

1. ① 创建归并段

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

② 第一趟归并

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

③ 第二趟归并

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

最佳排序方法是使用哈夫曼排序算法：将其中一个关系r加载到内存中，构建n个然后遍历另一个关系s。

需要  $\min\{s \text{大小}, r \text{大小}\}$

3. 关系大小  $r = \frac{40GB}{4KB} = 10^7$  块  
a. 内存大小  $m = \frac{40MB}{4KB} = 10^4$  块  
 $D_{bb} =$

a. 每块大小  $b = \frac{4KB}{4KB} = 1$  块  
扫描成本  $S = 5ms$

成本  $= 0.005 + \frac{40GB}{40MB/s} + \frac{40GB}{40MB/s}$

$b_b = 100$

成本  $= 0.005 + \frac{40GB}{40MB/s} + \frac{40GB}{40MB/s \times 100}$

b. 归并次数  $= T \log_{b_b} \left( \frac{40GB}{40MB} \right)$



$$b_b = 1$$

$$\text{成本} = 0.00002 + \frac{40\text{GB}}{400\text{MB/s}} + \frac{40\text{GB}}{400\text{MB/s}}$$

$$b_b = 100$$

$$\text{成本} = 0.00002 + \frac{40\text{GB}}{400\text{MB/s}} + \frac{40\text{GB}}{400 \times 100\text{MB/s}}$$

设有关系  $S$  和  $R$ .

先将  $R$  和  $S$  按其中一个属性排序.

将  $R, S$  分成相等大小的块, 每个块适应内存大小.

将  $R$  中以依据排序的属性 和  $S$  中依据排序的属性读入,

再利用两个属性排序合并, 得到属性排序后的元组.

通过对比属性, 找到连接属性相等的元组输出到结果.

利用将更多的数据存入到内存里处理, 从而减少 I/O 操作的次数, 从而减少并操作的运行成本.