

对象—关系数据库管理系统 原理与实现

李战怀 李红燕 徐秋元 等著



清华大学出版社





对象—关系数据库管理系统 原理与实现

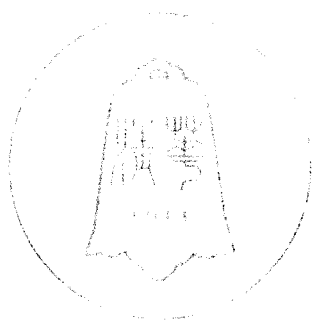
陈松海 李国杰 张立群 等著





对象-关系数据库管理系统 原理与实现

李战怀 李红燕 徐秋元 等著



清华大学出版社
北京

内 容 简 介

本书以完全从底层研究开发的对象-关系数据库管理系统 Angel 为背景,在深入分析国内外相关研究成果的基础上,针对对象-关系这种目前应用最为广泛、最为重要的数据库管理系统,论述了对象-关系数据库系统的原理与实现技术。作为国内第一本全面、系统而深入地论述对象-关系数据库管理系统设计思想与实现技术的专著,本书在基类扩充、对复杂对象的支持、继承语义及动态模式修改、引用语义、主动性规则、安全机制等方面实现了面向对象技术与关系技术的有机融合,其内容涵盖了对象-关系系统的数据模型、查询代数及查询语言等各个方面。

本书立论严谨,内容新颖,结构合理,语言流畅,可以作为计算机系研究生的数据库专业课教材,也可供有关数据库软件人员或高校教师参考。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

对象-关系数据库管理系统原理与实现/李战怀,李红燕,徐秋元等著. —北京:清华大学出版社,2006.6
ISBN 7-302-12781-6

I. 对… II. ①李… ②李… ③徐… III. 关系数据库—数据库管理系统—研究 IV. TP311.138

中国版本图书馆 CIP 数据核字(2006)第 029574 号

出 版 者: 清华大学出版社 地 址: 北京清华大学学研大厦
http://www.tup.com.cn 邮 编: 100084
社 总 机: 010-62770175 客户服务: 010-62776969

组稿编辑: 丁 岭

文稿编辑: 徐跃进

印 刷 者: 北京鑫丰华彩印有限公司

装 订 者: 三河市新茂装订有限公司

发 行 者: 新华书店总店北京发行所

开 本: 185×260 印张: 20.75 字数: 517 千字

版 次: 2006 年 6 月第 1 版 2006 年 6 月第 1 次印刷

书 号: ISBN 7-302-12781-6/TP·8140

印 数: 1~3000

定 价: 35.00 元

读者意见反馈

亲爱的读者：

感谢您一直以来对清华版计算机教材的支持和爱护。为了今后为您提供更优秀的教材，请您抽出宝贵的时间来填写下面的意见反馈表，以便我们更好地对本教材做进一步改进。同时如果您在使用本教材的过程中遇到了什么问题，或者有什么好的建议，也请您来信告诉我们。

地址：北京市海淀区双清路学研大厦 A 座 602 (100084) 信息分社营销室收

电话：010-62770175-4608/4409 邮购电话：010-62786544

电子邮件：jsjic@tup.tsinghua.edu.cn

教材名称：对象-关系数据库管理系统原理与实现

ISBN：7-302-12781-6/TP·8140

个人资料

姓名：_____ 年龄：_____ 所在院校/专业：_____

文化程度：_____ 通信地址：_____

联系电话：_____ 电子信箱：_____

您使用本书是作为：☐指定教材 ☐选用教材 ☐辅导教材 ☐自学教材

您对本书封面设计的满意度：

☐很满意 ☐满意 ☐一般 ☐不满意 改进建议_____

您对本书印刷质量的满意度：

☐很满意 ☐满意 ☐一般 ☐不满意 改进建议_____

您对本书的总体满意度：

从语言质量角度看 ☐很满意 ☐满意 ☐一般 ☐不满意

从科技含量角度看 ☐很满意 ☐满意 ☐一般 ☐不满意

本书最令您满意的是：

☐指导明确 ☐内容充实 ☐讲解详尽 ☐实例丰富

您认为本书在哪些地方应进行修改？（可附页）

您希望本书在哪些方面进行改进？（可附页）

数据库技术在商业领域的巨大成功刺激了其他领域对数据库技术需求的迅速增长。这些新的领域为数据库应用开辟了新的天地,而应用中提出的新的数据管理的需求也直接推动了数据库技术的研究与发展。20 世纪 80 年代,面向对象数据库系统(object oriented database system,OODBS)的研究与发展就是一个极好的例证。

