



普通高等教育“十一五”规划教材

高等院校计算机科学与技术系列教材

# 数据库系统 原理与设计

揭廷红 边芮 卞静 编著

```
</table>  
</td>  
<td align="left" width="100%" bgcolor="#FFEECC" valign="bottom">  
</tr>  
</tr>  
</TR>
```

冶金工业出版社

普通高等教育“十一五”规划教材  
高等院校计算机科学与技术系列教材

# 数据库系统原理与设计

揭廷红 边 芮 卞 静 编著

北 京

冶金工业出版社

# 内 容 简 介

本书是根据普通高等教育“十一五”国家级规划教材的指导精神而编写的。

数据库技术是计算机科学的重要分支，在各行各业得到了广泛的应用，成为检索、存储、处理信息资源的主要手段。数据库课程是计算机科学教育中一个必不可少的部分，是高等院校理工科类各专业一门重要的计算机技术基础课程。本书详细介绍了数据库的基本理论与技术，包括数据模型、查询语言、数据库设计、数据库系统实现以及数据库性能保证等问题。主要内容包括数据库系统概述、实体-联系模型、关系数据模型、关系数据库标准语言 SQL、查询处理和查询优化方法、关系数据理论、数据库保护技术、数据恢复技术、并发控制、对象数据库系统、XML 数据库以及 PowerBuilder 与数据库的应用开发。

本书适用面广，既可作为高等院校相关专业的教材，也适合数据库领域的研究人员和工程技术人员参考使用。

## 图书在版编目 (CIP) 数据

数据库系统原理与设计 / 揭廷红, 边芮, 卞静编著.

北京: 冶金工业出版社, 2007.8

普通高等教育“十一五”规划教材

ISBN 978-7-5024-4362-7

I. 数… II. ①揭…②边…③卞… 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 (2007) 第 109134 号

出版人 曹胜利 (北京沙滩嵩祝院北巷 39 号, 邮编 100009)

责任编辑 程志宏

ISBN 978-7-5024-4362-7

广州锦昌印务有限公司印刷; 冶金工业出版社发行; 各地新华书店经销

2007 年 8 月第 1 版第 1 次印刷

787mm × 1092mm 1/16; 20.5 印张; 474 千字; 320 页

35.00 元

冶金工业出版社发行部 电话: (010) 64044283 传真: (010) 64027893

冶金书店 地址: 北京东四西大街 46 号 (100711) 电话: (010) 65289081

(本社图书如有印装质量问题, 本社发行部负责退换)

# 前 言

## 一、关于本书

本书是根据普通高等教育“十一五”国家级规划教材的指导精神而编写的。

在全球信息化大潮的推动下，我国的计算机产业发展迅速，对专业人才的需求日益迫切。数据库技术是计算机科学技术中发展最快的领域之一，也是应用最广的技术之一，成为计算机信息系统与应用系统的核心技术和重要基础。数据库系统原理与设计是计算机和信息类学科的专业必修课，在专业人才培养系列中，起着举足轻重的作用。

为满足高等院校相关专业的教学需要，同时也为相关行业的工程技术人员提供系统、全面、先进、实用的学习参考书，作者根据多年来的教学、科研经验，结合当前主流的关系型数据库的发展状况，编写了本书。希望本书的问世，能在不同程度上满足广大学子学习数据库理论与技术的需要，促进数据库技术在各行业的应用与推广。

## 二、本书结构

本书共分 13 章，具体内容安排如下：

第 1 章：绪论。主要介绍数据库系统、数据库系统的组成、数据库系统的体系结构以及数据模型。通过总体介绍，使读者掌握数据库系统的基本概念和组成架构，并对数据库技术有一个全面的了解。

第 2 章：实体-联系模型。主要介绍实体-联系模型的基本概念、图形表示及其应用扩展，以及将 E-R 图转化为表的方法。

第 3 章：关系数据模型。主要介绍关系数据模型的基本概念、关系数据模型的数据操作、关系代数以及关系演算等内容。

第 4 章：SQL。主要介绍 SQL 语言的特点及功能、数据操纵语言（DML）、数据定义语言（DDL）和数据控制语言（DCL）等内容。

第 5 章：查询处理和查询优化。主要介绍查询处理和查询优化技术。

第 6 章：关系数据理论。主要介绍规范化理论、数据依赖以及模式分解等内容。

第 7 章：数据库设计。主要介绍需求分析、概念设计、逻辑设计、物理设计、验证设计，并给出按照数据库生命周期设计的数据库设计实例：考勤管理系统的实现。

第 8 章：数据库保护。主要介绍信息安全评估标准、数据库安全性控制技术、SQL 的用户授权、数据库的完整性等内容。

第 9 章：数据恢复。主要介绍事务概念、事务特征、事务状态、故障类型、数据恢复技术、数据恢复策略以及远程数据备份等内容。

第 10 章：并发控制。主要介绍并发调度的可串行化过程、时间戳协议、有效性检查协议、封锁管理以及多粒度等内容。

第 11 章：对象数据库系统。主要介绍面向对象数据库系统和对象-关系数据库系统等内容。

第 12 章: XML 数据库。主要介绍 XML 数据模型、XML 数据库查询语言 XPath 和 XQuery 查询语言等内容。

第 13 章: PowerBuilder 与数据库应用开发。主要介绍 PowerBuilder 的术语、PowerScript 语言以及使用 PowerBuilder 开发数据库应用程序等内容。

