



海量数据计算研究中心

# 设计篇

## 第四章 概念数据库设计

主讲：程思瑶

海量数据计算研究中心





# 数据库设计步骤

- 数据库设计概述与需求分析
- 概念数据库设计
- 逻辑数据库设计
- 物理数据库设计





## 4.1 数据库设计概述与需求分析





## 4.1 数据库设计

- 数据库设计

- 对于一个给定的应用领域，设计优化的数据库逻辑和物理结构，使之满足用户的信息管理要求和数据操作要求，有效地支持各种应用系统的开发和运行

- 信息管理要求

- 在数据库中应该存储和管理哪些数据对象。

- 数据操作要求

- 对数据对象需要进行哪些操作，如查询、增、删、改、统计等操作。





## 4.1 数据库设计

- 数据库设计目标

- 为用户和各种应用系统提供一个信息基础设施和效率高的运行环境

- 数据信息完整、冗余度低
    - 数据库的存取效率高
    - 数据库系统运行管理的效率高





## 4.1 数据库设计

- 数据库设计包括
  - 设计数据库模式
  - 设计访问和更新数据的程序
  - 设计数据访问的安全模式

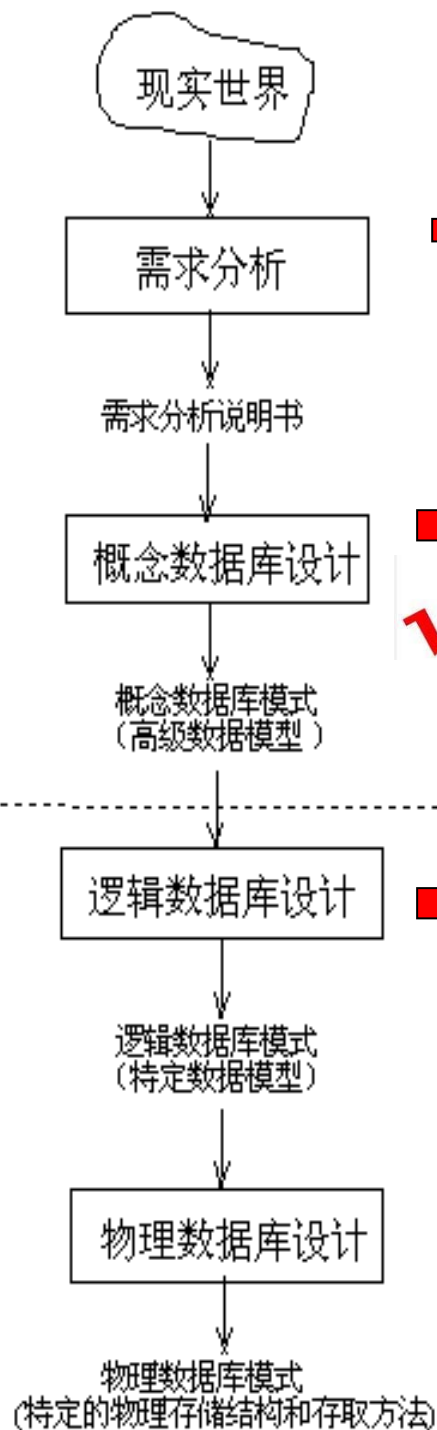
用户的需求在设计过程中扮演一个中心角色





# 数据库设计步骤

## 数据库各级模式的形成过程



综合各个用户  
的应用需求

形成独立于机器特  
点，独立于各个  
DBMS产品的概念  
模式(E-R图)

首先将E-R图转换成具体的数据  
库产品支持的数据模型，如关  
系模型，形成数据库逻辑模式  
然后根据用户处理的要求、安  
全性的考虑，在基本表的基础  
上再建立必要的视图(View)，  
形成数据的外模式

根据DBMS特点和处理的  
需要，进行物理存储  
安排，建立索引，形成  
数据库内模式

独立于数据  
库管理系统

与数据库管  
理系统相关



# 需求分析

- 需求分析就是分析用户的要求
  - 是设计数据库的起点
  - 结果是否准确地反映了用户的实际要求，将直接影响到后面各个阶段的设计，并影响到设计结果是否合理和实用







# 需求分析

- 需求分析的任务

- 详细调查现实世界要处理的对象（组织、部门、企业等）
- 充分了解原系统（手工系统或计算机系统）工作概况
- 明确用户的各种需求
- 在此基础上确定新系统的功能
- 新系统必须充分考虑今后可能的扩充和改变





# 需求分析

- 需求分析的重点是

- “数据”和“处理”，获得用户对数据库要求

- 信息要求

- 用户需要从数据库中获得信息的内容与性质

- 由用户的信息要求可以导出数据要求，即在数据库中需要存储哪些数据

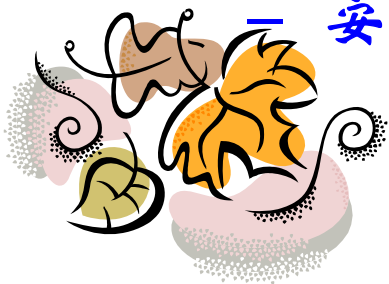
- 处理要求

- 用户要完成什么处理功能

- 对处理的响应时间的要求

- 对处理方式的要求(批处理 / 联机处理)

- 安全性与完整性要求





# 需求分析

- 需求分析的步骤
  - 应用领域的调查分析
  - 定义数据库支持的信息与应用
  - 定义数据库操作任务
  - 预测应用领域的未来改变





## 4.2 概念数据库设计





## 4.2 概念数据库设计

- 概述
- 实体联系模型
- 概念数据库设计的方法与策略
- 视图综合设计方法
- 事务的设计





## 4.2 概念数据库设计

- 概述
- 实体联系模型
- 概念数据库设计的方法与策略
- 视图综合设计方法
- 事务的设计





## 4.2.1 概述

- 概念数据库设计的任务包括两方面

- 概念数据库模式设计

- 以需求分析阶段所识别的数据项和应用领域的未来改变信息为基础，使用高级数据模型建立概念数据库模式

- 事务设计

- 考察需求分析阶段提出的数据库操作任务，形成数据库事务的高级说明





## 4.2.1 概述

- 概念数据库模式设计的目标
  - 准确描述应用领域的信息模式，支持用户的各种应用
  - 既易于转换为逻辑数据库模式，又容易为用户理解

不失真

易理解







## 4.2.1 概述

- 概念数据库模式**独立于**任何数据库管理系统，不能直接用于数据库的实现。
- 用于概念数据库设计的高级数据模型：

**实体联系模型**





## 4.2 概念数据库设计

- 概述
- 实体联系模型
- 概念数据库设计的方法与策略
- 视图综合设计方法
- 事务的设计





## 4.2.2 实体联系模型

- 实体联系模型

- Entity-Relationship model, 简称ER模型

- 被表示成“实体-联系”图，图中有三个主要元素

- 实体
    - 属性
    - 联系





# 基本概念

- 实体
- 属性
- 联系





# 基本概念(实体)

- 实体(Entity)

- **实体**是ER模型的基本对象，是现实世界中可区别所有其他对象的一个“事务”或“对象”。

- 实体可以是物理存在的事物，如人、汽车；
    - 也可以是抽象的概念，如学校、课程。

学生





# 基本概念(实体)

- 实体(Entity)

- 每个实体都有一组特征或性质，称为实体的属性。  
实体属性的一组特定值确定了一个特定的实体。实体的属性值是数据库中存储的主要数据。

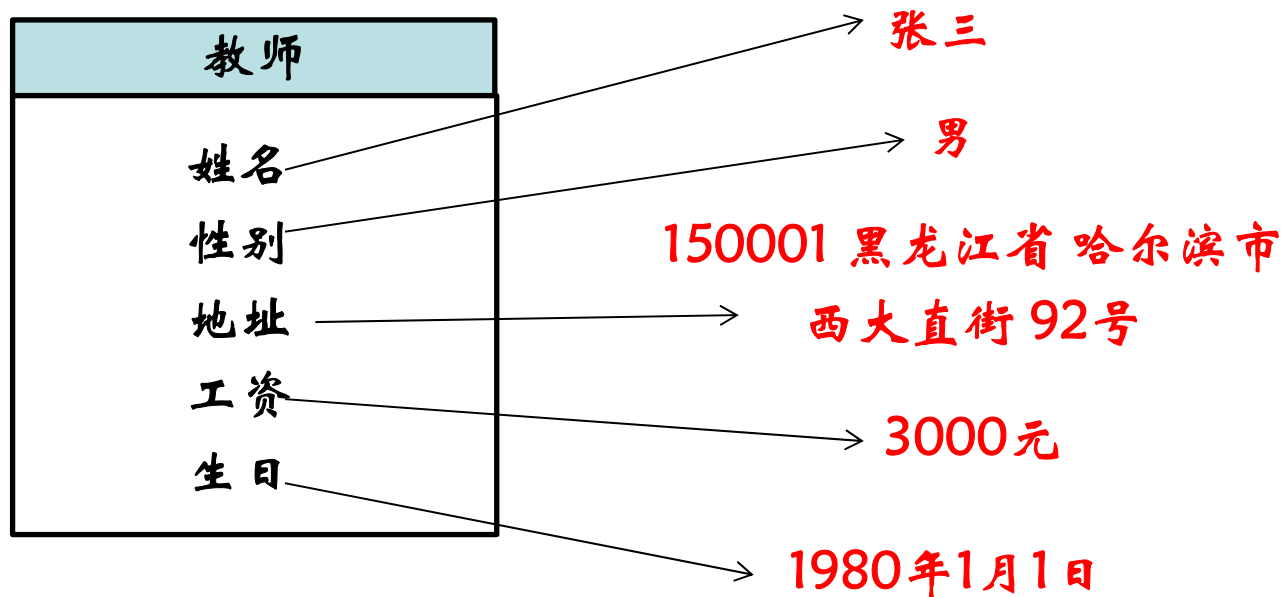
学生
学号
姓名
生日
性别
年龄
所属系





# 基本概念(实体)

- 实体例子





# 基本概念(实体)

- 实体集

- 是相同类型(即具有相同性质或属性)的实体集合

- 例如, 某个大学所有学生的集合可被定义为实体集 *student*.

