



丛书主编 陈松乔

高等学校计算机专业规划教材

(技术篇)

数据库技术与应用

——SQL Server 2005

主编 陆琳 罗明亮

 中南大学出版社
www.csupress.com.cn

数据库技术与应用(技术篇)

——SQL Server 2005

主 编	陆 琳	罗明亮		
副主编	刘 俐	彭铁光		
编 委	黄 胜	李 军	李 桥	
	马 翔	邹 竟	刘桂林	

中南大学出版社

图书在版编目(CIP)数据

数据库技术与应用:SQL server 2005. 技术篇/陆琳,罗明亮主编.

—长沙:中南大学出版社,2010

高等学校计算机专业规划教材

ISBN 978-7-5487-0008-1

I. 数... II. ①陆... ②罗... III. 关系数据库—数据库管理系统,SQL Server 2005—高等学校:技术学校—教材

IV. TP311.138

中国版本图书馆 CIP 数据核字(2010)第 036343 号

数据库技术与应用(技术篇)

——SQL Server 2005

主编 陆 琳 罗明亮

☐责任编辑 刘 辉

☐责任印制 文桂武

☐出版发行 中南大学出版社

社址:长沙市麓山南路

邮编:410083

发行科电话:0731-8876770

传真:0731-8710482

☐印 装 湖南大学印刷厂

☐开 本 787×1092 1/16 ☐印张 11 ☐字数 271 千字

☐版 次 2010 年 3 月第 1 版 ☐2010 年 3 月第 1 次印刷

☐书 号 ISBN 978-7-5487-0008-1

☐定 价 25.00 元

图书出现印装问题,请与出版社调换

前 言

数据库技术是现代社会各项科学技术中发展最快的技术之一，它综合应用了数学、计算机科学、管理科学等多学科知识，从基本原理、设计技术到开发应用形成了一套完整的知识体系，已成为现代信息系统的基础和核心，在当今社会国民经济各领域得到了广泛的应用。

本书以关系数据库系统为核心，按照“原理→设计→应用”循序渐进的模式，全面、系统地阐述了数据库系统的基本原理、设计技术和开发应用的主要知识。全书内容分为两大部分，分别为技术篇与应用篇：其中第1章、第2章，主要介绍关系数据库的主要基本概念与数据模型，为原理部分；第3章、第4章、第5章，讨论关系数据库的模式、保护以及开发设计等方面的原理、方法与技术，从基础理论知识渐次过渡到实用设计技术，为技术部分；第二部分是应用篇，从第6章开始，按照理论联系实际、重在实践操作的原则，以现今最流行的关系数据库管理系统——SQL Server 2005 及其使用的 Transact - SQL 语言为例，通过大量实例，全面介绍了关系数据库的程序设计基础、SQL Server 2005 的安装、配置和各种主要应用操作。

本书编者集多年的科研、开发与教学经验，深知数据库系统的学习的主要知识点、重点与难点，并以此为基础，构建了本书全面而不失系统性、翔实而不失严谨性的结构体系。本书内容丰富、重点突出，概念清楚、层次清晰，图文并茂、适合学习。每章都配有丰富的例题和习题，以及本章主要重点内容的小结，因而有助于巩固和加深读者对所学主要内容的掌握。

本书内容前后呼应，既有深入透彻的理论知识阐述，又有成熟实用的应用技术介绍，适合作为高等院校的计算机类专业、信息管理与信息系统专业的数据库原理与应用课程教材，也可作为数据库应用系统开发人员和从事信息领域工作的科技人员的技术参考书。

本书内容按照 96 课时的教学时间编写，立足全面系统，力求讲深讲透，不敢吝嗇笔墨。对于希望夯实基础但课时较少的本科专业，可以压缩前述应用篇部分的讲授课时，对于强调应用技术的专科层次的教学，可以少讲技术篇的较为深入的原理介绍。总之，可以根据自己的教学目的进行裁剪、取舍。

