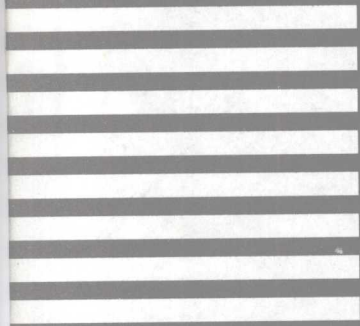


数据库系统原理

周志達 郭贵锁 等 编著



清华大学出版社



数据库系统原理

周志逵 郭贵锁 陆耀 孙新 张文耀 编著

清华大学出版社
北京

内 容 简 介

本书对数据库系统的概念、原理、技术和方法进行了系统和全面的阐述。全书共分 17 章,其中,第 1~3 章介绍了数据库的基本概念,包括数据库的发展过程、数据库的系统结构、数据模型和关系代数等;第 5~9 章对数据库管理系统中的查询优化、数据库安全性和完整性、数据库的恢复技术、并发控制等内容进行了阐述;第 10 章和第 11 章分别是数据库设计理论和数据库设计方法;第 4 章和第 12 章分别介绍了标准的 SQL 语言和 SQL Server 的 Transact-SQL 语言编程;第 13 章介绍了数据库的存储技术;第 14~17 章是数据库新技术的内容,分别介绍了分布式数据库、面向对象数据库、多媒体数据库、空间数据库、XML 数据库等。

本书在介绍理论的同时对 SQL Server 数据库中的具体实现进行了简要的描述,以达到理论与实际相结合的目的。

本书可以作为高等学校计算机专业、信息管理与信息系统等相关专业数据库原理课程的教材,也可作为从事数据库系统研究、开发和应用的工程技术人员和工程技术人员参考书。

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。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

数据库系统原理/周志逵等编著. —北京:清华大学出版社,2008.11

ISBN 978-7-302-18626-7

I. 数… II. 周… III. 数据库系统 IV. TP311.13

中国版本图书馆 CIP 数据核字(2008)第 145006 号

责任编辑:张瑞庆 李玮琪

责任校对:焦丽丽

责任印制:何 芊

出版发行:清华大学出版社

<http://www.tup.com.cn>

社 总 机:010-62770175

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

地 址:北京清华大学学研大厦 A 座

邮 编:100084

邮 购:010-62786544

印 刷 者:北京嘉实印刷有限公司

装 订 者:三河市李旗庄少明装订厂

经 销:全国新华书店

开 本:185×260

印 张:26.5

字 数:645 千字

版 次:2008 年 11 月第 1 版

印 次:2008 年 11 月第 1 次印刷

印 数:1~4000

定 价:36.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。

联系电话:010-62770177 转 3103 产品编号:028993-01

数据库技术是计算机科学领域中发展最快、应用最广泛的技术之一。自 20 世纪 60 年代末以来,在 40 年的发展历程中,数据库技术的研究和应用取得了巨大的成就。数据库技术被广泛应用于管理各种信息,已成为当今计算机信息系统的基础和核心,成为管理和利用信息资源不可缺少的工具。

数据库技术的广泛应用受到了人们的极大关注,学习和了解数据库技术成了广大科技工作者、管理人员和数据库使用者的共同需要。数据库是计算机专业大学生的必修课之一,非计算机专业的学生也需要学习和应用数据库技术来解决本专业的问题。从这些需求出发,作者编写了这本数据库系统原理教材。

本书是作者在多年从事教学和科研的基础上编写成的。在编写和组织内容时,注重了内容的全面性和实用性。本书内容包括数据模型、数据库语言、数据库的实现技术、数据库的设计理论和技术、数据库的存储结构等。除了介绍数据库的基础知识和技术外,还介绍了数据库的新技术,力求使读者能够全面了解数据库涉及的概念、原理、方法和技术,了解数据库技术的发展和研究成果。书中的大部分示例结合了商用数据库管理系统 SQL Server 数据库。从实用性出发,在介绍基本 SQL 语句功能的基础上,进一步介绍了这些命令语句的实际应用,以帮助读者加深对 SQL 语句的理解和应用。在数据库设计部分,介绍了实用的 IDEF1X 数据建模方法,这在同类教材中很少见。书中结合实际示例介绍了数据库系统的数据建模过程。

