



数据库系统概述

苏州大学计算机科学与技术学院





知识点1: 数据、数据库的概念

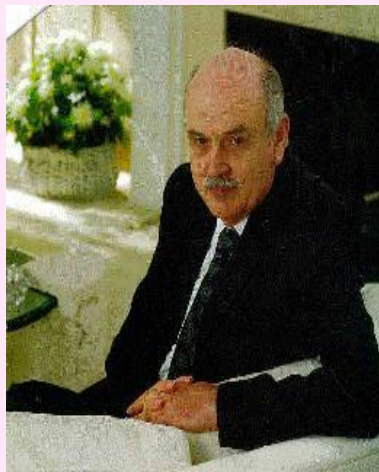


数据库技术的重要性

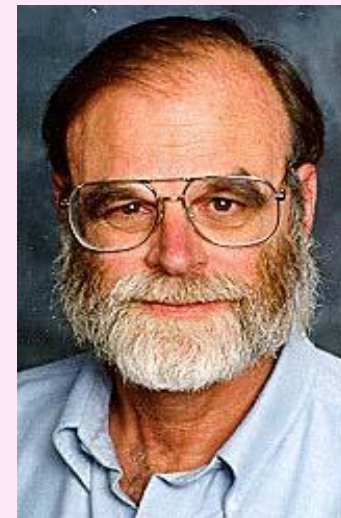
- 数据库技术产生于六十年代末，是计算机科学的重要分支。
- 经历了三代演变：层次/网状系统、关系系统、新一代数据库系统家族
- 造就了三位图灵奖Turing Award得主：
C.W.Bachman、E.F.Codd和James Gray



C.W.Bachman
查尔斯·巴赫曼，美国，
1973，网状数据库模型
以及数据定义（DDL）
和数据操纵语言（DML）
规范说明



E.F.Codd，
埃德加·科德，英国，
1981，关系模型理论



James Gray，
詹姆斯·尼古拉·格雷，
美国，**1998**，
数据库与事务处理



数据库技术的重要性(续)

- 当前，在大、中、小、微型计算机上都有数据库系统在运行。
- 信息资源已成为各部门的重要财富和资源，各行各业均需应用信息系统，而数据库是信息系统的核心。
- 数据库是其它很多系统的核心或重要组成部分，如：
 - **MIS (Management Information System)** 管理信息系统
 - **DSS (Decision Support System)** 决策支持系统，知识库，规则库，专家系统， 智能数据库
 - 计算机辅助设计与制造 工程数据库
 - 计算机集成制造



新型数据库系统

- 数据库技术与其他学科相结合，涌现出各种新型的数据库系统，如：
 - 分布式数据库系统
 - 并行数据库系统
 - 面向对象数据库系统
 - 知识库系统
 - 多媒体数据库系统
 - 模糊数据库系统
 - 网络数据库系统



关系数据库举例（1）

- 目前最为通用的数据库是关系数据库，关系数据库使用“表”来储存数据。

表 - dbo.Songs					
	SongID	Name	Lyricist	Composer	Lang
	S00001	传奇	左右	李健	Chinese
	S00002	后来	施人诚	玉城千春	Chinese
	S01001	Take me home	John Denver	John Denver	English
	S01002	Beat it	Michael Jackson	Michael Jackson	English

歌曲表

一行表示一首歌曲

表 - dbo.Singers					
	SingerID	Name	Gender	BirthYear	Nation
	GA001	Michael_Jackson	男	1958-07-01...	American
	GA002	John_Denver	男	1943-07-01...	American
	GC001	王菲	女	1969-07-01	China
	GC002	李健	男	1974-06-01...	China
	GC003	李小东	男	1969-07-01...	China

