

数据库系统原理

复杂关键软件环境全国重点实验室

郎 波

Tel: 82317656

E-Mail: langbo@buaa.edu.cn

数据库系统原理课程概述

- 讲什么——课程内容
- 怎么讲——教学安排
- 怎么学——学习建议

大数据时代的信息技术

- 大量信息化工具和系统，处理和产出大量数据
- 各领域海量数据的积累，使人们能够通过数据分析发现海量数据中的知识和规律
- 数据处理、数据分析都以数据存储和管理为基础，数据库就是数据管理的核心技术和软件

数据库技术的重要性

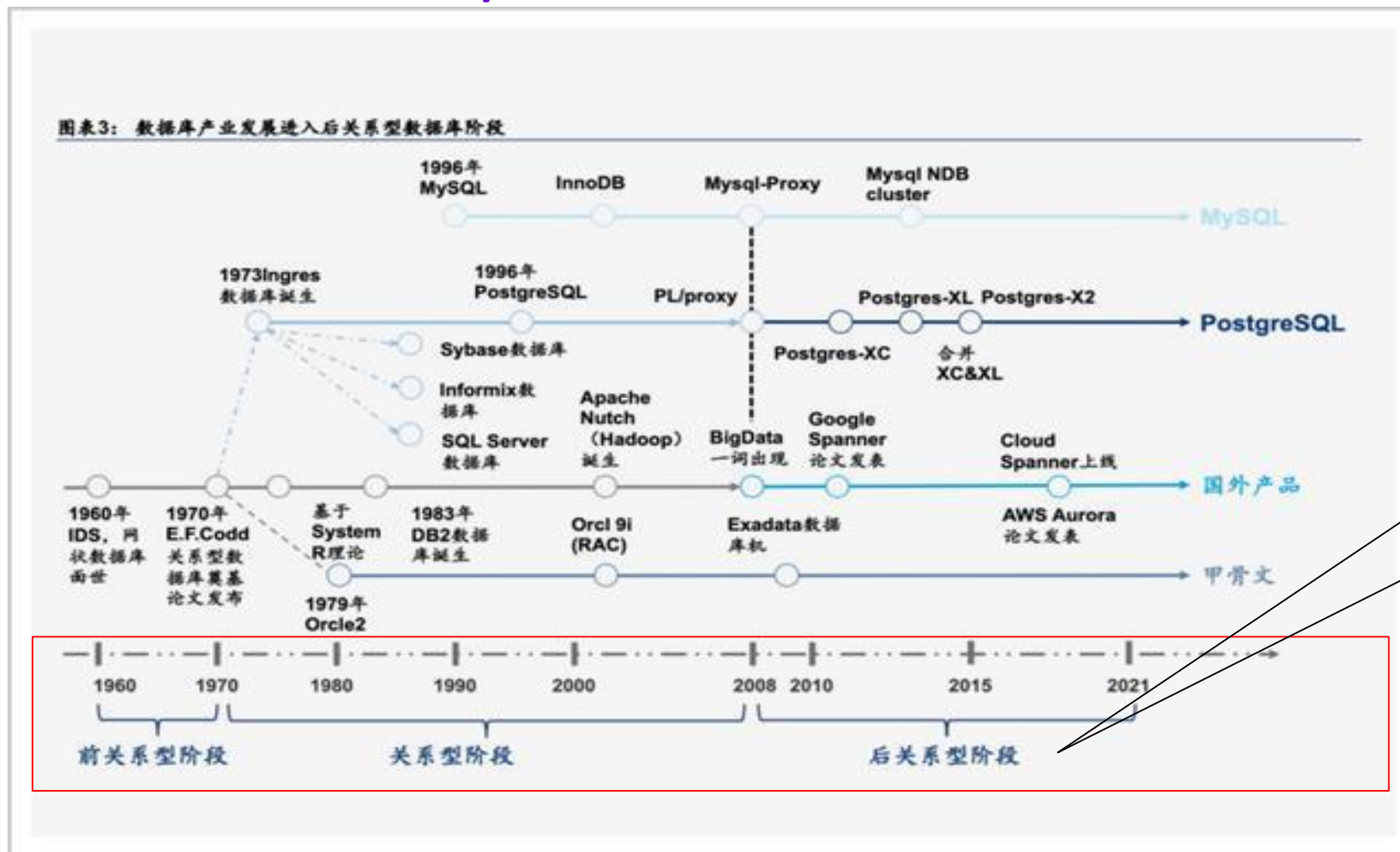
- 数据库被认为是数字经济的“根技术”，是关键基础软件



- 数据库是大国博弈的“杀手锏”
 - 2023年5月22日，华为完成全球88家子公司的MetaERP系统，成功实现大规模切换

数据库的发展历程

• 数据库的发展阶段



由关系型向非关系型、由集中式向分布式转型的数据库技术——NoSQL与NewSQL

国产数据库主要技术路线

- 从零开始自主研发，代表厂商为达梦等
- 基于PostgreSQL、MySQL等开源数据库二次开发，诸多大公司采用该技术路线
- 国产数据库发展趋势

— 国产化替代Oracle数据库已经迫在眉睫

— 云原生发展趋势：云原生数据库将存储和计算分离，部署于云基础设施之上，弹性可扩展是云原生数据库的本质

— 开源发展趋势：我国数据库逐步走向开源发展大趋势，商业产品市场推广的需要，也是掌握自主可控主导权的需要

图表30：国产数据库产业图谱



资料来源：《中国数据库行业研究报告》，艾瑞咨询（2021）、华泰研究

什么是数据库技术

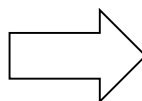
- 数据
 - 数据是描述现实世界各种事物的可以识别的符号
 - 有数字、文字、图形、图像、音视频等多种表现或存储形式
- 示例：
 - C1, 数据库, C5, C2, 数学, C3, 信息系统, C1, C4, 操作系统, C6, C5, 数据结构, C7, C6, 数据处理, C7, C语言, C6



数据与信息

- 数据只有被组织和加工形成信息，才具有使用价值

C1, 数据库, C5, C2, 数学, C3, 信息系统, C1, C4, 操作系统, C6, C5, 数据结构, C7, C6, 数据处理, C7, C语言, C6



课程编号	课程名称	先行课编号
C1	数据库	C5
C2	数学	
C3	信息系统	C1
C4	操作系统	C6
C5	数据结构	C7
C6	数据处理	
C7	C语言	C6

- 信息

- 信息是一种已经被加工为特定形式的数据，这些数据对现在与将来的决策有明显价值；
- 信息是各种数据所包括的意义；数据是信息的载体，是信息的具体表现形式。

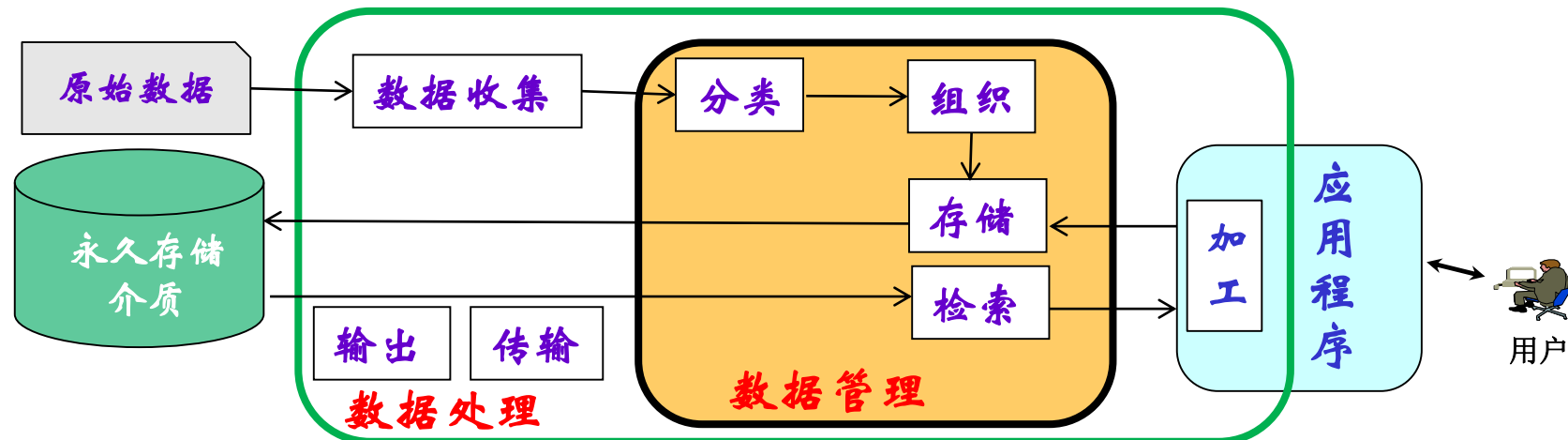
数据处理与数据管理

- 数据处理

- 从大量原始数据中抽取和推导出有价值信息的加工过程称为**数据处理**，包括过程：数据收集、组织、存储、加工、分类、检索、输出、传输等操作

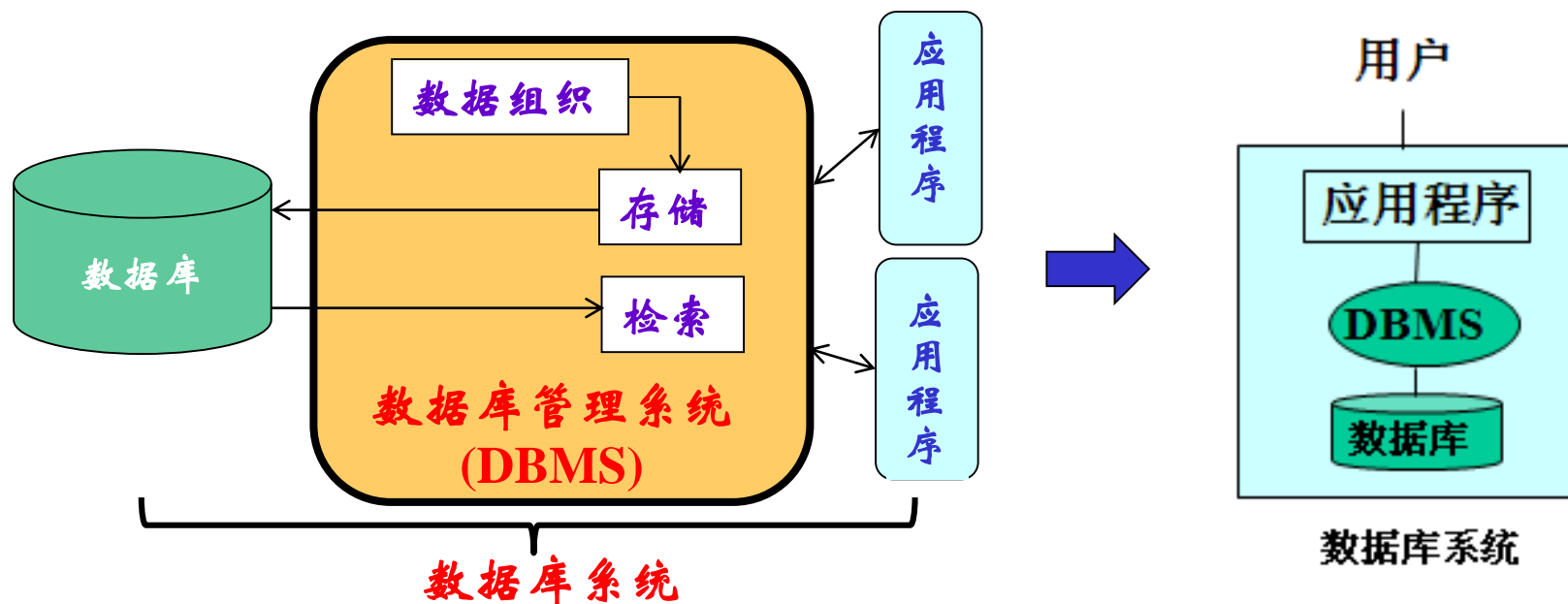
- 数据管理

- 数据处理一般性的基本操作，如数据分类、组织、存储、检索、维护等称为**数据管理**，并研究专门的技术——**数据管理技术**



数据库技术

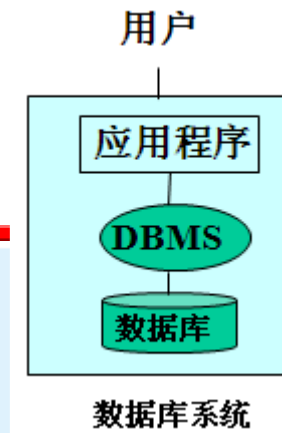
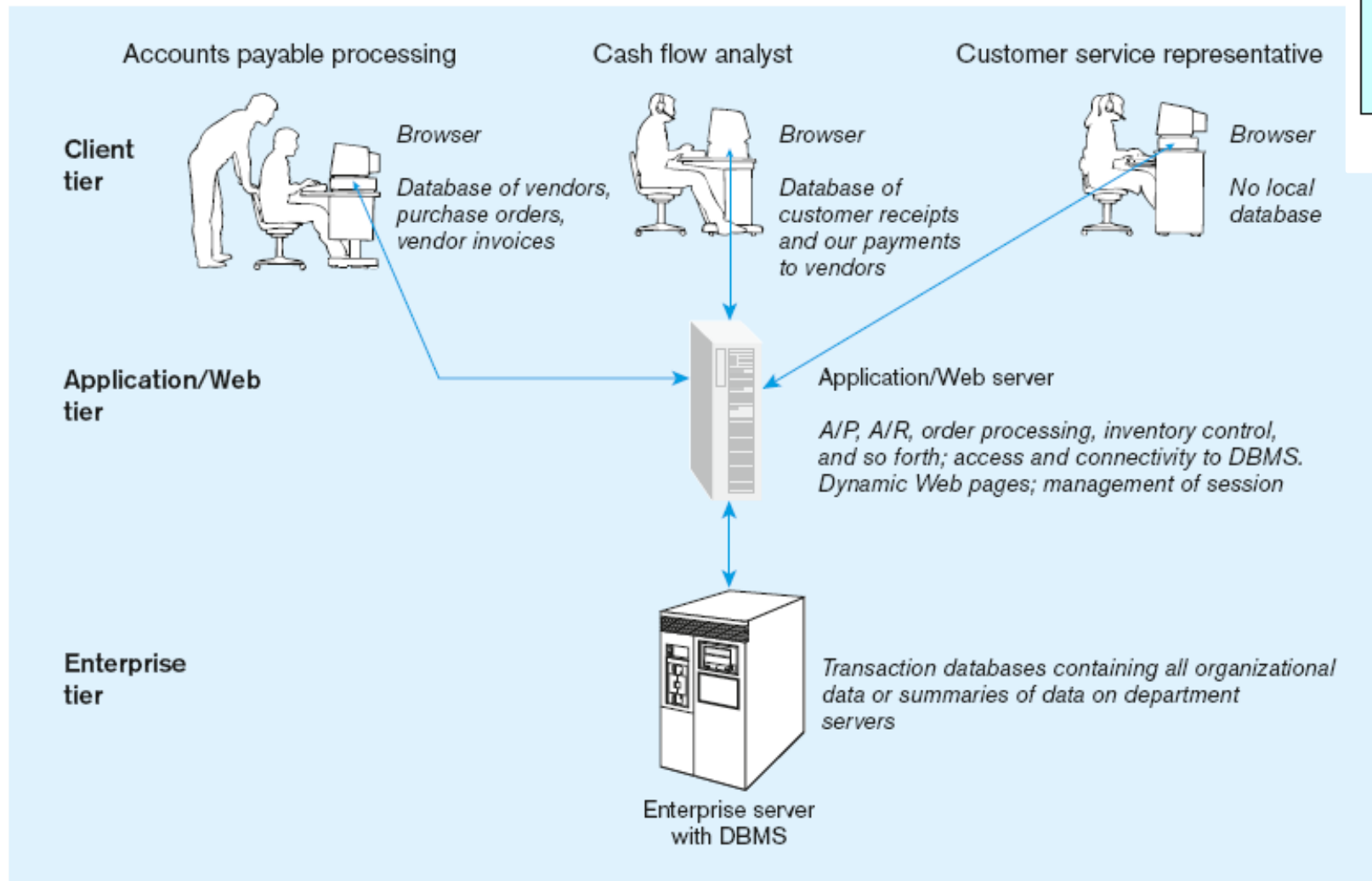
- 数据库技术
 - 是一种数据管理技术
 - 按照某种数据结构对数据进行组织后，存储在计算机的二级存储中，并可以提供数据共享操作
- 数据库技术的软件实现就是**数据库管理系统DBMS**
- **数据库系统**是基于数据库管理系统建立的具有特定数据处理功能的系统

