



数据库原理 及应用

任永功 编著



科学出版社

数据库原理及应用

任永功 编著

科学出版社

北京

内 容 简 介

本书系统地阐述了数据库系统的基础理论、基本技术和基本方法,共 11 章,主要包括绪论、关系数据库、关系数据理论、关系数据库标准语言 SQL、数据库安全性、数据库完整性、关系查询处理和优化、并发控制、数据库恢复技术、数据库设计、数据库技术发展动态。

本书可以作为高等院校计算机、软件工程、信息安全、信息管理与信息系统、信息与计算科学等相关专业本科生数据库课程的教材,可以作为电气工程相关专业研究生数据库课程及电力企业信息化教材,也可供从事数据库系统研究、开发和应用的工程技术人员参考。

图书在版编目 (CIP) 数据

数据库原理及应用 / 任永功编著. —北京: 科学出版社, 2016
ISBN 978-7-03-046958-8

I. ①数… II. ①任… 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 (2016) 第 006686 号

责任编辑:石悦 王迎春 / 责任校对:胡小洁
责任印制:徐晓晨 / 封面设计:华路天然工作室

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>

北京京华虎彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 2 月第 一 版 开本: 787×1092 1/16

2016 年 2 月第一次印刷 印张: 14 1/2

字数: 344 000

定价: 59.00 元

(如有印装质量问题, 我社负责调换)

前 言

从 20 世纪 60 年代末至今,数据库技术已经发展了 50 年。在这 50 年的历程中,人们在数据库技术的理论研究和系统开发上都取得了辉煌的成就,而且已经开始对新一代数据库系统进行深入研究。数据库系统已经成为现代计算机系统的重要组成部分。数据库技术是通过研究数据库的结构、存储、设计、管理以及应用的基本理论和实现方法,并利用这些理论来实现对数据库中的数据进行处理、分析和理解的技术。

本书是作者在教学实践的基础上,结合“数据库原理及应用”省级精品资源共享课建设内容,根据一些院校数据库原理与应用课程学时短、实践性强的教学需要编写而成。本书共 11 章,主要包括绪论、关系数据库、关系数据理论、关系数据库标准语言 SQL、数据库安全性、数据库完整性、关系查询处理和优化、并发控制、数据库恢复技术、数据库设计、数据库技术发展动态。书中对每个知识点都进行图示概括,有大量例题,每章后都有练习题。本书语言通俗、结构合理、图文并茂,具有较强的实用性。

本书由任永功编写提纲并统稿,刘洋、赵月、索全明、王玉玲、孙华阳、高鹏、郭健、张虹、钱海振、郎泓钰等参与了本书的编写。本书在编著过程中,参考了大量的国内外资料,努力跟踪数据库技术前沿技术,必有不足之处,请同仁批评指正。

编 者

2015 年 12 月于大连

目 录

第 1 章 绪论	1
1.1 数据库系统综述	1
1.2 数据模型	6
1.3 数据库系统的组成	17
1.4 小结	18
本章知识结构图	19
习题	19
本章参考文献	19
第 2 章 关系数据库	21
2.1 关系概述	21
2.2 关系操作与关系数据语言	23
2.3 关系的完整性	24
2.4 关系代数和关系演算	26
2.5 小结	36
本章知识结构图	36
习题	37
本章参考文献	38
第 3 章 关系数据理论	39
3.1 问题的提出	39
3.2 函数依赖	41
3.3 范式	44
3.4 规范化小结	54
3.5 模式的分解	55
3.6 小结	67
本章知识结构图	68
习题	68
本章参考文献	72
第 4 章 关系数据库标准语言 SQL	73
4.1 SQL 概述	73
4.2 学生-课程数据库	75
4.3 数据定义	75

