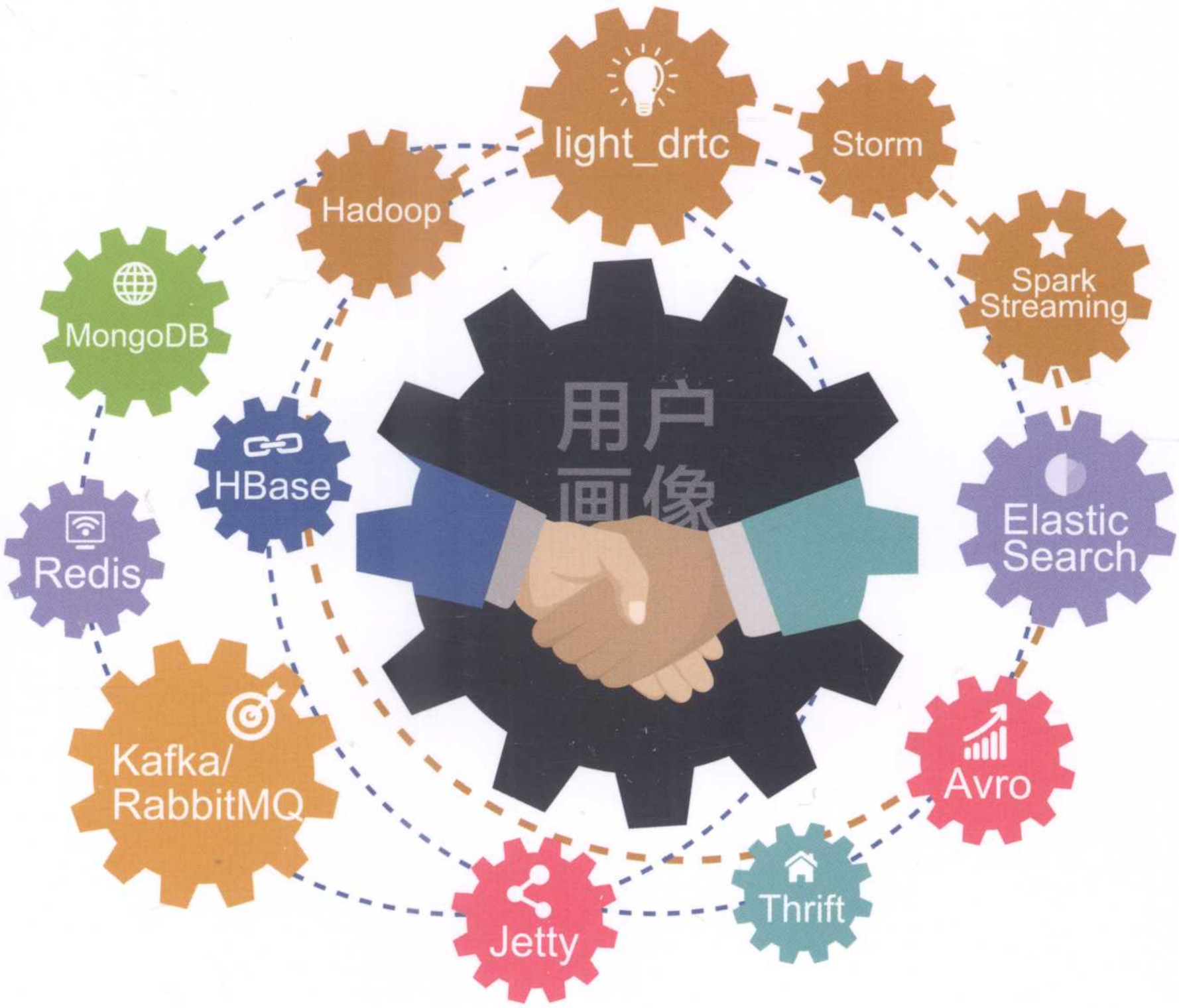
Broadview®
www.broadview.com.cn

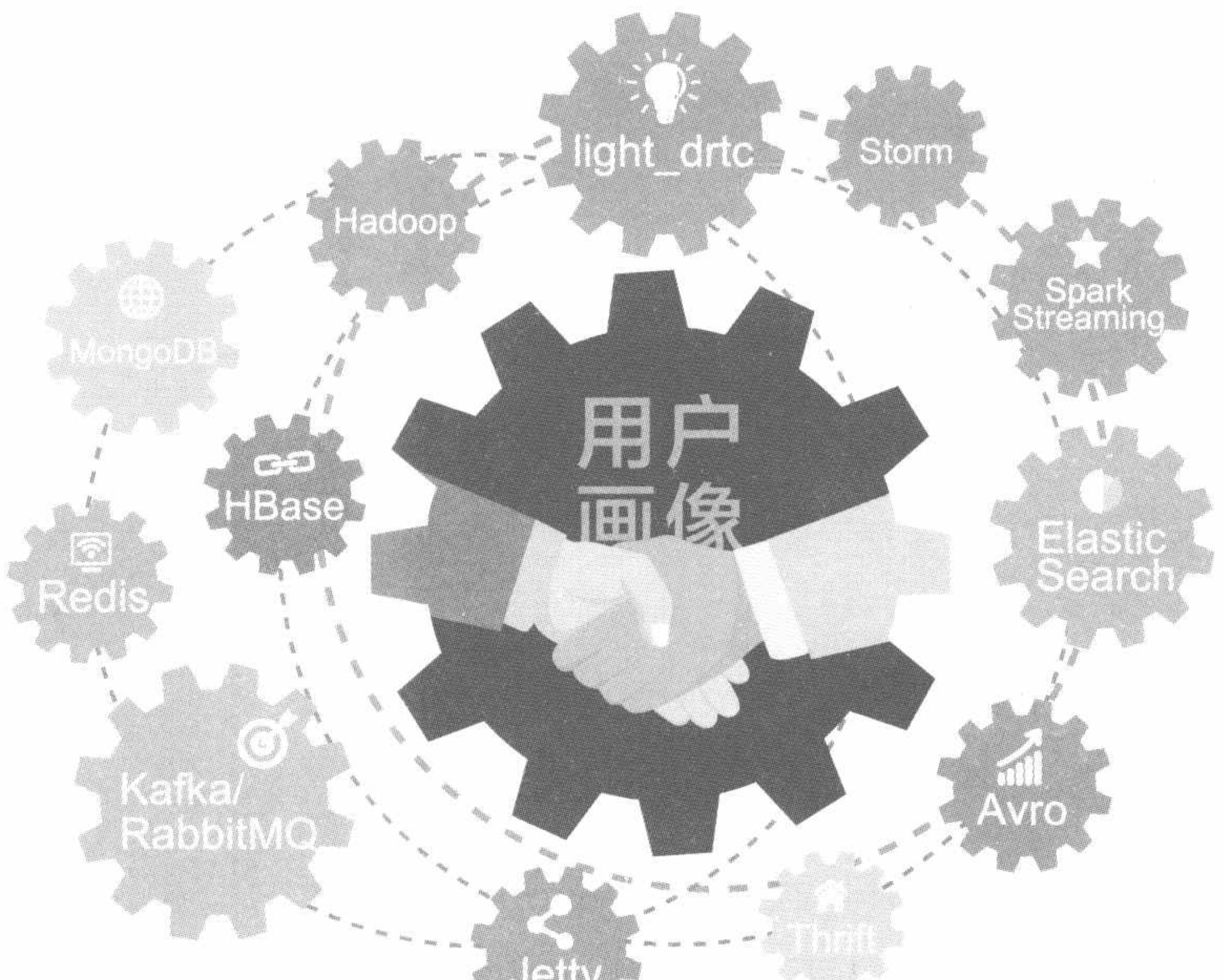
开源作者自研轻量级分布式实时计算服务框架——light_drtc，
简单易用，快速实现自定义的实时计算平台，快速实现企业所需计算实时性要求比较高的业务逻辑
学通本书，做大数据时代的“心灵捕手”！



分布式实时计算框架原理及实践案例

王成光 著

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn



分布式实时计算框架原理及实践案例

王成光 著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

“授人以鱼不如授人以渔”，本书是作者以此初心写成的，主要参考当前主流分布式实时计算框架 Storm 的任务分发和 Spark Streaming 的 Mini-Batch 设计思想，以及底层实现技术，开源了作者自研的轻量级分布式实时计算框架——Light_drtc，并且重点介绍设计思想和相关实现技术（Kafka/RabbitMQ、Redis/SSDB、GuavaCache、MongoDB、HBase、ES / Solr、Thrift、Avro、Jetty），最后从工程角度向大家介绍完整的个性化推荐系统，并实例介绍 light_drtc 在用户画像实时更新的应用。本书描述浅显易懂，希望读者理解分布式实时计算的实现原理，并快速上手解决实际问题。

