



# 数据库系统概论



PART 01  
数据库  
基本概念



PART 02  
数据模型



PART 03  
数据管理技术  
的产生和发展



PART 04  
数据库管理系统  
的功能与特点



PART 05  
数据库系统  
的结构



PART 06  
数据库语言



PART 07  
数据库技术  
的新发展



## 成绩构成：

平时： 5%（Mooc上章节作业+课堂布置作业）

期中： 10%（第五章结束后，方式待定）

实验： 25%（Power designer+SQL server）

期末： 60%



➤ 链接入口: <https://www.icourse163.org/course/UESTC-1458643166>



➤ 昵称修改:

注册登录后, 将慕课昵称修改为

特定规则昵称: “名F姓名F”

例: 张小飞 慕课ID = 飞F张小飞F

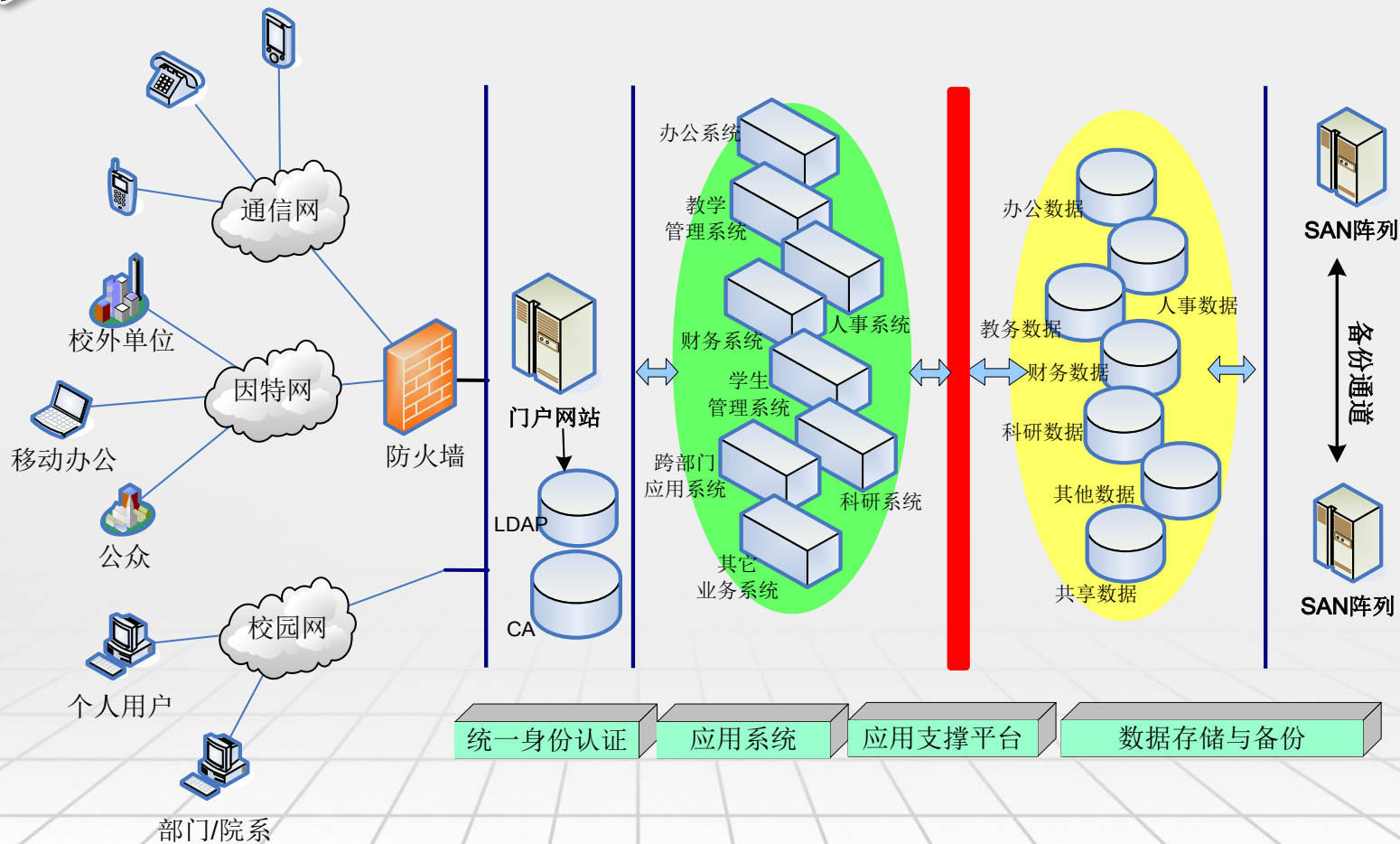




- 本章学习目标
  - 了解一些生活中的数据库应用实例
  - 掌握数据库的基本概念和相关术语
  - 了解数据库管理技术发展阶段
  - 掌握数据库系统构成
  - 理解并掌握数据库系统模式结构
  - 了解数据库管理系统的工作过程



## 0. 应用实例



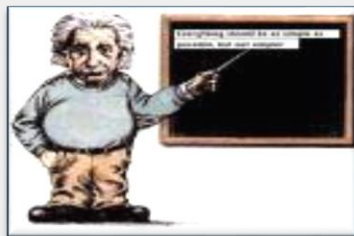


## 0. 应用实例

- 你还能举出生活中的哪些数据库应用实例？

- 校园学籍/成绩/财务信息管理系统
- 银行业务管理系统
- 电信业务管理系统
- 火车、飞机的订票管理系统
- 医院信息管理系统
- . . . . .

有哪些共性问题？



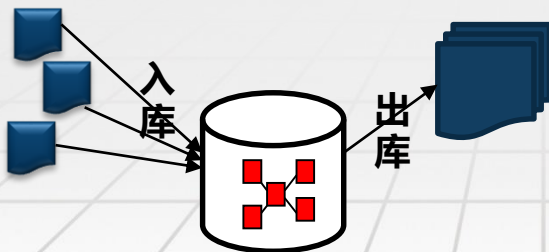
➤ 涉及大量数据的有效组织、存储和访问

- ◆ 减少数据存储冗余
- ◆ 实现数据共享
- ◆ 保障数据安全
- ◆ 高效检索和处理数据



## 0. 应用实例

- 医院信息管理系统（HIS）存取了哪些数据？
  - 业务对象（实体）
    - 患者、医生、药品、处方、设备。。。。
  - 业务活动（联系）
    - 挂号、化验、检查、领药、交费。。。。
  - 数据使用（检索）
    - 收入统计、病案分析、绩效评估、辅助决策。。。。