- 实体集不必互不相交







# 基本概念(实体)

## 实体集例子: *instructor*, *student*

—instructor\_ID instructor\_name

76766	Crick
45565	Katz
10101	Srinivasan
98345	Kim
76543	Singh
22222	Einstein

*instructor*

student-ID student\_name

98988	Tanaka
12345	Shankar
00128	Zhang
76543	Brown
76653	Aoi
23121	Chavez
44553	Peltier

*student*





# 基本概念

- 实体
- 属性
- 联系





# 基本概念(属性)

- 属性(Attribute)
  - 实体的特征或性质
  - 是实体集映射到域的函数
- 属性可包括
  - 单值(single-valued)、多值(multivalued)属性
  - 简单(simple)、复合(composite)属性
  - 派生(derived)属性





# 基本概念(属性)

- 实体属性

- ✓ 多数实体属性都是单值属性，即对于同一个实体只能取一个值。

- 例如，同一个人只能具有一个年龄，所以人的年龄属性是一个单值属性。

- ✓ 但是，在某些情况下，实体的一些属性可能取多个值。这样的属性称为多值属性。

- 例如：职务、联系方式等

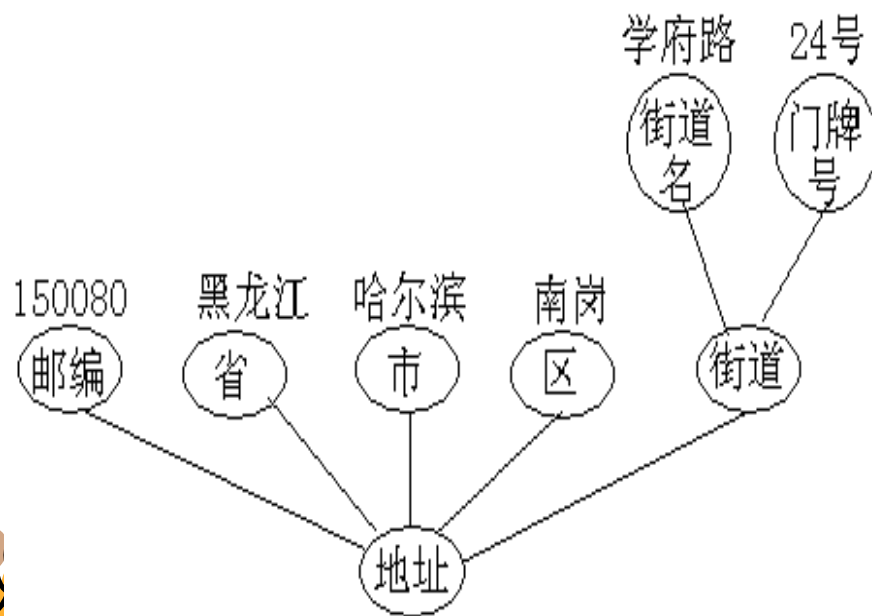




# 基本概念(属性)

## • 实体属性(续)

✓ 可以划分为多个具有独立意义的子属性。我们称这类属性为**复合属性**。复合属性具有**层次结构**



学生	
学号	
地址	
	邮编
	省市区街道
	街道名
	门牌号





# 基本概念(属性)

- 实体属性(续)

- ✓ 派生属性

- 可以由其他属性导出称为派生属性。

学生
学号
姓名
生日
性别
年龄
所属系





# 基本概念(属性)

- 实体属性(续)

- ✓ 空值

- 在某些情况下, 属性值为空值。

- ✓ 例如, 一个未获得任何学位的人的学位属性只能被设置为空值





# 基本概念(属性)

- 实体属性(码)

- ✓ 在ER模型中每个实体集具有一个由一个或多个属性组成的码，是足以区分每个实体的属性集

- 例如，由于不同学生不能具有相同的学号，学生实体集的学号属性是码。

- ✓ 由一个属性构成的键称为简单码。由多个属性构成的键称为复合码，统称为码。

学生
<u>学号</u>
<u>姓名</u>
<u>生日</u>
性别
年龄
所属系







# 基本概念

- 实体
- 属性
- 联系



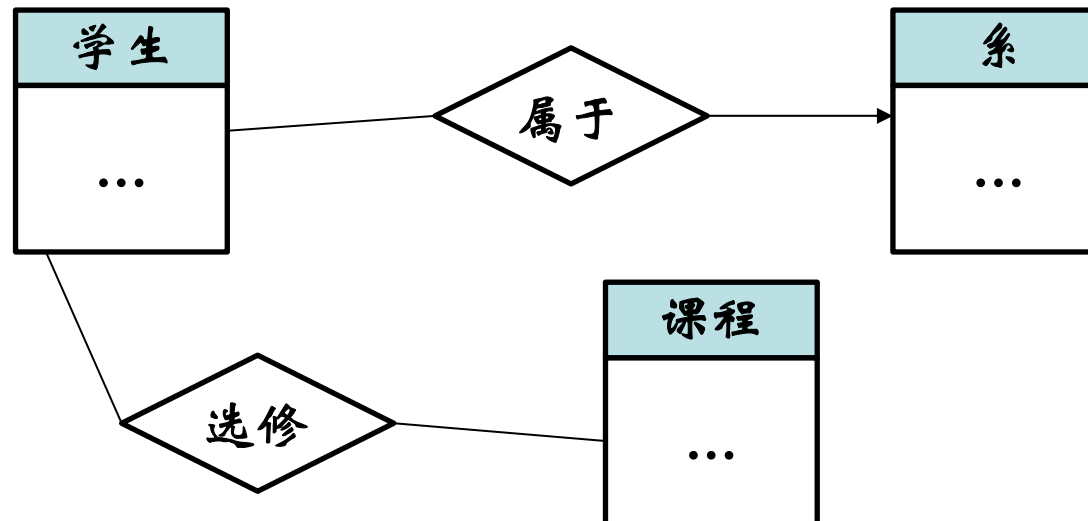


# 基本概念(联系)

- 联系 (Relation)

— 不同实体集的实体之间可能具有某种**关联**，我们称这种关联为**实体间的联系**

- 例如，一个学生必属于一个系；一个学生需选修多门课程

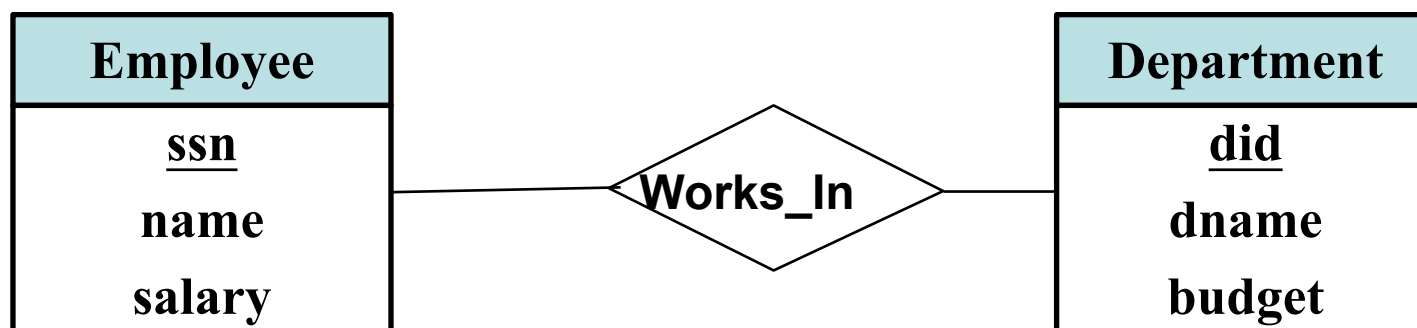




# 基本概念(联系)

- 联系集

— 同类联系的集合



称一个联系集所关联的实体集的数量为这个联系集的阶。阶为n的联系集称为n元联系集



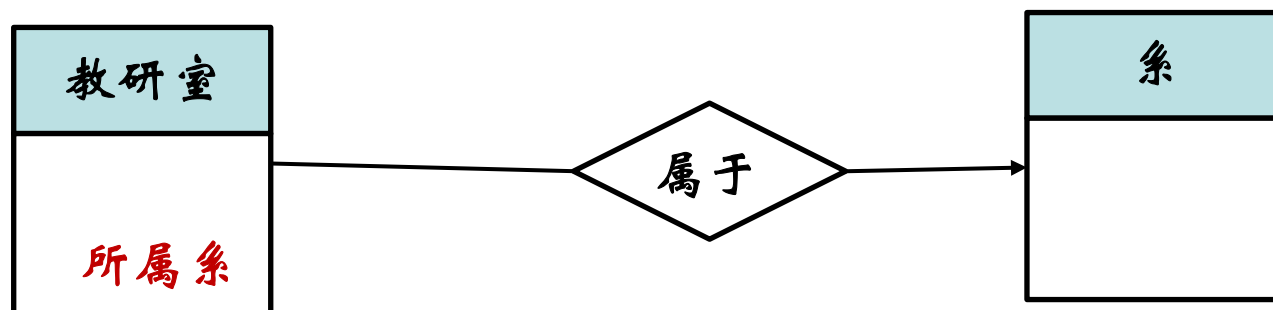
# 基本概念(联系)

