

基于 MS Analysis Services 的 OLAP 分析系统的设计与实现

邱怀姍 朱群雄

(北京化工大学信息科学与技术学院,北京 100029)

摘要:文中对 MS Analysis Services 中的多维数据建立和访问技术进行了讨论,提出了一种 OLAP 分析系统的设计与实现方法,在此基础上给出了一个从 OLAP 数据源设计到应用程序设计的实例。

关键词:多维分析;数据立方体;多维表达式;数据转换服务

中图分类号: TP311

引言

随着数据库技术的发展,数据库的规模越来越大,信息量也越来越多,人们希望从已有的数据中提炼出有用的信息服务于决策者。为此,数据仓库及 OLAP 技术成为数据库技术领域的一个新的研究热点。从逻辑上讲,数据仓库就是一个多维数据库,OLAP 就是以多维分析为基础,通过在管理与决策过程中对数据仓库的访问,实现对数据多层次,多角度的分析处理,并以直观易懂的形式将结果提交给决策者。

MS Analysis Services 是微软提供的一种数据仓库的中间件,作为 OLAP 系统的开发平台它有如下优点:(1)数据仓库技术集成在 SQL Server 2000 中,从业务数据库中获得数据更容易;(2)内置于 Microsoft 管理控制台(MMC),界面统一,易于使用;(3)SQL Server 集成了众多的支持工具,可能通过多种方式访问挖掘源数据;(4)支持多种数据挖掘模型,挖掘结果显示直观。所以,采用 MS Analysis Services 建立 OLAP 应用是一种流行的开发方案。在本文中,提供并实现了一个利用 MS Analysis Services 的 OLAP 引擎并结合 MDX (multi-dimensional expression),ADO MD,VB 建立 OLAP 分析系统的解决方案。

1 OLAP 分析系统的设计

经过多年的办公自动化建设,各个高校的许多部门都有了自己的一套业务管理系统(OLTP 系统)。随着时间的积累,部门数据库中的数据越来越多,但 OLTP 系统主要面向业务操作员,根据业务需要及时准确地对数据库数据进行修改以处理业务过程。随着“教育产业化”的发展趋势,学校的领导者迫切地需要充分挖掘已有数据,从中得出一些对学校发展有益的结论。对学校来说,最大工作重点是如何提高教师的主动性,如何提高教学质量。为了达到这个目的,首先要对教师的情况有一个充分的认识,通过多角度的比较对教师进行一个定位,从而对制定有益的政策倾斜给予科学的指导,给教师充分的发展空间,从而最大限度地挖掘出教师的潜力。本系统中建立了四个主题:教师基本数据集、项目集、论文集和教学任务集,目的是从四个方面考虑和评价教师的综合素质。OLAP 是一种以多维数据为基础,需要用户积极参与分析过程,动态地提出分析要求,对数据进行由浅及深的验证型分析工具,它以查询、分析为特征。本文所实现的 OLAP 分析系统是某高校科技信息管理平台的一部分,其中数据来源于教师的基本信息以及历年来的项目数据、论文数据以及教学数据。

1.1 系统的设计方案

系统利用三层客户机/服务器结构实现。系统从高校多个部门的多个 OLTP 数据库中采集数据到科技信息数据仓库。OLAP 服务器从数据仓库中抽取数据,提供给客户端使用,相关人员利用前端

收稿日期: 2003-02-20

第一作者: 女,1979 年生,硕士生

E-mail: shanqh@263.sina.com

PC 机上的客户软件,通过局域网访问数据仓库上的数据。系统利用 SQL Server 数据库系统建立数据仓库,用 Windows 2000 Server 组建和管理局域网,并以它作为 OLAP 服务器的工作平台。采用 VB 开发应用程序,采用 TCP/IP 与原有的部门管理系统连网。利用 ADOMD 进行数据连接,并利用 OLAP 服务器所提供的功能,尽可能使数据访问本地化,以提高响应速度^[1]。

