

5.1 数据管理

5.1.1 信息与数据

对每个人来说,“信息”和“数据”都非常重要。“信息”可以告知有用的事实和知识,“数据”可以更有效地表示、存储和抽取信息。

1. 信息、信息特征及作用

在日常生活中,经常可以听到“信息”(information)这个名词。什么是信息呢?简单地说,信息就是新的、有用的事实和知识,泛指人类传播的一切内容。创建宇宙万物的最基本单位是信息。信息是对客观世界中各种事物的运动状态和变化的反映,是客观事物之间相互联系和相互作用的表征,表现的是客观事物运动状态和变化的实质内容。

根据信息的概念,可以归纳出信息具有以下基本特征:

(1) 信息的内容是关于客观事物或思想方面的知识。信息的内容能反映已存在的客观事实,能预测未发生事物的状态和能用于指挥与控制事物发展的决策。信息对当前和将来的决策具有明显的或实际的价值。

(2) 信息源于物质和能量。信息不可能脱离物质而存在,信息的传递需要物质载体,信息的获取和传递要消耗能量。信息能够在空间和时间上传递。在空间上传递信息称为信息通信,在时间上传递信息称为信息存储。如信息可以通过报纸、电台、电视、计算机网络进行传递。

(3) 信息是有用的。信息是人们活动的必需知识,它可以提高人们对事物的认识,利用信息能够克服工作中的盲目性、增加主动性和科学性,可以把事情办得更好。

(4) 信息需要一定的形式表示,信息是可存储、加工、传递和再生的。信息与其表现符号不可分离。计算机存储技术的发展,进一步扩大了信息存储的范围。借助计算机,还可对收集到的信息进行管理。

信息对于人类社会的发展有重要意义。人通过获得、识别自然界和社会的不同信息来区别不同的事物,得以认识和改造世界。信息是社会机体进行活动的纽带,社会的各个组织通过信息网相互了解并协同工作,使整个社会协调发展;社会越发展,信息的作用就越突出;信息是管理活动的核心,要想管理好事物,需要掌握更多的信息,并利用信息进行工作。

在整个计算机应用中,信息系统所占比例高达70%~80%。一个国家的现代化水平越高,科学管理、自动化服务的要求就越迫切,各行各业的计算机信息系统所占的比例也越高。在信息系统的发展过程中,广泛使用到了数据管理技术。

2. 数据、数据与信息的关系及数据的特征

数据(data)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。数据是用于承载信息的物理符号。尽管信息有多种表现形式,它可以通过手势、眼神、声音或图形等方式表达,但信息的最佳表现形式和载体是数据。数据可以通过符号、文字、数字、语音、图像、视频等表示信息,因此数据能够被记录、存储和处理,从中挖掘出更深层的信息。

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,是对数据的语义解释。必须指出的是,在许多不严格的情况下,会把“数据”和“信息”两个概念混为一谈,称“数据”为“信息”。其实,数据不等于信息,数据只是信息表达方式中的一种,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。正确的数据可表达信息,而虚假、错误的数据所表达的是谬误,不是信息。

在计算机系统中,数据是用于输入计算机并被计算机程序处理,具有一定意义的数字、字母、符号的组合、图形、图像等符号介质的通称,这里的数字数据在某个区间内是离散的值。

通常数据具有以下特征:

(1) 数据有“型”和“值”之分。

数据的型是指数据的结构,而数据的值是指数据的具体取值。数据的结构指数据的内部构成和对外联系。例如,学生的数据由“学号”“姓名”“年龄”“性别”“所在系”等属性构成,其中“学生”为数据名,“学号”“姓名”等为属性名(或称数据项名);课程也是数据,它由“课程编号”“课程名称”“课时数”等数据项构成;“学生”和“课程”之间有“选课”的联系。“学生”和“课程”数据的内部构成及其相互联系就是学生课程数据的类型,而一个具体取值,如“2021936,张立,20,男,计算机系”就是一个学生数据值。

(2) 数据受数据类型和取值范围的约束。

数据类型是针对不同的应用场合设计的数据约束。根据数据类型的不同,数据的表示形式、存储方式及能进行的操作运算各不相同。在使用计算机处理信息时,应当对数据类型特别重视,为数据选择合适的类型,千万马虎不得。

