

DBMS

● 鞠时光 著

对象关系型数据库 管理系统的开发技术

000100001



科学出版社

对象关系型数据库管理 系统的开发技术

鞠时光 著

科学出版社

2001

内 容 简 介

本书重点介绍和分析国内外一些著名的对象关系型数据库管理系统在设计和研制中的创新特色以及一些典型的概念、方法、算法和技巧。主要内容包 括 DBMS 系统结构、数据存储机制、文件索引机制以及信息查询机制的开发技术。本书力求使读者熟悉对象关系型 DBMS 的设计和实现的全貌,学会怎样去设计和实现一个对象关系型 DBMS,并能在各种体系结构、算法及技巧中进行选择和取舍。

本书主要作为计算机学科科技人员从事对象关系型 DBMS 研究、设计、使用的参考书,也可作为计算机专业研究生及高年级本科生的数据库原理课程的教材。

图书在版编目(CIP)数据

对象关系型数据库管理系统的开发技术/鞠时光著. -北京:科学出版社,2001

ISBN 7-03-009224-4

I. 对… II. 鞠… III. 关系数据库-数据库管理系统-系统开发
N. TP311.134.3

中国版本图书馆 CIP 数据核字(2001)第 06752 号

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2001 年 3 月第 一 版 开本:787×1092 1/16

2001 年 3 月第一次印刷 印张:11 1/2

印数:1—3 000 字数:257 000

定价:17.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

前 言

自从 20 世纪 60 年代末世界上第一个数据库系统问世以来,数据库技术已经有了突飞猛进的发展。尤其是 80 年代以来,随着计算机技术的普及,数据库应用正在向各方面渗透。一些新兴的应用领域,如计算机集成制造系统(CIMS),办公信息系统(OIS),地理信息系统(GIS)及多媒体系统(Multimedia)等,正以惊人的速度将复杂数据计算机化。

在这些新兴的应用系统中,新的数据类型不断涌现。它们对数据库管理系统提出了许多新的要求。比如,需存储和处理各种复杂对象;支持新的数据类型和操作;需要有更多的语义概念;支持动态模式修改;处理时态数据(历史数据);具有基于高维关系上的查询功能等等。

这些新的要求使得传统的关系型 DBMS 或者面向对象型 DBMS 均难以有效地管理这些数据类型的数据。所以,数据库管理系统必须能提供一些特殊的存储结构及管理机制。这就要求我们去研究、引入新型的 DBMS 系统。其中最重要的研究方向之一就是面向对象技术与关系数据库技术相结合,构造新型的对象关系型数据库管理系统(ORDBMS),以满足应用系统的需求。

本书是作者多年设计和开发空间数据库管理系统以及完成南京大学计算机软件新技术国家重点实验室课题“对象关系型数据库管理系统的研究和开发”的经验总结,同时介绍和分析了国内外一些著名的对象关系型数据库管理系统的设计和研制中的创新特色以及一些典型的概念、方法、算法和技巧,并参照北京大学唐世谓、杨冬青等老师的部分研究成果。本书力求使读者熟悉对象关系型 DBMS 的设计和实现的全貌,学会怎样去设计和实现一个对象关系型 DBMS。并能在各种体系结构、算法及技巧中进行选择和取舍。本书可作为本科计算机专业高年级学生和研究生的教材,也可作为研究人员、系统设计和开发者的参考书。

本书内容分为八章。

第一章是数据库管理系统概论,回答了为什么要研究、发展对象关系型数据库技术。作为本书的开头,本章将简要归纳一下数据库应用领域所提出的新的要求,并分析关系数据库管理系统与面向对象数据库管理系统各自的优缺点,进而讨论研制对象关系数据库管理系统的必要性及其可行性。

第二章以 Illustra 和 Informix Universal Server 这两个对象关系型的数据库管理系统为例,从支持复杂对象、扩充基类、继承机制、规则系统等几个方面,讨论一个对象关系数据库所必须具有的基本特性。

第三章系统地阐述一个 DBMS 所必须具备的基本功能及系统结构。首先介绍 RDBMS 及 OODBMS 的典型体系结构,使 ORDBMS 系统开发人员能从宏观和总体的角度掌握 DBMS 的基本概念,并从中得到启发。

