

12.2

12.2
 $\Pi_{T.branch_name} ((\Pi_{T.branch, T.assets} (P_T (branch))) \bowtie$
 $T.assets > S.assets (S (branch_city = "Brooklyn") (P_S$
 $(branch))))$

这个表达式尽可能少的数据执行自然连接。将 join 的右侧操作数限制为 Brooklyn 中的那些分支，并从两个操作数中消除不必要的属性。

12.3

b.

12.3 [b]

$$20000/25 = 800$$

$$45000/30 = 1500$$

$$r_1 \text{ 外连接 } : \left\lceil \frac{800}{M-2} \right\rceil \cdot 1500 + 800$$

$$r_2 \sim : \left\lceil \frac{1500}{M-2} \right\rceil \cdot 800 + 1500$$

12.10

12.10

$$40 \text{ MB} / 4 \text{ kB} = 10^4 = M$$

$$40 \text{ GB} / 4 \text{ kB} = 10^7 = br$$

$$\nearrow 5 \times 10^{-3}$$

cost = number of disk seeks \times disk seek cost +
number of block transfers \times block transfer time

$$\Downarrow \\ 25 \times 10^6$$

$$\Downarrow \\ 10^{-4}$$

a. $b_b = 1$

$$\text{cost} = 5002 \times 10^3 \times 5 \times 10^{-3} + 25 \times 10^6 \times 10^{-4} = 27500 \text{ s}$$

$b_b = 100$

$$\text{cost} = 52 \times 10^3 \times 5 \times 10^{-3} + 25 \times 10^6 \times 10^{-4} = 2760 \text{ s}$$

$$b. \lceil \log_{M-1} (br/m) \rceil$$

$$= \lceil \log_{10^4-1} (10^3) \rceil$$

c. 和 a 一样

$$b_b = 1 : 2505.002 \text{ s} \Rightarrow 2506 \text{ s}$$

$$b_b = 100 : 2500.032 \text{ s} \Rightarrow 2501 \text{ s}$$

12.16

a. 使用流水线，对 r 的排序操作的输出被写入缓冲区 B 。当 B 满时，合并连接从 B 来的元组，将它们与从 s 到 B 的元组连接起来，直到 B 为空。此时，恢复排序操作并填充 B 。此过程将继续进行，直到合并联接完成。

b. 如果排序-合并操作是并行运行的，并且内存在两个操作之间平均共享，那么每个操作的内存缓冲区将只有 $M/2$ 帧；

这会增加合并数据所需的运行次数