全书共分 17 章,第 1 章数据库系统引论,介绍数据处理技术的发展、数据库系统结构和数据库管理系统等涉及的基本概念;第 2 章数据模型,介绍 E-R 模型、层次模型和网状模型、关系模型和面向对象模型中的基本概念;第 3 章关系数据库,介绍关系数据库涉及的基本概念及对关系的各种运算;第 4 章关系数据库标准语言 SQL,介绍 SQL 语言涉及的基本概念和 SQL 语言的主要功能;第 5 章查询处理和优化,介绍关系数据库的查询处理和查询优化技术;第 6 章至第 9 章,介绍数据库保护技术,分别介绍了数据库安全性概念和数据库采用的各种安全性措施、数据库完整性概念和完整性约束机制、数据库恢复技术和数据库并发控制技术,并介绍了实际数据库管理系统中所采用的各种数据保护技术;第 10 章至第 12 章介绍数据库设计理论和设计方法,结合 SQL Server 数据库介绍了数据库编程中的基本概念和方法;第 13 章数据库的存储结构,介绍数据库存储技术,包括记录的存储结构和数据库中使用的各种文件的存储结构;第 14 章至第 17 章介绍数据库的新技术,重点介绍了分布式数据库、对象和对象关系数据库、多媒体数据库、空间数据

库、XML 数据库等,并简要介绍了其他数据库的新技术和新应用。

本书在编写过程中注意取材合理,尽量反映国内外最新研究成果,力求做到理论联系实际,概念清晰,通俗易懂,以便于自学。

本书可作为大专院校数据库课程的使用教材,1~13 章为本科生教学的基本内容,建议学时 48~64,其中带 * 的部分为非主讲内容,其他章节可供本科生、研究生作为参考。本书也可作为从事数据库系统研究、开发和应用的研究人员和工程技术人员的参考书。

参加本书编写的人员和章节:第 1 章~第 4 章和第 10 章由周志逵编写,第 5 章~第 7 章由孙新编写,第 8 章和第 9 章由陆耀编写,第 11 章、第 12 章、第 16 章和第 17 章由郭贵锁编写,第 13 章~第 15 章由张文耀编写。周志逵对全书内容进行了统稿和审定。

在本书的编写过程中,2006 级研究生陈银美等参与了书稿的部分编辑工作,在此表示衷心的感谢。

由于作者水平有限,书中难免存在许多不足之处,恳请读者批评指正。

作 者

2008 年 5 月

第 1 章 数据库系统引论	1
1.1 数据管理技术的发展	1
1.1.1 人工管理阶段	1
1.1.2 文件系统阶段	2
1.1.3 数据库系统阶段	3
1.2 什么是数据库	5
1.3 数据模型	6
1.4 数据库系统结构	7
1.4.1 数据库系统的三级模式结构	8
1.4.2 三级模式结构的二级映像	9
1.5 数据库管理系统	9
1.5.1 数据库管理系统的功能	9
1.5.2 数据库管理系统的组成	12
1.5.3 数据库系统的工作过程	12
1.6 数据库系统的不同视图	13
1.7 数据库技术的发展	15
1.8 小结	16
习题	16
第 2 章 数据模型	18
2.1 E-R 概念模型	18
2.1.1 E-R 数据模型中的基本概念	18
2.1.2 E-R 数据模型	21
2.2 层次数据模型	25
2.3 网状数据模型	26
2.4 关系数据模型	28
2.4.1 关系模型的基本概念和结构	28
2.4.2 关系模型的数据完整性约束	30
2.4.3 关系模型的数据操纵	30
2.4.4 关系模型与网状和层次模型的比较	31
2.5 面向对象数据模型	32
2.6 小结	33

