

# 数据库系统原理

教程：数据库系统概论（第5版）

结合：CMU 15-445/645 INTRO TO DATABASE SYSTEMS

华中科技大学 计算机学院

左琼



# 第七章 数据库设计

*Principles of Database Systems*

# 第七章 数据库设计

## 7.1 数据库设计概述

## 7.2 需求分析

## 7.3 概念结构设计

## 7.4 逻辑结构设计

## 7.5 数据库的物理设计

## 7.6 数据库实施和维护

## 7.7 小结

# 7.1 概述

## 数据库设计的基本任务：

- 数据库设计是指对于一个给定的应用环境，构造优化的数据库模式，建立数据库及其应用系统，使之能够有效的存储和管理数据，满足用户的应用需求（包括信息要求和处理要求）。
- 数据库设计通常是在已有硬件和OS平台上，利用一个通用的DBMS来建立能够实现系统目标的数据库。



# 数据库设计的内容

## □ 数据库的结构设计（静态）

- 根据给定的应用环境，进行数据库的**模式/子模式**的设计。
- 包括数据库的**概念设计、逻辑设计和物理设计**。
- 数据库模式是各应用程序共享的结构，是静态的、稳定的，一经形成后通常情况下是不容易改变的，所以结构设计又称为静态模型设计。

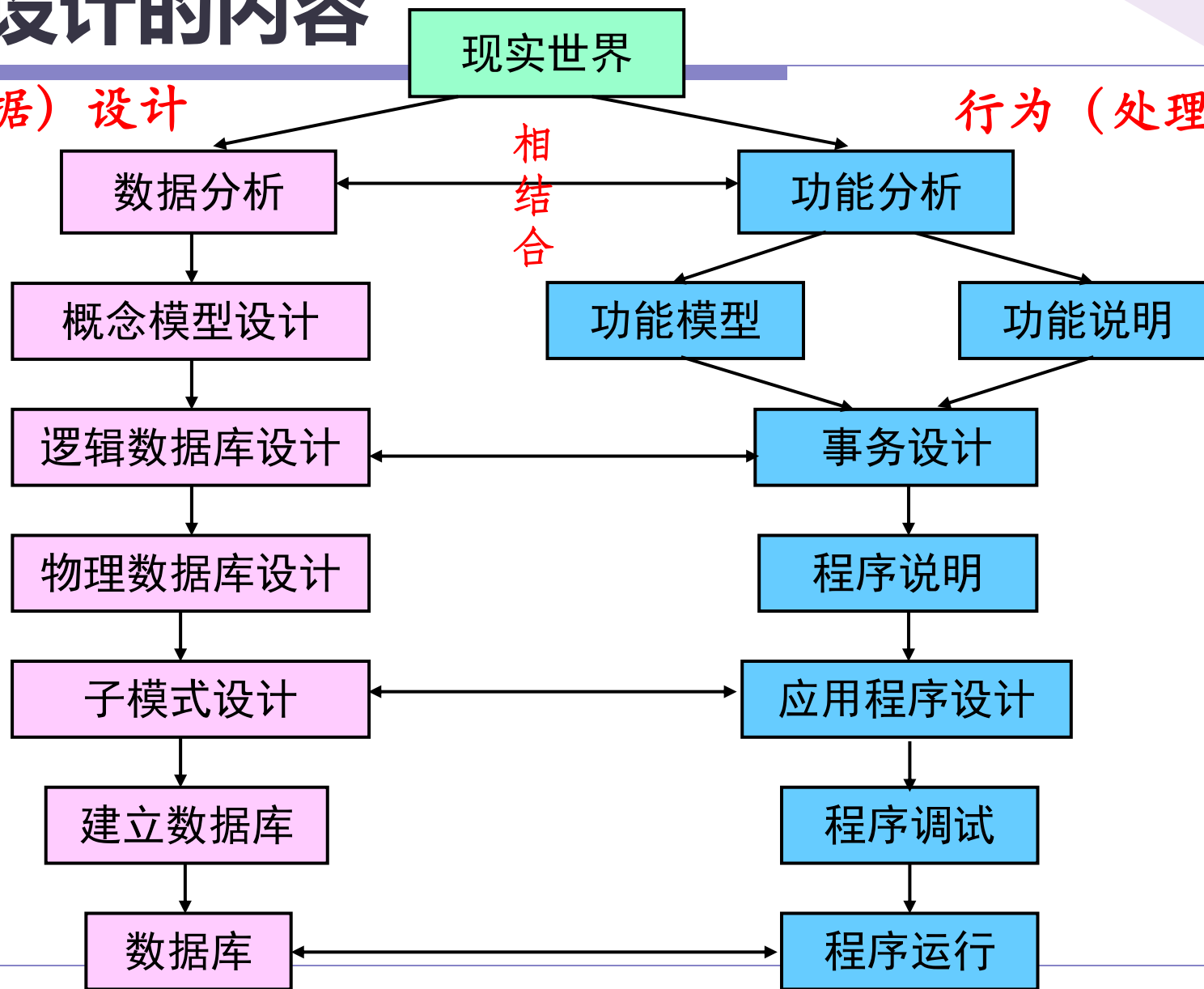
## □ 数据库的行为设计（动态）

- 是指确定数据库用户的行为和动作。而在数据库系统中，用户的行为和动作指用户对数据库的操作，这些要通过应用程序来实现，所以数据库的行为设计就是**应用程序的设计**。
- 用户的行为总是使数据库的内容发生变化，所以行为设计是动态的，行为设计又称为动态模型设计。

# 数据库设计的内容

结构（数据）设计

行为（处理）设计



# 数据库设计的特点

- 结构与行为设计相结合
  - 面向数据的设计方法（以信息需求为主）
  - 面向过程的设计方法（以处理需求为主）
- 分步进行
  - 数据库设计分为多个阶段
  - 前一阶段的设计结果作为后一阶段设计的依据
  - 后一阶段也可向前面的设计阶段反馈其要求
- 反复性
  - 反复设计、逐步求精
- 多解性
  - 设计结果不唯一，多种方案并存

## 7.1.2 数据库设计方法

□ 数据库设计方法可分为4类：

1. 直观设计法：也叫手工试凑法。
2. 规范设计法
3. 计算机辅助设计法
4. 自动化设计法

在数据库设计的某些过程中模拟某一规范化设计的方法，并以人的知识或经验为主导，通过人机交互方式实现设计中的某些部分。

- 目前许多计算机辅助软件工程 (CASE) 工具可以自动或辅助设计人员完成数据库设计过程中的很多任务。
- 如：SYBASE公司的 PowerDesigner 和 Oracle公司的Design 2000。