数据的存储方法是 ORDBMS 中最底层的部分。它设计的优劣,直接关系到一个数据

库的工作效率。

第四章系统地回答以下一些问题:怎样在域中表示 SQL 的数据类型?怎样用记录来表示元组?怎样在内存块中表示记录或元组的集合?怎样存储对象关系型数据?例如对象标识符(或别的指向记录的指针)和大对象(例如 2GB 的 MPEG 格式的视频信息)。当一些域被更新而引起该记录的长度改变时,应该做什么?同时怎样的结构才能适合于记录变长的需要。

第五章讨论索引技术。首先介绍散列方法,然后讨论线性索引方法及其树形索引方法。

实现一个对象关系型数据库管理系统的关键是如何开发一个通用的存储机制和索引机制。

第六章对基于树的索引方法给出一种通用算法。该算法是建立在类属 B 树的概念之上开发的,它将类型系统开放,使其索引系统能支持用户自定义的数据类型、函数和某些特殊的查询谓词的集合。并且,将新的数据类型、函数、查询谓词等登记到数据库管理系统中,使得这些新的数据类型就像内部构造的类型一样让用户使用。用这种索引方法所开发的对象关系型数据库管理系统能有效地解决新兴应用领域对数据库管理系统的需求。

第七章介绍查询优化技术,考察用于改善逻辑查询方案的各种代数法则,并考察每一种关系代数算符的可能算法。

查询处理通常包括以下三个阶段:

1. 把由诸如 SQL 语言编写的查询程序,转化为一种能有效地描述查询结构的分析树。
2. 把分析树转化为一种表示关系代数的符号树,即逻辑查询方案。
3. 把逻辑查询方案转化为物理查询方案。即不仅说明其查询过程,而且指出其查询规则。这个规则贯穿于查询过程的每一步中。

第八章按照以上顺序,系统地介绍每个转换过程中所涉及到的相关技术。

上海师范大学理工信息学院陆黎明副院长审阅了全书,并提出了许多很有价值的意见。本书的编写工作得到了张小红、陈伟鹤、蔡涛、李星毅、任国栋等同志的大力支持。另外,本书的出版还得到了南京大学计算机软件新技术国家重点实验室的资助。在此,谨向所有支持及帮助出版的单位及同仁表示诚挚的感谢!

由于对象关系型 DBMS 是一门较新的发展学科,加之时间仓促,本人水平有限,本书一定存在不足和欠妥之处,请同仁批评指正。

鞠时光

2000 年 6 月

目 录

前言

第一章 绪论	1
1.1 应用领域对数据库技术的新要求	1
1.1.1 计算机集成制造系统	1
1.1.2 办公信息系统	2
1.1.3 空间数据应用	3
1.1.4 巨型复杂数据库的应用	3
1.1.5 计算机辅助软件工程	3
1.1.6 多媒体系统	4
1.2 数据库管理系统的分类	6
1.2.1 简单数据,简单查询	7
1.2.2 简单数据,复杂查询	7
1.2.3 复杂数据,简单查询	8
1.2.4 复杂数据,复杂查询	9
1.3 关系数据库管理系统的基本概念	9
1.4 面向对象数据库管理系统的概念	11
1.4.1 基本术语	11
1.4.2 面向对象技术与数据库技术相结合的多种方法	13
1.4.3 面向对象数据库的不足	15
第二章 对象关系数据库的基本特性	16
2.1 对象关系数据库定义	16
2.2 支持复杂对象	18
2.2.1 复杂对象对系统的要求	18
2.2.2 创建复杂对象	19
2.3 基类扩充	20
2.3.1 创建基类	21
2.3.2 定义函数和操作符	21
2.3.3 基类扩充模块	23
2.4 继承机制	28
2.4.1 数据继承	28
2.4.2 函数继承	30
2.5 规则系统	31
2.5.1 修改-修改规则	32

