

基于 Analysis Services 的 OLAP 展现工具 在教务系统中的应用

OLAP Exhibition Tool Based on Analysis Services Applied to Educational Administration System

张选东 张小玲 李晋宏 (北方工业大学 100041)

摘要: 本文讲述了 OLAP 的特点, 建立数据仓库的步骤, OLAP 的设计和实施, 并完成了 OLAP 展现工具 proOLAP, proOLAP 系统结构和完成的主要功能, 并以教务系统为例用 proOLAP 对学生成绩进行了分析。

关键词: OLAP 数据仓库 展现工具

1 引言

目前数据仓库市场上的工具基本上没有国产的, 由于文化背景的差异, 西方人的数据展现习惯与中国人有很大差别, 因此, 几乎所有的前端工具都难以满足国内最终用户的需要。好在许多 OLAP 工具都提供了 API 接口, 可以开发定制的应用界面, 但是, 这不仅大大增加了数据仓库建立的复杂程度, 也延长了建设周期。为此, 基于 Analysis Services 作者完成了 OLAP 前端展现工具 proOLAP, 并成功的应用到了教务系统中。

本文讲述了 OLAP 的特点, 如何建立数据仓库, 将以教务系统为例用 proOLAP 对学生成绩进行分析。

2 OLAP

与数据仓库紧密联系在一起的是“联机分析处理”OLAP (On-line Analytical Processing), 它是基于数据仓库的信息分析处理过程, 是数据仓库的用户接口部分。根据 OLAP 委员会的定义: OLAP 是使分析人员、管理人员或执行人员能够从多种角度对从原始数据中转化出来的, 能够真正为用户所理解的并真实反映多维数据特性的信息进行快速、一致、交互的存取, 从而获得对数据更深入了解的一类软件技术。OLAP 的目标是满足决策支持或多维环境特定的查询分析需求, 它的核心是“维”的概念, 因此, OLAP 也可以说是多维数据分析工具的集合。

1993年, E.F.Codd在“Providing OLAP (On-

line Analytical Processing) to User-Analysts”一文中提出了有关 OLAP 的 12 条规则, 其后, 部分公司根据各自对 OLAP 的理解也提出了一些规则, 如表 1 所示。

近来, 随着人们对 OLAP 理解的不断深入, Nigel Pendse 和 Richard Creeth 提出了更为简明的定义 FASMI 准则, (即 Fast Analysis of Shared Multidimensional Information)。

2.1 快速“Fast”

指系统应以相当固定的速度向用户提供信息, 大多数查询应当在 5 秒或更短时间内提交给用户。

2.2 分析“Analysis”

指系统应能对于用户及应用相关的任何业

务逻辑进行统计分析,这一分析过程不用编程而且尽量利用已有的综合路径及统计公式。分析类型应包括时间序列分析、成本分配、货币兑换、非过程化建模、多维结构的随机变化等与应用相关的特征。

2.3 共享“Shared”

指多个用户存取数据时,系统应保证安全性,尤其是当存在多个用户向OLAP服务器写数据时,系统应在适当粒度上加更新锁。

2.4 多维“Multidimensional”

是应用的实质,系统必须提供数据的多维概念视图。

2.5 信息“Information”

访问应用程序是必须的,相关的所有数据

和信息,而不管它驻留在何处,并不受磁盘分区和卷的限制。

OLAP以多维形式表示数据,是分析人员能够从不同的角度对数据仓库中的数据进行观察,如下钻、旋转、切片与切块等。同关系模型相比,多维概念模式和用户视图更接近人类的思维习惯,有助于模型设计和分析操作。用户通过“切片和切块”或为旋转可以轻松地完成传统方法难以完成的分析工作,也可以使决策人员更容易验证和发现有用的信息,提高决策的科学性和及时性。

3 建立数据仓库

建立数据仓库的目的是为决策支持提供服

务的,数据仓库中存储的数据是面向决策分析目标的、经过提炼、加工后的数据集。这种数据存储结构为OLAP的实施提供了理想的环境;而OLAP作为一种多维查询和分析工具,是数据仓库功能的自然扩展,也是数据仓库中大容量数据得以有效利用的重要保障。

