

# 第一章 数据库系统概述

# 第一章 数据库系统概述

## **1.1 基本概念**

## **1.2 数据库系统的发展及趋势**

## **1.3 数据库系统的基本特点**

## **1.4 数据库内部结构体系**

# 1.1 基本概念

- 数据
- 数据库
- 数据库管理系统
- 数据库管理员
- 数据库系统
- 数据库应用系统

# 1.1 基本概念

## □ 数据 (data)

### ➤ 信息

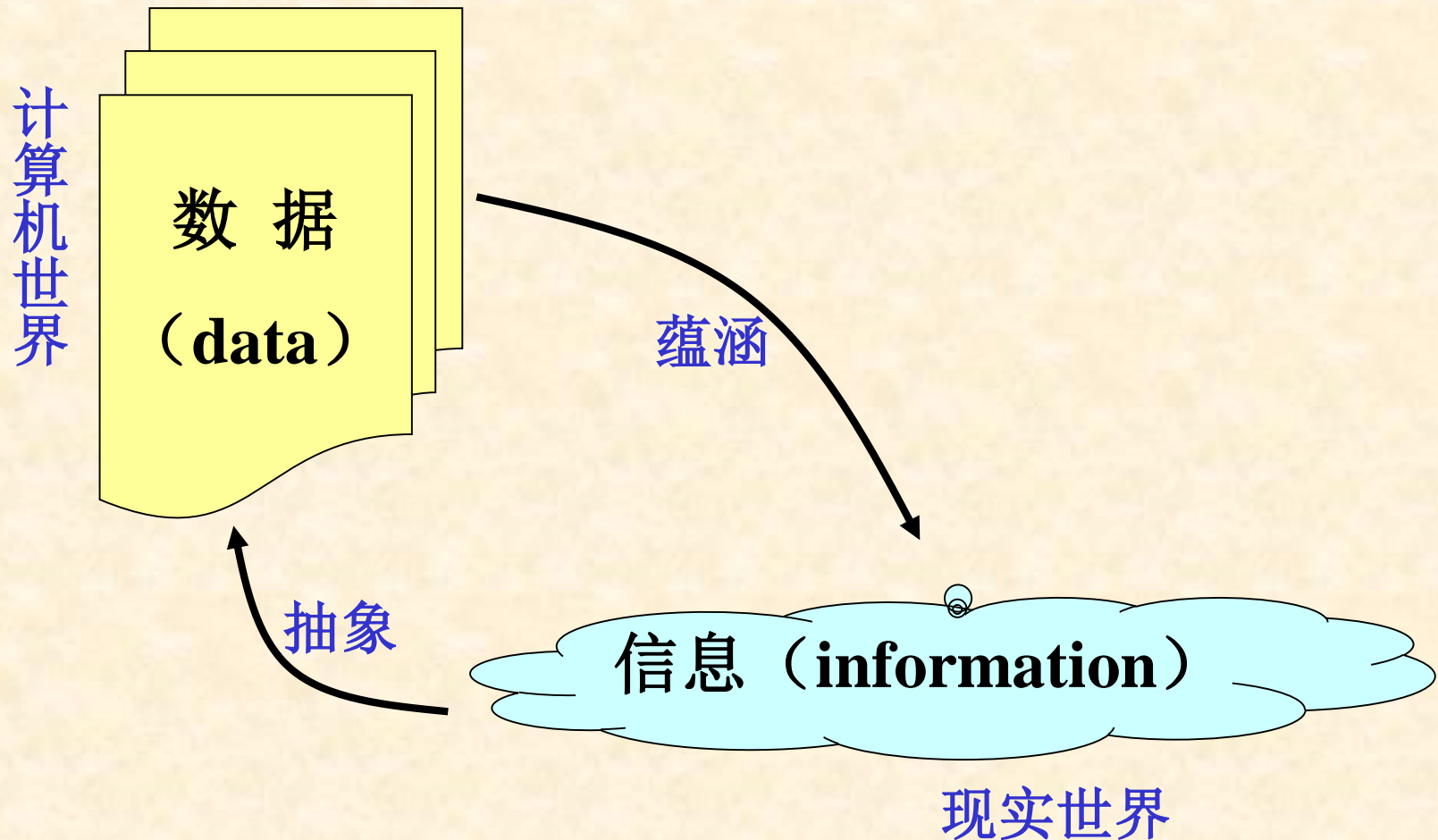
- 用于反映现实世界中事物的物理状态，向人们提供一些已知的、客观存在的事实和知识
  - 例：人、时间、空间的信息

### ➤ 数据

- 是指具有一定的语义含义，并且可以被记录下来的已知事实
- 在计算机中，数据被表示为具有一定格式（或结构）的符号串，它是计算机软件中程序加工的原料与结果，属于软件范畴
  - 例：一个人的姓名、电话号码、地址、照片等