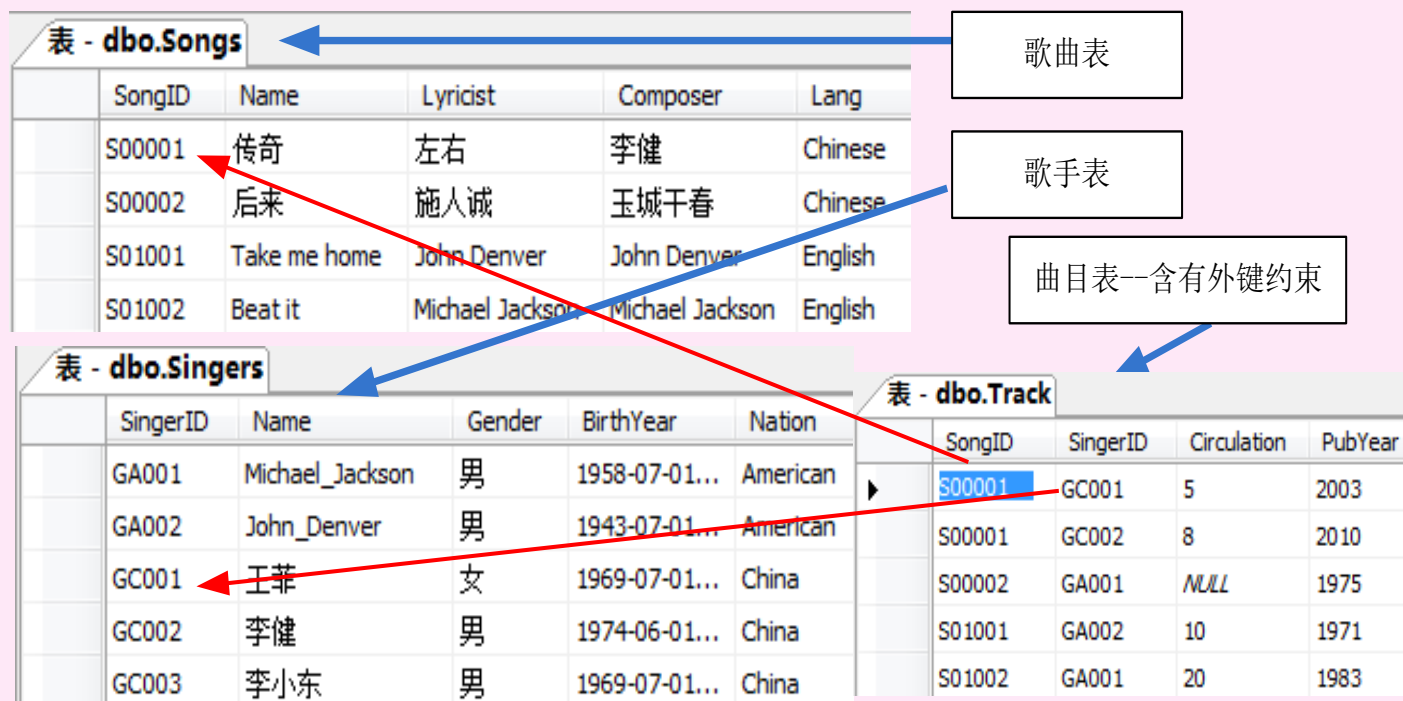
歌手表

这一列表示歌手的名字



关系数据库举例（2）

- 数据库不仅需要存储数据，还需要存储这些数据间的关系，从而可以传递一定的信息量。





数据库系统的基本概念

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



数据

- 信息：客观世界中事物的存在方式和运动状态及其变化的反映，是客观事物之间相互联系和相互作用的表征。
- 数据是对信息的符号化表示，即用一定的符号（数字、文字、图形、图象、声音等）来表示信息。



数据与信息的关系

- 数据与信息的关系：数据是信息的载体，信息是数据的内涵。
- 同一信息可以有不同的数据表示形式，而同一数据也可能有不同的解释。



数据举例

■ 数据:

(李明, 男, 1972, 江苏, 计算机系, 1990)

- 李明是个大学生, 男, 1972年出生, 江苏人, 1990年考入计算机系
- 李明是位老师, 男, 1972年参加工作, 江苏人, 计算机系, 1990年晋升为教授
- 数据的形式不能完全表达其内容, 必须经过解释, **数据的解释**即对数据语义的说明

信息



数据库（举例）

学生登记表

学 号	姓 名	年 龄	性 别	系 名	年 级
95004	王小明	19	女	社会学	95
95006	黄大鹏	20	男	商品学	95
95008	张文斌	18	女	法律学	95
...



数据库概念

- 人们收集并抽取出一个应用所需要的大量数据之后，应将其保存起来以供进一步加工处理，进一步抽取有用信息
- 数据库的定义
 - 数据库(Database,简称DB)是长期储存在计算机内、有组织的、可共享的大量数据集合



