

1. 考虑在公共属性 a 上连接关系 $R(a,b)$ 和 $S(a,c,d)$, 假设表上没有可用的索引来加速连接算法。

缓冲区中有 $B=75$ 页

表 R 包含 $M=2400$ 个页面, 每个页面包含 80 个元组

表 S 包含 $N=1200$ 个页面, 每个页面包含 100 个元组

请回答以下关于计算连接的 I/O 开销的问题。您可以假设最简单的开销模型, 即每次只读写一个页面。您还可以假设你需要一个缓冲块来保存演变中的输出块, 以及一个输入块来保存内部关系的当前输入块。你可以忽略写最终结果的成本。

A. 以 S 为外部关系, R 为内部关系的散列连接。您可以忽略递归分区和部分填充的块。划分阶段的成本是多少?

B. 以 R 为外部关系, S 为内部关系的块嵌套循环连接:

C. 以 S 为外部关系, R 为内部关系的块嵌套循环连接:

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 50 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. 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 40000 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。

4. 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

5. 给定以下关系模式，

Student (sid, sname, major)

Course (cid, cname, credit)

Enrollment (sid, cid, grade)

(1) 考虑以下的 SQL 查询语句，绘制其查询计划树。

SELECT C.name

FROM Student S, Course C, Enrollment E

WHERE S.sid = E.sid

AND C.cid = E.cid

```
AND S.major = 'Computer Science'
AND C.credit >= 90;
```

(2) 假设在 Student.major 和 Enrollment.sid 上建有索引, 绘制优化后的查询计划树。

6. 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 author FROM B NATURAL JOIN S NATURAL JOIN L
WHERE date = '2023-11-01' AND dept = 'Math';
```

(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。

将选择操作 $\sigma_{dept='Math'}$ 和 $\sigma_{date='2023-11-01'}$ 以及投影操作 $\Pi_{b.author}$ 进行下推可以尽早减少中间结果的大小。

T 元组数
B 块数
M 子数缓冲区
V 不同值数

1. 考虑在公共属性 a 上连接关系 R(a,b)和 S(a,c,d), 假设表上没有可用的索引来加速连接算法。

缓冲区中有 B=75 页