### 三、本书特点

本书内容充实、概念清晰、重点突出、深入浅出、实例丰富, 先进性、实用性与系统性并重。为了巩固所学知识, 每章配有综合练习题。通过本书的学习, 能够使读者充分掌握和理解数据库系统的基本概念、原理和方法, 掌握数据库设计、数据库语言和数据库实现技术, 能用数据库技术完成实际信息管理系统的开发与设计。

### 四、本书适用对象

本书适用面广, 既可作为高等院校相关专业的教材, 也适合数据库领域的研究人员和工程技术人员参考使用。

由于编写时间仓促, 编者水平有限, 书中疏漏和不足之处在所难免, 恳请广大读者批评指正。联系方式如下:

电子邮箱: [service@cnbook.net](mailto:service@cnbook.net)

网址: [www.cnbook.net](http://www.cnbook.net)

本书电子教案、习题参考答案可在该网站下载, 此外, 该网站还有其他相关书籍的介绍, 可以方便读者选购参考。

编者

2007 年 7 月

# 目 录

<b>第 1 章 绪论</b> .....	1	2.3.2 继承	36
1.1 数据库系统概述	1	2.3.3 约束	37
1.1.1 数据库系统的发展史	1	2.3.4 聚集	38
1.1.2 数据库系统的应用领域	3	2.4 E-R 模式转换为表	39
1.1.3 数据库系统的特点	4	2.4.1 实体的转换规则	39
1.2 数据库系统的组成	6	2.4.2 联系的转换规则	40
1.2.1 数据库	6	2.4.3 一般化的转换规则	40
1.2.2 数据库管理系统	6	2.4.4 聚集的转换规则	41
1.2.3 数据库用户及管理员	8	小结	41
1.2.4 其他	9	综合练习二	41
1.3 数据库系统体系结构	10	一、选择题	41
1.3.1 三级体系结构	11	二、填空题	42
1.3.2 应用程序体系结构	14	三、思考题	42
1.4 数据模型	15	四、上机题	43
1.4.1 实体-联系模型	16	<b>第 3 章 关系数据模型</b> .....	<b>44</b>
1.4.2 关系模型	17	3.1 关系数据模型概述	44
1.4.3 其他逻辑模型	18	3.1.1 关系数据结构	44
小结	21	3.1.2 关系操作	44
综合练习一	21	3.1.3 关系的完整性约束	45
一、选择题	21	3.2 关系数据模型的数据结构	45
二、填空题	22	3.2.1 关系	46
三、思考题	23	3.2.2 关系模式	49
四、上机题	23	3.2.3 关系数据库模式	50
<b>第 2 章 实体-联系模型</b> .....	<b>24</b>	3.2.4 E-R 模式转换为关系数据库模式	51
2.1 E-R 模型的基本概念	24	3.2.5 模式图	51
2.1.1 实体集 (Entity Set)	24	3.3 关系代数	52
2.1.2 联系集 (Relationship Set)	26	3.3.1 数据库的查询	52
2.1.3 码 (key)	28	3.3.2 数据库的修改	64
2.1.4 弱实体集 (Weak Entity Set)	29	3.3.3 视图	66
2.2 实体-联系图 (E-R 图)	30	3.4 关系演算	69
2.2.1 实体集的 E-R 图表示	31	3.4.1 元组关系演算	69
2.2.2 联系集的 E-R 图表示	31	3.4.2 域关系演算	76
2.2.3 E-R 图实例	33	小结	82
2.3 E-R 模型的扩展特性	35	综合练习三	83
2.3.1 特殊化和一般化	35	一、选择题	83

二、填空题.....	84	综合练习五.....	134
三、思考题.....	84	一、选择题.....	134
四、上机题.....	84	二、填空题.....	135
<b>第4章 SQL.....</b>	<b>86</b>	三、思考题.....	135
4.1 SQL 概述.....	86	四、上机题.....	136
4.1.1 SQL 语言概述.....	86	<b>第6章 关系数据理论.....</b>	<b>137</b>
4.1.2 SQL 语言的特点及功能.....	86	6.1 为什么需要范式.....	137
4.1.3 本章使用的示例.....	87	6.2 范式 (Normal Form).....	138
4.2 数据操纵语言 (DML).....	87	6.2.1 函数依赖 (Functional	
4.2.1 简单查询.....	88	Dependency).....	138
4.2.2 复杂查询.....	95	6.2.2 范式 (Functional	
4.2.3 修改.....	99	Dependency).....	139
4.3 数据定义语言 (DDL).....	101	6.3 Armstrong 公理系统.....	146
4.3.1 表的创建、删除与更改.....	101	6.4 模式分解.....	152
4.3.2 View (视图).....	104	6.4.1 模式分解的三个定义.....	152
4.3.3 UDF (用户自定义函数).....	106	6.4.2 模式分解的无损连接性.....	152
4.3.4 SP (存储过程).....	107	6.4.3 模式分解的保持函数依赖性.....	156
4.3.5 trigger (触发器).....	108	6.4.4 模式分解算法.....	157
4.3.6 动态 SQL.....	108	小结.....	159
4.3.7 cursor (游标).....	110	综合练习六.....	160
4.4 数据控制语言 (DCL).....	111	一、选择题.....	160
4.4.1 授予权限.....	111	二、填空题.....	160
4.4.2 收回权限.....	112	三、思考题.....	160
小结.....	113	四、上机题.....	161
综合练习四.....	113	<b>第7章 数据库设计.....</b>	<b>162</b>
一、选择题.....	113	7.1 数据库设计概述.....	162
二、填空题.....	114	7.1.1 数据库设计的内容.....	162
三、思考题.....	114	7.1.2 数据库设计的过程.....	162
四、上机题.....	115	7.2 需求分析.....	163
<b>第5章 查询处理和查询优化.....</b>	<b>116</b>	7.2.1 需求分析的主要工作.....	163
5.1 查询处理.....	116	7.2.2 数据字典.....	164
5.1.1 概述.....	116	7.3 概念设计.....	164
5.1.2 查询处理的代价.....	118	7.3.1 概念设计的重要性.....	164
5.1.3 单个关系运算符的计算.....	118	7.3.2 概念设计的方法.....	165
5.1.4 查询表达式的计算.....	123	7.3.3 概念设计的步骤.....	165
5.2 查询优化.....	126	7.4 逻辑设计.....	166
5.2.1 概述.....	126	7.5 物理设计.....	167
5.2.2 基于关系代数的优化.....	127	7.6 验证设计.....	169
小结.....	134	7.6.1 数据库实施.....	169

