

数据库基础知识

数据处理是目前计算机应用的主要方面,数据处理的核心是数据管理,而数据库技术是数据管理的最先进技术。在信息技术日益普及的今天,数据库技术已经深入人类社会的各个方面,并且随着计算机技术和互联网的迅猛发展,数据库技术的应用领域也在不断扩大。本章主要介绍数据管理技术的基础知识和关系数据库的基本概念等内容。

1.1 数据管理技术

1.1.1 数据和数据管理

1. 信息和数据

信息是指现实世界中事物的存在方式或运动状态的反映,数据则是描述现实世界事物的符号记录形式,是利用物理符号记录下来的可以识别的信息,这里的物理符号包括数字、文字、图形、图像、声音和其他特殊符号。

数据的概念包括两个方面:一是描述事物特性的数据内容;二是存储在某一种媒体上的数据形式。

数据处理是指将数据转换成信息的过程,从数据处理的角度来看,信息是一种被加工成特定形式的信息,这种信息形式是数据接收者希望得到的。

数据和信息之间的关系非常密切,数据是信息的符号表示或载体,信息则是数据的内涵,是对数据的语义解释。在某些不需要严格区分的场合,可以将两者不加区别地使用,如将信息处理说成数据处理。

2. 数据管理

数据管理包括对各种形式的数据进行收集、存储、加工和传输等的一系列活动。其目的之一是从大量原始数据中抽取、推导出对人们有价值的信息,将这些信息作为行动和决策的依据;另一目的是借助计算机科学地保存和管理复杂的、大量的数据,以使人们能够方便而充分地利用这些宝贵的信息资源。

1.1.2 数据管理技术的发展

数据管理技术随着计算机硬件和软件的发展而不断发展,在应用需求的推动下,数据管理技术经历了人工管理、文件系统和数据库系统 3 个发展阶段,如表 1-1 所示。

表 1-1 数据管理技术发展的 3 个阶段

发展阶段	主要特征
人工管理 (1950—1965 年)	(1)应用程序管理数据。 (2)数据不共享,一组数据只能对应一个程序,数据冗余度大。 (3)数据不具有独立性,数据与程序彼此依赖
文件系统 (1965—1970 年)	(1)数据由文件系统管理,应用程序通过文件系统访问数据文件中的数据。 (2)数据文件之间没有联系,数据共享性差,冗余度大。 (3)数据独立性差,数据仍高度依赖于程序,数据是为特定的应用服务的
数据库系统 (1970 年至今)	(1)数据由数据库管理系统统一管理和控制。 (2)数据是面向全组织的,共享性高,冗余度小。 (3)数据具有较高的逻辑独立性和物理独立性

1.2 数据库系统

1.2.1 数据库系统的组成

数据库系统(database system,DBS)是指带有数据库并利用数据库技术进行数据管理的计算机系统,它可以实现有组织地、动态地存储大量相关数据,提供数据处理和信息资源共享服务。

数据库系统由以下 5 部分组成。

(1)数据库(database,DB):存放数据的仓库。只不过这个仓库是在计算机存储设备上,而且数据是按一定的格式存放的。也就是说,数据库是数据的集合,并按照特定的组织方式将数据保存在存储介质上,同时可以被各种用户共享。数据库中的数据具有较小的冗余度、较高的独立性和扩展性。

数据库中不仅包括描述事物的数据本身,而且包括事物之间的联系。

(2)数据库管理系统(database management system,DBMS):数据库系统的核心,是一种系统软件,负责数据库中的数据组织、操纵、维护、控制、保护和数据服务等。数据库管理系统是位于用户与操作系统之间的数据管理软件。

数据库管理系统的主要功能如下:

- ①数据模式定义与数据的物理存取构建。
- ②数据操纵,包括数据更新(增加、删除、修改)和数据查询。
- ③数据控制,包括完整性和安全性定义、数据的并发控制与故障恢复。
- ④数据服务,包括数据保存、重组、分析等。

(3)硬件:支持系统运行的计算机硬件设备。

(4)软件:包括操作系统、应用开发工具和数据库应用系统。

(5)相关人员:数据库系统中的相关人员,包括数据库管理员、系统分析员、数据库设计人员、应用程序开发人员和最终用户。

1.2.2 数据库系统的特点

1. 数据结构化

在数据库系统中,每一个数据库都是为某一应用领域服务的,这些应用彼此之间都有着密切的联系。因此,在数据库系统中不仅要考虑某个应用的数据结构,还要考虑整个组织(多个应用)的数据结构。这种数据组织方式使数据结构化了,这就要求在描述数据时不仅要描述数据本身,还要描述数据之间的联系。

2. 数据共享性高、冗余度低

数据共享是指多个用户或应用程序可以访问同一个数据库中的数据,而且数据库管理系统提供并发和协调机制,保证在多个应用程序同时访问、存取和操作数据库数据时,不产生任何冲突,从而保证数据不遭到破坏。

数据冗余既浪费存储空间,又容易产生数据不一致等问题。

3. 具有较高的数据独立性

数据独立性是指应用程序与数据库的数据结构之间相互独立。在数据库系统中,因为采用了数据库的三级模式结构,保证了数据库中数据的独立性。在数据存储结构改变时,不影响数据的全局逻辑结构,这样保证了数据的物理独立性。在全局逻辑结构改变时,不影响用户的局部逻辑结构及应用程序,这样就保证了数据的逻辑独立性。

4. 有统一的数据控制功能

数据库管理系统提供了一套有效的数据控制手段,包括数据安全性控制、数据完整性控制、数据库的并发控制和数据库的恢复等,增强了多用户环境下数据的安全性和一致性保护。

1.2.3 数据库的结构体系

数据库领域公认的标准结构是三级模式结构及二级映射。三级模式包括概念模式、外模式和内模式,二级映射则分别是概念模式/内模式的映射及外模式/概念模式的映射。三级模式与二级映射构成了数据库的结构体系,如图 1-1 所示。

1. 数据库的三级模式

(1)概念模式。概念模式又称逻辑模式,或简称为模式,对应于概念级。它是由数据库设计者综合所有用户的数据,按照统一的观点构造的全局逻辑结构,是对数据库中全部数据的逻辑结构和特征的总体描述,是所有用户的公共数据视图(全局视图)。

(2)外模式。外模式又称子模式或用户模式,对应于用户级。它是某个或某几个用户所看到的数据库的数据视图,是与某一应用有关的数据的逻辑表示。外模式是从概念模式导出的一个子集,包含概念模式中允许特定用户使用的那部分数据。

