

《数据库原理及应用教程（MySQL）》



第1章 数据库系统概述

曹鼎鼎

15100297125

原理先行：内容涵盖数据库系统基本概念、原理、操作、管理、设计、SQL及Python数据库编程等知识
应用落地：全书共有200余实例，电商综合案例贯穿设计全过程，强化实践能力培养
资源丰富：提供慕课、微课、实验等资源，支持混合式教学

中国工信出版集团 人民邮电出版社
POSTS & TELECOM PRESS

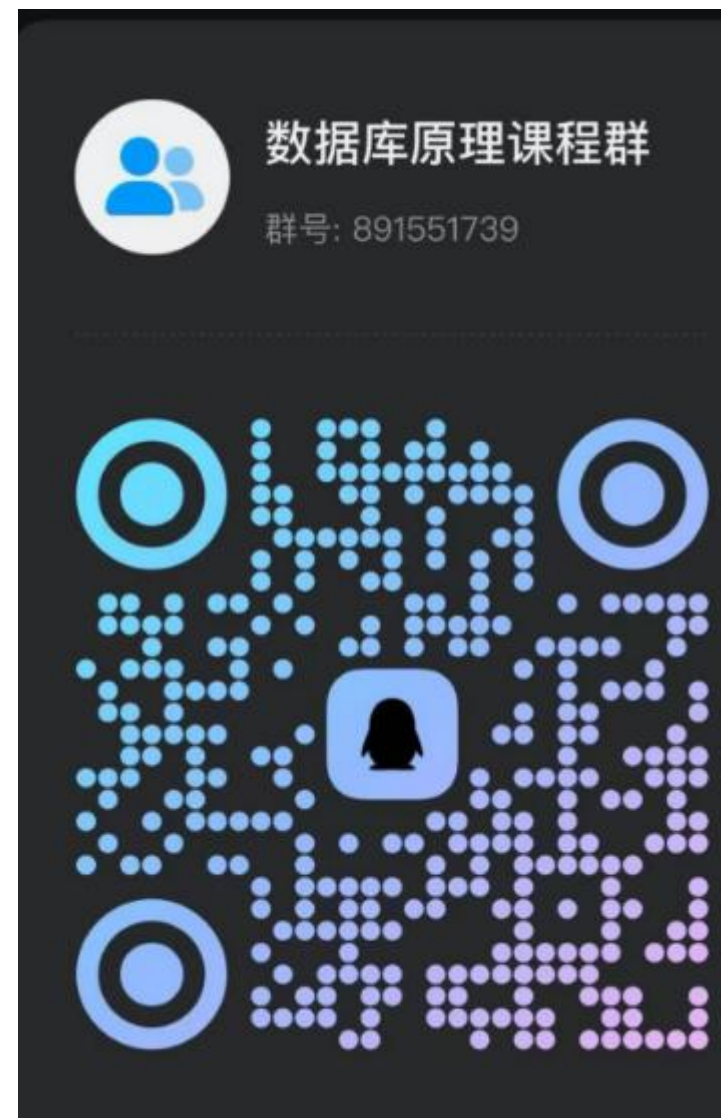
01
OPTION 数据库基本概念和数据库技术发展

02
OPTION 数据库系统的组成

03
OPTION 数据库系统的体系结构

04
OPTION 三个世界及有关概念

05
OPTION 数据库领域新技术



1. 信息、数据、数据处理和数据管理

(1) 信息及其特征

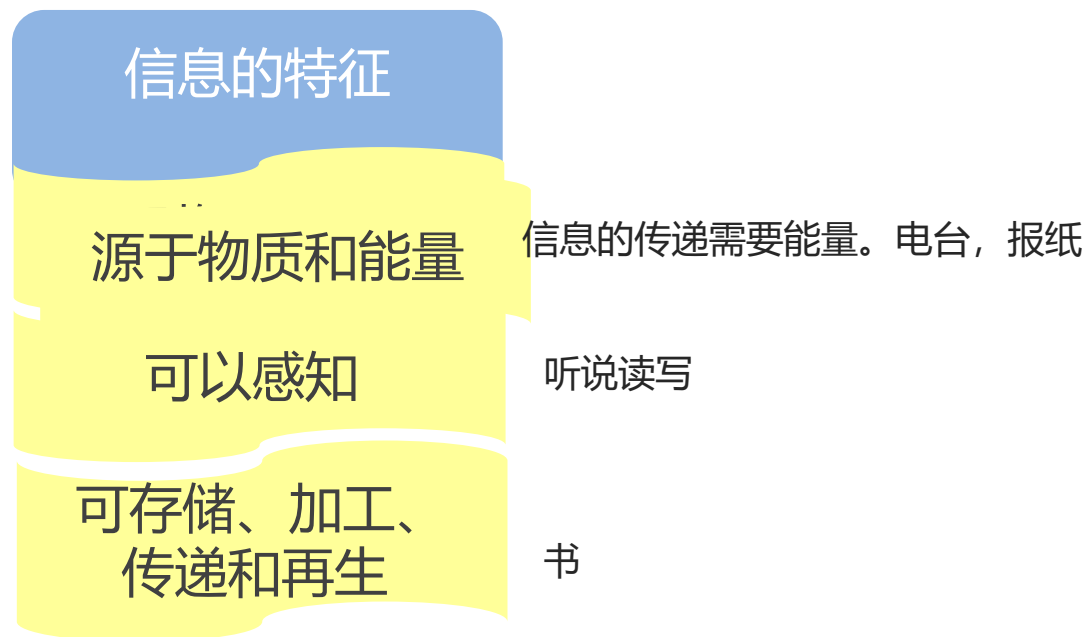
信息是人脑对现实世界事物的存在方式、运动状态以及事物之间联系的抽象反映。

“高等数学”课程的相关信息：

课程号：c7

课程名：高等数学

课 时：60



1. 信息、数据、数据处理和数据管理

(2) 数据及其表现形式

信息的一种，
一部分

- ✓ **数据**是用来记录信息的可识别的符号组合，是信息的具体表现形式。
- ✓ 数据和它的语义是不可分割的，给数据赋予不同的语义，对数据有不同的解释。
- ✓ 当给数据赋予特定语义后，它们就转换为可传递的信息。

有一天，教授在课堂上举起一杯水，说：
“同学们，这杯水的**数据**是：质量200克，体积200毫升。”

学生甲举手问：“那语义是什么呢？”
教授微笑着说：“对科学家来说，它是实验数据；对艺术家来说，它是生命的象征；对哲学家来说，它是存在的隐喻；对口渴的人来说，它是一口水。”

这时学生乙在后排小声嘀咕：“对我来说，它是一杯没写完作业的罚水。”

数据有不同的表现形式，包括数字，还包括文字、图形、图像、声音和视频等，它们都可以经过数字化后存储到计算机中。

如“参加了该考试的人数是500名，考试通过率是80%”，其中的数据“500”和“80%”可改为汉字形式“伍佰”和“百分之八十”，表达的信息是一致的。

1. 信息、数据、数据处理和数据管理

(3) 信息与数据的联系

- ✓ 数据是信息的符号表示
- ✓ 信息是数据的内涵，是对数据的语义解释

如，上例中的数据“500”和“80%”被赋予了特定的语义，此处的“500”表示“考试人数为500”，80%表示“考试通过率是80%”。

信息=数据+语义

1. 信息、数据、数据处理和数据管理

(4) 数据处理和数据管理

数据处理

将数据转换成信息的过程，
包括数据的收集、管理、
加工利用乃至信息输出、
演变推导等

数据管理

数据的分类、组织、编码、
存储、维护、检索等操作

产出
输出结果

原料
输入

信息 = 数据 + 数据处理

“数据处理”的真正含义应该是为了产生信息而处理数据

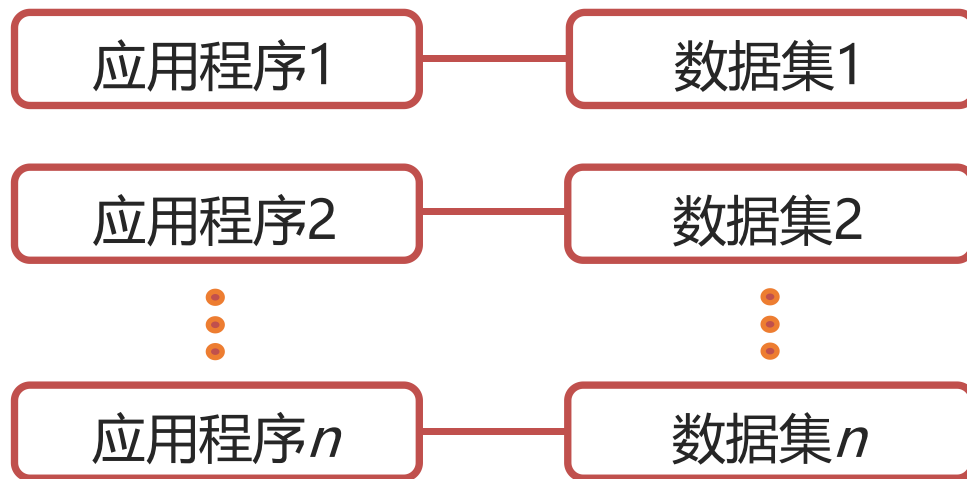
胡乱处理数据什么后果?

