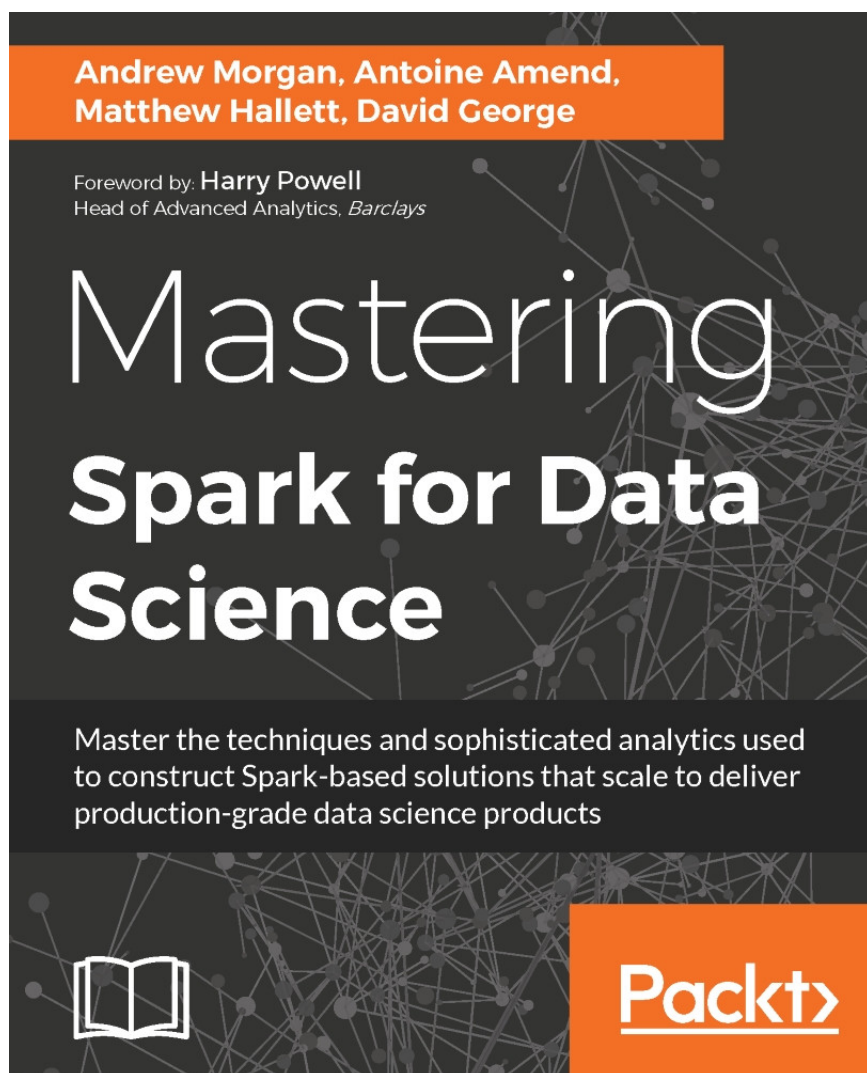


[\[电子书\]Mastering Spark for Data Science PDF下载](#)

本书由Andrew Morgan所著，全书共560页；Packt Publishing出版社于2017年03月出版。通过本书你将学习到以下的知识：

- 1、 Learn the design patterns that integrate Spark into industrialized data science pipelines
- 2、 See how commercial data scientists design scalable code and reusable code for data science services
- 3、 Explore cutting edge data science methods so that you can study trends and causality
- 4、 Discover advanced programming techniques using RDD and the DataFrame and Dataset APIs
- 5、 Find out how Spark can be used as a universal ingestion engine tool and as a web scraper
- 6、 Practice the implementation of advanced topics in graph processing, such as community detection and contact chaining
- 7、 Get to know the best practices when performing Extended Exploratory Data Analysis, commonly used in commercial data science teams
- 8、 Study advanced Spark concepts, solution design patterns, and integration architectures
- 9、 Demonstrate powerful data science pipelines



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 THE BIG DATA SCIENCE ECOSYSTEM
- 2 DATA ACQUISITION
- 3 INPUT FORMATS AND SCHEMA
- 4 EXPLORATORY DATA ANALYSIS
- 5 SPARK FOR GEOGRAPHIC ANALYSIS
- 6 SCRAPING LINK-BASED EXTERNAL DATA
- 7 BUILDING COMMUNITIES
- 8 BUILDING A RECOMMENDATION SYSTEM
- 9 NEWS DICTIONARY AND REAL-TIME TAGGING SYSTEM
- 10 STORY DE-DUPLICATION AND MUTATION
- 11 ANOMALY DETECTION ON SENTIMENT ANALYSIS
- 12 TREND CALCULUS
- 13 SECURE DATA

14 SCALABLE ALGORITHMS

下载地址

关注本微信公众号 iteblog_hadoop 并回复 m-Spark 获取本书的下载地址。或

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: **【】**（）