
Broadview
www.broadview.com.cn
精通实时大数据分析
Druid
实时大数据分析
原理与实践
欧阳辰 刘麒赟 张海雷 高振源 等著
腾讯、小米、优酷、云测等互联网公司的一线实践经验
为你解读海量实时OLAP平台
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

Druid
实时大数据分析
原理与实践

欧阳辰 刘麒赟 张海雷 高振源 许哲 著

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

内容简介

Druid 作为一款开源的实时大数据分析软件，最近几年快速风靡全球互联网公司，特别是对于海量数据和实时性要求高的场景，包括广告数据分析、用户行为分析、数据统计分析、运维监控分析等，在腾讯、阿里、优酷、小米等公司都有大量成功应用的案例。本书的目的就是帮助技术人员更好地深入理解 Druid 技术、大数据分析技术选型、Druid 的安装和使用、高级特性的使用，也包括一些源代码的解析，以及一些常见问题的快速回答。

Druid 的生态系统正在不断扩大和成熟，Druid 也正在解决越来越多的业务场景。希望本书能帮助技术人员做出更好的技术选型，深度了解 Druid 的功能和原理，更好地解决大数据分析问题。本书适合大数据分析的从业人员、IT 人员、互联网从业者阅读。

图书在版编目（CIP）数据

Druid 实时大数据分析原理与实践 / 欧阳辰等著. —北京：电子工业出版社，2017.3
ISBN 978-7-121-30623-5
I. ① D…II. ① 欧…III. ① 数据处理 IV. ① TP274

中国版本图书馆 CIP 数据核字（2016）第 304239 号

策划编辑：符隆美
责任编辑：葛　娜
印　　刷：三河市良远印务有限公司
装　　订：三河市良远印务有限公司
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：787×980　1/16　　印张：22　　字数：478 千字
版　　次：2017 年 3 月第 1 版
印　　次：2017 年 3 月第 1 次印刷
印　　数：4000 册　　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。

Foreword

Like many popular open source projects, Druid was initially created to solve a problem. We were trying to build an interactive analytics UI at a small advertising technology startup in San Francisco, and struggled to find a technology that could rapidly aggregate, slice and dice, and drill down into massive data sets. Eric Tschetter started the first lines of Druid to tackle this challenge, and that work has somehow led to an international community forming around the project.

I joined Eric on Druid soon after the project started, and for a while, the Druid world consisted of only 2 engineers. The first version of Druid was extremely minimalistic; there was a single process type, the "compute" node, and a handful of queries, but the core that was there was just enough to solve the problems with scale and performance we had at that time.

Our Druid cluster in the early years was less than 20 nodes, and we worked around the clock to aggressively develop features and fix bugs. There were a lot of late nights in those days. I can still very clearly recall waking up in the middle of the night to fix an outage, and occasionally cursing loudly because the only reason the pager went off was because it was out of batteries.

As Druid matured, and as data volumes grew, we continued to face challenges around performance at scale and operational stability. Running in the then notoriously finicky Amazon Web Services cloud environment wasn't always easy, and led us to make the decision to break up "compute" nodes into different components so that individual components could be fine tuned at scale, and any one component could fail without impacting the functionality of the other components. I am glad we made those decisions because it led us to sleep much more at night.

It has been extremely rewarding to watch the grassroots growth of the open source community. Unlike other popular open source projects, Druid was not developed at a major technology company or famous research lab. We open sourced the project without much attention, and the first open source version of the project almost didn't have querying capabilities. We weren't allowed to open source

many pieces of the codebase, including most of the queries we developed. The night before officially announced the project, Eric was up writing GroupBy queries in a hotel room just so people could have a way of getting data out of Druid. After we released Druid, the code repository was completely undocumented and barely functional. I don't think a single organization tried to use Druid when it was first open sourced.

I've long lost count of how many companies actually run Druid in production today, but I am glad people have found value from our work. I was very excited to learn that Qiyun Liu was writing a book on Druid. I hope through his book, you will learn much more about our project, and learn how to leverage it to bring value to your organization.

