

基于云计算的科技资源数据中心架构设计

林 珠 陈树敏 罗俊博

(广东省科技基础条件平台中心, 广东广州 510033)

摘 要: 以科技资源共享为目的, 结合云数据中心的特点, 将云计算引入科技资源数据中心建设, 提出基于云计算的科技资源数据中心的体系架构, 并分别对该架构中的资源及虚拟化层、中间管理层及应用服务层中的关键技术进行研究和阐述, 为科技资源数据中心建设提供有效的解决方案, 有效促进科技资源的开放共享。

关键词: 云计算; 数据中心; 科技资源; 资源及虚拟化; 资源共享

中图分类号: TP392

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2015.04.007

Data Center Architecture Design of Scientific and Technological Resources Based on Cloud Computing

Lin Zhu, Cheng Shumin, Luo Junbo

(Guangdong Science & Technology Infrastructure Center, Guangzhou 510033)

Abstract: Combined with the characteristics of cloud data centers, and cloud computing data centers are introduced to scientific and technological resources building in this paper, for the purpose of sharing resources. It is proposed of the infrastructure of scientific and technological resources data centers based on cloud computing, and it is researched on key technology of resources and infrastructure virtualization layer, middle management and application service layer separately, and so, effective solutions is provided for scientific and technological resources data centers construction, and open sharing of scientific and technological resources is promoted effectively.

Keywords: cloud computing, data center, science and technology resources, source and virtualization, resource sharing

1 引言

云数据中心是基于云计算的数据中心。它是运用云计算技术, 整合信息技术设备与基础设施, 能够更为动态地调动资源, 更加智能地管理基础设施, 同时, 对外提供软件即服务 (Software

as a Service, 简称 SaaS) 的新一代数据中心。数据中心通常是指在一个物理空间内通过计算机系统、存储、网络、安全监控等其它设备实现信息的集中处理、存储、传输、交换以及管理^[1]。云计算是通过网络通信技术将成千上万台电脑和服务器的连接成一片电脑云, 通过虚拟技术形成 IT 资

作者简介: 林珠* (1983-), 女, 广东省计算中心高级工程师, 主要研究方向: 数据挖掘与云计算; 陈树敏 (1982-), 女, 广东省计算中心工程师, 研究方向: 虚拟现实、并行计算与科技资源调查; 罗俊博 (1981-), 女, 广东省计算中心项目管理师, 研究方向: 科技资源数据整合、资源标准化建设。

基金项目: 广东省重大科技专项“现代信息服务业云平台关键技术研究及其产业化”(2012A080104019); 广东省科技计划项目“科技基础平台建设”(粤财教[2013]112号); 广东省行政类审批研发机构项目“广东省服务计算工程技术研究开发中心”(2011B080100031)。

收稿日期: 2014年11月7日。

源池，用户通过网络以按需服务、易扩展的方式获取所需资源^[2]。云数据中心具有高效低成本、虚拟化环境、灵活扩展、可靠等特点，它既可满足对带宽、存储和计算能力的突发需求，又可提供丰富的应用，满足科技资源集聚共享的需求^[3]。

本文针对科技资源共享现状和存在问题，提出建立开放共享、数据准确、体系完整的基于云计算的科技资源数据中心（下面简称“科技资源数据中心”）。该数据中心拥有高性能计算能力、大容量存储空间、高速数据处理能力，是一个高效、灵活、可扩展的科技资源数据中心，可解决资源共享过程中存在的跨地区、跨部门、跨学科、分布式资源共享不健全问题，进一步提高资源的开放服务水平，为科技基础条件共享与服务平台承载和优化公共服务提供信息网络和技术支撑。

