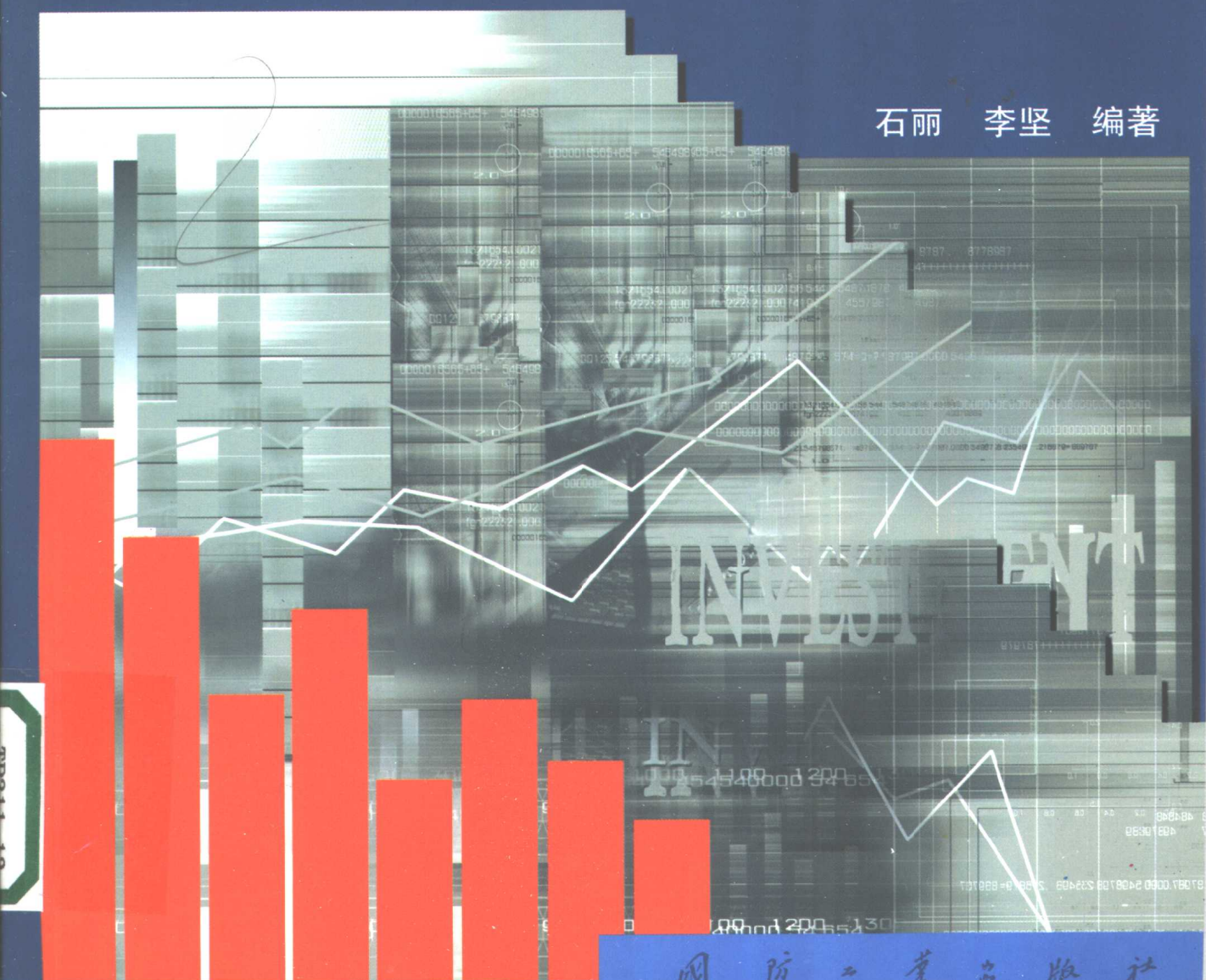


SHUJU CANGKU

数据仓库与决策支持

YU JUECE ZHICHI

石丽 李坚 编著



国防工业出版社

数据仓库与决策支持

石丽 李坚 编著

·

国防工业出版社

·北京·

内 容 简 介

本书从系统工程的角度全面系统地论述了数据仓库的理论与技术、决策支持与数据仓库的相互关系和如何构建基于数据仓库的决策支持系统。其内容分为3部分:第1部分介绍了数据库技术,并阐述了数据库与数据仓库的关系以及数据仓库产生和发展的必然性。第2部分详细而全面地介绍了数据仓库技术,从系统工程的角度全面地讲述了数据仓库与有关技术的使用与结合,并将数据仓库与决策支持作为统一体,较清楚地讲述了如何构建企业的数据仓库系统。第3部分介绍了实际应用实例,结合作者的科研工作,给出实际例子。本书主要特点,不是单纯介绍技术和概念,而是从工程的角度,用系统工程的思维方式,将自己实际科研工作中的真实体会告诉读者,具有一定的参考价值。

