

数据库系统原理与设计

(第 3 版)

认识自己

- 分析自己的兴趣、喜好、倾向、长处及强项：
 - ◆ 动力来源：内向还是外向；
 - ◆ 资讯处理：喜欢更多的数据还是凭直觉；
 - ◆ 如何决策：喜欢思考还是凭感觉；
 - ◆ 如何组织：谋定而后动还是兴之所至？

数据库系统原理与设计

(第 3 版)

第4章 数据库建模

数据模型与数据库三级模式 (回顾)

- **数据模型**是一个描述**数据结构**、**数据操作**以及**数据约束**的**数学形式体系**(即**概念及其符号表示系统**)
- 根据**数据抽象**的不同级别, 将**数据模型**划分为3类:
 - **概念模型**: **E-R模型**
 - **逻辑模型(关系模型)**: **数据库模式** (关系模式的集合)
 - **物理模型**: **物理存储结构和存取方法**
- **数据抽象与数据库三级模式**
 - **物理层抽象**: **内模式(存储模式)**, **全体数据**的**物理存储结构和索引**
 - **逻辑层抽象**: **模式(逻辑模式)**, **全体数据**的**逻辑结构**
 - **视图层抽象**: **外模式(子模式或用户模式)**, **局部数据**的**逻辑结构**

数据模型与数据库三级模式 (回顾)

数据库设计的任务

根据应用需求和软、硬件环境的约束, 设计最优的数据库概念模型、逻辑模型(模式和外模式)、物理模型, 使之能够有效地存储和管理数据, 满足用户的信息要求和处理要求。

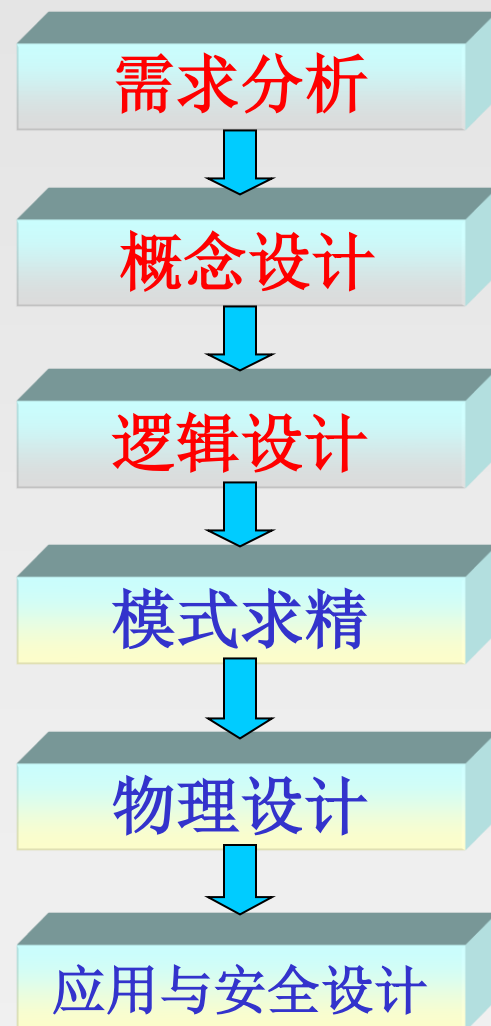
- 物理层抽象: 内模式(存储模式), 全体数据的物理存储结构和索引
- 逻辑层抽象: 模式(逻辑模式), 全体数据的逻辑结构
- 视图层抽象: 外模式(子模式或用户模式), 局部数据的逻辑结构

目 录

- 4.1 数据库设计过程
- 4.2 E-R模型基本概念及表示
- 4.3 约 束
- 4.4 弱实体集
- 4.5 扩展E-R特征
- 4.6 E-R建模问题
- 4.7 数据库概念设计实例——大学选课系统
- 4.8 逻辑设计——E-R模型转化为关系模型

数据库设计过程概述

- **数据库设计**就是根据各种应用处理的要求、硬件环境及操作系统的特性等，将现实世界中的数据**进行合理组织**，并利用已有的数据库管理系统(DBMS)来建立**数据库系统**的过程。
- 具体地说，对于一个给定的应用环境，构造出**最优的数据库逻辑模式**和**物理模式**，并建立**数据库及其应用系统**，使之能够有效地存储和管理数据，满足用户的信息要求和处理要求。



■ 整个数据库设计

■ 目标：了解和明确的数据需求，

● 功能需求

● 数据需求

● 数据使用业

● 数据联系及

● 性能需求

●

■ 方法：数据库设计者同应用领域的专家和用户进行深入沟通和交流。

■ 结果：形成用户的需求规格说明书。

运营：我需要国庆期间所有的订单数据...

数据员：要订单的哪些数据？订单号、用户ID、下单时间、预约取货时间、预约还货时间、订单状态？？

运营：哦哦，我要订单号、用户ID、下单时间、预约取货时间、预约还货时间、订单金额。

数据员：是下单时间在10.1-10.7号之间的订单数据？预约取货时间在10.1-10.7号之间的订单数据呢？还是实际取货时间在10.1-10.7号之间的订单数据？

运营：我要下单时间在10.1-10.7号之间的订单数据。

数据员：所有订单都要吗？还是只要预约成功的订单？

运营：只要预约成功的订单。

运营：你给我的数据，订单金额怎么那么大？是不是你拉错了？

数据员排查半天。。。

数据员：没有啊，这个就是订单金额。

运营思考半天。。。

数据员：你是不是没有排除优惠金额啊，订单金额要排除使用的优惠金额的。

概念设计

- **概念设计**是根据需求分析中得到的信息，运用适当的**工具**将这些需求转化为**数据库的概念模型**。
- **E-R模型**是Peter Chen于1976年提出的一种**语义模型**。该模型是基于对现实世界的这样一种认识：
 - 世界由一组称作**实体的基本对象及这些对象间的联系**组成。
- 由于**E-R模型**能将现实世界中**概念的含义和相互关联**映射到**数据库概念模型**，因此许多数据库设计工具都基于它进行扩展。
- 本章基于**E-R模型**进行**数据库概念设计**，其目的是通过**实体、联系、属性**等概念和工具精确地描述系统的**数据需求、数据联系及约束规则**。
- 结果：**E-R图（实体、联系及属性）**

逻辑设计

- 本步骤是将数据库的**概念模型**转化为所选择的数据库管理系统支持的**逻辑数据模型**，即**数据库模式**。
- 常见逻辑模型
 - 层次
 - 网状
 - **关系数据库**（本书讨论的对象）
 - 面向对象
 - XML

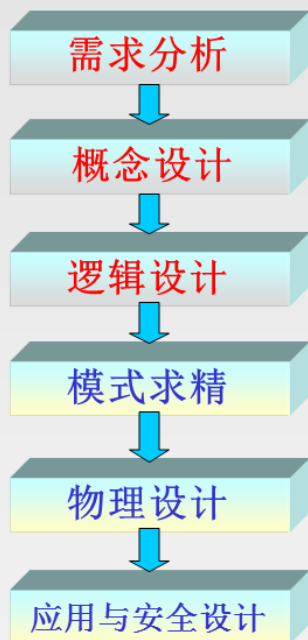
数据库模式-提炼

模式

studentNo	StudentName	courseNo	courseName	score
S0700001	李小勇	C001	高等数学	98
S0700001	李小勇	C002	离散数学	82
S0700001	李小勇	C006	数据库系统原理	56
S0700002	刘方晨	C003	计算机原理	69

■ **模式求精**的任务是分析并发现数据库**逻辑模式**存在的问题，并进行改进和优化。

■ 对于关系数据库而言，**模式求精**是以**关系数据理论**做指导，对已得到的**关系数据库模式**进行分析，找出潜在的可以改进和优化，如**减少数据冗余**，**消除更新、删除异常**等。

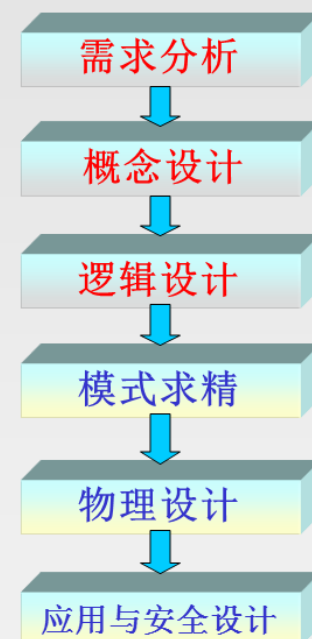


模式求精与**概念设计**是根据**主观认识**进行的，但**模式求精**是基于**关系理论**对相关**逻辑模式**进行优化。

物理设计

■ 考虑数据库要支持的负载和应用需求，为**逻辑数据库**选取一个最适合现实应用的**物理结构**，包括：

- 数据库文件的组织格式
- 文件内部的存储结构
- 建立索引
- 表的聚集
-



应用与安全设计

■ 一个数据库系统必须指出：

- 哪些用户可以访问数据库？
- 他们通过哪些存储过程访问数据库？
- 每个用户在每个过程中扮演什么样的角色？
- 对于每个角色，他们能够存取数据库的哪些部分和不能存取哪些部分（访问权限）？

完整的数据库设计是不可能一蹴而就的，
它往往是上述6个步骤的不断反复的过程！

目 录

- 4.1 数据库设计过程
- 4.2 **E-R模型基本概念及表示**
- 4.3 约 束
- 4.4 弱实体集
- 4.5 扩展E-R特征
- 4.6 E-R建模问题
- 4.7 数据库概念设计实例——大学选课系统
- 4.8 逻辑设计——E-R模型转化为关系模型

实体与实体集

■ **实体**是客观世界中可区别于其他事物的“事物”或“对象”。

■ 实体特征:

- **独立存在**。一个实体的存在不依赖于其它实体。例如，一名学生的存在不取决于其它学生实体是否存在。

- **可区别于其**
实体的特征
“李小勇”

实体/实体集?

一标识每个
地标识学生

■ **实体**可以是**有形的、实在的**事物，如一名**教师**、一本**书**等；

■ 也可以是**抽象的、概念上的**事物，如一门**课程**、一个**专业**，以及一次**订货**、**借书**、**选课**、**存款**或**取款**等业务产生的**单据**。

■ 但是，二者都应是组织或机构“感兴趣”的事物。

■ **实体集**是具有相同类型即相同性质(或属性)的实体组成的集合

属性

- **属性**是实体集中每个实体都具有的特征描述。
- 一实体集中所有实体都具有相同的属性。
 - 例如，**学生**实体集中的每个实体都具有：**学号、姓名、性别、出生日期、年龄、所学专业、电话号码、家庭住址、所在班级**等属性。
- 对每个属性来说，一个实体都拥有自己的**属性值**。
- 每个属性所允许的取值范围或集合称为该**属性的域**。

属性名	属性类别	域及约束	实例
学号	主码	char(10) , 10位数字组成, 其中第1位数字代表学生类别, 如: 1-本科生, 2-硕士生, 3-博士生, 4-独立学院本科生, 5-专科生; 接下来4位数字代表入学年份; 最后5位数字为序号。不允许取空值	1201600258
姓名		varchar(20) , 不允许取空值	李小勇
性别		char(2) , 取值范围: {'男', '女'}	男
出生日期		datetime , 取值范围: 1900-01-01~当前	1998-09-09
年龄	派生属性	smallint , 取值范围: 10~80	18
所学专业		varchar(30)	计算机科学与技术
电话号码	多值属性	varchar(13) , 每个电话号码由数字字符加连字符 '-' 组成	027-87009999
家庭地址	复合属性	varchar(60)	湖北省武汉市中山路56号
所在班级		char(8) , 前4位数字代表年级	20160803

图4-1 学生实体集的属性域及学生李小勇的属性值

属性分类

■ 简单属性和复合属性

- **简单属性**是指不能再分为更小部分的属性。
- **复合属性**是指可以进一步划分为更小部分的属性。
 - **学生**实体集的**家庭住址**可以进一步设计成包括**省份**、**城市**、**街道**等成分的属性。

■ 单值属性和多值属性

- 如果某属性对一个特定实体任何时候都只能有单独的一个值，则称该属性为**单值属性**，否则为**多值属性**。
 - **学号**为**单值属性**，因为一个学生的**学号**只有一个值。
 - **电话号码**为**多值属性**，因为一个学生可能有0个、1个或多个电话号码(如**实验室电话**、**宿舍电话**、**移动电话**等)。

属性分类

■ 派生属性

- 派生属性的值可以从其它相关属性或实体(集)派生出来。
 - 年龄为派生属性，因为一个学生的年龄属性的值可以由该学生的出生日期属性的值和当前日期计算得到。
 - 学生实体集的已修学分(该学生所选修课程的学分合计)也是派生属性，因为它的值可以通过统计该学生所选修课程的学分之和来获得。

学生 Student

<u>studentNo</u>	学号
studentName	姓名
sex	性别
birthday	出生日期
native	籍贯
nation	民族
creditSum	已修学分

课程 Course

<u>courseNo</u>	课程号
courseName	课程名
creditHour	学分
courseHour	课时数
priorCourse	先修课程

成绩 Score

<u>studentNo</u>	学号
<u>courseNo</u>	课程号
<u>term</u>	开课学期
score	成绩

成绩管理数据库 ScoreDB 的模式导航图

属性分类

■ 派生属性

- 派生属性的值可以从其它相关属性或实体(集)派生出来。
 - 年龄为派生属性，因为一个学生的年龄属性的值可以由该学生的出生日期属性的值和当前日期计算得到。
 - 学生实体集的已修学分(该学生所选修课程的学分合计)也是派生属性，因为它的值可以通过统计该学生所选修课程的学分之和来获得。

■ 空值(NULL)

- 当实体在某个属性上没有值时可使用NULL值，表示属性的值是未知的或不存在。
 - 不知道某学生的专业(所学专业的值未知)或该学生还没有分配专业(所学专业的值不存在)，则该学生所学专业的值为NULL。
 - 某学生的出生日期为NULL，则认为该值是缺失的(即值未知)。

实体集及属性的E-R图表示

■ 在E-R图中，实体集和属性分别表示如下：

- 实体集——矩形
- 属性——椭圆
- 多值属性——双椭圆
- 派生属性——虚线椭圆
- 属性与实体之间——连线

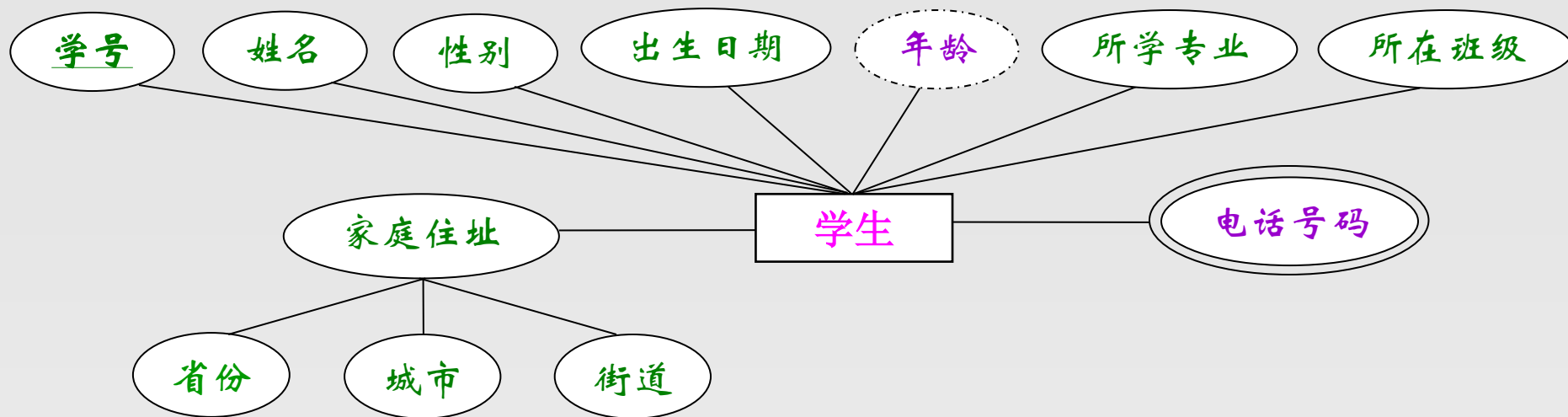


图4-2 学生实体集 E-R图表示

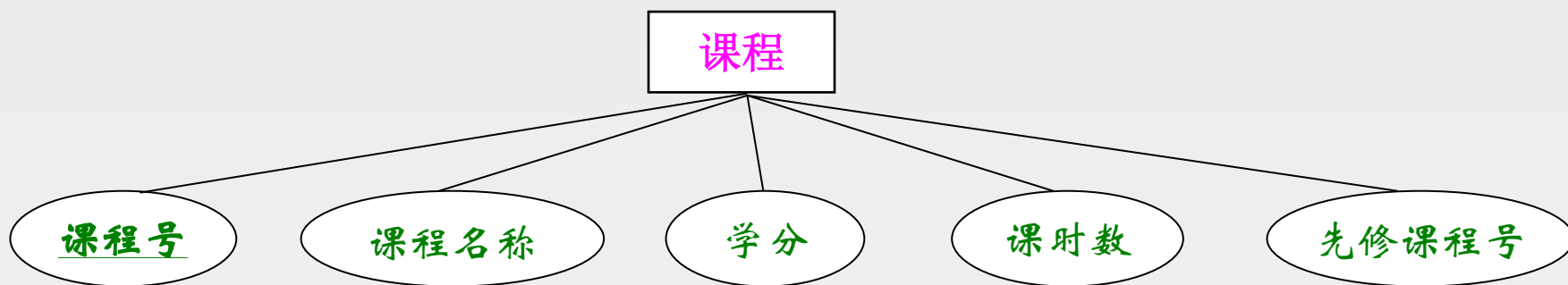


图4-3 课程实体集 E-R图表示

多值属性的变换

- 将多值属性转换为多个单值属性进行建模。例如，可将多值属性电话号码建模为移动电话、宿舍电话、实验室电话、家庭电话4个单值属性，如图4-4所示。

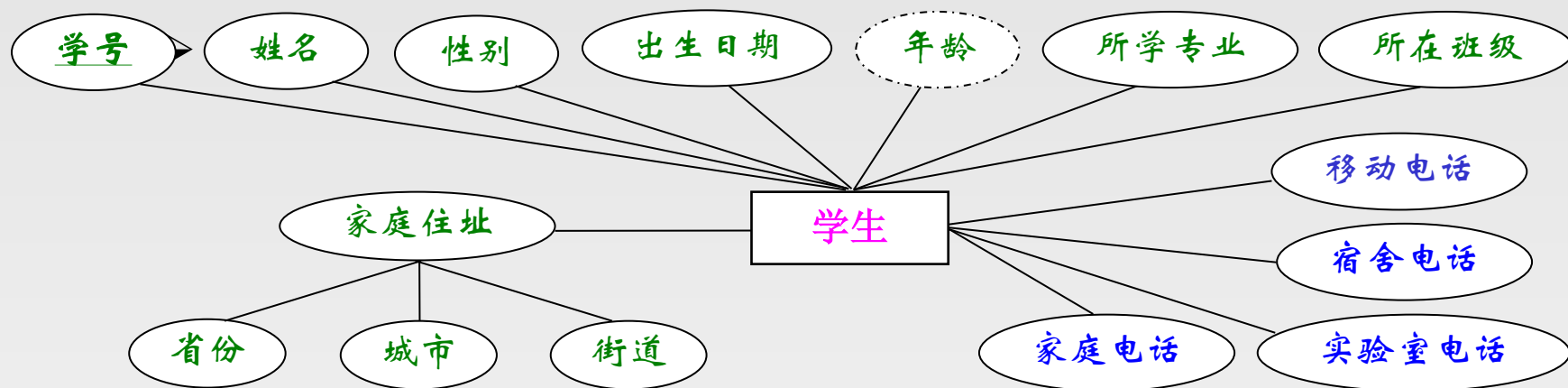


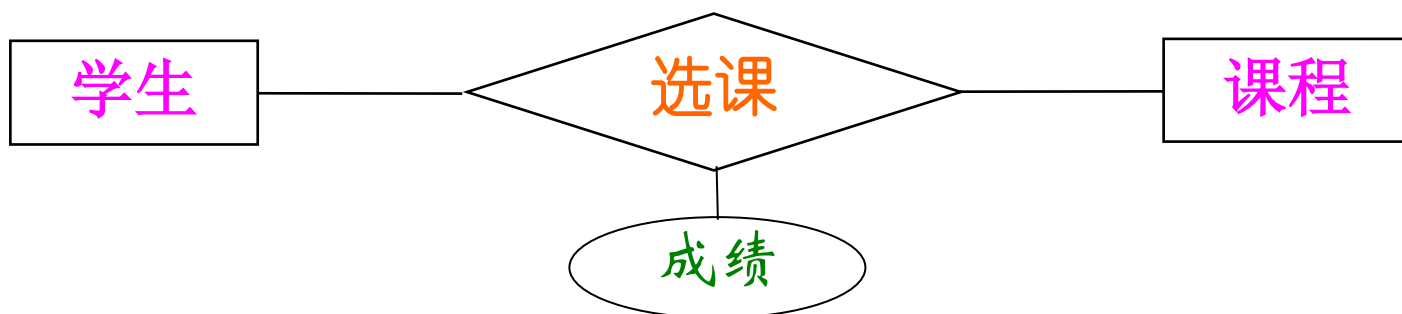
图4-4 学生实体集中的多值属性转换为多个单值属性表示

- 将多值属性单独建模为一个弱实体集，它依赖于原实体集而存在。请参见4.4节。

联系与联系集

- **联系**(relationship)是指多个**实体**间的**相互关联**。例如，**学生李小勇**选修了**数据库系统原理**课程。
- **联系集**是**同类联系**的集合。
- 形式化地说，**联系集**是 n ($n \geq 2$)个**实体集**上的**数学关系**，这些**实体集**不必互异。如果 E_1, E_2, \dots, E_n 为 n 个**实体集**，那么**联系集** R 是集合 $\{(e_1, e_2, \dots, e_n) \mid e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$ 的一个**子集**，而 (e_1, e_2, \dots, e_n) 是一个**联系**。
- **E-R图**用**菱形表示联系(集)**。**联系**可具有自身的**描述属性**。