2 意义和必要性

我国在“十一五”“十二五”交替之际着手打造“中国科技资源共享数据中心”，整合分散的科学数据资源，构建面向全社会的共享服务体系，实现对科学数据资源的规范化管理及其高效利用^[4]。但是，科技资源共享工作仍面临着一些突出的挑战，科技平台资源共享机制尚不健全，开放服务水平有待进一步提高，对重大科技创新活动和企业技术创新的支撑能力还不够强；科技资源配置与创新需求有效衔接不够，科技资源配置及开放共享围绕企业的需求设计不足；各类科技资源跨地区、跨部门、跨学科现象未得到有效解决^[5]。

在现阶段，科技资源共享利用不足，主要表现为：科技服务大多是根据自身的业务和职能部门的需求进行的，缺乏战略层面对数据的把握；大量的原始科技数据有待分析、提炼和挖掘，无法为科技管理和决策带来进一步的价值；大量的科技资源受地域与行业的限制，无法实现共享与重复利用，导致重复建设；各区域的数据标准与应用系统不一致，导致无法集成和利用等^[6]。

为解决现有科技资源共享方面的不足，充分

提高科技资源的利用率和服务效益，应整合现有的、分散的IDC资源，发挥云数据中心高效低成本、虚拟化环境、灵活扩展、可靠等特点，快速搭建科技资源数据中心，这将有利于进一步提高中国社会信息化水平，推动IT产业的整体升级，达到节能减排的总体目标，对科学发展有着重大的意义^[7]。方案将采用自建云计算数据中心的模式，而非采用租赁云服务的模式，主要是因为科技资源云数据中心是长期、持续的科技建设目标，数据中心建成后面向大范围的政、学、产、研、用以及各行各业广泛的用户群体，同时也将在该数据中心中扩展其他的科技研究工作，资源使用率较高，因此根据需求该云数据中心适合采用自建的模式。

3 基础架构

针对科技资源存在的跨地区、跨部门、跨学科资源分布情况，提出基于“物理上合理分布、逻辑上相对集中”的理念，整合现有科技基础条件资源，通过信息共享来带动实物共享。该中心包括：大型仪器、科技文献、科技成果、实验室共享体系、自然资源、科学数据库等资源和服务信息的数据库，成为跨资源类别联动、跨不同主管部门联动、市区联动、部市联动的科技资源数据中心。同时，建立不同科技资源信息间的动态关联，制订分级、分类的数据共享规则，使之成为本市科技资源信息采集、加工、发布、分析的数据中心，为广东省各高校、科研院所、企业等提供坚实的科技资源信息支撑，甚至成为珠三角，乃至国家科技资源共享数据中心的重要节点。

基于云计算的科技资源数据中心建设基础架构，如图1所示，主要分为资源及虚拟化层、中间管理层和应用服务层。

资源及虚拟化层将底层的物理资源作为统一的整体进行考虑，采用虚拟化屏蔽底层各硬件资源的异构性，即不管这些硬件资源是来自不同的厂商还是使用不同的内核，都使用虚拟化进行统一的管理，并实时监控这些资源的性能、负载等

