Broadview®
www.broadview.com.cn

# Spark 大数据处理技术

夏俊鸾 刘旭晖 邵赛赛
程 浩 史鸣飞 黄 洁 著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

# Spark 大数据处理技术

夏俊鸾　刘旭晖　邵赛赛
程　浩　史鸣飞　黄　洁　著

電子工業出版社
Publishing House of Electronics Industry

## 内 容 简 介

本书以Spark 0.9版本为基础进行编写，是一本全面介绍Spark及Spark生态圈相关技术的书籍，是国内首本深入介绍Spark原理和架构的技术书籍。主要内容有Spark基础功能介绍及内部重要模块分析，包括部署模式、调度框架、存储管理以及应用监控；同时也详细介绍了Spark生态圈中其他的软件和模块，包括SQL处理引擎Shark和Spark SQL、流式处理引擎Spark Streaming、图计算框架Graphx以及分布式内存文件系统Tachyon。本书从概念和原理上对Spark核心框架和生态圈做了详细的解读，并对Spark的应用现状和未来发展做了一定的介绍，旨在为大数据从业人员和Spark爱好者提供一个更深入学习的平台。

本书适合任何大数据、Spark领域的从业人员阅读，同时也为架构师、软件开发工程师和大数据爱好者展现了一个现代大数据框架的架构原理和实现细节。相信通过学习本书，读者能够熟悉和掌握Spark这一当前流行的大数据框架，并将其投入到生产实践中去。



**图书在版编目（CIP）数据**

Spark大数据处理技术 / 夏俊鸾等著. —北京：电子工业出版社，2015.1
ISBN 978-7-121-25081-1

Ⅰ. ①S… Ⅱ. ①夏… Ⅲ. ①数据处理软件 Ⅳ. ①TP274

中国版本图书馆CIP数据核字（2014）第288415号

策划编辑：张春雨
责任编辑：徐津平
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路173信箱　　邮编：100036
开　　本：787×980　1/16　　印张：22.25　　字数：346千字
版　　次：2015年1月第1版
印　　次：2015年1月第1次印刷
定　　价：65.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

服务热线：（010）88258888。

# 序

2009 年的时候，Netflix 公司举办了一个叫作 Netflix Prize 的推荐算法比赛。这个比赛匿名公布了 Netflix 五十万用户对近两万部电影的一亿个评分数据，希望参赛者能够开发出更好的推荐算法，以提高推荐系统的质量。这个比赛的奖金有一百万美元。一百万美元看似很多，但是和一个更好的推荐算法给 Netflix 带来的效益相比，实则九牛一毛。

高昂的奖金和 Netflix 提供的真实数据吸引了不少的参赛者，其中也包括了来自加州大学伯克利分校（UC Berkeley）的博士生 Lester Mackey。Lester 师从机器学习领域泰斗 Michael Jordan，在一个叫作 AMPLab 的大数据实验室里进行博士研究。AMPLab 和大多数学术界实验室不同的地方在于实验室内有多个教授和他们带领的学生一起合作。这些研究人员来自不同的领域，包括机器学习、数据库、计算机网络、分布式系统等。当时，要想提高算法研究迭代的效率，需要利用多台机器的分布式建模。在尝试了当时业界最流行的 Hadoop MapReduce 后，Lester 发现自己的时间并不是花在提高算法效率上，而是耗费在 MapReduce 的编程模型和低效的执行模式上。这个时候，他向实验室内部的另外一名进行分布式系统研究的学生 Matei Zaharia 求助。

当时年纪轻轻的 Matei 在业界已经小有名望。他在雅虎和 Facebook 实习期间做了很多 Hadoop 早期的奠基工作，包括现今 Hadoop 系统内应用最广的 fair scheduler 调度算法。在和 Lester 的思维碰撞中，Matei 总结了 Hadoop MR 的不足，开始设计了第一个版本的 Spark。这个版本完全为了 Lester 定制，只有几百行的代码，使得 Lester 可以高效率地进行分布式机器学习建模。

