

# 1

---

工作内存最多容纳的记录数为  $1024 * 1024 / 32 = 2^{15}$ .

在置换选择时, 如果工作内存全部用于存储记录, 平均形成长度为  $2 * 2^{15} = 2^{16}$  的顺串.

在进行多路归并时, 如果使用胜者树或败者树, 在进行  $k$  路归并时需要额外存储  $k - 1$  个内部结点, 结点中存放记录索引. 假设每个索引占用 4 字节, 则最多进行约  $2^{15} * \frac{32}{32+4} = \frac{1}{9} 2^{18}$  路归并.

在以上条件下, 第一轮归并后的顺串长度为  $2^{16} * \frac{1}{9} 2^{18} = \frac{1}{9} 2^{34}$ , 第二轮归并后的顺串长度为  $\frac{1}{9} 2^{34} * \frac{1}{9} 2^{18} = \frac{1}{81} 2^{52}$ .

综上, 所得到文件包含记录数最多为  $\frac{1}{81} 2^{52} \approx 1.6 * 2^{45}$ , 即长度为  $1.6 * 2^{50}$  个字节.

# 2

---

## (1)

我们知道, 用大小为  $M$  的堆进行置换选择排序时, 产生的顺串最小为  $M$ , 平均为  $2M$ , 最大为  $N$  (记录总数).

相应地, 产生的顺串个数最少为 1, 平均为  $\frac{N}{2M}$ , 最多为  $\frac{N}{M}$ .

代入题目数据, 即知产生的顺串个数最少为 1, 平均为 100, 最多为 200.

## (2)

手动模拟置换选择排序的过程, 可以知道共产生了三个顺串, 分别为:

串 1: 3, 12, 18, 22, 25, 48, 58, 61, 65, 72, 79, 90, 100;

串 2: 11, 14, 16, 18, 20, 22, 26, 27, 30, 38, 45, 50, 55, 107;

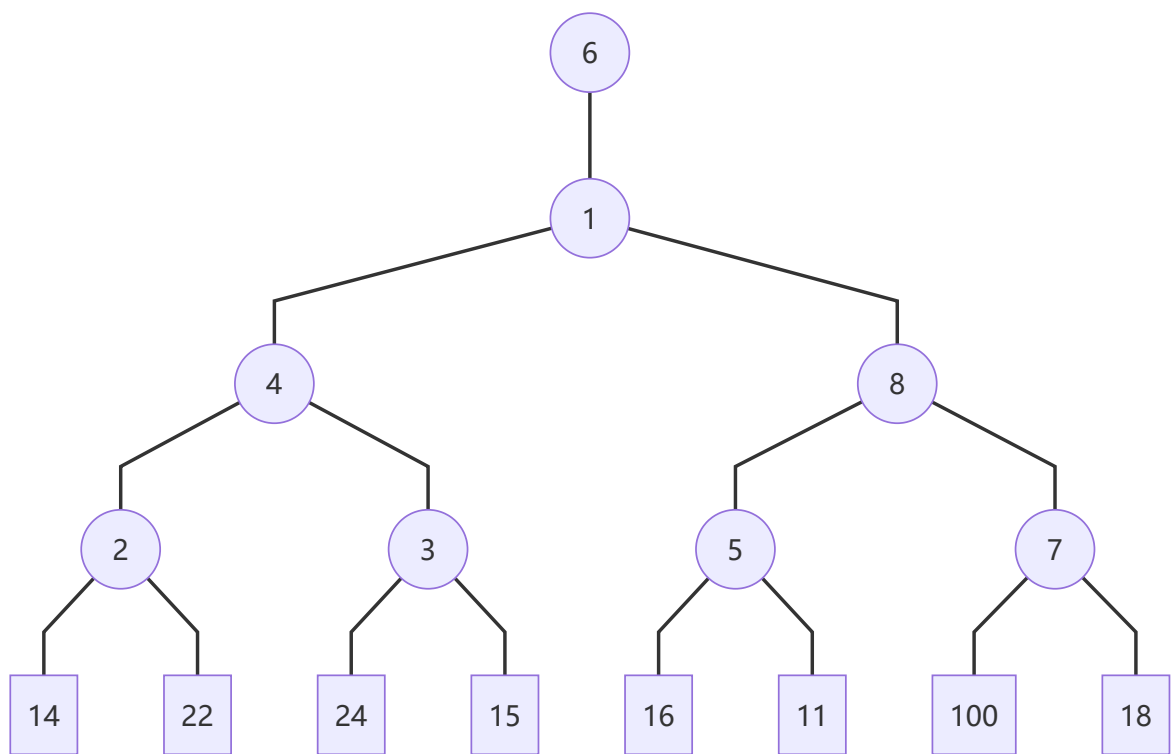
串 3: 17.