常见的数据类型有数值型、字符串型、日期型和逻辑型等,它们具有不同的特点和用途。数值型数据就是通常所说的算术数据,它能够进行加、减、乘、除等算术运算;字符串型数据是最常用的数据,它可以表示姓名、地址、邮政编码及电话号码等类数据,能够进行查找子串、取子串和连接子串的运算操作;日期型数据适合表达日期和时间信息;逻辑型数据能够表达“真”和“假”以及“是”和“否”等逻辑信息。

数据的取值范围也称数据的值域,例如学生性别的值域是{“男”,“女”}。为数据设置值域是保证数据的有效性、避免数据输入或修改时出现错误的重要措施。

(3) 数据有定性表示和定量表示之分。

在表示职工的年龄时,可以用“老”“中”“青”定性表示,也可以用具体岁数定量表示。由于数据的定性表示是带有模糊因素的粗略表示方式,而数据的定量表示是描述事物的精确

表示方式,因此在计算机软件设计中,应尽可能地采用数据的定量表示方式。

(4) 数据应具有载体和多种表现形式。

数据是客体(即客观物体或概念)属性的记录,它必须有一定的物理载体。当数据记录在纸上时,纸张是数据的载体;当数据记录在计算机的外存上时,保存数据的硬盘、U 盘或光盘就是数据的载体。数据的概念在数据处理领域中已大大地拓宽了,数据具有多种表现形式,它可以用报表、图形、语音及不同的语言符号表示。可用多种不同的数据形式表示同一信息,而信息不随数据形式的不同而改变。

众所周知,21 世纪是知识与信息爆炸的时代,人们生活的周围遍布信息。而如何更好、更充分地获取信息并将其以数据形式存储在计算机中是极其重要的。数据经过计算机的处理,可以使人们更好地交流与传递、进行分析与统计等。此外,随着计算机、通信及网络的广泛应用,数据管理已成为各个部门、机构或单位,甚至是国家政府机关的重要支撑以及决策的主要手段,发挥着越来越重要的作用,也是生活在信息时代每个人必备的信息素养能力之一。

5.1.2 数据管理的变迁

数据管理即对数据资源的管理,是利用计算机硬件和软件技术对数据进行有效的收集、存储、处理和应用的过程。数据管理已成为人类进行正常社会活动的一种需求。

随着计算机技术的发展,数据管理从手工记录的人工管理阶段,发展到以文件形式保存在计算机存储器中的文件系统管理阶段,再到数据库管理阶段。数据管理技术的发展,以数据存储冗余不断减小、数据独立性不断增强、数据操作与维护更加便捷为标志,每一阶段都有其特点,如图 5-1 所示。

1. 人工管理阶段

20 世纪 40 年代至 50 年代这段时间中,由于当时计算机结构简单,应用面狭窄且存储单元少,计算机内的数据管理非常简单,这时就由应用程序编制人员各自直接管理自身的数据,程序与要处理的数值数据放在一起,此阶段称人工管理阶段,如图 5-2 所示。

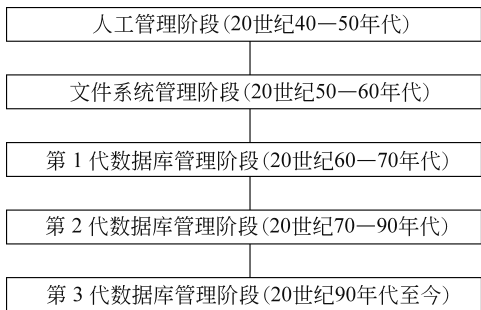


图 5-1 数据库管理变迁示意

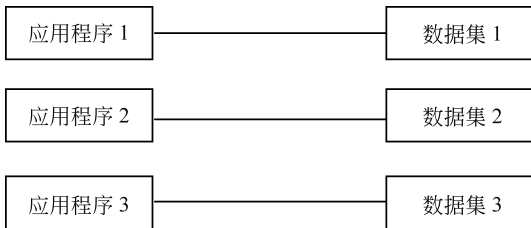


图 5-2 数据的人工管理阶段

在人工管理阶段,计算机主要用于科学计算,对于数据保存的需求尚不迫切,所以数据不加保存;每个应用程序都要包括数据的存储结构、存取方法、输入方式等,程序编制人员编写应用程序时,需要安排数据的物理存储和管理;数据面向程序一一对应,数据依赖于特定的应用程序,一组数据只能对应一个程序。多个应用程序涉及某些相同的数据时,也必须各