2.5.2	查询-修改规则	32
2.5.3	修改-查询规则	32
2.5.4	查询-查询规则	32
第三章	数据库管理系统结构	34
3.1	DBMS 的基本概念	34
3.1.1	数据库系统的层次结构	34
3.1.2	DBMS 的目标	37
3.1.3	DBMS 的基本功能	38
3.1.4	DBMS 与操作系统	39
3.2	RDBMS 的系统体系结构举例	40
3.3	OODBMS 的系统结构举例	43
3.3.1	对象子系统	43
3.3.2	事务子系统	44
3.3.3	存储子系统	45
3.4	客户/服务器结构中 DBMS 的功能划分	46
3.5	对象关系型数据库管理系统的两层体系结构	49
第四章	数据的存储方法	54
4.1	数据元素的表示	54
4.1.1	关系型数据库元素的定义	54
4.1.2	对象的表示	54
4.1.3	数据元素的表示	55
4.2	记录	57
4.2.1	建立固定长度的记录	57
4.2.2	记录的头信息	58
4.2.3	将定长记录组成数据块	59
4.3	块和记录地址的表示	60
4.3.1	客户机/服务器系统	60
4.3.2	逻辑和物理地址的转换	61
4.3.3	指针置换	62
4.3.4	将数据块写回磁盘	64
4.3.5	锁定记录	65
4.4	可变长度的数据和记录	65
4.4.1	带有可变长域的记录	66
4.4.2	带重复域的记录	67
4.4.3	格式可变的记录	68
4.4.4	长度大于数据块长的记录	69
4.4.5	大对象	70
4.5	记录的操作	70

4.5.1	添加记录	70
4.5.2	记录的删除	71
4.5.3	记录的更新	72
4.6	对象的存储	73
4.6.1	对象标识符	73
4.6.2	类的存储	73
4.6.3	对象实例的存储结构	73
4.7	对象关系型数据记录的存储	74
第五章	索引技术	76
5.1	散列方法	77
5.2	线性索引	79
5.3	B 树	81
5.3.1	B 树的定义	81
5.3.2	B 树结构中的三个层次	81
5.3.3	B 树的三大特点	82
5.3.4	B 树的搜索、插入与删除	83
5.4	多维索引技术	89
5.4.1	多维应用	89
5.4.2	k-d 树	92
5.4.3	R 树	94
第六章	类属 B 树索引技术	98
6.1	基本数据类型的扩充	98
6.1.1	用户自定义数据类型的原理	99
6.1.2	类型存储信息的定义	100
6.1.3	比较谓词的定義	100
6.1.4	操作函数的定义	101
6.2	类属的 B 树	101
6.2.1	类属 B 树的定义	102
6.2.2	GBT 树的查找算法	102
6.2.3	GBT 树的插入算法	103
6.2.4	GBT 树的删除算法	104
6.2.5	GBT 树的特点	105
6.3	地理数据类型的定义	105
6.3.1	点模型 POINT	106
6.3.2	曲线模型 CURVE	106
6.3.3	面模型 AREA	106
6.3.4	体模型 VOLUME	106
6.4	GBT 在 GIS 系统中的应用	107

6.4.1	比较谓词的定义	107
6.4.2	类型构造器	109
6.5	类型管理机制	113
6.5.1	系统目录表的定义	114
6.5.2	系统目录表的一个例子	114
第七章	查询优化技术	117
7.1	关系代数	118
7.1.1	并,交,差	119
7.1.2	选择运算	120
7.1.3	投影运算	121
7.1.4	笛卡儿乘积运算	122
7.1.5	连接运算	123
7.1.6	删去重复元组	124
7.1.7	分组和聚集	124
7.2	用于改善查询方案的代数定律	125
7.2.1	查询树	125
7.2.2	交换律和结合律	126
7.2.3	选择运算的特点	127
7.2.4	选择操作的移动	129
7.2.5	关于投影的规则	130
7.2.6	关于删去重复元组的规则	131
7.2.7	关于分组和聚集的规则	132
7.3	数据库操作的一次扫描算法	133
7.3.1	关系代数操作的计算模型	133
7.3.2	用于一元操作的主存算法	134
7.3.3	二元操作的主存算法	135
7.3.4	嵌套循环连接的算法	137
7.3.5	一次扫描算法的小结	138
7.4	基于分类的两次扫描算法	138
7.5	非代数优化的实现	140
7.6	基于复杂性估计的查询优化方法	143
第八章	查询处理	146
8.1	分析树	146
8.1.1	语法分析和分析树	146
8.1.2	SQL 语法	146
8.1.3	预处理程序	149
8.2	从分析树到查询方案的转换	150
8.2.1	向关系代数转换	150

