Hadoop数据

**一．**

**数据获取：**

1.代码：

获取的数据是2017年三明、泉州的某些月份的天气：

所以先获取获取时间，通过时间的增加来得到天气的遍历。



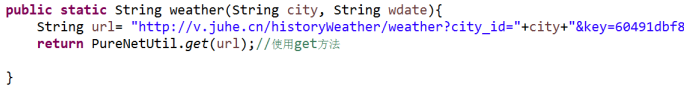




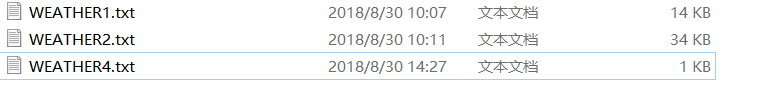


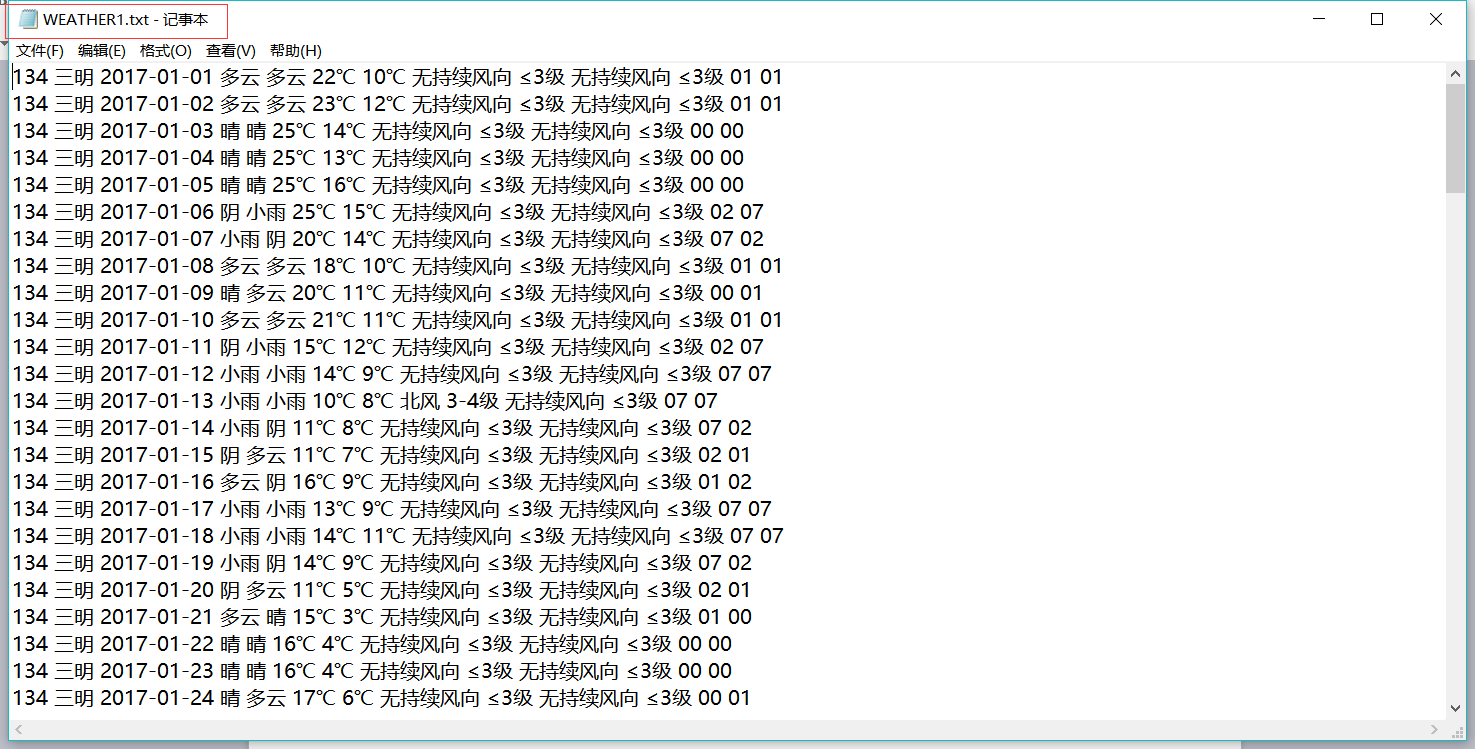
这文档名字可以改多次。

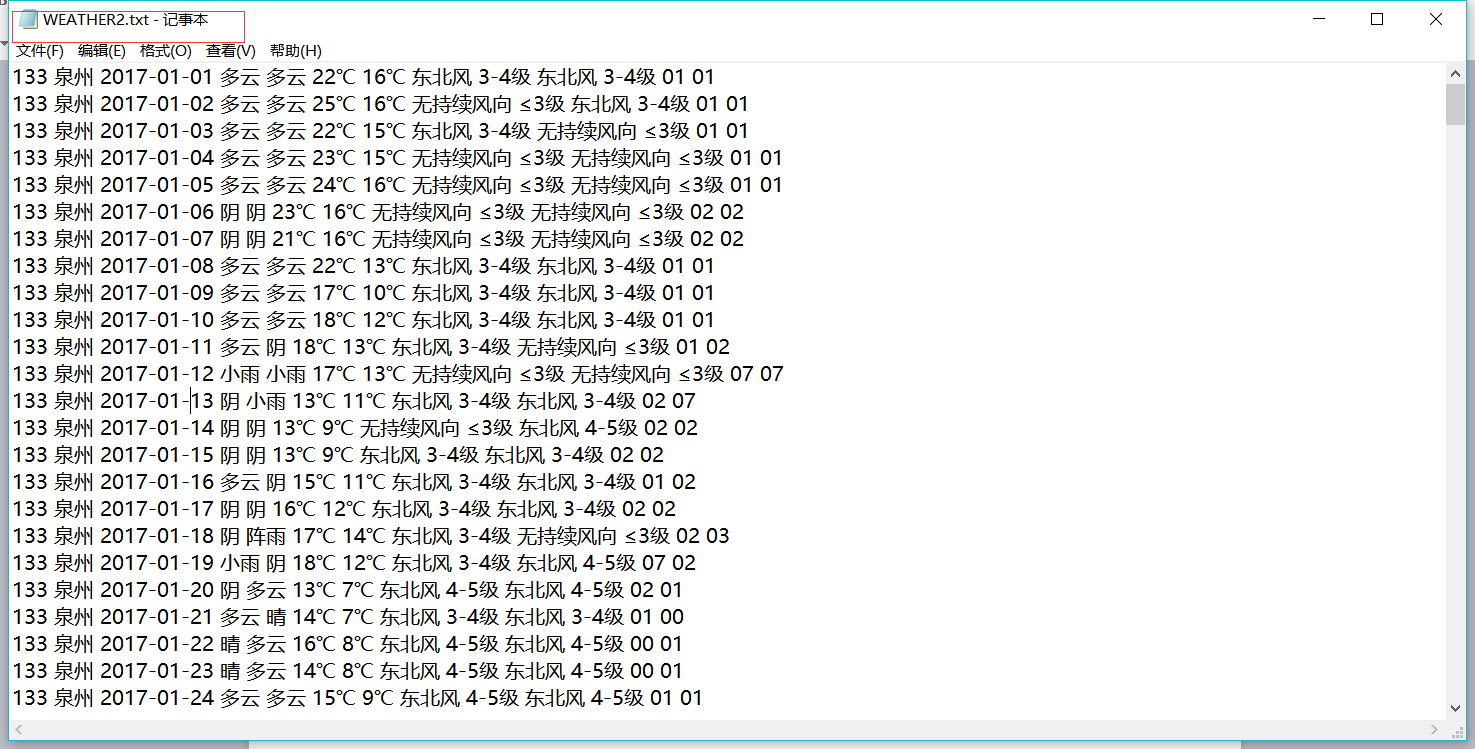
