# Spark 开发环境配置手册（windows）

Spark的开发可以通过Intellij或者Eclipse IDE进行，在环境配置的开始阶段，还需要安装相应的Scala插件。本次我们用Intellij IDEA。

## 相关下载安装

### 1、下载地址和版本

|  |  |  |
| --- | --- | --- |
| **名称** | **下载地址** | **版本** |
| Spark | http://spark.apache.org/downloads.html | 1.6 |
| Jdk | http://www.oracle.com/technetwork/java/javase/archive-139210.html | 1.7.67 |
| Scala | http://www.scala-lang.org/download/all.html | 2.10.5 |
| Intellij IDEA | https://www.jetbrains.com/idea/download/#section=windows | 当前版本 |

### 2、解压安装

（1）解压下载后的scala和spark到相应的位置（本次解压到D:\工具）。

（2）安装JDK和Intellij IDEA（直接点击下一步即可完成安装）。

## 配置系统环境变量

在系统环境变量里添加如下参数：

**1、JAVA\_HOME= C:\Java\jdk1.7.0\_67**

**说明**：“ C:\Java\jdk1.7.0\_67”是我本地Jdk的安装路径。

**2、SPARK\_HOME= D:\工具\spark-1.6.0-bin-hadoop2.6\spark-1.6.0-bin-hadoop2.6**

**说明**：“D:\工具\spark-1.6.0-bin-hadoop2.6\spark-1.6.0-bin-hadoop2.6”是我本地Spark解压的路径。

**3、SCALA\_HOME= D:\工具\scala-2.10.5\scala-2.10.5**

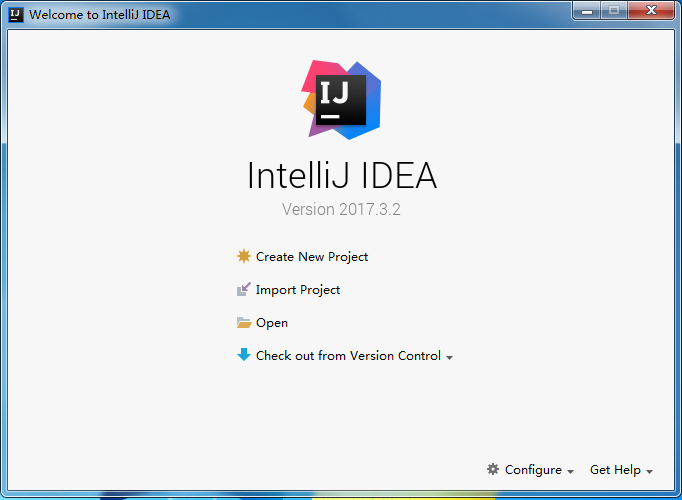
**说明**：“D:\工具\scala-2.10.5\scala-2.10.5”是我本地Scala解压路径。

最后在path变量添加：“;%JAVA\_HOME%\bin;%SPARK\_HOME%\bin;%SCALA\_HOME%\bin”

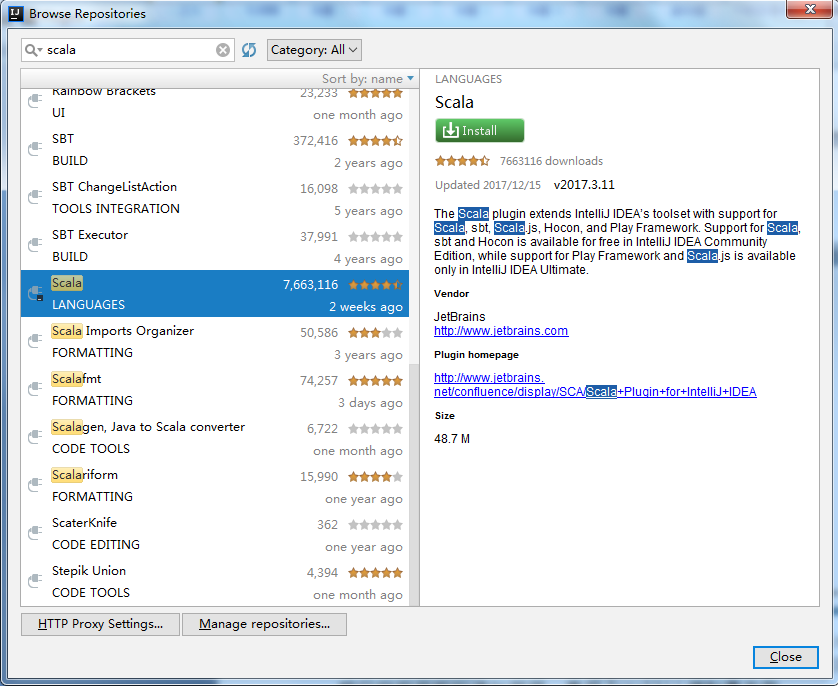
## Intellij IDEA相关设置

### 1、Scala插件配置

在“开始”-“程序”里打开我们的IDEA，一路默认进入到如下界面：



选择“Configure”—>“Plugins”—>“Browse repositories”命令，在弹出的界面中输入“Scala”搜索插件，如下图，然后点击相应的安装按钮进行安装，重启Intellij使配置生效。



