

云计算在数字图书馆中的应用研究

陈清华 浦洁

江苏省社会科学院

摘要 数字图书馆逐步解决了传统图书馆基础网络设施等硬件建设的高成本、传统架构模式的资源分享效率低和存储量小等一系列缺陷。云计算可以利用传统模式无法比拟的低成本与超大容量等优势，将图书馆数据进行转移，从而有效地解决了传统图书馆运行存在的一系列问题。本文在描述云计算及其优势的基础上，对云计算技术给图书馆信息构建带来的变化与问题进行了分析，研究了创建数字图书馆过程中的云计算应用问题。

关键词 云计算；数字图书馆；云存储基础框架；云服务模式

1 研究背景

数字图书馆处理各种各样的数据和信息，向用户提供快捷、便利的服务。但是传统意义上各自分离的数字图书馆拥有着属于自身的独立信息源，这种状况阻碍了数字图书馆的发展。随着分布式计算技术的发展，并行计算、网格计算、效用计算以及云计算等都已经十分普及，随之出现了类似于基础设施即服务（IaaS）、平台即服务（PaaS）和软件即服务（SaaS）等多种数字服务技术^①。云计算的特点在于其按需获取、多样性、扩展性和虚拟能力等方面的服务^②。同时，其在资源利用与管理方面的低成本与高效益的特点也必然会为图书馆的管理与服务领域带来深层次的改革与大范围的影响。

2 数字图书馆建设现状

2.1 基础设施架构的高成本

基础设施架构的成本和维护费高、利用率低是目前数字图书馆建设普遍存在的问题。传统图书馆 IT 改造基础设施的构架包括购买不间断电源、防火墙、交换机、服务器和存储器等各种设备，并且需要将软硬件进行搭建、运营、维护，而后期的这一笔费用是相当高的。同时，这些设备的性能与 IT 行业的发展共同导致了它们的生命活跃周期很短暂。即便是使用不太先进的设备，购买和扩展新的模块配件也要需要许多花费。

2.2 信息资源存储技术问题

信息资源的存储技术也是一个不得不面临的问题。在数字图书馆中，获取多规则开发语言的计算机科学技术是十分重要的。然而如今不同数据标准、不同软件或不同模块工具之间的数据交换技术已经更新地十分迅速，数据体量的爆炸式增长、标准规则的多应用拓展要求数字图书馆应该具备更强大、更灵活的资源存储设施，从而拥有可伸缩的存储资源空间、灵活可配的性能管理以及高可靠性的资料备份能力^③。

2.3 资源共享效率低

数字图书馆打破了空间与时间的限制。为了保证服务设备的正常功能与数据资源的安全性，传统图书馆对访问终端服务器的流量与最大支撑数量进行了控制，这样的做法不仅导致了服务器的访问限制，而且影响了图书馆之间的数据传输吞吐能力、压力承载能力与资源共享程度。

3 基于云计算的图书馆

3.1 云计算有效地解决了基础设施构架存在的高成本问题

云计算改变了信息服务的供应方式。在“云”中的存储资源、基础设备是可以无限制扩展的，通过云计算技术的利用，客户设备可以得到低成本化与资源绿色化的改善。云计算通过虚拟技术，利用虚拟软件融合成百上千的服务器，对外以统一、标准的平台服务用户。另外云计算是按需收费，在性能需求不变的情况下，软硬件更新迭代是无需额外花费的。并且云计算采用的“云端托管”模式，从开发、运营到后期维护，都提供成熟、低成本的服务方案，相较于自行项目建设，要节省很多成本。在这样的技术下，数据资源能够在满足高利用率的前提下运行得更长久，能源效用与数据资源的利用率会达到最高，成本也会控制得最低。

3.2 云计算有效地解决了资源存储技术面临的困难

云计算通过可视化界面，可以实时集中监控整个虚拟环境、存储负载服务的性能数据情况等。从而实现利用数据融合管理大规模服务器技术，根据用户需求，系统自动或人为配置，灵活地分配存储资源空间和对外服务性能，甚至可以通过资源调度来解决存储空间的问题。同时，云计算高性能、可扩展的存储计算技术，可以从容应对目前高速发展的数据交换技术。在数据资料备份方面，云计算的高存储效率造就了快速备份、一备三的备份技术，达到十个九的数据可靠性^④。