本书适合读者包括：高校师生及从事软件开发的中高级工程师、架构师及技术管理者等。

图书在版编目（CIP）数据

分布式实时计算框架原理及实践案例 / 王成光著. —北京：电子工业出版社，2016.9
ISBN 978-7-121-29620-8

Ⅰ. ①分… Ⅱ. ①王… Ⅲ. ①数据处理软件 Ⅳ.①TP274

中国版本图书馆 CIP 数据核字（2016）第 182107 号

责任编辑：孙学瑛
印　　刷：中国电影出版社印刷厂
装　　订：三河市华成印务有限公司
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：720×1000　1/16　印张：18.5　字数：280 千字
版　　次：2016 年 9 月第 1 版
印　　次：2016 年 9 月第 1 次印刷
定　　价：79.00 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）88254888，88258888。

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。

本书咨询联系方式：（010）51260888-819，faq@phei.com.cn。

专家力荐

互联网连续创业者　陈超仁（原基调网络创始人、前美丽说高级副总裁）

互联网深深渗入各个行业，消费级智能设备和 IoT 爆炸式增长，联网设备数量早已突破百亿，很快将达到千亿量级。伴随而来的是计算能力需求剧增、数据呈指数级积累，以及支撑架构设计和实施能力的巨大考验。企业的竞争力将高度等同于拥有的计算及数据处理能力。作为最风口浪尖的设计及实施者，软件工程师在现今变革的时代面临的机遇和挑战前所未有。

本书作者从自身实战经验总结并研发的分布式计算框架入手，对主流支撑框架做了高度简练的介绍，并集合框架对时下热门且实用的用户画像分析这一难题进行剖析，深入浅出，将晦涩难懂的概念通过案例完全展示在读者面前，让海量数据的价值在计算魔术中逐渐显现。本书实为难得的架构实战干货，相信可以帮助工程师及架构师避免踩坑，完成进阶。

在我互联网及软件技术行业近二十年的创业经历里，在技术道路上极度追求不断突破的极客不少见，成光给我的形象却犹为深刻。记得初次见面时间不长却一见如故，交谈甚欢。这些工程师是互联网时代最精华的组成部分，他们是互联网生产力的缔造者，也一定是成功者。

阿里巴巴技术专家　赵文旭

对于解决实际问题来说，一线的经验参考最为重要。在实际实现一个业务系

统、解决一个业务问题的时候，很少会仅使用一项单一技术就可以解决，往往需要多项技术相结合使用。

在大数据处理领域，Hadoop、Storm、Spark 等核心技术是必不可少的。但是要想构建一个完整的解决方案，还需要 RPC、消息队列、缓存系统、数据库系统，一个都不能少。以上每一项技术的书籍及网上的资料不可谓不多，也不缺乏精品。但从业务解决方案入手，综合多项技术，有针对性地介绍、分析、总结的书籍却很缺乏。

本书作者王成光同学，多年来一直奋战在大数据处理领域的开发一线，具有丰富的实战经验。他曾经从无到有一手搭建了优购网（百丽电商）的搜索推荐及 BI 系统，后曾在网易等一线互联网公司任职。他在解决业务需求的基础上，逐渐沉淀提炼出一套轻量级分布式实时计算服务框架——light_drtc。本书不仅对这个自研框架的架构原理做了介绍，同时分享了研发这套架构的心得体验。同时，对系统中应用到的各项技术进行了详细的、有针对性的介绍、对比及分析，不乏精准独到的个人见解，非常值得一读。

这本书是作者多年经验和智慧的结晶，细读它，你一定会有所收获。

TalkingData 研发副总裁　阎志涛

大数据技术发展到现在，实时流处理技术变得越来越重要。在这个大数据实时处理狂潮中，各种开源流计算框架也如雨后春笋般地涌现出来。比较常见的有 Storm、Samza、Spark Streaming、Flink，以及 Twitter 最新开源的 Heron 等等。比较遗憾的是如大部分开源技术一样，这些流计算框架都为国外公司所主导。国内虽然有 BAT 等互联网巨头，却没有在实时计算框架方面有对应自己地位的开源产品。令人稍感欣慰的是，light_drtc 作为纯粹国内技术人员主导开发的一个轻量级的流计算框架进入了开源世界，让实时计算开源框架中有了中国力量。

分布式实时计算技术对于大数据技术来讲非常重要，不过在技术上能讲解得非常透彻的中文技术书籍并不多。认识王成光时他正沉浸在开发 light_drtc 的状态中，关于分布式实时处理技术，我们聊了很多，有很多技术的见解也很一致。