20 多年来,国内外学术界、工业界努力探索数据库技术与对象技术的结合,沿着三条路线展开了面向对象(OO)数据模型和面向对象数据库系统的研究:

一条是以面向对象的程序设计语言为基础,研究持久的程序设计语言,支持 OO 模型;

一条是建立新的面向对象数据库系统支持 OO 数据模型;

一条是以关系数据库和 SQL 为基础,把面向对象技术融入数据库系统的对象关系数据库系统(object relational database system,ORDBS)。

回顾上述 3 个方向所做的研究及开发工作,可以看到各自的发展是不平衡的。其中 ORDBS 方向的成果最为卓著。它在传统关系数据库的基础上吸收了面向对象主要思想,同时又保持了关系数据库的优点,成功开发了诸如 Postgres、Illustra 等原型系统。近年来,各大 RDBMS 厂商也已推出了其产品的对象-关系版本,从而满足了许多新的数据库应用需求。

同样,面向对象数据库研究领域吸引了我国许多数据库工作者,取得了大量的研究成果,开发了多个面向对象数据库管理实验系统。本书介绍的就是其中之一。

从 1991 年起,西北工业大学数据库课题组开始了面向对象数据库技术的研究。该课题组以本书作者李战怀教授和他的导师徐秋元教授为核心,他们从探索新一代 DBMS 研制技术以及研制具有自主知识产权的数据库产品出发,在八五国防预研项目、国家自然科学基金、863、航空基金、陕西省自然科学基金等项目的支持下开发了客户机-服务器结构的对象-关系数据库系统 Angel。他们从系统底层做起,历时 8 年,投入大量精力和人力研发了 Angel-I、Angel-II。该系统技术先进,功能丰富,在国防科工委与航空工业部主持的技术鉴定会上得到专家们的高度评价,并且于 1996 年获得国防科技进步二等奖。

本书就是对 Angel 系统以及以此为背景所进行的研究成果(包括 3 篇博士论文)的系统介绍。这对于学习大型数据库系统软件的设计和开发有很好的参考价值,特别是对学习

ORDBS 的基本概念、实现技术、关键技术中难点要点以及实际应用的解决方案是极其有帮助的。

从本书所介绍的研究和开发工作充分说明,研制具有我国自主知识产权的数据库系统一直是我国数据库界长期的执着的追求。我相信这种追求一定会带来丰硕的成果。

中国人民大学 数据与知识工程研究所

王 珊

我们西北工业大学数据库研究组从 20 世纪 80 年代初期开始,在徐秋元教授和李战怀教授的领导下一直从事工程数据库系统的研究,20 世纪 80 年代中期以后,国际数据库界以新型应用领域(CAD、VLSI 设计、CASE 等)的需求为背景,广泛开展了新一代数据库管理系统的研究,由于这些研究的背景与我们一直关注的应用背景有很大的相似之处,因而引起了我们巨大的关注,经过反复讨论,我们决定自行研发一个数据库管理系统。

采用何种数据模型是我们首先需要解决的问题,当时国际数据库界主流的研究方向是面向对象系统和基于 PRLOG 的知识库管理系统,但也有一些观点认为要通过对关系模型增加面向对象特征来增加数据模型建模能力,如加州大学伯克利分校的 M. Stonebrake 教授就是这一观点的代表者。通过对各种模型,如面向对象模型、基于逻辑的模型、嵌套关系模型、复杂对象模型、Postgres 的数据模型等广泛的研究,我们提出了一种具有面向对象特征的非第一范式的模型,对这一模型的介绍以“一种适用于工程数据管理的数据模型”名称发表于第十一届全国数据库会议论文集中,基于此模型实现的数据库管理系统我们称之为 Angel。

