**Idea中Spark开发环境搭建**

目录

[一 Spark版本信息](#_Toc14498_WPSOffice_Level1) [3](#_Toc14498_WPSOffice_Level1)

[二 项目创建步骤](#_Toc4789_WPSOffice_Level1) [3](#_Toc4789_WPSOffice_Level1)

[三 配置依赖](#_Toc24732_WPSOffice_Level1) [4](#_Toc24732_WPSOffice_Level1)

[四 打包设置](#_Toc10178_WPSOffice_Level1) [7](#_Toc10178_WPSOffice_Level1)

**一 Spark版本信息**

Spark 1.6/Spark 2.0以上

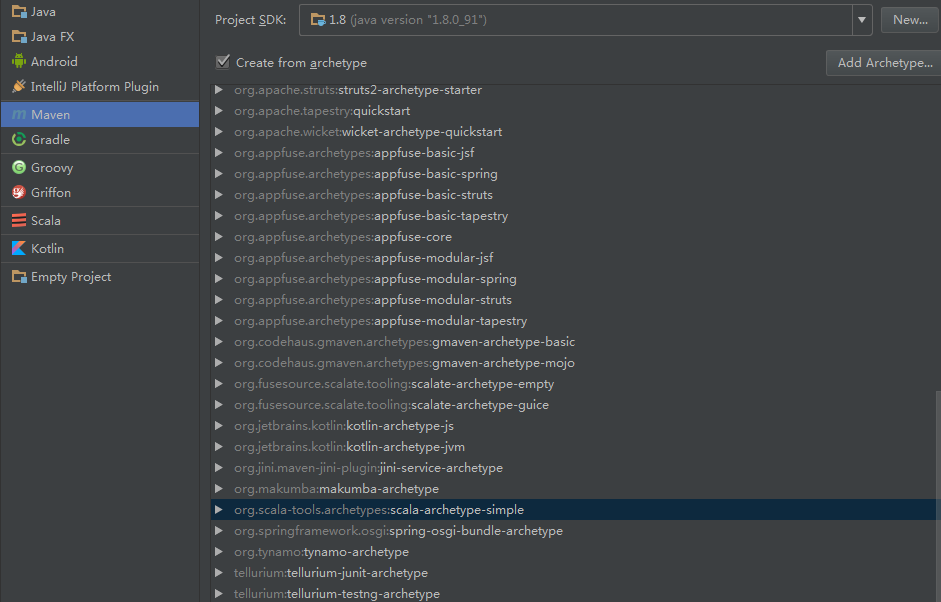
Java 1.8

Scala 2.11.7

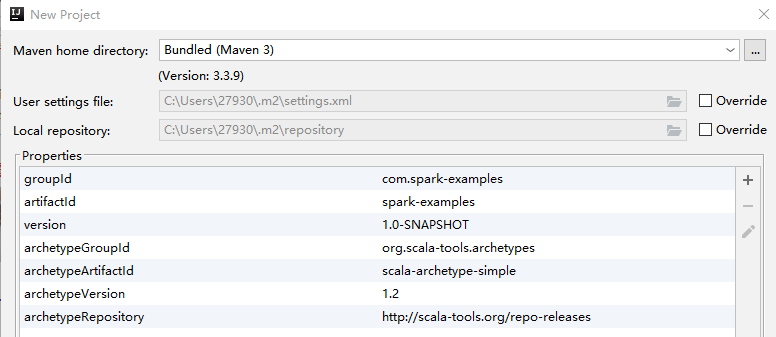
**二 项目创建步骤**

1. 步骤1

创建maven项目

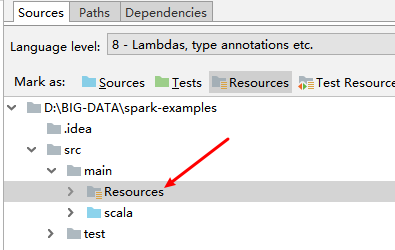


1. 尽量用默认的setting.xml

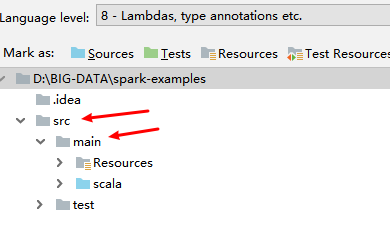


1. 创建的项目没有Resources包，手动在program-structure里面创建

步骤1：右键，在src下面创建resources文件夹，并标记为resources包，如图：



步骤2：取消src，main的sources包标记，只将scala包标记为sources标记。如图示：



**三 配置依赖**

1. **配置pom.xml依赖**

<properties>

<project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>

<spark.version>2.2.0</spark.version>

<scala.version>2.11</scala.version>

<hadoop.version>2.7.3</hadoop.version>

</properties>

<repositories>

<repository>

<id>scala-tools.org</id>

<name>Scala-Tools Maven2 Repository</name>

<url>http://scala-tools.org/repo-releases</url>

</repository>

</repositories>

<pluginRepositories>

<pluginRepository>

<id>scala-tools.org</id>

<name>Scala-Tools Maven2 Repository</name>

<url>http://scala-tools.org/repo-releases</url>

</pluginRepository>

</pluginRepositories>

<dependencies>

<dependency>

<groupId>org.scala-lang</groupId>

<artifactId>scala-library</artifactId>

<version>2.11.7</version>

</dependency>

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-core\_${scala.version}</artifactId>

<version>${spark.version}</version>

</dependency>

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-sql\_${scala.version}</artifactId>

<version>${spark.version}</version>

</dependency>

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-hive\_${scala.version}</artifactId>