自定义,程序之间有大量的冗余数据;数据不具有独立性,数据完全依赖于程序。数据维护存在共享性等问题。

显然,这种数据管理技术主要适用于数据量小、数据间无逻辑组织关系的应用。由此,数据管理进入文件系统阶段。

2. 文件系统管理阶段

文件系统管理出现于 20 世纪 50 年代,这一阶段计算机不仅用于科学计算,还大量用于信息管理,数据存储、检索和维护成为紧迫的需求。此时计算机中已出现有磁鼓、磁盘等大规模存储设备,计算机应用也逐步拓宽,数据管理技术也得益于计算机的处理速度和存储能力的提高。计算机内的数据已开始有专门的软件进行管理,这就是文件系统,如图 5-3 所示。

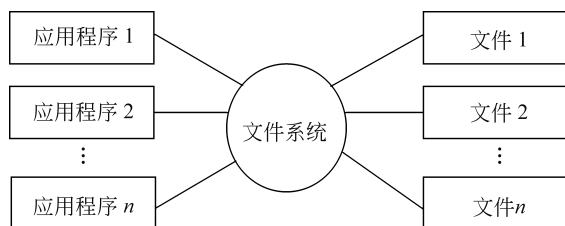


图 5-3 数据的文件系统管理阶段

在文件系统管理阶段,文件系统把计算机中的数据组织成相互独立的、可被命名的数据文件,同时可以按文件的名字来进行管理;此时的文件系统已能对数据进行初步的组织,数据便可以长期保存在计算机外存上,可以对数据进行反复处理,并支持文件的查询、修改、插入和删除等操作;文件形式多样化,包括顺序文件、倒排文件、索引文件等;文件系统实现了记录内部的结构化,但从文件的整体来看却是无结构的;数据文件是独立于程序而存在的,可以随意地增减要操作的数据。

与人工管理阶段相比,文件系统管理阶段对数据的管理有了很大的进步,但一些根本性问题仍没有彻底解决,文件管理方式的数据文件也是面向特定的应用程序,只是与应用程序分离;由于应用环境简单因此接口能力差,导致文件系统的的功能简单,且只能附属于操作系统而不能成为独立部分;具有数据独立性,数据管理简单化了,各数据文件之间缺少有机的联系。一个数据文件基本上对应于一个应用程序,数据仍然不能共享,数据冗余度大;由于相同数据的重复存储、各自管理,在进行更新操作时,容易造成数据的不一致性。可以将其看成是数据库系统的雏形,而不是真正的数据库管理系统。

3. 数据库管理阶段

自 20 世纪 60 年代起,随着硬件环境与软件环境的不断改善与提高,数据处理应用领域需求的持续扩大,计算机存储设备已出现大容量磁盘与磁盘组,且数据量已跃至海量,文件系统已无法满足新的数据管理要求。数据管理职能由附属于操作系统的文件系统而脱离成独立的数据管理机构,即数据库管理系统,由此数据管理进入了数据库管理系统阶段。

从文件系统到数据库系统,标志着数据管理技术质的飞跃。数据库的特点是数据不再只针对某一特定应用,而是面向全组织,具有整体的结构性,共享性高、冗余度小,具有一定的程序与数据间的独立性,并且实现了对数据进行统一的控制。数据库技术的应用使数据

存储量猛增,用户增加,而且数据库技术的出现使数据处理系统的研制从围绕以加工数据的程序为中心转向围绕共享的数据来进行,实现了整体数据的结构化,如图 5-4 所示。

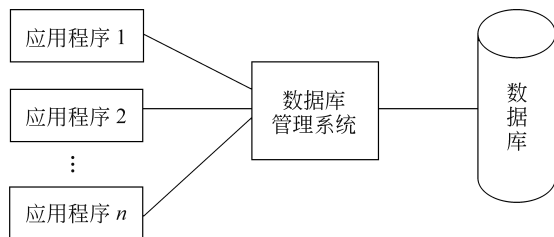


图 5-4 数据的数据库管理阶段

数据库技术与其他软件技术的加速融合,进一步促进了数据管理的模式与功能结构的改进。在数据库管理阶段,因不同的数据结构组织而分成为:

1) 层次与网状数据库管理时代

20 世纪 60 年代以后所出现的数据库管理系统是层次数据库与网状数据库,它们具有了真正的数据库管理系统特色,但是它们脱胎于文件系统,受文件的物理特性影响大,因此给数据库使用和应用带来诸多不便。

1969 年,IBM 公司 McGee 等开发的层次数据库系统的 IMS 发表。1971 年,美国数据库系统语言协会下属的数据库任务组对网状数据库方法进行了系统的研究、探讨,提出了网状数据库系统的许多概念、方法和技术,标志着数据库在理论上的成熟。

2) 关系数据库管理时代

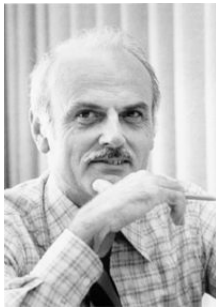


图 5-5 E.F.Codd

20 世纪 70 年代是关系数据库理论研究和开发原型的时代。随着其蓬勃的发展逐步取代层次与网状数据库系统。1970 年 IBM 公司 E.F.Codd(见图 5-5)发表了题为《大型共享数据库数据的关系模型》的论文,提出了关系数据模型,开创了关系数据库方法和关系数据库理论。由于 E.F.Codd 的杰出贡献,他于 1981 年获得图灵奖。

关系数据库管理系统用严格的数学理论来描述数据库的组织 and 操作,结构简单、使用方便、逻辑性强,被认为是最有前途的数据库管理系统,因此发展十分迅速。20 世纪 80 年代以后,关系数据库管理系统一直占据数据库领域的主导地位。各大计算机厂商先后推出 dBASE、FoxBASE、Oracle、FoxPro、Access 等多种商品数据库管理系统。

3) 新一代数据库管理时代

在 20 世纪 90 年代以后,数据库理论和应用进入成熟发展时期。数据库应用领域也在不断地扩大,数据库技术在商业领域的巨大成就刺激了其他领域对数据库需求的迅速增长。数据库逐步扩充至非事务处理领域与数据分析领域,传统关系数据库的应用受到了挑战。需要针对关系数据库管理系统做出必要的改造与扩充:

- (1) 引入面向对象概念,建立对象关系数据库管理系统,以适应非事务处理领域应用;
- (2) 扩充数据交换能力,以适应数据库在网络及互联网环境中的应用;
- (3) 引入联机分析处理概念,建立数据仓库,以适应数据分析处理领域的应用。

数据库技术与其他现代数据处理技术(如面向对象技术、分布式技术、时序和实时处理技术、人工智能技术、多媒体技术等)完美地集成,形成了“新一代数据库技术”,也可称为“现代数据库技术”,如时态数据库技术、分布式数据库技术、实时数据库技术和多媒体数据库技术等。新型数据库系统应运而生,也带来了一个又一个数据库技术发展的高潮,为数据的处理带来更便捷、更宽泛的管理。

新一代的数据库系统支持数据管理、对象管理和知识管理,保持和继承了关系数据库已有的技术,支持数据库语言标准,在网络上支持标准网络协议等。当然,由于新一代数据库系统的专业性要求高,对于中小数据库用户来说,其通用性受到一定的限制。

5.2 数据库基础

5.2.1 基本概念

1. 数据库

数据库(DB)是数据的集合,它具有统一的结构形式,存放于统一的存储介质内,并由统一机构管理。它由多种应用数据集成,并可被多个应用所共享,具有尽可能小的冗余度。

数据库存放数据,数据按所提供的数据模式存放,它能构造复杂的数据结构以建立数据间内在联系与复杂关系,从而构成数据的全局结构模式。

数据库中的数据具有“集成”“共享”的特点,也即数据库集中了各种应用的数据,并对其进行统一的构造与存储,而数据可为不同应用服务与使用。数据库本身不是独立存在的,它是组成数据库系统的一部分。

