

区域社区卫生卫生信息服务云平台构建*

李树春 李殿奎 刘晓敏 陈育德

姜志梅

(佳木斯大学信息电子技术学院 佳木斯 154007)

(佳木斯大学康复医学院 佳木斯 154007)

[摘要] 针对现有区域社区平台的居民健康档案查询效率低、用户等待时间过长及系统存储难的问题,提出构建 Hadoop 云环境实验平台,介绍基于云平台的居民健康档案平台架构及功能模块,实现数据快速查询,解决居民健康档案大量数据的存储问题。

[关键词] 居民健康档案; Hadoop; HBase; MapReduce

[中图分类号] R-056 [文献标识码] A [DOI] 10.3969/j.issn.1673-6036.2016.04.005

Construction of the Cloud Platform for Regional and Community Health Information Services LI Shu-chun, LI Dian-kui, LIU Xiao-min, CHEN Yu-de, College of Information and Electronics, Jiamusi University, Jiamusi 154007, China; JIANG Zhi-mei, Rehabilitation Medicine College, Jiamusi University, Jiamusi 154007, China

[Abstract] Regarding problems such as low query efficiency of resident health records on existing regional and community platforms, long user waiting time and difficulty in system storage, the paper proposes the construction of Hadoop cloud environment testing platform, introduces the architecture and functional modules of the cloud platform-based resident health records platform, realizes quick data query and solves the problem of storing mass data in resident health records.

[Keywords] Resident health records; Hadoop; HBase; MapReduce

1 引言

2009 年原卫生部颁布了《基于健康档案的区域卫生信息平台建设指南(试行)》、《基于健康档案

的区域卫生信息平台建设技术解决方案(试行)》等一系列技术标准和指导方案,随后颁布的《卫生部关于规范城乡居民健康档案管理的指导意见》指出,建立城乡居民健康档案的工作目标是:从 2009 年开始,逐步在全国统一建立居民健康档案,实施规范管理。可见,在全国统一建立居民健康档案是实现信息化的迫切需要和必然选择^[1]。然而,由于我国医疗健康档案发展较为滞后^[2],信息化程度还不高,基层医疗卫生机构的信息化发展也较为缓慢^[3],而且由于数据量增加,现行服务器要付出高昂代价才能满足存储需求,因此,该项工作目标的实现存在较大的障碍。但云计算技术的引入,可极大改变这一窘迫现状。云计算技术在全世界范围内

[修回日期] 2016-01-20

[作者简介] 李树春,硕士,发表论文 1 篇;通讯作者:李殿奎。

[基金项目] 2014 年佳木斯大学研究生科技创新项目(项目编号: LZR2014_013); 2015 年黑龙江省青年科学基金(项目编号: QC2015072); 2009 年黑龙江省教育厅重点项目(项目编号: 11551z015)。

已经成为 IT 技术中的一大热点^[4]，随着医院业务的持续扩展，医学影像存储将达到 TB（万亿字节）级数据容量^[5]，而云计算有高可靠性、高可扩展性、通用性、极其廉价等特点^[6]，非常适合构建区域医疗卫生信息平台。本文应用云计算相关技术对社区卫生中心的居民健康档案系统的构建进行研究，成功建立区域社区医疗卫生信息服务云平台。不仅可增大存储空间，而且可解决居民健康档案系统查询操作界面延时过长及海量数据的存储困难。

2 资料与方法

2.1 居民健康档案

居民健康档案是指居民身心健康（正常的健康状况、亚健康的疾病预防、健康保护促进、非健康的疾病治疗等）过程的规范及科学记录。它是以居民个人健康为核心、涵盖各种健康相关因素、实现信息多渠道动态收集、满足居民自身需要和健康管理的信息资源^[7]，是存储于计算机系统之中、面向个人提供服务、具有安全保密性能的终身个人健康档案^[8-9]。信息化是实现卫生改革跨越式发展的重要保障^[10]，而居民健康档案的建立、健全则成为医疗系统的有利补充^[11]。原卫生部印发的《健康档案基本架构和数据标准（试行）》提出，电子健康档案的基本内容包括两类：一类是个人基本信息，另一类是主要卫生服务记录^[12]。个人基本信息建档时，信息采集必须全面。编号和身份证号码唯一对应，是患者双向转诊的必要基础。居民电子健康档案建立和健全过程中，主要的卫生服务记录包括图像、影像、就诊记录等。其中，图像、影像的数据量较大，占用大量的存储空间。为方便数据的存储，采用云计算。