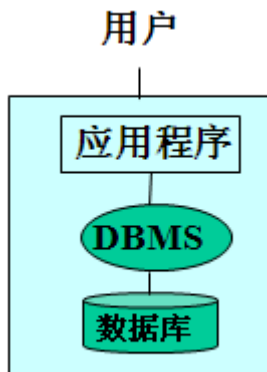


数据库系统示例



数据库系统示例(3)





数据库系统

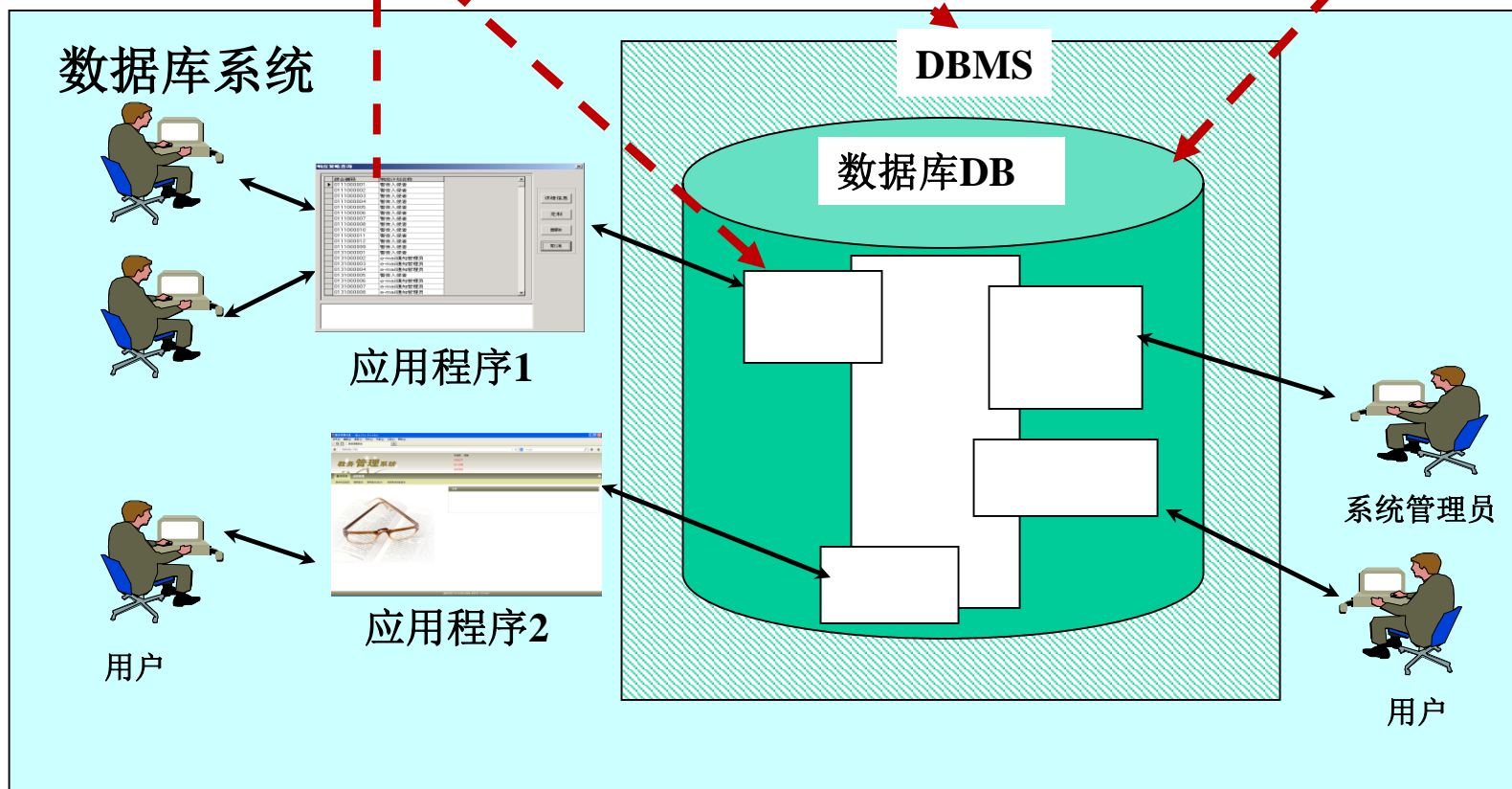
数据库系统原理

数据库系统原理

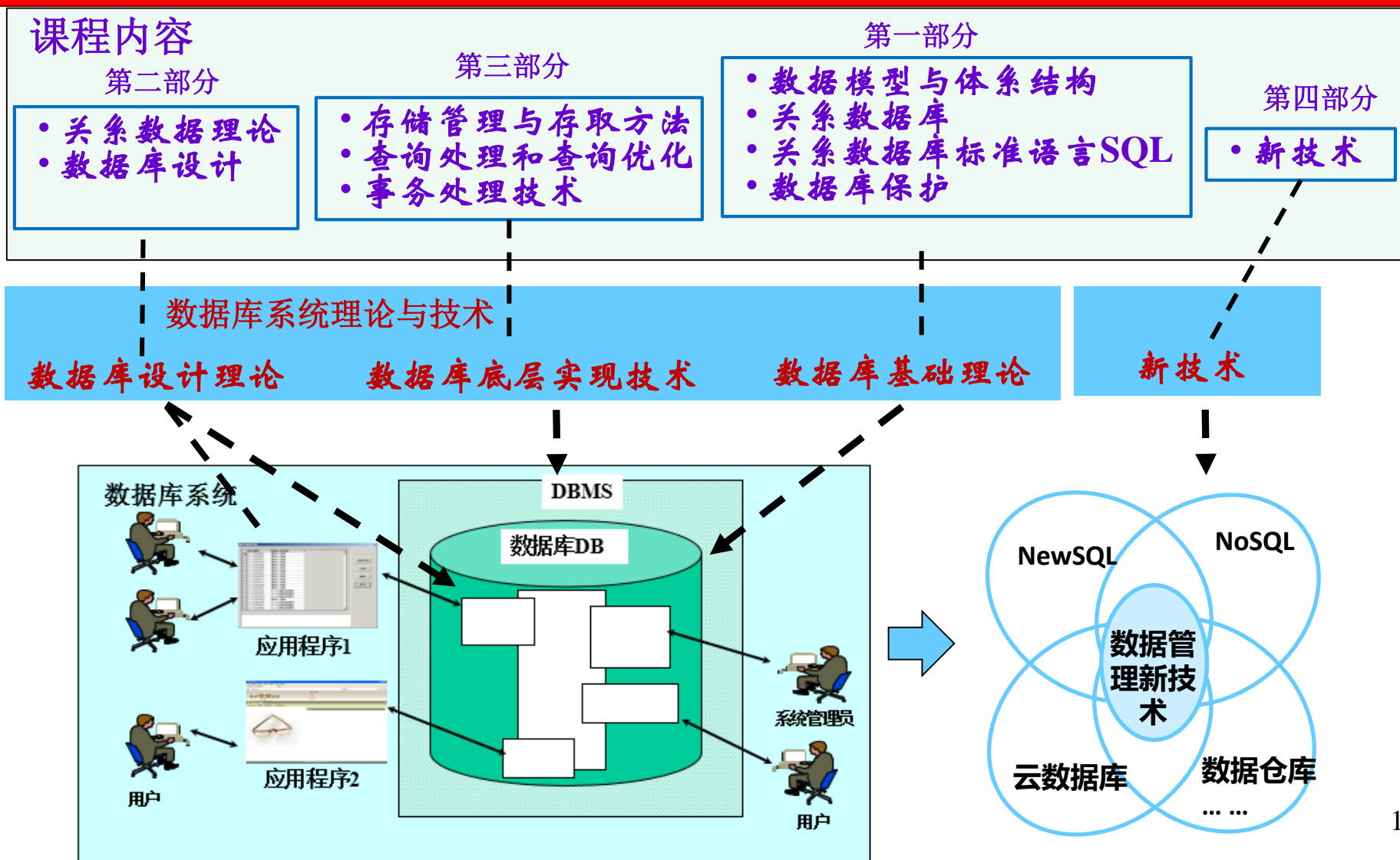
数据库设计理论

数据库底层实现技术

数据库基础理论



数据库系统原理课程内容



课程目标

- 课程目标：
 - 理解并掌握数据库系统的基本理论
 - 掌握并应用标准数据操纵语言与数据库系统的设计方法
 - 理解数据库管理系统的实现技术
 - 了解数据库管理技术的新发展

教学方式

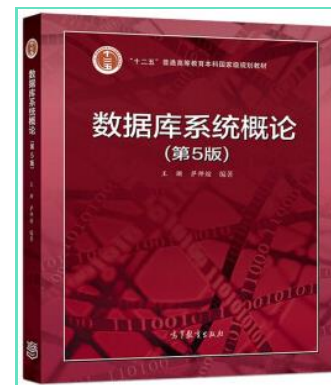
- 教学方式

- 课堂讲授、讨论、实验（课程设计，云数据库实验）、自学
- 结伴学习与独立学习相结合

- 教材

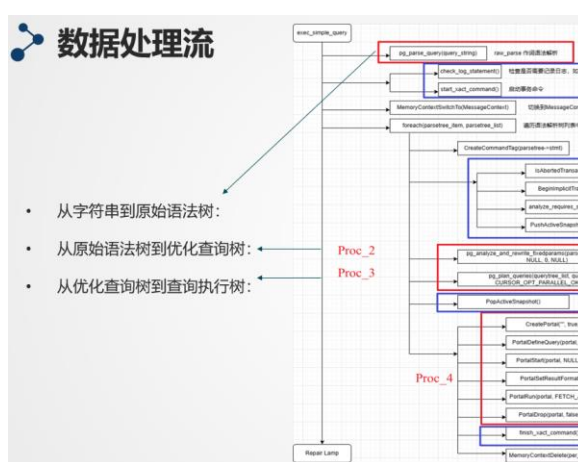
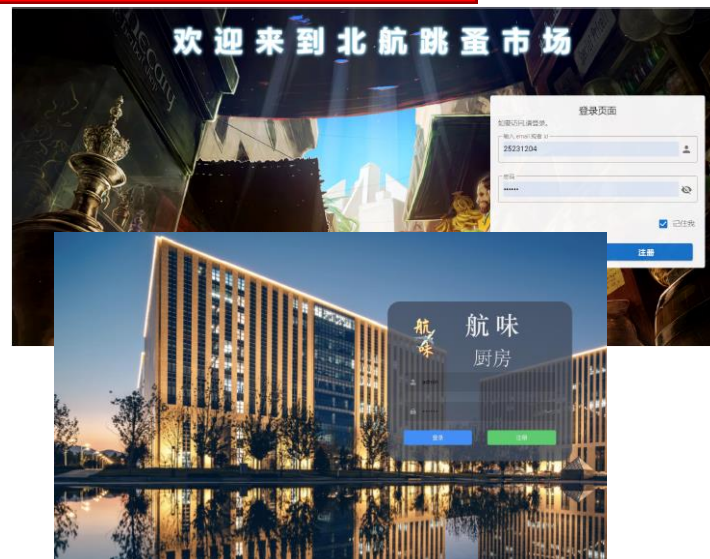
- 萨师煊,王珊编著, 数据库系统概论, 第六版/第五版/第四版/第三版, 高等教育出版社, 2023.3
- 杨冬青, 李红燕, 张金波等译, 数据库系统概念, 第七版, 机械工业出版社, 2021.6

(Abraham Silberschatz, Henry F. Korth, S. Sudarshan, Database System Concepts, 7th Edition, Mc Graw Hill)



课程大作业（课程设计）

- 32学时实验，与课程同步进行
- 题目类型与内容
 - 题目1：应用系统设计开发：设计与开发中等规模数据库应用系统
 - 分组进行，三人一组
 - 题目2：走进PostgreSQL—PG剖析、展示与调试
 - 分组进行，三-四人一组
 - 本轮课选两-三组同学做题目2，其余同学做题目1

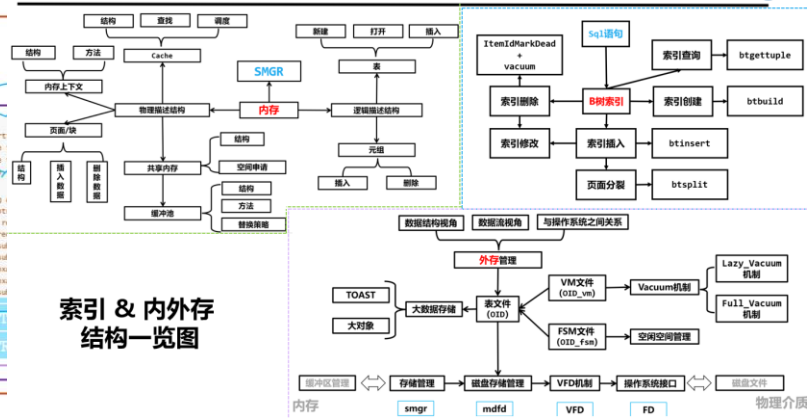


知识图谱-事务管理器的数据结构

● 事务块状态VS事务状态



概览图



索引 & 内外存 结构一览图

考核方式

- 考核方式

- 平时成绩 10%

- 作业、课堂参与

- 华为云数据库GAUSSDB实验

- 大作业（课程设计） 30%

- 期末考试 60%

课程管理

- 数据库课程平台
(访问方式稍后发布)



- 课程助教

贾惠婷, QQ: 2207143187

严宇皓, QQ: 709043514

陈 叹, QQ: 2406637273

邮箱: jiahuiting2021@163.com

邮箱: 709043514@qq.com

邮箱: nidhs1125@qq.com

课程学习建议

- 认真听课，积极参与
- 勤奋自学与实践
- 合理安排时间

课程内容

- 数据库系统基本概念（数据模型，体系结构）

- 关系数据库

- 关系数据库标准语言SQL

- 数据库保护

第一部分
基础理论

- 关系数据理论

- 数据库设计

第二部分
设计理论

- 存储管理与索引

- 查询处理与查询优化

- 事务处理技术

第三部分
实现技术

- 数据库技术新发展

第四部分
新技术

第一章 概述

- 本章主要内容

- 数据管理技术的发展

- 数据模型

- 数据库系统结构

- 本章要阐述的问题

- 1. 数据库与其他数据管理技术相比有什么特点?

- 2. 数据库技术要解决的基本问题是什么?

- 3. 有效解决这些基本问题的技术是什么?

