

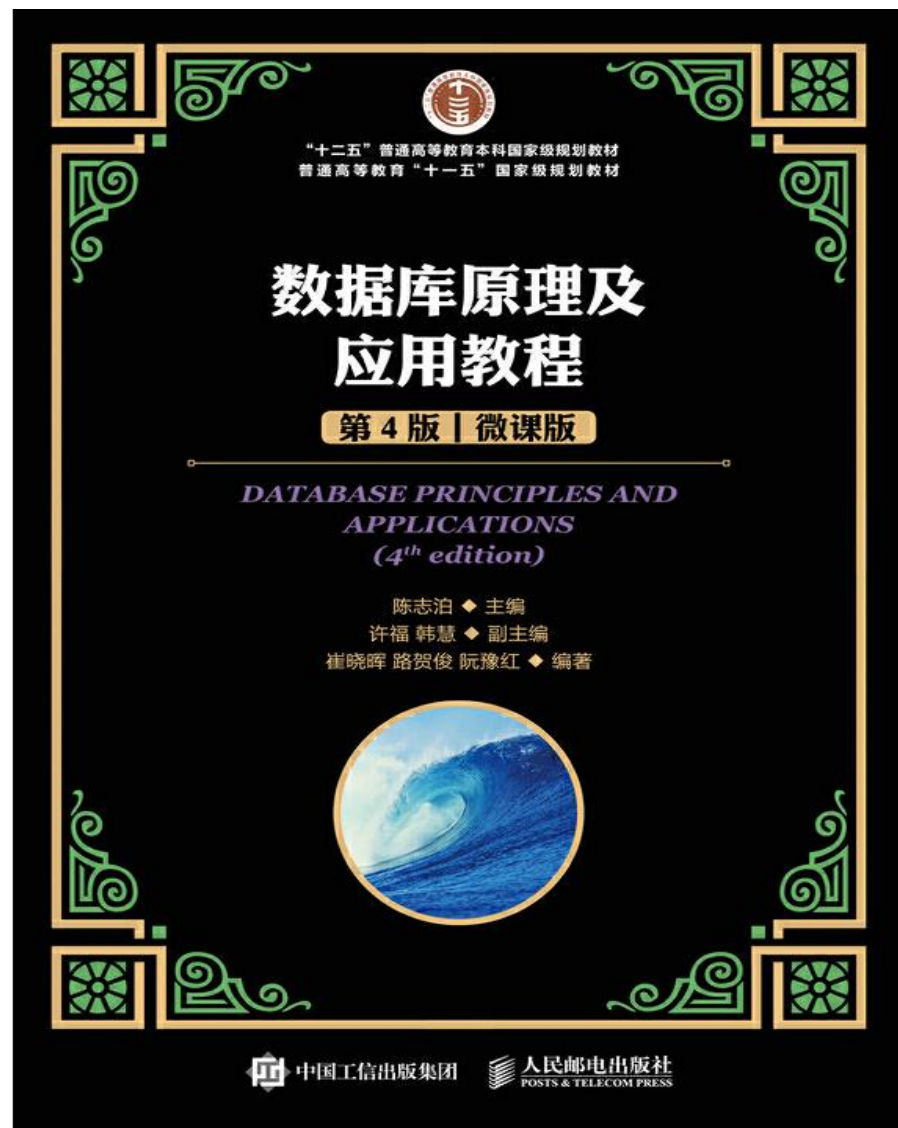
# 数据库原理及应用教程 (第4版)

“十二五”普通高等教育本科国家级规划教材  
微课版，对重点和难点进行视频详解

中国工信出版集团

人民邮电出版社

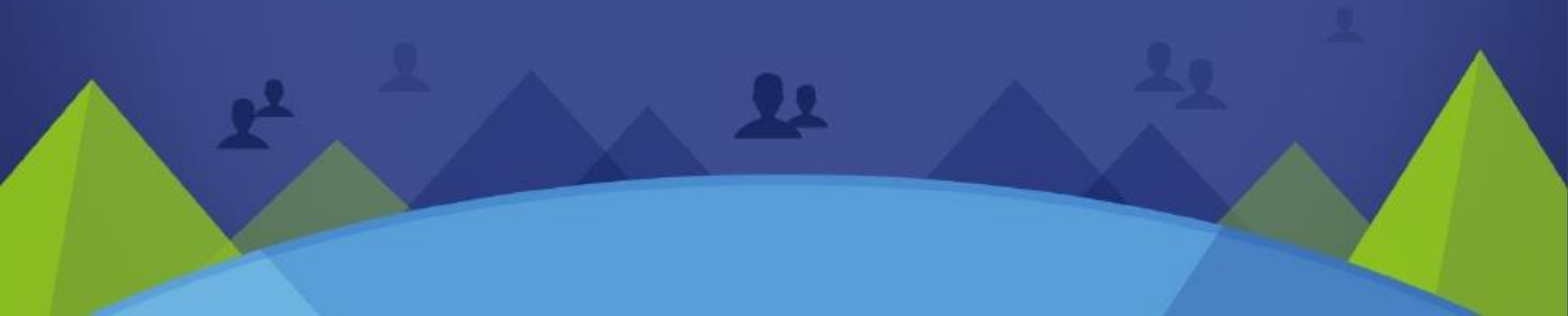
- 总共48学时，课堂讲授40学时，实践8学时
- 先导课程：程序设计、数据结构
- 平时成绩：作业、课内测试、实验
- QQ群：1054877013
- 我的邮箱：[515937274@qq.com](mailto:515937274@qq.com)



## – COURSE OBJECTIVES

- To study databases from four viewpoints, in particular, those of the **database user**, the **database designer**, the **database manager**, and the **database system implementer**.
- To understand the principles of relational database management systems and their languages – in particular SQL.
- To learn the methodology for building applications on top of database management systems – the so called data modeling process.
- To investigate the internal operations of relational database management systems.

# 第1章 数据库系统概述



- 1.1 信息、数据、数据处理与数据管理
- 1.2 数据库技术的产生、发展 (\*)
- 1.3 数据库系统的组成
- 1.4 数据库系统的内部体系结构 (\*)
- 1.5 数据库系统的外部体系结构
- 1.6 数据库管理系统 (DBMS)
- 1.7 数据模型 (\*)
- 1.8 三个世界及其有关概念 (\*)
- 1.9 四种数据模型
- 1.10 数据库领域的新技术



