

# 基于云计算的数字图书馆公共服务平台架构<sup>①</sup>

毛雪岷, 雷玲玲, 李 静

(合肥工业大学 管理学院, 合肥 230039)

(教育部过程优化与智能决策重点实验室, 合肥 230039)

**摘 要:** 数字图书馆公共服务平台为数字化全民阅读提供了一个有效的途径, 为广大用户提供大量数字版资源的在线阅读、下载和分享等. 随着数字化信息资源的爆炸式增长, 如何有效的解决海量数字图书馆信息资源的存储和处理, 并能将不同区域和类型的数字图书馆资源集成起来成为研究的热点. 依据安徽教育出版社的数字图书馆公共服务的全民阅读平台, 研究云计算下数字图书馆的平台架构, 利用云计算的分布式存储和并行计算的特点, 为出版社的数字图书馆建设提供可行性研究方案.

**关键词:** 云计算; 数字图书馆; 分布式存储; 公共服务; 平台

## Public Service Platform of Digital Library Based on Cloud Computing

MAO Xue-Min, LEI Ling-Ling, LI Jing

(School of Management, Hefei University of Technology, Hefei 230039, China)

(Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei 230039, China)

**Abstract:** Digital library has become an effective platform of national reading, people can read on line, download or shared digital resources and so on. With the explosive growth of digital information resources, how to effectively solve the massive digital library information storage and process resources, meanwhile make resources of digital libraries that lie in different regions and have different types integrated has become a hot research. This paper based on universal reading platform of Anhui Education Press digital library of public services, research the architecture platform of digital library based on cloud computing, in use of distributed storage and parallel computing characteristics of cloud computing, which providing a feasibility study for the construction of digital library of the Press.

**Key words:** cloud computing; digital library; distributed storage; public service; platform

随着计算机网络的普及和 Web2.0 的发展, 使得图书馆信息资源逐步实现了数字化, 然而在现行的服务模式, 昂贵的硬件扩展、软件系统不断的升级和维护、信息资源之间访问的封闭、数据的安全隐患等都严重制约了数字图书馆的快速发展. 云计算作为一种分布式处理技术可以很好的解决上述难题, 其基本原理是“数据计算被分布在许许多多的分布式计算机上, 用户根据自己的需要访问计算机和存储系统”<sup>[1]</sup>, 实际上 Web2.0 在图书馆中的应用包含大量云计算的成分<sup>[2]</sup>.

## 1 云计算定义与体系

由于云计算性能的扩张性, 现在还无法给云计算统一的定义, 维基百科中的定义为: 是一种基于互联网的计算方式, 通过这种方式, 共享的软硬件资源和信息可以按需求提供给计算机和其他设备. 云计算是继 1980 年代大型计算机到客户端-服务端的大转变之后的又一种巨变<sup>[3]</sup>. 云计算描述了一种基于互联网的新的 IT 服务增加、使用和交付模式, 通常涉及通过互联网来提供动态易扩展而且经常是虚拟化的资源.

Lamia Youseff 等人在 Toward a Unified Ontology

<sup>①</sup>基金项目:国家自然科学基金(70801025)

收稿时间:2014-01-23;收到修改稿时间:2014-02-24

of Cloud Computing 文献中提出了一种五层体系结构, 如图 1 所示:

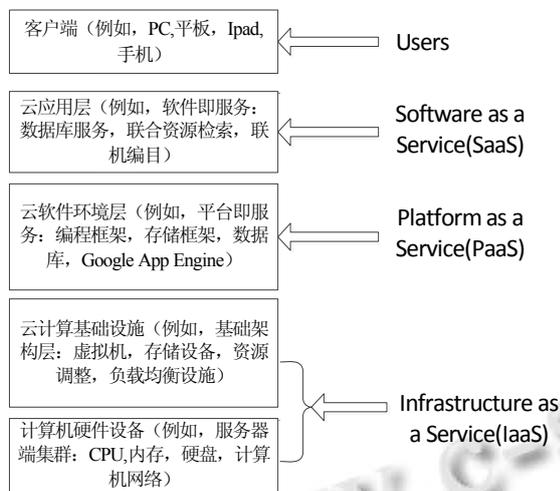


图 1 云计算五层体系结构

## 2 构建云计算下数字图书馆的关键技术

云计算是网格计算、分布式计算、并行计算、效用计算、网络存储、虚拟化、负载均衡等传统计算机技术和网络技术发展融合的产物, 其技术实质是计算、存储、服务器、应用软件等 IT 软硬件资源的虚拟化, 云计算在虚拟化、数据存储、数据处理、编程模式等方面具有自身独特的技术, 主要关键技术<sup>[4]</sup>有: ①虚拟化技术, 是通过新增的虚拟中间层截获上层软件对底层接口的调用, 并对该调用重新做出解释和处理, 以实现异构环境中资源的可共享、可管理和可协同, 同时支持应用大规模部署、迁移和运行维护. ②分布式数据存储技术, 采用可扩展的系统结构, 利用多台服务器分担存储负荷, 要求存储资源能被抽象表示和统一管理, 并且保证存储资源的读写操作安全、可靠等各方面的要求. ③海量数据管理技术, 对于海量数据, 云计算提出了基于 Map-Reduce 的编程模型, 来满足海量数据处理的性能和可靠性等方面的要求. ④分布式计算和并行计算, 分布式计算将大的问题分解为许多小的子问题, 然后将这些小的子问题分配给多台计算机(通常为一个集群)进行处理, 最终将计算得到的结果进行收集综合而得到大问题的结果. 并行计算和分布式计算思想基本类似, 分布式计算强调的是空间, 而并行计算强调的是时间的同步. 目前并行模式可粗

略归纳为 3 类, 即共享存储模式、分布存储模式、共享存储与分布存储混合模式. 相应的程序设计也可以归为 3 类, 即共享程序设计(如 OpenMP)、基于消息传递程序设计(如 PVM 和 MPI)和混合编程模式(如 MPI+OpenMP).

## 3 基于云计算的数字图书馆系统设计

### 3.1 基于云计算的数字图书馆体系架构

按照从底到上的设计思路, 技术架构主要包括: 图书馆实体层、虚拟化层、分布式集群层、数据处理接口层、服务应用平台层和客户端层.

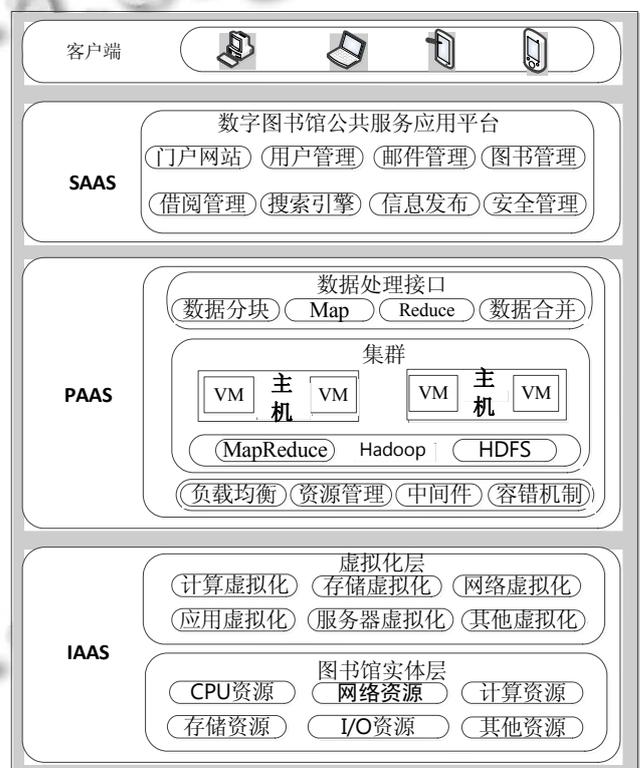


图 2 数字图书馆的体系架构

### 3.2 数字图书馆系统功能结构

数字图书馆系统主要包括四大功能模块: 数图构建、电子书阅读、电子书借阅和后台管理. 数图构建包括: 图书分类、创建机构馆、站点镜像发布和数据同步; 电子书阅读包括: 在线阅读和下载到客户端离线阅读; 电子书借阅包括: 借阅流程、U 盘借阅和移动硬盘借阅; 后台管理包括: 电子书管理、副本数管理、权限管理和用户身份管理.

