

# 厦门大学计算机科学系本科生课程

## 《数据库系统原理》

### 第七章 数据库设计 (2016版)

林子雨

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ▶▶

主页: <http://www.cs.xmu.edu.cn/linziyu>





# 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



# 数据库设计概述（续）

- 什么是数据库设计
  - 数据库设计是指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，使之能够有效地存储数据，满足各种用户的应用需求（信息要求和处理要求）
  - 在数据库领域内，常常把使用数据库的各类系统统称为数据库应用系统



# 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



## 7.1.1 数据库和信息系统

- 数据库是信息系统的核心和基础
  - 把信息系统中大量的数据按一定的模型组织起来
  - 提供存储、维护、检索数据的功能
  - 使信息系统可以方便、及时、准确地从数据库中获得所需的信息
- 数据库是信息系统的各个部分能否紧密地结合在一起以及如何结合的关键所在
- 数据库设计是信息系统开发和建设的重要组成部分



# 数据库设计人员应该具备的技术和知识

- 数据库的基本知识和数据库设计技术
- 计算机科学的基础知识和程序设计的方法和技巧
- 软件工程的原理和方法
- 应用领域的知识



# 数据库设计的基本步骤（续）

## 二、数据库设计的过程(六个阶段)

### 1.需求分析阶段

- 准确了解与分析用户需求（包括数据与处理）
- 是整个设计过程的基础，是最困难、最耗费时间的一步



# 数据库设计的基本步骤（续）

## 2.概念结构设计阶段

- 是整个数据库设计的关键
- 通过对用户需求进行综合、归纳与抽象，形成一个独立于具体**DBMS**的概念模型



# 数据库设计的基本步骤（续）

## 3.逻辑结构设计阶段

- 将概念结构转换为某个**DBMS**所支持的数据模型
- 对其进行优化



# 数据库设计的基本步骤（续）

## 4.数据库物理设计阶段

- 为逻辑数据模型选取一个最适合应用环境的物理结构（包括存储结构和存取方法）



# 数据库设计的基本步骤（续）

## 5.数据库实施阶段

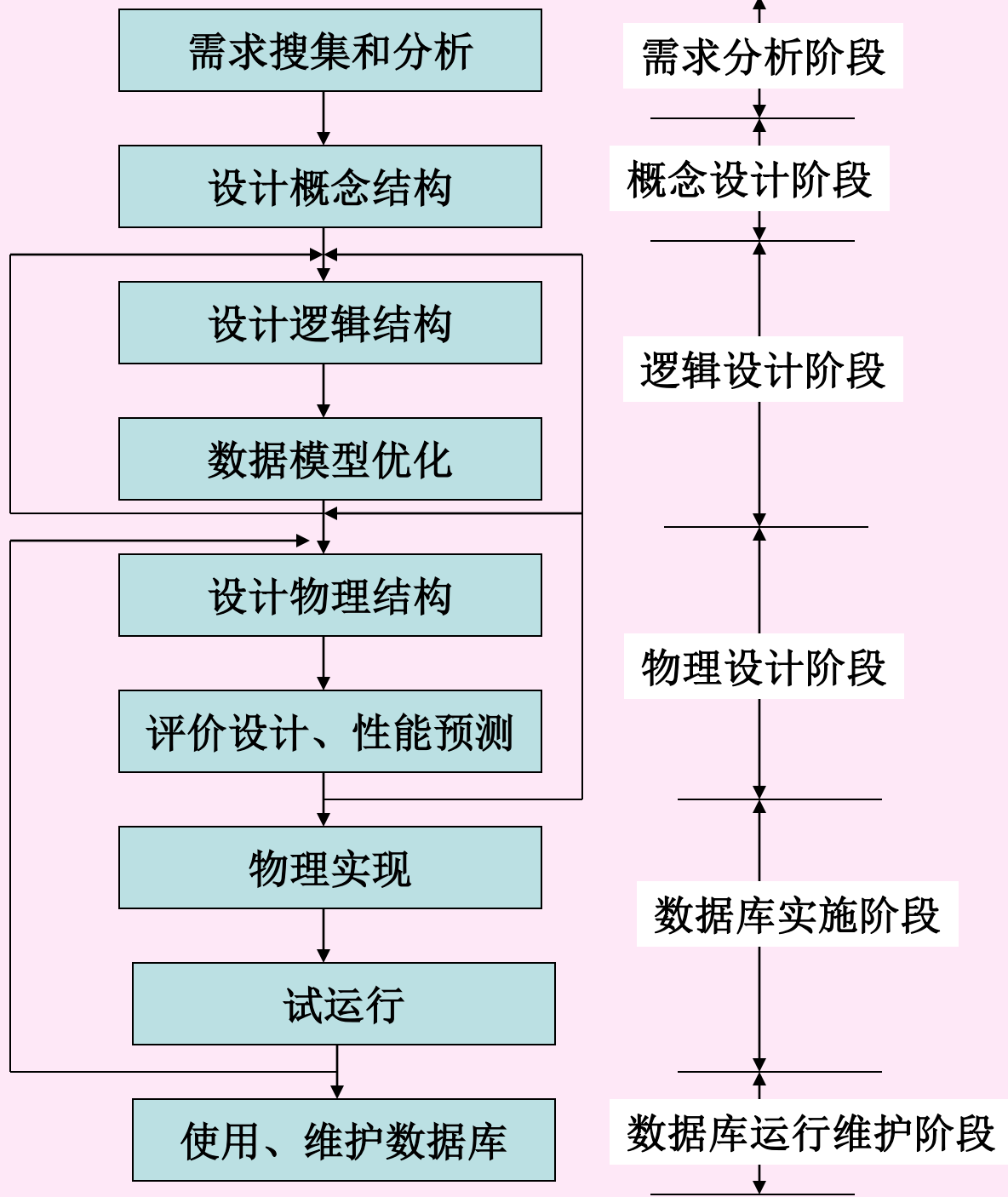
- 运用**DBMS**提供的数据库语言、工具及宿主语言，根据逻辑设计和物理设计的结果
  - 建立数据库
  - 编制与调试应用程序
  - 组织数据入库
  - 并进行试运行