4.4 数据查询	80
4.5 数据更新	88
4.6 视图	90
4.7 小结	93
本章知识结构图	94
习题	95
本章参考文献	96
第5章 数据库安全性	97
5.1 安全性	97
5.2 安全性控制的一般方法	102
5.3 定义视图	111
5.4 审计	112
5.5 数据加密	113
5.6 统计数据库安全性	114
5.7 小结	115
本章知识结构图	115
习题	115
本章参考文献	116
第6章 数据库完整性	118
6.1 实体完整性	119
6.2 参照完整性	121
6.3 用户自定义完整性	122
6.4 域完整性约束	124
6.5 完整性设计的原则	125
6.6 触发器	125
6.7 小结	128
本章知识结构图	129
习题	129
本章参考文献	130
第7章 关系查询处理和优化	131
7.1 关系数据库系统的查询处理	131
7.2 关系数据库系统的查询优化	134
7.3 代数优化	137
7.4 物理优化	143
7.5 小结	145

本章知识结构图	146
习题	146
本章参考文献	146
第 8 章 并发控制	147
8.1 事务的基本概念	147
8.2 并发控制概述	149
8.3 封锁	151
8.4 活锁和死锁	151
8.5 解决数据库不一致性的三级锁协议	154
8.6 并发调度的可串行性	155
8.7 两段锁协议	156
8.8 封锁的粒度	158
8.9 小结	160
本章知识结构图	161
习题	161
本章参考文献	162
第 9 章 数据库恢复技术	163
9.1 数据库恢复概述	163
9.2 故障的种类	163
9.3 恢复的实现技术	165
9.4 恢复策略	168
9.5 具有检查点的恢复技术	170
9.6 数据库镜像	172
9.7 小结	173
本章知识结构图	173
习题	173
本章参考文献	174
第 10 章 数据库设计	175
10.1 数据库设计概述	175
10.2 需求分析	179
10.3 概念结构设计	189
10.4 逻辑结构设计	196
10.5 物理结构设计	201
10.6 数据库的实施和维护	204
10.7 小结	206

本章知识结构图.....	207
习题.....	207
本章参考文献.....	209
第 11 章 数据库技术发展动态	210
11.1 本章概述.....	210
11.2 数据库技术与多学科技术的有机结合.....	210
11.3 数据库建设中应注意的几个问题.....	215
11.4 数据库技术的全新特性及发展趋势.....	216
11.5 数据挖掘.....	217
11.6 小结.....	222
本章知识结构图.....	223
习题.....	223
本章参考文献.....	223

第1章 绪 论

本章要点

数据库技术的应用需要有数据库原理知识作为基础与保障,数据库原理基础知识能指导用户更好地把握、更有效地应用数据库技能。本章以关系数据库管理系统相关知识为主,浓缩性地介绍作者认为读者需要了解与把握的数据库基础知识。本章沿着数据管理发展阶段—数据库系统组成—数据库管理系统—数据模型—数据库体系结构这条概念主线讲解。

数据库的建设规模、数据库信息量的大小和使用频度已成为衡量国家信息化程度的重要标志之一。因此,数据库课程不仅是计算机科学与技术专业、信息管理专业的重要课程,也是许多非计算机专业的选修课程。

本章介绍数据库系统的基本概念,包括数据管理的发展过程、数据库系统的组成部分等。读者从中可以学习到为什么要使用数据库技术以及数据库技术的重要性,本章是后面各章节的准备和基础。

1.1 数据库系统综述

计算机开始用于数值计算,同时在非数值计算中得到了很广泛的应用,显示了它强大的生命力,在现代计算机应用领域中,数据处理占据较大市场份额,数据库技术是数据处理的最新研究成果,它的出现使得计算机应用渗透到工业、农业等领域,下面首先介绍基本概念。

1.1.1 信息数据、数据库、数据库管理系统

1. 信息与数据

数据库处理的基本对象是信息或表示信息的数据。数字只是最简单的一种数据,是数据的一种传统和狭义的理解。广义的理解认为数据的种类很多,文本(text)、图形(graph)、图像(image)、音频(audio)、视频(video)、学生的档案记录、货物的运输情况等,这些都是数据。

