第七章 数据库设计

数据库设计概述

- 数据库设计的成果
  - 建立数据库
    - 构造最优的数据库模式
  - 应用系统
    - 有效地存储、处理数据，满足用户的需求
- 数据库是信息系统的核心和基础
  - 存储、维护、检索数据

数据库设计概述

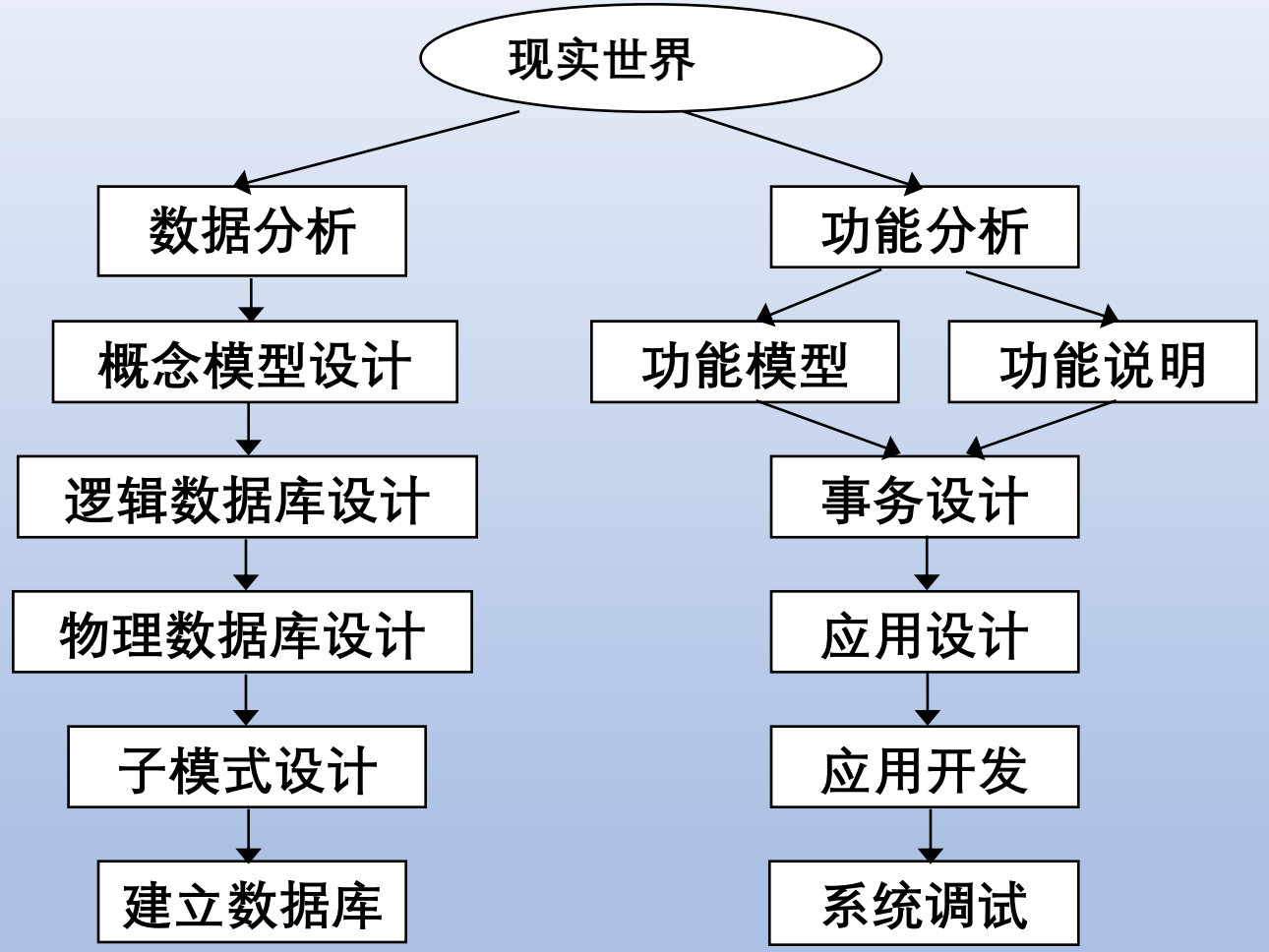
- 知识结构
  - 数据库的基本知识和数据库设计技术
  - 计算机科学的基础知识和程序设计的方法和技巧
  - 软件工程的原理和方法
  - 应用领域的知识