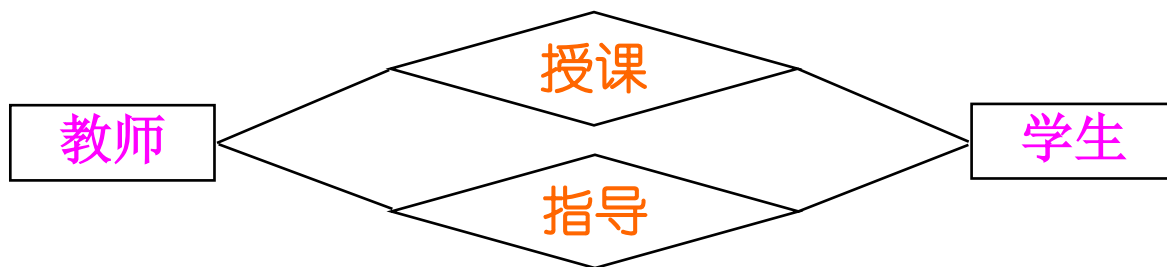
联系与联系集



- 形式化地说，**联系集**是 n ($n \geq 2$)个**实体集**上的**数学关系**，这些**实体集**不必互异。如果 E_1, E_2, \dots, E_n 为 n 个**实体集**，那么**联系集** R 是集合 $\{(e_1, e_2, \dots, e_n) \mid e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$ 的一个子集，而 (e_1, e_2, \dots, e_n) 是一个**联系**。
- E-R图**用菱形表示联系(集)**。**联系**可具有自身的**描述属性**。
 - 例如，可定义**学生**与**课程**实体集之间的**选课**联系集。联系属性**成绩**记录了**学生**选修**课程**的成绩。
 - 说明：图中省略了**实体集**的属性

多联系与联系中的实体角色

- **多联系**：在相同的实体集上可能存在多个不同的联系集。如教师与学生实体集之间同时存在授课、指导联系集。

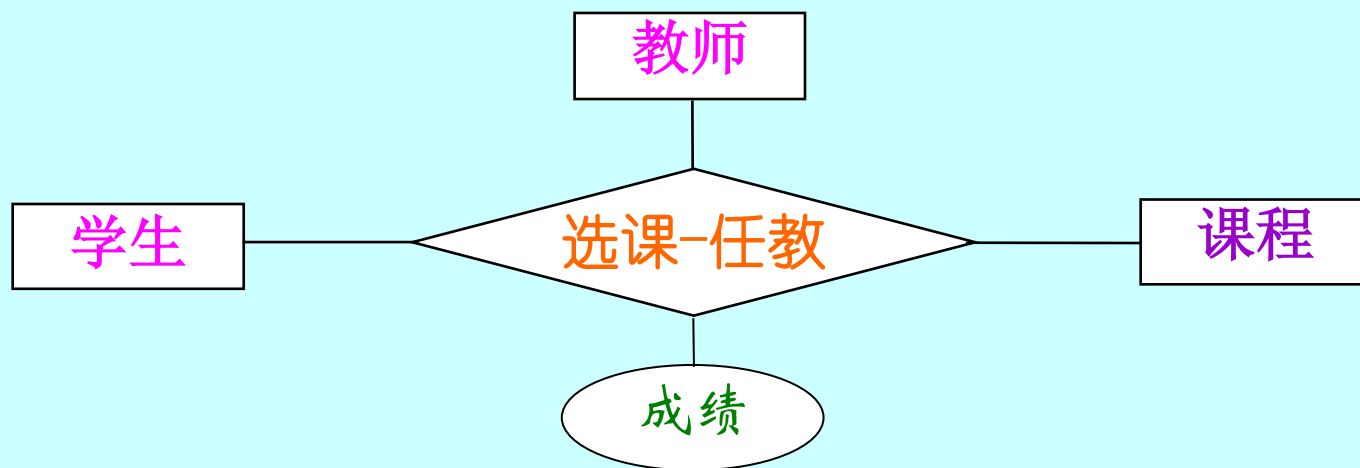


- **角色**：实体在联系中的作用。如课程实体集中的2门课程实体C1、实体C2之间存在先修要求联系，其中C1、C2扮演不同的角色，C1代表主课程，C2代表先修课程，分别表示实体所扮演的角色。



联系集的度

- 参与联系集的实体集的数目称为联系集的度
 - 如二元联系集的度为2，三元联系集的度为3。
- 联系集选课是二元(binary)联系集的例子，即涉及2个实体集的联系集。
- 如果还需要知道学生选修课程的任课教师信息，可以在联系集选课中增加教师实体集，形成一个新的三元联系集选课-任教。



目 录

- 4.1 数据库设计过程.....
- 4.2 E-R模型基本概念及表示.....
- 4.3 约 束.....
- 4.4 弱实体集.....
- 4.5 扩展E-R特征.....
- 4.6 E-R建模问题.....
- 4.7 数据库概念设计实例——大学选课系统.....
- 4.8 逻辑设计——E-R模型转化为关系模型.....

约 束

■ E-R模型的约束主要有：

- 映射约束
- 码约束
- 依赖约束
- 参与约束

■ 另外，还会介绍多值联系的概念。

映射约束

■ 映射基数

- 实体集 A 中的一个实体通过某联系集 R 能与实体集 B 中的实体相联系的数目，称为实体集 A 到实体集 B 之间的联系集 R 的映射基数(mapping cardinality)，简称为联系集 R 的映射基数。

- 说明：并不要求 A 与 B 是不同的实体集。

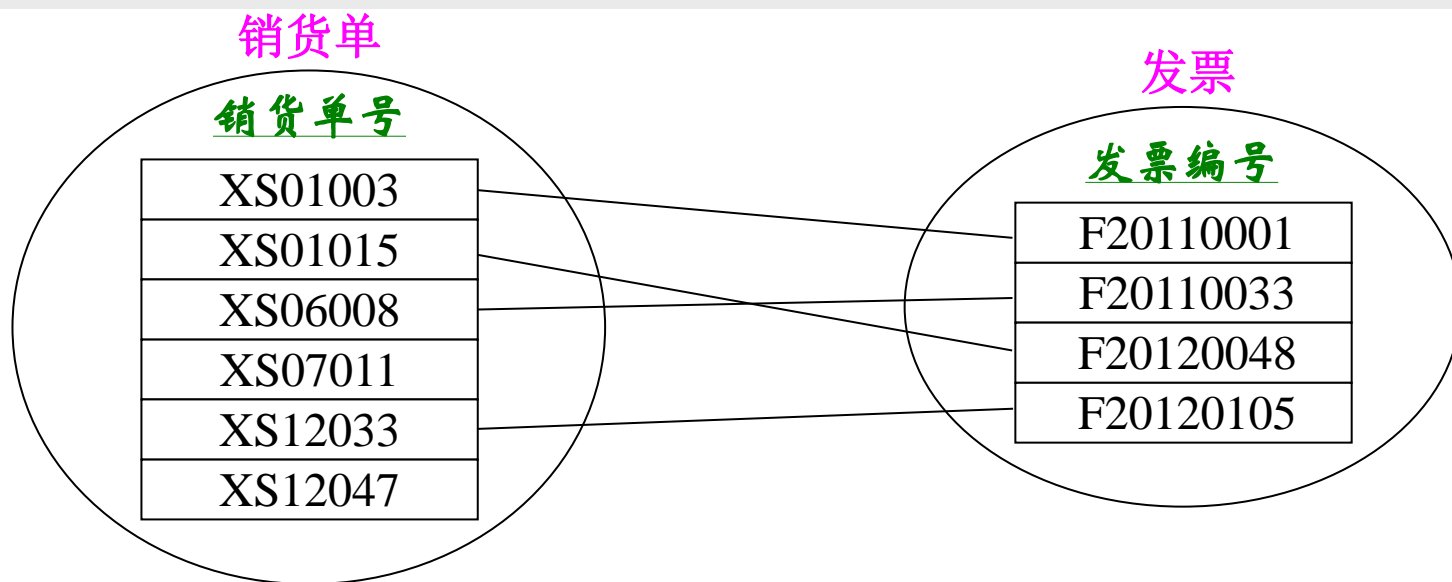
■ 在二元联系中，共有3种映射基数：

- 1:1（一对一）
- 1: m （一对多）——反过来看就是 $m:1$ （多对一）
- $m:n$ （多对多）

映射约束

■ **一对一**: A 中的一个实体至多(允许不)同 B 中的一个实体相联系, B 中的一个实体也至多(允许不)同 A 中的一个实体相联系。

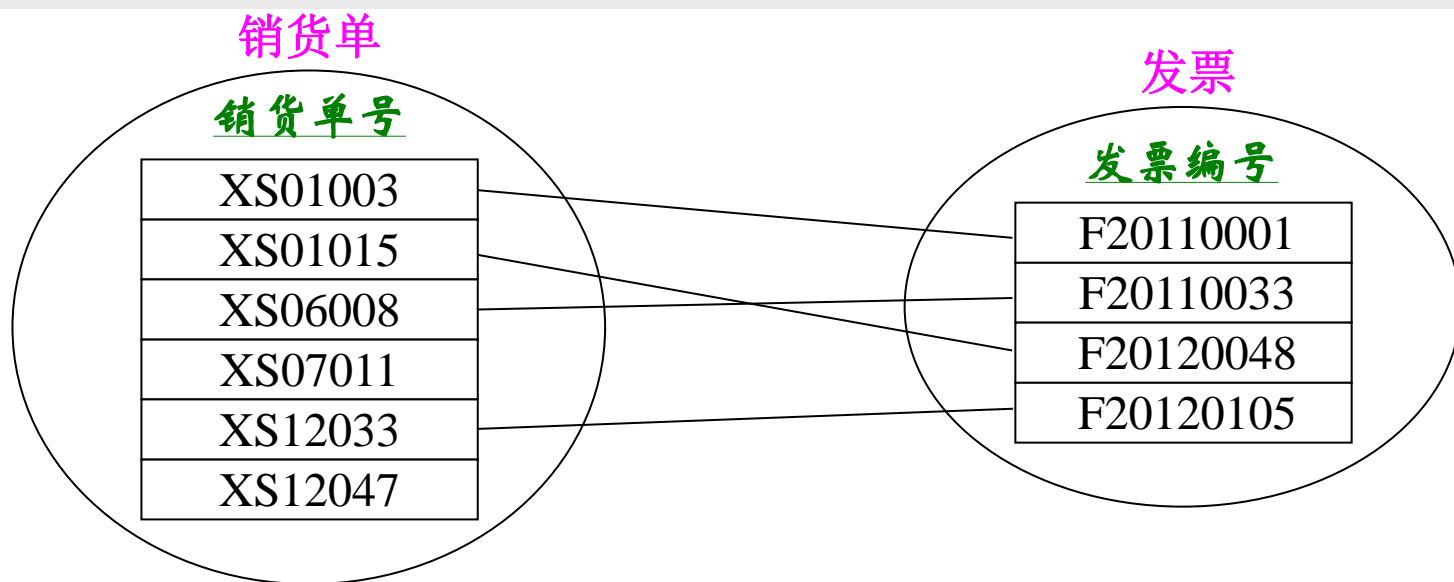
- 例如, 由实体集销货单和发票参与的联系集开发票, 假设一个销货单对应开一张发票, 则开发票为实体集销货单和发票之间的一对一联系集。



映射约束

■ **一对一**: A 中的一个实体至多(允许不)同 B 中的一个实体相联系, B 中的一个实体也至多(允许不)同 A 中的一个实体相联系。

◆ 说明: 存在部分销货单还没有开发票, 但每一张发票都有唯一对应的销货单。

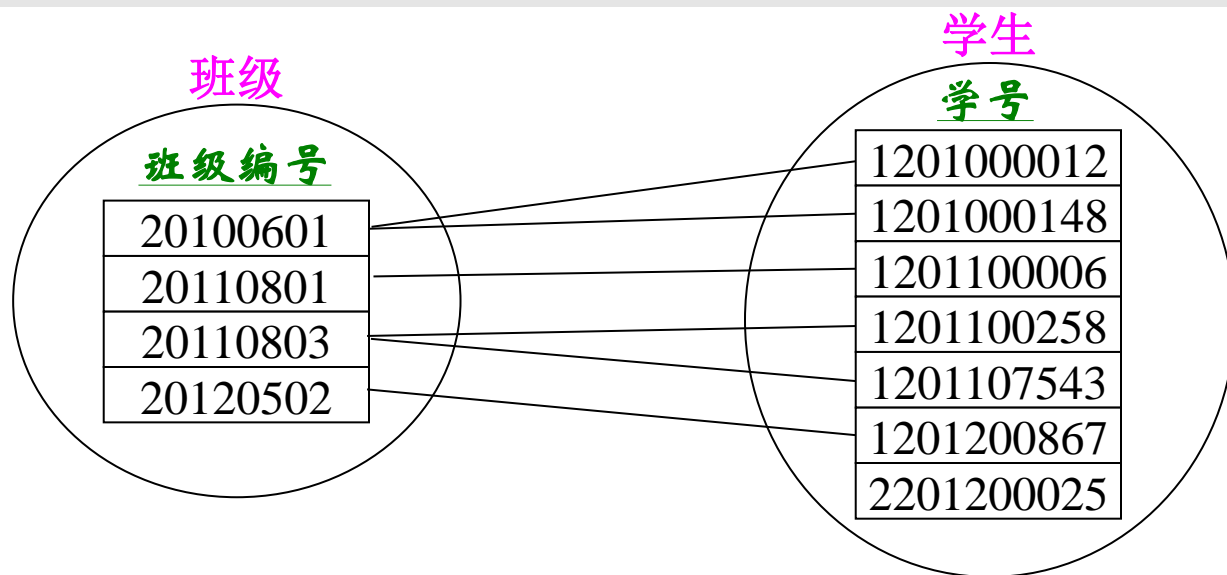


映射约束

- **一对多**：A中一个实体可以同B中任意数目(可以为0)的实体相联系，而B中一个实体至多(允许不)同A中一个实体相联系。
- 反过来看的**多对一**：A中一个实体至多(允许不)同B中一个实体相联系，而B中一个实体可以同A中任意数目(可以为0)的实体相联系。
- 例如，由实体集**班级**和**学生**参与的联系集**包含**，假设一个**班级**可以包含多名**学生**，但一名**学生**只能归属于某一个**班级**，则**包含**为从实体集**班级**到**学生**的**一对多**联系集，或为从实体集**学生**到**班级**的**多对一**联系集。

映射约束

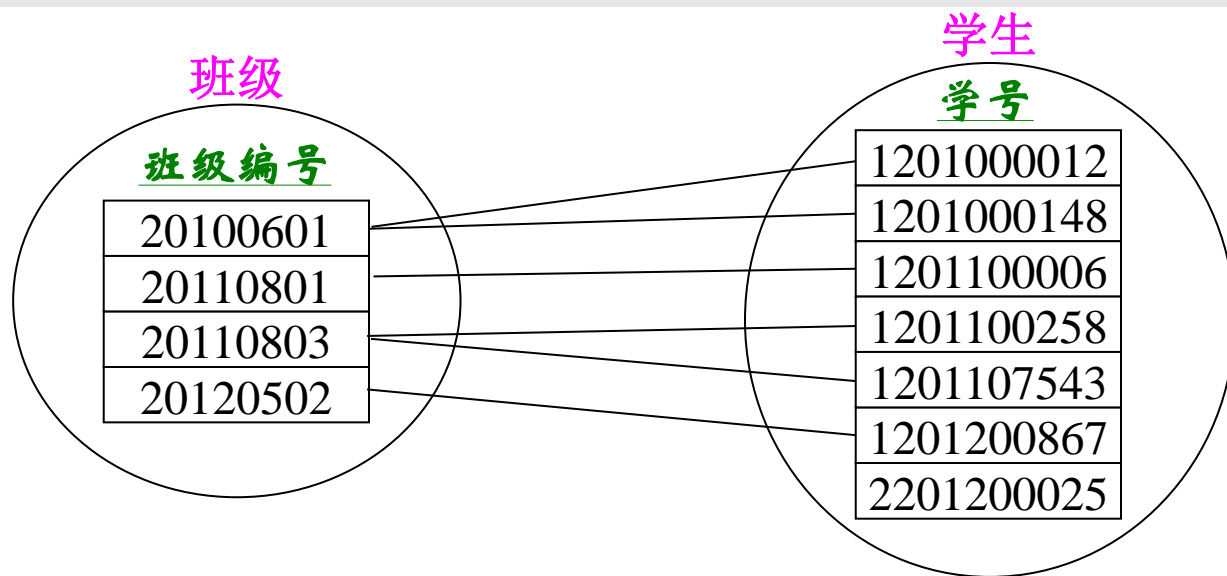
- **一对多**：**A**中一个实体可以同**B**中任意数目(可以为0)的实体相联系，而**B**中一个实体至多(允许不)同**A**中一个实体相联系。



- 例如，由实体集**班级**和**学生**参与的联系集**包含**，假设一个**班级**可以包含多名**学生**，但一名**学生**只能归属于某一个**班级**，则**包含**为从实体集**班级**到**学生**的**一对多**联系集，或为从实体集**学生**到**班级**的**多对一**联系集。

