

普通高等教育“十三五”规划教材

数据库技术与实战

大数据浅析与新媒体应用

潘瑞芳 徐芝琦 张宝军 编著



中国工信出版集团



电子工业出版社
Publishing House of Electronics Industry
<http://www.phei.com.cn>

普通高等教育“十三五”规划教材

数据库技术与实战

——大数据浅析与新媒体应用

潘瑞芳 徐芝琦 张宝军 编著



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书分为两部分,上篇为基础原理篇,全面讲述了数据库系统概念、关系数据库系统的相关知识、关系数据库标准语言 SQL、关系数据库的规范化理论、数据库设计过程、大数据与分布式数据库基本知识;下篇为技术应用篇,介绍了3个多媒体应用案例,分别是 SQLite 在 Android APP 开发中的应用、MySQL 在 Unity 网络游戏开发中的应用和 SQL Server 在图书管理系统开发中的应用。本书内容丰富,知识新颖,应用案例多样化,既包含关系数据库的基础理论,又介绍了大数据与分布式数据库的基本概念;既有典型的应用案例,又为学生毕业设计提供了应用开发的项目实例。

本书贴合当下数据库主流方向,可作为计算机专业本、专科数据库原理及应用的教材,也可作为数据库开发及应用人员的参考书籍。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

数据库技术与实战:大数据浅析与新媒体应用 / 潘瑞芳, 徐芝琦, 张宝军编著. — 北京: 电子工业出版社, 2018.3

ISBN 978-7-121-33609-6

I. ①数… II. ①潘… ②徐… ③张… III. ①数据库系统—高等学校—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2018)第 018931 号

策划编辑:戴晨辰

责任编辑:戴晨辰 文字编辑:刘 瑾

印 刷:天津嘉恒印务有限公司

装 订:天津嘉恒印务有限公司

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编:100036

开 本:787×1092 1/16 印张:18.25 字数:456 千字

版 次:2018 年 3 月第 1 版

印 次:2018 年 3 月第 1 次印刷

定 价:46.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: dcc@phei.com.cn。

前 言

数据库技术是计算机科学与技术学科下发展最快、应用最广的一个分支，它从产生发展到今天不过短短几十年，但其应用却已渗透到生活的各个方面。近年来，云计算、大数据的快速发展，进一步推动了数据库技术的变革，新一代数据库技术应运而生。

本书分为两部分，上篇为基础原理篇，包括1~6章；下篇为技术应用篇，包括7~9章，各章配有习题。本书还有配套的实验指导教材，本书中的各知识点在实验教材中均有体现。

本书上篇的第1章主要介绍了数据库系统的基本概念；第2章主要介绍了关系数据库系统的相关知识；第3章主要介绍了关系数据库标准语言SQL；第4章主要介绍了关系数据库规范化理论；第5章主要讨论了数据库设计的全过程；第6章概述了大数据与分布式数据库（NoSQL）的基本概念；下篇的第7章主要介绍了SQLite在Android APP开发中的应用，第8章主要介绍了MySQL在Unity网络游戏开发中的应用；第9章主要介绍了SQL Server在图书管理系统开发中的应用。

本书内容丰富，知识新颖，应用案例多样化，既包含关系数据库的基础理论，又介绍了大数据与分布式数据库的基本概念；既有典型的应用案例，又为学生毕业设计提供了应用开发的项目实例。本书贴合当下数据库主流方向，可作为计算机专业本、专科数据库原理及应用的教材，也可作为数据库开发及应用人员的参考书籍。

