

# 数据库系统概论

**An Introduction to Database System**

# 引言

- 数据库系统的发展经历了三代演变

层次/网状数据库系统、关系数据库系统、新一代数据库系统

- 造就了四位图灵奖得主

C.W.Bachman、E.F.Codd、James Gray、M.R.Stonebraker

- 发展了一门计算机基础学科

数据建模和**DBMS**核心技术为主，内容丰富领域宽广

- 带动了一个巨大软件产业

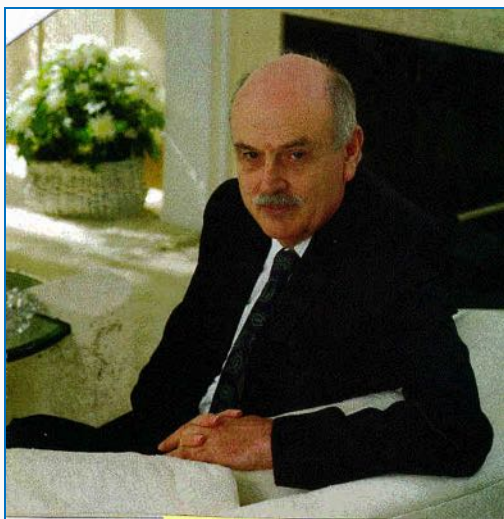
**DBMS**及其相关工具产品、应用套件、解决方案

**数据库技术和系统已经成为信息系统的核心技术和重要基础设施**

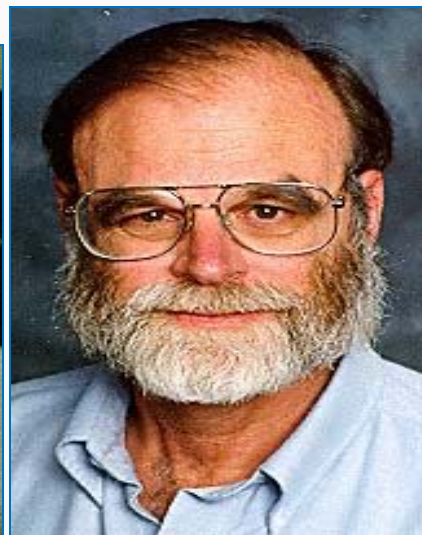
# 四位图灵奖得主



**C.W.Bachman (1973)**



**E.F.Codd (1981)**



**James Gray (1998)**



**M.R.Stonebraker (2014)**

# Charles.W.Bachman

## 网状数据库之父



❖ 1924年出生于美国堪萨斯州的曼哈顿。1970—1981年在Honeywell公司任总工程师，兼任Cullinet软件公司的副总裁。

他在数据库方面的杰出成就：

- 1 1960年为通用电气开发了世界上第一个网状数据库系统IDS
- 2 积极推动与促成了数据库标准的制定：DBTG报告，成为数据库历史上具有里程碑意义的文献。1971年第一版，73、78、81、84修订版。

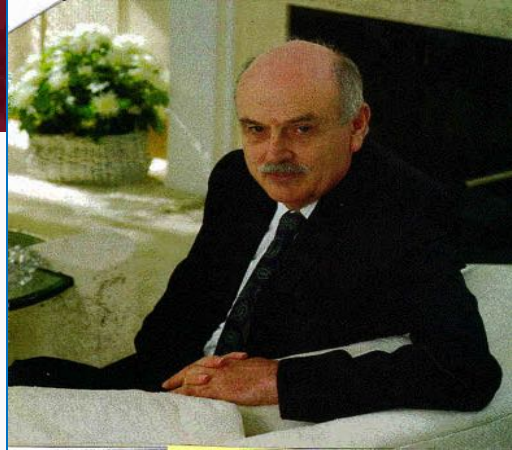
**巴赫曼在数据库技术的产生、发展与推广应用方面都发挥了巨大的作用**

❖ 1973获图灵奖

❖ 1983年成立自己的公司—Bachman Information System

# Edgar F.Codd 博士

## 关系数据库之父 美国工程院院士

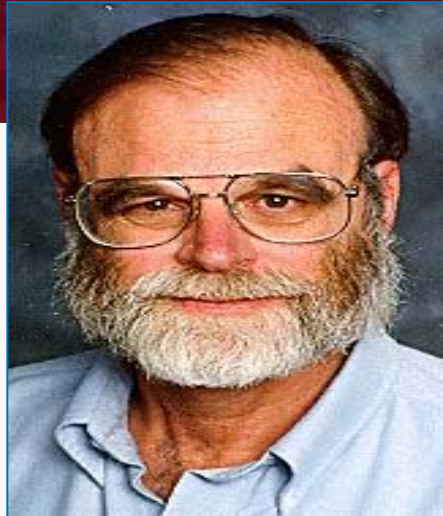


- ❖ 英国人，1923生于英格兰中部波特兰。  
第二次世界大战时应征入伍，在皇家空军服役。  
1942-1945年间任机长，参与了许多惊心动魄的空战。
- ❖ 大战结束后，到英国牛津大学 数学专业 理学士及硕士学位，48年远渡大西洋到IBM工作从事操作系统和自动机理论研究。
- ❖ 年近40重返密歇根大学进修计算机与通信专业  
1963年获得硕士学位，1965年又获得博士学位。
- ❖ 60年代后期开始数据库研究，1970年E.F.Codd 博士提出关系模型概念  
(CACM, Vol.13, Vol.6, 1970, “A Relational Model of Data for Large Shared Data Banks” ACM在1983年把这篇论文列为从1958年以来的四分之一世纪中具有里程碑式意义的最重要的25篇研究论文之一。)
- ❖ 1981年获图灵奖，84年从IBM公司退休。还创办了一个研究所和一个公司。



# James Gray

## 数据库技术和事务处理专家



- ❖ 1944年生，美国加州大学伯克利分校计算机科学系博士。
- ❖ 先后在贝尔实验室、IBM、Tandem、DEC等工作，研究方向转向数据库领域。
- ❖ 由于他在数据库事务处理研究方面的原创性贡献以及在将研究原型转化为商业产品的系统实现方面的技术领袖地位，1998年获奖（时任微软研究员）
- ❖ 2007年1月28日失踪。

# M.R.Stonebraker



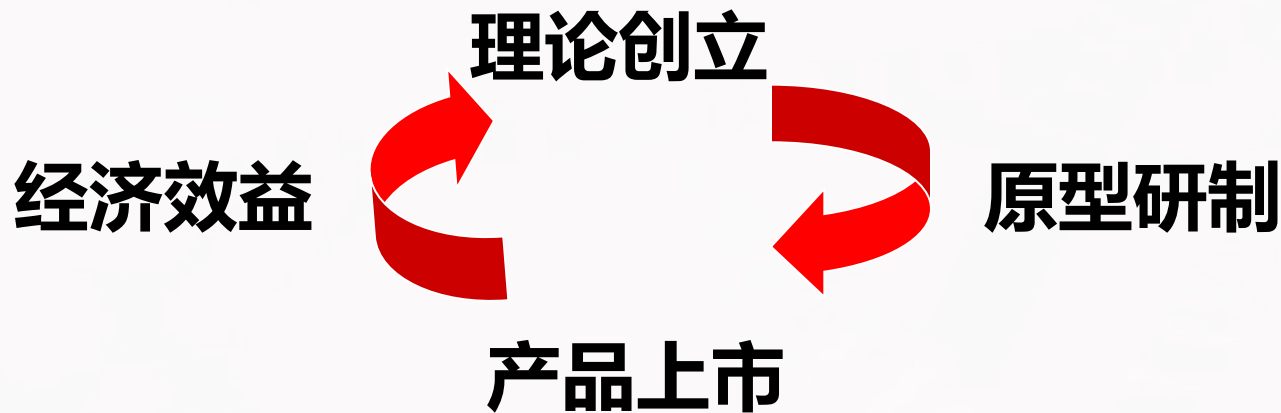
## 现代主流数据库系统架构的奠基人

- ❖ 1971年至2000年为**第一阶段**，从事关系数据库的体系架构与实现技术研究
- ❖ 2001年至2008年为**第二阶段**，在**One-size-does-not-fit-all**的理念下，开发了一系列新型数据库系统的体系架构设计与产品开发
- ❖ 2009年至今为**第三阶段**，大数据系统的体系架构设计与实践
- ❖ **2014年获图灵奖**

**2015年10月22日中国计算机大会上（合肥）做大会报告**

# 数据库:一个巨大的软件产业

## 形成良性循环



是理论成果转化为产品的成功范例





中国数据库开创者—萨师煊，中国人民大学教授（1922-2010）

# 数据库系统课程简介

- **教学内容**：本课程融数据库基础、数据库管理系统和数据库应用技术于一体，以关系数据库为重点，全面系统地介绍数据库的基本概念、数据库语言SQL、数据库设计的理论、数据库设计的步骤与方法、数据库管理系统各项功能的实现技术和数据库应用系统的开发方法。
- **教学目标**：学生全面了解数据库的基本概念、关系数据库语言SQL、数据库设计的理论和步骤与方法、数据库系统的组成及数据库管理系统的主要功能与实现技术、基于主流数据库管理系统的数据库应用系统开发方法，理解数据库应用系统软件的开发流程，为今后从事数据库系统的开发或在数据库技术方面继续深造奠定基础。

