

不积跬步，无以至千里。不积小流无，以成江海。

storm (5)

在界面选择离线安装即可:

1/13

linux (3)

IDE/Eclipse... (2)

scala (8)

python (3)

文章存档

2016年11月 (3)

2016年10月 (18)

2016年9月 (19)

2016年8月 (22)

2016年7月 (28)

展开

阅读排行

Ubuntu16.04下Java环境安装... (57742)

Java imageIO处理图像 (50996)

hibernate uniqueResult方法 (27462)

Java关键字abstract (26206)

Spring加载多个配置文件 (21524)

mybatis学习笔记 (八) Sqlse... (20999)

Spring基于HibernateTransac... (13504)

Java常用设计模式详解 (一) ... (13022)

Java程序数据库连接,数据源... (12808)

JAVA实现JNDI连接数据库 (11316)

最新评论

web.xml加载顺序和配置详解  
Yuwen\_forJava : 允许我转载吗?

spark学习笔记 (4) Intel...  
风湿水声 : 谢谢楼主

HDFS学习笔记 (5) IO读写操作...  
lijunnigedihu : 大牛您好,看到您的帖子,有个问题想请问一下,HIVE开启压缩后就没有办法控制压缩文件的大小,不知道...

Spring加载多个配置文件  
qq\_33460739 : 请问有spring 配置文件吗

spring学习笔记 (一) 事务的管...  
mayfla : (^\_^)不错嘛

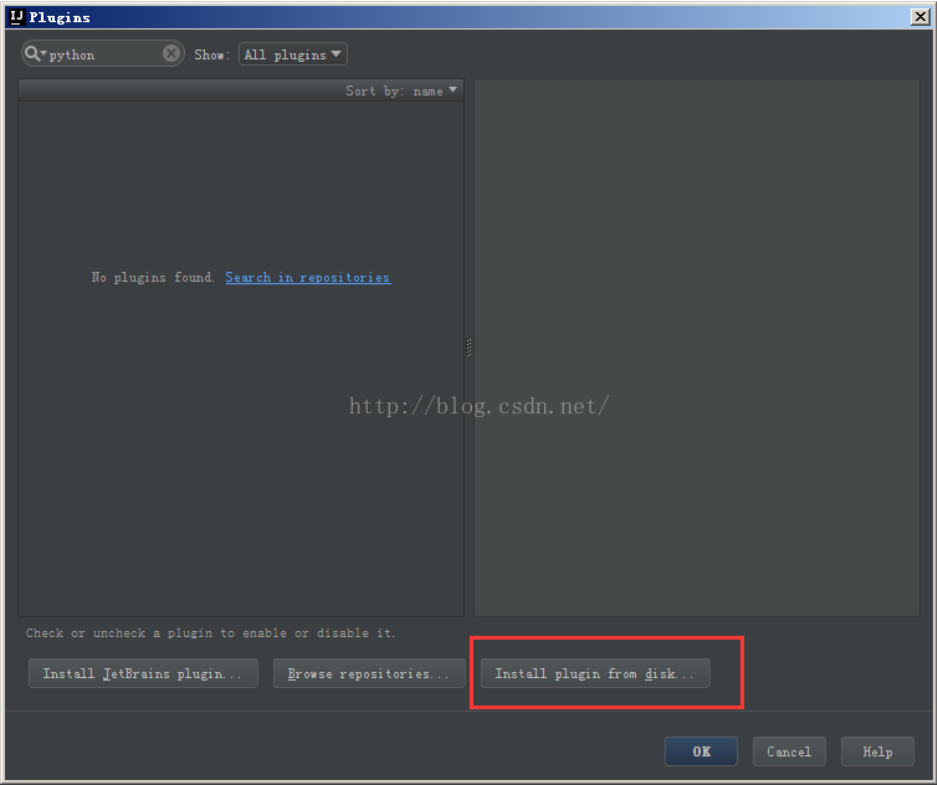
Java内部类模拟实现多继承  
maxcion : [reply]yizishou[/reply] 因为java本身就不支持多继承,java编程思想里...

spark学习笔记 (4) Intel...  
江北才子 : 写的非常详细,结合自己的版本稍作修改,配置成功

Java内部类模拟实现多继承  
yizishou : PencilWithEraser既不能转换为Pencil也不能转换为Eraser,说白了还是一个完...

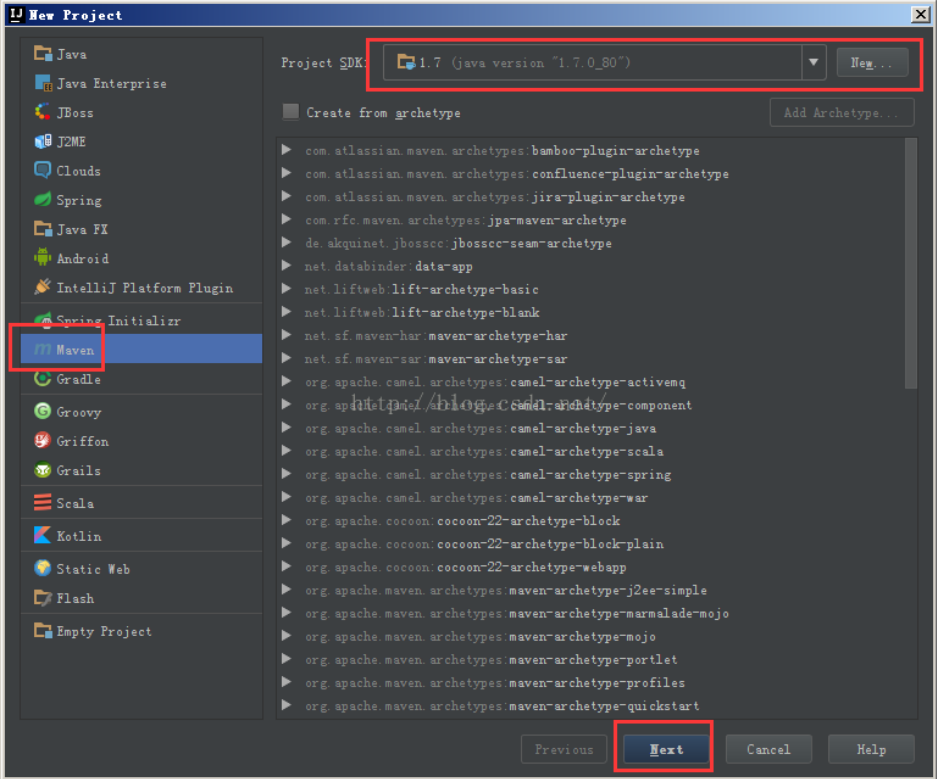
web.xml加载顺序和配置详解  
冬月二十一 : 学习了....

