

基于 ROLAP 医保查询系统设计与实现

徐沛东 邵佩英 中国科技大学计算机系(230026)

Abstract

This article dissertated that OLAP was used in medical insurance management system. Joined data warehouse with OLAP to provide decision-making analytical information for managers. Because data warehouse is a huge and complicated system, it is not quite extensive application in practice. Considering actual requirement in medical insurance management system, we put forward multi-dimension database modeling idea based on ROLAP and its concrete design and realization procedure.

Keywords: OLAP, OLTP, ROLAP

摘要

本文论述了将 OLAP 用于医疗保险管理系统。把 OLAP 与数据仓库结合起来,为管理者提供决策分析信息。由于数据仓库是一个庞大而复杂的系统,在实际中并没有得到十分广泛的应用。考虑到医疗保险管理系统的实际要求,我们提出了基于 ROLAP 的多维数据库建模的思想以及具体设计实现过程。

关键词:联机分析处理,联机事务处理,关系型联机分析处理

联机分析处理 (Online Analytical Processing, 缩写为 OLAP) 是用于对数据仓库进行查询的一种技术,它使管理人员通过对信息的多种可能的观察角度进行快速、一致和交互性的查询以获取对信息的深入理解,尤其通用于对大量数据进行快速的查询。ROLAP 表示关系型联机分析处理,ROLAP 遵循以下三个规则:支持基本的 OLAP 的概念;数据存储在某一关系型数据库中;支持某种形式的聚集导航。

1 系统数据流程

医保中心将医院数据采集后,首先进入医保数据库,该数据库是关系型数据库,然后经过装入,即对某些数据进行相应的处理,进入数据仓库。其数据流程如图 1 所示。

本系统构建的数据仓库不是完全意义上的数据仓库,它是借助于数据仓库的思想在关系型数据库引基础上构建的数据库,为了区别于 OLTP 所用数据库,我们称它为数据仓库。

2 数据仓库中的数据模式设计

一般数据模型可用三维空间来描述:详细维、时间维、区间维。

(1) 详细维(level of detail)

该维显示的是分析过程所需的数据形式:详细数据(detailed)和汇总数据(summary)。详细数据即是

未经过任何处理的原始数据值,这是整个分析过程的基础。汇总数据是在详细数据基础上加工处理后得到的新数据,往往是按照一定的层次对详细数据进行的聚合或累加。

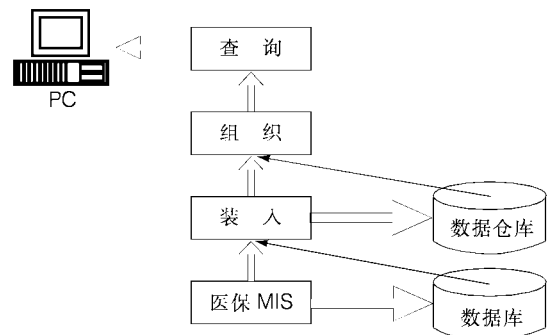


图 1 数据处理示意图

(2) 时间维(time)

时间维体现了数据的变化周期:历史数据(historical)、当前数据(current)和未来数据(future)。历史数据代表在某一特定时刻的数据状态,因此,它是静态的、不可变的;当前数据不仅包括正在操作中的各种数据,也包括近期内某一时刻数据的快照;未来数据则是对未来某一段时间内的预测信息。由于这种预测受到多种外界因素的影响(如,取值范围、挖掘算法等),所以对未来同一时刻也可能会有不同的结果。大多数的分析预测工作是基于历史数据的。

(3) 区间维(scope)

这一维反映了数据所在的区间范围。

这种数据模型具有以下几大特点:

- 这种模型非常适合于决策系统的分析过程。
- 该模型能够为复杂的数据管理提供有价值的信息。

例如,详细维能够体现通过不同汇总方式处理后,企业信息值的变化。

· 也可以用这种模型把一个事物分配到各个相应空间中进行处理。例如,可以把日常处理任务分配给当前详细数据空间;操作过程中的决策支持行为可在当前汇总数据空间进行;事物分析处理则由历史数据空间完成;战略性的决策支持行为则可分配给历史汇总数据空间。

