

C-RHIP: 云计算模式下的区域卫生信息平台

刘振鹏^{1,2}, 王虓¹, 尚开雨², 柴文磊², 吴凤龙¹

(1. 河北大学 数学与计算机学院, 河北 保定 071002; 2. 河北大学 网络中心, 河北 保定 071002)

摘要: 卫生机构现有的分散式系统建设模式存在资源配置调度任务繁重、信息格式不统一等问题导致区域卫生信息共享能力差, 因此提出采用云计算技术构建区域卫生信息平台 C-RHIP, 具有经济规模优势、可动态扩展、计算能力可管理等特点。依托河北大学医疗信息资源开发的原型系统, 表明 C-RHIP 系统有效解决了现有系统资源利用率低、可扩展性差、数据标准不统一等问题。

关键词: 区域卫生信息平台; 云计算; C-RHIP

中图分类号: TP399

文献标识码: B

文章编号: 1000-436X(2011)9A-0179-05

C-RHIP: the cloud regional health information platform

LIU Zhen-peng^{1,2}, WANG Xiao¹, SHANG Kai-yu², CHAI Wen-lei², WU Feng-long¹

(1. School of Math and Computer, Hebei University, Baoding 071002, China; 2. Network Center, Hebei University, Baoding 071002, China)

Abstract: It was known that the existing distributed model of system building used in the health agency has suffered the problems of heavy scheduler task of resource allocation, and the inconformity of information format, etc, which has caused poor sharing ability of the regional health information. Therefore a new method was proposed, using cloud computing technology to construct the regional health information platform C-RHIP, which has economic scale advantage, dynamically-scalable ability, and managed computing ability. The prototype system was developed on the health and IT resource of HBU, and it was demonstrated that C-RHIP effectively resolved the problems of low resource using rate, poor scalable ability, and inconformity of data standard.

Key words: regional health information platform; cloud computing; C-RHIP

1 引言

从现在起到 2020 年将是我国卫生事业发展的关键时期, 基于健康档案的区域卫生信息平台的建设工作方兴未艾。建设工作中存在资源利用率低、可扩展性差、数据标准不统一等问题。云计算作为近年来新兴技术之一, 发展势头迅猛, 引起了学界及社会各界的广泛关注, 它具有经济规模优势、可动态扩展、计算能力可管理等特点, 可以很好的解

决现存问题。本文描述了以云计算技术实现区域卫生信息平台的方案架构, 并搭建了原型系统。

2 区域卫生信息平台与云计算

区域卫生信息平台(RHIP), 是连接区域内的医疗卫生机构基本业务信息系统的信息交换和共享平台, 是不同系统间进行信息整合的基础和载体。RHIP 以区域内健康档案信息的采集、存储为基础, 向居民和医疗卫生机构提供数据和资源的共享, 并

收稿日期: 2011-07-05

基金项目: 国家自然科学基金资助项目(60873203); 国家社会科学基金资助项目(10BTQ039); 河北省自然科学基金资助项目(F2010000319); 河北大学自然科学校内基金资助项目

Foundation Items: The National Natural Science Foundation of China(60873203); The National Social Science Foundation of China(10BTQ039); The Natural Science Foundation of Hebei Province(F2010000319); The Natural Science Foundation of Hebei University

能够自动产生、分发、推送工作任务清单，支撑区域内各类卫生机构开展医疗卫生服务活动。平台主要以服务居民为中心，兼顾卫生管理和辅助决策的需要。健康档案是 R H I P 共享信息的载体，是居民健康管理（疾病防治、健康保护、健康促进等）过程的规范、科学记录。是以居民个人健康为核心，贯穿整个生命过程，涵盖各种健康相关因素、实现多渠道信息动态收集，满足居民自我保健和健康管理、健康决策需要的信息资源^[1]。

制约 R H I P 发展的主要原因有：现有系统资源利用率低、可扩展性差、数据标准不统一等问题。卫生机构现有的、分散式系统建设模式，即由各个医疗卫生机构独立建设自己的系统，导致无论是硬件建设，还是应用系统开发，以及数据的采集和应用，大都处于一种低水平的、重复的无序状态^[2]。我国医疗卫生机构不但基数大，而且每年的数目都有不小的变化。如果将服务器直接部署到各个医疗卫生机构中，机构的减少和增加等变更，将使得资源的重新配置和调度等任务变得异常繁重。由于各医疗机构系统独立建设，导致业务文档、数据标准、数据类型等不统一，给数据资源共享造成很大困难。

