

图书情报与信息管埋实验教材

决策支持系统实验教程

AN EXPERIMENTAL INSTRUCTION
TO DECISION SUPPORT SYSTEMS

陆泉 陈静 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

C934/85

2008

图书情报与信息管理实验教材

决策支持系统实验教程

AN EXPERIMENTAL INSTRUCTION
TO DECISION SUPPORT SYSTEMS

陆泉 陈静 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

决策支持系统实验教程/陆泉,陈静编著. —武汉:武汉大学出版社,
2008.6

图书情报与信息管理实验教材

ISBN 978-7-307-06245-0

I. 决… II. ①陆… ②陈… III. 决策支持系统—高等学校—教材
IV. TP399

中国版本图书馆 CIP 数据核字(2008)第 072351 号

责任编辑:白绍华

责任校对:黄添生

版式设计:马 佳

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北恒泰印务有限公司

开本:720×1000 1/16 印张:10.75 字数:208千字 插页:1

版次:2008年6月第1版 2008年6月第1次印刷

ISBN 978-7-307-06245-0/TP·295 定价:16.00元

版权所有,不得翻印;凡购我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。

内容简介

本书详细介绍了运用已经普及使用的软件环境快速进行决策支持系统开发的主要技术与具体过程。本教程内容以决策支持系统应用开发为核心,分别围绕数据仓库系统开发、商务智能软件包使用以及模型构建与数据分析软件应用进行组织。在介绍 Microsoft SQL Server 2005 与商业智能工具集以及 Microsoft Office Excel 2003 的安装与使用的基础上,设计了 6 个典型实验:数据仓库设计实验、DW/BI 型决策支持系统的系统设计实验、模型构造实验、回归分析实验、数据透视表实验以及使用 OLAP 数据集实验等。这 6 个实验的实验环境门槛低,方便普及;实验任务相对简单,易于掌握;既可独立进行,又可以构成一个完整而实用的决策支持系统,对促进决策支持系统课程教学与技术推广有现实意义。

本书适合作为信息管理与信息系统专业与电子商务专业高年级本科学学生及研究生学习的配套实验教材;同时,也适合作为希望快速掌握决策支持系统与商务智能系统开发工作人员的学习和参考用书。

前言

促进决策支持系统课程教学与技术推广是本书编写的主要动机。目前,由于研究内容的庞杂性、开发方法的多样性和软件环境要求的特殊性,决策支持系统无论在教学科研还是在企业应用中都处于一种“鸡肋”的尴尬地位,即明明很有价值,但是很难普及和推广。

本书力求将现代决策支持系统的理论与方法落到实处,特别强调决策支持系统在社会生产中的快速开发与运用,因此结合企业商业智能系统的开发和使用设计了一套典型的决策支持系统实验,可以让学习者在较短的时间内迅速掌握开发和使用实用的决策支持系统的能力,能够快速适应企业的决策支持系统和商务智能系统环境。

本书详细介绍了运用已经普及使用的软件环境快速进行决策支持系统开发的主要技术与具体过程,对促进决策支持系统课程教学与技术推广有现实意义。本教程内容以决策支持系统应用开发为核心,分别围绕数据仓库系统开发、商务智能软件包使用以及模型构建与数据分析软件应用进行组织。在详细介绍 Microsoft SQL Server 2005 与商业智能工具集以及 Microsoft Office Excel 2003 的安装与使用的基础上,本实验教程包括以下 6 个典型实验:利用 SQL Server 2005 进行数据仓库设计实验,再综合 Microsoft SQL Server 2005 商业智能工具集进行 DW/BI 型决策支持系统的系统设计实验,然后使用 Microsoft Office Excel 2003 进行模型构造实验、回归分析实验、数据透视表实验以及使用 OLAP 数据集实验等。这 6 个实验既可独立进行,又可以构成一个完整而实用的决策支持系统;既适合系统学习,又方便具有某一方面基础的用户进行选择学习。

