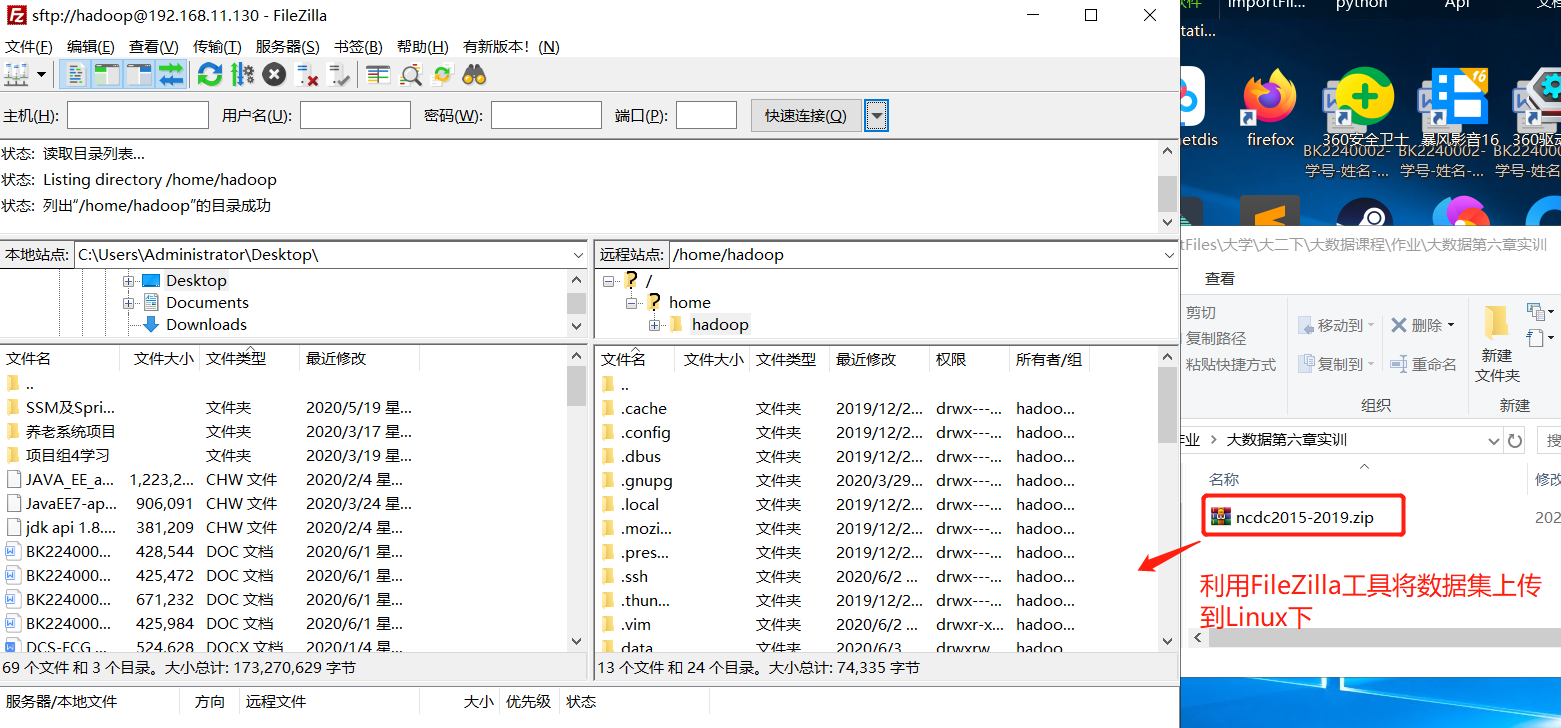
# 统计全球每年的最高气温和最低气温

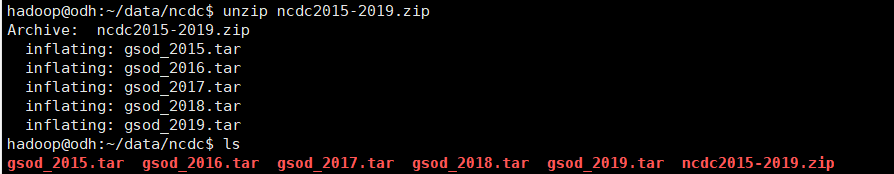
1. Linux下创建data/ncdc目录

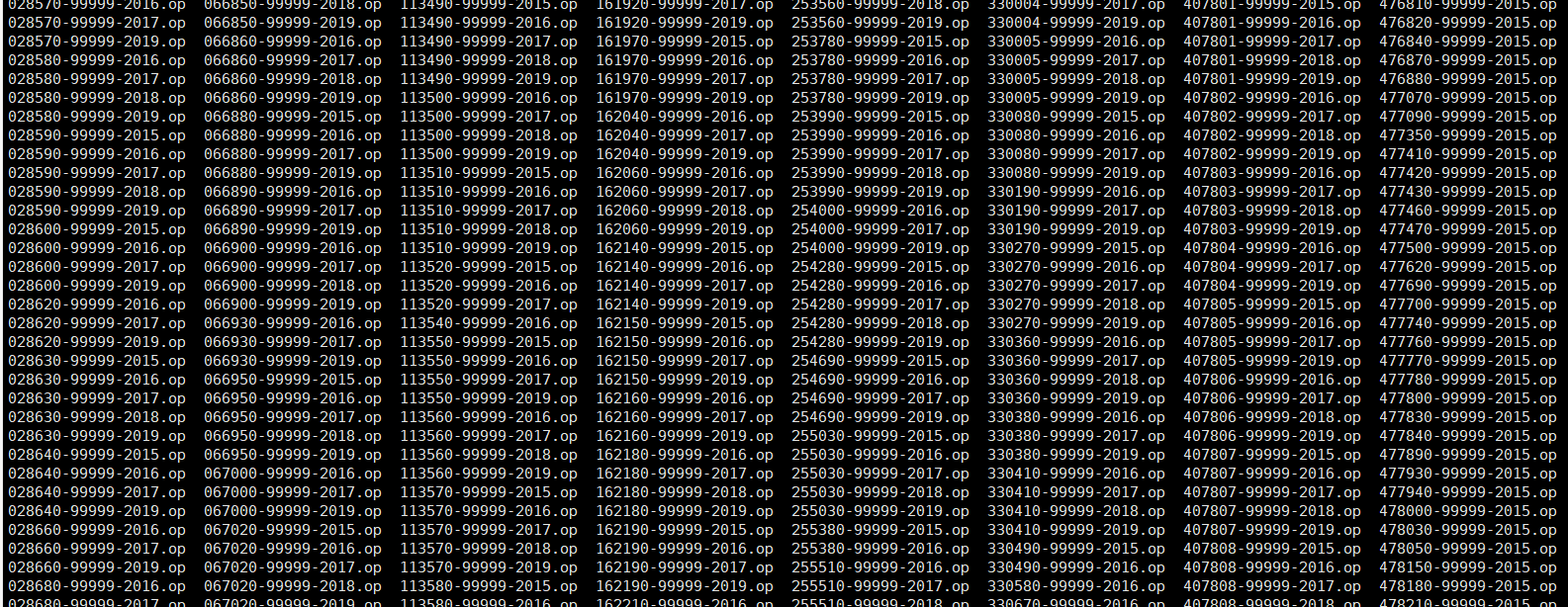


