 河南省图书馆 邮编：450052

联系电话：0371-67181468

联系人：周雪芹 崔 萌

汇款单请注明：**论文标题 手机号 作者详细地址**

编审费只能通过邮局汇款，不支持银行汇款，请保留邮局汇款单。

文章刊出后会给每位作者以快递到付的形式邮寄一本，如需平邮请注明，需要发票的不支持平邮。

《河南图书馆学刊》编辑部

2015 年 11 月 10 日

附录六

《农业网络信息》编辑部

论文刊用通知

《农业网络信息》由中国农业科学院农业信息研究所主办，国家农业部主管，是国家科技部和国家新闻出版总署批准的国内外公开发行的中央级刊物（刊号：ISSN 1672-6251 CN11-5065/TP；月刊），是中国核心期刊遴选数据库期刊、中国学术期刊全文数据库期刊、中文科技期刊数据库期刊，是CNKI与ASPT来源刊、中国期刊网来源刊等。

席亚军：你好。

您的论文“基于 RFID 技术的图书馆编目系统建设探索”（稿件编号：16-N060）一文，经有关专家审阅可以采用，需要收取版面费 550 元。请尽快办理版面费汇款，初步拟定于 2016 年 2 期刊登，2 月底出版。

顺祝：工作顺利。

汇款方式一（从邮局按地址汇款）：

邮 编：100081

地 址：北京市海淀区中关村南大街 12 号 中国农业科学院农业信息研究所

收款人：《农业网络信息》编辑部

（注：请在附言栏注明稿件编号以及发票抬头，推荐此种汇款）

汇款方式二（单位银行）

开户行：中国农业银行北京北下关支行

户名：中国农业科学院农业信息研究所

帐号：11050601040009874

备注：农网（必须写）

（注：通过银行汇出的版面费，用个人名字汇款的，发票只能开个人名字；，通过单位汇的款，发票就开单位。）

编辑部电话：010-82109657

投稿邮箱：nywlxx@caas.cn



附录七

《农业网络信息》编辑部

论文刊用通知

《农业网络信息》由中国农业科学院农业信息研究所主办，国家农业部主管，是国家科技部和新闻出版总署批准的国内外公开发行的中央级刊物（刊号：ISSN 1672-6251 CN11-5065/TP；月刊），是中国核心期刊遴选数据库期刊、中国学术期刊全文数据库期刊、中文科技期刊数据库期刊，是CNKI与ASPT来源刊、中国期刊网来源刊等。

席亚军：你好。

您的论文“新媒体背景下西部高校图书馆信息服务研究”（稿件编号：16-N081）一文，经有关专家审阅可以采用，需要收取版面费500元。请尽快办理版面费汇款，初步拟定于2016年3期刊登，3月底出版。

顺祝：工作顺利。

汇款方式一（从邮局按地址汇款）：

邮 编：100081

地 址：北京市海淀区中关村南大街12号 中国农业科学院农业信息研究所

收款人：《农业网络信息》编辑部

（注：请在附言栏注明稿件编号以及发票抬头，推荐此种汇款）

汇款方式二（单位银行）

开户行：中国农业银行北京北下关支行

户名：中国农业科学院农业信息研究所

帐号：11050601040009874

备注：农网（必须写）

（注：通过银行汇出的版面费，用个人名字汇款的，发票只能开个人名字；，通过单位汇的款，发票就开单位。）

编辑部电话：010-82109657

投稿邮箱：nywlxx@caas.cn



附录八

《农业网络信息》编辑部

论文刊用通知

《农业网络信息》由中国农业科学院农业信息研究所主办，国家农业部主管，是国家科技部和国家新闻出版总署批准的国内外公开发行的中央级刊物（刊号：ISSN 1672-6251 CN11-5065/TP；月刊），是中国核心期刊遴选数据库期刊、中国学术期刊全文数据库期刊、中文科技期刊数据库期刊，是CNKI与ASPT来源刊、中国期刊网来源刊等。

席亚军：你好。

您的论文“基于TPI系统的塔里木大学图书馆特色数据库建设实践”（稿件编号：16-N149）一文，经有关专家审阅可以采用，需要收取版面费700元。请尽快办理版面费汇款，初步拟定于2016年4期刊登，4月底出版。

顺祝：工作顺利。

汇款方式一（从邮局按地址汇款）：

邮 编：100081

地 址：北京市海淀区中关村南大街12号 中国农业科学院农业信息研究所

收款人：《农业网络信息》编辑部

（注：请在附言栏注明稿件编号以及发票抬头，推荐此种汇款）

汇款方式二（单位银行）

开户行：中国农业银行北京北下关支行

户名：中国农业科学院农业信息研究所

帐号：11050601040009874

备注：农网（必须写）

（注：通过银行汇出的版面费，用个人名字汇款的，发票只能开个人名字；通过单位汇的款，发票就开单位。）

编辑部电话：010-82109657

投稿邮箱：nywlxx@caas.cn



附录九

江苏科技信息杂志社



稿件录用通知

关于席亚军：你好！

来稿《边疆高校图书馆图书借阅量下降分析及对策——以塔里木大学图书馆为例》已收悉，经本刊编辑部审阅后，将在《江苏科技信息》2016年第13期予以刊登，特此通知。

《江苏科技信息》1984年创办，经国家新闻出版总署批准，由江苏省科学技术厅主管、江苏省科学技术情报研究所主办。

《江苏科技信息》为中国核心期刊（遴选）期刊、江苏省优秀期刊、江苏省一级期刊，被知网、维普、万方等多家数据库收录。

《江苏科技信息》中国标准连续出版物号：ISSN1004-7530，CN32-1191/T。

《江苏科技信息》杂志社

2016年4月5日

附录十

数字图书馆云存储系统 V1.0

设计说明书

一、项目背景

随着我国科学技术的飞速发展，云存储的计算技术也愈加成熟，高校的数字图书馆正向云存储建设出发，势必发展成为安全、经济、可靠、高效的云图书馆。显而易见，云图书馆的主要支撑技术就是云存储技术，既要想成功建设一个应用多、设备多、服务多的高效云图书馆集合体，就必须通过多种应用软件改造云存储数据中心，根据用户的需求为来访者提供服务。当然，要想让用户满意处于云存储环境下的数字图书馆，是依靠云虚拟化、个性化和能定制的业务访问和数据存储服务的。