7.6.2 数据库的运行和维护.....	170	9.2.2 稳定存储器的实现.....	214
7.7 数据库设计实例.....	171	9.2.3 数据存储和数据访问.....	216
7.7.1 考勤管理系统的需求分析.....	171	9.3 故障种类.....	217
7.7.2 考勤管理系统的概念设计.....	172	9.3.1 事务故障.....	217
7.7.3 考勤管理系统的逻辑设计和 物理设计.....	173	9.3.2 系统故障.....	217
7.7.4 考勤管理系统的验证设计.....	174	9.3.3 介质故障.....	218
小结.....	184	9.4 数据恢复技术.....	218
综合练习七.....	185	9.4.1 基于日志的数据恢复技术.....	218
一、选择题.....	185	9.4.2 基于数据转储的数据恢复技术.....	218
二、填空题.....	186	9.5 数据恢复策略.....	224
三、思考题.....	186	9.5.1 事务故障恢复.....	224
四、上机题.....	186	9.5.2 系统故障恢复.....	225
<b>第8章 数据库保护.....</b>	<b>187</b>	9.5.3 介质故障恢复.....	225
8.1 数据库安全性概述.....	187	9.6 远程数据备份.....	226
8.1.1 信息安全评估标准.....	187	小结.....	227
8.1.2 数据库安全性控制技术.....	191	综合练习九.....	227
8.2 SQL的用户授权.....	194	一、选择题.....	227
8.2.1 建立权限.....	194	二、填空题.....	228
8.2.2 权限授予与取消.....	195	三、思考题.....	228
8.2.3 角色对授权的简化.....	196	四、上机题.....	228
8.3 数据库的完整性.....	197	<b>第10章 并发控制.....</b>	<b>229</b>
8.3.1 实体完整性和键码.....	197	10.1 并发控制概述.....	229
8.3.2 参照完整性和外键码.....	199	10.1.1 并发的由来.....	229
8.3.3 用户定义的完整性.....	201	10.1.2 并发执行的问题.....	229
小结.....	204	10.2 并发调度的可串行化.....	231
综合练习八.....	205	10.2.1 可串行化的概念.....	231
一、选择题.....	205	10.2.2 冲突可串行化.....	231
二、填空题.....	205	10.2.3 视图可串行化.....	233
三、思考题.....	205	10.2.4 可串行化的判定.....	234
四、上机题.....	206	10.2.5 冲突可串行化(充要条件).....	234
<b>第9章 数据恢复.....</b>	<b>207</b>	10.3 封锁协议.....	235
9.1 事务概述.....	207	10.3.1 封锁.....	235
9.1.1 事务由来.....	207	10.3.2 一级封锁协议.....	236
9.1.2 事务概念.....	208	10.3.3 二级封锁协议.....	237
9.1.3 事务特性.....	208	10.3.4 三级封锁协议.....	237
9.1.4 事务状态.....	211	10.3.5 两段锁协议.....	238
9.2 存储器概述.....	213	10.3.6 树形协议.....	240
9.2.1 存储器类型.....	213	10.4 时间戳协议.....	241
		10.4.1 时间戳.....	241



