



# 数据库原理及应用

---

主讲：陈安龙

*[chenanlong@uestc.edu.cn](mailto:chenanlong@uestc.edu.cn)*

电子科技大学

# 一、教师个人简介

- 研究方向：数据库与机器学习
- 电话：13688434459
- Email: [chenanlong@foxmail.com](mailto:chenanlong@foxmail.com)
- 课程QQ群：1048115160      微信群：



群名称：数据库原理及应用-2020-陈安龙  
群 号：1048115160



2020-数据库原理及应用-陈安龙

该二维码7天内(2月21日前)有效，重新进入将更新

## 二、常见的企业应用系统

银行系统

股票交易系统

财务系统

医院HIS系统

人事档案管理系统

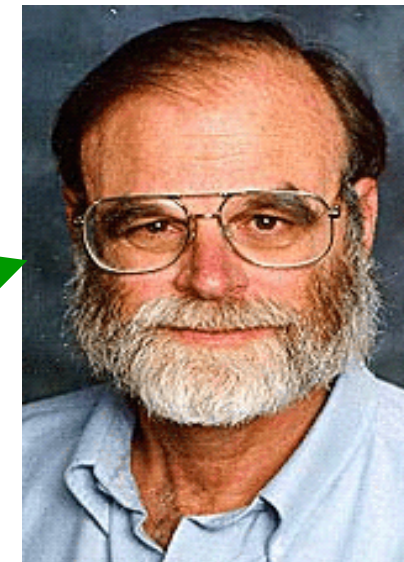
图书管理系统

电子商务系统

如何设计这些应用系统呢？系统如何存储信息？

### 三、数据库领域的图灵奖获得者

- 1973 **Charles W. Bachman** 因在数据库方面的杰出贡献而获奖
- 1981 **Edgar F. Codd** 因在数据库管理系统的理论和实践的贡献而获奖
- 1998 **James Gray** 因在数据库和事务处理方面的突出贡献而获奖
- 2014 **Michael Stonebraker** 开发了 Ingres、Postgres 和 Vertica、IBM System、Sybase 和 SQL Server 主流数据库。



## 四、课程学习目标

- **CO1:** 掌握数据库系统基本概念、数据模型原理、数据库编程语言、数据库系统管理、NoSQL数据库技术原理等基础知识。
- **CO2:** 掌握数据库建模设计技术，能够给出数据库的优化设计方案，培养数据库设计能力。
- **CO3:** 掌握数据库SQL语言和数据库管理技术，能够对实现结果给出合理性解释，培养数据库操作访问与数据库系统管理能力。
- **CO4:** 掌握数据库应用编程技术及其开发工具的使用，培养数据库应用编程能力，同时培养解决复杂工程问题的数据库应用系统开发能力。

## 五、本课程教学内容

- CM1: 数据库系统概论 3学时
- CM2: 数据库关系模型 4学时
- CM3: 结构化查询SQL语言 8学时
- CM4: 数据库设计与实现 8学时
- CM5: 数据库管理 6学时
- CM6: 数据库应用编程 7学时
- CM7: NoSQL数据库技术 4学时

## 六、本课程学分与学时

课程3个学分，48学时。课堂讲授40学时，实验8学时

## 七、本课程教学方式

- 采用“**翻转课堂教学**”方式进行课程教学，并以“**项目案例学习**”、“**挑战性问题研讨**”、“**边做边学**”等方式掌握数据库原理及应用技术。
- 学生还需基于**Mooc自主学习**，参与课堂讨论，并需要完成一定的课外上机实践作业。

# 八、本课程考核与评价方式

## 1、考核环节

课程考核分课程作业、单元测试、上机实验、期末考试四个环节，分别占课程总成绩的**20%、10%、20%、50%**。

## 2、课程评价

分为课程学习形成性评价、总结性评价

**课程学习形成性评价：**

课程作业成绩=（MOOC学习成绩+大作业成绩）/ 2 - 缺勤次数（每缺勤1次扣1分）

单元测试成绩=多次单元测试成绩的平均值

实验成绩=上机测评成绩\*0.8+实验报告成绩\*0.2

**课程学习总结性评价：** 期末考试成绩



## 九、本课程教材与参考书

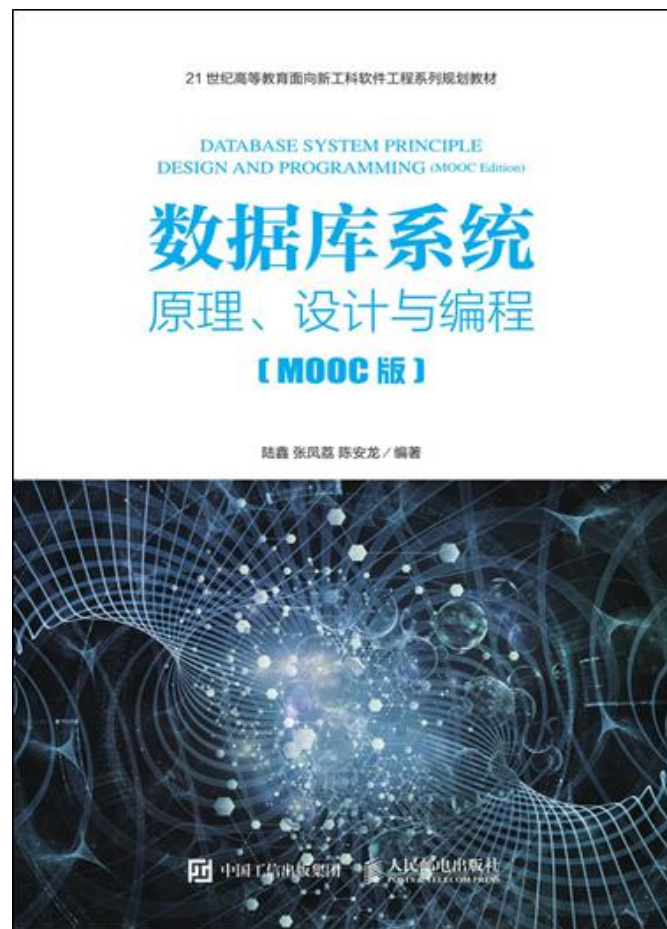
### 1、课程教材

“数据库系统-原理、设计及编程”  
(Mooc版) ，陆鑫，张凤荔，陈安龙，  
人民邮电出版社, 2019.1

### 2、参考书

[1] PostgreSQL实战，谭峰，张文升 著，  
机械出版社，2018.7

[2] NoSQL数据库入门与实践，刘瑜，刘  
胜松著，中国水利水电出版社，2018.2



# 本课程的在线电子书



4G 144 B/s 16:57 53%

× 可知

...

🏠



数据库系统——原理、设计与编程 (MOOC版)

作者：陆鑫 张凤荔 陈安龙

出版社/提供方：人民邮电出版社

出版时间：2019年03月 出版

☆☆☆☆☆

阅读量：582

内容简介

本书介绍了数据库的原理、设计与编程技术。全书共七章，内容包括数据库系统基础、数据库关系模型、数据库操作语言SQL、数据库设计与实现、数据库管理、数据库应用编程等。

图书信息

关键字

数据库系统—高等学校—教材

资源格式

pdf

ISBN

978-7-115-50274-2

加入书架

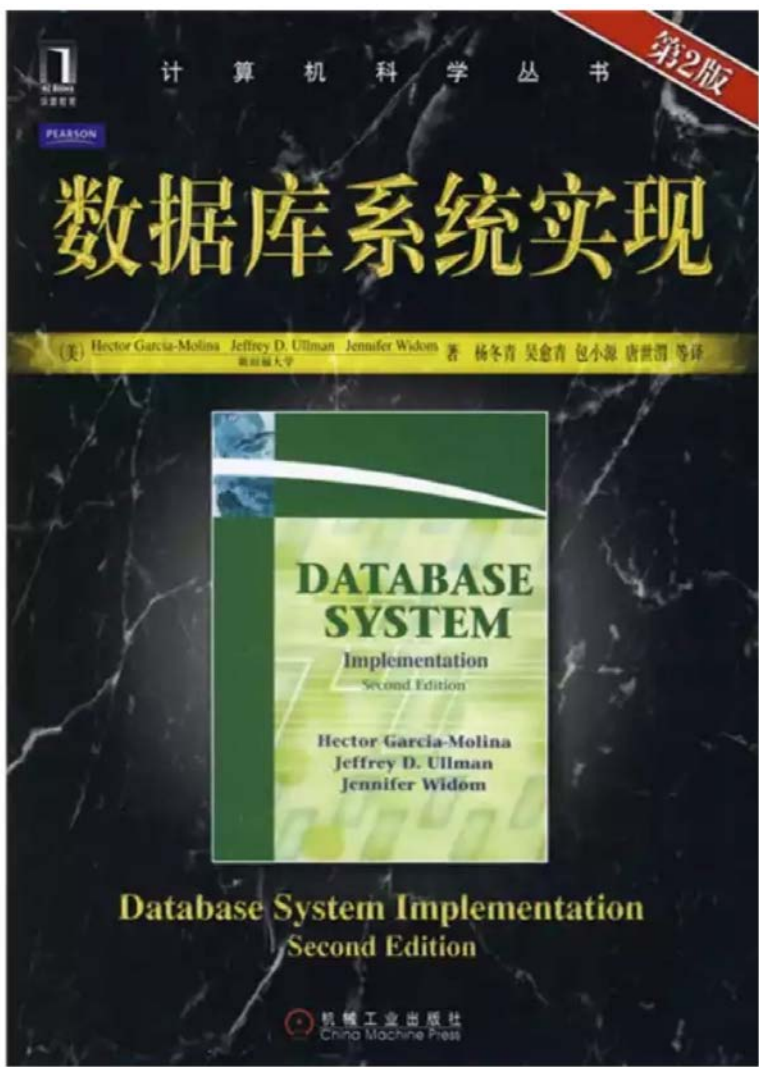
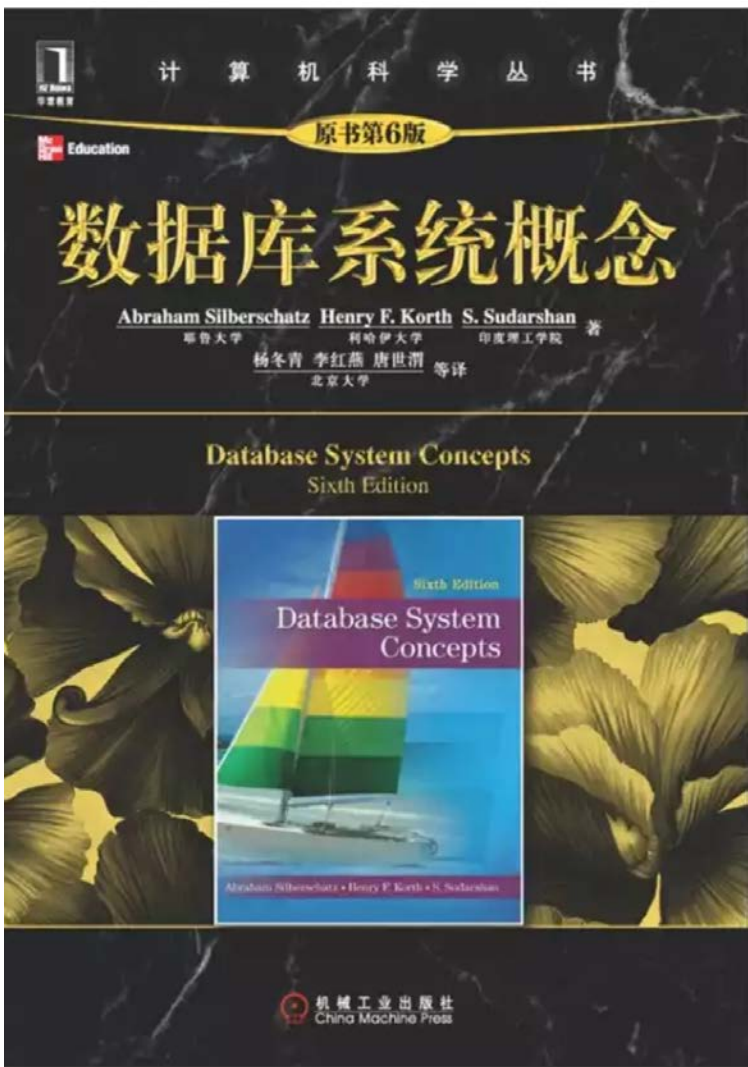
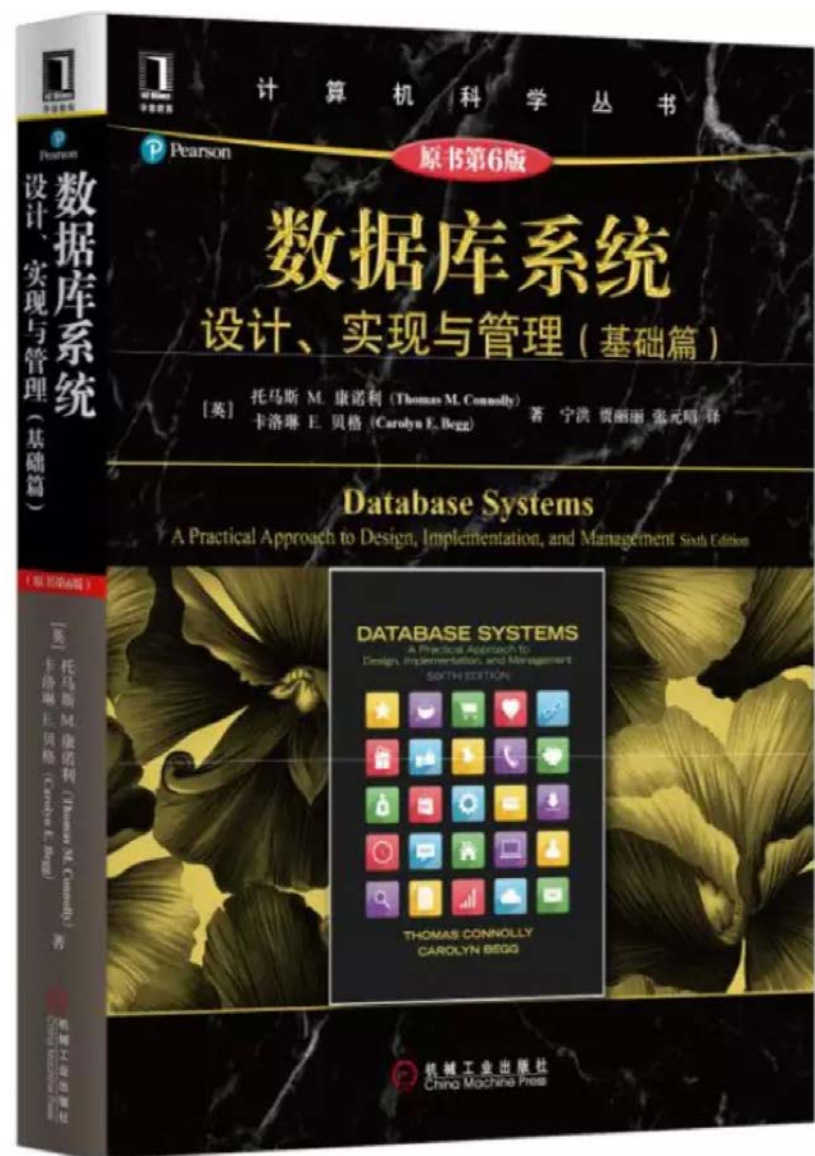
在线阅读