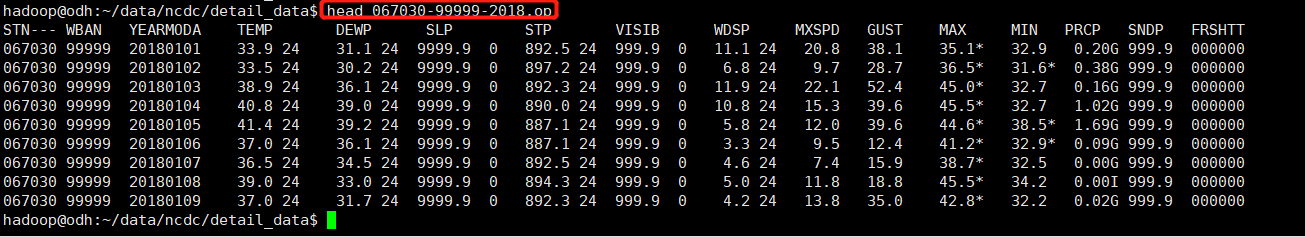
1. 上传数据集到Linux下



1. Unzip命令解压数据集，然后再用tar -xvf命令解压tar文件得到op.gz后缀的文件，最后用gunzip命令解压gz文件得到数万个.op文件







1. sed -i ‘1d’ \* ”删除所有文件的首行字段，然后用cat \* >> \*\*\*将所有数据汇总。

