System Setup: Mac

Install the JDK

Install Oracle's Java SE Development Kit 8

- oracle.com/technetwork/java/javase/downloads

- Click the **JDK Download** link

- Don't download the "Demo/Samples" file

Install Python

- Any version from 2.7.6 onward will do

- Enter the "python" command into the OS X Terminal to verify that Python is not already installed on your system.

- If Python is not installed, visit:

  python.org/downloads

Download the Spark Tarball

- http://spark.apache.org/downloads

- Release: 1.4.0

- Package Type: Hadoop 2.6

- Download Type: Direct

- Click "spark-1.4.0-bin-hadoop2.6.tgz" to begin the download process.

Extract the Tarball

- Move the tarball into your "home" directory.
  - If you are unsure where your "home" directory is located, just open your OS X Finder and then "shift-command-H".

- Extract the tarball by double-clicking on it with your left mouse button.

- The extracted folder is what we will refer to as your "Spark installation".

Eliminate annoying logs

- Enter your Spark installation using the OS X Finder.

- Enter the "conf" folder.

- Duplicate the "log4j.properies.template" file.

- Change the duplicated file's filename to "log4j.properties".

Eliminate annoying logs (2)

- Open "log4j.properties".

- Change the rootCategory from "INFO" to "WARN" - then save and close the file.

Launch the Python Shell

```
bin/pyspark
```

Run a simple command

Import a file into Spark.

```
testFile = sc.textFile("README.md")
```

Perform a line count.

```
testFile.count()
```

Exit your shell via control-D

Setup Complete

Congrats! Your Spark setup process is complete. If you have any colleagues who wish to learn Apache Spark, please refer them to these two sites:

liondatasystems.com/courses

liondatasystems.com/events