数据库的特征

■ 数据库的特征

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



知识点2： 数据库管理系统的概念

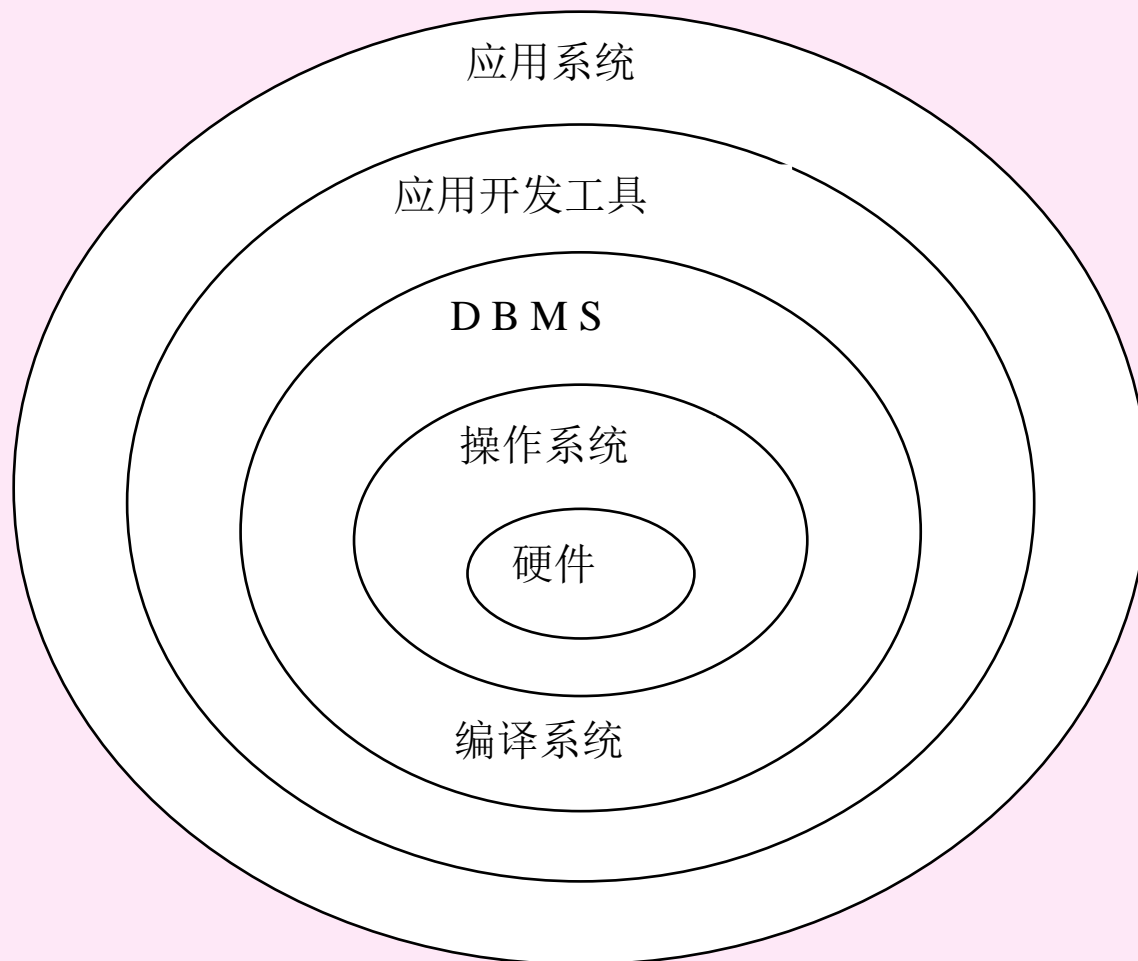


数据库管理系统

- **Database Management System, 简称 DBMS**
- 是位于应用程序与操作系统之间的一层数据管理软件。



DBMS在计算机系统中的地位





DBMS的主要功能（1）

■ 数据定义功能

- 提供数据定义语言DDL（Data Definition Language）
- 定义数据库中的数据对象（在关系数据库管理系统中就是指定义数据库、表、视图、索引等）
- 还包括定义完整性约束和保密限制等



DBMS的主要功能（2）

■ 数据操纵功能

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



DBMS的主要功能（2）

- 数据库的运行管理
 - 安全性控制
 - 完整性控制
 - 多用户对数据的并发使用时的并发控制



DBMS的主要功能 (3)

- 数据库的建立和维护功能(实用程序)
 - 数据库数据批量装载
 - 数据库转储
 - 介质故障恢复
 - 数据库的重组组织
 - 性能监视、分析等



知识点3： 数据库系统的概念

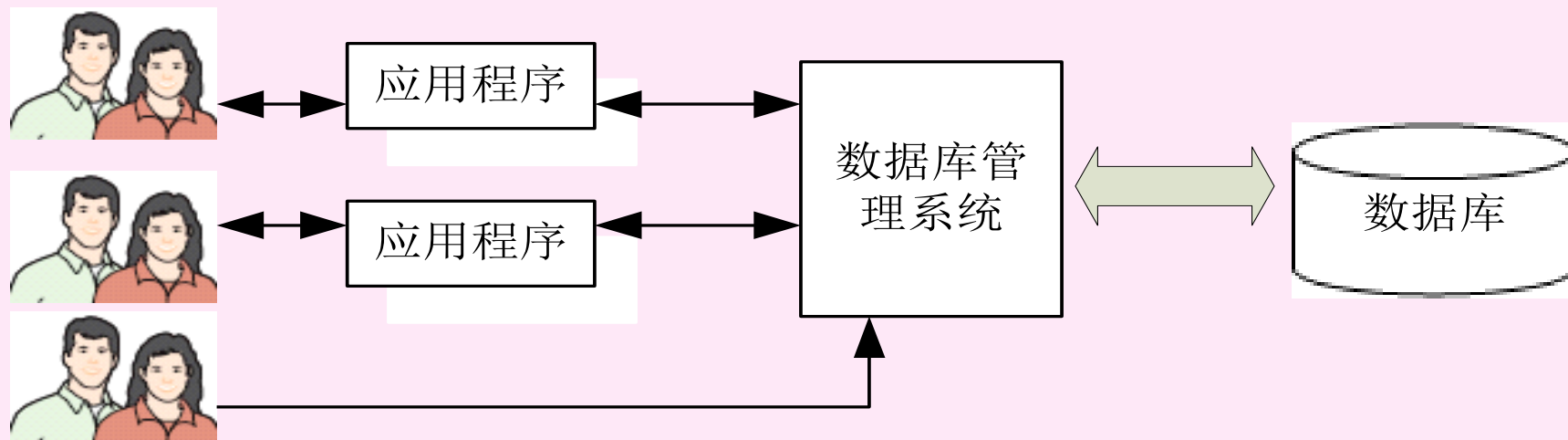


数据库系统

- 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
 - 由数据库、数据库管理系统、应用系统（及其开发工具） 和用户构成。



数据库系统构成图





数据库系统构成（1）

■ 数据库

- 数据库是数据库系统的基石。
- 数据库按一定的数据模型组织、描述和存储数据，具有较小的冗余度、较高的数据独立性和易扩展性，并可被多个用户共享使用。



数据库系统构成（2）

■ 数据库管理系统

- 数据库系统的核心
- 负责数据库存取、维护和管理系统软件
- 其功能的强弱是衡量数据库系统性能优劣的主要指标
- 数据库中的数据由数据库管理系统进行统一管理和控制
- 用户对数据库进行的各种操作是由数据库管理系统实现的



数据库系统构成（3）

■ 应用程序

- 在数据库管理系统的基础上，由用户根据应用的实际需要开发的、处理特定业务的应用程序。
- 普通用户主要利用应用程序来与数据库管理系统交互，以存取数据库内的数据。



数据库系统构成（4）

■ 用户

- 管理、开发、使用数据库系统的所有人员
- 包括数据库管理员、应用程序员和终端用户
- 数据库管理人员可以直接操纵数据库管理系统。数据库的实施、管理、维护一般都由数据库管理人员来完成。
- **DBA**（Database Administrator）——数据库管理员
 - 负责整个数据库系统的建立、管理、维护、协调工作的专门人员



