Software list:

For this book you need a personal computer (can be either Windows machine, Mac, or Linux). To run Apache Spark, you will need Java 7+ and an installed and configured Python 2.6+ or 3.4+ environment; we use the Anaconda distribution of Python in version 3.5, which can be downloaded from https://www.continuum.io/downloads.

The Python modules we randomly use throughout the book come preinstalled with Anaconda. We also use GraphFrames and TensorFrames that can be loaded dynamically while starting a Spark instance: to load these you just need an Internet connection. It is fi ne if some of those modules are not currently installed on your machine – we will guide you through the installation process.