完整的数据仓库实现方案包含下列要素,这些要素是成功建立数据仓库所必需的主要技术环节。

(1) 从数据源中抽取数据将管理决策需要的数据从各个数据源中抽取出来。当然,这些数据可以是外部异构数据,需经过进一步整理才能放入数据仓库。

(2) 对源数据进行清理和整合清理工作,包括对数据进行完整性和有效性检查,以保证数据仓库中数据的质量。而整合是指对数据的类型、格式、单位等进行统一。

(3) 按一定计划或规则进行数据仓库的加载和更新,保证数据仓库按照数据和信息使用的时间要求准时加载、更新。

(4) 按决策需要组织数据清理好的数据还要进行重组,按照决策的需要组织成不同主题的数据仓库表。

(5) 对数据进行分析处理提供各种函数功能和分析过程供编程使用,以方便决策支持数据分析。

4 OLAP的设计与实施

依据前文所述的OLAP的十二条评价规则,OLAP要对来自基层的操作数据进行多维化或与综合处理,因此不同于传统的两层客户/服务器结构,而是三层客户/服务器结构,所以设计OLAP要构建三层体系结构,要预定义维度、维表和事实表。

4.1 构建OLAP的三层体系结构

OLAP的三层体系结构是企业服务器/应用服务器(OLAP)/客户查询分析,本项目的体系

表1 关于OLAP的规则列表

Codd的12条规则	Gartner Group的9条规则
(1) 多维概念视图	(13) 多维数组
(2) 透明性	(14) OLAP连接操作
(3) 存取能力	(15) 数据库管理工具
(4) 一致的报告性	(16) 对象存储
(5) 客户机/服务器结构	(17) 子集选择
(6) 维的等同性	(18) 细节深入
(7) 动态的系数矩阵处理	(19) 局部数据支持
(9) 多用户支持	(20) 递增数据库刷新
(9) 非受限的跨维处理	(21) SQL接口
(10) 直观的数据操作	IFI软件公司的3条规则
(11) 灵活的报告	(22) 时间序列分析
(12) 不受限的维和聚合层次	(23) 过程语言和开发工具
	(24) 功能的集成化

