

第 5 章 数据库技术基础

数据库技术是信息管理的重要技术,是一种利用计算机辅助管理数据的方法,它研究的是如何组织和存储数据,高效地获取和处理数据。

5.1 数据库知识

5.1.1 数据库应用及发展

1. 数据库的常见应用

(1) 校园一卡通。主要实现校园卡的金融消费功能和身份识别功能,实现“一卡在手,走遍校园”的功能,应用于学校教学、借书、医疗、管理、生活等各个方面。校园一卡通的网络拓扑结构如图 5-1 所示。

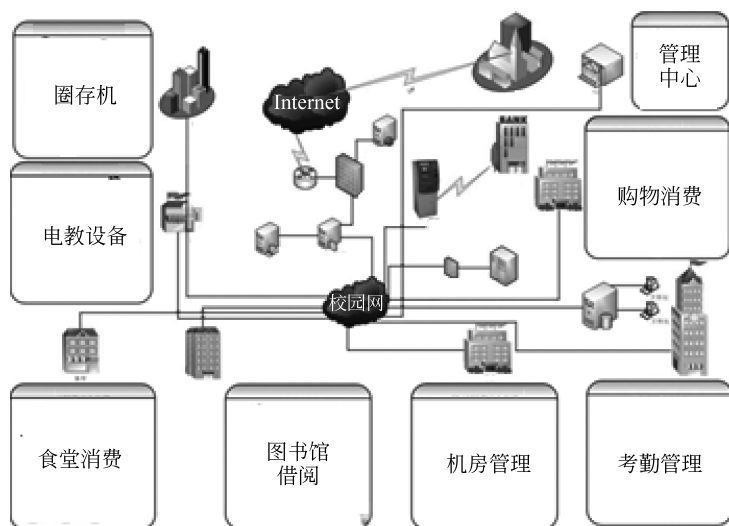


图 5-1 校园一卡通网络拓扑结构

(2) 企业进销存。企业进销存是对企业生产经营中采购、采购退货、采购付款、销售、销售退货、销售付款、退货情况、盘库、仓库调拨、借入、借出、借入还出、借出还入、供方客户资料、供方供货汇总、明细报表,采购付款汇总、明细报表进行多仓库、多币种、多结算方式、多报表的管理,进销存可以摆脱繁杂的数字统计等琐事。如图 5-2 为进销存业务流程示意图。

