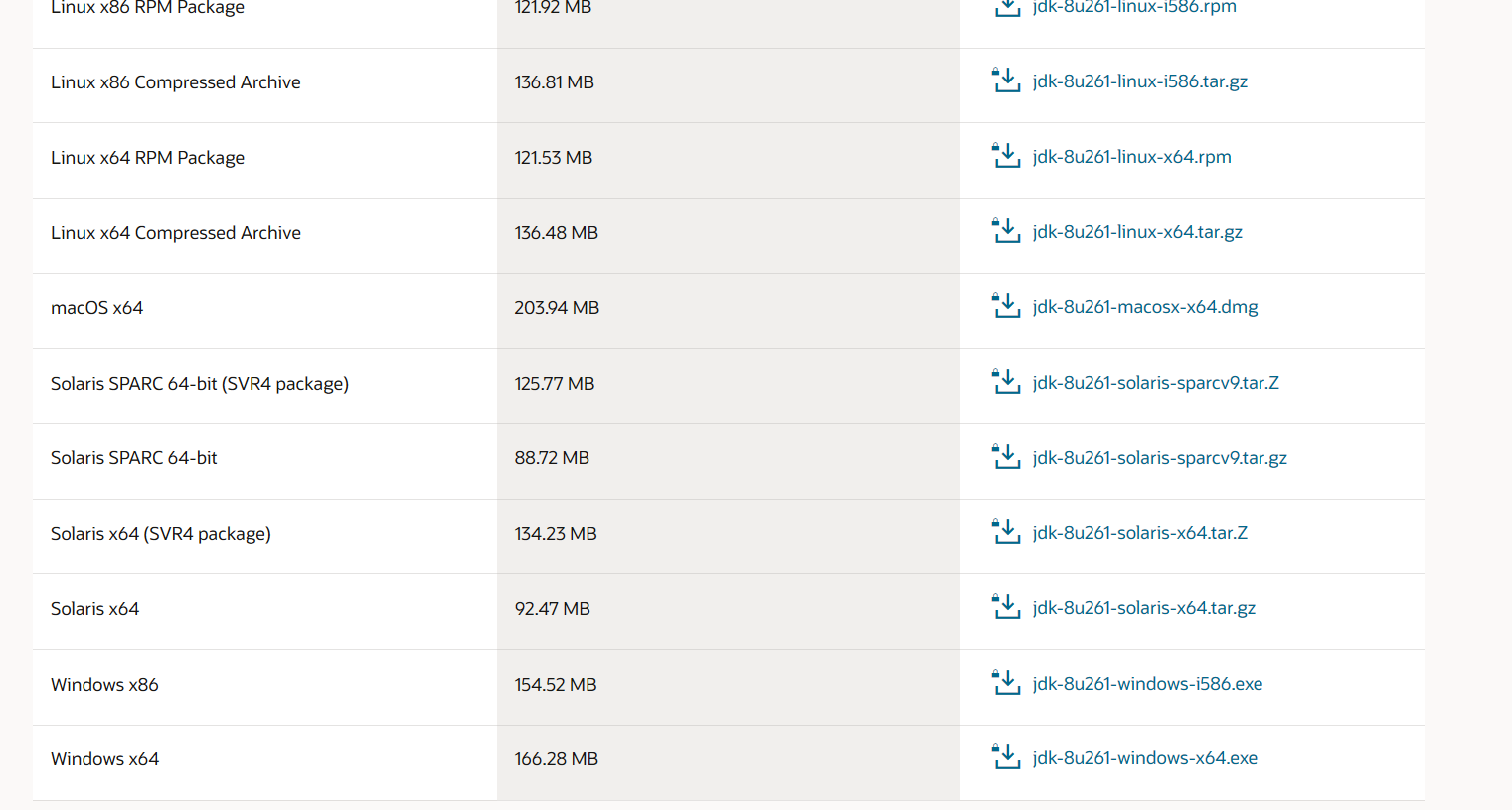
# 实验一、Spark运行与开发环境搭建

由于spark是由scala编写，而scala又是基于jvm环境的变成语言，所有首先要保证你的环境已经安装了JDK8

## 第一步：jdk的安装

百度直接搜索jdk1.8





点击接受，然后可以选择下载，可以根据自己系统进行下载，这里以windows64为例，点击下载（有时会让登录）  
下载完成之后双击exe文件进行安装，下一步，选择安装路径



输入oracle账户登入，进行下载



点击下一步



默认安装位置就可以啦，记住安装路径，后面配环境会用得到



可能有同学会有疑问怎么选择两次安装路径，这里解释一下，第一次选择的是jdk（java开发工具包），第二次选择的是jre（java的运行环境），安装完成，接下来要配置环境变量



