

15.2

$\pi_{T.branchname}((\pi_{branchname,assets}(\rho_T(branch))) \bowtie_{T.assets > S.assets} (\pi_{assets}(\sigma_{branch_city='Brooklyn'}(\rho_S(branch)))))$

15.3

- $r_1(A, B, C)$: 20,000 tuples, 25 tuples/block \rightarrow 800 blocks
- $r_2(C, D, E)$: 45,000 tuples, 30 tuples/block \rightarrow 1500 blocks

a. Nested-loop join

- r_1 as the outer relation
 - **Block transfers:** $B_{r_1} + N_{r_1} \times B_{r_2} = 800 + 20000 \times 1500 = 30000800$
 - **Seeks:** $B_{r_1} + N_{r_1} = 800 + 20000 = 20800$
- r_2 as the outer relation
 - **Block transfers:** $B_{r_2} + N_{r_2} \times B_{r_1} = 1500 + 45000 \times 800 = 36001500$
 - **Seeks:** $B_{r_2} + N_{r_2} = 1500 + 45000 = 46500$

b. Block nested-loop join

- r_1 as the outer relation
 - **Block transfers:** $\lceil B_{r_1}/(M-2) \rceil \times B_{r_2} + B_{r_1} = \lceil 800/(M-2) \rceil \times 1500 + 800$
 - **Seeks:** $2 \times \lceil B_{r_1}/(M-2) \rceil = 2 \times \lceil 800/(M-2) \rceil$
- r_2 as the outer relation
 - **Block transfers:** $\lceil B_{r_2}/(M-2) \rceil \times B_{r_1} + B_{r_2} = \lceil 1500/(M-2) \rceil \times 800 + 1500$
 - **Seeks:** $2 \times \lceil B_{r_2}/(M-2) \rceil = 2 \times \lceil 1500/(M-2) \rceil$

c. Merge join

- **Block transfers:** $B_{r_1} + B_{r_2} = 800 + 1500 = 2300$
- **Seeks:** $\lceil B_{r_1}/B_b \rceil + \lceil B_{r_2}/B_b \rceil = \lceil 800/B_b \rceil + \lceil 1500/B_b \rceil$

If not sorted, add sorting cost:

- **Sort r_1 :** $B_{r_1} \times (2 \times \lceil \log_{M-1}(B_{r_1}/M) \rceil + 2) = 800 \times (2 \times \lceil \log_{M-1}(800/M) \rceil + 2)$
- **Sort r_2 :** $B_{r_2} \times (2 \times \lceil \log_{M-1}(B_{r_2}/M) \rceil + 2) = 1500 \times (2 \times \lceil \log_{M-1}(1500/M) \rceil + 2)$

d. Hash join

Since r_1 is smaller, we use it as the build relation and r_2 as the probe relation

- **If recursive partitioning is not required**
 - **Block transfers:** $3 \times (B_{r_1} + B_{r_2}) + 4 \times n_h = 3 \times (800 + 1500) + 4 \times n_h$
 - **Seeks:** $2 \times (\lceil B_{r_1}/B_b \rceil + \lceil B_{r_2}/B_b \rceil) + 2 \times n_h = 2 \times (\lceil 800/B_b \rceil + \lceil 1500/B_b \rceil) + 2 \times n_h$
- **If recursive partitioning required**
 - **Block transfers:**
 $2 \times (B_{r_1} + B_{r_2}) \times \lceil \log_{M-1}(B_{r_1}) - 1 \rceil + B_{r_1} + B_{r_2} = 2 \times (800 + 1500) \times \lceil \log_{M-1}(800) - 1 \rceil + 800 + 1500$
 - **Seeks:**
 $2 \times (\lceil B_{r_1}/B_b \rceil + \lceil B_{r_2}/B_b \rceil) \times \lceil \log_{M-1}(B_{r_1}) - 1 \rceil = 2 \times (\lceil 800/B_b \rceil + \lceil 1500/B_b \rceil) \times \lceil \log_{M-1}(800) - 1 \rceil$

15.6

- 使用索引定位第一个 *branch_city* 字段值为“Brooklyn”的元组。从此元组开始，跟随指针链直至结束，检索所有符合条件的元组。
- 对于此查询，索引没有作用。我们可以顺序扫描文件，并选择所有 *branch_city* 字段值不为“Brooklyn”的元组。
- 此查询等价于以下查询：

$$\sigma_{(branch_city \geq Brooklyn \wedge assets \geq 5000)}(branch)$$

使用 *branch_city* 索引，我们可以通过从第一个“Brooklyn”元组开始跟随指针链，检索所有 *branch_city* 值大于或等于“Brooklyn”的元组。同时，对每个元组应用额外的条件 ($assets \geq 5000$)。

15.20

块传输与磁盘寻道次数估算（针对 $r_1 \bowtie r_2$ ）

根据练习题 15.3 和 15.19，估算执行连接操作 $r_1 \bowtie r_2$ 所需的块传输与磁盘寻道次数。

给定：

- r_1 占用 800 个块， r_2 占用 1500 个块
- 每个索引叶子块包含 n 个指针
- 内存页数 $M < 800$
- r_1 有 20000 条记录， r_2 有 45000 条记录

步骤 1：索引扫描所需块数

- r_1 的索引叶子块数：

$$B_1 = \left\lceil \frac{20000}{n} \right\rceil$$

- r_2 的索引叶子块数：

$$B_2 = \left\lceil \frac{45000}{n} \right\rceil$$

- 合并连接的索引读取代价：

$$B_3 = B_1 + B_2$$

步骤 2：连接输出元组数量

输出元组数量估算为：

$$n_0 = \left\lceil \frac{20000 \times 45000}{\max(V(C, r_1), V(C, r_2))} \right\rceil$$

连接结果块数（每块能容纳 n_1 条记录）：

$$B_{01} = \left\lceil \frac{n_0}{n_1} \right\rceil$$

合并连接总代价（不包括替换指针）：

$$B_j = B_3 + B_{01}$$

步骤 3：替换 r₁ 的指针

第一次排序访问代价（外部排序）：

$$B_{s1} = B_{01} \times \left(2 \times \left\lceil \log_{M-1} \left(\frac{B_{01}}{M} \right) \right\rceil + 2 \right)$$

读取 r₁ 的数据块数：

$$\min(800, n_o)$$

替换后中间结果块数（每块 n₁ 对形式为 (r₁ 元组, 指针)）：

$$B_{02} = \left\lceil \frac{n_o}{n_1} \right\rceil$$

替换 r₁ 指针的总代价：

$$B_f = B_{s1} + B_{01} + \min(800, n_o) + B_{02}$$

步骤 4：替换 r₂ 的指针

第二次排序代价：

$$B_{s2} = B_{02} \times \left(2 \times \left\lceil \log_{M-1} \left(\frac{B_{02}}{M} \right) \right\rceil + 2 \right)$$

替换 r₂ 指针总代价：

$$B_s = B_{s2} + B_{02} + \min(1500, n_o)$$

总磁盘访问次数

$$\text{Total} = B_j + B_f + B_s$$

输出页数

若每页可容纳 n₂ 条最终记录：

$$\left\lceil \frac{n_o}{n_2} \right\rceil$$

