

# 第1章 数据库系统概述

## 学习目标

- 掌握数据库4个基本概念
- 熟悉基本数据模型
- 掌握数据库系统的结构
- 掌握数据库系统的组成
- 了解数据库应用

## 学习内容

- 1.0 动机
- 1.1 数据库系统概述
- 1.2 数据模型
- 1.3 数据库系统的结构
- 1.4 数据库系统的组成
- 1.5 数据库系统的应用
- 1.6 本章小结

## 1.0 动机

现实场景：教师信息管理

804	李诚	男	计算机系	理科楼1121	51111
810	赵华伟	女	中文系	三院106	61234
816	陈宏林	男	电子工程系	理科楼3126	57890
825	王萍	女	计算机系	理科楼1121	51111
831	刘冰	女	电子工程系	理科楼3126	57890
856	张旭	男	电子工程系	理科楼3126	57890

存储信息

- 用文件
- 用电子表格

	A	B	C	D	E	F
1	tno	tname	tsex	dept	daddress	dphone
2	804	李诚	男	计算机系	理科楼1121	51111
3	810	赵华伟	女	中文系	三院106	61234
4	816	陈宏林	男	电子工程系	理科楼3126	57890
5	825	王萍	女	计算机系	理科楼1121	51111
6	831	刘冰	女	电子工程系	理科楼3126	57890
7	856	张旭	男	电子工程系	理科楼3126	57890
8						

管理外卖餐馆信息

	A	B	C	D	E	F	G	H
1	订单编号	客户姓名	客户地址	联系电话	所定菜品	要求送餐时间	备注	
2	DD-00001	张3先生	三号院3号楼3门303	张3的电话	青椒鸡蛋1份，米饭1份	2012/8/1	完成	
3	DD-00002	张3先生	三号院3号楼3门303	张3的电话	夫妻肺片2份，米饭2份	2013/8/2	完成	
4	DD-00003	张3先生	三号院3号楼3门303	张3的电话	青椒鸡蛋3份，米饭3份	2018/8/3		
5	DD-00004	李4先生	四号院4号楼4门404	李4的电话	饺子4份，啤酒4份	2010/8/4	完成	
6	DD-00005	李4先生	四号院4号楼4门404	李4的电话	夫妻肺片5份，饺子5份，啤酒5份	2012/8/5	完成	
7	DD-00006	李4先生	四号院4号楼4门404	李4的电话	夫妻肺片6份，饺子6份，啤酒6份	2016/8/6		
8	DD-00007	李4先生	四号院4号楼4门404	李4的电话	饺子6份，啤酒7份	2017/8/7		
9	DD-00008	王5先生	五号院5号楼5门505	王5的电话	蒜薹炒肉8份，夫妻肺片8份，米饭8份	2010/8/8	完成	
10	DD-00009	王5先生	五号院5号楼5门505	王5的电话	鱼香肉丝9份，米饭9份	2011/8/9	完成	
11	DD-00010	王5先生	五号院5号楼5门505	王5的电话	蒜薹炒肉10份，米饭10份	2016/8/10		
12	DD-00011	王5先生	五号院5号楼5门505	王5的电话	青椒鸡蛋11份，夫妻肺片1份，米饭11份	2017/2/11		
13	DD-00012	王5先生	五号院5号楼5门505	王5的电话	夫妻肺片12份，饺子12份，啤酒12份	2017/3/12		

思考：采用上述方式存储信息存在那些问题？

## 存在问题：

- 数据共享性差、冗余度大
- 数据结构化程度低：记录内有结构，整体无结构
- 数据独立性差：数据的逻辑结构改变必须修改应用程序
- 数据控制能力低：应用程序自己控制
  - 数据正确性
  - 数据的安全性
  - 并发访问
  - 备份和恢复

因此，我们需要数据库系统



## 1.1 数据库系统概述

### 一、数据存储和管理需求的增加

- 早期数据存储方式：在计算机技术刚刚兴起的20世纪50年代和60年代，数据存储和管理主要依赖于磁带和纸卡片，这种方式非常繁琐且容易出错。随着计算机应用的普及，尤其是商业领域对数据管理的需求日益增加，传统的数据存储方式已无法满足需求。
- 数据共享和集中管理的需求：随着多用户共享数据和快速检索数据的需求增加，传统的文件系统难以应对数据增长的挑战。因此，需要一种更为高效、集中的数据存储和管理方式。

## 二、计算机技术的进步

- 计算机硬件的发展：20世纪60年代，随着计算机硬件的发展，特别是磁盘存储技术的出现和成熟，数据存储的容量和速度大大提高。磁盘存储设备的出现使得随机访问数据成为可能，极大地提升了数据的存储和检索效率。
- 计算机软件的进步：随着计算机编程语言和操作系统的不断完善，软件技术也为数据库的发展提供了有力支持。软件技术的进步使得数据库管理系统（DBMS）的设计和实现成为可能。

## 三、信息处理复杂性的增加

- 数据处理需求的复杂化：随着企业和组织对数据处理需求的不断增加，简单的数据存储和检索已无法满足需求。他们需要对大量的数据进行复杂的分析和处理，以支持决策和业务运营。这种需求推动了数据库管理系统的进一步发展。
- 数据模型的发展：为了应对复杂的数据处理需求，数据模型逐渐从层次模型、网状模型发展到关系模型。关系模型以其简单、清晰、易于理解和使用的特点迅速成为主流。

## 四、数据库技术的诞生和发展

- 早期数据库系统：20世纪60年代，世界上第一个数据库系统——IDS（Integrated Data Storage，集成数据存储）诞生于通用电气公司。IDS是网状数据库，奠定了数据库发展的基础。随后，层次数据库系统如IBM的IMS（Information Management System，信息管理系统）等相继出现。
- 关系数据库的发展：1970年，IBM的研究员Edgar F. Codd提出了关系数据库模型的概念，并陆续发表了多篇论文奠定了关系数据模型的理论基础。关系数据库模型以其概念简单、数据独立性强等优点迅速得到发展。1976年，霍尼韦尔公司开发了世界上第一个商用关系数据库系统——Multics Relational Data Store。随后，Oracle、IBM DB2、Microsoft SQL Server等商业数据库产品相继出现。
- 新型数据库系统：随着计算机技术的不断发展，数据库技术也在不断演进。20世纪80年代和90年代，面向对象数据库、XML数据库、非关系型数据库等新型数据库系统相继出现，为特定类型的数据处理提供了更好的支持。

- 数据库技术产生于六十年代末，是数据管理的有效技术，是计算机科学的重要分支。
- 数据库技术是信息系统的核心和基础，它的出现极大地促进了计算机应用向各行各业的渗透。
- 数据库已经成为每个人生活中不可缺少的部分。

## ➤ 数据库的四个基本概念

- 数据 (data)
- 数据库 (Database)
- 数据库管理系 (DBMS)
- 数据库系统 (DBS)

数据 (Data) 是数据库中存储的基本对象

- 数据的定义

  - 描述事物的符号记录

- 数据的种类

  - 数字、文字、图形、图像、音频、视频、学生的档案记录等

数据的解释是指对数据含义的说明，数据的含义称为数据的语义，数据与其语义是不可分的。

例如: 93

语义1: 学生某门课的成绩

语义2: 某人的体重

语义3: 信管系某一级的人数

语义4: .....



## 学生档案中的学生记录

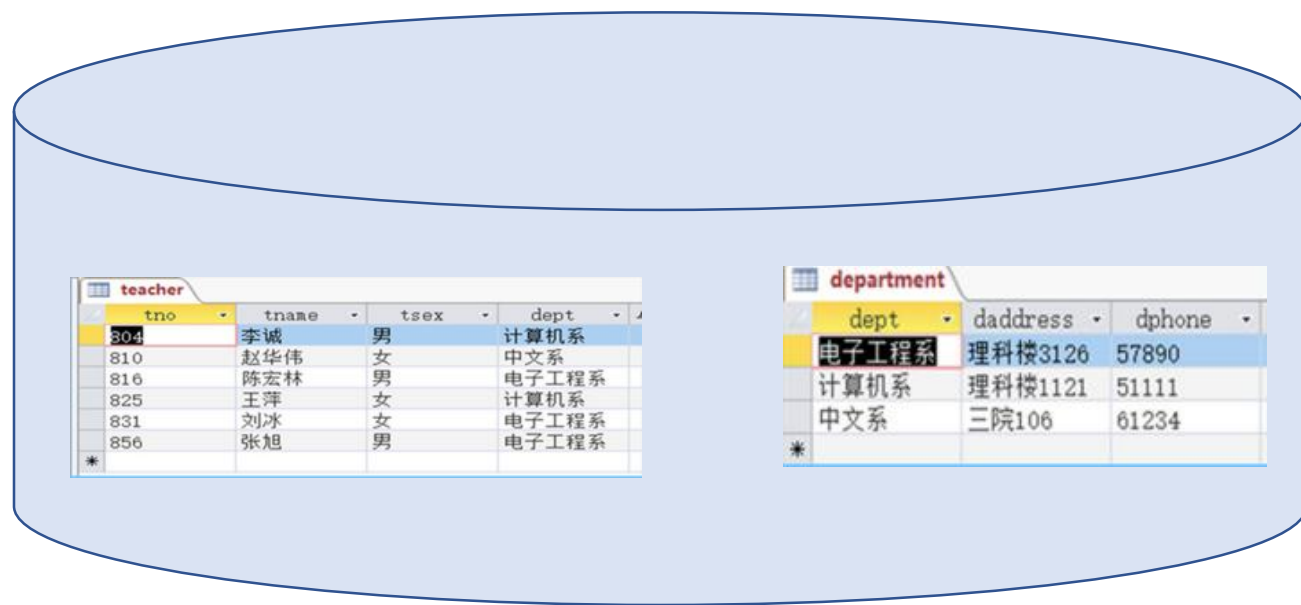
- 记录：（李明，男，199505，江苏南京市，计算机系，2013）
- 语义：学生姓名、性别、出生年月、出生地、所在院系、入学时间
- 解释：李明是个大学生，1995年5月出生，江苏南京市人， 2013年考入计算机系

