

ch7 数据库设计和E-R模型

7.1 设计过程概览

包括数据库模式，设计访问，更新数据，设计控制数据访问的安全模式

设计阶段

完整刻画用户需求

概念设计：定义数据库中的实体、实体的属性、实体之间的联系、实体和联系上的约束

功能需求规格说明

逻辑设计→物理设计

设计选择

实体：表示可以明确识别的个体。（rg，教师，课程，学生）

需要避免出现的问题：

1，冗余：多次、不必要。

2，不完整：例如只设置开课的表，没有设置已存在课表。

7.2 实体-联系模型

实体集

实体：现实世界中可以区别于其他对象的一个“食物”或者“对象”。例如大学的每一个人。

属性：实体集中每个成员所拥有的描述性性质。实体通过一组属性表示。

实体集：相同类型/具有相同属性的一个实体集合。（不必互不相交）

联系集

联系：多个实体之间互相关联。

联系集：相同类型联系的集合（也可有属性）

考虑图 7-1 中的两个实体集 *instructor* 和 *student*。我们定义联系集 *advisor* 来表示教师及学生之间的关联。这一关联如图 7-2 所示。

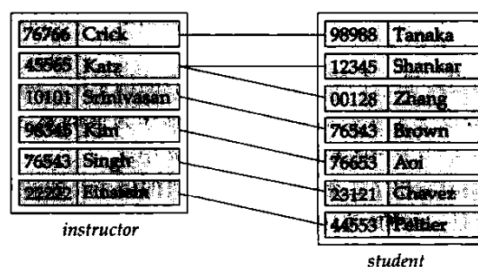


图 7-2 联系集 *advisor*

实体集和联系集之间是“参与”关系，实体集在联系中扮演的功能称为实体的“角色”。

参与联系集的实体集的数目叫做“联系集的度”。

属性

域/值集：属性的取值集合。

每一个实体可以用一组（属性，数据值）表示。

简单属性：不可再划分的属性。

复杂属性：可以再（多层）划分的属性（例如name可以划分成first_name.last_name）

单值属性：实体只能取一个值。

多值属性：实体可以取很多值例如一个人有很多电话号码。可以用{}表示。可以设定上下界。

派生属性：从别的属性（或实体）派生而来。例如我喜欢的男生的数量由我喜欢的男生这个实体集决定。值不存储，只是在需要的时候计算出来。

7.3约束

映射约束

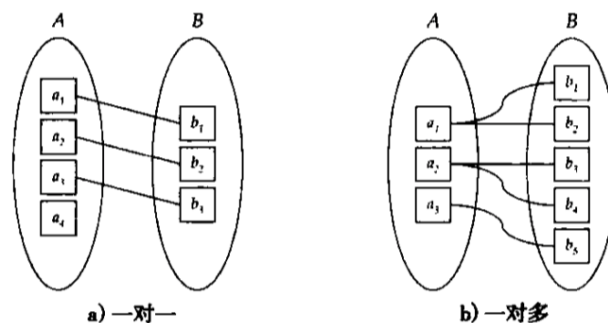


图 7-5 映射基数

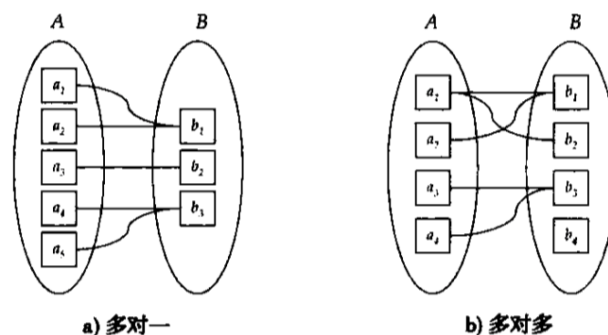


图 7-6 映射基数

参与约束

如果实体集 E 中的每个实体都参与到联系集 R 的至少一个联系中，实体集 E 在联系集 R 中的参与称为**全部(total)**的。如果 E 中只有部分实体参与到 R 的联系中，实体集 E 到联系集 R 的参与称为**部分(partial)**的。在图 7-5a 中， B 在联系集中的参与是全部的，而 A 在联系集中的参与是部分的；在图 7-5b 中， A 和 B 在联系集中的参与都是全部的。