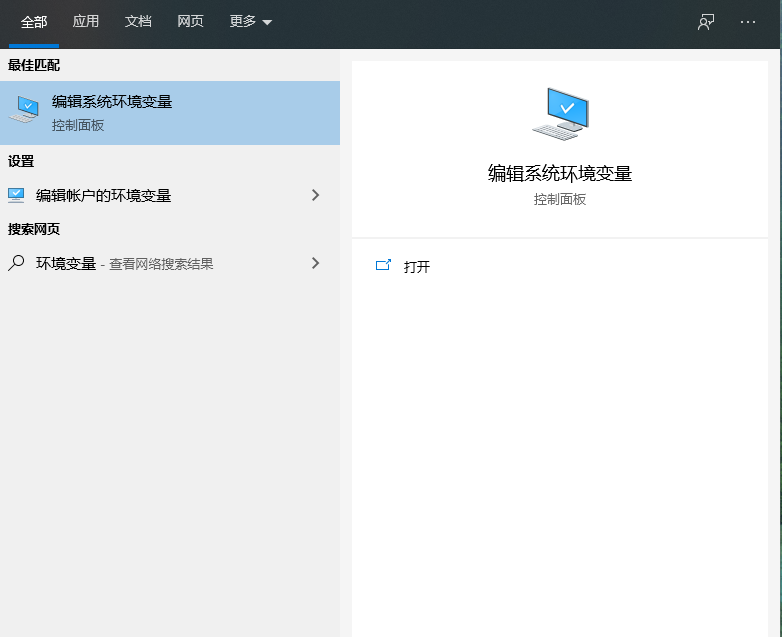




## 第二步：环境变量配置

左下角空白位置输入环境变量（右击我的电脑，点击属性，再点击高级系统变量也可）



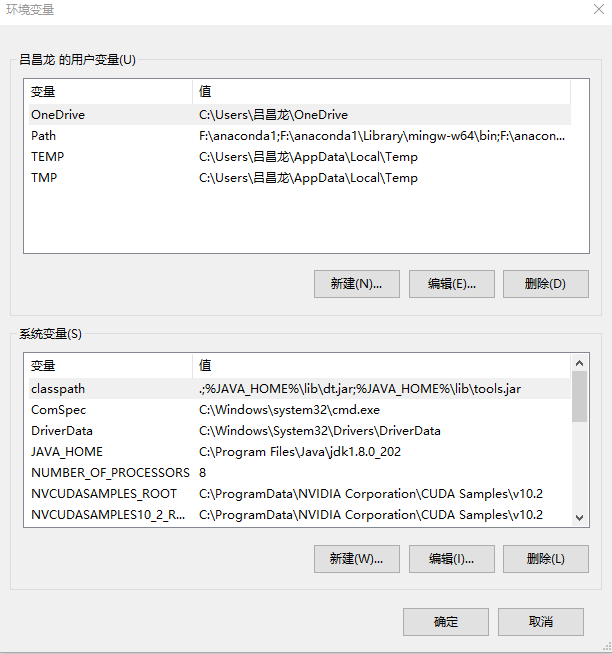




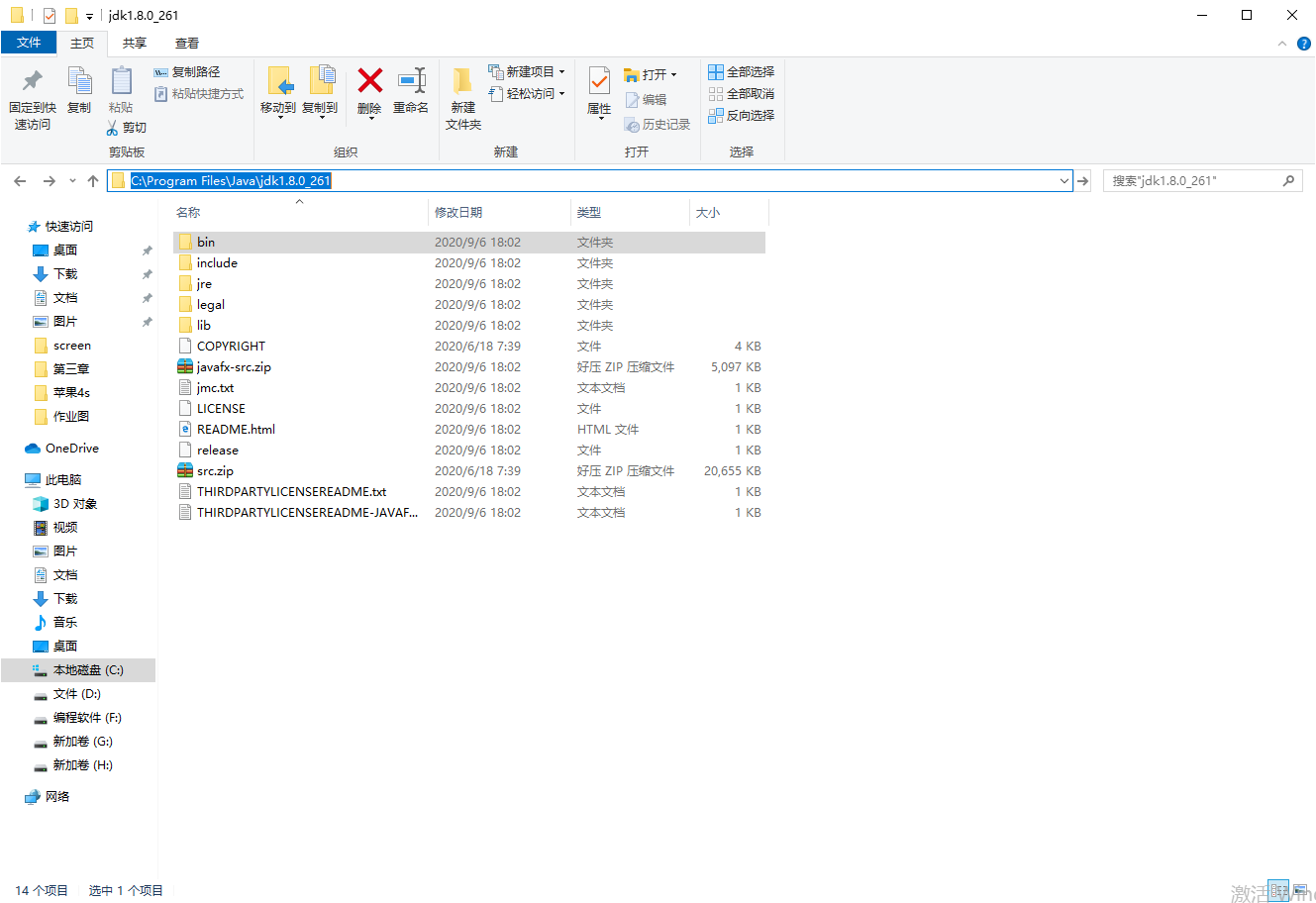
点击编辑系统环境变量



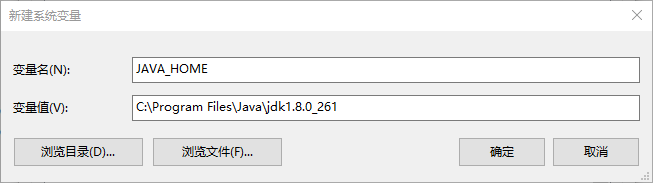
点击环境变量



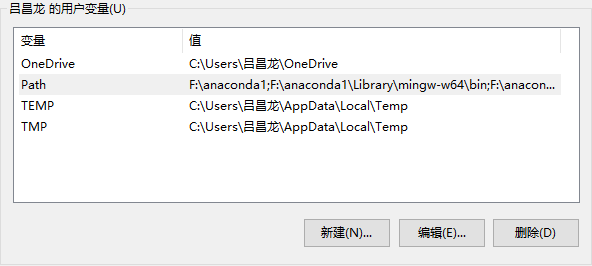
在环境变量对话框中的系统变量中点击“新建”按钮，在弹出的新建系统变量里的变量名中输入JAVA\_HOME，在变量值中输入你自己的JDK的根目录：C:\Program Files\Java\jdk1.8.0\_261 ，然后点击确定，返回环境变量对话框。如下图所示，打开jdk，复制路径。