知识点4： 数据库技术的发展历史



数据库技术的发展过程

■ 数据管理技术的发展过程

在应用需求的推动下，在计算机硬件、软件发展的基础上，数据管理技术经历了下面三个阶段：

- 人工管理阶段(40年代中--50年代中)
- 文件系统阶段(50年代末--60年代中)
- 数据库系统阶段(60年代末--现在)



人工管理阶段

■ 时期

- 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
- 数据处理方式 批处理

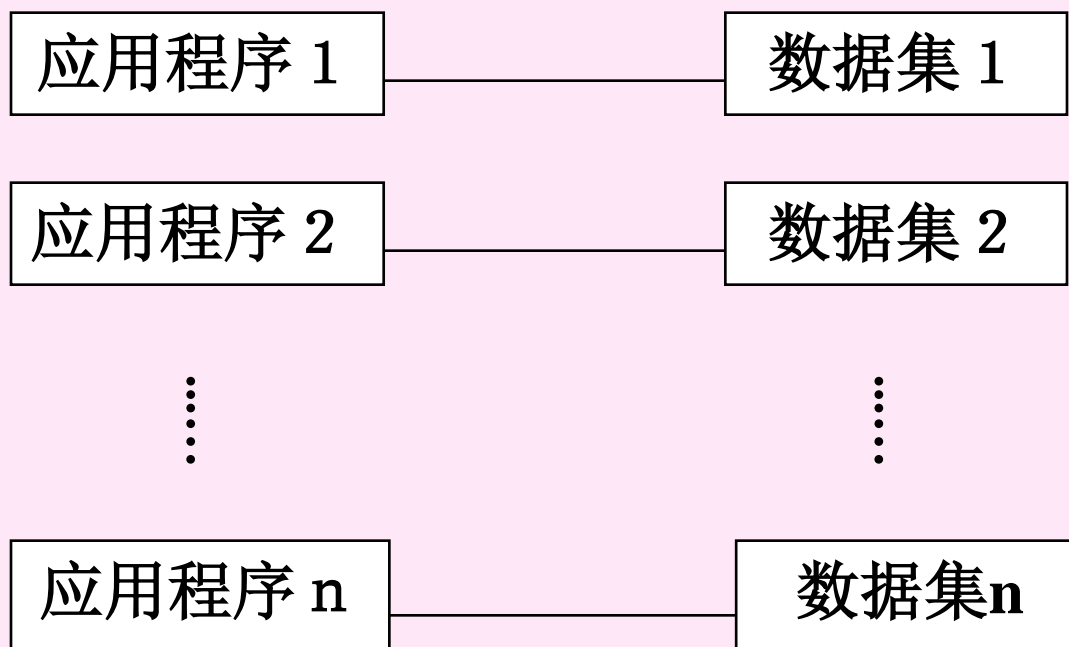


人工管理阶段特点

- 数据的管理者：用户（程序员），数据不保存
- 数据面向的对象：某一应用程序
- 数据的共享程度：无共享、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



应用程序与数据的对应关系(人工管理)





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、数据处理 |
| ■ 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 |

