

保险业务联机分析处理系统的实现*

詹敏, 孙玉芳

(中国科学院 软件研究所, 北京 100080)

E-mail: zhanmin@sinosoft.com.cn

http://www.ios.ac.cn

摘要: 数据库技术的广泛应用以及数据仓库技术的出现为决策支持系统提供了新的发展方向. 联机分析处理(Online analytical processing, 简称 OLAP)正是在数据仓库的基础上发展起来的一门崭新的决策支持技术. 在给出用于保险业务的中间数据库的设计、多维分析模型的实现和 OLAP 分析系统的模块结构框架之后, 具体讲述了 OLAP 功能设置子系统以及分析子系统的实现, 全面满足 E. F. Codd 提出的 OLAP 的 12 条准则, 并在实际应用的基础上作出了相应的提高.

关键词: 决策支持; 数据仓库; OLAP(online analytical processing); 钻取; 旋转; 切片; 漂移; 数据挖掘

中图分类号: TP311 **文献标识码:** A

联机分析处理技术(online analytical processing, 简称 OLAP)^[1]是信息技术领域近年来兴起的一种决策支持手段. 在众多的决策支持技术中, OLAP 技术以其直观的数据操作、灵活的分析功能、可视化的结果表达等特点, 在数据仓库技术^[2]的支持下脱颖而出.

OLAP 是一系列软件工具的集合, 这些软件技术能够从不同的角度对从原始的营运数据转换过来的、以用户所能理解的方式表达企业真实维结构的信息进行快速、一致、交互的访问, 并且通过访问实现对数据本质内容的分析、管理以及运作. 简单地说, OLAP 就是进行多维分析的一系列工具的集合.

早在 1993 年, 我们就与中保集团下属的一些分公司开始了保险业务管理信息系统的合作开发. 那时, 开发的主要内容是 OLTP(online transaction processing)系统, 即日常业务处理系统, 同时也提供一些简单的联机查询功能. 这一系统对保险公司业务的发展起了很大的作用^[3]. 当该系统连续运行多年, 积累了大量的历史数据之后, 合作双方都意识到有必要对已有的历史数据进行分析并实行电子化管理. 于是就有了开发一套经理查询软件以供决策层准确、及时地了解各个业务部门的运作状况、正确地决策运营的需求. 但从严格的意义上讲, 第 1 期的经理查询软件并没有提供足够的分析功能, 而只提供了一组工具, 使用户能够方便地利用它们来创建、组合各种各样的查询模型, 并且利用微机良好的图形用户界面为决策者提供直观的图表分析. 该系统先后在广州、深圳、厦门、新会等地安装运行, 并取得了良好的效果. 但是, 这套经理查询软件还存在一些缺陷: 早期查询的数据源并未和营运数据库分离, 不能提供为获取某种分析结果而进行的连续操作. 这种不足一方面影响了 OLTP 系统的运作效率, 另一方面也不利于用户的直观操作. 为了解决这些问题, 我们开始着手建立中间数据库, 把用于查询的数据源从 OLTP 系统中分离出来, 这不仅提高了两个系统的运作效率, 而且也为开发通用的 OLAP 工具打下了必要的基础.

本文第 1 节给出了 OLAP 的系统设计, 包括中间数据库设计、多维分析模型建立以及 OLAP 分析系统结构框架的组成. 第 2 节简要讲述了 OLAP 各子系统的主要模块的实现以及系统的主要特性. 第 3 节给出对 OLAP 系统的一些评价和今后的工作设想.

* 收稿日期: 1999-06-18; 修改日期: 1999-08-14

基金项目: 国家自然科学基金资助项目(19831020); 国家 863 高科技项目基金资助项目(863-306-ZD-07-4)

作者简介: 詹敏(1974—), 男, 福建闽清人, 博士生, 主要研究领域为大型数据库, 网络工程; 孙玉芳(1947—), 男, 江苏张家港人, 研究员, 博士生导师, 主要研究领域为计算机操作系统, 中文信息处理, 大型数据库, 网络工程.

1 OLAP 的系统设计

1.1 中间数据库的设计

我们的 OLAP 系统虽然是一种通用的 OLAP 分析工具, 但却是成型于中国人民保险公司广州市分公司的经理查询环境之中. 因此, 这里有必要先介绍一下该分公司的行政框架(如图 1 所示)和业务框架(如图 2 所示).

