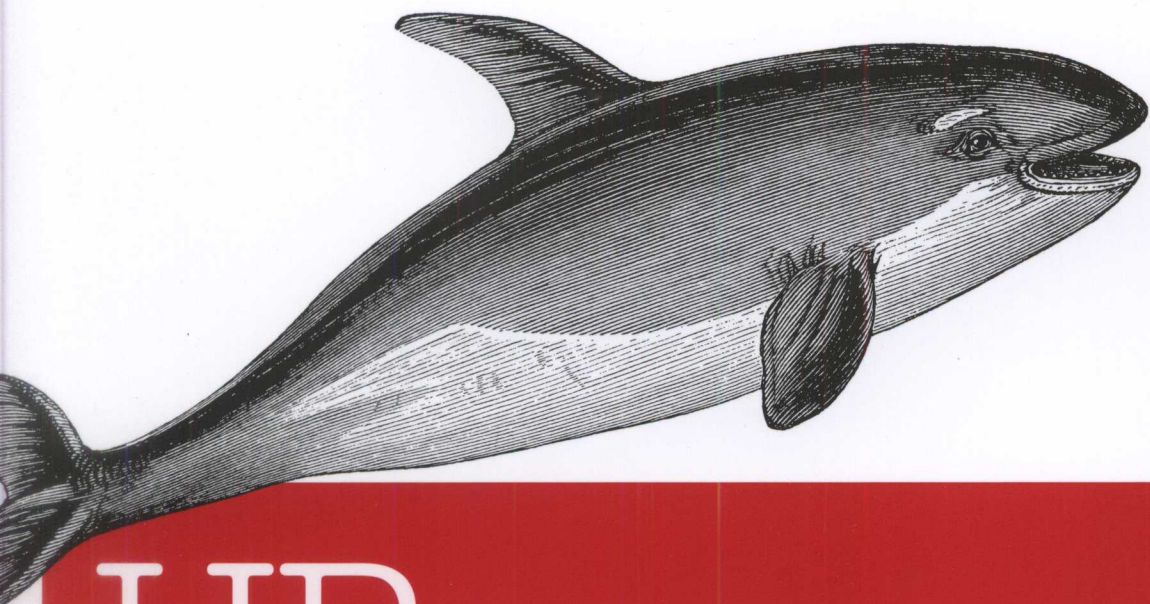


O'REILLY®



HBase 应用架构

Architecting HBase Applications

Jean-Marc Spaggiari, Kevin O'Dell 著
陈敏敏 夏锐 陈其生 译

中国电力出版社

HBase应用架构

Jean-Marc Spaggiari Kevin O'Dell 著

陈敏敏 夏锐 陈其生 译

Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

O'Reilly Media, Inc. 授权中国电力出版社出版

中国电力出版社

Copyright © 2016 Jean-Marc Spaggiari and Kevin O'Dell. All right reserved.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and China Electric Power Press, 2017.
Authorized translation of the English edition, 2016 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

英文原版由 O'Reilly Media, Inc. 出版 2016。

简体中文版由中国电力出版社出版 2017。英文原版的翻译得到 O'Reilly Media, Inc. 的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly Media, Inc. 的许可。

版权所有，未得书面许可，本书的任何部分和全部不得以任何形式重制。

图书在版编目 (CIP) 数据

HBase应用架构 / (美) 吉恩·马克·斯帕加里 (Jean-Marc Spaggiari), (美) 凯文·奥戴尔 (Kevin O'Dell) 著; 陈敏敏, 夏锐, 陈其生译. — 北京: 中国电力出版社, 2017.9

书名原文: Architecting HBase Applications

ISBN 978-7-5198-1121-1

I. ①H… II. ①吉… ②凯… ③陈… ④夏… ⑤陈… III. ①计算机网络—信息存贮 IV. ①TP393

中国版本图书馆CIP数据核字(2017)第219369号

北京市版权局著作权合同登记 图字: 01-2017-4340号

出版发行: 中国电力出版社

地 址: 北京市东城区北京站西街19号 (邮政编码100005)

网 址: <http://www.cepp.sgcc.com.cn>

责任编辑: 刘 焱 (liuchi1030@163.com)

责任校对: 常燕昆

装帧设计: Karen Montgomery, 张 健

责任印制: 蔺义舟

印 刷: 北京天宇星印刷厂

版 次: 2017年9月第一版

印 次: 2017年9月北京第一次印刷

开 本: 750毫米×980毫米 16开本

印 张: 14.5 980

字 数: 273千字

印 数: 0001—3000册

定 价: 48.00元

版 权 专 有 侵 权 必 究

本书如有印装质量问题, 我社发行部负责退换

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，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

译者序

HBase 是 Hadoop Database 的简称, 基于 Google Bigtable 实现的开源、分布式、可伸缩的列式存储数据库, 自其诞生之日起, 就受到业界的追捧, 而今已成为了 Hadoop 生态圈和各公司大数据平台不可或缺的组成部分。时间序列数据库 Opentsdb、阿里的 HQueue、OLAP 分析引擎 Kylin 等很多大数据开源项目也都是基于 HBase 的。