（最终汇总的数据共 628万余行）



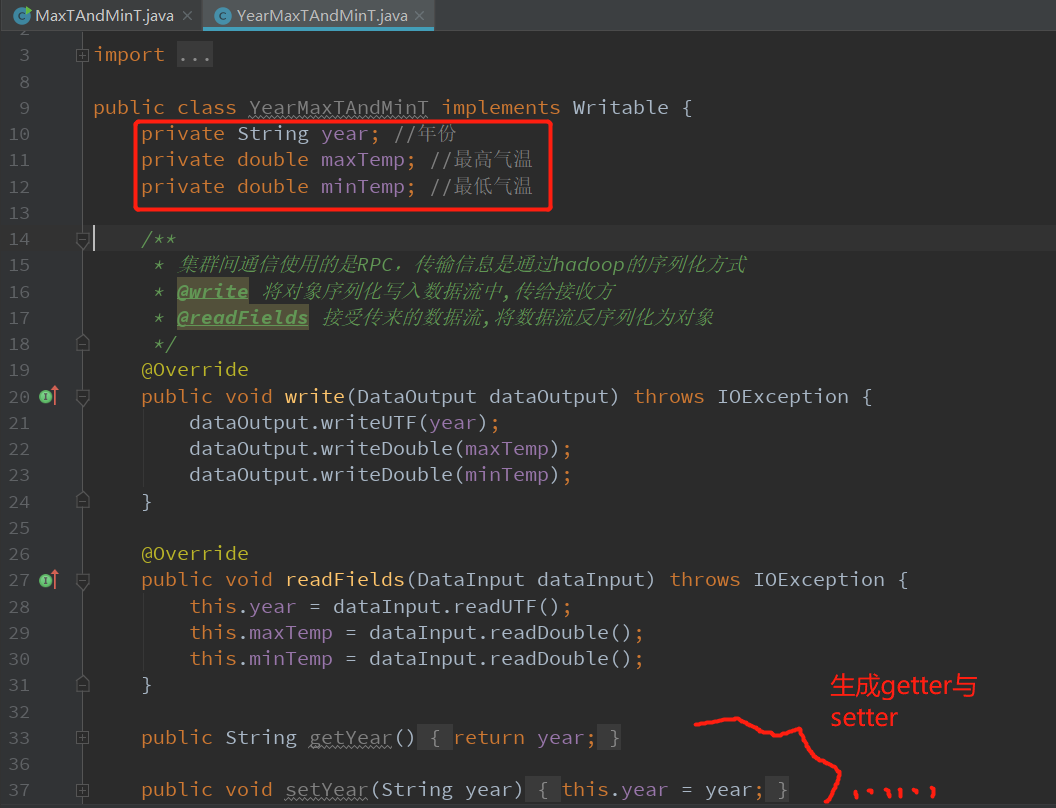


1. 上传数据到HDFS

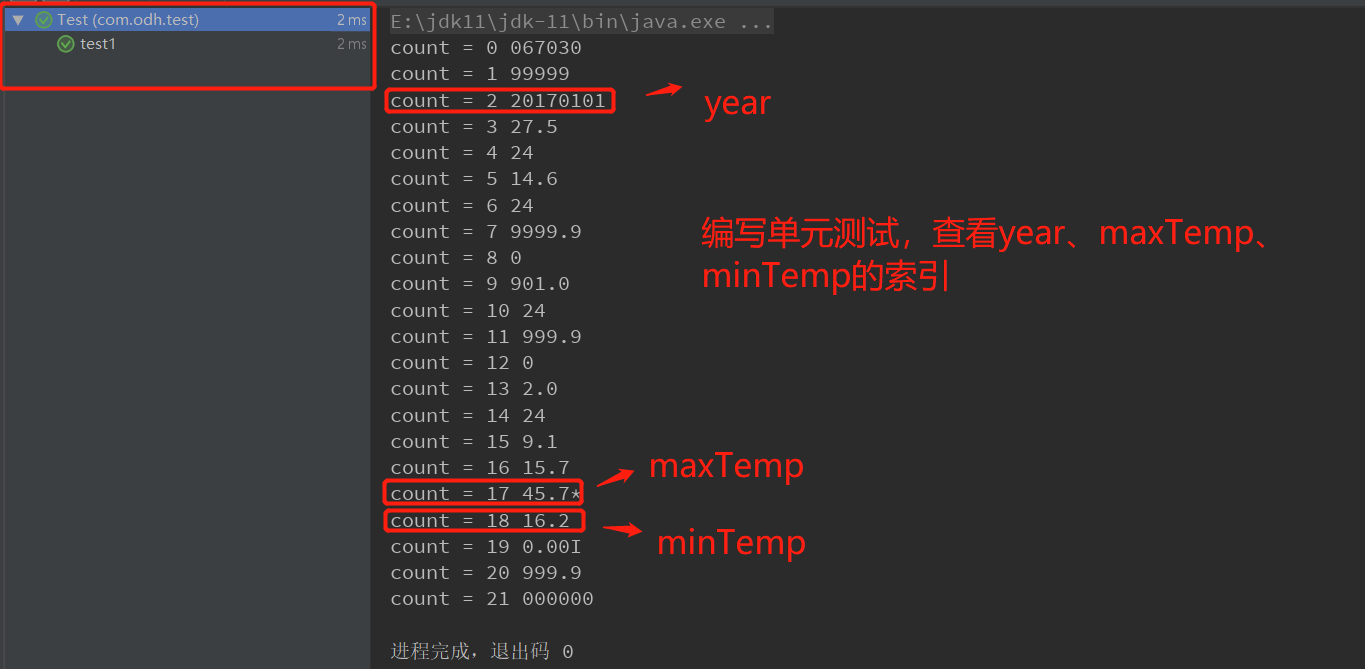


1. 创建entity包，在包下新建并编写YearMaxTAndMinT实体类



