

## [转发微博有机会获取《Spark大数据分析实战》](#)

即日起，关注@Spark技术博客 及@ 一位微博好友并转发本文章到微博有机会获取《Spark大数据分析实战》：

archives/1590。3月12日在微

博抽奖平台抽取1位同学并赠送此书。 **本活动已经结束，抽奖信息已经在[新浪微博抽奖平台公布](#)**



《Spark大数据分析实战》由高彦杰和倪亚宇编写，通过典型数据分析应用场景、算法与系统架构，结合6大案例，全面、深入讲解Spark大数据分析的各种技术和方法。通过这本书，你将了解到：

- (1)、Spark的机制与生态系统
- (2)、Spark算法分析与应用
- (3)、典型的数据分析场景
- (4)、Spark在日志分析、情感分析、协同过滤、个性化推荐、社交网络分析、搜索、文本分析等场景的应用

## (5)、Spark与其他开源工具及系统的集成

本书作者结合实战经验，编写了这本书。本书抽象出典型数据分析应用场景，适当简化，只进行主干介绍，讲解Spark应用开发、数据分析算法，并结合其他开源系统与组件，勾勒出数据分析流水线全貌。

Spark初学者Spark应用开发人员Spark机器学习爱好者开源软件爱好者其他对大数据技术感兴趣的人员如何阅读本书本书分为11章内容：

第1章 从Spark概念出发，介绍Spark的来龙去脉，阐述Spark机制与如何进行Spark编程。

第2章 详细介绍Spark的开发环境配置。

第3章 详细介绍Spark生态系统重要组件Spark SQL、Spark Streaming、GraphX、MLlib的实现机制，为后续使用奠定基础。

第4章 详细介绍如何通过Flume、Kafka、Spark Streaming、HDFS、Flask等开源工具构建实时与离线数据分析流水线。

第5章 从实际出发，详细介绍如何在Azure云平台，通过Node.js、Azure Queue、Azure Table、Spark Streaming、MLlib等组件对用户行为数据进行分析与推荐。

第6章 详细介绍如何通过Twitter API、Spark SQL、Spark Streaming、Cassandra、D3等组件对Twitter进行情感分析与统计分析。

第7章 详细介绍如何通过Scrapy、Kafka、MongoDB、Spark、Spark Streaming、Elastic Search等组件对新闻进行抓取、分析、热点新闻聚类等挖掘工作。

第8章 详细介绍了协同过滤概念和模型，讲解了如何在Spark中实现基于Item-based、User-based和Model-based协同过滤算法的推荐系统。

第9章 详细介绍了社交网络分析的基本概念和经典算法，以及如何利用Spark实现这些经典算法，用于真实网络的分析。

第10章 详细介绍了主题分析模型（LDA），讲解如何在Spark中实现LDA算法，并且对真实的新闻数据进行分析。

第11章 详细介绍了搜索引擎的基本原理，以及其中用到的核心搜索排序相关算法——Page Rank和Ranking SVM，并讲解了如何在Spark中实现PageRank和Ranking SVM算法，以及如何对真实的Web数据进行分析。

**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: [【】（）](#)**