

北京林业大学

# 数据库原理与应用

# 数据库设计概述

# 本章目录 CONTENTS

- | 数据库设计的任务
- | 数据库设计的内容
- | 数据库设计的特点
- | 数据库设计方法简述
- | 数据库设计的步骤





# 数据库设计的任务

**数据库设计是指根据用户需求研制数据库结构和行为的过程。**

对于一个给定的应用环境，构造最优的数据库模式，建立数据库及其应用系统；

有效地存储数据，满足用户的信息要求和处理要求。



# 数据库设计的内容

**结构设计** → **概念设计、逻辑设计和物理设计**

**行为设计** → **用户对数据库的操作**



# 数据库设计的特点

## 数据库设计的特点

结构源于行为——强调结构设计与行为设计相结合

行为总是变化——“反复探寻，逐步求精”





# 数据库设计方法简述

## 直观设计法

直观设计法也称为手工试凑法，它是最早使用的数据库设计方法。

这种方法依赖于设计者的经验和技巧，缺乏科学理论和工程原则的支持，设计的质量很难保证。

常常是数据库运行一段时间后又发现各种问题，这样再重新进行修改，增加了系统维护的代价。



# 数据库设计方法简述

## 常用的规范设计法：

### 基于E-R模型的数据库设计方法

用E-R图构造一个反映现实世界实体之间联系的概念模式

### 基于3NF的数据库设计方法

确定数据库中的全部属性和属性间的依赖关系，分析不符合3NF的约束条件，将其进行投影分解，规范成若干个3NF关系模式的集合。

### 基于视图的数据库设计方法

先从分析各个应用的数据着手，并为每个应用建立自己的视图，然后再把这些视图汇总起来合并成整个数据库的概念模式。





# 数据库设计方法简述

## 计算机辅助设计法

计算机辅助设计法是指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。



# 数据库设计方法简述

## 现代数据库设计方法

围绕软件工程的思想，通常以E-R图设计为主体，辅以3NF设计和视图设计实现模式的评价和优化，吸收各种设计方法的优势。

为提高设计的协同效率和规范化程度，现代数据库设计过程还会通过计算机辅助设计工具获得规范的数据库设计结果。



# 数据库设计的步骤



## 数据库设计分为六个阶段

1. 系统需求分析——收集信息内容和处理要求，进行分析
2. 概念结构设计——表达用户需求的概念模型
3. 逻辑结构设计——由概念模型得出的数据模型
4. 物理结构设计——存储结构和存取方法
5. 数据库实施——数据入库，编写数据库存取程序
6. 数据库运行与维护——收集和记录实际系统运行的数据