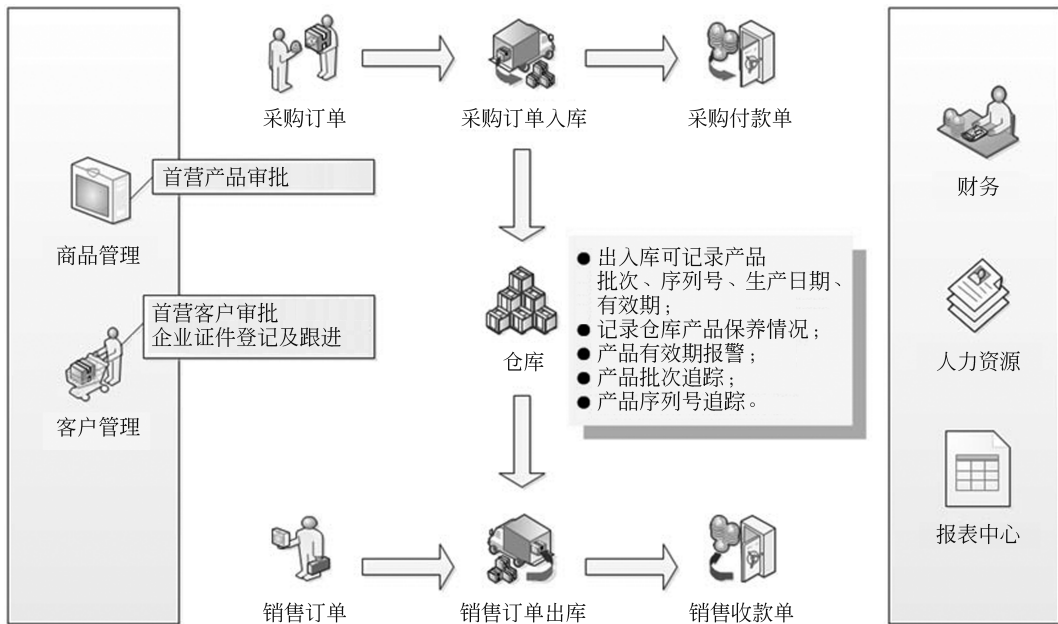


图 5-2 进销存业务流程示意图

校园一卡通及企业进销存的实现,离不开数据库技术的支持,各种信息以数字化的形式存储在数据库中,通过数据库达到数据共享的目的。

2. 数据库的历史

数据库及其技术出现在 20 世纪 60 年代后期,当时已经出现了大容量且价格低廉的磁盘且操作系统已成熟,数据处理的规模越来越大。为了解决数据独立和共享问题,实现数据的统一管理,数据库技术应运而生。数据库系统中应用程序与数据的关系如图 5-3 所示。

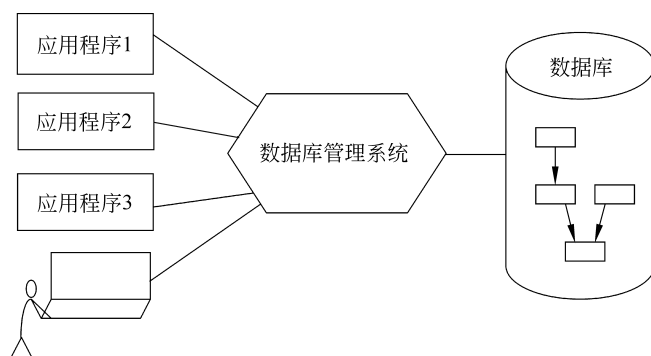


图 5-3 数据库与应用程序关系示意图

数据库技术满足了集中存储大量数据以方便众多用户使用的需求。具有如下特点。

(1) 采用一定数据模型。以数据为中心按一定的方式存储,即按一定的数据模型组织数据,最大限度地减少数据的冗余。

(2) 最低的冗余度。数据冗余是指在数据库中数据的重复存放。数据冗余不仅浪费

了大量的存储空间,而且还会影响数据的正确性。数据冗余是不可避免的,但是数据库可以最大限度地减少数据的冗余,确保最低的冗余度。

(3) 有较高的数据独立性。处理数据时用户所面对的是简单的逻辑结构,而不涉及具体的物理存储结构,数据的存储和使用数据的程序彼此独立,数据存储结构的变化尽量不影响用户程序的使用,使得应用程序保持不变。

(4) 安全性。并不是每一个用户都能允许访问全部数据,通过设置用户的使用权限可以防止数据的非法使用和丢失,在数据库被破坏时,可以把数据库恢复到可用状态。

(5) 完整性。系统采用一些完整性检验以确保数据符合某些规则,保证数据库中数据始终是正确的。

在数据库出现之前,人们对数据的管理依次经历了人工管理与文件系统管理两个阶段。

(1) 人工管理阶段处理数据的方式,是把程序和数据放在一起。如图 5-4 所示,Visual Basic 程序对 10 个数据从大到小排序,10 个数据是嵌入程序当中的,虽然是处理同批数据,但是程序之间没有共享数据。在 20 世纪 50 年代中期以前是人工管理阶段,这个阶段程序与数据的关系如图 5-5 所示。

```
Dim i%, j%, t%, a()  
a=Array(23,45,12,86,19,99,80,39,82,63)  
For i=1 To 10  
    Input #1, a(i)  
Next i  
For i=1 To 10-1  
    For j=i+1 To 10  
        If a(i)<a(j) Then  
            t=a(i): a(i)=a(j): a(j)=t  
        End If  
    Next j  
Next i
```

图 5-4 人工管理阶段程序与数据的关系

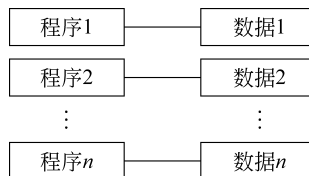


图 5-5 程序与数据在一起

(2) 文件系统阶段。20 世纪 50 年代中期至 60 年代中期,由于计算机大容量存储设备(如硬盘)的出现,推动了软件技术的发展,而操作系统的出现标志着数据管理步入一个新的阶段。在这个阶段,以数据文件为单位存储在外存,且由操作系统统一管理。操作系统为用户使用文件提供了友好界面。这一阶段数据管理的主要特征如下。