数据的仓库，即数据存放的地方

- ◆ 通讯录：小数据库可用手工管理
- ◆ 图书馆：大型数据库必须由计算机管理

计算机三大主要应用领域？

科学计算

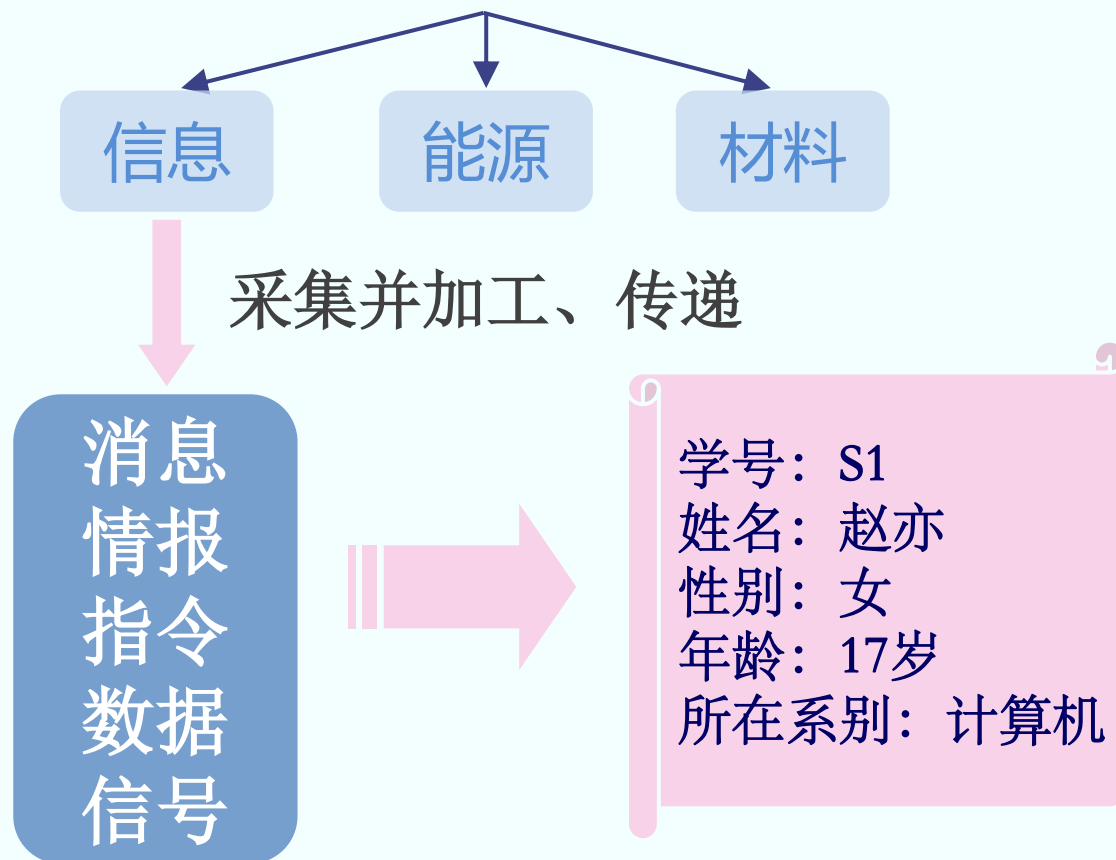
过程控制

数据处理

70%

### 1.1.1 数据与信息

客观世界的三大要素



信息的特征

源于物质和能量

可以感知

可存储、加工、  
传递和再生



### 数据

- 数据是信息的具体表现形式
- 数据的表现形式

数字  
字符  
图形  
图象  
声音等

S1  
赵亦  
女  
17岁  
计算机

### 数据与信息的联系

数据是信息的符号表示

信息是数据的内涵，是对数据的语义解释

### 1.1.2 数据处理与数据管理

#### 数据处理

将数据转换成  
信息的过程  
包括数据的收集、  
管理、加工利用  
乃至信息分析、  
演变推导

产出  
输出结果

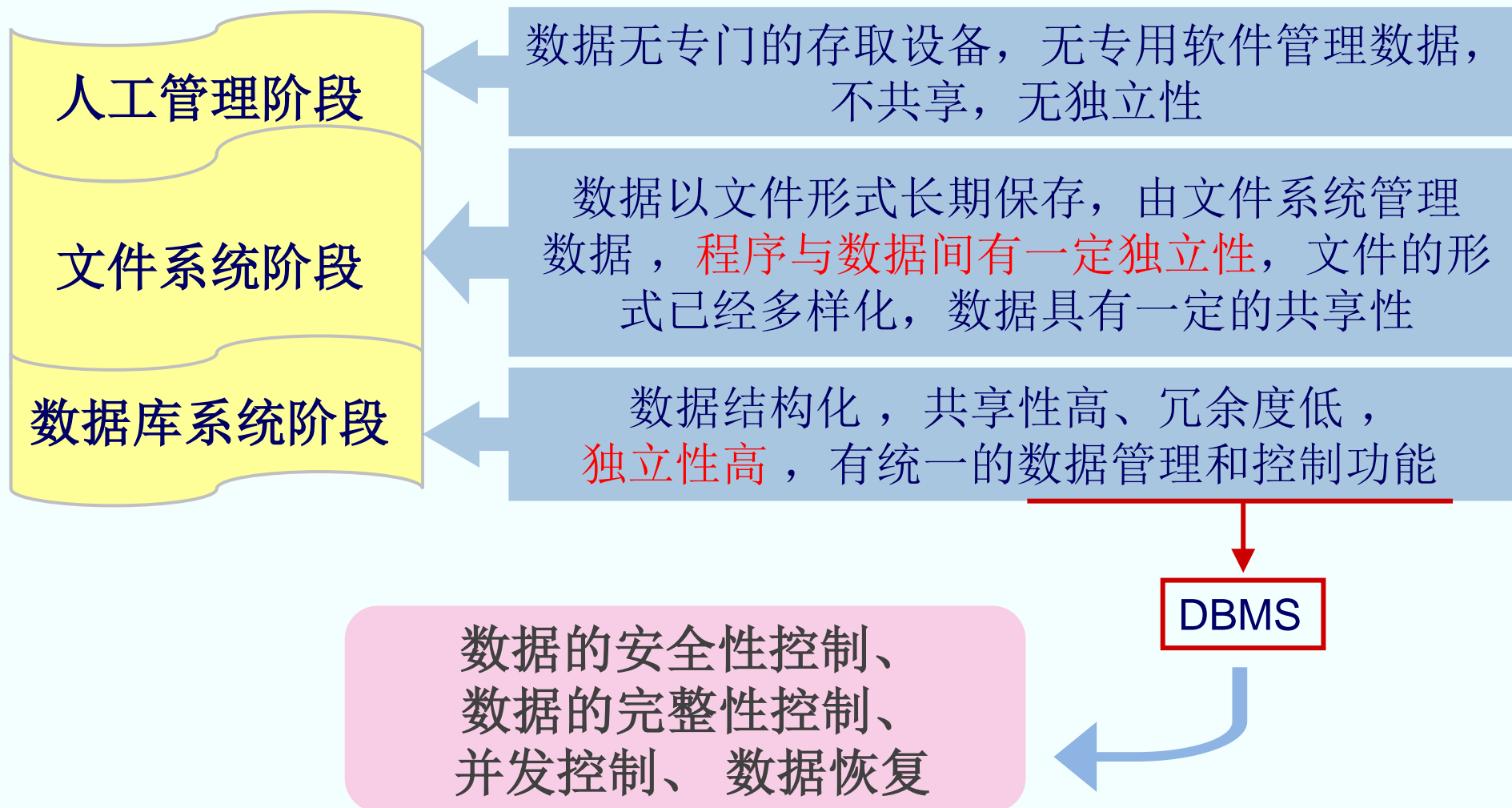
#### 数据管理

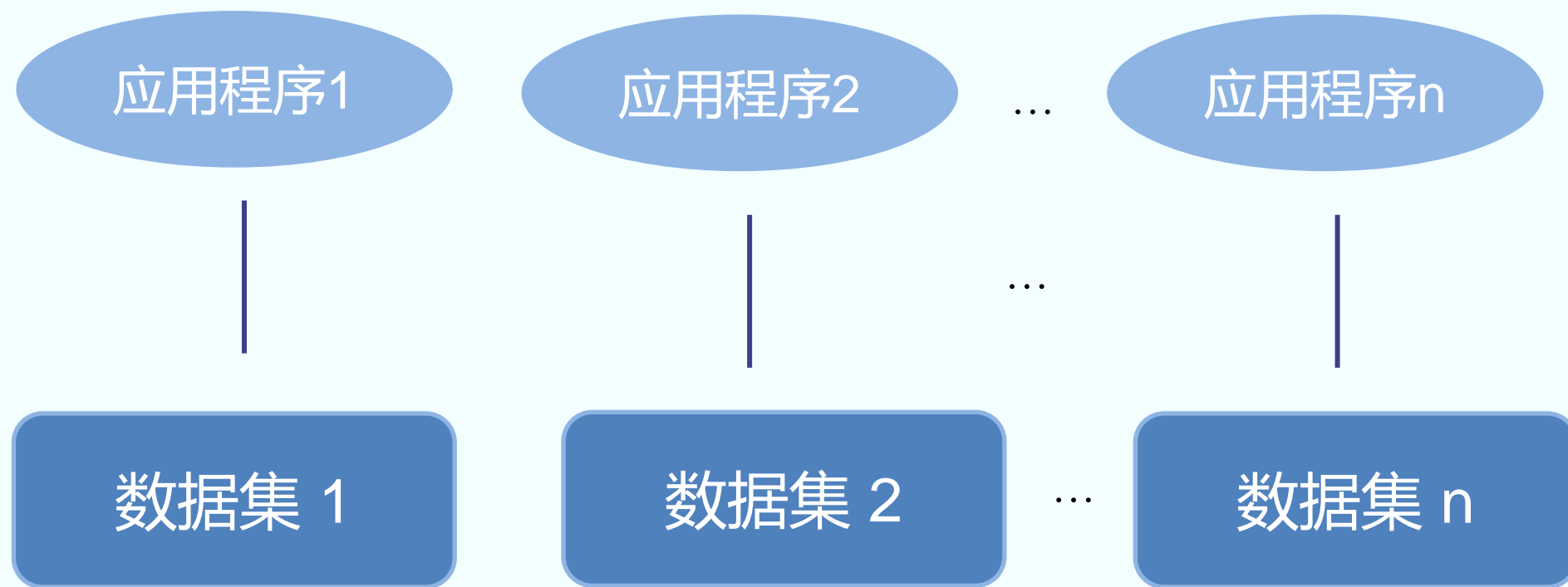
