一、 Scala 与 Spark 的安装和配置方法

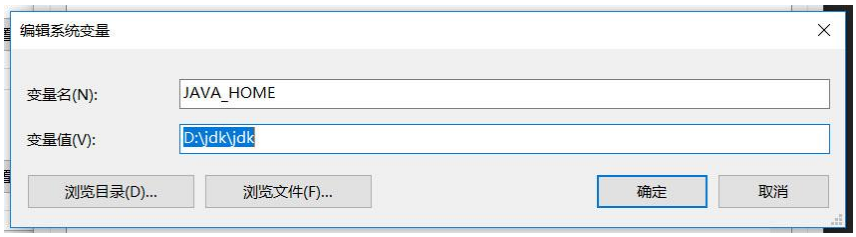
（一）、Scala 与 Spark 的安装

由于 Spark 是基于 Scala 开发的，而 Scala 又是基于 Java 虚拟机的，因而在安装 Spark 之前需要检查是否有匹配版本的 Java 和 Scala

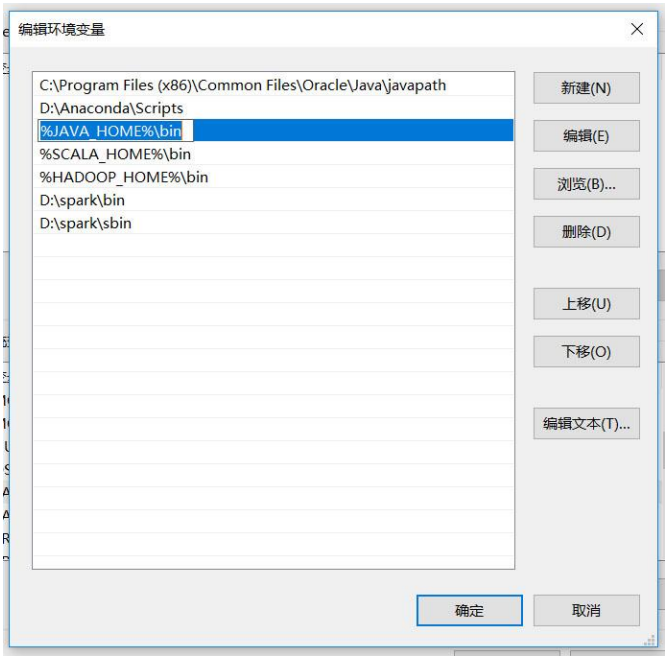
1、Java 环境检查与安装

Spark 目前不支持 Java 1.8 以后的版本，特别是 Java 1.9 与 Java 1.10，并且若系统内安装了多个版本的 Java 会引起 Spark 引用环境的混乱，因此在 Spark 安装之前需要卸载系统内所有的高版本 Java，仅保留或安装 Java 1.8。

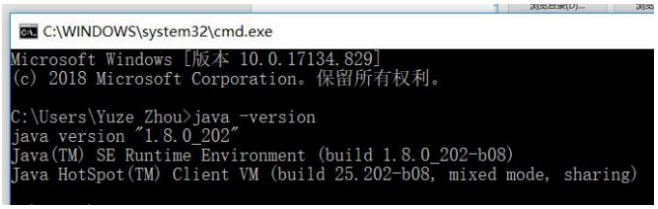
第二步需要配置 Java 相应的环境变量，需要添加新的 JAVA_HOME 变量以及在 PATH 变量中添加相应地路径。JAVA_HOME 的变量值应设为 JAVA 安装目录中包含 bin 文件夹的路径，如下图：



然后需要配置 Java 的路径，如下图所示：

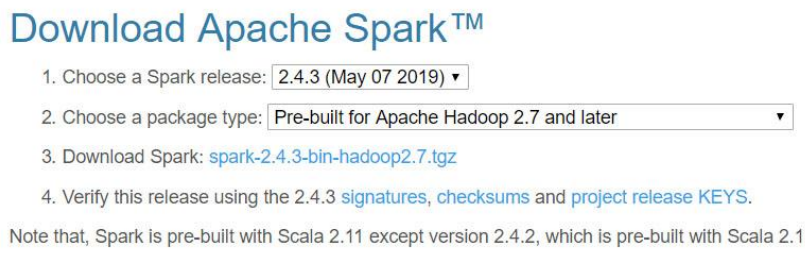


最后我们打开命令行，输入 java -version，如果能正常显示出 Java 的版本号为 1.8，那么要求的 Java 的配置就不会有任何问题，可以继续配置 Scala 了：