令人非常高兴的是他不仅将自己实践积累的产品 light_drtc 开源出来，同时将自己多年技术积累的经验以写书的形式奉献给了广大对分布式实时计算技术有兴趣的技术人员。《分布式实时计算框架原理与实践案例》这本书不止是一本对当前实时流计算技术进行解析的一本书，同时也是作者对在工作中实战经验总结的一本书。授人以鱼不如授人以渔，我相信这本书能够给大数据技术人员，尤其是对大数据流处理技术有兴趣的人带来很大的帮助。

京东首席技术顾问　翁志

本书不仅详实地推介了作者自主研发的实时计算框架，还基本涵盖介绍了当今主流的开源计算系统。内容浅显易懂，案例切合实际，分析精准到位。对于年轻的 IT 读者来说，不失为良师益友。

中国建筑电商 CTO　邓威

本书是作者基于自己多年的思考和实践经验，经过不断提炼而得的。一方面把当前业界在实时计算领域常用的产品和技术进行深入介绍，让读者对该领域中的各个产品如何使用不再迷茫，也为具有一定经验的从业者对下一步系统中关键服务如何选型和优化提供了新的方向；另一方面作者博采众家之长而创造的 light_drtc，降低了中小企业在分布式实时计算领域的门槛，使大数据处理能力触手可得。最后通过作者的实例，在了解大数据技术如何应用的同时，也能了解到在实例背后作者所体现出的思路、方法和方案。希望通过本书能让更多的人了解大数据处理，有更多的企业挖掘出自身的数据价值。

腾讯技术副总监　鞠奇

成光对于架构的不断钻研和踏实肯干给我留下了很深的印象，非常有幸见证了他这套分布式实时计算系统在新闻推荐领域中的应用。这本书结合他多年的一线实践经验，详细阐述了分布式计算系统、搜索架构和数据库等在企业应用的经

历，对于初学者和想深入理解这一方面知识的同学会起到很好的引导。感谢成光为国内的工程架构贡献自己的一份力量！

大码美衣 CTO　王伟涛

作者硕士从事计算机应用中文信息处理的研究工作，毕业后在百丽、好乐买、搜狐和网易等担任搜索、推荐架构师等关键岗位。多年来一直从事大数据相关的开发和研究工作，从未停止对技术的了解和钻研。作者结合当前主流开源软件的特点，以及中小型企业人才和资源不足的困难，利用多年来的技术积累和沉淀，独力研发了一套轻量级分布式实时计算服务框架——Light_drtc，包含实时数据收集服务、资源协调及任务管理服务和任务计算服务，可以帮助企业快速搭建自定义的实时计算平台，让企业聚焦于业务数据的分析和处理。

本书系统介绍了 Light_drtc 的设计思想、功能,以及核心技术。同时深入浅出地介绍了当前主流开源大数据处理技术，如 Hadoop，Spark 及 Storm 等主流计算框架,以及消息队列、内存数据库、NoSQL、搜索、RPC 框架等核心技术架构，帮助开发者系统全面地了解大数据平台的核心技术。

作为中小型企业的技术负责人，对此书的作者表示感谢，作者提供了一套轻量级的解决方案，并介绍了相关的技术环节，让小企业可以专注于数据的应用，而不是花大量的时间用于搭建平台，切实提供了很大的帮助。

感谢作者花了大量精力对中小企业的支持以及对开源的支持，希望本书可以帮助更多的企业，也为国家大数据平台的发展贡献一份力量。

前　言

“授人以鱼不如授人以渔。”——语出《淮南子 · 说林训》，道理很简单，鱼是目的，钓鱼是手段，一条鱼能解一时之饥，却不能解长久之饥，如果想永远有鱼吃，那就要学会钓鱼的方法。

随着时代发展，互联网，尤其是移动互联网的全民普及趋势，任何一个行业的相关产品都会有很多表示用户兴趣点的行为数据，像用户的浏览、收藏、分享、购买、评论、点赞和搜索等行为由此构建出海量用户行为数据。如何快速有效地使用上述大数据，挖掘出用户对产品的兴趣点，实时更新用户画像，进而向用户推荐其当前最感兴趣的产品及广告，是目前众企业所普遍关心的问题，也是大数据的价值所在。