北京林业大学

# 数据库原理与应用

# 系统需求分析

系统需求分析是数据库设计的起点，为以后的具体设计做准备。

需求分析的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。

系统需求分析的不正确或误解，直到系统测试阶段才发现许多错误，纠正起来要付出很大代价。



# 本章目录 CONTENTS

- | 需求分析的任务
- | 需求分析的方法
- | 案例的需求分析



## 需求分析的任务



调查分析用户活动，明确用户的需求目标



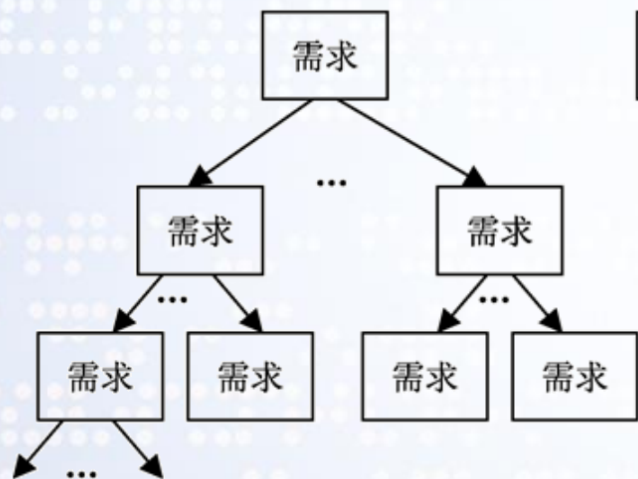
收集和分析需求数据，确定系统边界



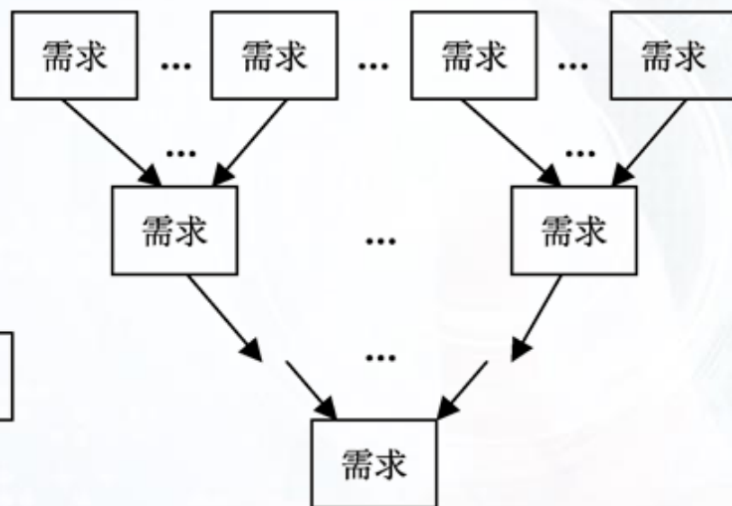
编写需求分析报告，组织专家评审



# 需求分析的方法



(a) 自顶向下的需求分析



(b) 自底向上的需求分析





# 需求分析的方法

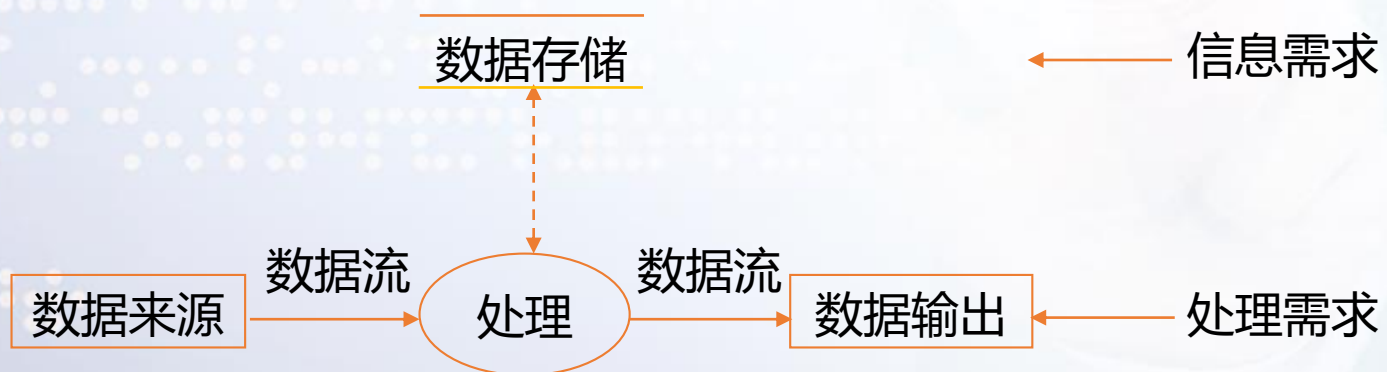
## 数据流图

用命名的箭头表示数据流

用圆圈表示处理

用不封闭的矩形或其他形状表示存储

用封闭的矩形表示来源和输出





# 需求分析的方法

## 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。在需求分析阶段，它通常包含以下五部分内容：**

数据项

数据结构

数据流

数据存储

处理过程

最终形成的数据流图和数据字典为系统分析报告的主要内容，这是下一步进行概念结构设计的基础。



# 案例的需求分析

## 案例描述

案例将围绕本科教学环节中的核心业务——**任课和选课**。



任课业务仅关心教师教授课程情况。