## 数据库 (Database)

数据库时长期储存在计算机内、有组织的、可共享的大量数据的集合。

### 数据库的基本特征

- 数据按一定的数据模型组织、描述和储存
- 可为各种用户共享
- 冗余度较小
- 数据独立性高
- 易扩展



## 数据库管理系统(DBMS)

- 数据库管理系统是一个以统一的方式管理、维护数据库数据的一系列软件的集合。
- DBMS实例
  - Microsoft: SQL Server, Access, Visual Foxpro
  - Oracle: Oracle Database, MySQL
  - IBM: DB2

## 主要功能

### ➤ 数据定义功能

- 提供数据定义语言 (DDL)
- 定义数据库中的数据对象

### ➤ 数据组织、存储和管理

- 分类组织、存储和管理各种数据
- 确定组织数据的文件结构和存取方式
- 实现数据之间的联系
- 提供多种存取方法提高存取效率

## ➤ 数据操纵功能

- 提供数据操纵语言 (DML)
- 实现对数据库的基本操作 (查询、插入、删除和修改)

## ➤ 数据库的事务管理和运行管理

- 数据库在建立、运行和维护时由数据库管理系统统一管理和控制
- 保证数据的安全性、完整性、多用户对数据的并发使用
- 发生故障后的系统恢复

## ➤ 数据库的建立和维护功能

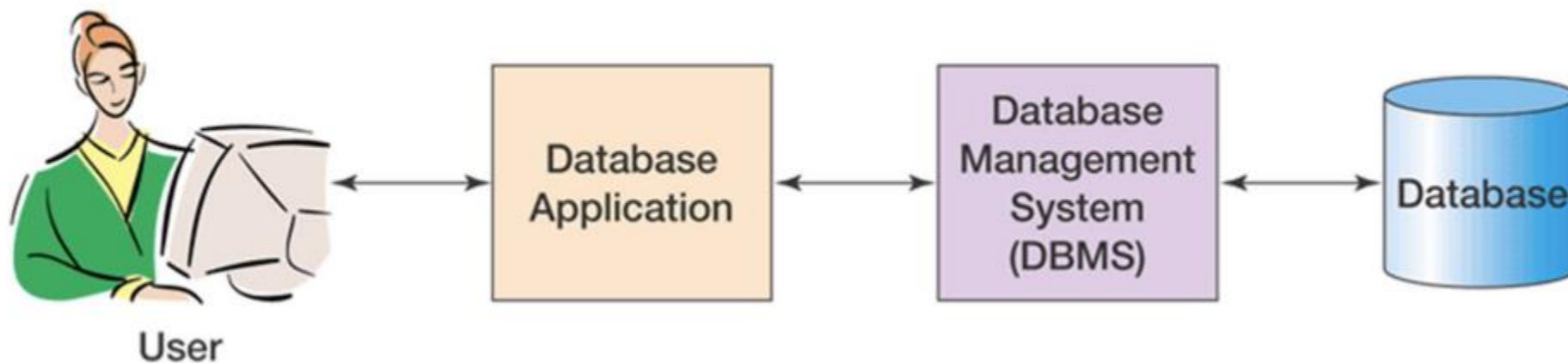
- 数据库初始数据的装载和转换
- 数据库转储、恢复功能
- 数据库的重组织
- 性能监视、分析等

