

一. 背景:

同学们经常听到“大数据”这个名词，大数据是现在非常热门的一个学科方向，但是大数据的学科发展不是一蹴而就的，大数据中非常重要的内容是海量数据的存储、管理。而针对数据的存储及管理是数据库技术初期就出现的重要研究内容，本门课程将介绍数据库技术的经典基础理论知识，当代先进的数据库管理与存储技术（包括大数据部分）都是在这经典的基础理论上发展而来。所以本门课程的开展是为了让同学们掌握最基础、最重要、最经典的数据库理论知识，为后续相关进阶课程打下基础。

掌握数据库的知识对大家后续考研、工作都有非常积极的意义，部分高校相关专业考研中会考数据库的知识点；BAT 等大型互联网公司、银行部门等单位也都需要掌握相关技能的人才，部分详细的招聘岗位要求可见我给大家的课件。

二. 任课老师:

罗晖，邮箱为 luohui2006@foxmail.com，研究方向为遥感图像智能处理，欢迎大家课余与我交流！

三. 课时安排:

本课程共 32+16=48 学时，即上课部分 32 课时，一周上 2 次课（2*2=4 课时），总共上 8 周课程（4*8=32 课时）。课内上机部分 16 课时，即 4 次上机课，一次上机课持续 3 小时，后续会跟大家商量上机时间。

由于课时的限制和这学期特殊的情况，本课程将讲授本教材（王珊，萨师煊：数据库系统概论(第五版)，高等教育出版社）中数据库最经典的理论知识部分。后续的内容，感兴趣的同学可以以后结合自己的需求自己学习。

上课时间按照新校区的作息时间表，每周二、五 3、4 节，10:15-11:00、11:05-11:50。

四. 考试:


考试为闭卷考试，期末总评成绩=40% 平时成绩+60% 期末试卷成绩。平时成绩根据作业，上机，考勤等综合决定。考勤时间和方式是随机的。

五. 上课内容及答疑:

每次上课的要求和方式及资料都不一定一样，我都会按照进度提前在群里通知大家。

课上和课下同学们有任何问题，都可以给我 QQ 留言或给我发邮件。

发给大家的课件说明： 点击该按钮进入详细解释的页面， 点击该按钮回到主

页面， 点击该按钮进入课件逻辑上的下一页。若课件页面无上述三个按钮，则直接使用 PDF 阅读器的下一页按钮来到逻辑上的下一页进行学习。

六. 今天的教学任务

今天的教学任务为第一章绪论部分的 1.1 节（教材的 P3-14）。本次教学内容的重点知识点为两大块：数据库领域内的 4 个基本概念、数据库技术的发展阶段和各自阶段的特点。

1. 在慕课网上，本课程设有人民大学的国家级精品课程《数据库系统概论（基础篇）》，但慕课可能会崩溃，这样会影响大家，这里提供给大家 bilibili 网址进行视频教学，bilibili 视频与慕课网上视频一致。

首先同学们请前往网址 <https://www.bilibili.com/video/av20449194?p=3>，观看该网址的 1.1 概述（1）-（2）两节。

2. 看完上一步的内容后，前往另一个网址 <https://www.bilibili.com/video/av83450228?p=4>，看 p2 数据库的 4 个基本概念，p3 数据技术的产生，数据系统的特点 这两小节，进一步加深对本次教学内容的重要知识点的理解程度。

3. 视频学习时，请结合教材和我发给大家本次课的课件（数据库系统概论-计科-第一章_1.pdf），课件部分内容比视频内容有所扩充，扩充部分也需要学习掌握和消化。

4. 上课时间没有完成学习任务的同学，请课下继续完成。

再次强调一下，大家学习的时候一定要将我发给大家的课件自学完成，课件的部分内容是教材和相应视频中某些知识点加深或者扩充多一些具体例子。需要掌握的知识点以发给大家的课件为准。每次上课的任务教学视频和课件没有学习完的，课下要自己花时间完成。

整个第一章的内容为绪论，很多为理论，虽然抽象，但是这些理论是很重要，在大家掌握了这些理论后，后续会进一步深入，后面几章的内容也会与第一章的知识点连接起来，大家最后学完后希望能达到融会贯通的效果。

一、第一次课堂作业讲解

1. 有人说：“我们在开发数据库，是一个学生管理系统”，该处数据库指的是（D）。

- A. 数据库
- B. 数据库管理系统
- C. 数据库系统
- D. 数据库应用系统

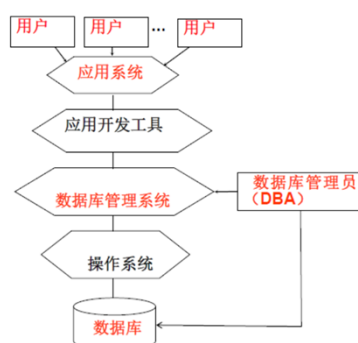
本道题目在发给你们的第一次上课课件的第 32 面，是一道原题，但是本道题错误率还是很高，很多同学都答错。这道题目的核心是大家要弄清楚数据、数据库（DB）、数据库管理系统（DBMS）、数据库系统（DBS）的概念，我们说学生管理系统不是上述的 4 个概念中的任何一个。现实情况中，学生管理系统是在 DBMS 平台基础上开发出来的一个应用系统。DBMS、DBS、数据库应用系统最后都有“系统”两个字，所以他们都是包含或者就是一个软件系统，而 DBS 跟其他两者系统之间明显不同的地方是包含人的概念，即 DBA。典型的 DBMS：微软的 SQL SERVER、甲骨文的 ORACLE、MYSQL、IBM 的 DB2 等等。人们可以在 DBMS 上使用 SQL 等语言进行数据库应用系统的开发。某个存储数据的文件如 1.db，db 是一种文件类型的后缀名，这个文件对应的概念为数据库。

