



第1章 数据库系统概述

北京理工大学 计算机学院

张文耀

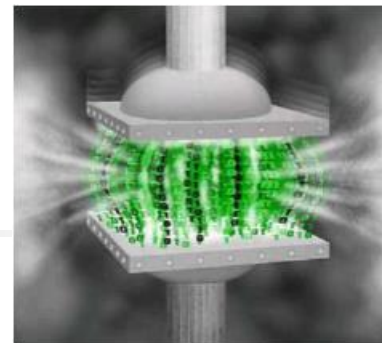
zhwenyao@bit.edu.cn



主要内容

- 数据
- 数据管理
- 数据库
- 数据库管理系统
- 数据库系统
- 数据库系统用户

1. 数据管理技术



■ 数据（Data）

- 描述事物的符号记录
（表示信息的符号，信息的载体）
- 形式多样：数字、文字、图形、图像、声音.....

■ 数据处理

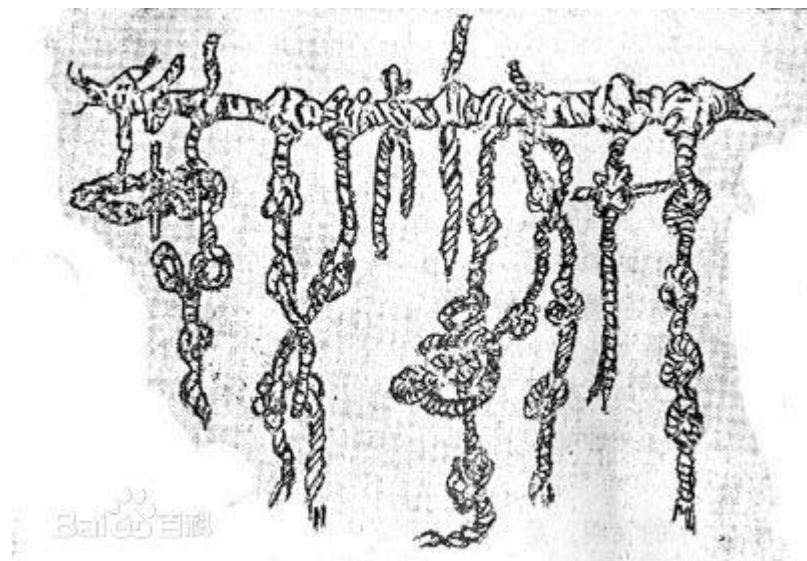
- 收集、加工、应用、储存、传播等过程

■ 数据管理

- 对数据进行分类、组织、编码、存储、检索和维护等；
- 是数据处理的中心问题
- 怎么管？
——数据管理技术

■ 数据的语义

- 数据的含义
- 数据的解释（对数据含义的说明）
- 数据与其语义是不可分的





■ 数据及其语义示例

■ 学生记录

（李明，男，**199505**，江苏南京市，计算机系，**2013**）

■ 语义

学生姓名、性别、出生年月、出生地、所在院系、入学时间

■ 解释

李明是个大学生，**1995**年**5**月出生，江苏南京市人，**2013**年考入计算机系。



■ 数据管理技术

- 人工管理(**20世纪40年代中-- 50年代中**)
- 文件系统(**50年代末-- 60年代中**)
- 数据库系统(**60年代末--现在**)
- **大数据 (Big Data, Now)**

■ ...

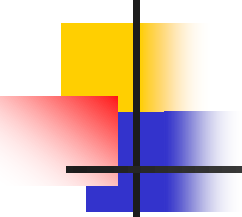
■ 数据管理技术的发展动力

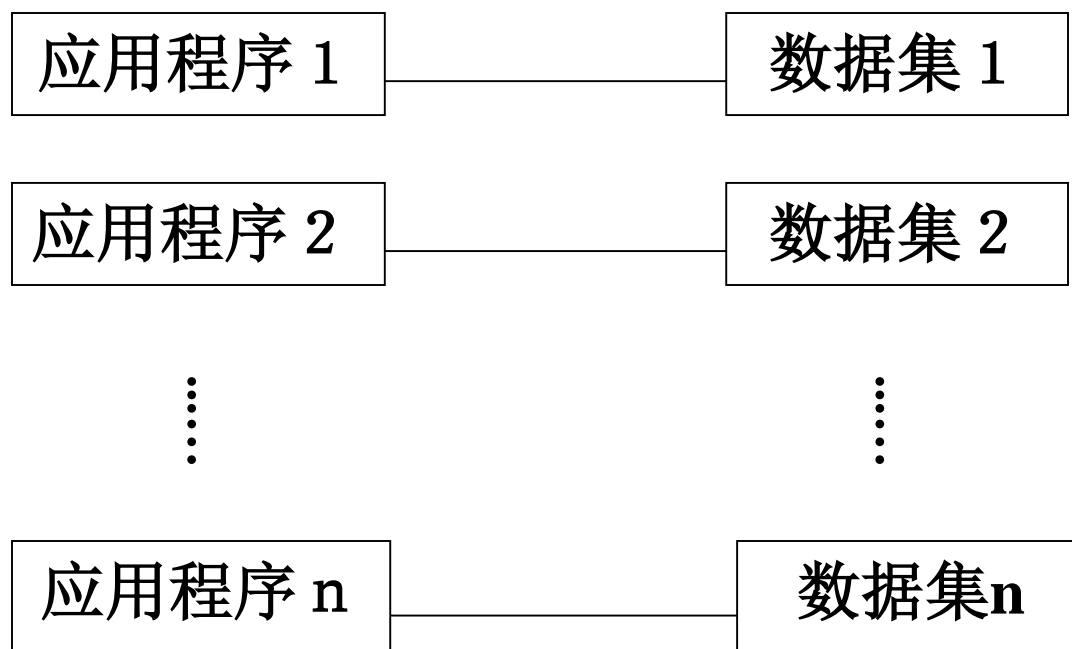
- 应用需求的推动
- 计算机硬件的发展
- 计算机软件的发展



人工管理阶段

- **20世纪40年代中--50年代中**
- **产生背景**
 - 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
- **特点**
 - 数据不保存
 - 没有专用的软件管理数据
 - 应用程序完全依赖于数据
 - 数据不能共享