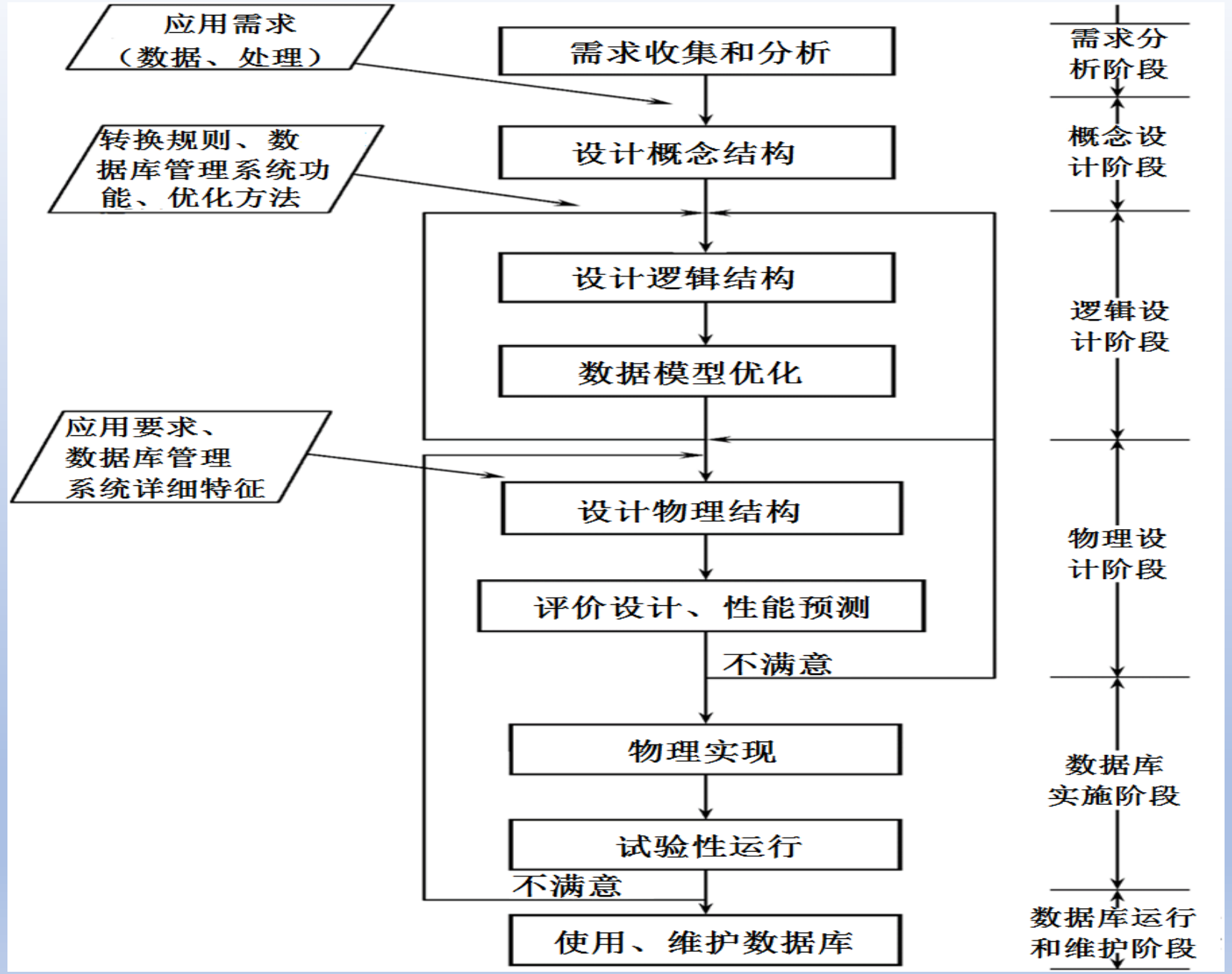
数据库设计概述——特点

- 硬件、软件、界面
- 数据（结构）、处理（行为）
- 数据库模式是公用的、共享的，影响整个应用的质量

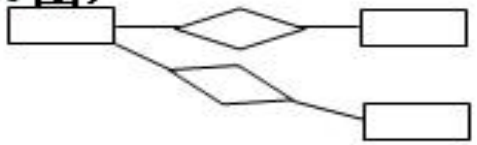
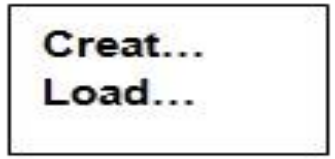
- 三分技术,七分管理,十二分基础数据.

数据库设计的特点


现实世界
数据分析
功能分析
概念模型设计
功能模型
功能说明
逻辑数据库设计
事务设计
物理数据库设计
应用设计
子模式设计
应用开发
建立数据库
系统调试


应用需求
（数据、处理）
转换规则、数据库管理系统功能、优化方法
应用要求、数据库管理系统详细特征
需求收集和分析
设计概念结构
设计逻辑结构
数据模型优化
设计物理结构
评价设计、性能预测
不满意
物理实现
试验性运行
不满意
使用、维护数据库
需求分析阶段
概念设计阶段
逻辑设计阶段
物理设计阶段
数据库实施阶段
数据库运行和维护阶段

数据库设计各个阶段的数据设计描述

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

数据库设计概述——方法

- 软件工程的思想和方法
- 新奥尔良方法
  - 需求分析
  - 概念分析
  - 逻辑设计
  - 物理设计

数据库设计概述——特性

- 反复性
  - 反复推敲、修改、迭代的过程
- 试探性
  - 结果不是唯一的
  - 在矛盾中求得平衡
- 分步进行
  - 小组、团队工作
  - 分段把关

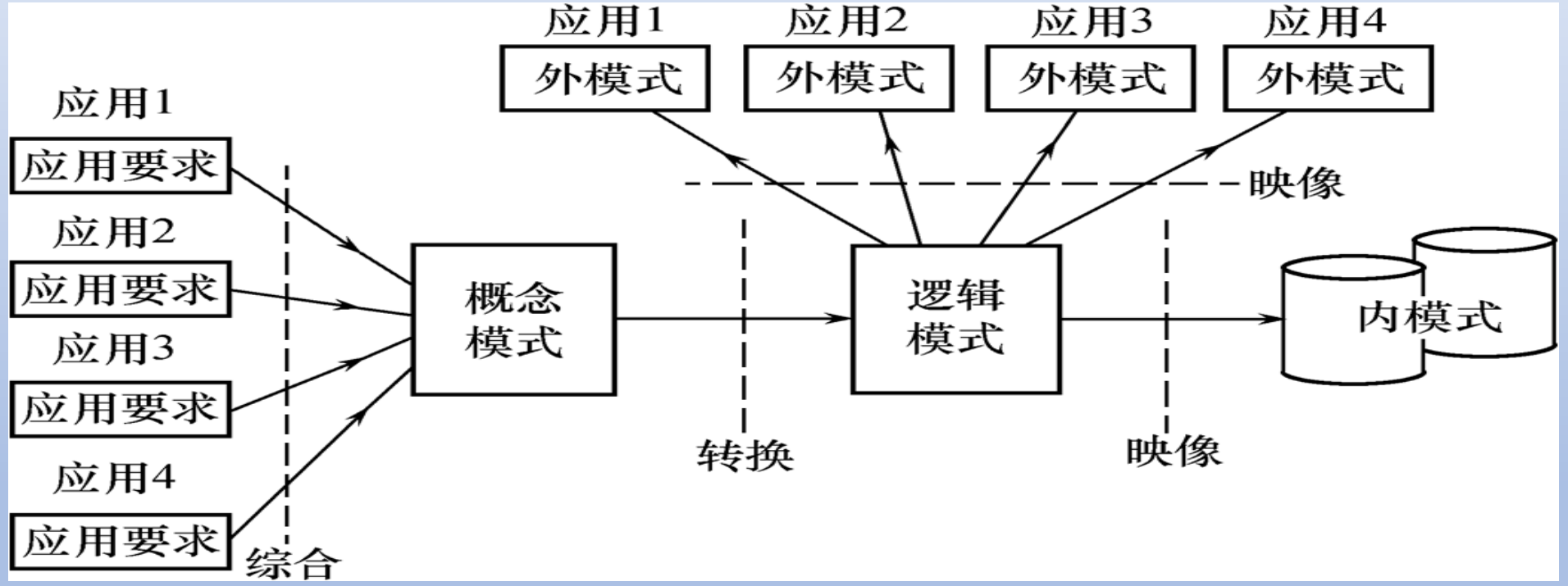
数据库设计概述——工具

- 手工迭代、优化
- 辅助工具（CASE）

  - Oracle：　　Design 2000
  - Sybase：　　Powerdesigner
  - Rational：　　Rational Rose

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

数据库设计概述——基本步骤

- 需求收集和分析
- 概念结构设计
- 逻辑结构设计
- 物理结构设计
- 物理实现

数据库设计概述——基本步骤

- 需求收集和分析
  - 用户/用途
  - 用户关心什么
  - 用户要什么结果
- 概念结构设计
- 逻辑结构设计
- 物理结构设计
- 物理实现

数据库设计概述——基本步骤

- 需求收集和分析
- 概念结构设计
  - 存什么，对现实世界的模拟
  - 关系（联系）如何，概念数据模型
  - E/R图、OO定义
- 逻辑结构设计
- 物理结构设计
- 物理实现

数据库设计概述——基本步骤

- 需求收集和分析
- 概念结构设计
- 逻辑结构设计
  - 转换成逻辑数据模型（与DBMS模型相关）
  - 数据库的模式（database schema）
  - 用户子模式（视图模式）
- 物理结构设计
- 物理实现

数据库设计概述——基本步骤

- 需求收集和分析
- 概念结构设计
- 逻辑结构设计
- 物理结构设计
  - 数据怎么存，物理数据模型
  - 根据DBMS产品、环境特点
  - 影响数据库的性能
- 物理实现