一般情况，语义环境描述的数据库是指哪种，需要自行结合上下文进行区分。

2. 数据库（DB），数据库系统（DBS）和数据库管理系统（DBMS）之间的关系是（A）。

- A. DBS 包括 DB 和 DBMS
- B. DBMS 包括 DB 和 DBS
- C. DB 包括 DBS 和 DBMS
- D. DBS 就是 DB，也就是 DBMS

4) 数据库系统（续）



该题错误率低，解析如图所示。

3.下列四项中，不属于数据库系统阶段的特点是（ C ）。

- A. 数据共享
- B. 数据完整性
- C. 数据冗余度高
- D. 数据独立性高

该题错误率低，数据库系统阶段的特点是共享性高、冗余度小、DBMS 控制数据完整性、高度的物理独立性和一定的逻辑独立性。

4.下面列出的数据库管理技术发展的三个阶段中，没有专门的软件对数据进行管理的是（ D ）。Ⅰ. 人工管理阶段 Ⅱ. 文件系统阶段 Ⅲ. 数据库阶段

- A. Ⅰ 和 Ⅱ
- B. 只有 Ⅱ
- C. Ⅱ 和 Ⅲ
- D. 只有 Ⅰ

该题错误率低，Ⅱ 和 Ⅲ 分别采用文件系统和 DBMS 进行数据管理。

第 3、4 题可参考下表。

表 1.1 数据管理三个阶段的比较

		人工管理阶段	文件系统阶段	数据库系统阶段
背景	应用背景	科学计算	科学计算、数据管理	大规模数据管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘、磁盘阵列
	软件背景	没有操作系统	有文件系统	有数据库管理系统
	处理方式	批处理	联机实时处理、批处理	联机实时处理、分布处理、批处理
特点	数据的管理者	用户（程序员）	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用	现实世界（一个部门、企业、跨国组织等）
	数据的共享程度	无共享，冗余度极大	共享性差，冗余度大	共享性高，冗余度小
	数据的独立性	不独立，完全依赖于程序	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据的结构化	无结构	记录内有结构、整体无结构	整体结构化，用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

二、第二次上课的教学任务

前往 <https://www.bilibili.com/video/av20449194?p=3> 观看 1.2 数据模型，这次课学习到该视频的 74 分 10 秒，边看视频边配套学习课件的 1.2.1 两类数据模型——1.2.6 网状模型（对应教材的 P14-24），剩下 10 分钟的内容下次课再观看。

本次课内容涵盖了 1.2 节数据模型的 1.2.1 到 1.2.6，其中 1.2 节中的 1.2.7 关系模型是下次课的教学任务。

把本次课件教学内容的展开思路介绍一下，方便大家在学习的时候理解。

- 为什么要建立数据模型？

- 数据模型的定义是什么？
- 数据模型分哪几类？
- 数据模型的建立过程是什么？从而引出三个世界、两级抽象。哪三个世界，哪两级抽象？
- 通过小区宠物调查的示例，深入理解数据模型的两个层次以及三个世界。
- 来到数据模型的第一个层次：概念模型，它的定义、用途、要求，引出**非常重要的一个知识点：概念模型的重要表示方法——E-R 模型（或 E-R 图）**。
- 信息世界中的一些基本概念（实体、属性、码、实体型、实体集、联系，码的概念贯穿本门课程，很重要）。通过分类讲解具体的 E-R 图中联系如何表达：两个实体型之间的联系、两个以上实体型之间的联系、同一个实体型之间的联系、哪三种类型的联系？
- E-R 模型怎么绘制，实例讲解：工厂物资管理。
- 数据模型的组成三要素：数据结构、数据操作、数据的完整性约束条件。三要素各自的含义是什么？
- 从逻辑模型的角度来说，常用的数据模型有哪几种？怎么分类的？
- 重点讲解几种在数据库领域有深远影响的经典模型：层次、网状、关系（关系模型下次课的内容）
- 层次和网状模型的展开都是按照数据模型的组成三要素详细展开，他们各自的特点、定义、异同点和优缺点是什么？

一、第二次课堂作业讲解

1. 数据库中，数据的物理独立性是指（C）。

- A. 数据库与数据库管理系统（DBMS）的相互独立
- B. 用户的应用程序与 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

错误率较低，这里考察的是数据的物理独立性的概念，另外注意相对应的另一个概念：逻辑独立性。该知识点为第一次上课的知识点，解释课件如下。

数据独立性

●物理独立性

指用户的应用程序与存储在磁盘上的数据库中数据是相互独立的。当数据的物理存储改变了，应用程序不用改变。

●逻辑独立性

指用户的应用程序与数据库的逻辑结构是相互独立的。数据的逻辑结构改变了，用户程序也可以不变。

2. 概念模型是现实世界的第一层抽象，这一类模型中最著名的模型是（D）。

- A. 层次模型
- B. 关系模型
- C. 网状模型
- D. 实体-关系模型

错误率低，实体-关系模型即 E-R 模型，是概念模型中非常重要的一种模型表达方式。而层次、关系、网状模型都是数据模型的第二个层次（逻辑模型和物理模型）中的具有代表性的不同种类的逻辑模型。

3. 区分不同实体的依据是（B）。

- A. 名称
- B. 属性
- C. 对象
- D. 概念

错误率低，考察 E-R 模型中的基本概念，依据属性的定义。

4. 判断该描述的正确性：层次模型是比网状模型更具有普遍性的结构，网状模型是层次模型的一个特例。（B）

- A. 对
- B. 错

部分同学答错，现实世界中实物之间的联系很多是非层次的，所以网状模型比层次模型更贴近现实的普遍情况。层次模型和网状模型都的组成单元为基本层次联系，网状模型去掉了层次模型的两个限制，允许多个结点没有双亲结点，允许结点有多个双亲结点，此外网状模型允许两个结点之间有多种联系。层次模型实际上是网状模型的一个特例。所以该题答案为‘错’。

