

# 数据库系统概论

**An Introduction to Database System**

**中国人民大学信息学院**

**School of Information,  
Renmin University of China  
2015**

# 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

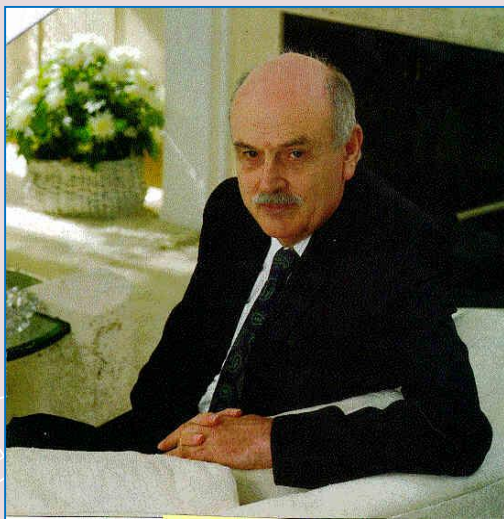
DBMS及其相关工具产品、应用套件、解决方案

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

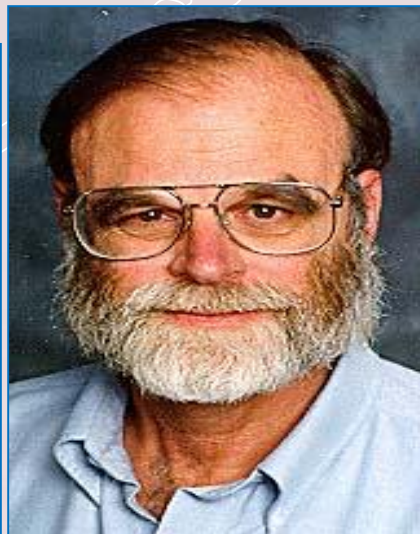
# 四位图灵奖得主



C.W.Bachman (1973)



E.F.Codd (1981)



James Gray (1998)



M.R.Stonebraker (2014)

参考：《数据库系统概论（第五版）习题解析与实验指导》  
附录A 数据库领域图灵奖获得者



# 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

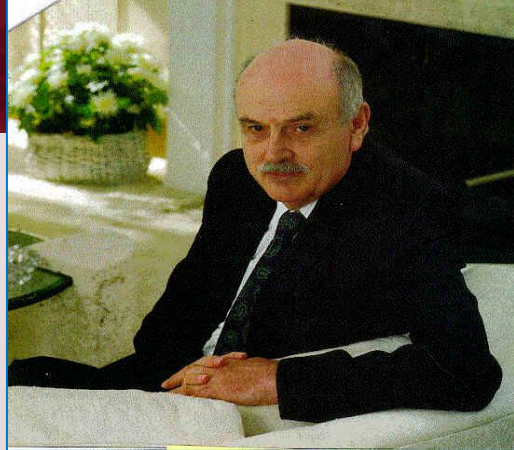




# Edgar F.Codd 博士

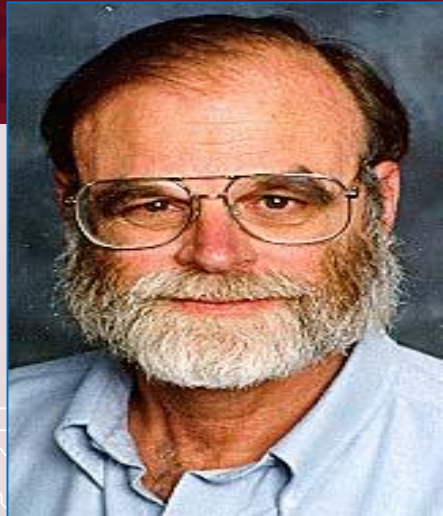
## 关系数据库之父 美国工程院院士

- ❖ 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





# 我国数据库发展历程

70年代	看	引进	1976年萨师煊教授将数据库概念引入国内，在全国讲学，在人大开设数据库课
80年代	学	请进来 走出去	外国专家来华讲学 中国专家出国进修
90年代	赶	研究 开发 应用	国家攻关、863高技术项目、国家自然科学基金等等
21世纪	创	创新研究 产品开发 应用集成	创新研究 与国际研究同步 立足应用、 开发自主的数据管理系统、