# 内容安排(1) 48+8

## 基础篇

- 第一章 绪论
- 第二章 关系数据库
- 第三章 关系数据库标准语言SQL
- 第四章 数据库安全性
- 第五章 数据库完整性

## 设计与应用开发篇

- 第六章 关系数据理论
- 第七章 数据库设计

# 内容安排(2)

## 系统篇

- 第九章 关系查询处理和查询优化
- 第十章 数据库恢复技术
- 第十一章 并发控制

## 实验计划：完成三个实验

实验一：SQL语言（2学时）

实验二：DBMS的数据库保护（2学时）

实验三：小型MIS开发（4学时）

# 教材及参考书

## 教材

- 王珊,萨师煊.数据库系统概论（第**5**版）.北京:高等教育出版社，**2014**

## 参考书

- 施伯乐，丁宝康. 数据库系统教程(第**3**版). 北京：高等教育出版社，**2008**



# 《数据库系统概论》--国家级优秀教材



第一版  
1983



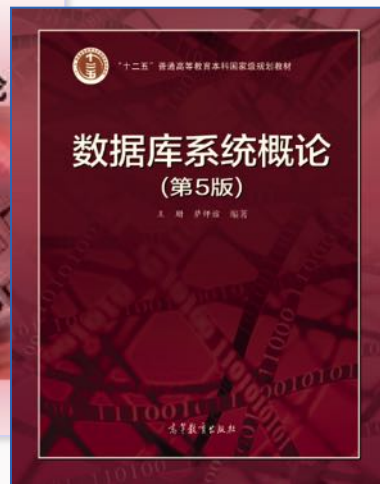
第二版  
1991



第三版  
2000



第四版  
2006



第五版  
2014



《数据库系统概论》已经出版第五版，**2次获得国家级优秀教材奖**，成为一本优秀的经典教材。为培养我国数据库专业人才、推动我国数据库技术的发展做出了突出贡献。

# 慕课资源

- ❖ 中国人民大学慕课课程
- ❖ 数据库系统概论（基础篇）
- ❖ <https://www.icourse163.org/course/RUC-488001>
- ❖ 数据库系统概论（高级篇）
- ❖ <https://www.icourse163.org/course/RUC-1001655006>

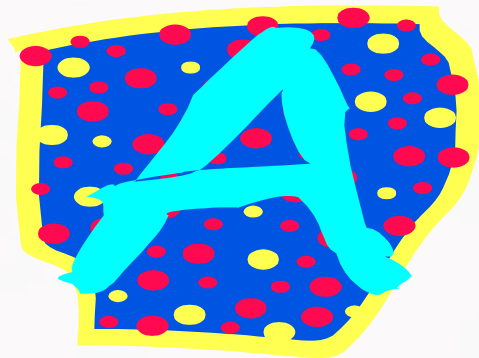
# 考核形式

**课程性质: 必修**

**考核方式: 考试、闭卷笔试**

**成绩比例: 期末70%，平时30%**

**平时包括：作业、到课率（点名和课堂练习）、实验和实验报告**



# 答疑方式

❖ E-MAIL: [bxia@njupt.edu.cn](mailto:bxia@njupt.edu.cn)

❖ 答疑时间: 任何时间, QQ

# 数据库系统概论

## 第一章 绪论

# 第一章 绪论

## 1.1 数据库系统概述

## 1.2 数据模型

## 1.3 数据库系统的结构

## 1.4 数据库系统的组成

## 1.5 小结



# 1.1 数据库系统概述

## 1.1.1 数据库的4个基本概念

## 1.1.2 数据管理技术的产生和发展

## 1.1.3 数据库系统的特点

# 1.1.1 数据库的4个基本概念

❖ 数据 (Data)

❖ 数据库 (Database, DB)

❖ 数据库管理系统

(DataBase Management System, DBMS)

❖ 数据库系统

( DataBase System , DBS )

# 1. 数据

❖ 数据（**Data**）是数据库中存储的基本对象

❖ 数据的定义

- 描述事物的符号记录

❖ 数据的种类

- 数字、文字、图形、图像、音频、视频、学生的档案记录、订单情况等等

# 数据举例

❖ 数据的含义称为数据的语义，**数据与其语义是不可分的。**

■ 例如 **93**是一个数据

语义1：学生某门课的成绩

语义2：某人的体重

语义3：计算机系**2014**级学生人数

语义4：请同学给出.....

# 数据举例

- ❖ 日常生活中，人们可以直接用自然语言（如汉语）来描述事物
- ❖ 计算机中常常用记录来描述，如**学生档案中的学生记录**：  
（李明，男，**199505**，江苏南京市，计算机系，**2013**）
- ❖ 数据的形式不能完全表达其内容
- ❖ 数据的解释
  - **语义**：学生姓名、性别、出生年份、籍贯、所在系别、入学时间
  - **解释**：李明是大学生，**1995**年**5**月生，男，江苏南京人，**2013**年考入计算机系
- ❖ **数据有结构**的：记录是计算机存储数据的一种格式或一种方法

