



数据库系统原理

别荣芳

实验助教: 王昊笛 杨晓一 蕾



数据库系统原理

第一次面授

课程内容：

- 对于网课及实验内容进行答疑。
- 梳理第一、二章网课知识点。
- 课堂习题小测及讲解。



第一章 绪论

1.1 数据库

- **数据库：**包含关于某单位、机构、部门，或是某领域、业务、主题，或者某对象的信息、互相关联的大量数据的集合。

(例：高等学校数据库、应急中心数据库、社交网络数据库、电子商务平台数据库、云平台检测管理系统、通信数据库、银行数据库、企业信息数据库、交通数据库)

本课中特指用专门通用软件管理，长期储存在计算机内、有组织、可共享的大量数据的集合。

- **数据类型：**数值、文本、图形、图像、音频、视频等。
- **数据库的形式：**打印或手写、电子文件存储 (txt、word、excel) 、软件系统管理 (PostgreSQL、DB2、Oracle) 。

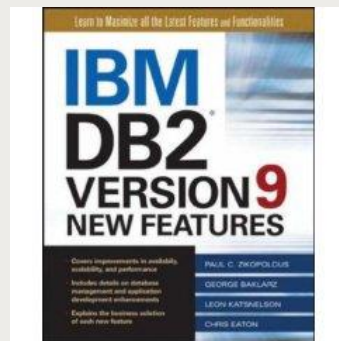


第一章 绪论

1.1 数据库

- **数据库管理系统 (DBMS) : Database Management System**, 数据库管理系统是位于用户与操作系统之间的、专门的、通用的管理数据库的大型复杂系统软件。

著名的DBMS: PostgreSQL、IBM DB2、Oracle、Microsoft SQL Server。





第一章 绪论

1.2 数据库管理系统

- **数据管理系统的目标：**简单、安全、高效地管理数据。
- **数据管理系统的基本功能：**允许用户逻辑地使用数据，而无需关注这些数据在计算机中是如何存放、如何处理的。
- **数据模型：**数据结构和语义的概括（层次模型、关系模型等）。
- **数据库模式：**面向特定数据模型、针对特定应用的数据库结构（如关系数据库中的关系模式或表模式）。
- **实例：**特定数据库中特定时刻存储的一个数据的集合。实例是模式的一个具体值，一个模式可以有多个实例。
- **E-R模型：**用E-R图表示。

