OLAP技术在统计信息系统中的应用

张晋锐 （上海海运学院1288信箱 200135）

摘要：本文讨论了目前统计信息系统功能上的局限性，概要介绍了OLAP技术，并结合具体案例就如何充分发挥OLAP技术中动态、多维等特性，扩充统计信息系统功能做了一些探讨。

关键词：OLAP 统计信息系统

1 前言

在企业信息化的进程中，管理信息系统（MIS）建设是一个重要的环节。很多企业都相继建立了各种各样的信息系统。然而，从功能角度来看，这些系统往往偏重于"事务处理"，如人事管理、会计信息管理、库存管理等等。虽然能够很好的满足数据处理自动化的要求，但在决策支持方面却做的很不够。尽管在现有的信息系统中一般都包含统计信息子系统，但这些子系统很多仅仅局限于"统计报表"和"基本统计指标计算"的功能。用户得到的统计信息很有限而且很被动，不能满足用户"任意统计"的需求。在实际应用中，用户往往想从多维(角度)对分析对象进行分析，例如"在1至5月华东地区各省市分别销售了多少某某品牌的商品"就是一个多维的问题。而且往往想从各个角度进行分析，以及不同角度的任意组合。对于此类问题，传统的以SQL语言为核心的数据库开发系统是无法解决的。

随着OLTP、data warehouse、data mining等技术的迅速发展为传统信息系统的进一步发展提供了可能，将这些技术应用于信息系统中能够很好的发挥信息系统的决策支持功能。本文针对OLAP技术在统计信息系统中的应用作一些探讨。

2 OLAP技术简介

2.1 OLAP技术发展背景

60年代，关系数据库之父E.F.Codd提出了关系模型，促进了联机事务处理(OLTP)的发展(数据以表格的形式而非文件方式存储)。1993年，E.F.Codd提出了OLAP概念，认为OLTP已不能满足终端用户对数据库查询分析的需要，SQL对大型数据库进行的简单查询也不能满足终端用户分析的要求。用户的决策分析需要对关系数据库进行大量计算才能得到结果，而查询的结果并不能满足决策者提出的需求，因此，E.F.Codd提出了多维数据库和多维分析的概念，即OLAP。

2.2 OLAP的定义

(1) 定义1：OLAP(联机分析处理)是针对特定问题的联机数据访问和分析。通过对信息(维数据)的多种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进行深入观察。

(2) 定义2：OLAP(联机分析处理)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、并真实反映企业维特性的信息进行快速、一致、交互地存取，从而获得对数据的更深入了解的一类软件技术。(OLAP委员会的定义)

OLAP的目标是满足决策支持或多维环境特定的查询和报表需求，它的技术核心是"维"这个概念，因此OLAP也可以说是多维数据分析工具的集合。

2.3 相关基本概念

维：是人们观察数据的特定角度，是考虑问题时的一类属性，属性集合构成一个维(时间维、地理维等)。

维的层次：人们观察数据的某个特定角度(即某个维)还可以存在细节程度不同的各个描述方面(时间维：日期、月份、季度、年)

维的成员：维的一个取值。是数据项在某维中位置的描述。("某年某月某日"是在时间维上位置的描述)

多维数组：维和变量的组合表示。一个多维数组可以表示为：(维1，维2，...，维n，变量)。(时间，地区，产品，销售额)

数据单元(单元格)：多维数组的取值。(2000年1月，上海，笔记本电脑，$100000)

2.4 OLAP多维数据结构

超立方结构(Hypercube)：超立方结构指用三维或更多的维数来描述一个对象，每个维彼此垂直。数据的测量值发生在维的交叉点上，数据空间的各个部分都有相同的维属性。(收缩超立方结构，这种结构的数据密度更大，数据的维数更少，并可加入额外的分析维)。

多立方结构(Multicube)：即将超立方结构变为子立方结构。面向某一特定应用对维进行分割，它具有很强的灵活性,提高了数据(特别是稀疏数据)的分析效率。

2.5 OLAP的分析方法

切片和切块(Slice and Dice)：在多维数据结构中，按二维进行切片,按三维进行切块，可得到所需要的数据。如在"城市、产品、时间"三维立方体中进行切块和切片，可得到各城市、各产品的销售情况。

