

7 数据库设计和E-R模型

Author: 中山大学 17数据科学与计算机学院 YSY

<https://github.com/ysyisyourbrother>

7.1 设计过程概览

实体 (entity)：我们使用这个术语来指示所有可明确识别的个体。在大学数据库的例子中，实体的例子包括教师、学生、系、课程和开课。

在设计一个数据库模式的时候，我们必须确保避免两个主要的缺陷。

冗余：一个不好的设计可能会重复信息。它的最大问题是更新困难，浪费空间

不完整：如果设计的不好，可能会导致有些地方会设置为了空值。

7.2 实体-联系模型

实体：现实世界中可区别于所有其他对象的一个事物或对象。每个实体有一组性质，其中一些性质的值可以唯一的标识一个实体。

实体集：是相同类型即具有相同性质的一个实体集合

属性：实体通过一组属性来表示，属性是实体集中每个成员的描述性质，每个属性有一个值

7.2.2 联系集

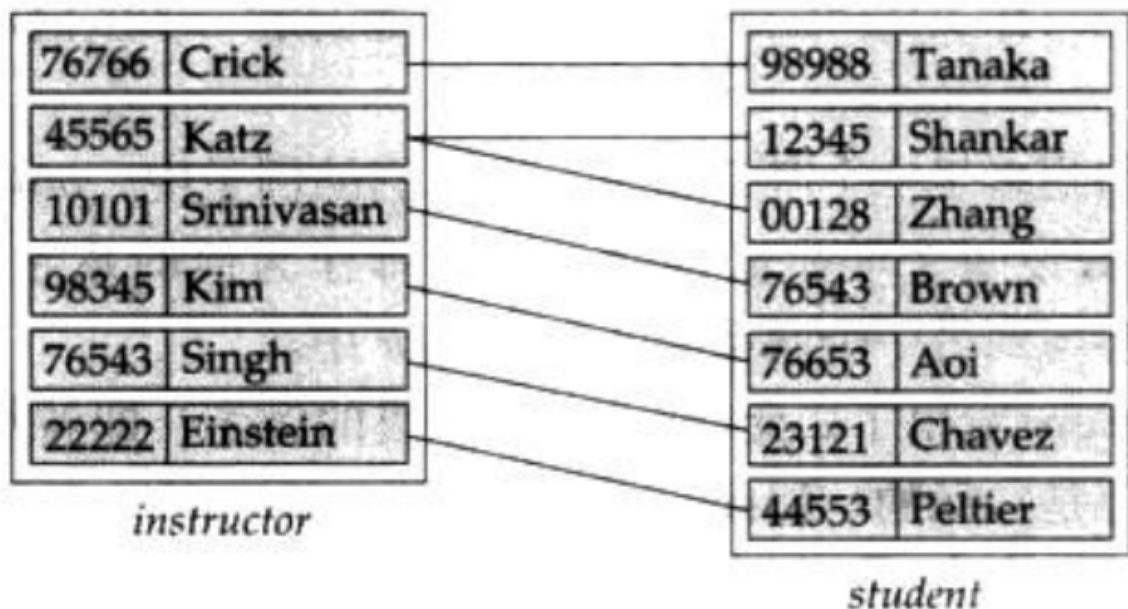
联系：是指多个实体间相互关联

联系集：是相同类型的联系的集合。联系集是 $n \geq 2$ 个（可能相同的）实体集上的数学关系，如果 $E_1 E_2 \dots E_n$ 为实体集，那么联系集R是

$$\{(e_1, e_2, \dots, e_n) \mid e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$$

的一个子集，而 (e_1, e_2, \dots, e_n) 是一个联系

如下图，一条连线就是一个联系，多条连线的集合就是一个联系集



参与：实体集之间的关联称为参与，也就是说，实体集 $E_1 E_2 \dots E_n$ 参与了联系集 R 。例如一个教师和一个学生实体分别参与到advisor的一个联系实例中。

角色：实体在联系中扮演的功能称为实体的角色

自环的：同样的实体集以不同的角色参与到一个联系集多于一次。有必要用显示的角色名来指明。

体是如何参与联系实例的。例如，考虑记录大学开设的所有课程的信息的实体集 *course*。我们用 *course* 实体的有序对来建模联系集 *prereq*，以描述一门课程 (C_2) 是另一门课程 (C_1) 的先修课。每对课程中的第一门课程具有课程 C_1 的角色，而第二门课程具有先修课 C_2 的角色。按照这种方式，所有的 *prereq* 联系通过 (C_1, C_2) 对来表示，排除了 (C_2, C_1) 对。