二、第三次上课的教学任务

本次上课内容涵盖 1.2.7-1.5 节，教材对应 P25-34。结合下面的视频与上传的最新课件配套学习。

1. 前往 <https://www.bilibili.com/video/av20449194?p=3> 观看上次上课看的 1.2 数据模型剩下 10 分钟的内容。从该视频的 74 分 10 秒开始看起，学习 1.2.7 关系模型。关系模型的讲解大纲与网状和层次模型一样：分别按照数据模型的三要素进行讲解（数据结构、数据操纵、完整性约束），最后总结优缺点。

2. 本次上课的**重点**任务，前往 <https://www.bilibili.com/video/av20449194?p=3> 观看 1.3 数据库系统结构。

1.3 节课件的讲解大纲如下：

- 型和值的概念，引申出模式和实例的概念
- 从数据库应用开发人员角度看，数据库系统通常采用三级模式结构，哪三级模式？
- 外模式、模式、内模式各自的含义
- 三级模式，两级映像
- 两级映像与数据库逻辑独立性、物理独立性的关系？
- 当模式与存储结构变化时，为什么要保证外模式的稳定性？通过什么机制保证的外模式的稳定性？
- **综上，请同学们思考数据库的三级模式结构的优点是什么？**

3. 前往 <https://www.bilibili.com/video/av83450228?p=8> 观看 P8 数据库系统的组成。

本次课后，第一章的教学就完成了，总结一下。本章涉及的重点知识点：

- **数据库系统概述**
 - 数据库的基本概念（4 个基本概念的定义分别是什么？之间的区别与联系是什么？）
 - 数据管理的发展过程（哪三个阶段？各阶段之间的区别与联系，都有什么特点？）
 - 数据库系统的特点（有什么特点？物理与逻辑的独立性的定义？）
- **数据模型**
 - 概念模型，E-R 模型（定义分别是什么，E-R 模型怎么画？）
 - 数据模型的三要素（哪三种要素？）
 - 三种主要数据库模型（层次、网状、关系模型各自的定义，涉及的基本概念？它们都有什么特点？优缺点？）
- **数据库系统内部的系统结构**
 - 数据库系统三级模式结构（哪三级模式结构？各自涉及的基本概念）
 - 数据库系统两层映像系统结构（定义是什么？两级映像与数据独立性的联系是？）
- **数据库系统的组成**（不同的角度看组成的结构，各组织结构的定义和所涉及的内容）

第三次教学任务课内时间没有完成的同学请课后完成视频和课件的学习，并自己总结复习，看看上述第一章所涉及的重点知识点的相关问题是否能回答？

三. 作业

第一道简单题在自己的 WORD 里面编辑答案，第二道 E-R 模型题目（包含 2 个小题）请大家用 PPT 或者 WORD 等其他软件画出下面题目要求的 E-R 模型。每位同学最后上交 1 个 WORD 文档，包含简答题答案，和自己用软件画的两道 E-R 模型题的图，WORD 文档以“姓名-班级”为文件名。3 月 14 号（本周六）晚上 20:00 之前，请学委收齐后打包发给我。没有电脑的同学就写在纸上，然后拍照把图像上交，图像名称也按照“姓名-班级”注明，一个同学涉及多个图像按照下列格式：“张三 1 班 1.jpg, 张三 1 班 2.jpg ……” **作业请独立完成，禁止抄袭！**

1. 试述数据库系统的三级模式结构，并说明这种结构的优点是什么？

2. E-R 模型

注：E-R 模型的构建需要结合当前的语义环境，所以注意阅读题目（即已知的语义环境，若已规定了某些语义环境，则构建的时候要完全按照已知的语义环境），联系实际情况，挖掘题干中没写的隐含信息，从而正确写出实体之间的对应联系类型。

1) 设有“产品”实体集，包含属性“产品号”和“产品名”，还有“零件”实体集，包含属性“零件号”和“规格型号”。每一产品可能由多种零件组成，通用零件可用于多种产品，有的产品需要一定数量的某类零件。因此，存在产品与零件的组成联系。请画出 E-R 图，并指出联系类型是 1:1, 1: n 还是 m:n。

2) 学校有若干个系，每个系有若干名教师和学生；针对若干学生，有辅导员、班主任、学务指导等等老师参与管理；每个教师可以承担若干门课程的教学，并参加多个项目；每个课程可以由多位老师任教，每个学生可以同时选修多门课程。请设计某学校的教学管理的 E-R 模型，要求给出每个实体、联系的属性。

实体及属性：

系（系编号，系名，系主任）

教师（教师编号，教师姓名，职称）

学生（学号，姓名，性别，班号）

项目（项目编号，名称，负责人）

课程（课程编号，课程名，学分）

联系：

任课：教师担任课程

参加：教师参加项目

选课：学生选修课程

领导：系、教师、学生之间为领导联系

一. 第三次课堂作业分析

1. 描述数据库全体数据的全局逻辑结构和特性的是 (A)。

- A. 模式
- B. 内模式
- C. 外模式

错误率低，同学们要搞清楚模式、内模式、外模式都分别是什么含义，这些含义课件和教材上都有详细的解释。

通过模式与外模式的定义思考模式（全局逻辑结构）与外模式（局部逻辑结构）之间的关系。

2. 要保证数据库的逻辑数据独立性，需要修改的是 (A)。

- A. 模式与外模式之间的映射
- B. 模式与内模式之间的映射
- C. 模式
- D. 三级模式

错误率低，该题目考察的是三级模式结构中的两层映象。数据的逻辑独立性和物理独立性由这两层映象关系保证。模式与外模式之间的映像、模式与内模式之间的映像分别保证了数据的逻辑独立性与物理独立性。这个知识点理解的核心是，数据库的使用者是用户，作为用户一般是不愿意看到使用的应用程序发生巨大变化的，因为对于数据库的应用系统直接关联的是外模式，也就是说在模式（逻辑上）和内模式（物理上）发生变化时，都希望看到外模式保证稳定，但是有时候模式发生变化（比如增加新的关系）、内模式发生变化（存储结构改变）是没法避免的，这时为了保证外模式的稳定，可以相应调整两级的映象关系，从而保证外模式和数据库应用程序的稳定。