在当今信息资源爆炸式增长的背景下，云计算可以确保数字图书馆在数据管理、数据存储、数据安全方面能力得到提升。

3.3 云计算扩展数字资源共享程度

云计算具有“云端资源库”的特点，人们可以在任何不同的时间和地点使用库中的资源。在云计算的环境下，数字图书馆的信息资源被存储在由成千上万个服务器构成的“云端”中，这些资源可以在全球范围内被使用。同时，云计算的高数据吞吐能力可以提升大型资料的传输效率。当资料面对用户高并发访问读取时，高性能的压力承载能力可以避免服务器瘫痪。云计算技术的应用，使得图书馆信息资源的共享度更高，信息使用者的需求更容易被满足。

4 数字图书馆中的云存储基础框架

与传统的数据存储相比，云存储具有很多优势。用户不仅不需要购买具体的存储设备，而且可以在任何地点利用网络接口获取已经存储的数据。云存储的核心是应用与存储装置的排列组合^⑤。对于数字图书馆终端用户来说，云存储并不是一种具体的设备，而是一种统一的、标准的服务，一种能够提供随时随地、图书资料数据存取的服务。

4.1 GFS、HDFS 分布式数据存储技术

GFS 和 HDFS 是两种主流的大规模分布式数据存储技术。云计算利用大规模分布式存储的途径来存储数据。为了满足 Google 在数据处理方面急速增长的需求，Google 设计并开发了文件处理系统 GFS (Google File System)。Hadoop 团队开发了开源系统 HDFS (Hadoop Distributed File System)。GFS 和 HDFS 是两种主流的大规模分布式数据存储技术。为了更加有效地利用云计算资源，Google 开发了基于 GFS 的 Map Reduce 运行模式。同时，Hadoop 团队基于 Google 的 GFS 和 Map Reduce，开发了属于自己的分布式并程序序设计模式，即 HDFS & Map Reduce。Map Reduce 是一个基于集群的高性能并行计算平台，运用并行计算与运行软件架构，提供了一个庞大但设计精良的并行计算存储软件框架，能自动完成计算存储任务的并行化处理，自动划分计算数据和计算任务，在集群节点上自动分配和执行任务以及收集计算结果，将数据分布存储、数据通信、容错处理等并行计算涉及到的很多系统底层的复杂细节交由系统负责处理，大大减少了软件开发人员的负担。

