

## 第四讲 数据库设计

### 【教学目的和要求】

1. 数据模型
2. 实体关系 E-R 图
3. 关系数据库范式

### 【主要内容】

#### 4.1 数据模型

- 4.1.1 三个世界的概念
- 4.1.2 不同世界概念的对应关系
- 4.1.3 数据抽象的基本过程

#### 4.2 实体关系 E-R 图

- 4.2.1 E-R 方法的概念
- 4.2.2 E-R 信息模型的设计
- 4.2.3 实体例子
- 4.2.4 实体之间的联系

#### 4.3 关系数据库范式

- 4.3.1 第一范式
- 4.3.2 第二范式
- 4.3.3 第三范式

小结

习题三

### 案例

#### 【电子教案】

参见：第四讲 数据库设计

#### 【重点与难点】

1. 数据的模型的作用
2. 实体关系 E-R 图的含义；
3. 关系数据库范式的分析。

#### 【教材和参考读物】

《管理信息系统——理论与实践》第三章

《管理信息系统》甘仞初 第三章

#### 【教学时数】 2

# 第 4 讲 数据库设计

## 4.1 数据模型

### 1. 数据模型及其分类

数据模型是对现实世界进行抽象的工具。现实世界是复杂多变的，目前任何一种数据建模技术都不可能原样复制实在中存在的对象，只能抽取其局部特征，构造反映现实实体主要特征的数据模型。图 3-5 描述了数据抽象的基本过程。

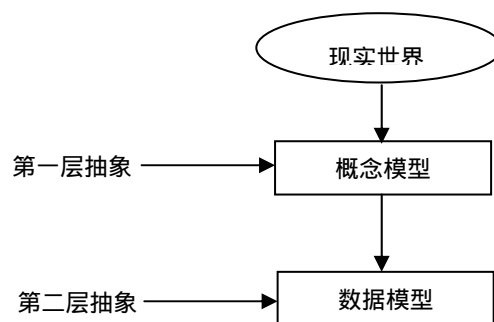


图 3-5 客观对象的抽象过程

概念模型是对现实世界的第一层抽象，也称信息模型。这一类模型中最著名的是实体联系模型（E-R 模型），它按照用户的观点对数据建模，具有较强的语义表达能力，概念清晰简单，易于理解，是数据库设计人员与用户之间交流的语言。

第二层抽象直接面向数据库的逻辑结构，描述了数据之间的逻辑组合，被称为数据模型。例如网状模型、关系模型和层次模型等。这类模型有严格的形式化定义，便于计算机存储和运算。数据模型应满足三方面的要求：一是能比较真实地模拟现实世界的实体；二是易于理解；三是便于在计算机中实现，并进行相关的运算。

### 2. 数据模型的组成要素

数据模型包括三部分：数据结构、数据操纵和数据的完整性约束。

数据结构是实体对象存储在数据库中的记录集合。例如建立一个人事档案数据库，每个人的基本情况，如姓名、单位、出生年月、工资、工作年限等是数据对象的某个特征，构成数据库的存储框架。其中每一个人属于某一个部门，某一个部门包含多个员工，这些“关系”也要存储在数据库中。数据库数据的结构不同，导致数据模型之间的存在显著差异，常见的数据库数据模型有：层次型、关系型和网状型。

数据操纵是指对数据库中各种对象实例的操作。例如根据用户的要求，检索、增加、删除、更新数据库中的数据。

数据的完整性约束是指在给定的数据模型中，数据及数据关联所遵守的一组通用的完整性规则。它能保证数据库中数据的正确性、一致性。例如，数据库主键值是惟一的，而且不能为空；关系数据库中，每个非空的外键值必须与某一主键值匹配。这类完整性约束是数据模型所必须遵守的通用