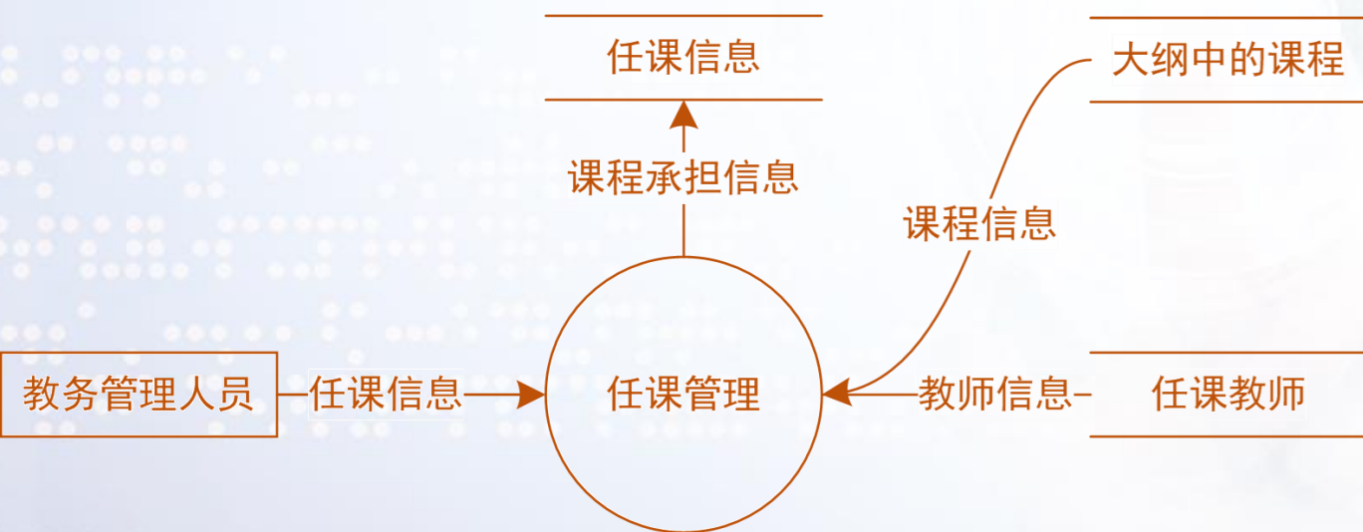
选课业务主要记录哪些学生选择了哪些课程，以及这门课程的得分。





# 案例的需求分析

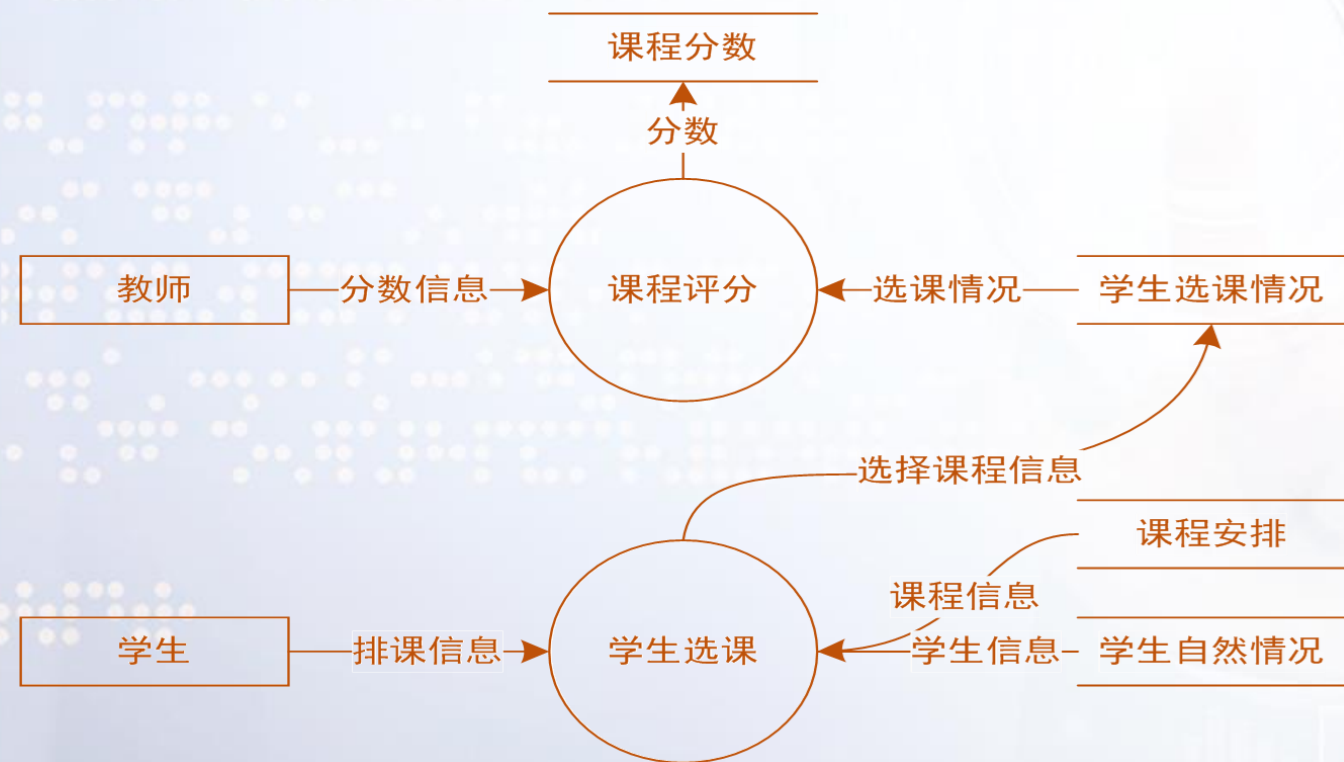
## 案例的数据流图——任课业务





# 案例的需求分析

## 案例的数据流图——选课业务





# 案例的需求分析

## 案例的数据字典

**学生自然情况信息：**学生的学号、姓名、年龄和院系等。

**课程信息：**课程的编号、名字和授课教师等。

**教师信息：**教师的编号、名字、教师的性别、职称和教授课程等。

**任课信息：**课程名和授课教师名等。

**学生选课信息：**学生名、课程名和教师名等。

**课程分数信息：**学生名、课程名和分数等。





# 案例的需求分析

## 案例的数据字典

除上述信息外，还需进一步分析该系统中是否还有隐含的数据结构。

”实际调研结果表明高校的管理通常以系别为单位，如果不划分系别，则各院系学生和教师的信息将混杂在一起，不便于开展各项业务。

因此，还需要对系的数据项进行定义。

**系别：**系的编号、名称、系的老师和系的学生。

北京林业大学

# 数据库原理与应用

# 概念结构设计





# 概念结构设计的必要性

概念结构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。



从逻辑设计中分离出概念设计以后，各阶段的任务相对单一化，设计复杂程度大大降低，便于组织管理。



概念模型不受特定的DBMS的限制，也独立于存储安排和效率方面的考虑，因而比逻辑模型更为稳定。