在数据库技术发展的起跑线上，中国晚了约15年



中国数据库开创者—萨师煊，中国人民大学教授（1922-2010）



# 教材及参考书

## 📖 教材

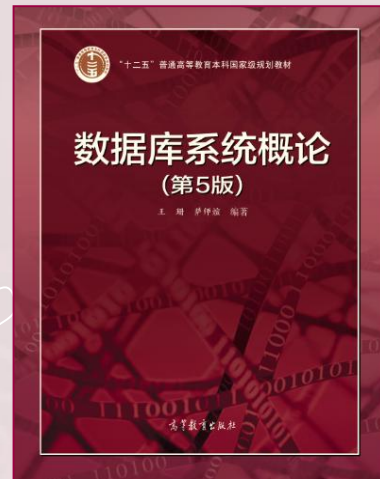
王珊,萨师煊. **数据库系统概论（第5版）** 高等教育出版社, 2014.9

## 📖 上机软件

**金仓数据库KingBaseES**

 北京人大金仓信息技术股份有限公司  
Beijing Kingbase Information Technologies Inc.

<http://www.Kingbase.com.cn/>



📖 中国人民大学 数据库系统概论精品课程网站：

<http://chinadb.ruc.edu.cn>

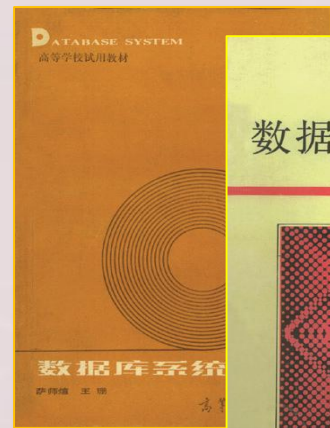
📖 高教出版社易课程网站：

<http://abook.hep.com.cn/187532>

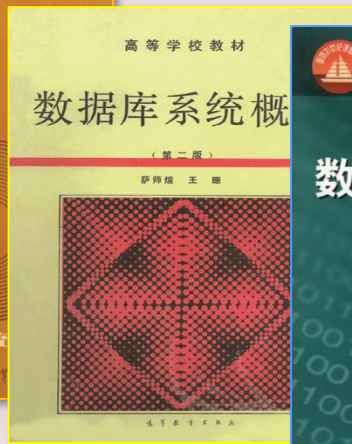




# 《数据库系统概论》--国家级优秀教材



第一版  
1983



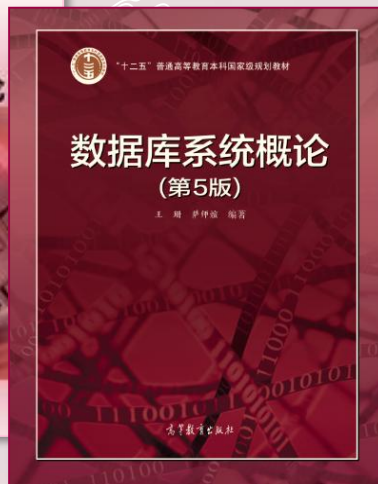
第二版  
1991



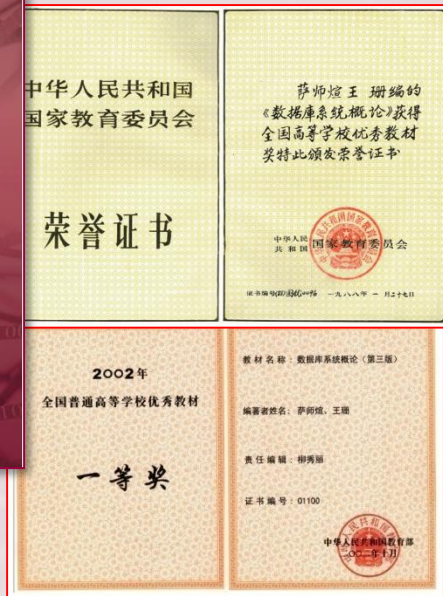
第三版  
2000



第四版  
2006



第五版  
2014



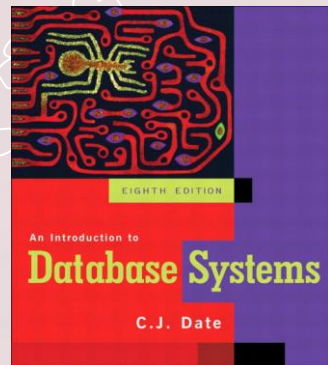