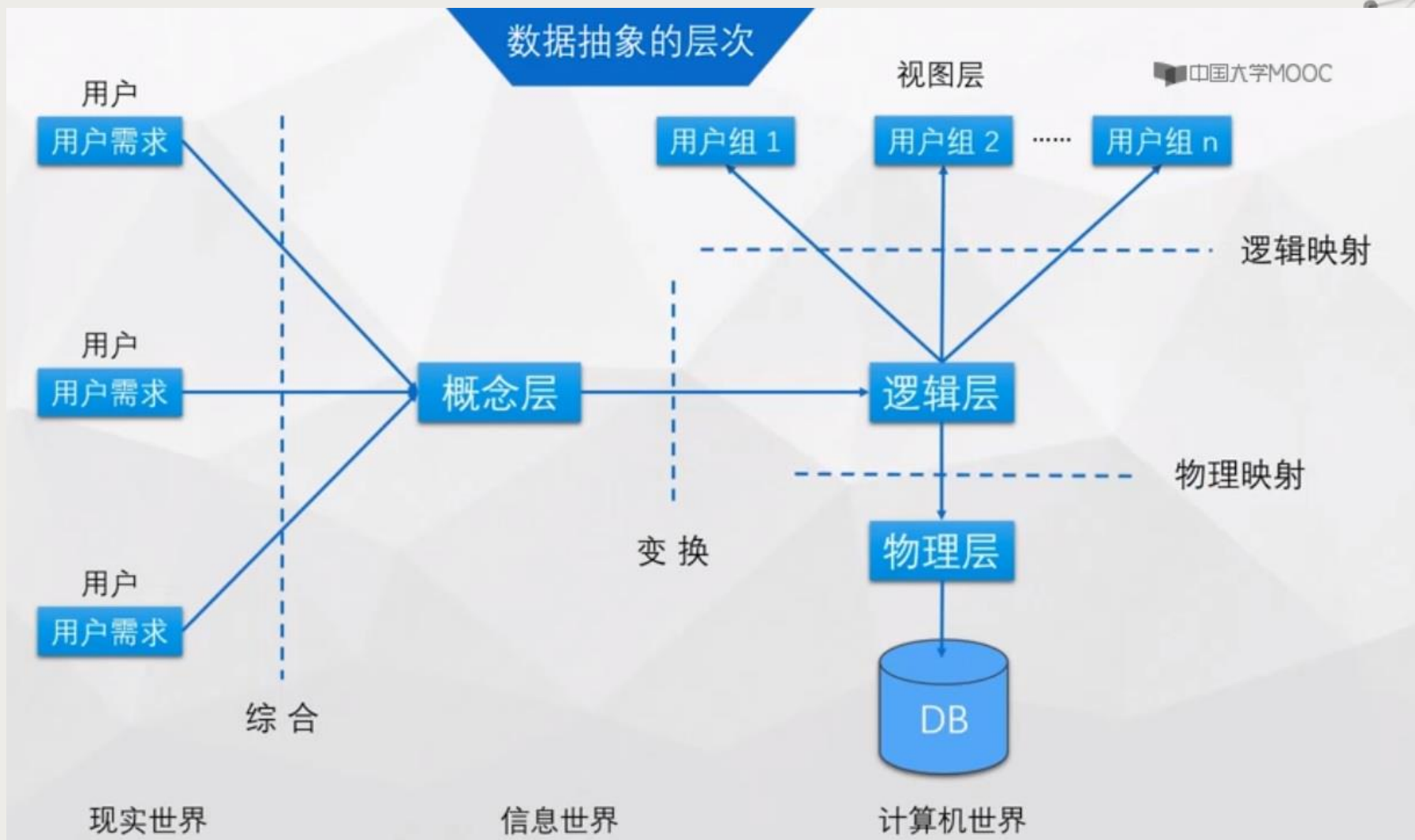


第一章 绪论

1.2 数据库管理系统

• 四层抽象:

- 概念层
- 物理层
- 逻辑层
- 视图层





第一章 绪论

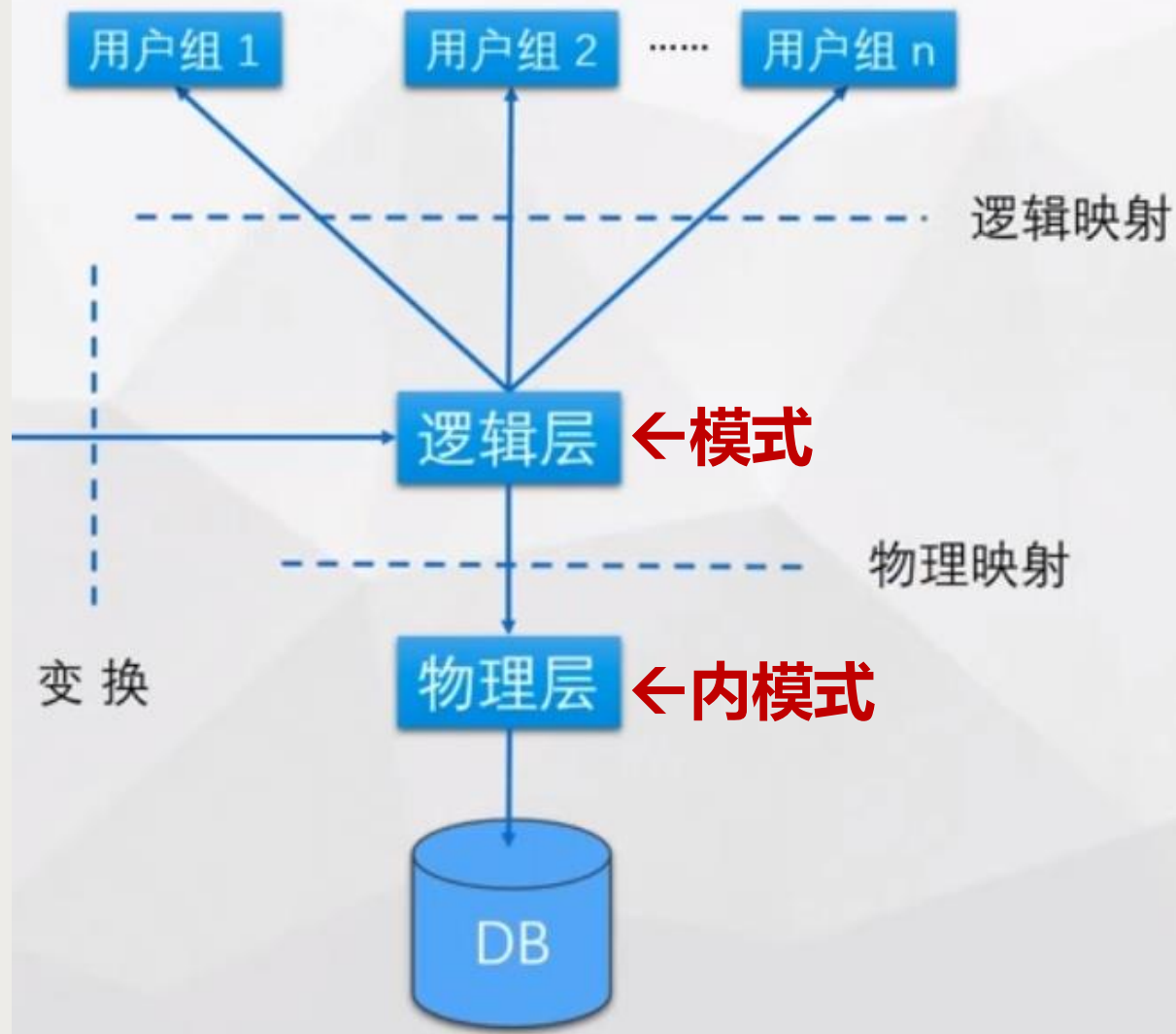
1.2 数据库管理系统

- **三层模式：**
 - 内模式：物理层的数据模式。
 - 模式：逻辑层的数据模式。
 - 外模式：视图层的数据模式。
- **两级映射：**
 - 物理映射（模式/内模式映射）：是模式与内模式之间的对应关系。
 - 逻辑映射（外模式/模式映射）：是外模式与模式之间的对应关系。

