

# 数据库概论

## The Introduction of Database System

### 第五章 数据库设计



中国矿业大学计算机学院



中国矿业大学数据库原理精品课程

# 主要内容

- 需求分析
- 概念结构设计
- 逻辑结构设计
- 数据库的物理设计
- 数据库的实施
- 数据库运行和维护



# 5.1 数据库设计概述

## ● 数据库设计

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，使之能有效地存储数据，满足各种用户的应用需求。

- 信息需求：数据库内容及结构的要求，静态
- 处理需求：数据库要进行的数据处理，动态

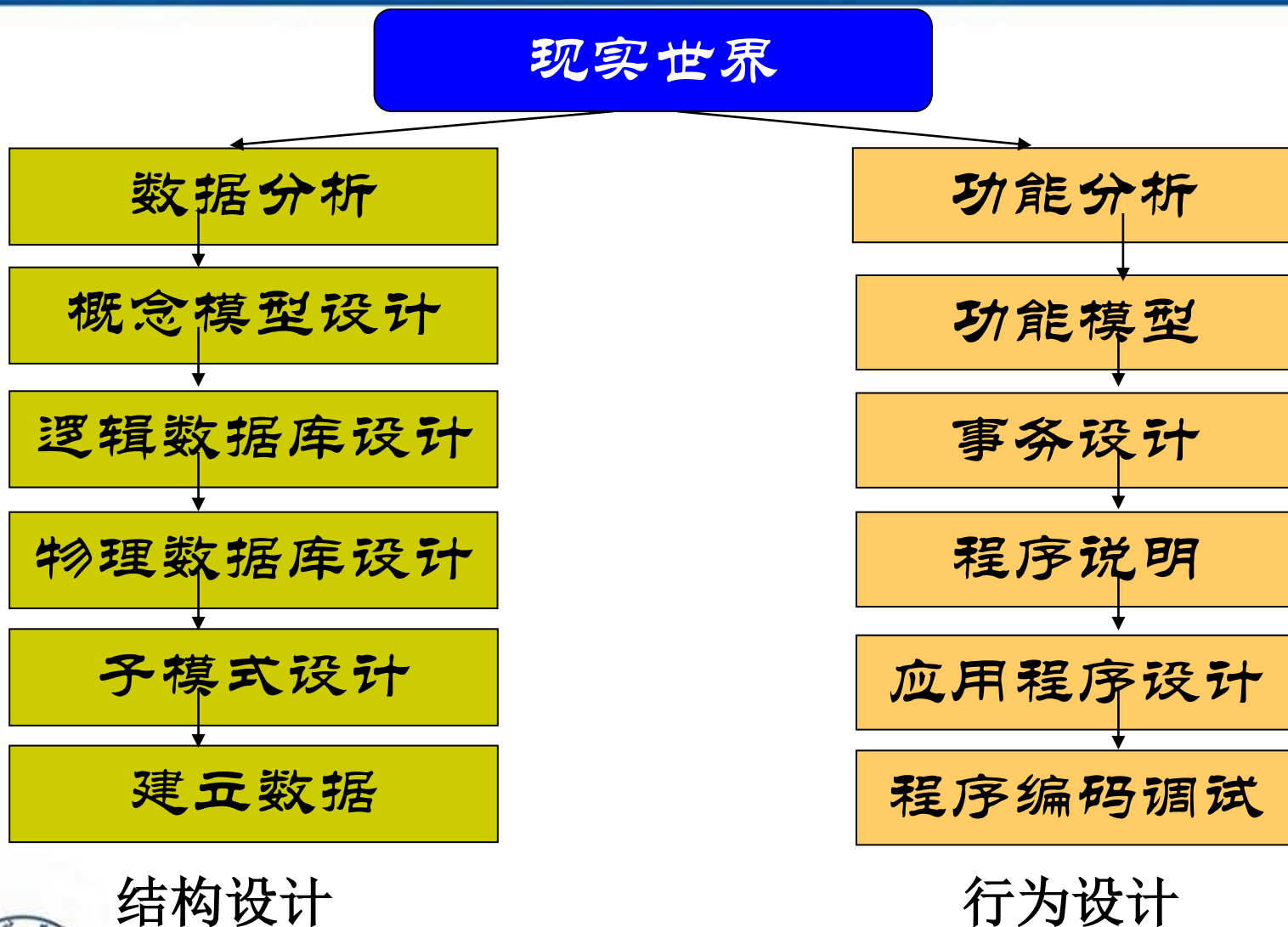


# 数据库设计人员应具备的技术和知识

- 1、计算机科学基础知识和程序设计技术
- 2、DB基本知识和DB设计技术
- 3、软件工程的原理和方法
- 4、应用领域的知识



## 5.1.2 数据库设计的内容



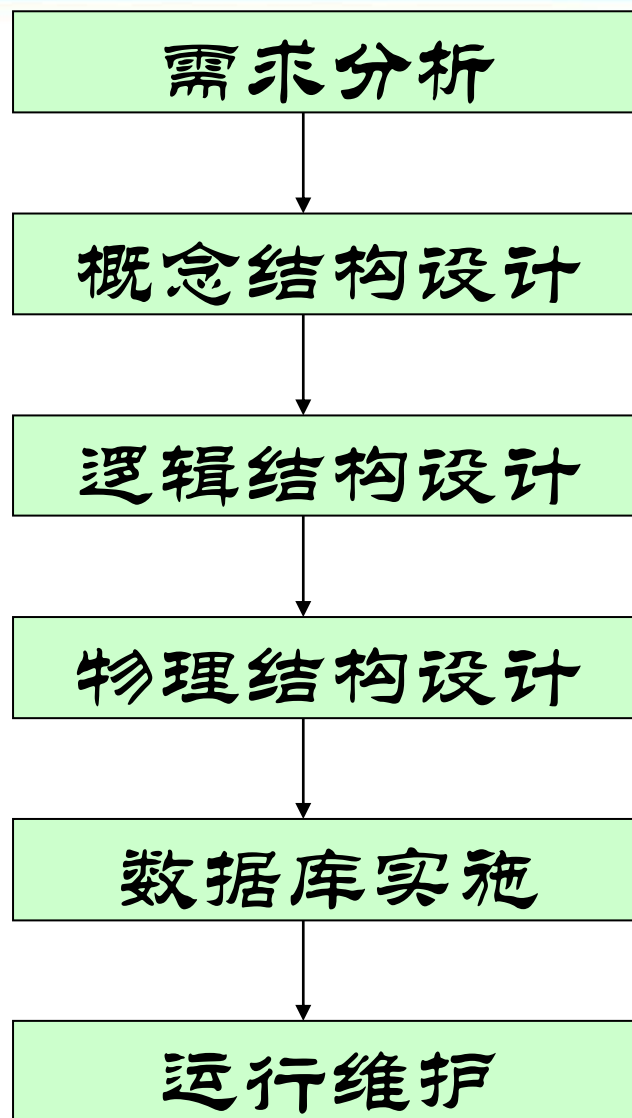
## 5.1.3 数据库的设计方法

- 直观设计法
- 规范设计法
- 计算机辅助设计法
- 自动化设计法。

基于实体联系的设计方法、基于3NF的数据库设计方法、基于视图概念的数据库设计方法等。



## 5.1.4 数据库设计的基本步骤



## 5.1.4 数据库设计的基本步骤

- **需求分析：**是整个设计过程的基础。

(1) 信息需求分析 (2) 操作需求分析

- **概念结构设计：**将需求分析的结果用一种工具进行形式化的定义和描述，是整个DB设计的关键。

如：E-R图

- **逻辑结构设计：**概念模型→数据模型

根据需求分析和概念设计的结果，选择合适的数据库模型，并选用某一设计方法构造一数据库模式。





## 5.1.4 数据库设计的基本步骤

- 物理设计：选择合适的物理结构。

包括存储结构和存取方法

- 实施阶段：根据逻辑设计和物理设计建立数据库；编制和调试应用程序；试运行；

- DB运行和维护阶段

根据运行记录对DB进行评价

根据评价对DB调整和修改



## 5.2 需求分析

需求分析的任务是通过详细调查现实世界要处理的对象，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，然后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善，不能仅仅按当前应用需求来设计数据库。



## 5.2.1 需求分析的任务

- 信息要求：DB中需存储的数据
- 处理要求：
  - 用户要求的处理功能
  - 对各处理的响应时间的要求
- 安全性与完整性要求
- 需求的优先级

确定用户的最终需求是  
非常困难的！



# 经验：

- 1、尽可能让自己成为用户；
- 2、倾听用户需要，理解用户需求；  
what-why-how
- 3、听听用户的解决方案；
- 4、了解需求发生的频度和强度；
- 5、模拟用户操作，补全缺失流程；



## 5.2.2 需求分析的方法和过程

### 1. 调查用户需求的具体步骤:

- 分析用户活动，产生业务流程图。
- 确定系统范围，产生业务范围图。
- 分析用户活动及所设计的数据，产生数据流图
- 分析系统数据，产生数据字典



## 2、需求分析的常用方法

- 跟班作业
- 开调查会
- 请专人介绍
- 询问
- 调查表
- 查阅记录

### 强调:

- 数据的收集要尽可能详尽，全面反映用户需求。
- 要考虑可能的扩充和改变。
- 必须强调用户的参与。



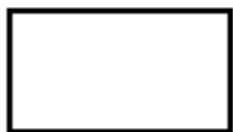
## 5.2.3 需求分析常用工具

### 1. 数据流图 (DFD)

DFD由四种基本符号组成。如下图所示。



数据加工(数据变换)



数据源或终点(外部实体)