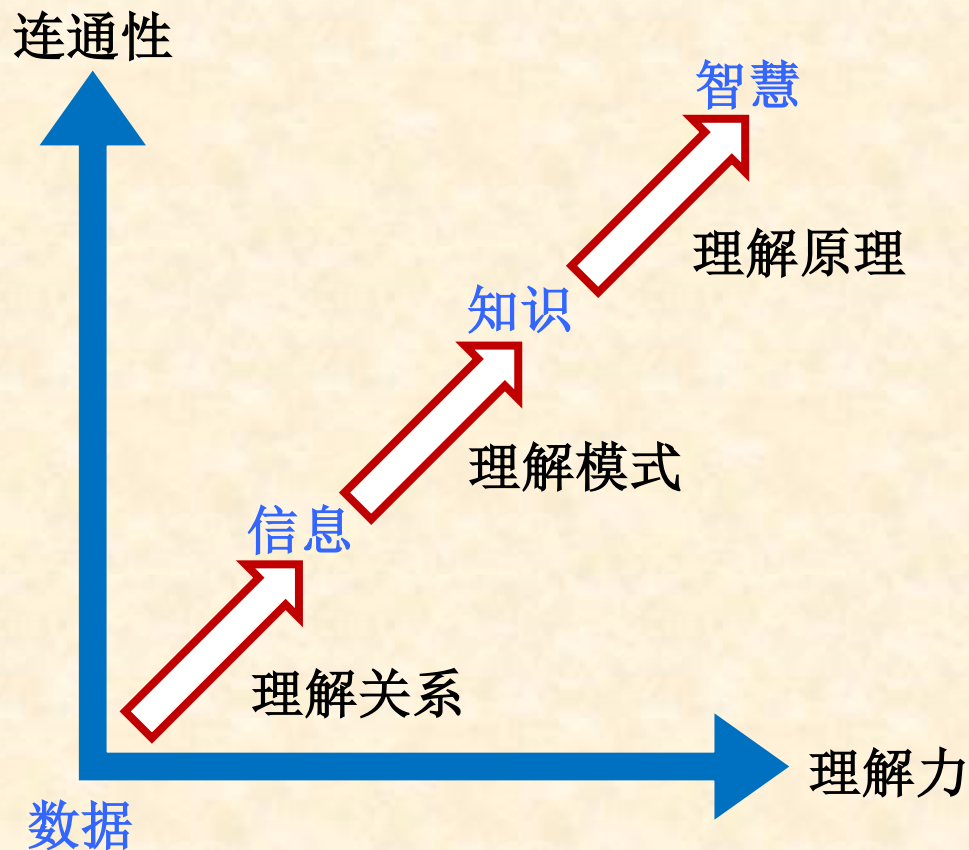
# 1.1 基本概念

□ 数据是信息的载体，信息则是数据的内涵



# 数据→信息→知识→智慧

- ❑ 数据：原始素材，对客观事物的形式化表示
- ❑ 信息：加工处理后有逻辑的数据
- ❑ 知识：从相关信息中过滤、提炼及加工而得到的有价值信息
- ❑ 智慧：关注未来，具有预测能力



# 1.1 基本概念

## □ 数据的特性

常用数据、多媒体数据、抽象数据

1) 数据表现的多样性

2) 数据的可构造性

型 type

值 value

数据类型 type

数据结构 structure

数据模式 schema

3) 数据的挥发性/持久性 (transient / persistent)

4) 数据的私有性/共享性 (private / share)

5) 数据‘量’的表示：少量/大量/海量

数据的‘量’是衡量与区别数据的重要标志，  
数据的‘量变’可能会引起数据的‘质变’

# 1.1 基本概念

## □ 数据特性的变化

### 1) 数据的量

少量 → 大量 → 海量 (big data)

### 2) 数据的结构

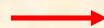
简单 → 复杂

### 3) 数据的服务范围

私有 → 共享

### 4) 数据在软件中的地位

附属地位  
(以程序为主体)



主导地位  
(以数据为中心)



