大数据数据初步分析

一、数据抓取：

（一）数据获取代码：

```
public class App {
    public static void main(String[] args) {
        String str = "";

        for (int dtime = -186; dtime < -2; dtime++) {
            java.text.SimpleDateFormat format = new java.text.SimpleDateFormat(
                    "yyyy-MM-dd");
            Calendar cal = Calendar.getInstance();// 取当前日期。
            cal = Calendar.getInstance();
            cal.add(Calendar.DAY_OF_MONTH, dtime);// 取当前日期的前N天.
            str = format.format(cal.getTime());

            String res = GetCityList.weather("119", str);//闽侯

            JSONObject obj = JSONObject.fromObject(res);

            String result = obj.getString("result");
            // 此时result中数据有多个key,可以对其key进行遍历,得到对个属性
            obj = JSONObject.fromObject(result);
            // 今日温度对应的key是today

String city_id = obj.getString("city_id");// 城市地区ID
String city_name = obj.getString("city_name");// 城市地区名称
String weather_date = obj.getString("weather_date");// 天气日期
String day_weather = obj.getString("day_weather");// 白天天气
String night_weather = obj.getString("night_weather");// 夜间天气
String day_temp = obj.getString("day_temp");// 白天最高温度
String night_temp = obj.getString("night_temp");// 夜间最低温度
String day_wind = obj.getString("day_wind");// 白天风向
String day_wind_comp = obj.getString("day_wind_comp");// 白天风力
String night_wind = obj.getString("night_wind");// 夜间风向
String night_wind_comp = obj.getString("night_wind_comp");// 夜间风力
String day_weather_id = obj.getString("day_weather_id");// 白天天气标识
String night_weather_id = obj.getString("night_weather_id");// 夜间天气标识
System.out.println(city_name + " " + weather_date + " "
        + day_weather + " " + night_weather + " " + day_temp + " "
        + night_temp + " " + day_wind + " " + day_wind_comp + " "
        + night_wind + " " + night_wind_comp + " " + day_weather_id
        + " " + night_weather_id);
```

```java
List<String> list = new LinkedList<String>();
list.add(city_id);
list.add(city_name);
list.add(weather_date);
list.add(day_weather);
list.add(night_weather);
list.add(day_temp);
list.add(night_temp);
list.add(day_wind);
list.add(day_wind_comp);
list.add(night_wind);
list.add(night_wind_comp);
list.add(day_weather_id);
list.add(night_weather_id);
```

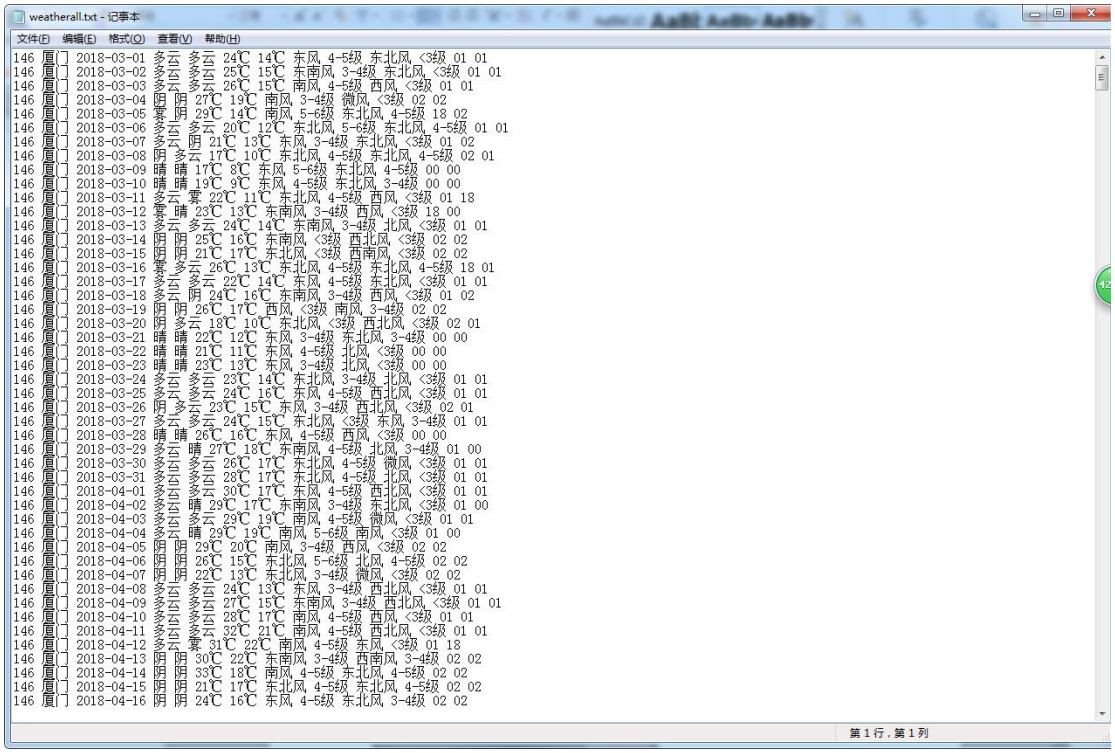
写入文件：

```java
File file1 = new File("F:\\weather.txt");
try {
    FileWriter fw = new FileWriter(file1, true);
    BufferedWriter bw = new BufferedWriter(fw);

    for (int i = 0; i < list.size(); i++) {
        bw.write(list.get(i).toString() + " ");
        bw.flush();
    }
    bw.newLine();
    bw.close();
    fw.close();

} catch (IOException e) {
    e.printStackTrace();
}
```