2. 数据库技术的发展

(1) 人工管理阶段

- ✓ 数据没有专门的存取设备
- ✓ 数据没有专门的管理软件
- ✓ 数据不共享
- ✓ 数据与应用程序不互相独立性

改了数据格式就要改软件

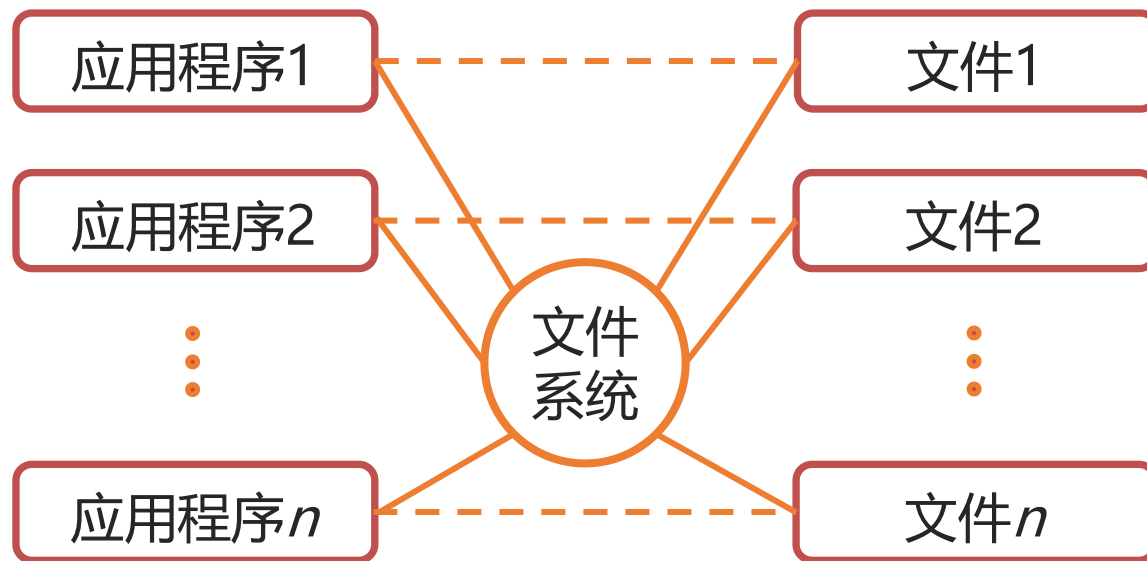


人工管理阶段应用程序与数据之间的对应关系

2. 数据库技术的发展

(2) 文件系统阶段

- ✓ 数据以文件形式长期保存
- ✓ 由文件系统管理数据 和名称打交道, 不用关心
数据储存在哪里
- ✓ 程序与数据间有一定独立性
- ✓ 文件的形式已经多样化 (储存
方式多样) 链表, 栈
- ✓ 数据具有一定的共享性 多个程序用一组数据

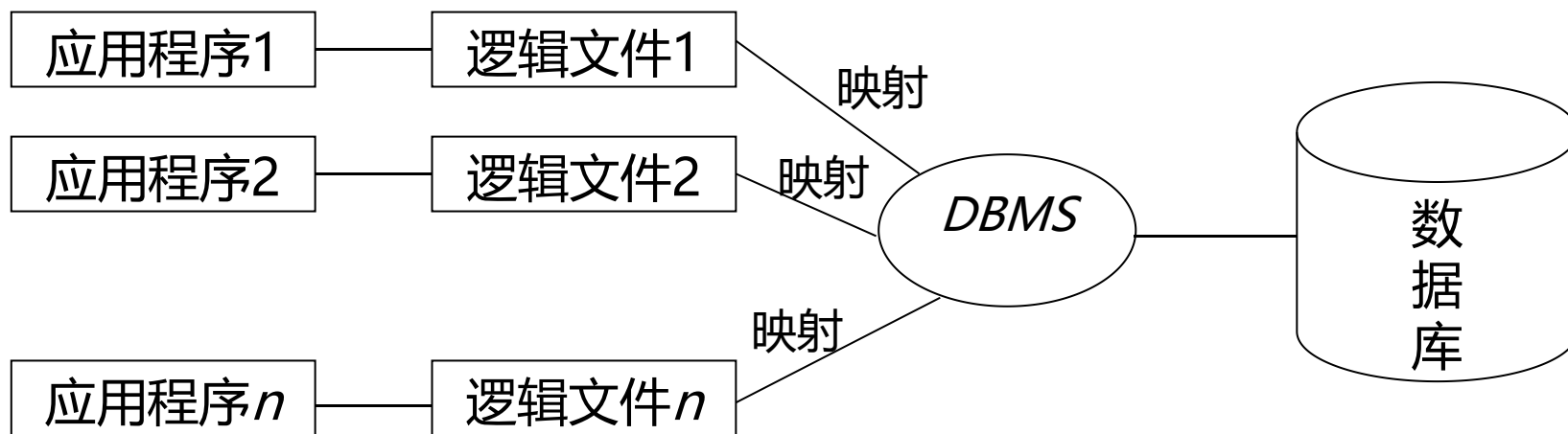


文件系统阶段应用程序与数据之间的对应关系

2. 数据库技术的发展

(3) 数据库系统阶段

- ✓ 结构化的数据及其联系的集合
- ✓ 数据共享性高、冗余度低
- ✓ 数据独立性高（更小的数据元）
- ✓ 有统一的数据管理和控制功能
 - 数据由DBMS统一管理和控制，包括：数据的安全性控制、完整性控制、并发控制、数据恢复



数据库系统阶段应用程序与数据之间的对应关系

(1) 人工管理阶段

例子：纸质档案室

- 想象一下在上世纪 60 年代，你在一家公司做文员。
- 所有客户资料都写在纸上，装进一个个档案袋，再放进柜子里。
- 如果领导要找“张三”的合同，你得跑去柜子翻一堆文件夹，慢慢找。
- **特点：**全靠人记忆和人工操作，效率低、容易出错，一旦文件丢了就很麻烦。

(2) 文件系统阶段

例子：U 盘里的 Excel 文件

- 后来有了计算机，大家把资料放在**文件**里，比如 客户资料.xlsx。
- 要找“张三”，就打开 Excel 用 **Ctrl+F** 搜索，比人工翻纸快多了。
- 但是文件多了之后，你会发现：
 - 数据分散在不同文件里，重复存放。
 - 修改一个地方，可能忘了同步更新其他文件。
 - 想做复杂统计（比如张三买了多少东西），要人工汇总。
- **特点：**有了电子化，但缺乏统一管理，容易产生冗余和不一致。

(3) 数据库系统阶段

例子：银行的客户管理系统

- 银行把所有客户的信息存放在数据库里。
- 你要找“张三”的账户，只要输入身份证号，一秒就能查到。
- 数据不会重复：张三的住址改了，所有业务都立即同步。
- 系统还能做复杂统计：比如“30 岁以下客户平均存款额”。
- **特点：**数据集中管理，有结构化存储、查询和权限控制，效率高、安全可靠。