在这一数据模型中,每一个数据空间都可以有不同的结构类型。对操作数据而言,最好采用规范化的、关系型结构,以减少数据冗余。由于汇总数据常常用于分析和报表,所以可采用星型模式或维模式。

3 数据仓库结构设计

根据上面所论述的建模思想,我们把医疗保险管理数据库的逻辑结构分为了以下几个部分:

1) 当前的详细数据。这一部分主要是未经过任何处理的原始详细数据,也就是日常业务所要处理的信息即 OLTP 数据。由于这些数据大多是在传统数据库所要存放的数据,因此对于这一部分数据的具体数据库设计,可以在传统数据库设计的基础之上稍加改动,除了原有的事实表外,适当的增加维表。这样就使得其在支持原来的 OLTP 功能的同时,还增加了对 OLAP 的支持。

2) 当前的汇总数据。这里的“当前”主要指近期的,可以是当天、本周或本月。但由于对这种当前较短范围之内汇总信息的查询或报表都属于简单查询,只需把详细数据进行简单计算即可得到,而且大多没有什么保存价值。因此,在实际的设计实现中,并不必设定单独的表保存这些信息。而对于有保存价值的月末汇总信息,则可在查询或报表后,将其结果加上时间戳作为历史汇总数据存入相应的表中。

3) 历史汇总数据。即保存按时间(如:月、年)汇总后的统计信息。这些信息主要用于数据比较和分析。

4) 未来的数据。这里所说的未来数据主要是指一些预定计划信息,比如:预定缴费额等。

对于历史的详细数据,在本数据库设计时并未提供历史详细表进行单独保存。首先,本数据库中的当

前详细数据表中存放的是相邻两年内的数据,在绝大多数情况下已经可以满足用户的查询需要;其次,如果在数据库中单独增加一些历史数据表去保存以往多年内的详细数据,其所需空间将会很大,从而影响整个数据库的运行效率。最后,考虑到每年都会对数据库进行一次备份和清库,即使用户真的需要查询五年前的详细信息,也可以从备份数据库中得到结果。

4 数据仓库的具体实现方案

4.1 建立详细表

构建传统数据库,首先应该确定的就是库中的主要详细信息表,也就是该数据库所要管理的主要内容。根据医疗保险的业务范围,我们建立了相应的详细表。例如,医疗数据表、基本医疗缴费表、高额补充缴费表、医院结算表、个人帐户收支表、医疗明细表、更新结果表、住院登记表、单位档案表、个人档案表、个人年度工资表、个人帐户状态变化记录表、个人缴费变动记录表等等,医疗数据表结构如表 1 所示。