# 数据库设计的基本步骤（续）

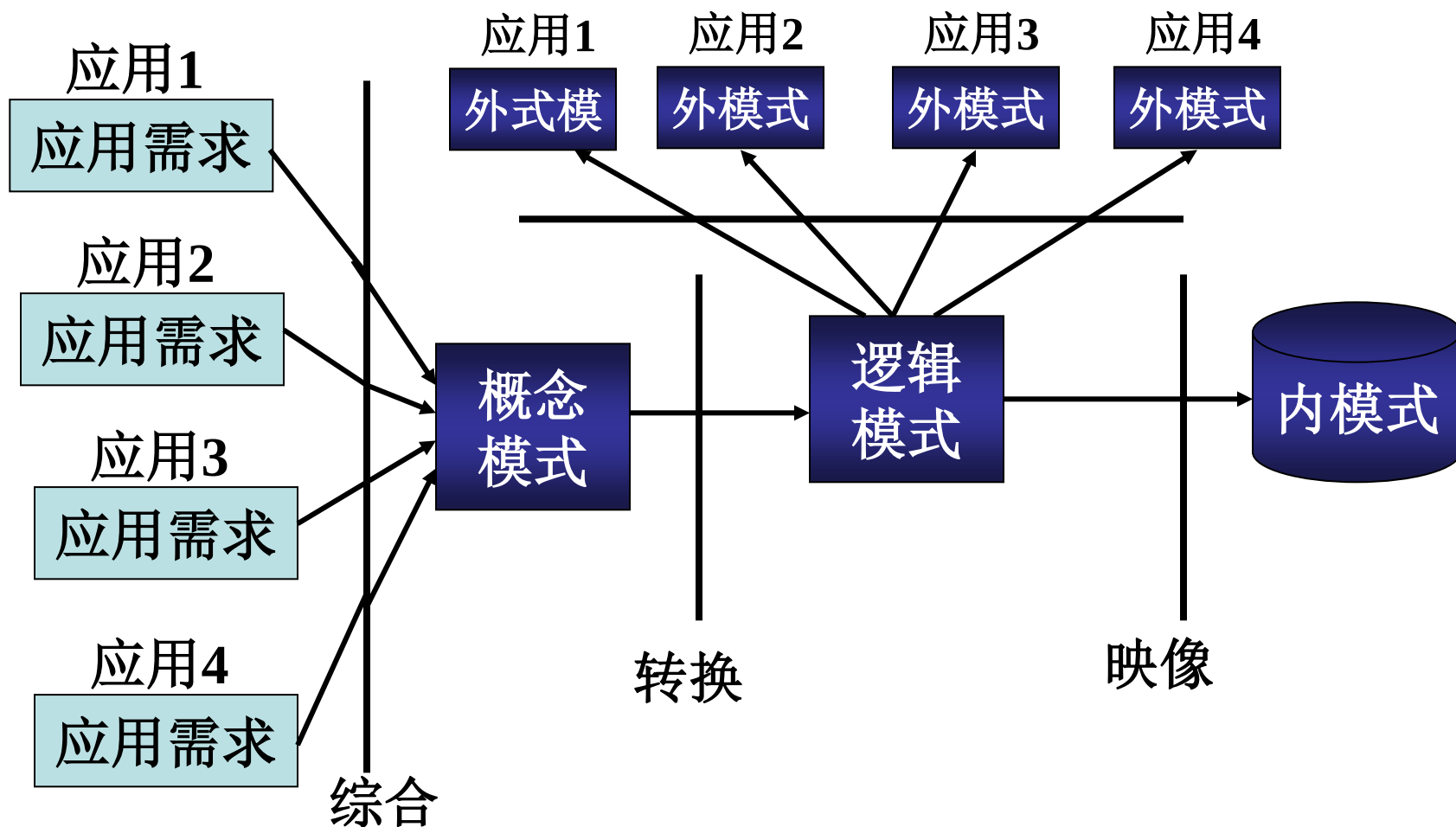
## 6.数据库运行和维护阶段

- 数据库应用系统经过试运行后即可投入正式运行。
- 在数据库系统运行过程中必须不断地对其进行评价、调整与修改。





# 数据库各级模式的形成过程





# 数据库设计的基本步骤（续）

## — 逻辑设计阶段

- 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



# 数据库设计的基本步骤（续）

## — 物理设计阶段

- 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



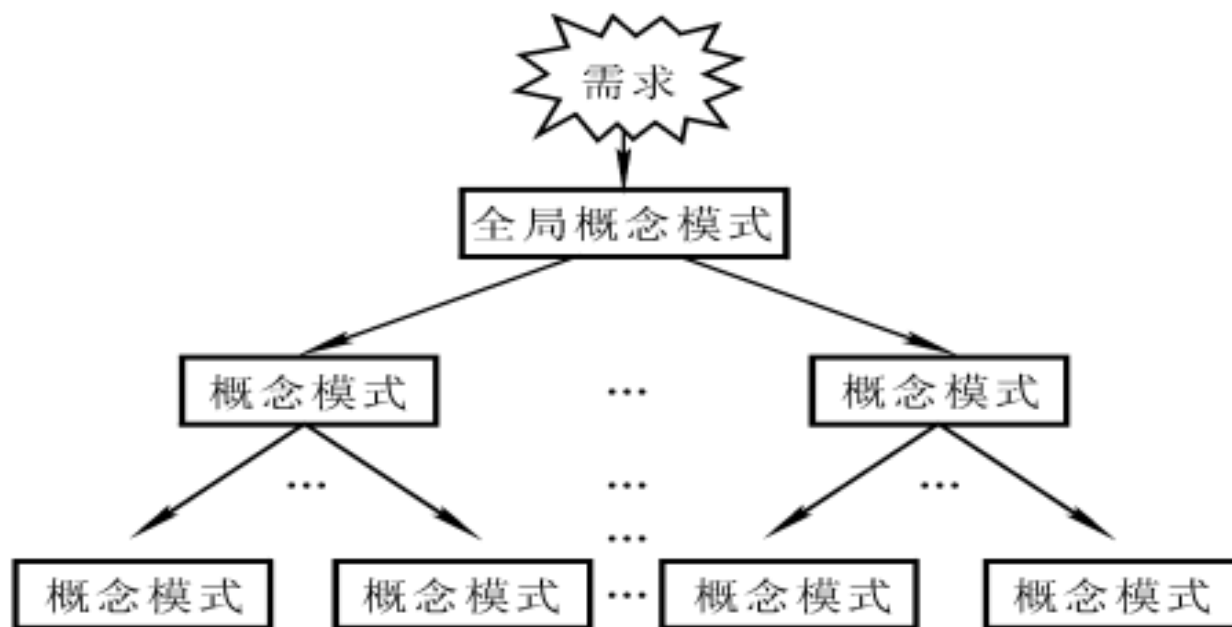
# 第七章 数据库设计

- 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



## 7.3.2 概念结构设计的方法与步骤

- 设计概念结构的四类方法
