

临床数据发布中心集成与开发

李 恺 吕寻伟

(北京怀柔医院 北京 101400)

〔摘要〕 以北京怀柔医院为例, 介绍临床数据发布中心集成与开发, 具体阐述平台的关键技术、体系结构、功能设计、应用效果, 指出该平台有助于改善患者就医体验, 提高医疗效率与质量。

〔关键词〕 大数据平台; 数据发布; 数据集成与开发; Hadoop; ETL

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.06.005

The Clustering and Development of the Release Center of Clinical Data LI Kai, LV Xunwei, Beijing Huairou Hospital, Beijing 101400, China

〔Abstract〕 With the example of Beijing Huairou Hospital, the paper describes the clustering and development of the release center of clinical data, and expounds the key technologies, system architecture, function designs and application effects of the platform, points out that this platform is conducive to improving medical services as well as the efficiency and quality of medical treatment.

〔Keywords〕 Big data Platform; data distribution; data clustering and development; Hadoop; ETL

1 引言

医院信息系统的开发、应用, 经过多年的历程之后正向广度和深度发展, 达到前所未有的新高度、新水平, 主要表现在建立大规模的“以病人为中心”, 具有网络分布、数据集中、一体化设计和数据共享等 4 个基本特征的一体化医院信息系统, 各应用软件服务商独立开发众多系统, 但缺乏建设主线, 各系统各自为政, 导致医院的整体信息系统缺乏融合, 无法充分发挥作用。

随着现代医疗信息化的普及与发展, 越来越多的现代化临床设备被引入到诊疗过程中, 各类应用系统日趋丰富繁杂, 系统间的数据关系、数据流向、接口实现等日趋复杂且缺乏有效的监控和统一

的调度, 给实际应用带来一系列的问题。对于医院管理者和信息管理者, 院内信息系统相对独立、数据分散, “信息孤岛”问题严重, 在推行分级诊疗政策、探索互联网医疗的过程中, 医疗数据的互联互通成为最大障碍。临床数据发布中心项目的主旨在于促进医疗资料的收集与共享, 为患者提供客观、连续的个人病历资料, 为医护人员全面准确地提供患者医疗信息, 实现医联体内患者病历信息的互联互通, 为电子处方推送、异地就诊、远程会诊、居民健康档案奠定基础^[1-4]。

2 关键技术

2.1 Hadoop

Hadoop 是能够对大量数据以一种可靠、高效、可伸缩的方式进行分布式处理的软件框架, 大幅度提升数据处理程序开发及运维的效率。利用成本比较低的中低端 PC 级服务器集群轻松存储和分析 PB

〔收稿日期〕 2019-01-31

〔作者简介〕 李恺, 副主任, 高级工程师, 发表论文 5 篇。

级别的海量数据。Hadoop 生态系统中集成大量的工具和组件来满足不同计算和存储需求,如 HDFS 分布式文件系统、HBase 列式数据库、Hive 数据仓库、Mahout 机器学习等,可以方便地进行数据存储和分析计算。采用 Hadoop 中分布式文件系统 HDFS 和分布式列式数据库 HBase 进行数据存储,每份文件存储至少 3 份,有效保证数据的可靠性。基于 Spark 和分布式消息队列 (KAFKA),实现具备实时计算能力的并行计算框架及任务的并行调度,提升临床数据中心的运行效率。采用多个机器节点组成的服务器集群,来构建分布式的存储系统,通过万兆网卡实现节点间的高速数据交换。Hadoop 共有 3 种角色节点,分别为 NameNode、CMNode 和 DataNode,其中 NameNode 负责任务调度和元数据存储,CMNode 负责基群监控和管理,DataNode 负责数据存储和计算。为保证系统的高可用性,NameNode 采用 2 台机器进行主从备份,CMNode 使用 1 台机器,即使宕机不影响集群使用,其他节点作为 DataNode,提供数据冗余存储和分布式计算能力。在 Hadoop 集群上端,分别是数据中心服务平台和应用服务器,通过主从备份方式部署,为外部客户端提供服务^[5-6]。

