

# 191098323-张峻琿-作业5

## 一. 在window上配置java、hadoop、maven、IDEA

### ■ java

之前有配置过，但当时的版本可能有问题，故卸载重来

```
C:\WINDOWS\system32\cmd.exe
有关详细信息，请参阅 http://www.oracle.com/technetwork/java/javase/documentation/index.html。

C:\Users\DELL>java -version
java version "1.8.0_152"
Java(TM) SE Runtime Environment (build 1.8.0_152-b16)
Java HotSpot(TM) 64-Bit Server VM (build 25.152-b16, mixed mode)

C:\Users\DELL>javac.exe
用法: javac <options> <source files>
其中，可能的选项包括:
-g          生成所有调试信息
-g:none     不生成任何调试信息
-g:{lines, vars, source} 只生成某些调试信息
-nowarn     不生成任何警告
-verbose    输出有关编译器正在执行的操作的消息
-deprecation 输出使用已过时的 API 的源位置
-classpath <路径> 指定查找用户类文件和注释处理程序的位置
-cp <路径> 指定查找用户类文件和注释处理程序的位置
-sourcepath <路径> 指定查找输入源文件的位置
-bootclasspath <路径> 覆盖引导类文件的位置
-extdirs <目录> 覆盖所安装扩展的位置
-endorseddirs <目录> 覆盖签名的标准路径的位置
-processor {none, only} 控制是否执行注释处理和/或编译。
-processor <class1>[, <class2>, <class3>, ...] 要运行的注释处理程序的名称; 绕过默认的搜索进程
-processorpath <路径> 指定查找注释处理程序的位置
-parameters 生成元数据以用于方法参数的反射
-d <目录> 指定放置生成的类文件的位置
-s <目录> 指定放置生成的源文件的位置
-h <目录> 指定放置生成的本机头文件的位置
-implicit: {none, class} 指定是否为隐式引用文件生成类文件

C:\WINDOWS\system32\cmd.exe
C:\Users\DELL>java.exe
用法: java [-options] class [args...]
      (执行类)
或 java [-options] -jar jarfile [args...]
      (执行 jar 文件)
其中选项包括:
-d32        使用 32 位数据模型 (如果可用)
-d64        使用 64 位数据模型 (如果可用)
-server     选择 "server" VM
            默认 VM 是 server。

-cp <目录和 zip/jar 文件的类搜索路径>
-classpath <目录和 zip/jar 文件的类搜索路径>
           用 ; 分隔的目录, JAR 档案
           和 ZIP 档案列表, 用于搜索类文件。

-D<名称>=<值> 设置系统属性
-verbose: {class, gc, jit} 启用详细输出
-version     输出产品版本并退出
-version:<值> 警告: 此功能已过时, 将在
              未来发行版中删除。
              需要指定的版本才能运行
-showversion 输出产品版本并继续
-jre-restrict-search | -no-jre-restrict-search
              警告: 此功能已过时, 将在
              未来发行版中删除。
              在版本搜索中包括/排除用户专用 JRE
```

### ■ hadoop

配置hadoop时出了一些问题，到最后一步始终无法通过，查了各种教程，最后才发现是我的java的路径中有空格导致的不行，故重新安装java

1 原因，路径不能包含空格

```
> 电脑 > 本地磁盘 (C:) > Program Files (x86) > Java > jdk1.8.0_144

1 C:\Program Files (x86)\Java\jdk1.8.0_144
```

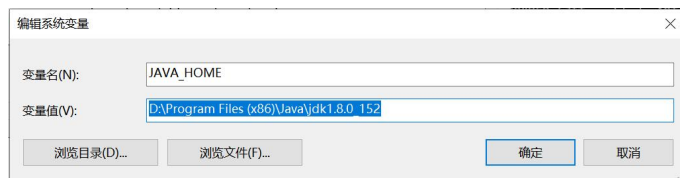
如果你安装jdk不是默认的路径，则不会有这个问题

这里的路径中含有 (x86)，所以weblogic在解析的时候是解析不了括号的。

2 解决办法

重新安装jdk，选择安装路径时候。不要有空格的目录名称。d:\java就ok





最后成功

```
C:\Users\DELL>hadoop version
Hadoop 2.9.1
Subversion https://github.com/apache/hadoop.git -r e30710aea4e6e55e69372929106cf119af06fd0e
Compiled by root on 2018-04-16T09:33Z
Compiled with protoc 2.5.0
From source with checksum 7d6d2b655115c6cc336d662cc2b919bd
This command was run using /D:/Hadoop/hadoop-2.9.1/share/hadoop/common/hadoop-common-2.9.1.jar

C:\Users\DELL>hadoop -version
java version "1.8.0_152"
Java(TM) SE Runtime Environment (build 1.8.0_152-b16)
Java HotSpot(TM) 64-Bit Server VM (build 25.152-b16, mixed mode)
```