不同于传统的数字图书馆，基于云计算技术和云存储技术下的数字图书馆拥有着更为先进的科学技术作为支撑。云存储系统的优势主要有计算能力强、存储效率高，运营环境的安全性高，自身功能也比较强健等，但同时云存储系统又是处于极为不安全的网络环境中，经常会出现非法用户的恶意攻击行为，再加上云存储的复杂特点，云存储业务的多样化办理，都给云图书馆的实际工作造成了一定的威胁，使之难以保证云存储数据高效、可靠、永久、真实地存储。

二、软件功能介绍

云图书馆是一个利用存储虚拟化、集群存储和 SAN+NAS 等技术，由网络设备、存储设备、服务器、应用软件、公用访问接口、接入网、和客户端程序等多个部分组成的存储区域系统，通过应用软件来对数字用户提供数据存储和云存储业务访问服务。其中应用软件部分主要实现了如下功能。

1、**底层基础存储功能**：基于云存储结构的底层基础存储，实现云站点之间数据同步、协同工作、线性扩容等，并提供上层应用层调用的接口，包括上传、下载、管理、浏览搜索等功能。

2、**PC 客户端**：同步客户指定的 PC 文件夹，保持和服务端文件的同步、文件共享、文件恢复、文件管理等功能，所有操作都是后台处理，对用户透明。

3、**Web 浏览器客户端**：提供 Web 浏览器模式的客户端，为用户提供上传、下载、文件共享、文件恢复、文件管理、图片预览等功能。

4、**Andrio 手机客户端**：实现基于 Andrio 系统的 JAVA 客户端，结合手机系统为用户提供上传、下载、文件共享、文件恢复、文件管理、图片预览等功能。

三、软件特性介绍

不言而喻，云计算与云存储的出现，给现有的数字图书馆注入了新鲜的力量，它推动了互联网下的数字图书馆的快速发展，为新时期云数字图书馆建设与服务提出了新的理念，创造出了 IT 界的一个里程碑。云存储技术在数字图书馆中的应用，解决了影响云图书馆的资源存储中碰到的诸多问题，例如：资源的共享、存取的困难、扩展范围的狭隘、动态存取数据的局限性等。云存储技术还运用硬件设备的共用和虚拟存储的科学技术，保证了云图书馆的数据的存储保密性、高效性以及共享性。云存储和云计算在不断地发展和完善后，使数字图书馆的发展进入一个全新的局面。

四、软件的运行环境介绍

软件应在以下两种环境中运行：

(1)硬件配置

本系统硬件表，如下表 1 所示：

表 1 硬件配置表

序号	类别	名称	配置	数量
1	PC_Server	应用服务器	四核*2/16G RAM/SAS 15K 300G*2/RAIDI/千兆网口*2/Linux AS5	4
2		缓存服务器	四核*2/16G RAM/SAS 15K 300G*2/RAIDI/千兆网口*2/Linux AS5	2
3		云存储服务器	四核*2/16G RAM/SAS 15K 300G*2/RAIDI/千兆网口*2/Linux AS5	2
4		数据库服务器	四核*2/16G RAM/SAS 15K 300G*2/RAIDI/千兆网口*2/Linux AS5	2
5		网盘数据 数据存储	NAS 存储, 1T*16, 300G*6	2
6	存储设备	数据库存储	光纤接口/控制器(1G Cache/4GB FC*2)*2/10*146GB (FC 15K)/4*4Gb 短波 模块	1
7	网络设备	交换机	48 千兆电口/4 光纤口	2
8		负载均衡	8 千兆电口/4 光纤口/4G 吞吐量/冗余电源	2
9		防火墙	4 千兆电口/4 光纤口/6G 吞吐量/220 万并发 连接/冗余电源	2