3. (B) 的存取路径对用户透明，从而具有更高的数据独立性、更好的安全保密性，也简化了程序员的工作和数据库开发建立的工作。

- A. 网状模型
- B. 关系模型
- C. 层次模型
- D. 以上都有

错误率低，层次模型和网状模型都需要知道记录的详细路径，这样加重了编写应用程序的负担，但是另一方面，由于路径明确，在进行记录查找的时候效率较高。而关系模型不需要了解关系表的存取路径，安全性和数据独立性都更好，程序员也简化了工作，但是查询效率不如格式化数据模型。

4. 下述 (D) 不是 DBA 数据库管理员的职责。

- A. 完整性约束说明

- B. 定义数据库模式
- C. 数据库安全
- D. 数据库管理系统设计

部分同学错误错选成 B。DBA 需要决定数据库中的信息内容和结构。DBA 需要参与数据库设计的全过程，与用户、应用程序员、系统分析员密切合作、共同协商，做好数据库的设计。而数据库设计的全过程是包括数据库模式设计。但是 DBMS 是一个成熟的开发平台，是一个成熟的商业软件系统，我们使用 DBMS 来开发，比如 SQL SERVER。你是某个数据库系统的 DBA，但是你并没有参与 DBMS 的设计，只是使用 DBMS。

二. 第四次课堂任务

今天，我们将开始第二章关系数据库的学习，本章主要包含两大内容：1) 关系数据库相关概念更为深入的的学习，仍旧是按照之前数据模型的三个组成要素进行主线展开，分别为 2.1 关系数据结构及形式化定义，2.2 关系操作，2.3 关系的完整性。2) 关系代数，关系代数会详细讲解各个运算符的含义，会计算各种复合关系代数式子的最终结果。另一方面，也是本章的**难点**，针对特定的查询任务会使用各种关系代数的运算符写出准确的关系代数的复合表达式。由于关系代数较为抽象，因此在学习的过程中一定要结合具体的实例进行学习。

今天的教学任务是大家学习关系数据库的相关概念，有同学会问我们已经在第一章中的 1.2.7 学习了关系模型，为什么现在还要再学习关系数据库。这是因为第一章的关系模型是为这一章的关系数据库打基础的。比如：1.2.7 的关系模型中我们介绍过关系的定义——一个关系对应通常说的一张表。这是因为那时我们属于初学的时候，将这个比较通俗易懂的概念介绍给大家，大家更容易理解。但是今天我们所需要学习的 2.1.1 中介绍关系的概念就是通过严密的数学理论知识的角度来定义了。这实际上也是一个由浅入深的过程。

关系数据模型具有严密的数学理论基础，这也为它后续的持续发展奠定了基础。1970 年 E.F.Codd 在美国计算机学会会刊上发表的《A Relational Model of Data for Shared Data Banks》,开创了数据库系统的新纪元，也可能可以算是关系数据库理论的开篇之作。**我们在学习了第二章之后，对大家的要求是：有关关系模型、关系数据库的许多基本的概念的问题都需要用第二章内这种更严密的，具有数学理论基础的答案进行回答。**

今天的教学任务是大家学习 2.1 至 2.3 节，教材对应内容为 P37-48 面。结合课件观看网页 <https://www.bilibili.com/video/av20449194?p=6> 内的 P6-P8 节（2.1-2.3）视频内容。

课件大纲及重难点：

➤ 2.1 节

- 域、笛卡尔积、关系的数学基础的定义，笛卡尔积的计算、基数的计算？
- 元组、属性、N 元关系的含义。
- 候选码和全码的定义。结合课件的例子
- 主码、主属性、非主属性的含义，容易混淆的知识点主码的属性和主属性是一样的么？
- 笛卡尔积在实际应用中是不符合现实情况的，引申出非常重要的定义，什么是关系？
- 哪三类关系？
- 基本关系的性质有哪些？
- 关系模式的定义与数学表达，关系模式和关系之间的区别和联系，这里注意联系第一章讲过的概念“型”与“值”的概念。

- 进一步，关系数据库中的“型”与“值”的分别代表什么？

➤ 2.2 节

- 关系操作的简单介绍，其中包含的关系代数部分是下次课的重点内容。

➤ 2.3 节

- 实体完整性含义，空值的含义是什么？
- 参照完整性含义，外码、参照关系的含义。外码定义中的 R 与 S 可以是同一张表么？外码的取值有哪两种情况？如果外码中包含的属性是主属性，外码可以取空值么？
- 用户自定义的完整性含义

课堂内没完成视频和课件学习任务的，请课后完成。

一. 上课任务

今天的教学任务为本章的重点及难点：**关系代数**。主要会学习关系代数的各类代数符号的含义，掌握各种符号的计算方式。今天学习结合教学视频学习课件（数据库系统概论-计科-第二章_2.pdf，课件 1 页到 51 页，剩下的课件内容下次课学习），今天上课任务覆盖教材的 48-55 页，**课件中对视频和教材中的内容有部分扩充，比如关系代数中除法的讲解与实例，扩充部分也要重点学习。学习中请结合具体示例理解各代数符号含义。**

1. 今天首先学习 <https://www.bilibili.com/video/av20449194?p=9> 视频 2.4 关系代数，学习到第 68 分钟。

2. 学习完上述视频后，前往 <https://www.bilibili.com/video/av83450228?p=12> 学习 P11 传统集合运算、P12 关系特有操作，进行重点知识点的总结和复习。

