

第7章 数据库设计 (1)

胡 敏

合肥工业大学

jsjxhumin@hfut.edu.cn

第 7 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 物理结构设计

7.6 数据库的实施和维护

7.7 小结



第 7 章 数据库设计

■ 掌握

- DBD的概念、DBD的主要内容和特点
- DBD各阶段的任务和工作步骤
- DBD中概念结构设计的重要性
- 概念模型(E-R模型)的设计方法
- 关系数据库逻辑设计的方法和步骤
- E-R模型向关系模型的转换

■ 了解理解

- 物理设计的性能及其表示
- 数据库实施、运行维护的主要工作



7.1 数据库设计概述（1）

广义地讲，是数据库及其应用系统的设计，即设计整个数据库应用系统；

狭义地讲，是设计数据库本身，即设计数据库的各级模式并建立数据库，这是数据库应用系统设计的一部分。

■ 1. 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，设计一个优良的数据库逻辑模式和物理结构，并据此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，满足各种用户的应用需求，包括信息管理要求和数据处理要求：

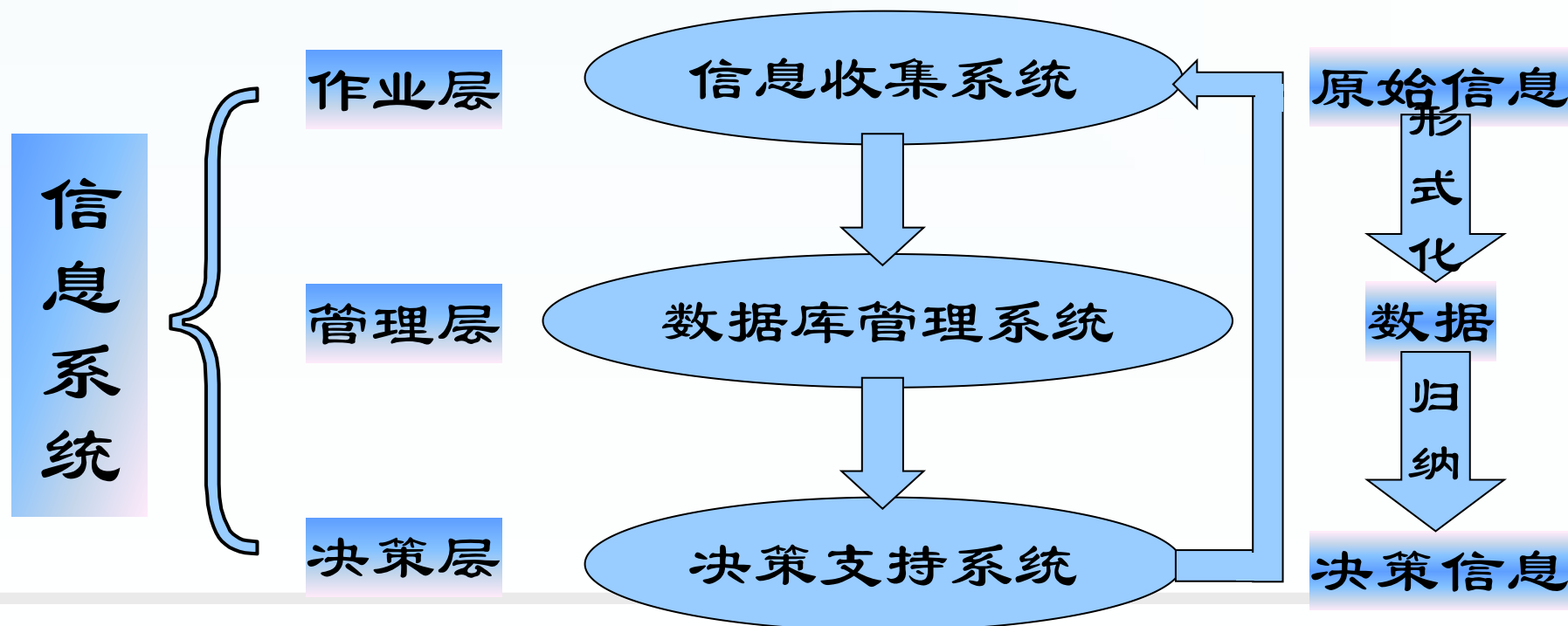
- 信息管理要求：在数据库中存储和管理需要的数据对象。
- 数据处理要求：对数据对象需要进行的处理，如查询、增删改、统计和分析等。（安全性与完整性要求）



7.1 数据库设计概述（2）

■ 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



7.1.1 数据库设计的特点

■ 1. 数据库建设的基本规律

➤ 综合性

★ 需要计算机专业知识及业务系统专业知识；解决技术及非技术两方面的问题。

★ 数据库建设是硬件、软件和干件的结合

— 三分技术，七分管理，十二分基础数据

— 技术与管理的界面称之为“干件”

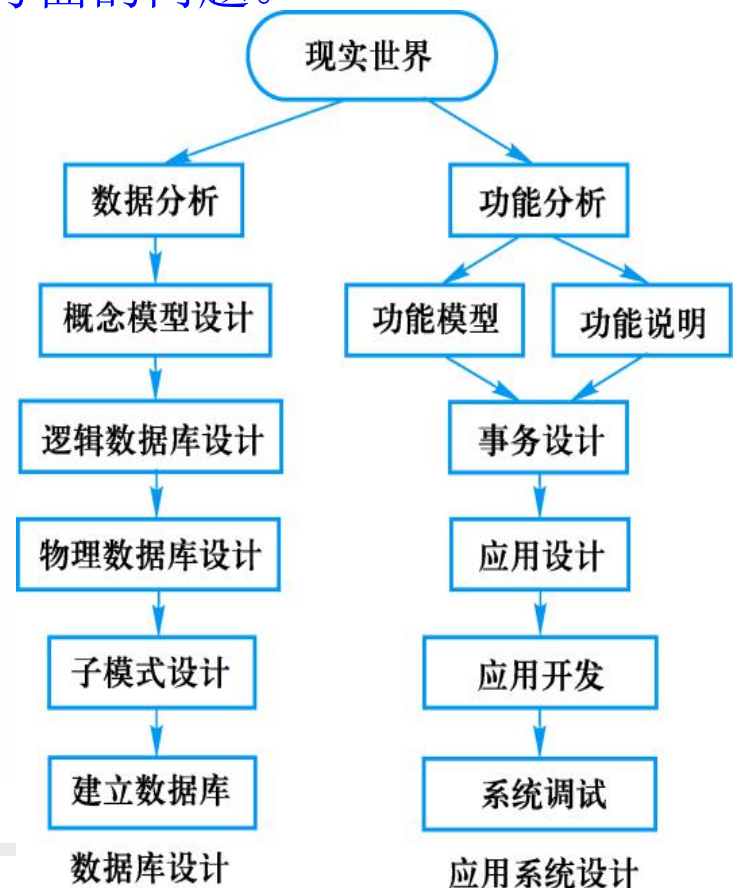
★ 管理：
┌ 数据库建设项目管理
└ 企业（即应用部门）的业务管理

★ 基础数据：数据的收集、整理、组织和不断更新

■ 2. 结构（数据）设计和行为（处理）设计相结合

➤ 结构（数据）设计：设计数据库框架或数据库结构

➤ 行为（处理）设计：设计应用程序、事务处理等



7.1.2 数据库设计方法

- 大型数据库设计是涉及多学科的综合性的技术，又是一项庞大的工程项目。
- 要求多方面的知识和技术。主要包括：
 - 计算机的基础知识
 - 软件工程的原理和方法
 - 程序设计的方法和技巧
 - 数据库的基本知识
 - 数据库设计技术
 - 应用领域的知识