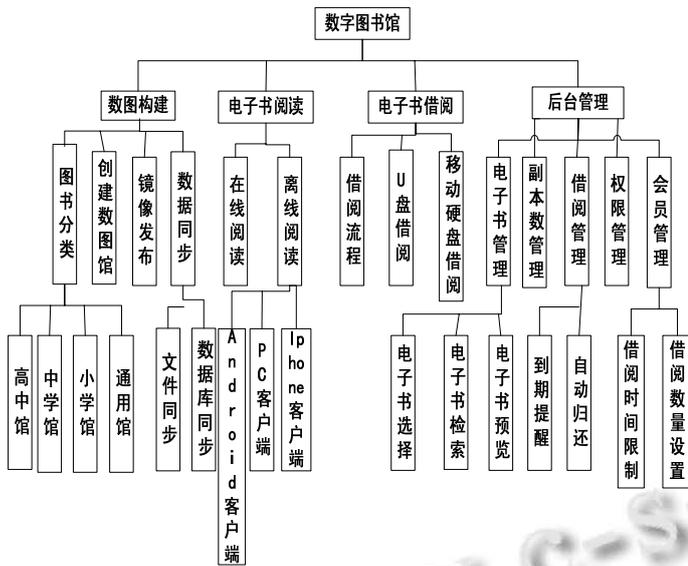


图 3 数字图书馆的系统功能结构

### 3.3 系统主要功能说明

#### 3.3.1 构建图书分类

构建符合要求的图书分类，图书分类按照中国图书分类法进行分类，提供给数字图书馆使用。

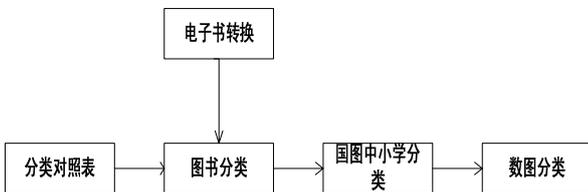


图 4 图书分类示意图

#### 3.3.2 构建机构数字图书馆

用户可根据需要，可以选择数图的版式，如新春特别版、书香版、科技版等，创建时还需选择运作模式并设定相应参数：是否限制副本数选择及设定副本数；是否限制会员数选择及设定会员数；是否限制时间选择及设定有效期。然后填写机构基本信息并提交，后台管理员接到申请后进行审核。如果审核通过，则创建机构图书馆；如果不通过，需要说明原因，更改后可重新提交。如下图 5 所示数字图书馆创建流程图：数字图书馆可以根据自己的需要，选取适合用户阅读的电子图书，选书过程支持电子书的缩略图查询、目录标签分类查询、关键字搜索查询，选取图书的过程还要支持电子书的预览功能，方便对图书内容的了解。