由于数据库技术发展迅速，加上编者水平有限、时间仓促，难免顾此失彼，对于书中存在的错误和不妥之处，敬请读者批评、指正。

编 者
2010 年 3 月

目 录

第 1 章 数据库概论	(1)
1.1 数据管理技术的发展	(1)
1.1.1 人工管理阶段	(1)
1.1.2 文件管理阶段	(2)
1.1.3 数据库阶段	(2)
1.2 数据描述	(3)
1.2.1 数据描述的 3 个领域	(3)
1.2.2 数据联系的描述	(4)
1.3 数据模型	(5)
1.3.1 数据模型的概念	(5)
1.3.2 实体联系模型与 E-R 图	(5)
1.3.3 结构数据模型	(6)
1.4 数据库体系结构	(8)
1.4.1 数据库的三级模式结构	(8)
1.4.2 数据库的二级映像技术	(10)
1.4.3 两级数据独立性	(10)
1.5 数据库管理系统	(10)
1.5.1 数据库管理系统的主要功能	(11)
1.5.2 数据库管理系统的组成	(12)
1.6 数据库系统	(12)
1.6.1 数据库系统的组成	(12)
1.6.2 数据库系统的前景和效益	(12)
第 2 章 关系模型	(14)
2.1 关系模型的基本概念	(14)
2.1.1 关系模型的基本术语	(15)
2.1.2 关系模式、关系子模式和存储模式	(18)
2.1.3 关系模型的 3 类完整性规则	(19)
2.2 关系代数	(20)
2.2.1 关系代数的基本操作	(20)
2.2.2 关系代数的组合操作	(24)
2.2.3 扩充的关系代数操作	(28)

2.3	关系演算	(29)
2.3.1	元组关系演算	(29)
2.3.2	元组关系演算语言 ALPHA	(31)
2.3.3	域关系演算	(34)
2.3.4	域关系演算语言 QBE	(36)
2.4	关系运算的安全性、等价性	(41)
2.4.1	关系运算的安全性	(41)
2.4.2	关系运算的等价性	(42)
2.5	关系查询优化	(42)
2.5.1	关系查询优化概述	(42)
2.5.2	一个基于代价分析的查询优化实例	(43)
2.5.3	查询优化的一般准则	(44)
2.5.4	关系代数等价变换规则	(45)
2.5.5	关系代数表达式的优化算法	(46)
2.5.6	优化的一般步骤	(47)
第3章	关系模式设计技术	(53)
3.1	关系模式的设计问题	(53)
3.1.1	问题的提出	(54)
3.1.2	问题分析	(55)
3.1.3	问题的解决方案	(55)
3.2	函数依赖	(55)
3.2.1	函数依赖的概念	(55)
3.2.2	函数依赖的分类	(56)
3.2.3	函数依赖的逻辑蕴涵与推理规则	(57)
3.2.4	函数依赖集的闭包与属性闭包	(57)
3.2.5	函数依赖集的覆盖和等价	(58)
3.2.6	函数依赖集的最小化	(59)
3.2.7	候选键	(61)
3.3	关系模式的范式	(62)
3.3.1	第一范式	(62)
3.3.2	第二范式	(63)
3.3.3	第三范式	(63)
3.3.4	BCNF 范式	(64)
3.3.5	多值依赖与第四范式	(64)
3.3.6	连接依赖与第五范式	(65)
3.3.7	关系模式规范化小结	(67)
3.4	关系模式的分解	(67)

3.4.1	无损联接	(68)
3.4.2	保持函数依赖的分解	(70)
3.4.3	关系模式的分解算法	(71)
第4章	数据库保护技术	(77)
4.1	数据库的安全性	(78)
4.1.1	数据库安全性概述	(78)
4.1.2	数据库安全性控制	(79)
4.2	数据完整性	(84)
4.2.1	数据完整性的概念与功能	(84)
4.2.2	数据完整性约束的分类	(84)
4.2.3	数据库完整性的定义与验证	(86)
4.2.4	主流 DBMS 的数据完整性规则	(86)
4.3	数据库的并发控制	(88)
4.3.1	并发控制与事务的概念	(88)
4.3.2	并发操作的副作用	(90)
4.3.3	隔离级别及其性能与选用	(92)
4.3.4	并发控制的方法	(94)
4.3.5	封锁的概念与类型	(96)
4.3.6	封锁协议	(99)
4.3.7	封锁出现的问题及解决方法	(102)
4.4	数据库的恢复	(104)
4.4.1	故障的类型	(104)
4.4.2	恢复的实现技术	(105)
4.4.3	恢复策略	(107)
4.4.4	具有检查点的恢复技术	(108)
4.4.5	数据库镜像简介	(110)
第5章	数据库设计技术	(114)
5.1	数据库设计概述	(115)
5.1.1	数据库与信息系统	(115)
5.1.2	数据库设计的特点	(115)
5.1.3	数据库设计方法学	(116)
5.1.4	数据库系统生存期	(117)
5.2	系统规划	(118)
5.2.1	系统规划的任务与步骤	(119)
5.2.2	可行性研究	(120)
5.2.3	项目开发计划	(122)

