

数据库系统 原理与应用教程 (第三版)

- ◆ 数据库基本概念
- ◆ 数据库建模方法
- ◆ 关系模型和关系代数
- ◆ SQL 语言基础和高级功能
- ◆ 索引技术和查询技术
- ◆ 事务处理和并发控制
- ◆ 分布式数据库技术
- ◆ 数据仓库和数据挖掘技术
- ◆ 面向对象数据库技术



闪四清 编著



清华大学出版社

高等学校计算机应用规划教材

数据库系统原理与应用教程

(第三版)

闪四清 编著

清华大学出版社

北 京

内 容 简 介

数据库技术是信息技术的一个重要组成部分。本书全面介绍了数据库技术基本原理、技术实现和应用等知识。全书包括4部分共16章。第一部分对数据库的基本概念和主流的数据库建模方法进行了研究,第二部分全面介绍了关系代数和SQL语言,第三部分涉及数据库技术深层次的实现,第四部分详细阐述了数据库技术的发展趋势。

本书内容全面、概念清晰、语言流畅、图文并茂,可作为高等院校计算机相关专业的数据库课程教材,也可供从事数据库应用程序开发工作的科研人员和工程技术人员参阅,例如管理信息系统、电子商务网站开发等。

本书的电子教案可以到<http://www.tupwk.com.cn/downpage> 网站下载。

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

数据库系统原理与应用教程/闪四清 编著. —3版. —北京:清华大学出版社,2008.4

(高等学校计算机应用规划教材)

ISBN 978-7-302-17185-0

I. 数… II. 闪… III. 数据库系统—高等学校—教材 IV. TP311.13

中国版本图书馆CIP数据核字(2008)第031397号

责任编辑:胡辰浩(huchenhao@263.net) 袁建华

封面设计:孔祥丰

版式设计:康 博

责任校对:成凤进

责任印制:王秀菊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦A座

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

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

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

印 刷 者:清华大学印刷厂

装 订 者:三河市金元印装有限公司

经 销:全国新华书店

开 本:185×260 印 张:23.25 字 数:580千字

版 次:2008年4月第1版 印 次:2008年4月第1次印刷

印 数:1~5000

定 价:35.00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:026664-01

前 言

当今是一个信息爆炸的时代，信息已经成为社会和经济发展的主要支柱之一。大量信息的产生、处理、存储、传播和使用推动了社会的进步和经济的发展。信息系统是一种以加工处理信息为主的计算机系统。数据库技术作为一种存储和使用信息的信息系统核心技术发挥着越来越重要的作用。例如，现在的银行无法再依赖手工记账来处理复杂的储蓄、结算、信贷和信用卡等业务，邮政电信也无法离开信息系统开展各种电信业务，航空运输也无法离开信息系统来管理庞大的分布式售票网点，加工企业无法离开信息系统来管理庞大的零部件仓库、计划、生产、销售和财务等，图书出版业无法依赖手工的方式管理大量图书的出版、库存和销售等业务。

本书对数据库技术进行了全面的阐述和研究。在结合了大量的实例和作者的教学体会基础上，对数据库技术的各个领域进行了深入浅出的剖析，对数据库技术的重点和难点进行了详细的描述，力求做到思路清晰、概念准确、结构合理、内容生动活泼。

本书包括了 4 部分，共 16 章内容。第一部分是建模篇，重点介绍了数据库的基本概念以及 ER 图、ODL 方法、IDEF1x 图、关系模型等数据库建模技术。第二部分是编程篇，对数据库技术中的关系代数、Datalog 语言等理论以及标准 SQL 语言进行了讲述。第三部分是实现篇，讨论了有关索引、查询、事务和并发控制等数据库技术内核的实现问题。第四部分对分布式数据库技术、数据仓库技术和面向对象数据库技术等内容进行了深入分析。

每一章的开始都对该章将要涉及的内容及其作用进行了分析，然后指出了学完本章读者应该掌握的重要内容。每一章的正文内容后的小结对本章所讲述的内容进行总结和评价。另外，每一章后面都附有一个思考和练习。思考题侧重于加深读者对本章涉及概念的理解，练习题侧重于培养学生应用本章学到的知识来解决实际问题的能力。

本教程另外一个显著的特点是以一个贯穿全书始末的图书管理数据库示例为主线，把各种数据库技术的知识要点，串联成一个逻辑严密的整体。该示例不断深化、丰富和完善的过程，实际上就是读者不断学习、理解和掌握数据库技术的过程。

