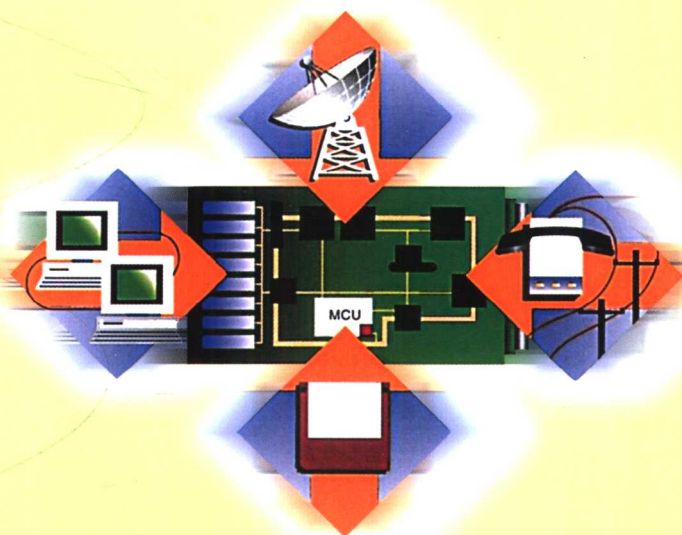





普通高等教育“十一五”规划教材

数据库原理教程

范明 叶阳东 邱保志 职为梅 编著



 科学出版社
www.sciencep.com

TP311.13/336

2008

普通高等教育“十一五”规划教材

数据库原理教程

范明 叶阳东 邱保志 职为梅 编著

科学出版社

北京

内 容 简 介

本书全面阐述了数据库系统的基本概念、理论、方法和技术。全书共分12章,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、完整性与安全性、关系数据库的设计理论、数据库设计、查询处理与优化、事务与并发控制、数据库的恢复技术、XML 和 ODBC 编程。附录给出了实验和课程设计,用于配合课堂教学。

本书涵盖了数据库系统原理入门课程的基本内容,同时包含了 XML 和 ODBC 编程等实用的较新知识点。本书理论与实践兼顾。关系数据库设计理论的形式化证明,以章后附录的形式给出,既保证了理论的完整性,又可以使得关注技术的读者摆脱繁琐证明的困扰。

本书是为高等学校计算机、信息科学及其相关专业本科生和软件学院学生的第一门数据库课程编写的教材,也适合希望掌握数据库理论、方法和技术的工程技术人员阅读。

图书在版编目(CIP)数据

数据库原理教程/范明等编著. —北京:科学出版社,2008

普通高等教育“十一五”规划教材

ISBN 978-7-03-021217-7

I. 数… II. 范… 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 027406 号

责任编辑:马长芳 潘继敏 / 责任校对:刘小梅

责任印制:张克忠 / 封面设计:陈 敬

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>

盛 立 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2008年4月第一版 开本:B5(720×1000)

2008年4月第一次印刷 印张:23 1/2

印数:1—4 000 字数:459 000

定价:35.00元

(如有印装质量问题,我社负责调换〈环伟〉)

前 言

数据库技术是现代信息处理的核心技术之一,数据库管理系统(DBMS)是使用最广泛的软件系统之一。因此,数据库系统原理已经成为计算机科学教育的一个必不可少的部分。本书是为本科生数据库课程编写的教科书,涵盖了数据库系统原理入门课程的基本内容。

数据库是计算机学科的主流领域之一。尽管数据库技术的内涵和外延都在不断扩展,新的学科分支还在迅速成长,但是作为入门知识的概念、原理和技术已经趋于稳定。这些入门知识包括数据库系统的三级模式和两级映像、实体-联系模型、关系数据模型、关系数据库设计理论、数据的完整性与安全性、查询处理与优化、事务处理、并发控制与恢复、SQL 语言等。

本书介绍数据库系统的一般概念和技术。书中的概念和算法基于商品化或实验数据库系统所采用的概念和算法,但并不与一个特定的数据库系统联系在一起。通过这些内容的学习,读者容易理解和掌握具体的数据库系统。