2.2 ETL

临床数据中心的数据采集分为两个环节:采集历史数据和增量数据。历史数据采用 Hadoop ETL 组件 Sqoop,将关系型数据库的历史数据一次性导入到 Hadoop 及 MySQL 中,由于 Sqoop 是通过 MapReduce 作业来实现的,而 MapReduce 是一种批处理方式,难以实现实时数据的导入和导出,加上增量数据业务繁琐,如修改后的数据、多服务器多数据源联合数据采集的方式,普通的 ETL 难以满足需求,所以增量数据需要借助二次开发的 ETL 来完成。抽取、转换、装载 (Extract - Transform - Load, ETL) 作为数据处理引擎,会对从数据源流入的数据进行清洗转换,再将处理后的数据传送到临床数据中心的数据仓库。采用与数据库底层存储方式无关的 ETL,能够实现复杂的数据清洗功能并降低数据库系统的压力^[7-9]。

3 体系结构

临床数据发布中心使用 Hadoop、HBase、ETL 等大数据技术,全面收集医院信息系统 (Hospital Information System, HIS)、影像存储与传输系统 (Picture Archiving and Communication System, PACS)、检验信息系统 (Laboratory Information System, LIS)、电子病历系统 (Electronic Medical Records, EMR) 以及各类小型医疗信息系统的数据信息;通过大数据处理引擎,对数据进行有序化处理,建立患者统一索引,形成患者就诊记录和报告的有序关联,构成临床数据发布中心,为数据的挖掘利用提供基础,见图 1。

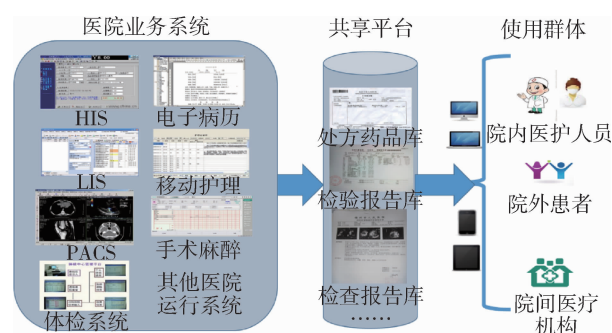


图 1 临床数据中心体系架构

4 功能设计

4.1 概述

患者、医务人员、医院管理者以上问题实际上是数据管理和数据利用问题。目前医院的数据量级已经达到大数据的规模,应该由信息化向数据化转变;由分散信息向聚合数据的转变;由单一的面向业务的数据展示,向数据整合利用与价值输出转变。临床数据发布中心以医院现有的信息系统为基础,基于大数据技术,通过互联网、移动互联网、APP 等主流技术和终端,充分释放数据的价值,唤醒沉睡在各个临床系统中的数据。

4.2 数据类别

临床数据发布中心的数据来自各业务系统,通

过 ETL 工具、虚拟打印技术、高拍仪、影像采集工具等技术手段收集各类系统的数据，见表 1。

表 1 临床数据发布中心数据来源及类别

1 级类别	2 级类别
检查类	CT、DR、MR、XA、CR、RF、DX、MG、NDP、内镜、呼吸内镜、肺功能、核医学、超声、病理、消化内镜
检验类别	常规检查、微生物检查
处方类别	门诊处方
影像格式文件	DCM、IMG、头文件
病历类别	门诊病历、住院病历 - 出院记录、住院病历 - 出院诊断证明

4.3 数据采集（表 2）

表 2 临床数据发布中心数据采集方式及时间

系统简称	对接内容	对接方式	任务执行日期
HIS	门诊处方	DB 视图	凌晨，每天 1 次
LIS	常规检验项目、微生物项目	DB 视图 FTP 共享目录	凌晨，每天 1 次
PACS	CT，DR，MR，XA，CR，RF，DX，MG，NDP，内镜，呼吸内镜，肺功能，核医学，超声，病理，消化内镜 影像文件	DB 视图 FTP 共享目录 DICOM Server	凌晨，每天 1 次 凌晨，每天 1 次
EMR	门诊病历、出院记录、出院诊断证明书	DB 视图	凌晨，每天 1 次
体检	体检报告	DB 视图	凌晨，每天 1 次

