

# 数据库系统原理

教程：数据库系统概论（第5版）

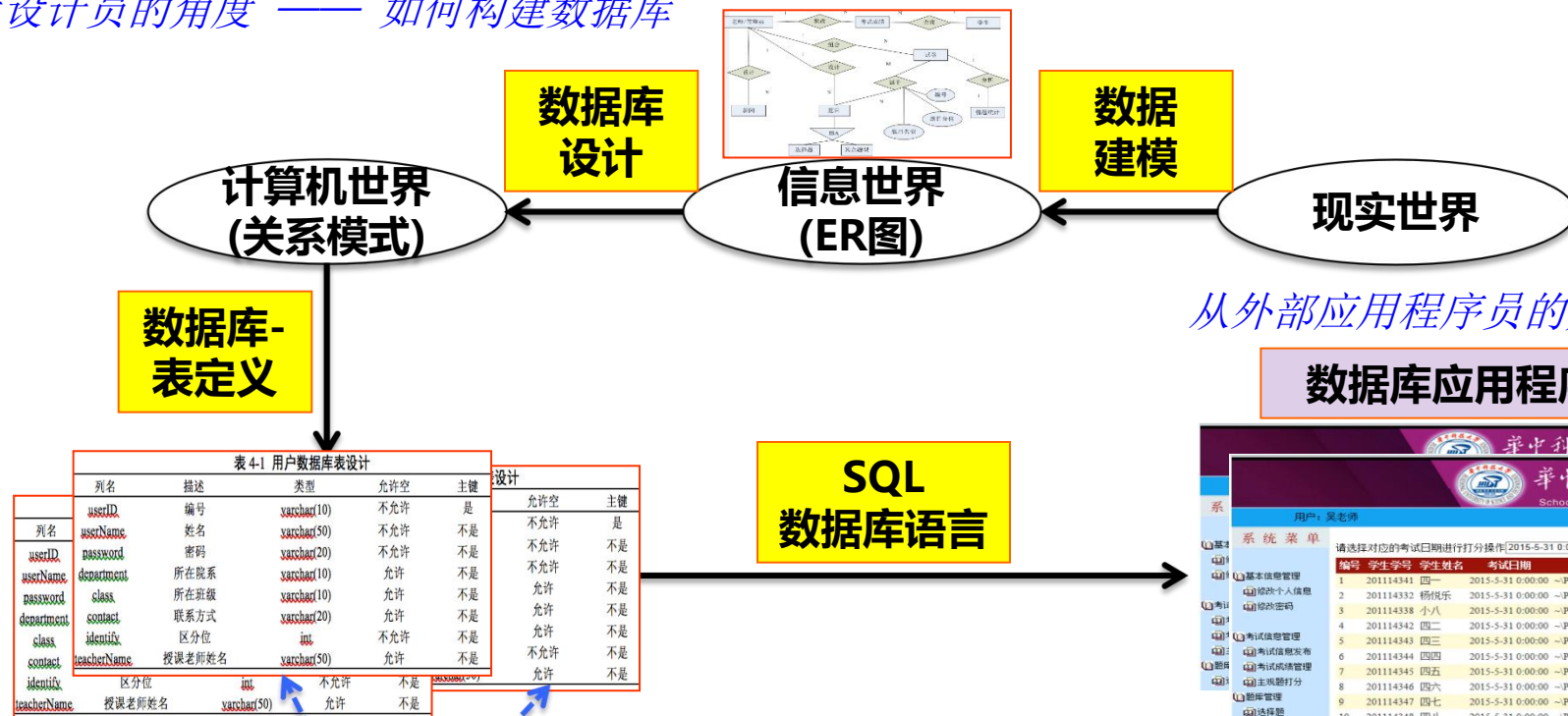
结合：CMU 15-445/645 INTRO TO DATABASE SYSTEMS

华中科技大学 计算机学院

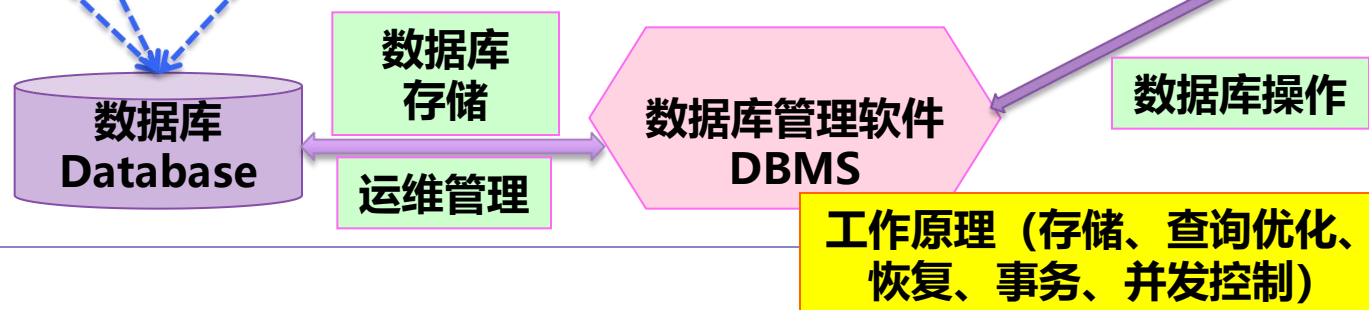
左琼

# 数据库系统原理课程规划

从数据库设计员的角度 —— 如何构建数据库



从内部系统组成的角度 —— 如何实现数据库系统软件



# 课程目标

- 了解并结合关系型数据库系统，深入理解数据库系统的基本概念，结构、工作原理和运行机制。
- 掌握关系数据模型及关系数据语言，能熟练应用SQL语言表达各种数据操作。
- 掌握E-R模型的概念和方法，关系数据库规范化理论和数据库设计方法，通过上机实践的训练，初步具备进行数据库应用系统开发的能力。
- 了解数据库系统的存储机制、查询优化、恢复、并发控制等技术以及最新的数据库技术。

