



数据库应用系列教材



# 数据库系统 教程

沈钧毅 侯迪 冯中慧 何亮 编著

 科学出版社  
[www.sciencep.com](http://www.sciencep.com)

数据库应用系列教材

# 数据库系统教程

沈钧毅 侯 迪 冯中慧 何 亮 编著

科学出版社

北 京

## 内 容 简 介

本书融数据库基础、数据库设计、数据库管理系统和数据库新技术为一体,以关系数据库为重点,全面系统地介绍数据库系统原理与技术的基本内容。本书在系统阐述数据库基本概念和原理的基础上,强调理论与应用相结合,力图反映当前数据库技术发展的水平和趋势。全书共分4部分14章。第1部分,分别介绍数据库系统的基本概念、关系数据库和数据库语言。第2部分,主要介绍关系数据库理论、数据库设计。第3部分,详细、深入、系统地介绍数据库管理系统内部机制。第4部分,分别简要介绍对象数据库系统、Internet环境下的数据库技术、数据仓库与数据挖掘。为方便读者学习,各章末均附有习题。

本书可作为高等学校计算机专业或其他专业的数据库基础课程教材,也可供广大科技人员自学、参考。

---

### 图书在版编目(CIP)数据

数据库系统教程/沈钧毅等编著. —北京:科学出版社,2006

(数据库应用系列教材)

ISBN 7-03-018079-8

I. 数… II. 沈… III. 数据库系统-教材 IV. TP 311.13

中国版本图书馆CIP数据核字(2006)第111887号

---

责任编辑:鞠丽娜/责任校对:赵燕  
责任印制:吕春珉/封面设计:三函设计

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

信浩彩色印装有限公司印刷

科学出版社发行 各地新华书店经销

\*

2006年11月第一版 开本:B5(720×1000)

2006年11月第一次印刷 印张:24 1/2

印数:1—4 000 字数:491 400

定价:32.00元

(如有印装质量问题,我社负责调换〈环伟〉)

销售部电话 010-62136131 编辑部电话 010-62138978-8002

# 前 言

“数据管理”是计算机应用领域中最广泛的分支之一，数据库系统是继文件系统进行数据管理以后的一种崭新技术，在信息化社会建设中有着重要的作用。当前，数据库系统课程在计算机及其相关专业的教学计划中占有重要的地位，本书是一部面向本科生教学的数据库系统教程。

全书共分4部分，内容涵盖了数据库系统基础、应用（设计）、实现及新技术4个部分。期望读者通过这4个部分的学习，全面掌握数据库系统基本概念及原理，以便为在各个领域信息化建设中熟练应用数据库技术奠定必要的基础。

第1部分为基础篇，共分3章阐述数据库系统的基础。第1章数据库系统概述，从数据库技术的产生和发展、数据库系统、数据库管理系统3个方面概括地阐述了数据库系统的基本概念。第2章关系模型，阐述了目前流行使用的数据模型——关系模型，从关系代数、关系演算两种数学概念出发，介绍了关系数据库的理论基础。第3章关系数据库语言SQL，从应用角度介绍了关系数据库语言，分别阐述了数据定义语言、操纵语言、查询语言、控制语言主要语句的功能和使用。

第2部分为设计篇，从如何应用数据库技术建立计算机信息系统出发，共分2章阐述数据库设计的原理与方法。第4章数据依赖与关系模式规范化，概要阐述关系数据库设计的理论依据。第5章数据库设计，全面阐述数据库设计的过程与方法。

第3部分为系统篇，共分6章阐述数据库系统的核心——数据库管理系统的内部实现机制。第6章存储结构，介绍了外存储器及相关的数据物理组织。第7章查询处理与优化，介绍了数据库查询操作的处理过程，并阐述了提高查询性能的优化策略。第8章事务管理，从事务这一重要概念出发，阐述故障恢复与并发控制功能的实现机制。第9章数据库完整性和安全性，介绍完整性、安全性的基本概念及其实现机制。第10章分布式数据库系统，简要阐述分布式数据库系统的组成、体系结构与实现机制。第11章数据库访问技术，分别阐述C/S、B/S两种体系结构以及ODBC访问技术。

第4部分新技术篇，共分3章从数据模型、Internet环境下的数据库技术以及数据库技术用于决策分析3个方面，简要地介绍了数据库新技术。第12章面向对象数据库系统/对象关系数据库系统，从数据模型角度阐述了关系数据库系统之后的新技术。第13章Internet环境下的数据库技术，着重阐述了WWW技术、WWW与数据库访问接口以及XML技术。第14章数据仓库与数据挖掘，概要地阐述了数据库技术应用于决策分析的新技术。