（二）获取的数据：



weatherall.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

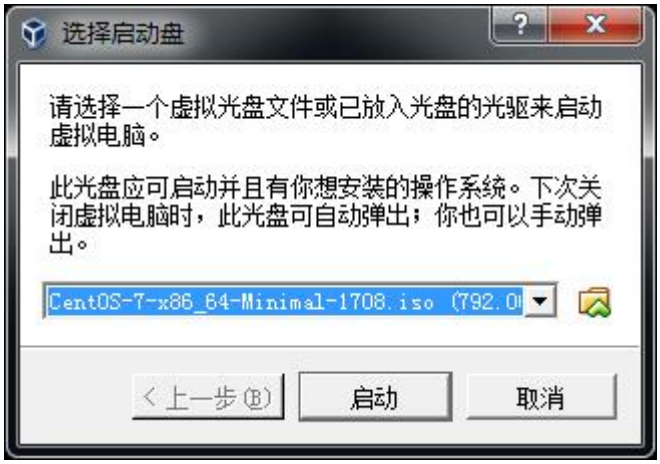
```
146 厦门 2018-03-01 多云 多云 24℃ 14℃ 东风,4-5级 东北风,<3级 01 01
146 厦门 2018-03-02 多云 多云 25℃ 15℃ 东南风,3-4级 东北风,<3级 01 01
146 厦门 2018-03-03 多云 多云 26℃ 15℃ 南风,4-5级 西风,<3级 01 01
146 厦门 2018-03-04 阴 阴 27℃ 19℃ 南风,3-4级 微风,<3级 02 02
146 厦门 2018-03-05 雾 阴 29℃ 14℃ 南风,5-6级 东北风,4-5级 18 02
146 厦门 2018-03-06 多云 多云 20℃ 12℃ 东北风,5-6级 东北风,4-5级 01 01
146 厦门 2018-03-07 多云 阴 21℃ 13℃ 东风,3-4级 东北风,<3级 01 02
146 厦门 2018-03-08 阴 多云 17℃ 10℃ 东北风,4-5级 东北风,4-5级 02 01
146 厦门 2018-03-09 晴 晴 17℃ 8℃ 东风,5-6级 东北风,4-5级 00 00
146 厦门 2018-03-10 晴 晴 19℃ 9℃ 东风,4-5级 东北风,3-4级 00 00
146 厦门 2018-03-11 多云 雾 22℃ 11℃ 东北风,4-5级 西风,<3级 01 18
146 厦门 2018-03-12 雾 晴 23℃ 13℃ 东南风,3-4级 西风,<3级 18 00
146 厦门 2018-03-13 多云 多云 24℃ 14℃ 东南风,3-4级 北风,<3级 01 01
146 厦门 2018-03-14 阴 阴 25℃ 16℃ 东南风,<3级 西北风,<3级 02 02
146 厦门 2018-03-15 阴 阴 21℃ 17℃ 东北风,<3级 西南风,<3级 02 02
146 厦门 2018-03-16 雾 多云 26℃ 13℃ 东北风,4-5级 东北风,4-5级 18 01
146 厦门 2018-03-17 多云 多云 22℃ 14℃ 东风,4-5级 东北风,<3级 01 01
146 厦门 2018-03-18 多云 阴 24℃ 16℃ 东南风,3-4级 西风,<3级 01 02
146 厦门 2018-03-19 阴 阴 26℃ 17℃ 西风,<3级 南风,3-4级 02 02
146 厦门 2018-03-20 阴 多云 18℃ 10℃ 东北风,<3级 西北风,<3级 02 01
146 厦门 2018-03-21 晴 晴 22℃ 12℃ 东风,3-4级 东北风,3-4级 00 00
146 厦门 2018-03-22 晴 晴 21℃ 11℃ 东风,4-5级 北风,<3级 00 00
146 厦门 2018-03-23 晴 晴 23℃ 13℃ 东风,3-4级 北风,<3级 00 00
146 厦门 2018-03-24 多云 多云 23℃ 14℃ 东北风,3-4级 北风,<3级 01 01
146 厦门 2018-03-25 多云 多云 24℃ 16℃ 东风,4-5级 西北风,<3级 01 01
146 厦门 2018-03-26 阴 多云 23℃ 15℃ 东风,3-4级 西北风,<3级 02 01
146 厦门 2018-03-27 多云 多云 24℃ 15℃ 东北风,<3级 东风,3-4级 01 01
146 厦门 2018-03-28 晴 晴 26℃ 16℃ 东风,4-5级 西风,<3级 00 00
146 厦门 2018-03-29 多云 晴 27℃ 18℃ 东南风,4-5级 北风,3-4级 01 00
146 厦门 2018-03-30 多云 多云 26℃ 17℃ 东北风,4-5级 微风,<3级 01 01
146 厦门 2018-03-31 多云 多云 28℃ 17℃ 东北风,4-5级 北风,<3级 01 01
146 厦门 2018-04-01 多云 多云 30℃ 17℃ 东风,4-5级 西北风,<3级 01 01
146 厦门 2018-04-02 多云 晴 29℃ 17℃ 东南风,3-4级 东北风,<3级 01 00
146 厦门 2018-04-03 多云 多云 29℃ 19℃ 南风,4-5级 微风,<3级 01 01
146 厦门 2018-04-04 多云 晴 29℃ 19℃ 南风,5-6级 南风,<3级 01 00
146 厦门 2018-04-05 阴 阴 29℃ 20℃ 南风,3-4级 西风,<3级 02 02
146 厦门 2018-04-06 阴 阴 26℃ 15℃ 东北风,5-6级 北风,4-5级 02 02
146 厦门 2018-04-07 阴 阴 22℃ 13℃ 东北风,3-4级 微风,<3级 02 02
146 厦门 2018-04-08 多云 多云 24℃ 13℃ 东风,3-4级 西北风,<3级 01 01
146 厦门 2018-04-09 多云 多云 27℃ 15℃ 东南风,3-4级 西北风,<3级 01 01
146 厦门 2018-04-10 多云 多云 28℃ 17℃ 南风,4-5级 西风,<3级 01 01
146 厦门 2018-04-11 多云 多云 32℃ 21℃ 南风,4-5级 西北风,<3级 01 01
146 厦门 2018-04-12 多云 雾 31℃ 22℃ 南风,4-5级 东风,<3级 01 18
146 厦门 2018-04-13 阴 阴 30℃ 22℃ 东南风,3-4级 西南风,3-4级 02 02
146 厦门 2018-04-14 阴 阴 33℃ 18℃ 南风,4-5级 东北风,4-5级 02 02
146 厦门 2018-04-15 阴 阴 21℃ 17℃ 东北风,4-5级 东北风,4-5级 02 02
146 厦门 2018-04-16 阴 阴 24℃ 16℃ 东风,4-5级 东北风,3-4级 02 02
```