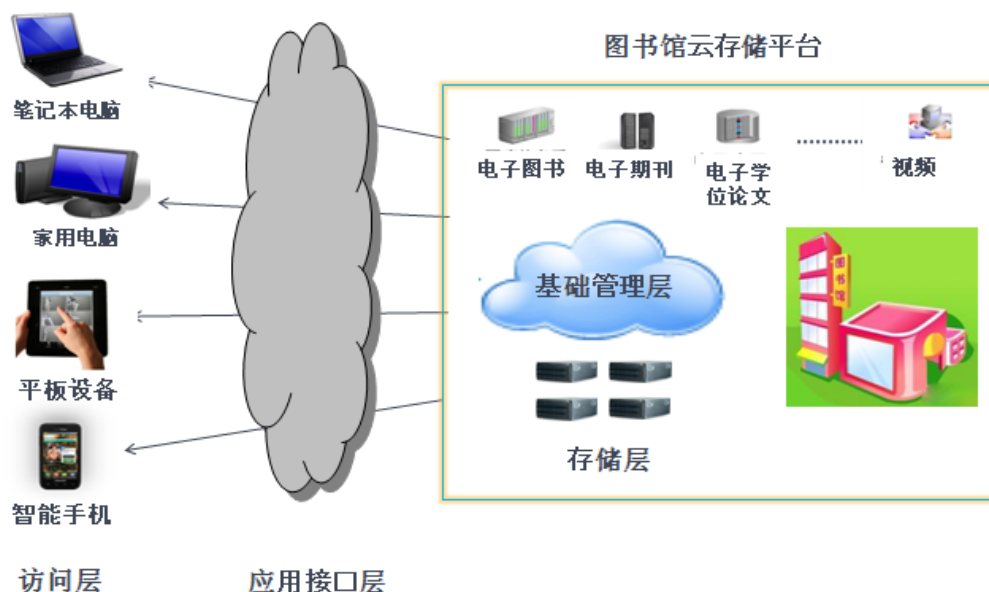
(2)计算机软件

采用 JAVA 开发平台,SQL Server2010 数据库,在 Windows 操作系统上运行。

五、系统结构

(1) 系统结构设计

数字图书馆的云存储系统是基于高性能网络传输技术,通过集群技术、分布式文件系统和网络计算技术等,根据客户服务需求将计算资源与已存储的数据资源按需分配。图书馆的数据资源云存储系统是一个由高性能网络传输设备、超大容量数据存储设备、超高性能服务器、特色服务应用软件和形式多样客户端等多个部分组成的复杂系统。通过虚拟化技术对云图书馆存储和数据操作进行管理,系统形成了以存储设备为核心,通过软件为客户提供数据存储和业务访问功能。



存储层是数字图书馆云存储系统的基础。图书馆将所有数字资源通过存储设备统一存放于存储层中，分布于世界不同地域的存储设备再通过各种网络设备连接在一起，由统一的存储设备管理系统管理，并利用虚拟化技术忽略掉硬件之间的差异，对外提供服务，实现海量数据的统一管理。存储设备包括：光纤通道存储设备、IP 存储设备和 DAS 存储设备等。

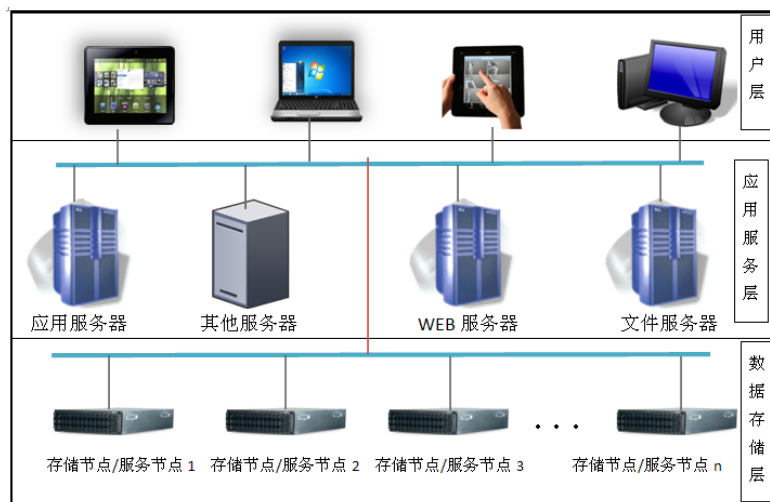
基础管理层是云存储的核心部分，也是最难以实现的部分。该层通过集群技术、分布式文件系统和网格计算技术等相关技术，实现系统中多个存储设备间的协同工作，并对外提供统一的服务标准，为存储系统提供更高效、更强大、更优越的数据访问性能打下基础。此层还具有内容分发系统、数据加密技术、数据备份和容灾技术和措施等保证了存储系统中的数据的安全性。

应用接口层是云存储系统中最灵活多变的的部分。不同的云存储运营单位可以根据实际业务类型，开发不同的应用服务接口，为不同的图书馆需求提供不同的应用服务。该层满足了云图书馆基于网络的跨平台的二次应用程序开发的要求，为图书馆建立特色平台，如：视频监控应用平台、IPTV 和视频点播应用平台、网络硬盘引用平台和远程数据备份应用平台等，提供了技术支持。

访问层中任何一个授权用户都可以通过标准的公用应用接口来登录数字图书馆云存储系统，享受相关服务。此时的数字图书馆馆员以及数字读者不再需要了解底层的硬件设备类型，也不需知道数字资源存放的物理位置，只需要在获得授权后，直接调用标准的 API，享受相关服务。此服务的使用不受读者所使用的终端设备限制。