映射约束

- **一对多**：A中一个实体可以同B中任意数目(可以为0)的实体相联系，而B中一个实体至多(允许不)同A中一个实体相联系。

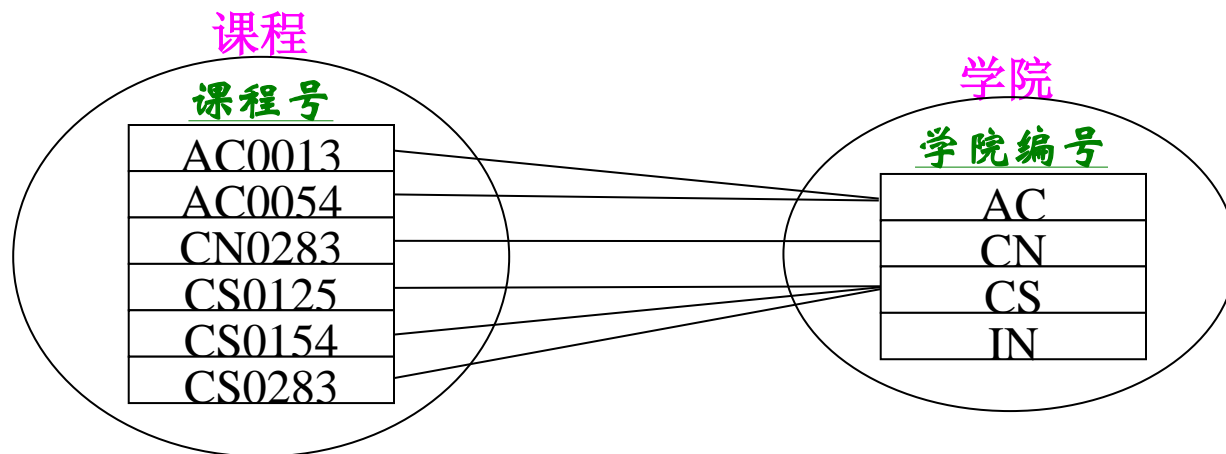


- 例如，由实体集**班级**和**学生**参与的联系集**包含**，**假设一个**

- ◆ 说明：每一个**班级**都包含**学生**，但**存在部分学生**没有对应的**班级**。

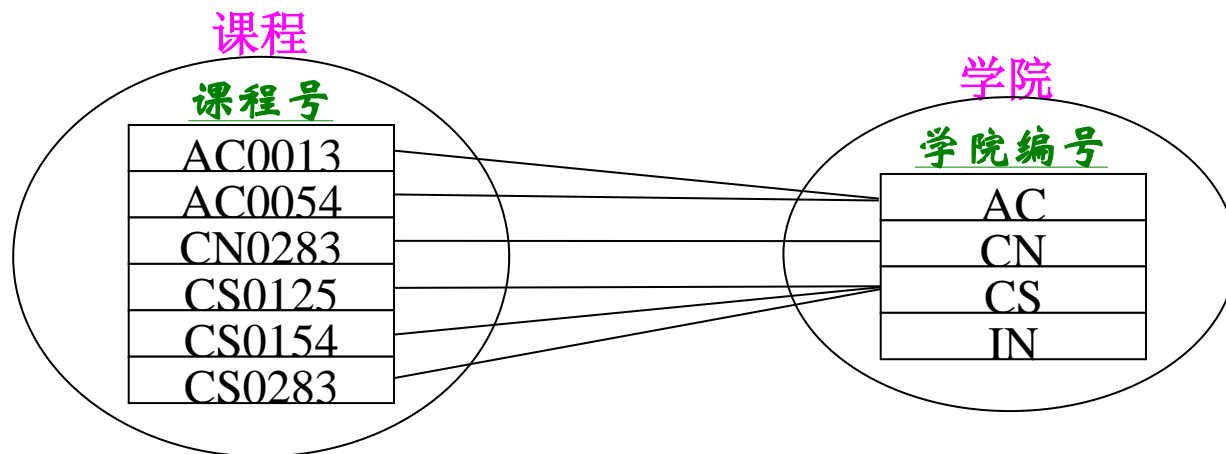
映射约束

- **一对多**：**A**中一个实体可以同**B**中任意数目(可以为0)的实体相联系，而**B**中一个实体至多(允许不)同**A**中一个实体相联系。
- 反过来看的**多对一**：**A**中一个实体至多(允许不)同**B**中一个实体相联系，而**B**中一个实体可以同**A**中任意数目(可以为0)的实体相联系。
- 再如，由实体集**课程**和**学院**参与的联系集**归属**，假设一门课程只能归属于一个学院，但一个学院可以负责多门课程，则**归属**为从实体集**课程**到**学院**之间的**多对一**联系集，或为从实体集**学院**到**课程**的**一对多**联系集。



映射约束

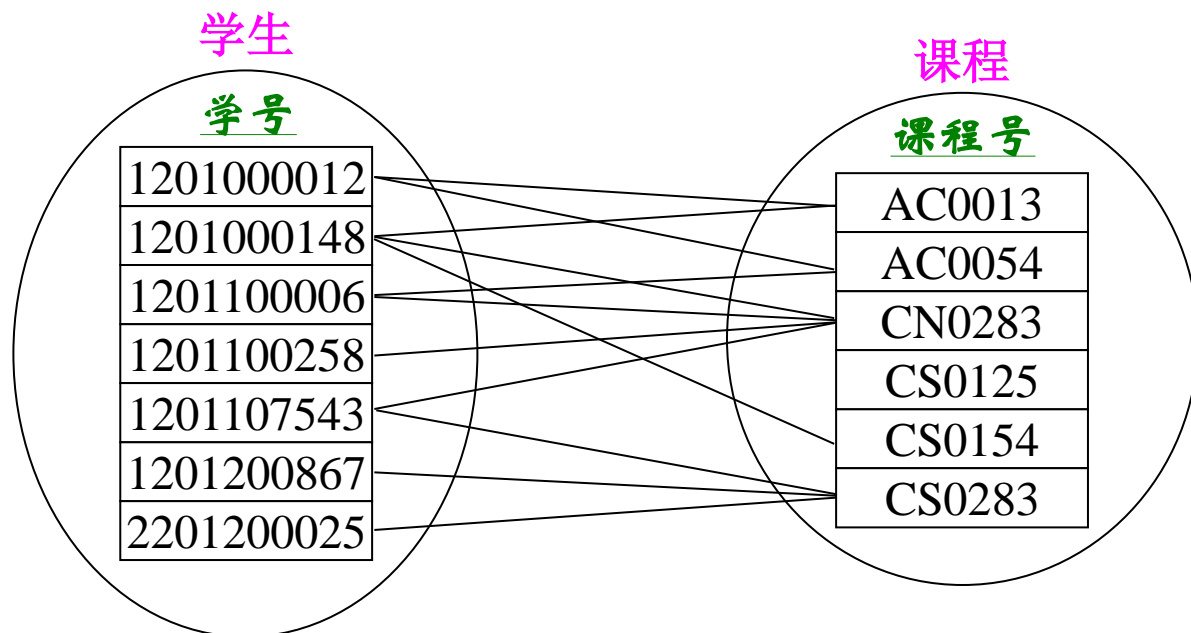
- **一对多**：**A**中一个实体可以同**B**中任意数目(可以为0)的实体相联系，而**B**中一个实体至多(允许不)同**A**中一个实体相联系。
- 反过来看的**多对一**：**A**中一个实体至多(允许不)同**B**中一个实体相联系，而**B**中一个实体可以同**A**中任意数目(可以为0)的实体相联系。
- 再如，由实体集**课程**和**学院**参与的联系集**归属**，假设一门
- ◆ 说明：每一门课程都有归属的学院，但存在部分学院没有负责的**课程**。



映射约束

■ **多对多**：A中的一个实体可以同B中任意数目(可以为0)的实体相联系，而B中的一个实体也可以同A中任意数目(可以为0)的实体相联系。

- 例如，由实体集学生和课程参与的联系集选课，假设一个学生可以选修多门课程，且一门课程允许多个学生选修，则选课为实体集学生与课程之间的多对多联系集。

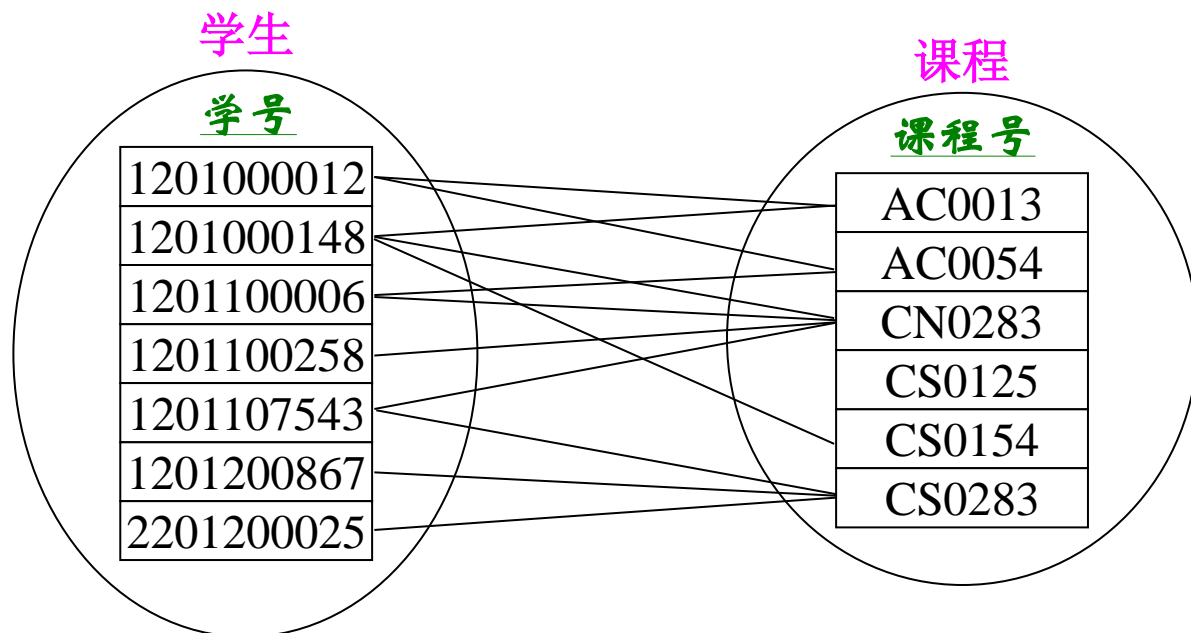


映射约束

■ **多对多**：A中的一个实体可以同B中任意数目(可以为0)的实体相联系，而B中的一个实体也可以同A中任意数目(可以为0)的实体相联系。

● 例如，由实体集学生和课程参与的联系集选课，假设一个

◆ 说明：每一个学生都选修了课程，但有的课程没有学生选修。

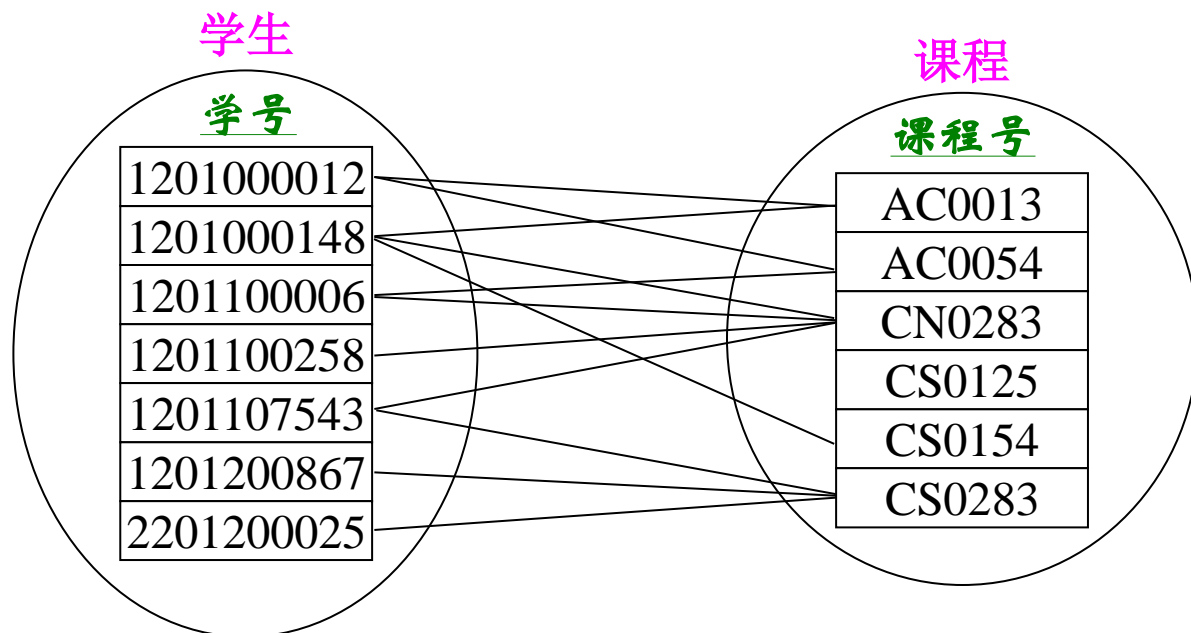


映射约束

■ **多对多**：A中的一个实体可以同B中任意数目(可以为0)的实体相联系，而B中的一个实体也可以同A中任意数目(可以为0)的实体相联系。

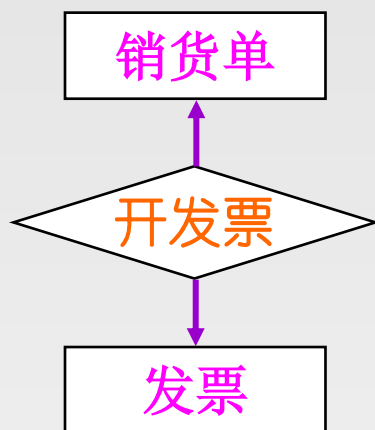
● 例如，由实体集学生和课程参与的联系集选课，假设一个

◆ 说明：每一个学生都选修了课程，但有的课程没有学生选修。

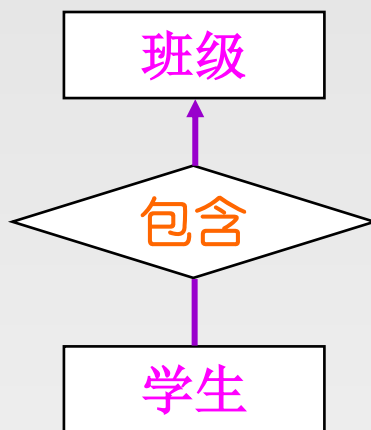


映射基数的表示

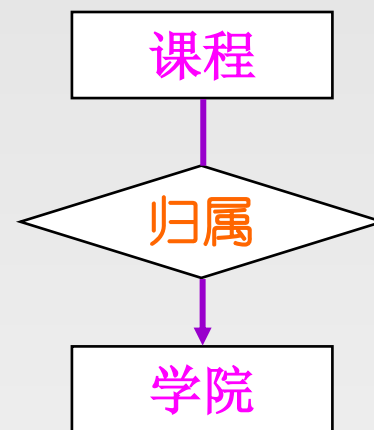
- 在E-R图中，为了反映联系集的映射基数，采用“→”表示指向参与联系集中的“一”方实体集，线段“——”表示参与联系集中的“多”方实体集。



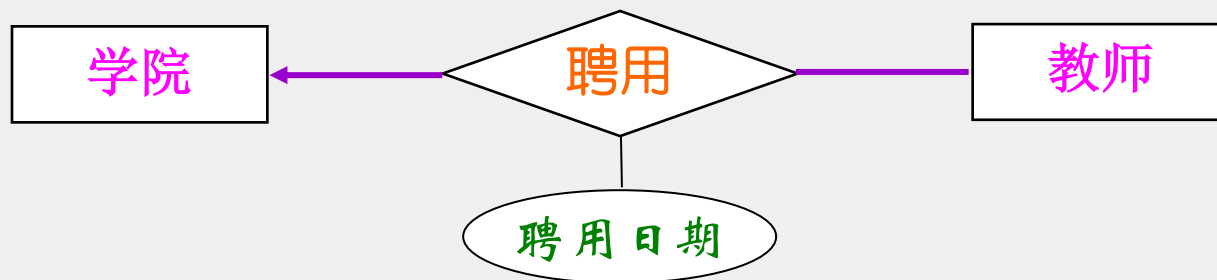
(a) 一对一联系集开发票



(b) 一对多联系集包含



(c) 多对一联系集归属



约 束

■ E-R模型的约束主要有：

- 映射约束小结（多种映射关系1:1,1:N,M:N）
- 码约束
- 依赖约束
- 参与约束

■ 另外，还会介绍多值联系的概念。

码约束与联系集的属性安置

■ 实体集的码

- **超码**(super key)是一个或多个**属性**的集合，且这些**属性**的集合在一**实体集**中能够**唯一标识**一个**实体**。
 - 如果**SK**是一个**超码**，那么**SK**的**任意超集**(即包含**超码**的**属性集**)也是**超码**
- 若一**超码**的**任意真子集**都不能成为**超码**，则称该**最小超码**为**候选码**(candidate key)。
 - 给定一组**属性集**，可能存在多个**候选码**。
 - 假设**学生**实体集中**不存在同名的学生**，则**学号**和**姓名**都可以用来**唯一区分学生**，因此它们都可以作为**学生**实体集的**候选码**。
 - **学号**和**姓名**的组合是候选码吗？

码约束与联系集的属性安置

- **候选码**和**超码**是实体集客观存在的特性，它们表示实体集中任意两个**实体**不允许在**码属性**上有相同的值。
- 可以从多个**候选码**中选择一个作为**实体集**的**主码**(primary key)。**主码**是被数据库设计者主观选中、用来区分同一实体集中不同实体的候选码。
- 当一**实体集**存在多个**候选码**时，**主码**选择原则：
 - 选择**属性**长度最短的**候选码**；
 - 选择**包含单个属性**的**码**，而不是复合**候选码**；
 - 选择在数据库系统生命周期内**属性值**最少变化的**候选码**；
 - 选择在数据库系统生命周期内**更可能包含唯一值**的**候选码**。

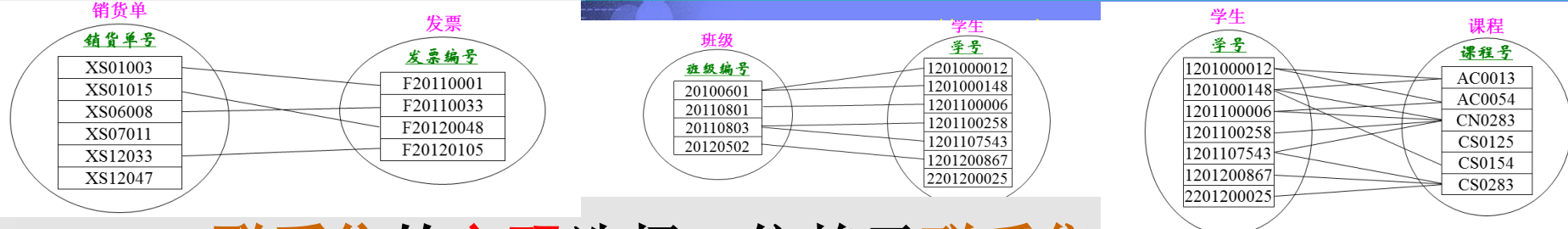
联系集
码约束

码约束与联系集的属性安置

- 一个联系集中的每一个联系也要求是可区分的，也必须存在候选码，并选择其中的一个候选码作为主码，用于唯一标识该联系集中的一个联系。

■ 联系集的码

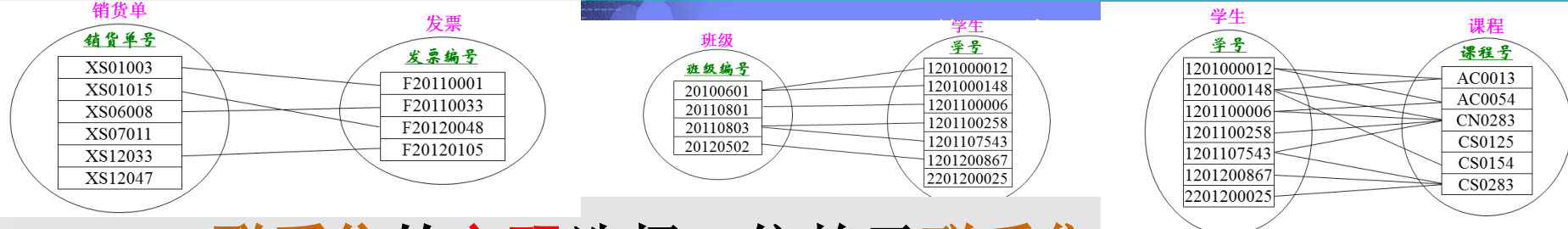
- 假设有一个涉及 n 个实体集 E_1, E_2, \dots, E_n 的 n 元联系集 R ， $PK(E_i)$ 代表构成实体集 E_i 的主码属性的集合，而 $(e_1, e_2, e_3, \dots, e_n)$ 是联系集 R 的一个联系，那么 $PK(E_1) \cup PK(E_2) \cup \dots \cup PK(E_n)$ 构成联系集 R 的一个超码。