- 
- 数据和应用程序的关系
 - 应用程序与数据的存储、存取方式密切相关——数据依赖



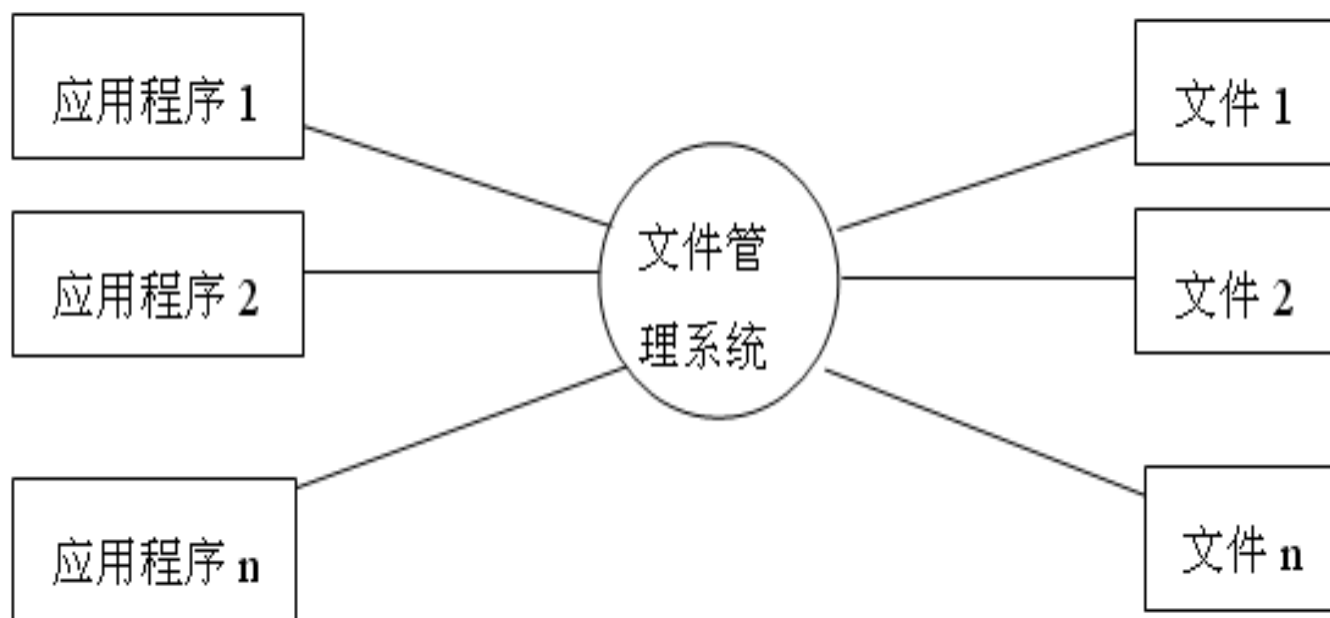


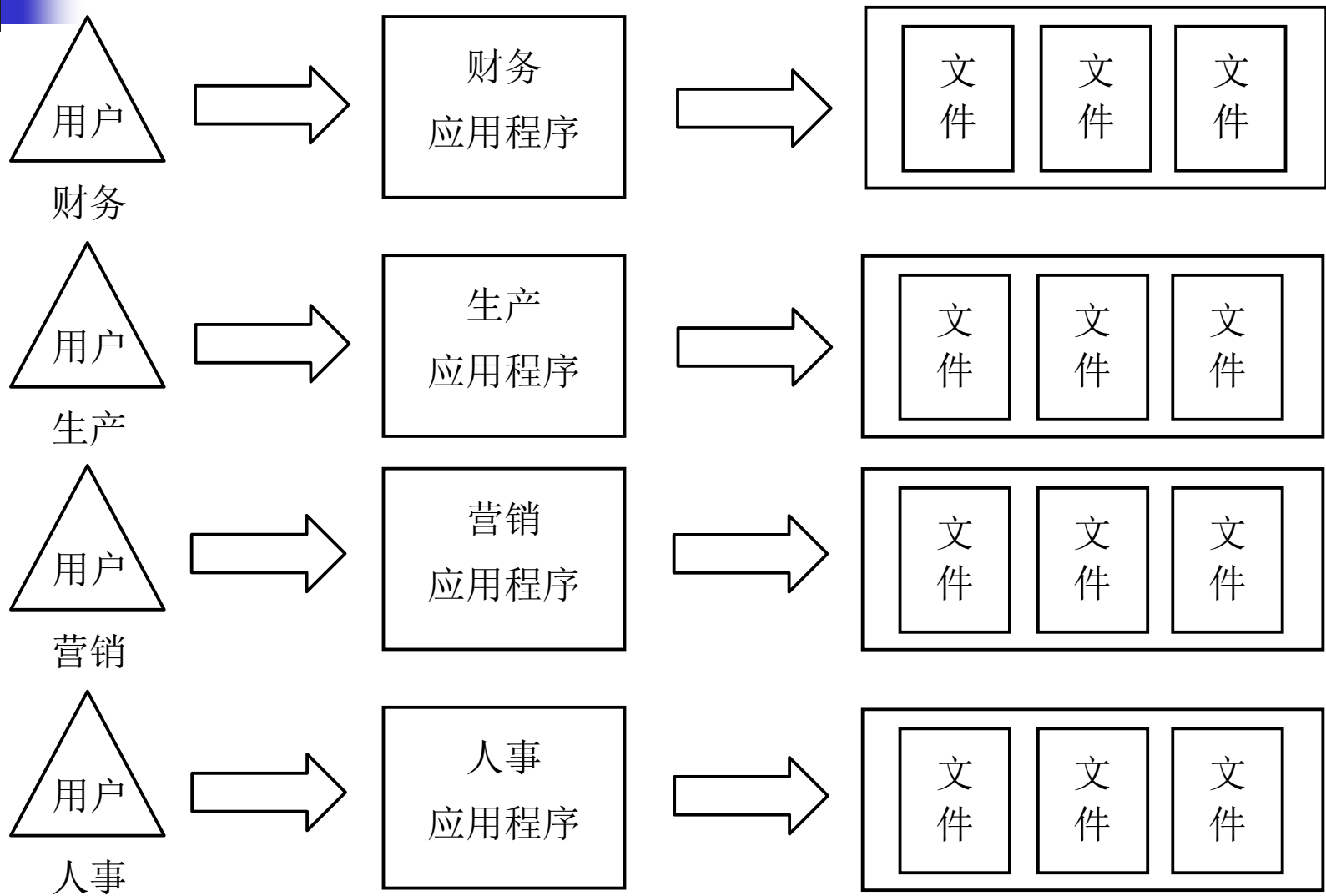
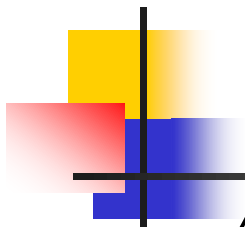
文件系统阶段

- **20世纪50年代末--60年代中**
- **产生背景**
 - 应用需求 科学计算、管理
 - 硬件水平 磁盘、磁鼓等直接存取设备
 - 软件水平 有文件管理系统
 - 处理方式 联机实时处理、批处理
- **特点**
 - 数据可以长期保存
 - 有专门的文件系统软件管理数据
 - 数据是面向应用的、分散的
 - 数据冗余度大
 - 缺乏对数据统一的控制机制