  - 自顶向下
    - 首先定义全局概念结构的框架，然后逐步细化



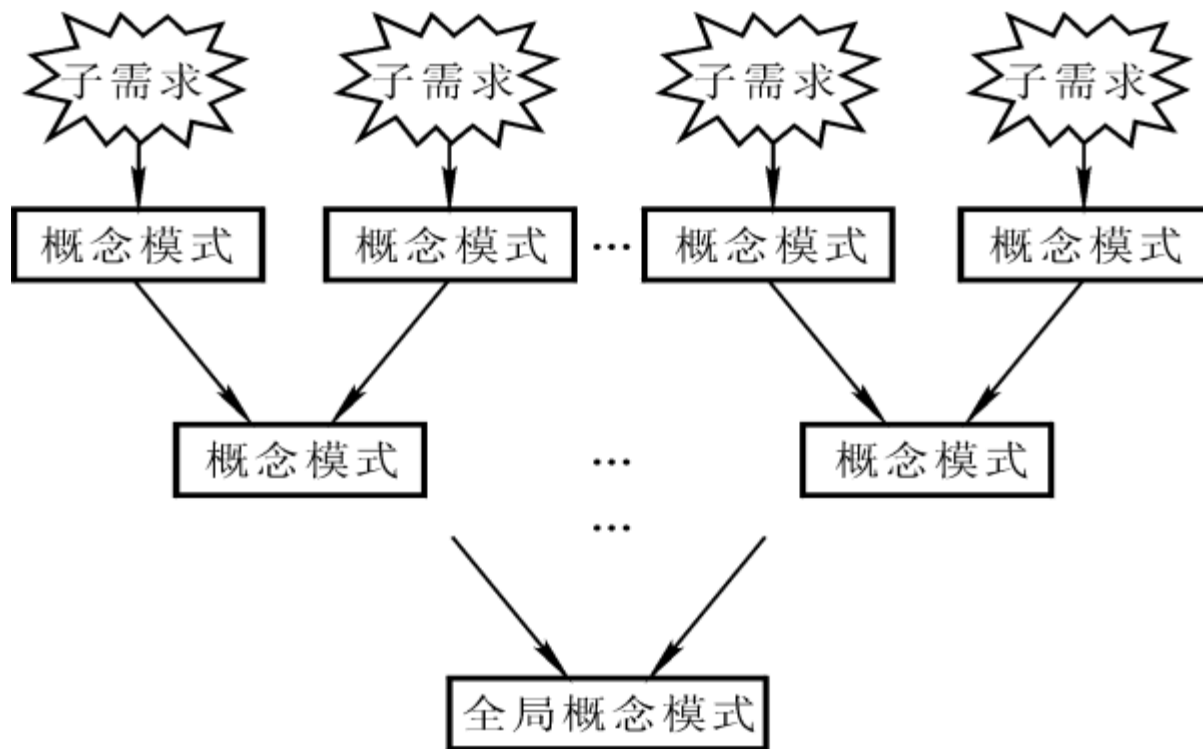


## 7.3.2 概念结构设计的方法与步骤

- 设计概念结构的四类方法

- 自底向上

- 首先定义各局部应用的概念结构，然后将它们集成起来，得到全局概念结构

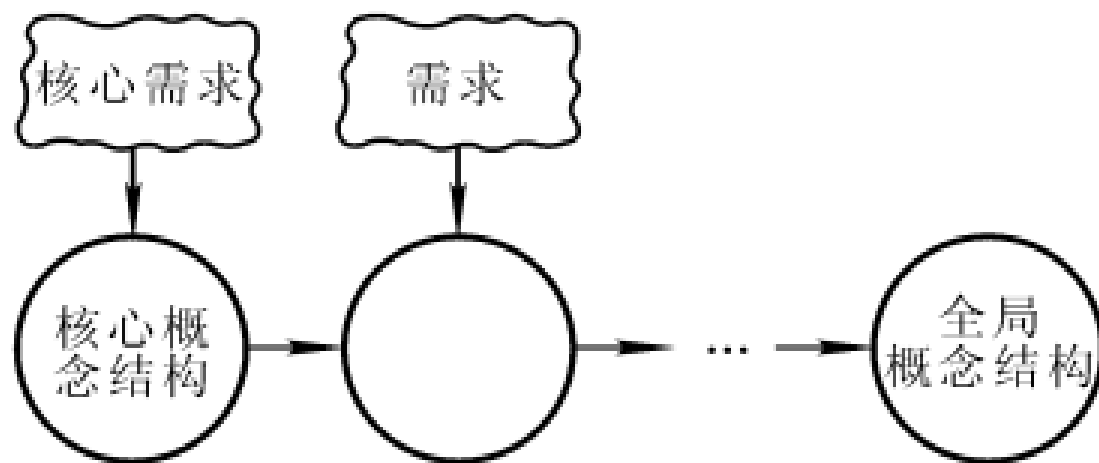




# 概念结构设计的方法与步骤（续）

## —逐步扩张

- 首先定义最重要的核心概念结构，然后向外扩充，以滚雪球的方式逐步生成其他概念结构，直至总体概念结构





# 概念结构设计的方法与步骤（续）

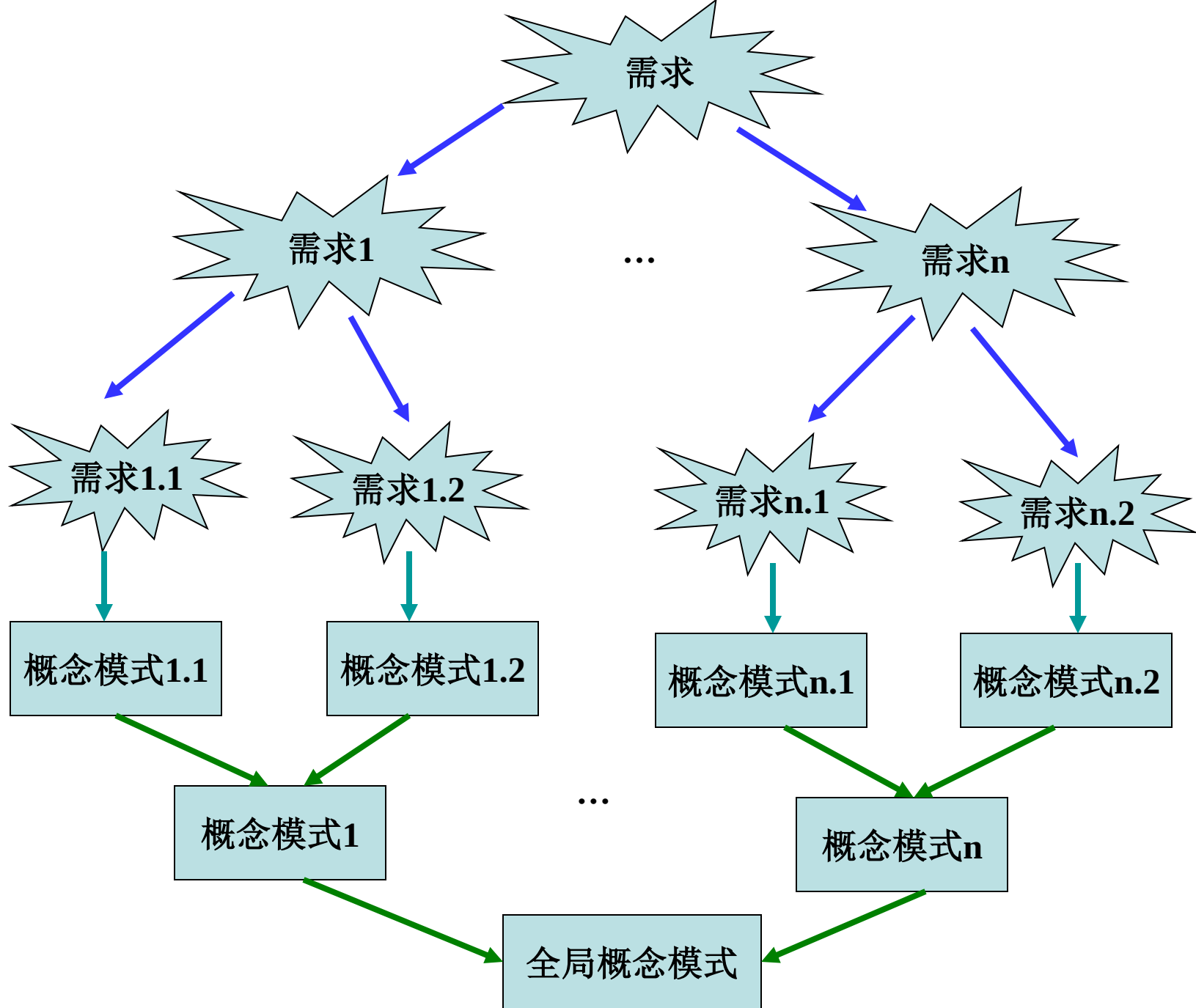
## —混合策略

- 将自顶向下和自底向上相结合，用自顶向下策略设计一个全局概念结构的框架，以它为骨架集成由自底向上策略中设计的各局部概念结构。