描述性属性：考虑实体集instructor和student之间的联系集advisor，我们可以将属性date与该联系关联起来，以表示教师成为学生导师的日期。

给定的联系集中的每一个联系，必须由参与实体唯一标识。例如一个教师是一个学生的导师这样一个联系，假如日期是这个联系的属性。我们不能用同一个教师和学生多个联系来表示多个日期，因为这些联系的参与的实体都是同一个教师和学生，是无法唯一标识每一个联系的。如果想对应多个日期，正确的处理方法是建立一个**多值属性**

相同的实体集可以参与到多于一个的联系集中

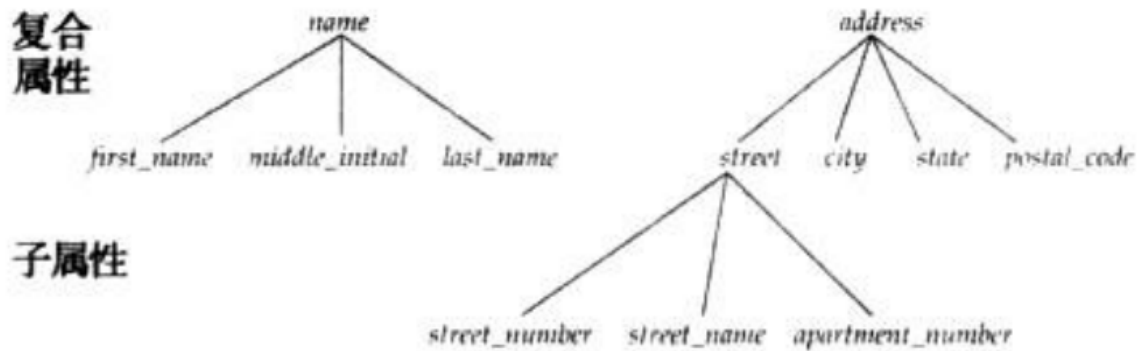
二元联系集：涉及两个实体集的联系集，如上面的advisor。参与联系集的实体集的数目称为**联系集**的**度**

7.2.3 属性

域：每个属性都有一个可取值的集合，称为该属性的域

简单属性：不能划分成更小的部分。

复合属性：复合属性可以划分成更小的部分。



单值属性：属性只能对应一个值

多值属性：属性可以对应多个值、

派生属性：这类属性的值可以从别的属性或实体得出。比如instructor实体集有一个属性students_advised表示一个教师指导多少学生，可以统计和一个教师关联的所有student实体来计算。

派生属性的值不需要存储，而是在需要的时候计算出来

7.3 约束

7.3.1 映射基数

映射基数：表示一个实体通过一个联系集关联的实体数。

对于实体集 A 和 B 之间的二元联系集 R 来说，映射基数必然是以下情况之一：

- **一对一 (one-to-one)**。 A 中的一个实体至多与 B 中的一个实体相关联，并且 B 中的一个实体也至多与 A 中的一个实体相关联 (如图 7-5a 所示)。
- **一对多 (one-to-many)**。 A 中的一个实体可以与 B 中的任意数目 (零个或多个) 实体相关联，而 B 中的一个实体至多与 A 中的一个实体相关联 (如图 7-5b 所示)。
- **多对一 (many-to-one)**。 A 中的一个实体至多与 B 中的一个实体相关联，而 B 中的一个实体可以与 A 中任意数目 (零个或多个) 实体相关联 (如图 7-6a 所示)。
- **多对多 (many-to-many)**。 A 中的一个实体可以与 B 中任意数目 (零个或多个) 实体相关联，而且 B 中的一个实体也可以与 A 中任意数目 (零个或多个) 实体相关联 (如图 7-6b 所示)。