经过本次修订，希望各位读者可以从本书中学习到自己需要的知识，真正掌握数据库系统的理论和应用。

本书是集体智慧的结晶，参加本书编写和制作的人员还有陈笑、洪妍、方峻、何亚军、王通、高娟妮、严晓雯、杜思明、孔祥娜、张立浩、孔祥亮、牛静敏、何俊杰、葛剑雄等人。由于作者水平有限，加之创作时间仓促，本书不足之处在所难免，欢迎广大读者批评指正。我们的邮箱是：huchenhao@263.net。

作者

2008年2月

目 录

第一部分 设计篇

第 1 章 步入数据库系统世界	3
1.1 概述	3
1.1.1 数据库系统应用示例	3
1.1.2 数据库系统的作用	6
1.1.3 常见的数据库系统	8
1.1.4 数据库应用开发过程 和数据库开发过程	12
1.2 数据库技术的演变	15
1.2.1 数据库技术环境的特点	15
1.2.2 传统的文件管理阶段	16
1.2.3 现代的数据库管理系统阶段	17
1.3 数据库系统架构	19
1.3.1 数据存储架构	19
1.3.2 数据视图管理架构	21
1.3.3 数据库应用架构	22
1.4 数据库管理系统组件	24
1.4.1 数据和元数据	24
1.4.2 输入	26
1.4.3 查询处理器	26
1.4.4 存储管理器	27
1.4.5 事务管理器	27
1.5 数据库技术的发展趋势	28
1.5.1 数据库技术面临的挑战	28
1.5.2 数据库技术的发展趋势	28
1.6 本章小结	30
1.7 思考和练习	30
第 2 章 数据库建模——ER 图	32
2.1 ER 图的基本概念	32
2.1.1 实体集和实体集中的属性	32

2.1.2 联系集和联系集中的属性	36
2.1.3 绘制 ER 图	37
2.2 ER 图设计原则	41
2.2.1 真实性原则	41
2.2.2 避免冗余原则	42
2.2.3 简单性原则	42
2.2.4 恰当性原则：用实体集 还是用属性	42
2.2.5 恰当性原则：用实体集 还是用联系集	43
2.3 对约束建模	43
2.3.1 映射约束	44
2.3.2 键码约束	47
2.3.3 参照完整性约束	48
2.3.4 其他类型的约束	49
2.4 弱实体集	50
2.4.1 弱实体集产生的原因	50
2.4.2 对弱实体集的要求	51
2.4.3 弱实体集表示方法	52
2.5 ER 图的高级技术	52
2.5.1 特殊化技术	52
2.5.2 概括技术	54
2.5.3 属性继承技术	54
2.5.4 概括约束设计	55
2.5.5 聚集技术	56
2.6 本章小结	57
2.7 思考和练习	57
第 3 章 数据库建模——IDEF1x 图	59
3.1 概述	59
3.1.1 IDEF 方法族简介	59
3.1.2 IDEF1x 图的概念和特点	60

3.2	IDEF1x 图的语法和语义	61	4.5	思考和练习	114
3.2.1	实体	62	第 5 章 关系模型		116
3.2.2	连接联系	63	5.1	基本概念	116
3.2.3	分类联系	66	5.2	从 ER 图到关系模型设计 的转换	122
3.2.4	非确定联系	68	5.2.1	ER 图和关系模式的比较	122
3.2.5	属性	69	5.2.2	实体集到关系模式的转换	122
3.2.6	主键码和次键码	71	5.2.3	联系集到关系模式的转换	123
3.2.7	外键码	72	5.2.4	弱实体集到关系模式的转换	125
3.3	使用 IDEF1x 图建立数据库 模型的过程	74	5.2.5	ER 图中其他元素的转换	127
3.3.1	准备阶段	75	5.2.6	键码的转换	128
3.3.2	定义实体阶段	77	5.3	从 IDEF1x 图到关系模型 设计的转换	129
3.3.3	定义联系阶段	78	5.4	从 ODL 设计到关系模型 设计的转换	130
3.3.4	定义键码阶段	80	5.4.1	ODL 设计和关系模式 设计的比较	130
3.3.5	定义属性阶段	85	5.4.2	从 ODL 属性到关系 模式的属性	131
3.3.6	文件编制阶段	86	5.4.3	ODL 设计中非原子属性 的转换	131
3.4	本章小结	88	5.4.4	ODL 设计中联系的转换	135
3.5	思考和练习	88	5.4.5	ODL 设计中联系和反向 联系问题	136
第 4 章 数据库建模——ODL 方法		89	5.4.6	ODL 设计中子类结构的转换	137
4.1	概述	89	5.5	本章小结	138
4.1.1	面向对象模型和关系 模型比较	89	5.6	思考和练习	138
4.1.2	面向对象基本概念	90	第 6 章 关系模式的规范化设计		141
4.2	ODL 的基本结构	92	6.1	概述	141
4.2.1	ODL 概念	92	6.1.1	异常问题	141
4.2.2	ODL 接口说明	93	6.1.2	泛关系模式和数据库模式	142
4.2.3	ODL 中的属性	94	6.2	函数依赖	143
4.2.4	ODL 中的联系	95	6.2.1	函数依赖的定义	143
4.2.5	反向联系的概念	96	6.2.2	函数依赖和键码	144
4.2.6	联系的多重性	98	6.2.3	逻辑蕴含	145
4.2.7	ODL 中的类型	100	6.2.4	函数依赖的推理规则	146
4.2.8	ODL 中的方法署名的说明	103			
4.3	子类和约束	105			
4.3.1	ODL 中的子类	105			
4.3.2	ODL 中的约束	109			
4.4	本章小结	114			