## ➤ 其它功能

- 数据库管理系统与网络中其它软件系统的通信
- 数据库管理系统系统之间的数据转换
- 异构数据库之间的互访和互操作

## 数据库系统(DBS)

- 数据库系统是由数据库、数据库管理系统（及其应用开发工具）、应用程序和数据库管理员（DBA）组成的存储、管理、处理和维护数据的系统。



## 数据库系统特点

- 数据结构化
- 数据的共享性高，冗余度低，易扩充
- 数据的独立性高
- 数据由DBMS统一控制



## (1) 数据结构化

数据的整体结构化是数据库的主要特征之一  
整体结构化

- 不再仅仅针对某一个应用，而是面向全组织
- 不仅数据内部结构化，整体是结构化的，数据之间具有联系
- 数据记录可以变长
- 数据的最小存取单位是数据项

数据是用数据模型描述，无需应用程序定义

## (2) 数据的共享性高，冗余度低，易扩充

数据面向整个系统，可以被多个用户、多个应用共享使用。

数据共享的好处

- 减少数据冗余，节约存储空间
- 避免数据之间的不相容性与不一致性
- 使系统易于扩充

## (3) 数据的独立性高

### 物理独立性

- 指用户的应用程序与数据库中数据的物理存储是相互独立的。当数据的物理存储改变了，应用程序不用改变。

### 逻辑独立性

- 指用户的应用程序与数据库的逻辑结构是相互独立的。数据的逻辑结构改变了，应用程序不用改变。

数据独立性由数据库管理系统的二级映像功能来保证。

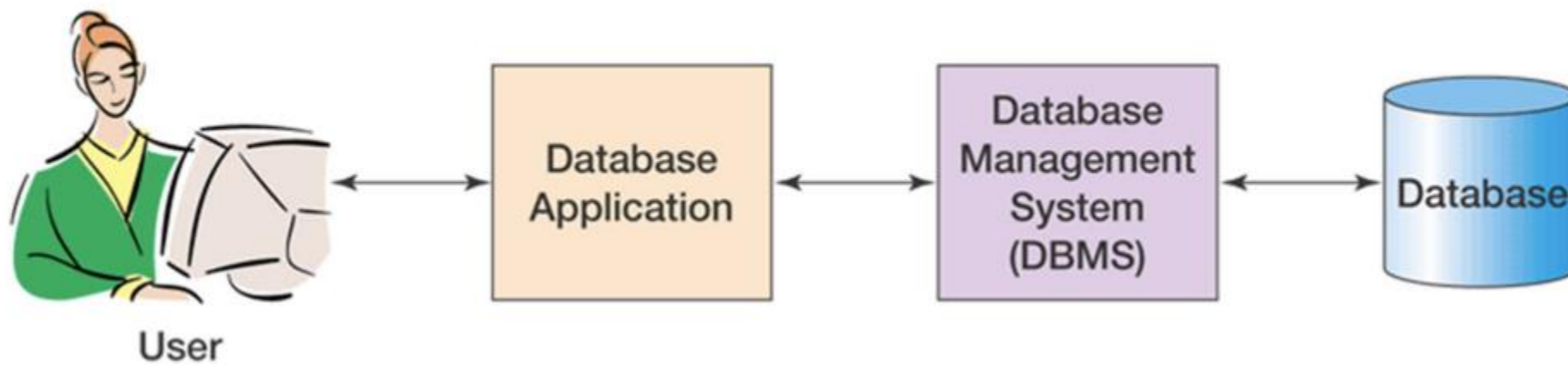
## (4) 数据由DBMS统一控制

数据库管理系统提供的数据控制功能

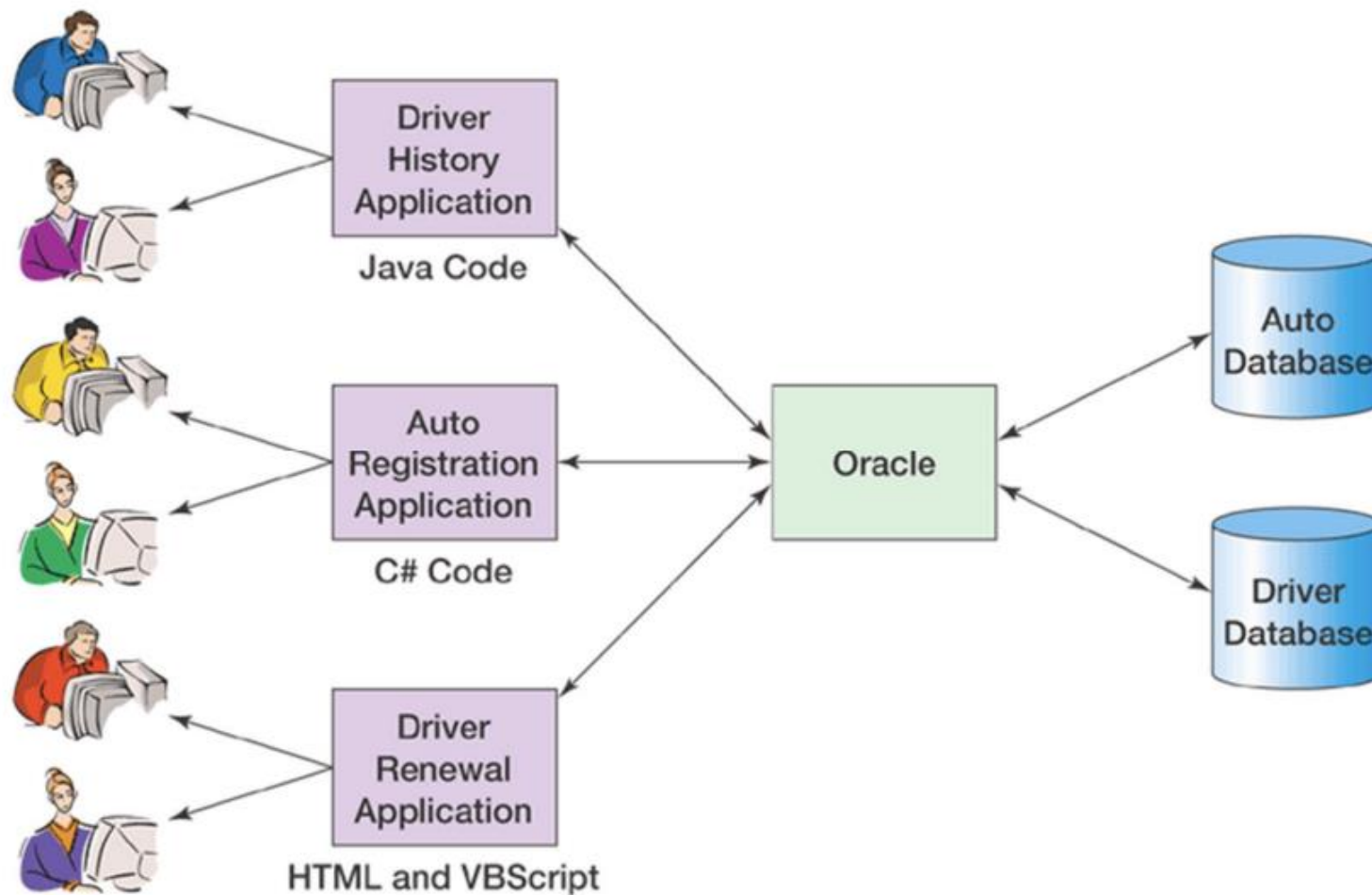
- 1) 数据的安全性 (Security) 保护  
保护数据以防止不合法的使用造成的数据的泄密和破坏。
- 2) 数据的完整性 (Integrity) 检查  
保证数据的正确性、有效性和相容性。
- 3) 并发 (Concurrency) 控制  
对多用户的并发操作加以控制和协调, 防止相互干扰而得到错误的结果。
- 4) 数据库恢复 (Recovery)  
将数据库从错误状态恢复到某一已知的正确状态。

## 数据库类型

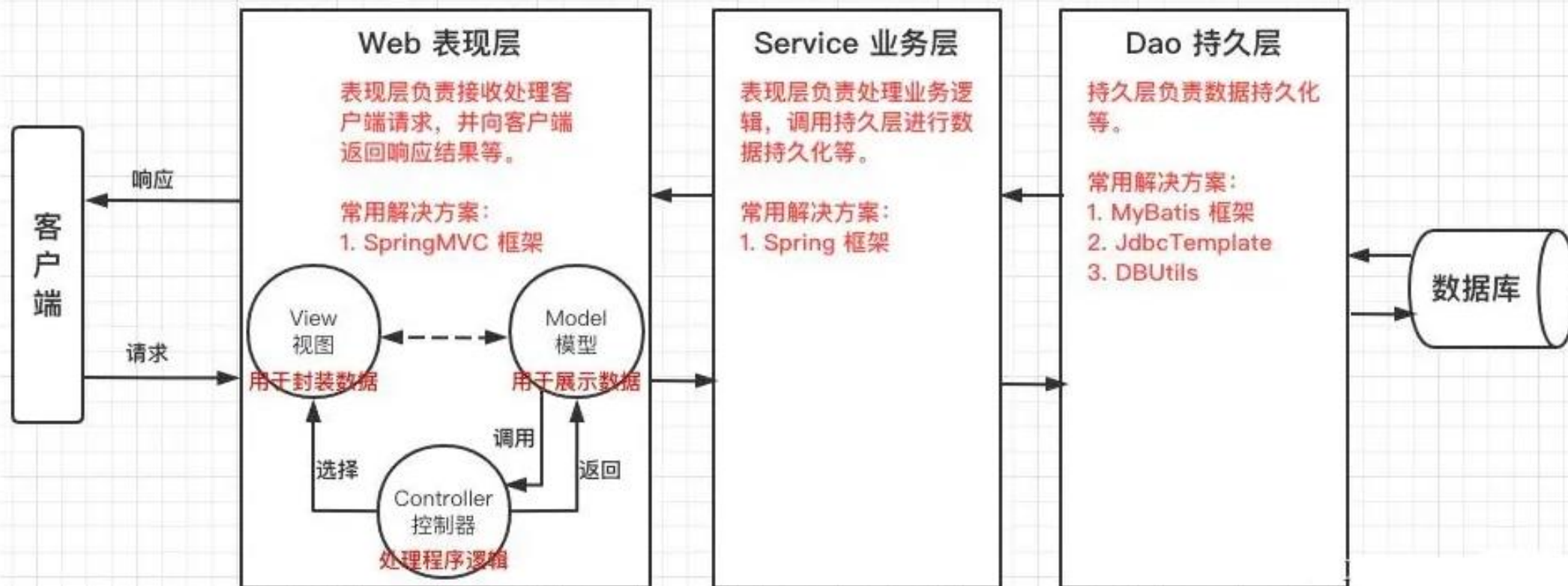
### 个人类型数据库



## 企业级数据库



## B/S 三层架构



## 1.2 数据模型

- 数据模型是对现实世界数据特征的抽象。
- 通俗地讲数据模型就是现实世界的模拟。
- 数据模型应满足三方面要求
  - 能比较真实地模拟现实世界
  - 容易为人所理解
  - 便于在计算机上实现
- 数据模型是数据库系统的核心和基础