## • 注意

— 实体之间的联系既可以使用联系集定义，也可以通过**实体属性**来表示。

• 例如，可以用属性表示实体教研室和系之间的**所属联系**。

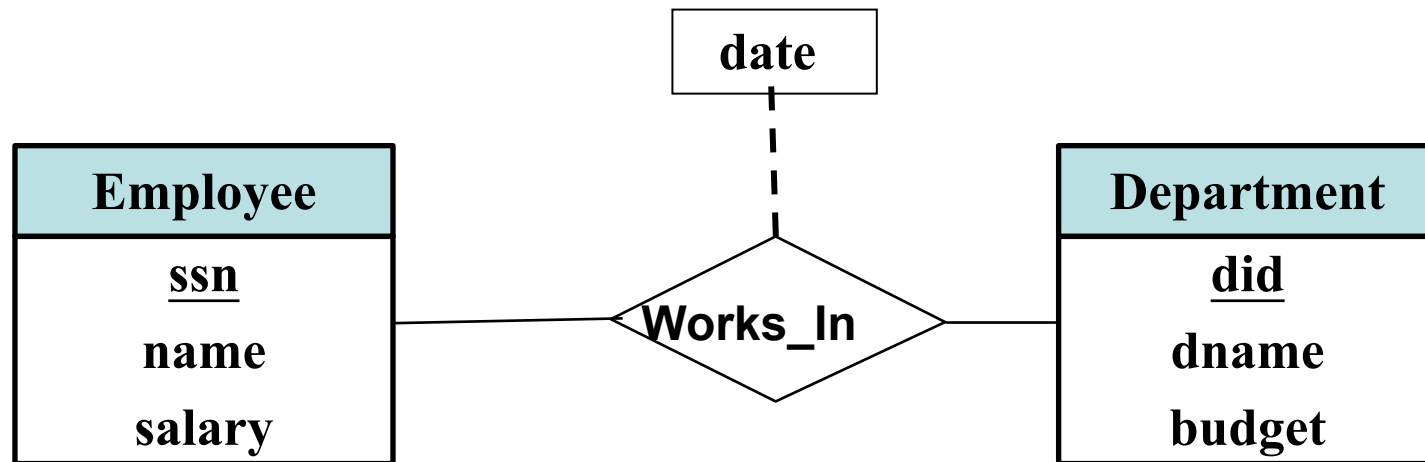
— 用教研室实体的属性“所属系”的值来表示这个实体所属的系实体。





# 基本概念(联系)

- 联系集的属性：描述性属性





# 基本概念(约束)

- 约束

- ER企业模式可以定义一些数据库中的数据必须满足的约束
- 映射基数
- 参与约束





# 基本概念(约束)

- 映射基数(基数比率)

- 一个实体通过一个联系集能关联的实体的个数

- 一对一

- 一个系只有一个系主任，一个大学只有一个校长

- 一对多 (或多对一)

- 一个系有很多学生

- 多对多

- 一个学生可选修多门课程，每门课程可被多个学生选修





# 其他概念(约束)

- 参与约束

- 全域关联约束

- 如果在大学数据库中规定每个教研室必须属于一个系，则教研室实体集与联系Belong\_To联系集有个全域关联约束。

- 部分关联约束

- 由于并非每个教师都是系主任，所以教师实体型上有一个与联系T\_D\_2有部分关联约束。







# 其他概念

- 弱实体集

- 弱实体集必须具有一个或多个属性，使得这些属性可以与主实体集的码相结合，形成相应弱实体的码
- 上述属性称为弱实体集的部分码



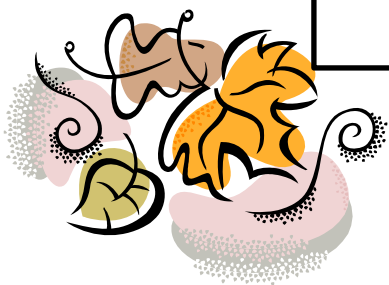
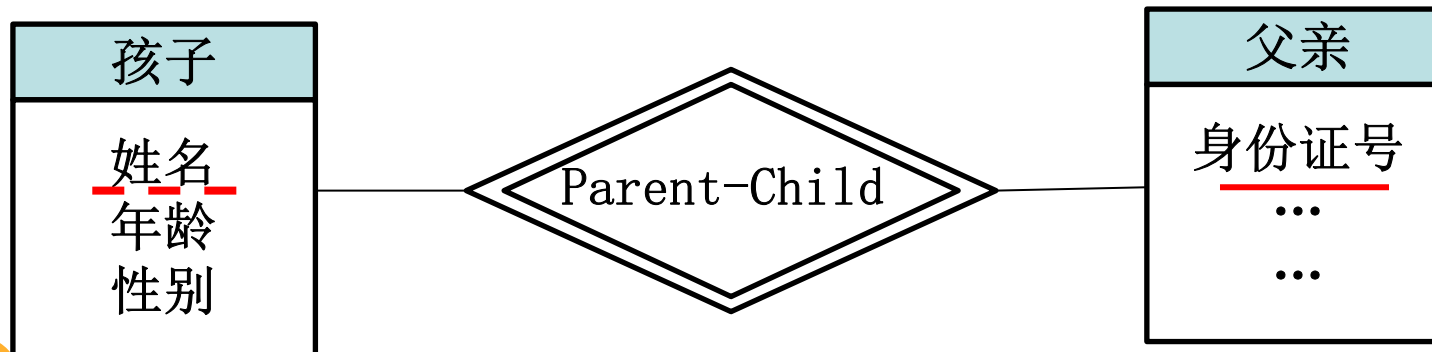


# 其他概念

## • 弱实体集(例)

### — 父亲实体集与孩子实体集

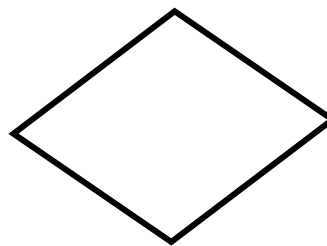
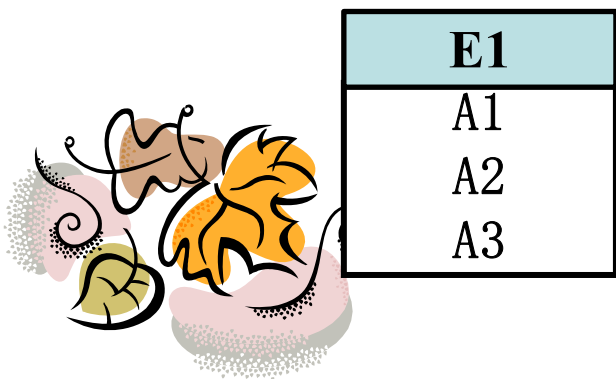
- 不同父亲的孩子可以具有相同的姓名、年龄和性别
- 同一个父亲的孩子一定具有不同的名字
- 显然，孩子实体集是弱实体集





# 实体联系图

- 实体联系图(ER图)
  - 是表示ER模型的图形工具
  - ER图用来表示实体集和实体联系集
    - 矩形：表示实体集。
    - 矩形里面文字：表示属性
    - 菱形：表示联系集。
    - 线段：将实体集连接到联系集。

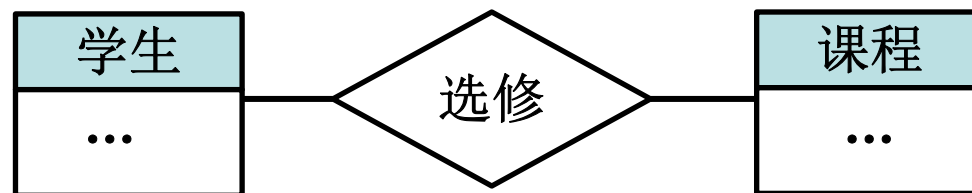
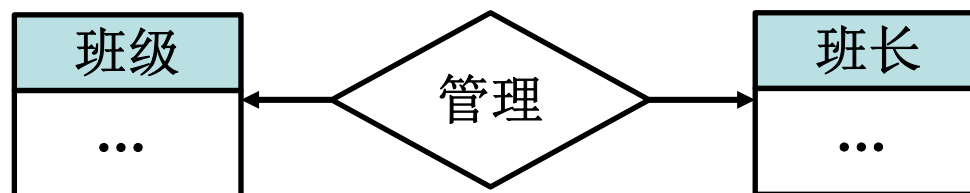




# 实体联系图

- 两个不同实体集间的联系

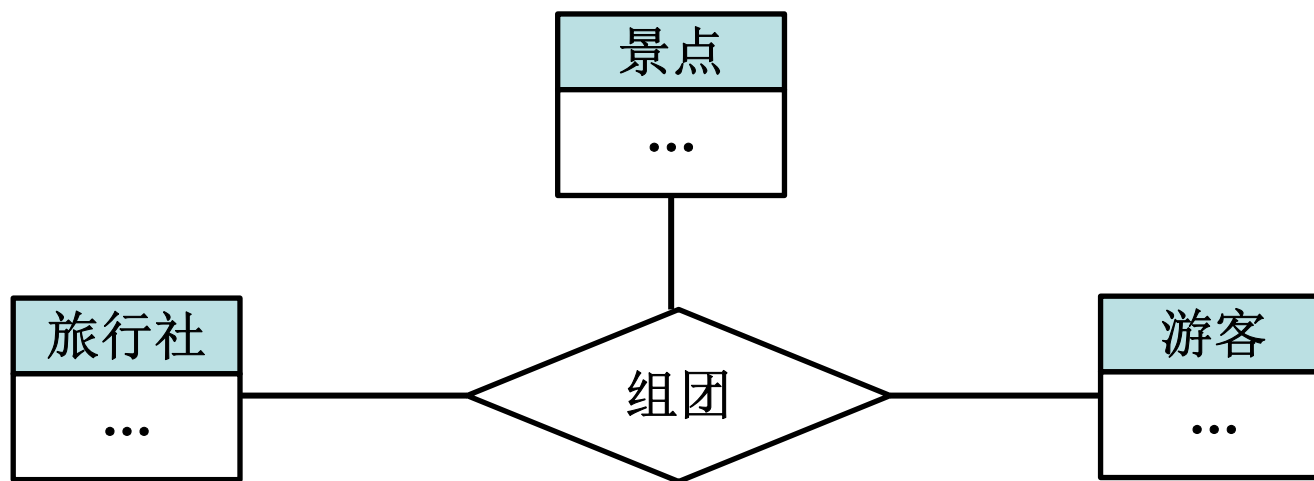
- 画出“学生”、“班级”的联系；
- 画出“班级”、“班长”的联系；
- 画出“学生”、“课程”的联系





