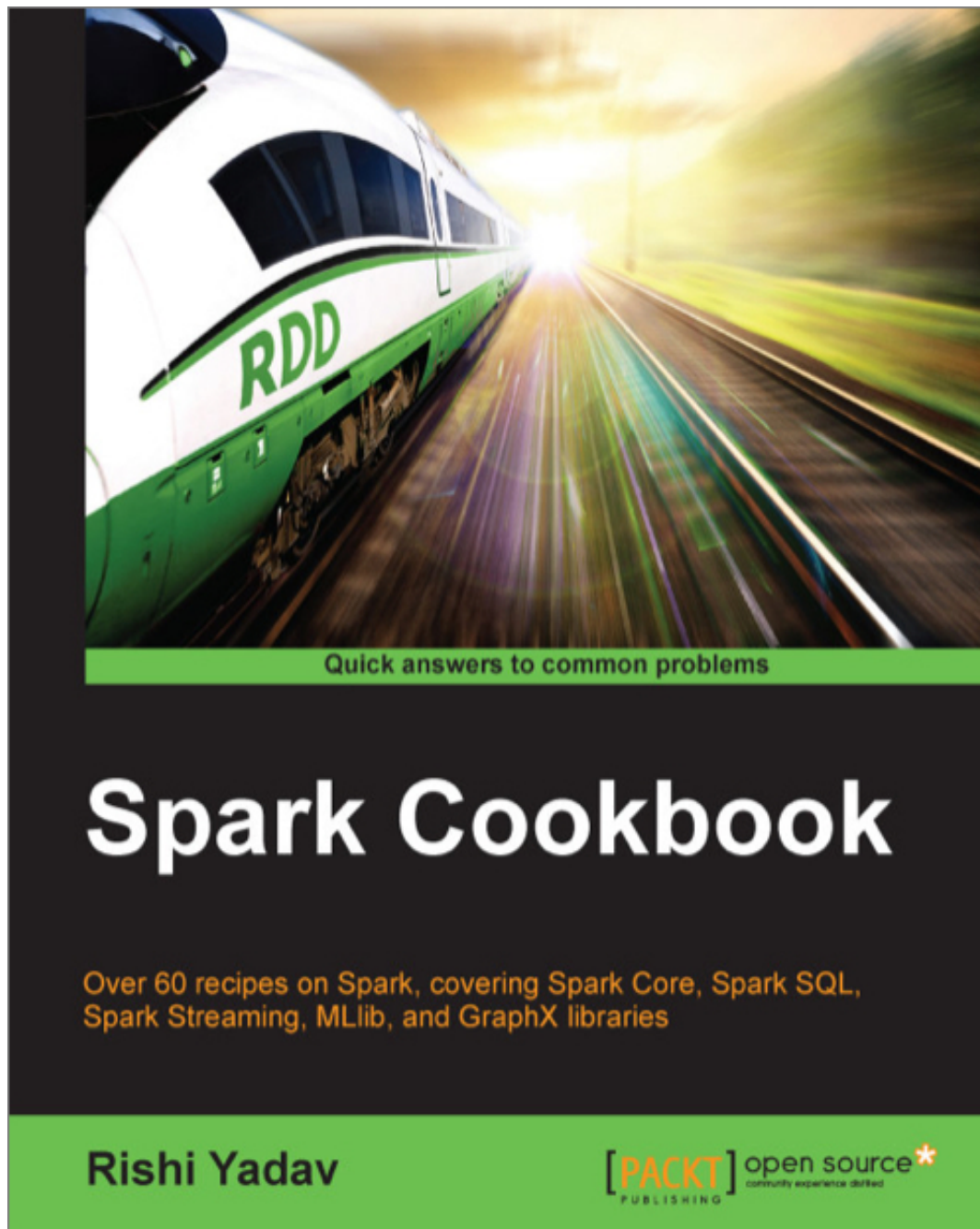


[\[电子书\]Spark Cookbook PDF下载](#)

本书重点介绍如何分析大量而且复杂的数据集。本书开头介绍了如何在各种集群管理上安装和配置Apache Spark，其中也会涵盖开发环境的设置。然后介绍了如何通过Spark SQL和实时流对各种数据源进行交互式查询，其中的实时流包括了Twitter Stream 和 Apache Kafka。然后，本书将专注于机器学习，包括监督学习，无监督学习和推荐引擎算法。在使用GraphX掌握图形处理后，本书将介绍集群优化和故障排除的各种技巧。

Spark Cookbook图书由Rishi Yadav所著，全书共226页；Packt Publishing出版社于2015年07月出版。



如果想及时了解Spark、Hadoop或者Hbase相关的文章，欢迎关注微信公共帐号：iteblog_hadoop

本书的章节

- Chapter 1: Getting Started with Apache Spark
- Chapter 2: Developing Applications with Spark
- Chapter 3: External Data Sources
- Chapter 4: Spark SQL
- Chapter 5: Spark Streaming
- Chapter 6: Getting Started with Machine Learning Using MLlib

Chapter 7: Supervised Learning with MLlib – Regression
Chapter 8: Supervised Learning with MLlib – Classification
Chapter 9: Unsupervised Learning with MLlib
Chapter 10: Recommender Systems
Chapter 11: Graph Processing Using GraphX
Chapter 12: Optimizations and Performance Tuning

下载地址

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。
本文链接: **【】**（**）**