本书提供配套电子课件、习题参考答案、案例源代码，任课教师可在华信教育资源网（<http://www.hxedu.com.cn>）注册后免费下载。

本书由浙江传媒学院新媒体学院潘瑞芳、徐芝琦和张宝军编著，其中第1~6章和第9章由潘瑞芳和张宝军编写，第7~8章由徐芝琦编写。

由于时间仓促，水平有限，本书难免存在缺点和错误，敬请广大读者批评指正。

编 者

目 录

上篇 基础原理篇

第 1 章 数据库系统概论	2
1.1 数据库技术的产生与发展	2
1.1.1 数据管理技术的发展	2
1.1.2 数据库技术的主要研究领域	3
1.2 数据库系统的基本概念	3
1.3 数据模型	5
1.3.1 现实世界的抽象过程	5
1.3.2 概念模型	5
1.3.3 数据模型	8
1.4 数据库体系结构	10
1.5 小结	12
1.6 习题	12
第 2 章 关系数据库	14
2.1 关系模型	14
2.1.1 基本概念	14
2.1.2 关系数据库的特点	15
2.2 数据完整性	16
2.2.1 实体完整性约束	16
2.2.2 参照完整性约束	16
2.2.3 用户自定义完整性约束	17
2.3 关系代数	17
2.3.1 传统的集合运算	17
2.3.2 关系运算	19
2.4 查询优化	22
2.4.1 查询优化的概念和策略	22
2.4.2 关系代数等价变换规则	22
2.5 小结	24
2.6 习题	24
第 3 章 关系数据库标准语言 SQL	26
3.1 SQL 语言概述	26
3.1.1 SQL 语言的基本概念	26
3.1.2 SQL 语言的分类	27

3.1.3	SQL 支持的数据库模式	28
3.1.4	标准 SQL 语言与数据库产品中的 SQL 语言	28
3.2	SQL Server 数据库简介	28
3.2.1	SQL Server 简介	28
3.2.2	SQL Server 2014 的安装	30
3.2.3	SQL Server 2014 的使用	36
3.3	数据定义	36
3.3.1	模式的定义和删除	36
3.3.2	创建基本表	37
3.3.3	修改表结构	39
3.3.4	删除基本表	40
3.3.5	创建索引	40
3.3.6	删除索引	41
3.4	数据更新	41
3.4.1	插入数据	41
3.4.2	修改数据	43
3.4.3	删除数据	44
3.5	数据查询	45
3.5.1	SELECT 的语法格式	45
3.5.2	简单查询	46
3.5.3	选择查询	48
3.5.4	分组查询	51
3.5.5	查询结果排序	53
3.5.6	连接查询	54
3.5.7	嵌套查询	55
3.5.8	使用聚集函数查询	57
3.5.9	子查询与数据更新	61
3.5.10	集合运算	63
3.6	视图	65
3.6.1	视图的作用	65
3.6.2	视图的定义	65
3.6.3	视图的删除	66
3.6.4	使用视图操作表数据	66
3.7	SQL 的数据完整性约束	67
3.7.1	事务	68
3.7.2	完整性约束	70
3.8	触发器	75
3.8.1	触发器的作用	76
3.8.2	触发器的组成	76

3.8.3	触发器的操作	76
3.9	存储过程	77
3.9.1	存储过程的基本概念	77
3.9.2	存储过程的定义	77
3.9.3	存储过程的执行	78
3.9.4	存储过程的删除	78
3.10	嵌入式 SQL 语言	78
3.10.1	嵌入式 SQL 语言的基本概念	78
3.10.2	嵌入式 SQL 语言需要解决的问题	79
3.10.3	嵌入式 SQL 语言的语法格式	79
3.10.4	嵌入式 SQL 与宿主语言之间的信息传递	79
3.10.5	游标	81
3.11	小结	82
3.12	习题	82
第 4 章	关系数据库规范化理论	84
4.1	问题的提出	84
4.1.1	存在异常的关系模式	85
4.1.2	异常原因分析	86
4.1.3	异常问题的解决	87
4.2	函数依赖	88
4.2.1	函数依赖基本概念	88
4.2.2	码的函数依赖表述	89
4.3	关系模式的规范化	89
4.3.1	第一范式	90
4.3.2	第二范式	91
4.3.3	第三范式	92
4.3.4	BCNF 范式	93
4.3.5	多值依赖与第四范式	94
4.3.6	连接依赖与第五范式	97
4.3.7	关系模式的规范化步骤	98
4.4	数据依赖的公理系统	99
4.5	关系模式的分解	102
4.5.1	模式分解中存在的问题	102
4.5.2	无损连接	103
4.5.3	保持函数依赖	106
4.6	小结	107
4.7	习题	108
第 5 章	数据库设计与管理	109
5.1	数据库设计概述	109