码

码：足以区分每个实体的属性。（实体集的主码）

如果 E_1, E_2, E_3 参与了联系集 R ，那么 R 的超码是三者的主码的并集。其结构也依赖于映射基数。

7.4从实体中删除冗余属性

例子：

- 实体集 *instructor* 包含属性 *ID*、*name*、*dept_name* 以及 *salary*，其中 *ID* 构成主码。
 - 实体集 *department* 包含属性 *dept_name*、*building* 以及 *budget*，其中 *dept_name* 构成主码。
- 我们用关联 *instructor* 和 *department* 的联系集 *inst_dept* 对每个教师都有一个关联的系的情况建模。
- 属性 *dept_name* 在两个实体集中都出现了。由于它是实体集 *department* 的主码，因此它在实体集 *instructor* 中是冗余的，需要将其移除。

7.5实体-联系图

基本结构

- 分成两部分的**矩形**代表实体集。本书中有阴影的第一部分包含实体集的名字，第二部分包含实体集中所有属性的名字。
- **菱形**代表联系集。
- **未分割的矩形**代表联系集的属性。构成主码的属性以下划线标明。
- **线段**将实体集连接到联系集。
- **虚线**将联系集属性连接到联系集。
- **双线**显示实体在联系集中的参与度。
- **双菱形**代表连接到弱实体集的标志性联系集（我们将在 7.5.6 节讲述标志性联系集和弱实体集）。

映射基数

- **一对一**：我们从联系集 *advisor* 向实体集 *instructor* 和 *student* 各画一个箭头（见图 7-9a）。这表示一名教师可以指导至多一名学生，并且一名学生可以有至多一位导师。
- **一对多**：我们从联系集 *advisor* 画一个箭头到实体集 *instructor*，以及一条线段到实体集 *student*（见图 7-9b）。这表示一名教师可以指导多名学生，但一名学生可以有至多一位导师。
- **多对一**：我们从联系集 *advisor* 画一条线段到实体集 *instructor*，以及一个箭头到实体集 *student*。这表示一名教师可以指导至多一名学生，但一名学生可以有多位导师。
- **多对多**：我们从联系集 *advisor* 向实体集 *instructor* 和 *student* 各画一条线段（见图 7-9c）。这表示一名教师可以指导多名学生，并且一名学生可以有多位导师。

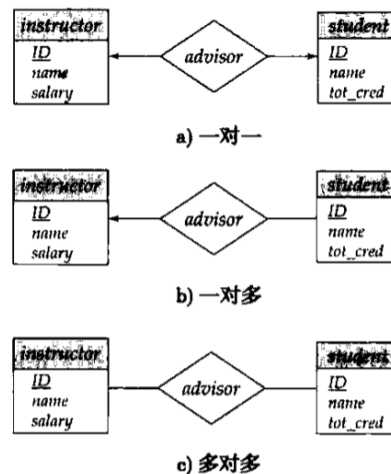


图 7-9 提供了三种解决两个实体集与联系集中的联系的方法。有箭头的为 1，没有箭头的为多。

用 $x..y$ 表示最大最小映射基数，* 表示没有限制

例如，考虑图 7-10，在 *advisor* 和 *student* 之间的边有 $1..1$ 的基数约束，意味着基数的最小值和最大值都是 1。也就是，每个学生必须有且仅有一个导师。从 *advisor* 到 *instructor* 边上的约束 $0..*$ 说明教师可以有零个或多个学生。因此，*advisor* 联系是从 *instructor* 到 *student* 的一对多联系，更进一步地讲，*student* 在 *advisor* 联系中的参与是全部的，表示一个学生必须有一个导师。