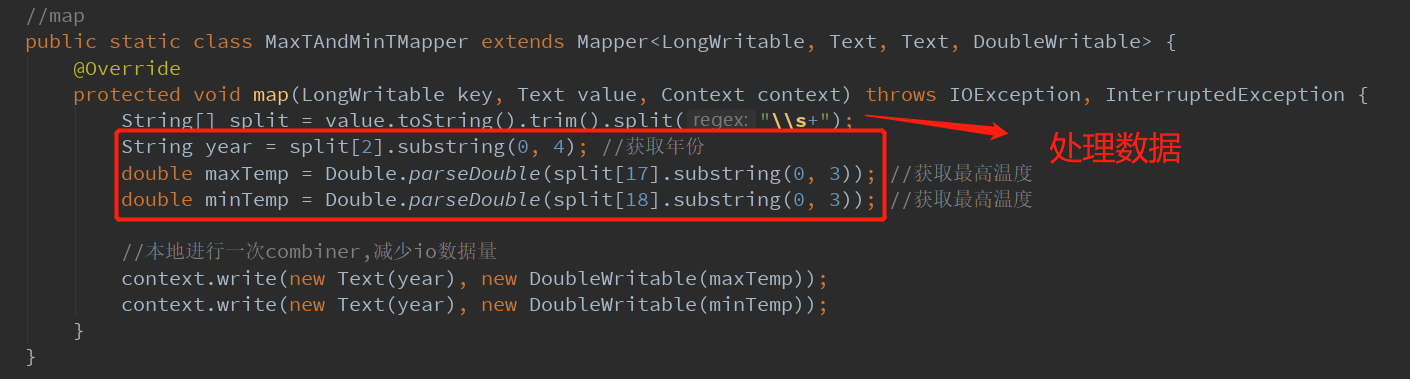


1. 分析数据



1. Map过程图以及编写Mapper类

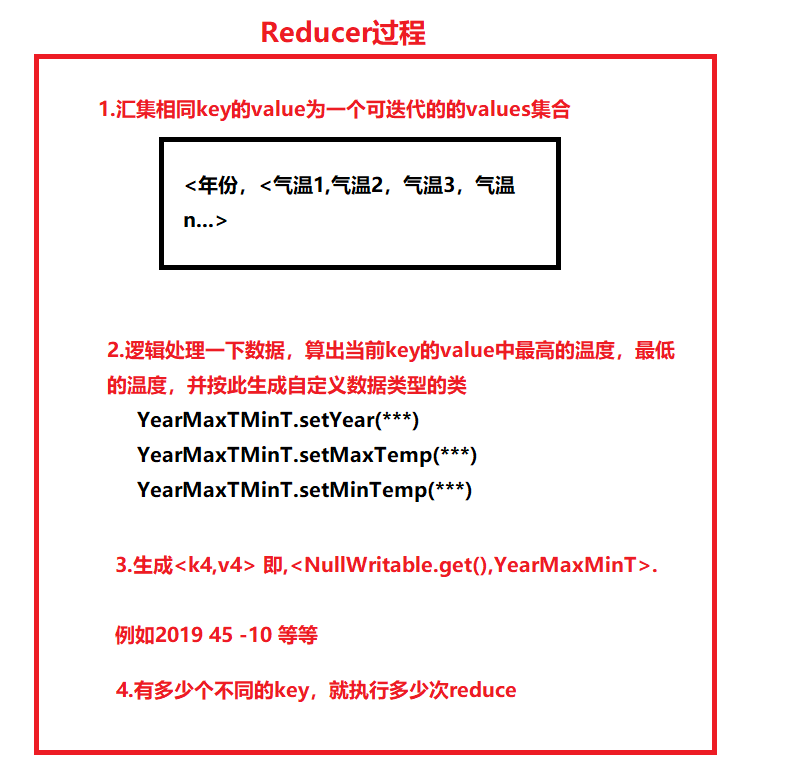


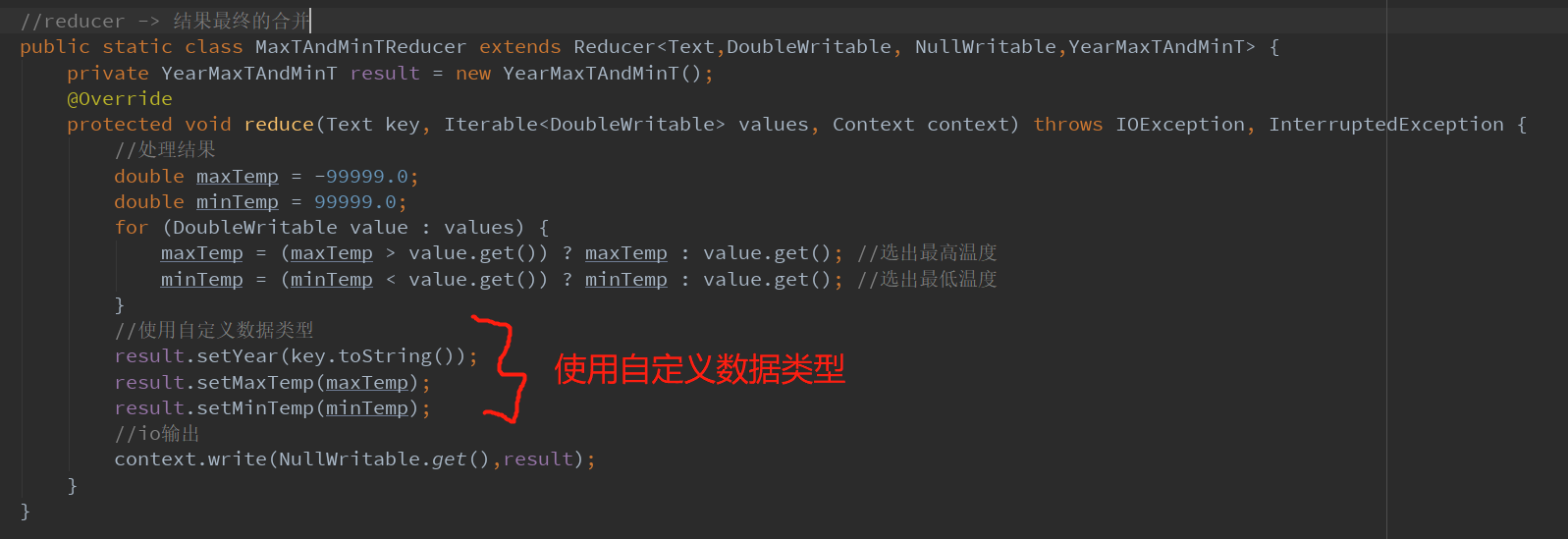


1. 编写Combiner类

