维普资讯 http://www.cqvip.com



# 基于方法库的联机分析系统框架的设计与实现

王未央　经　霄
（上海海事大学信息工程学院　上海 200135）

**摘　要**　联机数据分析是解决现行的实际统计工作中普遍存在的问题的有效方法。本文介绍了一个基于业务空间和方法库上的联机分析模型以及多维联机数据分析系统框架 XOLDAS 的实现技术，采用该框架可以快速开发 OLAP 系统并具有良好的可扩展性。

**关键词**　OLAP，业务空间，方法库，XOLDAS

## The Design & Implementation of OLAP Framework Based on the Method Base

WANG Wei-Yang　JING Xiao
(College of Information & Engineering, Shanghai Maritime University, Shanghai 200135)

**Abstract**　Online Analysis Process is an effective way in solving a lot of common statistics problems. A multidimensional online data analysis model based on the business space and method base is introduced, and a framework (called XOLDAS) based on the model is presented. The framework is characterized good expansibility, OLAP system can be developed rapidly based on it.

**Keywords**　OLAP, Business space, Method base, XOLDAS

## 1　引言

在我国现行的统计、分析和决策工作中，有些问题长期困扰着人们，主要有：

1. 机关或企事业单位统计分析或信息机构，对于用户或领导提出的预料不及的各种各样的分析要求，无法快速、及时地提供资料和分析结果。

2. 我国现行统计分析资料的采集，基本上是自下向上、层层汇总、逐级报告的，因此，统计资料或信息的拥有量和管理层次的关系是金字塔形的。重要的决策需要由高层机构来做，但他们在很多情况下缺少足够的、可靠的有用信息。基层单位保存了大量第一手资料，但很难汇总利用。

3. 很多企事业单位和管理部门积累了大量电子业务数据和电子文本资料。但目前缺少有力的方法、技术和系统，直接从中找出用户想要的、有价值的信息来支持决策分析。

4. 我国目前经济和社会的快速发展使政府和企事业单位的组织结构、职能岗位、统计、分析要求和口径等，经常重组和变化，现有的信息系统很难适应这种大幅度的变动。

数据挖掘技术和联机分析处理 OLAP 技术可以部分地解决上述第 2 和第 3 个问题。但目前已知的国内外的相关技术和商品化的系统在实用方面存在一系列的不足和问题：

1. 缺少必要的模型和手段来描述或定义用户的业务问题和多种多样的、可能的、包括潜在的分析要求，因而也不可能对已定义的问题和分析要求在用户层面上方便地进行修改。

2. 没有必要的技术和框架允许用户随时在方法库中加入必要的分析方法，并使新加入的方法能够和用户定义的业务问题和分析要求相匹配。

3. 无法把要分析的业务问题的定义和分析要求与已有的数据源进行方便的连接，从而在数据源中抽取必要的数据供指定的分析方法进行分析。

为解决上述问题，本文给出一个建立在业务空间和方法库基础上的联机分析模型，以及在它的基础上开发的多维联机分析系统框架 XOLDAS 的实现技术。开放式的 XOLDAS 通过允许系统设计者装入定义业务问题的“业务空间”、“方法库”、“数据源”和“分析界面”来定义、扩充和修改面向用户问题的联机分析系统。最终用户可以通过选择业务空间中的“指标维”、“类别维”和“时间维”坐标来描述自己各种各样的分析要求和结果显示要求，从而构造出它所需要的视图和计算出分析结果。

## 2　多维联机数据分析系统框架 XOLDAS 模型结构

XOLDAS 作为一个系统框架，需要提供如下基本功能[1,2]：

- 基础服务：为 XOLDAS 框架中的其它层次提供最基本的服务。主要包括：日志、缓存以及 XML 解析等等。
- 业务空间：建立用户的需求分析与分析服务数据库的映射关系；集成服务与数据源、数据仓库的映射关系。
- 数据集成服务：一组通用的、可扩展的 ETL 访问接口。
- 联机分析服务：一组通用的、可扩展的分析服务（Analysis Services）数据访问接口。

