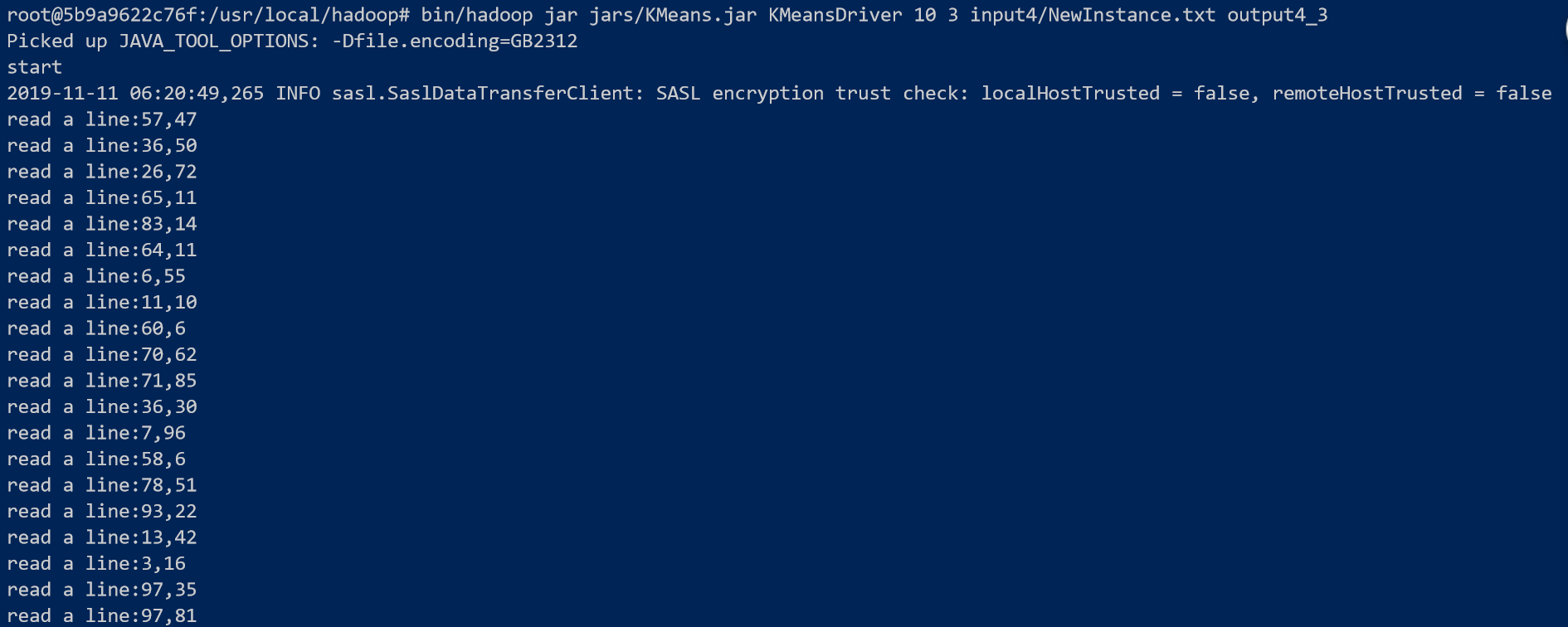
