

数据库系统之一

--基础知识与关系模型

战德臣

哈尔滨工业大学 教授·博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on **I**ntelligent
Computing for **E**nterprises & **S**ervices,
Harbin **I**nstitute of **T**echnology

第2讲 数据库系统的结构抽象与演变

战德臣

哈尔滨工业大学 教授·博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on Intelligent
Computing for Enterprises & Services,
Harbin Institute of Technology

本讲学习什么？



基本内容

1. 数据库系统的标准结构？
2. 数据模型？
3. 数据库系统的演变与发展？

重点与难点

- 一组概念的区分：三级模式两层映像，物理独立性和逻辑独立性
- 一组概念的区分：数据→模式→数据模型
- 几种数据模型的差异：网状/层次模型→关系模型→OO数据模型

数据库系统的标准结构

战德臣

哈尔滨工业大学 教授.博士生导师

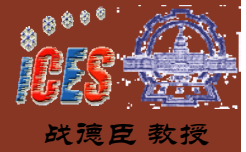
黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on **I**ntelligent
Computing for **E**nterprises & **S**ervices,
Harbin **I**nstitute of **T**echnology

数据库系统的标准结构

(1)数据库系统的分层抽象？

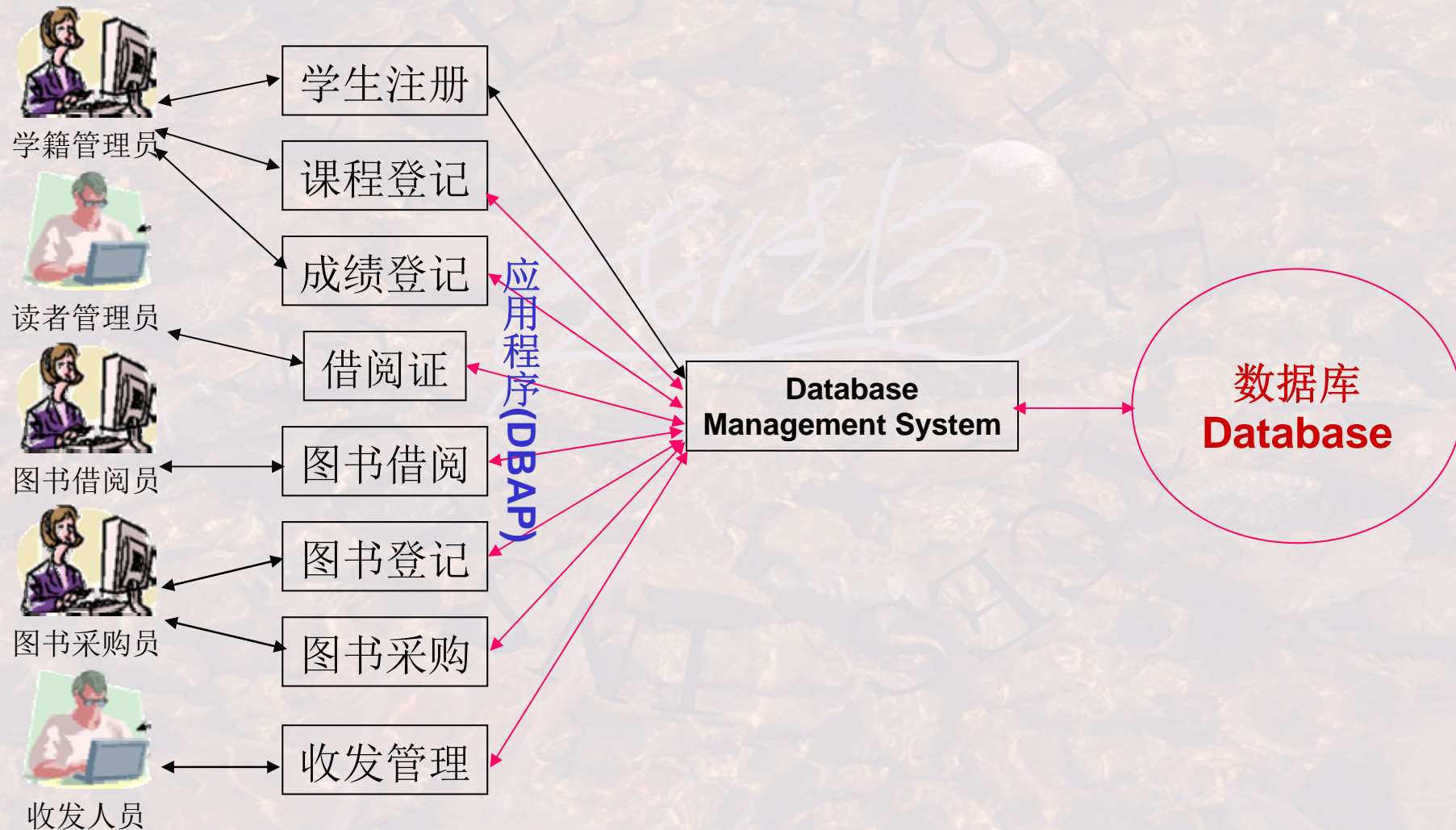


怎样抽象一个数据库系统？

数据库系统的标准结构

(1)数据库系统的分层抽象？

典型数据库系统的结构抽象



数据库系统的标准结构

(1)数据库系统的分层抽象？

