

基于数据仓库技术的医院信息系统(HIS) 的实现方案研究

The Research of Implement Solution of Hospital Information System(HIS) Based on Data Warehouse

谷 岩 (广州大学 510091)

摘要:数据仓库和联机分析处理(OLAP)技术是信息技术领域的新兴技术,而如何应用到医院的信息化建设中是医院信息系统(HIS)面临的问题。建立基于HIS的数据仓库并使用OLAP技术,可以将大量源数据有效地转化为有用的知识信息,并服务于决策过程。基于数据仓库的HIS就是在面向信息处理的联机事务处理(OLTP)系统的基础上并引入数据仓库和OLAP技术而发展起来的。本文结合医院医疗业务提出了系统实现的解决方案,并给出了基于医院医疗业务的数据仓库的体系结构、逻辑模型、粒度模型,以及基于医院医疗业务的联机分析处理(OLAP)系统的实现方法。

关键词:医院信息系统 数据仓库 联机分析处理 粒度模型

1 引言

通过近二十年的发展,基于联机事务处理(OLTP)的医院信息系统(HIS - Hospital Information System)的建设已初具规模,它不仅促进了医院管理水平、服务质量和经营效率的提高,同时也使医院的医疗水平迈上了新台阶。但随着业务规模的不断扩大,医院已积累了大量的数据,其容量呈几何级数增长,但由于传统的事务处理系统的局限性,新的问题也开始应运而生:由于产生了大量的历史数据,如何管理这些历史数据已成为迫在眉睫的问题;现有的信息处理系统只能实现业务流程的自动化,而如何对大量的历史数据进行深层次地挖掘,从中发现医疗业务和经营的内在规律,从而为管理人员的决策提供支持已成为医院管理层迫切需要解决的问题。因此采用数据仓库技术、联机分析处理(OLAP)技术、数据挖掘技术以及已成功应用的OLTP技术进行HIS的开发已成为新的趋势。基于此,本文将在已有的HIS工程实践的基础上,对基于数据仓库的HIS进行分析与设计,并最终提出一个可行的解决方案。

2 数据仓库技术

数据仓库是面向主题的、集成的、随时间变化的、非易失性的数据集合,用于支持管理层的决策过程。数据仓库的创建和使用都是围绕着主题的,如HIS中数据仓库是以医疗业务数据分析、病人结构分析、病人流动状况分析、床位占用分析、资金流动分析为主题而逐步展开的。数据仓库中的数据不是将业务处理系统中的数据简单集合,而是对各种源数据进行抽取、筛选、清理、综合而得到的数据集合。数据仓库中所存储的数据不经常进行更新处理,它主要用于查询和分析。一般数据仓库中的数据时限为5至10年。

OLAP是一种决策分析工具,它可以根据分析人员的要求,快速、灵活地对大量数据进行复杂的查询处理,并以直观的、易理解的形式将查询结果提供给各种决策人员,从而得到高度归纳的信息。OLAP是基于数据仓库的信息分析处理过程,是数据仓库的用户接口部分。

3 系统总体结构

3.1 基于数据仓库的 HIS 体系结构

通过对医院医疗业务处理过程的分析,我们可以将系统划分成 3 部分:

(1) 联机事务处理(OLTP)子系统:它是面向医院医疗业务层,完成日常的业务处理,其数据将成为数据仓库的数据源;

(2) 联机分析处理(OLAP)子系统:它是面向决策层,完成数据的统计与分析,其数据来源于数据仓库;

(3) 数据管理子系统:用于数据仓库中的元数据及数据仓库中数据的日常维护。

3.2 HIS 中数据仓库的体系结构

HIS 的 OLTP 子系统包括了医疗业务、药品、物品、设备、行政后勤等源数据库,根据需要从这些源数据库中抽取所需要的数据到数据准备区,然后对数据准备区中的数据进行净化处理,最后加载到数据仓库中。在数据仓库中不仅包含了医疗业务、药品、物品、设备、行政后勤等当前的和历史的明细信息,还包含有按照不同层次和粒度汇总而成的综合数据。数据仓库的体系结构见图 1。