云计算的基础硬件设施架构在大规模的廉价服务器集群上，通过多个节点之间的冗余，使系统获得高可用性^[3]。云平台的网络拓扑结构允许节点方便的添加和删除节点改变系统规模，也可以通过增强或减弱每个节点改变系统整体性能，由于虚拟化这些改变对用户而言都是透明的，使云平台具有很好的扩展性和动态配置能力。因此基于云计算的低成本区域卫生信息共享平台的系统构建，较之传统的建设模式有着明显的优势，它可以向用户集中提供共享的资源和应用，利用个用户业务高峰时间的不同，动态地部署、配置及回收计算资源，实时监控资源使用情况，并在需要时重新平衡资源分配。

3 基于云计算的区域卫生信息平台

为了解决上述问题，建立了一个基于云平台的区域卫生信息平台(C-RH I P)。C-RH I P 遵循卫生部《基于健康档案的区域卫生信息平台建设技术方案》^[2]（以下简称《技术方案》）所制订的准则和规范，依托体检中心、医院及社区卫生服务中心等医疗卫生机构，探索新的公共健

康管理服务模式，建立基于本体技术的健康知识库、个性化智能健康管理推荐服务等关键技术的公民健康档案公共服务平台，逐步形成集诊断与健康咨询指导、健康数据智能分析、健康知识库和健康信息检索为一体的新型远程医疗健康服务系统。

3.1 C-RH I P 系统架构

图 1 所示描述了 C-RH I P 系统架构。上部是云平台架构，下部是 C-RH I P 对应结构。云平台由下至上分为 4 层：构架层(Fabric)、统一资源层(U nified Resource)、平台层(P latform)、应用层(A p p l i c a t i o n)。通过高速网络将服务器连接成分布式集群即形成构造层，它包括基础硬件级的资源，例如计算资源、存储资源、和网络资源。统一资源层通过虚拟化技术将下层硬件封装成统一的资源池供上层使用，这里可以使用 X e n, V M W a r e, K V M 等虚拟机软件^[4,5]。第 3 层是平台层，相当于云平台的操作系统，这一层利用开源项目 H a d o o p, 它提供了分布式文件系统 H D F S、分布式数据库 H B a s e 和并行计算编程接口 M a p R e d u c e, 再加上负责整体资源调度的中间件如 J O 2 n a s 等^[6]，即可组建一个完整的云操作系统。第 4 层是应用层，用户可通过 W e b 页面或客户端程序调用平台所提供的功能或云平台提供者开发出的应用。

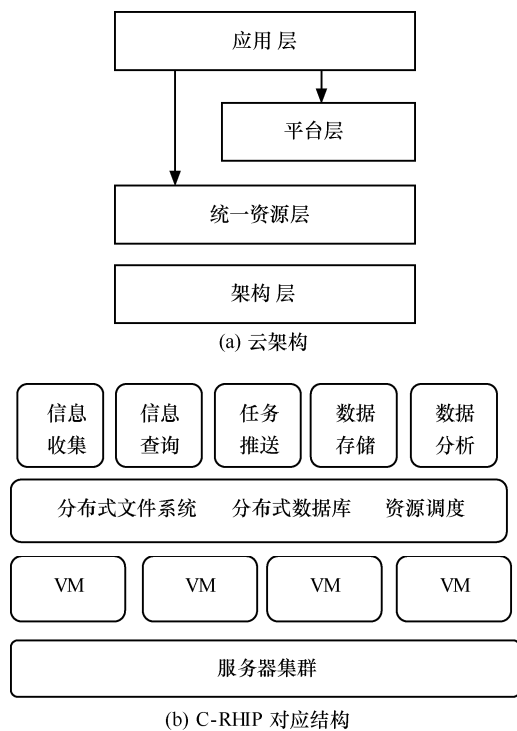


图 1 云架构及 C-RH I P 对应结构

云平台初始规模的确定是关键问题,既要保证运行性能良好又要避免资源过分闲置。C-RHIP 的初始规模定义为其覆盖的所有机构的平均需求的总和,并在此基础上增加 20%~40% [7]。这样既可以保证所有机构日常工作,又可以应对计算量突增的情况。