本书可作为数据仓库和决策研究人员的参考书,并作为研究生和相关专业高年级大学生的教材。

图书在版编目(CIP)数据

数据仓库与决策支持/石丽,李坚编著. —北京:国防工业出版社,2003.5
ISBN 7-118-03119-4

I. 数... II. ①石... ②李... III. ①数据库系统
②决策支持系统 IV. ①TP311.13 ②TP399

中国版本图书馆 CIP 数据核字(2003)第 019307 号

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

北京奥隆印刷厂印刷

新华书店经售

*

开本 787×1092 1/16 印张 13¼ 308 千字

2003 年 5 月第 1 版 2003 年 5 月北京第 1 次印刷

印数:1—4000 册 定价:19.00 元

(本书如有印装错误,我社负责调换)

前 言

数据仓库是近年来兴起的一种新的数据库应用。各大数据库厂商纷纷宣布自己的产品支持数据仓库,并提出一整套用以建立和使用数据仓库的方案。比如 Informix Gong-side 公司的数据仓库解决方案;ORACLE 公司的数据仓库解决方案;Sybase 公司的交互式数据仓库解决方案等。这些发展引起了学术界的极大兴趣,国际上许多重要的学术会议,如超大型数据库(VLDB)国际会议,数据工程(Data Engineering)国际会议等,都出现了专门研究数据仓库(Data Warehousing,简记为 DW)、联机分析处理(On-Line Analytical Processing,简记为 OLAP)、数据挖掘(Data Mining,简记为 DM)的论文。数据仓库虽然起源于 20 世纪 90 年代,但发展却非常迅速,特别是网络技术的飞速发展,为数据仓库的发展提供了更宽阔的数据跑道。中国的数据库应用已有 20 多年的历史,数据的积累和沉淀为数据仓库的发展提供了很好的数据平台,如何将数据演变成可以科学决策的信息是管理者要考虑的问题。对我国许多企业而言,在建立或发展自己的决策信息系统时,必须清楚以下的问题:在原有的数据库上建立数据仓库的原因、数据仓库能否代替传统的数据库和怎样建立数据仓库等。

随着各种计算机技术,如数据模型、数据库技术和应用开发技术的不断进步,数据仓库技术也在不断发展,并在实际应用中发挥了巨大的作用。因特网资料中心在 1996 年对 20 世纪 90 年代前期进行的 62 个数据仓库项目的调查结果表明:进行数据仓库项目开发的公司在平均 2.73 年的时间内获得了平均为 321% 的投资回报率。使用数据仓库所产生的巨大效益同时又刺激了对数据仓库技术的需求,数据仓库市场正以迅猛势头向前发展:一方面,数据仓库市场需求量越来越大,每年约以 400% 的速度扩张;另一方面,数据仓库产品越来越成熟,生产数据仓库工具的厂家也越来越多。

数据仓库知识作为生产力必不可少的信息商品已经成为,并将继续成为世界范围权力竞争的一个焦点,也许是唯一的焦点。可以预见,总有一天,国与国之间会为了控制信息而发动战争,就像他们在过去争夺领土,进而为了获得和利用自然资源和廉价的劳动力而发动战争一样。遍及世界的公司正试图利用多年来存储在他们计算机内的信息,建立适合自己企业的决策支持系统,而建立数据仓库是科学决策的前提。在未来大规模定制经济环境下,数据仓库将成为企业获得竞争优势的关键武器。