本书可作为48~56学时的数据库相关课程的教材使用，书中用\*号标出的部

分和章节为选讲内容，可根据具体的课程设置及课时数情况酌情增减。本书易于自学，既可作为相关课程的教材，也可供从事数据库领域相关工作的科技人员与工程技术人员参阅。

本书是在作者多年来从事数据库教学经验的基础上撰写而成，具体写作分工如下：第1章及第14章由沈钧毅执笔，第2章至第5章由何亮执笔，第7章至第9章由冯中慧执笔，第6章及第10章至第13章由侯迪执笔。在本书编写过程中，得到了科学出版社在济南主办教材审稿会议参与者富有建设性的建议，特别感谢南京大学徐洁磐教授、北京大学唐世渭教授、中国科学院研究生院邵佩英教授、山东大学董继润教授、山东财政学院聂培尧教授以及其他参会老师。

全书的编写得到了科学出版社鞠丽娜编辑的热情支持与具体帮助。南京大学徐洁磐教授精心审稿，提出了很多修改意见。在此，向以上老师表示衷心地感谢。

由于编者经验水平有限，缺点与错误在所难免，殷切期望广大读者批评指正。对本书的各种意见与建议，请发邮件至 [dbs\\_course@mail.xjtu.edu.cn](mailto:dbs_course@mail.xjtu.edu.cn)。

作 者

2006年3月于西安交通大学

# 目 录

## 第 1 部分 基础篇

<b>第 1 章 数据库系统概述</b> .....	1
1.1 数据库技术的产生和发展.....	1
1.1.1 数据管理的发展.....	1
1.1.2 数据库系统的特征.....	4
1.1.3 数据库技术的应用与发展.....	5
1.2 数据库系统.....	5
1.2.1 数据库系统的组成结构.....	5
1.2.2 数据库、数据模型与数据模式.....	6
1.3 数据库管理系统.....	9
1.3.1 DBMS 概述.....	9
1.3.2 DBMS 结构.....	10
1.3.3 DBMS 进程结构.....	12
1.3.4 数据目录.....	15
习题.....	16
<b>第 2 章 关系模型</b> .....	17
2.1 基本概念.....	17
2.1.1 概述.....	17
2.1.2 关系的定义.....	18
2.1.3 关系的完整性.....	21
2.2 关系代数.....	22
2.2.1 传统的集合运算.....	23
2.2.2 专门的关系运算.....	24
2.2.3 关系代数查询实例.....	28
2.3 关系演算.....	29
2.3.1 元组关系演算.....	29
2.3.2 域关系演算.....	32
习题.....	33
<b>第 3 章 关系数据库语言 SQL</b> .....	35
3.1 SQL 语言概述.....	35
3.2 SQL 语言的数据定义功能.....	37
3.2.1 基本表的建立.....	37
3.2.2 基本表的修改与删除.....	40

3.2.3	索引的建立和删除	41
3.2.4	模式的建立和删除	42
3.3	SQL 语言数据操纵功能	42
3.3.1	SQL 数据查询	43
3.3.2	SQL 数据插入	53
3.3.3	SQL 数据删除	55
3.3.4	SQL 数据修改	55
3.4	视图	56
3.4.1	视图的定义和删除	56
3.4.2	视图的查询	58
3.4.3	视图的更新	59
3.4.4	视图的应用	60
3.5	SQL 语言的数据控制功能	61
3.5.1	权限与角色	61
3.5.2	权限的授予和收回	62
3.6	嵌入式 SQL	63
3.6.1	嵌入式 SQL 概述	63
3.6.2	嵌入式 SQL 数据库访问过程	64
3.6.3	不使用游标的嵌入式 SQL 语句	65
3.6.4	使用游标的嵌入式 SQL 语句	67
3.6.5	动态 SQL 语句	69
习题		70

## 第 2 部分 设计篇

第 4 章	数据依赖与关系模式规范化	72
4.1	问题的提出	72
4.2	数据依赖	74
4.2.1	函数依赖	75
4.2.2	多值依赖	80
4.3	关系模式分解	81
4.3.1	无损连接分解	81
4.3.2	保持函数依赖分解	84
4.4	关系模式规范化	84
4.4.1	范式	85
4.4.2	规范化算法	87
习题		89