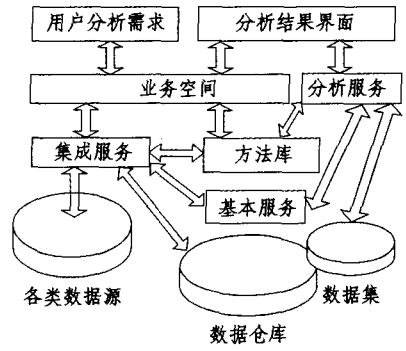


图 1　XOLDAS 模型结构

王未央　副教授，主要研究方向为信息管理技术；经　霄　硕士研究生，主要研究方向为计算机应用。

维普资讯 http://www.cqvip.com

·方法库:方法库将用户常用的分析问题进行分类,进行各种业务分析。

XOLDAS的模型结构如图1所示,分为五个层次,分别为:基本服务层、集成服务层、分析服务层、业务空间层和方法库。另外还有二个用户接口:用户分析需求和分析结果界面。层与层之间的调用是基于接口而不是基于实现,因此这种分层可以保证框架具有较好的可扩展性与稳定性。

## 3 XOLDAS设计

### 3.1 基础服务

基础服务层为XOLDAS模型中的其它层次提供一系列基础服务。主要分为三类[3]:

1)日志

日志模块用于记录程序在运行过程中的相关信息。这些信息可能是错误、警告或者其它类型。日志的记录方式可以记录在Windows事件查看器、数据库、文本文件或者XML文件中。用户可以在XML文件中配置日志记录类型和日志记录方式。

2)缓存

缓存模块是为了提高程序的运行效率,减少程序的交互时间。对于那些经常使用的数据,应该将它们放入缓存。这样可以避免每次读写硬盘的操作。在缓存中的数据如果被修改后,则在必要时需要写回磁盘。

3)XML解析

XML解析模块用于读取配置文件中的配置信息、将持久化的数据(XML格式)解析成一个或者多个复杂的对象并存入缓存或者将缓存中的数据写入XML文件。

### 3.2 集成服务

集成服务层用于进行数据的抽取、加载和转换。集成服务层又可分为三个模块:数据源模块、控制流模块和数据流模块。

·数据源模块定义数据源的类型,这些类型包括:数据库、文本文件、Excel文件、WebService服务或者数据仓库等等,数据源模块中的连接管理器负责对异构数据源的连接操作。

·控制流模块定义数据从源到目的数据仓库的业务流程。控制流模块提供以下基本功能:

1.添加在包中实现的工作流或将控制流划分为子集的容器;

2.添加支持数据流,准备数据流,执行工作流和商务智能功能以及实现脚本的任务;

3.使用优先约束把容器和任务连接为有序控制流。

·数据流模块定义了数据从源到目的地的过程中进行转换时的映射关系。数据流模块提供以下基本功能:

1.添加一个或多个数据源以便从文件或数据库中提取数据;

2.添加满足包的业务要求的转换;

3.通过将源和转换的输出连接到转换目标的输入来连接数据流组件;

4.添加一个或多个目标以便将数据加载道数据存储区;

5.配置组件的错误输出以处理问题,如被截断的错误或数据值。

### 3.3 联机分析层

联机分析服务层为商务智能解决方案提供联机分析处理功能。联机分析层包含三个模块:元数据模块,Adomd模块和AMO模块。

元数据模块用于查询OLAP数据库的元数据信息。元数据信息包括:数据库、维度、立方体、层次、属性、级别等等。

Adomd模块用于进行各种多维数据查询。查询可以返回多维数据集或者只返回一个只读的对象。

AMO(Analysis Management Object)模块用于对OLAP数据库的操作。客户端可以通过AMO模块创建或者删除OLAP数据库中的元数据,元数据可以是数据库、立方体、维度或者度量等等。

### 3.4 业务空间

业务空间描述了联机分析系统所要分析的业务问题的内容和范围。一个业务问题包含若干用户关心的、必须加以研究的因素,称为业务要素。如"时间","产品","价格","销售渠道"等等,都是"企业效益"这样一个业务问题的业务要素。有的业务要素可以再细分。在诸多业务要素中,那些不可再细分的、互相独立的业务要素称为"基本业务要素"。它们的集合构成了这个业务问题的业务空间。每一个基本业务要素就是这个业务空间的一个"维"。