第一章 概述

- 数据管理技术的发展
- 数据模型
- 数据库系统结构

数据库技术产生与发展背景

时间

1946

M1950s

M1960s

L1960s

今

计算机及相关技术的发展

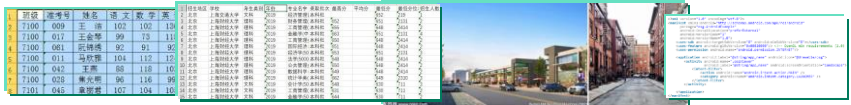
早期电子计算机

硬件发展
文件系统
高级程序设计语言

- 操作系统
- 算法与数据结构
- 软件工程
- 数据库技术（层次、网状、关系）

Internet
高性能数据采集设备

大规模业务数据管理技术（关系数据库和NewSQL）
大规模图像、音视频、文本数据管理技术（NoSQL）



计算机应用的发展

科学计算

业务数据
处理

科学计算

业务数据
处理

智能化
数据处理

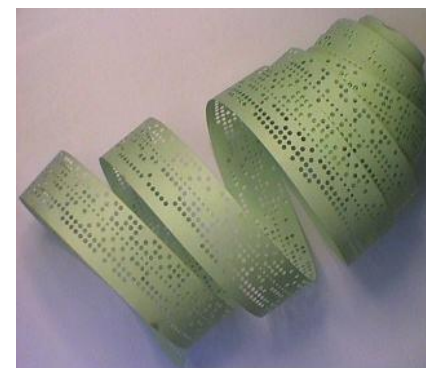
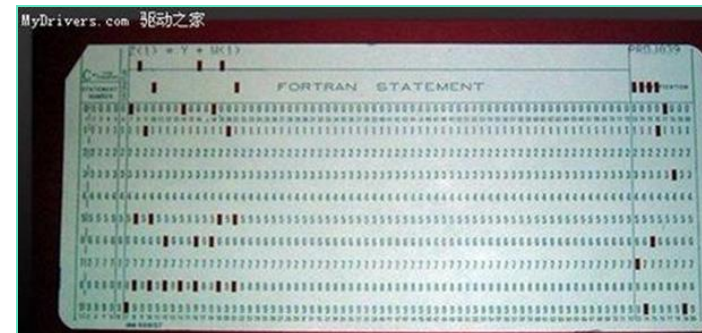
科学计算

数据管理技术的发展过程

- 人工管理阶段
- 文件系统阶段
- 数据库系统阶段

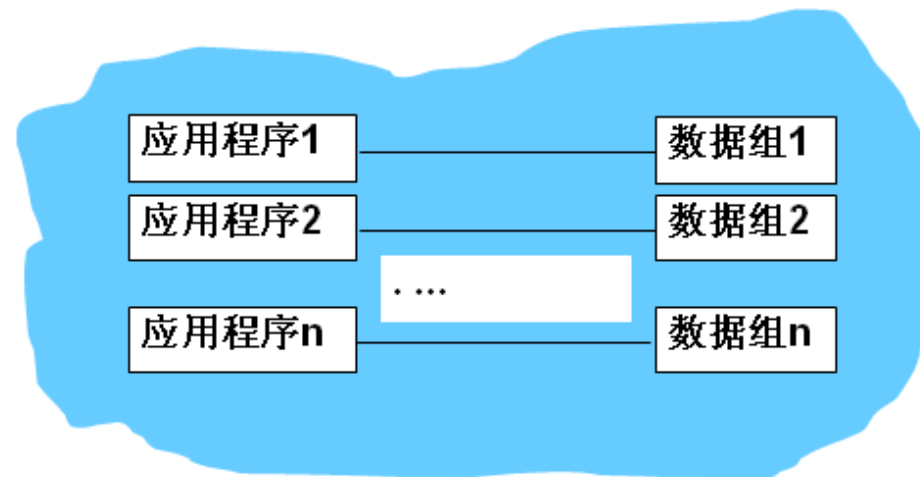
人工管理阶段

- 时间：20世纪50年代中期以前
- 背景：
 - 外存只有磁带、卡片、纸带等，没有磁盘等直接存取设备。
 - 没有操作系统，没有数据管理软件（用户用机器指令编码）。
 - 计算机主要用于科学计算。



人工管理阶段特点

- 数据不在计算机上保存；
- 没有软件系统对数据进行管理。程序规定数据的逻辑结构与物理结构。数据与程序不具有独立性；
- 基本没有文件概念，数据组织方式必须由程序员自行设计；
- 一组数据对应一个程序，数据是面向应用的，程序间不能共享数据。

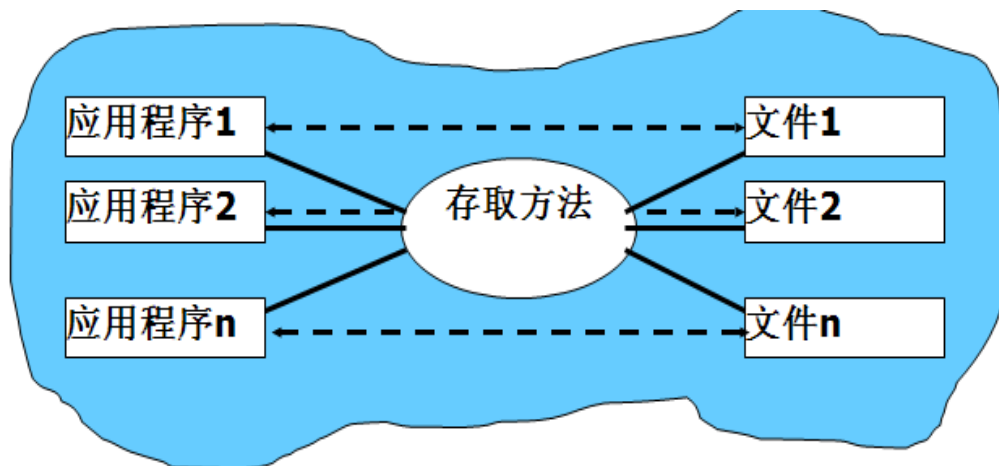


文件系统阶段

- 时间：20世纪50年代后期到60年代中期
- 背景：
 - 外存有了磁盘、磁鼓等直接存取设备；
 - 有了专门管理数据的软件，一般称为文件系统，包括在操作系统中；
 - 计算机不但用于科学计算，还用于管理。

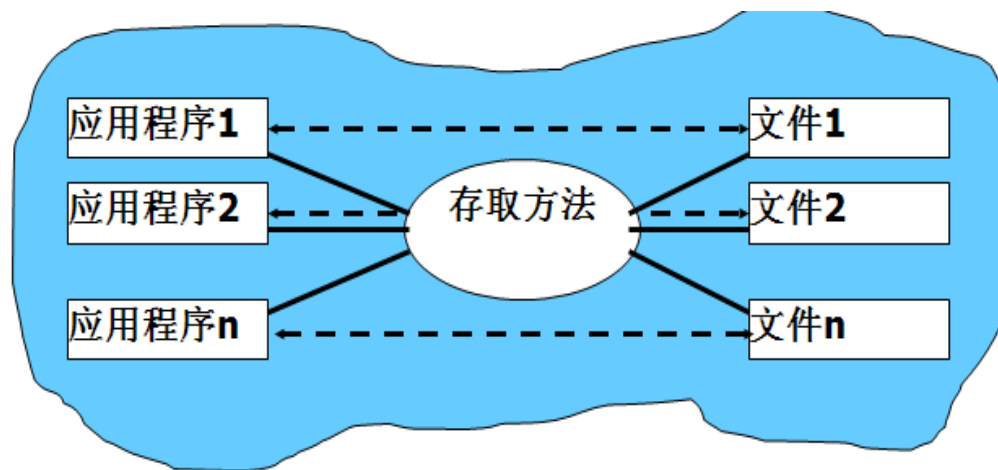
文件系统阶段特点

- 数据以文件形式保留在外存上；
- 文件多样化；
- 数据的存取基本上以记录为单位；
- 程序和数据有一定的独立性；
- 文件与应用程序基本上是一一对应，数据共享性差，冗余度大



文件系统存在的问题

- 文件分散、相互独立，数据冗余度大，数据和程序缺乏独立性
 - 存储空间浪费
 - 容易造成数据的不一致性
 - 数据维护难度大
 - 程序维护工作量大

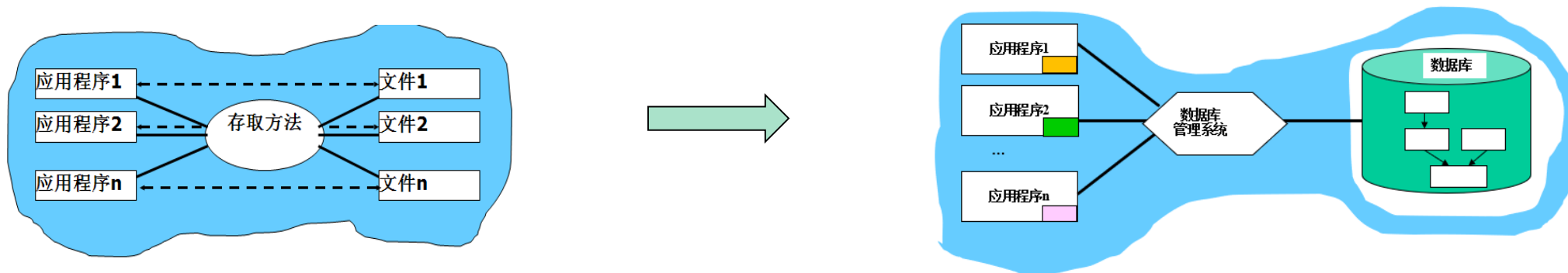


数据库系统阶段

- 时间：20世纪60年代后期开始
- 背景：
 - 外存有了大容量磁盘，光盘；
 - 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）
 - 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，其中维护的成本更高。

数据管理技术的新需求

- 突破文件系统分散管理的弱点，实现对数据的集中控制，统一管理，要求数据是“集成的”和“共享的”
 - ◆ “集成的”是指把特定应用环境中各种应用相关的数据及数据之间的联系，全部地、集中地并按照一定结构形式进行存储。
 - ◆ “共享的”是指数据可为多个不同的用户所共享。
- 这种需求，推动了数据库技术产生和发展



数据库系统基本问题与关键技术

- 数据库技术要解决的基本问题

- 集成数据的表示
- 统一管理下的数据共享

- 解决基本问题的核心技术

- 数据模型
- 数据独立性

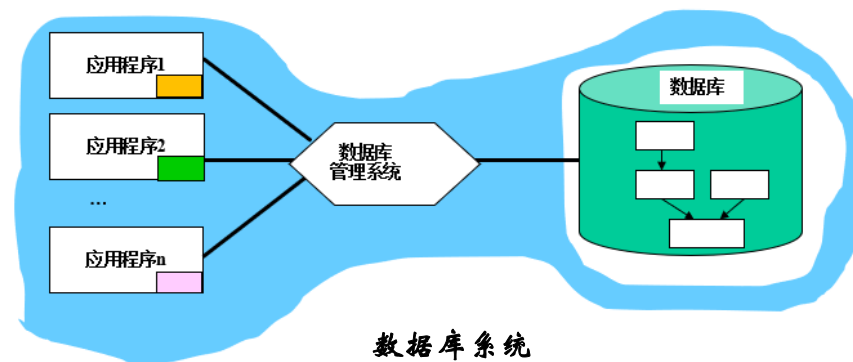
- 数据库系统中的核心软件——数据库管理系统DBMS

- 最早的DBMS —IMS(InformationManagementSystem),
IDS(Integrated Data Storage)



查尔斯·巴赫曼
Charles Bachman

IDS与DBTG报告

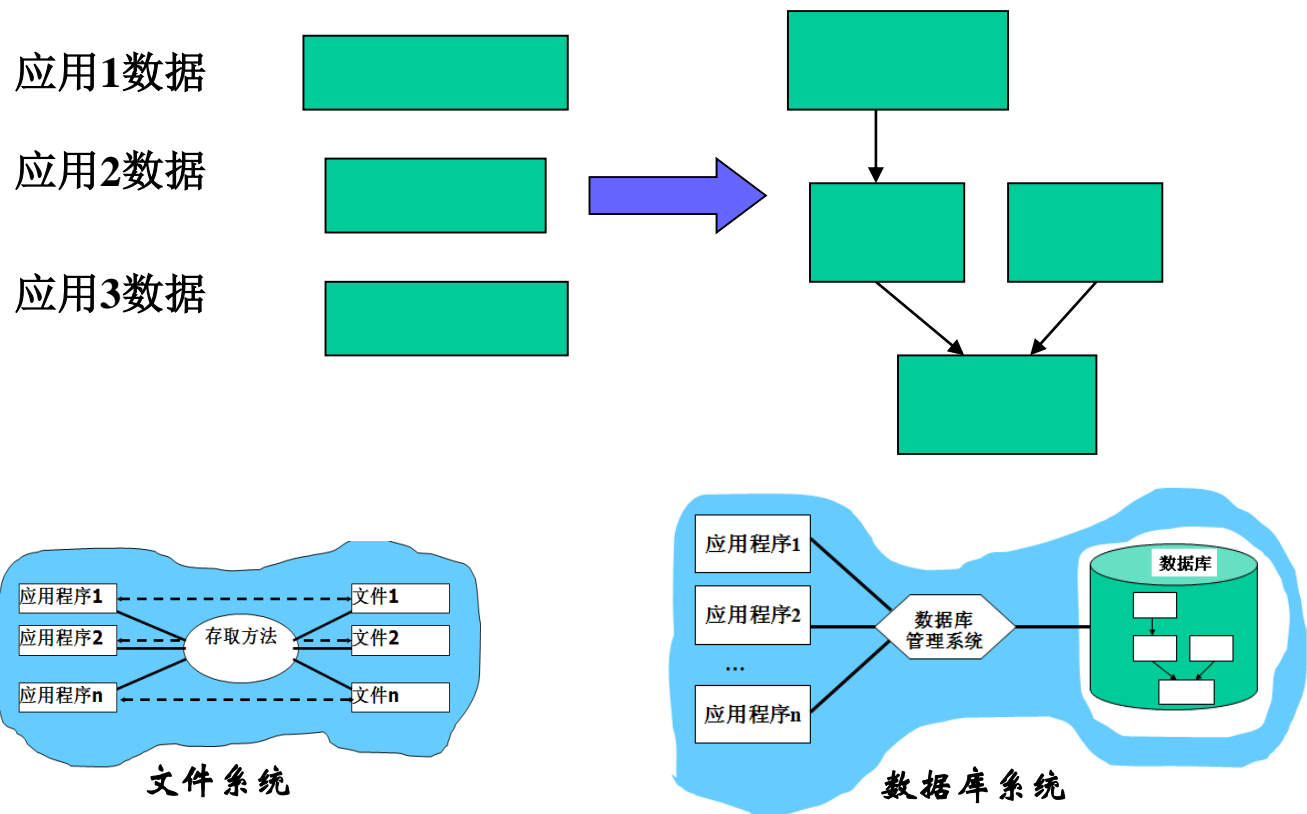


数据库系统

数据库系统数据管理特点 (1)

- 面向全组织的复杂的数据结构

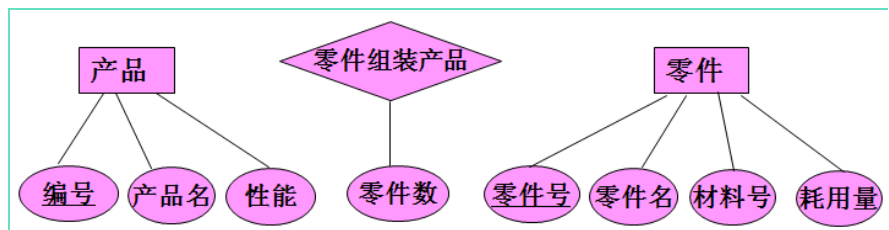
- 在描述数据时，不仅描述数据本身，还要描述数据之间的联系，使整个组织的数据结构化