1.2 功能设计

系统具有以下功能:根据前端应用程序,创建、生成不同的数据立方体,能够手动和自动刷新多维数据集。提供 OLAP 的各种操作,如上卷、下钻、切片、旋转等。提供灵活直观的图形或图表显示,整个系统界面友好,操作简单。

1.3 结构设计

本系统在逻辑结构上采用了先进的三层 C/S 模式。即数据仓库层、OLAP 引擎层和前端应用程序层。如图 1 所示。三层模式把工程项目(Project)划分为三部分服务:用户、事务、数据。用户服务是与用户应用程序通信有关的一切服务,通常是用户接口。事务服务包含事务逻辑,并通常组成事务对象。将三者分开可以增加系统的可重用性,事务服务组件可以重复利用,也就是说 OLAP 分析系统既可以用于运行在客户机上的 VB 应用程序中,也可以用于一个电子商店站点上,还可以用在远程客户机的 DHTML 应用程序中。

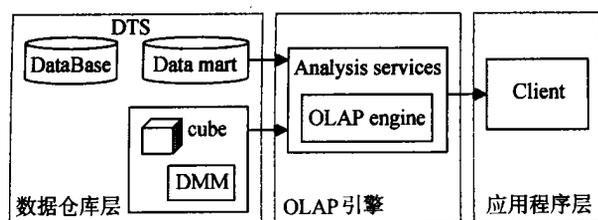


图 1 系统逻辑结构

Fig. 1 Logical structure of system

2 OLAP 分析系统的实现

一个完整的 OLAP 分析系统由以下几部分组成:1)提取、清洗数据组成数据仓库。2)根据需求利用 OLAP 引擎建立相应的 OLAP 立方体(也叫主题)。3)使用前端开发工具,通过 ADOMD 组件访问立方体包含的对象和集合^[2-3]。

2.1 数据准备

数据准备是成功进行 OLAP 分析的前提之一,

开发人员要从 OLTP 数据库中将数据提取到 OLAP 数据库中,在提取的同时进行了“校验”、“抽取”、“净化”以及“转换”操作。由于本系统所用的数据来自于不同的事务数据库,所以首先使用 DTS(微软数据转换服务)将所需的数据提取到一个库中,然后修改数据库结构,消除冗余,使之满足第三范式,最后进行数据的转换,在 OLAP 数据库产生两类表,一类是事实表,用来存储事实的度量值和各维的码值;另一类是维表,对于每一维来说,至少有一个表用来保存该维的元数据,即维的描述信息。事实表通过每一维的码值同维表联系在一起的,最终产生一个结构良好的 OLAP 数据库。

2.2 建立 OLAP 立方体

数据仓库和 OLAP 工具是基于多维数据模型的。多维数据模型将数据看作数据立方体(data cube)形式,该模型可以星型模式、雪花模式或父子模式存在。本系统采用的是雪花模式,即维表中生出分支,维表还带有附属的维表。

多维数据模型围绕中心主题组织。该主题用事实表表示,事实是数值度量的,把它们看作数量,是因为我们想根据它们分析维之间的关系。事实表包括事实的名称或度量,以及每个相关维表的外键。本系统根据需要包含四个主题:教师数据集,项目集,论文集和教学任务集。简洁起见,选择项目集作为例子,多维数据结构如图 2 所示。

项目主题中的度量值为合同总金额和项目数,建立这个主题,就可以从不同的维去查看事实表,比如:可以统计各个院系的合同数和合同金额,在这个基础还可加上时间范围以及其它的维。

从 OLAP 数据库到 OLAP 立方体这一步系统会自动进行,定义好多维数据结构后不需要用户过多的参与,只要 OLAP 数据库中的数据是经过校验、完整一致就能通过 OLAP 引擎生成正确的多维立方体。

2.3 OLAP 分析系统的关键技术

2.3.1 ADO MD 组件 ADO MD 是一套通过 OLE DB Provider 访问数据源的 COM 组件。ADO MD 是对 ADO 的扩展,它包含了访问 OLAP 立方体所需要的各种对象和集合。系统中所有对数据的操作都是通过 Visual Basic 6.0 中最新的 ADO 对象进行,其中 OLAP 立方体的访问通过 ADO MD 对象进行。ADO MD 对象模型如图 3 所示。