信息是向人们提供关于认识世界与改变世界的事实的知识。

数据的表现形式不可能完全表达其内容,需要经过解释。例如,10是一个数据,可以是一个人一双手的手指数,也可以是一个公交车号码。在日常生活中,人们可以直接用自然语言来描述事物。例如,可以这样来描述某校计算机系一名学生的基本情况:张华,男,1988年5月生,河南郑州人,2006年入学。在计算机中常常这样来描述:

(张华,男,198805,河南省郑州市,计算机系,2006)

2. 数据库

数据库(database,DB)是存放数据的仓库,具体来说,就是长期存放在计算机内的有组

织有顺序的数据集合。人们收集并抽取一个应用所需要的大量数据之后,应将其保存起来,以供进一步加工处理,进一步抽取有用信息。在科学技术飞速发展的今天,人们的视野越来越广,数据量急剧增加。

对一般用户来说数据库是什么样的呢?表面上,人们认为数据库内含有单位、企业或组织的形形色色的直观信息;在物理上,数据库实际上是存放在一个或多个磁盘上的若干物理文件。概括地讲,数据库数据具有永久存储、有组织和可共享三个基本特点。

3. 数据库管理系统

数据库管理系统(database management system, DBMS)是数据库系统的核心组成部分,也是一个大型复杂的软件系统,它的主要功能包括以下四方面。

(1)数据定义:DBMS提供数据定义语言(data definition language, DDL)来定义数据库的结构、数据库中数据之间的联系等。

(2)数据操作:DBMS还提供数据操作语言(data manipulation language, DML)来完成对数据的查询、插入、删除等操作。

(3)数据库的事务管理和运行管理:数据库在建立、运用和维护时由数据库管理系统统一管理、统一控制,以保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

(4)数据库的建立和维护功能:包括数据的输入和转换功能,数据库的转储、恢复功能,数据库的重组功能 and 性能监视、分析功能等。这些功能通常是由一些实用程序或管理工具完成的。

1.1.2 数据管理技术的进展

数据库技术是应数据管理任务的需要而产生的。要说数据管理,首先要提到数据处理,所谓数据处理是指对各种数据进行收集、存储、加工等一系列活动的总称。数据管理则是对数据进行分类、组织、编码、存储、检索和维护,它是数据处理的中心环节。

在应用需求的推动下,在计算机硬件、软件发展的基础上,数据管理技术经历了人工管理时期、文件系统时期、数据库系统时期3个时期。

1. 人工管理时期

20世纪50年代中期以前,计算机主要用于数值计算。当时的硬件有限:外存只有纸带、卡片、磁带,没有磁盘等直接存取的存储设备;软件当时没有操作系统,没有管理数据的专门软件。人工管理数据具有如下特点。

(1)数据不保存。由于当时计算机主要用于科学计算,一般不需要将数据长期保存,只是在计算某一课题时将数据输入,用完就删除。不仅对用户数据如此处理,对系统软件有时也是这样。

(2)应用程序管理数据。数据需要由应用程序自己设计、说明(定义)和管理,没有相应的软件系统负责数据的管理工作。应用程序中不仅要规定数据的逻辑结构,而且要设计物理结构,包括存储结构、存取方法、输入方式等,因此程序员负担很重。

(3)数据无法共享。数据是面向应用程序的,一组数据只能对应一个程序。

(4)数据独立性不强。数据的逻辑结构或物理结构发生变化后,应用程序也随之改变,

非常麻烦,加重了程序员的负担。

2. 文件系统时期

进入文件系统时期是在20世纪50年代末到60年代中期,这时硬件方面已有了磁盘、磁鼓等直接存取存储设备;软件方面,操作系统中已经有了专门的数据管理软件,这时程序与数据的关系如图1.1所示。