数据库设计概述——基本步骤

- 需求收集和分析
- 概念结构设计
- 逻辑结构设计
- 物理结构设计
- 物理实现
  - 运行DDL
  - 装入测试数据
  - 应用程序

需求分析的任务

- 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

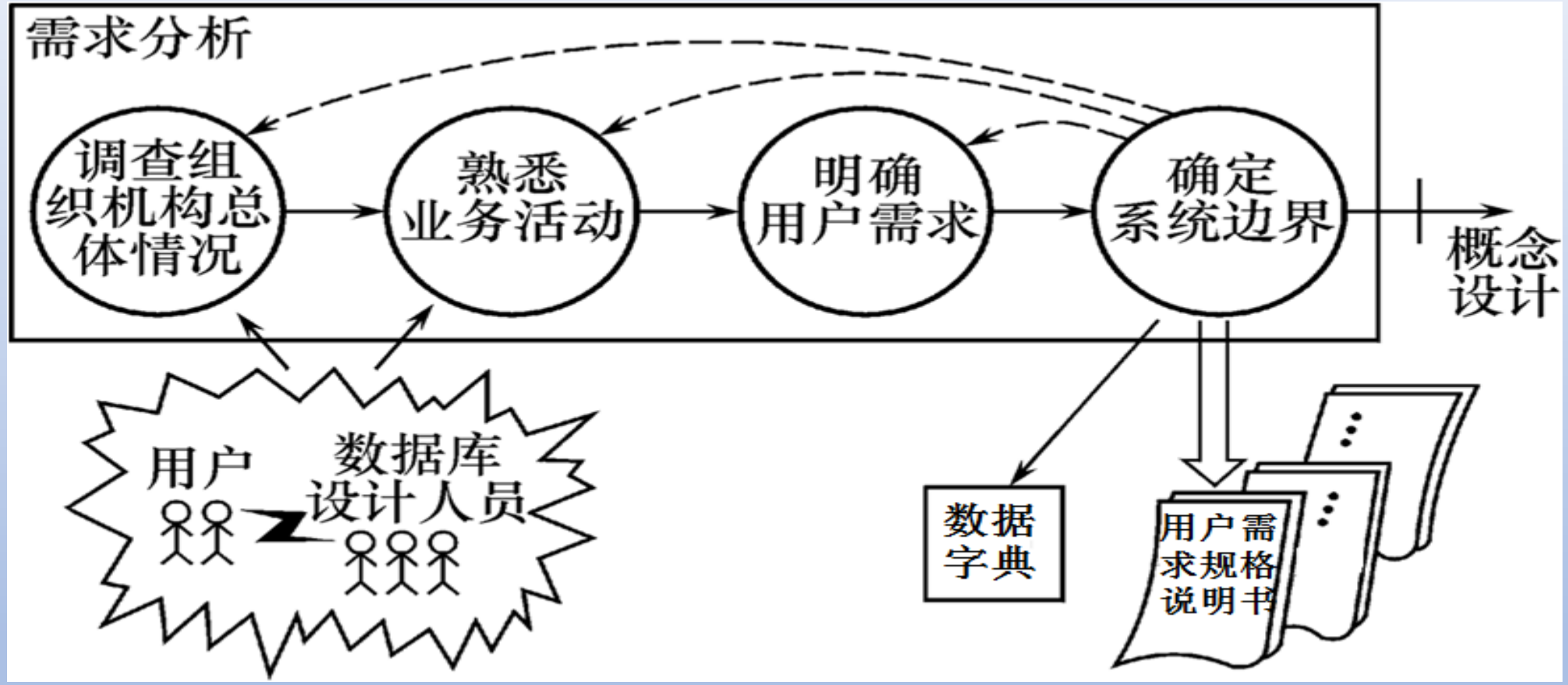
  - 用户需要从数据库中获得信息的内容与性质
  - 由信息要求可以导出数据要求，即在数据库中需要存储哪些数据

