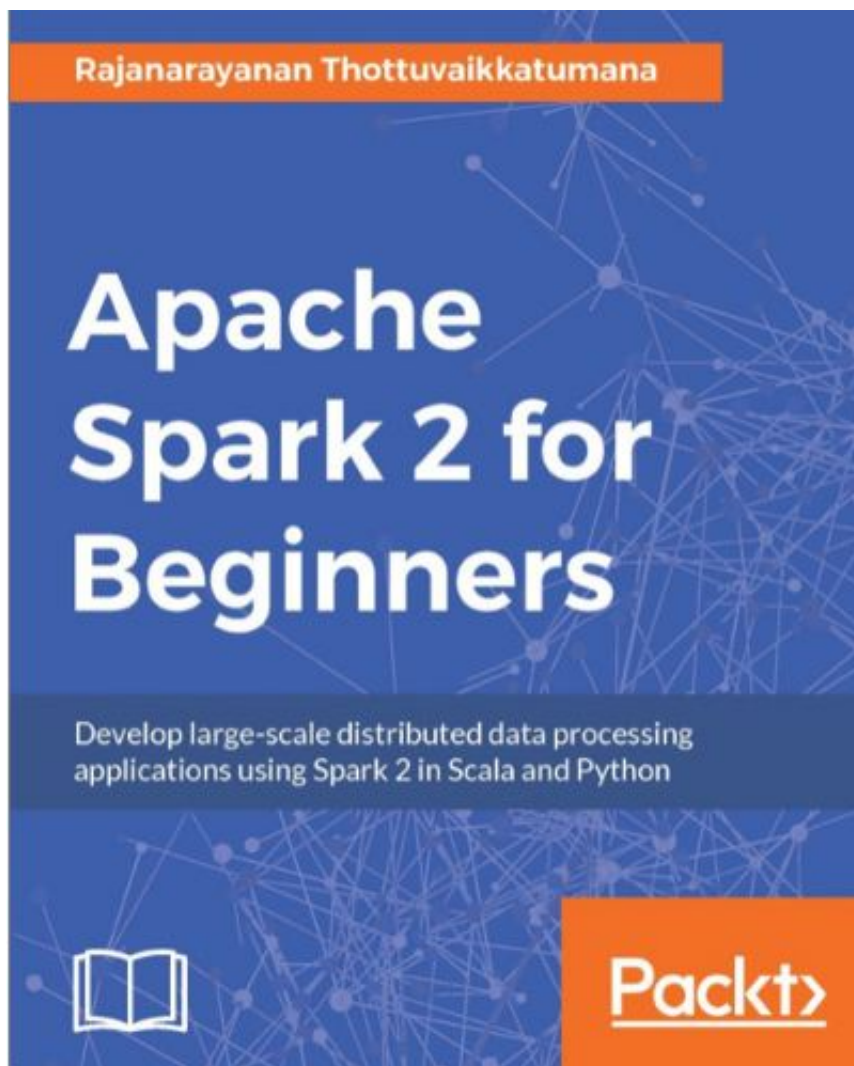


[\[电子书\]Apache Spark 2 for Beginners pdf下载](#)

本书由Packt出版，2016年10月发行，全书共332页。从标题可以看出这本书是适用于初学者的，全书的例子有Scala和Python两个版本，涵盖了Spark基础、编程模型、SQL、Streaming、机器学习以及图计算等知识。



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

本书的章节如下：

- Chapter 1: Spark Fundamentals
- Chapter 2: Spark Programming Model
- Chapter 3: Spark SQL
- Chapter 4: Spark Programming with R
- Chapter 5: Spark Data Analysis with Python

Chapter 6: Spark Stream Processing
Chapter 7: Spark Machine Learning
Chapter 8: Spark Graph Processing
Chapter 9: Designing Spark Applications