的完整性规则。另一类完整性约束是根据数据模型的完整性约束机制自己定义的。例如在流程管理中，经理的审批日期一定是在经办人的申请日期之后。

### 3. 概念模型

概念模型实际上是现实世界到计算机世界的一个中间层次。为了把现实世界中的具体事物抽象组织为某一数据库管理系统支持的数据模型，人们常常首先将现实世界抽象为概念模型，然后将概念模型转换为数据模型。也就是说，首先把现实世界中的客观对象抽象为某一种信息结构，这种信息结构并不依赖于具体的计算机系统及 DBMS，而是依赖于概念模型，然后再把概念模型转换为某一 DBMS 支持的数据模型。

使用概念模型描述现实实体涉及以下主要概念：

#### (1) 实体

实体是客观存在并以属性区分其差异的具体事物。数据就是客观现实的描述。为了抽象地描述客观现实相同的事情，可以使用实体的概念。实体中的一个具体事情的出现，就是一个实体的实例。常见的实体包括人、位置、对象、事件和概念等。表 3-2 列出了常见的实体。

表 3-2 常见的实体

实体分类	实体
人	学生、教师、客户、部门、雇员、经理、代理人、导师、供应商、摊主、主管、法官、工人、工程师、设计员、分析员、系统分析员、信息系统管理员
位置	大楼、房间、营地、仓库、分公司、机构、办公室、城市、街道、公路、宾馆、饭店等
对象	图书、机器、零件、产品、汽车、原材料、软件包、设备、飞机、轮船、计算机、电视机、手机、订单、合同等
事件	比赛、奖励、认证、分类、查询、飞行、邀请、订购、注册、请求、旅行、销售、预定、采购、运输、考试、评比等
概念	账户、总账、基金、周期、课程、应收、应付

#### (2) 属性

属性是实体所具有的特性，每一特性都称为实体的属性。例如学生的学号、班级、姓名、性别、出生年月等都是学生的属性。每一属性都有一个值域，例如学号的域为 7 位整数。

#### (3) 实体集

具有相同属性的实体集合称为实体集。例如全体教师是一个实体集，全体学生也是一个实体集。

#### (4) 主键

主键是能惟一标识一个实体的属性及属性值，主键也可称为关键字。例如，学号是学生实体的键。

#### (5) 联系

在现实世界中，实体与实体之间有各种联系，归纳起来，主要有三种情况，见图 3-6 所示。

一对一的关系。这是最简单的一种实体间的联系，它表示了两个实体集中的个体之间存在一种一一对应的关系。例如，每个班级有一个班长，班级实体与班长实体之间的联系是 1:1。

一对多的联系。实体间存在的另一种联系是一对多的联系。例如，一个班级有许多学生，一所医院有多个部门等，这种联系记为 1:M。

多对多的联系。实体间更多的是多对多的联系。例如教师和学生之间的联系。一个教师有多名学生，反之，一个学生同时上几个教师的课。多对多的联系表示了多个实体集，其中一个实体集中的

任一实体与另一实体集中的实体间存在一对多的联系。这种联系记为：M:N。

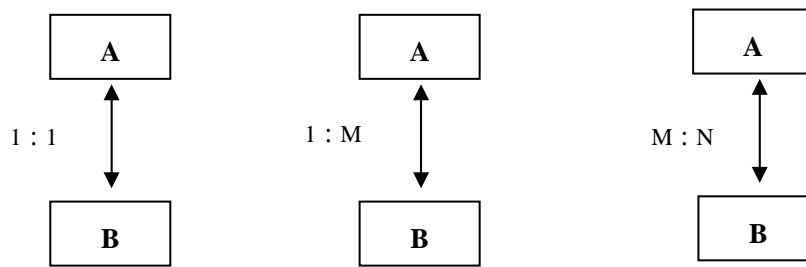


图 3-6 实体之间的联系

#### 4. 概念模型及其表示

概念模型的表示方法很多，其中最为著名的是 P. P. S. Chen 于 1976 年提出的实体-联系方法 (Entity-Relationship Approach)。该方法用 E-R 图描述了现实世界的概念模型，称为 E-R 模型。