□

○

◀

# 推荐几本从国外引进的优秀教材





# 十、本课程MOOC学习平台


## 学堂在线课程学习平台

<https://next.xuetangx.com/course/EST08091001150/1510780>



# 十一、本课程校内SPOC学习平台



 微信扫一扫，使用小程序



## 十二、本课程数据库学习软件

课程实践使用的数据库软件

(1) **Postgre SQL 12**

(2) **pgAdmin4**

(3) **Sybase PowerDesigner16.6**

# 大家思考下列问题

- 为什么要使用数据库？
- 数据库有什么特点？
- 引入数据库技术，主要是为了解决什么问题？

# 第1章 绪论

- 1、了解数据管理的发展过程
- 2、掌握数据库系统的基本概念和主要特点
- 3、掌握数据库系统的三级模式结构和数据库系统的组成
- 4、掌握数据库模型有关概念和三种数据模型
- 5、了解数据库技术发展方向



# § 1 数据库系统概述

## 一、基本概念

**1、数据：**描述事务的符号记录。可用文字、图形等多种形式表示，经数字化处理后可存入计算机。

**2、数据库（DB）：**按一定的数据模型组织、描述和存储在计算机内的、有组织的、可共享的数据集合。

**3、数据库管理系统（DBMS）：**位于用户和操作系统之间的一层数据管理软件。主要功能包括：

**数据定义功能：**DBMS提供**DDL**，用户通过它定义数据对象。

**数据操纵功能：**DBMS提供**DML**，用户通过它实现对数据库的查询、插入、删除和修改等操作。

**数据库的运行管理：**DBMS对数据库的建立、运行和维护进行统一管理、统一控制，以保证数据的安全性、完整性、并发控制及故障恢复。

**数据库的建立和维护功能：**数据库初始数据的输入、转换，数据库的转储、恢复、重新组织及性能监视与分析等。

**4、数据库系统（DBS）：**计算机中引入数据库后的系统，包括

数据库DB

数据库管理系统DBMS

应用系统

数据库管理员DBA和用户

## 二、数据管理与数据处理

### 1、数据管理：

对数据**收集、整理、组织、存储、维护、检索、传送**等

**对象**

**操作**

目标：在恰当的时候以恰当的形式给特定的人提供适当的数据。

2、**数据处理**：是对数据进行加工的过程。对数据进行的**查询、分类、修改、变换、运算、统计、汇总**等都属于加工。其目的是根据需要,从大量的数据中**抽取**出**有意义、有价值数据(信息)**,作为决策和行动的**依据**,其实质是**信息处理**。

**数据**



**信息**

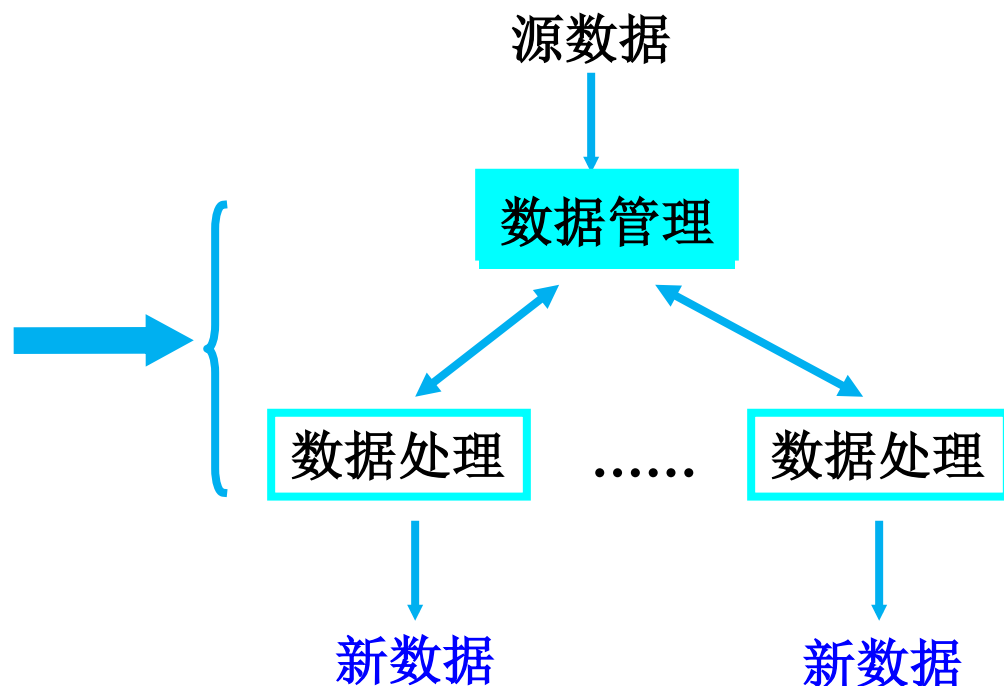
### 3、管理与处理的关系：

{ 管理是处理的基础  
处理为管理服务

注意：

数据处理与数据管理的区别：**狭义上**一般使数据发生**较大根本性变化的数据加工**称为**数据处理**（**其他称为数据管理**），如汇总，而**广义上**时常不加区别地统称为**数据处理**。