## 2. 数据库

### ❖ 什么是数据库

#### ■ 数据库（Database，简称DB）

是长期储存在计算机内、有组织的、可共享的大量数据的集合。

### ❖ 为什么要建立数据库

收集并抽取出一个应用所需要的大量数据，将其保存，以供进一步加工处理，抽取有用信息，转换为有价值的知识。

### ❖ 数据库的基本特征

#### ■ 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、易扩展

#### ■ 数据独立性较高



# 3. 数据库管理系统

## ❖ 什么是数据库管理系统（DBMS）

- 位于用户应用与操作系统之间的一层数据管理软件
- 是基础软件，是一个大型复杂的软件系统

## ❖ 数据库管理系统的用途

- 科学地组织和存储数据、高效地获取和维护数据

# 数据库管理系统的主要功能

## 1 数据定义功能

- 提供数据定义语言（DDL）
- 定义数据库中的数据对象

## 2 数据组织、存储和管理

- 分类组织、存储和管理各种数据
- 确定数据在存储级别上的结构和存取方式
- 实现数据之间的联系
- 提供多种存取方法提高存取效率

# 数据库管理系统的主要功能

## 3 数据操纵功能

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

## 4 数据库的事务管理和运行管理

- 数据的安全性、完整性、多用户对数据的并发使用
- 发生故障后的系统恢复数据库

由数据库管理系统统一管理和控制，保证事务正确运行

# 数据库管理系统的主要功能

## 5 数据库的建立和维护功能

提供实用程序/工具，完成数据库数据批量装载，数据库转储，介质故障恢复，数据库的重组织和性能监视等

## 6 其它功能

- 数据库管理系统与网络中其它软件系统的通信
- 数据库管理系统系统之间的数据转换
- 异构数据库之间的互访和互操作

# 4.数据库系统

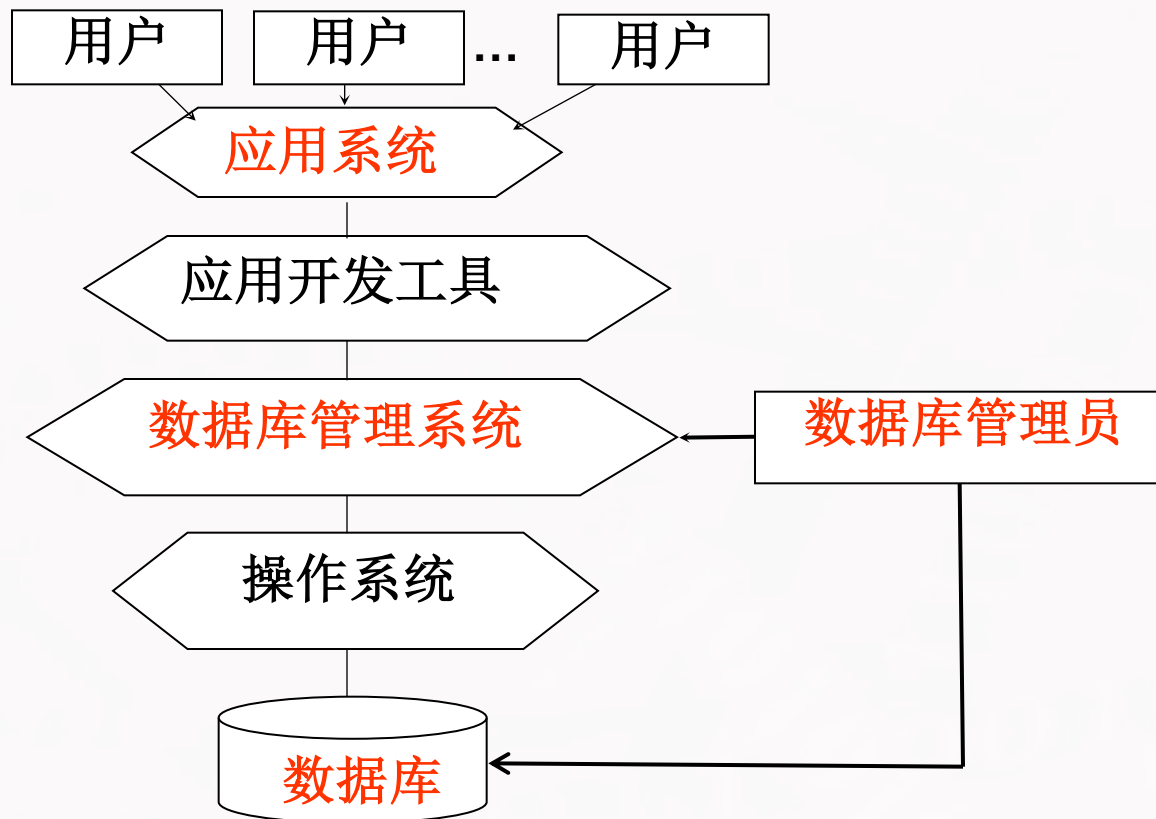
## ❖ 数据库系统（Database System，简称DBS）

- 是指在计算机系统中引入数据库后的系统构成。
- 在不引起混淆的情况下常常把数据库系统简称为数据库。

## ❖ 数据库系统的构成

- 数据库
- 数据库管理系统（及其应用开发工具）
- 应用程序
- 数据库管理员（DataBase Administrator，DBA）

# 数据库系统



# 1.1 数据库系统概述

## 1.1.1 四个基本概念

## 1.1.2 数据管理技术的产生和发展

## 1.1.3 数据库系统的特点



# 数据管理技术的产生和发展

## ❖ 什么是数据管理

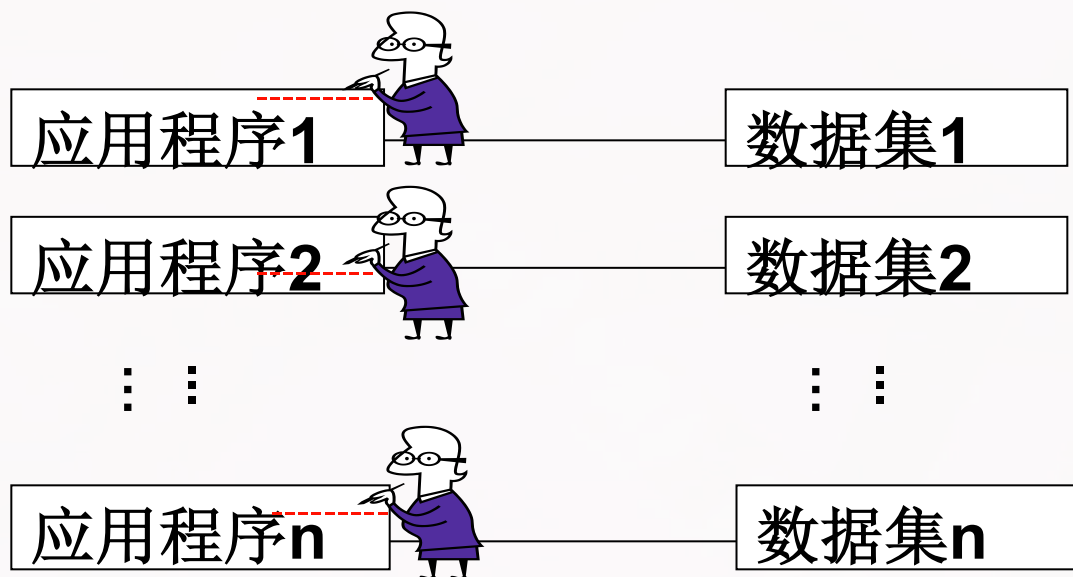
- 对数据进行**分类、组织、编码、存储、检索和维护**
- 数据处理和数据分析的中心问题

## ❖ 数据管理技术的发展过程

- 人工管理阶段（**20世纪50年代中之前**）
- 文件系统阶段（**20世纪50年代末--60年代中**）
- 数据库系统阶段（**20世纪60年代末--现在**）

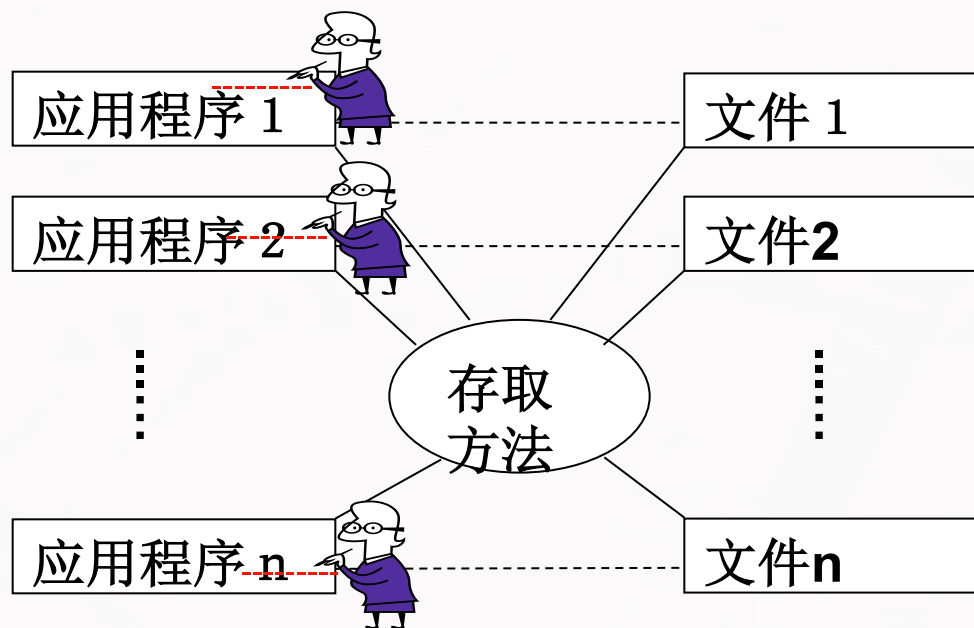
|    |         | 人工管理阶段      | 文件系统阶段         | 数据库系统阶段                    |
|----|---------|-------------|----------------|----------------------------|
| 背景 | 应用背景    | 科学计算        | 科学计算、管理        | 大规模数据管理                    |
|    | 硬件背景    | 无直接存取存储设备   | 磁盘、磁鼓          | 大容量磁盘、磁盘阵列                 |
|    | 软件背景    | 无操作系统       | 有文件系统          | 有数据库管理系统                   |
|    | 处理方式    | 批处理         | 联机实时处理，<br>批处理 | 联机实时处理，分布处理，<br>批处理        |
| 特点 | 数据的管理者  | 用户(程序员)     | 文件系统           | 数据库管理系统                    |
|    | 数据面向的对象 | 某一应用程序      | 某一应用           | 现实世界(一个企业、跨国公司)            |
|    | 数据的共享程度 | 无共享，冗余度极大   | 共享性差，冗余度大      | 共享性高，冗余度小                  |
|    | 数据的独立性  | 不独立，完全依赖于程序 | 独立性差           | 具有高度的物理独立性和一定的逻辑独立性        |
|    | 数据的结构化  | 无结构         | 记录内有结构，整体无结构   | 整体结构化，用数据模型描述              |
|    | 数据控制能力  | 应用程序自己控制    | 应用程序自己控制       | 由DBMS提供数据安全性、完整性、并发控制和恢复能力 |

# 应用程序与数据的对应关系（人工管理阶段）



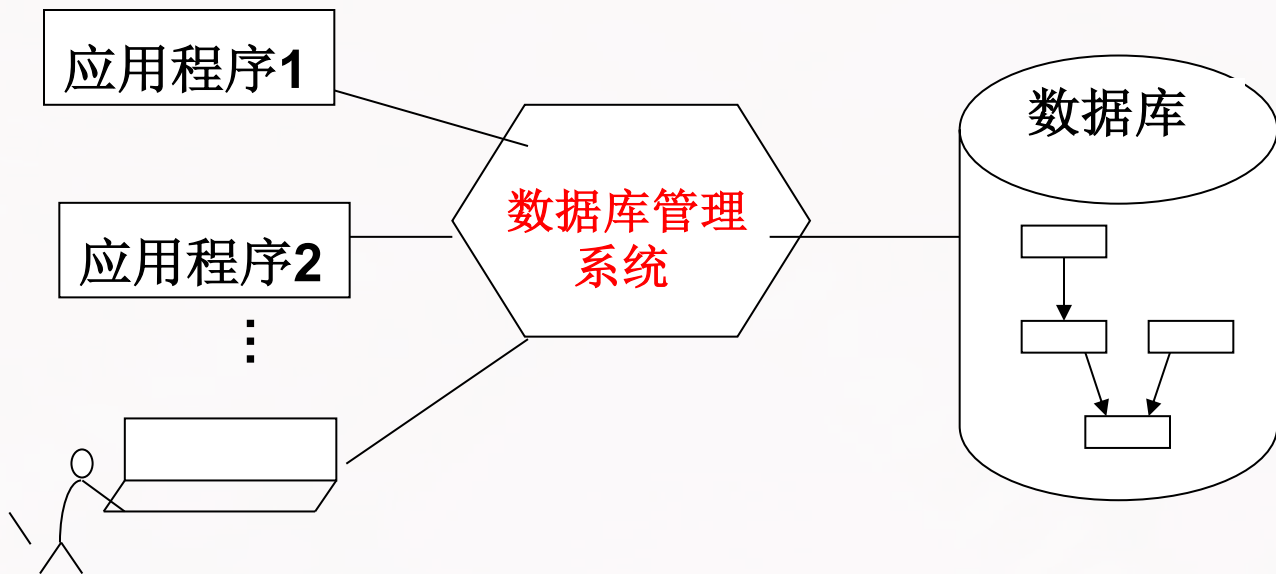
人工管理阶段 应用程序与数据之间的对应关系

# 应用程序与数据的对应关系（文件系统阶段）



文件系统阶段 应用程序与数据之间的对应关系

# 应用程序与数据的对应关系（数据库系统阶段）



数据库系统阶段 应用程序与数据之间的对应关系

# 1.1 数据库系统概述

1.1.1 四个基本概念

1.1.2 数据管理技术的产生和发展

1.1.3 数据库系统的特点

# 阿波罗登月计划

## ❖ 阿波罗飞船登月计划的需求

- 协调分散在全球制造的**200**万个阿波罗飞船零部件的生产进度
- 用文件系统开发了一个零部件生产计算机管理系统。
- 系统虽然可以工作，但由于文件系统分散管理的弱点，效率极低，**60%是冗余数据，维护十分困难。**
- 该系统曾一度成为实现阿波罗计划的重大障碍之一