数据的分类、  
组织、编码  
、存储、维  
护、检索等  
操作

原料  
输入

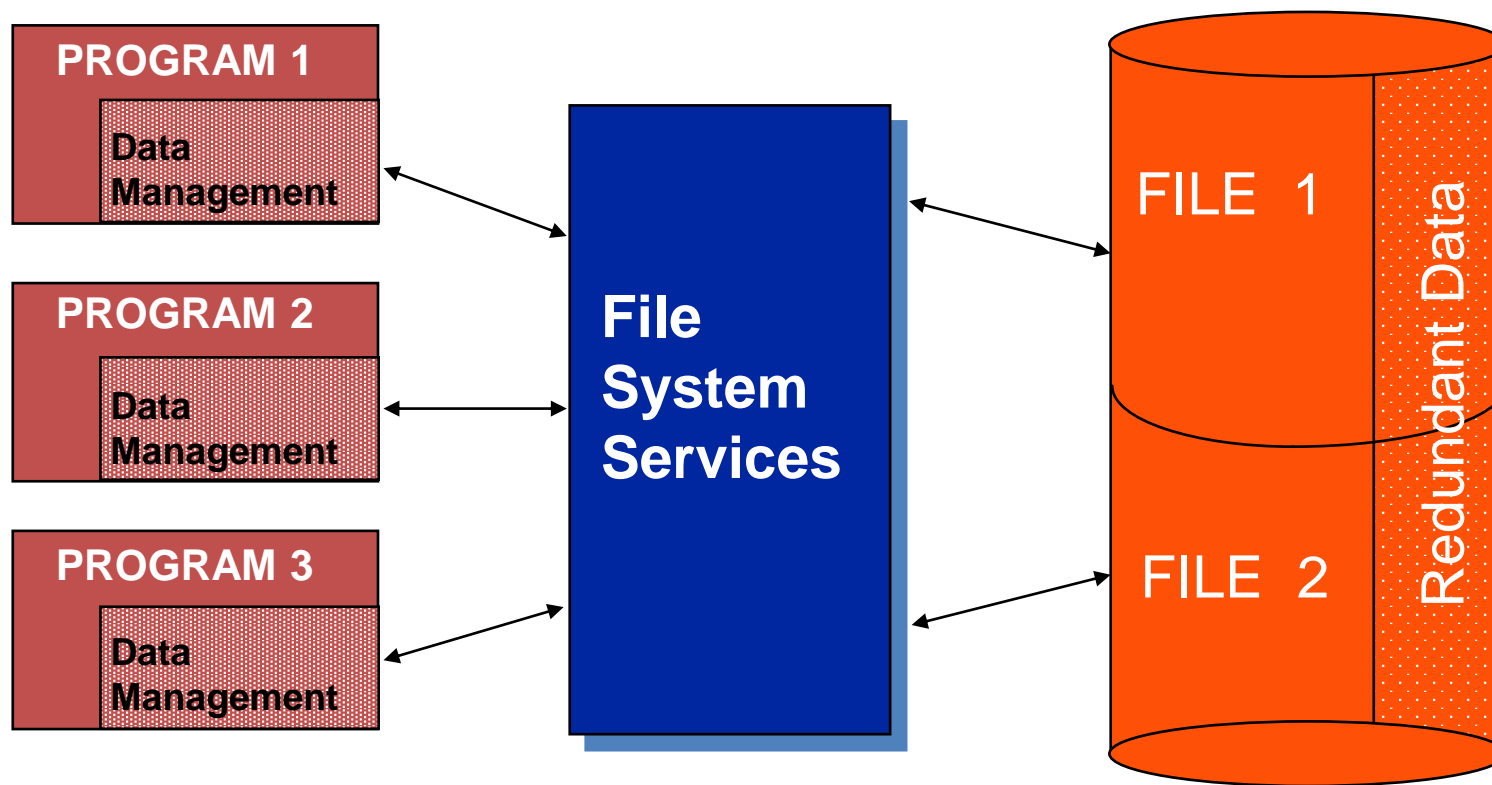
信息 = 数据 + 数据处理

“数据处理”的真正含义应该是为了产生信息而处理数据

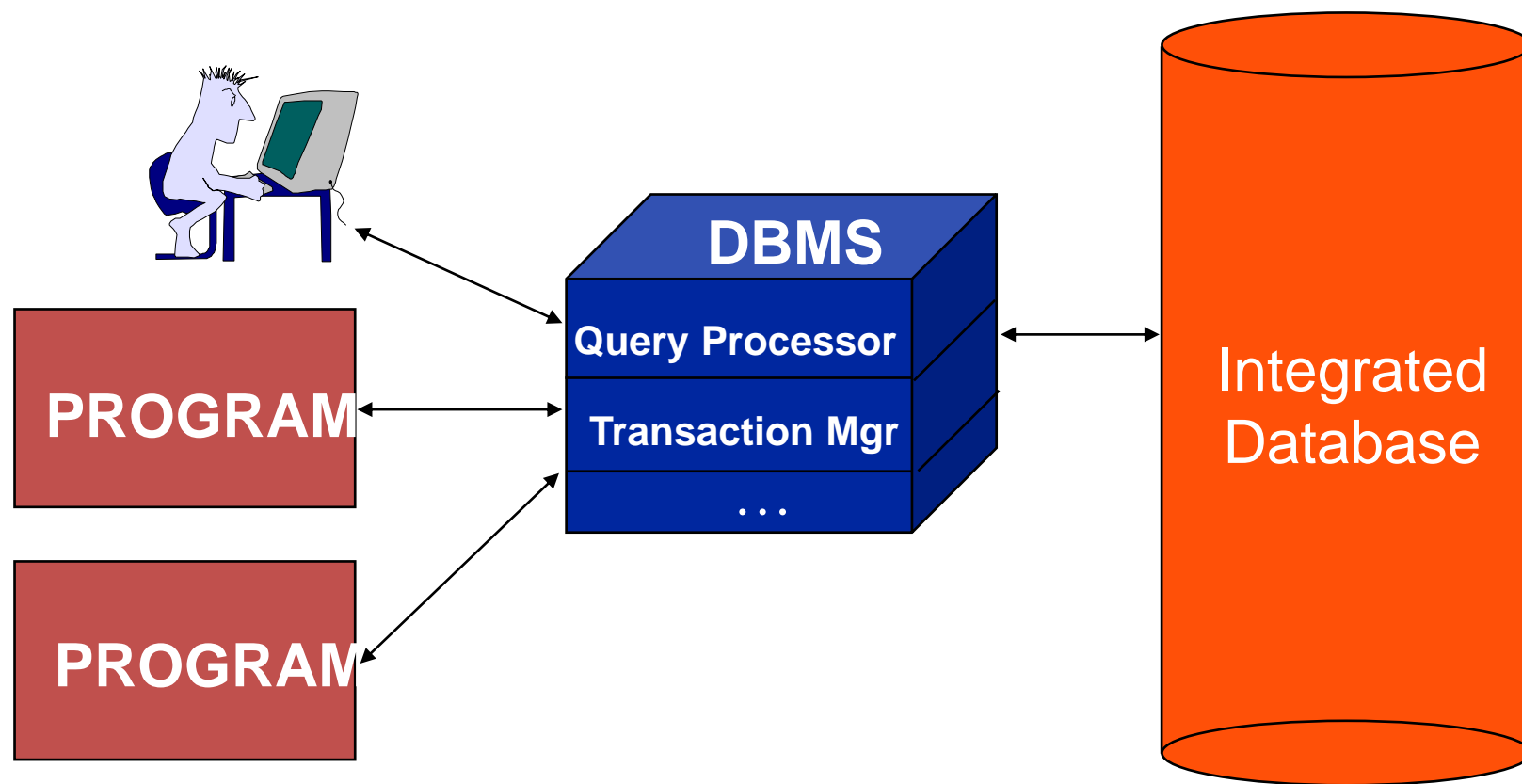




人工管理阶段应用程序与数据之间的对应关系



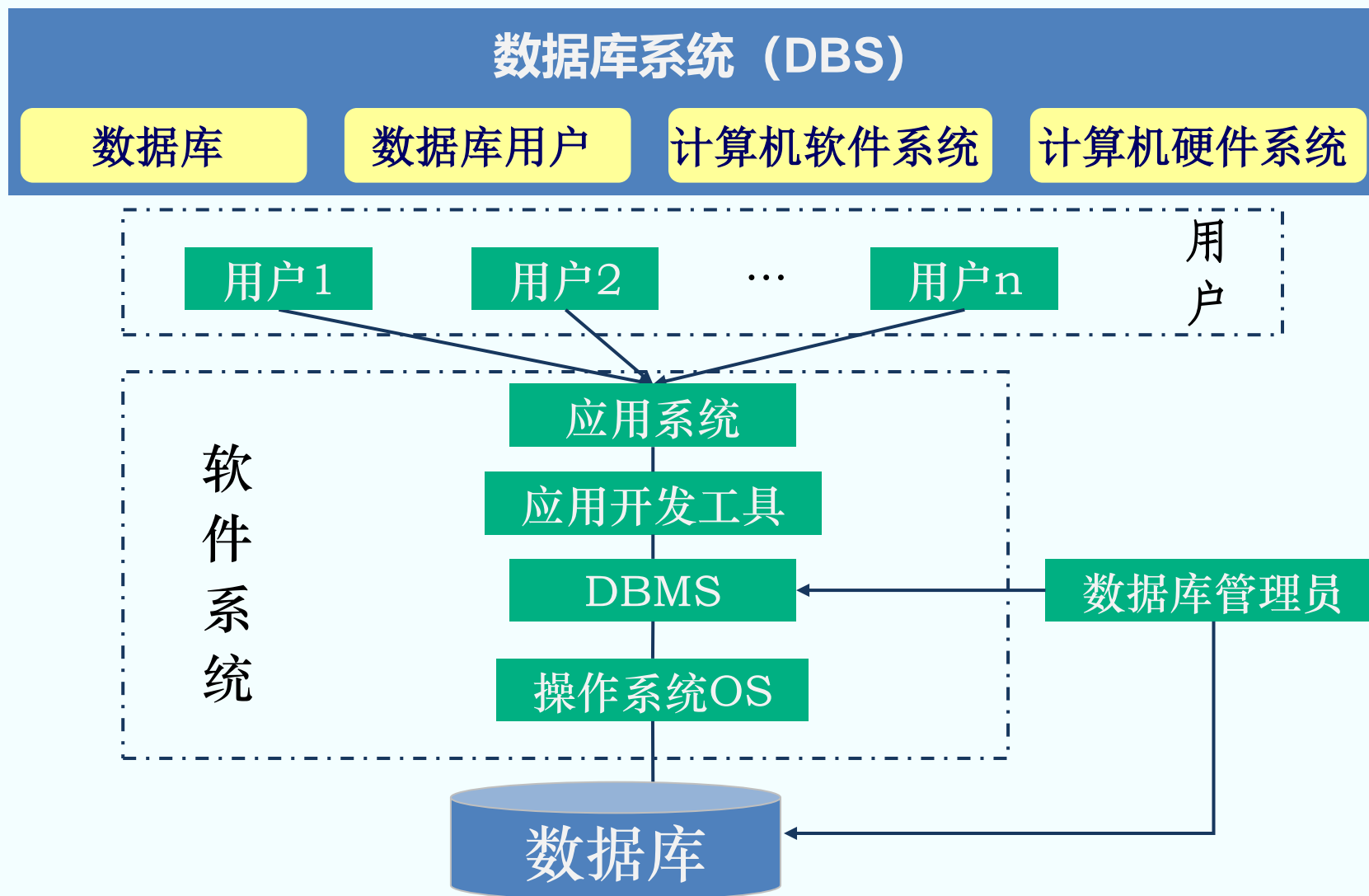
文件系统阶段应用程序与数据间的对应关系



数据库系统阶段程序与数据间的对应关系

## 1.3 数据库系统的组成

## 第1章



### 数据库

集成性，共享性

数据库 (DB) 是存储在计算机内、有组织的、可共享的数据和数据对象集合

### 用户

对数据库进行存储、维护和检索等操作

最终用户

非计算机专业人员  
使用数据库

应用程序员

设计和编制应用程序  
并进行调试和安装

数据库管理员

负责设计、建立、管理和维护数据库以及协调用户对数据库要求的个人或工作团队



### 软件 (Software) 系统

核心

软件系统:

DBMS

OS

应用开发工具

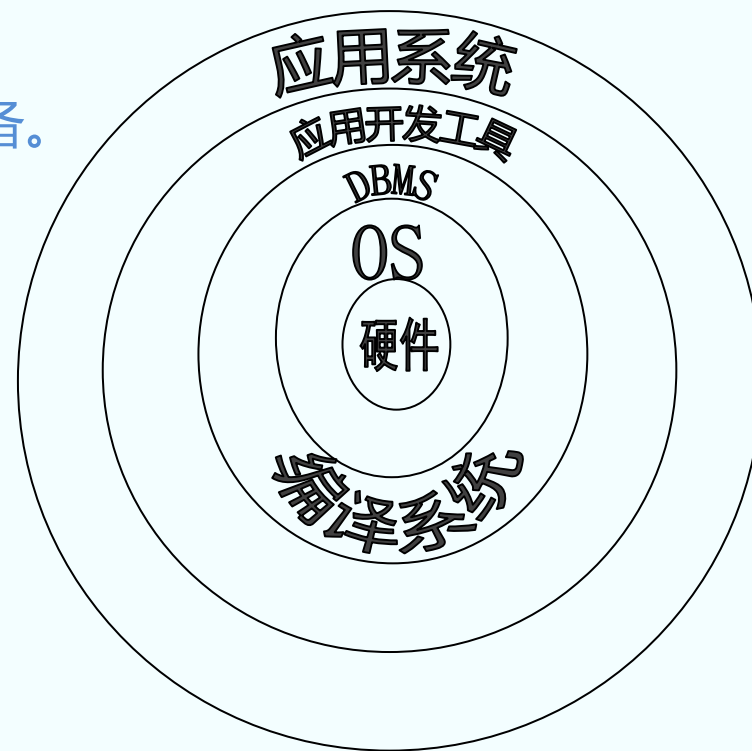
应用系统

### 硬件 (Hardware) 系统

硬件系统指存储和运行数据库系统的硬件设备。

数据库系统在整个计算机系统中的地位如图所示

DBMS在操作系统支持下工作，应用程序在DBMS支持下才能使用数据库。



### 1.4.1 数据库系统的三级模式结构

学号	姓名	性别	年龄	系别
001101	张立	男	20	计算机

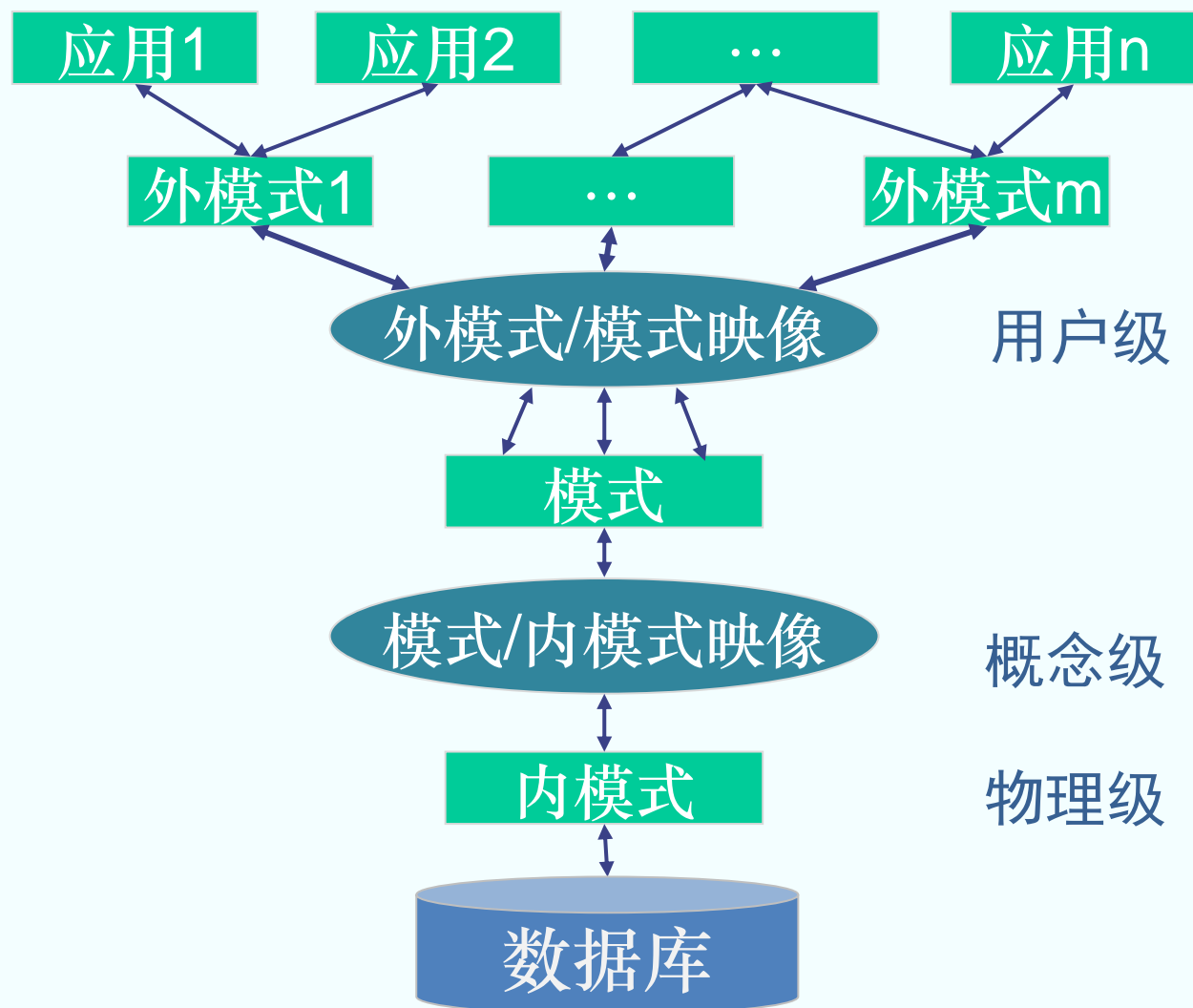
→ 型 → 模式  
→ 值 → 实例

数据库系统内部的体系结构从逻辑上分为三级

外模式

模式

内模式



数据库系统的三级模式结构  
和二级映像功能示意图

可有多個

外模式

数据库用户能看到并允许使用的那部分局部数据的逻辑结构和特征的描述

模式

数据库中全体数据的逻辑结构和特征的描述

只有一个

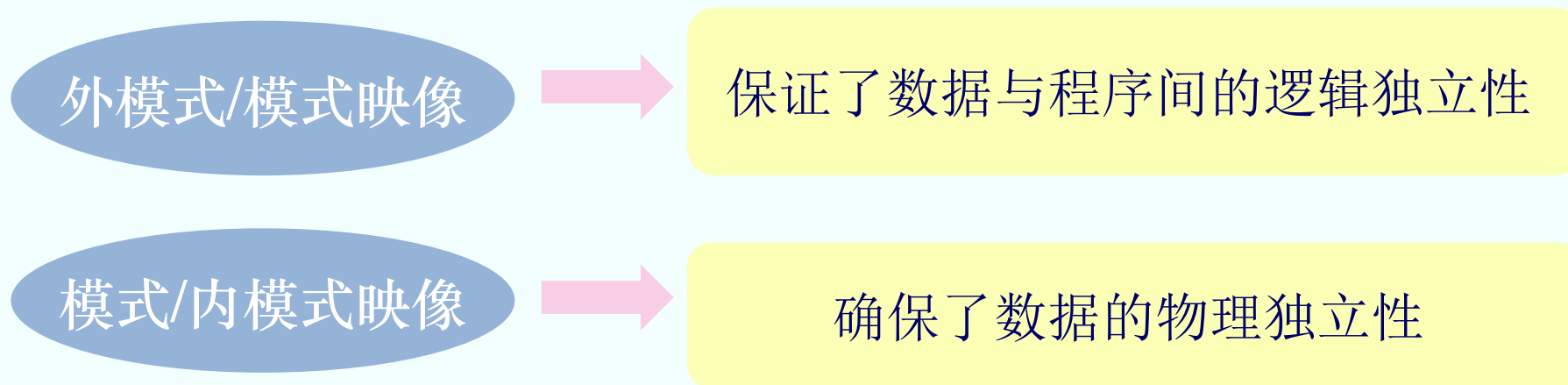
内模式

它是对数据库存储结构的描述，是数据在数据库内部的表示方式

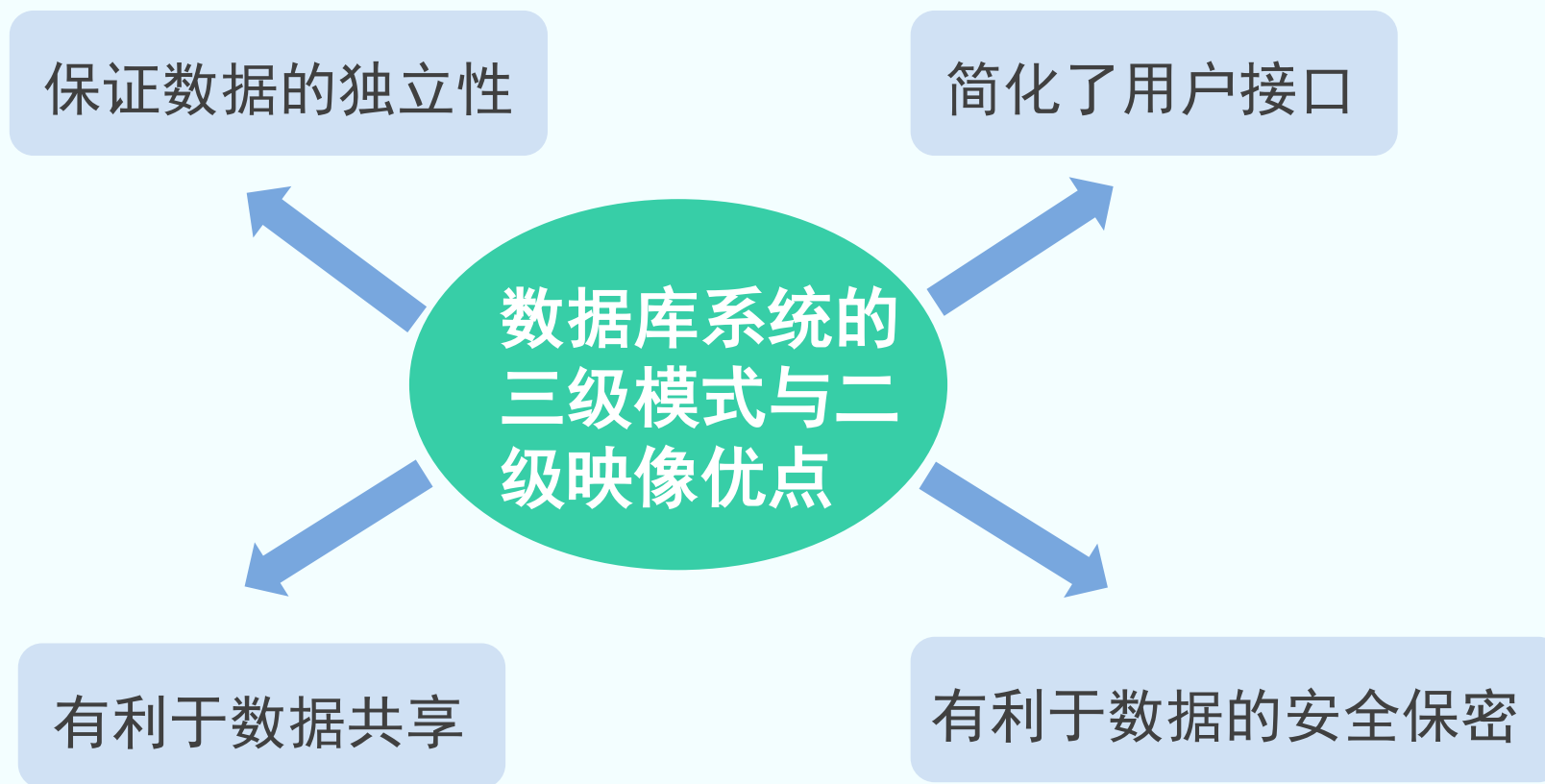
- 内模式是整个数据库实际存储的表示
- 模式是整个数据库实际存储的抽象表示
- 外模式是逻辑模式的某一部分的抽象表示

### 1.4.2 数据库系统的二级映像与数据独立性

- DBMS在三级模式之间提供了二级映像功能，保证了数据库系统中的数据能够具有较高的逻辑独立性与物理独立性。



### 1.4.3 数据库系统的三级模式与二级映像的优点



单用户结构的数据库系统

← 一台计算机，不能共享数据

主从式结构的数据库系统

← 大型主机带多个终端；  
主机处理，终端输出

分布式结构的数据库系统

← 是分布在计算机网络上的多个逻辑相关的数据库的集合

客户/服务器结构的数据库系统

← 把DBMS的功能与应用程序分开；  
管理数据的结点称为服务器  
应用DBMS的结点称为客户机

浏览器/服务器结构的数据库系统

← 将客户端运行的应用程序转移到应用服务器上，充当了客户机和数据库服务器的中介



### 1.6.1 DBMS的主要功能

#### 数据定义功能

定义数据的模式、外模式和内模式三级模式结构  
定义模式/内模式和外模式/模式二级映象  
定义有关的约束条件

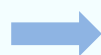


### 数据操纵功能

- 数据库的基本操作：检索、更新（包括插入、修改、删除）等。

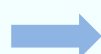
- DML有两类：

自主型



独立使用

宿主型



嵌入在如Fortran、Pascal、C等高级语言中

### 数据库运行管理功能

- 对数据库运行的管理是DBMS运行的核心部分。
- DBMS对数据库的控制主要通过四个方面实现：数据的安全性控制、数据的完整性控制、多用户环境下的数据并发控制和数据库的恢复。

# 1.6 数据库管理系统 (DBMS)

SQLQuery1.sql - LAPTOP-8V5NUPM3.GlobalToyz (LAPTOP-8V5NUPM3\sjf (52))\* - Microsoft SQL Server Management Studio

文件(F) 编辑(E) 视图(V) 查询(Q) 项目(P) 工具(T) 窗口(W) 帮助(H)

GlobalToyz 执行(X)

对象资源管理器

连接

LAPTOP-8V5NUPM3 (SQL Server 14.0.1000.169 - LAPTOP-8V5NUPM3\sjf)

- 数据库
  - 系统数据库
  - 数据库快照
  - CAPcopy
  - GlobalToyz**
- 安全性
  - 服务器对象
  - 复制
  - PolyBase
  - Always On 高可用性
  - 管理
- Integration Services 目录
- SQL Server 代理(已禁用代理 XP)
- XEvent 探查器

SQLQuery1.sql - ...V5NUPM3\sjf (52))\*

```
select * from Toys;
```

100 %

结果 消息

	cToyId	vToyName	vToyDescription	cCategoryId	mToyRate	cBrandI
1	000001	Robby the Whale	A giant Blue Whale with two heavy-duty handles ...	001	8.99	001
2	000002	Water Channel System	Children enjoy playing with water. The Water C...	001	33.99	001
3	000003	Parachute and Rocket	Step on the launch pad and the rocket is launch...	001	6.99	003
4	000004	Super Deluge	Create artificial rainfall in your garden with ...	001	35.99	005
5	000005	Light Show Lamp	Build a revolving, glowing mood lamp with a set...	002	15.99	001
6	000006	Glass Decoration	Make your own beautiful, art work. Set includes...	002	12.99	004
7	000007	Tie Dye Kit	CHECK THE ORIGINAL TOY LIST	002	19.99	002
8	000008	Alice in Wonderland	A huge coloring book.	002	14.99	001
9	000009	Glamorous Doll	She is ready for her acting audition in this bl...	002	18.99	001
10	000011	Sleeping Beauty Doll	Let your child enact this magical fairy tale wi...	002	18.99	005
11	000012	Pet Loving Doll	A beautiful doll with a small pet puppy.	002	10.99	001