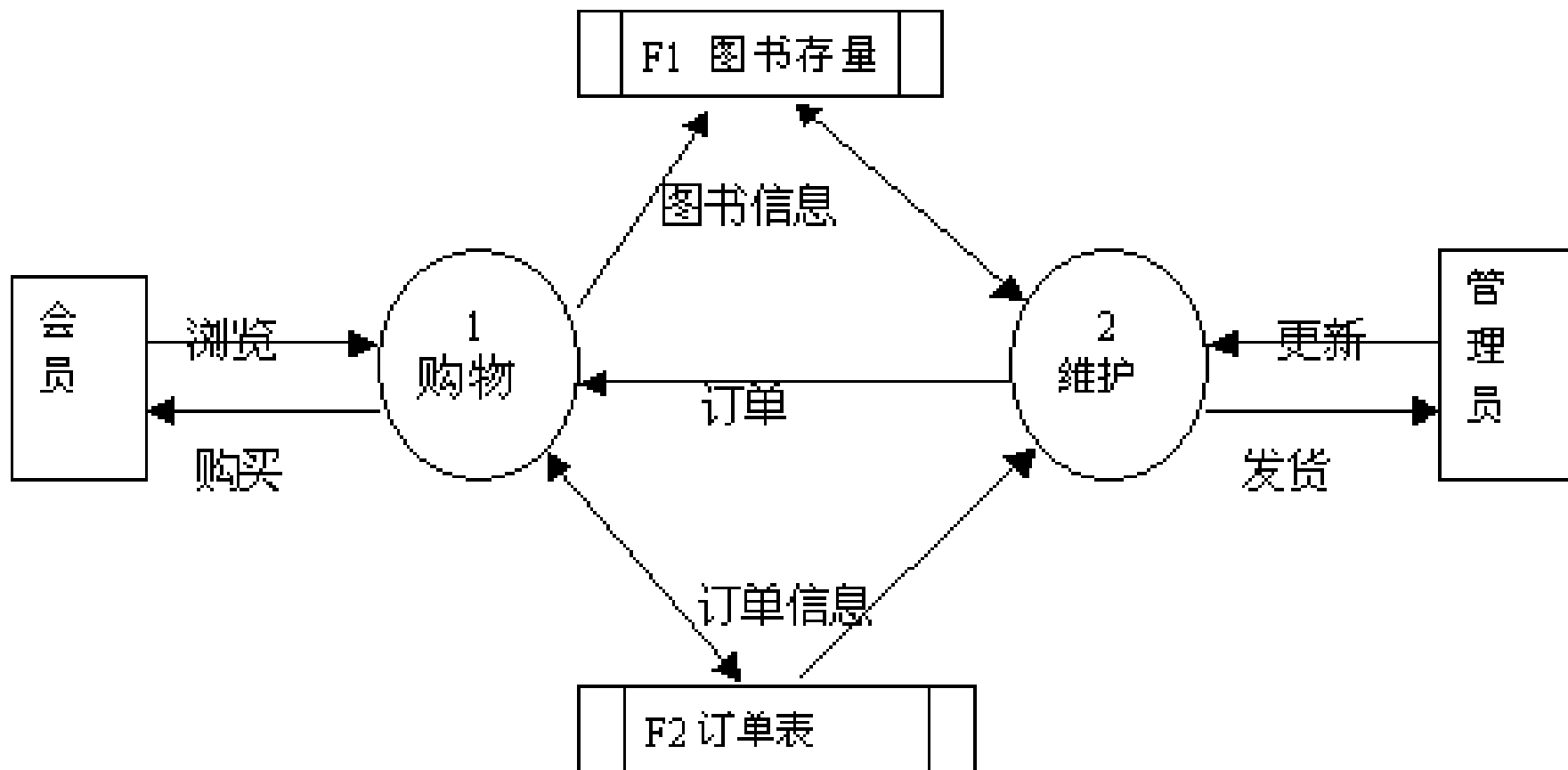
数据流



数据存储文件



# 例：网上书店的数据流图



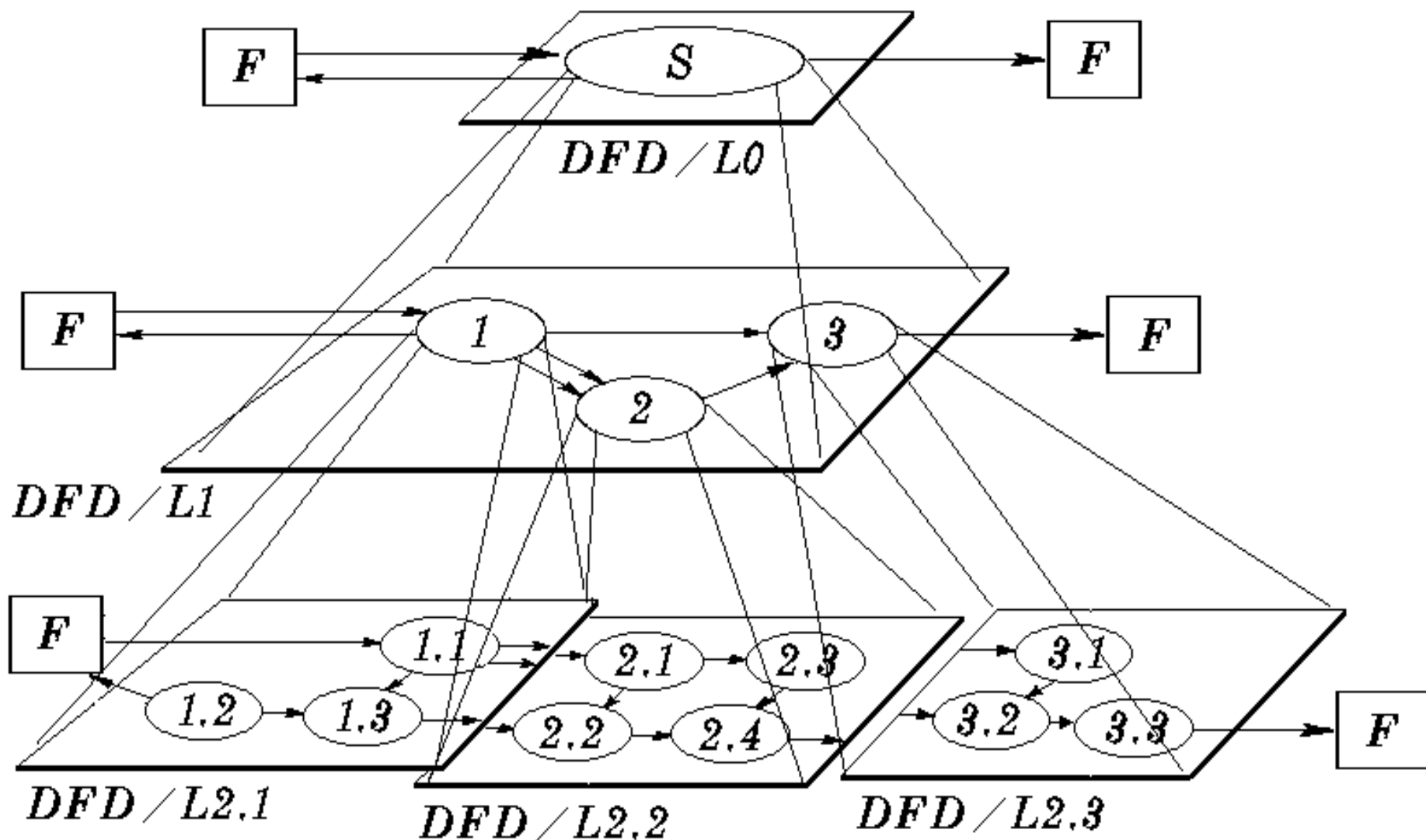


# 分层的数据流图

为了表达数据处理过程的数据加工情况，需要采用**层次结构**的数据流图。按照系统的层次结构进行**逐步分解**，并以分层的数据流图反映这种结构关系，能清楚地表达和容易理解整个系统



# 分层数据流图



- 在多层数据流图中，**顶层流图**仅包含一个**加工**，它代表被开发系统。它的输入流是该系统的输入数据，输出流是系统所输出数据
- **底层流图**是指其**加工不需再做分解**的数据流图，它处在最底层
- **中间层流图**则表示**对其上层父图的细化**。它的每一加工可能继续细化，形成子图。



# 仓库管理信息系统

## ——DFD实例



# 系统总体概述

仓库管理的物资主要是企业生产所需要的各种设备。

主要工作流程：

(1) 进货是经检查合同确定认为有效委托以后，进行验收入库，填写入库单，进行入库登记。

(2) 企业各个部门根据所需要的物资设备总额和部门生产活动需要提出物资需求申请。计划员根据整个企业的需求开发出物资设备出库单，仓库管理员根据出库单核对发放设备。

(3) 设备使用完毕需要及时归还入库，填写入库单。



# 入库单

No. 0000001

部门:

订单号:

日期: 20 年 月 日

序号	物料编号	品名	规格	单位	入库数量	实收数量	备注
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

一式三联 (白联仓库·红联财务·黄联生产)

核准:

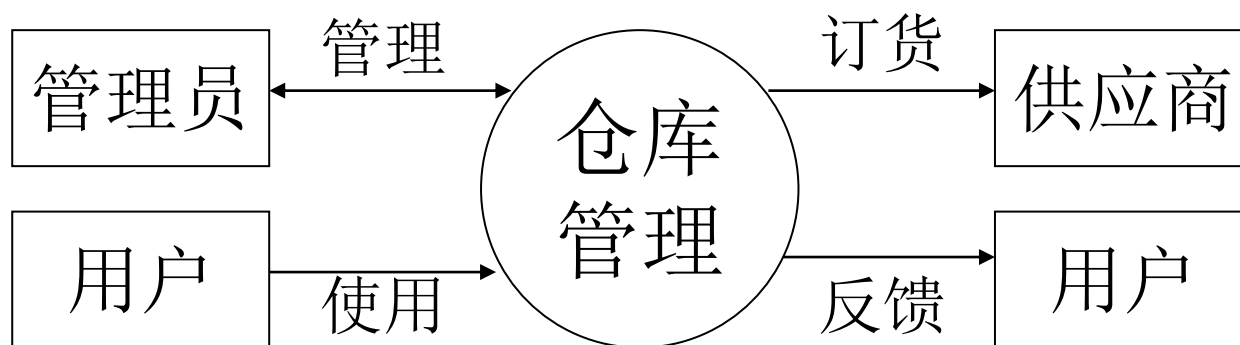
审核:

仓库:

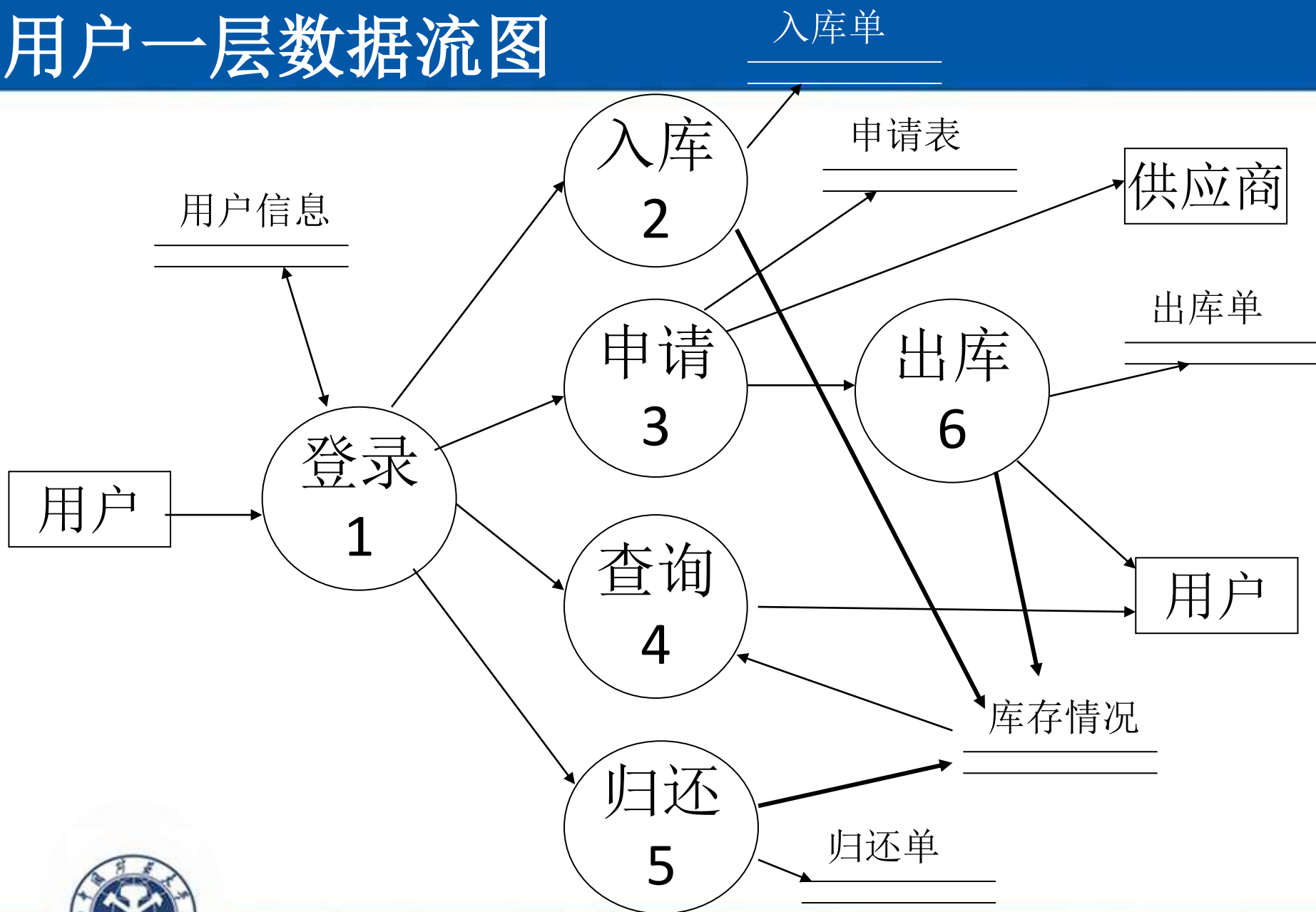
入库人:



# 顶层数据流图

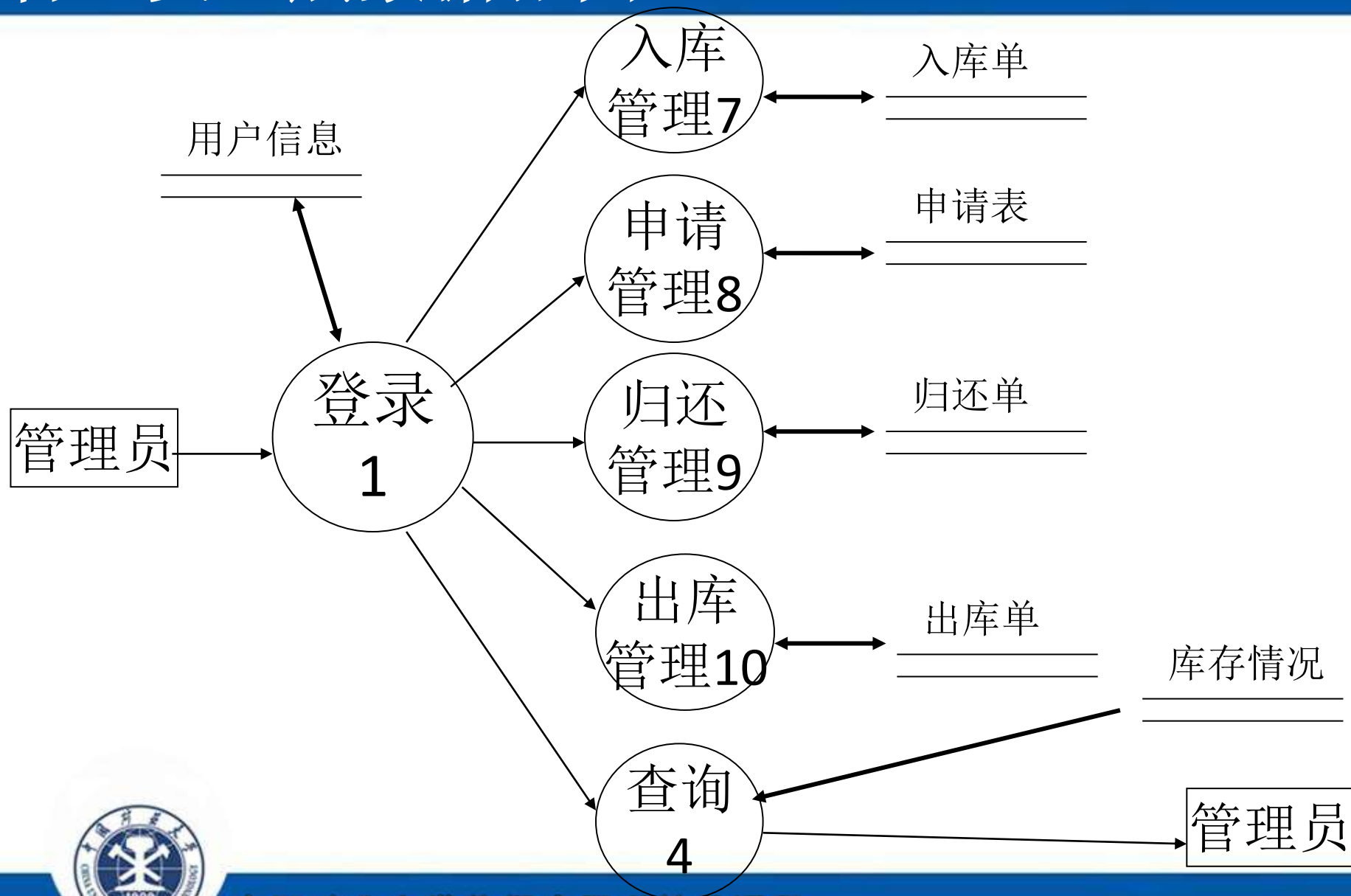


# 用户一层数据流图

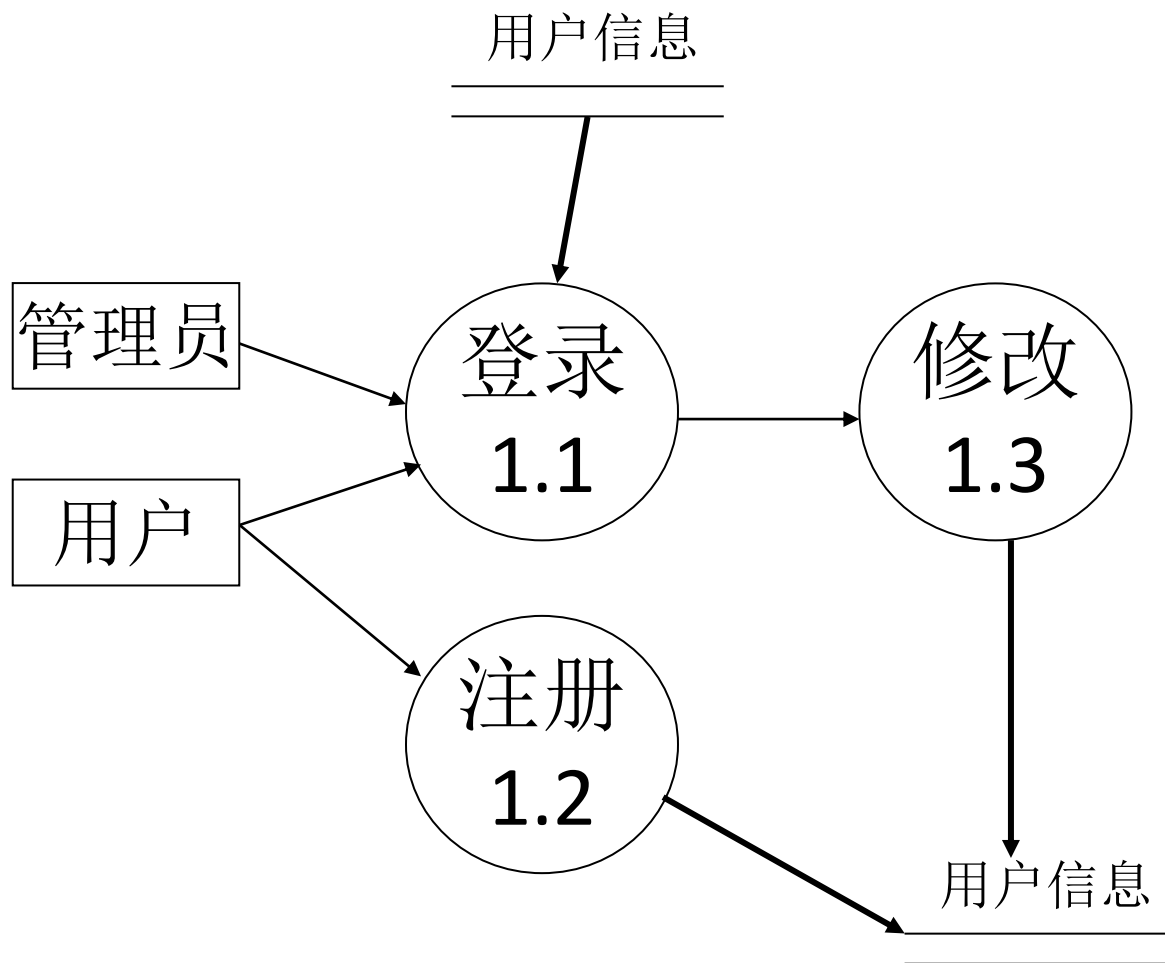




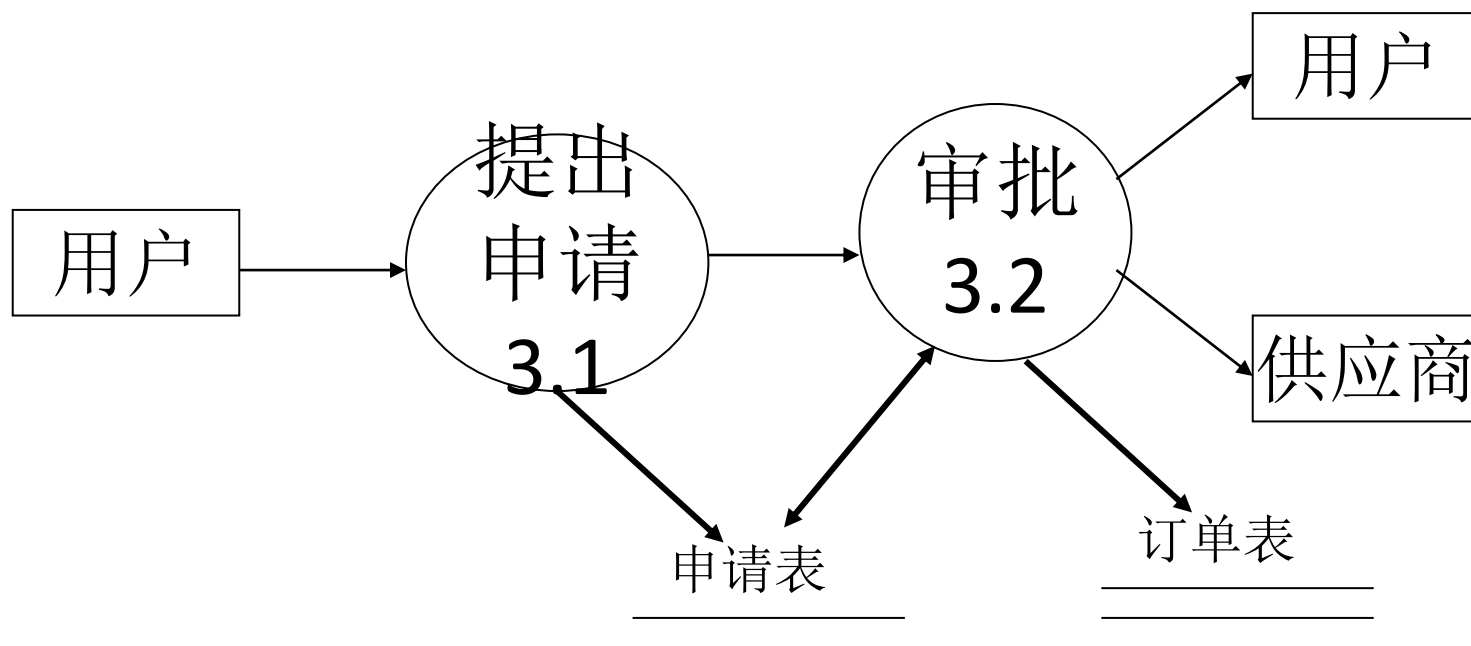
# 管理员一层数据流图



# 二层数据流图



# 二层数据流图



# 检查和修改数据流图的原则

- 数据流图上所有图形符号只限于前述四种基本图形元素
- 数据流图的主图必须包括前述四种基本元素，缺一不可
- 数据流图的主图上的数据流必须封闭在外部实体之间
- 每个加工至少有一个输入数据流和一个输出数据流



- 在数据流图中，需按层给加工框编号。编号表明该加工所处层次及上下层的亲子关系。
- 规定任何一个数据流子图必须与它上一层的一个加工对应，两者的输入数据流和输出数据流必须一致，此即父图与子图的平衡。