5.3	需求分析	(122)
5.3.1	需求分析的任务与步骤	(123)
5.3.2	需求分析的过程与方法	(124)
5.3.3	数据流图与数据字典	(126)
5.3.4	重要说明	(129)
5.4	概念设计	(130)
5.4.1	概念模型	(130)
5.4.2	概念设计的任务与步骤	(131)
5.4.3	数据抽象	(132)
5.4.4	局部概念结构设计	(135)
5.4.5	全局概念结构设计	(139)
5.5	逻辑设计	(142)
5.5.1	逻辑设计的任务与步骤	(143)
5.5.2	概念模型向关系模型的转换	(144)
5.5.3	子模式设计	(147)
5.5.4	模式优化	(148)
5.5.5	模式评价	(148)
5.5.6	模式修正	(149)
5.5.7	逻辑设计举例	(149)
5.6	物理设计	(151)
5.6.1	物理设计的任务与步骤	(151)
5.6.2	确定数据库的存储结构	(153)
5.6.3	选择关系模式的存取方法	(154)
5.7	实现	(155)
5.7.1	数据库的实现	(155)
5.7.2	其他相关工作	(157)
5.8	运行和维护	(158)
5.8.1	运行和维护阶段的主要工作	(158)
5.8.2	维护的类型与内容	(159)
5.9	数据库应用系统的设计	(159)
5.9.1	数据库系统的层次结构	(160)
5.9.2	数据库应用系统的设计方法及实例	(161)
参考文献		(168)

第 1 章 数据库概论

本章介绍数据管理技术的 3 个发展阶段,数据描述领域、术语和方法,数据模型的概念,数据库的体系结构,数据库管理系统(DBMS)的功能及组成,数据库系统的组成、前景和效益等内容。本章学习的重点内容有:数据描述术语和实体间的联系,数据模型的概念,数据库的体系结构,DBMS 的功能及组成和数据库系统的组成。本章的难点内容是:实体间的联系、概念模型的制作方法和数据库的三级结构、两级映像。

通过本章学习,应达到下述目标:

- 了解数据管理技术的 3 个发展阶段的特点。
- 领会数据描述的领域和术语,掌握数据描述的方法。
- 理解数据模型的概念。
- 领会数据库的体系结构,理解数据的两级独立性。
- 领会 DBMS 的功能和组成。
- 了解数据库系统的组成。

数据库技术作为数据管理技术,是计算机软件领域的一个重要分支,产生于 20 世纪 60 年代末,现已形成成熟的理论体系和实用技术,广泛应用于各行各业。

1.1 数据管理技术的发展

数据管理(data management)指数据的收集、整理、组织、存储、维护、检索、传送等操作。这些操作是数据处理业务的基本环节。数据管理技术的优劣直接影响数据处理的效率。用计算机进行数据管理是应用计算机技术将大量数据有组织地存储在存储介质中并进行维护、检索、重组和处理。

计算机的软、硬件发展水平,特别是存储技术的发展水平与数据管理技术的发展密切相关。数据管理技术的发展经历了 3 个阶段:人工管理阶段、文件系统阶段和数据库阶段。

1.1.1 人工管理阶段

人工管理阶段(20 世纪 50 年代中期以前)的计算机主要用于科学计算。外存只有磁带、

纸带和卡片等;软件设计语言只有汇编语言;数据处理方式是批处理。该时期数据管理的特点有:①数据不保存在机器中,数据处理时,将纸带中的程序和数据通过输入设备输入;②没有专用软件对数据进行管理,数据不具有独立性,数据组织结构依赖于程序;③只有程序(program)的概念,没有文件(file)的概念;④数据面向应用,数据与程序在一起,1组数据对应于1个程序。

