



System Setup: PC

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.
- If you don't already have Python installed, visit this site to obtain the MSI installer:
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

- Use 7-Zip to perform the extraction.
 - You can download 7-Zip at 7-Zip.org
- The extracted folder is what we will refer to as your "Spark installation".

Eliminate annoying logs

- Enter your Spark installation using the Windows Explorer.
- Enter the "conf" folder.
- Duplicate the "log4j.properties.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.

"cd" into your Spark installation

- Open your Windows Terminal.
- "cd" into your Spark installation.

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