5.1.1	数据库设计方法	109
5.1.2	数据库设计的一般步骤	110
5.2	需求分析	112
5.3	概念结构设计	114
5.3.1	概念结构设计概述	114
5.3.2	局部概念模型设计	116
5.3.3	全局概念模型设计	117
5.4	逻辑结构设计	119
5.4.1	E-R 模型到关系模型的转换	120
5.4.2	关系模型的优化	123
5.4.3	设计用户外模式	123
5.5	物理结构设计	123
5.5.1	选择存取方法	124
5.5.2	确定存储结构	125
5.5.3	物理结构设计的评价	126
5.6	数据库的管理	126
5.6.1	数据库的实施阶段	126
5.6.2	数据库的运行和维护	127
5.7	小结	128
5.8	习题	129
第 6 章	大数据与分布式数据库	130
6.1	大数据概述	130
6.1.1	大数据概念	130
6.1.2	大数据特征和技术特点	131
6.1.3	大数据发展	132
6.2	大数据应用	133
6.2.1	大数据应用的领域	133
6.2.2	大数据应用于行业	135
6.3	NoSQL 数据库	138
6.3.1	NoSQL 简介	138
6.3.2	NoSQL 数据库分类	140
6.3.3	NoSQL 与关系数据库的比较	141
6.4	小结	142
6.5	习题	142

下篇 技术应用篇

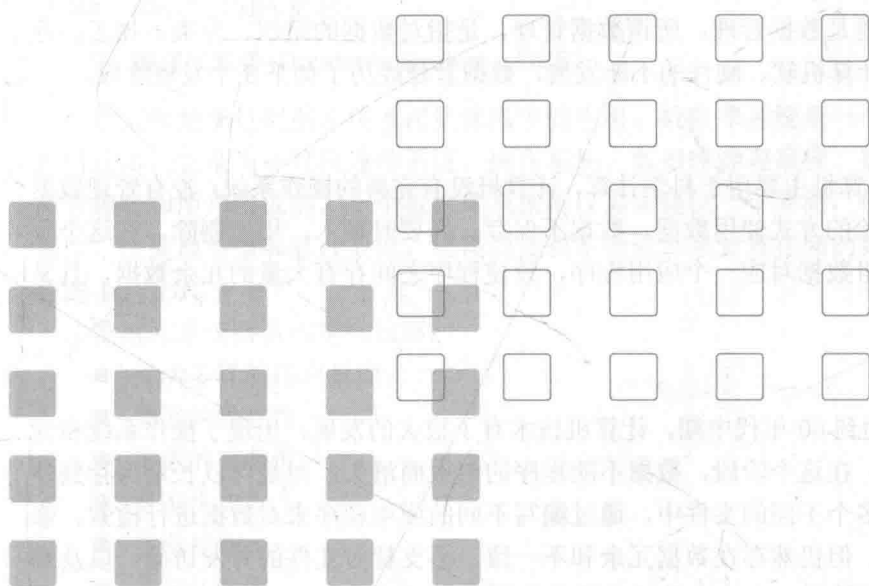
第 7 章	SQLite 在 Android APP 开发中的应用	144
7.1	SQLite 概述	144

7.1.1	SQLite 简介	144
7.1.2	SQLite 的特点	144
7.1.3	SQLite 的局限性	145
7.1.4	SQLite 基本语句	146
7.2	SQLite 的使用	146
7.2.1	SQLite 安装	146
7.2.2	SQLite 数据类型	149
7.2.3	SQLite 语法	151
7.2.4	SQLite 命令	155
7.3	Android SQLite 类和接口	157
7.3.1	SQLiteDataBase 类	158
7.3.2	SQLiteOpenHelper 类	161
7.4	搭建 Android SQLite 应用	162
7.4.1	创建新项目工程	162
7.4.2	定义 UI 界面	163
7.4.3	定义 schema	170
7.4.4	创建数据库相关内容	171
7.4.5	查看数据库文件	182
7.5	SQLite 应用的注意事项	183
7.6	小结	184
7.7	习题	184
第 8 章	MySQL 在 Unity 网络游戏开发中的应用	185
8.1	服务器的安装和配置	185
8.1.1	XAMPP 简介	185
8.1.2	XAMPP 的安装与运行	185
8.1.3	了解 Apache	190
8.1.4	了解 MySQL	191
8.1.5	了解 PHP	192
8.2	新建 Unity 项目	193
8.2.1	新建项目	193
8.2.2	创建 UI	194
8.3	创建数据库	199
8.3.1	定义数据库及相关表	199
8.3.2	插入测试数据	203
8.4	创建 PHP 脚本	205
8.4.1	login.php	206
8.4.2	insertUser.php	209
8.4.3	userData.php	212