### 2、配置Spark应用开发环境

（1）在Intellij IDEA中创建Scala Project,名称为SparkTest。

（2）选择菜单中的“File”—>“project structure”—>“Libraries”，然后选择“+”，导入spark-assembly-1.6.0-hadoop2.6.0.jar（在我们之前解压Spark的lib路径下）。

（3）如果IDE无法识别Scala库，需要以同样的方式将Scala库的jar包导入，之后可以进行开发了。可以把一下代码放到IDEA上测试IDEA是否配置成功。

测试代码：

|  |
| --- |
| **import** org.apache.spark.{SparkConf, SparkContext}  **object** SparkPi {   **def** main(args: Array[String]): Unit = {  **val** conf = **new** SparkConf().setAppName(**"Spark pi"**).setMaster(**"local"**)  **val** sc = **new** SparkContext(conf)  **val** slices = **if** (args.length > 0) args(0).toInt **else** 2  **val** n = 100000 \* slices  **val** count = sc.parallelize(1 to n, slices).map { i =>  **val** x = i \* 2 - 1  **val** y = i \* 2 - 1  **if** (x \* x + y \* y < 1) 1 **else** 0  }.reduce(\_ + \_)  *println*(**"pi is roughly "** + 4.0 \* count / n)  sc.stop()  } } |

### 3、运行Spark程序

（1）本地运行

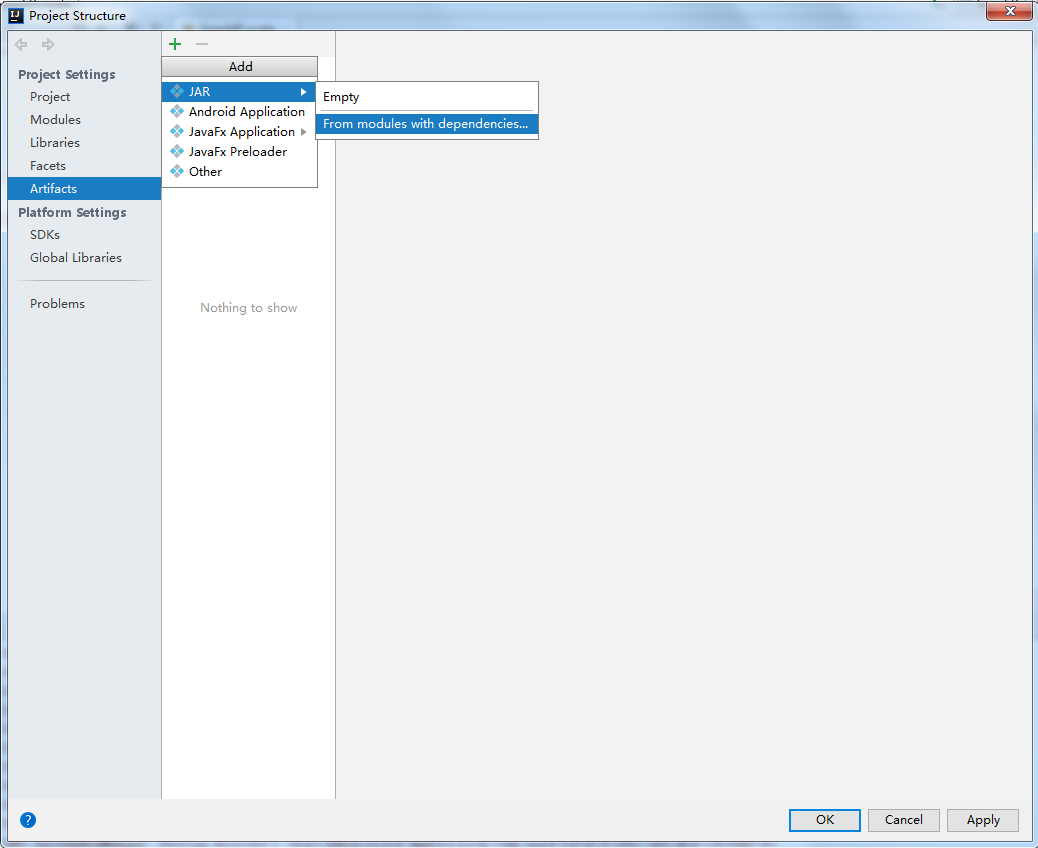
编写完Scala程序后，可以直接在IDEA中以本地（local）模式运行，方法是设setMaster，如代码标红处。“**new** SparkConf().setAppName(**"Spark pi"**).setMaster(**"local"**) ”。

