

在 DSS/EIS & OLAP 环境下构造 FoodMark 数据仓库的设计与实现

The Design and Implementation of Building FoodMark Data Warehouse under DSS/EIS&OLAP

钟乐海

(四川师范学院计算机科学系 四川南充637002)

Abstract The thesis introduces the theory of data warehouse and data architecture, describes the technology of data architecture and data warehouse. Combining FoodMark data base, the building procedure of data warehouse and analytical method of the data are introduced. The construction of the data warehouse is realized with Visual Basic.

Keywords Data warehouse, Data architecture, Decision supply system, Olin analytical processing

1 引言

DSS 的全称是 Decision Supply System, 即决策支持系统, 是分析数据环境的一种称呼, 也是数据仓库存在的环境; EIS 是 Executive Information Processing 的缩写, 就是指高级管理信息系统, 说得更准确一点, EIS 实现的功能是 DSS 中的一类功能子集; 而 OLPA 的全称是 Online Analytical Processing, 即联机分析处理, 在 Microsoft SQL2000 中称为 Analysis Services. FoodMark 是 Microsoft Corporation 为了展示其 Analysis Services 而提供的一个数据库例库。本文实现的不是一个完整的数据体系框架, 我们所做的工作是构造 FoodMark 这个体系框架的很小的一部分, 只是其中的一个主题域, 所以具有示范性。

2 结构化数据体系环境(数据流层模型)

结构化数据体系环境是由生产环境、操作环境和分析环境(数据仓库环境)三种小环境组成的, 生产环境和数据的操作环境很容易让人认为是一种环境, 但它们是有区别的。生产环境向分析环境提供数据有两种方式, 一种是直接提供, 另一种是通过数据操作环境提供。对生产环境中的数据进行计算机操作处理是目前计算机最为广泛的应用之一, 而在这之中产生的数据是数据分析环境最为重要的数据来源。

结构化数据体系环境(数据流层模型)具有层次性、数据多态敛聚性(集成性)和数据多元分割等特点。

3 数据仓库环境

3.1 数据仓库及其特点

美国科学家 W. H. Inmon 著的《Building the data Warehouse》一书中对数据仓库是这样定义的: “数据仓库是支持管理决策过程的, 面向主题的, 集成的, 随时间变化的, 持久的数据集合。”虽然数据仓库的概念和理论诞生的时间并不长, 但其发展是相当迅速的。正是由于这一理论的出现, 才使人们意识到数据还可能存在于另一种环境中, 这种环境有别于我们平时的数据操作环境, 在这种环境中, 数据更加适用于对数据的分析和查询。数据仓库具有面向主题、集成性、非易失性和随时间变化等特点。

数据仓库的应用主要有如下几类:

(1) 主管信息系统(EIS) 提供界面丰富, 定制容易的决策分析, 主要适合企业的高层决策者使用;

(2) 联系处理分析(OLAP) 提供灵活丰富的多维分析与查询, 可以从不同角度分析企业的运作情况, 并对未来进行预测, 主要适合企业的中层领导和业务分析人员;

(3) 专门查询(AD Hoc Query) 提供多角色的专门查询, 适合于分析人员使用;

(4) 灵活报表(Reporting) 提供灵活的报表设计, 适合于报表制作人员。

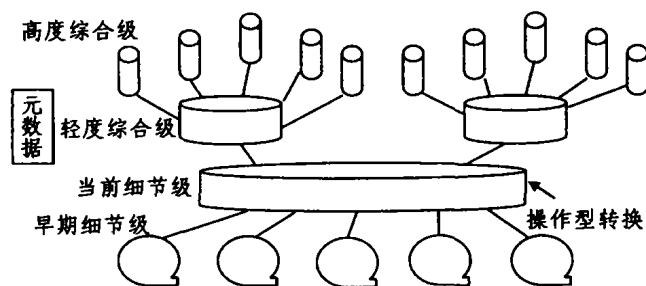


图1 数据仓库的结构

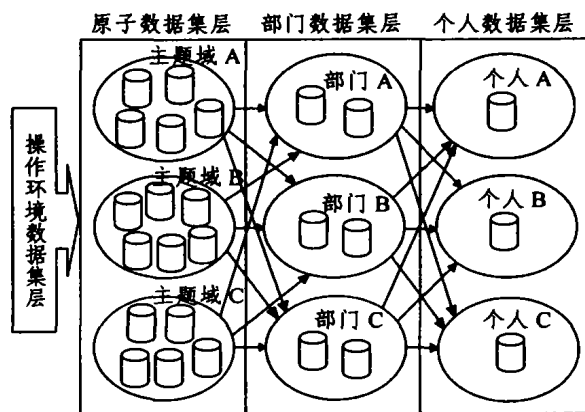


图2 数据仓库的体系结构

3.2 数据仓库的结构

在数据仓库中, 数据存在于不同的细节之中, 包括早期细节级、当前细节级、轻度细节级和高度综合数据级, 这些数据

是由操作型环境导入数据仓库的,相当数量的数据转换通常发生在由操作型级别向数据仓库级别传输过程中。图1所示的是一种经典的数据仓库体系结构,也是数据仓库最基本的结构,许多数据仓库都是以这种结构加以扩充发展而来的。

