

云计算在数字图书馆中的应用研究

浦 洁

摘 要 数字图书馆虽然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、传统架构模式上的资源分享效率低和存储量小等一系列缺陷,但无法克服各自独立的信息源分布储存和服务。云计算可以利用低成本与超大容量等优势,将图书馆数据进行迁移,从而有效地解决了传统图书馆运行存在的一系列问题。论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

分类号 G250.76

DOI 10.16810/j.cnki.1672-514X.2018.01.013

Research on the Application of Cloud Computing in Digital Library

Pu Jie

Abstract Although the digital library gradually solves a series of defects such as the high cost of traditional library infrastructure such as network and other hardware facilities, the low efficiency of resource sharing and low storage in traditional architecture model, it can not overcome the independent sources of information storage and distribution services. Cloud computing can make use of the advantages of low cost and super large capacity to transfer library data, thus effectively solve a series of problems existing in the operation of traditional libraries. Based on the description of cloud computing and its advantages, the changes and problems brought in library information construction by cloud computing technology have been analyzed, and the process of creating a digital library of cloud computing applications is studied.

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

数字图书馆虽然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,由于信息源各自分离、相对独立,阻碍了其进一步发展。随着分布式计算技术的应用,并行计算、网格计算、效用计算以及云计算的普及,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]。其中,云计算的特点在于按需获取、多样性、扩展性和虚拟能力服务^[2],在资源利用与管理方面有低成本、高效益的优势,这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。

1 数字图书馆建设中存在的问题

1.1 基础设施架构成本高

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,随着IT行业的发展,这些设施活跃周期很短,即便是使用不太先进的设备,购买和扩展新的模块配件也需要许多花费。

1.2 信息资源存储技术问题

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、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]。

1.3 资源共享效率低

尽管数字图书馆打破了空间与时间的限制,但受其安全性的限制,为保证服务设备的正常功能与数据资源的安全性,图书馆会对访问终端服务器的流量与最大支撑数量进行控制,这样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能力、压力承载能力与资源共享程度。

2 基于云计算的数字图书馆优势

2.1 有效降低基础设施构架成本

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,软硬件更新迭代无需额外花费。同时,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。

2.2 有效解决资源存储技术上的困难

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,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99.999999%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。

2.3 有效加强数字资源共享

云计算具有“云端资源库”的特点,人们可以在任何时间、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。

3 数字图书馆云存储技术及其基础框架

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,一种能够提供图书资料数据随时随地存取的服务。

3.1 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)。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行程序设计模式,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。

3.2 数字图书馆中的云存储基础框架

图1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。

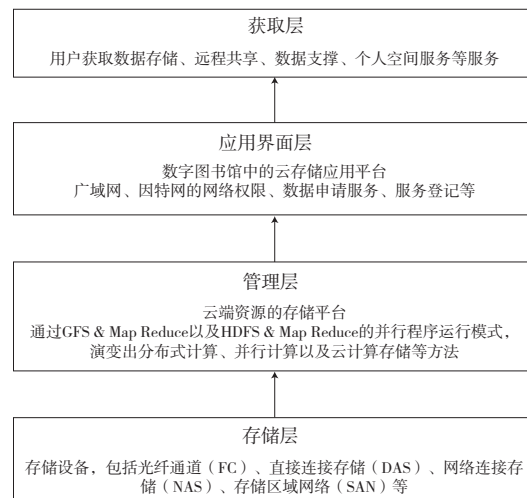


图1 基于云计算的数字图书馆云存储框架

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,实现空间资源地可伸缩、服务性能地可调度等特点^[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。应用平台提供不同服务对应于不同的操作界面和用户请求,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。

4 数字图书馆云服务模型

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,而是将资源和服务更紧密地聚集在一起。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,每一个更高层次的模型都提取了低程度模型的细节^[8]。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,用户都无需为服务器,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。

