

基于“云服务”架构的国家科技文献服务系统的设计与实践*

□ 吴广印 / 中国科学技术信息研究所 北京 100038

摘要: 文章主要目的是介绍由作者提出的“基于云服务的国家科技文献服务系统架构”,同时针对科技文献服务系统对“云服务”的基本概念、模式、本质作了简单阐述。同时为了更好地说明问题,文章还对我国科技文献服务系统的发展历程和现状作了简单回顾。最后通过对湖南省科技文献与创新资源共享服务平台的开发实施及架构作了简单介绍,是对文章提出的系统架构的具体实践和验证。

关键词: 云服务, 系统架构, 文献共享服务, 私有云服务, 公有云服务

DOI: 10.3772/j.issn.1673—2286.2012.03.001

1 “云服务”

介绍“云服务”之前先简单介绍一下云计算的基本原理。云计算 (cloud computing), 是一种基于互联网的計算方式, 通过这种方式, 共享的软硬件资源和信息可以按需提供给计算机和其他设备。云其实是网络、互联网的一种比喻说法。云计算的核心思想, 是将大量用网络连接的计算资源统一管理和调度, 构成一个计算资源池向用户按需服务。提供资源的网络被称为“云”。狭义云计算指IT基础设施的交付和使用模式, 指通过网络以按需、易扩展的方式获得所需资源; 广义云计算指服务的交付和使用模式, 指通过网络以按需、易扩展的方式获得所需服务。这种服务可以是IT和软件、互联网相关, 也

可是其他服务。云计算按照服务对象的不同, 一般分为公有云和私有云两大类, 前者指的是面向广域范围内的服务对象的云计算服务, 一般具有社会性、普遍性和公益性等特点, 而后者一般是指社会单位为自身需要所建设的自有云计算服务模式, 一般具有行业性特点。

通过云计算方式提供服务能力的服务称之为“云服务”, “云服务”让用户可以通过因特网存储和读取数据, 甚至可以直接调用不同的“云应用”直接嵌入自己的应用服务系统, 不需要直接的任何软硬件系统, 可以最大限度地共享硬件、系统、数据、和应用服务。基于“云服务”的信息服务商可以把自已的服务能力像水和电一样让外部随需使用, 同样用户也可方便地使用这些云服务, 这就是云服务的基

本服务模式 and 特点。

“云服务”带来的一个重大变革是从以设备为中心转向以信息为中心。设备包括应用程序不再为用户所重点关注, 只是服务中的一个配件, 而信息及人们在信息中的投资则是必须要长期保留的资产 (这一点对我们的科技信息服务更显而易见, 我们更换过无数次的设备、操作系统甚至应用服务系统, 但所有的数据却在一直使用)。无论多么新颖的、目前甚至可以是相当昂贵的前端硬件设备都会过时, 有的甚至会很快过时, 变为一文不值, 而云服务上的信息内容, 不仅不会过时, 而且时间越久越有价值。VMWare的CEO Paul Maritz解释说: “今后在云上每个人都将会会有一个伴随终生的个人数据体, 这样的个人数据体不会被捆绑到任何一种机器上,

* 本文得到国家高技术研究发展计划 (863计划) 重大专项支持, 课题编号为: 2011AA01A206。

随着机器的过期失效而失效。”硬件的过时, 应用软件的过时, 在“云服务”上都不再是一个问题, 这也是“云服务”的本质。

“云服务”的优势之一就是规模经济^[6]。利用“云服务”供应商提供的基础设施(软硬件、数据等), 为不同用户或开发者提供更好、更便宜甚至更可靠的应用。甚至有些用户应用能够直接利用云服务的全部资源而无须要求公司投资同类的任何物理资源和应用开发。“云服务”遵循一对多的模型, 与单独的桌面程序部署相比, 成本极大地降低了。“云服务”通常是“租用的”, 以用户的适用范围和规模为基础付费, 而不是采购系统(包括计算机软硬件), 这意味着用户将使用更少的前期投资得到一个更可预知的业务费预算。对开发者而言, 升级一个基于“云服务”的应用比传统的桌面系统或本地应用更容易。只需要升级集中的应用程序(或购买“云服务”商所提供的升级即可), 应用特征就能快速顺利地得到更新。有了“云服务”, 服务中心一个改变就能影响运行应用的每一个用户, 这大大降低了开发者的工作量^[5]。