  （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

需求分析过程



需求分析过程

数据字典

- 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

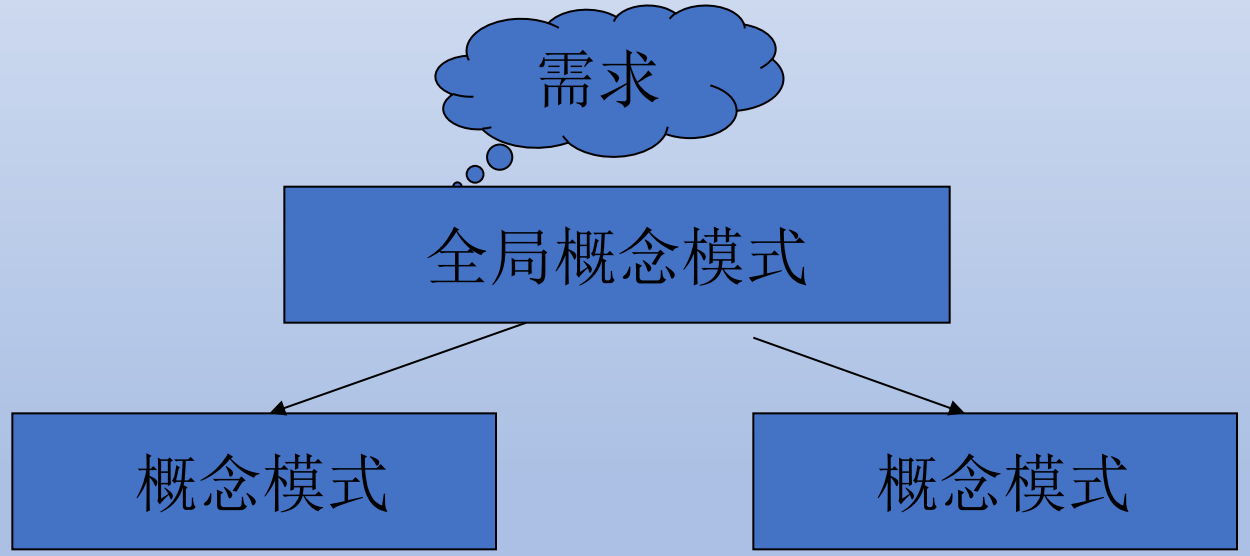
概念结构设计——ER图

概念结构设计的方法

- 自顶向下
- 自底向上：经常用的方法。
- 逐步扩张
- 混合策略

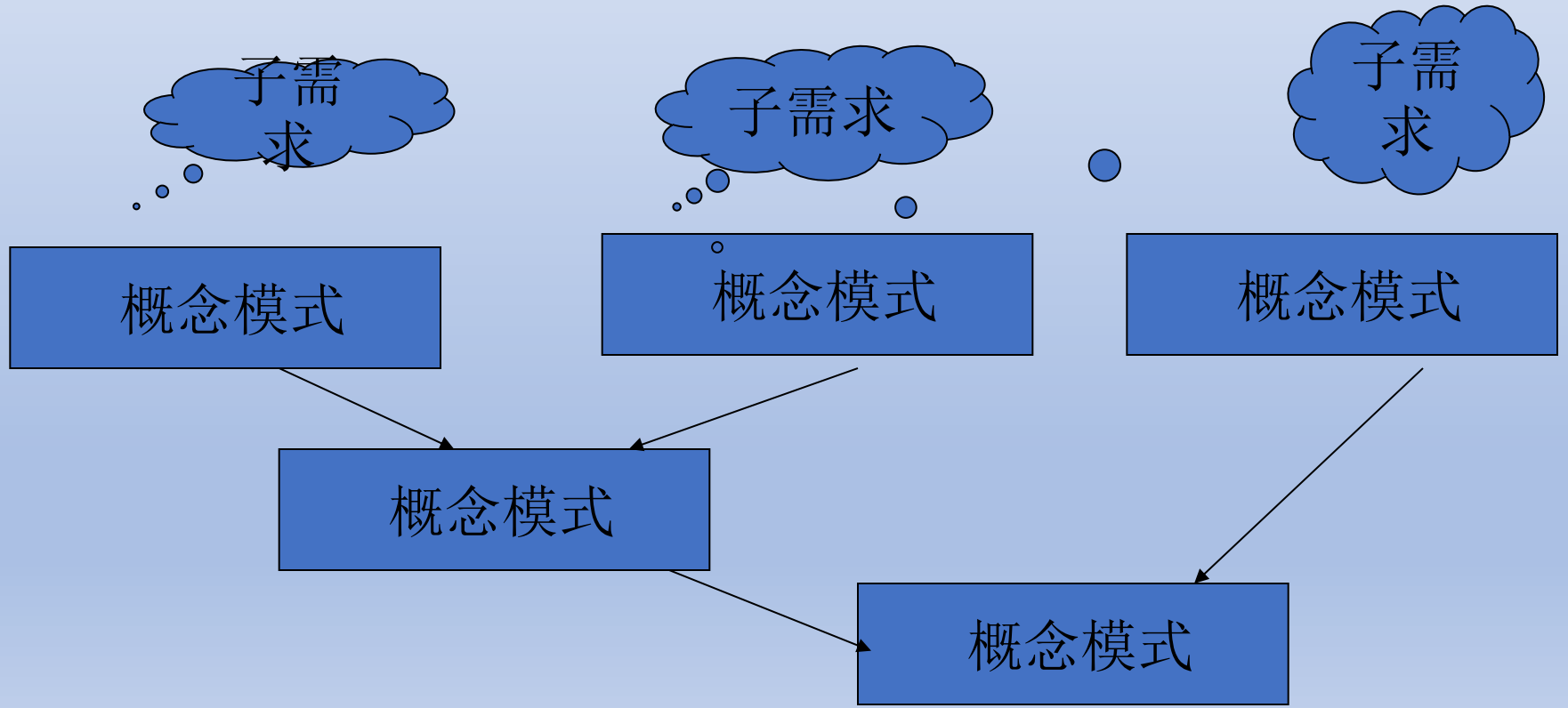
自顶向下

- 定义全局概念结构的框架,逐步细化.



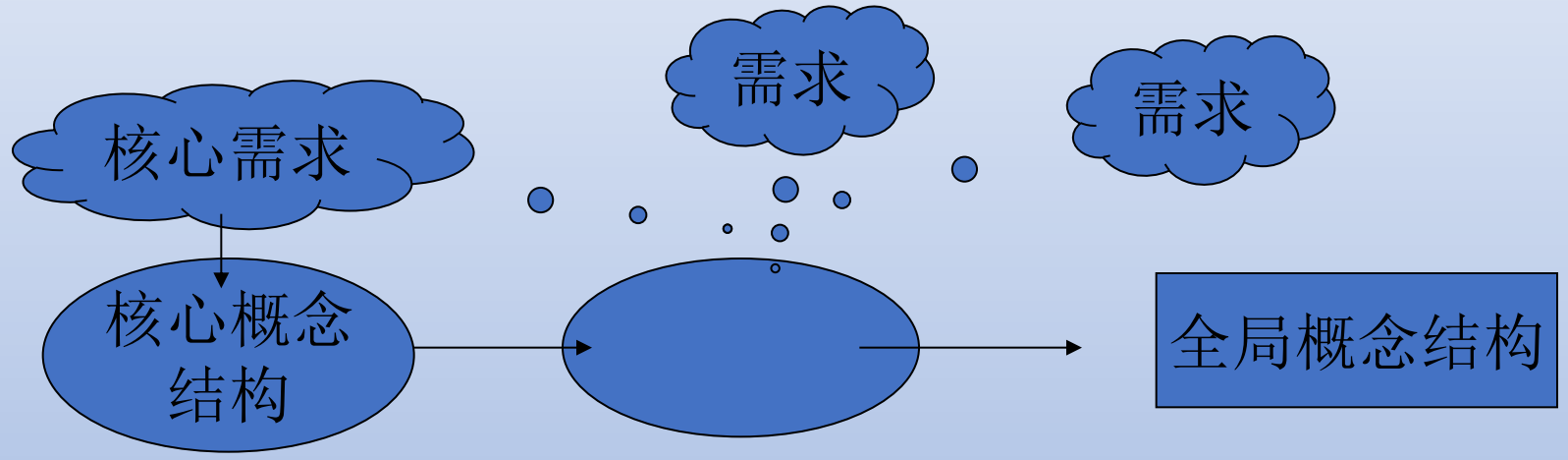
自底向上

- 定义局部应用的概念结构,再集成,形成全局概念结构.



其他策略

- 逐步扩展



- 混合策略

自底向上的ER图设计

- 选择局部应用，逐一设计分E-R图

  划分实体和属性的两条原则:

  （1）作为属性，不能再具有需要描述的性质，属性是不可分的数据项。

  （2）属性不能和其他实体具有联系。
- 合并分E-R图，生成初步的E-R图

  解决属性冲突，命名冲突，结构冲突。

自底向上的ER图设计（续）

- 消除冗余，设计基本的E-R图

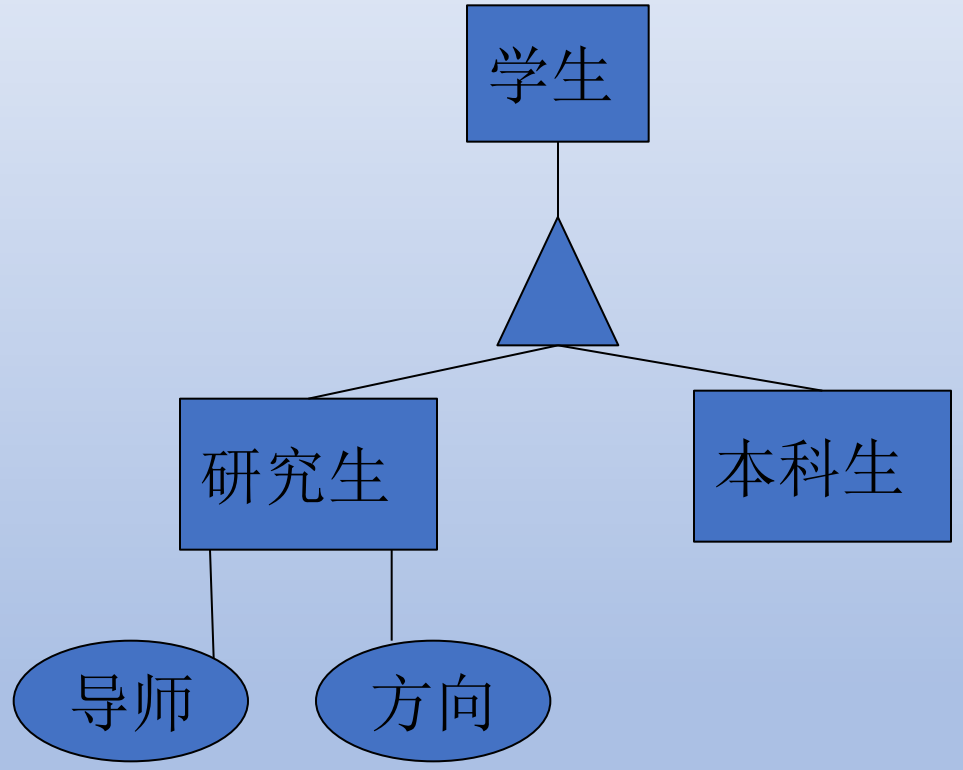
- 冗余数据: 可由基本数据导出.
- 冗余联系: 可由其它联系导出的联系.