10.4.2 时间戳排序协议.....	242	12.2 XML 数据模型.....	276
10.4.3 Thomas 写规则.....	243	12.2.1 XML DTD.....	276
10.5 有效性检查协议.....	244	12.2.2 XML Schema.....	277
10.6 封锁管理.....	245	12.3 XML 数据库查询语言 XPath.....	280
10.6.1 活锁.....	245	12.3.1 XPath 简介.....	280
10.6.2 死锁.....	246	12.3.2 数据模型.....	281
10.6.3 死锁预防.....	247	12.3.3 定位路径与定位步.....	282
10.6.4 死锁检测.....	248	12.3.4 基本表达式与函数调用.....	285
10.6.5 死锁恢复.....	248	12.4 XQuery 查询语言简介.....	291
10.7 多粒度.....	249	12.4.1 XQuery 简介.....	291
10.7.1 多粒度封锁.....	249	12.4.2 XQuery 语法与查询实例.....	291
10.7.2 意向锁.....	250	小结.....	293
小结.....	251	综合练习十二.....	293
综合练习十.....	251	一、选择题.....	293
一、选择题.....	251	二、填空题.....	294
二、填空题.....	252	三、思考题.....	294
三、思考题.....	252	四、上机题.....	295
四、上机题.....	252	<b>第 13 章 PowerBuilder 与数据库应用开发 ..</b>	<b>296</b>
<b>第 11 章 对象数据库系统 .....</b>	<b>253</b>	13.1 PowerBuilder 介绍.....	296
11.1 面向对象数据库系统.....	253	13.1.1 PowerBuilder 是什么.....	296
11.1.1 面向对象方法及基本概念.....	253	13.1.2 PowerBuilder 术语.....	299
11.1.2 面向对象的数据模型.....	256	13.1.3 PowerScript 语言.....	300
11.1.3 对象定义语言 ODL.....	258	13.2 使用 PowerBuilder 开发数据库应用	
11.1.4 对象查询语言 OQL.....	260	程序.....	302
11.2 对象-关系数据库系统.....	264	13.2.1 创建 PowerBuilder 应用程序的	
11.2.1 对象-关系数据库语言 SQL3.....	264	步骤.....	302
11.2.2 函数和过程.....	270	13.2.2 案例介绍: 图书馆管理系统.....	304
小结.....	271	13.2.3 数据库设计.....	304
综合练习十一.....	272	13.2.4 GUI 用户界面设计.....	306
一、选择题.....	272	13.2.5 业务逻辑实现.....	310
二、填空题.....	272	小结.....	316
三、思考题.....	272	综合练习十三.....	317
四、上机题.....	273	一、选择题.....	317
<b>第 12 章 XML 数据库 .....</b>	<b>274</b>	二、填空题.....	317
12.1 XML 与数据库技术的发展.....	274	三、思考题.....	317
12.1.1 XML 简介.....	274	四、上机题.....	318
12.1.2 XML 数据库技术.....	275	<b>参考文献.....</b>	<b>319</b>

# 第1章 绪 论

当今，信息资源已成为各个企业或组织的重要财富，而建立一个能满足各级部门信息需求且行之有效的信息系统成为其生存和发展的重要条件。数据库就是一个相互关联的数据的集合，包含了关于某个企业或组织的信息，它是信息系统的核心和基础。数据库管理系统是由一组用以组织、存储、访问和维护数据的程序组成，它是一种数据管理的软件。数据库管理系统提供一个可以方便、有效地存取数据库信息的环境。建立数据库系统的目的是为了管理大量信息资源。

数据库技术是计算机科学的重要分支并在各行各业得到了广泛的应用，它已成为存储、处理信息资源的主要手段。因此，数据库课程已成为计算机科学教育中一个不可少的部分。

本章首先概括讲述数据库系统，然后介绍数据库系统的组成和数据库系统的体系结构，最后介绍数据模型。通过总体介绍，使读者掌握数据库系统的基本概念和组成架构，并对数据库技术有一个全面的概括性的了解。本章是后面各章节的基础。

## 1.1 数据库系统概述

数据库系统是指引入数据库技术后的计算机系统。狭义地讲，数据库系统由数据库和数据库管理系统组成；广义地讲，数据库系统由数据库、数据库管理系统（及其开发工具）、应用系统、数据库管理员和用户构成。

在具体介绍上述概念之前，先来看看数据库系统的发展历程、应用领域以及数据库系统的特点。

### 1.1.1 数据库系统的发展史

从第一台计算机起，数据的处理就是计算机发展的源动力。数据处理是指对各类型数据进行收集、存储、加工和传播等一系列活动的总和。数据处理的核心问题——数据管理是指对数据进行分类、组织、编码、存储、检索和维护。数据库技术是应数据管理任务的需要而产生和发展的。

数据存储和处理技术的发展历程如下。

#### 1. 20世纪50年代

当时的硬件条件是：没有磁盘等直接存取的存储设备，数据存储磁带、卡片、纸带上；软件条件是：没有操作系统，没有管理数据的软件，数据需要由应用程序自己设计、说明（定义）和管理。由于数据是面向应用的，当多个应用程序涉及某些相同的数据时，必然造成大量的冗余数据，且当数据的逻辑结构或物理结构发生变化时，必然造成应用程序的相应修改，这给程序员带来了巨大的负担。

数据处理包括从一个或多个磁带上顺序读取数据，并将数据写回到新的磁带上。例如，工资增长的处理是通过读取表示增长的一叠穿孔卡片和保存主要工资细节的主磁带实现的，工资的增加额将被加入到从主磁带读出的工资中，并被写到新的磁带上，它将成

为新的主磁带。

### 2. 20 世纪 50 年代后期到 60 年代中期

当时的硬件方面已有了磁盘、磁鼓等直接存取存储设备；软件方面已有了操作系统，且操作系统中有了专门的数据管理软件（即文件系统）。由文件系统对数据进行管理，它将数据组织成相互独立的数据文件，采用“按文件名称访问，按记录进行存取”的管理技术，从而实现对数据的更新、插入和删除操作。此外，文件系统还提供了程序和数据间的存取方法，使得程序与数据之间具有了一定的独立性。

