

16337341(朱志儒)数据库系统作业 5

12.1 第一趟: {(emu, 1), (kangaroo, 17), (wallaby, 21)},

{(lion, 8), (platypus, 3), (wombat, 13)},

{(meerkat, 6), (warthog, 4), (zebra, 11)},

{(baboon, 12), (hornbill, 2), (hyena, 9)};

第二趟: {(emu, 1), (kangaroo, 17), (lion, 8), (platypus, 3), (wallaby, 21), (wombat, 13)};

{(baboon, 12), (hornbill, 2), (hyena, 9), (meerkat, 6), (warthog, 4), (zebra, 11)};

第三趟: {(baboon, 12), (emu, 1), (hornbill, 2), (hyena, 9), (kangaroo, 17), (lion, 8), (meerkat, 6), (platypus, 3), (wallaby, 21), (warthog, 4), (wombat, 13), (zebra, 11)}。

12.3 由题可知: r_1 有 800 块, r_2 有 1500 块。

a.嵌套循环连接

	块传输	磁盘搜索
最坏的情况下	$20000 \times 1500 + 800 = 30000800$	$20000 + 800 = 20800$
最好的情况下	$1500 + 800 = 2300$	2

b.块嵌套循环连接

	块传输	磁盘搜索
最坏的情况下	$800 \times 1500 + 800 = 1200800$	$2 \times 800 = 1600$
最好的情况下	$1500 + 800 = 2300$	2

c.归并连接

r_1 和 r_2 已排序:

	块传输	磁盘搜索
最坏的情况下	$800 + 1500 = 2300$	$800 + 1500 = 2300$
最好的情况下	$800 + 1500 = 2300$	2

r_1 和 r_2 未排序:

最坏的情况下:

对 r_1 排序的块传输次数: $800 \times \left(2 \times \left\lceil \log_2 \frac{800}{3} \right\rceil + 1\right) + 800 = 16000$;

磁盘搜索次数: $2 \times \left\lceil \frac{800}{3} \right\rceil + 800 \times \left(2 \times \left\lceil \log_2 \frac{800}{3} \right\rceil - 1\right) + 800 = 14934$;

对 r_2 排序的块传输次数: $1500 \times \left(2 \times \left\lceil \log_2 \frac{1500}{3} \right\rceil + 1\right) + 1500 = 30000$;

磁盘搜索次数: $2 \times \left\lceil \frac{1500}{3} \right\rceil + 1500 \times \left(2 \times \left\lceil \log_2 \frac{1500}{3} \right\rceil - 1\right) + 1500 = 28000$;

归并两个关系的块传输次数: $800 + 1500 = 2300$;

磁盘搜索次数: $800 + 1500 = 2300$;

总的块传输次数: $16000 + 30000 + 2300 = 48300$;

磁盘搜索次数: $14934 + 28000 + 2300 = 45234$;

最好的情况下:

对 r_1 排序的块传输次数: $800 \times (2 \times 1 + 1) + 800 = 3200$;

磁盘搜索次数: $2 + 800 + 800 = 1602$;

对 r_2 排序的块传输次数: $1500 \times (2 \times 1 + 1) + 1500 = 6000$;

磁盘搜索次数: $2 + 1500 + 1500 = 3002$;

归并两个关系的块传输次数: $800 + 1500 = 2300$;

磁盘搜索次数： $800 + 1500 = 2300$ ；

总的块传输次数： $3200 + 6000 + 2300 = 11500$ ；

磁盘搜索次数： $1602 + 3002 + 2300 = 6904$ ；

d.散列连接

不需要递归划分：

	块传输	磁盘搜索
最坏的情况下	$3 \times (800 + 1500) = 6900$	$2 \times 800 + 1500 = 3100$
最好的情况下	$800 + 1500 = 6900$	2

需要递归划分：

	块传输	磁盘搜索
最坏的情况下	$2 \times (800 + 1500) \times \lceil \log_2 1500 - 1 \rceil + 800 + 1500 = 48300$	$2 \times (800 + 1500) \times \lceil \log_2 1500 - 1 \rceil = 46000$
最好的情况下	$800 + 1500 = 6900$	2