7.1.2 数据库设计方法（续）

■ 3. 数据库设计的方法概述

➤ 手工试凑法

➤ 规范设计法

★ 典型方法

➤ 新奥尔良（New Orleans）方法

1978年10月
新奥尔良法

需求分析

概念结构设计

逻辑结构设计

物理结构设计

将数据库设计分为若干阶段和步骤：S.B.Yao（姚诗斌）方法

采用辅助手段实现每一过程

按设计规程用工程化方法设计数据库



7.1.2 数据库设计方法（续）

■ 基于E-R模型的设计方法

概念设计阶段广泛采用

■ 3NF（第三范式）的设计方法

逻辑阶段可采用的有效方法

■ ODL(Object Definition Language)方法

面向对象的数据库设计方法

■ UML(Unified Modeling Language)方法

面向对象的建模方法



数据库设计方法（续）

❖ 数据库设计工具

■ SYBASE PowerDesigner

数据库建模—UML工具

■ Rational Rose

UML工具—数据库建模

■ CA ERWin

ERwin全称是ERwin Data Modeler

功能强大、易于使用的数据库建模、数据库设计与开发工具



7.1.3 数据库设计的基本步骤

■ 数据库设计分6个阶段

■ 需求分析

■ 概念结构设计

■ 逻辑结构设计

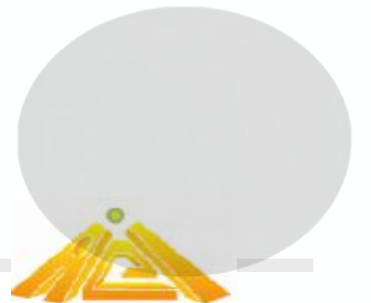
■ 物理结构设计

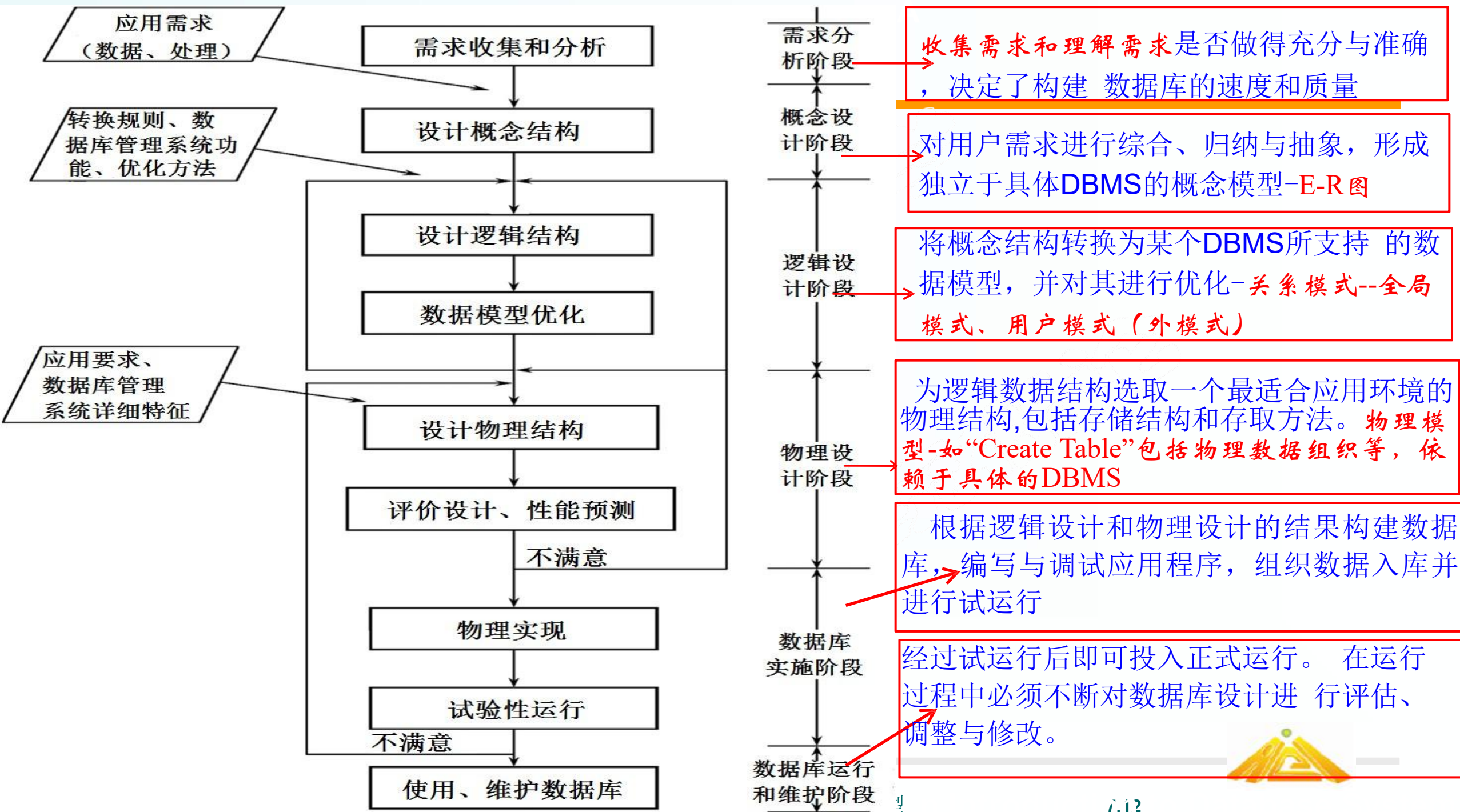
■ 数据库实施

■ 数据库运行和维护

独立于任何数据库管理系统

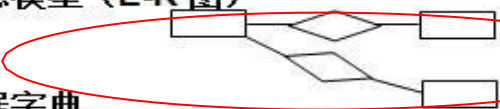
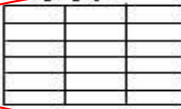
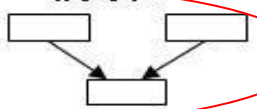

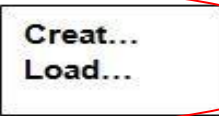
与选用的数据库管理系统密切相关





7.1.3数据库设计的基本步骤（续）

- 设计一个完善的数据库应用系统 往往是上述6个阶段的不断反复。
- 数据库的设计必须与数据处理的设计在每一个设计步骤中紧密结合，相互参照，相互补充

设计阶段	设计描述
需求分析	数据字典、全系统中数据项、数据结构、数据流、数据存储的描述
概念结构设计	概念模型（E-R图）  数据字典
逻辑结构设计	某种数据模型 关系  非关系 
物理结构设计	存储安排 存取方法选择 存取路径建立 
数据库实施	创建数据库模式 装入数据 数据库试运行 
数据库运行和维护	性能监测、转储/恢复、数据库重组和重构