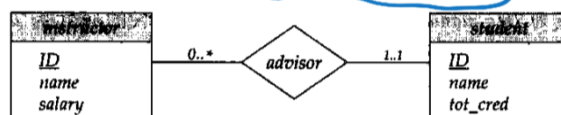


图 7-10 联系集上的基数约束

复杂的属性

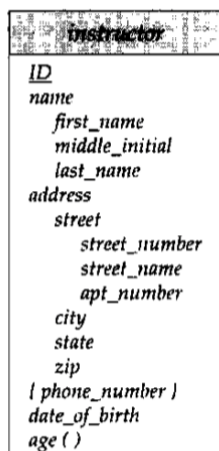


图 7-11 包含复合、多值和派生属性的 E-R 图

{phone_number}为多值属性

age()为派生属性

角色

在菱形和矩形连线上标注来表示角色（实体在联系中扮演的功能）

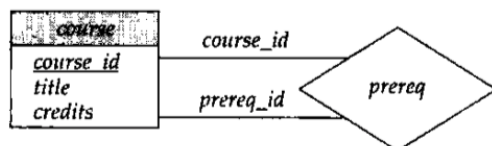


图 7-12 包含角色标识的 E-R 图

非二元的联系集

在一个联系集外我们至多允许一个箭头，因为在一个非二元的联系集外包含两个或更多箭头的 E-R 图可以用两种方法解释。假设实体集 A_1, A_2, \dots, A_n 之间有联系集 R ，并且只有指向实体集 $A_{i+1}, A_{i+2}, \dots, A_n$ 的边是箭头。那么，两种可能的解释为：

- 来自 A_1, A_2, \dots, A_i 的实体的一个特定组合可以和至多一个来自 $A_{i+1}, A_{i+2}, \dots, A_n$ 的实体组合相关联。因而联系 R 的主码可以用 A_1, A_2, \dots, A_i 的主码的并集来构造。
- 对每个实体集 $A_k, i < k \leq n$ ，来自其他实体集的实体的每个组合可以和来自 A_k 的至多一个实体相关联。于是对于 $i < k \leq n$ ，每个集合 $\{A_1, A_2, \dots, A_{k-1}, A_{k+1}, \dots, A_n\}$ 构成一个候选码。

若关系中的一个属性或属性组的值能够唯一地标识一个元组，且他的真子集不能唯一的标识一个元组，则称这个属性或属性组做候选码。

弱实体集

没有足够的属性形成主码的实体集（有主码的实体集叫强实体集）。

每个弱实体集要和“标识”或者“属实体集”关联。

标识性联系是从弱实体集到标识实体集多对一的，并且弱实体集在联系中的参与是全部的。标识性联系集不应该有任何描述性属性，因为这种属性中的任意一个都可以与弱实体集相关联。

在我们的例子中，*section* 的标识实体集是 *course*，将 *section* 实体和它们对应的 *course* 实体关联在一起的联系 *sec_course* 是标识性联系。

虽然弱实体集没有主码，但是我们仍需要区分依赖于特定强实体集的弱实体集中的实体的方法。弱实体集的**分辨符** (discriminator) 是使得我们进行这种区分的属性集合。例如，弱实体集 *section* 的分辨符由属性 *sec_id*、*year* 以及 *semester* 组成，因为对每门课程来说，这个属性集唯一标识了这门课程的一次开课。弱实体集的分辨符也称为该实体集的**部分码**。

弱实体集的主码由标识实体集的主码加上该弱实体集的分辨符构成。在实体集 *section* 的例子中，它的主码是 {*course_id*, *sec_id*, *year*, *semester*}，其中 *course_id* 是标识实体集 *course* 的主码，{*sec_id*, *year*, *semester*} 区分同一门课程的不同 *section* 实体。