4.2 数字图书馆中的云存储基础框架

如图 1 所示，图中展示的是基于云计算的数字图书馆中大规模数

据的云存储框架。

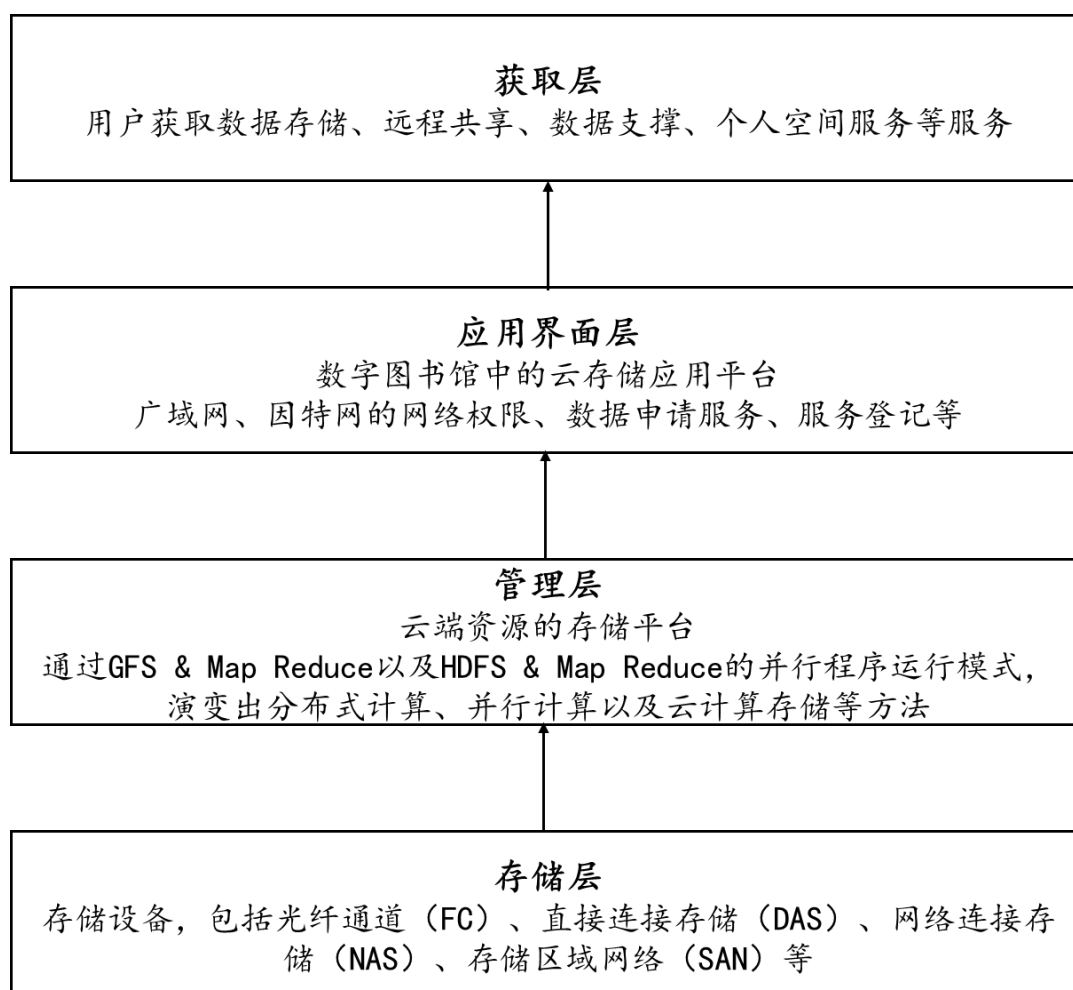


图 1 基于云计算的数字图书馆云存储框架

在云计算架构数字图书馆中，存储层是最基础的部分，书籍资料也是存储在这一层。存储设备可以是光纤通道，也可以是直接连接存储（DAS）、网络连接存储（NAS）和存储区域网络（SAN）等。云存储的存储装置通常涵盖许多不同地理区域的资料与数据。它们通过广域网、因特网或光纤网络等连接在一起，形成统一的存储服务。云端资源的存储平台是管理层，即虚拟化存储计算平台，也是数字图书馆中云存储的核心部分^⑥。通过 GFS & Map Reduce 以及 HDFS & Map Reduce 的并程序运行模式，存储平台可以演变出分布式计算、并行计算以及云计算存储等方法。云端资源的存储平台拥有处理大规模数据的能力，并且能够使多重存储设备中的云存储资源联合运作，实现空间资源可伸缩、服务性能可调度等特点^⑦。云存储应用平台是云计算系统的应用界面层，是整个云存储系统中最灵活的部分。对应于不同的操作界面和用户请求，应用平台提供相应的不同服务，包括统一的图书馆服务界面、服务登记和证明、数据申请服务、广域网、因特网的网

络权限等。在获取层，任何数字图书馆的被授权用户都可以登入云存储系统，并通过应用界面获得使用云存储设备的权限。通过获取层，用户可以获得书籍资料数据存储、远程读取共享、数据应用支撑和个人空间服务等服务。

5 数字图书馆中的云服务模型

随着云计算使用的普及，信息服务的定义方式也发生了相应的改变。事实上，“云”并不是将服务分开，而是将服务都聚集在一起。云计算供应者根据以下三种基本模式提供它们的服务：基础设施即服务（IaaS）、平台即服务（PaaS）、软件即服务（SaaS）。在这三种模式中，基础设施即服务（IaaS）是最基础的一种，每一个更高程度的模型都提取了低程度模型的细节^⑥。数字图书馆中的云服务模型包含硬件和应用程序托管层，它在提供基础服务的同时，还对计算服务和存储服务进行整合与链接。无论是哪里的资源，也无论是什么样的数据库，云服务模型都可以提供给用户最快、最直接的服务。平台即服务（PaaS）模式作为一个开发平台服务于软件开发者，开发者可以在一个包括 SDK，文档和测试环境等在内的开发平台上非常方便地编写应用，而且不论是在部署，或者运行的时候，用户无需为服务器，操作系统，网络和存储等资源的管理操心。此模块提供友好的开发环境、丰富的端口服务、可伸缩的资源调度、精细的管理和监控。软件即服务（SaaS）是将应用作为服务提供给软件使用者。通过 SaaS 这种模式，用户只要接上网络，并通过浏览器或者软件，就能随时随地直接享受数字图书馆的各类云服务。