# 3

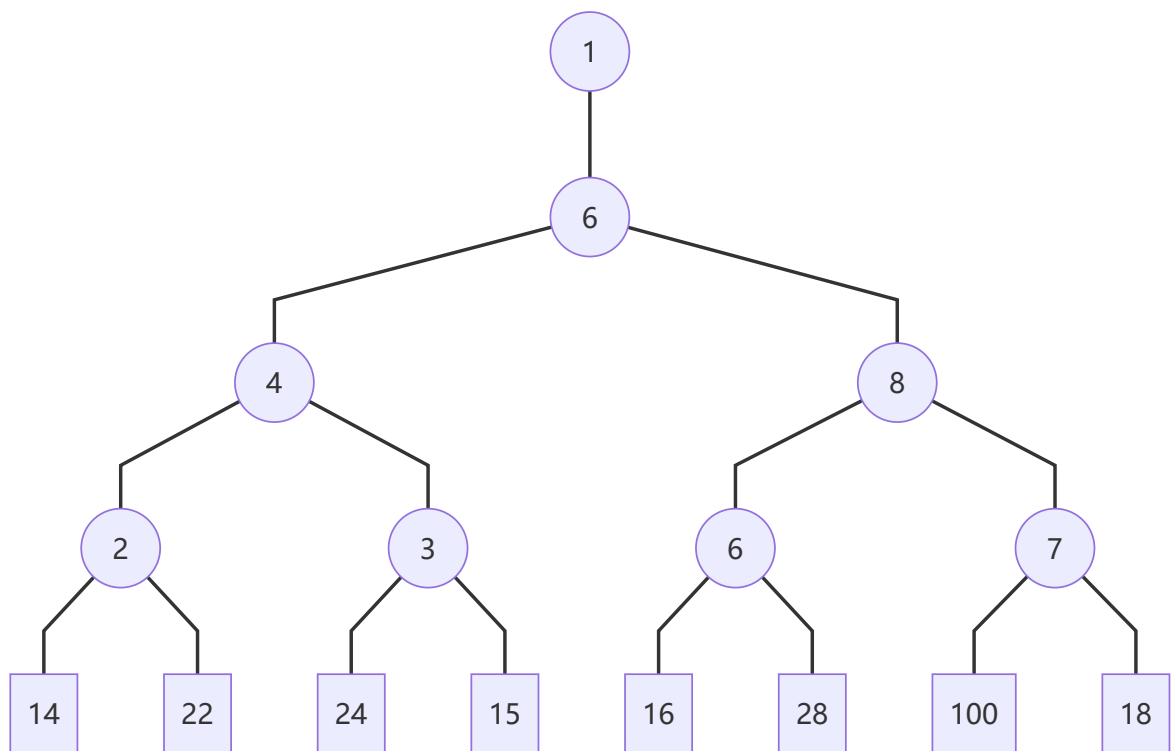
---

由于题目未加指明, 我按照从小到大排序完成的这道题目.

