

基于数据仓库技术和OLAP的决策支持系统

福州大学计算机科学与技术系 郭朝珍 杨俊杰

▲ 数据仓库技术和联机分析处理是近年来信息领域迅速兴起的计算机技术。本文以“菜篮子”工程决策支持系统(DSS)为背景,给出一种基于数据仓库技术和联机分析处理(OLAP)的专用DSS实现方案。应用数据仓库技术解决了长期困扰DSS的数据管理问题,OLAP则使用户能以统一界面多角度地快速观察分析数据。

一、引言

数据库仓库技术是一项新兴技术。传统数据库技术以单一的数据资源,即以数据库为中心进行从事务处理、批处理到决策分析等各种类型的数据处理工作。然而,不同类型的数据处理有不同处理特点,尽管数据库在事务处理方面的应用获得了巨大成功,但它对分析处理的支持一直不能令人满意,尤其是当以业务处理为主的联机事务处理(OLTP)应用与以分析处理为主的决策支持系统(DSS)应用共存于同一个数据库系统中时,这两种类型的处理就发生了明显冲突。应用数据仓库技术能有效地解决DSS的数据管理问题,数据仓库是进行分析决策的基础,因而还须有强有力的工具进行分析和决策。国外新兴软件技术联机分析处理(OLAP)专门设计用于支持复杂操作。

“菜篮子”工程有关信息繁多,用经济预测等定量方法决策工作量极大,以往“菜篮子”工程有关方面的领导主要是依靠大脑的分析和经验进行决策,决策质量同决策领导素质、经验有很大关系。这种决策方式效率低,时效性差。实行市场经济后,随着“菜篮子”产品种类日益增多,这种方式难以对日趋加速的市场变化快速反应,制定相应决策调控,也难以对农民进行积极引导,农民不了解经济整体情况,盲目一哄而上生产近期利润较大的产品,导致产品一段时间后严重供过于求,价格暴跌,挫伤生产积极性。而另一些产品生产规模缩小难以满足需求,价格急剧上涨,造成市场极不稳定。因此研制

“菜篮子”工程DSS显的很有必要。

充分考虑其数据来源复杂量大,市场反映要灵敏的诸因素,本文采用数据仓库技术和OLAP新技术来研究“菜篮子”工程DSS,主要实现“产、供、需”的预测与决策。利用数据仓库基本理论,设计了“菜篮子”工程DSS的数据仓库结构,提出了数据仓库数据析取与联机分析处理的有效方法,可以应分析人员要求快速、灵活地进行大数据量的复杂查询处理,使决策者能准确掌握经营状况,了解市场,判定正确方案,增加效益。

二、数据仓库技术

数据仓库是决策支持系统等分析型处理应用系统的基础,用来更好的支持分析型处理,它存储的是分析型数据。数据仓库是一个用以更好地支持企业或组织决策分析处理的、面向主题的、集成的、不可更新的、随时间不断增加的数据集合。主题是在较高层次上将企业信息系统中的数据综合、归类并进行分析利用的抽象,它对应企业中某一宏观分析领域所涉及的分析对象^[1]。

数据仓库中存放历史数据,这些数据通过数据析取程序从联机事务处理(OLTP)系统的数据库中抽取,如图1所示。数据析取可对OLTP系统数据库中数据拷贝、过滤,也可对OLTP数据库数据作变换、统计。一个典型的数据仓库系统一般由三层组成:最底下的一层是数据仓库,顶上的那层是决策支持工具,如联机分析处理(OLAP)、数据挖掘

等,中间的一层是中间件,如图2所示。

三、“菜篮子”工程DSS的分析与设计

“菜篮子”工程DSS的组成包括日常业务信息管理、决策支持、数据仓库管理。其中,日常业务信息处理包括产品信息处理、价格信息处理、供求信息处理和基地信息处理。决策支持分为预测、分析、基地计划决策和应急决策。预测又包括市场供应量预测、需求量预测和价格预测,分析又分为联机分析处理(以数据透视表的形式)和图形分析(以直方图的形式),两种分析都包括了市场供应量分析、需求量分析和价格分析。