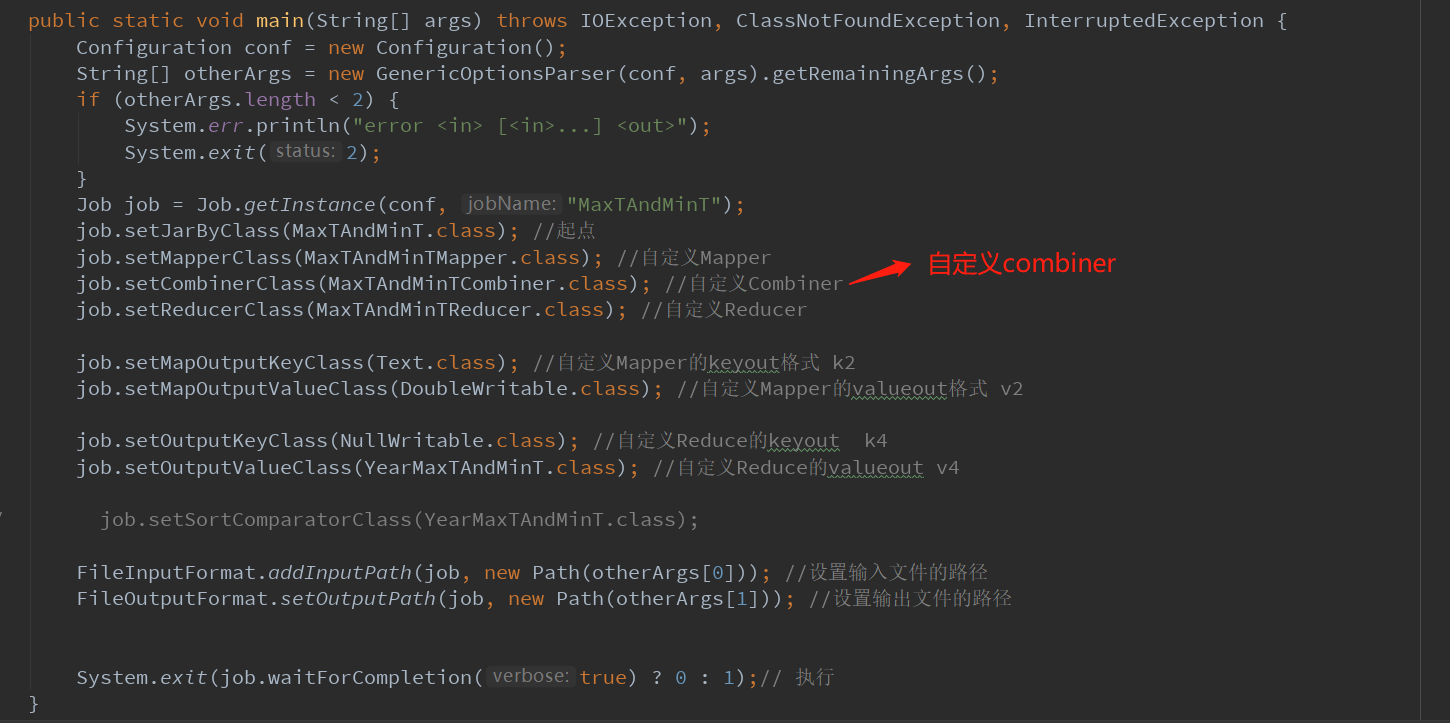


1. 编写Reducer类以及reducer过程流程图

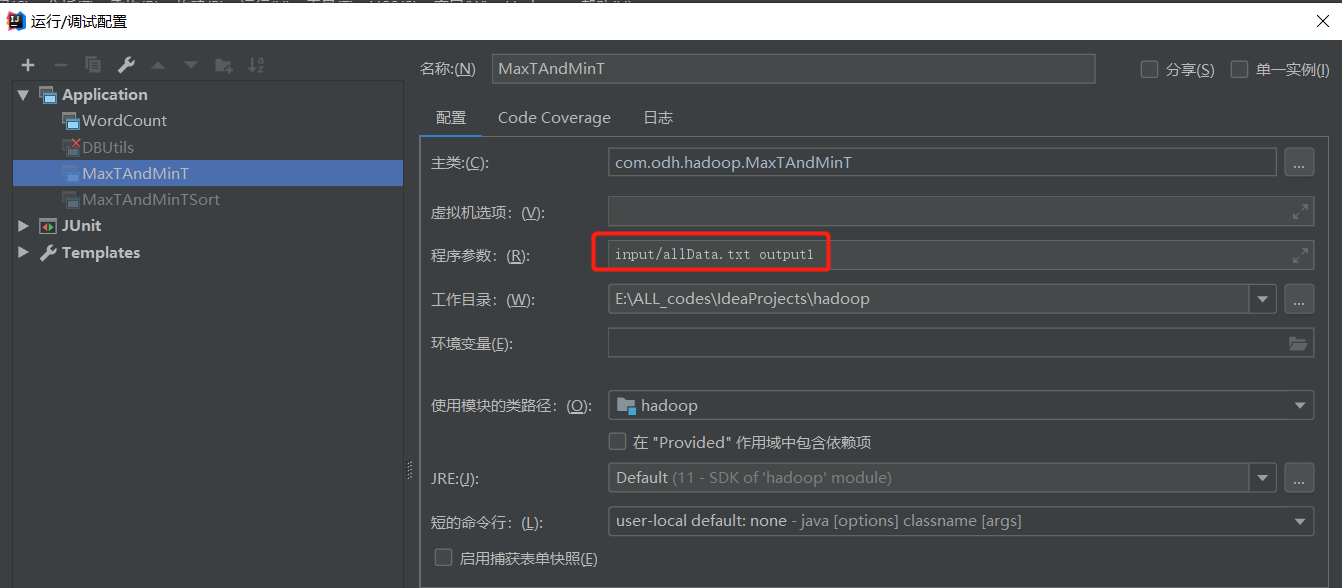




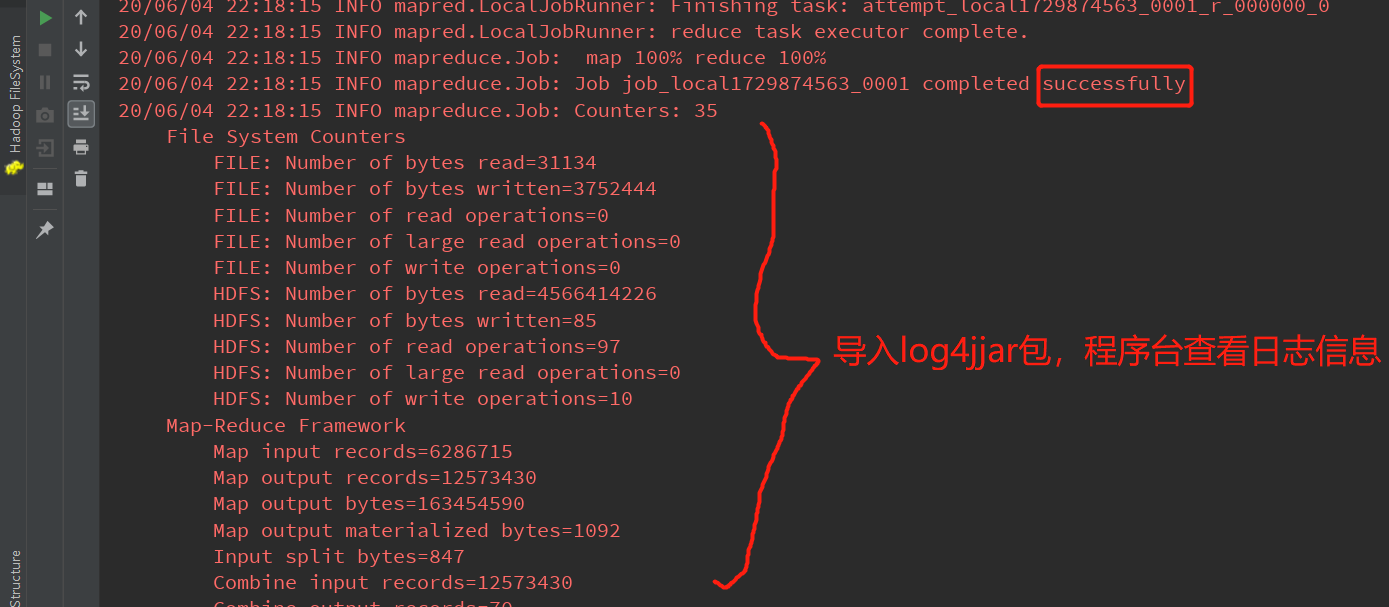
1. 构建Driver类，运行程序并查看结果