弱实体集的图片标识：

在 E-R 图中，弱实体集和强实体集类似，以矩形表示，但是有两点主要的区别：

- 弱实体集的分辨符以虚下划线标明，而不是实线。
- 关联弱实体集和标识性强实体集的联系集以双菱形表示。

在图 7-14 中，弱实体集 *section* 通过联系集 *sec_course* 依赖于强实体集 *course*。



图 7-14 包含弱实体集的 E-R 图

某些时候数据库设计者会把弱实体集表示为主实体集的多值复合属性（如果弱实体集属性不多，这样做更恰当）

大学的 E-R 图

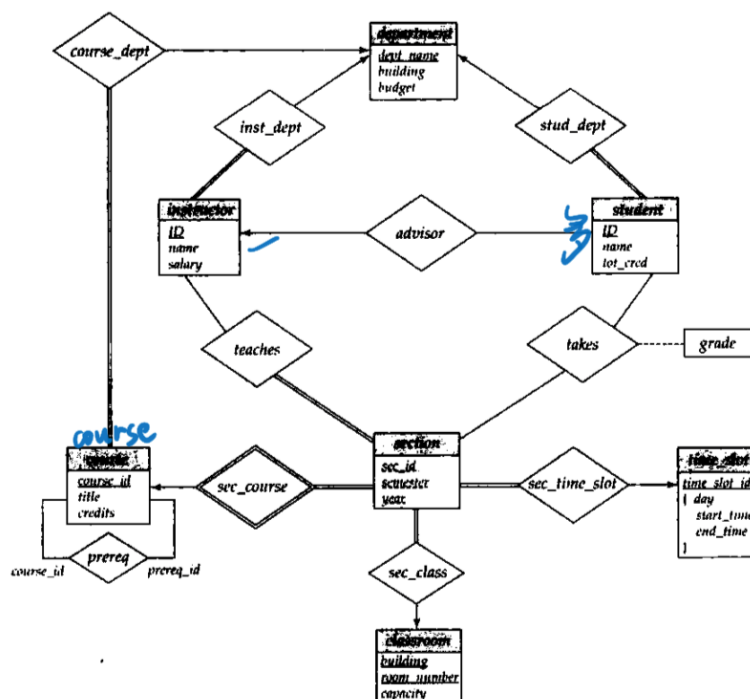


图 7-15 大学的 E-R 图

在我们的大学数据库中，我们限制每名教师必须有且仅有一个相关联的系。因此，如图 7-15 所示，在 *instructor* 和 *inst_dept* 之间有一条双线，表示 *instructor* 在 *inst_dept* 中全部参与；即每名教师必须和一个系相关联。另外，存在一个从 *inst_dept* 到 *department* 的箭头，表示每个教师可以有至多一个相关联的系。

类似地，实体集 *course* 和 *student* 分别与联系集 *course_dept* 和 *stud_dept* 用双线连接，实体集 *section* 和联系集 *sec_time_slot* 亦如此。前面的两个联系分别通过箭头指向另一个联系 *department*，而第三个联系通过箭头指向 *time_slot*。

此外，图 7-15 显示联系集 *takes* 具有一个描述性属性 *grade*，以及每个学生有至多一位导师。该图还表明目前 *section* 是一个弱实体集，由属性 *sec_id*、*semester* 以及 *year* 组成分辨符；*sec_course* 是连接弱实体集 *section* 和强实体集 *course* 的标识性联系集。

7.6转换为关系模式

简单属性的强实体集

强实体集的主码就是生成模式的主码

具有复杂属性的强实体集

当一个强实体集具有非简单属性时，事情更加复杂一点。我们通过为每个子属性创建一个单独的属性来处理复合属性；我们并不为复合属性自身创建一个单独的属性。例如，考虑图 7-11 中表示的 *instructor* 实体集。对于复合属性 *name*，为 *instructor* 生成的模式包括属性 *first_name*、*middle_initial* 和 *last_name*；没有单独的属性或模式表示 *name*。类似地，对于复合属性 *address*，产生的模式包括属性 *street*、*city*、*state* 以及 *zip_code*。由于 *street* 是一个复合属性，因此它被替换成 *street_number*、*street_name* 以及 *apt_number*。8.2 节将再次讨论这个问题。