本书是在实际工作的基础上,对数据仓库的发展、建设提出了许多建设性的意见。并试图尽可能地反映目前最新的技术。为了使读者对数据仓库有全面的了解,本书共分 3 个部分。第 1 部分:数据库技术,主要介绍数据库的起源,与数据仓库的关系。第 2 部分:数据仓库技术,较全面地介绍了数据仓库的发展与设计方法。第 3 部分:系统应用举例,介绍了实际应用的范例,并结合自己的科研工作,给出了电力系统数据仓库的解决方案。本书主要特点是从工程设计的角度,用系统工程的思维方式,将自己实际科研工作中的真

IV

实体会告诉读者,具有一定的参考价值。

本书面向的主要读者是计算机应用、信息系统、计算机科学、科学管理等专业大学本科高年级学生及硕士研究生。本书由石丽负责全文的编写与统稿工作,李坚参与编写第3部分,周龙君负责图形和表格,杨猛参与绘制了部分图表。

由于作者水平与时间有限,不当和疏漏之处,恳请专家与读者批评指正。

作 者

2003 年 2 月

目 录

第 1 部分 数据库技术

第 1 章 数据库概述	1
1.1 数据库系统的概念	1
1.2 数据库系统的发展	2
1.3 数据库的研究内容	3
1.3.1 数据库理论	3
1.3.2 数据模型	3
1.3.3 数据库语言	3
1.3.4 数据的安全性	4
1.3.5 事务管理	4
1.4 数据库研究的难点与展望	4
1.4.1 深度(智能化)	4
1.4.2 广度(多媒体)	4
1.4.3 宿主语言与查询语言的集成	5
1.4.4 各种数据库间的数据转换	5
1.5 数据库性能评价	5
第 2 章 数据库结构模型	6
2.1 客户机和服务器计算机的作用	6
2.2 客户机与服务器功能比较	7
第 3 章 SQL 语言	8
3.1 SQL 语言简介	8
3.2 SQL 语言及其优点	8
3.2.1 SQL 语言是非过程化语言	8
3.2.2 SQL 是统一的语言	9
3.2.3 SQL 是所有关系数据库的公共语言	9
3.3 SQL 语言的分类及语法	9
第 4 章 ODBC 与 ADO 技术	12
4.1 访问数据库的方式	12
4.2 ODBC 的定义	13
4.3 配置 ODBC 数据源	14

KJS/09/01

4.4 用 ADO 实现访问数据库	17
第 5 章 ASP 技术	19
5.1 ASP 简介	19
5.2 ASP 组件	20
第 6 章 数据库工具简介	24
6.1 Visual Basic 简介	24
6.2 Power Builder 简介	26
6.3 SQL Server 简介	28
6.3.1 SQL Server 特点	28
6.3.2 SQL Server 结构	29
第 7 章 数据库在数据仓库中的重要作用	33
7.1 从数据库到数据仓库	33
7.2 数据库与数据仓库的区别	35

第 2 部分 数据仓库技术

第 8 章 数据仓库的基础知识	38
8.1 数据仓库起源	38
8.2 数据仓库的技术趋势	40
8.3 数据仓库与网络技术	42
8.4 商务系统在因特网的应用	45
8.5 数据仓库及市场预测	46
8.6 建立数据仓库学科的必要性	46
第 9 章 数据仓库的有关概念	50
9.1 什么叫数据仓库	51
9.2 数据仓库数据组织	54
9.2.1 数据仓库的数据组织结构	54
9.2.2 粒度与分割	55
9.2.3 数据仓库的数据组织形式	55
9.2.4 数据仓库的数据追加	55
9.3 与数据仓库有关的名词解释	56
9.3.1 基本概念	56
9.3.2 关键技术	58
9.4 数据仓库系统的 3 个工具层	61
9.5 数据仓库平台的评测指标	62
第 10 章 数据仓库与数据集市	64
10.1 为什么需要数据集市	64
10.2 数据集市的概念	65