这不就是一个数据读写问题吗？我也学过了C或Java语言，编过数据读写程序，怎么还要开设这门数据管理课程呢？





## 0. 应用实例



➤ 大量数据在存储、访问、共享和管理过程中需要考虑哪些问题？

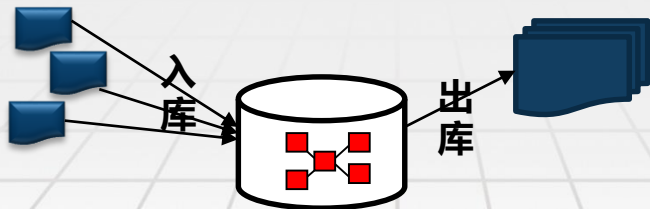
➤ 数据库中会存储哪些业务数据？请你举一个实例进行说明。





## 0. 应用实例

- 存取数据存在哪些问题？
  - 如何组织这些数据？ —— 数据模型、规范化理论、设计方法
  - 如何存取这些数据？ —— 数据定义和操作语言
  - 哪些人可以操作哪些数据？ —— 安全性控制
  - 多人如何操作同一数据？ —— 并发性控制
  - 出现故障后怎么办？ —— 数据恢复
  - 如何分析数据和发现数据价值？ —— 数据仓库、数据挖掘





# 1. 数据库基本概念

基本术语



数据

Pno	Pname	Pid	Pino	Pmno	Psex	Pbd	Padd
161	刘景	142201198702130061	1201676	6781121941	男	1987-2-13	新华路光源街
181	陈禄	142201196608190213	1204001	5461021938	男	1966-8-19	城建路茂源巷
201	曾华	142201197803110234	0800920	1231111932	男	1978-3-11	新建路柳巷
421	傅伟相	142202199109230221	0700235	4901021947	男	1991-9-23	高新区西源大道
481	张珍	142201199206200321	1200432	3451121953	女	1992-6-20	西湖区南街
501	李秀	142203198803300432	0692015	3341111936	女	1988-3-30	泰山大道北路

信息



## ● 数据(Data)

数据必须是保存在计算机中，能够被计算机识别、存储、处理的。

目前数据不仅包括数字、字母、文字和其他特殊字符，而且还包括图形、图像、声音等多媒体数据。

## ● 信息(Information)

信息是经过加工处理的数据，是对数据的具体描述。数据和信息既有联系又有区别。数据是信息的载体，而信息则是对数据的语义解释。

例如，我们可以从上表中得知刘景生于1987年，家住新华路光源街。这便是我们获取到的信息。



## ● 数据库

数据库（Database，简称DB）是长期存储在计算机内的、**有组织的、可共享的**大量**数据的集合**。





## ● 数据库管理系统

数据库管理系统是**位于用户与操作系统之间的一层数据管理软件**，主要目标是使数据成为方便各种用户使用的资源，并提高数据的安全性、完整性和可用性。



位于用户和操作系统之间



## ● 元数据

元数据即描述数据的数据，相当于**数据字典**。主要是**描述数据属性（property）的信息**，如数据的类型，格式，存储大小等。

Pno	Pname	Pid	Pino	Pmno	Psex	Pbd	Padd
161	刘景	142201198702130061	1201676	6781121941	男	1987-2-13	新华路光源街
181	陈禄	142201196608190213	120400	5461021938	男	1966-8-	城建路茂
201	曾华	142201197803110234					
421	傅伟相	142202199109230221					
481	张珍	142201199206200321					
501	李秀	142203198803300432					

属性名称	中文释义	类型	键描述
Pno	患者编号	Integer	PK
Pname	患者姓名	Char	
Pid	身份证号	Char	
Pino	社会保险号	Char	
Pmno	医疗卡识别号	Char	
Psex	性别	Char	
Pbd	出生日期	Date	
Padd	地址	Char	

数据

元数据



- 数据模型的基本概念

数据模型是**对现实世界数据特征的抽象**和**对现实世界的模拟**。

- 数据模型的组成要素

数据模型的组成要素之一是**数据结构**。

数据模型的组成要素之二是**数据操作**。

数据模型的组成要素之三是**数据完整性**。



### ● 概念数据模型

概念数据模型（Conceptual Data Model）是用户容易理解的、对现实世界特征的数据抽象，它与具体的DBMS无关，是数据库设计员与用户之间进行交流的语言。常用的概念数据模型是实体—联系（Entity-Relationship）模型。

E-R图的相关概念：

- 实体：客观存在并可相互区别的事物
- 属性：实体具有的若干特征
- 实体集：性质相同的同类实体的集合
- 键（码）：唯一标识实体集中每个实体的属性集合
- 域：属性的取值范围
- 联系：反映事物之间的联系

一对一联系 (1:1)  
一对多联系 (1:n)  
多对多联系 (m:n)



