

数据库原理与应用

信息系统系：李红

E-mail: Hong_lee@buaa.edu.cn

课程说明

1、期末考核

- 课程考核：闭卷考试：80%
 平时成绩：20%【出勤+作业】
- 作业部分：请安装SQL SERVER或其他数据库，按要求完成

2、教材与参考书

- 王珊，萨师焯. 数据库系统概论，高等教育出版
- 关于SQL Server或其他大型数据库的工具书

3、后续课程

- 《大型数据库应用实验》 信管专业必修课

课程知识体系

关系数据库
理论与方法

第 1、2、6、9、
10、11、12 章

SQL语言
标准及扩展

第 3、4、5、8 章

数据库
设计方法

第7章

数据库
技术新进展

第 13、14、
15、16 章

教学要求

- 理论、实践、应用并重
 - ✓ 理论：掌握关系数据模型和关系数据库基本原理
 - ✓ 实践：掌握SQL语言，学会实际的数据库操作
 - ✓ 应用：针对简单的需求，完成数据库设计、创建、操纵、管理

第 1 章 绪论

1.1 数据库系统概述

1.2 数据模型

1.3 数据库系统结构

1.4 数据库系统的组成

1.5 数据库技术研究领域

1.1 数据库系统概述

一、数据库的4个基本概念

数据

数据库

数据库管理系统

数据库系统

1. 数据(data): 描述事物的符号记录, 数据形式具有多样性

- 结构化数据: (李明, 男, 江苏省南京市, 计算机系, 2002)
- 非结构化数据: 文本、图数据 (graph)、图像、音频、视频

2. 数据库(DataBase, DB)

- 长期存储在计算机内、有组织、可共享的数据集合
- 按一定的数据模型组织、描述、存储
- 具有较小的冗余度、较高的数据独立性、易扩展性、为各种用户共享

1.1 数据库系统概述

一、数据库的4个基本概念

3. 数据库管理系统 (DBMS)：用户和操作系统间的数据管理软件

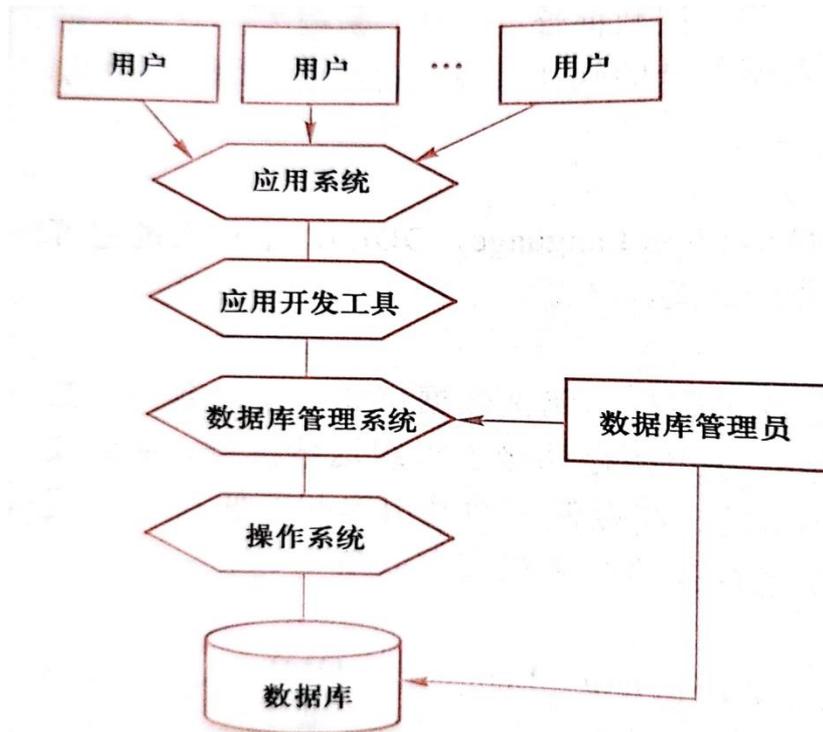
- **数据定义**：定义数据库中各个数据对象的组成与结构
- **数据组织、存储、管理**：确定文件结构、存取路径，提高存储空间利用率和存取效率
- **数据操纵**：支持用户对数据的增、删、改、查等操作
- **数据库的事务管理和运行管理**：保障事务运行的安全性、完整性、并发控制、故障恢复等
- **数据库建立和维护**：数据导入与导出，转储与恢复、性能监控、重组等
- **其他功能**：与其他系统的通信功能、异构系统间的互访和互操作等

1.1 数据库系统概述

一、数据库的4个基本概念

4. 数据库系统(DataBase System, DBS)

- 由数据库、数据库管理系统及其应用开发工具、应用程序、DBA及相关人员组成的存储、管理、处理和维维护数据的系统



1.1 数据库系统概述

二、数据管理技术的发展

数据管理：数据的分类、组织、存储、检索、维护

1. 人工管理阶段

- **技术背景**：只有纸带/卡片/磁带，无磁盘等直接存取设备，无操作系统
- **管理方式**：数据不保存，不共享，程序管理数据（程序和数据不独立）

2. 文件管理阶段

- **技术背景**：磁盘和OS成熟，由文件系统管理数据
- **管理方式**：数据长期保存，部分共享，冗余大，程序和数据部分独立

例子：学生信息相关文件（文件之间孤立，数据和程序的关联性强）

例子：学生相关文件（文件之间孤立，数据和程序的关联性强、共享性差）

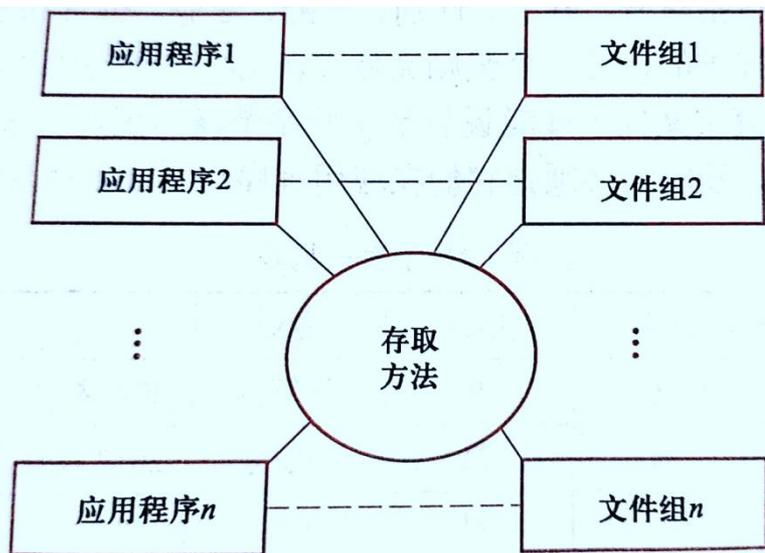


图 1.4 文件系统阶段应用程序与数据之间的对应关系

学号	姓名	性别	...	社团情况
S001	宋杰	男	...	IS
S002	武潭	男	...	IS
S003	李丽	女	...	IS
S004	凯林	男	...	MA