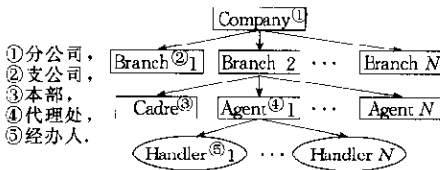


Fig. 1 Administration hierarchy
图1 行政框架图

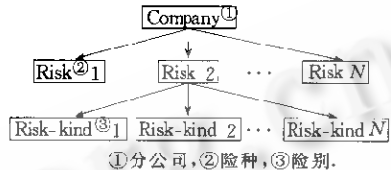
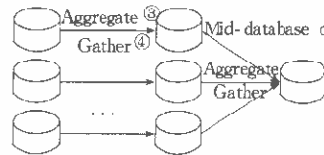


Fig. 2 Business hierarchy
图2 业务框架图

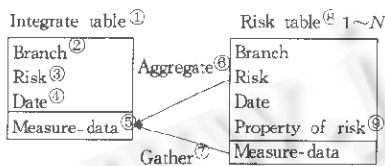
对应于图 1 的框架结构, 我们在各个支公司以及分公司本身设计了中间数据库. 这些中间数据库可以说是实际意义上的数据仓库和数据集市^[4]. 支公司的中间数据库完成对本支公司的日常营运数据的概括以及聚集, 并提供本支公司的元数据的管理. 由于受到主机存储空间的限制, 我们暂时无法在分公司的中间数据库中存储各个支公司中间数据库的所有信息. 因此, 分公司的中间数据库要对下属所有支公司的中间数据库进行更进一步的概括、聚集, 以浓缩数据量. 图 3 是该分公司中间数据库(简称中间库)的框架结构, 它为我们验证自行设计的 OLAP 系统的性能提供了一个理想的数据库环境. 我们设计的通用 OLAP 分析工具满足 E. F. Codd 所论述的 OLAP 系统的准则与特性^[5], 它可以在分公司和支公司两级中间库上顺利地进行分析操作, 并且没有操作功能上的差异. 同时, 由于有日常营运数据库的存在, 支公司级的 OLAP 分析操作还可以进行明细信息的挖掘. 图 4(其中的度量数据即表示经营情况的数据, 如保额、保费、赔款等)和图 5 分别是分公司和支公司中间库的框架结构.

Business database of branch ① Mid-database of branch ②



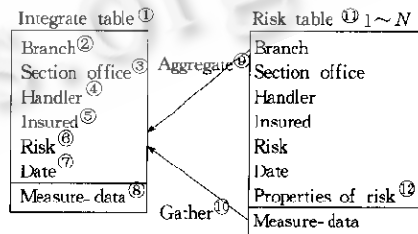
① 支公司营运库, ② 支公司中间库, ③ 聚集, ④ 概括, ⑤ 分公司中间库.

Fig. 3 Frame of mid-database
图3 中间库框架结构



① 综合表, ② 支公司, ③ 险种, ④ 时间, ⑤ 度量数据, ⑥ 聚集, ⑦ 概括, ⑧ 分险种表, ⑨ 险种特性.

Fig. 4 The structure of company's mid-database
图4 分公司中间库结构



① 综合表, ② 支公司, ③ 科室, ④ 经办人, ⑤ 被保人, ⑥ 险种, ⑦ 时间, ⑧ 度量数据, ⑨ 聚集, ⑩ 概括, ⑪ 分险种表, ⑫ 险种特性.

Fig. 5 The structure of branch's mid-database
图5 支公司中间库的框架结构

1.2 多维分析模型的设计

我们设计的两级中间数据库的引擎都是关系式数据库引擎, 而我们所追求的却是基于多维空间的分析操作. 因此, 很自然地需要一个逻辑模型, 该模型能够将二维平面的中间数据库数据映射为多维分析^[6]所需要的多维空间的数据. 另外, 我们允许用户自由地定义该逻辑模型, 也就是自由地组织多维空间的各个坐标轴以及各坐标轴上的粒度. 为了实现该目标, 我们提供了一个定义多维模型的工具以及从逻辑模型到中间数据库的平面数

据的映射.图6是OLAP系统的多维分析模型的一个示例.

Unit: branch section office agent hander^①
 Date: year quarter month week^②
 Risk: risk risky-kind^③
 Properties: properties of risk^④
 Other: measurement some independent properties^⑤
 Measurement: combined by computing data^⑥

①部门维:支公司 科室 代理处 经办人,
 ②时间维:年度 季度 月份 周,
 ③险种维:险种 险别,
 ④特性维:由险种特性组成,
 ⑤其他维:由另外一些独立的列组成(如出险原因等),
 ⑥度量维:由一些计算列组成.