<b>第 5 章 数据库设计</b> .....	92
5.1 数据库设计概述 .....	92
5.1.1 数据库系统生命周期 .....	92
5.1.2 数据库设计方法和步骤 .....	93
5.2 需求分析 .....	95
5.3 概念设计 .....	96
5.3.1 E-R 模型简介 .....	97
5.3.2 E-R 设计方法 .....	102
5.4 逻辑设计 .....	111
5.4.1 E-R 模型转换 .....	111
5.4.2 优化与调整 .....	114
5.5 物理设计 .....	116
5.5.1 存储结构概述 .....	116
5.5.2 存储结构与存取方式设计 .....	118
5.6 IDEF 设计方法简介 .....	121
5.7 计算机辅助数据库设计 .....	124
5.8 数据库运行与维护 .....	125
习题 .....	128

### 第 3 部分 系 统 篇

<b>*第 6 章 存储结构</b> .....	129
6.1 物理存储设备 .....	129
6.1.1 磁盘存储器 .....	129
6.1.2 磁带存储器 .....	131
6.1.3 磁盘缓冲处理 .....	131
6.1.4 第三级存储 .....	132
6.2 文件和文件记录 .....	132
6.3 无序文件 .....	133
6.4 顺序文件 .....	134
6.5 索引文件 .....	135
6.5.1 索引的基本结构 .....	135
6.5.2 主索引 .....	135
6.5.3 聚簇索引 .....	137
6.5.4 辅助索引 .....	138
6.5.5 多级索引 .....	139
6.5.6 索引的评价标准 .....	140



6.6	B 树与 B <sup>+</sup> 树索引 .....	141
6.6.1	索引树 .....	141
6.6.2	B 树索引 .....	142
6.6.3	B <sup>+</sup> 树索引 .....	144
6.7	散列文件 .....	147
6.7.1	简单散列方法 .....	148
6.7.2	散列函数 .....	148
6.7.3	散列文件的操作 .....	149
6.7.4	溢出处理 .....	150
6.7.5	动态散列方法 .....	150
	习题 .....	151
<b>第 7 章</b>	<b>查询处理与查询优化</b> .....	<b>153</b>
7.1	引言 .....	153
7.1.1	查询处理 .....	153
7.1.2	查询优化分类 .....	155
7.1.3	一个启发性的例子 .....	156
7.2	代数优化 .....	157
7.2.1	代数优化的基本原则 .....	158
7.2.2	代数优化的等价变换规则 .....	158
7.2.3	代数优化策略 .....	160
7.2.4	代数优化算法 .....	161
7.3	物理优化 .....	165
7.3.1	选择 .....	165
7.3.2	连接 .....	167
7.3.3	投影 .....	171
7.3.4	集合运算 .....	172
	习题 .....	175
<b>第 8 章</b>	<b>事务管理</b> .....	<b>176</b>
8.1	事务 .....	176
8.1.1	事务的概念 .....	176
8.1.2	事务的性质 .....	178
8.1.3	事务的操作 .....	179
8.1.4	事务的状态 .....	179
8.1.5	SQL 的事务管理 .....	180
8.2	数据库故障分类 .....	182
8.3	数据库恢复技术 .....	183
8.3.1	数据转储 .....	183

8.3.2	日志文件	184
8.3.3	恢复策略	186
8.4	事务的并发控制	188
8.4.1	并发引起的问题	189
8.4.2	事务调度及可串行性	191
8.5	基于锁的并发控制	196
8.5.1	封锁协议	196
8.5.2	两阶段封锁协议	199
8.5.3	活锁与死锁	200
8.5.4	封锁粒度	203
8.5.5	插入和删除操作	206
*8.6	其他并发控制技术	207
8.6.1	基于时间标记的并发控制	207
8.6.2	乐观并发控制	208
	习题	210
<b>第9章</b>	<b>数据库完整性与安全性</b>	<b>212</b>
9.1	引言	212
9.2	数据库的完整性	212
9.2.1	数据库完整性约束分类	213
9.2.2	完整性约束条件	213
9.2.3	DBMS 的完整性控制	215
9.2.4	完整性约束的说明	218
9.3	数据库的安全性	221
9.3.1	访问控制	222
9.3.2	视图机制	227
9.3.3	数据加密	228
9.3.4	数据库审计	229
	习题	230
<b>第10章</b>	<b>分布式数据库系统</b>	<b>231</b>
10.1	概述	231
10.1.1	分布式数据库的由来与发展	231
10.1.2	分布式数据库系统的特点	231
10.1.3	分布式数据库系统的组成	232
10.1.4	分布式数据库系统中存在的问题	233
10.2	分布式数据库系统的模式结构	235
10.3	分布式查询处理和优化	236
10.3.1	分布式查询优化的目标和准则	236

