

当选择云服务支撑图书馆业务时 需要考虑的因素

□ 孙卫 / 中国科学技术信息研究所 北京 100038
/ 科学出版社 北京 100717

摘要: 云计算和云应用已经受到越来越多图书馆用户的关注。但是,在选择云服务提供商的时候,需要考虑什么因素,很少有图书馆认真地思考。在发生很多云服务的事故以后,人们对于云服务事故使业务中断、云数据和系统的安全,以及针对云服务提供商的水平和能力等提出了一系列的问题进行思考。

关键词: 云计算, 云服务, 云故障, 云安全

DOI: 10.3772/j.issn.1673—2286.2012.03.002

在2011年4月从亚马逊租借计算资源类型云服务支撑机构应用的客户,由于云计算服务的故障^[1],从而造成了这些机构业务的中断,为此亚马逊专门进行了道歉。根据资料[2]披露,使用微软的Windows Live Messenger作为云应用的客户,由于Windows Live Messenger内部故障,造成了17355个账户出现问题,丢失了客户存储的10年数据、信息,同时使得有些客户失去了重要的商业机会,为此微软公司也进行了道歉。同样的资料披露,由于搭载在同一个云服务应用的某些客户受到司法部门的调查,司法部门采取相应的技术手段调查取证时,影响了在这组云服务中的其他正常用户,为此云服务的提供商无法向司法部门投诉,也只能向其他正常搭载的客户进行道歉。

由于这些应用云服务的客户受到了不同程度的影响,所以将要使用云服务的机构和个人开始了一场针对云服务适用性或者是云计算安全性的讨论。由于云计算、云应用、提供云环境基础设施是三类不同的云能力^[3],图书馆考虑比较多的是租借云计算资源和进行云应用,而管理云计算资源和已经具备云应用的机构都称之为云服务提供商,为此图书馆需要思考以下问题。

1 云服务的中断对业务影响程度的评估

针对图书馆要利用云服务支撑图书馆业务,图书馆首先要进行业务和云服务方式的若干分析。

(1) 在图书馆最经典的检索、借阅服务对比中,开架的图书馆能否提供读者正常的到馆借还服务使得比到馆的读者能否进行检索服务要重要。如果采用的是总分馆体系的云服务,借还业务的中断将会对图书馆的工作造成很大的影响。对于这类云服务,就要考虑是否需要构建混合模式的服务。例如,OCLC公司就提供把检索和借阅业务分别构建在远端(OCLC)和近端(图书馆自动化系统)的混合模式上^[4]。

(2) 对于云服务提供商的体系结构和能力的分析非常重要。

一种是专门提供云服务的提供商,这类服务商在体系架构、服务水平上是按照IT服务水平而建立的,如果图书馆选择这类云服务提供商提供的服务,系统的可靠性、系统的稳定性、系统的性能相对是有保障的。另一种是以自己应用为主,提供部分剩余能力进行云服务的云服务提供商,这类服务商的服务水平是按照满足

自己的应用需求为优先条件的,当外来搭载的用户影响到自己的使用时,会优先保证自己的应用,而非优先保证搭载客户的应用。

2 云服务的数据和系统安全对图书馆服务可能造成的影响的评估

(1) 图书馆的很多数据是具有隐私性质的,云服务如何保证这些隐私数据的安全?具体的措施有什么?这点是需要评估的。

从图书馆的角度,每个读者的个人信息、每个读者使用图书馆的信息是相对隐私的。对于云服务而言,读者个人信息的保护是最重要的,其次,对于读者个人使用信息的保护是有必要的。读者对图书馆是信任的,不能因为使用云服务而造成读者个人信息的泄漏,使得读者对图书馆的服务产生不信任。云服务可以对读者的使用信息进行宏观的统计分析与利用,但是不能对每个读者的行为进行跟踪分析,然后推送广告:这样会使读者收到过多针对个人专业、阅读习惯的广告,从而对图书馆产生反感。

(2) 图书馆的很多数据存放在云服务提供商的系统中,如果云服务提供商破产了,图书馆是否有能力导出这些数据到新的云服务提供商的系统中去,数据的可迁移性是重要的数据安全评价点。

图书馆的数据中,只有书目数据是基于MARC标准,但是这些数据进入云服务以后,变成了只符合特定的云服务要求的数据格式和类型。如何考量与测试云服务提供商能将数据再导出到MARC标准呢?

