

一、上次课堂作业分析

- 9道判断题，还是有很多同学做错，请大家仔细学习解析，掌握自己为什么做错。通过这几道判断题说明同学们对于关系代数的知识还是掌握不够的，需要加强复习。最终对大家的要求是给你一个查询任务的题目的叙述，你就要直接把正确的关系代数的查询式子写出来。
- 那什么是正确的关系代数的查询式子呢？通过这9道题目大家要明白，关系和关系模式是不一样的概念，关系模式是‘型’的概念，是静态的；而关系是‘值’的概念，是动态的，关系表的具体值都是可以变化的，但是无论怎么变它都需要满足一个关系模式的约束，关系模式可以体现完整性的约束条件（主码是什么？外码是什么？等等），正确的关系代数的式子是在某种关系模式下，无论你的关系表如何变，你都能通过这个关系代数的式子将正确的结果查询出来。
- 关系代数答案不唯一，关系代数中属性、表名要按照原本题目中的样子书写，即区分大小写，注意属性的分量值是否属于字符类型，若属于字符类型，判断时应当打单引号。

1. 课件例 2.10 中查询至少选修 1 号课程和 3 号课程的学生号码。

该关系代数表达式正确么？

$$(\prod_{Sno}(\sigma_{Cno='1'}(SC))) \cap (\prod_{Sno}(\sigma_{Cno='3'}(SC)))$$

对， $\prod_{Sno}(\sigma_{Cno='1'}(SC))$ 得到的是选了 1 号课程的学生的学号， $\prod_{Sno}(\sigma_{Cno='3'}(SC))$ 得到的是选了 3 号课程的学生的学号，两者求交就为至少选修了 1 号和 3 号课程的学生号码，为什么满足“至少”的条件呢？例如，如果是选了 1、3、4 号课程的学生学号也会被该关系代数的式子查询到，是符合题目要求的。

2. 课件例 2.10 中查询至少选修 1 号课程和 3 号课程的学生号码。

该关系代数表达式正确么？

$$\prod_{Sno} (\sigma_{Cno='1'}(SC) \cap \sigma_{Cno='3'}(SC))$$

错，比如

$\sigma_{Cno='1'}(SC)$ 查到 表 1

Sno	Cno	Grade
2001	1	90

$\sigma_{Cno='3'}(SC)$ 查到 表 2

Sno	Cno	Grade
2001	3	80

根据表 1 与表 2， $\sigma_{Cno='1'}(SC) \cap \sigma_{Cno='3'}(SC)$ 为空集，请同学们回头看 RNS 的计算原则。但是根据题意这个 2001 号同学应该是被查询出来的。

通过对比题目 1 与题目 2，这就是为什么题目一在求交之前要做投影处理：即对原表符合条件（选了某课程）的元组筛选出来之后再对 Sno 进行投影。

3. 课件例 2.10 中查询至少选修 1 号课程和 3 号课程的学生号码。
该关系代数表达式正确么？

$$\Pi_{\text{Sno}, \text{Cno}}(\text{SC}) \div \Pi_{\text{Cno}}(\sigma_{\text{Cno}='1' \vee \text{Cno}='3'}(\text{Course}))$$

对，课件与教材例 2.10 的答案就为 $\Pi_{\text{Sno}, \text{Cno}}(\text{SC}) \div K$ ，而临时表 K 为

Cno
1
3

$\Pi_{\text{Cno}}(\sigma_{\text{Cno}='1' \vee \text{Cno}='3'}(\text{Course}))$ 的含义等价于建立临时表 K，针对 Course 表选择 Cno 值满足 1 号或 3 号的元组，这些元组是包含 Course 表中所有的属性的，所以再进行 Cno 的投影。投影是要去掉重复的，所以最终得到的结果就是等价于临时表 K。

4. 课件例 2.10 中查询至少选修 1 号课程和 3 号课程的学生号码。
该关系代数表达式正确么？

$$\Pi_{\text{Sno}, \text{Cno}}(\text{SC}) \div \Pi_{\text{Cno}}(\sigma_{\text{Cno}='1' \wedge \text{Cno}='3'}(\text{Course}))$$