本书具有 5 个显著特点:将决策支持系统与商务智能系统较好地结合在一起;紧密联系企业决策支持系统的开发与应用;内容系统性强、重点突出;实验环境门槛低、方便普及;实验任务相对简单、易于掌握。在实验设计方面,本书的优势在于:(1) 将决策支持系统的模型与数据分析和商务智能的数据仓库与数据挖掘结合在一起,能快速发挥决策支持系统的支持作用;(2) 在实验中使用的软件工具 SQL Server 2005、Microsoft 商业智能工具集与 Microsoft Office Excel 2003 均为已普及使用的软件,避免了其他与决策支持系统有关的实验教材中必须使用特殊软件包作为实验环境的限制,更易于推广;(3) 实验环境与实验过

程说明详尽，提供了强有力的系统开发指导。

本书不仅适合作为信息管理、信息系统专业与电子商务专业高年级本科学生及研究生学习《决策支持系统》、《商务智能》或其他类似课程的配套实验教材,同时,也适合作为希望快速掌握决策支持系统与商务智能系统开发工作的多层次、多专业人员的学习和参考用书。

本书由武汉大学信息管理学院陆泉与华中师范大学信息管理系陈静两位教师共同编著。其中,陆泉负责全书的内容结构设计和第4、5、7、8、9章的编写,陈静负责第1、2、3、6章的编写,陆泉对初稿进行了最后的审查、修改和完善。衷心感谢武汉大学信息管理学院唐晓波教授对本书给予的建议和指导。此外,武汉大学信息管理学院与武汉大学出版社对此书的编写和出版给予了大力支持,书中还多次引用国内外同行的研究成果,谨此一并致谢。

由于编者时间和水平有限,书中错误或疏漏之处在所难免,敬请专家和广大读者给予批评指正。

作者

2008 年 4 月于珞珈山

目 录

1.1	实验目的与要求	1.1
1.2	实验内容	1.2
1.3	实验操作步骤	1.3
1.3.1	数据库概述	1.3.1
1.3.2	SQL Server 2005 技术简介	1.3.2
1.3.3	安装 SQL Server 2005	1.3.3
1.3.4	Microsoft SQL Server 2005 开发环境	1.3.4
1.4	练习题	1.4
2	使用 Microsoft SQL Server 2005 商业智能工具集	2
2.1	实验目的与要求	2.1
2.2	实验内容	2.2
2.3	实验操作步骤	2.3
2.3.1	商业智能概述	2.3.1
2.3.2	Business Intelligence Development Studio	2.3.2
2.3.3	Business Intelligence Development Studio 中的工具窗口	2.3.3
2.3.4	Business Intelligence Development Studio 中的菜单	2.3.4
2.3.5	使用解决方案和项目	2.3.5
2.3.6	自定义环境、工具和窗口	2.3.6
2.3.7	使用源代码管理服务	2.3.7
2.3.8	配置帮助	2.3.8
2.4	练习题	2.4
3	使用 Microsoft Office Excel 2003	3

3.1	实验目的与要求	31
3.2	实验内容	31
3.3	实验操作步骤	31
3.3.1	Excel 基础	31
3.3.2	Excel 2003 简介	40
3.3.3	安装与删除 Excel 2003	42
3.3.4	数据录入技巧	44
3.3.5	数据处理技巧	50
	练习题	57
4	数据仓库设计实验	58
4.1	实验目的与要求	58
4.2	实验内容	58
4.3	实验操作步骤	58
4.3.1	数据仓库设计的三级数据模型	58
4.3.2	数据仓库设计方法与步骤	60
4.3.3	多维表的数据组织	68
4.3.4	多维表设计	71
4.3.5	创建数据库	73
4.3.6	使用表设计器创建新表	75
4.3.7	创建主键	77
4.3.8	修改外键	78
4.3.9	使用数据库关系图设计器	80
4.3.10	使用查询编辑器	82
	练习题	83
5	DW/BI 型决策支持系统实验	84
5.1	实验目的与要求	84
5.2	实验内容	84
5.3	实验操作步骤	84
5.3.1	数据仓库与决策支持	84
5.3.2	熟悉 Analysis Services	86
5.3.3	使用 Analysis Services 进行开发与决策分析	89
	练习题	99

