

[\[电子书\]High Performance Spark下载](#)

如果你使用Apache

Spark解决了中等规模数据的问题，但是在海量数据使用Spark的时候还是会遇到各种问题。High Performance Spark将会向你展示如何使用Spark的高级功能，所以你可以超越新手级别。本书适合软件工程师、数据工程师、开发者以及Spark系统管理员的使用。本书作者Holden Karau, Rachel

Warren，由O'Reilly于2016年03月出版，

全书175页；本文提供的是本书预览版，只提供前四章，共91页（完整版请参见：[《\[电子书\]High Performance Spark完整版PDF下载》](#)）。通过本书你可以学到：

- 了解如何使Spark作业运行速度更快；
- 使用Spark探索数据；
- 使用Spark处理更大的数据集；
- 减少管道运行时间以获得更快的洞察力。



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 Introduction to High Performance Spark
- Chapter 2 How Spark Works
- Chapter 3 DataFrames, Datasets & Spark SQL
- Chapter 4 Joins (SQL & Core)
- Chapter 5 Effective Transformations

Chapter 6 Working with Key/Value Data
Chapter 7 Going Beyond Scala
Chapter 8 Testing & Validation
Chapter 9 Spark Components and Packages
Appendix Spark Tuning and Cluster Sizing

下载地址

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

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: **【】**（**）**