象的层次

视图层 ← 外模式

中国大学MOOC





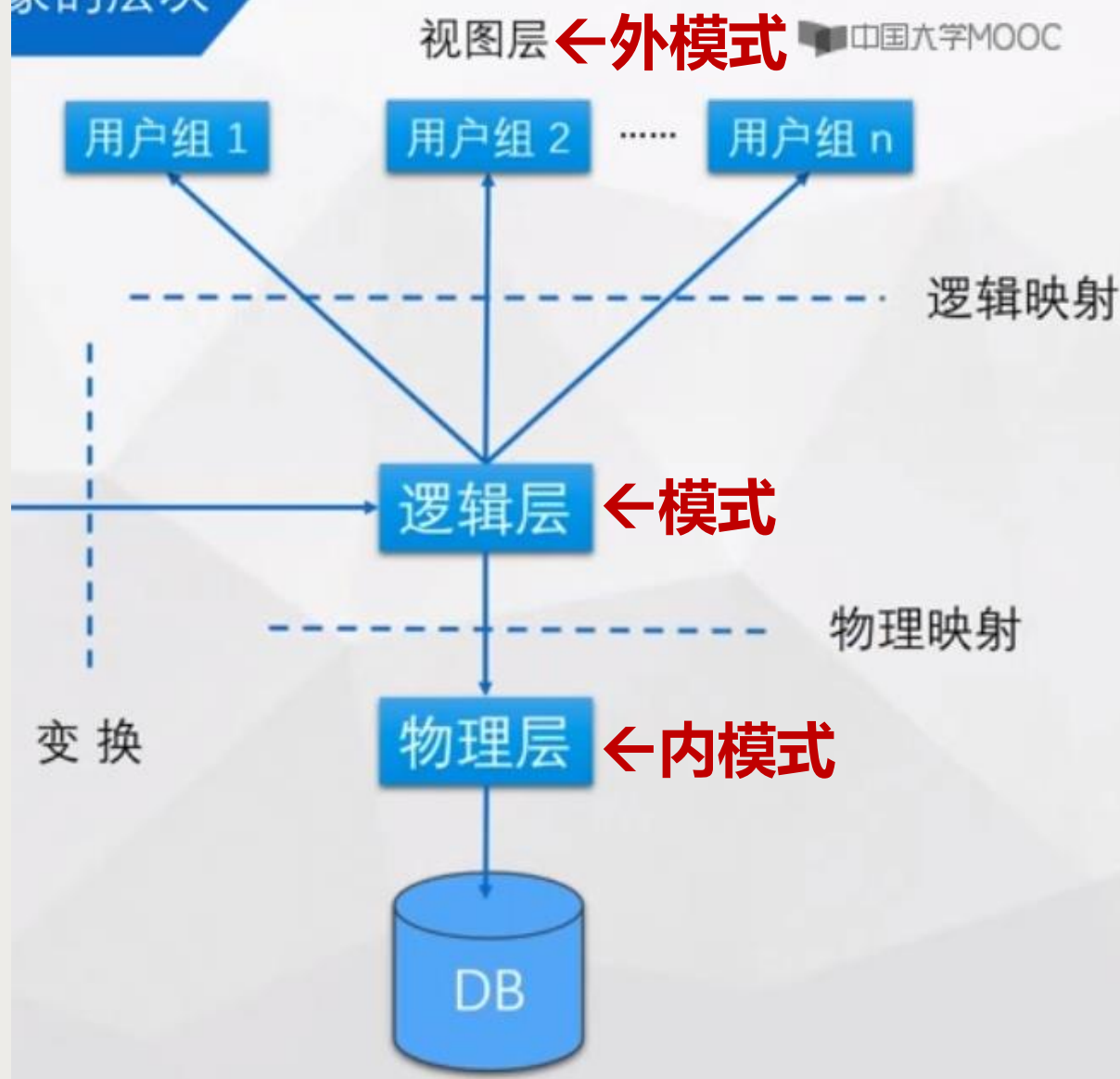
第一章 绪论

1.2 数据库管理系统

- **两个独立性：**

- **逻辑独立性：**当模式需要改变时，可保持外模式不变（用户视图不变），而外模式/模式映射（逻辑映射）进行改变。
- **物理独立性：**当内模式需要改变时，可保持模式不变（逻辑层不变），而对模式/内模式映射（物理映射）进行改变。