# 1.1 基本概念

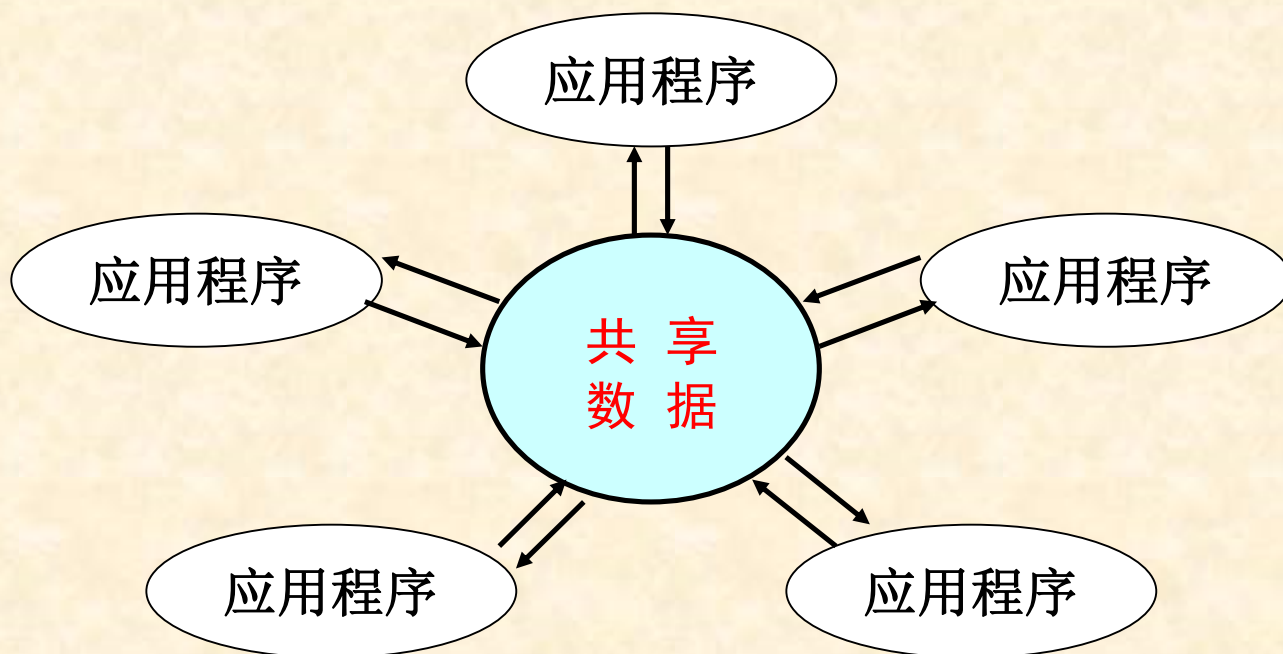


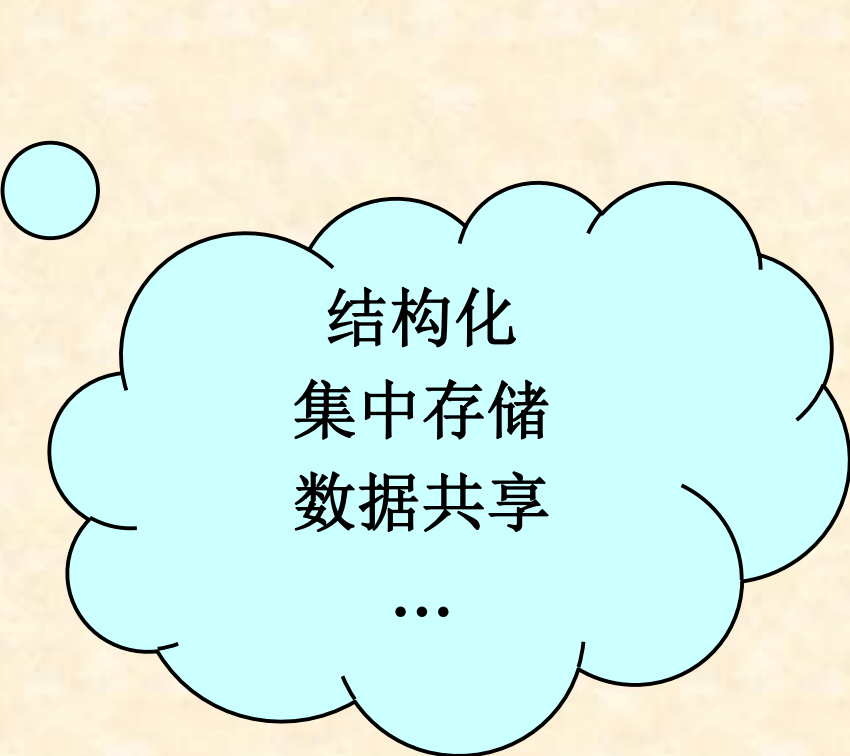
图1.3 以数据为主体的软件系统示意图

- ❑ 传统的数据管理方式已经无法适应上述数据特性的‘变化’情况，需要有新的数据管理技术，以便对数据作集中、统一的管理并使其对应用（程序）共享

# 1.1 基本概念

## □ 数据库 (Database, 简称DB)

- 是数据集合，具有统一的结构形式并存放于统一的存储介质内，它由多种应用数据集成，并可被应用所共享



结构化  
集中存储  
数据共享

...

# 一个关系数据库的例子

## students

sid	lname	fname	class	telephone
1	Jones	Allan	2	555-1234
2	Smith	John	3	555-4321
3	Brown	Harry	2	555-1122
5	White	Edward	3	555-3344

## courses

cno	cname	croom	time
101	French I	2-104	MW2
102	French II	2-113	MW3
105	Algebra	3-105	MW2
108	Calculus	2-113	MW4

## enrollment

sid	cno	major
1	101	No
1	108	Yes
2	105	No
3	101	Yes
3	108	No
5	102	No
5	105	No

# 1.1 基本概念

## □ 数据库管理系统 (Database Management System, 简称DBMS)

- 是一种管理数据库的系统软件
- DBMS是在文件管理系统的基础上发展起来的, 它区别于其它计算机软件系统的特点在于:
  - ① 能对持久性数据进行管理
  - ② 能对大量数据进行有效存取
  - ③ 可为众多使用者提供同一数据 (即数据共享)
- DBMS的作用
  - ① 是数据库的应用程序与数据库的接口
  - ② 在保证数据安全、可靠的同时, 提高数据库应用时的简明性和方便性

# 数据库管理系统

## □ DBMS的功能

- 数据组织(Definition)
- 数据操纵(Manipulation)
- 数据维护
- 数据控制及保护
- 数据交换
- 数据服务
- 数据字典

- 数据模式定义（为数据库构造数据框架）
- 数据存取的物理操作（为数据模式构造有效的物理存取方法与手段）

- 提供数据查询、插入、修改及删除的功能
- 还具有简单算术运算及统计等能力
- 此外，它还可以与某些过程性语言结合，进行过程性操作

- 数据的完整性、安全性定义与检查

- 数据库的并发控制
- 数据库的故障恢复

- 内置函数
- 拷贝、转储、重组、性能监测、分析...