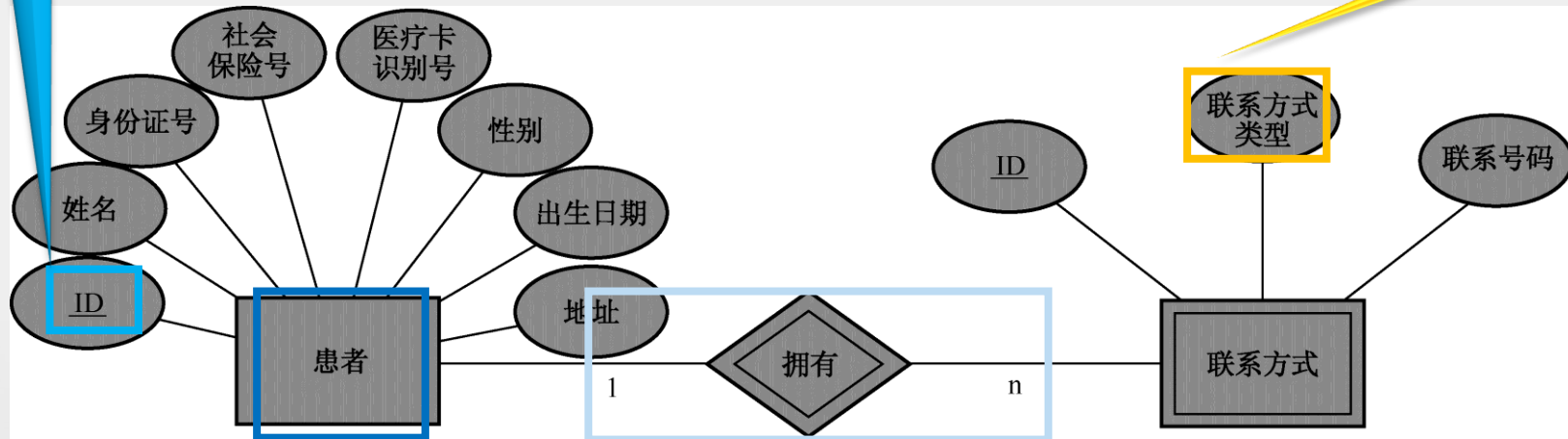


## 2. 数据模型



键

属性



实体

(一对多) 联系

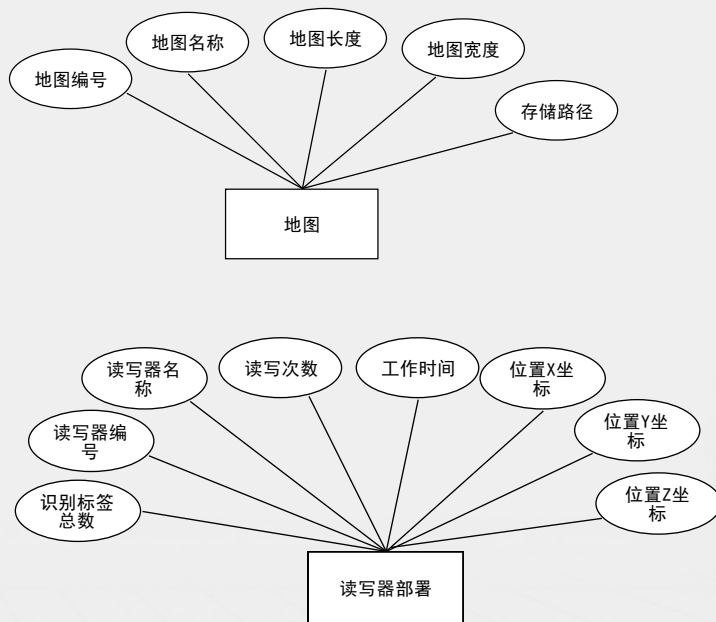
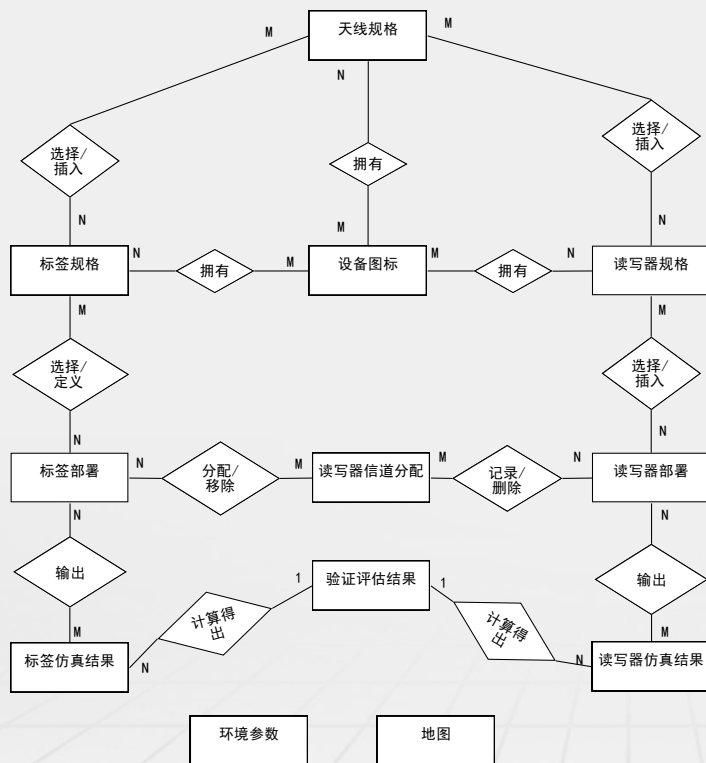
E-R图的表示规则:

实体集用**矩形**，属性用**椭圆形**，

联系用**菱形**，以上之间连接用**无向边**。



## 2. 数据模型



UESTC某实际科研项目中的E-R图设计



### ● 逻辑数据模型

逻辑数据模型即数据模型，是用户从数据库中所看到的数据模型，是具体的DBMS所支持的数据模型。

主要有层次模型，网状模型，关系模型，面向对象模型。

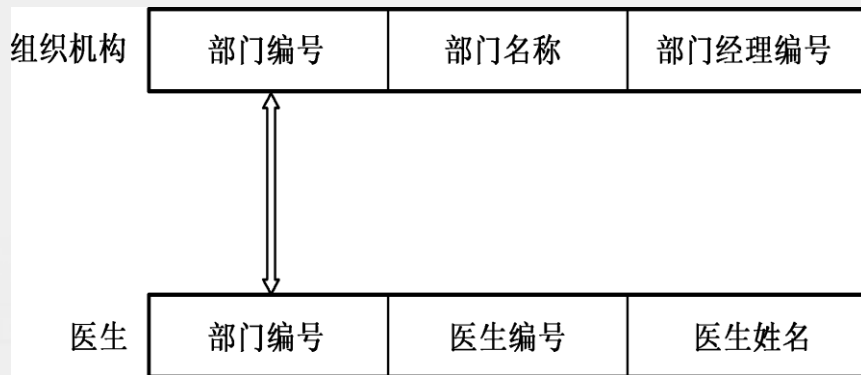
目前，数据库领域中常用的数据模型是关系模型和面向对象模型。

层次模型、网状模型已渐渐被淘汰。



### ● 层次模型

层次模型用**树形（层次）**结构来表示各类实体以及实体间的联系，每一节点表示一个记录类型（实体型），每个记录类型包含若干个字段（实体的属性）。**层次模型数据结构简单清晰。**





