Lesson 1: Introduction to the Spark Environment

1.5 Getting Set Up: Spark and Java

Installation: Requirements

- Spark binary (version 1.4.1)

- Java JDK 6/7

- Scientific Python (and Jupyter notebook)

- `py4j`

- (Optional) IRKernel (for Jupyter)

Installation: Requirements

- Spark binary (version 1.4.1)

- Java JDK 6/7

- Scientific Python (and Jupyter notebook)

- `py4j`

- (Optional) IRKernel (for Jupyter)

Installation: Requirements

NOTE Please do not install Spark with

- ***Homebrew*** on OSX

- ***Cygwin*** on Windows

Installation: Spark

- Find your OS here: http://spark.apache.org/downloads.html

  - Select "`Pre-built for Hadoop 2.4`" or earlier

    under "`Choose a package type`"

  - Download the tar package for `spark-1.4.1-bin-hadoop1.tgz`

    (If you are not sure pick the latest version.)

  Make sure you are downloading the binary version, not the source version.

Installation: Configuration

- Unzip the file and place it at your home directory
  (`/Users/jonathandinu/`)

- **Set PATH:** Include the following lines in your
  `~/.bash_profile` (or `~/.bashrc`):

Installation: Configuration

- Unzip the file and place it at your home directory (`/Users/jonathandinu/`)

- **Set PATH:** Include the following lines in your `~/.bash_profile` (or `~/.bashrc`):

```
export SPARK_HOME=/full/path/to/your/unzipped/spark/folder

export PYTHONPATH=$SPARK_HOME/python/:$PYTHONPATH
```

Installation: Java JDK

- http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

- Find download for your OS

- Follow install instructions/wizard

- Make sure you get **JDK** instead of **JRE**

Installation: Requirements

- Spark binary

- Java JDK 6/7

- Scientific Python (and Jupyter notebook)

- `py4j`

- (Optional) IRKernel (for Jupyter)