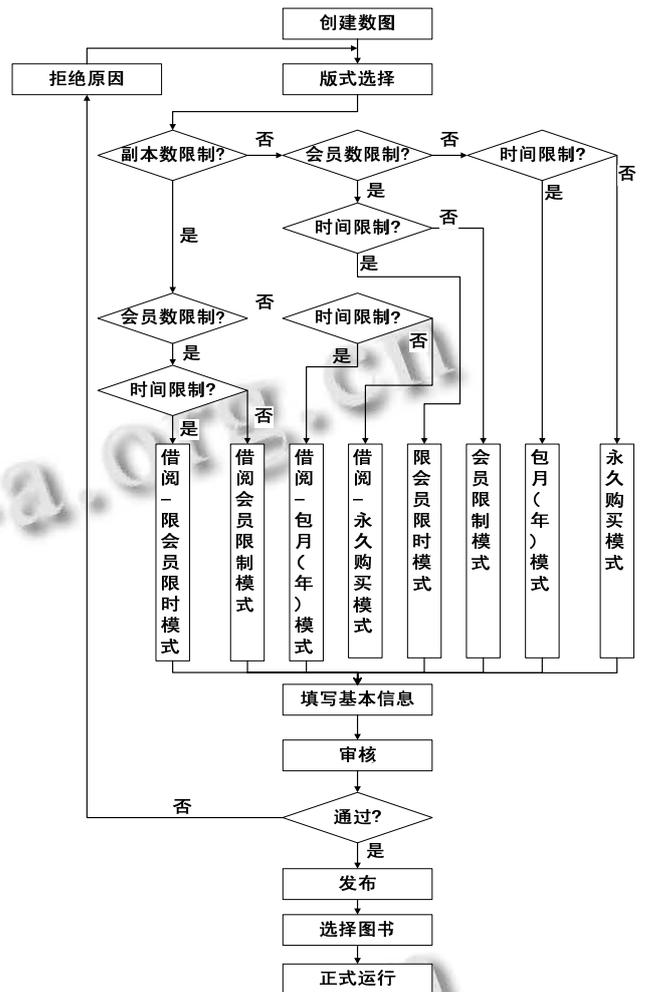


图 5 数字图书馆创建流程图

运作模式划分如下图 6 所示，主要考虑因素：副本数、会员数、时间。

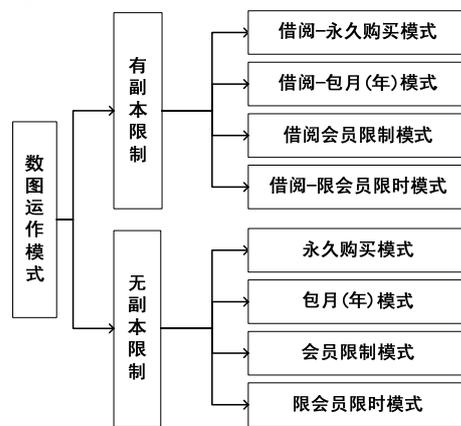


图 6 数字图书馆运作模式

数字图书馆运作模式首先分为两大类：借阅模式(有副本限制)和无副本限制模式。

1)、在借阅模式下,同一本书只能有限的几个会员借阅.跟现实图书馆运作方式一样,全面符合图书馆的借阅流程,管理员可以在后台设定数字图书的自动归还时间.学生可以自主归还图书,到期应该归还的图书,系统将自动归还,同时释放借书的额度.所有具有权限的用户看书都需要走借阅流程,且支持续借和预约功能.借阅后可支持在线阅读、客户端下载离线阅读和 U 盘(移动硬盘)下载阅读.

借阅-永久购买模式,不限制时间,不限会员数,但因副本数有限,所以只允许有限的几个会员同时借阅同一本书;

借阅-包月(年)模式,不限会员数,但限制时间.即在有效期内,只允许有限的几个会员同时借阅同一本书;

借阅-会员模式,不限制时间,但限制会员数.跟永久购买模式一样,但机构的会员人数限制被限制;

借阅-限会员限时模式,限制时间且限制会员数.跟包月(年)模式一样,但机构的会员人数限制被限制.

2)、在无副本限制模式下,机构所选取的电子书是不限制副本数的,即只要有权限的用户,都可以在线阅读机构图书馆里的电子书,且无需走借阅流程,也能下载离线阅读.

永久购买模式,不限制时间,不限会员数,所以只要满足权限的账号或 IP 都可以同时阅读同一本书或多本书;

包月(年)模式,不限会员数,但限制时间.即在有效期内,只要满足权限的账号或 IP 都可以同时阅读同一本书或多本书;

会员限制模式,不限制时间,但限制会员数.跟永久购买模式一样,但机构的会员人数限制被限制,只有使用账号登录才能阅读,使用 IP 登录只能预览;

限会员限时模式,限制时间且限制会员数.跟包月(年)模式一样,但机构的会员人数限制被限制,只有使用账号登录才能阅读,使用 IP 登录只能预览.