象的层次





第一章 绪论

1.2 数据库管理系统

- **数据管理系统模块：查询处理模块、存储管理模块、保护管理模块。**

- **查询处理模块：DDL/DPL解释器、DML编译器、查询执行引擎**

数据定义语言（DDL）、数据保护语言（DPL）、数据操作语言（DML，查插删更）

- **存储管理模块：文件管理器、缓存管理器**

- **保护管理模块：权限管理器、完整性管理器、事务管理器**

数据库同时存储数据文件、数据字典（存放元数据）、数据索引、统计信息、日志。

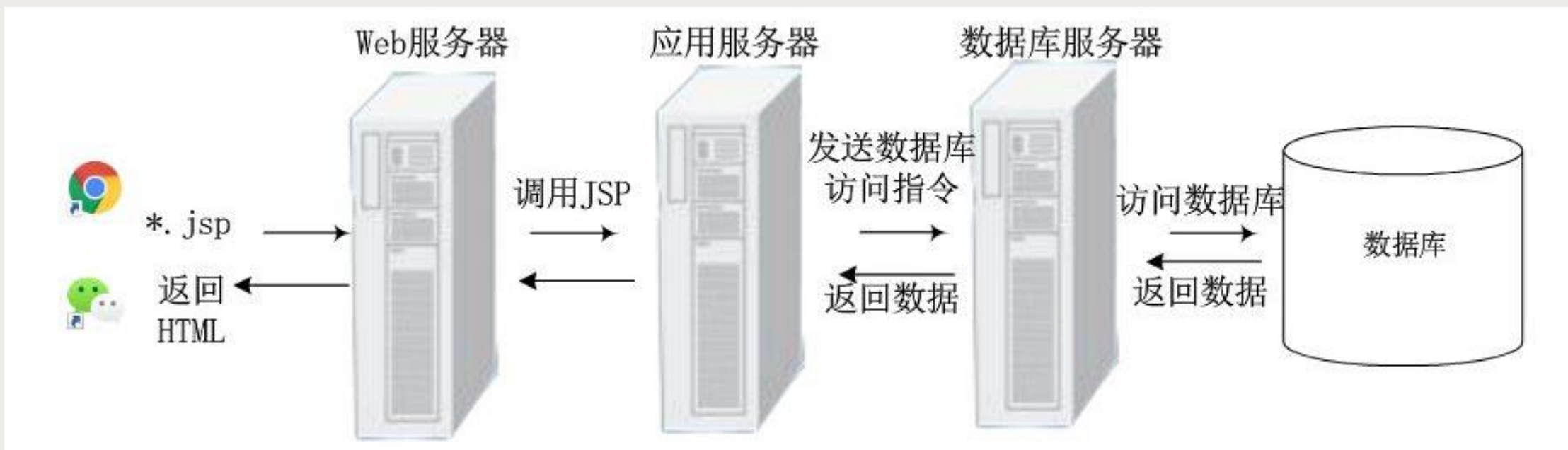
在日志帮助下，数据库以**事务**的方式维护多用户并发访问、故障情况下的数据一致性。



第一章 绪论

1.3 数据库系统

- **数据库系统 (DBS)**：由硬件、软件（操作系统、数据库、数据库管理系统、开发工具、应用系统等）和用户（常规用户、数据库管理员）三大部分组成。
- **大数据的目标**：高效、安全、共享数据库。





第一章 绪论

1.4 数据库管理技术发展趋势

- **数据库技术发展趋势：**

需重新进行设计以适应新架构、需尽快适应近几年硬件方面的快速发展。

- **RDBMSs：关系数据库管理系统**

- 简单性（关系模型，声明性语言SQL，三层模式两级映射）
- 高效性（查询优化，并发、恢复，索引、物化视图、反规范化）
- 安全性（视图、权限、约束，并发控制、恢复机制，数据加密）

- **NoSQL：Not only SQL，泛指非关系型的数据库**

- **NEWSQL：兼顾可扩展性、高性能、关系模型、声明性语言、完全ACID特性的系统**



第二章 关系模型

2.1 关系结构与约束

- **关系表：**关系数据库使用一个或多个表来存储数据。
 - 行：元组（表中的每一行数据，无序、不重复）。
 - 列：属性/字段（每一列的列名，无序、不重复）。
 - 域：属性可以具有的值的合法集合/属性的取值范围。
 - 表：关系（表名即为关系名）。
- **关系：**数学上把一系列域的笛卡尔积的子集成为关系。
- **约束：**对属性取值范围的限定或对属性之间相互关系的限定。



第二章 关系模型

2.1 关系结构与约束

- **超键：**在给定关系中，能唯一标识出各个元组的属性集合（任何包含候选键的属性集合）。
- **候选键：**在给定关系中，能唯一标识出各个元组的属性集合，且集合内不包含多余属性（即无关紧要的属性）。一个关系有一个或多个候选键。