E-R 模型有四个基本成分：矩形表示实体，椭圆形表示实体属性，菱形表示联系，连线表示实体之间以及属性之间的联系。矩形框、椭圆形框、菱形框内要标注实体、属性和联系的名字，连线两头标注联系的类型是一对一、一对多还是多对多的联系。

图 3-7 是一个篮球实体及其属性。

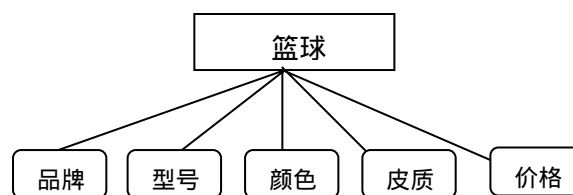


图 3-7 篮球实体及其属性

下面用 E-R 模型来描述某工厂物资管理的概念模型。物资管理所涉及的实体包括：职工、仓库、零部件以及供应商。其中每一个实体都具有相应的属性：

职工：属性有职工号、姓名、年龄、岗位。

仓库：属性有仓库号、面积、类型。

供应商：属性有供应商号、名称、地址、电话、账号、联系人、经理等信息。

零件：属性有零件编号、名称、规格、单价、计量单位、质量等级等。

以上实体之间的联系如下：

一个仓库可以存放多种零件，一种零件可以存放在多个仓库中，因此，仓库与零件之间具有多对多的联系。库存量表示某种零件在某个仓库中的数量。

一个仓库有多个仓库管理员，一个职工只能在某一个仓库工作，仓库与职工之间是一对多的关系。

领导与职工之间是一对多的关系，库存经理可以领导多名仓库管理人员。

供应商、零件之间是多对多的联系，一个供应商供应多种零部件，某一种零部件可以从多个供应商处采购。

依照实体以及实体之间的联系，工厂物资管理的 E-R 模型如图 3-8 所示。

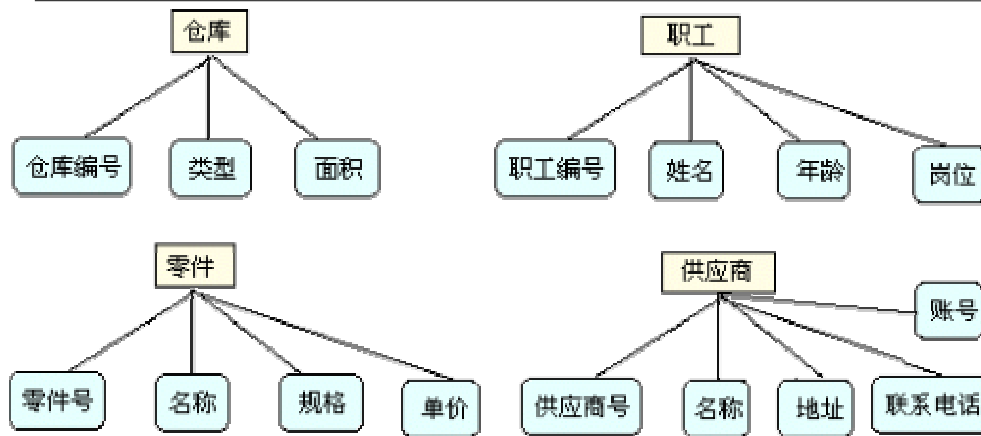


图 3-8(a) 实体及其属性

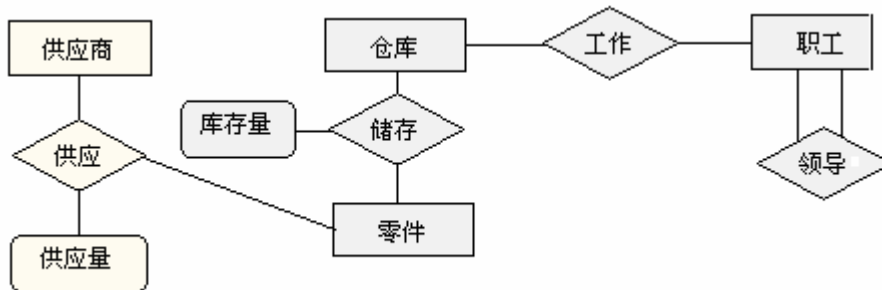


图 3-8(b) 实体联系图

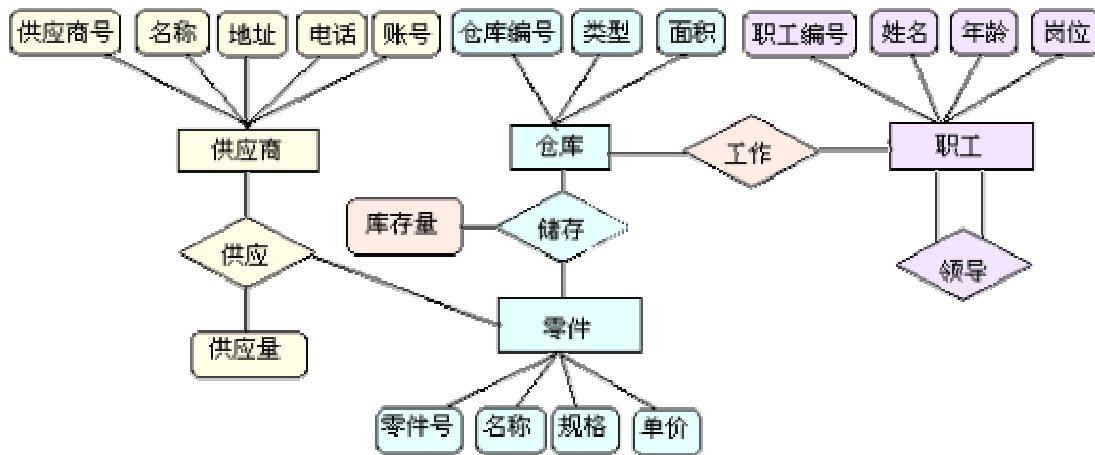


图 3-8(c) 完整的物资管理 E-R 模型