- 1978年10月提出新奥尔良法：将数据库设计分成多个阶段：需求分析、概念设计、逻辑设计和物理设计。

- 常用的规范设计法：

- 基于ER模型的数据库设计方法
- 基于3NF的数据库设计方法
- ODL(Object Definition Language)方法

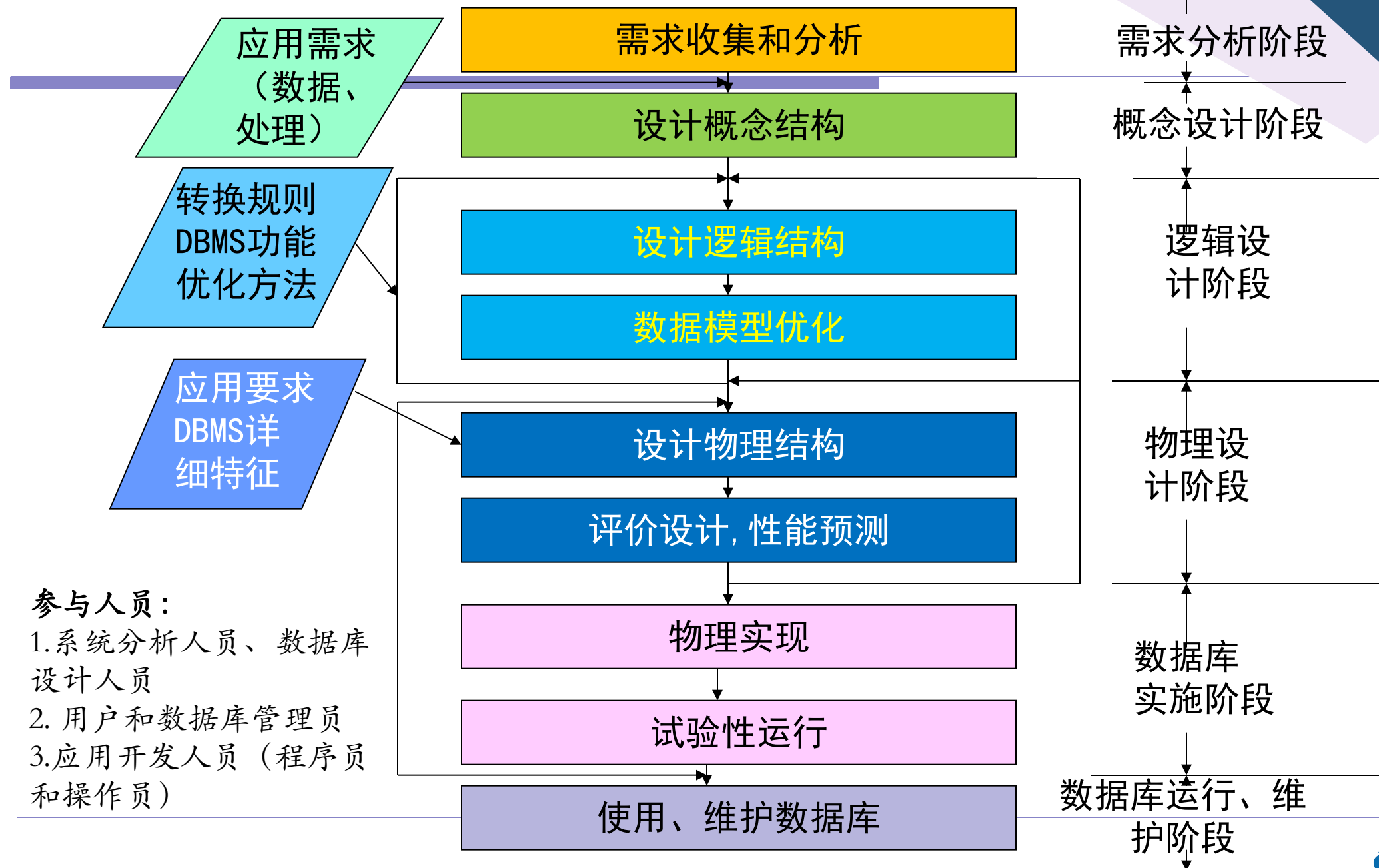
本质上仍是手工设计方法，基本思想是过程迭代和逐步求精。



## 7.1.3 数据库设计的步骤

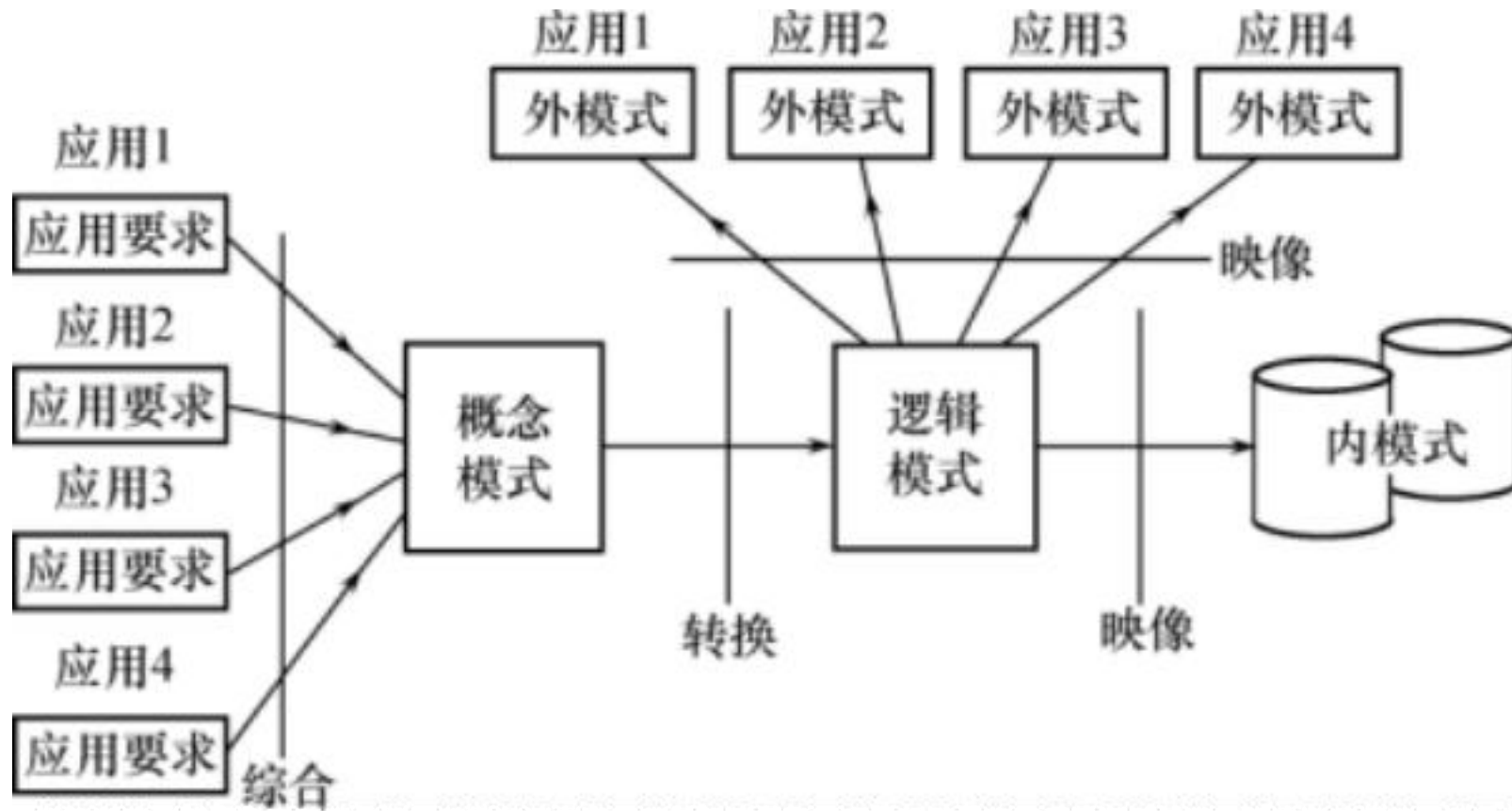
- 按规范设计法可将数据库设计分为六个阶段：
  - 需求分析阶段
  - 概念结构设计阶段
  - 逻辑结构设计阶段
  - 物理设计阶段
  - 数据库实施阶段
  - 数据库运行与维护阶段
- 需求分析和概念设计独立于任何数据库管理系统
- 逻辑设计和物理设计与选用的DBMS密切相关

# 数据库设计基本步骤



设计阶段	设计描述	
	数据	处理
需求分析	数据字典、全系统中数据项、数据流、数据存储的描述	数据流图和判定表（判定树） 数据字典中处理过程的描述
概念结构设计	概念模型（E-R图） 数据字典	系统说明书。包括： (1) 新系统要求、方案和概图 (2) 反映新系统信息的数据流图
逻辑结构设计	某种数据模型 关系模型或非关系模型	系统结构图（模块结构）
物理设计	存储安排、存取方法选择、存取路径建立	模块设计（IPO表）
实施阶段	编写模式、装入数据、数据库试运行	程序编码、编译联结、测试
运行维护	性能测试、转储/恢复、数据库重组和重构	新旧系统转换、运行、维护（修正性、适应性、改善性维护）

# 数据库设计过程中的各级模式



## 7.2 需求分析

### □ 需求分析是整个数据库设计的起点：

需求分析的结果是否准确反映了用户的实际要求，将直接影响后面各阶段的设计，并影响到设计结果是否合理和实用。

#### 7.2.1 需求分析的任务

##### ■ 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）；
- 明确用户的各种需求；
- 确定新系统的功能；
- 充分考虑今后可能的扩充和改变。

## 7.2.1 需求分析的任务

□ 调查的重点：“数据”和“处理”，获得用户对数据库要求：

■ **信息要求** 指目标范围内涉及的所有实体、实体属性以及实体间的联系等数据对象，也就是用户需要从数据库中获得信息的内容与性质。由信息需求可以导出数据要求。

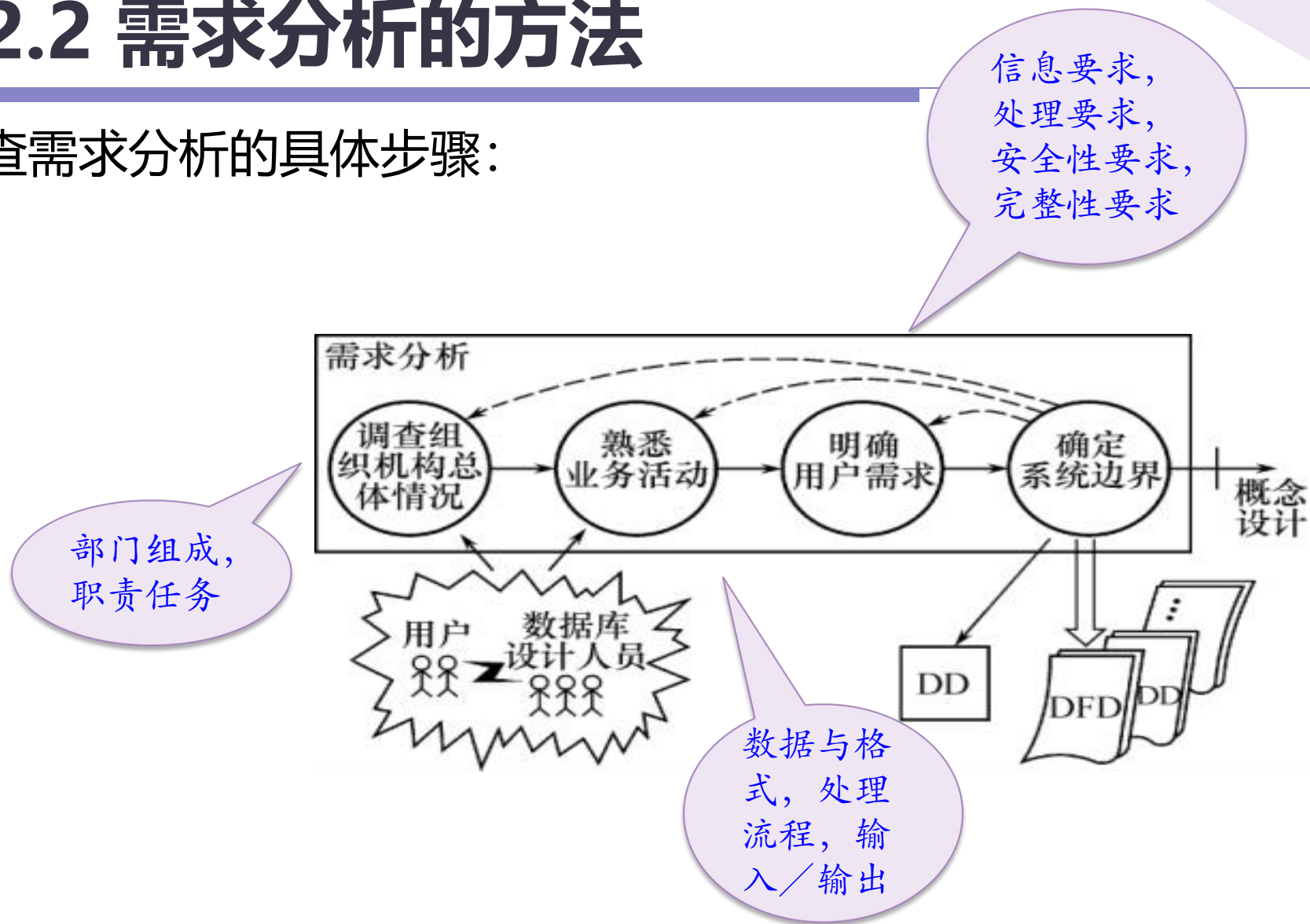
■ **处理要求** 指用户为了得到需求的信息而对数据进行加工处理的要求，包括对某种处理功能的响应时间，处理的方式（批处理或联机处理）等。

■ **安全性与完整性要求**

在定义信息需求和处理需求的同时必须相应确定安全性和完整性约束。

## 7.2.2 需求分析的方法

调查需求分析的具体步骤：



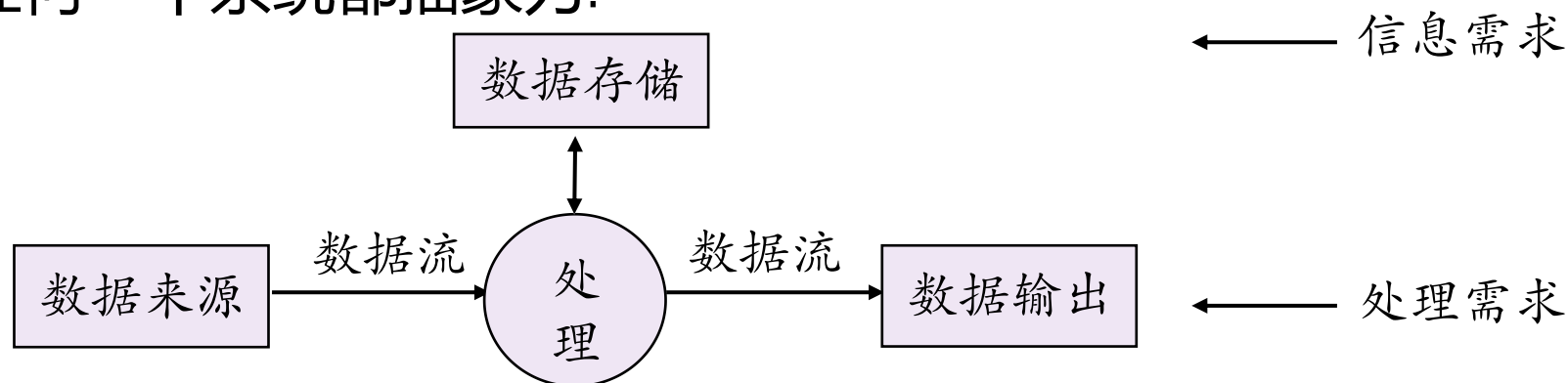
## 7.2.2 需求分析的方法

- 用于需求分析的方法有多种，主要方法有**自顶向下**和**自底向上**两种。
- 其中**自顶向下的分析方法**（Structured Analysis，简称SA方法）是最简单实用的方法。
  - 从最上层的系统组织机构入手
  - 采用逐层分解的方式分析系统
  - 用**数据流图**（Data Flow Diagram，DFD）和**数据字典**（Data Dictionary，DD）描述系统。



# 需求分析过程

1. 首先把任何一个系统都抽象为:



2. 分解处理功能和数据:

将处理功能的具体内容分解为若干子功能;

处理功能逐步分解同时, 逐级分解所用数据, 形成若干层次的数据流图DFD。

■ 表达方法:

处理逻辑: 用判定表或判定树来描述

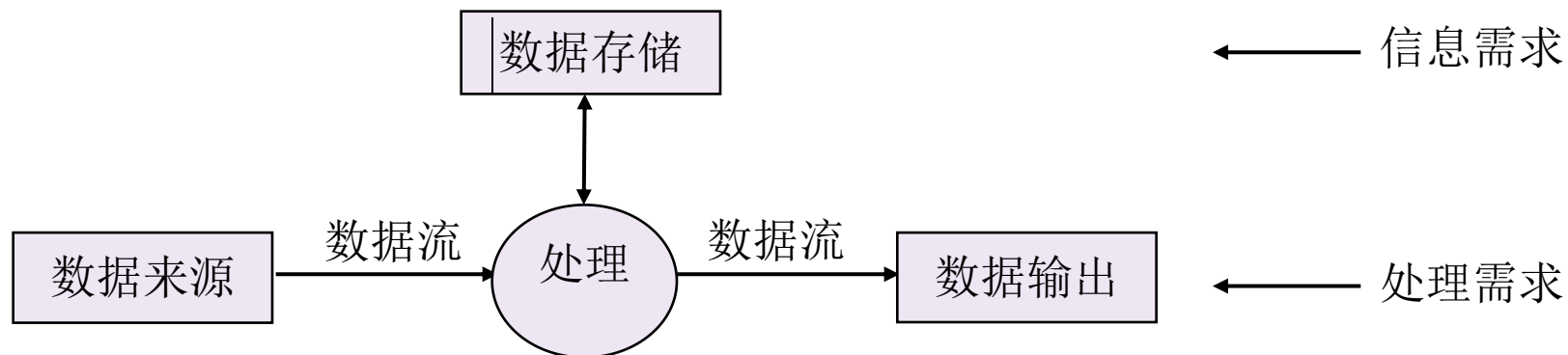
数据: 用数据字典DD来描述

3. 将分析结果再次提交给用户, 征得用户的认可

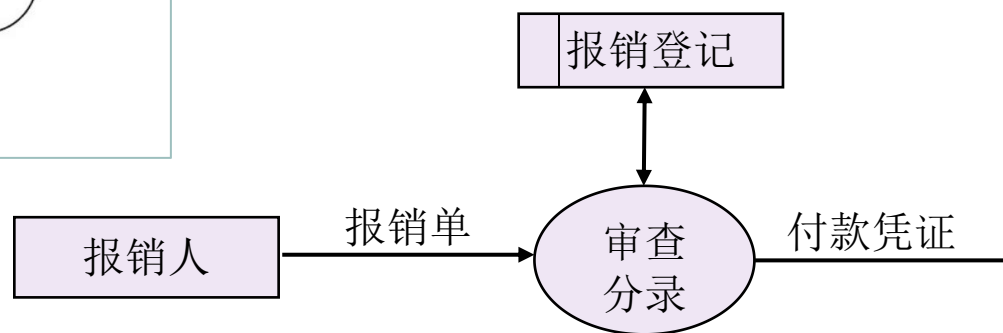
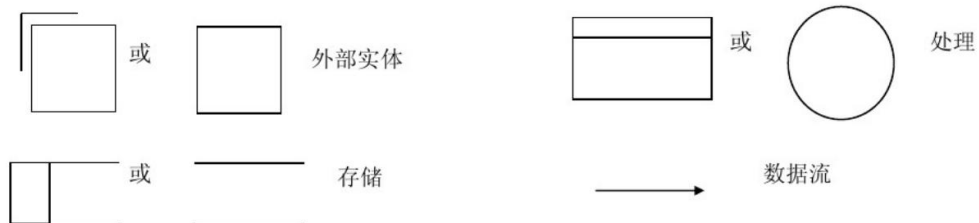
## 7.2.3 数据字典

- 数据字典是对系统中数据的详细描述，是各类数据结构和属性的清单。它与数据流图互为注释。
- 数据字典贯穿于数据库需求分析直到数据库运行的全过程，在不同的阶段其内容和用途各有区别。
- 在需求分析阶段，它通常包含以下五部分内容：
  1. **数据项**：不可再分的数据单位。
  2. **数据结构**：反映了数据之间的组合关系。
  3. **数据流**：表示某一处理过程中数据在系统内传输的路径。
  4. **数据存储**：数据的存放场所。
  5. **处理过程**：具体处理逻辑。

# 数据流图DFD

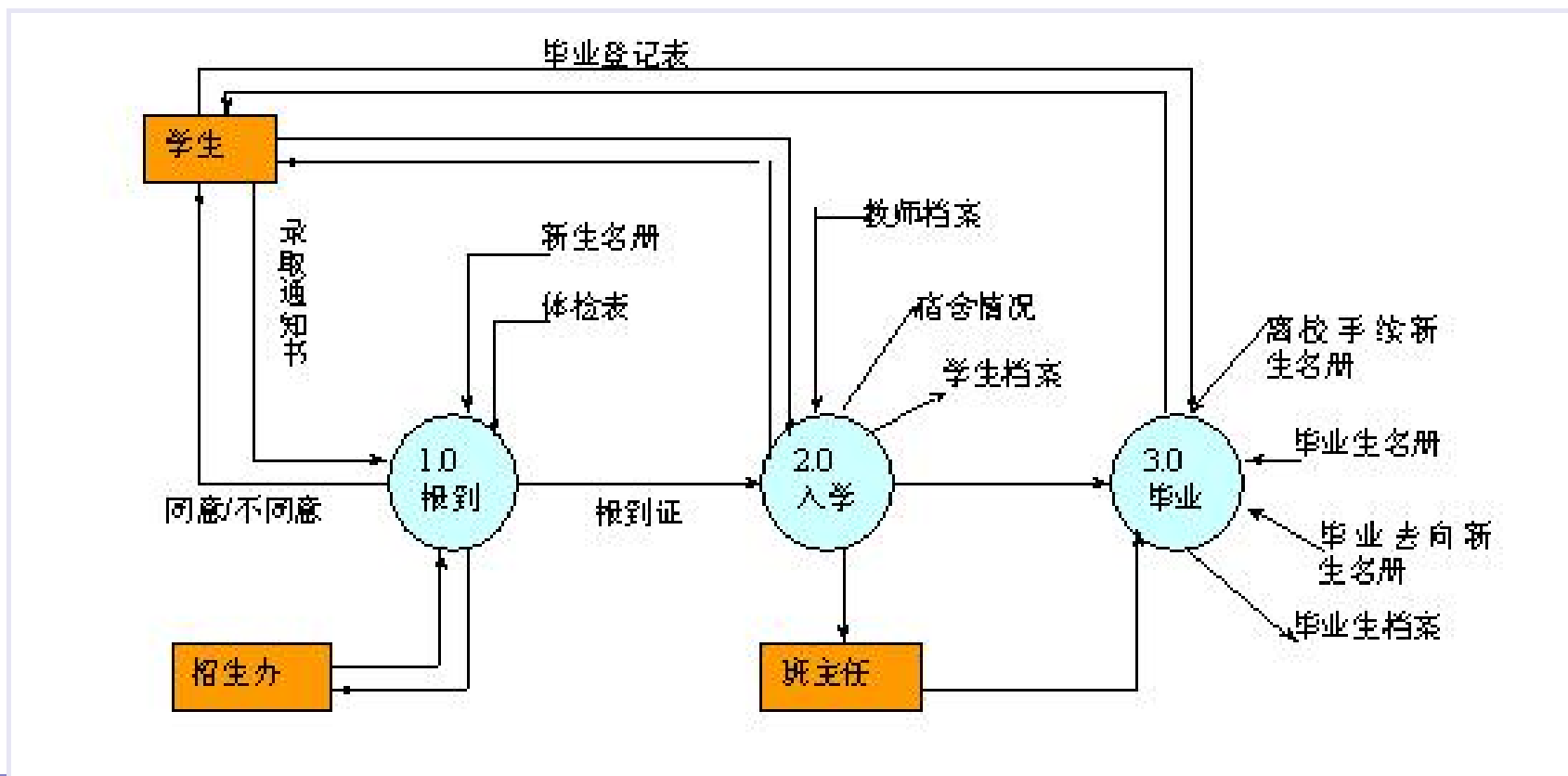


## 数据流图的表示

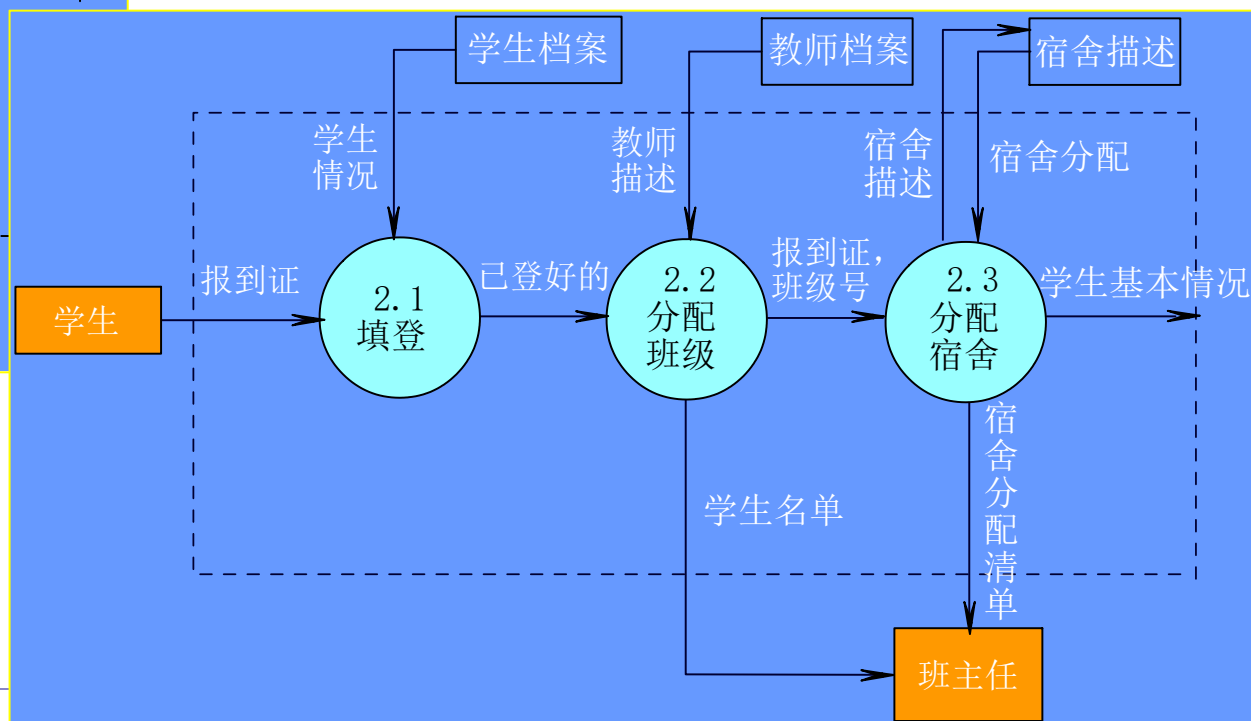
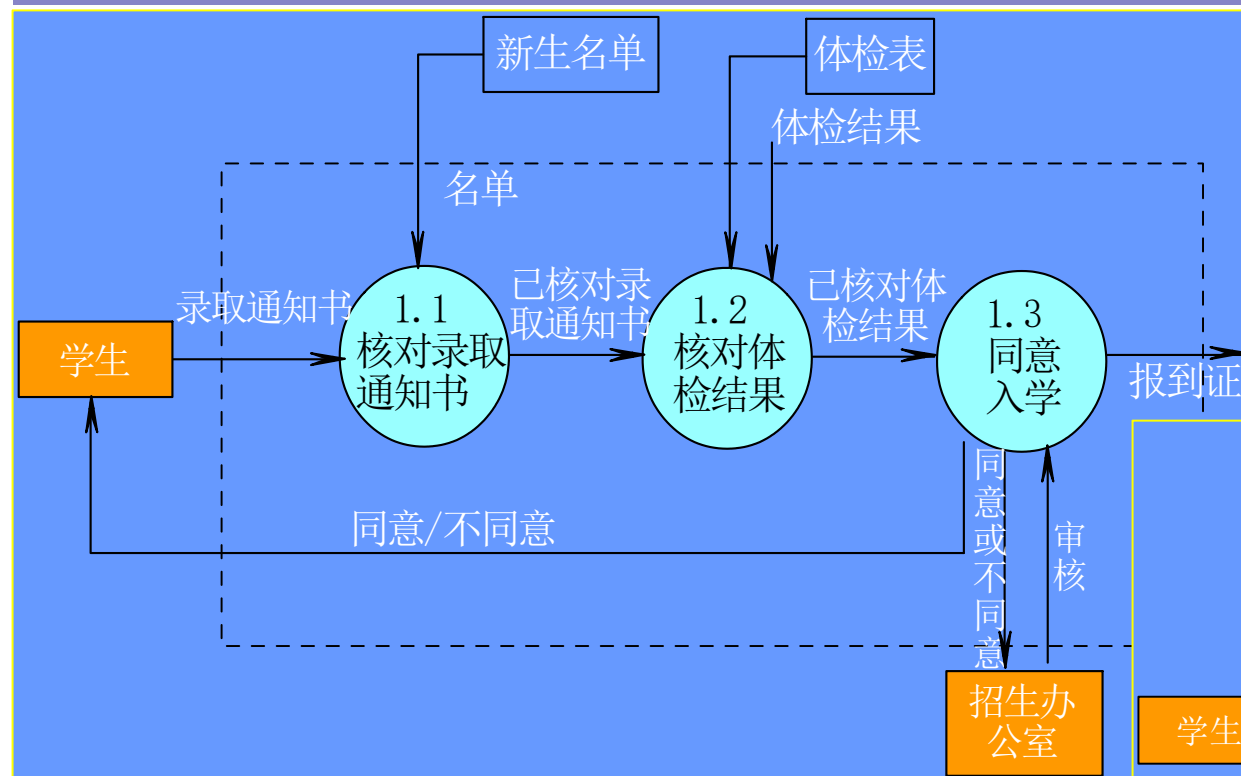