<version>${spark.version}</version>

</dependency>

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-streaming\_${scala.version}</artifactId>

<version>${spark.version}</version>

</dependency>

<dependency>

<groupId>org.apache.hadoop</groupId>

<artifactId>hadoop-client</artifactId>

<version>2.7.3</version>

</dependency>

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-streaming-kafka-0-10\_2.11</artifactId>

<version>2.1.0</version>

</dependency>

<dependency>

<groupId>org.apache.spark</groupId>

<artifactId>spark-mllib\_${scala.version}</artifactId>

<version>${spark.version}</version>

</dependency>

<dependency>

<groupId>mysql</groupId>

<artifactId>mysql-connector-java</artifactId>

<version>5.1.39</version>

</dependency>

<dependency>

<groupId>junit</groupId>

<artifactId>junit</artifactId>

<version>4.4</version>

<scope>test</scope>

</dependency>

<dependency>

<groupId>org.specs</groupId>

<artifactId>specs</artifactId>

<version>1.2.5</version>

<scope>test</scope>

</dependency>

<dependency>

<groupId>com.typesafe</groupId>

<artifactId>config</artifactId>

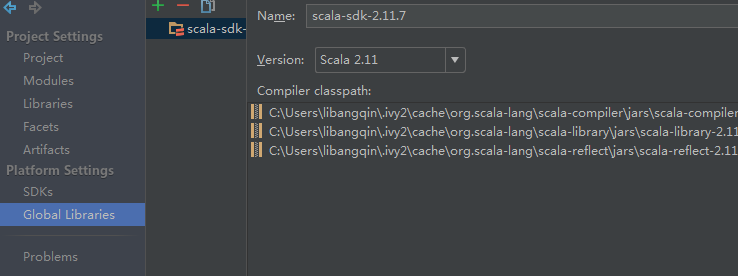
<version>1.3.0</version>

</dependency>

</dependencies>

1. **配置scala包**

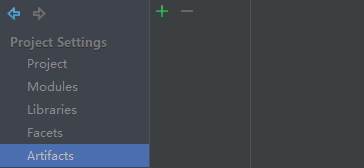
配置好pom.xml依赖之后需要配置下载 scala SDK 本环境所用的scala版本为2.11.7



点击”+”进行添加选择SBT下载 或者手动下载再进行添加。

**四 打包设置**

打包设置 File -> project structure -->选择 Artifacts --> Jars --> From modules with dependence



选择MainClass --> 勾选 build on make 然后删除spark的lib包 scala的sdk jar包添加额外的依赖包 点击完成

运行设置 VM-Option中 设置 -Dspark.master=local

以及sparkConf 设置spark.driver.host为本机IP

常用的配置如下：

driver.host = "10.0.0.159"  
executor.memory = "512m"  
serializer = "org.apache.spark.serializer.KryoSerializer"  
driver.memory = "512m"  
master = "spark://10.0.0.222:7077"  
appName = "CSV\_ComputedParams"