# 教材及参考书

## □ 教材：

- 王珊，萨师煊：数据库系统概论(第五版)，高教出版社，2014.9

## □ 参考资料：

- CMU 15-445/645 — INTRO TO DATABASE SYSTEMS
- 数据库系统概念Database System Concepts. Abraham, Silberschatz等著，杨冬青，唐世渭等译，机械工业出版社，任一版
- 杨冬青，唐世渭等译。数据库系统实现（Database System Implementation），原书第二版。机械工业出版社，2010
- 岳丽华，金培权等译。数据库系统基础教程（A First Course in Database Systems），原书第三版。机械工业出版社，2011
- 其它 数据库系统原理教程等

# 数据库系统原理实践

- 时间：10-16周（4次线上，4次线下）
- 上机：
  - Educoder（MySQL 8.0）
  - 内容包括：SQL练习、数据库设计、并发控制、B+树实现、部分数据库应用Java开发函数等
- 实践课计分：
  - Educoder成绩
  - 平时成绩（平时记录、平时检查）
- /\* 请同学们提前(第三章时)下载mysql 8.0，方便SQL练习 \*/

# 线上学习资源

- 华中科技大学计算机学院 **“数据库系统原理” 慕课** (第8次开课)
    - <https://www.icourse163.org/course/HUST-1449788170>
    - 用于课前预习、学习互动、课后复习，内含丰富的客观题题库，欢迎同学们按章节刷题，加深对知识点的理解。（不算到平时成绩中）。
  - **华为在线课程系列**
    - <https://edu.huaweicloud.com/roadmap/colleges.html>
    - 华为在线课程数据库系列有两类入口：
      - 一类通过在线课程的技术领域选择“数据库”进入；
      - 一类通过技术领域选择“鲲鹏”后，查找openGauss课程进入。
- 有兴趣的同学可以选看！

# 教学组织及成绩评定

- **学在华科大数据库课堂**（平时作业、课件）；
- **微助教**（签到） 课堂编号：MT429；
- 请推荐一位班干部协助管理课堂和班级，建立**QQ群**，方便课程日常交流及通知发布。
- **成绩评定**
  - 总评 = 平时成绩15% + 随堂测2次 15% + 期末考试卷面成绩\*70%



**学在华科大课程编号：**  
**0803574**



# 第一章 绪论

*Principles of Database Systems*



# 第一章 绪论

## 1.1 数据库系统概述

## 1.2 数据模型

## 1.3 数据库系统结构

## 1.4 数据库系统的组成

## 1.5 小结

### 本章思考：

- 什么是数据库系统？
- 产生DBMS的动机是什么？
- 如何设计数据库？
- DBMS的结构？
- 数据库系统的组成？

# 1.1 数据库系统概述

## 1.1.1 四个基本概念

- 数据(Data)
  - 数据库(Database)
  - 数据库管理系统(DBMS)
  - 数据库系统(DBS)
- 
- 为什么要学习四个基本概念?
    - 数据库={数据集, 操作集, 限制集}, 理解其中每个元素。
    - 不同人员, 不同视角。

# 1. 数据 (Data)

- **数据**的定义: 描述事物的符号记录。
- 数据的种类: 数字、文本、图形、图像、音频、视频等。

例1: “2023年” , “两千零二十三年” 。

- 数据的特点: **数据与其语义不可分**。

例2: (0005794, 601, 周济, 1, 19460826, 01)

- 语义1: (工号, 部门编号, 姓名, 性别, 生日, 民族)
- 语义2: (... , 办公室门牌 / 内部电话号, ....)

- **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。**

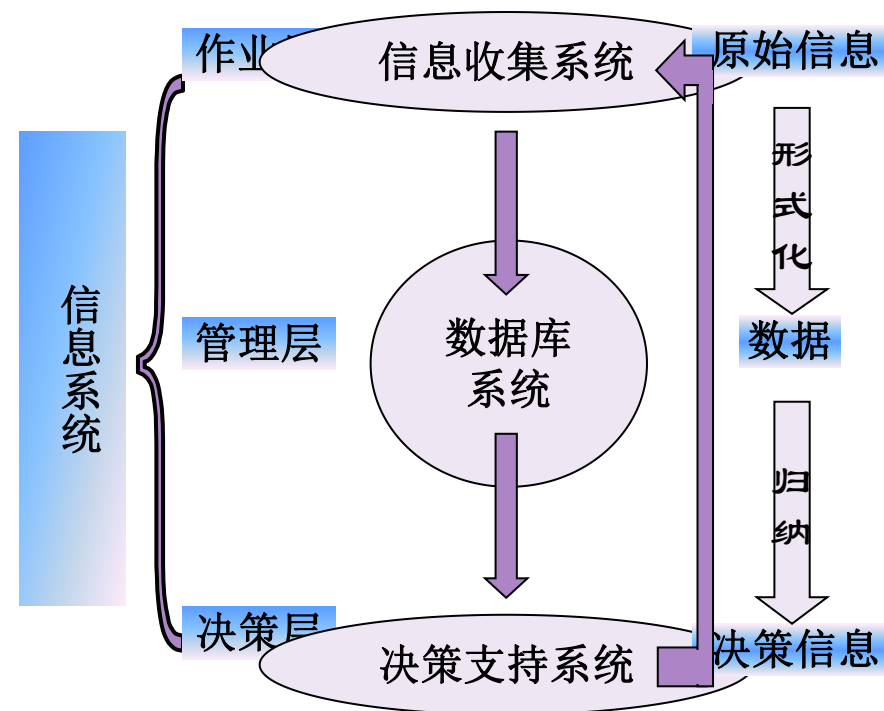
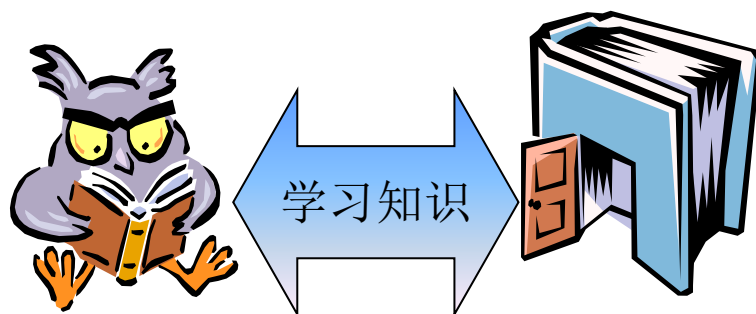
## 数据与信息的区别:

- 数据可以表示信息, 但不是任何数据都能表示信息, 同一数据也可以有不同的解释;
- 信息是抽象的, 不随数据形式而改变; 而同一信息可以有不同的数据表示方式。

# 更多的思考：信息与知识

## □ 知识

- “行动的能力”；
- “可以辅助我们作出决策或采取行动的有很高价值的一种信息形态”；
- 可以很快地将信息从一个地方转移到另一个地方，但将知识从一个人转移给另一个人经常是很困难和很慢的



## 2.数据库 (DB)

### □ 数据库 (DB, DataBase) 的定义:

- DB是长期存储在计算机内、有组织的、可共享的大量数据的集合。
- 数据是数据库中存储的基本对象。

### □ 数据库的基本特征:

- 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;
- 数据独立性较高;
- 易扩展。



### 3. 数据库管理系统 (DBMS)

- DBMS: DataBase Management System
- DBMS是**对数据库进行管理的****大型系统软件**，它是数据库系统的**核心**组成部分。
  - 位于**用户与操作系统之间**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。
- DBMS的用途：
  - 科学地组织和存储数据、高效地获取和维护数据。
- 用户在数据库系统中的一切操作，包括数据定义、查询、更新及各种控制，都是通过DBMS进行的。

# DBMS的主要功能

## □ 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的组成和结构 (逻辑结构、存储结构)

## □ 数据组织、存储和管理

- 分类组织、存储和管理各种数据 (数据字典、用户数据、数据存取路径等)
- 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

- 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)

## □ 数据库的事务管理和运行管理

- 数据库的建立、运行和维护
- 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

## □ 数据库的建立和维护功能(实用程序)

- 数据库初始数据装载
- 数据库转储
- 介质故障恢复
- 数据库的重组织
- 性能监视分析等

## □ 其它功能

- DBMS与网络中其它软件系统的通信
- 两个DBMS系统的数据转换
- 异构数据库之间的互访和互操作

# 什么是数据库管理系统?

□ 从OS角度看, DBMS是:

一个相互关联的数据集合, 一组访问这些数据的程序。

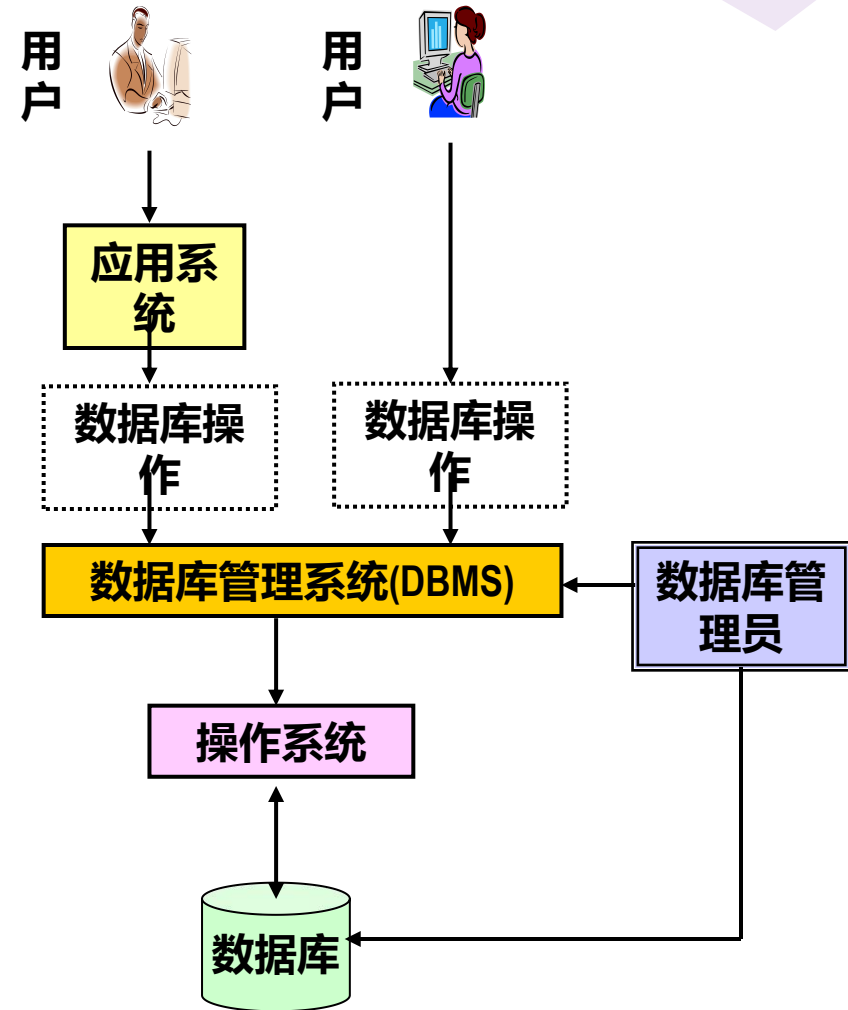
□ DBMS的目标是什么?

- 为应用提供一个方便、有效的数据支撑环境;
- 数据库技术是一种资源抽象技术, 将文件资源抽象成一种能存放结构化、关联性信息的虚拟数据库资源;
- 从资源共享角度考虑, DBMS应考虑数据一致性、操作原子性及异常处理等性能问题。



# 4.数据库系统 (DBS)

- DBS, DataBase System
- 数据库系统的构成:
  - 数据库
  - 数据库管理系统 (及其开发工具)
  - 应用系统
  - 数据库管理员 (DBA, DataBase Administrator)



## 1.1.2 数据管理技术的产生和发展

### □ 数据管理 ≠ 数据处理:

- 对数据进行分类、组织、编码、存储、检索和维护
- 是数据处理的中心问题

### □ DBMS的产生动机:

- 应用需求的推动
- 计算机硬件的发展
- 计算机软件的发展

推动计算机发展, DBMS出现

# 1.1.2数据管理技术的产生和发展

## □ 数据管理技术的发展过程:

### 手工管理

手工管理 (直接书写机器语言)

```
00101011 11100101 11001001
10101100 11010100 11110000
01010010 10010000 10000000
```