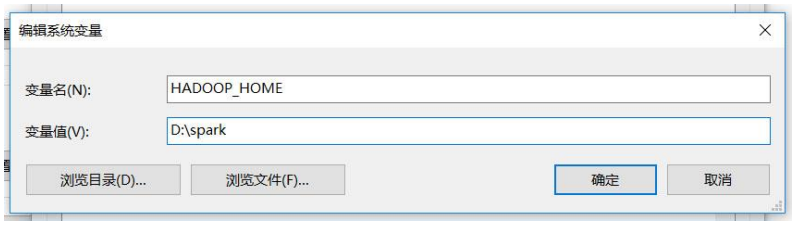


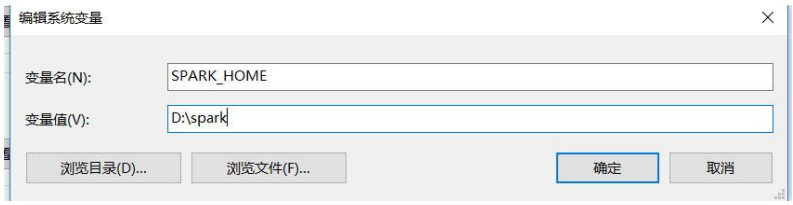
2、Spark 环境检查与安装

在安装之前，首先从 Apache Spark 的官网上下载最新的 Spark 版本（http://spark.apache.org/downloads.html），如下图所示选择 2.4.3 版，Pre-built for Apache Hadoop 2.7 and later。



在下载解压完毕后，需要设置 HADOOP_HOME 与 SPARK_HOME 两个新的环境变量，变量值均设为 Spark 安装目录下包含 bin 和 sbin 文件夹的路径：



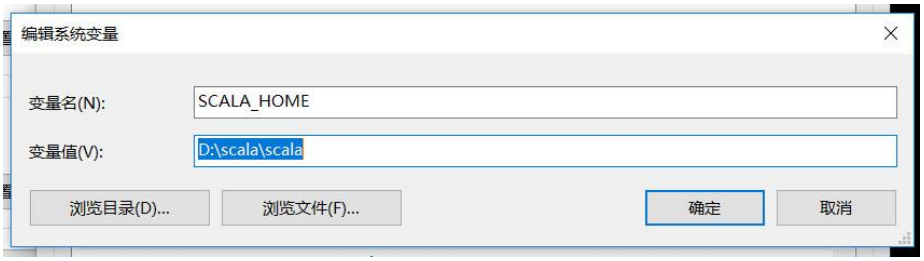


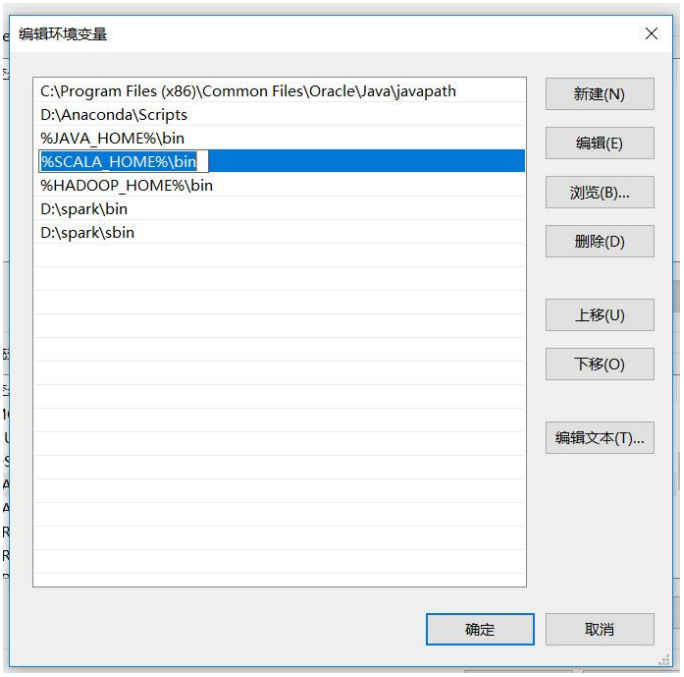
在 Path 变量中安装 spark 时需要修改的比较多，需要添加 %HADOOP_HOME%\bin，%SPARK_HOME%\bin，%SPARK_HOME%\sbin 三个路径至变量中，或者可以直接添加 spark 安装目录下 bin 与 sbin 两个文件夹的路径，同之前 Path 变量的图片中最末尾的三个路径所示。