spark学习笔记 (4) Intel...  
JeremyLeeY : 楼主你好,我想问一下,就是intelliJ开发spark1.x版本的时候,就是直接把spark中的...



3、创建Maven工程, File->New Project->Maven

选择相应的JDK版本, 直接下一步



设定Maven项目的GroupId及ArtifactId



联系我们



请扫描二维码联系客服

✉ webmaster@csdn.net

☎ 400-660-0108

💬 QQ客服    💬 客服论坛

关于    招聘    广告服务    百度

©1999-2018 CSDN版权所有

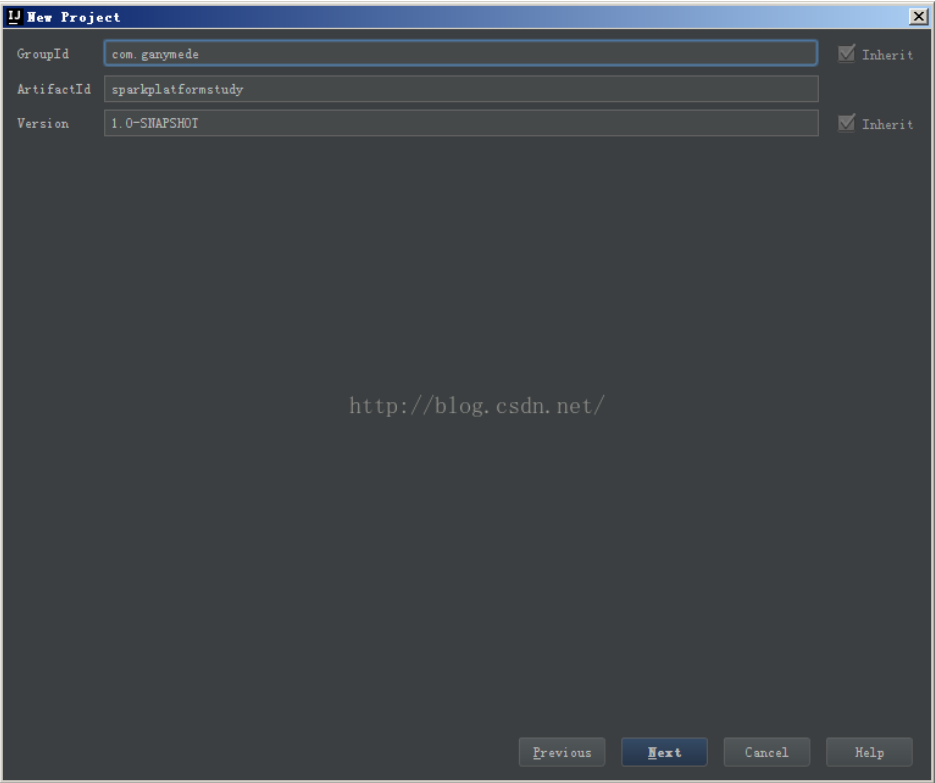
京ICP证09002463号

经营性网站备案信息

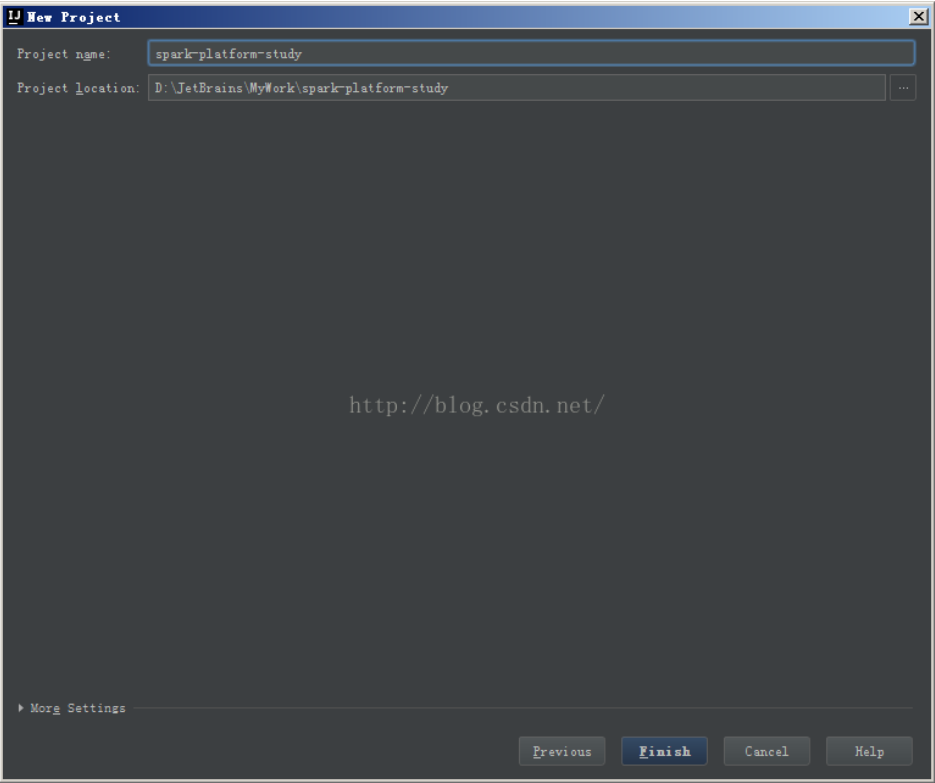
网络110报警服务

中国互联网举报中心

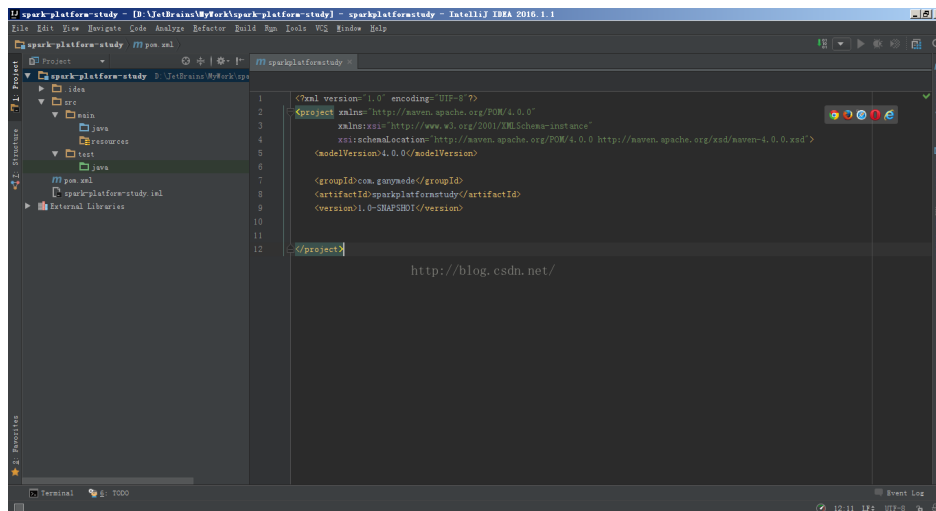
北京互联网违法和不良信息举报中心



创建项目的工程名称，点击完成即可



创建Maven工程完毕，默认是Java的，没关系后面我们再添加scala与spark的依赖



#### 4、修改Maven项目的pom.xml文件，增加scala与spark的依赖

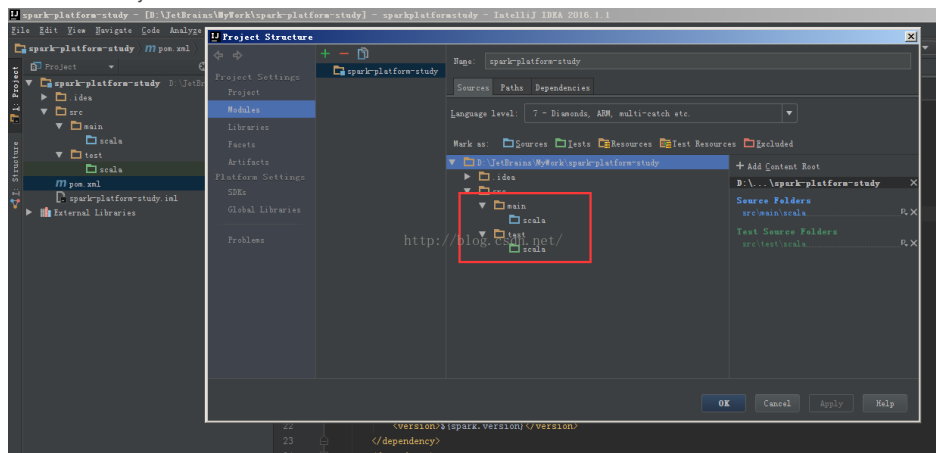
```
[java]
01. <?xml version="1.0" encoding="UTF-8"?>
02. <project xmlns="http://maven.apache.org/POM/4.0.0"
03.         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
04.         xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/xsd/
05.         <modelVersion>4.0.0</modelVersion>
06.
07.         <groupId>com.ganymede</groupId>
08.         <artifactId>sparkplatformstudy</artifactId>
09.         <version>1.0-SNAPSHOT</version>
10.
11.         <properties>
12.             <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>
13.             <spark.version>1.6.0</spark.version>
14.             <scala.version>2.10</scala.version>
15.             <hadoop.version>2.6.0</hadoop.version>
16.         </properties>
17.
18.         <dependencies>
19.             <dependency>
20.                 <groupId>org.apache.spark</groupId>
21.                 <artifactId>spark-core_${scala.version}</artifactId>
22.                 <version>${spark.version}</version>