■ 数据和应用程序的关系

- 应用程序和数据之间仍然存在依赖关系 (**m:n**)
- 数据改变应用程序也得改变
- 文件内有结构，文件间是独立的，数据整体无结构





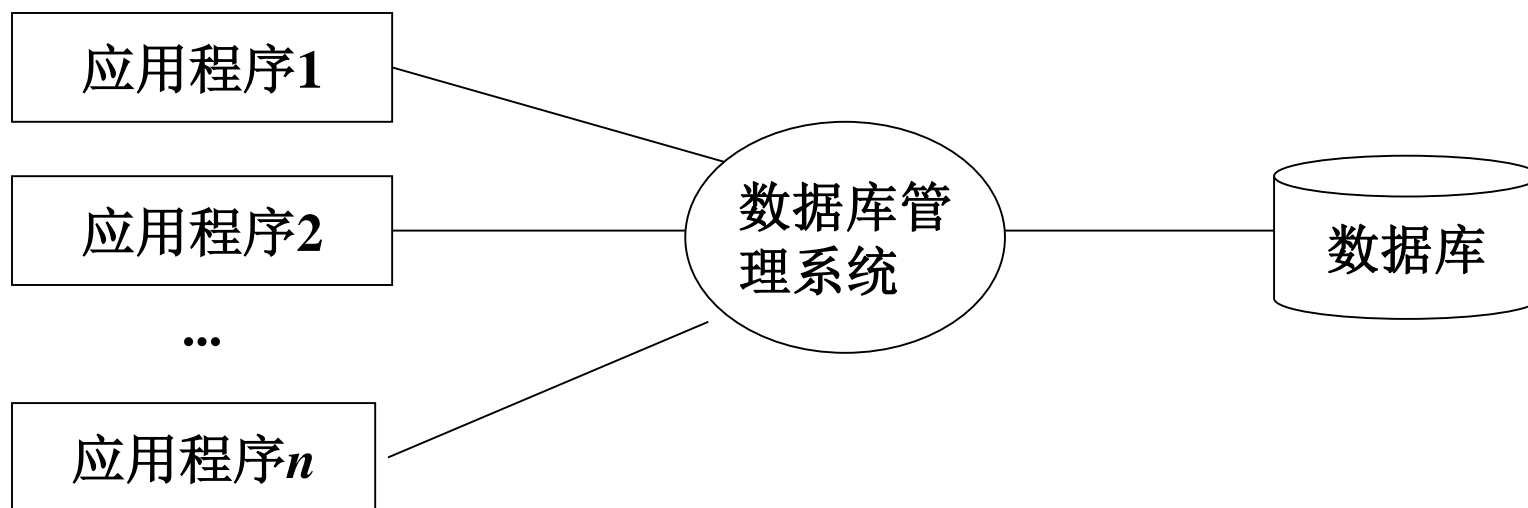


数据库系统阶段

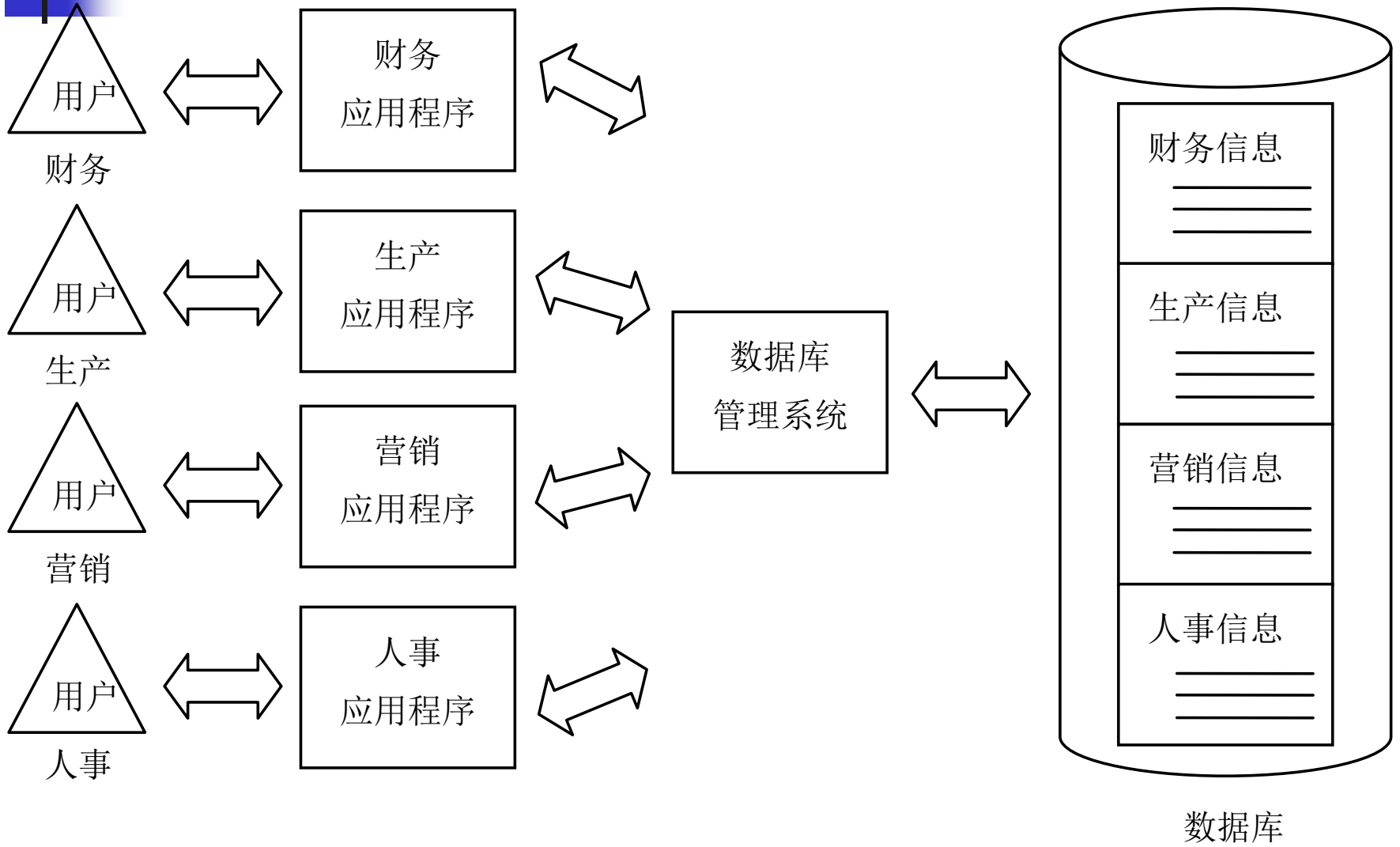
- 20世纪60年代末以来
- 产生背景
 - 应用背景 大规模数据管理
 - 硬件背景 大容量快速存取的磁盘
 - 软件背景 商品化的数据库管理系统
 - 处理方式 联机实时处理，分布处理，批处理
- 三件大事，标志着数据库时代的到来
 - 1968年，IBM研制了世界上第一个商品化的数据库管理系统——**层次数据库系统IMS**；
 - 1969年，美国数据库系统语言研究会下属的数据库任务组公布了**基于网状模型的DBTG报告**；
 - 1970年，IBM的E.F.Codd发表了题为《大型共享数据库数据的关系模型》等一系列**关系数据库论文**，奠定了关系数据库的理论基础