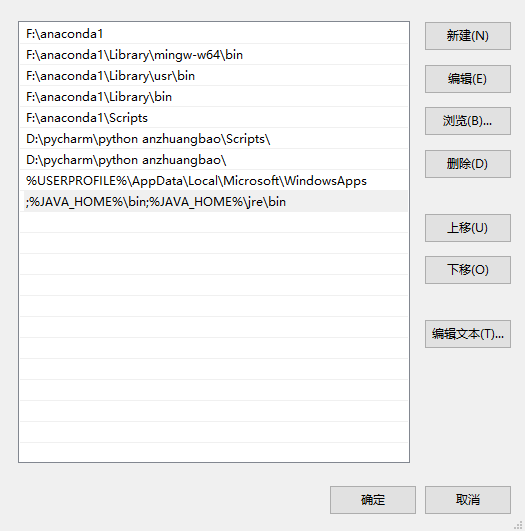
在系统变量中查看是否有“classpath”变量，如果没有，则新建这个变量，如果有，双击，进行添加。变量名classpath  变量值 .;%JAVA\_HOME%\lib; 注意，此变量值以英文句点符号开始，以分好结束。然后点击确定，返回环境变量对话框。



最后，在系统变量里面找到Path变量，注意，这次是点击编辑按钮，在弹出的对话框中的变量值的最后，一定是最后，添加如下字符串：;%JAVA\_HOME%\bin;%JAVA\_HOME%\jre\bin。注意，前面第一个是分号。如果没有Path变量，则添加Path变量，添加步骤和前面一样，不再重复。