错，因为 $\sigma_{\text{Cno}='1' \wedge \text{Cno}='3'}(\text{Course})$ 表示：对于 Course 表中选择逻辑表达式“ $\text{Cno}='1' \wedge \text{Cno}='3'$ ”为真的元组，但是对于每一个元组的 Cno 的值都只可能有一个值，不可能既等于 1 又等于 3。因此该关系代数式错误。

5. 课件例 2.10 中查询至少选修 1 号课程和 3 号课程的学生号码。
该关系代数表达式正确么？

$$\Pi_{\text{Sno}}(\sigma_{\text{Cno}='1' \wedge \text{Cno}='3'}(\text{SC}))$$

错，与第 4 题错误原因一致，那同学们会问如果把 \wedge 改为 \vee 呢？

$\Pi_{\text{Sno}}(\sigma_{\text{Cno}='1' \vee \text{Cno}='3'}(\text{SC}))$ 也为错，因为该表达式不能保证选出来的 Sno 是至少选择了 1 号和 3 号课程的，只能保证查询到选了 1 号课程或者 3 号课程的学生的学号。这也是课件例 2.10 为什么要用 \div ，因为除法表示包含的意思，可以体现出“至少”。

6. 课件例 2.10 中查询至少选修 1 号课程和 3 号课程的学生号码。
该关系代数表达式正确么？

$$\prod_{Sno} (SC \div \prod_{Cno} (\sigma_{Cno='1' \vee Cno='3'} (SC)))$$

错， $\prod_{Cno} (\sigma_{Cno='1' \vee Cno='3'} (SC))$ 与课件及教材的例 2.10 中的临时表 K 是等价的。那么我们看看 $SC \div K$ 等于什么？

假设 SC 表只有 2001 学号同学，如表 3，2001 学号的学生选了 1、3、4 号课程，2001 号同学肯定是查询结果，但是我们看如果是 $SC \div K = \Phi$ （空集）。SC 表中 (2001,80) 的象集为 1，(2001,75) 的象集为 3，(2001,90) 的象集为 4。没有任何一个 (Sno,Grade) 值的象集包含 K 表对共有属性 Cno 的投影结果 (1,3)。所以， $SC \div K = \Phi$ ，为空集。该题的核心考察大家对除法运算的掌握。

表 3

Sno	Cno	Grade
2001	1	80
2001	3	75
2001	4	90

7. 查询选修了全部直接先行课为 6 号课程的学生姓名。

该关系代数表达式正确么？

$$\prod_{Sname} (\sigma_{Cpno='6'} (Course) \bowtie \prod_{Sno, Sname} (Student))$$

错，题目的核心为全部直接先行课为 6 号课程，跟课件和教材的例 2.12 是不一样的。题目的意思是：假设以课程 6 为先行课的所有课程为课程 3、课程 4。所求的学生应当是至少同时选修了课程 3 与课程 4 的。而题目中的式子不能完成这个要求，题目中只是表达查询至少选修了一门先行课为 6 号课程的同学。该题正确的关系代数表达式为（课件上也有）

$$\prod_{Sname} ((\prod_{Sno, Cno} (SC) \div \prod_{Cno} (\sigma_{Cpno='6'} (Course))) \bowtie Student)$$

8. 查询选修了全部直接先行课为 6 号课程的学生姓名。

该关系代数表达式正确么？

$$\prod_{Sname} (\sigma_{Cpno='6'} (Course) \bowtie \prod_{Sno, Sname} (Student) \bowtie SC)$$

错，除了第 7 题的错误点外， $\sigma_{Cpno='6'} (Course) \bowtie \prod_{Sno, Sname} (Student)$ 是不能进行运算的，因为没有共同的属性，而自然选择需要具有同名属性。注意 $\sigma_{Cpno='6'} (Course)$ 的属性组与原 Course 一致。