第 1 行，第 1 列

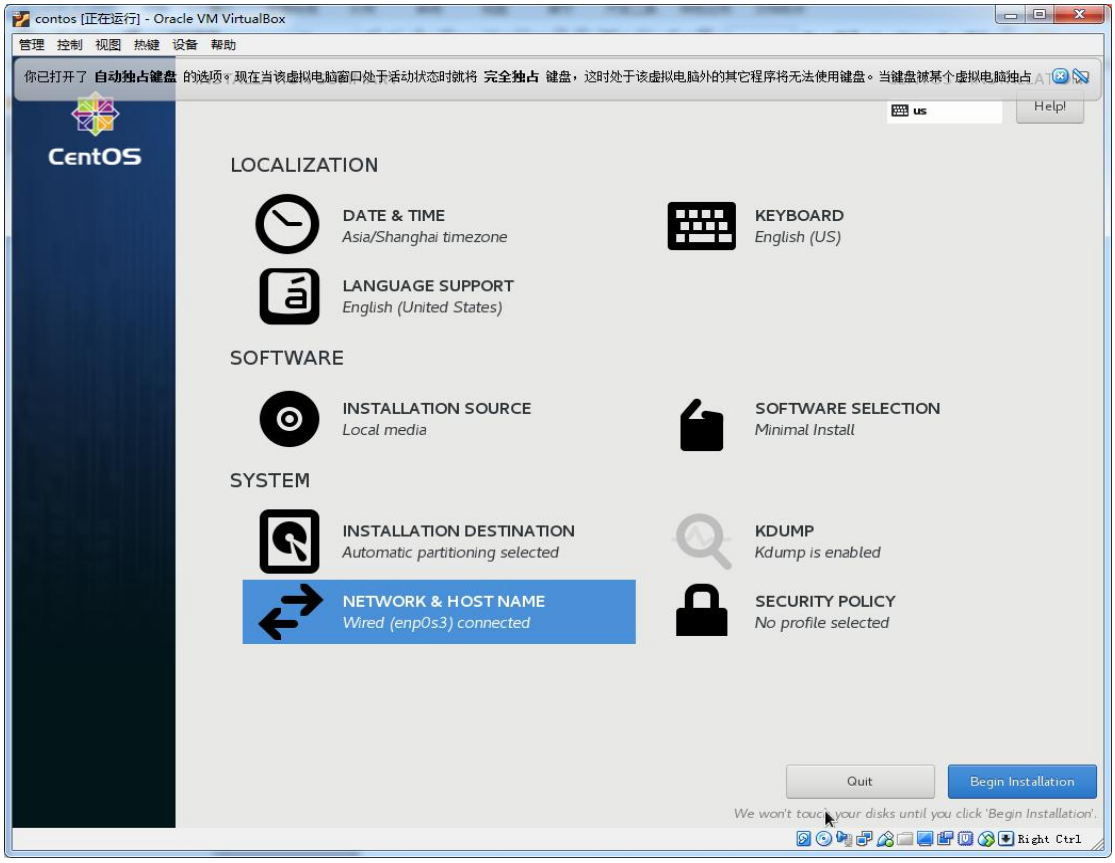
二、环境搭建：

（一）创建新的虚拟机

都为默认选项。



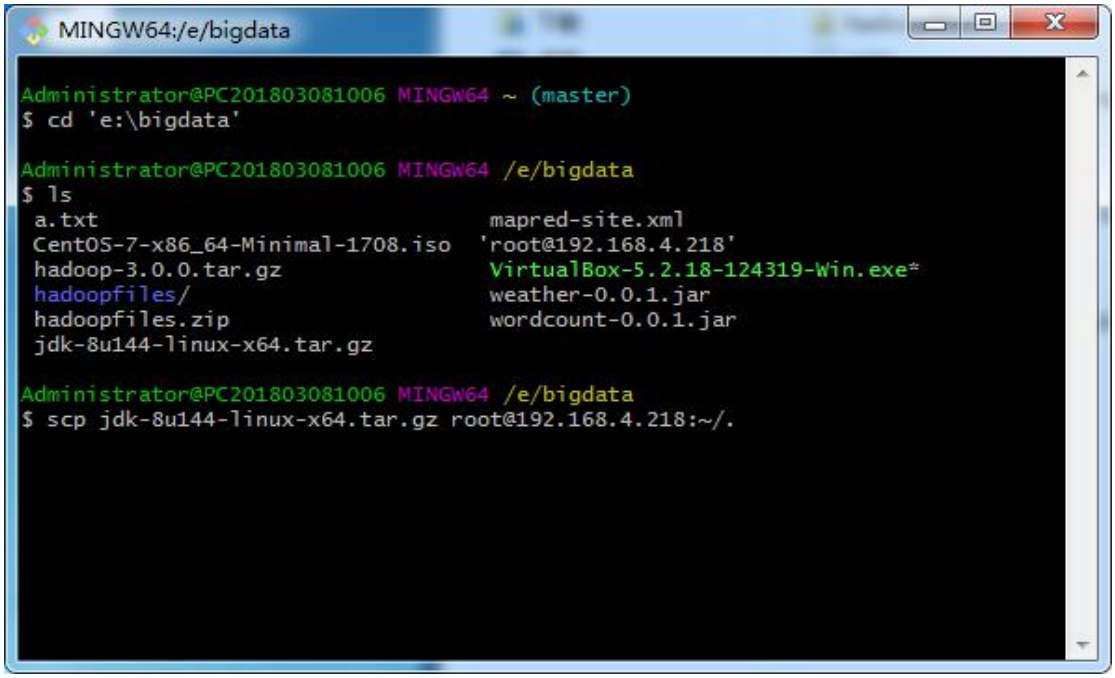
安装完成后启动，选择 LINUX 镜像进行系统配置



设置 DATA&TIME,打开 NETWORK&HOSTNAME，设置密码之后等待完成安装

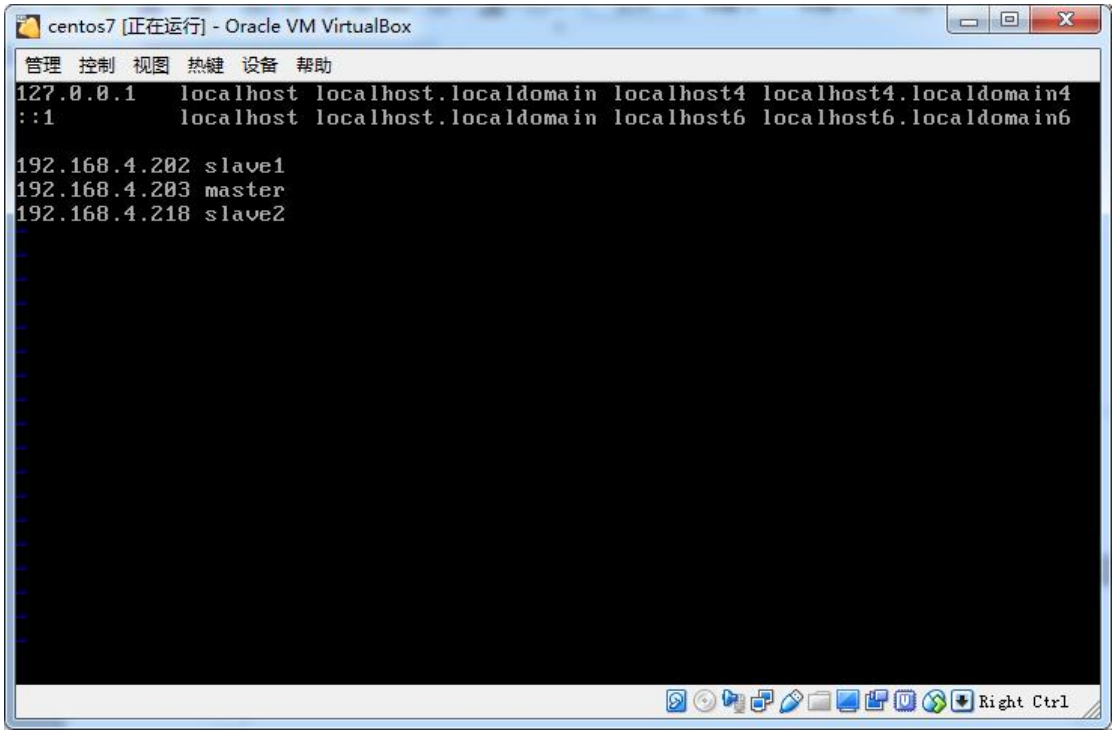