管理和处理又可看成一个问题  
的两个阶段，故可以统一  
起来，其中心是管理



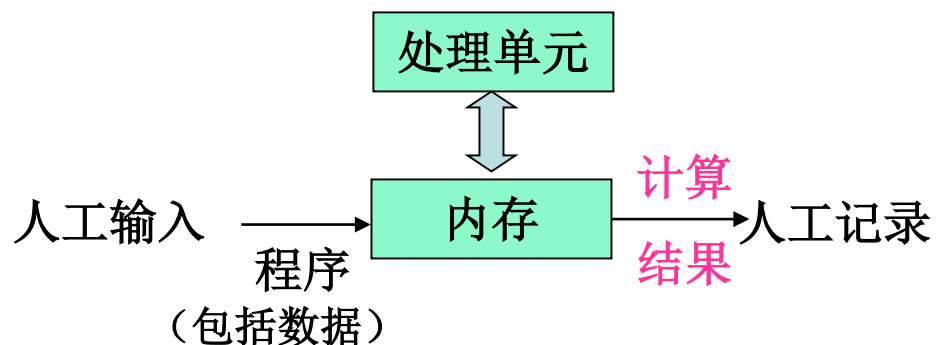
### 三、数据管理的发展阶段

- ✓ 人工管理阶段（50年代中期以前）
- ✓ 文件系统阶段（50年代中期至60年代后期）
- ✓ 数据库系统阶段（60年代后期以后）

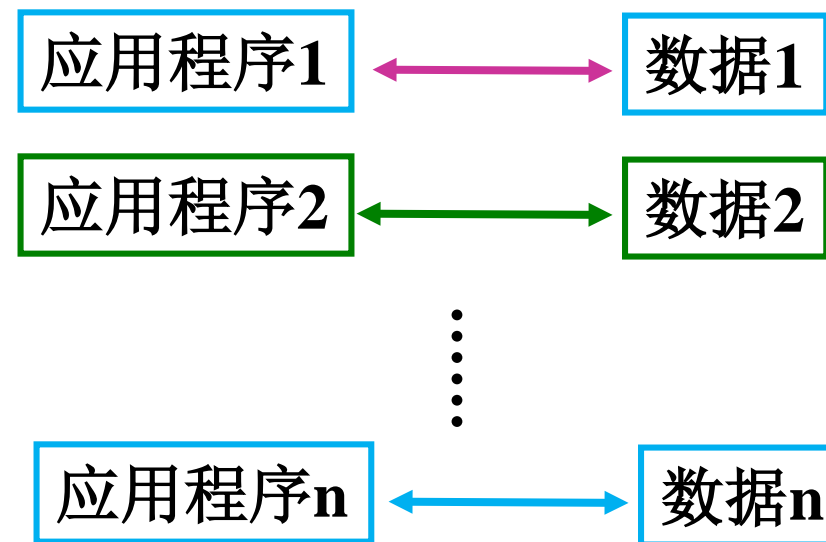
# 1、人工管理阶段（程序员管理阶段）

特点：

- ① 数据不保存
- ② 程序员负责数据管理的一切工作
- ③ 数据和程序一一对应，没有独立性和共享性



数据和程序的关系：



## 2、文件系统阶段

基础 { 硬件：有了大容量直接存储外存设备，如磁盘、磁鼓等  
软件：有了专门的数据管理软件—文件系统  
处理方式：有批处理、联机实时处理等

又可分为两个阶段

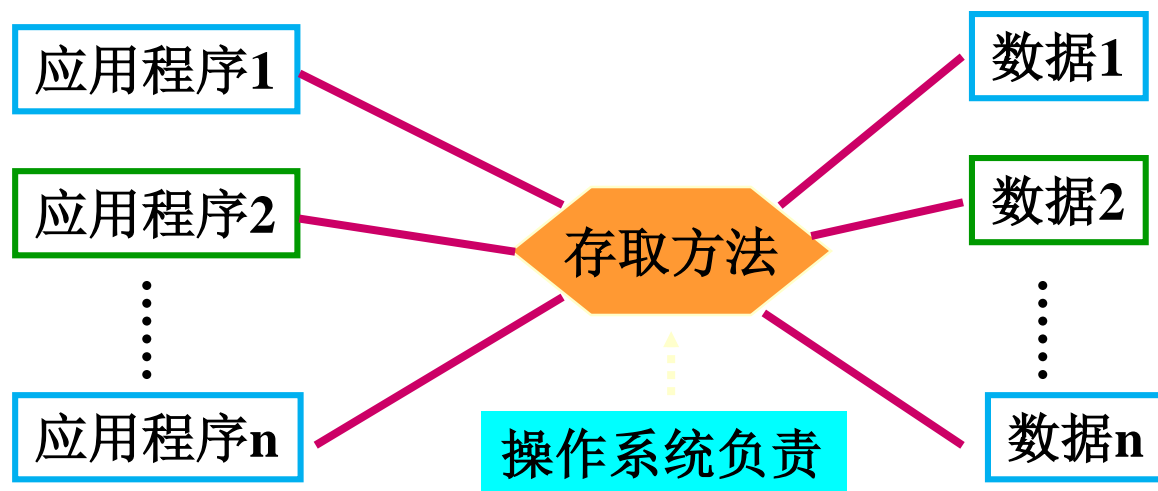
### (1) 60年代初期出现了初等的文件系统

主要特点：✓ 组织方式：顺序文件  
✓ 数据结构：物理结构 = 逻辑结构  
✓ 软件功能：仅有简单I/O操作

### (2) 60年代中期出现了成熟的文件系统

主要特点：✓ 组织方式：顺序和随机存取并用  
✓ 数据结构：物理结构和逻辑结构有了简单的变换  
✓ 软件功能：软件系统提供了存取方法

## 数据与程序的关系：

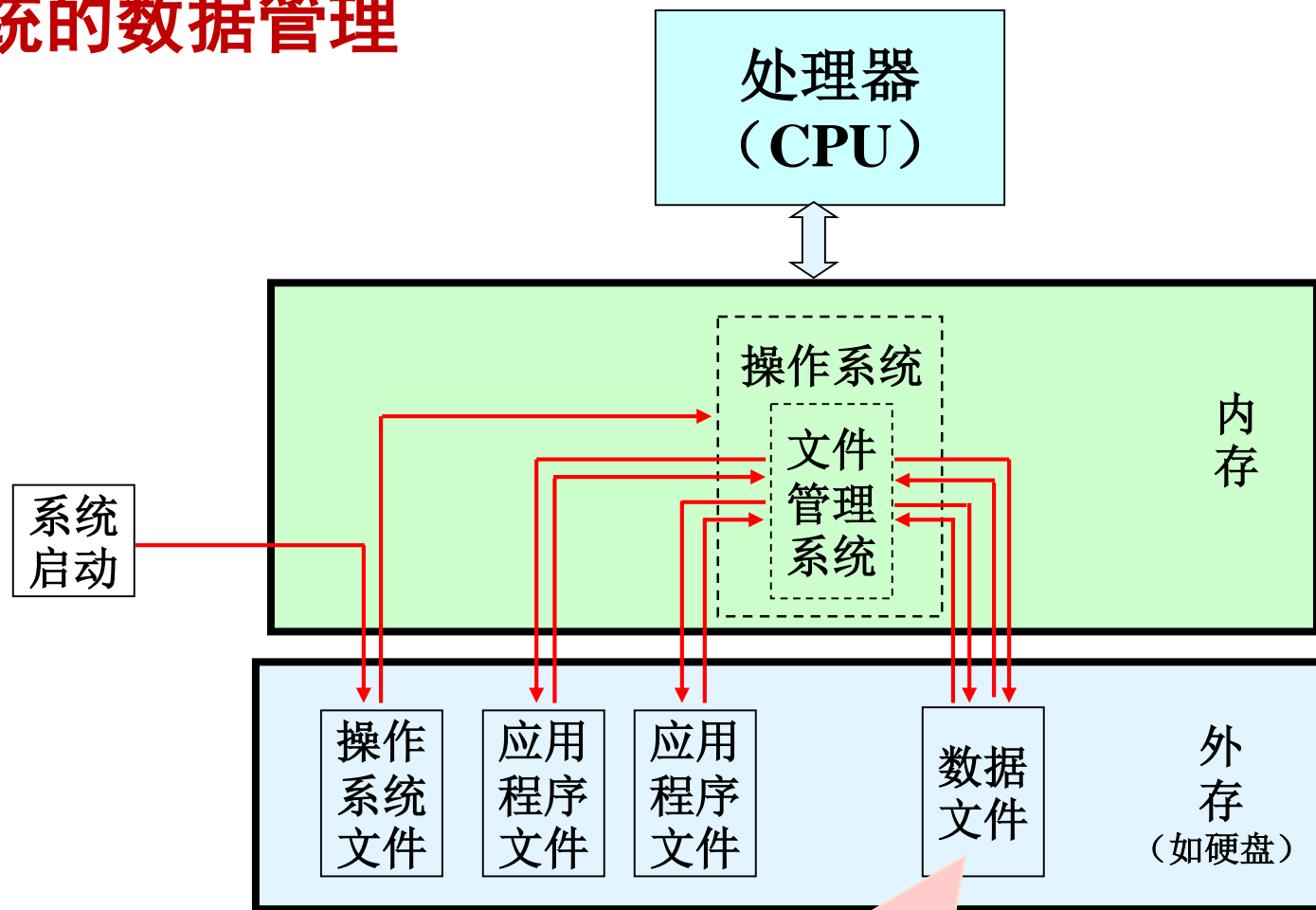


### 三个主要缺点：

- ① **数据高度冗余：** 数据基本上还是面向应用或特定用户的。
- ② **数据共享困难：** 文件基本上是私有的，只能提供很弱的文件级共享
- ③ **数据和程序缺乏独立性：** 只有一定的物理独立性，完全没有逻辑独立性。



# 基于文件系统的数据管理



多个程序共享数据文件，  
减少了数据的冗余！

### 3、数据库系统阶段

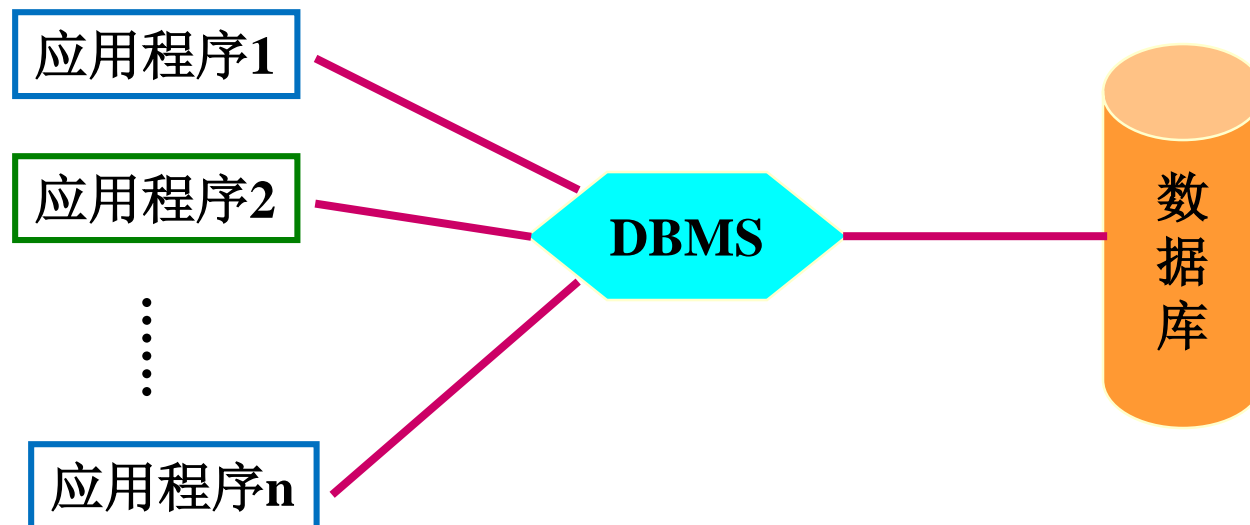
文件系统不能适应大数据量、多应用共享数据的根本原因：

数据没有集中管理

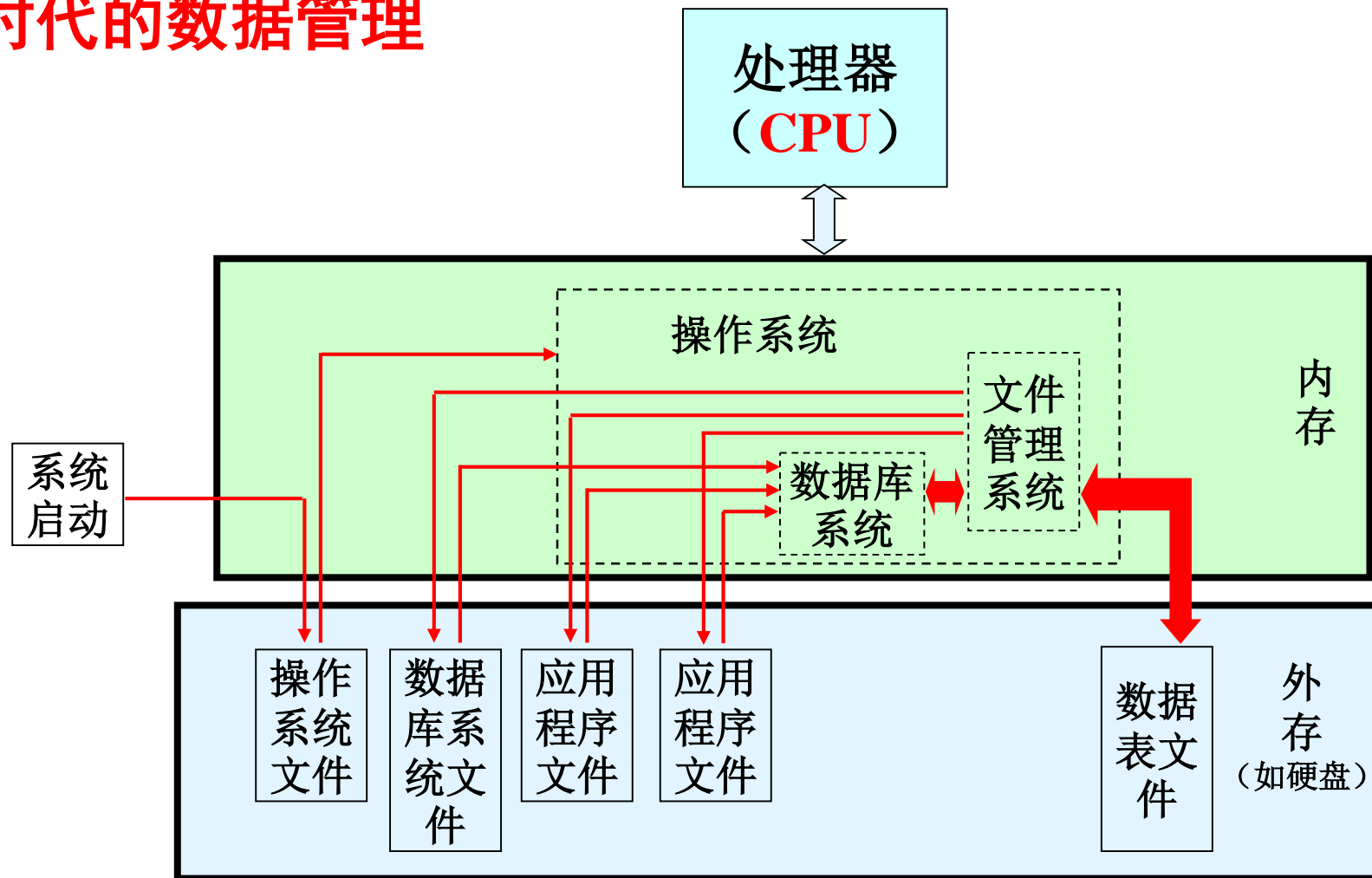
数据库方法的基本出发点：

把数据统一管理、控制，共享使用

数据与程序的关系：



# 数据库时代的数据管理



# 主要优点

- (1) 数据高度结构化集成，面向全组织
- (2) 数据共享性好。可为多个不同的用户共同使用
- (3) 数据冗余少，易扩充
- (4) 数据和程序的独立性高