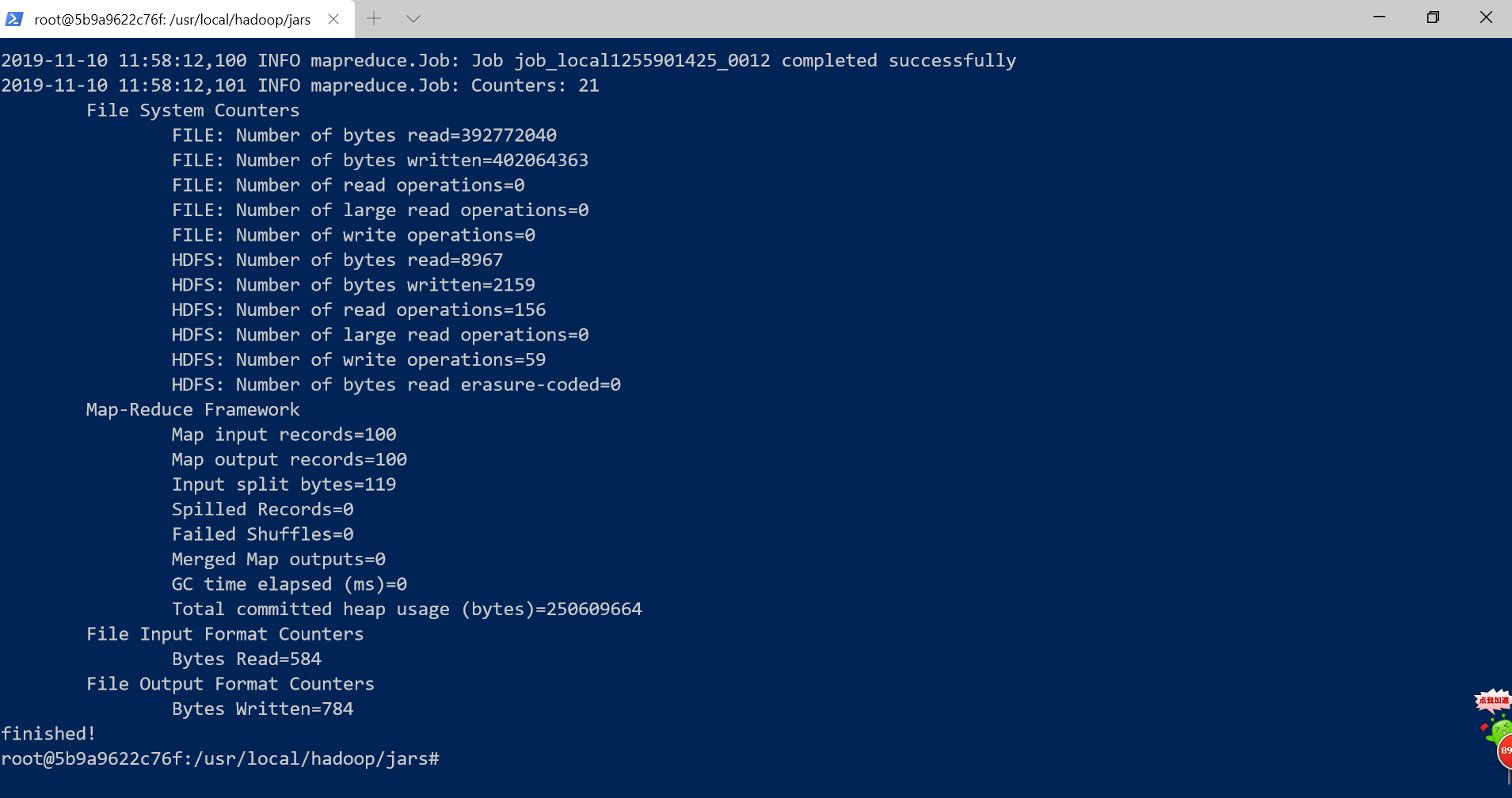