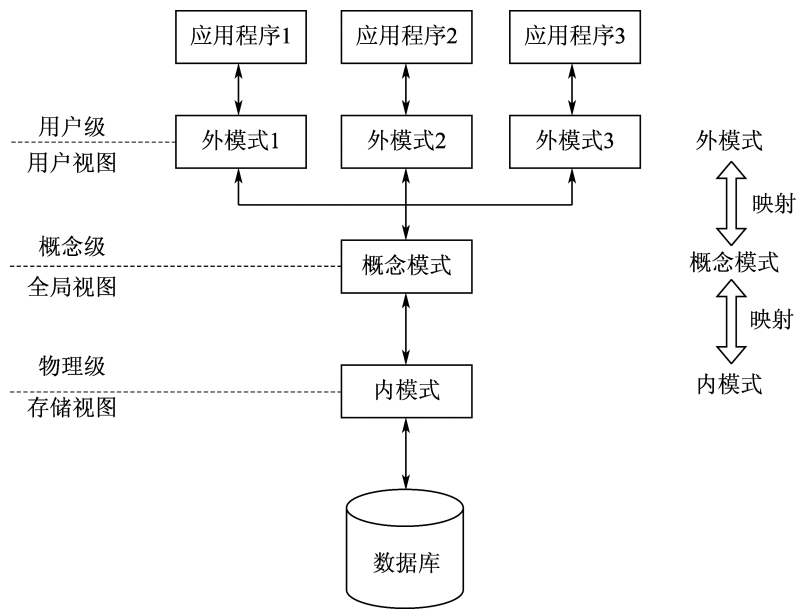


图 1-1 数据库的结构体系

(3)内模式。内模式又称存储模式或物理模式,对应于物理级。它是数据库中全体数据的内部表示或底层描述,是数据库最低一级的逻辑描述,它描述了数据在存储介质上的存储方式和物理结构,对应着实际存储在外存储介质上的数据库。

2. 三级模式间的二级映射

为了实现这 3 个抽象级别的联系和转换,数据库管理系统在三级模式之间提供了二级映射,正是这二级映射保证了数据库中的数据具有较高的物理独立性和逻辑独立性。

1.3 数据模型

在数据库技术中,用数据模型(data model)来对现实世界中的数据进行抽象和表示。

1.3.1 数据模型的组成要素

数据模型是一种描述数据、数据之间的联系及有关语义约束规则的形式化方法。这些规则分为 3 个方面:描述实体静态特征的数据结构、描述实体动态特征的数据操作规则及描述实体语义要求的数据完整性约束规则。因此,数据结构、数据操作及数据完整性约束也被称为数据模型的 3 个组成要素。

1.3.2 数据抽象的过程

数据模型是对现实世界进行抽象和转换的结果,这一过程如图 1-2 所示。

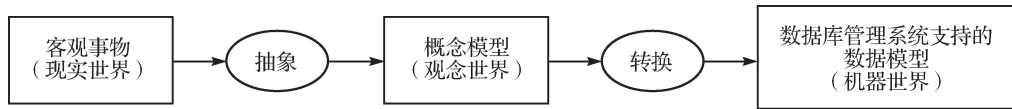


图 1-2 数据模型的抽象和转换

1. 对现实世界的抽象

计算机处理的对象是现实世界中的客观事物,在对现实世界实施处理的过程中,首先应了解和熟悉现实世界,从对现实世界的调查和观察中抽象出大量描述客观事物的事实,再对这些事实进行整理、分类和规范,进而将规范化的事实数据化,最终实现由数据库系统进行存储和处理。

2. 观念世界中的概念模型

概念模型的特征是按用户需求观点对数据进行建模,表达了数据的全局逻辑结构,是系统用户对整个应用项目涉及的数据的全面描述。概念模型主要用于数据库设计,它独立于实现时的数据库管理系统。

概念模型的表示方法很多,目前较常用的是 E-R 模型。

3. 机器世界中的逻辑模型和物理模型

在选定数据库管理系统后,就要将 E-R 图表示的概念模型转换为具体的数据库管理系统支持的逻辑模型。通常,也把数据的逻辑模型直接称为数据模型。数据库系统中主要的逻辑模型有层次模型、网状模型和关系模型。

物理模型是对数据最底层的抽象,用以描述数据在物理存储介质上的组织结构,与具体的数据库管理系统、操作系统和硬件有关。

1.3.3 概念模型

现实世界存在各种不同的事物,各种事物之间既存在联系又有差异,事物数据化过程就是要对事物的特征及事物之间的联系进行抽象化和数据化,在计算机内处理的各种数据实际上是客观存在的不同事物及事物之间的联系在计算机中的表示。

1. 实体与实体集

实体(entity)是现实世界中任何可以相互区分和识别的事物,可以是能触及的客观对象,还可以是抽象的事件。

实体是客观事物的真实反映,既可以是实际存在的对象,如一位教师、一本书、一台机器等,也可以是某种抽象概念或事件,如一门课程、一个专业、一次借阅图书、一个运行过程等。

性质相同的同类实体的集合称为实体集(entity set)。在关系数据库(如 Oracle、Visual FoxPro、Access 等)中,通常将同一种实体型的数据存放在一个表中,实体属性集合作为表结构,而一个实体属性值的集合作为表中一个数据记录,表示一个实体。

2. 属性

每个实体都具有一定的特征或性质,这样才能区分一个个实体。实体的特征称为属性(attribute),一个实体可用若干属性来描述。

能唯一标识实体的属性或属性集称为实体标识符。

3. 类型与值

属性和实体都有类型(type)和值(value)之分。属性类型就是属性名及其取值类型,属性值就是属性所取的具体值。属性类型是个变量,属性值是变量所取的值,而值域是变量的取值范围。

实体类型就是实体的结构描述,通常是实体名和属性名的集合。具有相同属性的实体有相同的实体类型。实体值是一个具体的实体,是属性值的集合。

因此,通过实体类型可以区分不同类型的事物。

例如,分别用教师(教师编号、姓名、性别、出生日期、职称、联系电话、所属院系),课程(课程编号、课程名称、开课学期、学时、学分)的形式来描述教师类实体和课程类实体。

4. 实体间的联系

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 中至多只有一个实体与之联系,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,一个班级只有一个正班长,一个正班长管理一个班级,因此,班级与正班长之间是一对一联系。

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 中可以有多个实体与之联系,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,一个班级有多个学生,而某个学生只隶属于一个班级,因此,班级与学生之间是一对多联系。

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 中可以有多个实体与之联系,而对于实体集 B 中的每一个实体,实体集 A 中也可以有多个实体与之联系,则称实体集 A 与实体集 B 之间有多对多联系,记为 m:n。例如,一个学生选修多门课程,而一门课程有多名学生选修,因此,学生与课程之间是多对多联系。

5. E-R 图

E-R 图是用一种直观的图形方式建立现实世界中实体及其联系模型的工具,也是一种数据库设计的基本工具。

E-R 图用矩形框表示现实世界中的实体,用菱形框表示实体间的联系,用椭圆形框表示实体和联系的属性。图 1-3 所示为读者实体和图书实体的多对多联系模型。

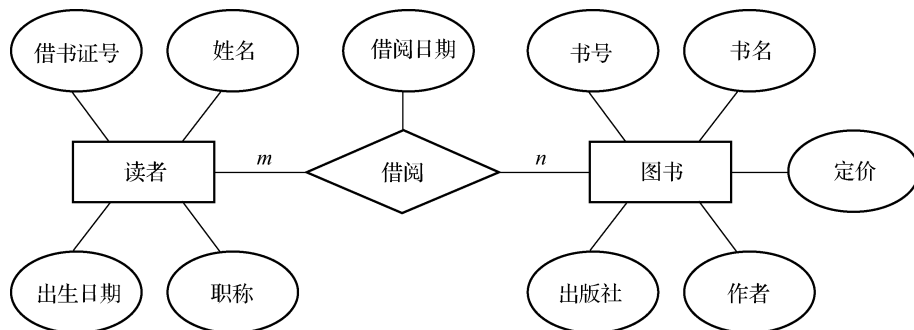


图 1-3 读者实体和图书实体的多对多联系模型

1.3.4 逻辑模型

数据模型是数据库管理系统中用于描述实体及其实体之间联系的方法,实体及其实体之间的联系用结构化数据体现出来,数据模型恰恰表示了这些结构化数据的逻辑关系,因此,任何一种数据库管理系统都需要用数据模型进行描述。用于描述数据库管理系统的数据模型有层次模型、网状模型和关系模型 3 种。