候选键一定是超键，但超键不一定是候选键。

只有任何真子集都不是超键的超键才是候选键。

- **主键：**从候选键中指定的唯一一个用来标识元组的属性。

主键一定是候选键，但候选键不一定是主键。

- **外键：**关系中要求其值必须与其他关系中的属性或属性组匹配的属性或属性组。



第二章 关系模型

2.1 关系结构与约束

- **关系代数：**关系操作的代数表示，包括一个运算集合，这些运算以一个或两个关系作为运算数，产生一个新的关系作为运算结果。

关系代数运算结果为关系 + 关系中不允许重复元组 = 关系代数运算结果的去重的

关系代数运算可以组合使用

- **基本关系代数运算**（选择、投影、集合并、集合差、笛卡尔积、更名）

基本关系代数运算是完备的，可表示任何普通关系代数查询

- **附加关系代数运算**（集合交、自然联接、属性联接、条件联接、赋值）

为解决基本关系代数运算查询表达式过于冗长、复杂而增添的运算

- **扩展关系代数运算**（广义投影、聚集、外联接）



第二章 关系模型

2.2 基本关系代数

运算名称	书写表达式 (关系名可替换为关系代数表达式)	运算定义
选择	$\sigma_{\text{条件表达式}}(\text{关系名})$	选出满足给定条件的元组，结果关系和原关系模式相同（及表的列及约束相同）。
投影	$\Pi_{\text{属性名1}, \dots, \text{属性名n}}(\text{关系名})$	用来从给定关系产生一个只有其部分列的新关系。
集合并	关系名1 \cup 关系名2	将两个相容的关系（属性同元、域相同或相容）中的元组合并。
集合差	关系名1 - 关系名2	将两个相容的关系作差，得出存在于关系1而不存在于关系2的所有元组。
笛卡尔积	关系名1 \times 关系名2	结果关系模式为两个关系模式的串接，元组为左侧每个元组与右侧每个元组的拼接。
更名	$\rho_{\text{新关系名}(\text{属性名1}, \dots, \text{属性名n})}(\text{原关系名})$	修改关系及对应属性的名称。



第二章 关系模型

2.3 附加关系代数



运算名称	书写表达式 (关系名可替换为关系代数表达式)	运算定义
集合交	关系名1 \cap 关系名2	选错两个相容的关系中共同存在的元组。
自然联接	关系名1 \bowtie 关系名2	等同于先将两个关系做笛卡尔积；选择两个关系中所有同名属性值对应相等的元组后去除重复列。
属性联接	关系名1 $\bowtie_{\text{属性名}}$ 关系名2	等同于先将两个关系做笛卡尔积；选择两个关系中指定的某个或某几个同名属性值对应相等的元组后去除重复列。
条件联接	关系名1 $\bowtie_{\text{条件表达式}}$ 关系名2	等同于先将两个关系做笛卡尔积；选择两个关系中符合条件的的元组（列不去重）。
赋值	关系变量名 \leftarrow 关系代数表达式	将右侧表达式的运算结果赋给左侧关系。



第二章 关系模型

2.4 扩展关系代数

运算名称		书写表达式 (关系名可替换为关系代数表达式)	运算定义
广义投影		$\Pi_{\text{表达式1}, \dots, \text{表达式n}}(\text{关系名})$	允许在投影列表中使用算数表达式（表达式也可以是单个属性或常量）。
聚集		$G_{\text{聚集函数名}(\text{属性名})}(\text{关系名})$	使用sum、avg、count、max、min聚集函数进行计算后返回计算结果，一般结果只有一个元组。
分组聚集		分组属性 $G_{\text{聚集函数名}(\text{属性名})}(\text{关系名})$	按照指定属性的不同取值进行分组，对每一组内的元组进行聚集函数的计算。
外联接	左联接	关系名1 \bowtie_L 关系名2	对两个关系进行联接运算，并保留左侧关系中的悬浮元组。
	右联接	关系名1 \bowtie_R 关系名2	对两个关系进行联接运算，并保留右侧关系中的悬浮元组。
	全联接	关系名1 \bowtie_F 关系名2	对两个关系进行联接运算，并保留两侧关系中的悬浮元组。