10.3.2	分布式查询的分类	237
10.3.3	分布式查询处理的层次结构	238
10.4	分布式事务	240
10.5	联邦数据库和多数据库系统	241
10.5.1	联邦数据库系统的基本功能	241
10.5.2	联邦数据库系统的数据共享	241
10.5.3	联邦数据库系统存在的问题	243
10.5.4	多数据库系统	243
习题		246
<b>*第 11 章</b>	<b>数据库访问技术</b>	247
11.1	数据库访问方式	247
11.1.1	数据库应用程序的组成部分	247
11.1.2	主机终端系统	248
11.1.3	基于网络的 C/S 系统结构	249
11.1.4	基于 Web 的 B/S 结构	253
11.2	中间件	253
11.2.1	中间件的定义	253
11.2.2	中间件的作用	254
11.3	SQL/CLI	255
11.3.1	SQL/CLI 的由来和发展	255
11.3.2	SQL/CLI 基本原理和方法	255
11.3.3	SQL/CLI 应用编程	257
11.4	ODBC	263
11.4.1	ODBC 的由来和发展	263
11.4.2	ODBC 的体系结构	264
11.4.3	ODBC 的特性	269
习题		269

## 第 4 部分 新技术篇

<b>第 12 章</b>	<b>面向对象数据库与对象关系数据库</b>	271
12.1	引言	271
12.2	面向对象的基本概念	273
12.2.1	面向对象方法的基本特征	273
12.2.2	对象	273
12.2.3	对象结构	274
12.2.4	消息	276

---

12.2.5	类	277
12.2.6	继承	277
12.2.7	动态联编	279
12.2.8	对象组合	279
12.3	面向对象的数据模型	280
12.3.1	基本特征	280
12.3.2	基本模型	281
12.4	持久化程序设计语言	284
12.4.1	对象的持久性	285
12.4.2	对象标识和指针	286
12.4.3	持久对象的存储和访问	286
12.5	对象-关系数据模型	287
12.5.1	关系数据模型扩充	289
12.5.2	嵌套关系	289
12.5.3	组合类型	291
12.5.4	聚集类型	292
12.5.5	引用类型	294
12.5.6	对象类型	295
12.6	ORDB 定义语言	296
12.6.1	数据类型定义	296
12.6.2	继承的定义	299
12.7	ORDB 查询语言	301
12.7.1	以关系为值的属性	301
12.7.2	路径表达式	301
12.7.3	嵌套与解除嵌套	302
12.7.4	函数的定义和使用	303
12.7.5	复合值的创建和查询	304
	习题	305
*第 13 章	Internet 环境下的数据库技术	306
13.1	网络基本概念	306
13.1.1	网络协议模型	306
13.1.2	Internet 与 WWW	308
13.2	WWW 与数据库访问接口	311
13.2.1	公共网关接口 CGI	312
13.2.2	ASP 技术	313
13.2.3	JSP 技术	316

13.2.4	PHP 技术	317
13.3	Java 和 JDBC	319
13.3.1	Java 简介	319
13.3.2	JDBC 的提出	320
13.3.3	JDBC 的基本功能	320
13.3.4	JDBC 数据库应用模型	321
13.3.5	JDBC 驱动程序	322
13.3.6	JDBC API	323
13.3.7	JDBC 应用程序的流程	325
13.4	XML 技术	327
13.4.1	XML 概述	327
13.4.2	XML 查询语言——XQuery	330
13.4.3	基于关系数据库的 XML 处理	335
	习题	341
<b>*第 14 章</b>	<b>数据仓库与数据挖掘</b>	<b>343</b>
14.1	概述	343
14.1.1	数据管理的层次结构	343
14.1.2	数据仓库的产生	344
14.1.3	从数据仓库到数据挖掘	344
14.2	数据仓库	345
14.2.1	概述	345
14.2.2	数据仓库的建立——数据模型、数据模式	348
14.2.3	OLAP 技术	351
14.3	数据挖掘	355
14.3.1	概述	355
14.3.2	数据挖掘的过程	357
14.3.3	数据挖掘的基本方法	360
14.3.4	复杂数据类型的挖掘	370
	习题	376
	主要参考文献	377