DBMS管理数据的三个层次

External Level = User Level

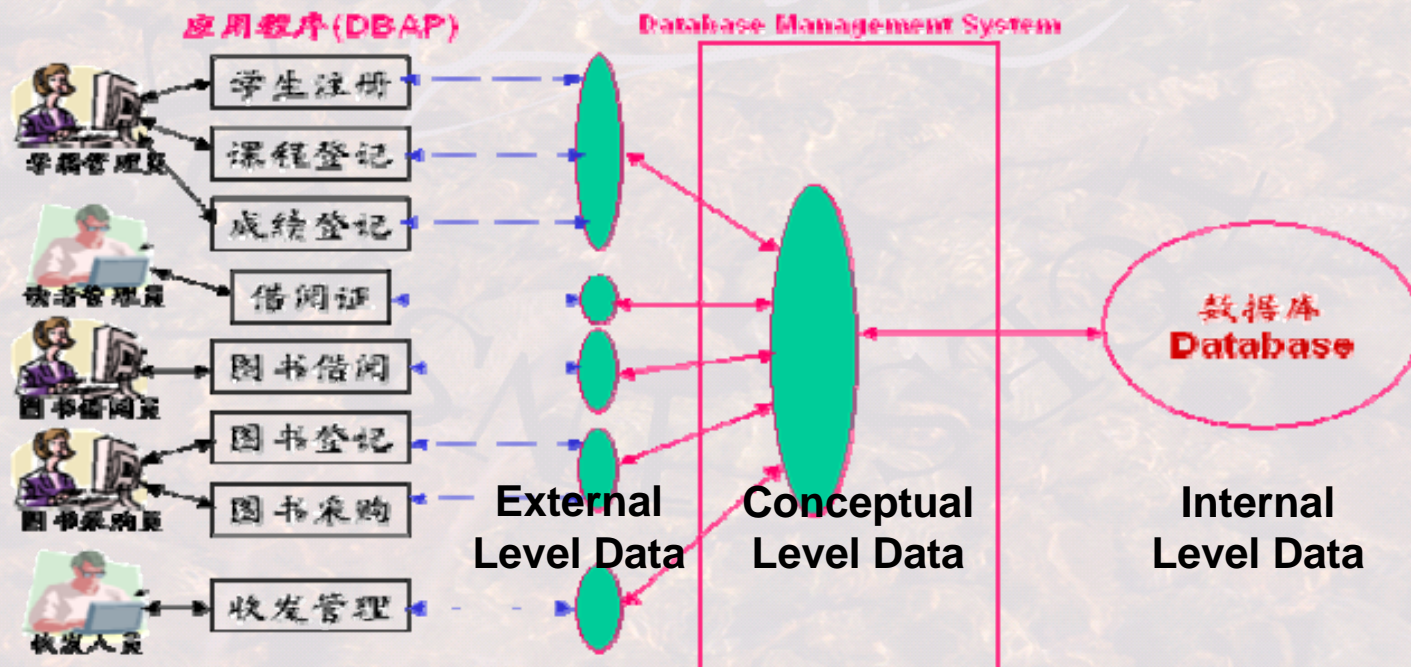
某一用户能够看到与处理的数据, 全局数据中的某一部分

Conceptual Level = Logic level

从全局角度理解/管理的数据, 含相应的关联约束

Internal Level = Physical level

存储在介质上的数据, 含存储路径、存储方式、索引方式等



数据库系统的标准结构

(2)数据(视图)与模式?



数据 与 数据的结构--模式

•模式(Schema)

对数据库中数据所进行的一种结构性的描述
所观察到数据的结构信息

•视图(View)/数据(Data)

某一种表现形式下表现出来的数据库中的数据

学生登记表(学号 **char(8)**, 姓名 **char(10)**,
性别 **Char(2)**, 出生年月 **datetime**, 入学日
期 **Datetime**, 家庭住址 **Char(40)**)

学生登记表

学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市

数据的结构----模式

展现的数据----视图

数据库系统的标准结构

(3)三级模式两层映像?

三级模式(三级视图)

External Schema ---- **(External) View**

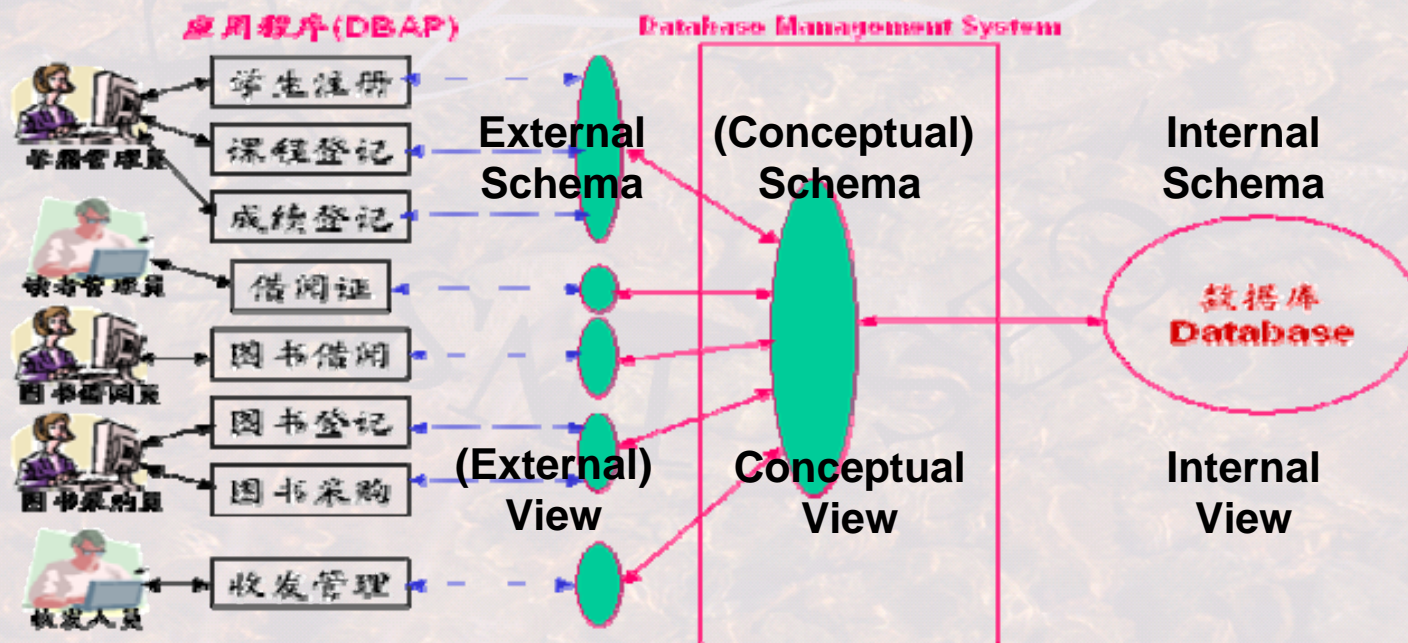
某一用户能够看到与处理的数据的结构描述

(Conceptual) Schema ---- **Conceptual View**

从全局角度理解/管理的数据的结构描述, 含相应的关联约束
体现在数据之间的内在本质联系

Internal Schema ---- **Internal View**

存储在介质上的数据的结构描述, 含存储路径、存储方式、索引方式等



数据库系统的标准结构

(3)三级模式两层映像?