4.4 数据生命周期

4.4.1 源数据 随着卫生机构信息化的发展，大大小小的系统从规范到紊乱，沉淀下来的数据不下几百 TB。此外有的系统自身具有灾备的能力，有的则仅此一份、没有备份，从数据安全的角度考虑，数据一旦丢失或毁坏后果严重。因此，当数据再次整合时需要从存贮成本、数据价值两个方面对入库数据进行评估。存贮方面，已具备一个或多个副本

的存储，数据再次聚合时需要减少多个副本的重复出现。数据价值方面，将具有利用价值的数据收录到临床数据发布中心，如针对处方数据、检查数据、检验数据、病历数据、医嘱数据等有价值历数据进行采集和处理。

4.4.2 网关服务 源数据的数据类型多样，可大致分为结构化数据、非结构化数据两种。常用的数据获取方式主要有 JDBC、FTP、共享目录、虚拟打印、扫描等。数据处理过程分为以下几个步骤：（1）索引数据。Hbase 最核心的工作不是表设计，而是行主键的设计，在 Hbase 中定位某条数据或某区域的数据都是通过行主键来定位的，由于 Hadoop 集群的优势是数据分散计算、分散存储，互不影响，而数据为什么能平均分散在各个节点中，取决于行主键的设计，行主键的来源与业务需求息息相关，每条进入到 Hbase 的数据都需要一个行主键，可理解为索引，而该索引会存储到 MYSQL 中，便于快速找到自己需要的数据。（2）结果数据。检验报告结果、检查报告结果等结构化数据会以增量和全量的形式进入写到 Hbase。（3）非结构化数据。检验版式文件、检查版式文件、病历版式文件等会以昂和全量的形式进入写到 Hbase。（4）科研数据。将科研用到的数据通过 Kafka 消息推送给科研平台，科研平台再进行针对行分析、处理。（5）缓存数据。为优化用户使用体验及响应速度，部分数据需要进行临时缓存，如影像数据和分析数据。

4.4.3 数据库 数据会被分别写入到 MySql 和 HBase 数据库中，两者相互补充弥补各自的不足：MYSQL 用来支持联机事务处理过程负载（OLTP Workload），就是大量的单行查询。MYSQL 本身不是分布式设计，即便使用传统的分表做法也无法与分布式的大数据运算相比。集成 HBase 的 Hadoop 是用来处理 OLAP workload 的，就是大量的单行查询。数据分析是 Hadoop 的优势。

4.4.4 应用服务 当数据被合理分配在各自的适当位置，数据利用就简单便捷，如患者的移动 APP 应用、临床和区域的患者统一视图系统等在数据质

量和数据体验上与以往基于关系型数据库模式都有明显的提升。

4.5 管理控制平台

临床数据发布中心通过管理控制平台进行集中的配置管理,包括控制台、信息发布管理、数据质量控制、报表管理等数据服务,对临床数据发布中心多种数据及应用的管理配置与控制。

4.6 患者统一视图

患者统一视图为医院管理者和临床医生提供患者信息全面展示、资料查阅功能,如可以通过统一视图所提供的全景资料快速获取患者的病历、检查检验资料等诊疗数据。统一视图提供就诊视图和分类视图两种高效的病历浏览方式。就诊视图按患者就诊时间方式进行病历浏览;分类视图可以结合报告类型和时间,按照报告类型针对性浏览。有 PC 版本和 Pad 版本,可嵌入 EMR、HIS 等集成使用。