# 实体联系图

- 两个以上实体间的多元联系  
— 画出“旅行社”、“景点”、“游客”的联系

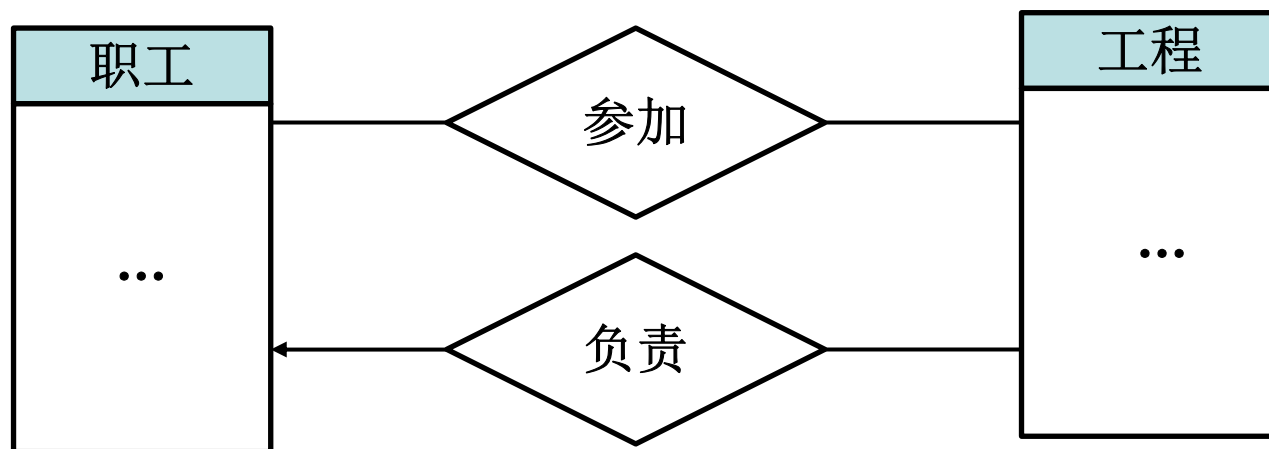




# 实体联系图

- 两个不同实体集间的多种联系

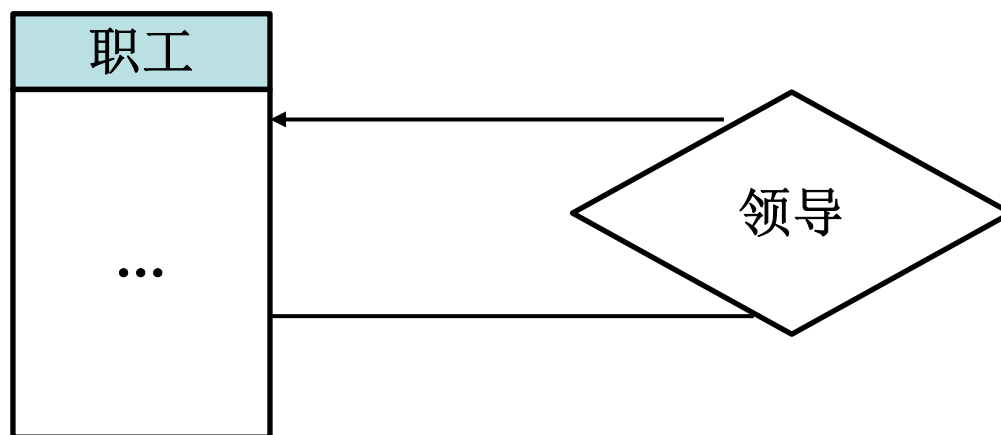
- “职工”与“工程”间，一个职工可以参加多个工程，一个工程可以有多个职工参加，同时一个工程由一个职工负责，一个职工可以负责多个工程