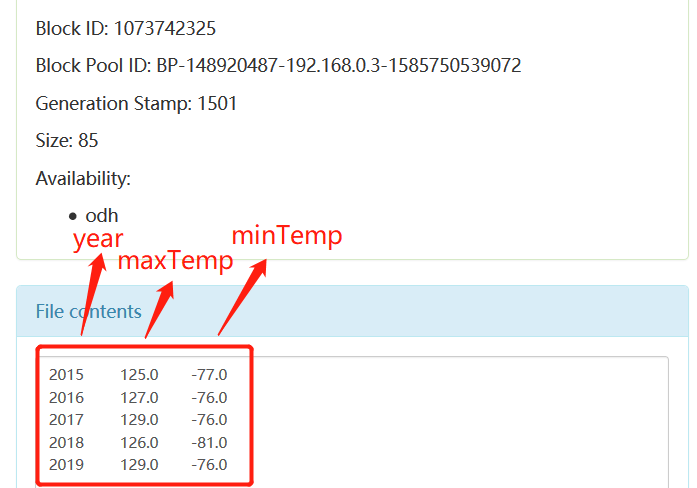


**设置args参数**



**控制台查看运行结果**



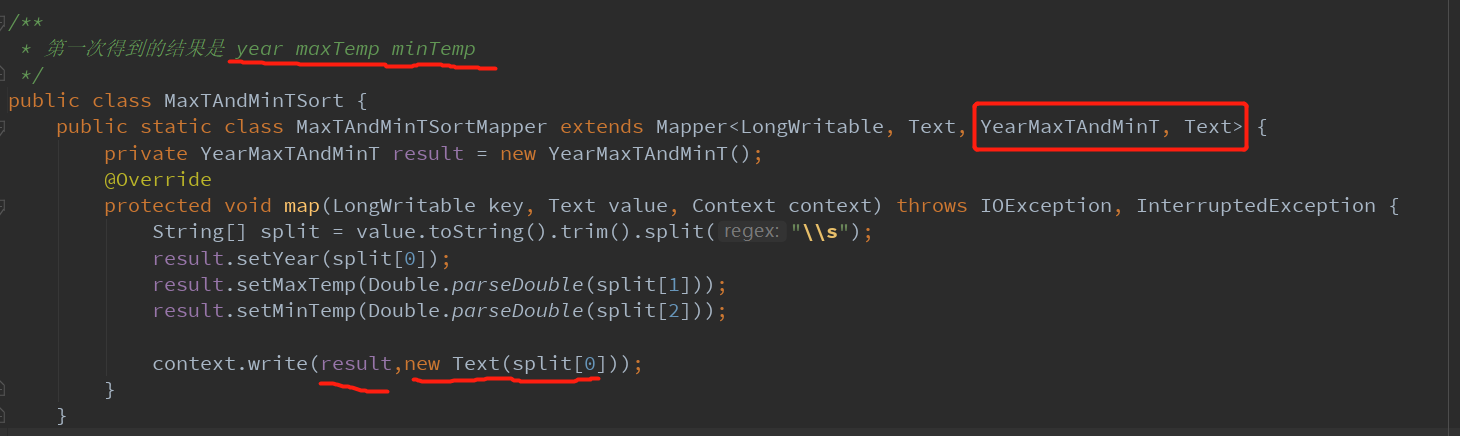


# 全球每年最高气温和最低气温的排序

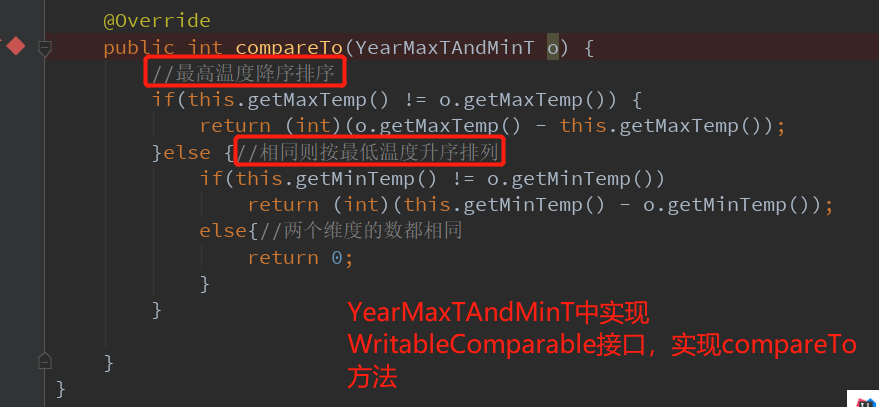