4.1 数字图书馆云服务模式

数字图书馆的云服务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),如图2所示。

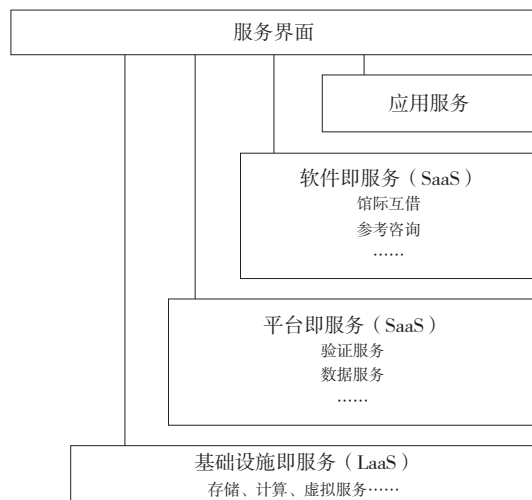


图2 数字图书馆中的云服务模式

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、第三方公共服务等地区性服务具有很大的便利性。

4.2 数字图书馆云服务平台架构

数字图书馆云服务平台包括基础框架服务、基础平台服务和软件服务三个方面,如图3所示。

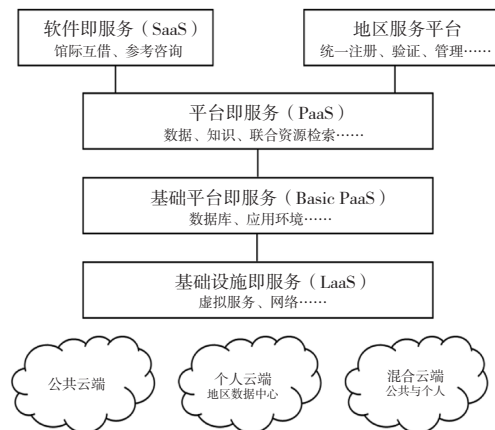


图3 数字图书馆中的云服务平台

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。

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):87-90.
- [5] 王丽敏,党卫红,王淑阁.云计算环境下个人数字图书

5 结语

数字图书馆的云计算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,云计算必将推动数字图书馆的发展。

馆发展探析[J].浙江高校图书情报工作,2011,1(105):1-5.

- [6] 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):5-8.
- [7] 湛伟.云存储系统的分析与应用探讨[J].数字技术与应用,2013(10):216.
- [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.

浦洁 南京图书馆馆员。江苏南京,210018。

(收稿日期:2017-08-01 编校:曹晓文)

(上接第47页)文更为严峻的技术问题。首先规范数据库的建库模式的选择及后期的维护与更新,都需要技术部门的大力支持。我们需要充分调研需引进的规范数据源,了解其更新方式、更新频率,评估其对本地系统的压力,在与技术部门保持良好沟通的基础上,借鉴已实施规范控制的西文编目经验,做好能够保证规范控制顺利实施的各项准备工作。其次,妥善处理日文编目平台的调整和OPAC显示的设置

问题。为实现目录的FRBR化呈现,极有可能需要将套录数据的内容类型、媒介类型及载体类型的日语文术语替换为英文术语,为了更好地揭示关系,也会考虑使用英文的关系说明语。“新NCR”虽然提供了日英对照的相关术语,但大量的本地化处理势必影响工作效率与准确度,因此,笔者认为除了套用模板的方式以外,还应尝试探索使用技术手段实现日英术语的自动化转换,以提高准确率与工作效率。

参考文献:

- [1] 王志君,霍晓伟.日语文献编目应用RDA设想[C]//编目:核心能力与挑战 第四届全国文献编目工作研讨会论文集.北京:国家图书馆出版社,2015.
- [2] 罗群.RDA全视角解读[M].北京:国家图书馆出版社,2015.
- [3] 崔健.日本图书馆界对RDA的研究[J].图书馆建设,2014(2):21-24.
- [4] 新しい『日本目録規則』の策定に向けて[EB/

OL].[2016-10-25].<http://warp.da.ndl.go.jp/info:ndljp/pid/9484238/www.ndl.go.jp/jp/library/data/newncr.pdf/>.

- [5] 新しい『日本目録規則』(新NCR)[EB/OL].[2016-12-30].<http://www.ndl.go.jp/jp/data/ncr/index.html>.

霍晓伟 中国国家图书馆馆员。北京,100081。

(收稿日期:2017-02-08 编校:刘明)