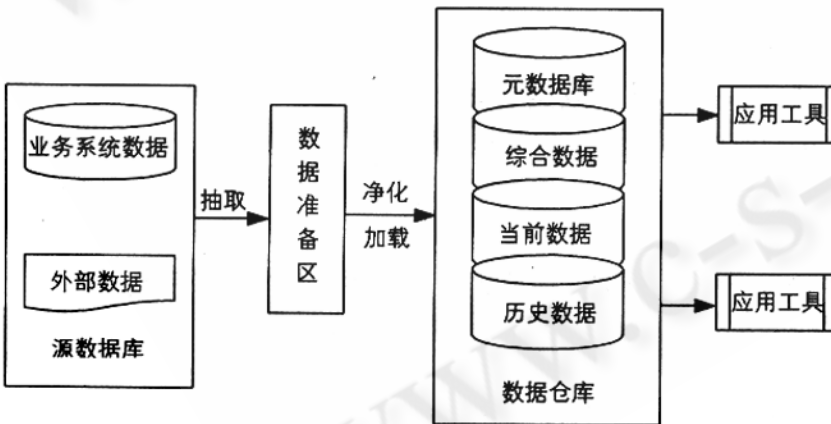


图 1 HIS 中数据仓库的体系结构

4 基于 OLTP 的 HIS

新的基于 OLTP 的 HIS 突破了过去“以财务为核心”的旧的管理模式,开始转向“以病人为中心”新的管理模式,全面支持医院的医疗和管理业务,同时面向医院管理诊断理念进行 HIS 的开发。在系统集成方面,全面实现医院业务数据的集成,实现医院数据的实时产生、更新、查找和使用。在实现技术上,采用先进

的网络通讯技术、大型数据库技术、系统安全技术和面向对象的软件开发技术,保证系统的先进性。新的 HIS 系统将涵盖临床信息管理、财务收费信息管理和行政信息管理等业务,因此通过系统分析,我们将 HIS 划分为 3 个业务子系统及 15 个功能模块。同时在三层 C/S 结构的基础上,我们将 HIS 划分为了四个逻辑层次:界面显示层、业务模块层、通用组件层和数据库系统层,其中界面显示层对应三层 C/S 结构的客户层,业务模块层和通用组件层对应三层 C/S 结构的业务逻辑层,数据库系统层对应三层 C/S 结构的数据服务层。系统的总体结构如图 2 所示。

5 数据仓库技术和 OLAP 在 HIS 中的实现方案

5.1 系统的软硬件平台

数据仓库的硬件平台:可采用 SUN 工作站作为数据仓库服务器,采用一般工作站作为 OLAP 应用服务器。

数据仓库的软件平台:服务器和前台的软件环境为 Windows 2000,数据库管理系统为 MS SQL Server 2000。

5.2 数据仓库逻辑模型的设计

常用的数据仓库建模方法有两种:关系数据库和多维数据库。在 HIS 系统中我们采用关系数据库对数据仓库进行建模,而数据仓库的逻辑模型则用雪花图来表示。

数据仓库中的数据是面向主题进行组织的。根据医院的业务特点,我们可以确定“医疗业务”是数据仓库的主题,通过对主题的展开,它所包含的子主题包括:医疗业务数据分析、病人结构分析、病人流动状况分析、床位占用分析、资金流动分析。根据主题集中涉及的数据信息,选择医疗业务、资金流动、床位占用、病人流动为发生事实,并选定与之相关的医疗业务维、病人维、账单维、处方维、床位维、时间维,从面向 OLTP 的业务系统中抽取有效数据。数据仓库的雪花形逻辑结构如图 3 所示。