# 需求分析实例I

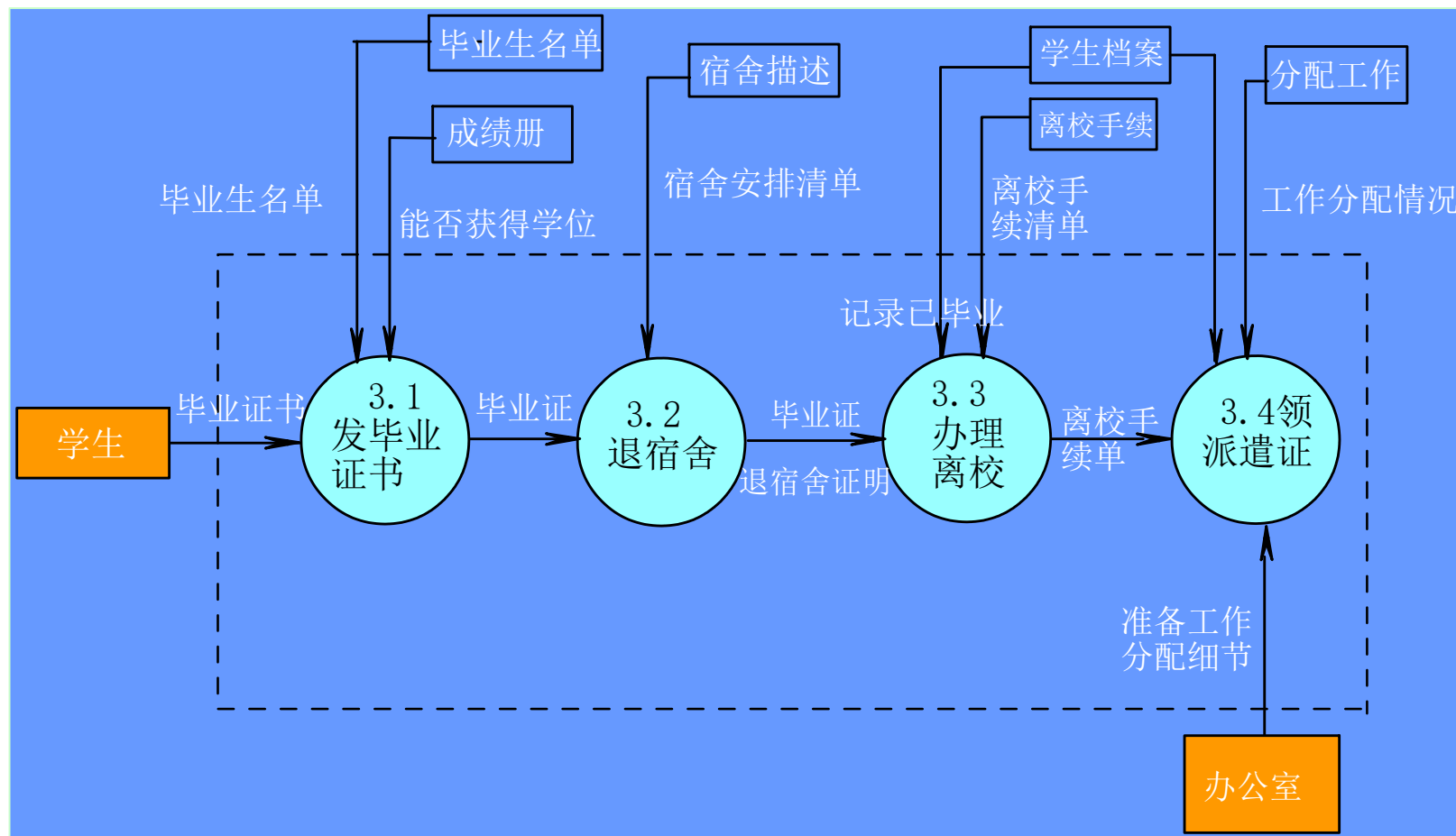
【例】开发某学校信息管理系统，经过可行性分析和初步调查，采用**自顶向下**的方法抽象出该系统高层数据流图，**逐步往下求精**得出各子系统。以学生学籍管理子系统为例向下求解。



# 需求分析实例I



# 需求分析实例I



# 数据字典

□ 数据字典是关于数据库中数据的描述，是元数据，而不是数据本身。

## (1) 数据项

- 数据项是数据的最小单位。
- 包括：{数据项名、含义说明、别名、类型、长度、取值范围、与其他数据项的关系等}。其中，取值范围、与其他数据项的关系：定义了完整性约束条件，是设计数据检验功能的依据。

例：数据项：以“学号”为例：

数据项：学号

含义说明：唯一标识每个学生

别名：学生编号

类型：字符型

长度：8

取值范围：00000000~99999999

取值含义：前两位标别该学生所在年级，后六位按顺序编号

与其他数据项的逻辑关系：……

# 数据字典

## (2) 数据结构

- 数据结构反映了数据之间的组合关系。
- 一个数据结构可以由若干个数据项组成，也可以由若干个数据结构组成，或由若干个数据项和数据结构混合组成。
- 内容包括：{数据结构名、含义说明，组成：{数据项或数据结构} }。

例：数据结构：以“学生”为例：

“学生”是该系统中的一个核心数据结构：

数据结构： 学生

含义说明： 是学籍管理子系统的主体数据结构，  
定义了一个学生的有关信息

组成：学号，姓名，性别，年龄，  
所在系，年级



# 数据字典举例

## (3) 数据流

- 数据流可以是数据项，也可以是数据结构，它表示某一处理过程中数据在系统内传输的路径。
- 内容包括：{数据流名、说明、流出过程、流入过程，组成：{数据项或数据结构}，平均流量，高峰值}。其中，流出过程说明该数据流由什么过程而来；流入过程说明该数据流到什么过程。

例：数据流：“体检结果”可如下描述：

数据流：体检结果

说明：学生参加体格检查的最终结果

数据结构名：体检表

数据流来源：体检

数据流去向：批准

组成：.....

平均流量：.....

高峰期流量：.....

# 数据字典举例

## (4) 处理过程

- 处理逻辑通常用判定表或判定树来描述。
- 处理过程包括 {处理过程名, 说明, 输入: {数据流}, 输出: {数据流}, 处理: {简要说明}}。

其中, 简要说明: 说明处理过程的功能及处理要求。

- 功能是指该处理过程用来做什么 (不是怎么做), 处理要求指该处理频度要求, 如单位时间里处理多少事务、多少数据量、响应时间要求等, 这些处理要求是后面物理设计的输入及性能评价的标准。

# 数据字典举例

例：处理过程：“分配宿舍”可如下描述：

- 处理过程：分配宿舍
- 说明：为所有新生分配学生宿舍
- 输入：学生，宿舍
- 输出：宿舍安排
- 处理：在新生报到后，为所有新生分配学生宿舍。
- 处理频率：对每个学生每学年处理一次。

□ 要求：

- 同一间宿舍只能安排同一性别的学生，
- 同一个学生只能安排在一个宿舍中。
- 每个学生的居住面积不小于3平方米。
- 安排新生宿舍其处理时间应不超过15分钟。

# 数据字典

## (5) 数据存储

- 处理过程中数据的存放场所，也是数据流的来源和去向之一。可以是手工凭证、手工文档或计算机文件。
- 包括：{数据存储名，说明，输入数据流，输出数据流，组成：{数据项或数据结构}，数据量，存取频度，存取方式。存取方法：批处理/联机处理；检索/更新}

例：数据存储，“学生登记表”可如下描述：

数据存储： 学生登记表  
说明： 记录学生的基本情况  
流入数据流： .....  
流出数据流： .....  
组成： .....  
数据量： 每年3000张  
存取方式： 随机存取

数据字典在需求分析阶段建立，在数据库设计过程中不断修改、充实、完善。

## 7.3 概念结构设计

### 7.3.1 概念模型

概念设计就是将需求分析得到的用户需求抽象为信息结构，即概念模型。




□ 描述工具：E-R图

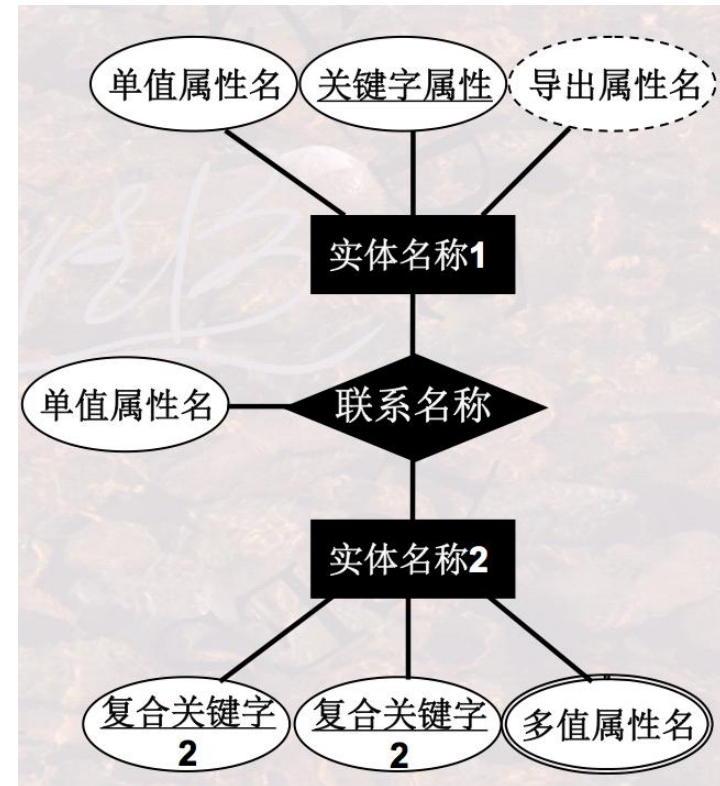
□ 特征：

- 独立于硬件；
- 独立于软件（OS、DBMS）。



## 7.3.2 E-R模型

- E-R数据模型提供了实体、属性和联系三个主要的抽象概念。
- E-R数据模式：用E-R数据模型对一个系统的模拟。
- E-R图：
  - 矩形：实体集 
  - 椭圆：属性 
  - 菱形：联系集 
  - 线段：将属性连接到实体集或将实体集联系到联系集，无向边上标明联系的类型（1: 1, 1: m, m: n）；