8.5	Unity 中的 WWW 应用	215
8.5.1	UserBean.cs	216
8.5.2	LoginScripts.cs	218
8.5.3	EnrollScripts.cs	221
8.5.4	DataScripts.cs	224
8.6	小结	226
8.7	习题	226
第 9 章	SQL Server 在图书管理系统开发中的应用	227
9.1	图书管理系统案例介绍	227
9.2	技术说明	227
9.2.1	ASP.NET	227
9.2.2	ADO.NET	227
9.2.3	使用 ADO.NET 进行数据库应用开发	229
9.3	需求分析	229
9.4	系统设计	230
9.4.1	系统数据流程图	230
9.4.2	功能模块设计	230
9.4.3	数据库设计	231
9.5	系统实现	237
9.5.1	创建数据库和数据表	237
9.5.2	创建项目	239
9.5.3	公共类设计	240
9.5.4	登录模块设计	248
9.5.5	主界面设计	251
9.5.6	系统管理	257
9.5.7	图书管理	264
9.5.8	读者管理	267
9.5.9	借阅服务	270
9.5.10	查询服务	276
9.6	小结	281
9.7	习题	281
参考文献		282

上篇

基础原理篇



第1章 数据库系统概论

1.1 数据库技术的产生与发展

随着人类社会的不断发展和进步，人们需要处理的数据量越来越大，如何对这些数据进行存储、加工、传输和使用，已日益受到人们的广泛重视。数据库技术就是在这种形势下产生并发展的。

1.1.1 数据管理技术的发展

数据(Data)，即人们用符号对客观事物的描述。数据的种类有很多，包括：文字、图像、声音、图形等。数据是事实或观察的结果，是对客观事物的逻辑归纳，是用于表示客观事物的未经加工的原始素材。

数据是信息的表现形式和载体，可以是符号、文字、数字、语音、图像、视频等。数据和信息是不可分离的，数据是信息的表达，信息是数据的内涵。数据本身没有意义，数据只有对实体行为产生影响时才成为信息。

数据处理的核心问题是数据管理，所谓数据管理，是指对数据的组织、分类、加工、存储、检索和维护。随着计算机软、硬件的不断发展，数据管理经历了如下3个发展阶段。

1. 人工管理阶段

20世纪50年代，计算机主要用于科学计算，计算机没有完善的操作系统，没有管理数据的软件，用户以极其原始的方式使用数据，数据不保存，需要时输入，用完删除。在这个阶段，数据面向应用，一组数据对应一个应用程序，致使程序之间存有大量的冗余数据，且易产生数据不一致等问题。

2. 文件系统阶段

20世纪50年代后期到60年代中期，计算机技术有了很大的发展，出现了操作系统和管理数据的文件管理系统。在这个阶段，数据不随程序的结束而消失，而是可以长期保存到外存，所需的数据存储在多个不同的文件中，通过编写不同的应用程序来对数据进行检索、修改、插入和删除等操作。但仍然存在数据冗余和不一致，不支持对文件的并发访问，以及难以满足系统安全性要求等弊端，无法满足越来越高的数据处理要求。

3. 数据库系统阶段

20世纪60年代后期，计算机软、硬件技术的飞速发展带来了数据管理的革命，出现了数据管理的新方式——数据库系统(DBS)。数据库系统主要由数据库和数据库管理系统组成，

在数据库系统中,数据以数据库方式存储,使用数据库管理系统可以管理数据库的创建、修改和使用。

与前两种数据管理方式相比,数据库系统具有数据独立性强、冗余小、共享性高、完整性和安全性好等特点。

1.1.2 数据库技术的主要研究领域

数据库技术是使用计算机管理数据的一项新技术,从开始发展到现在,数据库技术已在各行各业得到了广泛应用,是计算机应用的一个重要领域。

数据库是相互关联的数据的集合,但数据库不是简单的数据归集。数据之间包含了一定的逻辑关系,数据库就是根据数据之间的联系和逻辑关系,将数据分门别类地存储。数据库中的数据应具有较小的冗余和较高的数据独立性,可为广大用户所共享。