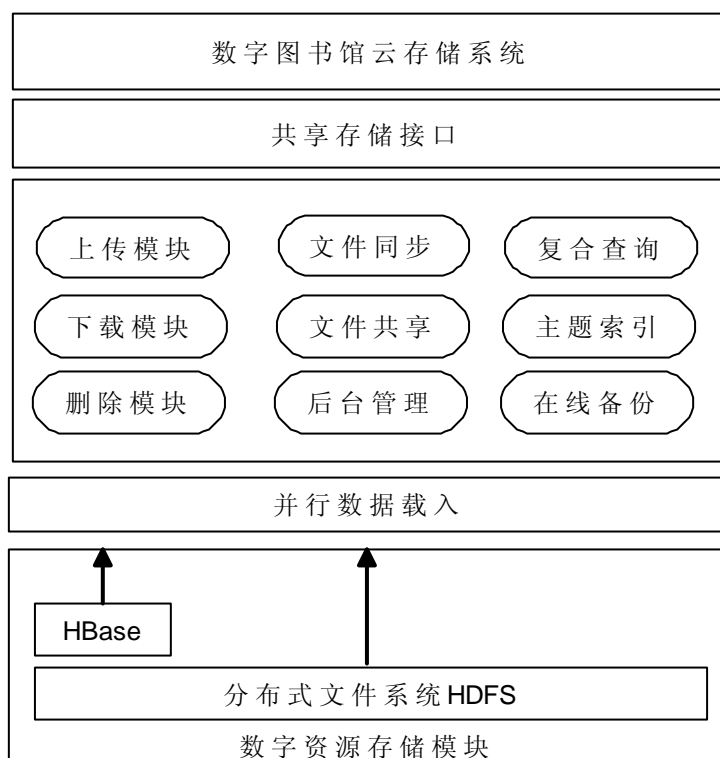
(2) 系统物理结构设计

本系统硬件节点部署分为三层：数据存储层、应用服务层和用户层。数据存储层由若干不同类型的存储节点或者服务节点组成，用于存储图书馆数字资源。应用服务层主要由具有不同功能的服务器组成，其中一部分服务器用于运行 NameNode、SecondNameNode、Hmaster 和 JobTracker 服务；一部分用于单独运行 ZooKeeper 服务；一部分用于运行 Katta 分布式索引检索服务；一部分用于运行 DataNode、TaskTracker 和 HregionServer 服务。用户层中的读者可以通过不同形式的平台使用该存储系统。



六、系统架构设计

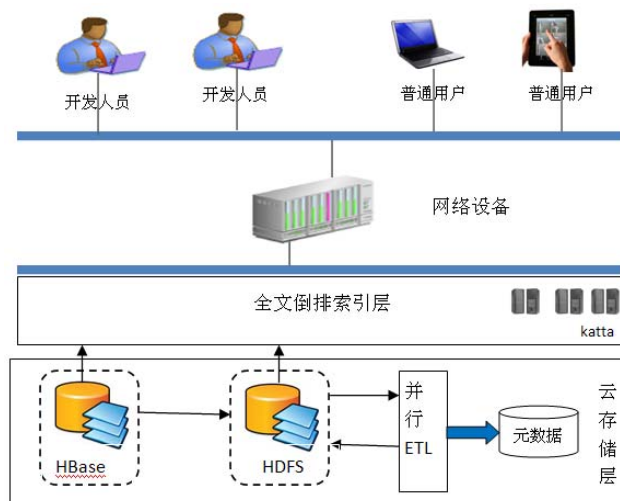
数字图书馆云存储系统以图书馆海量数字资源存储和共享为研究对象，以图书馆—读者为基本模式，从图书馆馆员、管理员和读者的不同角度设计该系统。系统包括数字资源存储模块、数据管理模块和索引检索模块三部分，其系统模块图如图 1。从图中可以看出，图书馆馆员将纸质版图书、期刊、论文、报纸等纸质资源，以及光盘等数字资源，通过传统手段录入到计算机中，形成数字资源。针对电子数据文件大小异构的特点，图书馆馆员通过数据存储模型，将视频、音频等大数据文件利用 HDFS 分布式文件系统存放；将图片、文档等较小的文件利用 HBase 直接存放。管理员利用数据管理模块，对数据文件进行上传，下载和删除等操作。做图书馆的服务对象，读者可以在 PC 机、手机、平板电脑等不同平台，利用系统提供的各种查询方式进行资源检索。



七、系统介绍

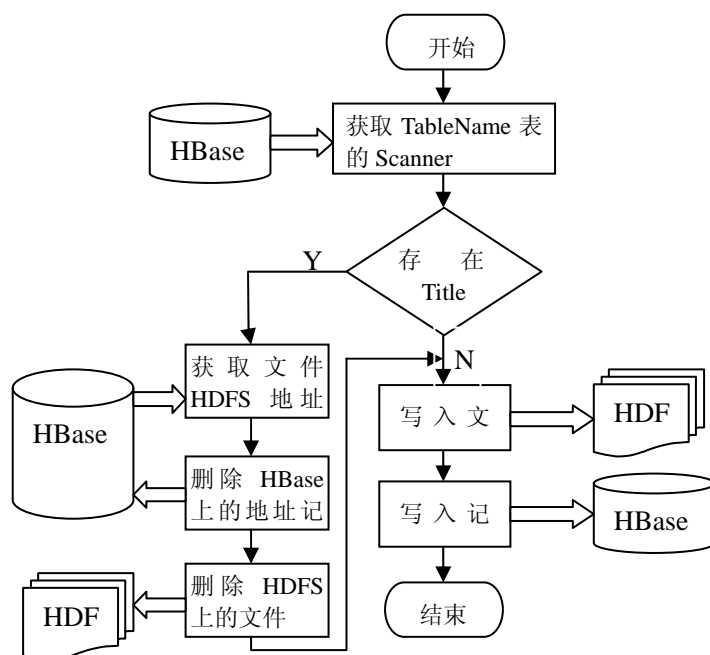
(1) 系统流程设计

系统对外提供图书馆馆员使用的管理接口和普通读者使用的查询接口。图书馆馆员借助于图形管理界面，实现数字资源的管理（上传、下载、删除等）操作。数据管理过程中，数据信息通过网络平台在当地计算机和存储集群之间的传递。数字资源按照类别，分别存储于 HBase 或直接存放于 HDFS。读者获得授权后，可以浏览当前数字资源，通过系统利用 HBase 自身的 RowKey 索引，进行精确查询、模糊查询和复合查询。