学号	姓名	专业	课程	成绩	选课时间
S001	宋杰	IS	8002	90	2022/03/01
S002	武潭	IS	8001	80	2022/03/01
S002	武潭	IS	8002	70	2022/03/01
S003	李丽	IS	NULL	NULL	2022/03/01
S101	凯林	MA	NULL	NULL	2022/03/01

二、数据管理技术的发展

3. 数据库管理阶段

■ 数据结构化

➢ 记录内部、记录之间、文件之间有结构

- ✓ 记录内部关联：如学时和学分之间的依赖关系，
- ✓ 记录之间关联：课程之间的先导关系
- ✓ 数据表之间：选课表和学生表、课程表直接的关系

➢ 存取粒度小：数据项

■ 低冗余、高共享、易扩充

■ 数据的独立性高：应用程序与数据逻辑结构、存储结构间的独立性

■ DBMS统一管理和控制：安全性、完整性、并发控制，故障恢复

学号	姓名	性别	专业
S001	宋杰	男	IS
S002	武潭	男	IS
S003	李丽	女	IS
S004	凯林	男	MA

学号	课号	成绩
S001	8002	90
S002	8001	80
S002	8002	70
S002	9001	75
S003	8001	76
S003	8002	87
S004	8001	79
S004	8002	69
S004	9001	86

课号	课名	先行课	学时	学分
8001	A	8002	40	2.5
8002	B	9001	32	2
9001	C	null	32	2

1.2 数据模型

一、数据模型的基本概念

1、什么是数据模型

- **模型**：事物特征的抽象
- **数据模型**：现实世界数据特征的抽象，它是数据库及应用系统抽象、表示、处理现实世界中的数据和信息的一种基本工具

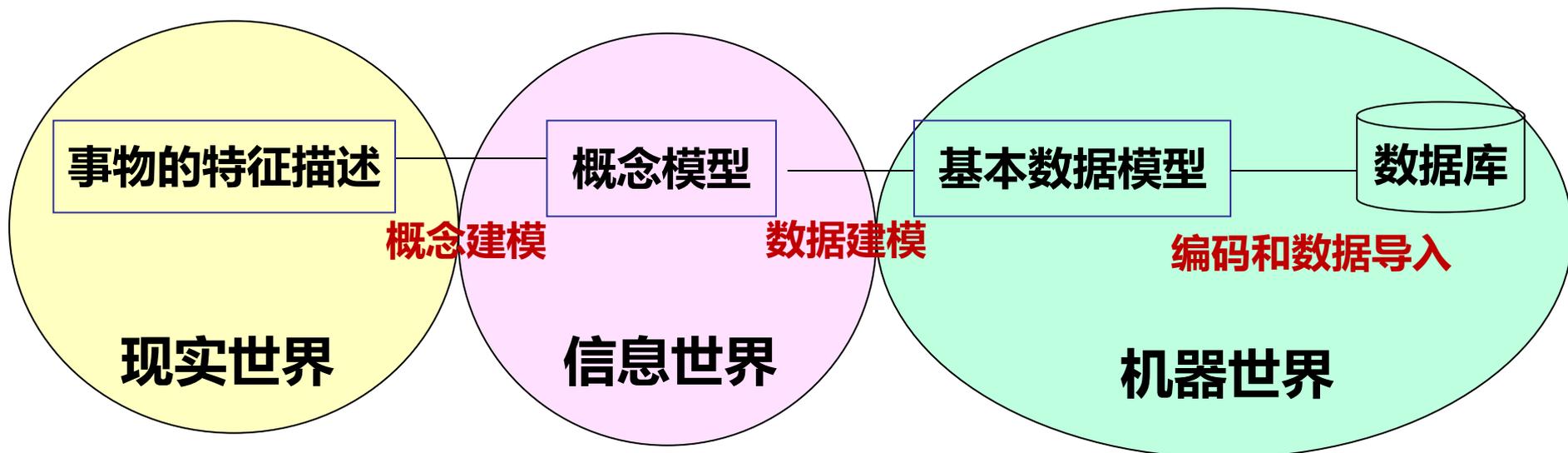
2、数据模型的要求

- 真实模拟世界
- 易于理解和交流
- 便于在计算机上实现

一、数据模型的基本概念

3、数据模型的分类

- **概念模型**：模拟性强，便于理解、交流、扩充
- **基本数据模型**：便于计算机实现



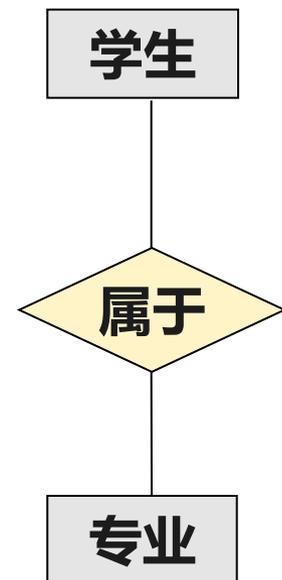
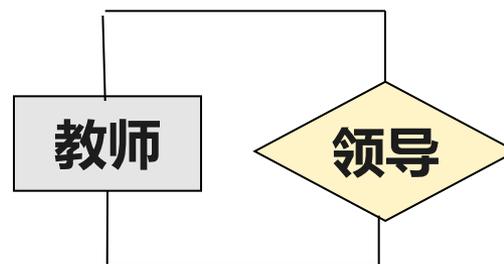
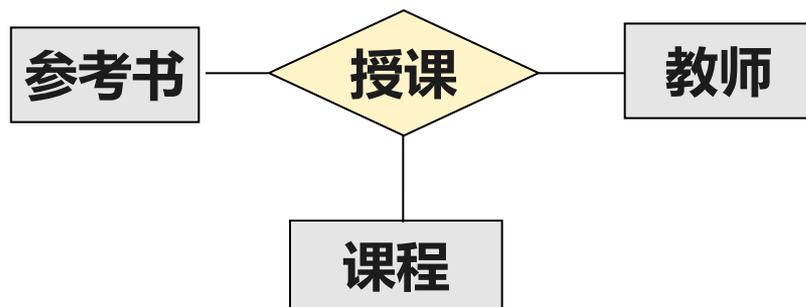
二、概念模型

1. 基本术语

- **实体**：客观存在并相互区别的事物，具体或抽象
- **实体集**：同类实体的集合
- **实体型**：同类实体所具有的共同特征和性质
- **属性**：实体集共同具有的某一特征
- **域**：属性的取值范围
- **码**：唯一标识实体的属性子集
- **联系**：实体内部各属性之间的联系
实体和实体之间的联系

实体间联系: 联系的阶

- 两个实体集间的联系: 一对一、 一对多、 多对多
- 多个实体集间的联系: 一对一、 一对多、 多对多
- 同一实体集内部的联系: 一对一、 一对多、 多对多