习题	34
第 3 章 关系数据库	35
3.1 关系模型的基本概念	35
3.1.1 关系的定义	35
3.1.2 关系模式和关系数据库	37
3.1.3 键	37
3.1.4 完整性约束	38
3.2 关系代数	39
3.2.1 传统的集合运算	40
3.2.2 专门的关系运算	41
3.2.3 扩充的关系运算	44
3.2.4 举例	46
3.2.5 ISBL 语言	47
3.3 元组关系演算	48
3.3.1 元组关系演算简介	48
3.3.2 元组关系演算语言 ALPHA	50
3.4 域关系演算	52
3.4.1 域关系演算简介	52
3.4.2 域关系演算语言 QBE	53
3.4.3 关系运算的安全限制和三种关系运算的等价性	56
3.5 小结	58
习题	58
第 3 章 关系数据库标准语言 SQL	60
4.1 SQL 简介	60
4.2 SQL 的系统结构	61
4.3 SQL 的数据定义	61
4.3.1 SQL 模式的定义和删除	62
4.3.2 基本表的定义、修改和删除	62
4.3.3 索引的建立和删除	65
4.4 SQL 的数据操纵	66
4.4.1 数据查询	66
4.4.2 数据更新	74
4.5 SQL 中的视图	76
4.5.1 视图的定义	77
4.5.2 视图上的操作	78
4.5.3 视图的优点	80
4.6 SQL 的数据控制	81

4.6.1 授权	82
4.6.2 权限回收	83
4.7 嵌入式 SQL	83
4.7.1 嵌入式 SQL 与主语言的接口	84
4.7.2 不用游标的嵌入式 SQL	85
4.7.3 用游标的嵌入式 SQL	87
4.7.4 嵌入式 SQL 应用实例	89
4.7.5 动态 SQL	91
4.8 小结	92
习题	92
第 5 章 查询处理和查询优化	94
5.1 关系数据库系统的查询处理	94
5.1.1 查询处理过程	94
5.1.2 执行查询操作的基本算法	95
5.2 关系数据库系统的查询优化	99
5.2.1 查询优化技术	99
5.2.2 查询优化实例	100
5.3 代数优化	102
5.3.1 关系代数表达式的等价变换规则	103
5.3.2 代数优化策略	105
5.3.3 代数优化算法	105
5.4 基于存取路径的优化	108
5.5 基于代价估算的优化	109
5.5.1 选择操作的代价估算	110
5.5.2 连接操作的代价估算	111
5.6 小结	112
习题	112
第 6 章 数据库的安全性	114
6.1 计算机安全性概述	114
6.2 数据库安全性概述	116
6.3 用户标识与鉴别	117
6.4 存取控制	118
6.4.1 自主存取控制	118
6.4.2 强制存取控制	123
6.5 视图机制	124
6.6 数据加密	126
6.7 数据库审计	127

6.8	统计数据库的安全性	127
6.9	SQL Server 的安全控制	128
6.9.1	SQL Server 的安全体系结构	128
6.9.2	登录管理	129
6.9.3	数据库用户管理	131
6.9.4	权限管理	132
6.9.5	角色管理	134
6.9.6	审计	136
6.10	小结	137
	习题	138
第 7 章	数据库的完整性	139
7.1	数据库的完整性概述	139
7.1.1	完整性约束条件	139
7.1.2	实现数据完整性的方法	141
7.2	实体完整性	141
7.2.1	实体完整性的定义	142
7.2.2	实体完整性检查和违约处理	143
7.3	参照完整性	143
7.4	用户定义的完整性	146
7.5	触发器	148
7.6	SQL Server 中数据库完整性的实现	152
7.7	小结	157
	习题	158
第 8 章	数据库恢复技术	160
8.1	事务的基本概念和特征	160
8.1.1	事务的基本概念	160
8.1.2	事务特征	160
8.1.3	事务状态	162
8.1.4	事务原子性和持久性的实现	163
8.1.5	事务的并发运行	163
8.2	数据库恢复的必要性	164
8.3	数据库恢复策略	164
8.4	数据转储与恢复	166
8.5	基于日志的数据库恢复	168
8.5.1	数据库系统日志文件	168
8.5.2	使用日志恢复数据库	168
8.6	检查点恢复技术	171

