第一题（个人作业）：

在新概念英语第二册上完成 word count

## 方法

实例代码的map函数中，根据英文单词之间的空格区分单词并记为key-value对。但是一些标点符号也作为单词列入了wordcount。所以需要先把一段文本的字符串中的标点符号过滤掉，替换成空格。

public static String rm\_punc(String str\_ori)

  {

    String str\_new=new String();

    int str\_len=str\_ori.length();

    for(int i=0;i<str\_len;i++)

    {

      char tmp\_char=str\_ori.charAt(i);

      if(tmp\_char>='a' && tmp\_char<='z')

      {

        str\_new+=tmp\_char;

      }

      else if(tmp\_char>='A'&&tmp\_char<='Z')

      {

        str\_new+=tmp\_char;

      }

      else

      {

        str\_new+=' ';

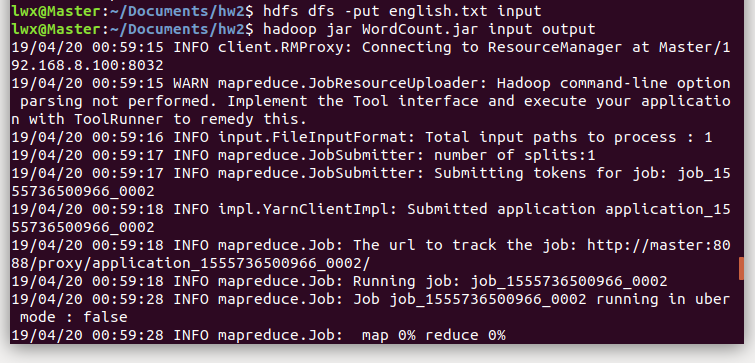
      }

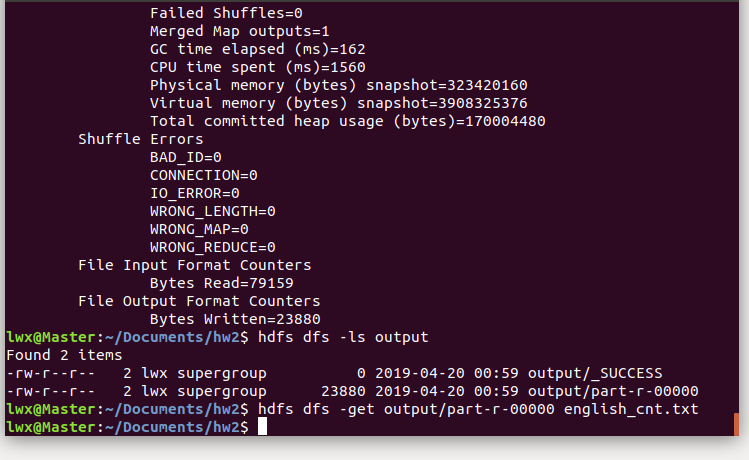
    }

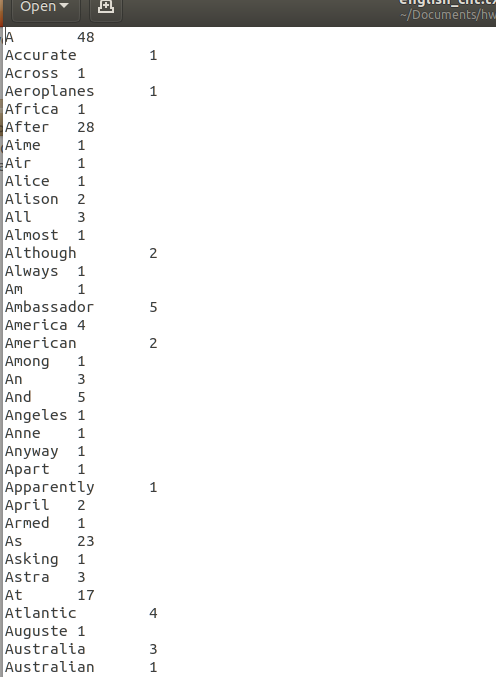
    return str\_new;

  }

## 截图







## 结果文件

见附件中的english\_cnt.txt

## 心得

该程序的输入是hdfs上的input文件夹，输出为output文件夹。主程序将input文件夹里的文件分片并map到各个节点，每个节点对收到的字符串进行处理，并声称一些键值对，最后reduce节点再将这些键值对汇总，写入output文件夹中的结果文件。