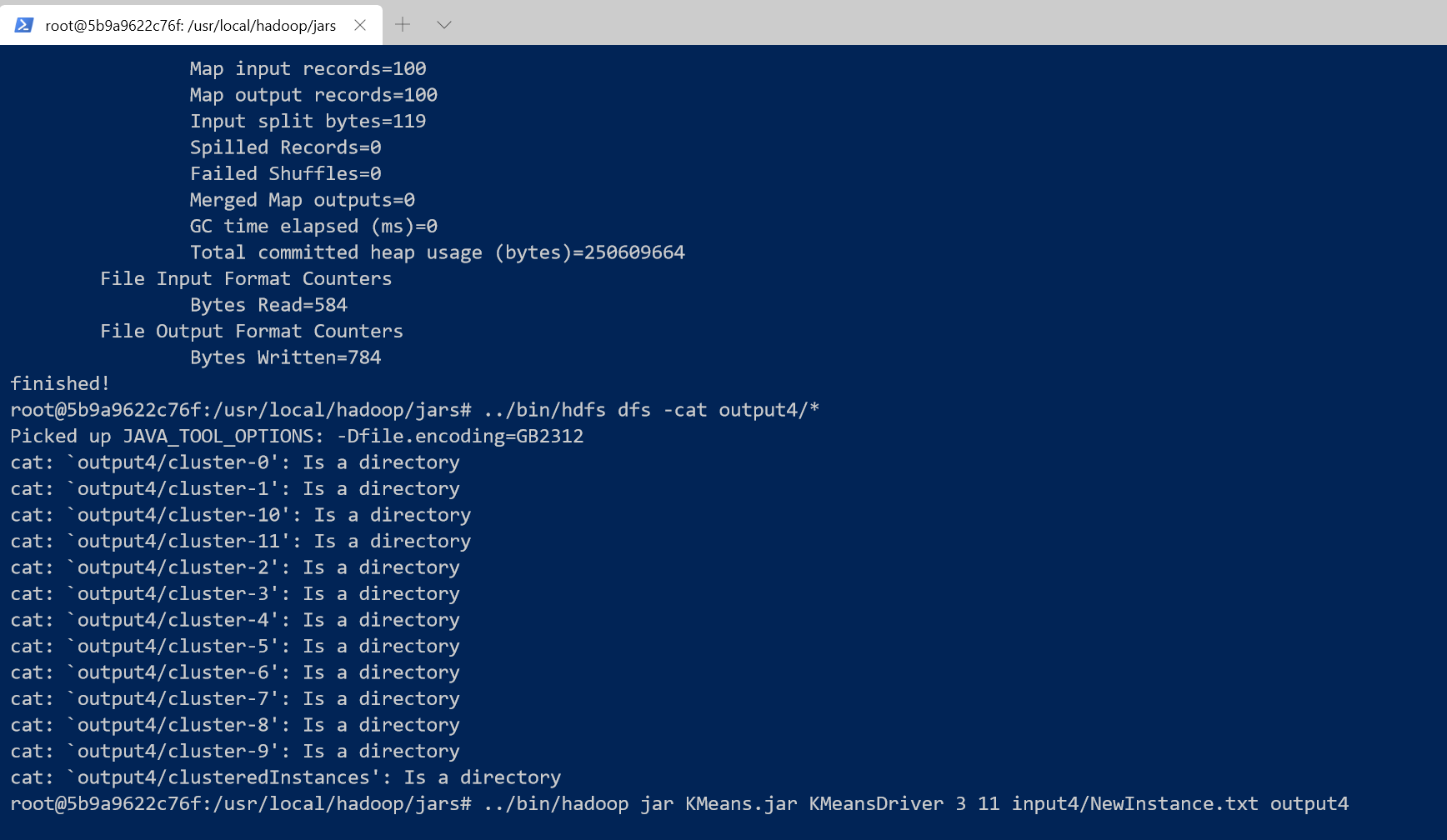
1. 运行（伪分布式）