① 数据可以长期保存。大量用于数据处理的数据可以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 程序与数据有了一定的独立性。程序与数据分开,有了程序文件与数据文件的区别,如图 5-6 所示的 Visual Basic 程序对 10 个数据从大到小排序,10 个数据是从文件中读取。

③ 用户的程序与数据可分别存放在外存储器上,各应用程序可以共享一组数据,实现了以文件为单位的数据共享。

文件系统阶段程序与数据的关系如图 5-7 所示。

```

Dim i%, j%, t%, a(10)
Open "d:\wenjian\wj0011.txt" For Input As #1
For i=1 To 10
    Input #1, a(i)
Next i
For i=1 To 10-1
    For j=i+1 To 10
        If a(i)<a(j) Then
            t=a(i):a(i)=a(j):a(j)=t
        End If
    Next j
Next i

```

图 5-6 文件管理阶段程序与数据的关系

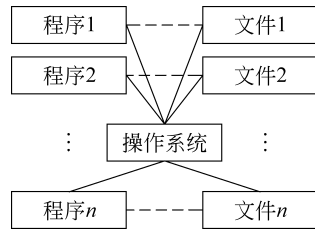


图 5-7 程序与数据分开

3. 数据库系统新技术

从 20 世纪 60 年代美国 IBM 公司开发的第一个数据库系统开始,数据库技术经过了几十年的发展,已成为一个数据模型丰富、新技术内容层出不穷、应用领域日益广泛的体系,是计算机科学技术中发展最快、应用最广泛的重要分支之一,也是计算机信息系统和计算机应用系统重要的技术基础和支柱。

20 世纪 80 年代,数据库技术成功应用于商业领域。这一时期出现了一系列重大的社会进步和科技进步,极大地刺激并影响了数据库的设计和发展。尤其是进入 20 世纪 90 年代,数据库的发展更加令人眼花缭乱,不断推陈出新,形成了一个庞大的数据库家族。

(1) 面向对象数据库管理系统。面向对象数据库管理系统将数据的特性和面向对象的特性结合起来。数据的特性包括数据的完整性、安全性、持久性、事务管理、并发控制、备份、恢复、数据操作和系统调节,面向对象的特性包括继承性、封装性和多态性等。

(2) 并行数据库管理系统。并行数据库管理系统是数据库技术与并行技术相结合的结果,其目的是提供一种高性能、高可用性、高扩展性的数据库管理系统。这种数据库可以充分发挥多处理机结构的优势,将数据分布存储在多个磁盘上,并且利用多个处理机对磁盘数据进行并行处理,从而解决了磁盘 I/O 瓶颈问题。如图 5-8 所示为共享内存并行数据库管理系统构成。并行数据库管理系统的基本思想在于通过先进的并行查询技术,开发查询间并行、查询内并行以及操作内并行来提高性能和查询的效率。

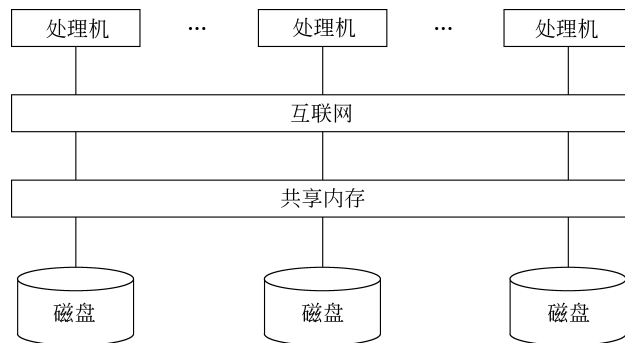


图 5-8 并行数据库管理系统构成

(3) 分布式数据库管理系统。分布式数据库管理系统用于对网络互相连接的计算机系统上的逻辑相关数据进行存储和处理。分布式数据库管理系统是由多个结点的数据库组成的,也就是说,一个单独的数据库被分成多个分片,这些分片存储于某个网络中的不同的计算机上,数据处理也分散到多个不同的网络结点上。分布式数据库系统是一种多重来源、多重位置的数据库,核心是多结点数据库。图 5-9 为分布式数据库管理系统构成。