查询已成功执行。 | LAPTOP-8V5NUPM3 (14.0 RTM) | LAPTOP-8V5NUPM3\sjf (52) | GlobalToyz | 00:00:00 | 29 行

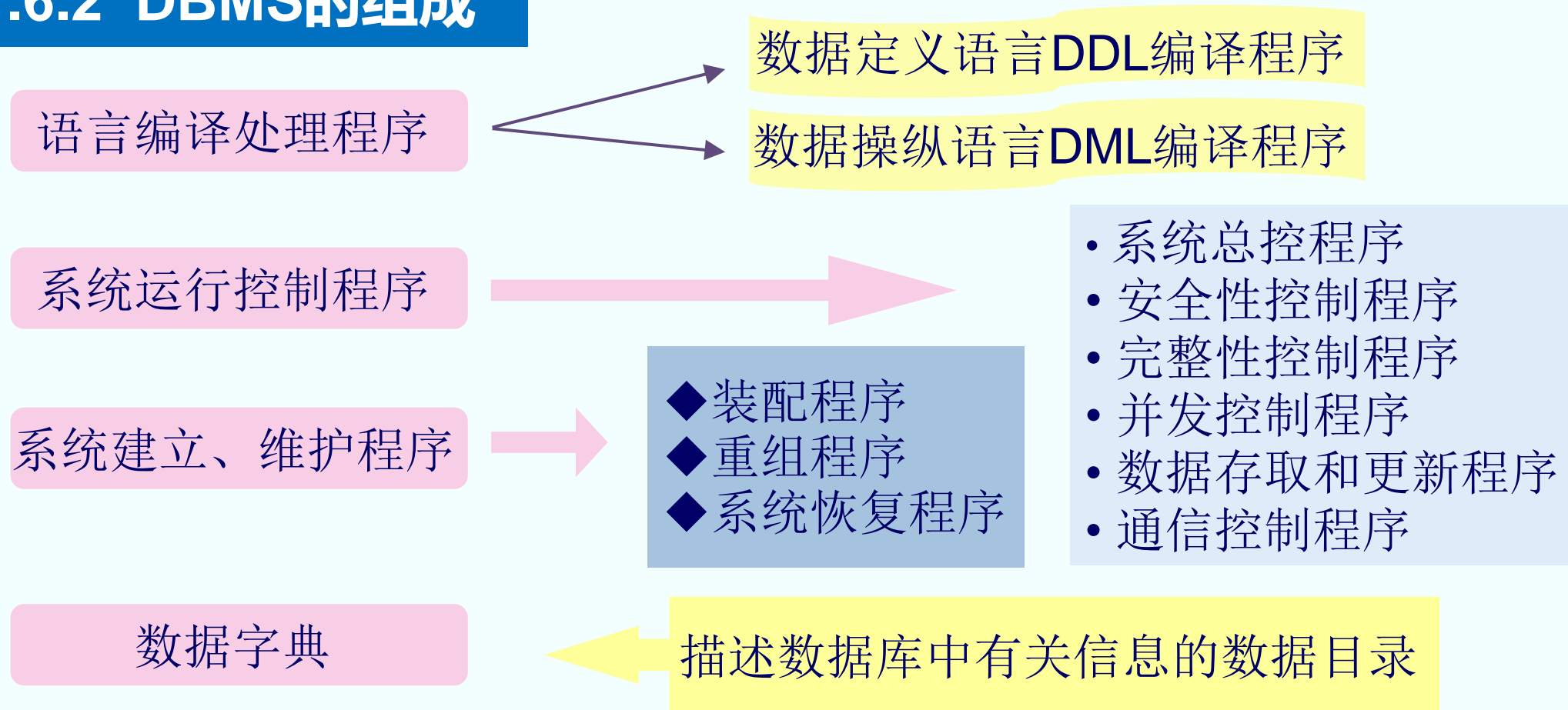
### – 数据库的建立和维护功能

- 数据库的建立包括数据库的初始数据的装入与数据转换等。
- 数据库的维护包括数据库的转储、恢复、重组织与重构造、系统性能监视与分析等。

### – 数据通信接口

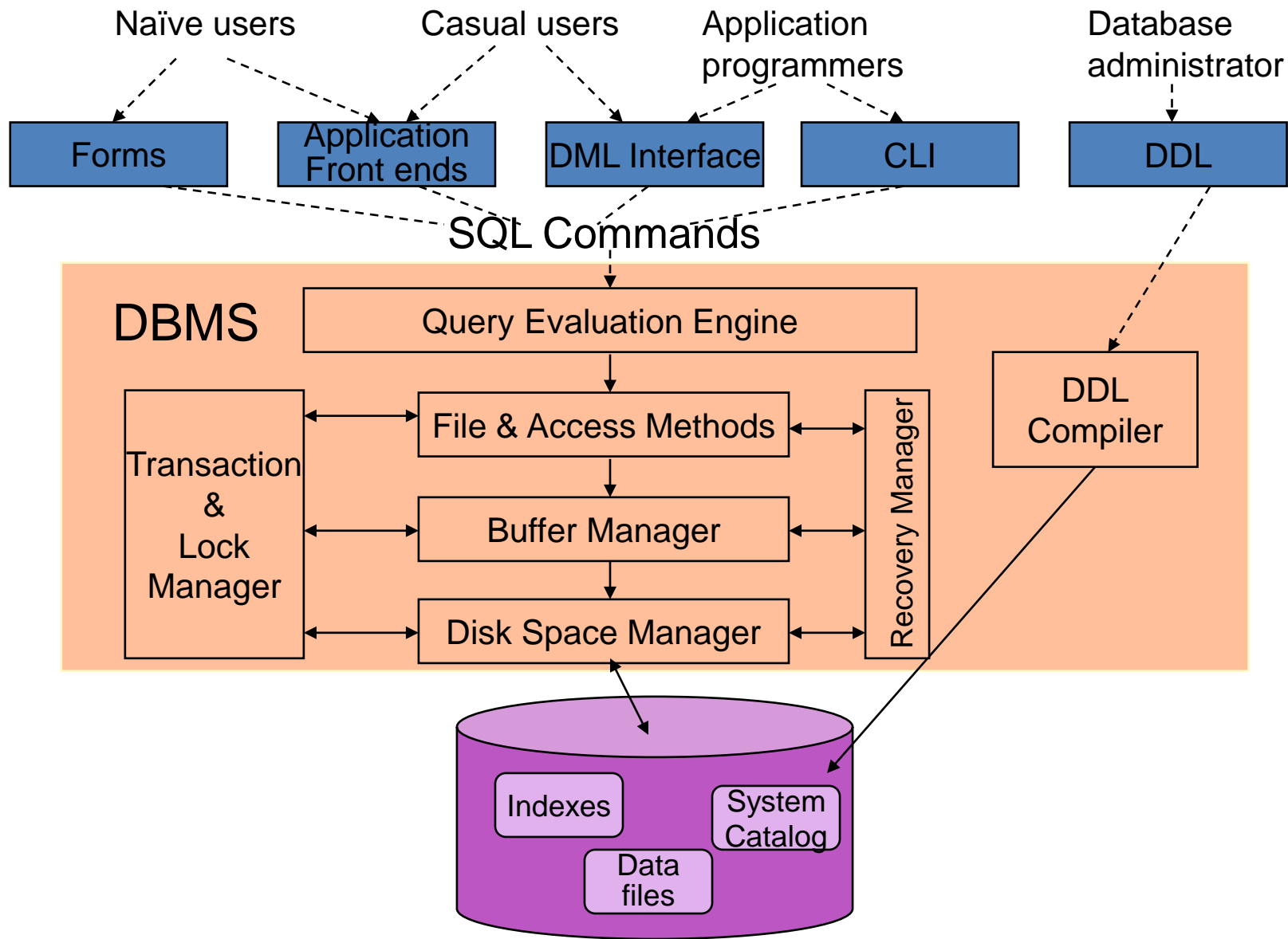
### – 数据组织、存储和管理

### 1.6.2 DBMS的组成



## 1.6 数据库管理系统 (DBMS)

## 第1章



### 1.6.3 DBMS的数据存取的过程

**DBMS对数据的存取通常需要以下几个步骤：**

- (1) 用户使用某种特定的数据操作语言向DBMS发出存取请求；
- (2) DBMS接受请求并将该请求解释转换成机器代码指令；
- (3) DBMS依次检查外模式、外模式 / 模式映象、模式、模式 / 内模式映象及存储结构定义；
- (4) DBMS对存储数据库执行必要的存取操作；
- (5) 从对数据库的存取操作中接受结果；
- (6) 对得到的结果进行必要的处理，如格式转换等；
- (7) 将处理的结果返回给用户。



图1-11 DBMS的工作方式

### 1.7.1 数据模型的概念及分类

- 数据模型是数据库的框架
  - 数据模型是数据库系统的核心和基础
- 数据及其联系 →
- 组织方式
  - 表达方式
  - 存取路径

### 三个世界的划分

- 首先将现实世界的事物及联系抽象成信息世界的信息模型
- 然后再抽象成计算机世界的数据库模型
- 数据加工经历了现实世界、信息世界和计算机世界三个不同的世界，经历了两级抽象和转换