02

OPTION

数据库系统的组成

03

OPTION

数据库系统的体系结构

04

OPTION

三个世界及有关概念

05

OPTION

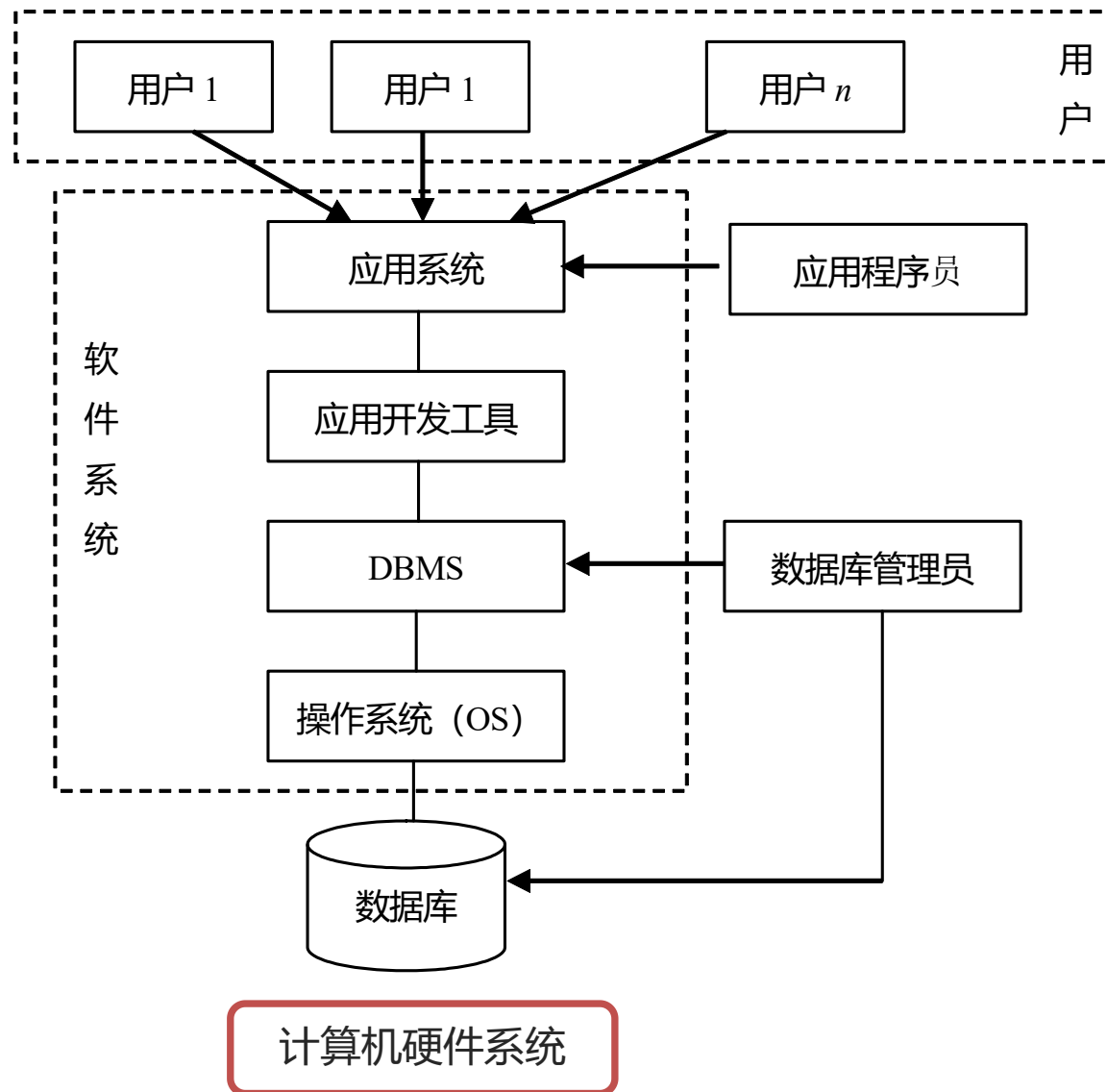
数据库领域新技术

1. 数据库

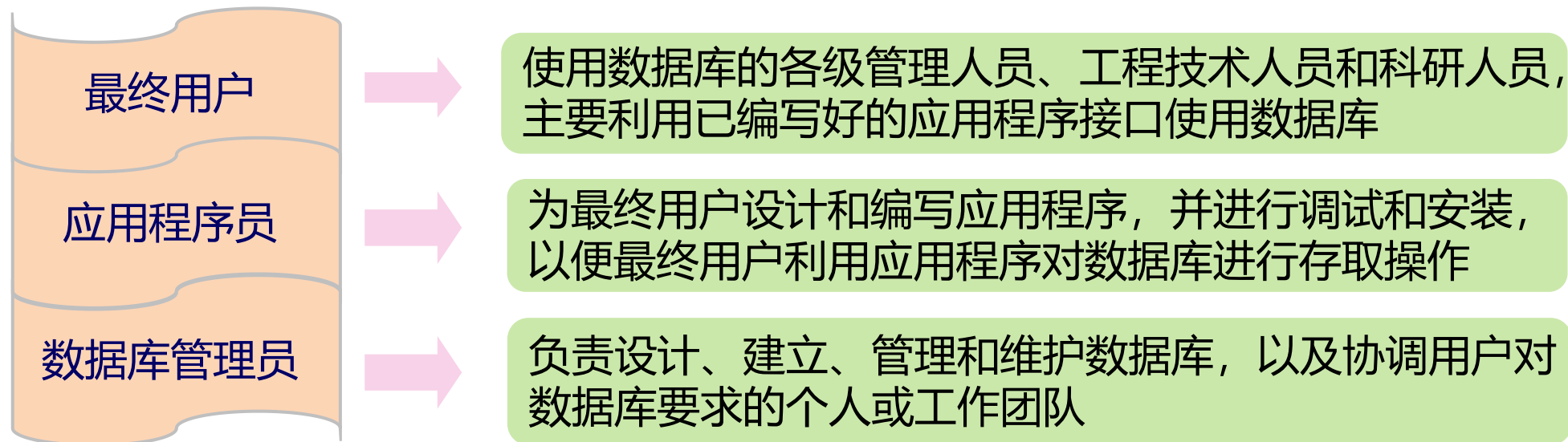
- ✓ 数据库 (DataBase, DB) 是存储在计算机内、有组织的、可共享的数据和数据对象 (如表、视图、存储过程和触发器等) 的集合
- ✓ 这种集合按一定的数据模型 (或结构) 组织、描述并长期存储, 同时能以安全和可靠的方法进行数据的检索和存储

数据库系统 (DataBase System, DBS) 的组成:

- ✓ 数据库
- ✓ 数据库用户
- ✓ 计算机硬件系统
- ✓ 计算机软件系统



2. 数据库用户

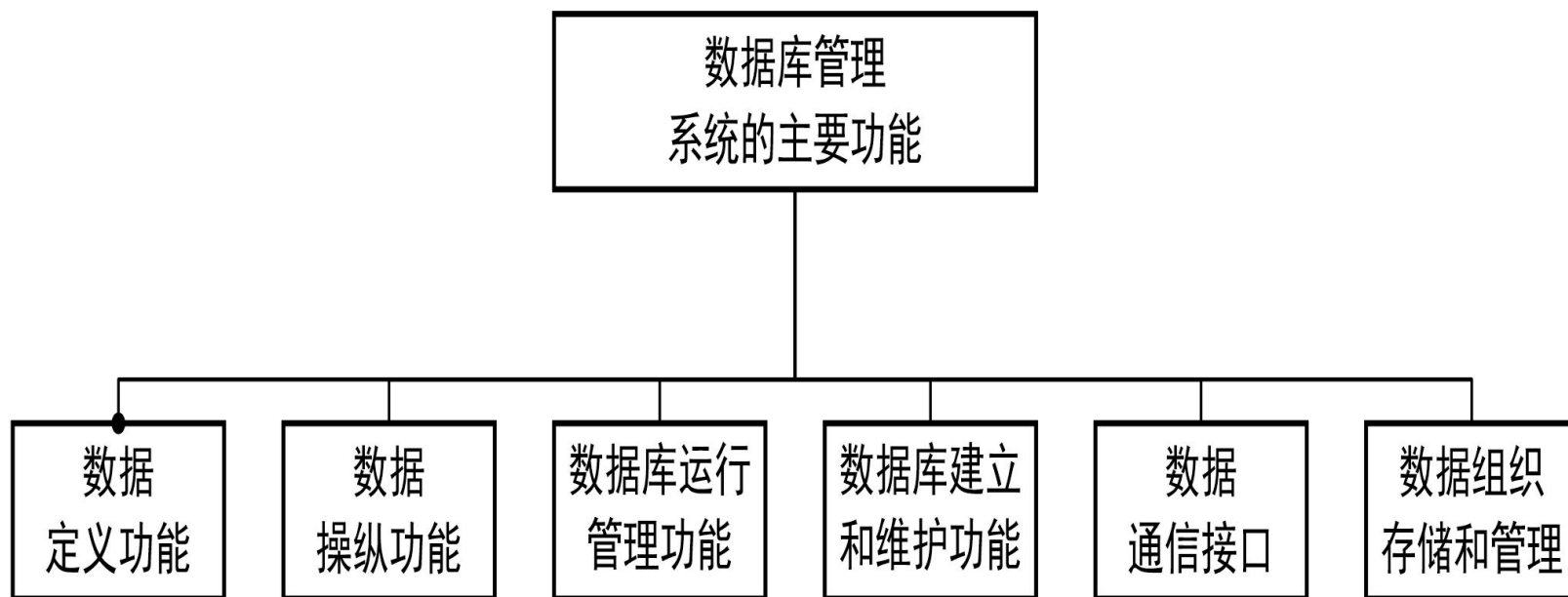


3. 软件系统

- ✓ 软件（Software）系统主要包括操作系统（Operating System, OS）、数据库管理系统（DBMS）及应用开发工具和应用系统等
- ✓ DBMS是数据库系统的核心软件，可借助操作系统对数据库的数据进行存取、维护和管理。数据库系统的各类人员、应用程序等对数据库的各种操作请求，都必须通过DBMS完成

3. 软件系统

数据库管理系统 (DBMS) 的主要功能



3. 软件系统

数据库管理系统 (DBMS) 的主要功能

✓ 数据定义功能

- DBMS提供数据定义语言 (Data Define Language, DDL), 定义数据的模式、外模式和内模式三级模式结构, 定义模式/内模式和外模式/模式二级映像, 定义有关的约束条件
- 例如, 为保证数据库安全而定义用户口令和存取权限, 为保证正确语义而定义完整性规则等
- 再如, DBMS提供的结构化查询语言 (SQL) 提供Create、Drop、Alter等语句可分别用来建立、删除和修改数据库

✓ 数据操纵功能

- DBMS提供数据操纵语言 (Data Manipulation Language, DML) 实现对数据库的基本操作, 包括检索、更新 (包括插入、修改和删除) 等
- 例如, DBMS提供的结构化查询语言SQL提供查询语句 (SELECT)、插入语句 (INSERT)、修改语句 (UPDATE) 和删除语句 (DELETE), 可分别实现对数据库中数据记录的查询、插入、修改和删除等操作

3. 软件系统

数据库管理系统 (DBMS) 的主要功能

✓ 数据库运行管理功能

- 数据的安全性 (Security) 控制: 防止不合法使用数据库造成数据的泄露和破坏, 使每个用户只能按规定对某些数据进行某种或某些操作和处理, 保证数据的安全
- 数据的完整性 (Integrity) 控制: 系统通过设置一些完整性规则等约束条件, 确保数据的正确性、有效性和相容性
- 并发 (Concurrency) 控制: 多个用户同时存取或修改数据库时, 系统可防止由于相互干扰而提供给用户不正确的数据, 并防止数据库受到破坏
- 数据恢复 (Recovery): 由于计算机系统的硬件故障、软件故障、操作员的误操作及其他故意的破坏等原因, 造成数据库中的数据不正确或数据丢失时, 系统有能力将数据库从错误状态恢复到最近某一时刻的正确状态