表 1 医疗数据表结构

医疗数据
保险编号
就诊日期
就诊类别代码
职工种类代码
帐户支出
社会统筹支出
个人负担
诊断编码
医师编码
……

4.2 确定事实、维和粒度

由于整个医疗保险管理系统中医疗数据管理是重要部分,因此,在数据库具体设计时,就以医疗数据表作为一个基本事实表,其主要内容如表 3 所示。

通过对该表的分析,不难确定出其对应的维,主要包括:时间、就诊类别、职工类别。而每个维所要划分的粒度如表 2 所示。

表 2 维的粒度划分

维	粒度
时间	日、月、季、年
就诊类别	类别代码、类别名称
职工类别	种类代码、种类名称

同样,按上述方法可以对其它部分的信息进行分析,以确定出其相应的事实、维和粒度的关系。用星型模型表示的数据库主要事实、维和粒度之间的关系,

如图 2 所示。

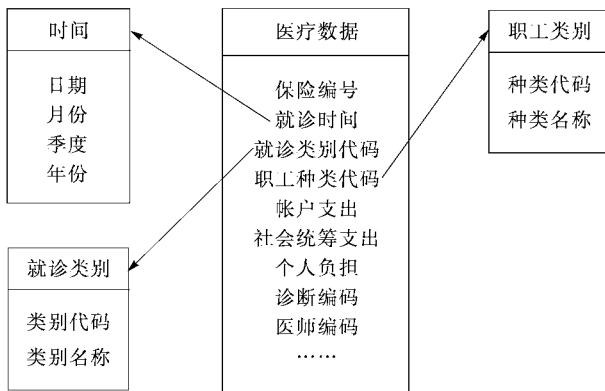


图 2 数据库的星型模式

表 3 是其对应的多维报表。

表 3 医疗数据库多维报表

	一月份			...	十二月份		
	帐户支出	社会统筹支出	个人负担	...	帐户支出	社会统筹支出	个人负担
门诊	在职						
	退休						
	退老						
住院	在职						
	退休						
	退老						
病床	在职						
	退休						
	退老						
合计							

4.3 建立合计表

由于传统的操作型数据库在设计时,要求采用规范化的、关系型结构,以减少数据冗余,因此,在传统数据库中很少建立合计表。而在支持 OLAP 的数据库中,合计表就成为非常重要的一部分内容。因为,在决策支持系统中,会涉及到许多复杂查询,其中包括对大量详细信息的汇总计算。如果能根据系统的具体要求,适当地增加合计表,虽然会产生一定的冗余信息,却可以在很大程度上提高系统的查询性能,这也正是决策支持系统所追求的目标。同时,值得注意的是,对于合计表的选择是非常重要的,并非是多益善。过多不必要的合计表只会浪费存储空间和合并更新时间,而性能提高不大。因此,在建立合计表时,一定要分析实际的查询情况,选择适当的合计表。一般情况下,汇总都是基于详细信息不同维和粒度的基础上进行的。通过以上步骤,整个数据库的基本结构就

已经建立起来了。当然,除了这些表外,还需要建立一些其它的系统表,以维护数据库的完整性。

5 结束语

把“基于 ROLAP 的数据仓库建模思想”应用于传统数据库建模过程之中所设计的医疗保险数据库系统,不仅保持了传统数据库所具有的 OLTP 功能,满足了医疗保险日常业务处理的需要,而且更加强了对 OLAP 的支持,为提高系统的查询决策能力提供了良好的基础。

在数据库的设计过程中,主要采用了“自底向上”的实施策略。首先,根据医疗保险的需要和业务特点,构建成一个完整的数据库。在设计具体的表结构时,充分考虑了 OLTP 与 OLAP 功能的结合。对于只需满足 OLTP 功能要求的数据,仍采用传统关系型数据库中表的设计原则;凡涉及到 OLAP 的信息内容,则在建表时把维、粒度的概念引入其中。同时,还适当地增加了合计表。

另一方面,由于该数据库是在传统关系型数据库的基础之上改造而成的,因此,它既不是完全意义上的数据仓库,又不同于传统数据库。虽然,它融合了传统数据库与数据仓库结构上的一些特点,但它对 OLTP 的响应不如传统数据库那么迅速——因为其数据结构已不再是完全规范化的关系型结构,而对于决策分析功能的支持又不可能象数据仓库那样强大——因为其数据存储量远远少于真正的数据仓库。所以,该数据库在性能上仍存在着许多不足之处,有待进一步改进。

下一步我们可以在 OLAP 的基础上,利用目前较为先进的数据挖掘技术(Data Mining),增加决策支持系统的预测和模拟功能。

参考文献

- 1 Boris Gendeleev, Closing the OLAP gap, Database Programming and design, April 1998:25~31
- 2 Tom Hammergren, Data Warehousing: Building the Corporate Knowledge Base, Ventana Communications Group, 1997
- 3 Joyce Bischoff, Ted Alexander, Data Warehouse, Practical Advice from the Experts, Prentice Hall Inc., 1997
- 4 Ralph Kinball, A Dimensional Modeling Manifesto, DBMS, 1997

[收稿日期:2001.11.11]