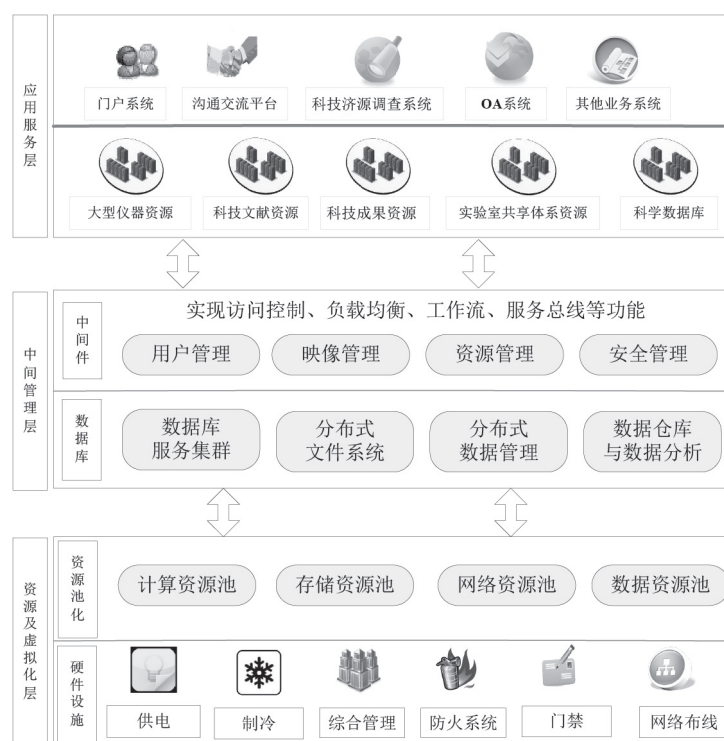


图1 科技资源数据中心基础架构图

相关数据^[8]，同时，能做到物理硬件设备的自动发现、添加和维护，使之具有良好的扩展能力，从而构建出一个能灵活组装、自适应、拥有巨大计算能力、容错能力的虚拟集群。

中间管理层主要负责云平台自身的服务功能及总线功能实现。该部分通过管理大量的中间件，实现平台的访问控制、负载均衡、工作流、服务总线等功能。中间件的管理主要可以分为用户管理、映像管理、资源管理、安全管理等。用户管理主要有用户身份管理、用户许可、用户请求管理、使用计费等功能；映像管理主要有映像创建、映像部署、映像库管理、映像生命期管理等功能；资源管理主要有负载均衡、故障检测、故障恢复、临视统计等功能；安全管理主要有身份认证、访问授权、综合防护、安全审计等功能。

应用服务层主要采用软件即服务的形式提供给科技资源共享平台的用户。该层主要围绕国家科技资源的整合情况对大型仪器资源、科技文献资源、科技成果资源、实验共享体系资源、科

学数据库等资源进行整合服务，在此基础上提供相关的应用服务，包括：门户系统、沟通交流平台、科技资源调查系统、OA系统以及其他业务系统等，平台还提供二次开发的接口、SDK等，为外界应用提供可以调用的应用适配器接口、服务适配器接口等。

基于云计算的科技资源数据中心是为科技资源共享服务的具有明显行业特色的数据中心，它与普通的企业数据中心有明显的区别，除了硬件设备和基础设施的建设以外，该数据中心还应考虑自身的特色，其中最明显的区别在于该数据中心设有专门为科技资源服务的数据资源池和数据中心应用服务层，根据科技资源的特征进行数据中心建设，更好地为科技资源共享服务，达到开放共享、创新驱动的目的。

4 虚拟化层

在科技资源数据中心架构中，将资源及虚拟化层分为两部分：一是硬件设施层，主要是指云计算的基础设施包括主机、存储、网络等；二是

资源池化，即通过虚拟化技术，将这些硬件设施进行池化后的各类资源池，主要包括：计算资源池、存储资源池、网络资源池、数据资源池等。

在硬件设施层,科技资源云数据中心采用光纤阵列（DS6310FE）作为主要的存储介质，提供4Gb/s FC连接链路，扩展为SAN存储系统，用SAN Appliances的专用虚拟化引擎实现存储网络层的虚拟化。具体控制形式采用带内方式直接位于主机服务器和存储设备的数据通道中间。

资源池化是通过虚拟化技术将底层的各种硬件设备进行更细粒度的划分，但同时又实现这些细粒度资源整合与管理，使得资源更加灵活地提供各种服务。随着科技资源整合工作的进展，资源的池化粒度会越来越细，对硬件资源进行虚拟化，容易满足科技资源工作的需求，支持日后的扩展。

其中，计算资源池和存储资源池，主要采用Red Hat Enterprise Virtualization（RHEV）虚拟化解决方案，通过RHEV Hypervisor技术将物理机关联，形成统一资源池，然后由RHEV Manager管理工具进行虚拟机的自由划分。RHEV以强大的基于内核的虚拟机（KVM）系统管理程序和oVirt开放虚拟化管理平台为基础,实现资源的离散化，使资源以更小粒度、更灵活的态势提供给云计算管理层。