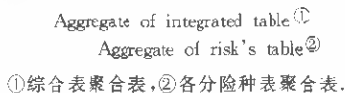
Fig. 6 An example of multi-dimension analytical model

图6 一个多维分析模型示例

通常,OLAP所要分析的数据集是相当庞大的,这必然会影响到OLAP的分析速度.而最终用户对OLAP的响应时间有一定的要求.因此,为了提高OLAP的分析速度,我们使用了聚合表技术.

一般而言,最终用户都有其常用的分析内容.也就是说,我们可以为特定的最终用户提供一些特定的聚合表,这些聚合表是根据最终用户常用的分析项,由基础表(综合表和分险种表)进一步聚合而成.聚合表可以分成多个聚合等级,每一个上级聚合表都是对其下级聚合表的进一步聚合.当最终用户进行OLAP分析时,系统将根据用户的分析项目在聚合表集中寻找包含所有分析项目的最小的聚合表,如果发现了这样的聚合表,那么当前的OLAP分析操作就在该聚合表上进行,否则,分析操作将依然在基础表上实现.

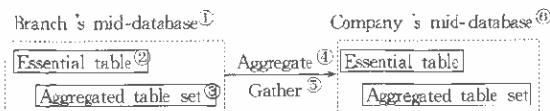
系统为中间数据库的每一张基础表预生成一系列的聚合表.图7是系统预定义的聚合表集合.这样,结合前述的中间数据库的框架结构(如图3所示),我们可以将该中间数据库结构细化为如图8所示的形式.



①综合表聚合表,②各分险种表聚合表.

Fig. 7 The system pre-defined aggregated table set

图7 系统预定义的聚合表集合



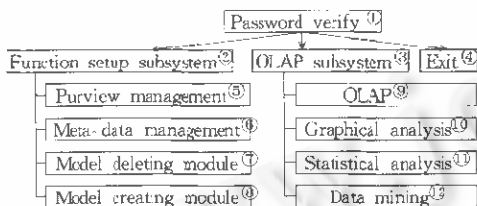
①支公司中间数据库,②基本表,③聚合表集合,④聚集,⑤概括,⑥分公司中间数据库.

Fig. 8 The detail structure of mid-database

图8 中间数据库的细化框架结构

1.3 OLAP 分析系统结构框架

一个完整的OLAP系统不仅要能够提供OLAP的各种分析手段,而且还要提供包括OLAP系统的初始准备、设置和连续操作等管理功能.因此,在设计时,我们将整个OLAP分析系统划分成相互独立的两大子系统:功能设置子系统和OLAP分析子系统.前者完成OLAP系统的初始准备、权限设置、模型创建等功能,后者完成与OLAP相关的分析操作.图9是整个OLAP系统的程序示意框架.



①口令校验,②功能设置子系统,③OLAP分析子系统,④退出,⑤权限管理模块,⑥元数据管理模块,⑦模型删除模块,⑧模型创建模块,⑨OLAP分析模块,⑩图表分析模块,⑪函数功能模块,⑫数据挖掘模块.

Fig. 9 The structure of OLAP system

图9 OLAP系统的程序示意框架

考虑到子系统间的独立性,我们将功能设置子系统和OLAP子系统分别予以封装,两者间通过模型描述文件进行通信.当用户在功能设置子系统中创建了一个用于OLAP分析的多维模型并保存后,系统将该模型的相关信息写入到模型描述文件的一个节中,节标识就是用户为该模型所取的名字.

OLAP分析子系统在进行分析操作前应当选择一个模型,以便实施分析操作.我们提供了一些相关的数据结构,它们能够将用户所选择模型的有关信息完整地映射到内存之中.本文不再赘述.

2 OLAP 系统的实现

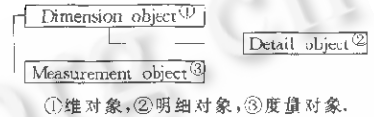
本节具体说明OLAP系统中一些模块的实现概况.

2.1 功能设置子系统

如前所述,功能设置子系统主要完成 OLAP 系统的初始准备、权限设置、模型创建等功能.它主要由权限管理模块、元数据管理模块、模型创建模块、模型删除模块等部分组成.

我们的 OLAP 系统基于中间数据库的“综合表-险种表”的结构,从而要求 OLAP 的分析操作能够在各表间自行跳跃.也就是说,一般综合性的查询分析操作可以在综合表及其聚合表上进行,当用户关注的视点转移到各险种所固有的自身特性时,OLAP 的分析操作将转移到险种表及其聚合表上进行.这种不固定于某张基本表的分析操作可以使用户更灵活地对数据仓库进行组织,另一方面也可以有效地节省数据存储空间.