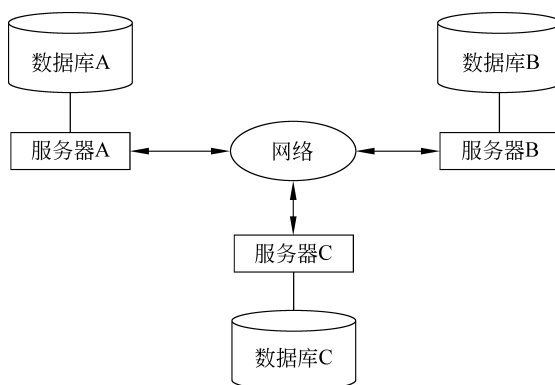


图 5-9 分布式数据库管理系统构成

(4) 空间数据库管理系统。空间数据库管理系统是以描述空间位置和点、线、面、体积特征的空间数据,以及描述这些空间对象特征属性的数据作为处理对象的数据库,其数据模型和查询语言能支持空间数据类型和空间索引,并提供空间查询和其他空间分析方法。空间数据库管理系统利用数据库技术实现空间数据的有效存储、管理和检索,为各种空间数据库的用户服务。

(5) 非关系型数据库。随着互联网网站的兴起,传统的关系数据库在应付超大规模、高并发及海量数据的高效率存储和访问的纯动态网络服务已经显得力不从心,暴露了很多难以克服的问题,而非关系型的数据库则由于其本身的特点得到了非常迅速的发展。

(6) 移动数据库管理系统。装配无线联网设备的移动计算机能够与固定网络甚至其他移动计算机相连,用户不再需要固定地连接在某一个网络中不变,而是可以携带移动计算机自由地移动,这样的环境称为移动计算。移动数据库管理系统是指在移动计算环境中的分布式数据库。可以从两个层面上来理解移动数据库:一个是用户在移动时可以存取后台数据库数据或其副本,另一个是用户可以带着后台数据的副本移动。图 5-10 为移动数据库管理系统构成。

5.1.2 数据知识

1. 数据

数据是事物特性的反映和描述,是数字、文字、声音、图形、图像等符号的集合。数据

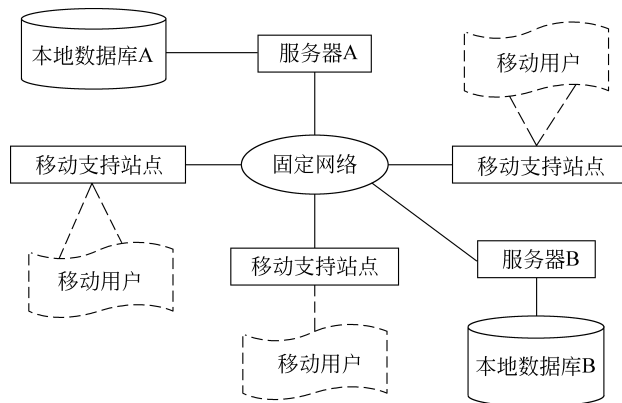


图 5-10 移动数据库管理系统构成

在空间上的传递称为通信(以信号方式传输),数据在时间上传递称为存储(以文件形式存取)。

2. 信息

信息是向人们或机器提供关于现实世界新的事实的知识,是数据中所包含的意义,是和数据关系密切的另一个概念,数据是信息的符号表示(或称为载体),信息则是数据的内涵,是对数据语义的解释。

3. 知识

知识是可用于指导实践的信息。知识不是数据和信息的简单积累,是人们在改造世界的实践中所获得的认识和经验的总和。知识又分为显性知识和隐性知识。显性知识是已经或可以文本化的知识,并易于传播。隐性知识是存在于个人头脑中的经验或知识,需要进行大量的分析、总结和展现,才能转化成显性知识。

4. 数据处理

数据处理是对数据的采集、存储、检索、加工、变换和传输。数据处理的基本目的是从大量的、可能是杂乱无章的、难以理解的数据中抽取并推导出对于某些特定的人们来说是有价值、有意义的数据。数据处理是系统工程和自动控制的基本环节。数据处理贯穿于社会生产和社会生活的各个领域。数据处理技术的发展及应用的广度和深度,极大地影响着人类社会发展的进程。

数据处理的基本过程包括数据收集、数据整理、数据描述、数据分析 4 个阶段。

5.1.3 数据库概念

1. E-R 图

E-R 图(entity-relationship approach)即实体-联系图,是数据关系分析的一种表示方法。E-R 图用矩形表示实体,矩形框内写明实体名;用椭圆形表示属性,并用无向边将其与相应的实体连接起来;用菱形表示联系,菱形框内写明联系名,如图 5-11~图 5-13 所示。E-R 图也可以说是现实世界到信息世界的一种抽象,抽象过程如图 5-14 所示。

