

基于云的区域电子健康档案存储和交换架构*

杨 剑 胡新平 董建成

(南通大学医学信息学系 南通 226001)

〔摘要〕 通过对区域电子健康档案基本内容、数据形式分类、存储分布、存储和交换需求的详细分析,利用云架构的特点,对区域电子健康档案的存储和交换架构进行重构,设计出区域电子健康档案的存储云分布、区域电子健康档案云存储和交换的结构模型,为区域电子健康档案共享提供指导性方案。

〔关键词〕 云架构;云计算;云存储;区域电子健康档案

Regional Electronic Health Records Storage and Exchange Architecture Based on Cloud Computing YANG Jian, HU Xin-ping, DONG Jian-cheng, Department of Medical Informatics, Nantong University, Nantong 226001, China

〔Abstract〕 The paper analyzes the basic contents, classification of data types, storage distribution and requirement of storage and exchange, then reconstructs the storage and exchange architecture of regional electronic health records (EHR) by utilizing the characteristics of cloud architecture. In the last it designs the storage cloud distributing and the structure model of cloud storage and exchange of regional EHR, provides a guiding program for sharing of regional EHR.

〔Keywords〕 Cloud architecture, Cloud computing, Cloud storage, Regional EHR

1 引言

2009 年 5 月 19 日,我国卫生部发布了我国第一部电子健康档案标准即健康档案基本架构与数据标准,这标志着我国在卫生信息化标准建设方面取得重大突破,具有非常现实而深远的意义^[1]。目前,

全球范围许多国家都在加紧进行区域电子健康档案信息平台的研究和部署,部分发达国家如美国、加拿大等,在区域电子健康档案信息平台的建设上已经走在了前面,无论标准的成熟度,还是架构的清晰度都达到了较高的水平^[2]。

近两年来,云概念和技术得到了越来越广泛的重视,在 IT 行业,国内国际各大巨头已经投入巨资进行相关的产业布局。如 Google 的 GFS (Google File System)、BigTable 和 MapReduce 架构、亚马逊的 S3、SimpleDB、EC2 以及开源的 Hadoop 框架^[3]。Google 公司提供的部分服务已经移植到了 Google 云中进行了整合,在云的应用上已经走在了前列,同时越来越多的问题也随着新概念和技术的运用而出现。以下从存储、交换的角度,深入剖析电子健康档案的数据性质、存储分布、交换需求等内外部特

〔收稿日期〕 2009-12-25

〔作者简介〕 杨剑,硕士,讲师,主要研究方向为医学信息学、云计算;胡新平,博士,副教授,主要研究方向为医学信息学;董建成,硕士,教授,主要研究方向为医学信息学。

〔基金项目〕 江苏南通大学自然科学基金(项目编号:09Z054)。

征,通过云技术和概念,对区域电子健康档案进行存储和交换的云重构,实现区域电子健康档案的可持续性共享。

2 区域电子健康档案

2.1 健康档案的基本内容

在我国刚颁布的健康档案基本架构与数据标准(以下简称标准)中,规定了健康档案的基本内容包括个人基本信息(人口学信息、社会经济学信息、亲属信息、社会保障信息、基本健康信息、建档信息)、主要卫生服务记录(儿童保健、妇女保健、疾病预防、疾病管理、医疗服务)。文章所讲的区域电子健康档案源自健康档案,但是侧重于信息的存储、交换和共享,在基本内容中,除了包含健康档案的基本内容,还必然存在索引数据、路由数据、架构数据、交换数据来支撑区域性信息存储、交换和共享。

2.2 区域电子健康档案数据形式分类

通过对区域电子健康档案基本内容的分析,在大粒度的层面上,索引数据、路由数据、架构数据等基本都适合于以严格结构化的数据形式存在;交换数据适合于以半结构化的数据形式存在,它基于各种协议、标准的规定以数据流的形式在各种节点中进行传递交换。个人基本信息、卫生服务记录的主要信息来源载体是卫生服务记录表单,这些表单从本质上接近于半结构化形式,所以这类数据除了以严格结构化的数据形式存在,在特定情况下,也需要以半结构化的数据形式存在;对于产生于各种服务环节的图片、文档比如 B 超图,则以非结构化的数据形式存在。

2.3 区域电子健康档案存储分布

目前我国关于区域电子健康档案存储分布的探索和实践,都属于星型网状结构的简单分布。这种分布以病人主索引 PMI 为中心,数据分布在各个医院、社区等大小医疗机构和医疗点。其中最主要的问题就在于不同医疗机构和医疗点,作为电子健康

档案的实际维护机构,它的 IT 技术能力完全不对等,这对于整个区域电子健康档案的数据安全、健壮、稳定性,都会产生很大的不可预知的隐患。以下称此类区域电子健康档案为“不可持续的区域电子健康档案”。