10.3 数据仓库与数据集市的区别	66
10.4 操作系统与数据仓库系统	69
10.4.1 操作系统	69
10.4.2 数据仓库系统	69
10.4.3 操作系统与数据仓库系统间的差别	70
10.4.4 操作系统与数据仓库系统的关系	70
第 11 章 数据仓库的组成	72
11.1 数据仓库的基本体系	72
11.1.1 数据源	73
11.1.2 创建数据准备区	74
11.1.3 数据抽取	75
11.1.4 数据刷新	77
11.1.5 元数据	78
11.1.6 数据转换和集成	78
11.2 数据仓库数据库	81
11.3 数据仓库数据的访问	81
11.4 查询工具和显示服务	83
11.5 应用程序编程接口	84
11.6 最终用户分析	84
11.7 数据仓库的用户	84
第 12 章 建立数据仓库的原则	85
12.1 系统的响应性是考评数据仓库优劣的主要因素	86
12.2 数据仓库系统要有较好的开放性	88
12.2.1 共享数据原则	88
12.2.2 实现统一的数据源原则	88
12.2.3 信息打包	89
12.3 系统要有很好的安全性	90
12.3.1 数据仓库的技术体系结构	90
12.3.2 数据获取模块的安全检查	91
第 13 章 构建基于数据仓库的 DSS	96
13.1 管理和决策制定	96
13.2 现代企业决策的挑战	97
13.3 决策支持系统的概念	98
13.3.1 决策支持系统的组件	99
13.3.2 决策支持系统的应用特点	100
13.4 决策支持系统的主要应用	100
13.5 企业管理信息决策分析系统的设计	105
13.5.1 传统决策系统的缺点	105
13.5.2 DSS 因数据仓库而实现了突破	106

13.5.3 基于数据仓库的 DSS 决策可以解决的问题	106
13.5.4 建立信息分析决策系统可增加企业的竞争优势	107
第 14 章 OLAP 技术	108
14.1 OLAP 概念	108
14.2 OLAP 的功能与实现结构	109
14.2.1 OLAP 的功能	109
14.2.2 存储结构	110
14.3 OLAP 技术的优点	110
14.4 OLAP 的多维数据概念	111
14.4.1 维	112
14.4.2 多维性	113
14.4.3 多维分析的基本分析动作	114
14.4.4 活动数据的存储及处理	114
14.4.5 多维数据库	115
14.5 OLAP 的数据模型及实现	116
14.5.1 数据模型的分类	116
14.5.2 表的分类	117
14.5.3 模型实现	118
14.5.4 物化视图	118
14.6 数据仓库和 OLAP 的决策支持技术	119
第 15 章 数据挖掘(DM)技术	120
15.1 数据挖掘技术概述	120
15.2 数据挖掘研究现状	121
15.3 数据挖掘的范围	122
15.4 数据挖掘的挖掘任务和挖掘方法	123
15.5 数据挖掘的概念	126
15.6 数据挖掘和知识发现的几种常用方法	127
15.6.1 基于数据仓库的数据挖掘	128
15.6.2 知识发现的常用方法	128
15.6.3 数据挖掘的分析方法	129
15.7 数据挖掘的数据分析过程	130
15.8 数据挖掘工具的评价标准	130
15.9 数据挖掘的优点和缺点	131
第 16 章 如何建立数据仓库系统	133
16.1 企业数据仓库建设的需求	133
16.2 数据仓库建立的步骤	138
16.2.1 数据仓库设计	138
16.2.2 事实表和维度表的设计	139
16.3 企业数据仓库系统的总体结构	145

16.4 BI(商业智能).....	147
第 17 章 各公司的解决方案及主要技术介绍	149
17.1 IBM	149
17.2 Oracle	149
17.3 Sybase	150
17.4 Informix	150
17.5 CA	151
17.6 NCR Teradata	151
17.7 Microsoft	151
17.8 SAS	154
17.9 Platinum Technology	154

第 3 部分 系统应用举例

第 18 章 产品数据管理	158
18.1 产品数据管理(PDM)的发展	158
18.2 采用数据仓库技术的产品数据管理支持决策.....	159
18.2.1 决策生产何种类型的产品.....	159
18.2.2 系统设计方案选择的决策.....	160
18.2.3 企业全局决策.....	160
第 19 章 电力系统数据仓库系统设计	163
19.1 项目背景.....	163
19.2 系统功能概述.....	164
19.2.1 目标分析.....	164
19.2.2 系统配置.....	165
19.3 数据仓库的方案设计.....	165
19.4 系统模块分析与设计.....	167
19.4.1 模块划分.....	167
19.4.2 系统网络图设计.....	168
19.5 系统设计与实现.....	170
19.5.1 系统功能实现.....	170
19.5.2 系统数据库设计.....	177
19.5.3 构建数据仓库.....	183
19.6 系统特点.....	196
19.7 系统运行报告.....	197
19.8 总结.....	198
第 20 章 电力系统关键技术简介	199
20.1 面向对象编程技术.....	199