■ 数据和应用程序的关系

- 数据以一定方式组织起来，由一个软件共同管理，作为应用程序与数据的接口。



文件系统到数据库系统，标志着数据管理技术的飞跃！





■ 特点

■ 数据结构化

- 整体从无序到有序

■ 数据独立性高

- 物理独立性：数据库物理结构的改变，不影响数据的逻辑结构，不影响应用程序。
- 逻辑独立性：数据库中的逻辑数据结构发生改变时，应用程序可以无需修改。

■ 减少数据冗余

- 集中管理，统一组织、定义和存储

■ 数据共享

- 数据库发展的主要原因

■ 统一的数据管理、控制和保护功能

- 安全性、一致性、有效性、并发控制、故障恢复等

数据管理示例

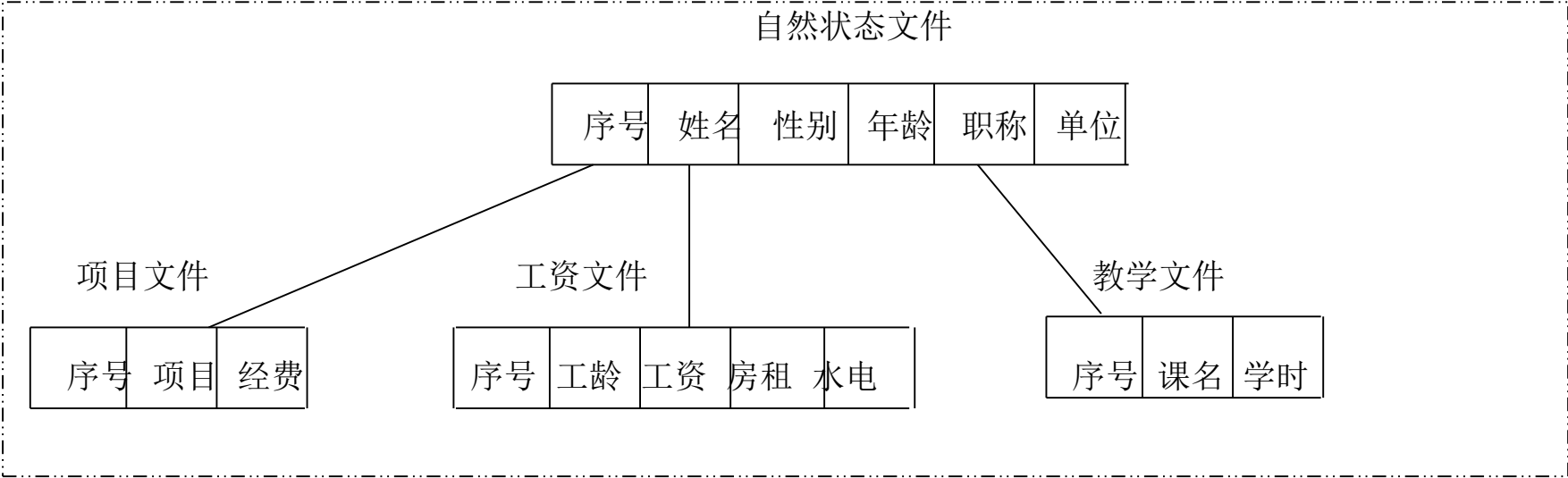
第一层

人事文件	教学文件
序号 姓名 性别 年龄 职称 单位 工龄 工资	序号 姓名 职称 年龄 单位 课程名 学时
科研文件	工资文件
序号 姓名 职称 年龄 单位 科研名称 经费	序号 姓名 单位 工资 工龄补贴 房租 水电

第二层

基本文件
序号 姓名 性别 年龄 职称 单位 工龄 工资 科研 经费 课名 学时 房租 水电

第三层





2.什么是数据库

- 数据库（**Data Base, DB, Database**）
 - 存放数据的仓库。
 - 存放在介质上的相关数据的集合。
 - 长期储存在计算机内、有组织的、可共享的大量数据的集合。
 - 长期存储在计算机内、有组织的数据集合，它根据数据间的联系组织在一起，具有较高的数据独立性，较少数据冗余，能够为各种用户共享。
 -

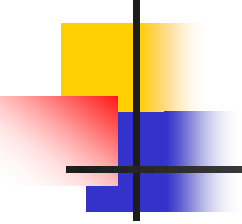


■ 数据库的特征

- 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

学生登记表

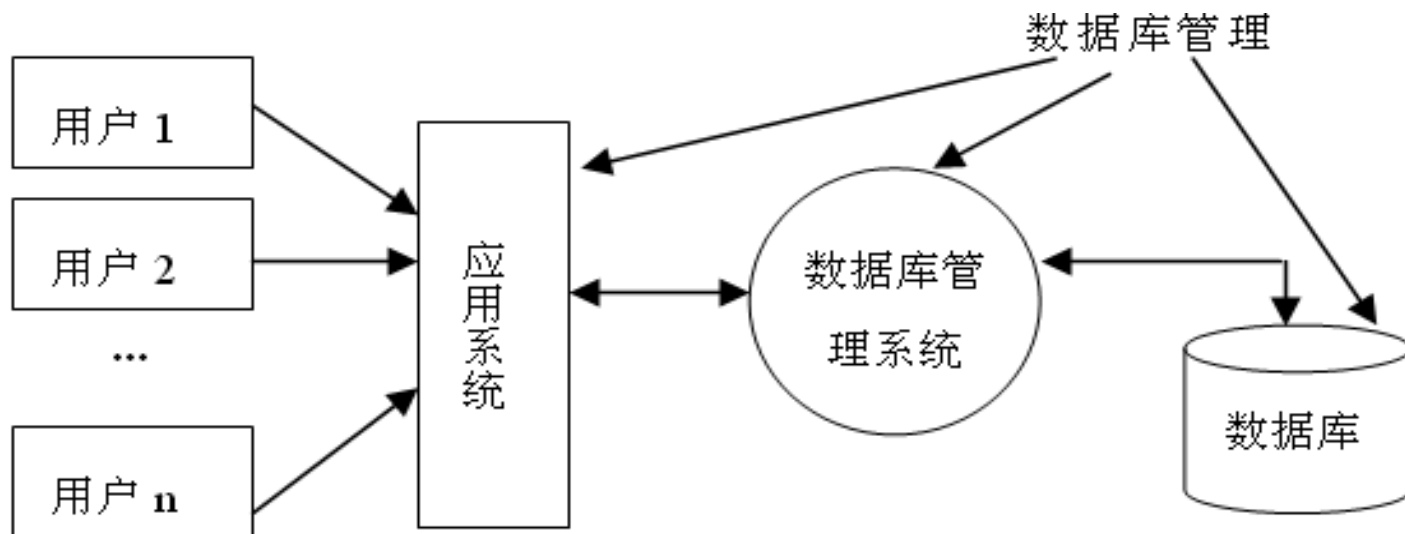
学 号	姓 名	年 令	性 别	系 名	年 级
95004	王小明	19	女	社会学	95
95006	黄大鹏	20	男	商品学	95
95008	张文斌	18	女	法律学	95
...

- 
- 数据库管理系统（**Database Management System, DBMS**）
 - 统一管理数据库的软件系统；
 - 位于用户与操作系统之间的一层数据管理软件；
 - 负责数据库的管理和维护，具有数据定义、数据操纵、运行管理和维护等功能；
 - 是数据库系统的核心。