在文件系统中，文件仍然是面向应用的，相同数据不能共享，数据的独立性差。当多个应用程序涉及某些相同的数据时，也必须建立各自的文件，数据的冗余度较大。同时，由于相同数据的重复存储、独立管理，极易造成数据的不一致性，这给数据的维护和修改带来了麻烦。此外，数据的逻辑结构改变必然导致应用程序的修改以及文件结构定义的变化，相应地，应用程序的改变也将引起文件数据结构的变化，程序与数据之间的独立性较差。

### 3. 20 世纪 60 年代后期到 70 年代

此时的硬件方面已有了大容量磁盘，硬件价格下降，磁盘得到了广泛的应用，而软件的开发及维护成本则上升。计算机的应用范围越来越广泛，待处理的数据量也随之急剧增长，对数据共享的要求越来越迫切。在这种背景下，文件系统已经不能满足实际应用对数据管理的要求，数据库技术应运而生，出现了统一管理数据的专门软件系统——数据库管理系统。数据库系统来管理数据解决了多用户、多应用共享数据的问题，提高了数据的独立性，降低了数据的冗余度。

最早出现的数据库系统是网状数据库和层次数据库，它们具有如表和树这样的数据结构，数据长期保存在磁盘上。程序员可以通过数据库管理系统创建和维护这些数据。1970 年由 Codd 发表的一篇具有里程碑意义的论文，定义了关系模型及在关系模型中非过程化的数据查询方法，至此关系数据库诞生了，数据库技术进入了一个新时代。以关系模型为基础的 System R 的出现，极大地推动了关系数据库的发展。完全功能的 System 原型导致了 IBM 的第一个关系数据库产品 SQL/DS 问世。最初的商用关系数据库系统，例如 IBM 的 DB2、Oracle、Ingres 和 DEC 的 Rdb，在推动有效地处理陈述式查询技术上起到了重要的作用。

### 4. 20 世纪 80 年代

关系数据库已经可以在性能上同网状/层次数据库进行竞争，并逐步取得优势，以至于最后完全取代了网状/层次数据库，而关系模型在数据模型中也取得了最高的统治地位。这是因为程序员在使用网状/层次数据库时不仅需要处理许多底层的实现细节，还要将查询任务编码成过程化的形式。更重要的是，它们在设计应用程序时必须时时考虑效率问题，这一切给程序员带来巨大的负担。相反，在关系数据库中，几乎所有的底层工作都由数据库自动完成，程序员只需考虑逻辑层的工作即可。关系数据库的简单易用是它占据市场的关键。

在 20 世纪 80 年代，人们还对并行和分布式数据库进行了大量研究，而在面向对象数据库方面也有初步的工作。

### 5. 20 世纪 90 年代以来

数据库的应用重心由事务处理、数据更新，转移到决策支持、数据查询上。决策支持

和查询再度成为数据库的一个主要应用领域。分析和处理大量数据的工具有了很大的发展。SQL 语言主要是为决策支持应用设计的，它是一个通用的、功能极强的关系数据库语言，已成为了国际标准语言。

20 世纪 90 年初，许多数据库厂商推出了并行和分布式数据库产品。不仅如此，数据库厂商还开始向数据库产品中加入对象 - 关系的支持。

20 世纪 90 年代末，互联网爆炸式发展，数据库比以前有了更加广泛的应用。现在的数据库系统不仅必须具有较高的事务处理速度、较高的可靠性和可用性，还必须支持网络接口，具有联机实时处理的能力。

### 1.1.2 数据库系统的应用领域

数据库的应用领域非常广泛，不管是企业、银行、医院、学校，还是政府部门，都需要使用数据库来存储数据信息。以下是一些数据库的典型传统应用：

(1) 销售业：用于存储供应商、商品、客户信息以及商品的销售信息。

(2) 航空业：用于存储航班和票务信息。航空业是最先以地理上分布的方式使用数据库的行业之一，分布于世界各地的终端通过通信网络或其他数字网络来访问中央数据库系统。

(3) 金融业：用于存储股票、债券等金融票据的持有、出售和买入等交易信息。

(4) 银行业：用于存储客户的信息、账户、贷款以及银行的交易记录。

(5) 电信业：用于存储通信网络的信息，存储通话记录，存储用户付费业务记录及产生每月通讯账单等。

(6) 制造业：用于存储产品的订单、产品原料的供应情况，跟踪产品的产量以及仓库产品的详细清单。

(7) 教育系统：用于存储教职员的信息，存储工资、津贴和纳税的信息，产生工资单；存储学生信息、课程注册和成绩信息等。

正如以上所列举的那样，数据库现已成为各个企业或组织的重要组成部分，同时数据库也直接或间接地给人们带来了前所未有的便利。例如，通过代理预定机票、通过银行付电话账单、查询考试成绩等等。尽管应用用户界面隐藏了用户访问数据库的细节，甚至大多数人没有意识到它们正在和一个数据库打交道，然而今天访问数据库已经成为几乎每一个人生活中的基本组成部分。