✓ 数据库的建立和维护功能

- 数据库的建立包括数据库的初始数据的装入与数据转换等, 数据库的维护包括数据库的转储、恢复、重组织与重构造、系统性能监视与分析等。这些功能分别由DBMS的各个实用程序来完成

3. 软件系统

数据库管理系统 (DBMS) 的主要功能

✓ 数据通信接口

- DBMS提供与其他软件系统进行通信的功能
- DBMS提供了与其他DBMS或文件系统的接口，从而使该DBMS能够将数据转换为另一个DBMS或文件系统能够接受的格式，或者可接收其他DBMS或文件系统的数据库，实现用户程序与DBMS、DBMS与DBMS、DBMS与文件系统之间的通信

✓ 数据组织、存储和管理

- DBMS负责对数据库中需要存放的各种数据（如数据字典、用户数据、存取路径等）的组织、存储和管理工作，确定以何种文件结构和存取方式物理地组织这些数据，以提高存储空间利用率和对数据库进行增、删、查、改的效率

3. 软件系统

数据库管理系统 (DBMS) 的DBMS的数据存取过程

- (1) 用户使用某种特定的数据操作语言向DBMS发出存取请求;
- (2) DBMS接受请求并将该请求解释转换成机器代码指令;
- (3) DBMS依次检查外模式、外模式 / 模式映象、模式、模式 / 内模式映象及存储结构定义;
- (4) DBMS对存储数据库执行必要的存取操作;
- (5) 从对数据库的存取操作中接受结果;
- (6) 对得到的结果进行必要的处理, 如格式转换等;
- (7) 将处理的结果返回给用户。

3. 软件系统

🏪 场景：小明去超市买饮料

(1) 用户请求 → 小明下单

小明走进超市，对服务员说：

☞ “我想买一瓶可乐。”

(相当于 用户用 SQL 发出存取请求)

(4) 存取数据库 → 去货架拿可乐

服务员走到货架，找到“可乐”并取下来。

(相当于 DBMS 在存储结构里执行取数操作)

(7) 返回用户 → 小明收到可乐

小明拿到打包好的可乐，开心地喝了一口。

(相当于 DBMS 把处理好的结果返回给用户)

(2) DBMS 接收请求 → 服务员理解需求

服务员听到请求后，心里翻译一下：

☞ “原来他要的是饮料区里的可乐。”

(相当于 DBMS 把 SQL 转换成机器能理解的操作)

(5) 得到结果 → 拿到可乐

服务员手里已经有了一瓶冰镇可乐。

(相当于 DBMS 得到了数据库里的数据)

- SQL 请求 = 小明下单
- 解析请求 = 服务员理解
- 外/模式/内模式 = 确认货架位置
- 存取操作 = 服务员去取货
- 结果处理 = 收银打包
- 返回结果 = 小明拿到可乐

(3) 检查外模式/模式/内模式 → 确认货架位置

服务员开始按超市的规则查：

- “饮料在哪个大类？”（外模式）
 - “可乐在饮料里的哪一类？”（模式）
 - “具体放在第几号货架、第几层？”（内模式）
- (相当于 DBMS 逐层检查映射关系，定位数据在存储中的位置)

(6) 结果处理 → 打包处理

服务员把可乐递到收银台，贴上价签，装到袋子里。

(相当于 DBMS 对结果做处理，如格式转换)