表 R 包含 M= 2400 个页面, 每个页面包含 80 个元组

表 S 包含 N = 1200 个页面, 每个页面包含 100 个元组

请回答以下关于计算连接的 I/O 开销的问题。您可以假设最简单的开销模型, 即每次只读写一个页面。你还可以假设你需要一个缓冲块来保存演变中的输出块, 以及一个输入块来保存内部关系的当前输入块。你可以忽略写最终结果的成本。

A. 以 S 为外部关系, R 为内部关系的散列连接。您可以忽略递归分区和部分填充的块。划分阶段的成本是多少?

B. 以 R 为外部关系, S 为内部关系的块嵌套循环连接:

C. 以 S 为外部关系, R 为内部关系的块嵌套循环连接:

A. 对 R 扫描一次 $2400 \times 2 = 4800$ 次 I/O
对 S 扫描一次 $1200 \times 1 = 1200$ 次 I/O
则成本为 7200 次 I/O

B. 外联 R 读一次, 产生 2400 次 I/O
内联 S 扫描 $\lceil 2400 / (75 - 1) \rceil = 33$ 次, 一共 $33 \times 1200 = 39600$ 次 I/O
∴ 一共产生 42000 次 I/O

C. 外联 S 读一次, 产生 1200 次 I/O
内联 R 扫描 $\lceil 1200 / (75 - 1) \rceil = 17$ 次, 一共 $17 \times 2400 = 40800$ 次 I/O
∴ 一共产生 42000 次 I/O

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 50 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。

(1) R 的每块读一次, 则一共 $\frac{1000}{20} = 50$ 次 I/O

由于聚簇索引, S 一共 $\frac{1500}{50} = 30$ 块, 平均一个元组能 $\lceil \frac{30}{20} \rceil = 2$ 次 I/O

一共 $1000 \times 2 = 2000$ 次 I/O

\therefore I/O 代价为 2050 次 I/O

(2) R 的每块读一次, 则一共 $\frac{1000}{20} = 50$ 次 I/O

对于 R 的每个元组 r , S 中平均有 $\frac{1500}{20} = 75$ 个元组能连接

最坏情况 $75 \times 1000 = 75000$ 次 I/O

\therefore I/O 代价为 75050

3. 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 40000 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。

$$T(R) = 40000 \quad T(S) = 60000$$

$$B(R) = \frac{40000}{20} = 2000 \quad B(S) = \frac{60000}{30} = 2000$$

$$M = 41$$

(1) ① 若采用基于元组的嵌套循环算法，外层为 S

