

数据库系统概论

An Introduction to Database System

第七章 数据库设计

中国人民大学信息学院

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 物理结构设计

7.6 数据库的实施和维护

7.7 小结



7.1 数据库设计概述

❖ 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（设计）优化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操作要求。
- 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。



数据库设计概述（续）

❖ 数据库设计

- 数据库设计的目标是为用户和各种应用系统提供一个**信息基础设施**和**高效率的运行环境**。
- 高效率的运行环境
 - 数据库数据的存取效率高
 - 数据库存储空间的利用率高
 - 数据库系统运行管理的效率高



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



7.1.1 数据库设计的特点

1. 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

- 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



数据库设计的特点（续）

2. 结构（数据）设计和行为（处理）设计相结合

- 将数据库结构设计和数据处理设计密切结合

❖ 结构和行为分离的设计

- 传统的软件工程：重 行为设计

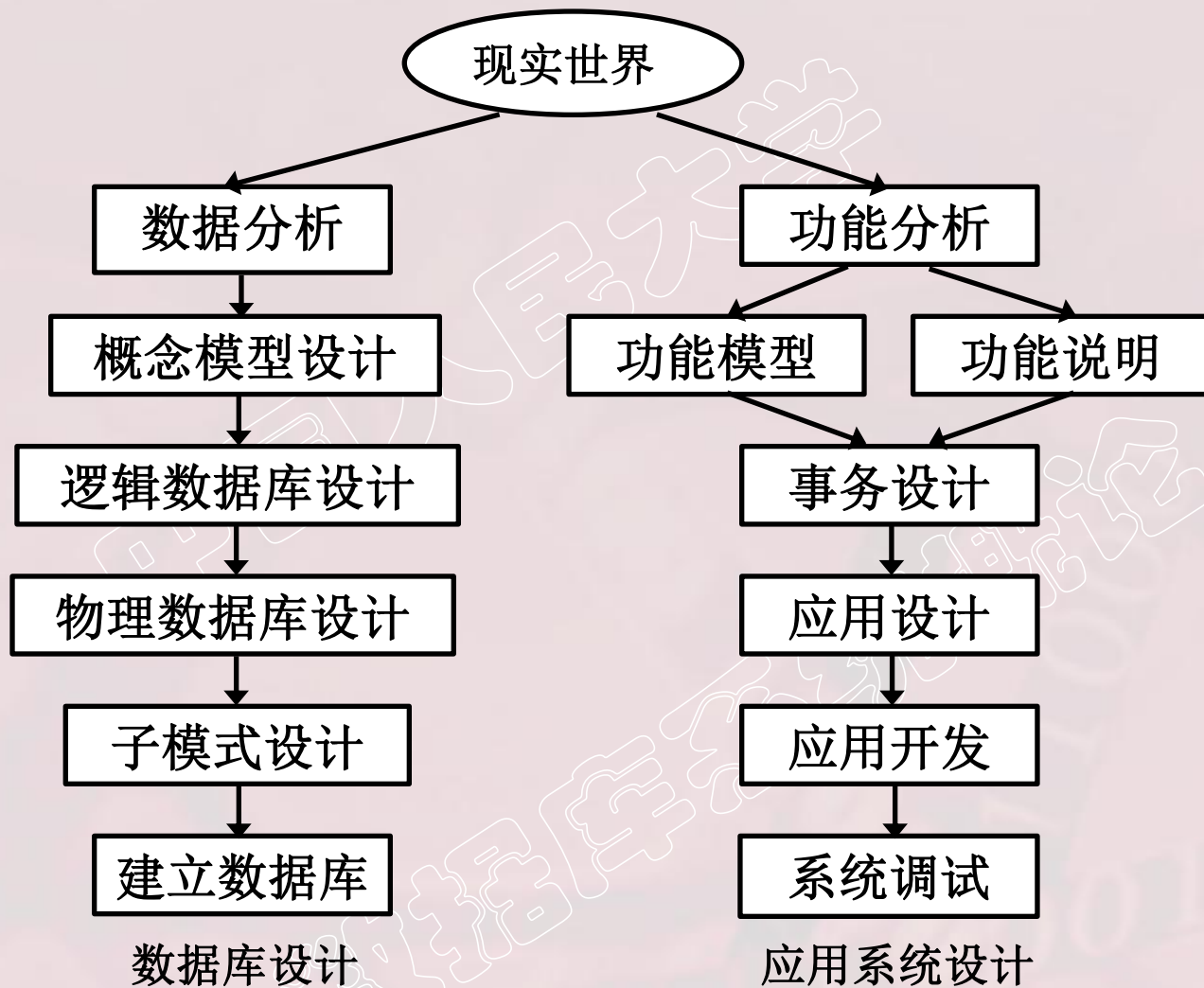
- 忽视对应用中数据语义的分析和抽象，只要有可能就尽量推迟数据库结构设计的决策

- 早期的数据库设计：重 结构设计

- 致力于数据模型和数据库建模方法研究，忽视了行为设计对结构设计的影响



数据库设计的特点（续）



结构和行为分离的设计



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



7.1.2 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。
- ❖ 它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



数据库设计方法（续）

❖ 手工试凑法

- 设计质量与设计人员的经验和水平有直接关系
- 缺乏科学理论和工程方法的支持，工程的质量难以保证
- 数据库运行一段时间后常常又不同程度地发现各种问题，增加了维护代价



数据库设计方法（续）

❖ 规范设计法

■ 手工设计方法

■ 基本思想

- 过程迭代和逐步求精

■ 典型方法

- 新奥尔良（**New Orleans**）方法
- 基于**E-R**模型的数据库设计方法
- **3NF**（第三范式）的设计方法
- 面向对象的数据库设计方法
- 统一建模语言（**UML**）方法



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



7.1.3 数据库设计的基本步骤

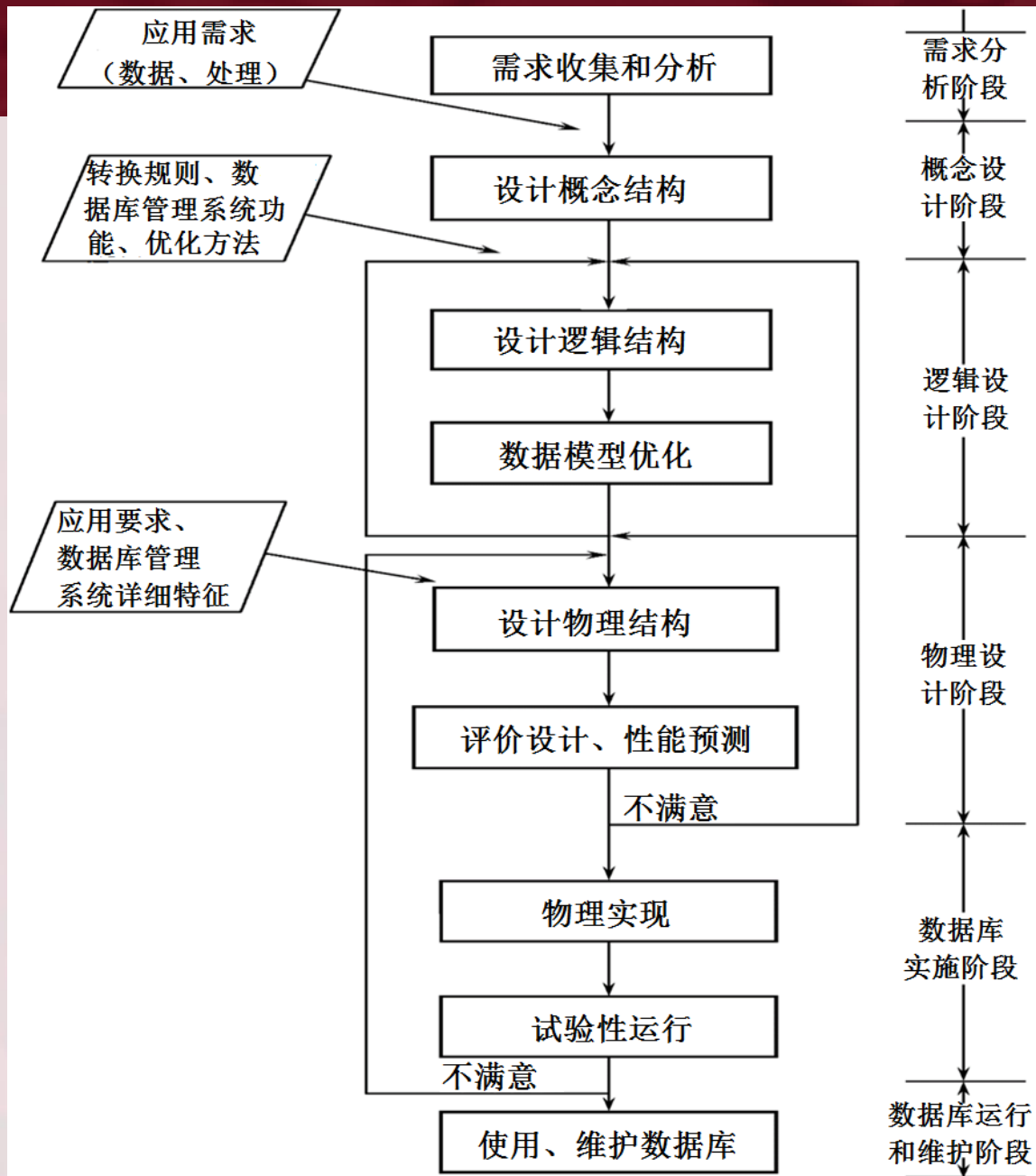
❖ 数据库设计分6个阶段

- 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

❖ 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数据库管理系统密切相关





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，其水平决定了数据库系统的质量

■ 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



数据库设计的基本步骤（续）

1. 需求分析阶段

- 是否做得充分与准确，决定了构建数据库的速度和质量

2. 概念结构设计阶段

- 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数据库管理系统的概念模型

3. 逻辑结构设计阶段

- 将概念结构转换为某个数据库管理系统所支持的数据模型，并对其进行优化



数据库设计的基本步骤（续）

4. 物理结构设计阶段

- 为逻辑数据结构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

5. 数据库实施阶段

- 根据逻辑设计和物理设计的结果构建数据库
- 编写与调试应用程序
- 组织数据入库并进行试运行

6. 数据库运行和维护阶段

- 经过试运行后即可投入正式运行
- 在运行过程中必须不断对其进行评估、调整与修改



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
- ❖ 这个设计步骤既是数据库设计的过程，也包括了数据库应用系统的设计过程
- ❖ 把数据库的设计和对数据库中数据处理的设计紧密结合起来，将这两个方面的需求分析、抽象、设计、实现在各个阶段同时进行，相互参照，相互补充，以完善两方面的设计



数据库设计的基本步骤（续）

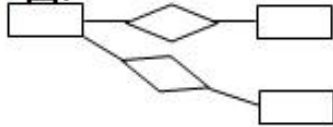
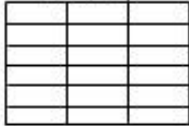
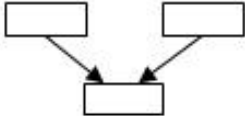


设计阶段	设计描述
需求分析	数字字典、全系统中数据项、数据结构、数据流、数据存储的描述
概念结构设计	概念模型 (E-R 图)  数据字典
逻辑结构设计	某种数据模型 关系  非关系 
物理结构设计	存储安排 存取方法选择 存取路径建立 
数据库实施	创建数据库模式 装入数据 数据库试运行 
数据库运行和维护	性能监测、转储/恢复、数据库重组和重构