主函数调用类GetCityList的weather方法：

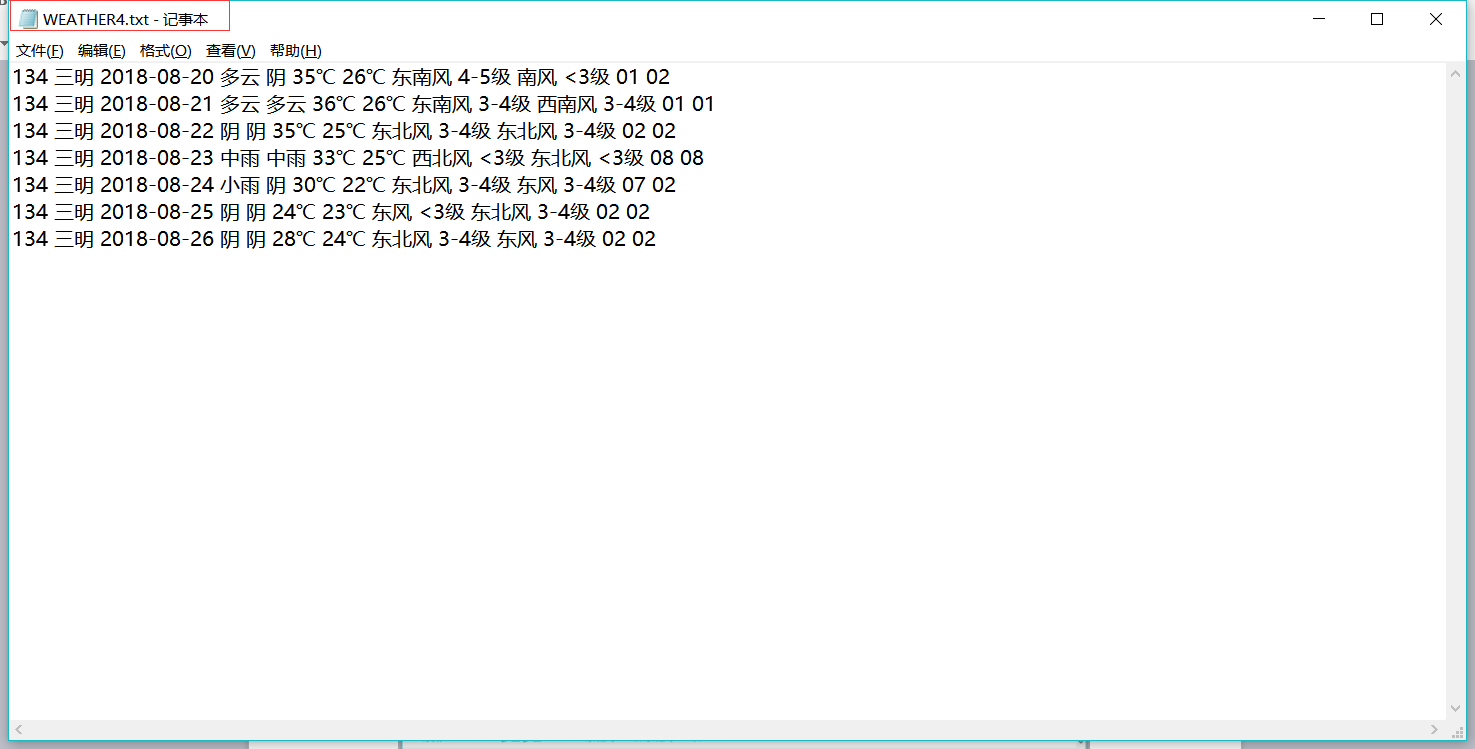


数据结果：







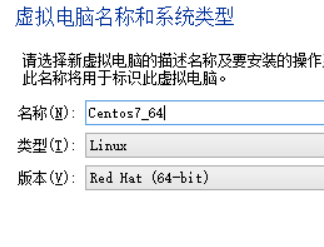


**环境搭建：**

新建虚拟机



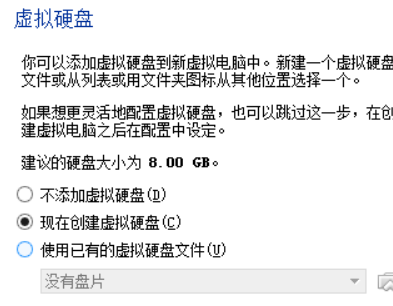
选择类型版本



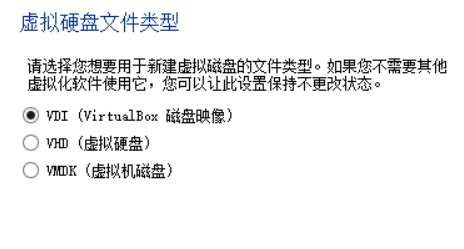
选择内存：



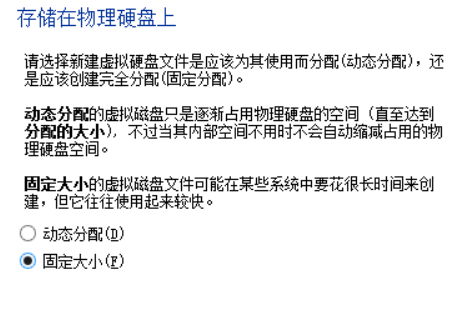