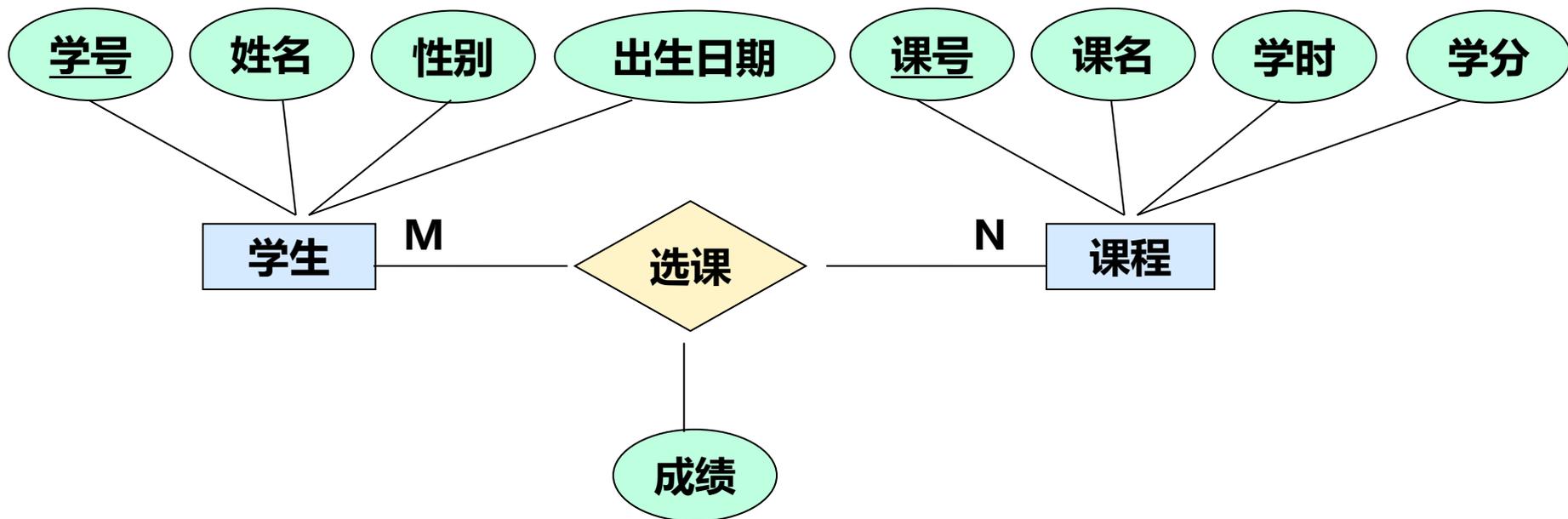


2. 概念建模的要素和表示

(1) 实体：用矩形表示

(2) 联系：菱形表示

(3) 属性：用椭圆表示



三、基本数据模型

1. 组成要素

(1) 数据结构：数据库静态特性的描述

- 描述数据库对象的类型、内容、性质
- 描述数据库对象间的联系

(2) 数据操作：数据库动态特性的描述

- 操作的含义、符号、规则、语言

(3) 数据的约束条件

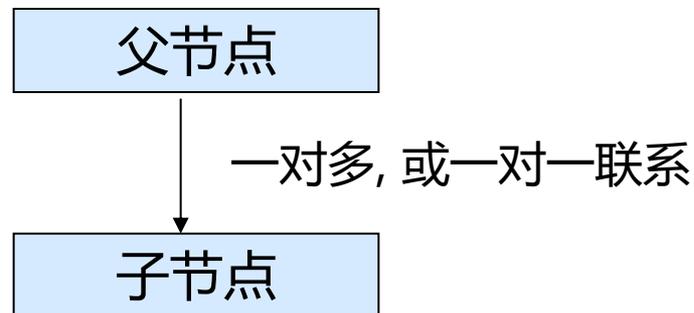
- 应用无关的约束条件。例如：实体应有唯一的标识
- 应用相关的约束条件。例如：年龄小于40岁

三、基本数据模型

2. 分类

- ◆ **非关系模型：层次模型和网状模型**

- **非关系模型的基本层次关系**



- ◆ **关系模型**

- ◆ **面向对象模型**

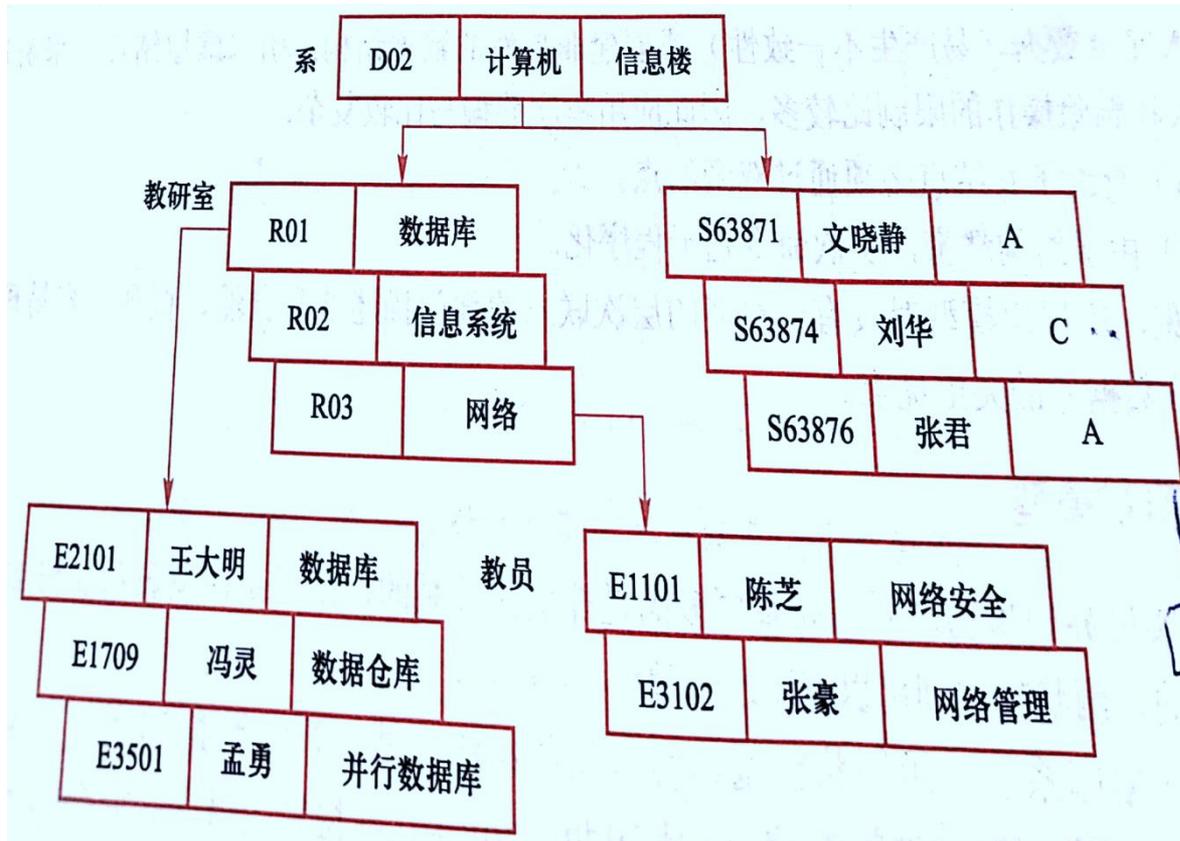
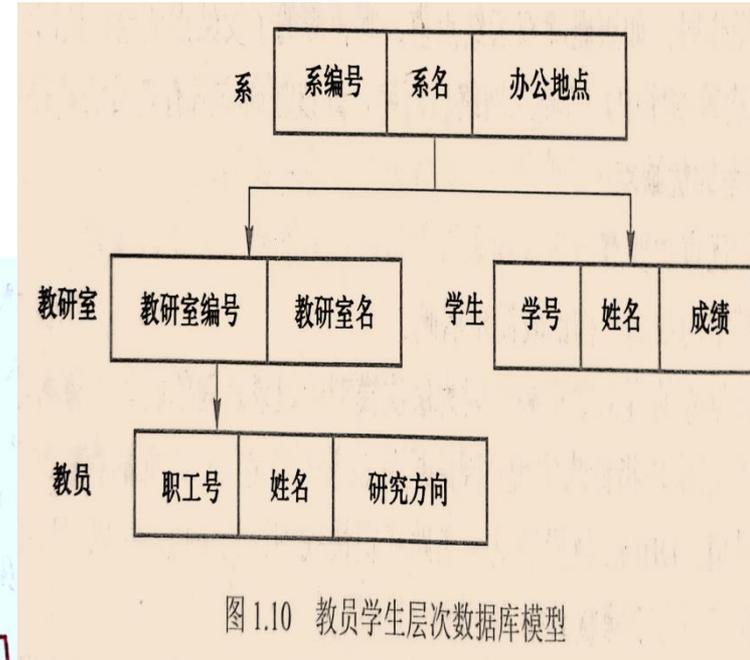
- ◆ **NoSql 模型**