■ 二元联系集的主码选择，依赖于联系集的映射基数

- 一对一：主码可以使用参与联系集中的任何一方实体集的主码；
- 一对多(多对一)：主码由“多”的一方实体集的主码组成；
- 多对多：主码由参与联系集中所有实体集的主码组成。

对于多元联系集，它的主码的确定要复杂得多，它依赖于各实体集参与联系的实际语义！



■ 二元联系集的主码选择，依赖于联系集的映射基数

- 一对一：主码可以使用参与联系集中的任何一方实体集的主码；
- 一对多(多对一)：主码由“多”的一方实体集的主码组成；
- 多对多：主码由参与联系集中所有实体集的主码组成。

■ 二元联系集的属性安置

- 一对一联系集的属性：安置于联系集或任一边的实体集上；
- 一对多联系集的属性：安置于联系集或多的一方实体集上；
- 多对多联系集的属性：它是描述相关联实体集间的交互性语义，因此，联系属性只能安置于联系集上，不能放到相关联的实体集上去。

主码的标识

例 1. 某田径运动会组委会需要一运动会管理系统，现提出如下需求。该系统中存在运动队和运动会两方面的实体。

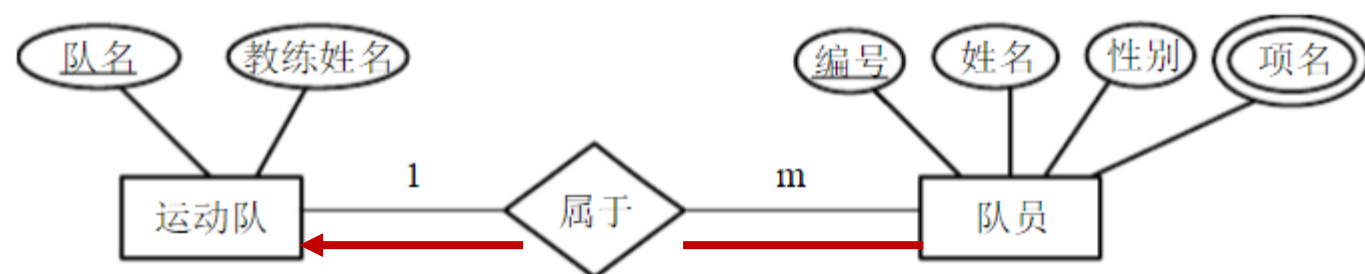
1. 运动队方面

运动队：队名、教练姓名

队员：编号、姓名、性别、项名

其中，一个运动队有多个队员，一个队员仅属于一个运动队，一个队一般有一个教练，一个队员可参加多个项目

运动队局部 ER 图：



约 束

■ E-R模型的约束主要有：

- 映射约束
- 码约束小结 实体的码，联系的码
- 依赖约束
- 参与约束

■ 另外，还会介绍多值联系的概念。

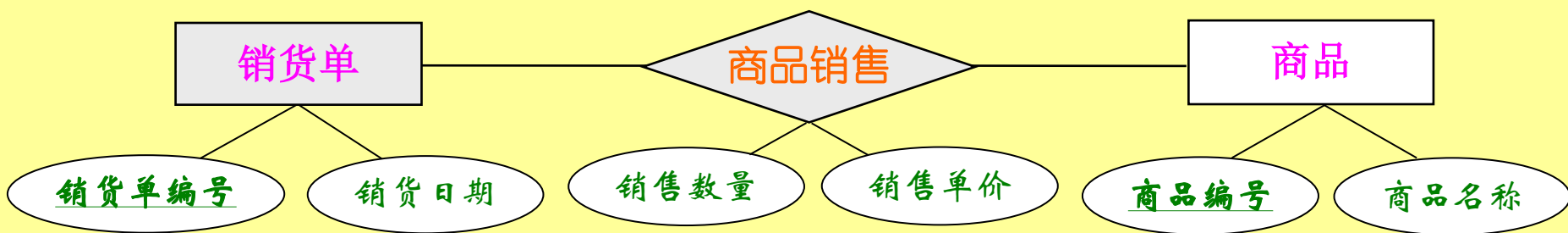
依赖约束

■ **依赖约束**是指联系中一种实体的存在依赖于该**联系集中联系**或其他**实体集中实体**的存在。

- 联系中一种实体的存在依赖于该**联系集中联系**的存在，称为**实体集与联系集之间的依赖约束**，并将**依赖于联系集而存在的实体集**称为**依赖实体集**；
- 联系中一种实体的存在依赖于其他**实体集中实体**的存在，称为**实体集之间的依赖约束**，并将**依赖于其他实体集而存在的实体集**称为**弱实体集**(见4.4节)。

实体集与联系集之间的依赖约束

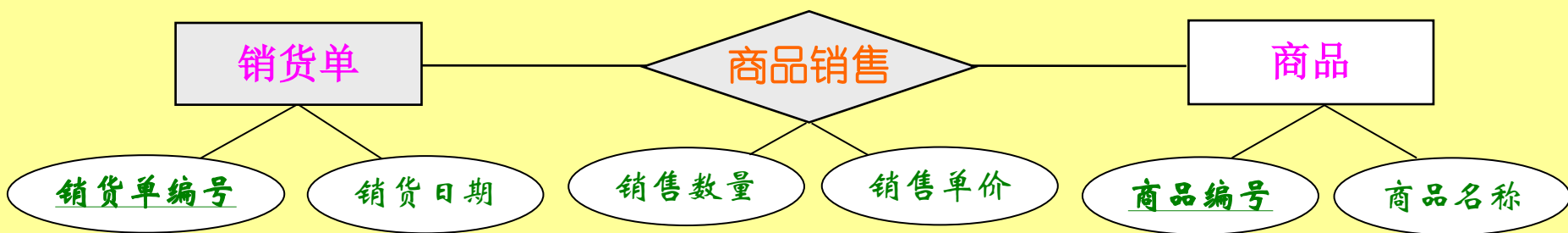
- 伴随着商品销售业务的发生,会产生销货单(或购货单)。
- 如果将销货单建模为实体集,则在销货单与商品实体集之间存在着多对多的商品销售联系集(反映一张销货单或一种商品的销售明细),联系属性有销售数量、销售单价等。
- 销货单实体集的存在是依赖于商品销售联系集的存在,也就是说,没有商品销售联系,就没有销货单实体,即销货单实体集与商品销售联系集之间存在依赖约束,销货单是依赖实体集。



实体集与联系集之间的依赖约束

◆ 为了区分，本书约定**依赖实体集**采用带填充背景的矩形表示，它所依赖的**联系集**用带填充背景的菱形表示(也可以不去区分依赖实体集和它所依赖的联系集)。

■ **销货单**实体集的存在是依赖于**商品销售**联系集的存在，也就是说，没有**商品销售**联系，就没有**销货单**实体，即**销货单**实体集与**商品销售**联系集之间存在**依赖约束**，**销货单**是**依赖实体集**。

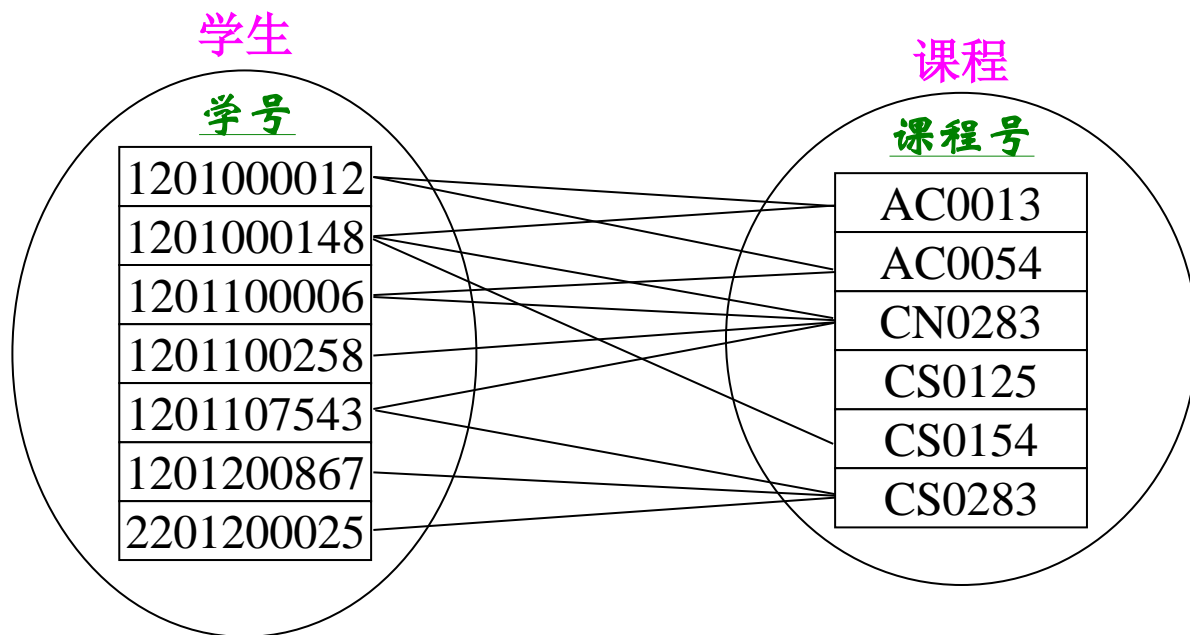


- 映射约束
- 码约束小结 实体的码, 联系的码
- 依赖约束
- 参与约束

参与约束

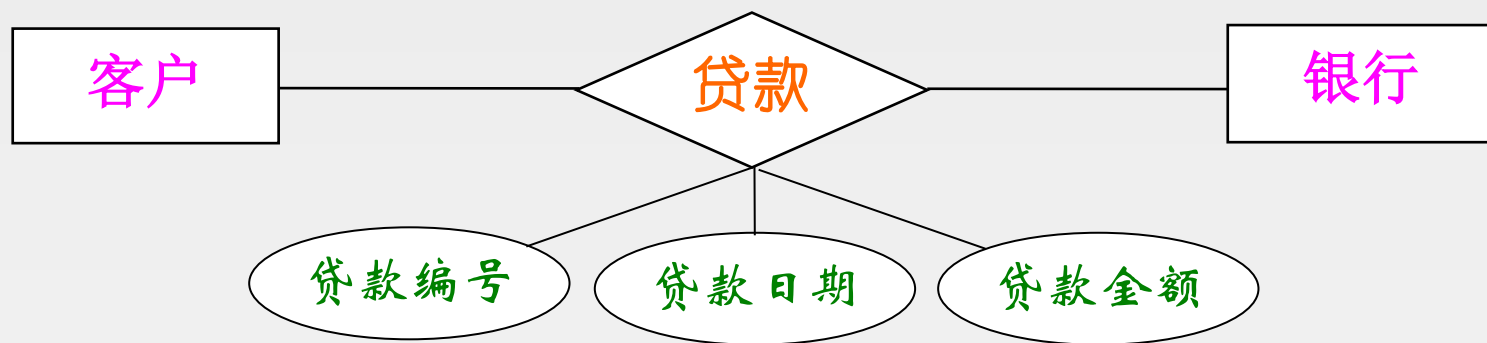
■ 另外, 还会介绍多值联系的概念。

- 如果**实体集A**中的**每个实体都参与**到**联系集R**中**至少一个联系**中, 则称**实体集A全部参与**联系集**R**。
- 如果**实体集A**中只有**部分实体参与**到**联系集R**的联系中, 则称**实体集A部分参与**联系集**R**。
- 例如, 对于**选课**联系集, **学生**实体集是**全部参与**, 而**课程**实体集是**部分参与**。



多值联系

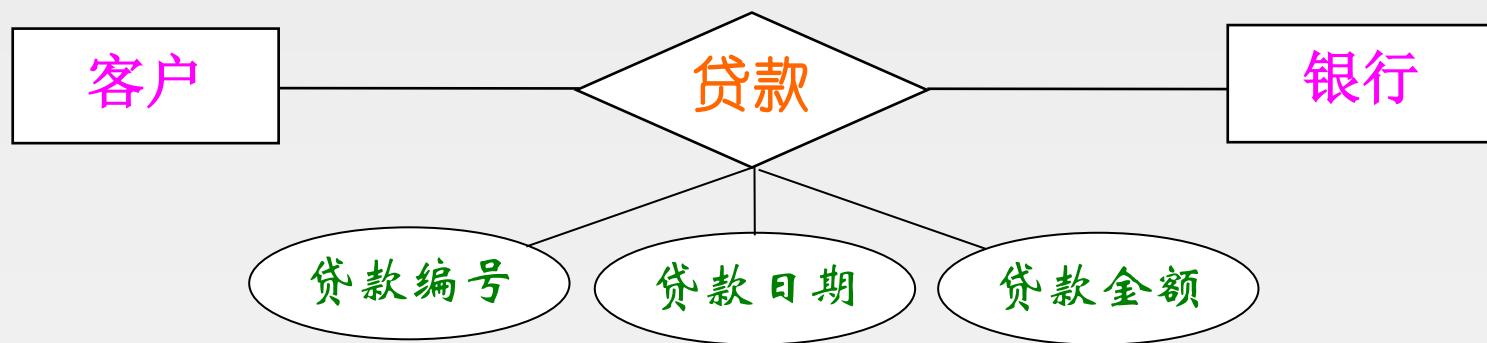
- **多值联系**是指在同一个给定的联系集中，相关联的相同实体之间可能存在多个联系。
- 例如，实体集**客户**与**银行**之间的**多对多贷款**联系集，表示一个客户可以向多个银行贷款，同时一个银行可以向多个客户发放贷款，联系集的属性有：**贷款编号**、**贷款日期**、**贷款金额**等。



多值联系

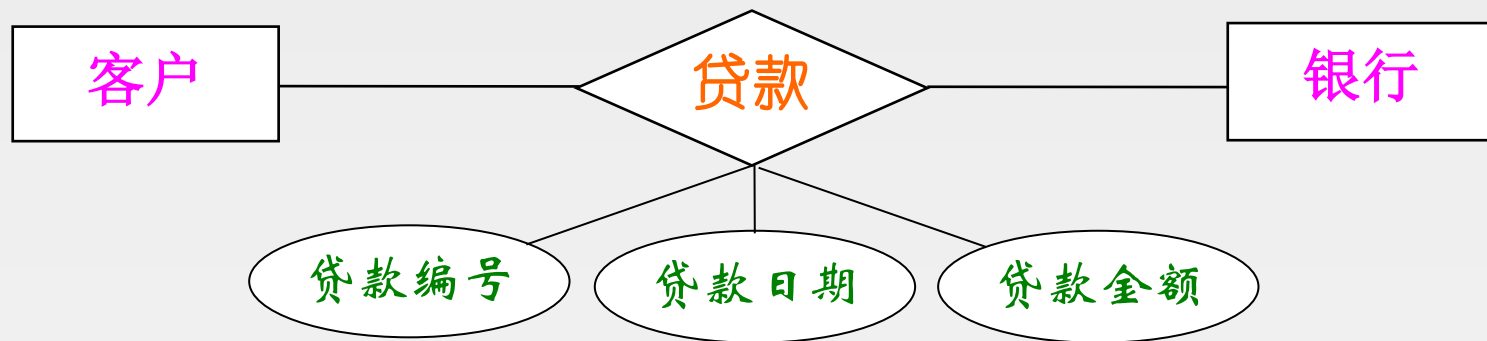
■ 该E-R模型存在如下问题:

- 当一个客户向同一个银行申请多笔贷款时, 则联系集中无法唯一标识一个联系。即贷款不仅是一个多对多联系, 而且是一个多值联系;
- 如果由多个银行联合发放一笔贷款, 或由多个客户共同借一笔贷款, 则会出现数据冗余问题(在联系集中反映该笔贷款的贷款编号、贷款日期等要重复多遍)。



■ 多值联系解决方案:

- 一种直观的解决办法: 从多值联系的联系属性中选择若干个标识性属性与联系集相关联的实体集的主码一道构成多值联系的主码。例如, 选择{客户编号, 银行编号, 贷款编号}作为联系集的主码, 其中贷款编号为联系属性, 它用于区分同一个客户在同一个银行发生的多次贷款业务。
- 该解决办法违背了E-R模型关于联系集主码的确定原则, 而且仍然没有解决数据冗余的问题。
- 一种更好的解决办法就是, 将多值联系建模为依赖实体集或弱实体集, 见4.6.3节。

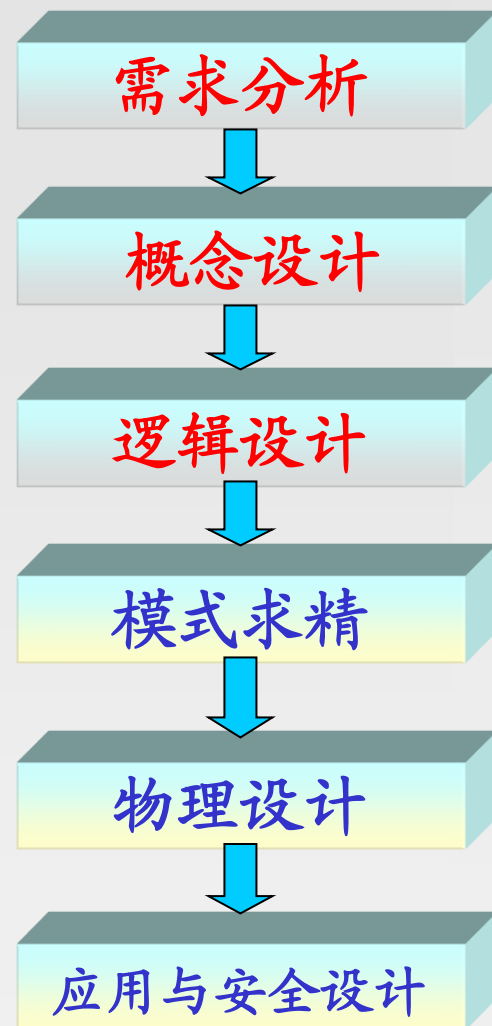


总结

■ 数据库设计过程

■ E-R模型

- 实体、属性与实体集（复合、多值属性）
- 联系、联系属性与联系集，联系集的超码
- 二元联系的主码
 - 一对一：参与联系的某一方实体集的主码；
 - 一对多：参与联系的“多”方实体集的主码；
 - 多对多：参与联系的所有实体集的主码。
- 二元联系的属性安置
- 映射基数（1:1、1:n、m:1、m:n联系）
- 依赖约束（依赖实体集、弱实体集）
- 参与约束（全参与、部分参与）
- 多值联系



数据库系统原理与设计

(第 3 版)

认识自己

- 分析自己的兴趣、喜好、倾向、长处及强项：
 - ◆ 动力来源：内向还是外向；
 - ◆ 资讯处理：喜欢更多的数据还是凭直觉；
 - ◆ 如何决策：喜欢思考还是凭感觉；
 - ◆ 如何组织：谋定而后动还是兴之所至？

