基于数据仓库的房地产数据分析系统

陈慧萍[1]， 唐志贤[1]， 陈岚峰[1,2]
(1. 河海大学 计算机信息工程学院，江苏 常州 213022；2. 无锡房产管理局，江苏 无锡 214001)

摘 要：针对房地产信息系统积累的大量房产数据，研究并实现基于数据仓库及 OLAP 技术的房地产智能数据分析系统。首先建立了分离的数据仓库，并对数据仓库进行多维建模和多维分析；然后对多维数据模型进行了切片、切块、旋转、上钻和下钻等多维分析，从多角度进行房产数据分析，并计算房产指数；最后利用前端开发工具开发了可视化的多维分析和数据展示平台。实践表明，基于数据仓库的房产数据分析系统可以为房地产管理层和决策层提供高效的决策支持。

关键词：房地产信息系统；数据仓库(DW)；OLAP 技术；多维数据模型；数据分析
中图法分类号：TP311 **文献标识码**：A **文章编号**：1000-7024 (2008) 17-4589-04

Real estate data analysis system based on data warehouse

CHEN Hui-ping[1], TANG Zhi-xian[1], CHEN Lan-feng[1,2]
(1. College of Computer Information and Engineering, Hohai University, Changzhou 213022, China;
2. Wuxi Real Estate Bureau, Wuxi 214001, China)

Abstract: Based on the large amount data accumulated in the information system, the real estate data analysis system based on data warehouse and OLAP technology is studied and implemented. First, the independent data warehouse of the real estate is established and the multi-dimensional model is set up to make multi-analysis. And then the multi-analysis analyses such as slice, dice, rotate, roll-up, and drill-down are made in order to analyze the real estate data from multi-dimension. The real estate indices are computed. On this basis, the front-end development tool is used to develop the display platform for the data, which can be used for integration of visualization analysis. The experiments show that the real estate data analysis system can help the administrators and the decision makers of real estate area to make the efficient decision supports.

Key words: real estate information system; data warehouse (DW); on-line analytical processing technology; multi-dimensional data sets; data analysis

0 引 言

为提高迅速发展的房地产业的信息化程度，近几年开发了不少房地产领域的信息系统[1-2]，其中积累了大量的数据。但目前房地产领域信息系统是以 OLTP(on-line transaction processing，在线事务处理)为主，OLTP 注重对数据库联机的日常操作如数据插入、删除、修改、查询和统计等功能，不能很好为较高层次的管理者和决策者提供决策支持。

从实际需求而言，房地产管理部门和决策者急需对房地产开发和销售的历史数据进行智能分析，从而制定相应的政策，规范市场行为和市场竞争，从宏观上给开发商予以指导。而数据仓库(data warehouse，DW)却能提供较好的决策支持，数据仓库面向 OLAP(on-line analytical processing，在线分析处理)，OLAP 注重对数据更高层次的分析，通常是对海量的历史数据进行查询和分析。这就促使将数据仓库技术应用于房地产行业。

在这样的背景下，本文针对无锡市房地产信息系统积累的大量房产数据，建立数据仓库并进行多维建模，将 OLAP 用于房产数据的多维分析中，并利用前端工具对房地产多维模型进行数据可视化展示。

1 数据仓库和 OLAP 技术

W.H.Inmon 在 1992 把数据仓库定义为："一个面向主题、集成的、随时间变化的、非易失性数据的集合，用于支持管理层的决策过程"[3]。数据仓库技术是在传统的数据库技术基础上发展而来的，其主要目的是为决策提供支持，为联机事务分析、数据挖掘(data mining，DM)等深层次的数据分析提供平台。目前，数据仓库技术已成功用于电信、银行、税收、零售业中。

1.1 数据仓库的组成

建立数据仓库的主要过程包括数据导入、数据存储和管理及数据分析与展现，为完成以上过程，数据仓库系统采用如图 1 所示的结构[4-5]。其中数据抽取、清洗、转换、装载和维护

收稿日期：2007-09-05 E-mail：chenhp@webmail.hhuc.edu.cn
作者简介：陈慧萍 (1964－)，女，江苏无锡人，硕士，副教授，研究方向为数据库和数据挖掘；唐志贤 (1983－)，男，硕士研究生，研究方向为空间数据库；陈岚峰 (1980－)，男，硕士研究生，助理工程师，研究方向为数据仓库技术。