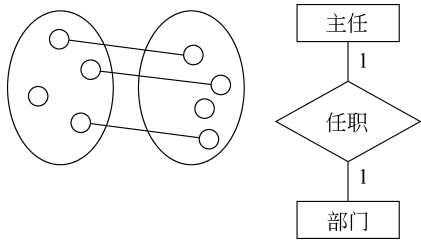


图 5-11 一对一联系

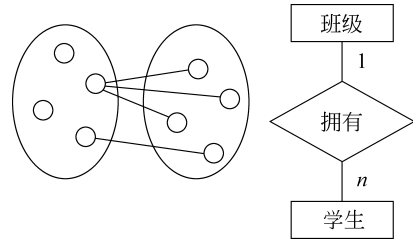


图 5-12 一对多联系

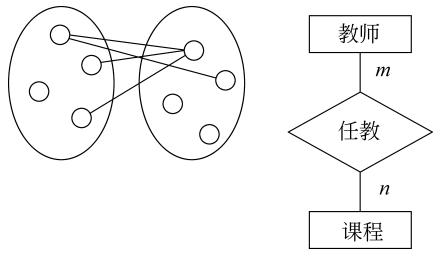


图 5-13 多对多联系

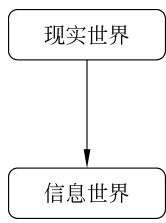


图 5-14 抽象过程

例如,学生实体具有学号、姓名、性别、年龄、学院等属性,用 E-R 图表示如图 5-15 所示。再如,用“供应量”来描述联系“供应”的属性,如图 5-16 所示用 E-R 图表示某供应商供应了多少数量的零件给某个项目。

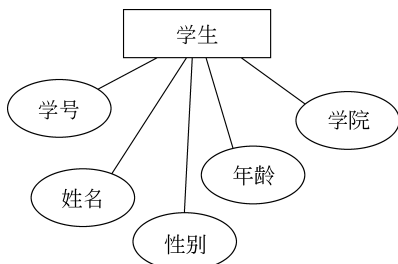


图 5-15 学生实体及属性

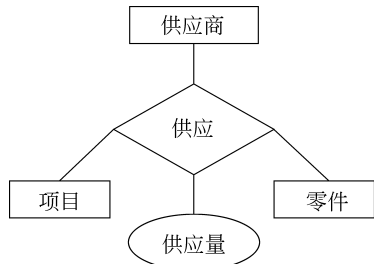


图 5-16 联系的属性

2. 数据库

数据库(database)是依照某种数据模型组织起来并以文件的形式存放在存储器中的数据集合。在生活中,手机的使用越来越普遍,手机里常存有很多亲戚和朋友的姓名和电话号码,用来方便地保持与他们的联系,这些姓名及电话号码通常记录在手机里的一个通讯簿(电话、联系人)中,当与亲戚和朋友打电话时,查找就很方便。这个通讯簿(电话、联系人)就是一个简单的数据库。这个数据库除了查找方便外,还可以根据需要随时在其中添加新朋友的姓名和电话号码数据,也可以修改和删除某个人的数据。

3. 数据库的数据模型

数据库的数据模型主要分 3 类:层次数据模型、网状数据模型、关系数据模型。

(1) 层次数据模型。层次数据模型是最早使用的一种数据模型,采用层次模型数据通过链接方式,将相互关联的记录组织起来,形成一种层次关系,构成树状结构。

这种树状结构就像一棵倒挂的树,把每条记录看成树上的一个结点,则最上一层的结

点类似于树根,称为根结点,结构中的每一结点都可以链接一个或多个结点,这些结点称后继结点或子结点。与子结点相链接的上一层结点称该子结点的前驱结点或亲结点。链接则表示结点之间的联系。如图 5-17 所示为某学校行政管理的数据模型。

(2) 网状数据模型。网状数据模型也是早期经常使用的一种数据模型,网状数据模型采用网状结构表示实体与实体之间的联系,网状数据库则是使用网络模型作为自己的存储结构。

在这种结构中,各数据记录便组成网络中的结点,有联系的各结点通过链接方式链接在一起,构成一个网状结构,这种结构的结点之间的联系较为复杂。某学院教学管理的数据模型如图 5-18 所示。