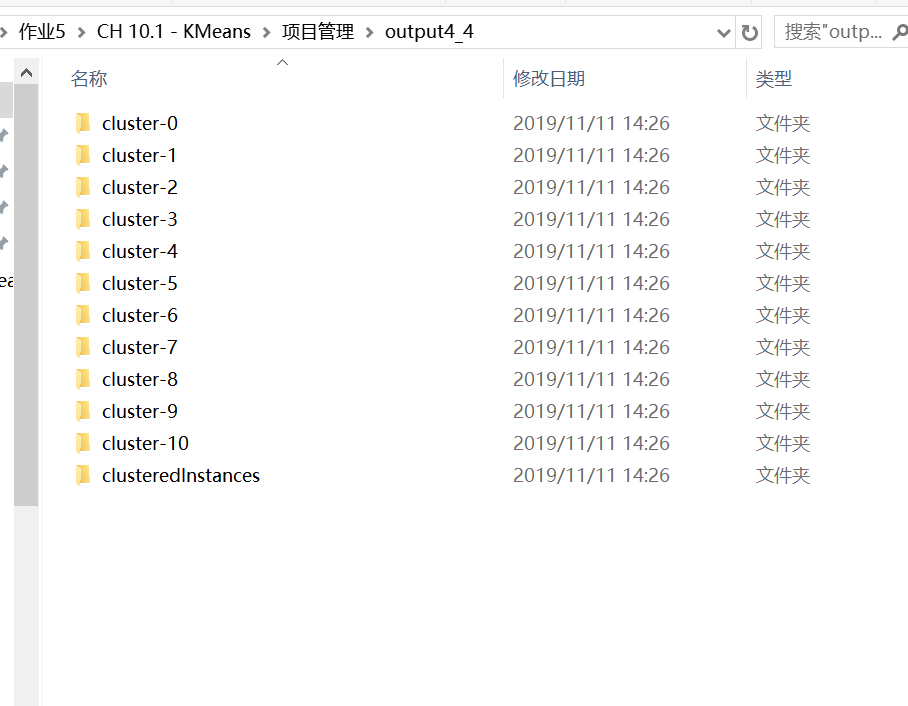


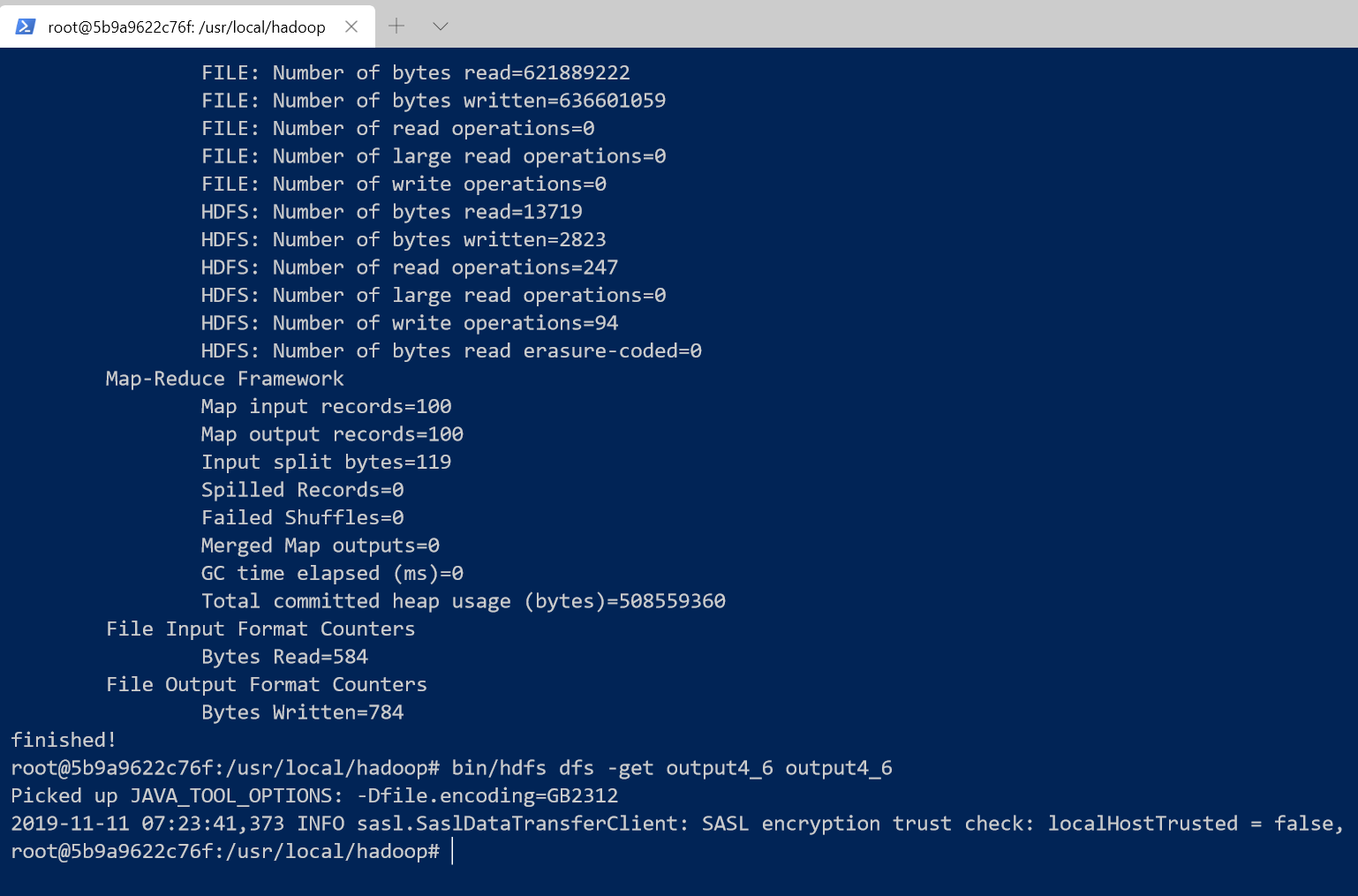


2、看一看输出的文件（hdfs -cat output4/\*）