弱实体集表示

设 *A* 是具有属性 a_1, a_2, \dots, a_n 的弱实体集，设 *B* 是 *A* 所依赖的强实体集，设 *B* 的主码包括属性 b_1, b_2, \dots, b_n 。我们用名为 *A* 的关系模式表示实体集 *A*，该模式的每个属性对应以下集合中的一个成员：

$$\{a_1, a_2, \dots, a_n\} \cup \{b_1, b_2, \dots, b_n\}$$

以图 7-15 所示 E-R 图中的弱实体集 *section* 为例说明。该实体集有属性：*sec_id*、*semester* 和 *year*。*section* 实体集所依赖的实体集 *course* 的主码是 *course_id*。因此，用来表示 *section* 的模式具有下面的属性：

section(course_id, sec_id, semester, year)

该主码由实体集 *course* 的主码和 *section* 的分辨符（即 *sec_id*、*semester* 以及 *year*）组成。我们还在模式

联系集表示

7.7实体-联系设计问题

用实体集or属性

但你希望这个属性的其他信息也存储进来时候，就把这个属性单独作为一个实体集

用实体集or联系集

二元集or n元集

联系属性的布局

7.8扩展的E-R特性

特化

例如，实体集 *person* 可进一步归类为以下两类之一：

- *employee*。
- *student*。

概化

一定数量的实体集共享一些共同特征。特化的反向对应“超类”和“子类”，即“高层实体”和“低层实体”

