

第1章 数据库技术基础

本章主要讲述数据库技术基础知识和相关概念,包括信息、数据和数据处理,数据管理技术的发展历史,数据库系统的基本概念,数据库的数据模式结构,数据库管理系统的功能等。

1.1 信息、数据和数据处理

信息和数据是两个比较容易混淆的概念,真正了解它们的含义很有必要。

1. 信息

信息是现实世界中各种事物(包括有生命的和无生命的、有形的和无形的)的存在方式、运动形态,以及它们之间的相互联系等诸多要素在人脑中的反映,是通过人脑抽象后形成的概念。人们不仅可以认识和理解这些概念,还可以对它们进行推理、加工和传播。信息甚至可为某种目的提供某种决策依据。例如,根据当前的天气进行天气预报。信息有多种表现形式,它可以通过手势、眼神、声音、图像等方式表达。

2. 数据

数据是信息的载体,是信息的一种符号化表示。符号是人为规定的,数据通过能书写的符号表示信息。数据的概念包括两方面的含义:一是数据的内容是信息;二是数据的表现形式是符号。数据不仅可以是数值,而且可以是文字、图形、动画、声音、视频等。由于数据能够被记录、存储和处理,因而可以从数据中挖掘出更深层次的信息。例如,用表1-1描述某高校计算机系学生的基本信息。

表1-1 某高校计算机系学生的基本信息

学 号	姓名	性别	出生日期	籍贯	所在院系	专业
202113020125	李梦瑶	女	2002-3-11	上海	计算机	软件工程
202113020127	王泽铭	男	2002-5-23	北京	计算机	网络工程
.....						

表1-1中的一行组成一条记录,这些符号被赋予了特定的语义,具体描述了一条学生信息,具有表示信息的功能。数据具有以下两个基本特征。

1) 数据有“型”和“值”之分

数据的型是指对某一类数据的结构和属性的描述,数据的值是型的一个具体值。

例如,表 1-1 中的学生信息是由“学号”“姓名”“性别”等数据项构成的。表 1-1 的栏目(学号、姓名、性别、出生日期、籍贯、所在院系、专业)可以看作学生数据的型;表 1-1 栏目下的数据(202113020125、李梦瑶、女、2002-3-11、上海、计算机、软件工程)等,可以看作学生数据型的值。

2) 数据有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。数据类型不同,则数据的表示形式、存储方式以及能进行的操作运算也各不相同。常见数据类型有字符型、数值型、日期型等。数据的取值范围也称数据的值域,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。

例如,表 1-1 中学生性别的值域是{“男”,“女”}。在使用计算机处理数据时,应该为数据选择合适的类型和值域。

3. 信息与数据的关系

信息与数据既有联系又有区别。数据是承载信息的物理符号,或称为载体,而信息是数据的内涵。同一信息可以有不同的数据表示方式。例如,描述某人的生日,可以用文字、照片、录像等多种形式作载体。

数据可以表示信息,但不是任何数据都能明确表达信息,同一数据也可以有不同的解释。例如,2021 可以理解为一个数值,也可以理解为 2021 年。

所以,数据和信息两个概念不能混为一谈,数据不等于信息,数据只是信息表达方式中的一种。正确的数据可表达信息,而虚假、错误的数字所表达的是谬误,不是信息。

4. 数据处理

数据处理与用数据表示信息不同。数据处理是指对各种形式的数据进行收集、存储、传播和加工,直至产生新信息输出的全过程。数据处理的目的是从大量、已知的原始数据出发,抽取和导出对人们有价值的、新的信息。

例如,学生的“出生日期”是基本特征之一,属于原始数据,而“年龄”是当前年份与出生日期相减而得到的数字,具有相对性,可视为二次数据。

5. 数据管理

数据管理是指数据的收集、整理、组织、存储、查询、维护和传送等各种操作。数据处理任务的矛盾焦点不是计算,而是如何把数据管理好。数据库技术正是瞄准这一目标而逐渐完善起来的一门计算机软件技术。

数据处理和数据管理是相互联系的,数据管理中各种操作都是数据处理业务必不可少的基本环节,数据管理技术的好坏,直接影响到数据处理的效率。

1.2

数据管理技术的发展历史

所谓数据管理,是指对数据进行收集、整理、存储、检索、加工和传递等一系列活动的总和。

1.2.1 手工管理阶段

手工管理阶段是指计算机诞生的初期,即20世纪50年代之前。这个时期的计算机主要用于科学计算。从硬件来看,外存只有纸带、卡片、磁带,没有直接存取的储存设备;从软件来看,那时还没有操作系统,没有管理数据的软件。数据的组织和管理完全由程序员手工完成,因此称为手工管理阶段。在手工管理阶段,应用程序与数据之间的关系是一一对应的。

1.2.2 文件系统阶段

20世纪50年代后期至60年代中期,在硬件方面,外存储器有了磁盘、磁鼓等直接存取的存储设备;在软件方面,操作系统中已经有了专门用于管理数据的软件,称为文件系统。在文件系统阶段,文件系统把数据组织成文件形式,存储在磁盘上,这些数据文件相互独立,长期保存在存储设备上。文件可以命名,应用程序利用“按文件名访问,按记录进行存取”的方式,对文件中的数据进行修改、插入和删除操作。文件系统阶段的数据还是面向应用程序的,数据文件基本上与各自的应用程序相对应。