## ❖ 各国计算机学术界和工业界纷纷开展研究

- 数据建模、数据模型研究与实现的探索
- 成果是出现了一种全新的高效的数据管理技术—数据库技术(**IMS\IDS**)

# 一个例子：用文件系统实现学籍管理

- 学生的信息包括学号、姓名、性别、年龄、专业和奖励
- 数据存储

“学生基本信息”和“奖励”文件的结构和内容

| 学号       | 姓名    | 性别    | 年龄    | 专业    | 位置    | 长度    |
|----------|-------|-------|-------|-------|-------|-------|
| 20100001 | 史玉明   | 女     | 20    | 计算机   | 0     | 30    |
| 20100100 | 李明虎   | 男     | 21    | 机械    | 30    | 15    |
| 20100234 | 张翔    | 男     | 21    | 化工    | 45    | 0     |
| .....    | ..... | ..... | ..... | ..... | ..... | ..... |

| 奖励                  |
|---------------------|
| 2011校奖学金, 2012国家奖学金 |
| 2012校优秀学生           |

缺点：程序员必须关注记录结构和不同文件中记录之间的联系，工作量大，编程复杂，开发速度慢。

- 查询数据

➤ 编写应用程序，实现数据的录入和查找



# 一个例子：用数据库系统实现学籍管理

存储数据：建立两张表，用**CREATE**命令：

**CREATE TABLE STUDENT (**            **// 存放学生的基本信息**

```
Sno    CHAR(8),
Sname  CHAR(10),
Ssex   CHAR(2),
Sage   SMALLINT,
Major  CHAR(20) );
```

**CREATE TABLE AWARD(**            **// 存放学生的奖励情况**

```
Sno    CHAR(8),
Details VARCHAR(2000) );
```

数据录入，用**INSERT**插入命令：

```
INSERT INTO STUDENT (Sno, Sname, Ssex, Sage, Major)
VALUES('20100001', '史玉明', '女', 20, '计算机') ;
```

**// 插入学生的基本信息**

```
INSERT INTO AWARD (Sno, Details)
```

```
VALUES('20100001', '2011校奖学金, 2012国家奖学金');
```

**// 插入学生获得的奖励**

查询功能：用一条查询语句实现：

```
SELECT A.Sno, Sname, Ssex, Sage, Major, Details
FROM STUDENT A LEFT JOIN AWARD B ON A.Sno=B.Sno
WHERE A.Sno = '20100001'
```

**优点：不要关注记录的存储和不同表之间的联系，不要编程，开发速度快。**

# 1.1.3 数据库系统的特点

- ❖ 数据结构化
- ❖ 数据的共享性高，冗余度低且易扩充
- ❖ 数据独立性高
- ❖ 数据由数据库管理系统统一管理和控制

# 数据结构化

## ❖ 数据的整体结构化是数据库的主要特征之一

- 不再仅仅针对某一个应用，而是面向整个企业或组织
- 不仅数据内部结构化，整体是结构化的，数据之间具有联系
- 数据记录可以变长
- 数据的最小存取单位是数据项

## ❖ 数据用数据模型描述，无需应用程序定义

学生基本记录

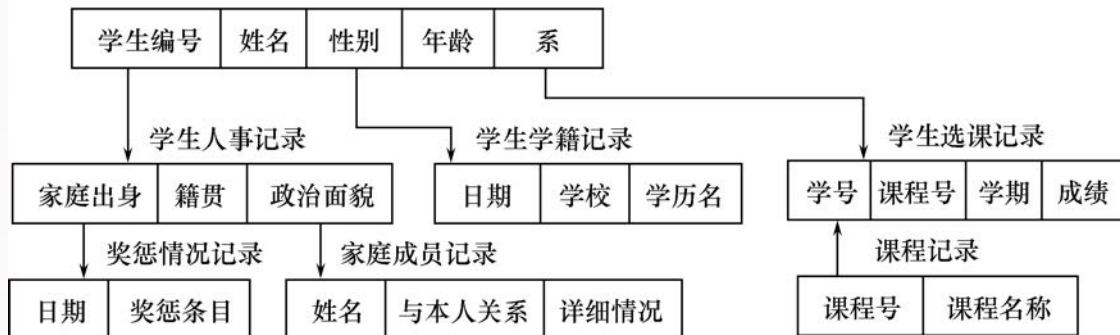


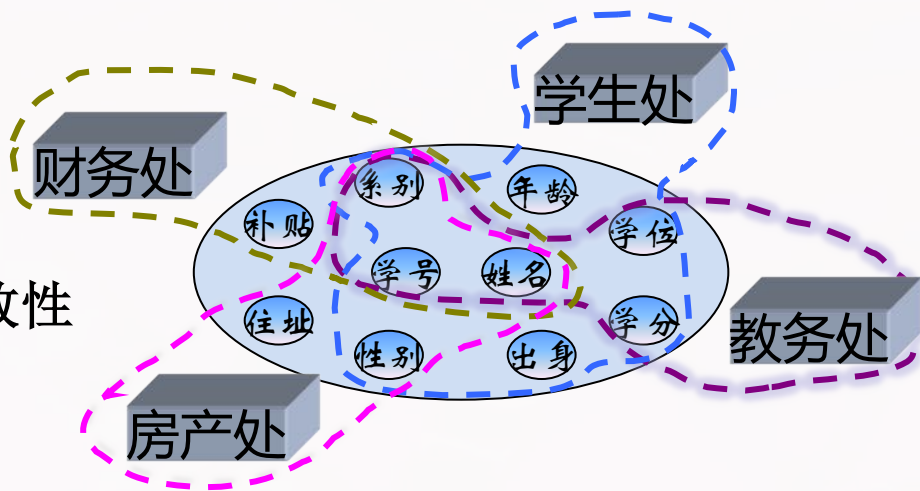
图1.5 某学校信息系统中的学生数据

## 数据的共享性高，冗余度低且易扩充

❖ 数据面向整个系统，可以被多个用户、多个应用共享使用。

❖ 数据共享的好处

- 减少数据冗余，节约存储空间
- 避免数据之间的不相容性与不一致性
- 使系统易于扩充



# 数据独立性高

## ❖ 物理独立性

- 指用户的应用程序与数据库中数据的物理存储是相互独立的。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，应用程序不用改变。

## ❖ 数据独立性由数据库管理系统的二级映像功能来保证。

# 数据由数据管理系统统一管理和控制

## ❖ 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

保证数据的正确性、有效性和相容性。

### (3) 并发控制 (Concurrency Control)

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。

# 1.1 数据库系统概述

1.1.1 掌握数据库的4个基本概念

1.1.2 了解数据管理技术的产生和发展概况

1.1.3 了解数据库系统的特点

❖目的：了解基本知识，初步掌握基本概念

❖难点：需要掌握数据库领域大量的基本概念

# 数据库定义

- ❖ 数据库是长期存储在计算机内有组织的共享的大量的数据集合。
- ❖ 可以供各种用户共享，具有最小冗余度和较高的数据独立性。
- ❖ 数据库管理系统在数据库建立、运用和维护时对数据库进行统一控制，以保证数据的完整性、安全性，并在多用户同时使用数据库时进行并发控制，在发生故障后对数据库进行恢复。



# 第一章 绪论

## 1.1 数据库系统概述

## 1.2 数据模型

## 1.3 数据库系统的结构

## 1.4 数据库系统的组成

## 1.5 小结

## 1.2 数据模型

- ❖ 数据模型是对现实世界数据特征的抽象。
- ❖ 通俗地讲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；
- ❖ 数据模型是数据库系统的核心和基础