### 文件管理 VS. 数据库管理

文件管理 (程序管理数据)

```
main () {
    int a, b, c;
    fopen(... ..);
}
```

数据库管理 (自主管理信息)

```
select * from S
insert
delete
```



OS




DBMS




- 各阶段的技术及应用背景
  - 计算机应用范围
  - 外存储设备
  - 数据管理软件
- 各个阶段的差别体现
  - 谁管理数据
  - 数据面向谁
  - 数据与应用的独立性

# 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  
数据量小、结构简单，如高阶方程、曲线拟和等
  - 硬件水平： 无直接存取存储设备  
外存为顺序存储设备，如：磁带、卡片、纸带，没有磁盘等
  - 软件水平： 没有OS，只有汇编语言
  - 处理方式： 数据批处理

# 1.人工管理阶段

- 特点：程序与数据是一个整体，一个程序中的数据无法被其他程序使用，因此程序与程序之间存在大量的重复数据。
- 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

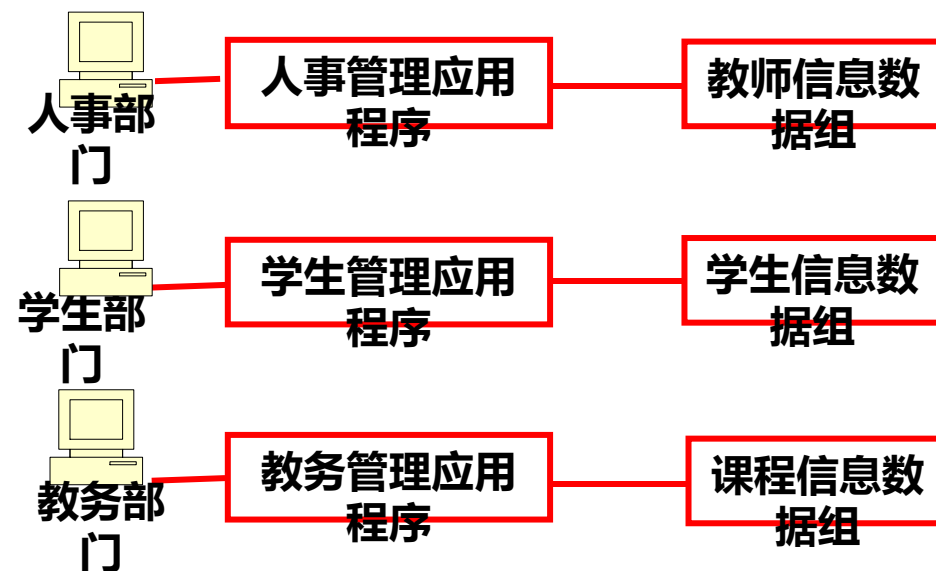


图 应用程序和数据的  
依赖关系

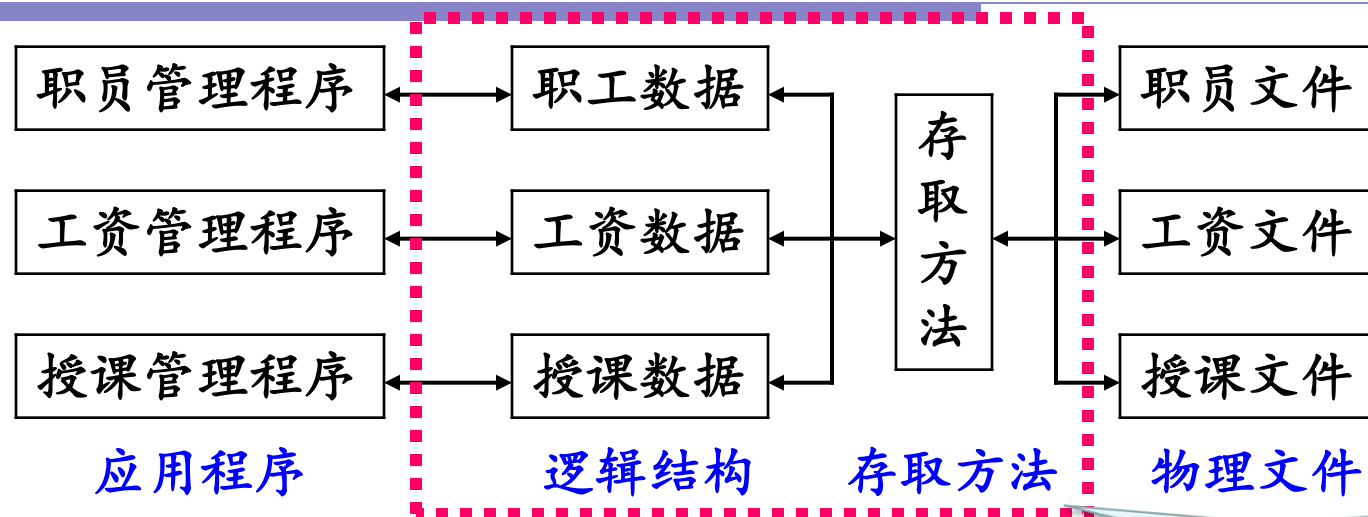
## 2. 文件系统阶段

- 时期: 20世纪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: 外存有了磁盘、磁鼓
  - 软件水平: 有OS和文件系统
  - 处理方式: 联机实时处理、文件批处理

## 2. 文件系统阶段

- 程序通过**数据文件**访问数据。文件系统阶段由操作系统的**文件系统**对数据进行管理，用户无需直接涉及物理设备及数据的物理组织细节。
- **文件系统提供一定的数据管理功能：**
  - 存取方法（索引、链接、直接存取、倒排文件等）；
  - 支持对文件的基本操作（增、删、改、查等），用户程序不必考虑物理细节；
  - 数据的存取基本上以**记录**为单位。
- 数据仍是**面向应用的**，一个数据文件对应一个用户程序。
- **数据与程序有一定的独立性**：实现了用户程序与数据的**物理存储结构**的分离。
- 数据可在文件级为**多用户共享**。