## 两类数据模型

- 数据模型分为两类（两个不同的层次）

- (1) **概念模型**也称信息模型，它是按用户的观点来对数据和信息建模，用于数据库设计。

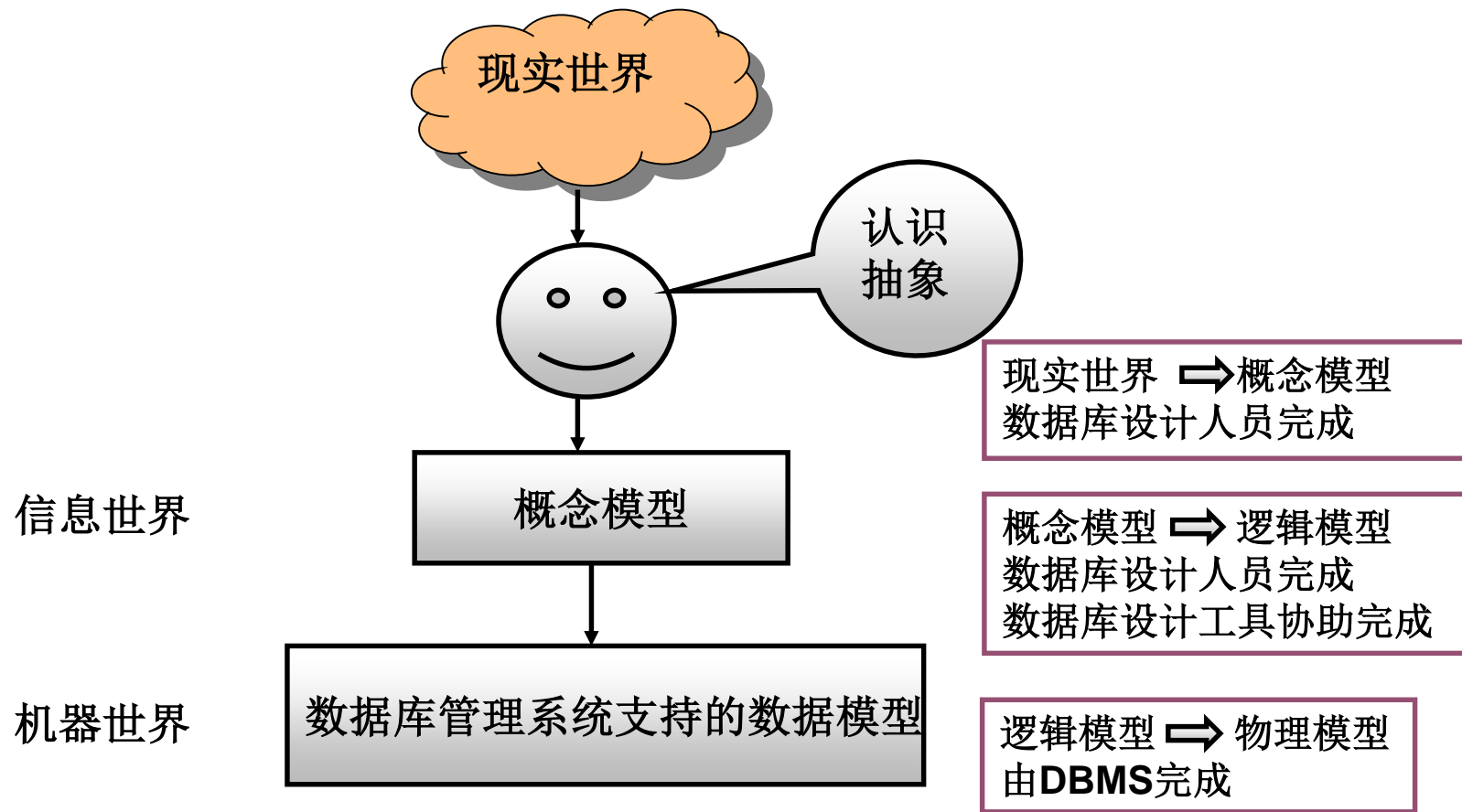
- (2) **逻辑模型和物理模型**

- 逻辑模型主要包括网状模型、层次模型、关系模型、面向对象数据模型、对象关系数据模型、半结构化数据模型等。按计算机系统的观点对数据建模，用于DBMS实现。
    - 物理模型是对数据最底层的抽象，描述数据在系统内部的表示方式和存取方法，在磁盘或磁带上的存储方式和存取方法。

## 两类数据模型

- 客观对象的抽象过程---两步抽象
  - 现实世界中的客观对象抽象为概念模型;
    - 将现实世界抽象为信息世界
  - 把概念模型转换为某一数据库管理系统支持的数据模型
    - 将信息世界转换为机器世界

## 两类数据模型



现实世界中客观对象的抽象过程

## 概念模型

1. 信息世界中的基本概念
2. 概念模型的一种表示方法：实体-联系方法

## 概念模型

### (1) 实体 (Entity)

客观存在并可相互区别的事物称为实体。

可以是具体的人、事、物或抽象的概念。

### (2) 属性 (Attribute)

实体所具有的某一特性称为属性。

一个实体可以由若干个属性来刻画。

### (3) 码 (Key)

唯一标识实体的属性集称为码。

### (4) 实体型 (Entity Type)

用实体名及其属性名集合来抽象和刻画

同类实体称为实体型

### (5) 实体集 (Entity Set)

同一类型实体的集合称为实体集

## 概念模型

### (6) 关系 (Relationship)

- 现实世界中事物内部以及事物之间的联系在信息世界中反映为实体（型）内部的关系和实体（型）之间的关系。
- 实体内部的关系通常是指组成实体的各属性之间的关系
- 实体之间的关系通常是指不同实体集之间的关系
- 实体之间的联系有一对一、一对多和多对多等多种类型

## (7) 实体-关系方法

- 实体-关系方法 (Entity-Relationship Approach)
  - 用E-R图来描述现实世界的概念模型
  - E-R方法也称为E-R模型

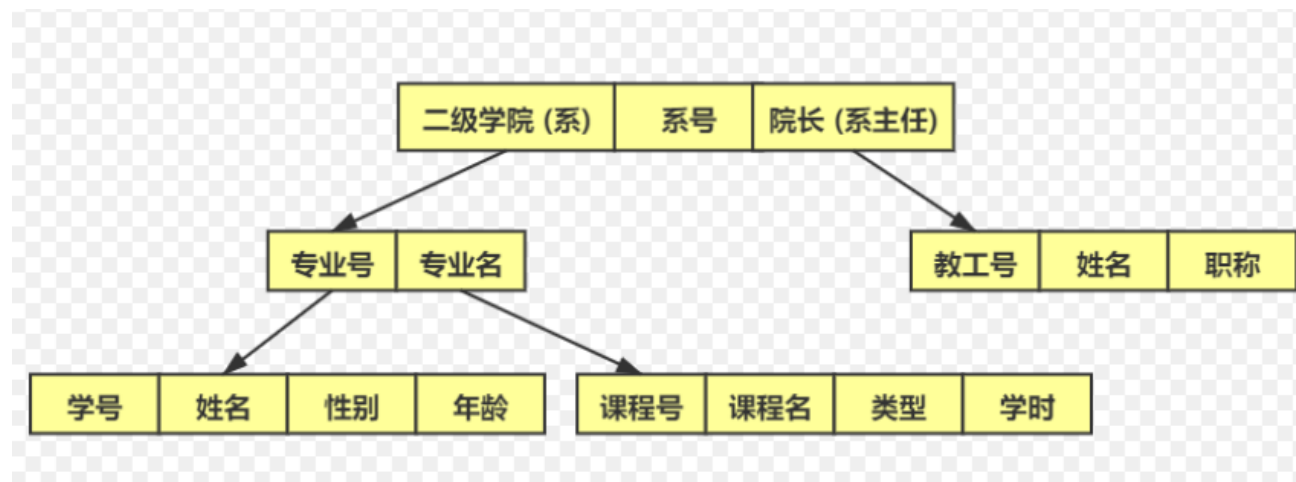
## 数据模型的组成要素

- 数据结构
- 数据操作
- 数据的完整性约束条件



## 数据结构

- 数据模型的数据结构  
描述数据库的组成对象，以及对象之间的联系
- 描述的内容
  1. 与对象的类型、内容、性质有关
  2. 与数据之间关系有关
- 数据结构是对系统静态特性的描述



## ➤ 数据操作

- 对数据库中各种对象（型）的实例（值）允许执行的  
操作的集合，包括操作及有关的操作规则

## ➤ 数据操作的类型

### ➤ 查询

### ➤ 更新（包括插入、删除、修改）

- 数据模型对操作的定义
  - 操作的确切含义
  - 操作符号
  - 操作规则（如优先级）
  - 实现操作的语言
- 数据操作是对系统动态特性的描述

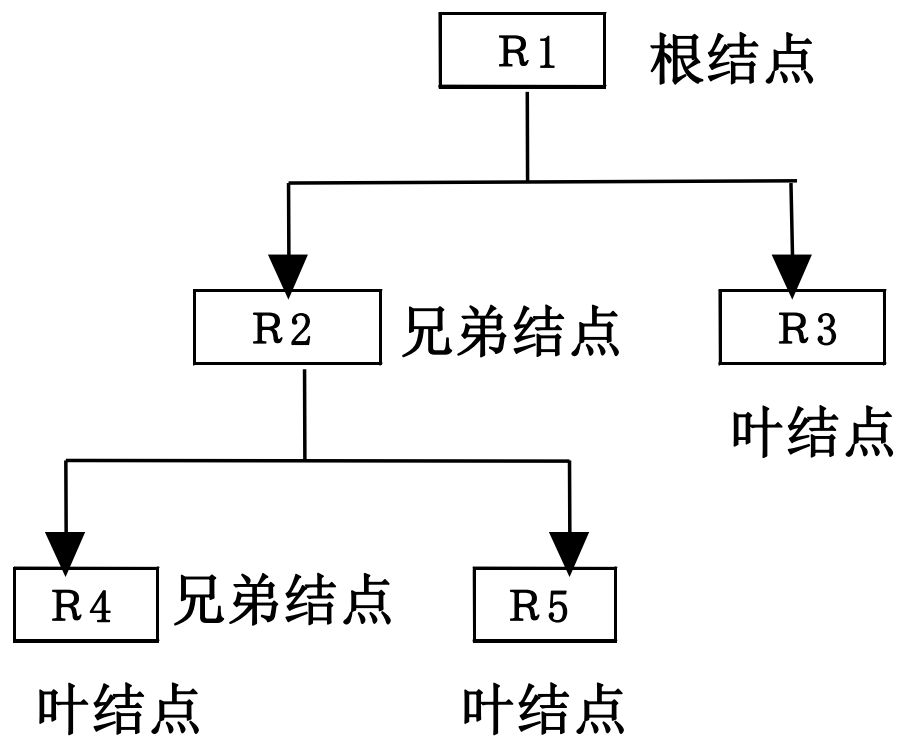
## ➤ 数据的完整性约束条件

- 一组完整性规则的集合
- 完整性规则：给定的数据模型中数据及其联系所具有的制约和依存规则
- 用以限定符合数据模型的数据库状态以及状态的变化，以保证数据的正确、有效和相容