23.             </dependency>
24.             <dependency>
25.                 <groupId>org.apache.spark</groupId>
26.                 <artifactId>spark-sql_${scala.version}</artifactId>
27.                 <version>${spark.version}</version>
28.             </dependency>
29.             <dependency>
30.                 <groupId>org.apache.spark</groupId>
31.                 <artifactId>spark-hive_${scala.version}</artifactId>
32.                 <version>${spark.version}</version>
33.             </dependency>
34.             <dependency>
35.                 <groupId>org.apache.spark</groupId>
36.                 <artifactId>spark-streaming_${scala.version}</artifactId>
37.                 <version>${spark.version}</version>
38.             </dependency>
39.             <dependency>
40.                 <groupId>org.apache.hadoop</groupId>
41.                 <artifactId>hadoop-client</artifactId>
42.                 <version>2.6.0</version>
43.             </dependency>
44.             <dependency>
45.                 <groupId>org.apache.spark</groupId>
46.                 <artifactId>spark-streaming-kafka_${scala.version}</artifactId>
47.                 <version>${spark.version}</version>
48.             </dependency>
49.             <dependency>
50.                 <groupId>org.apache.spark</groupId>
51.                 <artifactId>spark-mllib_${scala.version}</artifactId>
52.                 <version>${spark.version}</version>
53.             </dependency>
```

```

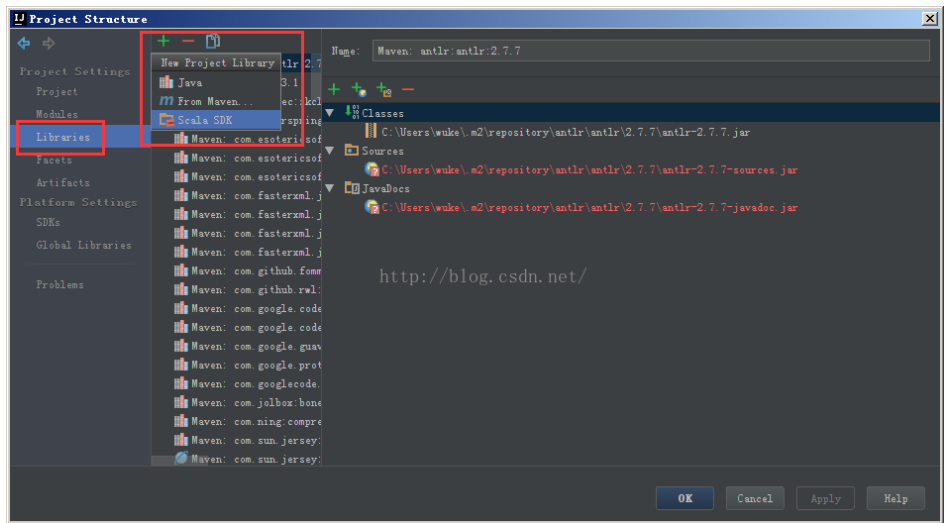
54.         <dependency>
55.             <groupId>mysql</groupId>