基于“云服务”的应用开发最大的不足就是给所有基于Web的应用带来一个显而易见的问题: “安全性”^[3]。长期以来, 业内专家一直对基于Web的应用所存在的潜在的安全风险给予高度重视。由于这一原因, 许多用户仍坚持将应用、数据和相关操作都保持在自己的掌控之下, 拒绝接纳“云服务”这一新的服务模式。为了解决这一问题, 我们针对我国科技文献服务系统的特点, 针对性地提出了“公有云”和“私有云”及“区域云”相结合的系

统架构, 本文将在后面的内容进行阐述。

2 我国科技文献服务系统的发展历程及现状

为了便于大家对本文提出的服务架构的技术背景有更好的理解, 现对我国科技文献服务系统的发展历程及现状作简单介绍。

2.1 系统发展历程

我国科技文献服务系统从技术层面上经历了单机、广域网、互联网三个阶段的发展过程。

单机时代基本上自上世纪80年代初期, 终于80年代末期。以微型计算机硬件为主体, 软件有自己开发、消化引进等类型, 典型的单机情报检索软件为Mini ISIS、CDS/ISIS和ISTIC/ISIS, 数据内容基本上是单位自建数据库, 以编目数据为主; 典型应用单位为各级科技情报研究所和一些大中型图书馆。

广域网时代基本上起源于上世纪80年代末期, 终于90年代末期。计算机硬件以中小型计算机为主, 典型机器为IBM43系列和DEC系列中小型计算机; 国内软件主体为引进消化, 最为代表的是列入国家“七五”攻关项目的“大型情报检索软件的汉化系统”, 国外应用主要是采用以Dialog为主的国外联机情报检索系统, 使用费用昂贵(包括信息服务和国际线路通讯费); 国内信息服务内容以自建为基础, 购买国外裸数据为辅。这一阶段的服务系统可以实现基于本机的数据资源共享(多用户访问), 但远程服务价格昂贵。能提供这种服务的国内单位基本上是国家

家部委级科技情报服务部门, 典型代表为中国科技情报研究所、机械部科技情报研究所、化工部科技情报研究所等。

互联网时代自上世纪90年代中期至今, 中国科技信息服务互联网时代到来的特征为中国科学技术信息研究所获取国际168网段地址并在原国家科委支持下启动的全国科技信息系统共建共享的www.chinainfo.org系统。Chinainfo可以说开中国科技信息服务互联网服务之先河, 得到各级科技管理部门的高度重视, 当时基本上每个省市都有Chinainfo的分中心。这一时代的计算机硬件基本上以小型计算机和微机服务器为主, 软件架构为基于Internet的HTTP协议应用, 又称三级架构体系, 即数据服务器、应用服务器和浏览器。和广域网时代相比, 这一时代的科技信息服务的服务和使用成本都有很大下降, 主要体现在远程服务通讯服务费用的下降和用户使用范围的扩大, 同时大大降低了信息服务部门进入这一领域的技术和资金门槛。这一时期基本上所有信息服务部门都开始对外提供信息服务, 只是在服务层次上有所区别, 我国的科技信息服务得到快速发展; 信息内容也更加丰富, 信息服务支撑平台也更加成熟和满足用户需求, 关系数据库系统、全文检索系统、各类内容管理系统等种类繁多。真正的信息资源共建、共享服务的时代已经到来。

