

1. 在教学管理数据库中，有如下三个关系表

学生信息表：S(S#, SNAME, AGE, SEX)

课程表：C(C#, CNAME, TEACHER)

选课表：SC(S#, C#, GRADE)

其中 S#、C#为 S、C 表的主键，(S#, C#)是 SC 表的主键，也分别是参照 S、C 表的外键。

请使用关系代数表达式回答下列问题：

- 1) 查询同时学习过课程号为 001 和 002 课程的学生的学号和成绩
- 2) 查询学习过课程号为 001 的学生学号、姓名、及这门课的成绩
- 3) 查询没有学习过课程号为 002 的学生姓名和年龄
- 4) 查询老师 gao 所教过的学生中成绩为 90 分以上(包括 90 分)的学生姓名
- 5) 查询选修了全部课程的学生姓名

2. 在工程管理数据库中，包括 S, P, J, SPJ 四个关系模式：

S(S#, SNAME, SCITY)

P(P#, PNAME, COLOR)

J(J#, JNAME, JCITY)

SPJ(S#, P#, J#, QTY)

供应商表 S 由供应商代码 (S#)、供应商姓名 (SNAME)、供应商所在城市 (SCITY) 组成；零件表 P 由零件代码 (P#)、零件名 (PNAME)、颜色 (COLOR) 组成；

工程表 J 由工程代码 (J#)、工程名 (JNAME)、工程所在城市 (JCITY) 组成；供应情况表由 SPJ 由供应商代码 (S#)、零件代码 (P#)、工程代码 (J#)、零件供应数量 (QTY) 组成。

其中 S#、P#、J#分别是 S、P、J 表的主键，(S#, P#, J#) 是 SPJ 的主键，也分别是参照 S、P、J 表的外键。

请使用关系代数表达式回答下列问题：

- 1) 检索使用了在北京且供应商名字为“S1”的供应商生产的蓝色零件的工程代码
- 2) 检索供应商与工程所在城市相同的工程代码和工程名
- 3) 检索长春的任何工程都不使用的零件代码
- 4) 检索使用了零件号=“P2”的零件的工程代码及工程名
- 5) 检索为工程代码=“J5”的工程供应绿色零件的供应商代码和供应商姓名

3. 设属性 K 是关系 R 的主键，关系 S 的外键 F 参照 R.K，写一个关系代数表达式来验证 R 和 S 的实例是否违反参照完整性约束，说明如何用该关系代数表达式的结果来验证。

4. 在关系代数运算器 (<https://dbis-uibk.github.io/relax>) 上加载数据集 “Database Systems The Complete Book - Exercise 2.4.1”，1) 用关系代数表达式表示下列查询，并用关系代数运算器进行验证；2) 使用域关系演算完成三个查询。

- a) Find the laptop model with the lowest price.
- b) Find the screen sizes that occur in two or more laptops.
- c) What manufacturers make both ink-jet and laser printers?

5. 在企业管理数据库中，存在如下两个关系表

员工信息表：Employee(E#, NAME, SALARY, D#)

部门信息表：Department(D#, Dname)

其中 E#为 Employee 表的主键，D#为 Department 表的主键和 Employee 表的外

键

员工信息表由 E# (员工 ID), NAME (员工姓名), SALARY (工资), D# (所属部门 ID) 组成;

部门信息表由 D# (部门 ID), NAME (部门名称) 组成;

请使用 SQL 语言回答下列问题:

- 1) 一号部门 (D# = 1) 员工的个数
- 2) 查询每个部门的部门 ID 和员工数量
- 3) 查询“技术部”员工工资超过 10000 的员工姓名
- 4) 查询所有部门的平均工资, 返回部门 ID 和平均工资 (avgSalary)
- 5) “技术部”中姓张的员工的个数

6. 在图书管理数据库中, 存在以下三个关系表:

图书信息表: Book(B#, Title, Publisher)

学生信息表: Student(Sno#, Sname, Sage, Sdept)

借阅信息表: Borrow(B#, Sno#, Time)

其中 B#、Sno# 为 Book、Student 表的主键, (B#, Sno#) 是 Borrow 表的主键, 也分别是参照 Book、Student 表的外键。

图书信息表由书号 (B#)、书名 (Title) 和出版社 (Publisher) 组成, 同一书名的书有多本, 且书号各不相同;

学生信息表由学号 (Sno#)、姓名 (Sname)、年龄 (Sage) 和所在系 (Sdept) 组成;

借阅信息表由书号 (B#)、学号 (Sno#)、借阅时长 (Time) 组成。

请使用 SQL 语言回答下列问题:

- 1) 查询借阅了超过 5 本书的学生学号
- 2) 查询借阅了“人民教育出版社”出版的书籍的学生姓名和年龄, 按年龄降序排列
- 3) 查询借阅的所有图书的借阅时长都超过 90 天的学生学号
- 4) 查询书名包含“Big%Date”的图书书名和对应的数量
- 5) 查询超过 5 名“CS”系的不同学生借阅的书的书名

7. 设有如下实体:

图书: 书号、书名、出版日期、出版社

读者: 借书证号、姓名、年龄、性别、家庭住址

书架: 书架号、房间号

上述实体中存在如下联系:

- a) 一本书只能放在一个书架上, 一个书架可以放多本书;
- b) 一位读者可以借多本书, 一本书只能被一位读者借阅

试完成如下工作:

- 1) 设计该图书管理系统的 E-R 图;
- 2) 将该 E-R 图转换为等价的关系模式表示的数据库逻辑结构。

8. 设有如下实体:

学生: 学号、单位名称、姓名、性别、年龄、选修课名

课程: 编号、课程名、开课单位、任课教师号

教师: 教师号、姓名、性别、职称、讲授课程编号

单位: 单位名称、电话、教师号、教师姓名

上述实体存在如下联系:

- 1) 一个学生可以选多门课程, 一门课程可被多名学生选修

- 2) 一个教师可讲授多门课程，一门课程可由多个教师讲授
- 3) 一个单位可有多个教师或学生，一个教师或学生只属于一个单位
- 4) 一个单位可开设多门课程，一门课程只能属于一个单位完成如下工作：
设计该系统的 ER 图，并写出对应的关系模式，标明主键

9. 工厂需建立一个管理数据库存储以下信息：

- 1) 工厂：厂名、厂长姓名；
- 2) 车间：车间号、车间主任姓名、地址、电话；
- 3) 仓库：仓库号、仓库主任姓名、电话；
- 4) 零件：零件号、重量、价格；
- 5) 产品：产品号、价格；

上述实体存在如下联系：

- c) 一个工厂内有多个车间和多个仓库，一个车间或一个仓库都只能属于一个工厂；
- d) 一个车间生产多种产品，每种产品只能产自一个车间；
- e) 一个车间生产多种零件，一种零件也可能为多个车间所制造；
- f) 一个产品由多种零件组成，一种零件也可装配出多种产品；
- g) 产品和零件均存入仓库。

根据上述要求，完成如下工作：

画出该系统的 E-R 图，并写出对应的关系模式，标明主键。

一、设有关系模式 $R(A,B,C,D,E,F)$ ，其函数依赖关系为

$F=\{AB\rightarrow C, BC\rightarrow AD, D\rightarrow E, CF\rightarrow B, AB\rightarrow D\}$

1.求 $(AB)_F^+$

2.求 R 的极小函数依赖集

3.求 R 的候选码

4. 将 R 进行分解，使其成为具有无损连接和保持函数依赖的 3NF

二、考虑关系模式 $R(A, B, C, D, E, F)$ ，其函数依赖集为 $F=\{AC\rightarrow EF, B\rightarrow CE, C\rightarrow B, AB\rightarrow D, E\rightarrow F\}$

1)求解 R 的最小函数依赖集

2)列举 R 的全部候选键

三、假设有以下关系模式 $R: R(A,B,C,D,E,F,G)$

其中，以下依赖关系成立：

$\{A\} \rightarrow \{B,C\}$

$\{B,C\} \rightarrow \{D,E\}$

$\{D\} \rightarrow \{F\}$

$\{E\} \rightarrow \{G\}$

请回答以下问题：

1.求出 R 的函数依赖集和候选码。

2.判断 R 是否符合第三范式。

四、有一个房屋租赁信息表：

$Rental(CNo, CName, PNo, PAddr, STime, ETime, Rent, ONo, OName, OPhone)$

表中各属性的含义为：

CNo	CName	PNo	PAddr	STime	ETime
客户编号	客户名	房屋编号	房屋地址	租赁起始时间	租赁结束时间
Rent	ONo	OName	OPhone		
租金	房东编号	房东名称	房东电话		

这些数据有如下语义：

- CNo 唯一标识一个客户，PNo 唯一标识一个房屋，ONo 唯一标识一个房东。
- 一个客户可以租赁多个房屋，一个房屋可以被多个客户租赁。
- 一个客户租赁一个房屋具有一个开始时间和一个结束时间。
- 每个房屋对应一个地址，每个地址可以有多个房屋。
- 每个房屋有一个租金，不同的房屋可以有相同的租金。
- 一个房东可以拥有多个房屋，每个房屋只能属于一个房东。
- 一个房东只能有一个电话，每个电话只能属于一个房东。