图7.3 数据库设计各个阶段的数据设计描述



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

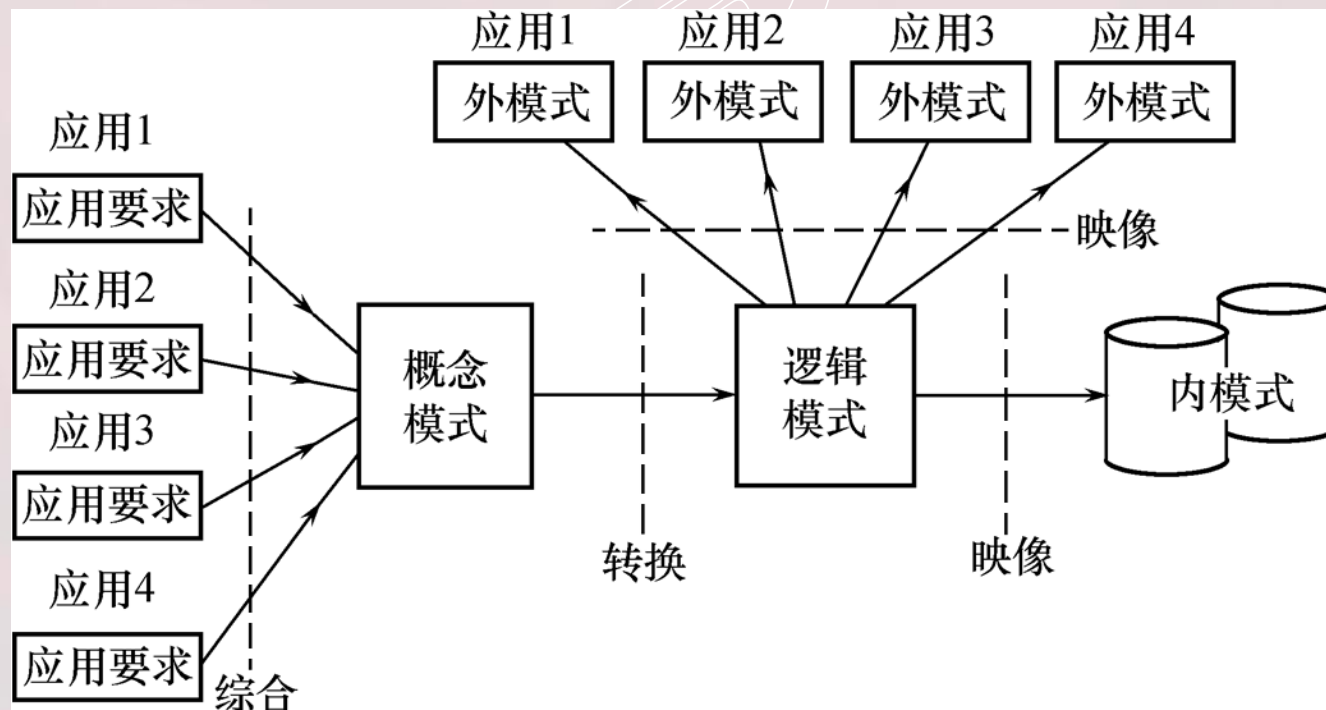
7.1.3 数据库设计的基本步骤

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

❖ 数据库设计不同阶段形成的数据库各级模式



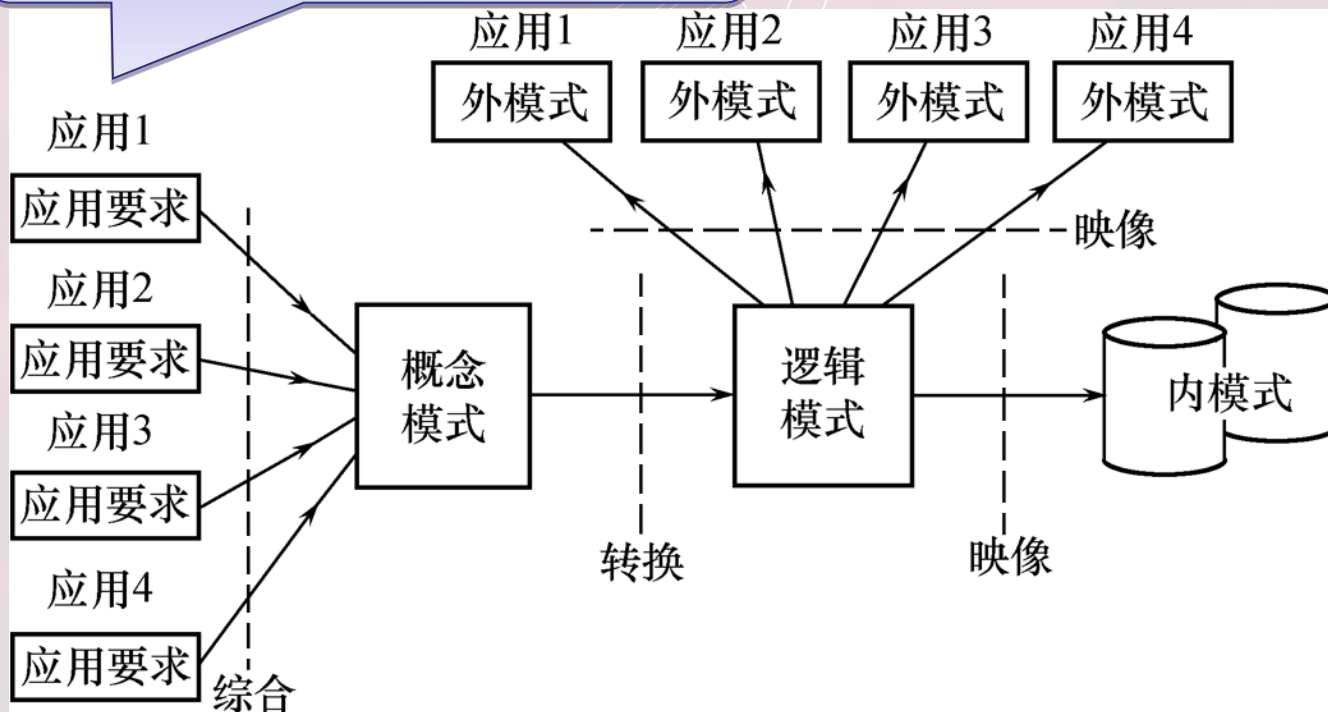
数据库的各级模式



数据库设计过程中的各级模式（续）

❖ 数据库设计不同阶段形成的数据库各级模式

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



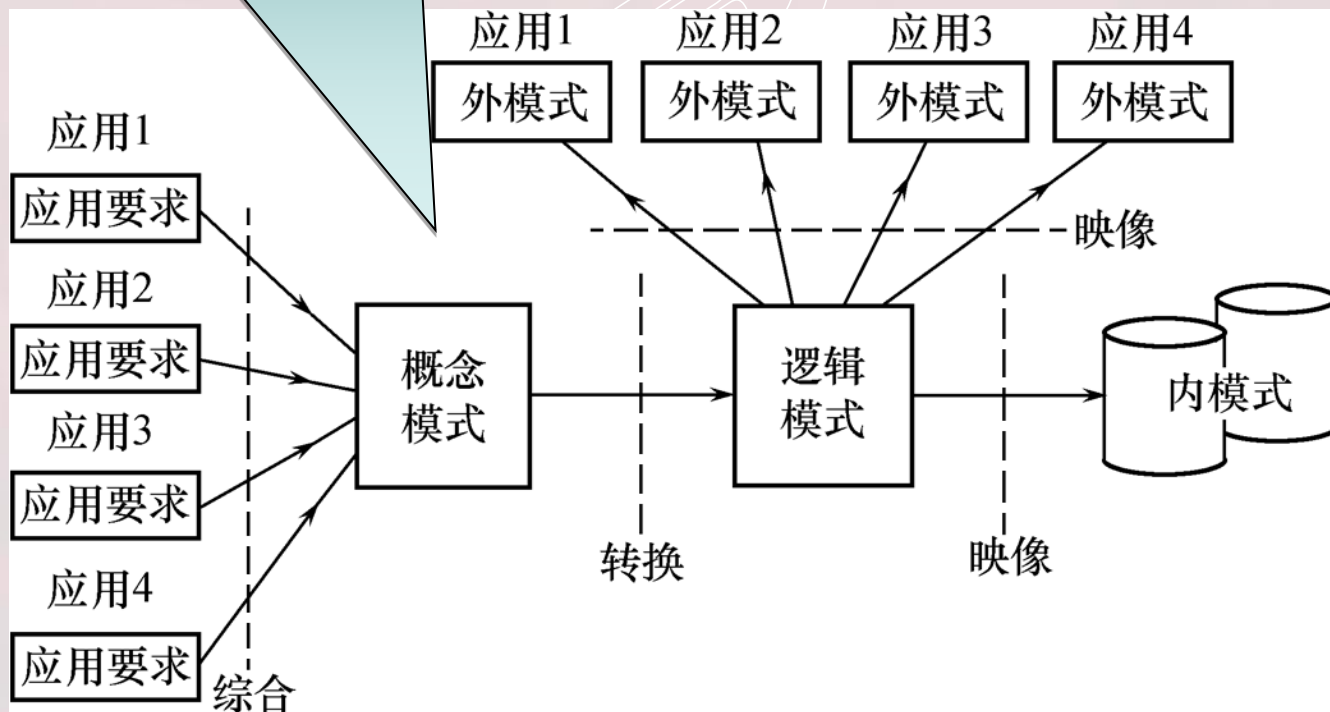
数据库设计过程中的各级模式 (续)

❖ 数据库

概念设计阶段:

形成独立于机器特点, 独立于各个数据库管理系统产品的**概念模式 (E-R图)**

数据库各级模式

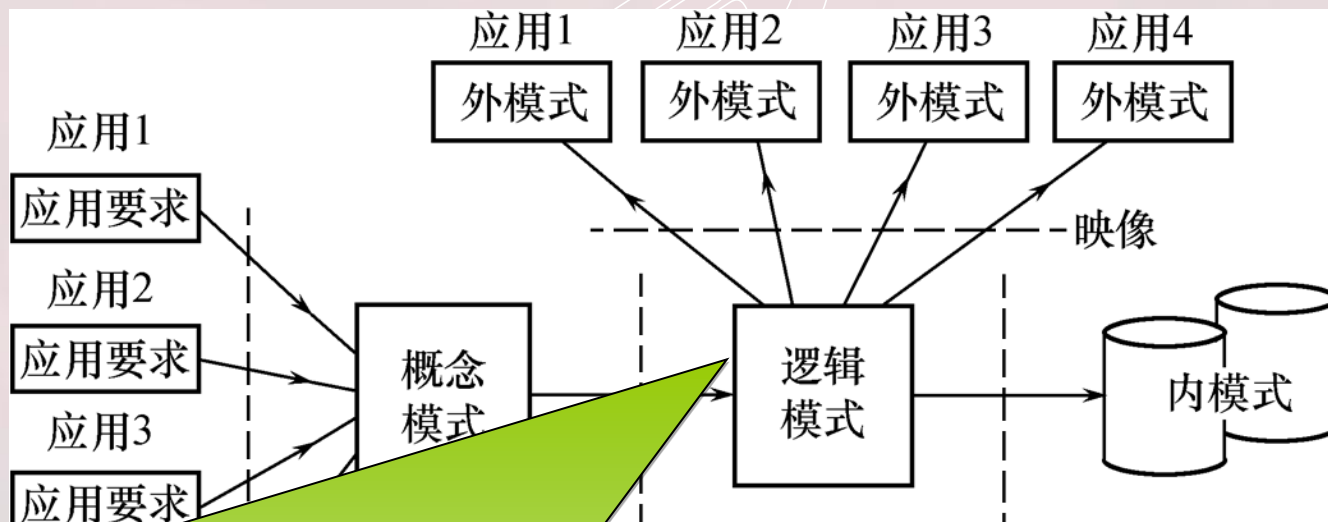


数据库的各级模式



数据库设计过程中的各级模式（续）

❖ 数据库设计不同阶段形成的数据库各级模式



逻辑设计阶段：

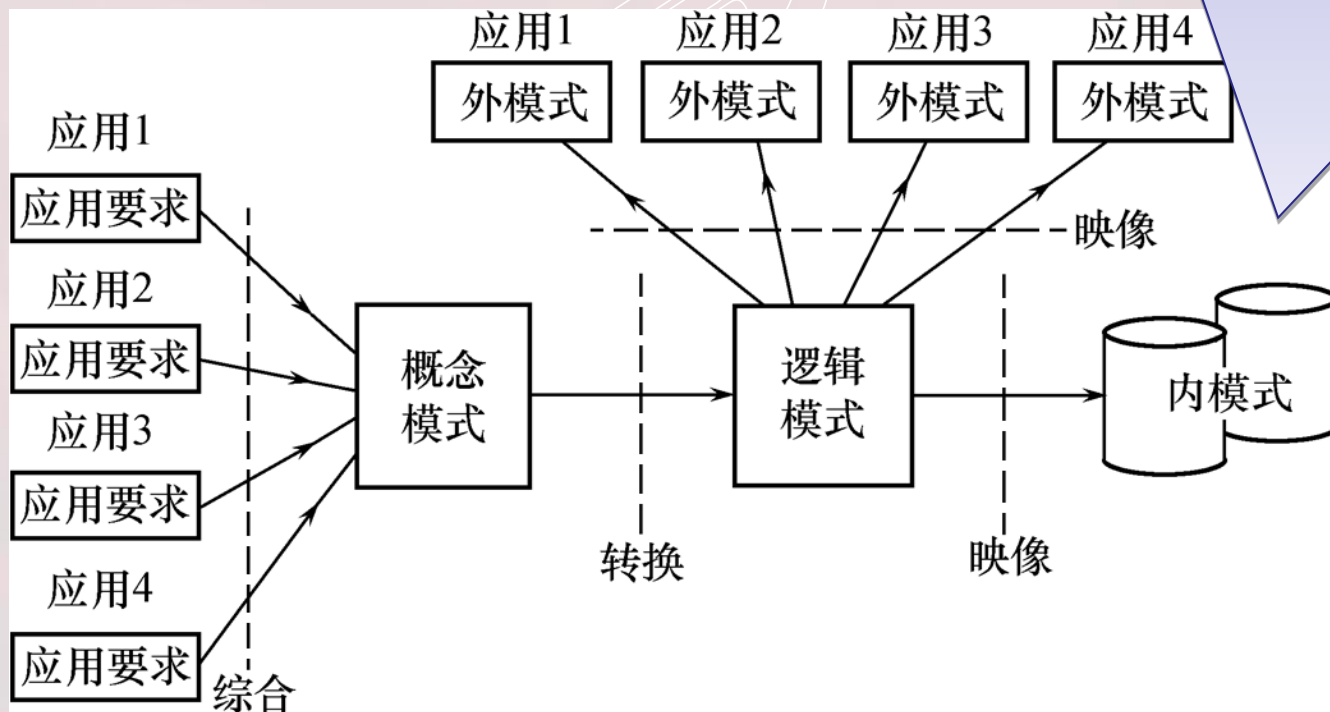
1. 首先将E-R图转换成具体的数据库产品支持的数据模型，如关系模型，形成数据库**逻辑模式**
2. 然后根据用户处理的要求、安全性的考虑，在基本表的基础上再建立必要的视图（**View**），形成数据的**外模式**

数据库设计过程中的各级模式（续）

❖ 数据库设计不同

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



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 物理结构设计

7.6 数据库的实施和维护

7.7 小结



7.2 需求分析

7.2.1 需求分析的任务

7.2.2 需求分析的方法

7.2.3 数据字典



需求分析（续）

❖ 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



7.2.1 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



需求分析的任务（续）

❖ 调查的重点是“数据”和“处理”，获得用户对数据库的要求

（1）信息要求

- 用户需要从数据库中获得信息的内容与性质
- 由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

（2）处理要求

- 用户要完成的处理功能
- 对处理性能的要求

（3）安全性与完整性要求



需求分析的任务（续）

❖ 确定用户最终需求的难点

- 用户缺少计算机知识，不能准确地表达自己的需求，他们所提出的需求往往不断地变化。
- 设计人员缺少用户的专业知识，不易理解用户的真正需求，甚至误解用户的需求

❖ 解决方法

- 设计人员必须不断深入地与用户进行交流，才能逐步确定用户的实际需求



7.2 需求分析

7.2.1 需求分析的任务

7.2.2 需求分析的方法

7.2.3 数据字典



7.2.2 需求分析的方法

- ❖ 调查清楚用户的实际需求并进行初步分析
- ❖ 与用户达成共识
- ❖ 分析与表达这些需求



调查用户需求的步骤

- (1) 调查组织机构情况
- (2) 调查各部门的业务活动情况
- (3) 协助用户明确对新系统的各种要求，包括信息要求、处理要求、完全性与完整性要求
- (4) 确定新系统的边界



常用调查方法

(1) 跟班作业

- 通过亲身参加业务工作了解业务活动的情况

(2) 开调查会

- 通过与用户座谈来了解业务活动情况及用户需求

(3) 请专人介绍

(4) 询问

- 对某些调查中的问题，可以找专人询问

(5) 设计调查表请用户填写

- 调查表设计合理，则很有效

(6) 查阅记录

- 查阅与原系统有关的数据记录



进一步分析和表达用户需求

❖ 分析方法

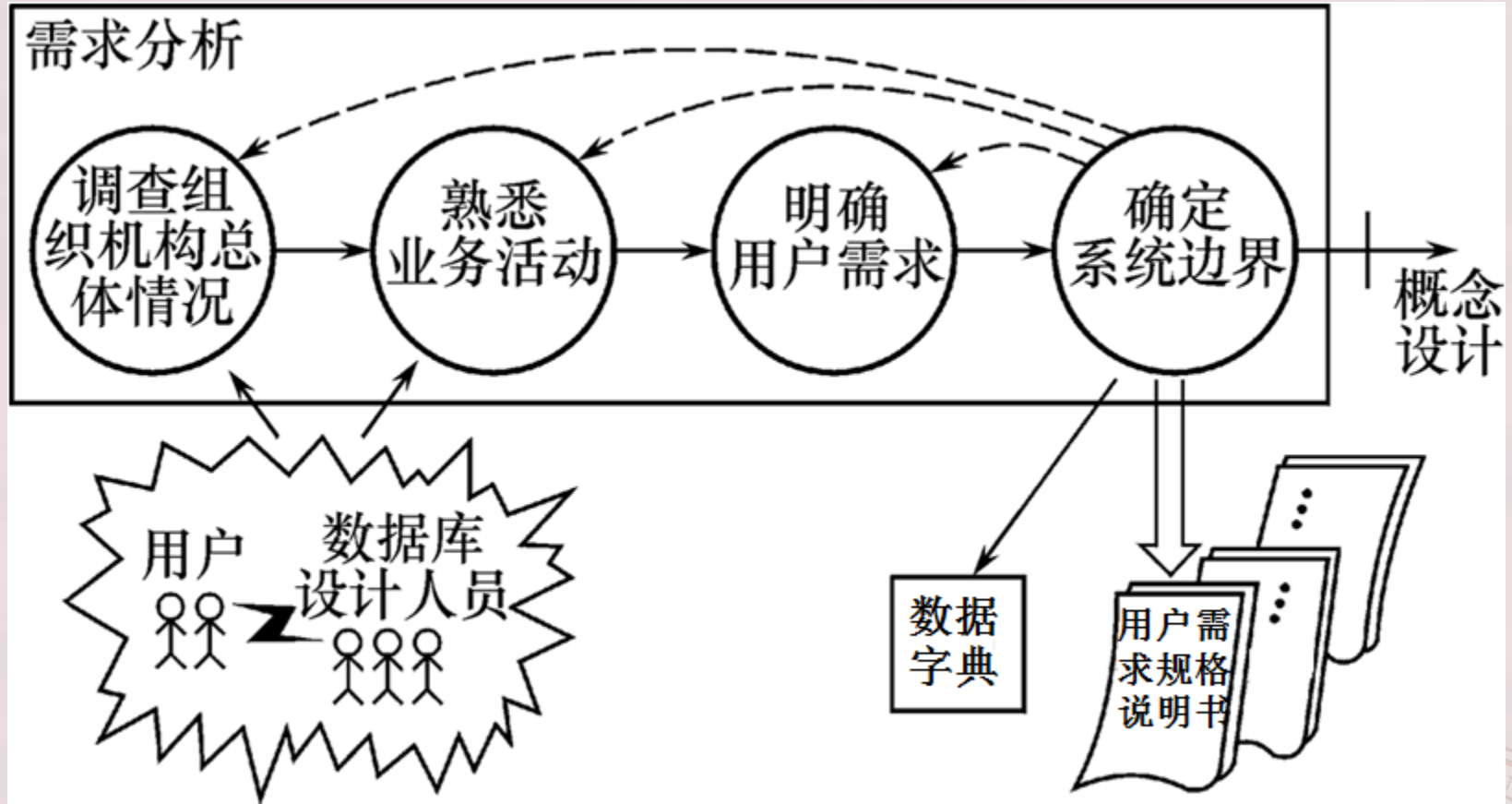
■ 结构化分析方法（**Structured Analysis**，简称**SA**方法）

- SA方法从最上层的系统组织机构入手
- 采用自顶向下、逐层分解的方式分析系统

❖ 对用户需求进行分析与表达后，需求分析报告必须提交给用户，征得用户的认可



需求分析过程



需求分析过程

7.2 需求分析

7.2.1 需求分析的任务

7.2.2 需求分析的方法

7.2.3 数据字典

中国人民大学
数据库系统概论



7.2.3 数据字典

- ❖ 数据字典是关于数据库中数据的描述，即元数据，不是数据本身
- ❖ 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
- ❖ 数据字典是进行详细的数据收集和数据分析所获得的主要结果

注意：

和关系数据库管理系统中数据字典的区别和联系



数据字典（续）

❖ 数据字典的内容

- 数据项
- 数据结构
- 数据流
- 数据存储
- 处理过程

❖ 数据项是数据的最小组成单位

❖ 若干个数据项可以组成一个数据结构

❖ 数据字典通过对数据项和数据结构的定义来描述数据流、数据存储的逻辑内容



1. 数据项

- ❖ 数据项是不可再分的数据单位
- ❖ 对数据项的描述

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,
与其他数据项的逻辑关系,
数据项之间的联系}

- “取值范围”、“与其他数据项的逻辑关系”定义了数据的完整性约束条件，是设计数据检验功能的依据
- 可以用关系规范化理论为指导，用数据依赖的概念分析和表示数据项之间的联系

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。
- ❖ 对数据结构的描述

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}}



3. 数据流

❖ 数据流是数据结构在系统内传输的路径。

❖ 对数据流的描述

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}

■ 数据流来源：说明该数据流来自哪个过程

■ 数据流去向：说明该数据流将到哪个过程去

■ 平均流量：在单位时间（每天、每周、每月等）里的传输次数

■ 高峰期流量：在高峰时期的数据流量



4. 数据存储

❖ 数据存储是数据结构停留或保存的地方，也是数据流的来源和去向之一。

❖ 对数据存储的描述

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}

- 存取频度：每小时、每天或每周存取次数，每次存取的数据量等信息
- 存取方法：批处理 / 联机处理；检索 / 更新；顺序检索 / 随机检索
- 输入的数据流：数据来源
- 输出的数据流：数据去向



5. 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。数据字典中只需要描述处理过程的说明性信息

❖ 处理过程说明性信息的描述

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：该处理过程用来做什么
- 处理要求：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
- 处理要求是后面物理设计的输入及性能评价的标准



需求分析小结

- ❖ 把需求收集和分析作为数据库设计的第一阶段是十分重要的。
- ❖ 第一阶段收集的基础数据（用数据字典来表达）是下一步进行概念设计的基础。
- ❖ 强调两点
 - (1) 设计人员应充分考虑到可能的扩充和改变，使设计易于更改，系统易于扩充
 - (2) 必须强调用户的参与



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 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

7.3.1 概念模型

7.3.2 E-R模型

*7.3.3 扩展的E-R模型

*7.3.4 UML

7.3.5 概念结构设计



7.3.1 概念模型

- ❖ 将需求分析得到的用户需求抽象为信息结构（即概念模型）的过程就是概念结构设计
- ❖ 概念模型的特点
 - (1) 能真实、充分地反映现实世界，是现实世界的一个真实模型。
 - (2) 易于理解，从而可以用它和不熟悉计算机的用户交换意见。
 - (3) 易于更改，当应用环境和应用要求改变时，容易对概念模型修改和扩充。
 - (4) 易于向关系、网状、层次等各种数据模型转换
- ❖ 描述概念模型的工具
 - E-R模型



7.3 概念结构设计

7.3.1 概念结构

7.3.2 E-R模型

*7.3.3 扩展的E-R模型

*7.3.4 UML

7.3.5 概念结构设计



7.3.2 E-R模型

1. 实体之间的联系

(1) 两个实体型之间的联系:

① 一对一联系 ($1:1$)

② 一对多联系 ($1:n$)

③ 多对多联系 ($m:n$)



E-R模型（续）

① 一对一联系（1：1）

- 如果对于实体集**A**中的每一个实体，实体集**B**中至多有一个（也可以没有）实体与之联系，反之亦然，则称实体集**A**与实体集**B**具有一对一联系，记为**1：1**。
- 例如，学校里一个班级只有一个正班长，而一个班长只在一个班中任职，则班级与班长之间具有一对一联系。