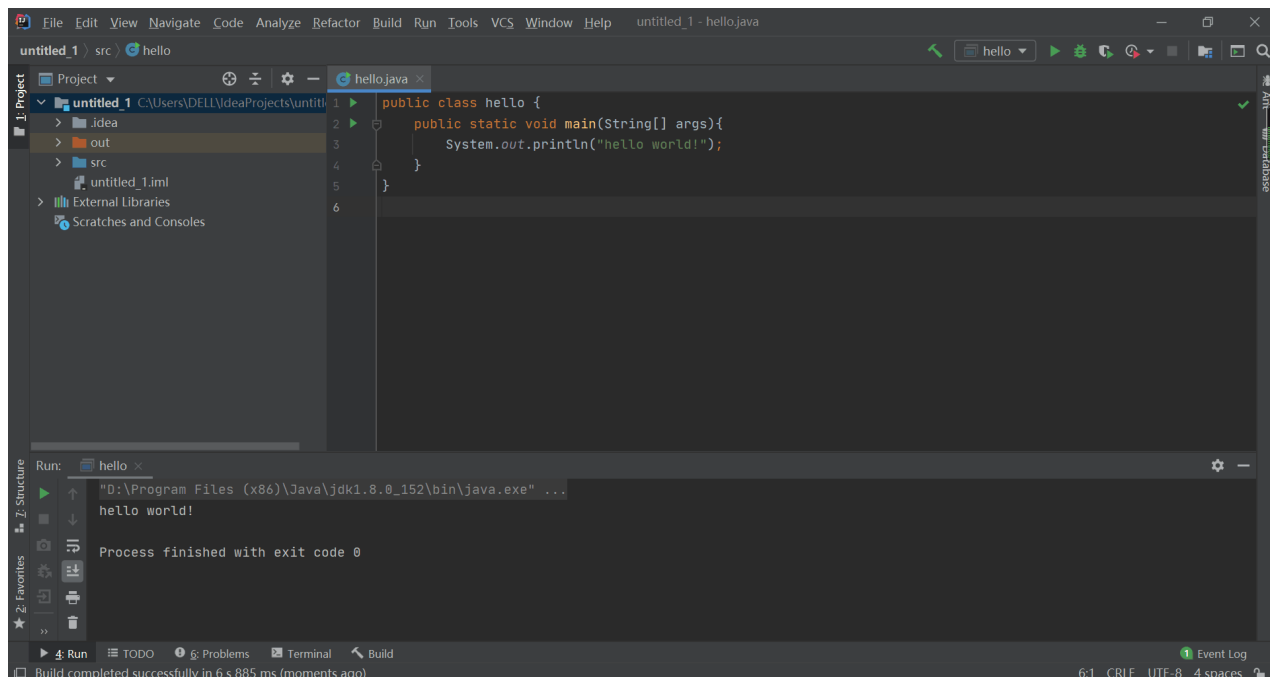
## ■ maven

```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [版本 10.0.19042.1288]
(c) Microsoft Corporation. 保留所有权利。

C:\Users\DELL>mvn -version
Apache Maven 3.6.3 (cecedd343002696d0abb50b32b541b8a6ba2883f)
Maven home: D:\apache-maven-3.6.3\bin\apache-maven-3.6.3\bin\..
Java version: 1.8.0_291, vendor: Oracle Corporation, runtime: C:\Program Files\Java\jre1.8.0_291
Default locale: zh_CN, platform encoding: GBK
OS name: "windows 10", version: "10.0", arch: "amd64", family: "windows"
```

## ■ IDEA

试了一个简单的可以



## 二. 运行情况

由于虚拟机的问题没有解决，并且问了有遇到该问题的同学说是需要重新配置虚拟机，最后来不及了，因此也就没有生成jar文件，而是在win上跑了。

通过参考文章(21条消息) [官网MapReduce实例代码详细批注\\_既认准这条路，又何必在意要走多久-CSDN博客](#)，大致看懂了mapreduce的思路，不过这个只能实现单文件的，由于实在不会java，多文件的处理就只能参考同学的代码，思路主要是：map的时候把key变成文件名、符号、单词，reduce的时候再根据特殊符号分开来，最后跑出如下结果，具体见Wordcount的输出文件夹。

电脑 > 本地磁盘 (C:) > 用户 > DELL > IdeaProjects > Wordcount > single-file-output

搜索"single-file-output"