### 3.3.3 镜像发布

如果学校数字图书馆用户过多,同时在线阅读时会严重增加服务器的负担和网络通信负担,影响服务质量.根据学校机构需要,数字图书馆可以选择部署在学校局域网,生成局域网站点的站点文件和数据库文件,生成子站点镜像,通过光盘独立发布到学校局域网,局域网的图书馆具有总站点的图书馆的所有功

能,并且局域网服务器定时将用户的数据发送在总站平台,实现二级交互,跟服务器之间只需少量的流量通信和定期的数据同步.

### 3.3.4 数据同步

数据同步的对象有主站总服务器数据同步和各个分布学校的子站数据同步,同步的内容有文件内容同步,通过文件清单.xml 记录传输内容和检测依据.数据同步是通过数据库管理软件的复制功能完成数据同步,复制是将一组数据从一个数据源拷贝到多个数据源的技术,是将一份数据发布到多个存储站点上的有效方式.利用数据同步技术确保分布在不同学校的数据自动同步更新,从而确保数据的一致性.

### 3.3.5 在线阅读

在线阅读,纯 WEB 方式的阅读,会员无需安装插件或者客户端,方便会员简单操作,阅读图书时,具有大纲及阅读、放大缩小、电子书签、高亮标记、键盘翻页、全文检索、划词翻译、书签定位等功能.

### 3.3.6 借阅流程

全面符合图书馆的借阅流程,管理员可以设定被借阅的数字图书的自动归还时间.数字图书馆可以根据自己的需要,选取适合用户阅读的电子图书,选书过程支持电子书的缩略图查询、目录标签分类查询、关键字搜索查询,选取图书的过程还要支持电子书的预览功能,方便对图书内容的了解.电子图书的借阅过程要符合图书馆的借阅流程;学生可以自主归还图书,到期应该归还的图书,系统能够自动归还,同时释放借书的副本数.

### 3.3.7 U 盘及移动硬盘借阅

用户把 U 盘查到电脑上,选中某本图书,通过借阅功能,自动将图书下载到 U 盘上.如果用户未登陆,弹出框输入用户名和密码后,选择 U 盘路径后开始下载图书.

### 3.3.8 电子书管理

数字图书馆可以根据自己的需要,选取适合读者阅读的电子图书,选书过程支持电子书的缩略图查询、目录标签分类查询、关键字搜索查询,选取图书的过程还要支持电子书的预览功能,方便对图书内容的了解.

### 3.3.9 副本数管理

将副本数作为一个可变化的参数,可设置每本书的借阅副本数,支持批量设置功能.电子书副本数用

来限制可同时阅读某一本书的读者数量。

### 3.3.10 权限与会员管理

对于不同的用户,比如学生和老师,系统可以设定不同的借阅权限。可设置不同用户身份的借阅数量、借阅时间等。数字图书馆平台系统为实现相应的安全访问控制机制,采用了目前国际上比较流行的基于角色的访问控制方法(Role-based Access Control, RBAC)设计。RBAC模型<sup>[5]</sup>其原理是以角色为中间媒介建立用户与权限的联系,克服了直接像用户授权的不足。RBAC实现模型控制流程大体为:用户登录,首先进行身份认证;认证通过后,根据用户身份进行角色与权限的关联;关联成功后,根据权限配置系统操控项并动态生操控界面。系统设置专门的接口用于对流程中的异常情况进行处理。

## 4 结语

近年来,通过不断创新和改革服务理念、服务模式及服务手段,基于云计算的数字图书馆集成服务平

台在界面设计、功能模块和系统优化三方面的建设,取得了较多成果。本文提供了基于云计算的云图书馆平台的系统架构设计,本研究可为未来的云计算数字图书馆平台结构设计和具体实现提供借鉴。

### 参考文献

- 1 钟盈寿.云计算环境中的数字图书馆架构探讨.图书情报工作,2010,(2):276-278.
- 2 胡小青,范并思.云计算给图书馆管理带来挑战.大学图书馆学报,2009,(4):7-12.
- 3 Danielson K. Distinguishing Cloud Computing from Utility Computing. Ebizq.net. [2010-08-22].
- 4 王平.云计算关键技术数字图书馆中的应用研究.报资料工作,2010,(5):52-56.
- 5 Powell AJ, Lyon L. The DNER technical architecture: Scoping the information environment. <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/dner-arch.html>. [2011-07-17].