MapReduce框架知识点梳理		
组件名	组件知识点	补充
Mapper	可以读取文件，默认是一行一行读取，把输入key和输入value通过map()传给程序员。输出key和输出value由业务来决定。MR框架会按Mapper的输出key做排序，输出key要实现WritableComparable接口	MapTask的数量=切片数量 切片是一个对象，封装文件块的描述信息，path，start,length
Reducer	接收Mapper组件的输出k,v然后按相同key做聚合	ReduceTask任务数量通过代码来指定
Partitioner	分区组件。分区概念等同于ReduceTask。即有几个ReduceTask就有几个分区。 默认的分区器是HashPartitioner，作用是按Mapper输出key的hash分区，可以确保相同key落到同一分区里。 可以自定义分区，写一个类继承Partitioner，最后在Driver里通过job.setPartitioner()	类名：HashPartitioner，底层用的是简单hash算法，这种分区算法可能会产生数据倾斜现象。
Combiner	合并组件，作用是让合并工作在MapTask端先合并，再发给reduce端。 开发方式，写一个类继承Reducer。然后在Driver里，通过job.setCombinerClass( )	