56.             <artifactId>mysql-connector-java</artifactId>
57.             <version>5.1.39</version>
58.         </dependency>
59.         <dependency>
60.             <groupId>junit</groupId>
61.             <artifactId>junit</artifactId>
62.             <version>4.12</version>
63.         </dependency>
64.     </dependencies>
65.
66.     <!-- maven官方 http://repo1.maven.org/maven2/ 或 http://repo2.maven.org/maven2/ （延迟低） -->
67.     <repositories>
68.         <repository>
69.             <id>central</id>
70.             <name>Maven Repository Switchboard</name>
71.             <layout>default</layout>
72.             <url>http://repo2.maven.org/maven2</url>
73.             <snapshots>
74.                 <enabled>false</enabled>
75.             </snapshots>
76.         </repository>
77.     </repositories>
78.
79.     <build>
80.         <sourceDirectory>src/main/scala</sourceDirectory>
81.         <testSourceDirectory>src/test/scala</testSourceDirectory>
82.
83.         <plugins>
84.             <plugin>
85.                 <!-- MAVEN 编译使用的JDK版本 -->
86.                 <groupId>org.apache.maven.plugins</groupId>
87.                 <artifactId>maven-compiler-plugin</artifactId>
88.                 <version>3.3</version>
89.                 <configuration>
90.                     <source>1.7</source>
91.                     <target>1.7</target>
92.                     <encoding>UTF-8</encoding>
93.                 </configuration>
94.             </plugin>
95.         </plugins>
96.     </build>
97. </project>

```

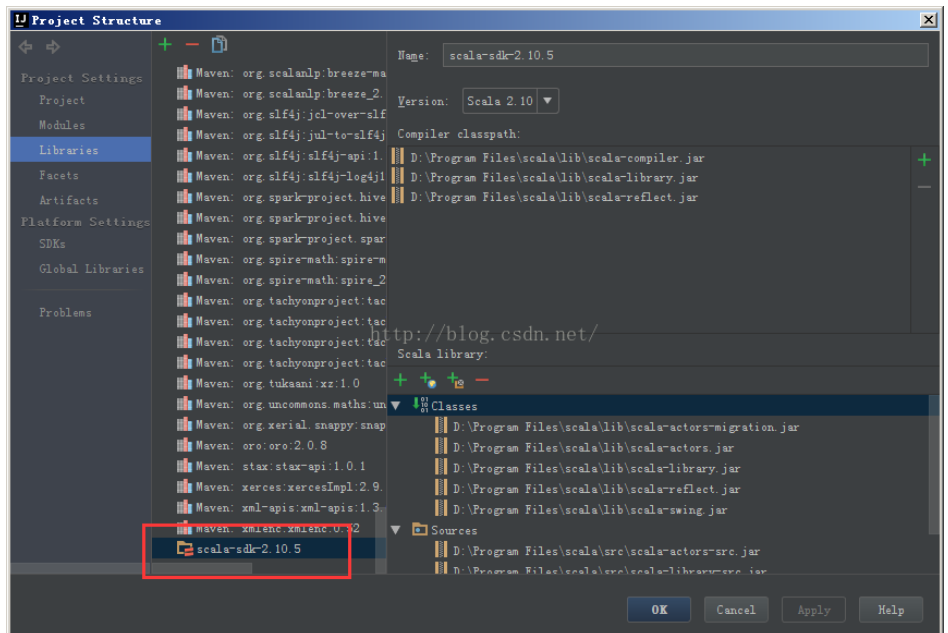
##### 5、删除项目的java目录，新建scala并设置源文件夹（没搞懂设置源文件夹什么意思）