2. 数据库管理系统

数据库管理系统(DataBase Management System,DBMS)是统一管理数据库、使用户可以定义、创建和维护数据库以及提供对数据库有限访问的软件(属系统软件),DBMS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。一般来说,DBMS通过提供统一的数据语言(data language)具体完成。

(1) 数据模式定义。数据库管理系统负责为数据库构造统一数据框架,这种框架称为数据模式,而这种功能称为数据组织。

(2) 数据操纵。数据库管理系统为用户定位与查找数据提供方便,它一般提供数据查询、插入、修改以及删除的功能,用于访问和操作数据库中的数据,处理用户请求。此外,它自身还具有一定的运算、转换及统计的能力和一定的过程调用能力。

(3) 数据控制。用于检索存储或保存的数据,授予或收回用户对数据库的访问权限。数据库管理系统负责数据语法、语义的正确性保护,称为数据完整性控制。数据库管理系统还负责数据访问正确性保护,称为安全性控制。此外,数据库管理系统还负责数据动态正确性保护,具体为并发控制与故障恢复。

(4) 数据交换。数据库管理系统为不同环境用户使用数据提供相应的接口,实现数据的交换。

(5) 数据的扩展功能。为使数据管理更好地为数据处理服务,在数据管理中增加一些对数据处理的延伸服务,这就是数据的扩展功能。它包括人机交互、嵌入式、自含式、调用层

接口以及 Web 数据库、XML 数据库等扩展功能。

(6) 数据服务(data service)。数据库管理系统提供对数据库中数据的多种服务功能称为数据服务。

(7) 数据字典。数据字典(data dictionary)是一组特殊的数据服务,它是信息服务的一种,又称元数据(metadata)。数据字典存放数据库管理系统中的数据模式结构、数据完整性规则、安全性要求等数据。

3. 数据库系统

数据库系统(DataBase System,DBS)是由数据库及其管理软件组成的系统。它是为适应数据处理的需要而发展起来的一种较为理想的数据处理系统,是一个为实际可运行的存储、维护和应用系统提供数据支撑的软件系统,是存储介质、处理对象和管理系统的集合体。

数据库系统由硬件系统、系统软件(包括操作系统、数据库管理系统等)、数据库应用系统和各类人员四部分组成,如图 5-6 所示。

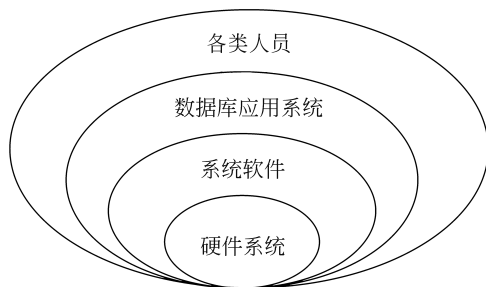


图 5-6 数据库系统的组成

1) 硬件系统

硬件系统的配置应满足整个数据库系统的需要。由于一般数据库系统数据量很大,加之 DBMS 丰富而强大的功能使得自身的体积很大,因此整个数据库系统对硬件资源提出了较高的要求:

(1) 具有足够大的内存以保证存放操作系统、DBMS 的核心模块、数据缓冲区和应用程序。

(2) 具有足够大的直接存取设备存放数据并完成数据备份。

(3) 要求计算机有较高的数据传输能力,以提高数据传送率。

2) 系统软件

系统软件主要包括操作系统、DBMS、与数据库接口的高级语言及其编译系统,以及以 DBMS 为核心的应用开发工具。

操作系统是计算机系统必不可少的系统软件,也是支持 DBMS 运行必不可少的系统软件。DBMS 是数据库系统不可或缺的系统软件。它提供数据库的建立、使用和维护等功能。一般来讲,DBMS 的数据处理能力较弱,所以需要与数据库接口的高级语言及其编译系统,以便于开发应用程序。以 DBMS 为核心的应用开发工具指的是系统为应用开发人员和最终用户提供的高效率、多功能的应用生成器、第四代语言等各种软件工具。

3) 数据库应用系统

数据库应用系统是为特定应用开发的数据库应用软件。DBMS 为数据的定义、存储、

查询和修改提供支持,而数据库应用系统是对数据库中的数据进行处理和加工的软件。例如,基于数据库的各种管理软件,管理信息系统、决策支持系统和办公自动化等都属于数据库应用系统。