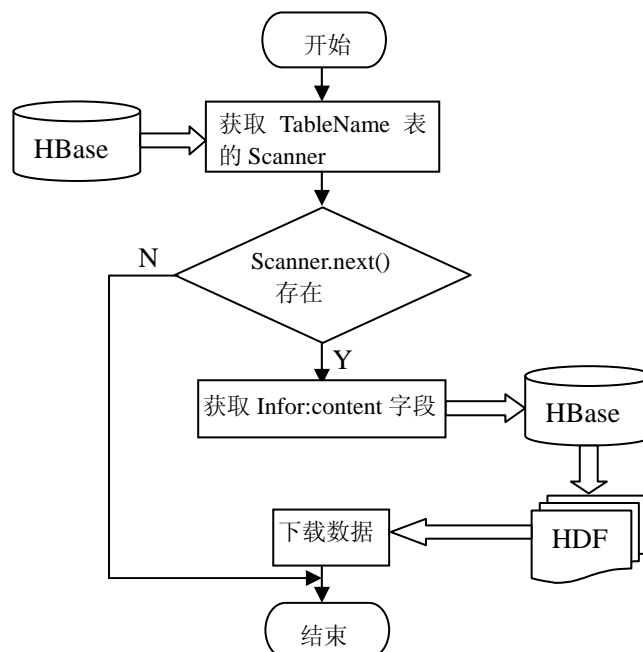
(2) 文件上传

数字资源的上传主要是对 HDFS 分布式系统和 HBase 数据库的操作。数字图书馆云存储系统根据数据文件大小进行了分类存储，对不同数据文件的上传进行了不同设计。图书馆馆员上传文件之前，先调用 HBase 数据库，获取 TableName 表的 Scanner，检查该文件是否存在。如果文件已经存在，则提示文件已经存在，是否需要更新。如果文件没有存在，则按照数据文件大小进行分类存储。



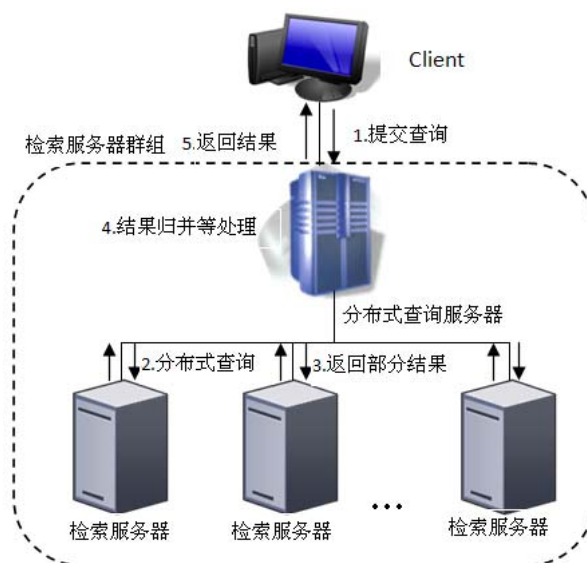
(3) 文件下载

文件的下载模块也是根据文件大小采用不同的方式进行下载的。对于相对较小的数据文件直接获取数据库 HBase 数据表 Infor 中的 content 字段中的内容，而对于相对较大的数据文件需要获取 Infor 中的 content 字段中的值后，从分布式系统 HDFS 中下载文件内容。



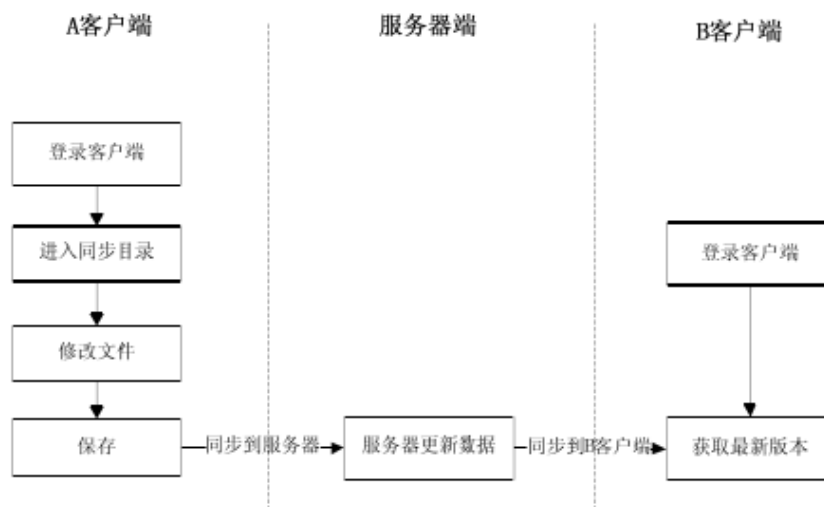
(4) 索引模块

针对文本文件内容的检索相对比较复杂，本系统通过分布式检索工具 Katta，利用分布式倒排索引 Lucene，将索引进行分片存储加以实现，检索流程图如图 6 所示。客户端首先向检索服务器群组中的分布式查询服务器发送查询请求，由分布式查询服务器根据检索服务器集群中设备的当前状态将查询任务发送，然后检索服务器将检索结果返回给分布式查询服务器，用户最终得到查询结果，完成查询任务。



(5) 文件同步

每个成功登陆系统的客户端，会自动更新文件状态，并提醒是否同步（可设定）。同步过程是增量备份过程，并在服务器保留一定数量的备份。同步是在后台多线程进行的，用户不需要干预，对于同步状态有四种：正在同步、无法同步（文件给其他程序锁定等原因）、完成同步、排除同步（文件超出限制）。