2.2 云计算 Hadoop 环境

在实验室用 7 台普通计算机搭建了小型云平台，云环境主要配置如下：操作系统为 Ubuntu 12.04 版；Hadoop 版本为 1.04 版；HBase 版本为 0.94.4；ZooKeeper 版本为 3.3.3 版。在 Ubuntu 系统下先安装 Hadoop，其次安装 HBase，最后安装 ZooKeeper

框架，搭建一个小型云环境。图 1 为 Hadoop 的系统环境^[13]，其中 HBase 位于结构化存储层，是面向列的开源分布式数据库系统。Hadoop HDFS 为 HBase 提供高可靠的底层存储支持，Hadoop MapReduce 为 HBase 提供高性能的计算能力，以 Key/Value 对的方式进行计算。ZooKeeper 则作为 HBase 集群的协同服务机制^[14]。

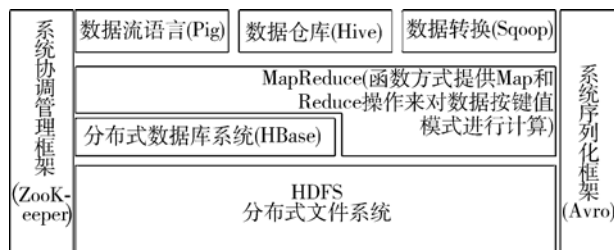


图 1 Hadoop 的系统环境

3 结果

3.1 基于云平台的居民健康档案平台架构

3.1.1 云平台架构 居民健康档案信息的导入分为自动和手动两种方式：自动导入通过 Sqoop 提供数据从传统数据库向 HBase 导入的能力，可以把原有系统中的各类信息导入到新的云平台上；手动方式是为新增居民信息通过人机交互界面导入信息而设计。所有的信息通过分类、提取形成各种不同的信息表，这些信息表都是通过系统设定好的规则、命名方式建立起来的，按列的方式存储在 HBase 数据库中，同时也方便今后的查询。将一些影像图片、B 超、CT 视频流、心电图等数据分别建立索引表，之后存储在数据库中，可加快存储速度同时也方便今后的查询。在数据库之上应用 MapReduce 对存储在数据库中的数据进行分析汇总，对社区内居民的各种常见疾病相关信息进行统计，例如发病率、病死率等。对各种常见传染病、流行病进行监测建模，及时做出预警，控制疾病的蔓延和扩散。基于居民健康档案的云平台架构，见图 2。

云平台架构的底层是数据库云存储系统，所有的数据都存在系统之中，系统主要包括 8 个子系统，8 个子系统分别对应 8 个居民健康模块，而模块和存储系统之间通过数据处理云平台连接一起构

成医疗卫生信息服务云平台。居民各类信息通过各模块录入系统中, 经过平台上的 MapReduce 编程模型以 key/value 对的方式分析、分类提取后存储在系统的相应子系统中。查询时, 通过关键字搜索, 查找存储在系统中的相关信息, 通过人机交互界面反馈给用户。