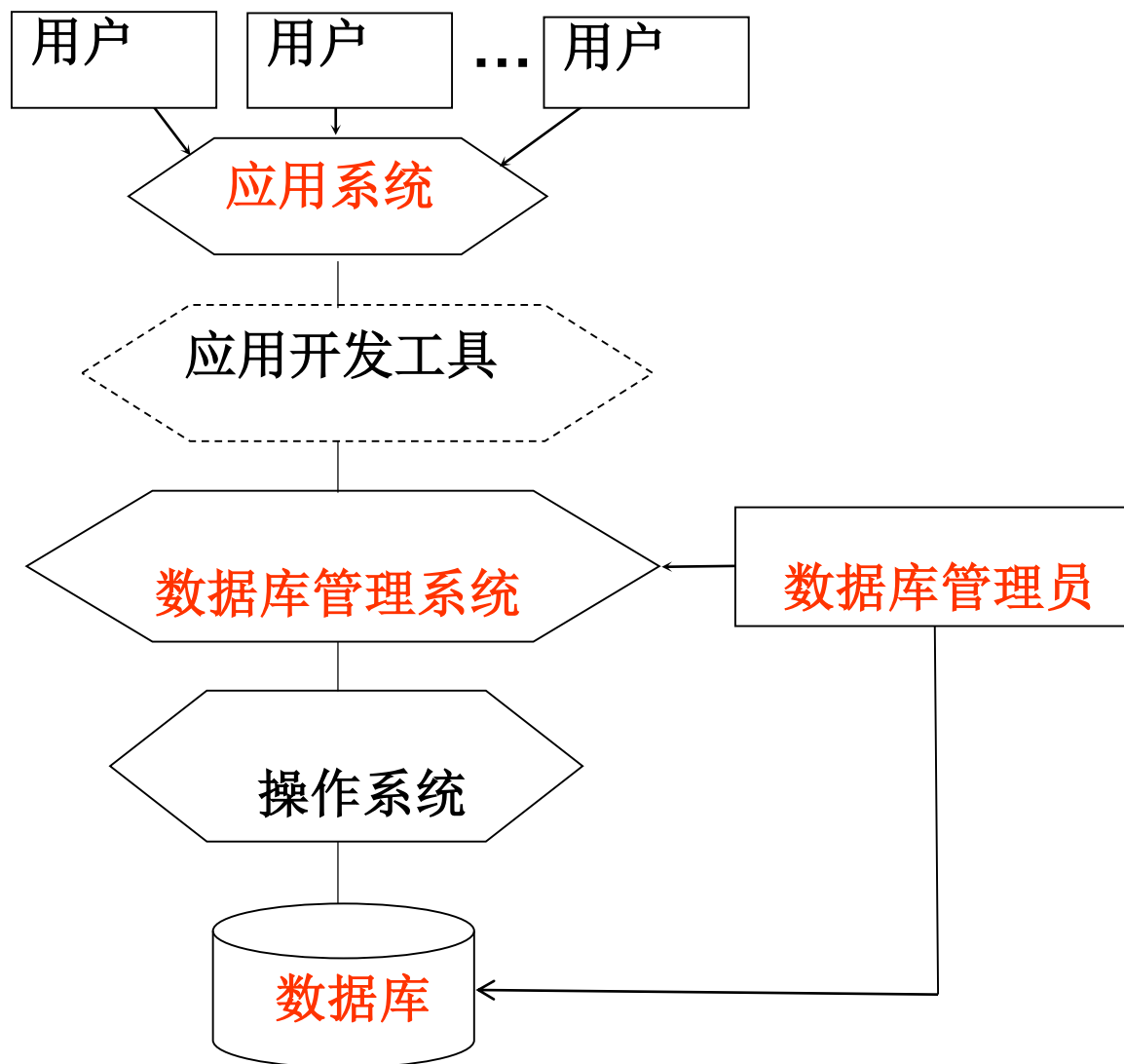
■ 数据库系统（**Database System, DBS**）

- 对数据提供存储、管理和应用的计算机系统，包括计算机硬件、软件、用户和相关管理人员（**DBA**）。
- 是指在计算机系统中引入数据库后的系统构成。

■ 数据库系统的组成

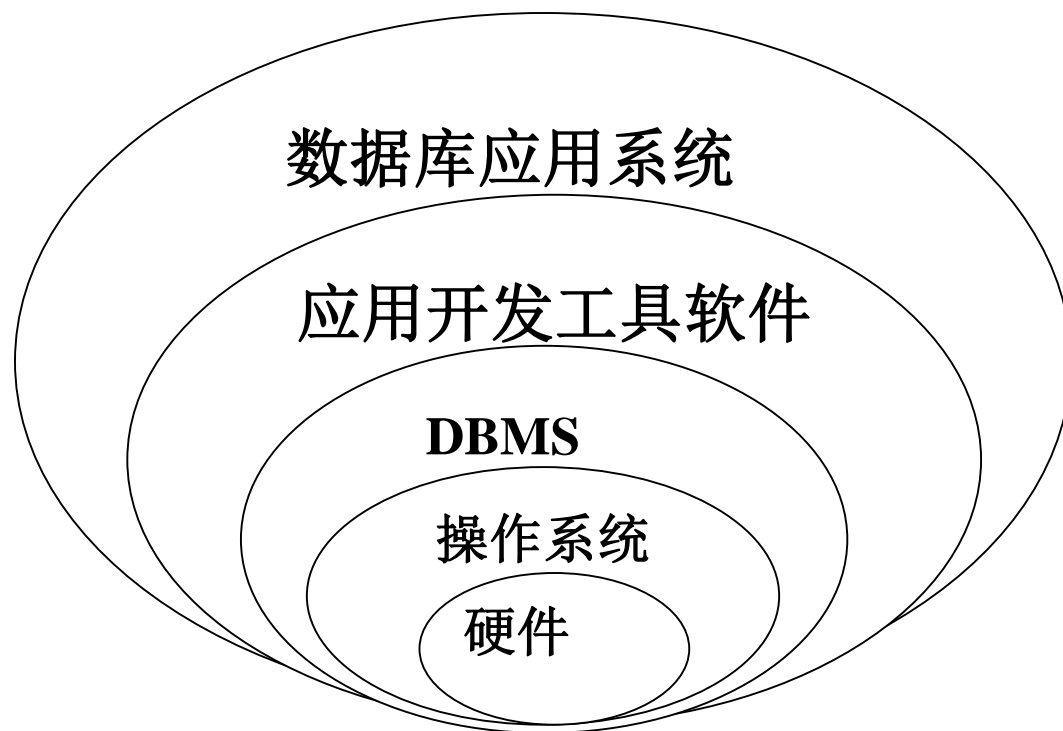


DBS=DB+DBMS+...





■ 数据库管理系统在计算机系统的位置图





■ 数据库的地位

- 数据库技术产生于**20世纪60**年代末，是数据管理的**重要技术**，是计算机科学的重要分支；
- 数据库技术是信息系统的核心和基础，它的出现极大地促进了计算机应用向各行各业的渗透；
- 数据库的广泛应用也促进了数据库技术的发展，使其成为计算机科学中发展最快的一个领域；
- 数据库的建设规模、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。



3.数据模型

- 数据库中的数据是按照一定的**数据模型**组织、描述和储存的。
- 模型
 - 对现实世界中某个对象特征的模拟和抽象
 - 对客观事物形式化的抽象描述
 - 简化问题，便于处理
 - 航模飞机



■ 数据模型（**Data Model**）

- 信息领域采用的模型
- 将现实世界的各种事物以及事物之间的联系，表示为数据以及数据之间的联系
- 是对现实世界数据特征的抽象和模拟
- 用来描述数据、组织数据和操作数据
- 是数据库系统的核心和基础



■ 数据模型的描述（数据模型三要素）

1. 数据结构

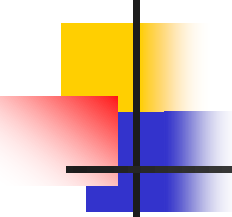
- 对数据静态特性的描述
- 包括应用所涉及的对象、对象具有的特征、对象间的联系