6.2.5	属性集的闭包	147
6.2.6	正则覆盖	149
6.3	关系模式的分解	149
6.4	关系模式的范式	151
6.4.1	第一范式	151
6.4.2	BCNF 范式	152
6.4.3	分解成 BCNF 模式的算法	152
6.4.4	函数依赖的投影	154
6.4.5	第三范式	154
6.4.6	第二范式	156
6.5	多值依赖	157
6.5.1	多值依赖的概念	157
6.5.2	多值依赖的推理规则	158
6.5.3	第四范式	158
6.6	范式之间的关系	159
6.7	数据库模式的实例	160
6.7.1	图书管理数据库模式	160
6.7.2	计算机产品信息管理 数据库模式	162
6.8	本章小结	164
6.9	思考和练习	164

第二部分 编程篇

第 7 章	关系代数基本理论	169
7.1	概述	169
7.2	关系代数的基本运算	170
7.2.1	集合运算	170
7.2.2	选择运算	173
7.2.3	投影运算	174
7.2.4	笛卡尔积运算	176
7.2.5	改名运算	177
7.3	关系代数的附加运算	178
7.3.1	自然连接运算	178
7.3.2	θ 连接运算	179
7.3.3	除法运算	181

7.3.4	赋值运算	182
7.4	扩充的关系代数运算	183
7.4.1	广义投影运算	183
7.4.2	外连接运算	183
7.4.3	聚集运算	185
7.5	关系演算	185
7.5.1	元组关系演算	186
7.5.2	域关系演算	186
7.6	关系代数的更新运算	187
7.6.1	插入运算	187
7.6.2	删除运算	188
7.6.3	更新运算	188
7.7	关系代数表达式的优化策略	189
7.8	本章小结	190
7.9	思考和练习	190
第 8 章	Datalog 语言	192
8.1	基本概念	192
8.1.1	基本结构	192
8.1.2	一般规则	193
8.1.3	安全规则	195
8.1.4	外延谓词和内涵谓词	197
8.2	关系代数向 Datalog 规则的 转换	197
8.2.1	从集合运算到 Datalog 规则	197
8.2.2	从投影运算到 Datalog 规则	198
8.2.3	从笛卡尔乘积到 Datalog 规则	199
8.2.4	从选择运算到 Datalog 规则	199
8.2.5	从连接运算到 Datalog 规则	201
8.2.6	从多重运算到 Datalog 规则	201
8.3	递归原理	202
8.3.1	关系代数存在的问题	202
8.3.2	计算最小固定点	203
8.3.3	使用 Datalog 规则表示 固定点公式	205
8.4	包的运算	205
8.4.1	包的意义	206