Fangjin Yang
Co-Founder, Druid
Co-Founder and CEO, Imply
San Francisco, California
2016.11.20

序言

正如许多广为应用的开源项目，Druid 是为解决某个特定问题而诞生的。几年前，在旧金山的一家广告技术创业公司，我们想要创建一个交互式分析的用户界面，同时也在寻找可以快速聚合、切片并深探海量数据集的技术。为了解决这个技术难题，Eric Tschetter 开始了 Druid 项目的第一行代码。到目前为止，这个项目已经拥有了跨越多个国家的社区，并且在不断发展壮大。

在 Eric 开始了 Druid 项目后，我很快便加入了这个团队。说是团队，实际上很长一段时间只有我和 Eric 两个人。Druid 的最初版本极其简单：只有一种分析类型、几个“计算”节点，以及一些简陋的查询功能。但是 Druid 的这些核心功能却已经足以解决我们当时所面对的性能与规模的难题。

在早期时，Druid 集群只有不到 20 个节点。那时我们会日以继夜地工作，不断地开发新的功能与改正代码中的错误。在那些日子里，我们经常会工作到深夜，直到现在我还很清楚地记得自己在半夜爬起来修复瘫痪的系统。不过有时半夜被叫醒的原因也会让我们感到暗暗生气，因为系统告警的原因竟然是由于呼叫器的电池没电了。

随着 Druid 的逐渐成熟，以及数据量的持续增长，我们在集群规模和运营稳定性方面不断面临新的挑战。众所周知，当时亚马逊网络服务云环境的状况不是很好，甚至有点儿“臭名昭著”，因此在它上面运行集群其实并不容易，从而促使我们决定将“计算”节点分解成不同的独立组件，以便可以在大规模的集群上对单个组件进行微调，同时保证任何一个组件的失败不会影响到其他组件的正常运行。现在回头看，我很高兴我们当初做了这些决定。正因为有了这些高容错的功能，我们终于可以不用时不时半夜起床，从而可以多睡一会儿了。

亲眼见证了这个项目的成长过程让我非常欣慰。与其他大型的开源项目不同，Druid 不是在一家前沿技术公司或是享誉盛名的研究实验室开发的。这个项目刚开源的时候并没有受到很多关注，而且第一个开源版本甚至几乎没有查询功能。许多代码库，包括我们当时开发的大多数查询功能，由于没有得到当时公司的允许从而没能开源。在正式宣布这个项目的前

一天晚上，Eric 还在写 GroupBy 查询，以便大家可以从 Druid 得到数据。在我们发布 Druid 后，代码存储库是完全没有记录的，也几乎不能用。Druid 刚开源的时候也没有公司要尝试用它。

今日此时，我实际上已经没有办法准确地知道到底有多少公司在他们的生产中使用 Druid，但我非常高兴很多人从我们的工作中发掘到了价值。得知刘麒赟他们在写一本关于 Druid 的书我非常兴奋。我希望通过这本书，您将更深入地了解 Druid，并用它为您的组织创造价值。

杨仿今
Druid 项目主要创始人
Imply 公司联合创始人，CEO
美国 旧金山
2016 年 11 月 20 日

推荐序一

“Druid 是一套非常棒的大数据软件，而本书是一本非常棒的 Druid 课本。”

阅读完欧阳辰等人写的原稿，我很快做出这样的判断，更感叹大数据技术已经彻底迈入一个全新的爆发时代。

作为曾经服务于大数据技术的先驱公司 Google 的从业者，我个人认为大数据技术有着明显的三个历史发展阶段：

1. 探索时代

大家知道，“大数据”与“数据”的核心区别在于数据的完整性。在互联网行业还不成熟的时代，传统行业的数据主要来自于“采样”，数据集并不完备。对小企业来说，数据采集是高成本、高门槛的；即使是对于信息化程度已经很高的大公司，当时的技术也没法很快速地处理 TB 级别的数据量。

互联网业务特别是数字广告，从第一天开始就尝试解决数据采集的完备性（考虑到按点击收费，客户的微观广告数据必须精细采集），也创新性地研发出能够快速处理大数据的技术解决方案。SSTable、MapReduce 和 BigTable 等非常成功的实践解决方案在这些探索中诞生。当然，还有很多探索性的研发都失败了。