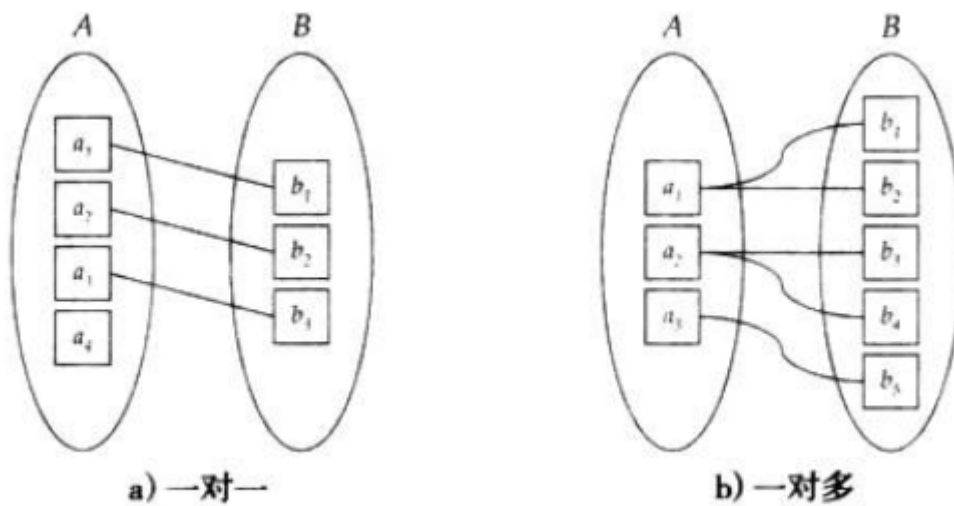
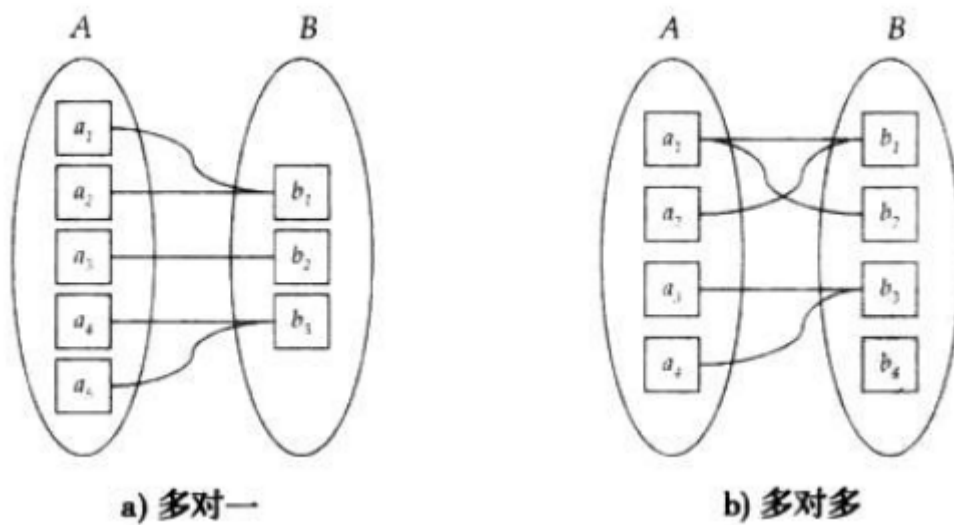


图 7-5 映射基数



7.3.2 参与约束:

全部参与: 实体集E中每个实体都参与到联系集R的至少一个联系中, 实体E在联系集R中的参与称为全部的。

部分参与: E只有部分实体参与到联系集R中

7.3.3 码

一个实体的属性值必须可以唯一标识该实体。各个实体是互异的, 不允许所有属性完全相同。

码: 足以区分每个实体的属性集

联系集的超码: 设R是一个涉及实体集 E_1, E_2, \dots 的联系集。设 $\text{primary_key}(E_i)$ 代表构成实体集 E_i 的主码的属性集合。

如果联系集 R 没有属性与之相关联, 那么属性集合

$$primary-key(E_1) \cup primary-key(E_2) \cup \cdots \cup primary-key(E_n)$$

描述了集合 R 中的一个联系。

如果联系集 R 有属性 a_1, a_2, \cdots, a_m 与之相关联, 那么属性集合

$$primary-key(E_1) \cup primary-key(E_2) \cup \cdots \cup primary-key(E_n) \cup \{a_1, a_2, \cdots, a_m\}$$

描述了集合 R 中的一个联系。

在以上两种情况下, 属性集合

$$primary-key(E_1) \cup primary-key(E_2) \cup \cdots \cup primary-key(E_n)$$

在以上两种情况下, 属性集合:

$$primary-key(E_1) \cup primary-key(E_2) \cup \cdots \cup primary-key(E_n)$$

构成了联系集的一个超码, 它唯一确定一个联系集。

联系集的主码结构还依赖于联系集的映射基数:

联系集的主码结构依赖于联系集的映射基数。例如, 考虑在 7.2.2 节中的实体集 *instructor* 和 *student* 以及具有属性 *date* 的联系集 *advisor*。假设联系集是多对多的, 那么 *advisor* 的主码由 *instructor* 和 *student* 的主码的并集组成。如果联系是从 *student* 到 *instructor* 多对一的, 即每个学生最多只能有一个导师, 则 *student* 的主码就是 *advisor* 的主码。而如果一名教师只能指导一名学生, 即联系是从 *instructor* 到 *student* 多对一的, 则 *instructor* 的主码就是 *advisor* 的主码。对于一对一的联系, 两个候选码中的任意一个可以用作主码。

