

# 大数据排序问题

---

日常生活中遇到的排序问题往往和课本上的排序问题不同。比如说，需要排序的数据可能不是整数，可能数据量巨大，可能不是以数组方式存储的。

题目要求：

1. 用g++ (GNU C++ 编译器)编译 generate.cpp 程序。如果使用Linux环境，可以使用如下命令编译： `g++ generate.cpp -o generate`
2. 运行编译出的执行文件generate 在Linux环境中： `./generate`
3. 观察生成的源数据文件source\_data.dat 在该文件中，每一行是一个记录<key, value>，其中key是记录的标号，value是记录的内容。key和value之间用空格分离，key是一个整数，value是一个字符串。整个文件一共有10000000行，即包含10000000个记录
4. 编写一个C/C++程序，读入source\_data.dat文件，生成一个sorted\_data.dat文件。在sorted\_data.dat文件中，所有的记录按key从小到大排列。注意：（a）每一个key所对应的value必须仍然在同一行。（b）该C/C++程序只能使用最基本的C/C++库，排序部分的算法必须自己实现。

例子：对于如下的source\_data.dat文件

```
234 78M71639 14 6M88173 1847 736M02472 738 6284M17
```

排序后生成的sorted\_data.dat文件应该为

```
14 6M88173 234 78M71639 738 6284M17 1847 736M02472
```