

云计算环境下数字图书馆信息安全研究*

□ 欧裕美 / 南京大学信息管理系 南京 210008
 / 江苏技术师范学院图书馆 常州 213001
 苏新宁 / 南京大学信息管理系 南京 210008

摘要: 云计算技术的迅速发展给图书馆的管理与服务带来深刻变革,给互联网环境下数字图书馆信息资源中心作用的发挥带来了机遇。但同时因为云计算作为一种新技术,还存在许多欠缺和问题,而云计算的安全问题是制约它的应用与发展的一个核心问题。云计算环境下的数字图书馆则同样面临着更为严峻的信息安全问题。针对一系列的安全问题,文章提出云环境下数字图书馆信息安全机制。

关键词: 云计算, 数字图书馆, 信息安全, 安全机制

DOI: 10.3772/j.issn.1673-2286.2011.09.006

1 引言

随着分布式计算、并行计算、网格计算、效用计算等计算模式的运用和不断的发展,近年来计算机界一种新的计算模式——“云计算”被提出并且在几年内得到快速发展和应用。

云计算技术的迅速发展使越来越多的用户感受到云计算所带来的巨大变化,给图书馆的管理与服务带来深刻变革。云计算不仅可以取代目前数字图书馆数据中心现有的功能,而且它的低成本优势,超大规模、高扩展能力等特点是现有数字图书馆数据中心所无法比拟的。云计算为我们提供了超大规模的计算和存储服务、便利的软件服务以及平台服务,但同时,它作为一种新技术,还存在很多欠缺和问题,而云计算的安全性和可靠性等问题成为云计算前进道路上的一道鸿沟。云计算环境下的数字图书馆则同样面临着较传统数字图书馆网络环境下的数据泄密、数据丢失以及知识产权纠纷等更为严峻的信息安全问题。

云计算是Parallel Computing(并行计算)、Distributed Computing(分布式计算)和Grid Computing(网格计算)的发展和相互融合的结果,是虚拟化

(Virtualization)、效用计算(Utility Computing)、IaaS(基础设施即服务)、PaaS(平台即服务)、SaaS(软件即服务)等概念的混合演进并跃升^[1]。云计算通过虚拟化等技术把成本不高的计算机群体整合具有强大的计算能力、海量的存储能力的计算系统,把它强大的计算能力以服务的模式(IaaS、PaaS、SaaS等服务)租借给它的用户。



图1 云计算的平台结构

* 本文系2010年国家自然科学基金《数字图书馆建设体制及其发展模式研究》(项目编号:10BTQ005)研究成果之一。

图1是典型的云计算平台结构。用户端只需要简单的终端设备就可以通过Internet访问到“云”中超大规模的计算和存储资源。服务层将云中的各种资源管理起来并组织成相应的服务(比如存储、软件即服务、平台及服务)提供给用户。虚拟资源层将经过虚拟化的资源进行整合管理,然后提供给服务层以供分配。

云计算具有按需服务、超大规模、通用性、可扩展性高以及虚拟能力强大等特点,具有客户端要求低、成本低以及能有效提高资源利用和管理效率等优势,但它同时具备着潜在的危险性。

云计算的出现,改变了信息服务的提供方式,“云”中的资源是可以无限扩展的,“云”并不是一个单独的服务,而是一个服务集合。图2是云计算中比较有共性的云计算技术框架,通过基础构架即服务(Infrastructure as a Service, IaaS)、平台即服务(Platform as a Service, PaaS)、软件即服务(Software as a Service, SaaS)三种形式,云计算以前所未有的规模为用户提供IT服务能力。