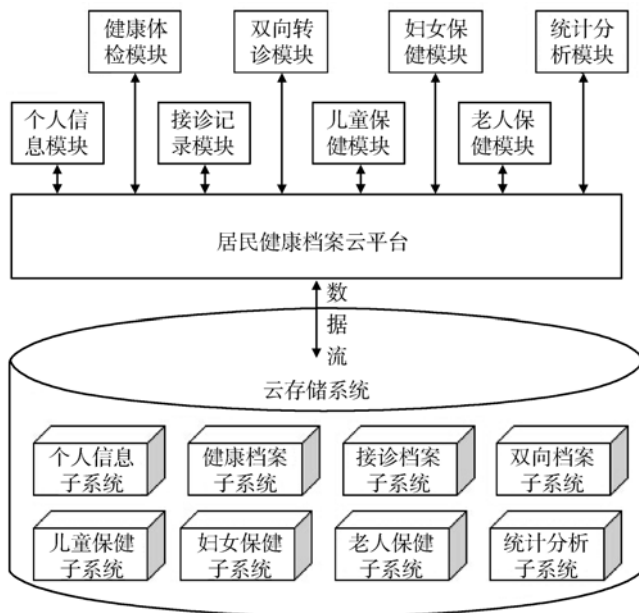


图2 居民健康档案的云平台架构

3.1.2 MapReduce 模型数据处理 当 MapReduce 函数被调用后, 会发生下列操作^[14]: (1) MapReduce 库首先将输入文件划分为 M 个数据分片, 然后复制 MapReduce 程序副本到集群所有节点上。(2) 程序副本有一个主控程序, 所在的节点是主控节点, 主控节点分配 Map 任务或 Reduce 任务到其他 Worker 节点上。(3) 分配 Map 任务的 Worker 节点读取一个被分配的数据分片并解析出 key/value 对, 然后将 key/value 对传递给自定义的 Map 函数。Map 函数经过处理会生成中间 key/value 对并缓存到内存中。(4) 每个 Map 任务生成的中间 key/value 对都通过分区函数分成 R 片, 然后周期性地写到本地磁盘上。这些存储位置被传给主控节点, 由主控节点负责把位置信息发送给 Worker 节点上的每个 Reduce 任务。(5) 当 Reduce 任务启动后, 从 Map Worker 所在主机的磁盘上读取中间数据。通过对 key 值进行排序合并具有相同 key 值的中间数据。

(6) Reduce Worker 程序将每一个唯一的中间 key 值以及对应的中间 value 值的集合传递给自定义的 Reduce 函数。Reduce 函数的运行结果输出到输出文件。任务完成后, 调用其他 MapReduce 程序合并生成的 R 个输出文件, 合并后的文件如果小于 64M, 主控节点会调用 HBase 架构进行数据存储。

3.1.3 基于 HDFS 支撑的 HBase 数据存储

HBase 中的数据以表的形式存储, 存储过程是, 首先通过远程过程调用 RPC 向和主控节点在同一个节点上的 NameNode 请求数据块存放的位置, NameNode 响应请求, 在 DataNode 中寻找存放该数据块的位置, 将 DataNode 和数据块位置信息返回给客户端, 客户端将位置信息存放在 HBase 上的 .META. 表中, 而 .META. 的 Region 信息存放在 ZooKeeper 中的 -ROOT- 表上。客户端将数据写入本地缓存, 当写满一个 64M 的块后, 客户端访问 ZooKeeper 获得 -ROOT- 表的地址, 进一步通过 -ROOT- 表访问 .META. 表获得需要通信的用户表的 HRegion 信息, 包括地址信息。之后客户端将数据块写入 DataNode 组成的 HDFS 框架上, 客户端通过 DFSOutputStream 对象把数据分成很多数据包, 再通过 FSDOutputStream 对象把数据包一份一份写进管道的第 1 个 DataNode 中, 然后管道将数据包转发给第 2 个 DataNode, 一直转发给最后一个 DataNode, 默认的副本因子是 3, 即一份数据被保存 3 份, 以防数据丢失, 文件存储成功。

3.2 居民健康档案的功能模块设计^[15]

3.2.1 个人基本信息档案 包括居民的姓名、性别、年龄、身份证号、编号、电话号、户籍地址、血型、文化程度及其他一些本人相关医疗信息等。

3.2.2 健康体检档案 包括体检人的姓名、性别、年龄、身份证号、编号等基本信息还有体检的时间, 检测项目的名称与结果, 以及体检结果与参考值、前几次体检结果的对比等信息。