本书理论与实践兼顾。关系数据库的设计理论不仅体现了数据库研究的重要成果,同时也为设计好的关系模式提供了具体方法和技术。关系数据库设计理论的形式化证明从正文中分离,以章后附录的形式给出,这样既保证了理论的完整性,又可以使得关注技术的读者摆脱繁琐证明的困扰。除了数据库系统原理入门课程的基本内容外,本书还包括 XML 和 ODBC 编程等实用的较新知识点。对于实际应用和数据库应用系统的开发,这些内容是有用的。为了加强实践环节,本书还在书后附录提供了一些可供选择的实验和课程设计的参考选题。

本书由范明主持编写。叶阳东编写了第 9、10 章,邱保志编写了第 5、7、11 章,职为梅编写了第 4、12 章和附录,其余部分由范明编写。范明还对各章做了修改并最终定稿。

本书的组织

全书共分 12 章和一个附录:

第 1 章是数据库系统的一般综述,所涉及的内容将在之后各章进一步展开讨论。

第 2 章介绍实体-联系模型(E-R 模型)。E-R 模型概念简单,并具有很强的语义表达能力,广泛用于对现实世界建模。

第 3 章介绍数据库系统的关系数据模型,包括与之相关的抽象语言关系代数和关系演算。这些抽象语言是实际数据库系统语言(如 SQL)的基础。

第4章介绍广泛使用的关系语言 SQL。SQL 是事实上的关系数据库标准语言,几乎所有的 RDBMS 都支持 SQL 语言。SQL 的介绍主要基于 SQL-92,同时也包含 SQL-99 的部分特色。

第5章介绍数据库的完整性和安全性,包括 SQL 对完整性和安全性的支持。

第6章介绍关系数据库的设计理论,包括数据依赖和规范化。为了适应不同读者的需要,理论结果的证明从正文分离,在该章附录中给出。

第7章介绍数据库应用系统的设计,涵盖从需求分析到数据库的建立和维护整个过程。

第8章介绍查询处理与优化,包括基本运算算法和基于保持等价性的查询变换和查询优化方法。

第9、10章介绍事务处理。其中,第9章介绍事务的一般概念,讨论并发控制;第10章介绍在出现故障的情况下保证事务正确执行的数据库恢复技术。

第11、12章分别介绍两个相对较新的主题:XML 和 ODBC。XML 已经成为事实上的异构数据库数据交换的标准语言,并广泛用于复杂数据类型的处理。ODBC 建立了一组规范,并提供了一组访问数据库的标准应用程序接口。对于实际数据库应用系统开发,这些内容都是十分有用的。

附录提供了一些实验和课程设计选题,用于配合课堂教学,加深对数据库原理的理解。

致教师

通过适当剪裁,本书可以用于计算机科学与技术、信息科学及其相关专业和软件学院本科生的第一门数据库课程教学。

在本书出版之前,我们在郑州大学计算机科学与技术专业和软件学院的本科生教学中使用过本书的内容。根据我们的经验,第1~5章和第9~10章适用所有这些学生。对于计算机科学与技术专业的学生,第6、8章是必需的,其他几章可以根据学时选讲。对于软件学院的学生,第6章的重点放在函数依赖、范式和规范化方法上,可以讲述第7、12章,其他章节根据情况考虑。

实验和课程设计是数据库课程教学的必要环节。实验可以配合课堂教学进行。课程设计可以在第7章之后布置,期末完成;也可以利用学校安排的实习时间,集中两周完成。在我们的教学实践中,两种方法都采用过。

课件的 ppt 文档和其他教学辅助材料将在本书网站(www.dbdm.zzu.edu.cn)提供。部分习题的参考答案正在准备,将向使用本书的教师提供。本书的网站正在建设,在此之前,你也可以向我们索取有关教学辅助材料。

致读者

本书仅要求读者熟悉基本的数据结构和一种高级程序设计语言(如 C、Java 或 Pascal)。集合论和数理逻辑的知识(如离散数学中讲述的)有助于该书的理解。书

中的概念都以直观的方式描述,并通过一些精心选择的例子加以解释。

本书是为高等学校计算机、信息科学及其相关专业本科生和软件学院学生的第一门数据库课程编写的教材,也适合希望比较完整地掌握数据库理论、方法和技术的工程技术人员阅读。

课件的 ppt 文档将在本书网站(www.dbdm.zzu.edu.cn)提供,其他辅助学习材料也正在准备,陆续在该网站提供。你也可以与我们联系,索取相关材料,提出你的意见和建议。

