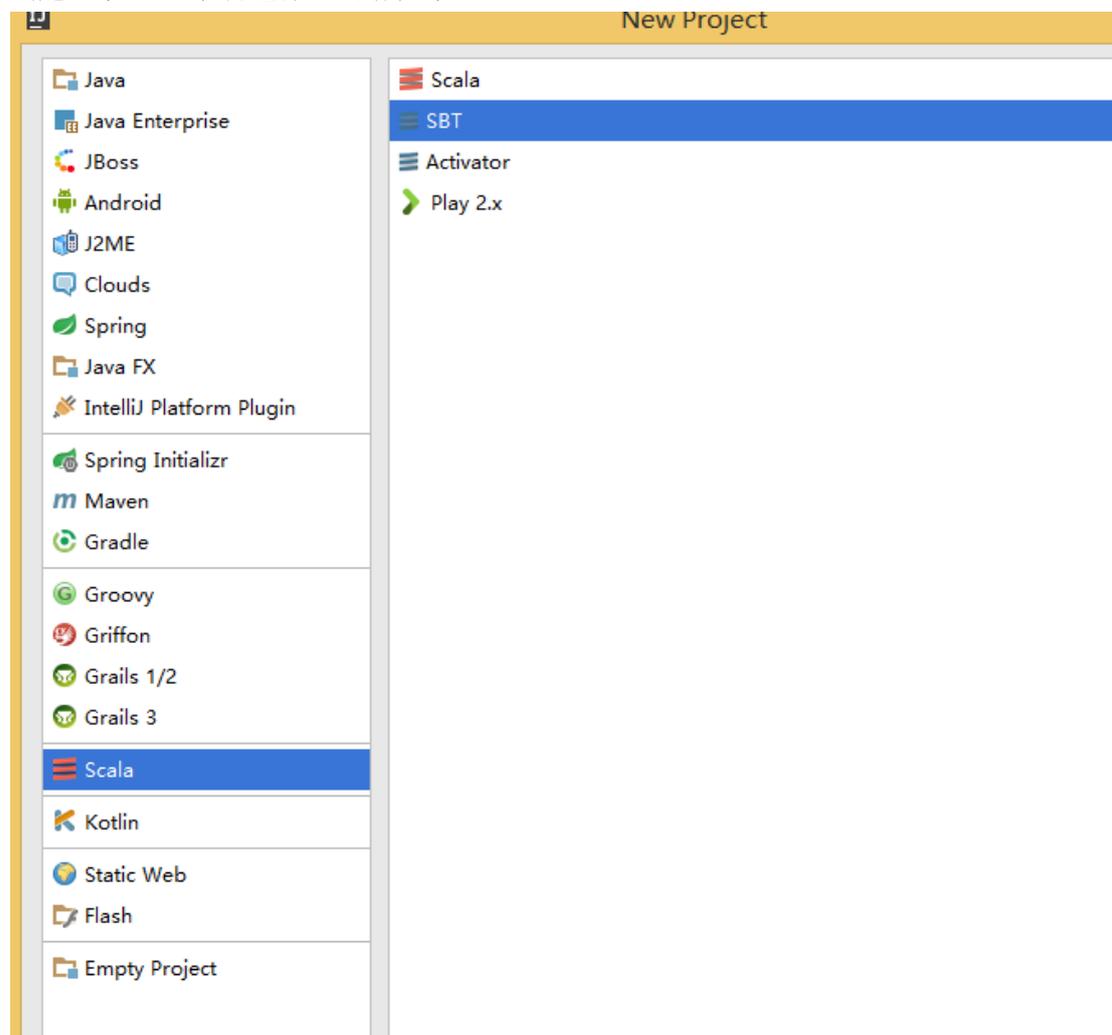


spark 在 windows 上以 local 模式运行

1. 创建项目

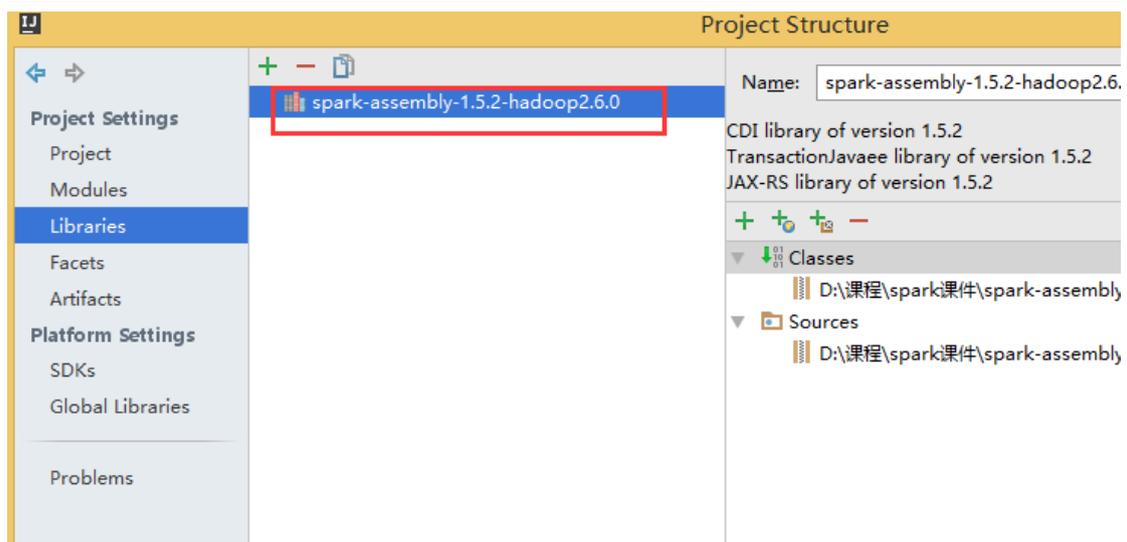
创建一个 scala 项目选择 sbt 的方式



或者 scala 也行
但是要注意选择 scala 的 SDK

2. 添加依赖

添加相应的 spark 依赖，一个 assembly 包含所有了



3. 下载解压 hadoop

下载解压 spark 包对应的 hadoop

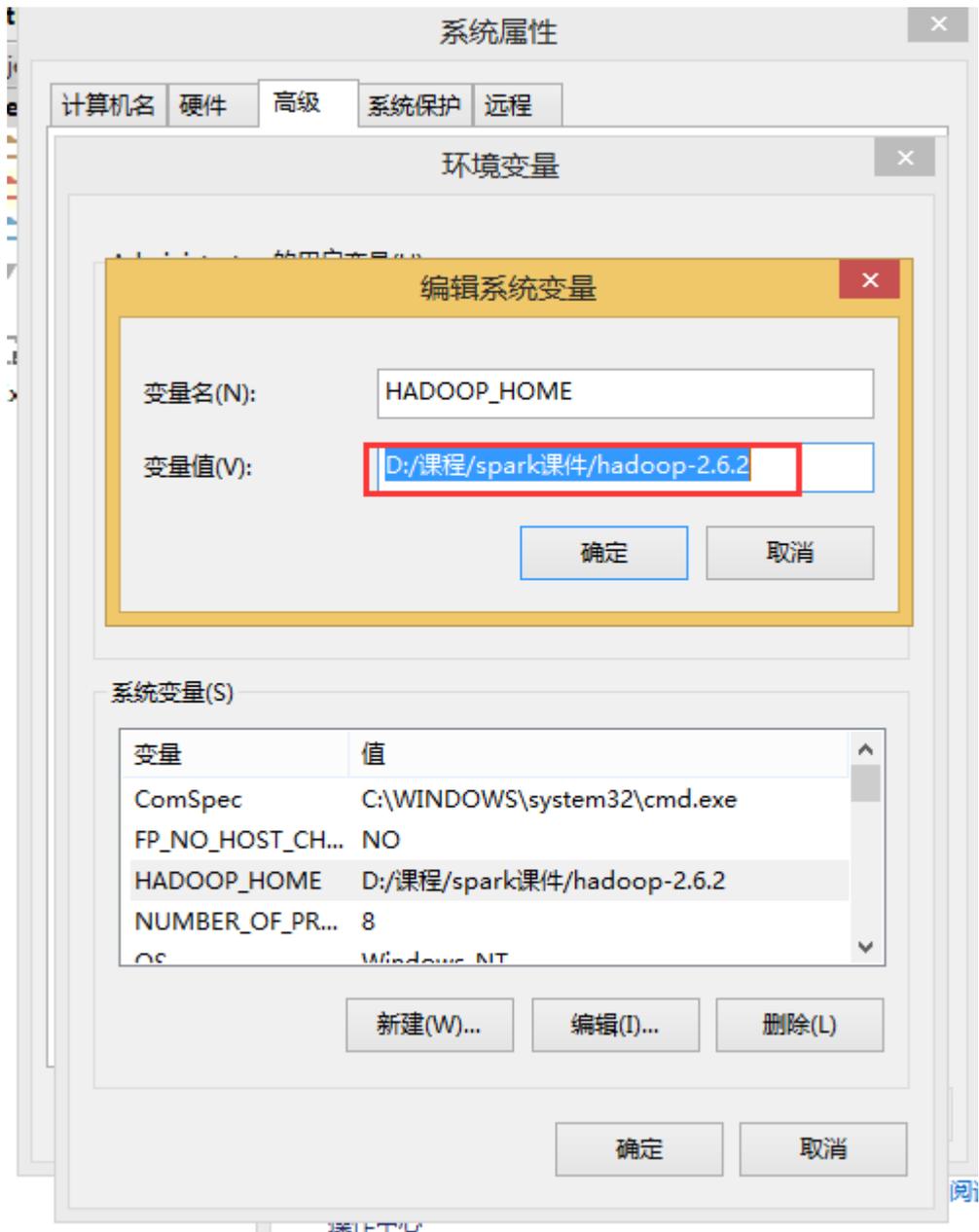
名称	修改日期	类型	大小
hadoop2.6(x64)V0.2	2016/3/5 星期六 ...	文件夹	
hadoop-2.6.2	2016/3/5 星期六 ...	文件夹	
spark-1.5.2-bin-hadoop2.6	2016/3/5 星期六 ...	文件夹	
hadoop2.6(x64)V0.2.zip	2016/3/5 星期六 ...	360压缩 ZIP 文件	698 KB
hadoop-2.6.2.tar.gz	2016/3/5 星期六 ...	360压缩	190,934 KB
spark-1.5.2.tgz	2016/2/29 星期...	360压缩	11,368 KB