数据库设计各个阶段产生的设计文档/设计说明

数据库设计各个阶段的数据设计描述



数据库设计的基本步骤（续）

■ 参加数据库设计的人员

➤ 系统分析人员和数据库设计人员

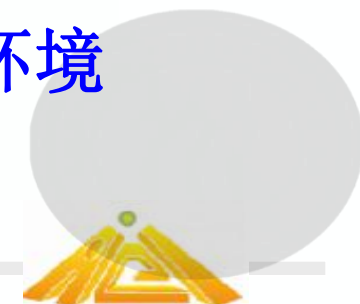
- 自始至终参与数据库设计

➤ 数据库管理员和用户代表

- 主要参加需求分析与数据库的运行和维护

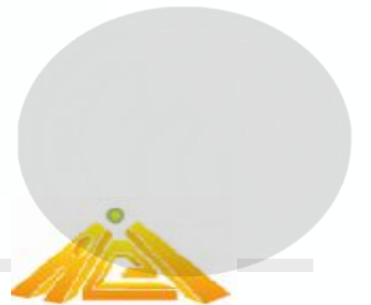
➤ 应用开发人员

- 包括程序员和操作员
- 在实施阶段参与进来，分别负责编制程序和准备软硬件环境



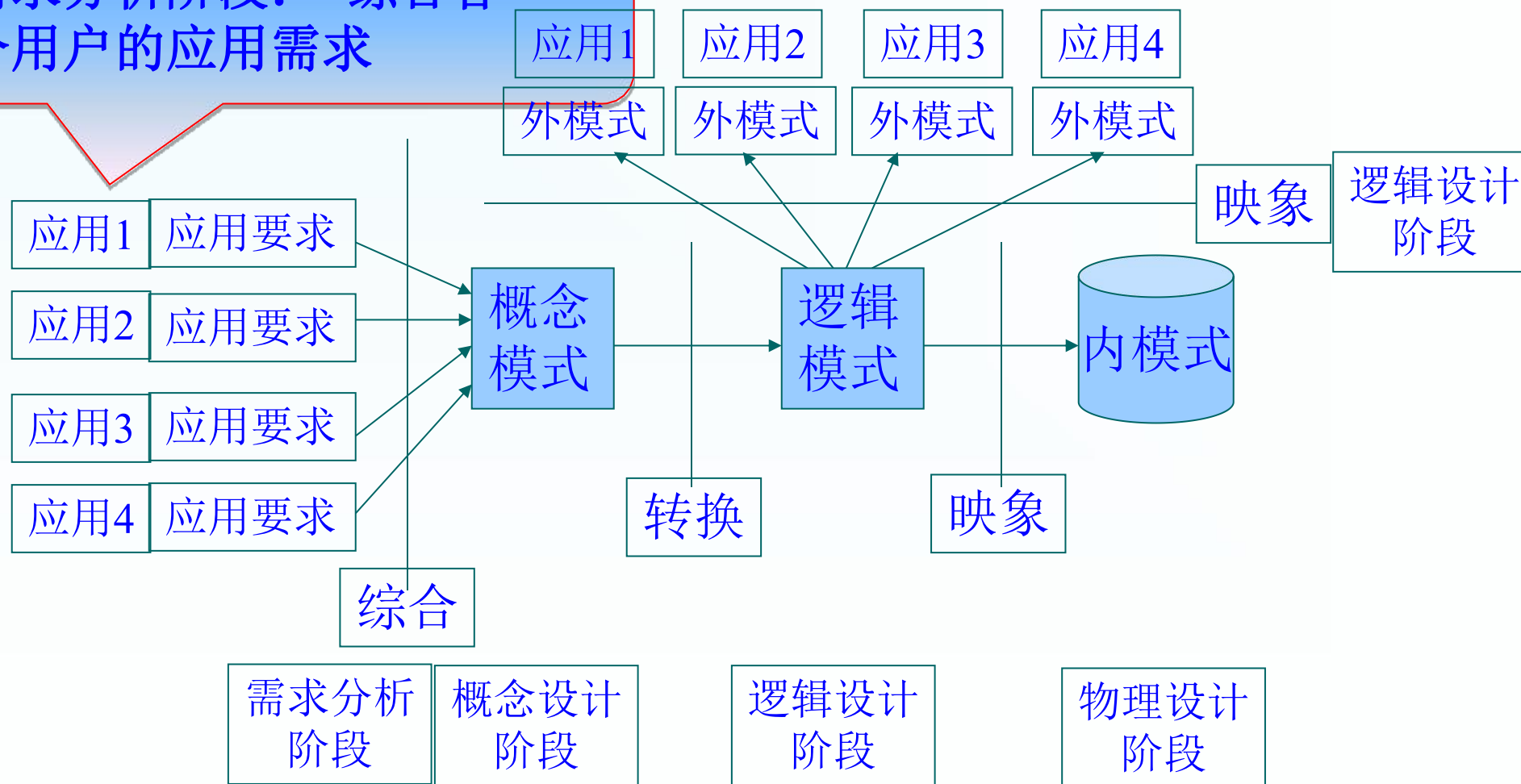
7.1 数据库设计概述

- 7.1.1 数据库设计的特点
- 7.1.2 数据库设计方法
- 7.1.3 数据库设计的基本步骤
- 7.1.4 数据库设计过程中的各级模式



7.1.4 数据库设计过程中的各级模式

需求分析阶段：综合各个用户的应用需求

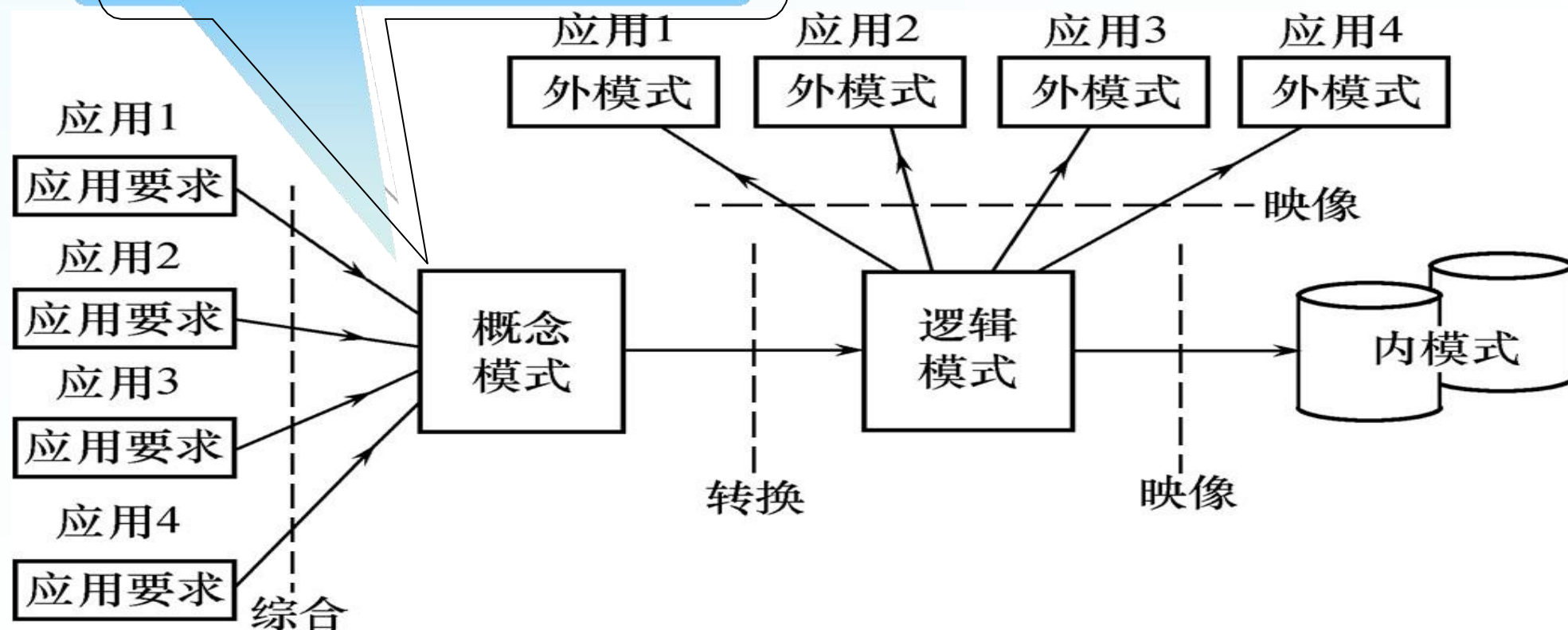


数据库的各级模式



数据库设计过程中的各级模式（续）

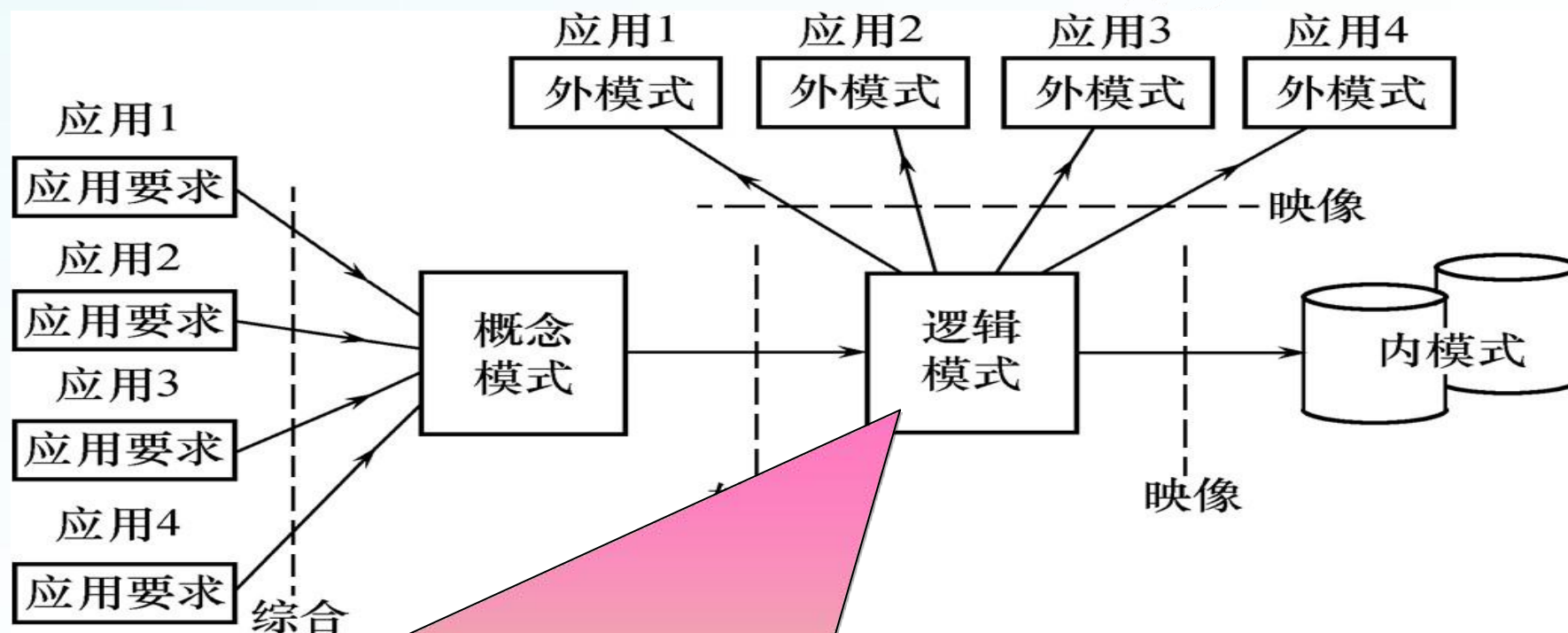
概念设计阶段：形成独立于机器特点，独立于各个 DBMS 产品的**概念模式**（E-R图）



数据库的各级模式



数据库设计过程中的各级模式（续）

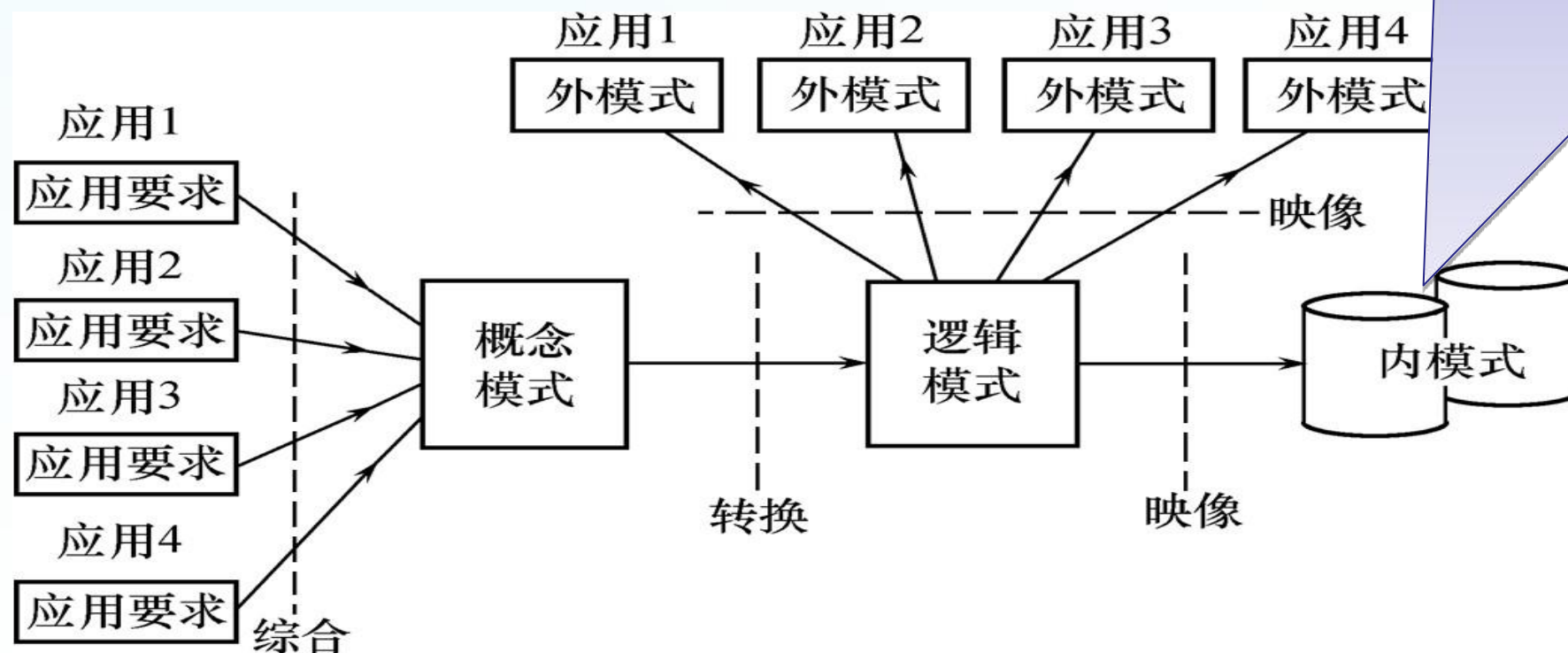


逻辑设计阶段:

1. 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**
2. 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**

数据库设计过程中的各级模式（续）

物理设计阶段：根据数据库管理系统特点和处理的需要，进行物理存储安排，建立索引，形成数据库内模式



数据库的各级模式



7.1 数据库设计概述回顾

❖ 数据库设计的特点

- 数据库建设的基本规律——三分技术，七分管理，十二分基础数据
- 将数据库结构设计和数据处理设计密切结合

❖ 数据库设计方法

❖ 数据库设计的基本步骤

- 数据库设计分6个阶段

❖ 数据库设计过程中的各级模式

- 数据库设计不同阶段形成了数据库的概念模式、模式、外模式、内模式



第 7 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 物理结构设计

7.6 数据库的实施和维护

7.7 小结



7.2 数据库设计的基本步骤

一.数据库设计的准备工作

选定参加设计的人员

1. 数据库分析设计人员
2. 用户
3. 程序员
4. 操作员



7.2 需求分析

二、数据库设计的过程(六个阶段)

1.需求分析阶段--起点

- 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

■ 需求分析常常被忽视

- 设计人员认为这是软任务，急于进行具体设计
- 用户嫌麻烦
- 领导不重视



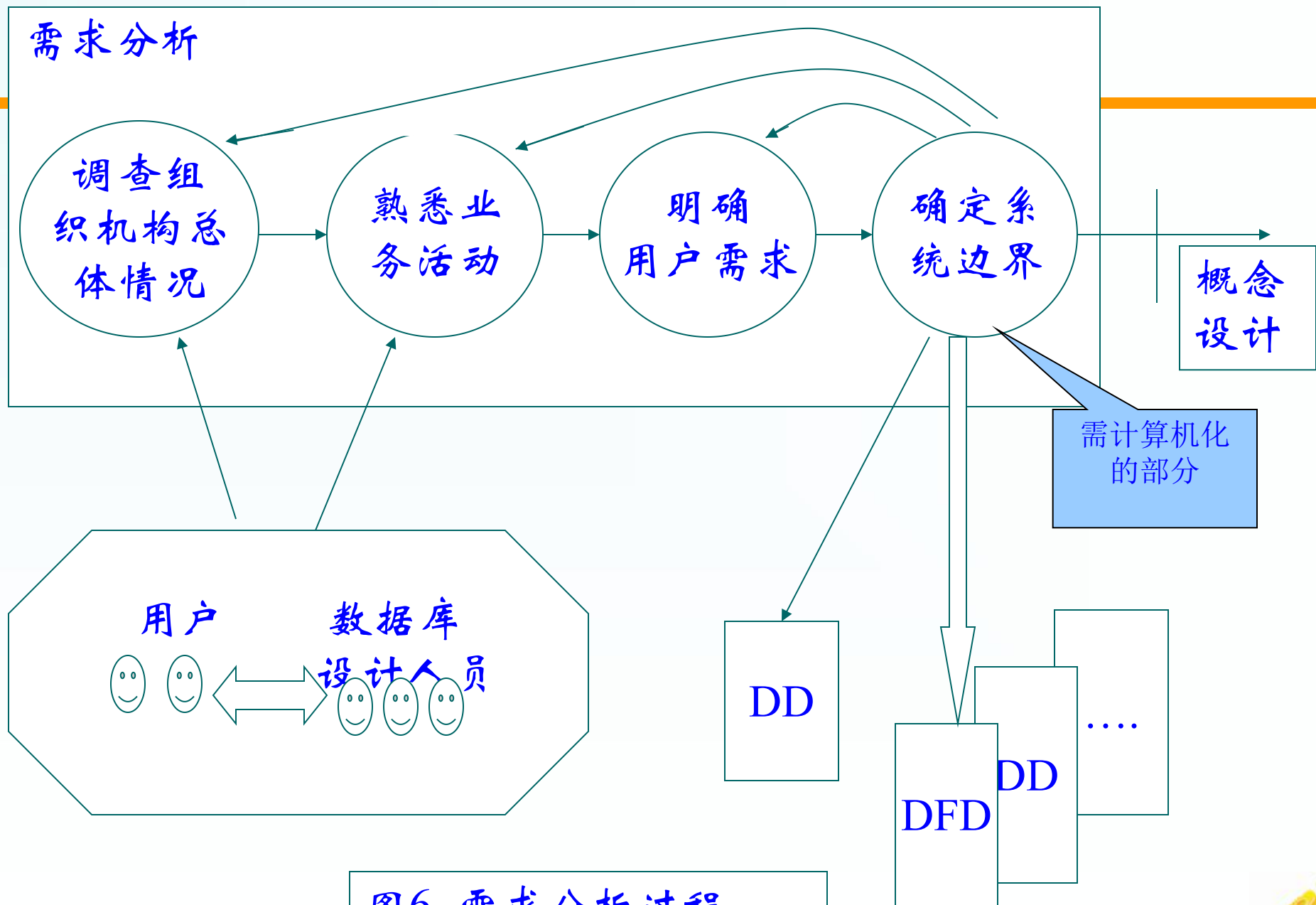


图6 需求分析过程



7.2 数据库需求分析（1）

■需求分析的三个步骤

- ✓通过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，不能仅仅按当前应用需求来设计数据库
- ✓需求分析的重点是调查、收集与分析用户在数据管理中的信息要求、处理要求、安全性与完整性要求。



需求调查的方法:

- **跟班作业** 参加业务工作来了解业务活动的情况，此种方法可以准确地了解用户的需求，但是比较耗费时间
- **开会调查** 通过与用户座谈来了解业务活动情况，座谈时，参加者之间可以相互启发
- **请专人介绍**
- **询问**
- **设计调查表请用户填写。** 如果调查表设计的合理，这种方法是很有效，也易于为用户接受。
- **查阅记录**

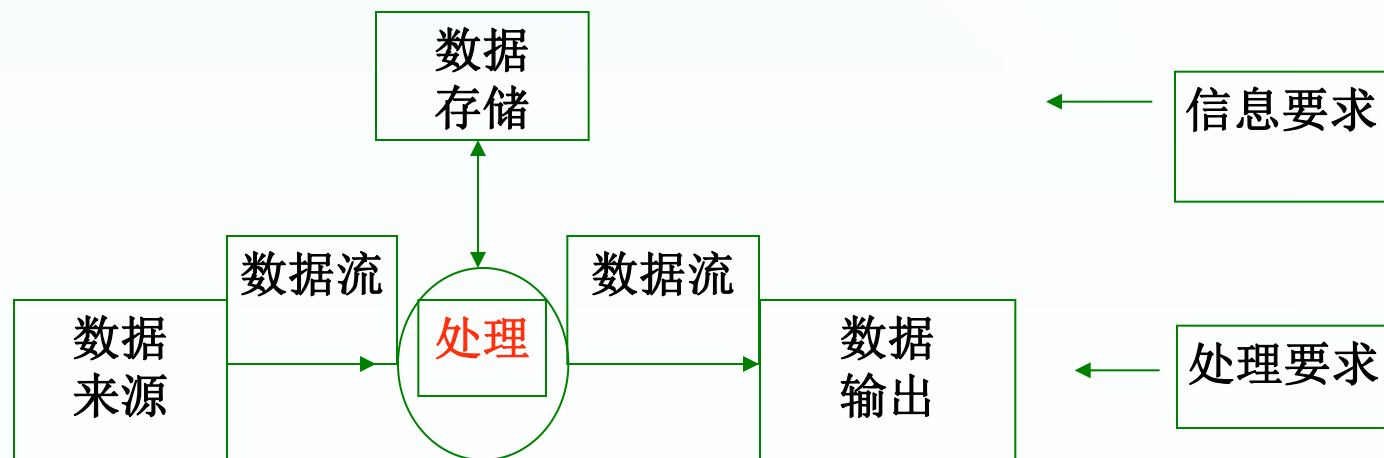
做需求调查时，往往需要同时采用上述多种方法。但是无论采用何种方法，都需要用户的配合。



7.2 数据库需求分析（2）

■需求分析的方法

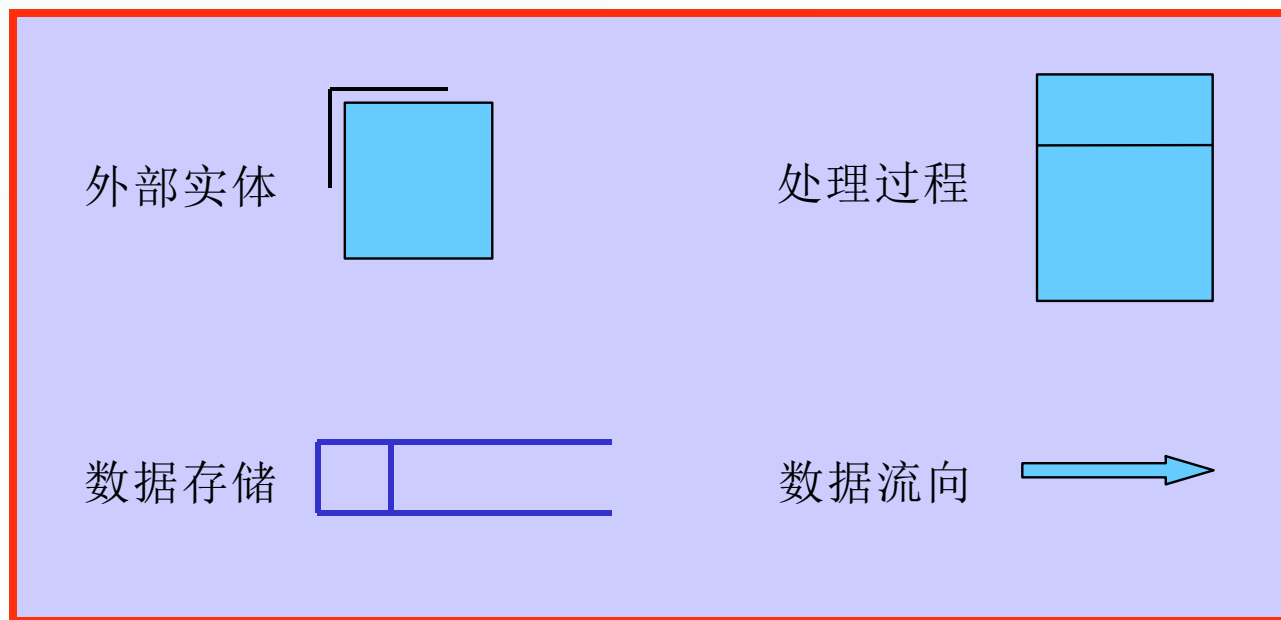
- 自顶向下的结构化分析方法（简称SA方法）
- 该方法从最上层的系统组织机构入手，采用逐层分解的方式分析系统，并用数据流图（DFD）和数据字典（DD）描述系统。
- 首先任何一个系统都抽象为数据流图



7.2 数据库需求分析（3）

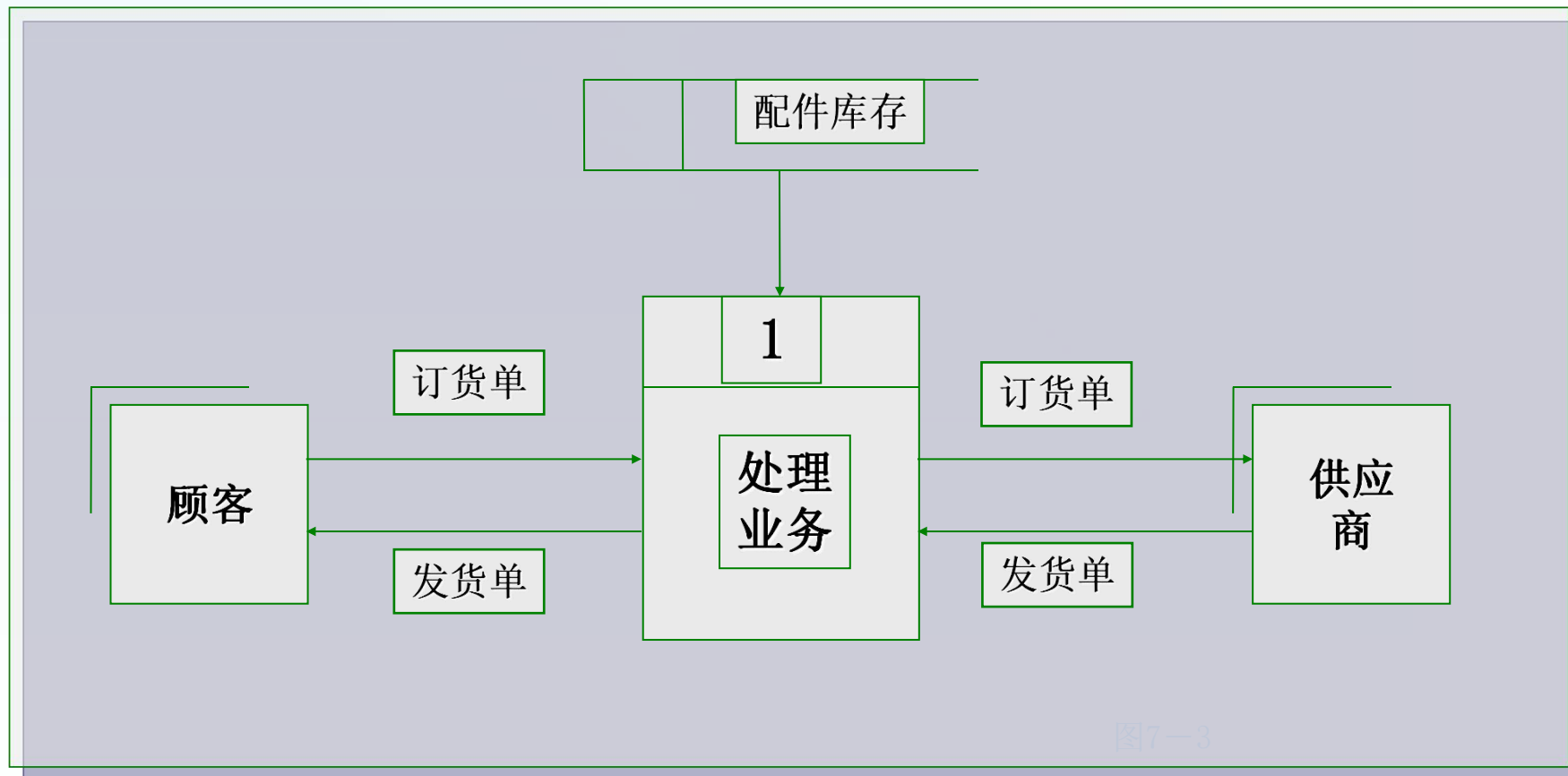
★数据流图是一种最常用的结构化分析工具，它从数据传递和加工角度，以图形的方式刻画系统内的数据运动情况。