# 实体联系图

- 同一实体内部个体间的二元联系  
— 一个职工负责人领导一个小组





# ER图设计实例

根据财经学院数据库应用需求，运用E-R模型的基本方法设计财经学院教学数据库的概念模型。

1、根据需求调查，初步确定所关心的数据对象：

学院、系、教师、班级、学生、课程、成绩

2、根据业务规则，设计初步E-R模型

(1) 学院有多个系，每个系只能属于一个学院。

(2) 每个系有多个班级，而每个班级只能属于一个系。

(3) 每个系聘任多名教师，而每个教师又只能属于一个系。

(4) 每个班级多名学生，而每个学生又只能属于一个班。

(5) 每门课程只由一位老师讲授，而每个教师却能讲多门课。

(6) 每个课由多名学生选修，每个学生可选多门课。





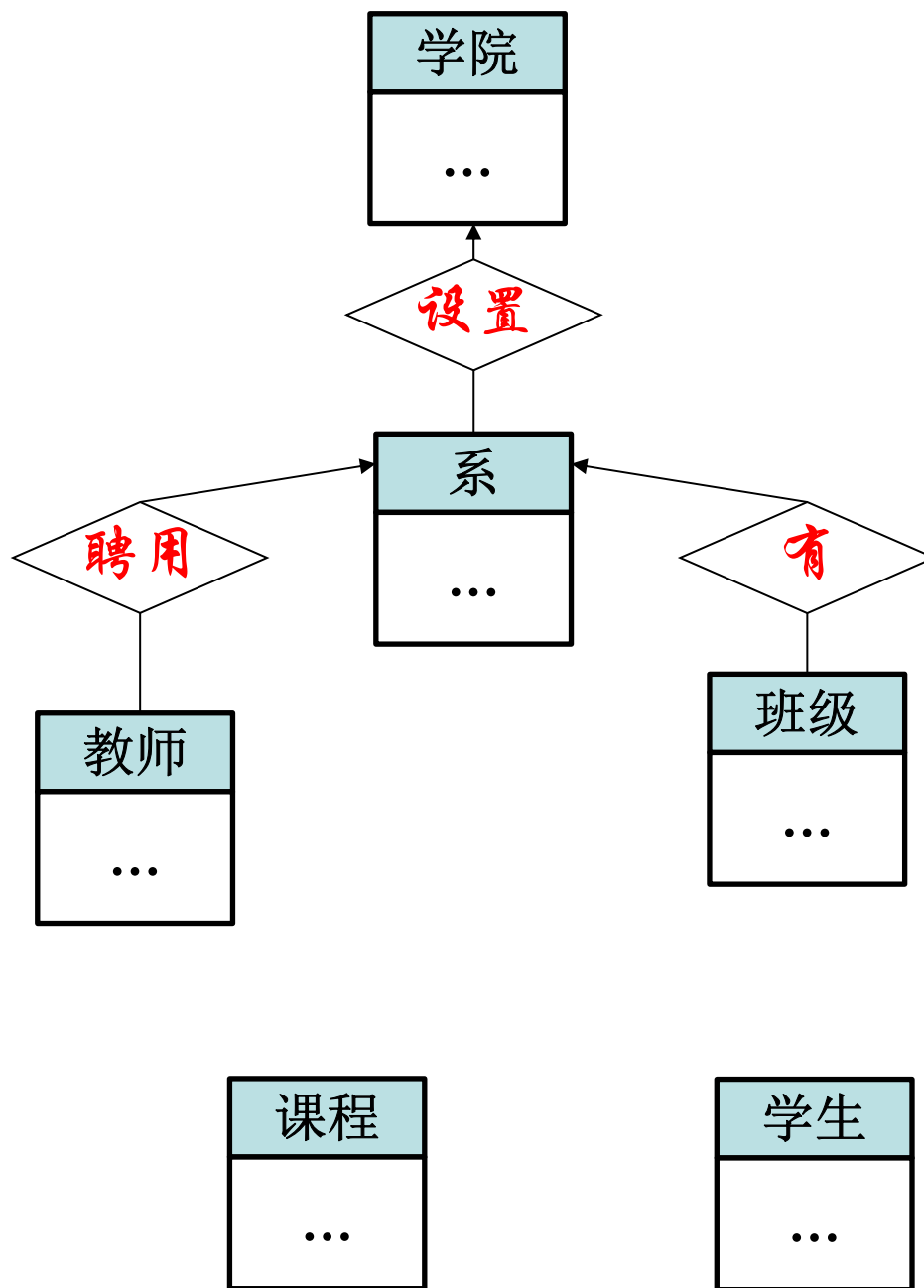


归纳上述6项可定义6个实体:学院、系、班级、教师、学生、课程。

(1) 学院有多个系, 每个系只能属于一个学院。

(2) 每个系有多个班级, 而每个班级只能属于一个系。

(3) 每个系聘任多名教师, 而每个教师又只能属于一个系



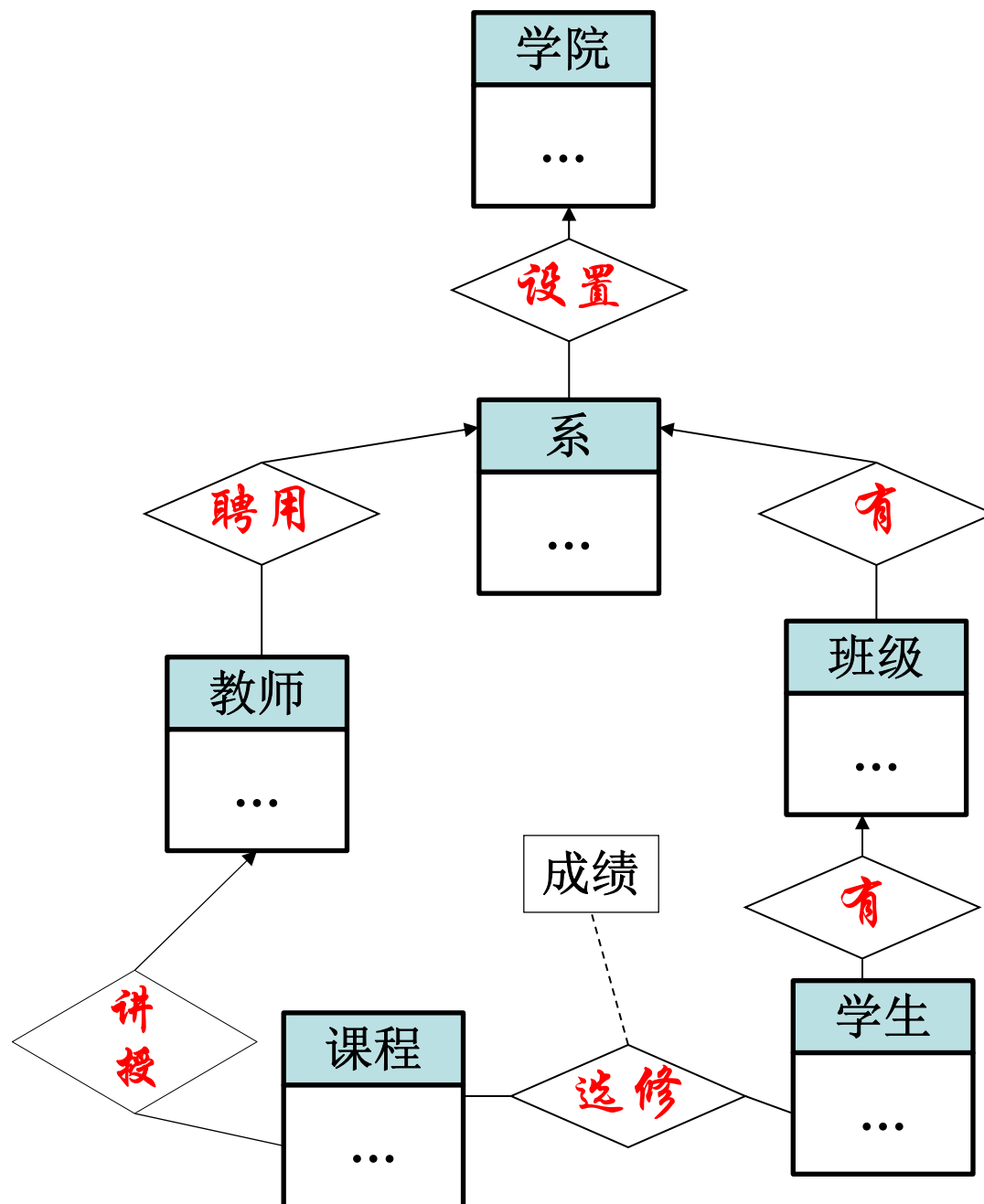


6个实体:学院、系、班级、教师、学生、课程。

(4) 每个班级多名学生,  
而每个学生又只能属于  
一个班。

(5) 每门课程只由一位老  
师讲授, 而每个教师却能讲  
多门课。

(6) 每个课由多名学生选  
修, 每个学生可选多门  
课。



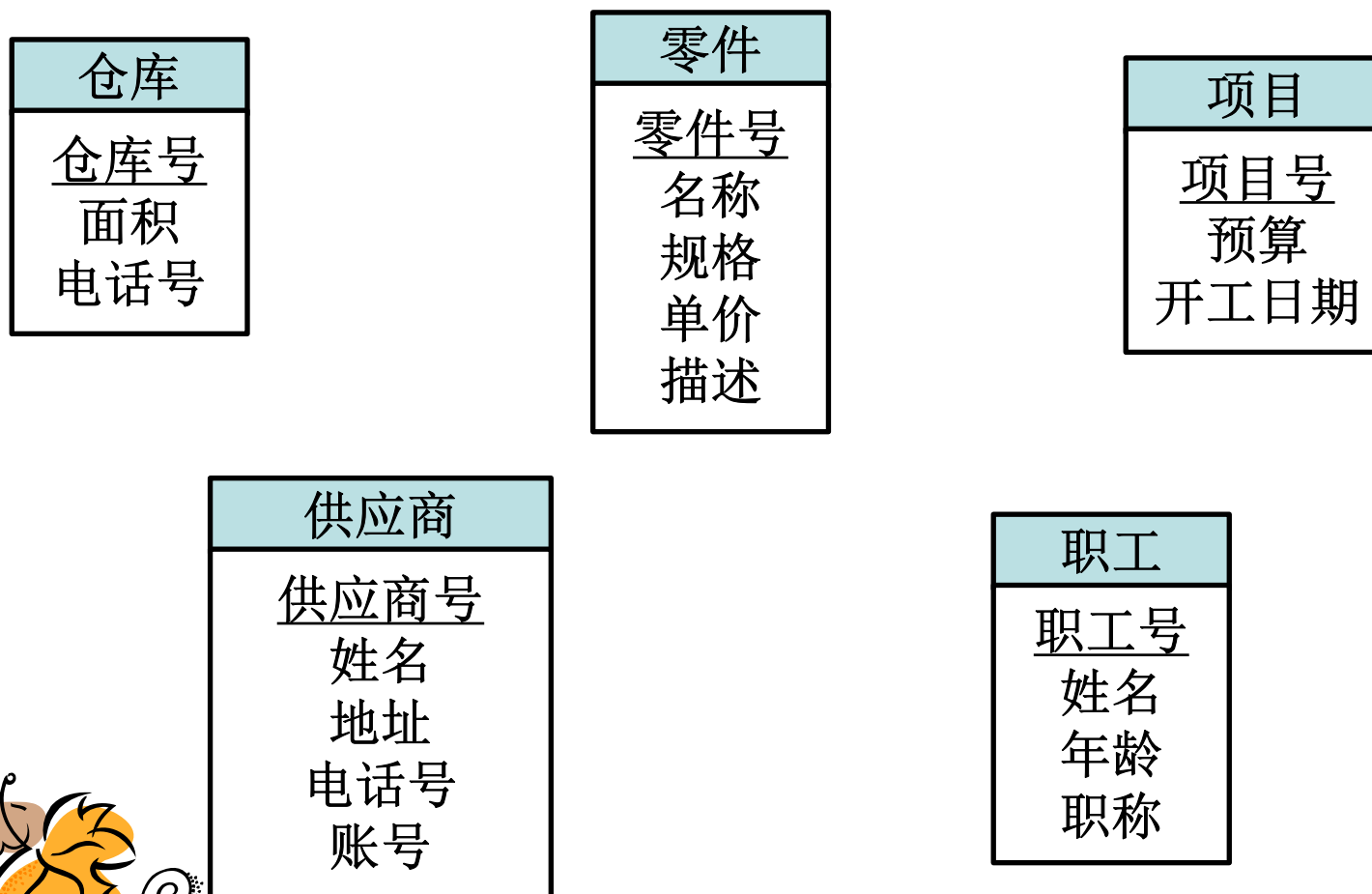
## • 实例

- 用E-R图来表示某个工厂物资管理的概念模型
- 物资管理涉及的实体有
  - 仓库 属性有仓库号、面积、电话号码
  - 零件 属性有零件号、名称、规格、单价、描述
  - 供应商 属性有供应商号、姓名、地址、电话号码、账号
  - 项目 属性有项目号、预算、开工日期
  - 职工 属性有职工号、姓名、年龄、职称

这些实体之间的联系如下：

- (1) 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，因此仓库和零件具有多对多的联系。用库存量来表示某种零件在某个仓库中的数量。
- (2) 一个仓库有多个职工当仓库保管员，一个职工只能在一个仓库工作，因此仓库和职工之间是一对多的联系
- (3) 职工之间具有领导-被领导关系。即仓库主任领导若干保管员，因此职工实体集中具有一对多的联系。
- (4) 供应商、项目和零件三者之间具有多对多的联系。即一个供应商可以供给若干项目多种零件，每个项目可以使用不同供应商供应的零件，每种零件可由不同供应商供给。

- 仓库 属性有仓库号、面积、电话号码
- 零件 属性有零件号、名称、规格、单价、描述
- 供应商 属性有供应商号、姓名、地址、电话号码、账号
- 项目 属性有项目号、预算、开工日期
- 职工 属性有职工号、姓名、年龄、职称



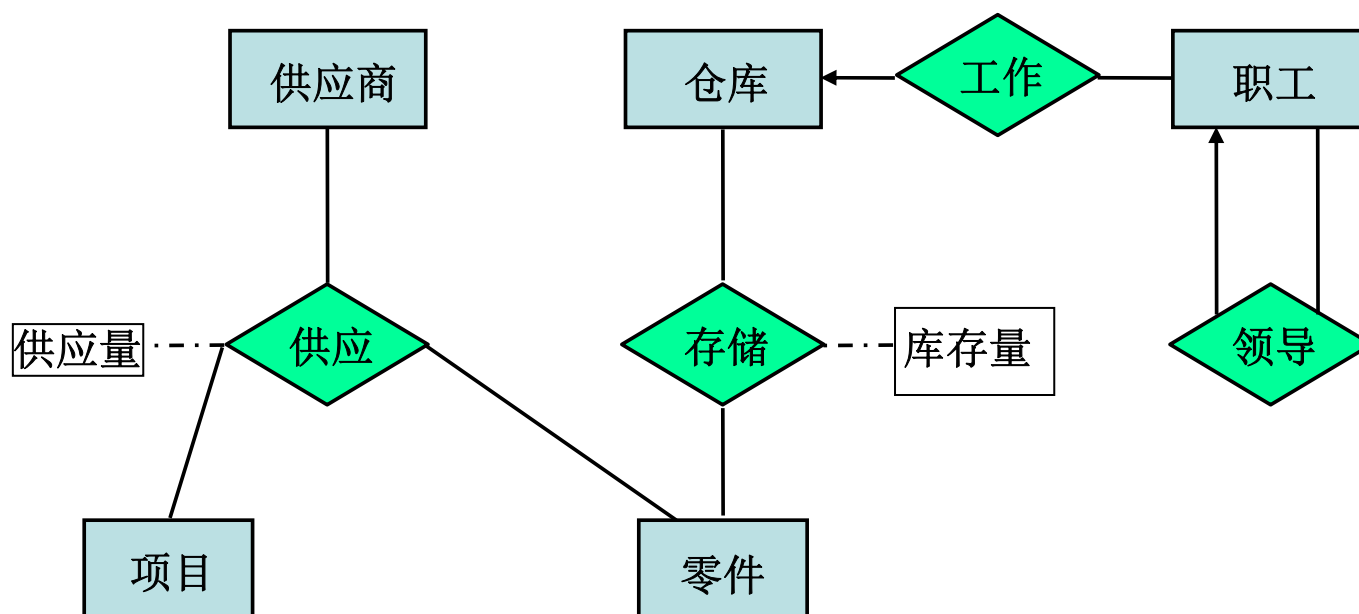
(a) 实体及其属性图

实体之间的联系：

- (1) 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**仓库和零件具有多对多的联系**。用库存量来表示某种零件在某个仓库中的数量。
- (2) 一个仓库有多个职工当仓库保管员，一个职工只能在一个仓库工作，因此**仓库和职工之间是一对多的联系**