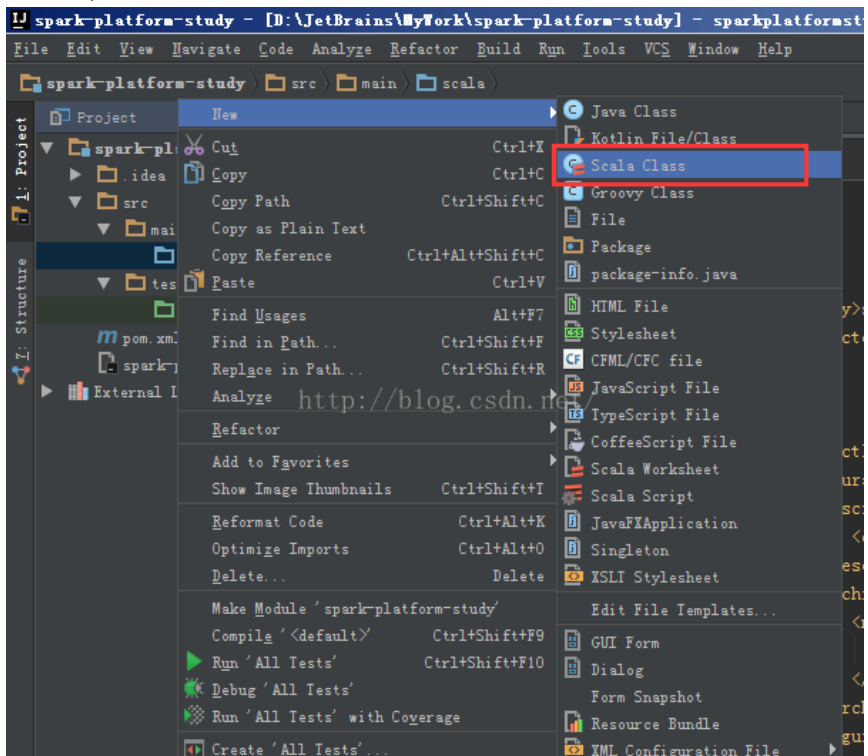
添加scala的SDK（这个地方莫名的花了我很长时间，主要download scala sdk时出的问题）



添加scala的SDK成功（要出现系列红框标记）



## 6、开发Spark实例

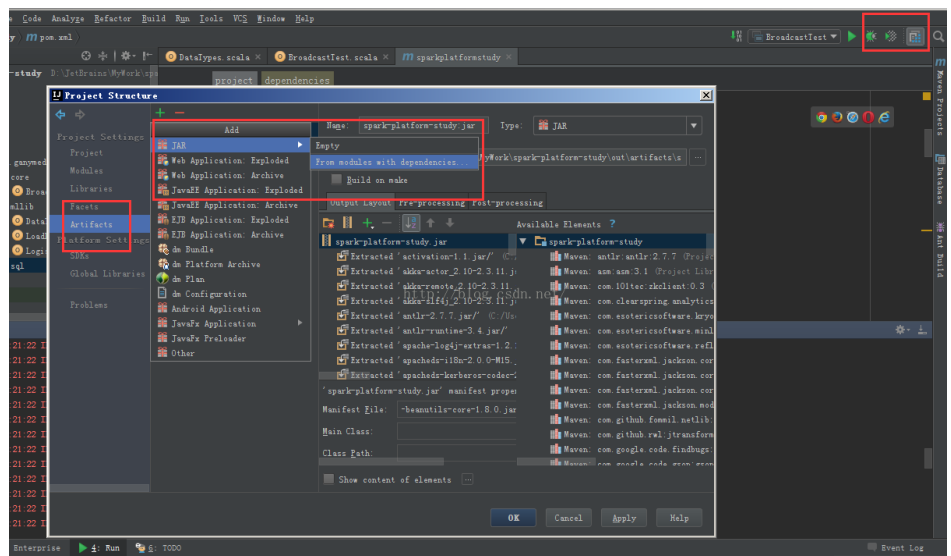


测试案例来自spark官网的mllib例子 <http://spark.apache.org/docs/latest/mllib-data-types.html>