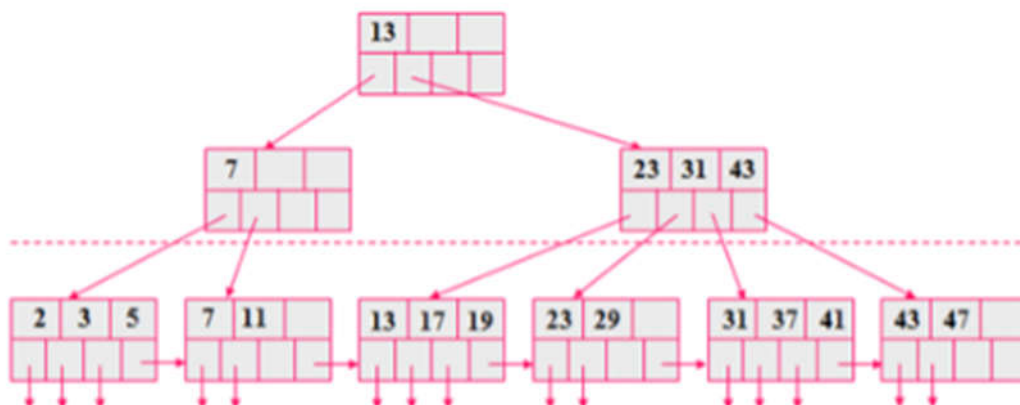
(1) 根据上述语义写出关系 $Rental$ 的函数依赖集 F ，并求 F 的极小函数依赖集 F_m 。

(2) 判断关系 $Rental$ 所达到的最高范式等级。

(3) 将 $Rental$ 分解为具有无损连接性和保持函数依赖的 3NF。

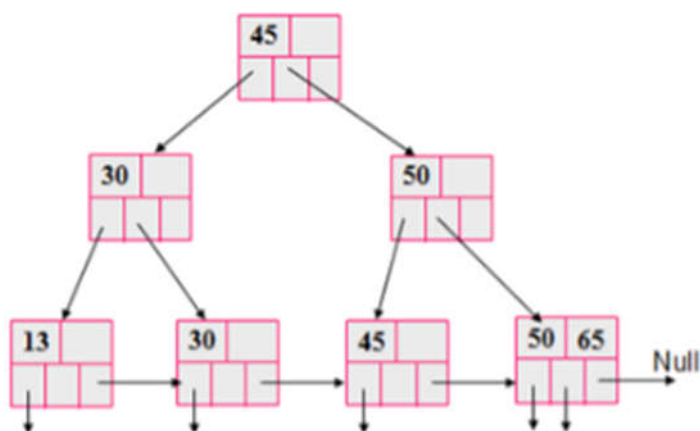
五、

1.已知一棵 B+树，如下图所示



请画出插入 38 后所得的 B+ 树。

2. 已知一棵 B+ 树，如下图所示



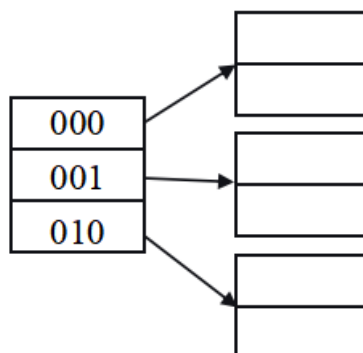
请画出删除 30 后的 B+ 树。

六、利用线性 hash 方法对以下记录进行 hash 存储：

5, 19, 24, 26, 35, 48, 53, 61

设 hash 函数 $h(x) = x \bmod 16$ ，请画出添加以上所有元素后，最终的索引结构以及关键步骤（进行桶的线性增长时）的索引结构。

注：线性 hash 表中最多容纳 $nb\theta$ 个记录， $\theta = 0.85$ ；



初始哈希桶结构为右图。

七、利用可扩展 hash 方法对以下记录进行 hash 存储：

3, 16, 20, 21, 30, 44, 47, 61

设 hash 函数 $h(x) = x \bmod 16$ ，其中散列函数 $h(k)$ 是一个 b (足够大) 位二进制序列，序列的前 d 位用作索引，来区分每个元素属于哪个桶。