习题

第一章：

1. (填空) 数据库系统中数据的独立性包括 ()、() 两个方面。
2. (填空) 数据库的逻辑数据独立性是由 () 映射提供的。
3. (单选) 数据库 (DB)、数据库系统 (DBS) 和数据库管理系统 (DBMS) 三者之间的关系是 ()。
 - A. DBMS包括DB和DBS
 - B. DB包括DBS和DBMS
 - C. DBS包括DB和DBMS
 - D. DBS就是DB, 也就是DBMS
4. (单选) 数据库管理系统能实现对数据库中数据表、索引等对象的定义、修改、删除, 这类语言称为 ()。
 - A. 数据定义语言 (DDL)
 - B. 数据操纵语言 (DML)
 - C. 数据查询语言 (DQL)
 - D. 数据控制语言 (DCL)



习题

第一章：

5. (单选) 数据库中，数据的物理独立性是指（ ）。

- A. 数据库与数据库管理系统的相互独立
- B. 用户程序与DBMS的相互独立
- C. 用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的
- D. 应用程序与数据库中数据的逻辑结构相互独立

6. (单选) 要保证数据库的数据独立性，需要修改的是（ ）

- A. 三层模式之间的两种映射
- B. 模式与内模式
- C. 模式与外模式
- D. 三层模式

7. (单选) （ ）用来记录对数据库中数据进行的每一次更新操作。

- A. 数据库文件
- B. 缓冲区
- C. 日志文件
- D. 后援副本



习题

第二章：

1. (单选) 同一个关系模型的任意两个元组值 () 。
A. 不能全同 B. 可全同 C. 必须全同 D. 以上都不是
2. (单选) 下面关于关系性质的叙述中, 不正确的是 () 。
A. 关系中元组的次序不重要 B. 关系中列的次序不重要
C. 关系中元组不可以重复 D. 关系不可以为空关系
3. (单选) 下列哪些运算是关系代数的基本运算 () 。
A. 交、并、差 B. 投影、选择、除、联接
C. 联接、自然联接、笛卡尔积 D. 投影、选择、笛卡尔积、差
4. (单选) 关系运算中花费时间可能最长的运算是 () 。
A. 投影 B. 选择 C. 笛卡尔积 D. 除



习题

第二章：

5. (单选) 候选码 (候选键) 中的属性可以有 ()。

- A. 0个 B. 1个 C. 1个或多个 D. 多个

6. (单选) 自然联接是构成新关系的有效方法。一般情况下, 当对关系R和S使用自然联接时, 要求R和S含有一个或多个共有的 ()。

- A. 元组 B. 行 C. 记录 D. 属性

7. (单选) 下面的两个关系中, 职工号和设备号分别为职工和设备的关键字 (主键) :

职工 (职工号, 职工名, 部门号, 职务, 工资)

设备 (设备号, 职工号, 设备名, 数量)

两个关系的属性中, 存在一个外关键字 (外键) 为 ()。

- A. 职工关系的 “职工号” B. 职工关系的 “设备号”
C. 设备关系的 “职工号” D. 设备关系的 “设备号”



习题

第二章：

8. (填空) 若商品关系 $G(GNO, GN, GQ, GC)$ 中, GNO 、 GN 、 GQ 、 GC 分别表示商品编号、商品名称、数量、生产厂家, 若要查询 “上海电器厂生产的其数量小于100的商品名称” 用关系代数可表示为 ()

9. (多选) 对于关系模式学生 $S(Sno, Sname, Age, Dept)$ 、课程 $C(Cno, Cname, Teacher)$ 、成绩 $SC(Sno, Cno, Score)$, 下列查询结果相同的是 ()