20.1.1	面向对象的程序设计简介.....	199
20.1.2	如何面向对象编程.....	200
20.2	OLE 技术	201
20.2.1	建立和设置 OLE 控件	202
20.2.2	激活修改 Windows 画板中的 OLE 对象	202
20.2.3	嵌入和链接 OLE 控件	202
20.2.4	OLE 控件的激活方式	203
20.2.5	设置和插入 OLE 对象	203
20.2.6	OLE 存储	203
20.3	预测技术.....	203
20.3.1	预测方法.....	204
20.3.2	一元线性回归模型.....	204
参考文献.....		208

第 1 部分 数据库技术

第 1 章 数据库概述

1.1 数据库系统的概念

数据库系统是一个实际可运行的软件系统，是存储介质、处理对象和管理系统的集合体，其目的是为存储、维护和应用系统提供数据。它通常由软件、数据库和数据管理员组成。其软件主要包括操作系统、各种宿主语言、实用程序以及数据库管理系统。数据库是依照某种数据模型组织起来并存放在二级存储器中的数据集合。这些数据为多个应用服务，独立于具体的应用程序。数据库由数据库管理系统统一管理，数据的插入、修改和检索均要通过数据库管理系统进行。数据库管理系统是一种系统软件，它的主要功能是维护数据库并有效地访问数据库中任意部分的数据。对数据库的维护包括：保持数据的完整性、一致性和安全性。数据管理员负责创建、监控和维护整个数据库，使数据能被任何有权使用的人有效使用。数据库管理员一般由业务水平较高、资历较深的人员担任。

数据库系统的个体含义是指一个具体的数据库管理系统软件和用它建立起来的数据库；它的学科含义是指研究、开发、建立、维护和应用数据库系统所涉及的理论、方法、技术所构成的学科。在这一含义下，数据库系统是软件研究领域的一个重要分支，常称为数据库领域。数据库研究跨越于计算机应用、系统软件和计算机理论三个领域，其应用促进新系统的研制开发，新系统带来新的理论研究，而新的理论研究又对前两个领域起着指导作用。

数据库系统的出现是计算机应用的一个里程碑，它使计算机应用从以科学计算为主转向以数据处理为主，从而使计算机在各行各业乃至家庭中普遍使用。在数据库系统之前的文件系统虽然也能处理持久数据，但是文件系统不提供对任意部分数据的快速访问，而这对数据量不断增大的应用来说是至关重要的。为了实现对任意部分数据的快速访问，就要研究许多优化技术，这些优化技术往往很复杂，是普通用户难以实现的，所以就由系统软件（数据库管理系统）来完成，而提供给用户的是简单易用的数据库语言。由于对数据库的操作都由数据库管理系统完成，所以数据库可以独立于具体的应用程序而存在，使得数据库可以为多个用户共享。因此，数据的独立性和共享性是数据库系统的重要特征。数据共享节省了大量人力、物力，为数据库系统的广泛应用奠定了基础。数据库系统的出现使得普通用户能够方便地将日常数据存入计算机并在需要的时候快速访问

它们，从而使计算机走出科研机构，进入各行各业，进入家庭。

1.2 数据库系统的发展

1969 年，美国的 IBM 公司开发了第一个数据库系统 IMS。这是一个层次数据库系统，在数据库系统发展史上有着重要的地位。同年，美国的数据系统语言委员会下属的数据库任务组提出了著名的 DBTG 报告，并在 1970 年提出了该报告的修订版。这份报告定义了数据库操纵语言、模式定义语言和子模式定义语言的概念。数据库操纵语言用于编写操纵概念视图的应用程序，模式定义语言用来编写概念视图和内部视图相结合的模式程序。在 20 世纪 70 年代，开发了许多遵循 DBTG 报告的网状数据库系统，如：IDMS，IDS，DMSII 等。20 世纪 70 年代初，E.F.Codd 提出了关系数据模型的概念，提出了关系代数和关系演算，使关系数据库从理论到实践都取得了辉煌的成果。在理论上，确立了完整的理论、数据依赖理论以及关系数据库的设计理论等；在实践上，开发了许多著名的关系数据库系统，如：system R，INGRES，ORACLE 等。1986 年，美国国家标准协会（ANSI）通过了关系数据库查询语言 SQL 的文本标准。进入 20 世纪 80 年代以后，随着计算机硬件技术的提高，使得计算机应用不断深入，从而产生了许多新的应用领域，如：计算机辅助设计、计算机辅助教学、计算机辅助制造、计算机辅助工程、计算机集成制造、办公自动化、物理信息处理、智能信息处理等。这些新的应用领域对数据库系统提出了新要求。由于没能设计出一个统一的数据模型来表示这些新型数据及其相互联系，所以出现了百家争鸣的局面，产生了演绎数据库（逻辑数据库，知识库）、面向对象数据库、工程数据库、时态数据库、地理数据库、模糊数据库、积极数据库等新型数据库的研究。到 20 世纪 80 年代后期和 20 世纪 90 年代初期，出现了面向对象数据库系统，如：GemStone，VBASE，ORION，Iris 等。但是，到目前为止，真正的新一代数据库系统还没有出现。