网络资源池采用虚拟交换机技术实现网络扩展和隔离需求^[9]，通过虚拟交换机技术，将原来需要运行在独立网络中对安全要求高的业务运行在统一的网络资源池中，实现网络资源的灵活调度，以及数据安全和节能减排，网络中的虚拟交换机之间是彻底隔离的，它们有各自独立的二层和三层协议栈和进程，有各自独立的管理人员，虚拟交换机之间是无法通过逻辑配置实现联通的^[10]，同时，由于它们的软件进程是完全独立的，当某个虚拟交换机出现问题的时候，不会影响到别的虚拟交换机，实现了故障的完全隔离。

数据资源池是科技资源云数据中心的特色资源池，它依据目前的科技资源整合情况而设定，该数据中心分为大型仪器资源、科技文献资源、

科技成果资源、实验共享体系资源、科学数据资源等，而这些数据来源并非完全统一到固定的存储地址，而是采用“逻辑上相对统一，物理上相对独立”的原则进行资源的整合。其采用的技术主要有两种：一是对于数据中心自主拥有的数据资源，将存于数据中心的数据资源池，其中包括结构化的数据库集群和非结构化的分布式文件系统；二是采用基于SOA/ESB的数据交换平台的方式，将企业独立拥有的资源，通过交换平台对外提供资源共享服务，实现逻辑上的统一，该交换平台可采用消息队列中间件、工作流引擎中间件和消息适配器中间件等，实现跨平台的数据资源对接。数据资源的使用用户也来自多个主机服务器，因此，为了应用服务层可以很好地调用这些数据资源，在存储虚拟化层应实现网络级别的存储虚拟化，提供多对多的访问模式，即多个服务器可对构成弹性资源池的多个异构存储设备进行访问。

5 中间管理层

科技资源云数据中心与其他云数据中心的不同之处在于，将科技资源全部应用部署在公共云上是不现实的，有些企业拥有的资源不希望完全共享，或者涉及某些安全性的问题，因此，采用公有云和私有云混合的模式。

对于私有云的管理将采用曙光Gridview v2.0服务器综合管理系统，通过该系统对数据中心进行统一监控，集中管理。通过该系统，管理员可以直接查看数据中心中各服务器终端的运行状态，提供各种状态视图和性能视图；可以实现故障检测和报警；可以查看各种统计列表；可以实现作业提交和负载均衡以及用户管理等功能。

而对于公有云以及混合云的管理，将采用云总线的管理方式。云总线将云环境下的信息的集成和服务进行整合，是一种不依赖于特定产品、特定语言、特定平台的通讯联系基础结构，是实现数据无障碍交换的枢纽。云总线的设计包括3个层次，分别是服务适配器层、总线层和应用适配器层。服务适配器层主要提供云环境中服务

跟总线通信的适配器,包括.NET平台服务适配器、J2EE平台服务适配器等;总线层主要提供服务注册、服务查找、服务监控、消息路由、安全验证等功能;应用适配器层主要提供云环境中软件应用跟总线通信的适配器,包括.NET平台应用适配器、J2EE平台应用适配器等;软件应用通过应用适配器跟总线进行交互,总线则通过服务适配器与具体服务进行交互。另外,出于性能和安全考虑,本总线将对用户的请求做特殊处理验证,从用户的身份标识符中判断用户是企业内部用户还是外部用户,如果是资源内部用户,将不对传输数据进行加密等操作,以提高性能;如果是资源外部用户,则对传输数据进行加密等操作,以提高安全性要求。

中间管理层建设采用混合模式,还体现在采用数据库服务集群与分布式文件系统相结合的方式实现资源数据的存储。由于结构化数据对I/O的要求很高,且通常以裸设备的方式来放置,一般会采用容量大、性能好的存储设备(如FC/FCoE)来整合。对于系统中更多的对I/O要求相对较低,但数据量巨大的非结构化数据,可以采用NAS或分布式文件系统(如HDFS)来整合。