- A. $\pi_{Sname}((S) \bowtie \sigma_{Score > 60}(SC))$
- B. $\pi_{Sname}(\sigma_{Score > 60}(S \bowtie SC))$
- C. $\sigma_{Score > 60}(\pi_{Sname}(S \bowtie SC))$
- D. $\sigma_{Score > 60}(\pi_{Sname}(S) \bowtie (SC))$



习题

第二章：

10. (单选) 关系R与关系S只有1个公共属性，T1是R与S作 θ 连接（等值联接）的结果，T2是R与S作自然连接的结果，则（ ）。

- A. T1的属性个数等于T2的属性个数
- B. T1的属性个数小于T2的属性个数
- C. T1的属性个数大于或等于T2的属性个数
- D. T1的属性个数大于T2的属性个数

11. (单选) 设有关系 R(A, B, C) 和 S(C, D)。与SQL语句select A, B, D from R, S where R.C = S.C等价的关系代数表达式是（ ）。

- | | |
|---|---|
| A. $\sigma_{R.C=S.C}(\pi_{A,B,D}(R \times S))$ | B. $\pi_{A,B,D}(\sigma_{R.C=S.C}(R \times S))$ |
| C. $\sigma_{R.C=S.C}((\pi_{A,B}(R)) \times (\pi_D(S)))$ | D. $\sigma_{R.C=S.C}(\pi_D((\pi_{A,B}(R)) \times S))$ |



习题

第二章:

R	A	B	C
	1	2	3
	4	5	6
	7	8	9

S	D	E	F
	6	5	1
	7	4	2
	8	3	3

12. (关系代数计算) 设有关系R和S如上, 试写出如下关系代数的结果值:

$$\Pi_{A,F}(\sigma_{E<5}(R \bowtie_{C<D} S))$$



习题

第二章：

R	A	B	C
	1	2	3
	4	5	6
	7	8	9

S	D	E	F
	6	5	1
	7	4	2
	8	3	3

12. (关系代数计算) 设有关系R和S如上，试写出如下关系代数的结果值：

$$\Pi_{A,F}(\sigma_{E<5}(R \bowtie_{C<D} S))$$

答案：

A	F
1	2
1	3
4	2
4	3



习题

第二章：

13. 设有三个关系：

学生：S (S#, SNAME, AGE, SEX)

课程：C (C#, CNAME, TEACHER)

成绩：SC (S#, C#, GRADE)

试用关系代数表达式表示下列查询语句：

- (1) 检索至少选修两门课程的学生学号 (S#) 。
- (2) 检索全部学生都选修的课程的课程号 (C#) 和课程名 (CNAME) 。
- (3) 检索选修课程包含“陈军”老师所授课程之一的学生学号 (S#) 。
- (4) 检索选修课程号为k1和k5的学生学号 (S#) 。



习题

第二章：

13. 设有三个关系：

学生：S (S#, SNAME, AGE, SEX)

课程：C (C#, CNAME, TEACHER)

成绩：SC (S#, C#, GRADE)

$$(1) \Pi_{S\#} (\sigma_{1=4 \wedge 2 \neq 5} (SC \times SC))$$

$$(2) \Pi_{C\#, CNAME} (C \infty (\Pi_{S\#, C\#} (SC) \div \Pi_{S\#} (S)))$$

$$(3) \Pi_{S\#} (SC \infty \Pi_{C\#} (\sigma_{TEACHER='陈军'} (C)))$$

$$(4) \Pi_{S\#, C\#} (SC) \div \Pi_{C\#} (\sigma_{C\#='k1' \vee C\#='k5'} (C))$$

试用关系代数表达式表示下列查询语句：

(1) 检索至少选修两门课程的学生学号 (S#) 。

(2) 检索全部学生都选修的课程的课程号 (C#) 和课程名 (CNAME) 。

(3) 检索选修课程包含“陈军”老师所授课程之一的学生学号 (S#) 。

(4) 检索选修课程号为k1和k5的学生学号 (S#) 。

THANKS