目 录

- 4.1 数据库设计过程
- 4.2 E-R模型基本概念及表示
- 4.3 约 束
- 4.4 弱实体集
- 4.5 扩展E-R特征
- 4.6 E-R建模问题
- 4.7 数据库概念设计实例——大学选课系统
- 4.8 逻辑设计——E-R模型转化为关系模型

弱实体集

- 在现实世界中存在一类实体集，其属性不足以形成主码，它们必须依赖于其它实体集的存在而存在，我们称这样的实体集为**弱实体集** (weak entity set)。与此相对应，其属性可以形成主码的实体集称为**强实体集**。
- 弱实体集所依赖的强实体集称为**标识实体集** (identifying entity set)。弱实体集必须与一标识实体集相关联才有意义，该联系集称为**标识联系集** (identifying relationship set)。
- 一个弱实体集中用来标识弱实体的属性(集) 称为该弱实体集的**部分码** (partial key)。
- **弱实体集**中的实体是由其**标识实体集**中的**主码**与其**部分码**共同标识。

弱实体集

- E-R图使用**双矩形**表示**弱实体集**，**双菱形**表示**标识联系集**，**虚下划线**表示弱实体集的**部分码**。

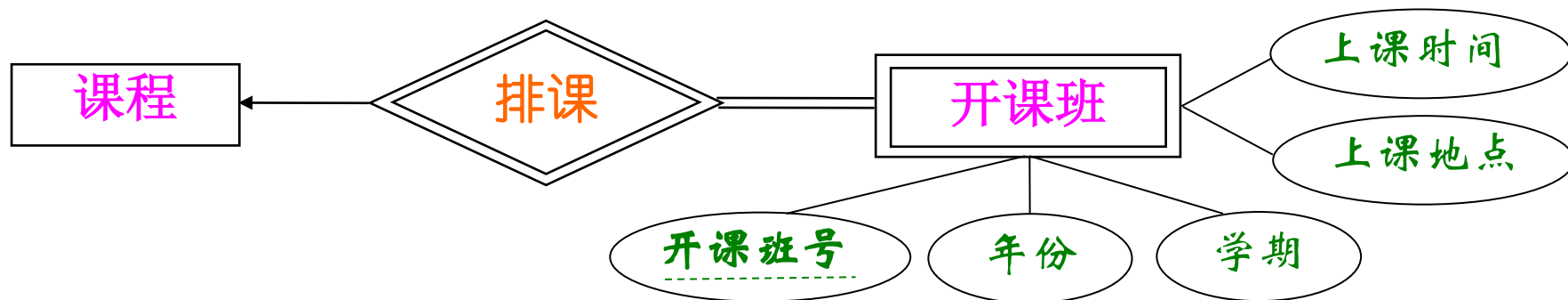


图4-17 弱实体集开课班与标识联系集排课

- 对于弱实体集，必须满足下列限制：
 - 标识实体集和弱实体集必须是“一对多”联系集。
 - 弱实体在标识联系集中是全部参与(双连线表示全部参与)。

弱实体集

- 考虑将多值属性电话号码建模为弱实体集联系电话，它有2个属性：电话号码、电话用途，其中电话号码为部分码。

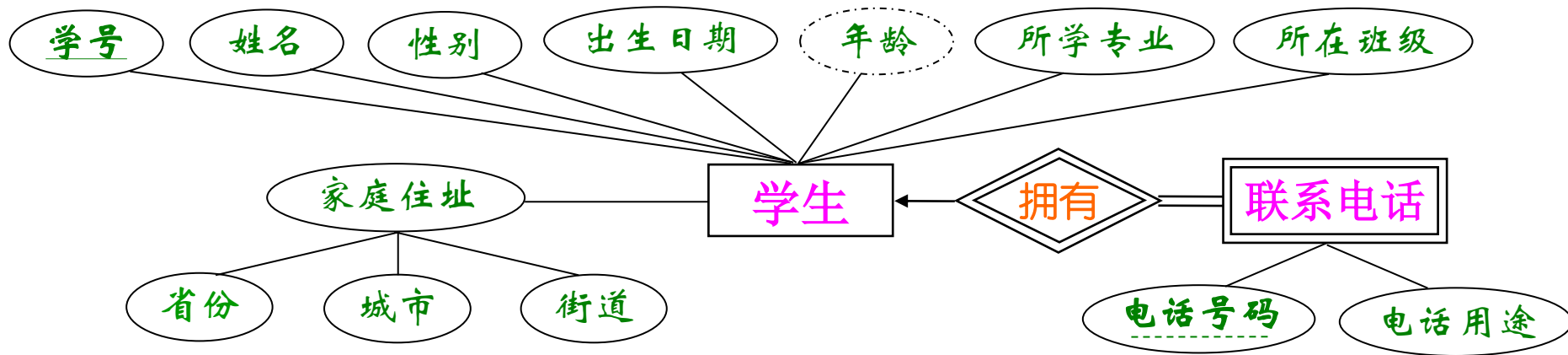


图4-18 学生实体集中的多值属性电话号码转换为弱实体集表示

- 弱实体集联系电话的属性也可以定义为：电话类别、电话号码，其中，电话类别为部分码，取值为：移动电话、宿舍电话、实验室电话、家庭电话等。

弱实体集

■ 再如，在一个大学选课管理系统中：

- 可将**学院**、**教师**、**学生**等建模为**强实体集**。
- 学院下设的**系**可考虑建模为**弱实体集**，它依赖于强实体集**学院**。
- 那么是否可以将**学院**所属的**教师(学生)**也认为是依赖于**学院**？即是否可将**教师(学生)**也建模为**弱实体集**？
 - 首先，**教师**、**学生**是客观存在的事物，不是依赖于**学院**的存在而存在的，因此**教师**、**学生**不宜作为**学院**的弱实体集进行建模。
 - 其次，一般可考虑将**没有必要**在系统的**全局范围内唯一标识**的实体集建模为**弱实体集**。

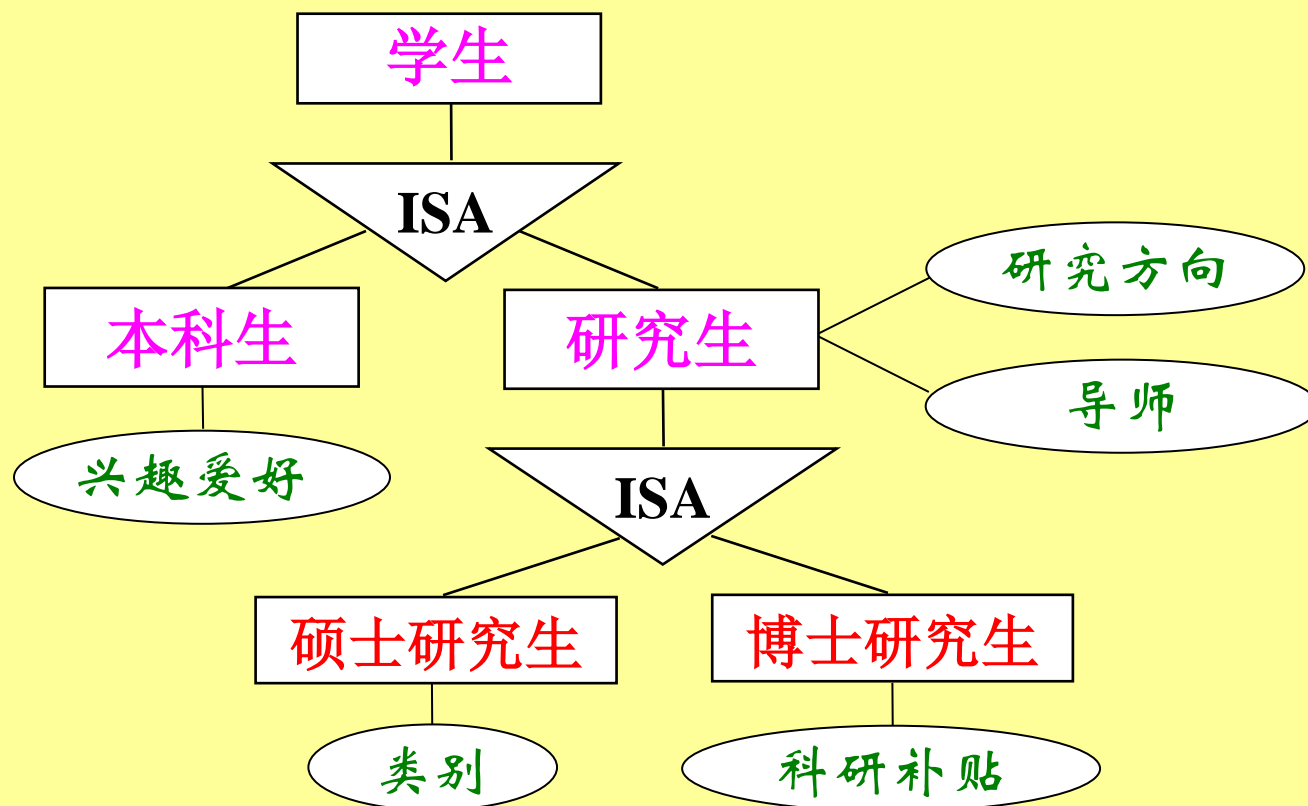
目 录

- 4.1 数据库设计过程.....
- 4.2 E-R模型基本概念及表示.....
- 4.3 约 束.....
- 4.4 弱实体集.....
- 4.5 扩展E-R特征.....
- 4.6 E-R建模问题.....
- 4.7 数据库概念设计实例——大学选课系统.....
- 4.8 逻辑设计——E-R模型转化为关系模型.....

类层次

- **实体集**中可能包含一些**子集**，该**子集**中的实体除了具有该**实体集**中所有实体所**共享的属性**外，可能还有一些自己**特有的属性**。
 - 例如，可将**学生**实体集划分为**本科生**和**研究生**两类。对于**本科生**可定义自己的特有属性**兴趣爱好**，而**研究生**可定义自己的特有属性**研究方向、导师**。
- **E-R模型**使用**实体集的继承**和**ISA联系(超类/子类联系)**来描述这种概念上的层次关系。

类层次



- **ISA**为“**is a**”的含义，表示高层实体和低层实体之间的“父类-子类”联系，称为“类层次”，也称为“特殊化”或“属性继承”(即子类会继承父类的所有属性)。

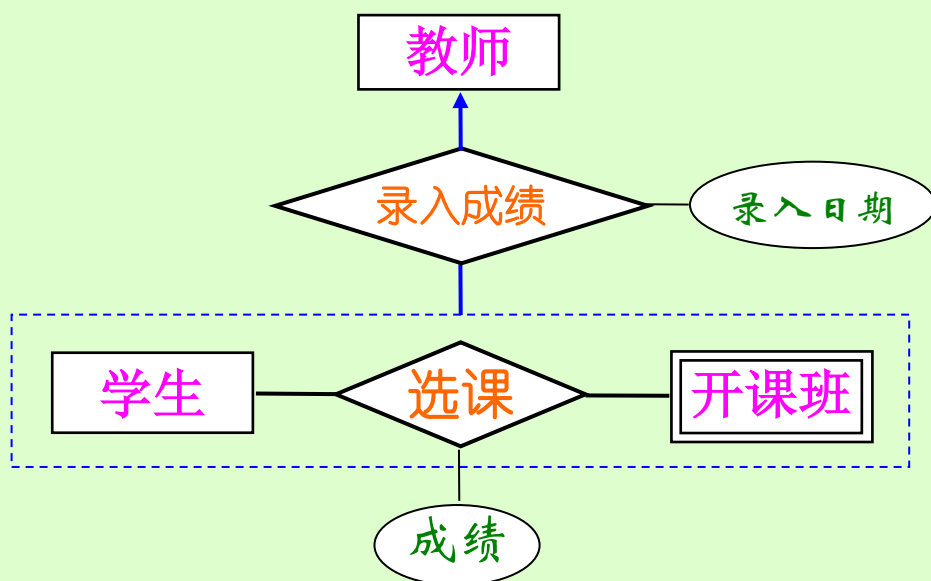
目 录

- 4.1 数据库设计过程.....
- 4.2 E-R模型基本概念及表示.....
- 4.3 约 束.....
- 4.4 弱实体集.....
- 4.5 扩展E-R特征.....
- 4.6 E-R建模问题.....
- 4.7 数据库概念设计实例——大学选课系统.....
- 4.8 逻辑设计——E-R模型转化为关系模型.....

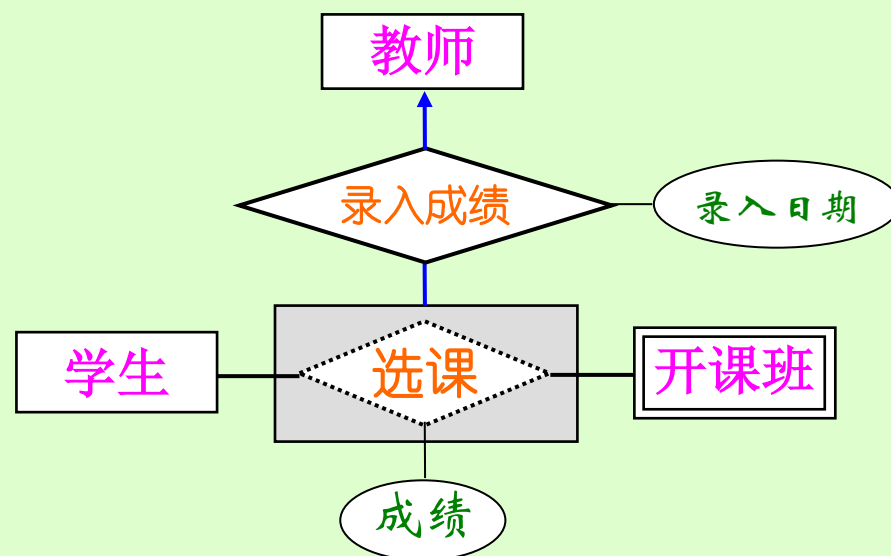
聚 合

- 问题：如何表示**联系集**之间的联系呢？
- **聚合**是一种抽象，它将一个**联系集**及其相关联的**实体集**抽象为一个**高层实体集**(称为**联系实体集**)对待，然后建立该高层实体集与其它实体集之间的联系集。
- 例如，考虑实体集**学生**和弱实体集**开课班**之间的**选课**联系集。学校教务部门需要安排教师**录入学生考试成绩**，并要求记录成绩的**录入日期**。
 - 直观上，**录入成绩**应是**选课**联系集与**教师**实体集之间的联系集，而不是**学生**或**开课班**(弱)实体集与**教师**实体集之间的联系集。

聚合



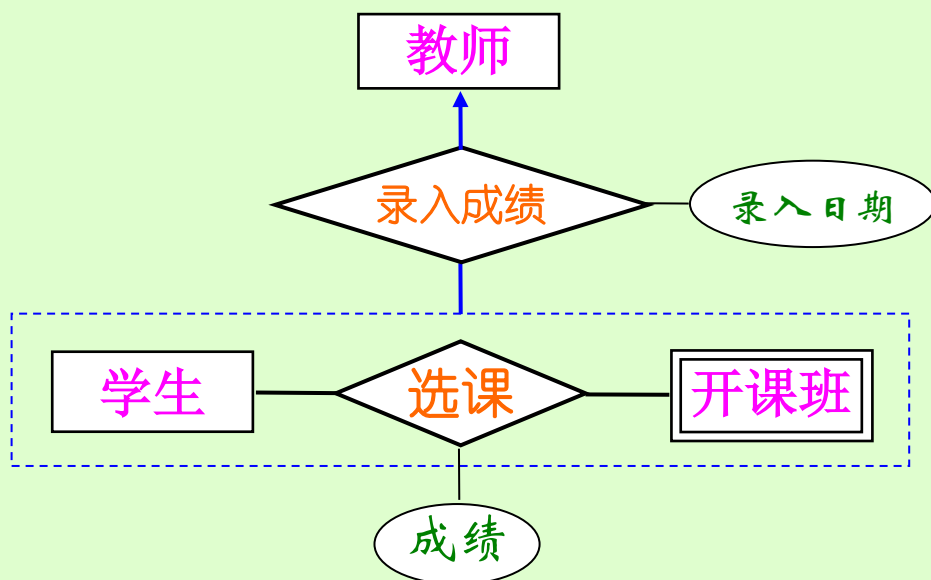
(a) 直接表示聚合的E-R建模



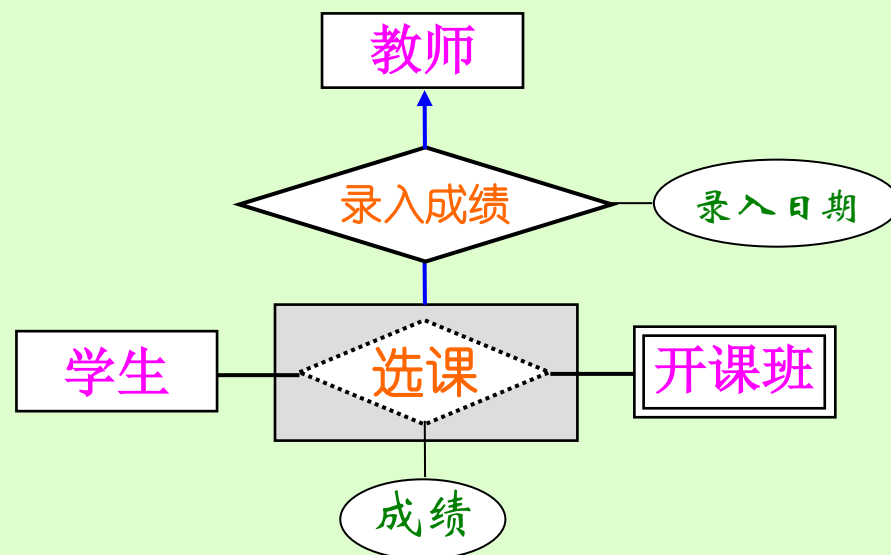
(b) 通过联系实体集的E-R建模

- 直观上，**录入成绩**应是**选课**联系集与**教师**实体集之间的联系集，而不是**学生**或**开课班**(弱)实体集与**教师**实体集之间的联系集。
- 可将包含联系集**选课**及其相关联的(弱)实体集**学生**和**开课班**的**聚合**(虚线框表示)参与到**录入成绩**联系集中(见图a)。
- 或直接将**联系实体集选课**参与到**录入成绩**联系集中(见图b)。

聚合



(a) 直接表示聚合的E-R建模



(b) 通过联系实体集的E-R建模

◆ 内部包含菱形框的带填充背景的矩形表示联系实体集，菱形框中标示的是联系集的名称，它可以同时作为联系实体集的名称。

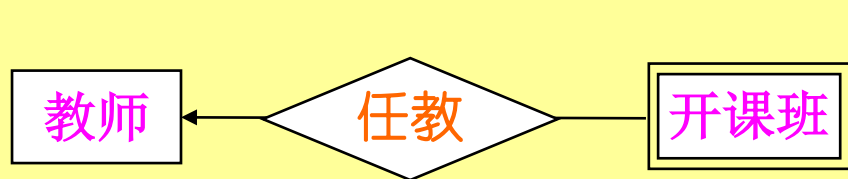
目 录

- 4.1 数据库设计过程.....
- 4.2 E-R模型基本概念及表示.....
- 4.3 约 束.....
- 4.4 弱实体集.....
- 4.5 扩展E-R特征.....
- 4.6 **E-R建模问题**.....
- 4.7 数据库概念设计实例——大学选课系统.....
- 4.8 逻辑设计——E-R模型转化为关系模型.....