## 7.3.2 E-R模型

### □ 实体与属性

- 实体：客观存在并可相互区分的事物；一类实体具有相同或相似的特性。
- 属性：实体所具有的某一方面特性。

#### □ 单一属性和复合属性：

复合属性示例：家庭住址:省份, 详细住址

在关系模型中，复合属性一定要转化为单一属性(1NF)

#### □ 单值属性和多值属性：

多值属性示例：电话号码, 一个人可能有多个电话号

在关系模型中，多值属性一定要转化为单值属性(1NF)

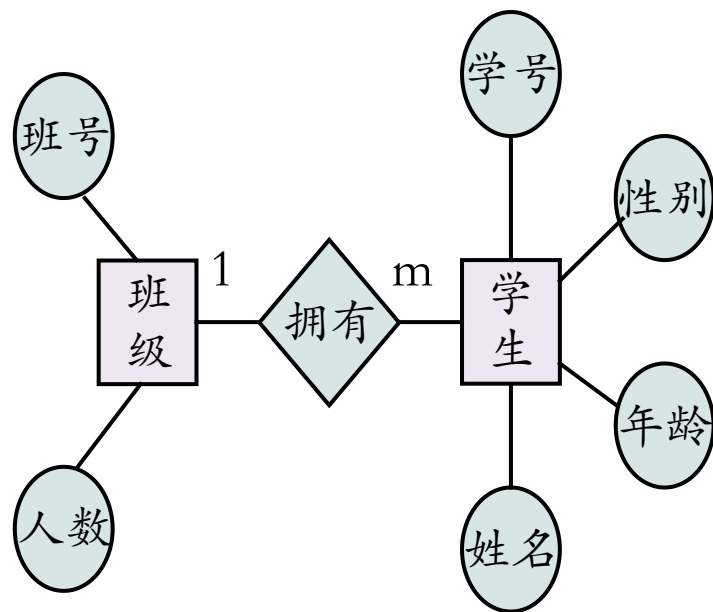
#### □ 可空值属性和非空值属性

#### □ 导出属性：由其他属性计算而得

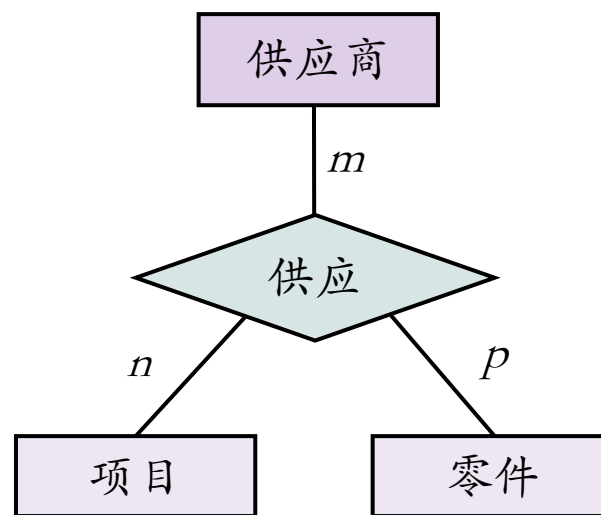
# E-R模型

## □ 联系的几种基本形式:

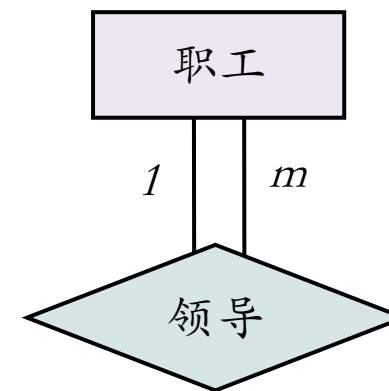
### ① 两个实体之间的联系



### ② 两个以上实体间的联系



### ③ 同一实体集内部各实体之间的联系。

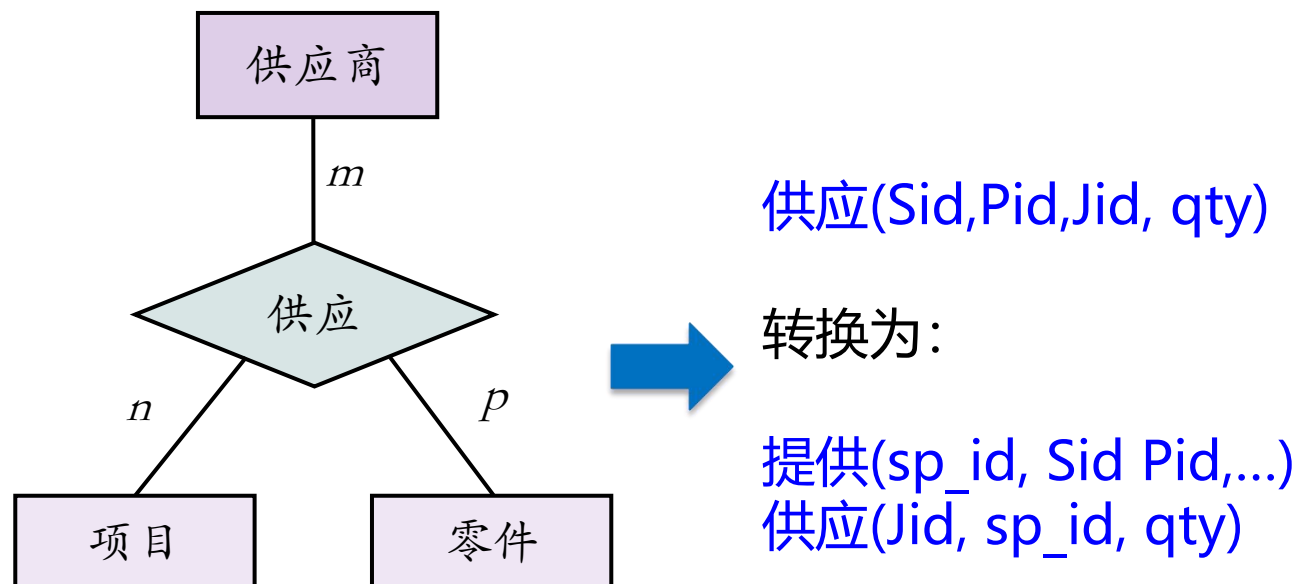


一个三元关系 = 2个二元关系?



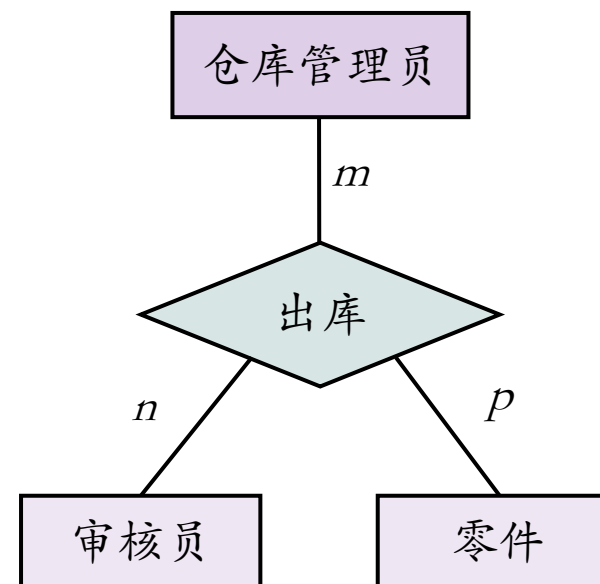
# E-R模型

## □ 三元关系的讨论：



一个三元关系 = 2个二元关系？

可以转换，而且解决了插入异常



本例没有插入异常，  
也可转换为2个二元关系：

- 出库单(id, uid, pid,...)
- 出库单审核(sid, id, time, ...)

# 概念模型设计的方法和步骤

## 1. 概念结构设计的方法

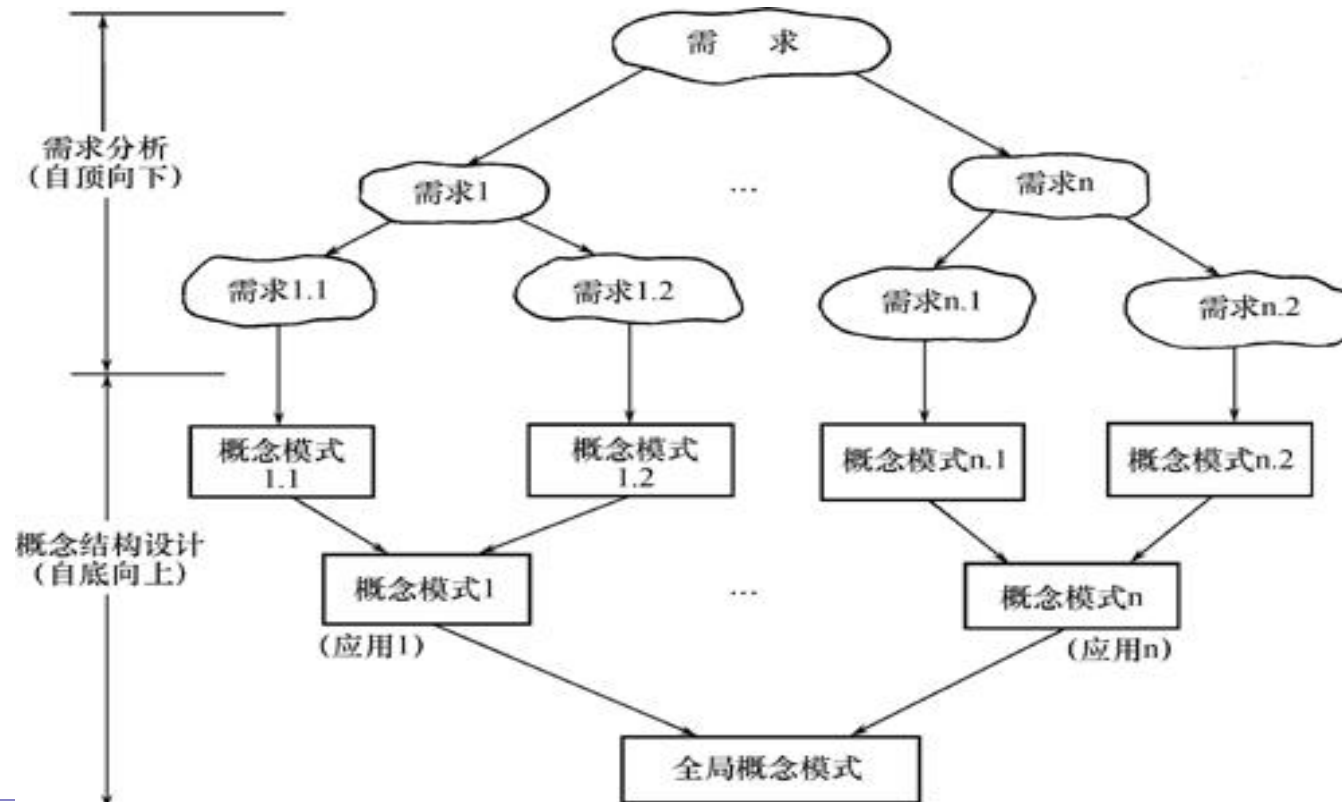
设计概念结构的E-R模型可采用四种方法。

- (1) **自顶向下** 先定义全局框架，再逐步细化。
- (2) **自底向上** 先定义局部E-R模型，然后将它们集成，得到全局E-R模型。
- (3) **逐步扩张** 先定义核心概念E-R模型，然后以滚雪球的方式逐步生成其他概念结构E-R模型。
- (4) **混合策略** 自顶向下和自底向上相结合的方法。