1. 以上就完成了对2015-2019年度最高气温与最低气温的统计，还要编写一个Sort类对maxTemp，minTemp维度进行二次排序，编写MaxTAndMinTSort类

(Hadoop会自动在map以后对数据的key进行排序，然后再分区，进入Reduce）

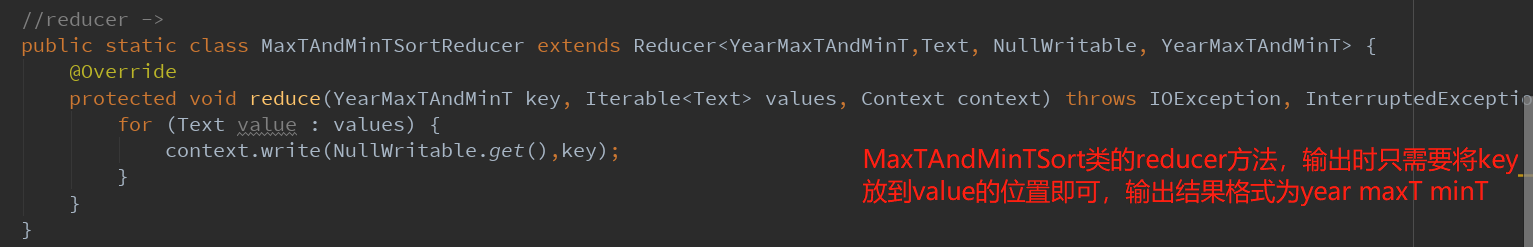
(所以这里的map输出的key,value是 <自定义键值对，Text> )



1. 利用Hadoop自动为key排序的机制，如果key实现了WritableComparable方法，可以重写WritableComparable中的compareTo方法来实现对key的排序



1. 编写MaxTAndMinTSort的reduce



1. 执行MaxTAndMinTSort类并输出，查看结果