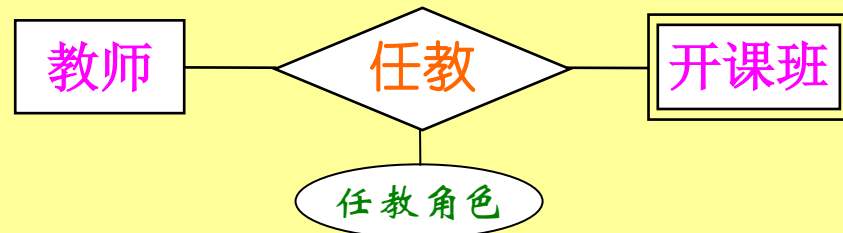
E-R建模的基本原则

■ 忠实性

- 设计应忠实于应用需求，这是首要的也是最重要的原则。即实体集、属性、联系集都应当反映现实世界及根据所了解的现实世界去建模。
- 例如，教师与开课班之间的联系集任教，是一对多还是多对多的联系集？如果规定一个开课班可能安排多名教师共同任教，则任教就是多对多联系集，联系属性为任教角色(如“主讲”、“指导实验”、“辅导”等)。



(a) 一对多的任教联系集



(b) 多对多的任教联系集

E-R建模的基本原则

■ 忠实性

- 设计应忠实于应用需求，这是首要的也是最重要的原则。即实体集、属性、联系集都应当反映现实世界及根据所了解的现实世界去建模。

■ 简单性

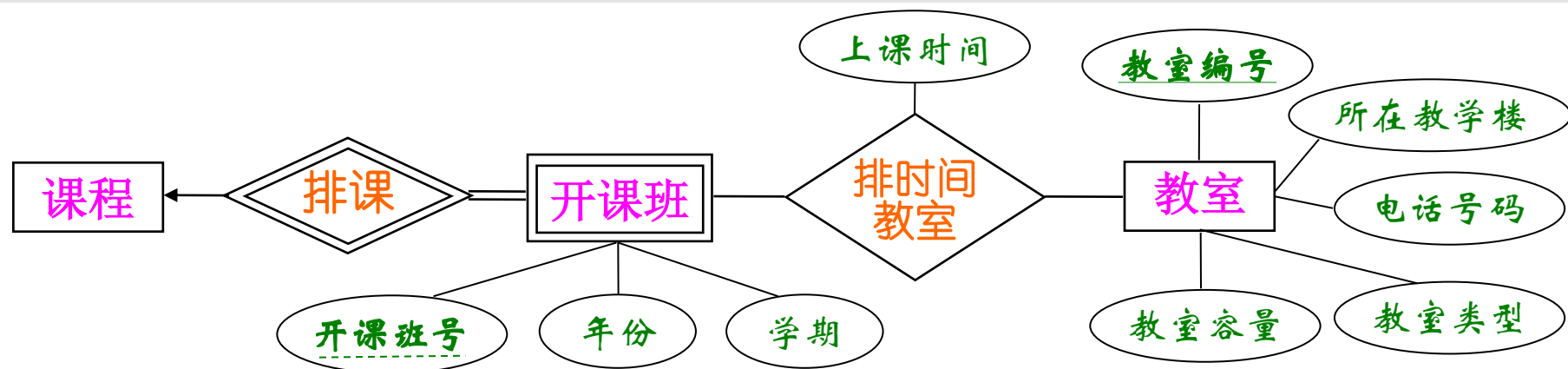
- 除非有绝对需要，否则不要在设计中增加更多成分；
- 只需要对数据库使用者所关心、感兴趣的属性建模。

■ 避免冗余

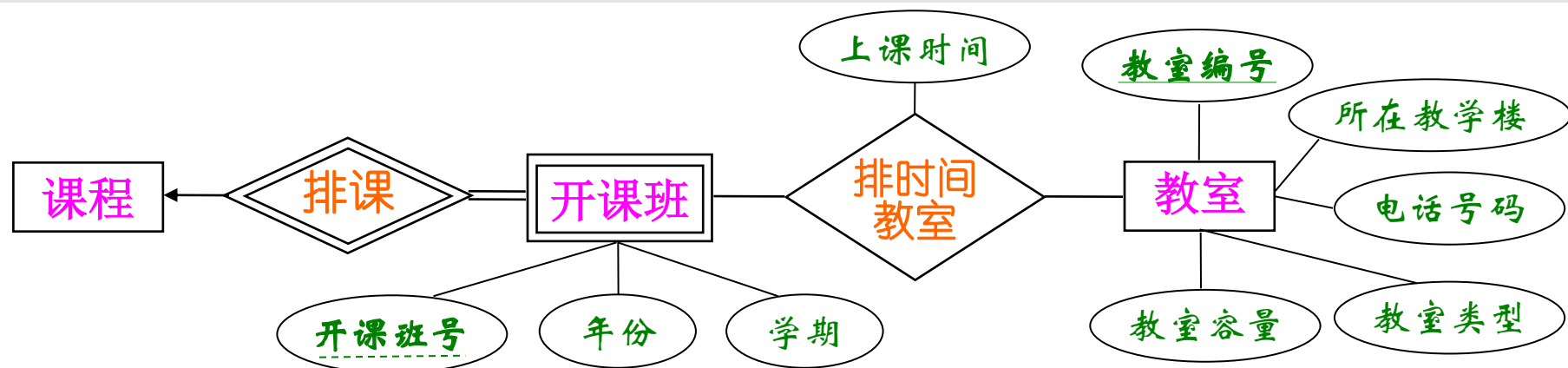
- 原则：一个对象只存放在一个地方。

选择实体集还是属性

- 通常满足下述两条规则，均可作为属性对待：
 - 作为属性，不能再具有要描述的性质；
 - 属性不能和其它实体相联系。
- 如，**开课班**弱实体集中的**上课地点**，如果除了**教室编号**之外，还需要描述更多信息，如**所在教学楼**、**电话号码**、**教室类型**、**教室容量**等，则需将属性**上课地点**转化为实体集**教室**，以实现教室管理功能。



- 如，**开课班**弱实体集中的**上课地点**，如果除了**教室编号**之外，还需要描述更多信息，如**所在教学楼**、**电话号码**、**教室类型**、**教室容量**等，则需将属性**上课地点**转化为实体集**教室**，以实现教室管理功能。
- 假设一个教室允许安排多个**开课班**上课(上课时间不能冲突)，一个**开课班**也需要安排多个时间上课，且不同时间可能安排在相同的或不同的教室上课，则**教室**实体集与**开课班**弱实体集之间存在多对多的**排时间教室**联系集，**上课时间**为联系属性。



- ◆ 说明：排时间教室不仅是多对多的联系集，而且是多值联系。因为一个开课班在不同的上课时间可能安排在同—个教室上课（即一个开课班与一个教室之间可能存在多个联系）。
- ◆ 如果一个开课班的每次上课时间都安排在不同的教室，那就不是多值联系了。
- ◆ 请参见4.6.3节的进一步讨论。

选择实体集还是属性

■ 选择实体集还是属性常犯两个错误:

- 将一实体集的主码作为另一实体集的属性，而不是使用联系；
- 将相关实体集的主码属性作为联系集的属性。因为联系集已隐含了实体集的主码属性。

选择实体集还是联系集

■ 一事物是描述为实体集还是联系集并没有一个绝对的标准。

■ 通常原则：

- 实体对应于现实世界中实际存在的事物，是名词。

- 如学生、教师和课程是名词，可作为实体集建模。

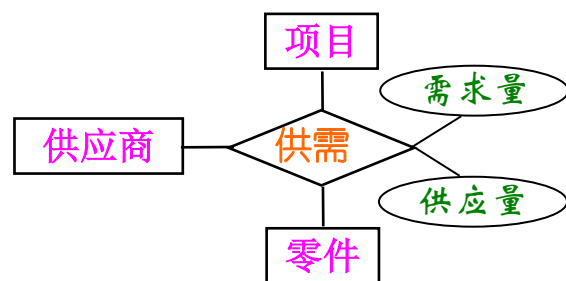
- 联系对应的概念一般为一种动作，即描述实体间的一种行为。

- 如选课、授课是动词，因此作为联系集建模。

■ 依赖约束和多值联系可能会导致将联系集建模为依赖实体集或弱实体集。

多元联系转化为二元联系

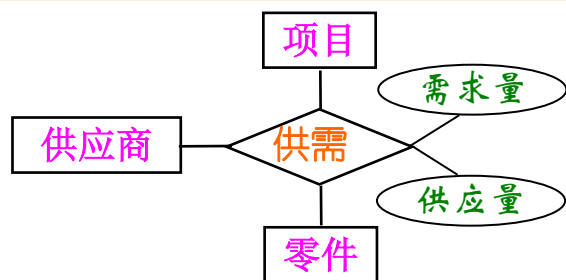
- 如图(a)所示的是**供应商**、**项目**、**零件**之间的**多对多三元联系集供需**，联系属性有**需求量**、**供应量**等。



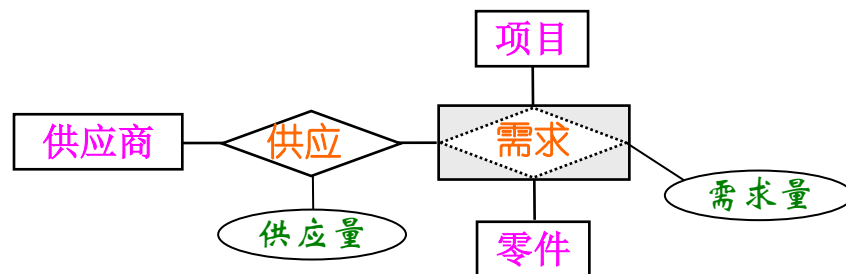
(a) 多元联系集**供需**

■ 三元联系转化为二元联系的一般方法：

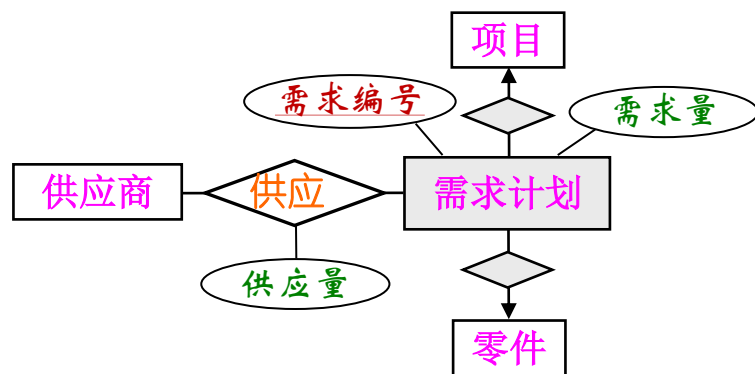
- 通过**聚合**将二元联系集建模成一个**联系实体集**，再加上它与原来联系的实体集之间的二元联系，如图(b)所示；
- 或者建立一个**依赖实体集**或**弱实体集**，再与原实体集之间建立二元联系，如图(c)、图(d)所示。



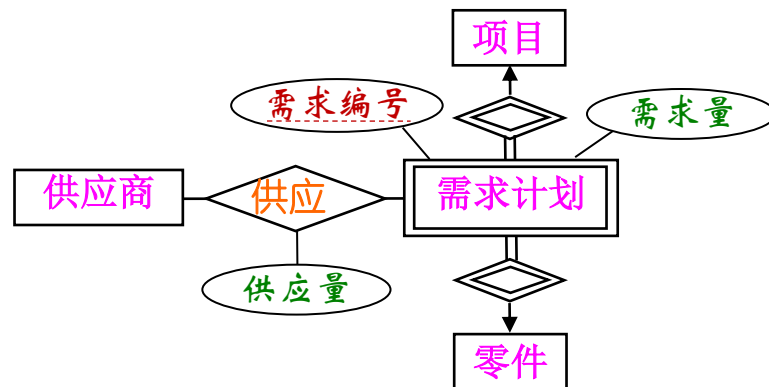
(a) 多元联系集供需



(b) 联系实体集需求



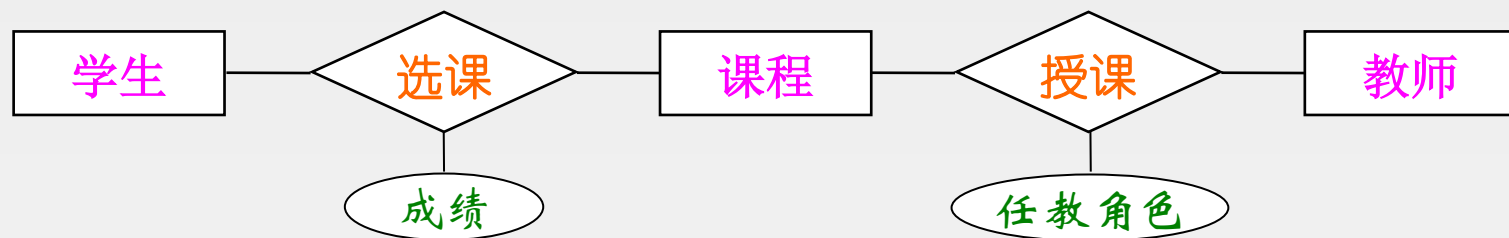
(c) 依赖实体集需求计划



(d) 弱实体集需求计划

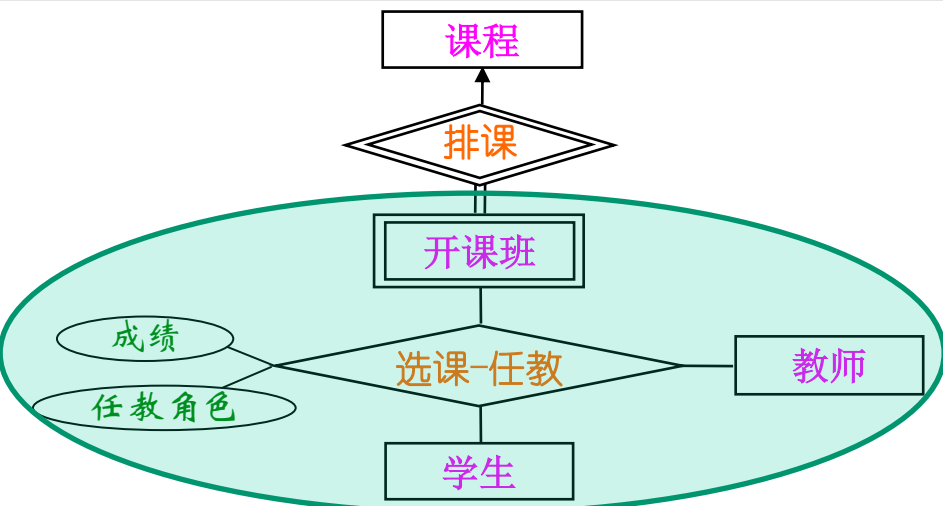
多元联系转化为二元联系

- 三元联系集选课-任教，描述了学生、课程、教师之间的多对多的联系语义。
 - 如果将其转化为学生与课程之间的选课以及教师与课程之间的授课2个二元联系，则这两个二元联系不能反映学生所选修课程是由谁授课的联系语义。
 - 问题出在一门课程可能会安排多个开课班，从而会安排多名教师授课(不同于一个开课班安排多名教师任教的语义)，而学生只是选择其中的一个开课班进行修读。

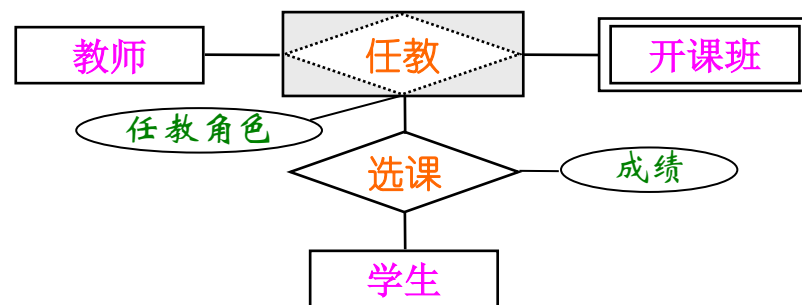


多元联系转化为二元联系

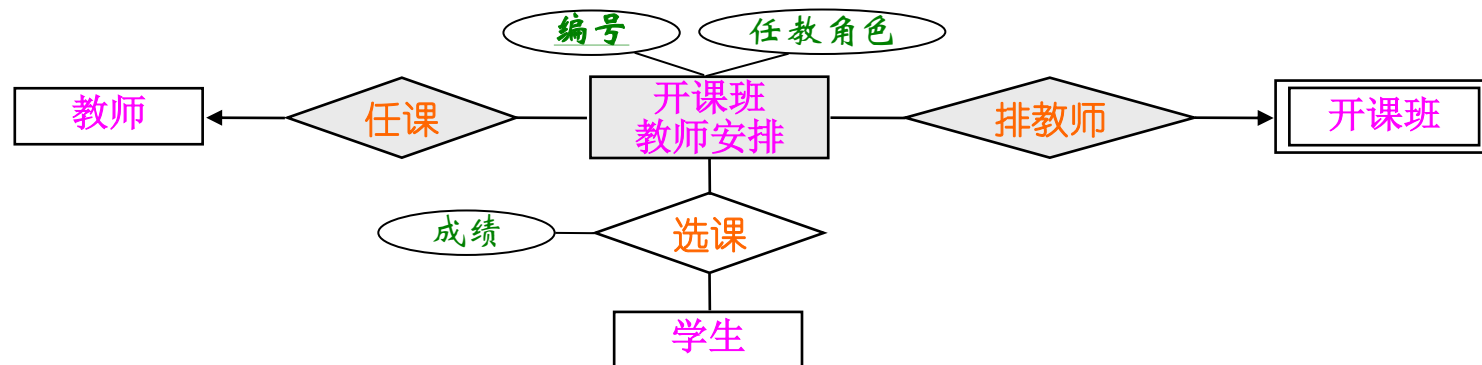
■ 考虑学生、开课班、教师之间的三元联系集选课-任教。



(a) 学生、开课班、教师之间的三元联系



(b) 通过联系实体集的转化方案



(c) 通过依赖实体集的转化方案

教师编号	开课班号	课程编号	任教角色
1	1	1	实验
2	1	1	理论

■ 考虑学生、开课班、教师之间的三

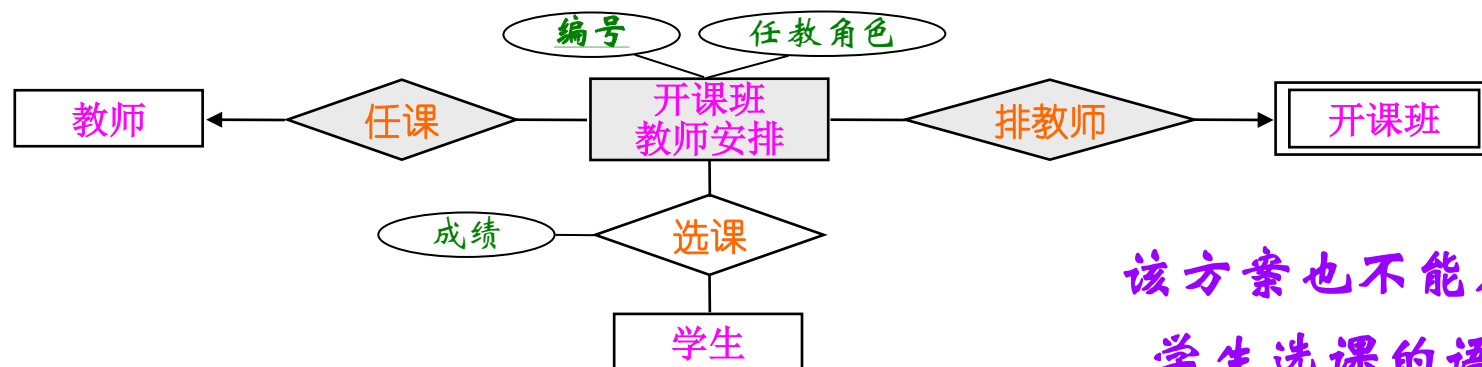


(b) 通过联系实体集的转化方案

- 先在实体集**开课班**与**教师**之间建立一个二元联系集**任教**，再在**联系实体集任教**与**学生**实体集之间建立二元联系集**选课**，如图(b)所示。
- 假设**任教**是多对多的联系语义，则**联系实体集任教**的主码是{**课程号**，**开课班号**，**教师编号**}。
- **学生**选课的语义是：**选择了某课程的某开课班，也就选择了为该开课班所安排的所有任课教师，而不能选择该开课班的某个(些)任课教师。**

