

1

关系数据库



主要内容

- 了解关系型数据库
- 掌握 E-R 模型到关系模型的转化
- 掌握数据库设计的三个规范

数据的结构化集合称为数据库。数据库模型可分为层次模型、网状模型和关系模型。关系模型是目前最普遍的一种数据模型,它采用人们所熟知的二维表格来描述现实世界中的实体以及各实体之间的各种联系,概念清晰、使用方便。在其被提出之后,就得到数据库开发商的积极支持,使其迅速成为继层次模型和网状模型之后一种崭新的数据库模型,如今在数据库技术领域占据了统治地位,市场上流行的数据库产品几乎都支持关系模型。

本章通过一张表和一个简单的教学管理系统案例来讲解关系模型以及如何利用关系模型来设计数据库。

1.1 关系模型基本概念

关系模型是由美国 IBM 公司的 E.F.Codd 博士提出的,是一种用二维表的形式来描述现实世界中实体以及实体间联系的数据模型。

现在通过图 1-1 所示的学生信息二维表来介绍关系模型中的基本概念。

- (1) 关系: 一个关系对应一张没有重复行、重复列的二维表。
- (2) 表结构: 即表中的第一行,表示组成该表的各个字段的名称。在表中,还应该包括

各字段的数据类型等。

(3) 属性：二维表中垂直方向的每一列称为一个属性，在表中对应一个字段。每一列有一个属性名，即字段名，如表中的学号、姓名、性别、生日等。

(4) 关系模式：是指对关系结构的描述，一般表示如下：

关系名(属性 1,属性 2,属性 3,...,属性 n)

例如图 1-1 所示的关系模式可以描述为：

学生(学号,姓名,性别,生日,邮箱,班级)

(5) 属性值：表中行和列的交叉位置表示某个属性的值，如“郭力月”“男”。

(6) 空值：表示属性值“不确定”“不知道”，用 NULL 表示。例如图 1-1 中学号为 0000000003 的学生没有给出邮箱，属性值为 NULL，因为不确定该学生的邮箱是什么。

(7) 域：属性的取值范围，如性别的域是(男,女)。

学号	姓名	性别	生日	邮箱	班级
0000000001	马文霞	男	1982-02-01	mawenxia@91xueit.com	测试
0000000002	郭力月	男	1987-01-15	guoliyue@91xueit.com	开发
0000000003	崔霄晨	女	1985-11-04	NULL	网络
0000000004	钟霁翔	女	1986-08-06	zhongaixiang@91xueit.com	开发
0000000005	夏可炎	男	1989-03-25	xiakeyan@91xueit.com	网络
0000000006	常亮思	女	1984-01-26	changliangsi@91xueit.com	开发
0000000007	邓咏桂	男	1982-11-18	dengyonggui@91xueit.com	开发
0000000008	方伊家	女	1985-03-02	fangyijia@91xueit.com	测试

图 1-1 学生信息表

(8) 主键：在关系中能够唯一标识表中不同行的属性或属性组合，并且这些属性值不包括空值和重复值，用 Primary Key 表示。如果是属性组合的主键，则称为联合主键。一张表中只能有一个主键，联合主键虽然包含了两列甚至多列，但也是一个联合主键。

例如在学生信息表中，一个学号值可以唯一地确定一个学生的所有信息，而且学号不能为空，也不可能重复，因而可以作为主键。因为学生可能重名，因此“姓名”列不能作为主键。

(9) 外键：如果 A 表中的一个字段引用了 B 表中的字段，两表之间根据这个引用的字段建立联系，那么引用的这个字段称为 A 表的外键，用 Foreign Key 表示。一般作为外键的字段具有唯一性和非空性，因此某个表的主键常常被引用为另一个表的外键。

例如在图 1-2 所示的学生成绩表中，“学号”“课程号”和“分数”都不能唯一地标识每个学生的成绩信息，只有学号和课程号组合起来才能区分每个学生的每科成绩，因此该表的主键是属性组(学号,课程号)组成的联合主键。学号是学生信息表的主键，而不是学生成绩表的主键，在两表相关联的时候学号称为学生成绩表的外键。