3. 层次模型

- 典型代表：IMS，1968年IBM推出的第一个大型商业数据库管理系统

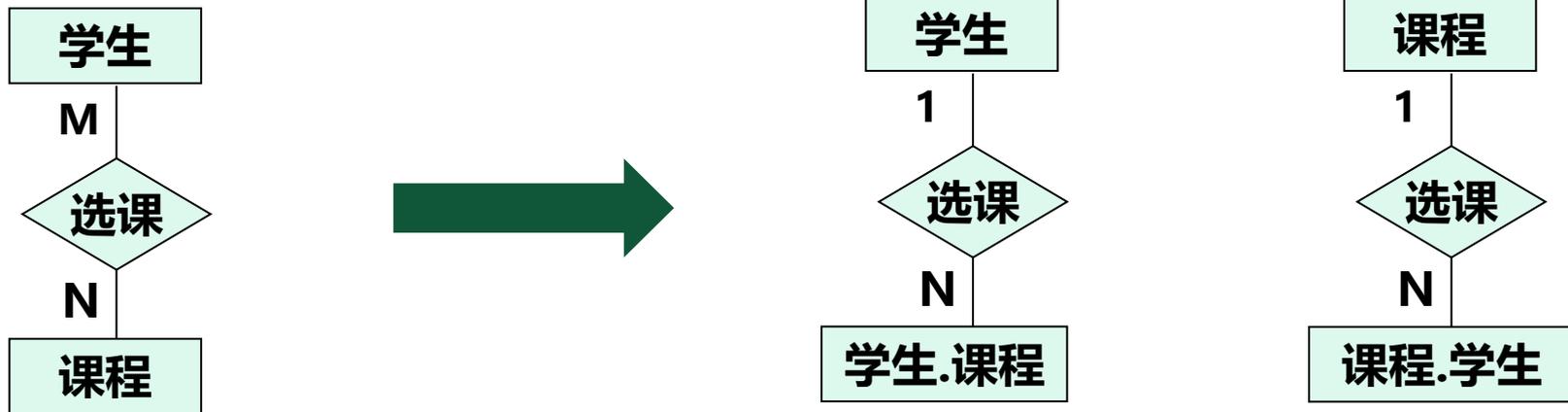
(1) 数据结构

- 有且仅有一个节点没有父节点（根节点）
- 根以外，其他节点有且仅有一个父节点



(1) 数据结构

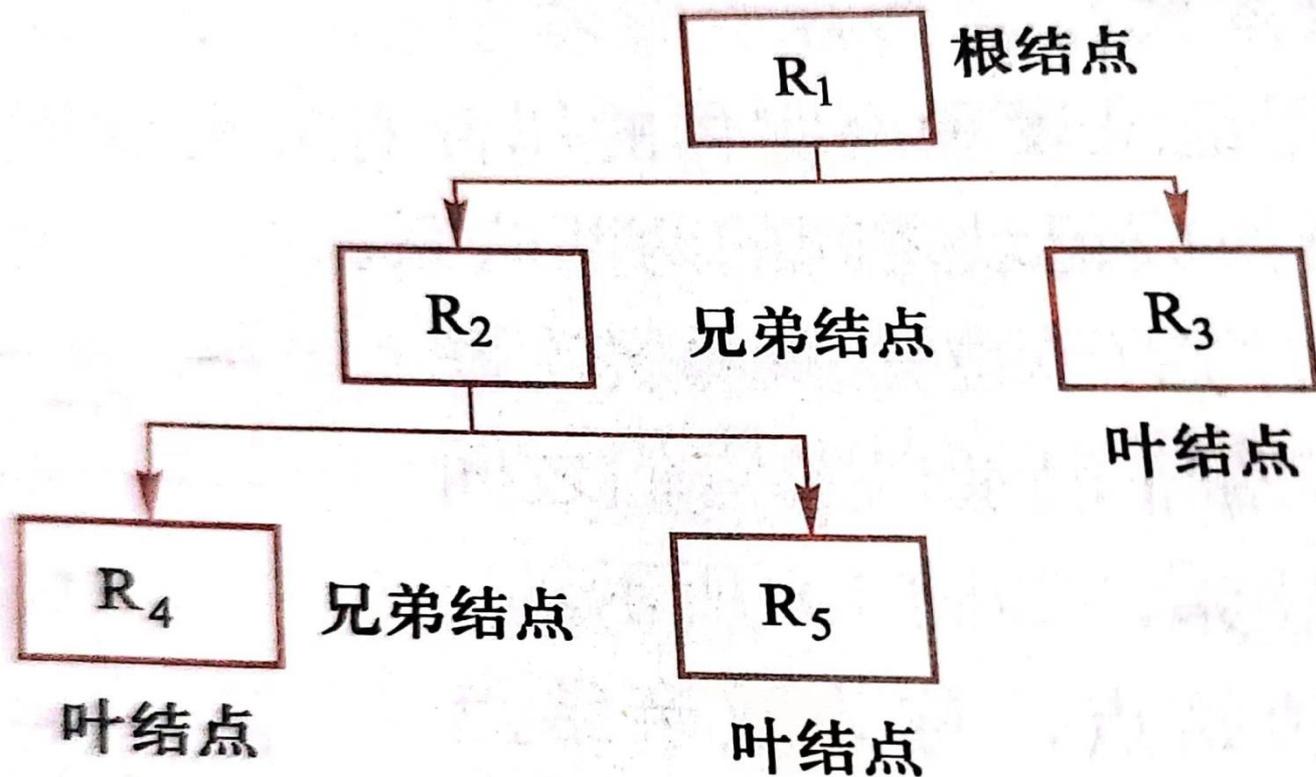
■ 多对多联系的表示



问题：如何在表达关联的前提下，设计学生和课程数据的存储策略？

- 冗余节点法：存储位置可任意变更，存在冗余
- 虚拟节点法：无冗余，但改变存储位置需修改指针

- **存储结构**
 - 邻接法：**记录值按特定顺序（先根、后根）依次存放
 - 链接法：**用指针代替顺序存储 [子女兄弟法，前序穿越法]