1.2.3 数据库系统阶段

从20世纪60年代中后期开始,计算机数据管理技术进入数据库系统阶段。与文件系统不同的是,数据库系统是面向数据的而不是面向程序的,各个处理功能通过数据管理软件从数据库中获取所需要的数据和存储处理结果。它克服了文件系统的弱点,为用户提供了一种方便的、功能强大的数据管理手段。

1.3 数据库系统的基本概念

数据库系统作为信息系统的核心和基础,涉及一些常用的术语和基本概念。

1. 数据库管理系统

数据库管理系统(database management system, DBMS)安装于操作系统之上,是一个管理、控制数据库中各种数据库对象的系统软件。数据库管理系统能够为数据库提供数据的定义、建立、维护、查询和统计等操作功能,并完成对数据完整性、安全性进行控制的功能。

数据库管理系统的目标是让用户能够更方便、更有效、更可靠地建立数据库和使用数据库中的信息资源。数据库管理系统不是应用软件,它不能直接用于诸如工资管理等事务管理工作。数据库管理系统是为设计数据管理应用项目提供的计算机软件,利用数据库管理系统设计事务管理系统。现今广泛使用的数据库管理系统有微软公司的 Microsoft SQL Server、甲骨文公司的 Oracle 和 MySQL 以及 IBM 公司的 DB2 和 Informix 等,其中,Oracle、DB2 属于大型 DBMS,Microsoft SQL Server、MySQL、Informix 属于中型 DBMS。

2. 数据库

数据库(database,DB)是按照一定数据结构来存储和管理数据的计算机软件系统,是用数据库管理系统定义的,是长期存储在计算机内的、可共享的大量数据的集合。概括起来说,数据库具有永久存储、有组织和可共享三个基本特点。

3. 数据库应用系统

使用数据库技术管理数据的系统都称为数据库应用系统(database application system)。一个数据库应用系统应携带有较大的数据量,否则它就不需要数据库管理。数据库应用系统的应用非常广泛,它可以用于事务管理、计算机辅助设计、计算机图形分析和处理及人工智能等系统中,即所有数据量大、数据成分复杂的地方,都可以使用数据库技术进行数据管理工作。

4. 数据库系统

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。数据库系统应包括计算机硬件、操作系统、数据库管理系统、数据库、数据库应用系统和数据库管理员,即由计算机硬件、软件和使用人员构成。数据库的建应、使用和维护等工作只靠一个DBMS是不够的,还需要专业人员协助完成。DBS简化表示如下:

DBS=计算机系统(硬件、软件平台、人)+DBMS+DB

数据库系统包含了数据库、DBMS、软件平台与硬件支撑环境及各类人员;DBMS在操作系统(operating system,OS)的支持下,对数据库进行管理与维护,并提供用户对数据库的操作接口。一般在不引起混淆的情况下,常常把数据库系统直接简称为数据库。

1.4 数据库的数据模式结构

数据有“型”和“值”之分。模式是对数据逻辑结构和特征的描述,它仅为型的描述,不涉及具体的值。模式的一个具体值称为模式的一个实例。模式反映的是数据的结构及其联系,而实例反映的是数据某一时刻的状态。因此,数据模式是稳定的,而实例是在不断变化和更新的。

1.4.1 数据库系统的三级模式结构和优点

为了有效地组织和管理数据,提高数据库系统的逻辑独立性和物理独立性,数据库采用三级模式结构来组织和管理数据。数据库的三级模式结构由外模式、模式和内模式三级模式构成,它们分别代表了看待数据库的3个不同角度。在三级模式之间还提供了二级映像,即外模式/模式映像和模式/内模式映像,以保证数据的逻辑和物理独立性。数据库系统的三级模式结构如图1-1所示。

1. 数据库系统的三级模式结构

(1) 模式

模式也称逻辑模式、概念模式,是数据库中所有数据的逻辑结构和特征的描述。它通常以某种数据模型为基础,定义数据库全部数据的逻辑结构。例如,数据记录的名称,数据项的名称,类型和值域等。另外,它还要定义数据项之间的联系,不同记录之间的联系,以及与数据有关的安全性和完整性。

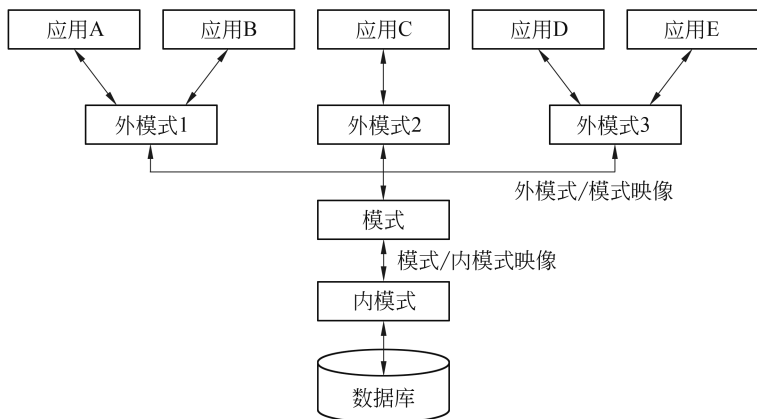


图 1-1 数据库系统的三级模式结构

模式是数据库系统模式结构的中间层。模式与硬件和软件环境无关,也不与任何计算机语言有关。一个数据库只有一个逻辑模式,是数据的逻辑表示,即描述数据库中存储具体的数据及其之间存在的联系。