6 模型构造实验	100
6.1 实验目的与要求	100
6.2 实验内容	100
6.3 实验操作步骤	100
6.3.1 DSS 模型概述	100
6.3.2 数学模型的分类	101
6.3.3 使用 EXCEL 的投资决策函数	104
6.3.4 构建投资指标决策分析模型	116
练习题	118
7 回归分析实验	119
7.1 实验目的与要求	119
7.2 实验内容	119
7.3 实验操作步骤	119
7.3.1 统计分析工具概述	119
7.3.2 使用“数据分析”进行回归分析	122
7.3.3 使用直线回归函数	124
7.3.4 使用预测函数	129
7.3.5 使用指数曲线趋势函数	130
7.3.6 用趋势线进行回归分析	132
练习题	138
8 数据透视表实验	139
8.1 实验目的与要求	139
8.2 实验内容	139
8.3 实验操作步骤	139
8.3.1 数据透视表概述	139
8.3.2 创建数据透视表	141
8.3.3 显示或隐藏数据透视表或数据透视图字段中的项	145
8.3.4 创建数据透视图	148
练习题	151
9 使用 OLAP 数据集实验	152
9.1 实验目的与要求	152
9.2 实验内容	152

001	9.3 实验操作步骤	152
001	9.3.1 OLAP 概述	152
001	9.3.2 Microsoft Excel 中的 OLAP 功能	155
001	9.3.3 访问 OLAP 所需的软件组件	156
001	9.3.4 连接和使用 OLAP 多维数据集	157
101	9.3.5 Excel 可访问的数据源	159
101	练习题	162
011	主要参考文献	163
011		1
011		1.7
011		2.7
011		3.7
011		4.7
011		5.7
011		6.7
011		7.7
011		8.7
011		9.7
011		10.7
011		11.7
011		12.7
011		13.7
011		14.7
011		15.7
011		16.7
011		17.7
011		18.7
011		19.7
011		20.7
011		21.7
011		22.7
011		23.7
011		24.7
011		25.7
011		26.7
011		27.7
011		28.7
011		29.7
011		30.7
011		31.7
011		32.7
011		33.7
011		34.7
011		35.7
011		36.7
011		37.7
011		38.7
011		39.7
011		40.7
011		41.7
011		42.7
011		43.7
011		44.7
011		45.7
011		46.7
011		47.7
011		48.7
011		49.7
011		50.7
011		51.7
011		52.7
011		53.7
011		54.7
011		55.7
011		56.7
011		57.7
011		58.7
011		59.7
011		60.7
011		61.7
011		62.7
011		63.7
011		64.7
011		65.7
011		66.7
011		67.7
011		68.7
011		69.7
011		70.7
011		71.7
011		72.7
011		73.7
011		74.7
011		75.7
011		76.7
011		77.7
011		78.7
011		79.7
011		80.7
011		81.7
011		82.7
011		83.7
011		84.7
011		85.7
011		86.7
011		87.7
011		88.7
011		89.7
011		90.7
011		91.7
011		92.7
011		93.7
011		94.7
011		95.7
011		96.7
011		97.7
011		98.7
011		99.7
011		100.7

1 Microsoft SQL Server 2005的安装与使用

1.1 实验目的与要求

- (1) 了解数据仓库的概念、特点和结构。
- (2) 了解 SQL Server 2005 包含的主要技术。
- (3) 掌握 SQL Server 2005 的安装方法。
- (4) 掌握 SQL Server 2005 的主要开发环境。

1.2 实验内容

- (1) 分析概述部分提供的案例。
- (2) 安装 SQL Server 2005。
- (3) 练习使用 SQL Server Management Studio。

1.3 实验操作步骤

本实验包括以下 4 部分：数据仓库概述、SQL Server 2005 技术简介、安装 SQL Server 2005、Microsoft SQL Server 2005 开发环境。