# 第 1 部分 基础篇

## 第 1 章 数据库系统概述

数据库系统的目标是管理大量的、持久的、共享的数据，并支持用户查找和更新所需要的信息，是计算机科学的重要分支。本章主要讲述数据库技术的产生背景、发展历史以及未来的发展方向；数据库系统的组成结构、基本概念和技术；数据库管理系统的基本功能、进程结构和工作原理。

### 1.1 数据库技术的产生和发展

#### 1.1.1 数据管理的发展

众所周知，电子计算机从 1946 年诞生以来，已成为现代社会各种应用领域中的重要工具。以前，人们经常在讲到计算机应用领域时，都会列举哪些领域能使用电子计算机，可能会列举几千种应用方面，将其分类可以概括为：科学计算、过程控制、数据管理等几大领域。现在，人们往往会这样提出问题：哪些部门没有使用电子计算机。这一问题表明：现代社会的方方面面几乎无处不用电子计算机。几十年的应用表明：在这些应用领域中，电子计算机用于数据管理几乎占到了绝大多数。

电子计算机如何进行数据管理，在历经将近 50 年的过程中，大体可划分为以下 3 个阶段。

##### 1. 人工管理阶段

在 20 世纪 60 年代计算机操作系统问世以前，人们使用计算机进行数据管理，必须用人工方式组织数据，把用户的数据直接存放在外存设备的有关地址里，然后编写应用程序对这些数据进行管理。在这个阶段，应用程序与数据之间的关系如图 1.1 所示。

显然，这样一种数据管理方式无法解决数据管理的如下需求：① 数据量很大。如果管理几百、几千个数据，还可以对付，如果数据量是几十 KB、MB，甚至 GB，将是无法解决的；② 保存数据。每次使用都得将数据输入到外存，无法持久保存，下次使用必须重新保存；③ 数据共享。数据只能对应于某一个应用程序，其他应用程序使用这些数据时必须自己重新定义，造成多个应用程序使用同一数

据时的冗余现象。

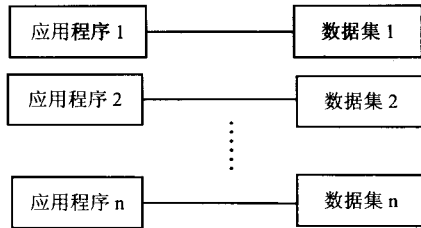


图 1.1 人工管理阶段应用程序与数据之间的相互关系

正因为这样，操作系统中的文件系统，作为管理外存的一种手段，应运而生，为数据的存放和应用带来了若干方便。

### 2. 文件系统阶段

20 世纪 60 年代，出现了操作系统，其中的文件系统作为管理外存的系统软件，解决了外存上数据的按“名”存取，为数据管理发展奠定了基础。

文件系统存取数据，用户不必自己指定外存地址来存放，而且只需给定数据集的“名字”，由文件系统加以统一管理。例如，对学生数据进行管理，只要建立学生文件即可。至于这些数据放在磁盘的什么位置，以后如何取出数据，都不需要用户来管理，而由文件系统来管理。这个阶段的应用程序与数据之间的对应关系，可用图 1.2 来表示。

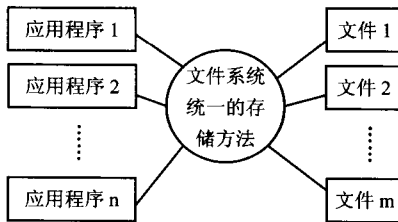


图 1.2 文件系统阶段应用程序与数据之间的对应关系

注意，图 1.2 中应用程序中的  $n$  与文件  $m$  不一定相同。

可以看出，这一阶段的数据管理，比人工管理阶段的好处是：① 解决了数据量大的问题。用户的数据通过文件来存放，其数据量大小与用户自己分配空间的难度完全不同。按“名”存取文件，对数据量大小没有什么影响；② 解决了数据的持久保存问题。文件系统解决了按“名”持久保存数据的难题，用户一旦建立起了所需的文件，文件系统将统一管理这些数据；③ 部分解决了数据共享问题。用户所建立的文件，可由不同的应用程序加以调用，当然用户必须了解其文件的结构。图 1.2 表明：程序与文件之间由文件系统的统一存取方法加以联系，部分