(2) 外模式

外模式也称子模式、用户模式,是局部数据的逻辑结构和特征的描述。外模式是模式的子集,一个数据库可以有多个外模式,是各个用户的数据视图。不同用户的外模式也是不同的,但不同用户的外模式可以相互覆盖。外模式是数据库安全的一个有力保障措施,每个用户只能访问与外模式中对应的数据。

外模式完全是根据用户自己对数据的需求,站在局部的角度进行设计的。由于一个数据库系统有多个用户,因此就可能有多个数据外模式。由于外模式是面向用户和应用程序的,所以它被称为用户数据视图。从逻辑关系上来看,外模式是模式的一个逻辑子集,从一个模式可以推导出多个不同的外模式。

(3) 内模式

内模式也称存储模式、物理模式,是数据物理结构和存储方式的描述,是数据在数据库内部的表达方式,对应于实际存储在外存储介质上的数据库。例如,记录的存储方式是顺序存储还是链式存储,数据是否压缩、加密存储等。一个数据库只有一个内模式。

2. 三级模式结构的优点

三级模式结构是数据领域的标准结构,是数据库实现数据逻辑独立性和物理独立性的基础。将外模式和模式分开来保证数据的逻辑独立性;将模式和内模式分开来实现数据的物理独立性。三级模式结构主要有以下3个优点:

- (1) 有利于数据的安全性;
- (2) 有利于数据共享,减少数据冗余;
- (3) 简化了用户接口。

1.4.2 数据库的二级映像

在数据库系统三级模式结构的基础上,DBMS在三级模式之间提供了二级映像来保证

数据的物理独立性和逻辑独立性。

1. 外模式/模式映像

外模式/模式映像定义了外模式和模式之间的对应关系。外模式描述数据的局部逻辑结构,模式描述数据的全局逻辑结构。数据库中的同一模式可以有多个外模式,对于每个外模式,都存在一个外模式/模式映像。

外模式和模式之间的对应关系称为映像。这些映像确定了数据的局部逻辑结构与全局逻辑结构之间的对应关系。当模式改变时(如增加新关系、新属性等),由数据库管理系统对这个外模式/模式映像进行相应的改变,可保证外模式不变。应用程序是根据数据的外模式编写的,外模式没有变,则应用程序不需要修改,保证了数据的逻辑独立性(数据与程序之间的逻辑独立性)。

2. 模式/内模式映像

模式/内模式映像定义了数据的全局逻辑结构与存储结构之间的对应关系。数据库中的模式和内模式都只有一个,因此模式/内模式映像也是唯一的。应用程序依赖于数据的外模式,独立于数据的模式和存储结构。当数据库的存储结构发生变化时,只需要数据管理员对模式/内模式映像进行相应的改变,就可以使模式保持不变,从而保证用户程序不需要改变,保证了数据的物理独立性。

1.5 数据库管理系统的功能

在计算机系统中,对数据的管理是通过 DBMS 和数据库实现的。DBMS 是提供建立、管理、维护和控制数据库功能的一组计算机软件。DBMS 的目标是使用户能够科学地组织和存储数据,能够从数据库中高效地获得需要的数据,能够方便地处理数据。DBMS 能够提供以下几个方面的主要功能。

1. 数据定义功能

DBMS 提供数据定义语言(data definition language, DDL),对系统中的数据及结构进行定义,DBMS 根据其定义执行建库、表、视图、索引、存储过程等操作。

2. 数据操纵功能

DBMS 提供数据操纵语言(data manipulation language, DML)。用户可以使用 DML 操纵数据,实现对数据库的基本操作,如插入、修改、删除、查询、统计等数据存取操作。

3. 数据库的运行管理

数据库在建立、运用和维护时由 DBMS 统一管理、统一控制,以保证数据的安全性、完整性和多用户对数据库使用的并发控制及发生故障后的系统恢复等。数据库的运行管理功能是 DBMS 的核心功能。

4. 数据库的建立和维护功能

数据库的建立和维护包括初始数据的输入、转换,数据库的转储、恢复,数据库的重组织和性能检测分析等功能。上述功能主要由一些实用软件或管理工具完成。

5. 其他功能

其他功能主要包括 DBMS 与其他软件系统的数据通信功能、不同 DBMS 或文件系统的数据转换功能以及异构数据库之间的互访和互操作功能等。

1.6 信息的三种世界及描述

计算机系统不能直接处理现实世界中的客观事物,只有将它们数据化后,计算机才能处理。将现实世界中的信息转换为数据库中的数据,不可能一步到位,通常分为三个阶段,称为三种世界,即现实世界、信息世界和计算机世界(也称数据世界)。