5.1 数字图书馆云服务模式

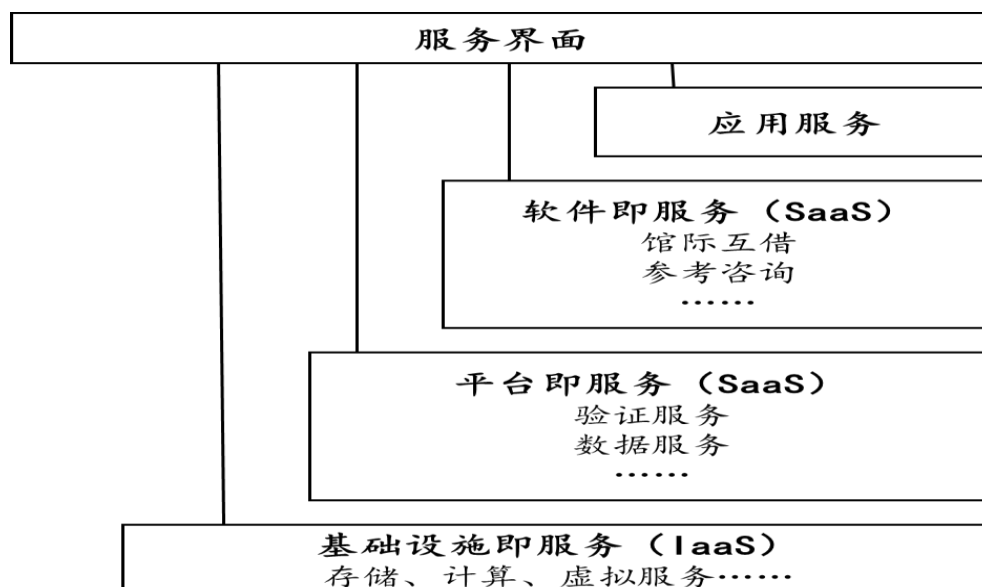


图 2 数字图书馆中的云服务模式

图 2 展示了数字图书馆中较为普遍的云服务模型。数字图书馆的云平台包括基础框架服务平台、软件服务平台、地区服务平台（包括地区应用基础框架平台和地区应用系统）和服务平台（即应用、程序、信息资源的整合）。数字图书馆的这种开放式、面向服务的框架，可以通过定义良好的接口和契约，将应用程序的不同功能服务之间联系起来。对于整合类似于区域性公共服务、不同交易之间的公共服务、第三方公共服务等此类数字图书馆的不同地区性服务是具有很大的便利性的。

