

数据库原理

第1章 数据库基础

辽东学院 鲁 琴

本章要点

数据库，数据库管理系统，数据库系统，数据库管理员

数据库和信息

数据库系统的体系结构

数据模型

数据管理技术的产生和发展

数据库基础概念

数据库原理

关系数据库

数据库新技术

1. 数据和信息

(1) 信息

- 信息是客观存在的，是关于现实世界事物的存在方式或运动状态反映的综合
- 例如，辽东学院的学生
- 信息与材料和能源一个层次,是人类社会赖以生存和发展的三大资源之一

(2) 数据

- 数据是用来记录信息的可识别的符号，是信息的具体表现形式
- 例如：计算机中的一张二维表，该数据辽东学院学生的基本信息
- 其表现形式包括数字、文字、图形、图像和声音等等

(3) 数据与信息的联系

- 数据是信息的符号表示或载体
- 信息则是数据的内涵，是对数据的语义解释

sid	sname	sex	tel
0911160103	耿直	男	13470071013
0911417002	王婷	女	13470071011
0911417003	刘文远	男	13470071012
0911417005	薛国成	男	13470071013
0911417006	尹博涵	女	13470071014
0911417007	吴丽秋	女	13470071015
0911417008	曲思乐	女	13470071016
0911417009	张雪	女	13470071017
0911417010	刘健	女	13470071018
0911417012	宿仕博	女	13470071019

第1章 数据库基础

1.数据和信息

2.数据库, 数据库管理系统, 数据库系统, 数据库管理员

3.数据库系统的体系结构

4.数据模型

5.数据管理技术的产生和发展

2. 数据库, 数据库管理系统, 数据库系统, 数据库管理员

(1) 数据库

- 数据库(Database,简称DB)是长期储存在计算机内、有组织的、可共享的数据集合

数据库的特征

- 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
- 数据独立性较高
- 易扩展

sid	sname	sex	tel
0911160103	耿直	男	13470071013
0911417002	王婷	女	13470071011
0911417003	刘文远	男	13470071012
0911417005	薛国成	男	13470071013
0911417006	尹博涵	女	13470071011
0911417007	吴丽秋	女	13470071011
0911417008	曲思乐	女	13470071011
0911417009	张雪	女	13470071011
0911417010	刘健	女	13470071011
0911417012	宿仕博	女	13470071011

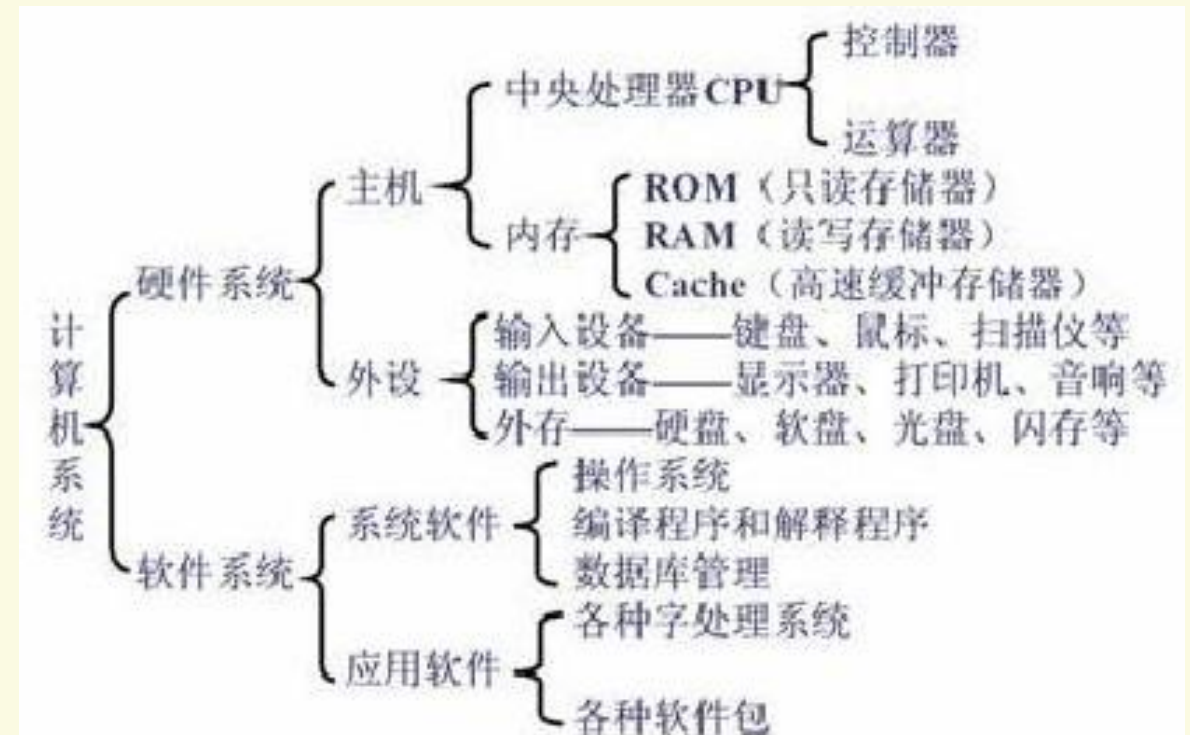
2. 数据库, 数据库管理系统, 数据库系统, 数据库管理员

(2) 数据库管理系统

- 数据库管理系统 (Database Management System, 简称DBMS) 是位于用户与操作系统之间的一层数据管理**软件**
- 科学地组织和存储数据、高效地获取和维护数据

讨论

- 什么是OS?
- DBMS和OS的关系?