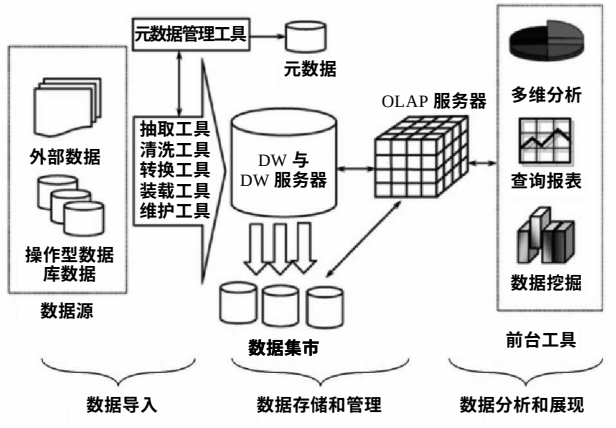

图 1　数据仓库系统结构

工具负责将所需数据从数据源导入数据仓库 DW 中；DW 服务器负责数据的存储管理和数据存取，并给 OLAP 服务器和前台工具提供存取接口；OLAP 服务器则透明地为前台工具和用户提供多维数据视图；前台工具包括查询报表工具、多维分析工具、数据挖掘工具负责进行数据分析，并以直观的方式向决策层展现数据，提供决策支持。

1.2　OLAP 多维分析技术

OLAP 支持决策人员从不同的角度、快速灵活地对数据仓库中的数据进行复杂查询和多维分析。

OLAP 技术的基础是多维数据模型。所谓多维模型就是数据分析时用户的数据视图，是面向分析的数据模型，用于为分析人员提供多种观察的视角和面向分析的操作[6]。其中分析的核心数据称为多维模型的度量值，这些数据一般是销售量、成本和费用等，例如楼盘数据中的建筑面积、绿化面积等。观察的视角即为多维模型的维度，例如在研究楼盘时，从开发商角度来分析和研究度量值建筑面积，开发商就是楼盘的一个维度。在同一个维度上，可以存在多个不同的细节，这些细节就是维的层次，它是对维的进一步细化。例如楼盘的销售时间就有年、季度、月和日这 4 个层次。多维数据模型是包含维度和度量值的多维结构，也被称为数据立方体或超立方体。图 2 就是一个三维的楼盘多维模型示意图，其中度量值为楼盘建筑面积，维度为楼盘所在的行政区域、开发商和楼盘类型。实际分析中，维度数量往往高于 3 维。

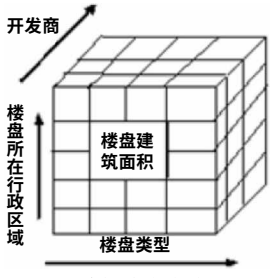

图 2　三维楼盘数据模型

OLAP 的多维分析，是指采用切片(slice)、切块(dice)、旋转(pivot)、向上综合(roll-up)和向下钻取(drill-down)等基本操作[7]，对多维数据模型进行深入的研究，从而使客户达到从多个角度，多个细节分析数据的目的。

2　房地产数据仓库的设计

2.1　房产数据仓库需求分析

根据对现有房产信息系统和无锡房产局管理层的调研，了解到决策层希望从楼盘数据、销售数据和汇总数据等多个主题来分析数据，因此数据仓库的需求是多方面的，包括功能需求、数据需求、数据安全性和系统性能方面，以下重点分析数据需求和功能需求。

2.1.1　数据需求

数据仓库的数据来自于现有的房产信息系统如商品房备案系统、销售管理系统等，所以首先要将与决策分析相关的数据从原有系统中导出并导入到一个分离的数据仓库中，供决策分析使用。数据仓库是面向主题的，所以数据的抽取也必须以分析主题为中心。根据楼盘数据、销售数据和汇总数据等 3 个不同的主题，必须建立不同主题的多个多维数据模型，针对每个多维模型拟从原数据源中抽取如下数据：

(1)度量值：楼盘数据主题下的度量包括楼盘的占地面积、绿化面积、建筑面积、可售面积和楼盘建筑成本；销售数据主题下的度量包括楼盘的销售面积、销售利润和销售金额等；汇总数据主题下的度量包括楼盘销售面积、楼盘销售金额和楼盘销售利润等。