**物理独立性：**存储结构变，逻辑结构可以不变，从而应用程序也不必改变。

**逻辑独立性：**宏观逻辑结构变，局部逻辑结构可以不变，从而应用程序不必改变。

**优点：**简化应用程序的编写和维护

- (5) 数据控制统一

**安全性控制：**防止泄密和破坏

**完整性控制：**正确、有效、相容

**并发控制：**多用户并发操作的协调控制

**故障恢复：**发生故障时，将数据库恢复到正确状态

## 4、从四个方面各个阶段的比较：

	人工管理	文件系统	数据库系统
谁管理数据	程序员	操作系统提供存取方法	系统集中管理
面向谁	特定应用	基本上是特定用户	面向系统
共享性	不能	共享很弱	充分共享
数据独立性	没有	一定的物理独立性	较高的独立性

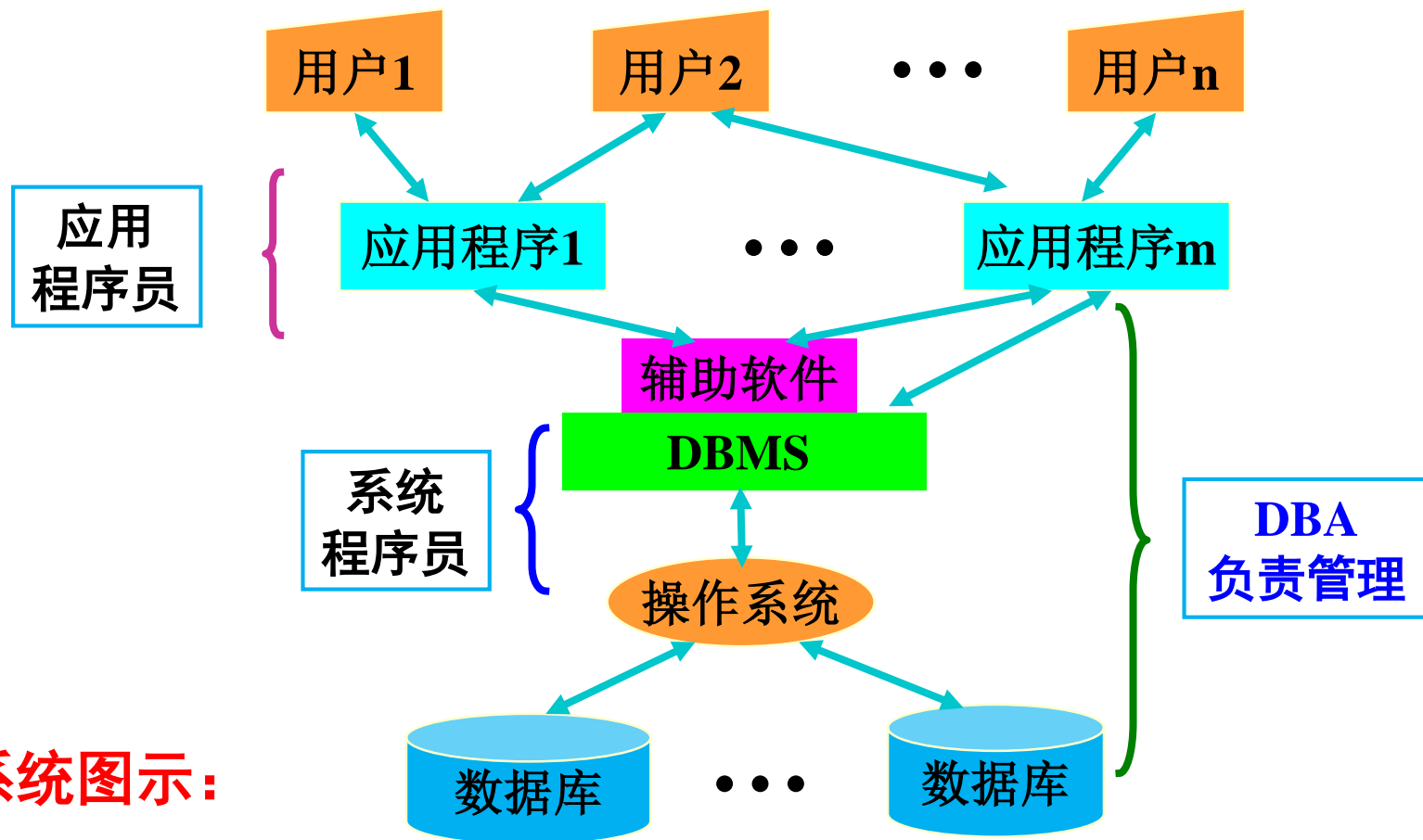
### 文件系统和数据库系统的本质区别：

**内部：**数据库的数据是结构化的，有联系的文件系统的各记录无联系

**外部：**数据库系统是共享的，文件系统基本上是面向特定用户的

## § 2 数据库系统的构成

**数据库系统**是一个**采用**数据库技术的计算机系统。是按照数据库方式存储、管理、维护并可**提供**数据支持的系统，**一个典型的数据库系统包括**数据库、DBMS、应用程序、用户和数据库管理员(DBA)**五个部分**。



数据库系统图示：

**用户 (User)** 是指使用数据库的人员。用户可分为**终端用户**、和**应用程序员**。

**终端用户 (End User)** 是指在终端按权限使用数据库的各类人员。**应用程序员 (Application Programmer)**负责为终端用户设计和编制数据库应用程序，以便终端用户对数据库进行操作。

**数据库管理员 (DataBase Administrator, DBA)**是数据库所属机构的**专职管理员**。

**DBA主要职责为：**

- (1) **参与数据库分析设计或引进**的整个过程，**决定**数据库的结构和数据内容。
- (2) **定义 (建立-设置)** 数据的安全性和完整性，负责**分配**用户对数据库的使用权限和口令管理。
- (3) **监督控制**数据库的使用和运行，**改进和重新构造**数据库系统.当数据库受到意外破坏时，负责**进行恢复**；当数据库的结构需要改变时，负责对其**修改**。

# 数据库、数据库系统与数据管理系统（DBMS）区别

**数据库**(DataBase,DB)是存储在计算机上的**结构化的相关数据集合**。可理解为“按一定结构**存管数据的仓库**”,是在计算机内的、有组织（结构）的、可共享、长期存储的**数据集合**。数据库中的数据可按一定的**数据模型（结构）**进行**组织、描述和存储**,具有较高的数据独立性和易扩展性,较小的冗余度,并可共享。数据库还具有集成性、共享性、海量性和持久性等**特点**。数据库技术**主要用于**根据需求自动处理、管理和控制大量业务数据。

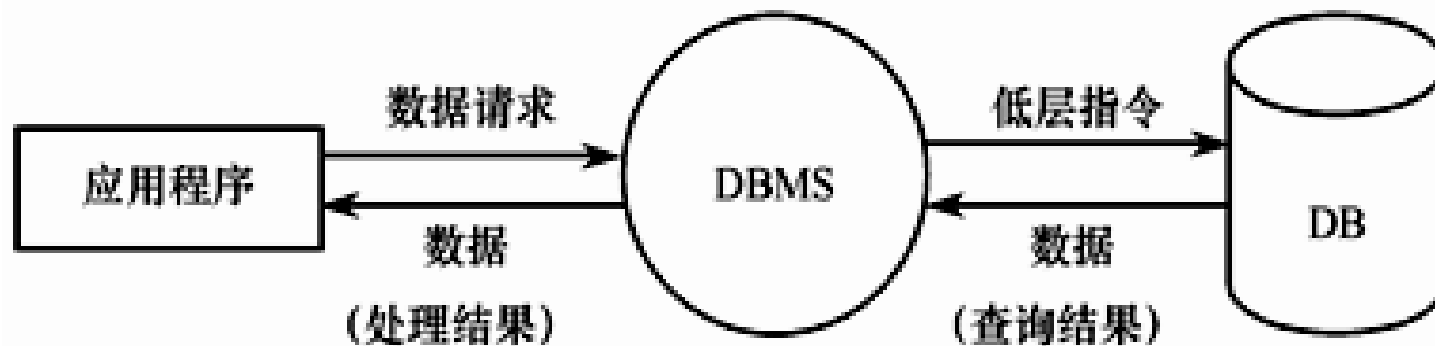
**数据库系统** (DataBase System, DBS) 是**具有数据库功能特点的计算机系统**。是**实现**有组织地、动态地**存储**大量关联数据、方便多用户**访问**的计算机软硬件和数据资源**组成**的系统。

**特性**: 实现数据共享,减少数据冗余度; 保持数据一致性和独立性;提高系统安全性, 并发控制及故障恢复。



## 数据库管理系统DBMS

数据库管理系统（DBMS）是对数据库及其数据进行统一管理控制的软件系统。是数据库系统的核心和关键的部分，用于统一管理控制数据库系统中的各种操作，包括数据定义、查询、更新及各种管理与控制，都是通过DBMS进行的.DBMS的查询操作工作示意图如图所示。



DBMS的查询工作示意图

# DBMS的主要功能和机制

## 1. DBMS的主要功能

- (1) 数据定义功能-数据定义语言DDL,定义库、表、视图等。
- (2) 数据操作功能-数据操作语言DML、数据查询语言DQL。
- (3) 事务与运行管理(核心)-数据控制语言DCL,事务管理语言TML,系统运行控制程序。
- (4) 组织、管理和存储数据-分类、管理、数据字典、存取路径/方式
- (5) 数据库的建立和维护功能-数据载入、存储、重组、恢复、维护
- (6) 其他功能-与其他系统通信、互访、互操作、转换。

## 2. DBMS的工作机制

**DBMS的工作机制**是将用户对数据的操作**转化为**对系统存储文件的操作，有效地**实现**数据库三级模式结构之间的**转化**。

**数据库管理系统的主要职能**有数据库的定义和建立、数据库的操作、数据库的控制、数据库的维护、故障恢复和数据通信。

# DBMS的模块组成

1) 按程序实现的功能可分为以下4部分

- (1) 语言编译处理程序。
- (2) 系统运行控制程序。
- (3) 系统建立与维护程序。
- (4) 数据字典。



2) 按照模块结构分, 可将DBMS分成查询处理器和存储管理器两大部分

①查询处理器有4个主要成分:

DDL编译器、DML编译器、嵌入式DML的预编译器及查询运行核心程序。

②存储管理器有4个主要成分:

权限和完整性管理器、事务管理器、文件管理器及缓冲区管理器。

## 常见的数据库管理系统

**小型：** Dbase、Foxbase、Foxpro、Visual Foxpro、Access

**中型：** SQL Server、MySQL、PostgreSQL

**大型：** Oracle、Sybase、DB2

**NoSQL数据库：** MongoDB、CouchDB、Cassandra、Redis、Memcached

**嵌入式数据库：** SQLite

# 国产数据库产品情况

**POLARDB 阿里云**

**TBASE/TSQL 腾讯云**

**GaussDB(高斯数据库) 华为技术有限公司**

**CynosDB 腾讯云计算（北京）有限责任公司**

**DM(达梦数据库) 武汉达梦数据库有限公司**

**Gbase 天津南大通用数据技术股份有限公司**

**Kingbase 北京人大金仓信息技术股份有限公司**

**OSCAR(神通数据库) 天津神舟通用数据技术有限公司**

**Obase 上海丛云信息科技有限公司**

**EsgynDB(易鲸捷数据库) 贵州易鲸捷信息技术有限公司**

**K-DB 浪潮集团有限公司**

**OpenBASE 东软集团有限公司**

**SequoiaDB(巨杉数据库) 广州巨杉软件开发有限公司**

**OceanBase 阿里/蚂蚁金服**

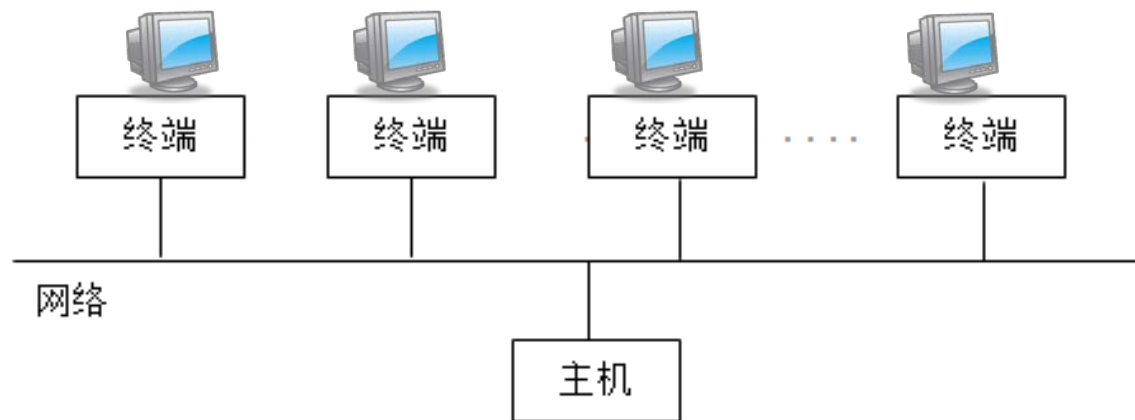
**TiDB 北京平凯星辰科技发展有限公司**

.....