数据库技术主要应用在处理密集型数据的领域,这些领域涉及的数据量大,数据需要长时间保存且为多个应用服务,数据库技术所研究的问题就是如何科学地组织和存储这些数据,以及如何高效地处理和使用这些数据。

1.2 数据库系统的基本概念

1. 数据(Data)

数据,是指用符号记录下来的可区别的信息。在数据库系统中,数据实际上就是可以被计算机存储、识别的信息。

2. 数据库系统(Database System, DBS)

数据库系统是数据库技术在计算机中的应用。数据库系统是一个有机结合的人机系统,严格地讲,它是由计算机硬件系统、操作系统、数据库管理系统、数据库、应用程序、数据库管理员和用户组成的。一个数据库系统不仅需要提供一个界面,使用户可以方便地建立数据库,快捷地检索和修改数据,还需要提供系统软件来管理存储的数据。数据库系统的组成如图 1-1 所示。

数据库系统必备的特性包括:

- 灵活多样的用户界面;
- 数据的独立性;
- 数据的完整性;
- 查询优化;
- 并发控制;
- 备份与恢复;
- 安全性。

3. 数据库(Database, DB)

在数据库中,数据与数据的逻辑结构同时存储,各数据文件中数据项的逻辑定义都记录

在“数据字典”中，通过数据库管理系统，用户可以方便地访问数据库中的数据，数据可高度共享。

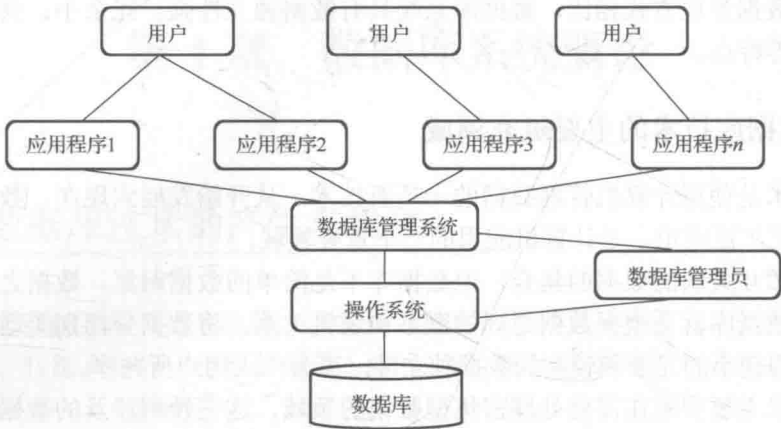


图 1-1 数据库系统组成

4

4. 数据库管理系统(Database Management System, DBMS)

数据库管理系统是数据库系统的核心，在操作系统的支持下，其对内负责管理数据库中的数据，对外负责对用户提供操作数据库的界面。数据库管理系统的主要功能如下。

(1) 数据定义

DBMS 提供了数据定义语言(Data Definition Language, DDL)，用于定义数据库中数据的逻辑结构。

(2) 数据操纵

DBMS 提供了数据操纵语言(Data Manipulation Language, DML)，用于对数据库进行检索、插入、修改和删除等基本操作。数据操纵语言一般分为两类：一类为自主型，一类为宿主型。自主型可独立使用，无须依赖其他程序设计语言；而宿主型则需嵌入到其他程序设计语言(如 C 语言等)中。

(3) 数据库运行控制

DBMS 提供了运行控制机制，包括：数据完整性控制、并发控制、安全性控制及数据备份和恢复功能。

5. 数据库管理员(DataBase Administrator, DBA)

数据库管理员不仅要熟悉数据库管理软件的使用，还应熟悉本行业的业务工作，其主要职责是：管理用户对数据库及相关软件的使用，对数据库进行维护，确保数据库正常运行。

6. 用户(User)

用户，即数据库的使用者，不同的用户可通过不同的方式访问数据库，既可通过良好的用户界面访问数据库，也可使用数据库语言直接访问，但必须是已经授权的用户，不同的用户被授予的访问权限也可能不同。