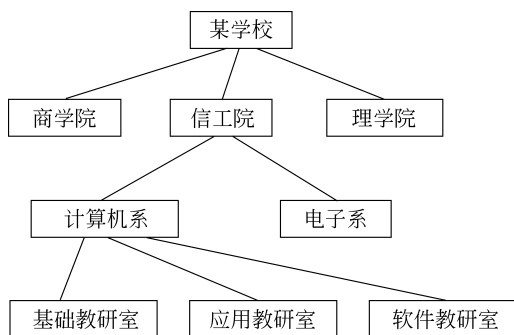


图 5-17 某学校行政管理的数据模型

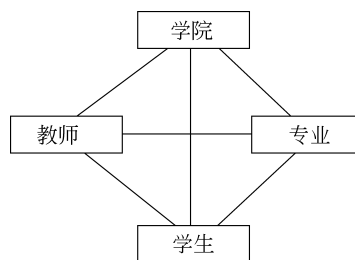


图 5-18 某学院教学管理的数据模型

(3) 关系数据模型。关系数据模型简称关系模型,是目前使用最广泛的一种数据模型,关系模型以其概念简单清晰、操作直观方便、易学易用等优势,受到了众多用户的青睐。现在的数据库产品 90%以上都是以关系模型为基础的。

关系模型采用关系作为逻辑结构,实际上关系就是一张二维表,一般简称表。一张二维表都是由行和列构成,每行称为一条记录,每列称为一个字段,如表 5-1 所示。

表 5-1 教工登记表

工作证编号	姓名	性别	年龄	职称	基本工资	部门
22001	江海	男	30	讲师	650.00	信息系
22002	张大山	男	52	副教授	750.00	计算机系
22003	王兰英	女	45	副教授	750.00	电子系
22004	张柳	女	38	讲师	650.00	通信系
22005	王坤	男	55	教授	850.00	外语系
22006	李天洋	男	28	助教	500.00	历史系

4. 数据库的数据组织

数据库中的数据组织分 4 个级别: 字段、记录、表和数据库文件,它们之间的关系如图 5-19 所示。

(1) 字段。字段是定义数据库数据的最小单位。字段与现实世界实体的属性对应,每个字段都有一个名称,被称为字段名。字段的值可以是数值、字母、字母数字、汉字等形

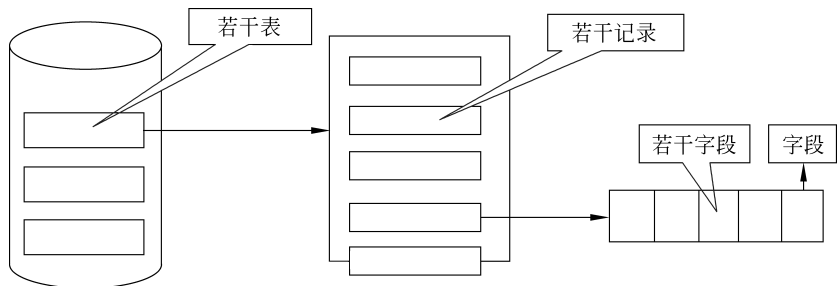


图 5-19 数据库数据组织透视图

式。字段的取值有一定的范围,被称为域,域以外的任何值对该字段都是无意义的。

(2) 记录。记录由若干相关联的字段组成,是处理和存储信息的基本单位,是关于一个实体的数据总和。构成记录的字段表示实体的若干属性。为了标识每个记录的唯一性,就必须有记录标识符(也称为键)。能标识记录唯一性的键称为候选键,确定作为记录的唯一标识的某一候选键称为主键。

(3) 表。表是一个给定类型的记录的全部具体值的集合,用表名称来标识。由于表可以看成具有相同性质的记录的集合,因而具有以下特性:表的记录格式相同,长度相等;不同的行是不同的记录,因而具有不同的内容;不同的列则代表不同的字段,相同一列中的数据属性(性质)相同;每一行各列的内容是不能分割的,行的顺序和列的顺序不影响表内容的表达。

(4) 数据库文件。数据库文件是比表更大的数据组织形式,是具有特定关系的表的集合。

5. 数据库的体系结构

数据库的体系结构分 3 个层:数据库的视图结构、数据库的逻辑结构、数据库物理结构。数据库的视图结构是用户的数据视图,最接近用户,称为外模式;数据库的逻辑结构介于物理结构和视图结构两者之间,称为模式;数据库的物理结构是数据的物理存储方式,称为内模式,如图 5-20 所示。