(2) 维度：多维模型的维度即为观察数据的角度。3 个分析主题尽管有所不同，但有时会从共同的角度去分析不同的主题，所以维度可以共享。3 个分析主题下的维度包括楼盘所在行政区域、楼盘所在地理区域、开发商企业性质、开发商企业注册登记的行政区域、楼盘的类型及楼盘的户型等。

2.1.2　功能需求

针对上述数据，建立房产数据仓库后应该实现如下功能：①按楼盘的开发商企业性质分析楼盘数据、销售数据和汇总数据；②按楼盘的开发商企业注册登记的行政区域分析楼盘数据、销售数据和汇总数据；③按楼盘的所在行政区域分析楼盘数据、销售数据和汇总数据；④按楼盘的所在地理区域分析楼盘数据、销售数据和汇总数据；⑤按楼盘的类型和户型分析楼盘数据、销售数据和汇总数据；⑥在上述的多个维度上分析楼盘数据、销售数据和汇总数据；⑦对楼盘数据、销售数据和汇总数据进行旋转、切片、切块、向上综合和向下钻取等多维分析，以获得多角度、多粒度历史数据；⑧进行多种房产指数计算；⑨实现分析数据的可视化展示平台。

2.2　房产数据仓库设计与实现

由于大部分现有房地产信息系统的底层数据库为 SQL Server，因此考虑到兼容性和数据导入的便捷性，在实现数据仓库时采用 Microsoft SQL Server 2000 Analysis Services[8]平台。先根据数据需求将现有信息系统中的数据抽取出来，进行简单的预处理如数据缺失值、数据不一致的处理等操作后进行数据集成，然后设计多维模型，再在 Analysis Services 平台上建立多维模型，然后进行存储设计与实现。

2.2.1　房产数据仓库的多维模型的设计与建立

数据仓库的设计常常采用的是星型模型和雪花模型。星型模型通常采用一个包含主题的事实表和多个包含事实的维度表来支持各种决策查询，但星型模型不能很好提供对属性

层次的支持。雪花模型是在星型模型的基础上改进而来的，可以提供对属性层次的支持[7]。在雪花模型中，维度表除了具有星型模型中维度表的功能外，还与详细类别表相连，详细类别表可以在相关维上进行详细分析描述，以缩小事实表、提高查询效率的目的。由于本文研究的房地产信息较为复杂，所以均采用雪花模型。

在房产数据仓库中共建立了3个多维模型即楼盘数据模型、销售数据模型、汇总数据模型，均为雪花模型。如楼盘数据模型的需花模型图如图3所示，图4是在平台上实现后的示意图。

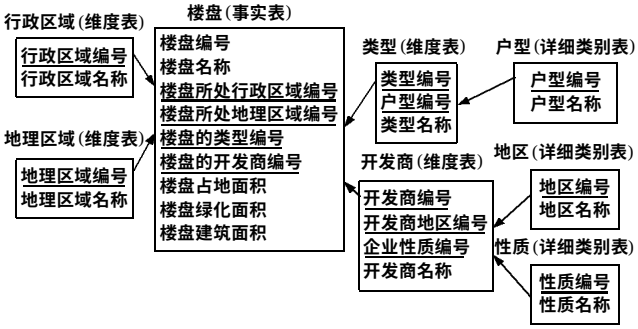

图3　楼盘数据雪花模型

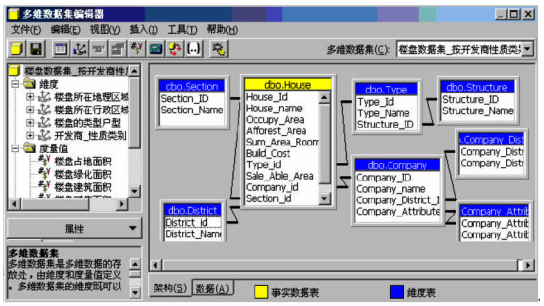

图4　楼盘数据雪花模型实现

2.2.2　房地产数据仓库多维模型的存储设计

在对多维模型进行处理时，需要解决是采用多维数据库系统还是采用关系数据库系统存储数据的问题。如果采用多维数据库系统存储、显示数据，那么这种OLAP系统就是基于多维的OLAP，即MOLAP(multidimensional OLAP)。如果采用关系数据库系统存储、显示数据，那么这种OLAP系统就是基于关系的OLAP，即ROLAP(relational OLAP)。