环境变量配置完成之后打开命令行，若使用的是分布式集群，输入 spark-shell，若仅使用本地资源，输入 spark-shell2。命令行中如果显示有如下所示的图标，那么就表明 Spark 已经成功安装完毕了。

2、Scala 环境检查与安装

在之前安装完 Spark 的命令行中我们可以发现，目前最新的版本是基于 Scala 2.11.12 开发的。因而，为方便以后更加便捷的开发 Spark，则相应的 Scala 版本也需要安装。在完成下载和解压后，在环境变量中设置新变量 SCALA_HOME 为 SCALA 安装目录下包含 bin 文件夹的路径；另外，在 Path 中再额外添加 bin 文件夹的路径。





再次打开命令行，输入 scala -version，如果能正常显示出 scala 版本为 2.11.12，则 Scala 的配置也已经达到要求：

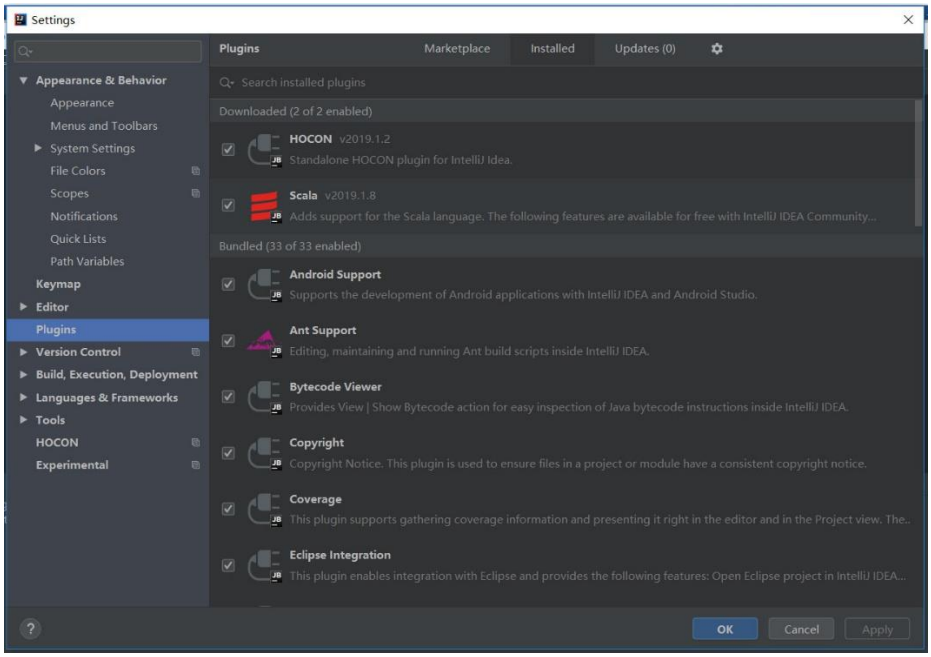
```
C:\Users\Yuze Zhou>scala -version
Scala code runner version 2.11.12 -- Copyright 2002-2017, LAMP/EPFL
```