业务空间中的维被分为三种类型,即"时间维","分类维"和"指标维"。分类维用于对业务要素或维坐标值的细分;指标维用于说明所要分析的内容。任何一个用户要分析的问题的内容和范围,都可以看作是业务空间中的某个子空间。每一个要分析的问题至少应该包括一个指标维的坐标值以及若干个分类维和时间维坐标。

业务空间面向用户,用户是很难理解数据源中数据意义的,因此在业务空间中要增加数据对应的解释词汇。

### 3.5 方法库

方法库位于业务空间层之下,集成服务、基础服务和联机分析服务层之上。用户提出分析需求后,通过将业务空间层将用户需求合成一个模式,该模式封装了调用方法库类型和多维查询语句,在方法库中,通过该模式与数据立方体中的数据集关联。

XOLDAS框架目前为用户提供了五类常用的统计分析方法,并把这些统计方法组织在方法库中。每一类分析方法下又可以细分为多种具体的分析方法。每一个具体的方法又与一个具体的模式相对应。XOLDAS框架的方法库的五类统计分析方法包括[4]:

1.基本情况分析。它是最基本的一种分析方法,它将用户在业务空间中所关心的类型数据通过二维表格的形式显示给用户。它不涉及任何的计算。

2.计算。用于对可计算的类型进行简单或者复杂的数学计算并生成临时类型。常用的计算有结构分析、总量分析、支持度分析等。结构分析用于说明类型在总体分布中所占的比例及整体分布情况。例如:计算平均价、利润和占有率百分比等等。总量分析用于计算类型的总和。支持度分析用于计算所选类型占全部指标值合计的比例。

3.分组。分组可以从多角度、多层次灵活地定制数据聚集地范围。

4.比较。对于可计算的类型,可以灵活地定制可比系列。常用的比较有同期比和环比。

同期比主要是为了消除季节变动的影响,用以说明本期发展水平与去年同期发展水平对比而达到的相对发展速度。如,本期2月比去年2月,本期6月比去年6月等。环比是报告期水平与前一时期水平之比,表明现象逐期的发展速度。如计算一年内各月与前一个月对比,即2月比1月,3月比2月,4月比3月……12月比11月,说明逐月的发展程度。

5.高级分析。对于一些复杂的分析,可以使用高级分析

维普资讯 http://www.cqvip.com

方法。高级分析方法包括:改善度分析、相关分析以及差异显著性检验。改善度分析是一种相对较为复杂的分析方法,又称为相互影响度分析。它用于说明类型之间的相互影响程度;相关性分析用于说明可计算类型 $i$ 与可计算类型 $j$ 之间关系的密切程度;差异显著性检验通过对各种类型的指标值的多重比较。找出具有显著大或者显著小或一般情况等特征的类别。

这些基本分析方法可以满足用户大多数情况下的分析要求。此外,用户也可以综合运用这五种分析方法进行复杂的统计分析。

### 3.6 用户接口

XOLDAS框架有二个用户接口:用户分析需求和分析结果界面。

用户分析需求接口用来接受用户的分析请求,通过业务空间,用户可能方便地选择或输入期望得到的结果,业务空间层在组合其需求后调用方法库中的方法进行数据分析服务。

分析结果界面接口是向用户反馈数据分析结果,用户也可在该界面中灵活地变换分析需求得到新的结果以及图形曲线。

**结束语** XOLDAS框架是建立在业务空间和方法库基础上的,通过对业务空间的重新定义,可以快速建立一个新业务的OLAP系统,通过对方法库的扩充,可以方便地增加数据分析方法,因此系统具有良好的快速开发能力和可扩展性。

## 参 考 文 献


1 Thomsen E. OLAP Solution : Building Multidimensional Information System. 北京:电子工业出版社,2004. 4~5

2 Nguyen T B, Tjoa A M, Wagner R R. An Object Oriented Multidimensional Data Model for OLAP. In: Proc. WAIM. Shanghai, China, June 2000. 130~132

3 徐忠健,袁捷,陆菊康,陈毛狗. 超市决策支持系统的数据仓库的设计与实现. 计算机工程与应用,2003(18):56~57

4 肖昭媛. 统计学原理与应用. 上海:上海交通大学出版社,2002


(上接第 115 页)