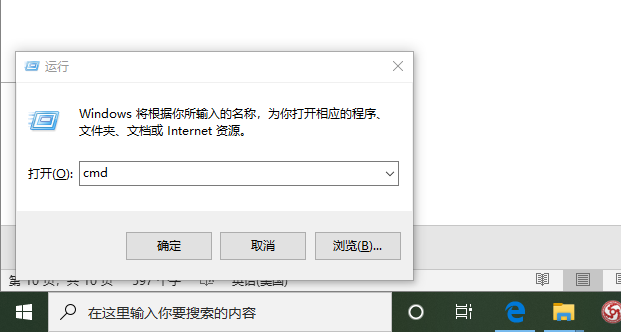


点击新建进行添加



最后，点击确定。

Win+r，输入cmd，打开终端。



输入，Java -version

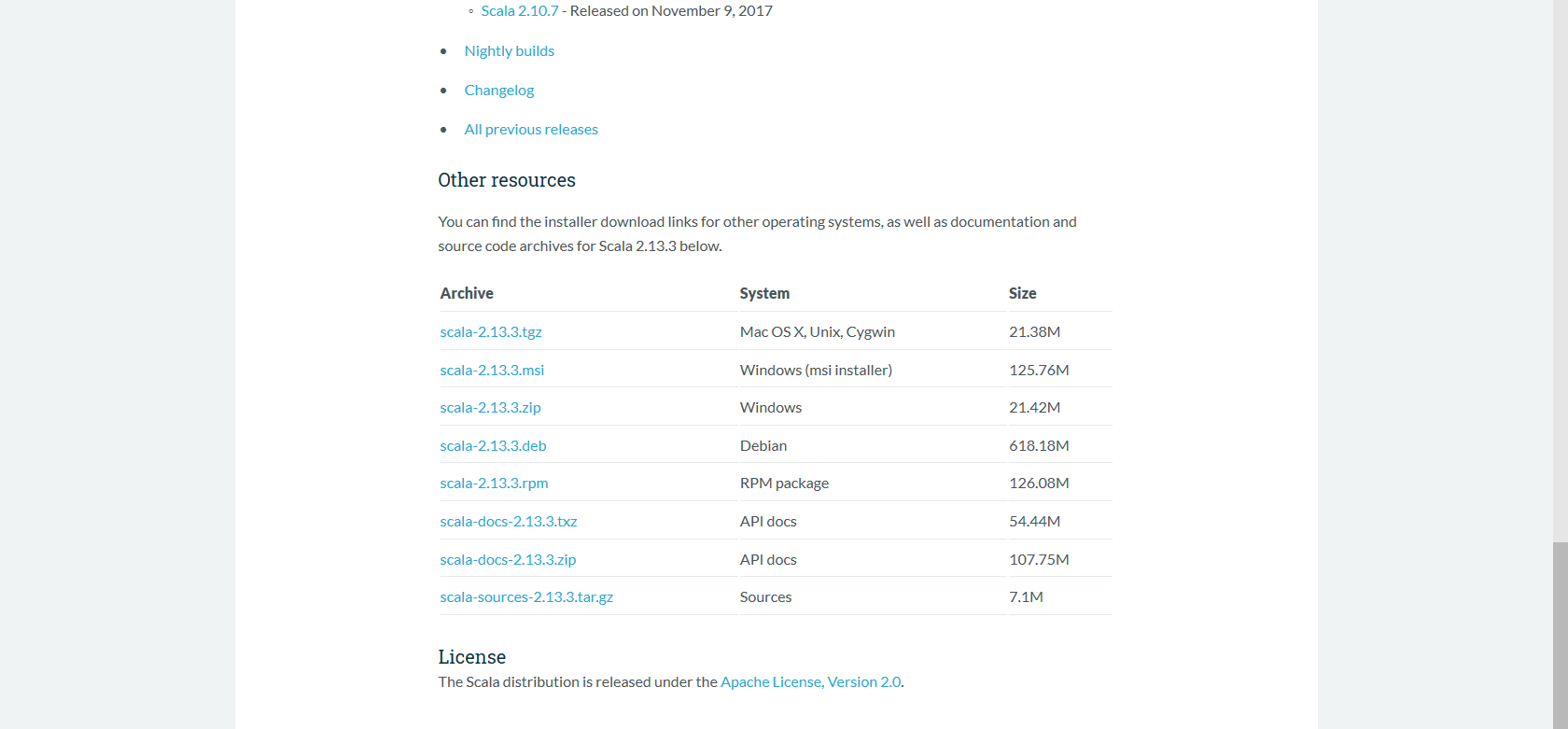


说明安装成功。

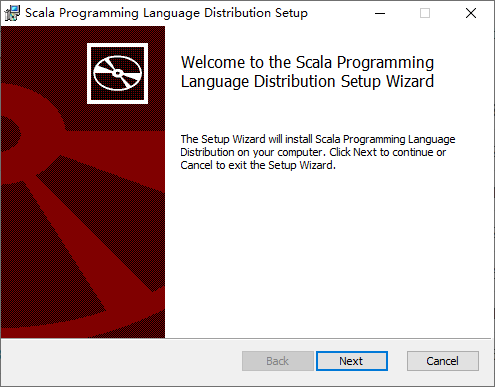
## 第三步：scala的安装

下载地址：<https://www.scala-lang.org/download/>

打开后，划到最下面，找到适合你系统的点击进行下载。



依旧是傻瓜式安装，一直点击下一步就行。路径按照你自己的喜好进行选择。



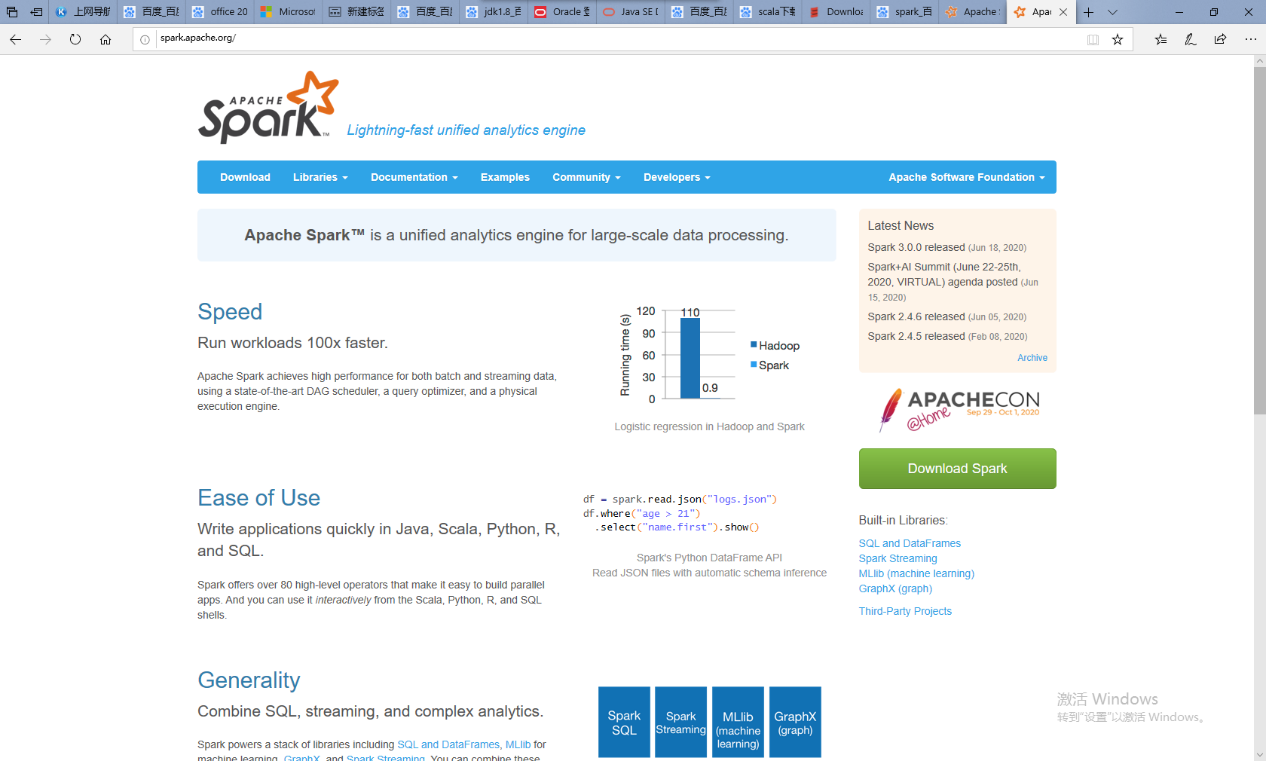
Scala安装好之后呢，它会自动在环境变量PATH里面配置。我们不需要进行手动配置了。下面就打开cmd输入Scala看下安装后控制台的输出效果。



