外排序算法（External Sorting）：指在内存不够的情况下，对存储在一个或多个大文件中的数据进行排序的算法。在海量数据处理算法中十分重要。

外排序的两个基本步骤：

1. 将大文件分为若干个小文件，并分别使用内存拍好序：根据内存的大小，尽可能多的分批次的将数据 Load 到内存中，并使用系统自带的内存排序函数（或者自己写个快速排序算法），将其排好序，并输出到一个个小文件中。比如一个文件有1T，内存有1G，那么我们就这个大文件中的内容按照1G的大小，分批次的导入内存，排序之后输出得到 1024个1G 的小文件。
2. 使用K路归并算法将若干个排好序的小文件合并到一个大文件中：将 K 个文件中的第一个元素加入到堆里，假设数据是从小到大排序的话，那么这个堆是一个最小堆（Min Heap），如果这个元素来自第x个文件，则从第x个文件中继续读入一个新的数进来放到堆里，并重复上述操作，直到所有元素都被输出到目标结果文件中。如果每次每个文件只读入1个元素，那么一次就只能排序1024个元素然后输出到结果文件中，显然这样效率太低了。为了避免频繁读取和写入文件操作，可以为每个文件设置1个缓冲Buffer，每次需要从文件中读取数据时，就把这个buffer读满，当Buffer中的数据用完时，再从文件中读取，同理输出时也设置一个Buffer来写入文件。