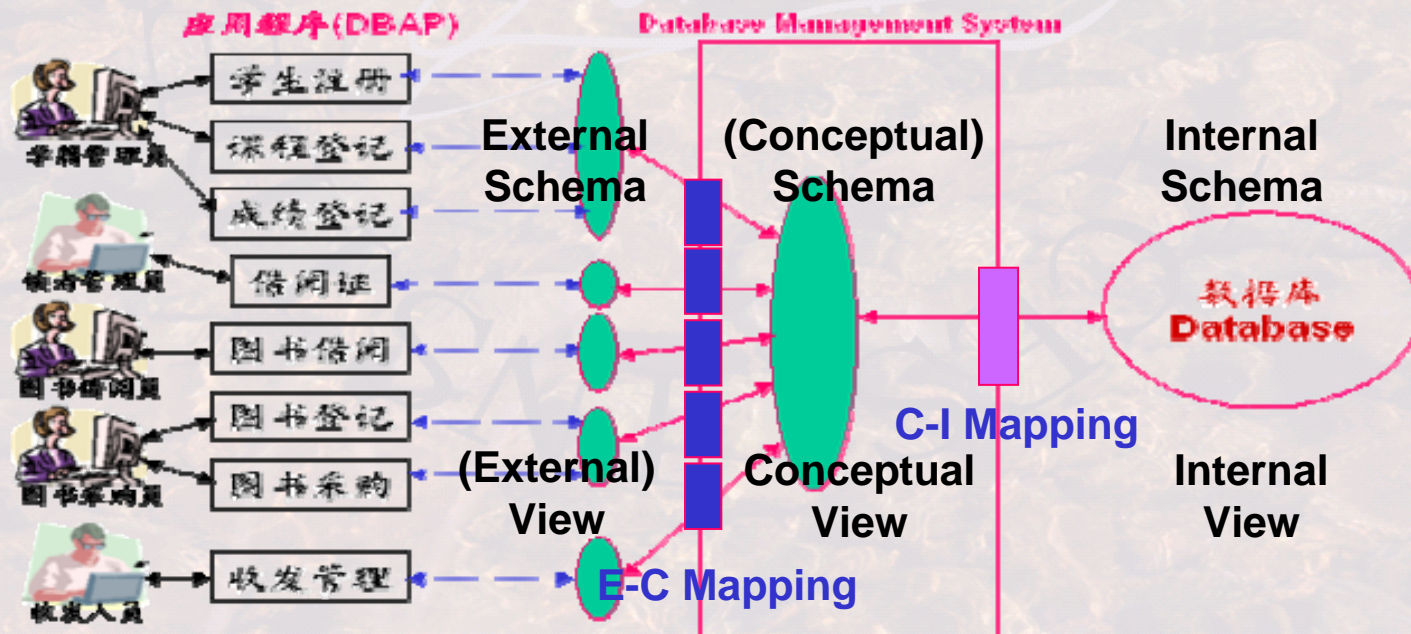
两层映像

E-C Mapping : External Schema-Conceptual Schema Mapping

- 将外模式映射为概念模式，从而支持实现数据概念视图向外部视图的转换
- 便于用户观察和使用

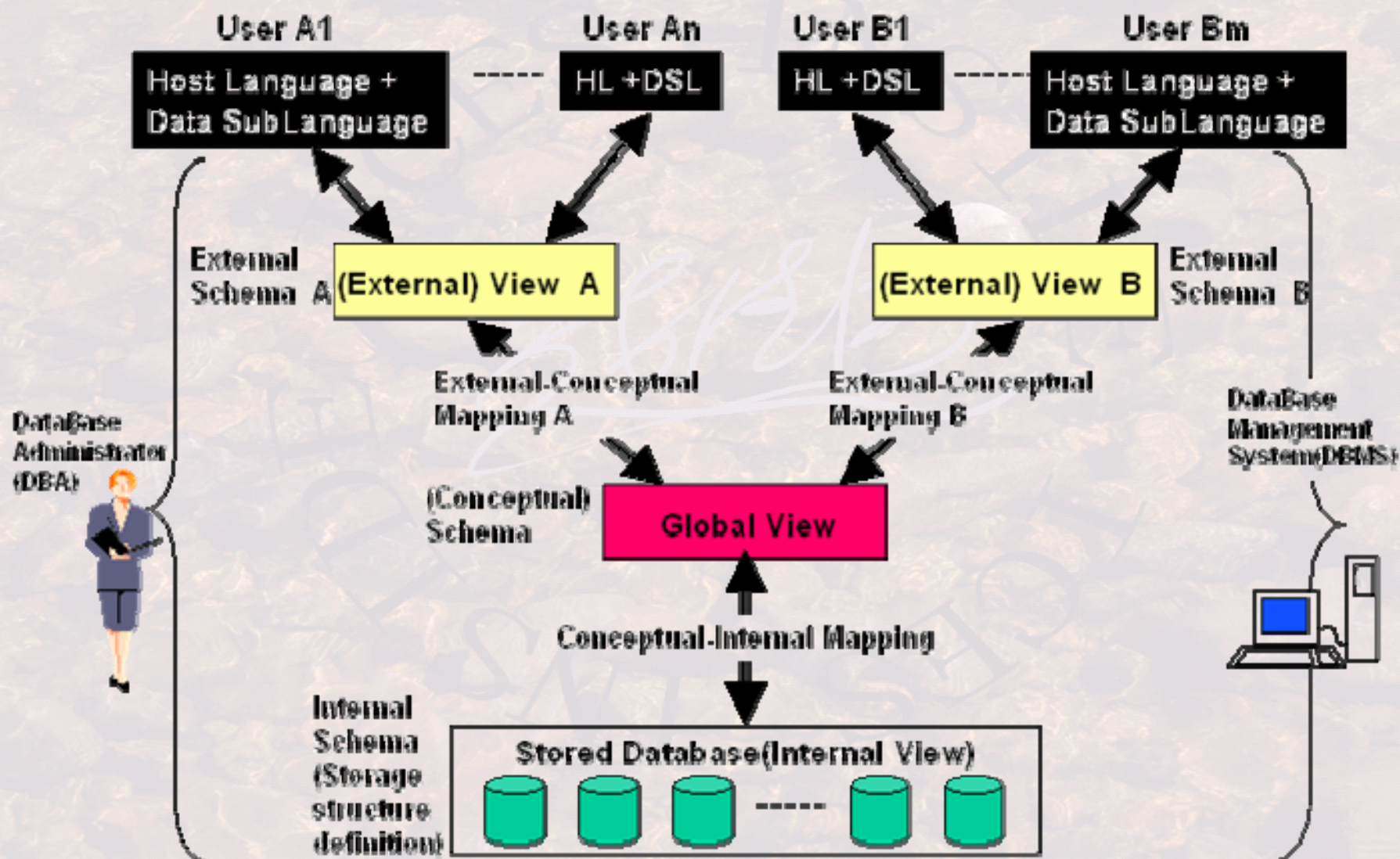
C-I Mapping : Conceptual Schema-Internal Schema Mapping

- 将概念模式映射为内模式，从而支持实现数据概念视图向内部视图的转换
- 便于计算机进行存储和处理



数据库系统的标准结构

(4)数据库系统的标准结构?



为什么要按照标准结构进行
数据库系统的抽象？

数据库系统的标准结构

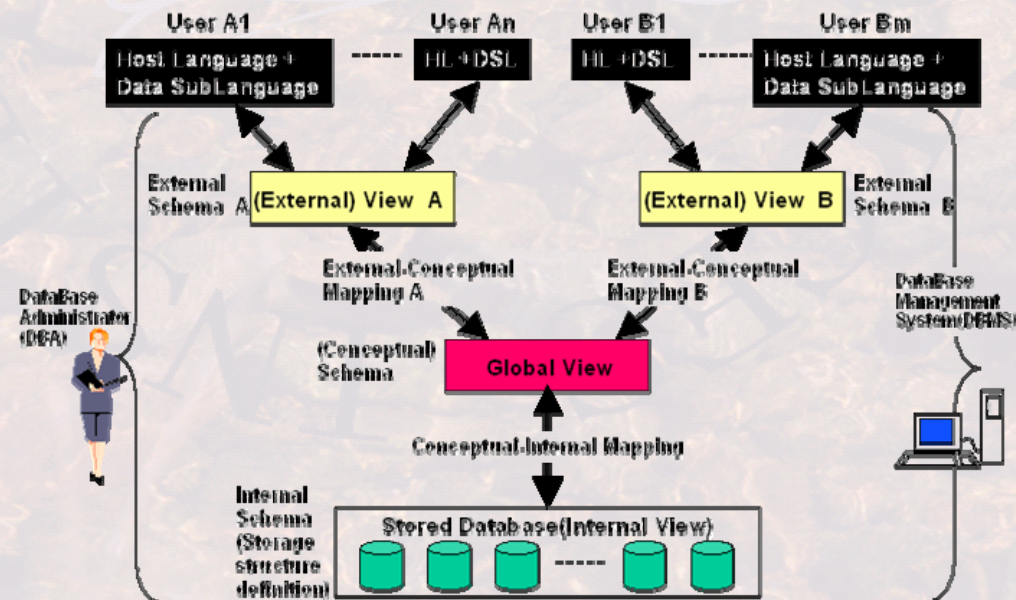
(5)两个独立性?