E-R模型（续）

②一对多联系（1： n ）

- 如果对于实体集**A**中的每一个实体，实体集**B**中有 n 个实体（ $n \geq 0$ ）与之联系，反之，对于实体集**B**中的每一个实体，实体集**A**中至多只有一个实体与之联系，则称实体集**A**与实体集**B**有一对多联系，记为1： n 。
- 例如，一个班级中有若干名学生，而每个学生只在一个班级中学习，则班级与学生之间具有一对多联系。



E-R模型（续）

③多对多联系（ $m:n$ ）

- 如果对于实体集**A**中的每一个实体，实体集**B**中有 **n** 个实体（ $n \geq 0$ ）与之联系，反之，对于实体集**B**中的每一个实体，实体集**A**中也有 **m** 个实体（ $m \geq 0$ ）与之联系，则称实体集**A**与实体集**B**具有多对多联系，记为 **$m:n$** 。
- 例如，一门课程同时有若干个学生选修，而一个学生可以同时选修多门课程，则课程与学生之间具有多对多联系。



E-R模型（续）

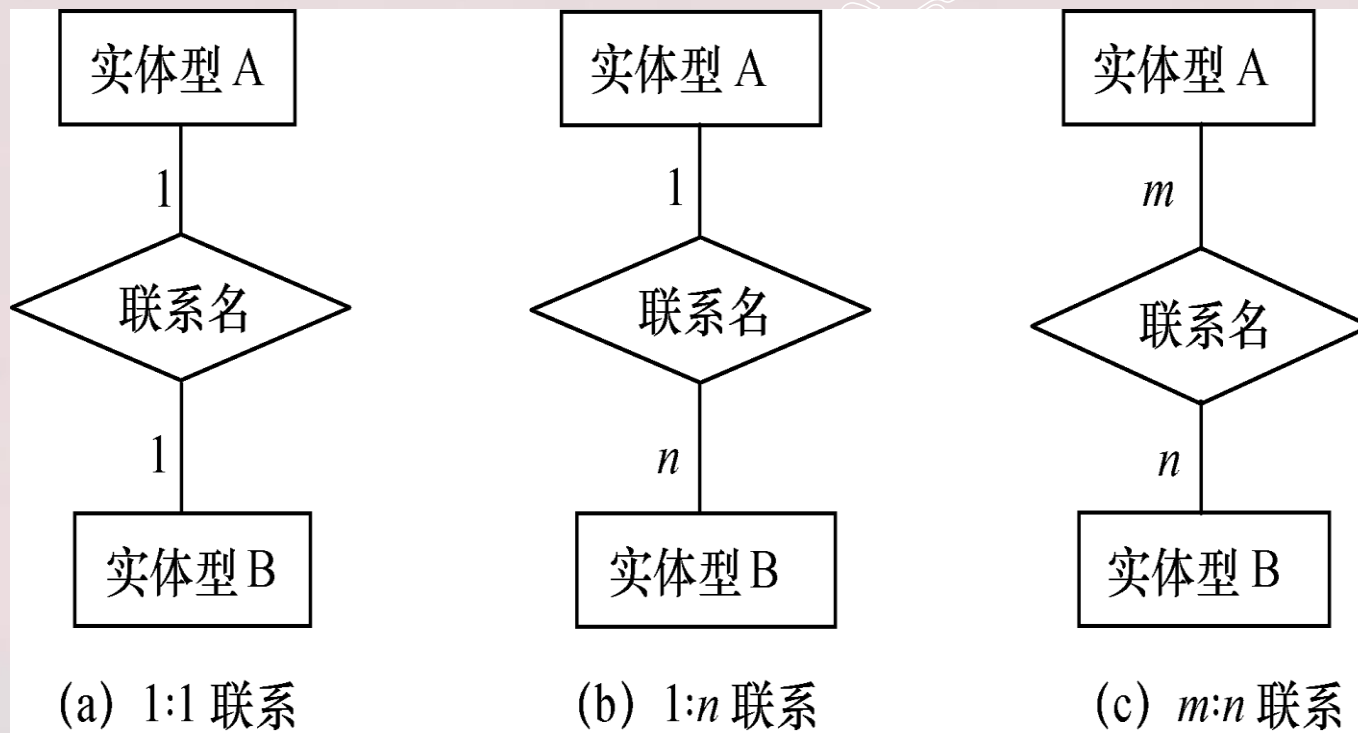


图7.6 两个实体型之间的三类联系



E-R模型（续）

(2) 两个以上的实体型之间的联系

- 一般地，两个以上的实体型之间也存在着一对一、一对多、多对多联系。
- 对于课程、教师与参考书3个实体型，如果一门课程可以有若干个教师讲授，使用若干本参考书，而每一个教师只讲授一门课程，每一本参考书只供一门课程使用，则课程与教师、参考书之间的联系是一对多的，如图7.7(a)所示。



E-R模型 (续)

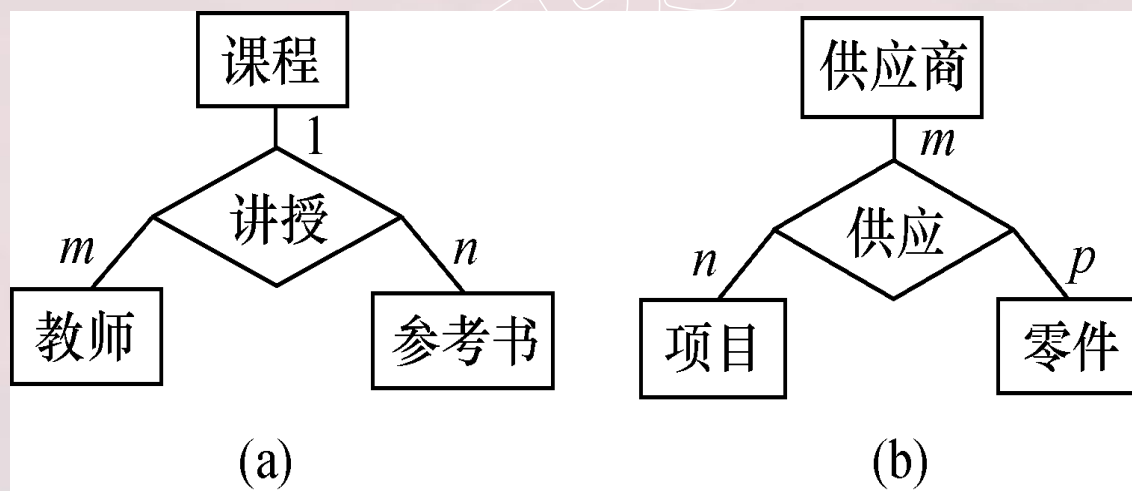


图7.7 三个实体型之间的联系示例



E-R模型（续）

(3) 单个实体型内的联系

- 同一个实体集内的各实体之间也可以存在一对一、一对多、多对多的联系。
- 例如，职工实体型内部具有领导与被领导的联系，即某一职工（干部）“领导”若干名职工，而一个职工仅被另外一个职工直接领导，因此这是一对多的联系，如图7.8所示。



E-R模型（续）

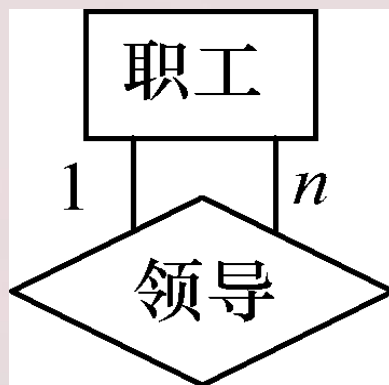


图7.8 单个实体型内的一对多联系示例

■联系的度：参与联系的实体型的数目

- 2个实体型之间的联系度为2，也称为二元联系；
- 3个实体型之间的联系度为3，称为三元联系；
- N个实体型之间的联系度为N，也称为N元联系



E-R模型（续）

2. E-R图

■ E-R图提供了表示实体型、属性和联系的方法：

- 实体型：用矩形表示，矩形框内写明实体名。
- 属性：用椭圆形表示，并用无向边将其与相应的实体型连接起来。

▶ 例如，学生实体具有学号、姓名、性别、出生年份、系、入学时间等属性，用E-R图表示如图7.9所示

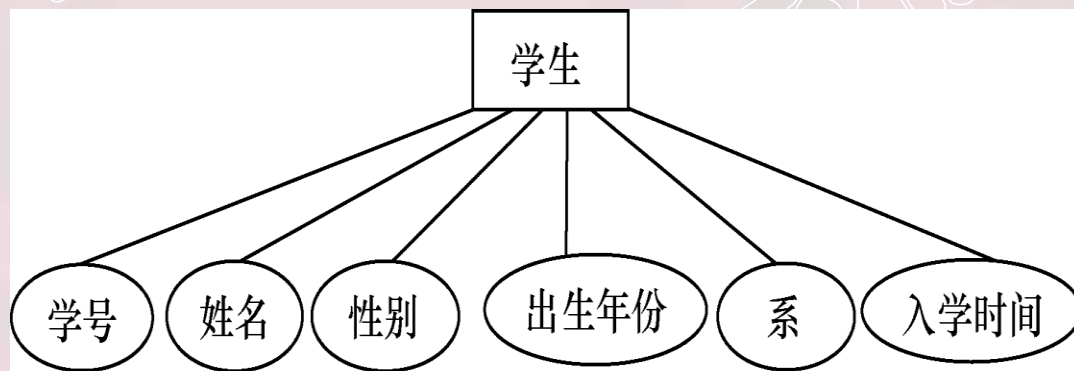


图7.9 学生实体及属性



E-R模型（续）

- 联系：用菱形表示，菱形框内写明联系名，并用无向边分别与有关实体型连接起来，同时，在无向边旁标上联系的类型（1:1, 1:n或m:n）。
- 联系可以具有属性

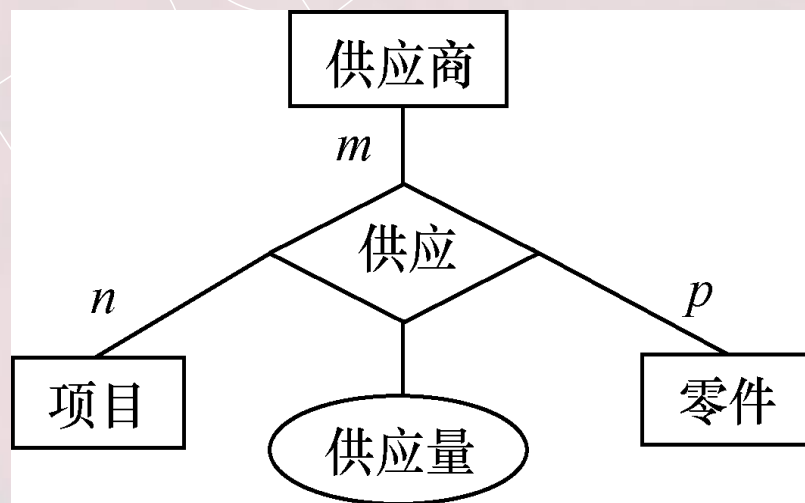


图7.10 联系的属性



E-R模型（续）

❖ 3. 一个实例

■ 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



E-R模型（续）

■ 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。



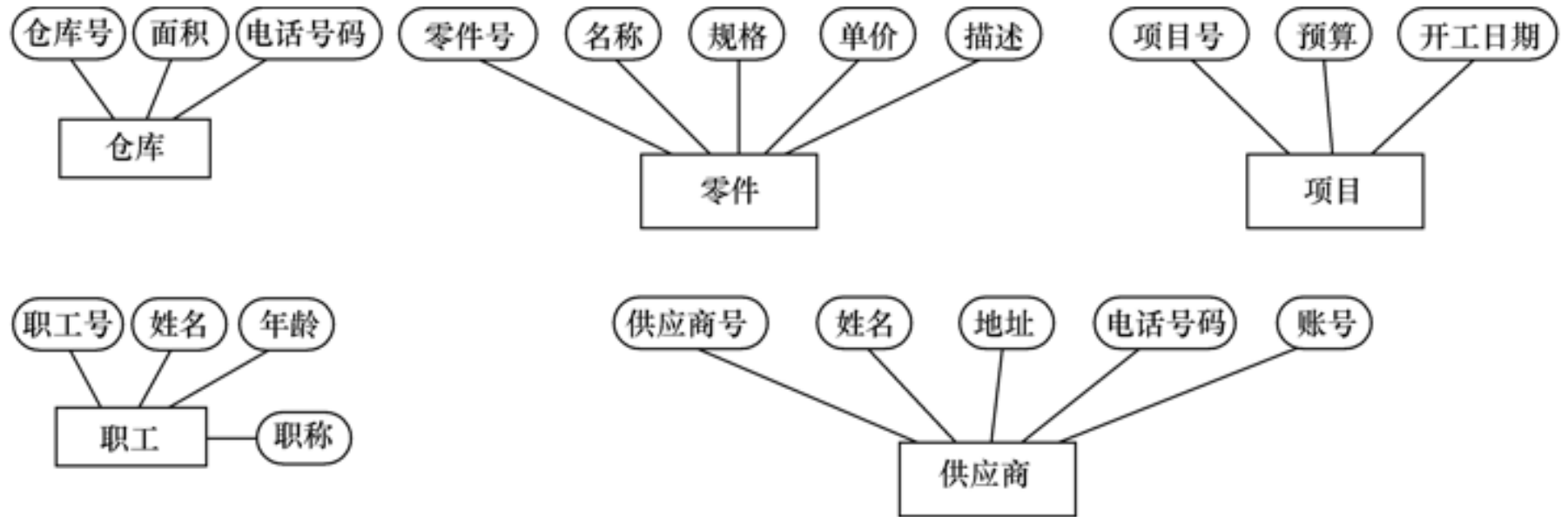
E-R模型（续）

■ 这些实体之间的联系如下（续）：

- （3） 职工之间具有领导与被领导关系。即仓库主任领导若干保管员，因此职工实体型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。



