O'REILLY®

Kudu:
构建高性能实时数据分析存储系统

Getting Started with Kudu:
Perform Fast Analytics on Fast Data

[美] Jean-Marc Spaggiari, Mladen Kovacevic,
Brock Noland, Ryan Bosshart 著
常冰琳 译

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

O'REILLY®

Kudu：构建高性能实时数据分析存储系统

Getting Started with Kudu: Perform Fast Analytics on Fast Data

【美】Jean-Marc Spaggiari, Mladen Kovacevic,
Brock Noland, Ryan Bosshart 著

常冰琳 译

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

内 容 简 介

要在Hadoop生态系统中实现数据的快速输入和快速分析，一直以来只有少数可用但是不够完美的解决方案。它们要么以缓慢的数据输入为代价实现快速分析，要么以缓慢的分析为代价实现快速的数据输入。这个问题现在有了解决办法，使用Apache Kudu基于列的数据存储，可以很容易地对快速输入的数据进行快速的分析。这就是本书的内容。

在这本书中，你将学习Kudu设计中的关键概念，以及如何用它构建快速、可扩展和可靠的应用程序。通过实际的示例，你将了解Kudu是如何与其他Hadoop生态系统组件（如Apache Spark、Spark SQL和Impala）集成的。

本书适合大数据系统的架构师、开发者和咨询师阅读。

版权贸易合同登记号 图字：01-2018-4157

图书在版编目（CIP）数据

Kudu：构建高性能实时数据分析存储系统 /（美）吉恩-马克·斯帕加里（Jean-Marc Spaggiari）等著；常冰琳译. —北京：电子工业出版社，2019.3
书名原文：Getting Started with Kudu: Perform Fast Analytics on Fast Data
ISBN 978-7-121-29541-6

Ⅰ. ①K… Ⅱ. ①吉… ②常… Ⅲ. ①数据处理 Ⅳ. ①TP274

中国版本图书馆CIP数据核字(2019)第030205号

责任编辑：许　艳
封面设计：Randy Comer　　张　健
印　　刷：三河市华成印务有限公司
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：787×980　1/16　印张：12.5　字数：160.6千字
版　　次：2019年3月第1版
印　　次：2019年3月第1次印刷
定　　价：69.00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。

O'Reilly Media, Inc.介绍

O'Reilly Media通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自1978年开始，O'Reilly一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了Make杂志，从而成为DIY革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版、在线服务或者面授课程，每一项O'Reilly的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar博客有口皆碑。”

——Wired

“O'Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference是聚集关键思想领袖的绝对典范。”

——CRN

“一本O'Reilly的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去，Tim似乎每一次都选择了小路，而且有几次都是转瞬即逝的机会，尽管大路也不错。”

——Linux Journal

献给所有没日没夜、绞尽脑汁为那些行为诡异、似乎故意出错的软件做架构设计、开发和咨询的人。

献给我们的家人，他们可能根本不关心技术，但仍然给予我们巨大的耐心和支持，让我们能投入时间和精力写书。没有他们，这本书是不可能完成的。爱你们！

前言

选择存储引擎是实施所有大数据项目时要做的最重要的决定之一，而且更换存储引擎的成本也是最高的。Apache Kudu 是 Hadoop 生态系统中的一个全新存储系统。它的灵活性使我们能够更快地搭建和维护应用程序。在 Hadoop 开发者的大数据工具箱中，Kudu 是一个关键工具。它解决了一些使用目前的 Hadoop 存储技术很难实现或不可能实现的常见问题。

在这本书中，你将学习 Kudu 设计中的关键概念，以及如何用它构建快速、可扩展和可靠的 Kudu 应用程序。通过实际的示例，你将了解 Kudu 如何与其他 Hadoop 生态系统组件（如 Spark、Spark SQL 和 Impala）集成。

本书假设读者对 Hadoop 生态系统组件（如 HDFS、Hive、Spark 或 Impala）有一些使用经验，有 Java 或 Scala 编程经验，还有 SQL 和传统关系型数据库管理系统“使用”经验，熟悉 Linux shell。

本书使用的约定

本书中使用了下列排版约定：

斜体（*Italic*）

表示新的术语、URL、电子邮件地址、文件名和文件扩展名。

等宽字体（Constant Width）

表示程序清单、在段落内引用的程序变量或函数名、数据库名、数据类型、环境变量、语句和关键字。

等宽粗体（**Constant Width Bold**）

表示应该由用户输入的命令或其他文本。

等宽斜体（*Constant Width Italic*）

表示该内容应由用户提供或由上下文决定。

这个图标表示提示或建议。

这个图标表示一般性注释。

这个图标表示警告或注意。

使用代码示例

可以在 *https://github.com/kudu-book/getting-started-kudu* 下载本书的补充材料（代码示例、练习等）。

这本书就是为了帮助你完成工作的。一般来说，如果本书提供了示例代码，你可以在自己的程序和文档中使用它。除非复制了相当多的示例代码，否则你不需要与我们联系。例如，你编写一个程序，其中使用了本书中的几个代码块，这种情况是不需要获得我们的许可的。但是如果你销售或发行一张光盘，其中包含了 O'Reilly 图书中的例子，则需要获得许可。引用本书和引用示例代码来回答问题，不需要许可。将大量示例代码嵌入到你的产品文档中则需要获得许可。

我们将十分感激你在使用我们的代码时标注相关的版权信息，但是并不强求你这么做。版权信息通常包括书名、作者、出版商和 ISBN。例如："Getting Started with Kudu by Jean-Marc Spaggiari, Mladen Kovacevic, Brock Noland, Ryan Bosshart (O'Reilly). Copyright 2018 Jean-Marc Spaggiari, Mladen Kovacevic, Brock Noland, Ryan Bosshart, 978-1-491-98025-5"。