多元联系转化为二元联系

- 考虑学生、开课班、教师之间的三元联系集选课-任教。
- 先在(弱)实体集开课班与教师之间引入一个依赖实体集(或弱实体集)开课班教师安排，再在依赖实体集开课班教师安排与学生实体集之间建立一个二元联系集选课，如图(c)所示。
- 该方案本质上与图(b)所示的方案相同，差别在于联系实体集与依赖实体集(或弱实体集)的主码不同。
- 而依赖实体集开课班教师安排的主码是编号，{课程号，开课班号}和教师编号分别是参照(弱)实体集开课班和教师的外码。



(c) 通过依赖实体集的转化方案

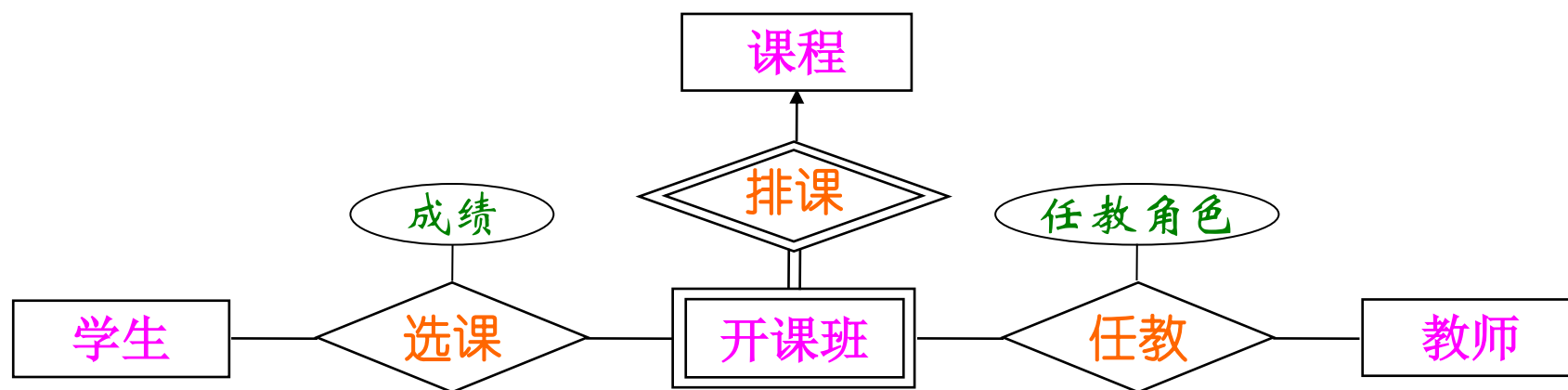
该方案也不能反映
学生选课的语义

开课编号	课程编号	教师	开课编号	课程编号	角色
1	1	1	1	1	实验
2	1	2	2	1	理论
3	2				
学生编号	开课编号	课程编号	分数		
1000	1	1	100		
1200	3	2	98		

多元联系转化为二元联系

■ 考虑学生、开课班、教师之间的三元联系集选课-任教。

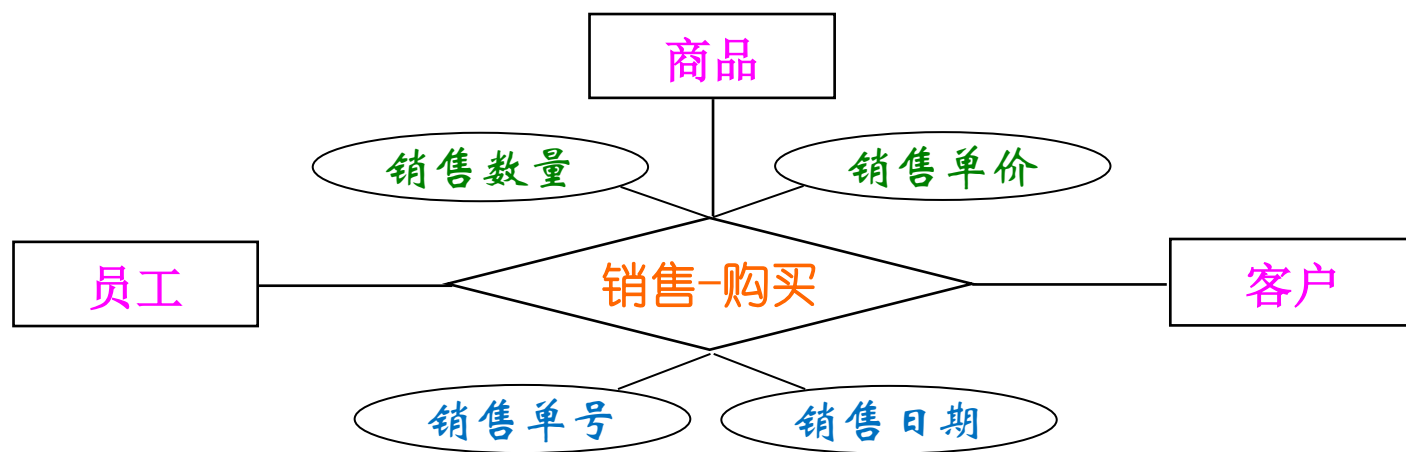
- 正确的转化方案如下图所示，它间接地表示了学生、开课班、教师之间的多对多三元联系选课-任教。
- 这是因为，若学生选修了某课程的某开课班，则可间接地通过开课班与教师之间的联系集任教来获得为该开课班所安排的所有任课教师(即为该学生授课的教师)。



依赖约束的建模

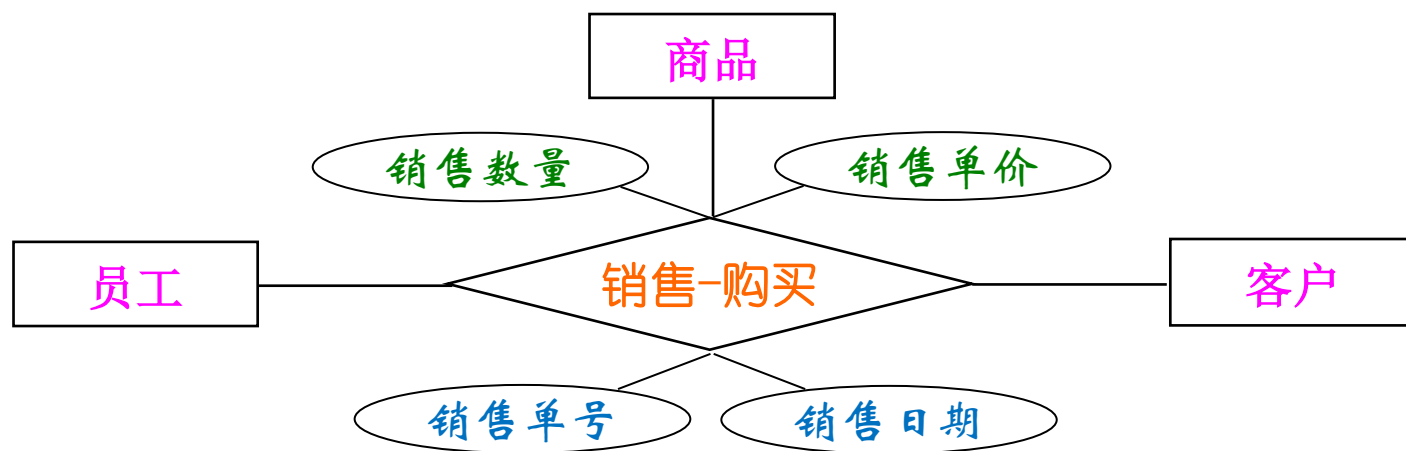
■ 对于商品销售业务，直观上的建模思路有：

- 在员工、客户与商品实体集之间建立多对多的 3 元联系集“销售-购买”。



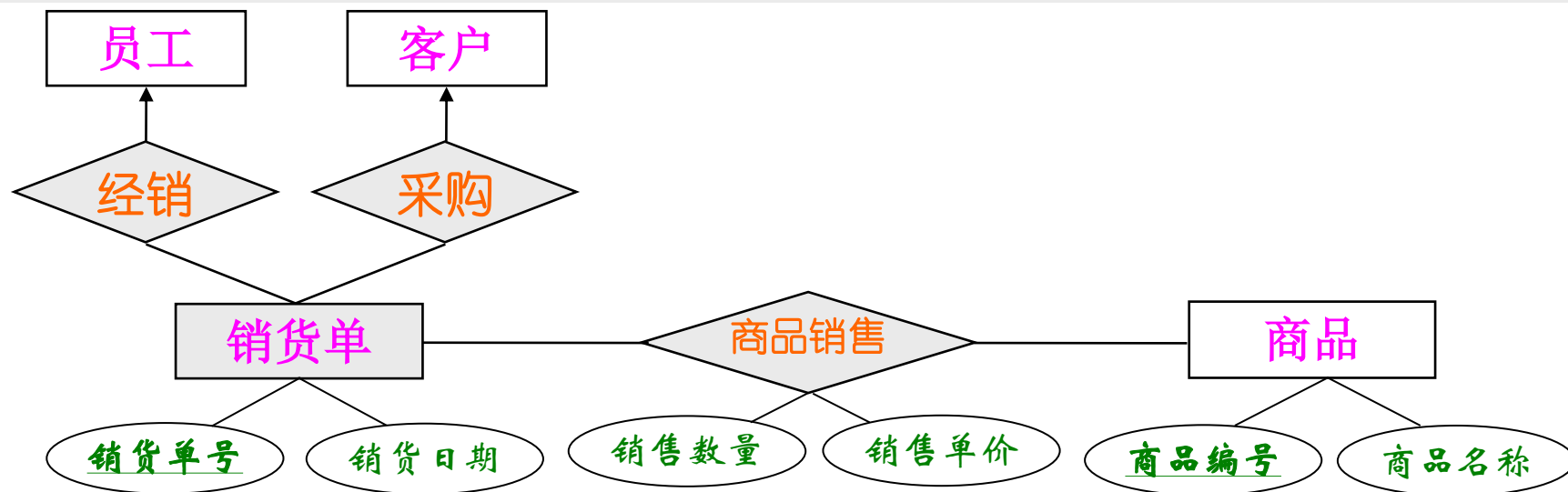
◆该建模思路存在如下2个问题:

- **数据冗余**。在**销售-购买**联系集中, 由于有的属性只依赖于一次**销售-购买**业务, 而不依赖于该次业务中的每一件**商品**, 如**销售单号**、**销售日期**等属性, 这样将造成**数据冗余**。
- **多值联系**。由于一个**员工**、一个**客户**与一件**商品**之间可能发生多次**销售-购买**, 即多对多的**3元联系集销售-购买**是**多值联系**。



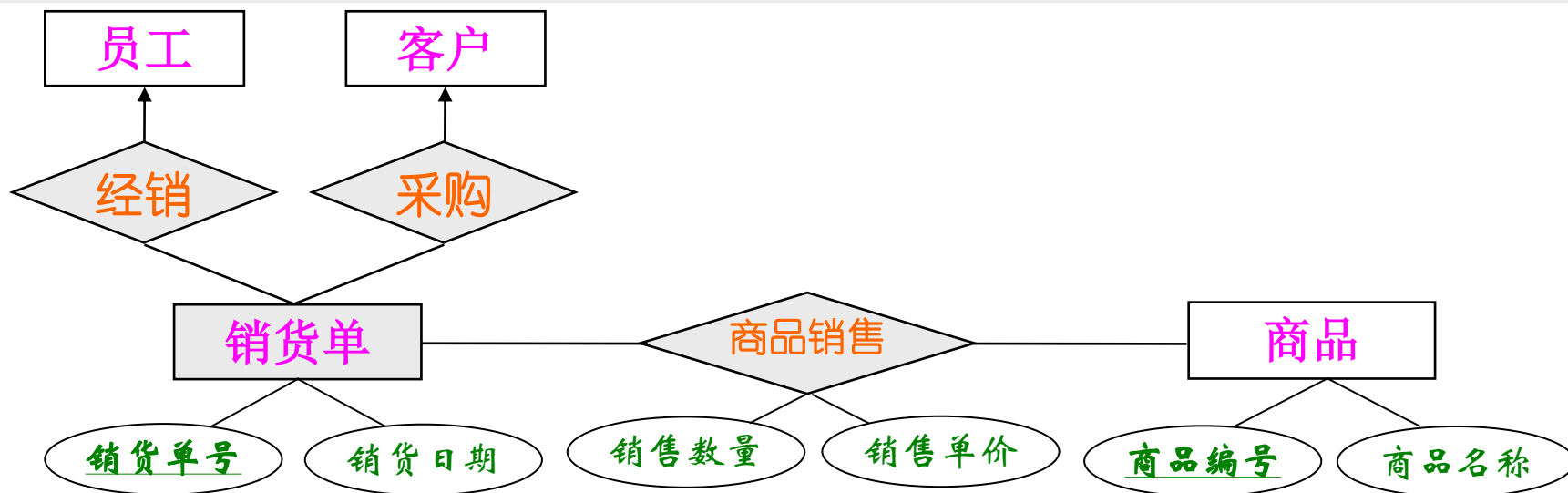
依赖约束的建模

- 伴随着商品销售业务的发生, 会产生销货单(或购货单)。
- 如果将销货单建模为实体集, 则在销货单与商品、员工、客户实体集之间分别存在着商品销售、经销、采购联系集。
- 销货单实体集与商品销售、经销、采购联系集之间存在依赖约束, 销货单是依赖实体集。



依赖约束的建模

- ◆ 依赖于联系集而存在的实体集一般是指伴随着业务发生而形成的单据。如员工、客户、商品之间发生销售/购买商品等业务时，伴随着会产生销货单/购货单。
- ◆ 在E-R建模时，一般将依赖于业务的发生而产生的销货单/购货单等直接建模为依赖实体集(而不是联系集)，并将它直接与所依赖的联系集关联起来。



依赖约束的建模

◆ 依赖于联系集而存在的实体集一般是指伴随着业务发生而形成的单据。如员工、客户、商品之间发生销售/购买商品等业务时，伴随着会产生销货单/购货单。

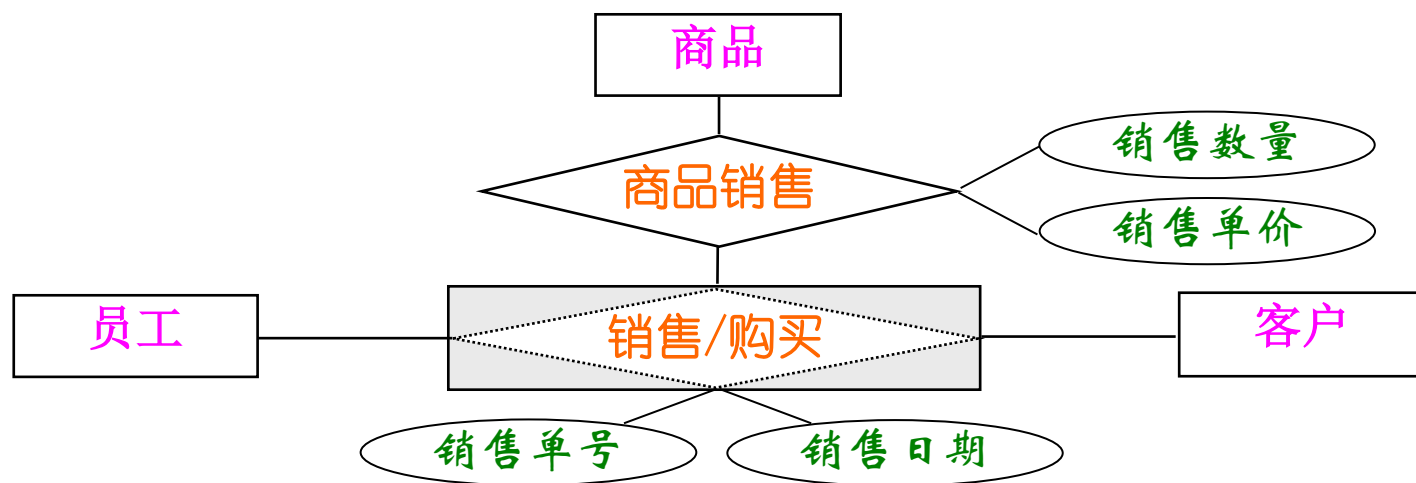
类似的业务有：

- 领料员/采购员、仓库保管员、材料之间发生的出库/入库业务会伴随着产生出库单/入库单；
- 读者、图书管理员、图书之间发生的借书业务会伴随着产生借书单；
- 客户、员工、现金之间发生的存款/取款业务会伴随着产生存款单/取款单；
- 病人、医生、药品之间发生的诊断业务会伴随着产生病历记录-处方单；
- 旅客、员工、客房之间发生的入住业务会伴随着产生入住单；
- 司机、警察、违章处罚目录之间发生的违章处罚业务会伴随着产生违章处罚单；
- 员工、游客、景点之间发生的旅游业务会伴随着产生旅游安排单；
- 公交车、车站之间发生的运行安排业务会伴随着产生公交线路。

依赖约束的建模

■ 在商品销售业务中，再对直观上的建模思路进行分析：

- 方案一：第一步先在员工与客户实体集之间建立多对多的销售/购买联系集，第二步再通过聚合在销售/购买联系集(即联系实体集)与商品实体集之间建立商品销售联系集。



- 由于一个员工与一个客户之间可能会发生多次销售/购买业务，因此，多对多的销售/购买联系集是多值联系。



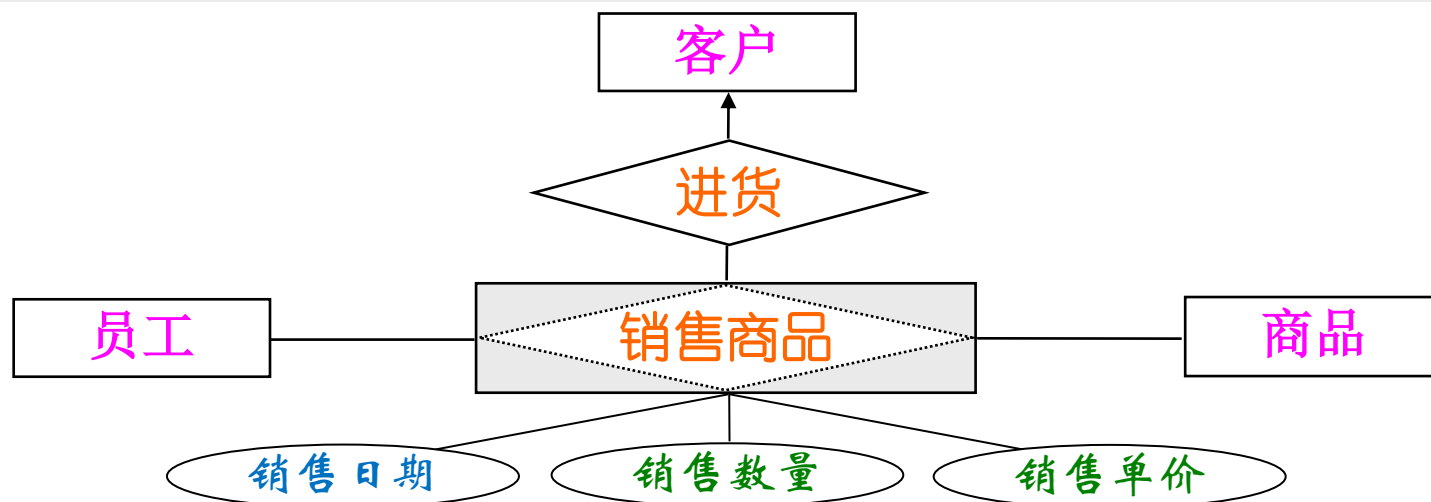
◆ 在将多值联系集销售/购买转化为销货单/购货单依赖实体集建模之后，该E-R图将转化为图4-27所示的了。