现在创建虚拟硬盘

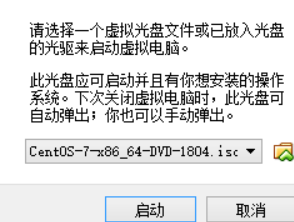


选择VDI

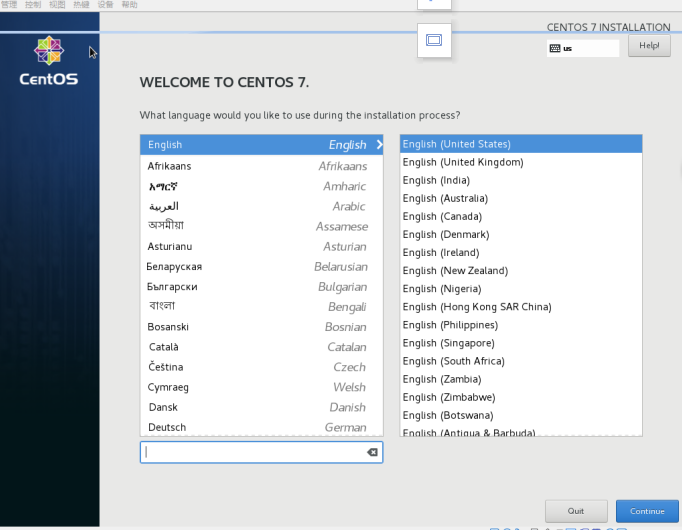


硬盘足够就选择 固定大小启动之后选择Centos的安装文件

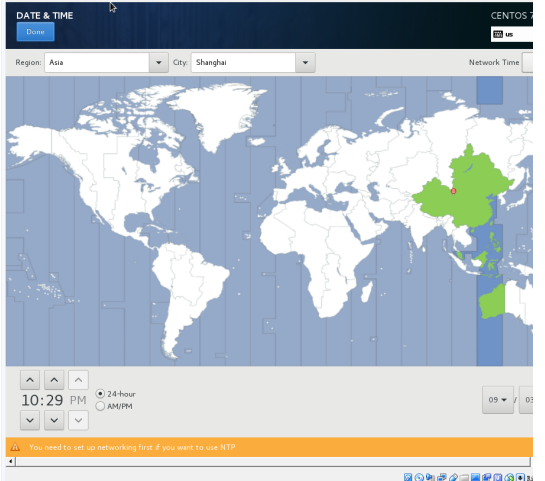




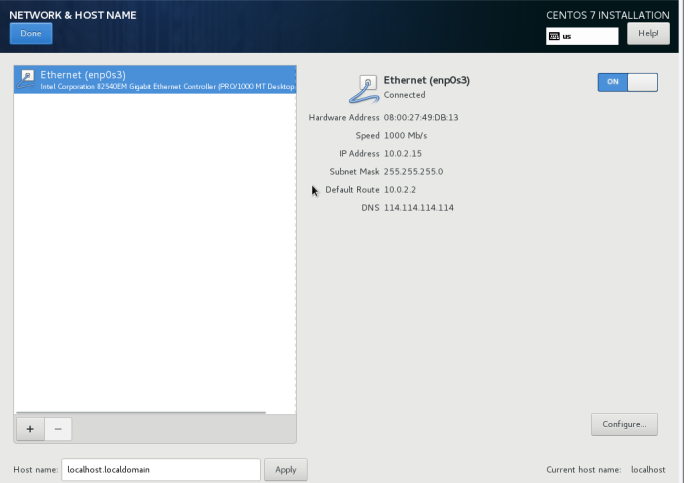
稍后就回出现一个图形化界面，选择语言



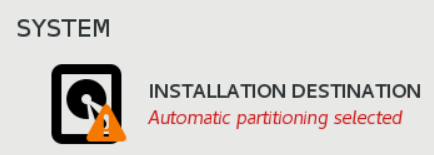