8.4.2	包的关系运算	207
8.4.3	包的逻辑运算	211
8.5	本章小结	211
8.6	思考和练习	211
第9章	SQL 语言初步	213
9.1	概述	213
9.2	简单查询语句	214
9.2.1	最简单的查询语句	215
9.2.2	投影	216
9.2.3	选择	218
9.2.4	字符串比较	221
9.2.5	日期和时间的比较	222
9.2.6	排序结果	222
9.2.7	副本	224
9.3	高级查询语句	225
9.3.1	连接查询	225
9.3.2	集合查询	231
9.3.3	合计查询	234
9.3.4	子查询	236
9.4	数据库操纵语句	238
9.4.1	插入数据	239
9.4.2	删除数据	240
9.4.3	更新数据	240
9.5	数据定义语句	241
9.5.1	数据类型和域	242
9.5.2	管理表	243
9.5.3	默认值	245
9.5.4	索引定义	246
9.5.5	管理视图	246
9.6	递归查询	249
9.7	动手练习	250
9.8	本章小结	251
9.9	思考和练习	251
第10章	SQL 语言高级功能	253
10.1	约束	253
10.1.1	键码约束	253

10.1.2	外键码约束	254
10.1.3	非空约束	256
10.1.4	CHECK 约束	256
10.1.5	断言约束	258
10.2	触发器	260
10.2.1	触发器的结构	260
10.2.2	触发器示例	262
10.3	游标	262
10.3.1	定义游标	262
10.3.2	游标的选项	264
10.4	嵌入式 SQL 语言	265
10.4.1	接口和声明	265
10.4.2	动态的 SQL 语句	266
10.5	安全控制和权限管理	267
10.5.1	权限的类型	267
10.5.2	授权和收权	268
10.6	动手练习	269
10.7	本章小结	270
10.8	思考和练习	270

第三部分 实现篇

第11章	索引和散列技术	275
11.1	概述	275
11.1.1	文件组织方式	275
11.1.2	顺序文件组织	276
11.1.3	聚集文件组织	277
11.2	索引技术	278
11.2.1	基本概念	278
11.2.2	顺序索引	278
11.2.3	B ⁺ 树索引文件	282
11.3	散列技术	283
11.3.1	基本概念	283
11.3.2	散列索引	284
11.4	Microsoft SQL Server 系统中的索引	284

11.4.1	索引类型	284	13.3.2	事务的工作原理	309
11.4.2	创建索引的方法	286	13.3.3	锁的粒度	311
11.5	本章小结	287	13.3.4	锁的类型	314
11.6	思考和练习	287	13.3.5	死锁和死锁的解决方案	316
第 12 章	查询处理技术	288	13.4	本章小结	317
12.1	概述	288	13.5	思考和练习	317
12.2	查询处理的代价模型	290	第四部分 提高篇		
12.3	单个关系运算的代价估计	291	第 14 章	分布式数据库技术	321
12.3.1	选择运算的代价估计	291	14.1	概述	321
12.3.2	连接运算的代价估计	292	14.2	分布式数据库的概念和特点	323
12.3.3	集合运算的代价估计	292	14.2.1	分布式数据库的概念	323
12.4	表达式运算的代价估计	293	14.2.2	分布式数据库的特点	324
12.4.1	实体化方法	293	14.3	分布式数据存储技术	326
12.4.2	流水线方法	294	14.3.1	数据复制	326
12.5	Microsoft SQL Server 系统 的查询处理器	294	14.3.2	数据分片	327
12.5.1	特性概述	294	14.3.3	数据混合存储	329
12.5.2	交互式操作	296	14.4	分布式数据查询技术	331
12.5.3	执行计划	296	14.4.1	分布式数据查询示例	331
12.5.4	优化索引	298	14.4.2	基于半连接的优化策略	332
12.5.5	优化隐藏	298	14.4.3	基于连接的优化策略	333
12.6	本章小结	300	14.5	Microsoft SQL Server 的分布式计算特点	333
12.7	思考和练习	300	14.5.1	Microsoft Distributed Transaction Coordinator	334
第 13 章	事务和并发控制	301	14.5.2	链接服务器	334
13.1	事务	301	14.5.3	数据复制	334
13.1.1	事务的基本概念	301	14.6	本章小结	337
13.1.2	事务的状态变迁图	302	14.7	思考和练习	338
13.1.3	事务的并发执行	303	第 15 章	数据仓库技术	339
13.1.4	事务的一致性级别	305	15.1	概述	339
13.2	并发控制	306	15.2	数据挖掘和数据中心库	341
13.2.1	并发操作带来的问题	306	15.2.1	数据挖掘技术	341
13.2.2	锁	306	15.2.2	数据中心库	344
13.2.3	死锁	307			
13.3	Microsoft SQL Server 系统 的事务和锁	308			
13.3.1	事务的类型	308			