# 1.2 数据模型

## 1.2.1 两类数据模型

## 1.2.2 概念模型

## 1.2.3 数据模型的组成要素

## 1.2.4 常用的数据模型

## 1.2.5 层次模型

## 1.2.6 网状模型

## 1.2.7 关系模型

# 1.2.1 两类数据模型

❖ 数据模型分为两类（两个不同的层次）

## （1）概念模型，也称信息模型

它是按用户的观点来对数据和信息建模，用于数据库设计。

## （2）逻辑模型和物理模型

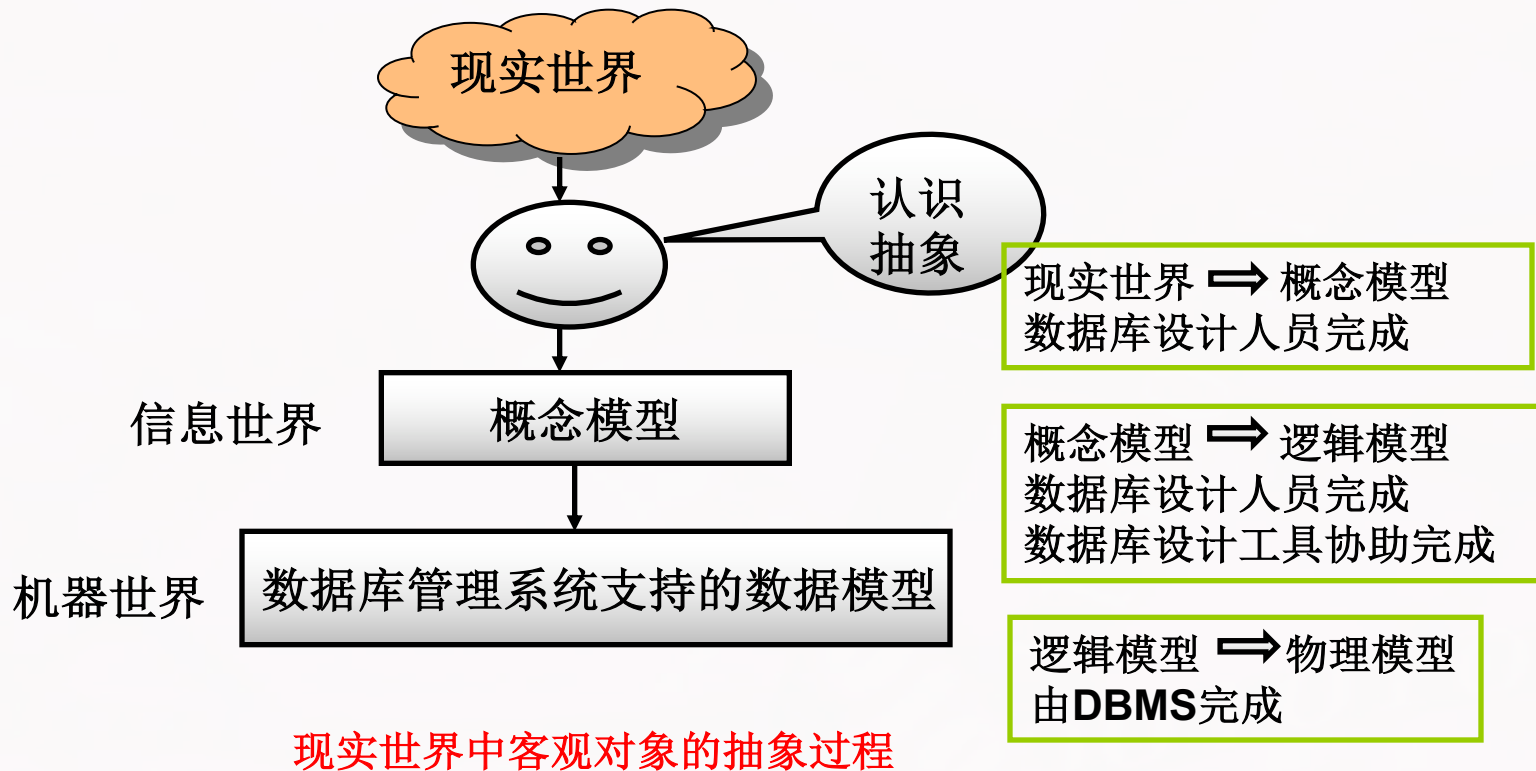
- 逻辑模型主要包括网状模型、层次模型、关系模型、面向对象数据模型、对象关系数据模型、半结构化数据模型等。

按计算机系统的观点对数据建模，用于DBMS实现。

- 物理模型是对数据最底层的抽象

描述数据在系统内（磁盘上）的表示方式和存取方法。

# 两类数据模型



# 1.2 数据模型

1.2.1 两大类数据模型

1.2.2 概念模型

1.2.3 数据模型的组成要素

1.2.4 最常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

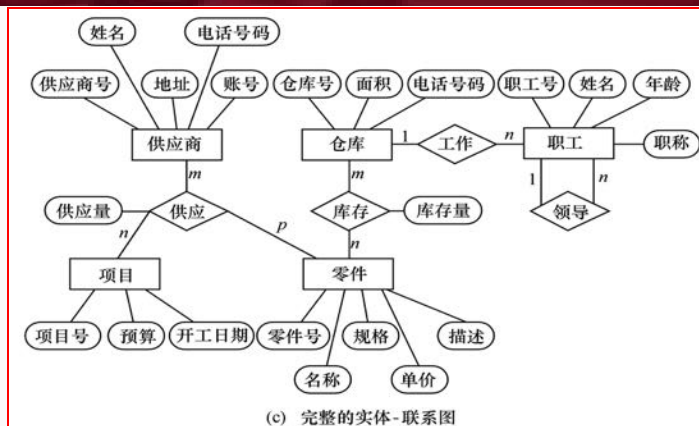
## 1.2.2 概念模型

### ❖ 概念模型的用途

- 概念模型用于信息世界的建模
- 是现实世界到机器世界的一个中间层次
- 是数据库设计的有力工具
- 数据库设计人员和用户之间进行交流的语言

### ❖ 对概念模型的基本要求

- 较强的语义表达能力
- 简单、清晰、易于用户理解



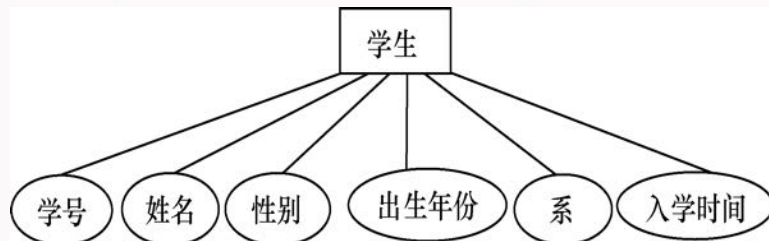
例：工厂物质管理的概念模型

# 1. 信息世界中的基本概念

## (1) 实体 (Entity)

客观存在并可相互区别的事物称为实体。

可以是具体的人、事、物或抽象的概念。



## (2) 属性 (Attribute)

实体所具有的某一特性称为属性。一个实体可以由若干个属性来刻画。

## (3) 码 (Key)

唯一标识实体的属性集称为码。

## (4) 实体型 (Entity Type)

用实体名及其属性名集合来抽象和刻画同类实体称为实体型

## (5) 实体集 (Entity Set)

同一类型实体的集合称为实体集



# 信息世界中的基本概念（续）

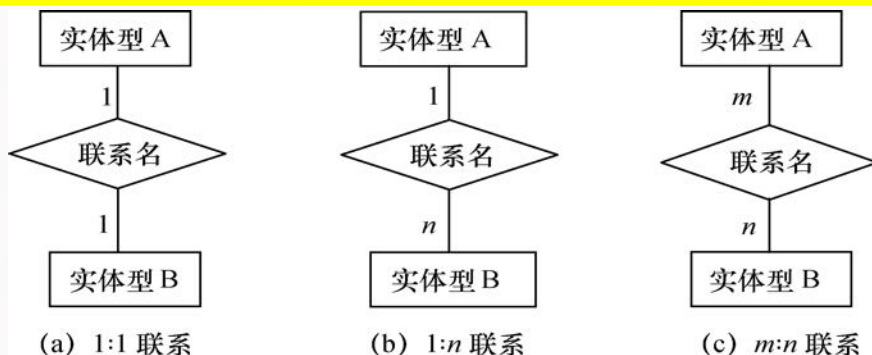
## （6）联系（Relationship）

■ 现实世界中事物内部以及事物之间的联系在信息世界中反映为实体（型）内部的联系和实体（型）之间的联系。

■ 实体内部的联系：是指组成实体的各属性之间的联系

■ 实体之间的联系：通常是指不同实体集之间的联系

实体之间的联系有一对一（1:1）、一对多（1:n）和多对多（m:n）等多种类型



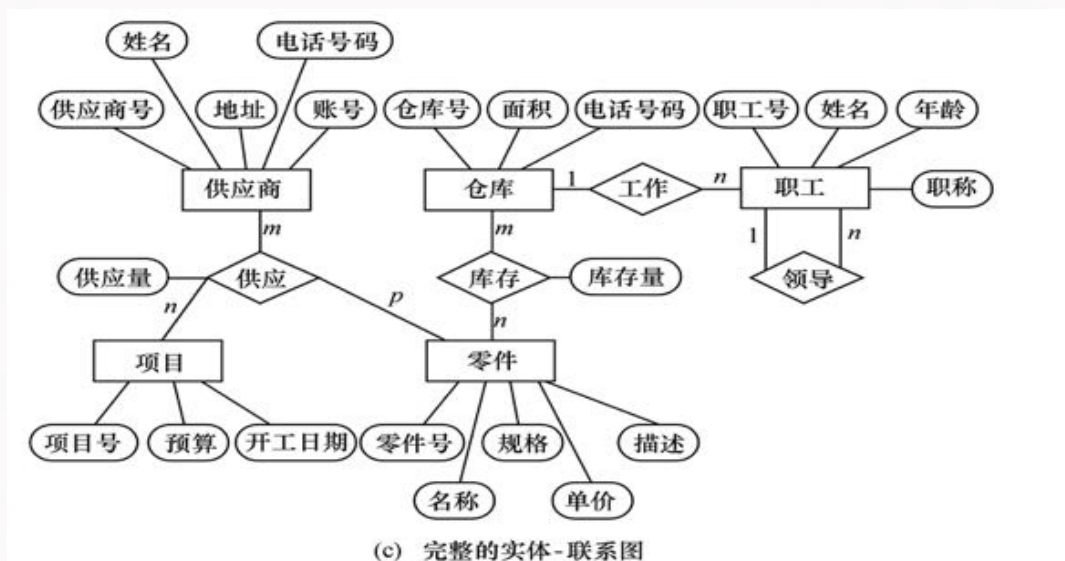
# 实体-联系方法

❖ 概念模型的一种表示方法:

❖ **实体-联系方法 (Entity-Relationship Approach)**

■ 用E-R图来描述现实世界的概念模型

■ E-R方法也称为E-R模型



# 1.2 数据模型

1.2.1 两大类数据模型

1.2.2 概念模型

1.2.3 数据模型的组成要素

1.2.4 最常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

## 1.2.3 数据模型的组成要素

### ❖ 数据模型是严格定义的一组概念的集合

精确地描述了系统的静态特性、动态特性和完整性约束条件 (Integrity Constraints)。

### ❖ 数据模型由三部分组成

- 1 数据结构--描述系统的静态特性
- 2 数据操作--描述系统的动态特性
- 3 完整性约束

# 1. 数据结构

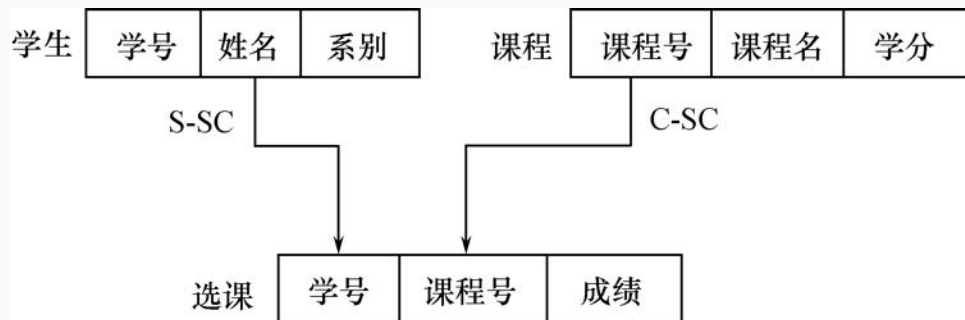
## ❖ 刻画数据模型性质的重要方面

### 数据结构的类型来命名数据模型

层次结构-层次模型、网状结构--网状模型、关系结构—关系模型

## ❖ 描述数据库的组成对象--对象的类型、内容、性质

## ❖ 描述对象之间的联系



一个网状数据模型实示例

记录:

学生--由学号、姓名所在的专业系名等组成

**SET TYPE:**

**S-SC**--学生记录和学生选课记录之间的联系

## 2. 数据操作

### ❖ 数据操作

- 对数据库中各种对象的实例允许执行的操作的集合  
包括操作及有关的操作规则

### ❖ 数据操作的类型

- 查询
- 更新（包括插入、删除、修改）

### ❖ 数据操作语言

- 定义数据操作的确切含义、符号、优先级别
- 实现数据操作的语言
  - 查询语言——**Query Language**
  - 更新语言——**Insert、Delete、Update**

# 3. 数据的完整性约束条件

## ❖ 一组完整性规则的集合

- 完整性规则：给定的数据模型中数据及其联系所具有的**制约和依存规则**。
- 用以限定符合数据模型的数据库状态以及状态的变化，以保证数据的**正确、有效和相容**。

## ❖ 数据模型对完整性约束条件的定义

- 反映和规定必须遵守的**基本的通用的**完整性约束条件。
- 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**。

## 1.2 数据模型

1.2.1 两大类数据模型

1.2.2 数据模型的组成要素

1.2.3 概念模型

**1.2.4 常用的数据模型**

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型



## 1.2.4 常用的数据模型

- ❖ 层次模型 (Hierarchical Model)
  - ❖ 网状模型 (Network Model)
  - ❖ 关系模型 (Relational Model)
  - ❖ 面向对象数据模型 (Object Oriented Data Model)
  - ❖ 对象关系数据模型 (Object Relational Data Model)
  - ❖ 半结构化数据模型 (Semi-structure Data Model) —如XML
  - ❖ 非结构化数据模型、图模型 .....
- 格式化模型
- 对象模型

# 1.2 数据模型

1.2.1 两大类数据模型

1.2.2 数据模型的组成要素

1.2.3 概念模型

1.2.4 最常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

## 1.2.7 关系模型

- ❖ 关系数据库系统采用关系模型作为数据的组织方式
- ❖ 1970年美国IBM公司San Jose研究室的研究员E.F.Codd首次提出了数据库系统的关系模型
- ❖ 计算机厂商推出的数据库管理系统几乎都支持关系模型

# 1. 关系模型的数据结构

❖ 在用户观点下，关系模型中数据的逻辑结构是一张二维表。

学生登记表

| 学 号     | 姓 名 | 年 龄 | 性 别 | 系 名 | 年 级  |
|---------|-----|-----|-----|-----|------|
| 2013004 | 王小明 | 19  | 女   | 社会学 | 2013 |
| 2013006 | 黄大鹏 | 20  | 男   | 商品学 | 2013 |
| 2013008 | 张文斌 | 18  | 女   | 法律  | 2013 |
| ...     | ... | ... | ... | ... | ...  |

属性

元组

# 关系模型的数据结构（续）

- 关系（**Relation**）—— 一个关系对应通常说的一张表
- 元组（**Tuple**）—— 表中的一行即为一个元组
- 属性（**Attribute**）—— 表中的一列即为一个属性，给每一个属性起一个名称即属性名
- 码（**Key**）—— 也称码键。表中的某个属性组，它可以唯一确定一个元组
- 域（**Domain**）—— 是一组具有相同数据类型的值的集合。

属性的取值范围来自某个域。

例：学生年龄属性的域（**15~45岁**），  
性别的域是（男，女），  
系名的域是一个学校所有系名的集合；

- 分量——元组中的一个属性值。
- 关系模式——对关系的描述

关系名（属性1，属性2，...，属性n）

学生（学号，姓名，年龄，性别，系名，年级）

# 关系模型的数据结构（续）

❖ 关系必须是规范化的，满足一定的规范条件

最基本的规范条件：关系的每一个分量必须是一个不可分的数据项，**不允许表中还有表**。图1.15中**工资和扣除是可分的数据项**，不符合关系模型要求

| 职工号   | 姓名 | 职称 | 工 资  |      |      | 扣 除 |       | 实 发  |
|-------|----|----|------|------|------|-----|-------|------|
|       |    |    | 基本工资 | 岗位津贴 | 业绩津贴 | 三险  | 个人所得税 |      |
| 86051 | 陈平 | 讲师 | 3305 | 2200 | 2850 | 360 | 812   | 7183 |
| ⋮     | ⋮  | ⋮  | ⋮    | ⋮    | ⋮    | ⋮   | ⋮     | ⋮    |

图1.15 一个工资表（表中有表）实例

# 关系模型的数据结构（续）

表1.5 术语对比

| 关系术语  | 一般表格的术语       |
|-------|---------------|
| 关系名   | 表名            |
| 关系模式  | 表头（表格的描述）     |
| 关系    | （一张）二维表       |
| 元组    | 记录或行          |
| 属性    | 列             |
| 属性名   | 列名            |
| 属性值   | 列值            |
| 分量    | 一条记录中的一个列值    |
| 非规范关系 | 表中有表（大表中嵌有小表） |

## 2. 关系模型的操纵与完整性约束

- ❖ 数据操作是集合操作，操作对象和操作结果都是关系
  - 查询
  - 插入
  - 删除
  - 更新
- ❖ 存取路径对用户隐蔽，用户只要指出“找什么”，不必详细说明“怎么找”。
  - ➡ 提高了数据的独立性，提高了用户生产率。



# 关系模型的操纵与完整性约束（续）

## ❖ 关系的完整性约束条件

- 实体完整性
  - 参照完整性
  - 用户定义的完整性
- 关系的两个不变性

# 3. 关系模型的优缺点

## 优点

- ❖ 建立在严格的数学概念的基础上
- ❖ 概念单一
  - 实体和各类联系都用关系来表示
  - 对数据的检索结果也是关系
- ❖ 关系模型的存取路径对用户透明
  - 具有更高的数据独立性，更好的安全保密性
  - 简化了程序员的工作和数据库开发建立的工作

## 缺点

- ❖ 存取路径对用户透明，查询效率往往不如格式化数据模型
- ❖ 为提高性能，必须对用户的查询请求进行优化，增加了开发数据库管理系统的难度

# 1.2 数据模型

## ① 自学了解的

层次数据模型及网状数据模型的基本概念

层次和网状数据库的内容十分精简

这两类系统虽然有它们的缺点，但是执行效率高是他们的显著优点

## ② 需要牢固掌握的

概念模型的基本概念

数据模型的**3**个组成要素

关系数据模型的相关概念

## ③ 难点

数据模型的概念对于刚刚学习数据库的读者来说会感到比较抽象。

# 第一章 绪论

## 1.1 数据库系统概述

## 1.2 数据模型

## 1.3 数据库系统的结构

## 1.4 数据库系统的组成

## 1.5 小结

# 1.3 数据库系统的结构

## ❖ 从数据库应用开发人员角度看

数据库系统采用三级模式结构，是数据库系统内部的系统结构

## ❖ 从数据库最终用户角度看

数据库系统的结构有：

- 单用户结构
- 主从式结构
- 分布式结构
- 客户-服务器
- 浏览器-应用服务器 / 数据库服务器
- 等

# 数据库系统的结构（续）

## 1.3.1 数据库系统模式的概念

## 1.3.2 数据库系统的三级模式结构

## 1.3.3 数据库的二级映像功能与数据独立性

# 数据库系统模式的概念（续）

## ❖ 模式（Schema）

- 是对数据库逻辑结构和特征的描述
- 是型的描述，不涉及具体值
- 模式是相对稳定的

## ❖ 实例（Instance）

- 数据库某一时刻的状态——模式的一个具体值
- 同一个模式可以有很多实例
- 实例随数据库中的数据更新而变动

# 数据库系统模式的概念

- ❖ “学生选课数据库” **模式**：学生、课程和学生选课3个关系模式：
  - 学生表：Student(Sno, Sname, Ssex, Sage, Sdept)
  - 课程表：Course(Cno, Cname, Cpno, Ccredit)
  - 学生选课表：SC(Sno, Cno, Grade)
- ❖ **2014年的学生选课数据库实例**：
  - 2014年学校中所有学生的记录
  - 2014年学校开设的所有课程的记录
  - 2014年所有学生选课的记录
- ❖ **2013年的学生选课数据库实例**：
  - 2013年学校中所有学生的记录
  - 2013年学校开设的所有课程的记录
  - 2013年所有学生选课的记录
- ❖ **2014年度和2013年度“学生选课数据库”模式对应的2个实例是不同的**