钻取(Drill)：钻取包含向下钻取(Drill-down)和向上钻取(Drill-up)/上卷(Roll-up)操作，钻取的深度与维所划分的层次相对应。

旋转(Rotate)/转轴(Pivot)：通过旋转可以得到不同视角的数据。

2.6 OLAP分类及体系结构

按照存储方式的不同可以将OLAP分为ROLAP(关系型OLAP)、MOLAP(多维型OLAP)以及HOLAP(混合型OLAP)。ROLAP的体系结构图如图1。

MOLAP的体系结构与上图所示不同的是用MOLAP Server代替了ROLAP Server，并且MOLAP Server将所需的数据装入一个特殊的多维数据库中。HOLAP则是将两者结合起来。

3 OLAP技术在统计信息系统中的应用

从以上的介绍中可以看出，OLAP技术的"多维"特性大大扩展了传统描述统计的视野，所要处理的数据不仅仅是"平面"的，而是变成"立体"的。下面笔者结合一个"信访案件统计分析系统"谈谈如何在统计信息系统中应用OLAP技术实现多维统计分析。

3.1 系统实施背景及目标

该系统实施单位为一个纪检监察部门，主要负责信访及各类违纪案件的处理工作。以前使用的一个MIS系统虽然能够实现一定的动态管理及统计报表的功能，但随着信访件和违纪案件数量的增多，如何去分析信访和发案的规律则是现有系统不能满足的。用户迫切需要实现的功能是从各个角度以及不同角度的组合去观察信访件和案件的总数以及比例，各个要素与案发的关

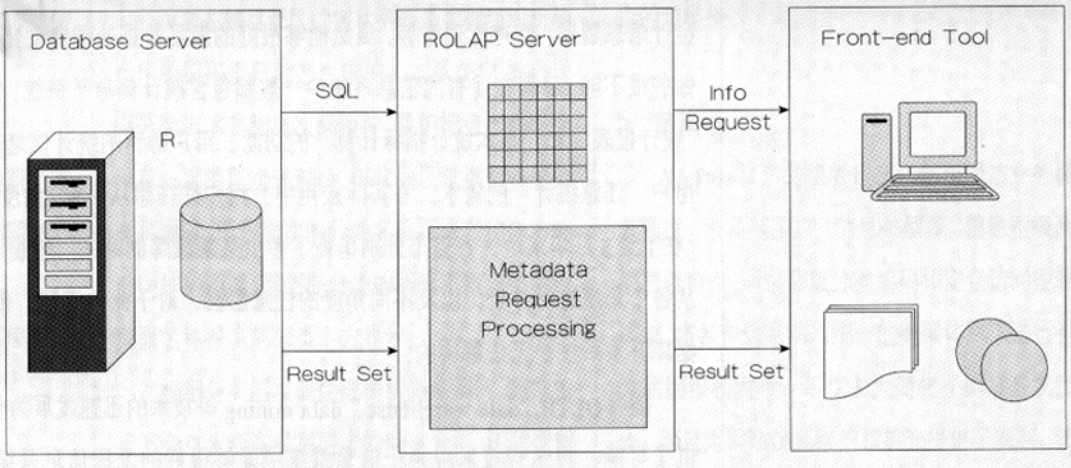

图1

联程度，从而为纪检监察工作提供科学的决策依据。总之，灵活的统计分析是用户提出的总的要求。

根据用户的要求，该系统在数据库系统基础之上主要采用了OLAP技术实现了多维统计表(图)的自动生成。

3.2 系统主要功能

系统的主要功能包括以下几个部分：

(1) 基本情况分析，基本情况分析的目的是向用户展示在某一时间段内信访或案件工作的总量和工作成果。

(2) 基本情况分类分析，基本情况分类分析的目的是向用户展示在某一时间段内各种不同类型信访或案件工作的总量和工作成果。

(3) 联动分析，联动分析说明各种不同类别组合的信访或案件数量，用以找出发案数量与各种类别组合的关系。

(4) 结构分析，用于说明在总体分布中各类案件(人员等)所占比例及整体分布情况。

(5) 趋势分析和动态平均指标，表明不同类别的信访或案件在不同年别的发案件数，用以分析案件发生变化的趋势、水平和速度。

(6) 比较变动分析，说明某类现象在两个不同时间段之间变化的相对程度。

(7) 特征值动态分析，说明某一范围内不同类别发案件数的最大、最小值、均值、中位数、四分位数差等，最容易发案的类别，并找出某时间、范围内最大的案件、作案人和相关材料。