1.6.1 现实世界

现实世界是指我们要管理的客观存在的各种事物、事物之间的相互联系及事物的发生、变化过程。现实世界通过实体、特征、实体集及联系进行划分和认识。

1. 实体

现实世界中存在的可以相互区分的事物或概念称为实体。实体可以分为事物实体和概念实体。例如,一个学生、一个工人、一台机器、一部汽车等是事物实体,一门课、一个班级等称为概念实体。

2. 实体的特征

每个实体都有自己的特征,利用实体的特征可以区别不同的实体。例如,学生通过姓名、性别、年龄、身高、体重等许多特征来描述自己。尽管实体具有许多特征,但是在研究时,只选择其中对管理及处理有用的或有意义的特征。例如,对于人事管理,职工的特征可选择姓名、性别、年龄、工资、职务等。现实世界就是通过每个实体所特有的特征来相互区分的。

3. 实体集及实体集之间的联系

具有相同特征或能用同样特征描述的实体的集合称为实体集。例如,学生、课程、汽车、计算机等都是实体集。实体集不是孤立存在的,实体集之间有着各种各样的联系,例如学生和课程之间有“选课”联系,教师和所在系之间有“工作”联系。

1.6.2 信息世界

信息世界是现实世界在人们头脑中的反映,人们以现实世界为基础,用思维对事物进行认识、选择、命名、分类等抽象工作之后,并用文字符号表示出来,从而得到了信息。当用信息来描述事物时,就形成了信息世界。信息世界对现实世界的抽象重点在于构造数据结构。信息世界主要涉及以下3个概念。

1. 实例

实体通过其特征(属性)的表示称为实例。实例与现实世界的实体相对应。实体与实例是不同的,例如,张三是一个实体,而“张三,男,25岁,计算机系学生”是实例。现实世界中

的张三除了姓名、性别、年龄和所在系外还有其他的特征,而实例仅对需要的特征通过属性进行了描述。

2. 属性

实体的特征在人们思想意识中形成的知识称为属性。一个实例可能拥有多个属性,其中能唯一标识实例的属性或属性集合称为码。每个属性的取值是有范围的,称为该属性的域。属性与现实世界的特征相对应。例如,学生李四有学号、姓名、性别、出生日期等属性。其中学号能唯一标识该学生,则学号就是该学生实例的码。性别的取值不是男就是女,则该属性的域就是(男,女)。

3. 对象及对象间联系

同类实例的集合称为对象,对象即实体集中的实体用属性表示,得出的信息集合。实体集之间的联系用对象间联系表示。对象及对象间联系与现实世界的实体集及实体集间的联系相对应。例如,所有学生实例的集合就是学生对象,即全体学生。每个学生之间都可能发生联系,例如,同班的学生,班干部和普通学生之间有管理联系。

按用户的观点对现实世界的抽象,即对现实世界的数据库信息建模就称为概念模型(也称信息模型)。信息世界通过概念模型以及过程模型、状态模型反映现实世界,它要求对现实世界中的事物之间的联系和事物的变化情况能准确、如实、全面地表示。

1.6.3 计算机世界

计算机世界又称数据世界,是将信息世界中的信息经过抽象和组织,按照特定的数据结构,即数据模型,将数据存储在计算机中。数据模型应符合具体的计算机系统和DBMS的要求。计算机世界主要涉及以下5个概念。

1. 数据项

数据项是对象属性的数据表示,数据项与信息世界的属性相对应。数据项有型和值之分,数据项的型是对数据特性的表示,它通过数据项的名称、数据类型、数据宽度和值域等来描述;数据项的值是其具体取值。数据项的型和值都要符合计算机数据的编码要求,即都要符合数据的编码要求。

2. 记录

记录是实例的数据表示。记录有型和值之分:记录的型是结构,由数据项的型构成;记录的值表示对象中的一个实例,它的分量是数据项值。例如,“姓名,性别,年龄,所在系”是学生数据的记录型;而“张三,男,23,计算机系”是一个学生的记录值,它表示学生对象的一个实例,“张三”“男”“23”“计算机系”都是数据项值。

3. 文件

文件是对象的数据表示,是同类记录的集合,即同一个文件中的记录类型应是一样的。例如,将所有学生的登记表组成一个学生数据文件,文件中的每条记录都要按“姓名,性别,年龄,所在系”的结构组织数据项值。文件与信息世界中的对象相对应。文件的存储形式有很多种,如顺序文件、链接文件、索引文件等。

4. 文件集

文件集是若干文件的集合,即由计算机操作系统通过文件系统来组织和管理。它与信息世界中的对象集相对应。文件与文件之间是有联系的。文件系统通过对文件、目录、磁盘的管理,可以对文件的存储空间、读写权限等进行管理。

5. 数据模型

现实世界中的事物反映到计算机世界中就形成了文件的记录结构和记录,事物之间的相互联系就形成了不同文件间的记录联系。记录结构及其记录联系的数据化的结果就是数据模型。

1.6.4 三种世界的转换及关系

1. 三种世界的术语对应关系

现实世界、信息世界和计算机世界这 3 个领域是由客观到认识,由认识到使用,由使用到管理的 3 个不同层次,后一领域是前一领域的抽象描述。每种世界都有自己对象的概念描述,但是它们之间又相互对应。信息的三种世界之间术语的对应关系如表 1-2 所示。