3.2 C-RHIP 功能

C-RHIP 的主要功能模块包括信息收集,数据存储,数据分析,信息查询,任务推送等。下面介绍各模块的具体实现。

1) 信息收集

信息收集是在区域卫生平台之上开展其他工作的基础。信息由服务终端收集,并发送到云平台。信息收集的基本内容包括个人基本信息和主要卫生服务记录,主要来源是各类卫生服务记录。根据其提供医疗服务的种类和现有设施的不同,为其配置不同的客户端软件。对于已有 HIS 系统的中心医院,在其服务器上安装代理自动检测数据库更新情况,将更新信息实时上传。对于没有 HIS 系统的医院和社区卫生中心,为其安装数据录入和上传的软件。软件具备的功能包括:①身份认证,只有经过认证的具有相应权限的人员才能使用;②数值检测,对于具有一定范围的数据进行边界检测,防止工作人员输入错误;③文档生成,自动根据采集到的数据生成符合 XDS 规范,具有 CDA 结构业务文档;④文档发送,将业务文档自动发送到 RHIP;⑤文档清除,将发送后的文档自动清除。如图 2 所示。

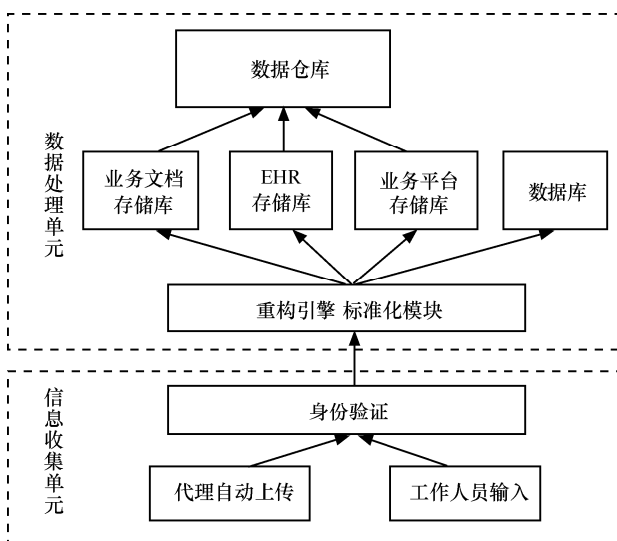


图2 C-RHIP 信息收集单元与数据处理单元

发送文件时首先通过业务人员的数字证书和 CA 认证技术与平台建立连接,通过认证后利用 VPN 技术在终端和平台之间建立虚拟安全通路,对文档加密后发送到平台。根据文档大小和网络状况的不同,对文档进行不同处理。对于体检报告、门诊病历、保健记录、疾病记录等纯文本文档采用实时传送的方式;对于医疗影像等较大的文档在网络状况良好的情况下立即上传,网络较差时则只将索引记录上传至平台,而将影像加密保存,等网络畅通或接到调用指令时再上传;在医疗下乡等连接不到网络的情况下,将文档自动加密保存,稍后再进行上传。文档上传完毕,自动删除本地保存的文档。

2) 数据处理

数据的存储与二次使用是整个平台工作的重点。参考 IBM、Google、Amazon 等公司的商业云平台的构架 [8],并结合《技术解决方案》中对数据存储和处理的要求,给出 C-RHIP 中数据处理的实现。

各医疗卫生机构信息标准不统一的问题,是区域卫生信息共享困难的主要原因,为了解决这一问题采用 2 种解决方案。第一、对于已有 HIS 系统的医院,将采集上来的数据通过标准化模块将其转换为符合《健康档案数据标准》的数据。第二、对于没有 HIS 系统的医院,在客户端提供符合《技术解决方案》规范的业务文档格式给业务人员使用。

数据缓冲区负责接收从终端发送来的业务文档。业务文档与 EHR 文档并非一一对应,需要重构引擎将缓冲区的文档转换为 EHR 文档,将业务文档与 EHR 文档分别存储于业务文档存储库和 EHR 存储库。与业务文档存储库和 EHR 存储库类似,业务平台存储库存储经过处理后的业务数据。此外,其他大量信息需要存入数据库中,包括病人主索引(MPI)、EHR 索引、健康档案摘要、标准数据、其他注册数据等。这些数据是 RHIP 数据交换、检索和统计的基础。为了对业务数据进行综合分析实现辅助决策,还需建立数据仓库。数据仓库从不同的存储库中抽取数据,并经过转换、去隐私化、装载等加工处理后,提供平台范围内各种数据分析的数据源。如图 2 所示。