# 概念结构设计的方法与步骤（续）

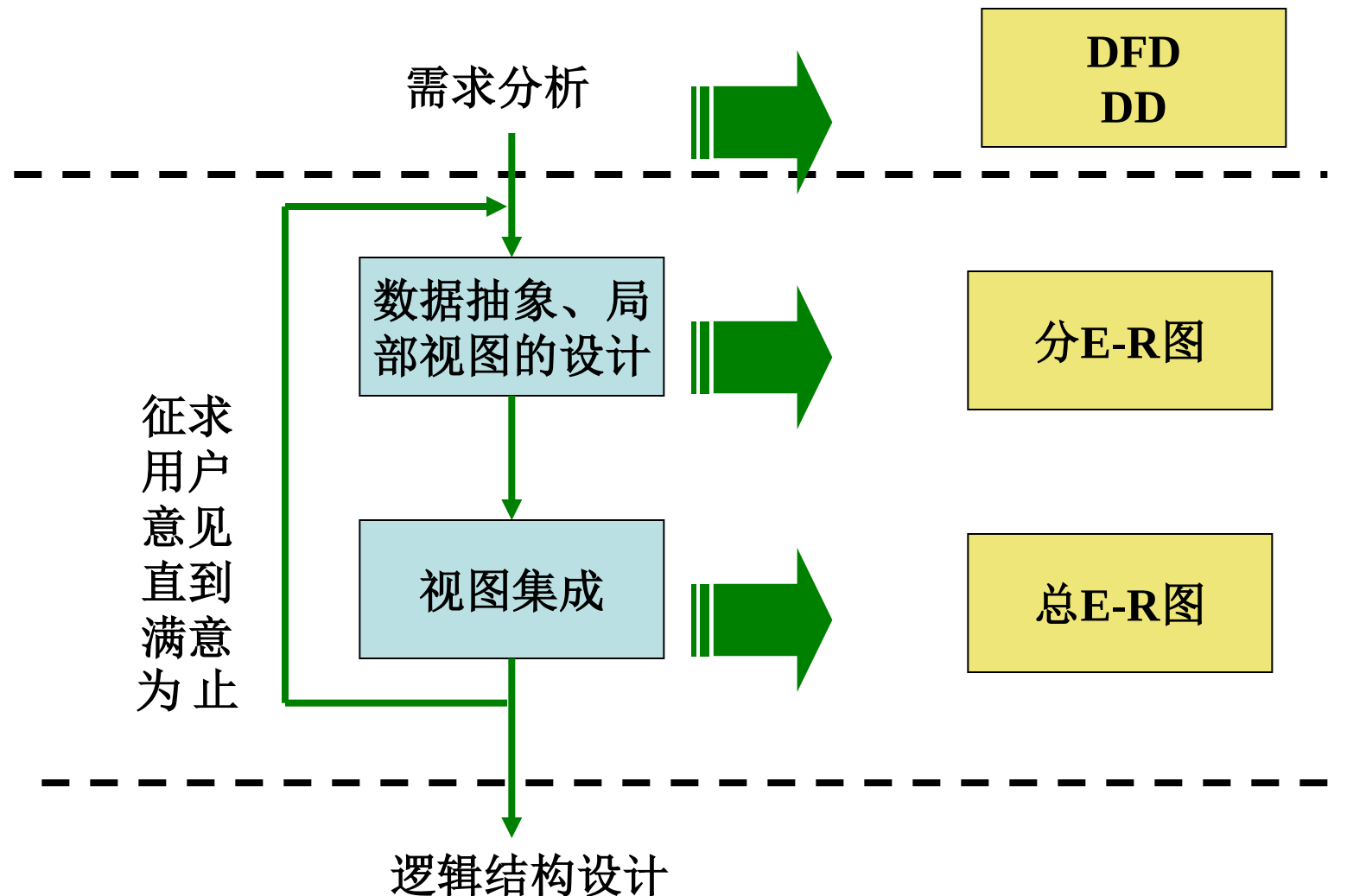
- 常用策略
  - 自顶向下地进行需求分析
  - 自底向上地设计概念结构





# 概念结构设计的方法与步骤（续）

- 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





## 7.3.3 数据抽象与局部视图设计

- 数据抽象
- 局部视图设计



# 一、数据抽象

- 概念结构是对现实世界的一种抽象
  - 从实际的人、物、事和概念中抽取所关心的共同特性，忽略非本质的细节
  - 把这些特性用各种概念精确地加以描述
  - 这些概念组成了某种模型



# 数据抽象（续）

- 数据抽象的用途

- 对需求分析阶段收集到的数据进行分类、组织（聚集），形成

- 实体

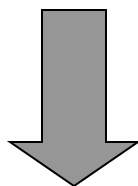
- 实体的属性，标识实体的码

- 确定实体之间的联系类型(**1:1**, **1:n**, **m:n**)