8.2.2	从条件子句中去除子查询	151
8.2.3	优化逻辑查询方案	153
8.3	对操作开销的估算	154
8.3.1	估算中间关系的规模	154
8.3.2	投影结果规模的估算	155
8.3.3	选择结果规模的估算	155
8.3.4	连接结果规模的估算	156
8.3.5	具有多连接属性的自然连接	158
8.3.6	多关系连接	159
8.3.7	其他操作的规模估算	159
8.4	基本开销的优化	160
8.4.1	参数大小的估算	161
8.4.2	逻辑查询方案的枚举和开销计算	162
8.5	连接顺序的选择	163
8.5.1	左右连接变量的意义	163
8.5.2	连接树	164
8.5.3	左深度连接树	164
8.5.4	确定连接顺序的动态程序设计	165
8.6	物理查询方案的选择	168
8.6.1	缓冲区管理模块	168
8.6.2	缓冲区管理策略	169
8.6.3	确定选择算法	170
8.6.4	确定连接算法	171
8.6.5	管道操作与介质化	171
8.6.6	一元运算的管道操作	172
8.6.7	二元运算的管道操作	172
参考文献		174

第一章 绪 论

在系统地介绍对象关系数据库管理系统的基本概念及实现技术之前,我们首先应回答一个问题:为什么要研究、发展对象关系型数据库技术。作为本书的开头,本章将简要地归纳一下数据库应用领域所提出的新的要求,并分析一下关系数据库管理系统与面向对象数据库管理系统各自的优缺点,进而讨论研制对象关系数据库管理系统的必要性及可行性。

1.1 应用领域对数据库技术的新要求

自从 20 世纪 60 年代末世界上第一个数据库系统问世以来,数据库技术已经有了突飞猛进的发展。尤其是 80 年代以来,随着计算机技术的普及,数据库应用正在向各方面渗透。这一节我们首先介绍一下这些新兴的应用领域及其特征。

1.1.1 计算机集成制造系统

计算机集成制造系统(CIMS)是集现代信息技术、机器人技术、自动控制和管理科学为一体的综合性技术,是在计算机辅助设计(CAD)、计算机辅助制造(CAM)、计算机辅助工程(CAE)、计算机辅助生产过程规划(CAPP)、柔性制造系统(FMS)、机器人学(Robotics)和计算机辅助生产管理(CAPM)等基础上发展起来的高技术。CIMS 实现从产品需求、分析、设计、制造和管理到成品装备的综合自动化过程,涉及到大量的、可能是分布的单功能的主机之间的信息共享、传递、动态修改、快速检索以及一致性维护等等。因此,数据库技术是实现 CIMS 的关键之一。

CIMS 对 DBMS 的新功能要求包括:能定义新的面向工程环境的数据模型;能修改和重新定义已有的数据结构;能存储和有效地检索设计图形和工程数据;能管理设计过程中对象演变的历史;具有实施复杂对象间语义完整性、一致性约束的能力等等。

这就要求数据库管理系统对产品生成的不同阶段提供相适应的信息表示和数据处理能力,有效地支持如下的数据库应用:

- 市场预测与销售管理;
- 产品规划、生产管理、投资决策和成本核算;
- 产品设计与工程实施计划、原材料需求分析与规划;
- 产品制造的计划和管理;
- 制造过程的自动控制。

目前大量的研究工作主要集中在计算机辅助设计和制造中的数据管理问题,重点开发适合于 CAD 应用的数据库管理系统。

1.1.2 办公信息系统

随着大众对企业网兴趣的增加,办公自动化将会向无纸办公室转化。由于办公环境处理信息的多样性(图形、图像、报表、文字、声音)及存储介质的多样性(纸张、声音、录音、录像、磁盘、光盘等),多介质的 Intranet 和 Internet 信息管理是办公室自动化信息系统的重要课题。由于办公信息具有很多一般信息所不具有的特点,所以办公信息系统(OIS)对数据库系统提出很多新的要求。