概念模型不含具体的DBMS所附加的技术细节，更容易为用户所理解，因而更有可能准确反映用户的信息需求。



# 概念模型的特点

## 概念模型特点

语义表达能力丰富

易于交流和理解

易于修改和扩充

易于向数据模型转换

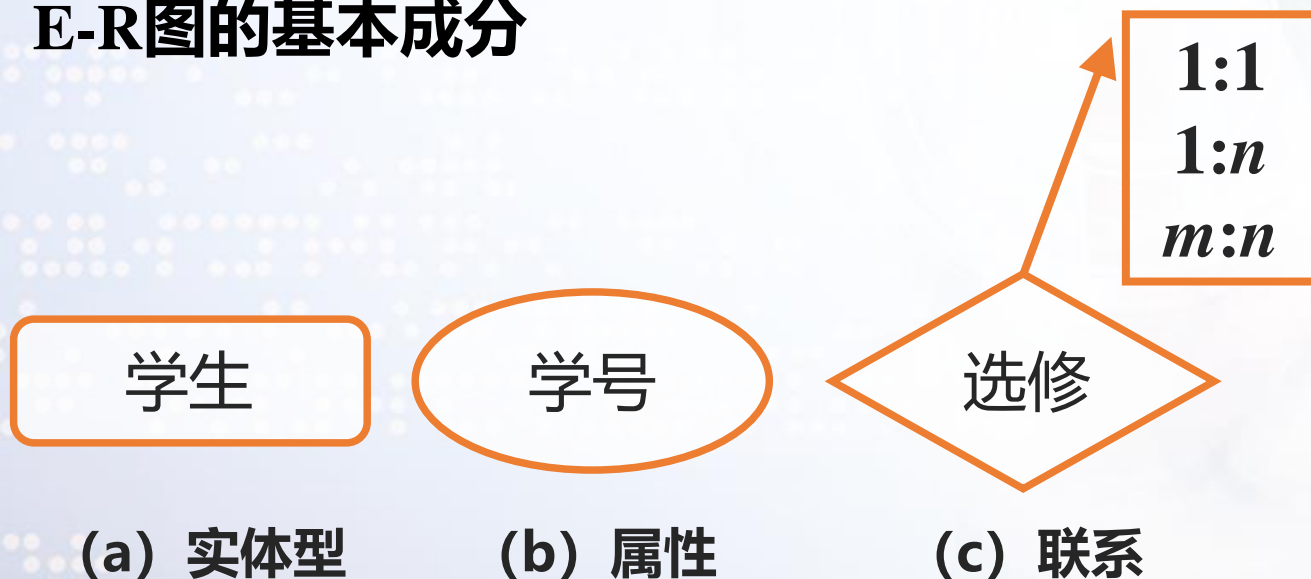


# 概念模型的E-R表示方法



E-R模型是最著名、最实用的一种是概念模型。

## E-R图的基本成分

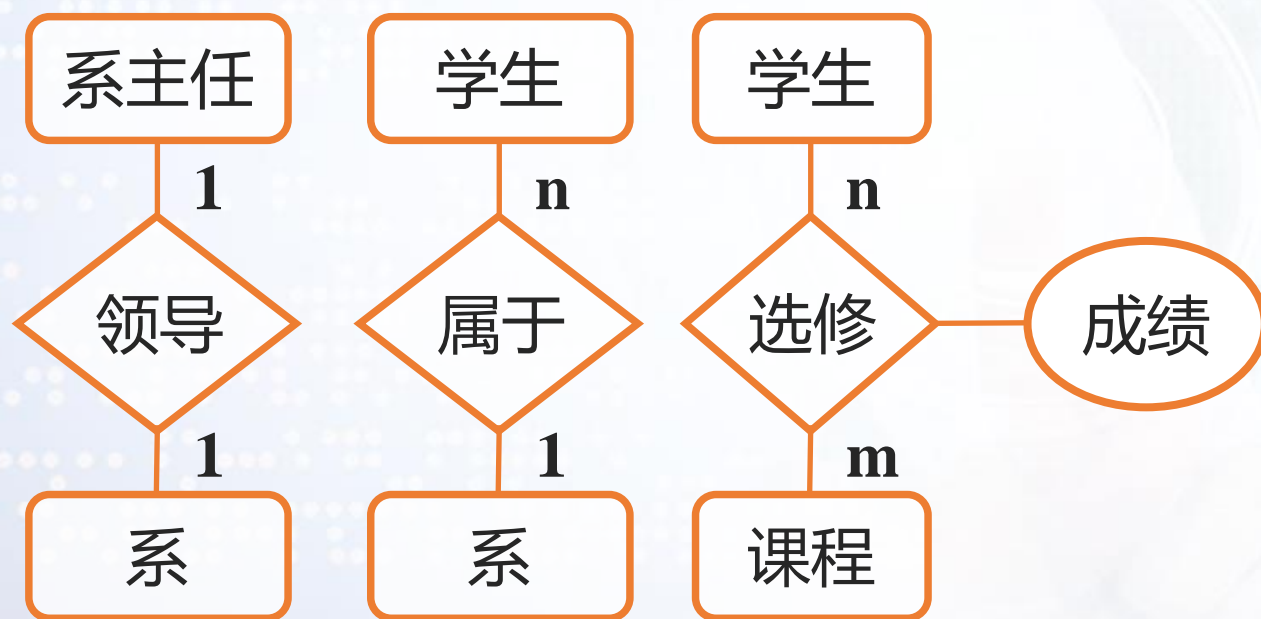






# 概念模型的E-R表示方法

## 实体及其联系图的形式

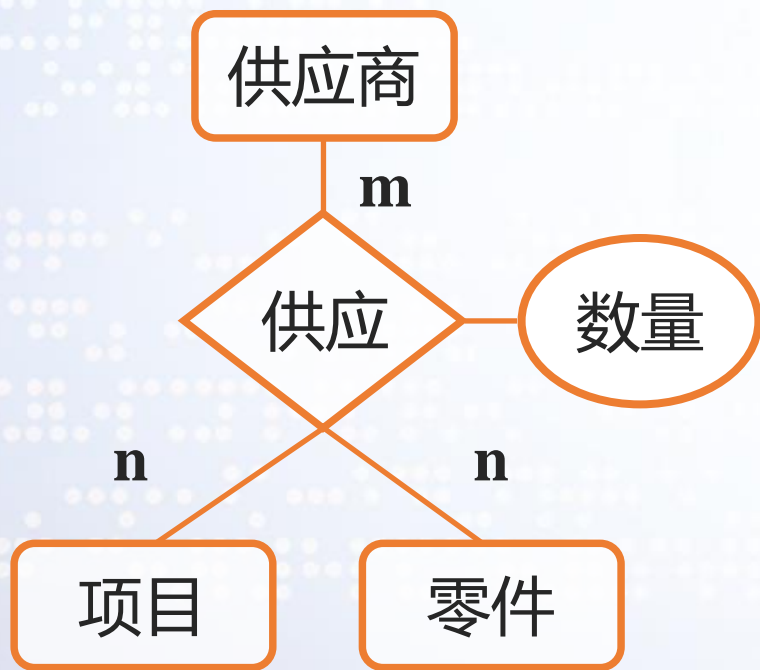


(a) 两个实体型之间的联系

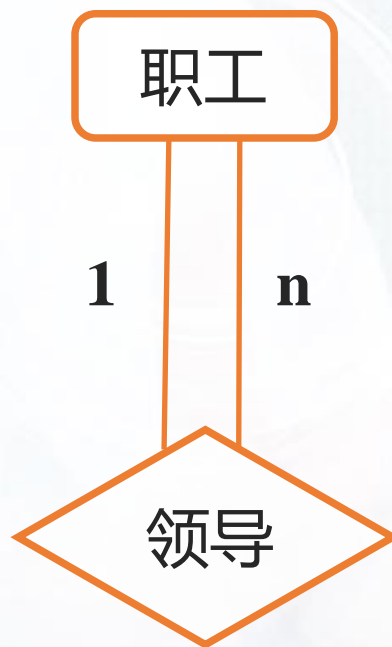


# 概念模型的E-R表示方法

## 实体及其联系图的形式



(b) 多个实体型之间的联系

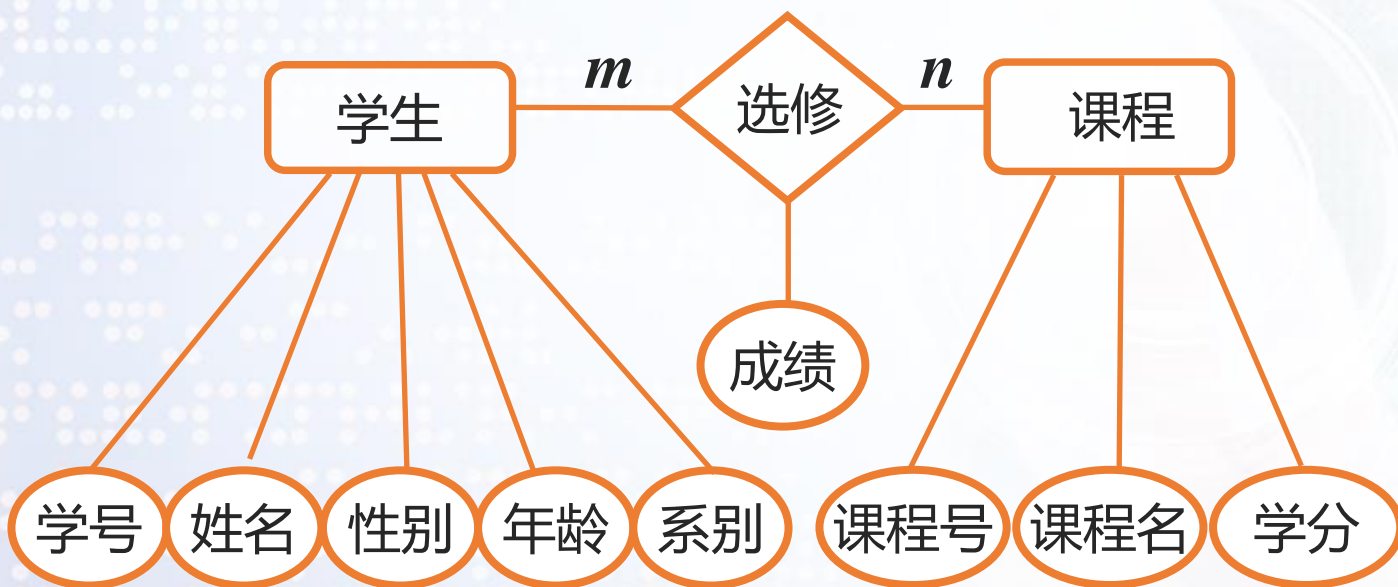


(c) 同一实体集内部的联系



# 概念模型的E-R表示方法

## 举例说明



学生与课程联系的完整的E-R图





# 概念结构设计的方法与步骤

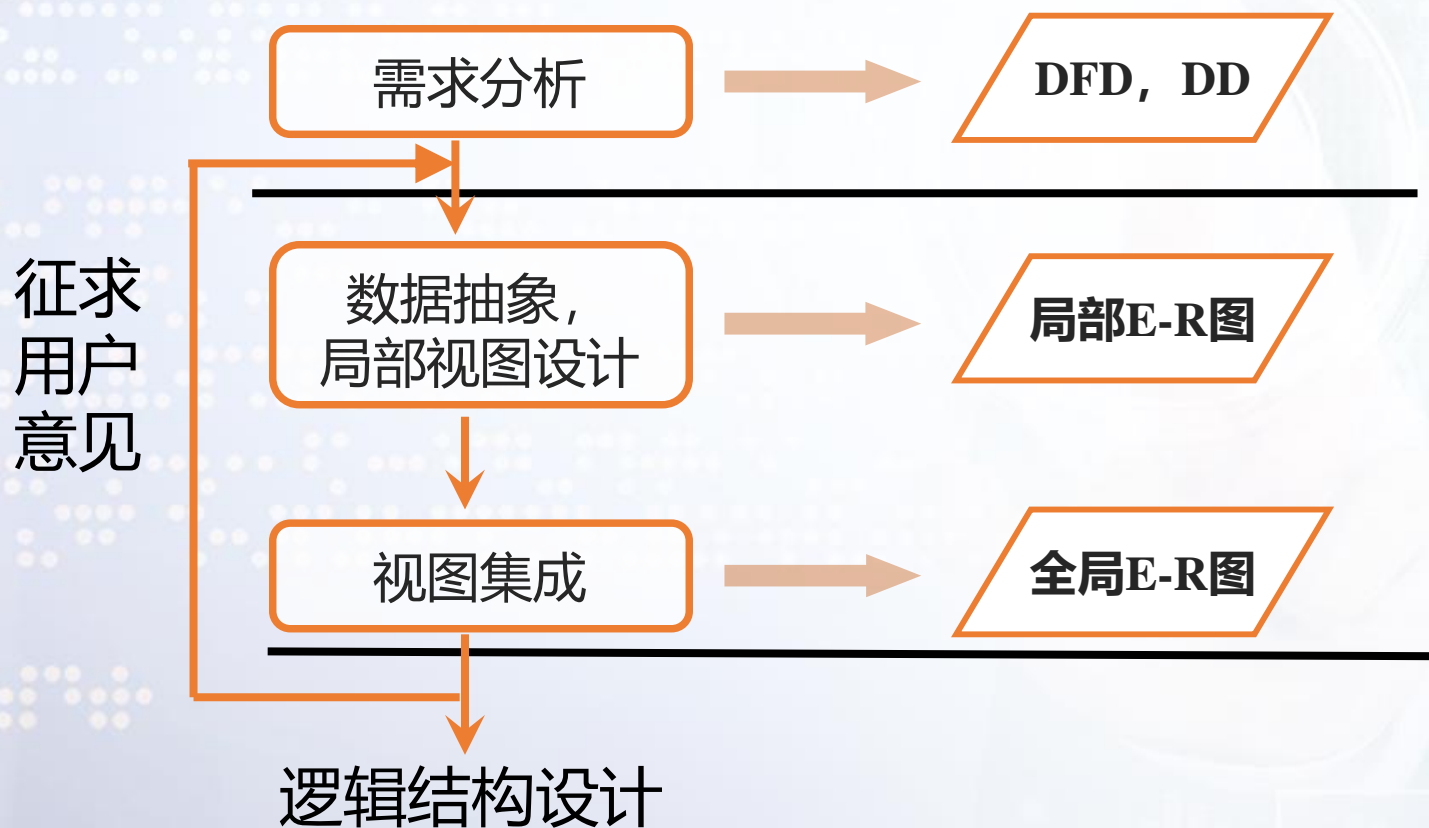
## 概念结构设计的方法

- ◆ 自顶向下 —— 逐步求细
- ◆ 自底向上 —— 由精细到整体
- ◆ 逐步扩张 —— 由核心逐渐向外扩充
- ◆ 混合策略 —— 自顶向下+自底向上



# 概念结构设计的方法与步骤

## 概念结构设计的步骤





# 概念结构设计的方法与步骤

## ◆ 数据抽象

分类



将一组具有共同特性和行为的对象抽象为一个实体

聚集



