

12.3 设关系 $r_1(A, B, C)$, $r_2(C, D, E)$ 有如下特性： r_1 有20000个元组， r_2 有45000个元组，一块中可容纳25个 r_1 元组或30个 r_2 元组。估计使用以下连接策略计算 $r_1 \bowtie r_2$ 需要几次块传输和磁盘搜索：

- 嵌套循环连接
- 块嵌套循环连接
- 归并连接
- 散列连接

关系 r_1 需要800个块，关系 r_2 需要1500个块。如果内存大小 M 大于801个块，那么只需要 $800 + 1500 = 2400$ 次块传输和2次磁盘搜索。如果内存大小 M 至多800个块时：

- 如果 r_1 是外层关系，那么需要 $20000 * 1500 + 800 = 30000800$ 次块传输， $20000 + 800 = 20800$ 次磁盘搜索；如果 r_1 是内层关系，那么需要 $45000 * 800 + 1500 = 36001500$ 次块传输， $45000 + 1500 = 46500$ 次磁盘搜索。

- 如果 r_1 是外层关系，那么需要 $\left\lceil \frac{800}{M-2} \right\rceil * 1500 + 800$ 次块传输， $2 * \left\lceil \frac{800}{M-2} \right\rceil$ 次磁盘搜索；如果 r_1 是内层关系，那么需要 $\left\lceil \frac{1500}{M-2} \right\rceil * 800 + 1500$ 次块传输， $2 * \left\lceil \frac{1500}{M-2} \right\rceil$ 次磁盘搜索。

- 如果 r_1 和 r_2 都没有排序，那么需要 $1600 \log_{M-1} \frac{800}{M} + 3000 \log_{M-1} \frac{1500}{M} + 800 + 1500$ 次块传输， $2 \left\lceil \frac{800}{M} \right\rceil + \left\lceil \frac{800}{b_b} \right\rceil * \left(2 \left\lceil \log_{\left\lceil \frac{M}{b_b} \right\rceil - 1} \frac{800}{M} \right\rceil - 1 \right) + 800 + 2 \left\lceil \frac{1500}{M} \right\rceil + \left\lceil \frac{1500}{b_b} \right\rceil * \left(2 \left\lceil \log_{\left\lceil \frac{M}{b_b} \right\rceil - 1} \frac{1500}{M} \right\rceil - 1 \right) + 1500 + \left\lceil \frac{1500}{b_b} \right\rceil + \left\lceil \frac{1500}{b_b} \right\rceil$ 次磁盘搜索（设 b_b 为分配给归并段或者每个关系的缓冲块）；

- 假设没有散列表溢出发生。当内存足够大即不用递归划分时，需要 $3 * (800 + 1500) = 6900$ 次块传输， $2 \left(\left\lceil \frac{800}{b_b} \right\rceil + \left\lceil \frac{1500}{b_b} \right\rceil \right)$ 次磁盘搜索；当内存不足需要递归划分时，需要 $2(800 + 1500)[(\log_{M-1} 800) - 1] + (800 + 1500)$
 $= 4600[(\log_{M-1} 800) - 1] + 2300$ 次块传输， $2 \left(\left\lceil \frac{800}{b_b} \right\rceil + \left\lceil \frac{1500}{b_b} \right\rceil \right) [(\log_{M-1} 800) - 1]$ 次磁盘搜索（设 b_b 为用于划分的缓冲块数量）。