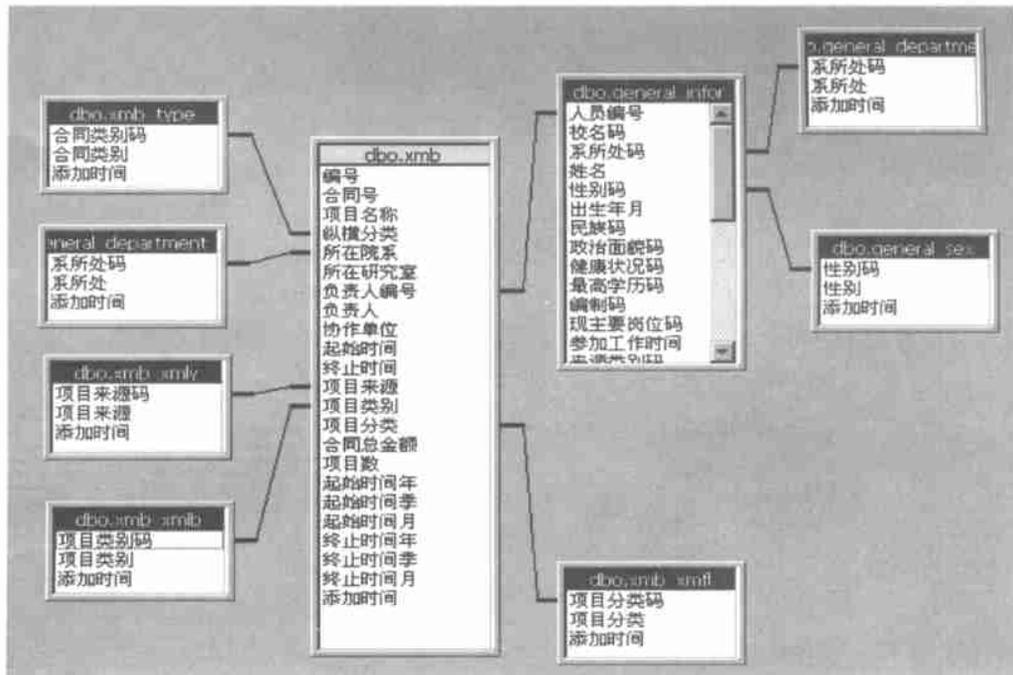


图 2 项目主题的多维数据结构

Fig.2 Multi dimension data structure of project item

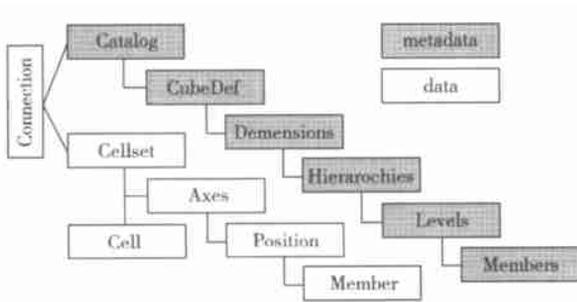


图 3 ADO MD 对象模型

Fig.3 Object model of ADO MD

其中 Connection 维护对象与数据源的连接, Catalog 维护立方体的结构和模式, CellSet 维护 MDX 语句的查询结果。实际编程时,应先创建一个 ADOMD. Connection 对象,并使用多维数据提供者 MSOLAP 打开与 OLAP 数据源的连接,然后通过 Catalog 和 CellSet 对象来查询立方体的模式和数据。

2.3.2 多维表达式 MDX MDX 是 OLAP 多维立方体查询语言,在功能上类似于关系数据库查询语言 SQL,是 OLAP 服务器与外界交互的专用语言,主要语法如下

```
SELECT sets ON axes (n-axes)
FROM a source cube
WHERE the data is sliced by some coordinates
```