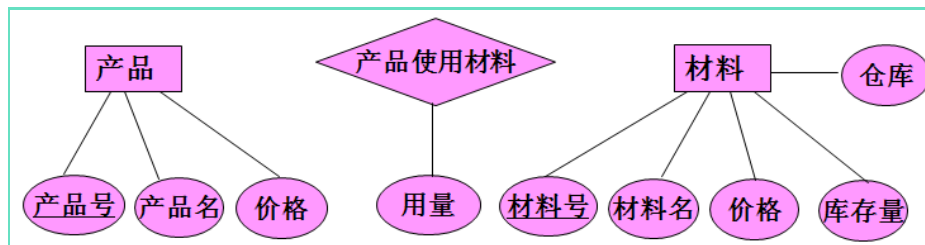
数据库系统数据管理特点 (1)

• 面向全组织的复杂的数据结构 — 示例

文件系统

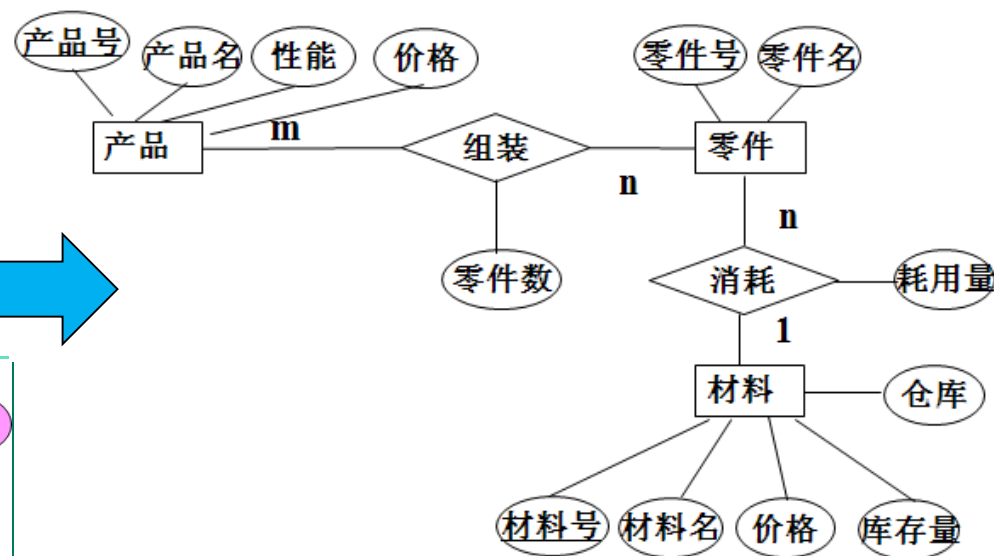


生产系统中的数据



原材料供应系统中的数据

数据库系统

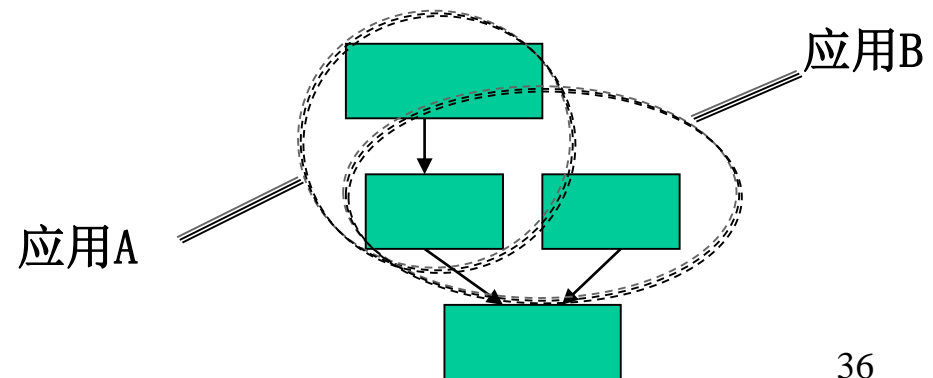


企业中的数据

— 数据结构化是数据库主要特征之一，是数据库与文件系统的根本区别

数据库系统数据管理特点 (2)

- 数据冗余度小，易扩充
 - 数据库从整体观点描述数据。数据不再面向某个应用，而是面向整个系统，从而大大减小数据的冗余度；
 - 数据库数据的应用可以有很灵活的方式，可以取整体数据的各种合理子集用于不同的应用系统，并可以根据应用需求的变化，重新选取不同子集。

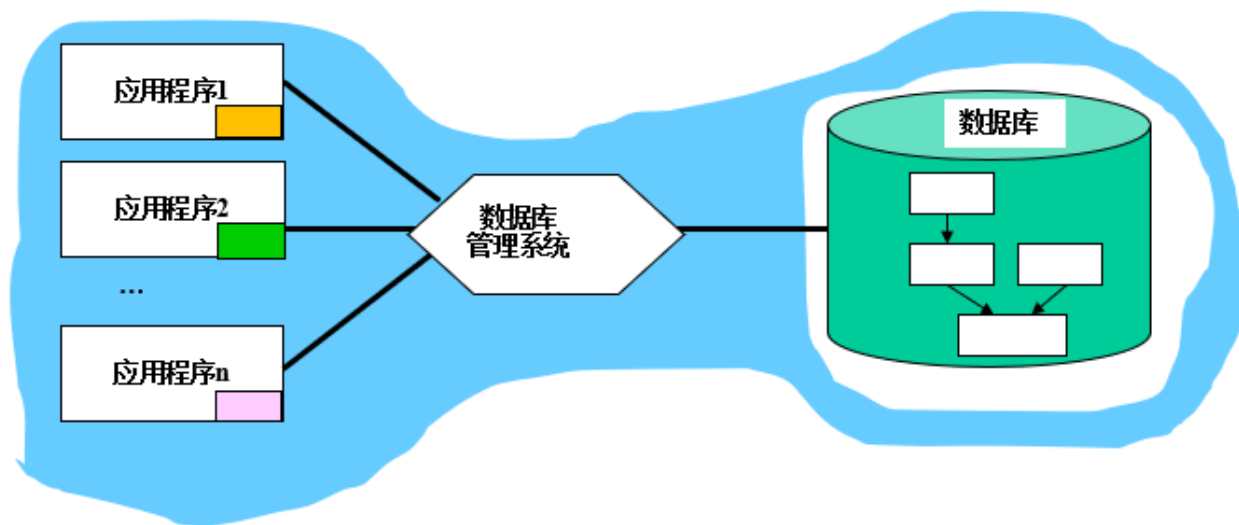


数据库系统数据管理特点 (3)

- 具有较高的数据和程序的独立性

- 数据独立性

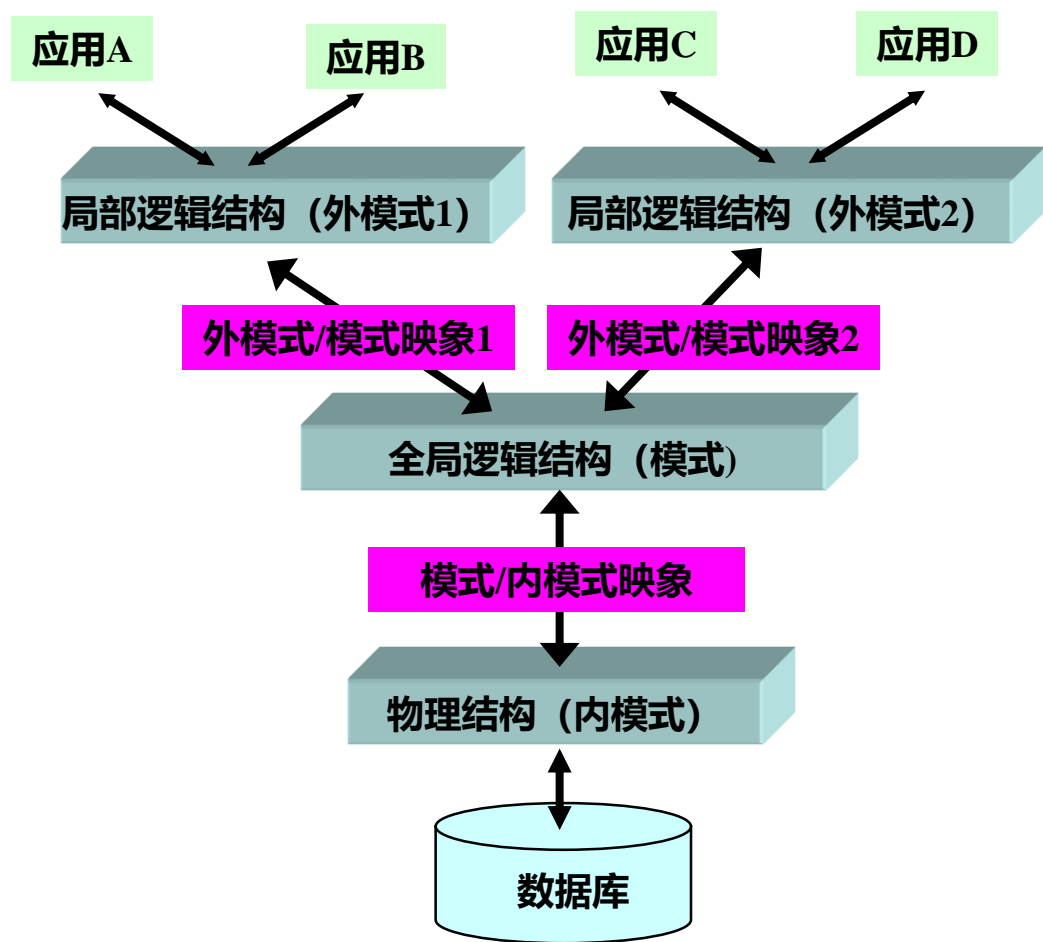
- **数据的物理独立性**：数据的存储结构（物理结构）改变时，数据的逻辑结构可以不变，从而应用程序也不必改变；
 - **数据的逻辑独立性**：数据的逻辑结构改变时，应用程序可以不变。



数据库系统数据管理特点 (3)

数据库系统提供了两方面的映象（转换）功能：

- 数据的存储结构与逻辑结构之间的映象——实现数据的物理独立性
- 数据的全局逻辑结构与某类应用所涉及的局部逻辑结构之间的映象——实现数据的逻辑独立性



数据库系统体系结构

数据库系统数据管理特点 (4)

- 统一的数据控制功能

- 数据的安全性控制

- 保护数据以防止不合法的使用所造成数据的泄密和破坏

- 数据的完整性控制

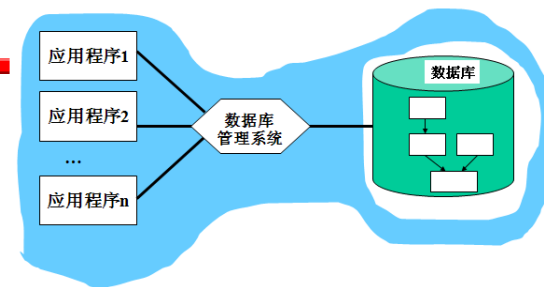
- 指数据的正确性与相容性

- 并发控制

- 对多用户的并发操作进行控制、协调，保护数据的完整性

- 数据库恢复

- 将数据库从错误状态恢复到某一已知的正确状态



数据库系统数据管理特点 (5)

- 数据的最小存取单位是数据项
 - 既可以存取一个或一组记录，也可以数据库中某个或一组数据项。

课程编号	课程名称	先行课编号
C1	数据库	C5
C2	数学	
C3	信息系统	C1
C4	操作系统	C6
C5	数据结构	C7
C6	数据处理	
C7	C语言	C6