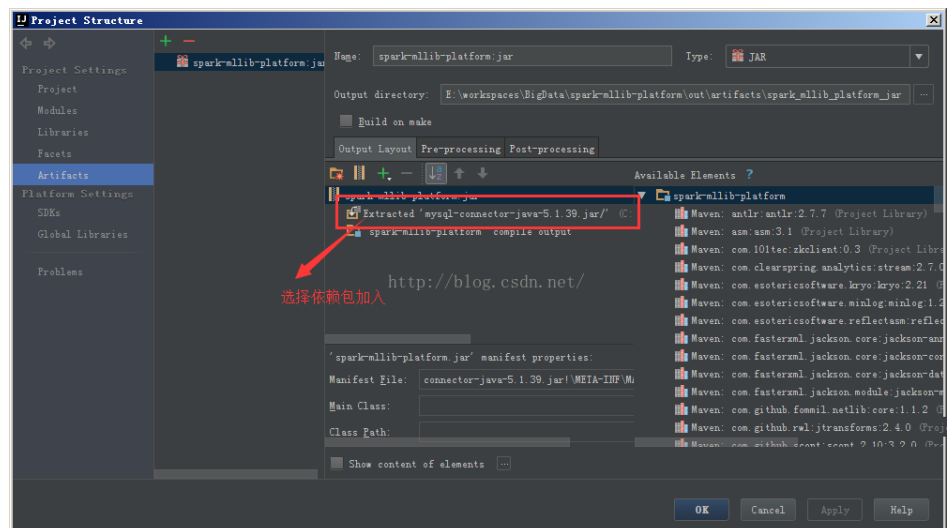
```
[java]
01. import org.apache.spark.{SparkConf, SparkContext}
02.
03. /**
04.  * Created by wuke on 2016/7/5.
05.  */
06. object LoadLibSVMFile extends App{
07.     import org.apache.spark.mllib.regression.LabeledPoint
08.     import org.apache.spark.mllib.util.MLUtils
09.     import org.apache.spark.rdd.RDD
10.
11.     val conf = new SparkConf().setAppName("LogisticRegressionMail").setMaster("local")
12.
13.     val sc = new SparkContext(conf)
14.     val examples: RDD[LabeledPoint] = MLUtils.loadLibSVMFile(sc, "data/mllib/sample_libsvm_data.txt")
15.
16.     println(examples.first)
17. }
```

测试通过

## 7、打包编译，线上发布



注意选择依赖包



项目搭建完成后，运行报错：Failed to locate the winutils binary in the hadoop binary path  
java.io.IOException: Could not locate executable null \bin\winutils.exe in the Hadoop binaries。大概的意思就是说找不到hadoop的环境变量，因此有查了一通，最后解决办法是这样的：

1、下载hadoop包，我下载的hadoop-2.7.3.tar.gz，解压后放到c盘

- 2、下载spark包，我下载的是spark-2.0.0-bin-hadoop2.7（这里指明了hadoop的版本），解压放到C盘；
- 3、下载hadoop在windows下运行的插件包，我已经分享到云盘上，<http://pan.baidu.com/s/1kU5q0Ub>，将所有内容拷贝到hadoop下的bin目录下
- 4、配置hadoop和spark环境变量，主要是home和path。
- 5、重启（应该可以不用重启）

下面是截图配置完成后的截图信息：

hadoop-2.7.3	2016/9/29 18:07	文件夹
JetBrains	2016/9/29 11:32	文件夹
mongodb	2016/7/26 8:58	文件夹
mongodb_1	2016/7/26 10:42	文件夹
mongodb_2	2016/7/26 14:45	文件夹
mongodb_3	2016/8/15 15:09	文件夹
mysql-5.7.12-winx64	2016/4/19 14:49	文件夹
my-workspace	2016/8/2 15:37	文件夹
ncworkspace	2016/9/28 13:57	文件夹
Program Files	2016/9/28 11:40	文件夹
redis-3.0.5	2016/8/8 17:59	文件夹
spark-2.0.0-bin-hadoop2.7	2016/9/29 17:36	文件夹

环境变量：

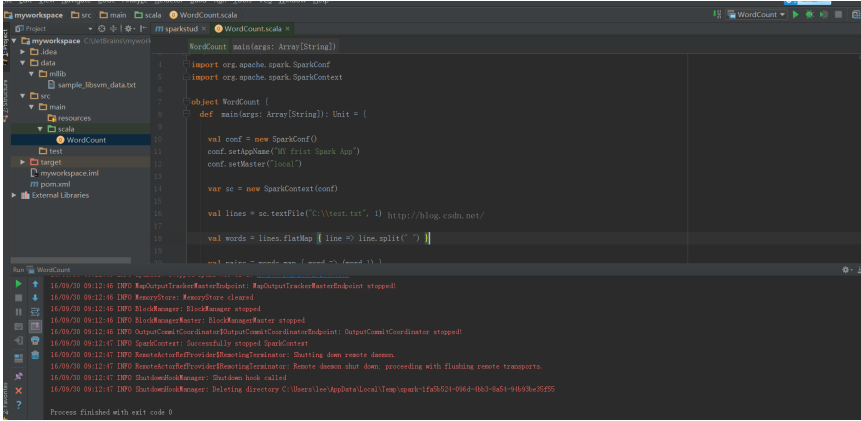
HADOOP_HOME	C:\hadoop-2.7.3
JAVA_HOME	C:\Program Files\Java\jdk1.8.0_101
SPARK_HOME	C:\spark-2.0.0-bin-hadoop2.7
C:\spark-2.0.0-bin-hadoop2.7\bin;C:\hadoop-2.7.3\bin	

windows下插件

Windows7_OS (C:) > hadoop-2.7.3 > bin		
共享 新建文件夹		
名称		修改日期
container-executor		2016/8/18 9:49
hadoop		2016/8/18 9:49
hadoop.cmd		2016/8/18 9:49
hadoop.dll		2014/2/18 14:48
hadoop.exp		2014/2/18 14:48
hadoop.lib		2014/2/18 14:48
hadoop.pdb		2014/2/18 14:48
hdfs		2016/8/18 9:49
hdfs.cmd		2016/8/18 9:49
libwinutils.lib		2014/2/18 12:16
mapred		2016/8/18 9:49
mapred.cmd		2016/8/18 9:49
rcc		2016/8/18 9:49
test-container-executor		2016/8/18 9:49
winutils.exe		2014/2/18 12:16
winutils.pdb		2014/2/18 12:16
yarn		2016/8/18 9:49
yarn.cmd		2016/8/18 9:49

