

# 基于数据仓库的综合查询系统的设计和实现\*

张凤荔\*\* 葛晓峰 卢显良

(电子科技大学计算机学院 成都 610054)

**【摘要】** 在详细论述数据仓库思想、体系结构与工作原理以及相关的重要技术的基础上,给出了厂长综合查询系统的数据仓库结构和数据仓库中数据采集、数据的可视化的方法和实现示例。最后讨论了建立行之有效的决策支持系统的构想。

**关键词** 数据仓库; 联机事务处理; 综合查询; 数据挖掘

**中图分类号** TP392; TP311

## 1 数据仓库

### 1.1 数据仓库

随着数据量的迅速增大以及查询要求的复杂化,人们对服务于 OLTP(On-Line Transaction Processing)的数据库中的原始数据进行再加工,形成一个综合的、面向分析的环境以支持决策的产生,逐渐形成数据仓库的思想。数据仓库最主要的四个特征为面向主题的、集成的、稳定的、随时间变化的,它的一种典型的结构如图1所示<sup>[1]</sup>。

数据仓库分为早期细节级、当前细节级、轻度综合级和高度综合级四级。原始数据首先进入当前细节级,进一步进行综合后进入轻度综合级乃至高度综合级,老化的数据进入早期细节级。数据仓库中存在着不同的细节级别,称之为“粒度”。粒度越高表示细节程度越低、综合程度越高,回答细节问题能力下降。粒度既影响到数据仓库中数据量的多少,也影响到数据仓库所能回答询问的种类。在数据仓库中,多重粒度是必不可少的。

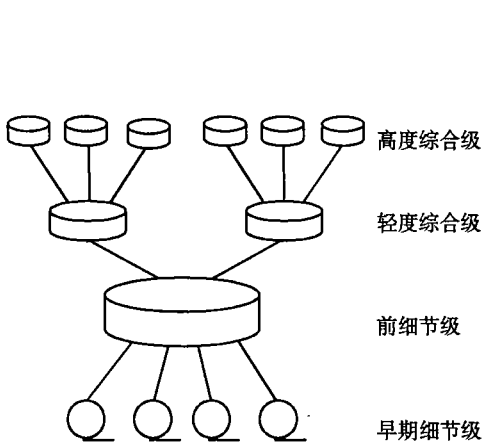


图1 数据仓库结构

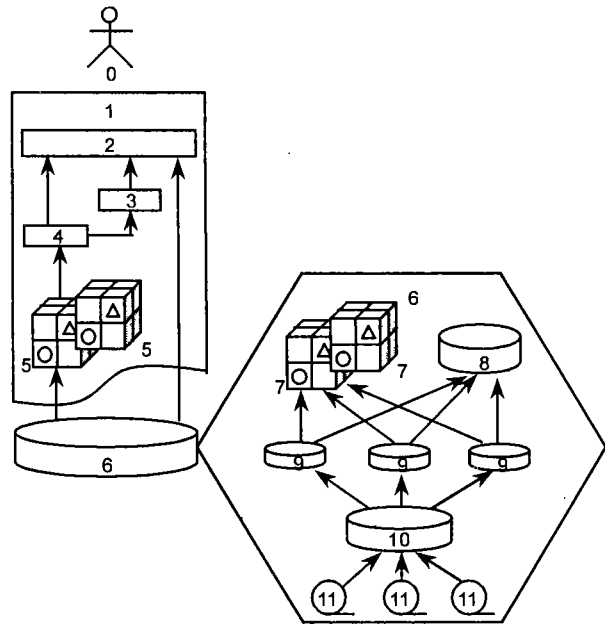


图2 数据仓库体系结构

### 1.2 DW的体系结构

数据仓库系统由三部分组成<sup>[1, 2]</sup>: 数据仓库 (DW)、数据仓库管理系统 (DWMS) 和数据仓库

1998年9月25日收稿,1998年10月6日修改定稿

\* 电子部预研基金资助项目

\*\* 女 35岁 硕士 副教授

工具。在图 2 中可以看出：数据仓库居于核心地位，是信息挖掘的基础；数据仓库管理系统负责管理整个系统的运转，起着承上启下的作用；而数据分析工具则是整个系统发挥作用的关键，只有通过高效的工具，数据仓库才能真正发挥数据宝库的作用，如图 2 所示。图中：1 为数据仓库工具；2 为可视化工具；3 为数据挖掘工具；4 为多维分析工具；5 为多维数据；6 为数据仓库；7 为多维数据；8 为关系型数据库；9 为轻度综合数据；10 为当前细节数据；11 为早期细节数据。

