Software specifications

Chapter number	Software required (With version)	Free/Proprietary	If proprietary, can code testing be performed using a trial version	If proprietary, then cost of the software	Download links to the software	Hardware specifications	OS required
1-10	Java JDK 6.0+	Free	N/A	N/A	Oracle JDK: http://www.oracle.com/technetwork/java/javase/downloads/index.html OpenJDK: http://openjdk.java.net/install/index.html		Linux / Unix / Mac OS X / Windows
1, 10	SBT	Free	N/A	N/A	http://www.scala-sbt.org/release/docs/Getting-Started/Setup.html		Linux / Unix / Mac OS X / Windows
1	Maven	Free	N/A	N/A	http://maven.apache.org/download.cgi		Linux / Unix / Mac OS X / Windows
1-10	Apache Spark (Version 1.2.0 with Hadoop	Free	N/A	N/A	Spark Downloads page: http://spark.apache.org/downloads.html		Linux / Unix / Mac OS

	2.4)				Apache mirror: http://www.apache.org/dyn/closer.cgi/spark/spark-1.2.0/spark-1.2.0-bin-hadoop2.4.tgz		X. Windows supported by Spark but not specifically covered in the book.
1-10	Python (Version 2.6+)	Free	N/A	N/A	https://www.python.org/downloads/		Linux / Unix / Mac OS X – Python is normally a part of the OS distribution. Windows requires Python to be downloaded and installed separate

							ly.
3,6,8	NumPy (Version 1.4+)	Free	N/A	N/A	http://www.scipy.org/install.html		Linux / Unix / Mac OS X / Windows
3,6,8	Matplotlib	Free	N/A	N/A	http://matplotlib.org/downloads.html		Linux / Unix / Mac OS X / Windows
3,6,8	IPython (Version 2.0+)	Free	N/A	N/A	http://ipython.org/install.html		Linux / Unix / Mac OS X / Windows

Detailed installation steps (software-wise)

The steps should be listed in a way that it prepares the system environment to be able to test the codes of the book.

1. Java JDK – required for running Java and Scala Spark code:
    a. Most Linux, Unix and Mac OS X distributions come with Java installed. This can be tested by running java –version.
    b. If required to be installed, for Linux and Unix follow the instructions for OpenJDK here: http://openjdk.java.net/install/index.html.
    c. Alternatively, the Oracle JDK can be installed from the following instructions: http://www.oracle.com/technetwork/java/javase/downloads/index.html.
2. Apache Spark:

a. Download the pre-built Apache Spark binaries from *http://www.apache.org/dyn/closer.cgi/spark/spark-1.2.0/spark-1.2.0-bin-hadoop2.4.tgz*.
b. Unzip the downloaded file using *tar xfvz spark-1.2.0-bin-hadoop2.4.tgz*.
c. Change directory into the created folder using *cd spark-1.2.0-bin-hadoop2.4*.
d. Test the Scala shell using *./bin/spark-shell*.
e. Test the Python shell using *./bin/pyspark*.
f. Run an example using *./bin/run-example SparkPi*.
g. If all of it works, then that means Java and Python are both installed correctly on your system.

3. Python scientific stack:
    a. For #4, #5, #6 and #7 below, a **highly recommended** way of installing all the required Python components is to install Anaconda (http://continuum.io/downloads) or Enthought's Canopy (https://www.enthought.com/downloads/).
    b. These packages provide Python, Numpy, Scipy, matplotlib and IPython all in one easy to install package.
4. Python:
    a. If using Linux, Unix or Mac OS X, Python is usually installed already as part of the OS. You can test the version by running *python –version* from the command line.
    b. If using Windows, download and install the Python installation from *https://www.python.org/downloads/*.
5. NumPy and SciPy:
    a. If using *Anaconda*, NumPy will already be installed as part of the distribution.
    b. If using *Enthought Canopy*, NumPy will already be installed as part of the distribution.
    c. If installing directly, follow the instruction located at *http://www.scipy.org/install.html*.
6. Matplotlib:
    a. If using *Anaconda*, NumPy will already be installed as part of the distribution.
    b. If using *Enthought Canopy*, NumPy will already be installed as part of the distribution.
    c. If installing directly, follow the instructions located at http://matplotlib.org/1.4.2/users/installing.html
7. IPython:
    a. If using *Anaconda*, run *conda update conda*, followed by *conda update ipython* to install the latest version of IPython.
    b. If using *Enthought Canopy*, run *enpkg ipython* to install the latest version of IPython.
    c. If using Ubuntu, run *apt-get install ipython* (or *sudo apt-get install ipython* if necessary) to install IPython.
    d. Otherwise, follow the instructions here: http://ipython.org/install.html