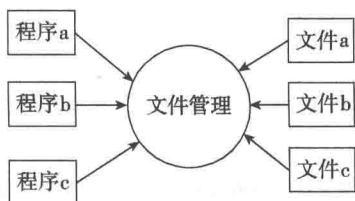


图 1.1 文件系统阶段应用程序与数据之间的对应关系

用文件系统管理数据具有如下特点。

(1)数据能长期保存。由于计算机大量用于数据处理,数据需要长期保留在外存上反复进行查询、修改、插入和删除等操作。

(2)由文件系统进行数据管理。由专门的软件即文件系统进行数据管理,文件系统把数据组织成相互独立的数据文件,利用“按文件名访问、按记录进行存取”的管理技术,可以对文件进行修改、插入和删除操作。文件系统实现了记录内的结构性,但整体无结构。程序和数据之间由文件系统提供存取方法进行转换,使应用程序与数据之间有了一定的独立性,程序员可以不必过多地考虑物理细节,将精力集中于算法。而且数据在存储上的改变不一定反映在程序上,大大节省了维护程序的工作量。

同时,文件系统仍存在以下缺点。

(1)数据共享性差,冗余度大。在文件系统中,一个(或一组)文件基本上对应于一个应用程序,即文件仍然是面向应用的。当不同的应用程序具有部分相同的数据时,必须建立各自的文件,而不能共享相同的数据,因此数据的冗余度大,浪费存储空间。同时由于相同数据的重复存储、各自管理,容易造成数据的不一致性,给数据的修改和维护带来困难。

(2)数据独立性差。文件系统中的文件是为某一特定应用服务的,文件的逻辑结构对该应用程序来说是优化的,因此,要想对现有的数据再增加一些新的应用会很困难,系统不容易扩充。

一旦数据的逻辑结构改变,必须修改应用程序,修改文件结构的定义。应用程序的改变,如应用程序改用不同的高级语言编写,也将引起文件数据结构的变化。因此,数据与程序之间仍缺乏独立性。可见,文件系统仍然是一个不具有弹性的无结构的数据集合,即文件之间是孤立的,不能反映现实世界事物之间的内在联系。

3. 数据库系统时期

20世纪60年代后期以来,计算机管理的对象规模越来越大,应用范围越来越广泛,数据量急剧增长,同时多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。

这时硬件已有大容量磁盘,硬件价格下降;软件则价格上升,因为编制和维护系统软件及应用程序所需的成本相对增加;在处理方式上,联机实时处理要求更多,并开始提出和考

虑分布式处理。在这种背景下,以文件系统作为数据管理手段已经不能满足应用的需求,于是为解决多用户、多应用共享数据的需求,使数据为尽可能多的应用服务,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——数据库系统。

用数据库系统来管理数据比文件系统具有更明显的优势,从文件系统到数据库系统,标志着数据管理技术的飞跃。下面详细讨论数据库系统的特点。

1.1.3 数据库系统的特点

与人工管理和文件系统相比,数据库系统的特点主要包括以下几方面。

1. 数据结构化

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

所谓“整体”结构化是指在数据库中的数据不再仅仅针对某一个应用,而是面向全组织的;不仅数据内部是结构化的,而且整体是结构化的,数据之间是具有联系的。

在文件系统中每个文件内部是有结构的,即文件由记录构成,每个记录由若干属性组成。例如,学生文件 student 的记录是由学生编号、姓名、性别、年龄、系、家庭住址、联系电话等属性组成;课程文件 course 和学生选课文件 sc 的结构如图 1.2 所示。