这些成功的新技术慢慢在互联网技术圈传播，随着 Apache Hadoop 框架的成功，大数据技术开始在行业普及。

2. 普及时代

随着社交通信、数字广告、电子商务、网络游戏等商业模式的发展，越来越多的互联网企业诞生。他们都享受了大数据基础技术的红利，从初始就具备比较强大的数据收集、分析和处理能力，并且可以用在业务优化上。

很显然，因为行业的多样性，业务场景变得越来越复杂，对数据处理的要求已经不仅是体量大和速度快，还要数据结构灵活、编程接口强大、系统可扩展、原子化操作、高效备份、读性能加速或者写性能加速等。在这个技术普及的时代，不仅互联网行业有越来越多的技术人员和数据人员开始参与到大数据工作中，而且很多传统软件从业者也慢慢受到吸引，双方互相借鉴，进一步扩大了大数据技术的能力和影响。可以看到，传统的数据库、操作系统、编程语言等技术思想被引入来解决各种复杂的需求。因此而诞生的包括 NoSQL、SQL on Hadoop、ElasticSearch 这样的新事物，逐渐把我们推进到一个全新的时代。

3. 创新时代

本书所介绍的 Druid，是大数据技术新时代的产物。现在的新技术，并不只是解决各种技术问题，而是更加贴近复杂的创新型业务的需求场景。我们看到，业内的新框架和新产品，都在探索如何让大数据能为各种不同类型的业务带来更多的优化，解决数据可用性、垂直性、实时性、灵活性、可视化等问题。

如本书所介绍的，Druid 以及相关配套的工作，使我们可以非常灵活地实时分析数据，做复杂的维度切割和条件查询，而且可以非常方便地做可视化展示。无论是在互联网企业，还是传统企业，这个工具的使用场景都是非常丰富的，如监控报警、诊断排错、生成业务报表、对接机器学习及策略优化等。

在这个创新时代，还有很多新技术涌现出来，比如强调可编程与实时性的 Spark、与 Druid 类似的 Pinot，还有 A/B 测试（比如我们吆喝科技提供的解决方案）等。

可以看到，大数据相关技术的发展速度是逐渐加快的。原因自然是相关应用的普及（本书有很多详尽的相关案例介绍多家成功公司的应用场景），以及因此而带来的从业人员规模的增长（感谢互联网行业招募和培养了大批人才）。

从 MapReduce 论文 2004 年问世到 Apache Hadoop 框架被广泛使用，经过了 5 年以上的时间。而从 Dremel 论文问世到 Druid 被广泛认可，只用了 3 年时间。值得指出的是，在这几年时间内，还有很多公司借鉴了 Lambda Calculus 思想自己研发了闭源系统（Microsoft Dryad、阿里巴巴等）。不过经过几年的实践摸索，业内逐渐形成了以 Apache 的一系列项目为核心的统一解决方案。大家逐渐意识到，与其对同一问题采取不同的解决方案，不如一个问题一个解决方案，然后大家一起来探索解决更多不同场景的问题。这是现代互联网时代特有的网络效应和规模效应。

本书很大的贡献就是普及 Druid（以及如 Pinot 这样的相似框架），让更多的技术人员、数据人员和互联网业务人员可以快速地熟悉和尝试这个成功的新技术，将它应用在更多场景中，然后能激发更多的创新，进一步推动 Druid 以及新技术的持续发展。

我在 Google 总部工作的时候，经常使用 Dremel（和 Druid 类似的工具），也用过基于 Dremel 的可视化系统 PowerDrill。当时的感觉是，一个像 SQL 一样好用的工具，却能快速查询海量的实时的数据，对业务帮助非常大。举个例子，当时某个广告新产品上线测试后数据不佳，Dremel 从实时的数据里发现在某些浏览器里没有点击，于是进一步发现在这些浏览器里渲染有问题，马上改正。如果没有 Dremel，这个问题的解决可能需要至少 1 周以上，而不是几个小时。相信开源的 Druid 也会像 Dremel 一样，在很多企业内成为业务数据分析的利器，大幅度提高大家的工作效率。

本书特别出色的地方在于，不仅对 Druid 的架构以及细节有深入的阐述，而且有非常详尽的代码例子（codelab），甚至有一章专门介绍怎么安装和配置，非常适合工程师一边学习，一边上机实践。在 Druid 项目文档还不是特别完善的情况下，这本书不仅适合作为大家的学习材料，还能当作日常工作中的手册，已备随时查询。