这部分的难度相对比前面的课程大一点，所以希望同学们在好好消化。按照课件的讲解顺序，请大家思考下列问题。

- 今天的教学知识点主要分为两个部分，第一个是传统的集合运算，另一个是关系特有的操作。
- 传统集合运算中同学们应该掌握并、交、差、广义笛卡儿积的计算方法。给出任意一个关系 R 与 S ，这两个关系在任何条件下都可以做传统集合运算么？如果不是需要满足什么条件？
- 其中本次课程中广义笛卡儿积的计算与上一次课程 2.1.1 关系中笛卡儿积有什么区别与联系？
- 广义笛卡尔积中的核心表示 $t_r t_s$ 是什么意思？怎么计算？
- 专门关系运算中关于关系中常用的符号：选择、投影、连接、除。
- 选择与投影是针对关系表的什么方向来计算的？选择运算符中 F 为复合的逻辑表达式，进行选择条件的筛选。复合逻辑表达式怎么表达？括号表示的优先级最高。不等于通常用 $<>$ 表示。
- 等值连接与自然连接的区别与联系？
- 外连接是什么意思，分哪两种外连接。悬浮元组的含义是什么？
- 除法的运算理解需要基于象集的概念，象集是怎么计算的？除法怎么计算？
- 除法计算的一般步骤： $R(X, Y) \div S(Y, Z) = P(X)$ ，确定 Y 属性组是什么？ X 属性组是什么？

$$R \div S = \{t_r[X] \mid t_r \in R \wedge \Pi_Y(S) \subseteq Y_x\}$$

- 元组在 X 上分量值 x 的象集 Y_x 包含 S 在 Y 上投影的集合，这个包含指的什么意思？能大于么？能等于么？
- 学完了这次课后，所有关系代数的运算符号（传统集合运算+关系特有操作）的计算都需要掌握，任意给你几个关系表，需要能够计算出复合的关系代数运算的结果。

一. 课堂作业解析

1. 在通常情况下，下面关系中不可以作为关系数据库的关系的是 (D)

- A. R(学生号, 学生名, 性别)
- B. R(学生号, 学生名, 班级号)
- C. R(学生号, 学生名, 宿舍号)
- D. R(学生号, 学生名, 简历)

错误率低，关系要满足的基本原则是每个属性都是最小的、不可再分的，但是显然简历这个属性可以再分，所以 D 不能作为关系。

2. 根据关系模式的完整性规则，一个关系中的主码 (C)

- A. 不能有两个
- B. 不能成为另一个关系的外码
- C. 不允许为空
- D. 可以取相同的分量值

部分同学错误，如果一个关系有多个候选码，在特定的语义环境下选择一个作为主码，可以在另外一种语义环境下选择另外一个候选码作为主码，所以说主码在不同的语义环境下可能是不同的变化的，因此 A 不对。B 选项，外码的定义就直接说明了 B 选项是错误的。D 选项由于之前出题的时候打掉了部分字，这里 D 选项内容已更正，关系中主码的分量值不允许相同。一个关系中主码是不允许为空的，一定要有属性作为主码，并且主属性不能为空值。所以正确答案为 C。

3. 一般情况下，当对关系 R 和 S 使用自然连接时，要求关系 R 和 S 含有一个或多个共有的 (D)

- A. 元组
- B. 行
- C. 记录
- D. 属性

错误率低，自然连接，一定要求 R 与 S 有共有的属性。R 与 S 自然连接的结果并不一定是非空的，若结果为空集，说明 R 的元组与 S 中的元组进行自然连接时，它们的共有属性的取值不相等。

4. 设关系 R (A,B,C) 和关系 S (B,C)，A、B、C 为属性名，同名属性取自同一个域，为下列各关系代数表达式不成立的是 (C)

- A. $R \div S$
- B. $\prod_{B,C}(R) \cup S$
- C. $R \cap S$
- D. $R \bowtie S$

部分同学错误，A 选项的考察点在于 $R(X,Y) \div S(Y,Z) = P(X)$ ，S 中的 Z 是否一定要有？答案是不一定，Z 属性组是否存在 S 中并不影响 $R \div S$ 的计算，重要的是 R 与 S 一定要有共有的属性组 Y。B 选项考察的是 \cup 的要求，若 $R \cup S$ ，R 与 S 具有相同的属性个数，并且 R 与 S 相应的属性都取自同一个域，那么就可以进行 \cup 的运算，所以对 R 进行 B、C 属性的投影后再与 S 进行‘并运算’完全符合并运算的要求。C 选项中考察的是 \cap 运算的前提要求，与 \cup 一样，需要 R 与 S 具有相同的属性个数，并且 R 与 S 相应的属性都取自同一个域，显然 $R \cap S$ 不符合条件，R 有 3 个属性，但是 S 只有 2 个属性。D 选项为自然连接，R 与 S 具有共有属性 B,C，所以可以进行计算。

5. 设 $R \bowtie_{A\theta B} S = W$, W 、 R 、 S 都为关系名, 元组个数分别为 p 、 m 、 n , 那么三者之间满足 (D)

A. $p < (m+n)$

B. $p \leq (m+n)$

C. $p < (m*n)$

D. $p \leq (m*n)$

错误率较高,

笛卡尔积的定义: $R \times S = \{ \widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S \}$

θ 连接的定义:

$R \bowtie_{A\theta B} S = \{ \widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S \wedge t_r[A] \theta t_s[B] \}$

对比两个定义, θ 连接的结果从两个关系的笛卡尔积中选取属性间满足一定条件的元组。那么我们知道如果 $R \times S$ 中, R 的元组个数为 m , S 的元组个数为 n , $R \times S$ 的元组个数为 $m*n$ 。而 θ 连接的结果是笛卡尔积的子集, 所以答案为D。