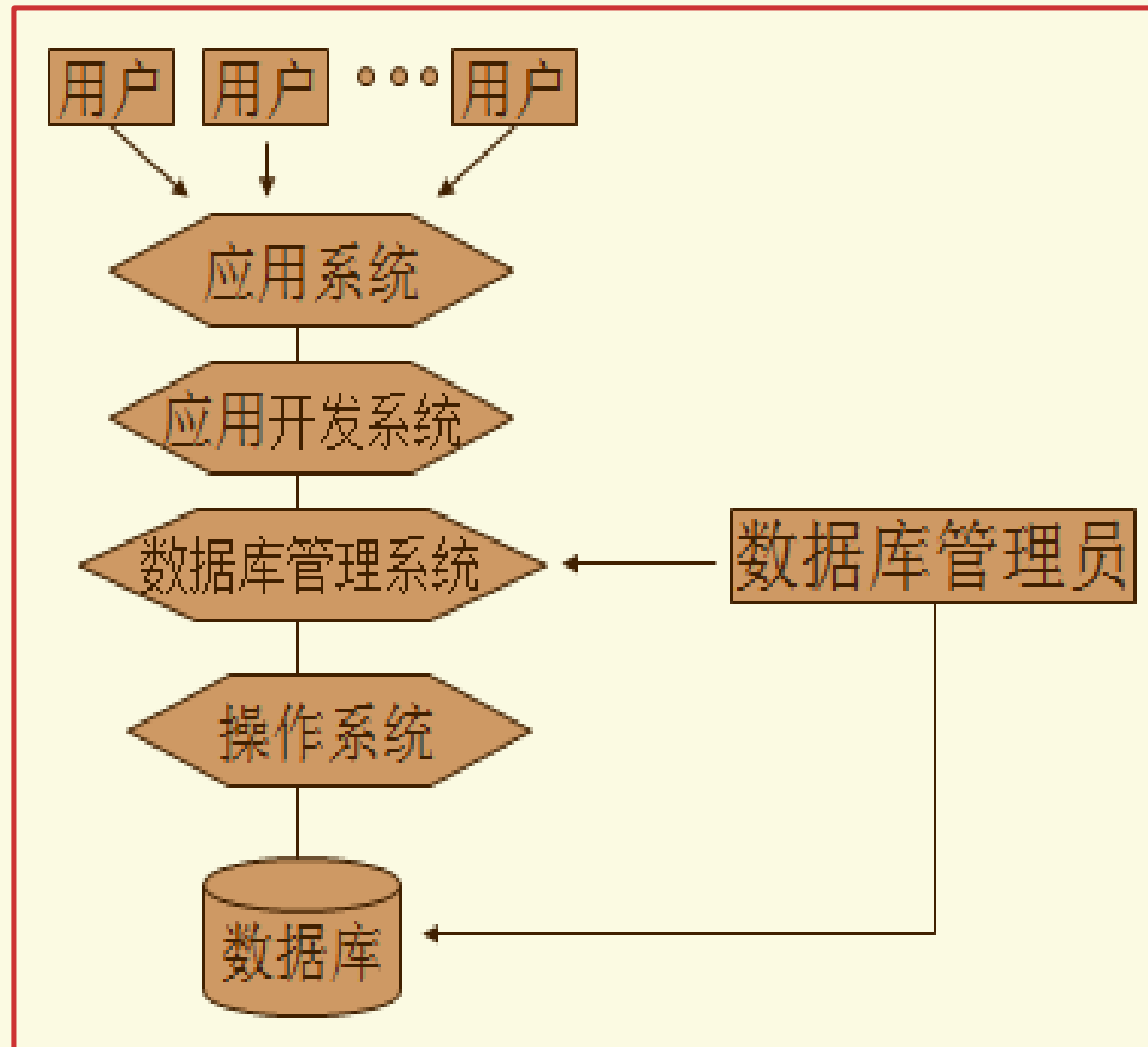


2. 数据库, 数据库管理系统, 数据库系统, 数据库管理员

DBMS的主要功能

- 数据定义
 - 提供数据定义语言(DDL), 用于定义数据库中的数据对象
- 数据操纵
 - 提供数据操纵语言(DML), 用于操纵数据实现对数据库的基本操作(查询、插入、删除和修改)
- 数据库的运行管理
 - 保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复
- 数据库的建立和维护功能
 - 提供实用程序, 完成数据库数据批量装载, 数据库转储, 介质故障恢复, 数据库的重组和性能监视等

2. 数据库, 数据库管理系统, 数据库系统, 数据库管理员



(3) 数据库系统

(Database System, 简称 DBS) 是指在计算机系统中引入数据库后的系统构成

数据库系统的构成

- 数据库
- 数据库管理系统
- 应用系统（及其开发工具）
- 相关人员（数据库管理员、应用系统开发人员、用户）

2. 数据库, 数据库管理系统, 数据库系统, 数据库管理员

(4) 数据库管理员

- 数据库管理员(Database Administrator,简称DBA)
- 负责全面管理和控制数据库系统，是数据库系统最重要的人员。

DBA的主要职责

- 设计和定义数据库系统
- 帮助最终用户使用数据库系统
- 监督与控制数据库系统的使用和运行
- 改进和重组数据库系统
- 调优数据库系统的性能
- 转储和恢复数据库
- 重构数据库

第1章 数据库基础

1.数据和信息

2.数据库, 数据库管理系统, 数据库系统, 数据库管理员

3.数据库系统的体系结构

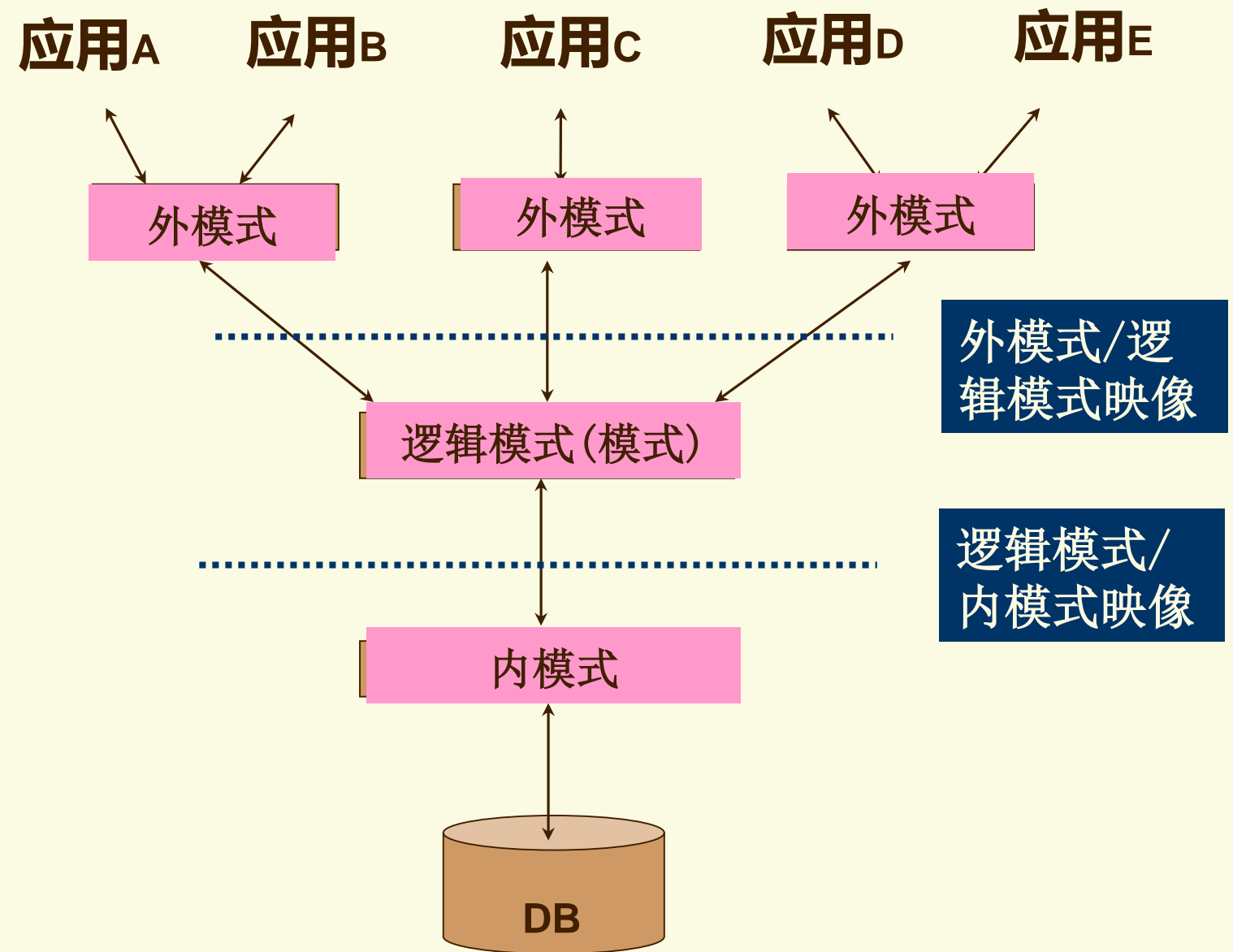
4.数据模型

5.数据管理技术的产生和发展

3. 数据库系统的体系结构 三层模式和两级映像

在用户(或应用程序)到数据库之间,DB的数据结构有3个层次: **外部模型,逻辑模型和内部模型**

这三个层次要用DB的数据定义语言(DDL)定义,定义以后的内容,称为**“模式”**,



① 逻辑模式(也称模式)(Schema)