表 1-2 信息的三种世界术语的对应关系

现实世界	信息世界	计算机世界
实体	实例	记录
特征	属性	数据项
实体集	对象	数据或文件
实体间联系	对象间联系	文件集(数据间的联系)
—	概念模型	数据模型

2. 三种世界的转换关系

现实世界、信息世界和计算机世界的转换关系如图 1-2 所示。

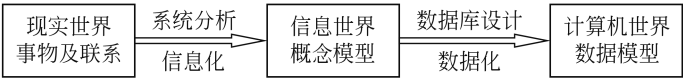


图 1-2 信息的三种世界之间的转换关系

从图 1-2 中可以看出,人们首先将现实世界的事物及联系抽象为概念模型,然后将概念模型经过数据化处理转换为数据模型。也就是说,首先将现实世界中客观存在的事物及联系抽象为某一种信息结构,这种结构并不依赖计算机系统,是人们认识的概念模型,这个过程由数据库设计人员完成;然后将概念模型转换为计算机上某一具体的 DBMS 支持的数据模型,则成为计算机世界的数据库,这个过程由数据库设计人员和数据库设计工具共同完成。

1.7

概念模型

概念模型是对信息世界的建模,概念模型应当能够全面、准确地描述出信息世界中的基本概念。在把现实世界抽象为信息世界的过程中,实际上是抽象出现实系统中有应用价值的元素及其关联。这时所形成的信息结构就是概念模型。这种信息结构不依赖具体的计算机系统。

1.7.1 概念模型的基本概念

概念模型用于信息世界的建模,是对现实世界的抽象和概括,是现实世界到信息世界的第一层抽象,是数据库设计人员进行数据库设计的有力工具,也是数据库设计人员和用户之间进行交流的语言。

1. 对象和实例

现实世界中,具有相同性质,服从相同规则的一类事物(或概念,即实体)的抽象称为对象,对象是实体集信息化(数据化)的结果。对象中的每一个具体的实体被抽象为该对象的实例。

2. 属性

属性是对象的某一方面特征的抽象表示。例如,学生可以通过学生的姓名、学号、性别、年龄等特征来描述。此时,姓名、学号、性别、年龄等就是学生的属性。属性值是属性的具体取值。例如某一学生,其姓名为李梦瑶,学号为20125,性别为女,年龄为21,这些具体描述就称为属性值。

3. 码、主码和次码

码也称关键字,它能够唯一标识一个实体。码可以是属性或属性组,如果码是属性组,则其中不能含有多余的属性。例如在学生的属性集中,学号确定后,学生的其他属性值也都确定了,学生记录也就确定了。由于学号可以唯一地标识一个学生,所以学号为码。在有些实体集中,可以有多个码。例如学生实体集,假设学生姓名没有重名,那么属性“姓名”也可以作为码。当一个实体集中包括多个码时,通常要选定其中的一个码为主码,其他的码就是候选码。

实体集中不能唯一标识实体属性的叫次码。例如,年龄、性别这些属性都是次码。一个主码值对应一个实例,而一个次码值会对应多个实例。

4. 域

属性的取值范围称为属性的域。例如,学生的年龄为16~35内的正整数,其数据的域为(16~35);性别的域为(男,女);姓名的域为字符串集合,学院名称的域为学校所有学院名称的集合。

5. 实体型和实体集

在许多教科书中,还有实体型和实体集的概念。

具有相同属性的实体具有共同的特征和性质,用实体名及其属性名集合来抽象和刻画

的同类实体称为实体型。例如,学生(学号,姓名,性别,出生日期,学院名称)是一个实体型。相同类型的实体集合称为实体集。例如,全体学生就是一个实体集。对象是实体集按其实体型抽象的结果。

6. 实体联系的类型

现实世界的事物之间是有联系的,这种联系必然要在信息世界中加以反映。这些联系在信息世界中反映为实体(型)内部的联系和实体(型)之间的联系。实体(型)内部的联系主要表现在组成实体的属性之间的联系。实体(型)之间的联系主要表现在不同实体集之间的联系。

1) 两个实体集之间的联系

两个实体集之间的联系可概括为以下3种。

(1) 一对一联系(1:1)。设有两个实体集A和B,如果实体集A与实体集B之间具有一对一联系,则对于实体集A中的每一个实体,在实体集B中最多有一个(也可以没有)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也最多有一个实体与之联系。两实体集间的一对一联系记作1:1。

【例1-1】 一个学校只能有一位校长,一位校长也只能在一个学校任职,所以学校与校长之间的联系即为一对一的联系。还有总经理与公司之间也都是一对一的联系。

(2) 一对多联系(1:n)。设有两个实体集A和B,如果实体集A与实体集B之间具有一对多联系,则对于实体集A的每一个实体,实体集B中有一个或多个实体与之联系;而对于实体集B的每一个实体,实体集A中至多有一个实体与之联系。实体集A与实体集B之间的一对多联系记作1:n。