《数据库系统概论》已经出版第五版，**2次获得国家级优秀教材奖**，成为一本优秀的经典教材。为培养我国数据库专业人才、推动我国数据库技术的发展做出了突出贡献。



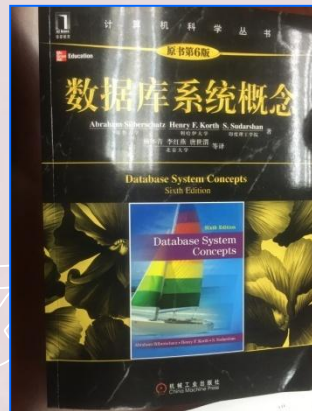
# 教材及参考书

## 参考书


 **An Introduction to Database System**  
Date C J, Addison-Wesley



 **A First Course in Database Systems**  
Jeffrey.D.Ullman, Jennifer Widom



Dept. Of Computer Science Stanford University

 **Database System Concepts**

Silberschatz A, Korth Henry F, Sudarshan S



# 进度安排

## 数据库系统概论（基础篇）

数据库系统概论 基础课	讲授视频	习题与实验
1. 绪论	8个	习题
2. 关系数据库系统	7个	习题、实验准备
3. SQL语言	12个	习题、实验
4. 数据库安全性	4个	习题、实验
5. 数据库完整性	5个	习题、实验



# 进度安排

## 数据库系统概论（高级篇）

数据库系统概论 高级课	讲授视频	习题与实验
6. 关系数据理论	待定	习题
7. 数据库设计	待定	大作业
8. 数据库编程	待定	习题、实验
9. 关系查询处理和查询优化	待定	习题、实验
10. 数据库恢复技术	待定	习题
11. 并发控制	待定	习题