# 数据库管理系统

## □ 数据子语言 (data sub\_language)

### ➤ 数据定义语言

- Data Definition Language, 简称**DDL**, 负责数据的模式定义与数据的物理存取构造

### ➤ 数据操纵语言

- Data Manipulation Language, 简称**DML**, 负责数据的操纵, 包括查询及增、删、改等操作

### ➤ 数据控制语言

- Data Control Language, 简称**DCL**, 负责数据的完整性、安全性的定义与检查以及并发控制、故障恢复等功能

SQL 语言



# 数据库管理系统

## □ 数据子语言的使用方式

### ➤ 交互式命令语言

- 能在终端上即席操作，又被称为自含型或自主型语言

### ➤ 宿主型语言

- 需要嵌入到某种宿主语言中

一种高级程序设计语言，  
比如FORTRAN，COBOL，  
C/C++，JAVA，Python  
等

# 1.1 基本概念

## □ 数据库管理员（Database Administrator, 简称DBA）

➤ 对数据库进行规划、设计、维护、监视的专职人员

### ➤ DBA的主要工作

- 数据库设计，建立与调整
- 数据库维护
- 改善系统性能，提高系统效率



# 1.1 基本概念

## □ 数据库系统（Database System，简称DBS）

- 是一个以对海量的、具有复杂数据结构的、可以持久保存的、可供多用户共享的数据进行统一管理为目标的计算机系统

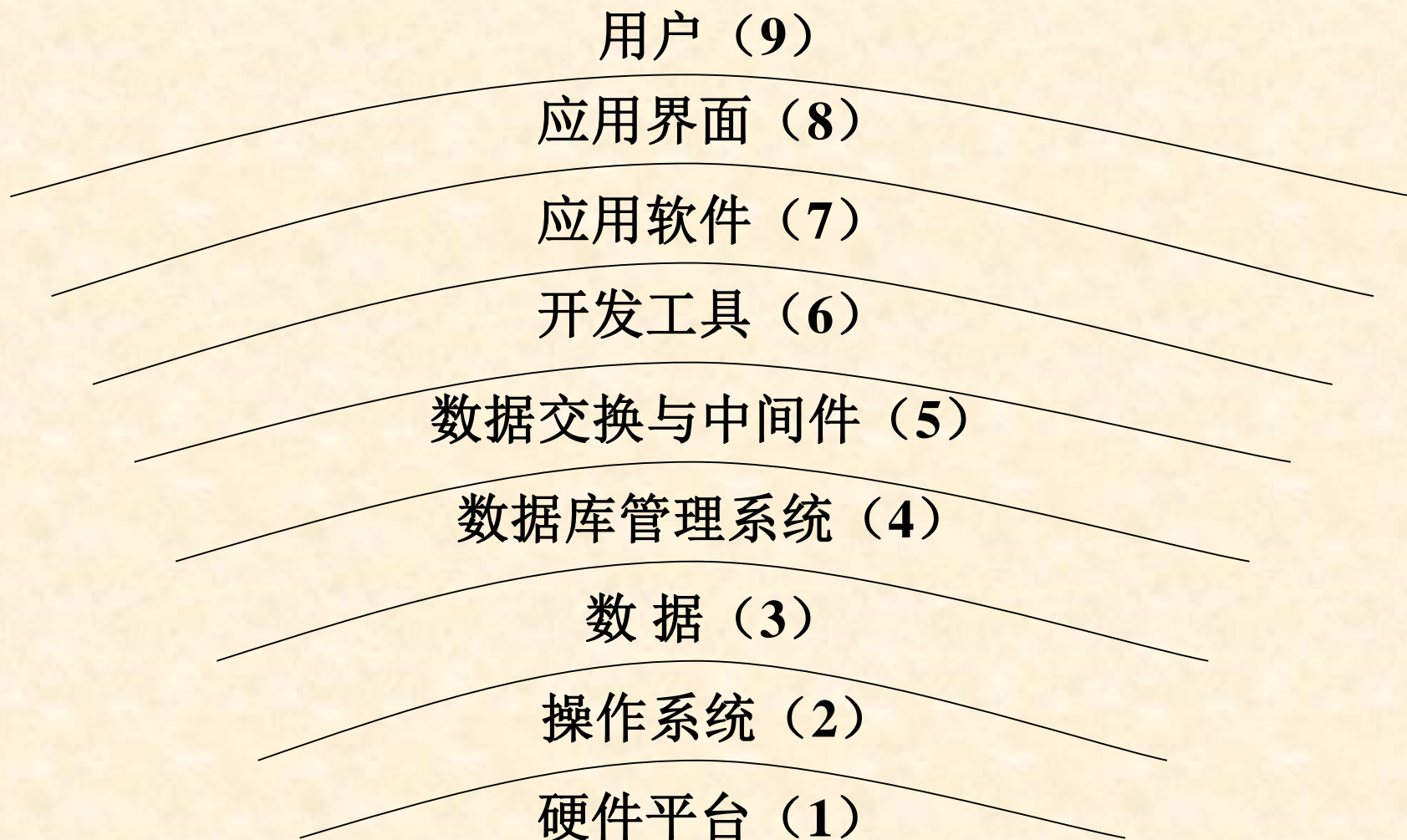
### — DBS的组成部分

- 数据库
- 数据库管理系统
- 数据库管理员
- 软件平台
  - 操作系统，语言，数据库应用开发工具，通用的数据库访问接口
- 硬件平台

# 1.1 基本概念

## □ 数据库应用系统（Database Application System, 简称DBAS）

- 利用数据库系统作应用开发所构成的集成化的独立运行系统
- DBAS的组成
  - 数据库系统
  - 应用软件
  - 应用界面
  - 用户



**图1.4 数据库应用系统层次结构示意图**

# 1.1 基本概念

## □ 数据库用户 (Database Users)

### ① 最终用户 (End users)

- 终端查询用户 (Casual users)

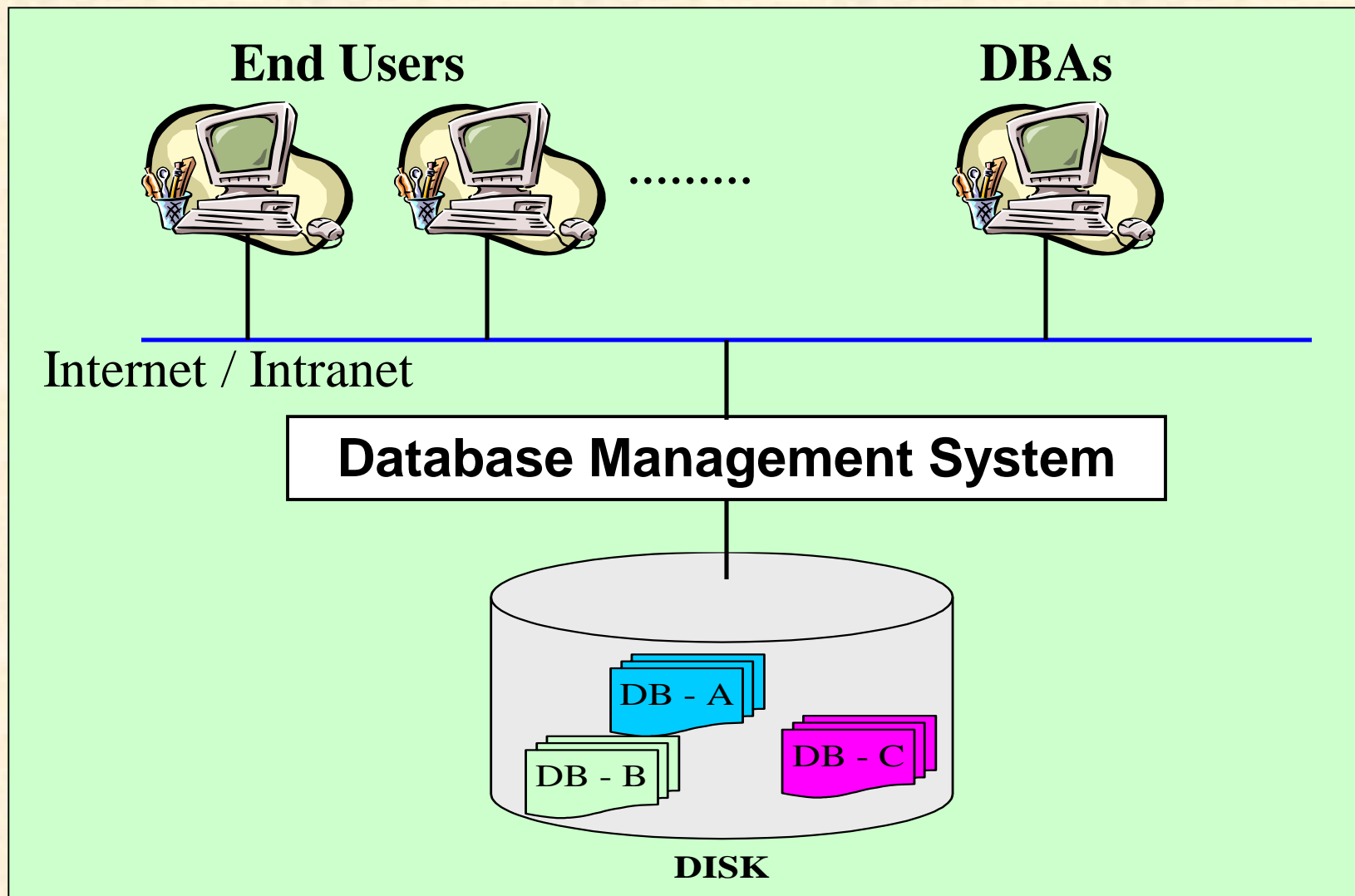
- 应用程序的使用者 (Naive users)