8.7 数据库镜像恢复技术	172
8.8 SQL Server 的数据恢复机制	173
8.8.1 SQL Server 中的事务	173
8.8.2 备份和恢复	174
8.9 小结	179
习题	180
第9章 并发控制	181
9.1 并发事务运行存在的异常问题	181
9.2 并发调度的可串行性	183
9.2.1 可串行化调度	183
9.2.2 调度的冲突等价性	184
9.2.3 调度的状态等价性	185
9.2.4 调度的可串行性测试	186
9.3 基于封锁的并发控制技术	187
9.3.1 锁	187
9.3.2 封锁协议	188
9.3.3 活锁	189
9.3.4 死锁	190
9.3.5 两阶段封锁协议	192
9.3.6 锁表	192
9.4 多粒度封锁	194
* 9.5 基于时间戳协议的并发控制	196
9.5.1 时间戳	196
9.5.2 时间戳协议	196
* 9.6 基于有效性确认的并发控制	198
9.7 插入与删除操作对并发控制的影响	200
9.8 SQL Server 中的并发控制	202
9.8.1 事务的隔离级别	202
9.8.2 专用锁	203
9.8.3 锁的使用与管理	204
9.9 小结	205
习题	206
第10章 关系数据库设计理论	208
10.1 关系模型的存储异常	208
10.2 函数依赖	210
10.2.1 函数依赖的定义	210
10.2.2 函数依赖的蕴涵性	212

10.3	函数依赖公理	212
10.3.1	Armstrong 公理	212
10.3.2	函数依赖集的等价和覆盖	216
10.4	模式分解	218
10.4.1	无损连接分解	219
10.4.2	分解的保持依赖性	222
10.5	关系模式的规范化	223
10.5.1	第一范式	224
10.5.2	第二范式(2NF)	225
10.5.3	第三范式	226
10.5.4	Boyce-Codd 范式(BCNF)	227
10.5.5	模式分解算法	228
10.6	多值依赖和 4NF	230
10.6.1	多值依赖	230
10.6.2	4NF	233
* 10.7	连接依赖和投影-连接范式(Project-Join NF)	234
10.7.1	连接依赖	234
10.7.2	投影-连接范式(Project-Join NF)	235
10.8	小结	236
	习题	236
第 11 章	数据库设计	238
11.1	数据库设计方法	238
11.2	数据模型与数据建模	240
11.3	IDEF1X 数据建模方法	242
11.3.1	数据模型的结构	243
11.3.2	逻辑模型	243
11.3.3	物理模型	244
11.4	IDEF1X 的语法和语义	244
11.5	IDEF1X 建模过程	250
11.5.1	阶段 0——设计的开始	250
11.5.2	阶段 1——定义实体	251
11.5.3	阶段 2——定义联系	252
11.5.4	阶段 3——定义键	253
11.5.5	阶段 4——定义属性	254
11.6	ERwin 数据建模	254
11.6.1	ERwin 的工作空间	254
11.6.2	建立实体联系	255
11.6.3	两个实体的多个联系的处理	256

11.6.4	递归联系	258
11.6.5	分类联系	258
11.6.6	使用域简化数据类型的设置	260
11.6.7	将数据模型导入到数据库	261
11.7	合同管理系统数据建模	263
11.7.1	合同管理应用需求	263
11.7.2	合同管理应用系统功能需求	265
11.7.3	实体的确定	267
11.7.4	联系的确定	267
11.7.5	确定属性	268
11.8	小结	269
	习题	269
第 12 章	数据库编程	271
12.1	Transact-SQL	271
12.1.1	Transact-SQL 元素	272
12.1.2	过程的类型	275
12.1.3	变量和参数	277
12.1.4	控制流程	277
12.1.5	错误处理	278
12.2	Transact-SQL 游标	280
12.2.1	游标的基本概念与操作	281
12.2.2	处理游标中的行	283
12.3	Transact-SQL 存储过程	285
12.3.1	什么是存储过程	285
12.3.2	存储过程的类型	286
12.3.3	设计与实现存储过程	287
12.4	Transact-SQL 用户定义函数	290
12.4.1	多语句表值函数	292
12.4.2	标量函数	293
12.4.3	内联表值函数	294
12.5	Transact-SQL 触发器	295
12.5.1	Transact-SQL 触发器基本概念	295
12.5.2	DML 触发器	296
12.5.3	deleted 表和 inserted 表	297
12.5.4	AFTER 触发器	297
12.5.5	INSTEAD OF 触发器	300
12.6	ADO.NET	303
12.6.1	数据提供程序	306