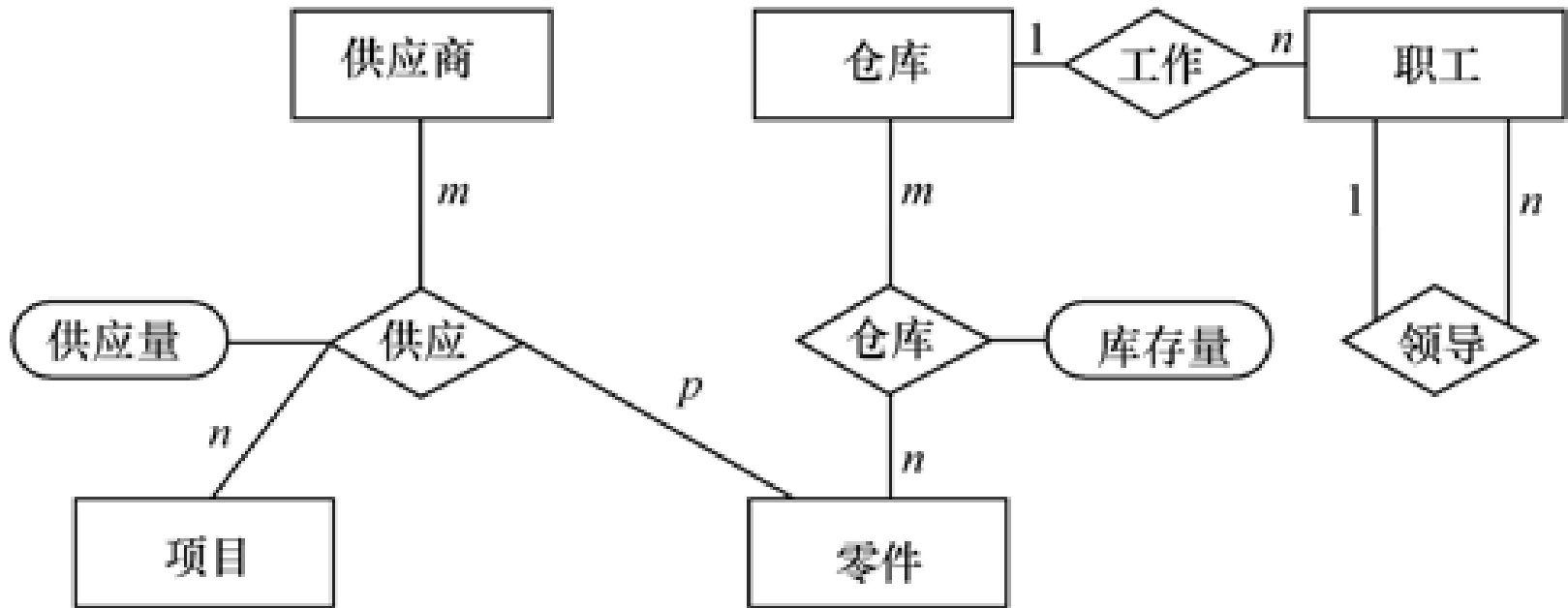
E-R模型 (续)



(a) 实体及其属性图



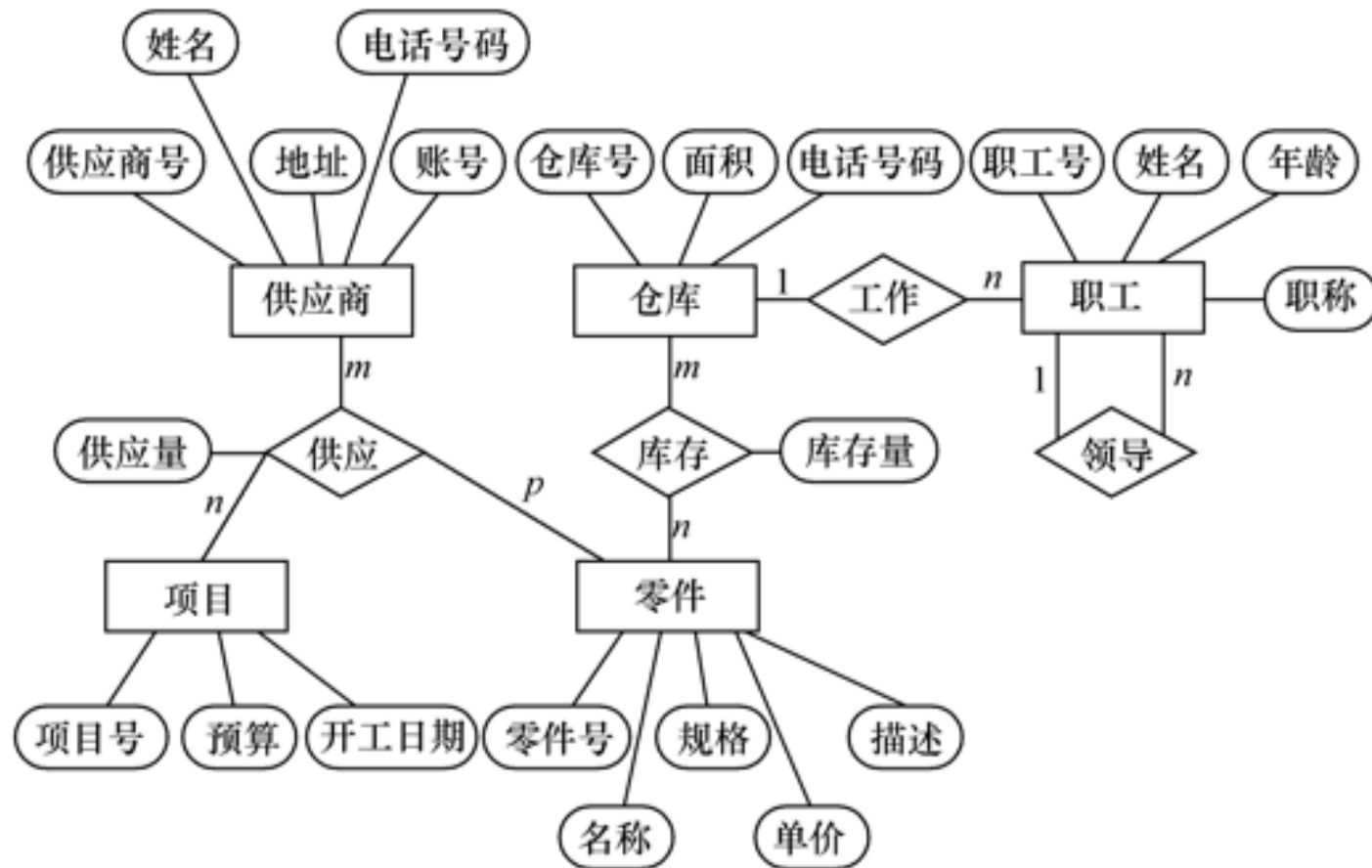
E-R模型 (续)



(b) 实体及其联系图



E-R模型 (续)



(c) 完整的实体-联系图

E-R模型（续）

❖ E-R图，可以参见：

■ 爱课程网数据库系统概论1.2节动画

《E-R图难点解析（1）》

《E-R图难点解析（2）》

《E-R图难点解析（3）》



7.3 概念结构设计

7.3.1 概念结构

7.3.2 E-R模型

*7.3.3 扩展的E-R模型

*7.3.4 UML

7.3.5 概念结构设计



7.3.5 概念结构设计

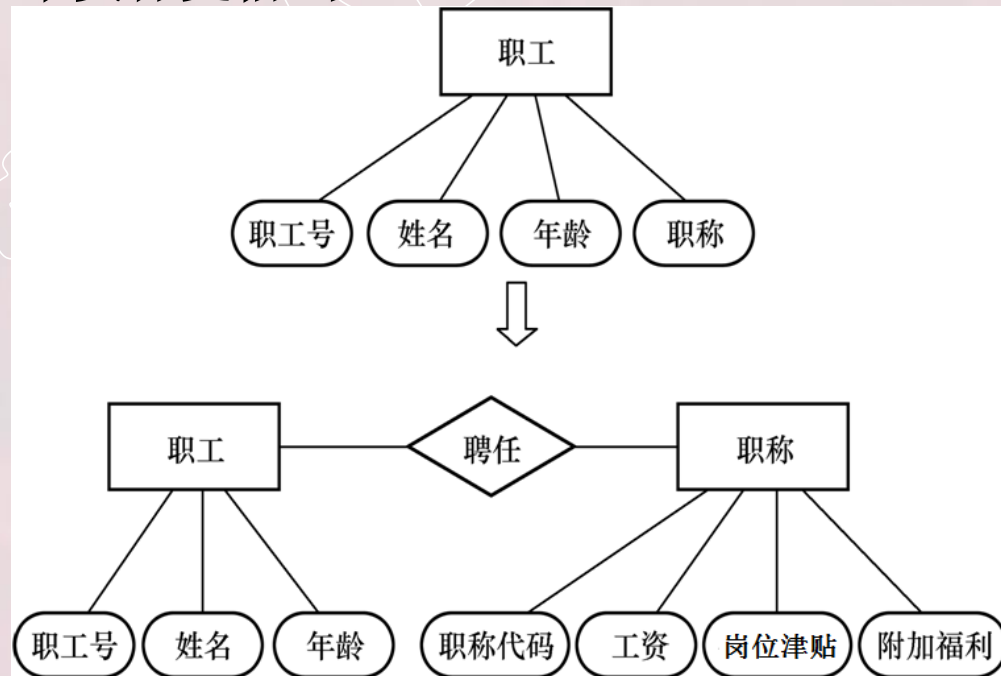
1. 实体与属性的划分原则

- 为了简化E-R图的处置，现实世界的事物能作为属性对待的，尽量作为属性对待。
- 两条准则：
 - (1) 作为属性，不能再具有需要描述的性质。属性必须是不可分的数据项，不能包含其他属性。
 - (2) 属性不能与其他实体具有联系，即E-R图中所表示的联系是实体之间的联系。