### ② 应用程序开发人员 (Application programmers)

### ③ 数据库管理员

这三类用户需要  
掌握与其工作相  
关的数据库技术

# 1.1 基本概念



用户与数据库系统之间的网络结构示意图

# 第一章 数据库系统概述

## 1.1 基本概念

## 1.2 数据库系统的发展及趋势

## 1.3 数据库系统的基本特点

## 1.4 数据库内部结构体系

# 1.2 数据库系统的发展历史

## □ 数据管理技术的三个发展阶段

- 人工管理（50年代以前）
- 文件管理系统（60 — 70年代）
- 数据库系统管理（70年代至今）



# 1.2 数据库系统的发展历史

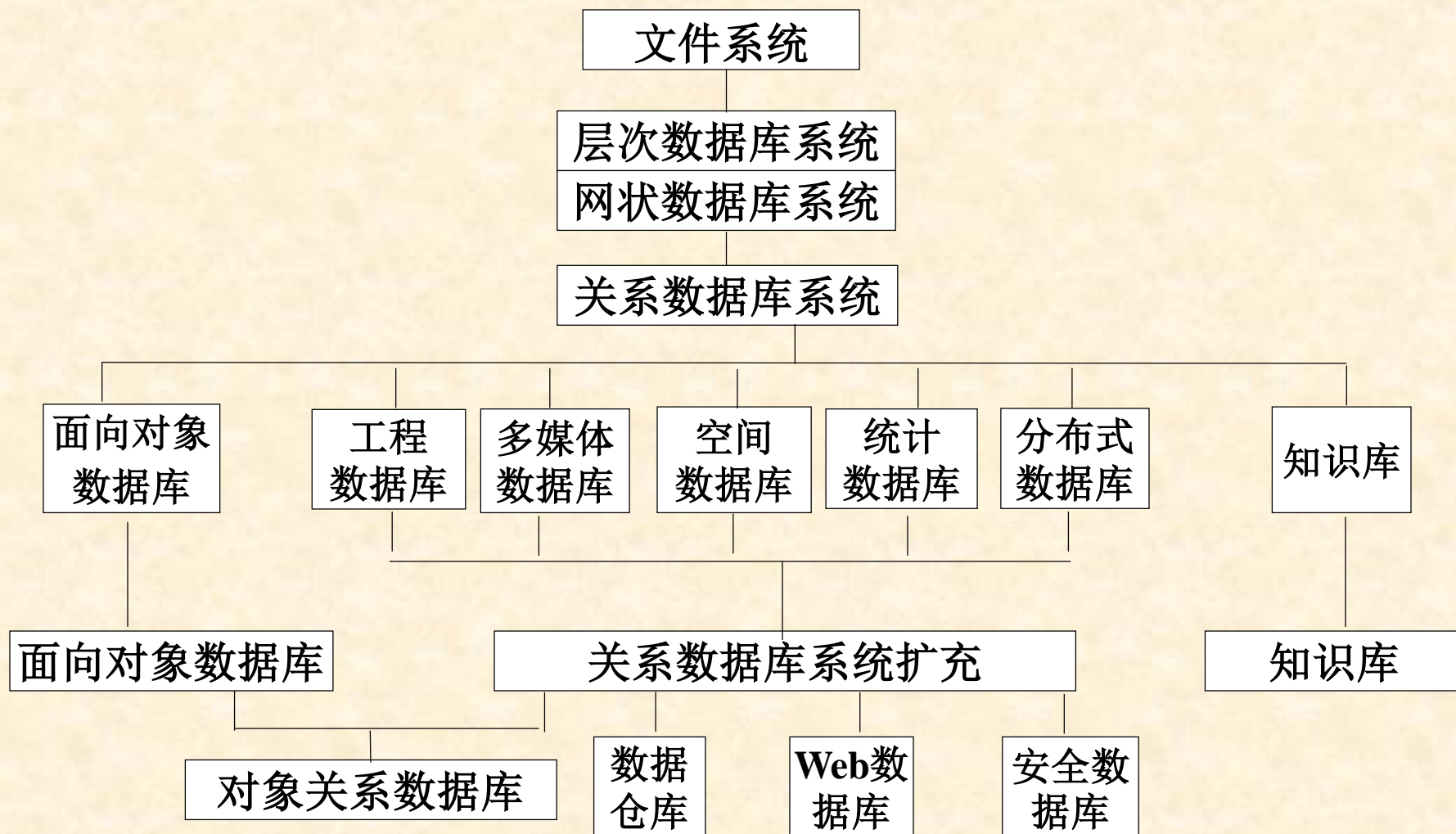


图1.5 数据库系统发展简图



# 1.2 数据库系统的发展历史

## □ 文件系统阶段

- 文件系统是数据库系统发展的初级阶段，出现于20世纪50年代末期。目前一般将其看成仅是数据库系统的雏形，而不是真正的数据库系统

### — 优点

- 利用文件系统来参与数据管理，向用户提供简单的数据管理和共享能力

### — 缺点

- 无法提供完整统一的数据管理功能和较强的数据共享能力
- 文件系统只是操作系统的一个附属软件，因此也不利于数据库系统在不同平台之间的移植

## 1.2 数据库系统的发展历史

### □ 层次数据库与网状数据库

- 层次数据库与网状数据库于20世纪60年代末开始发展。它们为统一管理与共享数据提供了有力支撑，  
是真正的数据库系统

#### — 优点

- 数据库系统为用户提供了统一的数据管理工具，向用户提供了数据共享能力

#### — 缺点

- 它们脱胎于文件系统，受文件的物理影响较大，对数据库的使用带来不便
- 数据模式构造烦琐，不利于推广使用

# 1.2 数据库系统的发展历史

## □ 关系数据库

- 关系数据库于20世纪70年代开始问世，80年代初进入潮流

### — 优点

- 结构简单、使用方便、逻辑性强、物理性少

### — 缺点

- 模型的描述能力不足
- 数据操纵功能有待进一步的扩展

专用数据库系统（关系数据库的扩充）

缺点：专用性有余而通用性不足

## 1.2 数据库系统的发展历史

### □ 通用数据库系统

➤ 20世纪90年代的研究与发展重点集中于具有通用性的三类数据库系统

— 面向对象数据库系统

— 知识库系统

— 关系数据库系统扩充

## 1.2 数据库系统的发展历史

### □ 新一代数据库系统

- 进入21世纪，数据库技术的发展集中于对传统**关系数据库系统**的进一步扩充与改造上
  - 对象关系数据库系统
  - 数据仓库(Data Warehouse)
  - Web数据库
  - 安全数据库
  - 嵌入式数据库，移动数据库，实时数据库，网格数据库，传感器网络数据库， .....