二. 第六次上课任务

本次课程涵盖2个内容, 第二章的收尾和第三章的开头。

其中, 第二章的收尾为关系代数的综合训练, 这个部分内容是第二章 关系代数的**重点以及难点**, 希望同学们通过例子理解后举一反三。最后需要同学们达到的要求是给定的查询任务, 写出正确的关系代数综合表达式。**查询任务中关系代数的表达式通常并不唯一**。第二章中的2.5节 关系演算部分不做要求, 感兴趣的同学可以自己去看看。

1. 关系代数综合举例, 前往 <https://www.bilibili.com/video/av83450228?p=13> 学习p13 关系代数3 综合训练, 配套从上次课件(数据库系统概论-计科-第二章_2.pdf)的52面开始学习课件。涉及教材的P56-57。

关系代数综合举例以及第三章SQL语言学习时大家要学生-选课数据库的三个表很熟悉: student表、sc表、course表, 其中student表中存储所有的学生信息, sc表中存选了课的学生的选课情况, course表中存放所有的课程信息, course表中的cpno表示这门课的先行课的课程号(比如数据库课程的先行课课号为5, 5号代表数据结构, 即要先上数据结构, 才能有基础上数据库)。student主码为sno, sc表主码为sno、cno这个属性组, course表中主码为cno。可以看出sc表的中sno参照于student表中的sno, sc表中的cno参照于course表中的cno, sc表中的sno\cno是外码。course表中的cpno参照于course表中的cno, 所以cpno也是course表的外码。

第二章完成后, 关系代数的各个符号怎么运算的大家都需要掌握。如果给定任意的关系表, 要能够将关系代数的表达式正确计算结果写出来。另一方面, 针对特定的查询任务, 要能将正确的关系代数的符合表达式写出来。

关系代数有严格的语法要求, 都要按照正确的语法要求来写,比如某属性值为字符类型, 比如查询学号为20125100的同学的学生信息, $\sigma_{Sno='20125100'}(Student)$, 这里学号应当是一个字符类型的, 因此这里20125100需要打单引号。如果是数字类型, 比如查询年龄等于20岁同学的学生信息, 则为 $\sigma_{Sage=20}(Student)$, 这里的20没有单引号。另外, 通过综合训练的视频, 大家可以看到很多查询任务需要采用关系代数的复合表达式完成, 这时要注意运

算的先后顺序，即运算的优先级问题，一般具体计算时按照从左至右的顺序计算，但是如果
有括号，则先计算括号内的，也建议大家写式子时采用括号，因为括号的优先级最高。
再比如，若写成 $\sigma(\text{Student})_{\text{Sno} = '20125100'}$ 这种形式也属于不符合语法的错误写法。

**大家学完了综合训练后，请思考下面的问题，对于例2.10 查询至少选修1号和3号课程的学生
的号码（课件52页），以及查询选修了全部直接先行课为6号课程的学生姓名（课件55
页，课件中的扩充题目，视频中没有）这两道题目，还有没有其他的能完成这两项查询任
务关系代数的式子？关于这两道题目本次课内会布置相关的课堂作业。**


2. 开始第三章的学习，第三章的重点是掌握SQL语法，最终能使用正确的SQL语句完成题
目的任务。

首先前往 <https://www.bilibili.com/video/av83450228?p=16>，学习 P16 SQL概述。

再前往另一个页面 <https://www.bilibili.com/video/av20449194?p=12>，学习3.2 数据定义。
这两个视频内容都配套学习今天上传的课件（数据库系统概论-计科-第三章_1.pdf）。涉及
教材的P75-89。

**有些不好理解的知识点，课件中有具体举例和知识扩充，可能视频中是没有的，所以
也需要好好学习课件，比如SQL语言支持的数据库系统三级模式结构。**

课件的展开思路及重难点：

- SQL语言是关系数据库的标准语言。SQL对应中文是什么？（结构化的查询语
言），那可以看到顾名思义‘结构’和‘查询’是它非常大的特征，说明这个语
言肯定是结构化的，查询是它非常重要的功能，当然它的功能不止查询。
- SQL语言经过多年的发展，已经有很多版本的标准库，但是不同的DBMS中对标
准的SQL基本命令集进行了不同程度的扩充和修改，本章的SQL语言讲的都是
SQL的基本概念和功能，当使用不同的DBMS运行SQL时，还需查阅该DBMS的具
体用户手册。比如SQL SERVER基本支持本章所讲的SQL语法功能，但是个别不
支持，如drop table中SQL SERVER没有cascade 和 restrict的区别。
- SQL语言有什么特点？其中最重要的特点是它为高度非过程化的语言，用户不需
要了解存储路径，只用关心做什么。
- SQL语言怎样体现关系数据库的三级模式？这里与第一章的知识点进行了联系。
结合课件的具体举例理解。
- 数据定义部分就正式开始讲解SQL语法。整个讲解的过程会有大量的例子，对应
的数据库就是第二章大家熟悉的学生-课程数据库。
- SQL语言中的数据的定义，定义包含3个类别：创建（CREATE）、删除
（DROP）、修改（ALTER），三个命令都是属于对处理数据对象（模式、表、视
图、索引等）逻辑上的调整。不同处理对象对应的语句都是什么？
- 3.3节中首先出现的是模式的定义，注意这里的模式与第一章所讲的三级模式结构
中的模式概念并不一样。这里的模式，你们可以用命名空间或者架构（SQL
SERVER中用架构）来代替理解。比如SQL SERVER中默认的架构为dbo，如果默
认在对应数据库中建立一个表student，会生成  **dbo.student**。
- 基本表的创建、删除、修改的语法是非常重要的，后面的上机练习中会大量的用
到。