1. 层次模型

层次模型是通过树型结构表示实体及其实体之间联系的数据模型,“树”中每个结点表示一个实体类型,如图 1-4 所示。

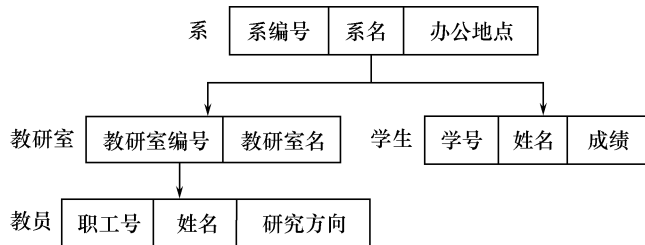


图 1-4 层次模型

层次模型的特点是:有且仅有一个结点,没有父结点,称之为根结点;每个非根结点有且仅有一个父结点。

在数据库技术中将支持层次数据模型的数据库管理系统称为层次数据库管理系统。

2. 网状模型

网状模型是通过网状结构表示实体及其实体之间联系的数据模型。“网”中每个结点表示一个实体类型,如图 1-5 所示。

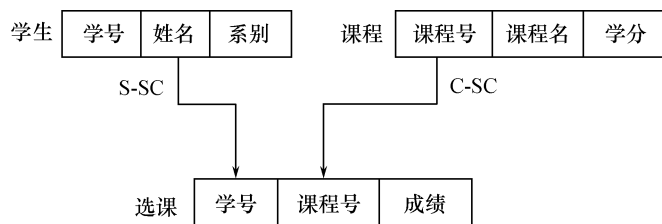


图 1-5 网状模型

网状模型的特点是:可能有多个结点,没有父结点,即有多个根结点;某个非根结点可能有多个父结点。

在数据库技术中将支持网状数据模型的数据库管理系统称为网状数据库管理系统。

3. 关系模型

关系模型是通过二维结构表示实体及其实体之间联系的数据模型,用一张二维表(学生表)来表示一种实体类型,表中一行数据描述一个实体,如图 1-6 所示。

学号	姓名	性别	出生日期	民族	籍贯	入学成绩	有否奖学金	所属学院
20190001	潘石屹	男	2000/2/16	汉族	江苏	585.00	<input type="checkbox"/>	商学院
20190002	王石	男	2000/4/25	汉族	湖北	575.00	<input type="checkbox"/>	外国语学院
20190003	蒋锡培	男	1999/10/8	汉族	江苏	598.00	<input checked="" type="checkbox"/>	物理学院
20190004	张丽敏	女	2000/6/25	汉族	贵州	569.00	<input type="checkbox"/>	外国语学院
20190005	孙建芹	女	2000/1/2	汉族	云南	576.00	<input type="checkbox"/>	数学科学学院
20190006	叶茂中	男	1999/9/16	汉族	广东	580.00	<input type="checkbox"/>	商学院
20190007	王冉	女	1999/11/21	汉族	湖南	583.00	<input checked="" type="checkbox"/>	物理学院
20190008	熊晓鸽	女	2000/4/2	汉族	湖南	605.00	<input checked="" type="checkbox"/>	计算机学院

图 1-6 关系模型

1.4 关系数据库

关系数据模型具有坚实的数学理论基础,通过实践证明:它是简单的、易于人们理解的、容易实现的一种数据模型。因此,目前广泛使用的 Visual FoxPro、Access、Oracle 等都采用了这种关系模型,即它们都是关系数据库管理系统。

1.4.1 关系数据库的基本概念

1. 关系

一个关系就是一张二维表,表是属性及属性值的集合。

通常将一个没有重复行、重复列,并且每个行列的交叉点只有一个基本数据的二维表格看成一个关系。二维表格包括表头和表中的内容,相应地,关系包括关系模式和记录的值,表包括表结构(记录类型)和表的记录,而满足一定条件的规范化关系的集合就构成了关系模型。

关系是一种规范化的二维表格。在关系模型中,对关系做了种种规范性限制,关系具有以下 6 条性质:

- (1)关系必须规范化,每一个属性都必须是不可再分的数据项。
- (2)每一列中的分量是同一类型的数据,来自同一个域。
- (3)在同一关系中不允许出现相同的属性名。
- (4)关系中不允许有完全相同的元组。
- (5)在同一关系中元组(tuple)的次序无关紧要。
- (6)在同一关系中属性的次序无关紧要。

2. 元组

二维表格的每一行在关系中称为元组,相当于表的一条记录(record)。二维表格的一行描述了现实世界中的一个实体。在关系数据库中,行是不能重复的,即不允许两行的全部元素完全对应相同。

3. 属性

二维表格的每一列在关系中称为属性(attribute),相当于记录中的一个字段(field)或数据项。每个属性有一个属性名,一个属性在其每个元组上的值称为属性值,因此,一个属性包括多

个属性值,只有在指定元组的情况下,属性值才是确定的。同时,每个属性有一定的取值范围,称为该属性的值域。同理,在关系数据库中,列是不能重复的,即关系的属性不允许重复。属性必须是不可再分的,即属性是一个基本的数据项,不能是几个数据的组合项。

表中每一列称为一个属性(字段),每列都有属性名,其也称为列名或字段名。例如,学号、姓名和出生日期都是属性名。

4. 关键字

关系中能唯一区分、确定不同元组的单个属性或属性组合,称为该关系的一个关键字。关键字又称为键或码(key)。单个属性组成的关键字称为单关键字,多个属性组合的关键字称为组合关键字。需要强调的是,关键字的属性值不能取“空值”,因为“空值”无法唯一地区分、确定元组。所谓“空值”,就是“不知道”或“不确定”的值。

关系中能够作为关键字的属性或属性组合可能不是唯一的。在关系中能够唯一区分、确定不同元组的属性或属性组合称为候选关键字(candidate key)。例如,学生表的候选关键字有学号和身份证号。

在候选关键字中选定一个作为关键字,其称为该关系的主关键字或主键(primary key)。关系中主关键字是唯一的。

5. 外部关键字

如果关系中某个属性或属性组合并非本关系的关键字,但却是另一个关系的关键字,则称这样的属性或属性组合为本关系的外部关键字或外键(foreign key)。在关系数据库中,用外部关键字表示两个表之间的联系。

1.4.2 关系运算

将关系与二维表进行比较可以看出两者存在简单的对应关系,关系模式对应一个二维表的表头,而关系的一个元组就是二维表的一行。在很多时候,甚至不加区别地使用这两个概念。

对关系数据库进行查询时,需要找到用户需要的数据,这就需要对关系进行一定的关系运算。关系的基本运算有两类:传统的集合运算和专门的关系运算。

1. 传统的集合运算

进行传统集合运算的两个关系必须具有相同的模式,即元组具有相同的结构。

(1)并运算(union)。设 R 和 S 同为 n 元关系,且相应的属性取自同一个域,则 R 和 S 的并也是一个 n 元关系,记作 $R \cup S$ 。 $R \cup S$ 包含了所有分属于 R 和 S 或同属于 R 和 S 的元组。因为集合中不允许有重复元素,因此,同时属于 R 和 S 的元组在 $R \cup S$ 中只出现一次。