数据抽象与局部视图设计

三种抽象:

- 分类: 定义某一类概念作为现实世界中的一组对象的类型 is a member of
- 聚集:定义某一类型的组成成分 is a part of
- 概括: 定义类型之间的一种子集关系 is a subset of

扩展的E-R模型（子类描述）



- 不相交与可重叠约束：父类中的实体是否最多属于一个子类
- 完备性约束：父类中的实体是否必须属于某一个子类

扩展的E-R模型 （part-of联系）

- 强联系 part-of:

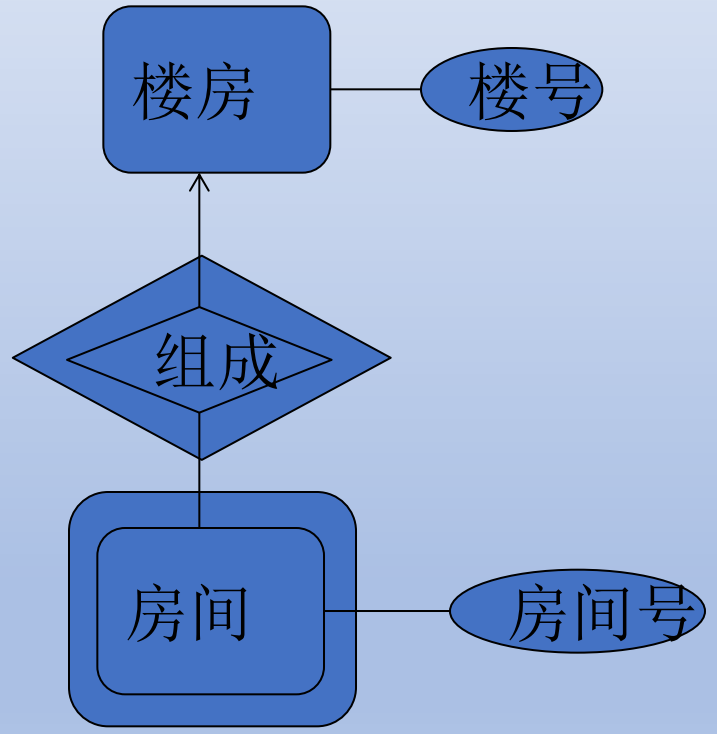
整体实体被破坏，部分实体不存在。如窗口由菜单，按钮等组成，一旦窗口不存在，菜单也没有了，称为强part-of.

- 弱联系 part-of

整体实体被破坏，部分实体还存在。

汽车与车轮的关系。汽车由车轮等组成，该组成关系是弱part-of。

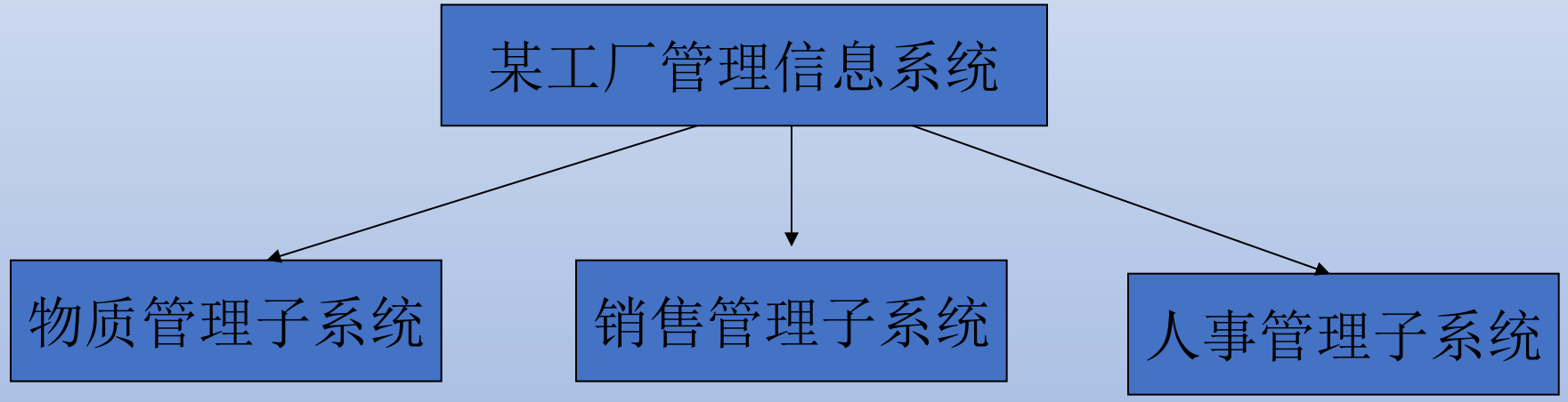
扩展的E-R模型 （弱实体集）



- 房间是一个弱实体集，因为它必须是在某一个楼里，房间号与楼号共同组成房间的码。
- 组成关系是一个弱关系。
- 房间的码：房间号+楼号

自底向上设计的一个实例:

- 先设计子系统的E-R图
- 合并



视图的合并

- 属性冲突: 属性类型,取值范围不同
- 命名冲突:同名异议,异名同异
- 结构冲突:

1. 同一对象具有不同的抽象(实体,属性)
2. 实体中属性个数,次序不同
3. 实体联系类型不同

- 消除冗余

E-R模型（实例）

- 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：

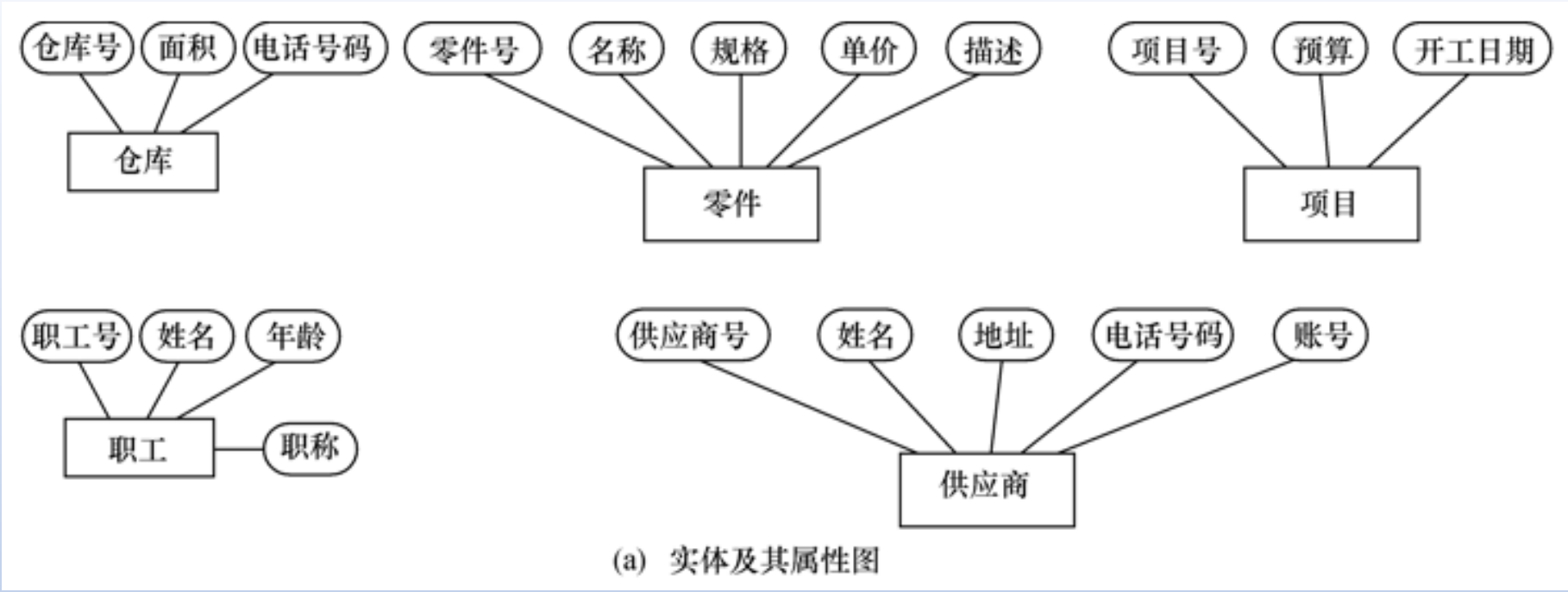
1. 一个仓库可以存放多种零件，一种零件可以存放在多个仓库中，因此仓库和零件具有多对多的联系。用库存量来表示某种零件在某个仓库中的数量。
2. 一个仓库有多个职工当仓库保管员，一个职工只能在一个仓库工作，因此仓库和职工之间是一对多的联系。