（一）虚拟机搭建

上传 JDK 等需要用到的工具，进行安装



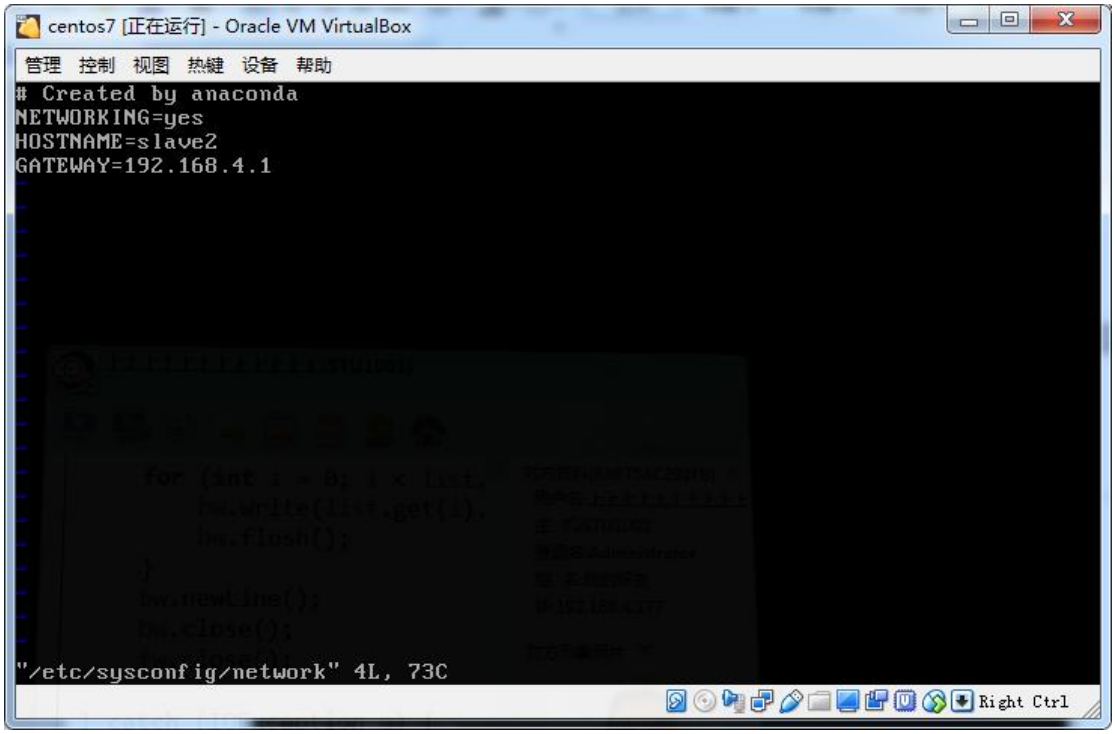
分别设置三台虚拟机的 hosts

```
[root@slave2 ~]# vi /etc/hosts_
```



设置网关

```
[root@slave2 ~]# vi /etc/sysconfig/network
```



分别配置三台虚拟机的 hadoop

```
[root@slave2 hadoop]# cd ~/hadoop
[root@slave2 hadoop]# vi sbin/start-dfs.sh
[root@slave2 hadoop]# vi sbin/stop-dfs.sh
[root@slave2 hadoop]# vi sbin/start-yarn.sh
[root@slave2 hadoop]# vi sbin/stop-yarn.sh
[root@slave2 hadoop]#
```