近几年来云计算的研究引起学界的广泛关注,相关论文的数量急剧上升,发文量呈快速增长趋势。图书情报界的云计算研究的刊文量尽管也呈增长趋

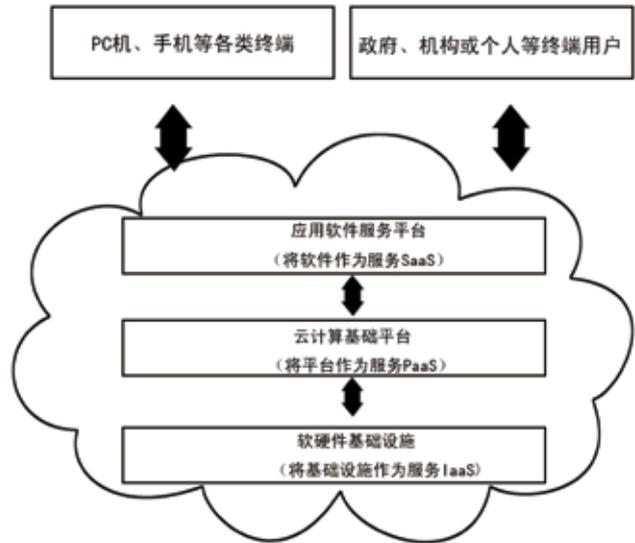


图2 云计算的技术框架

势,但尚处于起步阶段,学术论文为数不多,博硕士学位论文更是寥寥无几,少有图书情报方面的云计算研究,图书情报界安全方面的云计算研究更是为数不多。表1是中国期刊网中查阅到2007年至2011年各年份篇名分别为“云计算”、“云计算、数字图书馆”、“云计算、数字图书馆及安全”的论文篇数。

表1 论文分布表

年份 \ 篇名	云计算	云计算、数字图书馆	云计算、数字图书馆 安全
2007年	2	0	0
2008年	86	0	0
2009年	348	4	0
2010年	869	19	3
2011年	217	6	1

短短几年间,研究云计算的相关论文呈急剧上升的趋势,但篇名为“云计算、数字图书馆”的文章为数不多,2009年仅4篇;2010年有所增加,为19篇,可见我国图书情报界的关于数字图书馆运用云计算技术的研究尚处于起步阶段,而数字图书馆运用云计算技术的安全问题的研究还几乎是空白,篇名为“云计算、数字图书馆及安全”的论文2007-2009年都为0篇,2010年至今4篇,因此本文对云环境下数字图书馆的信息安全进行研究具有很大的意义。

2 数字图书馆信息资源建设

2.1 现阶段数字图书馆信息资源建设中存在的一些问题

(1) 基础设施建设、维护费用高但利用率不高。目前IT基础设施随着信息化建设的推进与日俱增,例如,购买大量的服务器和存储设备甚至是小型机,这些设备每笔都是数万、数十万甚至数百万,而据统计

资料显示, 这些基础设施的平均利用率不到50%, 并且每年还需要庞大的费用维护服务器、存储器等设备。同时, 由于设备本身的性能和信息技术的迅速发展, 设备使用寿命实际上非常短。也就是说, 即使仅仅为维持原来的服务, 在较短的周期内也都需要投入大量的资金购买新的设备。

(2) 随着数字资源增长, 硬件设施越来越成为瓶颈。图书馆作为信息资源中心, 长期保存大量数据资源。近几年来数字资源急剧增长, 按照目前现状来看, 硬件的增长速度将会越来越跟不上数字资源的增长速度。数字资源的海量增加对各个图书馆的数据管理、备份、存储能力造成了非常大的负担。

(3) 信息资源存储技术问题。这是数字图书馆面临的一个很大的问题。数字图书馆的多语言检索、人机交互技术、不同数据标准之间的数据交换, 数字图书馆需要的软件、工具、技术等有待进一步发展和完善。这对数字图书馆的信息资源海量增加对数字图书馆数据管理备份、存储设备能力提出越来越高的要求。

(4) 重复建设概率高。馆与馆之间信息资源相互独立, 重复建设现象严重, 高校间的学科、专业重叠交叉较大, 造成高校数字化资源建设的不断重复和浪费。数字信息资源的购买或数字资源镜像的设立也同样存在着大量的重复建设现象。

(5) 资源共享度不高, 数字图书馆打破了原有的图书馆服务的时空限制, 但为确保服务器的可靠运行及服务器内数据资源的安全性, 图书馆对服务器的接入终端数量及最大服务响应等都进行了相关的限制, 形成了服务器访问限制瓶颈的问题。另外图书馆的资源共享往往局限于一些对口的、接口相同的数据资源, 因此出现馆和馆之间的数据不能相通, 资源无法实现共享。

(6) 资源管理不到位, 资源维护欠缺。数字化资源随着新技术、多媒体的应用, 数据结构日益复杂、类型日益丰富, 数据量飞速增长, 然而资源管理系统对资源的有效组织和管理存在很多不足。数字化资源还必须不断进行维护, 需要推陈出新、与时俱进。