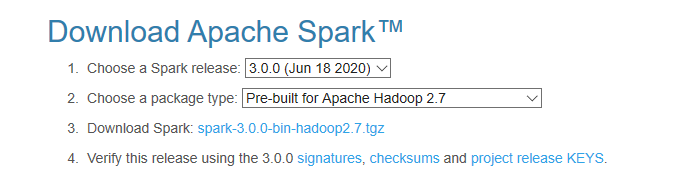
这样，就说明scala安装成功，下面我们就开始安装spark

## 第四步：spark的安装

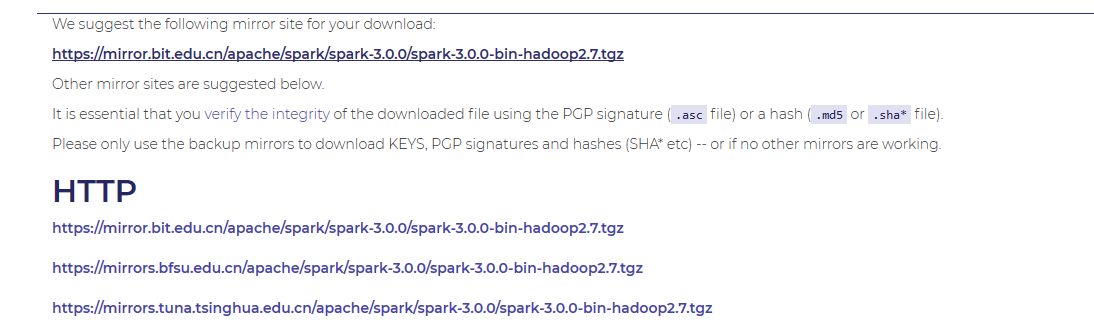
Spark的安装地址：http://spark.apache.org/



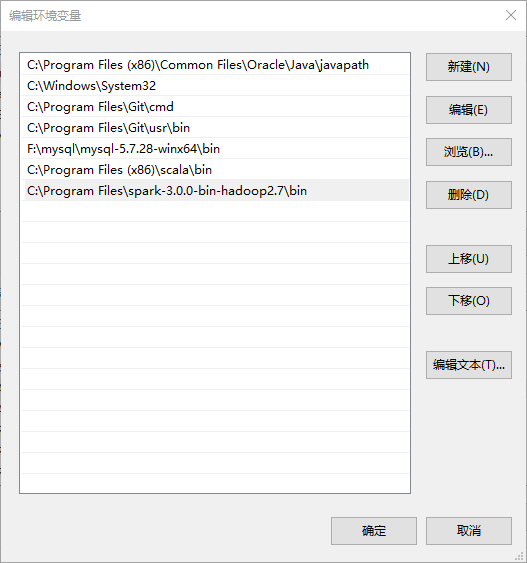
点击左上角的download



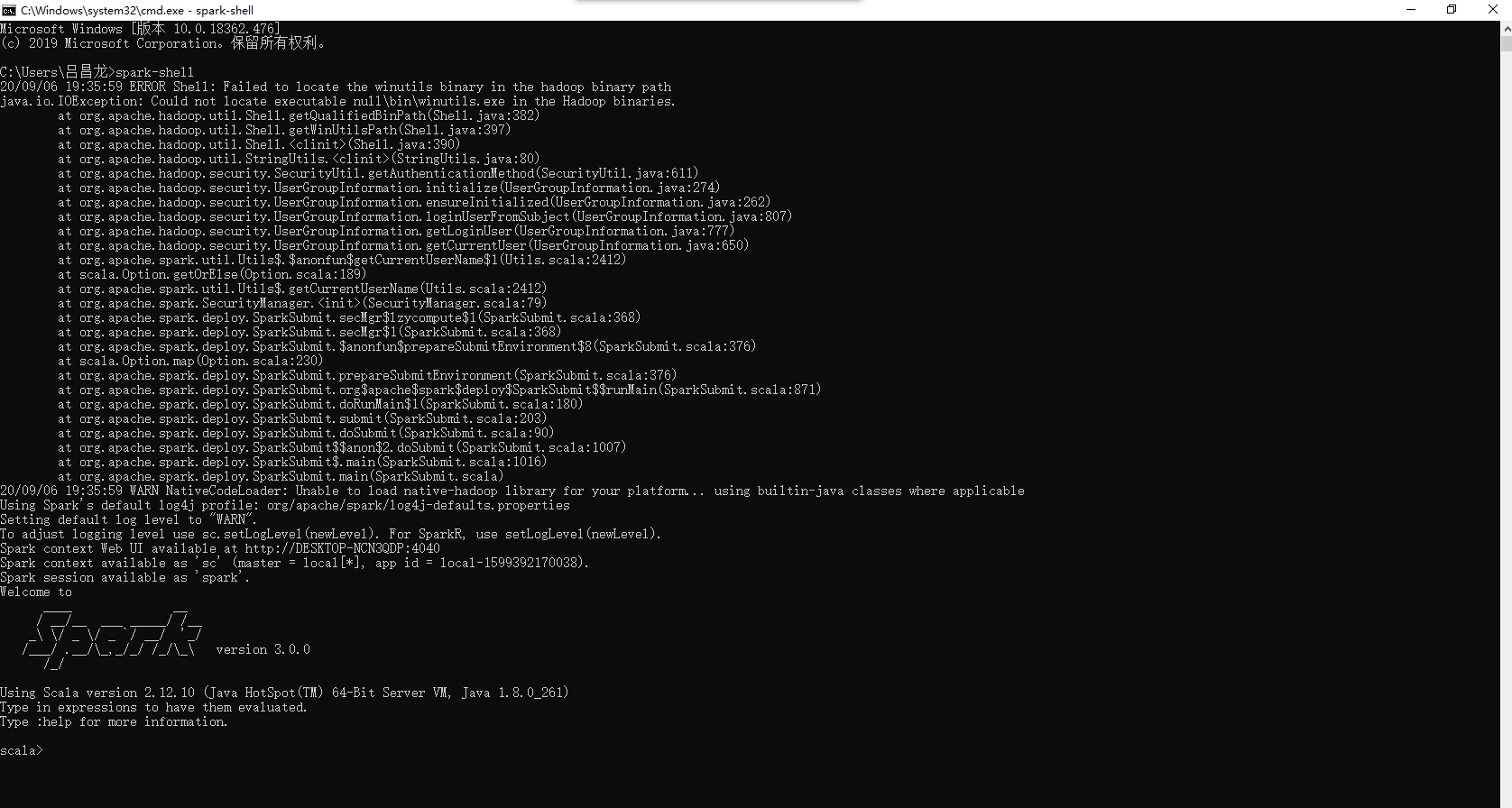
点击以后会出现如下界面，选择一个镜像，进行下载



下载并安装完毕后，对spark进行环境配置，同jdk一样，把路径添加到系统变量的path即可。



下面我们来验证一下看spark是否能正常启动。   
在控制台输入一下命令：spark-shell   
注意spark-shell之间是没有空格的，千万不要输错了。   
情况看图片