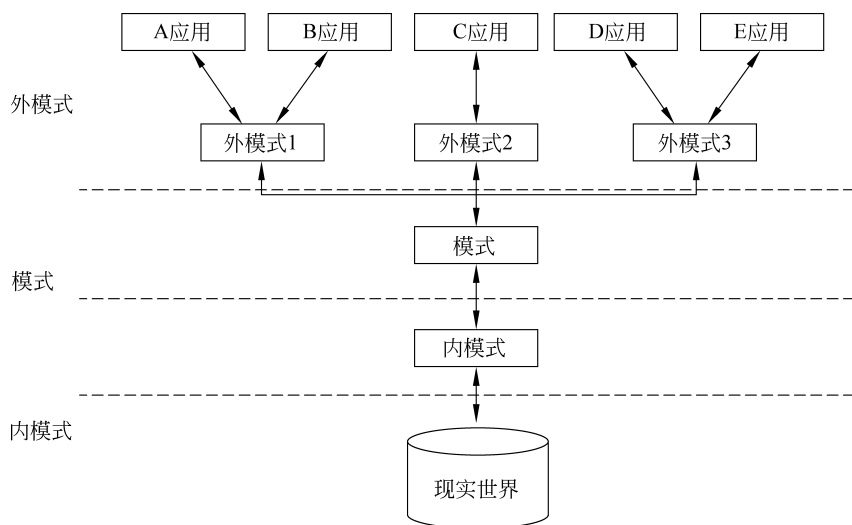


图 5-20 数据库三层体系结构

(1) 外模式对应于不同的用户,用户的应用目的不同、使用权限不同,对应的外模式的定义就不同,每个用户只能使用自己权限范围内的外模式的数据,而无法涉及其他用户定义的外模式数据。

(2) 模式表示了数据库的全部信息内容,定义了数据库的全部数据的逻辑结构,主要描述数据库中存储什么数据以及这些数据之间有何种关系。

(3) 内模式用于定义数据的存储方式和物理结构,处在最底层。

5.1.4 数据库管理系统

数据库管理系统(database management system,DBMS)是一种操纵和管理数据库的大型软件,用于建立、使用和维护数据库。它对数据库进行统一的管理和控制,以保证数据库的安全性和完整性,如图 5-21 所示。

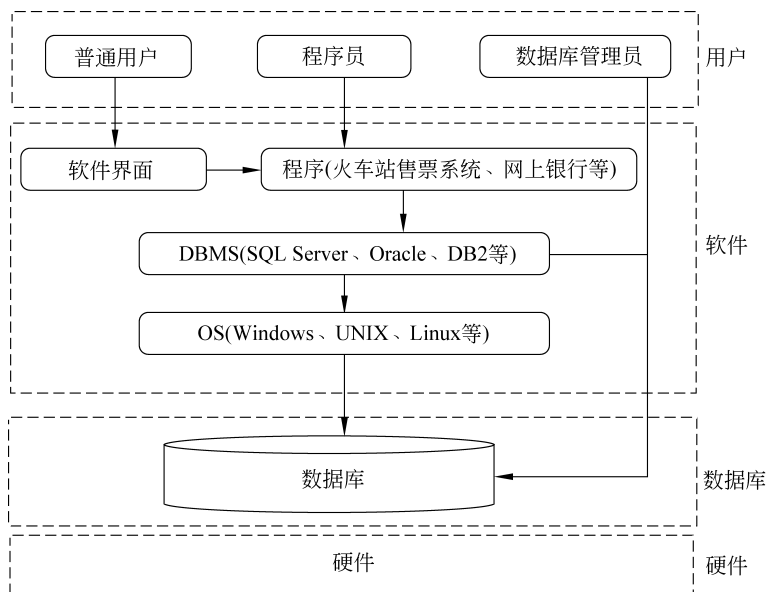


图 5-21 数据库管理系统的构成

5.2 关系数据库

5.2.1 关系数据库概念

本节主要讲解关系数据库的相关概念。

1. 关系

在关系数据库中,基本数据结构被限制为二维表格,数据在用户观点下的逻辑结构就是一张二维表,每张二维表称为一个关系(relation)。关系数据模型中最基本的概念是关系,表 5-2 给出的工资表就是一个关系。