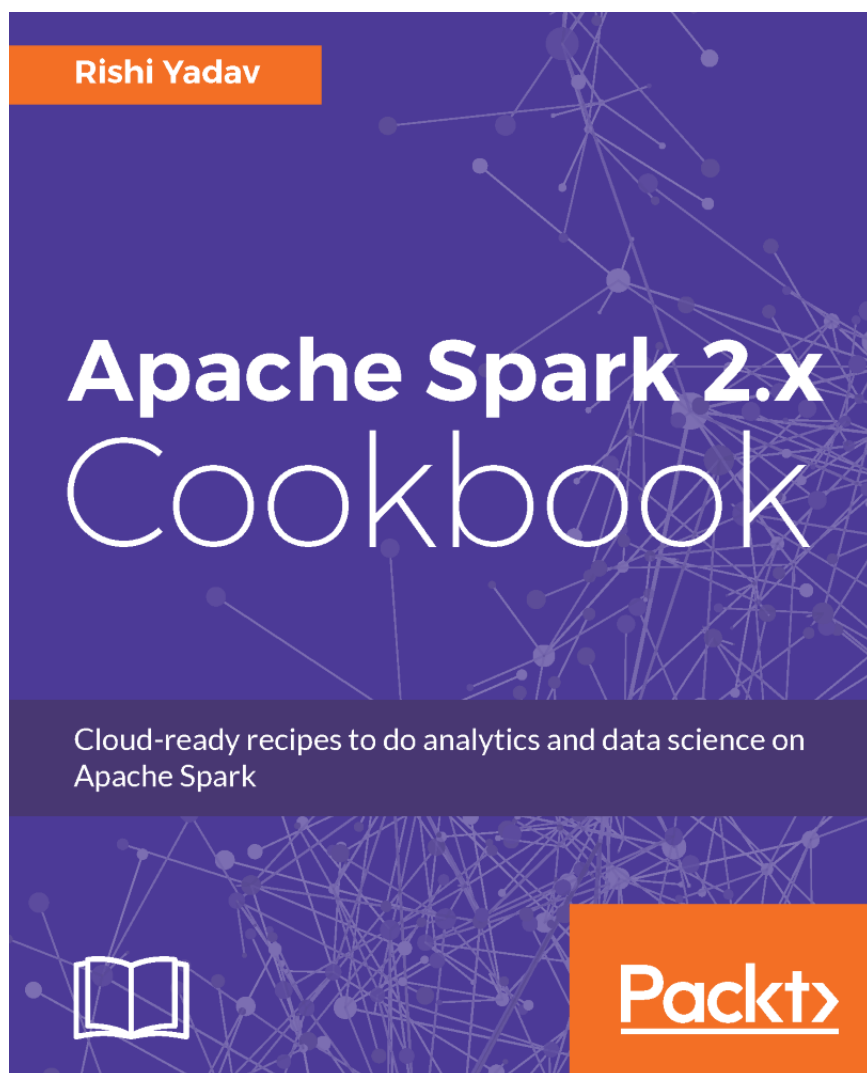


[\[电子书\]Apache Spark 2.x Cookbook, 2nd Edition PDF下载](#)

本书于2017-05由Packt Publishing出版，作者Rishi Yadav，全书294页。从书名就可以看出这是一本讲解技巧的书。本书副标题：Over 70 recipes to help you use Apache Spark as your single big data computing platform and master its libraries。本书适合数据工程师，数据科学家以及那些想使用Spark的读者。阅读本书之前最好有Scala的编程基础。通过本书你将学到以下知识：

- Install and configure Apache Spark with various cluster managers & on AWS
- Set up a development environment for Apache Spark including Databricks Cloud notebook
- Find out how to operate on data in Spark with schemas
- Get to grips with real-time streaming analytics using Spark Streaming & Structured Streaming
- Master supervised learning and unsupervised learning using MLlib
- Build a recommendation engine using MLlib
- Graph processing using GraphX and GraphFrames libraries
- Develop a set of common applications or project types, and solutions that solve complex big data problems



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

- 1、Getting Started with Apache Spark
- 2、Developing Applications with Spark
- 3、Spark SQL
- 4、Working with External Data Sources
- 5、Spark Streaming
- 6、Getting Started with Machine Learning
- 7、Supervised Learning with MLlib - Regression
- 8、Supervised Learning with MLlib - Classification
- 9、Unsupervised learning
- 10、Recommendations Using Collaborative Filtering
- 11、Graph Processing Using GraphX and GraphFrames
- 12、Optimizations and Performance Tuning

下载地址

提供了PDF、azw3 二种格式的下载。

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: **【】**（ ）