【例1-2】 一个学校里有多名学生,而每名学生只能在一个学校里学习,则学校与学生之间具有一对多联系。还有公司和职工、球队和球员之间也都是一对多的联系。

(3) 多对多联系(m:n)。设有两个实体集A和B,如果实体集A与实体集B之间具有多对多联系,则对于实体集A的每一个实体,实体集B中有一个或多个实体与之联系;反之,对于实体集B中的每一个实体,实体集A中也有一个或多个实体与之联系。实体集A与实体集B之间的多对多联系记作m:n。

【例1-3】 一名学生可以选修多门课程,一门课程可以被多名学生选修,所以学生和课程之间的联系即为多对多的联系;一名教师教过许多学生,一名学生也被许多老师教过,教师和学生之间的联系也是多对多的联系。

一对一联系是一对多联系的特例,而一对多联系又是多对多联系的特例。用图形可以表示两个实体集之间的1:1、1:n或m:n联系,如图1-3所示。

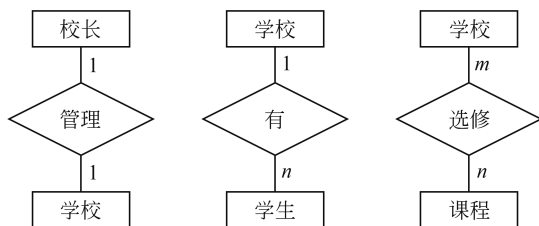


图1-3 两个实体集联系的样子

2) 多实体集之间的联系

两个以上的实体集之间也会存在联系,其联系类型一般为一对多和多对多。

(1) 多实体集之间的一对多联系。设实体集 $E_1, E_2, \dots, E_n$, 如果 $E_j (j=1, 2, \dots, n)$ 与其他实体集 $E_1, E_2, \dots, E_{j-1}, E_{j+1}, \dots, E_n$ 之间存在有一对多的联系, 则 E_j 中的一个给定实体, 可以与其他实体集 $E_i (i \neq j)$ 中的一个或多个实体联系, 而实体集 $E_i (i \neq j)$ 中的一个实体最多只能与 E_j 中的一个实体联系。

【例 1-4】在图 1-4(a) 中, 一门课程可以由若干教师讲授, 一名教师只讲授一门课程; 一门课程使用若干本参考书, 每一本参考书只供一门课程使用。所以课程与教师、参考书之间的联系是一对多的。

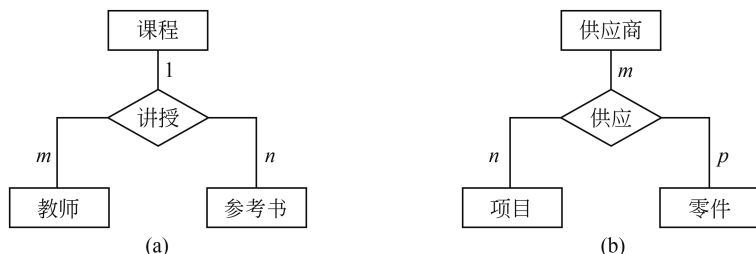


图 1-4 三个实体集联系的实例

(2) 多实体集之间的多对多联系。在两个以上的实体集之间, 当一个实体集与其他实体集之间均存在多对多联系, 而其他实体集之间没有联系时, 这种联系称为多实体集间的多对多联系。

【例 1-5】有三个实体集: 供应商、项目、零件, 一个供应商可以供给多个项目多种零件; 每个项目可以使用多个供应商供应的零件; 每种零件可由不同供应商供给。因此, 供应商、项目、零件三个实体型之间是多对多的联系, 如图 1-4(b) 所示。

3) 实体集内部的联系

在一个实体集中的实体之间也可以存在一对多或多对多的联系。

【例 1-6】职工是一个实体集, 职工中有领导, 而领导自身也是职工。职工实体集内部具有领导与被领导的联系, 即某一个职工领导若干名职工, 而一个职工仅被一个领导所管, 这种联系是一对多的联系, 如图 1-5 所示。

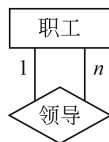


图 1-5 同一实体集内的一对多联系实例

1.7.2 概念模型的表示方法

概念模型的表示方法有很多, 其中最常用的是实体—联系法, 简称 E-R 图法。该方法用 E-R 图来描述现实世界的概念模型, E-R 图也称为 E-R 模型。E-R 模型是抽象和描述现实世界的有力工具, 是各种数据模型的基础。

1. E-R 图中的图示

E-R 图提供了表示实体、实体的属性以及实体之间(或内部)联系的方法。在 E-R 图中, 用长方形、椭圆形、菱形分别表示实体集、属性、联系, 联系上标注联系类型。

1) 实体集

实体集用长方形表示,并在长方形中标注实体集名。

【例 1-7】 教师、课程、学生实体集如图 1-6 所示。

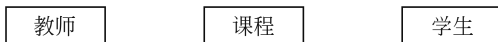


图 1-6 实体集

2) 实体集的属性

实体集的属性用椭圆形表示,并在椭圆中标注属性名,再用线段将该属性与对应实体集连接起来。在多个属性中,如果有一个(组)属性可以唯一表示一个实体,则可以在该属性下边画出下划线,用来标识该属性,即主属性,也就是主码。