★数据流图的基本符号



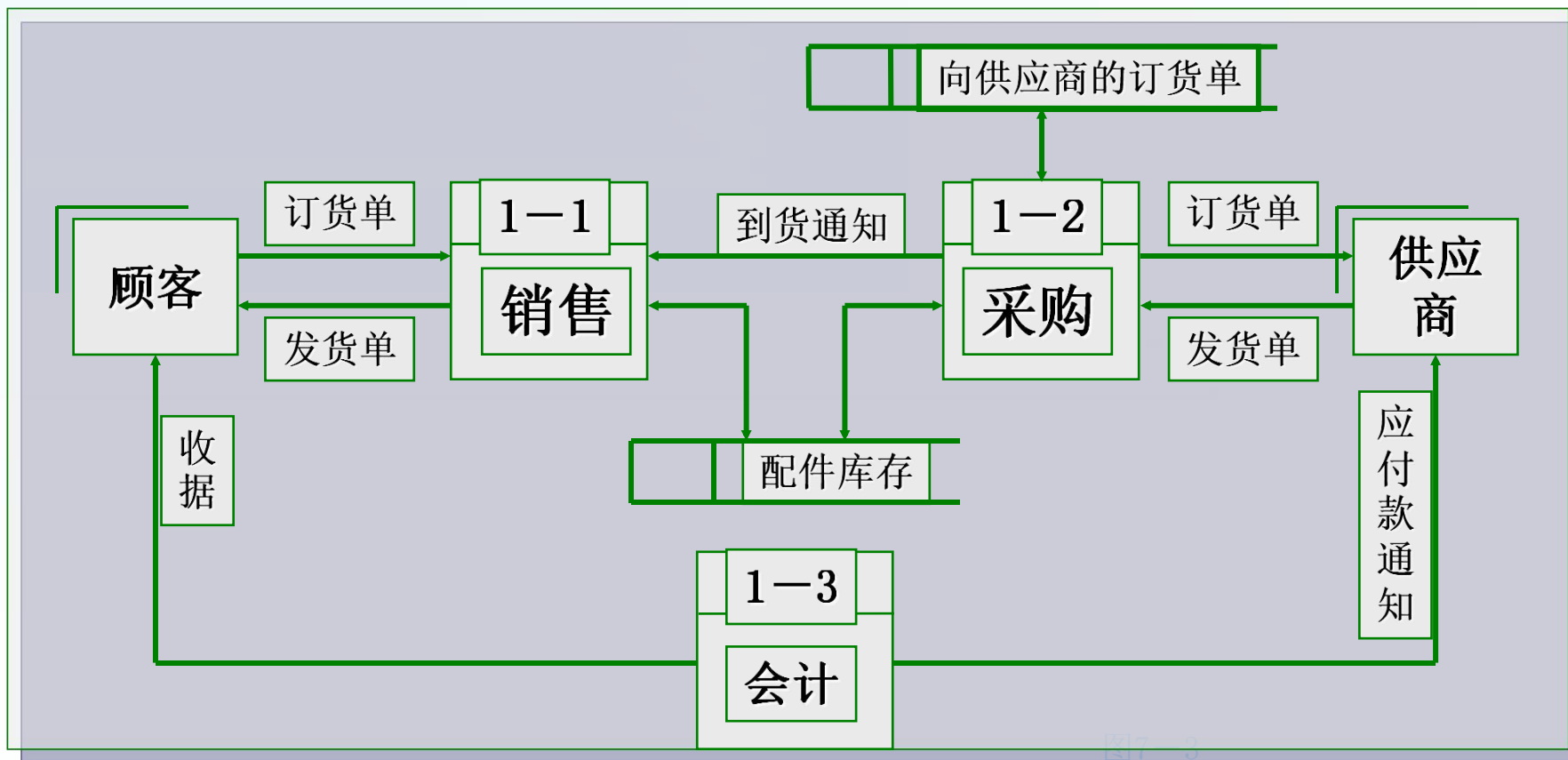
7.2 数据库需求分析（4）

汽车配件公司：第一层数据流程图



7.2 数据库需求分析（5）

汽车配件公司：第二层数据流程图



7.2 数据库需求分析（6）

★数据字典（Data Dictionary，简称DD）

✓ 什么是数据字典

- 数据流图上所有成分的定义和解释的文字集合就是数据字典。
- 数据词典和数据流图密切配合，能清楚地表达数据处理的要求。数据词典用于对数据流图中出现的所有成分给出定义，它使数据流图上的数据流名字、加工名字和数据存贮名字具有确切的解释。每一条解释就是一个词条，按一定的顺序将所有词条排列起来，就构成了数据词典，就象日常使用的英汉词典、新华词典一样。



7.2 数据库需求分析（7）

✓数据字典的内容

– 数据项： 数据项是数据的最小单位。

数据项描述= {数据项名, 数据项含义说明, 别名, 数据类型, 长度, 取值范围, 取值含义, 与其它数据项的逻辑关系, 数据项之间的联系 }

数据结构： 数据结构是数据项有意义的集合。反应了数据之间的组合关系。

数据结构描述= { 数据结构名, 含义说明, 组成: {数据项或数据结构}

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。

数据流描述={数据流名, 说明, 数据流来源, 数据流去向, 组成: {数据结构}, 平均流量, 高峰期流量 }



7.2 数据库需求分析（7）

✓数据字典的内容

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。

数据存储描述={ 数据存储名，说明，编号，输入的数据流，输出的数据流，组成：{数据结构}，数据量，存取频度，存取方式 }

— 处理过程：处理过程的具体处理逻辑一般用判定表或判定树来描述。

处理过程描述= { 处理过程名，说明，输入：{ 数据流}，输出：{ 数据流 }，处理：{ 简要说明 } }

简要说明：说明该处理过程的功能及处理要求，功能是指该处理过程用来干什么处理要求包括处理频度要求，如单位时间里处理多少事务、多少、的数据量、响应时间。



第 7 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 物理结构设计

7.6 数据库的实施和维护

7.7 小结



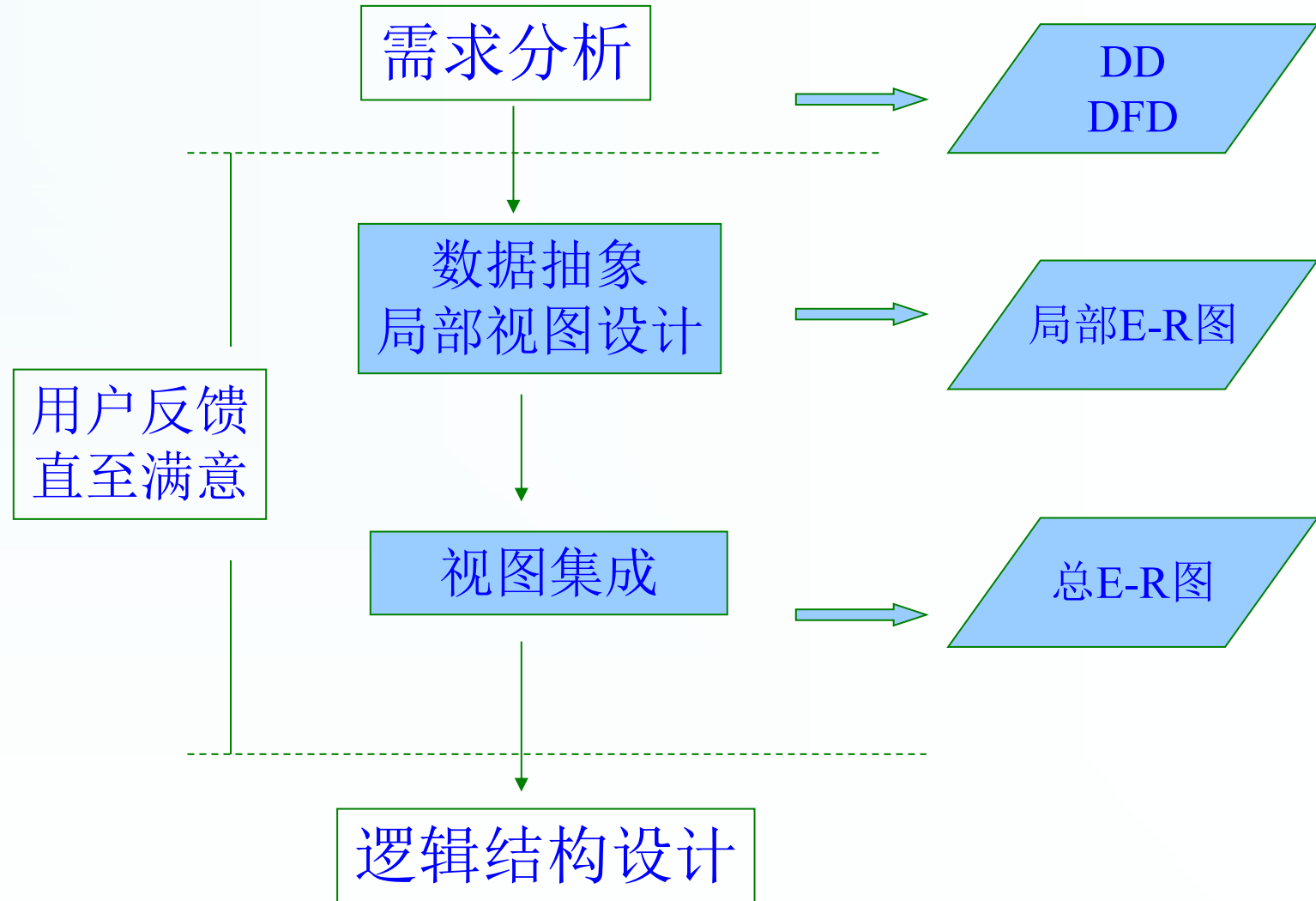
7.3 概念结构设计（1）

■ 1. 什么是概念结构设计

★ 将需求分析得到的用户需求抽象为**概念模型**的过程就是概念结构设计

★ 概念模型是不依赖于计算机和具体的DBMS，从而更加稳定。

★ 概念结构设计是整个数据库设计的关键



7.3 概念结构设计（2）

■2. 概念结构设计的特点

- ★ **有丰富的语义表达能力** 能真实、充分地反映现实世界，包括事物和事物之间的联系，能满足用户对数据的处理要求。是对现实世界的一个真实模型。
- ★ **易于交流和理解** 从而可以用它和不熟悉计算机的用户交换意见，用户的积极参与是数据库的设计成功的关键。
- ★ **易于更改** 当应用环境和应用要求发生变化时，能很容易对概念模型进行修改，以反映这些变化。
- ★ **易于向各种数据模型转换** 易于导出与DBMS有关的逻辑模型。



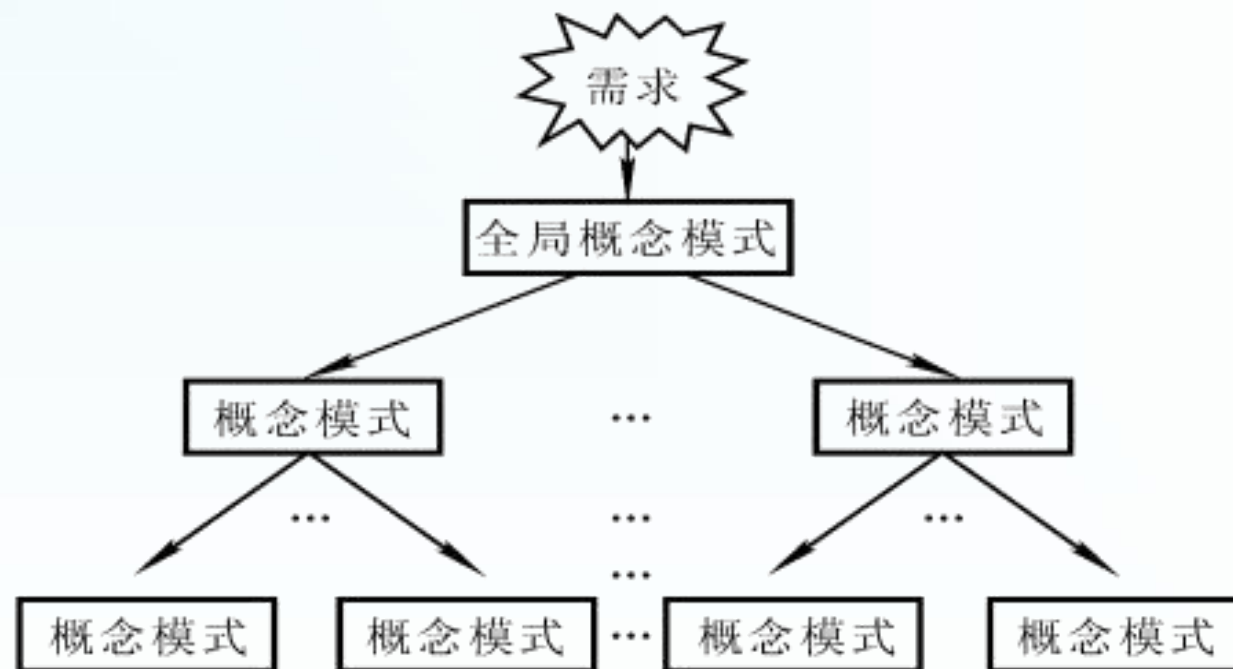
7.3 概念结构设计（3）

■3. 概念结构设计的策略

- **自底向上**。先定义每个局部应用的概念结构，然后按一定的规则把它们集成起来，从而得到全局概念模型。
- **自顶向下**：先定义全局概念模型，然后再逐步细化。
- **逐步扩张**：先定义最重要的核心概念结构，然后再逐步向外扩展。以滚雪球的方式逐步生成其它概念结构，直至总体概念结构。
- **混合策略**：将自顶向下和自底向上结合起来使用。先用自顶向下设计一个概念结构的框架，然后以它为框架再用自底向上设计局部概念结构，并把它们集成。
- 最常用的设计策略是**自底向上策略**。