本书的作者欧阳辰是大数据领域的顶级专家，他现在服务的小米公司在大数据创新上非常积极，对于 Druid 的使用和贡献也处于业内领先的地位。所以，本书里有非常多的真实业务场景相关的解析，不仅对技术人员，而且对数据人员和业务人员也非常有借鉴价值。

如果你想拥抱大数据的新时代，Druid 是你的必学，本书是你的必读。

王晔

AdHoc 吆喝科技 创始人 CEO

推荐序二

向在大数据行业从事多年的架构师、正在如火如荼地开展大数据相关工作的工程师，以及正在准备步入大数据行业的新手推荐《Druid 实时大数据分析原理与实践》这本书。

在从北京到苏州的高铁上花了 5 个小时读了这本书，虽然还没有读完，但是我已经可以非常确定地告诉大家这本书“非常引人入胜”。对我这样一个在软件行业做了 30 年的“码农”、15 年以上互联网从业的老兵来说，也别开生面地学了很多新知识，又把脑中的和大数据有关的各种系统知识重新更新了一遍。

本书非常清晰、明确地介绍了 Druid 是什么、是为什么设计的、特点和特长是什么，以及怎么使用。

本书在介绍美国 MetaMarkets 公司为什么会设计 Druid 的同时介绍了业界流行的和大数据有关的大部分系统，以及这些系统诞生的原因及相互之间的比较和特长，比如经典的 Hadoop、飞速发展的 Spark、用于实时数据流的 kafka，非常引人入胜。

本书在介绍为什么和如何使用 Druid 的同时介绍了 Druid 的源代码结构，对那些心里痒痒地想给 Druid 做点贡献的工程师开启了一条入门的道路。

本书最后一章“Druid 生态与展望”很好地介绍了在先行使用 Druid 的用户中逐渐开发的配套设施，以及这些配套设施如何反过来帮助 Druid 的发展。想使用或者评估 Druid 的用户都能从这一章得到很多新的启示，并节省用来评估和寻找 Druid 相关配套设施的时间。

Sherman Tong
微软中国研发中心，高级研发总监

推荐语

（排名不分先后）

只有久经考验又乐于分享的大数据架构师，才有这样的功力，把实时大数据分析技术的原理与实践讲得这么系统与透彻。书中随处可见来自实践的真知灼见。阅读这本书，就如同由一位老司机带着开启的美妙旅程，一路轻松、兴奋、风景无限。

鲁肃

蚂蚁金服 CTO

无论是数据量总量还是数据增量都在急速增长的背景下，急需一种技术能够快速地对海量数据进行实时存储和多维度实时分析，Druid 作为一款优秀的实时大数据分析引擎应运而生。Druid 非常强大，与之伴随的是使用上的复杂性，因此理解 Druid 的架构和运行机制原理对于更好地使用 Druid 及定制化扩展显得尤为重要。《Druid 实时大数据分析原理与实践》这本书正好可以满足读者的需求。诸位作者理论功底深厚，实践经验丰富，本书可以帮助大家快速地了解和学习 Druid，强烈推荐。

张雪峰

饿了么 CTO

开源软件已经成为了构建现代软件系统的重要基石，特别是在大数据和云计算等热门领域，开源软件更是独领风骚。作为一家技术驱动的公司，Testin 云测一直是开源软件坚定的倡导者和实践者，在 Testin 各个产品线中都使用了开源技术，Testin 云测是开源的受益者。其中，大数据实时多维度分析场景充满技术挑战，很高兴看到 Druid 最终完美地解决了我们客户的问题。大数据时代已经到来，Druid 无疑是解决大数据多维度实时分析的最佳选择，本书则是一把打开该技术之门的钥匙。

徐琨

Testin 云测总裁

2015 年，我们因大数据实时分析的业务需求而开始接触 Druid。在做架构选型时，Druid 因其在快速查询、水平扩展、实时数据摄入和分析这三方面都有良好的支持而很好地满足了我们的需求。2016 年年初，我们几个 Druid 技术爱好者和 Druid 联合创始人 Fangjin Yang 一起组建了 Druid 中国用户组的微信群，并举办了多次 Druid Meetup。靠着技术圈同学的口口相传，Druid 中国用户组从最初十几个人的小群，已发展为 500 人的大群；与此同时，阿里、腾讯、小米、滴滴等众多公司也都开始使用 Druid。本书的几位作者都是 Druid 中国用户组中非常活跃的技术专家，他们在社区中的口碑是本书质量的保证，如果你对 Druid 感兴趣，这本书一定不能错过。