将对象类型的组成成分抽象为实体的属性





# 概念结构设计的方法与步骤

## ◆ 局部E-R模型设计

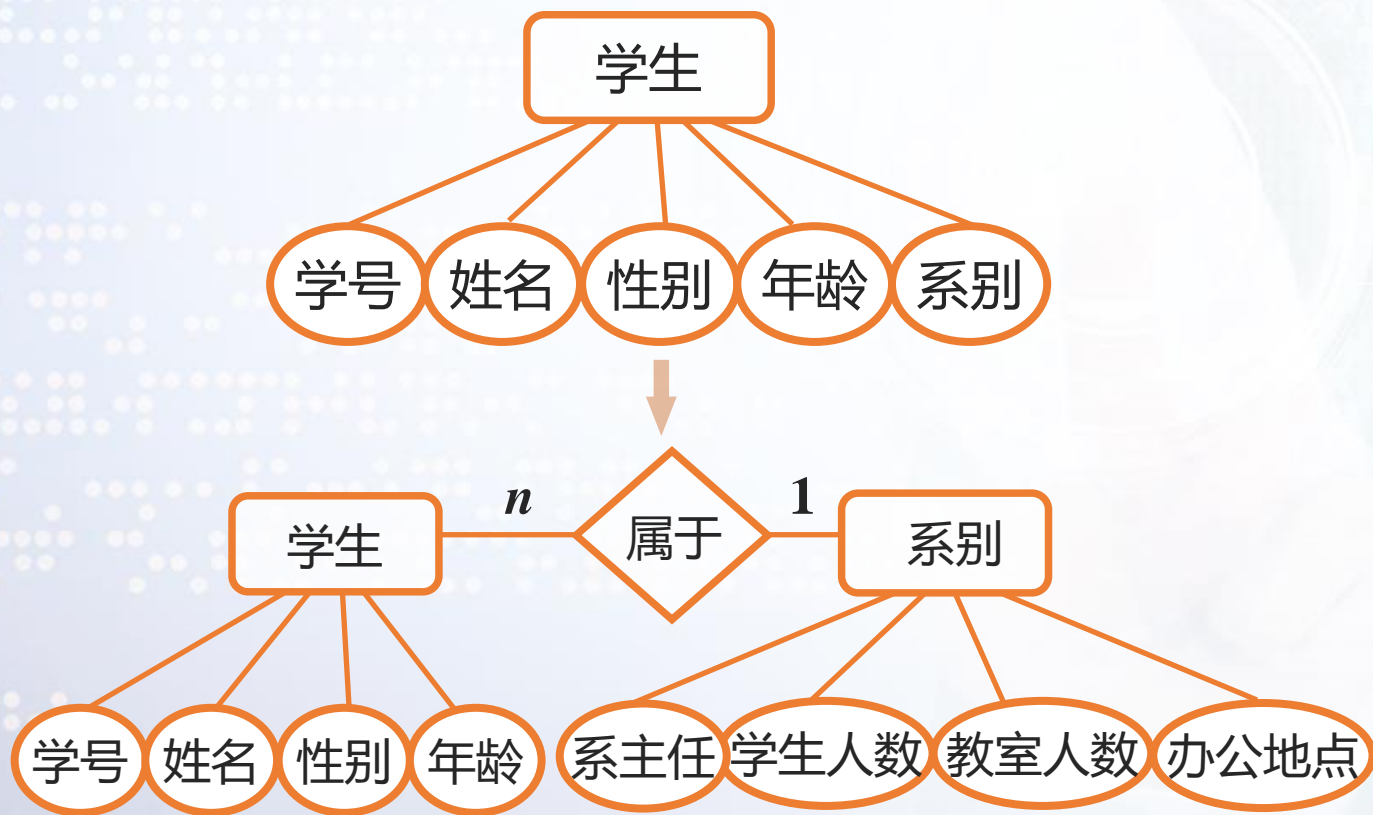
### 局部E-R模型设计

- ① 属性必须是不可分的数据项，不能再由另一些属性组成。
- ② 属性不能与其他实体具有联系，联系只能发生在实体之间。



# 概念结构设计的方法与步骤

## 局部E-R模型设计



系别作为一个属性或实体



# 概念结构设计的方法与步骤

## 局部E-R模型设计





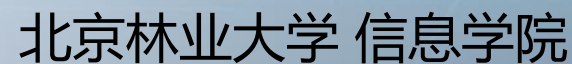


# 概念结构设计的方法与步骤

## 全局E-R模型设计

### 视图集成的方法:

- ◆ 多元集成法，一次性将多个局部E-R图合并。
- ◆ 二元集成法，首先集成两个重要的局部E-R图，以后用累加的方法逐步集成新的E-R图。





# 概念结构设计的方法与步骤

**合并** 合并局部E-R图，消除局部E-R图之间的冲突，生成初步E-R图







# 概念结构设计的方法与步骤

**优化** 消除不必要的冗余，生成基本E-R图。



- ◆ 冗余指冗余的数据和实体之间冗余的联系。
- ◆ 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。



# 案例的概念结构设计

## 案例的局部E-R模型设计

0

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。

0

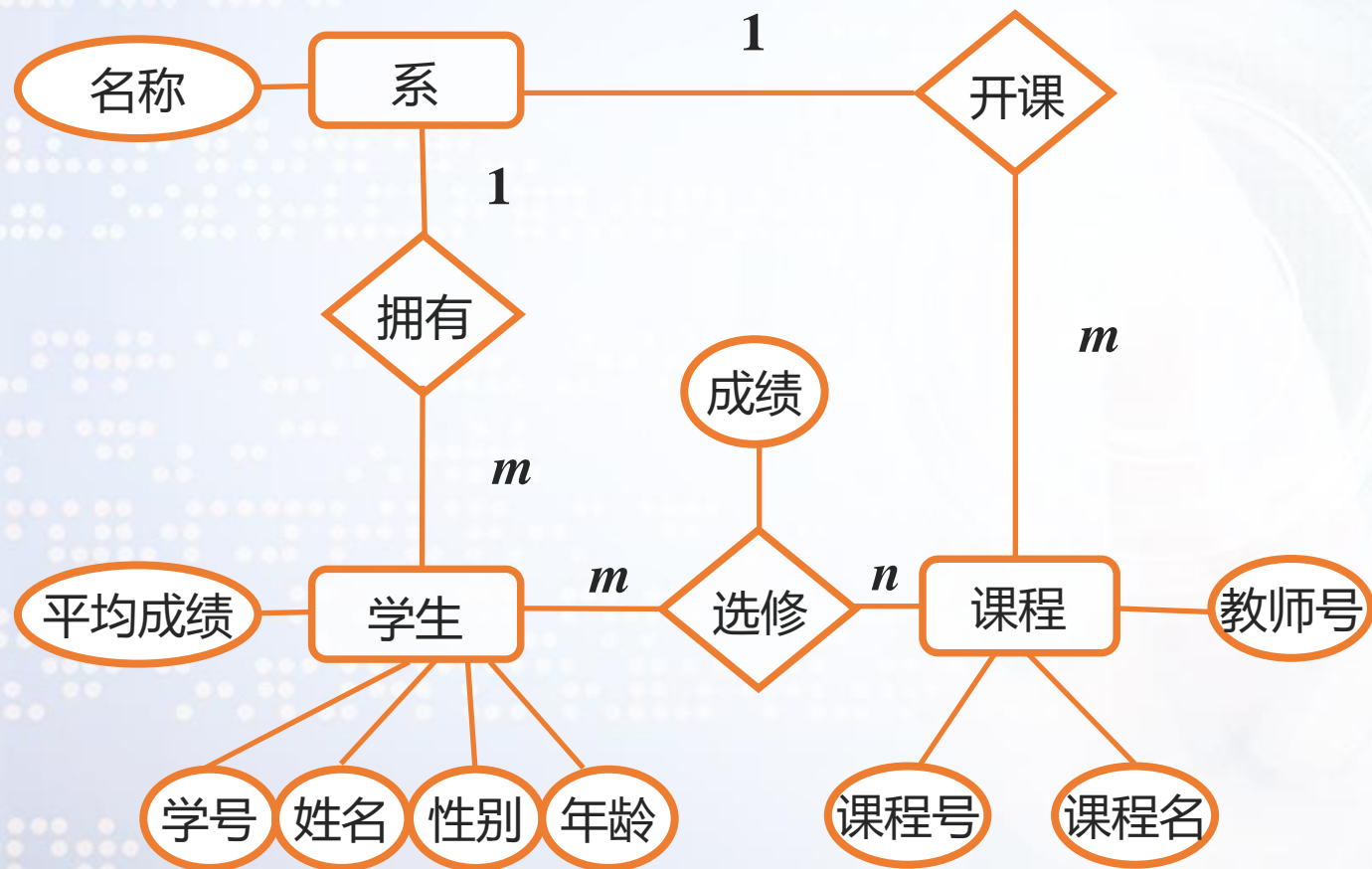
一个教师可讲授多门课程，一门课程可为多个教师讲授，因此，教师和课程也是多对多的联系。