本系统需要的数据有产品信息、生产科技信息、基地信息、供求信息、价格信息、地区编码信息、市场编码信息和流通行业信息八大类信息。系统采用数据仓库作为决策支持基础,日常业务信息处理子系统的基础是业务数据库。为了“菜篮子”工程有关信息能通过Internet从各地快速采集,日常业务信息处理子系统采用Internet浏览方式。由于系统维护的需要,本系统还建立日常业务数据库维护模块、数据析取模块和数据仓库管理系统模块。“菜篮子”工程DSS系统结构图如图3所示,下面主要介绍数据仓库的设计与数据析取和OLAP分析。

1. 数据仓库设计与数据析取分析

数据仓库中的数据面向主题进行组织^[2],主题对应企业中某一宏观分析领域所涉及的分析对象,经过分析,系统数据

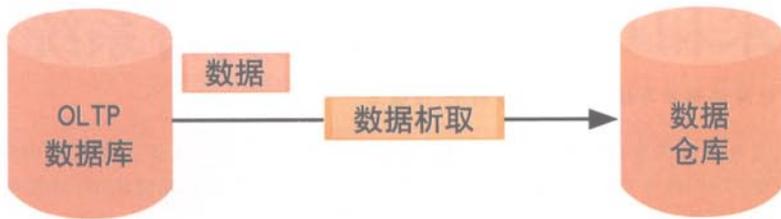


图1 数据仓库与OLTP数据库的关系

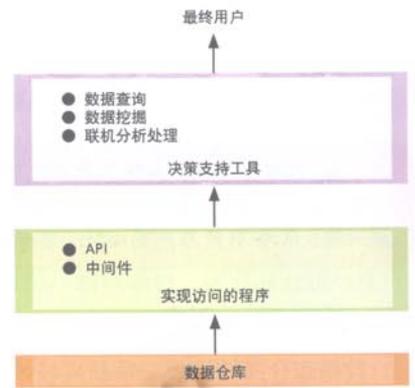


图2 数据仓库系统的一般结构

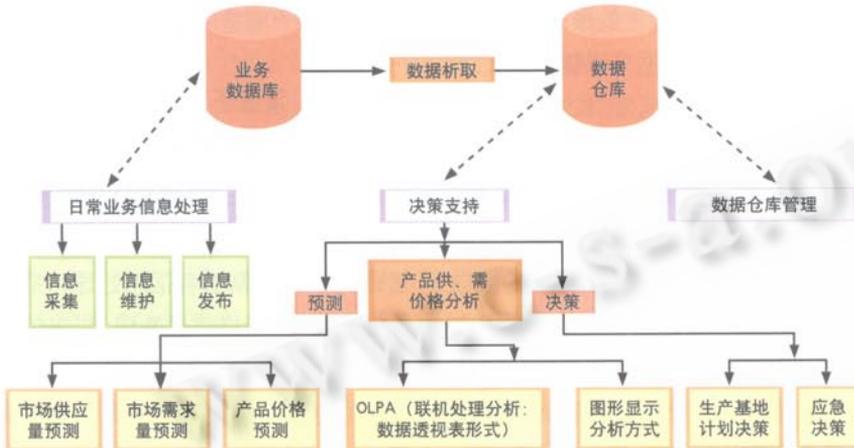


图3 “菜篮子”工程DSS系统结构图

蔬菜公司经营情况信息(地区号, 公司号, 年份, 月份, 公司自营量, 批发市场交易量, 销售外地及出口量, 公司合同约定购量, 外地调入量)

(3) 需求量

产品需求信息(地区号, 产品号, 年份, 月份, 数量)

数据析取模块包括三部分: 数据仓库数据的初始装入、产品价格的析取统计、数据仓库中主题库信息的刷新。产品价格的析取统计又包括产品价格旬析取统计、月析取统计和年析取统计。