### ● 层次模型的缺点

- (1) 实现复杂
- (2) 难于管理
- (3) 实现的限制
- (4) 缺乏标准
- (5) 缺乏结构独立性
- (6) 应用程序编写和使用复杂性



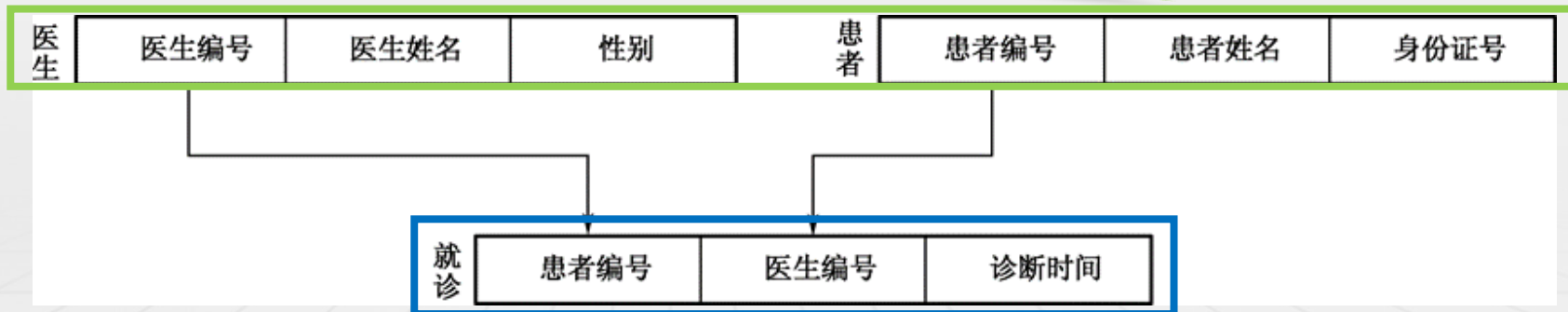
## 2. 数据模型



### ● 网状模型

网状模型是一种比层次模型更具普遍性的结构。它去掉了层次模型的两个限制，**允许多个结点没有双亲节点，允许结点有多个双亲节点。**

允许多个节点没有双亲节点



允许节点有多个双亲节点



### ● 网状模型的**优点**

网状模型**保留**了许多层次数据库模型的优点，同时，它也对层次数据库进行了改进。

- (1) 能够更加方便地描述现实世界，结点之间可以有多种联系。
- (2) 具有良好的性能，存取效率高。

### ● 网状模型的**缺点**

- (1) 结构复杂性
- (2) 系统复杂性
- (3) 缺乏结构独立性
- (4) 用户不容易掌握和使用



### ● 关系模型

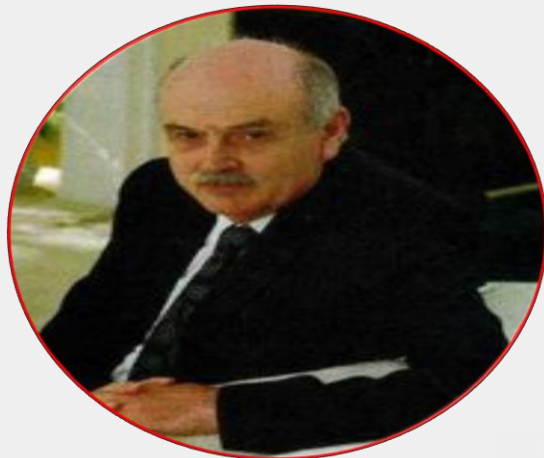
关系模型是目前最重要的一种数据模型。

**关系数据库系统**采用**关系模型**作为数据的组织方式。

### ● 关系模型的主要优点

(1) 结合了层次和网状的优点，摒弃了它们的缺点；

(2) 强烈的数学支撑 - 关系代数

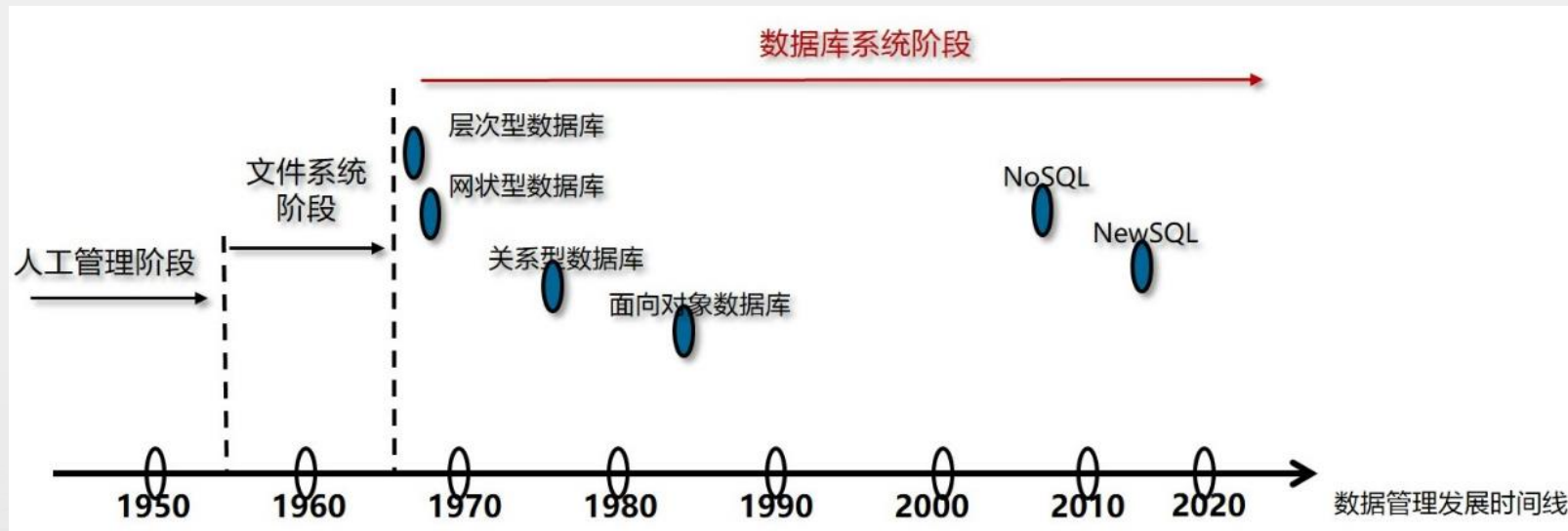


"A Relational Model of Data for  
Large Shared Data Banks"  
(大型共享数据库的关系数据模型)