(6) 文件备份

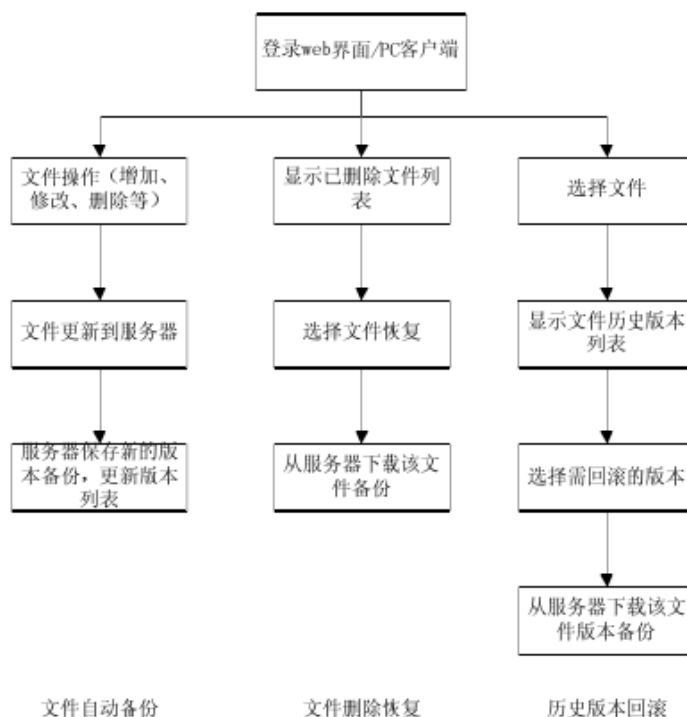
实现目前备份技术的四个主要功能，文件备份的副本数目以及时效长短，由不同的产品类型来决定，在后台管理可以控制。

1) 文件自动备份：对 PC 客户端文件的任何操作，将更新至服务器内容中，并后台进行文件的同步备份。

2) 文件删除回复：对于误删除的文件，提供类似回收站功能，实现在线文件备份恢复。

3) 文件历史版本回滚：根据用户选择的不同套餐，提供指定备份数量的历史版本，可以自由选择回滚。

4) 文件时光机：对用户指定的文件，自动快捷恢复到最近一次的备份。



八、算法设计

本系统采用编程算法设计，使用最多的是逻辑结构的判断语句，还有顺序结构，还有当型循环结构。本系统的算法简单，方便操作，容易上手，方便操作工人的操作。

当型循环算法的算法程序实例如下：

```
var i, k: integer;
begin
  i=10;
  k=0;
  while (i>=10) and(i<=30) do
  begin
    k=k+i
    i=i+2
  end;
  writeln(s:5, a:5, n:5);
end
```

九、网络带宽的设计

网络带宽的需求如下表所示：

表 2 网络带宽需求列表

服务器	并发数	每并发需要带宽(K)	共需带宽	计算 30%冗余
Web 服务器	900	50	45000	64286
WAP 服务器	400	50	20000	28571
外链共享服务器	500	100	50000	71429
断点续传服务器	1500	100	150000	214286

十、测试计划

本系统采用单元测试的方法，以文件查找模块为例。文件查找模块的功能是能够根据给定的文件名，检索系统当前存储的所有数据，如果该文件已存在，返回 True，否则返回 False。需要注意的是，用户必须指定所要查找的文件类型，

否则，由于不同类型的文件可能重名，无法判断是否存在。在上传、下载、删除某个文件之前，都需要使用文件查找模块判定该文件是否已经存在，然后才可以做进一步的处理。

文件查找模块的测试用例及结果如表 3 所示。给出需要查找的文件名及类型，该模块即会检索系统所存储的数据，返回是否存在的判定结果。这里的类型，是通过文件的扩展名来识别的。目前，系统内置可以识别的图片文件类型包括：jpg、bmp、gif；文本类型包括：txt、doc；视频文件类型包括：rmb、avi 等。如果查找时，未指定类型，指示“指定类型”。如果指定的文件类型超出上述范围，则提示“无效类型”。

表 3 查找模块测试用例表

测试用例	测试要求	输入	操作步骤	预期输出
查找图片文件	图片已存在，输出 true	Image_1.jpg		输出 true
	图片不存在，输出 False	Image_2.jpg	SearchType	输出 False
	必须输入文件类型	Image_1	选择“file”，	未指定类型
	文件类型必须满足要求	Image_1.err	在“Search	无效类型
查找文本文件	文件已存在，输出 true	Text_1.txt	Exp”中输入值，点	输出 true
	文件不存在，输出 False	Text_2.txt	击“Search”	输出 False
查找视频文件	视频已存在，输出 true	Video_1.rmb		输出 true
	视频不存在，输出 False	Video_2.rmb		输出 False

十二、系统维护设计

软件本身操作简单，占用资源少，维护只需对所用文件夹进行定期备份处理，以确保数据不必要丢失。

十三、系统功能实现

1、登录界面的实现

浏览器中输入访问地址，便可以实现访问，进入登录页面。输入用户名和密码后，系统进行匹配，输入正确，便可以查看和管理数字图书馆云存储系统了。



2、账户管理

系统管理员有权对系统用户进行数据管理，包括：增加、修改、查询和删除操作。在界面中，管理员可以点击增加、删除、启用/停用等相关按钮，完成操作，还可以获取账户对应的 AccessKey 和 SecretKey。

账户标识: 账户状态:

账户标识	账户名称	邮箱地址	账户状态	申请容量	已用容量
1111	testdongdong	xxx@xxx.xxx	已启用	2096MB	18MB
47442	47442		已启用	100MB	0MB
47588	47588		已启用	1MB	0MB
48285	whynot		已启用	100MB	0MB
48415	48415		已启用	1MB	0MB
guofeng	guofeng		已启用	100000MB	3306MB
helliming	helliming		已启用	10240MB	9MB
jiazy29	jiazy29		已启用	100000MB	133MB
john DOE	John Doe	john@example.com	已启用	1024MB	0MB
lichanghao	lichanghao	lichanghao@gcidesign.com	已启用	1024MB	0MB