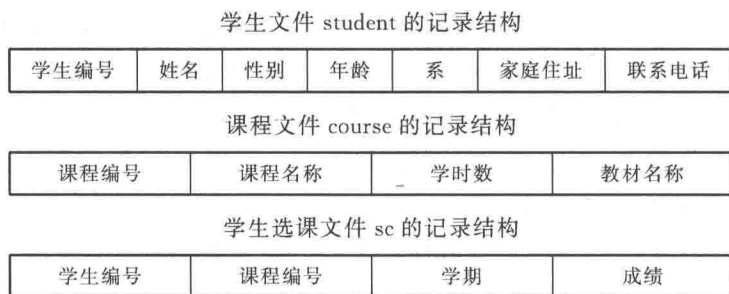


图 1.2 学生、课程、学生选课文件结构

在文件系统中,尽管其记录内部已有了某些结构,但记录之间没有联系。例如,学生文件 student、课程文件 course 和学生选课文件 sc 是 3 个独立的文件,但实际上这 3 个文件的记录之间是有联系的,sc 的学生编号必须是 student 文件中某个学生的学号,sc 的课程编号必须是 course 文件中某门课程的编号。

在关系数据库中,关系表的记录之间的这种联系是可以用参照完整性(将在第 2 章中详细讲解)来表述的。如果向 sc 中增加一个学生的考试成绩,但是这个学生并没有出现在 student 关系中,那么关系数据库管理系统(relational database management system, RDBMS)将拒绝执行这样的插入操作,从而保证了数据的正确性。而在文件系统中要做到这一点,必须由程序员编写一段代码在应用程序中实现。

在数据库系统中实现了整体数据的结构化。也就是说,不仅要考虑某个应用的数据结构,还要考虑整个组织的数据结构。例如,一个学校的信息系统中不仅要考虑教务处的学生学籍管理、选课管理,还要考虑学生处的学生人事管理,同时要考虑研究生院的研究生管理、人事处的教员人事管理、科研处的科研管理等。

在数据库系统中,数据是整体结构化的,存取数据的方式也很灵活,可以存取数据库中的某一个数据项、一组数据项、一个/一组记录。而在文件系统中,数据的存取单位是记录,粒度不能细到数据项。

2. 数据的共享性高,冗余度低,易扩充

数据库系统从整体角度看待和描述数据,数据不再面向某个应用,而是面向整个系统,因此数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余,节约存储空间,还能够避免数据之间的不相容性与不一致性。

所谓数据的不一致性,是指同一数据不同复制副本的值不一样。采用人工管理或文件系统管理时,由于数据被重复存储,当不同的应用使用和修改不同的复制副本时很容易造成数据不一致。在数据库中,数据共享减少了由于数据冗余造成的一致现象。

由于数据面向整个系统,是有结构的数据,不仅可以被多个应用共享使用,而且容易增加新的应用,这就使得数据库系统弹性大,易于扩充,可以适应各种用户的要求。可以选取整体数据的各种子集用于不同的应用系统,当应用需求改变或增加时,只要重新选取不同的子集或加上一部分数据,便可以满足新的需求。

3. 数据独立性高

数据独立性是数据库领域中的一个常用术语和重要概念,包括数据的物理独立性和数据的逻辑独立性。

物理独立性是指用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的。也就是说,数据在磁盘上的数据库中怎样存储是由 DBMS 管理的,用户程序不需要了解,应用程序要处理的只是数据的逻辑结构,当数据的物理存储结构改变时,应用程序不需要改变。

逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的,也就是说,数据的逻辑结构改变了,用户程序也可以不变。

数据独立性是由 DBMS 的二级映像功能来保证的,这部分内容将在下面讨论。

数据与程序独立,把数据的定义从程序中分离出去,加上存取数据的方法又由 DBMS 负责提供,从而简化了应用程序的编制,大大减少了应用程序的维护和修改。

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的数据集合。它可以供各种用户共享,具有最小的冗余度和较高的数据独立性。DBMS 在数据库建立、运用和维护时对数据库进行统一控制,以保证数据的完整性、安全性,并在多用户同时使用数据库时进行并发控制,在发生故障后对数据库进行恢复。

