

基于大数据分析的医疗质量管控系统设计与应用

李鹏泉 黄碧香 陈粤宁

(中国人民解放军联勤保障部队第九二三医院信息科 南宁 530021)

〔摘要〕 分析医疗数据开发、利用需求, 阐述基于 Hadoop 架构的医疗质量管控系统总体架构设计、Hadoop 集群配置、关键医疗数据分析结果展示以及系统创新性、特点, 指出该系统应用有助于医院加强医疗质量管理, 辅助决策制定。

〔关键词〕 Hadoop; 医疗数据; 医院信息系统

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.06.016

Design and Application of the Medical Quality Control System Based on Big Data Analysis LI Pengquan, HUANG Bixiang, CHEN Yuening, Department of Information, The 923th Hospital of the Joint Logistics Support Force of the Chinese People's Liberation Army, Nanning 530021, China

〔Abstract〕 The paper analyzes the demand of medical data development and utilization, expounds the overall architecture design of the medical quality control system based on Hadoop architecture, Hadoop cluster configuration, analysis results display of key medical data, innovation and characteristics of the system, and points out that the application of the system is helpful for the hospital to strengthen medical quality management and assist decision-making.

〔Keywords〕 Hadoop; medical data; Hospital Information System (HIS)

1 引言

近年来随着医疗信息化快速发展, 医疗数据日益复杂多样, 医疗服务质量要求不断提升。医院信息系统之间缺乏互联互通、数据共享, 导致医院医疗质量管控效果不佳。海量医疗数据亟需大数据技术支撑, 挖掘潜在价值, 提升医疗管理质量。随着大数据技术引入医学领域, 通过模型分析和数据分析发现医疗决策中的规律^[1], 为医院决策者和医务

人员提供重要数据支持^[2]。吕道明^[3]通过 Hadoop 平台将医院临床数据整合实现各数据互连互通。朱甜甜^[4]运用 Hadoop 平台与数据挖掘技术对肿瘤数据进行分析挖掘, 获得肺癌与致病因素的关系。马骁^[5]通过 Hadoop 技术在集群上分析医疗数据生命周期, 有效提升存储设备的存储能力。本文从提升医院医疗质量管控角度出发, 设计基于 Hadoop 架构的医疗管控系统。

2 需求分析

2.1 问题描述

医疗数据具有体量大、多样性等特征, 其数据

〔修回日期〕 2021-08-30

〔作者简介〕 李鹏泉, 硕士, 工程师, 发表论文 2 篇。

分析没有得到充分利用^[6]。例如医院的药占比、收入情况、病种分布情况等得不到及时查询，主要原因是缺少一套能够实时查询科室、治疗过程关键数据并能进行数据分析的系统，无法迅速做出调整。只能使用结构化查询语言（Structured Query Language, SQL）查询提供简单报表，其时效性不高且数据分析质量存在不足^[7]。需要构建医院大数据平台并在一定程度上实现数据整合、统计分析 & 报表导出功能^[8]。

2.2 技术分析

Hadoop 是分布式系统的一种基础架构^[9]，是大数据分析平台常用开源架构，其特点是对海量数据进行分布式管理与存储。Hadoop 技术已成为顶级公司的分布式系统架构首选。Hadoop 具有以下优点^[10]：高可靠性^[11]、高可扩展性、高效性；成本低，可以采用普通 PC 机构建集群，不需要购买小型机、刀片服务器等较昂贵的设备。系统开发使用 Java 语言、J2EE 技术，结合 SSM（Spring + SpringMVC + Mybatis）开发模式，确保良好兼容性。

2.3 建设目标

利用大数据及智能设备技术将医疗数据转换成管理人员需要的医疗指标，提供线上线下相结合的公众健康服务，为领导层、医生等提供辅助决策，满足相关人员多层次、多方位的需求。将现有和历史医疗数据盘活，开展有价值的数据分析。系统应集成数据中心，提供高并发、高并行的读写和计算能力^[12]。每日产生新增数据，数据中心对历史数据进行采集^[13]。充分利用医院现有数据，融合既往指标，开展医院精细化管理^[14]。

3 系统总体架构

3.1 技术框架

3.1.1 数据层 主要是通过 Sqoop 采集相关数据，将医院信息系统（Hospital Information System, HIS）历史数据导入，采用分布式文件系统（Hadoop Distributed File System, HDFS）对 HIS 数据进行分布式

存储^[15]。HIS 数据可分为两类：一类是常年积累的历史数据，一类是业务查询数据。

3.1.2 控制层 主要是利用 MapReduce 对离线数据进行预处理、分析、离线计算。患者医疗费用数据分析和挖掘通常在 Map 函数和 Reduce 函数中执行，最终执行结果供应用层使用。