1.3.1 数据仓库概述

随着计算机技术的飞速发展和企业界不断提出新的需求，数据仓库技术应运而生。传统的数据库技术是以单一的数据资源，即数据库为中心，进行事务处理、批处理到决策分析等各种类型的数据处理工作。然而，不同类型的数据处理有着其不同的处理特点，以单一的数据组织方式进行组织的数据库并不能反映这种差异，满足不了数据处理多样化的要求。近年来，随着计算机应用，特别是数据库应用的广泛普及，人们对数据处理的这种多层次特点有了更清晰的认识。总结起来，当前的数据处理可以大致地划分为两大类：操作型处理和分析型处理（或信息型处理）。操作型处理也叫事务处理，是指对数据库联机的日常操作，通常是对一个或一组记录的查询和修改，主要是为企业的特定应用服务的，人们关心的是响应时间、数据的安全性和完整性。分析型处理则用于管理人员的决策分析。例如，在 DSS 中经常要访问大量的历史数据，当以业务处理为主的联机

事务处理 (OLTP) 应用与以分析处理为主的决策支持系统 (DSS) 应用共存于同一数据库系统中时, 这两种类型的处理发生了明显的冲突。人们逐渐认识到, 事务处理和分析处理具有极不相同的性质, 直接使用事务处理环境来支持 DSS 是行不通的。两者之间的巨大差异使得操作型处理和分析型处理的分离成为必然。这种分离, 划清了数据处理的分析型环境与操作型环境之间的界限, 从而由原来的以单一数据库为中心的数据环境发展为一种新环境: 体系化环境。

1.3.1.1 数据仓库的概念

数据仓库 (data warehouse, DW) 的概念形成是以 Prism Solutions 公司副总裁 W. H. Inmon 在 1992 年出版的《建立数据仓库》(Building the Data Warehouse) 一书为标志的。数据仓库的提出是以关系数据库、并行处理和分布式技术的飞速发展为基础的, 是解决信息技术 (IT) 在发展中存在的拥有大量数据却有用信息贫乏 (Data rich-Information poor) 这一问题的综合解决方案。

从目前的形势看, 数据仓库技术已紧跟 Internet 而上, 成为信息社会中获得企业竞争优势的又一关键。美国 MetaGroup 市场调查机构的资料表明, 《幸福》杂志所列的全球 2000 家大公司中已有 90% 将 Internet 网络和数据仓库这两项技术列入其企业计划, 而且有很多企业为使自己在竞争中处于优势已经率先采用之。

传统数据库用于事务处理, 也叫操作型处理, 是指对数据库联机进行日常操作, 即对一个或一组记录的查询和修改, 主要为企业特定的应用服务。用户关心的是响应时间、数据的安全性和完整性。数据仓库用于决策支持, 也称分析型处理, 用于决策分析, 它是建立决策支持系统 (DSS) 的基础。

操作型数据 (DB 数据) 与分析型数据 (DW 数据) 之间的差别如表 1.1 所示。

表 1.1 操作型数据 (DB 数据) 与分析型数据 (DW 数据) 的区别

DB 数据	DW 数据
细节的	综合或提炼的
存取时准确的	代表过去的数据
可更新的	不更新的
操作需求事先可知道	操作需求事先不知道
事务驱动	分析驱动
面向应用	面向分析
一次操作数据量小	一次操作数据量大
支持日常操作	支持决策需求
生命周期符合 SDLC	不同的生命周期
对性能要求高	对性能要求宽松

1.3.1.2 数据仓库特点

(1) 数据仓库是面向主题的。

与传统数据库面向应用进行数据组织的特点相对应，数据仓库中的数据是面向主题进行组织的。什么是主题呢？首先，主题是一个抽象的概念，是在较高层次上将企业信息系统中的数据综合、归类并进行分析利用的抽象。在逻辑意义上，它对应于企业中某一宏观分析领域所涉及的分析对象。主题是数据归类的标准，每一个主题基本对应一个宏观的分析领域。例如，保险公司的数据仓库的主题为：客户、政策、保险金、索赔等。基于应用数据库的组织则完全不同，它的数据库只是为处理具体应用而组织在一起的。保险公司按照应用组织的数据库是：汽车保险、生命保险、健康保险、伤亡保险等。面向主题的数据组织方式，就是在较高层次上对分析对象的数据的一个完整、一致的描述，能完整、统一地刻画各个分析对象所涉及的企业的各项数据，以及数据之间的联系。所谓较高层次是相对面向应用的数据组织方式而言的，是指按照主题进行数据组织的方式具有更高的数据抽象级别。