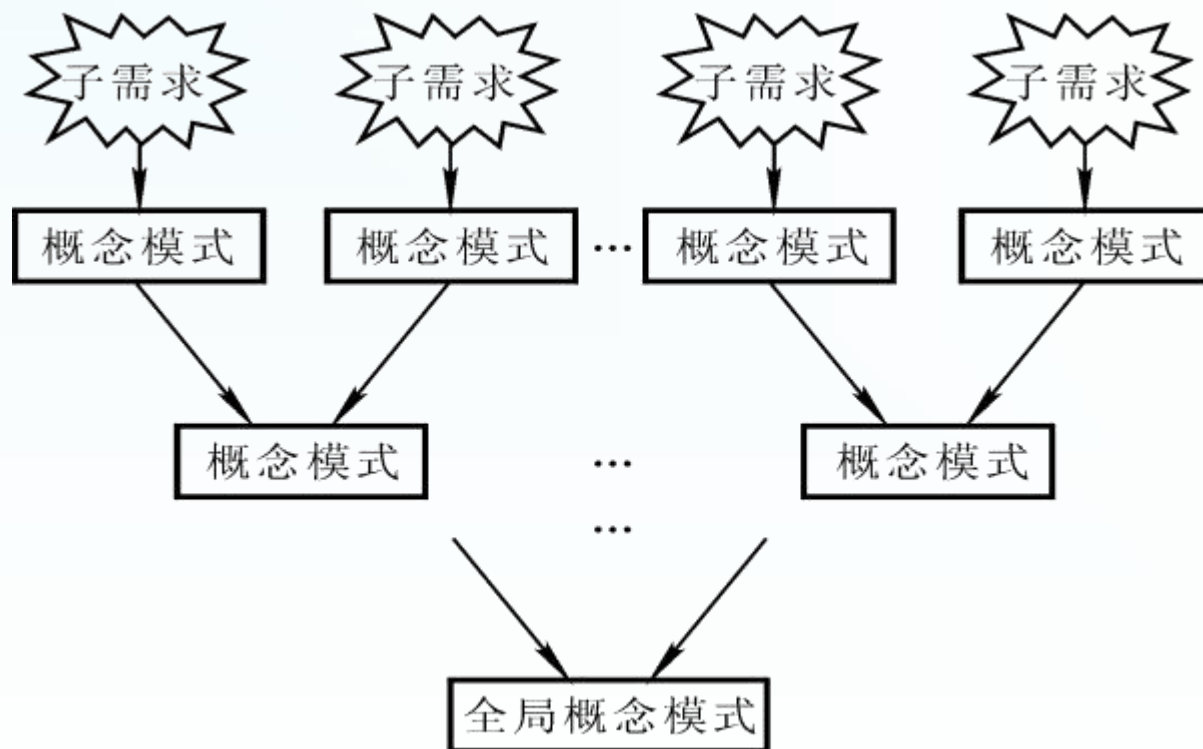
7.3 概念结构设计（4）



自顶向下策略



7.3 概念结构设计（5）



自底向上策略（常用）



7.3 概念结构设计（6）

■4. 采用E-R模型方法的概念结构设计

采用E-R方法的概念结构设计可分为如下三步：

- （1）设计局部E-R模型 局部E-R模型的设计内容包括确定局部E-R模型的范围、定义实体、联系以及它们的属性。
- （2）设计全局E-R模型 这一步是将所有局部E-R图集成为一个全局E-R图，即全局E-R模型。
- （3）优化全局E-R模型



7.3 概念结构设计（7）

■ (1) 设计局部E-R模型

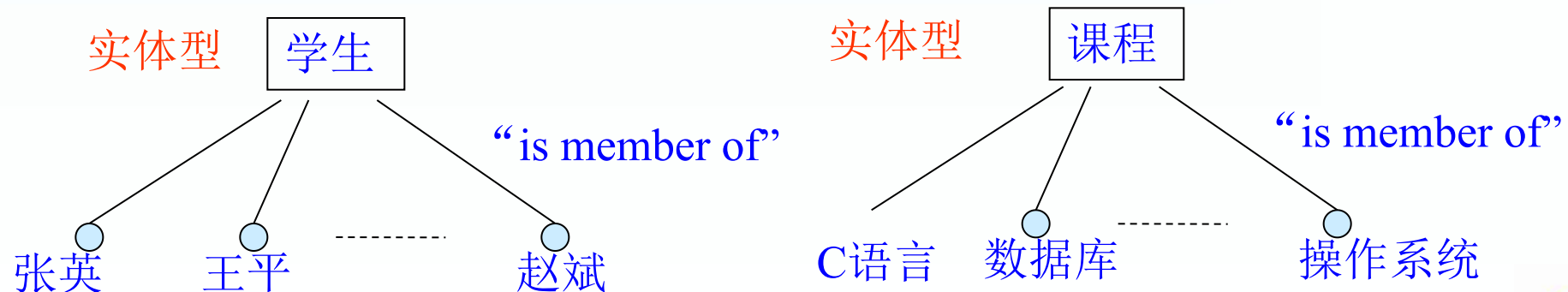
- 设计局部E-R模型的关键就是正确划分实体和属性。
- 将现实世界中的事物进行数据抽象，可以得到实体和属性。
- ★ 三种常用抽象：

- ✓ 分类 (Classification)

定义某一类概念作为现实世界中一组对象的类型

这些对象具有某些共同的特性和行为

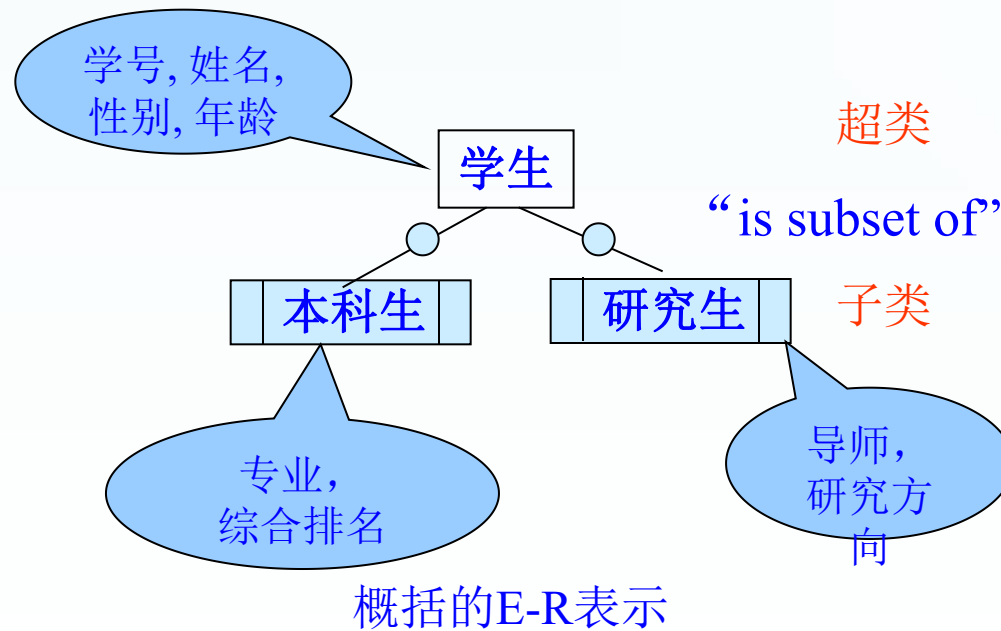
它抽象的是对象值和型之间的“is member of”的语义



7.3 概念结构设计（8）

✓ 概括 (Generalization)

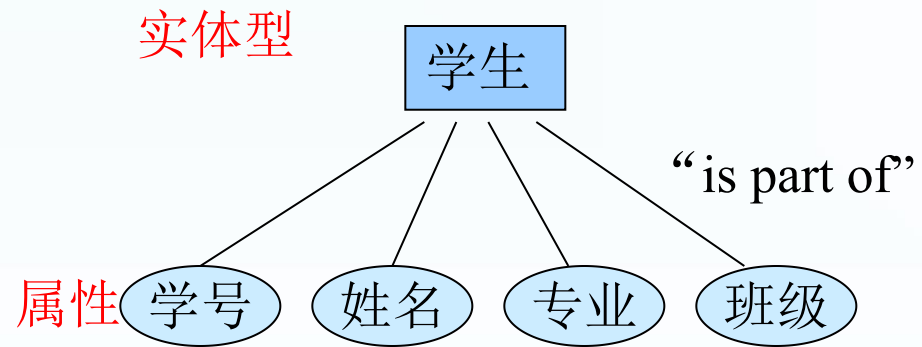
定义类型之间的一种子集联系，它抽象了类型之间的“is subset of”的语义
概括有一个很重要的性质：继承性。子类继承超类上定义的所有抽象。



7.3 概念结构设计 (9)

✓ 聚集 (Aggregation)

定义某一类型的组成成分 它抽象了对象内部类型和成分之间
“is part of”的语义

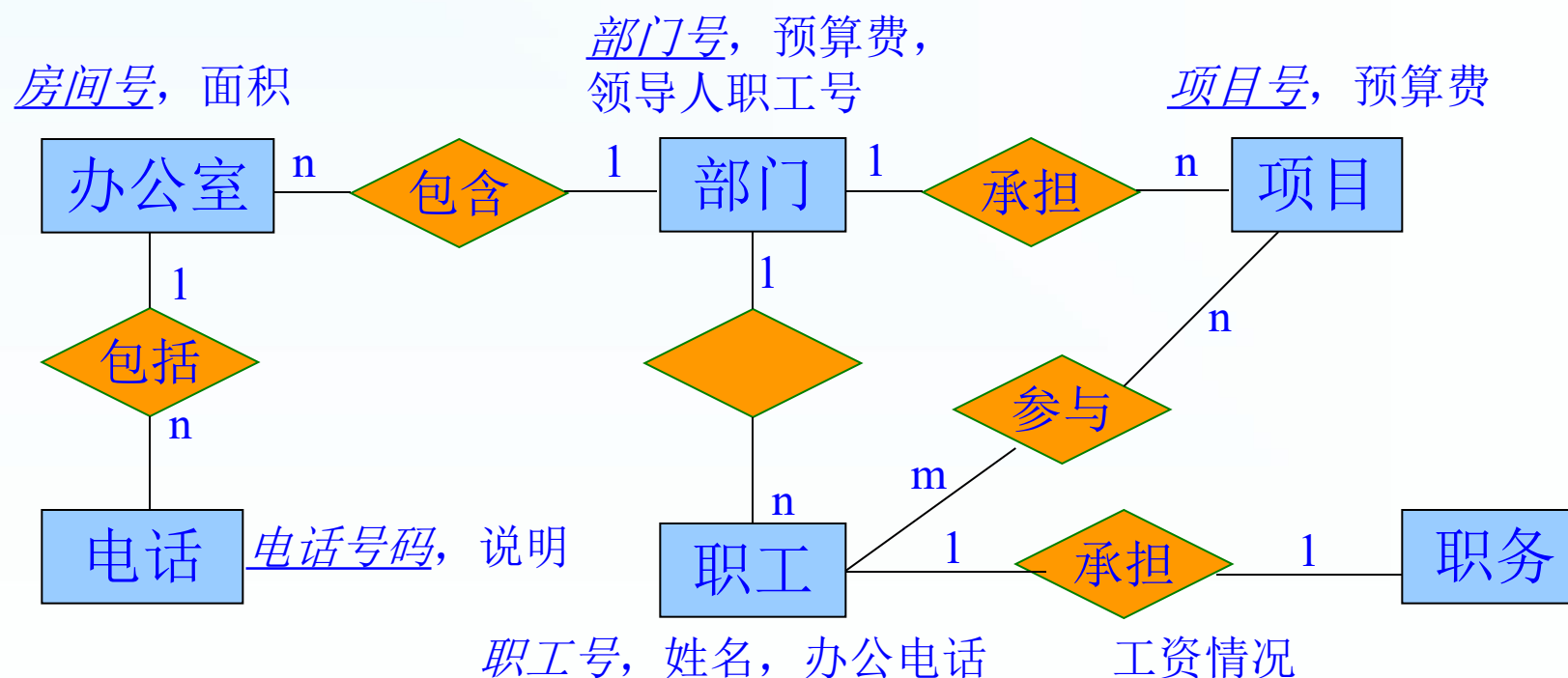


- 设计局部E-R模型就是利用抽象机制对需求分析阶段收集的数据进行分析，标定 局部应用中的实体、属性、码，实体间的联系, 设计分E-R图



7.3 概念结构设计（10）

如公司管理系统的局部E-R图



7.3 概念结构设计（11）

■ (2) 设计全局E-R模型

方法：

- **一次集成：**一次性将多个局部图合并为一个全局E-R图
- **逐步累积式：**首先集成两个局部E-R图（通常是比较关键的两个局部E-R图）以后每次将一个新的E-R视图集成进来
- 当将局部E-R图集成为全局E-R图时，需要消除各分E-R图合并时产生的冲突。



