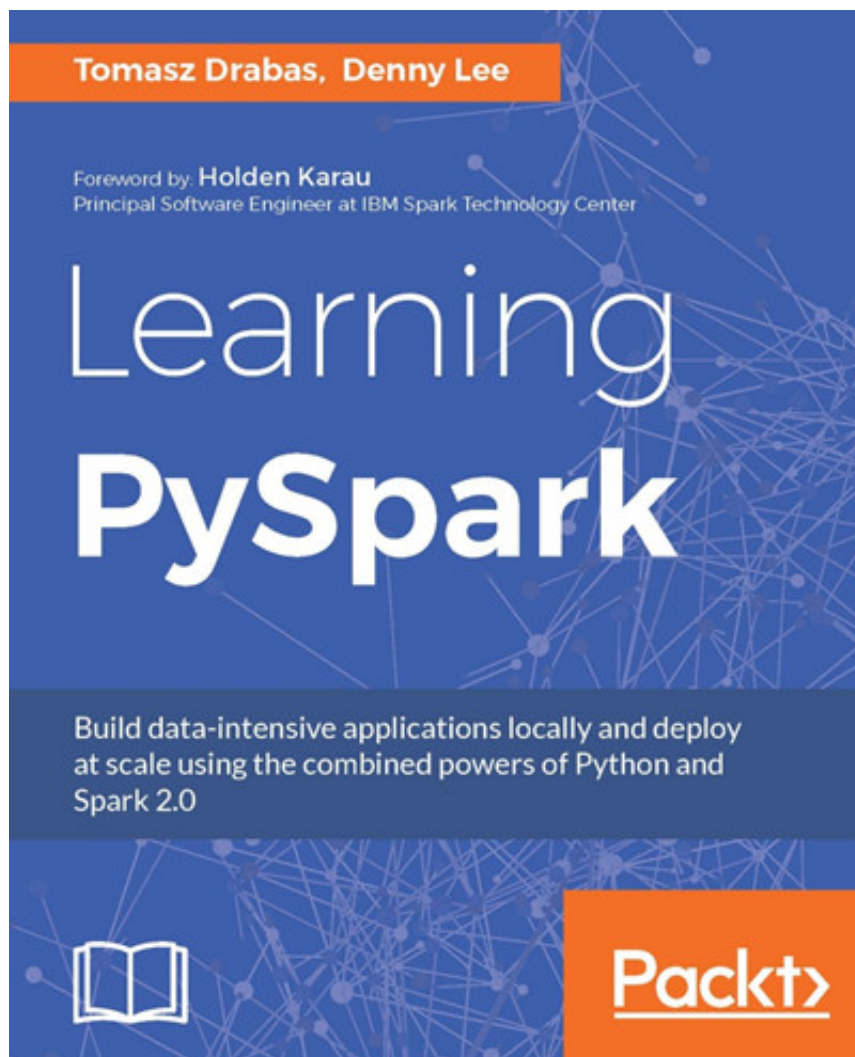


## [\[电子书\]Learning PySpark PDF下载](#)

本书将向您展示如何利用Python的强大功能并将其用于Spark生态系统中。您将首先了解Spark 2.0的架构以及如何为Spark设置Python环境。通过本书，你将会使用Python操作RDD、DataFrames、MLlib以及GraphFrames等；在本书结束时，您将对Spark Python API有了全局的了解，并且学习到如何使用它来构建数据密集型应用程序。通过本书你将学习到以下的知识：

- Learn about Apache Spark and the Spark 2.0 architecture
- Build and interact with Spark DataFrames using Spark SQL
- Learn how to solve graph and deep learning problems using GraphFrames and TensorFrames respectively
- Read, transform, and understand data and use it to train machine learning models
- Build machine learning models with MLlib and ML
- Learn how to submit your applications programmatically using spark-submit
- Deploy locally built applications to a cluster

本书由Tomasz Drabas所著，全书共273页；Packt Publishing出版社于2017年02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

- 1 UNDERSTANDING SPARK
- 2 RESILIENT DISTRIBUTED DATASETS
- 3 DATAFRAMES
- 4 PREPARE DATA FOR MODELING
- 5 INTRODUCING MLLIB
- 6 INTRODUCING THE ML PACKAGE
- 7 GRAPHFRAMES
- 8 TENSORFRAMES
- 9 POLYGLOT PERSISTENCE WITH BLAZE
- 10 STRUCTURED STREAMING
- 11 PACKAGING SPARK APPLICATIONS

## 下载地址

关注本微信公众号 iteblog\_hadoop 并回复 Learning\_PySpark 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: [【】](#)（）