致谢

在近 20 年的数据库教学中,我们使用和参考过许多国内外优秀教材。这些教材不仅为我们过去的教学提供了方便和支持,也深深影响着本书的编写。我们在此向这些教材(见参考文献)的所有作者和译者表示敬意和感谢。

感谢郑州大学选修数据库系统原理课程的历届学生。在 20 年的教学相长过程中,他们的求知欲推动我们不断思考如何选择和组织教学内容,如何用简洁而又不失严谨的表述方式讲述数据库的基本概念和技术,最终形成这本教材。

感谢科学出版社的编辑们,感谢他们对出版本书的立项支持,感谢他们在我们几度拖延交稿时表现出的耐心。没有他们的支持、鼓励和辛勤工作,本书不可能这么快与读者见面。

希望读者喜欢这本书,希望本书能够帮助读者掌握数据库系统的基本概念、原理和技术,希望本书能够成为将来步入数据库领域的读者的垫脚石,为他们以后更好地工作打下坚实的基础。

书中的错误和不当之处,敬请同仁和读者朋友指正。意见和建议请发往 mfan@zzu.edu.cn,我们将不胜感激。

作者

2008 年 1 月于郑州大学

目 录

前言

第 1 章 数据库系统概述	1
1.1 引言	1
1.2 数据模型	4
1.3 数据库系统的结构	8
1.4 数据库语言	12
1.5 数据库管理员与数据库用户	14
1.6 数据库技术的发展	16
1.7 小结	17
习题	18
第 2 章 实体联系模型	19
2.1 基本概念	19
2.2 实体-联系图	24
2.3 一个例子	26
2.4 弱实体集	28
* 2.5 扩展的 E-R 图	30
2.6 小结	35
习题	36
第 3 章 关系数据模型	38
3.1 关系数据库的结构	38
3.2 从 E-R 模型到关系模型	42
3.3 关系的完整性约束	48
3.4 关系代数	50
3.5 关系演算	61
* 3.6 关系数据库的更新	67
3.7 小结	69
习题	69
第 4 章 关系数据库标准语言 SQL	73
4.1 引言	73
4.2 数据定义	76

4.3	数据查询	85
4.4	数据更新	103
4.5	视图	106
4.6	嵌入式 SQL	112
4.7	小结	124
	习题	125
第 5 章	完整性与安全性	128
5.1	完整性概述	128
5.2	实体完整性和参照完整性	129
5.3	域约束	133
5.4	用户定义的完整性	135
5.5	触发器	138
5.6	安全性概述	141
5.7	用户标识与鉴别	143
5.8	存取控制	144
5.9	SQL 的授权	149
5.10	其他安全措施	154
5.11	小结	155
	习题	156
第 6 章	关系数据库的设计理论	159
6.1	问题提出	159
6.2	函数依赖	162
6.3	函数依赖的推导	165
6.4	关系模式的分解	171
6.5	关系模式的范式	176
6.6	将关系模式分解成高级范式	179
6.7	多值依赖与 4NF	183
6.8	在设计中使用规范化理论	190
6.9	小结	194
	习题	195
	附录: 本章引理和定理证明	197
第 7 章	数据库设计	206
7.1	数据库设计概述	206
7.2	需求分析	210
7.3	数据库概念设计	215

7.4	逻辑结构设计	222
7.5	数据库物理设计	225
7.6	数据库的实施和维护	229
7.7	小结	232
	习题	232
第8章	查询处理与优化	234
8.1	查询处理概述	234
8.2	选择运算的实现	237
8.3	连接运算的实现	240
8.4	查询优化	245
8.5	代数优化	247
8.6	物理优化	256
8.7	小结	261
	习题	262
第9章	事务与并发控制	264
9.1	事务的概念	264
9.2	并发控制概述	268
9.3	并发调度的可串行化	271
9.4	基于锁的协议	276
9.5	两阶段锁协议	281
9.6	多粒度封锁	283
9.7	小结	286
	习题	286
第10章	数据库的恢复技术	288
10.1	数据库恢复概述	288
10.2	存储器结构	290
10.3	基于日志的恢复技术	292
10.4	基于检查点的恢复技术	297
*10.5	缓冲技术	299
10.6	介质故障恢复技术	301
*10.7	其他恢复技术	302
10.8	小结	304
	习题	305
第11章	XML	306
11.1	概述	306