逻辑数据独立性

当概念模式变化时，可以不改变外部模式(只需改变E-C Mapping)，从而无需改变应用程序

物理数据独立性

当内部模式变化时，可以不改变概念模式(只需改变C-I Mapping)，从而不改变外部模式



数据模型

战德臣

哈尔滨工业大学 教授.博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on **I**ntelligent
Computing for **E**nterprises & **S**ervices,
Harbin **I**nstitute of **T**echnology

数据模型vs.模式vs.数据？

模式 与 模式的结构：数据模型

•数据模型

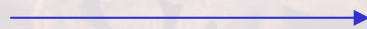
- 规定模式统一描述方式的模型，包括：数据结构、操作和约束
- 数据模型是对模式本身结构的抽象，模式是对数据本身结构形式的抽象

比如：

- 关系模型：所有模式都可为抽象表(Table)的形式[数据结构]，而每一个具体的模式都是拥有不同列名的具体的表。对这种表形式的数据有哪些[操作]和[约束]

关系模型：

Table/Relation
Relation Calculation
Constraints about Table



模式1：

学生登记表(学号 char(8), 姓名 char(10), 性别 Char(2), 出生年月 datetime, 入学日期 Datetime, 家庭住址 Char(40))

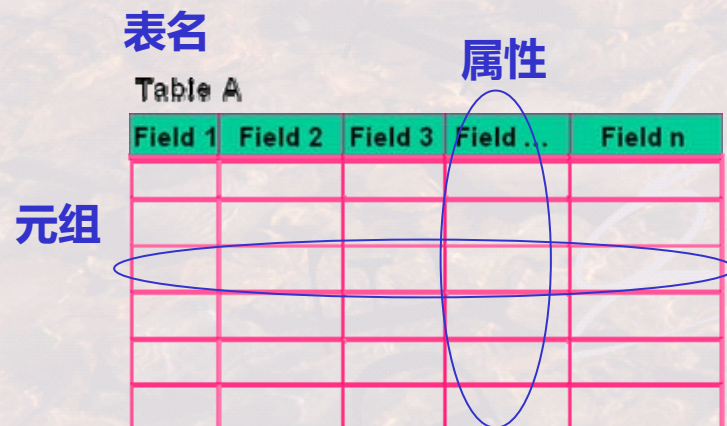
模式2：

学生成绩单(学号 char(8), 姓名 char(10), 班级 Char(6), 课程 char(40), 学期 Char(4), 成绩 Number)

三大经典数据模型

- 关系模型：**表**的形式组织数据
- 层次模型：**树**的形式组织数据
- 网状模型：**图**的形式组织数据

关系模型的例子



**关系模型
(表的形式)**

学生登记表(学号 char(8), 姓名 char(10), 性别 Char(2), 出生年月 datetime, 入学日期 Datetime, 家庭住址 Char(40))

关系模型的模式

学生登记表

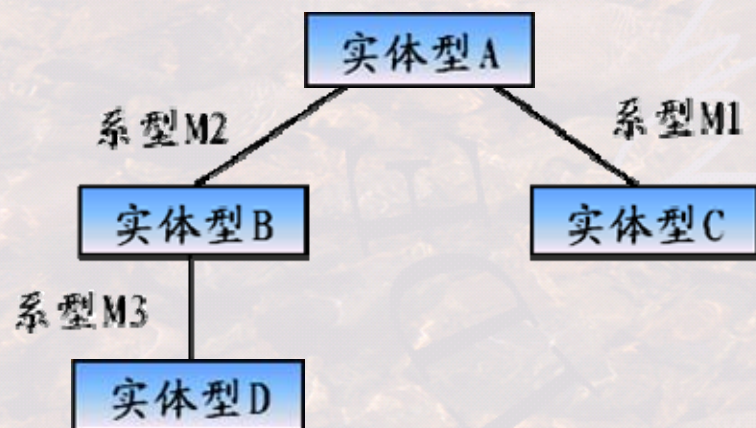
学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市

关系模型的数据

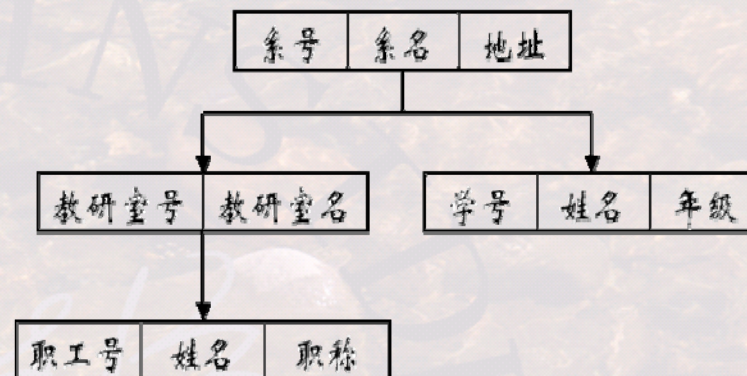
数据模型

(4)层次模型?

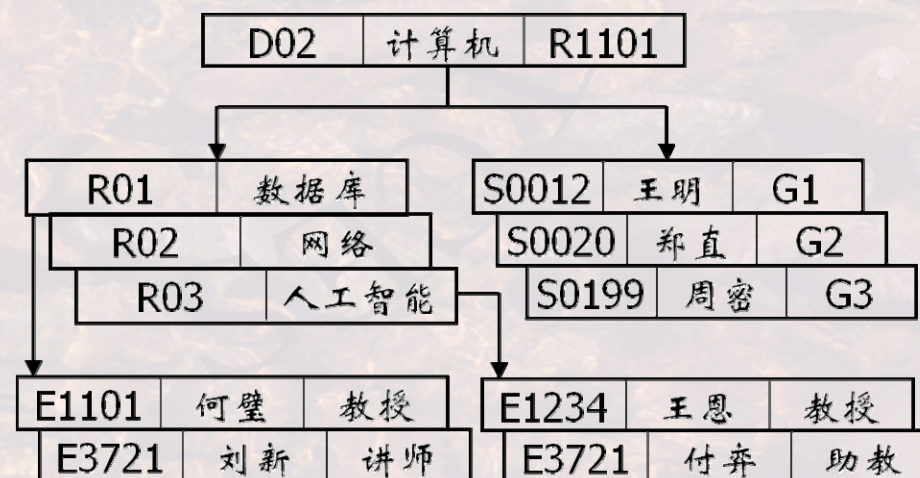
层次模型的例子



层次模型
(树的形式)



层次模型的模式

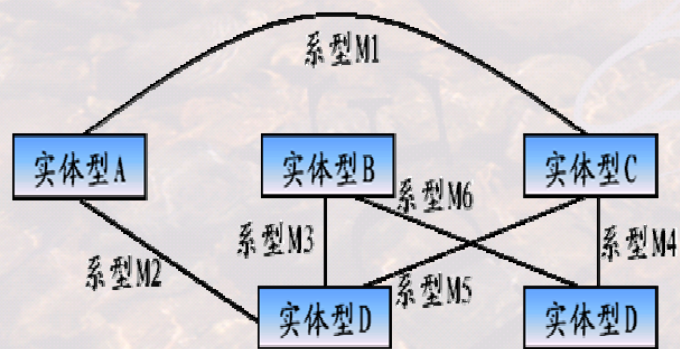


层次模型的数据

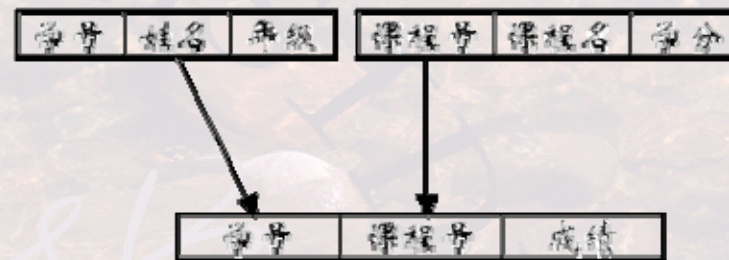
数据模型

(5)网状模型?