配置 etc/hadoop 中的 core-size.xml 配置为当前虚拟机的 hostname

以上环境就基本搭建完成

三、数据分析

数据分析包括了对获取到进行每个月的平均温度的分析，还有每个月白天以及夜间最高和最低温度的分析。

（一）数据分析代码：

```java
@Override
protected void reduce(Text key, Iterable<Text> values,
        Reducer<Text, Text, Text, Text>.Context context)
        throws IOException, InterruptedException {

    Integer sum = 0;
    Integer nisum = 0;
    String s = null;
    int avgtemperture = 0;
    int niavgtemperture = 0;
    Text t = null;
    int i = 0;
    List<Integer> listDay = new ArrayList<Integer>();// 存白天温度
    List<Integer> listNight = new ArrayList<Integer>();// 存夜间温度

    for (Text value : values) {
        s = value.toString();
        String[] words = s.split("-");// 切割，words[0]为白天温度，words[1]为夜间温度
        sum += Integer.parseInt(words[0]);
        nisum += Integer.parseInt(words[1]);

        listDay.add(Integer.parseInt(words[0]));
        listNight.add(Integer.parseInt(words[1]));

        i++;
    }

    // 计算平均气温
    avgtemperture = sum / i;
    niavgtemperture = nisum / i;

    // 气温排序
    Collections.sort(listDay);
    Collections.sort(listNight);

    int MinNight = listNight.get(0);
    int MaxNight = listNight.get(listNight.size() - 1);
    int MinDay = listDay.get(0);
    int MaxDay = listDay.get(listDay.size() - 1);

    t = new Text("白天平均温度是" + avgtemperture + "℃ 夜间平均温度是"
            + niavgtemperture + "℃ 白天最低温度是" + MinDay + "℃ 白天最高温度是"
            + MaxDay + "℃ 夜间最低温度是" + MinNight + "℃ 夜间最高温度是" + MaxNight
            + "℃");
    context.write(key, new Text(t));
```

```java
public class WordCountMapper extends Mapper<LongWritable, Text, Text, Text> {

    @Override
    protected void map(LongWritable key, Text value,
            Mapper<LongWritable, Text, Text, Text>.Context context)
            throws IOException, InterruptedException {
        //例:146 厦门 2018-03-01 多云 多云 24℃ 14℃ 东风 4-5级 东北风 <3级 01 01
        String line = value.toString();
        String[] words = line.split(" ");
        String id = words[0];
        String cityname = words[1];
        String datetime = StringUtils.substringBeforeLast(words[2], "-");
        String dayTemperture = words[5];
        String nightTemperture = words[6];
        String dayTemp = StringUtils.substringBefore(dayTemperture, "℃");
        String nightTemp = StringUtils.substringBefore(nightTemperture, "℃");
        context.write(new Text(cityname + "--" + datetime), new Text(dayTemp + "-" + nightTemp));
    }
}
```