现要求每个桶至多包含 2 个元素，以上元素按从左往右的顺序依次添加。开始时只使序列的前 1 位作索引（即 $d=1$ ），当桶满时进行分裂， d 相应增大。请画出添加完以上所有元素后，最终的索引结构。

1. 已知 2 个关系 $R(A, B)$ 和 $S(B, C)$ ，其主键分别为 $R.A$ 和 $S.B$ 。 R 有 20000 个元组， S 有 60000 个元组，一块中可以容纳 20 个 R 元组或 30 个 S 元组。设 2 个关系均采用聚簇存储，且每个关系中的元组均已按照其主键值递增排序。现在要执行自然连接操作 $R \bowtie S$ 。设缓冲区中可用内存页数为 $M = 41$ 。回答下列问题：

(1) 采用嵌套循环连接算法执行 $R \bowtie S$ 分别需要进行多少次 I/O？给出具体分析过程。

(2) 采用归并连接算法执行 $R \bowtie S$ 分别需要进行多少次 I/O？给出具体分析过程。

(3) 设 $R.B$ 是关系 R 的外键，参照 $S.B$ 。如果 $R \bowtie S$ 的结果中元组的平均大小是 R 中元组平均大小的 1.2 倍， $R \bowtie S$ 的结果中元组的平均大小是 S 中元组平均大小的 2 倍，那么在外存中存储 $R \bowtie S$ 的结果需要占用多少个块（页）？给出具体分析过程。

2. 设关系 $R(X, Y)$ 和 $S(Y, Z)$ ， R 共有 1000 个元组， S 共有 1500 个元组，每个块中可容纳 20 个 R 元组或 30 个 S 元组。 S 中 Y 不同值的个数为 20。

(1) 若在 $S.Y$ 上建有聚簇索引，估计 R 和 S 基于索引连接的 IO 代价。

(2) 若在 $S.Y$ 上建有非聚簇索引，估计 R 和 S 基于索引连接的 IO 代价。

3. 设教学管理数据库有如下 3 个关系模式：

$S(S\#, SNAME, AGE, SEX)$

$C(C\#, CNAME, TEACHER)$

$SC(S\#, C\#, GRADE)$

其中 S 为学生信息表、 SC 为选课表、 C 为课程信息表； $S\#$ 、 $C\#$ 分别为 S 、 C

表的主码， $(S\#, C\#)$ 是 SC 表的主码，也分别是参照 S 、 C 表的外码

用户有一查询语句：

Select SNAME

From S, SC, C

Where $SC.S\#=S.S\#$ and $SC.C\#=C.C\#$ and CNAME= “数据库”

检索选学“数据库”课程的学生的姓名。

(1) 写出以上 SQL 语句所对应的关系代数表达式。

(2)画出上述关系代数表达式所对应的查询计划树。使用启发式查询优化算法，对以上查询计划树进行优化，并画出优化后的查询计划树。

(3)设 SC 表有 10000 条元组，C 表有 50 条元组，S 表中有 1000 条元组，SC 中满足选修数据库课程的元组数为 150，计算优化前与优化后的查询计划中每一步所产生的中间结果大小

4. 给定以下关系模式，

```
Rabbit (rid, name, color)
Carrot (cid, weight, type)
Allocation (rid, cid, date)
```

(1) 考虑以下的 SQL 查询语句，绘制其查询计划树。

```
SELECT C.type
FROM Rabbit R, Carrot C, Allocation A
WHERE R.rid = A.rid
AND C.cid = A.cid
AND R.name = 'Tutu'
AND C.weight > 100
AND C.weight < 500
```

(2) 假设在 Rabbit.name 和 Allocation.rid 上建有索引，绘制优化后的查询计划树。

5. 已知一个关系数据库的模式如下：

关系 B(bno, bname, author)为图书表，其中 bno 为书号，bname 为书名，author 为作者；

关系 S(sno, sname, dept)为学生表，其中 sno 为学号，sname 为姓名，dept 为学生所在系；

关系 L(sno, bno, date)为借书表，其中 sno 为学号，bno 为书号，date 为借书时间。

回答下列问题：

(1) 绘制下面的 SQL 查询语句的逻辑查询计划树。

```
SELECT bname FROM B NATURAL JOIN S NATURAL JOIN L
WHERE date = '2021-06-04' AND college = 'CS';
```


(2) 使用启发式查询优化方法对上面的逻辑查询计划树进行优化，绘制优化后得到的逻辑查询计划树，具体说明你进行这些优化的理由。

6. 设 $ri(X)$ 与 $wi(X)$ 分别表示事物 T_i 读、写数据单元 X ，则一个并发调度可以抽象为读、写串。基于上述表示，请判断下面两个并发调度是否是可串行化的，为什么？