数据库系统基本问题与关键技术

- 数据库技术要解决的基本问题

- 集成数据的表示
- 统一管理下的数据共享

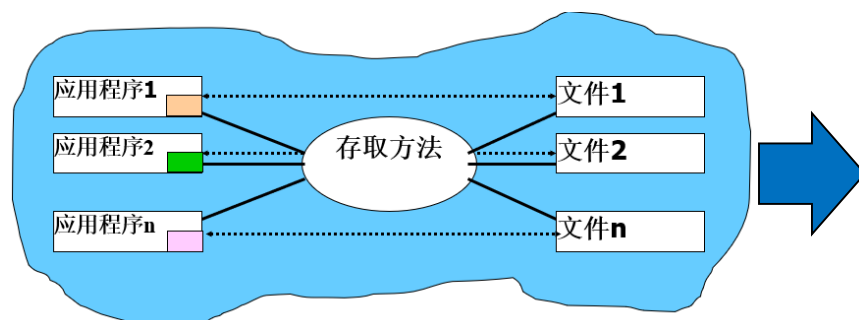
- 解决基本问题的核心技术

- 数据模型
- 数据独立性

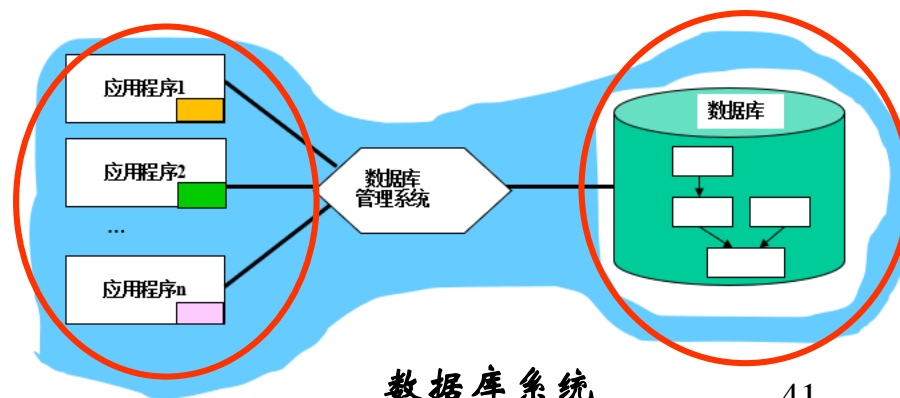


查尔斯·巴赫曼
Charles Bachman

IDS与DBTG报告



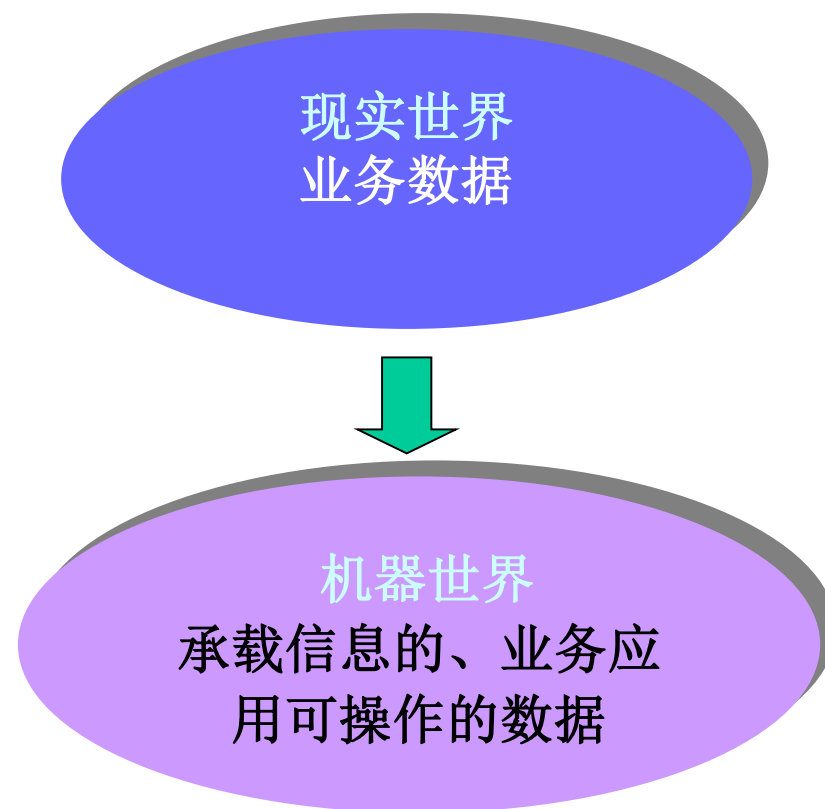
文件系统



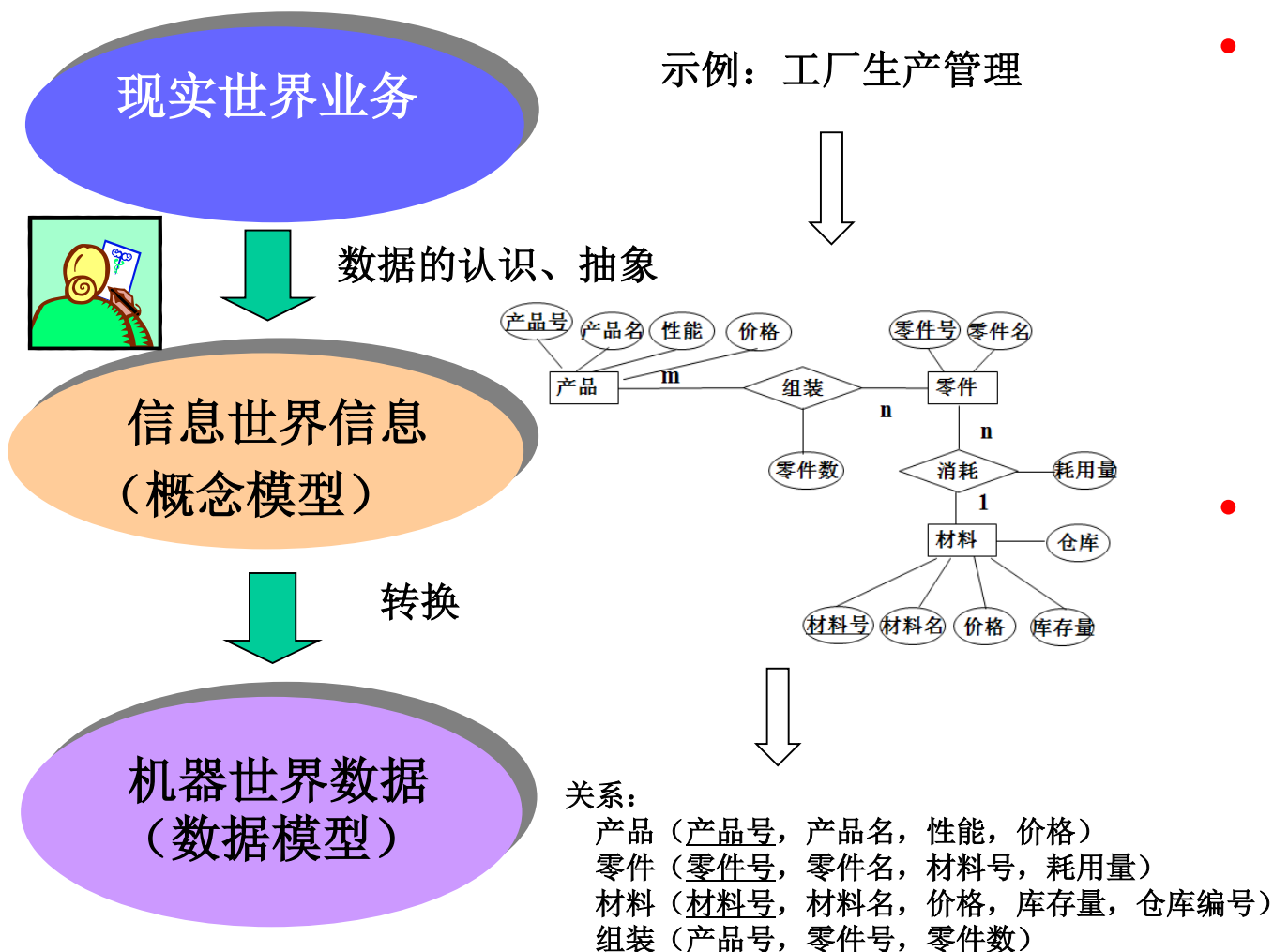
数据库系统

数据模型

- 数据模型用来抽象和表示现实世界中的数据和信息
 - 描述的对象
 - 描述的目标



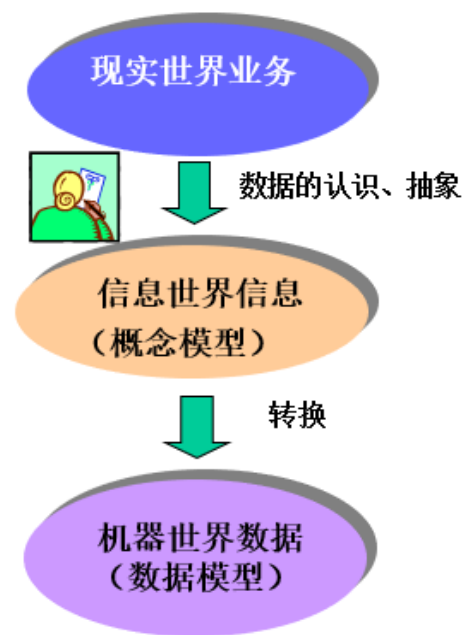
数据模型的层次



- **概念模型**用于信息世界建模，是现实世界到信息世界的抽象，是用户和数据库设计人员进行交流的语言。
- **数据模型**用于机器世界，按计算机系统的观点对数据建模。

数据模型

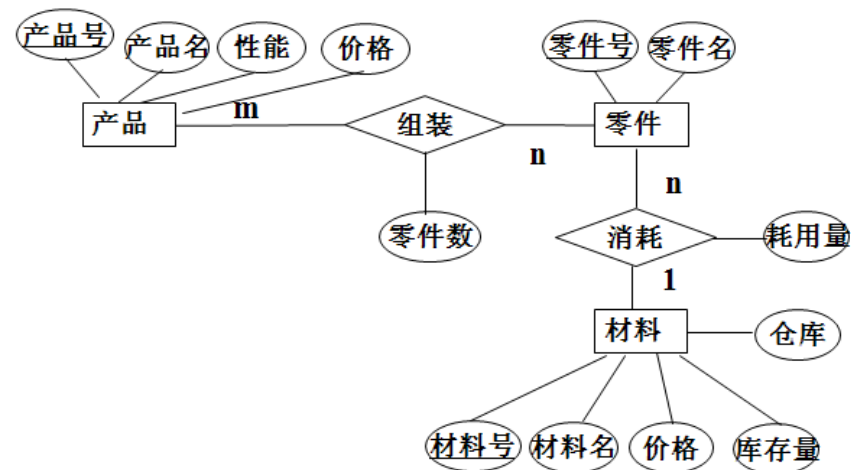
- 在数据库中，根据模型应用的不同目的，将模型分为两类或两个层次：
 - 概念模型（也称信息模型）
 - 数据模型（如层次、网状、关系模型）
- 概念模型
- 数据模型
- 数据模型的分类



概念模型

- 概念模型基于信息世界的主要概念，表达应用中的各种语义（信息）；
- 具有较强的语义表达能力，能够方便、直接表达应用中的各种语义；
- 概念模型应该简单、清晰、易于理解；
- 概念模型最常用的表示方法：

实体—联系方法（E-R）法。



概念模型中的基本概念 (1)

- 实体 (Entity) : 客观存在并可相互区分的事物;
- 属性 (Attribute) : 实体所具有的某一特性;
- 码 (Key) : 唯一标识实体的属性集;
- 域 (Domain) : 某个 (些) 属性的取值范围;
- 实体型 (Entity Type) : 表示一类实体, 用实体名及其属性名集合来抽象、刻画。

概念模型中的基本概念 (2)

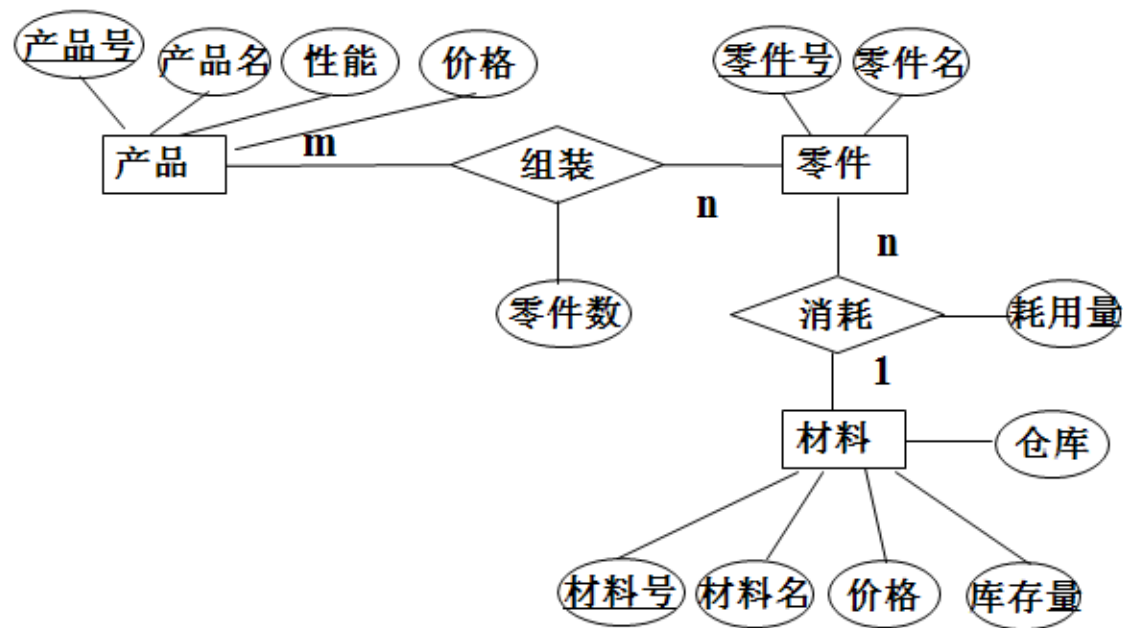
- 联系 (Relation)：实体型之间的联系，是实体之间的相互关联。

- 名称

- 类型

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

- 可以具有属性



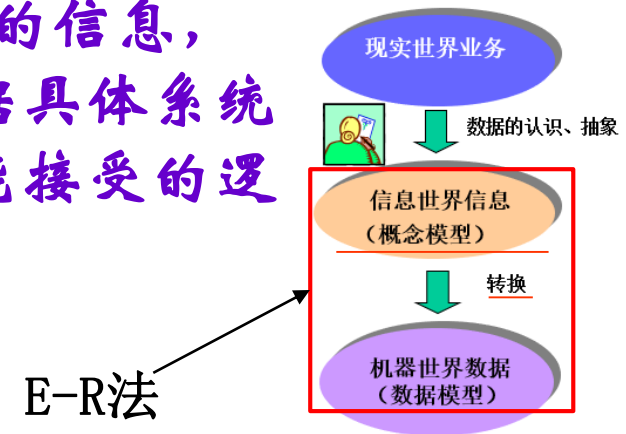
E-R法概述

- Peter Chen在1976年提出E-R法（Entity-Relation Approach，实体联系方法）也称为E-R模型；



Dr. Peter Chen (陳品山) at
Louisiana State University (LSU)

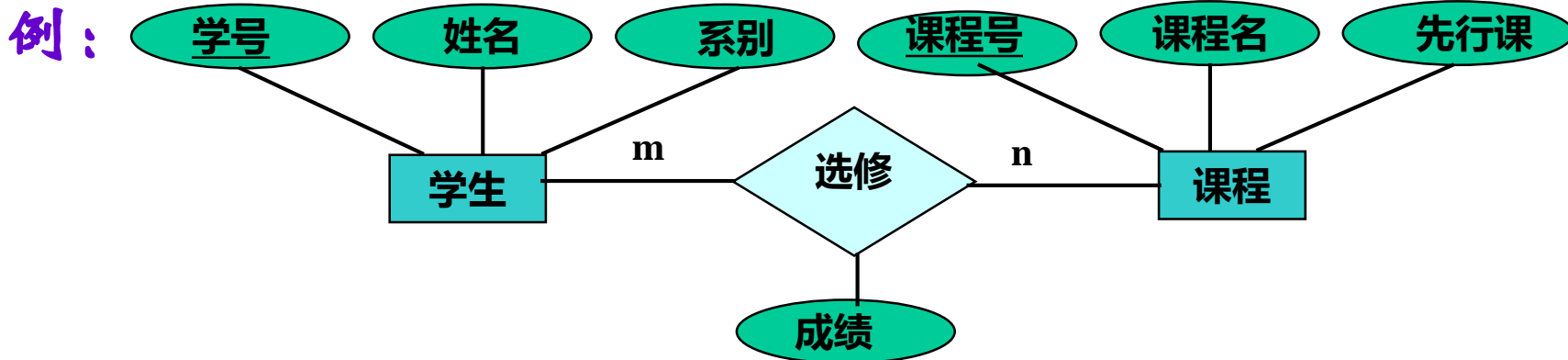
- E-R法的思想是用**E-R图**描述现实世界的信息，这种信息结构称为**概念结构**，然后根据具体系统的要求将概念结构转换成特定系统所能接受的逻辑结构（层次、网状、关系）；
- E-R法由**两部分**组成：
 - 用E-R图描述现实世界；
 - 将E-R图转换成相应的数据模型。



E-R图 (1)

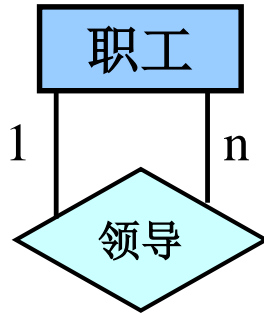
- E-R图的组成：实体、联系、属性

- 实体：用长方形表示实体型，在框内写上实体名。
- 属性：用椭圆形表示实体的属性，并用无向边把实体与其属性连接起来。
- 联系：用菱形表示实体间的联系，菱形框内写上联系名。用无向边把菱形分别与有关实体相连，在无向边旁标上联系的类型。若联系也具有属性，则属性和菱形也用无向边连接上。

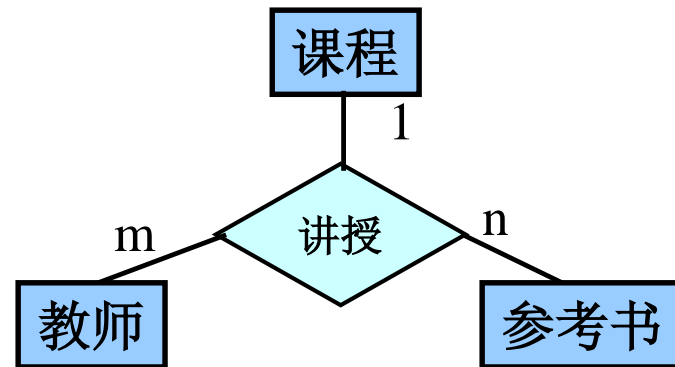


E-R图 (2)

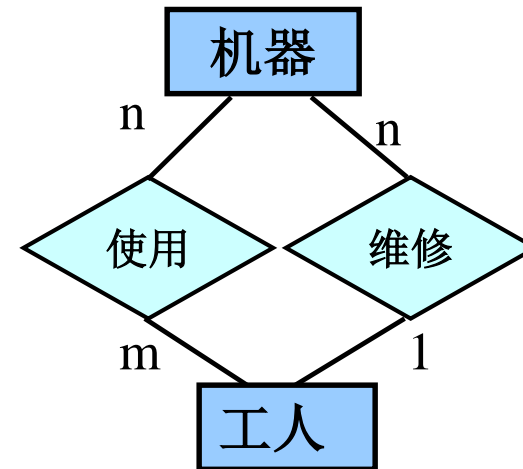
- 同一个实体集内部各实体之间也可以存在一对一、一对多、多对多的联系。(a)
- 三个或多个实体型间可能具有联系。(b)
- 两个实体型之间可具有多种联系。(c)