地解决了相互依赖问题，但没有完全解决数据共享问题。

对于文件系统而言，程序与数据有了一定的物理独立性，数据在物理存储上的改变没有反映在程序上。但是，文件为特定的应用程序服务，文件的逻辑结构改变会直接影响应用程序，数据与程序之间缺乏逻辑独立性。而且，前面已说明，文件系统在数据共享上存在问题，现举一个例子，如图 1.3 所示。

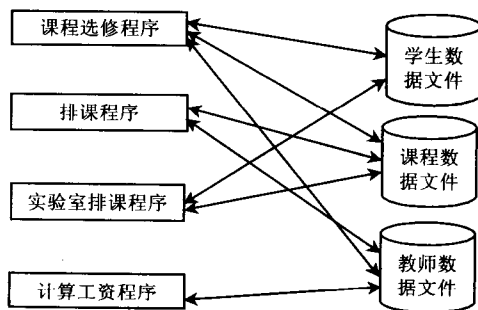


图 1.3 应用程序与文件之间的对应例子

针对文件系统在数据管理上存在的缺陷，在数据统一管理中以共享数据为主要特征的数据库系统应运而生。

### 3. 数据库系统阶段

20 世纪 60 年代末期，计算机用于数据管理，规模急剧增长，数据共享要求越来越强烈。在文件系统基础上，出现了统一管理数据的专门性系统软件——数据库管理系统 (database management system, DBMS)。从数据模型角度看，可以认为：数据库系统是在文件系统基础上增加了文件之间的联系而出现的。(本章 1.2 节将阐述数据模型概念)。图 1.4 表明了 DBMS 是文件系统发展而得来的。

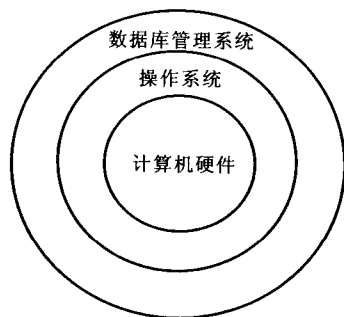


图 1.4 DBMS 是文件系统的发展

数据库系统阶段，应用程序与数据之间的对应关系，可用图 1.5 来表示。



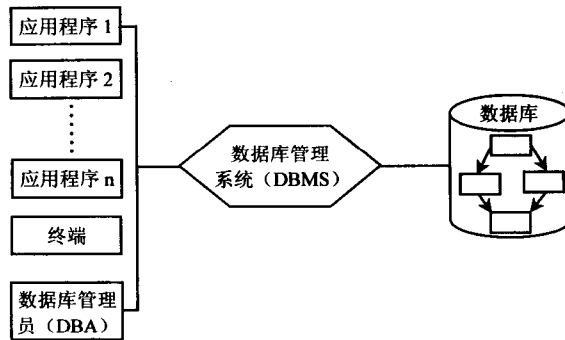


图 1.5 数据库系统阶段应用程序与数据之间的对应关系

### 1.1.2 数据库系统的特征

数据库系统是现阶段计算机数据管理的主要形式，是数据管理的最新技术。与数据管理的前期阶段即人工管理与文件系统相比较，它具有以下几个方面的特征。

#### 1. 良好的数据模型

回顾数据库系统阶段以前的数据管理手段，在数据模型上存在严重的缺陷：人工管理阶段，数据都是由用户自行组织的，没有组织结构；文件系统阶段，对于文件内部有了一定的组织结构（如：顺序文件与索引文件），记录之间都有一定的联系，但是不同文件之间一般是相互独立的，使用时由应用程序来完成不同文件记录之间的调用。

数据库系统阶段，数据库模型是数据库存储数据的基础，从开始出现的层次、网状数据库模型，到目前流行的关系数据库模型，直到新一代数据模型的提出，无不反映出构成数据库模型基础的文件之间的联系，数据模型所能描述数据的特性更为良好。

#### 2. 数据共享性好

数据库系统中，数据模型是从所建数据库设计的企事业单位整体角度来描述数据的，所建的数据将被多个用户与应用共享。从数据存储角度看，数据共享大大减少了有关数据分别存放时的重复现象，节省了存储空间，用计算机专业术语说是减少了冗余。减少冗余，带来了重要的特性：避免数据之间的不一致。例如，同一数据如果分别存放，有可能出现不同的值，造成同一数据的不一致，这是数据管理中必须避免的问题。因为数据库的不一致性，使这些数据成为垃圾数据，对这样的数据加以管理显然是没有意义的。数据管理中流行的话，垃圾进垃圾出 (garbage in, garbage out)，就是指这种现象。