# 师资队伍

## 授课教师

- 王珊教授
- 杜小勇教授
- 陈红教授



## 习题和实验教师

- 卢卫副教授
- 焦敏高工







# 長白山天池

# 数据库系统概论

**An Introduction to Database System**

**中国人民大学信息学院**

**School of Information,  
Renmin University of China  
2015**

# 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



# 数据库在计算机系统中的位置





# 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。

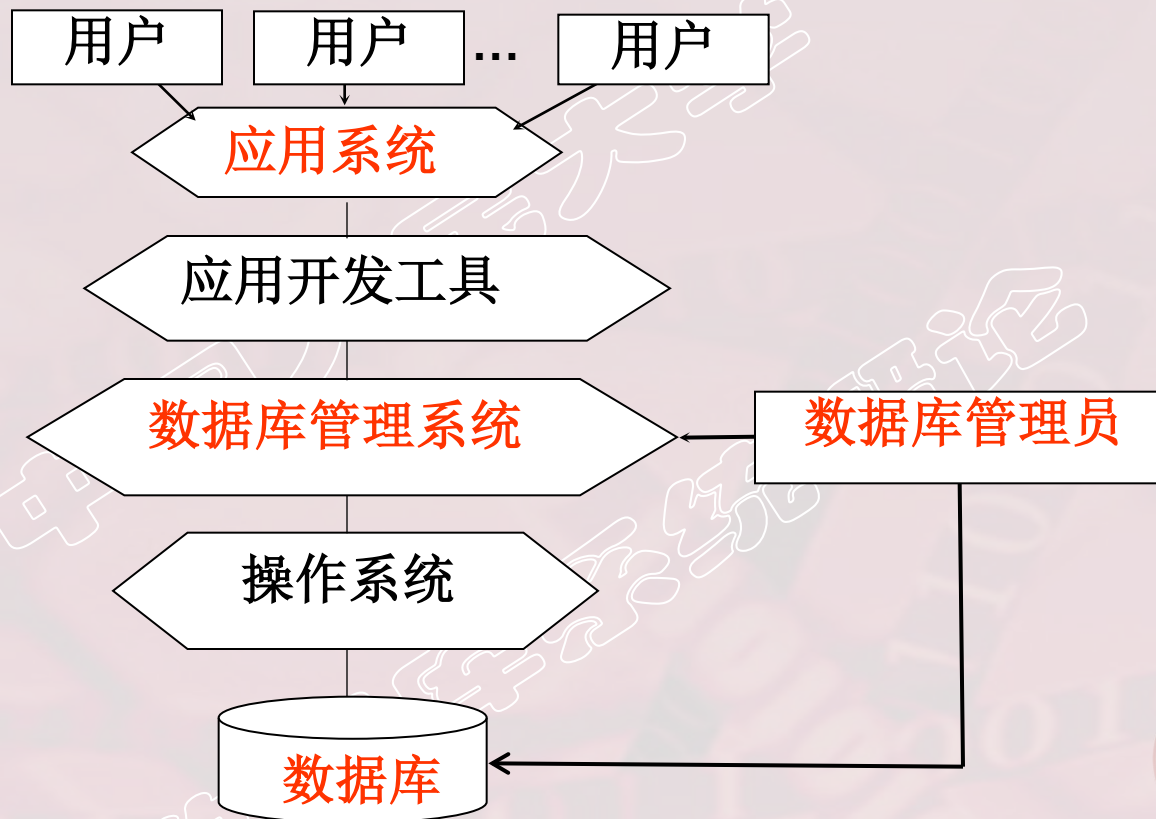
## ❖ 数据库系统的构成

- 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

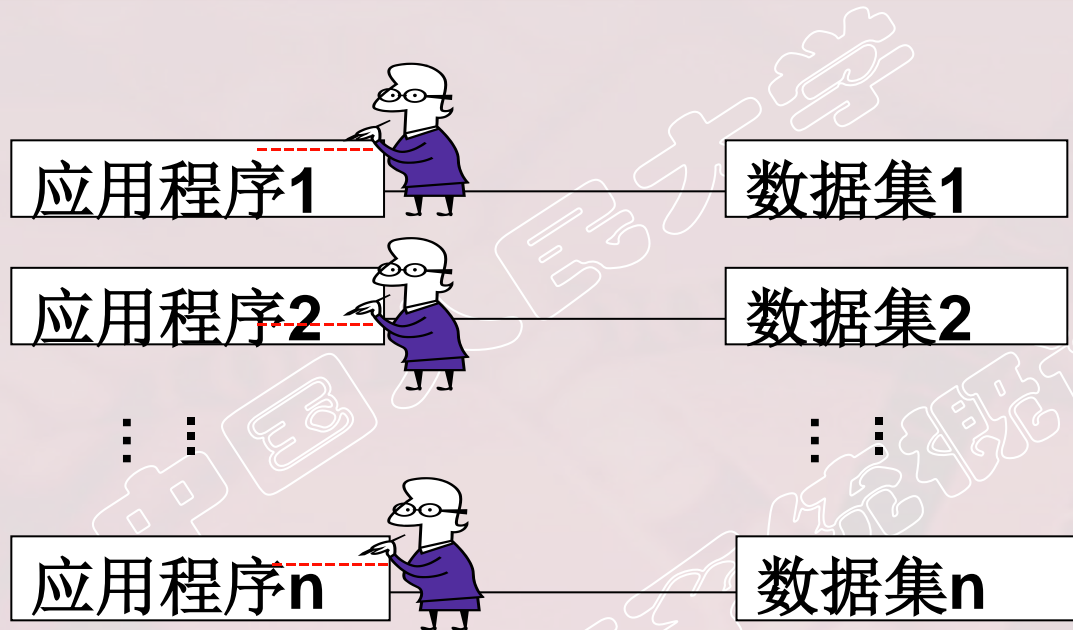
- 人工管理阶段（**20世纪50年代中之前**）
- 文件系统阶段（**20世纪50年代末--60年代中**）
- 数据库系统阶段（**20世纪60年代末--现在**）