# 数据库系统的结构（续）

**1.3.1** 数据库系统模式的概念

**1.3.2** 数据库系统的三级模式结构

**1.3.3** 数据库的二级映像功能与数据独立性

# 数据库系统的三级模式结构（续）

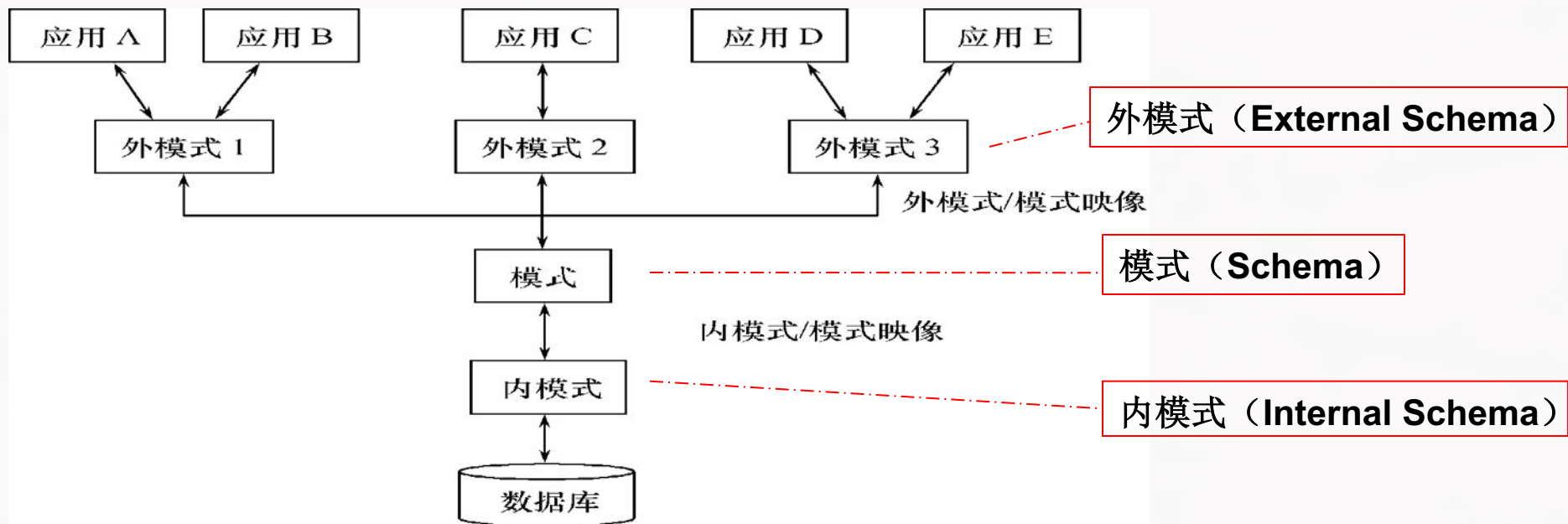


图1.16 数据库系统的三级模式结构

# 1. 模式 (Schema)

## ❖ 模式 (也称逻辑模式)

- 数据库中全体数据的逻辑结构和特征的描述
- 所有用户的公共数据视图

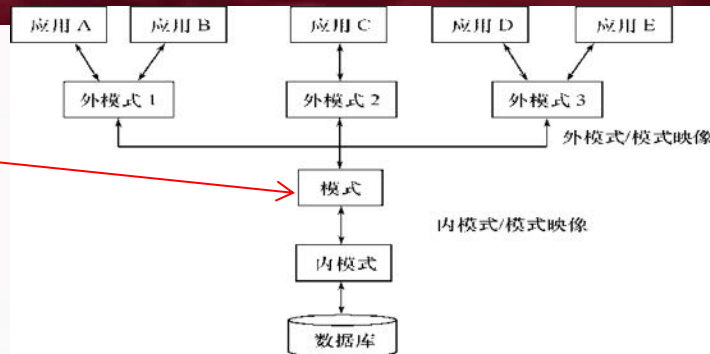
## ❖ 一般, 某个应用的数据库有一个模式

## ❖ 模式是数据库系统模式结构的中心

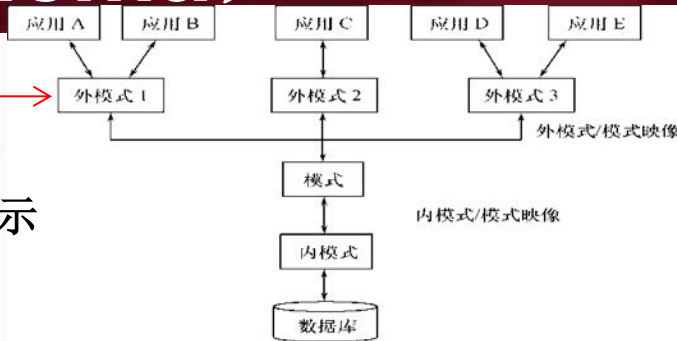
- 与数据的物理存储细节和硬件环境无关
- 与具体的应用程序、开发工具及高级程序设计语言无关

## ❖ 定义模式

- **DDL**定义数据的逻辑结构, 以某种数据模型为基础  
数据记录由哪些数据项构成, 数据项的名字、类型、取值范围等
- 定义数据之间的联系
- 定义与数据有关的安全性、完整性要求



## 2. 外模式（External Schema）



### ❖ 外模式（也称子模式或用户模式）

数据库用户使用的**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

### ❖ 外模式与模式的关系

- 外模式通常是模式的子集、一个模式可以有多个外模式  
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
- 对模式中某一数据，在不同的外模式中 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

- 一个外模式可以为 **多个** 应用系统所使用，一个应用程序只能使用一个外模式

### ❖ 外模式的用途

- 每个用户只能看见和访问所对应的外模式中的数据，简化用户视图
- 保证数据库安全性的一个有力措施

# 3. 内模式 (Internal Schema)

## ❖ 内模式 (也称存储模式)

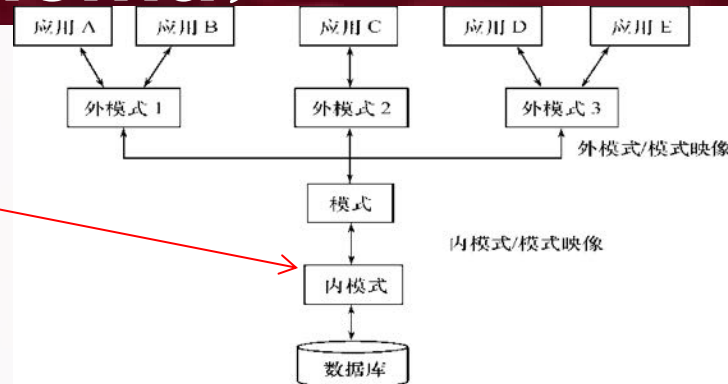
- 是数据物理结构和存储方式的描述
- 是数据在数据库内部的表示方式

- 记录的存储方式

(例如, 顺序存储, 堆存储, **Cluster**按hash方法存储等)

- 索引的组织方式 (**B+**树, **Bitmap**, **Hash**)
- 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—如定长/变长, 记录是否可以跨页存放等