2.2 系统现状

2004年7月3日, 国务院办公厅转发了科技部、发展改革委、教育部、财政部《2004-2010年国家科技基础条件平台建设纲要》, 纲要任

务第一条明确提出：构建和完善物质与信息保障系统。制定科学、合理、统一的技术标准和规范，研究开发相关技术，对现有的大型科学仪器、设备、设施、科学数据、科技文献、自然资源等进行整合、重组和优化，充分利用国际资源，加快实现资源的信息化、网络化，建立适当集中与适度分布相结合的资源配置格局。在这一“纲要”精神指导下，各部委、省市相继提出开发建设自己的科技条件保障平台，其中包括文献共享平台、大型仪器设备共享、知识产权交易、科技成果管理等科研条件保障子平台；科技文献共享平台基本上由各地科技信息服务部门承担。到“十一五”结束，各地文献保障平台相继建设完成，多数通过地方科技管理部门验收。自2010年初开始，部分科技信息服务部门开始从文献共享服务平台向为产业科技创新服务的产业服务平台升级。北京万方软件公司多年来一直从事科技文献服务平台建设的技术开发与实施工作，承担了大部分省市科技文献共享服务平台的技术开发和技术支持工作。其中包括云南、广西、福建、湖南、湖北、河南、山东、河北、天津、上海、黑龙江、吉林、辽宁等省级文献共享平台的设计开发工作。

3 基于“云服务”架构的科技文献服务系统的总体设计

北京万方软件公司多年来一直从事科技信息服务系统相关的技术研究开发工作，开发出了系列相关产品，包括非结构数据资源管理系统RMS、万方数据资源整合服务平台、科技文献自动分类与摘要服务

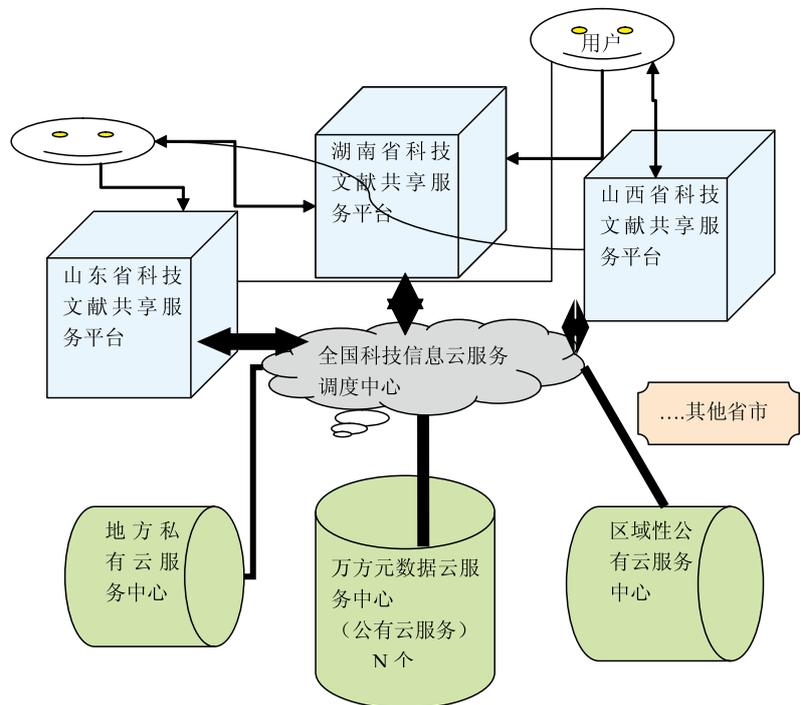


图1 基于“云服务”的科技文献服务系统总体架构图

系统、万方学术搜索、科技创新文献共享支撑平台等系列产品和服务系统。经过近十年的研究与开发实践，结合目前承担的国家“863”计划重大专项“以科技文献为主的搜索引擎研制”部分成果，我们提出了基于“云服务”的科技文献服务系统总体架构，并通过部分省市系统建设得到实施验证。

图1是我们在多年研究开发基础上设计提出的基于“云服务”的科技文献服务系统总体架构图。下面将对这一系统架构作详细功能解释说明。

万方元数据云服务中心：该中心是本系统架构的核心，它包括规范化的元数据仓储中心、相关知识库中心和管理这些数据的基于Web Service架构的资源管理与服务系统^[1]，该系统为万方软件自主研发，也是“863”专项支持的“科技文献搜索引擎”的核心系统。其中：