1.3 数据模型

1.3.1 现实世界的抽象过程

现实世界是指实际存在的事物或现象。各种事物都有着自己的许多特性，在众多的事物之间，它们又存在着千丝万缕的联系。

现实存在的事物，例如：桌子、人，桌子有高低、有方有圆、有黄有红等；人有男女、有胖有瘦、有白有黑等。这些都是事物自身拥有的特性，这些事物用计算机是无法直接处理的，只有将这些事物的特性数据化以后，才能被计算机所接受，才能被计算机处理。但是如何将现实世界的这些事物转换成计算机所能处理的数据，也就是如何将这些事物的特性及事物之间的联系转换成数据，就是本节要讨论的现实世界的抽象过程。

现实事物是不可能自动转换成计算机所能处理的数据的，它必须通过人的帮助才能转换。首先，应是对现实世界的事物有了发现，这种发现通过人们的头脑反应、理解后，转换成信息，然后将这些在人的头脑中反应的信息转换成计算机所能处理的数据。一般来说，我们把现实世界实际存在的东西称为事物，每件事物都有其基本特征，现实世界中的事物在人脑中的反应称为信息，这些信息被具体描述成一个个实体，这些实体就对应于现实世界的一件件事物，而事物的特征即被描述成实体的属性，再把信息在计算机中的物理表示称为数据，对应于实体和属性，在数据世界中称为记录和数据项。现实世界的抽象过程见图 1-2。

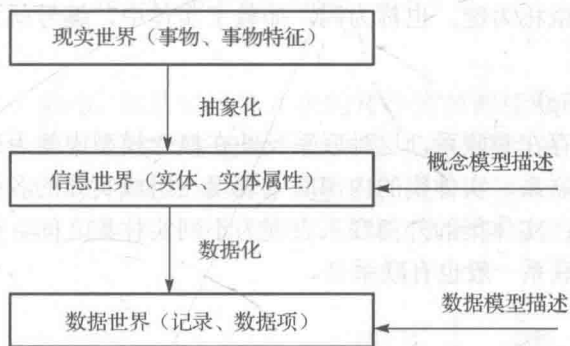


图 1-2 现实世界的抽象过程

1.3.2 概念模型

信息世界是现实世界转换到数据世界（又称机器世界）的中间环节，是人们对现实世界的认识和理解。信息世界用概念模型描述，概念模型不依赖于具体的机器世界，而是与现实世界紧密联系。要进行数据库设计，首先必须给出概念模型，概念模型能很好地体现出设计人员的思想，且设计简单，有利于设计人员与用户交流。

1. 基本概念

(1) 实体(Entity)

实体即客观存在且可区别的事物在信息世界的反映, 实体既可以是实际的事物, 又可以是一种概念或现象, 例如, 一个教师、一本书、一堂课、一个程序等都可称为实体。

(2) 实体集(Entity Set)

具有相同属性名, 而属性值可有所不同的实体的集合即为实体集。在实体集中, 不能存在两个或两个以上相同的实体, 如学校的全体教工、书店的全部书籍、工厂的所有设备等都是实体集。为了区别不同的实体集, 应给每个实体集取一个名字, 称为实体名。

(3) 实体型(Entity Type)

实体型即实体集的命名表示, 由实体名和实体集的各属性名构成。如教工登记表(编号, 姓名, 性别, 年龄, 婚否, 职称, 部门)就是全体教工实体集的实体型。

(4) 属性(Attribute)

属性即事物具有的具体特征, 在实体中称为属性, 实体是由若干个属性来描述的。如教工实体是由编号、姓名、性别、年龄、婚否、职称、部门这些属性来描述的。

(5) 域(Domain)

某个属性的取值范围称该属性的域。如性别的域为“男”和“女”, 姓名的域取 8 个字节长的字符串, 职称的域定义为“教授”、“副教授”、“讲师”、“助教”等。域限制属性的取值。

(6) 键(Key)

在实体集中, 不允许完全相同的两个实体存在, 即在同一个实体集中的实体, 相互间至少应有一个属性(或属性组)的值不同, 也就是应有一个能唯一区分一个实体的属性或属性组存在, 该属性或属性组就称为键, 也称为码。如教工实体中, 编号就可作为键, 因为每个编号只对应一个教工实体。

(7) 联系(Relationship)