5.3 数据仓库粒度模型设计

在数据仓库的开发过程中还要解决的重要问题就

是构造数据仓库的粒度模型。粒度划分层次适当与否

中,主要面向的处理是分析型处理,因此一般将粒度划

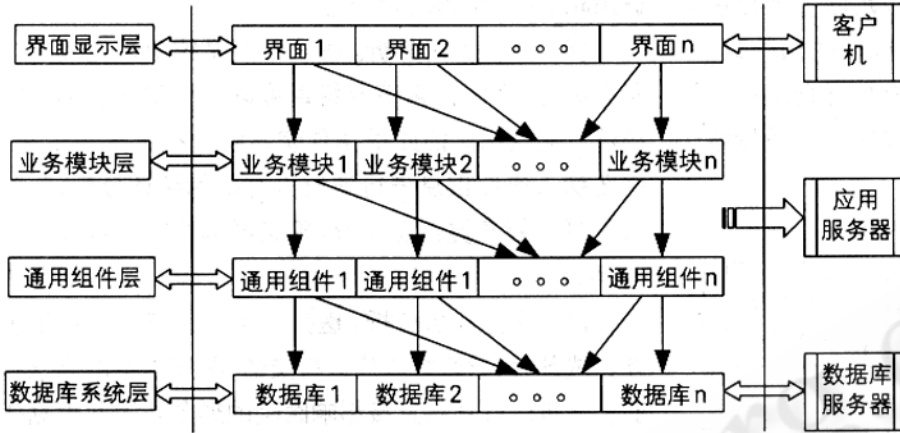


图 2 HIS 的总体结构图

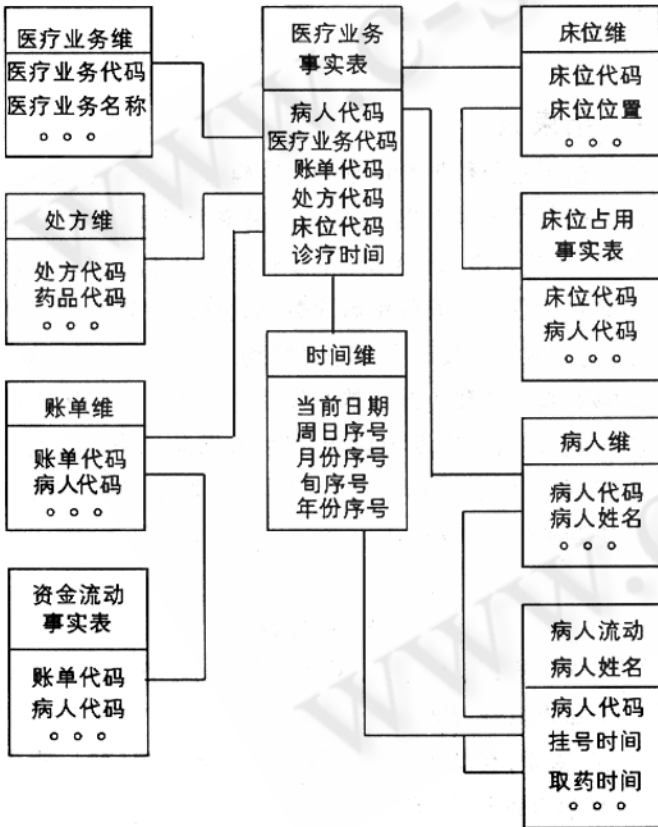


图 3 数据仓库的雪花模型

直接影响到数据仓库中的数据量和所适合的查询类型。所谓粒度是指数据仓库中数据单元的详细程度和级别。数据越详细,粒度就越小,级别也就越低;数据综合程度越高,粒度越大,级别也就越高。在数据仓库

分为详细数据、轻度综合、高度综合三级。在对医院业务数据综合时,按时间汇集粒度可分为如下粒度层次:按年综合、按旬综合、按月综合、按日综合。

5.4 数据仓库的物理实现

利用 SQL Server 2000 可以进行数据仓库的物理实现,具体的实现步骤如下:

(1) 创建数据准备区。为能对准备装入数据仓库的数据进行提取、清理和加载,在数据仓库中需要创建一些独立的数据库作为数据准备区。在数据准备区中可以从数据源中析取数据并转换为数据仓库的常用格式,检查数据的一致性和引用完整性。数据准备区的创建可以采用 SQL Server2000 中的数据库与表创建工具实现。

(2) 创建数据仓库。数据仓库的框架通常由事实表和维表组成。数据仓库的创建同样可以采用 SQL Server2000 中的数据库与表创建工具实现。

(3) 从业务系统提取数据。数据仓库中的数据主要是从业务系统中的数据源获取,在提取过程中需要对数据源的数据进行抽取。在 SQL Server2000 中可以使用 Transact - SQL、DTS、分布式查询、命令行应用程序、BCP 实用工具及 BULK Insert 语句和 ActiveX 脚本实现数据的提取。

(4) 清理和转换数据。提取出来的数据还需要进行数据的一致性检查、格式化处理等清理工作,并对数据进行必要的转换。在 SQL Server2000 中可以使用 Transact - SQL、DTS、命令行应用程序及 ActiveX 脚本实现数据的清理和转换。

(5) 将数据加载到数据仓库。可以使用 Transact - SQL、DTS 和 BCP 实现数据的加载。

6 联机分析处理(OLAP)系统的实现

6 联机分析处理(OLAP)系统的实现

OLAP 技术通过对数据仓库进行综合、统计、分析,以专业报表、查询结果的形式提供给管理人员的决策过程,最终形成决策数据。OLAP 是一种多用户的三层 Client/Server 结构,它由数据仓库、OLAP 服务

