先用Hash做预处理，把大文件转换为小文件。首先采用url%100的方式（(i +1)\*(int) url[i]\* 10007 %100 这样可能分的均匀一点），把Url分成100小份，如果生成的文件大小仍然大于内存大小，则使用另一种哈希方法继续划分数据直到每个文件的大小都小于1G，

对每个文件进行排序（根据url数目的大小由大到小），取每个文件的前100个ur（每个url最大不超过2500B 哈希3次 300个文件 300\*2200B\* 100 使用的内存为71.5MB<1G)