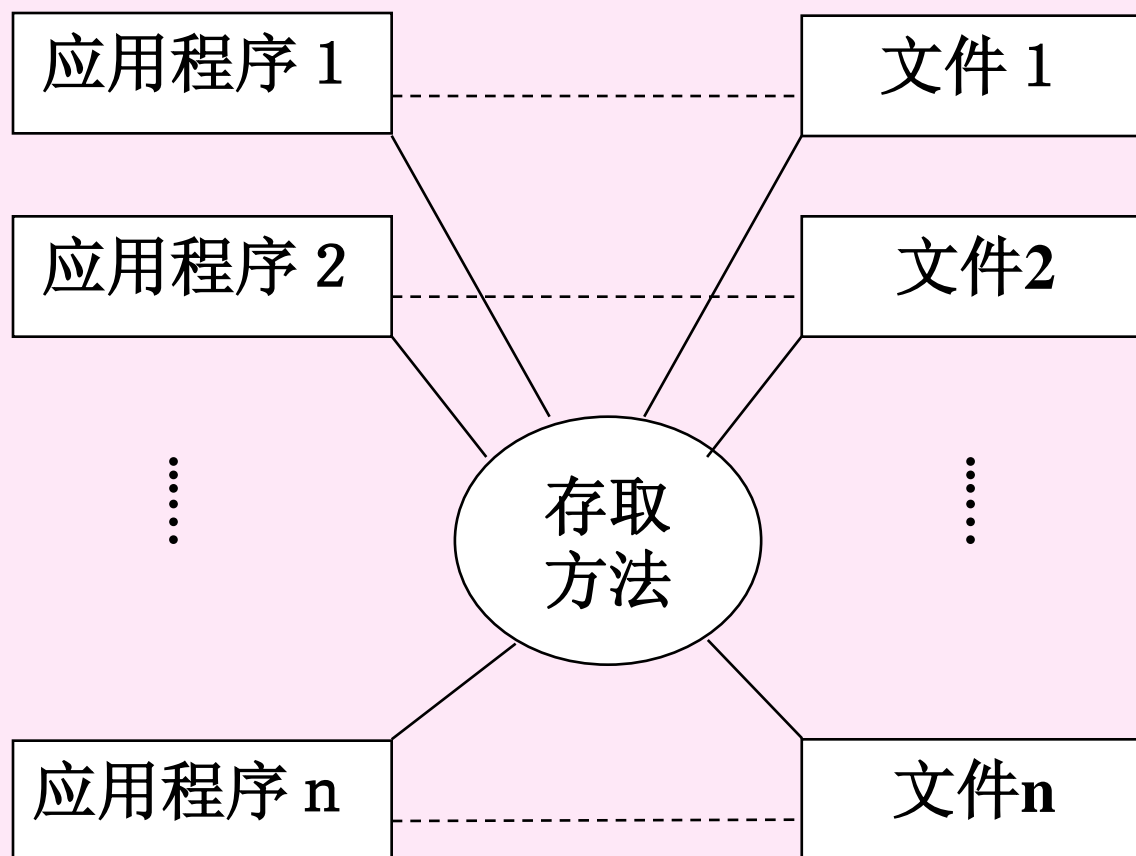


文件系统阶段特点

- 数据可长期保存
- 数据由文件系统管理
- 数据面向某一应用程序
- 数据的共享性差、冗余度大
- 数据记录内有结构, 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



应用程序与数据的对应关系(文件系统)





数据库系统阶段

■ 时期

- 60年代末以来

- 标志:

- IBM公司1968年成功研制层次数据管理系统(IMS)。
- 美国CODASYL (Conference On Data System Language, 数据系统语言协商会) 1971年公布的DBTG报告提出网络数据库系统 (CODASYL系统或DBTG系统)。
- IBM公司E.F.Codd 1970发表一系列论文, 奠定了关系数据库系统 (RDBMS) 理论基础



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



数据库系统阶段特点（1）

■ 结构化的数据模型

- 数据的结构用数据模型描述
- 数据库系统不仅描述数据本身，而且描述数据之间的联系
- 数据结构化是数据库与文件系统的根本区别



数据库系统阶段特点（2）

■ 较高的数据独立性

- 数据和程序彼此独立，数据存储结构的变化尽量不影响用户程序的使用
- 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。
- 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，用户程序也可以不变。



数据库系统阶段特点（3）

- 最低的冗余度
 - 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

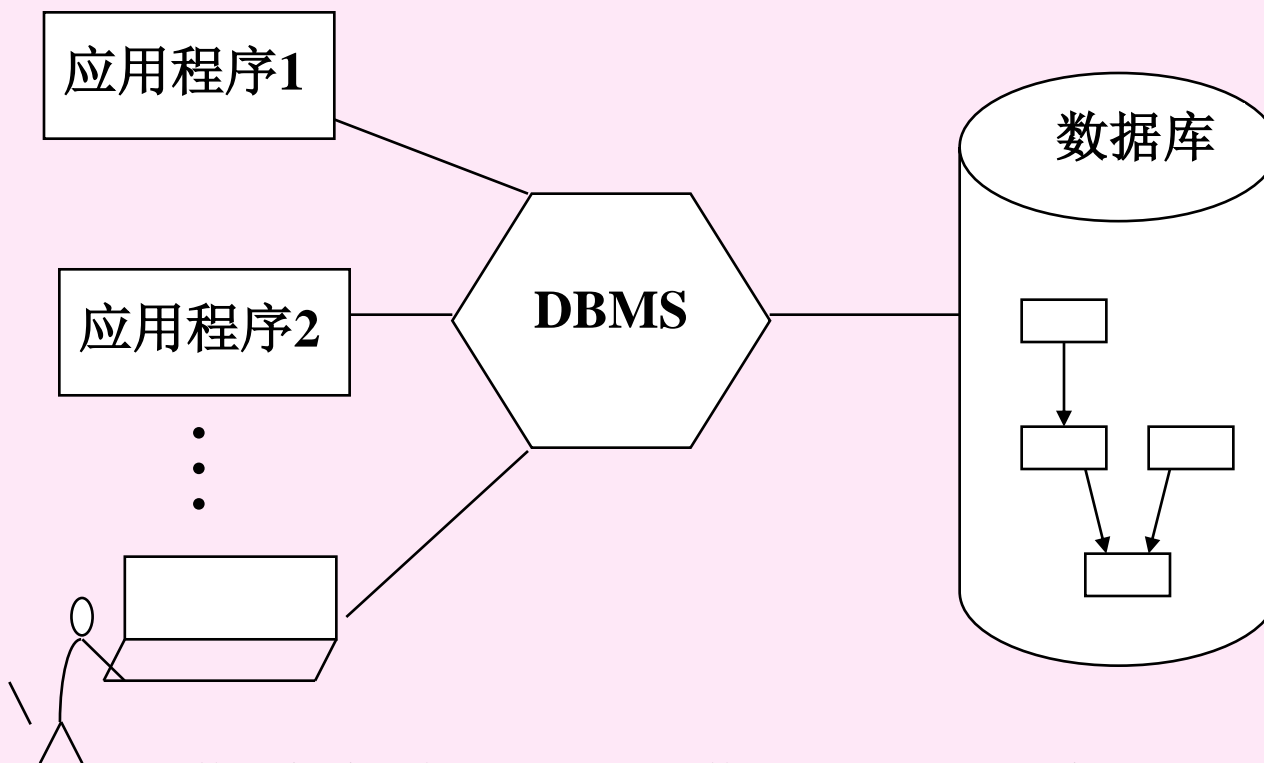


数据库系统阶段特点（4）

- DBMS（数据库管理系统）提供的数据库控制功能
 - (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：
 - 将数据控制在有效的范围内，或保证数据之间满足一定的关系。
 - (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。



应用程序与数据的对应关系(数据库系统)