5.2 数字图书馆云服务平台架构

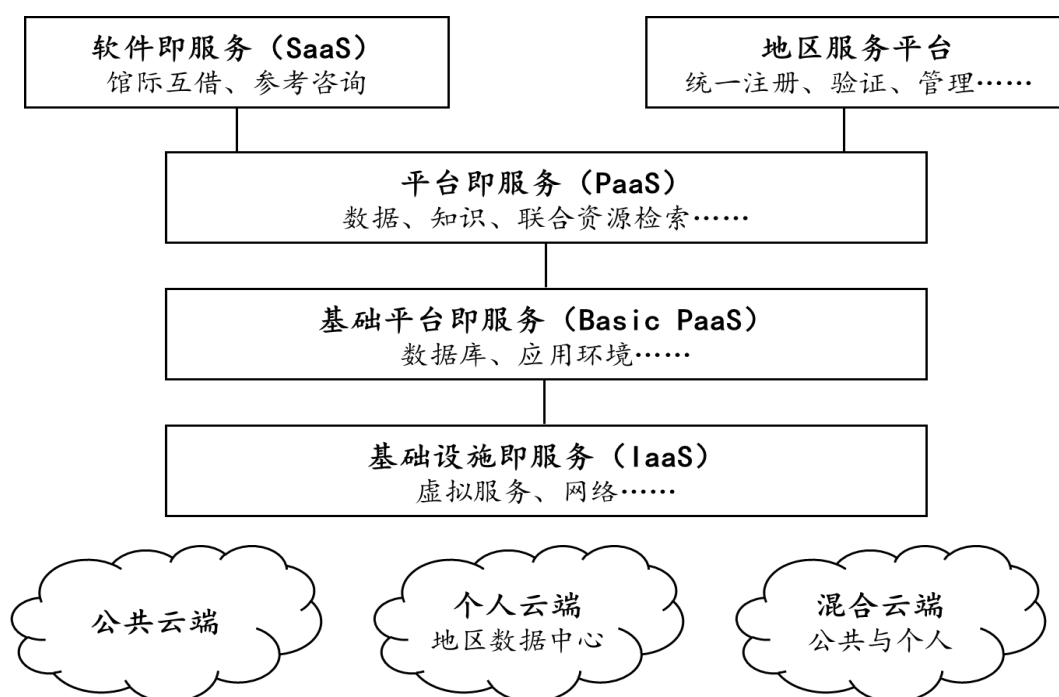


图 3 数字图书馆中的云服务平台

我们可以建立一个像图 3 中展示的数字图书馆云服务平台。这个平台包括基础框架服务、基础平台服务和软件服务。我们可以通过按需购买 IaaS 的云存储服务，建立属于自己的云服务中心。公共服务平台（平台即服务 PaaS）面向数字图书馆软件开发者，由一系列软件构成，它提供类似数据服务、验证和指示等服务。终端应用服务（即软件即服务 SaaS）面向数字图书馆用户，包括数字图书馆的服务平台和地区服务平台，提供数字图书馆的一系列软件应用服务，如参考咨询、馆际互借和专题服务等^⑨。数字图书馆云服务平台针对不同的场景应用和数据安全要求，对外输出形式分为公共云端、个人云端以及混合云端，实现云计算服务定制化。

6. 结语

数字图书馆的云计算的定义目前仍处在研究的初级阶段。随着云计算在图书馆领域的不断发展与应用,数字图书馆的发展将会迈上新的台阶。同时,随着云计算技术和相关政策法规的逐步成熟与完善,云计算必将在数字图书馆的发展中发挥巨大的推动作用。

参考文献:

[1] Fox A, Griffith R, Joseph A, et al. Above the clouds: A Berkeley view of cloud computing[J]. Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS, 2009, 28: 13.

[2] Ou Y, Zhou Z. Cloud Storage and Cloud Service in the Digital Library[C]/Proceedings of the 9th International Symposium on Linear Drives for Industry Applications, Volume 1. Springer Berlin Heidelberg, 2014: 295-302.

[3] 张东. 云存储技术研究与应用[J]. 科研信息化技术与应用, 2012, 3:85-90.

[4] 孔陶茹. 云存储应用探讨[J]. 互联网天地, 2013(12).

[5] 王丽敏, 党卫红, 王淑阁. 云计算环境下个人数字图书馆发展探析[J]. 浙江高校图书情报工作, 2011, 1(105):1-5.

[6] 比特网. 云存储时代来临:吹响后存储时代的号角[J]. 计算机光盘软件与应用, 2013 (21).

[7] 谌伟. 云存储系统的分析与应用探讨[J]. 数字技术与应用, 2013(10).

[8] 江伟玉, 刘丽敏, 查达仁. 面向云存储的访问控制服务研究[J]. 信息网络安全, 2013(10):34-37.

[9] 李屹炬, 魏宝林. 分布式云存储技术的分析[J]. 信息系统工程, 2013(9):98-98.

Abstract Cloud computer draws increasing attention in the field of information. The development of the current digital library is prevented for its high cost of infrastructure construction, its low rate of resource sharing, and the problem of storage technology. Cloud computing can effectively solve

these problems by replacing the recent function of library database with the incomparable advantage of low cost and super-scale. The paper briefly describes the cloud computing and its advantages and has an analysis on the changes and the problems in library information construction. Clouding computing technology brings about and mainly deals with the application of cloud computing in the digital library on the basis of cloud storage infrastructure and cloud service models.

Keywords Cloud computing, Digital library, Cloud storage infrastructure, Cloud service model

（作者信息：陈清华，江苏省社会科学院图书馆；浦洁，南京图书馆。

邮编：210018。收稿日期：2018-4-16）

通联：210016 南京市建邺路168号3号楼3403室 陈清华 收

18705161666, cqh088@163.com