# 一、合并分E-R图，生成初步E-R图

- 各分E-R图存在冲突
  - 各个局部应用所面向的问题不同由不同的设计人员进行设计



各个分**E-R**图之间必定会存在许多不一致的地方

- 合并分**E-R**图的主要工作与关键所在：合理消除各分**E-R**图的冲突



# 合并分E-R图，生成初步E-R图（续）

- 冲突的种类
  - 属性冲突
  - 命名冲突
  - 结构冲突

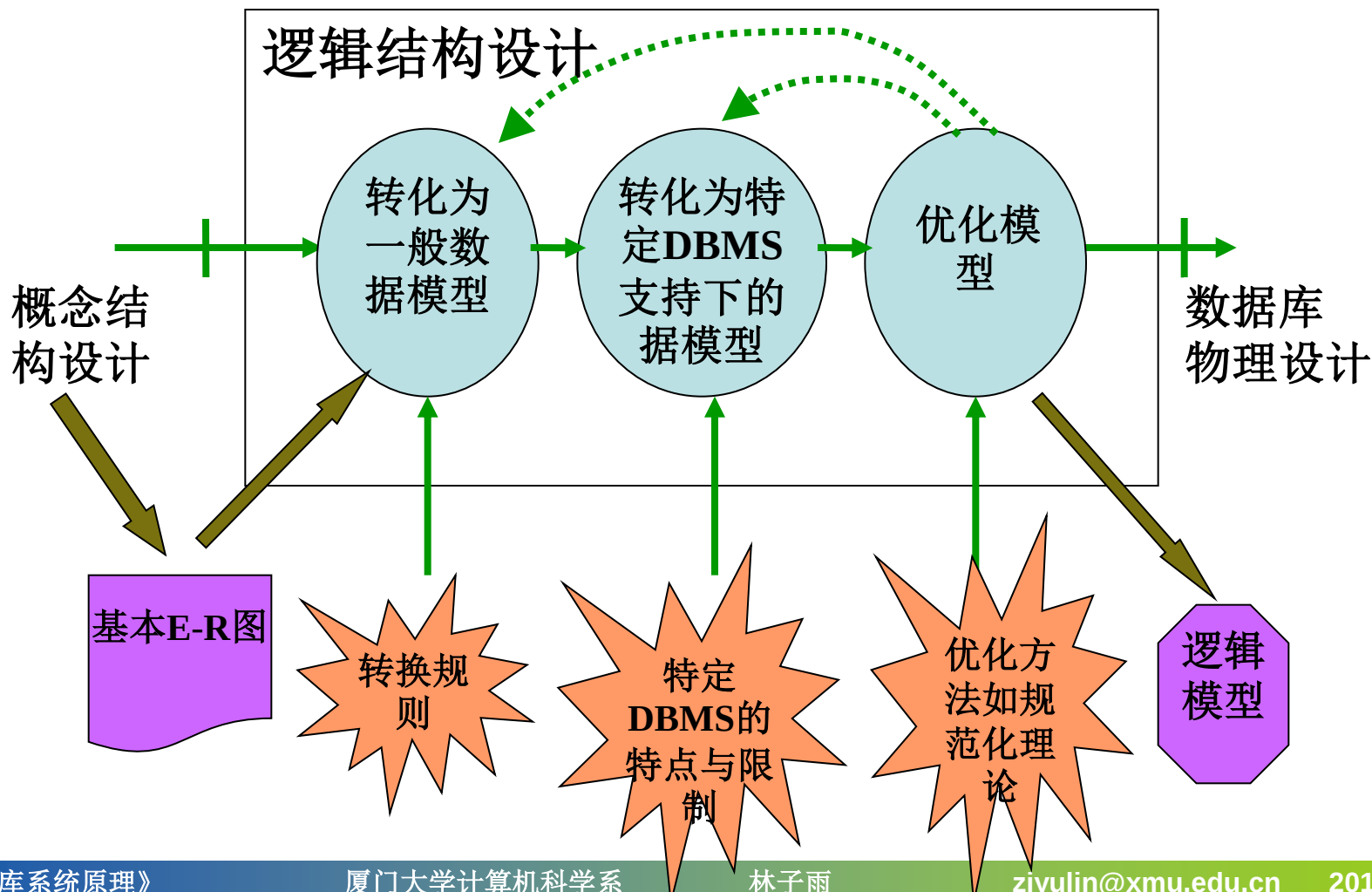


# 第七章 数据库设计

- 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



# 逻辑结构设计具体过程





# E-R图向关系模型的转换（续）

## • 转换内容

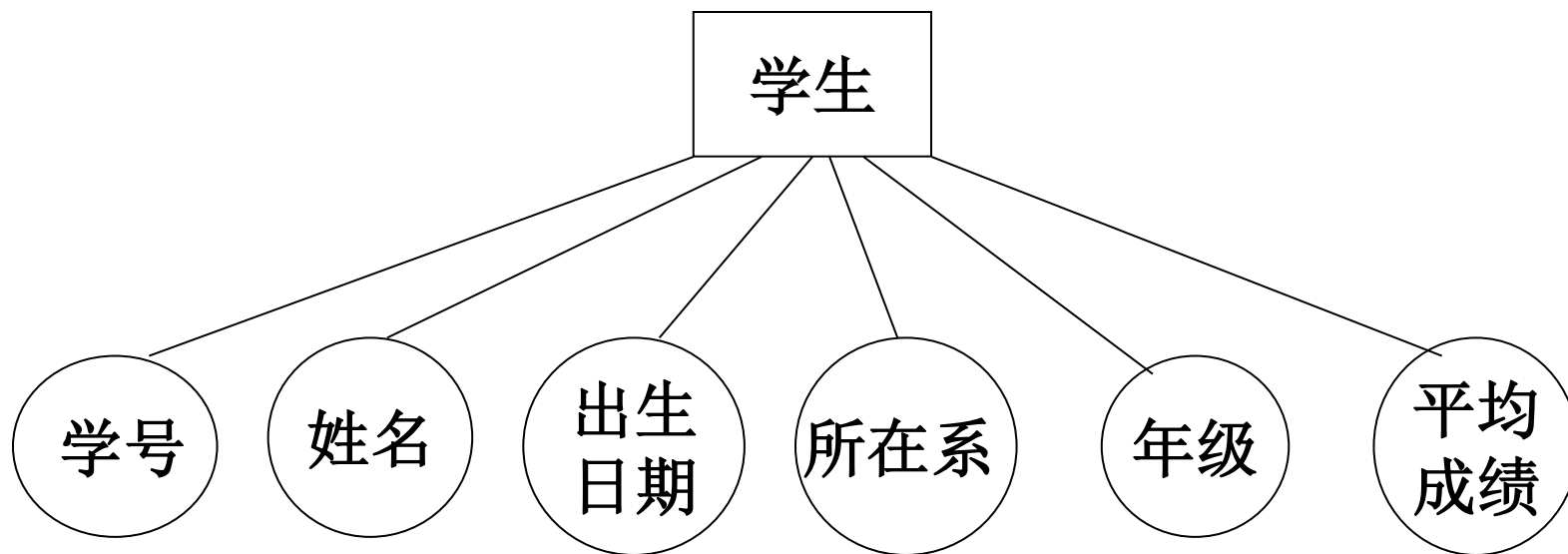
- **E-R图**由实体、实体的属性和实体之间的联系三个要素组成
- 关系模型的逻辑结构是一组关系模式的集合
- 将**E-R图**转换为关系模型：将实体、实体的属性和实体之间的联系转化为关系模式。



# E-R图向关系模型的转换（续）

## • 转换原则

1. 一个实体型转换为一个关系模式。
    - *关系的属性*: 实体型的属性
    - *关系的码*: 实体型的码
- 例，有一个学生实体



学生实体可以转换为如下关系模式：

学生（学号，姓名，出生日期，所在系，  
年级，平均成绩）



# E-R图向关系模型的转换（续）

## 2. 一个m:n联系转换为一个关系模式

### — 关系的属性:

与该联系相连的各实体的码以及联系本身的属性

### — 关系的码:

各实体码的组合

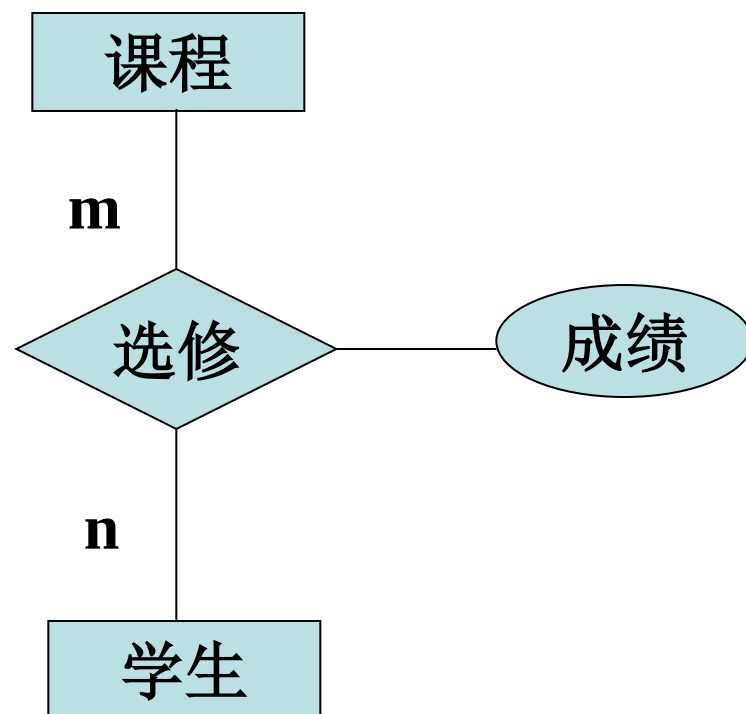
例，“选修”联系是一个m:n联系，它的E-R图为：



## E-R图向关系模型的转换（续）

按照上面规则可以将它转换为如下关系模式，其中学号与课程号为关系的组合码：

选修（学号， 课程号， 成绩）





## E-R图向关系模型的转换（续）

3. 一个1:n联系可以转换为一个独立的关系模式，也可以与n端对应的关系模式合并。

– 1) 转换为一个独立的关系模式

- 关系的属性:

与该联系相连的各实体的码以及联系本身的属性

- 关系的码:

n端实体的码



## E-R图向关系模型的转换（续）

3. 一个1:n联系可以转换为一个独立的关系模式，也可以与n端对应的关系模式合并

– 2) 与n端对应的关系模式合并

- 合并后关系的属性：在n端关系中加入1端关系的码和联系本身的属性

- 合并后关系的码：不变

– 可以减少系统中的关系个数，一般情况下更倾向于采用这种方法



# E-R图向关系模型的转换（续）

例，“组成”联系为1:n联系。

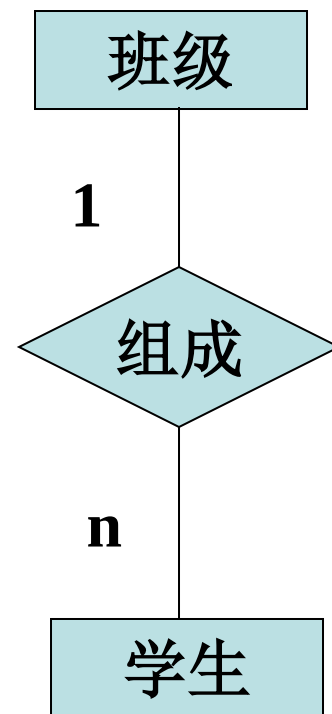
将其转换为关系模式的两种方法：

1)使其成为一个独立的关系模式：

组成（学号，班级号）

2)将其学生关系模式合并：

学生（学号，姓名，出生日期，所在系，  
年级，班级号，平均成绩）





## E-R图向关系模型的转换（续）

4. 一个1:1联系可以转换为一个独立的关系模式，也可以与任意一端对应的关系模式合并
- 1) 转换为一个独立的关系模式
    - *关系的属性*: 与该联系相连的各实体的码以及联系本身的属性
    - *关系的候选码*: 每个实体的码均是该关系的候选码



## E-R图向关系模型的转换（续）

4. 一个1:1联系可以转换为一个独立的关系模式，也可以与任意一端对应的关系模式合并。

### – 2) 与某一端对应的关系模式合并

- 合并后关系的属性：加入对应关系的码和联系本身的属性
- 合并后关系的码：不变

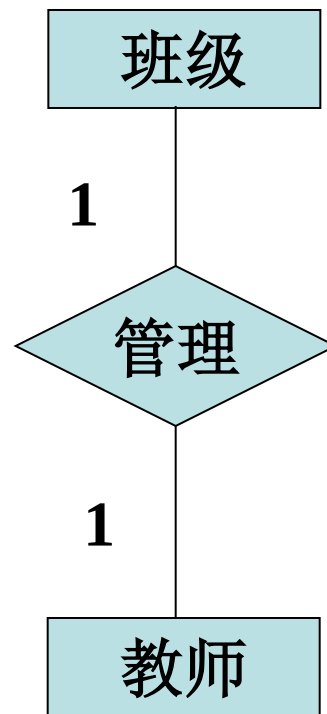


## E-R图向关系模型的转换（续）

例，“管理”联系为1:1联系，可以有三种转换方法：

（1）转换为一个独立的关系模式：

管理（职工号，班级号）  
或  
管理（职工号，班级号）

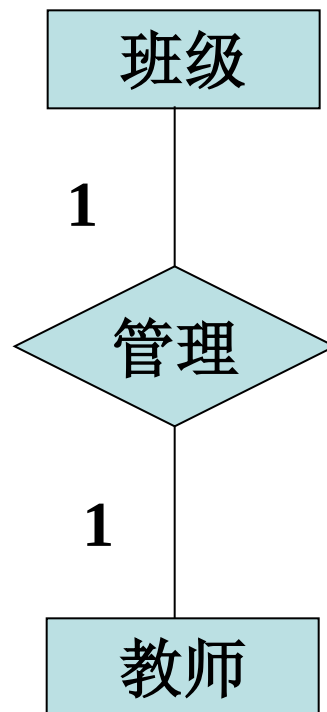




## E-R图向关系模型的转换（续）

例，“管理”联系为**1:1**联系，可以  
有三种转换方法：

**(2)** “管理”联系与班级关系模式  
合并，则只需在班级关系中加入教  
师关系的码，即职工号：



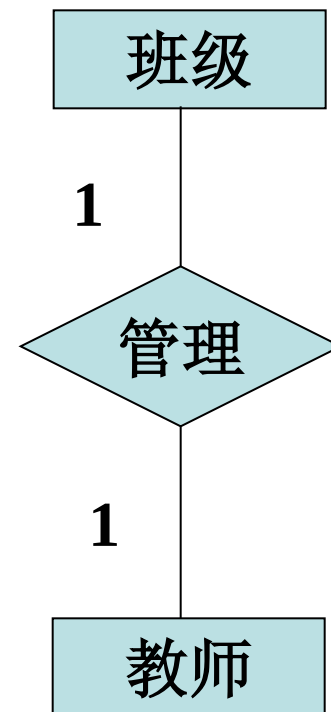
班级：（班级号，学生人数，职工号）



## E-R图向关系模型的转换（续）

例，“管理”联系为**1:1**联系，可以有三种转换方法：

**(3)** “管理”联系与教师关系模式合并，则只需在教师关系中加入班级关系的码，即班级号：



教师：（职工号，姓名，性别，职称，班级号，是否为优秀班主任）



# E-R图向关系模型的转换（续）

注意：

- 从理论上讲，**1:1**联系可以与任意一端对应的关系模式合并
- 但在一些情况下，与不同的关系模式合并效率会大不一样。因此究竟应该与哪端的关系模式合并需要依应用的具体情况而定
- 由于连接操作是最费时的操作，所以一般应以尽量减少连接操作为目标
- 例如，如果经常要查询某个班级的班主任姓名，则将管理联系与教师关系合并更好些