## 数据结构的类型

- 层次模型 (Hierarchical Model)
- 网状模型 (Network Model)
- 关系模型 (Relational Model)
- 面向对象数据模型 (Object Oriented Data Model)
- 对象关系数据模型 (Object Relational Data Model)
- 半结构化数据模型 (Semi-structure Data Model)

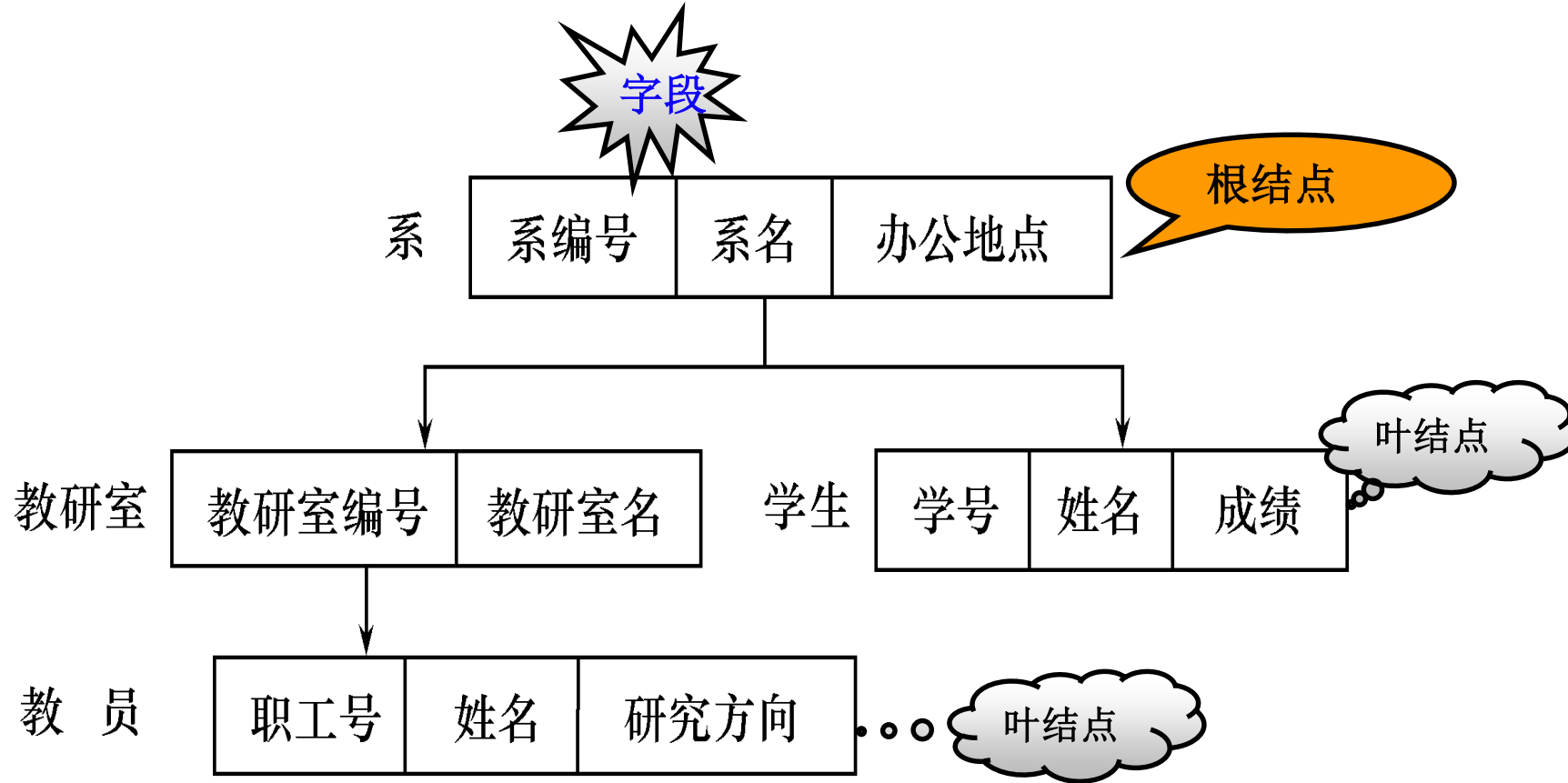
- 层次模型是数据库系统中最早出现的数据模型
- 层次数据库系统的典型代表是IBM公司的IMS (Information Management System) 数据库管理系统
- 层次模型用树形结构来表示各类实体以及实体间的联系