# 数据库系统的结构类型

- 1、集中式系统

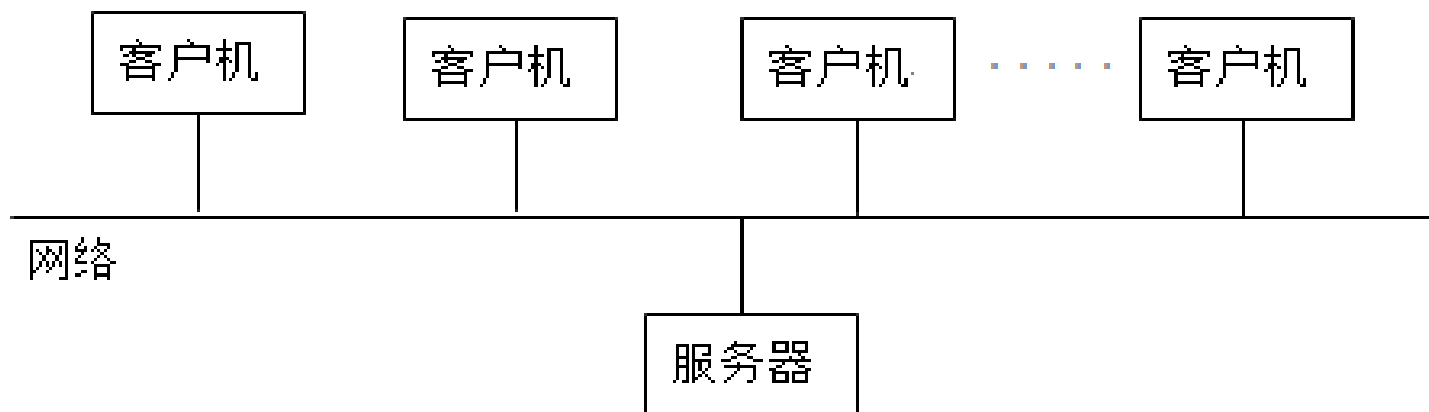
**集中式（Centralized）结构**是指一台主机带有多个用户终端的数据库系统。终端一般只是主机的扩展（如显示屏），并非独立的计算机。终端本身并不能完成任何操作，完全依赖主机完成所有的操作。如图所示。



集中式数据库系统结构

## 2. 客户机 / 服务器系统

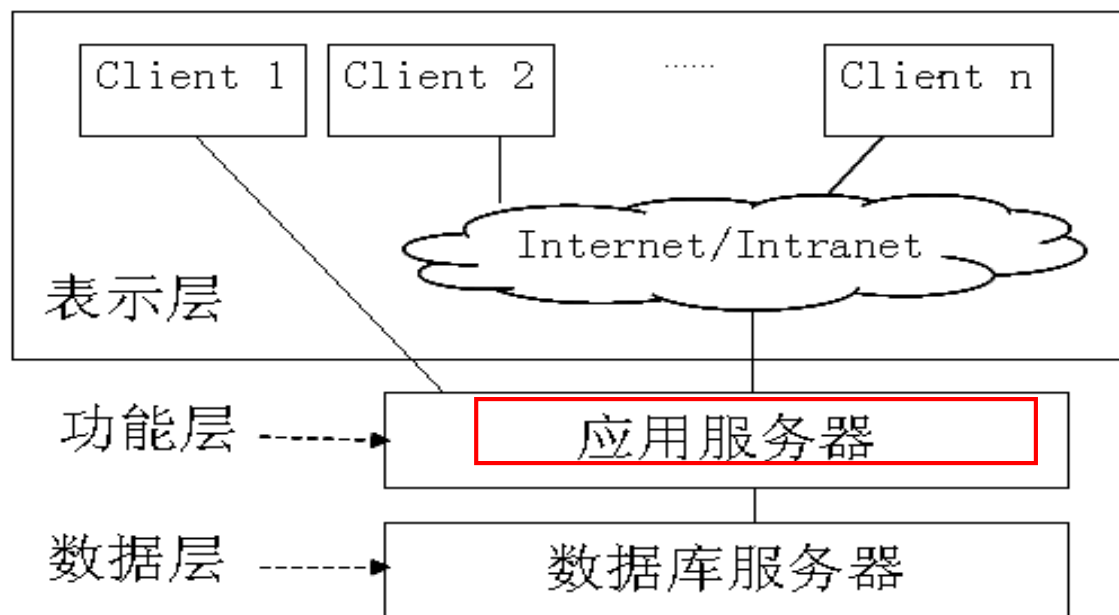
在**客户机/服务器**（Client/Server, C/S）结构中，将计算机应用任务**分解成**多个子任务，由多台计算机**分工完成**，即采用“**功能分布**”原则。**客户端**完成数据处理、数据表示和用户交互功能，**服务器端**完成DBMS的核心功能。**C/S系统模式**，如图所示。



C/S系统的一般结构

## 2. 客户机 / 服务器系统

- 三层结构的C/S体系结构比二层结构增加一个应用服务器层，如图。
- 三层C/S结构优点主要包括：整个系统被分成不同的逻辑块，层次清晰，一层的改动不会影响其他层次，可减轻的客户机负担，开发和管理工作向服务器端转移，使得分布的数据处理成为可能，管理和维护变得相对简单。

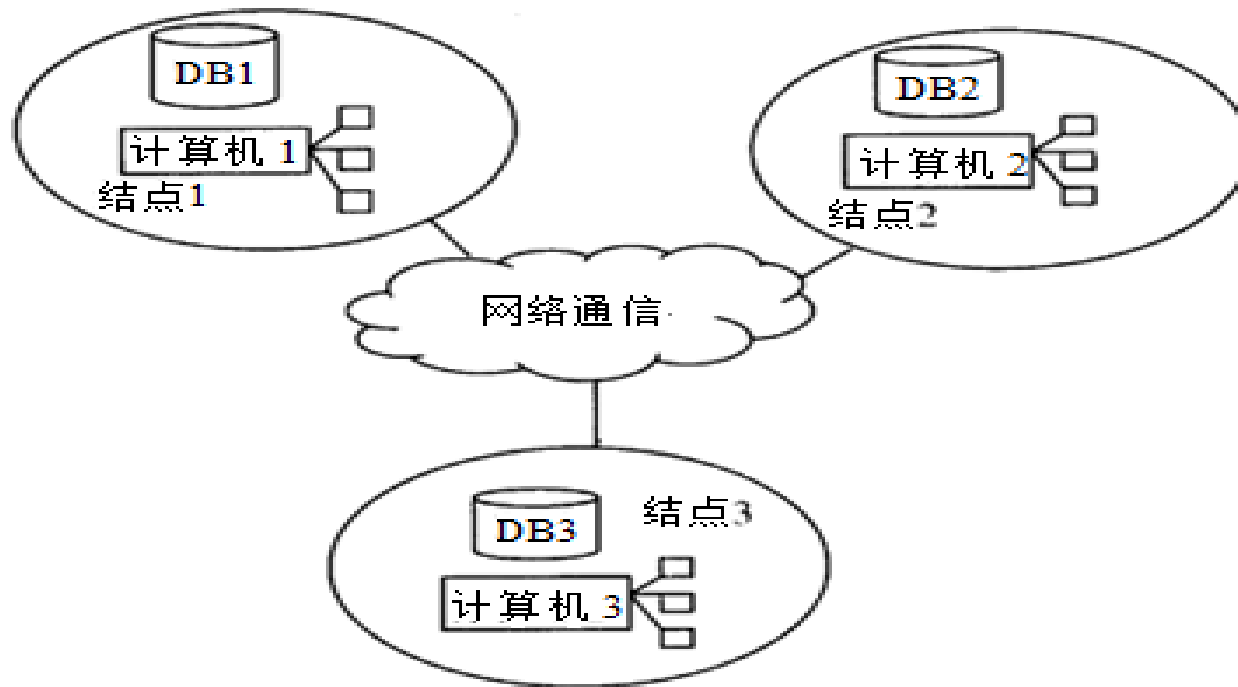


三层C/S体系结构图



### 3. 分布式系统

**分布式**（Distributed）数据库的数据具有“**逻辑整体性**”，**分布**在各地（结点）的数据逻辑上是一个整体，由计算机网络、数据库和多个结点**构成**，用户使用起来如同一个集中式数据库。如分布在不同地域的**大型银行和企业**等，采用的就是这种数据库。**分布式数据库结构**如图所示



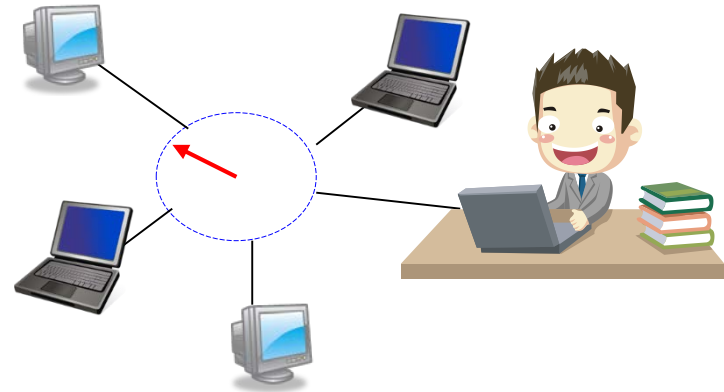
分布式数据库系统结构

## 4. 并行式系统

**并行式**（Parallel）计算机系统**使用多个CPU和多个磁盘进行并行操作**，**提高数据处理和I/O速度**。**并行处理时，许多操作同时进行，而不是采用分时的方法。**

并行DBS有**两个重要的性能指标**：

- (1) 吞吐量
- (2) 响应时间



 **讨论思考**

- (1) 数据库系统是哪几部分构成？
- (2) 数据库系统的外部系统结构有哪几种类型？
- (3) 什么是C/S系统的一般结构？画图表示。

# 数据库系统研究的对象

如何高效巧妙地进行数据管理，而又花费最少

数据库方向的主要研究领域：

- DBMS及其辅助软件
- 数据库设计
- 数据库相关理论
- 数据库的知识发现
- NoSQL等新型数据库

如：占用空间少  
查询快  
维护方便等

## § 3 数据库技术的特点及应用

- 1. 数据库技术的主要特点

- 1) 数据高度集成
- 2) 数据广泛共享
- 3) 数据独立冗余低
- 4) 实施统一的数据标准
- 5) 提高数据安全性和完整性
- 6) 保证数据一致性
- 7) 应用程序开发与维护效率高

- 2. 数据库技术应用

产品销售业务数据管理

随着IT技术的快速发展，数据库技术的应用从数据处理与管理，扩展到计算机网络应用、决策支持系统、商务智能和计算机辅助设计等新领域。在21世纪现代信息化社会，由于信息（数据）无处不在，数据库技术的应用非常广泛深入，遍布各个领域、行业、业务部门和各个层面。网络数据库系统及数据库应用软件已成为信息化建设和应用中的重要支撑性信息产业，得到广泛应用。

# 数据库技术应用行业。

电子商务、电子政务办公、金融业、制造业、电信业、航空业、教育系统等。

数据库技术是数据管理的最新技术，新的应用领域包括：

- (1) 电子商务平台
- (2) 多媒体数据库。
- (3) 空间数据库（地理信息系统）。
- (4) 移动数据库。
- (5) 信息检索系统。
- (6) 智能决策系统。