进一步分析物资管理系统的 E-R 图，可以发现，实体所反映的是企业的“数据节点”，每一个节点由一系列属性值描述其特征。比如，零部件实体是企业的一个数据节点，这个数据节点由零件号、名称、规格、单价以及其他数据描述。而 E-R 模型中的联系，事实上对应于企业活动中的业务流程，通过这些流程所有的数据节点联系在一起，构成企业的数据流。比如，供应商和零部件之间的联系，实际上对应于企业的采购流程，每当企业根据生产计划采购某种零件的时候，采购流程启动，购买一定量的零部件之后，企业的库存数据随之发生变化。E-R 图将企业的数据和活动从纷繁复杂的企业

经营活动中抽象出来，在此基础上，按照一定的数据模型，就可以构造企业的数据库结构和数据流。因此，E-R 图是描述企业经营管理活动的一个很好的工具。

## 4.2 三种数据模型

E-R 图描述了现实抽象的概念模型，但是，将概念模型转化为具体的、逻辑表达的数据库，还取决于数据库系统采用怎样的数据模型。数据模型是数据库系统的一个核心问题，数据库管理系统大都是基于某种数据模型的。目前常用的数据模型包括：层次模型、网状模型和关系模型。

### 4.2.1 层次模型

层次模型是最早用于数据库技术的一种数据模型，它是按层次结构来组织数据的。层次结构也称为树形结构，树中的每一个节点代表一种实体类型。这些节点满足两个条件：

- (1) 有且仅有一个节点无双亲。
- (2) 其他节点有且仅有一个双亲节点。

在层次模型中，根节点处在最上层，其他节点都有上一级节点作为其双亲节点，这些节点称为双亲节点的子女节点，同一双亲节点的子女节点称为兄弟节点。没有子女节点称为叶子节点。双亲节点和子女节点表示了实体之间一对多的关系。

层次模型是现实世界存在的各种数据关系的抽象和反映，企业组织、产品结构以及家庭关系都呈现层次结构。例如，ERP 系统中使用的产品结构图，就是非常典型的层次结构。图 3-9 是自行车制造厂有关自行车生产的产品结构图。

