15.1 Assume (for simplicity in this exercise) that only one tuple fits in a block and memory holds at most three blocks. Show the runs created on each pass of the sort-merge algorithm when applied to sort the following tuples on the first attribute: (kangaroo, 17), (wallaby, 21), (emu, 1), (wombat, 13), (platypus, 3), (lion, 8), (warthog, 4), (zebra, 11), (meerkat, 6), (hyena, 9), (hornbill, 2), (baboon, 12).

解：

15.1 解：将这些元组进行标号，设为如下所示：

t_1: (kangaroo, 17) ; t_2: (wallaby, 21) ; t_3: (emu, 1)

t_4: (wombat, 13) ; t_5: (platypus, 3) ; t_6: (lion, 8)

t_7: (warthog, 4) ; t_8: (zebra, 11) ; t_9: (meerkat, 6)

t_{10}: (hyena, 9) ; t_{11}: (hornbill, 2) ; t_{12}: (baboon, 12)

进行 create runs，对每个内部进行排序：

（一个块）

(t_3, t_1, t_2)

(t_6, t_5, t_4)

(t_9, t_7, t_8)

(t_{12}, t_{11}, t_{10})

∵需要1个block来存放结果，所以每次最多合并2个 runs

第1个 pass：$(t_3, t_1, t_6, t_5, t_2, t_4)$ and $(t_{12}, t_{11}, t_{10}, t_9, t_7, t_8)$

产生 2个runs

第2个pass：$(t_{12}, t_3, t_{11}, t_{10}, t_1, t_6, t_9, t_5, t_2, t_7, t_4, t_8)$

得到最终的合并结果

15.3 Let relations $r_1(A, B, C)$ and $r_2(C, D, E)$ have the following properties: r_1 has 20,000 tuples, r_2 has 45,000 tuples, 25 tuples of r_1 fit on one block, and 30 tuples of r_2 fit on one block. Estimate the number of block transfers and seeks required using each of the following join strategies for $r_1 \bowtie r_2$:

a. Nested-loop join.

b. Block nested-loop join.

c. Merge join.

d. Hash join.

解：

r1 需要 800 个 block，r2 需要 1500 个 block

设内存大小为 M 个 block

（1）M>800

对于 a,b,c,d，均只需要 1500+800=2300 次 transfer 和 2 次 seek

（2）M≤800

a. Nested-loop join

若使用 r1 作为外循环的关系，则

transfer = 20000*1500+800 = 30000800

seek = 20000+800 = 20800

若使用 r2 作为外循环的关系，则

transfer = 45000*800+1500 = 36001500

seek = 45000+800 = 45800

b. Block nested-loop join

若使用 r1 作为外循环的关系，则

transfer = $\lceil \frac{800}{M-2} \rceil * 1500 + 800$

seek = $2\lceil \frac{800}{M-2} \rceil$

若使用 r2 作为外循环的关系，则

transfer = $\lceil\frac{1500}{M-2}\rceil * 800 + 1500$

seek = $2\lceil\frac{1500}{M-2}\rceil$

c. Merge join

若 r1,r2 有序，则

transfer = 800+1500 = 2300

设每个关系分配 br 块 block，则

seek = $\left\lceil\frac{800}{br}\right\rceil + \lceil\frac{1500}{br}\rceil$

若 r1,r2 无序，设 bn 为缓冲区大小，则将其排好序时，

transfer1 = $800 * (2\lceil\log_{\lfloor\frac{800}{bn}\rfloor}\frac{800}{M}\rceil + 1) + 1500 * (2\lceil\log_{\lfloor\frac{1500}{bn}\rfloor}\frac{1500}{M}\rceil + 1)$

假设每次从归并段读取 bn 块数据，则

$$seek1 = 2 * \lceil\frac{800}{M}\rceil + \lceil\frac{800}{bn}\rceil\lceil\log_{\lfloor\frac{M}{bn}\rfloor-1}\frac{800}{M}\rceil - 1) + 2 * \lceil\frac{1500}{M}\rceil + \lceil\frac{1500}{bn}\rceil\lceil\log_{\lfloor\frac{M}{bn}\rfloor-1}\frac{1500}{M}\rceil - 1)$$

所以 transfer = 800+1500+transfer1 = 2300+transfer1

seek = $\left\lceil\frac{800}{bn}\right\rceil + \left\lceil\frac{1500}{bn}\right\rceil + seek1$

d. Hash join

若不需要递归划分：

transfer = 3*(800+1500) = 6900

假设输入输出缓冲区大小为 bn：

seek = $2(\left\lceil\frac{800}{bn}\right\rceil + \left\lceil\frac{1500}{bn}\right\rceil)$

若需要递归划分：

transfer = $2(800 + 1500)\left\lceil\log_{\lfloor\frac{M}{bn}\rfloor-1}\frac{800}{M}\right\rceil + 800 + 1500$

$$= 4600 \left\lceil \log_{\lfloor \frac{M}{bn} \rfloor - 1} \frac{800}{M} \right\rceil + 2300$$

$$\text{seek} = \ 2(\lceil \frac{800}{bn} \rceil + \lceil \frac{1500}{bn} \rceil) \lceil \log_{\lfloor \frac{M}{bn} \rfloor - 1} \frac{800}{M} \rceil$$