器、OLAP 客户机及客户端应用程序构成。复杂的应用逻辑集中存放在应用服务器上,由应用服务器提供高效的数据存储,并安排后台处理及报表的预处理。OLAP 服务器设计的重点在于如何组织数据仓库中的综合数据,如何满足前端用户的查询要求。其逻辑结构图见图 4。

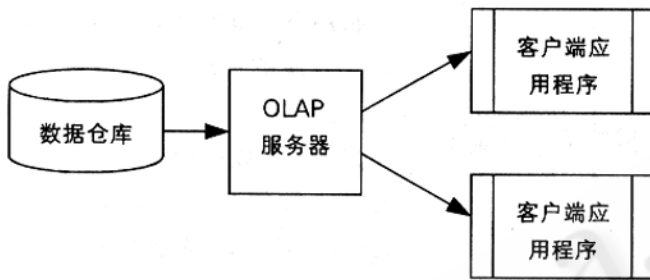


图 4 OLAP 的三层逻辑结构

我们可以采用 MS SQL Server 2000 Analysis Services 提供的 OLAP、Microsoft English Query 等工具进行即时查询,同时采用 Visual Basic、Powerbuilder、Delphi 等开发工具开发应用程序,实现对数据仓库中数据的查询。开发的客户端应用程序包括如下功能:

(1) 病人结构分析。可以按照性别、年龄、文化程度等对病人进行分类,根据每项分类,系统将对病人的经济状况、需求特征和购买行为进行分析,从而得到不同性别、不同年龄、不同文化程度病人的经济水平、需求状况及主要医疗服务类型等信息。根据这些信息,医院管理者可以分析病人差异对医院收益的影响,帮助医院进行市场定位、确定营销策略,从而使医院的经营活动更具主动性和目的性。

(2) 资金流动分析。针对医院的资金流动状况,按不同的时间维度(包括按年综合、按旬综合、按月综合、按日综合)对医院的各种资金流动情况进行分析,进行资金的同期对比动态分析,门诊和住院收入因素分析,并以各种专业报表、查询结果的形式反映给决策者。

(3) 病人流动状况分析。根据门诊病人从挂号到取药在医院逗留的时间进行时间数列动态分析,掌握影响病人诊疗效率的因素,帮助医院管理者进行业务流程的更新和改造,以提高医院的经营效率。同时帮

助医院“以病人为中心”整合医院所有对外服务,以促进医院的医疗和管理水平。

(4) 床位占用分析。医院住院系统中最重要的对象是床位,它的占用情况将直接影响医院住院部分的经济效益状况。因此根据住院系统的病区床位信息、病区病人床位信息进行床位占有率的统计分析,帮助医院管理者掌握经营状况,确定影响医院经济效益的因素。

(5) 医疗业务分析。医疗业务是医疗活动中的重要组成部分,它直接与病人打交道,医疗水平的高低和服务质量的好坏将直接影响医院的形象。因此根据病人诊断及治疗方面的历史数据,按不同的时间维度(包括按年综合、按旬综合、按月综合、按日综合),以已开展的医疗业务为类别,统计出各个医疗业务的开展数量,从而对医院重点开展哪些医疗服务类型进行辅助决策。

7 结束语

由于基于数据仓库的 HIS 规模大、涉及的问题多,因此在整个开发过程中必须严格遵循从上到下、先粗后精、化整为零、分而治之的现代软件开发思想。应用该方法不仅可以有效地解决数据仓库的管理和使用问题,同时能及早发现系统整体结构上存在的问题,从而能够圆满完成整个系统的开发工作。本文所提出的解决方案,是数据仓库和 OLAP 技术在医院信息化建设中的应用的一个初步探索,我们相信对 HIS 的开发具有十分重要的借鉴作用。

参考文献

- 1 陈京民等,数据仓库与数据挖掘技术[M],北京电子工业出版社,2002。
- 2 王珊等,数据仓库技术与联机分析处理[M],北京科学出版社,1998。
- 3 Lou Agosta (2000). The Essential Guide to Data Warehousing[M], Prentice - Hall.
- 4 William A Giovino(2000). Object - Oriented Data Warehousing Design[M].
- 5 刘同明等,数据挖掘技术及其应用[M],北京国防工业出版社,2001。