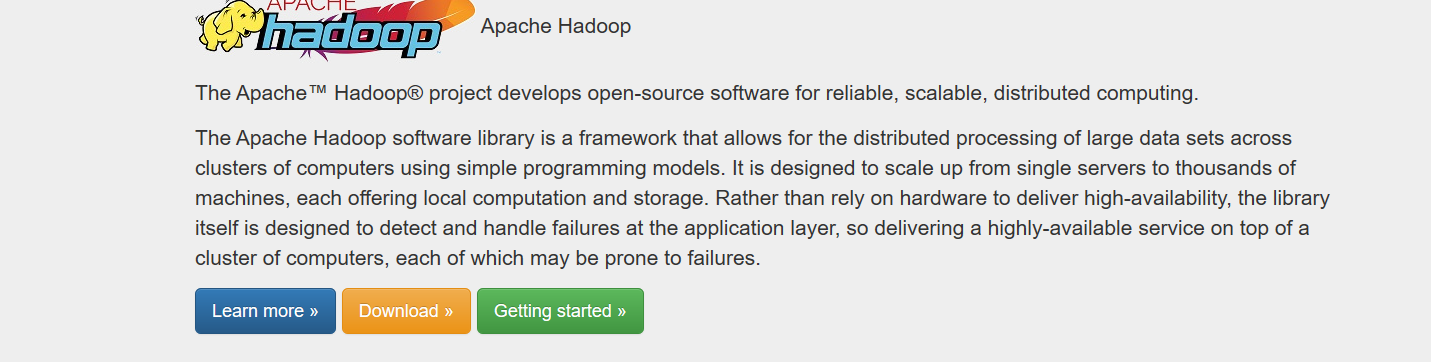


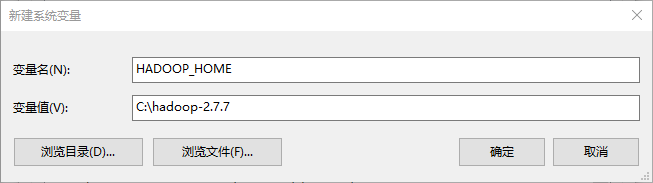
如图所示，spark安装成功。

## 第五步：Hadoop的安装及配置

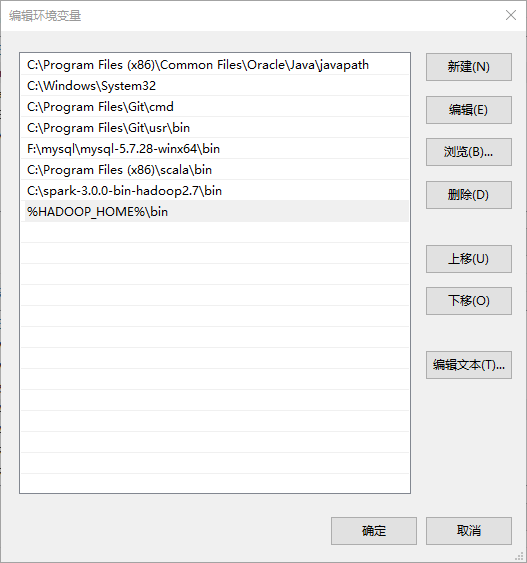
由于前面我们下载的spark是基于hadoop2.7的，所以我们下面将下载Hadoop 2.7

首先到官方下载官网的hadoop2.7.7,链接如下  
https://mirrors.tuna.tsinghua.edu.cn/apache/hadoop/common/  
找网盘的hadooponwindows-master.zip  
链接如下  
https://pan.baidu.com/s/1vxtBxJyu7HNmOhsdjLZkYw  
提取码：y9a4  
把hadoop-2.7.7.tar.gz解压后  
使用hadooponwindows-master的bin和etc替换hadoop2.7.7的bin和etc（这一步很重要）

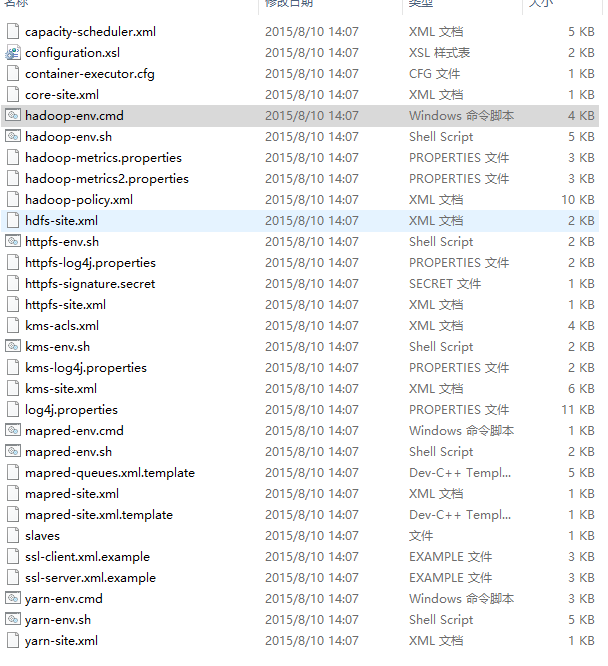


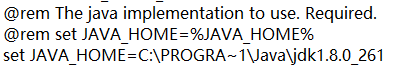


在path中增加%HADOOP\_HOME%\bin

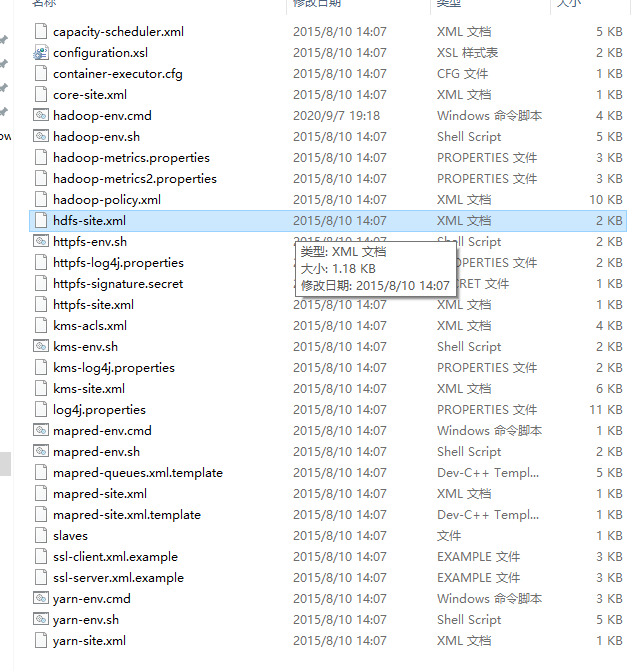


使用编辑器打开C:\hadoop-2.7.7\etc\hadoop\hadoop-env.cmd  
修改JAVA\_HOME的路径  
把set JAVA\_HOME改为jdk的位置  
注意其中PROGRA~1代表Program Files  
set JAVA\_HOME= C:\ PROGRA~1 \Java\jdk1.8.0\_261