E-R模型（续）

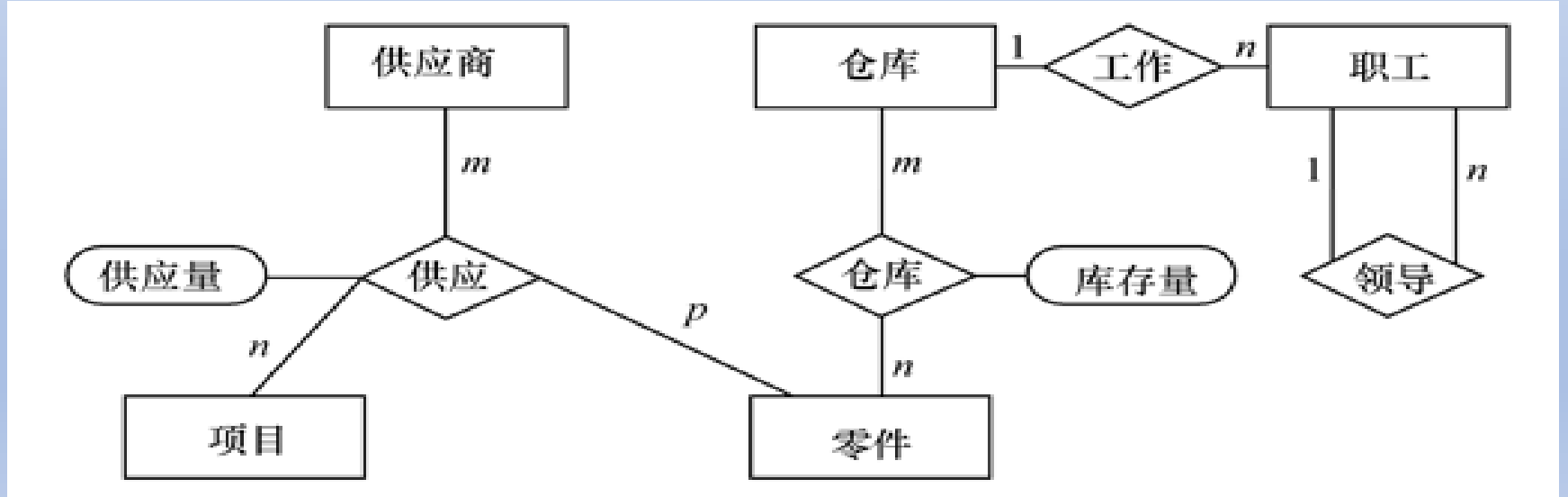
- 这些实体之间的联系如下（续）：

（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。

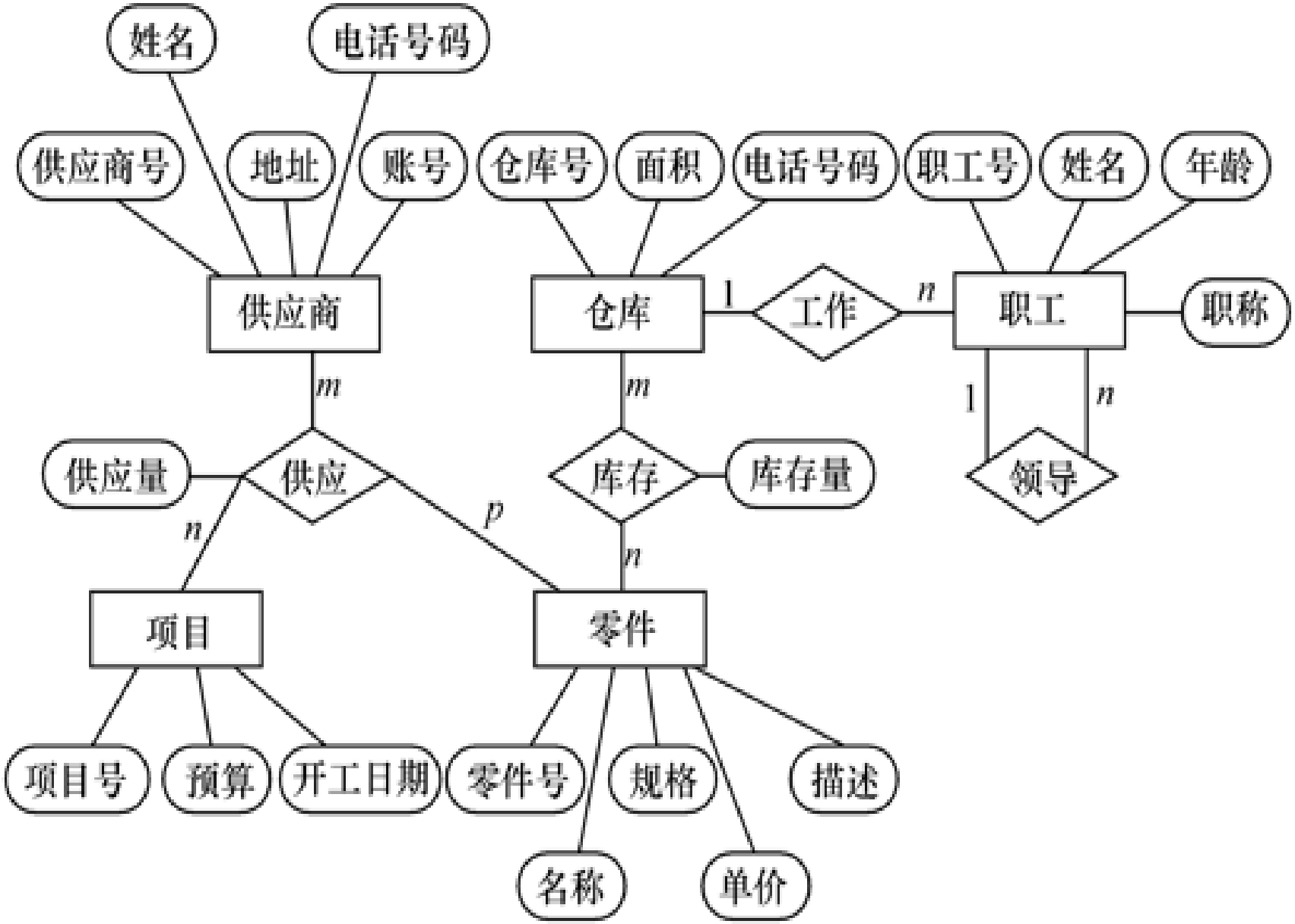
（4） 供应商、项目和零件三者之间具有多对多的联系。即一个供应商可以供给若干项目多种零件，每个项目可以使用不同供应商供应的零件，每种零件可由不同供应商供给。