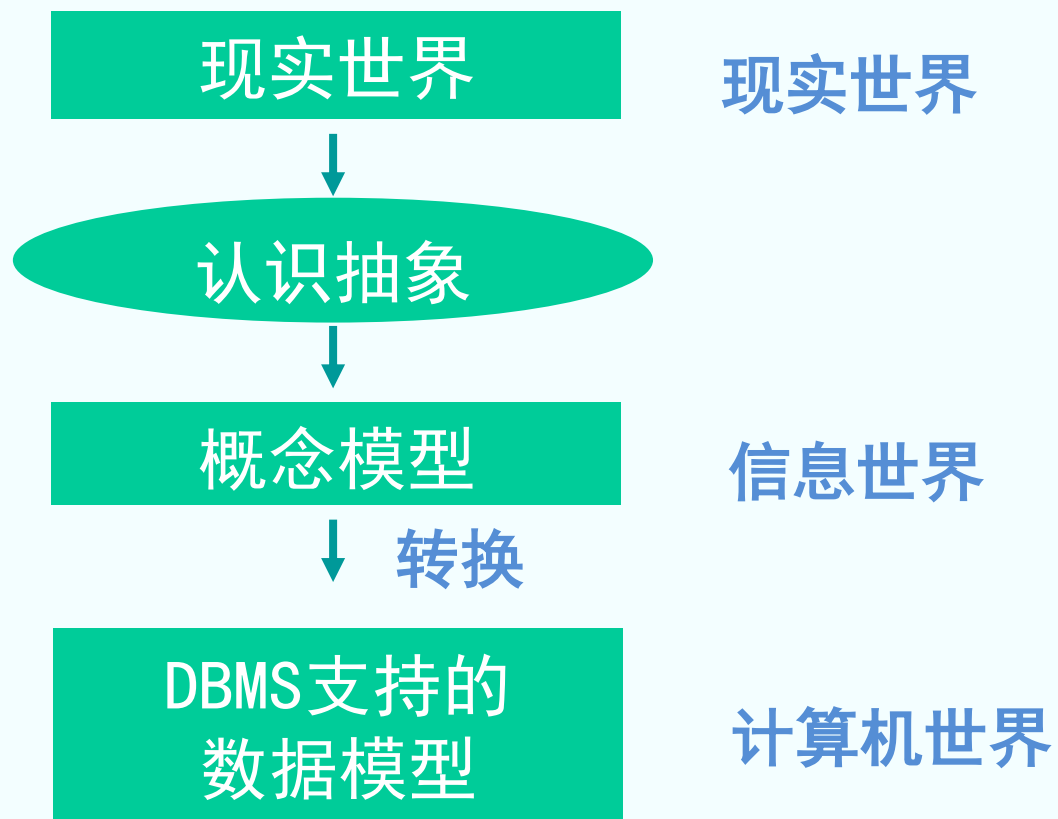


图1-13 数据处理的抽象和转换过程

### 1.7.2 数据模型的组成要素

数据结构

数据对象的类型、内容、性质等  
数据对象间的联系

数据操作

对数据库中的各种数据允许执行的操作的集合，  
包括操作及相应的操作规则，描述了数据库的动态特性

数据的  
完整性约束

数据模型应该反映和规定本数据模型必须遵守的基本的通用的完整性约束条件  
数据模型还应该提供定义完整性约束条件的机制，  
以反映具体应用所涉及的数据必须遵守的特定的语义约束条件

数据结构

层次结构

网状结构

关系结构

数据操作

查询

插入

删除

修改

更新

数据的完整性约束

正确

有效

相容

### 1.8.1 现实世界

现实世界即客观存在的世界  
人们总是选用感兴趣的最能表征一个事物的若干特征来描述该事物

### 1.8.2 信息世界

#### 信息世界及其有关基本概念

**实体**

客观存在并且可以相互区别的“事物”称为实体  
实体可以是具体的人、事、物，也可以是抽象的事件

**属性**

实体所具有的某一特性称为属性

型

——> 属性名

值

——> 具体值

学生（学号，姓名，年龄，性别，系）

实体型

具有相同属性的实体必然具有共同的特征

实体集

同型实体的集合称为实体集

码

能唯一标识一个实体的属性或属性集称为实体的码

域

某一属性的取值范围称为该属性的域

联系

单个实体型内部的联系通常是指组成实体的各属性之间的联系  
实体型之间的联系通常是指不同实体集之间的联系

### 两个实体间的联系

两个实体型之间的联系是指两个不同的实体集间的联系，有如下三种类型：

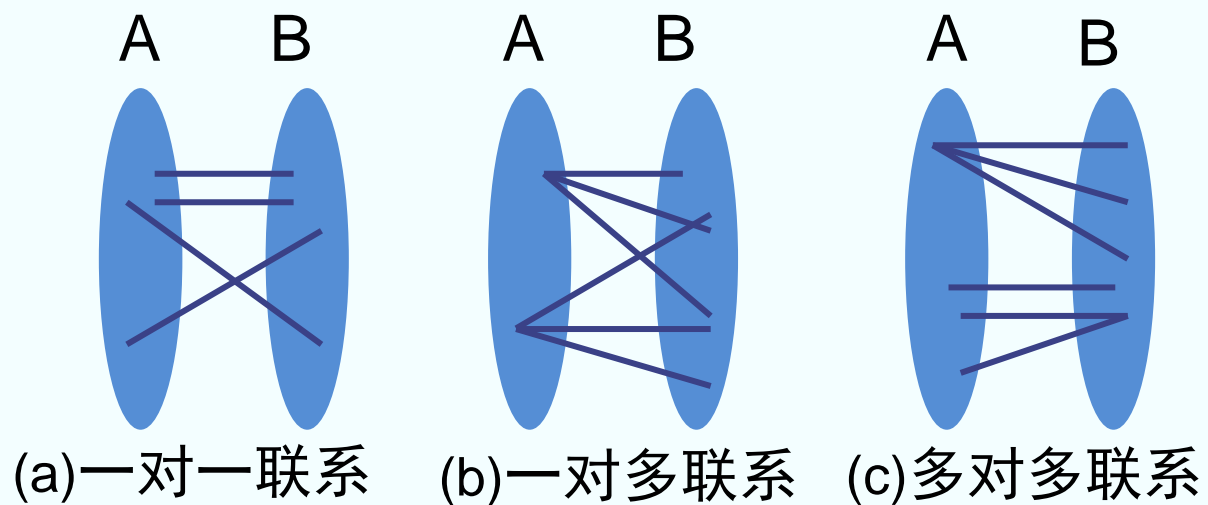
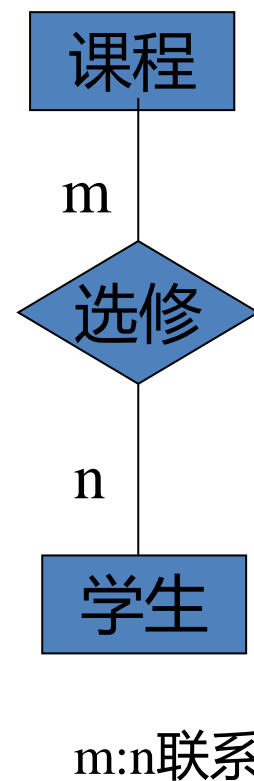
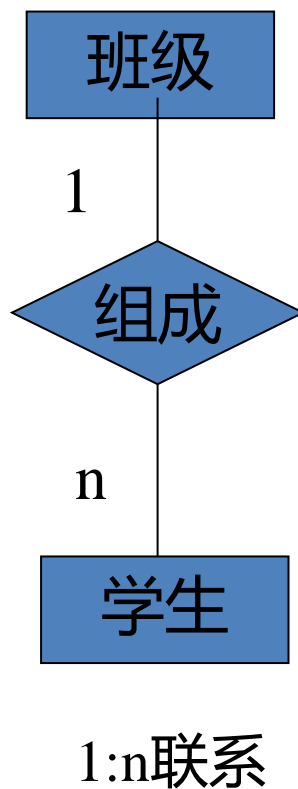
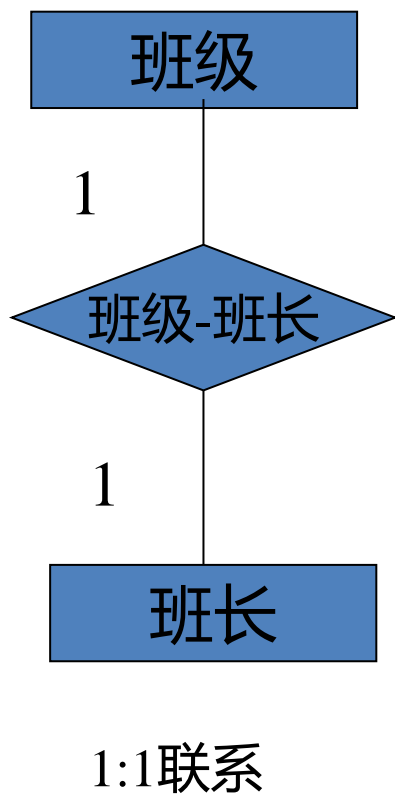