(10) 主表和从表：是指通过外键相关联的两个表，其中被引用字段所在的表称为主表，另一张表称为从表。

例如图 1-2 中，学生信息表以学号作为主键，而学号是学生成绩表的外键，所以学生信息表是主表，学生成绩表是从表。

学生信息表					学生成绩表		
学号	姓名	性别	生日	班级	学号	课程号	分数
0000000001	马文霞	男	1982-02-01	测试	0000000001	0001	59
0000000002	郭力月	男	1987-01-15	开发	0000000001	0002	55
0000000003	崔霄晨	女	1985-11-04	网络	0000000001	0003	70
0000000004	钟雷翔	女	1986-08-06	开发	0000000002	0001	58
0000000005	夏可炎	男	1989-03-25	网络	0000000002	0002	77
0000000006	常亮思	女	1984-01-26	开发	0000000002	0003	63

图 1-2 学生信息表和学生成绩表

(11) 关系数据库：是指一些相关的表和其他数据库对象的集合。一个关系数据库包含多个数据表，这些表之间的关联性是由主键和外键所体现的参照关系来实现的。数据库不仅包含表，而且包含了其他数据库对象，如视图、存储过程、索引等。

1.2 E-R 模型到关系模型的转化

数据库设计案例：某学校设计学生教学管理系统，学生实体包括学号、姓名、性别、生日、民族、籍贯、登记照，每名选择选择一个主修专业，专业包括专业号、专业名和专业类别，一个专业属于一个学院，一个学院可以有若干个专业。学院信息要存储学院号、学院名、院长。教学管理还要管理课程表和学生成绩。课程表包括课程号、课程名、学分，每门课程由一个学院开设。学生学习每门课程获得一个成绩。

如图 1-3 所示，有学生、专业、学院和课程 4 个实体和每个实体对应的属性，实体用方框表示，属性用椭圆表示，这种模型称为 E-R 模型。所谓 E-R 模型，它是一种工具，将实体与属性转化为关系模型，作为连接实际对象与数据库的桥梁。



图 1-3 4 个实体的 E-R 图



图 1-3 4 个实体的 E-R 图 (续图)

针对案例,可以初步设计成 4 张二维表,分别为学生表、专业表、学院表和课程表。每张表都有一个主键,分别是学号、专业号、学院号和课程号。

根据上面的案例需求,一个学生只能主修一个专业,一个专业可以由多名学生主修,因此专业和学生是一对多的关系;一个学院可以设置多个专业,一个专业只能属于一个学院,因此学院和专业是一对多的关系;一个学院可以开设多门课程,一门课程只能由一个学院开设,因此学院和课程是一对多的关系;一个学生可以学习多门课程,一门课程可以由多名学生学习,因此课程和学生的关系是多对多的关系。

1. E-R 模型到关系模型的转化

(1) E-R 模型中的实体映射为关系表。

(2) E-R 模型中的属性映射为关系属性,即字段。

2. 一对多 (1:n) 到关系模型的转化

要转化 1:n 的关系,需要在 n 方实体表中增加一个属性,该属性对应 1 方的主键属性。例如专业和学生两个实体,学生实体是 n 方,专业实体是 1 方,需要在学生实体表中增加专业表中的主键,即专业号。通过增加该属性可以唯一确定双方之间的联系,知道了学号可以立即对应到该学生所修的专业。

图 1-4 所示的模型转化为关系模型为:

学生(学号,姓名,性别,生日,民族,籍贯,登记照,专业号)

专业(专业号,专业名,专业类别)

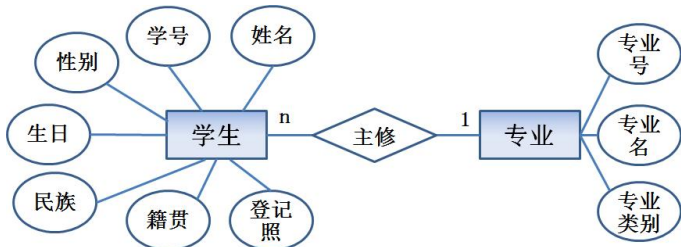


图 1-4 1:n 到关系模型的转化

3. 多对多 (m:n) 到关系模型的转化

转化一个 m:n 的关系,除了对两个实体分别转化以外,还需要单独建立一个关系模式,该关系模式以两个实体的主键作为它的外键,以表示两个实体间的 m:n 关系。例如,学生和

课程之间的关系是 $m:n$ ，新建一个关系成绩表，成绩表中有学号和课程号两个属性。

图 1-5 所示的模型转化为关系模型为：

学生(学号,姓名,性别,生日,民族,籍贯,登记照)

课程(课程号,课程名,学分)

成绩(学号,课程号,成绩)