4.7 移动 APP 患者端

4.7.1 基本服务 临床数据发布中心的患者端移动 APP 支持 iOS、Android 主流手机平台;基于发布中心向患者开放的数据类别,围绕病历资料,为患者提供基于病历信息的查阅浏览功能。(1) 患者身份验证。患者注册并登录 APP 后,填入身份信息和医院的就诊卡号实现就诊卡的认证和绑定,以便获取病历信息。(2) 最近就诊信息快捷查看入口。在首页直观列出患者最近就诊的科室和诊断信息,显示最近用药的处方信息;点击可以查看对应的详细信息。(3) 家庭健康档案管理。患者可以建立多个健康档案,绑定家人的就诊卡号,实现对家人病历信息的获取和查阅,对家人的健康情况进行关注和查阅。(4) 就诊记录。患者可以查阅自己在医院的就诊记录,根据其就诊时间从新到旧的显示就诊信息记录;点击可以查看该次就诊的详细信息。(5) 就诊详细信息查阅。就诊详细信息可以显示某次就诊的类别、日期、科室和就诊医生;显示该就诊的病历、处方、检验、检查等报告。点击可以查看各类报告的详细内容。(6) 报告查阅。检验检查报告可以显示该份报告的电子版内容;影像类检查,可

以查看影像图像。(7) 分类查阅。检验检查报告提供分类视图,可以分类展示患者就诊记录中各类别的报告,可以按照病历、处方、检验、检查等类别进行针对性的浏览。(8) 病历分享。对某个报告进行分享,支持分享到 QQ 好友、微信好友、邮件等。(9) 就诊医生。可以查看患者就诊过的医生列表,点击某位医生可以查看这位医生为其诊断的详细记录。

4.7.2 扩展应用服务 临床数据发布中心面向多应用场景的应用扩展机制,可以在患者端扩展以下应用服务。(1) 智能可穿戴设备接入。接入指定品牌的智能电子血压仪产品,实现血压测量后自动上传,将测量记录显示在移动 APP 的患者端。(2) 院内就诊导引。在移动 APP 的患者端加入院内就诊导引模块,实现在 APP 内查看院内的路线图、楼层分布等。(3) 患者满意度调查。通过移动 APP 进行患者满意度调查。患者可以通过移动 APP 患者端对最近就诊进行满意度评分。(4) 健康宣教。在移动 APP 患者端展示医院的新闻动态和健康宣教内容。(5) 移动病案复印。患者通过移动 APP 进行病案复印的申请、缴费,医院采用快递的方式将复印后的病案送达患者。(6) 医联体课题管理。如已开展医联体慢病课题,通过移动 APP 进行展示,借助 APP 的慢病管理入口告知患者疾病相关信息和健康宣教内容。(7) 网络预约与结算。患者可以通过移动 APP 进行预约挂号和缴费。其他需要缴费的业务使用移动 APP 进行支付结算。

4.8 移动 APP 医生端

临床数据发布中心的医生端移动 APP 支持 iOS、Android 主流手机平台;发布中心向医务人员开放的数据类别,具有为医务人员提供患者病历信息的查阅浏览和检索功能。一是就诊信息同步,提供与医院信息一致的就诊患者列表,可以查看患者的就诊类型、时间和患者姓名、年龄等基本信息及其详情。二是就诊记录,提供患者详细就诊记录信息的查阅,根据患者就诊时间从新到旧的显示就诊信息记录;点击可以查看某次就诊的详细信息。三是筛选功能,在就诊患者中按就诊类型、时间进行筛选。四是搜索功能,在就诊患者中按患者 ID 或诊断信息进行搜索。五是分类视图,分类展示患者就诊记录中各类别的报

告,方便医生按照病历、处方、检验、检查等类别进行针对性的浏览。六是就诊详情,提供某次就诊的类别、日期、科室和就诊医生等就诊详细信息的展示;显示该次就诊的病历、处方、检验、检查等报告。点击可以查看各类报告的详细内容。七是查看报告,医务人员可以查看电子版检验、检查、病理等报告;影像类检查还可以查看影像图像。八是分组功能,提供患者分组功能,根据患者不同的属性、维度进行分组管理。在患者列表中通过关注患者将其分配到某个分组中。九是临床随访扩展应用,如果开展有临床随访课题,通过医生端二维码扫码方式邀请患者入组,入组后进行患者随访跟踪,按照随访计划分发随访问卷给患者填写。十是常用工具和知识库扩展,提供药品库、法律法规库、医疗护理技术操作常规、治愈好转标准库知识库以及常用医学计算公式。