时间选上海



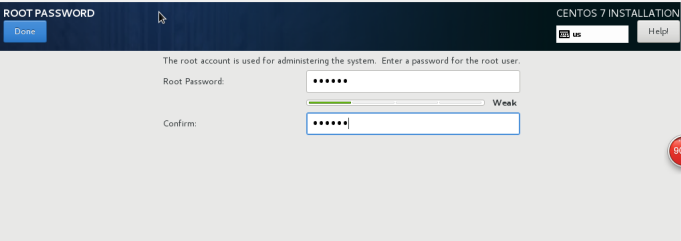
网络打开



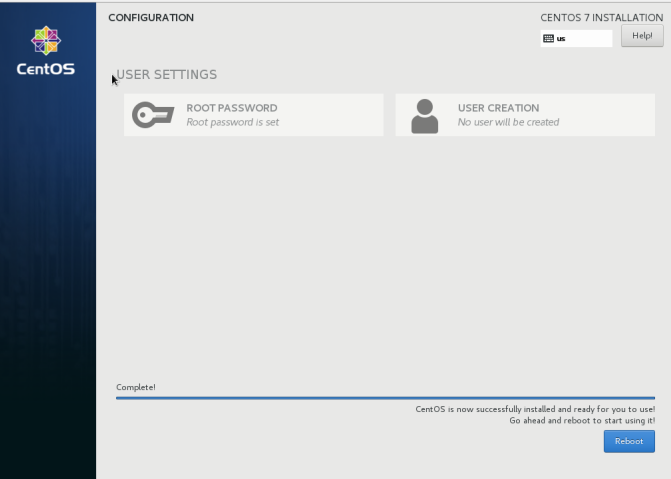
点进去在重新点done之后“！”这标志就没有了



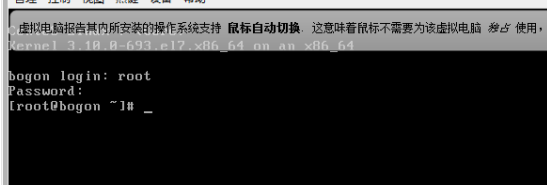
设置账号密码：



点done自动安装



安装完成后面启动登陆



**登陆进去进行配置：**

1.安装ifconfig

yum install net-tools

2.安装jdk

//把本机公钥送入到目标机器的authorized\_keys文件中

ssh [root@192.168.4.223](mailto:root@192.168.4.223) ‘ cat >> .ssh/authorized\_keys' < ~/.ssh/id\_rsa.pub

上传文件

scp jdk-8u144-linux-x64.tar.gz root@192.168.3.181:~/.

解压jdk

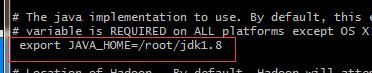
tar xvf jdk-...