12.5 将较小的关系存储在内存中，逐块读取较大的关系，将较大的关系作为外部关系执行

嵌套循环连接，I/O 操作的次数为 $b_r + b_s$ ，内存要求为 $\min(b_r, b_s) + 2$ 页。

13.4 由于 C 是 r_2 的码， r_1 参照 r_2 的外码，所以 $r_1 \bowtie r_2$ 的估计值最大为 1000，又由于 E 是 r_3 的码， r_2 参照 r_3 的外码，所以 $r_1 \bowtie r_2 \bowtie r_3$ 的估计值最大为 1000。

策略：在 r_2 的属性 C 和 r_3 的属性 E 上创建索引，对 r_1 中的每个元组，使用 r_2 的索引查找匹配 r_1 中 C 值最多的元组，使用 r_3 的索引查找至多一个与 r_2 中的 E 唯一匹配的元组。

13.5 由 $V(C, r_1) = 900, V(C, r_2) = 1100$ ，得到 $r_1 \bowtie r_2$ 的两个估计值 $\frac{1000 \times 1500}{900} =$

$1667, \frac{1000 \times 1500}{1100} = 1364$, 取较小者 1364 为 $r_1 \bowtie r_2$ 的估计。由 $V(E, r_2) = 50, V(E, r_3) = 100$,

得到 $r_1 \bowtie r_2 \bowtie r_3$ 的两个估计值 $\frac{1364 \times 750}{50} = 20460, \frac{1364 \times 750}{100} = 10227$, 取较小者 10227 为 $r_1 \bowtie$

$r_2 \bowtie r_3$ 的估计。最佳的连接顺序为 $r_1 \bowtie r_2 \bowtie r_3$ 。

13.6

a. 使用索引查找 building = “Watson” 的第一个元组，从这个元组开始，按照指针链直到结束，检索所有的元组。

b. 按顺序扫描文件并选择 building ≠ “Watson” 的所有元组。

c. 由于

$$\sigma_{\neg(\text{building} < \text{"Watson"} \vee \text{budget} < 50000)}(\text{department}) = \sigma_{\text{building} \geq \text{"Watson"} \vee \text{budget} \geq 50000}(\text{department})$$

使用索引查找 building = “Watson” 的第一个元组，从这个元组开始，按照指针链直到结束，检索所有的元组，选择 budget ≥ 5000 的元组。

13.15 使用 (dept_name, building) 属性上的索引，找到第一个具有 (dept_name = “Music”, building = “Watson”) 属性的元组，然后按照指针检索连续的元组，选择满足条件 (building < “Watson”, budget < 55000) 的元组。

13.16

a. 根据等价规则 1，可得： $\sigma_{\theta_1 \wedge \theta_2 \wedge \theta_3}(E) = \sigma_{\theta_1}(\sigma_{\theta_2 \wedge \theta_3}(E)) = \sigma_{\theta_1}(\sigma_{\theta_2}(\sigma_{\theta_3}(E)))$

b. 根据等价规则 1 和 7.a，可得： $\sigma_{\theta_1 \wedge \theta_2}(E_1 \bowtie_{\theta_3} E_2) = \sigma_{\theta_1}(\sigma_{\theta_2}(E_1 \bowtie_{\theta_3} E_2)) = \sigma_{\theta_1}(E_1 \bowtie_{\theta_3} (\sigma_{\theta_2}(E_2)))$

13.20 找到均匀划分两个直方图区间的最大单位 u，将每个直方图划分为大小为 u 的区间，并假设区间内的值均匀分布。用值落入该区间的行数来代替 n_r 和 n_s，用该区间的不同取值

个数代替 $V(A, r)$ 或 $V(A, s)$ ，然后使用 13.3.3 中的方法计算估计值，最后将每个区间大小估计值加起来得到总的大小估计值。