我们选取了 6 个城市，从今年的 3 月份到 8 月份的天气数据进行分析

（二）数据分析结果：

```
[root@slave2 hadoop]# bin/hdfs dfs -cat /w1/output/*
厦门--2018-03   自天平均温度是23℃   夜间平均温度是14℃   自天最低温度是17℃   自天最高温度是29℃   夜间最低温度是8℃   夜间最高温度是19℃
厦门--2018-04   自天平均温度是27℃   夜间平均温度是18℃   自天最低温度是21℃   自天最高温度是33℃   夜间最低温度是13℃   夜间最高温度是23℃
厦门--2018-05   自天平均温度是30℃   夜间平均温度是23℃   自天最低温度是24℃   自天最高温度是36℃   夜间最低温度是18℃   夜间最高温度是28℃
厦门--2018-06   自天平均温度是29℃   夜间平均温度是24℃   自天最低温度是25℃   自天最高温度是34℃   夜间最低温度是21℃   夜间最高温度是26℃
厦门--2018-07   自天平均温度是32℃   夜间平均温度是26℃   自天最低温度是29℃   自天最高温度是36℃   夜间最低温度是23℃   夜间最高温度是28℃
厦门--2018-08   自天平均温度是31℃   夜间平均温度是25℃   自天最低温度是25℃   自天最高温度是35℃   夜间最低温度是23℃   夜间最高温度是27℃
杭州--2018-03   自天平均温度是18℃   夜间平均温度是9℃   自天最低温度是10℃   自天最高温度是28℃   夜间最低温度是2℃   夜间最高温度是15℃
杭州--2018-04   自天平均温度是24℃   夜间平均温度是14℃   自天最低温度是15℃   自天最高温度是31℃   夜间最低温度是6℃   夜间最高温度是23℃
杭州--2018-05   自天平均温度是27℃   夜间平均温度是20℃   自天最低温度是19℃   自天最高温度是37℃   夜间最低温度是14℃   夜间最高温度是27℃
杭州--2018-06   自天平均温度是29℃   夜间平均温度是22℃   自天最低温度是22℃   自天最高温度是36℃   夜间最低温度是18℃   夜间最高温度是29℃
杭州--2018-07   自天平均温度是34℃   夜间平均温度是26℃   自天最低温度是28℃   自天最高温度是38℃   夜间最低温度是24℃   夜间最高温度是28℃
杭州--2018-08   自天平均温度是33℃   夜间平均温度是26℃   自天最低温度是25℃   自天最高温度是37℃   夜间最低温度是23℃   夜间最高温度是29℃
海口--2018-03   自天平均温度是26℃   夜间平均温度是19℃   自天最低温度是17℃   自天最高温度是35℃   夜间最低温度是14℃   夜间最高温度是23℃
海口--2018-04   自天平均温度是28℃   夜间平均温度是21℃   自天最低温度是17℃   自天最高温度是36℃   夜间最低温度是15℃   夜间最高温度是27℃
海口--2018-05   自天平均温度是33℃   夜间平均温度是25℃   自天最低温度是30℃   自天最高温度是35℃   夜间最低温度是23℃   夜间最高温度是27℃
海口--2018-06   自天平均温度是31℃   夜间平均温度是25℃   自天最低温度是26℃   自天最高温度是36℃   夜间最低温度是23℃   夜间最高温度是28℃
海口--2018-07   自天平均温度是31℃   夜间平均温度是25℃   自天最低温度是28℃   自天最高温度是36℃   夜间最低温度是24℃   夜间最高温度是28℃