SELECT 子句用来选择维成员和度量,由于检索的是多维数据集,ON 子句用来确定所选或度量显示于哪个轴,FROM 子句用来指出数据来源的立方体,这样,MDX 就可以访问任何结构良好的数据立方体了。

MDX 还可以用“DrillDown Member”等关键字对检索的数据进行上卷、下钻等操作,并且还可以通过数学函数和自定义运算对所得数据进行计算分析。可以很轻松地计算出和、百分率、线性回归、集合排序等等。

2.4 OLAP 分析服务与前端应用程序的结合

根据图 2 的雪花型模式建立好 OLAP 数据立方体,利用以下代码可以将建立好的 OLAP 数据立方体嵌入到前端 VB 应用程序中^[4]。

```
Dim catOLAP As New adomd. Catalog
Dim cbfOLAP As adomd. CubeDef
连接 OLAP SERVER
With cnnOLAP
.Provider = "MSOLAP"
.ConnectionString = "Integrated Security =
SSPI; Persist Security Info = False; Data Source =
QHS; Connect Timeout = 60; Initial Catalog = teach
er; Client CacheSize = 25; Auto Synch Period =
```

10 000 ”

. Open

End With

Set catOLAP.ActiveConnection = cnnOLAP

OLAP SERVER 连接好后,即可用 MDX 语句进行多维查询,通过 ADOMD 打开一个数据集,并将结果保存在单位集对象 (cellset) 中。在本系统中,MDX 语句可以根据用户的动作动态生成。例如,用户在 Contract 数据立方体中选择了“所有院系”作为行,“合同数、合同金额”作为列,即查看所有院系项目数量和项目总金额。那么动态生成 MDX 语句为

SELECT

[[([Measures].[项目数]),([Measures].[合同金额])] ON COLUMNS,

[[([院系部门],[系所处].members)] ON ROWS

FROM 项目

如果用户又对该数据立方体的某一字段实行下钻操作,即查看更具体的数据,各个院系中,每个研究所有多少合同,多少合同金额。那么生成的 MDX 语句为

SELECT

[[([Measures].[项目数]),([Measures].[合同金额])] ON COLUMNS,

DrillDownMember([院系部门].[系所处].Members).

[院系部门].[全部].[研究所] ON ROWS

FROM 项目

MDX 语句生成后,就可以通过 ADOMD 访问查询结果:

Dim CellOLAP AS ADOMD. CellSet

重置 CellSet

IF Not CellOLAP IS Nothing Then

Set CellOLAP = Nothing

End IF

Set CellOLAP = New Cellset

执行生成的 MDX 语句,获取结果

CellOLAP.Open strMDXTest, cnnOLAP

此外,系统还采用 TreeView (树型) 控件呈现 OLAP 数据立方体的多维层次结构,它用各个节点来显示 OLAP 数据立方体的立方体定义、维和层次等结构信息,方便用户浏览和观察。

3 结束语

该系统目前已开发调试完成,多维分析部分用 MS Analysis Services 开发,客户端调用 VB6.0 实现,数据抽取、转换部分由 DTS 和自行编写代码完成。图 4 是以项目为主题的分析界面,以院系部门、负责人和项目类别作为维度来查看项目数和项目金额,通过点击统计值可得到相应的具体数据,图 5 是图型化分析界面。



图 4 以项目为主题的分析系统界面

Fig. 4 Analysis system's interface of project

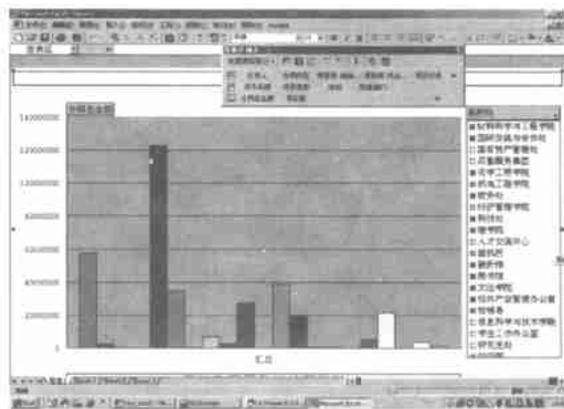


图 5 图型化分析系统界面

Fig. 5 Analysis system's interface of graph

本文通过对高校教师的科技信息数据仓库进行多维分析,帮助决策者多角度地深入分析,而不是只局限于固定的报表。本文对系统设计与实现等技术层面进行了深入的研究,对以后的高校科技数据仓库中的数据分析应用有一定的指导意义。