3) 信息共享和任务推送

建立区域卫生信息平台的目的是实现区域内的信息共享,在完成信息的标准化存储后信息的共享也变得较为容易。用户通过 Web 形式登陆 C-RHIP 平台后即可查询自己的体检信息、就诊记录等信息,医务

人员可以通过客户端软件或 HIS 系统向 C-RHIP 查询。用户通过身份验证, 获得系统相关信息查询权限, C-RHIP 前端向数据库发送查询语句, 系统按照 EHR 文档名将文档读出返回给用户。

任务推送通过触发器监听上传的业务文档类型, 按业务类型的不同生成不同的任务并发送给相关机构。例如, 当急救中心接到急救电话后记录病人的身份及病情, C-RHIP 系统监听到这个事件, 自动生成一份急救任务单发送到病人所在社区医院, 包括姓名、性别、年龄、住址、病情、病史、过敏史、禁忌药物等相关信息, 以便急救工作快速高效地展开。图 3 描述了 C-RHIP 工作流程。

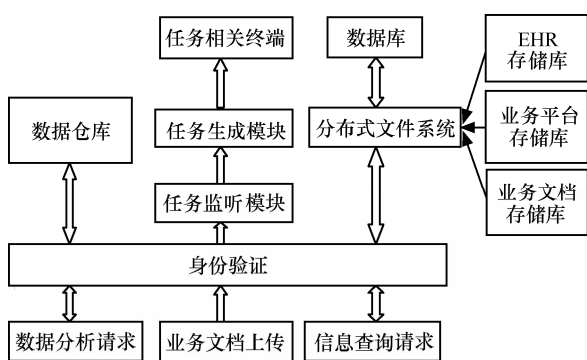


图 3 C-RHIP 工作流程

随着信息标准化工作的逐步开展和居民对医疗质量要求的提高, 可以利用 Hadoop 等软件的程序调用接口在 C-RHIP 平台上实现更多的功能。

3.3 安全性分析

C-RHIP 系统平台中存储着区域内所有居民的健康信息, 系统的安全性极为重要。

网络安全方面。当身份认证系统被入侵者通过身份冒用、滥用特权、非法夺取控制权限的手段突破后, 入侵检测就成为系统的第二道防线。C-RHIP 采取 2 种云平台下的入侵检测框架。一种是基于 DHT 的分布式入侵检测系统, 包括 DDOS pushback scheme、Worm Shield、异常和入侵系统监测系统(CAIDS)等功能模块, 分布于云平台各处的警报关联传感器将搜集到的信息汇集到警报关联模块, 生成高等级的入侵报告。另一种是基于融合的入侵检测系统, 通过对多种传感器信息的融合, 以及不同等级的信息抽取形成比单一传感器更准确的入侵报告。

系统和应用安全方面。为了保证终端系统的安全性, 在涉及相关业务的终端机上采取防毒措施。C-RHIP 引入基于多版本杀毒软件的云杀毒思想,

在 C-RHIP 外围终端机上安装代理, 每次程序执行之前, 首先生成程序的唯一标识符, 将其上传至安装在云中虚拟机的杀毒软件, 通过与多种版本杀毒软件特征库的交叉对比, 有效扩大了查杀范围、缩短了缺陷窗口期。

数据安全方面。由于云平台使用虚拟化技术, 多个用户可能共享同一物理资源, 用户数据可能通过虚拟机的安全漏洞被他人获取。C-RHIP 向医疗卫生系统内部提供数据上传和分析服务, 对其他用户只提供查询服务, C-RHIP 对不同的存储库采取不同的加密方式以保证数据安全。业务文档存储库、EHR 存储库和业务平台数据库中的数据是作为存档保存的, 采用全文加密方式, 数据库中涉及个人身份标识的字段加密存储。数据仓库中的数据经过去隐私处理, 并为分析结果中较为重要的信息划出独立的物理空间进行存储以保证其安全性。

4 原型系统实现

以河北大学网络中心为依托, 构建了覆盖河北大学附属医院、河北大学体检中心、社区卫生中心等卫生机构的 C-RHIP 平台——厚德健康档案服务系统。具有开放性、智能性及可延伸性, 可以提供健康档案查询以及健康诊疗信息咨询和检索等服务功能。