初始败者树:



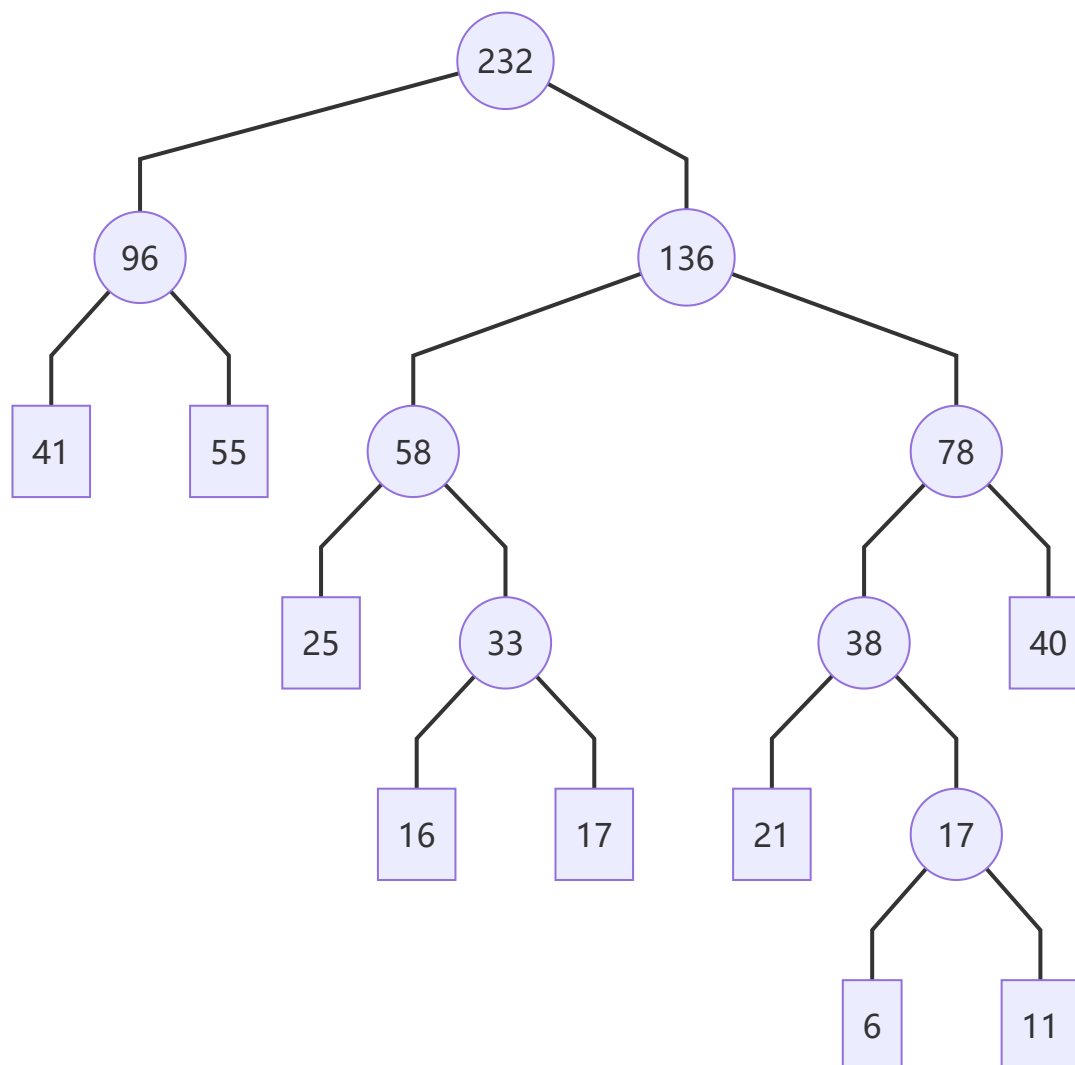
第一次重构后的败者树：



图中方形结点为记录关键字，圆形结点为败者索引（根结点为冠军索引）。

**(1)**

以顺串长度为权值构造的哈夫曼树即为最佳归并树：



其中方形结点表示原始顺串，圆形结点为过程中产生的顺串或结果顺串。

**(2)**

假设每轮合并需要读所涉及的记录各一次。

第一趟： $6 + 11 + 16 + 17 + 41 + 55 = 146$ 。

第二趟： $17 + 21 + 22 + 25 = 96$ 。

第三趟： $38 + 40 = 78$ 。

第四趟： $78 + 58 = 136$ 。

第五趟： $136 + 96 = 232$ 。

总计： $146 + 96 + 78 + 136 + 232 = 688$ 。