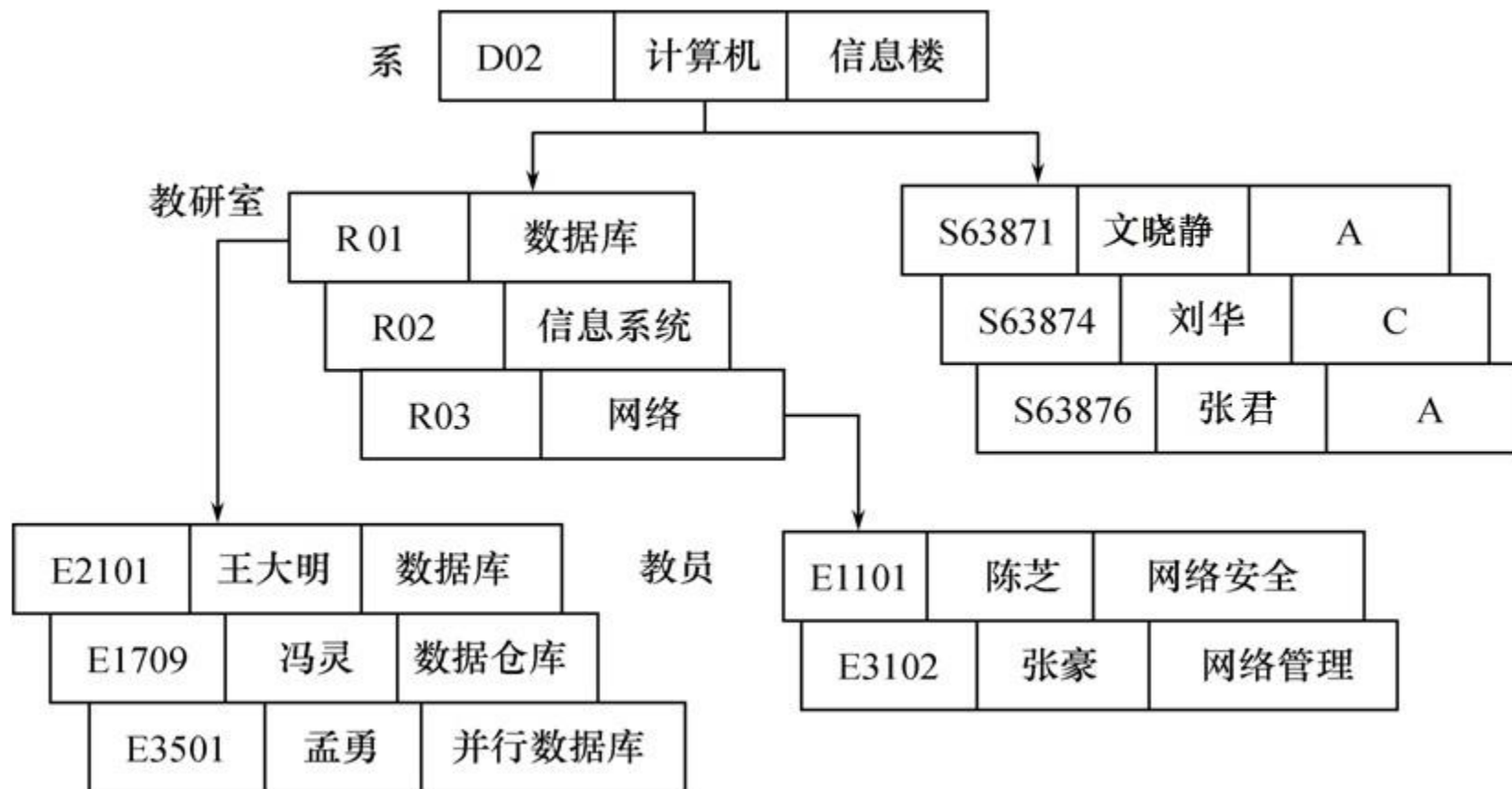
## 层次模型

满足下面两个条件的基本层次联系的集合为层次模型

1. 有且只有一个结点没有双亲结点，这个结点称为根结点
2. 根以外的其它结点有且只有一个双亲结点







- 网状模型

满足下面两个条件的基本层次联系的集合：

1. 允许一个以上的结点无双亲；
2. 一个结点可以有多于一个的双亲。

- 表示方法（与层次数据模型相同）

实体型：用记录类型描述

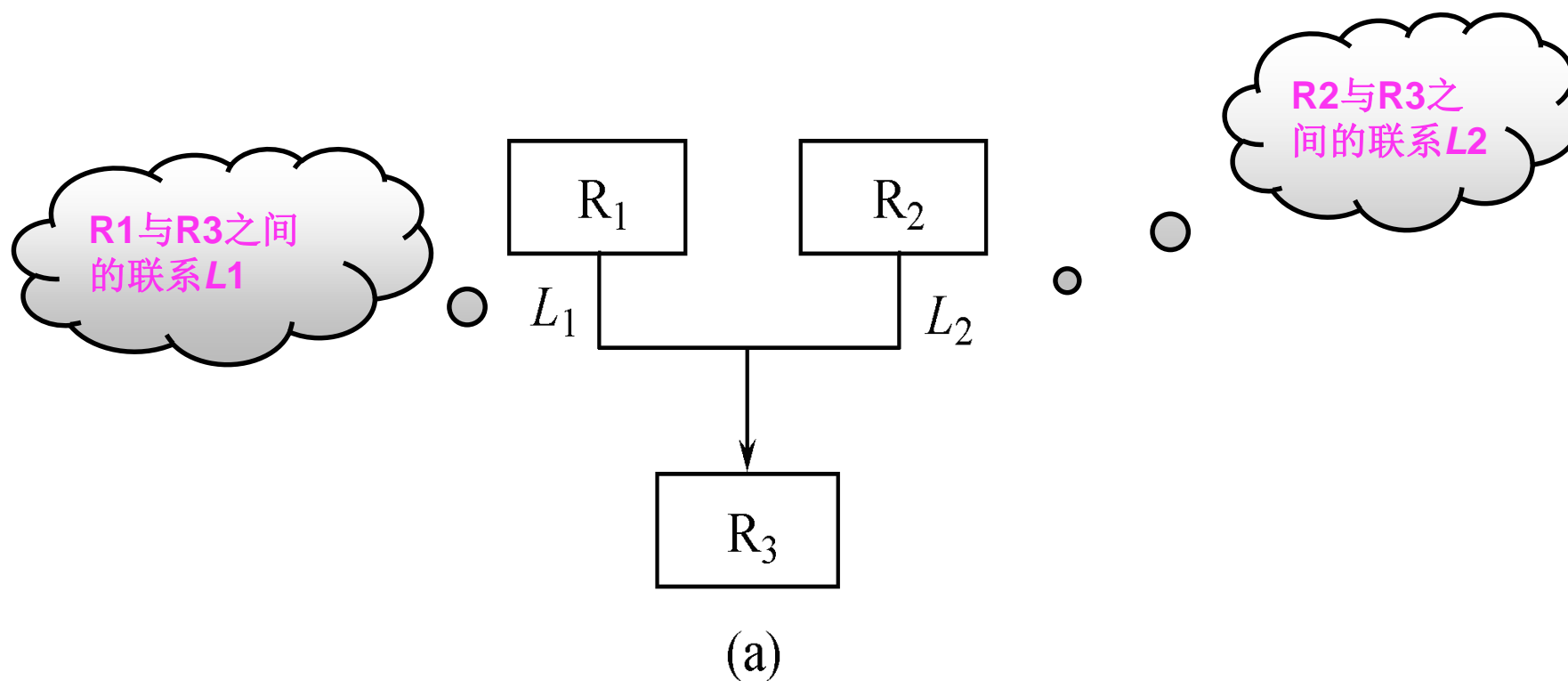
每个结点表示一个记录类型（实体）

属性：用字段描述

每个记录类型可包含若干个字段

联系：用结点之间的连线表示记录类型（实体）之间的一对多的父子联系

- 网状模型中子女结点与双亲结点的联系可以不唯一  
要为每个联系命名，并指出与该联系有关的双亲记录和子女记录

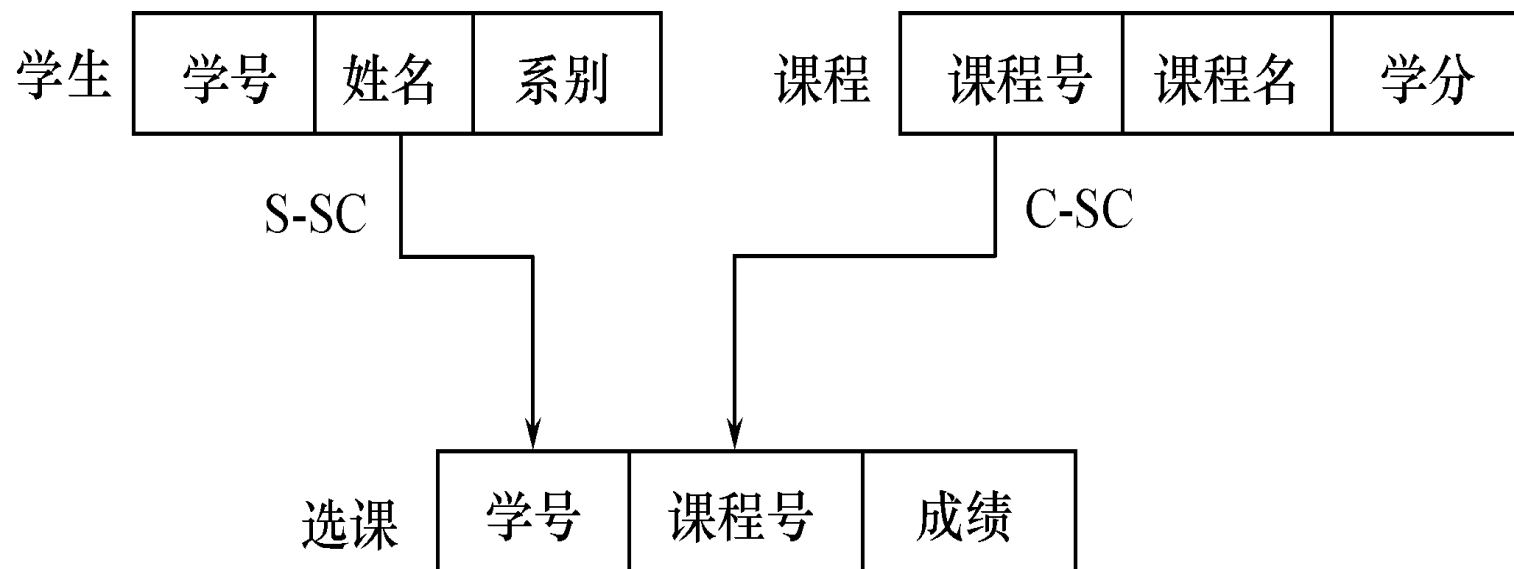


## 多对多联系在网状模型中的表示

- 用网状模型间接表示多对多联系
- 方法：将多对多联系直接分解成一对多联系

例如：一个学生可以选修若干门课程，某一课程可以被多个学生选修，学生与课程之间是多对多联系

- 引进一个学生选课的联系记录，由3个数据项组成
  - 学号
  - 课程号
  - 成绩
  - 表示某个学生选修某一门课程及其成绩



学生/选课/课程的网状数据模型

## 关系模型

- 关系数据库系统采用关系模型作为数据的组织方式
- 1970年美国IBM公司San Jose研究室的研究员E.F.Codd首次提出了数据库系统的关系模型
- 计算机厂商新推出的数据库管理系统几乎都支持关系模型

## 关系模型

学生登记表

属性

元组

学 号	姓 名	年 龄	性 别	系 名	年 级
2013004	王小明	19	女	社会学	2013
2013006	黄大鹏	20	男	商品学	2013
2013008	张文斌	18	女	法律	2013
...	...	...	...	...	...