由于实体集的属性比较多,有些实体可具有多达上百个属性,所以在 E-R 图中,实体集的属性可不直接画出,而通过数据字典的方式表示(即文字说明方式)。无论使用哪种方法表示实体集的属性,都不能出现遗漏属性的情况。

【例 1-8】 学生实体集有学号、姓名、性别、出生日期、学院名称属性,其中学号为主属性。课程实体集有课程号、课程名、学分属性,其中课程号为主属性,如图 1-7 所示。

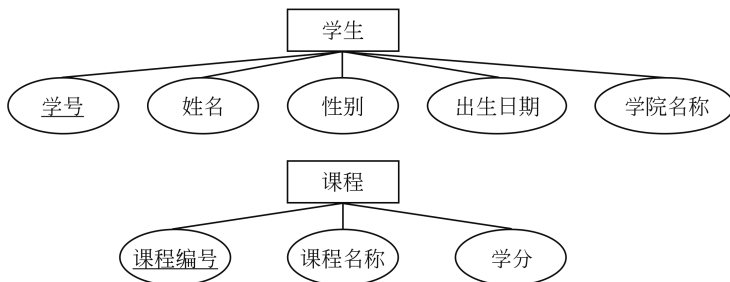


图 1-7 学生、课程实体集及属性

3) 实体集间的联系

实体集间的联系用菱形表示。即在菱形中标注联系名,再用线段将该联系与联系实体集连接起来,同时在线段旁标注联系的类型。如果联系具有属性,则该属性仍用椭圆框表示,需要用线段将属性与其联系连接起来。联系的属性必须在 E-R 图上标出,不能通过数据字典说明。通常,如果实体之间有同名属性,并且同名属性表示的含义也相同,则实体之间有联系。

【例 1-9】 学生实体集与课程实体集之间存在联系。因为每门课程都有许多个学生选修,但一个学生也可以选修多门课程,所以课程实体集和学生实体集之间有联系,联系类型为 $m:n$ (即多对多),如图 1-8 所示。

如果一个 E-R 图中的实体比较多,实体的属性也比较多。为了使 E-R 图简单明了,可以先分别绘制各实体的 E-R 图,最后只将所有实体联系起来。

2. E-R 图设计原则与设计步骤

1) E-R 图设计原则

(1) 属性应该存在于且只存在于某一个地方(实体集或者联系)。该原则确保了数据库

中的某个数据只存储于某个数据库表中(避免同一数据存储于多个数据库表),避免了数据冗余。

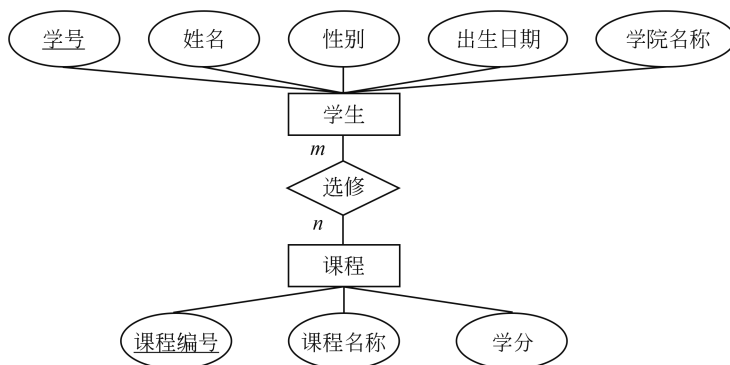


图 1-8 学生实体集与课程实体集之间的联系

(2) 实体集是一个独立的对象,不能存在于另一个实体集中,成为其属性。该原则确保了一个数据库表中不能包含另一个数据库表,即不能出现“表中套表”的现象。

(3) 同一个实体集在同一个E-R图中仅出现一次。例如,同一个E-R图中的两个实体集间存在多种关系时,为了表示实体集间的多种关系,不要让同一个实体集出现多次。

2) E-R图设计步骤

- (1) 划分和确定实体集。
- (2) 划分和确定联系。
- (3) 确定属性。属性不能再有需要描述的性质或与其他实体集有联系。
- (4) 重复前面步骤,找出所有实体集、关系集、属性和属值集,然后绘制E-R图。
- (5) 优化E-R模型。对E-R图进行优化,消除实体集间冗余的联系及属性,形成基本的E-R模型。

在绘制E-R图时,可以先绘制部分E-R分图,在此基础上综合各E-R分图,形成E-R总图。

1.8 数据模型

虽然概念模型不依赖计算机系统,但现实世界的的数据最终还是要存放到计算机的数据库中。这时就需要将概念模型转化为与具体计算机数据库相关的数据模型。

1.8.1 数据模型的基本概念

数据模型是严格定义的一组概念的集合,这些概念精确地描述了系统的静态和动态特性,是数据库中用来对现实世界进行抽象的工具,是数据库系统的核心与基础,是描述数据的结构以及定义在其上的操作和约束条件。

数据模型是一组严格定义的概念集合,这些概念精确地描述了系统的数据结构、数据操作和数据完整性约束条件,称为数据模型的三要素。

1. 数据结构

数据结构是所研究的对象类型的集合。这些对象是数据库的组成成分,它们包括两类:一类是与数据类型、内容、性质有关的对象;另一类是与数据之间联系有关的对象。根据对象元素之间关系的不同特性,常用的基本结构有树形结构、图状结构(或网状结构)、关系结构等。在数据库系统中,通常按照数据结构的类型来命名数据模型,例如,层次结构、网状结构和关系结构的数据模型分别被命名为层次模型、网状模型和关系模型。