(a)



(b)

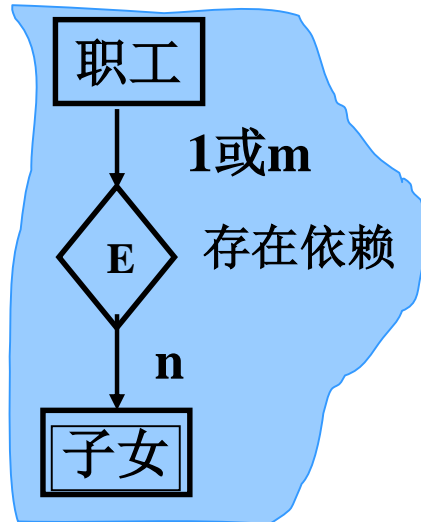


(c)

E-R图 (3)

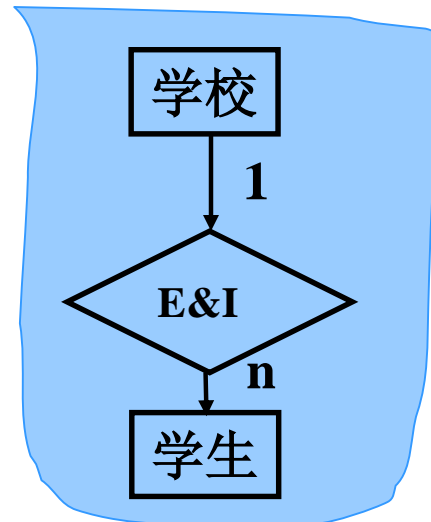
• E-R图实体之间联系的语义扩充

(1) 存在依赖



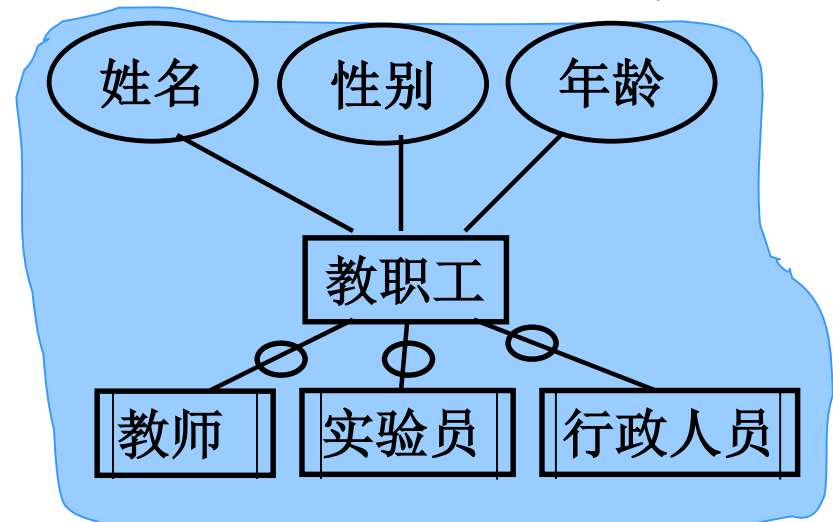
子女实体存在依赖于职工实体的存在，则称子女实体是弱实体。

(2) 标识依赖



如果实体不能由它自己的属性来唯一标识，而必须通过与它相联系的另一实体一起来标识，那么称该实体标识依赖于另一个实体。

(3) 实体的子类



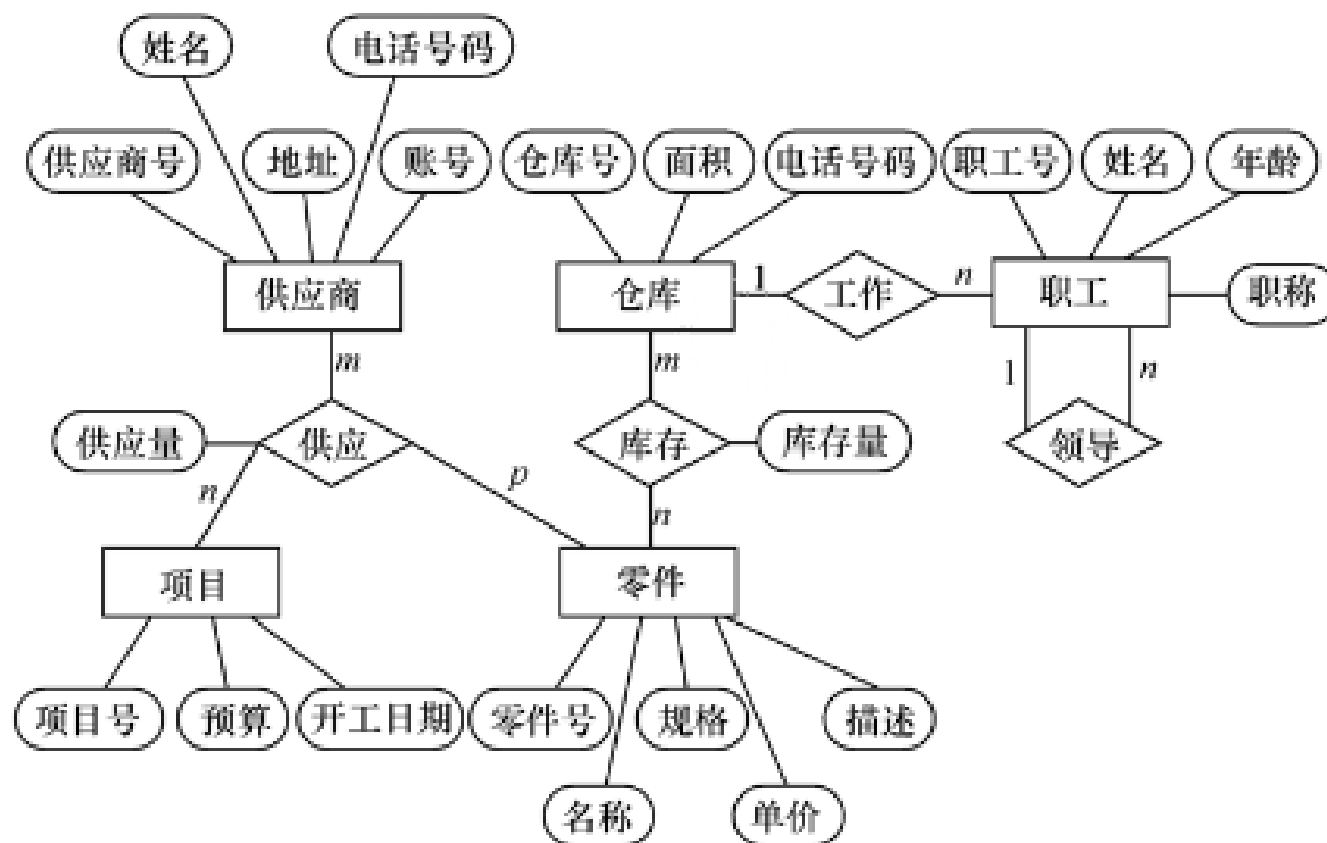
- 子类可继承父类的属性，子类也可以附加某些属性；
- 子类之间的交不一定为空。

概念模型示例

- 工厂物资管理系统——需求描述

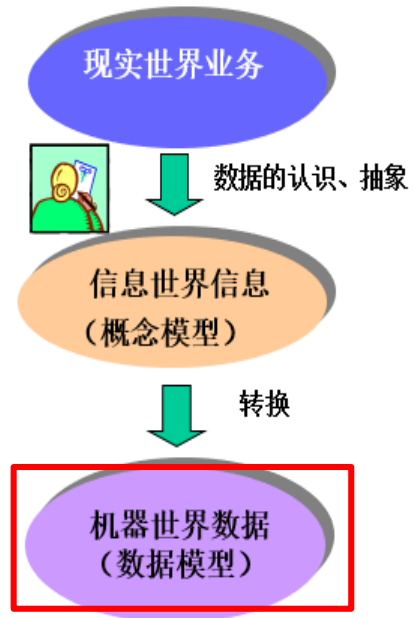
某工厂要开发数据库系统进行物资管理，需要管理仓库、仓库中的职工、保存的零件、零件所用于的项目以及供应商信息。仓库的信息包括...,职工的信息包括...。一个仓库可存放多种零件，一种零件可存放在多个仓库中；一个仓库有多名职工，但一名职工只能在一个仓库工作；仓库主任领导多名职工；一个供应商可以供给多个项目多种零件，每个项目可以使用多个供应商供应的零件，每种零件可由不同供应商供给。

• 工厂物资管理系统—概念模型 (E-R图)



数据库中的数据模型

- 数据模型是严格定义的概念集合。这些概念精确地描述系统中数据的静态特性、动态特性和完整性约束。
- 数据模型的三要素：
 - 数据结构
 - 数据操作
 - 完整性约束



数据模型三要素——数据结构

- **数据结构**由描述数据对象以及对象之间联系的一组概念组成。包括：
 - 描述**对象**的类型、内容、性质的概念，如关系模型中的域、属性等；
 - 描述**对象之间联系**的概念，如关系模型中的关系；
- 是数据**静态特性**的描述；
- **数据结构是刻画数据模型最重要的方面**；通常按照数据结构的类型来命名数据模型。

数据模型三要素——数据操作

- 是对数据库中各种数据对象（型）的实例（值）允许执行的操作集合，包括操作及操作规则；
- 定义操作的确切含义、操作符号、操作规则及操作语言；
- 是数据动态特性的描述；
- 数据库主要有检索和更新（插、删、改）两大类操作。

数据模型三要素——数据的约束条件

- 是完整性规则的集合。完整性规则是给定的数据模型中数据及其联系所有的制约和依存规则，用以保证数据的正确、相容；
- 完整约束条件包括：
 - 符合这种数据模型所必须遵守的基本的通用的完整性约束条件；
 - 针对具体数据的特定语义约束条件。

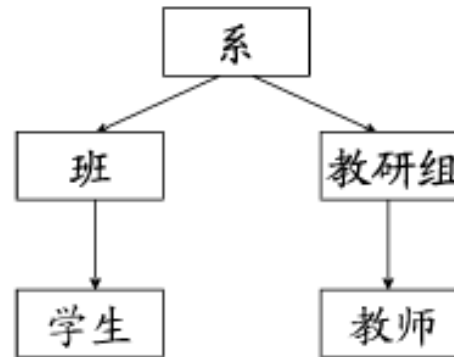
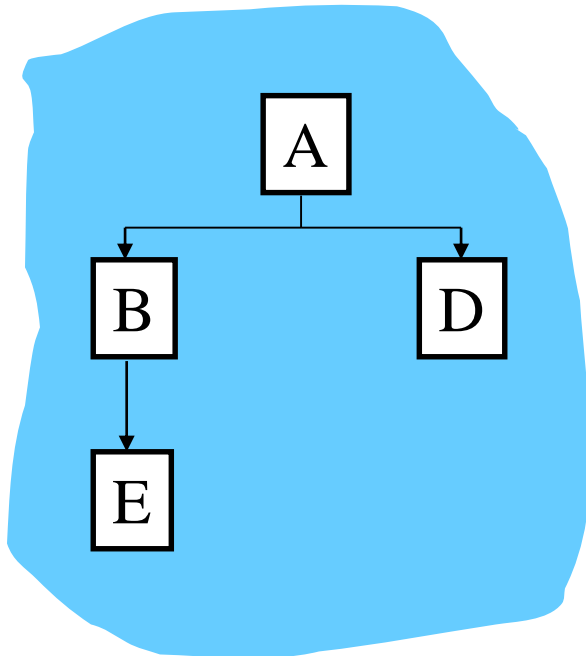


数据模型的分类

- 当前公认的基本数据模型有三类：层次模型、网状模型、关系模型。
- 各种数据模型之间的根本区别在于数据之间联系的表示方式不同：
 - 层次模型：用“树结构”表示数据之间的联系；
 - 网状模型：用“图结构”表示数据之间的联系
 - 关系模型：用“二维表”表示数据之间的联系

层次模型

- 是数据库系统中最早出现的数据模型；
- 数据结构是有向树。节点代表实体型，连线表示两实体型间的一对多联系；

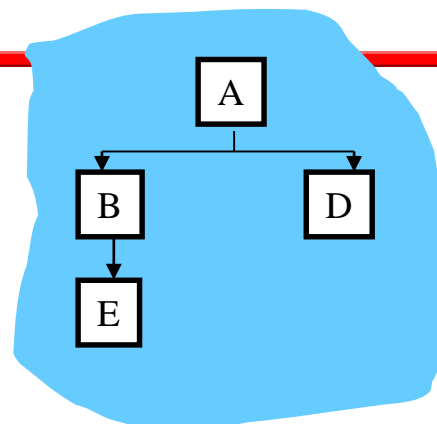


特征：

- 有且仅有一个结点没有双亲；
- 其它结点有且仅有一个双亲。

层次模型

- 优点：
 - 结构简单，易于实现。
- 缺点：
 - 支持的联系种类太少，只能直接表示二元一对多联系。
 - 数据操纵不方便，子结点的存取只能通过父结点来进行。
- 代表产品：
 - IBM的IMS数据库，1968年研制成功。



网状模型

- 数据结构是有向图。节点代表实体型，连线表示两实体型间的一对多联系。

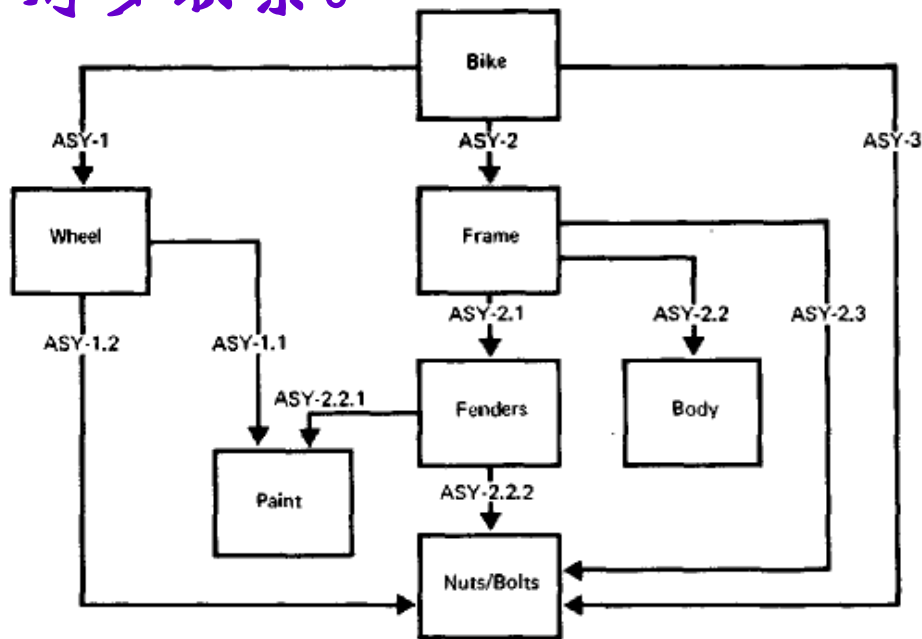
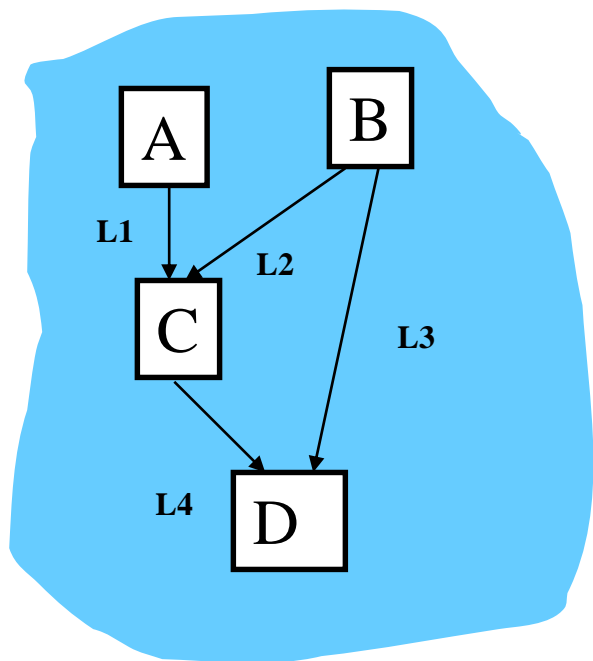


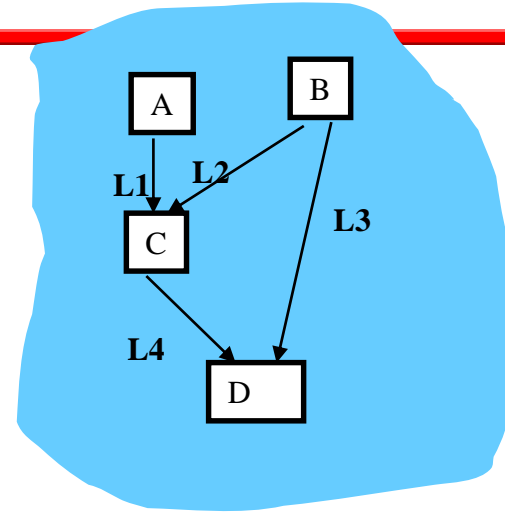
FIGURE 14. Parts explosion as a graph.