在许多应用中,关联了很多非MARC记录的部分,每个云服务的应用针对这类关联也是完全不一样的,在导出MARC标准的数据时,如何正确导出这种关联关系也是在做云服务迁移时需要考虑的因素之一。

在云服务中还有很多使用过程的数据,而积累使用过程的数据是帮助提高图书馆信息服务能力很重要的因素。例如,进入云服务的用户的IP信息、使用的栏目信息、检索信息、发表书评与议论的信息等。但是,每个云服务提供商处理这类信息的方式是不同的,很大程度是不兼容的。如果简单的迁移丢失了这类数据,那么利用新建立的云服务吸引已经稳固的用户群是很困难的。

(3) 云服务系统崩溃以后,是否具备原始数据的系统恢复能力是需要评价的。

从资料中可以看到,微软的Windows Live Messenger内部故障使得系统崩溃以后无法进行数据恢复,而给用户造成不可逆的损失。那么图书馆把全部业务搭载在云端,也一定存在由于云端崩溃而造成的对服务不可逆的损失与影响的可能性。

3 对于云服务提供商的一些评价分析^[5]

(1) 对于图书馆而言,减少IT基础设施,提高服务水平 and 能力是使用云服务的原因之一。那么云服务提供商是否具备足够的IT基础设施呢^[6]?云服务提供商在带宽、存储能力、系统的冗余能力上是否有足够的商业资本支撑,这点是最重要的考察内容之一。

图书馆的末端浏览器和客户端可能来自于不同互联网网络,要保证每个方向的使用畅通无阻,云服务提供商的网络带宽和接入网的多样性是投入之一。同时,网络的繁忙程度和图书馆的峰值应用是否相关也是需要数据统计分析进行评价的。数据的增加、运营的稳定都需要云计算服务商提供更多的IT基础资源,但是又不能为了备份而收图书馆很多费用,因此是否能动态调整IT基础资源和动态租借IT资源成为考量云服务提供商管理水平的一个平衡点。

(2) 对于云服务提供商而言,是否有能力判断客户的类型、客户之间的相关关系、客户之间的安全性^[2],也是需要考量的。

在资料中显示,由于云服务提供商不能判断客户的类型、相互之间的关系、客户的安全性,因而产生了恶意的客户,使得正常的客户使用受到了影响。

首先,图书馆类型的客户可以搭载在一起。如果非图书馆类型的商业信息服务类型客户与图书馆客户搭载在一起,就有可能窥探图书馆用户的使用数据,从而搭载上他们的广告和应用。

其次,如果有恶意的搭载客户,使用同一组计算环境是没有相互隔离的,那么这个恶意客户在这个环节进行图书馆相关数据的窃取就很容易。

(3) 云服务提供商对于内部人员管理机制是否严格,如何避免内部人员在本地复制、窃取数据,也是需要关注的问题。

在国内已经发生过这样的事件:信息服务商利用和客户合作的机会,窃取客户的数字内容,然后自己开展信息服务提供数字内容营销。此类事件在国内这种环境下尤为突出,IT人员的流动大,云服务提供商是否

具备防止内部人员复制、窃取数据的能力也是需要评估的。

(4) 对于图书馆租借云服务而言,更多的是租借计算能力和应用系统的搭载。而无论是计算能力还是应用系统搭载,图书馆的数字内容、用户数据、使用过程的日志文件等都是在云端的,那么如何保证提供这些数据导入后的原始数据保存、导入后数据的导出、这些原始数据和导入数据的安全存放能力,也是需要评估的。

例如,图书馆提供的数字内容,是保存在云端还是保存在图书馆本地?如果保存在云端,评估安全性就包含很多因素,最主要的因素是防范病毒、原始内容的检查与修复、原始数字内容被窃取的预防措施等方面需要评估。如果是放在图书馆本地,如何证实在云端卸载以后,在云端没有保存或者复制原始的数据内容呢?