## 关系模型

- 关系 (Relation)
  - 一个关系对应通常说的一张表
- 元组 (Tuple)
  - 表中的一行即为一个元组
- 属性 (Attribute)
  - 表中的一列即为一个属性，给每一个属性起一个名称即属性名
- 主码 (Key)
  - 也称码键。表中的某个属性组，它可以唯一确定一个元组
- 域 (Domain)
  - 是一组具有相同数据类型的值的集合。属性的取值范围来自某个域。



## 关系模型

- 分量

- 元组中的一个属性值。

- 关系模式

- 对关系的描述

关系名 (属性1, 属性2, ..., 属性n)

学生 (学号, 姓名, 年龄, 性别, 系名, 年级)

# 关系模型

- 关系必须是规范化的，满足一定的规范条件
- 最基本的规范条件：关系的每一个分量必须是一个不可分的数据项, 不允许表中还有表
- 图中工资和扣除是可分的数据项 ,不符合关系模型要求

职工号	姓名	职称	工 资			扣 除		实 发
			基本工资	岗位津贴	业绩津贴	三险	个人所得税	
86051	陈平	讲师	1305	1200	1850	160	112	4083

## 术语对比

关系术语	一般表格的术语
关系名	表名
关系模式	表头（表格的描述）
关系	（一张）二维表
元组	记录或行
属性	列
属性名	列名
属性值	列值
分量	一条记录中的一个列值
非规范关系	表中有表（大表中嵌有小表）

## 1.3 数据库系统的结构

### ➤ “型” 和“值” 的概念

型 (Type)

- 对某一类数据的结构和属性的说明

值 (Value)

- 是型的一个具体赋值

例如

学生记录:

型: (学号, 姓名, 性别, 系别, 年龄, 籍贯)

一个记录值:

值: (201315130, 李明, 男, 计算机系, 19, 江苏南京市)

## ➤ 模式 (Schema)

- 数据库系统中全体数据的逻辑结构和特征的描述。
- 仅涉及到型的描述，不涉及具体值，相对稳定

学生(姓名,性别,出生年月,出生地,系,入学年)

## ➤ 模式的一个具体的值称为模式的一个实例(Instance)

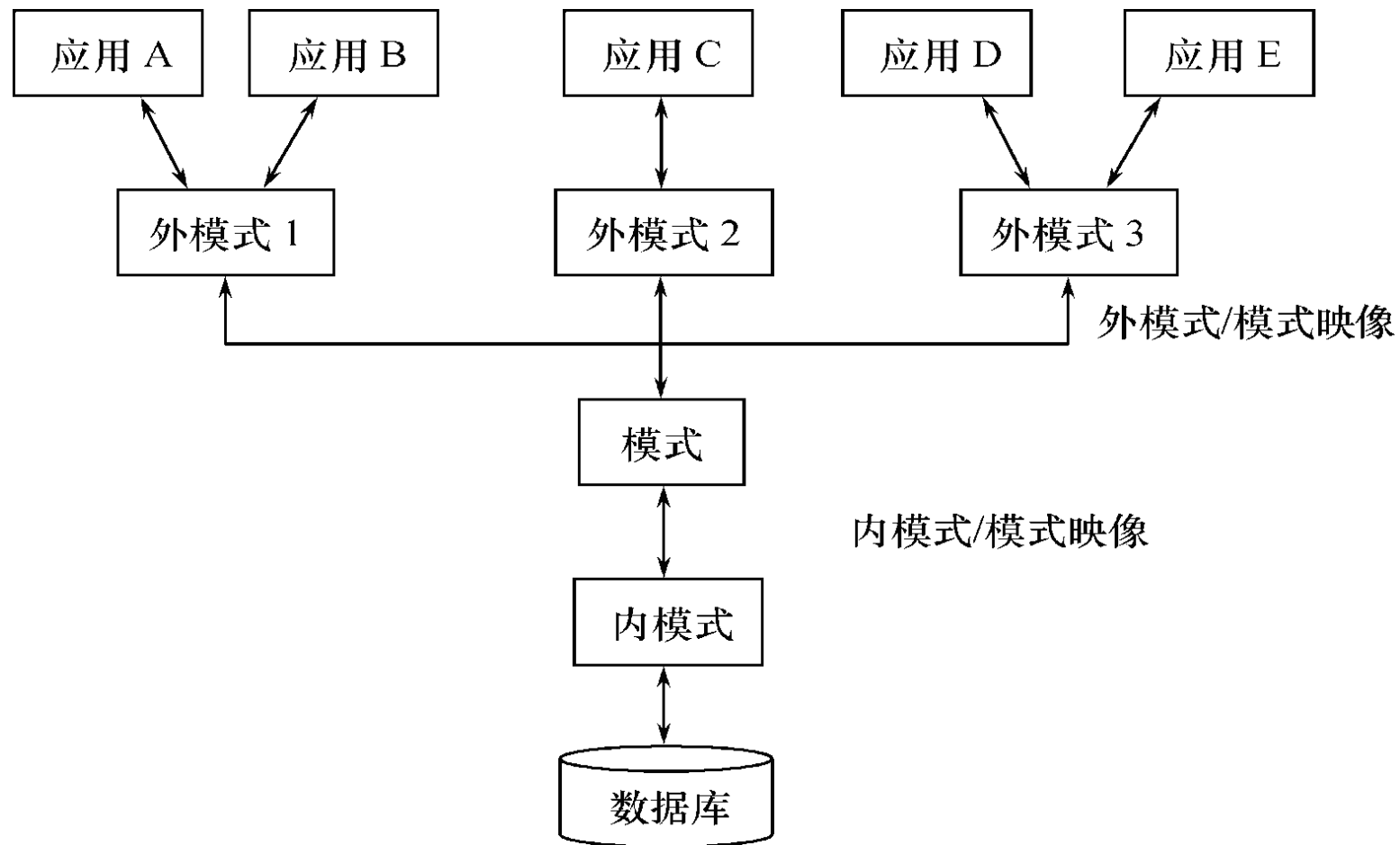
(李明, 男, 198505, 南京市, 信息管理系, 2005)

- 同一模式在不同时刻有不同的实例

为了提高数据的物理独立性和逻辑独立性，使数据库的用户观点，即用户看到的数据库，与数据库的物理方面，即实际存储的数据库区分开来，数据库系统的模式是分级的

## □ 数据库系统的三级模式结构 (Schema)

- 模式 (Schema)
- 外模式 (External Schema)
- 内模式 (Internal Schema)
- 三级模式之间有两级映象



## 1.模式(Schema)

- 是数据库中全体数据的全局逻辑结构和特性的描述，是所有用户的公共数据视图。
- 由DDL定义
- 数据库设计的关键

## 2.外模式(External-Schema)定义

- 也称为子模式 (Subschema) 或用户模式，是用户的数据视图
- 是数据的局部逻辑结构，模式的子集。一个模式可以有多个外模式。
- 由子模式DDL定义。

## 3.内模式(Internal Schema)

- 又称存储模式 (Storage Schema) ，是数据的物理结构及存储方式的描述，是数据在数据库内部的表示方式。
- 由内模式DDL定义。



## 二级映象功能

- 三级模式是对数据的三个抽象级别
- 二级映象在数据库管理系统内部实现这三个抽象层次的联系和转换
  - 外模式 / 模式映像
  - 模式 / 内模式映像

## 二级映象功能

### 外模式/模式映象

- 定义某一个外模式和模式之间的对应关系，映象定义通常包含在各外模式中
- 当模式改变时，修改外模式/模式映象，使外模式保持不变，从而应用程序可以保持不变，称为数据的逻辑独立性

外模式：teacherview ( teachername, teachersex, dept, deptaddress, deptphone )

模式：Teacher ( tno, tname, tsex, dept,daddress, dphone )

映像：

teachername  $\leftarrow$  Teacher.tname  
teachersex  $\leftarrow$  Teacher.tsex  
dept  $\leftarrow$  Teacher.dept  
deptaddress  $\leftarrow$  Teacher.daddress  
deptphone  $\leftarrow$  Teacher.dphone

修改后模式：Teachernew ( tno, tname, tsex, dept)  
Department(dept, daddress, dphone )