4) 各类人员

参与分析、设计、管理、维护和使用数据库的人员均是数据库系统的组成部分。他们在数据库系统的开发、维护和应用中起着重要的作用。各类人员主要包括:

(1) 数据库管理员(DataBase Administrator,DBA)。

由于数据库的共享性,因此对数据库的规划、设计、维护、监视都需要有专人管理。数据库管理员负责数据库的总体信息控制。

数据库管理员的具体职责包括:根据数据库中的信息内容和结构,决定数据库的存储结构和存取策略;定义数据库的安全性要求、完整性约束条件、并发控制及系统恢复;监控数据库的使用和运行;负责数据库的性能改进,数据库的调整、重组及重构,保证其运行的效率,以提高系统的性能;负责制定与使用数据库有关的规章制度、检查落实人员培训和咨询等工作。

(2) 系统分析员(System Analyst,SA)和数据库设计人员。

系统分析员是数据库系统建设期的主要参与人员,负责应用系统的需求分析和规范说明,要和最终用户相结合,确定系统的数据库结构和应用程序的设计,以及软硬件的配置,并参与组织整个数据库系统的概要设计。

数据库设计人员负责数据库中数据的确定、数据库各级模式的设计。

(3) 应用程序员(Applications Programmer,AP)。

应用程序员根据系统的功能需求,负责设计和编写数据库的应用程序。这些应用程序可对数据进行建立、删除、修改或检索,并参与对程序模块的测试。

(4) 最终用户。

数据库系统的最终用户是有不同层次的。不同层次的用户其需求的信息以及获得信息的方式也是不同的。最终用户利用系统的接口或查询语言访问数据库。一般可将最终用户分为操作层、管理层和决策层。

5.2.2 数据抽象

数据库管理的对象(数据)存在于现实世界中,抽象于现实世界中的事物及其各种关系。在数据处理中,数据抽象涉及许多范畴。数据从现实世界到计算机数据库里的具体表示要经历三个阶段,即现实世界、信息世界和计算机世界,并通过三个阶段的二次抽象将现实世界的事物抽象为计算机中的数据描述,如图 5-7 所示。

1. 现实世界

现实世界是指客观存在世界中的事物及其联系。在这一阶段要对现实世界的事物进行收集、分类,并抽象成信息世界的描述形式。

2. 信息世界

信息世界是现实世界在人们头脑中的反映。经过人脑的分析、归纳和抽象所形成的信息是对客观

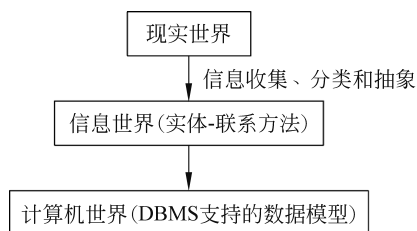


图 5-7 数据处理的三个阶段

事物及其联系的一种抽象描述,把这些信息进行记录、整理、归类和格式化后,就构成了信息世界。在数据库设计中,这一阶段又称为概念设计阶段。

1) 常用术语

(1) 实体。客观存在并可相互区别的事物。实体可以是现实世界中具体的人、事、物,也可以是抽象的概念或联系,例如一个学生、一个教师、一所学校、一门课、一次会议、一堂课、一场球赛等。这里从建立信息结构的角度出发,强调实体是被认识的客观事物,未被认识的客观事物就不可能找出它的特征,也就无法建立起相应的信息结构。

(2) 实体集。性质相同的同类实体的集合叫实体集,如教师、学生、课程等实体集。研究实体集的共性是信息世界的基本任务之一。

(3) 属性。实体的某一特征称为属性。每个实体都有许多特征,以区别于其他实体。如一本书的主要特征是书名、作者名、出版社、出版年月和定价等;一次会议的主要特征是会议名称、会议时间、会议地点、参加对象及参加人数等。特征是在对客观事物进行深入分析的基础上归纳出来的。属性也称为“型”。实体集中实体具有相同的性质,即指的是具有相同的属性(或相同的型)。

(4) 元组。实体的每个属性都有一个确定值,称为属性的值。当某实体有多个属性时,它们的值就构成一组值,称为元组。实体在信息世界中就是通过元组来表示的。属性的取值有一定的范围,这个范围称为属性域(或值域)。如描述人的年龄属性,可定在1~200的整数范围内;若对于具体某个人的年龄值,可能取值为50。