需要指出一点，目前数据仓库仍是采用关系数据库技术来实现的，也就是说数据仓库的数据最终也表现为关系。因此，要把握主题和面向主题的概念，需要将它们提高到一个更高的抽象层次上来理解，也就是要特别强调概念的逻辑意义。

为了更好地理解主题与面向主题的概念，说明面向主题的数据组织与传统的面向应用的数据组织方式的不同，在此引用一例：一家采用“会员制”经营方式的商场，按业务已建立起采购、销售、库存管理以及人事管理子系统。按照其业务处理要求，建立了各自的数据库模式：

采购子系统：
 订单（订单号，供应商号，总金额，日期）
 订单细则（订单号，商品号，类别，单价，数量）
 供应商（供应商号，供应商名，地址，电话）

销售子系统：

顾客（顾客号，姓名，性别，年龄，文化程度，地址，电话）
 销售（员工号，顾客号，商品号，数量，单价，日期）

库存管理子系统：
 领料单（领料单号，领料人，商品号，数量，日期）
 进料单（进料单号，订单号，进料人，收料人，日期）

库存（商品号，库房号，库存量，日期）

库房（库房号，仓库管理员，地点，库存商品描述）

人事管理子系统：

员工（员工号，姓名，性别，年龄，文化程度，部门号）

部门（部门号，部门名称，部门主管，电话）

以上述数据模式为例，我们可以看出传统的面向应用的数据组织具有如下特点：

第一，面向应用进行数据组织，是指对企业中相关的组织、部门等进行详细调查和收集数据库的基础数据及其处理的过程。调查的重点是“数据”和“处理”，在进行数据组织时应充分了解企业的部门组织结构，考虑企业各部门的业务活动特点。

第二，面向应用进行数据组织应反映一个企业内数据的动态特征，即它要便于表达企业各部门内的数据流动情况以及部门间的数据输入输出关系，通俗地讲是要表达每个部门的实际业务处理的数据流程，即从哪儿获取输入数据，在部门内进行什么样的数据处理，以及向什么地方输出数据。按照实际应用即业务处理流程来组织数据，其主要目的是通过进行联机事务处理来提高日常业务处理的速度与准确性等。

第三，这种数据组织方式生成的各项数据库模式与企业中实际的业务处理流程中所涉及的单据或文档有很好的对应关系，这种对应关系使得数据库模式具有很强的操作性，因而可以较好地在这类数据库模式上建立起各项实际的应用处理。如库存管理中的领料单、进料单和库存等是实际管理中就存在的单据或报表，并且其各项内容也是相互对应的。在有些应用中，这种数据组织方式只是对企业业务活动所涉及的数据存储介质的改变，即从纸介质到磁介质的转变。

第四，面向应用进行数据组织的方式并没有体现出提出数据库这一概念时的原始意图：把数据与处理分开，即要将数据从数据处理或应用中抽象、解放出来，组织成一个和具体的应用相独立的数据世界。所以说，实际中的数据库建设由于偏重对联机事务处理的支持，无论是在设计方法还是在使用上都将数据应用逻辑与数据于一定程度上又重新捆绑在一起了，造成的后果是使得本来是描述同一客观实体的数据由于与不同的应用逻辑捆绑在一起而变得不统一；使得本来是一个完整的客观实体的数据分散在不同的数据库模式中。

总的来说，面向应用来进行企业数据的组织，其抽象程度还不够高，没有完全实现数据与应用的分离。但是这种方式能较好地数据库模式和企业的现实业务活动对应起来，从而具有很好的操作性，便于实现企业原来的各项业务从手工处理的方式向计算机处理方式的转变。所以在进行 OLTP 数据库系统的开发时，面向应用的数据组织方式也不失为一种有效的数据组织方式，它可以较好地支持联机事务处理。