1.1.2 文件管理阶段

文件管理阶段(20世纪50年代后期至60年代中期),计算机不仅用于科学计算,还用于信息管理。数据结构和数据管理软件迅速发展起来。外存已有磁盘、磁鼓等直接存取存储设备。软件领域出现了高级语言和操作系统。操作系统中的文件系统是专门管理外存的数据管理软件。处理的方式有批处理,也有联机实时处理。这一阶段数据管理的特点有:①数据可长期保存在磁盘上;②数据的逻辑结构与物理结构有了区别;③文件组织已呈现多样化,有索引文件、链接文件和散列文件等,但文件之间相互独立,缺乏联系;④数据不再属于某个特定的程序,可重复使用。

随着数据管理规模的扩大和数据量急剧增加,文件系统显露出3个缺陷:①数据冗余性(redundancy):由于文件之间缺乏联系,造成每个应用程序都有对应的文件,有可能同样的数据在多个文件中重复存储;②数据不一致性(inconsistency):往往由数据冗余造成,在更新操作时,稍有不慎,就可能使同样的数据在不同文件中不一样;③数据联系弱(poor data relationship):这是由文件之间相互独立缺乏联系造成的。

1.1.3 数据库阶段

20世纪60年代末,磁盘技术取得重要进展,具有数百兆容量和快速存取的磁盘陆续进入市场,为数据库技术的产生提供了必要的物质条件。

数据管理技术进入数据库阶段的标志是20世纪60年代末发生的3件事:①1968年美国IBM公司推出层次模型的IMS(Information Management System)系统;②1969年10月美国数据系统语言协会(CODASYL)的数据库任务组(DBTG)发表关于网状模型的DBTG报告(1971年通过);③1970年美国IBM公司的E. F. Codd连续发表论文,提出关系模型,奠定了关系数据库的理论基础。

20世纪70年代以来,数据库技术迅速发展,不断有新的产品投入运行。数据库系统克服了文件系统的缺陷,使数据管理更有效、更安全。

1. 数据库阶段数据管理的特点

①采用复杂的数据模型表示数据结构(data structure)。数据不再面向特定的某个或多个应用,而是面向整个应用系统。数据冗余明显减少,实现了数据共享。

②有较高的数据独立性(data independence)。数据库的结构分成用户的逻辑结构、整体逻辑结构和物理结构3级,使得数据库具有物理数据独立性(当数据的物理结构改变时,不影响整体逻辑结构和用户逻辑结构以及应用程序)和逻辑数据独立性(当数据整体逻辑结构改变时,不影响用户的逻辑结构以及应用程序)。

③数据库系统为用户提供方便的用户接口,用户可以使用查询语言或终端命令操作数据

库,也可以用程序方式操作数据库。

④系统提供4个方面的数据控制功能:数据库的恢复、并发控制、数据完整性和数据安全性,以保证数据库中数据安全、正确和可靠。

⑤对数据的操作不一定以记录为单位,也可以以数据项为单位,增加了系统的灵活性。

2. 数据库技术常用名词术语

①数据(data):是描述事物的符号记录。记录形式可以是文字、图形、图像、声音等。

②数据库(database, DB):是存储在存储介质上、按一定结构组织在一起、可共享的相关数据的集合。DB具有较小的数据冗余、较高的数据独立性和易扩展性。通常由两大部分组成,一部分是应用数据的集合,称为物理数据库;另一部分是关于各级数据结构的描述,称为描述数据库。

③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,为用户或应用程序提供访问DB的方法,包括DB的建立、查询、更新及各种数据控制,能够确保数据的完整性和安全性。

④数据库管理员(database administrator, DBA):是控制数据整体结构的技术人员,负责DB的正常运行。DBA必须熟悉企业全部数据的性质和用途,充分了解用户需求,熟悉系统性能等方面的素质。DBA的主要职责是定义和修改数据库模式、管理数据库用户和用户权限、完成DB的日常管理(备份、恢复、性能监测、性能优化等)。

⑤数据库系统(database system, DBS):是实现有组织地、动态地存储大量关联数据,方便多用户访问的计算机软件、硬件和数据资源组成的系统,即采用了数据库技术的计算机系统。由DB、DBMS、应用系统(包括数据库应用软件、相关系统软件和硬件)、DBA和用户(user)组成。

⑥数据库技术(database technology):是研究数据库的结构、存储、管理和使用的软件学科。数据库技术是在操作系统(OS)的文件系统基础上发展起来的,而DBMS本身要在OS的支持下才能工作。数据库技术涉及到数据结构、集合论、数理逻辑中的知识,是一门综合性较强的学科。