现实世界中的事物存在着联系, 这种联系反映在概念模型中就表现为实体集本身内部的联系和实体集间外部的联系。实体集的内部联系表现在组成实体的各属性之间, 如姓名与职称之间是“拥有”关系; 实体集的外部联系表现为不同实体集之间, 如教师实体与学生实体之间是“教学”关系。联系一般也有联系名。

2. 实体集间的联系

两个实体集间的联系一般分为 3 类。

(1) 一对一联系(1:1)

假设有两个实体集 A 和 B , 如果实体集 A 中的每个实体至多与实体集 B 中的一个实体相联系, 而实体集 B 中的每个实体也至多与实体集 A 中的一个实体相联系, 则称实体集 A 与实体集 B 或实体集 B 与实体集 A 是一对一联系, 一般可记为 1:1, 如图 1-3 所示。

例如, 一个部门只有一个主任, 而一个主任也只有一个部门任职。则可认为主任与部门之间是一对一联系。

(2) 一对多联系(1:n)

假设有两个实体集 A 和 B , 如果实体集 A 中的每个实体都可以与实体集 B 中的多个实体

相联系，而实体集 B 中的每个实体却至多只能够与实体集 A 中的一个实体相联系，则称实体集 A 与实体集 B 是一对多联系，一般可记为 $1:n$ 。例如，一个班级可以有多名学生，而每个学生只能在一个班级中。则可认为班级与学生之间是一对多联系，如图 1-4 所示。

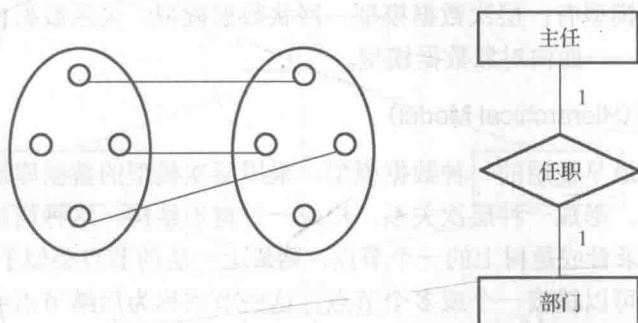


图 1-3 两个实体集间的一对一联系

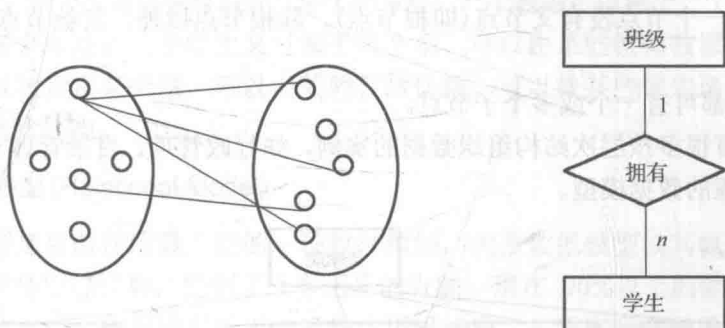


图 1-4 两个实体集间的一对多联系

(3) 多对多联系 ($m:n$)

假设有两个实体集 A 和 B ，如果实体集 A 中的每个实体都能够与实体集 B 中的多个实体相联系，而实体集 B 中的每个实体也能够与实体集 A 中的多个实体相联系，则称实体集 A 与实体集 B 或实体集 B 与实体集 A 是多对多联系，一般可记为 $m:n$ 。例如，一个教师可以教多门课程，一门课也可由多个教师任教。则可认为教师与课程之间是多对多联系。如图 1-5 所示。

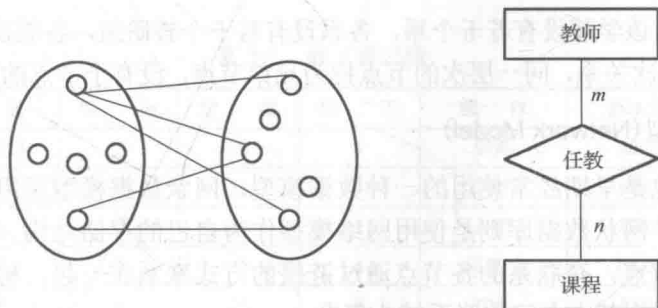


图 1-5 两个实体集间的多对多联系