2.2 云计算技术给数字图书馆资源建设带来种种优势

云计算环境下, 能够解决数字图书馆资源建设中

的以上问题。

(1) 云计算有效解决了目前图书馆基础设施建设、维护费用高的问题。云计算对用户端的设备要求最低, 使用起来也最方便。例如数字图书馆可以按需购买相应的云计算, 购买基础设施即服务(IaaS)、平台即服务(PaaS)或者软件即服务(SaaS), 像水、电一样付费, 而不必把大量的资金投入基础设施建设中。

云计算中的核心技术是虚拟化技术, 可以说虚拟化为我们带来了“云”^[2], 云计算环境下所有虚拟化解决方案都是集服务器、存储系统、网络设备、软件及服务于一体的系统整合方案, 因此云计算具有低成本、绿色节能的优势, 用户不需要负担日益增加的数据中心管理的高昂成本, 不需要购买或租赁设施等。在海量数据处理等场景中, 云计算以PC集群分布式处理方式替代小型机加盘阵的集中处理方式, 大大地降低了成本^[3]。

(2) 云计算能有效提高数字图书馆基础设施利用率, 扩大数字资源共享度, 避免资源重复建设。在云计算环境下, 数字图书馆的信息资源存储在“云”的成千上万台服务器中, 可以在全球存取^[4], 图书馆的信息资源得到较高级别的共享, 云用户的信息需求得到较高级别的满足。

在云计算环境下, 传统的信息技术基础设施建设和交付使用模式发生了改变, 云计算为资源共享带来了变革。馆和馆之间的数字化信息资源不再是各自独立, 而是可以共享的。云计算提供对终端设备的跨平台、多样化支持, 为资源共享带来了天然优势。另外, 云计算可以无限整合国内外高校优质数字化资源, 有效地解决了数字化资源建设分布不均, 甚至重复建设问题^[5]。

(3) 云计算能有效解决图书馆信息资源存储技术问题和提高数字化信息资源管理和应用水平。云计算采用海量分布式存储的方式来存储数据, 采用冗余存储的方式来保证存储数据的可靠性^[6]。云计算采用数据管理技术, 对海量的数据存储、读取后进行大量的分析、处理, 向用户提供高效的服务。云计算资源规模庞大, 云计算系统管理技术是云计算的“神经网络”, 通过这些技术能够使大量的服务器协同工作, 提高海量增加的数字图书馆信息资源的管理备份、存储等能力, 使得数字图书馆数字化资源的管理和应用水平不断地提高。

3 云计算环境下数字图书馆面临的信息安全问题

先进的云计算技术给数字图书馆的建设带来种种优势,但因为它本身的一些特点以及技术上的缺陷可能会导致数据的泄露,对图书馆的信息资源安全构成严重的威胁。云计算环境下的数字图书馆面临着更为严峻的信息安全问题,不仅仅是传统数字图书馆网络环境下的数据泄密、数据丢失以及知识产权纠纷等问题,还有云计算模式的特点以及新技术带来的更具威胁性的数据泄密、数据丢失及知识产权纠纷的信息安全问题。云计算是一个公共存储平台,云环境下,图书馆将自己的数据交给“云”托管、将自己的数据与资料存放在公共存储平台中,对存储设备和信息数据不能进行直接的物理控制,这就增加了数据泄漏和丢失的风险。另外云供应商会不会因各种原因破产倒闭,云服务商有没有足够的力量、足够技术力量保证数字图书馆的数据安全,这些问题还没有得到有效解决。云计算环境下数字图书馆面临如下一系列信息安全问题:

3.1 云供应商问题

(1) 云供应商宣布破产倒闭,服务器被没收、停止运转或者云服务商改做其他生意,那么数字图书馆就面临着无法访问自己的系统的风险。

(2) 云供应商没能正确地处理某部分基础设施的安全问题,尤其没能正确地维护物理访问控制,导致图书馆的系统被破坏或者是控制图书馆数据的云服务商在保护数据方面做得不够,做得并不像它之前承诺的服务水平条款那么好。

(3) 云供应商没能很好地保护网络,没有严格的保密措施,没能使用严格的安全措施、没有做足工作来保障数据安全等等。

3.2 对数据不知情以及不可控制问题

数据安全性问题是数字图书馆将数字资源转移到“云”的一个最大的挑战。云计算的超大规模性、高扩展性以及强大的虚拟能力足以保证数字图书馆数字资源的存储和备份。但是,由于云是虚拟的,在现在运行的系统中,用户本身不清楚数据存储的具体位