### 3. 数据管理技术的产生和发展



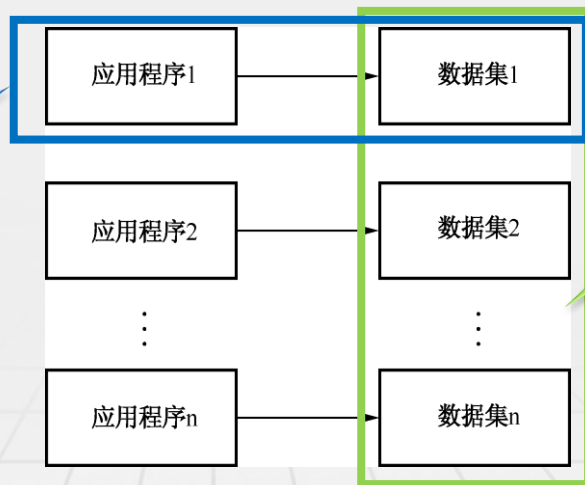


### 3. 数据管理技术的产生和发展

- 人工管理阶段的特点

(1) 数据不能共享

(2) 数据无独立性



数据无独立性

数据不能共享



### 3. 数据管理技术的产生和发展

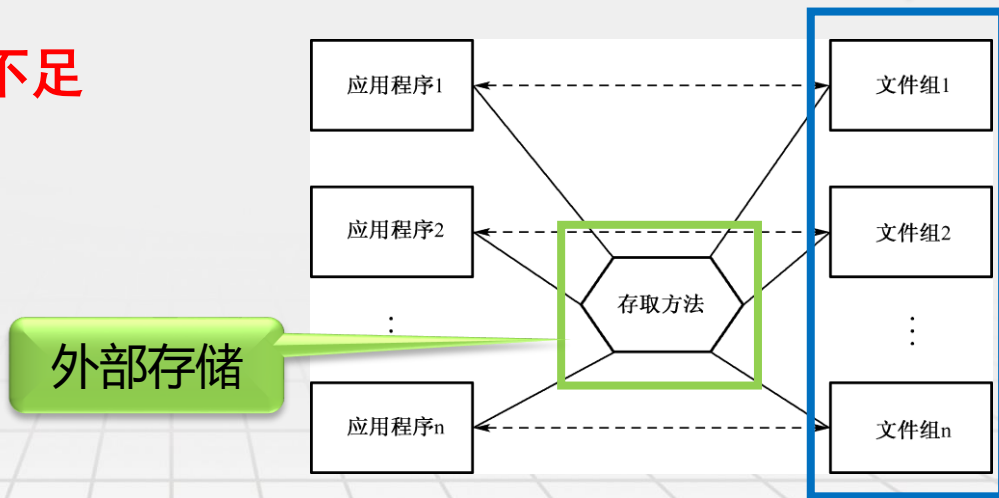
#### ● 文件系统阶段的**特点**

- (1) 数据以“文件”形式可长期保存在外部存储器的磁盘上。
- (2) 数据的逻辑结构与物理结构有了区别，但比较简单。

数据联系弱

#### ● 文件系统阶段的**不足**

- (1) 数据冗余
- (2) 不一致性
- (3) 数据联系弱



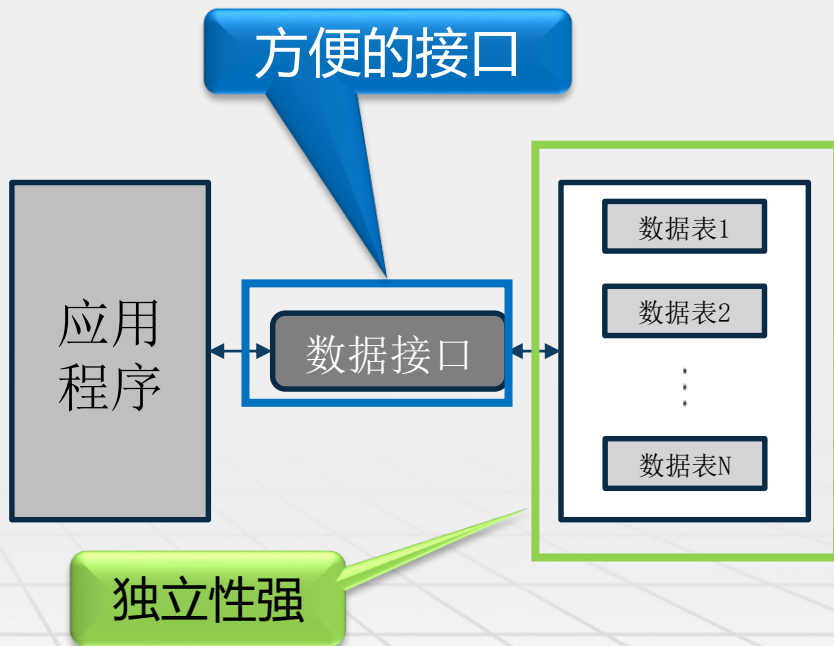


### 3. 数据管理技术的产生和发展

#### ● 数据库管理系统阶段的特点

- (1) 采用数据模型表示复杂的数据结构。
- (2) 有较高的数据独立性。
- (3) 数据库系统为用户提供了方便的用户接口。
- (4) 数据库系统提供了数据控制功能。

- 数据库的并发控制
- 数据库的恢复
- 数据完整性
- 数据安全性





### 3. 数据管理技术的产生和发展

阶段对比



		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	大规模管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘
	软件背景	没操作系统	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	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
	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	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力



### 3. 数据管理技术的产生和发展

阶段对比





## 4. 数据库管理系统的功能和特点



### ● 数据定义功能

提供数据定义语言**DDL (Data Definition Language)**，对各级数据模式进行精确定义。

### ● 数据操纵功能

提供数据操纵语言**DML (Data Manipulation Language)**，可以对数据库中的数据进行**追加、插入、修改、删除、检索**等操作。







## 4. 数据库管理系统的功能和特点

### ● 数据库运行控制功能

提供数据控制语言**DCL (Data Control Language)**，可以对数据库中的数据进行并发控制功能、数据的安全性控制、数据的完整性控制。

### ● 数据组织、存储和管理功能

DBMS分类组织、存储和管理各种数据，包括数据字典、用户数据、数据的存取路径。确定以何种文件结构和存取方式在存储级上组织这些数据，实现数据之间的联系。

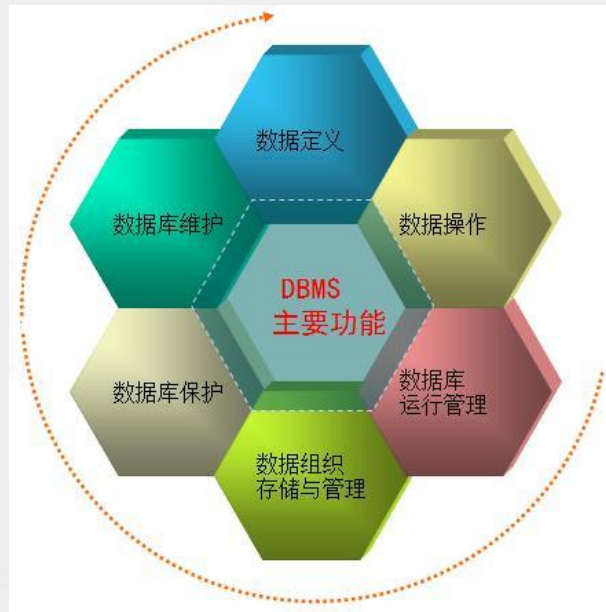






### ● 数据库的建立和维护功能

- (1) 数据库初始数据的输入
- (2) 转换功能
- (3) 数据库的转储
- (4) 恢复功能
- (5) 数据库的重组功能 and 性能监视
- (6) 分析功能