- ◆ 数据库中**全体数据**的逻辑结构和特征的描述
- ◆ 是综合了**所有用户**的需求的**公共数据视图**
- ◆ 一个数据库只有一个逻辑模式

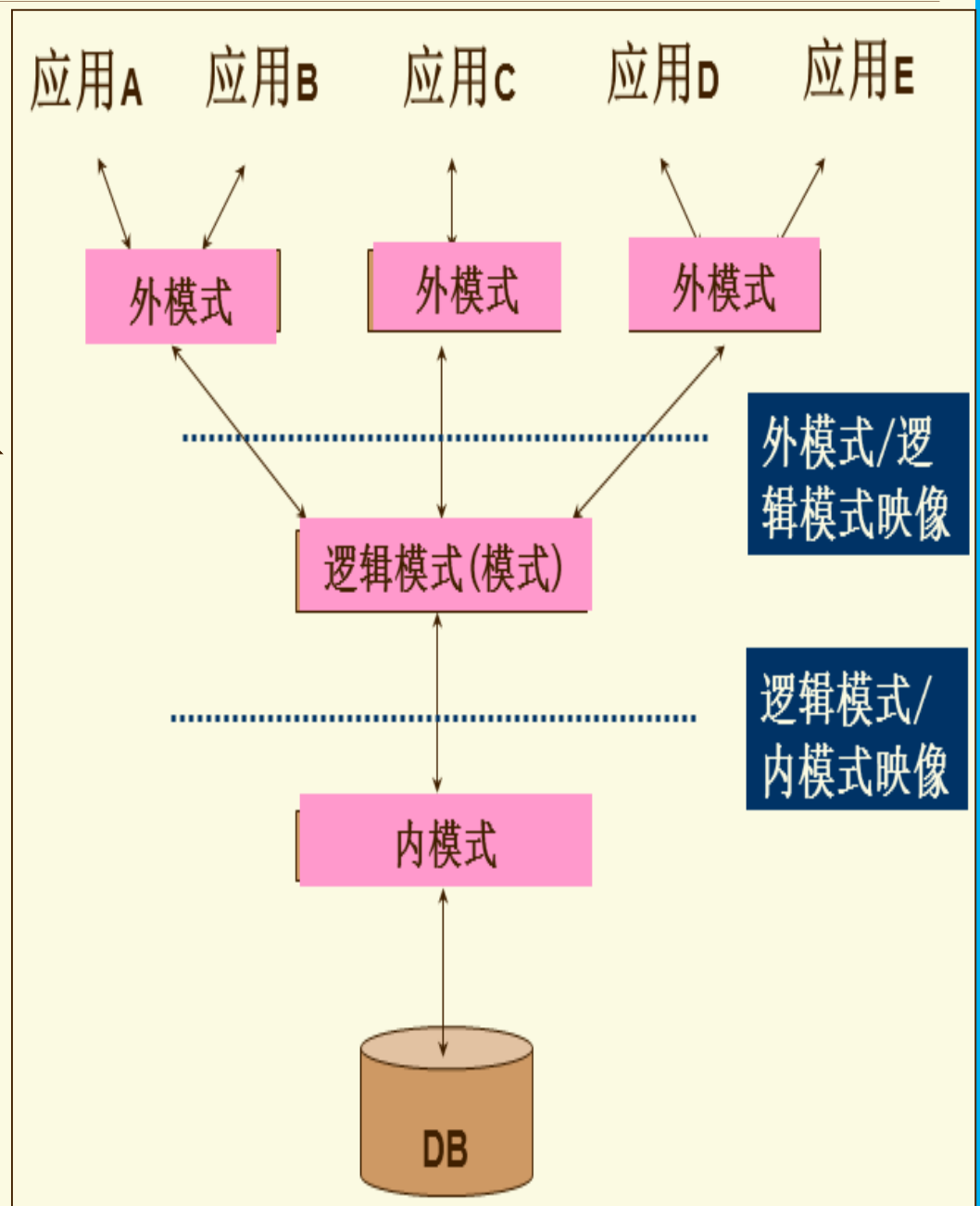
逻辑模式的地位：是数据库系统模式结构的中间层

- 与数据的物理存储细节和硬件环境无关
- 与具体的应用程序、开发工具及高级程序设计语言无关

逻辑模式定义的内容

Student(sid,sname,sex,tel)

- 数据的逻辑结构（数据项的名字、类型、取值范围等）
- 数据之间的联系
- 数据有关的安全性、完整性要求



② 外模式(也称子模式或用户模式)(External Schema)