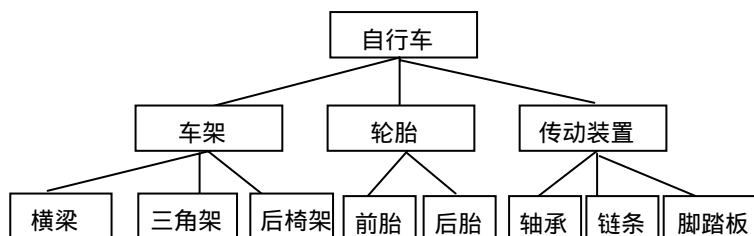


图 3-9 自行车生产层次结构图

### 4.2.2 网状模型

层次模型中一个节点只能有一个双亲节点，且节点间的联系只能是 1:M 的关系，在描述现实世界中自然的层次结构关系是比较简单、直观，易于理解，但对于更复杂的实体间联系就很难描述了。网状模型中，节点间的联系是任意的，任意两个节点间都能发生联系，更适于描述客观世界。

虽然，网状模型适于“描述”现实实体，但在计算机处理时却增加了复杂度，比如如果实体之间存在 M:N 这样的联系，数据结构的组织和处理将十分繁杂。因此，在已经实现的网状数据库中，一般只处理 1:M 的联系；对于 M:N 关系，要先转换成 1:M 联系，然后再处理。通常也将只有 1:M 关系的网称为简单网，而将存在 M:N 关系的网称为复杂网。

网状模型的典型代表是 DBTG 系统，这是 20 世纪 70 年代数据系统语言研究会 CODASYL (Conference On Data System Language) 下属的数据库任务组 (Data Base Task Group, 简称 DBTG) 提出的一个方案。它虽然不是实际的软件系统，但对网状数据库系统的研究和发展起着重大的影响，现有网状数据库系统大都是基于 DBTG 报告文本开发的。

### 4.2.3 关系模型

关系模型是目前最重要的一种数据模型。20 世纪 80 年代以来，几乎所有的数据库系统都支持关系模型，非关系模型也增加了关系模型接口，关系型数据库已经成为数据库系统的工业标准。

关系模型的数据结构是用二维表格表达的实体集，与前两种数据模型相比，其数据结构相对简单。在现实世界中，人们经常使用各种表格，如履历表、报名表、统计表、调查表来表示信息，这些表格可以非常直接地转化为计算机存储信息。

表 3-3 和表 3-4 是学生成绩表和教师任课表。

表 3-3 学生成绩表

姓名	数据结构	操作系统	数据库原理和应用
江丰年	78	86	73
张照	90	88	76
王旭	89	91	79
汝以山	78	88	76
胡延平	76	83	81

表 3-4 教师任课表

姓名	年龄	学院	所任课程
杨老师	38	计算机	数据结构
姜老师	29	计算机	操作系统
张老师	55	经济与管理	信息资源管理
蒋老师	39	计算机	数据库原理与应用
钟老师	46	计算机	计算机接口与编程

从这两个表中可以得到数据之间的一些“联系”，江丰年、张照、王旭、汝以山、胡延平是计算机学院的学生，他们的数据结构、操作系统和数据库原理和应用分别是由杨、姜、张三位老师担任的。在关系模型中，数据被组织成一张二维表格，每一张二维表格称为一个关系。二维表中存放了两类数据——实体本身以及实体之间的联系，因此关系模型的数据组织是非常简单的，它将数据和关系一体化地存储在二维表格中。