元数据仓储中心包括中外文期

刊、会议、学位论文、专利、标准、法律法规、科技成果、科技人物、机构等以事实数据为基础的元数据仓储，该仓储中心的数据规范原则，以本人提出的“知识获取五要素”为指导思想。该数据仓储数据规范的主要工作目标是解决科技信息服务中的人物重名和机构名称变迁、机构合并等引起的“查全/查准”问题^[2]。目前该仓储的元数据记录数达2.5亿规模，几乎涵盖所有科技文献所涉及的元数据记录。

知识库中心：包括知识获取五要素中涉及的学科、人物、主题、机构、基金等相关知识库，其中包括420多万的主题知识库和1200万作者相关的知识库，其中作者的科研合作网络和学术网络知识库是通过数据挖掘及其相关技术由计算机自动生成的。

资源管理与服务系统：万方资源管理与服务系统（以下简称R Service）是万方软件完全自主

开发的非结构化数据库管理系统,可对各类信息资源进行管理,支持嵌入式COM组件调用和标准Web Service服务,是“863”专项支持的科技文献搜索引擎的核心,是真正意义上的情报检索系统^[1]。在元数据服务中心中R Service负责元数据的接收、存储、索引,并提供标准的Web Service服务。R Service具有分段索引、多库检索、自定义交换格式等特殊功能,尤其在支持海量数据分布式检索服务方面。关于R Service详细技术特点,本文不作详细介绍。

在本架构中,万方元数据云服务中心属于公共云服务中心范畴,它除了管理万方软件自己的仓储数据外,还可以为用户提供可以共享服务的数据服务。目前该中心支持15种标准格式的元数据交换,涵盖期刊、会议、图书、方志、报告、视频等科技文献数据。同时该服务中心属于本架构方案中的最底层,除了网络和系统上的安全措施外,对存储在中心的所有数据均采用了高强度的128位加密算法进行磁盘级保护。目前该服务中心已正式对图书、情报等信息服务部门提供服务,用户通过接口直接调用本中心(中间经过云调度中心的认证和管理)提供的80多种服务。该中心提供的服务包括数据库管理、检索、数据交换、自动标引与分类、基于“知识获取五要素”的导航服务等。另外云服务中心的硬件设备也可根据用户规模、资源规模进行快速扩展部署。

地方私有云服务中心:就像第一部分介绍的那样,云服务模式的确存在潜在的用户关键数据的安全风险(起码从技术上是这样的),用户的关键数据放在公共云服务中

心,虽然节省投资,提高了效率,但毕竟放在别人那里[6]。因此我们在总体架构里面提供了对私有云服务中心的支持,私有云服务中心可提供和公共云服务中心一样的功能。不同之处是该中心还支持对其他关系数据库的管理,这样也可以方便将原有老架构的系统纳入新的云管理服务架构继续使用。私有云服务中心主要用来管理用户的本地关键数据,规避云服务潜在的安全问题。

区域性公有云服务中心:目前,部分省市信息服务机构已经开始建设区域性重点行业科技创新服务系统,这类系统不同于现有的文献服务系统,主要表现如下特征:

区域性:是为对本地区重点产业的关键业务提供支撑;

内容涵盖面广:不仅仅是科技文献服务,还包括基于互联网信息产业动态、研究报告、政策法规、专家互动、竞争情报、成果转化与服务等一系列产业信息服务。

交叉性:虽然产业服务是某个省市根据自己区域业务需求提出的,但在全国范围内和部分区域仍然存在一定的交叉性。鉴于这种情况,万方软件提出的区域性云服务的架构思想,主要是为了避免不同省市间产业信息的重复建设。

比如,山东省计划建设汽车产业服务平台,吉林等其他省市也需要这样的产业服务,我们建议以山东为主,其他省市参与共建共享。

区域性云服务中心的管理架构和万方元数据云服务中心一样,只是内容的归属有所区别,也可体现多个信息服务机构的共建共享宗旨。

云服务调度中心:云服务调度中心是本架构中“云服务”特征最为突出的部分,所有基于“云服务”

的管理、调度模式都在这里得到体现。它主要包括整个云的安全防护与认证、用户管理、服务机构管理、云数据服务中心配置、管理与调度、服务缓存、服务负载均衡等功能。无论公有云、私有云,还是区域性云服务,都通过本调度中心进行管理调度。

