# Software specifications

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Chapter no. | 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 |
| 2 | Sbt 0.13.1 | Free | n/a | n/a | scala-sbt.org | n/a | CentOS 6.5 64 bit |
| 2 | Apache Spark 1.0 | Free | n/a | n/a | cloudera.com | n/a | CentOS 6.5 64 bit |
| 2 | Cloudera CDH 5.3 | Free | n/a | n/a | cloudera.com | n/a | CentOS 6.5 64 bit |
| 2 | Apache Maven 3.2.5 | Free | n/a | n/a | repos.fedorapeople.org | n/a | CentOS 6.5 64 bit |
| 2 | Apache Spark 1.3 | Free | n/a | n/a | spark.apache.org | n/a | CentOS 6.5 64 bit |
| 3 | Apache Flume 1.5 | Free | n/a | n/a | Flume.apache.org | n/a | CentOS 6.5 64 bit |
| 3 | Apache Kafka 1.2 | Free | n/a | n/a | archive.cloudera.com | n/a | CentOS 6.5 64 bit |
| 4 | Postgresql 9.4 | Free | n/a | n/a | jdbc.postgresql.org | n/a | CentOS 6.5 64 bit |
| 5 | Docker 1.4.1 | Free | n/a | n/a | Yum based install | n/a | CentOS 6.5 64 bit |
| 5 | Mazerunner 1.x | Free | n/a | n/a | Docker based install | n/a | CentOS 6.5 64 bit |
| 5 | Neo4j 2.1.7 | Free | n/a | n/a | Docker based install | n/a | CentOS 6.5 64 bit |
| 6 | Titan 0.9.0-M2 | Free | n/a | n/a | s3.thinkaurelius.com | n/a | CentOS 6.5 64 bit |
| 6 | HBase 0.98.6 | Free | n/a | n/a | CDH cluster install | n/a | CentOS 6.5 64 bit |
| 6 | SparkOnHBase-cdh5-0.0.2 | Free | n/a | n/a | github.com | n/a | CentOS 6.5 64 bit |
| 6 | Cassandra 2.0.13 | Free | n/a | n/a | datastax.com | n/a | CentOS 6.5 64 bit |
| 6 | spark-cassandra-connector\_2.10 | Free | n/a | n/a | com.datastax.spark | n/a | CentOS 6.5 64 bit |
| 7 | Spark 1.2 | Free | n/a | n/a | CDH 5.1.3 cluster install | n/a | CentOS 6.5 64 bit |
| 7 | H2O 0.2.12-95 | Free | n/a | n/a | 0xdata.com/download | n/a | CentOS 6.5 64 bit |

# 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.

Note: Once a tool has been installed it is available for all chapters

1. Software SBT:

Chapter 2 – install described in book

1. Software Apache Spark 1.0

Chapter 2 – install described in book

1. Software Cloudera CDH 5.3

Note to editor – generally I have used HDFS in this book but I don’t consider it to be what the book is about. If the user wants to know how to install and configure Hadoop they can read my first book ☺

1. Software Apache Maven 3.2.5

Chapter 2 – install described in book

1. Software Apache Spark 1.3

Chapter 2 – install described in book

1. Software Apache Flume 1.5

Chapter 3 – step 1 -> yum install flume-ng

Chapter 3 – step 2 -> yum install flume-ng-agent

Chapter 3 – step 3 -> yum install flume-ng-agent-doc

Chapter 3 – step 4 -> cd /etc/flume-ng/conf

Chapter 3 – step 4 -> cp flume-conf.properties.template flume.conf

1. Software Apache Kafka 1.2

Chapter 3 – install described in book

1. Software Postgresql 9.4

Chapter 4 – install described in book

1. Software Docker 1.4.1

Chapter 5 – install described in book

1. Software Mazerunner 1.x

Chapter 5 – install described in book

1. Software Neo4j 2.1.7

Chapter 5 – install described in book

1. Software Titan 0.9.0-M2

Chapter 6 – install described in book

1. Software HBase 0.98.6

Chapter 6 – install described in book

1. Software SparkOnHBase-cdh5-0.0.2

Chapter 6 – install described in book

1. Software Cassandra 2.0.13

Chapter 6 – install described in book

1. Software spark-cassandra-connector\_2.10

Chapter 6 – install described in book

1. Software Spark 1.2

Chapter 7 – automated install – provided as part of CDH 5.3 hadoop cluster automatically

Software H2O 0.2.12-95 (also called sparkling water , h2O + spark = sparkling water)

Chapter 7 – install described in book