Lester 所在的 The Ensemble 团队最后和 BellKor's Pragmatic Chaos 设计了在效率上并列第一的算法，可惜因为晚了 20 分钟提交，与一百万美元奖金失之交臂。5 年之后，Lester 和 Matei 都变成了学术界和业界杰出的人物。Lester 成为了斯坦福大学计算机系的教授，带领着自己的学生攻克一个又一个机器学习和统计的难题。Matei 成为了麻省理工计算机系的教授，也是 Databricks 公司的 CTO。

2009 年之后的 4 年里面，AMPLab 以 Spark 为基础展开了很多不同的学术研究项目，其中包括了我参与和主导的 Shark 和 GraphX，还有 Spark Streaming、MLlib 等。4 年里随着 Hadoop 的发展，Spark 也逐渐从一个纯学术研究项目发展到了开始有业界敢于吃螃蟹的用户。

2013 年，包括 Matei 和我在内的 Spark 核心人员共同创立了 Databricks 公司，立志于提高 Spark 的发展速度。过去两年，Spark 的发展超越了我们所有人的想象。一年半以前 Spark 还是一个连监控界面都不存在的系统，很难放进生产线部署。而一年半后的今天，它已经变成了整个大数据生态圈和 Apache Software Foundation 内最活跃的项目，活跃程度远远超出了曾经 Spark 只能望其项背的 Hadoop。

在从 Hadoop 转向 Spark 的道路上，我个人感觉国内的速度甚至超越了国外的社区。一年以前我第一次在中国的大数据会议上宣讲 Spark，当时台下的大多数人对这个新的项目还有很大的质疑，认为其只会昙花一现。一年之后，Spark 的每个新版本中都有不少华人贡献的代码，国内很多高科技和互联网公司也都有了 Spark 的生产作业，不少用户直接减少了在 Hadoop MapReduce 上的投资，把新的项目都转移到了 Spark 上。

今天正好是 Databricks 公司成立一年半，也是 Spark 1.2 版本第一个 release candidate 发布的日期。Spark 的高速发展导致了中文信息的脱节。这本书深入浅出地介绍了 Spark 和 Spark 上多个重要计算框架，希望它的问世可以更好地在大中华地区普及 Spark，增进华人 Spark 社区的发展。

辛湜 Reynold Xin
2014 年 11 月 30 号
Berkeley, CA

# 前言

"Use of MapReduce engine for Big Data projects will decline, replaced by Apache Spark."

——Hadoop 之父 Doug Cutting

受到 Google 两篇经典论文（*GFS* 和 *MapReduce*）的启发，Hadoop 项目诞生于 2005 年。起初，Hadoop 只是用来支撑 Nutch 搜索引擎项目。2006 年后 Hadoop 脱离了 Nutch，且变为 Apache 的顶级项目，并在 2008 年打破 TeraSort 的世界纪录。从此 Hadoop 变为大数据处理的事实标准，在工业界和学术界得到蓬勃发展。

随着数据规模的不断扩大，以及使用场景的不断丰富，用户对于大数据处理系统的要求也越来越高。而 Hadoop 生态中的数据处理引擎 MapReduce，则越来越不能满足用户的需求。在这样的背景下，Spark 于 2009 年诞生于 AMPLab 实验室。一经推出，其精妙的数据集抽象、数据重用机制，以及人性化的用户接口，都给人们留下深刻印象。尤其是 Spark 特别擅长的迭代式计算，相较于 MapReduce，性能有上百倍的提升。

2013 年成为 Apache 顶级项目后，Spark 基于自身的核心 API，发展出适应大数据处理的多种场景生态组件，包括 Shark/Spark SQL、Spark Streaming、Spark GraphX、

Spark MLlib 等，极大地满足了用户的需求。Spark 生态使得构建端到端的大数据应用成为可能，在处理各种场景时，提供给用户统一的编程体验，可极大提高编程效率，所以 Spark 被称为大数据处理的“瑞士军刀”。

Spark 目前在 Github 上安家落户，它的蓬勃发展得益于开发者的不断壮大。目前已经有来自 50 多家机构的近 400 位开发者贡献代码，使得 Spark 社区成为目前最为活跃的大数据处理开发社区。俗话说“工欲善其事，必先利其器”，当今世界已经迎来大数据的时代，如果想要从大数据中挖掘“金矿”，那么作为“工具”的大数据处理框架，在很大程度上决定了挖掘“金矿”的成败和效果。所以，我们有理由相信 Spark 拥有一个美好的未来!