7.3 概念结构设计（12）

关键：合理消除各局部E-R图合并时产生的冲突

属性冲突

命名冲突

结构冲突

a. 属性值域冲突，即属性值的类型、取值范围或取值集合不同。

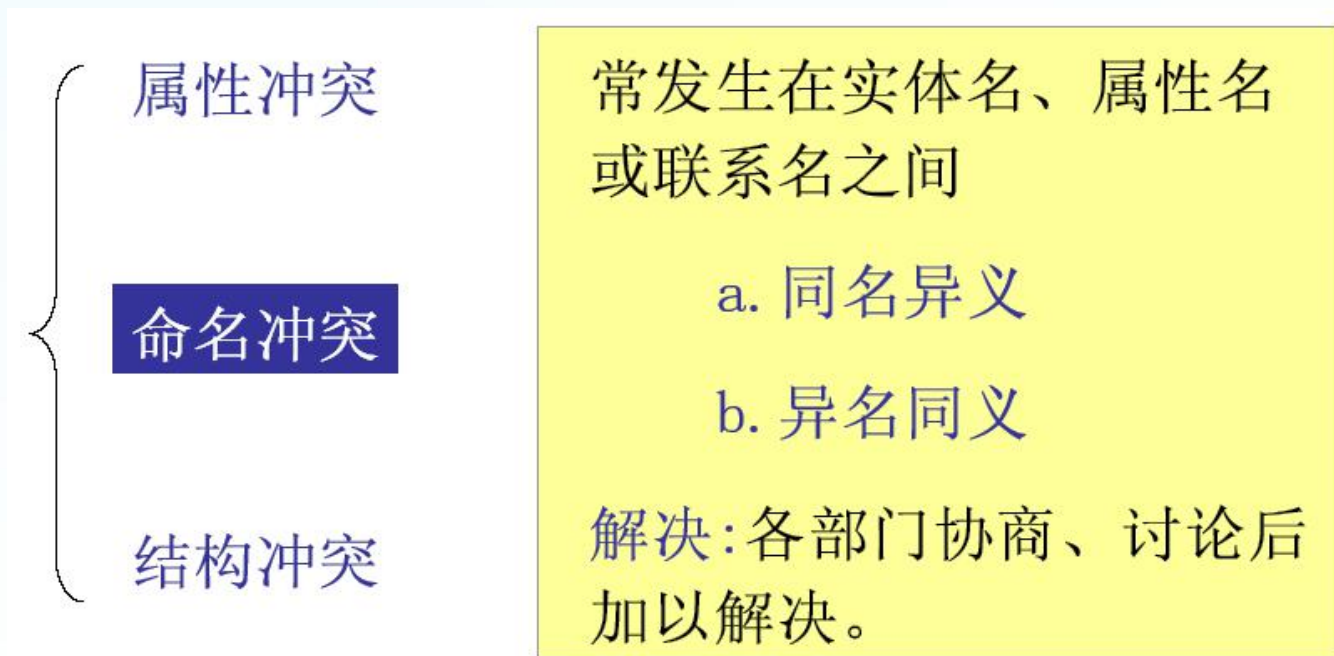
b. 属性的取值单位冲突。

解决：与用户协商后解决。



7.3 概念结构设计（13）

关键：合理消除各局部E-R图合并时产生的冲突



7.3 概念结构设计（14）

关键：合理消除各局部E-R图合并时产生的冲突

属性冲突

命名冲突

结构冲突

a. 同一对象在不同应用中有不同的抽象。

b. 同一实体在不同应用中属性组成不同。

c. 同一联系在不同应用中呈现不同的类型。



7.3 概念结构设计（15）

■ (3) 优化全局E-R模型

一个好的全局E-R模型除了能反映用户功能需求外，还应满足如下条件：

- 实体个数尽可能少；
- 实体所包含的属性尽可能少；
- 实体间联系无冗余。

优化的目的就是要满足上述三个条件。

消除冗余的方法：

1. 分析方法：即以数据字典和数据流图为依据，根据数据字典中关于数据项之间的逻辑关系的说明来消除冗余。
2. 用规范化理论中函数依赖的概念来消除冗余

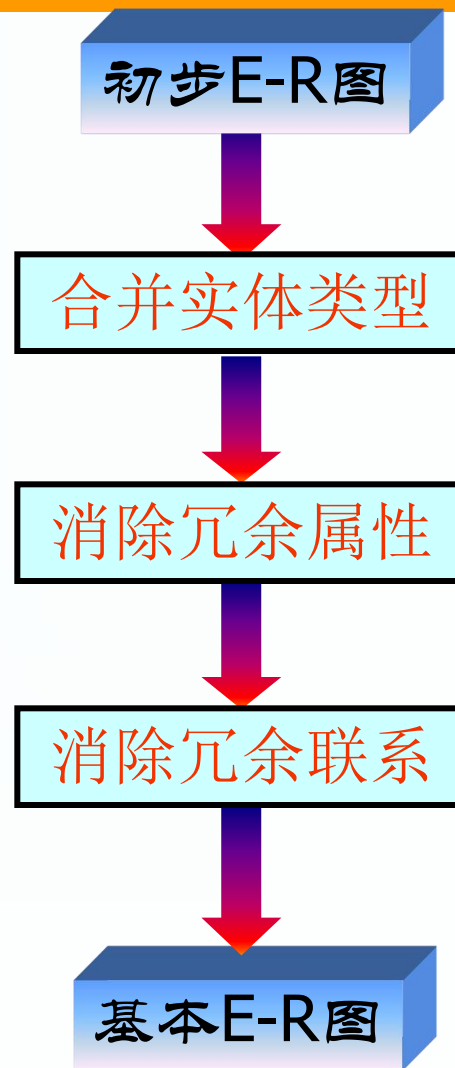


7.3 概念结构设计（16）

■ (3) 优化全局E-R模型

方法:一般是把具有相同主码的实体进行合并,消除冗余的属性(数据)和冗余联系

- 并不是所有的冗余数据与冗余联系都必须加以消除,有时为了提高某些应用的效率,不得不以冗余信息作为代价。
- 设计数据库概念结构时,哪些冗余信息必须消除,哪些冗余信息允许存在,需要根据用户的整体需求来确定。



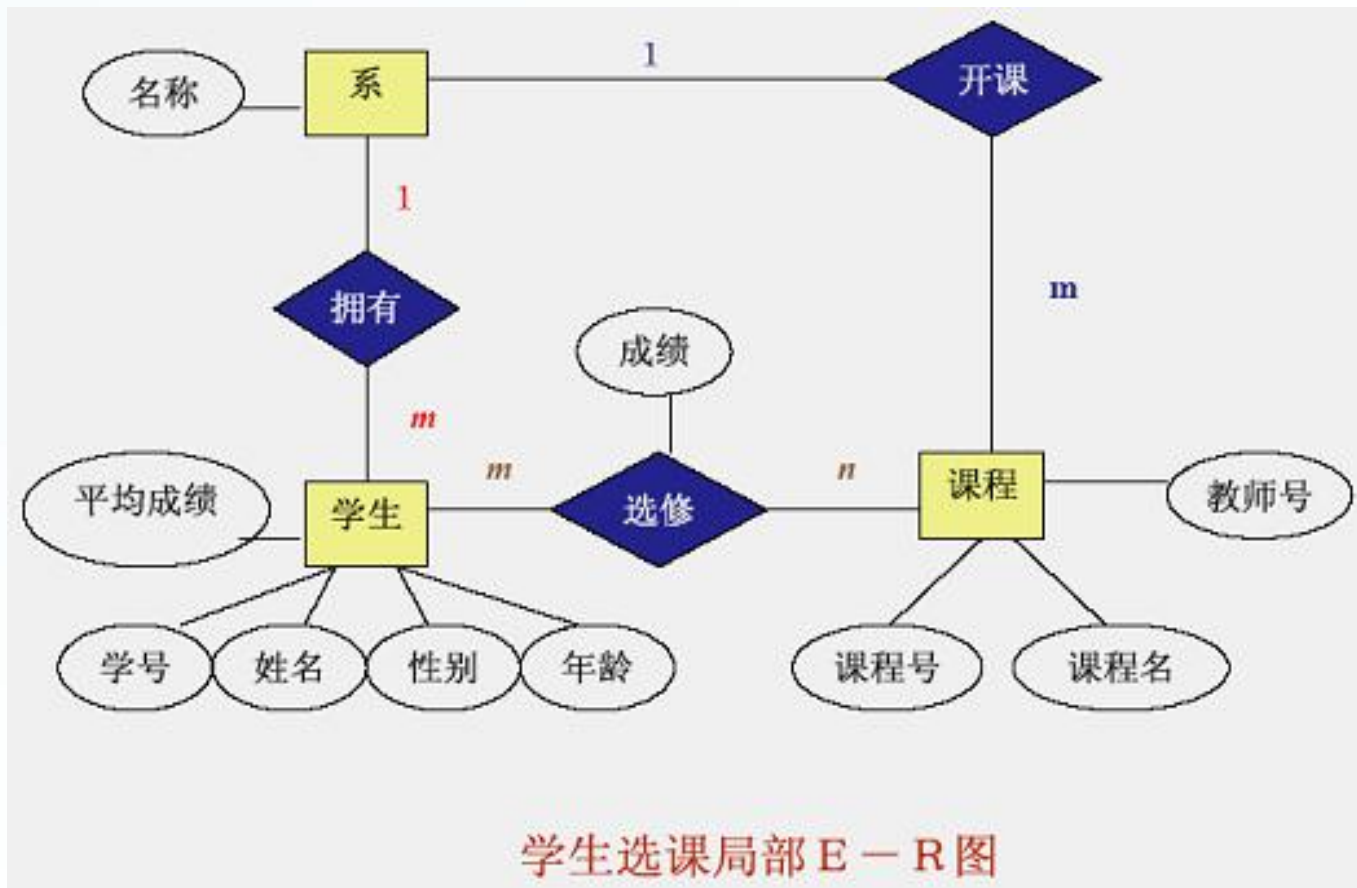
举例:教务管理系统E-R图设计

■ 设计教务系统管理中得到局部E-R图

(1) 一个学生可选修多门课程, 一门课程可为多个学生选修 (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 (教师和课程是多对多的联系);

(3) 一个系有多个教师 (学生), 一个教师 (学生) 只能属于一个系 (系和教师、学生是一对多的联系)