12.6.2	数据集	308
12.7	小结	309
	习题	310
第 13 章	数据库的存储结构	311
13.1	数据库存储设备	311
13.1.1	物理存储设备概述	311
13.1.2	存储器的层次结构	312
13.1.3	数据库的存储体系	313
13.1.4	磁盘容错技术	314
13.2	记录的存储结构	315
13.3	文件的存储结构	317
13.3.1	无序文件	318
13.3.2	顺序文件	319
13.3.3	散列文件	320
13.3.4	多表聚集文件	322
13.4	索引文件	323
13.4.1	索引概述	323
13.4.2	稀疏索引和稠密索引	325
13.4.3	聚集索引和辅助索引	325
13.4.4	B+树索引	327
13.4.5	散列索引	330
13.5	典型 DBMS 的存储结构	331
13.5.1	SQL Server 的存储结构	331
13.5.2	Oracle 的存储结构	332
13.6	小结	333
	习题	334
第 14 章	分布式数据库系统	335
14.1	分布式数据库系统概述	335
14.1.1	分布式数据库系统的定义	335
14.1.2	分布式数据系统的基本特征	336
14.1.3	分布式数据库系统的组成	337
14.1.4	分布式数据库的模式结构	338
14.1.5	分布式数据库系统的分类	339
14.2	数据分布和分布透明性	339
14.2.1	数据分片	339
14.2.2	数据分布	340
14.2.3	分布透明性	341

14.3	分布式查询处理和优化	343
14.3.1	分布式查询的分类	343
14.3.2	分布式查询处理过程	344
14.3.3	分布式查询优化	345
14.4	分布式事务管理	347
14.4.1	分布式事务恢复	348
14.4.2	分布式并发控制	350
14.5	分布式目录管理	352
14.6	小结	352
	习题	353
第 15 章	对象和对象关系数据库	354
15.1	概述	354
15.2	面向对象数据库	356
15.2.1	面向对象数据模型	357
15.2.2	面向对象数据库语言	359
15.2.3	面向对象数据库系统	362
15.3	对象关系数据库	363
15.3.1	对象关系数据模型	363
15.3.2	对象关系数据库系统	368
15.4	小结	368
	习题	368
第 16 章	多媒体数据库	369
16.1	多媒体数据库的特点	369
16.2	系统体系结构	370
16.2.1	多媒体数据库系统的层次结构	370
16.2.2	多媒体数据库系统的组织结构	371
16.3	多媒体数据模型	373
16.3.1	数据模型的需求	373
16.3.2	通用数据模型	373
16.4	多媒体数据的查询	374
16.5	特征提取、索引和相似性度量	376
16.6	QoS 保证	377
16.7	多媒体数据库的实现	378
16.8	其他问题	380
16.9	小结	381
	习题	382

第 17 章 数据库新技术与新应用	383
17.1 数据库新技术	384
17.1.1 面向对象数据库	384
17.1.2 实时数据库	385
17.1.3 主动数据库	386
17.1.4 分布式数据库	386
17.1.5 数据挖掘	387
17.1.6 多媒体数据库	387
17.2 并行数据库	387
17.2.1 并行数据库系统的体系结构	387
17.2.2 并行处理技术	389
17.2.3 商用并行数据库系统的并行策略	390
17.3 主动数据库	391
17.4 空间数据库	395
17.4.1 基本概念	395
17.4.2 空间数据操作	398
17.4.3 空间数据建模	398
17.4.4 空间数据索引	400
17.5 XML 数据库	401
17.5.1 原生 XML 数据库	402
17.5.2 XML 数据库的研究问题	405
17.6 小结	407
习题	407
参考文献	408

数据库系统引论

数据库是 20 世纪 60 年代末发展起来的数据处理新技术,它的出现使数据处理进入了一个崭新时代。数据库技术被广泛地用于信息管理、办公自动化、计算机辅助设计、计算机辅助决策等。时至今日,数据库已成为管理和应用信息资源不可缺少的重要工具。数据库的应用提高了人们的工作效率,产生了极大的社会 and 经济效益。同时,数据库的广泛应用也进一步促进了数据库技术的发展,使其成为计算机科学中发展最快的一个领域。本章介绍数据库系统涉及的基本概念,通过本章的学习,将对数据库有一个概括的了解。首先,回顾一下数据处理技术的发展。