9. 查询选修了全部直接先行课为 6 号课程的学生姓名。

该关系代数表达式正确么？

$$(\prod_{Sname, Cno} (Student \bowtie SC)) \div (\prod_{Cno} (\sigma_{Cpno='6'} (Course)))$$

错，假设以课程 6 为先行课的所有课程为课程 3、课程 4， $\prod_{Cno} (\sigma_{Cpno='6'} (Course))$ 等价于建立临时表 S，如表 4。

表 4

Cpno
3
4

假设 Student 为表 5:

Sno	Sname	Ssex	Sage	Sdept
2001	李勇	男	17	MA
2004	李勇	男	18	MA

假设 SC 为表 6:

Sno	Cno	Grade
2001	3	80
2004	4	75

Student \bowtie SC 为表 7:

Sno	Sname	Ssex	Sage	Sdept	Cno	Grade
2001	李勇	男	17	MA	3	80
2004	李勇	男	18	MA	4	75

$\Pi_{Sname, Cno}(Student \bowtie SC)$ 为表 8:

Sname	Cno
李勇	3
李勇	4

$\Pi_{Sname, Cno}(Student \bowtie SC) \div S$ 为

Sname
李勇

但是大家可以看到叫李勇的有 2 位同学，实际上任何一位李勇同学都没有同时修课程 3、4，按理说不符合题目的要求的，但是最后的查询结果却将其查出来了。所以本题表达式错误，这也就是为什么正确答案（见习题 7）中在除法的左端为 $\Pi_{Sno, Cno}(SC)$ ，这里的投影属性含有主属性 Sno。Sname 并不是 Student 表的主码。所以为什么先讲完整性约束的概念再来到关系代数的讲解，所有知识点一环套一环。整个书本知识点是由浅入深。

二. 本次课的课堂任务

本次课继续学习 SQL 语言，本次课将开始学习 SQL 最重要的知识点：数据的查询。

1. 今天首先学习视频 P18-P20，网址为

<https://www.bilibili.com/video/BV13J411J7Vu?p=18>

涉及教材的 P89-103 面，配套学习课件“数据库系统概论-计科-第三章_2.pdf”

今天涉及两个部分：单表查询、连接查询。

- 首先从最简单的单一关系表的查询开始学起 SQL 查询的语法，分为：

1) 选择表中的若干列: *代表什么含义? 对属性名怎么采用别名的方法?
2) 选择表中的若干元组: 用什么关键词能达到关系代数中 Π 的作用? 比较大小运算符, IN\NOT IN, 字符匹配查询中 LIKE、ESCAPE、%、_ 都分别是什么含义? 怎么使用? 空值查询和符合逻辑表达式查询怎么书写?

3) ORDER BY 子句: 什么时候用? 默认为正序还是倒序? 只能放在查询语句的什么位置?

4) 聚集函数: 聚集函数有哪些? 都有什么含义? 什么时候使用?

5) GROUP BY 子句: 课件中扩充了一个书上及视频中没有的例题, 请注意学习。

SELECT 相当于对属性的操作, 即关系表中的列提取。WHERE 相当于对满足条件的行操作, 类似于关系代数中的 σ 。聚集函数与 GROUP BY 子句是难点, 什么是聚集函数?

GROUP BY 经常与 HAVING 连用, 大家要重点区分 WHERE 和 HAVING 的作用对象, 都是满足条件的筛选出来, 但是两者的作用对象有什么区别? WHERE 子句中能出现聚集函数么?

- 如果查询涉及到 2 个及 2 个以上的表时, 就可能面临着连接查询。

任意两个表进行连接查询的执行步骤是什么? 有哪两种类型? 怎样用 SQL 语言表达关系代数中 $R \bowtie S$? 有些查询任务必须通过自身表的连接进行, 那对自身表进行连接时, 是将其当做不同的表还是相同的表进行操作呢? 自身表连接与不同的两个表连接在执行步骤原理上是一样的么? 3 个及 3 个以上的表连接的关键 WHERE 的筛选条件一般需要怎么写?