15.3 数据仓库的支持工具 345

 15.3.1 OLAP 服务 345

 15.3.2 数据转换服务 346

 15.3.3 PivotTable 服务 347

15.4 数据仓库的实现步骤 347

 15.4.1 确定用户需求 347

 15.4.2 设计和建立数据库 348

 15.4.3 提取和加载数据 350

15.5 数据的粒度、分割和元数据 352

15.6 本章小结 352

15.7 思考和练习 353

第 16 章 面向对象数据库技术 354

 16.1 概述 354

 16.2 面向对象数据库技术架构 355

 16.2.1 大对象和外部软件 356

 16.2.2 专用媒体服务器 356

 16.2.3 对象数据库中间件 357

 16.2.4 用户定义类型的
 对象关系 DBMS 358

 16.2.5 面向对象 DBMS 359

 16.3 本章小结 360

 16.4 思考和练习 360

参考文献 361

第一部分 设计篇

- 步入数据库系统世界
- 数据库建模——ER 图
- 数据库建模——IDEF1x 图
- 数据库建模——ODL 方法
- 关系模型
- 关系模式的规范化设计

第1章 步入数据库系统世界

本书全面讲述数据库系统的基本原理和应用。通过本书的学习，读者在学习数据库系统的基本工作原理的同时，还可以掌握数据库产品的使用技术和应用开发技术。对于数据库技术专业人员来说，掌握“怎样做”的应用技术和“为什么这样做”的基本原理都非常重要。

本章提供了一个有关数据库系统的整体轮廓。首先，通过介绍一些典型的数据库应用，使读者了解数据库技术的作用。为了加深读者对数据库技术的理解，本章快速浏览了当前市场上的主流数据库产品主要特点，讨论了数据库和数据库应用开发过程。然后，回顾了数据库及其系统的发展简史。通过回顾数据库系统的发展历史，读者可以更好地理解数据库系统当前的应用状况、存在的问题和发展趋势。接下来，对数据库系统架构进行了深入研究，具体内容包括数据存储架构、数据管理架构和应用开发架构。第四，对数据库管理系统的组件进行了剖析，把数据库系统的核心内容展示给读者。最后，总结数据库技术面临的挑战和机遇，展望数据库技术的发展趋势。

本章的学习目标：

- 了解数据库技术的应用现状和作用；
- 了解当前数据库市场上的主流产品的特点；
- 了解数据库和数据库应用开发的基本过程；
- 掌握数据库技术的发展历史和主要阶段的特征；
- 理解数据库中数据的存储、管理和应用架构；
- 理解数据库管理系统的架构和各个组成部分的作用；
- 了解当前数据库技术面临的挑战；
- 理解数据库技术的发展趋势。

1.1 概述

信息技术的发展极大地推动了社会的发展。作为信息技术主要支柱之一的数据库技术在社会各个领域中有广泛和深入的应用。数据库技术可以为各种用户提供及时的、准确的、相关的信息，满足用户的不同需要。

1.1.1 数据库系统应用示例

现在，通过研究几个数据库系统应用的典型示例，使读者对该信息技术的支撑技术有一个感性的认识。

1. 航空售票管理信息系统

航空售票管理信息系统是最早使用数据库技术的应用领域之一。在该系统中,主要包括以下数据项:

- (1) 座位预定信息: 座位分配、座位确认、座位状态、餐饮选择等;
- (2) 航班信息: 航班号、飞机型号、机组号、起飞地、目的地、起飞时间、到达时间、飞行状态等;
- (3) 机票信息: 票价、折扣、有无等。

对该系统的查询主要是在某一段时间内从某个指定城市到另一个指定城市的航班,是否还有可以选择的座位、是否有其他飞机型号、飞机票价以及是否折扣等信息。对该系统的主要更新操作包括为乘客登记航班、分配座位和选择餐饮等。任何时候,都会有许多航空售票代理商访问这些数据,并且要避免出现多个代理商同时卖出同一个座位的情况。这些数据还可以自动统计和分析经常乘坐某一个航班的乘客信息,为旅客提供特殊的优惠服务。

如果没有使用数据库技术,那么系统会因为数据量庞大和数据更新缓慢,航空部门无法为旅客提供及时、准确、有竞争力的服务。

2. 银行业务管理信息系统

银行业务的繁忙状态是社会经济发展速度的重要标志之一。银行业务管理信息系统也是最早使用信息技术(包括数据库技术)的系统之一。在银行业务管理信息系统中,典型的数据项包括以下几方面:

- (1) 顾客信息: 姓名、身份证号码、地址、电话等信息;
- (2) 账户信息: 账号、存款、余额、取款、日期等信息;
- (3) 顾客和账户关系信息: 顾客、账户等信息。

对银行业务管理信息系统的操作既可以通过各地的银行营业网点,也可以通过安装在各地的 ATM(Automated Teller Machine, 自动出纳机)进行。该系统的主要查询操作包括询问顾客的账户、账户的余额以及更新账户的数据等。同航空售票管理信息系统类似,允许对同一个账户进行并发访问而不会出现任何错误,非常重要。即使系统发生了故障,例如 ATM 突然断电了,正在处理的账户数据也不会出现任何不一致的记录。当前的数据库技术已经可以完全解决这种表面简单而实质复杂的事务并发难题。

使用了信息技术和数据库技术的银行业务管理信息系统为人们的生活和工作带来了巨大的便利。例如,人们可以在不同地点的任何银行网点存款和取款,从而避免随身携带大量的货币,保证安全。可以快速地异地汇兑和结算,减少资金的在途时间,提高资金的使用效率。

3. 超市的商品管理和销售信息系统

超市购物由于其种类繁多的商品、合理的价格、选物的便利已经成为人们日常生活的一个重要组成部分。数据库技术是超市成功管理大量信息的重要技术之一。在超市的商品管理和销售信息系统中,主要的数据项如下。

(1) 销售信息：连锁商店、销售日期、销售时间、顾客、商品编码、商品名称、销售数量、总价等；

(2) 商品信息：商品编码、商品名称、单价、库存数量、进货数量、供应商名称、商品类型、商品货位等；

(3) 供应商信息：供应商名称、地址、商品、信誉等。

在超市的商品管理和销售信息系统中，主要操作是记录顾客的购买信息、查询超市现有商品的名称和数量、分析当天连锁商店的销售情况，确定进货的内容和商品的摆放位置等。超市的经营决策主要是依赖商品管理和销售信息系统中存储的大量数据，从这些表面似乎孤立的大量数据中发掘出真正有价值的商品销售规律，从而提高经营者的管理和决策水平。

4. 工厂管理信息系统

工厂管理信息系统是最早依据数据库技术建立的一种集成的信息系统。在这种管理信息系统中，主要包括下面一些数据项。

(1) 销售记录：产品名称、服务类型、客户名称、销售人员、销售日期、销售价格等；

(2) 产品信息：产品名称、产品规格、产品型号、产品成本等；

(3) 雇员信息：姓名、地址、工资、津贴等；

(4) 财务信息：合同编码、合同金额、销售价格、应收货款、应付货款等。

在该系统中，典型的查询操作包括打印雇员的工资、打印应收应付货款清单、打印销售人员的业绩、打印工厂的各种统计报表等。每进行一次采购和销售，收到每个账单、收据，雇员的聘用、解聘、提职、加薪等都将导致对数据库的更新。

一个典型的工厂管理信息系统的窗口如图 1-1 所示。在该典型的工厂管理信息系统导航窗口中，包括了进货管理、销售管理、仓库管理、账务系统、人事工资、系统维护等功能。在销售管理功能中，包括了销售订单、销售开票、销售收款、现款销售、销售退货等业务。使用该导航窗口，操作人员可以执行各种业务操作。在工厂管理信息系统的背后，是存储了大量业务数据的数据库管理系统。



图 1-1 一个典型的工厂管理信息系统的业务导航窗口

如果没有工厂管理信息系统,那么许多企业就会陷入混乱的状态:货款未到账却没有及时发现、财务报表不能及时提供、企业高层领导不了解确切的库存零部件和产品的数量和状态、作业计划的安排不符合实际情况等。

5. 学校教学管理信息系统

学校教学管理信息系统主要是涉及学生、教师、教室、课程和排课等信息的管理。该系统包括的典型数据项如下。

(1) 学生信息:学号、学生姓名、性别、班级、年龄、宿舍、电话、E-mail 地址等;

(2) 教师信息:教师编号、教师姓名、性别、年龄、学历、教研室、住址、电话、E-mail 地址等;

(3) 教室信息:教室编号、位置、教室类型、座位编号等;

(4) 课程信息:课程名称、教材名称、学时、学分等;