11.2	XML 数据结构	308
11.3	XML 文档模式	311
11.4	查询和转换	316
11.5	应用程序接口	324
11.6	XML 数据的存储	325
11.7	XML 应用	327
11.8	小结	328
	习题	329
第 12 章	ODBC 编程	332
12.1	ODBC 简介	332
12.2	ODBC 的工作原理	333
12.3	ODBC API 基础	340
12.4	ODBC 的工作流程	343
12.5	ODBC 应用实例	347
12.6	小结	351
	参考文献	352
	附录 实验与课程设计	358
1	实验	358
2	课程设计	362

第 1 章 数据库系统概述

数据库技术产生于 20 世纪 60 年代中期,经历了 40 余年的发展,已经成为计算机学科的重要分支。数据库技术是信息系统的核心和基础,它的出现极大地促进了计算机应用向各行各业的渗透。现在,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一个国家信息化程度的重要标志。

数据库管理系统(database management system, DBMS)是一种重要的程序设计系统,它由一个相互关联的数据集合和一组访问这些数据的程序组成。这个数据集合称为数据库,它包含了一个企业、政府部门或一个单位的全部信息。DBMS 的基本目标是提供一个方便、有效地访问这些信息的环境。DBMS 是最广泛使用的软件系统之一。为了设计和实现 DBMS,已经建立了一些原理、方法和技术。理解这些原理、方法和技术对于有效地使用 DBMS 是至关重要的。

本章是数据库系统的概述,简要介绍数据库的一些基本概念。这些内容的进一步讨论遍及本书的其余章节。

1.1 引言

今天,数据库已经无处不在。本节我们简略介绍 DBMS 产生的原因、DBMS 的主要功能、什么是数据库和使用数据库的优点。

1.1.1 为什么需要数据库管理系统

“需要是发明之母”。数据库技术和数据库管理系统的产生和发展的直接动力源于数据管理的需要。

1. 数据管理的例子

几乎所有的行业和部门都存在并且不断产生大量数据。为了维持正常运作,这些行业和部门都需要持久地存储、维护和管理它们的数据。下面是数据管理的一些典型例子:

- 零售业:管理产品、客户和购买信息。
- 银行业:管理客户、账户和存贷款信息。
- 制造业:管理供应、订单、库存、销售等信息。
- 交通:例如,航空公司管理航班和订票信息,铁路部门管理火车票销售信息,

公路交通部门管理班车和客车票销售信息。

- 电信业:管理通信网络信息、存储通话记录、维护电话卡余额。
- 图书馆:管理图书资料、读者和借阅信息。
- 政府部门:例如,税务部门需要管理纳税人信息和纳税信息,交管部门需要管理车辆信息和驾驶员信息。
- 学校:管理学生、教师、课程信息和学生成绩。

要将这些数据存储在计算机系统中,所面临的共同问题是,如何合理地组织数据、如何有效地访问数据。

2. 数据管理的基本操作

不同的应用涉及不同的数据。许多应用表面看来很不相同,似乎没有什么共同点。例如,超市存储的数据与银行存储的数据内容很不相同,用法也很不相同。然而,对于数据管理,所有的应用都需要如下基本操作:

(1) 数据查询:从计算机系统(暂且称它为数据库)中查找用户需要的信息。例如,查找图书,查阅学生成绩,查找商品价格等。

(2) 数据插入:将新的数据输入到数据库中。用数据库的术语,这种操作称作“插入”。例如,新产品的信息需要输入数据库,新的银行账户信息需要输入数据库,等等。

(3) 数据删除:从数据库删除不再需要的数据。例如,车辆报废,其相关信息要从数据库中删除;账户注销,其信息要从相关数据库中删除;等等。

(4) 数据修改:修改数据库中某些数据。例如,某种商品降价,修改它的价格。本质上,修改可以用删除+插入实现。然而,修改作为一种单独的操作是方便的。

数据的插入、删除和修改统称**数据更新**。几乎所有的更新操作都涉及隐含的查询。例如,订购8月1日从郑州到广州CZ3971航班的机票,本质上是导致该航班的机票数减1。但是,我们必须首先查询是否有8月1日CZ3971航班;有的话,是否有剩余机票。

3. 实际应用对数据操作的要求

实际应用是在并发、充满故障和错误的环境下运行的。这对数据操作的实现提出了很高的要求。这些要求包括:

- (1) 并发访问:允许多个用户同时对数据库中的数据进行访问。
- (2) 面临故障:各种各样的故障都可能发生,必须确保数据在任何情况下都不被破坏。对于许多应用,这一点至关重要。例如,银行存款数据不能因突然停电而丢失或破坏。
- (3) 数据的安全性:防止用户对数据进行未经授权的访问。例如,我们可以允

许银行客户查看自己的存款,但不能允许他们修改。

(4) 数据的完整性:防止不符合语义的数据进入数据库。例如,我们不能将负数作为学生的成绩输入到数据库中。

(5) 数据的一致性:防止数据库进入不一致状态。有些操作必须作为一个整体(原子性)。例如,从账户 A 转 1000 元到账户 B,涉及将账户 A 的存款额减去 1000 元,将账户 B 的存款增加 1000 元。这两个操作要么都做,要么都不做。否则,数据库将进入不一致状态。故障可能导致两个操作中的一个完成,而另一个未完成。必须保证即使发生故障也不会影响数据库的一致性。

这些要求,加上查询条件的多样性和复杂性,使得数据查询和更新的实现很复杂。共同的需要值得开发专门的软件程序实现,不必每个应用都写类似的程序。开发专门的软件系统管理数据,提供数据的组织和基本操作是必要的。这种软件系统就是数据库管理系统,即 DBMS。

1.1.2 数据库管理系统与数据库

DBMS 是位于用户和计算机操作系统之间的数据管理软件,它提供如下功能:

(1) 数据定义:提供数据定义语言(data definition language, DDL),用于定义数据库中的数据对象和它们的结构。

(2) 数据操纵:提供数据操纵语言(data manipulation language, DML),用于操纵数据,实现对数据库的基本操作(查询、插入、删除和修改)。

(3) 事务管理和运行管理:统一管理数据、控制对数据的并发访问,保证数据的安全性、完整性,确保故障时数据库中数据不被破坏,并且能够恢复到一致状态。

(4) 数据存储和查询处理:确定数据的物理组织和存取方式,提供数据的持久存储和有效访问;确定查询处理方法,优化查询处理过程。

(5) 数据库的建立和维护:提供实用程序,完成数据库数据批量装载、数据库转储、介质故障恢复、数据库的重组和性能监测等。

(6) 其他功能:包括 DBMS 与其他软件通信、异构数据库之间数据转换和互操作等。

数据库(database, DB)是持久储存在计算机中有组织的、可共享的大量数据的集合。数据库中的数据按一定的数据模型组织、描述和存储,可以被各种用户共享,具有较小的冗余度、较高的数据独立性,并且易于扩展。

在数据库中,使用数据模型对数据建模,所产生设计结果称为**数据库模式**。数据库模式描述数据库的数据结构(型),具有相对稳定性。特定时刻数据库中的数据称为数据库的**实例(值)**。数据库的值是随时间推移不断变化的。

数据库系统由数据库、DBMS(及其开发工具)、应用系统和数据库管理员组成。使用数据库进行信息管理具有如下优点:

(1) 数据整体结构化:在数据库中,数据的组织面向整个机构、面向所有可能的应用,而不是某个具体部门或某个特定的应用。数据结构化是整体结构化,数据结构不仅描述现实世界的对象,而且描述对象之间的联系。

(2) 数据共享:数据库中的数据面向整个机构组织使得它能够更好地被多个用户、多个应用程序共享。不仅已有的应用可以共享数据库中的数据,而且新的应用也能对这些数据进行操作。共享的好处是节省存储空间,尽量避免同一数据不必要地重复存放(冗余)。从而在某种程度上避免了同一数据的不同副本具有不同值(数据的不一致性)。

(3) 数据独立性:数据独立性是指数据与应用程序相互独立,包括数据的物理独立性和数据的逻辑独立性。数据的结构用数据模型定义,无需程序定义和解释。数据库系统的三级模式和两级映像使得数据的存储结构和逻辑结构的改变不会影响应用程序(见 1.3 节)。

(4) 数据由 DBMS 同一管理和控制,使得系统能够为数据管理提供更多的支持。这些支持包括:

① 提供事务支持:事务是一个逻辑单元,包括一系列操作,这些操作要么都做,要么都不做,即便发生故障也如此(见第 9 章)。

② 增强安全性:DBMS 提供数据的安全性保护,使每个用户只能按指定方式使用和处理指定数据,保护数据以防止不合法的使用造成的数据的泄密和破坏(见第 5 章)。

③ 保持完整性:DBMS 提供数据的完整性检查,将数据控制在有效的范围内,保证数据之间满足一定的关系(见第 5 章)。

④ 平衡相互冲突的请求:并发控制对多用户的并发操作加以控制和协调,防止相互干扰而导致错误的结果(见第 9 章)。

⑤ 面对故障的弹性:当系统发生故障时,防止数据库的数据丢失,并将数据库从错误状态恢复到某种已知的正确状态(见第 10 章)。

(5) 标准化:使用数据库进行信息管理有利于制定部门标准、行业标准、工业标准、国家标准和国际标准,促进数据库管理系统和数据库开发工具的研制、开发,推动数据管理应用的健康发展。

1.2 数据模型

数据模型是数据库技术的核心概念。所有的 DBMS 都基于某种数据模型实现,所有的数据库应用都建立在某种数据模型之上。**数据模型**是一种形式机制,用于数据建模,描述数据、数据之间的联系、数据的语义、数据上的操作和数据的完整性约束条件。一种好的数据模型要能准确地描述现实世界,容易理解和易于实现。

对数据建模的模型分成两个不同的层次:概念模型也称信息模型,按用户的观

点来对现实世界进行数据建模。数据模型按计算机系统的观点对信息世界进行数据建模。

1.2.1 实体-联系模型

实体联系(entity-relationship, E-R)模型是一种广泛使用的概念模型,用于对现实世界建模。E-R模型基于这样的认识:现实世界由一些称为实体的基本对象和这些对象之间的联系组成。

实体可以是实际存在的事物(如学生),也可以是抽象概念(如课程)。实体用一些称为属性的特征刻画。例如,学生可以用学号、姓名、性别、出生年月、专业等属性刻画;课程可以用课程号、课程名、学时、学分等属性刻画。具有相同属性的实体被汇集在一起,形成实体集。例如,所有的学生形成一个实体集,所有的课程也形成一个实体集。

联系是实体之间的关联。它反映了现实世界中实体之间客观存在的关联关系。例如,如果学生李明选修了数据库原理这门课程,李明与数据库原理之间就存在一种联系。同一类联系汇集在一起,形成联系集。例如,所有表明学生与课程之间选修关系的联系形成一个联系集“选修”。联系也可能需要属性刻画。例如,选修联系可以包含一个属性“成绩”,描述特定的学生选修特定课程取得的成绩。

E-R模型用一种称为E-R图的图形对现实世界建模。图1.1是一个E-R图示例。图中,矩形框代表实体集,椭圆代表属性,菱形框代表联系集,一些线段将属性与对应的实体集或联系相连接,而另一些将参与联系的实体集连接到联系集上。

E-R模型的更多细节将在第2章给出。

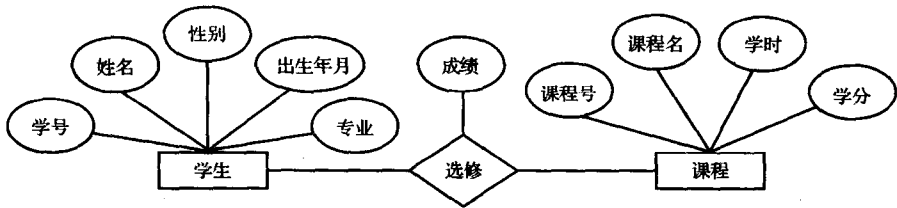


图 1.1 E-R 图示例

1.2.2 数据模型的三要素

数据模型是实际 DBMS 支持的模型。数据模型有数据结构、数据操作和完整性约束三个基本要素。

1. 数据结构

数据结构描述数据库中的对象和对象之间的联系,是对系统静态特性的描述。

数据结构定义基本数据项的类型,如何用基本数据项构造更大数据对象,如何表示数据对象之间联系,以及联系具有的类型等。

尽管不同的数据模型使用不同的术语,但是用基本数据项构造更大数据对象的方法是类似的——把描述现实世界同一对象的数据项组织成记录(关系模型称之为元组)。然而,对于描述对象之间的联系,不同的数据模型提供了不同的方法。例如,层次模型要求所有的记录型组织成一棵树,并且父节点记录型与子节点记录型之间的联系是一对多联系。网状模型则把记录型的组织放宽为有向无环图。而在关系模型中,对象本身和对象之间的联系都用关系表示,并且可以直接表示多对多联系。

