第四次作业答案

习题：6.3（1）~（4）

6.3 考虑一个包含 10,000 页的关系 Executives(ename:string, title:string, dname:string, address: string)，及针对它的查询 SELECT **DISTINCT** E.title, E.ename FROM Executives E。

如果可用的缓存页数为 10，**并假设该关系的 4 个属性等长度，每个页可存储 10 个元组**。同时采用如下的排序投影算法：初始排序阶段读入关系，并创建只包含 ename/title 属性的排序子表；随后的归并阶段将附带删除重复元组。试回答以下问题：

（1）初始排序阶段将产生的子表数，以及每个子表的平均长度。

（2）计算排序的 I/O 代价。为计算最终的投影，还需要多少额外的 I/O 代价？

（3）如果有 title 上的聚集 B+树索引，则该索引是否能为排序提供更便宜的代价？如果索引是非聚集的，或是一个散列索引，则结果又如何？

（4）如果有 ename 上的聚集 B+树索引，则该索引是否能为排序提供更便宜的代价？如果索引是非聚集的，或是一个散列索引，则结果又如何？

【解答】

（1） B=10，初始阶段将产生 5000 个排序子表，每个子表长度为 10 个页；
读入 10，000 个页，投影后写出 5000 个页，需要总代价=10000＋5000 ＝15000。

（2） 为合并 1000 个子表，我们还需另外 3 个归并阶段，代价为 2*3*5000=30000I/0s

（3） 可合理假设每页可存储 10*4 个 title 属性， B+树至少会有 100,000/(10*4)=2500 个叶节点。因此，扫描 B+树本身至少需要 2500I/0s 代价。利用 title 上的聚集 B+树索引扫描关系的代价为 12500（超过简单堆文件扫描的 10000 次）。利用 title 上的非聚集索引扫描的代价更高，可能会超过 2500＋100000,达到 2500＋100000*10 次 （假定每页元组数 10)。如果散列索引是聚集的且散列桶中直接存储元组，则使用散列索引检索并完成排序代价可能会很好。

（4） 利用 ename 上的聚集 B+树索引，扫描代价为 12500。因为 ename 为主码，扫描 B+树检索出的<ename,title>对不会有重复，不需要在进行排序消除重复，因此，产生查询结果的总估计代价也是 12500,代价远远低于简单排序归并的(15000+30000)次 I/0s。但非聚集 B+树检索所有目标元组的代价可能达到：1500＋10000*10=102500。

习题 6.4（1）~（3）

6.4　考虑连接 $R\bowtie_{R.a=S.b}S$，已知：

- 关系 R 有 10,000 个元组,每页可存 10 个元组,其数据文件为简单堆文件；
- 关系 S 有 2,000 个元组，每页可存 10 个元组，b 是它的主键，其数据文件为简单堆文件；
- 有 52 个可用缓存页。试回答以下问题：

（1）　分别计算采用简单嵌套循环连接、页嵌套循环连接和块嵌套循环连接算法时的代价，实现相应算法需要的最小缓存页数分别是多少？

（2）　若采用排序-归并连接算法，则其代价和需要的最小缓存页数分别是多少？

（3）　若采用散列连接算法，则其代价和需要的最小缓存页数分别是多少？

【解答】

令关系 R 和 S 的总页数分别为 M、N，可用缓存页数为 B，由题中已知条件，有：M=1000、N=200、B=52。

（1）简单嵌套循环连接算法，总代价＝$N+(N*P_R)*M$= 200+(200*10)*1000 = 2000200

需要的最小缓存页数＝ 3。

页嵌套循环连接算法，总代价=N+(N*M)=200200，需最小缓存页数＝3。

块嵌入循环连接，一次可读入外存关系的 B-2 个页，只需扫描内层关系⌈200/50⌉＝4

总代价=N+ M*⌈200/50⌉ = 200＋1000×4＝4200，需要的最少主存数＝52。

（2） 若 $B>(M)^{1/2}>(N)^{1/2}$，我们可以使用改进的排序-归并算法：排序阶段划分R为20个子表（每个子表50个页），划分S为4个子表，每个近似50页。这24个子表可一次完成归并。另外，需留下一个缓存页作为输出。

总代价＝(M+N)*3=3600；最少需要的缓存页数是25。

注意，如果S.b不是一个键，则在最坏情况下，排序-归并算法的归并阶段可能需要探测整个关系，这会导致在归并阶段产生M*N次I/O。

（3）若 $B>(f*N)^{1/2}$ ，f 主存散列因子，则散列连接算法的

总代价＝(M+N)*3=3600

若假设 f=1.2，最少需要的缓存页数＝$(1.2*500)^{1/2}$=＝25