03

OPTION

数据库系统的体系结构

04

OPTION

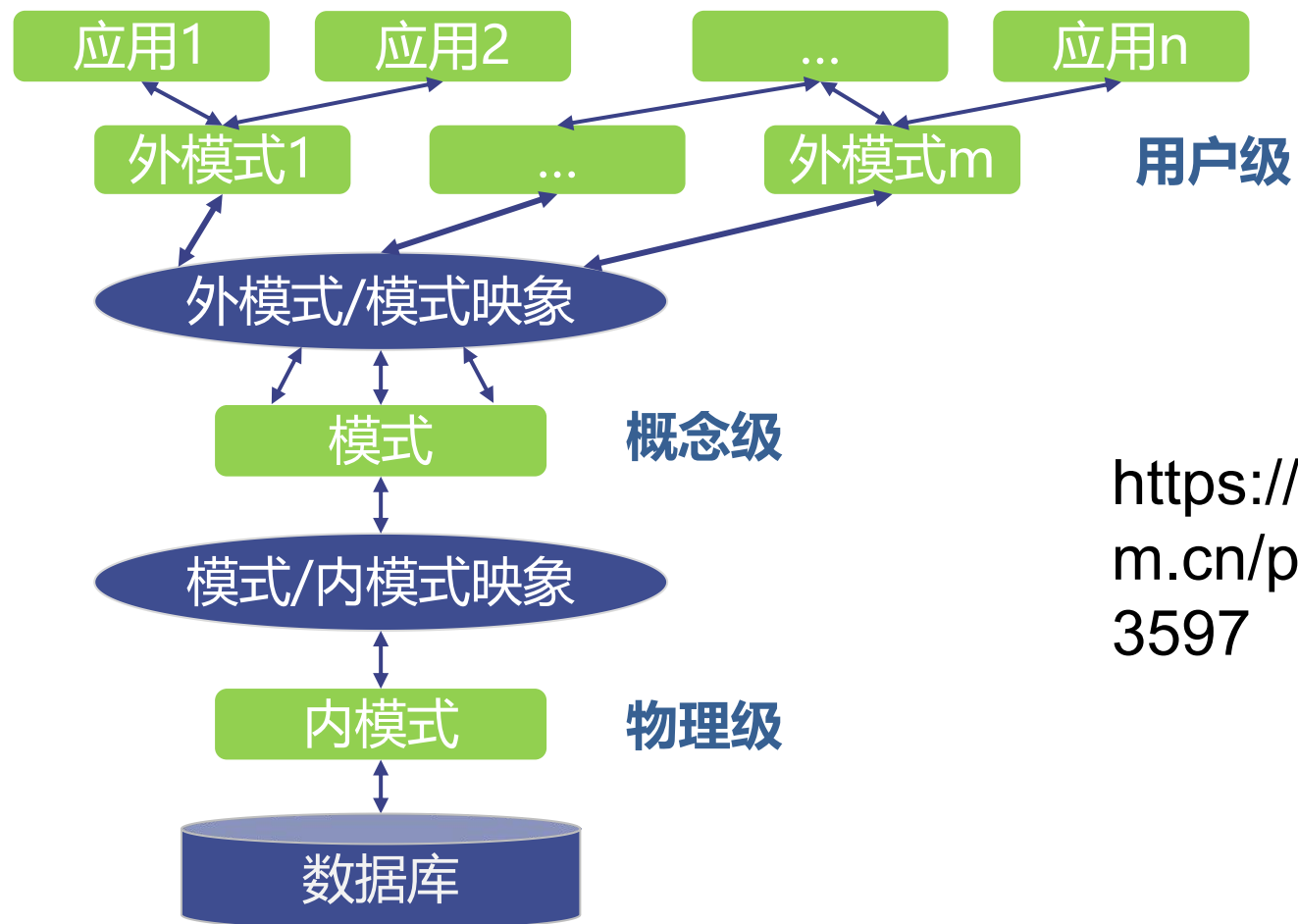
三个世界及有关概念

05

OPTION

数据库领域新技术

1. 数据库系统的内部体系结构

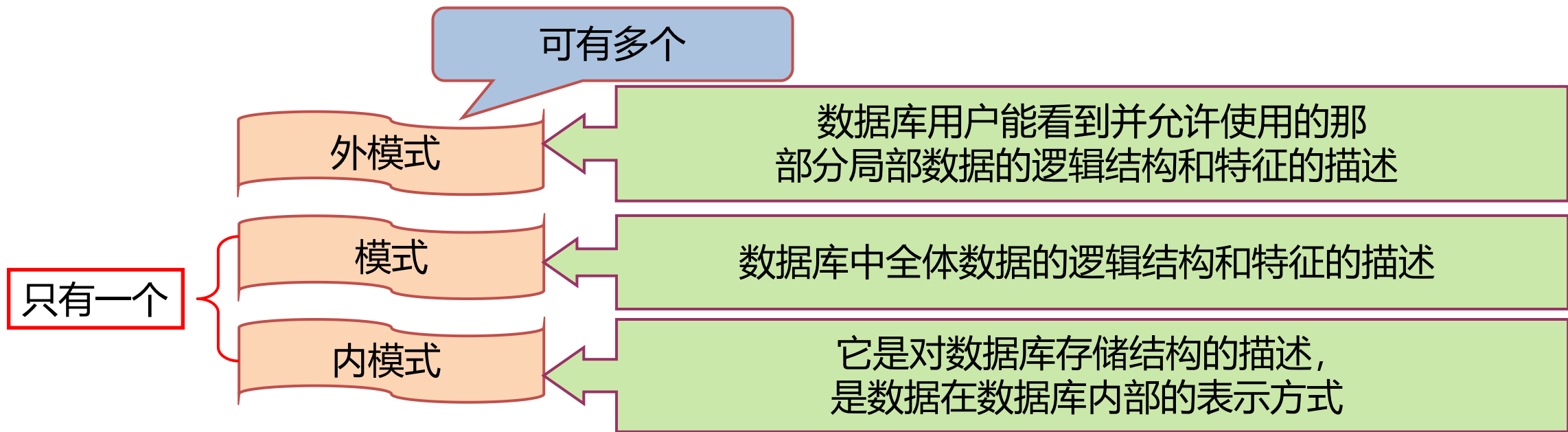


<https://ewm.proedu.com.cn/previewIndex/1133597>

数据库系统的三级模式结构和二级映像功能示意图

1. 数据库系统的内部体系结构

(1) 三级模式结构

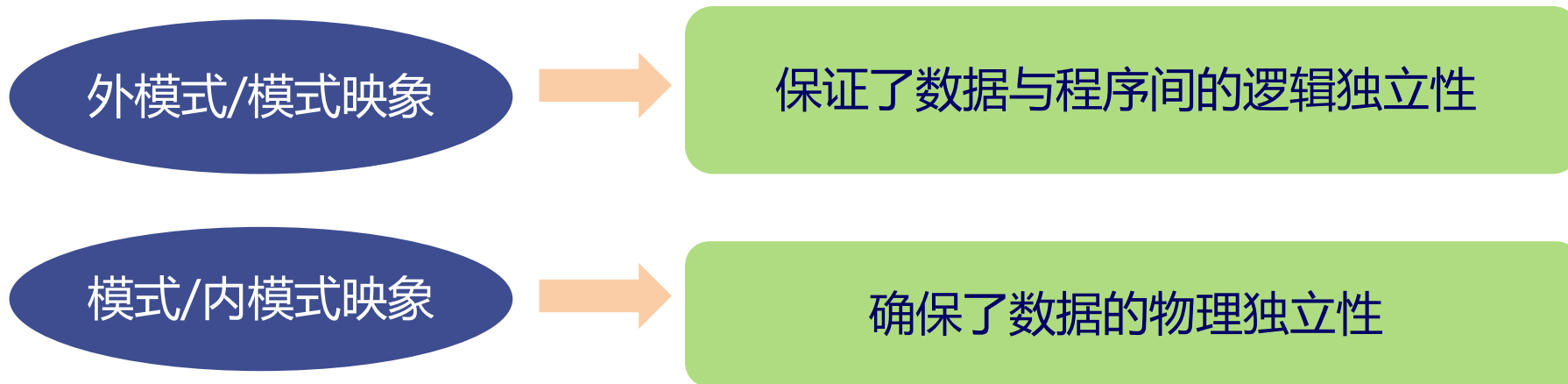


- 内模式是整个数据库实际存储的表示
- 模式是整个数据库实际存储的抽象表示
- 外模式是概念模式的某一部分的抽象表示

1. 数据库系统的内部体系结构

(2) 二级映像

- ✓ DBMS在三级模式之间提供了二级映像功能，保证了数据库系统中的数据能够具有较高的逻辑独立性与物理独立性。



用奶茶店讲数据库

角色设定

- 顾客小明：点单的人（应用/用户）
- 前台点单屏：看到的那一小块数据世界（外模式）
- 总配方手册：店里所有饮品的标准定义与规则（模式/概念模式）
- 后厨仓库与机器：原料怎么摆、机器怎么调、冰块几克（内模式）

二级映像（怎么连起来）

A. 外模式 / 模式 映像

- 把“顾客屏幕上的‘少糖’”翻译成配方里的 `sugar_level = 25`；
- 把“‘去冰’按钮”映射为 `ice = 0g`；
- “小杯/中杯/大杯”映到 `volume = {300, 450, 600}`。
- 作用：连接“各类用户视图 → 统一概念结构”。
- 带来的独立性（逻辑独立）：
 - 如果概念模式改名“`sugar_level → sweetness_ratio`”，只要调整映像，顾客的点单屏不用改。

B. 模式 / 内模式 映像

- 把“配方=抹茶拿铁(450ml, 少糖, 去冰)”翻译成后厨操作：
 - 去 B21 取抹茶 10g、B05 取牛奶 250ml、程序03 12 秒；
 - 数据层面就是“查哪个文件、走哪个索引、读哪几页”。
- 作用：连接“统一概念结构 → 物理存储细节”。
- 带来的独立性（物理独立）：
 - 把珍珠从 B17 挪到 B09、把 HDD 换成 SSD、给“原料名”新建索引，顾客屏幕和配方手册都不用改。

三级模式结构（是什么）

1) 外模式（External Schema）= 前台点单屏/菜单视图

- 小明点单屏上只看到“少糖/去冰/中杯可乐/芝士抹茶”等选项。
- 经理的屏幕会多出“成本价/毛利/库存预警”，外送平台看到的是“外卖价/配送范围”。
- 要点：不同角色看到不同“子视图”；字段可重命名、隐藏、格式化（如把 `sugar_level=25` 显示为“少糖”）。

2) 模式 / 概念模式（Conceptual Schema）= 总配方手册

- 统一定义“饮品”“原料”“规格”“加料”“价格计算规则”“约束（最多两种加料/咖啡因上限）”。
- 要点：全店唯一的逻辑结构与约束（实体-属性-关系-规则）。

3) 内模式（Internal Schema）= 后厨仓储与机器细节

- 珍珠在 B17 号桶、牛奶在冷柜第 3 层；“标准冰=200g”“程序03 = 12 秒高速摇匀”；有的原料建了“快速拿取索引”。
- 要点：物理存储与访问方式（文件/页/记录布局、索引、压缩、分区、SSD/HDD...）。

1. 小明：“芝士绿茶，中杯，少糖，去冰！”（外模式上的操作）
2. 前台把它翻译成：`drink='matcha_tea', size=450, sugar=25, ice=0`（外/模式映像）
3. 后厨据此执行：去 B21/B05 拿料、机器跑程序03、取 0g 冰（模/内映像）
4. 一杯成品上桌（结果返回）。

极简对照表（备课好用）

层次	奶茶店比喻	数据库含义	典型变化	受影响层	
外模式	点单屏/角色菜单	各用户/应用的视图	把“少糖”改成“25%糖”显示	只改外→模映像	
模式	总配方手册	全局逻辑结构与约束	新增“燕麦拿铁”实体/关系	外→模映像可能需更新系	
内模式	仓储布局/机器程序	物理存储与访问方式	换 SSD、建索引、分区	只改模→内映像	