置,更不知道数据有多少副本,以及云提供商是否遵循当地的隐私协议等等,使得数据的泄密成为问题。对于数字图书馆来说,数据的泄密就将会导致知识产权等方面的一系列纠纷问题,而如果是业务数据的泄密的话,将会带来更严重的问题。

3.3 云计算平台导致的数据安全问题

Gartner公司发布的一份报告《云计算安全风险评估》(Assessing the Security Risks of Cloud Computing)中提出:“云计算平台服务提供商权利过大,致使云计算平台中的用户数据可能会被非法访问与泄露。”

云计算以虚拟机技术作为云计算平台的基本组成部分,虚拟机技术带来非常明显的优势,服务器运行在虚拟的服务器上面而不再依赖于物理设备。但是虚拟的服务器存在着很大的安全问题。云计算中一台服务器存在着好几台虚拟的服务器,虚拟服务器之间有互相攻击的可能性,这对虚拟服务器构成很大的安全威胁。另外管理与维护超大规模云计算平台的硬件和软件不容易,云计算平台的硬件和软件故障问题可能会对其提供的服务产生影响,使得数据的安全性受到威胁。

在云计算体系下,所有用户的数据都位于共享环境之中。加密能够起一定作用,但是仍然不够。服务商是否将一些数据与另一些隔离开?加密服务是否由专家设计并测试?加密系统会不会出现问题?所有这些都导致数据安全问题。

还有来自外部的恶意攻击也构成了对数据安全的威胁。2005年著名的在线付费提供商PayMaxx泄露了多于250000个用户的信息,就是因为其系统的安全漏洞受到了来自外部的恶意攻击^[7]。

3.4 特权用户的接入

对数据进行访问时需要有限制控制,用户应该能够控制谁能访问自己的数据,每次对数据进行访问需要进行用户认证与授权,并对用户情况进行审计。随着云计算的使用客户及业务流程和运用向越来越多的终端用户开放,安全威胁不仅来自外部互联网,也来自内部的安全隐患。云提供商有责任对数据进行管理和维护,超级用户的存在简化了数据管理功能,因此威胁到用户的隐私。超级权限在给用户提供方便的同

时也给用户带来了威胁。

3.5 数据的持久性问题

云计算提供商一旦破产或是被大公司收购,图书馆的数据会不会受到影响?是否能拿回自己的数据,以及拿回的数据是否能够被导入到替代的应用程序中?一旦发生灾难,图书馆的数据和服务将会面临什么样的情况?任何没有经过备份的数据和应用程序都将出现问题!云服务提供商是否有能力恢复数据?以及需要多长时间?这些也都是云环境下数字图书馆的数据安全隐患。

3.6 知识产权保护问题

数字图书馆的知识产权问题在云环境下不仅依旧存在,而且有了新变化^[8]。在云计算环境中,各个图书馆之间通过有关的协议可以共享彼此的信息资源,“云”中的用户可以共享各个图书馆中的信息资源。各个图书馆运用“云”方便了用户,提高了信息资源的共享,但同时知识产权纠纷问题也就越来越突出。数字图书馆使用云计算服务,将自己的数据交给云托管,本来用户完全拥有被托管数据的知识产权。但云计算环境下,这些数据可能会以数据整合、数据挖掘、知识服务的名义被合法化而引发知识产权纠纷问题。

4 云计算环境下的数字图书信息安全机制

针对以上问题,我们提出以下云计算环境下的数字图书馆信息安全机制。

4.1 云计算模式与传统模式相结合机制

由于作为新技术的云计算本身的一些特点以及技术上的缺陷可能会导致数据的泄露和丢失,对图书馆的信息资源安全构成严重的威胁,还可能会因供应商的问题导致云服务中断,导致数据的泄露,甚至可能会在瞬间数据灰飞烟灭。因此我们运用先进技术的同时不丢弃传统的数字图书馆模式,采用云计算模式与传统模式相结合的灵活机制,对图书馆馆藏的各种数

据资源进行及时、全面的备份并长期、可靠地保存,保证万无一失。一旦云计算发生问题,服务中断、数据泄露甚至是数据全部丢失,也还有传统的服务,还有备份的数据。另外图书馆也可以搭建自己的私有云,私有云相对来说会更安全。私有云与公共云相结合使用,享用云计算带来的巨大优势和便利服务的同时,也可以避免可能发生的意外。