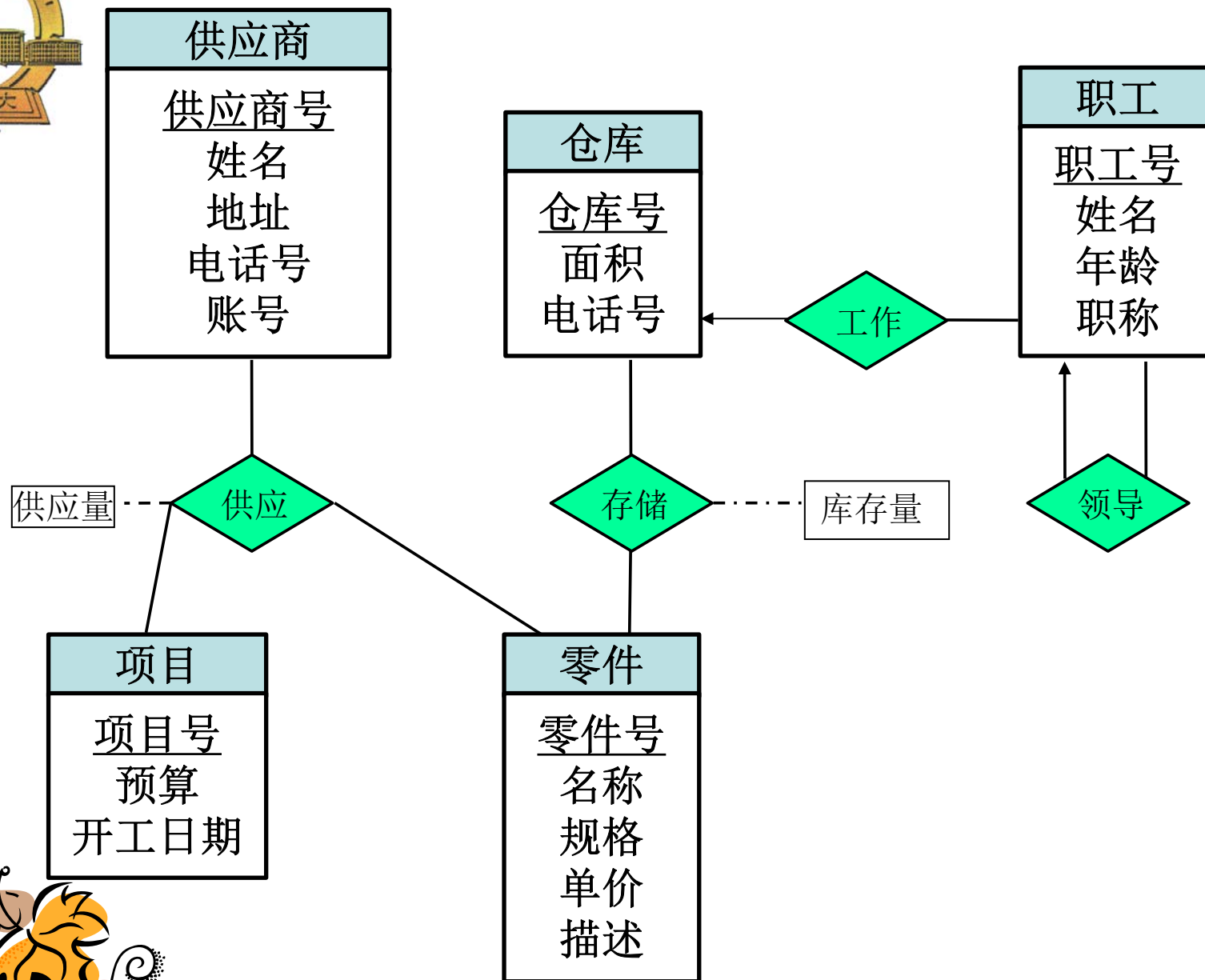
(3) 职工之间具有领导-被领导关系。即仓库主任领导若干保管员，因此**职工实体集中具有一对多的联系**。

(4) 供应商、项目和零件三者之间**具有多对多的联系**。即一个供应商可以供给若干项目多种零件，每个项目可以使用不同供应商供应的零件，每种零件可由不同供应商供给。



