《Spark大数据处理：技术、应用与性》

图书基本信息

书名：《Spark大数据处理：技术、应用与性能优化》

13位ISBN编号：9787111483863

出版时间：2014-11

作者：高彦杰

页数：268

《Spark大数据处理：技术、应用与性》

内容概要

《Spark大数据处理：技术、应用与性能优化》根据最新技术版本，系统、全面、详细讲解Spark的各项功能使用、原理机制、技术细节、应用方法、性能优化，以及BDAS生态系统的相关技术。
作为一个基于内存计算的大数据并行计算框架，Spark不仅很好地解决了数据的实时处理问题，而且保证了高容错性和高可伸缩性。具体来讲，它有如下优势：
打造全栈多计算范式的高效数据流水线
轻量级快速处理
易于使用，支持多语言
与HDFS等存储层兼容
社区活跃度高
……
Spark已经在全球范围内广泛使用，无论是Intel、Yahoo!、Twitter、阿里巴巴、百度、腾讯等国际互联网巨头，还是一些尚处于成长期的小公司，都在使用Spark。本书作者结合自己在微软和IBM实践Spark的经历和经验，编写了这本书。站着初学者的角度，不仅系统、全面地讲解了Spark的各项功能及其使用方法，而且较深入地探讨了Spark的工作机制、运行原理以及BDAS生态系统中的其他技术，同时还有一些可供操作的案例，能让没有经验的读者迅速掌握Spark。更为重要的是，本书还对Spark的性能优化进行了探讨。

《Spark大数据处理：技术、应用与性》

作者简介

高彦杰　毕业于中国人民大学，就职于IBM，精通Hadoop相关技术，较早接触并使用Spark，对Spark应用开发、Spark系统的运维和测试比较熟悉，深度阅读了Spark的源代码，了解Spark的运行机制，擅长Spark的查询优化。

《Spark大数据处理：技术、应用与性》

书籍目录

《Spark大数据处理：技术、应用与性》

精彩短评

1、计算机科学
2、2016 NO.7 还是有不少收获的，要是用 Java 代码就好了
3、章节的安排，先难后易，抛出一大堆概念和原理。。有点谭浩强，太不注重实践了
4、朋友的书，支持一下~
5、真的是很差,配图和文字都有诸多错误.
6、毕业纪念书籍
7、分布式系统
8、错误有点多 居然没有找到网上勘误
9、第三，四两章不错，包含rdd操作还有spark的工作机制。
10、初学者看一看还是蛮有帮助，不过还是看官方手册来的靠谱些。
11、可以入门，但是编辑的细节错误不少
12、这本书里边有比较多的细节错误或者描述不清。但是总体还行，读过之后可以对spark的某些方面的实现和优化，有一定的理解。
13、比王家林什么的好太多了。
14、书的内容还不错，就是内容顺序不太好，总感觉读起来不顺畅，有些章节顺序对调下，可能会比较好
15、适合初学者学习Spark，讲的都比较浅显。
16、很全面，也很技术。
17、内容很少，且大多泛泛而谈，很多东西都是网上能够找到的东西，不推荐
18、理论还要联系实践呀
19、浏览了第4章 ...
20、内容还不错，整体上把Spark的运行逻辑和注意点都理清了，但不少地方有些粗糙。
21、了解spark系统原理的较好的入门书
22、作为新手入门来说还不错，内容挺浅的。

《Spark大数据处理：技术、应用与性》

精彩书评

1、有误，比如join，spark中的join是inner join，书中对着源码讲成了outer join..跑题，花了大量篇幅在FIFO，HASHMap的原理上...以上不过很多地方还是很详细的，而且通俗易懂

《Spark大数据处理：技术、应用与性》

版权说明