2. 数据操作

- 对数据动态特性的描述。
- 对数据库中对象实例执行的一组操作，包括检索、插入、删除、修改等

3. 数据的完整性约束

- 对数据静态和动态特性的限定
- 反映了数据间的制约和依存关系
- 一般所指的不同数据模型主要是由数据结构来表征。

- 
- 数据模型应满足三方面的基本要求
 - 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

针对不同的使用对象和应用目的，采用不同的数据模型。



■ 数据模型的分类

■ 独立于计算机的

- 概念模型（信息模型）
- 实体-联系模型、语义数据模型
- 描述特定范围内的信息结构
- 与信息在计算机中的表示无关，与**DBMS**也无关
- 比较稳定，由现实需求决定

■ 直接面向计算机的

- 基本数据模型（数据模型）
- 按照计算机的观点对数据建模，与具体的**DBMS**有关
- 层次模型、网状模型、关系模型
面向对象数据模型、对象关系数据模型、半结构化数据模型等



■ （基本）数据模型的层次

■ 逻辑数据模型

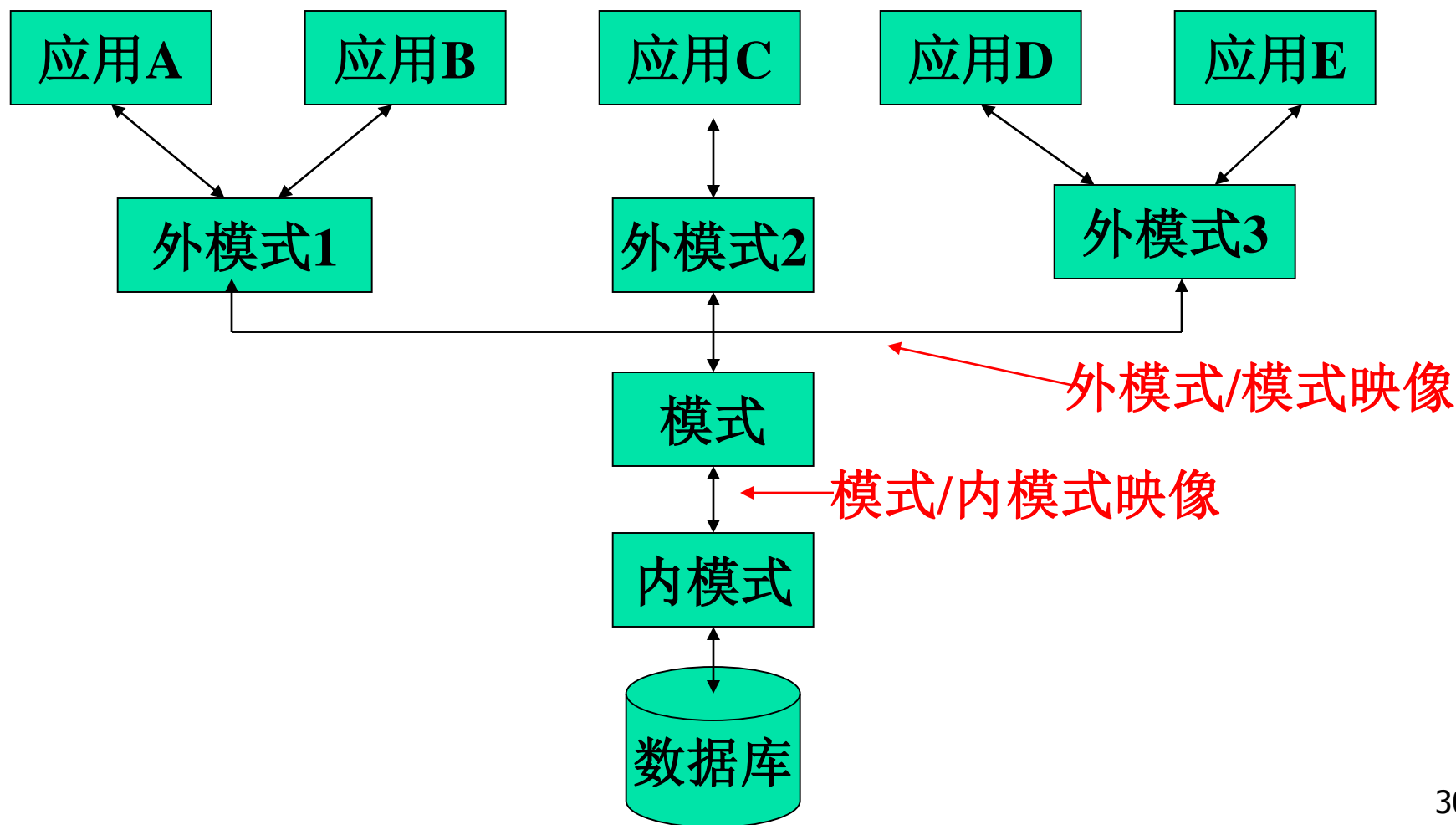
- 用户所看到的数据模型；
- 仅表示数据的逻辑结构；
- 现有**DBMS**都是以其所支持的逻辑数据模型来分类的。

■ 物理数据模型

- 表示数据存储结构和存取方法的数据模型；
- 反映逻辑数据模型在物理存储器上的具体实现；
- 与**DBMS**有关，还与操作系统和硬件设备有关。

4.数据库系统结构

- 体系结构——三级模式结构（二级映像）





■ 模式（**Schema**）

- 概念模式、逻辑模式
- 是全体数据的逻辑结构和特征的描述
- 独立于应用程序和物理存储
- 一个数据库对应一个模式
- 是数据库的总框架，所有数据都按照这一模式进行装配
- 设计数据库模式结构时应首先确定数据库的逻辑模式
- 数据库系统提供模式描述语言 (**Data Description Language, DDL**)



■ 外模式（**External Schema**）

- **子模式（Subschema）**：模式的子集
- 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。
- 是数据库用户的数据**视图**，是对用户数据结构的描述。
- 应用程序根据外模式对数据的描述存取数据。
- 一个数据库可以有多个外模式；一个应用程序只能使用一个外模式；一个外模式可以对应多个应用程序。
- 外模式对数据的描述可以不同于模式，但受模式的制约。
- 数据库系统提供外模式描述语言（**Subschema DDL**）



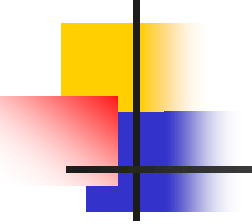
■ 内模式（**Internal Schema**）

- 也称存储模式，是数据库的存储结构和存取方法的描述，是数据在数据库内部的表示方法。
- 它规定数据在存储介质上的物理组织方式、记录寻址方式，定义物理存储块的大小、溢出处理方法等内容。
- 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。
- 与模式对应，一个数据库只有一个内模式。
- 内模式描述语言（内模式**DDL**）



■ 三级模式结构的意义

- 内模式是整个数据库实际存储的表示；
概念模式是整个数据库实际存储的抽象表示；
外模式是概念模式的某一部分的抽象表示。
- 三级模式结构将数据库的全局逻辑结构、用户的局部逻辑结构和物理组织结构分开。
- 用户能逻辑抽象地处理数据，不必关心数据在计算机中的内部表示。
- 数据的存储也不必考虑用户对数据的使用。
- 提供了三种级别的数据抽象：视图级、概念级和物理级