依赖约束的建模

■ 在商品销售业务中，再对直观上的建模思路进行分析：

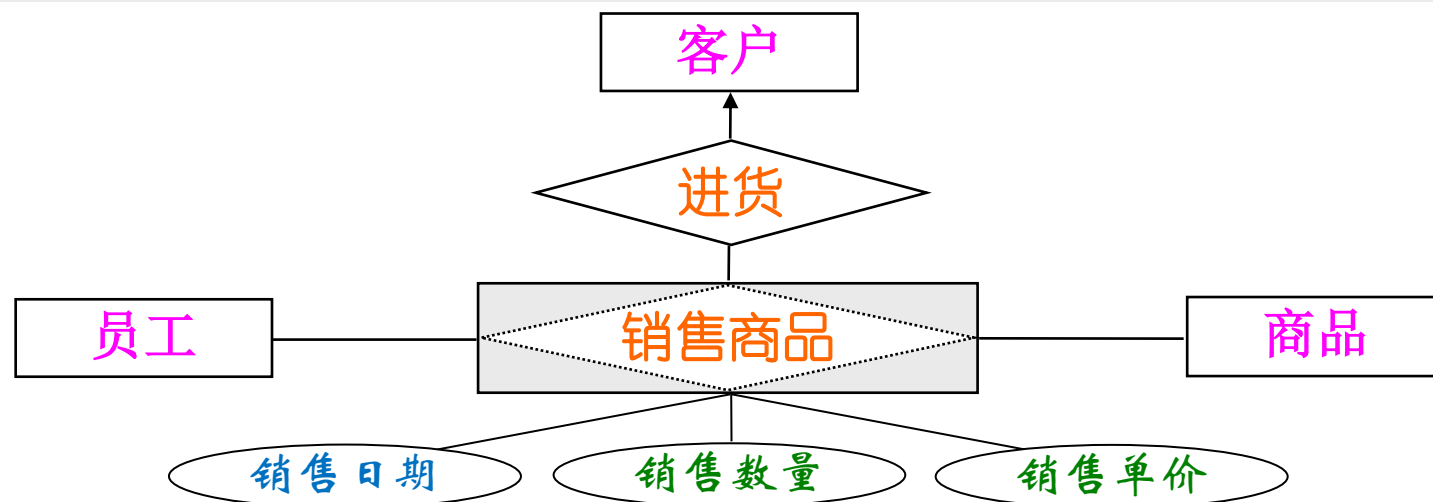
- 方案二：第一步先在员工与商品实体集之间建立多对多的销售商品联系集，联系属性有销售日期、销售数量、销售单价等；第二步再通过聚合在销售商品联系集(即联系实体集)与客户实体集之间建立进货联系集。



依赖约束的建模

◆ 该建模思路存在如下2个问题：

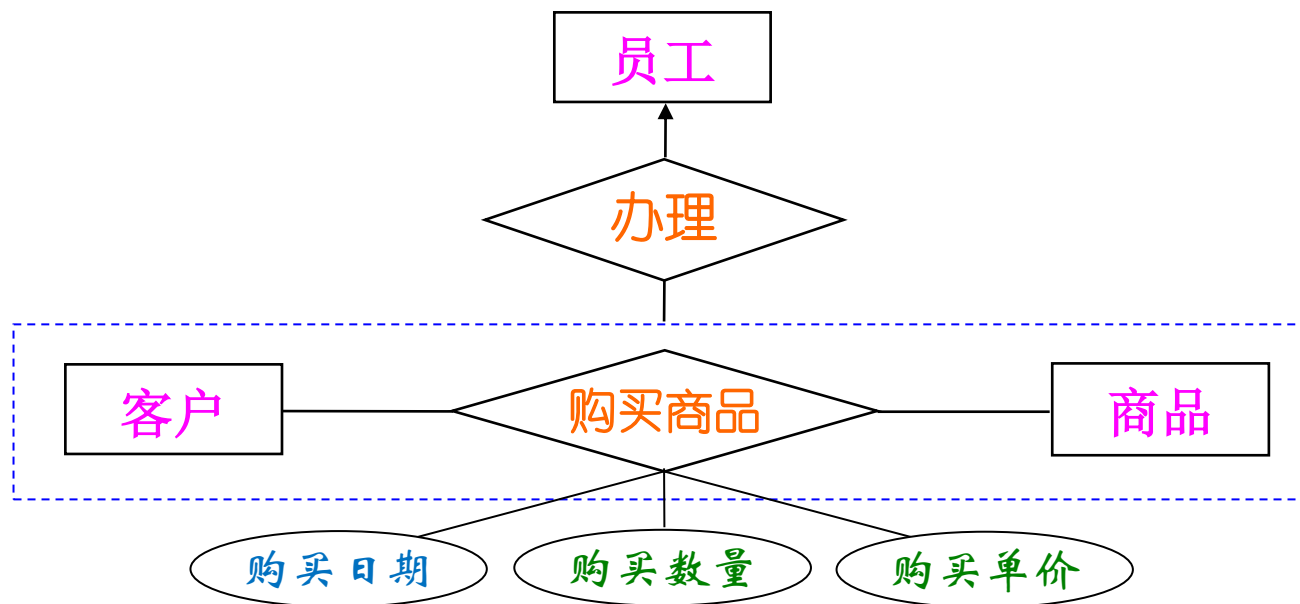
- **数据冗余**。由于**销售商品**联系集中，有的属性只依赖于一次**商品销售**业务，而不依赖于该次业务中销售的每一件**商品**，如**销售日期**等属性，这样将造成**数据冗余**。
- **多值联系**。由于一个**员工**与一件**商品**之间可发生多次销售，因此，多对多的联系集**销售商品**是**多值联系**。



依赖约束的建模

■ 在商品销售业务中，再对直观上的建模思路进行分析：

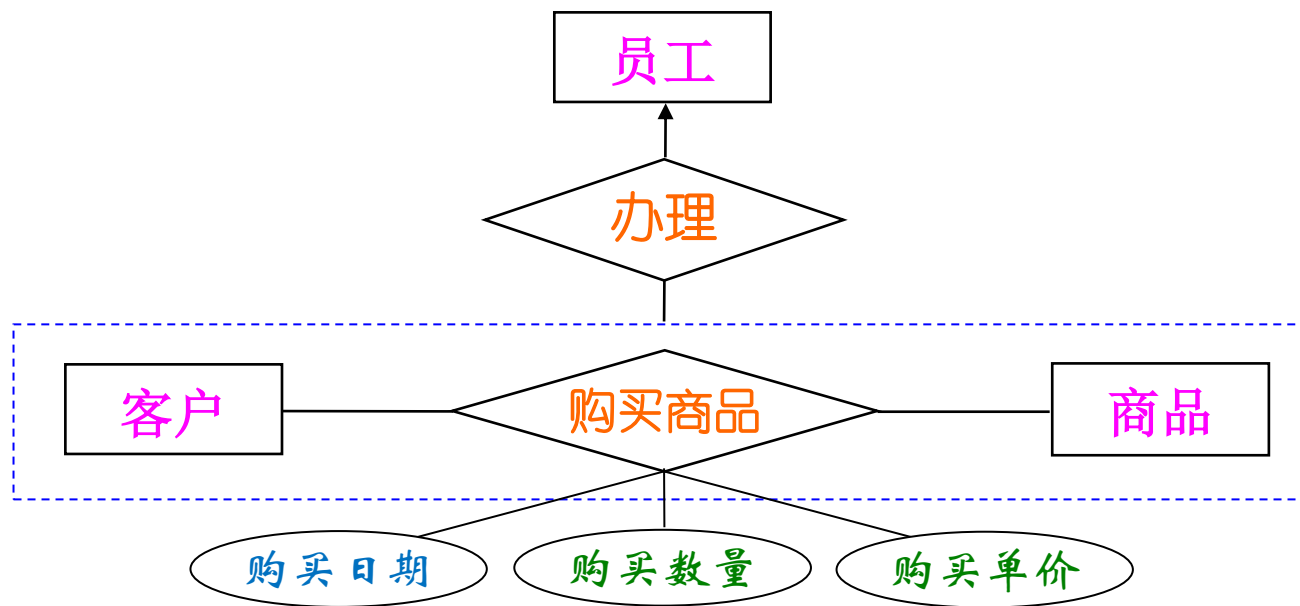
- 方案三：第一步先在客户与商品实体集之间建立多对多的购买商品联系集，联系属性有购买日期、购买数量、购买单价等；第二步再通过聚合在购买商品联系集与员工实体集之间建立办理联系集。



依赖约束的建模

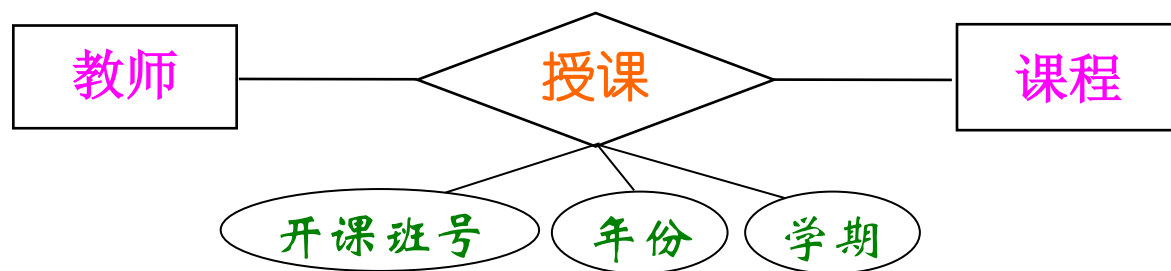
■ 在商品销售业务中，再对直观上的建模思路进行分析：

◆ 方案三的建模思路与方案二类似，存在着相同的问题。



多值联系的建模

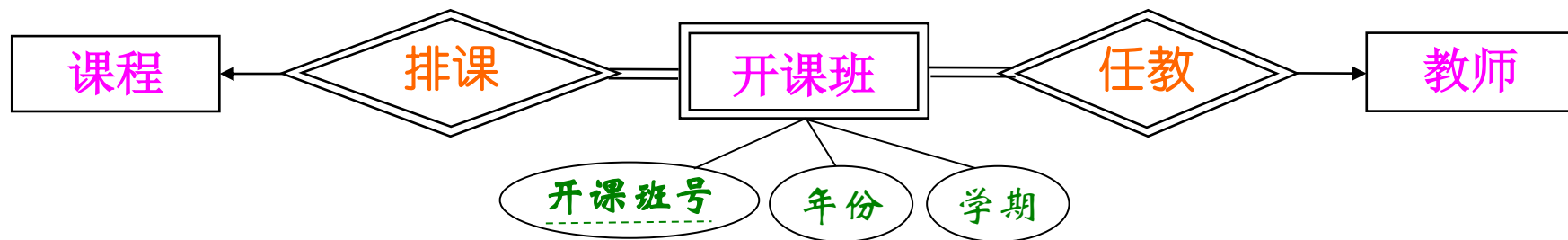
- 考虑实体集**教师**与**课程**之间的多对多**授课**联系集。由于一个教师可能会讲授同一门课程多次，即**授课**联系集是多值联系。



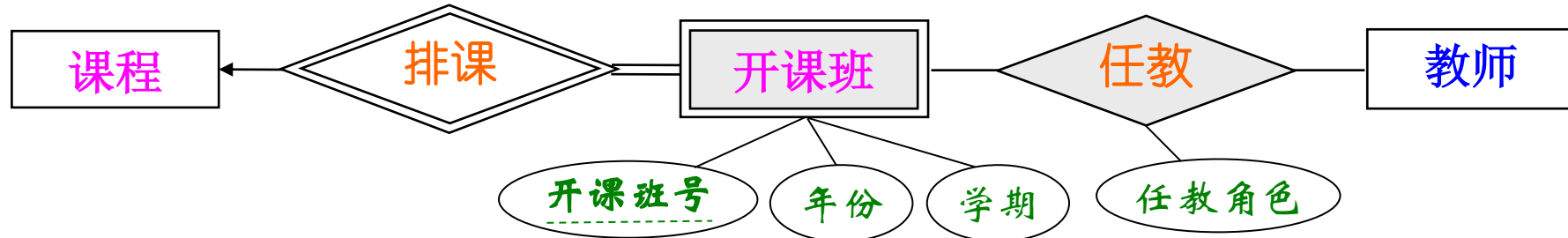
- 为了唯一标识多值联系中的多个联系，可以考虑将多值联系建模为一个**依赖实体集**或**弱实体集**，该**弱实体集**依赖于与它相关联的各个实体集，或该**依赖实体集**依赖于与它相关联的各个联系集。也就是说，多值联系的建模问题可转化为**依赖约束的建模问题**。

多值联系的建模

■ 将多值联系建模为弱实体集



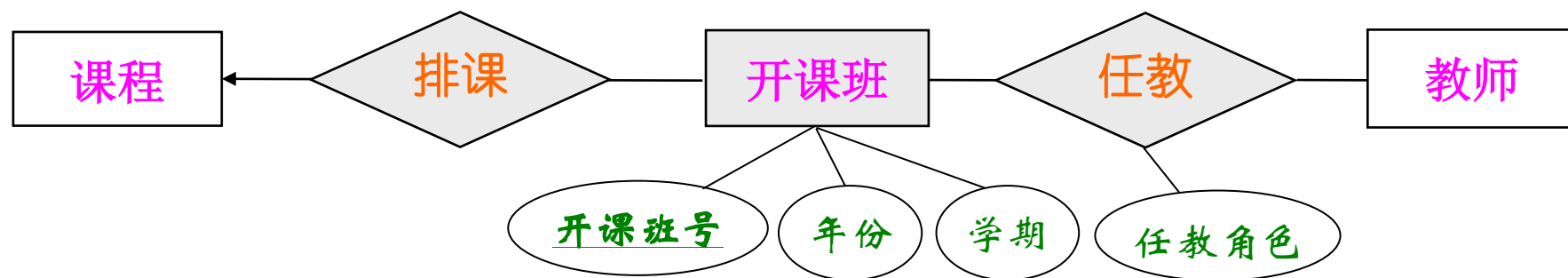
- 一方面，如果开课班还没有明确任课教师，则该开课班无法存在；
- 另一方面，如果一个开课班需要安排多名教师任教，则无法安排，因为弱实体集与其所依赖的实体集之间只能存在多对一的联系集。
- 因此，应该将开课班建模为仅依赖于课程实体集的弱实体集，同时弱实体集开课班也依赖于联系集任教。



多值联系的建模

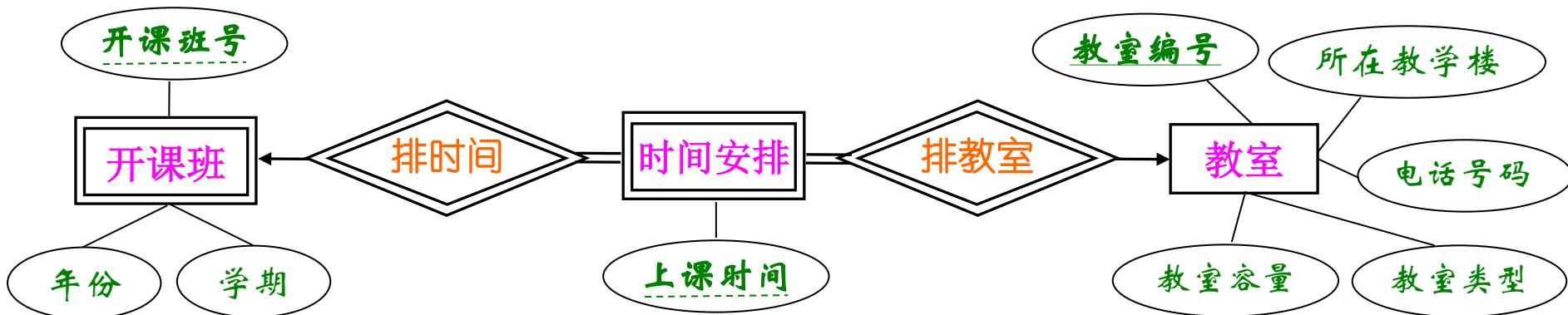
■ 将多值联系建模为依赖实体集

- 为了唯一标识多值联系中的多个联系，也可以将开课班直接建模为一个同时依赖于排课、任教联系集的依赖实体集，此时开课班号为主码，要求能够唯一标识所有课程在所有学期开设的教学班(即开课班号全局不允许出现重号)。



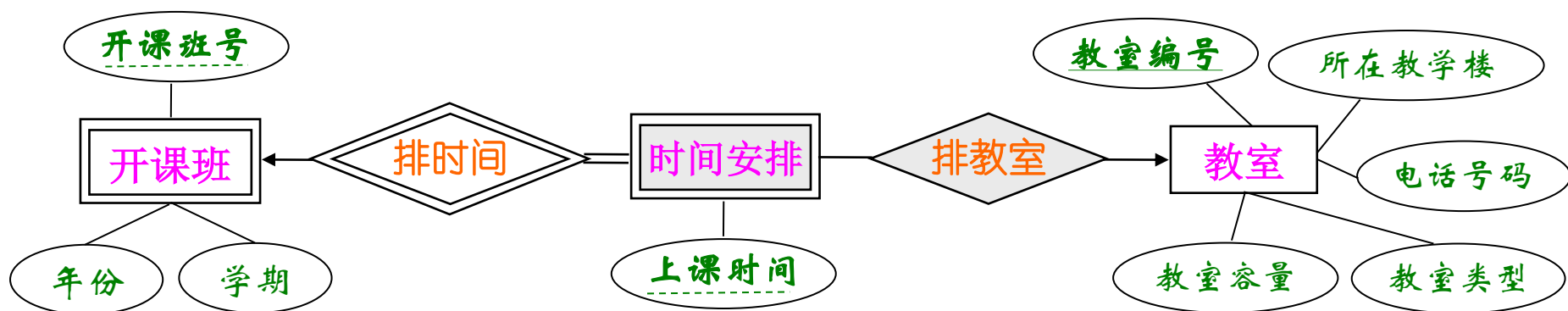
多值联系的建模

- 考虑多对多的排时间教室联系集，假设一个开课班可能安排多个时间上课，且不同时间可能安排在相同的或不同的教室上课，则排时间教室联系集可能是多值联系。
- 因此，可以考虑将排时间教室联系集建模为一个同时依赖于开课班和教室(弱)实体集的时间安排弱实体集，属性有上课时间(作为部分码)。



多值联系的建模

- 同时依赖于开课班和教室(弱)实体集的时间安排弱实体集，要求排上课时间和排上课教室必须同时完成，显然这样的依赖约束不满足教学管理的需要。
- 教学管理语义：先安排开课班的上课时间，再安排上课教室。
- 应该将时间安排建模为仅依赖于开课班的弱实体集，同时弱实体集时间安排也依赖于联系集排教室。



总结

■ E-R模型

- 实体、属性与实体集（复合、多值属性）
- 联系、联系属性与联系集，二元联系的主码与联系属性的安排
- 映射基数(1:1、1:n、m:1、m:n联系)、依赖约束、多值联系
- 弱实体集、部分码
- 扩展特征：类层次与聚合(建模为联系实体集)
- 依赖约束的建模：建模为依赖实体集或弱实体集
- 多值联系的建模：转化为依赖约束的建模

■ E-R模型设计原则

- 忠实性、简单性、避免冗余
- 选择实体集还是属性？
- 选择实体集还是联系集？（依赖约束、多值联系的建模）
- 多元联系转化为二元联系：联系实体集、依赖实体集或弱实体集