1. 信息类型丰富

办公信息类型相当丰富,而且都属于非结构化信息,如正文、注册、口语会话等。办公信息中还包括很多印刷信息和具有标准格式的报表等。

2. 时间因素

在办公信息系统中有很多与时间有关的信息,如一个文件的传递时间,一项工作所持续的时间、日历和日程表,一个应答信息所允许的响应时间,闹钟信息等。所以办公信息系统的数据库模型必须包括时间因素和时间约束。

3. 不规则

办公活动常常是无规则的。完成某些工作的指令经常是不完整且无规则的。相应的信息也具有不规则性。经常性的通信和对话是必不可少的。

4. 交互式界面

办公信息系统是一个交互式系统。在各类办公室人员中,部分人对计算机和信息系统不太熟悉,这样用户界面就成了办公信息系统的一个具有挑战性的实际任务。

5. 信息的过滤与综合

多数办公信息系统都具有一个内在的层次结构。相应地,底层工作人员需要把他们的信息经过选择、过滤、综合整理后传递给上层工作人员。办公信息系统需要支持金字塔形的自底向上的信息过滤、综合和传递过程。

6. 优先级调度和触发器

工作优先级和调度复杂是办公信息系统的重要特点。办公信息系统需要定义工作优先级和完成调度工作的机构。在任何办公室工作的过程中都可能出现一些紧急事件需要立即处理。这样办公信息系统需要很多触发器。这些触发器用来在必要的时刻提醒工作人员处理紧急事件。

显然,办公信息系统是一个复杂的系统,作为办公信息的基础数据库管理系统必须满足上述要求。

1.1.3 空间数据应用

目前,很多计算机应用系统都以一维、二维和三维空间中的对象为基础。最早的空间数据应用是计算机辅助机械设计和几何应用领域。最近,空间数据应用的范围已经扩展到了机器人、计算机视觉、图像识别、环境保护、地理信息处理等领域。空间数据应用对数据库系统提出了新的空间数据管理要求。这些要求包括空间对象的表示、空间数据的存取方法、空间对象查询语言和查询优化等。更为困难的问题是不同的应用需要不同的技术。例如计算机辅助超大规模集成电路设计应用的几何要求是二维离散空间数据的管理。它需要表示、存储和处理的空间对象是点、线段、长方形和多边形。计算机视觉要解决的基本问题是根据二维图像来识别和确定三维空间对象的位置。然而,在地理信息管理系统中,要求以数据形式表示图形、图像数据,提供从市政计划、区域经济计划、交通控制、公安指挥到作战指挥等应用系统。这些都需要有良好的图形处理和检索功能。

1.1.4 巨型复杂数据库的应用

最近几年,数据库的规模增长很快,要求处理数亿字节的数据的应用已经屡见不鲜。宇宙空间研究是一个典型的巨型数据库应用实例。多年来,美国国家宇航局已经从宇宙空间获取了大量的数据。据估计,仅卫星图像信息在几年内就可达到 10^{16} 字节。人类基因的研究是另一个巨型数据库应用实例。每个基因序列由数亿个元素构成,而且每个元素都是一个复杂可变的对象。DNA 序列的分析研究要求新的数据表示和索引技术。

1.1.5 计算机辅助软件工程

数据库是计算机辅助软件工程环境的核心与基础。很多研究表明,传统的数据库系统难以有效地管理软件工程数据,软件工程应用需要新的数据库技术:

- 软件工程数据库需要存储很多新型数据,如源程序代码、软件模块间的联系、变量的定义和使用情况、图形数据、设计文档、软件项目的设计历史等。软件工程数据库中存储的数据项往往都具有信息量大、结构复杂、可变长等特点。

- 软件工程数据库中存储的实体型很多,但是每个实体型的实例却很少。实体型之间的关系十分复杂,而且这种关系经常是动态变化的。

- 软件工程数据库中存储的有关软件环境和设计项目的信息是持续变化的。

- 软件工程数据库存储的数据对象的模式经常被用户改变。

- 软件工程数据库的初始数据加载量很小,但是在软件项目的设计过程中数据库的增长非常迅速。这种增长包括数据内容和数据模式两个方面。

- 软件工程数据库中的很多数据项具有多个版本。版本之间的依赖性信息和相关性信息必须被严格地记录存储。

- 很多软件工程应用事务都是长事务。

- 软件工程应用需要在数据库上定义复杂的完整性约束。例如,我们要求软件工程师必须按照用户的规定设计给定的软件项目。传统数据库系统很难支持这类复杂完整性约束的定义和验证。