(a) 实体及其属性图



(b) 实体及其联系图


姓名
电话号码
供应商号
地址
账号
仓库号
面积
电话号码
职工号
姓名
年龄
供应商
仓库
1
工作
n
职工
职称
m
m
1
n
供应量
供应
库存
库存量
领导
n
p
n
项目
零件
项目号
预算
开工日期
零件号
规格
描述
名称
单价

(c) 完整的实体-联系图

概念结构设计（实例）

- 某工厂管理信息系统的视图集成。

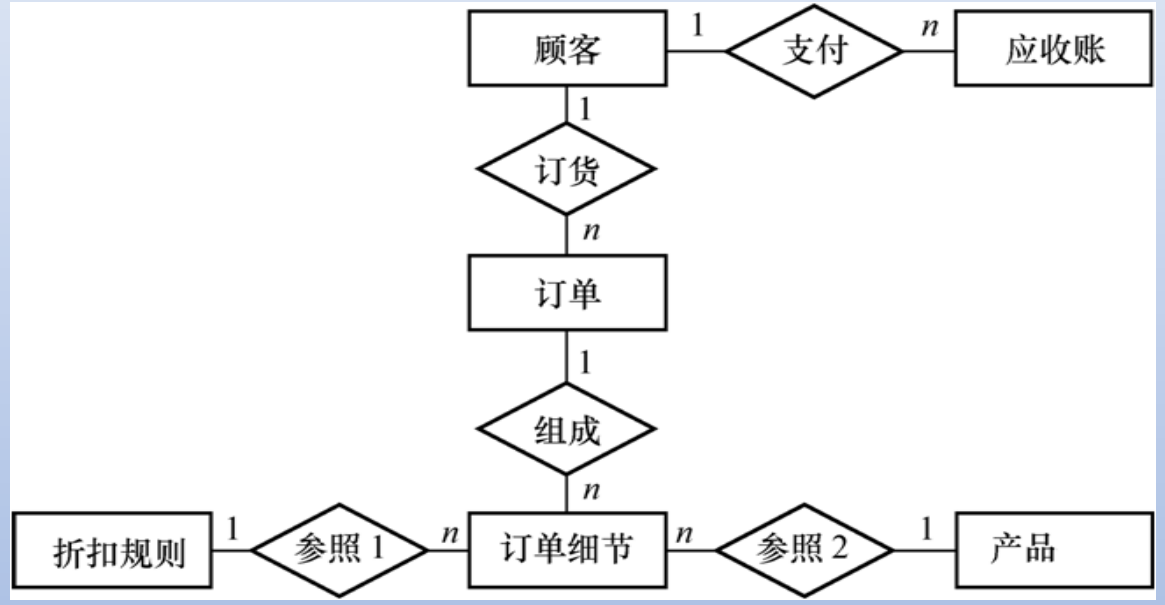


图7.23 销售管理子系统的E-R图

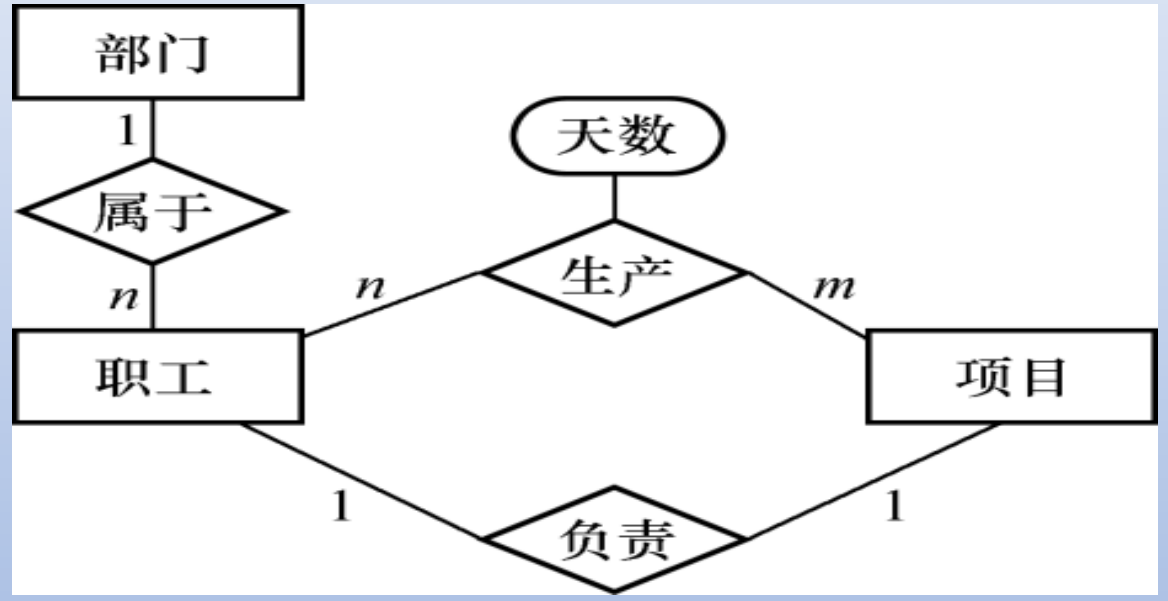
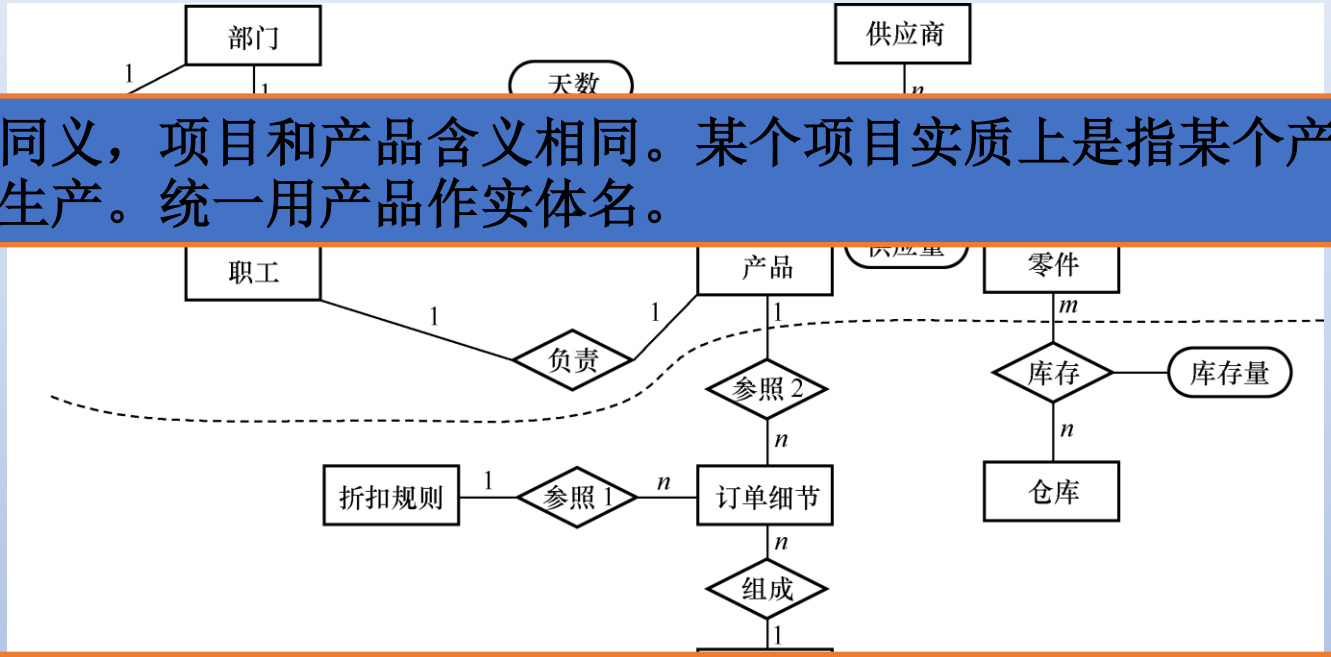