建立软连接

ln -s jdk1.8.0\_144/ jdk1.8

3.设置JAVA\_HOME

设置path路径



4.在每个节点建立目录

cd /root/hadoop/data

mkdir -p datanode namenode tmp localdir logdir

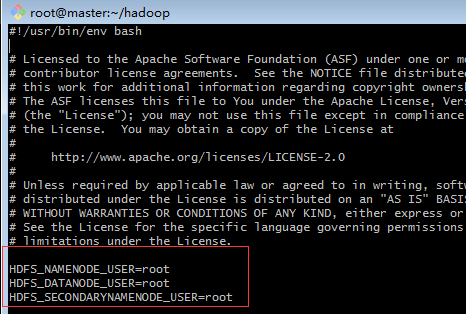
//清空目录内容， 为 下次启动准备

rm -rf datanode/\* namenode/\* localdir/\* logdir/\* tmp/\*

5修改/bin/hdfs中的start-dfs.sh、stop-dfs.sh文件，添加以下内容

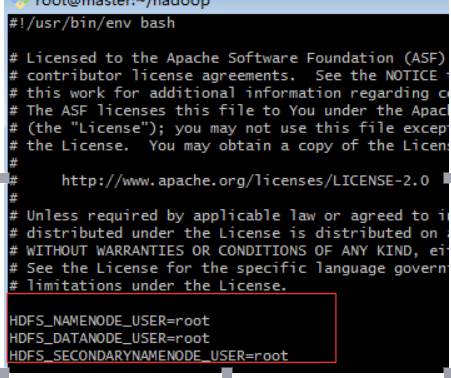
start-dfs.sh：





stop-dfs.sh：