3.2.3 接诊记录档案 包括居民基本信息外, 还包括接诊时间、检查项目、影像资料存储、初步诊断结果等信息。

3.2.4 双向转诊档案 包括居民基本信息外, 还

包括转院时间、接收医院信息、初步诊断结果以及转回时间、病情情况、康复意见等信息。

3.2.5 儿童保健档案 包括儿童基本信息、父母基本信息、出生日期、健康情况、疫苗接种、营养膳食等信息。

3.2.6 妇女保健档案 包括妇女基本信息及婚姻状况、健康状况、生育状况等信息。

3.2.7 老人保健档案 包括老人基本信息及情感认知状况、健康状况、体育锻炼、饮食习惯、血压、血糖、老人常见病等信息。

3.2.8 统计分析 对存储在数据库中的数据进行常见病统计、分年龄段性别统计、常见传染病统计与检测分析等。

4 结语

随着医学科学技术的发展,各种医学检测技术产生的数据量越来越大,图像、视频流、心电图等等数据占用了大量的存储空间,而脑电技术的发展使得数据量增长的速度进一步加快。目前脑电采集仅仅几分钟、十几分钟,随着技术的成熟,将来实时脑电监测技术的成型应用,这种数据量的增长将是爆炸性的,只有云存储技术才能更好地解决这种日趋增长的数据存储的问题。而区域卫生信息化已成为各医疗机构发展的必然趋势,但在偏远地区的社区卫生中心信息化程度十分低,由于各方面条件的限制,原有的区域卫生信息化非云平台查询效率低下、存储困难的现状十分严重。云计算的分布式并行计算特点可很好地解决查询效率低下问题,为医疗信息化的实现奠定了平台基础。

本研究基于 Hadoop 云计算平台实现了区域社区云平台的居民健康档案各类信息的快速查询与存储。平台的实现为病人双向转诊提供了条件,形成“首诊在社区、小病进社区、大病到医院、康复回社区”的就医格局,可避免耽误病情和资源浪费。也为区域卫生信息化提供了解决思路,实现真正意义上的资源共享。平台虽然仅在实验环境中搭建起来,但提出了一种性价比很高的平台横式,是新的

尝试,解决区域社区卫生信息化,云计算技术的应用将成为一种趋势,云技术不仅能解决运算分析问题,能解决大数据量的存储问题。

参考文献

- 1 刘振鹏,王坤瑞,卞昭玲,等. 基于云计算的区域电子健康档案服务系统研究 [J]. 档案学通讯, 2012, (4): 17-20.
- 2 马晓君. 大学生电子健康档案管理系统功能设计及保障机制研究 [D]. 长春: 吉林大学, 2015: 73.
- 3 曲晓慧. 基于云服务的区域医疗信息共享协同平台的研究 [D]. 青岛: 中国海洋大学, 2014: 59.
- 4 张迪, 霍妍. 云计算技术在医院信息化中的应用 [J]. 信息技术, 2011, (5): 171-173.
- 5 史森中, 曾凡, 黄昊, 等. 云计算在医疗信息化建设中的应用分析 [J]. 中国卫生质量管理, 2014, (6): 6-8.
- 6 周品. 云计算实战 [M]. 北京: 清华大学出版社, 2012: 4-5.
- 7 董建成, 杨剑, 蒋天民, 等. 基于云计算的电子健康档案系统建设理念与实践 [J]. 中国卫生信息管理杂志, 2010, (6): 43-46.
- 8 Tang PC, Ash JS, Bates DW, et al. Personal Health Records: definitions, benefits, and strategies for overcoming barriers to adoption [J]. Journal of the American Medical Informatics Association, 2006, 13 (2): 121-126.
- 9 Hayrinen K, Saranto K, Nykanen P. Definition, Structure, Content, Use and Impacts of Electronic Health Records: a review of the research literature [J]. International Journal of Medical Informatics, 2008, 77 (5): 291-304.
- 10 胡新平, 张志美, 董建成. 基于云计算理念与技术的医疗信息化 [J]. 医学信息学杂志, 2010, (3): 6-9.
- 11 柴文磊. 云计算模式下电子健康档案服务系统的开发 [D]. 保定: 河北大学, 2013: 48.
- 12 和征. 基于健康档案的区域卫生信息共享平台的研究 [J]. 计算机技术与发展, 2011, (10): 190-192, 96.
- 13 怀特著, 周敏奇, 等译. Hadoop 权威指南 [M]. 第 2 版. 北京: 清华大学出版社, 2011: 126.
- 14 鲍亮, 陈荣. 深入浅出云计算 [M]. 北京: 清华大学出版社, 2012: 82-83, 102.
- 15 严格. 云计算在区域医疗信息化中的应用研究 [D]. 北京: 北京交通大学, 2011: 45.