3.1.3 应用层 对数据整理结果进行可视化呈现。要做好接口工作、前后端数据交互，使用可视化工具 Echarts，系统结合 SSM 框架，采用模型 - 视图 - 控制器（Model View Controller, MVC）开发模式，实现数据信息的传输和查询，见图 1。

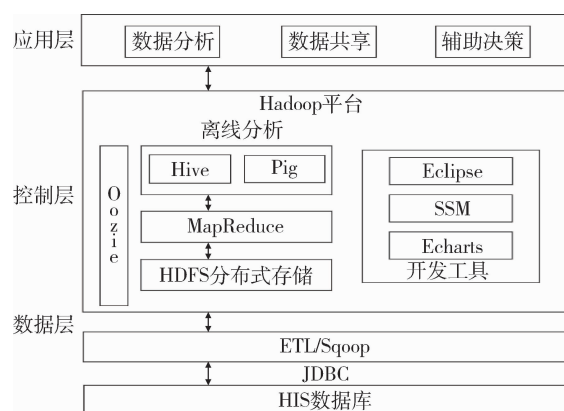


图 1 数据处理框架

3.2 医疗大数据处理

3.2.1 数据采集 通过 Sqoop 组件对 HIS 数据库中患者费用、医嘱等相关数据进行采集。Sqoop 第 1 次为全量采集，之后每天 01:00 定时导入增量数据并将其存放在 HDFS 中。

3.2.2 数据预处理 通过 Sqoop 导入的 HIS 数据存在不完整、不一致情况，为提高医疗数据分析质量，保证后续处理数据格式统一，获得规则的结构化数据，必须进行数据预处理。本系统使用 MapReduce 对元数据进行预处理，具有以下优势：一是 MapReduce 应用 Java 语言开发，熟悉度较高；二是 MapReduce 的分布式计算方式并发处理效率高。

3.2.3 数据入库 将预处理好的结构化数据导入数据仓库 Hive。使用 MapReduce 进行数据分析效率不高，且写代码工作量巨大，因此使用 HiveQL 对

相关数据进行分析。数据入库主要步骤如下：从数据源导入数据到数据仓库的操作数据存储（Operational Data Store，ODS）层（源数据），数据经过提取、转换、加载（Extract - Transform - Load，ETL）到数据仓库（Data Warehouse，DW）层并确定分析主题，基于主题寻找相关联事实表数据，根据业务需求确定分析维度，对数据进行采集、预处理，再把数据填充到数据仓库创建好的事实表中并映射成功，展开后续分析工作。

3.2.4 数据分析 将数据映射成 Hive 当中的表，根据业务需求、业务指标编写相关 HiveQL，分析计算出需要的指标和结果。所编写 HiveQL 分析语句被 Hive 编译成 MapReduce 任务提交到 Hadoop 的 YARN 上运行。数据分析是根据业务需求分析提供的指标在系统中不断更新、维护的过程，根据需求编写 HiveQL。

3.2.5 数据可视化 使用图形表格的形式将分析结果和规律进行展示，也可称为数据报表。HiveQL 分析结果通过 Sqoop 导出到 HIS 的 Oracle 数据库，数据导出到持久层，使用前端可视化工具 Echarts 对分析结果做报表。Echarts 属于前端技术开发的基于 Javascript 的可视化工具。根据数据分析的需求选择创建折线图、饼状图、柱状图等多图表混合展示。

3.3 数据可视化后端 Web 架构

使用 Maven 的 Tomcat 插件启动项目，将导出在 Oracle 的数据按照需要的格式查询返回给前端。当用户使用浏览器访问页面发起请求，请求被表现层接收，再调用业务层，对具体业务进行操作，持久层对数据进行增删改查，确保业务落实在数据库中。医院的 Oracle 数据创建好相应数据库表，数据仓库 Hive 将数据分析结果通过 Sqoop 导入 Oracle 存储。DAO 层即可访问 Oracle 中的数据分析结果，返回给业务层，业务层对相关数据进行转换调整返回给表现层，最后展示在浏览器。在表现层使用 SpringMVC，业务层使用 Spring 及其生态圈，持久层使用 Mybatis，见图 2。