(5) 码。唯一标识实体的属性集称为码,例如学号是学生实体的码,座位所在的行号和列号也是座位实体的码。

(6) 联系。实体间的“联系”反映了现实世界中客观事物之间的关联。这种联系是复杂的、多种多样的,但归纳起来可分一对一、一对多和多对多三类。

① 一对一联系。

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,描述学校的客观事物时,对于班级和(正)班长两个实体集,一个班级只有一个(正)班长,而一个(正)班长只在一个班级中任职,则实体集班级与(正)班长之间具有一对一联系。

② 一对多联系。

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 中至多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,对于班级和学生两个实体集,一个班级中有若干名学生,而每个学生只在一个班级中学习,则实体集班级与学生之间具有一对多联系。

③ 多对多联系。

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 中也有 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,对于课程和学生两个实体集,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,则实体集课程与学生之间具有多对多联系。

2) 概念模型

从现实世界抽象到信息世界后,抽象的结果是通过概念模型来表达的。它是面向现实世界建模,是面向用户的模型。它按用户的观点对数据和信息建模,描述现实世界的概念化结构,这时与具体的 DBMS 和具体的计算机平台无关。概念模型只是用来描述某个特定组织所关心的信息结构。它使设计人员在设计初始阶段摆脱计算机系统及 DBMS 的具体技术问题,集中精力分析数据本身的特性及数据之间的联系。它是系统分析员、程序设计员、维护人员、各级用户之间相互理解的共同语言。

概念模型是在了解了用户的需求、用户的业务领域工作情况以后,经过分析和总结,提炼出来的用以描述用户业务需求的一些概念性的东西。概念模型的表示方法很多,其中最著名、最常用的是麻省理工学院的陈品山(Peter Pin-Shan Chen)于 1976 年提出的实体-联系方法(Entity-Relationship Approach,E-R 方法)。该方法用 E-R 图来描述现实世界的概念模型,也称为 E-R 模型,如图 5-8 所示。

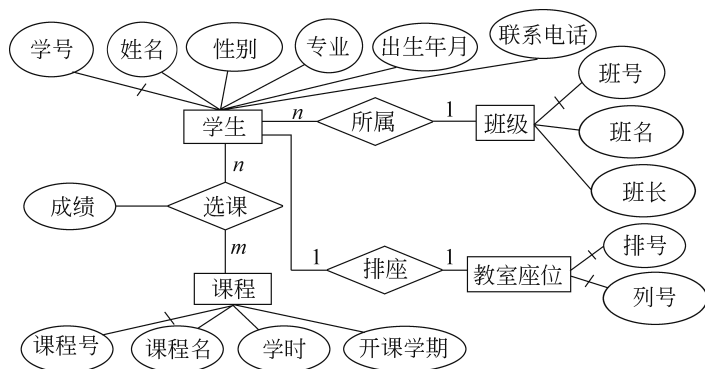


图 5-8 学生选课系统的 E-R 图

E-R 图主要是由实体,属性和联系三个要素构成,其中:

- (1) 实体用矩形框表示,矩形框内写明实体名。
- (2) 属性用椭圆表示,并用无向边将其相应的实体连接起来。例如,学生实体集具有学号、姓名、性别、出生年月、专业等属性。
- (3) 联系用菱形表示,菱形框内写明联系名,并用无向边与有关实体集连接起来,同时,在无向边旁边标注上联系的类型。

3. 计算机世界

这一阶段的数据处理是在信息世界对客观事物的描述基础上做进一步抽象,将其信息化,使得信息能够存储在计算机中,是对信息世界中信息的数据化,所以又叫数据世界。它将信息用字符和数值等数据表示,具体使用计算机存储并管理信息世界中描述的实体集、实体、属性和联系的数据。信息世界抽象到计算机世界,则概念模型被抽象为数据模型。数据模型是对数据及其联系的描述。概念模型中的实体内部的联系抽象为数据模型中同一记录内部各字段间的联系,实体之间的联系抽象为记录与记录之间的联系。这一阶段的数据处理在数据库的设计过程中也称为逻辑设计。