调度 S ：

$r_2(A); r_1(B); w_2(A); r_3(A); w_1(B); w_3(A); r_2(B); w_2(B)$

调度 S' ：

$r_2(A); r_1(B); w_2(A); r_2(B); r_3(A); w_1(B); w_3(A); w_2(B)$

7. 已知下面两个事务：

```
T1:
read(A);
read(B);
if A > B then B := A;

T2:
read(B);
read(A);
if B < 0 then A := B * B;
```

回答下列问题：

- (1) 简述两段锁协议（简称 2PL），并给事务 T_1 和 T_2 添加加锁和解锁指令，使其遵从 2PL 协议。
- (2) 按照你添加的加锁和解锁指令，给出 2PL 下对 T_1 和 T_2 的一个正确的调度。
- (3) 什么情况下， T_1 和 T_2 会发生死锁？请给出一个会发生死锁的调度。
- (4) 在数据库的实际应用中经常会遇到的与锁相关的异常情况，请简述一下常见的死锁检测方法

8. 已知某数据库采用即时更新方法（undo-redo 方法）记录 WAL 日志。设故障发生时 WAL 日志文件内容如下：

```

<T1, begin>
<T1, A, 114, 114514>
<T2, begin>
<T1, B, "hit", "hits">
<T1, commit>
<T3, begin>
<T3, B, "hits", "hitsdb">
<T2, A, 114514, 1919810>
    
```

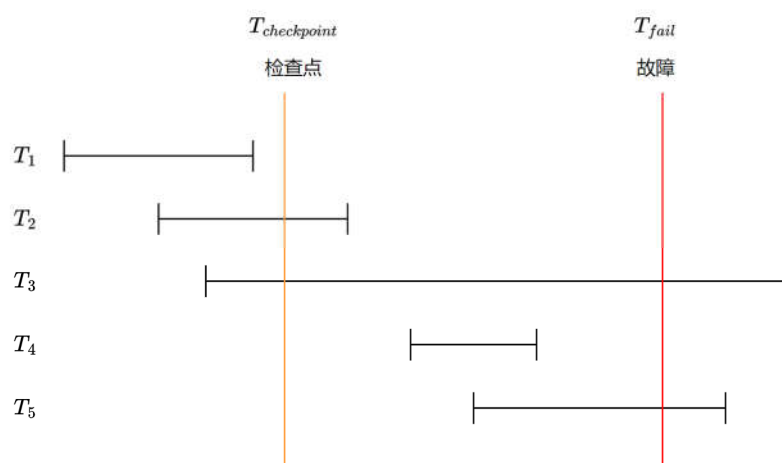
当系统重启后，DBMS 基于该 WAL 日志文件进行故障恢复。回答下列问题：

(1) 该数据库系统的日志恢复策略为 Undo/Redo 型，那么其对应的缓冲区处理策略是什么？该策略的每一项的具体内容都有什么？

(2) 当 DBMS 进行故障恢复时，需要对那些事务进行 undo？对那些事务进行 redo？给出具

(3) 当故障恢复完成时，对象 A 和 B 的值分别是什么？描述故障恢复的具体过程。

9. 使用检查点的数据库恢复系统将根据事务的状态和检查点的关系采取相对应的恢复策略，现在有事务 T_1 - T_5 其执行过程如下图所示（线段左端和右端分别表示事务开始与提交），其执行过程中数据库系统发生如图所示的故障，请回答下列问题



(1) 请问在故障恢复时事务 T_1 - T_5 那些需要撤销，那些需要重做，那些不需要操作？

(2)事务 T_6 - T_8 的日志文件如下图所示,< T_i , begin>表示事务 T_i 开始执行, < T_i , commit>表示事务 T_i 提交, < T_i , D, V_1 , V_2 >表示事务 T_i 将数据项 D 的值由 V_1 修改为 V_2 , <crash>表示数据库发生故障

```
<T6, begin>
<T6, X, 100, 1>
<T7, begin>
<T7, X, 1,3>
<T8, begin>
<T7, Y, 50, 6>
<T8, Y, 6, 8>
<T8, Z, 10, 9>
<checkpoint>
<T6,commit>
<T8,Z,9,10>
<crash>
```

数据库系统发生故障时, 请给出恢复子系统时需要 undo 的事务列表和需要 redo 的事务列表

(3)请简述事务 T_6 - T_8 在系统故障后, 基于检查点的故障恢复过程