(b) 实体及其联系图





(c) 完整的实体联系图



# 目录

- 概述
- 实体联系模型
- 概念数据库设计的方法与策略
- 视图综合设计方法
- 事务的设计





## 4.2.3 概念数据库设计方法

- 集中式设计方法

- 合并在需求分析阶段得到的各种应用需求
- 在上述基础上设计一个概念数据库模式，满足所有应用的需求

- 视图综合设计法

- 不要求应用需求合并

- 视图设计阶段

- 根据每个应用的需求，独立地为每个用户和应用设计一个概念数据库模式。每个应用的概念数据库模式称为一个视图

- 视图合并阶段

- 把所有视图有机合并成一个统一的概念数据库模式，支持所有应用

视图综合设计方法已经成为目前的重要概念设计方法





## 4.2.3 概念数据库设计策略

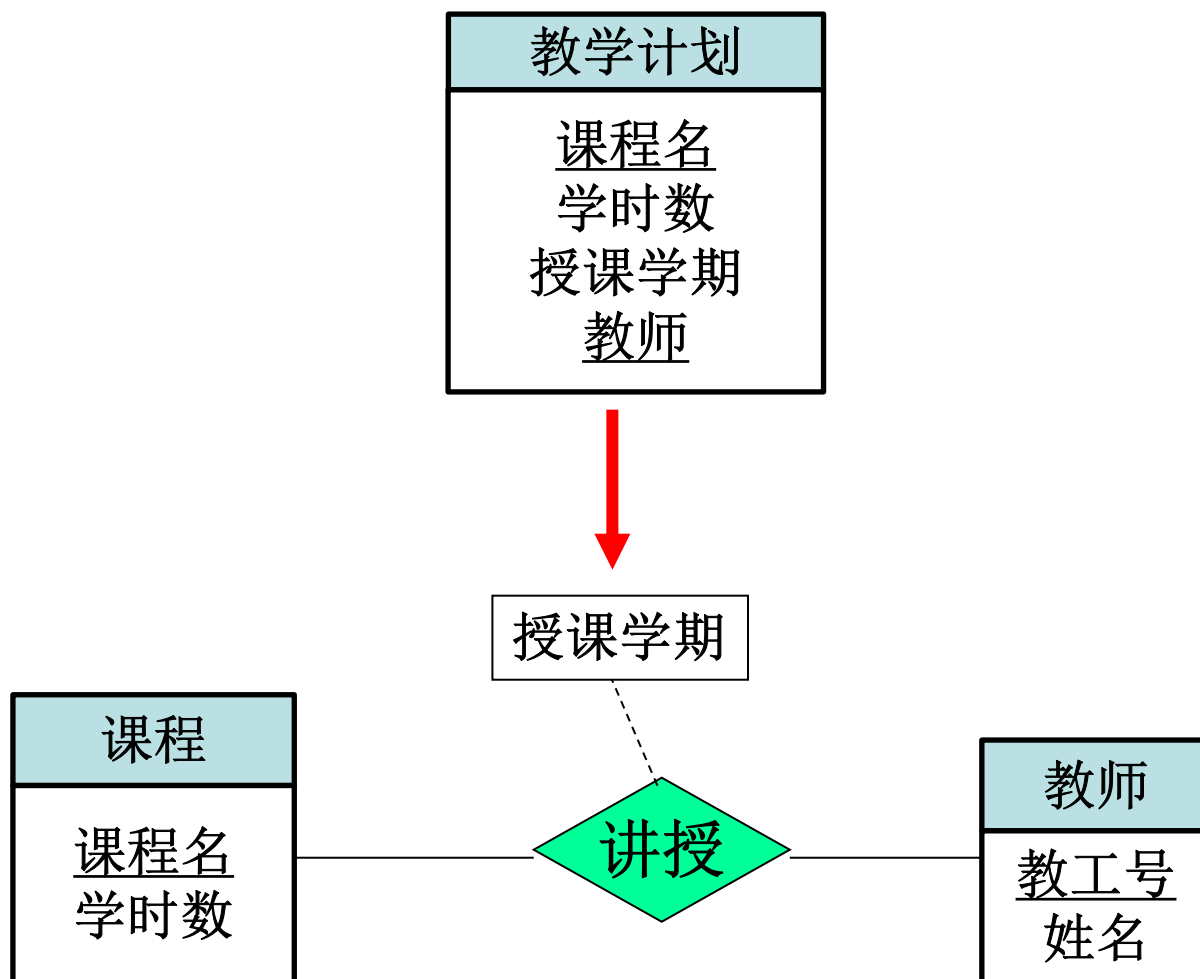
- 自顶向下的策略

- 从一个包含高级抽象概念结构的模式出发，对这些高级抽象概念结构逐步求精，形成最终的概念数据库模式





## 一个实体集细分为两个实体集和一个联系集的实例





# 概念数据库设计策略

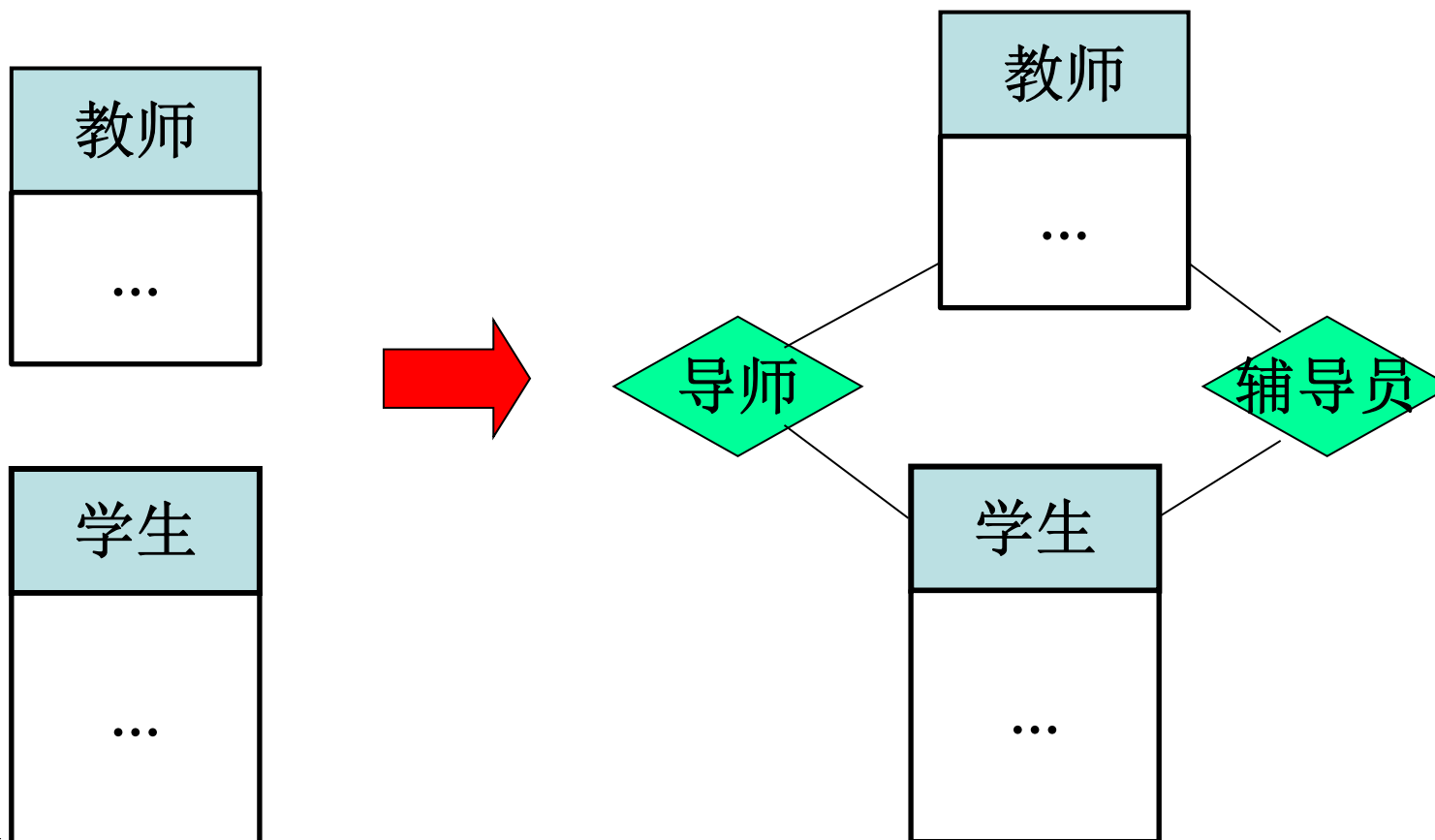
- 自底向上的策略

- 从包含基本概念结构的模式出发，逐步组合这些基本概念结构，形成最终的概念数据库模式。





# 概念数据库设计策略



增加两个联系的实例





# 概念数据库设计策略

- 混合策略

- 首先使用**自顶向下**的策略把应用需求划分为多个需求子集合；
- 然后，对于每个需求子集合，使用**自底向上**的策略设计局部模式；
- 最后，组合局部模式，形成最后的概念数据库模式。





# 目录

- 概述
- 实体联系模型
- 概念数据库设计的方法与策略
- **视图综合设计方法**
- 事务的设计





## 4.2.4视图综合设计方法

- 视图综合设计方法
  - 第一步设计局部概念模式。
  - 第二步把局部概念模式合并成一个完整的全局概念模式，即最终的概念数据库模式。





## 4.2.4视图综合设计方法

- 局部概念模式设计

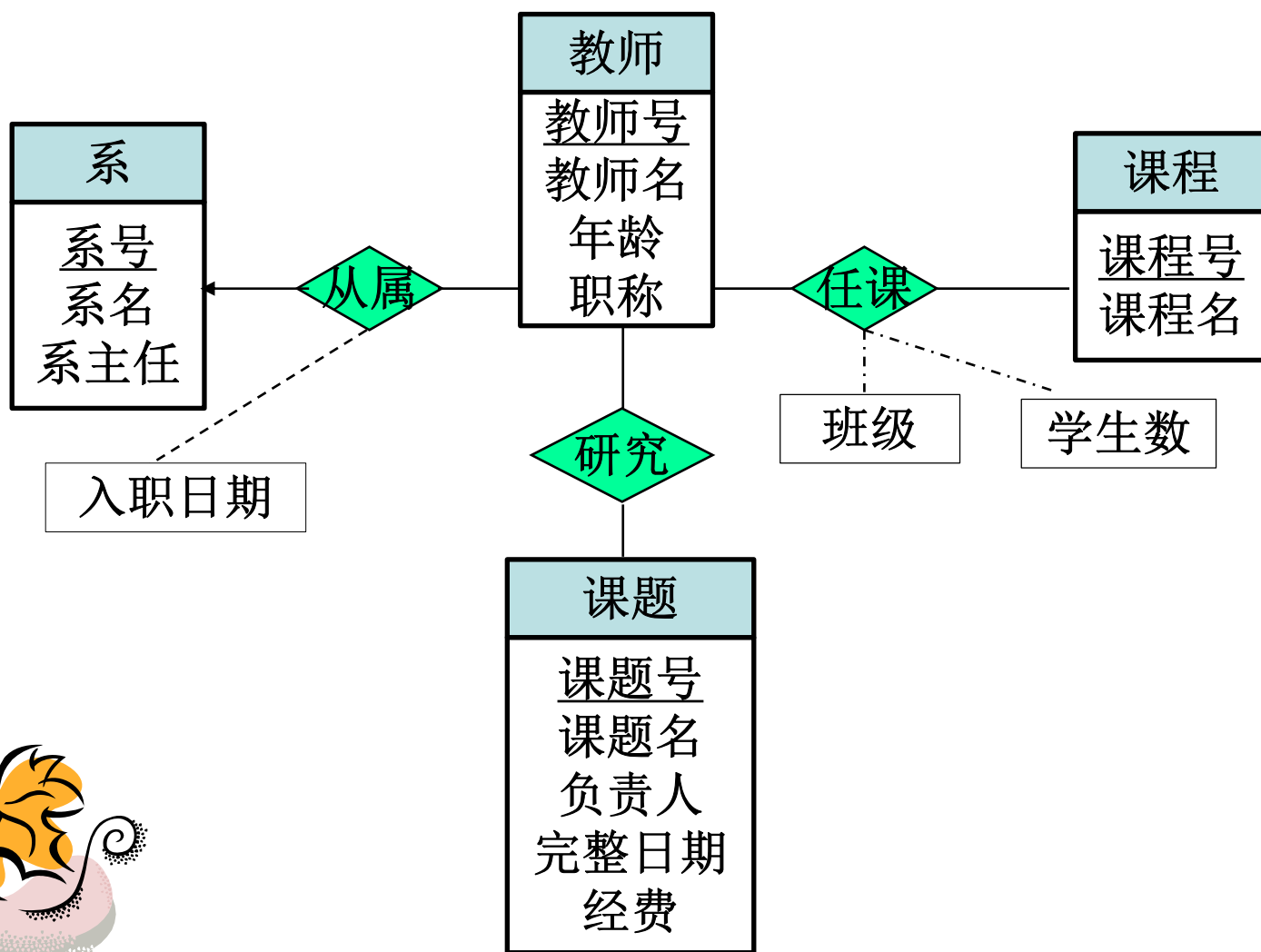
- 抽取各局部应用涉及的数据，标定各局部应用中的实体、实体的属性、标识实体的键，确定实体之间的联系及其约束 (1:1, 1:n, m:n)
- 形成局部E-R图