Angel 的实现分两个阶段。1991—1995 年是第一阶段,在这个阶段我们完全从底层做起,在 Sun 工作站上实现了一个基于客户机-服务器结构的对象关系数据库管理系统,我们称之为 Angel-I。1995 年年底在国防科工委与原航空工业部主持下开了鉴定会,主要结论是:国内领先,有关研究与国际同步。1996 年获得国防科技进步二等奖。1996—1998 年是第二阶段,这一阶段主要的工作是提高系统的健壮性并开发客户端开发工具,在此期间,我们完全重新设计了整个系统的体系结构,实现中更多地采用了面向对象的实现方法,在存储管理、查询执行和优化方面做了大量工作,实际上重新实现了整个系统。同时在客户端,我们为 Angel 开发了 ODBC DRIVER,嵌入 C 的预编译器,以及导航式图形化查询界面。这一阶段开发的系统我们称之为 Angel-II。整个 Angel 系统的实现总计花费了四十多人年的工作。

Angel 系统的研究与实现,得到国防预研、国家自然科学基金、国家 863 网络与数据库专题、航空基金、陕西省自然科学基金等多方支持,在此向有关单位致以深深的谢意。

本书以李战怀、李红燕、邓江的博士论文为基础,加上这些年来我们围绕 Angel 系统发表的论文和技术报告编著而成。虽然主要内容是现成的,但要整理成一本书还是需要花费很大的精力。李红燕副教授从 2002 年开始整理有关资料,更系统地对以前的工作进行了总

结,到 2003 年年底,经过 1 年多时间的努力,终于完成了这本书。作者要特别感谢蒋泽军教授等对 Angel 系统实现作出贡献的所有项目组成员,但由于人员众多,恕不一一列名致谢。

由于众所周知的原因,出版一本专著性质的书是非常不容易的,作者特别感谢清华大学出版社对本书出版给予的支持,感谢编辑们的辛勤劳动,没有他们的帮助,本书的出版是不可能的。

出版本书的目的是为了把我们在研究对象-关系数据库管理系统方面的心得体会介绍给大家,希望起到抛砖引玉的作用。由于作者的水平与能力所限,书中肯定存在很多不足,如蒙赐教,将不胜感激。

作 者

2006 年 2 月

第 1 章 绪论	1
1.1 研究目的及意义	1
1.1.1 扩展关系数据库	2
1.1.2 面向对象数据库	2
1.1.3 对象-关系数据库	3
1.2 对象-关系技术的研究进展	3
1.2.1 对象-关系技术的萌发	3
1.2.2 对象-关系技术的现状	4
1.2.3 对象-关系技术的前景	6
1.3 相关研究工作概述	6
第 2 章 数据模型	8
2.1 设计思想	8
2.1.1 关于值与对象的讨论	8
2.1.2 对象标识	10
2.1.3 对象结构	11
2.1.4 继承性	12
2.1.5 封装	13
2.1.6 方法	13
2.2 形式描述	14
2.2.1 型与值	14
2.2.2 类与关系	16
第 3 章 查询代数	20
3.1 保持模式结构的操作	20
3.1.1 并	21
3.1.2 差	22

3.1.3	交	23
3.1.4	选择	23
3.2	改变模式结构的操作	25
3.2.1	投影	25
3.2.2	联结	27
3.2.3	NEST 和 UNNEST	28
3.3	代数操作的嵌套使用	29
3.3.1	扩展选择	29
3.3.2	扩展投影	30
第 4 章	体系结构	32
4.1	客户机-服务器体系结构	32
4.1.1	设计原则	32
4.1.2	平台及通信工具	33
4.1.3	功能分布	33
4.1.4	服务器的多进程结构	34
4.2	系统功能划分	35
4.2.1	客户端	35
4.2.2	服务器端	35
4.2.3	语言处理	35
4.2.4	模式管理	36
4.2.5	事务处理	36
4.2.6	查询执行	36
4.2.7	对象管理	36
4.2.8	存储管理	36
4.3	系统环境	37
4.3.1	系统运行环境	37
4.3.2	用户界面	37
第 5 章	数据库语言	38
5.1	主要特点	38
5.2	ADBL 的数据定义功能	39
5.2.1	数据库的创建	39
5.2.2	表的创建	40
5.2.3	索引的创建	44
5.2.4	视图的创建	44
5.3	ADBL 的模式操纵功能	45
5.3.1	表的修改	46
5.3.2	表的删除	47