## E-R图向关系模型的转换（续）

5. 三个或三个以上实体间的一个多元联系转换为一个关系模式。

- *关系的属性*: 与该多元联系相连的各实体的码以及联系本身的属性
- *关系的码*: 各实体码的组合

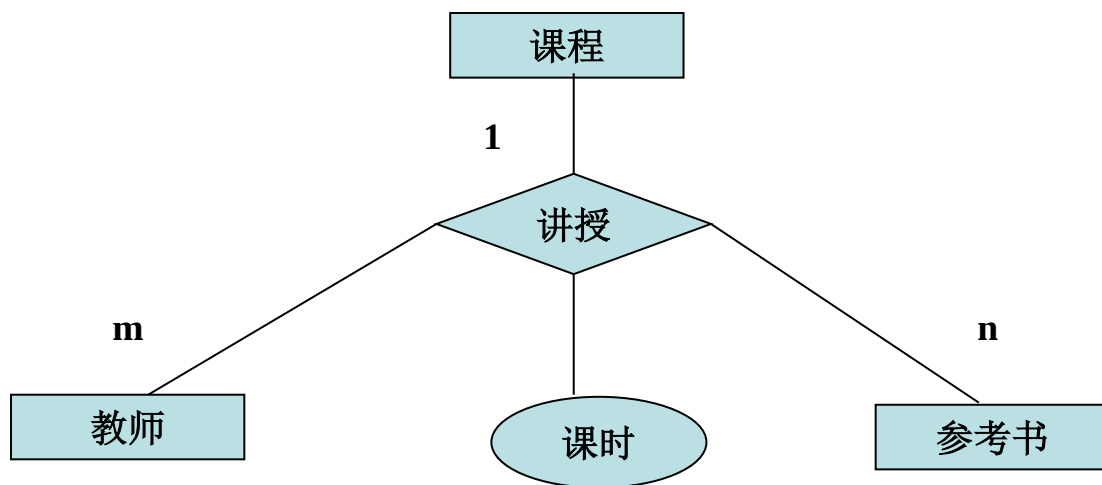
例，“讲授”联系是一个三元联系



## E-R图向关系模型的转换（续）

“讲授”联系是一个三元联系，可以将它转换为如下关系模式，其中课程号、职工号和书号为关系的组合码：

讲授（课程号，职工号，书号，课时）





## E-R图向关系模型的转换（续）

6. 同一实体集的实体间的联系，即自联系，也可按上述**1:1**、**1:n**和**m:n**三种情况分别处理。

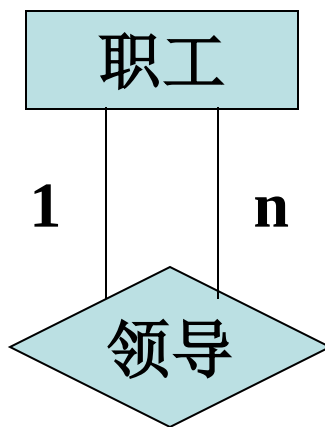
例，如果教师实体集内部存在领导与被领导的**1:n**自联系



## E-R图向关系模型的转换（续）

教师实体集内部存在领导与被领导的1:n自联系，我们可以将该联系与教师实体合并，这时主码职工号将多次出现，但作用不同，可用不同的属性名加以区分：

教师： { 职工号，姓名，性别，职称，系主任职工号 }





# E-R图向关系模型的转换（续）

## 7. 具有相同码的关系模式可合并

- 目的：减少系统中的关系个数
- 合并方法：将其中一个关系模式的全部属性加入到另一个关系模式中，然后去掉其中的同义属性（可能同名也可能不同名），并适当调整属性的次序



# 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



## 7.5 数据库的物理设计

- 什么是数据库的物理设计
  - 数据库在物理设备上的存储结构与存取方法称为数据库的物理结构，它依赖于给定的计算机系统
  - 为一个给定的逻辑数据模型选取一个最适合应用环境的物理结构的过程，就是数据库的物理设计

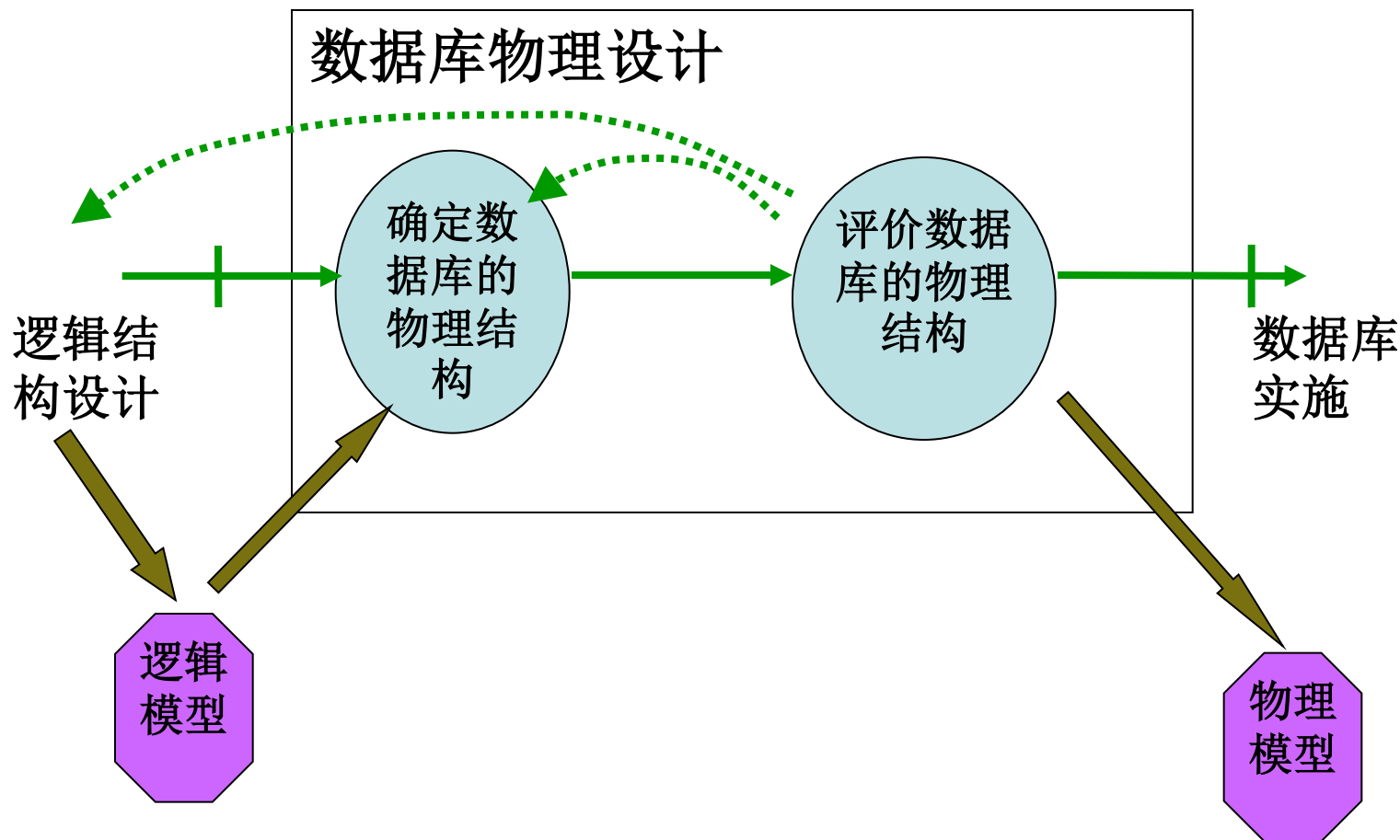


## 7.5 数据库的物理设计

- 数据库物理设计的步骤
  - 确定数据库的物理结构
  - 对物理结构进行评价，评价的重点是时间和空间效率
  - 如果评价结果满足原设计要求则可进入到物理实施阶段，否则，就需要重新设计或修改物理结构，有时甚至要返回逻辑设计阶段修改数据模型。



# 物理设计过程





## 7.5.2 关系模式存取方法选择

- 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。
- 物理设计的第一个任务就是要确定选择哪些存取方法，即建立哪些存取路径。



# 关系模式存取方法选择（续）

- **DBMS常用存取方法**
  - 索引方法，目前主要是**B+**树索引方法
  - 聚簇（**Cluster**）方法
  - **HASH**方法



# 一、索引存取方法的选择

- 选择索引存取方法的主要内容

根据应用要求确定

- 对哪些属性列建立索引
- 对哪些属性列建立组合索引
- 对哪些索引要设计为唯一索引



# 索引存取方法的选择（续）

- 选择索引存取方法的一般规则
  - 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



## 索引存取方法的选择（续）

- 关系上定义的索引数过多会带来较多的额外开销
  - 维护索引的开销
  - 查找索引的开销



# 建立聚簇索引

- 聚簇索引

- 建立聚簇索引后，基表中数据也需要按指定的聚簇属性值的升序或降序存放。也即聚簇索引的索引项顺序与表中元组的物理顺序一致。

例：

**CREATE CLUSTER INDEX Stusname ON Student(Sname);**

在**Student**表的**Sname**（姓名）列上建立一个聚簇索引，而且**Student**表中的记录将按照**Sname**值的升序存放



# 建立聚簇索引

- 在一个基本表上最多只能建立一个聚簇索引
- 聚簇索引的用途：对于某些类型的查询，可以提高查询效率
- 聚簇索引的适用范围
  - 很少对基表进行增删操作
  - 很少对其中的变长列进行修改操作