特征：

- 可有一个以上的结点没有双亲；
- 至少有一个结点有多于一个的双亲。

网状模型

- 特点：
 - 表达的联系种类丰富。
 - 结构复杂。
- 代表产品—DBTG系统：
 - 1969年，由美国CODASYL组织（Conference On Data System Language，数据系统语言协商会）提出DBTG报告，确立了网状数据库系统的概念、方法、技术。



关系模型

- 用二维表格结构（关系）表示实体及实体之间的联系。

学生

学号	姓名	年龄	系别
s1	A	18	CS
s2	B	18	CS
s3	C	18	MA

课程

课号	课名	先行课号
c1	aaa	
c2	bbb	c1
c3	ccc	

学生选课

学号	课号	成绩
s1	c1	80
s1	c2	90
s2	c1	95

注意：要求关系中每个分量是不可分的数据项

例，下列关系：职工（职工号，姓名，职称，工资，扣除，实发），不满足关系模型要求

职工号	姓 名	职 称	工资			扣除		实 发
			基本工资	岗位津贴	业绩津贴	三 险	个人所得 税	
86051	陈平	讲师	1 305	1 200	1 850	160	112	4 083
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

关系模型

- 特点
 - 用关系描述实体及实体间的联系。这种描述一致性使数据结构大大简化，概念简单。
 - 可直接表示多对多联系。
 - 关系必须是规范化关系。即每个分量是不可分的数据项，或不许表中套表。
 - 关系模型是建立在数学概念基础上，有较强的理论基础。
- 早期代表IBM开发的System R和加州大学Berkley分校开发的INGRES
- 关系数据库产品：Oracle, MySQL, PostgreSQL, SQL Server等，国内：达梦、人大金仓、OpenGauss等。



数据库系统基本问题与关键技术

- 数据库技术要解决的基本问题

- 集成数据的表示

- 统一管理下的数据共享

- 解决基本问题的核心技术

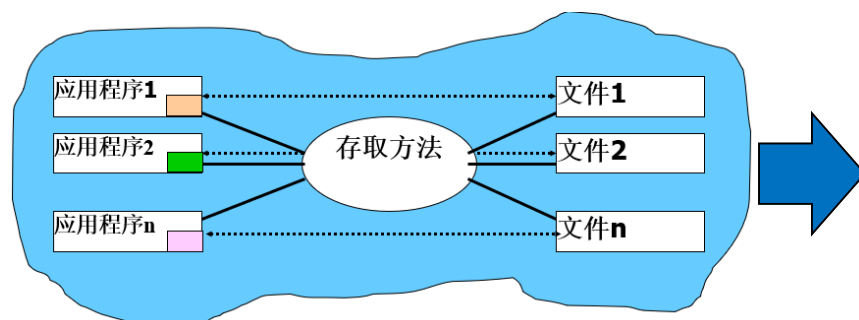
- 数据模型

- 数据独立性

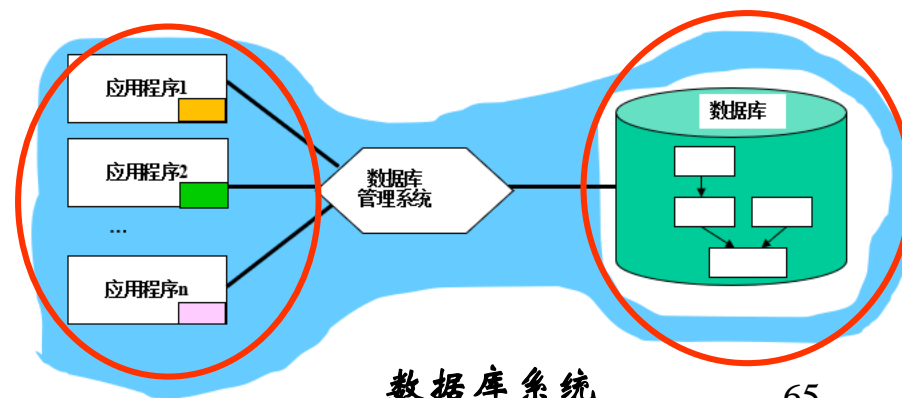


查尔斯·巴赫曼
Charles Bachman

IDS与DBTG报告



文件系统



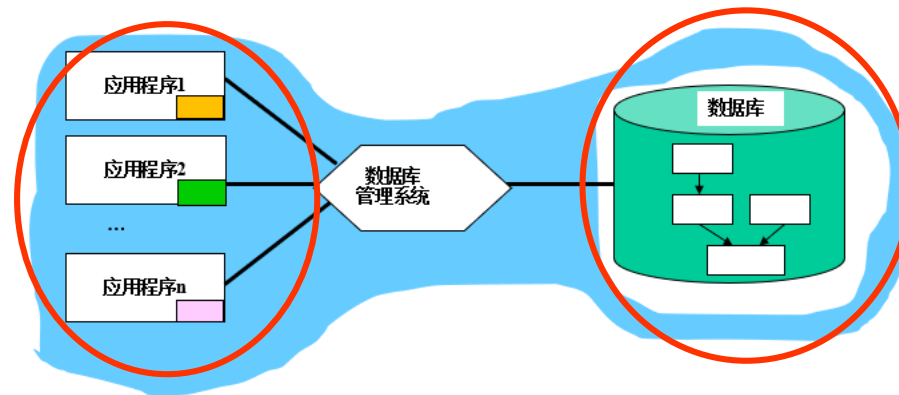
数据库系统

数据独立性与数据共享

- 数据独立性是指应用程序与数据结构之间互相独立的关系

- 数据独立性的层次

- 数据的物理独立性
- 数据的逻辑独立性

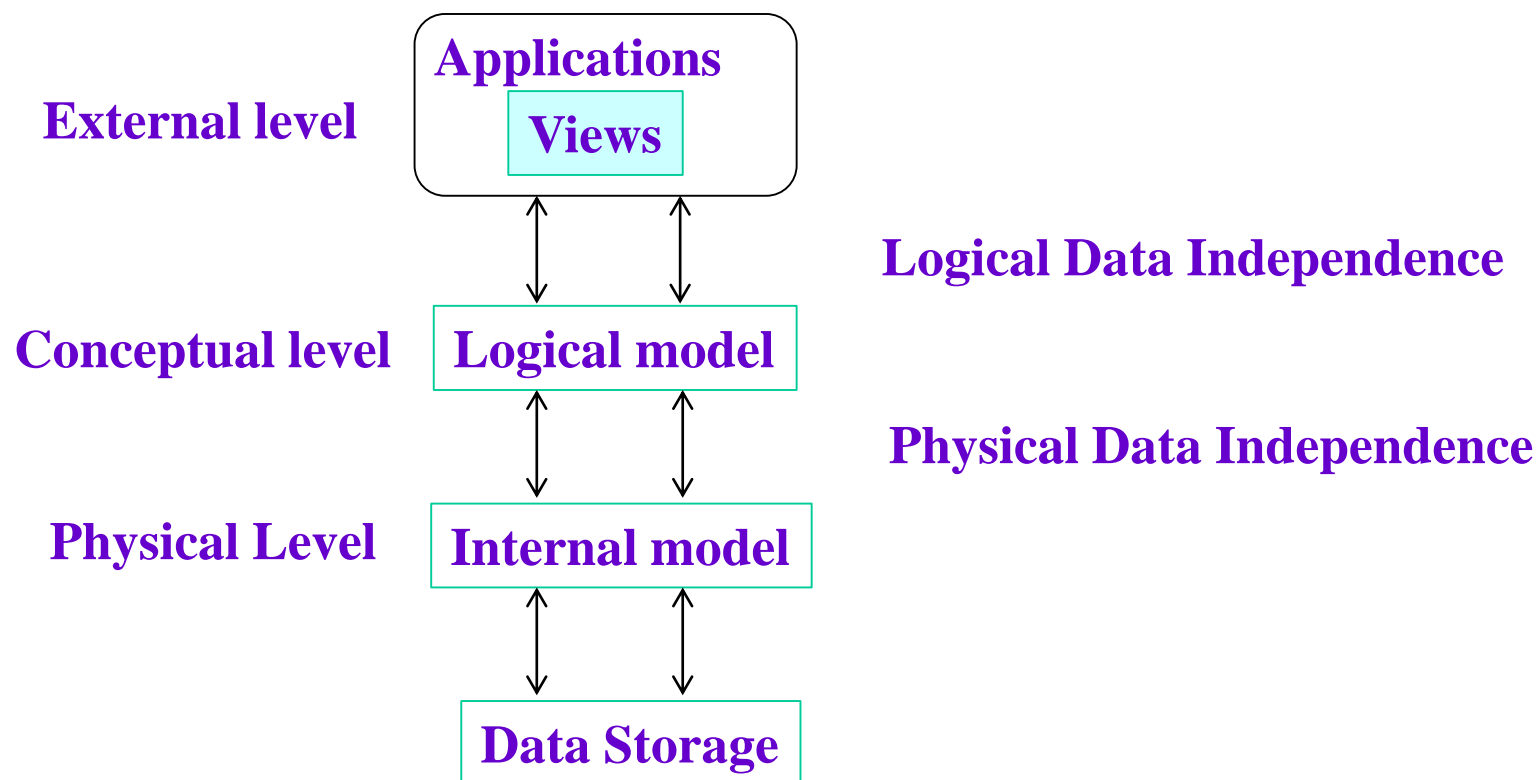


- 数据逻辑独立性使得数据库逻辑结构变化后，程序中使用数据结构不变

- 即应用中使用的数据结构可以与数据库统一数据逻辑结构不同
- 也即意味着数据库统一逻辑结构可以对应不同应用使用的多个数据结构
- 从而使得数据库可以支持多种应用特定的数据表示和操作，从而实现数据共享