概念结构设计（续）

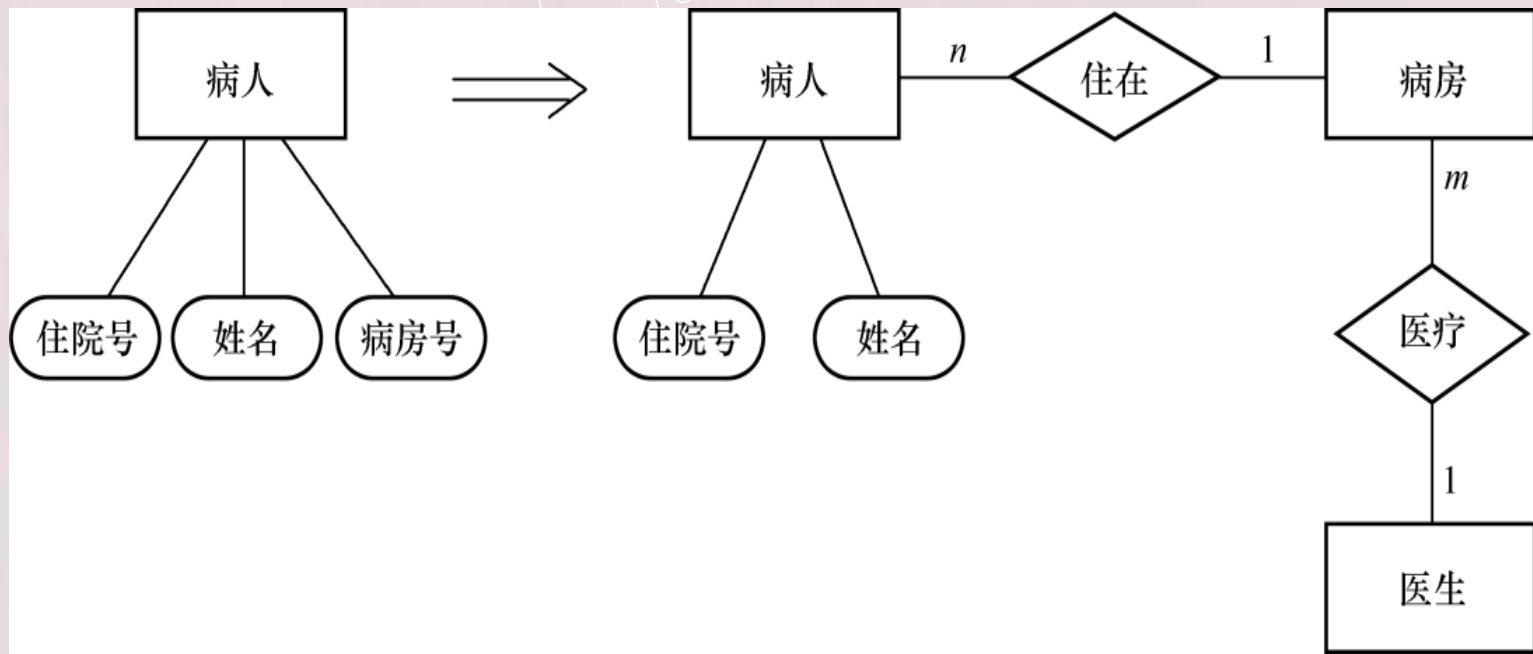
- [例1]** 职工是一个实体，职工号、姓名、年龄是职工的属性。
- 职称如果没有与工资、福利挂钩，根据准则（1）可以作为职工实体的属性
 - 如果不同的职称有不同的工资、住房标准和不同的附加福利，则职称作为一个实体更恰当



概念结构设计（续）

[例2] 在医院中，一个病人只能住在一个病房，病房号可以作为病人实体的一个属性；

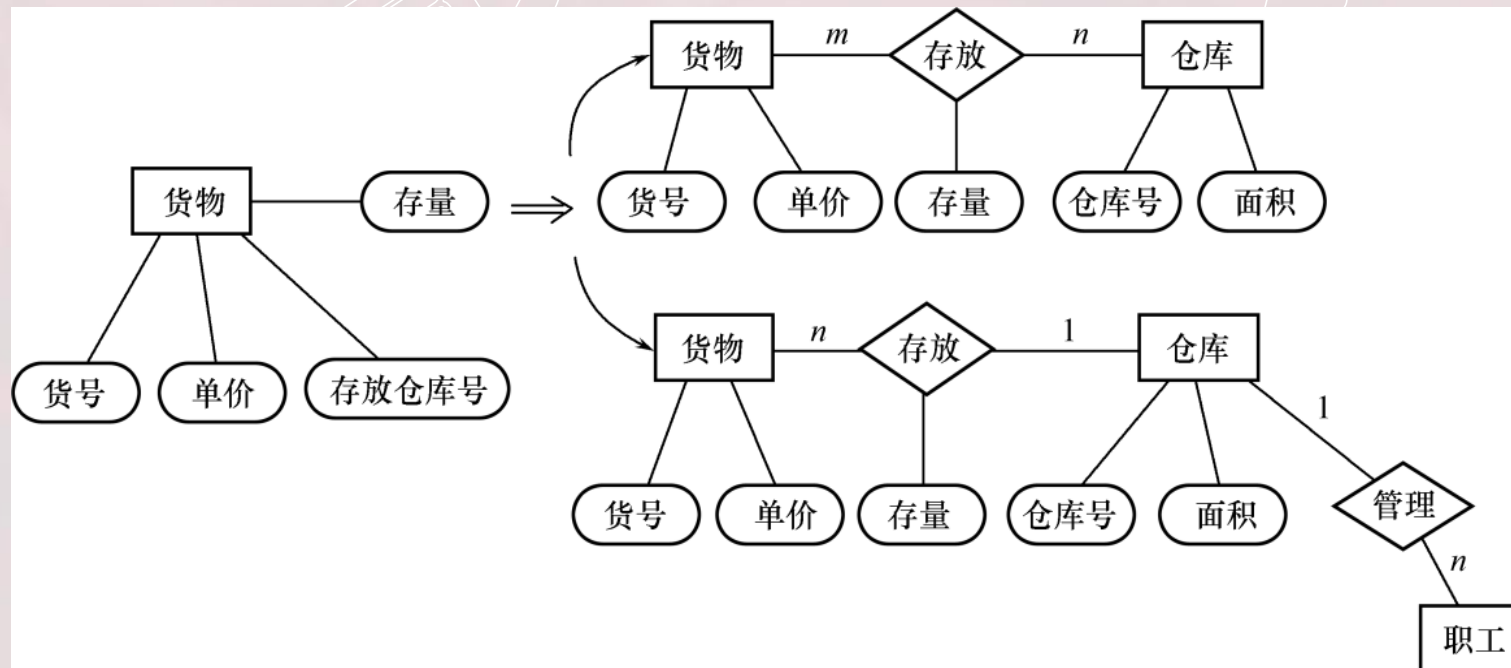
如果病房还要与医生实体发生联系，即一个医生负责几个病房的病人的医疗工作，则根据准则（2）病房应作为一个实体。



概念结构设计（续）

[例3] 如果一种货物只存放在一个仓库，那么就可以把存放货物的仓库的仓库号作为描述货物存放地点的属性。

如果一种货物可以存放在多个仓库中，或者仓库本身又用面积作为属性，或者仓库与职工发生管理上的联系，那么就应把仓库作为一个实体。



概念结构设计（续）

❖ [例7.1] 销售管理子系统E-R图的设计。

■ 该子系统的主要功能是：

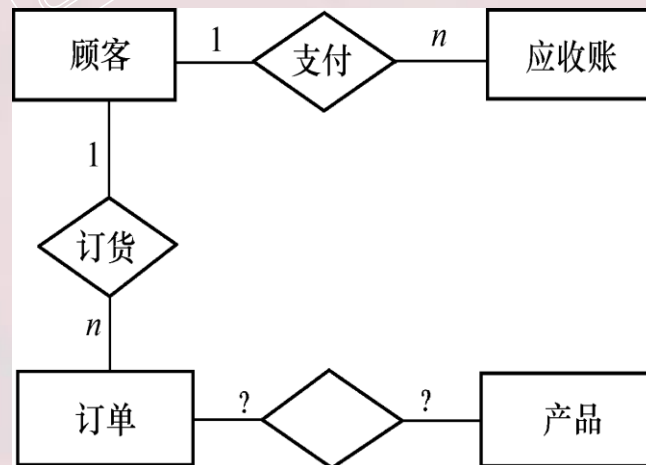
- 处理顾客和销售员送来的订单
- 工厂是根据订货安排生产的
- 交出货物同时开出发票
- 收到顾客付款后，根据发票存根和信贷情况进行应收款处理



概念结构设计（续）

❖ 参照需求分析和数据字典中的详尽描述，遵循前面给出的两个准则，进行了如下调整：

(1) 每张订单由订单号、若干头信息和订单细节组成。订单细节又有订货的零件号、数量等来描述。按照准则(2)，订单细节就不能作订单的属性处理而应该上升为实体。一张订单可以订若干产品，所以订单与订单细节两个实体之间是1:n的联系。



概念结构设计（续）

- (2) 原订单和产品的联系实际上是订单细节和产品的联系。每条订货细节对应一个产品描述，订单处理时从中获得当前单价、产品重量等信息。
- (3) 工厂对大宗订货给予优惠。每种产品都规定了不同订货数量的折扣，应增加一个“折扣规则”实体存放这些信息，而不应把它们放在产品实体中。



概念结构设计（续）

❖ 最后得到销售管理子系统E-R图如图7.23所示。

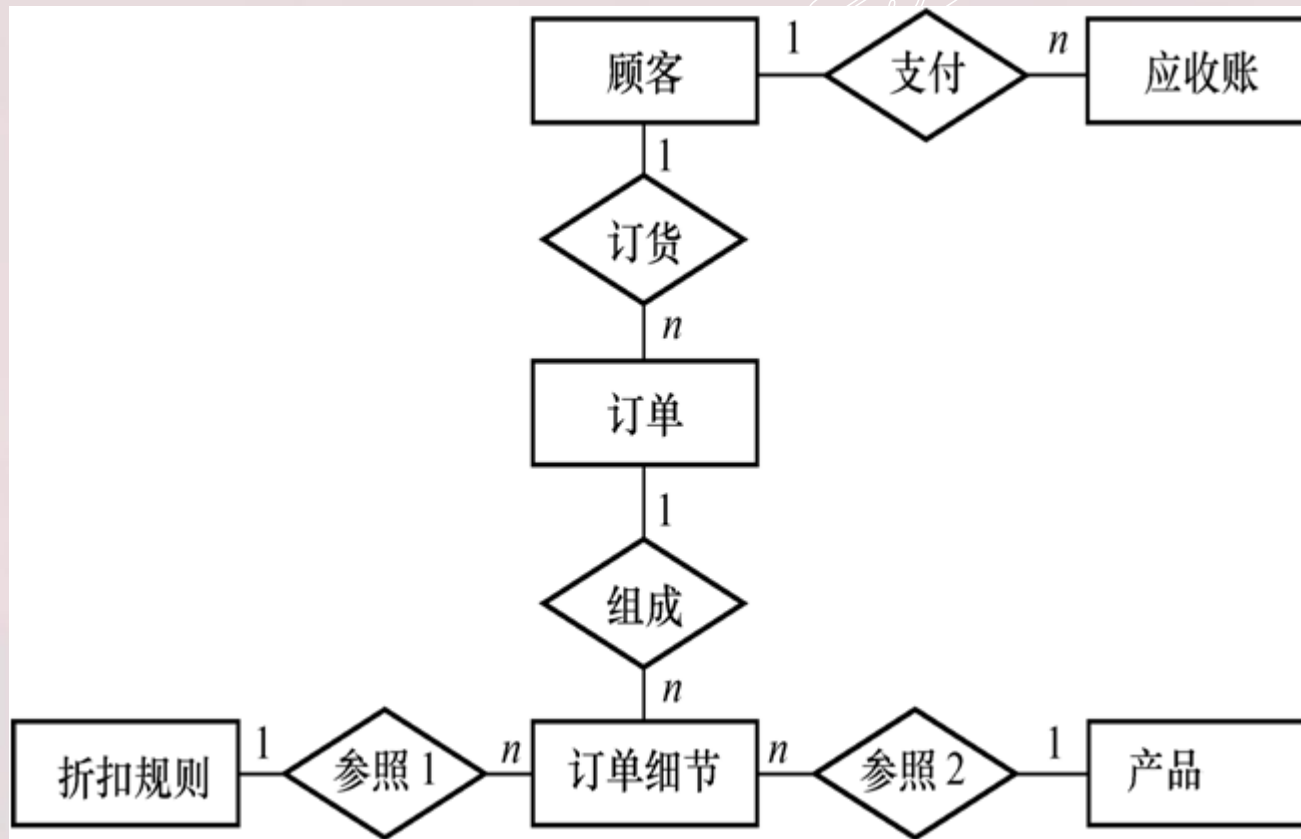


图7.23 销售管理子系统的E-R图



概念结构设计（续）

❖ 对每个实体定义的属性如下：

- 顾客：{顾客号，顾客名，地址，电话，信贷状况，账目余额}
- 订单：{订单号，顾客号，订货项数，订货日期，交货日期，工种号，生产地点}
- 订单细则：{订单号，细则号，零件号，订货数，金额}
- 应收账款：{顾客号，订单号，发票号，应收金额，支付日期，支付金额，当前余额，贷款限额}
- 产品：{产品号，产品名，单价，重量}
- 折扣规则：{产品号，订货量，折扣}