先根顺序: $R_1-R_2-R_4-R_5-R_3$

中根顺序: $R_4-R_2-R_5-R_1-R_3$

后跟顺序: $R_4-R_5-R_2-R_3-R_1$

3. 层次模型

(2) **数据操作：支持数据的增、删、改、查**

(3) **完整性约束**

- **插入约束：无父不能插子**
- **删除约束：删父则自动删子**
- **修改约束：数据一致性**

(4) **特点：**

- **优点：简单、查询效率高，实现了完整性约束**
- **缺点：复杂情况(多对多联系、多个父节点)，表达方法复杂**
完整性约束严格、数据操纵不灵活
有时数据操作不方便（必须沿树的路径查询）

4. 网状模型: DBTG系统方案

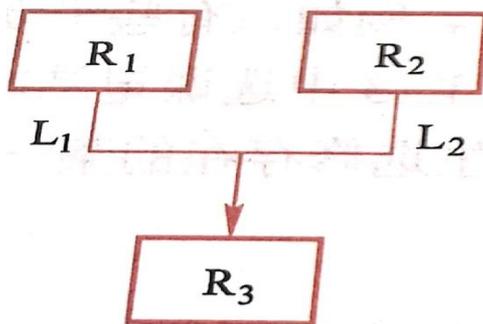
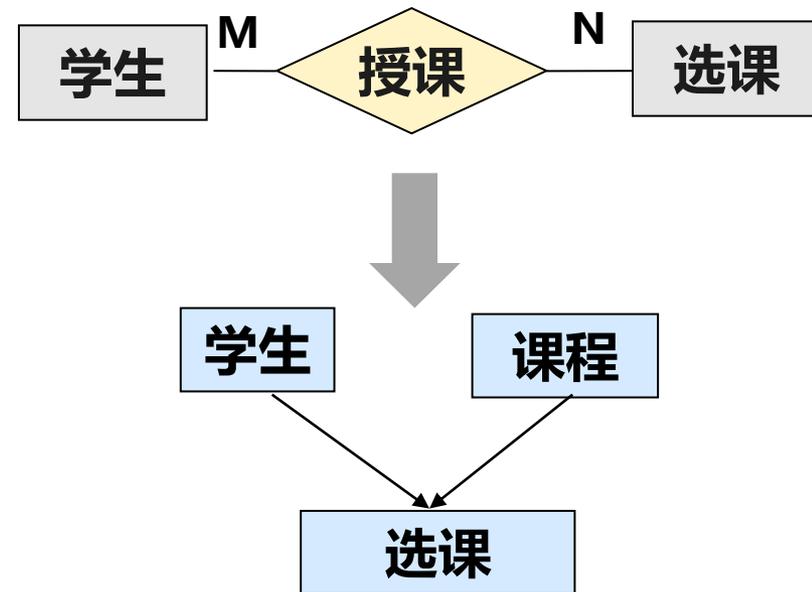
(1) 数据结构

- 允许一个以上节点没有父节点 (根节点)
- 节点可以有多个父节点

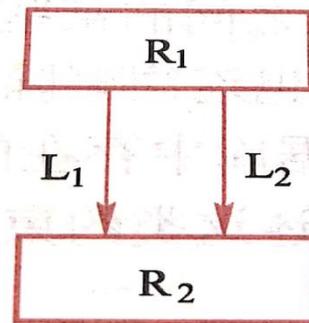
(2) 联系的表示

- 多对多联系: 引进联结记录
- 两个节点间可有多多个联系, 分别命名

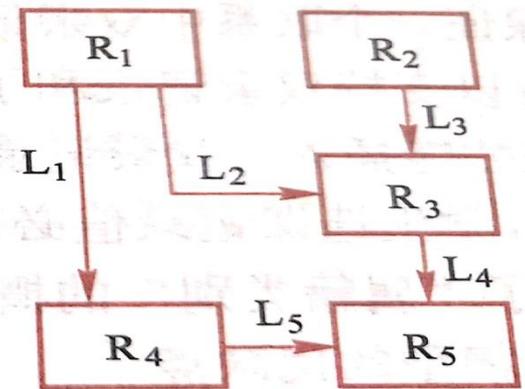
(3) 数据操作: 增、删、改、查



(a)



(b)



(c)

4. 网状模型: DBTG系统方案

(4) 完整性约束

- 支持码的概念
- 联系是一对多的
- 支持灵活的约束条件

(5) 存储结构

- 记录间的联系用单向链实现（链接法）

(6) 特点

- 直接描述世界、性能好
- 模型实现复杂
- 检索需要指明路径，语言编程负担重

5. 关系模型

1970年，E.F.Codd发表系列论文，提出关系模型

(1) 数据结构：建立在数学概念的基础上

- 关系模型由一组“关系”组成
- 每个“关系”的数据结构是一张规范化的“二维表”
- 相关术语：关系——元组——属性——域——码——分量
- 关系模式：关系名（属性1，……，属性n）

学号	姓名	年龄	专业编号
S001	王晓明	19	08
S002	张文	20	09

5. 关系模型

表 1.5 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）

5. 关系模型

(2) 数据操作

- 类型：增、删、改、查
- 特点：集合操作，非过程化

(3) 完整性约束

- 实体完整性、参照完整性、用户自定义完整性

(4) 存储结构：文件形式

(5) 优点：数学基础严密

概念单一

存取路径透明

缺点：查询需要内部优化模块

1.3 数据库系统结构

■ 应用体系结构

- 集中式结构
- 客户/服务器结构
- 分布式结构
- ○ ○ ○ ○ ○ ○

■ 数据库系统结构：指数据层面的系统架构

1.3 数据库系统结构

一、基本概念

1. **型(Type)**: 某一类数据的结构和属性的说明
2. **值(Value)**: 型的一个具体赋值
3. **模式(Schema)**: 数据库中全体数据的逻辑结构和特征的描述, 仅涉及型, 不涉及值
4. **实例(Instance)**: 模式的一个具体值称为模式的一个具体实例