⑦数据库应用软件(database applications software):是采用数据库技术、专门为某一应用目的开发的应用软件。通常使用某种程序设计语言(如C#、Visual FoxPro)及其支持的DBMS(如SQL Server、Visual FoxPro),为完成数据库系统最终用户的某些特定用途而开发。

1.2 数据描述

1.2.1 数据描述的3个领域

从事物特征到计算机中数据表示,数据描述要经历现实世界、信息世界和机器世界3个领域。

①现实世界(real world)指存在于人们头脑之外的客观世界。它是原始数据(记录)的来源。

②信息世界(information world)指现实世界在人们头脑中的反映。人们用文字、符号、图

形、图像、声音等方式记载下来现实世界的信息,称为信息世界。信息世界中常用的名词有:

实体(entity):客观存在可以相互区别的事物称为实体。实体可以是具体的对象,例如一个员工、一本教材等;也可以是抽象的事件,例如一次旅游、一场球赛等。

实体集(entity set):性质相同的同类实体的集合称为实体集,例如学生、公司员工等。

属性(attribute):实体有很多特性,每一个特性称为一个属性。属性具有数据类型和值域。例如:学生有学号、姓名、年龄等属性。学号、姓名的数据类型是字符串,年龄的数据类型是整数。

实体标识符(identifier):能唯一标识每个实体的属性或属性集的称为实体标识符,有时简称为“键”。例如,学生的学号可以作为学生实体标识符

③**机器世界(machine world):**信息世界的信息在机器中以数据形式存储,成为机器世界。机器世界中常用的名词有:

字段(field):标记实体属性的命名单位称为字段或数据项。它是可以命名的最小信息单位。字段的命名往往与属性名相同。例如:一个学生记录中有学号、姓名、年龄、性别等字段。

记录(record):有限个字段的集合称为记录。一般用一个记录描述一个实体,所以记录又可定义为能完整地描述一个实体的字段集。例如一个学生记录由字段集(学号,姓名,年龄)组成。

基本表(basic table, 常简称为“表”):描述一个实体集的所有记录的集合。有的 DBMS, 一个基本表就对应一个数据文件(file)。

候选键(key, 简称为“键”):能唯一标识基本表中每个记录的字段或字段集。

机器世界与信息世界术语的对应关系如表 1-1 所示。

数据描述有 2 种形式:物理描述和逻辑描述。物理描述是具体的、实际的描述,如物理设备、物理路径、物理结构。逻辑描述是抽象的、形式的描述,如逻辑设备、逻辑文件、逻辑结构。

表 1-1 机器世界与信息世界术语对应关系

信息世界	机器世界
实体	记录
属性	字段
实体集	基本表(文件)
实体标识符	候选键

1.2.2 数据联系的描述

现实世界中的事物是相互联系的,这种联系要在信息世界中体现出来。实体内部的联系体现在同一记录各字段之间的联系,实体间的联系体现在各记录之间的联系。

实体间的联系分为 2 种:一种是同一实体集中各实体间的联系,另一种是不同实体集的各个实体之间的联系。数据库中一般研究后一种。

不同实体的实体间联系有以下 3 种情况。

①**一对一联系:**如果实体集 E1 中每个实体至多和实体集 E2 中 1 个实体有联系,反之亦然,那么实体集 E1 对 E2 的联系称为“一对一联系”,记为“1:1”。

【例 1-1】 考场座位与考生之间的关系:1 个座位对应 1 个考生,1 个考生对应 1 个

座位。

②一对多联系：如果实体集 E1 中每个实体与实体集 E2 中任意个(零或多个)实体有联系，而 E2 中每个实体至多和 E1 中一个实体有联系，则称 E1 对 E2 是“一对多联系”，记为“1:N”。

【例 1-2】 班级与学生之间的关系，1 个班级有多个学生，每个学生只属于 1 个班级。

③多对多联系：如果实体集 E1 中每个实体与实体集 E2 中任意个(零个或多个)实体有联系，反之亦然，那么称 E1 和 E2 的联系是“多对多联系”，记为“M:N”。