(2)差运算(difference)。设 R 和 S 同为 n 元关系,且相应的属性取自同一个域,则 R 和 S 的差也是一个 n 元关系,记作 $R - S$ 。 $R - S$ 包含了所有属于 R 但不属于 S 的元组。

(3)交运算(intersection)。设 R 和 S 同为 n 元关系,且相应的属性取自同一个域,则 R 和 S 的交也是一个 n 元关系,记作 $R \cap S$ 。 $R \cap S$ 包含了所有同属于 R 和 S 的元组。

实际上,交运算可以通过差运算的组合来实现,如 $A \cap B = A - (A - B)$ 或 $B - (B - A)$ 。

下面通过实例说明上述3种运算,已知两个关系 R 和 S ,如图1-7所示。

R		
a	b	c
1	2	3
4	5	6
7	8	9

S		
a	b	c
1	2	3
7	8	9
10	11	12

图 1-7 关系 R 和关系 S(一)

关系 R 和关系 S 的并运算、交运算和差运算的结果如图 1-8~图 1-10 所示。

$R \cup S$

a	b	c
1	2	3
4	5	6
7	8	9
10	11	12

图 1-8 并运算

$R \cap S$

a	b	c
1	2	3
7	8	9

图 1-9 交运算

$R - S$

a	b	c
4	5	6

图 1-10 差运算

(4) 广义笛卡尔积。设 R 是一个包含 m 个元组的 j 元关系, S 是一个包含 n 个元组的 k 元关系, 则 R 和 S 的广义笛卡尔积是一个包含 $m \times n$ 个元组的 $j+k$ 元关系, 记作 $R \times S$, 并定义

$$R \times S = \{(r_1, r_2, \dots, r_j, s_1, s_2, \dots, s_k) \mid (r_1, r_2, \dots, r_j) \in R \text{ 且 } \{s_1, s_2, \dots, s_k\} \in S\}$$

即 $R \times S$ 的每个元组的前 j 个分量是 R 中的一个元组, 而后 k 个分量是 S 中的一个元组。

例如, 已知关系 R 和关系 S, 如图 1-11 所示。

关系 R			关系 S		
A	B	C	E	F	D
a1	b2	c1	e1	f2	d2
a2	b1	c3	e2	f3	d1
a3	b3	c2	e3	f1	d3

图 1-11 关系 R 和关系 S(二)

笛卡尔积结果如图 1-12 所示。

A	B	C	E	F	D
a1	b2	c1	e1	f2	d2
a1	b2	c1	e2	f3	d1
a1	b2	c2	e3	f1	d3
a2	b1	c3	e1	f2	d2
a2	b1	c3	e2	f3	d1
a2	b1	c3	e3	f1	d3
a3	b3	c2	e1	f2	d2
a3	b3	c2	e2	f3	d1
a1	b3	c2	e3	f1	d3

图 1-12 笛卡尔积结果

2. 专门的关系运算

对于关系数据库,已经有了结构化查询语言(structured query language, SQL),它对表具有很强的操纵能力。在多数关系数据库管理系统中除了支持 SQL 外,自身也提供了许多操作表的功能,不同关系数据库管理系统提供的功能可能有些差异,但它们检索数据的操作都是以选择、投影和连接 3 种关系为核心的。

1) 选择操作(selection)

设 $R = \{(a_1, a_2, \dots, a_n)\}$ 是一个 n 元关系, F 是关于 (a_1, a_2, \dots, a_n) 的一个条件, R 中所有满足 F 条件的元组组成的子关系称为 R 的一个选择,记作 $\sigma F(R)$,并定义

$$\sigma F(R) = \{(a_1, a_2, \dots, a_n) \mid (a_1, a_2, \dots, a_n) \in R \text{ 且 } (a_1, a_2, \dots, a_n) \text{ 满足条件 } F\}$$

简言之,对 R 关系按一定规则筛选一个子集的过程就是对 R 施加了一次选择运算。

选择操作是从表中选取满足某种条件的元组(记录)进行操作。通常在命令中加上条件子句和逻辑表达式来完成选择操作。

例如,从图 1-6 所示的“学生表”中选出“男”同学,结果如图 1-13 所示。

学号	姓名	性别	出生日期	民族	籍贯	入学成绩	有否奖学金	所属学院
20190001	潘石屹	男	2000/2/16	汉族	江苏	585.00	<input type="checkbox"/>	商学院
20190002	王石	男	2000/4/25	汉族	湖北	575.00	<input type="checkbox"/>	外国语学院
20190003	蒋锡培	男	1999/10/8	汉族	江苏	598.00	<input checked="" type="checkbox"/>	物理学院
20190006	叶茂中	男	1999/9/16	汉族	广东	580.00	<input type="checkbox"/>	商学院

图 1-13 选择运算结果

2) 投影操作(projection)

设 $R = R(A_1, A_2, \dots, A_n)$ 是一个 n 元关系, $\{i_1, i_2, \dots, i_m\}$ 是 $\{1, 2, \dots, n\}$ 的一个子集,并且 $i_1 < i_2 < \dots < i_m$, 定义

$$\pi(R) = R_1(A_{i_1}, A_{i_2}, \dots, A_{i_m})$$

通俗地讲,关系 R 上的投影是从 R 中选择出若干属性列组成新的关系。

投影是从表中选取若干列进行操作。选取列时不受表中列顺序的约束,可按实际需要安排各列顺序。通常在命令中加上要选取的各个列名称来完成投影操作。

例如,显示“学生表”中的“学号”“姓名”“性别”,结果如图 1-14 所示。

学号	姓名	性别
20190001	潘石屹	男
20190002	王石	男
20190003	蒋锡培	男
20190004	张丽敏	女
20190005	孙建芹	女
20190006	叶茂中	男
20190007	王冉	女
20190008	熊晓鸽	女

图 1-14 投影运算结果

3) 连接操作(join)

连接是从两个关系的笛卡尔积中选取属性间满足一定条件的元组,记作 $R \bowtie_{A \theta B} S$ 。

其中 A 和 B 分别为 R 和 S 上维数相等且可比的属性组, θ 是比较运算符。连接运算从 R 和 S 的笛卡尔积 $R \times S$ 中选取 $(R$ 关系)在 A 属性组上的值与 $(S$ 关系)在 B 属性组上的值满足比较关系 θ 的元组。

连接操作是对两张表进行连接,同时生成一张新表,新表中所有的列是被连接的两张表中列的并集或是该并集的子集,新表中包含的元组(记录)是满足连接条件的所有元组(记录)集合。连接运算有等值连接和自然连接两种。连接条件中的运算符为比较运算符,当运算符取“=”时为等值连接,而自然连接是去掉重复属性的等值连接。

例如,连接教学管理中的“学生表”“选课表”,显示学生选课情况,结果如图 1-15 所示。

学号	姓名	课程编号	成绩
20190001	潘石屹	501001	87
20190001	潘石屹	701001	88
20190001	潘石屹	801002	91
20190002	王石	501001	76
20190002	王石	701001	68
20190003	蒋锡培	701001	77
20190003	蒋锡培	501001	78
20190004	张丽敏	501001	95
20190005	孙建芹	701001	77
20190005	孙建芹	501001	91
20190006	叶茂中	701001	76
20190006	叶茂中	501001	69
20190006	叶茂中	801001	94
20190006	叶茂中	801002	76
20190007	王冉	501001	65
20190008	熊晓鸽	501001	86
20190008	熊晓鸽	901001	74
20190008	熊晓鸽	901002	88

