

15.2 解:

$\pi_{T.branch_name} ((\pi_{branch_name, assets} (P_T(branch)))$
 $\bowtie_{(T.assets > S.assets)}$
 $(\pi_{assets} (\sigma_{branch_city='Brooklyn'} (P_S(branch)))$

通过减少连接中涉及的属性可以提升查询效率,
体现在表达式中即使用投影将必要的属性留下参与连接,
其余除去。

15. 解 a. 嵌套-循环连接. 最坏情况下

r_1 有 800 个块,

r_2 有 1500 个块。

将 r_1 视作外层关系, 需要

$$20000 \times 1500 + 800 = 30000800 \text{ 次块传输}$$

$$20000 + 800 = 20800 \text{ 次扫描.}$$

b. 块嵌套-循环连接.

假设内存有 M 个块: 需要

$$r_1 \text{ 为外层关系: } \lceil \frac{800}{M-2} \rceil \times 1500 + 800 \text{ 次块传输}$$

$$2 \times \lceil \frac{800}{M-2} \rceil \text{ 次扫描.}$$

c. 归并连接, 每个总分配 1 个缓冲块, 内存有 M 个块

假设 r_1, r_2 并未排序, 需排序

$$800 \times (2 \lceil \log_{M-1}(\frac{800}{M}) \rceil + 2) + 1500 \times (2 \lceil \log_{M-1}(\frac{1500}{M}) \rceil + 2)$$

$$\text{总共 } 800 \times (2 \lceil \log_{M-1}(\frac{800}{M}) \rceil + 2) + 1500 \times (2 \lceil \log_{M-1}(\frac{1500}{M}) \rceil + 2) + 1500 + 800 \text{ 次传输}$$

$$\text{扫描次数: } 2 \lceil \frac{800}{M} \rceil + \lceil 800 \rceil (2 \lceil \log_{M-1}(\frac{800}{M}) \rceil - 1) +$$

$$2 \lceil \frac{1500}{M} \rceil + \lceil 1500 \rceil (2 \lceil \log_{M-1}(\frac{1500}{M}) \rceil - 1) +$$

$$800 + 1500$$

d. ~~散列~~ 散列连接

不考虑散列表的溢出.

不需要递归归区, 即 $m > \sqrt{800}$.

块传输. $3 \times (1500 + 800)$ 次块传输

寻道. $b_{bf} = 1$, $2(\lceil \frac{1500}{1} \rceil + \lceil \frac{800}{1} \rceil) \approx 4600$ 次寻道

需要递归归区 $m \leq \sqrt{800}$.

块传输 $2(1500 + 800) \lceil \log_{m-1}(\frac{800}{m}) - 1 \rceil + 1500 + 800$

寻道. $2(\lceil \frac{1500}{1} \rceil + \lceil \frac{800}{1} \rceil) \lceil \log_{2, m-1}(\frac{800}{m}) \rceil$