1.1.6 多媒体系统

多媒体(Multimedia)应用的发展正以惊人的速度将复杂数据计算机化。据估计,现在世界上 85% 的有用信息还未转变成电子形式,并且大部分是多媒体信息。如果将这些信息变成电子形式,由计算机进行管理,这将给 DBMS 带来巨大的应用市场。例如,利用 Web 作为传输机制发布信息内容,提供即时查询。这种情况下,每个 Web 节点都存在信息的管理和存取问题。

另一个例子是数字胶片。有人估计,在今后 10 年中,传统胶片作为数据存储体可能会消失。像 X 光片、家庭照片,均可采用数字化设备,把产生的数字图像存入计算机中,并提供浏览、查询等功能。这些应用具有数以万计的潜在用户。

在以上这些系统中,新的数据类型不断涌现。这些数据类型可统一到四维 $\{x, y, z, t\}$ 坐标下取值,即空间数据+时间序列。我们把这些二维以上的数据称为高维数据,如图 1.1 所示。

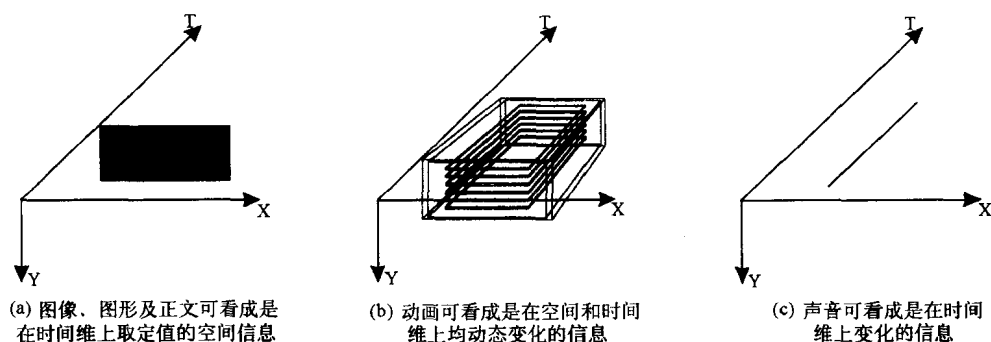


图 1.1

这些高维数据对数据库管理系统提出了许多特殊的要求,例如:

(1) 有时一个对象,比如,一幅图像,不能用存储字符编码的方式来存储。同样,检索一个图像也不像检索一个字符串或一个字符那样方便。

(2) 用户查询或检索某个对象,有时需据空间关系来进行。例如,CAD, GIS 等系统中常用的拓扑操作:“包含”,“相联”等。这些操作与空间物体的拓扑结构相关。

几何操作:“靠近”,“远离”,“覆盖”等。这些操作基于两个实体间距离的概念之上。

方位操作:“在---右边”,“在---上面”,“在---下面”等。这些操作基于实体间相互所处的空间位置之上。

(3) 虽然数据库管理语言有能力描述高维信息和字符信息的处理过程,但由于两种过程在本质上的差异,所以在诸如 DDL(数据库定义语言)、DML(数据操纵语言)、SQL(查询语言)当中,描述两种过程的语义规则不能混合。其过程如图 1.2 所示。

在这些新兴的应用领域中,由于高维数据的数据类型比传统的数据类型要复杂得多,所以数据库管理系统必须能提供一些特殊的存储结构及管理机制。其查询系统也需能定义与传统的查询系统完全不同的查询机制。为了能对高维信息系统的特征给出一个定性的描述,在此,我们以 CAD 系统为例,将其与传统的数据库系统进行比较。其比较结果如