二、数据库的三级模式结构

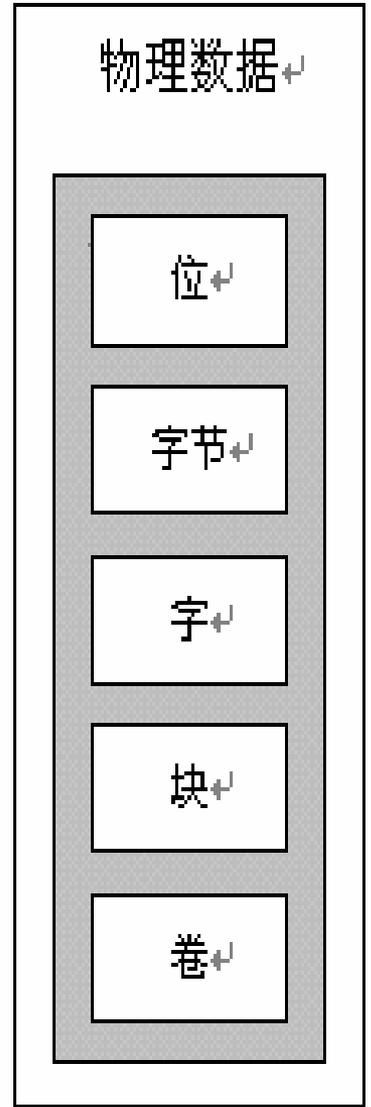
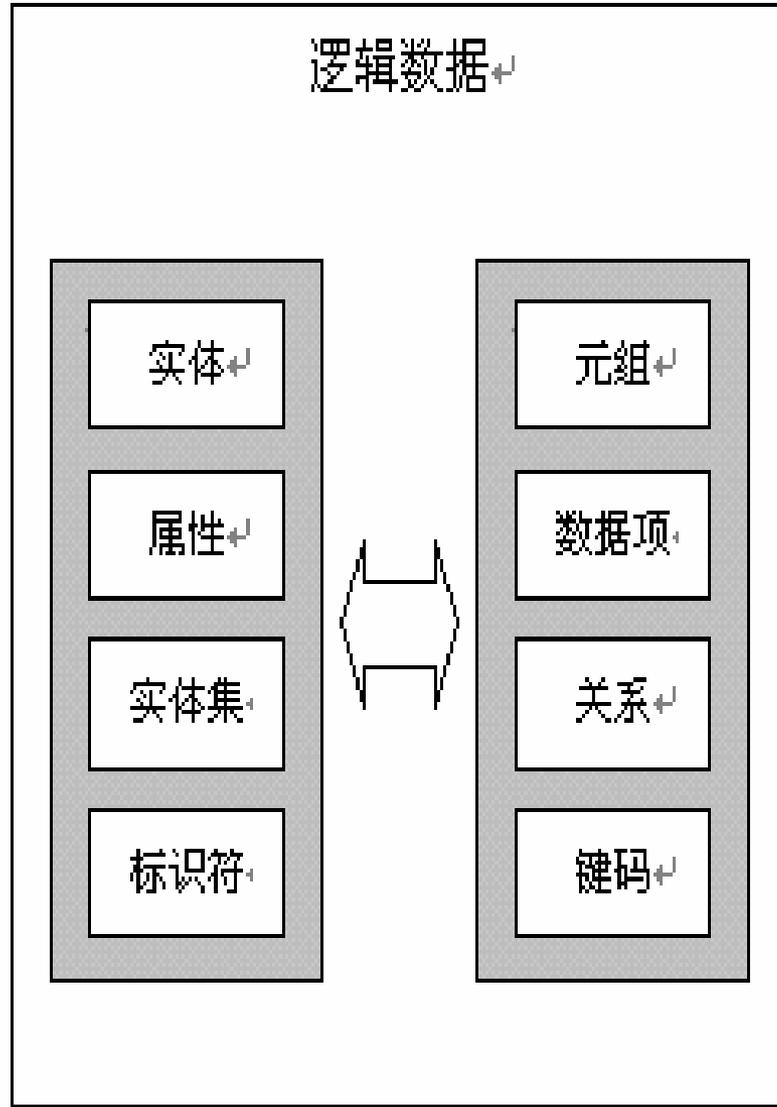
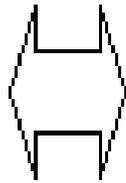
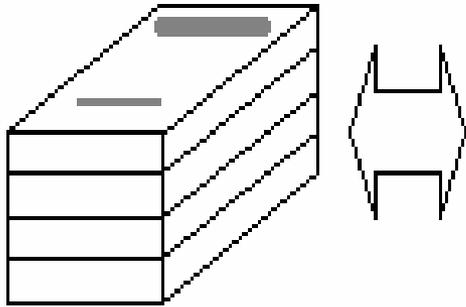
1. 模式（逻辑模式、全局模式）：唯一的

- 概念：数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图，全局视图
- 定义：由DDL定义

2. 内模式（物理模式，存储模式）：唯一的

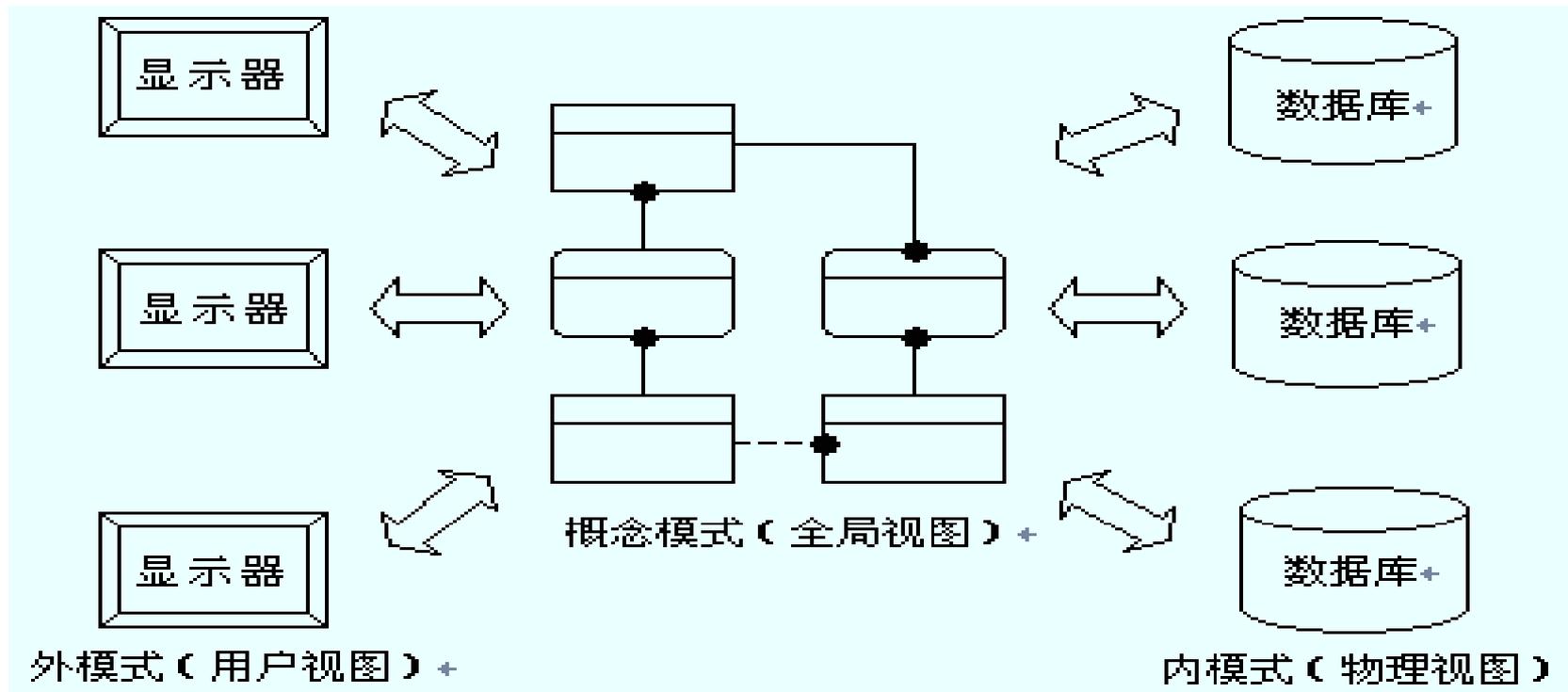
- 概念：数据物理结构和存储方式的描述，是数据在数据库内部的表现方式
- 定义：由DDL定义