由于MOLAP结构能迅速地响应决策分析人员的分析请求并快速地将分析结果返回给用户，而房产数据仓库中数据量很大，所以为提高响应速度，本文研究的多维数据模型均采用MOLAP存储。

3　房产数据智能分析

建立房产数据仓库的最终目的是为了对房产历史数据进行多方面的智能分析和指数计算，本系统中分别在Analysis Services、Microsoft Excel和MDX这3个平台上对房地产数据仓库中的多维数据进行多维分析和计算，限于篇幅，以下只介绍部分功能。

3.1　楼盘数据模型的浏览和分析

通过Analysis Services可以对楼盘数据集进行上钻、下钻、切片和切块等多维分析。Analysis Services提供了多维数据浏览器，利用它可以对多维数据进行浏览。

图5展示了数据分析的常用操作。在多维数组的某一维上选定一个维成员的动作称为切片，即在多维数组中选一维，并取其一维成员，所得的多维数组的子集称为在该维上的一个切片。对于楼盘数据来说，只研究无锡市市中心的楼盘数据这就是一个切片。切块是指在多维数组的某一维上，选取某一区间的维成员的动作。切块是对多维数组在某一维上取值区间的限制。对于楼盘数据，若研究类型户型中所有的别墅数据就是一个切块。钻取是改变维的层次，变换分析的粒度。它包括上钻或上卷(roll up)和下钻(drill down)。上卷是在某一维上将低层次的细节数据概括到高层次的汇总数据，或者减少维数；而下钻则相反，它从汇总数据深入到细节数据进行观察或增加新维。要了解所有开发商开发的房地产数据就可以采用上钻操作，要了解无锡市惠山区的房地产数据就可以采用下钻操作，在这里只下钻了一个层次，如需要更详细的数据还可以进一步下钻。

图5　楼盘数据模型的多维分析

3.2　房产数据展示前台平台

Microsoft的数据透视表服务作为客户端工具，起着与OLAP服务器通讯和为客户程序提供访问OLAP数据接口的作用。透视表服务应用中，最典型的应用是Microsoft Excel。首先在Excel中启动数据透视表服务，然后将数据透视表的数据源设为OLAP中的具体多维模型。

使用Microsoft Excel应用程序对OLAP中多维数据集进行旋转、上钻、下钻、切片和切块等多维分析，同时Microsoft Excel应用程序还提供了数据透视图的功能，可以将表转化成图的形式更直观地表示出来。Excel提供的图种类很多，有折线图、柱形图、条形图、饼图等。

图6给出了楼盘数据模型的一个数据透视表，在透视表

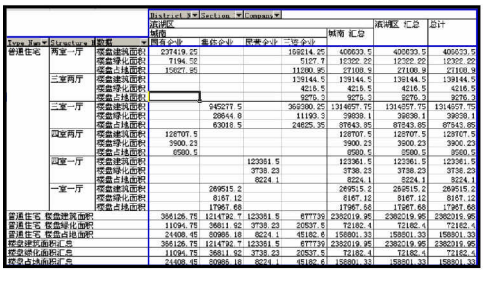

图6　楼盘数据集的数据透视表实例

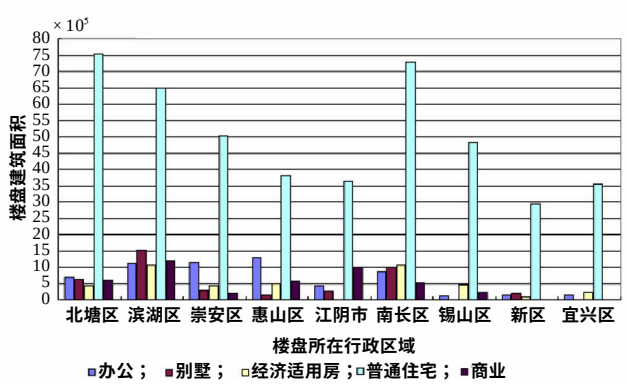

图 7 楼盘数据集可视化展现实例

上可以方便地进行各种操作。图 7 给出了无锡市七区两市的各种类型楼盘的建筑面积数据对应的柱形图，这完全是在数据透视表基础上建立的，还可以根据分析需要建立其它的图表和进行房产指数计算。

4 结束语

本文基于数据仓库技术建立了房产数据分析系统，从传统信息系统中分离出历史数据导入到数据仓库中，然后在数据仓库中根据数据分析需求建立不同的多维模型，在多维模型的基础对数据进行智能分析，大大提高了历史数据的使用价值，为决策者提供了决策支持，也为进行更深层次的数据分析如数据挖掘建立了良好的基础。

参考文献:

[1] 曹新建,张鹏,王小东,等.房地产信息管理系统开发研究[J].计算机工程与设计,2004,25(9):1520-1522.