陈冠诚

Druid 中国用户组发起人

从 2011 年创业开始，TalkingData 就是开源技术社区的重度参与者，因为我们始终面临海量数据的压力，仅靠自己闭门造车完全行不通。我们自 2013 年开始关注 Druid 项目，因为它的特性非常契合分析的业务场景，能解决海量数据的多维交叉分析问题。同时，为了增强其分析能力，我们也在把基于 Bitmap 的自研分析引擎 Atom Cube 融合到 Druid 中。拥抱开源社区的各种曲折，有苦有乐，不足道也，但是庆幸有许多热情的领路人，给予大家无私的帮助。本书作者之一欧阳辰就是这样一位乐于分享的人，文理兼修，对技术和数据都有深厚的积累和独到的见解，让人敬佩。相信这本书一定能够带大家领略 Druid 的魅力，让大家少走弯路，真正聚焦在对数据的探索上。

肖文峰

TalkingData CTO

Druid 正在开创海量数据实时数据分析的时代，作为一家以技术创新驱动的公司，OneAPM 幸运地在正确的时间选择了正确的技术构筑自己的后端处理平台，我希望 OneAPM 的经验能够给后来者以借鉴，本书作者之一麒赟是 Druid 技术在 OneAPM 落地生根的实践者，这本书一定能够给大家更多的启迪。

何晓阳

OneAPM 创始人，董事长

开源软件在过去十年中蓬勃发展，特别是在大数据等新兴领域，开源软件逐渐在企业级应用中占有一席之地。我们很欣喜地看到 Druid 这样有中国元素的开源项目在这个过程中茁壮成长，被企业客户接受并在核心系统应用中部署。

刘隶放

Cloudera 大中华区技术总监

我用“大、全、细、时”四个字来总结大数据，传统数据库在这种数据特性下根本无法支撑。而 Druid 的出现，正好比较完美地满足了这四点，特别是对于维度变换不频繁的场景，非常适用。本书既讲解了 Druid 技术本身，也讲解了多维数据分析相关的知识，并对业内的分布式存储和查询系统都做了对比。想要系统掌握 Druid 技术，推荐阅读本书。

桑文锋

神策数据公司创始人 & CEO

Druid 是一个分布式的支持实时分析的数据存储系统（DataStore），Druid 设计之初就是为分析而生的，主要应用于大数据实时查询和分析的高容错、高性能开源分布式系统，旨在快速处理大规模的数据，并能够实现快速查询和分析。本书让读者能够深入了解 Druid 的架构设计、设计理念、安装配置、集群管理和监控，书中还介绍了一些高级特性和核心源码的导读，最后深入分析了 Druid 的最佳实践。本书采用由浅入深、循序渐进的方式介绍 Druid，是一本非常难得的 OLAP 的实时分析系统经典书籍。

卢亿雷

AdMaster（精硕科技）技术副总裁

前言

大数据的繁荣已经来到，Druid 是数据分析的一把利剑，以开源之道，高效解决了大数据实时分析的众多场景！希望此书成为 Druid 宝剑秘籍，帮助读者利用 Druid 解决业务问题。

大数据，相比传统数据，具有四个典型特征，即形式多、体量大、速度快及价值高，最终以产生商业和社会价值为目的。Druid 在解决大数据问题时，能够比较好地处理其中两个方面，一是大数据量；二是实时处理速度。Druid 在设计上支持 PB 级别的数据处理能力，在实践中，不少公司都有几百 TB 数据量的成功应用。实时性是 Druid 的一个内置特性，能够轻松应对每秒数万的流式实时事件，并且支持水平扩展。Druid 的大多数据查询的响应时间也在亚秒级。在数据多样性方面，Druid 还是有些局限性的，它只支持强类型的数据结构，原因是为了保证数据索引的高效性和查询性能。