名称	修改日期	类型	大小
shakespearealls11-r-00000	2021/10/30 23:02	文件	2 KB
shakespeareantony23-r-00000	2021/10/30 23:02	文件	2 KB
shakespeareas12-r-00000	2021/10/30 23:02	文件	2 KB
shakespearecomedy7-r-00000	2021/10/30 23:02	文件	2 KB
shakespearecoriolanus24-r-00000	2021/10/30 23:02	文件	2 KB
shakespearecymbeline17-r-00000	2021/10/30 23:02	文件	2 KB
shakespearefirst51-r-00000	2021/10/30 23:02	文件	2 KB
shakespearehamlet25-r-00000	2021/10/30 23:02	文件	2 KB
shakespearejulius26-r-00000	2021/10/30 23:02	文件	2 KB
shakespeareking45-r-00000	2021/10/30 23:02	文件	2 KB
shakespearelife54-r-00000	2021/10/30 23:02	文件	2 KB
shakespearelife55-r-00000	2021/10/30 23:02	文件	2 KB
shakespearelife56-r-00000	2021/10/30 23:02	文件	2 KB
shakespearelovers62-r-00000	2021/10/30 23:02	文件	2 KB
shakespeareloves8-r-00000	2021/10/30 23:02	文件	2 KB
shakespearemacbeth46-r-00000	2021/10/30 23:02	文件	2 KB
shakespearemeasure13-r-00000	2021/10/30 23:02	文件	2 KB
shakespearemerchant5-r-00000	2021/10/30 23:02	文件	2 KB
shakespearemerry15-r-00000	2021/10/30 23:02	文件	2 KB
shakespearemidsummer16-r-00000	2021/10/30 23:02	文件	2 KB

例如，用记事本打开后一个后，输出格式如下：

shakespearealls11-r-00000 - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

1:lord,211  
2:parolles,177  
3:bertram,136  
4:helena,125  
5:king,124  
6:lafeu,117  
7:shall,115  
8:countess,99  
9:sir,97  
10:thou,95  
11:good,90  
12:thy,84  
13:know,83  
14:thee,79  
15:second,70  
16:clown,67  
17:say,61  
18:love,60  
19:diana,58  
20:tis,52  
21:ill,51  
22:hath,51  
23:make,49  
24:soldier,49  
25:enter,48  
26:come,47  
27:great,46  
28:let,45  
29:madam,45  
30:sneak,41

90° 0.68/s CPU 60°C

第 1 行, 第 1 列 100% Unix (LF) UTF-8

三. 由于上次虚拟机有问题没来得及跑，这次重新装了跑了一下jar

All Applications

+

← → ↺

localhost:8088/cluster

☆

🔍

☰

Cluster

About

Node Labels

Applications

NEW

NEW\_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used
4	0	0	4	0	0 B

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[MEMORY]	<memory:2048, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State
<a href="#">application_1636009744412_0004</a>	severus	sort file	MAPREDUCE	default	0	Thu Nov 4 01:57:30 -0700 2021	Thu Nov 4 01:58:59 -0700 2021	FINISHED
<a href="#">application_1636009744412_0003</a>	severus	word count	MAPREDUCE	default	0	Thu Nov 4 01:51:02 -0700 2021	Thu Nov 4 01:57:25 -0700 2021	FINISHED
<a href="#">application_1636009744412_0002</a>	severus	sort file	MAPREDUCE	default	0	Thu Nov 4 00:40:53 -0700 2021	Thu Nov 4 00:41:36 -0700 2021	FINISHED
<a href="#">application_1636009744412_0001</a>	severus	invertedIndex	MAPREDUCE	default	0	Thu Nov 4 00:25:10 -0700 2021	Thu Nov 4 00:40:49 -0700 2021	FINISHED

```

Activities  Terminal  Thu 02:04
severus@ubuntu: ~/hadoop_installs/hadoop-2.9.1

File Edit View Search Terminal Help

Job Counters
  Failed reduce tasks=4
  Launched map tasks=1
  Launched reduce tasks=4
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=9949
  Total time spent by all reduces in occupied slots (ms)=49645
  Total time spent by all map tasks (ms)=9949
  Total time spent by all reduce tasks (ms)=49645
  Total vcore-milliseconds taken by all map tasks=9949
  Total vcore-milliseconds taken by all reduce tasks=49645
  Total megabyte-milliseconds taken by all map tasks=20375552
  Total megabyte-milliseconds taken by all reduce tasks=101672960
Map-Reduce Framework
  Map input records=208827
  Map output records=208827
  Map output bytes=7785318
  Map output materialized bytes=8202978
  Input split bytes=125
  Combine input records=0
  Spilled Records=208827
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=209
  CPU time spent (ms)=2400
  Physical memory (bytes) snapshot=251179008
  Virtual memory (bytes) snapshot=1904762880
  Total committed heap usage (bytes)=153620480
File Input Format Counters
  Bytes Read=7377958
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ sbin/stop-dfs.sh
Stopping namenodes on [localhost]
localhost: stopping namenode
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
0.0.0.0: stopping secondarynamenode
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$ sbin/stop-yarn.sh
stopping yarn daemons
localhost: stopping nodemanager
localhost: nodemanager did not stop gracefully after 5 seconds: killing with kill -9
no proxyserver to stop
severus@ubuntu:~/hadoop_installs/hadoop-2.9.1$

```