# 编号以及父子图平衡问题

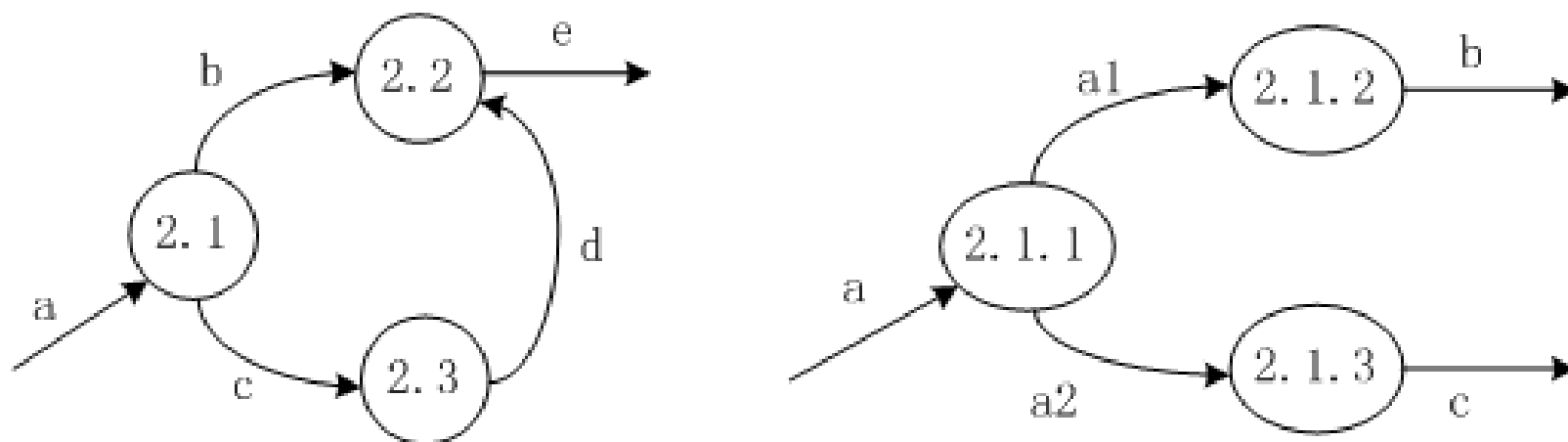


图3-5 父图与子图



- 图上每个元素都必须有名字
- 数据流图中不可夹带控制流
- 初画时可以忽略琐碎的细节，以集中精力于主要数据流

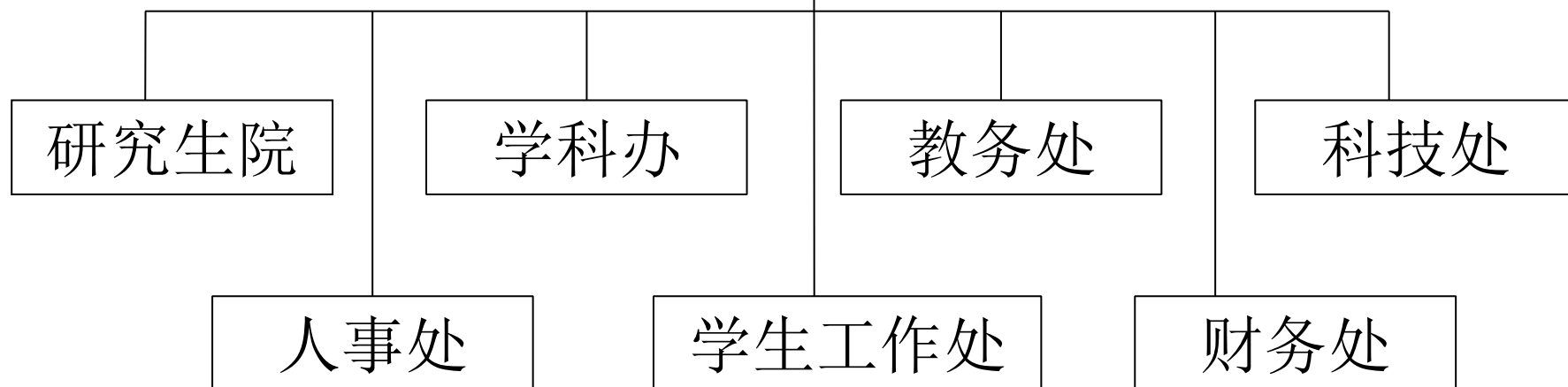


## 学校管理信息系统



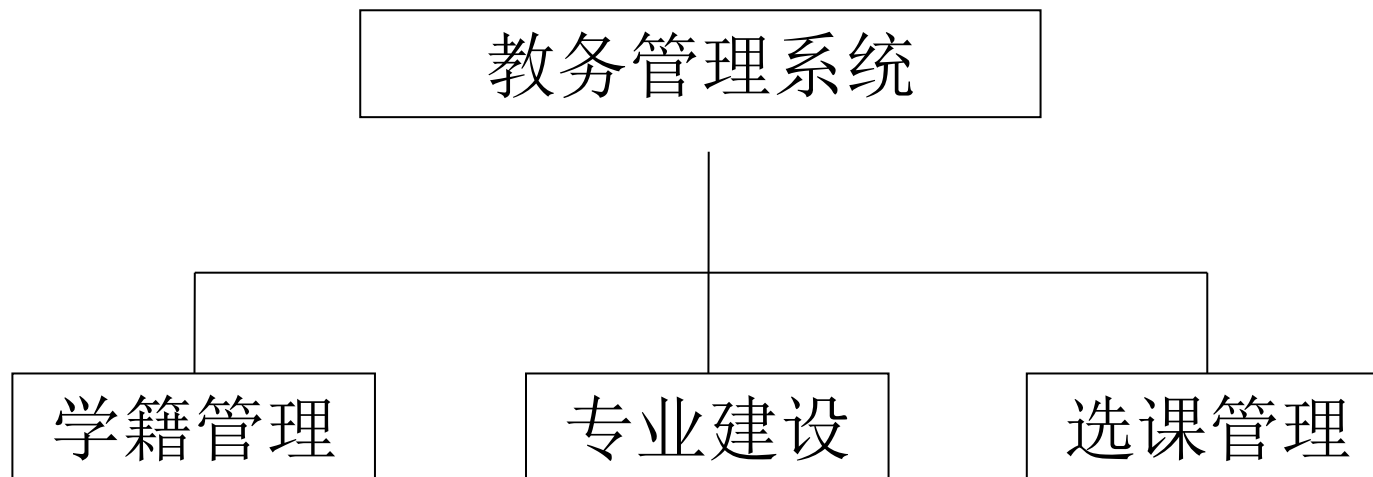


# 学校管理信息系统



学校组织结构图





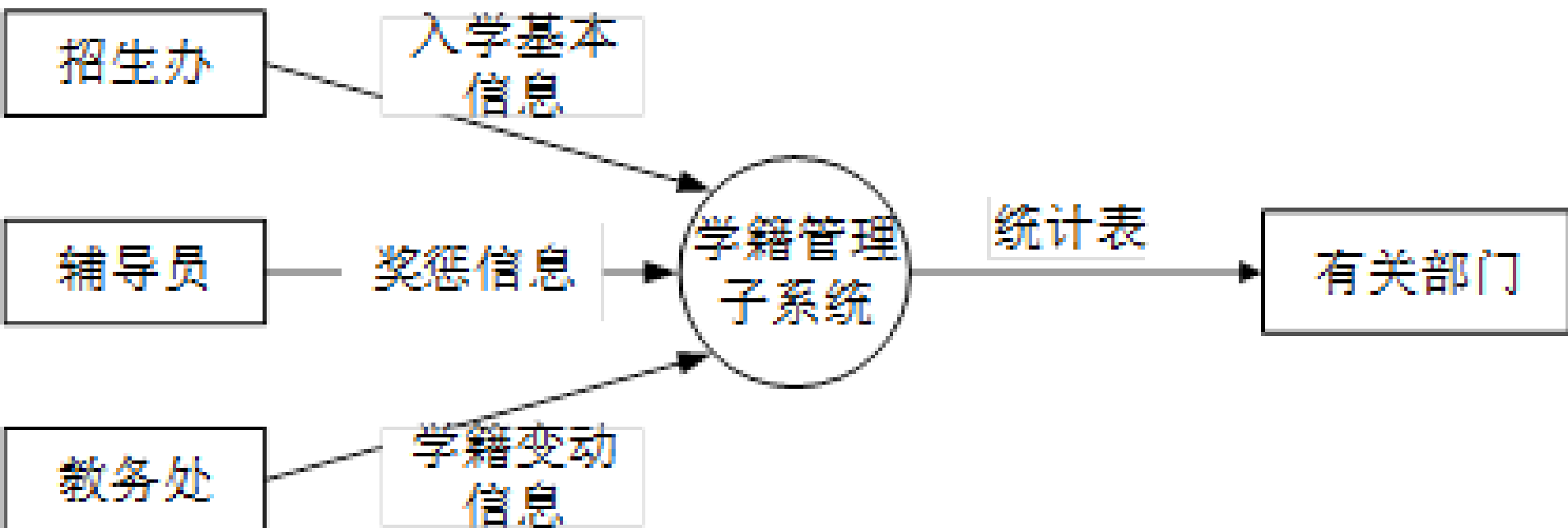
教务管理系统功能结构图



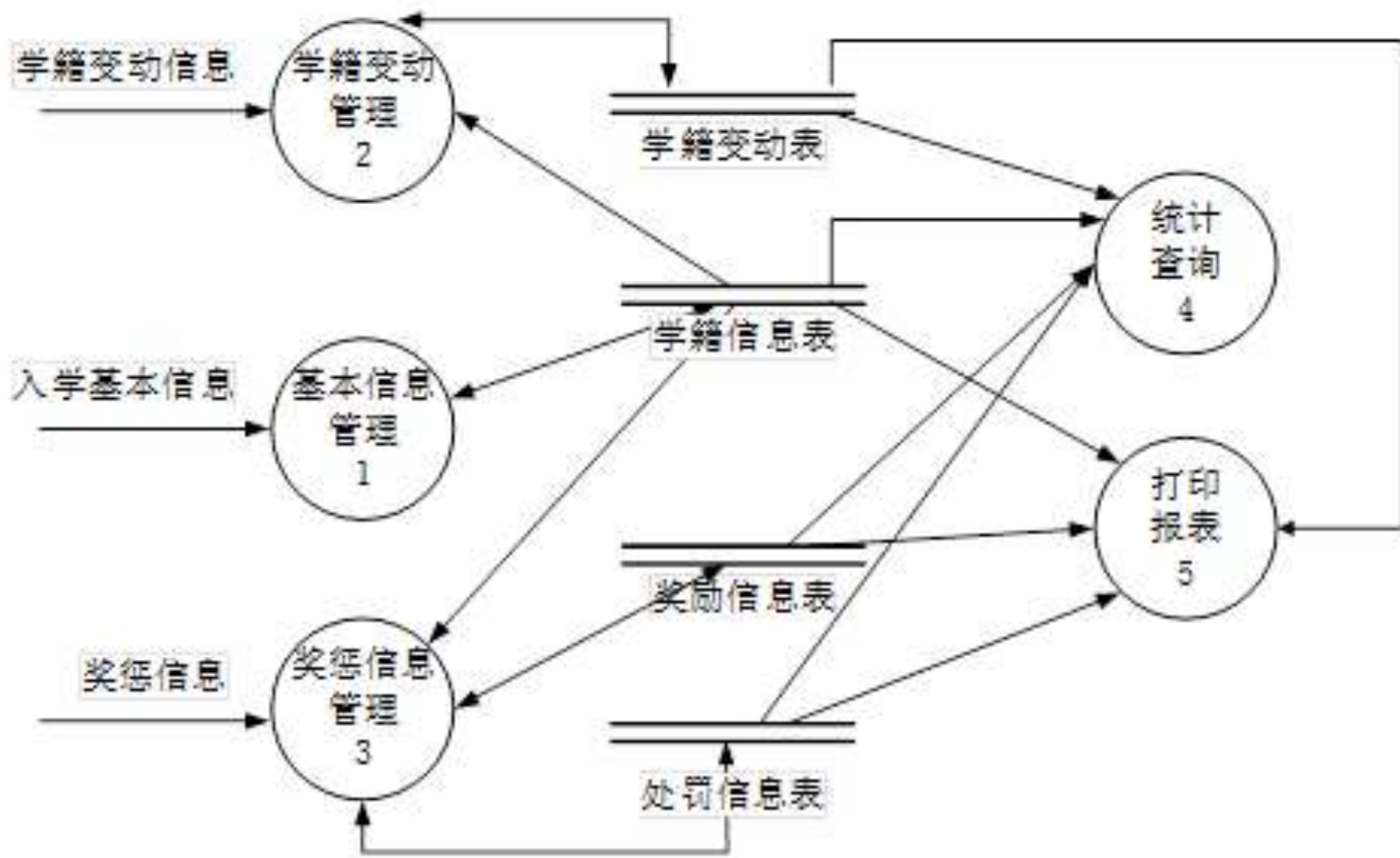
# 学籍管理模块数据流图



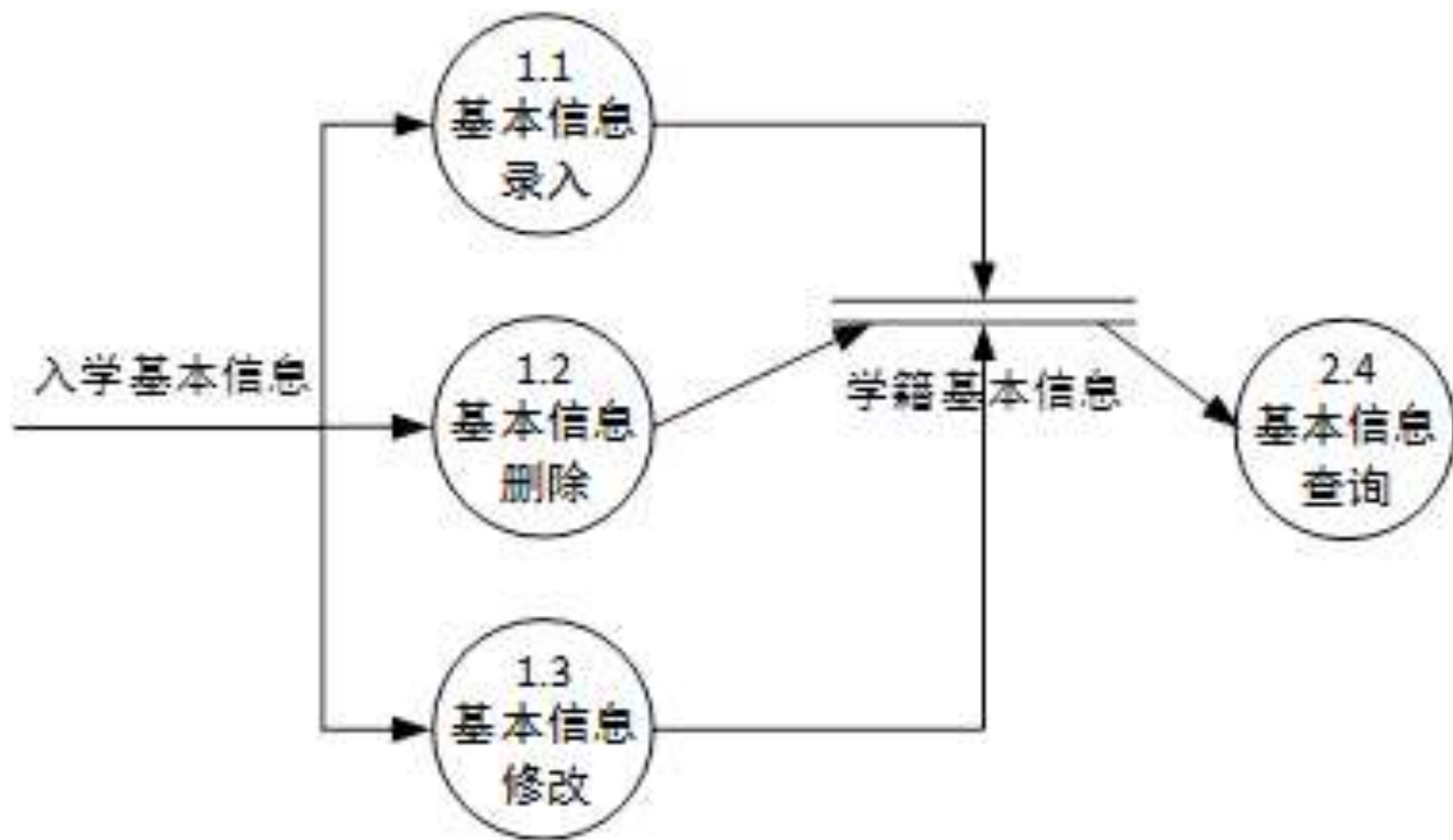
# 学籍管理子系统顶层数据流图



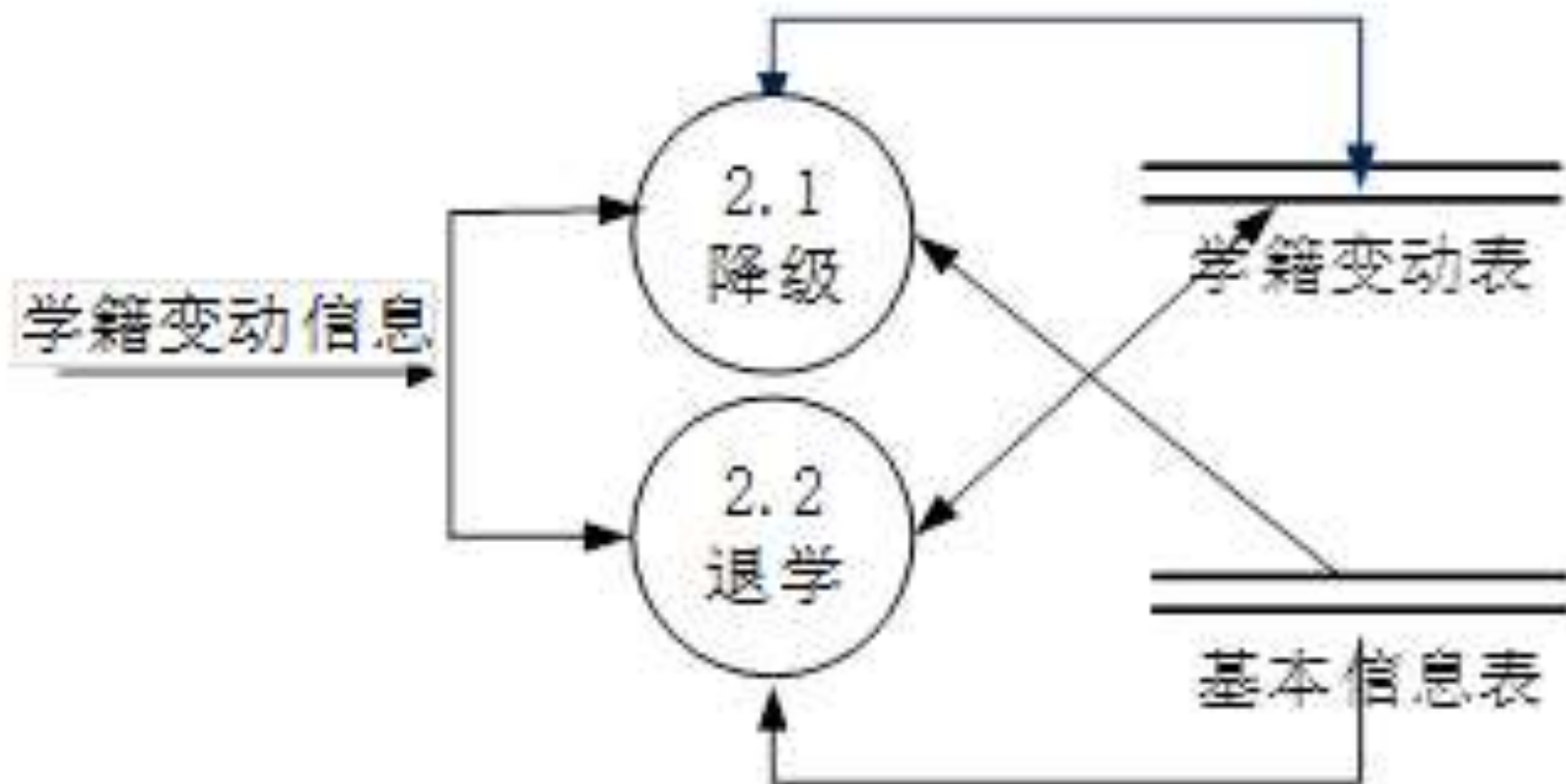
# 学籍管理子系统一层数据流图



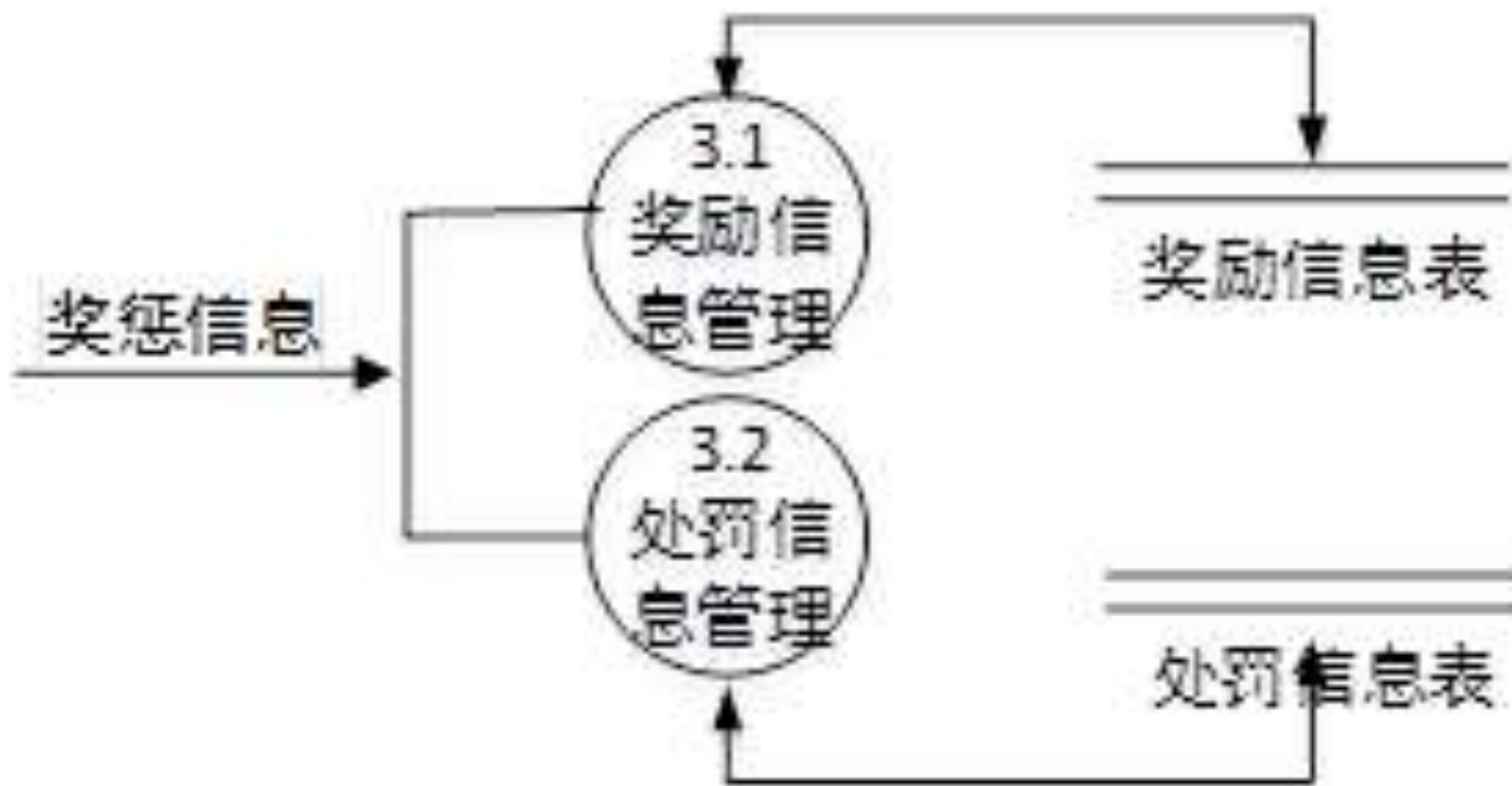
# 基本信息管理二层数据流图



# 学籍变动管理二层数据流图



# 奖惩信息管理二层数据流图





## 2. 数据字典 (DD)

数据字典是关于数据的数据库，它是对数据流程图上各个元素作出详细的定义和说明。

数据字典	数据项
	数据结构
	数据流
	数据存储
	处理过程



# (1) 数据项

➤ 数据项是不可再分的数据单位。

➤ 数据项的描述=

{数据项名, 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}



**数据项：**学号

**含义说明：**唯一标识某个学生

**别名：**学生编号

**类型：**字符型

**长度：**10

**取值范围：**000000000000至999999999999

**取值含义：**前两位标别该学生所在院系，3、4位表示所在年级，5、6位表示所在班级，后4位按顺序编号

**与其他数据项的逻辑关系：**



## (2) 数据结构

- 数据结构反应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。