$$I/O \text{ 次数为 } T(S)(T(R)+1) \approx 6 \times 10^4 \times 4 \times 10^4 = 2.4 \times 10^9$$

② 若采用基于块的嵌套循环算法，外层为 S

$$\text{一共需要 } \lceil \frac{20}{41-1} \rceil = 50 \text{ 轮}$$

$$I/O \text{ 次数为 } B(S) \times 50 \times B(R) = 2000 + 100000 = 102000$$

(2) 设 R 和 S 有序，

R 创建时排序需要每块读字各一次，一共 $2B(R)$ 次，

S 同理，一共 $2B(S)$ 次，

归并阶段 R 和 S 各扫描一次，一共 $B(R) + B(S)$ 次，

$$I/O \text{ 次数为 } 3 \times (2000 + 2000) = 12000 \text{ 次}$$

(3) 由于 $R.B$ 为外键，所以连接后一共 $T(R) = 40000$ 个元组

$$\frac{30}{2} = 15 < \frac{20}{1.2}$$

\therefore 一块可存 15 个 $R \bowtie S$ 元组

$$\therefore \text{有 } \lceil \frac{40000}{15} \rceil = 2667 \text{ 块}$$

4. 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”

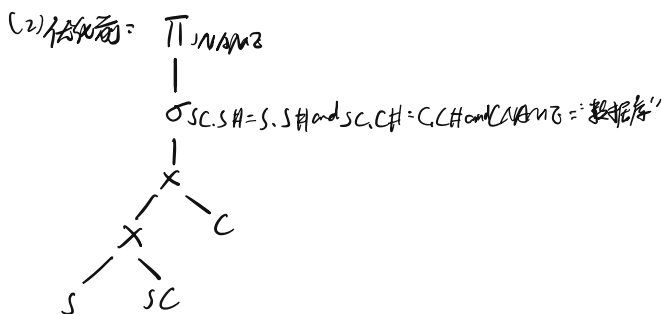
检索选学“数据库”课程的学生的姓名。

(1)写出以上 SQL 语句所对应的关系代数表达式。

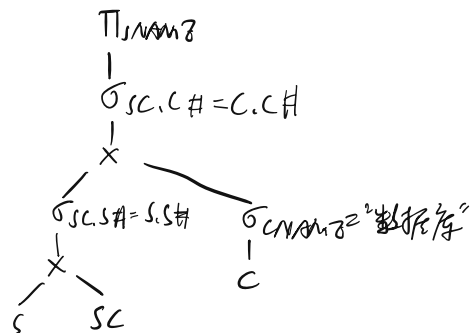
(2)画出上述关系代数表达式所对应的查询计划树。使用启发式查询优化算法，对以上查询计划树进行优化，并画出优化后的查询计划树。

(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

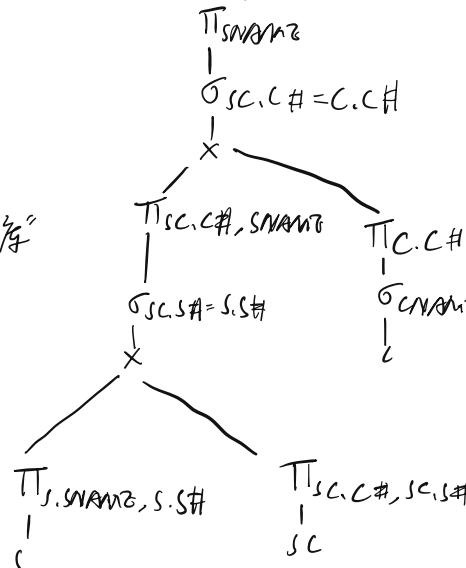
(1) $\pi_{SNAME} \sigma_{SC.S\#=S.S\# \text{ and } SC.C\#=C.C\# \text{ and } CNAME='数据库'} (S \times SC \times C)$



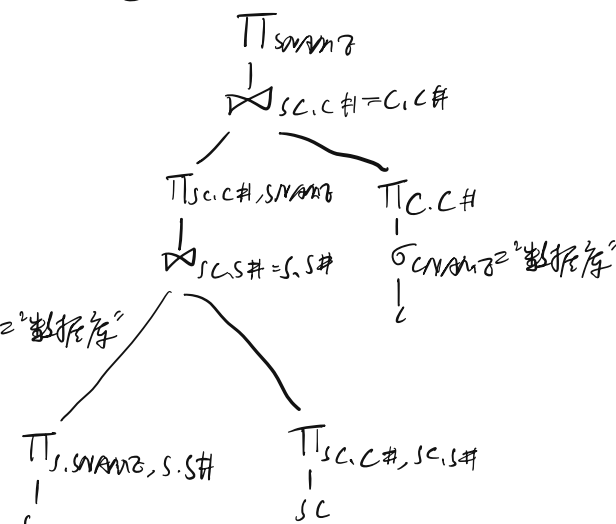
优化-①移动选择操作

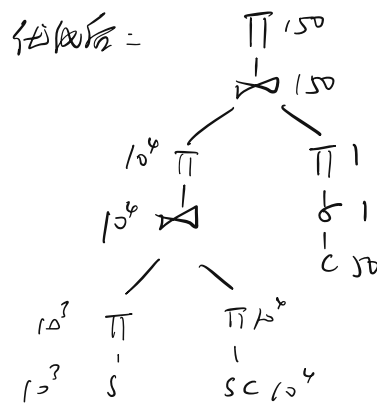
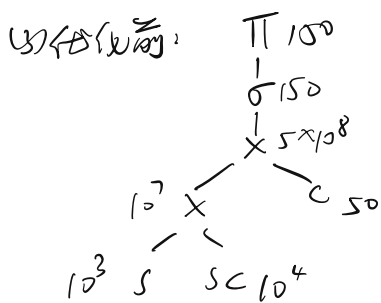


②投影合并



③合并选择和投影操作





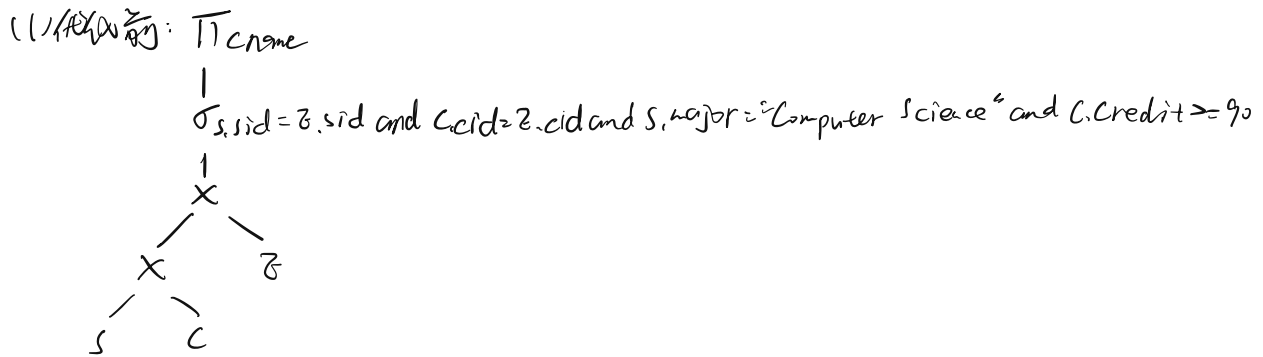
5. 给定以下关系模式,

Student (sid, sname, major)
 Course (cid, cname, credit)
 Enrollment (sid, cid, grade)

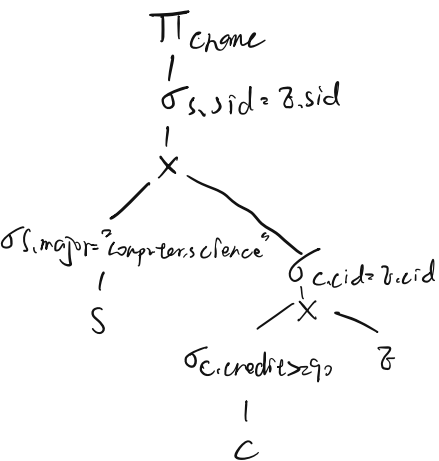
(1) 考虑以下的 SQL 查询语句, 绘制其查询计划树。