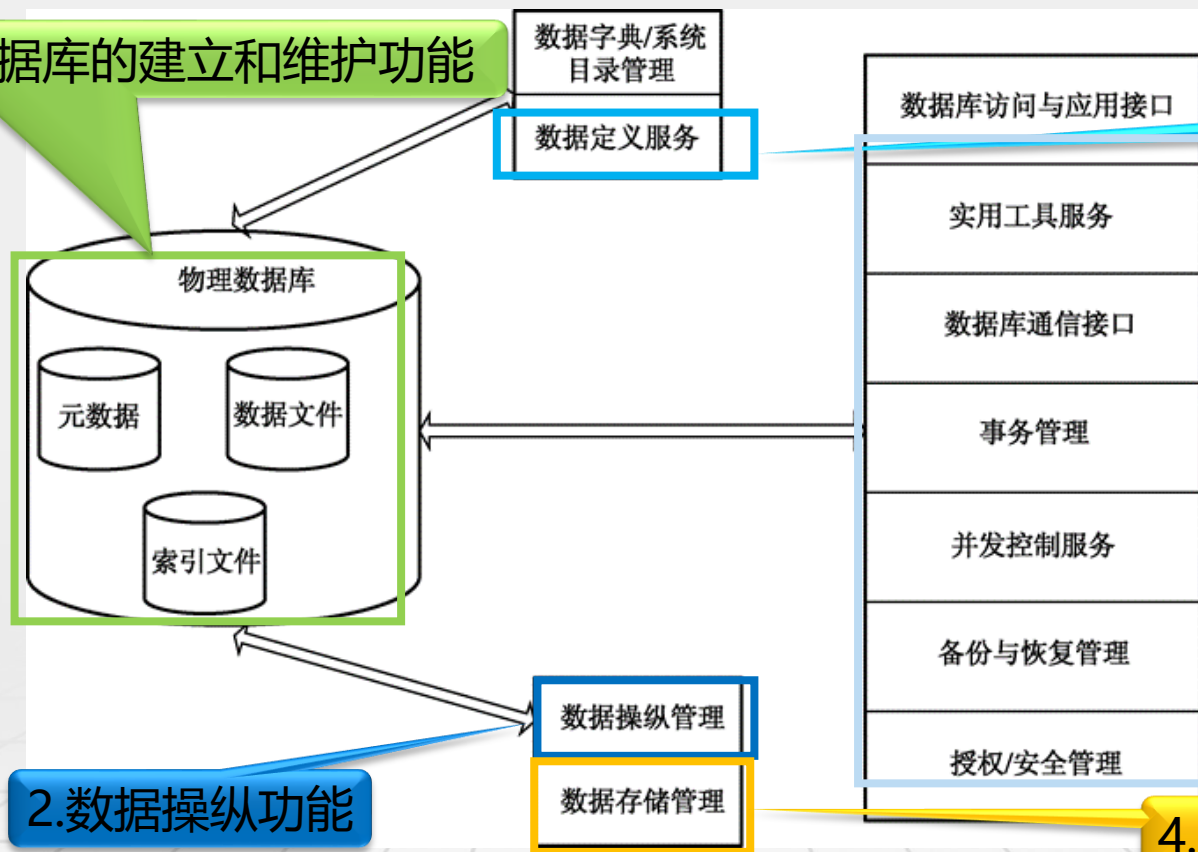


## 4. 数据库管理系统的功能和特点

DBMS的功能



### 5. 数据库的建立和维护功能



### 1. 数据定义功能

### 3. 数据库运行控制功能

### 2. 数据操纵功能

### 4. 组织, 存储和管理功能



### ● 数据结构化

数据库系统实现**整体数据的结构化**，这是数据库的主要特征之一，也是数据库系统与文件系统的本质区别。

### ● 数据的共享性高、冗余度低，易扩充

数据库从整体的观点来看待和描述数据，数据不再是面向某一应用，而是面向整个系统。这样就减小了数据的冗余，节约存储空间，缩短存取时间，避免数据之间的不相容和不一致。



## 4. 数据库管理系统的功能和特点



- 数据独立性高

数据的独立性主要包含**逻辑独立性**和**物理独立性**。

- 数据由DBMS统一管理和控制

统一的数据管理功能，包括数据的**安全性控制**、**数据的完整性控制**及**并发控制**。

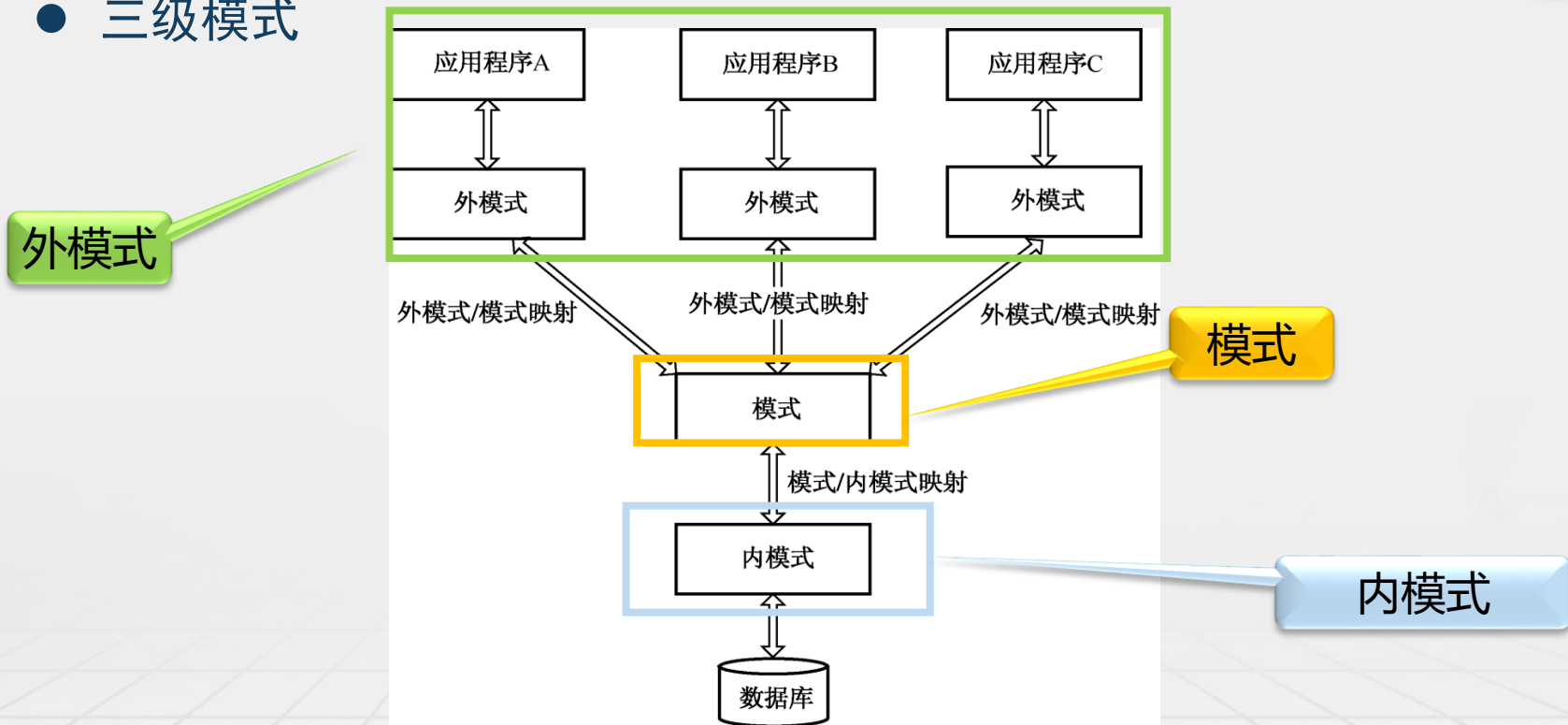


## 5. 数据库系统的结构

三级模式



### ● 三级模式



数据库系统的三级模式结构是指数据库系统是由**外模式**、**模式**、**内模式**三级构成。



### ● 外模式

外模式也称子模式或用户模式，它是数据库用户看见和使用的**局部**数据的**逻辑结构****和特征的描述**，是数据库用户的数据视图，是与某一应用程序有关的数据的逻辑表示。