1. 数据库系统的内部体系结构

(3) 三级模式与二级映像的优点

保证数据的独立性 + 

简化了用户接口 + 



 + 有利于数据共享

 + 有利于数据的安全保密

2. 数据库系统的外部体系结构

✓ 从最终用户角度来看，数据库系统分为：

单用户结构

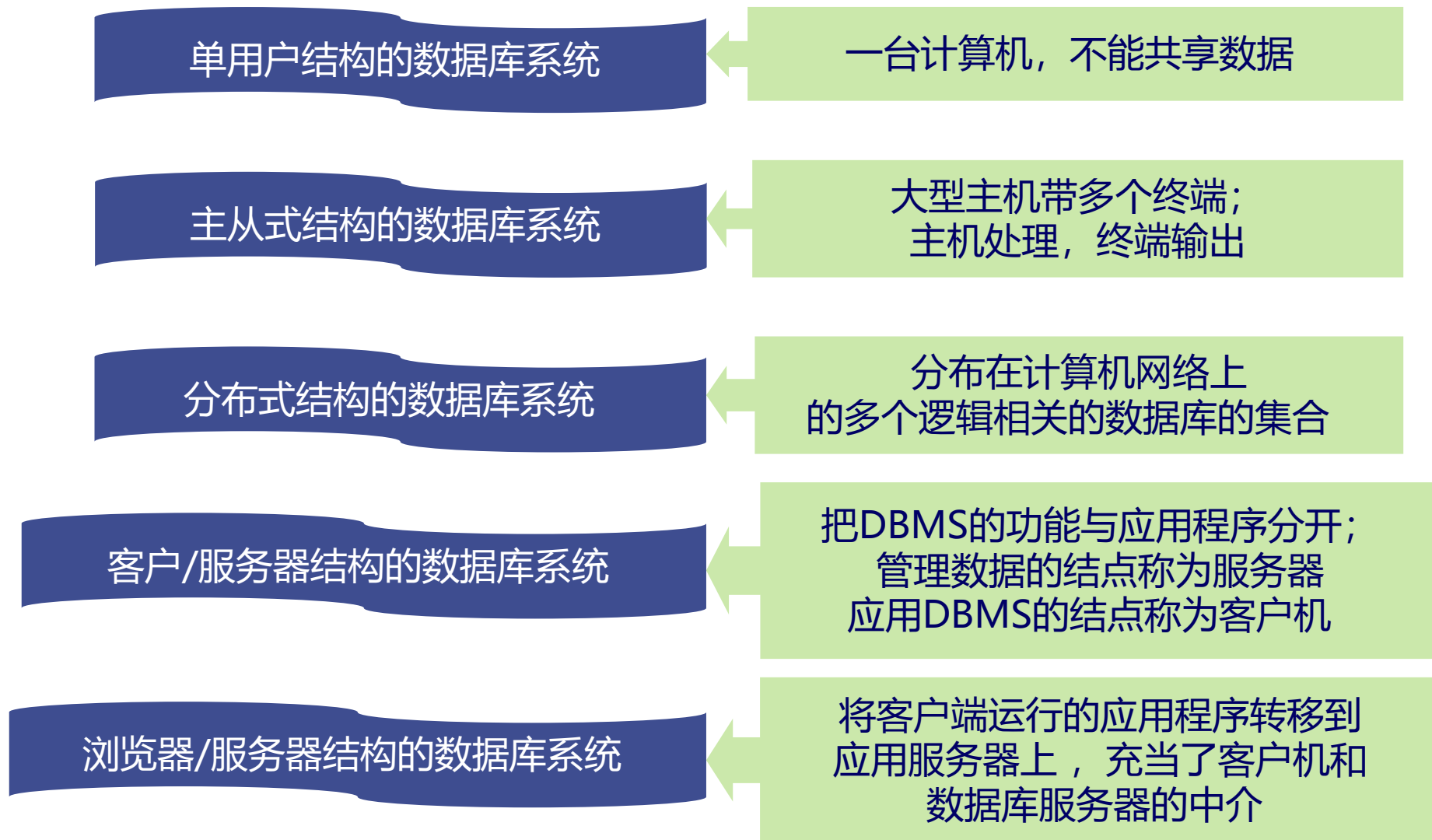
主从式结构

分布式结构

客户/服务器结构

浏览器/服务器结构

2. 数据库系统的外部体系结构



04

OPTION

三个世界及有关概念

05

OPTION

数据库领域新技术

1. 现实世界

- ✓ 现实世界，即客观存在的世界，其中存在着各种事物及它们之间的联系，每个事物都有自己的特征或性质
- ✓ 现实世界中，人们总是选用感兴趣的最能表征一个事物的若干特征来描述该事物
 - 例如，要描述一门课程，常选用课程号、课程名、课时等来描述，有了这些特征，就能区分不同的课程
- ✓ 现实世界中，事物之间是相互联系的
 - 例如，可以选择“教师讲授课程”这一联系表示教师和课程之间的关系

2. 信息世界

- ✓ 是现实世界在人们头脑中的反映，经过人脑的分析、归纳和抽象，形成信息，人们把这些信息进行记录、整理、归类和格式化后，就构成了信息世界
- ✓ 信息世界是对客观事物及其联系的一种抽象描述
 - 例如，学生信息、教师信息、学生选课信息、教师授课信息等