# 开发一个教职员管理系统，采用文件系统如何做？



数据独立性问题

文件系统fopen

## 职工数据

职工号	姓名	单位	性别	年龄	工龄	职称	工资
-----	----	----	----	----	----	----	----

## 工资数据

数据共享问题

数据一致性问题

职工号	单位	姓名	职称	工龄	工资	房租	水电
-----	----	----	----	----	----	----	----

数据结构化问题

## 教课数据

职工号	姓名	单位	职称	课程名	学时
-----	----	----	----	-----	----



# 文件系统管理数据的缺陷

## □ 存在问题：

- **数据完整性问题**：数据库中数据必须满足一致性约束，如学号唯一，只能通过应用程序来保证。
- **故障恢复**：计算机系统出现异常，数据库中数据应恢复到故障前状态；文件系统不保证这种原子性问题。例如，转账问题。
- **共享数据并发访问异常**：文件系统只提供互斥访问共享数据，不考虑并发中的一致性保证。如，对同一账号的两个并发取款操作。
- **安全性问题**：数据库中共享数据要求更高的共享保护，文件系统中由于应用程序与数据密切结合，难于保证。

# 如何解决文件系统缺陷？

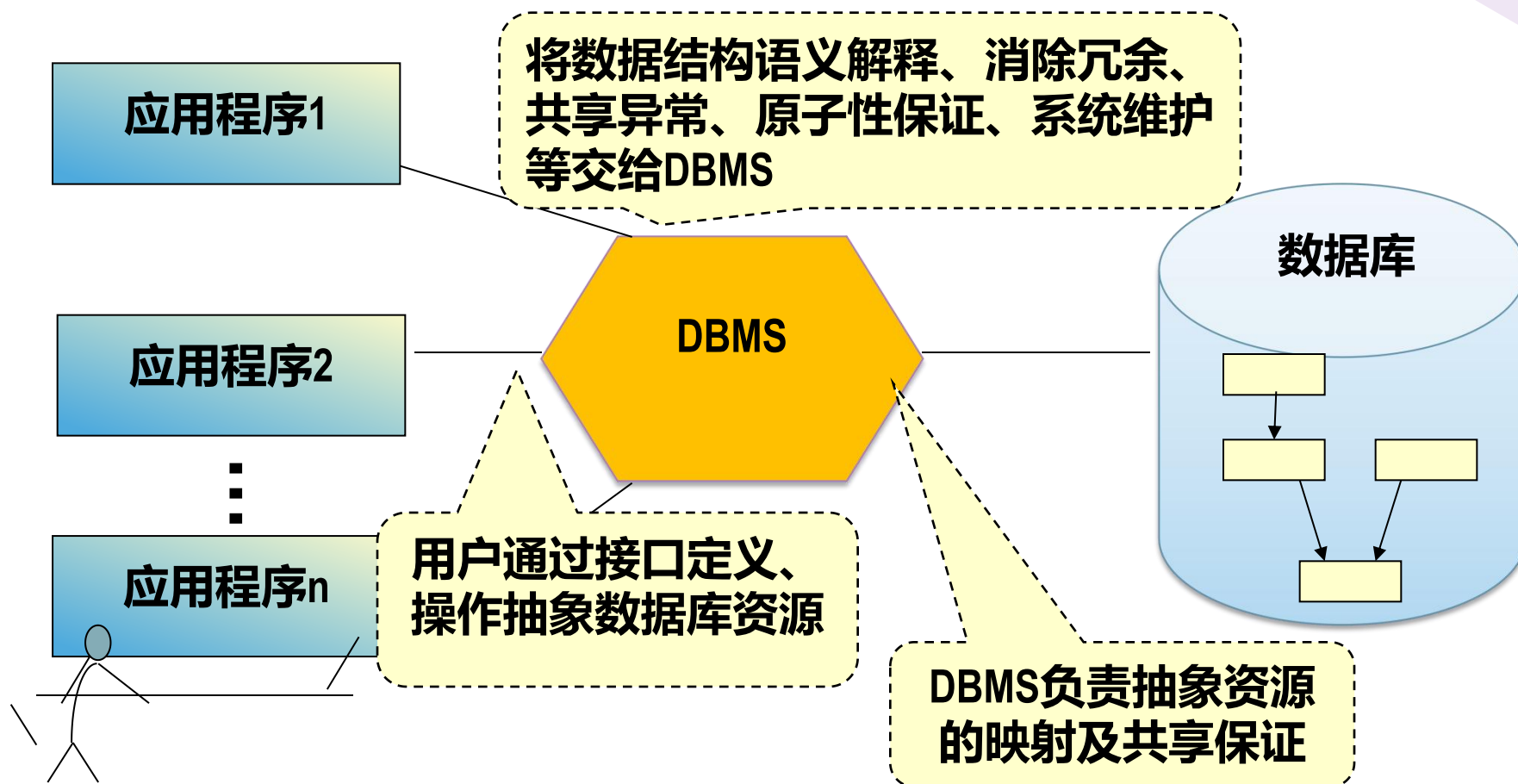
## □ 解决方案：

- 开发一种比文件系统更加强大的数据管理系统，将数据的数据结构语义、关联关系、冗余消除等交给该系统管理，并提供一次一集合的数据访问方式。
- 使用资源抽象、资源共享等虚拟技术，自动保证数据并发共享访问的安全性、隔离性。每个应用对各自的虚拟数据库操作，系统保证虚拟资源到物理资源的映射，保证物理数据库的一致性、完整性。
- 能自动解决共享异常、原子性、完整性问题。

### 3.数据库系统阶段

- 时期：20世纪60年代末以来
- 产生背景：
  - **应用背景**：大规模管理：计算机管理的数据量大，关系复杂，共享性要求强（多种应用、不同语言共享数据）
  - **硬件背景**：大容量磁盘、磁盘阵列、光盘
  - **软件背景**：有**数据库管理系统**；软件价格上升，硬件价格下降，编制和维护软件及应用程序成本相对增加，求降低
  - **处理方式**：联机实时处理,分布处理,批处理