图 1-15 连接运算结果

3. 关系的完整性

关系模型的完整性规则是对关系的某种约束条件。关系模型中有 3 类完整性约束:实体完

完整性规则、用户定义的完整性规则和参照完整性规则。

1) 实体完整性规则

关系的主键可以标识关系中的每条记录,而关系的实体完整性要求关系中的记录不允许出现两条记录的主键值相同,既不能有空值,也不能有重复值。实体完整性规则规定关系的所有主属性都不能为空值,而不是整体不能为空值。

例如,学生表(学号、姓名、性别、出生日期、民族)中,“学号”为主关键字,则“学号”都不能取空值,而不是整体不能为空。

2) 用户定义的完整性规则

不同的关系数据库系统根据其应用环境的不同,通常需要针对某一具体字段设置约束条件。

例如,性别字段的取值只能是“男”或“女”。

3) 参照完整性规则

参照完整性是相关联的两个表之间的约束。对于具有主从关系的两个表来说,从表中每条记录对应的外键值必须是主表中存在的,如果在两个表之间建立了关联关系,则对一个关系进行的操作要影响到另一个表中的记录。

例如,在学生表和选课表之间用“学号”建立了关联关系,学生表是主表,选课表是从表,那么在向从表添加新记录时,系统要检查新记录的“学号”是否在主表(学生表)中已存在,如果存在则允许执行输入操作,否则拒绝输入。

1.5 数据库设计

数据库设计是指对于一个给定的应用环境,构造优化的数据库逻辑模式和物理结构,并据此建立数据库及其应用系统,使之能够有效地存储和管理数据,满足各种用户的应用需求,包括信息管理要求和数据操作要求。

为了合理地组织数据,应遵从以下基本设计原则:

- (1) 关系数据库的设计应遵从概念单一化的原则。
- (2) 避免在表之间出现重复字段。
- (3) 表中的字段必须是原始数据和基本数据元素。
- (4) 用外部关键字保证有关联的表之间的联系。

1.5.1 数据库设计的基本步骤

考虑数据库及其应用系统开发全过程,可以将数据库设计分为6个阶段:需求分析、概念设计、逻辑设计、物理设计、数据库实施、数据库运行和维护。

1. 需求分析阶段

需求分析的任务是通过详细调查现实世界要处理的对象(组织、部门、行业等),充分了解用户单位目前的工作状况,明确用户的各种需求,然后在此基础上确定新系统的功能。

2. 概念设计阶段

将需求分析得到的用户需求抽象为信息结构,即概念模型的过程就是概念设计,它是整个数

数据库设计的关键。

在需求分析阶段所得到的应用需求应该首先抽象为概念模型,以便更好、更准确地用某一数据库管理系统实现这些需求。

概念模型是各种逻辑模型的共同基础,它比逻辑模型更独立于机器、更抽象,从而更加稳定。描述概念模型的有效工具是 E-R 图。

3. 逻辑设计阶段

数据库逻辑设计是将概念模型转换为逻辑模型,也就是被某个数据库管理系统所支持的数据模型,并对转换结果进行规范化处理。关系数据库的逻辑结构由一组关系模式组成,因而从概念模型结构到关系数据库逻辑结构的转换就是将 E-R 图转换为关系模型的过程。

4. 物理设计阶段

数据库在物理设备上的存储结构与存取方法称为数据库的物理结构,它依赖于给定的计算机系统。为一个给定的逻辑模型选取一个最适合应用要求的物理结构的过程,就是数据库的物理设计。

5. 数据库实施阶段

完成数据库的物理设计之后,就要用数据库管理系统提供的数据库定义语言和其他实用程序将数据库逻辑设计和物理设计结果严格地描述出来,而成为数据库管理系统可以接收的源代码,再经过调试产生目标代码,然后就可以组织数据入库了,这就是数据库实施阶段。

数据库实施阶段包括两项重要的工作:一是数据的载入,二是应用程序的编码和调试。

6. 数据库运行和维护阶段

数据库系统经过试运行合格后,数据库开发工作就基本完成,即可投入正式运行了。在数据库系统的运行过程中,对数据库设计进行评价、调整、修改等维护工作是一个长期的任务,也是设计工作的继续和提高。

在数据库运行阶段,对数据库经常性的维护工作主要是由数据库管理员完成的,它包括数据库的转储和恢复、数据库的安全性及完整性控制、数据库性能的分析及改造、数据库的重组与重构造。

1.5.2 E-R 图向关系模型的转换

1. 要解决的问题

- (1)如何将实体和实体间的联系转换为关系模型。
- (2)如何确定关系模型的属性和码。

2. 转换内容

将 E-R 图转换为关系模型:将实体、实体的属性和实体间的联系转换为关系模式。

3. 一般原则

将一个实体转换为一个关系模式,关系的属性就是实体的属性,关系的码就是实体的码。

4. 独立实体转换为关系模型

识别主码属性,唯一标识一个实体,用下划线标示出来,一般设置为编号。学生实体如图 1-16 所示。

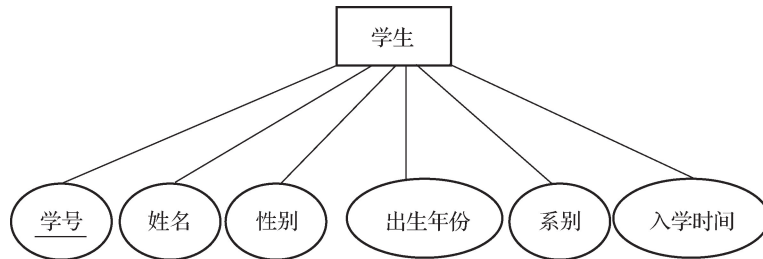


图 1-16 学生实体

学生实体转换为关系模型为：

学生(学号,姓名,性别,出生年份,系别,入学时间)

5. 实体间的联系的不同情况

(1)一个 1:1 联系的两个实体可以各自转换为独立的关系模型,也可以与任意一端对应的关系模型合并。1:1 的 E-R 图如图 1-17 所示。

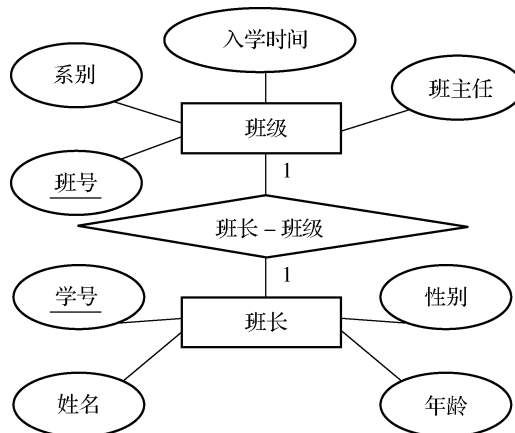


图 1-17 1:1 的 E-R 图

将 1:1 联系的两个实体各自转换为独立的关系模型,加入对方的主码属性。转换后的关系模型为：

班级(班号,系别,班主任,入学时间,班长学号)

班长(学号,姓名,性别,年龄)

或

班级(班号,系别,班主任,入学时间)

班长(学号,姓名,性别,年龄,班号)