(8) 关联分析，分析和显示任意两类在各种不同状态组合下的发案率和相互影响的程度，通过相互比较容易确定发案率为最高、最低和一般水平的状态组合。例如，对年龄和职务进行关联分析，得到各种职务和年龄组合的发案率，在不同年龄(职务)固定条件下，职务(年龄)在不同状态下的发案率大小等。

上面列出的功能中，有的可以通过SQL及动态SQL实现，如"基本情况分析"，大部分功能则必须采取OLAP技术实现多维统计，灵活生成透视表。

3.3 系统结构及主要技术

该系统采用了SQL Server数据库，在前端开发工具采用Visual Basic，并集成Excel，充分应用了ADO、VBA、Microsoft Office Web Components技术。(OLAP系统既可用于数据仓库，也可以用于标准的关系型数据库，该系统建立在关系型数据库基础之上)该系统结构如图2所示。

该系统采用的主要技术：

(1) 多维扩展，多维扩展(MDX)是由微软和其他厂商一起开发的SQL扩展语言，MDX由指定用于评价度量(evaluate measures)的位于立方体上的点来创建数据视图，也可以在一个立方体中使用MDX定义临时维和度量。

从形式上看，MDX类似于SQL，例如下面的语句就是从一个预先定义好的名为Sales的立方体(cube)中显示各个季度的销售额：

```
SELECT
{([Measures].[Unit Sales])} on
COLUMNS,
{([Time].[Quarter].MEMBERS)} on
ROWS FROM Sales
```

(2) 多维ActiveX数据对象(ADO MD)，ADO MD通过对ADO的扩展，使之包括了特定于多维数据的对象，如CubeDef和Cellset对象。使用ADO MD，用户能够浏览多维模式、查询立方并检索结果。在ADO MD中，中心元数据对象是"立方"，立方由相关的维、分级结构、级别和成员所构造的集合组成。其中：

"维"是多维数据库中数据的独立目录，由业务实体产生，通常，维包含用作查询标准以度量数据库的项目。

"分级结构"是合计维的路径，维可以有多个间隔级别，级别具有父子关系，分级结构定义这些级别之间的关系。

"级别"是分级结构中进行合计的一个步骤，对具有多层信息的维，每一层就是一个级别。

"成员"是维中的数据项目，通常，使用成员来创建标题或描述数据库的度量。

(3) Microsoft Office Web Componet，MOWC由图表工作区对象模型(ChartSpace)、数据源控件对象模型(DataSourceControl)、数据透视表对象模型(PivotTable)、电子表格对象模型(Spreadsheet)，这些控件可以很方便的应用于窗体程序或HTML页面中，通过对其编程可以实现非常强大的OLAP分析功能。

3.4 系统实现效果

该系统建成后，很好的满足了用户分析需求，用户可根据时间、案发部门、涉案人员特征、部门特征、职务特征等分析案发规律，起到了一定的预防、警示作用。

4 结语

通过以上的分析可以看出，随着数据仓库技术在数据库技术基础之上发展的越来越成熟，相对应的，SQL也扩展到了MDX(或类似的标准)，数据库系统中增、删、改、查的核心地位将被分析所代替，把OLAP技术以及数据仓库、数据挖掘技术应用于统计信息系统中则会大大提到统计系统的"分析"功能，更好的为用户提供决策支持。

参考文献

1 (美) Jake Strum 著，刘汉宇译，SQL Server 7 数据仓库技术指南，机械工业出版社，2000。

2 Microsoft 联机帮助。

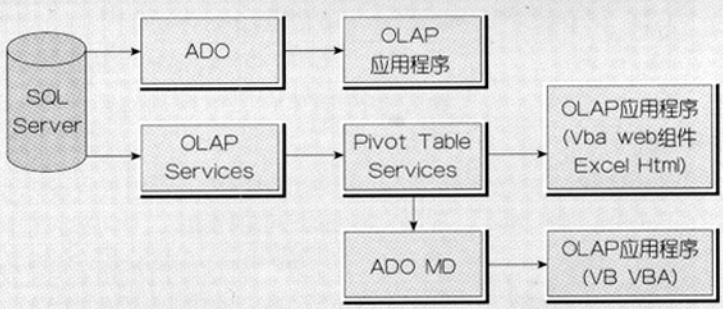

图2