7.4 从实体集中删除冗余属性

7.5 实体联系图

7.5.1 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)。

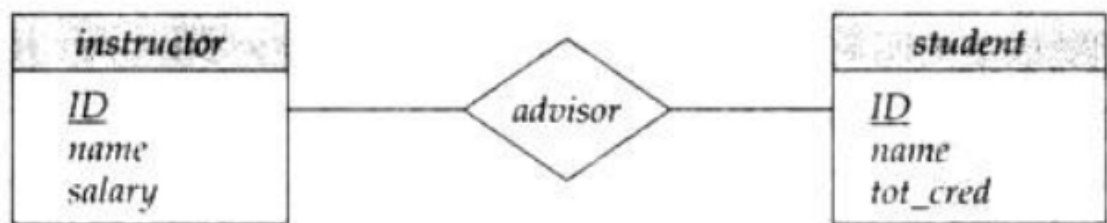
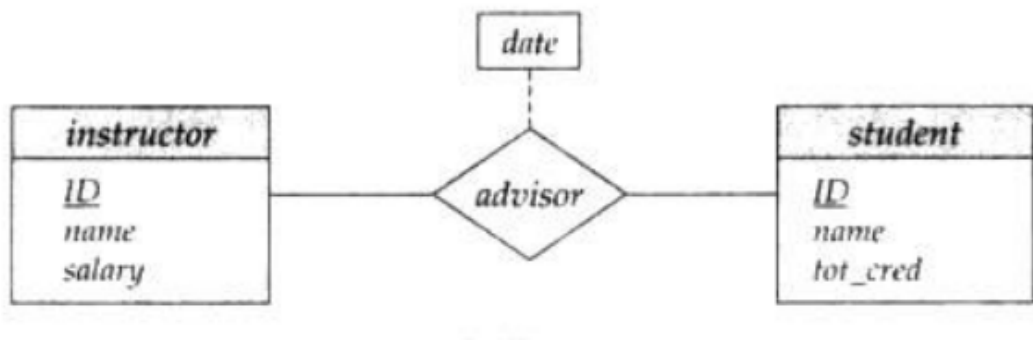
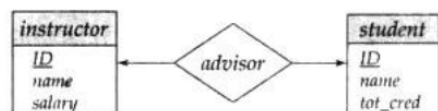


图 7-7 对应于教师和学生的 E-R 图

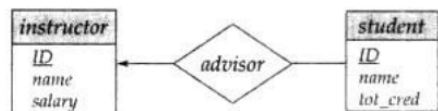


7.5.2 映射基数

- **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。



a) 一对一



b) 一对多



ER图还有更复杂的描述实体参与联系集中联系次数的约束方法：

实体集和二元联系集的边可以有一个**关联的最大和最小的映射基数**，而h表示最大的映射基数。

1. 最小值为1表示这个实体集在该联系集中全部参与，而实体集中的每个实体在联系集中的至少一个联系中出现。
2. 最大值为1表示这个实体最多参与一个联系
3. *表示没有限制
4. 如果两边都有最大值为1，说明联系是1对1的

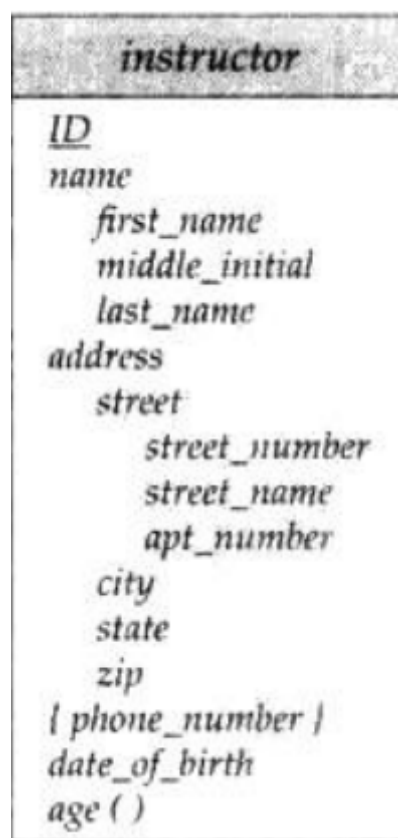
例如下面的导师和学生表中的约束1..1意味着基数的最小和最大都是1，也就是说每个学生有且只有一个老师。从0..*可以知道，每个导师可以有0个或多个学生。因此导师和学生是一对多的关系：



图 7-10 联系集上的基数约束

7.5.3 复杂的属性

- name street等都是复合的属性
- phone_number是多值属性
- age()表示派生属性



7.5.4 角色

在 E-R 图中，我们通过在菱形和矩形之间的连线上进行标注来表示角色。图 7-12 给出了 *course* 实体集和 *prereq* 联系集之间的角色标识 *course_id* 和 *prereq_id*。

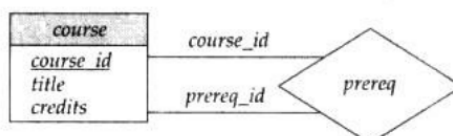


图 7-12 包含角色标识的 E-R 图