◆数据库用户(包括应用程序员和最终用户)使用的局部数据的逻辑结构和特征的描述

◆数据库用户的数据视图,是与某一应用有关的数据的逻辑表示

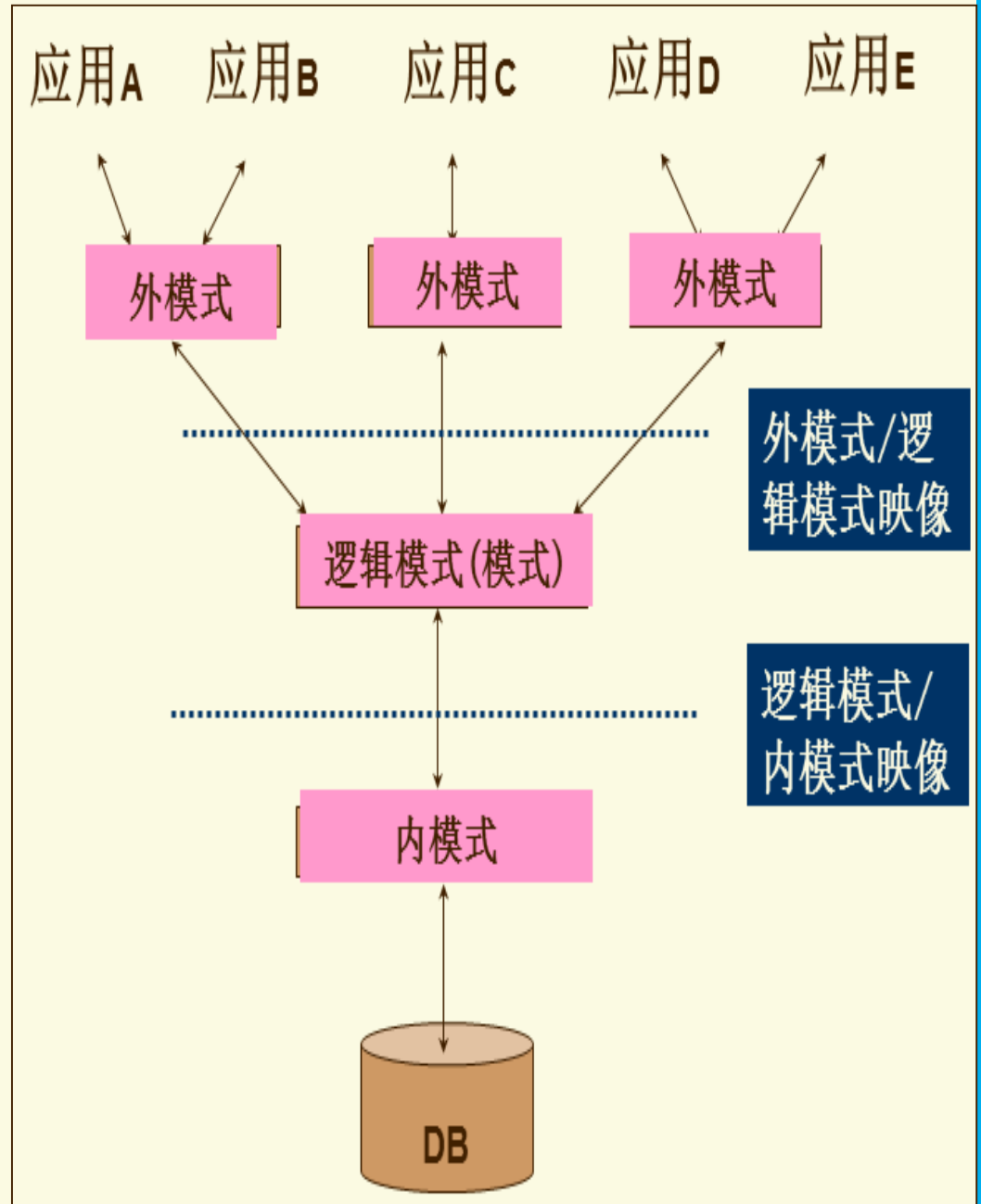
外模式的地位: 介于模式与应用之间

- 逻辑模式与外模式的关系: 一对多
- 外模式通常是逻辑模式的子集

外模式与应用的关系: 一对多

外模式的用途

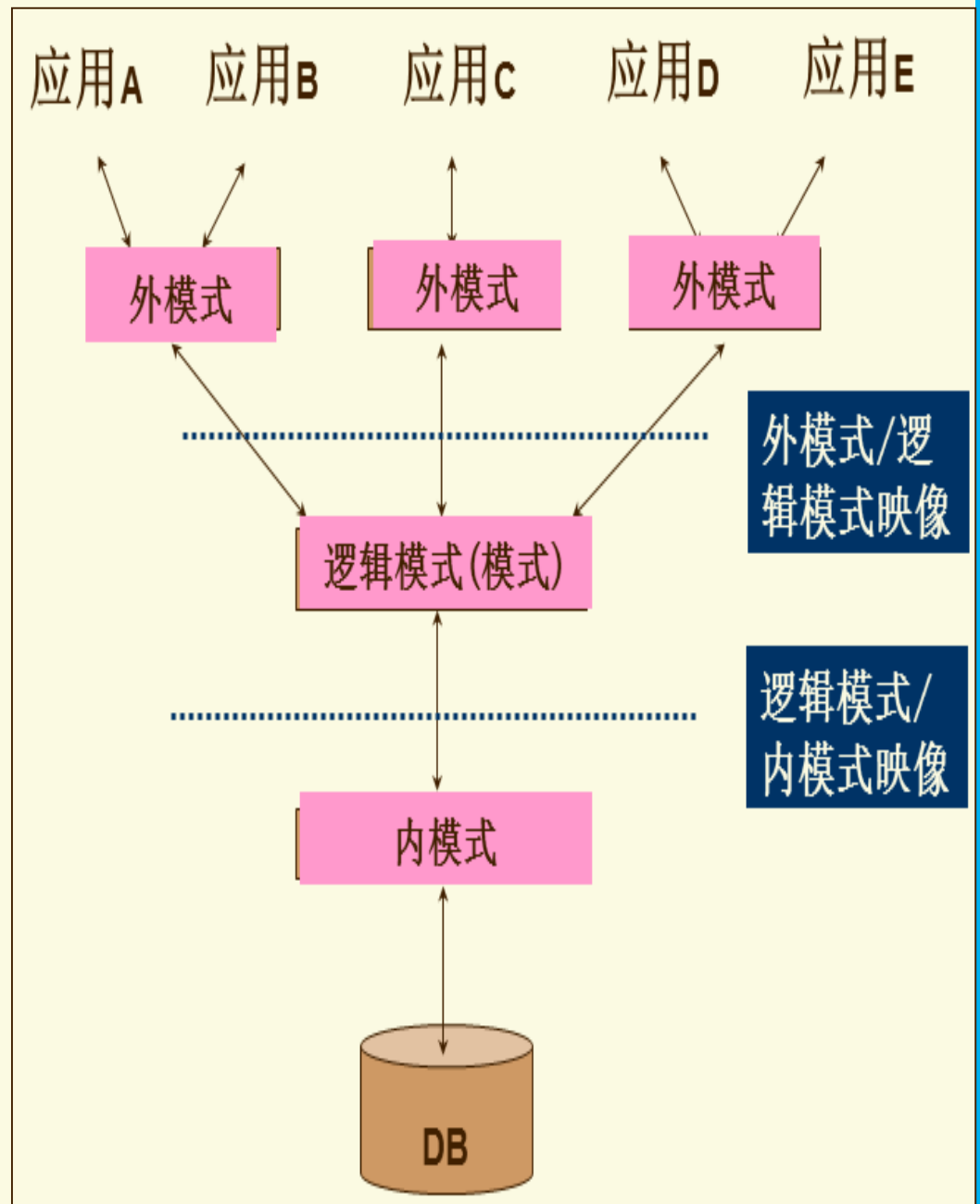
- ◆保证数据库安全性的一个有力措施。每个用户只能看见和访问所对应的外模式中的数据
- ◆保证数据独立性的一个有力措施



③内模式(也称存储模式)(Internal Schema)

- ◆ 是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示方式
 - 记录的存储方式（顺序存储，按照B树结构存储，按hash方法存储）
 - 索引的组织方式
 - 数据是否压缩存储
 - 数据是否加密
 - 数据存储记录结构的规定