## ■ 讨论思考

- (1) 什么是数据管理？与数据处理有何区别。
- (2) 数据库系统与数据库管理系统的区别有哪些？
- (3) 数据库技术的主要特点有哪些？

## § 4 数据库技术的发展

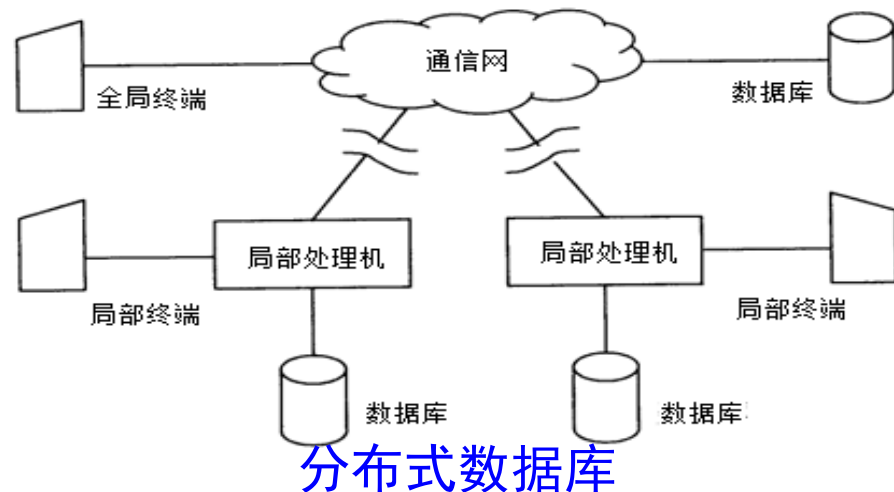
从20世纪80年代以后，数据库技术在商业领域取得巨大成功，激发了其他领域对其需求的快速增长，开辟了新的应用领域。

### 1. 分布式数据库技术

具有如下5个主要特点：

- (1) 大部分数据在本地进行分布处理,提高了系统处理效率和可靠性.数据复制技术是分布式数据库的重要技术。
- (2) 解决了中心数据库的不足，减少了数据传输代价。
- (3) 提高系统的可靠性，局部系统发生故障,其他部分仍可继续工作。
- (4) 各地终端由数据通信网络相联。
- (5) 数据库位置透明，方便系统扩充。

分布式数据库系统兼顾集中管理和分布处理两项任务，具体结构如图所示



## 2. 面向对象数据库技术

主要有2个特点：

(1) 对象数据模型借用面向对象程序设计的思想，使用对象来描述现实世界的实体，能完整地描述现实世界的数据结构，表达数据间 嵌套、递归的联系。

(2) 具有面向对象技术的封装性（数据与操作定义一起）和继承性（继承数据结构和操作）的特点，提高了软件的可重用性。

如：Postgres SQL就是对象-关系数据库。

## 3. 面向应用领域的数据库技术

为了适应应用多元化的需求，结合各应用领域的特点，将数据库技术应用到特定领域，产生了工程数据库、地理数据库、统计数据库、科学数据库、空间数据库等多种数据库，同时也出现了数据仓库和数据挖掘等技术，使数据库领域中的新技术不断涌现。

如ORACl 12c、SQL Server 2012 等为云计算做好平台准备。

## ■ 数据库技术的发展趋势

- 1) 混合数据**快速发展**
- 2) 数据**集成**与**数据仓库**倾向内容管理
- 3) 主数据管理
- 4) 数据仓库将向内容展现和战术性分析方面**发展**
- 5) 基于网络的自动化**管理**
- 6) **NoSQL**数据库产品应用
- 7) 数据库将与业务语义的数据内容**融合**

## ■ **讨论思考：**

- (1) 数据管理技术经历了哪几个阶段？其特点如何？
- (2) 分布式数据库的主要特点有哪些？
- (3) 数据库技术的发展趋势是什么？



# NoSQL数据库技术

NoSQL最常见的解释是“**non-relational**”，“**Not Only SQL**”也被很多人接受。NoSQL仅仅是一个概念，泛指非关系型的数据库，区别于关系数据库，它们不保证关系数据的ACID特性。NoSQL是一项全新的数据库革命性运动，运用非关系型的数据存储，是一种全新的思维的注入。

## NoSQL有如下优点：

**(1) 易扩展：**NoSQL数据库种类繁多，但是一个共同的特点都是去掉关系数据库的关系型特性。数据之间无关系，这样就非常容易扩展。无形之间也在架构的层面上带来了可扩展的能力。

**(2) 大数据量，高性能：**NoSQL数据库都具有非常高的读写性能，尤其在大数据量下，同样表现优秀。这得益于它的无关系性，数据库的结构简单。

# NoSQL数据库的类型

## (1) 键值(Key-Value)存储数据库

使用哈希表，有一个特定的键和一个指针指向特定的数据。Key/value模型优势在于简单、易部署。缺点：如果DBA只对部分值进行查询或更新的时候，Key/value就显得效率低下了。举例如：Tokyo Cabinet/Tyrant, Redis, Voldemort, Oracle BDB。

## (2) 列存储数据库

用来应对分布式存储的海量数据。键值是指向了多个列，这些列是由列族来排列的。如：Cassandra, HBase, Riak。

# NoSQL数据库的类型

## (3) 文档型数据库

文档型数据库的灵感是来自于Lotus Notes办公软件的，而且它同键值存储相类似。该类型的数据模型是版本化的文档，半结构化的文档以特定的格式存储，比如JSON。文档型数据库可以看作是键值数据库的升级版，允许之间嵌套键值。而且文档型数据库比键值数据库的查询效率更高。如：CouchDB, MongoDB. 国内也有文档型数据库SequoiaDB，已经开源。

## (4) 图形(Graph)数据库

图形结构的数据库同其它结构的SQL数据库不同，它是使用灵活的图形模型，并且能够扩展到多个服务器上。NoSQL数据库没有标准的查询语言(SQL)，因此进行数据库查询需要制定数据模型。许多NoSQL数据库都有REST式的数据接口或者查询API。如：Neo4J, InfoGrid, Infinite Graph。

# NewSQL数据库技术

(1) NewSQL 是对各种新的可扩展/高性能数据库的简称，这类数据库不仅具有NoSQL对海量数据的存储管理能力，还保持了传统数据库支持ACID和SQL等特性。

(2) NewSQL是指这样一类新式的关系型数据库管理系统，针对OLTP（读-写）工作负载，追求提供和NoSQL系统相同的扩展性能，且仍然保持ACID和SQL等特性（scalable and ACID and (relational and/or sql -access)）。

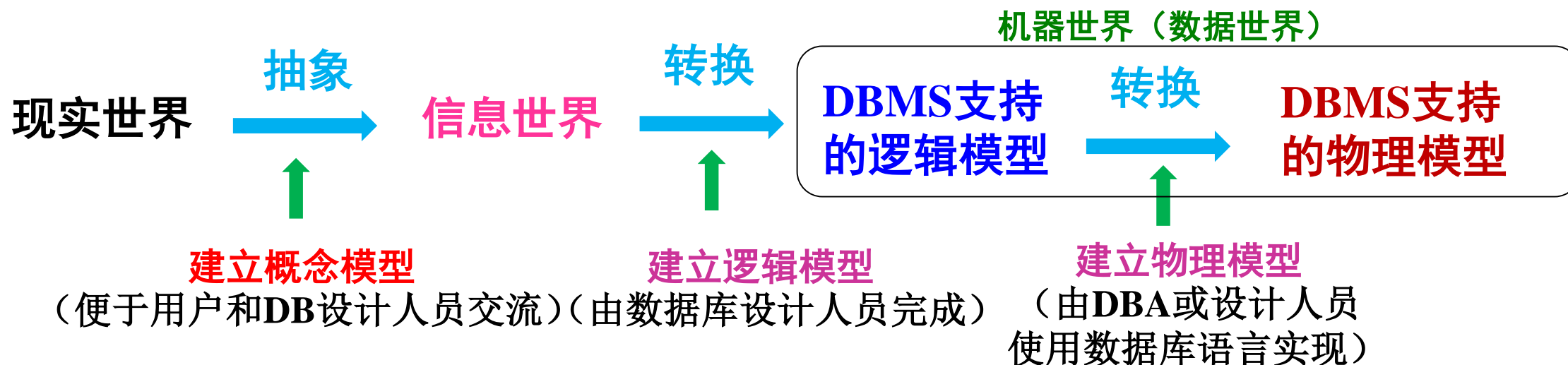
(3) NewSQL应是RDBMS中的数据模型和NoSQL中的计算框架、分布式存储、开源特定的结合产物。关于NewSQL是否从属于NoSQL仍有争议，但SQL化或者准确来说是relational化是趋势。NewSQL中的典型代表有Google Spanner和国内的TiDB。

(4) NewSQL两个显著的共同特点：1)它们都支持关系数据模型，2)它们都使用SQL作为其主要的接口。已知的第一个NewSQL系统叫做H-Store，它是一个分布式并行内存数据库系统。

## § 5 数据模型

**数据模型 (Data Model)** 是一种表示数据特征的抽象模型，是数据处理的关键和基础。专门用于抽象、表示和处理现实世界中的数据（信息）的工具，DBMS的实现都是建立在某种数据模型基础上的。数据模型通常由数据结构、数据操作和完整性约束（数据的约束条件）三个基本部分组成，称为数据模型的三要素。

数据处理的抽象过程（涉及三个领域）



# 数据模型的类型

根据模型的不同应用，可将模型分为三类。

(1) **概念模型**。概念模型也称**信息模型**，位于客观现实世界与机器世界之间。只用于描述某个特定机构所关心的数据结构，实现数据在计算机中表示的转换，是一种**独立于**计算机系统的数据模型，**用于数据库设计初期信息建模**。

(2) **逻辑（结构）模型(LDM)**。包括**网状模型、层次模型和关系模型等**是以计算机系统的观点对数据建模，是直接面向数据库**逻辑结构**，是对客观现实世界的第二层抽象。这类模型直接与DBMS有关，称为“**逻辑数据模型**”，简称为**逻辑模型**，又称为“**结构模型**”；**每种数据库管理系统都会采用其中某种逻辑模型**。

(3) **物理数据模型(PDM)**，用于**存储结构和访问机制**的更高层描述，描述数据是如何在计算机中**存储的**，如何表达**记录结构、记录顺序和访问路径**等信息。使用物理数据模型，可以在系统层实现数据库。数据库的物理设计阶段必须在此基础上进行详细的后台设计，包括数据库的存储过程、操作、触发、视图和索引表等。

# 概念模型的基本概念

## 1、实体与记录

### 信息世界

**实体**：客观存在并可相互区分的事物。

**实体集**：性质相同的同类实体的集合。

**属性**：实体具有的某一特性。

**实体标识符**：能将一个实体与其它实体区分开来的一个或一组属性。

### 数据世界

**记录**  $\longleftrightarrow$  实体（抽象表示）

**文件**  $\longleftrightarrow$  实体集

**字段或数据项**  $\longleftrightarrow$  属性

**关键字**  $\longleftrightarrow$  实体标识符, 唯一地标识一个记录, 又称**码**、**键**。

## 2、型与值

在DBS中，每一个对象广义上讲都有型与值之分：

**型**是对象的结构或特性描述，

**值**是一个具体的对象实例。

类似于程序设计语言中**数据类型**与**数据值**的概念。

**(1) 实体型**：对实体固有特性或结构的描述, 用实体名及其属性名集合来抽象和刻画。

**如：** 汽车（车牌号，车型，车主）

**实体值**：实体型的一个实例，即一个具体的实体。

**如：** （ 川A00001，丰田，张三）

**(2) 记录型**：记录格式。

**记录值**：一个具体的记录。



如：

车牌号	名称	车主
-----	----	----

豫A00001	丰田	张三
---------	----	----

### (3) 几点说明

- 区分型与值的实质
- DBS中讨论的重点是型（结构）
- 通常只说实体、记录，含义根据上下文自明

## 3、实体间的联系

实体内部的联系（属性间的联系）：

反映在数据上就是记录内部数据项间的联系

实体之间的联系：反映在数据上就是记录之间的联系

实体之间的联系可归结为三类：

(1) 1对1联系 (1 : 1) : 两个实体集中的每一个实体至多和另一个实体集中的有一个实体有联系。

如 国家 —— 总统  
学员队 —— 队长

(2) 1对多联系 (1 : n) : 若实体集A中的每个实体与实体集B中0个或多个实体有联系，而B中每个实体至多与A中的一个实体有联系，则称从A到B为1对多的联系。