		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， 批处理	联机实时处理，分布处理， 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

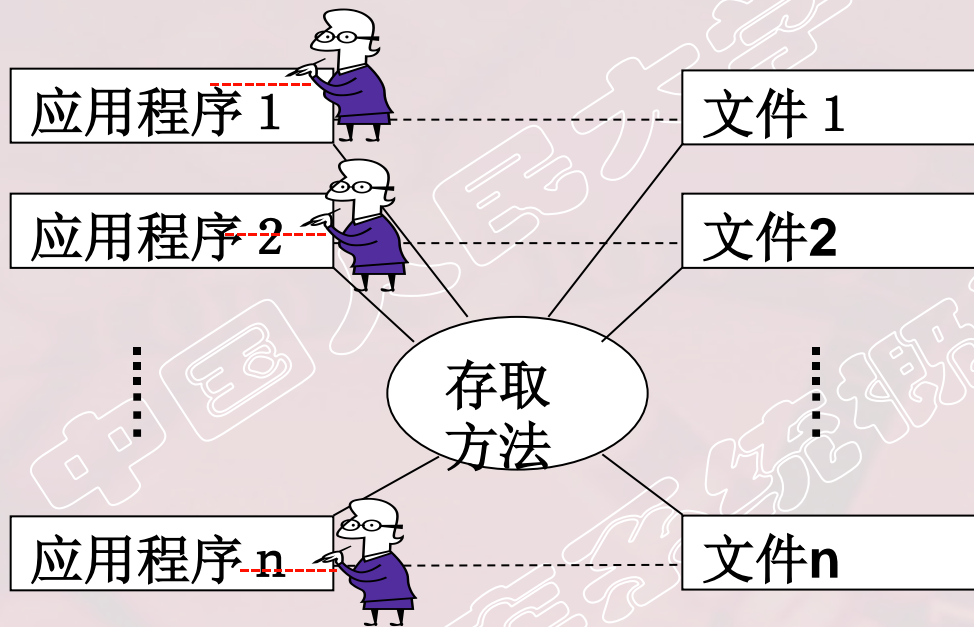
# 应用程序与数据的对应关系（人工管理阶段）



人工管理阶段 应用程序与数据之间的对应关系



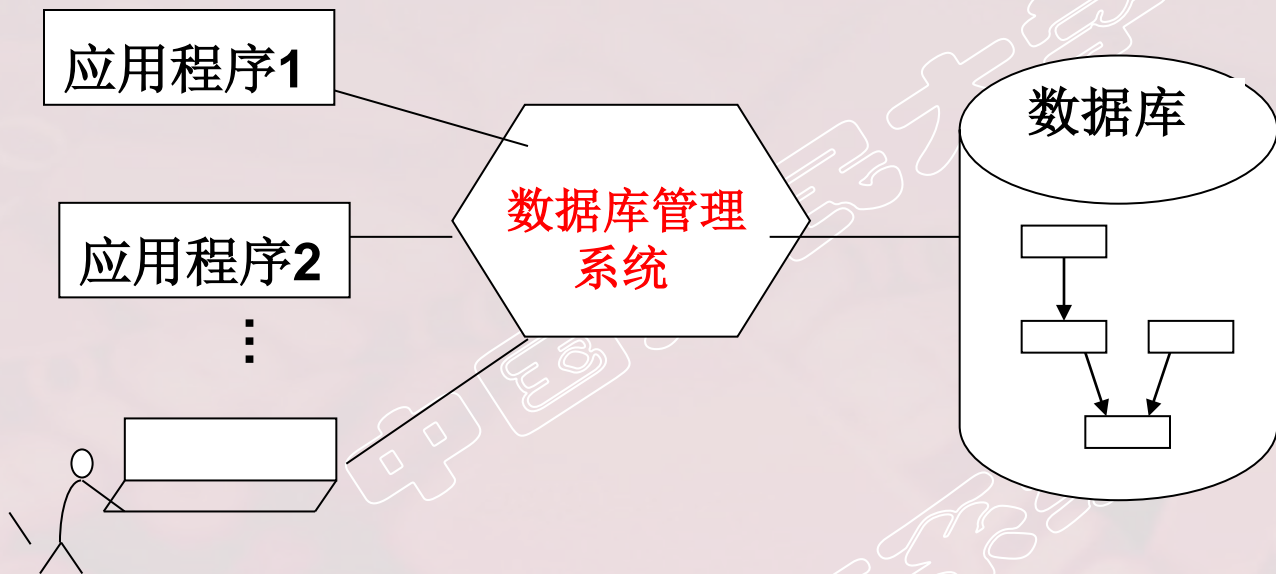