5.4	ADBL 的查询功能	47
5.4.1	对构造层次的查询	49
5.4.2	对继承层次的查询	52
5.4.3	对传递闭包的查询	53
5.5	ADBL 的数据操纵功能	54
5.5.1	INSERT 语句	54
5.5.2	UPDATE 语句	56
5.5.3	DELETE 语句	56
5.5.4	对视图的操作	57
5.6	ADBL 的数据控制功能	59
5.6.1	完整性保护	59
5.6.2	安全性保护	61
5.6.3	事务管理	65
5.6.4	封锁机制	67
第 6 章	基类扩充	68
6.1	对 built-in 类型的扩充	68
6.1.1	域	68
6.1.2	模式类型	69
6.1.3	抽象数据类型	69
6.2	用户定义函数机制	70
6.2.1	UDF 在 ADBL 中的使用	71
6.2.2	UDF 的继承机制	72
6.2.3	UDF 的实现	73
6.3	利用 UDF 实现操作符定义及操作符重载	76
第 7 章	嵌入式 ADBL	78
7.1	语法结构	78
7.1.1	可执行语句	78
7.1.2	说明性语句	79
7.2	预编译实现方法	81
7.3	嵌入式 ADBL 的程序首部处理	83
7.3.1	SQL 变量	83
7.3.2	SQL 通信区	87
7.4	嵌入式 ADBL 的程序体处理	87
7.4.1	不含输出宿主变量的 AEDBL 语句	88
7.4.2	含有输出宿主变量的 AEDBL 语句	88
7.4.3	特殊 AEDBL 语句的处理	89
7.5	游标机制	89

7.5.1	游标的声明	90
7.5.2	游标的打开	91
7.5.3	对当前实例的取值	91
7.5.4	游标的关闭	92
7.5.5	带游标的 ADML 操作	92
7.5.6	游标的撤销	95
第 8 章	复杂对象	96
8.1	扩展的非第一范式结构	96
8.2	复杂对象语义	97
8.3	对象引用	99
8.3.1	引用作为完整性约束	99
8.3.2	引用作为类型构造器	101
8.3.3	两种特殊的引用方式	106
8.3.4	引用信息的外存表示	107
8.4	复杂对象信息组成	108
8.5	复杂对象表示形式	109
8.5.1	复杂对象语言形式	109
8.5.2	复杂对象内存形式	109
8.5.3	复杂对象外存形式	112
8.5.4	复杂对象显示形式	116
8.6	对象管理	116
8.6.1	对象缓冲区管理	117
8.6.2	对象存取	119
8.6.3	驻留对象管理	121
8.7	对象聚集	122
8.7.1	传统聚集策略	122
8.7.2	Angel 聚集策略	122
8.8	对复杂对象操作的实现	125
8.8.1	ADML 操作	125
8.8.2	授权操作	127
8.8.3	并发控制操作	128
第 9 章	查询处理与优化	130
9.1	查询语义	130
9.1.1	路径语义	130
9.1.2	对层次的讨论	131
9.1.3	层次的确定	132
9.1.4	层次的匹配	136