数据库系统的出现使信息系统从以加工数据的程序为中心转向以共享的数据库为中心的新阶段。这样既便于数据的集中管理,又有利于应用程序的研制和维护,提高了数据的利用率和相容性,同时提高了决策的可靠性。

目前,数据库已经成为现代信息系统的重要组成部分。具有数百吉字节、数百太字节,甚至数百皮字节的数据库已经普遍应用于科学技术、工业、农业、商业、服务业和政府部门的信息化系统中。

数据库技术是计算机领域发展最快的技术之一。数据库技术的发展是沿着数据模型的主线展开的。下面讨论数据模型。

1.2 数据模型

模型是现实世界的模拟和抽象。数据模型是对现实世界数据特征的抽象。例如,飞机模型抽象了飞机的基本特征,包括机头、机身、机翼和机尾,飞机模型还模拟了飞机的起飞、飞行和降落。特别是具体模型,人们并不陌生。一张地图、一组建筑设计沙盘、一架精致的航模飞机都是具体的模型。

数据模型应反映和规定本数据模型必须遵守的基本的通用的完整性约束条件。通俗地讲,数据模型就是现实世界的模拟。数据模型可分为层次模型、网状模型和关系模型。

现有的数据库系统均是基于某种数据模型的,数据模型是数据库系统的核心和基础。因此,了解数据模型的基本概念是学习数据库的基础。

根据模型应用的不同目的,可以将这些模型划分为两类,它们分别属于两个不同的层次:第一类是概念模型,第二类是逻辑模型和物理模型。

概念模型(conceptual model)也称信息模型,它是按用户的观点来对数据和信息建模,主要用于数据库设计。

逻辑模型主要包括层次模型(hierarchical model)、网状模型(network model)、关系模型(relational model)、面向对象模型(object oriented model)和对象关系模型(object relational model)等。它按计算机系统的观点对数据建模,主要用于 DBMS 的实现。

物理模型是对数据最低层的抽象,它描述数据在系统内部的表示方式和存取方法,是面向计算机系统的。物理模型的具体实现是 DBMS 的任务,数据库设计人员要了解和选择物理模型,一般用户则不必考虑物理级的细节。

数据模型是数据库系统的基础和主要部分。各种机器上实现的 DBMS 软件都是在某种数据模型的基础上或者说是支持某种数据模型的。

为了把现实世界中的具体事物抽象、组织为某一 DBMS 支持的数据模型,人们常常首先将现实世界抽象为信息世界。也就是说,首先把现实世界中的客观对象抽象为某一种信

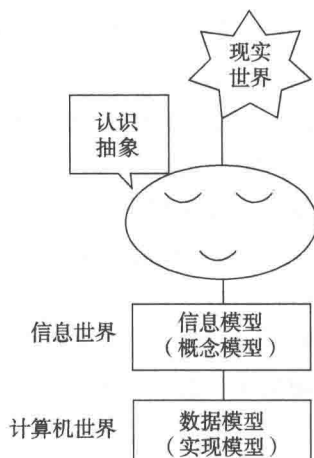


图 1.3 现实世界中客观对象的抽象过程

息结构,这种信息结构并不依赖于具体的计算机系统,不是某一个 DBMS 支持的数据模型,而是概念级的模型;然后把概念模型转换为计算机上某一 DBMS 支持的数据模型,这一过程如图 1.3 所示。

从现实世界到概念模型的转换是由数据库设计人员完成的,从概念模型到逻辑模型的转换可以由数据库设计人员完成,也可以由数据库设计工具协助设计人员完成,从逻辑模型到物理模型的转换一般是由 DBMS 完成的。