6 应用服务层

应用服务层是科技资源数据中心的特色所在,它主要将资源及虚拟化层中的数据资源池进一步细化,提供多样化的服务和应用,以实现科技资源的开放共享。

在资源数据的整合方面,采用应用整合的方式,着眼于科技服务的类型进行整合。根据广东省的科技资源拥有现状,整合后的资源包括:大型仪器资源、科技文献资源、科技成果资源、实验共享体系资源、科学数据库等。这些应用都以用户、功能、数据类型为基础,对开发技术进行统一规划,整合相同的应用程序和功能,并达到负载均衡。通过资源的汇聚和整合,为科技资源的开放共享打好坚实的基础,在此基础上扩展了广东省科技资源共享网、广东省科技资源需求调

查、广东省实验室体系管理信息统计系统等围绕科技资源的一系列应用。

在资源软件的应用方面,根据应用整合的结果,在科技资源云数据中心门户系统中展现各类资源,同时根据资源的分类,个性化定制相应的软件,实现SaaS的应用效果。用户可以在该门户系统中直接使用完全公开的信息资源,也可以通过该门户对未完全开放的资源进行申请使用,主要包括各类科技信息资源,以及数据中心本身具有的计算资源,如在进行科研实验时除了利用门户系统提供的免费计算节点外,还可以填写申请表向数据中心申请指定CPU数量的计算资源。

7 结论与展望

科技资源数据中心针对科技资源共享现状和存在问题提出基础架构,在资源及虚拟化层采用红帽虚拟化解决方案和虚拟交换机技术实现计池化,通过虚拟交换机技术实现硬件资源的池化;在中间管理层通过曙光Gridview v2.0服务器综合管理系统管理私有云,并通过云总线设计实现公有云以及混合云的管理;在应用服务层围绕资源数据整合和资源软件应用实现资源的充分利用。

该架构设计着力解决科技资源共享中存在的问题,与其他数据中心相比具有明显的适用性和优势。在资源及虚拟化层设有科技资源数据资源池,在中间管理层采用混合模式以适应科技资源异构性、分布性和共享权限不统一的情况,在应用服务层针对科技资源扩展多样化的应用等,整合分散的科技资源,实现资源的先整合再按需分配,能够有效促进资源共享服务体系的构建,促进科技资源的共享水平。

在科技资源数据中心构建过程中,存在一些亟待解决的问题,如云数据中心尚未有成型的建设标准;在科技数据资源池实现过程中,可研究冷、热资源的不同结合方式,以促进资源池化,达到更好的资源整合效果。

(下转第79页)