9.2 查询处理	137
9.2.1 总体处理方案	137
9.2.2 嵌套查询处理	143
9.2.3 继承层次查询处理	150
9.2.4 传递闭包查询处理	152
9.3 查询优化	155
9.3.1 代数查询优化	156
9.3.2 索引机制	163
9.3.3 基于代价评估的优化方法	165
第 10 章 模式管理与动态模式修改	176
10.1 模式管理	176
10.1.1 数据字典	176
10.1.2 ANGEL\$TABLES 表	177
10.1.3 模式管理器	181
10.1.4 模式管理的应用	184
10.2 动态模式修改	185
10.2.1 ORDBMS 中的 DSM	185
10.2.2 DSM 在 Angel 系统中的实现	190
第 11 章 安全机制与完整性保护	206
11.1 安全机制	206
11.1.1 安全性定义	207
11.1.2 安全性检查	218
11.1.3 安全机制的实现	218
11.2 完整性保护	220
11.2.1 ORDBMS 主动性规则的研究	221
11.2.2 Angel 对规则系统的支持	224
11.2.3 S E-C-A 模型的实现	226
11.2.4 S E-C-A 模型中的规则行为分析	229
11.2.5 语义完整性约束	232
第 12 章 事务管理	236
12.1 事务	236
12.1.1 基本概念	236
12.1.2 事务管理器	239
12.2 常规事务管理	241
12.2.1 封锁与并发控制	242
12.2.2 日志与恢复	247

12.2.3	事务管理器的实现	250
12.3	长事务管理	250
12.3.1	保存点机制	251
12.3.2	具体实现	252
第 13 章	开放数据库连接驱动程序的设计与实现	256
13.1	开放数据库连接简介	256
13.1.1	ODBC 接口	257
13.1.2	ODBC 组件	258
13.1.3	驱动程序类型	259
13.1.4	一致级	260
13.2	设计目标	262
13.2.1	驱动程序结构	263
13.2.2	Winsock 简介	265
13.3	重要数据结构	266
13.3.1	三级句柄	266
13.3.2	数据类型	267
13.3.3	返回状态与出错信息	268
13.4	Angel 数据传输协议	269
13.4.1	连接 Angel 服务器	269
13.4.2	发送命令	270
13.4.3	返回结果	270
13.5	具体实现	272
13.5.1	初始化连接	273
13.5.2	终止连接	275
13.5.3	处理 SQL 语句	276
13.5.4	取回结果	279
13.5.5	处理终止语句或事务	282
13.5.6	信息查询与目录函数	283
13.6	数据源管理	284
第 14 章	复杂对象推理的逻辑语言	285
14.1	A-logic 的语法	285
14.1.1	类型	285
14.1.2	A-logic 语言	286
14.2	A-logic 的语义	299
14.3	讨论	302
附录 A	Angel 数据语言文本	304
参考文献		315

对象-关系数据库技术是目前数据库领域的主流技术,本书将完全从低层研究开发的对象关系数据库管理系统 Angel 为背景,全面深入地论述了对象关系数据库系统的原理与实现技术。作为绪论,本章介绍研究对象-关系数据库管理系统的目的及意义,并对对象-关系技术的研究进展进行总结和回顾。

1.1 研究目的及意义

应用的需求始终是推动技术进步的动力。自 20 世纪 70 年代初 E. F. Code 提出关系数据库理论,定义关系数据模型以来,关系数据库管理系统 (relational database management system, RDBMS) 以其简洁、严密的数据模型,简单易学、功能强大的描述性查询语言而备受用户青睐,并迅速取代网状及层次型数据库系统,成为数据库技术发展的主流。之后,随着 RDBMS 理论基础及其实现方法的日趋成熟,关系数据库技术日臻完善。

近年来,随着计算机辅助设计 (CAD)、计算机辅助制造 (CAM)、计算机辅助软件工程 (CASE)、全球信息系统 (GIS)、图像处理、超文本应用等领域的飞速发展及其在传统领域中应用的深化,要求数据库管理系统 (database management system, DBMS) 能够有效地管理复杂对象。比如在工程应用领域,一个客观复杂实体往往由数十个,甚至成百上千个简单实体组成,为了减小数据库应用系统的设计复杂度、提高其执行效率,要求 DBMS 不但能根据实体丰富的语义进行建模、提供有效的存储与操纵手段,以及模拟复杂实体的复杂行为,而且在逻辑上还要将一个复杂实体的表示和操纵作为一个整体看待,在操纵数据的同时考虑实体间的复合语义,即各简单实体的存在方式 (独立或依赖) 以及实体间的引用方式 (共享或排他)。