为了方便使用,在创建多维空间模型时,我们将参与模型创建的中间库数据列(我们称之为对象)划分成 3 种类别:维对象、度量对象和明细对象.维对象是用户所关注的分析项目(如部门),度量对象是用户所关注的分析内容(如保费),明细对象是对维对象的细节性的描述.它们间的关系如图 10 所示.

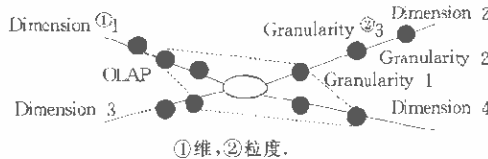


①维对象,②明细对象,③度量对象.
Fig. 10 The relationship of three objects
图10 3类对象的关系示意图

一个像保险公司这样的大型企业,使用 OLAP 分析工具进行决策支持的人员可能很多,他们的权限级别也有所不同.用户显然不能容许这种情况的存在:即某个下级部门的员工对上级部门管理的业务情况进行分析操作.因此,在设计之初,有关 OLAP 查询分析的权限问题就受到了足够的重视.

2.2 OLAP 分析子系统

OLAP 分析子系统由 OLAP 分析模块、图表分析模块、函数功能模块等构成.为了说明的方便,我们将用于多维分析的超级立方体简化成如图 11 所示的模型.我们知道,OLAP 所要实现的最重要的操作就是分析数据的钻取、旋转、切片、漂移以及一些辅助的分析函数等.从图 11 中我们可以看出,这些多维分析操作实际上可以看成是由各维上的粒度所组成的多边形的变换操作.



①维,②粒度.
Fig. 11 The simplified model of the multi-dimension analysis
图11 多维分析的简化模型

2.3 系统的实现特性

E. F. Codd 博士指出,OLAP 作为一种多维的分析工具,它的分析操作应当尽量直观、易行.具体来说,OLAP 应有如下特点:(1)多概念视图,(2)透明性,(3)易访问性,(4)一致的报表操作,(5)C/S 结构,(6)通用的维结构,(7)动态稀疏矩阵处理,(8)多用户支持,(9)非限定的定义操作,(10)直接数据操作,(11)弹性报表,(12)不受限制的维和聚集级别.另外,结合具体应用实际,我们认为 OLAP 工具还应具有以下特点:(1)明确信息的挖掘,(2)报表数据的本地化支持.我们的 OLAP 系统正是基于这些准则来实现的,因此,它具有如下一些特点:

(1) 简明的报表显示

我们采用了最新的 OCX 技术——Formula One 来组织分析结果的表达,同时,保留用户分析路径上的各个节点的分析条件,以便用户可以随时了解当前的分析环境.

(2) 便捷的数据导航

我们在分析过程中将用户所定义的多维数据模型图形化地展示给用户,并标注出用户当前所处的分析位置.根据这张导航图,用户可以轻易地在该多维数据模型中自由地遍历、分析.

(3) 通用的逻辑模型

我们在功能设置子系统中提供了供用户自由定义多维数据模型的工具,并将用户定义的模型存放在模型描述文件中.同时,我们在 OLAP 分析子系统中提供的数据结构具有很强的通用性,可以提供维的通用的逻辑操

作,也就是说,对用户而言,各个维在操作方式上不存在任何差别。

(4) 组合的分析工具

我们将 OLAP 的相关操作,诸如钻取、旋转、切片、漂移等,组合在同一个操作界面下,并且分别关联于鼠标动作,用户在打开 OLAP 分析开关之后,只须在表达分析结果的报表上使用鼠标就可以进行连续的 OLAP 分析操作,而不必进行菜单及操作界面之间的切换,这使得分析动作更为直观、自然和连续。

(5) 辅助的分析函数

用户所需要的不仅仅是查询数据的简单罗列,还需要对这些数据进行简单的统计分析,以提供更加直观的分析比较结果,在系统中,我们封装了诸如均值、极值等函数,只要用户需要,它们可以随时加入分析结果之中。

(6) 生动的图表表达

嵌入于 OLAP 分析过程中的图表分析工具能够灵活地、随时地体现分析报表上的数据内容,用户可以根据自己的爱好以及需要来组织图表的风格和内容。

(7) 可靠的报表存储

我们提供了现场保存/恢复功能,它允许用户随时中断他们的查询分析操作,并且在下一次进入系统时能够从上一次的中断点开始继续查询分析。

(8) 自由的数据旋转

OLAP 工具不应该对维的交叉分析操作作任何限制,为此,我们允许用户在多维数据模型中任意选择分析所需要的维,也就是说,图 11 中的分析多边形是可以任意变幻的,只要形成多边形顶点的粒度参与了本次分析模型的构成。