如 国家 —— 部长  
学员队 —— 学员

(3) 多对多联系 (m : n) : 两个实体集中的每一个实体都和另一个实体集中0个或多个实体有联系。

如 学员 —— 课程

## DBS的核心问题之一：如何表示和处理实体及实体间的联系。

### 4、概念模型的表示方法之一：

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

用ER图（Entity-Relationship Diagram）描述：

**实体型**：用长方形表示

**联系**：用菱形表示

**属性**：用椭圆形表示

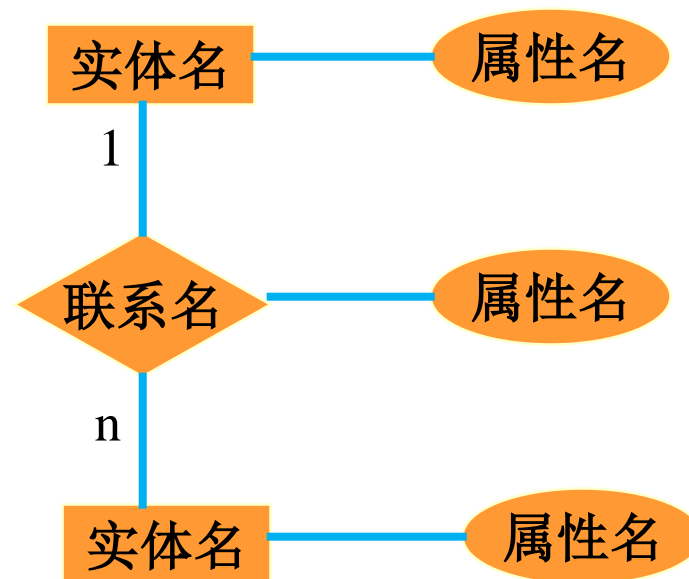
框内写上相应的名称

用无向边连接：

实体与其属性

联系与其属性

联系与有关实体，并标上联系类型



# 数据模型的组成要素

形式化描述数据、  
数据之间的联系  
以及数据操作  
和有关的语义  
约束规则的方法

数据结构（静态）

数据操作（动态）

完整性约束

如何表示  
实体及联系  
（难点是表示联系）

如何实现  
增删改查？

如何保证数据的  
约束条件得到满足

在数据库不同的发展阶段，采用了不同的数据模型

层次模型  
网状模型  
关系模型  
面向对象模型

**数据模型**通常由数据结构、数据操作和完整性约束**三要素**组成。

## 1. 数据结构

对于各种数据模型都规定了一种数据结构，即信息世界中的实体和实体之间联系的表示方法。数据结构描述了系统的**静态特性**，是数据模型本质的内容。**数据结构**是所研究的对象类型的集合。其对象是数据库的组成部分，包括**两类**，一类是与数据类型、内容、性质有关的对象；另一类是与数据之间联系有关的对象。

## 2. 数据操作

**数据操作**描述了系统的**动态特性**，是对于数据库中的各种对象（型）的实例（值）允许执行的操作的集合，包括**操作**及有关的**操作规则**。对数据库的操作主要有数据维护和数据检索**两大类**，是数据模型都必须规定的操作，**包括**操作符、含义和规则等。

## 3. 数据的约束条件

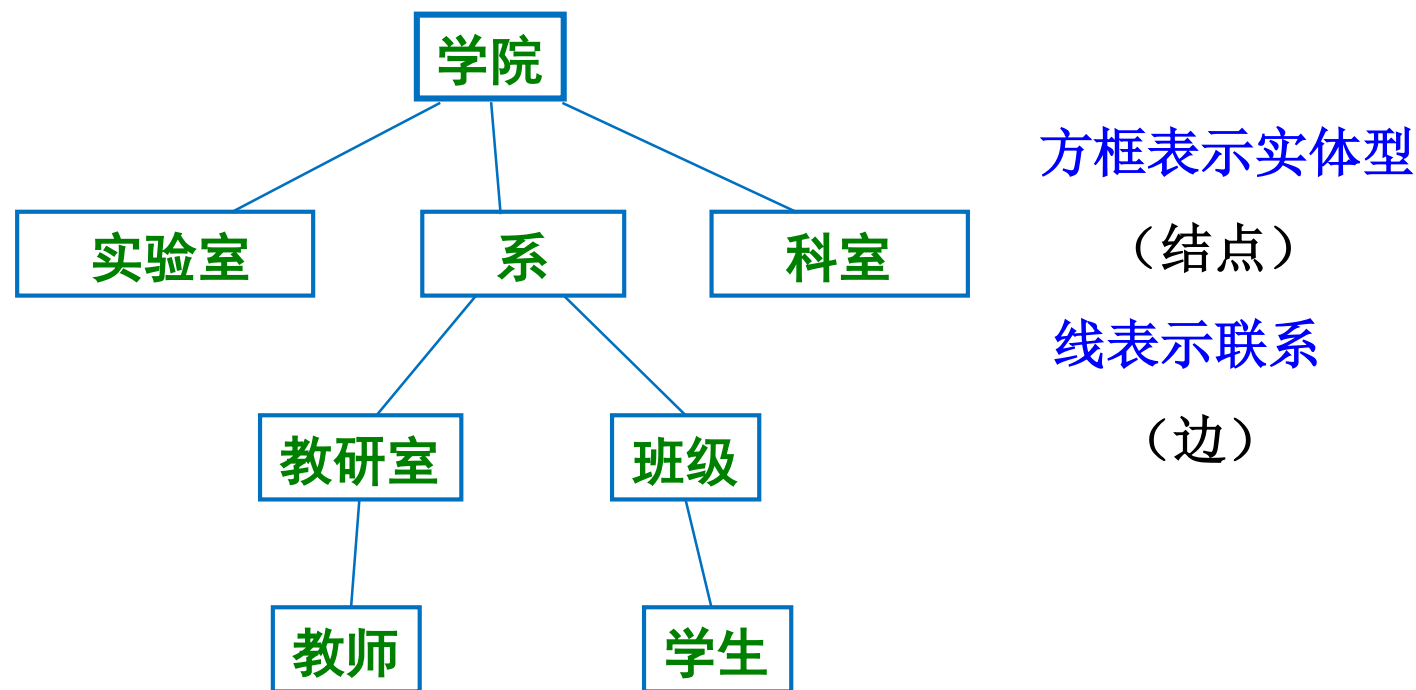
数据的约束条件是一组**完整性规则**的集合。完整性规则是给定的数据模型中的数据及其联系所**具有的制约和依存规则（条件和要求）**，用于限定符合数据模型的数据库状态以及状态的变化，以保证数据的正确、相容和有效。

# 层次数据模型

## 1).层次模型的结构

**层次模型** (Hierarchical model) 是一个**树状结构模型**，有且只有一个**根节点**，其余节点为其**子孙**；每个节点（除根节点外）只能有一个**父节点**（也称**双亲节点**），却可以有一个或多个**子节点**，当然也允许无子节点**被称为叶**；每个节点对应一个记录型，即对应概念模型中的一个实体型，每对节点的父子联系隐含为1对多(含1对1)联系。

2). 层次模型的**特点** 在这种模型的数据库系统中，要**定义和保存**每个节点的记录型及其所有值和每个父子联系。

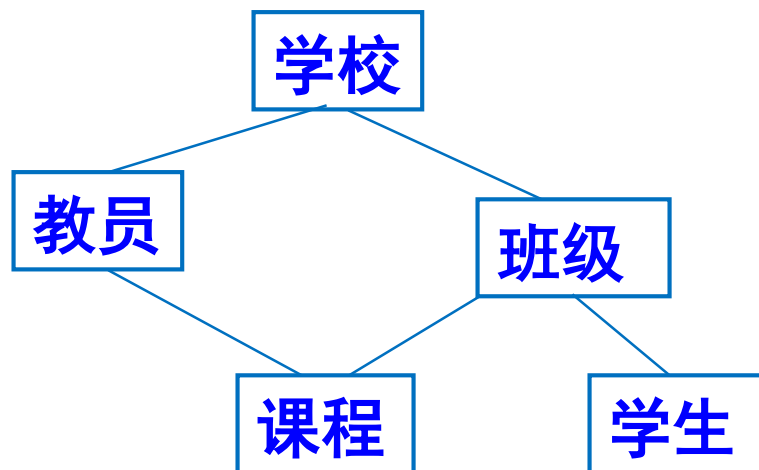


## 网状数据模型

(1) 定义：用图结构来表示实体以及实体间联系的模型。

其特征是：任一结点都可以无双亲或有一个以上的双亲。

例：



(2) 优：可表示m:n的联系，运行效率高

缺：过于复杂，实现困难

基于层次模型或网状模型的数据库称为第一代数据库

# 关系数据模型

**关系模型**（Relational model）是一种简单的二维表结构，概念模型中的每个实体和实体之间的联系都可以直接转换为对应的二维表形式。每个**二维表**称**做一个关系**，一个二维表的表头**称为关系的型**（结构），其表体（内容）称**做关系的值**。关系中的每一行**数据（记录）**称**做一个元组**，其**列数据**称**做属性**，**列标题**称**做属性名**。

属性(字段)

客户 ID	客户名称	注册日期	联系人 ID	类型
20120146	汽车公司	2008/2/12	31002894	大客户
20121832	大众宾馆	2009/6/13	20121832	大客户
20121845	王永贵	2007/9/18	20121845	流动
20122481	李为明	2003/7/25	20122481	流动

元组(记录)





# 面向对象模型

**面向对象模型**（Object-Oriented Model, OOM）是用面向对象观点来描述实体的逻辑组织、对象间限制、联系等模型。将客观世界的实体都模型化为一个对象，每个对象有一个惟一的标识。共享同样属性和方法集的所有对象构成一个对象类，简称为**类**，而一个对象就是某一类的一个实例。

## 1. 面向对象的概念

在面向对象的方法中，基本概念主要有**对象、类、方法和消息**。

**（1）对象。**对象是含有数据和操作方法的独立实体，是数据和行为的统一体。如一个城市、一座桥梁或高楼大厦，都可作为地理对象。对于一个对象具有如下**特征**：

- ①以一个唯一的标识，表明其存在的独立性；
- ②以一组描述特征的属性，表明其在某一时刻的状态；
- ③以一组表示行为的操作方法，用于改变对象的状态。

(2) **类**。类是共享同一属性和方法集的所有对象的集合构成类。从一组对象中抽象出公共的方法和属性，并将它们保存在一类中，是面向对象的核心内容。如汽车均具有共性，如品牌、颜色、长度等，以及相同的操作方法,如查询、计算长度、求数量等，因而可抽象为汽车类。被抽象的对象，称为实例，如轿车、公共汽车等。

(3) **消息**。消息是对象操作的请求,是连接对象与外部世界的唯一通道。

(4) **方法**。方法是对象的所有操作方式，如对对象的数据进行操作的函数、指令、例程等。

## 2) 基本思想

**面向对象的基本思想**是通过对问题领域进行自然的分割，以更接近人们通常思维的方式建立问题领域的模型，并进行结构模拟和行为模拟，从而使设计的软件能尽可能地直接表现出问题的求解过程。因此，面向对象的方法是将客观世界的一切实体模型化为对象。每一种对象都有各自的内部状态和运动规律，不同对象之间的相互联系和相互作用就构成了各种不同的系统。