spark-1.5.2-bin-hadoop2.6.tgz	2016/2/29 星期...	360压缩	274,324 KB
spark-assembly-1.5.2-hadoop2.6.0.jar	2015/11/4 星期...	Executable Jar File	179,682 KB
spark之dataframe.pptx	2016/3/5 星期六 ...	Microsoft Power...	696 KB

4. 添加 winutils.exe

从 <http://files.cnblogs.com/files/yjmyzz/hadoop2.6%28x64%29V0.2.zip> 下载包，解压将里面东西全部拷贝到 hadoop 目录的 bin 下面

5. 配置环境变量

配置 HADOOP_HOME 的环境变量，然后重启 idea



6. 设置 scala 的 SDK

将 scala 的 SDK 的版本设置为 2.10.4，这里用 2.11 版本的 scala 可能和 spark 不兼容

The screenshot shows the 'Project Structure' dialog in IntelliJ IDEA for a project named 'scala-sdk-2.10.4'. The 'Global Libraries' tab is selected in the left sidebar. The main area shows the project name 'scala-sdk-2.10.4' highlighted with a red box. The right pane displays the following configuration:

- Name:** scala-sdk-2.10.4
- Version:** Scala 2.10
- Compiler classpath:**
 - D:\soft\scala\lib\scala-compiler.jar
 - D:\soft\scala\lib\scala-library.jar
 - D:\soft\scala\lib\scala-reflect.jar
- Scala library:**
 - Classes
 - D:\soft\scala\lib\scala-actors-migra
 - D:\soft\scala\lib\scala-actors.jar
 - D:\soft\scala\lib\scala-library.jar
 - D:\soft\scala\lib\scala-reflect.jar
 - D:\soft\scala\lib\scala-swing.jar
 - JavaDocs
 - www.scala-lang.org\api\2.10.4\