# 概念模型设计的方法和步骤

## □ 常用策略:

- 自顶向下地进行需求分析
- 自底向上地设计概念结构

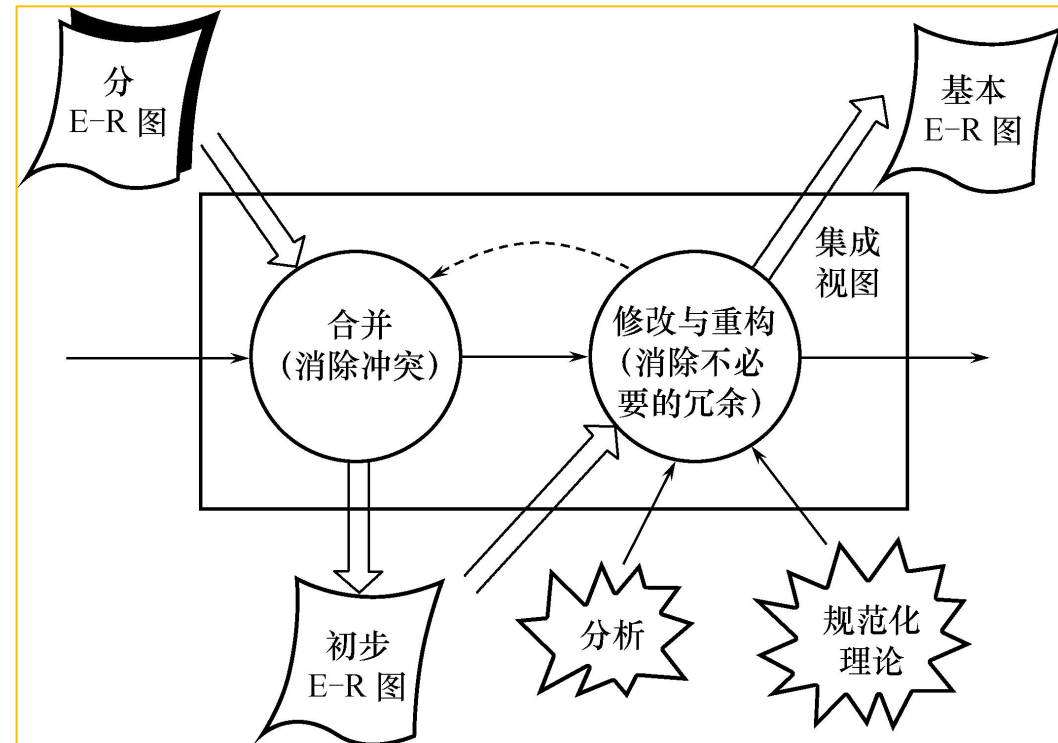
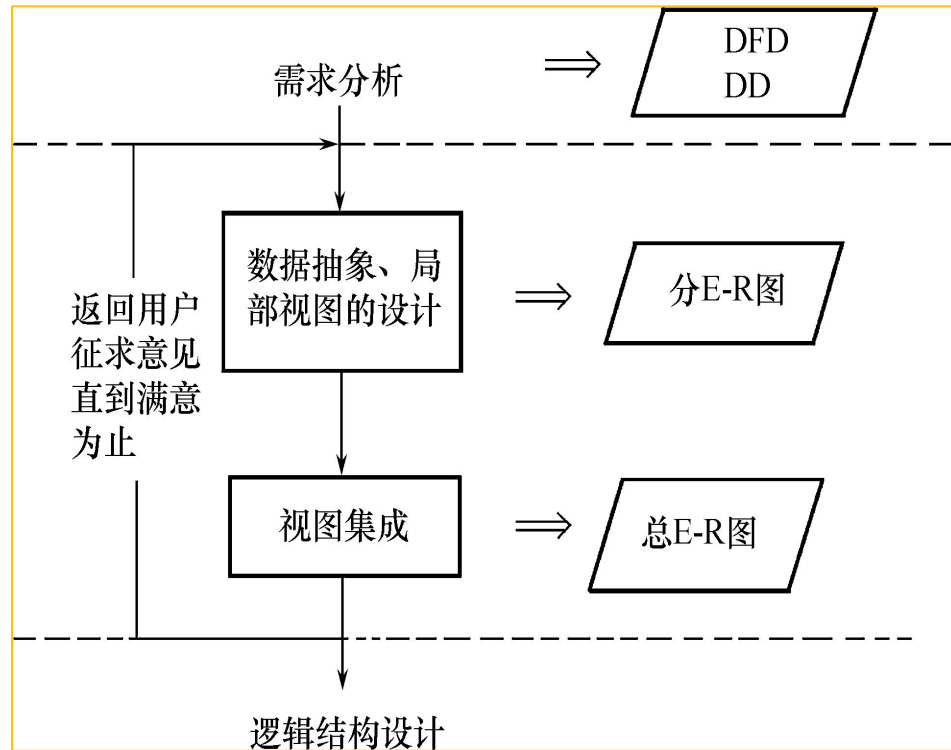


# 概念模型设计的方法和步骤

## □ 自底向上设计概念结构的步骤

第1步：抽象数据并设计局部视图

第2步：集成局部视图，得到全局概念结构



# 步骤1:数据抽象和局部设计

□ 概念设计的第一步：利用抽象机制对需求分析阶段收集的数据进行分类、组织形成实体、实体的属性，标识实体的码，确定实体之间的联系类型，设计分E-R图。

□ 设计分E-R图：

选择局部应用，确定范围原则：

- 1) 相对独立；
- 2) 内部联系较紧密；
- 3) 与外部联系相对较少。

逐一设计分E-R图：利用抽象机制形成实体、属性，确定联系等。

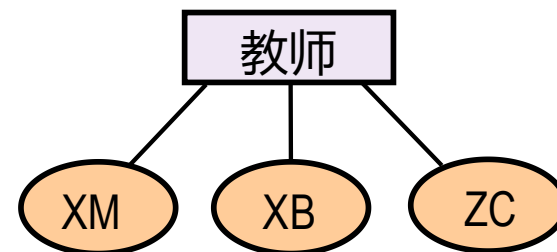
# 划分实体和属性的准则

## □ 划分遵循两条原则：

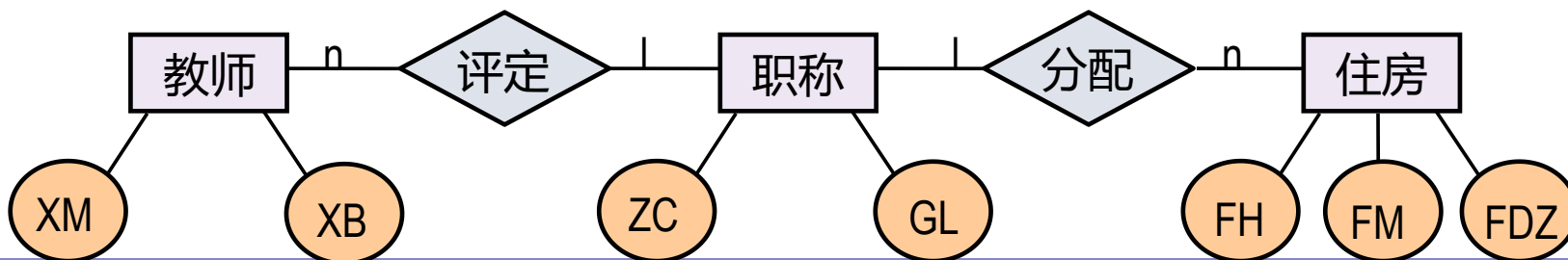
- ① 实体具有描述信息，而属性没有。属性必须是不可分的，不能包含其他属性。可再分者为实体。
- ② 联系只能发生在实体之间。

一般情况下，凡能作为属性对待的，应尽量作为属性，以简化E-R图的处理。

【例】一般“ZC（职称）”做属性：



若职称与住房有联系则作实体：



## 步骤2：集成局部设计，生成全局概念结构

### □ 集成的方式：

- ① **多元集成法**：一次性将多个局部E-R图合并为一个全局E-R图（复杂）。
- ② **二元集成法**：首先集成两个重要的局部视图，以后用累加的方法逐步将一个新的视图集成进来。

### □ 无论使用哪一种方法，集成均分成两个步骤：

- ① **合并**：消除局部E-R图之间的冲突，生成初步E-R图。
- ② **优化**：消除不必要的冗余，生成基本E-R图。

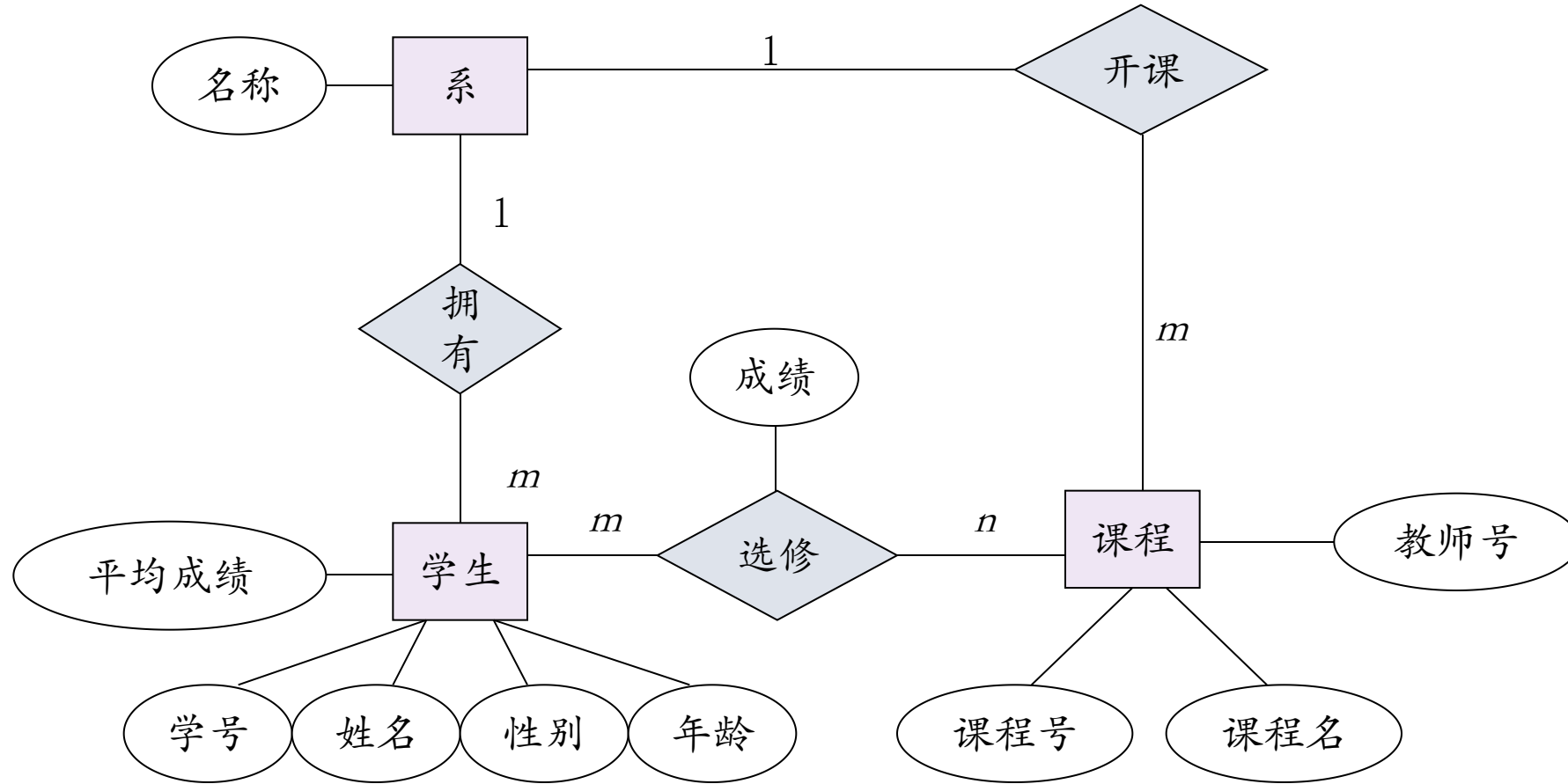
# 局部E-R模型设计

- 举例：局部E-R模型设计。
- 在一个简单的教务管理系统中，有如下语义约束：
  - ①一个学生可选修多门课程，一门课程可为多个学生选修，因此学生和课程是多对多的联系；
  - ②一个教师可讲授多门课程，一门课程可为多个教师讲授，因此教师 and 课程也是多对多的联系；
  - ③一个系可有多个教师，一个教师只能属于一个系，因此系和教师是一对多的联系，同样系和学生也是一对多的联系。