【例 1-3】 学生与课程之间的关系，一个学生要选修多门课程，一门课程有多个学生选修。

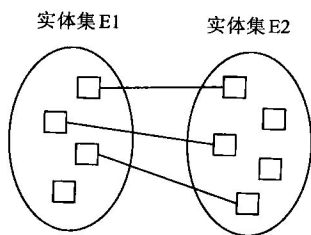


图 1-1 一对一联系

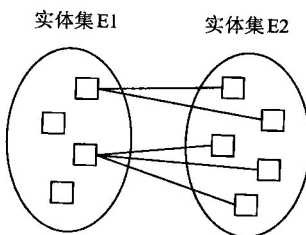


图 1-2 一对多联系

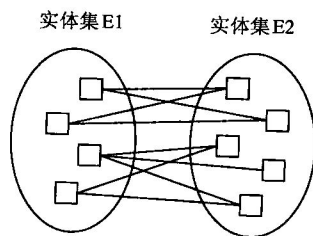


图 1-3 多对多联系

以上 3 种联系是实体之间最基本的联系。多个实体之间的联系可转化为两两之间的联系。

1.3 数据模型

1.3.1 数据模型的概念

模型是对现实世界的抽象。数据库中用模型描述 DB 的结构与语义，对现实世界进行抽象。表示实体类型及实体间联系的模型称为“数据模型”(data model)，常用的有两种：

①概念数据模型：独立于计算机系统，不涉及信息在计算机系统中的表示，只描述人们所关心的信息结构；是现实世界的第一层抽象，是用户和数据库设计人员之间进行交流的工具，用于建立信息世界的数据库模型，强调其语义表达功能，应该概念简单、清晰，易于用户理解。这类中著名的模型是“实体联系模型”。

②结构数据模型：直接面向数据库逻辑结构，是现实世界的第二层抽象。这类模型涉及到计算机系统和 DBMS，例如层次模型、网状模型、关系模型面向对象模型等。它们有严格的形式化定义，以便于在计算机系统中实现。其中的面向对象模型(object-oriented model)是在数据库技术中引入面向对象技术所产生，具有丰富的表达能力，但模型相对比较复杂。

1.3.2 实体联系模型与 E-R 图

实体联系模型(entity relationship model, E-R 模型)由 p. p. Chen 于 1976 年提出，是从现

实世界中抽象出实体类型及实体间联系、用 E-R 图表示的数据模型。设计 E-R 图的方法称 E-R 方法。

E-R 图有 4 个基本成分: ①矩形框, 表示实体类型(问题的对象); ②菱形框, 表示联系类型(实体间的联系); ③椭圆形框, 表示实体类型和联系类型的属性。关键码的属性, 在属性名下划一横线; ④直线, 联系类型与其涉及的实体类型之间用直线连接, 并在直线端部标上联系的种类(1:1, 1:N, M:N)。相应的命名均记入各种框中。

可以按下述步骤制作 E-R 图: ①确定实体类型; ②确定联系类型; ③把实体类型和联系类型组合成 E-R 图; ④确定实体类型和联系类型的属性, 并标上联系种类; ⑤确定实体类型的键, 在键的属性名下划一横线(详见本书第 5 章)。

【例 1-4】 学生与课程的联系

学生和课程两个实体的联系是修读。学生的属性有学号、姓名、性别、年龄; 课程的属性有课程号、课程名称; 修读的属性有成绩。学生与课程 M:N 联系。学生的键是学号, 课程的键是课程号。其 E-R 图见图 1-4。

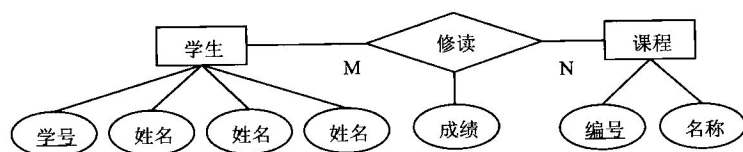


图 1-4 E-R 图示意图

1.3.3 结构数据模型

结构数据模型由数据结构(指对实体类型和实体间联系的表达和实现)、数据操作[指对数据库的检索(查询)和更新(插入、删除、修改)操作的实现]、数据完整性约束(是给出数据及其联系应具有制约和依赖规则)3 部分组成。

常用的结构数据模型有层次、网状、关系模型 3 种。

1. 层次模型(hierarchical model)