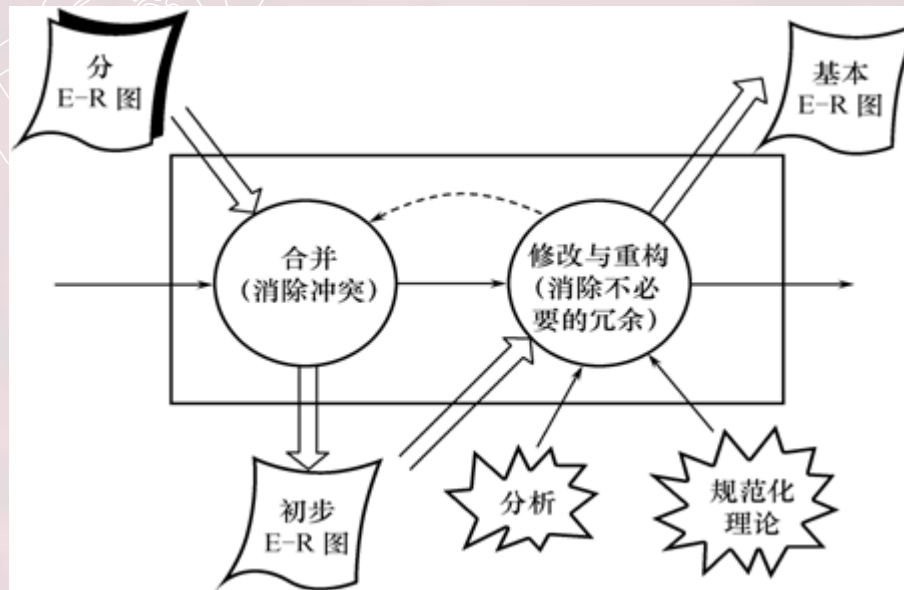


概念结构设计（续）

2. E-R图的集成

■ E-R图的集成一般需要分两步

- 合并。解决各分E-R图之间的冲突，将分E-R图合并起来生成初步E-R图。
- 修改和重构。消除不必要的冗余，生成基本E-R图。



概念结构设计（续）

（1）合并E-R图，生成初步E-R图

- 各个局部应用所面向的问题不同，各个子系统的E-R图之间必定会存在许多不一致的地方，称之为冲突。
- 子系统E-R图之间的冲突主要有三类：
 - ① 属性冲突
 - ② 命名冲突
 - ③ 结构冲突



概念结构设计（续）

①属性冲突

- 属性域冲突，即属性值的类型、取值范围或取值集合不同。
 - 例如零件号，有的部门把它定义为整数，有的部门把它定义为字符型。
 - 年龄，某些部门以出生日期形式表示职工的年龄，而另一些部门用整数表示职工的年龄。
- 属性取值单位冲突。
 - 例如，零件的重量有的以公斤为单位，有的以斤为单位，有的以克为单位。



概念结构设计（续）

②命名冲突

- 同名异义，即不同意义的对象在不同的局部应用中具有相同的名字。
- 异名同义（一义多名），即同一意义的对象在不同的局部应用中具有不同的名字。
 - 如对科研项目，财务科称为项目，科研处称为课题，生产管理处称为工程。
- 命名冲突
 - 可能发生在实体、联系一级上
 - 也可能发生在属性一级上
 - 通过讨论、协商等行政手段加以解决



概念结构设计（续）

③结构冲突

- 同一对象在不同应用中具有不同的抽象。
 - 例如，职工在某一局部应用中被当作实体，而在另一局部应用中则被当作属性。
 - 解决方法：把属性变换为实体或把实体变换为属性，使同一对象具有相同的抽象。
- 同一实体在不同子系统的E-R图中所包含的属性个数和属性排列次序不完全相同。
 - 解决方法：使该实体的属性取各子系统的E-R图中属性的并集，再适当调整属性的次序。



概念结构设计（续）

③结构冲突（续）

- 实体间的联系在不同的E-R图中为不同的类型。
 - 实体E1与E2在一个E-R图中是多对多联系，在另一个E-R图中是一对多联系
 - 解决方法是根据应用的语义对实体联系的类型进行综合或调整。



概念结构设计（续）

图7.25(a)中零件与产品之间存在多对多的联系“构成”

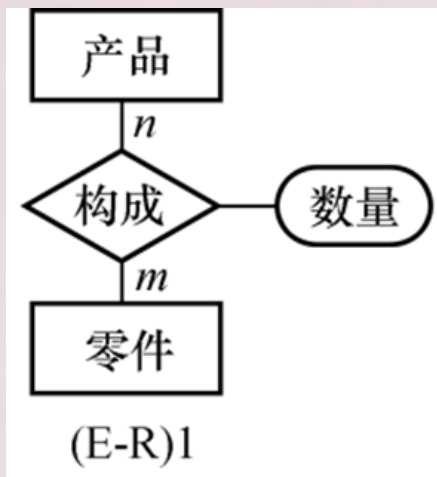
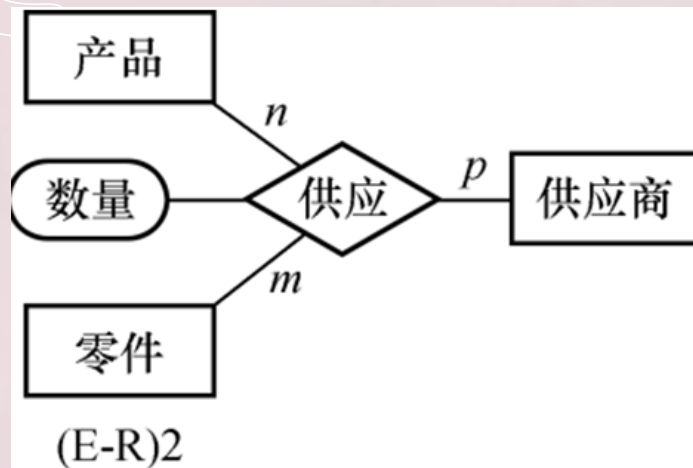
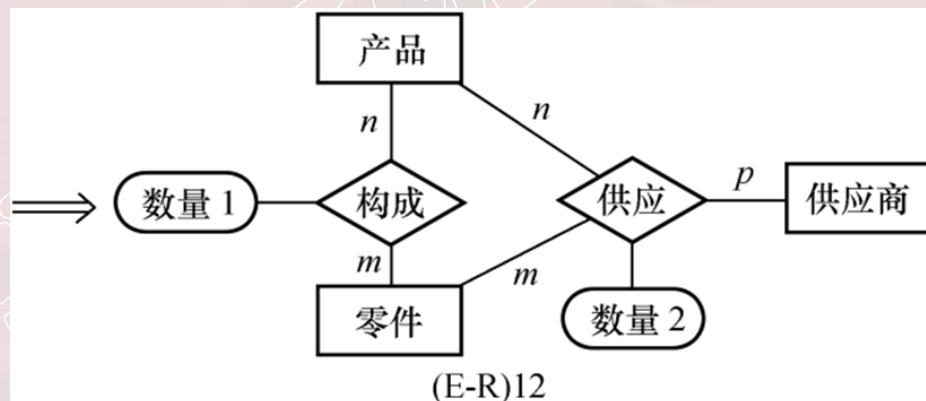


图7.25(b)中产品、零件与供应商三者之间还存在多对多的联系“供应”



合并两个E-R图，
如图7.25(c)



概念结构设计（续）

（2）消除不必要的冗余，设计基本E-R图

- 所谓冗余的数据是指可由基本数据导出的数据，冗余的联系是指可由其他联系导出的联系。
- 消除冗余主要采用分析方法，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逻辑关系的说明来消除冗余。



概念结构设计（续）

- 如图7.26中， $Q_3=Q_1 \times Q_2$ ， $Q_4=\sum Q_5$ 。所以 Q_3 和 Q_4 是冗余数据，可以消去。并且由于 Q_3 消去，产品与材料间 $m:n$ 的冗余联系也应消去。

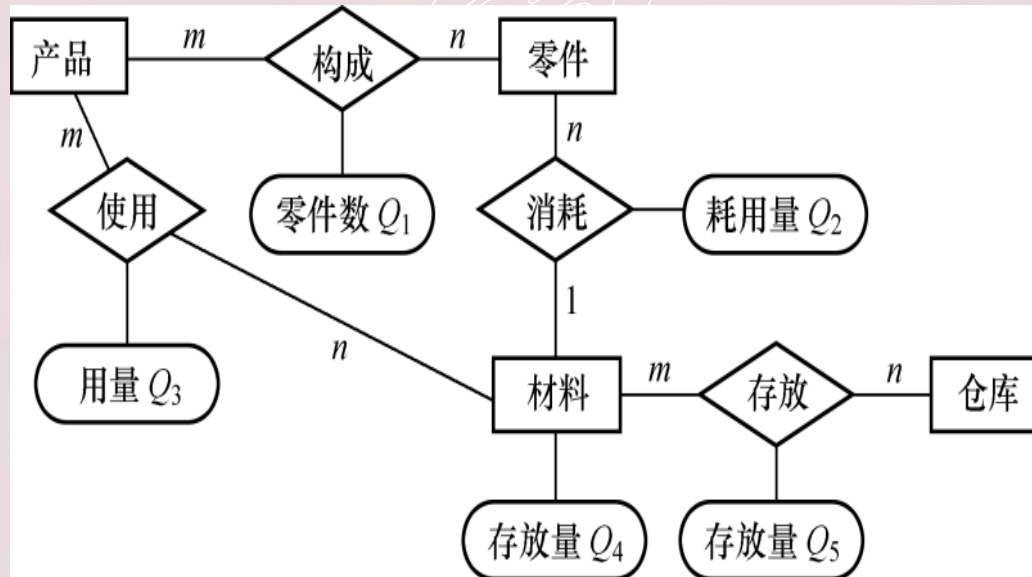


图7.26 消除冗余

并不是所有的冗余数据与冗余联系都必须加以消除，有时为了提高效率，不得不以冗余信息作为代价。

概念结构设计（续）

❖ 用规范化理论来消除冗余

① 确定分E-R图实体之间的数据依赖。

- 实体之间一对一、一对多、多对多的联系可以用实体码之间的函数依赖来表示。于是有函数依赖集 F_L 。
- 如图7.27中：
部门和职工之间一对多的联系可表示为职工号 \rightarrow 部门号
职工和产品之间多对多的联系可表示为
(职工号, 产品号) \rightarrow 工作天数等。



概念结构设计（续）

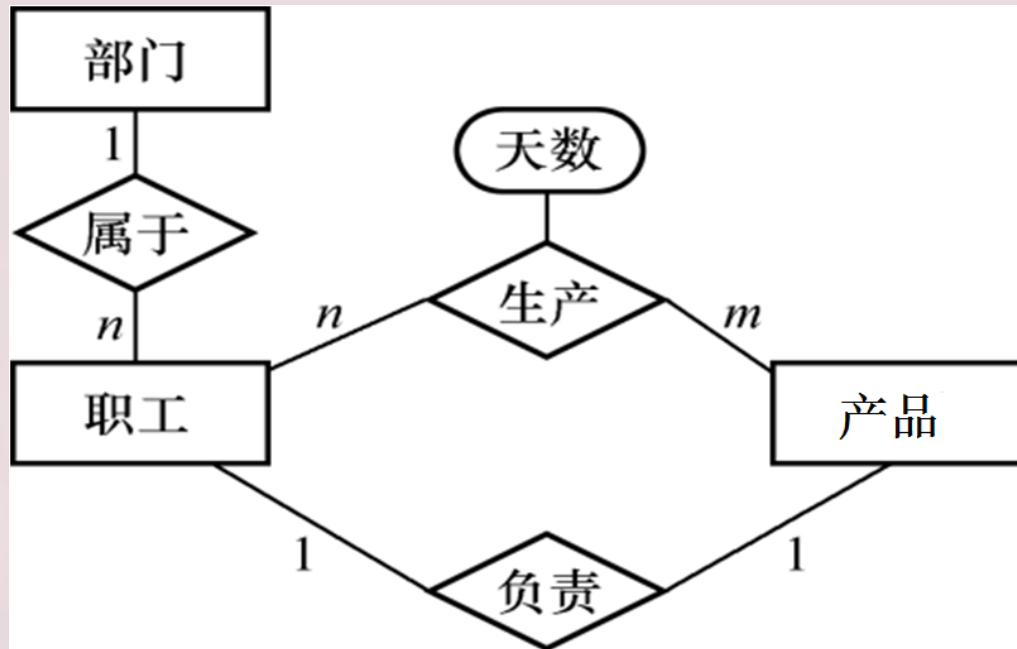


图7.27 劳动人事管理的分E-R图

②求FL的最小覆盖GL，差集为 $D=FL-GL$ 。

- 逐一考察D中的函数依赖，确定是否是冗余的联系，若是，就把它去掉。



概念结构设计（续）

- 由于规范化理论受到泛关系假设的限制，应注意下面两个问题：
 - 冗余的联系一定在D中，而D中的联系不一定是冗余的；
 - 当实体之间存在多种联系时，要将实体之间的联系在形式上加以区分。