数据管理是数据库的核心任务，内容包括对数据的分类、组织、编码、存储、检索和维护等。

数据库技术的发展可分为 3 个阶段。

第一代数据库管理系统：层次和网状数据库管理系统（DBMS）。

层次 DBMS 是数据库系统的先驱。层次 DBMS 以树型结构存储数据，它假定数据之间的父子关系，树的顶端称为根，它可能有很多后代，后代又有任意多个下一代，层次数据库现已消失。

网络数据库是数据库概念、方法、技术的奠基。网络 DBMS 以记录和链接的形式存储数据，这样的系统使多对多的关系比层次 DBMS 更灵活，网络 DBMS 非常快而且存储效率很高，网络 DBMS 允许使用复杂的数据结构，但非常不方便，并要求繁重冗长的设计。

第二代数据库管理系统：关系数据库管理系统（RDBMS）。

RDBMS 具备数据库可能有的最简单结构，在 RDBMS 中，数据被组织成表，而表又由记录组成，记录由字段（域）组成，每个字段对应一个数据项。如果有一个或多个共同字段，则两个以上的表可能链接在一起。RDBMS 易于使用，目前非常流行。

第三代数据库管理系统：面向对象数据库管理系统。

面向对象的 DBMS 用于处理数字与词之类的数据，近几年，已出现面向对象的 DBMS，这些系统可以处理视频、图像、图片等对象。

数据库管理系统是为数据库的建立、使用和维护而配置的软件，它建立在操作系统之上，对数据进行统一的管理和控制。数据仓库是数据库发展到一定阶段的必然产物，它和数据库有紧密的联系，又有重要的区别，数据库技术的发展会推动数据仓库技术的发展，同时数据库又是数据仓库的主要引擎。

1.3 数据库的研究内容

数据库系统作为一门学科，其主要的研究内容为：数据库理论、数据库模型、数据库语言、查询处理及其优化技术、数据的安全性（存取控制、可恢复性）和事务管理（并发控制）。

1.3.1 数据库理论

数据库理论的主要内容为关系数据库理论（依赖理论、泛关系理论、超图理论等）、事务理论、逻辑与数据库理论和面向对象数据库理论。关系数据库理论开始于 E.F.Codd 在 1970 年所写的论文。数据依赖是定义在关系上的约束条件，或者说数据依赖用于定义合法的数据库，以维护数据的完整性和一致性。泛关系理论将数据库中的所有关系都看作为包含所有属性的大关系的投影，这隐含了这样的假设——脱离具体的关系而讨论属性是没有意义的。泛关系思想为关系模式规范化提供了基础，而规范化是关系数据库设计的依据。但是，泛关系数据库系统的含义是：用户对数据库的操作都是在泛关系上进行的。到目前为止，还没有出现商品化的泛关系数据库系统。超图理论将数据库模式描述为超图，其主要目的为研究有效的查询处理算法，如：把无环超图用于分布式数据库的查询优化。事务理论的研究内容是如何维护数据的一致性。某些操作被意外中断后会造成数据的不一致，如同一数据在某关系中作了修改而在另一关系中却没改，为了避免这种情况，引入了事务。一个事务是一组数据库操作命令，它们或者没有执行，或者全部执行完毕。在有多用户同时访问数据库的情况下，就要考虑并发控制，如二段加锁、事务的串行化。逻辑与数据库理论主要研究如何将逻辑程序设计技术与数据库技术有机结合，如演绎数据库系统的研究。面向对象数据库理论主要处理大规模的复杂对象。