举例：教务管理系统E-R图设计

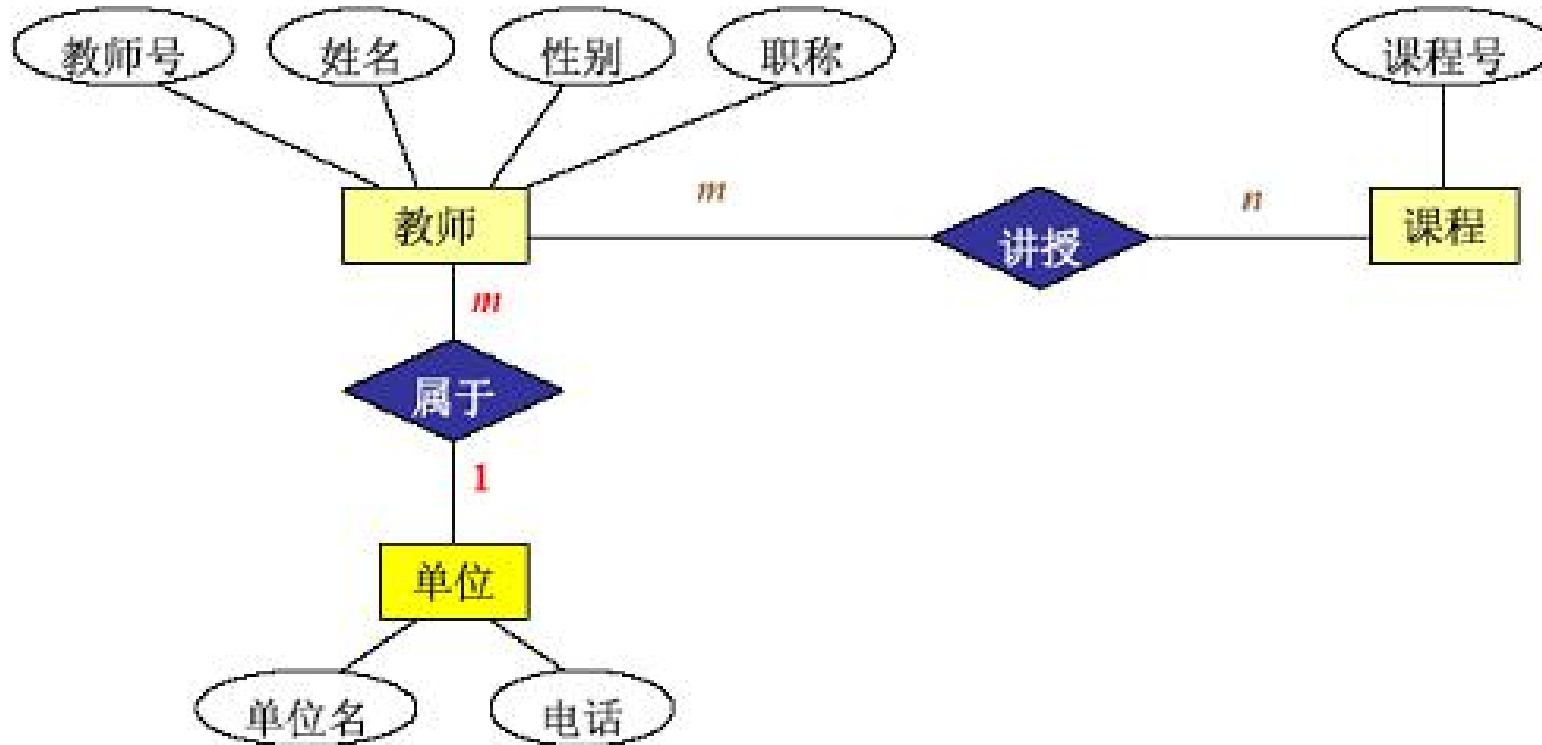


图 教师任课局部E-R图



举例:教务管理系统E-R图设计

■ 步骤一:生成初步ER图

(1)命名冲突:学生图中的实体“系”与教师图中的实体“单位”,都是指“系”,即同名同义

解决:合并后统一为“系名”。

(2)结构冲突:实体“系”和实体“课程”在两个不同应用中的属性组成不同。

解决:合并后的组成为原来局部E-R图中的同名实体属性的并集

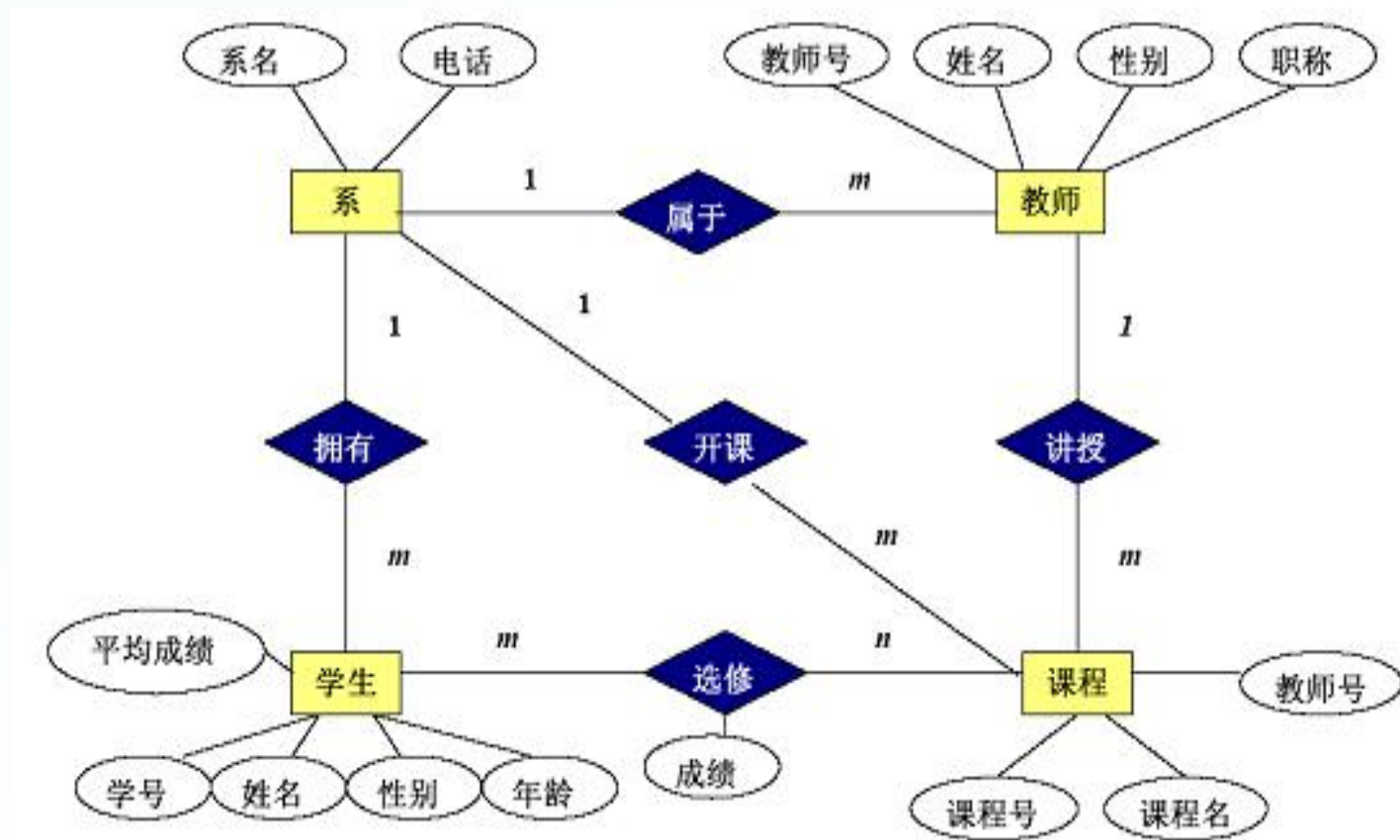


图 教务管理系统的初步E-R图



举例:教务管理系统E-R图设计

■ 步骤二:消除冗余,生成基本E-R图

(1)冗余数据

“课程”实体中的属性“教师号”,可由“讲授”这个教师与课程之间的联系导出。

学生的平均成绩,可由“选修”联系中的属性“成绩”计算出来

(2)冗余联系

“系”和“课程”之间的联系“开课”可由“系”和“教师”之间的“属于”联系与“教师”和“课程”之间的“讲授”联系推导出来。

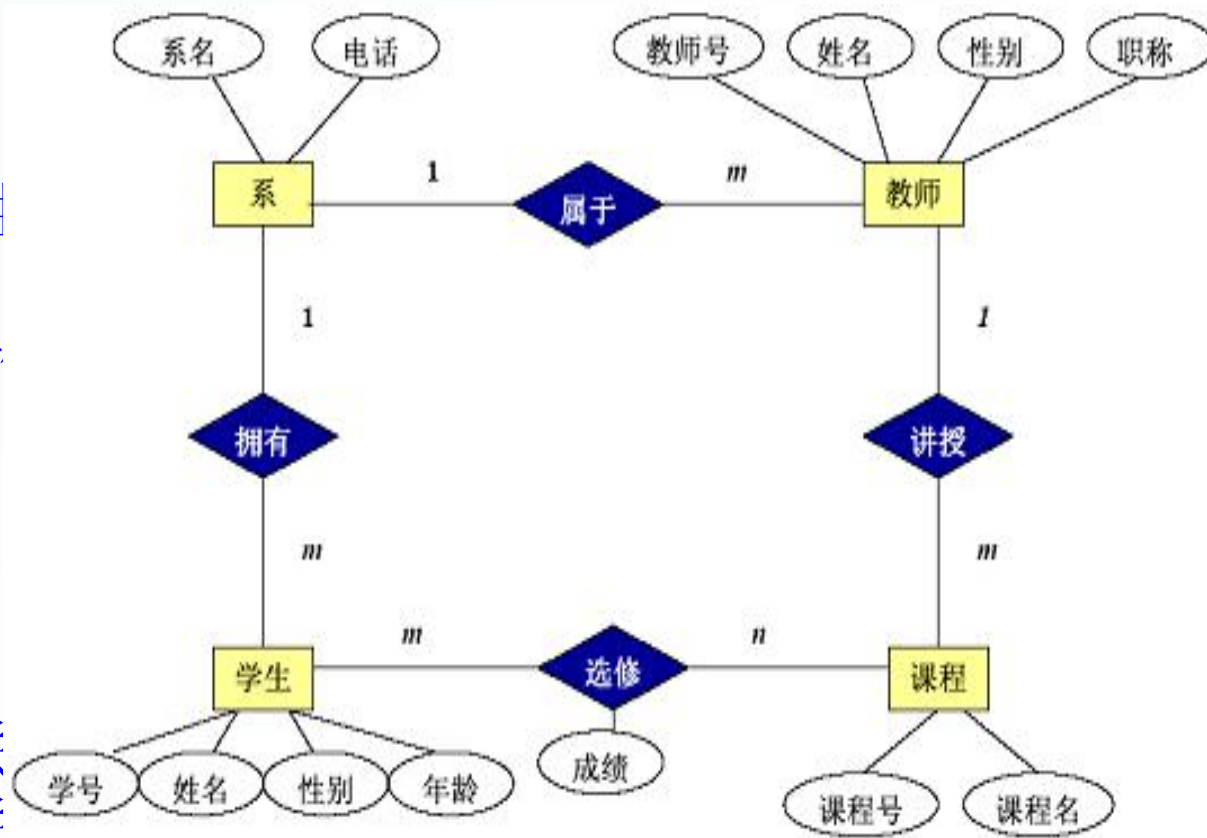


图 教务管理系统的基本E-R图

图7-3



下课了。。



休息一会儿。。。