0

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。



## 案例的概念结构设计



学生选课局部E-R图





# 案例的概念结构设计



教师任课局部E-R图



## 案例的概念结构设计



教师管理系统的初步E-R图



## 案例的概念结构设计



教师管理系统的基本E-R图



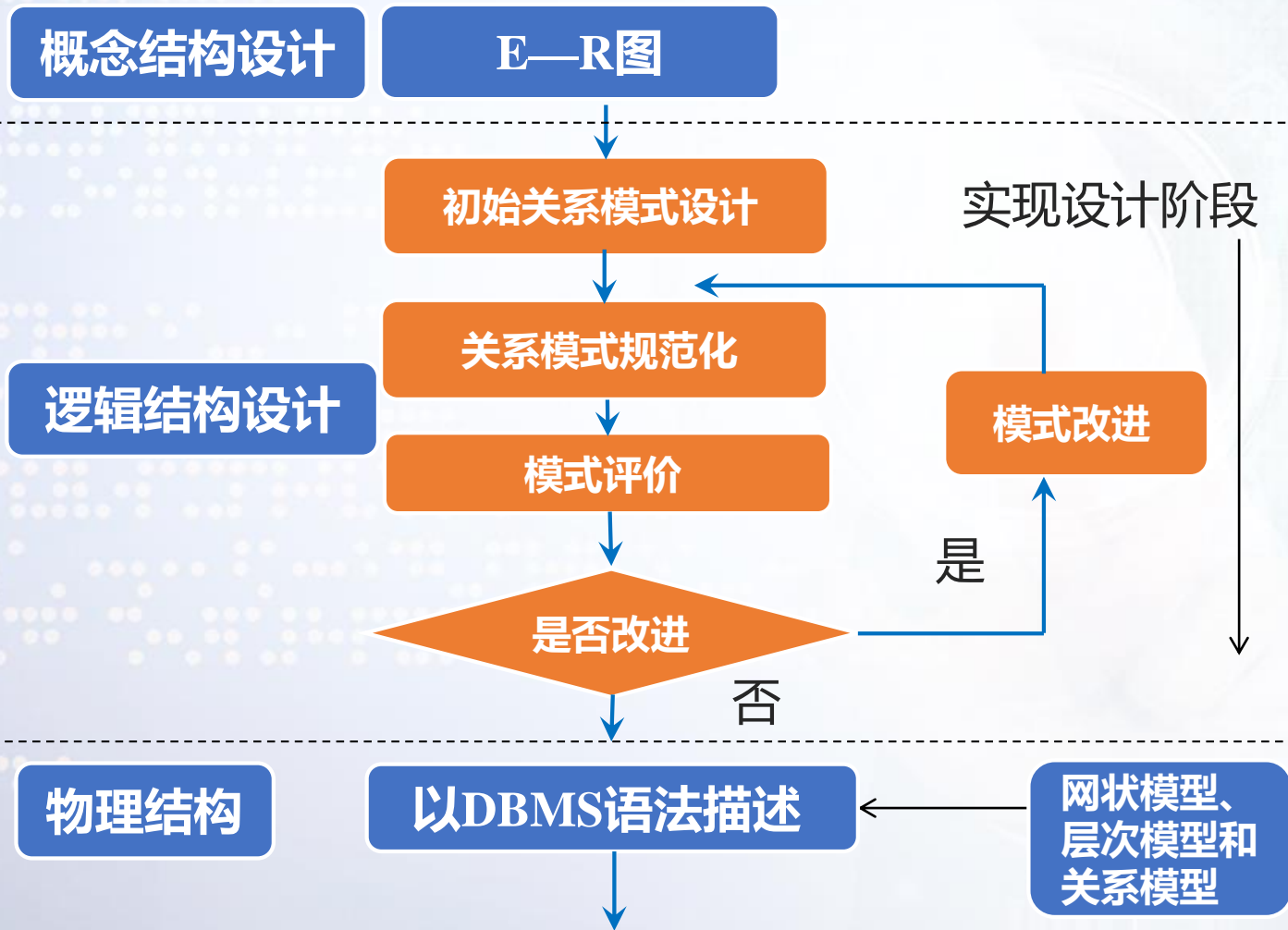
北京林业大学

# 数据库原理与应用

# 逻辑结构设计



# 逻辑结构设计的任务和步骤

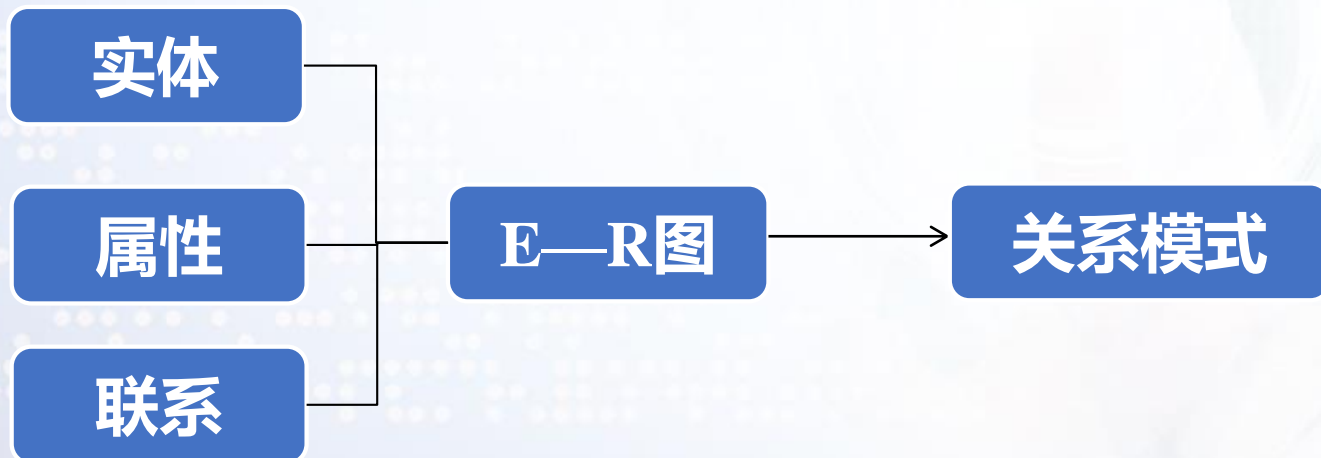






# 初始关系模式设计

## 转换原则





# 初始关系模式设计

## 转换原则



实体：一个实体转换为一个关系模式，实体的属性就是关系的属性，实体的码为关系的主码。

**联系的处理：一个联系转换为一个关系模式，有三种情况：**

- 如果联系为1:1，则每个实体的主码都是关系的候选码；
- 如果联系为1: $n$ ，则 $n$ 端实体的主码是关系的主码；
- 如果联系为 $n:m$ ，则每个实体的主码的组合是关系的主码。

