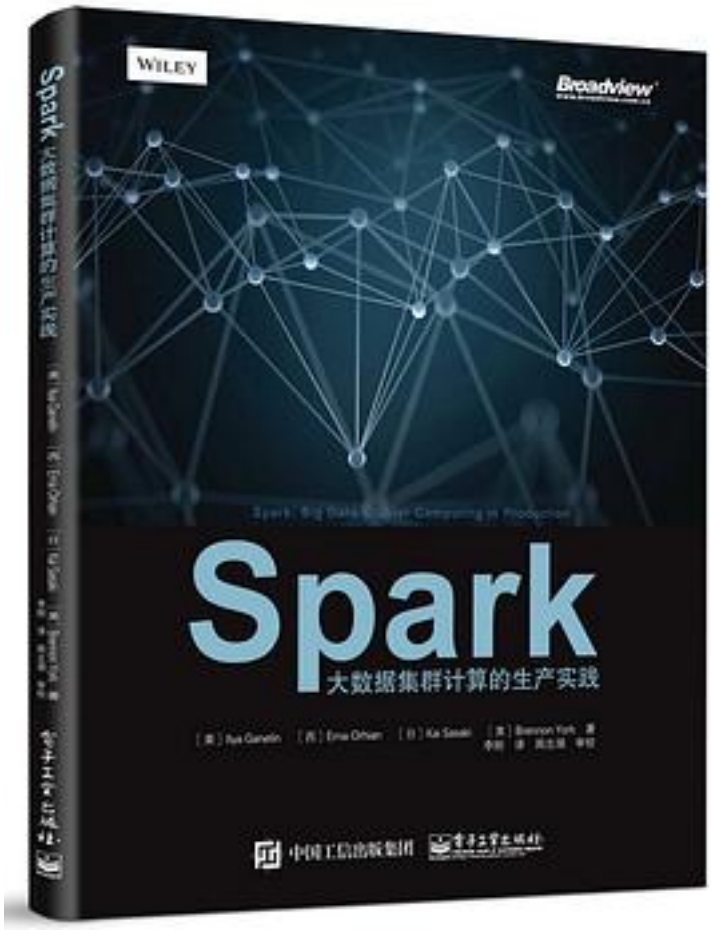
Spark：大数据集群计算的生产实践



Spark：大数据集群计算的生产实践_下载链接1_

著者:【美】Ilya Ganelin

出版者:电子工业出版社

出版时间:2017-5

装帧:平装

isbn:9787121313646

《Spark：大数据集群计算的生产实践》涵盖了开发及维护生产级Spark应用的各种方法、组件与有用实践。全书分为6章，第1～
2章帮助读者深入理解Spark的内部机制以及它们在生产流程中的含义；第3章和第5章

阐述了针对配置参数的法则和权衡方案，用来调优Spark，改善性能，获得高可用性和容错性；第4章专门讨论Spark应用中的安全问题；第6章则全面介绍生产流，以及把一个应用迁移到一个生产工作流中时所需要的各种组件，同时对Spark生态系统进行了梳理。

《Spark：大数据集群计算的生产实践》不会讲述入门级内容，读者在阅读前应已具备Spark基本原理的知识。《Spark：大数据集群计算的生产实践》适合Spark开发人员、Spark应用的项目经理，以及那些考虑将开发的Spark应用程序迁移到生产环境的系统管理员（或者DevOps）阅读。

作者介绍:

Ilya Ganelin
从机器人专家成功跨界成为一名数据工程师。他曾在密歇根大学花费数年时间研究自发现机器人（self-discovering
robot），在波音公司从事手机及无线嵌入式DSP（数据信号处理）软件开发项目，随后加入Capital One 的数据创新实验室，由此进入大数据领域。Ilya是Apache
Spark核心组件的活跃贡献者以及Apache
Apex的提交者（committer），他希望研究构建下一代分布式计算平台。同时，Ilya还是一个狂热的面包烘焙师、厨师、赛车手和滑雪爱好者。

Ema Orhian
是一位对伸缩性算法充满激情的大数据工程师。她活跃于大数据社区，组织会议，在会上发表演讲，积极投身于开源项目。她是jaws-spark-sql-rest（SparkSQL数据仓库上的一种资源管理器）的主要提交者。Ema一直致力于将大数据分析引入医疗领域，开发一个对大型数据集计算统计指标的端到端的管道。

Kai Sasaki
是一位日本软件工程师，对分布式计算和机器学习很感兴趣。但是一开始他并未从事Hadoop或Spark相关的工作，他最初的兴趣是中间件以及提供这些服务的基础技术，是互联网驱使他转向大数据技术领域。Kai一直是Spark的贡献者，开发了不少MLlib和ML库。如今，他正尝试研究将机器学习和大数据结合起来。他相信Spark在大数据时代的人工智能领域也将扮演重要角色。他的GitHub地址为：https://github.com/Lewuathe
。

Brennon
York既是一名特技飞行员，也是一位计算机科学家。他的爱好是分布式计算、可扩展架构以及编程语言。自2014年以来，他就是Apache
Spark的核心贡献者，目标是通过发展GraphX和核心编译环境，培育一个更强大的Spark社区，激发更多合作。从为Spark提交贡献开始，York就一直在用Spark，而且从那个时候开始，就使用Spark将应用带入生产环境。

目录:

Spark：大数据集群计算的生产实践_下载链接1_

标签

大数据

Spark

DEV

评论

spark版本比较老，2017年出版的书，还是1.5; 内容都比较浅，偏向介绍

这是目前看过的对Spark介绍最完整的书，有使用技巧，还有原理分析和调优，非常值得推荐

Spark：大数据集群计算的生产实践_下载链接1_

书评

Spark：大数据集群计算的生产实践_下载链接1_