随着信息时代的进一步发展，数据库也相应产生了一些新的应用领域，主要表现在下面几个方面：

(1) 多媒体数据库：这类数据库主要存储与多媒体相关的数据，如声音、图像和视频等数据。多媒体数据最大的特点是数据量比较大、数据类型较多且数据类型间差距较大，这样多媒体数据库需要较大的存储空间，需要较复杂的数据管理。

(2) 空间数据库：这类数据库应用于地理信息系统（即 GIS）和计算机辅助设计与制造（CAD/CAM）。那些直接或间接与地球上的空间位置相关的空间数据是地理信息系统的重要组成部分；而设计数据（如机械、集成电路以及电子设备设计图等）则是计算机辅助设计与制造的核心。

(3) 移动数据库：这类数据库是在移动计算机系统上发展起来的，如笔记本电脑、

掌上电脑等。该数据库最大的特点是通过无线数字通信网络进行传输，用户可以随时随地地访问和获取数据，为一些商务应用和紧急情况带来了更大的便利。

(4) 信息检索系统：信息检索就是根据用户输入的提示信息，从数据库中查找相关的文档，并把查找的结果反馈给用户。信息检索领域和数据库技术是同步发展的，它是一种典型的联机文档管理系统。

(5) 决策支持系统：决策系统主要是指联机分析处理（即 OLAP）。OLAP 是数据仓库系统的主要应用，它支持复杂的分析操作，侧重决策支持，并且提供直观易懂的查询结果。OLAP 可使分析人员、管理人员或执行人员能够从多角度对信息进行快速、一致、交互地存取，从而获得对数据更深入的了解。

由此可见，随着科技的发展，将有越来越多新的应用领域采用数据库来存储和处理信息资源。

### 1.1.3 数据库系统的特点

数据库系统是指在计算机系统中引入数据库之后的系统。狭义地讲，数据库系统由数据库和数据库管理系统组成；广义地讲，数据库系统一般由数据库、数据库管理系统（及其开发工具）、应用系统、数据库管理员和数据库用户组成。数据库系统的组成如图 1-1 所示。本章的第 2 节将对数据库系统的组成作详细介绍。

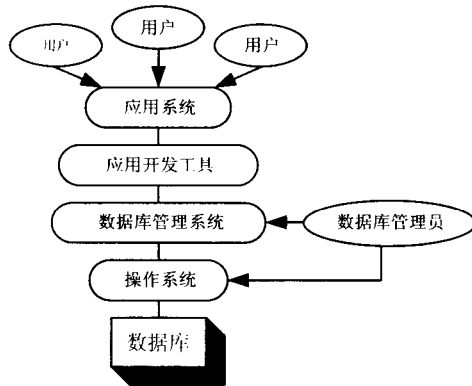


图 1-1 数据库系统的组成

在不引起混淆的情况下，通常把数据库系统简称为数据库。

与操作系统中专门的数据管理软件——文件系统相比，数据库系统具有以下主要特点：

#### (1) 数据结构化。

数据库系统实现整体数据的结构化是数据库系统与文件系统的本质区别，也是数据库系统的主要特征之一。

在文件系统中，尽管每个文件内部都是有结构的，即一个文件是由若干记录组成，每个记录是由若干属性组成，但是记录之间没有联系。并且文件是面向具体应用的，因此它只考虑某个具体应用的数据结构即可。

数据库系统实现了整体数据的结构化，即不仅要考虑某个具体应用的数据结构，还要考虑整个组织的数据结构。例如，在一个企业的信息管理系统中，不仅要考虑人力资源部门的职员管理，还要考虑生产部门的人事调配，同时要考虑财务部门的薪酬发放等等。因

此,企业的信息管理系统中的职员数据就要面向不同的应用,有效的数据组织方式为各个部门的应用提供了必要的数据库,此时就要求在描述数据时不仅要描述数据本身,还要描述数据之间的联系。

此外,在文件系统中数据的最小存取单位是记录,而在数据库系统中,不仅可以存取数据库中的一组记录或一个记录,还可以存取记录中的一组数据项或某一个数据项。数据库系统具有灵活的数据存取方式。

### (2) 数据具有较高的共享性。

在文件系统中,不同的应用程序各有自己独有的文件,而且文件之间基本上是不能共享的,从而导致了数据的冗余度较高,相同数据的重复存储、独立管理又造成了数据的不一致性。

数据库系统从全局的角度组织和描述数据,数据不再面向某个具体应用,而是面向整个系统,因此数据为多个用户、多个应用所共享。数据共享极大地减少了数据的冗余,从而避免了数据的不一致性。

此外,数据库中的数据是面向数据库系统整体的,这样不仅可以被共享使用,而且更容易向数据库系统中添加新的应用。数据库系统的扩展性好,可以更好地满足用户的需求。

### (3) 数据具有较高的独立性。

在文件系统中,尽管提供了程序和数据之间的存取转换方式,但是数据逻辑结构的变化必然导致应用程序以及文件结构定义的修改。

数据库系统提供了使应用程序与数据的逻辑结构和物理存储相独立的数据组织方式,数据具有较高的独立性。也就是说,即使数据的逻辑结构或者物理存储发生了改变,应用程序亦不需改变。数据与应用程序之间的独立,将数据的定义与数据的存储从程序中分离出来,从而简化了应用程序的负担,大大减少了应用程序的修改和维护。