在结构化数据体系框架中,用数据集层的观点来描述数据仓库的结构,如图2所示。从这里我们可以清楚地看到,在数据仓库中不同层(或不同细节)之间的关系。

Microsoft Corporation 对数据仓库提出了一种具体而细致的解决方案,如图3所示,在这个框架结构中,清楚地表述了该公司对数据仓库所提供的解决方案,对各个环节都作了比较细致的说明,对使用的相应组件也作了必要的功能和特性介绍。

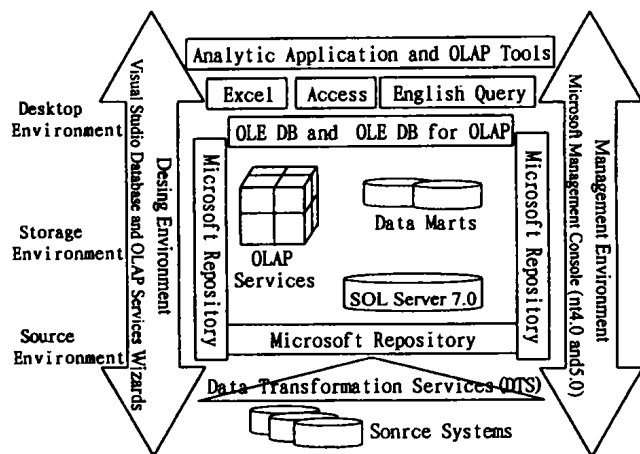


图3 Microsoft 数据仓库框架

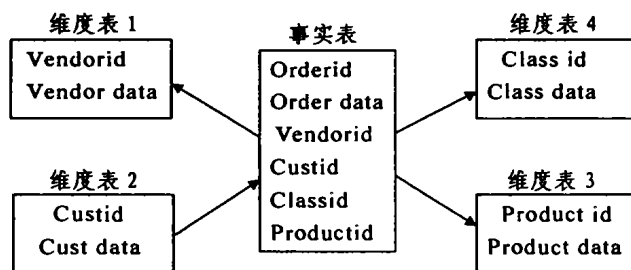


图4 星型数据仓库数据模型

4 数据仓库的设计

在数据仓库的设计中,主要做与操作型系统接口的设计和数据库本身的设计这两方面的工作。在某种程度上说,设计并不能精确描述在启发式下建造数据仓库时发生了什么,数据仓库的构造是一种反馈过程,首先载入一部分数据,供DSS分析人员使用和查看,然后,根据用户的反馈,在数据仓库中修改,增加一些数据。这种反馈循环贯穿于整个数据仓库的开发过程,在数据仓库部分载入并且为DSS分析员使用之前,数据仓库的需求有可能是不知道的,不同的人员对数据仓库中的信息的使用也可能是不同的,因此,在设计数据仓库时,不能采用传统的“需求-驱动”法,当然,那种认为数据仓库的构件不需预计需求的说法也是不正确的。在现实中,通常采用折中的办法。

建立数据仓库,从操作型数据开始。在操作型数据基础上建立一个星型数据仓库数据模型,如图4所示。位于星型连接中央的是“事实表”,而周围的表称为“维度表”,一般来说,事

实表中存放的都是标识数据,当然也包括事实表本身的数据和指向维度表的外键,而维度表中存放的是大量的具体数据,并且这些数据都是关于某一维的。理论上可以允许任意多个维度表与事实表相关,但由于支持数据仓库的系统不同,存在着数量上的限制和差异。

5 数据仓库的实现

下面以 FoodMark 数据仓库为例来介绍数据仓库的设计和实现。建立 FoodMark 数据仓库的目的在于了解职员的一些情况,分析促销活动对销售的影响以及商场的销售情况(天、月、季度和年)等,从而对市场作预测和决策。在构造 FoodMark 数据仓库时,在操作型数据的基础上建立一个星型数据仓库数据模型,对于事实表于维度表中的数据采用两种不同的方式得到,一种是 DTS Wizard,另一种使用的是 DTS Package Designer。我们结合 FoodMark 数据仓库的情况建立了以1997年事实活动为中心的星型结构,如图5所示。这是数据仓库的数据模型,也是进行数据分析的基础。