省市科技文献共享服务平台:省市科技文献共享服务平台是各省科技信息服务机构根据自身业务特点提出的面向本省市的科技信息服务平台,具有明显区域特征和个性化服务模式。目前大部分信息服务平台在公共性文献信息服务方面基本上都是利用万方元数据云服务中心所提供的数据和相关服务,自己拥有的特色数据存放在自己的私有云服务系统中。平台采用万方软件提供的科技创新文献共享支撑平台,该平台可直接调用万方云服务平台提供的所有服务,同时可以调用私有云的所有服务。

以上是对基于“云服务”架构的科技文献服务系统的总体设计介绍。

4 系统实施典型案例:湖南省科技文献与创新资源共享服务平台

湖南省科技文献与创新资源共享服务平台(HNSTI)是由湖南省科学技术厅组织,湖南省科学技术信息研究所具体承担,围绕政府相关部门和各市州科技局、中央在湘有关单位、科研院所、大专院校、企业等多层面需要,建立的以数字资源为主的“一站式”科技文献与创新资源服务保障体系。该项目的主要宗旨是围绕湖南省科技发展需要,按照“建设、整合、共享、完善、

提高”的原则,紧扣“与时代发展同步、与科技工作同拍,与读者利益相符”的工作理念,利用现代网络技术整合集成湖南省自有的科技文献信息资源,构建门类齐全、结构合理、面向全省、服务中部的科技文献信息资源协同保障体系和网络快速传递服务体系。平台的建立旨

在面向湖南省广大科技工作者和科技管理部门,为社会各界充分利用科技文献信息资源提供快捷便利的服务,为湖南省科技、经济与社会发展提供有效的科技文献信息支撑。湖南省科技文献与创新资源共享服务平台的主要功能如图2所示。

湖南省科技文献与创新资源共享服务平台的开发和实施由北京万方软件有限公司承担。该系统的主要功能说明如下:

a.一站式科技文献检索服务:该服务的主要功能是利用万方元数据云服务中心提供的元数据检索服务、人物科研关系分析、主题趋势、原文定位和原文传递等云服务功能和本地自建特色数据库服务(私有云服务)等功能进行服务。湖南省科技信息研究所的自建特色数据库工作,也是“863”“以科技文献服务为主的搜索引擎研制”项目的子课题。利用本功能的原文定位功能,用户可以直接使用湖南省科技信息研究所购买的各服务厂商的原文资源。

◆ 自定义检索服务:系统通过私有云管理功能提供用户自建库的跨库检索等检索服务。

◆ 创新服务:该服务主要是利用万方公有云服务中心提供的“科技创新辅助决策支持系统”对外提供科技创新辅助决策服务。该服务可提供基于机构、人物、主题、学科、基金、项目等方面的研究、评估报告,格式可以是Word、HTML等格式。关于“科技创新辅助决策支持系统”可通过网络获取:<http://stads.infosoft.so>。

◆ 服务加盟:这是湖南省平台的一大特色,它允许区域专家加盟该服务平台,可以提供自己的科研成果利用平台共享,同时也可通过平台与用户互动。这一功能可以有效解决科技信息服务机构在对用户提供创新支撑服务时无专家团队支撑的系列问题。