## 2 应用数据仓库实现厂长综合查询

### 2.1 厂长综合查询系统的结构与工作原理

厂长综合查询系统的结构为：各个部门的 OLTP 数据库位于三个不同的服务器上，两个 Foxpro 数据库运行在服务器 Nov1 和 Nov2 上，一个 Sybase 数据库运行在小型机上，涉及到全厂的销售、生产、财务、设备、物料、供应、人才、仓储管理等诸多方面(财务用 Nov1，其他部门公用 Nov2)的管理。厂长综合查询的 OLTP 数据库适用于厂长了解各个部门的业务情况，位于小型机(UNIX)上。两个 Foxpro 服务器通过网络交换机和小型机相连，进行数据交换。厂长查询数据仓库位于厂长查询的数据服务器上，它定期从 OLTP 数据库中取得数据并予以综合，得到面向全厂的查询和数据分析，其完整的工作过程如图 3 所示。(虚线框表示在这里用的是自己开发的模型化的工具)。

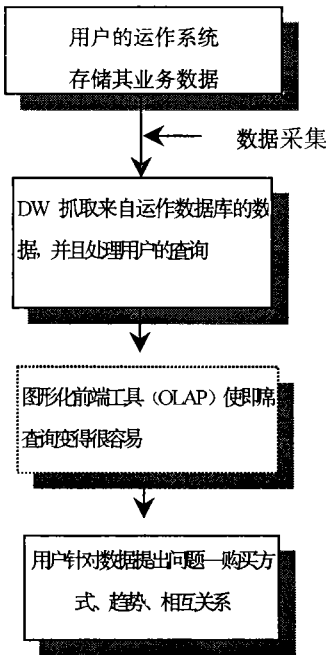


图 3 DW 工作流程示意

厂长查询数据仓库最初主题为“销售”、“生产”、“财务”和“人才”，它的逻辑设计采用了流行的星型模式(Star Schema)。星型模式是基于关系型数据库，面向 OLAP(Online Analytical Process)的一种多维化的数据组织方式。关系数据库将多维数据库中的多维结构划分为两类表：一类是事实表，用来存储事实的度量值及各个维的码值；另一类是维表，每个维至少有一个表保存该维的元数据，包括维的层次及成员类别等。这里以厂长查询数据仓库的主题之一“销售”为例来说明。选择销售金额与销售费用和与之相关的四个维——时间、市场、产品和客户组成事实表结构。事实表的时间维精确到每天。市场维分为省、地区、县三个层次；产品维表除产品代码、时间外，还含有名称、规格、单价等信息；市场维表和产品维表的时间粒度定为一年。客户维表由 DW 定期根据 OLTP DB 客户档案进行刷新。DW 中还存放其他和销售有关的基础信息和统计信息，为 DW 拓展主题、加大规模时，提供充实的资源。

### 2.2 厂长查询数据仓库的逻辑结构

数据采集是数据仓库建设阶段的关键所在。它的实现需考虑：采集的周期和时机，以满足数据的及时性和实时性要求，同时净化、集成数据，生成纯净、统一的企业数据视图。目前，将旧数据转换为数据仓库数据采用的方法有：综合、重构、筛选、重新格式化、转换和统一风格。数据采集的工作是通过用 PowerBuilder 编制的前端应用来完成的，支持异质/异构数据库间的批量数据传输。

### 2.3 数据的可视化呈现

由于后端数据表中存放的数据未必适合于用户从多种角度来观察，因此数据在到达统计图之前形成一种多维的数据视图，即数据的多维化。数据多维化成什么形式需视数据本身的情况而定，多维化的途径仍然是通过 SQL 语句来重组数据。例如表 1 所示的人员的多维数据视图。

表 1 人员多维视图

项目	分科	人数
性别	男	785
性别	女	724
学历	博士	5
.....		
部门	销售	1 108
.....		

只要有项目和分科两个维即可确定一个数据对象(人数), 它是一个二维的数据视图。其重组的 SQL 语句如下:

```

SELECT DISTINCT '性别',性别, count(*)
      FROM 人事档案表  GROUP BY 性别
UNION
SELECT DISTINCT '学历',学历,count(*)
      FROM 人事档案表  GROUP BY 学历
UNION
SELECT DISTINCT '部门',部门,count(*)
      FROM 人事档案表  GROUP BY 部门
.....

