《Apache Spark源码剖析》

图书基本信息

书名：《Apache Spark源码剖析》

13位ISBN编号：9787121254204

出版时间：2015-3

作者：许鹏

页数：304

《Apache Spark源码剖析》

内容概要

《Apache Spark源码剖析》以Spark 1.02版本源码为切入点，着力于探寻Spark所要解决的主要问题及其解决办法，通过一系列精心设计的小实验来分析每一步背后的处理逻辑。
《Apache Spark源码剖析》第3~5章详细介绍了Spark Core中作业的提交与执行，对容错处理也进行了详细分析，有助读者深刻把握Spark实现机理。第6~9章对Spark Lib库进行了初步的探索。在对源码有了一定的分析之后，读者可尽快掌握Spark技术。
《Apache Spark源码剖析》对于Spark应用开发人员及Spark集群管理人员都有极好的学习价值；对于那些想从源码学习而又不知如何入手的读者，也不失为一种借鉴。

《Apache Spark源码剖析》

作者简介

许鹏长期致力于电信领域和互联网的软件研发，在数据处理方面积累了大量经验，对系统的可扩展性、可靠性方面进行过深入学习和研究。因此，累积了大量的源码阅读和分析的技巧与方法。目前在杭州同盾科技担任大数据平台架构师一职。对于Linux内核，作者也曾进行过深入的分析。

《Apache Spark源码剖析》

书籍目录

《Apache Spark源码剖析》

精彩短评

1、代码贴太多，没看代码的我很快就过了一遍
2、有所收获，读源码的方式值得学习。篇幅比较短，贴了比较多的代码，有些地方解释的不够。domain相关的几个lib的细节等到做过一些实际应用后再对照或许会有更好的理解
3、贴代码太多，图例和讲解太少，但好在基本讲清楚，后面对Streaming、SQL、GraphX、Mllib基本讲解也不错。（2015.6.2jd）
4、我错了我再也不买国人写的书了
5、有几个问题：一，图太少，流程图类图这些几乎没有，描述很无力；二，贴代码不写明在哪个文件内，很难结合书和代码一起看；三，版本略久，和最新版代码差别较大
6、我的意见是很一般般！讲基本原理的部分没有讲太清楚，含含糊糊的，讲底层的时候全是大段的代码，让人云里雾里。
7、支持的是1.0版本，有点老，配合官方文档一起食用
8、有所收获，但是这本书的讲解确实太粗略了。我觉得如果三四百页都用来讲解spark-core差不多。简单地列出代码加上些许讲解，这能称为源码剖析？

版权说明