## 本书目的

2014 年初，Spark 在国内已经发展得如火如荼。但是放眼全球还没有一本系统介绍 Spark 生态系统的书籍，所以项目组的小伙伴们决定写一本关于 Spark 生态的书籍。由于 Spark 社区的发展太快，所以我们避免将本书写成 Spark 的源码分析。而是对 Spark 中不变的思想进行深入剖析，使得读者能够更加透彻地了解 Spark 的运行机制，从而能更容易地写出高效的 Spark 应用程序。

## 内容速览

本书以 Spark 0.9.1 为蓝本，在后续写作中加入了部分 Spark 1.0 和 1.1 的内容。全书一共分为十二章，主要的章节内容如下。

**第 1 章　Spark 系统概述：**简要介绍当前大数据处理框架面临的挑战和 Spark 相应的解决方案。

**第 2 章　Spark RDD 及编程接口：**介绍了 Spark 中非常重要的概念 RDD，以及基于 RDD 的相关操作。

**第 3 章　Spark 运行模式及原理：**深入阐述了 Spark 的各种运行模式的部署和使用方法，介绍了各种模式的内部实现原理以及环境参数配置、文件依赖、序列化、权限控制等各种实现细节。

**第 4 章　Spark 调度管理原理：**分析了支撑 Spark 程序运转的核心调度管理逻辑，以作业调度模块为中心，介绍整个 Spark 调度管理系统的原理。

**第 5 章　Spark 存储管理模块：**阐述了 Spark 的存储管理模块的整体架构和实现细节，与 RDD 之间的映射关系，以及 shuffle 数据是如何管理以及运作的。

**第 6 章　Spark 监控管理：**介绍了目前 Spark 提供的两种监控管理方式，包括 Spark 应用程序的监控 UI 以及 Spark Metrics 系统的架构、原理和使用。

**第 7 章　Shark 架构与安装配置：**介绍了 Shark 的架构、安装、使用，以及数据缓存表等概念。同时，对 Shark 使用过程中的一些常见问题进行了分析。

**第 8 章　Shark 程序开发与扩展：**介绍了应用程序如何通过 API 访问 Shark 查询结果，如何通过编写自定义函数、自定义 SerDe 以及 StorageHandler 来扩展 Shark。

**第 9 章　Spark SQL：**围绕 SparkSql 的 Catalyst 模块，从软件架构和用户 API 两个角度来阐述，抽丝剥茧地介绍 Spark SQL 的全貌。

**第 10 章　Spark Streaming 流数据处理框架：**介绍了 Spark Streaming 的使用、调优以及内部原理，使读者对 Spark Streaming 有一个感性的认识。

**第 11 章　GraphX 计算框架：**介绍了 Spark GraphX 的设计、实现及应用，深入分析了 GraphX 中图的存储方式、执行策略，以及执行优化的实现和原理。

**第 12 章　Tachyon：**介绍了基于内存的分布式文件系统 Tachyon 的原理和实现，

包括整体框架的设计、部署方式以及在 Spark 中的应用等。

## 写作分工

第 1、2 章由夏俊鸾编写；第 3、4 章由刘旭晖编写；第 5、6、10 章由邵赛赛编写；第 7、8、9 章由程浩编写；第 11 章、12 章分别由史鸣飞、黄洁编写。

## 致谢

感谢英特尔亚太研发有限公司大数据部门首席架构师戴金权先生，他将整个大数据部门带入了 Spark 的世界，使其成为 Spark 社区最重要的贡献力量，然后部门的小伙伴们才有了写这本书的原始动机。

感谢英特尔亚太研发有限公司段建刚先生，在整本书籍的撰写过程中，得到了他的很多支持及建议。

感谢电子工业出版社的张春雨先生和编辑贾莉，他们严谨认真的工作态度为这本书的成功出版奠定了坚实的基础。

感谢为本书撰写序言的辛湜先生，以及撰写评论的曾宪杰先生、连城先生、黄明先生，他们在百忙之中阅读了书籍的样稿并提出了很多中肯的建议。

# 目　录