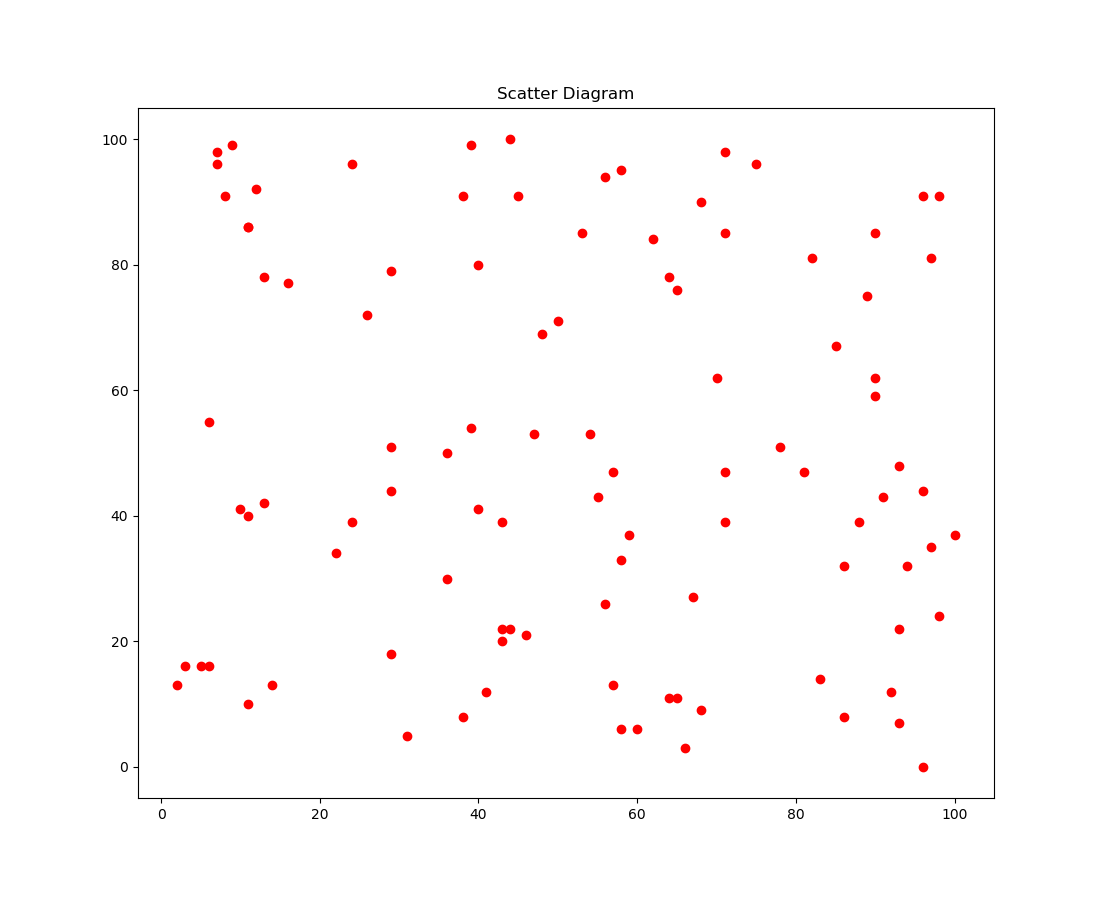
3、把输出文件复制到win10系统上看一下





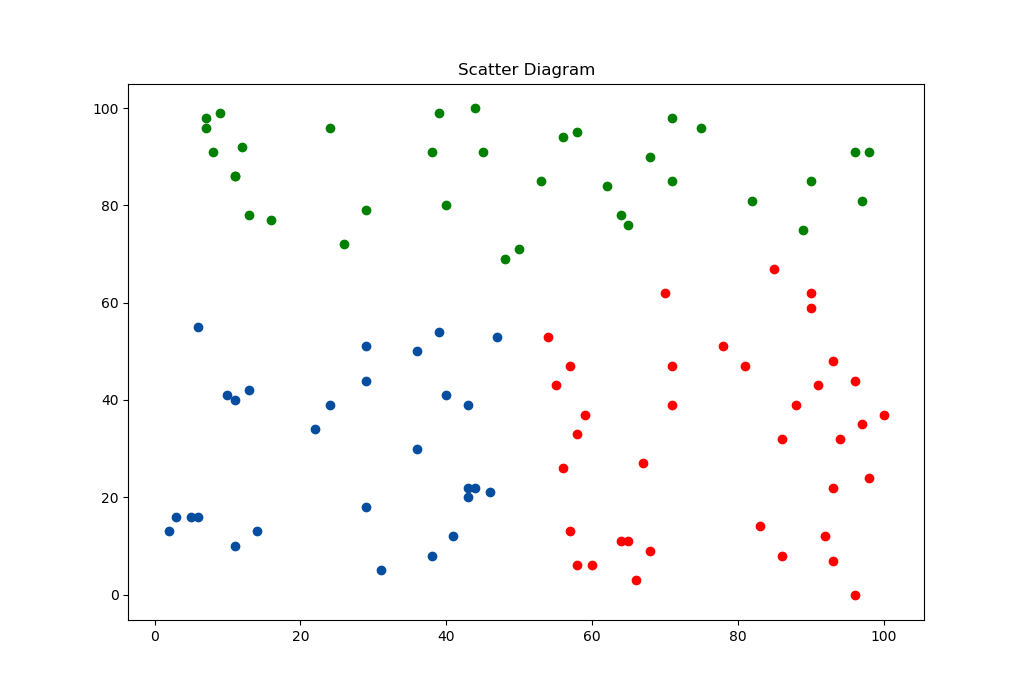
4、KMeans聚类结果可视化（通过python）

没有做聚类的时候散点图是这样的：