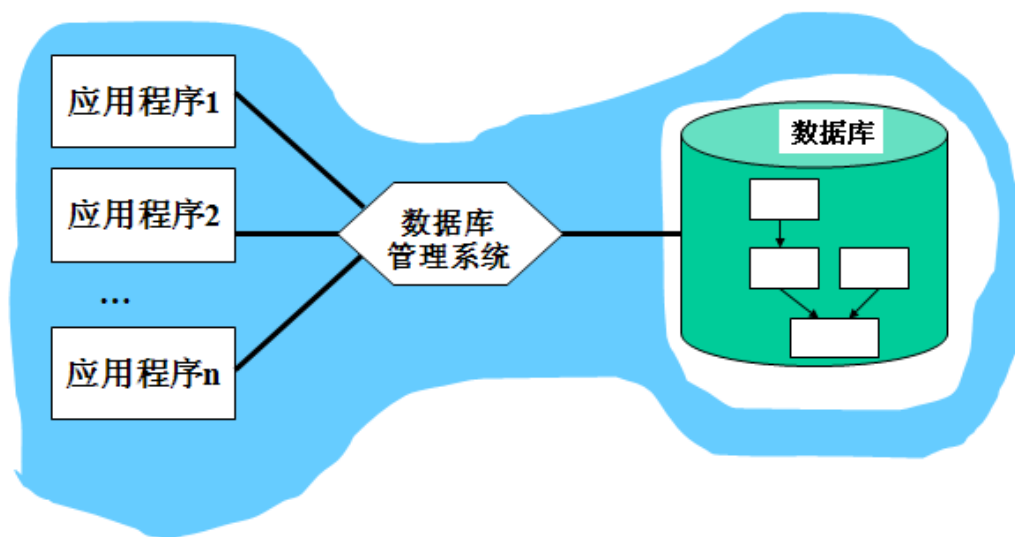
数据独立性的实现方法

- 数据库结构多层次抽象，层次之间建立映射/对应关系

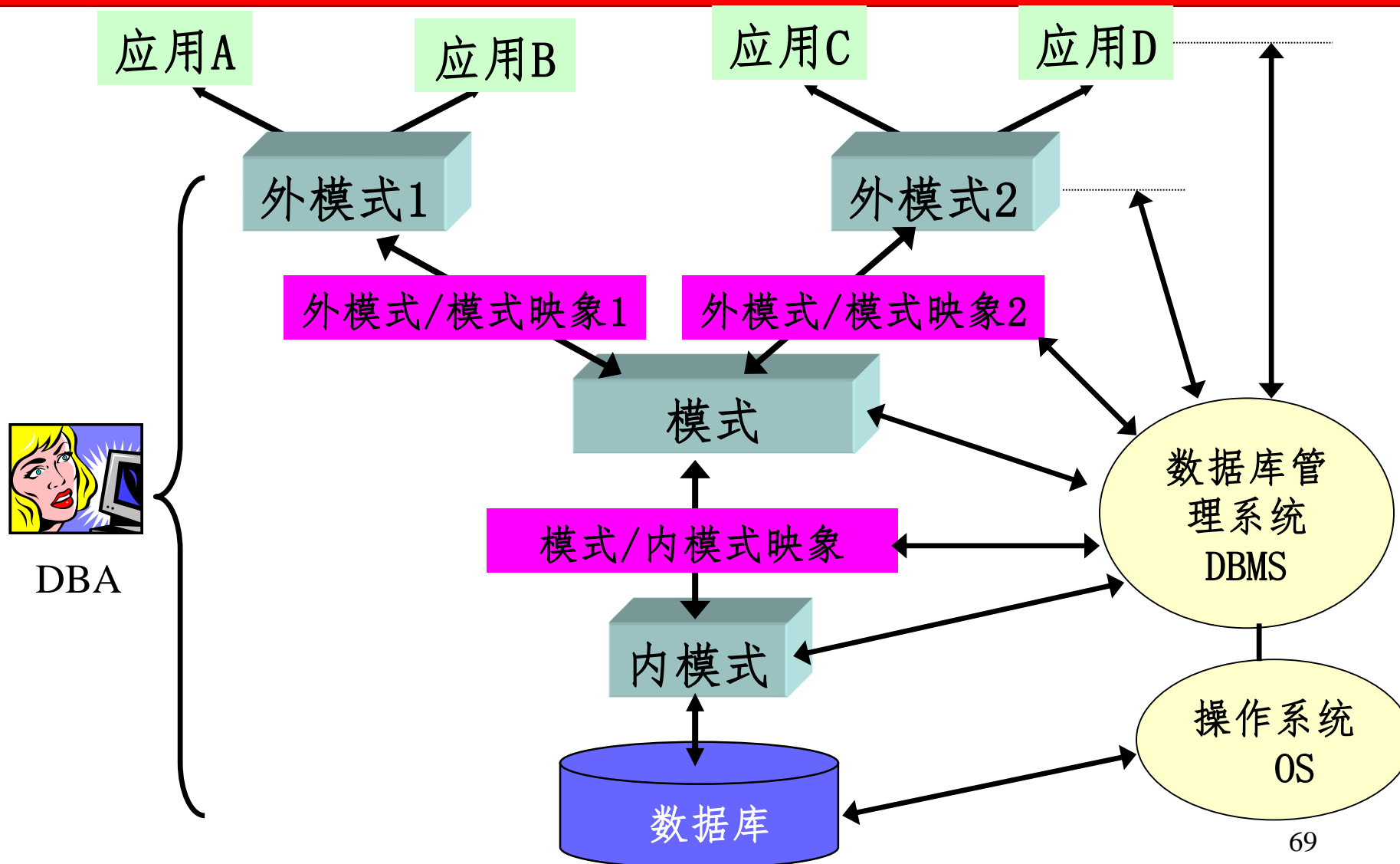


数据库系统的结构

- 数据库系统体系结构上具有三级模式、两级映像的结构特征
 - 模式
 - 外模式
 - 内模式
 - 两级映像
- DBMS和DBA

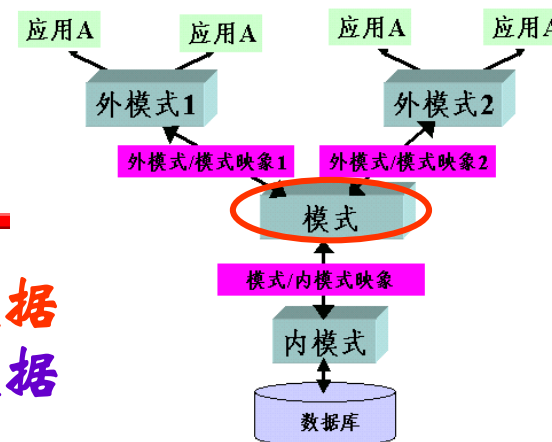


数据库系统的结构



数据库系统三级模式结构——模式

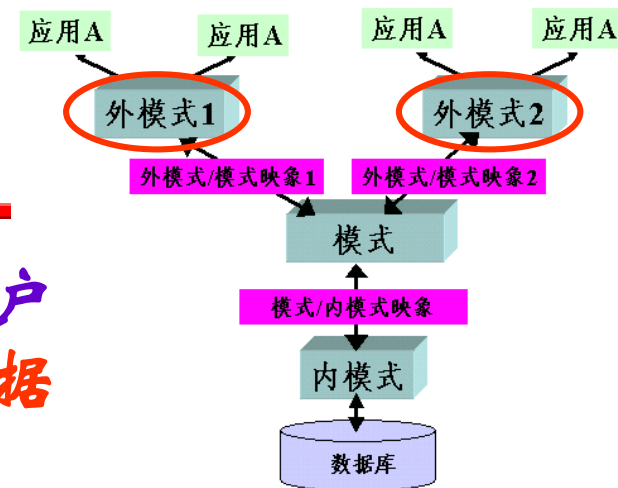
- 也称为逻辑模式、概念模式，是数据库中全体数据的逻辑结构和特性的描述。是所有用户的公共数据视图；
- 是三级模式的核心。不涉及数据物理存储细节，与具体的应用程序与编程语言无关；
- 具体定义数据的逻辑结构（数据记录结构、数据之间的联系）、数据安全性、完整性要求；
- 数据库系统提供模式描述语言（模式DDL，Data Description Language）进行模式定义。用模式DDL写出的一个数据库逻辑定义的全部语句，称为某个数据库的模式。



数据库系统三级模式结构

——外模式

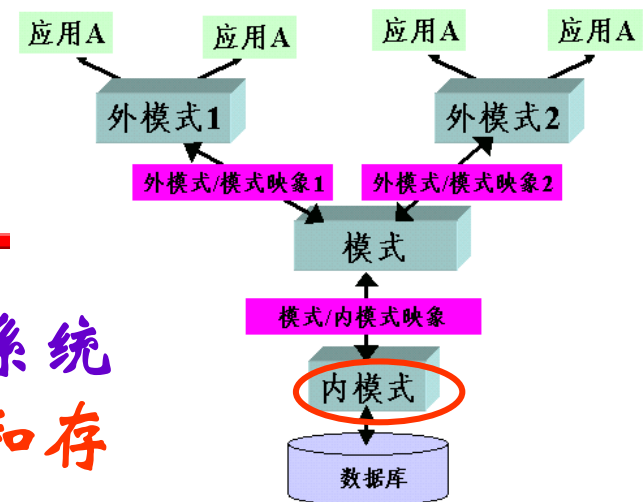
- 也称为子模式或用户模式。是个别用户的数据视图，即与某一应用有关的数据的逻辑表示；
- 通常是模式的子集。不同应用的外模式可以相互覆盖，一个应用只能启用一个外模式；
- 数据库系统提供外模式描述语言（外模式DDL）定义外模式。外模式DDL和用户选用的程序设计语言具有相容的语法。



数据库系统三级模式结构

——内模式

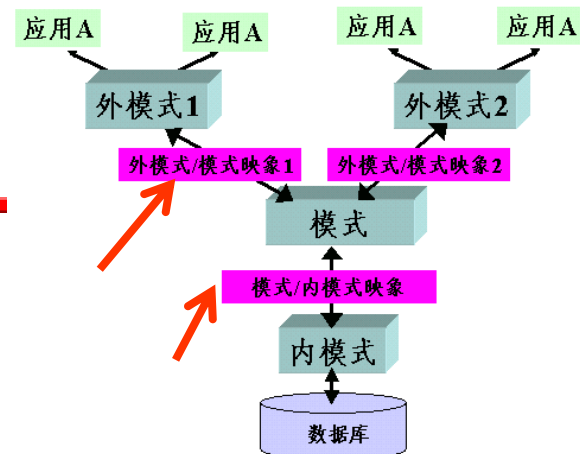
- 也称为存储模式，是数据在数据库系统内部的表示，即对数据的物理结构和存储方式的描述；
- 内模式通常用内模式数据描述语言（内模式DDL）来描述和定义。



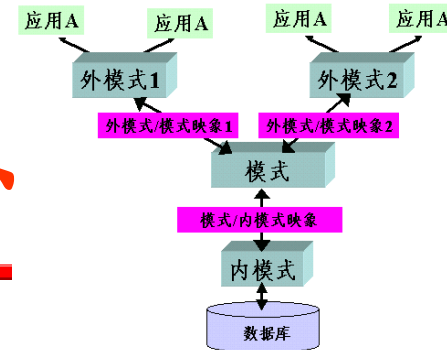
数据库系统三级模式结构

——两级映像

- 外模式/模式映像定义某个外模式与模式之间的对应关系。当模式改变时，外模式/模式映像做相应改变，可以保证外模式不变。——数据的逻辑独立性
- 模式/内模式映像定义数据逻辑结构与存储结构之间的对应关系。当内模式改变时，模式/内模式映像做相应修改，使得模式保持不变。——数据的物理独立性

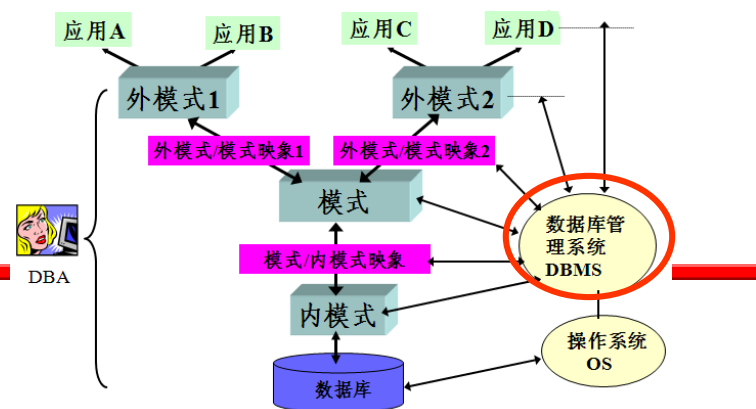


数据库系统三级模式结构优点

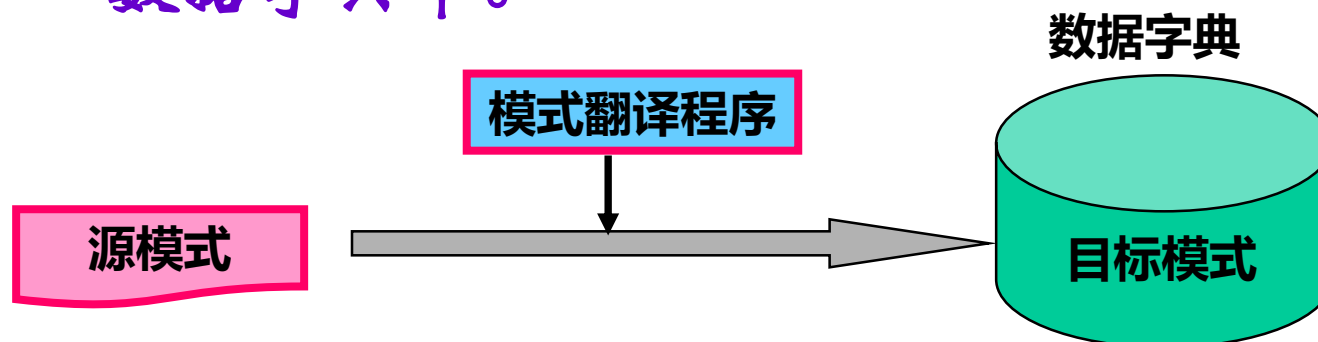


- 保证数据的独立性
 - 模式与内模式分开——数据物理独立性
 - 外模式与模式分开——数据逻辑独立性
- 简化用户接口，方便用户使用
 - 用户只按照外模式操作，无需了解数据库的总体逻辑结构与物理存储结构。
- 有利于数据共享
 - 从模式产生不同的外模式，外模式间可相互覆盖。
- 有利于数据的安全保密
 - 应用程序只能操作其对应的外模式

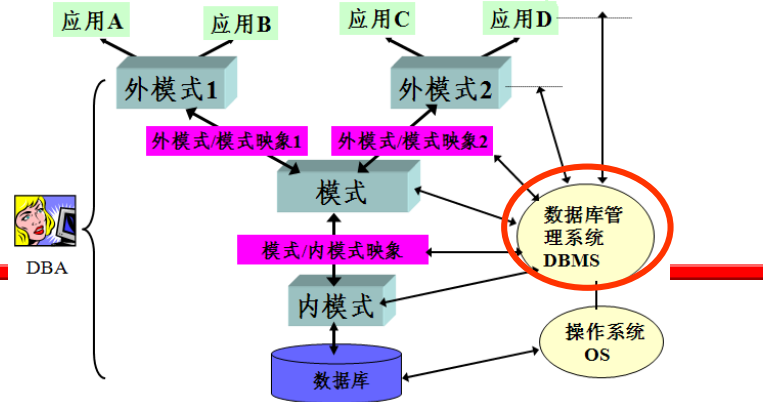
DBMS主要功能 (1)



- DBMS是数据库系统的核心软件，在操作系统支持下工作
- 数据库定义功能
 - 提供DDL语言描述外模式、模式、内模式（源模式）。
 - 模式翻译程序把源模式翻译成目标模式，存入数据字典中。

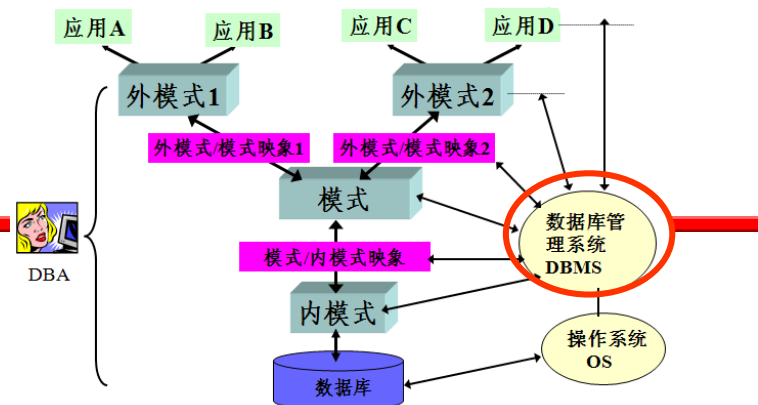


DBMS主要功能 (2)



- 数据存取功能
 - 提供DML语言 (Data manipulation language) 对数据库进行检索、插入、修改、删除。
- 数据库运行管理
 - 并发控制、存取控制、完整性约束条件检查和执行, 日志组织和管理, 事务管理和自动恢复。

DBMS主要功能 (3)



- 数据组织、存储和管理

- 用户数据、索引、数据字典的组织、存储和管理，包括文件结构、存取方式、数据之间联系的实现等；

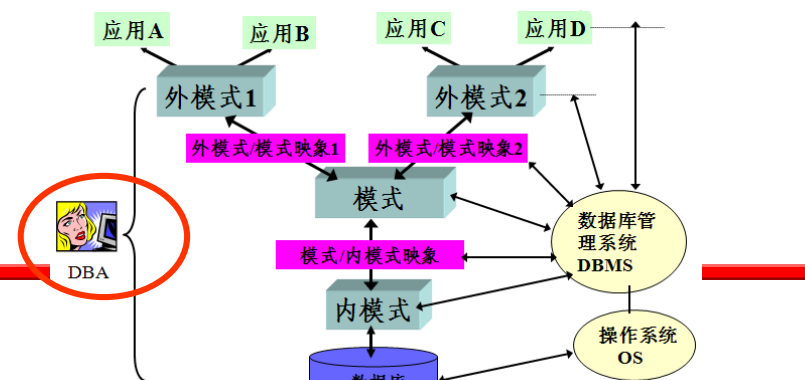
- 数据库的建立和维护功能

- 数据的装入、转换，数据库的转储、恢复、性能监视和分析等。

DBMS的组成

- 语言编译处理程序
- 系统运行控制程序
 - 包括系统总控、存取控制、并发控制、完整性控制、保密性控制、数据存取和更新、通信控制等程序。
- 系统建立和维护程序
 - 数据装入、数据库系统恢复、性能监督、工作日志等程序。
- 数据字典
 - 也称为数据目录或系统目录，由一系列表组成，存储着数据库中有关信息的当前描述，包括数据库的三级模式、用户名表、用户权限等信息。

DBA



- 建库方面：
 - 确定模式、外模式、存储结构、存取策略、负责数据的整理和装入。
- 用库方面：
 - 定义完整性约束条件，规定数据的保密级别、用户权限，监督和控制数据库的运行情况，制定后援和恢复策略，负责故障恢复。
- 改进方面：
 - 监督分析系统的性能（空间利用率，处理效率）；
 - 数据库重组织，物理上重组织，以提高性能；
 - 数据库重构造，设计上较大改动，模式和内模式修改。



本章小结

- 数据库技术的产生与发展，数据管理技术各阶段的特点
- 数据模型：两个层次模型的基本概念
- 数据库系统结构，包括数据独立性概念，三级模式体系结构，DBMS的组成与功能等