- **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}  
}



**数据结构：**学生

**含义说明：**是学籍管理模块的主体数据结构，  
定义了一个学生的有关信息

**组 成：**学号，姓名，性别，年龄，  
所在院系，年级，专业



### (3) 数据流

➤ 数据流是数据结构在系统内传输的途径。

➤ 数据流的描述=

{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}



数据流名：学籍变动信息

含义说明：学籍变化的相关信息

数据流来源：教务人员

数据流去向：学籍变动管理

组成：{{学生}，变动内容}

数据量：较少

峰值：学期末和开学初10次/天



## (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。它可以是手工文档或手工凭单，也可以是计算机文档。

➤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}





文件名：学生基本信息表

含义说明：学生学籍基本信息

输入的数据流：新的学籍信息

输出的数据流：查询的学籍信息

组成：{学号，姓名，性别，入学年份，学院，宿舍}

存取要求：顺序存取

存取量：每年4000个新记录



## (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。数据字典只描述处理过程的说明性信息。

➤ 处理过程描述=

{数据过程名, 说明, 输入: {数据流}, 输出: {数据流}, 处理: {简要说明}}



**处理过程：** 基本信息管理

**含义说明：** 实现新生在校学籍信息的增删改

**输入数据流：** 学生卡片

**输出数据流：** 学生在校基本信息

**处理：** 新生报到后，根据录取专业，为所有新生录入宿舍，院系，专业，班级等在校信息。



# 5.3 概念结构设计

- 概念结构
- 设计方法
- 设计步骤
- 局部概念结构设计
- 全局概念结构设计



## 5.3.1 概念结构设计的定义

- 不同DBMS基于不同的数据模型，而现实应用环境复杂多变，将现实世界中的事物直接转换为机器中的对象，非常不方便。
- 概念模型是现实世界到机器世界的一个中间层次，使用接近计算机存储的方式表示数据，同时又不涉及具体的DBMS。