数据结构刻画了数据模型最重要的方面。在数据库系统中,数据模型通常按它所使用数据结构来命名。例如,将数据按照树状结构加以组织的数据模型被称为层次模型,用有向无环图组织数据的数据模型被称为网状模型,而用关系组织数据的模型被称为关系模型。

2. 数据操作

数据操作定义数据库中各种数据对象的实例上允许执行的操作和操作规则,是对系统动态特性的描述。

数据库操作主要包括查询和更新(包括插入、删除、修改)两大类。数据模型定义这些操作的运算对象、运算符、运算的确切含义和运算规则(如优先级),并且提供实现这些操作的语言。

3. 数据的完整性约束条件

数据的完整性约束条件是一组规则,用以限定符合数据模型的数据库状态和状态的变化,保证数据的正确、有效和相容。

数据的完整性约束条件可以分通用的约束条件和与具体应用相关的专用约束条件。所谓**通用完整性约束条件**是指该模型下的所有数据库都必须满足的约束条件。例如,在关系模型中,任何关系必须满足实体完整性和参照完整性两个约束条件。而与具体应用相关的**专用完整性约束条件**取决于实际问题的语义。例如,零件的库存量必须是非负整数,学生的成绩必须在0~100取值。

通常,数据模型只明确规定该模型的通用完整性约束规则,而对专用完整性约束条件,由DBMS提供定义和检查机制。

1.2.3 关系模型

目前,关系模型是主流数据模型。几乎所有的DBMS都建立在关系模型上,或支持关系模型。长期以来,关系数据库也一直是数据库的研究与开发的重点。

关系模型具有坚实的数学基础、简洁的数据表示形式,并且支持易学易用的非过程化语言。所有这些都使关系模型具有其他数据模型无法比拟的优点。

关系模型只有一种数据结构——关系。现实世界中的对象和对象之间的联系都用关系表示。关系是元组的集合。从用户角度来看,关系是一张二维表。表的第一行称为表头,它列出关系的属性,刻画关系的模式;其余各行是关系的元组。

使用 E-R 模型的术语,实体集和联系集都用关系表示。如果关系表示实体集,则表的每一行(第一行除外,下同)代表一个实体,表的每一列代表刻画实体特征的一个属性。如果关系表示联系集,则表的每一行代表实体之间的一个联系,而表的列给出参与联系的实体的码属性和联系的属性。例如,图 1.1 的 E-R 模型可以用图 1.2 所示的三个表表示。

学号	姓名	性别	出生年月	专业
200705001	王万里	男	1988.12	计算机科学与技术
⋮	⋮	⋮	⋮	⋮

课程号	课程名	学时	学分
CS101	计算科学导论	32	2
⋮	⋮	⋮	⋮

学号	课程号	成绩
200705001	CS101	88
⋮	⋮	⋮

图 1.2 关系的二维表表示

关系模型要求关系必须是规范化的:关系的每个属性只能取原子值(不能再分的值)。换句话说,表中不能包含子表。图 1.3 就是一个非规范化的表,其中工资和扣除都是子表,可以划分成更小的数据项。

职工号	姓名	职称	工资			扣除			实发
			基本工资	岗位津贴	职务工资	所得税	医疗保险	失业保险	
05010	陈海华	讲师	1305	1200	50	298.35	127.5	127.5	2001.65
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

图 1.3 非规范化的表

在关系模型中,定义数据操作的方法有两种:关系代数和关系演算。

关系代数显式地定义了一些关系运算(如并、交、差、选择、投影、连接等)。关系运算的运算对象和运算结果都是关系,这使得我们可以将上一步的运算结果作为下一步的运算对象,通过运算的复合构造关系代数表达式。查询用关系代数表达式表示,而更新用关系代数表达式向关系变量的赋值表示。

关系演算并不显式定义基本运算,而是用一个逻辑公式表示查询结果必须满足的条件。关系演算又分元组关系演算和域关系演算,它们的主要区别是前者公