# 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

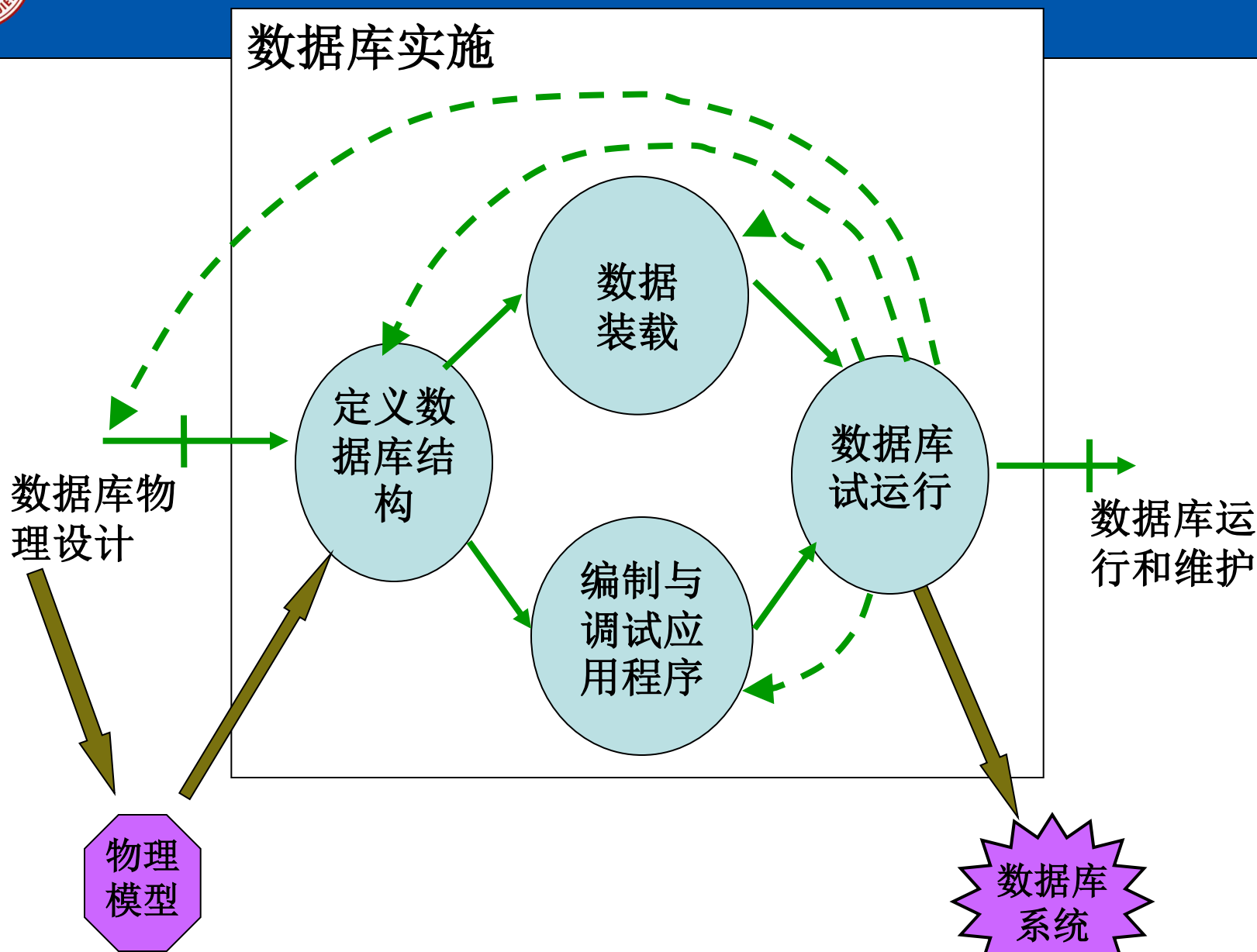


## 7.6 数据库的实施和维护

- 数据库实施的工作内容
  - 用**DDL**定义数据库结构
  - 组织数据入库
  - 编制与调试应用程序
  - 数据库试运行



## 7.6 数据库的实施过程





# 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



## 7.7 小结

- 数据库的设计过程

- 需求分析
- 概念结构设计
- 逻辑结构设计
- 物理设计
- 实施
- 运行维护

设计过程中往往还会有许多反复。



## 小结（续）

- 数据库各级模式的形成
  - 数据库的各级模式是在设计过程中逐步形成的
  - 需求分析阶段综合各个用户的应用需求（现实世界的需求）。
  - 概念设计阶段形成独立于机器特点、独立于各个DBMS产品的概念模式（信息世界模型），用E-R图来描述。



## 小结（续）

- 在逻辑设计阶段将**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**。
- 在物理设计阶段根据**DBMS**特点和处理的需要，进行物理存储安排，设计索引，形成数据库**内模式**。



## 小结（续）

- 整个数据库设计过程体现了结构特征与行为特征的紧密结合。



## 小结（续）

- 目前很多**DBMS**都提供了一些辅助工具（**CASE**工具），为加快数据库设计速度，设计人员可根据需要选用。

例如需求分析完成之后，设计人员可以使用**ORACLE DESIGNER**画**E-R**图，将**E-R**图转换为关系数据模型，生成数据库结构；画数据流图，生成应用程序。



# 附录：主讲教师



主讲教师：林子雨

单位：厦门大学计算机科学系

E-mail: [ziyulin@xmu.edu.cn](mailto:ziyulin@xmu.edu.cn)

个人网页: <http://www.cs.xmu.edu.cn/linziyu>

数据库实验室网站: <http://dblab.xmu.edu.cn>



扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），现为厦门大学计算机科学系助理教授（讲师），曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革局副局长。中国高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度厦门大学奖教金获得者。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学基金青年基金项目(No.61303004)、1项福建省自然科学基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。编著出版中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》并成为畅销书籍，编著并免费网络发布40余万字中国高校第一本闪存数据库研究专著《闪存数据库概念与技术》；主讲厦门大学计算机系本科生课程《数据库系统原理》和研究生课程《分布式数据库》《大数据技术基础》。具有丰富的政府和企业信息化培训经验，曾先后给中国移动通信集团公司、福州马尾区政府、福建省物联网科学研究院、石狮市物流协会、厦门市物流协会、福建龙岩卷烟厂等多家单位和企业开展信息化培训，累计培训人数达2000人以上。



# 附录：课程助教



**助教：谢荣东**

单位：厦门大学计算机科学系数据库实验室2014级硕士研究生

E-mail: [xrdxmu@sina.com](mailto:xrdxmu@sina.com)



**助教：薛倩**

单位：厦门大学计算机科学系数据库实验室2015级硕士研究生

E-mail: [xueqian\\_victoria@163.com](mailto:xueqian_victoria@163.com)



# 附录：班级网站

林子雨主讲《数据库系统原理》2016班级主页

<http://dblab.xmu.edu.cn/post/5550/>



扫一扫访问班级网站  
支持手机浏览

The background is a solid blue color. It features several faint, light-blue silhouettes of people. In the top left, there is a group of three people holding hands. In the top right, there is a group of six people, some holding hands. On the right side, there is a large silhouette of a person standing with their hand on their chin. In the bottom left, there is a silhouette of a person's head and shoulders.

# Thank You!

Department of Computer Science, Xiamen University, 2016