下面首先介绍数据模型的共性,即数据模型的组成要素,然后分别介绍两类不同的数据模型,分别是概念模型和逻辑模型。

一般来讲,数据模型是严格定义的一组概念的集合。数据模型通常由数据结构、数据操作和完整性约束条件三部分组成。

(1)数据结构。数据结构是所研究的对象类型的集合,包括两类:一类是与数据类型、内容、性质有关的对象;另一类是与数据之间联系有关的对象。

数据结构刻画了一个数据模型性质最重要的方面。因此在数据库系统中,人们通常按照其数据结构的类型来命名数据模型。例如,层次结构、网状结构和关系结构的数据模型分别命名为层次模型、网状模型和关系模型。

总之,数据结构是所描述的对象类型的集合,是对系统静态特性的描述。

(2)数据操作。数据操作是指对数据库中各种对象(型)的实例(值)允许执行的操作的集合,如插入、删除、修改等。

数据操作是对系统动态特性的描述。

(3)数据完整性约束条件。数据完整性约束条件是一组完整性规则。数据模型应反映和规定本数据模型必须遵守的基本的通用的完整性约束条件。完整性规则是给定的数据模型中数据及其联系所具有的制约和依存规则,用以限定符合数据模型的数据库状态以及状态的变化,以保证数据的正确、有效、相容。

此外,数据模型还应该提供定义完整性约束条件的机制,以反映具体应用所涉及的数据必须遵守的特定的语义约束条件。例如,在某大学的数据库中规定学生成绩如果有 7 门以上不及格将不能授予学士学位,教授的退休年龄是 65 周岁,男职工的退休年龄是 60 周岁,女职工的退休年龄是 55 周岁等。

1.2.1 概念模型

概念模型也称为信息模型,是人们为正确直观地反映客观事物及其联系,对所研究的信息世界建立的一个抽象的模型,是现实世界到信息世界的第一层抽象,是数据库设计人员和用户之间进行交流的语言,概念模型实际上是现实世界到机器世界的一个中间层次。

概念模型一方面应该具有较强的语义表达能力,能够方便、直接地表达应用中的各种语义知识;另一方面还应该简单、清晰、易于用户理解。

1. 信息世界中的基本概念

信息世界涉及的概念主要有以下几个。

1) 实体(entity)

客观存在并可相互不同的事物称为实体。实体可以是具体的人、事、物,也可以是抽象

的概念或联系,例如,一个职工、一个学生等都是实体。

2) 属性(attribute)

属性就是实体所具有的特性,一个实体可以由若干属性来刻画。例如,工人实体可以由工号、姓名、性别、出生年月、所在车间、入厂时间等属性来描述。

3) 实体型(entry type)

用实体名及其属性名集合来抽象和刻画同类实体。具有相同属性的实体必然具有共同的特征和性质。例如,工人(工号,姓名,性别,出生年月,所在车间,入厂时间)就是一个实体型。

4) 键(key)

键能够唯一地标识出一个实体集中每一个实体的属性或属性组合,键也称为关键字或码。例如,学号是学生实体的码。

5) 域(domain)

属性的取值范围称为域。域是一组具有相同数据类型的值的集合。属性的取值范围来自某个域。例如,学号的域为8位整数,姓名的域为字符串集合,学生年龄的域为整数,性别的域为(男,女)。

6) 实体集(entity set)

具有相同属性的实体的集合称为实体集。例如,全体学生就是一个实体集。

7) 联系(relationship)

在现实世界中,事物内部以及事物之间是有联系的,这些联系在信息世界中反映为实体(型)内部的联系和实体(型)之间的联系。实体内部的联系通常是指组成实体的各属性之间的联系;实体之间的联系通常是指不同实体集之间的联系。

2. 两个实体型之间的联系

两个实体型之间的联系可以分为三种。

1) 一对一联系(1:1)

如果对于实体集 A 中的每一个实体,实体集 B 中至多有一个(也可以没有)实体与之联系,反之亦然,则称实体集 A 与实体集 B 具有一对一联系,记为 1:1。

例如,学校里面,一个班级只有一个班长,而一个班长只在一个班级中任职,则班级与班长之间具有一对一联系。

2) 一对多联系(1:n)