## 局部E-R设计实例：学校教学管理,假定它涉及到 师资、教务等部门。

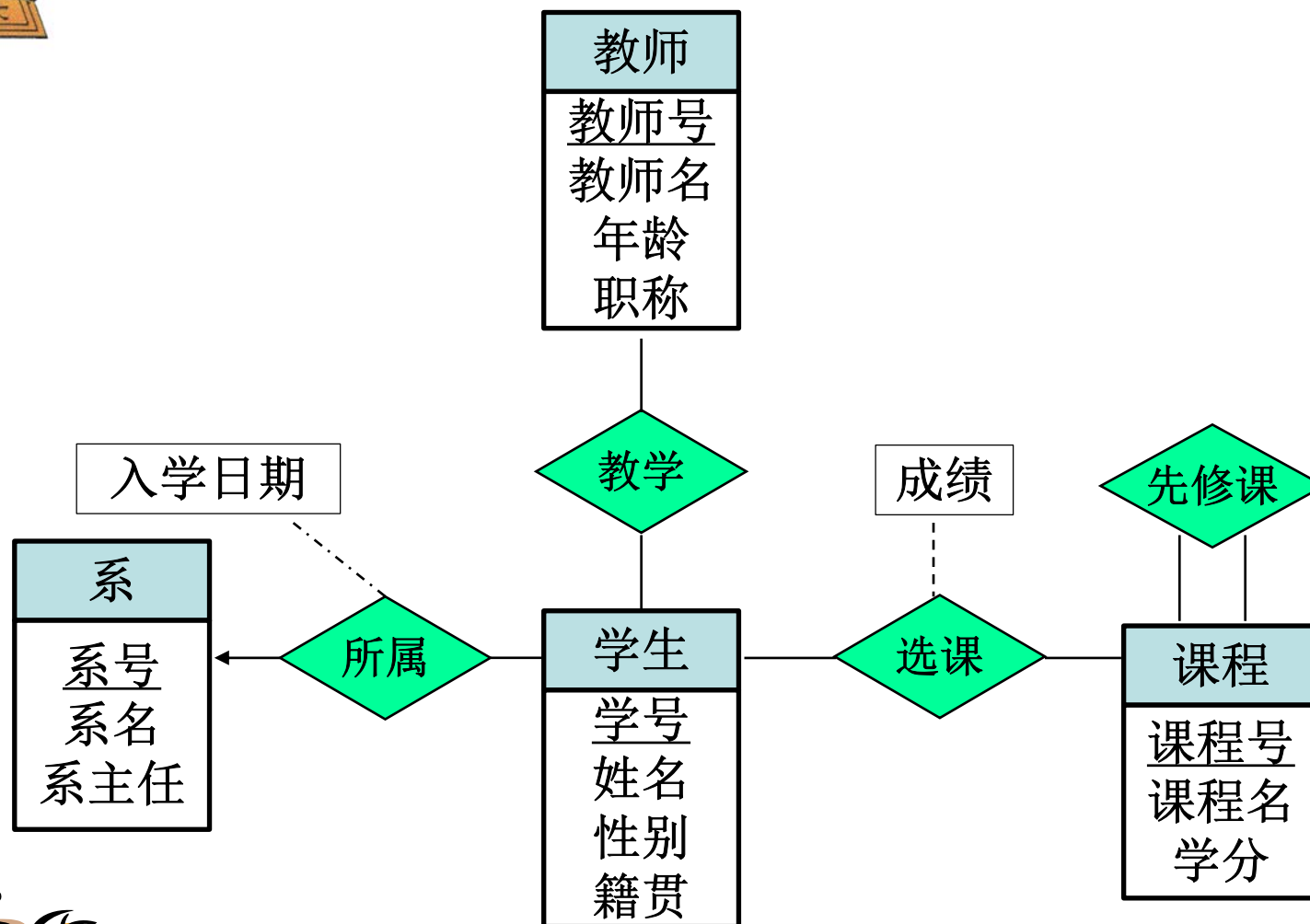


师资部门局部E-R图





## 教务部门局部E-R图





## 4.2.4视图综合设计方法

- 全局概念模式的合成
  - 各个局部概念模式即局部E-R图建立好后，还需进行合并，集成为一个总E-R图
  - 集成E-R图的步骤
    - 1、识别局部概念模式间的冲突
    - 2、修改局部模式
    - 3、局部模式合并
    - 4、优化全局模式





## 4.2.4视图综合设计方法

- 冲突的种类

- 命名冲突：异名同义，同名异义

- 模式结构冲突

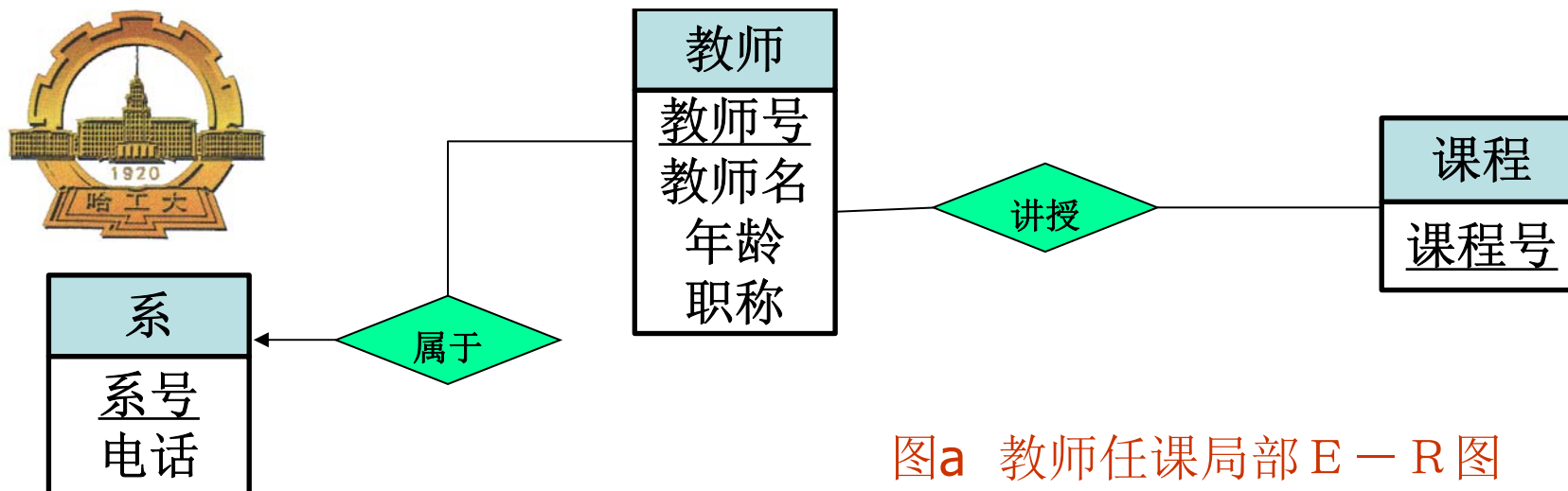
- 相同概念在不同的局部模式中使用不同的概念模式表示

- 例如，“系”在一个局部模式中表示为一个实体，而在另一个模式中则被表示成一个复合属性

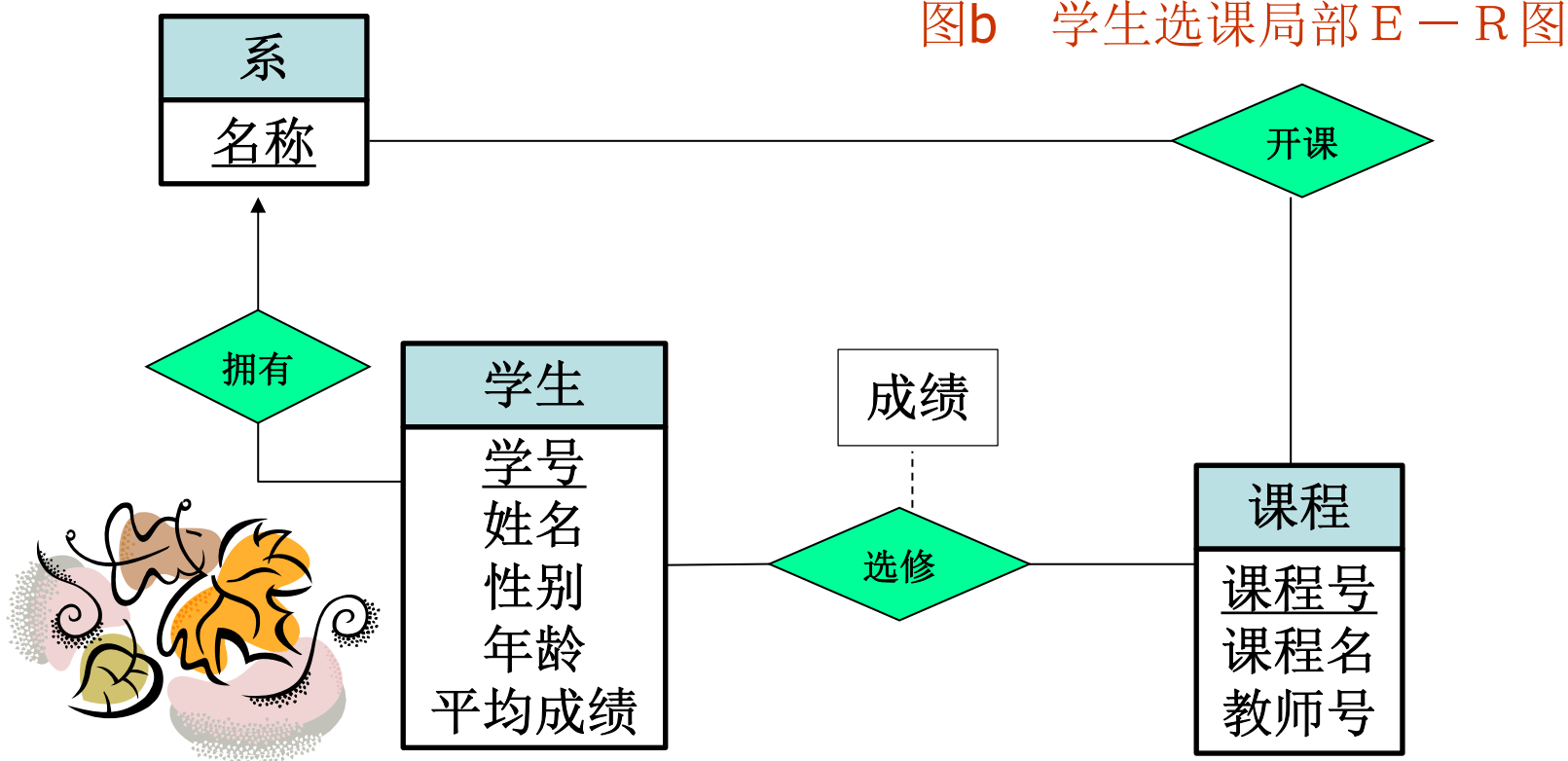
- 值域冲突

- 约束冲突





图a 教师任课局部 E—R 图



图b 学生选课局部 E—R 图



## 4.2.4视图综合设计方法



- 全局模式合成方法
  - 二元阶梯合成法
  - N元合成法
  - 平衡二元合成法
  - 混合合成法





# 目录

- 概述
- 实体联系模型
- 概念数据库设计的基本步骤
- 视图综合设计方法
- 事务的设计





# 事务的设计

- 事务

- 一个或多个数据操作构成的集合，这组操作满足原子性。

- 例如，银行从账户A到账户B的一次资金转账操作

- 事务设计任务：定义事务功能

- 说明事务的输入、输出







# 小结

- 概念数据库设计：
  - 实体联系模型、ER模型
  - 实体、实体集
  - 实体的属性、实体的属性值、复合属性、单值属性、多值属性、导出属性、空值
  - 实体间的联系
  - 弱实体、ER图





# 小结

- 本章重点:

- 熟练掌握实体-联系模型
- 熟练掌握实体-联系图
- 掌握概念数据库设计方法





**Now let's go to  
Next Chapter**