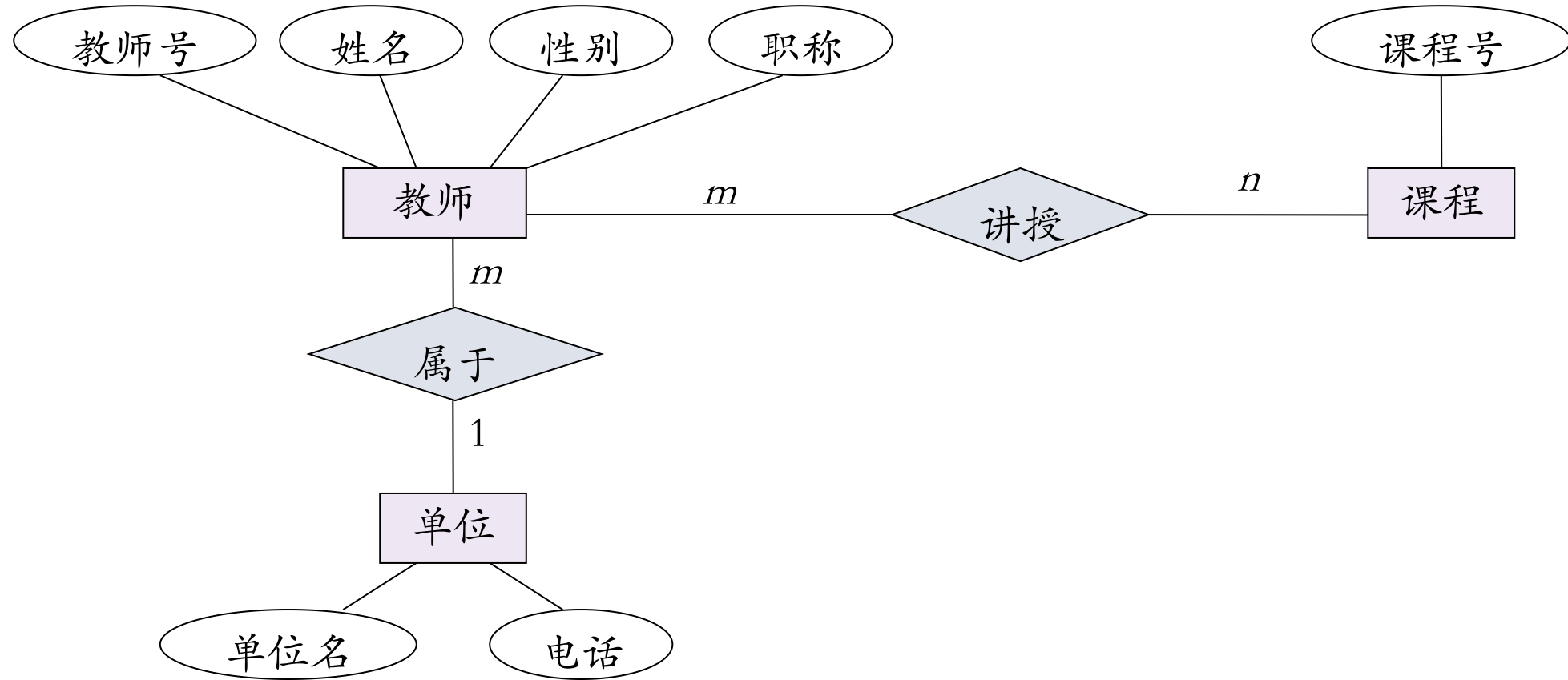
# 学生选课局部E-R图

## □ 学生选课:



# 教师任课局部E-R图

## □ 教师任课:



# 全局ER模型设计

□ E-R图中的冲突有三种：属性冲突、命名冲突和结构冲突。

## ① 属性冲突

■ **属性值域冲突**——即属性值的类型、取值范围或取值集合不同。

如：学号，数值型 或 字符型。

年龄，出生年月 或 整数。

■ **属性的取值单位冲突**

如：零件的重量，以：公斤为单位 或 斤为单位 或 克为单位。

属性冲突属于用户业务上的约定，必须与用户协商后解决。

# 全局ER模型设计

## ② 命名冲突

命名不一致可能发生在**实体名、属性名或联系名之间**，其中属性的命名冲突更为常见。

- **同名异义**——即同一名字的对象在不同的部门中具有不同的意义。

例如：“单位”在某些部门表示为人员所在的部门，而在某些部门可能表示物品的重量、长度等属性。

- **异名同义**——即同一意义的对象在不同的部门中具有不同的名称。

例如：对于“科研项目”，在财务部门被称为“项目”，而在科研处被称为“课题”。

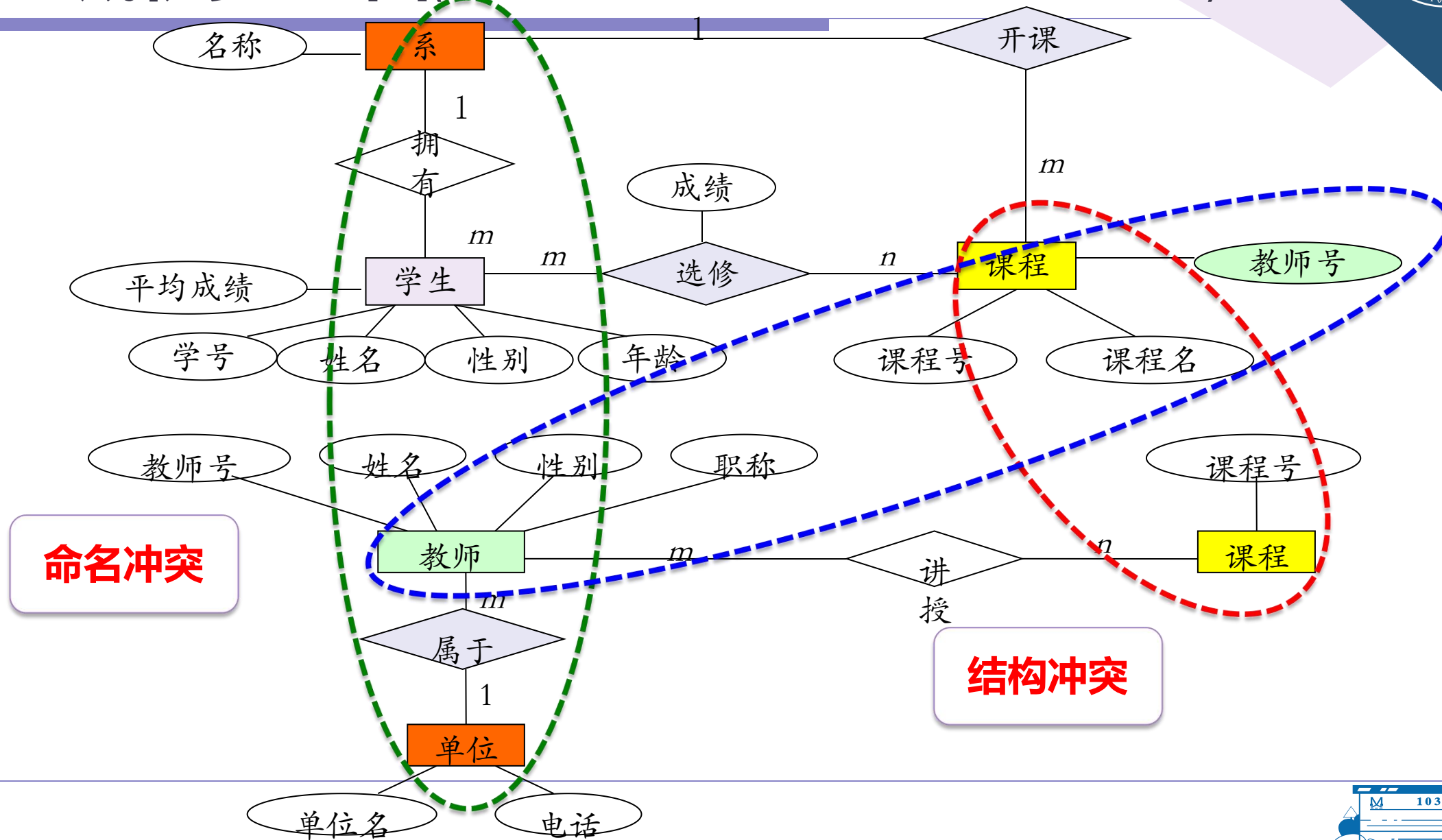
命名冲突的解决方法同属性冲突，需要与各部门协商、讨论后加以解决。

# 全局ER模型设计

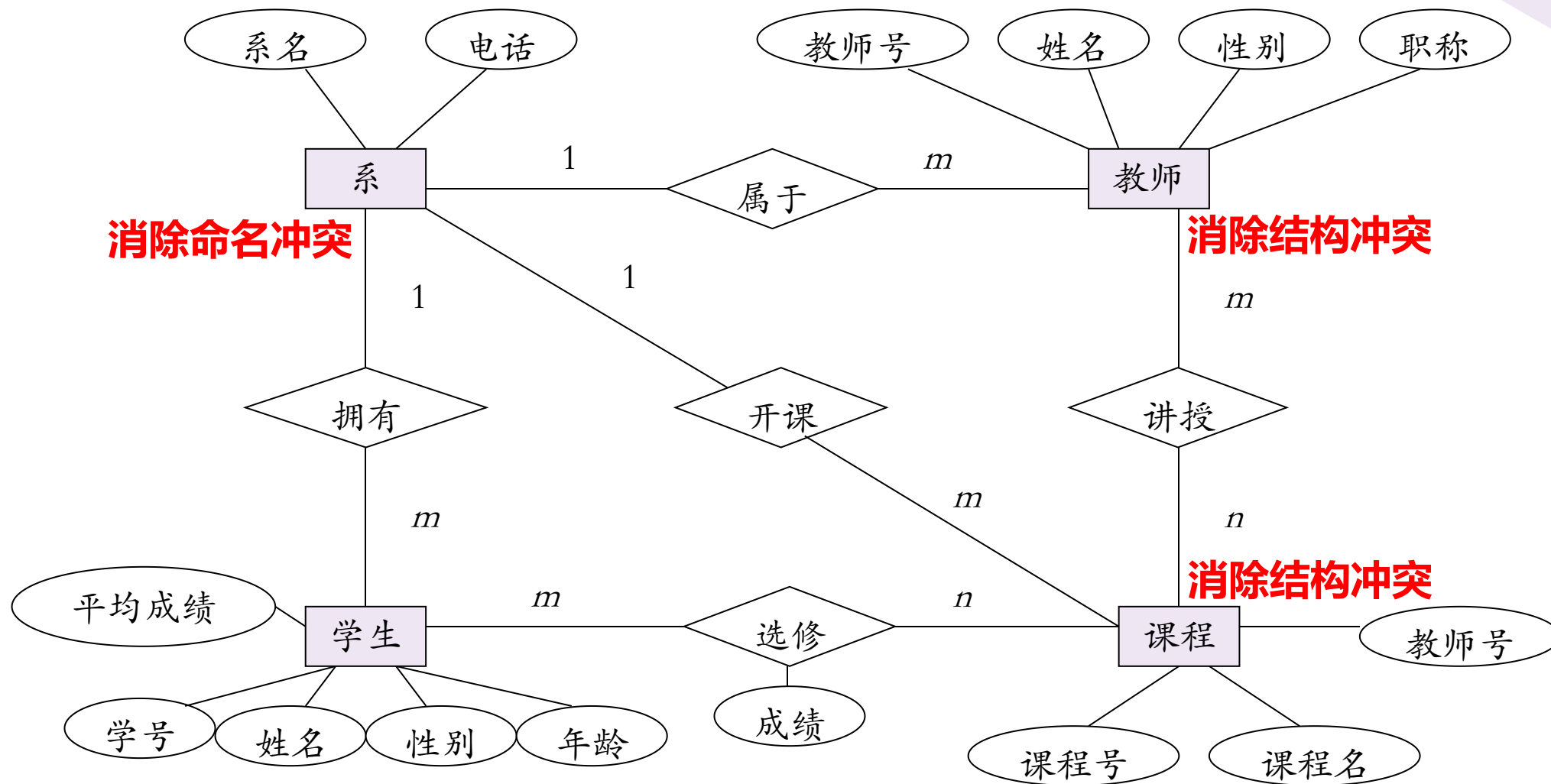
## ③ 结构冲突

- 同一对象在不同应用中有不同的抽象，可能为实体，也可能为属性。如，教师的职称在某一局部应用中被当作实体，而在另一局部应用中被当作属性。
- 解决方法：统一抽象级别（要符合7.3.3中所介绍的准则）
- 同一实体在不同应用中属性组成不同，可能是属性个数或属性次序不同。  