- 析:1994-2007[J]. 科学学与科学技术管理, 2009, 30 (7): 32-37.
- [2] 孙海生. 国内图书情报研究机构科研产出及合作状况研究[J]. 情报杂志, 2012, 31(2): 67-74.
- [3] 袁玉琳. 某学术机构发表的学术论文分析研究[D]. 重庆: 重庆大学, 2012.
- [4] Okubo Y. Bibliometric Indicators and Analysis of Research Systems: Methods and Examples[J]. OECD science, technology and industry working papers, 1997(1): 24-31.
- [5] 朱军文. 基于 SCIE 论文的我国研究型大学基础研究产出表现研究: 1978-2007[D]. 上海: 上海交通大学, 2009.
- [6] 谢彩霞, 刘则渊. 科研合作及其科研生产力功能[J]. 科学技术与辩证法, 2006, 23(1): 99-102.
- [7] 赵君, 廖建桥. 科研合作研究综述[J]. 科学管理研究, 2013, 31(2): 117-120.
- [8] 朱丽娟, 李丽娜. 科研合作计量指标研究述评[J]. 情报杂志, 2013, 32(6): 76-79, 52.
- [9] 易明, 毛进, 曹高辉, 等. 互联网知识传播网络结构计量研究[J]. 情报学报, 2013, 32(1): 44-57.
- [10] 吴金闪, 狄增如. 从统计物理学看复杂网络研究[J]. 物理学进展, 2004, 24(1): 18-46.
- [11] 李亮, 朱庆华. 社会网络分析方法在合著分析中的实证研究[J]. 情报科学, 2008, 26(4): 549-555.
- [12] Woo S H, Kang D J, Martin S. Seaport Research: An Analysis of Research Collaboration using Social Network Analysis[J]. Transport Reviews, 2013, 33(4): 1-16.
- [13] 冯祝斌, 华薇娜. 基于 Web of Science 的中国图书情报学科研成果调研与分析[J]. 图书情报知识, 2013 (4): 42-50.
- [14] Tomassini M, Luthi L. Empirical Analysis of the Social Network of Scientific Collaborations[J]. Physica A: Statistical Mechanics and Its Applications, 2007, 385 (2): 750-764.
- [15] 刘利, 宋歌, 袁曦临, 等. 论文合著视角下的科研机构合作网络测度分析——以我国电信学科为例[J]. 现代情报, 2014, 34(1): 94-99.
- [16] 张玉涛, 李雷明子, 王继民, 等. 数据挖掘领域的科研合作网络分析[C]//第25届全国计算机信息管理学术研讨会论文集. 2011: 327-335.
- [17] 苗蔚. 第十五个世界癌症日到来及《世界癌症报告》发表[J]. 中华医学杂志, 2014, 94(12): 888.
- [18] 冯祝斌, 赵丹群. 我国图书情报学研究机构合作网络演变分析(2002-2012)[J]. 情报杂志, 2014, 33(8): 92-98.
- [19] 刘军. 整体网分析讲义: UCINET 软件实用指南[M]. 上海: 上海人民出版社, 2009: 99.
- [20] 李辉, 赵海, 徐久强, 等. 基于 k-核的大规模软件宏观拓扑结构层次性研究[J]. 电子学报, 2010, 38(11): 2635-2643.
- [21] 李彪. 网络事件传播空间结构及其特征研究——以近年来 40 个网络热点为例[J]. 新闻与传播研究, 2011 (3): 90-99.
- [22] 邱均平, 温芳芳. 作者合作程度与科研产出的相关性分析——基于“图书情报档案学”高产作者的计量分析[J]. 科技进步与对策, 2011, 28(5): 1-5.

(上接第 44 页)

参考文献

- [1] 余侃. 云计算时代的数据中心建设与发展[J]. 信息通信, 2011(6): 116-102.
- [2] 赵吉志, 李金, 姚萃南. 云计算数据中心及标准化发展[J]. 标准化研究, 2011(3): 30-34.
- [3] 陈志峰, 王洁萍, 李海波, 等. 云计算数据中心参考架构及标准研究[J]. 信息技术与标准化, 2013(5): 39-41.
- [4] 邹佳利, 山红梅. 基于云计算的科技资源共享问题研究[J]. 科技管理研究, 2013(6): 186-189.
- [5] 戴国强. 加强科技平台建设, 推动科技资源共享[J]. 科技信息化, 2013(4): 468-475.
- [6] 何亮, 周琼琼. 大数据时代我国科技资源领域发展探析[J]. 科技进步与对策, 2014(2): 21-23.
- [7] 陈伟, 马严, 肖波, 等. 资源整合打造高效数据中心[J]. 中国教育网络, 2011(2): 58-60.
- [8] 钱琼芬, 李春林, 张小庆, 等. 云数据中心虚拟资源管理研究综述[J]. 计算机应用研究, 2012(7): 2411-2415.
- [9] 黄大川. 云计算数据中心网络的关键技术[J]. 邮电设计技术, 2011(10): 14-18.
- [10] 李晨, 许辉阳. 云计算数据中心组网技术研究[J]. 电信网技术, 2012(6): 10-13.