一个数据库只有一个内模式



三层模式与二级映像

- ◆ 三层模式是对数据的三个抽象级别
- ◆ 二级映像 **在DBMS内部实现**这三个抽象层次的联系和转换

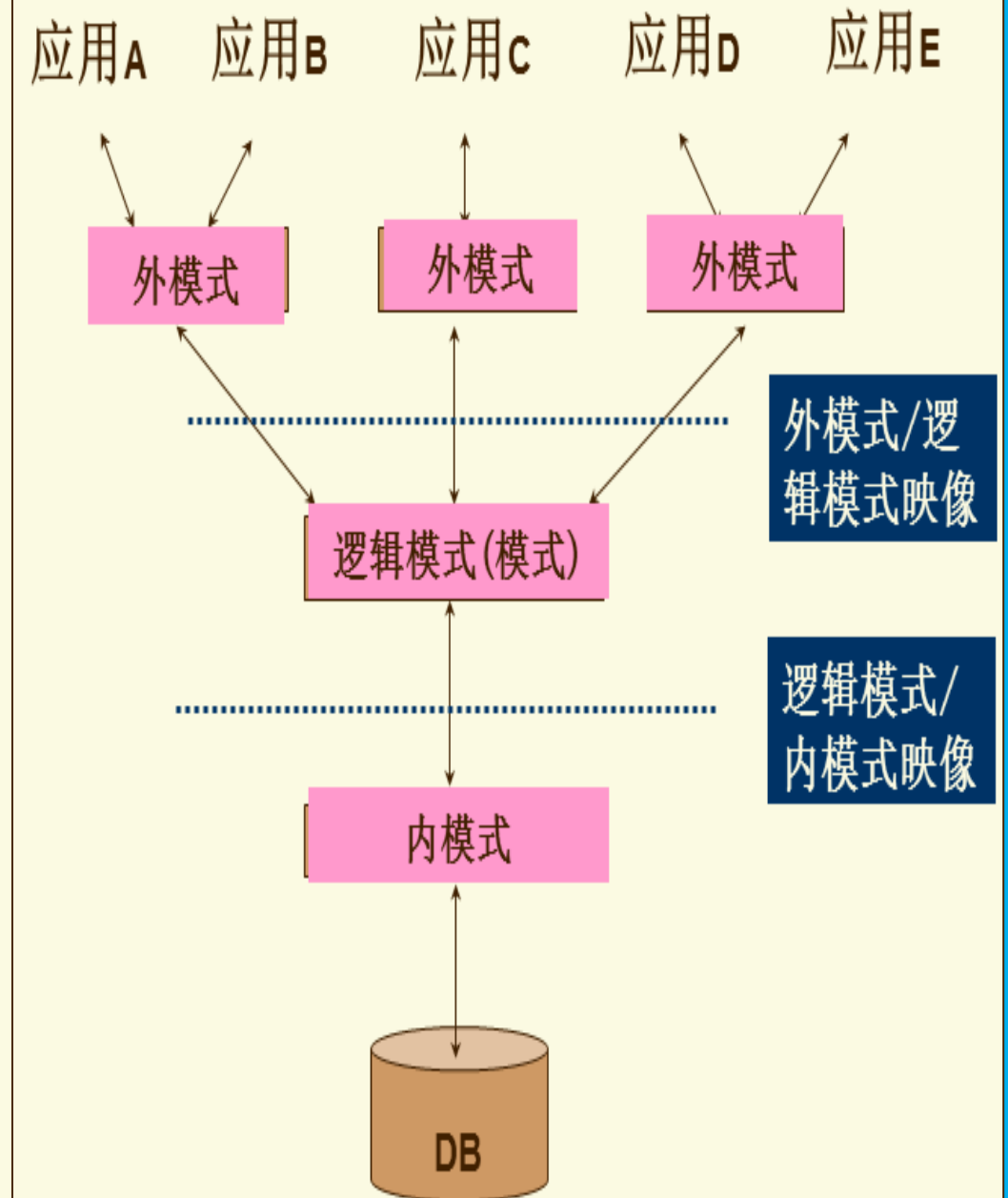
二级映像

① 外模式 / 逻辑模式映像

② 逻辑模式 / 内模式映像

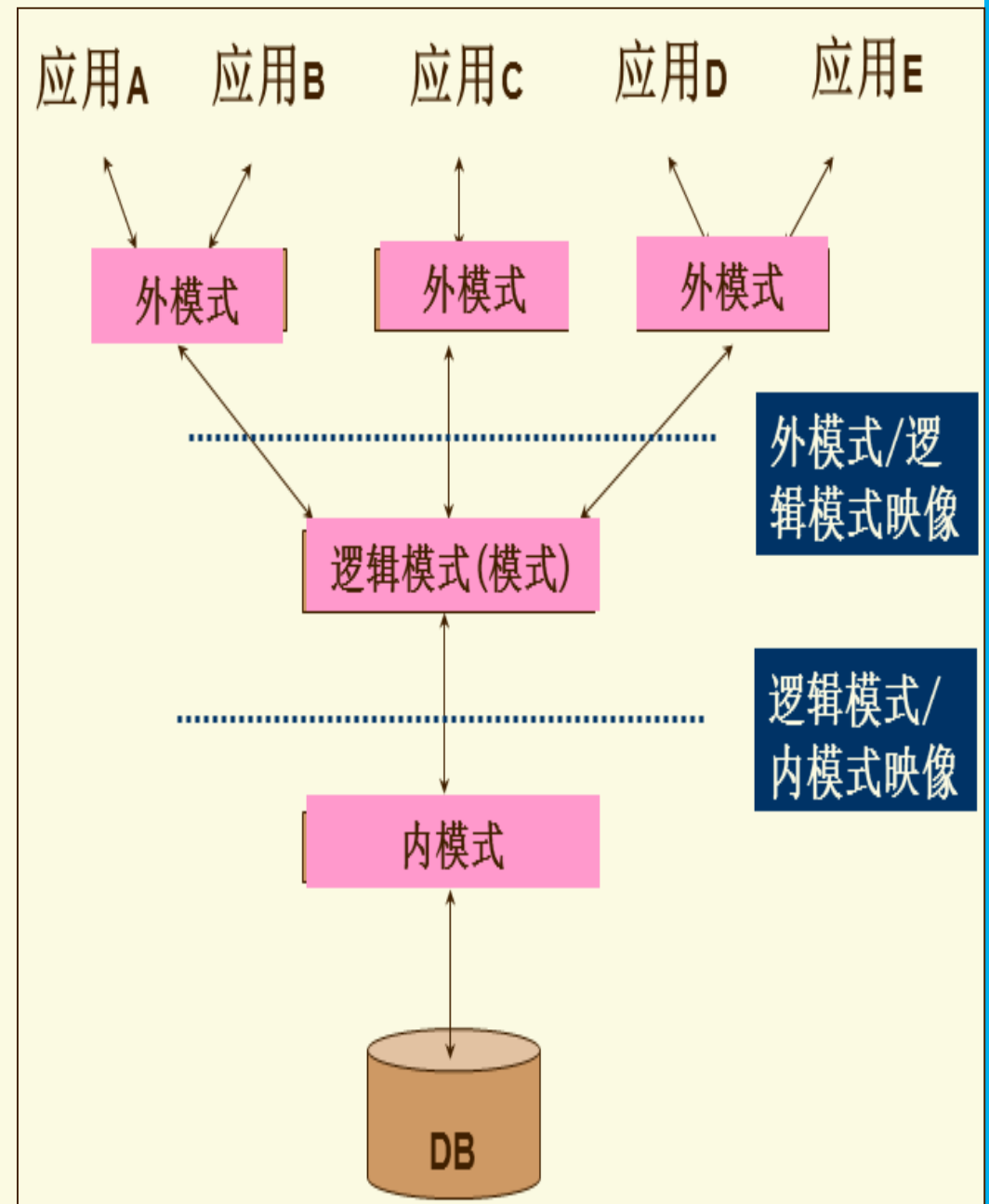
① 外模式 / 模式映像

- ◆ 定义外模式与逻辑模式之间的对应关系
- ◆ 每一个外模式都对应一个外模式 / 模式映像
- ◆ 映像定义通常包含在各自外模式的描述中



② 逻辑模式 / 内模式映像

- ◆ 逻辑模式 / 内模式映像定义了数据**全局逻辑结构**与**存储结构**之间的对应关系。
 - **例如**，说明逻辑记录和字段在内部是如何表示的
- ◆ 数据库中模式 / 内模式映像是**唯一**的
- ◆ 该映像定义通常包含在逻辑模式描述中



数据独立性

应用程序与数据之间相互独立

逻辑数据独立性:

protection from changes in the
logical structure of the data

*I.e. should not need to ask:
can we add a new entity or
attribute without rewriting
the application?*

物理数据独立性:

protection from *physical layout*
changes

*I.e. should not need to ask:
which disks are the data
stored on? Is the data
indexed?*