那么按照面向主题的方式，数据应该怎样来组织呢？数据的组织应该分为两

个步骤：获取主题以及确定每个主题所应包含的数据内容。

前面提到，主题是对应某一分析领域的分析对象，所以主题的抽取，应该是按照分析的要求来确定的。这与按照数据处理或应用的要求来组织数据的主要不同在于同一部门关心的数据内容不同。如在商场中，同样是商品采购，在 OLTP 数据库中，人们所关心的是怎样更方便更快捷地进行“商品采购”这个业务处理；而在进行分析处理时，人们就应该关心同一商品的不同采购渠道。

①在 OLTP 数据库中，在进行数据组织时要考虑如何更好地记录下每一笔采购业务的情况，如我们可以用采购管理子系统中组织的“订单”、“订单细则”以及“供应商”三个数据库模式，来清晰完整地描述一笔采购业务所涉及的数据内容，这就是面向应用来进行数据组织的方式。

②在数据仓库中，由于主要是进行数据分析处理，那么商品采购时的分析活动主要是要了解各供应商的情况，显然“供应商”是采购分析时的分析对象。所以我们并不需要组织像“订单”和“订单细则”这样的数据库模式，因为它们包含的是纯操作性的数据；但是仅仅只用 OLTP 数据库的“供应商”中的数据又是不够的，因而要重新组织“供应商”这么一个主题。

概括各种分析领域的分析对象，我们可以综合得到其他的主题。仍以商场为例子，它所应有的主题包括：商品、供应商、顾客等。每个主题有着各自独立的逻辑内涵，对应了一个分析对象。这三个主题所应包含的内容列出如下：

商品：

商品固有信息（商品号、商品名、类别、颜色等）

商品采购信息（商品号、供应商号、供应价、供应日期、供应量等）

商品销售信息（商品号、顾客号、售价、销售日期、销售量等）

商品库存信息（商品号、库房号、库存量、日期等）

供应商：

供应商固有信息（供应商号、供应商名、地址、电话等）

供应商品信息（供应商号、商品号、供应价、供应日期、供应量等）

顾客：

顾客固有信息（顾客号、顾客名、性别、年龄、文化程度、住址、电话等）

顾客购物信息（顾客号、商品号、售价、购买日期、购买量等）

我们以“商品”主题为例，关于商品的各種信息已综合在“商品”这一个主题中，主要是两个方面的内容：第一，它包含了商品固有信息，如商品名称、商品类别以及型号、颜色等商品的描述信息；第二，“商品”主题也包含有商品流动的信息，如“商品”主题也描述了某商品采购信息、商品销售信息及商品库存信息等。比照商场原有数据库的数据模式，可以看到：首先，在从面向应用到面

向主题的转变过程中, 丢弃了原来不必要的、不适于分析的信息, 如有关订单信息、领料单等内容就不再出现在主题中。其次, 在原有的数据库模式中, 关于商品的信息分散在各子系统中, 如商品的采购信息存在采购子系统中, 商品的销售信息则存在于销售子系统中, 商品库存信息却又在库存管理子系统中, 根本没有形成一个有关商品的完整一致的描述: 面向主题的数据组织方式所强调的就是要形成关于商品的一致信息集合, 以便在此基础上针对“商品”这一分析对象进行分析处理。

值得注意的是, 不同的主题之间也会有一些内容的重叠。这种重叠是逻辑上的重叠, 而不是同一数据内容在物理上的重复储存; 主题之间的重叠一般是在细节级上的重叠, 因为不同主题中的综合方式是不同的。

总结起来, 面向主题的数据组织方式是根据分析要求将数据组织成一个完备的分析领域即主题域。主题域具有以下特性:

①独立性。如针对商品进行的各种分析所要求的是“商品”主题域, 这一主题域可以和其他的主题域有交叉部分, 但它必须具有独立内涵, 即要求有明确的界限, 规定某项数据是否该属于“商品”主题。