外模式通常是模式的子集，**一个数据库可以有多个外模式。**



### ● 模式

**模式也称逻辑模式，是数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图。**

它是数据库系统模式结构的中间层，既不涉及数据的物理存储细节和硬件环境，也与具体的应用程序、所使用的应用开发工具及高级程序设计语言无关。

**一个数据库只有一个模式。**



### ● 内模式

内模式也称存储模式，一个数据库只有一个内模式。

它使用一个物理数据模型，全面描述了数据库中数据存储的全部细节和存取路径，是数据在数据库内部的表示方式。

例如，记录的存储方式是顺序存储、按照B树结构存储还是按hash方法存储；索引按照什么方式组织。





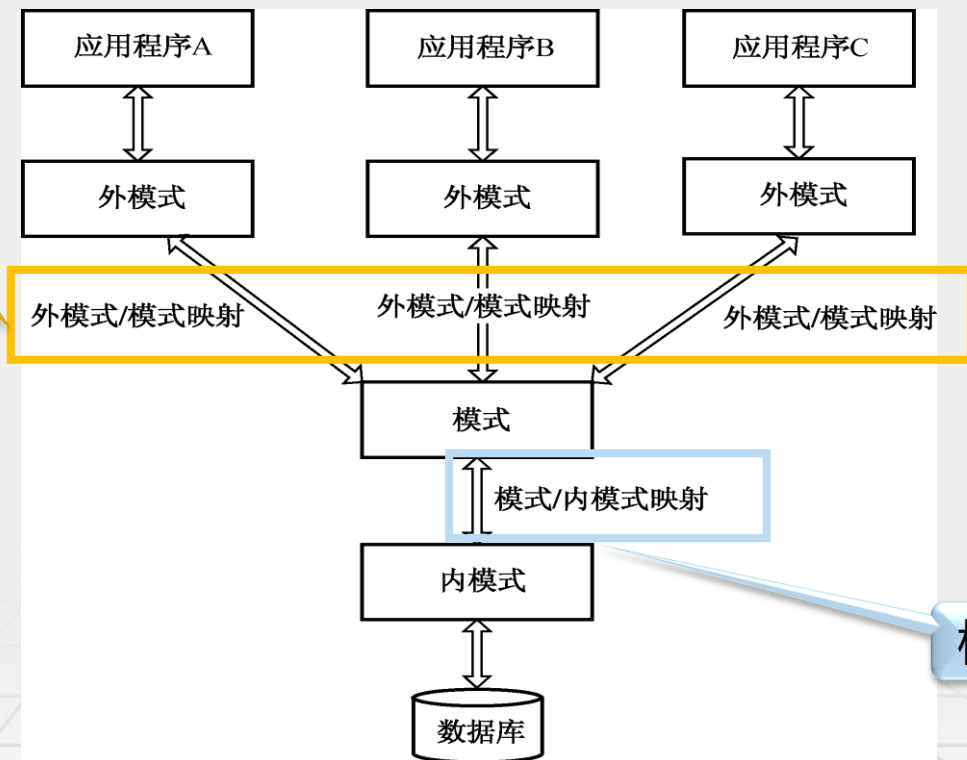
## 5. 数据库系统的结构

二级映射



### ● 二级映射

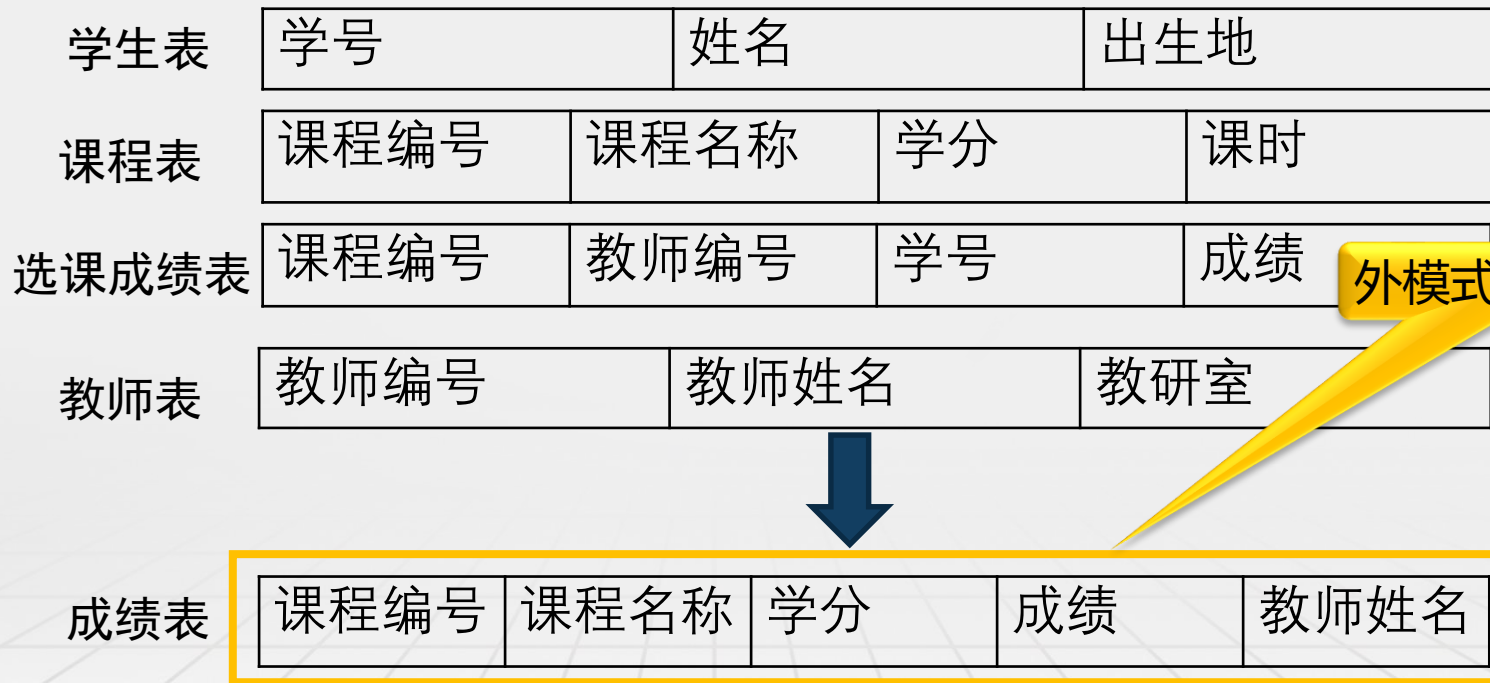
外模式/模式映射



模式/内模式映射



### ● 外模式/模式映射





### ● 模式/内模式映射

数据库只有一个模式和一个内模式，所以模式/内模式映像是唯一的。

它定义了数据全局逻辑结构和存储结构之间的对应关系。

当数据库的存储结构改变了，由DBA对模式/内模式作相应改变，可以使模式保持不变，从而保证了数据的物理独立性。



### ● 数据独立性

三层体系结构的一个主要目的是保证数据的独立性，这意味着对较低层的修改不会对较高层造成影响。数据的独立性分为**逻辑独立性**和**物理独立性**两类。

### ● 逻辑独立性

逻辑数据独立性是指外部模式不受概念模式变化影响。

### ● 物理独立性

物理数据独立性是指概念模式不受内部模式变化的影响。



### ● 客户-服务器结构

客户-服务器（Client/Server, C/S）结构是常见的数据库系统结构。在这种结构中，客户机提出请求，服务器对客户机的请求做出回应。

### ● 浏览器-服务器结构

浏览器-服务器（Browser/Server, B/S）结构是针对C/S结构的不足而提出的。在B/S结构中，客户机端仅安装通用的浏览器软件，实现用户的输入/输出，而应用程序不再安装在客户机端，而是在服务器端安装与运行。