- 
- 三级数据抽象的实现依赖于二级映像
 - 外模式/模式映像
定义局部数据逻辑结构与全局逻辑结构之间的对应关系
 - 模式/内模式映像
定义全局数据逻辑结构和物理数据存储间的对应关系
 - 三级结构和二级映像提供了数据库的**数据独立性**。
 - **数据的逻辑独立性**
当模式结构改变时，只需修改外模式与模式间的对应关系，外模式中的局部逻辑结构和应用程序不必修改。
 - **数据的物理独立性**
当数据库的物理存储结构改变时，仅需修改模式与内模式间的映像关系，模式及其以上的内容可以保持不变。

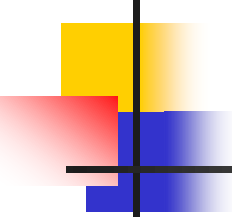


5.数据库管理系统

- **DBMS**的主要功能

- 数据库的定义功能

- **DBMS**提供数据定义语言（**DDL**），定义数据库的数据结构，包括模式定义语言、外模式定义语言和内模式定义语言。
- **DBMS**一般仅提供一种或两种**DDL**。
- 数据定义语言描述的模式成为**源模式**，需要借助编译程序翻译成机器代码形式的**目标模式**，供**DBMS**使用。



■ 数据组织、存储和管理

- 分类组织、存储和管理各种数据
- 确定组织数据的文件结构和存取方式
- 实现数据之间的联系
- 提供多种存取方法提高存取效率

■ 数据操纵功能

- **DBMS**提供数据操纵语言(**Data Manipulation Language, DML**), 实现对数据库的插入、删除、修改和检索查询等基本操作。
- **DML**分为自含型**DML**和宿主型**DML**。



■ 数据库的运行管理和控制功能

- 数据库在建立、运行和维护时由数据库管理系统统一管理
- 数据库的控制功能
 - 数据库的安全性控制
 - 数据库的完整性控制
 - 数据库的并发控制
 - 数据库的故障处理与恢复



■ 数据库的维护功能

DBMS通过一些实用例程实现对数据库的维护

- 转储程序
- 数据装入程序
- 统计分析程序
- 重组程序
- 以及无用数据的收集及空间的再分配等程序

■ 数据库的其他功能

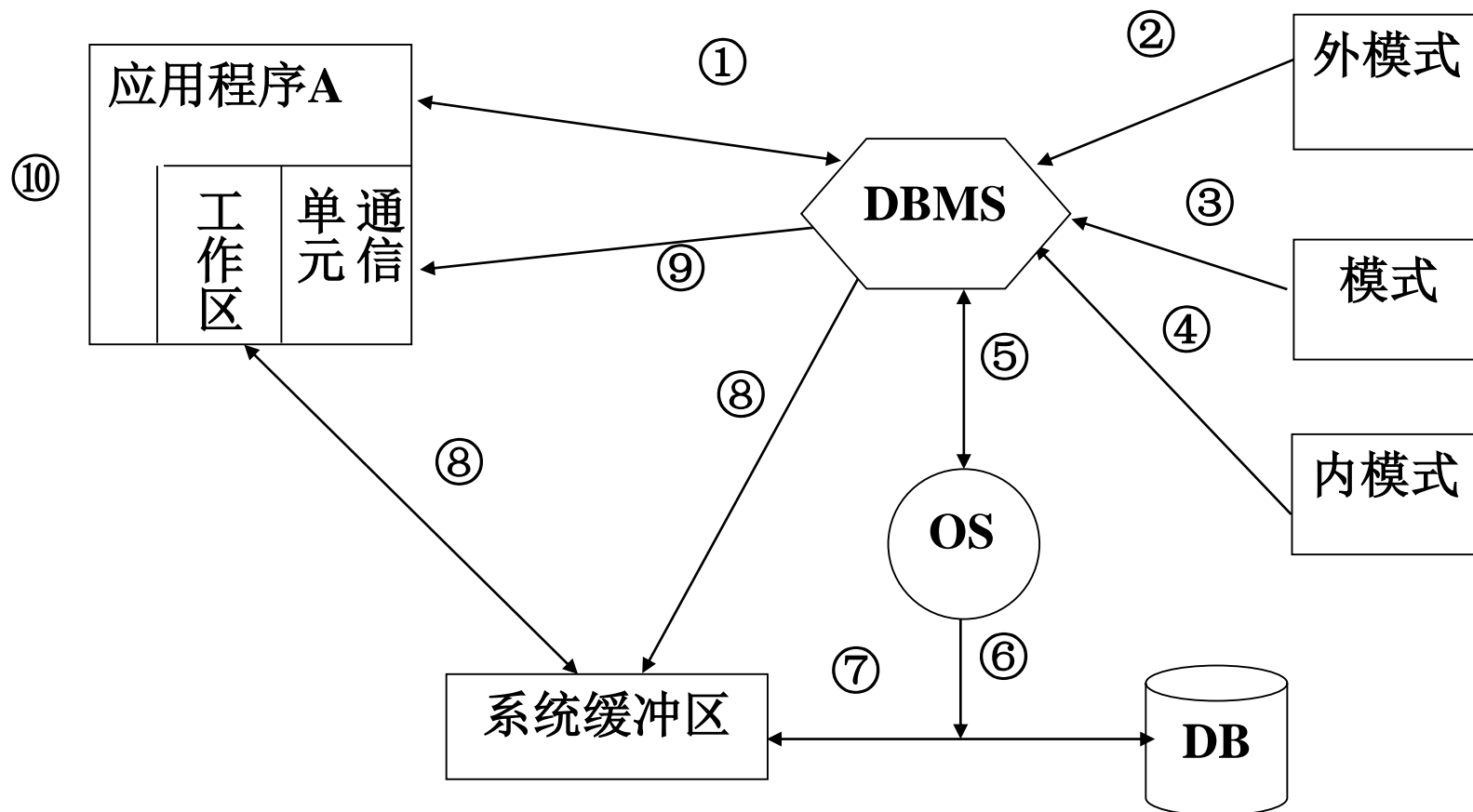
- 与其它系统的**通信功能**
- 系统之间的**数据转换功能**
- 异构数据库之间的互访和**互操作功能**



■ DBMS的组成

1. 系统主控程序
2. 存取控制程序
3. 并发控制程序
4. 数据有效性检查程序
5. 数据保护程序
6. 查询处理程序
7. 数据更新程序
8. 目录管理程序
9. 通讯程序
10.

■ DBMS的工作过程(参见电子教材)



- ① 应用程序A通过DML命令向DBMS发读请求，并提供读取记录参数。如记录名，关键字值等。
- ② DBMS根据应用程序A对应的外模式中的信息，检查用户权限、决定是否接受应用程序A的读请求。
- ③ 如果是合法用户，则查看模式，根据模式与外模式间数据的对应关系，确定需要读取的逻辑数据记录。
- ④ DBMS查看内模式，根据模式与内模式间的映像关系确定需要读取哪些物理记录。
- ⑤ 向操作系统发读取记录的命令。
- ⑥ 操作系统执行该命令，控制存储设备从数据库中读取物理记录数据。
- ⑦ 在操作系统控制下将读出的记录送入系统缓冲区，并通知DBMS数据已读出。
- ⑧ DBMS比较模式和外模式，从系统缓冲区中得到所需的逻辑记录，经过必要的数据库变换后，将数据送入用户工作区。
- ⑨ DBMS设置通信单元以向应用程序A发送读命令执行情况的状态信息。
- ⑩ 应用程序A根据状态信息确定DML命令执行是否成功，若成功则对工作区中读出的数据进行相应处理，否则按出错处理。