表 1.1 所示。

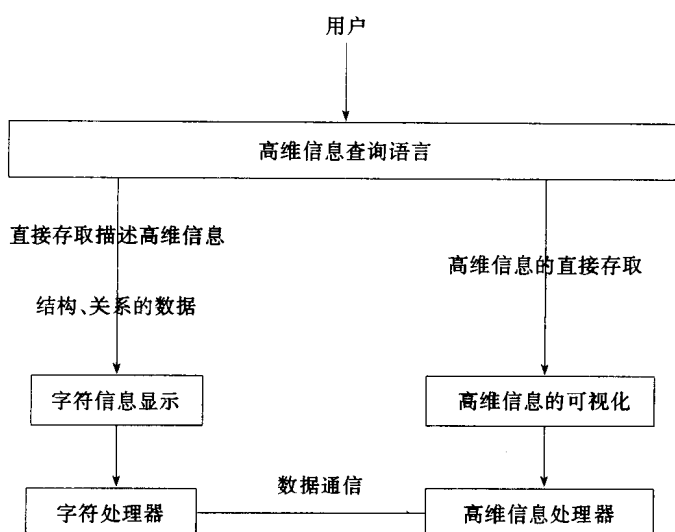


图 1.2 高维信息和字符信息的处理过程不同

表 1.1 CAD 数据库与传统数据库的比较

	传统数据库	CAD 数据库
被模仿的实际事物	被模仿的实际事物是静态的,清楚了解实体及实体间的关系,因此能对此进行描述,并被用于数据库的初始装入	必须处理下列两类信息: 1. 设计环境(规则,方法,标准,元件等) 2. 有关设计对象的数据是高度动态的,因为预先并不了解它们,在设计过程中才被确定
概念模式的性质	信息可预先描述,因此描述是静态和可编译的	因大部分信息相继地逐渐形成,模式必须是动态的
数据更新	数据更新虽然频繁,但只涉及值的改动	数据更新频繁,而且既更改数据值,也更改数据结构
谁执行更改	用户只能执行数据值的更改,但任何结构更改均需由数据库管理员实现	既然设计者使用包括数据库在内的 CAD 系统,必须允许它执行数据值与数据结构的更改
数据模型化的复杂性	一个管理模型可包含许多实体,但在大多数情况下,其间只是微弱相关	在某一对象的设计过程中,要涉及很多实体和许多关系,从而导致一个复杂的网状结构
数据结构的演变	一旦对实际数据进行了描述,并将数据库装入后,数据的使用是经常的,但结构的演变缓慢或不变	当对象没有完全被设计好时,对其在结构和值这两方面的描述是不断演变的
信息量	数据库相当稳定地管理大量数据	CAD 数据库开始较小,但可在设计过程中增大

通过比较, 我们可归纳出这些新兴的数据库应用需求所具有的主要特征:

1. 存储和处理各种复杂对象

这些对象一般具有复杂的层次结构, 比如说描述对象的某些属性本身又是另一个对象。这就很难用普通的平面关系结构来表示, 而且还有对象共享的要求。

2. 支持新的数据类型和操作

数据库管理系统需要支持新的数据类型和操作, 并能支持抽象数据类型, 提供用户定义新的数据类型和操作的可扩展工具。

3. 需要有更多的语义概念

比如泛化关系(*generalization*)以及聚集关系(*aggregation*)。

4. 支持动态模式修改

在工程设计系统中, 设计数据是逐步生成, 逐步增加的, 某些实体的结构以及实体间的联系有时事先未能考虑到, 因而需要对模式不断地扩充。这就要求 DBMS 能支持动态模式修改。

5. 时态数据(历史数据)

在有的控制系统中, 需要分析系统的演变过程, 这时需要用到一些历史数据。在工程设计系统中, 要求 DBMS 能保存和管理多版本数据。

6. 基于高维关系上的查询功能

高维数据之间, 必然存在一些高维关系。新兴的数据库应用要求能提供基于高维关系上的查询功能。如在设计某个机械系统时, 模拟两个相邻部件在运动过程中是否会发生干涉现象, 则是基于三维空间的查询。另外, 在时-空数据库中, 比方说长江江堤, 随着时间的变迁, 其形状、位置都会发生一定的变化。这是一个典型的基于四维关系上的查询。

这些新的特征使得传统的关系型 DBMS 或面向对象型 DBMS 均难以有效地管理这些类型的数据, 从而促使我们去研究、引入新型的 DBMS 系统。其中最重要的研究方向之一就是面向对象技术与关系数据库技术相结合, 构造新型的对象-关系型数据库管理系统, 用来组成高维信息应用系统。