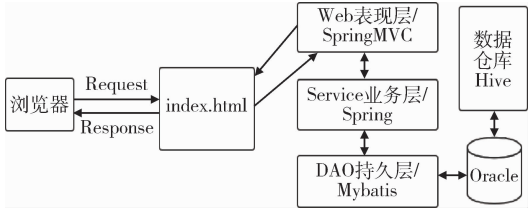


图 2 后端 Web 架构设计

3.4 系统结构

3.4.1 整体架构 基于 Hadoop 的医院管控系统主要分为 4 个功能模块，见图 3。

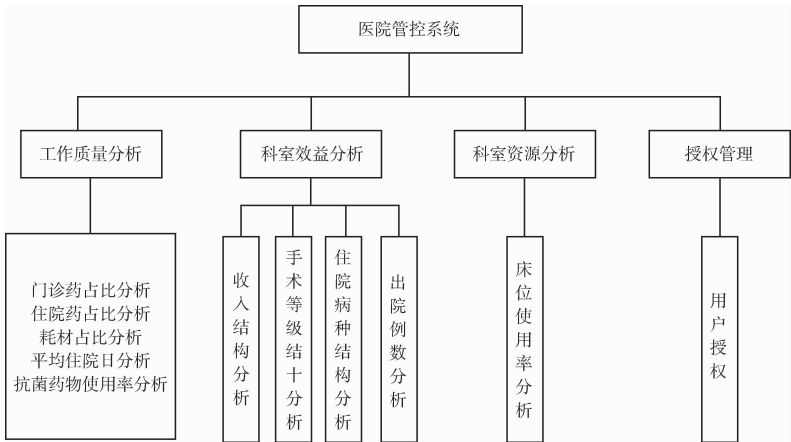


图 3 系统结构

3.4.2 功能分析 （1）药占比分析。查看全院上周、上月、上季度、上一年的药占比趋势；查询

目前全院、某个科室药品使用比例；能够按时间粒度查询；查询个人药品总费用、西药和中药占比；

能够下载打印报表。(2) 材料占比分析。查看全院上周、上月、上季度、上一年耗材占比趋势；查询目前全院、某个科室、个人耗材使用占比；能够按时间粒度查询，比较今年与去年的耗材占比；能够下载打印报表。(3) 平均住院日分析。查看全院和各临床科室平均住院日，对科室今年趋势与去年同一时期比较，可设置目标平均住院日。(4) 抗菌药物使用率分析。查看全院以及所有科室抗菌药物使用率，今年跟去年同一时期趋势比较；能够按住院号、药物名称、日期时间检索；查看在院和出院情况；查看抗菌药物使用率、医生排名和药物排名。(5) 收入结构分析。查看全院和各科室总收入情况，科室今年趋势与去年同一时期比较，收入汇总包括总收入、纯收入、住院总收入、门诊总收入。(6) 手术等级结构分析。查看全院和各科室手术等级结构，包括手术例数、1 级手术例数、2 级手术例数、3 级手术例数、4 级手术例数、3 级以上手术率。(7) 病种结构分布分析。查看单病种患者在各科室分布情况，统计全院病种总例数，统计病种名称分布。(8) 出院例数分析。查看全院和各临床科室出院例数情况，科室今年趋势与去年同一时期比较。(9) 手术例数分析。查看全院周次手术例数变化和上周、上月、上季度手术例数，同比和环比变化。(10) 床位使用率分析。查看全院或临床科室床位使用率情况，信息包括床位使用率、床位周转数、实际床位工作日、实际平均床位工作日。

4 Hadoop 集群配置

4.1 硬件环境

系统数据分析模块应用 Hadoop 技术，搭建环境为医院网络中心，数据源基于医院 HIS 平台 Oracle 数据库。医院中心机房已实现虚拟化管理，光纤交换机和存储等设备已相互连接，只需使用管理集群虚拟化软件 Vmware vCenter Server 分配资源、创建服务器即可。

4.2 集群配置

安装 3 台虚拟机 Linux CentOS7 操作系统，主服

务器 1 台作为 Namenode 主节点，两台 Datanode 为从节点，见表 1。安装 JDK8u161、Hadoop 2.7.4、Hive 2.1.1、Sqoop 等关键组件，开启服务，对 HIS 数据库数据抽取测试。Oracle 数据库与 HDFS、Hive 之间导入导出主要通过 Sqoop 组件完成。

表 1 集群配置信息

服务器名	类型	CPU (Hz)	内存 (G)	硬盘 (T)	IP
Mbigdata1	NameNode	2.4	64	2	192.1.176.111
Sbigdata2	DataNode	2	32	1	192.1.176.112
Sbigdata2	DataNode	2	32	1	192.1.176.113

5 关键医疗数据分析结果展示

5.1 系统导航

在系统首页导航中可以清晰看到全院住院药占比、耗材比、抗菌药物使用率、床位使用率、当天手术例数、出院人数、平均住院日、总收入等关键数据。

5.2 药占比趋势分析

用户可以选择以周、季度、年为时间单位查看全院各临床科室药占比趋势分析。过度医疗或者高额医疗费用通常都是在用药部分占较大比例，应合理管控各住院科室药占比，避免出现天价药费情况。同时对药占比较为突出的科室进行整顿或者经济处罚，使管理层有据可依、奖罚分明。