# 第一章 数据库系统概述

## 1.1 基本概念

## 1.2 数据库系统的发展及趋势

## 1.3 数据库系统的基本特点

## 1.4 数据库内部结构体系



## 1.3 数据库系统的基本特点

- ❑ 数据的集成性
- ❑ 数据的高共享性与低冗余性
- ❑ 数据独立性
- ❑ 数据的统一管理与控制

# 1.3 数据库系统的基本特点

## □ 集成性 (integration)

- 集多种应用数据于一体
- 集成性的表现
  - 采用统一的数据结构
  - 建立一个全局统一的数据模式
  - 根据每个应用的数据需要构造局部模式



## 1.3 数据库系统的基本特点

### □ 高共享性与低冗余性

#### ➤ 数据共享

- 可供多个应用程序使用，并可用于不同的目的
- 可以在已有的数据库系统上开发新的应用程序
- 可向外界提供信息服务功能

#### ➤ 数据冗余

- 同一个数据在不同的地方出现了重复存储

#### ➤ 数据库系统所具有的高共享性和低冗余性不仅可以减少不必要的存储空间，更为重要的是可以避免数据的不一致性

## 1.3 数据库系统的基本特点

### □ 高共享性与低冗余性（cont.）

#### ➤ 数据的一致性

- 在系统中，同一数据的不同出现应保持相同的值

#### ➤ 数据的不一致性

- 同一数据在系统的不同拷贝处有不同的值
- 数据的不一致性会造成系统的混乱，因此减少冗余性、避免数据的不同出现是保证系统一致性的基础

# 1.3 数据库系统的基本特点

## □ 数据独立性

➤ 指数据库中的数据与使用这些数据的应用程序之间的互不依赖性，即数据或数据结构的改变不会导致对使用这些数据的应用程序的修改，反之亦然

— 物理独立性

— 逻辑独立性

## 1.3 数据库系统的基本特点

### □ 物理独立性

- 数据的物理结构（包括存储结构、存取方式等）的改变，不影响数据库的逻辑结构，从而不致引起应用程序的变化

### □ 逻辑独立性

- 数据库总体逻辑结构的改变，如修改数据模式、增加新的数据类型、改变数据间联系等，不需要相应修改应用程序

# 数据独立性示例

## students

sid	lname	fname	class	telephone
1	Jones	Allan	2	555-1234
2	Smith	John	3	555-4321
3	Brown	Harry	2	555-1122
5	White	Edward	3	555-3344

## courses

cno	cname	croom	time
101	French I	2-104	MW2
102	French II	2-113	MW3
105	Algebra	3-105	MW2
108	Calculus	2-113	MW4

## enrollment

sid	cno	major
1	101	No
1	108	Yes
2	105	No
3	101	Yes
3	108	No
5	102	No
5	105	No

# 数据独立性示例 (cont.)

## Students-Courses

<b>sid</b>	<b>Iname</b>	<b>fname</b>	<b>cname</b>	<b>major</b>
<b>1</b>	<b>Jones</b>	<b>Allan</b>	<b>French I</b>	<b>No</b>
<b>1</b>	<b>Jones</b>	<b>Allan</b>	<b>Calculus</b>	<b>Yes</b>
<b>2</b>	<b>Smith</b>	<b>John</b>	<b>Algebra</b>	<b>No</b>
<b>3</b>	<b>Brown</b>	<b>Harry</b>	<b>French I</b>	<b>Yes</b>
<b>3</b>	<b>Brown</b>	<b>Harry</b>	<b>Calculus</b>	<b>No</b>
<b>5</b>	<b>White</b>	<b>Edward</b>	<b>French II</b>	<b>No</b>
<b>5</b>	<b>White</b>	<b>Edward</b>	<b>Algebra</b>	<b>No</b>



Students-Courses

si d	lname	fname	cname	majo r
1	Jones	Allan	French I	No
1	Jones	Allan	Calculus	Yes
2	Smith	John	Algebra	No
3	Brown	Harry	French I	Yes
3	Brown	Harry	Calculus	No
5	White	Edward	French II	No
5	White	Edward	Algebra	No

物理实现  
结构

用户视角  
中的结构

DISK

File B

File A

逻辑独立性

students

si d	lname	fname	class	telephone
1	Jones	Allan	2	555-1234
2	Smith	John	3	555-4321
3	Brown	Harry	2	555-1122
5	White	Edward	3	555-3344

enrollment

si d	cn o	majo r
1	10 1	No
1	10 8	Yes
2	10 5	No
3	10 1	Yes
3	10 8	No
5	10 2	No
5	10 5	No

物理独立性

courses

cn o	cname	croom	time
10 1	French I	2-104	MW 2
10 2	French II	2-113	MW 3
10 5	Algebra	3-105	MW 2
10 8	Calculus	2-113	MW 4

系统层次上  
的结构

# 1.3 数据库系统的基本特点

## □ 数据的统一管理与控制

### ➤ 数据的完整性检查

- 对数据库中数据正确性作检查以保证数据的正确

### ➤ 数据的安全性保护

- 对数据库访问者作检查以防止非法访问

### ➤ 并发控制

- 对多个应用并发访问所产生的相互干扰作控制以保证其正确性

### ➤ 数据库故障恢复

- 对遭受破坏的数据具有恢复能力，使数据库具有抗破坏性



# 回顾

## □ 数据、数据库、数据库管理系统、数据库系统

### ➤ 数据库系统的组成部分

- 数据库
- 数据库管理系统
- 数据库管理员
- 软件平台
- 硬件平台

## □ 数据库系统的特点

- 数据的集成性
- 数据的高共享性与低冗余性
- 数据独立性
- 数据的统一管理与控制

# 第一章 数据库系统概述

## 1.1 基本概念

## 1.2 数据库系统的发展及趋势

## 1.3 数据库系统的基本特点

## 1.4 数据库内部结构体系

## 1.4 数据库内部结构体系

### □ 数据库系统的三级模式

- 数据模式是数据库系统中数据结构的一种表示形式，它具有不同的层次与结构方式
  - 概念模式（简称模式）
  - 外模式（也称子模式、用户模式）
  - 内模式（也称物理模式）
- 数据库系统的三级模式是对数据在三种不同的抽象级别上进行的描述，它把数据的具体物理实现留给物理模式，使用户与全局设计者能不必关心数据库的具体实现与物理背景