解决办法：取各局部E-R图中的同名实体属性的并集，然后再适当调整属性的次序。
- 同一联系在不同应用中呈现不同的类型。如E1与E2在某一应用中是一对一联系，而在另一应用中是一对多或多对多联系，或在E1、E2、E3三者之间有联系。  
解决办法：根据应用语义对实体联系的类型进行综合调整。

# 1. 生成初步E-R图(以教务管理系统中的两个局部E-R图为例)



# 初步E-R图



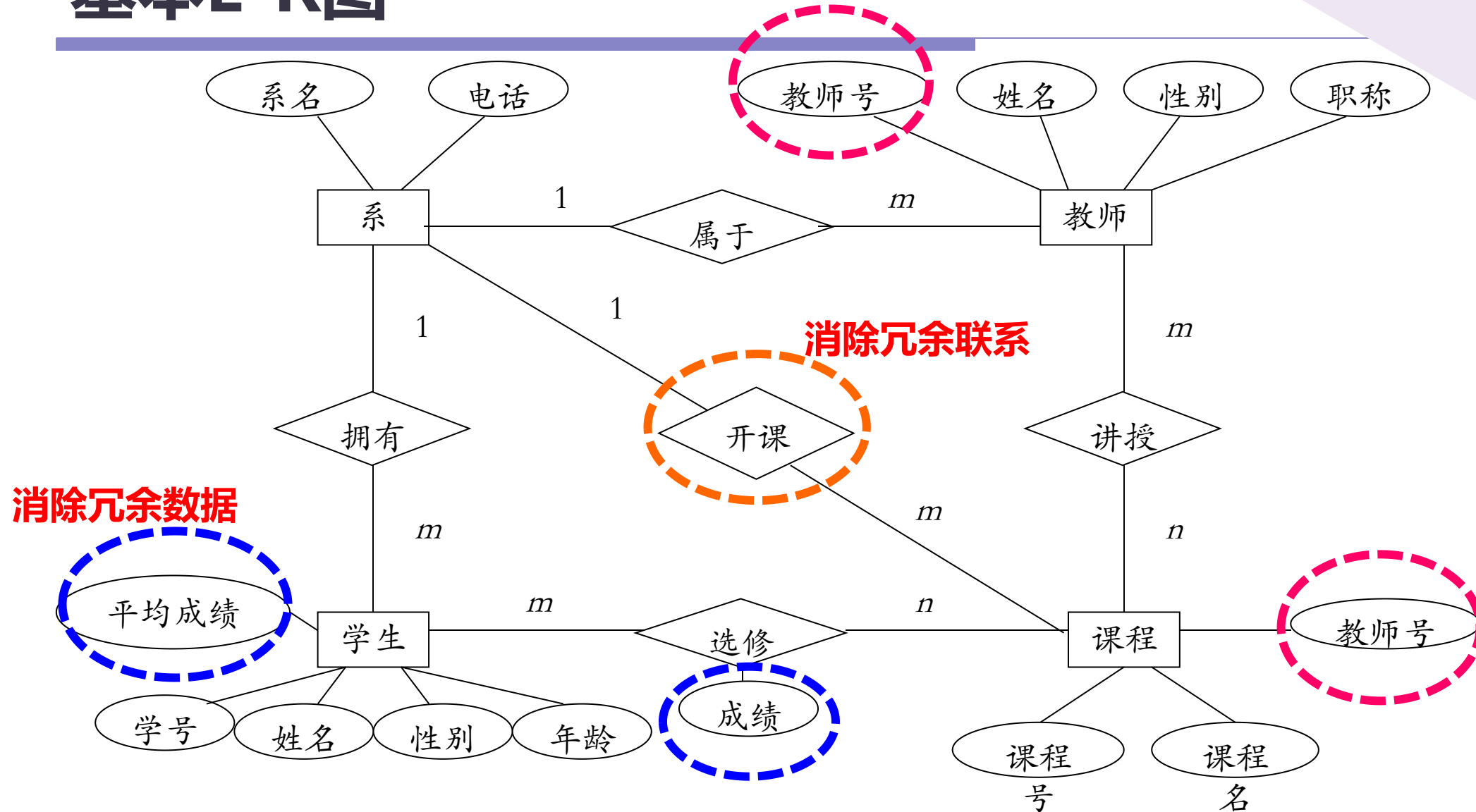
# 全局E-R图设计

## (2) 消除不必要的冗余，生成基本E-R图

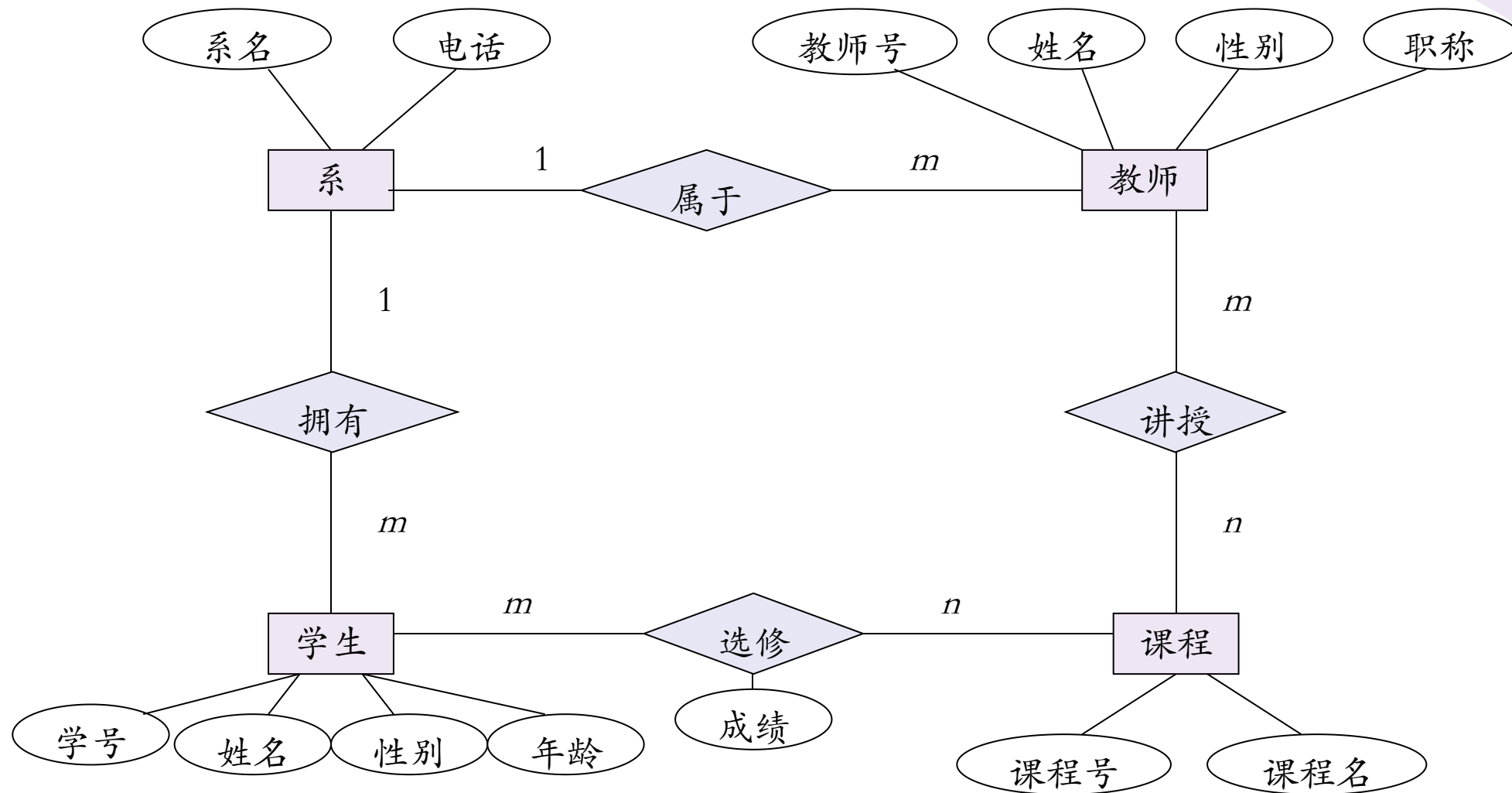
- 冗余包括冗余的数据和冗余的联系。
  - 冗余的数据是指可由基本的数据导出的数据。
  - 冗余的联系是由其他的联系导出的联系。
- 消除了冗余的初步E-R图称为基本E-R图。
- 通常采用分析的方法消除冗余。数据字典是分析冗余数据的依据，还可以通过数据流图分析出冗余的联系。



# 基本E-R图



# 最终全局E-R图



# 课堂练习：图书馆管理系统

请给出如下场景（满足1~4）的图书管理数据库的ER图设计：

- 用户类型：读者、管理员
- 管理员负责图书的管理和读者的管理；
- 一个读者可以借多本书；一本书可以借给多个读者（馆藏多本）；
- 一本书最迟1个月要归还；如果逾期还书，需按逾期天数罚款。

题目升级（如果要满足5~8，ER图如何设计？）：

- 读者要有所属单位信息；
- 图书管分为若干室，书按图书类型分别放在不同室的不同书架上；
- 若第4点改为：读者可分为：学生、老师；学生最多可借有图书5本，教师最多可以借阅图书10本；
- 借期30天，可续借1次，续借15天；逾期每册书每天罚款0.2元。

