

Chapter number	Software required (With version)	Hardware specifications	OS required
1-8, 10	<ol style="list-style-type: none"> <li>1. Spark 2.0.0 (or higher)</li> <li>2. Hadoop 2.7 (or higher)</li> <li>3. Java (JDK and JRE) 1.7+/1.8+</li> <li>4. Scala 2.11.x (or higher)</li> <li>5. Python 2.6+/3.4+</li> <li>6. R 3.1+ and RStudio 0.99.879 (or higher)</li> <li>7. Eclipse Mars or Luna (latest)</li> <li>8. Maven Eclipse plugin (2.9 or higher)</li> <li>9. Maven compiler plugin for Eclipse (2.3.2 or higher)</li> <li>10. Maven assembly plugin for Eclipse (2.4.1 or higher)</li> </ol> <p>Most importantly, re-use the provided pom.xml file with Packt supplementary and change the above mentioned version and APIs accordingly and everything will be sorted out.</p>	<ul style="list-style-type: none"> <li>• Processor Core i3, Core i5 (recommended) ~ Core i7 (to get best result). However, multicore processing would provide faster data processing and scalability.</li> <li>• At least 8GB RAM (recommended) for a standalone mode</li> <li>• At least 32 GB RAM for a single VM and higher for cluster</li> <li>• Enough storage for running heavy jobs (depending the dataset size you will be handling) preferably at least 50GB of free disk storage (for stand-alone and for SQL warehouse)</li> </ul>	<ul style="list-style-type: none"> <li>• Linux distributions are preferable (including Debian, Ubuntu, Fedora, RHEL, CentOS etc.). To be more specific, for example, for Ubuntu it is recommended to have 14.04 (LTS) 64-bit complete installation or VMWare player 12 or Virtual box.</li> <li>• Windows (XP/7/8/10)</li> <li>• Mac OS X (10.4.7+)</li> </ul>
9	Same as above	For the Streaming related job, for example, if you want to do window operations over 10 minutes, then even more data needs to be kept buffered $600 \text{ sec} * 10\text{MB/sec} = 6 \text{ GB}$ . Therefore, large-scale RAM and storage is recommended although it really depends on the input dataset to be processed. Moreover, if you are targeting 500-1000 records per second and your workload is simple enough, total of 10 - 20 cores should be more than enough.	Same as above