```

维的操作主要有切片和切块。当对某一维只保留单一值时称为对该维切片, 例指定日期 = 1997/07/24 是对日期的切片; 而指定如“日期在 1997/06/01 到 1997/06/30 之间”一类范围的过滤称为切块。上例中, 用户只关心本厂所有职员的一种构成分布, 使得可操作的维只有“项目”, 而操作方式只可能为切片。在其他的情况下, 必须仔细地分析各数据维的可操作性以及固有的相关程度, 有时统计图类型(折线图、条形图、堆积图、饼图以及散点图等)也不得不作为一个维来对待, 在制作多样化动态报表、全面进行数据分析与挖掘的工作中, 数据多维化都是强有力的基础工具和必不可少的技术手段。

### 3 数据挖掘与决策支持系统—数据仓库的最终目标

数据挖掘主要依赖于以下技术的支持程度<sup>[3]</sup>: 大规模数据采集、功能强大的多处理器计算机、数据挖掘算法。数据挖掘可自动预测趋势、行为并自动发现以前未知的模式。

图 4 示出了对大型数据仓库进行分析的结构。目前, DW+OLAP+DM 的结构已被公认为是 DSS 的有效解决方案。这种构架的真正意义在于重新展示信息的本质, 表明了信息系统的设计观念从处理驱动到数据驱动的转变。今后, 信息的重点将转移到数据模式分析, 处理技术则是应数据分析的需要而产生的。

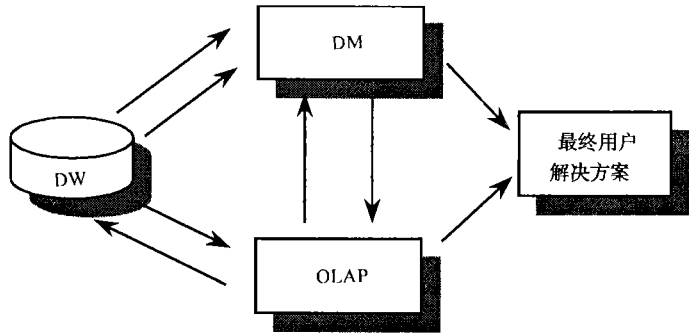


图 4 集成的数据挖掘结构

## 参 考 文 献

- 1 Dewire D, Bedell J, 王仲谋, 刘书舟译. 数据仓库—客户/服务器计算指南. 北京, 清华大学出版社, 1997
- 2 Red Brick Systems. The data warehouse: archieving batter decisions faster. Database and Network Journal, 1998, 28(3): 3~7
- 3 Guderian Dave, Leer Douge, Molini Steve. Keeping the data warehouse on track. Database Programming and Design, 1998, 11(1): 39~46

## Design and Implement of Comprehensive Query System Based on Data Warehouse

Zhang Fengli    Ge Xiaofeng    Lu Xianliang

(College of Computer Science and Engineering,, UEST China    Chengdu    610054)

**Abstract** This paper introduces the origination of the DW ideology and its architecture and principles, as well as relative techniques. The structure of DW of comprehensive query system of factory manager is also discussed. The implementation methods of data collection and data visual appears are presented. Finally, this paper proposes the ways of building an effective decision support system.

**Key words** data warehouse; on-line transaction processing; comprehensive query; data mining

· 科研成果介绍 ·

### 3 mm 雪崩振荡器系列

主研人员: 唐小宏 吴正德 樊 勇 张永鸿 张显静

该项目完成如下研究内容:

1) 3 mm 连续雪崩振荡器及注入锁定放大器完成了边续波雪崩振荡器的电路设计, 研制了适用于国内外器件的振荡器结构。在 3 mm 频段上, 输出功率大于 100 mW, 注入锁定放大器输出功率大于 700 mW, 注锁增益大于 17 dB, 锁定带宽大于 500 MHz。

2) 3 mm 脉冲雪崩振荡器及收发组件完成了脉冲雪崩振荡器和脉冲驱动源的电路设计, 研制了适用于国内外器件的振荡器结构。在 89 GHz~95 GHz 范围内点频工作时, 输出功率大于 10W, 脉宽 20 ns~50 ns, 重复频率为 10 kHz。研制成功非相干脉冲收发组件, 在 94.8 GHz 上, 脉冲发射功率大于 8 W, 脉冲宽度为 26 ns, 重复频率为 11 kHz, 接收机中频 1 GHz, 射频—中频变损耗 11 dB。

该成果性能优良, 工作稳定、实用, 具有很好的应用前景。

· 科 下 ·