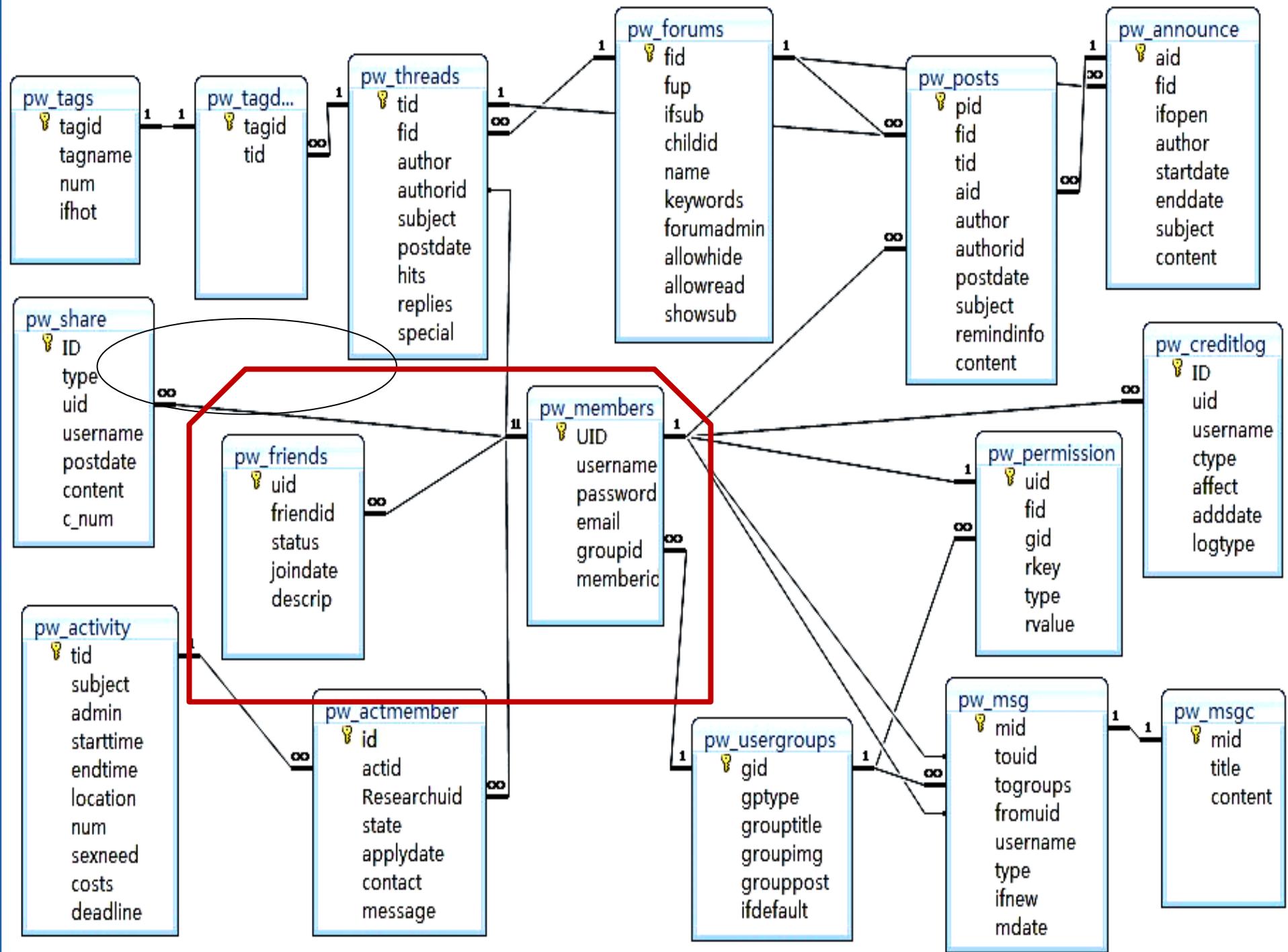
实体 (书籍)



3. 外模式（子模式，用户模式）：面向应用，面向用户

- 概念：某一数据库用户看见和使用的局部数据逻辑结构和特征的描述，一个外模式是模式的一个子集，用户视图
- 定义：由DDL定义



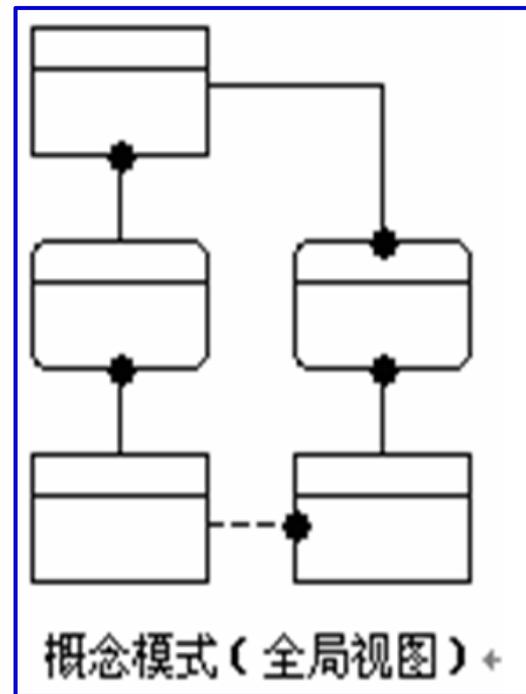
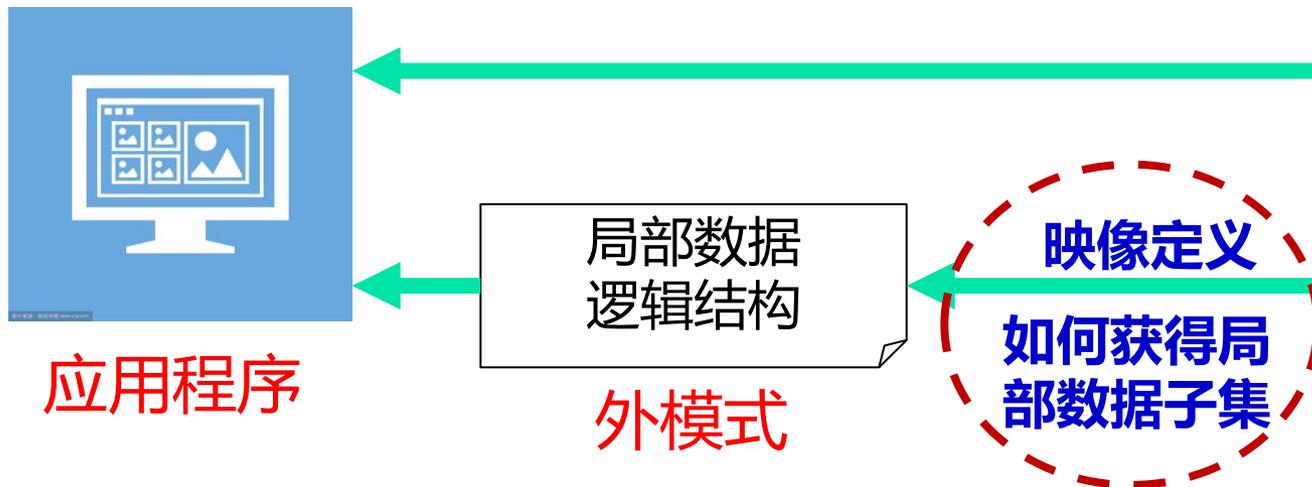