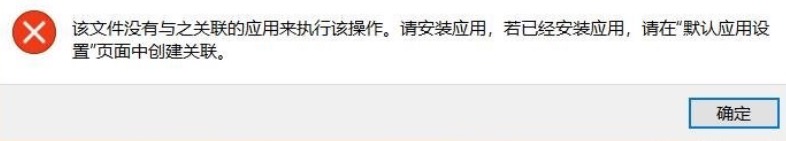




注意我标红的地方，不是简单的进行复制路径哦，这个地方复制了路径，后面报错很麻烦的。p下的namenode和datanode

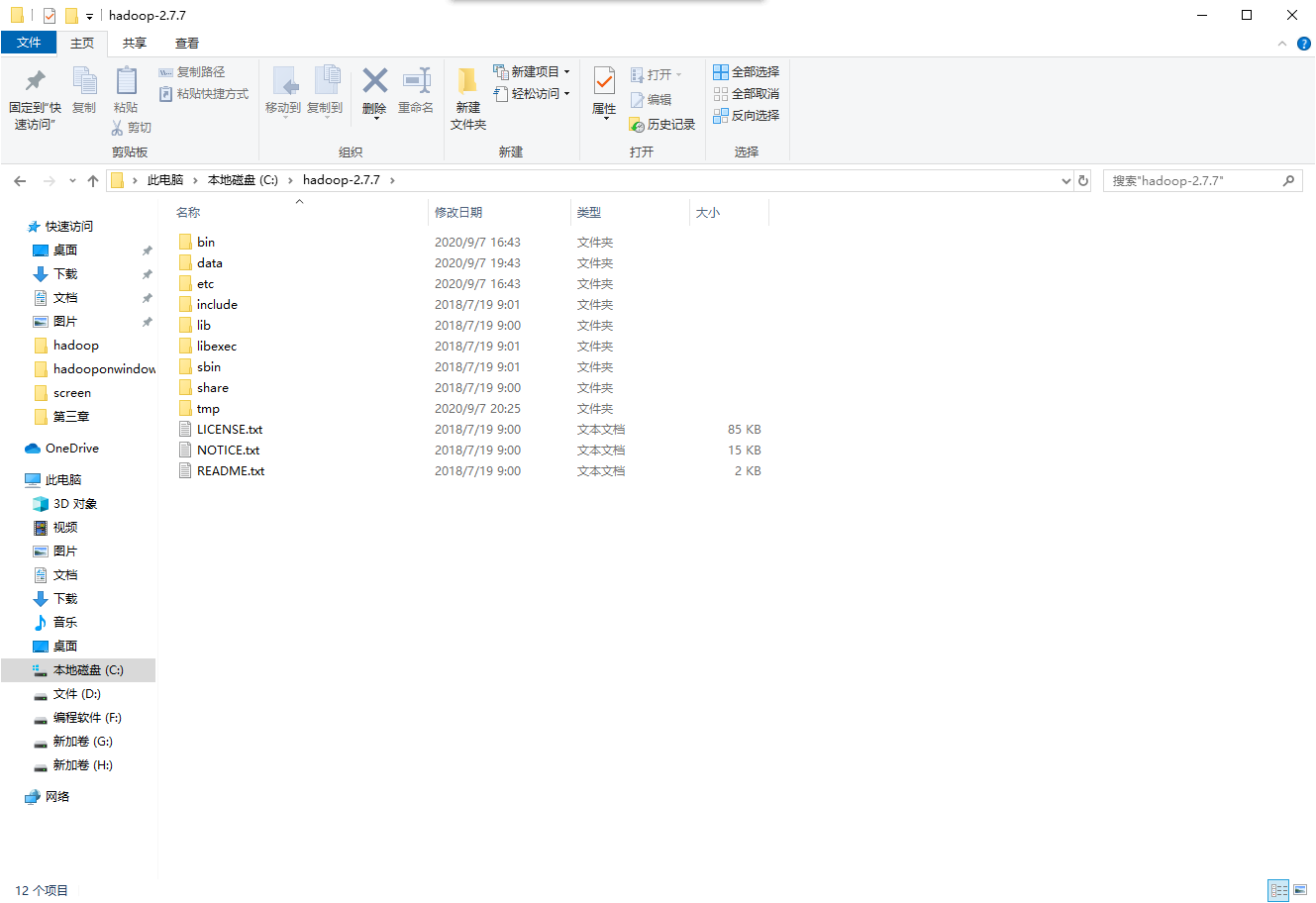


双击打开hdfs-site-xml,在这里我遇到一个错误

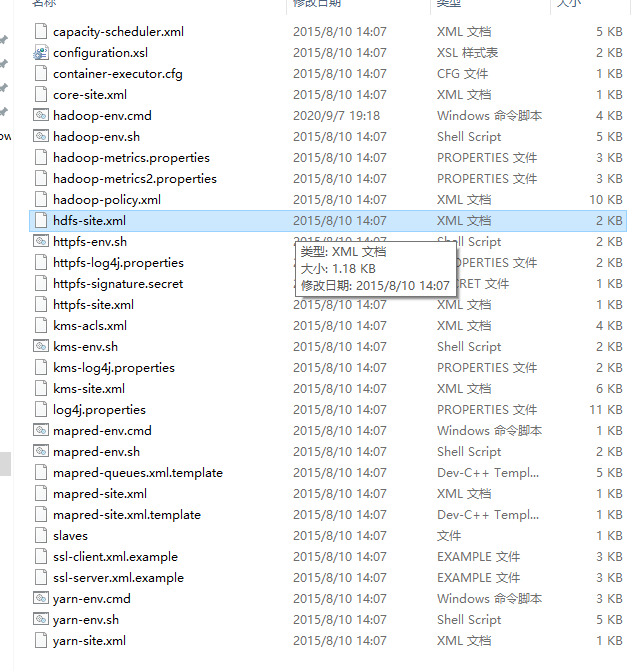


遇到这个问题看，那就用记事本打开。

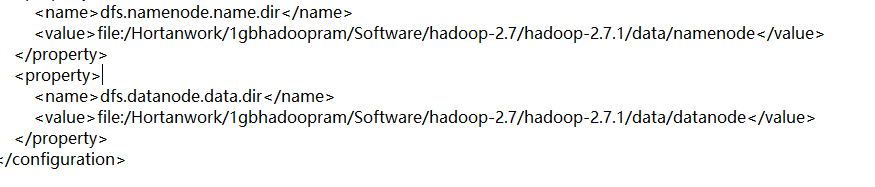
在hadoop文件夹下创建data和tmp文件夹，并在data文件夹下创建datanode和namenode两个子文件夹。