图7.27 劳动人事管理的分E-R图

- 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。

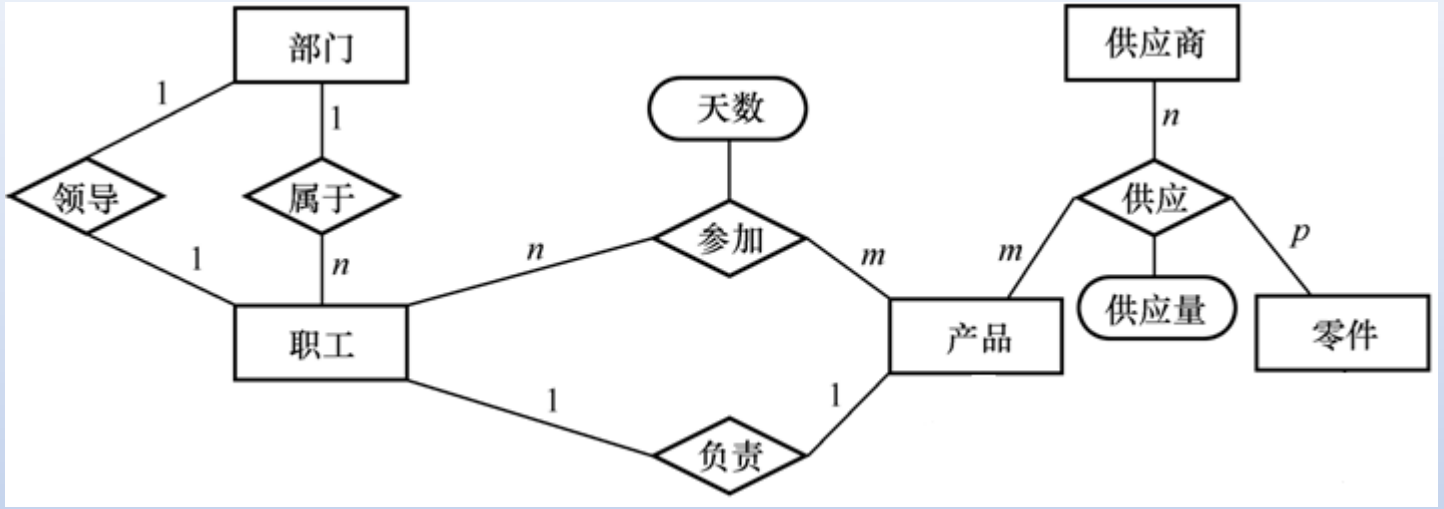
逻辑结构的设计——ER到关系

- 实体集→表
  - 属性→属性
  - 主码→主码

逻辑结构的设计——ER到关系

- 联系集→表
  - M：N，联系集必须存在
    - Key包含参与实体集的主码
  - N：1，联系集可以去除
    - 将1端的主码属性，引入N端的表中，成为外码
    - 外码参照1端的主码
  - 1：1，联系集可以去除
    - 两个实体集可合并
    - 或一端的主码引入到另一端中，作为外码

E-R图向关系模型的转换（实例）



- 部门（部门号，部门名，经理的职工号，…）
- 职工（职工号、部门号，职工名，职务，…）
- 产品（产品号，产品名，产品组长的职工号，…）
- 供应商（供应商号，姓名，…）
- 零件（零件号，零件名，…）
- 职工工作（职工号，产品号，工作天数，…）
- 供应（产品号，供应商号，零件号，供应量）

数据库的物理设计

分为两步：

- 确定数据库的物理结构：设计存取方法和存贮结构
- 对物理结构进行评价：时间和空间效率

→ 根据不同DBMS，确定系统的配置

数据库的物理设计：需要考虑

- 数据库的查询事务

查询关系，查询条件所涉及的属性，连接条件所涉及的属性，投影属性

- 数据库的更新事务

被更新的关系，更新操作条件所涉及的属性，以及要改变的属性值

数据库的物理设计：存取方法

存取方法包括：

- 索引方法：决定那些属性要建立索引，B+树
- HASH方法：通过计算找到实际的存放位置
- 聚簇方法：存放在连续的物理块

B+树索引存取方法的选择

- 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

HASH存取方法的选择

- 选择Hash存取方法的规则
  - 如果一个关系的属性主要出现在等值连接条件中或主要出现在等值比较选择条件中，而且满足下列两个条件之一
    - 该关系的大小可预知，而且不变；
    - 该关系的大小动态改变，但所选用的数据库管理系统提供了动态Hash存取方法。

聚簇存取方法的选择

- 什么是聚簇
  - 为了提高某个属性（或属性组）的查询速度，把这个或这些属性（称为聚簇码）上具有相同值的元组集中存放在连续的物理块中称为聚簇。
  - 该属性（或属性组）称为聚簇码（cluster key）
  - 许多关系型数据库管理系统都提供了聚簇功能
  - 聚簇存放与聚簇索引的区别

聚簇存取方法的选择（续）

- 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。
  - 在一个基本表上最多只能建立一个聚簇索引
- 聚簇索引的适用条件
  - 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

聚簇存取方法的选择（续）

- 选择聚簇存取方法
  - 设计候选聚簇

    （1）常在一起进行连接操作的关系可以建立组合聚簇

    （2）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簇；

    （3）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簇。

数据库的物理设计: 存贮结构

- 综合考虑*存取时间*，*存贮空间利用率*和*维护代价*

- 确定数据的存放位置
- 确定系统的配置

- 评价物理结构

确定数据的存放位置

- 基本原则
  - 根据应用情况将
    - 易变部分与稳定部分分开存放
    - 经常存取部分与存取频率较低部分分开存放
- [例]
  - 可以将比较大的表分别放在两个磁盘上，以加快存取速度，这在多用户环境下特别有效。
  - 可以将日志文件与数据库对象（表、索引等）放在不同的磁盘以改进系统的性能。

确定系统配置

- 数据库管理系统一般都提供了一些存储分配参数
  - 同时使用数据库的用户数
  - 同时打开的数据库对象数
  - 内存分配参数
  - 缓冲区分配参数（使用的缓冲区长度、个数）
  - 存储分配参数
  - 物理块的大小
  - 物理块装填因子
  - 时间片大小
  - 数据库的大小
  - 锁的数目等

评价物理结构

- 对数据库物理设计过程中产生的多种方案进行评价，从中选择一个较优的方案作为数据库的物理结构。
- 评价方法
  - 定量估算各种方案
    - 存储空间
    - 存取时间
    - 维护代价
  - 对估算结果进行权衡、比较，选择出一个较优的合理的物理结构。

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
  - 物理结构设计
  - 数据库实施
  - 数据库运行维护
  - 设计过程中往往还会有许多反复

课堂练习：

根据交大选课系统，以及下面需要完成的功能，我们设计一个选课系统，由学生，课程，老师等组成，请画出其ER模型，并设计各个模式，以及各种必要的完整性约束。假设一门课程可以有多个教师上，一个教师可以上多门课程。系统主要针对学生，完成功能如下：

1）学生选课

2）学生查询成绩