2. 信息世界

信息世界的有关概念

实体

- ✓ 客观存在并且可以相互区别的“事物”称为实体
- ✓ 实体可以是具体的人、事、物，也可以是抽象的事件

属性

- ✓ 实体所具有的某一特性称为属性

型

→ 属性名

值

→ 属性的具体内容

2. 信息世界

信息世界的有关概念

实体型

具有相同属性的实体必然具有共同的特征

- ✓ 用**实体名及其属性名集合**来抽象和描述同类实体，称为**实体型**，
如：**课程（课程号，课程名，课时）**

实体集

同型实体的集合称为**实体集**

- ✓ 如所有的学生、所有的课程等。

2. 信息世界

信息世界的有关概念

码

在实体型中，能唯一标识一个实体的属性或属性集称为实体的码

- ✓ 如学生的学号就是学生实体的码

域

某一属性的取值范围称为该属性的域

- ✓ 如学号的域为6位整数，姓名的域为字符串集合，年龄的域为小于40的整数，性别的域为男或女等

2. 信息世界

信息世界的有关概念

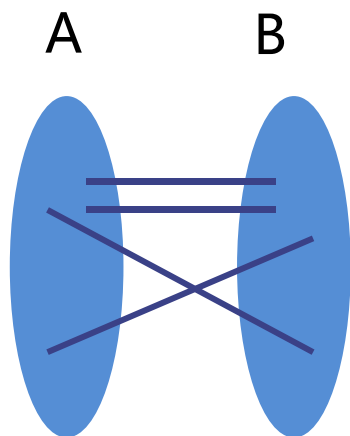
联系

- ✓ 两个实体型之间的联系：两个不同的实体集间的联系
- ✓ 两个以上实体型之间的联系
- ✓ 单个实体型内部的联系：同一个实体集内的各个实体之间的联系

2. 信息世界

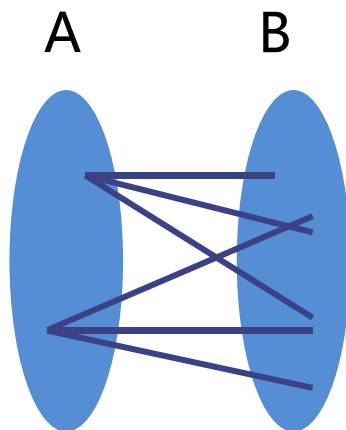
信息世界的有关概念

✓ 两个实体型之间的联系是指两个不同的实体集间的联系，有如下三种类型：



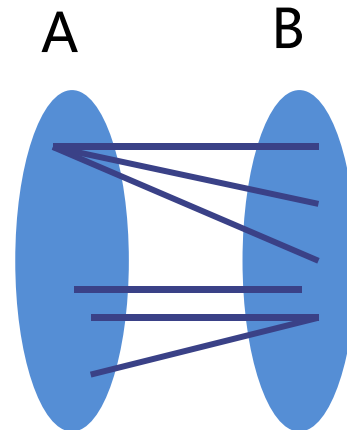
一对一联系

班级与班长
观众与座位
病人与床位



一对多联系

班级与学生
公司与职员
省与市



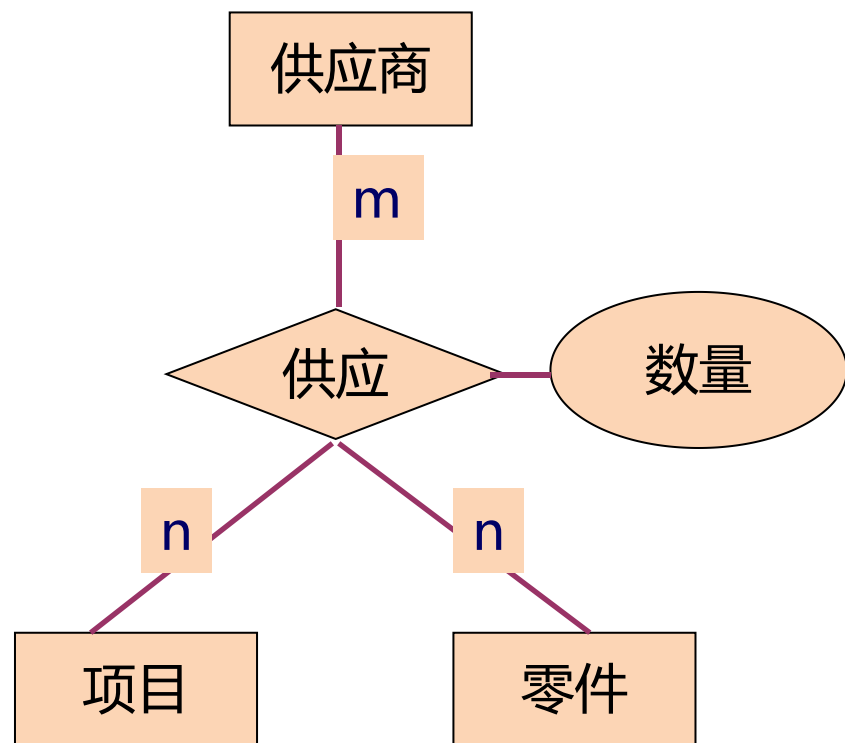
多对多联系

教师与学生
学生与课程
工厂与产品

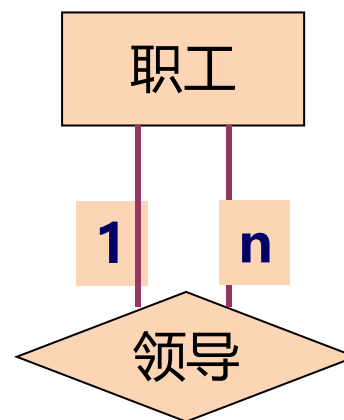
2. 信息世界

信息世界的有关概念

✓ 两个以上实体型间的联系



✓ 单个实体型内部的联系



3. 计算机世界

计算机世界的有关概念

字段

标记实体属性的命名单位称为字段，也称为数据项

✓ 如课程有课程号、课程名和课时等字段

记录

字段的有序集合称为记录

✓ 如一门课程 (c7, 高等数学, 60) 为一个记录

3. 计算机世界

计算机世界的有关概念

文件

同一类记录的集合称为文件

- ✓ 如，所有课程的记录组成了一个课程文件

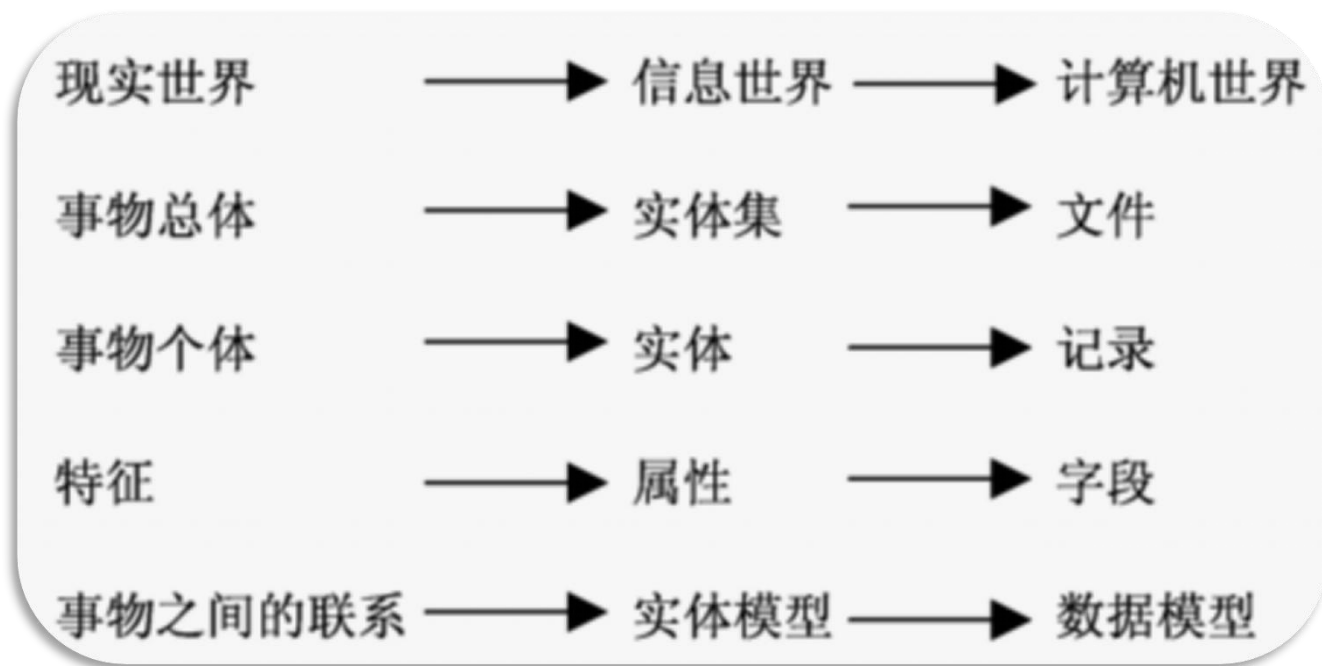
关键字

能唯一标识文件中每个记录的字段或字段集，称为记录的关键字，或简称键（码）

- ✓ 如课程号可作为课程记录的关键字

3. 计算机世界

三个世界中各术语的对应关系

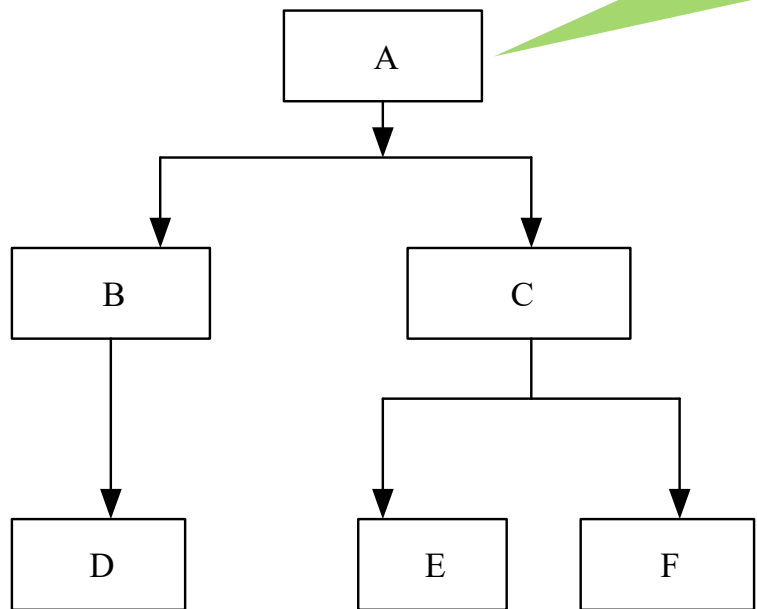


3. 计算机世界

计算机世界的数据库模型

层次模型（最早的数据模型）有向树的示意图

无双亲，根节点



- ✓ 每棵树有且仅有一个结点没有双亲，该结点就是根结点
- ✓ 根结点以外的其他结点有且仅有一个双亲结点
- ✓ 父子结点之间的联系是一对多（1:n）的联系

3. 计算机世界

计算机世界的数据库模型

✓ 层次模型的数据操纵与数据完整性约束

- 进行插入操作时，如果没有相应的双亲结点值就不能插入子女结点值
- 进行删除操作时，如果删除双亲结点值，则相应的子女结点值也被同时删除
- 修改操作时，应修改所有相应的记录，以保证数据的一致性

✓ 层次模型的优缺点

优点

结构简单，层次分明

查询效率高

提供良好的数据完整性支持

缺点

不能直接表示多对多联系

插入和删除数据限制太多

查询子女结点必须通过双亲结点

3. 计算机世界

计算机世界的数据库模型

网状模型的数据结构

- ✓ 网状模型是采用有向图结构表示记录型与记录型之间联系的数据模型

网状模型的特点:

- 有一个以上的结点没有双亲结点
- 允许结点有多个双亲结点
- 允许两个结点之间有多种联系 (复合联系)

3. 计算机世界

计算机世界的数据库模型

✓ 网状模型的数据操纵与完整性约束

- 插入数据时，允许插入尚未确定双亲结点值的子女结点值
- 删除数据时，允许只删除双亲结点值
- 修改数据时，只需更新指定记录即可

✓ 网状模型的优缺点

优点

可表示实体间的多种复杂联系

具有良好的性能和存储效率

缺点

数据结构复杂

数据定义语言、数据操纵语言复杂

用户需要了解网状模型的实现细节

3. 计算机世界

计算机世界的数据库模型

✓ 关系模型的数据结构

- 规范化的二维表，由表名、表头和表体三部分构成
- 表名即二维表的名称，表头决定了二维表的结构（即表中列数及每列的列名、类型等）
- 表体即二维表中的数据。每个二维表又可称为关系
- 关系模型与层次模型、网状模型不同，它是建立在严格的数学概念之上的，严格的定义将在第2章给出
- 教材中表1-1~表1-5所示为教学数据库teaching的关系模型及其实例，分别为教师关系t、学生关系s、课程关系c、选课关系sc和授课关系tc

3. 计算机世界

计算机世界的数据库模型

✓ 关系模型的数据操纵与完整性约束

- 关系模型中的数据操作是集合操作，操作对象和操作结果都是关系，即若干元组的集合
- 关系模型把对数据的存取路径隐蔽起来，用户只要指出“干什么”，而不必详细说明“怎么干”，从而大大地提高了数据的独立性，提高了用户操作效率

✓ 关系模型的优缺点

优点

有严格的数学理论根据

用关系描述实体间的联系

具有更高的数据独立性、
更好的安全保密性

缺点

查询效率不如非关系模型

3. 计算机世界

计算机世界的数据模型

✓ 面向对象模型

- 对象是现实世界中实体的模型化，如一个学生、一门课程等都可以看作对象
- 每个对象都包含属性和方法
- 属性用来描述对象的静态特征，方法用以描述对象的行为特性
 - ▶ 如一辆机动车，它不仅具有描述其静态特征的属性：高度、重量等，还具有加速、减速等动态特征
- 在面向对象模型中，可以继承操作形成新的类，新的类是对已有的类定义的扩充和细化，从而形成了一种类间的层次结构，有了超类和子类的概念