## ❖ 一个数据库只有一个内模式

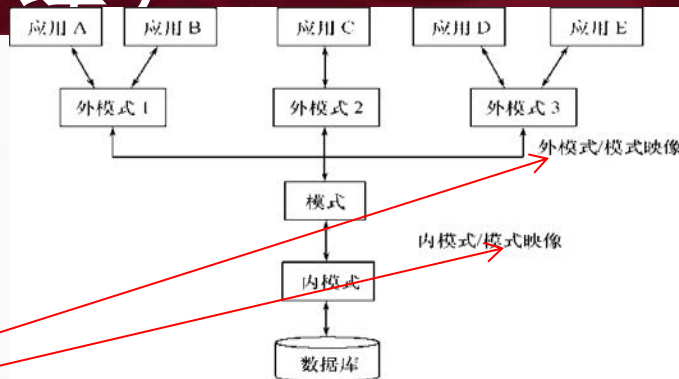


# 数据库系统结构（续）

## 1.3.1 数据库系统模式的概念

## 1.3.2 数据库系统的三级模式结构

## 1.3.3 数据库的二级映像（mapping）功能与 数据独立性



# 数据库的二级映像功能与数据独立性

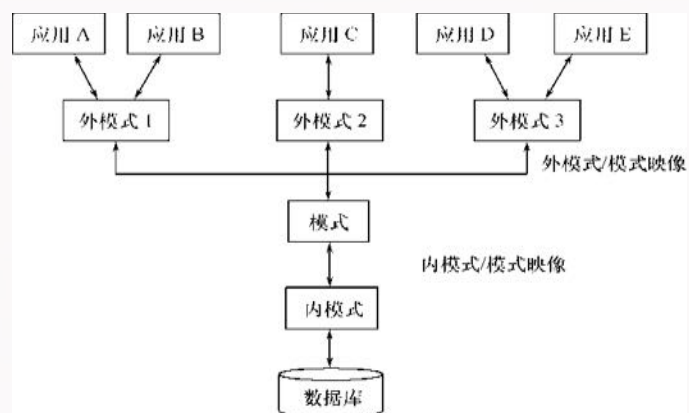
❖ 三级模式是对数据的三个抽象级别

❖ 数据库管理系统内部提供二级映像

- 外模式 / 模式映像

- 模式 / 内模式映像

❖ 三个抽象层次的联系和转换



# 1. 外模式 / 模式映像

❖ 对每一个外模式，有一个外模式 / 模式映像

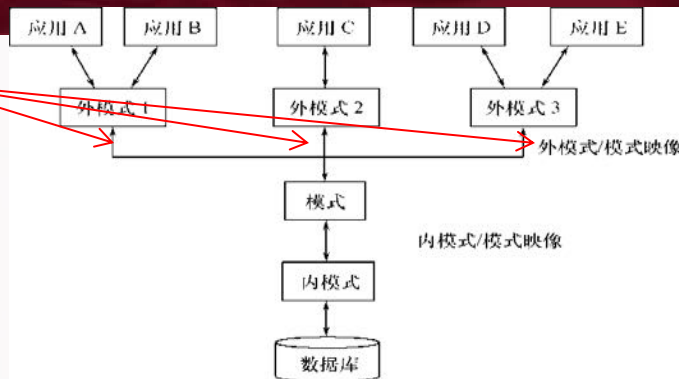
定义外模式与模式之间的对应关系

映像定义通常包含在各外模式的描述中

❖ 保证数据的逻辑独立性

■ 当模式改变时，数据库管理员对外模式 / 模式映像作相应改变，使外模式保持不变

■ 应用程序是依据数据的外模式编写的，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





## 2. 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。

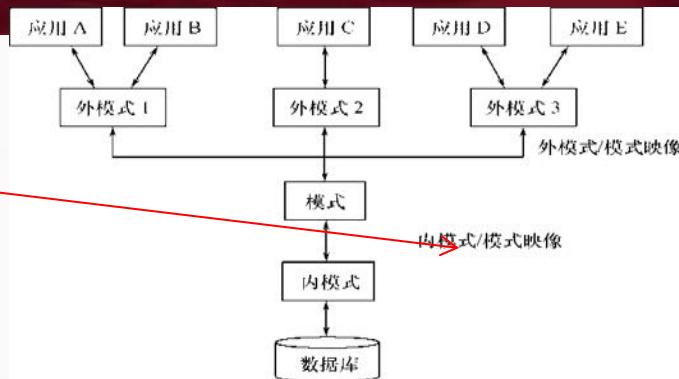
如，说明某个逻辑记录对应何种存储结构。

### ❖ 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模式描述中。

### ❖ 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 / 内模式映像，使模式保持不变。
- 模式不变，则应用程序不变。保证了数据与程序的物理独立性，简称数据的物理独立性。



# 数据库的二级映像功能与数据独立性

## ❖ 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。

## ❖ 从程序为中心——发展为 以数据为中心

具有了数据与程序之间的独立性，使得数据的定义和描述可以从应用程序中分离出去。

## ❖ 数据的存取由数据库管理系统管理

- 简化了应用程序的编制
- 大大减少了应用程序的维护和修改

# 第一章 绪论

## 1.1 数据库系统概述

## 1.2 数据模型

## 1.3 数据库系统的结构

## 1.4 数据库系统的组成

## 1.5 小结

# 1.4 数据库系统的组成

❖ 数据库

❖ 数据库管理系统（及其开发工具）

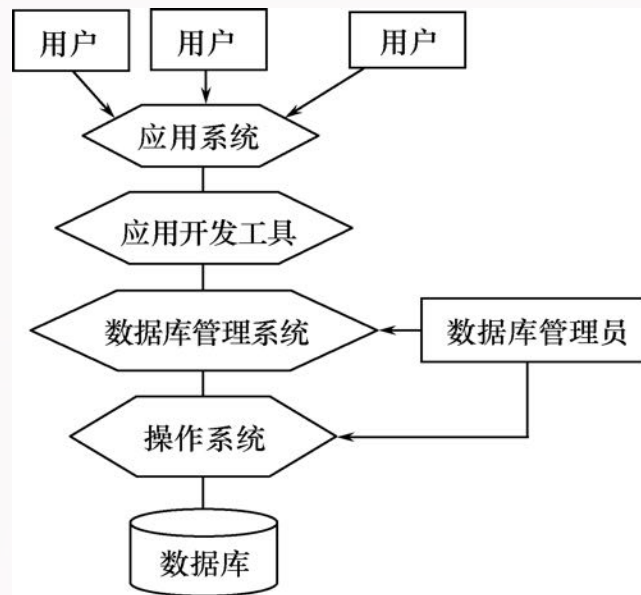
❖ 应用程序

❖ 数据库管理员

1. 硬件平台及数据库

2. 软件

3. 人员



# 1. 硬件平台及数据库

## ❖ 硬件平台及数据库

### 数据库系统对硬件资源的要求

- 足够大的内存
- 足够的大的磁盘或磁盘阵列等外部设备
- 较高的通道能力，提高数据传送率

## 2. 软件

### ❖ 数据库管理系统

❖ 支持数据库管理系统运行的操作系统

❖ 与数据库接口的高级语言及其编译系统

❖ 以数据库管理系统为核心的应用开发工具

❖ 为特定应用环境开发的数据库应用系统

### 3. 人员

- ❖ 数据库管理员
- ❖ 系统分析员和数据库设计人员
- ❖ 应用程序员
- ❖ 最终用户

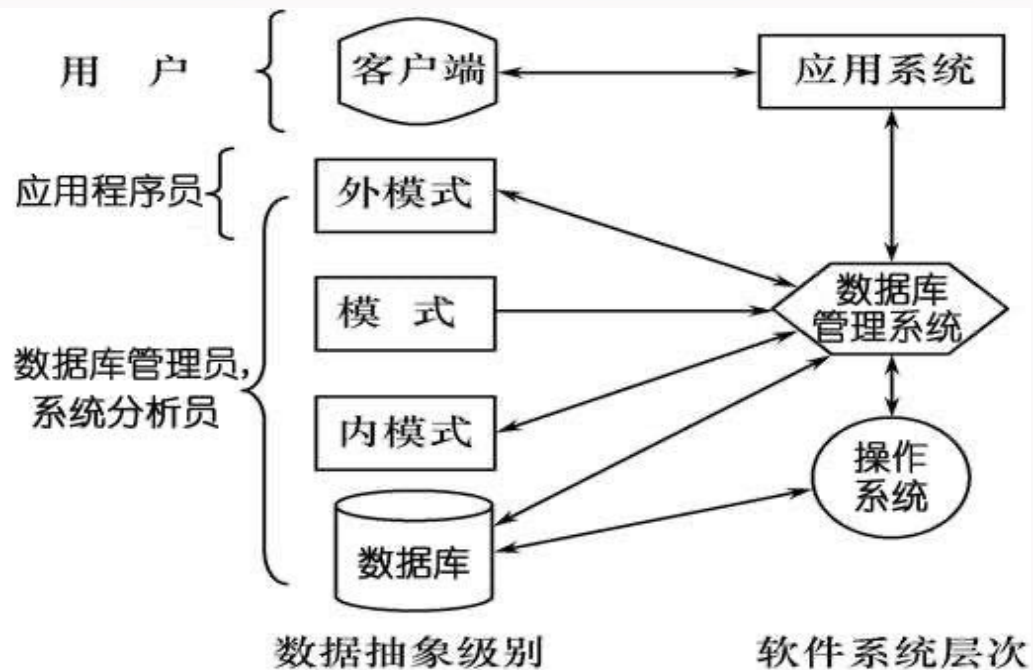


图1.17 各种人员的数据视图

不同的人员涉及不同的数据抽象级别，具有不同的数据视图

# 1. 数据库管理员 (DBA)

## 职责:

1. 参与确定数据库中的信息内容和结构
2. 参与数据库的存储结构和存取策略的设计
3. 参与确定数据安全性要求和完整性约束条件
4. 监控数据库的使用和运行
  - 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

5. 数据库的改进和重组
  - 性能监控和调优
  - 定期对数据库进行重组，以提高系统的性能
  - 需求增加和改变时，数据库须需要重构造



## 2. 系统分析员、数据库设计人员    3.应用程序员

### 2. 系统分析员

- 负责应用系统的需求分析和规范说明
- 与用户及数据库管理员结合，确定系统的硬软件配置
- 参与数据库系统的概要设计

### 2. 数据库设计人员

- 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

### 3. 应用程序员

- 设计和编写应用系统的程序模块
- 进行调试和安装

# 4. 用户

用户是指最终用户（**End User**）。他们通过应用系统的用户接口使用数据库。

## 1. 偶然用户

- 不经常访问数据库，但每次访问数据库时往往需要不同的数据库信息
- 企业或组织机构的高中级管理人员

## 2. 简单用户

- 主要工作是查询和更新数据库
- 银行的职员、机票预定人员、旅馆总台服务员

## 3. 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应用程序接口编制自己的应用程序

# 第一章 绪论

## 1.1 数据库系统概述

## 1.2 数据模型

## 1.3 数据库系统的结构

## 1.4 数据库系统的组成

## 1.5 小结

# 第一章 绪论小结

## 本章目标：

学习本章的目的在于了解基本知识，初步掌握基本概念，  
为以后的学习打好基础。

## 重点：

牢固掌握概念模型的基本概念；数据模型的3个组成要素；  
数据库系统三级模式和两层映像的体系结构；  
数据库系统的逻辑独立性和物理独立性等。

## 难点：

介绍了大量的基本概念。

数据模型及数据库系统的体系结构也是本章的难点。