数据库系统阶段应用程序与数据之间的对应关系



知识点5： 数据库系统的三级模式结构

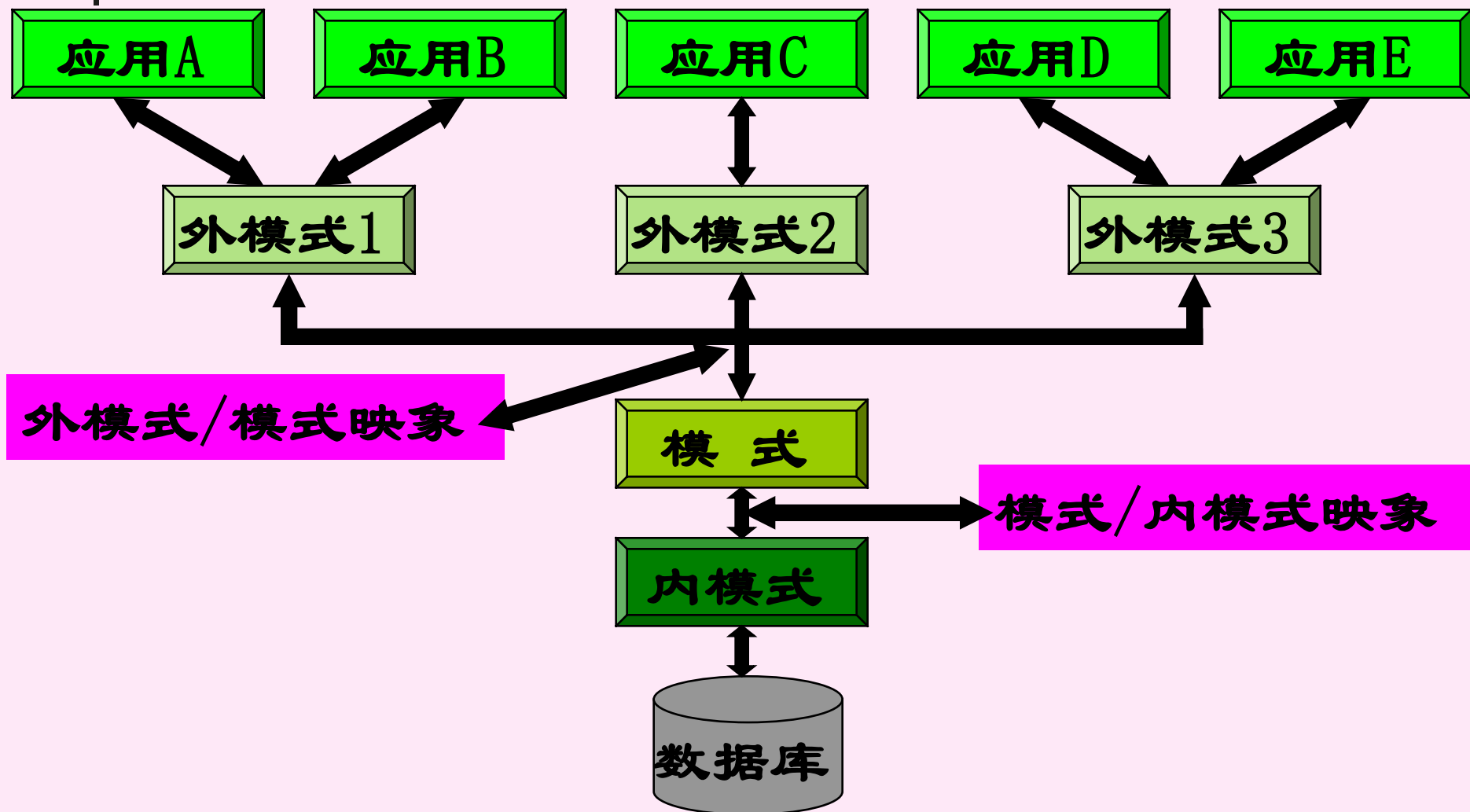


三级模式

- 外模式（External Schema）
- 模式（Schema）
- 内模式（Internal Schema）



数据库系统的三级模式结构（续）





模式 (Schema)

- 模式（也称逻辑模式、数据库模式、概念模式）
 - 数据库中全体数据的逻辑结构的描述
 - 所有用户的公共数据视图，综合了所有用户的需求
- 一个数据库只有一个模式
- 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



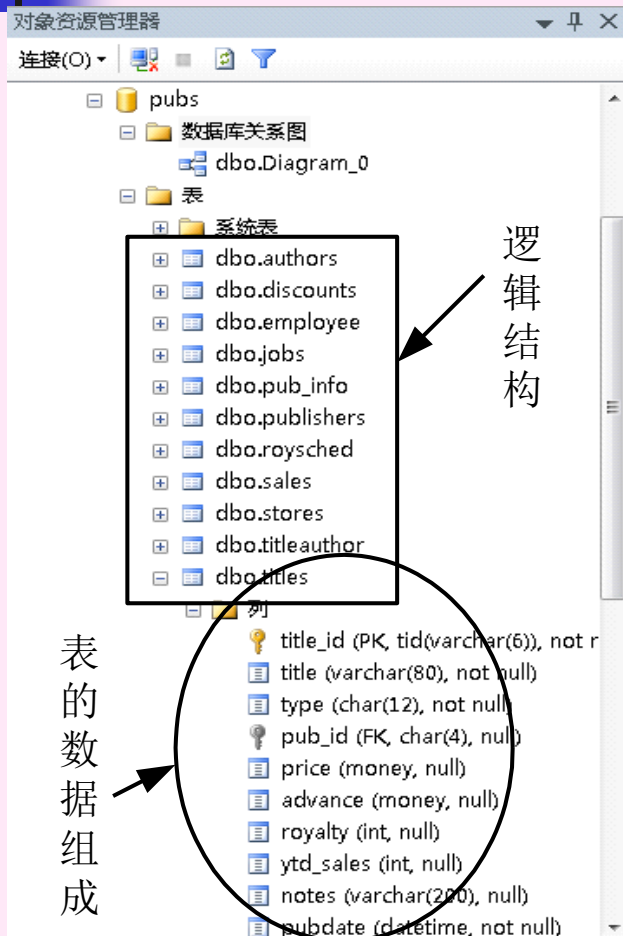
模式的定义

■ 模式的定义

- 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
- 有关数据的安全性、完整性要求
- 数据库系统提供数据定义语言DDL来表示这些内容