概念模型的描述工具通常是E-R模型，该模型不依赖于具体的硬件环境和DBMS。



## 5.3.2 概念结构设计方法

- 1、自顶向下：先定义全局，然后再逐步细化。
- 2、自底向上：先定义局部，然后再集成起来。
- 3、逐步扩张：先定义核心，然后再逐步向外扩充。
- 4、混合策略：自底向上和自顶向下相结合。



### 5.3.3 局部（分）E-R图设计

- 选择局部应用。一般而言，中层数据流图能较好反应系统各局部应用的子系统组成，因此通常以中层数据流图作为设计分E-R图的依据。
- 从数据字典中抽取数据。依据数据流图，标定局部应用中的实体、实体属性、标识实体的码，确定联系及类型。



# 1、确定局部概念结构的范围

参考原则：

- 1) 联系密切的数据
- 2) 实体(集)数要适中

## 2、确定实体（集）

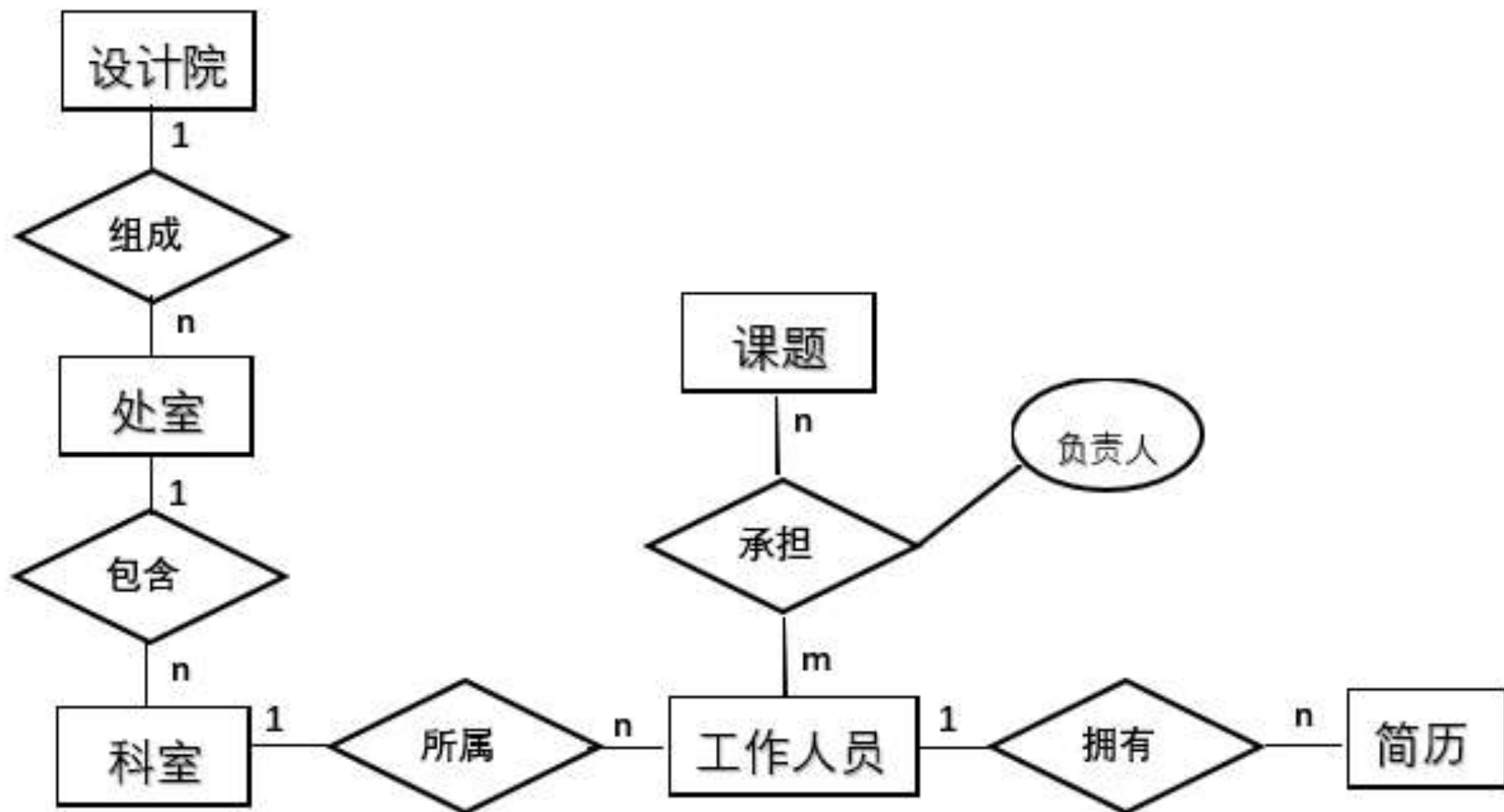
## 3、确定实体（集）的属性

## 4、定义实体（集）间的联系

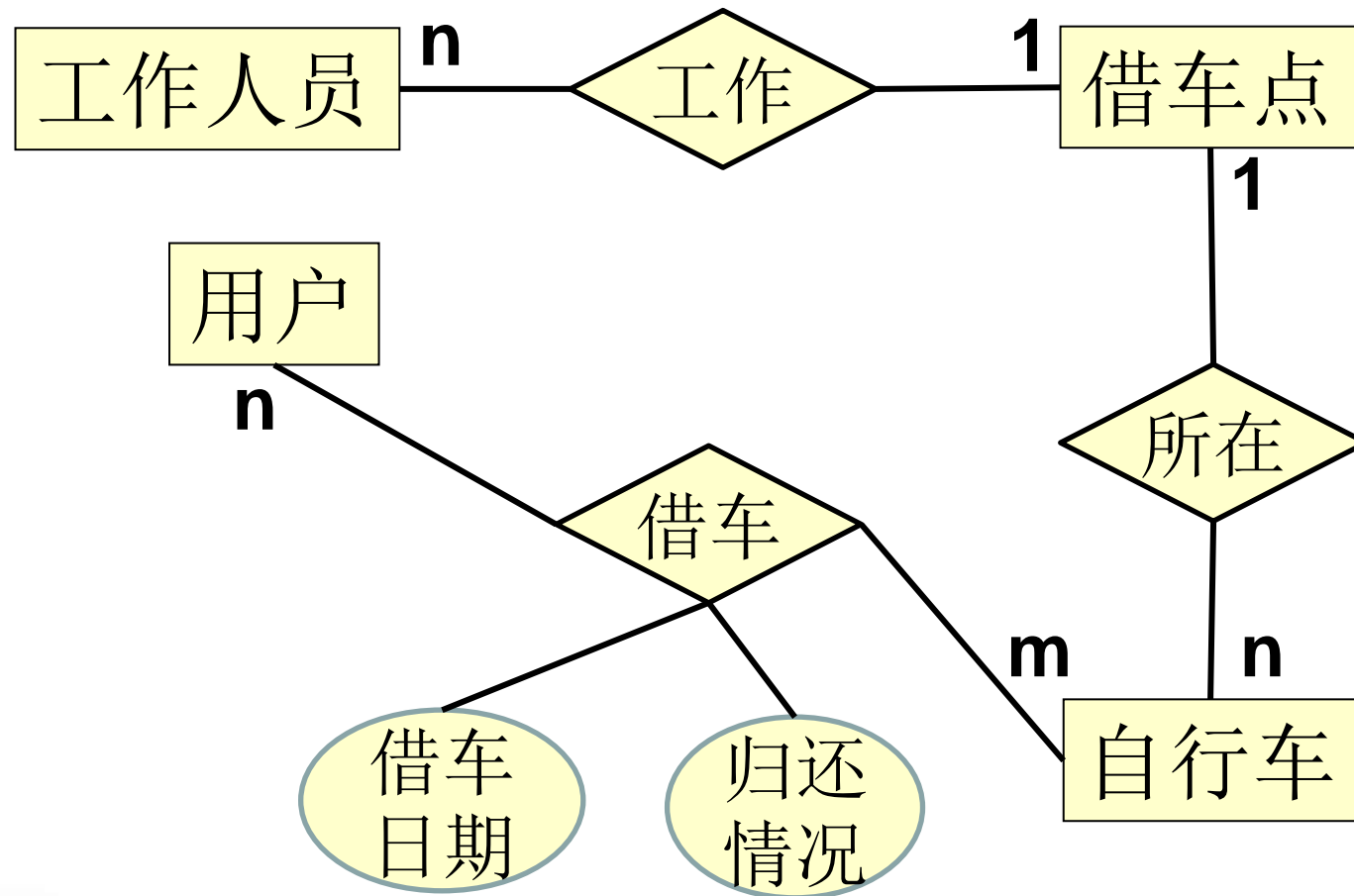




# 设计院信息管理系统



# 校园公共自行车管理系统



# 属性方面的讨论：

- 简单/组合属性

- 简单属性：由独立存在的单个部分组成的属性。
- 组合属性：由多个部分组成的属性，每个部分都可以独立存在。

- 例如：

- 姓名，地址



# 属性方面的讨论：

- 单值/多值属性

- （强实体类型与弱实体类型）

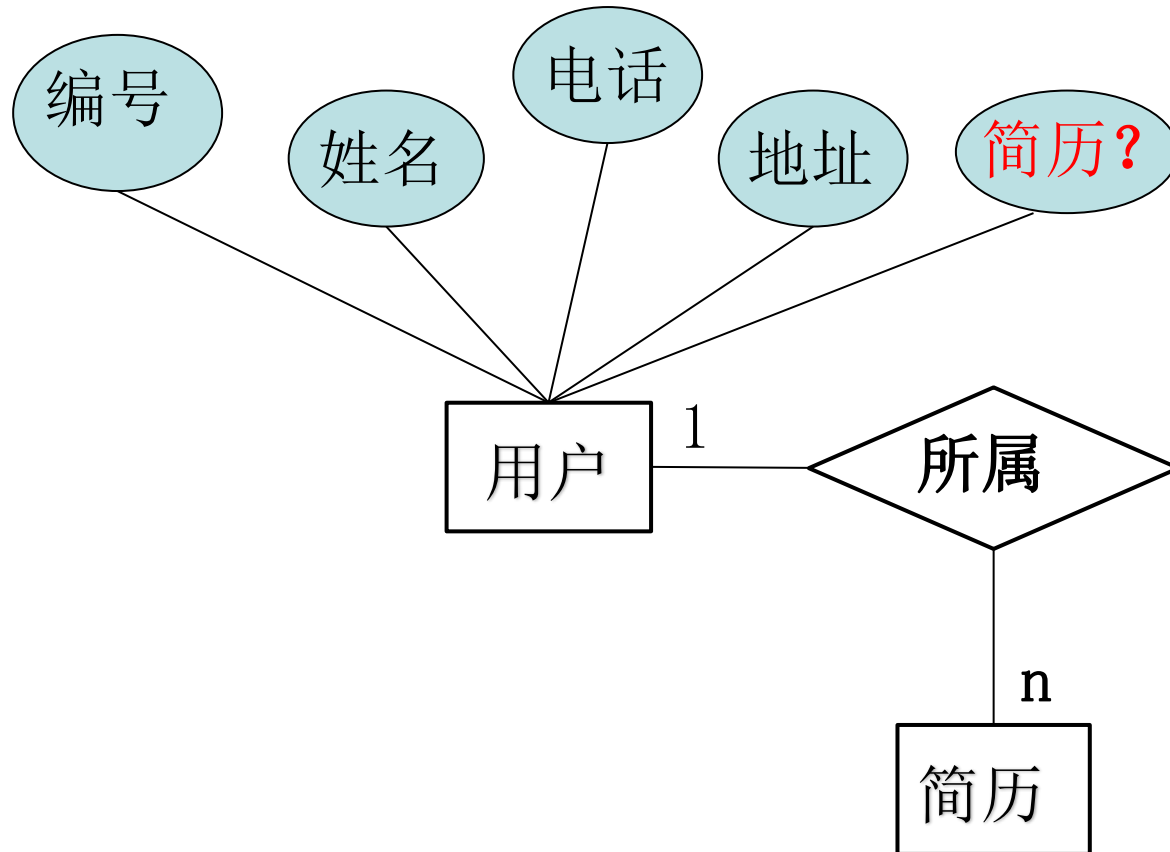
- 单值属性：对于实体类型的每个实例，都只取一个单值的属性。
    - 多值属性：对于实体类型的某些实例，出现可能取多个值的属性。