## 1.4 数据库内部结构体系

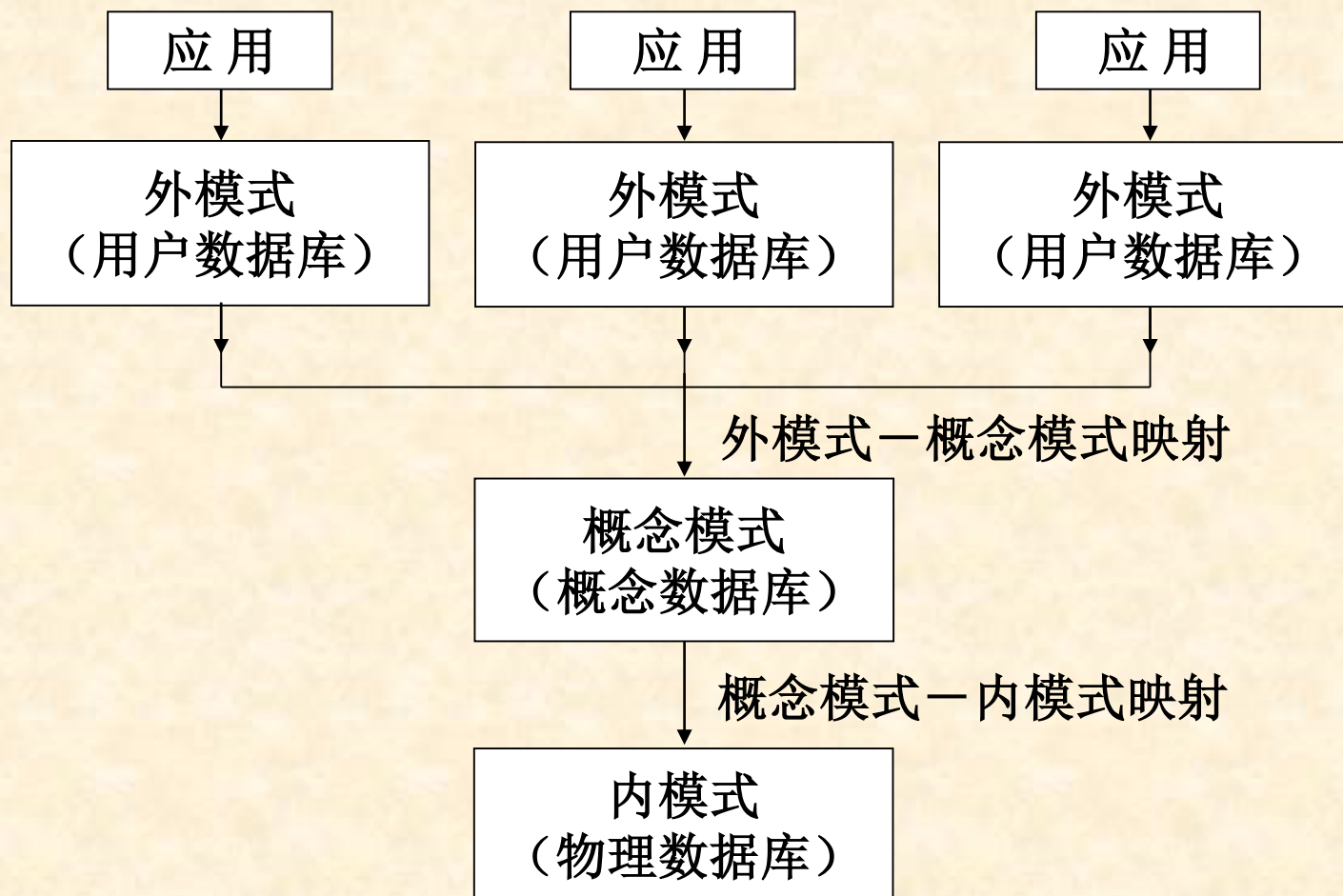


图1.6 三级模式两种映射关系图

## 1.4 数据库内部结构体系

### □ 概念模式

- 是关于整个数据库中数据的全局逻辑结构的描述
- 是面向数据库系统的，它一般以某种数据模型为基础，利用具体的DBMS所提供的数据定义语言（DDL）来描述
  - 数据的类型、长度、特征
  - 数据间的联系
  - 安全性、完整性等方面的要求

# 1.4 数据库内部结构体系

## □ 外模式（子模式、用户模式）

- 是关于某个用户所需数据的逻辑结构的描述
- 外模式可由概念模式推导而出，是概念模式的一个子集
  - 一个概念模式可对应着多个外模式
  - 一个用户只关心并通过与其相关的外模式来使用数据库，其有利之处在于
    - 简化用户接口，便于使用
    - 降低数据冗余度
    - 有利于数据的安全性保护和保密



# 1.4 数据库内部结构体系

## □ 内模式（物理模式）

- 是关于数据库中数据的物理存储结构和物理存取方法的描述

## □ 利用上述的三种模式可构造出三个层次上的数据库概念

- 概念数据库
- 用户数据库
- 物理数据库

- 在这三种数据库中，只有‘**物理数据库**’是真实存在于计算机的外存中的，其他两种数据库并不真正存在于计算机中，而是由物理数据库通过数据库管理系统构造而成的



# 1.4 数据库内部结构体系

## □ 数据库系统的两级映射

### ➤ 概念模式到内模式的映射

- 该映射给出了概念模式中数据的全局逻辑结构到数据的物理存储结构间的对应关系，此种映射一般由DBMS实现

- 可实现‘物理独立性’

### ➤ 外模式到概念模式的映射

- 概念模式是一个全局模式，而外模式则是用户的局部模式。一个概念模式中 can 定义多个外模式，而每个外模式是概念模式的一个基本视图
- 外模式到概念模式的映射给出了外模式与概念模式的对应关系，这种映射一般也由DBMS实现

- 可实现‘逻辑独立性’