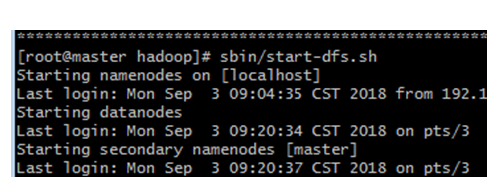




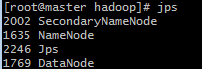
6、格式化



7启动

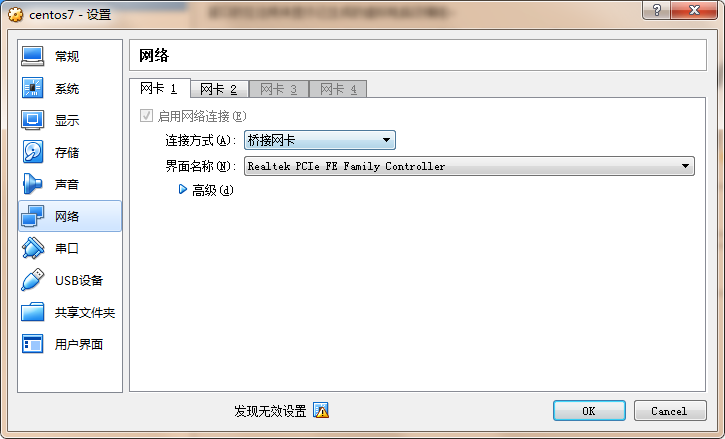


8、查看结果

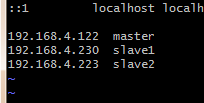


**联机操作**

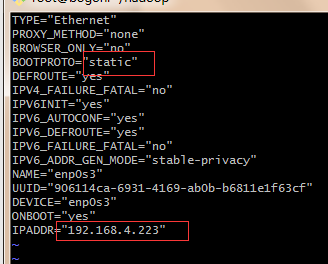
1.网卡设置成桥接网卡



1. 三台机子的虚拟IP



1. 配置静态ip



1. slave1、slave2中的公钥传到master中的authorized\_key里面





master中的公钥传到slave1和slave2中的authorized\_key





