**Config the environment of Pyspark.**

Firstly, java se 8, 安装好后，设置JAVA\_HOME, and Path加入$JAVA\_HOME$\bin

Note: java must install in the C:\program files...... that is the default path, otherwise configing hadoop will have bug. %JAVA\_HOME%\bin ;%JAVA\_HOME%\jre\bin; %HADOOP\_HOME%\bin;



Config anaconda 到Path

E:\Anaconda3\;E:\Anaconda3\Scripts\;E:\Anaconda3\Library\bin;

在CMD中pip install pyspark，然后在输入spark就可以激活spark。可以配置变量直接用jupyter notebook 打开。

安装hadoop，参见视频https://www.bilibili.com/video/BV1JT4y1e7Jb/?spm\_id\_from=333.337.search-card.all.click