模式举例



逻辑结构

完整性定义





外模式 (External Schema)

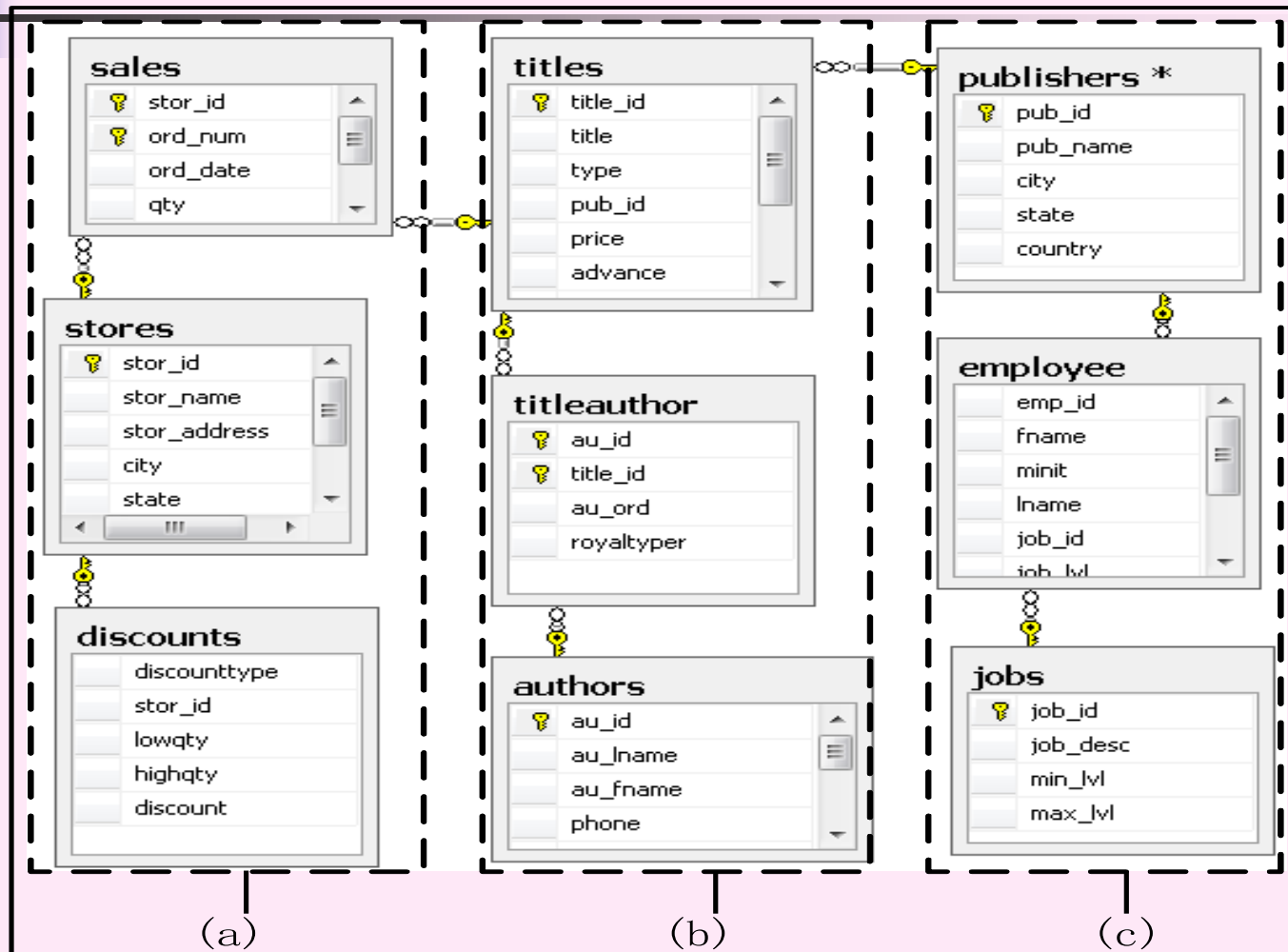
- 外模式 (也称子模式或用户模式)
 - 数据库用户 (包括应用程序员和最终用户) 使用的局部数据的逻辑结构的描述
 - 数据库用户的数据视图, 是与某一应用有关的数据的逻辑表示



外模式（续）

- 外模式的地位：介于模式与应用之间
 - 模式与外模式的关系：一对多
 - 外模式通常是模式的子集
 - 一个数据库可以有多个外模式。反映了不同的用户的应用需求、看待数据的方式、对数据保密的要求
 - 对模式中同一数据，在外模式中的结构、类型、长度、保密级别等都可以不同
 - 外模式与应用的关系：一对多
 - 同一外模式也可以为某一用户的多个应用系统所使用，
 - 但一个应用程序只能使用一个外模式。

外模式举例





外模式的用途

- 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



内模式 (Internal Schema)