- 例如：

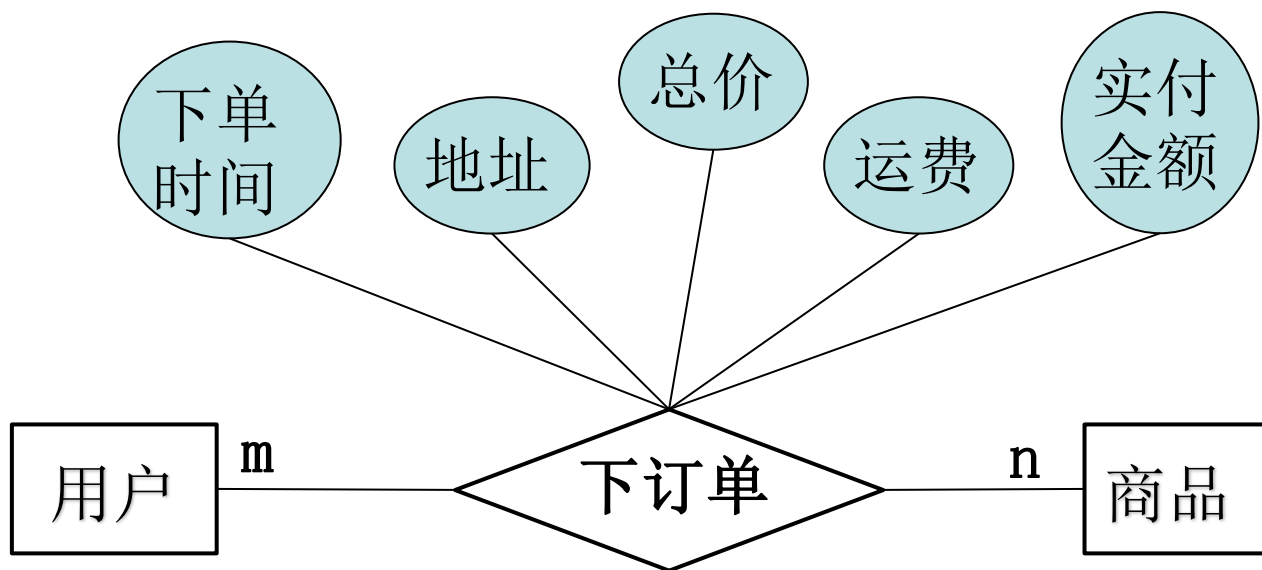
- 联系方式，简历



# 简历



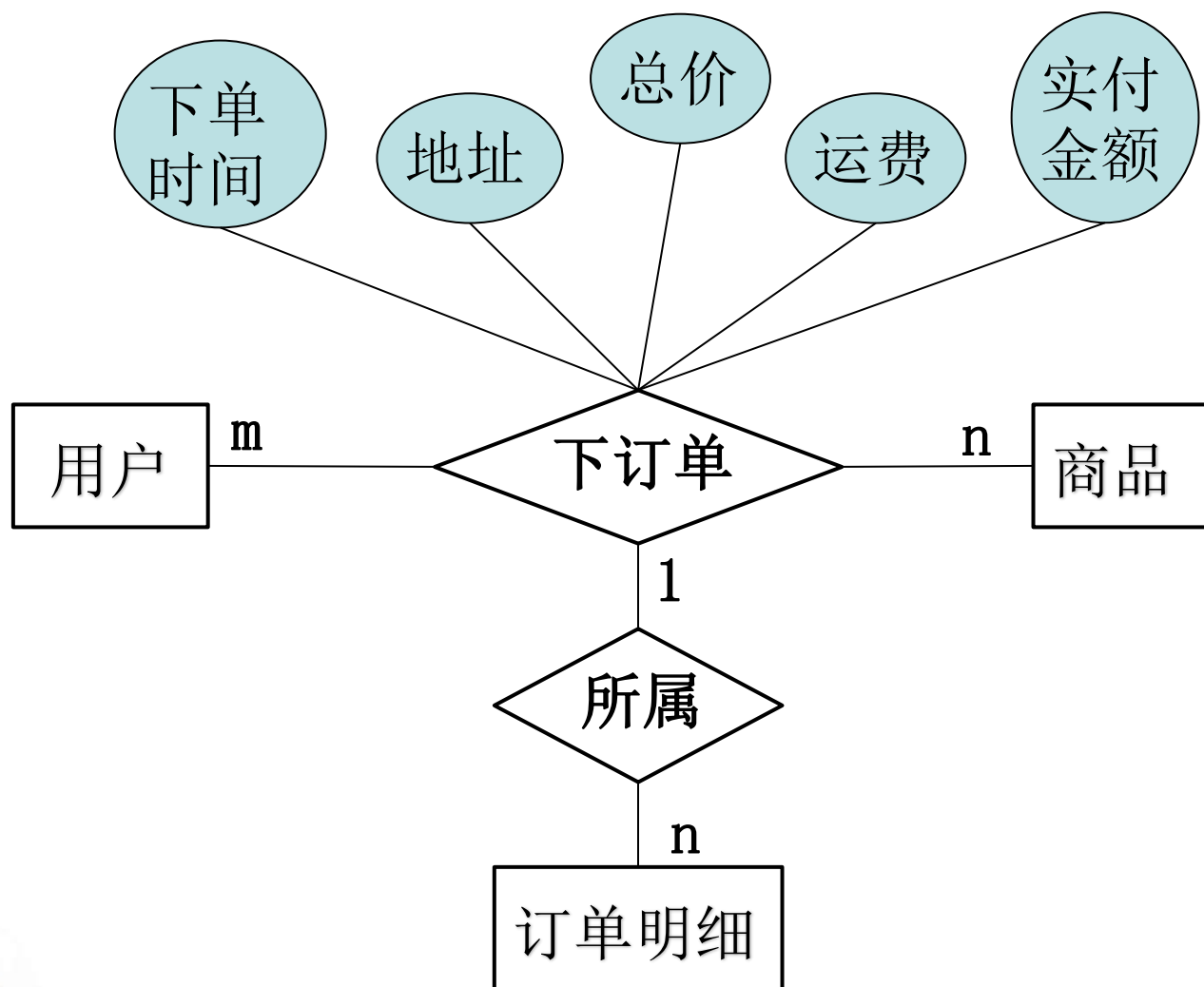
# 订单



订单编号	用户id	下单时间	地址	总价	运费	实付金额



# 订单



# 属性方面的讨论：

- 导出属性

- 属性的值是从相关的一个或者一组属性（不一定来自同一个实体类型）的值导出出来的属性。

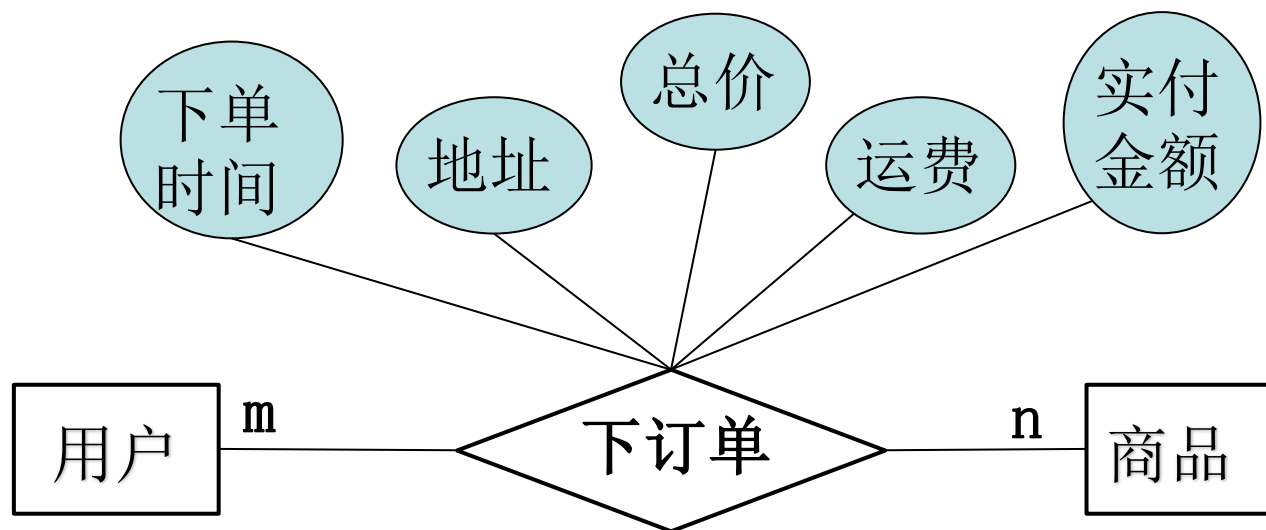
- 例如：

- 租用时间，总人数，成绩





# 订单



订单编号	用户id	下单时间	地址	总价	运费	实付金额

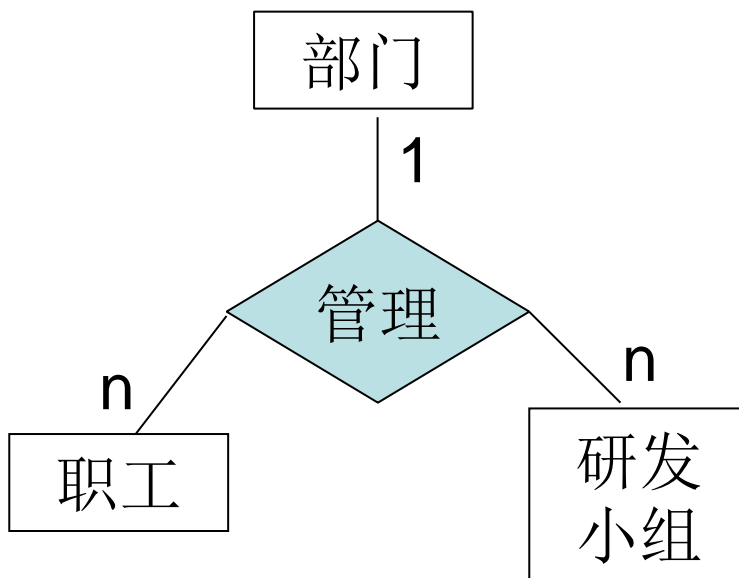


# 联系方面的讨论：

- 扇形陷阱

- 模型给出了两实体类型之间的一种联系，但在某些实体之间存在着多条通路。

例如：一个部门包含多名职工以及多个研发小组，每个职工和每个研发小组只属于一个部门。



# 联系方面的讨论：

- 断层陷阱

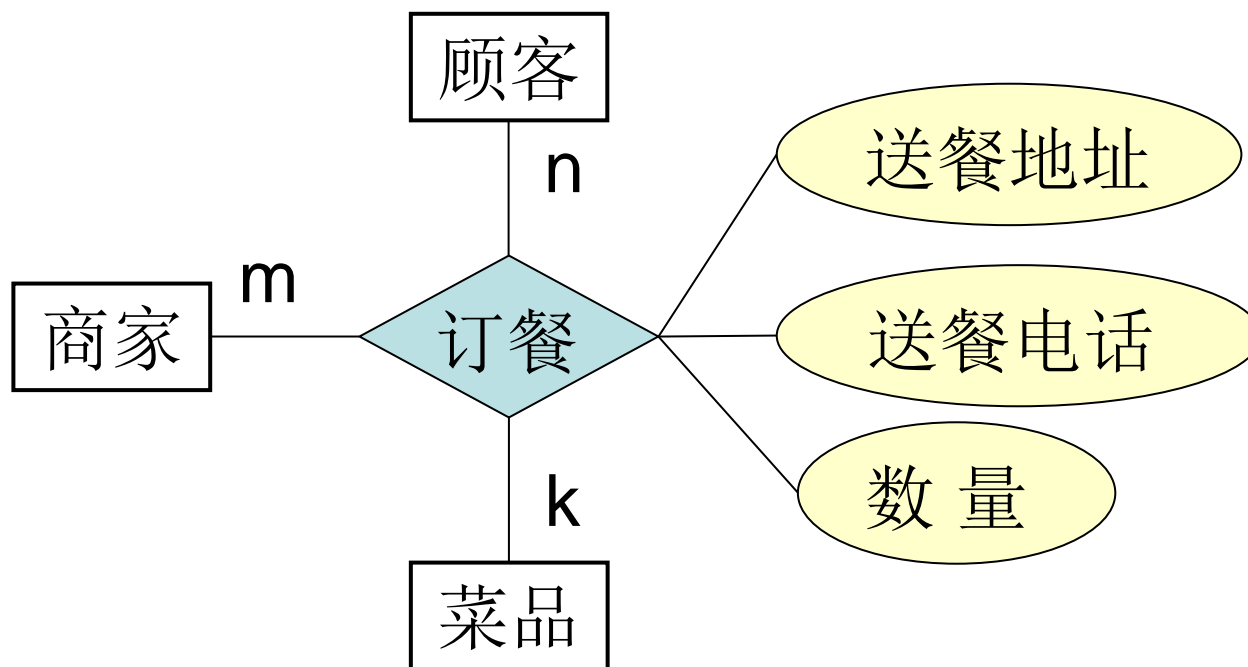
- 模型表明某些实体类型之间存在着联系，但某些实体之间却不存在通路。

例如：一个部门包含多名职工，负责多个项目，每个职工和每个项目只属于一个部门。



# 订餐系统

不同的顾客可以在不同的商家订购多种菜品。



## 5.3.4 集成全局视图

### 视图集成要解决的问题

(1) 确定模式之间的对应和冲突

- 属性冲突 (属性域冲突 取值单位冲突)
- 命名冲突 (同名异议 异名同义)
- 结构冲突

(2) 修改视图使得相互一致

(3) 合并视图

- 消除冗余



# 1、属性冲突

- 两类属性冲突
  - 属性域冲突：属性值的类型、取值范围不同。
  - 属性取值单位冲突。
- 属性冲突的解决方法
  - 通常用讨论、协商等行政手段加以解决



## 2、命名冲突

- 两类命名冲突
  - 同名异义：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）：同一意义的对象在不同的局部应用中具有不同的名字



## 2、命名冲突

### ➤ 属性层面

s(sno, sname, .....)

s(s#, sname, .....)