4.2 图书馆资源的用户认证与授权机制

用户认证与授权管理旨在授权合法用户进入系统访问数据,同时保护这些资产免受非授权的访问。通过集中化的身份和访问控制,可以保证信息安全,降低安全风险。用户认证与授权措施具有以下能力:

(1) 身份管理。有效地管理用户身份和访问资源的权限,用户身份认证在云计算环境下对于数据安全起到至关重要的作用,通过了认证的授权用户方可以访问“云”中的相应信息资源,通过对用户身份的认证、对资源访问的控制,能有效地保证数据的安全。

(2) 访问授权。访问授权在用户生命周期以内提供及时访问,从而加强安全和保护IT资源。通过集中的身份、访问、认证和审查服务,检测、管理并降低身份识别和访问的风险。云环境下的数字图书馆可以根据用户信息需求,将用户划分为若干个层次,并严格控制用户访问资源的权限。有用户权限的合法用户才可以进行操作,登录到云环境下的数字图书馆,进行浏览、检索、下载、创建、更新及整合等等。

4.3 数据隔离机制

基础构架云的一个核心技术是虚拟化,这意味着不同用户的数据可能存放在共享的物理存储器上。云计算系统对于客户数据的存放一般采取两种方式:提供统一共享的存放设备或者提供单独的存储设备。理论上共享存储设备存放数据时,可以通过存储自身的安全措施(比如存储影射)等功能以确保数据的隔离性。共享存储设备的优点是可以节约存储空间。单独的数据存储设备从物理层面隔离保护了用户重要信息,其优点是有效地保护用户的数据,但缺点是存储无法有效利用^[9]。数字图书馆绝对要选择能确保数据的有效隔离的存储设备,以保证数据的安全。

4.4 数据加密机制

数据加密的目的是防止他人拿到数据的原始文件后进行数据的窃取。在云计算环境中,数据的隔离机制如果可以防止其他用户对数据的访问,那么数据加密的主要用途是防止“内鬼”,即避免服务提供者对数据进行窃取。即便隔离机制不能有效防止其他用户对数据的访问,数据加密机制也能有效保护数据的安全。数据加密在云计算中的具体形式为:数据在用户侧使用用户密钥进行加密,然后上传至云计算环境中,之后使用时进行实时解密,避免将解密的数据存放在任何物理介质上。数据加密有很多成熟的算法,如对称加密、公钥加密等。

在云计算环境中,与数据加密配合使用的方式还有数据切分,也就是把数据在客户端打散,经过加密后分散在几个不同的云服务器上,这样对于任何一个服务提供商来说,都无法获取到完整的数据。

4.5 数据保护机制

比如把一份数据存放在多个地点,这样即便一份数据出现问题,图书馆也可以从其他位置获取数据。采用快照、备份和容灾等重要保护手段保护重要数据,即便受到黑客、病毒等逻辑层面的攻击或者地震、火灾等物理层面的灾害也都可以有效恢复数据。图书馆还可以选择多个云计算服务提供商,选择不同地点的数据中心提供服务,这样即便一个服务停止了,甚至服务提供商破产倒闭了,图书馆也可以保留自己的数据,继续运行自己的业务。

4.6 非技术手段

技术不是万能的,特别是对于安全这个领域,我们无法判断一个云计算在技术上是否真正做到安全可靠,我们也无法保证我们自己的技术能确保数据的安全可靠。因此,我们可以采用一些传统的非技术手段来约束云服务提供商^[9]。例如采用第三方认证,即采用一个中立机构对信任双方进行约束,该机构主要负责对云计算服务提供商的服务进行安全认证,采用标准化的技术手段和非技术手段对服务进行检测,找出其安全漏洞,对其安全级别进行评价。再比如运用合同、协议进行强制约束。目前已经有很多云计算服务

提供商推出自己的云计算服务水平协议,这些协议从服务质量、技术支持和知识产权等方面对服务进行了规范,对双方的权利和义务进行了明确。图书馆作为云计算用户,要认真斟酌协议,判断其是否合理,争取通过协商使协议合理,以保护自身权益,确保数据安全问题的有效解决。在很多情况下,合同应该详细地列出灾难准备程序、数据丢失和泄露的具体解决方案;图书馆要十分谨慎地认真仔细检查安全等级以及合同条款。还有一点非常重要,那就是“信誉”,一般越大的企业对于自身的信誉越看重,不会因为利益去盗取客户的数据,对其员工有更加严格的要求。因此,图书馆应该选择有实力、有信誉的云服务提供商提供云服务。