②完备性。就是要求对任何一个商品的分析处理要求, 我们应该能在“商品”这一主题内找到该分析处理所要求的一切内容; 如果对商品的某一分析处理要求涉及现存“商品”主题之外的数据, 那么就应当将这些数据增加到“商品”主题中来, 从而逐步完善“商品”主题。或许有人担心, 要求主题的完备性会使得主题包含有过多的数据项而显得过于庞大。这种担心是完全不必要的, 因为主题只是一个逻辑上的概念, 实现时, 如果主题的数据项多了, 可以采取各种划分策略来化大为小。

主题是一个在较高层次上对数据的抽象, 这使得面向主题的数据组织可以独立于数据的处理逻辑, 因而可以在这种数据环境上方便地开发新的分析型应用。同时, 这种独立性也是建设企业全局数据库所要求的, 所以面向主题不仅是适用于分析型数据环境的数据组织方式, 而且是适用于建设企业全局数据库的数据组织方式。

(2) 数据仓库是集成的。

数据仓库的数据是从原有的分散的数据库数据中抽取来的。在前面我们已经看到, 操作型数据与 DSS 分析型数据之间差别甚大。第一, 数据仓库的每一个主题所对应的源数据在原有的各分散数据库中有许多重复和不一致的地方, 且来源于不同的联机系统的数据都和不同的应用逻辑捆绑在一起; 第二, 数据仓库中的综合数据不能从原有的数据库系统直接得到。

数据进入数据仓库之前, 必须经过加工与集成。对不同的数据来源要统一数据结构 and 编码, 统一原始数据中的所有矛盾之处, 如字段的同名异义、异名同

义、单位不统一、字长不一致等。总之,要将原始数据结构进行一个从面向应用到面向主题的大转变,所要完成的工作包括:

①要统一元数据中所有矛盾之处。如字段的同名异义、异名同义、单位不统一、字长不一致,等。

②进行数据综合和计算。数据仓库中的数据综合工作可以在从原有数据库抽取数据时生成,但许多是在数据仓库内部生成的,即进入数据仓库以后进行综合生成的。

(3) 数据仓库是稳定的。

数据仓库包括了大量的历史数据。数据经集成进入数据仓库后是极少或根本不更新的。数据仓库的数据主要供企业决策分析之用,所涉及的数据操作主要是数据查询,一般情况下并不进行修改操作。数据仓库的数据反映的是一段相当长的时间内历史数据的内容,是不同时点的数据库快照的集合,以及基于这些快照进行统计、综合和重组的导出数据,而不是联机处理的数据。数据库中进行联机处理的数据经过集成输入到数据仓库中,一旦数据仓库存放的数据已经超过数据仓库的数据存储期限,这些数据将从当前的数据仓库中删去。因为数据仓库只进行数据查询操作,所以数据仓库管理系统 DWMS 相比 DBMS 而言要简单得多。DBMS 中许多技术难点,如完整性保护、并发控制等,在数据仓库的管理中几乎可以省去。但是由于数据仓库的查询数据量往往很大,所以就对数据查询提出了更高的要求,它要求采用各种复杂的索引技术;同时由于数据仓库面向的是企业的高层管理者,他们会对数据查询的界面友好性和数据表示提出更高的要求。

(4) 数据仓库是随时间变化的。

数据仓库中的数据不可更新是针对应用来说的,也就是说,数据仓库的用户进行分析处理时是不进行数据更新操作的。但并不是说,在从数据集成输入数据仓库开始到最终被删除的整个数据生存周期中,所有的数据仓库数据都是永远不变的。

数据仓库的数据是随时间的变化不断变化的,这是数据仓库数据的第四个特征,表现在以下三方面:

①数据仓库随时间变化不断增加新的数据内容。数据仓库系统必须不断捕捉数据库中变化的数据,追加到数据仓库中去,也就是要不断地生成 OLTP 数据库的快照,经统一集成后增加到数据仓库中去;但对于每次的数据库快照确实是不再变化的,捕捉到新的变化数据,只不过又生成一个数据库的快照增加进去,而不会对原来的数据库快照进行修改。

②数据仓库随时间变化不断删去旧的数据内容。数据仓库的数据也有存储期限,只要超过了这一期限,过期数据就要被删除。只是数据仓库内的数据时限要远远长于操作型环境中的数据时限。在操作型环境中一般只保存有 60~90 天的