层次模型用树型结构表示实体类型及实体间联系。树的结点是记录类型; 有且只有 1 个根结点; 每个非根结点有且只有 1 个父结点, 可有多个子结点; 父结点与子结点的联系都是 1:N 联系。

层次模型的优点是记录之间的联系通过指针实现, 查询效率较高; 缺点是只能表示 1:N 的联系, 查询、更新、编程复杂。

【例 1-5】 层次模型: 某大学下设多个学院, 每个学院下设多个系, 共有 3 个层次。某大学是根结点, 是各学院的父结点; 各学院子结点是它所属的系; 系是叶子结点, 其父结点是它所属的专业。某大学、学院、系之间的联系分别都是 1:N 联系, 构成层次模型(图 1-5)。

1968 年 IBM 公司推出的 IMS(Information Management System)是典型的层次模型系统, 20 世纪 70 年代在商业上得到广泛应用。

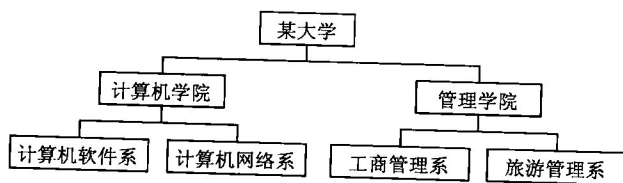


图 1-5 层次模型示意图

2. 网状模型(network model)

网状模型用有向图结构表示实体类型及实体间联系。图中的顶点是记录类型，箭头(有向边，称为弧)表示从箭尾(弧尾)一端的记录类型到箭头(弧头)一端的记录类型间联系是1:N。图中每个顶点都可以与其他多个顶点存在联系，因此网状模型的记录类型(顶点)之间构成M:N联系，可以理解为每个顶点都可有多个父结点和多个子结点。

网状模型优点是记录之间的联系通过指针实现，M:N联系也容易实现(每个M:N联系可拆成两个1:N联系)，查询效率较高；缺点是结构复杂，编程难度大，不易掌握和使用。

【例 1-6】 网状模型：教师、学生与课程

该模型有教师、学生和课程有三个顶点(实体)。每个教师可授多门课程，每门课程可有多个教师讲授；每个学生可修多门课程，每门课程可有多个学生修读；每个教师可辅导多个学生，每个学生可有多个教师辅导；教师、学生、课程之间的联系由6条弧构成M:N联系，属于网状模型(图1-6)。

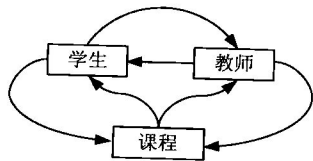


图 1-6 网状模型示意图

网状数据模型的典型代表是 DBTG 系统(也称 CODASYL 系统)，它是 20 世纪 70 年代数据系统语言研究会 CODASYL(Conference on Data Systems Language)下属的数据库任务组(Data Base Task Group, DBTG)提出的一个系统方案。

3. 关系模型(relational model)

关系模型的主要特征是用二维表格结构表达实体集，用外键表示实体间联系。一个关系模型是由若干个关系模式组成的集合。一个关系模式就是一个二维表，它的实例称为关系，每个关系实际上是一张二维表。

关系模型的优点是有严格的数学理论基础，概念简单，容易理解；缺点是查询效率低。

【例 1-7】 关系模型：学生、课程和成绩的联系

学生模式：Student(学号，姓名，性别，年龄)

课程模式：Course(课程号，课程名称)

选修模式：Grade(学号，课程号，成绩)

上述 3 个关系模式的实例关系分别参见表 1-1、表 1-2、表 1-3。

表 1-2 学生关系

学号	姓名	性别	年龄
0001	张三	男	19
0002	李四	女	18
0003	王五	男	20

表 1-3 课程关系

课程号	课程名称
01	数据库原理与应用
02	计算机基础
03	大学物理

表 1-4 选修关系

学号	课程号	成绩
0001	01	80
0001	02	75
0001	03	86
0002	01	90
0002	02	92
0002	03	88
0003	01	81
0003	02	77
0003	03	96