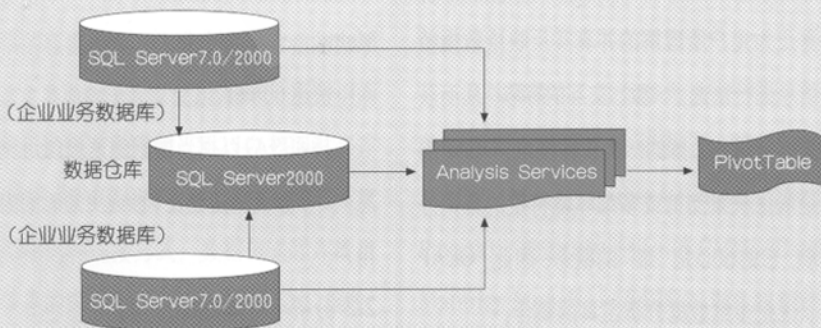


图1 OLAP的体系结构

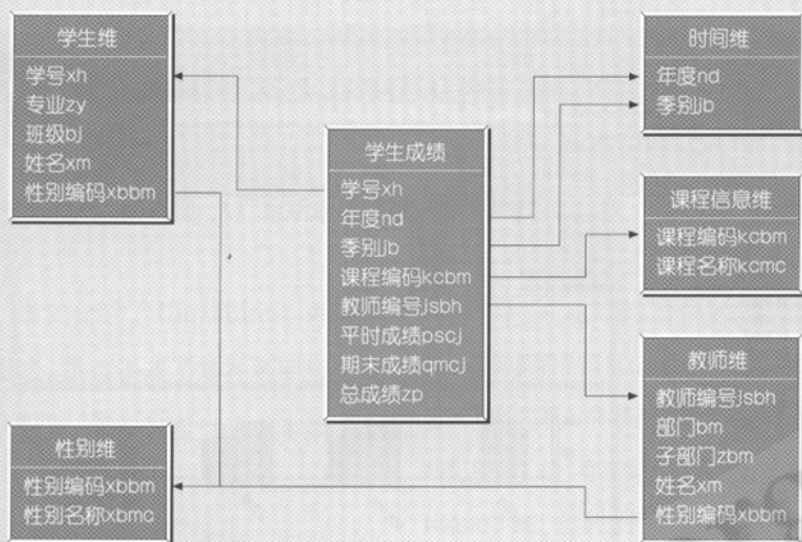


图2 成绩分析总体结构

结构如图1:

图1中第一层是企业服务器和数据仓库细节数据,采用MS的SQL Server 7.0或SQL Server 2000作为企业业务数据库,SQL Server 2000作为企业数据仓库数据库;第二层是OLAP应用服务器,采用MS的Analysis Services构建;第三层是面向企业高层和部门及人员的前端分析界面,采用MS的PivotTable作为数据调用的管道。

4.2 确定 OLAP 的维度

维度是设计OLAP过程十分重要的,维度的划分要有利于前端方便地分析查询,同时要简洁、清晰、便于最终用户理解。以成绩分析为例,在设计中把学生的平时成绩、期末成绩、总成绩作为度量之;学生、教师、课程、时间、性别作为维度。要注意:一个多维数据集最多可包含128个维度(每个维度中可包含数百万成员)和1,024个度量值。具有适当数目的维度和度量值的多维数据集通常能够满足最终用户的要求。

4.3 维度表和事实表的生成

作者设计的OLAP是基于关系数据库实现的,关系数据库将多维数据库的多维结构划分为两类表:一类是事实表(1),用来存储事实的度量(1)值及各个位的码值;另一类是维表,对每一个维来说,至少有一个表用来保存该维的

元数据,即维的表述信息,包括维的层析及成员类别等。

事实表是通过每一个维的值和维表联系在一起的,该结构有时被称之为“雪花模式(snowflake schema)”,如图2所示。

学生成绩为事实表,学生维、教师维、课程维、时间维和性别维为维度,学生成绩通过xh、nd、jb、kcbm、jsbh和学生成绩表中的各个字段一一对应。教师维和学生维通过xbbm和性别维中的xbbm对应。

5 展现工具的实现

OLAP分析工具是对多维形式组织起来的数据采取切片、切块、旋转等各种分析动作,以求剖析数据,使最终用户能够从多个角度、多侧面地观察数据库中的数据,从而深入地了解包含在数据中的信息、内涵,proOLAP实现的功能:

- (1) 无缝连接 Microsoft OLAP Server,
- (2) 提供目录树结构模型描述体系,
- (3) 提供多维表格和丰富图形两种数据展现方式,
- (4) 任意选择维度及维值组合查询分析,
- (5) 提供上下钻、透视表旋转、切片等基本多维分析操作。

- (6) 提供最大值、均值、求和等辅助统计功能,
- (7) 预留自定义指标接口,分析用在线添加计算指标项,

- (8) 即席打印功能,

- (9) 支持分析模板的保存加载使用,它包括几个主要模块: Connect; MetaData; Results; Slicer,

· Connect 模块

实现从 Browser 端到 Analysis Services 服务器的连接,并实现安全性和用户的身份验证,用户需指定待连接服务器的 IP 地址或机器名称,提供用户名和口令,并指定待连接数据库,完成连接。

· MetaData 模块

这个模块以树型结构显示选定 Cube 的元数据,每个 Cube 都有一个特殊的维——Measures,这个维的成员都是度量值,提供最终用户感兴趣的数据,元数据的层次结构为 Cube、Dimension、Hierarchy、Level、Member,这个模块可以帮助用户更好地理解数据仓库中的数据,方便的在 Cube 中导航,也可以帮助数据仓库管理员更好地管理仓库中的数据。