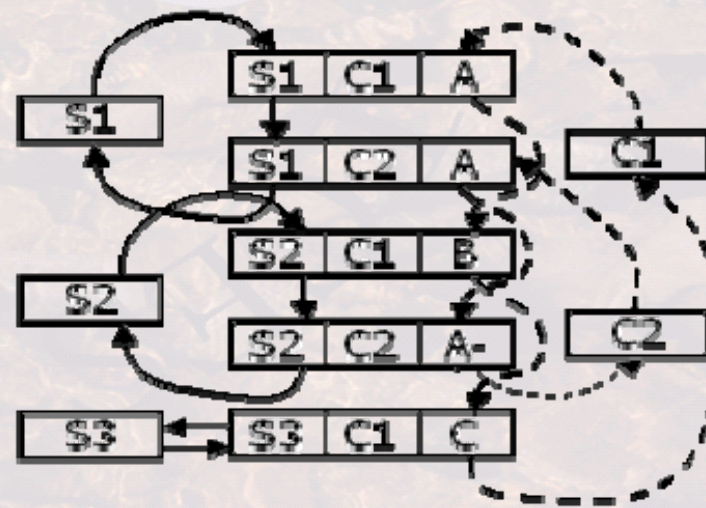
网状模型的例子



网状模型
(图的形式)



网状模型的模式



网状模型的数据

数据库系统的演变与发展

战德臣

哈尔滨工业大学 教授.博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on **I**ntelligent
Computing for **E**nterprises & **S**ervices,
Harbin **I**nstitute of **T**echnology

---- 简要发展史：四个阶段

- 第一阶段：数据库技术探索阶段(59-65/67)
 - ❑ 研制成功格式文件系统
 - ❑ 正式提出“Data Base”，并开始进行研究
- 第二阶段：数据库技术确立阶段(65/68-75)
 - ❑ 三大数据库：层次、网状及关系数据库相继提出并进行了深入研究
 - ❑ 商用数据库出现并应用，但多为网状及层次型系统
 - ❑ 数据库研究形成理论基础：关系数据库理论

---- 简要发展史：四个阶段(续)

➤ 第三阶段：数据库技术成熟阶段(76-80s前期)

- 提出了标准化数据库系统结构模型
- 关系DB系统迅速发展：如SQL, QBE, System R, Ingres等
- 关系理论日臻完善，包括规范化理论，关系语言，RDB的设计与实现，新型关系模型等；
- 数据库应用已十分普及，渗透到社会各个方面，出现众多DB的技术分支，DB走向全面成熟，人称70年代为“数据库的年代”

➤ 第四阶段：数据库技术深化发展阶段(85年以来)

- 数据库方法逐步理论化、数据库设计理论不断完善
- 新型数据模型、专用数据模型, 专用型、新型数据库系统，不断涌现
- 数据库技术+其他计算机技术结合 == 面向各行各业的专用数据库

数据库系统的演变与发展

(2)由文件系统到数据库

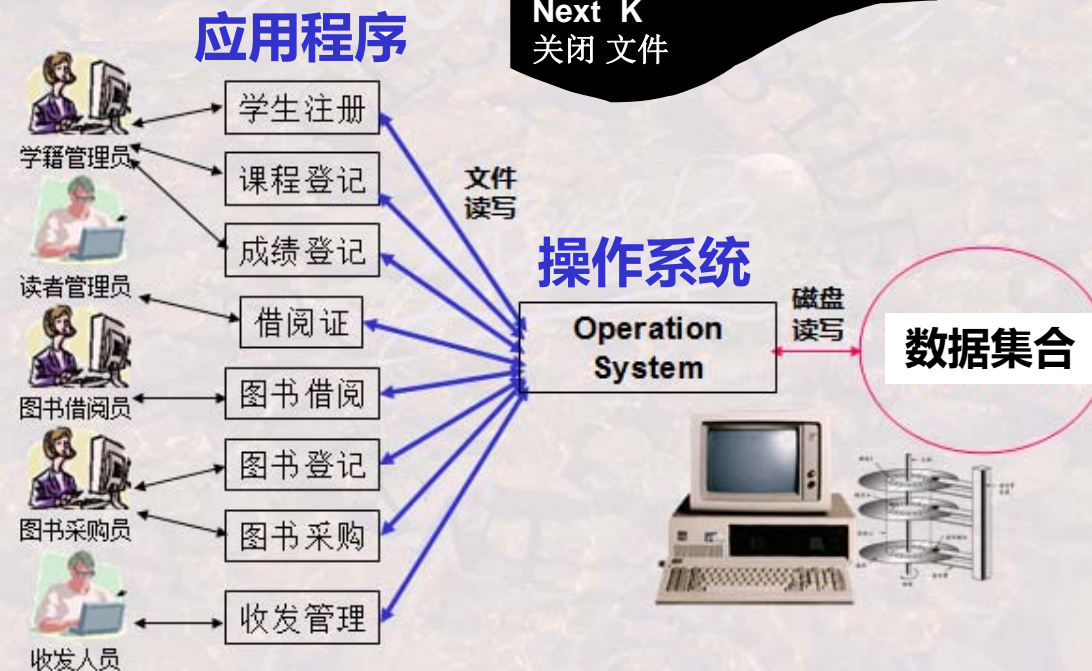
文件系统的示例

学生登记表

学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市

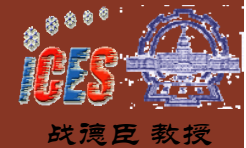
定义“学生登记表”文件的数据结构;

```
打开"学生登记表"文件
For K=1 to 最后一条记录
  读第K条记录
  识别出 性别 属性的值
  If 性别 = '男' then
    显示第K条记录
  End if
Next K
关闭 文件
```