数据库系统的数据独立性包括数据的逻辑独立性和数据的物理独立性。从数据库管理系统角度来讲,数据库系统通常抽象为三级模式,在这三级模式之间提供两层映像功能,正是这两层映像功能保证了数据库系统的独立性。将在第4节详细讨论。

### (4) 数据由数据库管理系统统一管理和控制。

数据库中的数据是数据库系统用户的共享资源,它由数据库管理系统统一管理和控制。数据库管理系统保证了科学地组织和存储数据、高效地存取和维护数据,它的主要功能包括:数据定义功能、数据操纵功能、数据的组织、存储和管理功能、数据库的运行管理、数据库的建立和维护功能以及数据转换功能等等。

数据库管理系统实现了应用程序与数据的独立,将程序开发人员从复杂的数据组织和维护中解放出来,极大地减轻了程序开发人员的负担。

数据库系统中应用程序、数据库管理系统以及数据库的关系如图1-2所示。

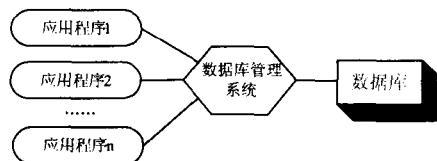


图1-2 应用程序、数据库管理系统以及数据库的关系

综上所述,数据库是长期存储在计算机内有组织的、大量的共享数据集。数据库中的数据是整体结构化的,具有较高的共享性、独立性、易扩充性及较低的冗余度。数据库中的数据由数据库管理系统在数据库的建立、运行和维护时对其进行统一的管理和控制,以保证数据的科学组织和高效存储,并提供数据共享的并发控制、故障恢复等功能。

数据库系统的优势使其成为现代信息系统中不可分离的重要组成部分,数据库技术也成为计算机领域中发展最快的技术之一。下面讨论数据库系统的组成。

## 1.2 数据库系统的组成

上一节介绍了数据库系统一般由数据库、数据库管理系统(及其开发工具)、应用系统、数据库管理员和数据库用户组成。下面分别介绍数据库系统的主要组成部分。

### 1.2.1 数据库

数据库中存储的基本对象是数据。狭义的理解,数据就是数字,例如 100、73.5、-20、 $\pi$  等等;广义的理解,数据可以是文本(text)、图形(graph)、图像(image)、音频(audio)、视频(video)等等。确切地讲,数据是描述事物的符号记录。随着科技的进步,描述事物的符号已由狭义的理解——数字扩展至文字、图形、声音、图像等多种形式。人们将应用中所需要的大量数据收集、抽取出来,并将其存储在计算机中形成数据库,以供进一步加工处理,进一步提取有用信息。数据库中的数据是由数据库管理系统统一进行管理和控制的。

数据表示在数据库中实际存储的内容,而关于这些数据的解释称为数据的语义。例如,85 是一个数据,它可以解释为某个商品的价格,也可以解释为仓库中某种商品的库存量,还可以解释为某个同学的某门课程的成绩。数据与其语义是不可分割的,它们共同描述了一个事物。

确切地讲,数据库(DataBase)是长期储存在计算机内、有组织、可共享的数据集合。所谓数据库中的数据是长期储存在计算机内的,这是因为一旦数据进入数据库被数据库管理系统所接受,就只有向数据库管理系统提出某些明确的操作请求并得到允许时,才能对数据库中的数据进行插入、更新或删除操作。数据库中的数据区别于某些程序的输入数据、输出数据、工作队列、中间结果等临时数据。所谓数据库中的数据是有组织的和可共享的,这是指数据库是按照某一数据模型对数据进行组织、描述和储存的,它不仅包含数据本身,还包含数据之间的联系。这使得数据库中的数据具有较小的冗余度和较高的易扩展性,并可为各种用户所共享。

最后,由于数据库系统通常抽象为三级模式,它所提供的映像功能保证了数据库具有较高的数据独立性。

### 1.2.2 数据库管理系统

数据库管理系统(DataBase Management System, DBMS)是位于用户与操作系统之间的一层数据管理软件。

下面详细地介绍一下 DBMS 的功能,主要包括以下几个方面。

#### 1. 数据定义功能

DBMS 必须提供数据定义语言(Data Definition Language, DLL),用户通过它可以方

便快捷地定义数据库中的各类对象。同时，DBMS 必须包含支持数据定义语言的 DDL 解释器。DDL 解释器解释 DDL 语句并将数据定义存储在数据字典中。DBMS 通过它可以 将数据定义的源形式转化成相应的目标形式，并调用操作系统，存储在数据字典当中。

DBMS 包括数据字典。数据字典本身也可以看作为一个数据库（系统数据库而不是用户数据库），它存储的是数据的描述（元数据，即系统中其他实体的定义）。例如，数据库模式的定义、数据的各种安全性和完整性约束的定义等等。当然，查询数据字典与查询其他数据库是相同的。

## 2. 数据操纵功能

DBMS 还需提供数据操纵语言（Data Manipulation Language, DML），用户通过它可以方便快捷地操作数据库中的各类数据，实现对数据的查询、插入、删除或更新等操作。同时，DBMS 必须包含支持数据操纵语言的 DML 编译器。DML 编译器将 DML 语句翻译为一个求值方案，包括一系列求值引擎能理解的低级指令。DBMS 通过它们可以将数据操作的源形式转化成相应的目标形式，并调用操作系统，实现对数据库中数据的物理操作。

