# Step 1: Install JDK

For windows machine, you also need to set Java home after you install JDK.

* Right-click the My Computer icon on your desktop and select Properties
* Click the Advanced tab
* Click the Environment Variables button
* Under System Variables, click New
* Enter the variable name as JAVA\_HOME
* Enter the variable value as the installation path for the Java Development Kit

# Step 2: Download and Test Apache Spark

* Untar the file
* Test PySpark
* cd spark-1.4.1-bin-hadoop2.6
* ./bin/pyspark
* if everything goes well, you can see the PySpark interactive shell

# Step 3: Install Anaconda Scientific Python Distribution

Open “Launcher” after installing anaconda python. Choose to update python notebook and then Launch it. Now you should see python notebook (Jupyter) is open in your browser.

# Step 4: Run Spark in Notebook

You can follow the steps in the following link to configure Notebook. Then Spark will be automatically loaded when you open a new notebook.<http://thepowerofdata.io/configuring-jupyteripython-notebook-to-work-with-pyspark-1-4-0/>

Or the script in notebook cell to start spark.

<http://kdd2015-sparktutorial.droppages.com/images/pyspark.png>

The first cell is to start PySpark. The second cell is a test for PySpark.