属性继承

高层实体被低层继承

概化上的约束

聚集

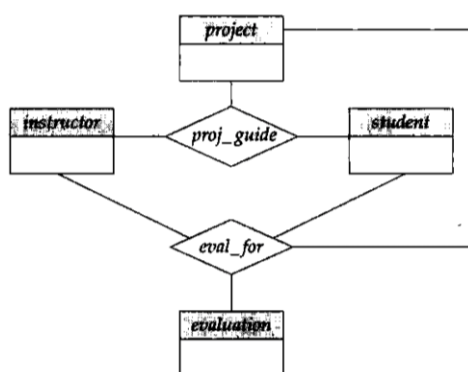


图 7-22 包含冗余联系的 E-R 图

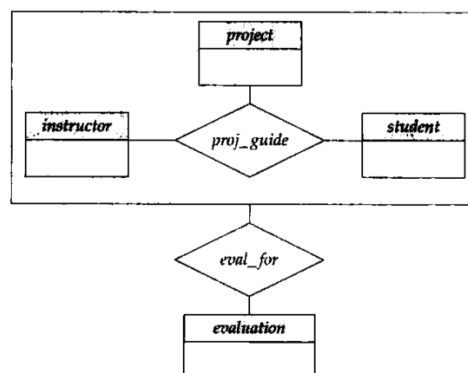


图 7-23 包含聚集的 E-R 图

转换为关系模式

7.9 数据建模的其他表示方法

E-R图的其他表示方法

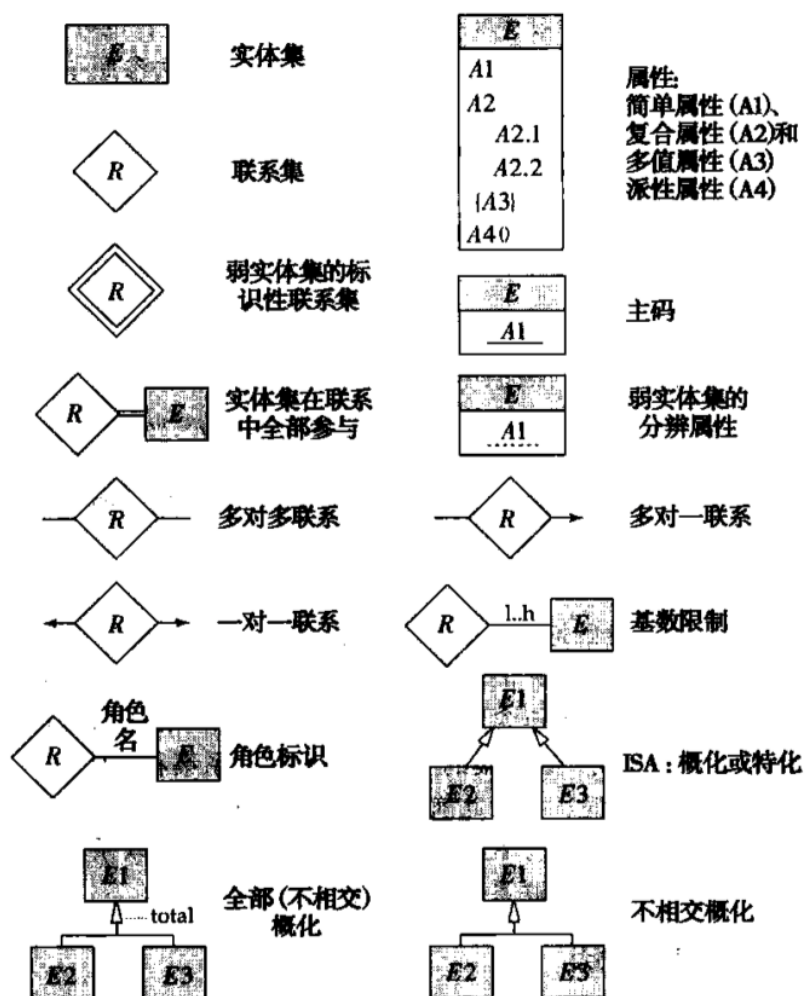


图 7-24 E-R 图表示法中使用的符号

实体集 E 包含简单属性 $A1$ 、复合属性 $A2$ 、多值属性 $A3$ 、派性属性 $A4$ 、以及主码 $A1$

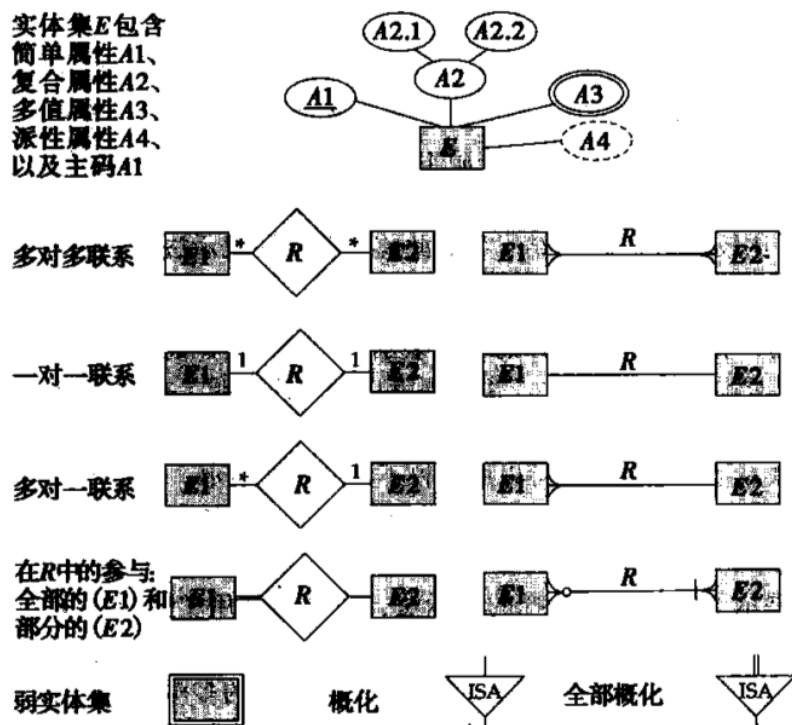


图 7-25 其他可选择的 E-R 图表示法

7.10数据库设计的其他方面

数据库约束和关系数据库设计

使用需求：查询、性能

授权需求

数据流、工作流

数据库设计的其他问题

7.11总结