2. 数据操作

数据操作是指对数据库中各种数据对象允许执行的操作集合。数据操作包括操作对象和有关的操作规则两部分。数据库中的数据操作主要有数据检索和数据更新(即插入、删除或修改数据的操作)两大类操作。数据模型必须对数据库中的全部数据操作进行定义,指明每项数据操作的确切含义、操作对象、操作符号、操作规则以及对操作的语言约束等。数据操作是对系统动态特性的描述。

3. 数据完整性约束条件

数据约束条件是一组数据完整性规则的集合。数据完整性规则是指数据模型中的数据及其联系所具有的制约和依存规则。数据约束条件用以限定符合数据模型的数据库状态以及状态的变化,以保证数据库中数据的正确、有效和相容。每种数据模型都规定有基本的完整性约束条件,这些完整性约束条件要求所属的数据模型都应满足。同理,每个数据模型还规定了特殊的完整性约束条件,以满足具体应用的要求。例如,在关系模型中,基本的完整性约束条件是实体完整性和参照完整性,特殊的完整性条件是用户定义的完整性。

1.8.2 常用的数据模型

当前,数据库领域最常用的数据模型主要有三种,它们是层次模型、网状模型和关系模型。

1. 层次模型

用树状结构来表示实体及实体之间联系的模型称为层次模型。层次数据库系统采用层次模型作为数据的组织方式。层次模型是数据库系统中最早出现的数据模型。层次数据库系统的典型代表是IBM公司的IMS数据库管理系统。

2. 网状模型

现实世界中,许多事物之间的联系是非层次结构的,它们需要使用网状模型表示。用网状结构来表示实体及实体之间联系的模型称为网状模型,网状数据库系统是采用网状模型作为数据组织方式的数据库系统,如HP公司的IMAGE、Univac公司的DMS1100、Honeywell公司的IDS/2等。

3. 关系模型

关系模型是三种模型中最重要的一种。关系数据库系统采用关系模型作为数据的组织方式,现在流行的数据库系统大多是关系数据库系统。

1.9 习题1

一、选择题

1. 在数据管理技术发展的三个阶段中,数据共享最好的是()。
A. 人工管理阶段 B. 文件系统阶段
C. 数据库系统阶段 D. 三个阶段相同
2. 在数据库系统的三级模式结构中,面向某个或某几个用户的数据视图是()。
A. 外模式 B. 模式
C. 内模式 D. 概念模式
3. 数据库是存储在计算机上的()相关数据集合。
A. 结构化的 B. 特定业务
C. 具体文件 D. 其他
4. DBS的中文含义是()。
A. 数据库系统 B. 数据库管理员
C. 数据库管理系统 D. 数据定义语言
5. 数据库管理系统是()。
A. 操作系统的一部分 B. 在操作系统支持下的系统软件
C. 一种编译系统 D. 一种操作系统
6. 数据库、数据库管理系统和数据库系统三者之间的关系是()。
A. 数据库包括数据库管理系统和数据库系统
B. 数据库系统包括数据库和数据库管理系统
C. 数据库管理系统包括数据库和数据库系统
D. 不能相互包括
7. 信息的3种世界是指现实世界、信息世界和()世界。
A. 计算机 B. 虚拟 C. 物理 D. 理想
8. 所谓概念模型,就是()。
A. 客观存在的事物及其相互联系
B. 将信息世界中的信息数据化
C. 实体模型在计算机中的数据化表示
D. 现实世界到机器世界的一个中间层次,即信息世界
9. 下列()不能称为实体。
A. 班级 B. 手机 C. 图书 D. 姓名
10. 绘制E-R图的3个基本要素是()。
A. 实体、属性、关键字 B. 属性、数据类型、实体
C. 属性、实体、联系 D. 约束、属性、实体
11. 一间宿舍可住多名学生,则实体宿舍和学生之间的联系是()。
A. 一对一 B. 一对多 C. 多对一 D. 多对多

12. 数据库中,用来抽象表示现实世界中数据和信息的工具是()。

- A. 数据模型 B. 数据定义语言 C. 关系范式 D. 数据表

二、简答题

1. 什么是信息? 什么是数据? 什么是数据处理?
2. 什么是数据库管理系统? 什么是数据库? 什么是数据库应用系统? 什么是数据库系统?
3. 什么是数据库系统的三级模式结构? 并画图表示。
4. 信息有哪三种世界? 分别具有什么特点? 它们之间有什么联系?
5. 什么是概念模型?
6. 实体的联系有哪三种?
7. 解释概念模型中常用的概念:实体、属性、码、域、实体型、实体集、联系。
8. 绘制 E-R 图,描述学生选课,包括学生(学号,姓名,性别,出生年月)、班级(班级编号,班级名称,所在学院,所属专业,入学年份)和课程(课程编号,课程名称,课程学时,课程学分),学生选修课程后得到平时成绩和期末成绩。