■ 主要的DBMS

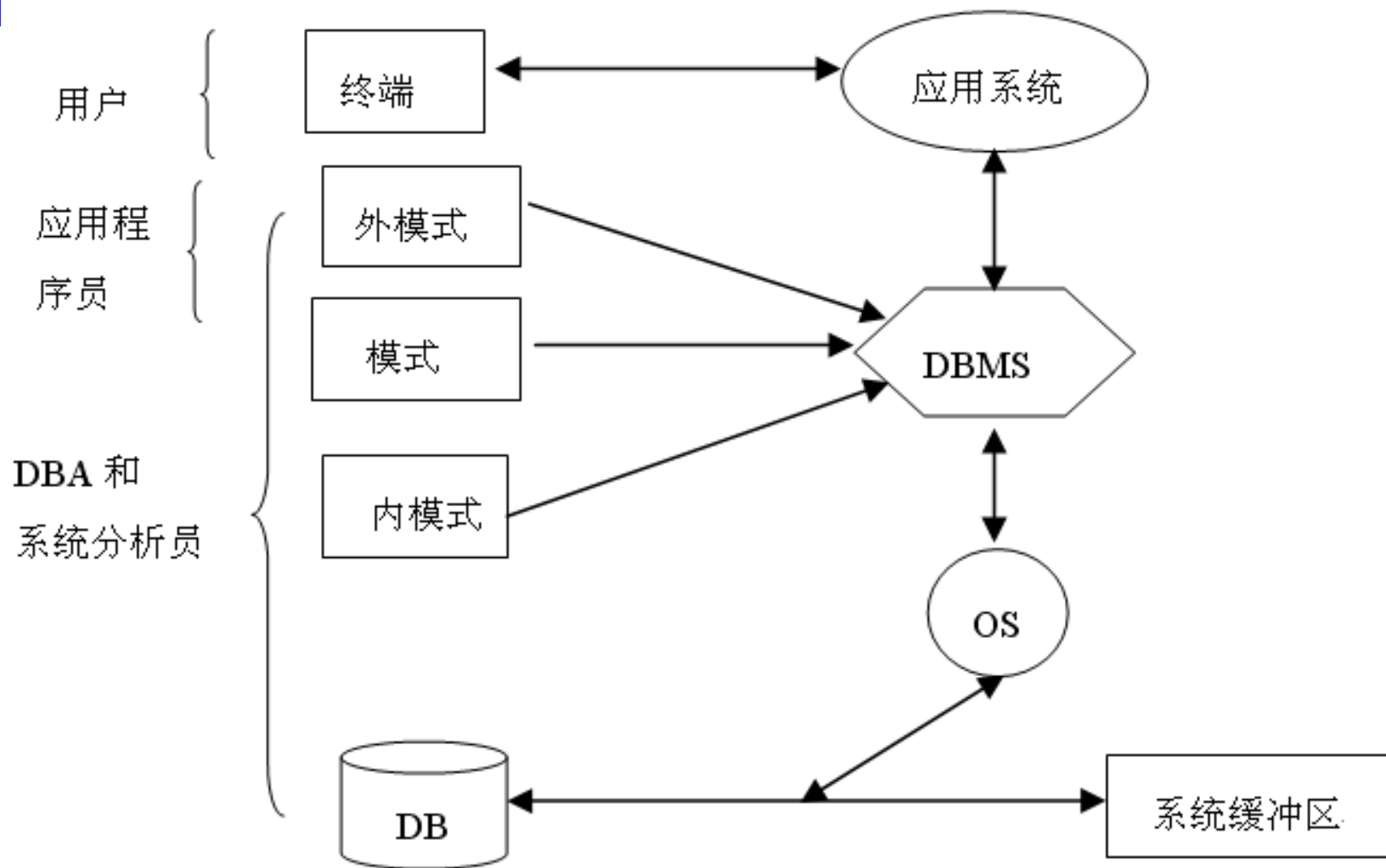
- IBM DB2
- Oracle
- SQL Server
- Sybase
- Informix
- Teradata
- PostgreSQL
- MySQL
- ...





6.数据库系统的不同视图

- 一个数据库系统的设计、建立、使用和维护涉及许多人，这些人员可以分为四类：
 - 数据库管理员
 - 系统分析员
 - 应用程序员
 - 用户
- 不同人员职责不同，所涉及数据的抽象级别也不同，因而具有不同的数据库视图。





■ 数据库管理员(**DBA, Database Administrator**)

- 决定数据库的信息内容, 参与数据库设计的全过程;
- 决定数据库的存储结构和存取策略;
- 定义数据的安全性要求和完整性约束条件
- 建立数据库
- 监控数据库的运行
 - 周期性转储数据库(数据文件/日志文件)、系统故障恢复、介质故障恢复、监视审计文件
- 数据库的改进(性能监控和调优)和重组
- 数据库重构



■ 系统分析员

- 负责应用系统的需求分析和规范说明
- 与用户及**DBA**协商，确定系统的硬软件配置
- 参与数据库各级模式的概要设计

■ 数据库设计人员（常由数据管理员担任）

- 参加用户需求调查和系统分析
- 确定数据库中的数据
- 设计数据库各级模式

■ 应用程序员

- 负责应用系统的程序设计
- 可以同**DBA**和系统分析员一起完成子模式的设计



■ 用户（最终用户）

- 通过应用系统使用数据库
- 偶然用户：企业或组织机构的高中级管理人员
- 简单用户：银行的职员、机票预定人员等
- 复杂用户
 - 工程师、科学家、经济学家、科技工作者等
 - 直接使用数据库语言访问数据库，甚至能够基于数据库管理系统的**API**编制自己的应用程序

■ 其他相关人员

- **DBSM**设计者和实现者
- 数据库系统工具开发者
-



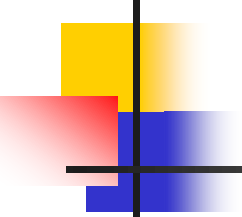
7.数据库技术的发展

■ 第一代：层次、网状数据库系统

- 20世纪60年代末，**IBM**的层次数据库系统**IMS**，是世界上最先出现的商品化数据库管理系统，该系统的不同版本至今还在使用。
- 美国数据库系统语言研究会下属的数据库任务组在70年代初公布了**DBTG**报告，由此发展了许多网状数据库系统。

■ 第二代：关系数据库系统

- 1970年**E.F.Codd**提出了关系数据模型的概念，之后相继发表了一系列关系模型的论文，奠定了关系数据库的理论基础。
- **IBM**用了5年时间研制出了关系数据库系统**System R**
- 加州大学**Berkeley**分校也研制出了关系数据库系统**INGRES**



■ 第三代：新一代数据库系统

- 关系数据库系统难以满足新的应用领域的需求
- 数据库技术与其他领域的技术相结合，出现了新的数据库分支
 - 网络数据库
 - 分布式数据库
 - 面向对象数据库
 - 多媒体数据库
 - **XML数据库等等**
 - **NoSQL**
 - **NewSQL**
 -

百花齐放 百家争鸣



本章小结

- 数据库系统的基本概念
 - 数据、数据库、数据库管理系统、数据库系统
- 数据管理
- 数据模型
- 数据库系统的三级模式结构和二级映像
- 数据库管理系统的功能、组成和工作过程
- 数据库系统的用户视图
- 数据库技术的发展