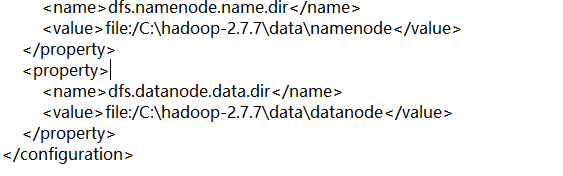
打开hdfs-site.xml



打开hdfs-site.xml



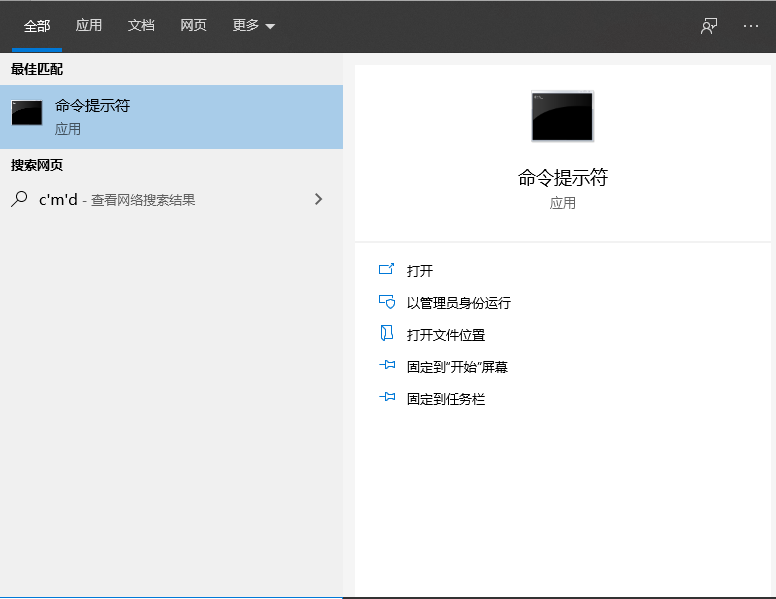
修改路径为Hadoop下的namenode和datanode：如下所示：



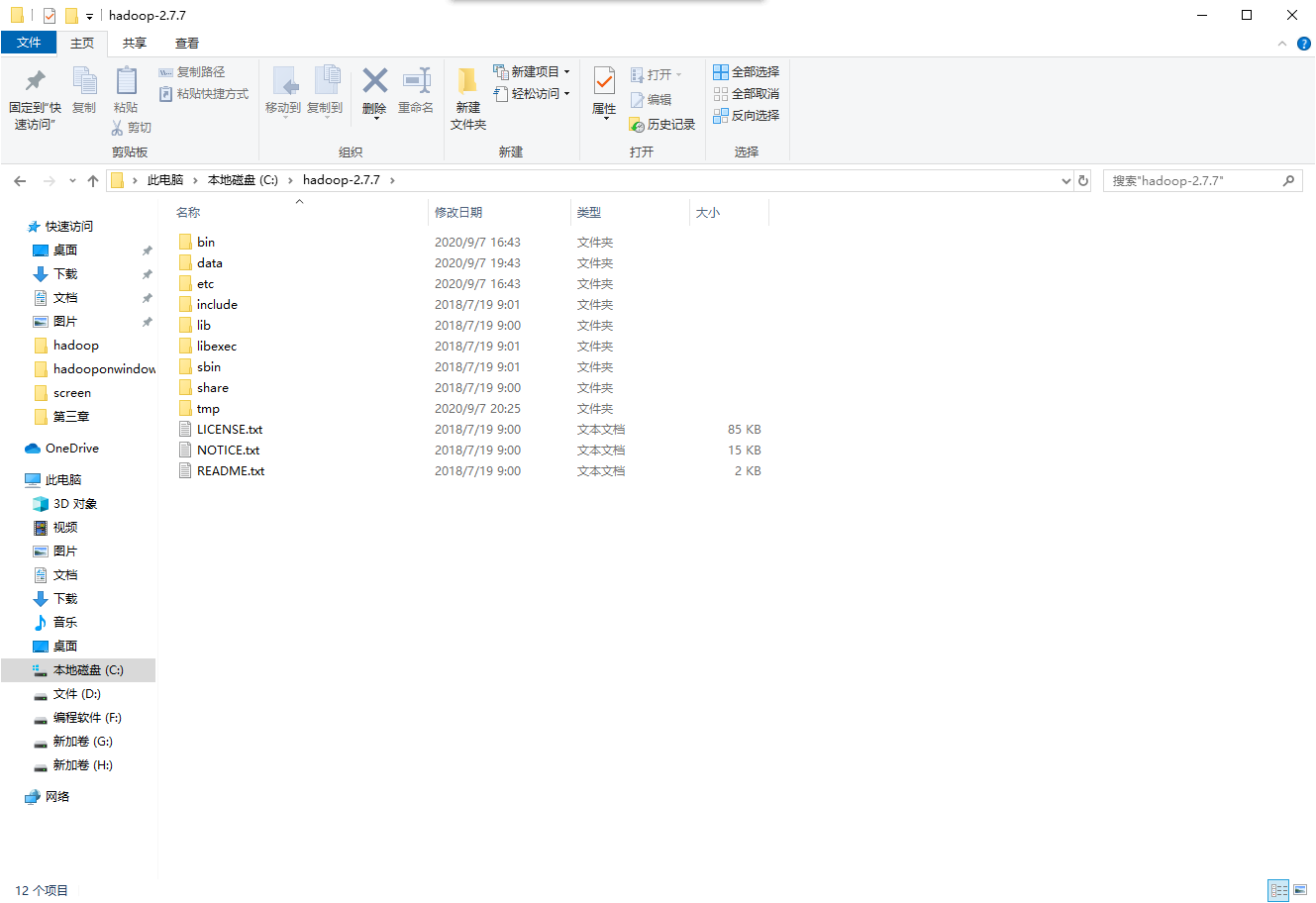
将hadoop.dll（在bin里面）拷贝到c:\Windows\system32中，不然在windows中运行mapreduce测试时会报错

以管理员身份打开命令提示符；输入hdfs namenode -format,看到successfully就说明format成功了。

点击这个，输入cmd，



点击理员身份运行



同志们，终于成功了，这玩意真的巨麻烦，想起来后后面竟然六个实验，我就头疼，看看镜子里面的自己，头顶又吗明亮了许多。

## 第六步：Scala IDE