# 应用程序与数据的对应关系（文件系统阶段）



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



# 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

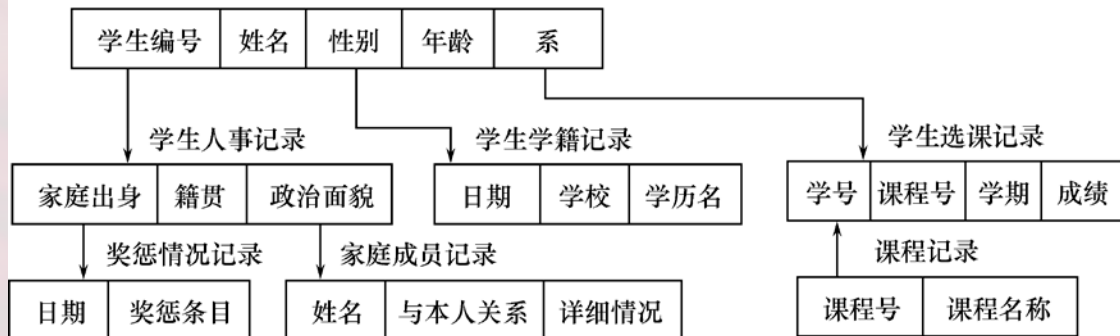


图1.5 某学校信息系统中的学生数据

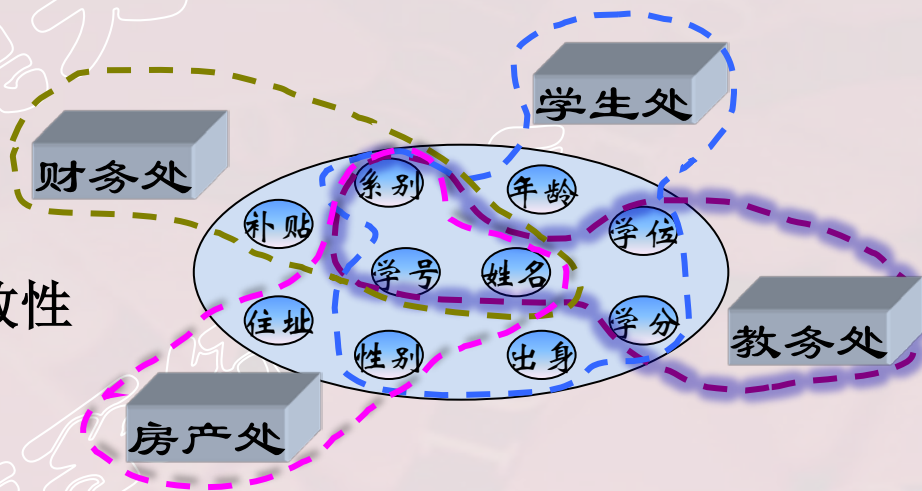


# 数据的共享性高，冗余度低且易扩充

❖ 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。