## 2. 面向对象的特性

面向对象方法的包括特性具有**抽象性、封装性、多态性**等。

**(1) 抽象性。**抽象是对现实世界的简明表示。形成对象的关键是抽象，对象是抽象思维的结果。抽象思维是通过概念、判断、推理来反映对象的本质，揭示对象内部联系的过程。

**(2) 封装性。**封装是指将方法与数据放于同一对象中，以使对数据的操作只可通过该对象本身的方法来进行。

**(3) 多态性。**多态是指同一消息被不同对象接收时，可解释为不同的含义。

## § 6 数据库的三级模式结构

### 1. 数据模式

**数据模式**（Data Schema）是数据库中所有数据的逻辑结构和特征的描述。**型**（Type）是对某一类数据的结构和属性的具体描述说明，

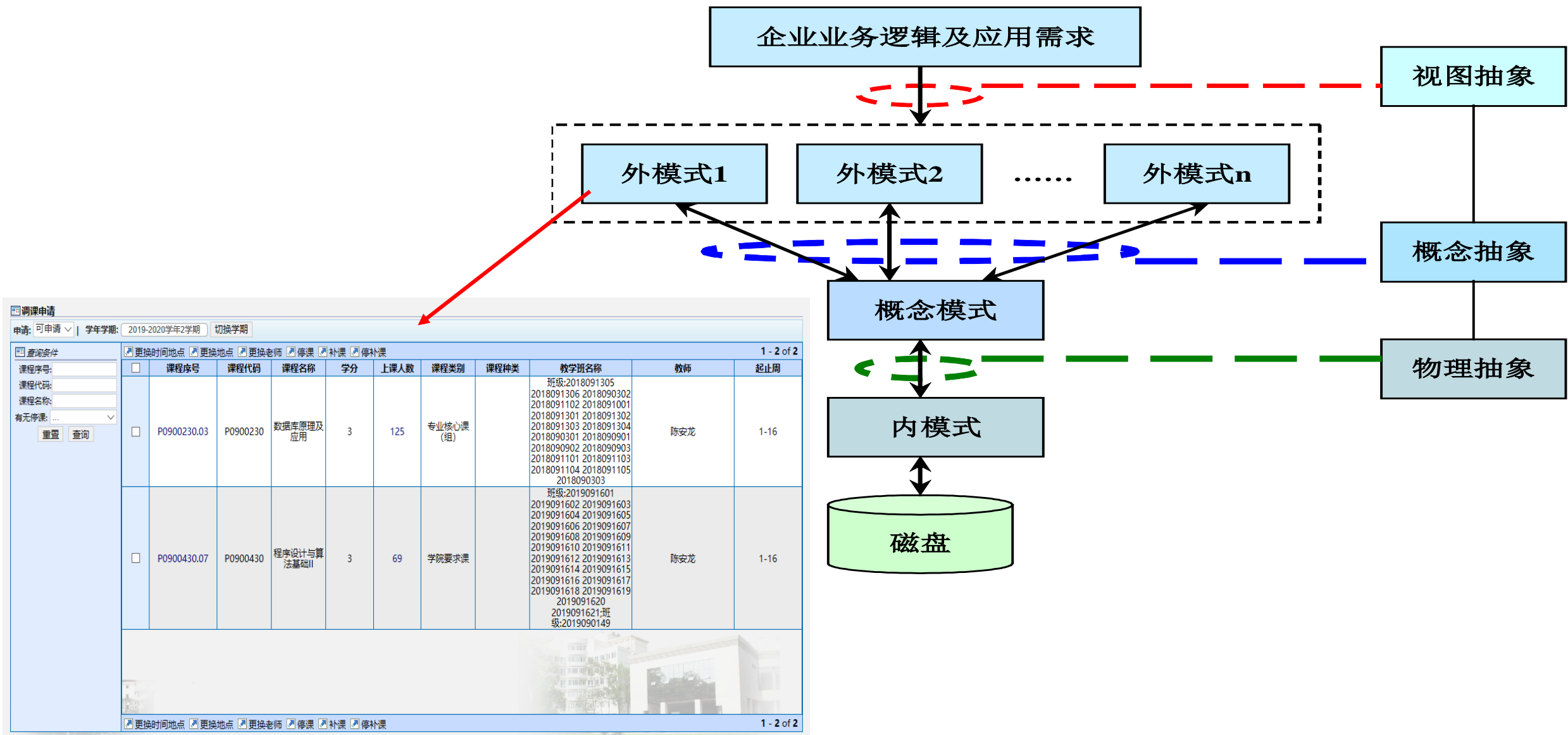
**值**（Value）是型的一个具体值。如**货物记录**的**型**定义为（货物编号，名称，种类，型号，颜色，产地，价格）称为**记录型**，而（K01101，服装，西服，XXL，黑色，上海，2800）则是该记录型的一个**记录值**。

**模式只涉及型的描述，而不涉及具体的值**。某数据模式下的一组具体的数据值称为数据模式的一个**实例**（Instance）。

### 2. 数据库的三级模式结构

**数据库系统的三级模式结构**，从逻辑上主要是指数据库系统由**内模式**、**模式**（**概念模式**）和**外模式**三级构成，且在这三级模式之间还提供了**外模式 / 模式映像**、**模式 / 内模式映像**，分别反映看待数据库的三个角度。

# 数据库的三级结构



- (1) **外模式** (External Schema) 也称**子模式** (Subschema) 或**用户模式**、**外视图**, 用于描述数据库数据的**局部逻辑结构和特征**。
- (2) **模式** (Schema) 也称**逻辑模式** (Logic Schema)、**概念模式** (Conceptual Schema) 或**概念视图**, 是数据库中所有数据的逻辑结构和特征的**描述**。如**关系型**。
- (3) **内模式** (Internal Schema) 也称**内视图**或**存储模式** (Storage Schema), 是三级模式结构中的最内层, 是靠近物理存储的一层, 即与实际存储的数据方式有关的一层, 是数据在数据库内部的**表示方式**, 详细**描述**了数据复杂的物理结构和存储方式, 由多个存储记录组成, 不必关心具体的存储位置。如**数据表**。
- (4) **三级模式结构的优点**
- ① 三级模式结构是数据库系统**最本质的系统结构**。
  - ② 数据共享。
  - ③ 简化用户接口 (交互)。
  - ④ 数据安全。

# 数据库的二级映像

数据的独立性由DBMS的二级映像功能实现，一般分为物理独立性和逻辑独立性两种。  
**物理独立性**是指数据的物理结构(包括存储结构、存取方式等)的改变，如更换存储设备或物理存储、改变存取方式等都不影响数据库的逻辑结构，从而不致引起应用程序的改变。**逻辑独立性**是指数据的总体逻辑结构改变时，如修改数据模式、改变数据间的联系等，不需要修改相应的应用程序。

## 1. 外模式 / 模式映像

**模式 / 内模式映像**位于概念级和内部级之间，用于定义概念模式和内模式之间的对应性。外模式描述数据的局部逻辑结构，模式描述数据的全局逻辑结构。数据库中的同一模式可以有任意多个外模式，对于每一个外模式，都存在一个外模式 / 模式映像。

## 2. 模式 / 内模式映像

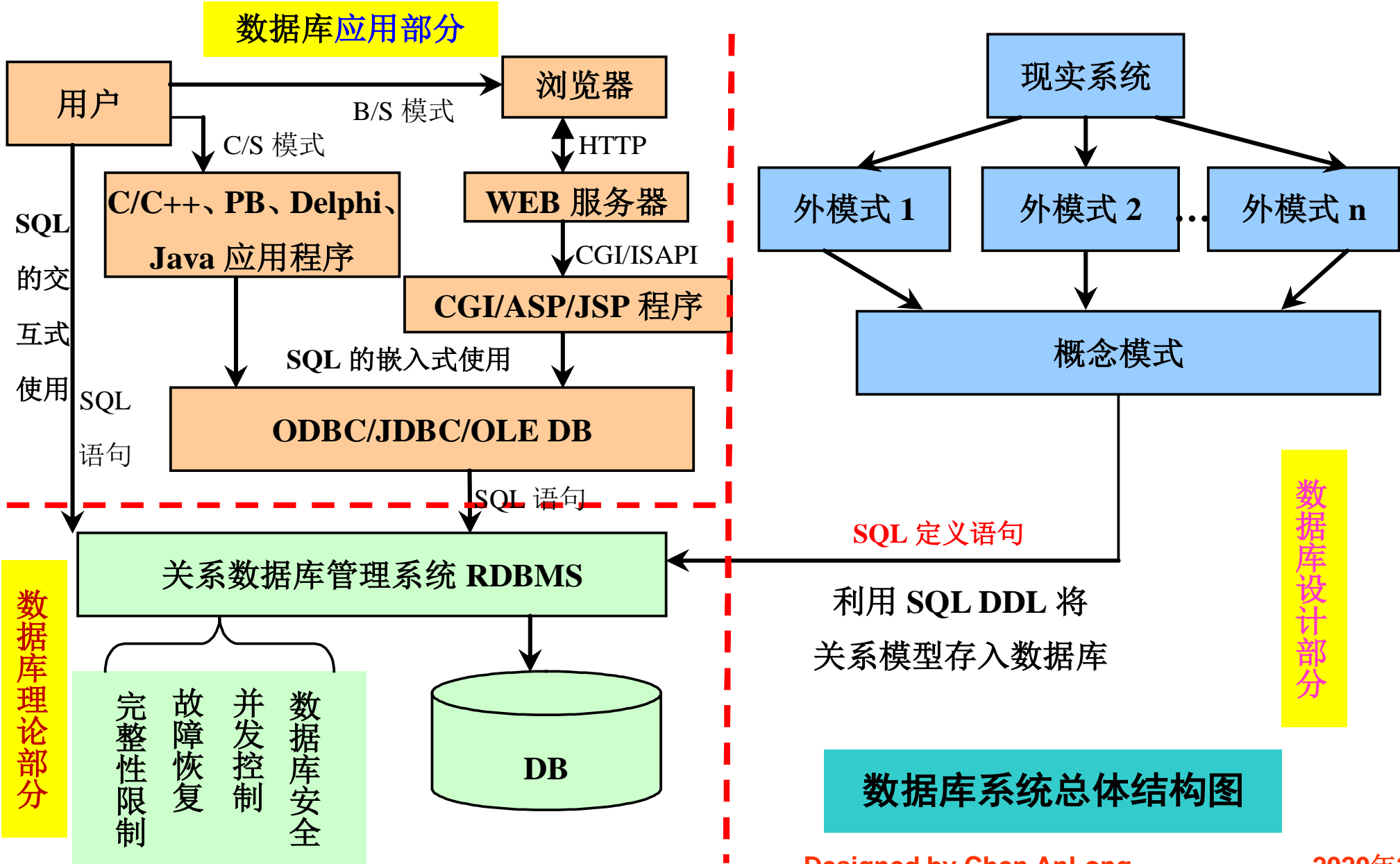
**外模式 / 模式映像**介于外部级和概念级之间，用于定义外模式和概念模式之间的对应性。数据库中的模式和内模式都只有一个，所以模式 / 内模式映像是唯一的。确定了数据的全局逻辑结构与存储结构之间的对应关系。

## 讨论思考

- (1) 什么是数据模式？请举例说明。
- (2) 什么是数据库系统的三级模式结构？并画图表示。
- (3) 数据的独立性如何由DBMS的二级映象功能实现？



# 数据库理论与应用、设计之间的关系？



## § 7 数据库应用系统开发阶段

### 1. 系统需求分析

系统分析人员与用户交流，按照软件工程方法获取**系统数据需求信息**，并采用模型定义系统**数据组成**，及其**数据字典**。

### 2. 系统数据架构设计

系统架构设计人员根据系统功能和性能需求，对系统数据架构模型进行设计，包括**系统数据概念模型**、**系统数据逻辑模型**和**系统数据物理模型**。

### 3. 系统数据库实现

系统数据库开发人员根据系统设计方案，对系统数据库进行部署和模型实现，包括**系统数据库创建**、**数据表创建**、**索引和约束等对象创建**。

## 4. 系统测试

系统测试人员将测试数据上载到数据库中，对数据库对象进行测试操作访问，实现数据库**功能和性能测试**。

## 5. 系统运行与维护

系统运维人员在系统投入运行过程中，对系统数据库进行**定期维护和优化**，如：**数据库服务器磁盘整理、数据备份**等，以保证数据库系统正常地、高效地运行。

## 本讲主要内容回顾

- (1) 理解为什么需要使用数据库？
- (2) 理解数据库的基本概念；
- (3) 理解数据库的数据模型；
- (4) 数据库ER数据模型和关系数据模型
- (5) 理解数据库系统的基本构架；
- (6) 理解并掌握数据库分层结构；

## 下一讲将讲解的主要内容

- (1) 数据库管理系统环境安装；
- (2) SQL语言及其使用。