1.3.2 数据模型

任何一个数据库管理系统都至少提供一种数据模型，因此数据模型是数据库研究的基础。根据某种数据模型，人们可以用数据世界来合理表示现实世界的某一部分，并且将数据世界映照成一个意识世界（用户界面）。数据模型有 2 方面含义：数据以何种形式存储；用户以何种形式看待数据。常见的数据模型有层次模型、网状模型、关系模型、逻辑模型、实体联系（E-R）模型和面向对象模型等。

1.3.3 数据库语言

与一般的程序设计语言不同，在数据库语言中，描述性部分和过程性部分是分开的。

其过程性部分是一个通用的程序设计语言，称为宿主语言；而描述性部分包括数据定义语言和数据操纵语言。数据定义语言用于说明数据库的逻辑模式；数据操纵语言，亦称为查询语言，用于说明对数据库的操作。为了提高对数据库操作的效率，采用了大量的查询优化技术。从而查询处理及其优化技术的研究就成为数据库研究的重要内容。这方面的工作主要包括索引技术和连接技术。对传统的数据库而言，这两项技术已趋完善。由于数据库查询语言和宿主语言之间存在阻抗不匹配问题，所以在新型数据库系统中（如：面向对象数据库系统和知识库系统），倾向于两者的有机集成，构成一个数据库程序设计语言或持久性程序设计语言。

1.3.4 数据的安全性

数据的安全性（存取控制、可恢复性）是指数据不被非法使用、在意外事件中不被破坏或丢失，这是存取控制和可恢复性的研究内容。存取控制的通常做法是为不同用户设置不同的数据存取特权并设立视图机制，使得每个用户只能访问到允许他访问的数据。可恢复性是指在意外事件（软件或硬件方面）破坏了当前数据库状态后，系统有能力恢复数据库，使损失减少到最低限度。数据恢复采用的方法通常是建立日志和经常性地作数据库的备份。

1.3.5 事务管理

在多用户共享的系统中，许多事务可能同时对同一数据进行操作（并发操作），这样数据库的完整性就可能遭到破坏。主要有丢失更新问题、不一致分析问题和尚未提交的更新问题等。因此，要对事务进行管理、控制并发操作，其基本作法是对数据实行加锁及事务调度。

1.4 数据库研究的难点与展望

1.4.1 深度（智能化）

计算机科学的主要目标是使计算机与人的界面尽量靠近人，因此，要尽量提高计算机的智能水平。智能化是计算机科学各个分支的研究前沿。在数据库方面，智能化的工作是将人工智能技术与数据库技术相结合，即演绎数据库、知识库研究。目前的主要困难在于递归查询处理无法取得满意的性能，硬件技术的革命（大内存、并行机、高速存取的外存储器）将是提高知识库查询效率的重要因素。

1.4.2 广度（多媒体）

多媒体数据处理的困难很多，即使是对于一般的复杂对象，目前也还不能很好地处理。多媒体数据的建模、存储和多媒体数据库的查询及查询处理等研究都很困难。信息分布方面主要是分布式数据库系统的研究。分布式数据库从 20 世纪 70 年代开始研究，但是一直没有出现商品化的分布式数据库系统，这说明了它的难度。当前比较好的具有数据分布特征的数据库管理系统是客户/服务器体系结构的系统（如：Sybase, ORACLE 7

等)。但新的计算机应用又对它提出了新的要求,智能化、新型事务模型、多媒体数据的处理、高速信息通信、数据源的高度透明性等将是新型的分布式数据库系统的重要研究内容。

1.4.3 宿主语言与查询语言的集成

查询语言与宿主语言之间的“阻抗不匹配”问题,长期以来一直困扰着“数据库操纵语言+宿主语言”形式的数据库系统。阻抗不匹配问题主要是指二者支持不同的编程风范和数据类型。编程风范的不匹配将导致用户必须学会使用两种完全不同的语言,并领会二者之间的连接规则,这对用户是一个负担。数据类型的不匹配将导致用户必须处理不同类型数据间的转换,并且这种转换的工作将由系统的接口程序来实现,这对于数据密集型应用来说也是一个沉重的负担。有机地将操纵语言与宿主语言集成是知识库系统和面向对象数据库系统研究的初始原因之一,但问题还远没有解决。