- 内模式 (也称存储模式)
 - 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
 - 索引的组织方式
 - 数据是否压缩存储
 - 数据是否加密
- 一个数据库只有一个内模式



知识点6： 数据库系统的二级映像

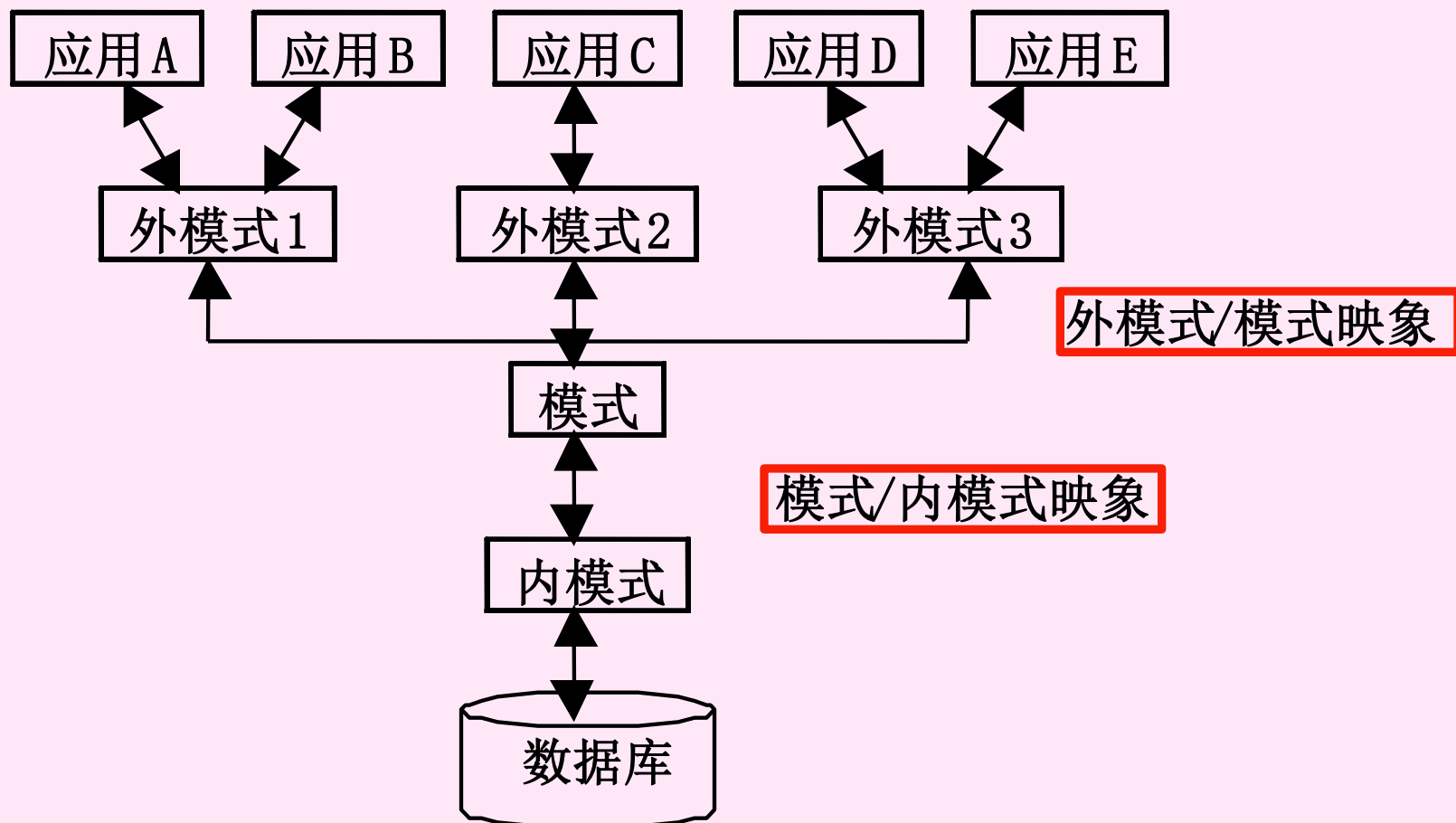


三级模式与二级映像

- 三级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



数据库系统的三级模式结构





外模式 / 模式映像

- 定义外模式与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



外模式 / 模式映象的用途

- 保证数据的逻辑独立性
 - 当模式改变时，数据库管理员修改有关的外模式 / 模式映象，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，简称数据的逻辑独立性。



模式 / 内模式映像

- 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 is 唯一的
- 该映像定义通常包含在模式描述中



模式 / 内模式映象的用途

■ 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，简称数据的物理独立性。



三级模式结构总结 (1)

■ 模式

- 是数据库的中心与关键
- 独立于数据库的其它层次
- 设计数据库模式结构时应首先确定数据库的逻辑模式



三级模式结构总结（2）

■ 内模式

- 依赖于全局逻辑结构，但独立于数据库的用户视图即外模式，也独立于具体的存储设备。
- 它将全局逻辑结构中所定义的数据结构及其联系按照一定的物理存储策略进行组织，以达到较好的时间与空间效率。



三级模式结构总结（3）

■ 外模式

- 面向具体的应用程序，定义在逻辑模式之上，但独立于存储模式和存储设备
- 设计外模式时应充分考虑到应用的扩充性。当应用需求发生较大变化，相应外模式不能满足要求时，该外模式就得做相应改动



三级模式结构总结（4）

■ 二级映像

- 保证了数据库外模式的稳定性，从而从底层保证了应用程序的稳定性，除非应用需求本身发生变化，否则应用程序一般不需要修改。