再次运行：





到此windows下spark开发环境搭建基本ok。  
感谢一下博主提供的宝贵经验  
<http://www.cnblogs.com/davidwang456/p/5032766.html>  
<http://www.cnblogs.com/eczhou/p/5216918.html>

- [上一篇](#) IntelliJ Idea
- [下一篇](#) spark学习笔记（5）WordCount for Spark



您还没有登录,请[登录](#)或[注册](#)

查看评论



风湿水声

3楼 2018-01-09 11:32发表

谢谢楼主



江北才子

2楼 2017-08-02 09:29发表

写的非常详细，结合自己的版本稍作修改，配置成功



JeremyLeeY

1楼 2017-02-15 10:52发表

楼主你好，我想问一下，就是intellij开发spark1.x版本的时候，就是直接把spark中的assmbly.jar这个包导入就能开发，但是spark2.0版本之后这个jar包就找不到了，我想问能不能用之前的导入Jar包的方法来开发scala？

IntelliJ Idea搭建Spark开发环境

在Spark快速入门指南 – Spark安装与基础使用 u012877472 2016年03月30日 14:24 14554  
中介介绍了Spark的安装与配置，在那里还介绍了使用spark-submit提交应用，不过不能使用vim来开发Spark应用，放着IDE的方便不用。...

spark配置IntelliJ开发环境详解



bitcarmanlee 2016年07月27日 20:06 7402

1.花一天半时间配置spark开发环境RD同学开发代码肯定需要开发环境。为了配置好spark的本地开发测试环境，宝宝前后花了一天半的时间。各种踩坑各种无奈各种崩溃。文章以下内容都是宝宝一天半时间的实践...

## Windows下用IDEA进行Spark开发

当集群搭建好了,接下来就是将自己的代码写好,扔到集群上进行跑了。安装软件1、JDK  
2、IntelliJ IDEA 3、xshell这三部安装过程这里不介绍,下一步下一步即可。4、IntelliJ...

## Spark开发环境配置(windows/IntelliJ IDEA 篇)

IntelliJ IDEA是一个蛮不错的IDE,在java/scala/Groovy领域深得人心。笔者之前使用的是Eclipse那一套开发环境,虽然也不错,但忍不住好奇心的驱使,折腾了一下IDEA,发...

## IDEA SPARK SCALA 搭建简单的helloworld

新建maven工程: 在第一次使用idea时,如下图所示选择新建工程 Paste\_Image.png 如下图,选择maven工程, sdk通过new, 选择安装的jdk, 然后点击next: Pa...

## 码农不会英语怎么行? 英语文档都看不懂!

软件工程出身的英语老师,教你用数学公式读懂天下英文→



## IntelliJ IDEA使用Maven搭建spark开发环境 ( scala )

如何一步一步地在IntelliJ IDEA使用Maven搭建spark开发环境,并基于scala编写简单的spark中wordcount实例。...

**用IntelliJ IDEA提交pyspark程序** FishSeeker 2017年04月15日 23:04 1784  
在idea上安装Python组件并利用idea写pyspark的程序,然后关联pyspark的源码。

## spark编程python实例

spark编程python实例ValueError: Cannot run multiple SparkContexts at once; existing SparkContext (app=PySp...


## Apache Spark MLlib学习笔记 ( 四 ) IntelliJ上部署Spark源码

从这篇开始,就正式开始进行源码分析,源码分析是一件工作量很大的工作,尤其是spark这样的大工程,因此一定要使用编辑器才能较好的理顺各个类的关系。看别人经验,都推荐使用intelliJ编辑器。在此也要...

## Spark ( 八 ) -- 使用IntelliJ Idea搭建Spark开发环境

IntelliJ Idea下载地址: 官方下载 选择右下角的Community Edition版本下载安装即可  
本文中用的是windows系统 环境为: jdk1.6.0\_45 scala2...

## spark开发环境搭建 ( 基于idea 和maven )

使用idea构建maven 管理的spark项目，默认  212272215 2016年11月22日 22:26 0 12222222  
已经装好了idea 和Scala,mac安装Scala 那么使用idea 新建maven 管理的spark 项目有以下几步: scala插件...

## 免费云主机试用一年

云服务器免费试用


百度广告

## Spark2.x学习笔记：6、在Windows平台下搭建Spark开发环境（Intell...


第6章 在Windows平台下搭建Spark开发环境（IntelliJ IDEA+Maven）6.1 集成开发环境IDE为了方便应用程序开发与测试，提高开发效率，一般使用集成开发工具IDE。同样，为了...

 chengyuqiang 2017年09月10日 21:15 0 2363


## 如何使用IntelliJ IDEA搭建spark开发环境（上）

本文转自<http://www.beanmoon.com/2014/10/11/%E5%A6%82%E4%BD%95%E4%BD%B0/11/%E5%A6%82%E4%BD%95%E4%BD%B0/11/%E5%A6%82%E4%BD%95%E4%BD%B0/>  sdujava2011 2015年07月08日 11:22 0 1974  
F%E7%94%A8intellij%E6%90%AD%E5%BB%B...