如果对于实体集 A 中的每一个实体,实体集 B 中 n 个实体($n \geq 0$)与之联系,反之,对于实体集 B 中的每一个实体,实体集 A 中至多只有一个实体与之联系,则称实体集 A 与实体集 B 有一对多联系,记为 1:n。

例如,一个班级中有若干名学生,每个学生只在一个班级中学习,则班级与学生之间具有一对多联系。

3) 多对多联系(m:n)

如果对于实体集 A 中的每一个实体,实体集 B 中有 n 个实体($n \geq 0$)与之联系,反之,对于实体集 B 中的每一个实体,实体集 A 中有 m 个实体($m \geq 0$)与之联系,则称实体集 A 与实体集 B 具有多对多联系,记为 $m:n$ 。

例如,一门课程同时有若干学生选修,而一个学生可以同时选修多门课程,则课程与学生之间具有多对多联系。

实际上,一对一联系是一对多联系的特例,而一对多联系又是多对多联系的特例。

可以用图形来表示两个实体型之间的这三类联系,如图 1.4 所示。

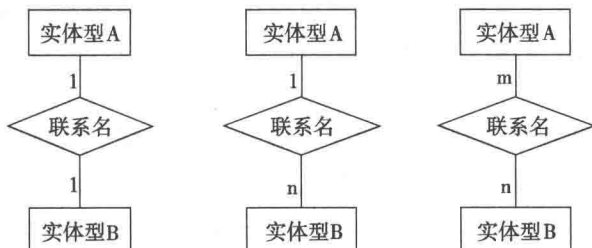


图 1.4 实体之间的三种联系

3. 两个以上的实体型之间的联系

一般地,两个以上的实体型之间也存在着一对一、一对多、多对多联系。

若实体型 E_1, E_2, \dots, E_n 之间存在联系,对于实体型 $E_i (j=1, 2, \dots, i-1, i+1, \dots, n)$ 中的给定实体,最多只和 E_i 中的一个实体相联系,则说 E_i 与 $E_1, E_2, \dots, E_{i-1}, E_{i+1}, \dots, E_n$ 之间的联系是一对多的。

例如,对于课程、教师与参考书 3 个实体型,如果一门课程可以由多个教师讲授,使用多本参考书,而每一个教师只讲授一门课程,每一本参考书只供一门课程使用,则课程与教师、参考书之间的联系是一对多的,如图 1.5 所示。

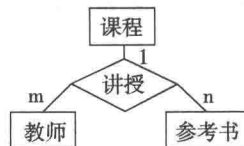


图 1.5 3 个实体之间的联系示例

4. 概念模型的一种表示方法:实体-联系方法

概念模型是对信息世界建模,所以概念模型应该能够方便、准确地表示出上述信息世界中的常用概念。概念模型的表示方法很多,其中最著名、最常用的是 P. P. S. Chen 于 1976 年提出的实体-联系方法(entity-relationship approach)。该方法用 E-R 图(E-R diagram)来描述现实世界的概念模型,E-R 方法也称为 E-R 模型。

E-R 图提供了表示实体型、属性和联系的方法。

实体型:用矩形表示,矩形框内写明实体名。

属性:用椭圆形表示,并用无向边将它与相应的实体型连接起来。

例如,学生实体具有学号、姓名、性别、出生年份、系、入学时间等属性,用 E-R 图表示如图 1.6 所示。

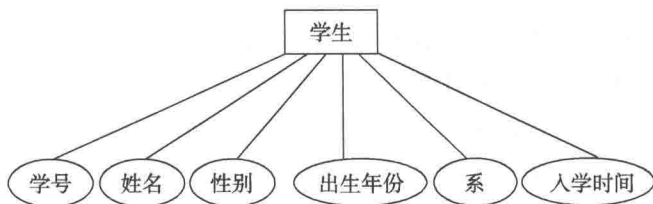


图 1.6 学生实体及其属性