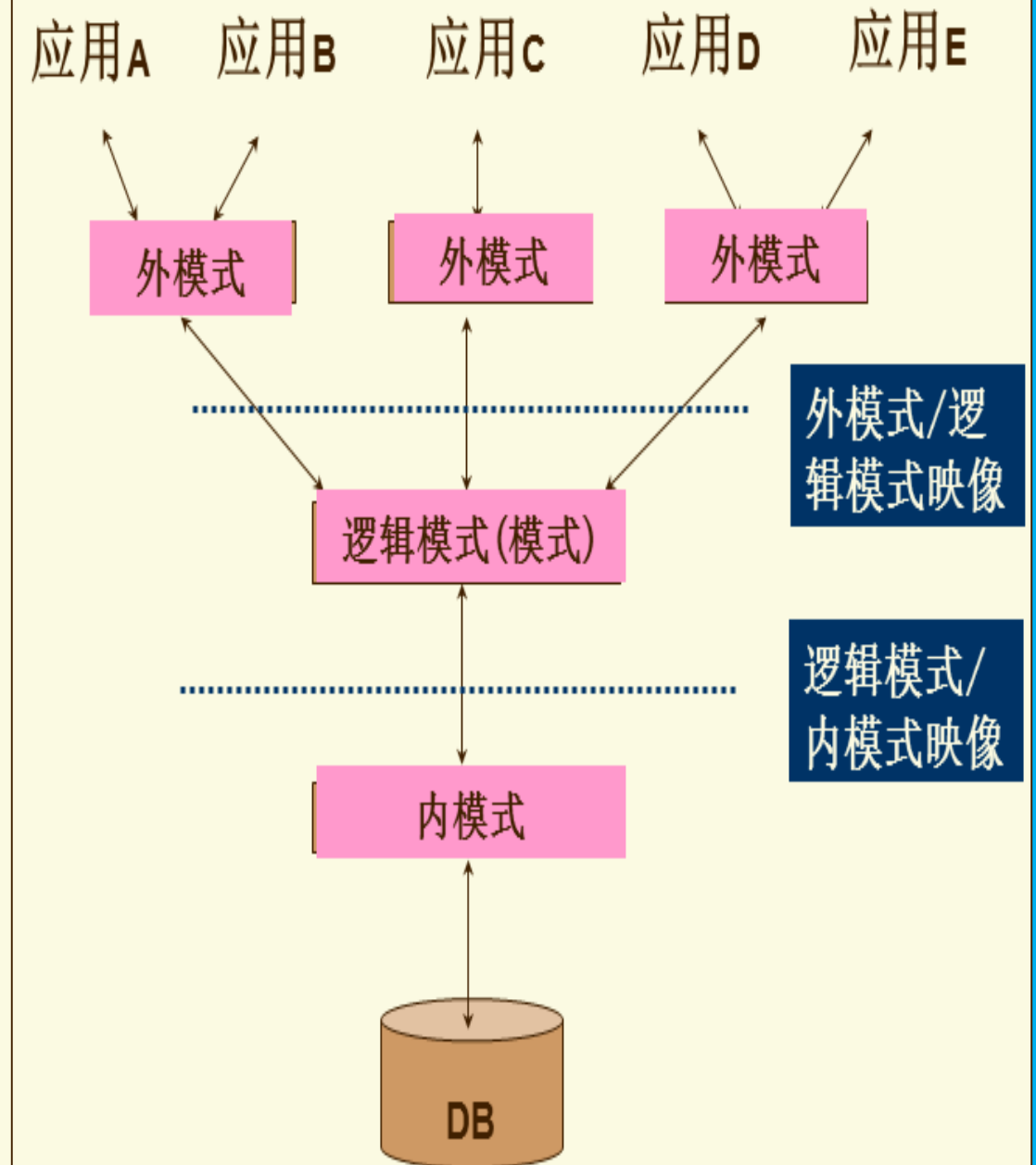
One of the most important reasons to use a DBMS

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，应用程序也可以不变。

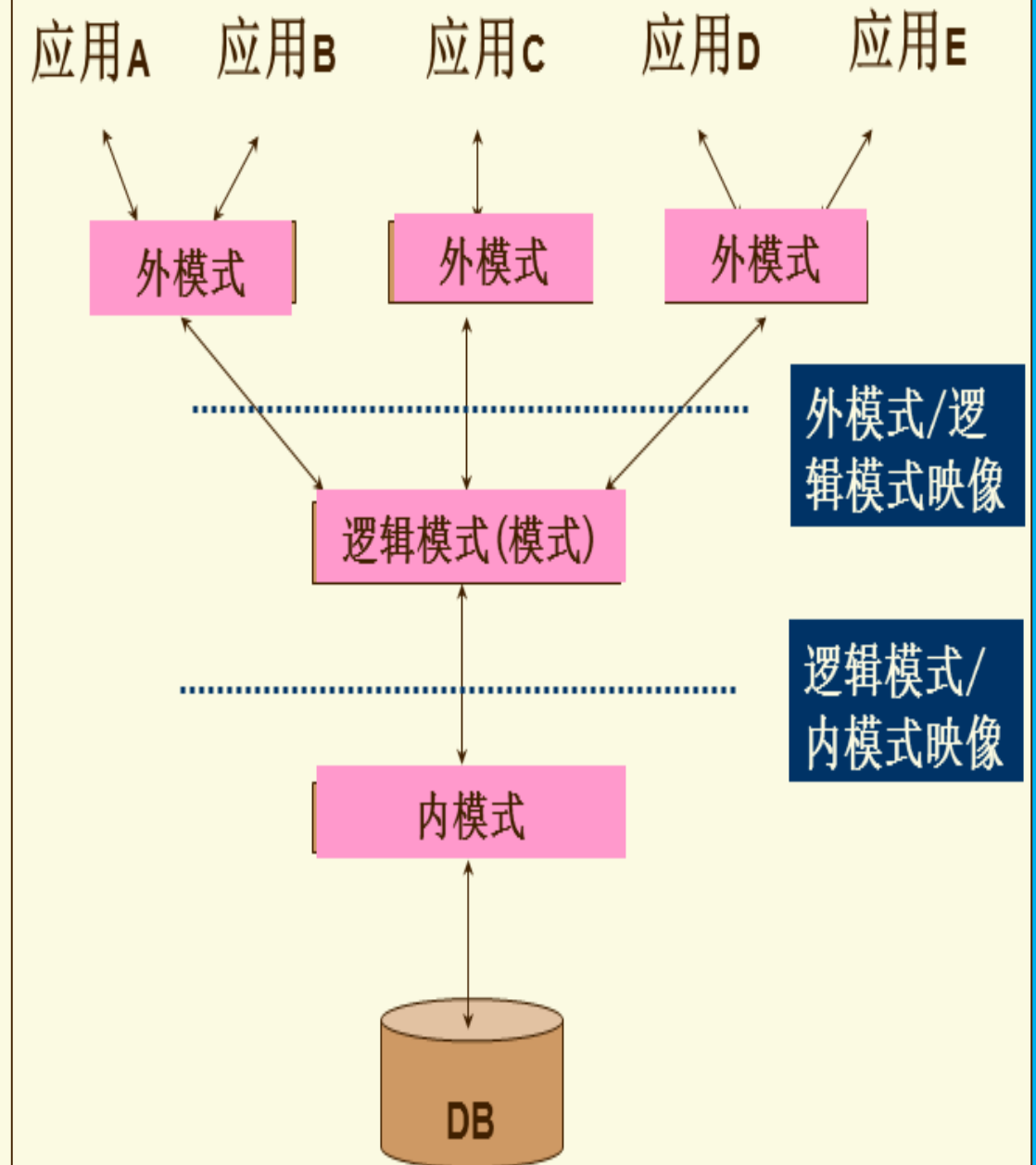
外模式 / 模式映像保证了数据的**逻辑独立性**

- 当模式改变时，数据库管理员修改有关的外模式 / 模式映像，**使外模式保持不变**
- 应用程序是依据数据的外模式编写的，从而应用程序不必修改，保证了数据与程序的**逻辑独立性**。