大数据是时代产物，除了 BAT、网易、搜狐、新浪、京东等一线互联网企业需要相关处理，国内更多中小企业或者传统行业的巨无霸也都需要大数据处理技术。一线互联网企业由于自己本身就做互联网业务拥有人才优势，处理大数据相对简单。但中小企业及传统行业巨无霸在面对自己日积月累的大数据时困难重重，主要是本身没有处理能力，也没有相关硬件设施支持。

目前大数据处理技术，以开源界大名鼎鼎的 Hadoop、Spark、Storm 及后起之秀 Flink 为代表，当然国内一线互联网企业像 BAT 也都有各自独特的处理技术，但由于国内大环境所致，商业公司在运作时首先考虑的是商业及保密措施，使得国内开源界在分布式处理方面相对薄弱，基本上是空白战场。

目前看来，Hadoop 比较适合用于离线数据处理，Spark 及 Storm 的实时处理

技术正好弥补了 Hadoop 实时处理的欠缺。虽然 Hadoop、Spark 及 Storm 都在快速发展，但国内真正深入理解这三者的人毕竟凤毛麟角，而且 Hadoop 集群本身动辄就需要上百台服务器的集群，这对中小企业来说基本上是不可能的，而且中小企业也很少能够拥有精通 Hadoop、Spark 及 Storm 的资深技术人员，这就导致绝大部分中小企业的大数据处理基本上处于停滞状态，它们眼睁睁地看着大量数据产生，而无法进行相应处理。即使众中小企业花费巨资招聘了 Storm、Spark Streaming 开发人员，由于自身平台的限制，开发者基本上只是简单地调用其封装好的 API，很难从源码跟踪到问题根源，也就造成生产环境下有很多问题难以解决。

时代在变，技术也在变，没有任何一项技术会解决所有问题，对企业及个人开发而言，适合自己的才是最好的。Hadoop3.x 以后将会调整方案架构，将 MapReduce 基于内存+IO+磁盘，共同处理数据，这点和当前的 Spark 很像；Storm1.0 也打破了之前一直存在的诟病：Nimbus 单点问题，实现了 HA，这点和阿里推出的 JStorm 很像，而 JStorm 本身也是源自 Storm；最近 Twitter 又推出 Heron，兼容 Storm。Storm 也提供了类似 Spark Streaming 的微批处理方式——trident（以一组 tuple 为单位）。上述开源项目之间都在互相学习对方优点，取其精华为己所用。

鉴于上述原因，本书作者经过多年的深入思考，结合自己硕士毕业 8 年多的一线互联网开发经验，根据 Hadoop 的 Map/Reduce 及当前主流实时计算框架 Storm 的任务分发和 Spark Streaming 的 Mini-Batch 处理思想，利用时下比较流行的 MQ、RPC、NoSQL 等，独力研发了一套**轻量级分布式实时计算服务框架——light_drtc**。其最大特点就是简单易用，它可以帮助开发者快速实现自定义的实时计算平台，其设计目的是为了降低当前大数据时代的分布式实时计算入门门槛，方便初、中级读者上手，快速实现企业所需计算实时性要求比较高的业务逻辑。它本身既可以作为独立的分布式实时计算平台，也可以以嵌入式方式，作为其他项目的基础类库存在。

Light_drtc 目前以 Java8 为基础设计和实现，框架主要包括三部分：实时数据收集服务（CollectNode，简称 CN）、资源协调及任务管理服务（AdminNode，简称 AN）和任务计算服务（JobNode，简称 JN）。这套框架扩展灵活，各个相关组件可以自由扩展（目前框架已整合 Kafka 和 RabbitMQ，物理分布上可以将 CN 和

AN 整合在一起），集群节点所完成计算所需要的开发语言不仅限于 Java，也可以用时下流行的各种开发语言。

对中小企业而言，利用 light_drtc 搭建分布式实时计算平台，不需要庞大的服务器集群规模，完全可以根据自己业务需要。例如抽取 9 台服务器：其中 3 台服务器同时兼做 CN 和 AN，6 台服务器做 JN，每个 AN 独立管理 2 个 JN，即可搭建自己的高可用分布式实时计算集群。在 light_drtc 框架中，每个 AN 都有自己独立管理的 JN，且每个 AN 至少独立管理 1 个 JN。框架使用比较方便，尤其是如果开发者也选用 Kafka 或 RabbitMQ，对于 CN 和 AN 而言，仅需要开发者自定义 MQ 相关配置及实现框架所定义的流数据解析接口，将实现类传递给框架 AN 节点启动类即可。对于 JN，则需要开发者按照自己业务需求自行实现，框架中有丰富的实例可供参考。