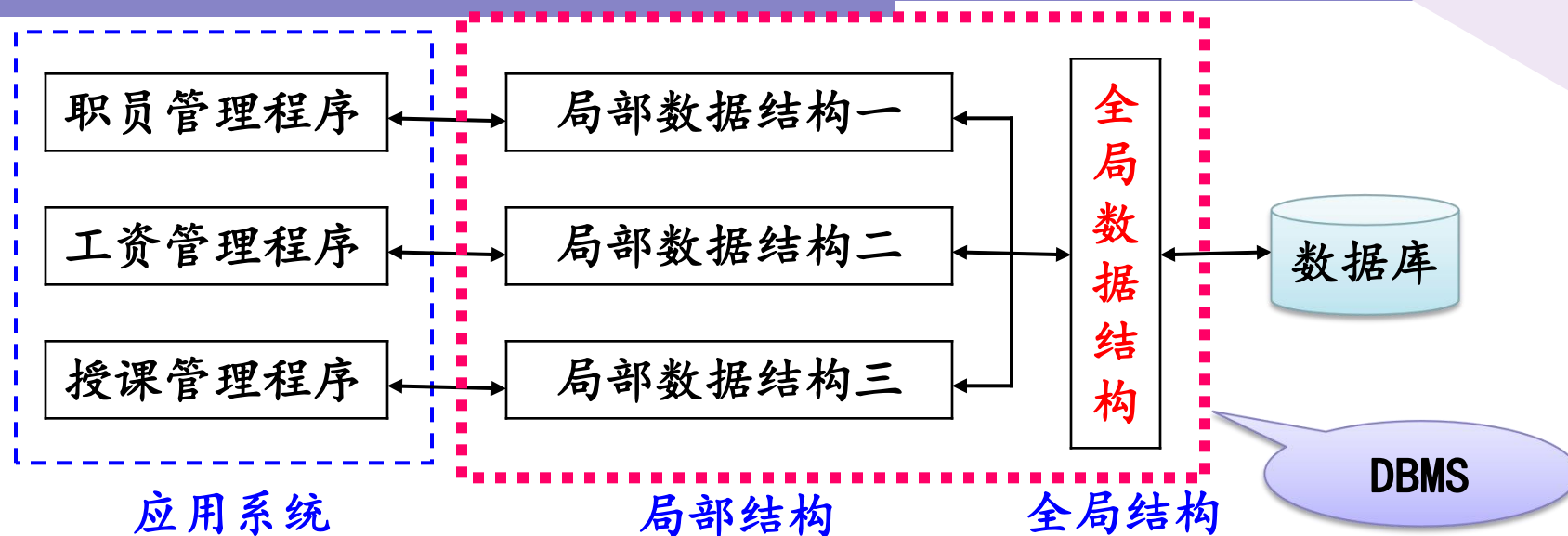
# 应用程序与数据的对应关系(数据库系统)



第四代编程 1、设计数据库、建立数据库 2、使用数据库

图 数据库系统阶段应用程序与数据之间的对应关系

# 开发一个教职员管理系统，采用数据库系统如何做？



## 全局数据结构范例一

工号	姓名	单位	性别	年龄	工龄	职称	工资	房租	水电	课号	学时
----	----	----	----	----	----	----	----	----	----	----	----

