1. 安装spark，下载：<http://spark.apache.org/downloads.html>
   1. 加环境变量：SPARK\_HOME，并把SPARK\_HOME\bin加入到PATH环境变量
   2. 下载：<https://github.com/SweetInk/hadoop-common-bin.git>，加环境变量：HADOOP\_HOME（如：D:\big\_data\hadoop-common-bin\2.7.3），并把HADOOP \_HOME\bin加入到PATH环境变量
   3. 现在就可以运行pyspark了，本目录下有一个用python写spark应用的代码模板：spark\_tpl.py