

Setting Up our Environment Locally

Things to Install

- Oracle VirtualBox (for running virtual machines)
- Ubuntu (Linux Operating System for virtual machine)
- Python 3 (Python 3.5, specifically. Not 2.7 or 3.6)
- Pip3 (for installing Python 3 packages)
- Jupyter Notebooks (for ease of coding and testing)
- Java 8 (Spark 2 does NOT support Java 9 yet)
- Scala 2.11 (version is important)
- Py4j (Python package that allows Python to access Java objects)
- Spark 2.1.0 & Hadoop 2.7 (These work together)

...followed by setting the environment variables and file permissions.

Things to Install

- Install Scala <https://www.scala-lang.org/download/>
- Install Java 8
<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>
- Spark 2.1.1 <https://spark.apache.org/downloads.htm>
- Python 3.6 with Jupyter <https://www.anaconda.com/download/>

Things to Install

- Install Scala <https://www.scala-lang.org/download/>
- Install Java 8
<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>
- Spark 2.1.1 <https://spark.apache.org/downloads.htm>
- Python 3.6 with Jupyter <https://www.anaconda.com/download/>
- (*important for Windows users*) Windows Utilities
<https://github.com/jleetutorial/sparkTutorial/blob/winutils/winutils.exe>

Things to Install

- Install Scala <https://www.scala-lang.org/download/>
- Install Java 8
<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>
- Spark 2.1.1 <https://spark.apache.org/downloads.htm>
- Python 3.6 with Jupyter <https://www.anaconda.com/download/>
- (*important for Windows users*) Windows Utilities
<https://github.com/jleetutorial/sparkTutorial/blob/winutils/winutils.exe>
- Git <https://git-scm.com/downloads>

Things to Install

- Install Scala <https://www.scala-lang.org/download/>
- Install Java 8
<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>
- Spark 2.1.1 <https://spark.apache.org/downloads.htm>
- Python 3.6 with Jupyter <https://www.anaconda.com/download/>
- (*important for Windows users*) Windows Utilities
<https://github.com/jleetutorial/sparkTutorial/blob/winutils/winutils.exe>
- Git <https://git-scm.com/downloads>
- (*optional*) Sublime Text 3 <https://www.sublimetext.com/3>
- (*optional*) Microsoft Visual Studio <https://code.visualstudio.com/>
- (*optional*) PyCharm <https://www.jetbrains.com/pycharm/download/>