(5) 排课信息:课程名称、教室、班级、教师名称、上课时间等。

除了上面的信息之外,还包括学生选课、考试成绩等信息。典型的查询操作包括提供教室安排、学生成绩统计清单、教师工作量统计等,典型的更新操作包括记录学生选课、登记考试成绩、自动排课等操作。此类系统的关键在于保证正确存储和处理大量的教务数据,为学校各部门和院系的工作安排提供及时、准确的数据支持,减少错误发生。

6. 图书管理信息系统

图书管理信息系统也是数据库技术应用的又一个典型领域。在图书管理信息系统中,主要包括以下数据项。

(1) 图书信息:书号、书名、作者姓名、出版日期、类型、页数、价格、出版商名称等;

(2) 作者信息:姓名、身份证号、性别、出生日期、学历、住址、电话等;

(3) 出版社信息:名称、地址、社长、成立日期等;

(4) 读者信息:姓名、借书证号、书名、借书日期、借书数量等。

图书管理信息系统中典型的查询操作包括查看某种类型和书名的图书、浏览指定出版商出版的图书、检索指定作者的图书等,典型的更新操作包括登记新书信息、作者信息等。作为一个动辄存储几百万图书的大学图书馆,如果没有管理图书的信息系统,那么借阅一本书可能需要耗费大量的时间和精力。这种管理大量图书信息的管理信息系统的技术基础之一正是数据库技术。

1.1.2 数据库系统的作用

当初步了解了数据库系统的应用之后,下面总结一下数据库系统的作用。数据库是许多数据的集合,数据库管理系统(database management system, DBMS)是管理这些数据集合的计算机软件系统,数据库管理系统也可以简称为数据库系统。此处需要指出的是,数据库技术泛指与数据库系统直接相关的各种技术,如果没有特别声明,本书将数据库技术和数据库系统两个术语同等看待。一般地,数据库系统主要用来管理大量业务数据、定义数据库、执行数据库操作、控制多用户访问等。

1. 管理大量数据

前面那些数据库技术的应用示例表明数据库技术应用在管理大量数据的地方。使用数据库的第一个原因即存储大量的数据。大量的数据到底有多大呢？现在数据库市场上最先进的数据库产品能支持 10TB 量级的数据。10TB 级的数据是一个怎样的概念呢？下面是一些理解 10TB 数据概念的例子：

- 10TB 的数据是 10,995,116,277,760 个字节大小，大约可以存储 5,497,558,138,880 个汉字信息；
- 在航空售票管理信息系统中，大约可以存储 183 亿个航班信息；
- 在银行业务管理信息系统中，大约可以存储 38 亿个流水账户信息；
- 在超市商品管理和销售信息系统中，大约可以存储 26 亿个商品信息；
- 在工厂管理信息系统中，大约可以存储 32 亿个客户信息；
- 在学校教学管理信息系统中，大约可以存储 42 亿个学生信息；
- 在图书管理信息系统中，大约可以存储 45 亿本书的简要信息，即大约可以存储 2800 万本 400 页厚的图书的所有内容(文字部分)。

2. 数据库定义功能

为了存储大量数据，需要定义数据库和数据库中的各种架构，例如表、视图和索引等对象。对一个具体的数据库系统来说，通常允许用户使用一种被称为数据定义语言(data definition language, DDL)的专门语言来建立数据库，定义数据库的架构。并且还可以对这些数据库和数据库架构进行修改和删除。

例如，在 Microsoft SQL Server 数据库系统中，可以使用 Transact-SQL 语言中的下列 DDL 语句定义数据库和数据库中的各种对象。

- CREATE DATABASE: 创建数据库；
- ALTER DATABASE: 修改数据库；
- DROP DATABASE: 删除数据库；
- CREATE TABLE: 创建表；
- ALTER TABLE: 修改表；
- DROP TABLE: 删除表；
- CREATE VIEW: 创建视图；
- ALTER VIEW: 修改视图；
- DROP VIEW: 删除视图。

3. 执行数据库操作

当数据存储到数据库中之后，用户就可以使用其中的数据。数据库系统提供了可以查询数据库中数据的查询(Query)语言，该语言经常被称为数据操纵语言(data manipulation language, DML)。使用 DML 语言可以在数据库中执行以下操作：检索指定的数据、插入需要的数据、更新已经变化的或错误的数据库、删除无用的或不再需要的数据等。