2005 年以前，数据分析主要是以关系型数据库为基础，包括多维数据库（OLAP），支持中小规模数据的复杂维度的分析查询。2005 年以后，很多分析场景的数据量大大增加，对实时性要求高，对计算总成本敏感，主流关系型数据库开始有些力不从心。同时，数据行业涌现出一批基于新硬件架构设计的数据存储系统，常常统称为 NoSQL，很多系统都高效地解决了 Key-Value 的存储和访问问题。

2010 年后，基于大数据的分析场景越来多、越来越重要。大数据驱动的业务模式开始深入人心，没有数据指标，就无法进行优化。很多数据软件系统都无法支持 TB 到 PB 级别数据量的实时分析，Druid 就是在这种背景下脱颖而出的开源软件，帮助大家解决业务中的实时数据分析问题。

2011 年，MetaMarkets 公司为了解决广告交易中海量实时数据的分析问题，在尝试了各种 SQL 和 NoSQL 方案后，最后决定自行设计且创建了 Druid，该项目于 2013 年开源。Druid 是一个支持在大型数据集上进行实时查询而设计的开源数据分析和存储系统，提供了低成本、高性能、高可靠性的解决方案，整个系统支持水平扩展，管理方便。实际上，Druid 的很多设计思想来源于 Google 的秘密分析武器 PowerDrill，在功能上，和 Apache 开源的 Drill 也有几分相似。Druid 被设计成支持 PB 级别的数据量，现实中有数百 TB 级别的数据应用实

例，每天处理数十亿流式事件。Druid 之所以保持高效，有这样几个原因：一是数据进行了有效的聚合或预计算；二是数据结构的优化，应用了 Bitmap 的压缩算法；三是可扩展的高可用架构，灵活支持部署的扩展；四是社区的力量，Druid 开发和用户社区保持活跃，不断推动 Druid 的完善和改进。

Druid 成功应用于众多互联网和非互联网公司中，特别是用户行为分析、个性化推荐的数据分析、物联网的实时数据分析、互联网广告交易分析等领域。国内的主流广告技术公司，都曾尝试或开始采用 Druid 支持实时数据分析。传统技术公司如 Cisco, SK Telecom，也都在使用 Druid 进行用户行为分析等项目。Druid 帮助这些业务场景实现了高效数据存储和流式数据分析。

另外，Druid 项目中也有不少中国元素，其创始人之一为中国工程师杨仿今，其他核心开发工程师也包括阿里的宾莉金、谷歌的郭秉坤等。杨仿今曾多次来到中国进行 Druid 的技术交流。Druid 项目初期，不少中国广告技术公司参与了 Druid 的技术评估。目前该技术也广泛应用于中国互联网公司中，例如腾讯、阿里、小米、优酷土豆、蓝海讯通等。

本书的目的就是介绍 Druid，让读者能够深入了解 Druid 的架构设计、使用管理，也介绍了一些高级特性和核心源码的导读。本书采用由浅入深、循序渐进的方式介绍 Druid，内容组织如下：

第 1 章，介绍 Druid 的初级概念；第 2 章，对行业中不同的数据分析软件进行介绍和对比，包括一些时序数据库；第 3 章，Druid 的设计理念和架构介绍；第 4 章，Druid 的安装和配置；第 5 章，Druid 的数据摄入；第 6 章，查询详解；第 7 章，介绍 Druid 的一些高级特性，包括正在积极完善的一些功能；第 8 章，核心代码的导读和分析；第 9 章，集群管理中的安全和监控；第 10 章，介绍几个公司的 Druid 最佳实践；第 11 章，Druid 的生态介绍和展望。附录 A 简要回答了一些常见的问题；附录 B 列出了各个服务模块的参数含义和建议值，方便系统管理。

Druid 本身也在不断升级中，大约每 3~6 个月都有一次升级，每年都有一个大变化，支持更多的业务场景，不断支持各种主流开源生态，同时围绕着 Druid 的开源生态也在慢慢崛起，包括灵活数据摄入、数据可视化、标准 SQL 查询等。目前 Druid 在中国发展非常迅速，几乎都是口口相传的推广，很多基于 Druid 的项目都是线上产品的一部分。正因如此，我们有理由相信 Druid 必将会在开源世界里拥有一个更为繁荣和光彩的明天！

本书创作组

目录