参 考 文 献

[1] 孙宇涵,邓胜春,徐晓飞. 基于 MS Analysis Services 的多维查询分析系统的研究与实现[J]. 计算机应用,

- 2002, 20(9): 89 - 91 京:机械工业出版社, 2000
- [2] Claude Seidman. SQL Server 2000 数据挖掘技术指南 [M]. 北京:机械工业出版社, 2002 [4] Microsoft. OL EDB for OLAP, Version 1.0 specification [M]. New York: Microsoft Company, 1998
- [3] Jake Strum. SQL Server 7 数据仓库技术指南[M]. 北

Design and implement of OLAP system based on MS analysis services

Qiu Huai-shan Zhu Qun-xiong

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: OLAP and relative concepts were introduced, and how to build and access multi-dimension database in MS Analysis Services was discussed. A method of design and implement of OLAP analysis system was proposed, with an example from database design to application design.

Key words: multi-dimension analysis; data cube; MDX; DTS

(责任编辑 刘同帅)

北京化工大学两项具有高水平的水处理项目通过鉴定验收

由魏刚教授、熊蓉春教授为首分别承担的国家“十五”科技重点攻关项目“工业蒸汽锅炉节水成套技术开发及应用研究技术”及“工业水处理用绿色药剂环氧琥珀酸中试开发及应用研究”近日通过了教育部组织的技术鉴定和国家专用化学品办公室组织的项目验收。来自中国国家工程咨询公司、中国锅炉水处理协会、天津化工研究设计院、蓝星清洗工程有限公司等单位的专家对项目的资料进行审查,并对现场进行了考核,对项目的整体工作进行了认真的评价。

专家一致认为,“工业蒸汽锅炉节水成套技术开发及应用研究技术”课题组通过科技攻关,对工业蒸汽锅炉的汽水系统的运行规律、腐蚀与结垢机理进行了大量研究,突破锅炉必须排污的观念和传统,首次开发了排污率接近于零的蒸汽发生技术,制备了超分子型结构的缓蚀剂与防腐蚀技术,国际首创地实现了工业锅炉在接近零排污工况下的安全、经济运行。整体技术达到国际领先水平。课题组开发并建成的 60 t/h 的工业示范装置,技术路线先进,工艺合理,机构紧凑,采用计算机监控、运行平稳,成功消除离子交换工艺的废水,基本消除锅炉排污和被腐蚀产物污染的凝结水等废水排放,经济效益、社会效益和环境效益十分明显。

目前,蒸汽锅炉使用软化水,生产 1 吨蒸汽需补水 1.4 ~ 1.8 t,排污水 1.4 ~ 1.8 t。采用本技术,应用到目前的 34.91 万台蒸汽锅炉,可节水 54 万吨,减少排污水 50 ~ 64 万吨。聚环氧琥珀酸(PESA)是近年国外发现的两种绿色聚合物(另一种是聚天冬氨酸)之一,主要用作阻垢剂和洗涤剂。PESA 具有无磷、非氮结构,不会引起水体富营养化。其可生物降解性好,是公认的绿色阻垢剂。其对钙、镁、铁等离子的螯合力强,适用于高碱高固水系,是阻垢剂的更新换代产品。专家对“工业水处理用绿色药剂环氧琥珀酸中试开发及应用研究”项目进行评价后认为,课题组开发了合成聚环氧琥珀酸的三种催化剂、两条技术路线和定向引导工艺,技术达到国际先进水平,工艺流程、生产效率达到国际领先水平。课题组建成的 300 t/a 的中试装置,流程合理、工艺先进、运行平稳,生产过程无“三废”,产品性能优良,经济效益明显。