三. 第二章课后作业

与第一次课后作业要求一致，每位同学在 WORD 中写作业答案，WORD 文档以“姓名-班级”为文件名。**3月27号（下周五）晚上 20:00 之前**，请学委或班长收齐后打包发给我。
没有电脑的同学就手写纸上，然后拍照把图像上交，图像名称也按照“姓名-班级”注明，一个同学涉及多个图像按照下列格式：“张三 1 班 1.jpg, 张三 1 班 2.jpg” **作业请独立完成，禁止抄袭！**

1. 请说明关系模式、关系、关系数据库的区别与联系。

2. 关系代数表达式 $R \times S \div T - U$ 的运算结果是？

关系R

A	B
1	a
2	b
3	a
3	b
4	a

关系S

C
x
y

关系T

A
1
3

关系U

B	C
a	x
c	y

第3题与第4题类似“英译汉”与“汉译英”的关系。

3. 设有如下图所示的3个关系S, C和SC，将下列关系代数表达式用汉语表示出其查询含义

(即答案类似第4题中的题目)，并求其关系代数表达式得到最终运算的结果。

	学号 (Sno)	姓名 (Sname)	年龄 (Age)	性别 (Sex)	籍贯 (Origin)
S	98601	王晓燕	20	女	北京
	98602	李波	23	男	上海
	98603	陈志坚	21	男	长沙
	98604	张兵	20	男	上海
	98605	张兵	22	女	武汉

	课程号(Cno)	课程名 (Cname)	教师姓名 (Tname)	办公室 (Office)
C	C601	高等数学	周振兴	416
	C602	数据结构	刘建平	415
	C603	操作系统	刘建平	415
	C604	编译原理	王志伟	415

	学号 (Sno)	课程号 (Cno)	成绩 (Grade)
SC	98601	C601	90
	98601	C602	90
	98601	C603	85
	98601	C604	87
	98602	C601	90
	98603	C601	75
	98603	C602	70
	98603	C604	56
	98604	C601	90
	98604	C604	85
	98605	C601	95
	98605	C603	80

- 1) $\Pi_{Sname, Cno, Grade}(\sigma_{origin='上海'}(S \bowtie SC))$
- 2) $\Pi_{Sname, Cno, Grade}(S \bowtie SC \bowtie \sigma_{Cname='操作系统'}(C))$
- 3) $\Pi_{Sname, Age}(S \bowtie (\Pi_{Sno, Cno}(SC) \div \Pi_{Cno}(C)))$

4. 教材70面第6题

6. 设有一个 SPJ 数据库，包括 S、P、J 及 SPJ 4 个关系模式：

S(SNO, SNAME, STATUS, CITY);

P(PNO, PNAME, COLOR, WEIGHT);

J(JNO, JNAME, CITY);

SPJ(SNO, PNO, JNO, QTY)。

供应商表 S 由供应商代码 (SNO)、供应商姓名 (SNAME)、供应商状态 (STATUS)、供应商所在城市 (CITY) 组成。

零件表 P 由零件代码 (PNO)、零件名 (PNAME)、颜色 (COLOR)、重量 (WEIGHT) 组成。

工程项目表 J 由工程项目代码 (JNO)、工程项目名 (JNAME)、工程项目所在城市 (CITY) 组成。

供应情况表 SPJ 由供应商代码 (SNO)、零件代码 (PNO)、工程项目代码 (JNO)、供应数量 (QTY) 组成，表示某供应商供应某种零件给某工程项目的数量为 QTY。

今有若干数据如下：

S 表

SNO	SNAME	STATUS	CITY
S1	精益	20	天津
S2	盛锡	10	北京
S3	东方红	30	北京
S4	丰泰盛	20	天津
S5	为民	30	上海

P 表

PNO	PNAME	COLOR	WEIGHT
P1	螺母	红	12
P2	螺栓	绿	17
P3	螺丝刀	蓝	14
P4	螺丝刀	红	14
P5	凸轮	蓝	40
P6	齿轮	红	30

J 表

JNO	JNAME	CITY
J1	三建	北京
J2	一汽	长春
J3	弹簧厂	天津
J4	造船厂	天津
J5	机车厂	唐山
J6	无线电厂	常州
J7	半导体厂	南京

SPJ 表

SNO	PNO	JNO	QTY
S1	P1	J1	200
S1	P1	J3	100
S1	P1	J4	700
S1	P2	J2	100
S2	P3	J1	400
S2	P3	J2	200
S2	P3	J4	500
S2	P3	J5	400
S2	P5	J1	400
S2	P5	J2	100
S3	P1	J1	200
S3	P3	J1	200
S4	P5	J1	100
S4	P6	J3	300
S4	P6	J4	200
S5	P2	J4	100
S5	P3	J1	200
S5	P6	J2	200
S5	P6	J4	500

请写出能够完成如下查询的关系代数的式子：

1) 求供应工程 J1 零件的供应商号 SNO。

2) 求供应工程 J1 零件 P1 的供应商号 SNO。

3) 求供应工程 J1 零件为红色的供应商号 SNO。

4) 求没有使用天津供应商生产的红色零件的工程号 JNO。

(注意：S 表和 J 表的 CITY 字段不是一个含义)

5) 求至少用了 S1 供应商所供应的全部零件的工程号 JNO。(明确题意，根据上表查询结果应该为 J4)

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?