三、二级映象与数据独立性

- ◆ **数据独立性：程序与数据间的独立性，数据结构改变，代码不变**

1. 外模式/模式映像与数据的逻辑独立性

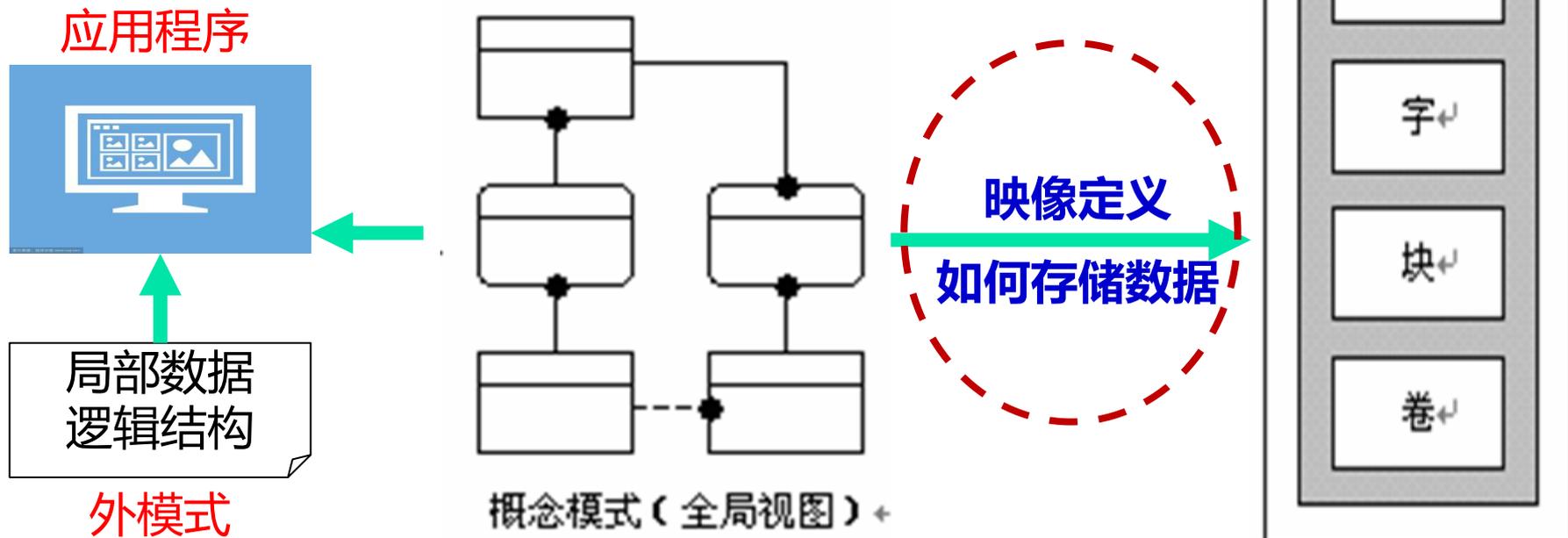
- **映像：描述外模式与模式之间的对应关系，包含在外模式定义中**
- **逻辑独立性：模式结构改变时，修改映像，保持外模式不变，进而保持应用程序不变**



三、 二极映象与数据独立性

2. 模式/内模式映像与数据的物理独立性

- 映象：描述数据库模式与内模式之间的对应关系，有些包含在模式定义中，有些独立存在
- 物理独立性：修改内模式时，只需修改模式/内模式映像，模式保持不变，应用程序保持不变



数据独立性含义

独立性	特性
逻辑独立性	数据库逻辑结构改变，代码不变
	实现：定义和维护外模式/模式映像
物理独立性	数据库存储模式改变，代码不变
	实现：定义和维护模式/内模式映像
分布独立性	数据库片段的存储场地变化，代码不变
	实现：定义和维护数据分片和分布模式

1.4 数据库系统的组成

一、硬件系统

- ◆ **大内存**：加载OS、DBMS核心模块、应用程序，分配各种缓冲区
- ◆ **足够大的磁盘或磁盘阵列**：存储组织级的大规模数据
- ◆ **处理器速度和I/O通道能力**：提高数据处理能力和数据传送效率

二、软件系统

- ◆ 操作系统
- ◆ 数据库管理系统
- ◆ 开发工具和开发环境
- ◆ 数据库应用系统

1.4 数据库系统的组成

三、数据库

- ◆ 系统数据库、用户数据库、数据库镜像或快照等

四、人员

- ◆ 数据库管理员 (Database Administrator, DBA)
- ◆ 系统分析员和数据库设计人员 (System Analyst, SA)
- ◆ 应用程序员 (Programmer & Programmer leader)
- ◆ 用户：偶然用户、简单用户、复杂用户

1.4 数据库系统

五、DBA职责：全面管理和控制数据库系统

- ◆ 决定数据库中的信息内容和结构
- ◆ 决定数据库的存储结构和存取策略
- ◆ 定义数据库的安全性要求和完整性约束条件
- ◆ 监控数据库的使用和运行
- ◆ 数据库的改进和重组、重构

1.5 数据库技术的研究领域

1. DBMS软件研制

- ◆ 研究目标：开发DBMS及配套的工具软件、中间件等
- ◆ 研究问题：新的应用领域、新的数据格式、能力提升等

2. 数据库设计

- ◆ 设计方法：数据模型，建模方法，设计规范等
- ◆ 设计工具：计算机辅助设计，CASE工具等

3. 数据库理论与模型

- ◆ 关系数据理论，OO模型，并行数据库
- ◆ NoSql模型、内存数据库
- ◆ OLAP、数据挖掘、机器学习