2. OLAP 的需求分析

OLAP是对特定问题的联机数据访问分析。通过对信息(这些信息已经从原始数据进行了转换, 以反映用户所能理解的企业的真实的“维”[3])的很多可能的观察形式进行快速、稳定一致和交互性的存取, 允许管理决策人员对数据深入观察。在分析型处理中需要对信息统计综合, 所以联机分析处理除了提供直观、统一的数据访问界面外, 内部还有数据综合引擎。

下面对OLAP应用需求分析。“菜篮子”工程DSS的OLAP的数据流图见图4所示。分析联机分析处理对象的“维”, 即用户希望从哪些角度观察数据。联机处理的数据流图已说明系统联机分析处理对象有三类: 价格、市场供应量、需求量。其中价格包括平均价、批发价、零

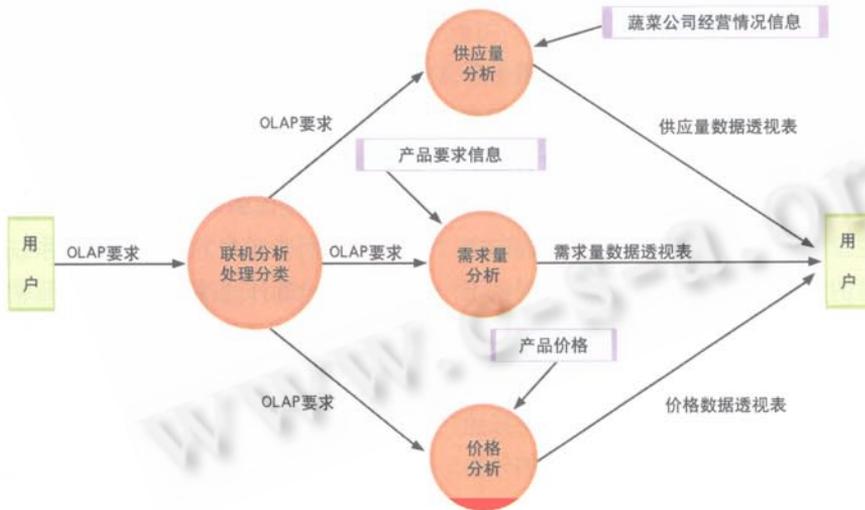


图4 联机分析处理的数据流图

仓库的主题域主要有价格、市场供应量、需求量三个。根据这三个主题域设计的数据仓库结构具体如下:

(1) 价格

产品价格旬统计信息(地区号, 产品号, 年份, 月份, 旬, 平均价, 批发价, 零售价)

产品价格月统计信息(地区号, 产品号, 年份, 月份, 平均价, 批发价, 零售价)

产品价格年统计信息(地区号, 产品号, 年份, 平均价, 批发价, 零售价)

(2) 市场供应量

产品供给信息(地区号, 产品号, 年份, 月份, 数量)

售价,市场供应量包括公司自营量、批发市场交易量、外销及出口量、合同订购量和外地调入量。经过调查、分析,对以上三类指标的维设计如下:价格设有地区、产品、年份、月份、旬五个维,市场供应量设有地区、年份、月份四个维,需求量设有地区、公司、年份、月份、旬四个维。

四、“菜篮子”工程DSS的

数据析取与OLAP的实现

由于采用了数据仓库技术,且本系统是C/S模式(其中日常业务信息处理是B/S模式),所以数据库管理系统是由Microsoft SQL Server 7.0,前端采用Delphi X.0开发。Microsoft SQL Server 7.0是Microsoft SQL Server的第一个提供数据仓库解决方案的版本,是通过其中的两个服务来支持数据仓库技术:数据转换服务(Data Transformation Service,简称DTS)和联机分析处理服务(OLAP Service)。

1. 数据析取的实现

“菜篮子”工程DSS采用DTS实现数据析取,完成数据仓库建立、数据的初始装和追加。通过DTS可在多个使用OLE DB为基础的体系结构的异质数据源间导入和导出数据,可以变换联机事务处理系统中的数据以建立数据仓库。