4.7 政策法规制定与完善机制

云计算环境下,信息安全和隐私保护问题是云计算面临的首要问题。其次是标准问题。云计算需要借助标准化才能规模化和集约化。目前云计算厂商各自为战,尚未在业界形成一个统一的标准,这就给云计算产业的发展带来了瓶颈。再次,就是法律法规问题。各国对数据保护和信息跨界流动的法律法规不尽一致,会直接导致云计算的应用效果大打折扣。

我国目前的信息化立法,尤其是信息安全立法,相对国外已成普及之势的网络立法情况而言,仍旧处于起步阶段,对于刚刚兴起的云计算技术中的信息安全问题更是几乎一片空白。我国政府和法律界必须认识到制定与完善相关的政策法规的紧迫性,积极推进这一方面的工作,确保用户信息安全、服务安全。云服务提供商和用户应各自负起应负的法律 responsibility,建立互信关系,确保用户利益。

只有制定云计算环境下相关的政策法规,并不断地进行完善,云计算环境下的数字图书馆才能保障其信息资源的安全可靠,才会得到长足发展。

5 结束语

目前云计算的应用还处在初步阶段,我们的数字图书馆的云计算应用也还处于初步阶段甚至是探索阶段,随着云计算技术的进一步发展,随着图书情报界对云计算技术的广泛应用,我国数字图书馆的发展将进入一个崭新的阶段。随着云计算技术的日益成熟

以及相关政策法规的逐步健全,云计算将更好地发挥其优势作用,云计算的安全问题将会迎刃而解。云计算安全问题的解决则会将云计算的发展推上另一个高

峰,云计算环境下的数字图书馆也将得到更加有力地发展。

参考文献

- [1] [OL]. [2011-08-20]. <http://www.chinacloud.cn/show.aspx?id=741&cid=17>.
- [2] CHEN HAIBO, CHEN RONG, ZHANG FENGZHE. Live updating operating system using Virtual [C]// Proceedings of the 2nd International Conference on Virtual Execution Environments, 2006:33-42.
- [3] 张为民,唐剑锋,罗治国,等. 云计算深刻改变未来[M]. 北京:科学出版社,2009:55.
- [4] 石美清. 云计算对图书馆数字信息资源建设的影响[J]. 图书馆理论与实践,2001(5).
- [5] 张润. 基于云计算的高校数字化资源和平台建设[J]. 计算机教育,2010(22).
- [6] 张为民,唐剑锋,罗治国,等. 云计算深刻改变未来[M]. 科学出版社,2009:76.
- [7] A Chronology of Data Breaches Since the ChoicePoint Incident [OL]. [2011-08-20]. <http://www.privacyrights.org/ar/ChronDataBreaches.htm>.
- [8] 王长全,艾雾,姚建. 云计算环境下数字图书馆信息资源安全策略研究[J]. 情报杂志,2010(3).
- [9] 朱近之. 智慧的云计算[M]. 北京:电子工业出版社,2010:233.

作者简介

欧裕美 (1971-), 南京大学管理学硕士, 江苏技术师范学院图书馆馆员, 主要从事数字图书馆技术和管理研究。
苏新宁, 南京大学信息管理系教授、博士生导师。E-mail: xnsu@nju.edu.cn

Research on Information Security of the Digital Library Based on Cloud Computing

Ou Yumei / Department of Information Management, Nanjing University, Nanjing, 210008
/ Library, Jiangsu Teachers University of Technology, Changzhou, 213001
Su Xinning / Department of Information Management, Nanjing University, Nanjing, 210008

Abstract: The rapid development of cloud computing brings deep reform to the management and service of digital libraries and development opportunity to the function of information center on the base of internet. But as a new technology, cloud computing is not mature enough with quite a lot of disadvantages, and the issue of security is the key hinder of the application and development of cloud computing. Digital Library under cloud computing environment is also facing much severer information security issues. For a range of security issues, the paper tries to deal with the security mechanisms of the digital library information.

Keywords: Cloud computing, Digital library, Information security, Security mechanism

(收稿日期: 2011-08-26)