### ● 单用户结构

单用户结构适合早期的最简单的数据库系统。



### ● 主从式结构

主从式结构也称为集中式结构，指的是一台主机连接多个用户终端的结构

### ● 分布式结构

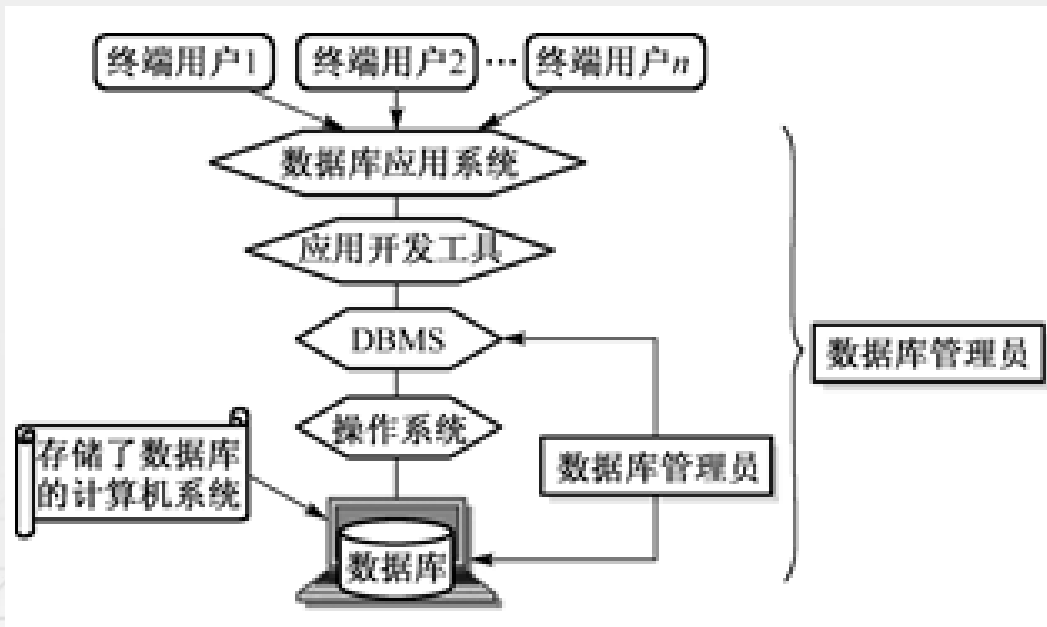
分布式结构的数据库系统是指数据库中的数据逻辑上是一个整体，物理上分布在计算机网络中的不同节点上。



## 5. 数据库系统的结构

数据库系统由五部分组成：

硬件系统、数据库集合、系统软件、数据库管理员和用户





### ● 硬件系统

运行数据库系统的计算机需要有足够大的**内存**、足够大容量的**磁盘**等联机直接存取设备和较高的通道能力及支持对外存的频繁访问；还需要足够数量的**脱机存储介质**，如光盘、磁盘等存放数据库的备份。

### ● 数据库集合

系统包括若干个设计合理、满足应用需要的数据库。

### ● 系统软件

系统软件指的是数据库系统中被计算机使用的程序的集合。需要三种类型的软件来实现数据库系统的全部功能：**操作系统**、**DBMS软件**和**应用程序**。





- 数据库管理员 (Database Administrator, DBA)

数据库系统一般需要专人来对数据库进行管理，这个人称为数据库管理员 (DBA)。数据库管理员负责数据库系统建立、维护和管理。

- 用户

数据库系统必然涉及不同的用户。数据库的用户分为两类：一类是最终用户，另一类是专业用户，即应用程序员。



DB、DBMS和DBS三者之间的关系是\_\_\_\_\_？

- ☐ A DB包括DBMS和DBS
- ☐ B DBS包括DB和DBMS
- ☐ C DBMS包括DB和DBS
- ☐ D 不能相互包括



### ● 数据定义语言

数据定义语言用来定义数据的结构，如创建、修改或者删除数据库对象。

### ● 数据操纵语言

用户通过数据操纵语言可以实现对数据库中的数据进行追加、插入、修改、删除、检索等操作，主要包括向数据库中插入新的信息、从数据库中删除信息和修改数据库中存储的信息。

### ● 数据控制语言

数据控制语言用来授予或回收访问数据库的某种权限，并控制数据库操纵事务发生的时间及效果，对数据库实行监视等。

### ● 事务控制语言

事务控制语言用于提交或回滚事务。



- 趋势一：对于异构数据的支持

为了让异构数据能够像结构化数据那样进行管理和查询。

- 趋势二：对于网络计算的支持

网络运算技术可以让用户更好地在网络环境中分享存储资源，并且同时可以保障数据在安全方面的需求。网络数据库最大的优势在于，数据库可以利用这个技术，将一个数据库应用部署在多台独立的服务器中，实现一个高容错的运算平台，以提高数据库应用的稳定性，减少数据库宕机的时间。

- 趋势三：管理的智能化

如何让数据库系统自动优化资源用量。



- 新型的NoSQL数据库Redis
- Key-Value数据库

以Key-Value数据存储为基础的数据库管理系统，以键值对的数据模型存储数据，并提供持久化机制和数据同步等功能。

Key-Value数据库系统的**主要特性**：

- (1) 高吞吐和海量存储
- (2) 较强的扩展性能
- (3) 数据格式灵活

Key(ID)	Value(Object)
1	Byte[]
2	Byte[]



### ● Key-Value数据库产品分类

#### (1) 满足高读写性能需求

此类产品以Redis、Memcached为代表。Redis是一种Key-Value缓存数据库服务器，其查询操作是通过缓存数据库完成的。

#### (2) 满足文档操作

此类产品以MongoDB、CouchDB为代表。MongoDB是一种功能较为强大的文档型数据库，具有扩展性强和操作灵活等特点。

#### (3) 满足列操作特性

此类产品以Cassandra、Bigtable为代表。Cassandra是由多个面向列的Key-Value数据库子节点所构成的分布式网络数据库系统。



### ● 内存数据库

将数据放在内存中直接操作的数据库。典型的内存数据库有SQLite、eXtremeDB和Oracle TimesTen。

### ● 云数据库

一种即开即用、稳定可靠、可弹性伸缩的在线数据库服务。典型的云数据库有Google Cloud SQL、Microsoft Azure、阿里云、华为云等。



### ● 大数据的特点

#### (1) 海量的数据规模 (Volume) :

大数据技术处理的数据量往往超过PB级，数据容量增长的速度大大超过硬件技术的发展速度，从而引发了数据存储和处理危机。 **TB、PB、EB、ZB、YB、BB、NB、DB、CB、XB**

#### (2) 处理速度快 (Velocity) :

这是大数据区别于传统数据挖掘的最显著的特征，在海量的数据面前，处理数据的效率就是企业的生命。

#### (3) 数据类型繁多 (Variety) :

除了结构化数据外，如存储在关系数据库二维表中的数据，目前，互联网中如图片、声音和视频等非结构化数据占到了很大的比重。

#### (4) 价值密度低 (Value) :

价值密度的高低与数据总量的大小成反比。以视频为例，一部1小时的视频，在连续不间断的监控中，有用数据可能仅有一二秒。