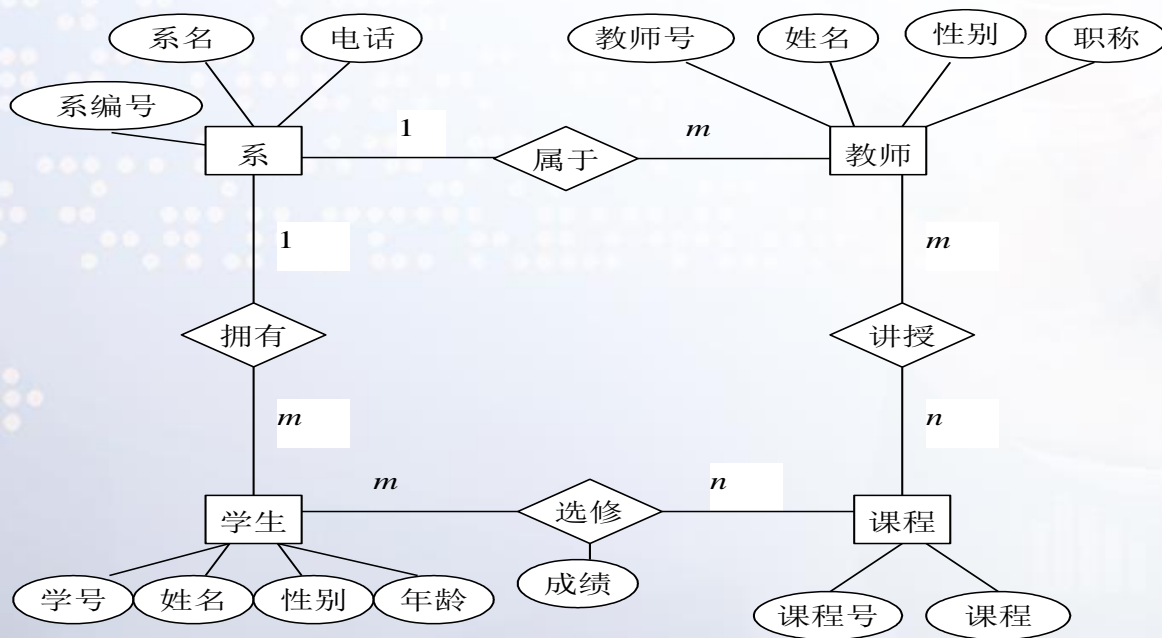


# 初始关系模式设计

## 具体做法

(1) 把每一个实体转换为一个关系模式。

(2) 把每一个联系转换为关系模式。



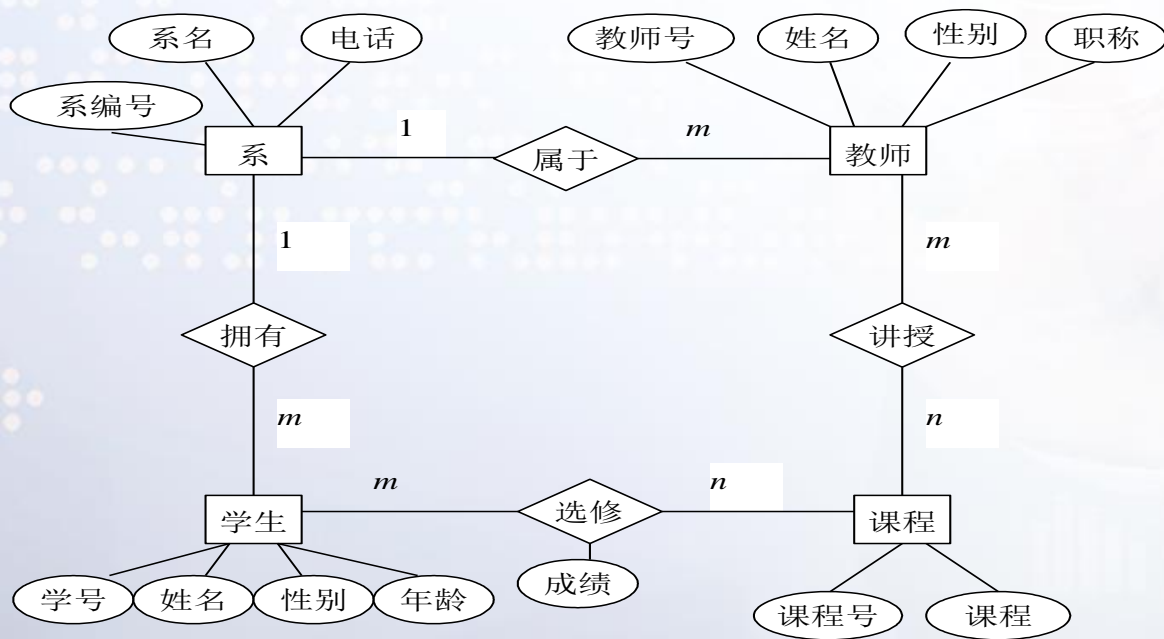




# 初始关系模式设计

## (1) 把每一个实体转换为一个关系模式。

- 学生 (学号, 姓名, 性别, 年龄)
- 系 (系编号, 系名, 电话)
- 课程 (课程号, 课程名)
- 教师 (教师号, 姓名, 性别, 职称)

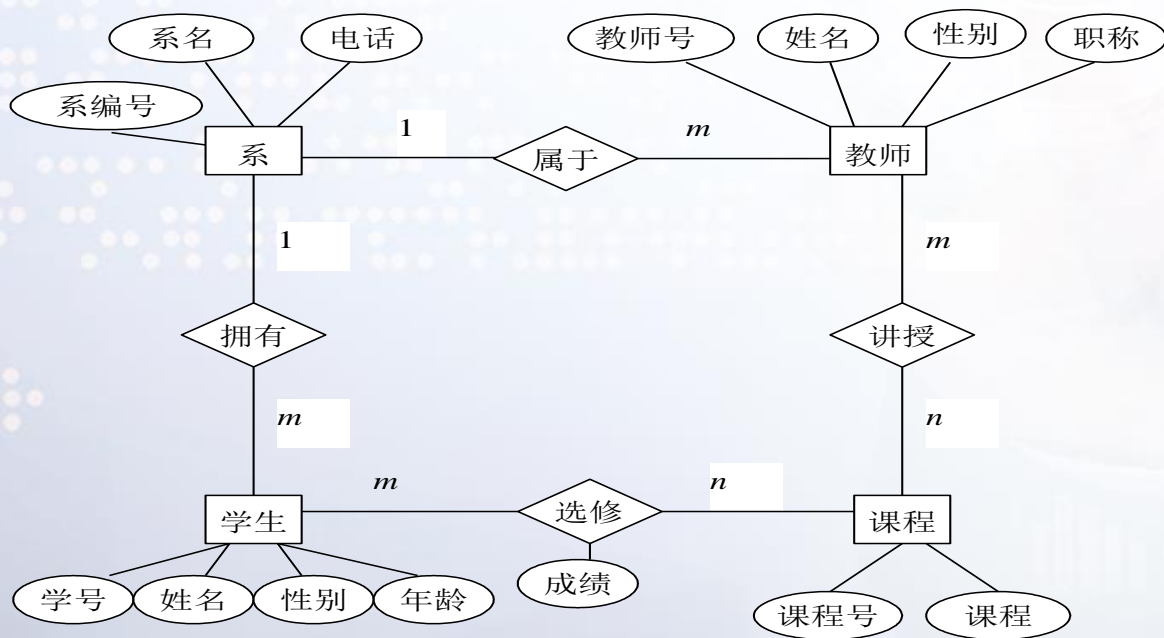




# 初始关系模式设计

## (2) 把每一个联系转换为关系模式。

- 属于 (教师号, 系编号)
- 讲授 (教师号, 课程号)
- 选修 (学号, 课程号, 成绩)
- 拥有 (系编号, 学号)