图1-14 两个实体型之间的联系

### 两个实体间的联系（举例）



### 多个实体间的联系（举例）

- 实例1

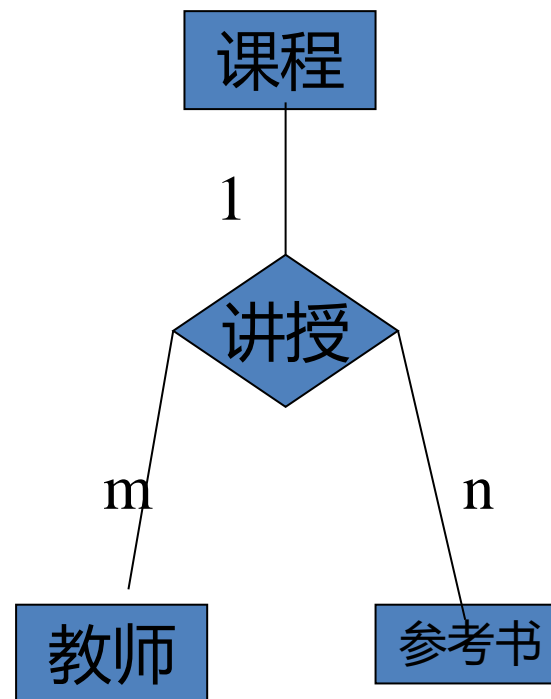
课程、教师与参考书三个实体型

一门课程可以有若干个教师讲授，

使用若干本参考书，

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



两个以上实体型间1:n联系



### 多个实体间的联系（举例）

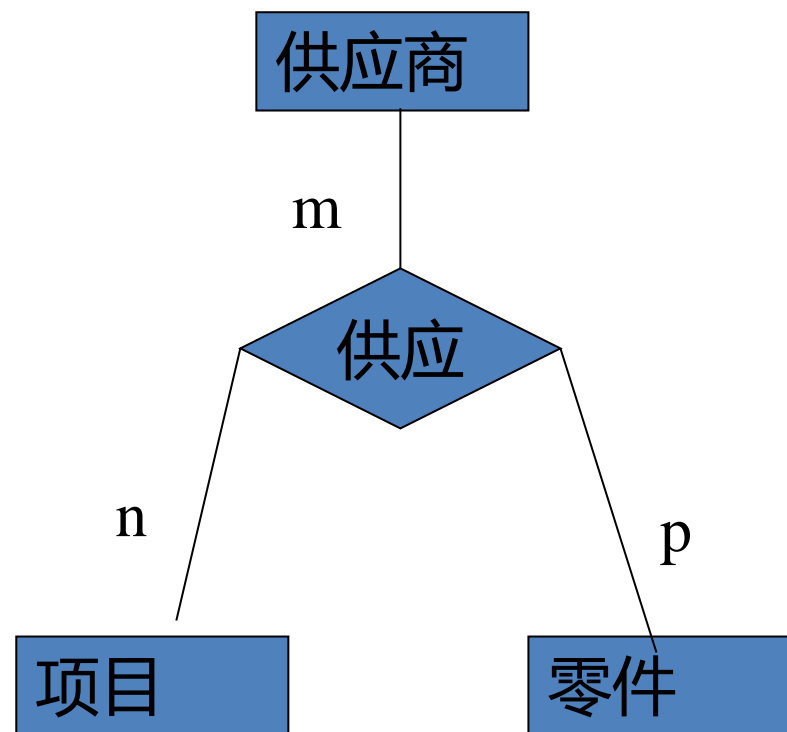
#### ● 实例2

供应商、项目、零件三个实体型

一个供应商可以供给多个项目多种零件

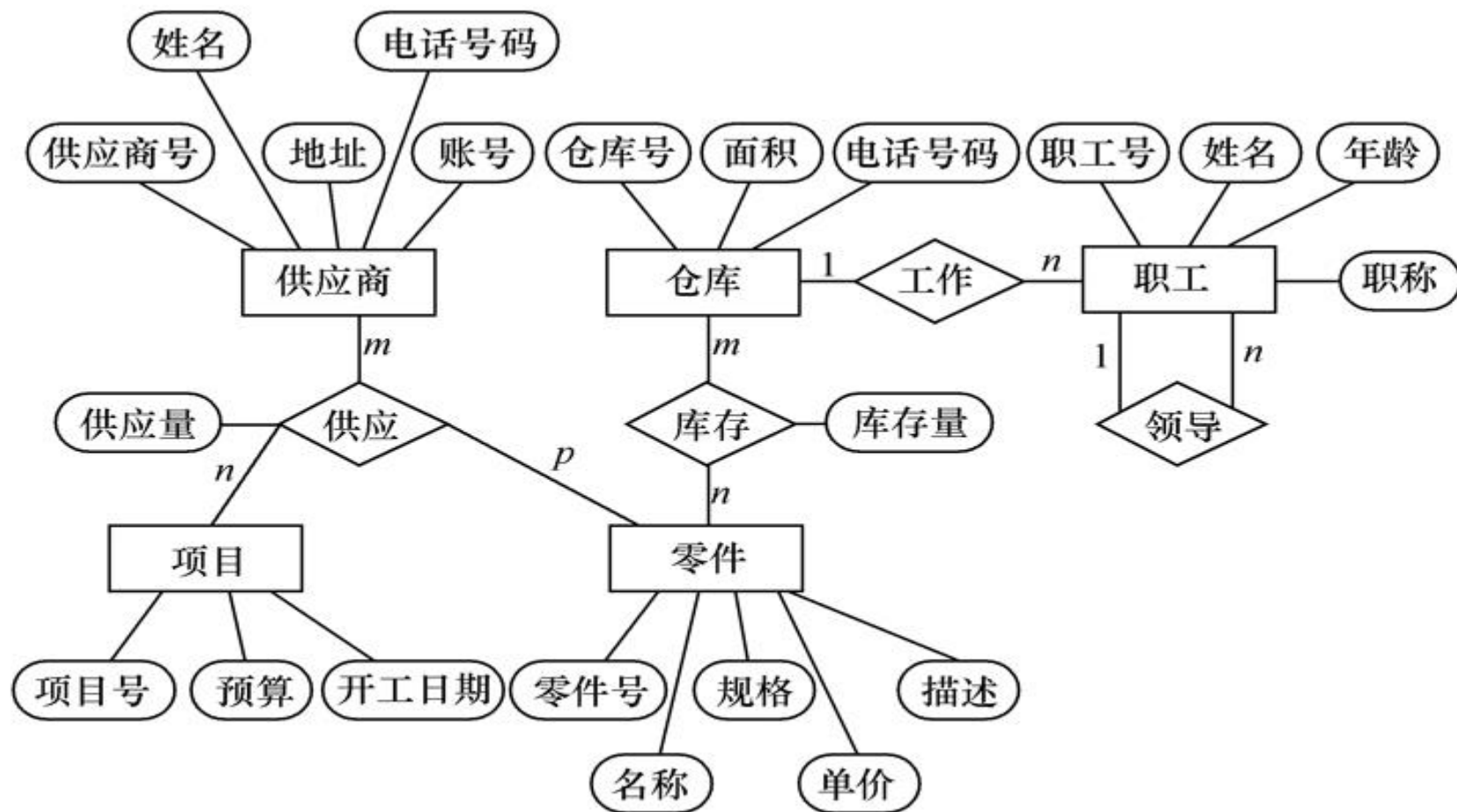
每个项目可以使用多个供应商供应的零件

每种零件可由不同供应商供给



两个以上实体型间m:n联系

## 一个ER模型的实例



(c) 完整的实体-联系图

### 1.8.3 计算机世界

字段  
Field

标记实体属性的命名单位称为字段。

记录  
Record

字段的有序集合称为记录。如，  
一个学生（990001，张立，20，男，计算机）为一个记录。

文件  
File

同一类记录的集合称为文件。如，  
所有学生的记录组成了一个学生文件。

关键字  
Key

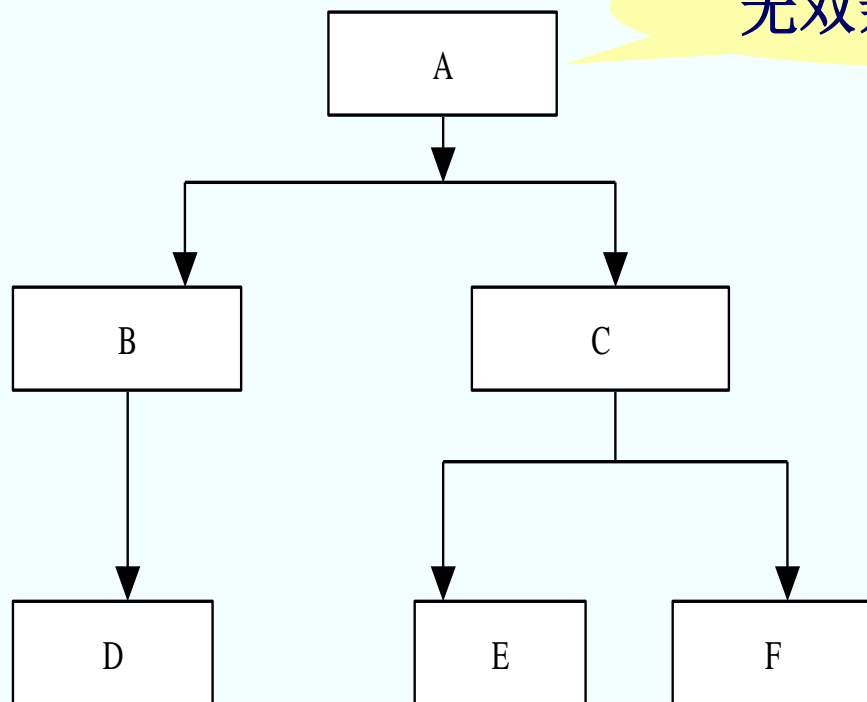
能唯一标识文件中每个记录的字段  
或字段集，称为记录的关键字。



图1-15 三个世界各术语的对应关系

### 1.9.1 层次模型

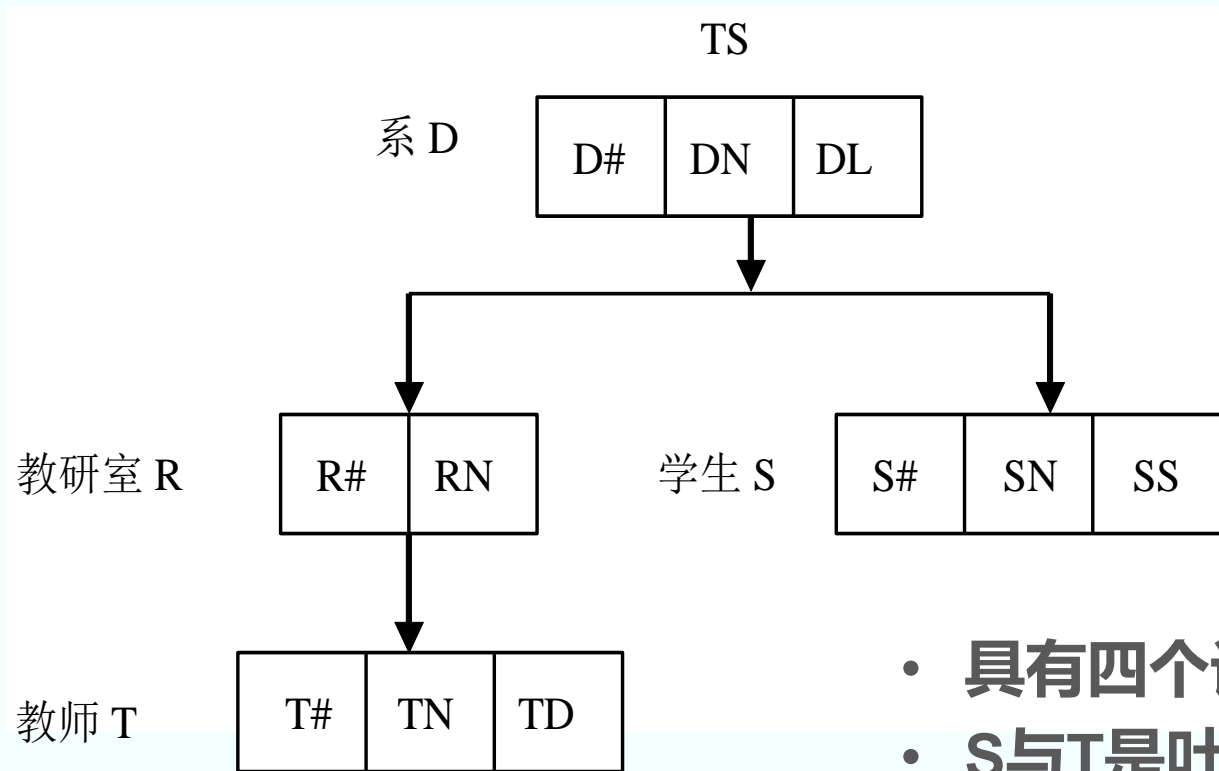
#### 层次模型有向树的示意图



无双亲，根节点

- 每棵树有且仅有一个结点没有双亲，该结点就是根结点
- 根以外的其他结点有且仅有一个双亲结点
- 父子结点之间的联系是一对多（**1:n**）的联系

### TS数据库模型



- 具有四个记录型：D, R, S, T
- S与T是叶结点，它们没有子女结点
- 由D到R, R到T, 由D到S均是一对多的关系

### — 层次模型的数据操纵与数据完整性约束

- 进行插入操作时，如果没有相应的双亲结点值就不能插入子女结点值。
- 进行删除操作时，如果删除双亲结点值，则相应的子女结点值也被同时删除。
- 修改操作时，应修改所有相应的记录，以保证数据的一致性。

### — 层次模型的优缺点

#### 优点

结构简单，层次分明

查询效率高

提供良好的数据完整性支持

#### 缺点

不能直接表示多对多联系

插入和删除数据限制太多

查询子女结点必须通过双亲结点



### 1.9.2 网状模型

#### 网状模型的数据结构

网状模型是采用有向图结构表示记录型与记录型之间联系的数据模型。

- 每个结点表示一个记录型，每个记录型可包含若干个字段，记录型描述的是实体。
- 结点间的带箭头的连线（或有向边）表示记录型间的1:  $n$ 的父子联系。

网状模型的特点：

- 有一个以上的结点没有双亲结点
- 允许结点有多个双亲结点
- 允许两个结点之间有多种联系（复合联系）



### — 网状模型的数据操纵与完整性约束

- 插入数据时，允许插入尚未确定双亲结点值的子女结点值。
- 删除数据时，允许只删除双亲结点值。
- 修改数据时，只需更新指定记录即可。

### — 网状模型的优缺点

#### 优点

可表示实体间的多种复杂联系

具有良好的性能和存储效率

#### 缺点

数据结构复杂

数据定义语言、数据操纵语言复杂

用户需要了解网状模型的实现细节

### 1.9.3 关系模型

小型数据库系统：  
Foxpro、Access

大型数据库系统：  
Oracle、SQL Server、  
Informix、Sybase

#### — 关系模型的数据结构及有关概念

SNO 学号	SN 姓名	SEX 性别	AGE 年龄	DEPT 系别
S1	赵亦	女	17	计算机
S2	钱尔	男	18	信息
S3	孙珊	女	20	信息
S4	李思	男	21	自动化
S5	周武	男	19	计算机
S6	吴丽	女	20	自动化