(MDR)和每5秒钟内低许可级事务被高许可级事务阻塞的数目(LBN)。其中MDR的定义如下:MDR=(错过截止期事务数)/(系统接纳的事务数)。MDR反映了系统实时性能,而LBN则用来衡量违犯安全性的严重程度。在我们的仿真实验中,优先级分派采用EDF(Earliest Deadline First)策略。事务的截止期($DL$)按如下公式计算:$DL=AT+SF\times ET$,这里,$AT$ 表示事务的到达时间;$SF$ 表示松弛因子,为满足均匀分布的随机变量;$ET$ 表示事务的估计执行时间。仿真测试的主要参数设置如表1所示。

表1 仿真测试参数

| 参数 | 值 | 含义 |
|---|---|---|
| R | [5, 40] | 事务到达率 |
| $P_U$ | 0.4 | 更新操作的概率 |
| SF | U[2.0, 6.0] | 松弛因子 |
| NCL | 6 | 不同许可级的级别数目 |
| RCL | 1/6 | 不同许可级事务的比率 |
| ω | 0.5 | 确保数据安全的权值 |

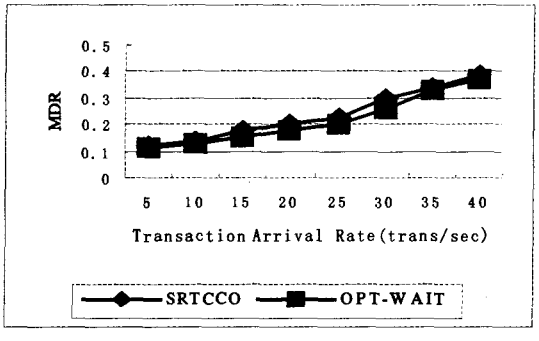


图1 MDR的比较

在表1中,$U[i, j]$表示在区间$[i, j]$上满足均匀分布的随机变量。性能测试结果如图1和图2所示。

图1显示了OSRTCC和OPT-WAIT在MDR方面的比较,随着事务到达率的增大,OSRTCC和OPT-WAIT的MDR都相应增加,但OPT-WAIT稍稍优于OSRTCC。图2给出了OSRTCC和OPT-WAIT在LBN方面的比较,显然,相对于OPT-WAIT,OSRTCC具有明显的优势。

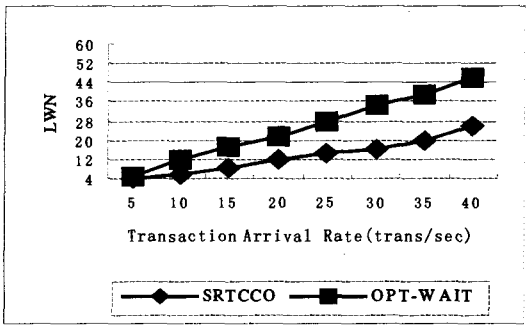


图2 LWN的比较

**结束语** 一个安全实时数据库系统必须同时满足两个目标:确保数据安全和尽可能减低实时事务错过截止期的比率。然而,这两方面需求通常是相互冲突的,满足一方面是以牺牲另一方面为代价。本文提出了一种基于乐观方法的安全实时并发控制协议(OSRTCC)。OSRTCC在验证阶段引入了安全检查机制,并通过定义安全违背因子和实时影响因子来有效地实现在安全性和实时性两方面的折中。仿真实验显示,OSRTCC在确保数据安全的同时仍能保持良好的实时性能。

## 参 考 文 献


1 刘云生. 现代数据库技术. 北京:国防工业出版社,2001

2 Sha L, Rajkumar R, Son S H, et al. A real-time locking protocol. IEEE Transactions on Computers, 1991. 793~800

3 Haritsa J, Carey M, Livay M. Data access scheduling in firm real-time database systems. Real-Time Systems Journal, 1992(4)

4 George B, Haritsa J. Secure transaction processing in firm real-time database systems. ACM SIGMOD Record, 1997, 26(2): 462~473

5 David R, Son S H, Mukkamala R. Supporting security requirements in multilevel real-time databases. IEEE Transactions on Knowledge and Data Engineering, 2000. 865~879

6 David R, Son S H. A secure concurrency control protocol for real-time database. In: Proc. of Annual IFIP WG 11.3 Conference of Database Security, August 1995

7 朱虹,冯玉才. 避免隐通道的并发控制机制. 小型微型计算机系统,2000,21(8):844~846