Recommended books

Big data management tools

This is a round-up of the important books to have on your bookshelf. All will help you learn more about the Apache Hadoop ecosystem, projects and related topics.

**Hadoop in Action**  
Learn how to write programs using MapReduce.

**Hadoop. The Definitive Guide**  
This is one of the best books on the market. It summarises what you need to know to master HDFS, MapReduce, Pig and also Spark.

**Hadoop in Practice**  
*Hadoop in Practice* collects 85 Hadoop examples and presents them in a problem/solution format.

**Programming Pig**  
Alan Gates, 2011.  
Do you want to know more about Apache Pig? This book will give you great insights into how to analyse large datasets that consist of a high-level language for expressing data analysis programs, coupled with infrastructure. It will also help you learn Pig Latin.

**Learning Spark: Lightning-Fast Big Data Analysis**  
This book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. With Spark, you can tackle big datasets quickly through simple APIs in Python, Java and Scala.

**MapReduce Design Patterns**  
Donald Miner & Adam Shook, 2013.  
If you are data scientists, design patterns are like recipes in a cookbook for you: very useful. Design patterns describe how to approach your data analytics problems once you know how your data looks.

Find the book in a library near you with Worldcat | [https://www.worldcat.org/](https://www.worldcat.org/)