海口--2018-08   自天平均温度是30℃   夜间平均温度是25℃   自天最低温度是26℃   自天最高温度是34℃   夜间最低温度是23℃   夜间最高温度是28℃
深圳--2018-03   自天平均温度是24℃   夜间平均温度是17℃   自天最低温度是15℃   自天最高温度是29℃   夜间最低温度是10℃   夜间最高温度是21℃
深圳--2018-04   自天平均温度是27℃   夜间平均温度是20℃   自天最低温度是20℃   自天最高温度是31℃   夜间最低温度是15℃   夜间最高温度是25℃
深圳--2018-05   自天平均温度是31℃   夜间平均温度是25℃   自天最低温度是25℃   自天最高温度是35℃   夜间最低温度是21℃   夜间最高温度是29℃
深圳--2018-06   自天平均温度是30℃   夜间平均温度是26℃   自天最低温度是26℃   自天最高温度是34℃   夜间最低温度是24℃   夜间最高温度是28℃
深圳--2018-07   自天平均温度是31℃   夜间平均温度是26℃   自天最低温度是27℃   自天最高温度是34℃   夜间最低温度是24℃   夜间最高温度是28℃
深圳--2018-08   自天平均温度是30℃   夜间平均温度是25℃   自天最低温度是27℃   自天最高温度是34℃   夜间最低温度是24℃   夜间最高温度是29℃
连江--2018-03   自天平均温度是20℃   夜间平均温度是11℃   自天最低温度是14℃   自天最高温度是25℃   夜间最低温度是4℃   夜间最高温度是16℃
连江--2018-04   自天平均温度是24℃   夜间平均温度是16℃   自天最低温度是18℃   自天最高温度是31℃   夜间最低温度是9℃   夜间最高温度是24℃
连江--2018-05   自天平均温度是29℃   夜间平均温度是22℃   自天最低温度是18℃   自天最高温度是35℃   夜间最低温度是15℃   夜间最高温度是26℃
连江--2018-06   自天平均温度是29℃   夜间平均温度是23℃   自天最低温度是21℃   自天最高温度是36℃   夜间最低温度是19℃   夜间最高温度是27℃
连江--2018-07   自天平均温度是32℃   夜间平均温度是25℃   自天最低温度是28℃   自天最高温度是37℃   夜间最低温度是24℃   夜间最高温度是28℃
连江--2018-08   自天平均温度是32℃   夜间平均温度是25℃   自天最低温度是26℃   自天最高温度是36℃   夜间最低温度是24℃   夜间最高温度是28℃
闽侯--2018-03   自天平均温度是22℃   夜间平均温度是12℃   自天最低温度是15℃   自天最高温度是29℃   夜间最低温度是4℃   夜间最高温度是18℃
闽侯--2018-04   自天平均温度是26℃   夜间平均温度是16℃   自天最低温度是18℃   自天最高温度是32℃   夜间最低温度是8℃   夜间最高温度是24℃
闽侯--2018-05   自天平均温度是31℃   夜间平均温度是22℃   自天最低温度是19℃   自天最高温度是38℃   夜间最低温度是15℃   夜间最高温度是27℃
闽侯--2018-06   自天平均温度是30℃   夜间平均温度是23℃   自天最低温度是22℃   自天最高温度是38℃   夜间最低温度是19℃   夜间最高温度是27℃
闽侯--2018-07   自天平均温度是34℃   夜间平均温度是25℃   自天最低温度是29℃   自天最高温度是38℃   夜间最低温度是23℃   夜间最高温度是28℃
闽侯--2018-08   自天平均温度是33℃   夜间平均温度是24℃   自天最低温度是25℃   自天最高温度是38℃   夜间最低温度是23℃   夜间最高温度是27℃
```