2.4 区域电子健康档案数据存储和交换需求

区域电子健康档案是一个复杂的庞大的体系,对于内部不同数据的特性以及对数据的不同需求,使得区域电子健康档案数据交换也需要存在不同的存储交换需求:实时存储交换、队列存储交换、独立存储交换。

3 基于云的存储和交换架构

3.1 云和基于云的存储和交换架构的概念

云的概念目前并没有进行专门的定义,简单地说,云是一种通过网络提供资源或服务的架构^[4]。存储云提供存储服务,计算云提供计算服务。结合不同的云,组合在一起形成一系列的云服务,提供云平台用于搭建基于云的应用。

云概念具有以下几大特点,只有满足以下 3 大特点才能叫做真正意义上的云:按需扩展,云可以根据业务需求,随时扩展自己的存储和计算容量,满足消费者(Consumer)对云中资源或服务的膨胀式需求;无限易扩展,云对消费者而言是陌生的黑箱,它可以基于云构建者的不同架构方案,但是架构必须很容易扩展,并且无限制扩展;低成本扩展,这是区分云架构与传统分布式架构的一个很重要的特点,云架构必须能够进行低成本扩展,利用其能力来整合各种硬件、软件资源,包括整合现有的资源以及本身不具备高级功能的廉价资源两个方面,需要特别指出的是,低成本扩展并不是低成本架构,并不是指云中所有构件都是廉价设备,云架构主要强调的是扩展的低成本性;扩展与成本成反比,随着云的不断扩大,单位成本逐渐降低,这是云理念得以发展和存在的基石。基于云的存储和交换架构指的是提供存储和交换服务的云,在云中各种构件的组合作方式。

3.2 云存储和交换的基础环境

云存储和交换架构的存在必须基于一定的基础环境,包括网络基础设施、通信协议和交换标准。网络基础设施:充分利用现有网络环境,针对具体的场景,建立存储和交换架构。通信协议:根据应用环境和需求,存在着丰富的通信协议集,如 TCP/IP、HTTP、SOAP、P2P 等。一些特殊的针对云存储和交换架构的协议也已经进行了设计,如用于高速广域环境传输的基于 UDP 的 UDT 协议^[5]、基于 P2P 的网络路由协议 Chord^[6]。交换标准:互联互通性要求交换标准的完善,针对于云的具体应用,有不同的交换标准需求。基于云的区域电子健康档案存储和交换架构,也需要完善的标准。

4 重构区域电子健康档案存储和交换架构

4.1 重构目标

对区域电子健康档案的存储和交换进行重构,最终目标是要实现区域电子健康档案的可持续性共享,这包含以下几个方面的含义:可维护:不会因为数据源机构、数据存储机构等不具备足够的 IT 技术能力造成数据丢失、数据异常、数据崩溃等不可维护性异常。按需扩展:区域共享可根据实际需求动态扩展终端;区域共享可根据需求动态接入上级区域,实现更大范围的共享;终端可根据实际需求动态获取更多存储。低成本:通过云存储和云交换架构,最大限度地减少机构在本地硬件、软件、人员方面的建设和维护成本;根据区域电子健康档案各种数据组成的特点,重构到不同类型的存储云中。

4.2 区域电子健康档案存储云(EHR-Cloud)分布

通过对区域电子健康档案的数据组成、存储分布、数据形式分类、存储和交换需求进行分析,结合云架构的特点,把区域电子健康档案存储云构建为两个存储云层,3 种存储云:公共存储云、私有存储云、第 3 方存储云,见图 1。基于云的区域电子健康档案存储和交换架构,通过合理利用各种云资源,可以实现可持续性的区域电子健康档案共享。

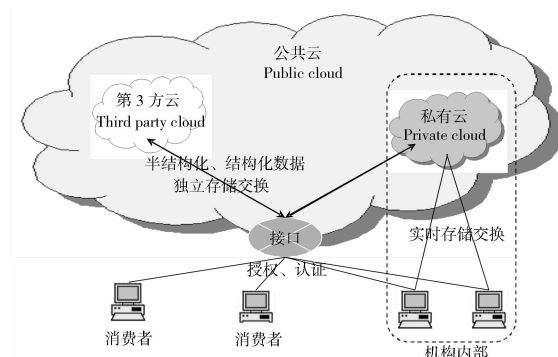


图 1 区域电子健康档案存储云分布

私有存储云存在于医疗机构内部,特别是大型医院。私有存储云一般情况下只允许机构内部的消费者访问和利用。私有存储云提供机制整合内部分布在各个存储点的电子健康档案数据,满足区域电子健康档案交换需求的实时存储交换。私有存储云在特定情况下,与公共存储云、其它存储云之间进行云数据交换。