概念结构设计（续）

❖ [例7.2] 某工厂管理信息系统的视图集成。

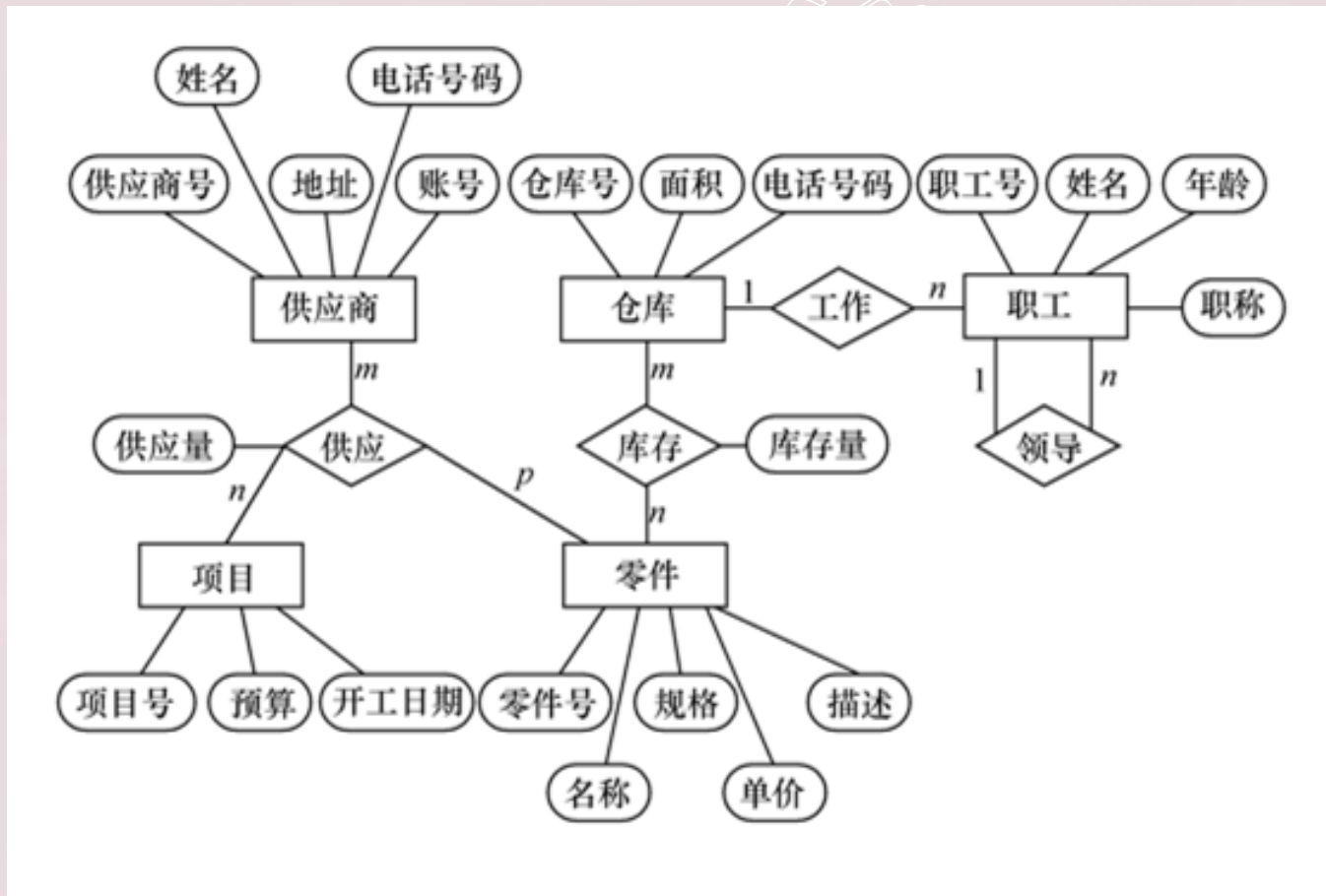


图7.11 工厂物资管理E-R图

概念结构设计（续）

❖ [例7.2] 某工厂管理信息系统的视图集成。

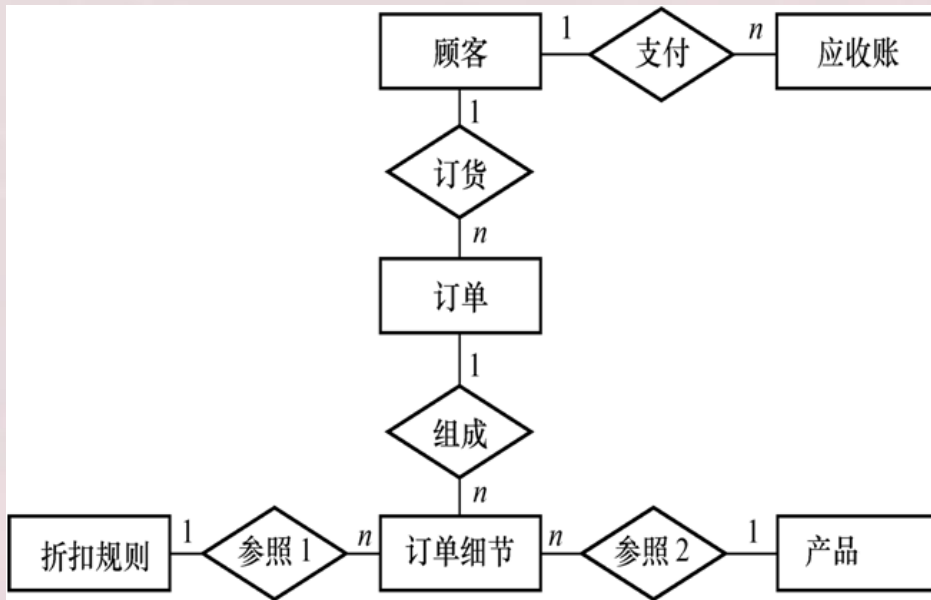


图7.23 销售管理子系统的E-R图

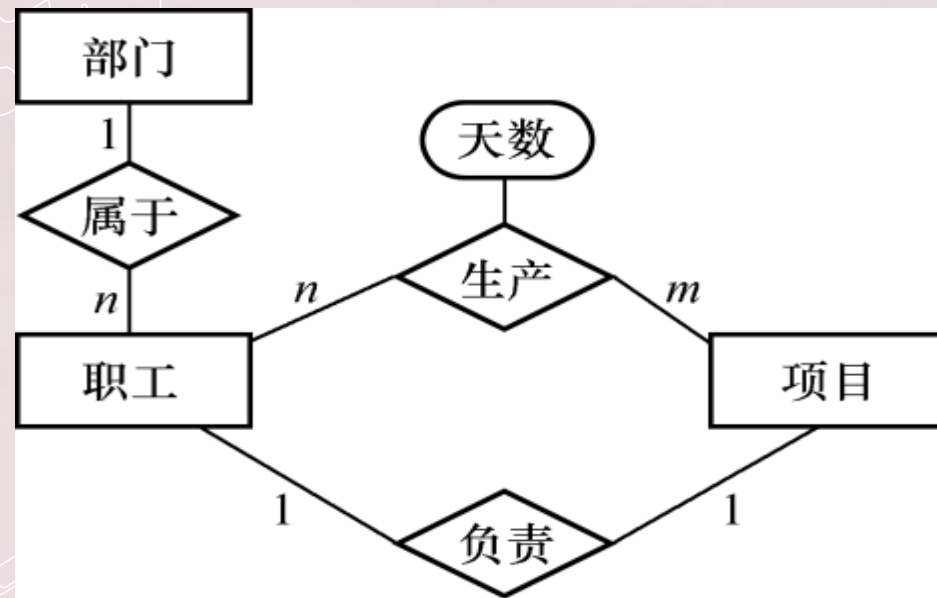
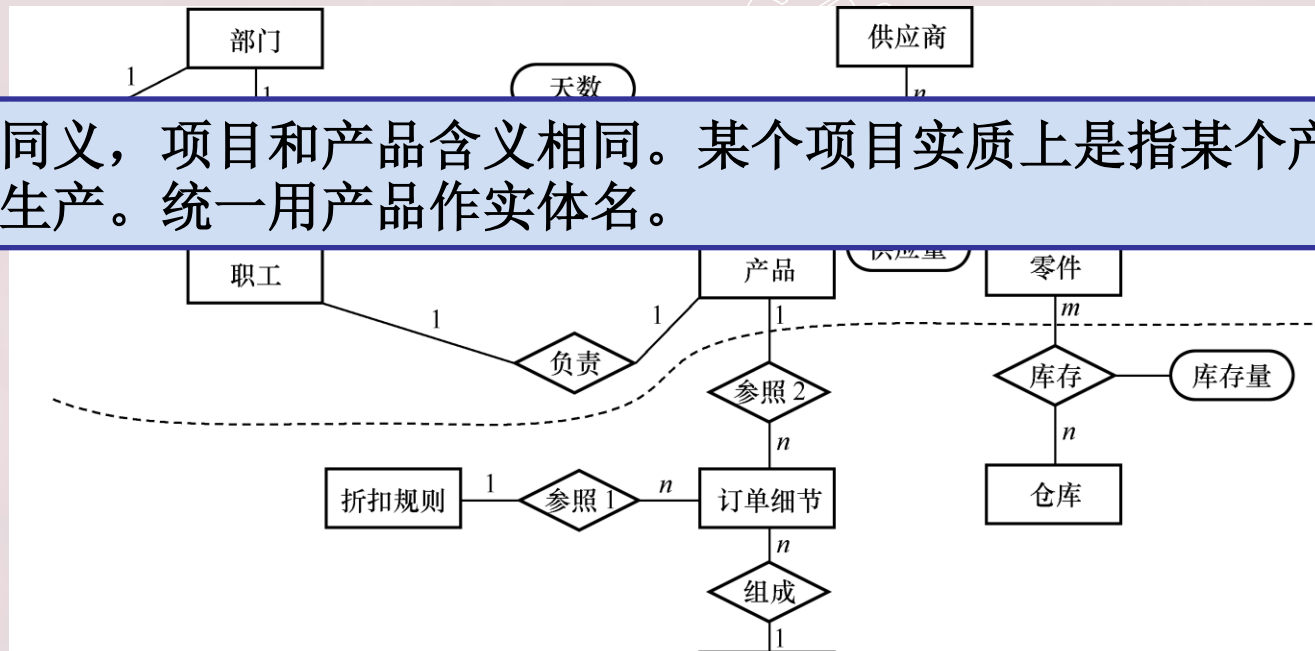


图7.27 劳动人事管理的分E-R图

概念结构设计（续）

❖ [例7.2] 某工厂管理信息系统的视图集成。

异名同义，项目和产品含义相同。某个项目实质上是指某个产品的生产。统一用产品作实体名。



库存管理中职工与仓库的工作关系已包含在劳动人事管理的部门与职工之间的联系之中，所以可以取消。职工之间领导与被领导关系可由部门与职工（经理）之间的领导关系、部门与职工之间的从属关系两者导出，所以也可以取消。

图7.28 某工厂管理信息系统的基本E-R图