本书偏重工程架构方面，主要内容除了作者自研的 light_drtc 详细介绍，还会以作者多年一线互联网开发经验角度，陆续向读者介绍当前主流 MQ、NoSQL、全文检索 Elasticsearch/Solr，及常用微服务架构技术实现 RPC 和 Web Service 的多种框架，最后介绍整个新闻个性化推荐系统的各个组成部分，对核心模块用户画像实时更新做了详细设计，并对比 Storm、Spark Streaming 和 light-drtc 不同实现方式。作者希望读者通过阅读本书，让您对分布式实时计算系统的设计原理及相关实现技术有更加清晰的理解，也希望让众多中小企业可以快速组建自己的分布式实时计算平台，也同时为国内分布式处理技术贡献一点自己的力量。

最后给读者朋友分享一下个人多年学习的一点心得：万丈高楼平地起，任何一项别人看似游刃有余的技能都是经过时间打磨后的熟能生巧，希望读者朋友们经过自己的奋斗都可以实现人生目标，达到人生顶峰。

王成光

2016/8/5

目　录

第1章 分布式实时计算框架介绍

目前分布式计算框架以大名鼎鼎的 Hadoop Map/Reduce、Spark Streaming 和 Storm 为代表，可以说 Hadoop 奠定了最近 10 年来开源分布式计算框架的基石。

1.1 分布式计算 Hadoop

Hadoop 是原 Yahoo 的资深技术专家 Doug Cutting 根据 Google 发布的学术论文研究而来的。Hadoop 是一个能够对大量数据进行分布式处理的软件框架，它以一种可靠、高效、可伸缩的方式进行数据处理。

Hadoop2.0 框架最核心的设计就是：HDFS、MapReduce 和 Yarn。HDFS 是一个分布式文件系统，为海量数据提供了存储；MapReduce 是一个离线处理框架，由编程模型（新旧 API）、运行时环境（JobTracker 和 TaskTracker）和数据处理引擎（MapTask 和 ReduceTask）三部分组成，为海量数据提供了计算框架；Yarn 是一个框架管理器，为各种框架进行资源分配和提供运行时环境。MRv2 则是运行在 YARN 之上的第一个计算框架。

Hadoop 2.0 中对 HDFS 进行了改进，使 NameNode 可以横向扩展成多个，其中，每个 NameNode 分管一部分目录，这不仅增强了 HDFS 的扩展性，也使 HDFS

具备了隔离性。利用 Hadoop 进行分布计算的核心是如何优化分解一个大计算任务为若干独立子任务，从而并行运行。作者自研分布式计算框架在 AN 的资源管理和任务调度方面也是参考了这点。

Hadoop 是一个能够让用户轻松架构和使用的分布式计算平台。用户可以轻松地在 Hadoop 上开发和运行处理海量数据的应用程序。它主要有以下几个优点：

（1）高可靠性：Hadoop 按位存储和处理数据的能力值得人们信赖。

（2）高扩展性：Hadoop 是在可用的计算机集簇间分配数据并完成计算任务，这些集簇可以方便地扩展到数以千计的计算节点中。

（3）高效性：Hadoop 能够在节点之间动态地移动数据，并保证各个节点的动态平衡，因此处理速度非常快。

（4）高容错性：Hadoop 能够自动保存数据的多个副本，并且能够自动将失败的任务重新分配。

（5）低成本：与一体机、商用数据仓库及 QlikView、Yonghong Z-Suite 等数据集市相比，Hadoop 是开源的，项目的软件成本因此会大大降低。

即将推出的 Hadoop3.0 将会调整方案架构，将 Mapreduce 基于内存+IO+磁盘，共同处理数据，其实最大改变的是 HDFS，HDFS 通过最近 black 块计算，根据最近计算原则，本地 black 块，加入到内存，先计算，通过 IO，共享内存计算区域，最后快速形成计算结果。虽然性能改善让人很期待，不过距离使用还有一段时间，目前尚不能应用到实际生产环境。

诚然，Hadoop 在分布式计算框架方面的贡献无人可出其右，但它也不是万能的，Hadoop 适合离线批处理计算任务，而且为了凸显 Hadoop 集群优势，业内一般都是几百台甚至上千台服务器集群规模。对于中小企业而言，一般不会组建这么大的计算集群，反而更需要一个小而精的分布式计算平台。此外，Hadoop 一般只用来离线批处理计算，因为它必须将所有输入数据都处理完才返回最终计算结果，这对于实时性要求比较高的在线学习显然不合适。