## 全局数据结构范例二

职工号	单位	姓名	性别	年龄	职称
-----	----	----	----	----	----

职工号	工龄	工资	房租	水电
-----	----	----	----	----

职工号	课程名	学时
-----	-----	----

# 数据库系统的目的

## □ 数据库系统的主要目的之一：

给用户提**整体数据的抽象视图**，将磁盘上的所有物理数据集合抽象成 **整体结构化的虚拟数据**，隐藏了细节。

## □ 例如：类似编译器，用户可以用类型定义语句定义一个存放数据空间，

- `int a;`

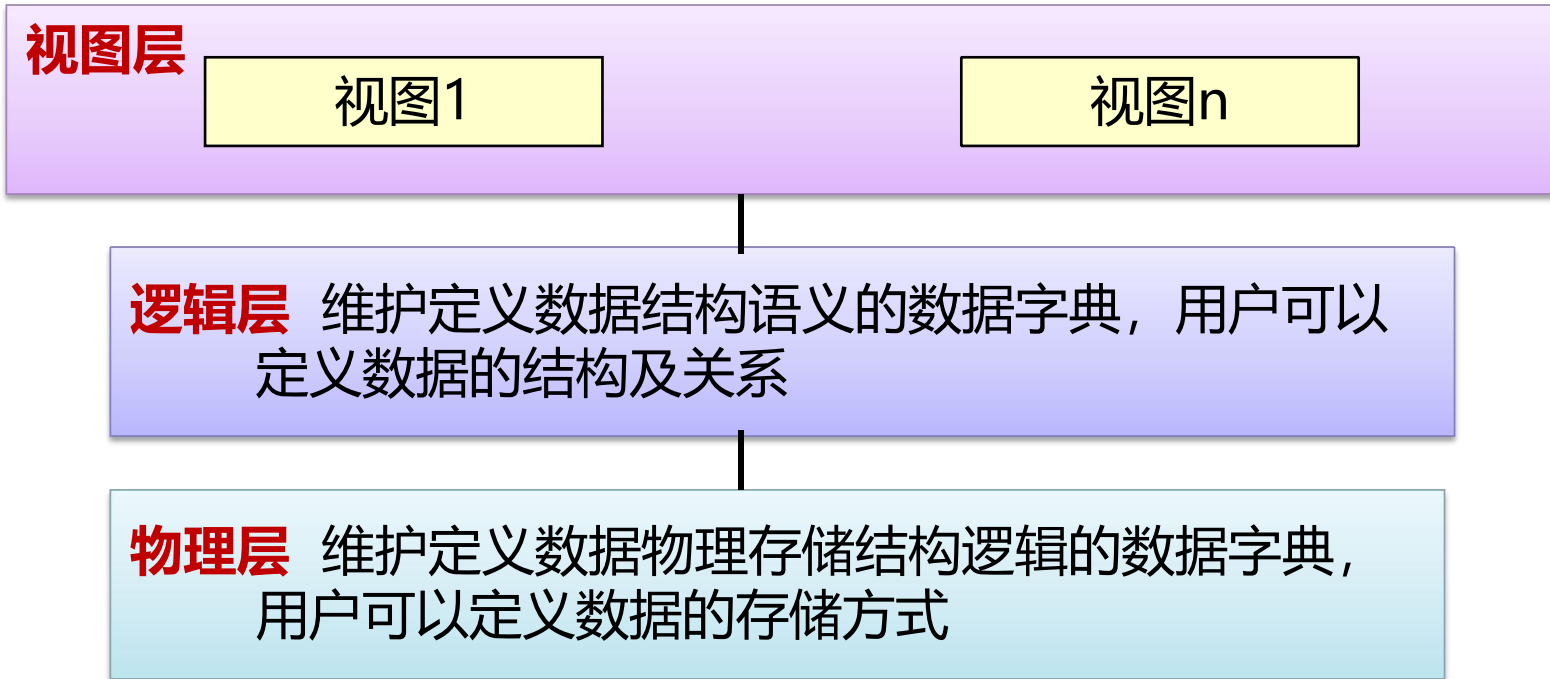
- 类似文件系统，用户可以用`fopen()`来创建一个文件；

- DBMS可以用CREATE语句创建数据库；

## □ 系统通常采用三层抽象来完成：

视图层、逻辑层、物理层

# 数据库系统的三层数据抽象



- CREATE VIEW AS 可以定义一个视图类型的数据
- CREATE TABLE 可以定义一个表类型的数据
- 模式和实例 定义的数据可类比变量定义，称为**模式**。某时刻的数据称为**实例**，类似值

# 3.数据库系统阶段

## □ 特点:

数据库观点: 数据不是依赖于处理过程的附属品, 而是现实世界中独立存在的对象。

- 数据结构化
- 数据的共享性高, 冗余度低, 易扩充
- 数据独立性高
- 产生了数据库管理系统 (DBMS) ,数据由DBMS统一管理



# 文件系统与数据库系统的区别

- 从文件系统管理发展到数据库系统管理是信息处理领域的一个重大变化，表现在：
  - 在文件系统阶段，人们关注的是系统功能的设计，因此程序设计处于主导地位，数据服从于程序设计；
  - 在数据库系统阶段，数据的结构设计成为信息系统首先关心的问题。

数据管理技术经历了以上三个阶段的发展，已有了比较成熟的数据库技术，但随着计算机软硬件的发展，数据库技术仍需不断向前发展。

## 1.1.3 数据库系统的特点

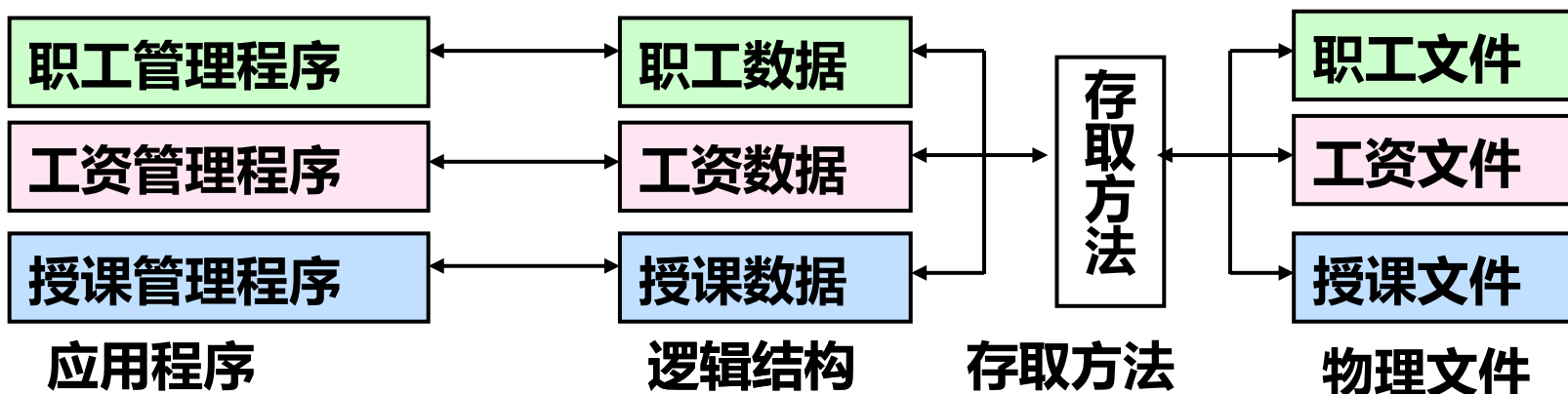
1. 数据结构化
2. 数据的共享性高，冗余度低，易扩充
3. 数据独立性高
4. 数据由DBMS统一管理和控制

# 1. 数据的结构化

例：设有如下数据：

<u>职工基本记录：</u>
(职工号, 姓名, 单位, 性别, 年龄, 工龄, 职称, 工资)
<u>工资记录：</u>
(职工号, 姓名, 职称, 工龄, 工资, 房租, 水电)
<u>授课记录：</u>
(职工号, 单位, 姓名, 职称, 课程名, 学时)

文件系统中应用程序与数据的使用对应方式：



# 1. 数据的结构化

例：设有如下数据：(职工号，姓名，单位，性别，年龄，工龄，职称，工资，房租，水电，课程名，学时)

职工基本记录：

(职工号，姓名，单位，性别，年龄，工龄，职称，~~工资~~)