## IntelliJ搭建spark开发环境

 pirage 2015年12月08日 11:18 0 17248  
spark怎么学习呢？在一无所知的前提下，首先去官网快速了解一下spark是干什么的，官网在此。然后，安装开发环境，从wordcount开始学习。第三，上手以后可以学习其他算法了。最后，不要放弃，继续...

## IntelliJ IDEA配置Spark使用

 haluoluo211 2016年10月08日 18:25 0 1842  
第二步下载插件 首先启动intelliJ IDEA:在命令行终端中，进入\$IDEA\_HOME/bin目录，输入./idea.sh进行启动,进入如下界面，然后选择右下角“plugins” ...

## Spark学习: Spark-Scala-IntelliJ开发环境搭建和编译Jar包流程

spark-scala使用IntelliJ搭建的开发环境。  databatman 2016年08月29日 22:55 0 6181


## Spark Streaming+IntelliJ Idea+Maven开发环境搭建

国内关于Spark流处理方面的资料实在是少之又少，开发环境搭建上一些细节上的说明就更少了，  
本文主要介绍在Windows下通过IntelliJ Idea连接远程服务器的Spark节点，接收FlumeNG...

## 使用PyCharm配置Spark的Python开发环境（基础）

在本地搭建好Spark 1.6.0后，除了使用spark-submit提交Python程序外，我们可以使用PyCharm这个IDE在本地进行开发调试,提升我们的开发效率。配置过程也十分简单，在stac...

## IntelliJ Idea开发spark程序及运行

版本：spark-1.6.01.利用创建一个SBT项目参  fz2543122681 2016年08月25日 11:48 2905  
考步骤：<http://blog.csdn.net/jameshadoo>  
[p/article/details/522992502](http://blog.csdn.net/jameshadoo/p/article/details/522992502).编写简单代码pac...

## Windows 上面搭建 Spark + Pycharm/idea scala/python 本地编写S...

Windows 上面搭建 Spark + Pycharm/idea scala/python 本地编写Spark程序，测试通过后再提交到Linux集群上...


 CLHugh 2017年07月06日 18:36 702

## 短信验证码接口

短信验证码是什么

百度广告


## 使用IDEA进行Spark开发（二）-第一个scala程序

上面一篇文章博主已经给大家演示好了如何去  qq\_31794773 2016年12月01日 21:57 6252  
配置一个本机的scala开发环境，现在我们就一起  
去写我们的第一个spark开发的scala程序吧！打开IDEA，选择创建一个新的工程文件。点击scala，创...

## 基于IntelliJ IDEA开发Spark的Maven项目——Scala语言

基于IntelliJ IDEA开发Spark的Maven项目  kwu\_ganymede 2016年07月05日 16:41 15259  
——Scala语言 Maven管理项目在JavaEE  
普遍使用，开发Spark项目也不例外，而Scala语言开发Spark项目的首选。因此需要...


## Spark+IntelliJ IDEA创建项目

 oMrApollo 2017年03月24日 09:48 1435  
创建scala项目1.选择scala:2.scalaSdk版本选择，我选择的是如下图：3.新建Module，选择scala，Simple module with attached Scala SDK。...

## 001-使用IDEA 工具创建Spark项目 shenfuli

2016年05月29日 23:20 2151  
1. 环境准备 1 2. 创建第一个SBT项目 1 2.1. File->Project 1 2.2. SBT项目工程结构 4 2.3. SBT项目配置依赖 4 3. Spark编写WordCount...

## idea创建Spark Maven项目

 gao634209276 2016年08月16日 22:30 956  
在Spark快速入门指南 – Spark安装与基础使用中介绍了Spark的安装与配置，在那里还介绍了使用spark-submit提交应用，不过不能使用vim来开发Spark应用，放着IDE的方便不用。...

## 5分钟完成加壳，防止代码反编译

Virbox Protector加壳工具，堪比VMP安全强度的加壳工具



## Spark学习（二）---kafka+SparkStreaming的搭建与连接

kafka+Sparkstreaming环境搭建与配置说明以及  ASIA\_kobe 2015年11月19日 21:18 8298  
相关的测试代码的编写

## windows本地sparkstreaming开发环境搭建及简单实例

windows本地spark开发环境搭建及简单实例  u012551524 2017年05月28日 23:00  1077

1：开发环境IDEA选择：IntelliJ IDEA Community Edition 2017.1.1 （相比eclipse更方便...

---

## Spark教程（2）Spark Streaming 介绍

问题导读 1、如何利用DStream所提供的api，而在数据流上实时进行操作？ 2、你如何理解Spark Streaming启动后的流程？ 3、如何创建StreamingContext对象？ ...  javastart 2016年06月18日 23:00  661

问题导读 1、如何利用DStream所提供的api，而在数据流上实时进行操作？ 2、你如何理解Spark Streaming启动后的流程？ 3、如何创建StreamingContext对象？ ...

---

## Spark-Spark Streaming例子整理(一)

（摘自王家林）流（Streaming），在大数据时代为数据流处理，就像水流一样，是数据流；既然是数据流处理，就会想到数据的流入、数据的加工、数据的流出。日常工作、生活中数据来源很多不同...

日常工作、生活中数据来源很多不同...

---

## Spark 开发环境搭建（2）IDEA 安装新版本IntelliJ IDEA 2017.1

Spark 开发环境搭建（2）IDEA 安装新版本IntelliJ IDEA 2017.11，卸载IDEA 老版本IntelliJ IDEA 2016.3 2，安装新版本IDEA 2017.1下载地址 ...

 duan\_zhihua 2017年04月23日 18:15  1429