1.2 数据库管理系统的分类

应用需求是数据库管理系统和其他数据管理软件发展的动力, 不同的应用领域需要管理的数据类型、数据复杂程度不同, 对数据所需进行的查询复杂程度也不同; 对数据的其他操作处理, 以及数据安全性、并发控制、故障排除的要求也不同。没有一种万能的数据管理软件能够适应所有的应用需求, 能够有效地为各个不同的应用领域服务。因此受应用需求的驱动, 多种多样的数据管理软件被研制出来, 推向市场。下面, 我们借用文献[10]中

对 DBMS 的分类方法,对现有的商用数据管理软件进行分类及讨论。在图 1.3 中,横坐标表示被处理数据的复杂程度,在此分成简单和复杂两类;纵坐标表示查询命令的复杂程度,分成简单和复杂查询两类。这样任一 DBMS,按照其应用特性,至少会落入该图中 4 个区域中的某一个。

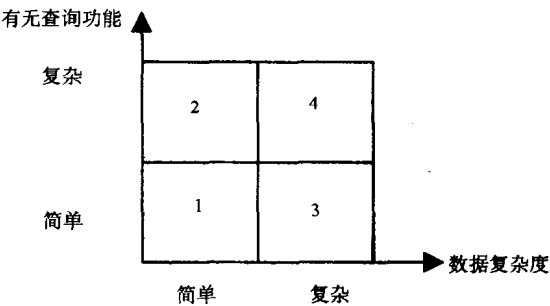


图 1.3 DBMS 的分类矩阵

1.2.1 简单数据,简单查询

在图 1.3 中,区域 1 表示处理的是简单数据,且只需一些简单查询或无需什么查询。这一类应用系统,正是通常的正文处理系统。常见的正文处理系统有 Word, Framemaker, Word Perfect, Vi 等。这些正文系统可归类为最简单的 DBMS 软件系统。它们的工作方式非常简单,使用者用名字打开文件,文件内容被拷贝到虚存中。在编辑正文时,更新虚存对象,对象被定期地保存到硬盘中。在结束正文处理时关闭文件,这时,虚存中的拷贝被存到文件系统中。

上面描述的这种正文处理系统代表了一类应用需求。它的数据结构非常简单,仅仅是字符的序列。它没有什么查询要求,如果有也很简单,仅仅是读文件和写文件。因此,对于此类应用,操作系统厂商提供的文件系统就可满足其功能要求。

1.2.2 简单数据,复杂查询

在图 1.3 中,区域 2 表示系统处理的是简单数据,但需能进行复杂查询的一类应用。关系数据库管理系统(RDBMS)正是适于具有这种特性的应用。目前常见的保险业务处理系统、企业的人事管理、生产管理、物资管理等信息管理系统都是基于这种 RDBMS 之上开发的。

下面,我们以公司职工管理信息系统为例,来看一看这类应用的需求和解决方案。

假设你需要为公司的每一个职工保存如下信息:职工的姓名、开始工作的时间、工资及所在的部门。为每一个部门,需要保存部门名称、楼层号、年度预算等信息。这些信息都可以通过使用符合 SQL-92 标准的 SQL 语句创建数据库表来完成。表中的属性都是 SQL-92 标准中的标准数据类型。我们称这些标准数据为“简单数据”。在系统的运行中,可能需要进行如下的查询:列出 1980 年以后参加工作、年工资超过 2 万元的职工姓名;或列出在同一楼层工作的职工姓名;或列出在市场部工作的职工工资表。查询要求可能很复杂,但是 SQL 语句很容易把这些查询表达出来。

满足这一类应用需求的是成熟的关系数据库技术。对于查询语言,国际标准组织相继推出 SQL-86、SQL-89 和 SQL-92 标准。为了方便数据库应用系统的开发,数据库厂商开发了各种客户工具,除了可以在 SQL 语句中嵌入第三代语言(如 C, COBOL 等),还提供 4GL 第四代语言,使得用户能够以交互的方式创建用于录入和显示信息的 Form,并通过控制流逻辑建立 Form 间的顺序关系。另外,还有一些报表书写工具、数据库设计工具、性能监视器等可供使用。