属性

关系 ↔ 二维表

元组

域：  
属性的取值范围，  
(男，女)

分量

每一行对应的列的属性值，即为元组中的一个属性值。

关键字

可惟一标识一个元组的属性或属性集，也称为关系键或主码，如S表中学号可以惟一确定一个学生，为学生关系的主码。

关系模式

关系模式是对关系的描述，是关系模型的“型”。  
一般表示为：关系名（属性1，属性2，……属性n）  
如：学生（学号，姓名，性别，年龄，系别）

关系实例

关系实例是关系模式的“值”，是关系的数据，  
相当于二维表中的数据。

### — 关系模型的数据操纵与完整性约束

- 关系模型中的数据操作是集合操作，操作对象和操作结果都是关系，即若干元组的集合。
- 关系模型把对数据的存取路径隐蔽起来，用户只要指出“干什么”，而不必详细说明“怎么干”，从而大大地提高了数据的独立性，提高了用户操作效率。

### — 关系模型的优缺点

#### 优点

有严格的数学理论根据

用关系描述实体间的联系

具有更高的数据独立性、  
更好的安全保密性

#### 缺点

查询效率不如非关系模型

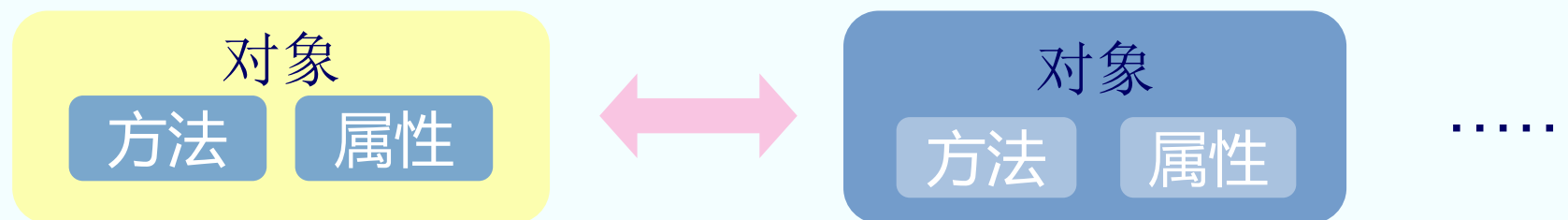


### 1.9.4 面向对象模型

一个学生  
一门课程  
一次考试记录

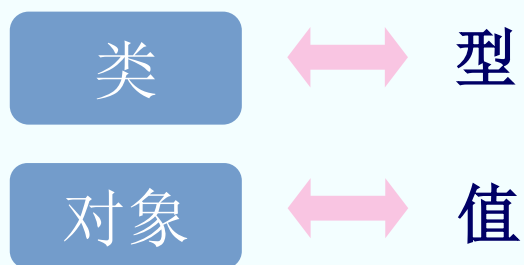
#### — 对象和对象标识

- **对象**是现实世界中实体的模型化。
- **对象标识**独立于对象的内容和存储位置，是一种逻辑标识符，通常由系统产生，它在整个系统范围内是惟一的。
- 两个对象即使内部状态值和方法都相同，如标识符不同，仍认为是两个相等而不同的对象。
- 每个对象都包含一组属性和一组方法。



### – 类 (Class) 和继承 (Inheritance)

- 具有同样属性和方法集的所有对象构成了一个对象类，一个对象是某一类的实例。



- 类的属性域可以是基本数据类型（如整型、实型、字符型等），也可以是类。
- 类的表示具有层次性和继承性。

- 面向对象模型的优缺点

### 优点

能完整地描述现实世界的数据结构

具有丰富的表达能力

### 缺点

模型相对比较复杂

涉及的知识比较多

因此，面向对象数据库尚未达到关系数据库的普及程度！

### 1.10.1 分布式数据库

#### 集中式系统和分布式系统

**集中式数据库**就是集中在一个中心场地的电子计算机上，以统一处理方式所支持的数据库。

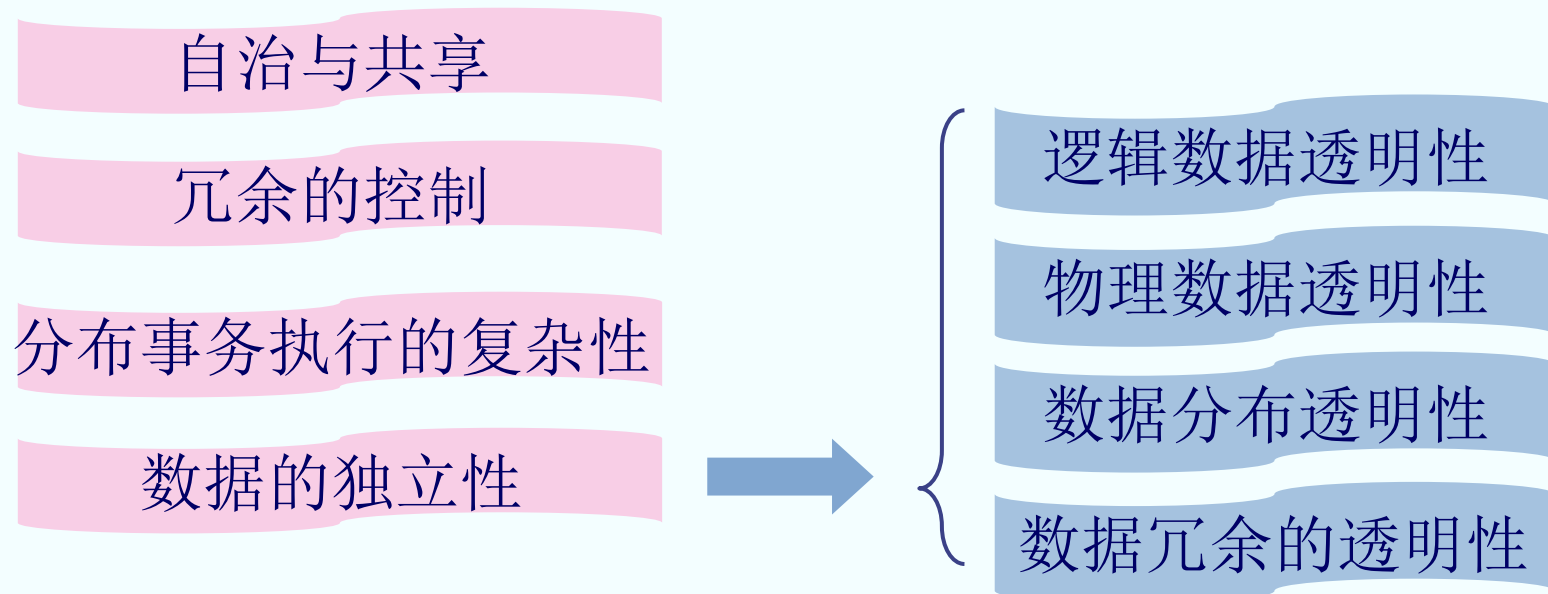
- » 集中控制处理效率高，可靠性好；
- » 数据冗余少，数据独立性高；
- » 易于支持复杂的物理结构去获得对数据的有效访问。



### — 分布式数据库的定义

- 分布式数据库是一组结构化的数据集合，它们在**逻辑上**属于同一系统，而在**物理上**分布在计算机网络的不同结点上。

### — 分布式数据库的特点



### 1.10.2 数据仓库与数据挖掘技术

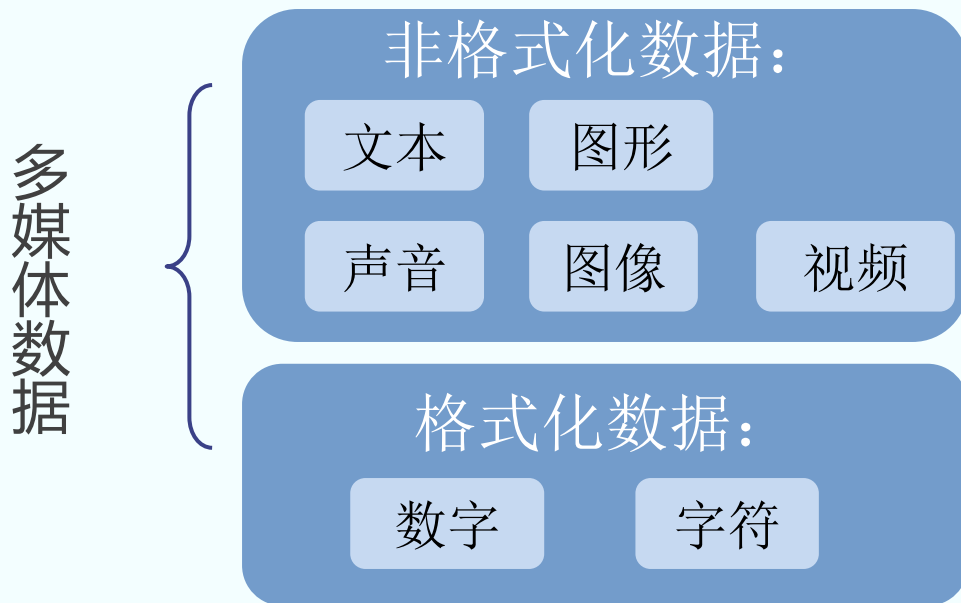
#### 数据仓库的定义

数据仓库就是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合，通常用于辅助决策支持。

#### 数据仓库的体系结构

对多个异构数据源的有效集成，集成后按照主题进行重组，包含历史数据。存放在数据仓库中的数据通常不再修改，用于做进一步的分析型数据处理。

### 1.10.3 多媒体数据库



多媒体数据库系统能够有效实现对结构化和非结构化的多媒体数据进行存储、管理和操纵等功能。

- **多媒体数据库目前有三种结构**

- 第一种结构：由单独一个多媒体数据库管理系统来管理不同媒体的数据库以及对象空间。
- 第二种结构：主辅DBMS体系结构。
- 第三种结构：协作DBMS体系结构。

### 1.10.4 DBMS的改进

数据库管理系统软件的研制

提高系统的性能和提高用户的生产率。

数据库设计

在 DBMS 的支持下，按照应用的要求，为某一部门或组织设计一个结构合理、使用方便、效率较高的数据库及其应用系统。

数据库理论

关系规范化理论、关系数据理论的研究。

- 中国国产数据库的历史  
<https://zhuanlan.zhihu.com/p/573116957>

- 信息、数据、数据处理与数据管理的基本概念；
- 数据管理技术发展的三个阶段及各自的优缺点；
- 数据库系统主要包括数据库、数据库用户、计算机硬件系统和计算机软件系统等几部分；
- 数据库系统内部的体系结构：三级模式结构；
- 数据库系统的三级抽象和二级映象保证了数据库系统的逻辑独立性和物理独立性；
- 层次模型、网状模型、关系模型和面向对象模型。