第 3 方存储云指的是大型 IT 企业提供的商业存储云。两类情况下有必要把私有存储云中的数据转移到第 3 方存储云中,以获得更低的成本和更高的效用。一类是医疗机构自身不具备 IT 技术部门或者 IT 技术能力不强,此时,运营一只技术能力强的 IT 人员队伍来进行私有云的维护以及私有云与其它云的云数据交换,变得低效和浪费,有必要和需求把私有云的数据转移到第 3 方存储云。另一类是私有云中的大量半结构化、非结构化数据,特别是具有独立存储交换需求的数据,可以转移到第 3 方非结构化存储云中,获得更低的成本,实现更好的数据备份、灾难恢复等功能。通过第 3 方存储云的架构,很好地解决了目前区域电子健康档案共享的不可持续性问题。

外部云层是公共存储云,直接与消费者(Consumer)进行数据交互,对消费者而言,公共存储云内部的架构是看不见的,只需通过适当的授权认证,通过公共存储云提供的接口就可以获取利用公共存储云中的电子健康档案数据。区域电子健康档案的公共存储云提供机制整合内部的私有存储云和第 3 方存储云。公共存储云同时具有私有云的特

征,与更高一级云层进行整合。

4.3 区域电子健康档案云存储和交换的结构模型

区域电子健康档案的存储云,从不同层级的粒度来看,有几种类型,但是从存储云本身来说,它的结构基本是一致的。EHR - Cloud 是一个复杂的体系,按照层次划分,可以划分为存储层、管理层和接口层。

EHR - Cloud 的存储层是云存储和交换的最基础部分。存储设备包括了各个医疗机构、卫生行政区域已有的各种存储设备。存储层甚至可以基于 PC 机进行低成本扩展。存储层的存储设备往往分布在不同的物理地点、医疗机构,彼此通过广域网、局域网或光纤通道连接在一起,通过特殊的交换协议进行数据交换。存储设备之上是一个统一存储设备管理系统,可以实现存储设备的逻辑虚拟化管理、多链路冗余管理,以及硬件设备的状态监控和故障维护。

EHR - Cloud 的管理层是云存储和交换的最核心部分,也是云存储和交换架构中最难以实现的部分。管理层通过集群、分布式文件系统和网格计算、云关系数据库等技术,实现云存储中多个存储设备之间的协同工作,使多个存储设备可以对外提供同一种服务,并提供更大更强更好的数据访问性能。CDN 内容分发系统、数据加密技术保证云存储中的数据不会被未授权的用户所访问,同时,通过各种数据备份和容灾技术和措施可以保证云存储中的数据不会丢失,保证云存储自身的安全和稳定。

EHR - Cloud 的应用接口层是云存储和交换最灵活多变的部分。不同的存储云比如私有云、第 3 方云、公共云,根据消费者需求情况,以及云交换的需求情况,基于 HL7、IHE、DICOM 等交换标准,提供相应的应用服务接口。

5 结语

基于云的区域电子健康档案存储和交换架构充分利用了云架构的特点,实现了区域电子健康档案的可持续性共享。通过对区域电子健康档案的基本内容、数据形式分类、存储分布、存储和交换需求的详细分析,利用云架构的特点,对区域电子健康档案的存储和交换架构进行了重构,确定了重构目标,设计出了区域电子健康档案的存储云分布、区域电子健康档案云存储和交换的结构模型。EHR - Cloud 也面临着更大的挑战,云架构的复杂性,为区域电子健康档案的安全、隐私方面提出了新的问题,后续将对基于云的区域电子健康档案做详细的安全、隐私方面的研究。

参考文献

- 1 杨剑,胡新平,董建成. 中国卫生信息系列基础标准解读 [J]. 中国数字医学, 2008, 3 (11): 55 - 56.
- 2 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.
- 3 Grossman R L, Gu Y H, Sabala M, Zhang W Z. Compute and Storage Clouds Using Wide Area High Performance Networks [J]. Future Generation Computer Systems - the International Journal of Grid Computing Theory Methods and Applications, 2009, 25 (2): 179 - 183.
- 4 Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg, Ivona Brandic. Cloud Computing and Emerging IT Platforms: vision, hype, and reality for delivering computing as the 5th utility [J]. Future Generation Computer Systems, 2009, 25 (6): 599 - 616.
- 5 Yunhong Gu, Robert L. Grossman. UDT: UDP - based data transfer for high - speed wide area networks [J]. Computer Networks, 2007, 51 (7): 1777 - 1799.
- 6 Stoica I, Morris R, Karger D, Kaashoek M F, Balakrishnan H. Chord: a scalable peer to peer lookup service for Internet applications [C]. Proceedings of the ACM SIGCOMM, New York: Association for Computing Machinery, 2001.