一. 第八次课教学任务

本次课继续学习 SQL 语言的重要知识点嵌套查询。

1. 请前往网址：<https://www.bilibili.com/video/BV13J411J7Vu?p=21>

学习 P21 节。学习完后，学习同一网页中 P22 节，22 节视频只用学习到 4 分 45 秒。

这部分内容请配套学习上传的课件《数据库系统概论-计科-第三章_3.pdf》。

课件展开思路及重点：

1) 什么是嵌套查询？

2) 嵌套查询分为哪两种类型：相关子查询、非相关子查询，分别是什么解算思路？

3) 嵌套查询中常用的 4 种子查询：3 种谓词作为子查询和单纯比较运算符的子查询。

分别用例子示范讲解 IN 谓词子查询怎么用？比较运算符的子查询怎么用？ANY/SOME/ALL 谓词的子查询怎么用？ANY/SOME/ALL 与聚集函数、IN 谓词的等价转换关系是什么？

EXISTS 谓词的子查询怎么用？（见下面步骤 2 的描述）

2. 然后，请同学们下载并学习群里上传的《EXISTS 子查询.mp4》视频文件（注意不要下成《人大-例 3.62-EXISTS 查询动画.mp4》）。EXISTS 子查询.mp4 是我提前录制好的关于 EXISTS 子查询知识讲解部分的视频，该视频是为同学们自己参考学习的，若同学想将该视频发给非本课程班上的人或者公开发到网上，请事先征得我的同意。录这个视频的初衷是：因为这部分知识点较难，一般同学初学的时候都有很多困难，而 MOOC 这部分讲得并不深入，所以我录制了一个关于 EXISTS 子查询讲解很详细的视频，请同学们认真学习。我录制的视频中有用到人民大学制作的一个动画，文件名为《人大-例 3.62-EXISTS 查询动画.mp4》，该视频也上传到群中，大家可以自行下载观看学习。

我录制的视频部分请大家配套上传的课件《数据库系统概论-计科-第三章_4.pdf》学习。

这部分内容很难的知识点为：如何使用 EXISTS 谓词表达全称量词、逻辑蕴涵的表达。请大家结合例子深入理解。

对嵌套查询做一个如下的简要总结：

嵌套查询总结

问题 1：WHERE 的判断单元是什么？

元组

问题 2：子查询的结果是什么？

表（集合）

问题 3：子查询所得结果在哪些情况下是一个属性值？

- 集合中只有一个元素
- 聚集函数

外层WHERE条件中，究竟可以写入什么内容？

1. 带有IN谓词的子查询 判断元素与集合的属于关系
2. 带有比较运算符的子查询 比较元素与元素
3. 带有ANY或ALL谓词的子查询 比较元素与集合
4. 带有EXISTS谓词的子查询 判断集合是否非空

一. 第九次上课任务

本次课程将是第三章 SQL 语言的最后一次课程，涉及到教材的 P115-129。今天需要学习下面网页：<https://www.bilibili.com/video/BV13J411J7Vu?p=23> 的 P23-P27，共 5 节内容。大家首先学习 P23 节集合查询，然后请在群里下载 3.4.5-派生表查询.mp4（仅供同学们自己学习使用，若想发送至网上，需要先征得我的同意），学习 3.4.5 节，因为 3.4.5 节 MOOC 中没有讲授。看完我录制的视频后继续学习网页中的 P24-27 这 4 节内容。本次课涉及的课件为“数据库系统概论-计科-第三章_5.pdf”，请结合视频对课件学习。

本次课件展开的大纲及重点是：

1. SQL 语言的数据查询部分的收尾：集合查询、基于派生表的查询。

1) 集合查询部分同学们学习时需要和第二章的关系代数联系起来，SQL 语言中支持一些传统集合运算，那么这些关键词分别是什么含义，并、交、差的关键词分别对应什么？UNION 与 UNION ALL 有什么区别？

2) 派生表的子查询的核心为在父查询的 FROM 子句中嵌入一个子查询块，这个子查询块形成了一个临时派生表，在父查询的筛选条件中比如 WHERE 子句后面可以利用这个临时的派生表完成某些查询任务的筛选。

3) 大家注意总结学习 3.4.6 节，SELECT 语句的一般形式，它将前面数据查询 3.4 节的所有讲到的知识点进行了一个总结。

2. 数据更新：插入、修改、删除。

1) 数据的增删改就意味着要对基本表的进行变化，变化时需要满足什么约束性条件？

2) 插入两种形式：插入元组、插入子查询结果的命令怎么写？插入的 INTO 子句什么时候要指定插入属性列或者部分属性列，或者什么时候不指定属性列？VALUES 子句的值与 INTO 子句中属性需要满足什么要求，怎么匹配？

3) 修改语句中 WHERE 子句有没有会带来什么不同的影响？怎样进行带有子查询的修改？

4) 删除用的语句 DELETE，与我们前面数据定义部分基本表、视图、索引的删除 DROP 有什么区别？带有子查询的删除语句怎么书写？删除多个元组和单个元组分别怎么写？能删除某个元组内某属性的值么？

3. 空值的处理

空值怎样产生的？空值一般表示什么？有些查询中还要结合题意考虑查询等于空值的情况。

4. 视图，视图的讲解也是与我们前面学的 3.4-3.5 节针对基本表的 SQL 一样的结构展开：定义、查询、更新、视图的作用

1) 视图是什么？视图跟基本表有什么区别和联系？（重点理解视图是虚表，什么是虚表？）

2) 视图的定义：建立与删除。

- 怎么建立视图？视图建立语句中属性值什么时候省略、什么时候需要明确指定？建立语句中 WITH CHECK OPTION 的含义是什么？什么是行列子集视图？视图的建立讲解分为：基于单表的视图、基于多表的视图、基于视图的视图。另一种形式：带表达式的视图、分组视图。

- 删除视图，标准库的 SQL 语言中对视图的删除中级联删除（CASCADE）什么时候

使用？

3) 视图的查询：视图的查询本质上要转换为基本表的查询，因为视图是一个虚表，而这个转换的过程要用到视图的消解法，视图的消解法是什么含义？所有的视图都能正确查询么？什么类型的视图查询采用视图消解法会有局限？

4) 视图的更新

视图的更新要转换为对基本表的更新。所有的视图都能更新么？什么类型的视图更新有限制？这里跟前面行列子集视图的含义以及分组视图的定义联系起来。

5) 视图的作用

通过前面的学习，根据视图的特点总结视图的 5 大作用，通过例子去体会。