如果你觉得自己对代码的使用不在合理使用或以上许可的范围内，请通过 permissions@oreilly.com 与我们联系。

O'Reilly Safari

Safari（以前的 Safari Books Online）是企业、政府、教育者和个人的会员制培训及参考平台。

订阅者可以从一个完全可搜索的数据库中获得来自 250 多家出版商的成千

上万的书籍、培训视频、互动教程，这些出版商包括 O'Reilly Media、Harvard Business Review、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、Adobe、Focal Press、Cisco Press、John Wiley & Sons、Syngress、Morgan Kaufmann、IBM Redbooks、Packt、Adobe Press、FT Press、Apress、Manning、New Riders、McGraw-Hill、Jones & Bartlett 和 Course Technology。

若想获得更多资讯，请访问 *http://oreilly.com/safari*。

联系我们

请将对本书的评价和发现的问题通过如下地址告知出版社。

美国：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472

中国：

北京市西城区西直门南大街 2 号成铭大厦 C 座 807 室（100035）
奥莱利技术咨询（北京）有限公司

我们提供了本书网页，上面列出了勘误表、示例和其他信息。请通过 *http://bit.ly/getting-started-with-kudu* 访问。

要给本书提意见或者询问技术问题，请发送邮件到 *bookquestions@oreilly.com*。

更多有关书籍、课程、会议和新闻的信息，请见网站 *http://www.oreilly.com*。

在 Facebook 找到我们：*http://facebook.com/oreilly*。

在 Twitter 上关注我们：*http://twitter.com/oreillymedia*。

在 YouTube 上观看：*http://www.youtube.com/oreillymedia*。

致谢

感谢 Apache Kudu 社区的帮助，包括 Apache Kudu 的创始人、代码提交者、贡献者、早期采纳者和用户。感谢我们的技术审稿人 David Yahalom、Andy Stadtler 和 Attila Bukor 对细节的关注和反馈。还要感谢非官方的技术审稿人，包括 Nipun Parasrampuria、Mac Noland、Sandish Kumar、Tony Foerster、Mike Rasmussen、Jordan Birdsell 和 Gunaranjan Sundararajan。

Ryan Bosshart 和 Brock Noland 感谢他们在 phData 和 Cloudera 的同事对本书的支持和投入。

Mladen Kovacevic 感谢他在 Cloudera 的同事，包括解决方案架构师、工程师、技术支持人员、产品管理人员以及其他人，感谢他们的热情和支持。Mladen 还要感谢家人在他写作时给予的耐心、支持和鼓励——没有他们，他是不可能完成写作的！

Jean-Marc Spaggiari 要感谢所有在此过程中支持他的人。

目录

读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- **提交勘误**：您对书中内容的修改意见可在 提交勘误 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动**：在页面下方 读者评论 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：*http://www.broadview.com.cn/29541*

第 1 章

为什么会有Kudu

Kudu 为什么重要

随着大数据平台的不断创新和发展，无论是在企业内部还是在云上，新产品的发布都让人应接不暇，大家对此已经很习惯了。不过，2017 年在一些大公司和各种业务应用中使用过 Kudu 之后，我们比以往任何时候都更加确信 Kudu 很重要，让这个项目加入开源大数据世界是非常有价值的。

原因可以归结为以下三点：

1．大数据仍然是一项很难的技术——由于对数据的需求以及用户规模持续增长，Hadoop 和大数据系统相对来说还是很难使用的，其大部分的复杂性来源于存储系统的局限性。而 Kudu 的局限性就小得多，在我们公司，从前冗长的架构讨论现在变成简单的一句话：“用 Kudu 就好了。”

2．新的应用场景需要 Kudu——Hadoop 的应用场景正在发生变化，其中的一个变化是越来越多的应用集中在机器生成的数据和实时分析领域。为了展示这种变化带来的复杂性，我们会讨论几个使用现有大数据存储技术进

行实时分析的架构方案，然后讨论 Kudu 能如何简化这些方案。

3．硬件环境正在发生变化——Hadoop 诞生时的很多硬件现在发生了变化。因此，我们有新的契机来为新的硬件环境创建新的存储系统，从而提升性能和应用的灵活性。

在本章中，我们会详细讨论设计 Kudu 的以上三点动机。如果你已经熟知这些，可以跳到本章的后半部分，在那里我们会讨论 Kudu 的设计目标，将 Kudu 和其他大数据存储系统对比。在本章的最后，我们会总结为什么世界需要 Kudu 这个新的大数据存储系统。

易用性驱动接纳度

分布式系统过去一直既昂贵又难用。21 世纪头 10 年的中期，我们曾经为一家大型媒体和信息提供商工作，负责搭建一个输入、处理、搜索、存储和检索数百 TB 在线内容数据的平台。搭建这样一个平台是一项庞大的工程，需要数百名工程师和不同的团队来构建平台的各个部分。我们有独立的团队分别负责实现不同功能的分布式系统，包括数据的输入、存储、搜索等。为了实现系统的高可扩展性，我们为关系数据存储、搜索索引和存储系统都做了分片，然后在上层构建精心设计的元数据系统，以保持所有数据都整齐一致。这个平台是建立在昂贵的专利技术之上的，这就把那些想要做同样事情的小型竞争对手挡在了外面。

大约在同一时间，Doug Cutting 和 Mike Carafella 开发了 Apache Hadoop。由于他们以及整个 Hadoop 生态社区的努力，搭建高可扩展性的基础架构不再需要数百万美元的成本和庞大的分布式系统工程师团队。Hadoop 带来的最主要的进步在于，只要软件工程师对分布式系统有一定的了解，他就