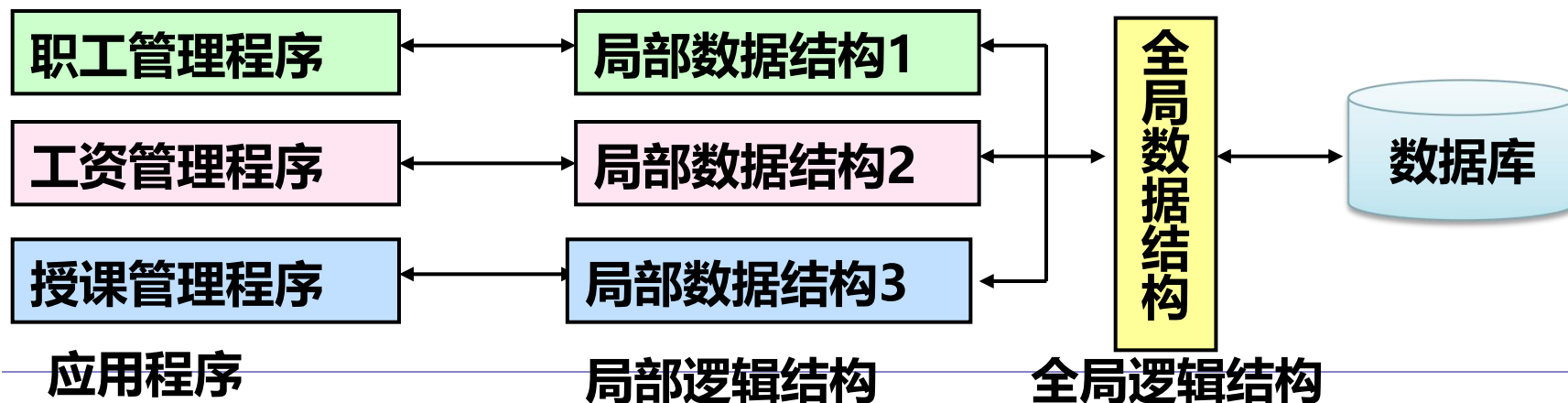
工资记录：

(职工号，~~姓名~~，~~职称~~，工龄，工资，房租，水电)

授课记录：

(职工号，~~单位~~，~~姓名~~，~~职称~~，课程名，学时)

数据库系统中应用程序与数据的使用对应方式：



## 数据结构化：

按照某种数据模型，将全系统的各种数据组织到一个结构化的数据库中，整个组织的数据不是一盘散沙，可表示出数据之间的有机关联。

# 1. 数据结构化

- 整体数据的结构化是数据库的主要特征之一
- “整体” 结构化
  - 不再仅仅针对某一个应用，而是面向全组织
  - 不仅数据内部结构化，整体是结构化的，数据之间具有联系
- 数据库中实现的是数据的真正结构化
  - 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

## 2. 数据的共享性高、冗余度低，易扩充

- **数据面向整个系统**，而不是面向某一应用，数据集中管理，数据共享，因此冗余度低；
- 节省存储空间，减少存取时间，且可避免数据之间的不一致性；
- 每个应用选用数据库的一个子集，只要重新选取不同子集或者加上一小部分数据，就可以满足新的应用要求，这就是易扩充性。

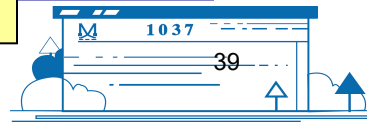
**注意：**

**数据库中并非完全消除冗余！** 有时为了提高数据的存取效率，同一数据可保留多个副本，但是这种冗余是在数据库系统的控制之下的。

### 3. 数据独立性

- 数据独立性指数据与使用数据的程序分离的特性。分为**逻辑独立性**和**物理独立性**。
- **逻辑独立性**是指当数据的总体逻辑结构改变时，数据的局部逻辑结构不变，由于应用程序是依据数据的局部逻辑结构编写的，所以应用程序不必须修改，从而保证了数据与程序间的逻辑独立性。
  - 例如，在原有的记录类型之间增加新的联系，或在某些记录类型中增加新的数据项，而不需要修改应用程序。
- **数据的物理独立性**是指当数据的存储结构改变时，数据的逻辑结构不变，从而应用程序也不必改变。
  - 例如，改变存储设备和增加新的存储设备，或改变数据的存储组织方式，而不需要修改应用程序。

**良好的数据独立性可降低程序开发和维护的代价。**



## 4. 统一的数据管理和控制功能

### □ 数据的**安全性控制** (Security)

- 保护数据以防止不合法的使用所造成的数据泄露和破坏
- 措施：用户标识与鉴定，存取控制、数据加密、审计等

### □ 数据的**完整性控制** (Integrity)

- 数据的正确性、有效性、相容性  
例如：身高 = 17.5米
- 措施：完整性约束条件定义和检查  
例如：身高约束 (身高 < 3米)

### □ **并发控制** (Concurrency)

- 对多用户的并发操作加以控制、协调，防止其互相干扰而得到错误的结果并使数据库完整性遭到破坏
- 措施：封锁

### □ **数据库恢复** (Recovery)

- 把数据库从错误状态恢复到某一已知的正确状态(亦称一致性状态或完整性状态)
- 措施：转储，镜像，日志



# 1.1 回顾

1. 认识了一种新的计算机系统——**数据库系统**。
2. **数据库系统 (DBS)** 与其它系统的主要区别在于其中包含了**数据库 (DB)** 和 **数据库管理系统 (DBMS)** 。
3. **数据库**是特定应用环境中的相关数据按照一定的方式**组织**形成的数据整体，具有**冗余低、数据独立性高、易扩展、可共享**等传统数据文件所不具备的优点。
4. **数据库管理系统**是负责管理数据库的系统软件。它使得应用程序能简单、快速、共享地使用数据库中的数据，并能在此过程中保证数据的**安全性、完整性和一致性**。
5. 数据管理技术发展的3个阶段：**人工、文件系统、数据库系统阶段**。
6. 数据库系统的4大特点。

# 课堂练习

1. 长期存储在计算机内、有组织、可共享的大量数据的集合是（ ）。  
A. 数据    B. DB    C. DBMS    D. DBS
2. 数据库管理系统是一种系统软件，它建立在( )系统之上。  
A. 应用系统    B. 编译系统    C. 操作系统    D. 硬件系统
3. 判断题：在文件系统管理阶段，由文件系统提供数据存取方式，所以数据已经达到了很强的独立性，不需要应用程序来管理数据。 （ ）
4. 判断题：引入数据库的计算机系统称为数据库系统。 （ ）