数据库系统的演变与发展

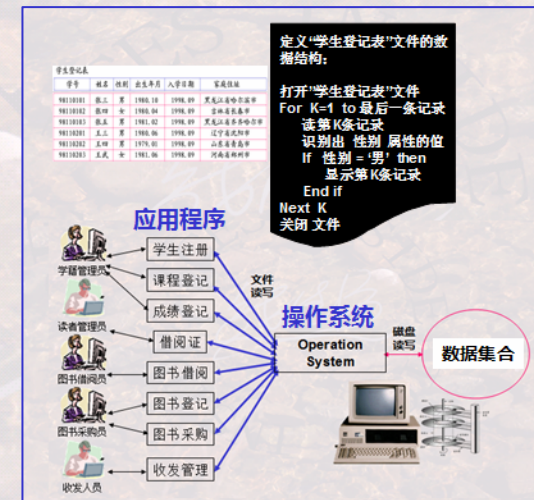
(2)由文件系统到数据库



重要发展：由文件系统到数据库

❖ 文件系统

- 文件存储空间的管理、目录管理、文件读写管理、文件保护、向用户提供操作接口
- 提供了不同的存取方法(索引文件、链接文件、直接存取文件、倒排文件等), 支持对文件的基本操作(增、删、改、查等), 数据存取基本上以**记录**为单位



- **(优点)** 用户(程序)不必考虑文件存储的物理细节, 解脱了对物理设备存取复杂性处理的负担
- **(不足)** 数据与程序紧密结合, 数据的组织及语义紧密依赖于处理该文件的应用程序, 数据结构发生改变则必须修改应用程序, 文件之间无联系, 文件的记录之间无联系, 共享性差, 冗余度大, 不一致性高

数据库系统的演变与发展

(2)由文件系统到数据库

数据库系统的示例

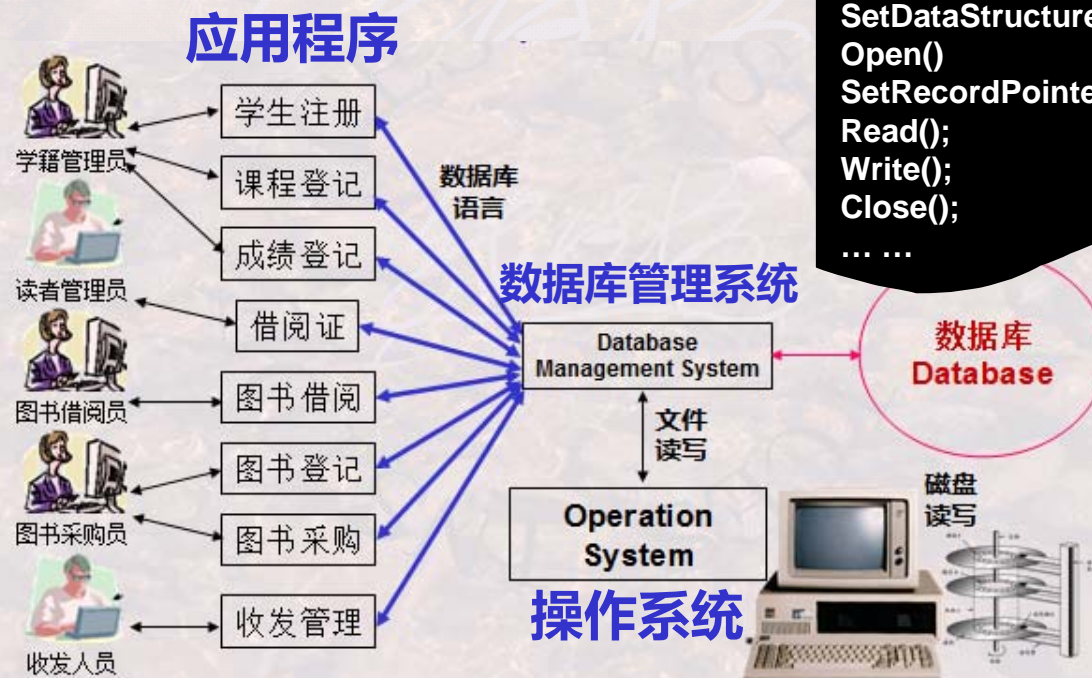
学生登记表

学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市

读取哪一个表的
哪一行哪一列

定义表的格式

读取数据的细节
例如：
SetDataStructure()
Open()
SetRecordPointer()
Read();
Write();
Close();
... ..



数据库系统的演变与发展

(2)由文件系统到数据库

重要发展：由文件系统到数据库

❖ 数据库系统

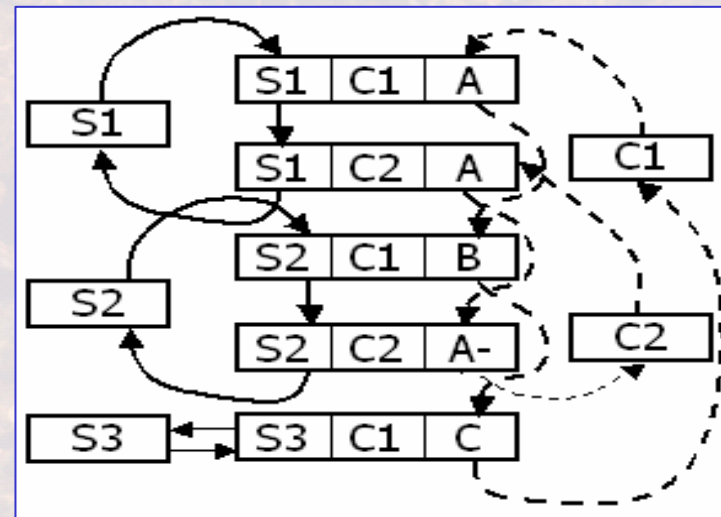
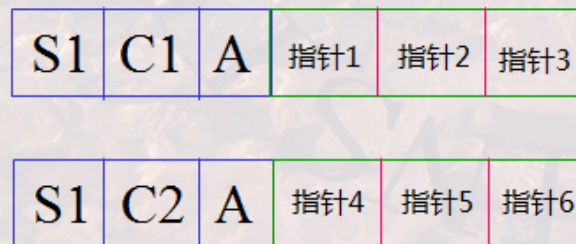
- 由**DBMS**统一存取、维护数据组织形式及语义，可较强地独立于应用程序(数据的物理独立性和逻辑独立性)
- 把数据及数据结构的定义和描述从应用程序中分离出去，交给**DBMS**，使得多个应用程序可共享数据及数据结构的操作，数据存取可以**记录**为单位，也可以以**数据项**和**记录集合**为单位
- 统一的数据控制功能，数据共享程度高：系统可自动检查安全性、完整性和并发正确性
- 整体数据结构化，文件(**Table**)之间、记录之间相互有关联，数据的冗余度小，易扩充
- 独立于应用程序的高效率查询/统计操作