(9) 初步的数据挖掘

OLAP 的分析逻辑是演绎逻辑,它可以将统计数据一步步细化分解成用户所需要的明细信息,但它只能罗列一些已经存在的简单事实,另外再进行一些相对简单的统计分析,OLAP 分析工具无法回答诸如“为什么会这样”或“到底将来会发生什么”这样的问题,这一类问题在往需要对已有的历史数据建模分析,从而发掘出隐藏在数据背后的数据间的联系或趋势,显然,这样的分析操作需要由归纳逻辑来完成,我们把应用归纳逻辑进行分析操作的方法称为数据挖掘,利用这种方法,我们在系统中预建了一个机动车辆的赔付率预测分析模型,用于预测保户投保机动车辆可能的赔付情况。

3 小 结

我们的 OLAP 系统(约上万行代码)目前已被嵌入于自行设计的保险风险评估地理信息分析系统(GIS)以及财务分析系统之中,其灵活性、可用性已经在实践中得到了证明,整个保险综合业务与网络信息处理系统在技术上有重大的突破,并取得了巨大的经济和社会效益,分别获得 1997 年度中国科学院科技进步二等奖和 1999 年度国家科技进步三等奖。

随着应用的深入,我们发现,OLAP 系统的生命力在于其强大的分析功能以及灵活的应用环境,我们将在以下几个方向继续努力。

(1) Web 应用。随着 Internet 技术和 Intranet 技术的发展,各个大型企业都拥有了自己的企业内部网,随之诞生了新类型的应用环境,用户希望把有价值的分析结果、企业内部分析报表等放到企业内部网或 Internet 站点上,以供相关人员共同分析、查询,基于此,我们设想提供这样一种功能,它能够将用户的分析结果以报表方式存放到企业内部网的 Web Server 上,其余的用户可以在 Web Server 上对该报表进行在线分析,也可以将该分析报表下载到本地微机上进行分析,这样,企业内部就可以共享分析结果。

(2) 数据挖掘的深入。数据挖掘是知识发现技术的应用,数据挖掘实现的关键是分析模型的建立,一个合适的、识别率高的分析模型对于发掘数据后面隐藏的信息有着重要的作用,我们准备提供多种建模算法,以满足用户不同场合的分析需要。

References:

- [1] Erik, T. Dimensional Modeling, An Analytical Approach. San Mate, CA: Miller Freeman Inc., 1997. 29~35.
- [2] Inmon, W. H., Welch, J. D., Glassey, K. L. Managing the Data Warehouse. New York: John Wiley & Sons. Inc., 1997.
- [3] Sun, Yu-fang, Zuo, Chun, Wang, Mu-lan. Integrated management information system of insurance business. In: The Scientific Reports of Institute of Software of the Chinese Academy of Science (ISCAS), Vol 3. Beijing: ISCAS, 1995. 82~90 (in Chinese).
- [4] Inmon, W. H. Building the Data Warehouse. Wellesley: QED Technical Publishing Group, 1992.
- [5] Codd, E. F., Codd, S. B., Salley, C. T. Beyond Decision Support. USA: Computer World, 1993.
- [6] Inmon, W. H., Hackathorn, R. Using the Data Warehouse. New York: John Wiley & Sons. Inc., 1994.

附中文参考文献:

- [3] 孙玉芳,左春,汪木兰. 保险业务综合管理信息系统. 见:中国科学院软件研究所学术报告文集,第3卷.北京:中国科学院软件研究所,1995. 82~90.

Implementation of Online Analytical Processing System for Insurance Business

ZHAN Min, SUN Yu-fang

(Institute of Software, The Chinese Academy of Science, Beijing 100080, China)

E-mail: zhanmin@sinosoft.com.cn

<http://www.ios.ac.cn>

Received June 18, 1999; accepted September 14, 1999

Abstract: The wide use of database and the appearance of data warehouse technology provide new direction for Decision Support System(DSS). Online Analytical Processing (OLAP) is one of these new DSS technologies based on data warehouse. The mid-database design, multidimensional data analytical models and OLAP analytical system module structure for insurance business are presented in this paper, and the implementation of OLAP function setup subsystem and main analytical subsystem are outlined. It meets the requirement of the 12 rules of E. F. Codd, and makes improment based on practical application.

Key words: decision support; data warehouse; OLAP (online analytical processing); drill; rotation; slice; shift; data mining