物理独立性

- ◆ 指用户的**应用程序**与存储在磁盘上的数据库中**数据**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。
- ◆ 模式/内模式映象的用途是保证数据的**物理独立性**
 - 当数据库的存储结构改变了（例如选用了另一种存储结构），数据库管理员修改模式/内模式映象，使模式保持不变
 - 应用程序不受影响，保证了数据与程序的**物理独立性**



第1章 数据库基础

1.数据和信息

2.数据库, 数据库管理系统, 数据库系统, 数据库管理员

3.数据库系统的体系结构

4.数据模型

5.数据管理技术的产生和发展

4. 数据模型

- ◆ 在数据库中用**数据模型**这个工具来抽象、表示和处理现实世界中的数据和信息。
- ◆ **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

数据模型的三要素

①数据结构

- 用于描述静态特性

②数据操作

- 用于描述动态特性

③数据的约束条件

- 是一组完整性规则的集合

数据模型的分类

①关系模型

②半结构化的数据模型

✓ 网状模型

✓ 层次模型

网状模型和层次模型是早期的

✓ XML

✓ JSON

XML和JSON是两种经常在网络使用的数据表示格式

几种数据模型说明

◆ 这些数据模型的根本区别在于数据结构不同

◆ 新的数据模型

关系模型

数据结构 – 表

<i>title</i>	<i>year</i>	<i>length</i>	<i>genre</i>
Star Wars	1977	124	<i>sciFi</i>
Mighty Ducks	1991	104	<i>drama</i>
Wayne's World	1992	95	<i>comedy</i>

Figure : The relation Movies

XML

XML是可扩展标记语言（eXtensible Markup Language）的缩写
常用于传输和存储数据

<Movies>

<Movie title="Star Wars">

<Year>1977</Year>

<Length>124</Length>

<Genre>sciFi</Genre>

</Movie>

<Movie title="Mighty Ducks">

<Year>1991</Year>

<Length>104</Length>

<Genre>drama</Genre>

</Movie>

<Movie title="Star Wars">

<Year>1977</Year>

<Length>124</Length>

<Genre>sciFi</Genre>

</Movie>

</Movies>

title	year	length	genre
Star Wars	1977	124	sciFi
Mighty Ducks	1991	104	drama
Wayne's World	1992	95	

XML

它是一种数据表示格式，可以描述非常复杂的数据结构

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE note SYSTEM "book.dtd">
<book id="1">
  <name>Java核心技术</name>
  <author>Cay S. Horstmann</author>
  <isbn lang="CN">1234567</isbn>
  <tags>
    <tag>Java</tag>
    <tag>Network</tag>
  </tags>
  <pubDate/>
</book>
```

XML有几个特点：

一是纯文本，默认使用UTF-8编码，

二是可嵌套，适合表示结构化数据。

如果把XML内容存为文件，那么它就是一个XML文件，例如book.xml。

此外，XML内容经常通过网络作为消息传输。

JSON

XML的特点是功能全面，但标签繁琐，格式复杂。

在**Web**上使用**XML**现在越来越少，取而代之的是**JSON**这种数据结构

JSON是**JavaScript Object Notation**的缩写，它去除了所有**JavaScript**执行代码，只保留**JavaScript**的对象格式。

一个典型的JSON