5.三台机器已ping通，可以无密码访问：

IMG_256

IMG_256

IMG_256

IMG_256

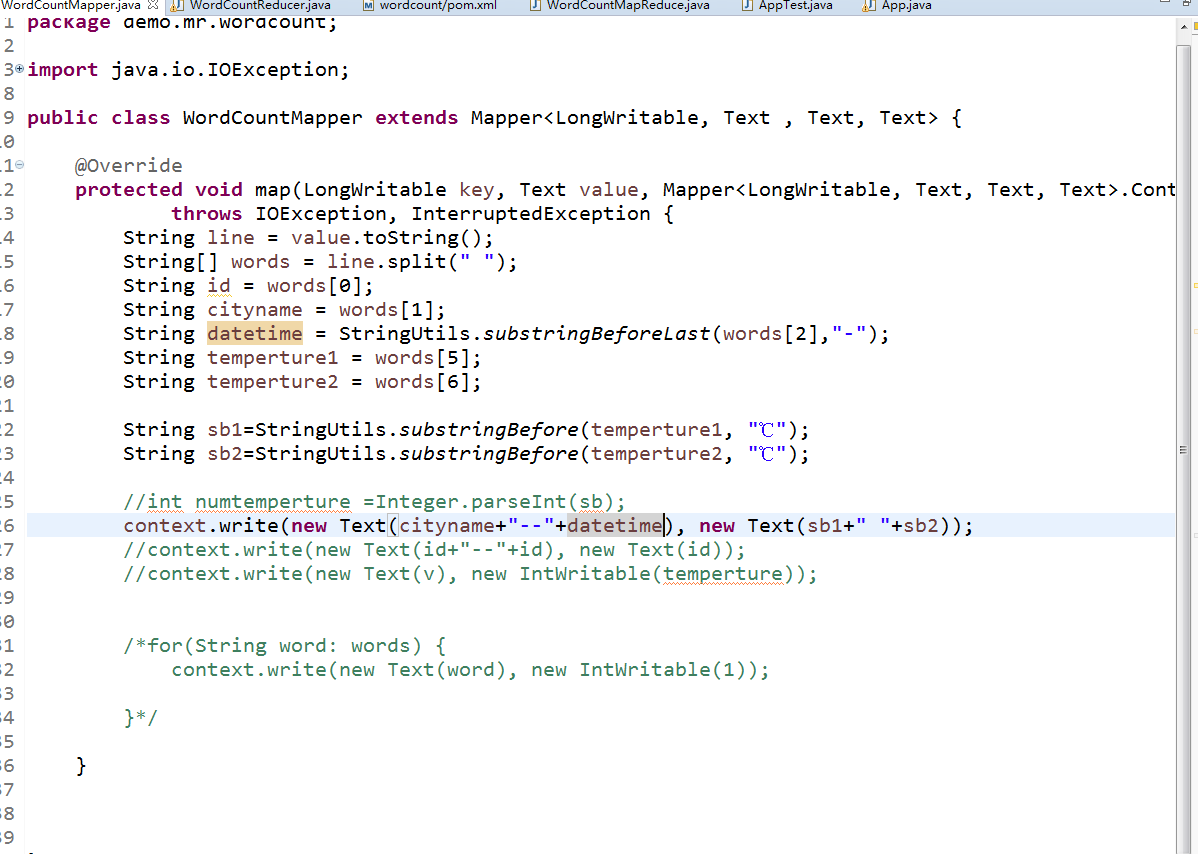
到这配置基本完成

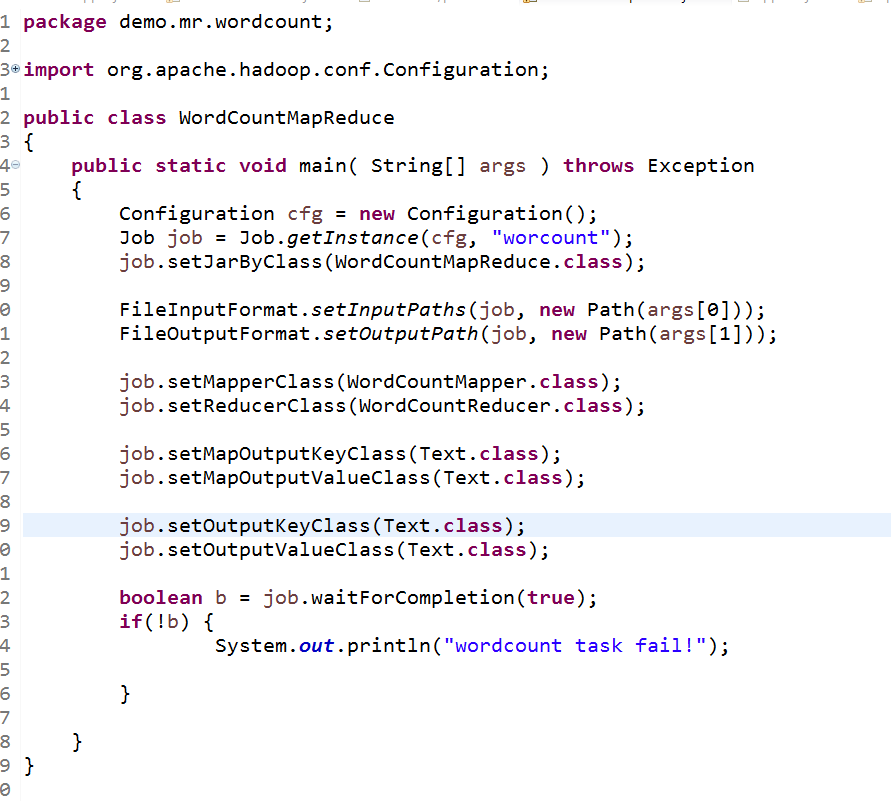
三、

**数据分析**

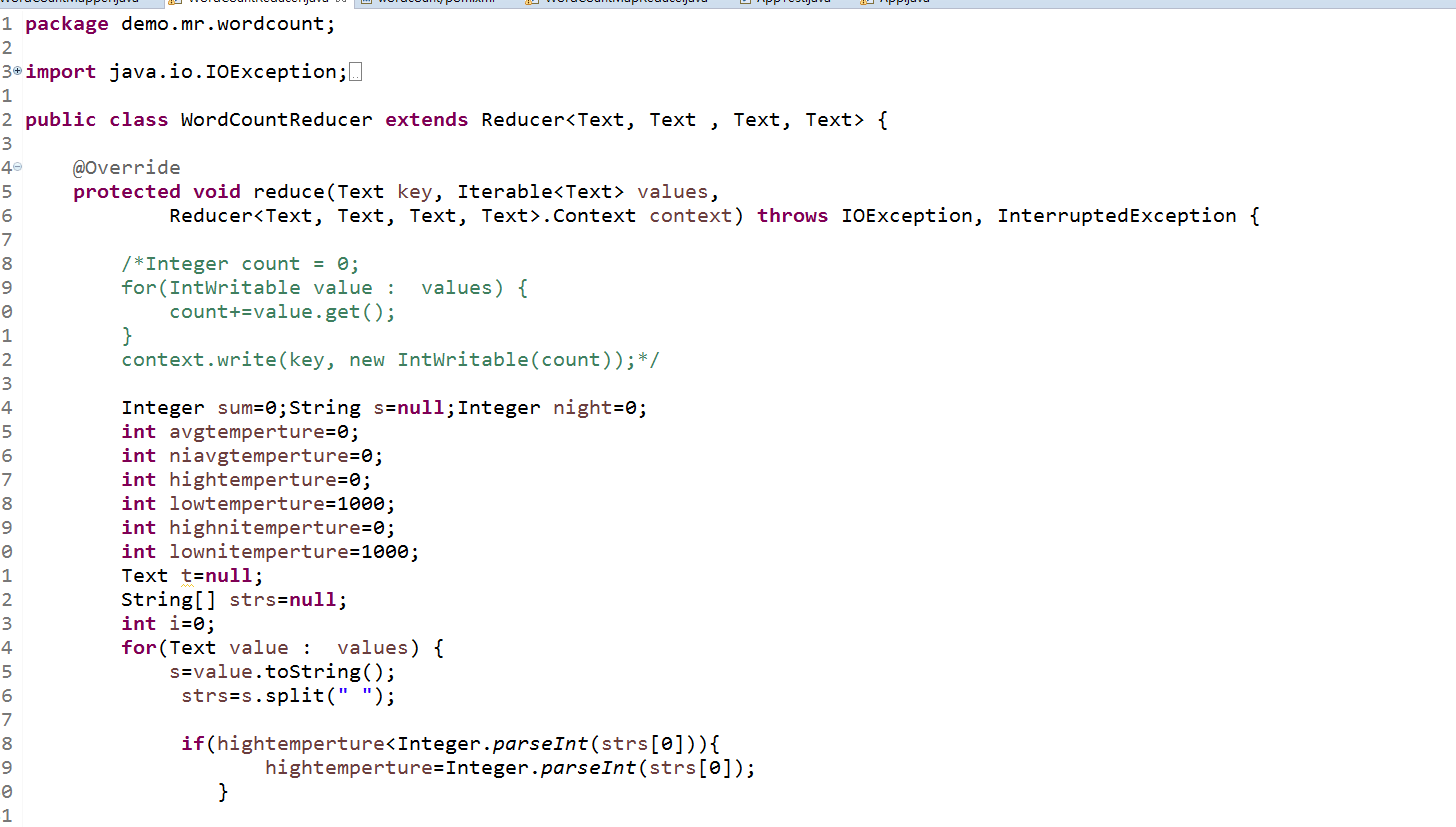
1. 由于前面已经得到数据了，再把这些转换成我们需要的或者是统计的数据。
2. 代码：

把文档数据传值进来





利用传进来的值进行比较数据或者统计数据。





把这份代码上传虚拟机上，并创建目录：



bin/hdfs dfs -put \*.txt /wordcount/input 把所有文档放在input目录

进行数据分析



结果：