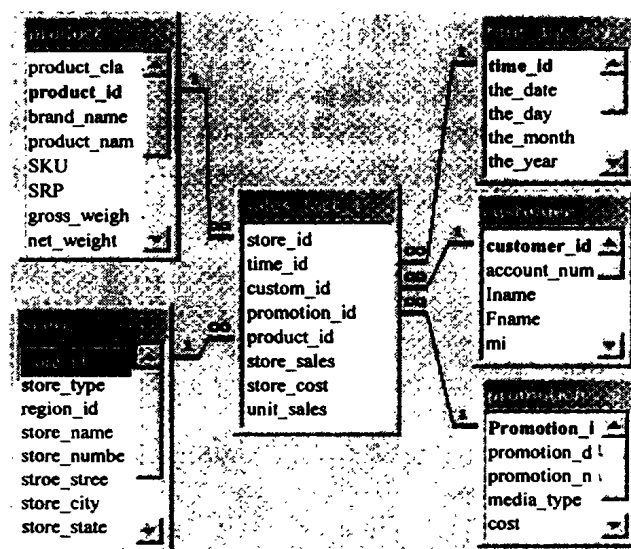


图5 FoodMark 数据仓库的星型连接

5.1 数据分析

在这个星型连接中,sales_fact_1997充当了事实表的角色,其他表都是与它相关的维度表。在这个星型连接的基础上,就可以进行数据分析了,根据对 FoodMark 数据仓库提出的要求,在此以公司每个商店每个季度营业情况(1997年)和公司每季度促销与收入的情况(1997年)两个方面为例,作数据分析介绍。

(1) 公司每个商店每个季度营业情况(1997年)

在这个功能的数据分析中,首先将 store 表、sales_fact_1997 表和 time-by-day 表之间建立联系,使用 order by 和 compute by 来排序和求和,具体代码如下:

```
select
  store_name as storename, [quarter] as 'quarter', store_sales as
  'store sales'
from
  store, sales_fact_1997, time-by-day
where
  (store.store_id=sales_fact_1997.store_id)and
  (time-by-day.time_id=sales_fact_1997.time_id)
order by store_name,[quarter]
compute sum(store_sales) by store_name,[quarter]
compute sum(store_sales) by store_name
```

(下转第126页)

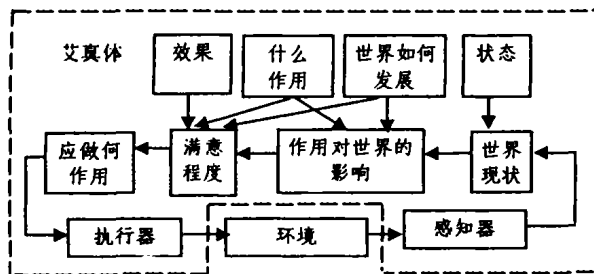


图6 基于效果的艾真体结构

(5) 复合式艾真体 复合式艾真体即在一个艾真体内组合多种相对独立和并行执行的智能形态,其结构包括感知、动作、反应、建模、规划、通信和决策等模块^[13],如图7所示。艾真体通过感知模块来反映现实世界,并对环境信息做出一个抽象,再送到不同的处理模块。若感知到简单或紧急情况,信息被送入反射模块,做出决定,并把动作命令送到行动模块,产生相应的动作。

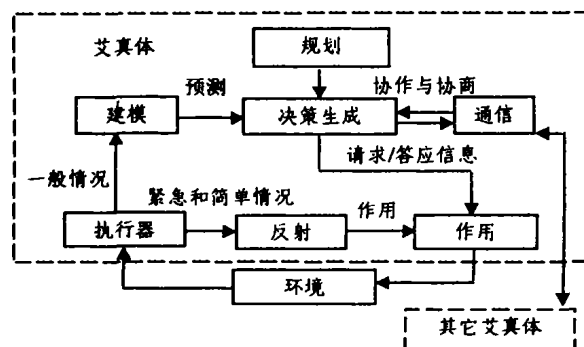


图7 复合式艾真体的结构

结束语 本文所讨论的艾真体和多艾真体系统是分布式人工智能的重要研究方向,具有重要的理论意义和应用价值。特别是当前国内外分布式计算和网络技术的快速发展,使得

艾真体和多艾真体系统具有广阔的应用背景,在工业和商业等领域得到广泛的应用。

文中对 agent 的中文译法给出的建议,即译为艾真体,是在与许多专家和高校师生讨论后得出的。在中文文献中,长期使用一个外文术语是欠妥的;当该术语的出现频度很高时,更是如此。希望并建议能够引起讨论,集思广益,并在此基础上得到共识。