```
{  
  "id": 1,  
  "name": "Java核心技术",  
  "author": {  
    "firstName": "Abc",  
    "lastName": "Xyz"  
  },  
  "isbn": "1234567",  
  "tags": ["Java", "Network"]  
}
```

JSON

JSON作为数据传输的格式，有几个显著的优点：

JSON只允许使用**UTF-8**编码，不存在编码问题；

JSON只允许使用双引号作为**key**，特殊字符用****转义，格式简单

浏览器内置**JSON**支持，如果把数据用**JSON**发送给浏览器，可以用**JavaScript**直接处理。

第1章 数据库基础

1.数据和信息

2.数据库, 数据库管理系统, 数据库系统, 数据库管理员

3.数据库系统的体系结构

4.数据模型

5.数据管理技术的产生和发展

5. 数据管理技术的产生和发展

◆ 数据管理技术的产生动力

- 应用需求的推动
- 计算机硬件的发展
- 计算机软件的发展

◆ 数据管理技术的发展阶段

- 人工管理阶段
- 文件系统阶段
- 数据库系统阶段

① 人工管理阶段

- 时期
 - 20世纪40年代中--50年代中
- 产生的背景
 - 应用需求 科学计算
 - 硬件水平 无直接存取存储设备
 - 软件水平 没有操作系统
 - 处理方式 批处理

人工管理阶段的特点

— 数据的管理者是应用程序，数据不保存

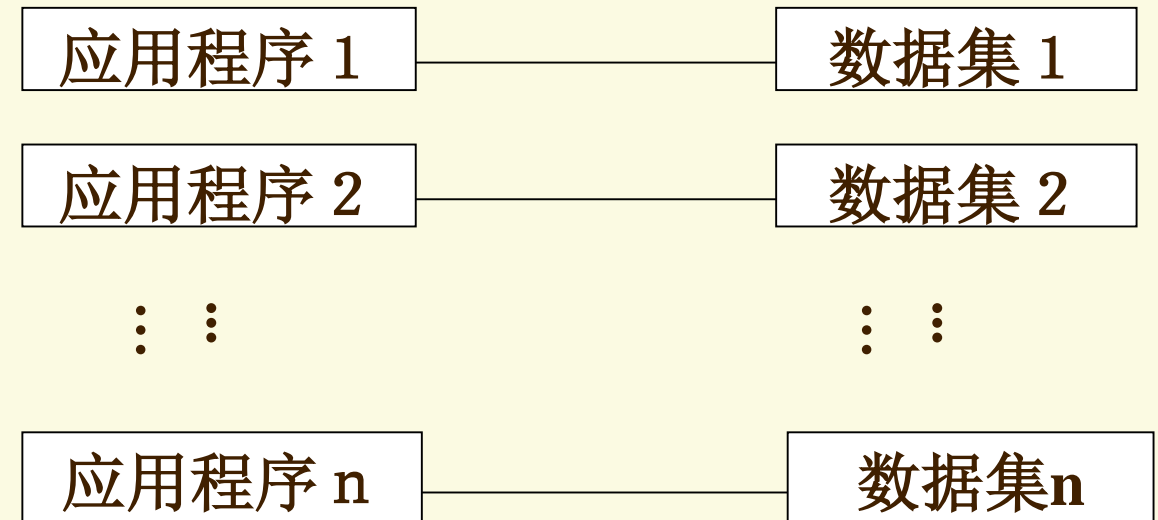
— 数据面向的对象是某一应用程序

— 数据的共享程度：不共享、冗余度极大

— 数据的独立性：不独立，完全依赖于程序

— 数据的结构化：无结构

— 数据控制能力：应用程序自己控制

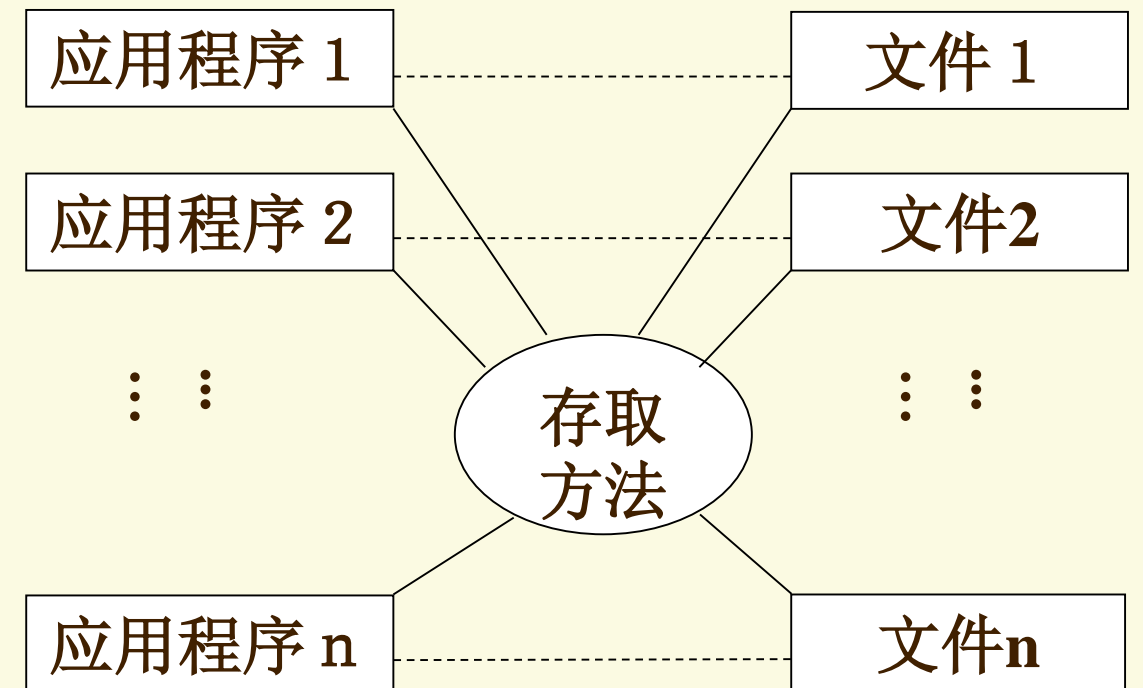


② 文件系统阶段

- 时期
 - 50年代末--60年代中
- 产生的背景
 - 应用需求 科学计算、管理
 - 硬件水平 磁盘、磁鼓
 - 软件水平 有文件系统
 - 处理方式 联机实时处理、批处理

文件系统阶段的特点

- 数据的**管理者**是文件系统，数据可**长期保存**
- 数据**面向的对象**是某一应用程序
- 数据的**共享程度**：共享性差、冗余度大
- 数据的**结构化**：记录内有结构,整体无结构
- 数据的**独立性**：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**：应用程序自己控制



③ 数据库系统阶段

- 时期

- 20世纪60年代末以来

- 产生的背景

- 应用背景 大规模管理

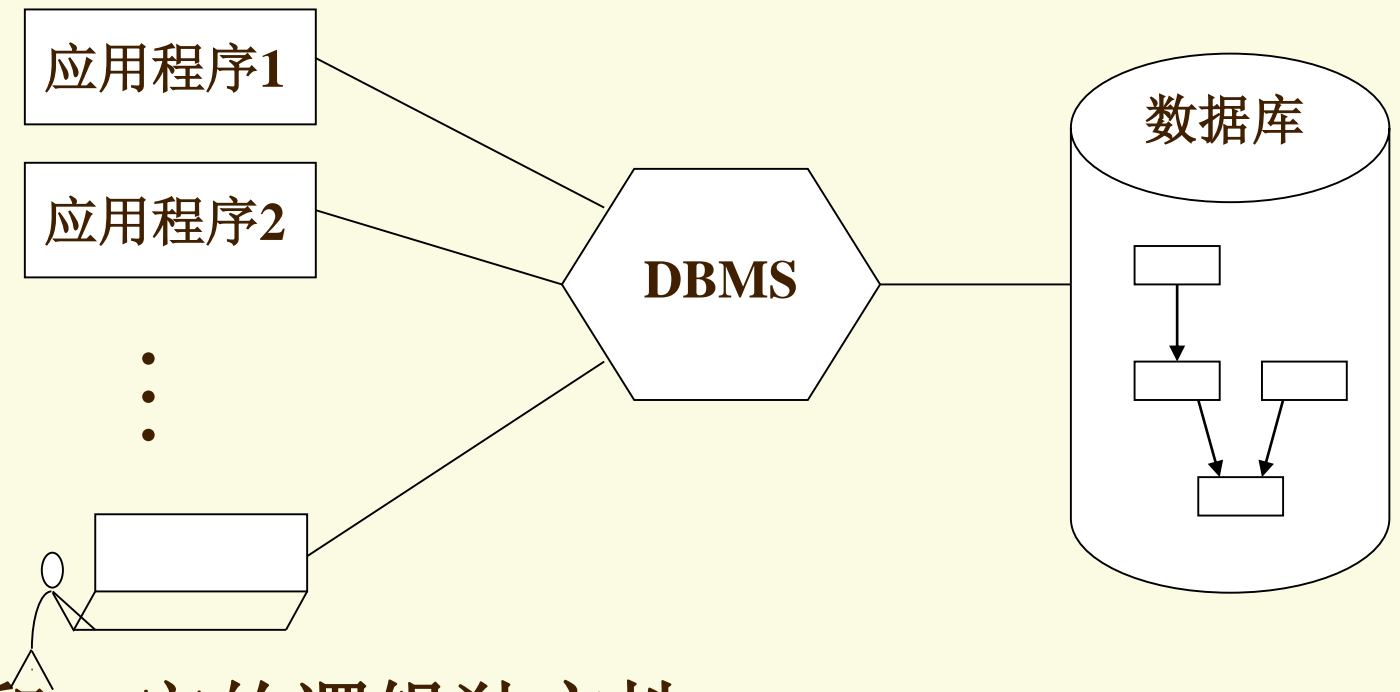
- 硬件背景 大容量磁盘

- 软件背景 有数据库管理系统

- 处理方式 联机实时处理，分布处理，批处理

数据库系统阶段的特点

- 数据的管理者：**DBMS**
- 数据面向的对象：现实世界
- 数据的共享程度：共享性高
 - 降低数据的冗余度，节省存储空间
 - 避免数据间的不一致性
 - 使系统易于扩充
- 数据的独立性：高度的物理独立性和一定的逻辑独立性
- 数据的结构化：整体结构化
 - 数据的结构用数据模型描述，无需程序定义和解释
 - 数据可以变长
 - 数据的最小存取单位是数据项
- 数据控制能力：由**DBMS**统一管理和控制



本章小结

数据库, 数据库管理系统, 数据库系统, 数据库管理员

数据库和信息

数据库系统的体系结构

数据模型

数据管理技术的产生和发展

数据库基础概念

数据库原理

关系数据库

数据库新技术