重要发展：由层次模型数据库、网状模型数据库到关系数据库

❖ 层次模型与网状模型数据库

- 数据之间的关联关系由复杂的指针系统来维系，结构描述复杂
- 数据检索操作依赖于由指针系统指示的路径
- 逐一记录的操作，不能有效支持记录集合的操作



重要发展：由层次模型数据库、网状模型数据库到关系数据库

❖ 关系模型数据库

➤数据之间的关联关系由**Table**中属性的值来表征，结构描述简单：

Table/relation

➤数据检索操作**不依赖于**路径信息或过程信息，支持非过程化的数据操作

➤有效支持记录集合的操作

➤较为完善的理论基础

学生登记表

学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市

数据库系统的演变与发展

(3)由层次模型数据库、网状模型数据库到关系数据库

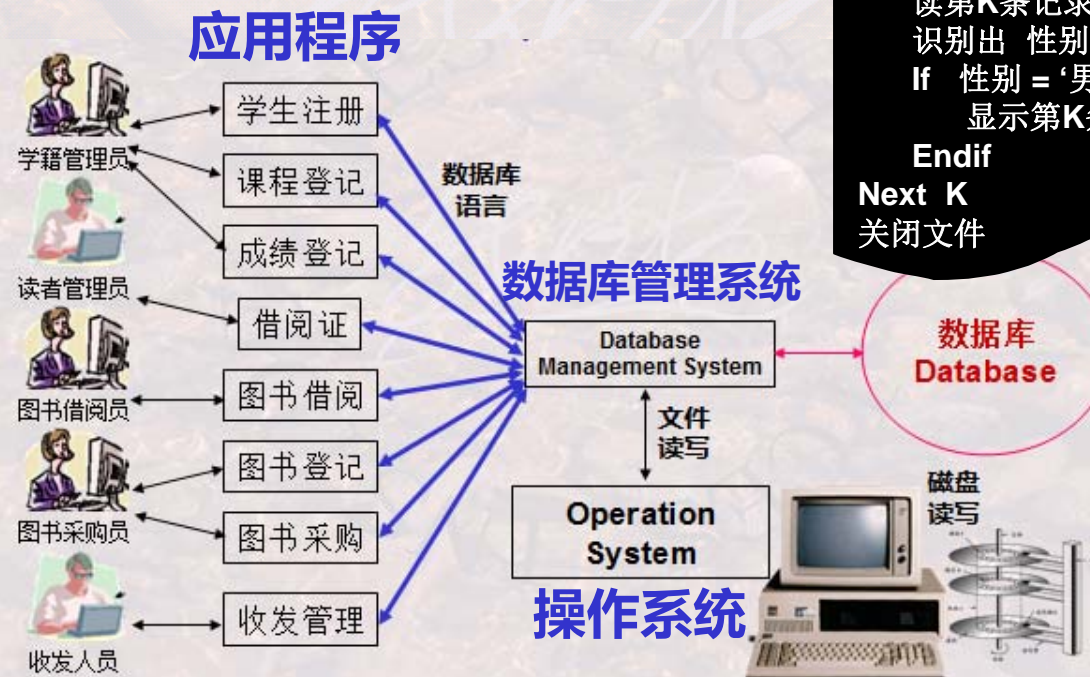
关系数据库系统的示例

学生登记表

学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市

```
Select 学号, 姓名  
From 学生登记表  
Where 性别 = '男';
```

```
打开文件  
For K=1 to 最后一条记录  
  读第K条记录  
  识别出 性别 属性的值  
  If 性别 = '男' then  
    显示第K条记录  
  Endif  
Next K  
关闭文件
```



重要发展：由关系数据库到对象关系数据库、面向对象数据库

❖ 关系数据库

- 按行按列形式组织数据：关系的第1范式
- 数据项的不可再分特性
- 关系运算：关系代数、元组演算、域演算--→标准SQL
- 关系数据库设计理论

学生登记表

学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市

重要发展：由关系数据库到对象关系数据库、面向对象数据库

❖ 对象-关系数据库

- 可有效支持不满足关系第1范式的数据项
- 以对象来封装需分解的数据项
- 行对象与列对象；聚集对象与结构对象

列对象

Head: structured type

Students						
sid	name		class	telephone	enrollment	
	lname	fname			cno	major
1	Jones	Allan	2	555-1234	101	No
					108	Yes
2	Smith	John	3	555-4321	105	No
3	Borwn	Harry	2	555-1122	101	Yes
					108	No
4	White	Edward	3	555-3344	102	No
					105	No

行对象

结构对象
聚集对象

Value: structured value
collection of values

重要发展：由关系数据库到对象关系数据库、面向对象数据库

❖ 面向对象数据库

- 面向对象技术(O-O)与集合/聚集操作技术(SQL)的结合
- 支持复杂的数据类型，数据封装与抽象数据结构
- 支持面向对象的一些特性：类、继承、封装、多态...

列对象

Head: structured type

行对象

Students						
sid	name		class	telephone	enrollment	
	lname	fname			cno	major
1	Jones	Allan	2	555-1234	101	No
					108	Yes
2	Smith	John	3	555-4321	105	No
3	Borwn	Harry	2	555-1122	101	Yes
					108	No
4	White	Edward	3	555-3344	102	No
					105	No

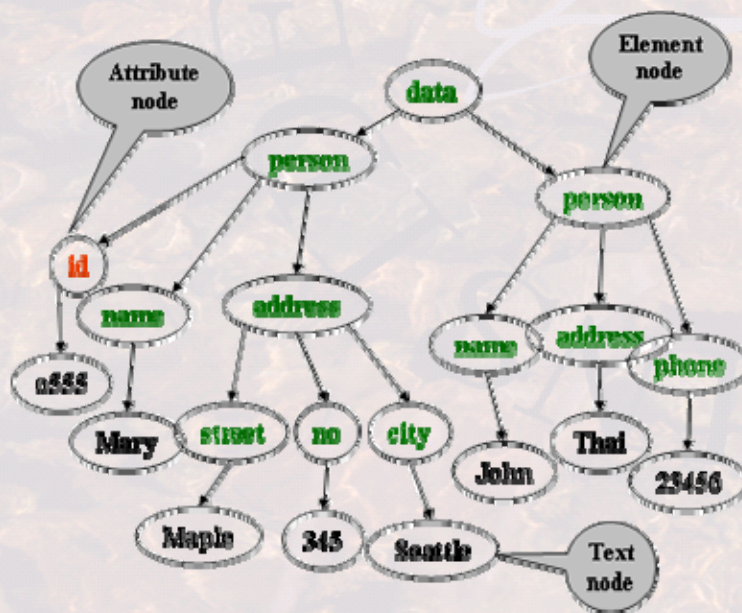
结构对象
聚集对象

Value: structured value
collection of values

重要发展：由关系数据库到对象关系数据库、面向对象数据库

❖ XML数据库

- 是数据库的另一种形式, 被称为半结构化数据库;
- 数据** 与 **数据的语义** 合并在一起进行存储和处理;
- 面向**数据交换**而提出, 在**互联网世界**得到广泛应用.