### ➤ 实体层面

s(sno, sname, .....)

user(sno, sname, .....)

### • 命名冲突的解决方法

- 通过讨论、协商等行政手段加以解决





# 3、结构冲突

## 三类结构冲突

- 同一实体在不同局部视图所包含的属性不完全相同。

- 同一对象在不同应用中具有不同的抽象。

在应用A中作为实体，在应用B中作为联系

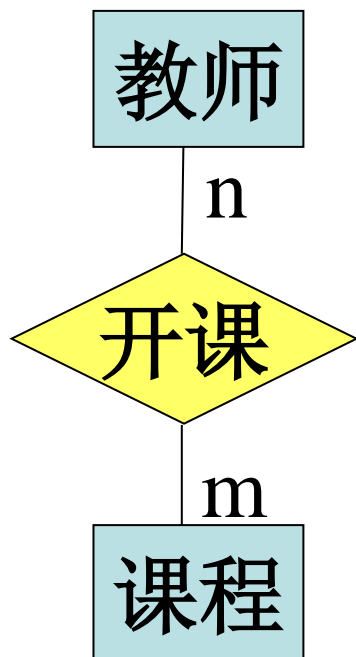
例如：开课



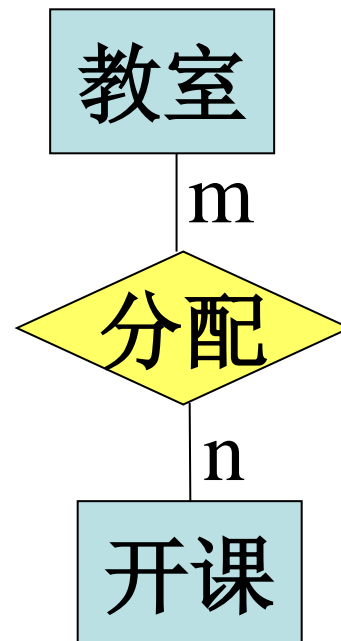
- 实体之间的联系在不同局部视图中呈现不同的类型。



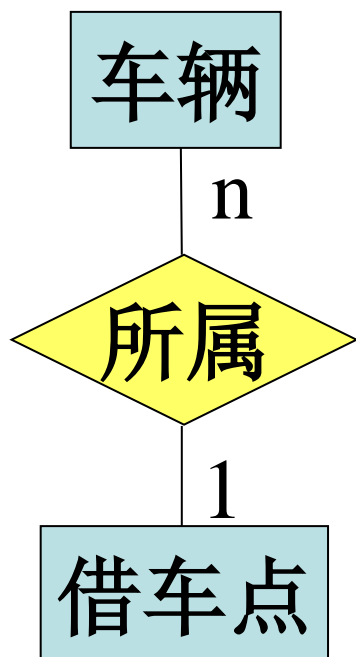
## 局部应用1:



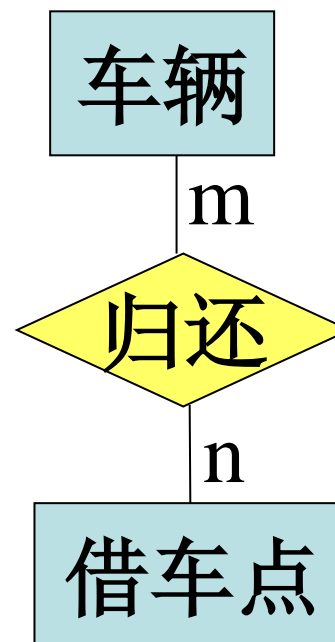
## 局部应用2:



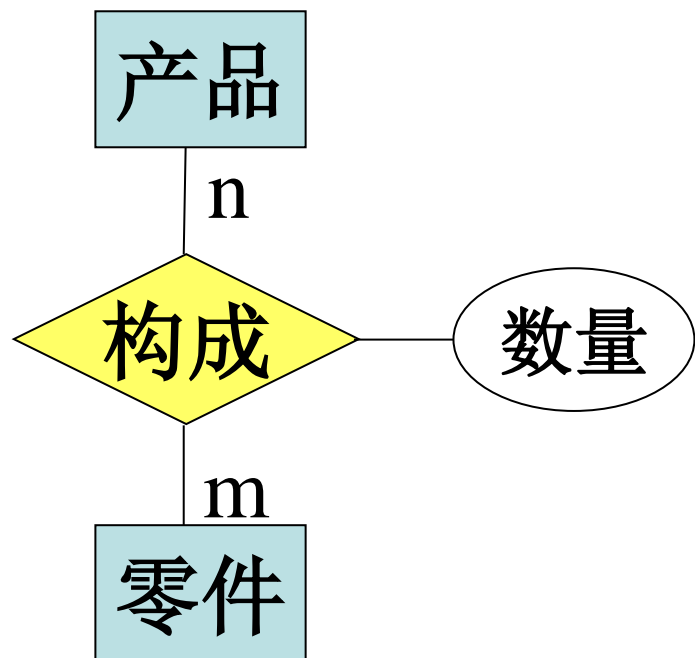
## 局部应用1:



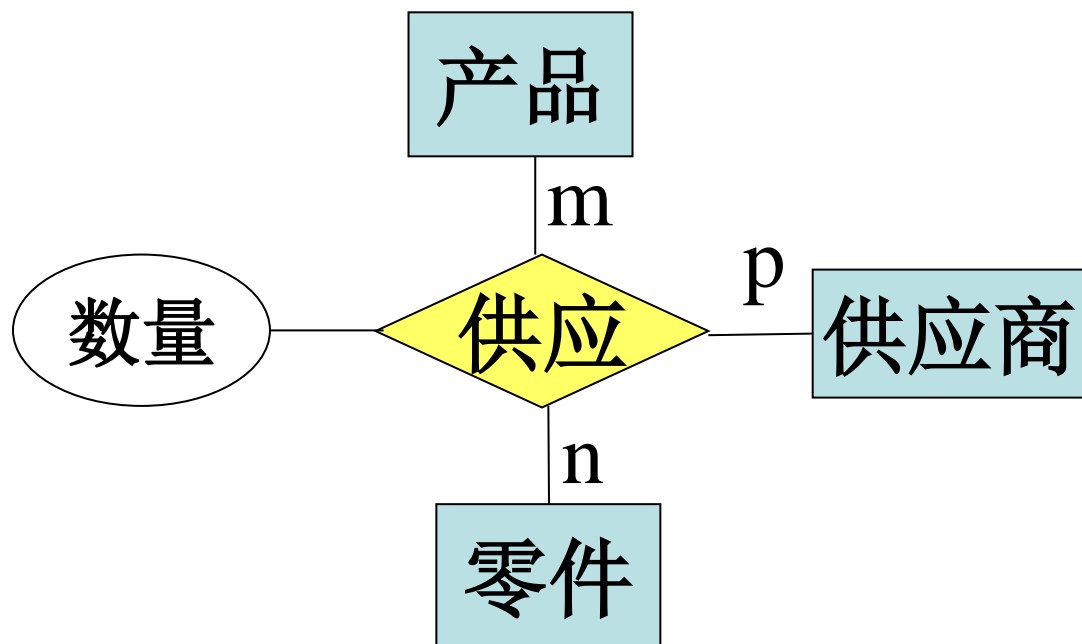
## 局部应用2:



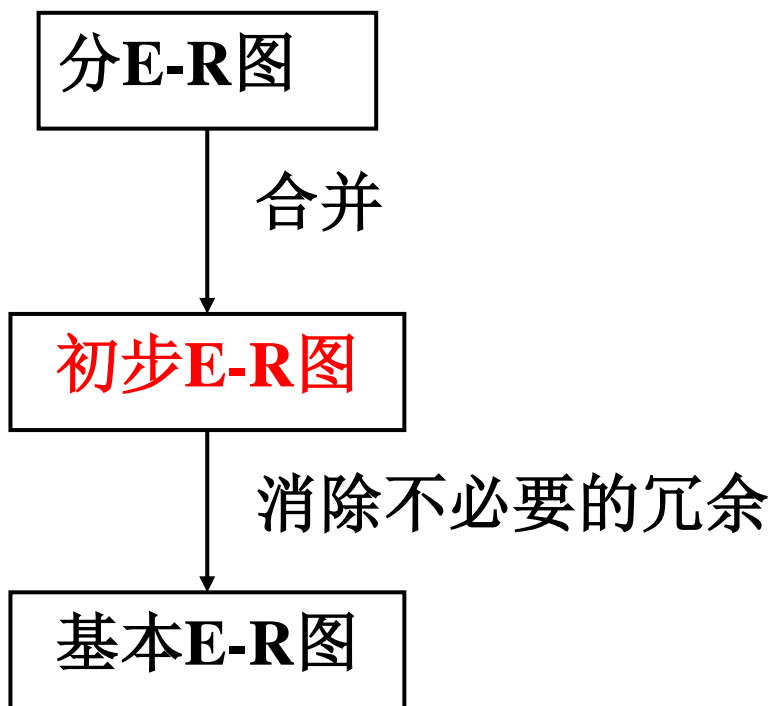
## 局部应用1:



## 局部应用2:



# 合并视图：消除冗余



可能存在冗余的数据  
和冗余的实体间联系



# 消除冗余的方法

## 1. 语义分析法

- 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。

## 2. 规范化理论法

- 函数依赖的概念提供了消除冗余联系的形式化工具
  - 求 $F$ 的最小覆盖 $F_m$ ，差集为 $D = F - F_m$ 。逐一考察 $D$ 中的函数依赖，确定是否是冗余的联系。



# 第一章作业

设计一个学生档案管理系统，学生的信息主要包括学生的学号、姓名、性别，入学年份、出生日期、联系电话，宿舍等，和学生相关的信息还有学生所在学院，所学专业，所在班级，班主任等。其中一个学院可以有多个专业，一个专业可以有多个班级，一个专业只属于某个学院，一个班级也只属于某个专业，一个班级只能有一个班主任，一个班主任也只能带一个班级，学院信息包括学院代号、学院名称、学院负责人，专业信息包括专业代号、专业名称、专业负责人，班级信息包括班级代号、班级名称，班主任信息包括工号、姓名、职称，和班主任有关信息包括所在学院和所在系部，一个教师只能在一个学院下的一个系部工作。

要求：(1) 确定有哪些实体，每个实体包括哪些属性。

(2) 找出实体间的联系，并画出E-R图。



## ➤ 实体及其属性

- 学生 {学号, 姓名, 性别, 入学年份, 出生日期, 联系电话, 宿舍}
- 学院 {学院代号, 学院名称, 学院负责人}
- 专业 {专业代号, 专业名称, 专业负责人}
- 班级 {班级代号, 班级名称}
- 班主任 {工号, 姓名, 职称}
- 教师 {职工号, 姓名, 性别, 职称}
- 系部 {系部代号, 系部名称}





# 视图集成