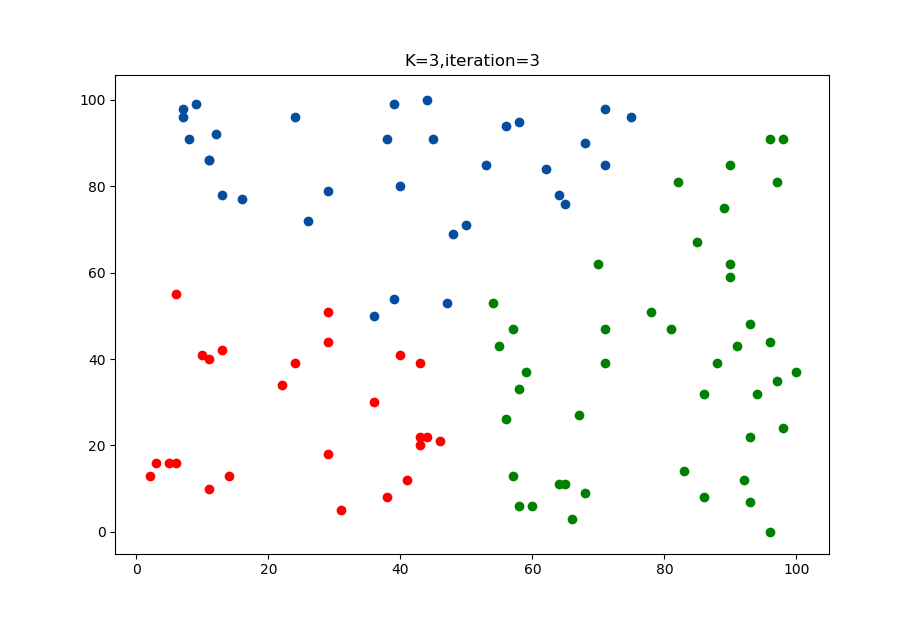


设置K=3，迭代次数为10的聚类结果是这样的：

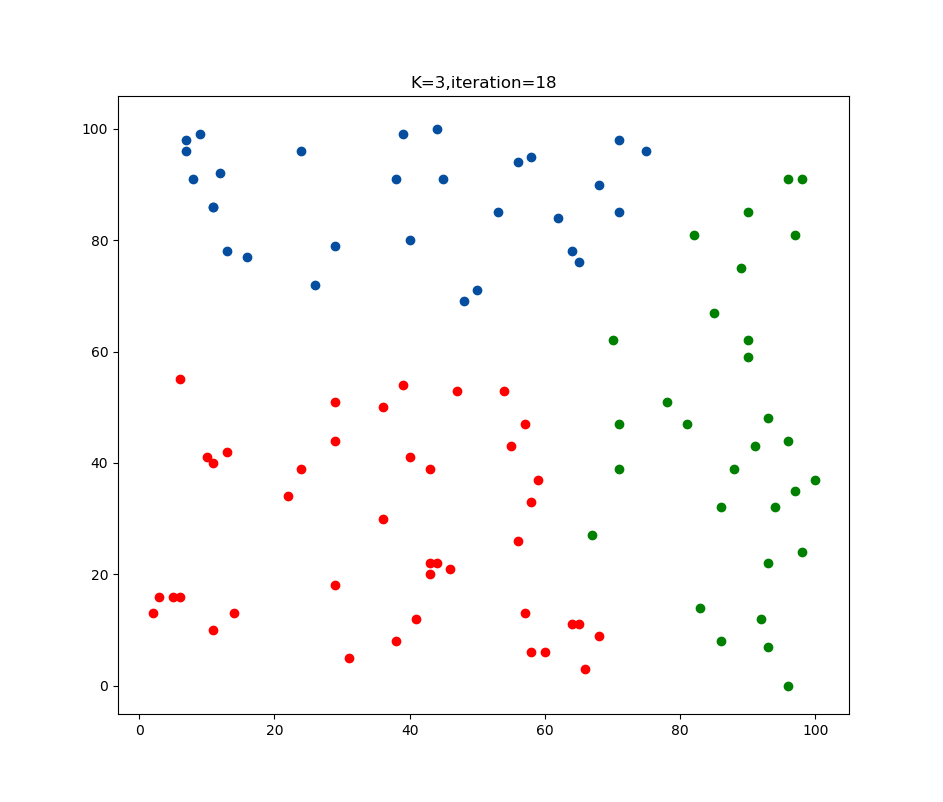
其中用不同的颜色代表不同的簇