用户可以选定元数据并拖到 Results 模块中,实现图表的多维显示,如图3:

· Results 模块

Results 模块是整个系统中最重要模块,



图3 成绩分析的元数据

该模块实现数据的灵活展现，用户可以拖拉 MetaData 模块中的元数据到这个模块中的适当位置，实现图表的多维显示。

报表数据显示形式多种多样，可用表格，而且表格有 Explorer 和 Reporter 两种浏览方式，Explorer 方式下钻或上钻时不保留同一级别的其他数据，Reporter 方式下钻或上钻时保留同一级别的其他数据，也可用图形，如直方图、饼图、曲线图、棒图等，且表、图可以同屏显示。

· Slicer 模块

完成切片和切块操作。

6 结束语

利用本系统的 OLAP 分析工具，复杂的分析功能被简化为直观的、鼠标拖放式的数据导航——包括旋转透视表和交互式图表。灵活的 OLAP 分析包括即席的任意钻取，同时也可以对一个预定义路径钻取和细节钻取。图表和旋转透视表通过强大的格式化和异常提示功能来揭示数据中的显示趋势，数据被激活了。

此分析工具具有友好的图形界面为数据探查和决策制定提供了一个直观的环境。其一致的旋转透视表、图表和报表的设计模式，使得各

个层次的用户都可以流畅的在该产品中移动。

本系统集成查询、分析、报表于一个容易使用的工具。本系统的实施和维护，大大减少了 IT 人员在管理上的麻烦和时间。这些分析应用在整个企业组织中最大下限度的提高了组织的生产力和工作效率，因此比以往任何时候都有更多的用户可以通过一致的按钮点击方式访问他们所需要的关键信息，以便更好的进行分析决策。 ■

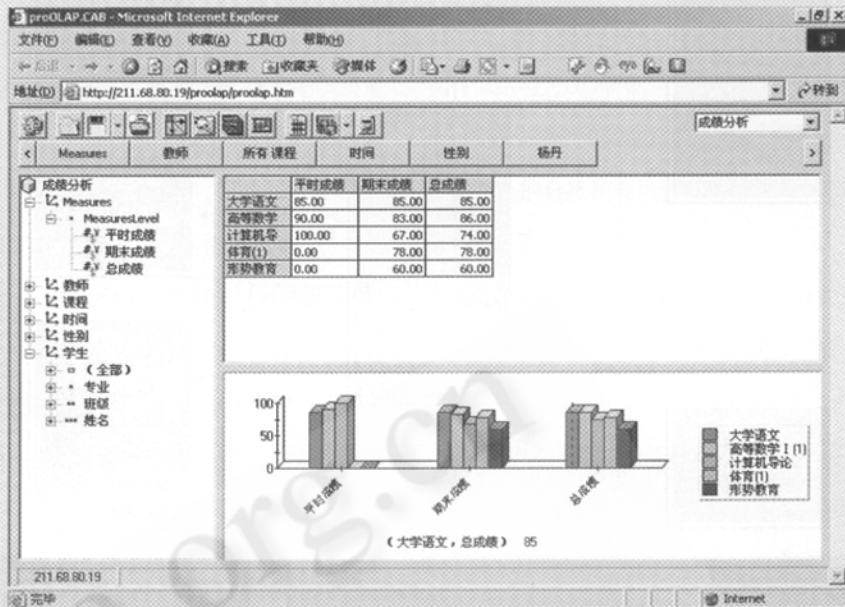


图 4 某学生成绩示意图

参考文献

- 1 Erik Thomsen. Microsoft OLAP 解决方案, 人民邮电出版社, 2000.
- 2 William C. Amo. SQL Server OLAP 开发指南, 电子工业出版社, 2000.
- 3 Mike Gunderloy Tim Sneath, SQL Server 开发指南——OLAP (联机分析处理), 电子工业出版社, 2001.