[2] 马智亮,傅思源,李恒,等.房地产开发项目市场分析信息管理系统[J].清华大学学报, 2007,47(9):1409-1413.

[3] Inmon W H.数据仓库[M].王志海,译.北京:机械工业出版社, 2000.

[4] 王锁柱,孙玉芳.基于数据仓库的 EIS 数据管理体系结构[J].计算机工程与应用,2003,39(19):17-20.

[5] 王珊,萨师煊.数据库系统概论[M]. 4 版.北京:高等教育出版社, 2006:408-417.

[6] Jiawei Han, Micheline Kamber.数据挖掘:概念与技术[M].北京:机械工业出版社,2001:223-260.

[7] 陈京民.数据仓库原理、设计与应用[M].北京:中国水利水电出版社,2004.

[8] 刘爽英, 张静. 基于 SQL Server 2000 的数据仓库和数据挖掘[J].中北大学学报, 2004,25(5):322-324.

(上接第 4570 页)

的存储和分析，并设置了相应的服务，其它联网的客户机可以浏览监控各水质监测站的监测数据和工作状态[9]。中心服务器的软件系统可以采用组态网或者其它组态软件来开发。

3 结束语

近年来，水污染监测不仅出现了一些新的方法，同时也出现了一些新材料、新的监测物。尽管新监测手段不断出现，除 MEL 微生物检测仪外，其它利用水生物监测水质的分析仪投入使用的还很少[1-9]。本系统结构包含控制中心站系统、监测基站系统两大部分组成。中心站系统由前端计算机、应用软件组成并与 N 台计算机组成局域网。前端计算机通过可接收来自基站系统的数据和向基站系统发出来自中心站的指令。基站系统与中心站通信，并接受来自异构水质传感器的数据信号。在本系统中，一个中心站系统可监控若干个基站系统，从而组成一个庞大的水质无线计算机自动监控网络。

根据设计的 zigbee 无线水质监测网络平台，对各种偏远环境下的水质参数进行连续采集，并在监控中心服务器上实时显示。Zigbee 网络是低功耗、低成本、高可靠性的无线传感器网络，其在无线远程环境检测中有着广阔的应用前景[7-9]。本文在研究 Zigbee 无线传感器网络的基础上，提出了基于 Zigbee 协议的无线传感器水质监测网络系统的构成方案，并在由此方案构建的无线网络平台上进行了水质参数检测收集和分析测试。实验验证了通过该系统进行远程无线水质监测的可行性。

参考文献:

[1] 叶湘滨,陈利虎,胡罡.传感器网络在环境监测中的应用[J].计算机测量与控制,2004,12(11):1033-1035.

[2] 任丰原,黄海宁,林闯.无线传感器网络[J].软件学报,2003,14(7):107-111.

[3] ZHENG J, LEE M J. A comprehensive performancestudy of IEEE802.15.4[Z].2004.

[4] Chipcon, Acket. Sniffer for IEEE802.15.4 and Zigbe[S]. Oslo, Norway:User Manual,2004.

[5] Akyildiz I, Su W, Sankarasubramaniam Y, et al. Wireless sensor networks:A survey[J].Computer Networks,2002,38(4):393-422.

[6] Solis I,Obraczka K.The impact of timing in data aggregation for sensor networks[C].Proceedings of the IEEE International Conference on Communications:2004:640-3645.

[7] 郭世富,马树元,吴平东,等.基于 Zigbee 技术的无线传感器网络在远程家庭监护系统中的应用研究[J].电子技术应用,2006(6):28-30.

[8] 任秀丽, 于海斌. 基于 ZigBee 技术的无线传感网的安全分析[J].计算机科学,2006(10):111-113.

[9] 张宏锋, 李文锋. 基于 ZigBee 技术的无线传感器网络的研究[J].武汉理工大学学报(信息与管理工程版),2006(8):12-15.