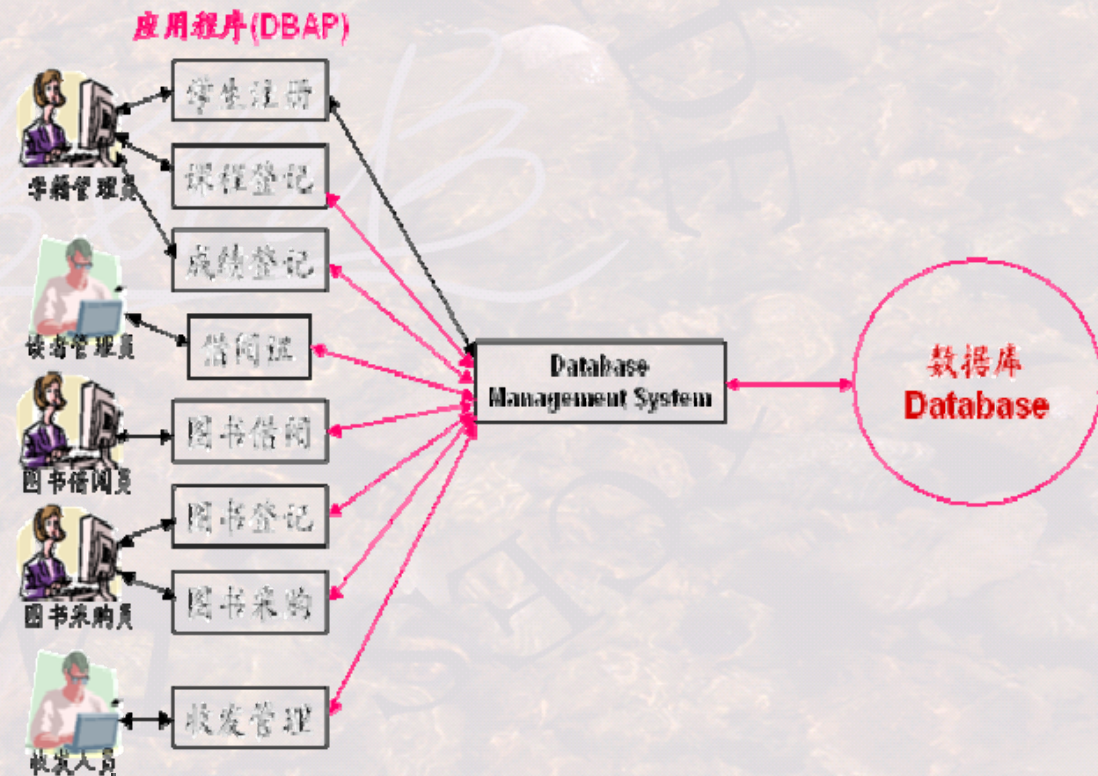


```
<data>
  <person id="o555" >
    <name> Mary </name>
    <address>
      <street>Maple</street>
      <no> 345 </no>
      <city> Seattle </city>
    </address>
  </person>
  <person>
    <name> John </name>
    <address>Thailand
    </address>
    <phone>23456</phone>
  </person>
</data>
```


重要发展：由多种多样的数据库到多数据库开放式互连

❖ 多种多样的数据库

- Oracle
- Sybase
- Ingres
- DB 2
- MS Access
- Informix
-

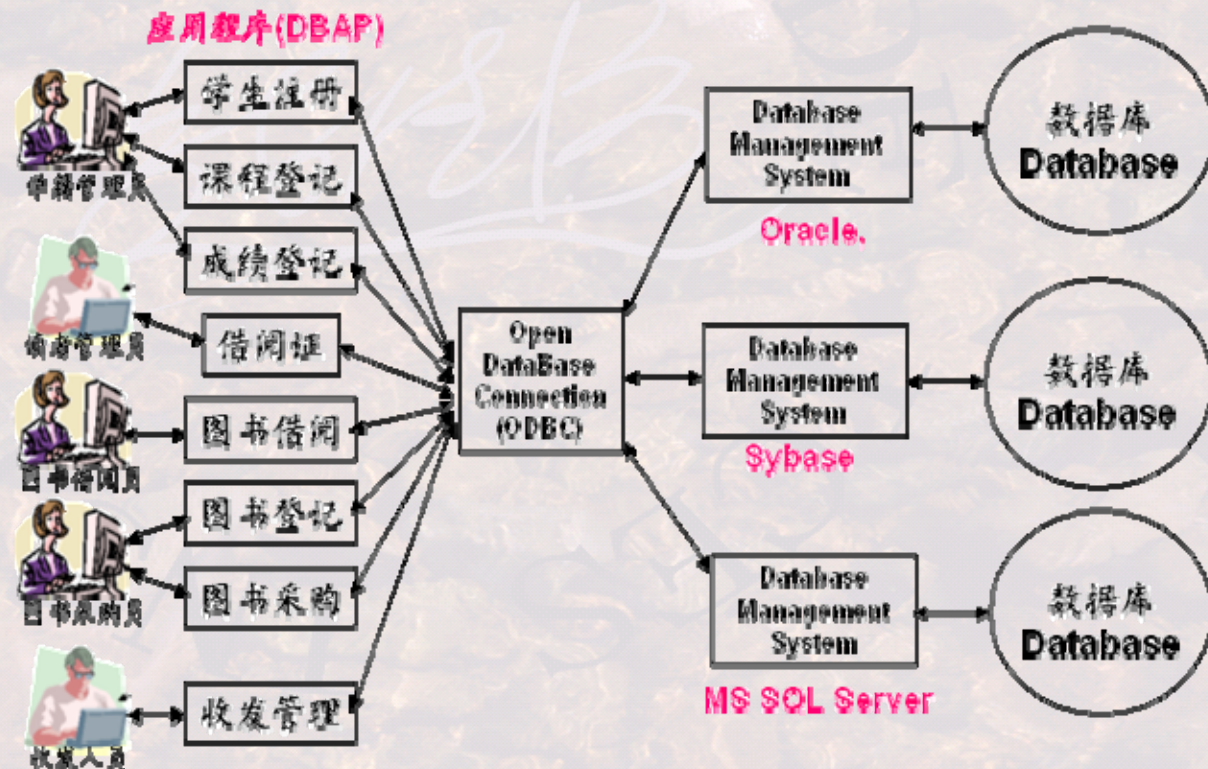


重要发展：由多种多样的数据库到多数据库开放式互连

❖ 开放互连多种多样的数据库

➢ ODBC

➢ JDBC



重要发展：由普通数据库到与各种先进技术结合所形成的新型数据库

❖ 新型数据库

- **OA**: DB + Management Information System
- **Database Machine** \Leftarrow DB + Computer Architecture
- **Intelligent Database** \Leftarrow DB + Artificial Intelligence
- **Distributed Database (DDB)** \Leftarrow DB + Computer Network。
- **Image Database / Multimedia Database** \Leftarrow DB + Image processing / Multimedia processing。
- **Temporal Database** \Leftarrow DB + 时态技术处理。
- **Mobile Database** \Leftarrow DB + 移动计算技术。
- **Active Database** \Leftarrow DB + 产生式规则/触发器技术。

重要发展：由普通数据库到与各种先进技术结合所形成的新型数据库

❖ 新型数据库

- **Fuzzy Database** \Leftarrow DB + 模糊处理技术。
- **Real-Time Database** \Leftarrow DB + 实时处理技术。
- **Engineering Database** \Leftarrow DB + CAD/CAPP/CAM技术。
- **Geographical Databas**和空间数据库(**Spacial Database**) \Leftarrow DB + 数字地图、全球定位、空间分析技术。
- **Statistical Database** \Leftarrow DB + 统计学。
- **Internet Database** \Leftarrow DB + Internet/WWW(网页/HTML文档)。
- **Data Warehouse/Data Mining** \Leftarrow DB + OLAP + 统计学。
- **NoSQL**

回顾本讲学了什么？

战德臣

哈尔滨工业大学 教授.博士生导师

黑龙江省教学名师

教育部大学计算机课程教学指导委员会委员

Research Center on **I**ntelligent
Computing for **E**nterprises & **S**ervices,
Harbin **I**nstitute of **T**echnology

回顾本讲学习了什么？