### 4.2.4 关系模型的特征

- 结构单一化是关系模型的一大特点。从上面两个数据库可以看出，学生成绩表仅仅为学生情况，而教师数据表反映教师授课信息。
- 关系模型建立在严格的数学理论基础之上，这是关系模型的又一大特点。关系是数学上集合论的一个概念，对关系可以进行各种运算，运算结果形成新关系。在关系数据库中，对数据的全部操作都可以归结为关系的运算。关系模型的逻辑结构实际上是一个二维表，关系数据库的逻辑结构实际上也是一个二维表，这个二维表即是通常所说的关系。每个关系由一组元组成，而每个元组又是由若干属性和域构成的。只有两个属性的关系称为二元关系，有三个属性的关系称为三元关系，以此类推，有  $N$  个属性的关系称为  $n$  元关系。

#### 4.2.5 关系数据库与其他数据库相比的优点

关系模型、网状模型和层次模型是常用的三种数据模型，它们的区别是表示信息的方式。关系模型只用了数据记录的内容，而层次模型和网状模型要用到记录间的联系以及它们再存储结构中的布局。关系模型中记录之间的联系通过多个关系模式的公共属性来实现。如果建立了关系数据库，用户只要用关系数据库提供的查询语言发出查询指令，通知系统查询目标，具体实现的过程由系统自动完成，用户不需要了解记录的联系及顺序。关系数据库提供了较好的数据独立性。层次数据库和网状数据库中，记录之间的联系用指针实现，数据处理只能是过程化的，程序员的角色类似于导航员，程序与数据结构之间的联系过于紧密，大大降低了数据的独立性。

综合起来，关系型数据库的优势在于：

- 使用简便，处理数据效率高；
- 数据独立性高，有较好的一致性和良好的保密性；
- 数据库的存取不必依赖索引，可以优化；
- 可以动态地导出和维护视图；
- 数据结构简单明了，便于用户理解和维护；
- 可以配备多种数据库接口。

由于关系模型有严格的数学基础，许多专家及学者在此基础上发展了关系数据理论。目前，关系模型已是非常成熟的数据模型，深受用户欢迎，20世纪80年代以来，几乎所有的数据库管理系统都是关系数据库管理系统。

## 小结

1. 本章的主要内容：

- a. 掌握从现实世界到概念模型和数据模型抽象的含义，不同的数据库模型，实体关系图画法，
- b. 理解数据库管理系统（DBMS）的功能及其工作过程

2. 本章的重点、难点：

实体关系图画法：

- 确定实体
- 决定各实体的属性和主键
- 确定实体之间的关系
- 完成 E-R 图

## 习题

### 思考题

掌握从数据模型、概念模型到数据模型的过程，掌握关系数据库系统的基本概念和 E-R 图的基本画法；了解多媒体数据库和数据仓库的基本概念。

1. E-R 图的作用是什么，请画出学生、课程和成绩这三个实体之间的 E-R 图。
2. 为什么说数据库系统具有程序和数据的独立性。
3. 说明关系数据库中域、元组、属性的关系。





综合题

某图书馆业务中，现有如下三张表格

(1) 借书证

借书证号	姓名	性别	电话	签发年月	是否有效
1001	冬冬	女	65450019	1983-1-9	是
1002	平平	男	65450020	1984-4-8	是
1003	海航	男	65450021	1985-2-5	是
1004	海平	女	65450023	1984-10-10	否
1005	海风	女	65450045	1986-3-20	是

(2) 图书

书号	书名	作者	单价
ENG01	精读	英名	14.40
ENG02	听力	登辉	23.00
CMP01	计算机基础	老板	34.43
HUM01	中国语文	哈桑	12.70
CMP02	ACCESS	近代	33.00
ECO01	政治经济学	潇洒	44.00

(3) 借阅登记

借阅流水号	借书证号	书号	借阅日期	还书日期
1	1001	ENG01	1983-1-9	1983-2-9
2	1002	ENG01	1984-4-8	1984-5-8
3	1003	CMP01	1985-2-5	1985-3-5
4	1001	CMP02	1984-10-10	1984-11-10
5	1002	CMP01	1986-3-20	1986-4-20
	.....	.....	.....	

要求：

- (1) 将每张表作为一个实体，写出每个实体包含的属性。
- (2) 指出每个实体的主关键字。
- (3) 画出这三个实体之间 E-R 图，标出实体之间的关联关系。