1.4 数据库体系结构

数据库的体系结构可以概括为三级模式和两级映像,如图 1-7 所示。

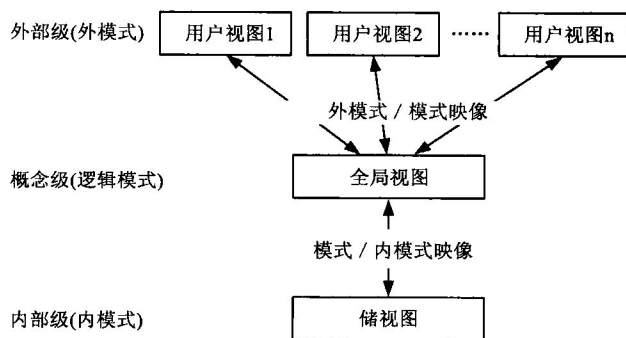


图 1-7 数据库体系结构

在此,我们从某个角度看到的数据特征称为“数据视图”(data view)。

1.4.1 数据库的三级模式结构

DB 的模式结构可分为 3 级,称为“三级模式结构”,即:内部级(internal)的“内模式”、概念级(conceptual)的“逻辑模式”、外部级(external)的“外模式”。

1. 逻辑模式与概念数据库

逻辑模式(logical schema)常简称“模式”(schema),是对 DB 中数据的整体逻辑结构和特征的描述,是涉及所有用户的数据定义,即全局的数据视图。逻辑模式使用模式 DDL(data

definition language)进行定义,其定义的内容包括对DB的记录和数据项的逻辑类型、记录间的联系等的描述,以及对数据的安全性要求(保密方式、保密级别和数据使用权)、完整性条件和数据寻址方式的说明。逻辑模式不涉及到存储结构、访问技术等细节。

逻辑模式是为减小数据冗余、实现数据共享并对所有用户的数据进行综合抽象而得到的统一的全局数据视图。一个DBS只能有一个逻辑模式,以逻辑模式为框架的DB称为概念数据库。

2. 外模式与用户数据库

外模式(external schema)也称“子模式”(sub schema)或“用户模式”,它是对各个用户或程序所涉及到的数据的逻辑结构和数据特征的描述,是单个用户所看到的数据特性,即用户数据视图。外模式使用子模式DDL(sub schema DDL)进行定义,该定义主要涉及对子模式的数据结构、数据域、数据构造规则及数据的安全性和完整性等属性的描述。子模式可以在数据组成(数据项的个数及内容)、数据间的联系、数据项的型(数据类型和数据宽度)、数据名称上与逻辑模式不同,也可以在数据的安全性和完整性方面与逻辑模式不同。

子模式是完全按用户自己对数据的需要、站在局部的角度进行设计的。由于一个DBS有多个用户,所以就可能有多个子模式。从逻辑关系看,子模式是模式的一个子集,从一个模式可以导出多个不同的子模式。使用子模式具有下列优点:

①由于使用子模式,用户不必考虑那些与自己无关的数据,也无需了解数据的存储结构,简化了用户使用数据的工作和程序设计的工作。

②由于用户使用的是子模式,使得用户只能对自己需要的数据进行操作,DB的其他数据与用户是隔离的,这样有利于数据的安全和保密。

③由于用户可以使用子模式,而同一模式又可派生出多个子模式,所以有利于数据的独立性和共享性。

以子模式为框架的DB称为用户数据库。每个用户数据库都是概念数据库的部分抽取。

3. 内模式与物理数据库

内模式(Internal schema)也叫存储模式(Access Schema)或物理模式(Physical Schema),它最接近于物理存储设备,是对数据的内部表示或底层描述,涉及实际数据存储的结构。内模式使用内模式DDL(Internal schema DDL)定义,不仅要定义所有的内部记录类型、索引和文件的组织方式、存取路径在内的一切物理组织方式等属性,还要规定数据存储、数据访问和数据控制方面的细节,例如数据的优化性能、响应时间和存储空间需求、数据的记录位置、块的大小与数据溢出区等。

以物理模式为框架的DB为物理数据库。在DBS中,只有物理数据库才是真正存在的,它是存放在外存的实际数据文件;而概念数据库和用户数据库在计算机外存上是不存在的。

用户数据库、概念数据库和物理数据库三者的关系是:概念数据库是物理数据库的逻辑抽象形式;物理数据库是概念数据库的具体实现;用户数据库是概念数据库的子集,也是物理数据库子集的逻辑描述。

三级模式结构之间往往差别很大,为了实现这3个抽象级别的联系和转换,DBMS在三级模式结构之间提供2个层次的映像(mappings):外模式/模式映像;模式/内模式映像。