1.4.4 各种数据库间的数据转换

数据库之间的转换有 2 方面的需要:数据库管理系统软件的更新和计算机网络上不同数据库间的数据共享。当数据库管理系统软件更新时,希望将原先的数据库直接转到新系统中来,以保护以前的投资。在计算机网络上,常常运行着多种类型的数据库系统,它们是在不同时间、不同地点建立的,联网后要保护这些投资,所以要进行数据库转换。

1.5 数据库性能评价

一个好的数据库应该是:数据冗余尽可能少、数据库访问效率尽可能高并且数据库易于维护。这涉及到数据语义和许多人为因素,因此很难对其进行定量地评价。找到良好的评价方法或衡量模型将产生好的数据库设计方法,因而具有重大的意义。

到目前为止,数据库设计与应用程序设计是分离的,并且在具体的应用系统开发中,往往由两个小组各行其事。其后果是不得不放弃其中一种优良的设计,这将导致系统性能低下或用户要求无法满足。面向对象技术是二者结合的有效手段,但是,良好地结合还需要很长时间的努力。

需要处理的数据越来越庞大、计算机网络越来越复杂、系统的智能水平越来越高是计算机系统发展的总趋势。因此,未来的信息管理系统的特征将是处理复杂对象、数据分布和智能化。在复杂对象的处理方面,面向对象数据库、多媒体数据库将会由于广泛的应用背景和强大的系统实验而迅速发展。在数据分布方面,客户/服务器数据库系统将快速发展,并在应用上取得良好效果。在智能化方面,数据库和人工智能将在各自的领域不断发展、不断取得新的成果。二者结合方面的研究将不断地利用二者的成果研制出新型的系统,任何时候两方面的结合都是必要的。

第2章 数据库结构模型

在诸多定义形式中，客户/服务器模型是描述服务客户（Client）与服务的提供者（Server）之间通信的一个概念。这种定义的特点就是客户通过发送一条消息或调用一个操作来启动与服务器之间的交互。类似地，服务器通过返回一条消息来响应。这种模型的一个基本结构如图 2-1 所示。

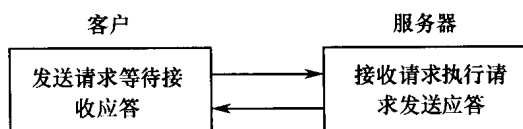


图 2-1 简单客户/服务器模型

客户/服务器系统又被称作分布式计算系统，它的含义是：程序的数据处理并不像通常在基于小型机或基于主机的计算机系统（终端方式）中那样在单个的计算机上发生，而是把程序的不同部分在多台计算机上同时运行。例如，对于SQL Server来讲，将数据存放在服务器计算机上，客户端界面作为程序的另一部分（完成商业逻辑和显示逻辑）存在于客户端桌面计算机上。客户/服务器系统的这两个部件通过网络连接相互通信，并且可以扩展到任意规模。SQL Server是一个真正的客户/服务器关系型数据库系统。它使企业可以设计出能够满足不断改变的信息需求的分布式数据库系统。客户端通过网络向服务器发送SQL语句，服务器返回客户端结果集（见图2-2）。

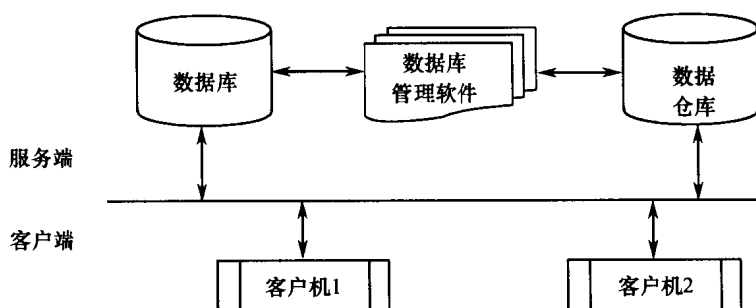


图 2-2 客户/服务器模式

2.1 客户机和服务器计算机的作用

客户机运行那些使用户能阐明其服务请求的程序，并将这些请求传送到服务器。由客户机执行的计算称为前端处理（Front-end processing）。前端处理具有所有与提供、操