然而,传统 RDBMS 由于采用满足第一范式 (first normal form, 1NF) 的平关系模型,在面对各种新的应用领域时存在以下不足。

(1) 不能完整地描述复杂实体。对一个逻辑上作为整体的复杂实体只能采用分解成多个关系的方法进行描述,从而在内部数据库结构和由这种结构所描述的外部实体之间缺乏一对一的对应关系,在查询时这将导致系统效率的严重降低。

(2) 缺乏稳定的元组标识。RDBMS 中完全由关键字属性值作为元组的惟一标识,当关键字值不稳定时,引用时会产生数据的一致性问题的,给维护带来很大困难。

(3) 缺乏数据抽象。在关系模型中只有一个非常简单的结构构造概念——关系,缺乏

诸如聚合(把由多个子实体组成的复合实体抽象成一个聚合实体)、概化(把相似实体归类为一个更一般的实体型)与特化(把抽象的实体型细化成更特殊的实例)等在许多实际应用中非常重要的抽象概念。

(4) 缺乏丰富的类型系统。

这些不足使得传统关系型数据库技术在面对来自“非经典”应用领域中的巨大挑战时显得力不从心,新的数据库技术应运而生。

1.1.1 扩展关系数据库

这类研究相对来说是“改良性”的:即在传统关系模型上增添各种特点,或打破原有约束以支持各种新型的、复杂结构的数据。早期对关系模型的扩展主要包括非第一范式(non-first normal form, NF^2)和抽象数据类型(abstract data type, ADT)。

非第一范式打破了关系模型中第一范式的要求,从而能够有效地描述复杂实体。这方面的工作主要集中在 20 世纪 80 年代,所作的扩充也各有不同,包括: P. Pistor 等提出的 AIM-P 中,允许关系的属性为任意域^[1]; Jaeschke 与 Schek 所扩充的关系模型允许关系带有集合值属性^[2]; 由 Thomas 和 Fischer 提出的“嵌套关系模型”^[3]。其中,嵌套关系模型融合了前两种模型,它允许关系中集合操作与聚合操作以任意层次相互嵌套,建模能力最强,因此后续的研究大多围绕嵌套关系进行,包括代数和演算系统^[4]、查询语言^[5]、规范化^{[4][6]}等。

抽象数据类型原本是程序设计中的重要概念,它强调数据的封装,是面向对象思想的起源之一。而在数据库系统中,ADT 表示用户定义类型,用户不仅要定义类型的规格(数据长度及适用函数),还需要用外部语言予以实现。DBMS 对新类型进行登记,并在适当时机调用相应函数。将 ADT 引入 DBMS 的先驱是加州伯克利大学的 ADT-Ingres^[7],并在其后续系统 Postgres^[8]中进一步阐明了相关的优化方法。ADT 是扩展数据库类型系统的强大武器,在很多系统中都采用这种方法。

关系模型的这类扩展,克服了平关系不能自然地表达复杂实体结构的缺点,并因为与平关系有着自然的联系,从而可以充分利用平关系系统的研究成果和原有的理论基础。另一方面,这种方法虽然确实增强了关系数据库(RDB)的建模能力,但在数据聚合、概化和特化、刻画复杂对象行为以及对象标识的稳定性等方面,仍未能有根本性改善。

1.1.2 面向对象数据库

这类研究将面向对象思想融入数据库技术,从而发展为全新概念的面向对象数据库管理系统(object-oriented database management system, OODBMS)。尽管 OODB 缺乏通用数据模型、形式基础及强有力的实验活动,但一般说来,它至少包含有如下核心概念^[9]:

(1) 对象和对象标识。在面向对象系统和语言中,现实世界的客体统一地用对象来模拟,每一对象有与之相联系的惟一标识符。

(2) 属性和方法、封装和消息传递。每一对象都有一个状态和行为。对象的状态是对象属性值的集合,而对象的行为是作用于对象状态上的方法(程序代码)集。对象的属性值