DTS主要功能有导入和导出数据、变换数据和传送数据库对象。DTS允许在一个过程中导入、导出和变换数据。这个过程的定义可以保存在一个叫作包(Package)的东西中。DTS包含三种类型对象:连接(Connection)对象、任务(Task)对象和步骤(Step)对象。连接对象定义每一个是数据的源或目标的OLE DB数据源;任务对象定义工作项目,例如执行SQL语句、拷贝一个表的内容或执行一段ActiveX脚本;步骤对象定义任务对象执行的次序。

定义包方式有三种:使用DTS设计器(DTS Designer)、DTS导入和导出向导、

DTS编程接口。系统采用DTS设计器定义包,包可以三种方式保存: DTS基于COM的文件(.dts文件)保存,保存在MS SQL Server的msdb数据库中,保存在Microsoft Repository中。定义好的包可以通过MS SQL Server代理(SQL server Agent)服务让SQL Server自动、定期执行。

在数据析取中,系统价格预测方法采用同期数据比较法,将价格预测分为旬价格预测、月价格预测和年价格预测,在数据仓库中设计了价格旬统计信息、月统计信息和年统计信息以支持上述三个预测。系统设计了五个DTS包完成数据析取。“初始装入”包用于数据仓库的建立和数据初始装入,“旬析取”、“月析取”、“年析取”三个包分别用于价格的旬、月、年析取统计,实现数据仓库的数据追加。这三个包利用Microsoft SQL Server 7.0中的SQL Server Agent服务由Microsoft SQL Server自动定期执行,“主题库信息”包实现数据仓库主题库信息的刷新、追加。

2. 联机分析处理的实现

“菜篮子”工程DSS使用Delphi中用于联机分析处理的Decision Cube类组件编制其联机分析处理应用的。Decision Cube类组件包括有Decision Cube组件、Decision Query组件、Decision Source组件、Decision Pivot组件、Decision Grid组件和Decision Graph组件。

Decision Cube是一个多维数据库。Decision Cube组件维护一个交叉数据分析,并提供给Decision Source组件。Decision Cube分析从数据集中取出数据,把数据变成一个多维表的结构,该结构中的每一维与数据表中需要统计的某一个域对应。Decision Query类似于Query组件,用于对数据库进行查询。Decision Source类似于Data Source组件,为数据透视表、栅格和图表提供数据源。Decision Pivot是用于对栅格的图表重新

组织的工具。Decision Grid以栅格的形式显示多维的数据。Decision Graph以图表的形式显示数据,可以按不同的字段重新组织图表。

本系统的分析处理包含了价格分析、市场供应量分析和需求量分析。所以,本系统联机分析处理中对价格(包括平均价、批发价、零售价)设计了地区、产品、年份、月份、旬五个维。市场供应量指标、需求量指标类似。

五、结束语

本文给出了一种基于数据仓库技术和OLAP的实现“菜篮子”工程DSS的解决方案,将数据仓库技术和联机分析处理(OLAP)引入系统,解决了传统系统忽视历史数据和DSS数据管理难等问题。有效地实现了数据仓库的建立和数据析取,使对系统的市场供需预测、产品价格预测,生产基地决策提供了良好的基础。并使得用户能以一致的界面向快速地从多个角度观察,深入分析数据。

文中介绍的分析设计和实现方法具有一定的通用性,可用于其他专用DSS的实现。

参考文献

- [1] Joyce Bischoff\, Ted Alexander, Data warehousing, Beijing: Electronic Industrial Press, 1998) (Joyce Bischoff\, Ted Alexander, 数据仓库技术, 北京: 电子工业出版社, 1998)
- [2] Inmon W H. Building the data warehouse. A Wiley\|QED Publication, John Wiley & Sons, inc., 1992
- [3] Wang Shan, Data warehousing and on\|line analytic processing, Beijing: Science Press, 1998