3. 计算机世界

计算机世界的数据库模型

✓ 面向对象模型的优缺点

优点

能完整地描述现实世界的数据结构

具有丰富的表达能力

缺点

模型相对比较复杂

涉及的知识比较多

05

OPTION

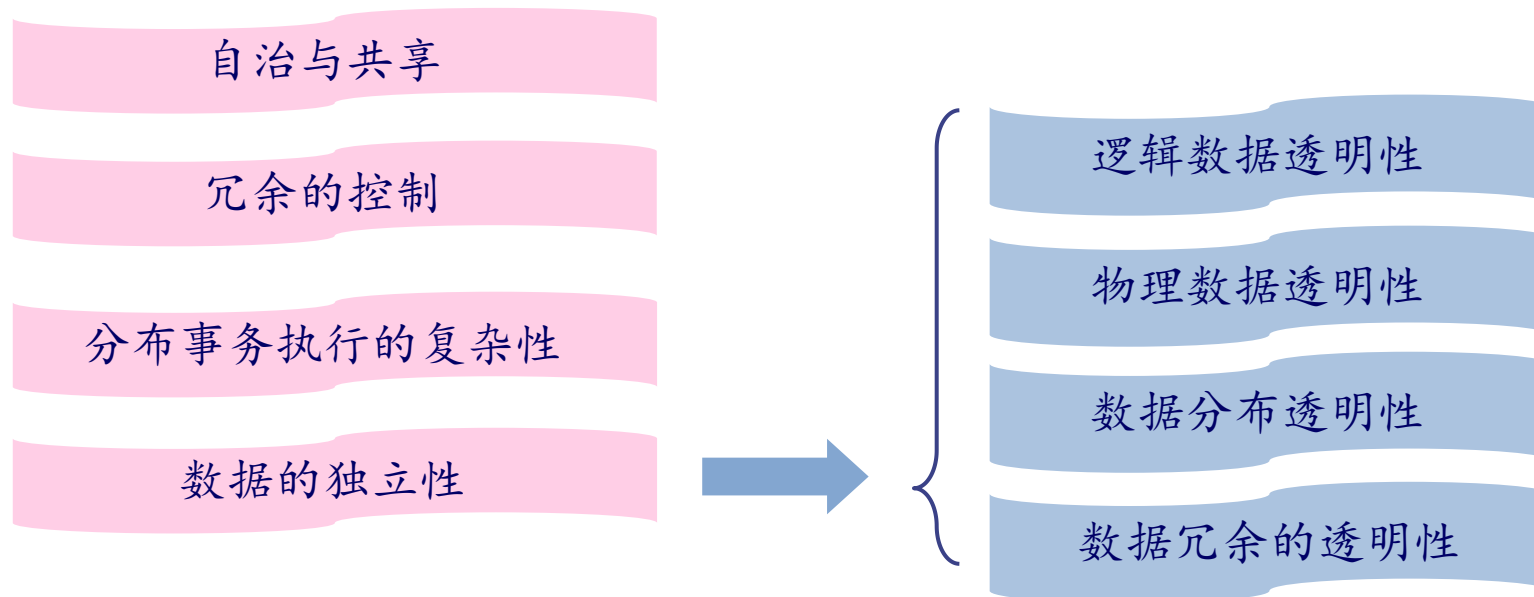
数据库领域新技术

1. 分布式数据库

✓ 分布式数据库的定义

- 分布式数据库是一组结构化的数据集合，它们在逻辑上属于同一系统，而在物理上分布在计算机网络的不同结点上

✓ 分布式数据库的特点



2. 数据仓库与数据挖掘

✓ 数据仓库的定义

- 数据仓库就是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合，通常用于辅助决策支持

✓ 数据仓库的体系结构

- 对多个异构数据源的有效集成，集成后按照主题进行重组，包含历史数据；存放在数据仓库中的数据通常不再修改，用于做进一步的分析型数据处理

✓ 数据挖掘的定义

- 从大量数据中获取有效的、新颖的、潜在有用的、最终可理解的模式非平凡过程，又被称为数据库中的知识发现 (Knowledge Discovery in Database, KDD)

✓ 数据挖掘的方法

- 直接数据挖掘：利用可用的数据建立一个模型，这个模型对剩余的数据，比如对一个特定的变量进行描述，直接数据挖掘包括分类 (Classification)、估值 (Estimation) 和预言 (Prediction) 等分析方法
- 间接数据挖掘：在所有的变量中建立起某种关系，如相关性分组或关联规则 (Affinity Grouping or Association Rules)、聚集 (Clustering)、描述和可视化 (Description and Visualization) 及复杂数据类型挖掘 (文本、网页、图形图像、音视频和空间数据等)

3. 大数据技术

✓ 大数据的概念

- 人们普遍采用大数据的4V特性来描述大数据，即“数据量大 (Volume)” “数据类型繁多 (Variety)” “数据处理速度快 (Velocity)” 和“数据价值密度低 (Value)”

✓ 大数据的关键技术

- 大数据所涉及的关键技术主要包括数据的采集和迁移、数据的存储和管理、数据的处理和分析、数据安全和隐私保护

✓ 大数据技术的应用场景

- 大数据技术的应用已经非常普遍，涉及的领域包括传统零售业、金融业、医疗业和政府机构等

1. 数据库系统 (DBS) 的组成最准确的是 ()。

A.DBMS = DBS

B.DBMS = DB + DBMS

C.DBMS = 数据库 (DB) + 数据库管理系统 (DBMS) + 硬件/操作系统 + 应用程序 + 数据库管理员与用户等人员 + 管理制度与方法

D.DB = 与数据库相关的所有软硬件

1. 数据库系统 (DBS) 的组成最准确的是 ()。

A.DBMS = DBS

B.DBMS = DB + DBMS

C.DBMS = 数据库 (DB) + 数据库管理系统 (DBMS) + 硬件/操作系统 + 应用程序 + 数据库管理员与用户等人员 + 管理制度与方法

D.DB = 与数据库相关的所有软硬件

答案：C

解析：DBS 是“人一机—数据”的综合体：除 DB 与 DBMS 外，还包含运行环境（硬件、OS、网络）、应用程序以及 DBA/开发/用户 等人员与管理制度。A、B 将 DBS 简化为两件软件；D 把 DB 扩大成所有软硬件，概念错误。

2. 一个数据模型通常由三部分构成，它们是（ ）。

- A. 数据结构、数据操作、完整性约束
- B. 实体、属性、联系
- C. 逻辑结构、物理结构、外部结构
- D. 安全策略、备份策略、恢复策略

2. 一个数据模型通常由三部分构成，它们是（ ）。

- A. 数据结构、数据操作、完整性约束
- B. 实体、属性、联系
- C. 逻辑结构、物理结构、外部结构
- D. 安全策略、备份策略、恢复策略

答案：A

解析：数据结构说明“存什么与如何关联”，数据操作规定“能做什么及语义”，完整性约束保证“数据正确”。B 是概念模型要素而非模型三要素；C 将三级模式误作模型组成；D 属数据库管理而非数据模型。

3. 下列选项按从外到内正确列出数据库三级模式的是（ ）。

- A. 模式—外模式—内模式
- B. 外模式—模式—内模式
- C. 内模式—模式—外模式
- D. 外模式—内模式—模式

3. 下列选项按从外到内正确列出数据库三级模式的是（ ）。

- A. 模式—外模式—内模式
- B. 外模式—模式—内模式
- C. 内模式—模式—外模式
- D. 外模式—内模式—模式

答案：B

解析：三级模式自外向内依次为 **外模式（用户视图）—模式/概念模式（全局逻辑）—内模式（物理存储）**。该分层实现“对上抽象、对下屏蔽”。

4. 在二级映像与数据独立性关系中，逻辑独立性主要依赖于（ ）。

- A. 模式/内模式映像
- B. 外模式/模式映像
- C. 外模式/内模式映像
- D. 外模式/外模式映像

4. 在二级映像与数据独立性关系中，逻辑独立性主要依赖于（ ）。

- A. 模式/内模式映像
- B. 外模式/模式映像
- C. 外模式/内模式映像
- D. 外模式/外模式映像

答案：B

解析：修改概念模式（如增加属性、拆分关系）不应迫使应用程序改写，靠 **外模式/模式映像** 将变化映射到原有外模式；而 **A** 保证的是 **物理独立性**（屏蔽存储变化）。

5. 某表由“堆表”改为“按时间戳聚簇索引”，不改变表结构与视图，体现的独立性是（ ）。

- A.逻辑独立性
- B.物理独立性
- C.表示独立性
- D.安全独立性

5. 某表由“堆表”改为“按时间戳聚簇索引”，不改变表结构与视图，体现的独立性是（ ）。

- A.逻辑独立性
- B.物理独立性
- C.表示独立性
- D.安全独立性

答案：B

解析：仅发生在物理层（页组织、索引、文件布局）的调整，通过模式/内模式映像屏蔽对上层的影响，属于物理独立性。A 涉及概念层结构变化；C、D与题意无关。

7. 下列哪项不属于数据库管理员 (DBA) 的典型职责 ()。

- A. 定义与维护模式、存储结构与安全策略
- B. 制定备份恢复方案、监控与性能调优
- C. 并发控制策略配置与权限管理
- D. 编译器的词法/语法优化

7. 下列哪项不属于数据库管理员 (DBA) 的典型职责 ()。

- A. 定义与维护模式、存储结构与安全策略
- B. 制定备份恢复方案、监控与性能调优
- C. 并发控制策略配置与权限管理
- D. 编译器的词法/语法优化

答案：D

解析：D 属编译器/语言实现范畴；A、B、C 均为 DBA 的核心工作（模式管理、安全与权限、备份恢复、性能、并发等）。

9. 关于数据抽象层次，属于物理层的是（ ）。

- A. 课程—学生的多对多联系
- B. 视图 V_PublicStudent 的可见列集合
- C. B+树索引的页大小与节点分裂策略
- D. 概念模式中的主键与外键约束

9. 关于数据抽象层次，属于物理层的是（ ）。

- A. 课程—学生的多对多联系
- B. 视图 V_PublicStudent 的可见列集合
- C. B+ 树索引的页大小与节点分裂策略
- D. 概念模式中的主键与外键约束

答案：C

解析：物理层关注 **存储实现细节**（文件、页、块、索引实现等）。A、D 属概念/逻辑层；B 属外部层（用户视图）。

10. 某数字媒体平台把“作品(Work)”表拆分为“作品元数据(WorkMeta)”与“发布信息(PublishInfo)”，应用的查询视图保持不变。该变更主要考查()。

- A.物理独立性
- B.逻辑独立性
- C.事务隔离性
- D.可串行化

10. 某数字媒体平台把“作品(Work)”表拆分为“作品元数据(WorkMeta)”与“发布信息(PublishInfo)”，应用的查询视图保持不变。该变更主要考查()。

- A.物理独立性
- B.逻辑独立性
- C.事务隔离性
- D.可串行化

答案：B

解析：拆分概念模式的关系是逻辑层变化；若通过外模式/模式映像保持原有视图与程序不变，即体现逻辑独立性。A是存储层变化；C、D为并发控制术语。

11. 在某校“6人间”宿舍建模中，若假定同一时刻每位学生唯一入住一个宿舍房间，而每个房间可住多名学生，则“宿舍房间-学生”的联系类型是（ ）。

A.1:1

B.1:n (房间→学生)

C.n:1 (房间←学生)

D.m:n

11. 在某校“6人间”宿舍建模中，若假定同一时刻每位学生唯一入住一个宿舍房间，而每个房间可住多名学生，则“宿舍房间-学生”的联系类型是（ ）。

A.1:1

B.1:n (房间→学生)

C.n:1 (房间←学生)

D.m:n

答案：B

解析：“一个房间对应多名学生”，且“每个学生当前只对应一个房间”，因此 房间：学生 = 1:n。若允许学生同时住多房或一房只住一人，才会变更为 m:n 或 1:1。