然后点击菜单中“Run”按钮即可运行。

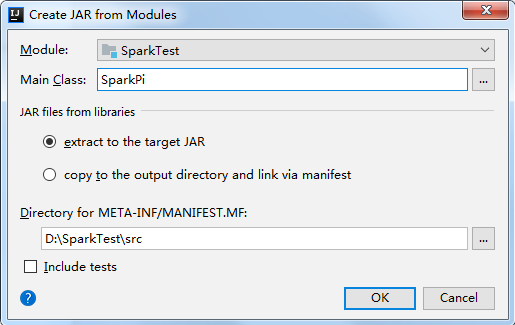
（2）在集群中运行Spark应用程序Jar包

如果想把程序打成Jar包，在Spark集群中运行，可以按照一下步骤操作。

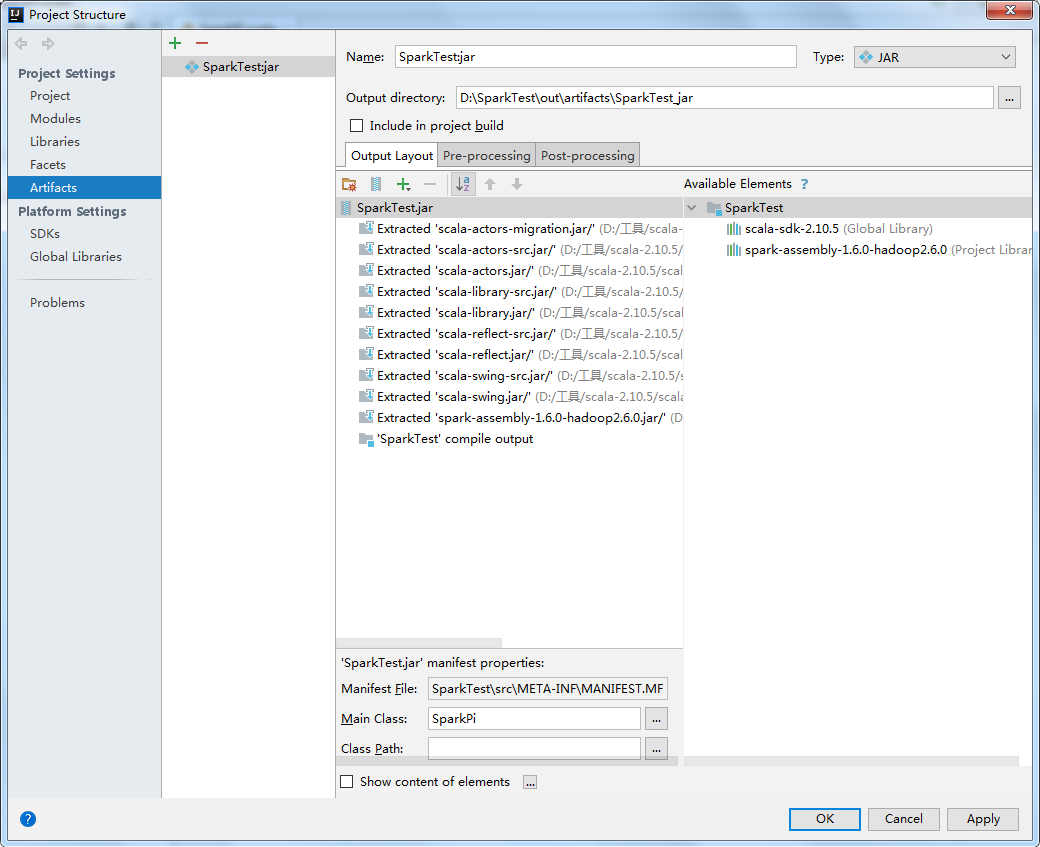
1. 选择“File”—>“Project Structure”命令，然后选择“Artifact”，单击“+”按钮，选择“Jar”—>“From Modules with dependencies”，如下图所示。



点击“OK”按钮，进入如下界面：



选择Main Class，在弹出的对话框中选择输入Jar位置，并单击“OK”按钮。如下图：



在上图所示对话框中通过OutPut layout中的“+”选择依赖的Jar包后，点击“OK”按钮即可。

1. 在主菜单中选择“Build”—>“Build Artifact”命令，编译生成Jar包。
2. 在集群的主节点，通过下面命令执行生成的Jar包SparkTest.jar。