(2)一个 1:n 联系可以转换为一个独立的关系模型,也可以与 n 端对应的关系模型合并。

1:n 的 E-R 图如图 1-18 所示。

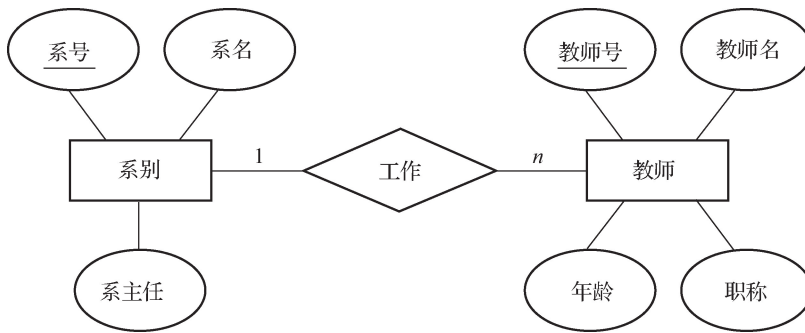


图 1-18 1:n 的 E-R 图

将 1:n 联系转化为关系模型,在多方加入 1 方的主码属性。转换后的关系模型为:

系别(系号,系名,系主任)

教师(教师号,教师名,年龄,职称,系号)

(3)一个 $m:n$ 联系可以各自转换为一个独立的关系模型,必须为联系产生一个新的关系。该关系中至少包含被它所联系的双方实体的主码,若联系有属性,也要并入该关系中。新关系的主码可以为双方实体的主码的组合,也可以另外设置一个属性作为主码。 $m:n$ 的 E-R 图如图 1-19 所示。

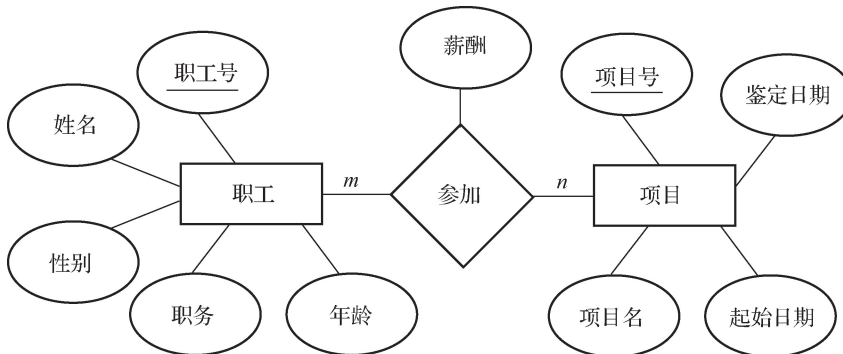


图 1-19 m:n 的 E-R 图

将 $m:n$ 联系转换为关系模型。转换后的关系模型为:

职工(职工号,姓名,性别,年龄,职务)

项目(项目号,项目名,起始日期,鉴定日期)

参加(职工号,项目号,薪酬)

1.5.3 数据库设计实例

某电子商务网站要求提供以下服务:可随时查询库存中现有物品的名称、数量和单价,所有物品均应由物品编号唯一标识;可随时查询顾客订货情况,包括顾客编号、顾客名、所订物品编号、订购数量、联系方式、交货地点,所有顾客编号不重复;当需要时,可通过数据库中保存的供应商名称、电话、邮编与地址信息向相应的供应商订货,一个编号货物只由 1 个供应商供货。

根据以上要求,完成如下任务:根据语义设计出 E-R 图,将该 E-R 图转换为的一组等价的关系

模型,并标出各关系模型的主码。

第一步,根据以上描述内容,得到 E-R 图,如图 1-20 所示。

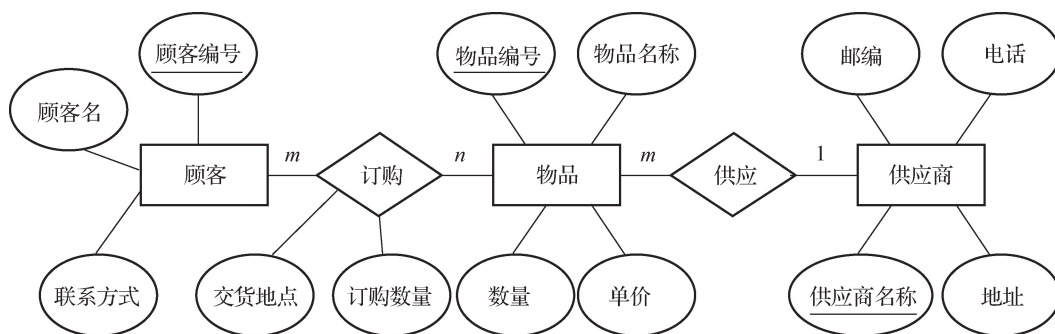


图 1-20 E-R 图

第二步,根据 E-R 图得到关系模型为:

顾客(顾客编号,顾客名,联系方式)

物品(物品编号,物品名称,数量,单价,供应商名称)

供应商(供应商名称,地址,电话,邮编)

订购(顾客编号,物品编号,订购数量,交货地点)

1.6 Access 2010 操作基础

Access 诞生于 20 世纪 90 年代初期,历经多次升级改版,其功能越来越强,而操作越来越直观、方便。

2010 年 6 月,Microsoft Office 2010 正式在中国发布,这是 Microsoft 公司推出的新一代办公软件,Microsoft Access 2010 是其中的重要组件。

Access 2010 是 Access 的较新版本,与原来的版本相比,Access 2010 除了继承和发扬以前版本功能强大、界面友好、操作方便等优点外,在界面的易操作性方面、数据库操作与应用方面进行了很大改进。

1.6.1 Access 2010 的安装、启动与退出

Access 2010 是 Microsoft 公司 Office 2010 办公套件中的一款数据库管理系统软件,在安装 Office 2010 时通过选中 Access 2010 选项即可轻松地将其安装在系统中。

1. Access 2010 的安装

通过执行 Microsoft Office 2010 安装程序 setup.exe 文件来启动安装过程,然后按照系统提示,逐步进行操作即可。安装完成后,就可以正常使用 Access 2010。

2. Access 2010 的启动

(1)使用“开始”菜单启动 Access 2010。

(2)使用快捷方式启动 Access 2010。

(3)使用已有的数据库文件启动 Access 2010。

启动 Access 2010 之后,屏幕显示 Access 2010 的启动窗口,其也称为 Microsoft Office Backstage 视图,如图 1-21 所示。但使用第三种方法,即双击 Access 2010 数据库文件图标启动 Access 2010,进入的界面是 Access 2010 主窗口。

整个 Access 2010 初始界面可以分为 3 部分,左侧是文件操作区,在其中列出了各种针对文件设立的常用操作,用户可以根据自身的需要选择不同的操作来完成指定任务。模板的中间区域是“可用模板”区域,在其中用户可以选择创建空数据库或者使用不同模板创建数据库。在创建新数据库时,模板的右侧用来设置新数据库保存位置及定义新数据库的名称。在查看信息时,用来查看和编辑数据库属性。



图 1-21 Access 2010 启动窗口

3. Access 2010 的退出

- (1)在 Access 2010 窗口中,执行“文件”→“退出”命令。
- (2)单击 Access 2010 窗口右上角的“关闭”按钮。
- (3)双击 Access 2010 窗口左上角的控制菜单图标;或单击控制菜单图标,从打开的菜单中选择“关闭”命令;或按组合键 Alt+F4。
- (4)右击 Access 2010 窗口标题栏,在弹出的快捷菜单中选择“关闭”命令。