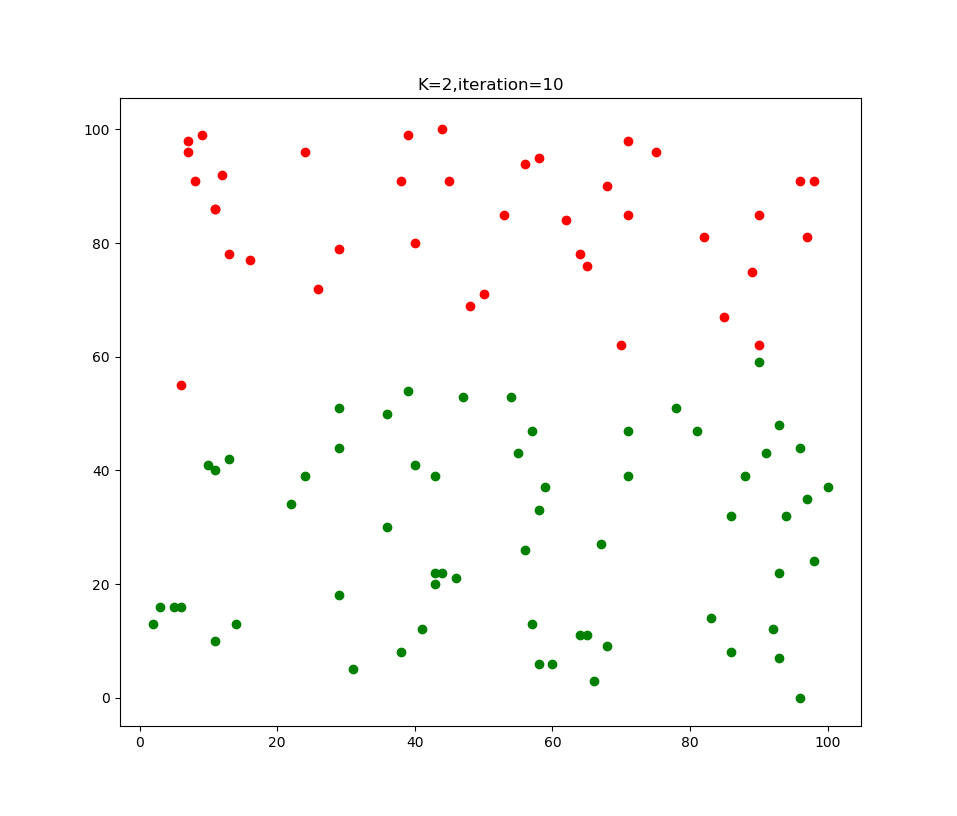
改变迭代的次数，保持K=3，设置迭代次数为3次，结果如下：



再次改变迭代的次数，保持K=3，设置迭代次数为18次：



改变簇中心数，设置K=2(迭代次数为10次)：



再次改变簇中心数，设置K=5（迭代次数为10次）：