3、日志管理

点击“日志管理”，可以查看系统的日志，实施对系统登录信息和数据的操作查看，完成系统监督任务。这个功能是保证系统安全的一项重要举措。

日志内容: 日志类型: 操作用户登陆IP:

日志表ID	日志内容	日志类型	用户登陆IP	用户ID	用户名称	创建时间
1	用户【admin】登陆系统	登陆	172.16.13.201	1	admin	2013-11-11 22:34:20
2	修改服务器信息操作, 对象名称: ServiceInfo	修改	172.16.13.201	1	admin	2013-11-11 22:35:30
3	修改服务器信息操作, 对象名称: ServiceInfo	修改	172.16.13.201	1	admin	2013-11-11 22:36:17
4	修改服务器信息操作, 对象名称: ServiceInfo	修改	172.16.13.201	1	admin	2013-11-11 22:37:01
5	添加服务器信息操作, 对象名称: ServiceInfo	新增	172.16.13.201	1	admin	2013-11-11 22:37:44
6	添加服务器信息操作, 对象名称: ServiceInfo	新增	172.16.13.201	1	admin	2013-11-11 22:38:18
7	添加服务器信息操作, 对象名称: ServiceInfo	新增	172.16.13.201	1	admin	2013-11-11 22:39:38
8	添加服务器信息操作, 对象名称: ServiceInfo	新增	172.16.13.201	1	admin	2013-11-11 22:40:16
9	新增文件系统用户操作, 新增用户信息: UserId:	新增	172.16.13.201	1	admin	2013-11-11 23:11:15
10	新增文件系统用户操作, 新增用户信息: UserId:	新增	172.16.13.201	1	admin	2013-11-11 23:26:17

4、对象管理

数字图书馆云存储系统的对象主要是图书馆馆员需要放在存储空间上管理的数据或文件。云盘系统中的对象功能提供了基于 Web 端和手机端的文件操作功能，包括文件上传、下载、删除和分享，以及文件夹的创建、删除和修改等。文件以对象存储方式存放在云端，支持随时随地共享文件。图书馆馆员登录进入系统后，选择要进入的存储器，点击进入，便可以查看该存储器下的所有对象。对象根据文件夹来分层显示。

bucket列表 全部文件 上一步 当前位置: 888888 >

序号	文件名称	文件大小	文件类型	操作
1	我的测试	0 B	filepackage	
2	002.JPG	1.61 MB	JPG	<input type="button" value="分享"/> <input type="button" value="下载"/>
3	24854	245.23 KB	24854	<input type="button" value="分享"/> <input type="button" value="下载"/>
4	artifacts.xml	100.64 KB	xml	<input type="button" value="分享"/> <input type="button" value="下载"/>
5	edipse.exe	184.00 KB	exe	<input type="button" value="分享"/> <input type="button" value="下载"/>
6	image.jpg	114.31 KB	jpg	<input type="button" value="分享"/> <input type="button" value="下载"/>
7	谢霆锋 - 估计亚洲.mp3	5.99 MB	mp3	<input type="button" value="分享"/> <input type="button" value="下载"/>