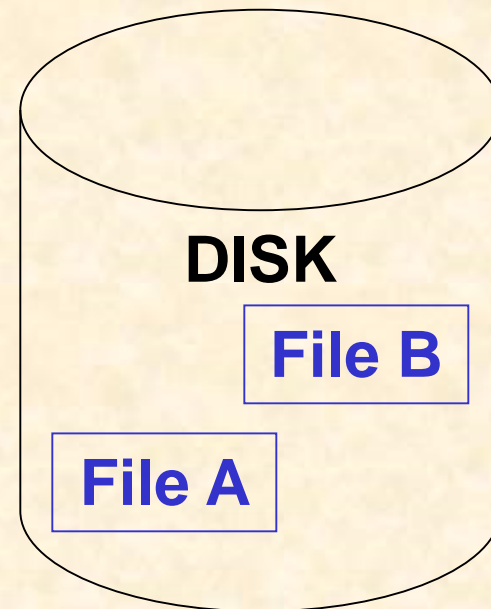
## 1.4 数据库内部结构体系

### □ 数据库系统的两级映射

- 数据库系统通过两级映射建立三级模式间的**联系与转换**，使得概念模式与外模式虽然在物理上并不存在，但也能通过映射而获得其存在的实体
- 三级模式之间的两级映射也保证了数据库系统中**数据独立性的**实现，亦即数据的内模式或概念模式的改变，并不影响用户使用到的外模式，只要调整它们之间的映射方式即可

# Students-Courses

sid	lname	fname	cname	major
1	Jones	Allan	French I	No
1	Jones	Allan	Calculus	Yes
2	Smith	John	Algebra	No
3	Brown	Harry	French I	Yes
3	Brown	Harry	Calculus	No
5	White	Edward	French II	No
5	White	Edward	Algebra	No



逻辑独立性

students

sid	lname	fname	class	telephone
1	Jones	Allan	2	555-1234
2	Smith	John	3	555-4321
3	Brown	Harry	2	555-1122
5	White	Edward	3	555-3344

courses

cno	cname	croom	time
101	French I	2-104	MW2
102	French II	2-113	MW3
105	Algebra	3-105	MW2
108	Calculus	2-113	MW4

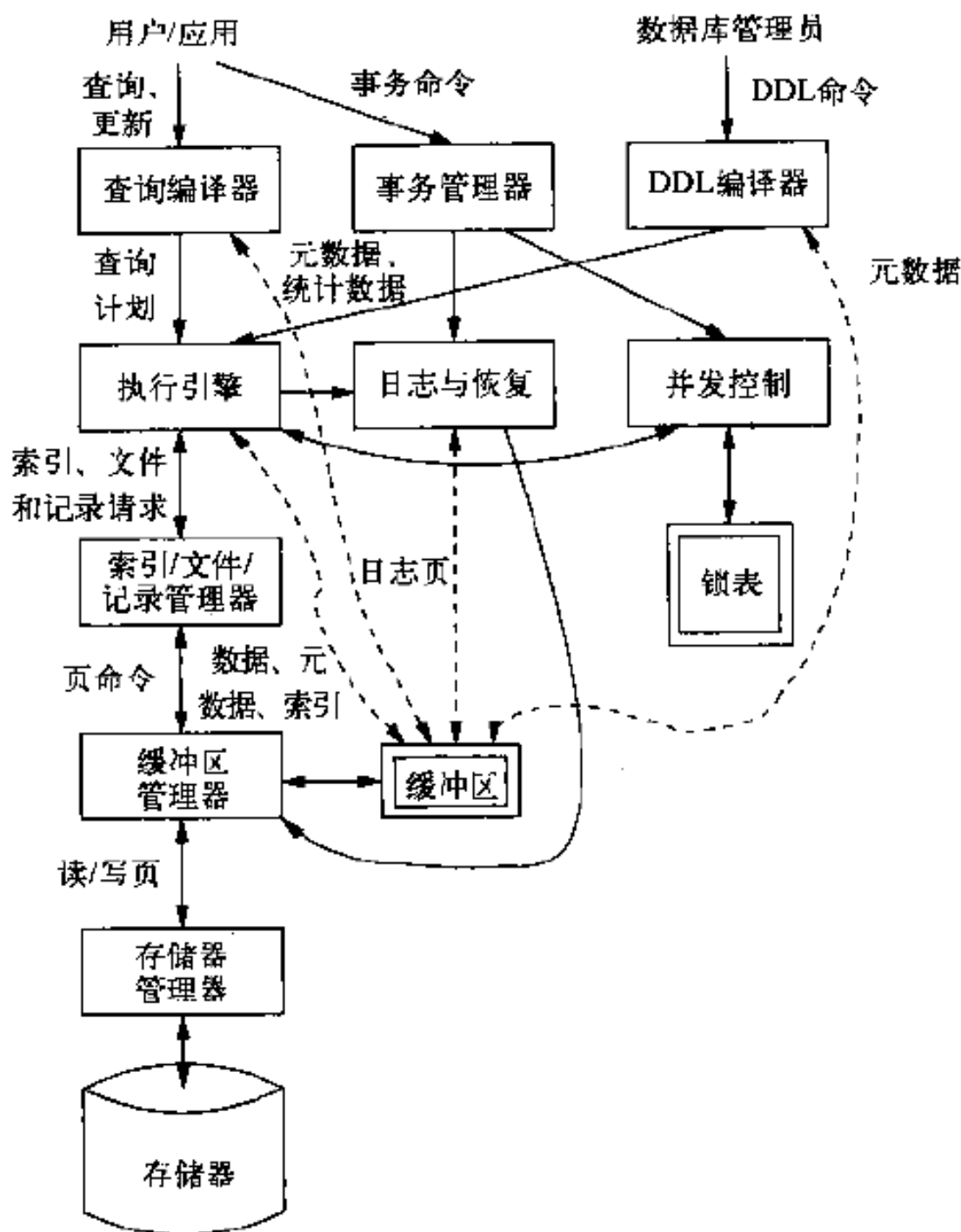
enrollment

sid	cno	major
1	101	No
1	108	Yes
2	105	No
3	101	Yes
3	108	No
5	102	No
5	105	No

物理独立性

# 数据库管理系统

□ 单线框——系统构成，双线框——内存中的数据结构  
实线——控制和数据流，虚线——数据流



# 本章小结

## □ 基本概念

- 数据库，数据库管理系统，数据库管理员
- 数据库系统，数据库应用系统

## □ 数据库系统的发展历史

## □ 数据库系统的基本特点

- 数据集成性
- 数据的一致性
- 数据独立性
- 完整性，安全性，并发控制，故障恢复

## □ 三级模式与两级映射

- 概念模式，外模式，内模式
- 两级映射