详细目录

Preface

Chapter 1: Spark Fundamentals

- An overview of Apache Hadoop
- Understanding Apache Spark
- Installing Spark on your machines
 - Python installation
 - R installation
 - Spark installation
 - Development tool installation
 - Optional software installation
 - IPython
 - RStudio
 - Apache Zeppelin
- References
- Summary

Chapter 2: Spark Programming Model

- Functional programming with Spark
- Understanding Spark RDD
 - Spark RDD is immutable
 - Spark RDD is distributable
 - Spark RDD lives in memory
 - Spark RDD is strongly typed
- Data transformations and actions with RDDs
- Monitoring with Spark
- The basics of programming with Spark
 - MapReduce
 - Joins
 - More actions
- Creating RDDs from files
- Understanding the Spark library stack
- Reference
- Summary

Chapter 3: Spark SQL

- Understanding the structure of data
- Why Spark SQL?
- Anatomy of Spark SQL

- DataFrame programming

 - Programming with SQL

 - Programming with DataFrame API

- Understanding Aggregations in Spark SQL

- Understanding multi-datasource joining with SparkSQL

- Introducing datasets

- Understanding Data Catalogs

- References

- Summary

Chapter 4: Spark Programming with R

- The need for SparkR

- Basics of the R language

- DataFrames in R and Spark

- Spark DataFrame programming with R

 - Programming with SQL

 - Programming with R DataFrame API

- Understanding aggregations in Spark R

- Understanding multi-datasource joins with SparkR

- References

- Summary

Chapter 5: Spark Data Analysis with Python

- Charting and plotting libraries

- Setting up a dataset

- Data analysis use cases

- Charts and plots

 - Histogram

 - Density plot

 - Bar chart

 - Stacked bar chart

 - Pie chart

 - Donut chart

 - Box plot

 - Vertical bar chart

 - Scatter plot

 - Enhanced scatter plot

 - Line graph

- References

- Summary

Chapter 6: Spark Stream Processing

- Data stream processing

- Micro batch data processing

 - Programming with DStreams

- A log event processor

 - Getting ready with the Netcat server

 - Organizing files

 - Submitting the jobs to the Spark cluster

- Monitoring running applications
- Implementing the application in Scala
- Compiling and running the application
- Handling the output
- Implementing the application in Python
- Windowed data processing
 - Counting the number of log event messages processed in Scala
 - Counting the number of log event messages processed in Python
- More processing options
- Kafka stream processing
 - Starting Zookeeper and Kafka
 - Implementing the application in Scala
 - Implementing the application in Python
- Spark Streaming jobs in production
 - Implementing fault-tolerance in Spark Streaming data processing applications
 - Structured streaming
- References
- Summary
- Chapter 7: Spark Machine Learning
 - Understanding machine learning
 - Why Spark for machine learning?
 - Wine quality prediction
 - Model persistence
 - Wine classification
 - Spam filtering
 - Feature algorithms
 - Finding synonyms
 - References
 - Summary
- Chapter 8: Spark Graph Processing
 - Understanding graphs and their usage
 - The Spark GraphX library
 - GraphX overview
 - Graph partitioning
 - Graph processing
 - Graph structure processing
 - Tennis tournament analysis
 - Applying the PageRank algorithm
 - Connected component algorithm
 - Understanding GraphFrames
 - Understanding GraphFrames queries
 - References
 - Summary
- Chapter 9: Designing Spark Applications
 - Lambda Architecture
 - Microblogging with Lambda Architecture

- An overview of SfbMicroBlog
- Getting familiar with data
- Setting the data dictionary
- Implementing Lambda Architecture
 - Batch layer
 - Serving layer
 - Speed layer
 - Queries
- Working with Spark applications
- Coding style
- Setting up the source code
- Understanding data ingestion
- Generating purposed views and queries
- Understanding custom data processes
- References
- Summary
- Index

下载地址

关注本微信公众号iteblog_hadoop并回复Spark2电子书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: **【】**（**）**