（二）、Scala 与 Spark 的 IDE 安装

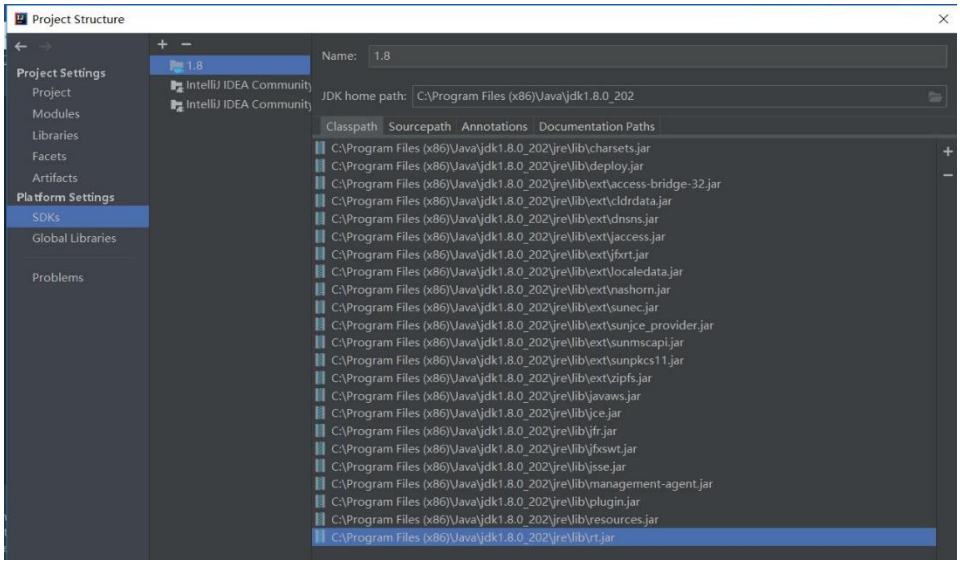
1、IntelliJ Idea 的安装

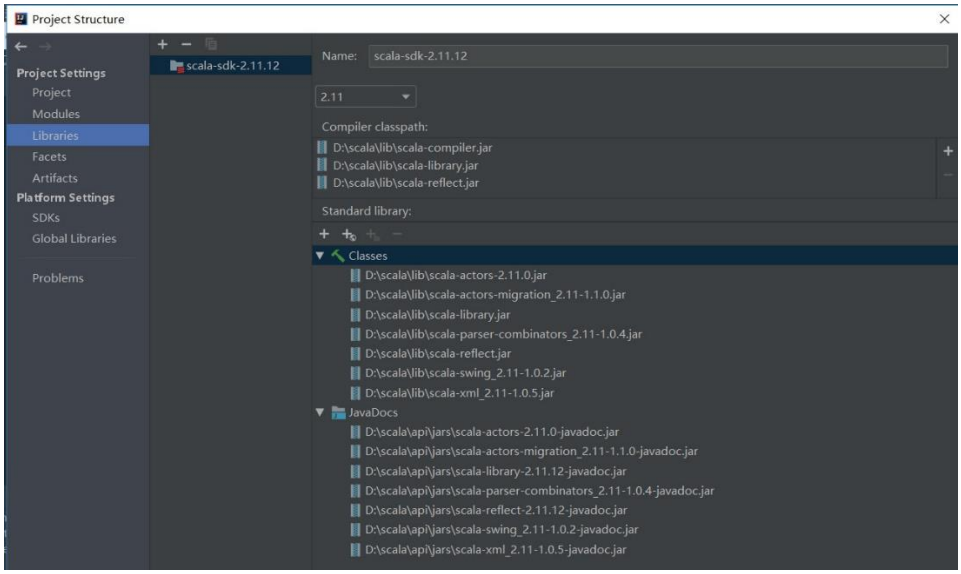
在完成所有软件的安装之后，我们会发现当前只能通过打开 cmd 使用极其不方便的交互式编程来开发 Spark 程序。因而必须要额外配置可以与 Spark 无缝衔接且方便进行开发调试的 Scala IDE。

我选择的是 IntelliJ Idea 作为开发用的 IDE，在配置 Spark 工作环境之前，需要事先确认相应版本的 Java 和 Scala 已经安装完毕并下载解压最新版本的 IntelliJ Idea。打开 IntelliJ Idea 并创建新项目时，第一步先要打开如下目录：**File -> Settings -> Plugins -> Marketplace** 选择并安装 Scala 插件：



安装完 Scala 插件后 Idea 会提醒是否重启，这时选择重启并且在重启后的新项目中选择 **Scala -> sbt**。在新项目页面打开之后，我们需要打开: **File -> Project Structure** 进一步确认 Java 与 Scala 的环境是否正确。其中在 **SDK** 一栏，Classpath 中需包含 java 安装目录下 lib 文件夹下的所有文件。在 Libraries 一栏中需添加好 Scala 安装目录下所有 lib 和 api 文件夹下的所有文件，分别如下图所示：





在确认好环境安装无误后，打开左侧 **Project -> src -> main -> scala** 右键选择新建 Scala Class，然后就可以自由创建 Scala 程序了。