1.6.2 Access 2010 工作窗口

1. Access 2010 启动窗口

1) “文件”菜单命令

“文件”菜单命令是对数据库文件进行各种操作及对数据库进行设置的命令。

2) 快速访问工具栏

快速访问工具栏中的命令始终可见,可将最常用的命令添加到此工具栏中。通过快速访问工具栏,只需一次单击即可访问命令。默认的快速访问工具栏包括“保存”“恢复”和“撤销”命令。

可以自定义快速访问工具栏,以便将经常使用的命令加入其中。还可以选择显示该工具栏的位置和最小化功能区。

单击快速访问工具栏右侧的下拉箭头,将弹出“自定义快速访问工具栏”菜单,选择“其他命令”菜单项,弹出“Access 选项”对话框,默认显示“自定义快速访问工具栏”界面,如图 1-22 所示。在其中选择要添加的命令,然后单击“添加”按钮。若要删除命令,在右侧的列表框中选择该命令,然后单击“删除”按钮。也可以在列表框中双击该命令实现添加或删除。完成后单击“确定”按钮。



图 1-22 “自定义快速访问工具栏”界面

也可以执行“文件”→“选项”命令,然后在弹出的“Access 选项”对话框的左侧窗格中选择“快速访问工具栏”选项,进入“自定义快速访问工具栏”界面。

2. Access 2010 主窗口

在 Access 2010 的启动窗口,提供了创建数据库的导航,当选择新建空数据库、空白 Web 数据库或选择某种模板后,就正式进入 Access 2010 主窗口,如图 1-23 所示。

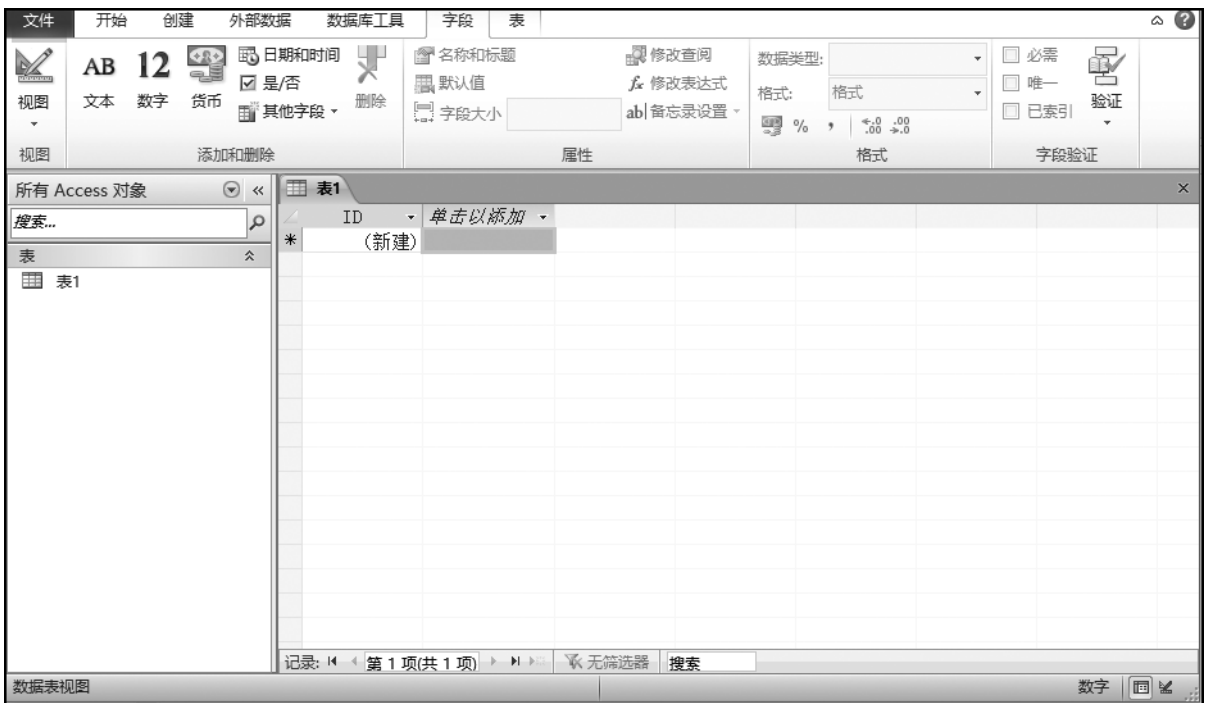


图 1-23 Access 2010 主窗口

1) 功能区

Access 2010 中的新用户界面由多个元素构成,这些元素定义了用户与产品的交互方式。

最重要的新界面元素称为功能区,它是 Microsoft Office 用户界面的一部分。功能区是一个横跨程序窗口顶部的条形带,其中包含多组命令。

功能区是一个存放各种命令的单一位置,同时也是菜单和工具栏的主要替代部分。功能区中有多个选项卡,这些选项卡按照合理的方式将选项组组合在一起。

(1)功能区的组成。Access 2010 功能区是一个横跨 Access 2010 主窗口顶部的带状区域,它由选项卡、选项组和各组的命令按钮 3 部分组成。单击选项卡可以打开此选项卡所包含的选项组,显示各组相应的命令按钮。

在 Access 2010 中,主要的选项卡包括“文件”“开始”“创建”“外部数据”和“数据库工具”,每个选项卡都包含多组相关命令。

①“文件”选项卡包括 3 部分,左侧包含“保存”“对象另存为”“数据库另存为”“打开”“关闭数据库”“信息”“最近所用文件”“新建”“打印”“保存并发布”“帮助”“选项”“退出”等多个选项;中间区域是“可用模板”区域,在其中用户可以选择创建空数据库或者使用相应的模板创建数据库;右侧在创建新数据库时,用来设置新数据库保存位置及定义新数据库名称,在查看信息时,用来查看和编辑数据库属性,如图 1-24 所示。



图 1-24 “文件”选项卡

②“开始”选项卡包括视图、剪贴板、排序和筛选、记录、查找、文本格式、中文简繁转换 7 个选项组,用户可以在“开始”选项卡中对 Access 2010 进行数据复制粘贴、修改字体和字号、数据排序等操作,如图 1-25 所示。



图 1-25 “开始”选项卡

③“创建”选项卡包括模板、表格、查询、窗体、报表、宏与代码 6 个选项组,“创建”选项卡中包含的命令主要用于创建 Access 2010 的各种元素,如图 1-26 所示。

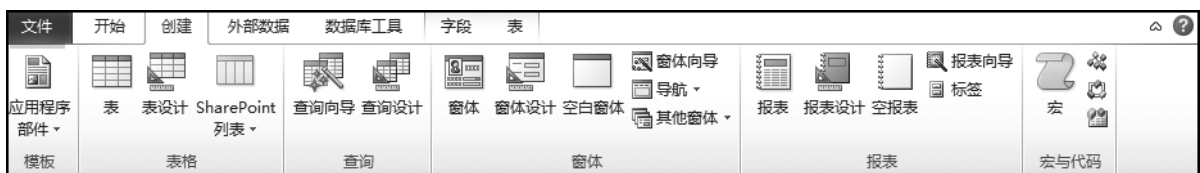


图 1-26 “创建”选项卡

④“外部数据”选项卡包括导入并链接、导出、收集数据 3 个选项组，在“外部数据”选项卡中主要对 Access 2010 以外的数据进行相关处理，如图 1-27 所示。