原型系统以 x86 服务器, Linux, Xen, Hadoop, 和 JO²nAS 为基础搭建。其中 Hadoop 是一个开源项目, 它包括分布式文件系统 HDFS, 分布式数据库 HBase, 和并行处理软件框架 MapReduce, 它提供了大量的接口和抽象类, 供开发人员调用。JO²nAS 是一个负责整体资源调度的中间件, 它可以在系统负载较轻的时候将多个实例迁移到一台服务器上, 关闭其他服务器以节约资源, 也可以在系统繁忙时快速启动新的服务器以分担工作量保证性能。身份验证, 数据标准化模块, 重构引擎等模块以 Java 实现, 数据分析利用中间件 Mahout, 实现了聚类、分类、协同过滤、进化编程等数据挖掘算法, 并且允许根据应用层的业务定制算法库进行扩展。入侵检测系统基于分布式入侵检测思想建立, 病毒检测系统基于多版本云平台杀毒思想建立。信息搜集和信息查询等客户端软件以及用户查询网站用 NET 平台实现。

经过运行测试, C-RHIP 平台可以平稳保证各机构日常工作需要, 客户端软件和查询网站运行平稳、反应迅速。同时, 通过接入大量虚拟机同时发

送处理请求,模拟了工作量突增的情况,服务器启动迅速、平台动态扩展性能良好;工作量下降后,管理系统可自动将任务合并到少数服务器,使其他服务器进入停机状态,有效节约资源。



图4 C-RHIP 原型系统首页

5 结束语

本文提出了云计算下的区域卫生信息平台 C-RHIP 的构建方案。此方案提供了信息共享平台的集中建设机制,利用云计算的规模效益、动态扩展性和强大的计算能力,解决卫生机构现有的、分散式系统建设模式下的低水平重复建设、资源配置调度任务繁重、资源利用率低、信息格式不统一等问题。通过原型系统的搭建和运行模拟,表明本方案有效解决了上述问题,较好地实现了预期目的。

参考文献:

[1] 中华人民共和国卫生部. 健康档案基本架构与数据标准[EB/OL]. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohbgt/s6718/200905/40699.htm>. 2009-5-19.
The Ministry of Public Health of the People's Republic of China, The Basic Framework and Data Standards[EB/OL]. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohbgt/s6718/200905/40699.htm>. 2009-5-19.

[2] 中华人民共和国卫生部. 基于健康档案的区域卫生信息平台建设技术方案[EB/OL]. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohbgt/s6718/200912/45413.htm>. 2009-12-31.
The Ministry of Public Health of the People's Republic of China, Construction Technical Solutions of RHIP Based Health Record[EB/OL].

<http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohbgt/s6718/200912/45413.htm>. 2009-12-31.

[3] FOSTER I, ZHAO Y, RAICU I, et al. Cloud Computing and Grid Computing 360-Degree Compared[A]. IEEE Grid Computing Environments Workshop[C]. Texas, USA. 2008.1-10.

[4] PAUL B, BORIS D, KEIR F, et al. Xen and the Art of Virtualization[A]. The 19th ACM Symposium on Operating System Principles[C]. New York, USA, 2003.164-177.

[5] FENN M, MURPHY M, MARTIN J, et al. An Evaluation of KVM for Use in Cloud Computing[A]. The 2nd International Conference on the Virtual Computing Initiative[C]. North Carolina, USA. 2008.

[6] 北京大学信息科学技术学院. 面向网构软件的云平台的资源管理工具技术白皮书[R]. 2010.18-24.
School of Electronics Engineering and Computer Science, Peking University. Technical White Papers of Internetware-oriented Cloud Platforms Resource Managing Tools[R]. 2010.18-24.

[7] ARM BRUST M, FOX A, GRIFITH R, et al. Above the Cloud: A Berkeley View of Cloud Computing[R]. Technical Report No. UCB/EECS-2009-28, University of California at Berkeley, USA, 2009.10-12.

[8] ABADID J. Data Management in the Cloud: Limitations and Opportunities[J]. The Bulletin of IEEE Computer Society Technical Committee on Data Engineering, 2009, 32(1):3-12.

作者简介:



刘振鹏(1966-),男,河北安国人,博士,河北大学教授,主要研究方向为云计算、对等网络、Web服务。

王斌(1986-),男,河北巨鹿人,河北大学硕士生,主要研究方向为云计算等。

尚开雨(1974-),男,河北保定人,河北大学网络中心实验师,主要研究方向为Web服务。

柴文磊(1982-),男,河北高碑店人,河北大学网络中心实验师,主要研究方向为网络环境、Web服务。

吴凤龙(1985-),男,河北唐县人,河北大学硕士生,主要研究方向为云计算。