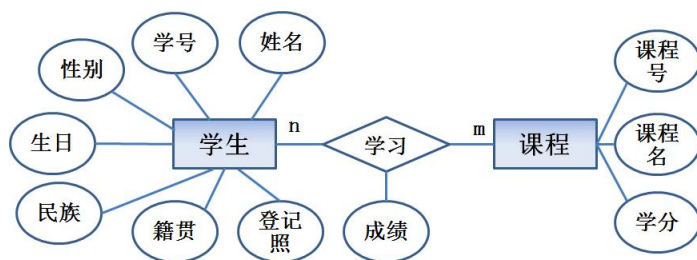


图 1-5 $m:n$ 到关系模型的转化

4. 一对一 (1:1) 到关系模型的转化

在转化这种关系的时候，在任意一方实体表里添加另一方的主键属性即可满足要求。例如班级和班长的关系，一个班级只能有一个班长，一个班长只能属于一个班级，它们的关系是一对一的关系。

图 1-6 所示的模型转化为关系模型为：

班级(班号,专业,人数,学号)

班长(学号,姓名,专长)

也可以转化为：

班级(班号,专业,人数)

班长(学号,姓名,专长,班号)

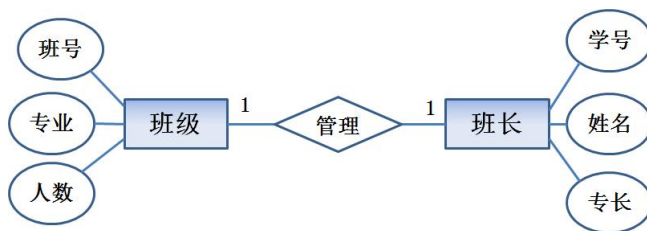


图 1-6 1:1 到关系模型的转化

通过上面的分析，案例中的模型可以如图 1-7 所示。

由 E-R 模型转化的关系模型是：

学生(学号,姓名,性别,生日,民族,籍贯,登记照,专业号)

专业(专业号,专业,专业类别,学院号)

学院(学院号,学院,院长)

课程(课程号,课程名,学分,学院号)

成绩(学号,课程号,成绩)

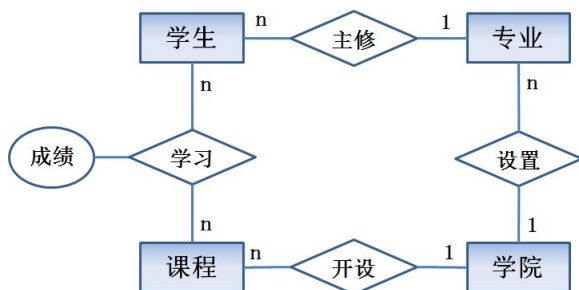


图 1-7 E-R 模型转化为关系模型

1.3 数据库设计规范

上面的设计已经满足了这个简单的教学管理系统数据库设计的大部分要求。像这样简单的小规模数据库处理起来比较轻松，但随着数据库规模的扩大，用户操控数据库将变得笨拙、复杂，性能也不好，更糟糕的是很有可能导致数据不完整、不准确。所以针对复杂的数据库，需要更规范更严谨的设计。

一个良好的数据库设计，能够节省数据存储空间，确保数据完整性，方便数据库应用系统的开发。而一个糟糕的数据库设计，往往存在数据冗余、数据更新、数据插入和数据删除的缺陷。

怎样使数据库设计更加规范呢？一般来说，满足以下三个范式即可避免糟糕的数据库设计：第一范式、第二范式和第三范式。

1.3.1 第一范式 (1NF)

如图 1-8 所示，该表的地址栏包含了省份、城市和详细地址信息，如果对省份或地址有操作需求，比如需要按省份挑选学生，这样的设计就不合理。这样的设计需求称为第一范式。也就是说第一范式要求数据库表中的字段都是单一属性且字段值不可再分解。第一范式的目标是保证每一列的原子性。

可以将该表设计成如图 1-9 所示，这样的设计满足第一范式。

当然，第一范式是根据实际需求来决定的。同样是图 1-8 所示的数据库表，如果关于这个表的操作都不涉及省份、城市，那么对地址这个字段也就没有进一步分解的要求。在这样的情况下，图 1-8 所示的表设计也符合 1NF。

学号	姓名	性别	地址
2012001	韩立刚	男	河北省石家庄市南二环东路20号
2012002	韩力挥	男	河北省冀城市廉州西路002
2012003	张静	女	河北省晋州市滨河路28号
2012004	王国涛	男	贵州省贵阳市南明区中华南路99号
2012005	靳灵	女	山东省临沂市兰山区通达路18号
2012006	李华东	男	福建省厦门市集美区盛光路56号
2012007	许东	男	浙江省温州市鹿城区五马街11号
2012008	崔栋	男	广东省广州市白云区天明路148号