图 1-27 “外部数据”选项卡

⑤“数据库工具”选项卡包括工具、宏、关系、分析、移动数据、加载项 6 个选项组，主要针对 Access 2010 数据库进行比较高级的操作，如图 1-28 所示。



图 1-28 “数据库工具”选项卡

除了上述选项卡之外，还有一些隐藏的选项卡默认没有显示。只有在进行特定操作时，相关的选项卡才会显示出来。例如，在对已有表进行操作时，会自动打开“表”选项卡和“字段”选项卡，在利用设计视图创建表时会出现“设计”选项卡。

有些选项组的右下角有一个对话框启动器按钮，单击该按钮可以打开相应的对话框或任务窗格。

除标准选项卡之外，Access 2010 还有上下文选项卡，即根据正在进行操作的对象及正在执行的操作的不同而在标准选项卡旁边出现的选项卡。

上下文选项卡可以根据所选对象状态的不同而自动显示或关闭，具有智能特性，给用户的操作带来很大方便。

(2)功能区的操作。在 Access 2010 中，执行命令的方法有多种。一般可以选择功能区选项卡，再在相关选项组中单击相关命令按钮。也可以使用与命令关联的键盘快捷方式，如果用户知道早期 Access 版本中所用的键盘快捷方式，那么也可以在 Access 2010 中使用此快捷方式。此外，按下并释放 Alt 键，将显示命令的访问键，此时按下所提示的键也可以执行相应的命令。

2) 导航窗格

(1)导航窗格的组成。导航窗格按类别和组对数据库对象进行组织。可以从多种组织选项中进行选择，还可以在导航窗格中创建用户的自定义组织方案。在默认情况下，新数据库使用“对象类型”类别，该类别包含对应于各种数据库对象的组。“对象类型”类别组织数据库对象的方式与早期版本中的默认“数据库窗口”显示界面相似。

(2)打开数据库对象。若要打开数据库对象，则在导航窗格中双击该对象，或在导航窗格中选择对象，然后按 Enter 键。或在导航窗格中右击对象，再在弹出的快捷菜单中选择菜单命令，该快捷菜单中的命令因对象类型而不同。

(3)显示或隐藏导航窗格。单击导航窗格右上角的“百叶窗开/关”按钮,将隐藏导航窗格。若要再显示导航窗格,则再单击导航窗格条右上角的“百叶窗开/关”按钮。

3)其他界面元素

(1)对象编辑区。对象编辑区位于 Access 2010 主窗口的右下方、导航窗格的右侧,是用来设计、编辑、修改,以及显示表、查询、窗体和报表等数据库对象的区域。对象编辑区的最下面是记录定位器,其中显示共有多少条记录,当前编辑的是第几条。

(2)选项卡式文档。启动 Access 2010 后,可以用选项卡式文档代替原来 Access 版本中的重叠窗口来显示数据库对象。单击选项卡中不同的对象名称,可以切换到不同的对象编辑界面。右击选项卡,将弹出快捷菜单,选择其中的相应命令可以实现对当前数据库对象的各种操作,如保存、关闭、视图切换等。

(3)状态栏。状态栏是位于 Access 2010 主窗口底部的条形区域。右侧是各种视图切换按钮,单击各个按钮可以快速切换视图状态,左侧显示了当前视图状态。

(4)获取帮助。在使用 Access 2010 的过程中,如有疑问,可以按 F1 键或单击功能区右侧的问号按钮来获取帮助。在进入 Access 帮助界面后,可以根据目录或关键字来查找帮助信息。还可以执行“文件”→“帮助”命令,单击按钮进入相关的帮助界面。

1.6.3 Access 2010 数据库的组成

Access 2010 将数据库定义为一个扩展名为 accdb 的文件,并包括 6 种不同的对象,它们是表(table)、查询(query)、窗体(form)、报表(report)、宏(macro)和模块(module)。

1. 表

表又称数据表,它是数据库的核心与基础,用于存放数据库中的全部数据。

2. 查询

查询是按照一定的条件从一个或多个表中筛选出所需要的数据而形成的一个动态数据集,并在一个虚拟的数据表窗口中显示出来。

3. 窗体

窗体是数据库和用户联系的界面。

4. 报表

利用报表可以将数据库中需要的数据提取出来进行分析、整理和计算,并将数据以格式化的方式打印输出。

5. 宏

宏是一系列操作命令的集合,其中的每个操作命令都能实现特定的功能。

6. 模块

模块是用 VBA 语言编写的程序段,使用模块对象可以完成宏不能完成的复杂任务。

习题 1

一、填空题

1. 支持数据库系统的 3 种数据模型是_____、_____、_____。
2. 关系中的某个属性组,它可以唯一标识一个元组,该属性组称为_____。
3. 数据库系统的核心组成部分是_____。
4. 两个实体间的联系有_____、_____和_____类型。
5. 在关系数据库中,一个属性的取值范围为_____。
6. 二维表中的列称为关系的_____,二维表中的行称为关系的_____。

二、选择题

1. DBMS 是()。
A. 数据库
B. 数据库管理系统
C. 数据库系统
D. 数据库处理系统
2. 在关系数据库系统中,所谓的“关系”是指一个()。
A. 表
B. 文件
C. 二维表
D. 实件
3. 数据库的对象不包括()。
A. 表、查询
B. 宏、模块
C. 窗体、报表
D. 编程语言
4. 下列关于数据库系统的叙述,正确的是()。
A. 数据库系统只是比文件系统管理的数据更多
B. 数据库系统中数据的一致性是指数据类型一致
C. 数据库系统避免了数据冗余
D. 数据库系统减少了数据冗余
5. 下列关于实体完整性的叙述,正确的是()。
A. 实体完整性由用户来维护
B. 关系的主键可以有重复值
C. 主键不能取空值
D. 空值即空字符串
6. 数据库(DB)、数据库系统(DBS)、数据库管理系统(DBMS)三者之间的关系是()。
A. DBS 包括 DB 和 DBMS
B. DBMS 包括 DB 和 DBS
C. DB 包括 DBS 和 DBMS
D. DBS 就是 DB,也就是 DBMS
7. 关系模型支持的 3 种基本运算是()。
A. 选择、投影、连接
B. 选择、查询、连接
C. 投影、编辑、选择
D. 投影、选择、索引
8. Access 是一个()。
A. 数据库文件系统
B. 数据库系统
C. 数据库管理系统
D. 数据库应用系统

9. 层次型、网状型和关系型数据库的划分原则是()。
- A. 记录长度
B. 文件的大小
C. 联系的复杂程度
D. 数据之间的联系方式
10. 在关系型数据库中要显示“学生表”中的姓名和性别信息,应采用的关系运算是()。
- A. 选择
B. 投影
C. 连接
D. 关联

三、简答题

1. 数据库管理技术经历了哪几个发展阶段? 各阶段的主要特征是什么?
2. 数据库管理系统的主要功能是什么?
3. 目前常用的数据模型有哪几种? 它们的主要特征是什么?
4. 什么是关系数据库? 其主要特点有哪些?
5. 关系运算包括哪些运算方式?
6. 关系完整性约束有哪几种? 其作用是什么?
7. 数据库设计的步骤是什么?