映像：

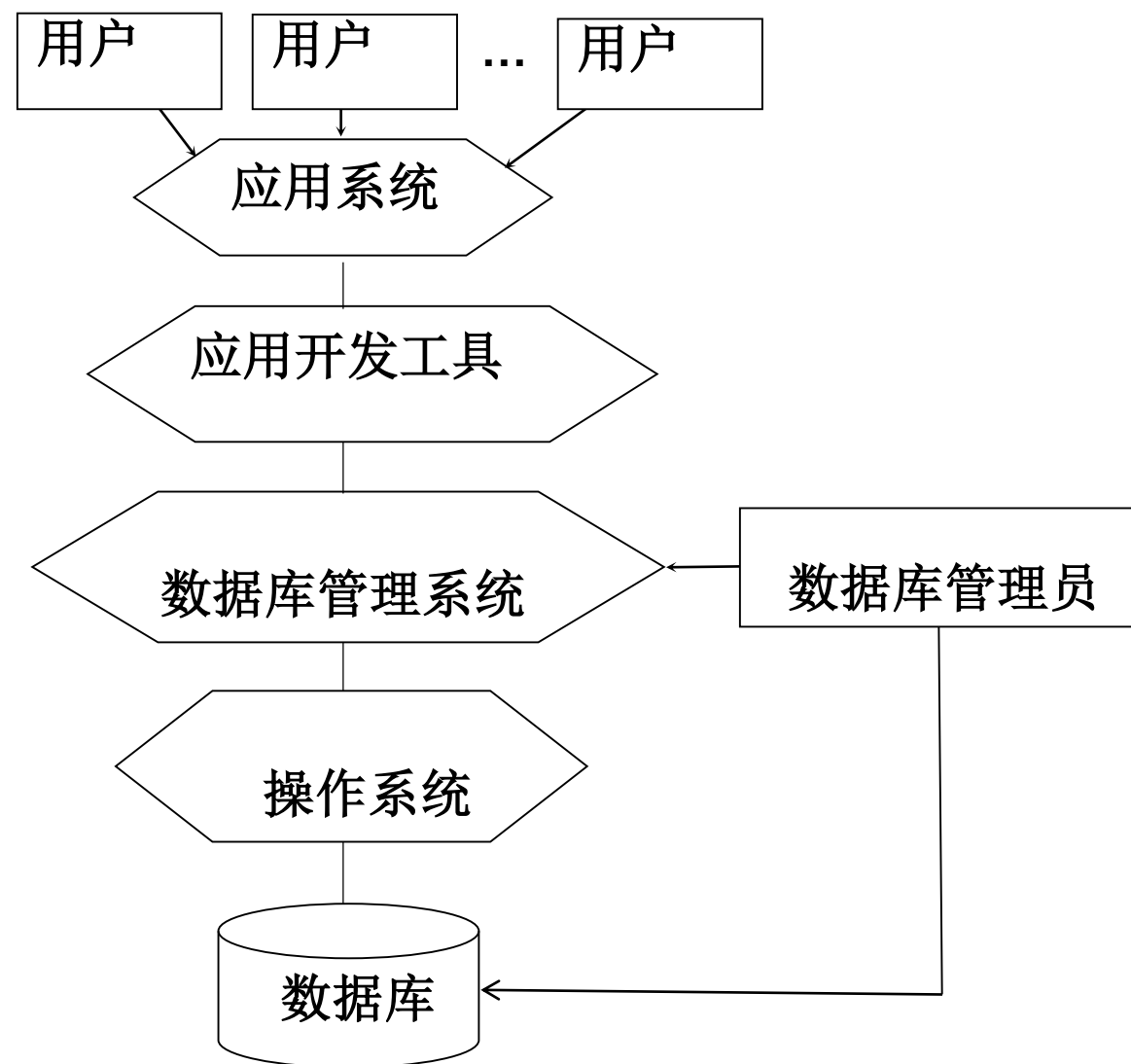
Teachername  $\leftarrow$  Teachernew.tname  
teachersex  $\leftarrow$  Teachernew.tsex  
dept  $\leftarrow$  department.dept  
deptaddress  $\leftarrow$  department.daddress  
deptphone  $\leftarrow$  department.dphone

## 模式/内模式映象

- 定义数据逻辑结构与存储结构之间的对应关系
- 存储结构改变时，修改模式/内模式映象，使模式保持不变，从而应用程序可以保持不变，称为数据的物理独立性

## 1.4 数据库系统的构成

- 数据库
- 数据库管理系统（及其开发工具）
- 应用系统
- 数据库管理员



## 硬件平台要求

- ( 1) 足够大的内存, 存放OS, DBMS核心模块, 数据缓冲区, 应用程序等
- ( 2) 大容量的磁盘等直接存取的外存设备, 以及作数据备份的磁带等。
- ( 3) 较高的通道能力, 提高数据传送率。

软件包括：

- (1) DBMS
- (2) OS
- (3) 具有与数据库的接口的高级语言编译系统
- (4) 以DBMS为核心的应用开发工具
- (5) 应用系统

## 数据库应用系统：

- 银行: 存取款...
- 超市: 进销存...
- 交通旅行: 火车飞机票旅馆预定, ...
- 医院: 病历, 检查诊断记录, ...
- 大学: 学籍, 成绩管理, ...
- 图书馆管理: 图书采购、借还书; 图书资料查找...
- 制造型企业管理: 生产计划, 原材料和产品库存, 销售订单, 人力资源管理, ...
- 电子商务: 顾客, 商品, 销售管理, ...
- 电子政务



人员:

数据库管理员(Database Administrator—DBA) 负责数据库的全面管理和控制

DBA职责

①决定数据库中的信息结构和内容

②决定数据库的存储结构、存取策略

③定义数据的安全要求和完整性约束条件

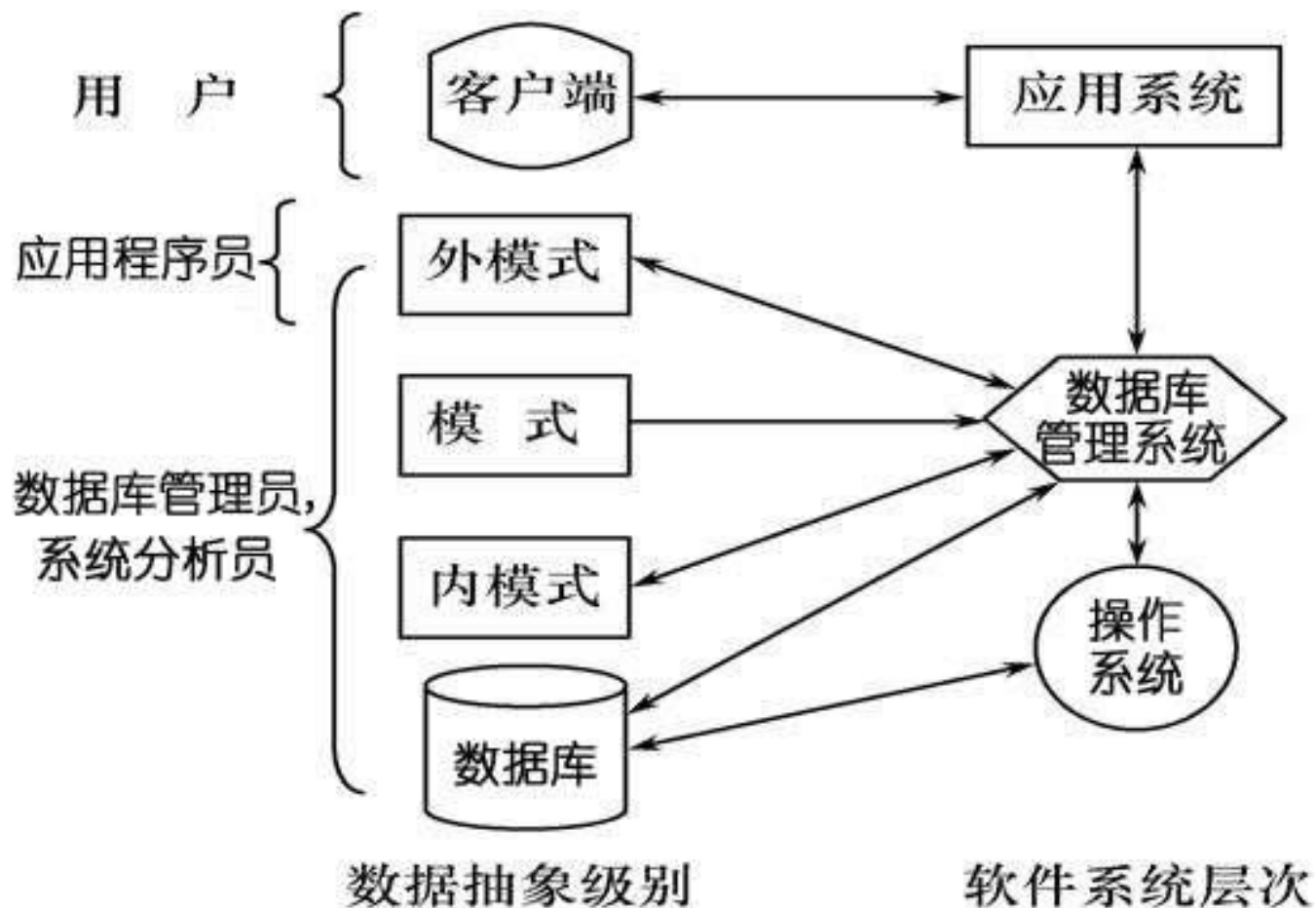
规定数据的保密级别、用户权限

④监控数据库的使用和运行情况

制定后援和恢复策

⑤ 数据库的改进和结构重组 监督分析系统的性能（空间利用率，处理效率） 数据库重组，物理上重组织，以提高性能 数据库重构造，设计上较大改动，模式和内模式修改略，负责故障恢复

人员： 不同的人员涉及不同的数据抽象级别，具有不同的数据视图，如下图所示



## 2. 系统分析员和数据库设计人员

系统分析员负责应用系统的需求分析和规范定义，确定系统的软硬件配置，参与数据库模式设计。数据库设计人员负责数据库中数据的确定、数据库各级模式的设计。可由DBA担任。

## 3. 应用程序员

负责编写应用程序

## 4. 用户

最终用户(End User)。通过应用系统的用户接口（菜单、表格等）使用数据库

## 小结

1.0 动机

1.1 数据库系统概述

1.2 数据模型

1.3 数据库系统的结构

1.4 数据库系统的组成

1.5 数据库系统的应用