参考文献

- 1 Bode A H, Gasser L. Readings of Distributed Artificial Intelligence. San Francisco: Morgan Kaufmann, 1998
- 2 Mataric M. A Distributed Model for Mobile Robot Environment Learning and Navigation: [Technical Report]. AI Lab, MIT, Cambridge, MA, 1990
- 3 石纯一,王克宏,等. 分布式人工智能进展. 模式识别与人工智能, 1995, 8(s): 73~92
- 4 祝明发. 分布式人工智能. 计算机研究与发展, 1990, 27(10): 7~18
- 5 Nilsson N J. Artificial Intelligence: A New Synthesis. San Francisco: Morgan Kaufmann, 1998
- 6 Lux S D. Understanding cooperation: an agent perspective. In: Proc. of the First Intl. Conf. on Multiagent Systems, San Francisco, 1995
- 7 史忠植. 智能体及其应用. 北京: 科学出版社, 2000
- 8 王怀明, 吴泉源, 高洪奎, 等. 基于 Agent 的分布式计算环境. 计算机学报, 1996, 19(3): 197~201
- 9 Wooldridge M, Jennings N R. Intelligent Agents: Theory and Practice. The Knowledge Engineering Review, 1995, 10(2): 115~152
- 10 Russel S J, Norvig P. Artificial Intelligence, A Modern Approach. New Jersey: Prentice Hall, 1995
- 11 Shoham Y, Tennenholtz M. On Social Laws for Artificial Agent Societies: Off-line Design. Artificial Intelligence, 1995, 73: 231~252
- 12 吴喜明. 基于智能体的分布式智能机器人体系结构的研究: [浙江大学博士学位论文]. 浙江杭州, 1995
- 13 史忠植. 高级人工智能. 北京: 科学出版社, 1998

(上接第119页)

(2) 公司每季度促销与收入的情况(1997年)

在这个功能数据分析中,包括两方面的数据分析,第一个是公司的季度收入情况分析,代码如下:

```
select
[quarter] as 'quarter', store_sales as 'store sales'
from
store_sales_fact_1997, time_by_day
where
(store.store_id=sales_fact_1997.store_id) and
(time_by_day.time_id=sales_fact_1997.time_id)
order by [quarter]
compute sum(store_sales) by [quarter]
```

第二个数据分析是公司每季度的促销情况分析,程序为:

```
select
datepart(qq, end_date) as promotiondate, cost as promotioncost
from promotion
where (datepart(yy, end_date)=1997)
order by datepart(qq, end_date)
compute sum(cost) by datepart(qq, end_date)
```

5.2 程序实现

根据前面的分析,用 Visual Basic 语言程序实现了数据仓库的数据管理和数据分析功能,可以作为数据分析的工具。程序功能体现在两个方面:建立于数据仓库的连接和数据分析,包括 OLTP 数据显示、OLAP 数据分析和 EIS 图表分析等功能。在主程序界面上,分成数据仓库的登陆区域及数据分

析和结果显示区域,程序实现的系统平台是 Microsoft Windows 2000 Server (或 NT),数据库支撑平台是 Microsoft SQL Server Developer。

结束语 本文对数据体系结构的理论进行了介绍和研究,并对数据仓库的具体构建进行了实践探索,这仅是结合实际情况进行的一点应用方面的尝试,数据仓库的应用性很强,数据信息处理的方法和技术也很多,结合实际情况,使用相应的开发工具和软件,可以开发出实用的应用系统。

参考文献

- 1 inmon W H. Building the data warehouse. Cole Publishing Company, 2000
- 2 Amo W C. SQL Server OLAP Developer's Guide. Microsoft Corporation (USA), 1998
- 3 Sperle E. The Enterprise Data Warehouse, Volume I: Planning, Building and Implementation, McGRAW-HILL book Company, 2000
- 4 Hunter P. Microsoft SQL Server 7.0 Data Warehousing, Prentice-Hall, 1999
- 5 Microsoft Corporation. Microsoft SQL Server 简介 (SQL Server, 包含 OLAP 服务). 北京: 微软(中国)有限公司
- 5 萨师煊, 王珊. 数据库系统概论(第三版). 北京: 高等教育出版社出版, 2002