5.3 抗菌药物使用率分析

可对比分析今年和去年抗菌药物使用率。滥用抗菌药物会引发患者产生耐药性、菌群失调、身体不良反应等危害。系统对滥用抗菌药物提供数据支撑^[16]，帮助医院严格管控抗菌药物使用。

5.4 病种结构分析

近年来开始实施按病种收付费用，效果并不理想，符合病种付费标准的患者费用差异较大。病种分布饼状图和条形图帮助管理者了解各病种患者数

量和病种的占比率,图形显示包括疾病名称、患者例数、病种百分比。

5.5 床位使用率分析

床位使用率 = 实际使用床位数/实际床数。床位使用率分析界面反映的是床位使用情况和负荷情况。显示内容包括科室名称、床位使用率、床位周转次数、实际床位工作日、实际平均床位工作日。

6 创新性特点

传统方式主要依靠统计人员经验或简单的查询系统统计临床指标,针对多个部门需要做不同报表,时效性难以保证,难以得到满意的结果。大数据医疗质量管控平台硬件投资成本较低,多层次处理和自动负载均衡,自动收集、清洗、转换所需数据保存到大数据平台数据仓库中。提供完整的数据信息链处理和深度数据挖掘,智能化信息交互和工作协同体系,操作便捷快速,可检索所需的结果、打印报表。同时大数据医疗质量管控平台能够将所有医疗指标分类排列,根据不同角色人员灵活分配所需数据指标,还能够对系统中对所选指标窗口自由布局,满足不同用户浏览习惯。医疗管控平台充分整合和挖掘医院历史数据,从多个角度对运营指标进行全方位、精细化管控,改变传统、简单的管控模式,美观、高质量的数据图表有效辅助医院管理决策,加强了医疗质量管控。

7 结语

医疗质量管控系统应用能够及时监控医院药占比、耗材占比、抗菌药物使用率等关键数据。对病种结构、平均住院日、床位使用率、出院例数、手术例数分析等能够辅助医院管理层做决策。系统能够满足医院医疗质量管理需求。随着大数据技术日益成熟,基于 Hadoop 架构难以充分满足数据检索查询速度需要,而基于 Spark 内存的批处理模式要

比 MapReduce 更加高效,今后将引入 Spark 框架提升系统运行效率。

参考文献

- 1 马敬东,张学高,李岳峰,等.国内外健康医疗大数据资源标准体系研究进展[J].中国卫生信息管理杂志,2019,16(3):257-262.
- 2 衡反修,王力华.医疗机构医疗大数据平台建设指南[M].北京:电子工业出版社,2019.
- 3 吕道明.基于 Hadoop 平台的临床数据集成方案设计与实现[D].杭州:浙江大学,2015.
- 4 朱甜甜.基于医疗大数据的肿瘤疾病模式分析与研究[D].青岛:青岛科技大学,2018.
- 5 马骁.大数据环境下的医疗数据生命周期研和实现[D].北京:北京邮电大学,2019.
- 6 孟若谷,杨羽,张路霞.健康医疗大数据质量评估方法进展与展望[J].中国卫生信息管理杂志,2019,16(6):677-681.
- 7 温煜,唐丹,徐双平,等.基于大数据架构的医院数据中心管理[J].中国数字医学,2018,13(12):5,15-16.
- 8 曾航齐,黄桂新,伍君贤,等.基于 Hadoop 的医疗大数据智能辅助诊疗平台的构建[J].中国数字医学,2018,13(9):96-99.
- 9 许欢.安徽省中医院信息化建设的现状及对策[D].合肥:安徽大学,2015.
- 10 Hurwitz J S, Nugent A, Halper F, et al. Big Data for Dummies [M]. Hoboken: John Wiley & Sons, 2013.
- 11 刘晓群,邹欣,范虹.基于并行云计算模式的建筑结构设计[J].电子技术应用,2011,37(10):123-125,130.
- 12 周颖.基于大数据机器学习的客服系统的研究[D].贵阳:贵州大学,2018.
- 13 张祥儒.医疗数据信息平台数据采集系统的设计与实现[J].电子技术与软件工程,2019(1):147-148.
- 14 沈贝敏.基于数据中心的医院精细化管理系统的建立及应用[J].中国管理信息化,2017,13(14):34-35.
- 15 生慧.大数据背景下中医电子病历关键问题研究[D].济南:山东中医药大学,2017.
- 16 林顺和,黄璇璇,林亚忠.基于微信公众号的移动决策支持系统的设计与应用[J].中国数字医学,2019,14(12):39-41.