4 如何应对未来的技术变化

(1) 很多人在是否使用云计算或者云服务能力时,更多考虑的是节约成本、提高基础设施的利用率。但是,技术是变化的^[3],特别是计算机技术和网络技术处于变化的时代。可以预测的是未来云端的应用不会离本地太远,因为离本地太远,多个网络环境交融

造成的不可控因素大于少数几个网络交融的不可控因素。这是技术发展到尽可能实时所必须的一个路径。目前,云计算针对低速和没有延迟要求的应用是可以满足的,而对于及时响应要求的用户是无法满足的。目前的云服务和当时出现的网格计算一样,是没有统一标准的,那么一旦发生技术变化时,图书馆是否能保证业务平滑地迁移^[2]?

(2) 数字内容的安全、读者信息的隐私、读者使用过程的记录的相对独立等是应用从一个云服务迁移到下一个云服务应用最基本需要保护的部分。为了对这部分进行保护,图书馆是否需要建立基础设施对内容进行长期的保存、检查、实时更新、修复、迁移呢?如果要做到这些,图书馆的IT设施也是必不可少的。如何做到这一点呢?传统的纸介质只需要保存MARC记录就可以,而数字内容仅仅保存MARC记录是不够的。这点图书馆在考虑自己的IT基础系统架构时需要认真思考并作出战略规划。

(3) 如果租借的是隔离型的云服务(即每个租借者的用户数据自己进行管理),那么图书馆要培养掌握这种云平台操作和数据处理能力的IT人员,特别是需要特定人员对云服务产生的日志文件进行收集、分析和整理。所以,数据分析师的培养成为对图书馆业务发展的新要求。

参考文献

- [1] INDUS. Amazon Apologises for Cloud Computing Fault [OL]. [2012-02-10]. <http://www.gev.com/2011/04/amazon-apologises-for-cloud-computing-fault/>.
- [2] BOOTH N. Ten cloud computing problems that could break an implementation [OL]. [2012-02-10]. <http://www.cloudpro.co.uk/cloud-essentials/1745/ten-cloud-computing-problems-could-break-implementation>.
- [3] 孙卫.图书馆在云时代的思考[J].数字图书馆论坛,2009(6).
- [4] RAPP D. With U of Tennessee Rollout, OCLC's Cloud-Based ILS Enters Early-Adoption Phase [OL]. [2012-02-10]. http://www.libraryjournal.com/lj/home/885999-264/with_u_of_tennessee_rollout.html.csp.
- [5] BINNING D. Top five cloud computing security issues [OL]. [2012-02-10]. <http://www.computerweekly.com/news/2240089111/Top-five-cloud-computing-security-issues>.
- [6] Southern African Internet Governance Forum [OL]. [2012-02-10]. <http://www.apc.org/en/system/files/1.CloudComputing.pdf>.

作者简介

孙卫,中国科学技术信息研究所,计算机高级工程师,中国计算机学会高级会员,北京通信学会理事。目前主要在知识组织、知识挖掘、知识处理技术上进行研究,在科学出版社(中国科技出版传媒股份有限公司)从事数字出版的相关工作。E-mail: sunwei0125@gmail.com

Factors to be Considered When the Cloud Services Support Library Operations

Sun Wei / Institute of Science & Technology Information of China, Beijing, 100038
/ Science Press, Beijing, 100717

Abstract: Many libraries began to pay attention to cloud computing and cloud services. Few libraries think about the factors when the cloud services support library operations. People began to think about there are the business interruption, cloud data and system security and level and ability for cloud service providers, when a few incidents of cloud failure happened.

Keywords: Cloud computing, Cloud services, Cloud failure, Cloud security

(收稿日期: 2012-02-17)