◆ 湖南省科技文献与创新资源共享服务平台中提供了大量与业务相关的科技创新支撑服务功能,



图2 湖南省科技文献与创新资源共享服务平台主页面截图

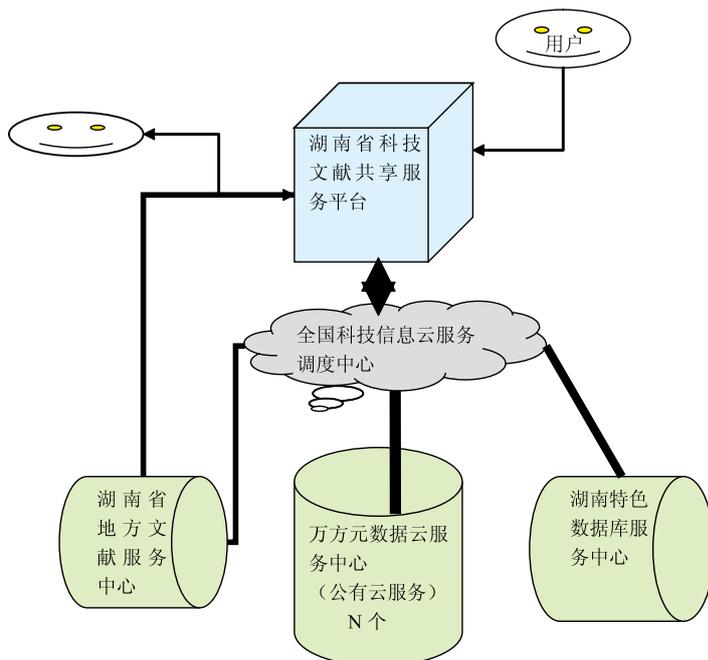


图3 湖南省科技文献与创新资源共享服务平台架构图

由于篇幅有限,且和本文主题关系不大,不再详述。详细内容可访问<http://www.hnsti.ac.cn/>。

湖南省科技文献与创新资源共享服务平台的具体系统架构如图3所示。

湖南省科技文献与创新资源共享服务平台的上述架构是以前面介绍的国家科技文献服务系统的架构为基础,通过个性化设计、开发而完成的,因此这里不再对该平台

的架构作详细说明。湖南省科技文献服务平台已经于2011年5月份正式对外服务,是支撑湖南省科技创新服务的重要平台。另外,该项目2010年底通过鉴定,2011年获湖南省“科技进步二等奖”。

5 结语

通过上述介绍可以得出结论,“云服务”模式更加适合全国科技

信息行业所倡导的“共建、共享、共用”的系统建设和服务理念,有利于减少投资,避免重复建设和快速形成我国科技文献服务网络。本文所介绍的内容属我国“863”领域“云计算”重大专项支持项目。本文介绍的架构方案已在全国多数省级科技信息服务机构中的文献共享服务平台建设中得到应用推广,其中包括湖南、湖北、云南、河南、山东、山西、天津、辽宁、吉林、黑龙江等省市。

参考文献

- [1] 吴广印.基于WEB SERVICE构架的资源共享技术研究与实现[J].情报学报,2007(6):851-857.
- [2] 吴广印.浅谈“查全/查准率”[J].数字图书馆论坛,2010(7):30-34.
- [3] 林小勇.云服务下信息用户隐私权保护[J].图书馆学研究,2010(7):99-100.
- [4] 黎春兰,邓仲华.信息资源视角下云计算面临的挑战[J].图书与情报,2011(3):17-22.
- [5] 彭静.近年来国内信息服务理念研究综述[J].新世纪图书馆,2011(1):48-51.
- [6] 卢敏.众说纷“云”(产业篇)[J].软件和信息服务,2010(8):66-69.

作者简介

吴广印(1965-),男,中国科学技术信息研究所研究员,北京万方软件有限公司总经理。研究方向:非结构数据库管理系统和中文信息检索,RMS系统的总体设计师和主要开发人员。本文“863”专项的技术负责人。E-mail:gywu@wanfangdata.com.cn

The Design and Implementation of National Scientific and Technical Literature Service System Based on the "Cloud Service" Architecture

Wu Guangyin / Institute of Scientific and Technical Information of China, Beijing, 100038

Abstract: The primary purpose of this paper is to present "National Scientific and Technical Literature Cloud Service System" based on cloud service architecture; meanwhile the paper gives brief introduction of basic concept, pattern and essence of "cloud service" from scientific and technical literature service system's point of view. For better understanding, the paper reviews the development course and present situation of national scientific and technical literature service system. Finally, to verify the system infrastructural architecture proposed by the paper, we illustrate a case study according to research and implementation of Scientific and Technical Literature and Innovation Resource Sharing Platform, Institute of Scientific and Information of Hunan province, China.

Keywords: Cloud services, System architecture, Literature sharing services, Private cloud services, Public cloud services

(收稿日期: 2012-01-04)