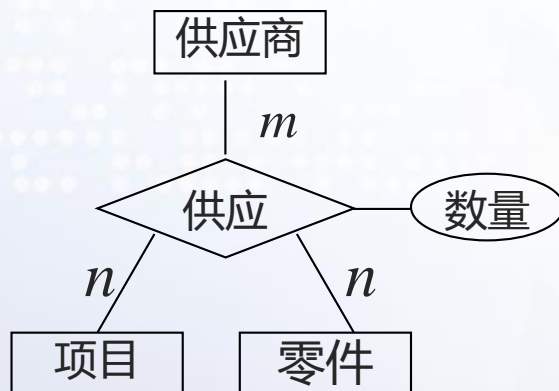


# 初始关系模式设计

## 具体做法

### (3) 特殊情况的处理。

三个或三个以上实体间的一个多元联系，  
与该多元联系相连的各实体的主码  
及联系本身的属性均转换成为关系的属性，  
转换后主码为各实体码的组合。



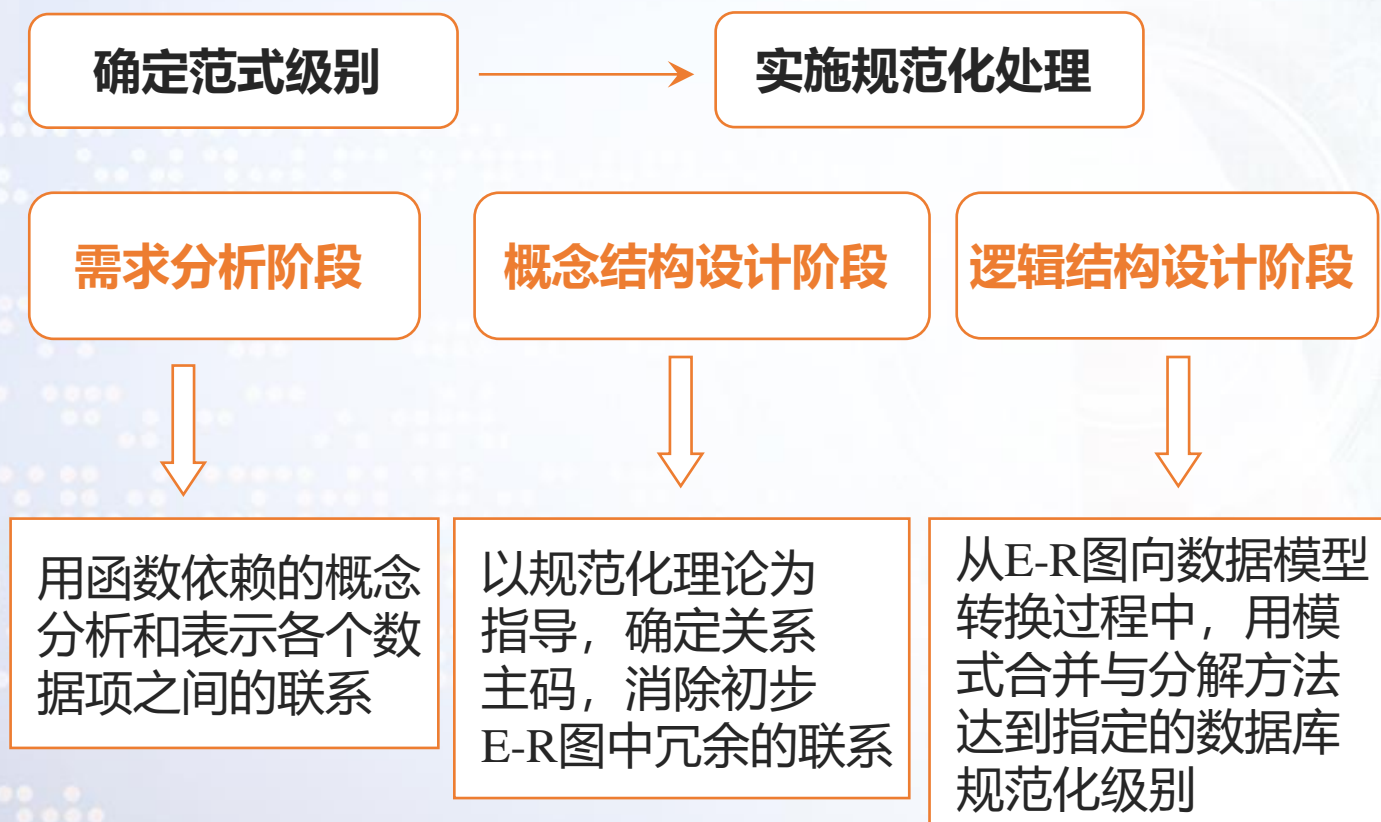
多个实体之间的联系

- 供应 (供应商号, 项目号, 零件号, 数量)





# 关系模式规范化





# 模式评价与改进

## 模式评价

### 功能评价（出现问题，回溯分析）

对照需求分析的结果，检查规范化后的关系模式集合是否支持用户所有的应用要求。

### 性能评价（执行效果）

对实际性能进行估计，包括逻辑记录的存取数、传送量以及物理结构设计算法的模型等。



# 模式评价与改进

## 模式改进



如果因为系统需求分析、概念结构设计的疏漏导致某些应用不能得到支持，则应该增加新的关系模式或属性。



如果因为性能考虑而要求改进，则可采用合并或分解的方法。

- (1) **合并**：减少连接操作，提高查询效率。
- (2) **分解**：提高数据操作的效率和存储空间的利用率。





# 案例的逻辑结构设计

## 案例的初始关系模式设计

依据转换原则，将概念模型中得到的全局E-R模型中四个实体和四个联系分别转换成四个关系模式。

## 案例关系模式的规范化

由于上述转换基于的是全局E-R模型，因此，上述转换得到的模式满足3NF。



# 案例的逻辑结构设计

## 案例关系模式的评价和改进

对关系模式进行合并处理，  
合并具有相同主码的关系模式。

- 学生 (学号, 姓名, 性别, 年龄)
- 系 (系编号, 系名, 电话)
- 课程 (课程号, 课程名)
- 教师 (教师号, 姓名, 性别, 职称)
- 属于 (教师号, 系编号)
- 讲授 (教师号, 课程号)
- 选修 (学号, 课程号, 成绩)
- 拥有 (系编号, 学号)

- ↓
- 学生 (学号, 姓名, 性别, 年龄, 系编号)
  - 系 (系编号, 系名, 电话)
  - 课程 (课程号, 课程名)
  - 教师 (教师号, 姓名, 性别, 职称, 系编号)
  - 讲授 (教师号, 课程号)
  - 选修 (学号, 课程号, 成绩)

北京林业大学

# 数据库原理与应用



# 物理结构设计



## 物理结构设计

”

对于给定逻辑模型，选取一个最适合应用环境的物理结构的过程。

任务

有效地实现逻辑模式，确定所采取的存取策略。



# 物理结构设计

## 确定物理结构

在关系数据库中主要指存取方法和存储结构。

## 评价物理结构

评价的重点是时间和空间效率。





# 确定物理结构

## 存取方法的设计

- ◆ 数据库系统是多用户共享的系统，对同一个关系要建立多条存取路径才能满足多用户的多种应用要求。物理结构设计任务之一是根据数据库管理系统支持的存取方法确定选择哪些存取方法。
- ◆ 存取方法是快速获取数据库中数据的技术。

**聚集**——针对重复的存储和高频的查询。

**索引**——保证数据的完整性，提高查询效率，但注意维护成本。



# 确定物理结构

## 存取方法的设计

**聚集**——针对重复的存储和高频的查询。

为了提高查询速度，把在一个（一组）属性上具有相同值的元组集中存放在一个物理块中，这个（这组）属性称为聚集码。

- ◆ 对经常在一起进行连接操作的关系可以建立聚集。
- ◆ 如果一个关系的一组属性经常出现在相等比较条件中，则该单个关系可建立聚集。
- ◆ 如果一个关系的一个（或一组）属性上的值重复率很高，则此单个关系可建立聚集。即对应每个聚集码值的平均元组数不能太少，太少则聚集的效果不明显。



# 确定物理结构

## 存取方法的设计

**索引**——保证数据的完整性，提高查询效率，但注意维护成本。

- ◆ 如果一个（或一组）属性经常在查询条件中出现，则考虑在这个（或这组）属性上建立索引（或组合索引）。
- ◆ 如果一个属性经常作为最大值和最小值等聚集函数的参数，则考虑在这个属性上建立索引。
- ◆ 如果一个（或一组）属性经常在连接操作的连接条件中出现，则考虑在这个（或这组）属性上建立索引。





## 确定物理结构

### 数据存放位置的设计

为了提高系统性能，应该根据应用情况将数据的**易变部分、稳定部分、经常存取部分和存取频率较低部分**分开存放。

多磁盘下：将表和索引、日志和数据库对象分开存储。

### 系统配置的设计

DBMS产品一般都提供了一些系统配置变量、存储分配参数，供设计人员和DBA对数据库进行物理优化。



## 评价物理结构

评价重点是**时间效率**和**空间效率**。

评价物理数据库的方法完全依赖于所选用的数据库管理系统，主要是从定量估算各种方案的存储空间、存取时间入手，对估算结果进行权衡、比较，选择一个较优的、合理的物理结构。

北京林业大学

# 数据库原理与应用



# 数据库实施与维护



# 数据库实施

数据库实施指根据逻辑结构设计和物理结构设计的结果，在计算机上建立实际的数据库结构、装入数据、进行测试和试运行的过程。

- ◆ 建立实际数据库结构
- ◆ 装入数据
- ◆ 应用程序编码与调试
- ◆ 数据库试运行
- ◆ 整理文档



## 建立实际数据库结构

DBMS提供的数据库定义语言（DDL）可以定义数据库结构。



## 装入数据

装入数据又称为数据库加载（Loading），是数据库实施阶段的主要工作。为了保证装入数据库中数据的正确无误，必须高度重视数据的校验工作。





## 应用程序编码与调试

- ◆ 数据库应用程序的设计属于一般的程序设计范畴，但数据库应用程序有自己的一些特点。
- ◆ 集成开发环境。



## 数据库试运行

- (1) **功能测试**。实际运行应用程序，测试它们能否完成各种预定的功能。
- (2) **性能测试**。测试系统的性能指标，分析是否符合设计目标。



## 整理文档

技术说明书，使用说明书。

完整的文件资料是应用系统的重要组成部分，但这一点常被忽视。必须强调这一工作的重要性，引起用户与设计人员的充分注意。



# 数据库实施

逻辑设计结果

物理设计结果

建立实际数据库结构  
(DDL—Table&View)

可在物理阶段  
设计子系统

装入数据-耗时  
(筛选、转换、校验)

部分测试原则  
注意备份工作

数据库试运行  
(功能和性能的测试)

应用程序编码与调试  
(借助开发工具)

整理文档

编写使用说明书  
编写技术说明书





# 数据库运行和维护

数据库运行和维护阶段的主要任务包括以下三项内容：

○ **维护数据库的安全性与完整性：** 监管权限，调整转储计划。

○ **监测并改善数据库性能：** 按照性能监控调整功能，最小化的影响现有业务。

○ **重新组织和构造数据库：** 存储位置，回收垃圾等