图 1-8 违反第一范式的表

学号	姓名	性别	省份	城市	详细地址
2012001	韩立刚	男	河北省	石家庄	南二环东路20号
2012002	韩力挥	男	河北省	冀城	廉州西路002
2012003	张静	女	河北省	晋州市	滨河路28号
2012004	王国涛	男	贵州省	贵阳市	南明区中华南路99号
2012005	靳灵	女	山东省	临沂市	兰山区通达路18号
2012006	李华东	男	福建省	厦门市	集美区盛光路56号
2012007	许东	男	浙江省	温州市	鹿城区五马街11号
2012008	崔栋	男	广东省	广州市	白云区天明路148号

图 1-9 符合第一范式的表

1.3.2 第二范式（2NF）

第二范式要求一张表只包含一个实体信息，并且每行记录由主键唯一标识这一行。

如图 1-10 所示，学生课程表描述了学生及主修的课程信息。表中所有字段不可再分，因此该表满足第一范式。同时在该表中，字段“学号”“姓名”和“性别”属于学生信息，这是该表描述的第一件事，字段“课程编号”“课程”和“学分”属于课程信息，这是该表描述的另一件事。一张表中同时描述两件事，这样的设计不满足第二范式。

学生课程表

字 段	例 子
学号	2010452011
课程编号	001
课程	数学
姓名	韩立刚
学分	2
性别	男
...	...

图 1-10 不满足第二范式的设计

可以改进表设计，将该表分成学生信息表和课程信息表，学生信息表中包含字段学号、

姓名和性别，学号作为该表的主键，课程信息表中包含字段课程编号、课程和学分，课程编号作为该表的主键，如图 1-11 所示，这样的设计就能满足第二范式。

学生信息表		
字 段	例 子	
学号	2010452011	主键
姓名	韩立刚	
性别	男	

课程信息表		
字 段	例 子	
课程编号	001	主键
课程	数学	
学分	2	

图 1-11 满足第二范式的设计

1.3.3 第三范式 (3NF)

第三范式要求数据表只能引用其他表中的主键，而不能引用其他表中的非主键列。

在 1:n 的关系中，从表的设计要满足第三范式，即从表只引用主表的主键。

一个学生可能有多个手机号，因此需要设计一个联系方式表来存储学生的多个手机。其中学生信息表是主表，联系方式表是从表，从表引用主表的主键。图 1-12 所示不满足第三范式，图 1-13 所示满足第三范式，因为在联系方式表中不应该出现姓名列，通过联系方式表的学号可以从学生信息表中找到学生姓名，姓名在这张从表中属于冗余字段，因此在从表中只需引用主表的主键字段即能满足第三范式。

从表不满足第三范式案例

学生信息表 (主表)	
字 段	例 子
学号	2010452011
姓名	韩立刚
性别	男

联系方式表 (从表)	
字 段	例 子
学号	2010452011
姓名	韩立刚
手机	13803367892

图 1-12 不符合第三范式的设计

从表满足第三范式案例

学生信息表 (主表)	
字 段	例 子
学号	2010452011
姓名	韩立刚
性别	男

联系方式表 (从表)	
字 段	例 子
学号	2010452011
手机	13803367892

图 1-13 符合第三范式的设计

在 m:n 的关系中，关系表的设计也要满足第三范式，即关系表只引用实体表的主键。

学生信息表记录学生信息，课程信息表记录课程信息，如果想记录学生考试成绩，这个表应该怎样设计呢？

要记录考试成绩，需要新增一张关系表——成绩表。成绩表中需要记录学生信息、考试的课程信息和考试的成绩。虽然在这张表中记录三个信息，但是学生信息和课程信息可以通过

外键来查找具体的值。因此在图 1-14 中,姓名字段属于冗余字段,这样的设计不满足第三范式。成绩表中只需引用课程信息表和学生信息表的主键,如图 1-15 所示,这样就能满足第三范式。



图 1-14 不符合第三范式的设计



图 1-15 符合第三范式的设计

综上所述,主、从表之间的设计和实体、关系表之间的设计都需要满足第三范式。

上面介绍了三个范式,更严格的数据库设计还有第四范式、第五范式和 BC 范式。所有的范式都在前提范式的基础上进一步满足了更多要求。需要注意的是,更严格的范式可能会带来性能上的一些不足,因为对数据的操作可能需要频繁地进行表结合。一般来说,数据库只需满足第三范式(3NF)就足够了。