1.1 数据管理技术的发展

所谓数据是表示信息的符号,可以是数字、文字、图形、图像、声音等。对数据进行收集、加工、应用、储存和传播等的一系列过程称为数据处理。

用计算机进行数据处理始于 20 世纪 50 年代初,那时的计算机不但能够处理数字,也具有处理字母的能力。人们用卡片、纸带或磁带作为存储介质把数据输入到计算机中进行处理,然后把处理结果打印出来。随着计算机软、硬件的发展,数据管理技术也在不断地发展。从最初的数据处理到现在的数据库管理,数据管理技术经历了人工管理、文件系统和数据库系统三个阶段。

1.1.1 人工管理阶段

20 世纪 50 年代中期以前,计算机主要用于科学计算。这一阶段,计算机除硬设备外没有任何数据管理软件,数据处理方式是批处理。这一阶段的数据处理具有如下特点:

(1) 数据不保存

处理的数据由应用程序读入内存,经过处理后即将结果输出。数据和程序都不保存在计算机中。

(2) 没有专用的软件管理数据

应用程序中除了要规定数据的逻辑结构外,还要考虑数据在计算机中如何存储和组织,

为数据分配空间、决定存取方法等。

(3) 应用程序完全依赖于数据

由于应用程序需要管理数据的逻辑结构和物理结构,数据结构的改变,存储与存取方法的变化,都会使应用程序跟着改变。这种应用程序与数据的存储、存取方式密切相关的情况,称为“数据依赖”。

(4) 数据不能共享

应用程序与数据是一一对应的。如果几个应用要用到同一数据,这些数据要重复存储,数据的冗余度很大。

这一阶段,数据的管理还是手工的、分散的,处理效率很低。数据与程序之间的关系如图 1-1 所示。

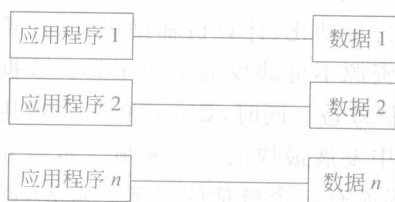


图 1-1 手工管理阶段数据与程序间的关系

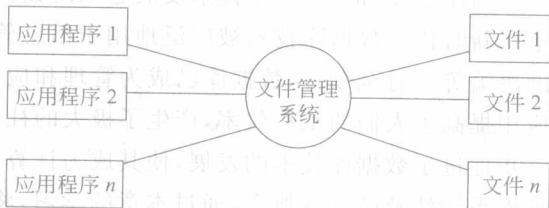


图 1-2 文件系统阶段数据与程序间的关系

1.1.2 文件系统阶段

20 世纪 50 年代后期,外设有了磁盘、磁鼓等直接存取的存储设备。与此同时,计算机软件也在迅速发展,出现了操作系统软件,数据可以以文件的形式存储在外存中。随着数据处理的日渐增多,有了专用于数据处理的软件,称为文件管理系统。

文件的应用,是计算机数据处理的重大进展。数据文件可以按名引用,应用程序通过文件管理系统与数据文件发生联系,数据的物理结构与逻辑结构间有了简单变换。在应用程序中,可以不必过多地考虑数据的物理存储细节,简化了程序员的数据管理工作。同时,一个应用程序可以和几个数据文件发生联系,增加了数据处理的灵活性。数据处理的方式有批处理,也有联机处理。但是,数据仍然是分散的,是面向应用的。文件系统阶段数据与程序间的对应关系如图 1-2 所示。

文件系统阶段数据处理有如下特点:

(1) 数据可以长期保存

大量的数据以文件形式保存在磁介质上,可以对数据进行查询、插入、删除和修改等操作。

(2) 有专门的文件系统软件管理数据

应用程序与数据间的依赖关系得到了改善。应用程序通过文件系统存取数据文件,不再管理数据的物理存储,程序与数据间具有一定的物理独立性。

(3) 数据是面向应用的

在文件系统阶段,应用程序与数据间不再是一一对应,部分系统允许应用程序存取多个文件中的数据。但数据还是面向应用的,文件间相互独立,缺乏联系。数据的记录结构与应用相对应,应用程序与数据间仍然存在着依赖关系。一旦数据结构改变,相应的应用程序就