5 应用效果

经过 1 年的调研和方案研究,北京怀柔医院 2018 年年初开始进行基于大数据平台的临床数据发布中心项目调研、规划、研发与建设,于 2018 年 10 月投入试运行。此 APP 不仅增强了患者病历资料的获得感,还为每个患者建立一个可以随时随地访问的诊疗清单,在其中存储 X 光、核磁共振、CT 等检查的历史资料。患者不论是在基层医疗单位还是在大型医院就医,医生都能够及时了解患者的过往病史,减少大量重复性的医疗检查,不仅提高诊疗效率,还能节省患者的开支,也使患者在不同的医疗机构之间获得连续的诊疗服务,从而促进城市社区和农村卫生服务体系建设与协调发展。医生还可通过临床数据中心,在院前急救过程中及时获取患者的诊疗史、过敏史等信息,避免因患者病史不清楚而耽误抢救时间。后续还有许多便民惠民的功能正在规划和开发。如近期已开发、还未上线的病案复印业务,患者可以通过 APP 提交病案复印申请并在线完成付费,医生审核、复印后,直接通过快递的方式邮寄给患者,为患者免去来回奔波及排队等候的麻烦,非常便捷。APP 最终将实现以患者为中心,覆盖诊前、诊中、诊后全流程的服务体系,让患者足不出户,就能享受到优质的医疗服务。

6 结语

临床数据中心通过基于电子病历的医院信息平台数据集成,支持医院信息系统中纵向和横向的数据交换及信息共享。医院内部患者连续、准确的电子病历信息是高质、高效诊疗过程的基础,而在医院外需要及时向区域人口健康信息平台提供诊疗记录和检查检验数据。基于临床数据中心完成医疗信息数据整合,可实现对各院区分散在不同地域、不同系统的临床数据和个人健康数据的汇聚和梳理,搭建起大数据中心,同时为医院的检查检验结果共享、运营数据分析及再利用提供支持,从而提升医院诊疗质量,协助医院进行临床科研决策。建设开放式临床数据中心,将医院内部或者医院集团间甚至是一个地区的各级医疗机构的非结构化数据以标准的方式接入并且以灵活的方式存储,以患者为中心、以患者的诊疗时间为轴线统一展现,实现单个体医院内部、医疗集团之间、甚至是地区级、区域级医疗结构之间数据共享,可减少重复检查,改善患者就医体验,也有利于提高医疗质量与效率。

参考文献

- 1 廖淑华. 健康档案管理系统的设计与实现 [D]. 成都: 电子科技大学, 2013.
- 2 夏槌. 面向社区用户的健康档案信息服务系统设计与分析 [D]. 南京: 南京邮电大学, 2015.
- 3 李冉. 居民健康档案管理系统的设计与实现 [D]. 济南: 山东大学, 2013.
- 4 顾晓晖. 社区居民电子健康档案系统的设计与实现 [J]. 医学信息学杂志, 2010, 31 (10): 8-12
- 5 李军华. 云计算及若干数据挖掘算法的 MapReduce 化研究 [D]. 成都: 电子科技大学, 2010.
- 6 王超鹏, 梁正科, 李强. 基于云计算的分布式数据挖掘算法研究 [J]. 硅谷, 2012, (4): 14, 92.
- 7 计虹, 李维, 贾末. 基于大数据的医院集成平台协同与数据融合利用 [J]. 中国卫生信息管理杂志, 2017 (4): 525-529.
- 8 朱立峰, 刘淑君, 陈德华, 等. 多中心临床大数据平台建设及深度应用 [J]. 大数据, 2018 (3): 46-53.
- 9 屈晓晖, 孟群. 基于电子病历的医院信息协同服务研究 [J]. 中国卫生信息管理杂志, 2011 (5): 16-20.