通常情况下，DML 操作请求可以分为“计划”的和“非计划”的两种：

“计划”的请求是指可预见的需求请求。计划的请求通常来自预先写好的应用程序，数据库管理员可以据此调整数据库的物理设计以保证请求具有较好的执行性能。

“非计划”的请求是指不可预知的需求请求，是一种特殊查询。非计划的请求通常来自“决策支持”类应用，它是通过查询语言处理器交互提交的，因此数据库的物理设计不一定能够适当处理这种请求。

“计划”的或“非计划”的数据操纵请求是在解释器/编译器、优化器部件处理之后，在查询求值引擎的控制下执行。在实际操作过程中，查询求值引擎是调用文件管理等来实现对数据的访问与存储。优化器部件是用来决定有效执行请求的方式，优化问题将在查询优化与查询处理中进行详细讨论。

## 3. 数据库的建立、运行和维护功能

数据库在建立、运用和维护时是由数据库管理系统进行统一的管理和控制的。

数据库的建立和维护功能包括数据库初始数据的输入功能、数据库的备份功能、数据库的性能监测功能、数据库的重组功能等等。

此外，为实现数据库的运行管理，DBMS 必须保证数据的安全性、完整性、数据的并发控制及故障后的数据库恢复：

（1）数据的安全性（Security）和完整性（Integrity）。

数据的安全性是指保护数据库中的数据，以防止不合法的使用造成的数据泄密、更改或破坏，即保护数据库防止恶意的破坏和非法的存取。数据的安全性检查使得非法用户不能访问数据库，每个合法用户也只能对指定数据按照指定权限进行访问和存取。

数据的完整性指数据的正确性和相容性，即防止数据库中不存在不符合语义的数据。数据的完整性检查使得不合语义的、不正确的数据不能进入数据库。

为保证数据的安全性和完整性，DBMS 要监控所有用户的操作请求，拒绝一切破坏数据库安全性和完整性的请求，只有通过安全性和完整性检查的操作请求才被允许执行。数据的安全性和完整性检查由权限及完整性管理器完成。

（2）并发（Concurrency）控制。



数据库中的数据是高度共享的，当多个用户的并发进程同时访问数据库时，可能会发生相互干扰而得到错误的结果，或者造成数据的不一致，因此 DBMS 必须具有对多用户的并发操作进行控制的功能。DBMS 通过并发控制管理器控制并发进程的相互影响，保证数据的一致性。事实上，并发控制机制以及数据库恢复机制是事务处理技术。事务（Transaction）是用户定义的一个数据库操作序列，它具有原子性、一致性、隔离性和持久性。并发控制机制保证了事务执行的一致性和隔离性。

### （3）数据库恢复（Recovery）。

尽管数据库系统采取了各种保护措施来防止数据库的数据遭到破坏，但是计算机系统硬件故障、软件错误、操作员的失误以及恶意破坏仍然会影响到数据库中数据的正确性，更严重地，甚至会造成数据库中数据的部分或全部丢失。因此，DBMS 必须具有将数据库从错误状态恢复到故障发生前的某一已知正确（一致）状态的功能，即数据库恢复功能。DBMS 通过事务管理器实现数据库恢复功能，它保证即使发生了故障，数据库也能保持在正确的（一致的）状态下。数据库恢复机制保证了事务执行的原子性和持久性。

### 4. 数据的组织、存储和管理

数据库中的数据主要有数据文件、数据字典以及索引三种结构。数据文件存储数据库自身的数据（即用户数据）；数据字典存储关于数据库结构的元数据；索引提供对包含特定值的数据项的快速访问。数据的组织、存储和管理的目标就是确定文件的结构，提高存储空间的利用率，提供多种存取方式以提高存取效率等。DBMS 通过文件管理器和缓冲区管理器来实现上述功能。

此外，DBMS 还具有数据库的转储、与其他软件系统的通信、与其他文件系统的数据库转换等功能。数据库管理系统是数据库系统的重要组成部分。

## 1.2.3 数据库用户及管理

数据库系统的使用人员可以分为数据库用户和数据库管理员。数据库用户主要查询和修改数据库中的数据，而数据库管理员负责全面管理和控制数据库系统。

根据数据库用户与数据库系统交互方式的不同，将其分为以下三类。

### 1. 初级用户

初级用户没有专业的数据库领域知识，它们只能通过应用系统的用户界面或者调用先前写好的应用程序与数据库系统进行交互。常用的应用系统用户界面有菜单驱动、图形显示、报表输入等等。

例如，一个生产部门经理要查询本月的生产计划，他会访问一个用来输入查询要求的窗口，该窗体上面的应用程序就会用输入的查询请求提取数据库中的生产计划数据，并将这个信息反馈给用户。或者，此用户可以调用一个称为 plan 的应用程序，该应用程序要求用户输入月份，并将生产计划作为输出。

初级用户只需按照用户界面的提示填写相应项，或者按照应用程序说明填写相应的程序参数就可以了。此类用户可以很简单地获得和阅读数据库的数据，而不需掌握数据库的专业知识。

### 2. 高级用户

高级用户具有专业的数据库领域知识，他们不仅可以通过调用应用程序或用户界面与