```
SELECT C.name
FROM Student S, Course C, Enrollment E
WHERE S.sid = E.sid
AND C.cid = E.cid
AND S.major = 'Computer Science'
AND C.credit >= 90;
```

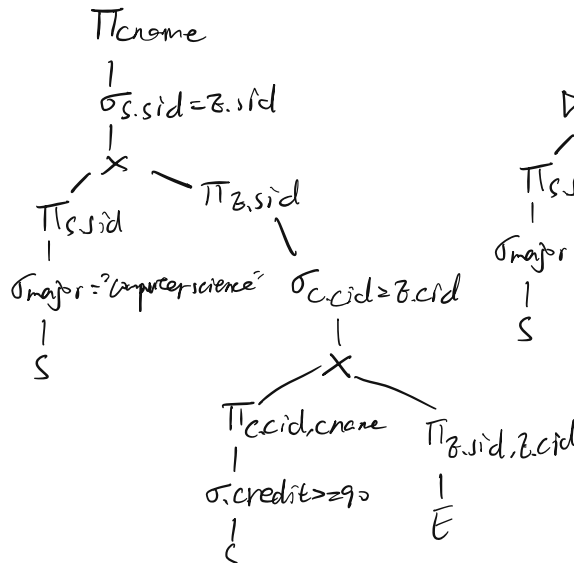
(2) 假设在 Student.major 和 Enrollment.sid 上建有索引, 绘制优化后的查询计划树。



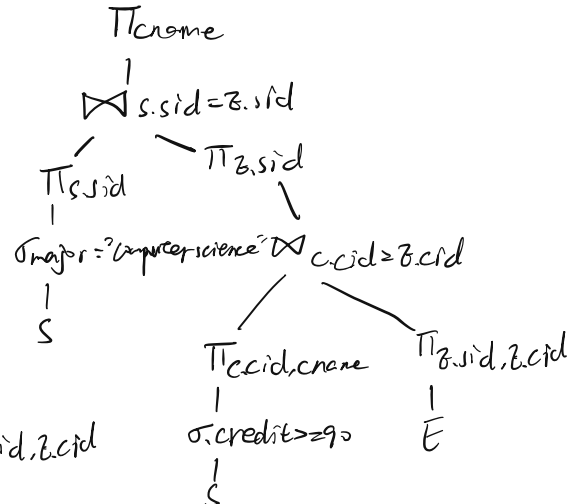
(2) 优化: ① 移动选择操作



② 投影合并



③ 合并选择和笛卡尔积



6. 已知一个关系数据库的模式如下:

关系 $B(\underline{bno}, bname, author)$ 为图书表, 其中 bno 为书号, $bname$ 为书名, $author$ 为作者;

关系 $S(\underline{sno}, sname, dept)$ 为学生表, 其中 sno 为学号, $sname$ 为姓名, $dept$ 为学生所在系;

关系 $L(\underline{sno}, bno, date)$ 为借书表, 其中 sno 为学号, bno 为书号, $date$ 为借书时间。

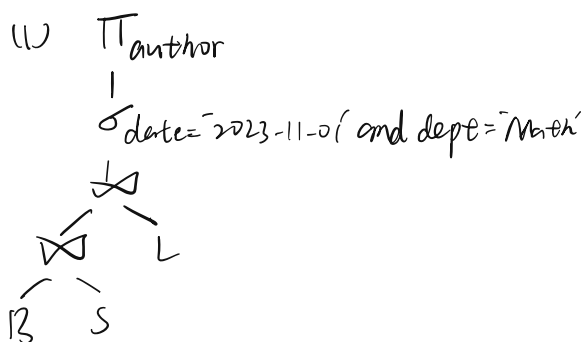
回答下列问题:

(1) 绘制下面的 SQL 查询语句的逻辑查询计划树。

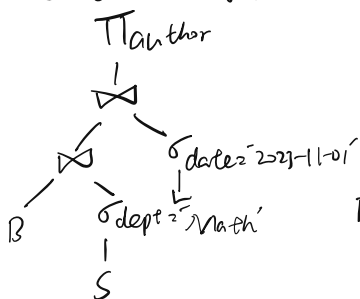
```
SELECT author FROM B NATURAL JOIN S NATURAL JOIN L
WHERE date = '2023-11-01' AND dept = 'Math';
```

(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。

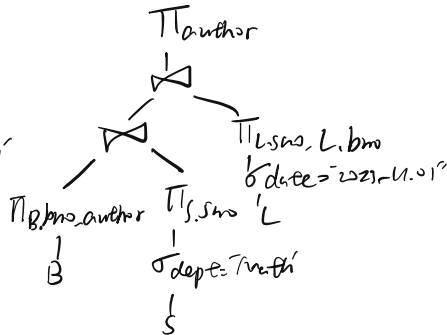
将选择操作 $\sigma_{dept='Math'}$ 和 $\sigma_{date='2023-11-01'}$, 以及投影操作 $\Pi_{b.author}$ 进行下推可以尽早减少中间结果的大小。



(2) ① 将选择操作



② 投影合并



③ 将 S 和 L 连接后投影