5、文件上传

文件支持多线程多文件同时上传功能。



6、文件下载

点击下载按钮，可以实现文件本地下载功能。



7、文件共享

提供了多种文件共享的方式。用户可以根据实际应用，选择自己的方式。



8、系统主界面

普通读者登录系统后，首先进入主界面。可以点击相关菜单完成相应任务。



9、文件查询

本系统为读者提供了多种查询方式。读者进入查询界面，完成查询任务。



附录十一

```
import com.lzw.login.Login;
public class JXCFrame {
    private JPanel sysManagePanel;
    private JDesktopPane desktopPane;
    private JFrame frame;
    private JLabel backLabel;
    // 创建窗体的 Map 类型集合对象
    private Map<String, JInternalFrame> ifs = new HashMap<String,
JInternalFrame>();
    public JXCFrame() {
        frame = new JFrame("数字图书馆云存储系统");
        frame.getContentPane().setBackground(new Color(170, 188, 120));
        frame.addComponentListener(new FrameListener());
        frame.getContentPane().setLayout(new BorderLayout());
        frame.setBounds(100, 100, 800, 600);
        frame.setDefaultCloseOperation(JFrame.EXIT_ON_CLOSE);
        backLabel = new JLabel();// 背景标签
        backLabel.setVerticalAlignment(SwingConstants.TOP);
        backLabel.setHorizontalAlignment(SwingConstants.CENTER);
        updateBackImage(); // 更新或初始化背景图片
        desktopPane = new JDesktopPane();
        desktopPane.add(backLabel, new Integer(Integer.MIN_VALUE));
        frame.getContentPane().add(desktopPane);
        JTabbedPane navigationPanel = createNavigationPanel(); // 创建导航标签面
板
        frame.getContentPane().add(navigationPanel, BorderLayout.NORTH);
        frame.setVisible(true);
    }
    public static void main(String[] args) {
        SwingUtilities.invokeLater(new Runnable() {
            public void run() {
                new JXCFrame();//Login();
            }
        });
    }
    package com.wsy.struts.util;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.Statement;

import javax.sql.DataSource;
```

```
import com.wsy.struts.datasource.DB;

public class DownTable {

    int totalPages=1;
    int pageSize=5;
    int currentPage=0;
    ResultSet rs=null;
    Statement s=null;
    int totalRows;
    DB d;
    public DownTable(DataSource datasource) {
        d=new DB(datasource);
    }
    public int getTotalPage() {
        if(getRows()%getPageSize()==0)
            return getRows()/getPageSize();
        else
            return getRows()/getPageSize()+1;
    }
    public void setpageSize(int size) {
        this.pageSize=size;
    }
    public int getPageSize() {
        return pageSize;
    }
    public void setCurrentPage(int current) {
        this.currentPage=current;
    }
    public int getCurrentPage() {
        return currentPage;
    }
    public int getRows()
    {
        try{

            rs=d.OpenSql("select count(*) from tb_forum");
            if(rs.next())
            {
                totalRows=rs.getInt(1);
            }
        }
    }
}
```

```
        s.close();
        d.close();
    }
    catch(Exception e)
    {
    }
    return totalRows;
}
// 普通JDBC的连接
/*public static Connection getConnection1() {
    int icon=0;
    try{
        Class.forName("oracle.jdbc.driver.OracleDriver");
        do{
            conn=DriverManager.getConnection(ORACLECON, USER, PWD);
            if(conn==null) Thread.sleep(1000);
            //icon++;
            //System.out.println(icon);
        }while(conn==null&&icon<10);
        //logger.debug("==== 获得数据库资源 =====");
    }catch(Exception e){
        e.printStackTrace();
        try {
            if (conn != null)
                conn.close();
        } catch (Exception ex1) {
        }
        conn = null;
        return null;
    }
    return conn;
}
//连接池的连接
/*public static Connection getConnection2() {
    int iconn = 0;
    try {
        do {
            conn = DriverManager.getInstance().getConnection("cmsdb");
            if (conn == null)
                Thread.sleep(1000);
        } while (conn == null && iconn <= 10);

        if (conn == null)
            return null;
```

```
        //logger.debug("==== 获得数据库资源 =====");

    } catch (Exception ex) {
        ex.printStackTrace();
        try {
            if (conn != null)
                conn.close();
        } catch (Exception ex1) {
        }
        conn = null;
        return null;
    }
    return conn;
}

public static void main(String args[]) {
    DownTable d=new DownTable();
    d.setCurrentPage(0);
    d.setpageSize(3);
    System.out.println("rows"+d.getRows());
    System.out.println("current"+d.getCurrentPage());
    System.out.println("total"+d.getTotalPage());
}*/
}

    return tabbedPane;
}
/** *****辅助方法***** */
// 为内部窗体添加 Action 的方法
private JButton createFrameButton(String fName, String cname) {
    String imgUrl = "res/ActionIcon/" + fName + ".png";
    String imgUrl_roll = "res/ActionIcon/" + fName + "_roll.png";
    String imgUrl_down = "res/ActionIcon/" + fName + "_down.png";
    Icon icon = new ImageIcon(imgUrl);
    Icon icon_roll = null;
    if (imgUrl_roll != null)
        icon_roll = new ImageIcon(imgUrl_roll);
    Icon icon_down = null;
    if (imgUrl_down != null)
        icon_down = new ImageIcon(imgUrl_down);
    Action action = new openFrameAction(fName, cname, icon);
    JButton button = new JButton(action);
    button.setMargin(new Insets(0, 0, 0, 0));
    button.setHideActionText(true);
    button.setFocusPainted(false);
}
```

```
        button.setBorderPainted(false);
        button.setContentAreaFilled(false);
        if (icon_roll != null)
            button.setRolloverIcon(icon_roll);
        if (icon_down != null)
            button.setPressedIcon(icon_down);
        return button;
    }
    // 获取内部窗体的唯一实例对象
    private JInternalFrame getIFrame(String frameName) {
        JInternalFrame jf = null;
        if (!ifs.containsKey(frameName)) {
            try {
                Class fClass = Class.forName("internalFrame." + frameName);
                Constructor constructor = fClass.getConstructor(null);
                jf = (JInternalFrame) constructor.newInstance(null);
                ifs.put(frameName, jf);
            } catch (Exception e) {
                e.printStackTrace();
            }
        } else
            jf = ifs.get(frameName);
        return jf;
    }
    // 更新背景图片的方法
    private void updateBackImage() {
        if (backLabel != null) {
            int backw = JXCFrame.this.frame.getWidth();
            int backh = frame.getHeight();
            backLabel.setSize(backw, backh);
            backLabel.setText("<html><body></img></body></html>");
        }
    }
    // 窗体监听器
    private final class FrameListener extends ComponentAdapter {
        public void componentResized(final ComponentEvent e) {
            updateBackImage();
        }
    }
    // 主窗体菜单项的单击事件监听器
    protected final class openFrameAction extends AbstractAction {
```

```
private String frameName = null;
private openFrameAction() {
}
public openFrameAction(String cname, String frameName, Icon icon) {
    this.frameName = frameName;
    putValue(Action.NAME, cname);
    putValue(Action.SHORT_DESCRIPTION, cname);
    putValue(Action.SMALL_ICON, icon);
}
public void actionPerformed(final ActionEvent e) {
    JInternalFrame jf = getIFrame(frameName);
    // 在内部窗体关闭时，从内部窗体容器 ifs 对象中清除该窗体。
    jf.addInternalFrameListener(new InternalFrameAdapter() {
        public void internalFrameClosed(InternalFrameEvent e) {
            ifs.remove(frameName);
        }
    });
    if (jf.getDesktopPane() == null) {
        desktopPane.add(jf);
        jf.setVisible(true);
    }
    try {
        jf.setSelected(true);
    } catch (PropertyVetoException e1) {
        e1.printStackTrace();
    }
}
}
static {
    try {
        UIManager.setLookAndFeel(UIManager.getSystemLookAndFeelClassName());
    } catch (Exception e) {
        e.printStackTrace();
    }
}
}
```