本书是 HBase 应用架构的手册指南, 它涵盖了 HBase 实践应用的各个方面, 从 HBase 底层的基本原理到整个生态系统, 从最初的应用实例到性能优化, 以及后续的疑难问题的定位与排错等方面。同时, 作者也列举了大量的实例场景及实现代码, 读者都可直接从中获取实际应用原型。

本书主要由陈敏敏 (苏宁云商大数据中心总监)、夏锐 (奇虎 360 技术专家)、陈其生 (唯品会技术专家) 翻译, 由于译者水平所限, 翻译难免有疏漏之处, 愿广大读者给予批评指正。

译者

目录

感谢我父亲，我希望你能看到这本书……

——Jean-Marc Spaggiari

感谢我日思夜想的母亲，感谢每天一直陪在我身边的父亲，以及我美丽的妻子Melanie和女儿Scotland，是她们的陪伴帮助我度过枯燥且漫长的写作生活。

——Kevin O'Dell

第一部分 HBase概述

第1章 什么是HBase 11

面向列与面向行 11

实现和使用案例 13

第2章 HBase原理 15

表格之列表 15

表的内部操作 22

复制 27

HBase角色 31

第3章 HBase生态系统 32

系统工具 32

SQL 40

框架 42

第4章 HBase性能预估和优化概述 45

硬件 46

目录

序	1
前言	3
第一部分 HBase概述	
第1章 什么是HBase	11
面向列与面向行	13
实现和使用示例	13
第2章 HBase原理	15
表格式列族	15
表的内部操作	22
依赖	27
HBase 角色	27
第3章 HBase生态系统	32
监控工具	32
SQL	40
框架	42
第4章 HBase规模预估和调优概述	45
硬件	46

存储	46
网络	47
操作系统调优	48
Hadoop调优	49
HBase调优	50
负载不均调优	52
第5章 环境设置	54
系统要求	55
HBase单机安装	58
虚拟机中的HBase	61
本地与VM	62
故障排除	63
第一步	65
伪分布式模式和全分布式模式	73
第二部分 用例	
第6章 用例：HBase作为一个记录系统	77
摄取/预处理	78
处理/服务	79
用户体验	84
第7章 底层存储引擎的实现	87
表设计	87
数据转换	92
HFile校验	98
批量加载	99
数据索引	104
数据检索	107
更进一步	109

第8章 用例：近实时事件处理	111
摄取/预处理	114
近实时事件处理	115
处理/服务	116
第9章 近实时实现事件处理	119
应用流	121
实施	125
进一步	144
第10章 用例：HBase作为主数据管理工具	145
摄取	146
处理	148
第11章 主数据管理工具HBase的实现	150
MapReduce与Spark	150
Spark与HBase交互	151
Spark结合HBase实现	152
进一步	162
第12章 用例：文档存储	163
数据服务	165
数据摄取	166
清理	168
第13章 文档存储的实现	169
MOB	169
数据一致性	174
进一步	175

第三部分 疑难问题的定位和排错

第14章 region过多	179
后果	179
原因	180
解决方案	181
防范	188
第15章 列族过多	191
后果	192
原因、解决方案和预防	193
第16章 热点	196
后果	196
原因	196
防范和解决方案	200
第17章 超时和垃圾回收	201
后果	201
原因	203
解决方案	205
预防	205
第18章 HBase和不一致	210
HBase文件系统布局	210
查看META表	211
在HDFS上查看HBase	212
HBase概述	214
使用HBase	215

序

在我25年软件从业生涯里，经历了许多颠覆性的变化：互联网、万维网、大型机、C/S架构模型等。我曾经在一个保障炼油厂安全的软件研发团队工作。我们40人的团队共享唯一的一台DEC VAX机器，而且性能也远不及我现在使用的手机。

我还记得20世纪90年代初的一天，我们被安排接收新的机器。要更换的机器位于三楼，是一个很大很重的像洗衣机大小的庞然大物。我们一群人在“机房”里等着，想看看他们是怎么把新机器搬上楼梯的。我们可以想象一下一个巨大的起重机，外面的街道被封锁的场景……总之，这是一个浩大的工程！

但实际上出乎意料。一个男的提着一个小箱子走进房间。他把它放在那台像老式“洗衣机”的机器的顶部，换了几根电缆，做了一些测试，便离开了。这样，就完成了！

这是技术产业带来的喜悦：如果我们愿意学习新事物，并与之同行，我们将永远不会感到厌倦，不断地遇见惊喜。几年前似乎不可能的事情突然会变得司空见惯。

大数据就是这样在变化，大数据无处不在。由谷歌公司的Google File System和BigTable引发的技术变革几乎影响了所有的高科技公司、银行和政府。无论是好是坏，这些系统直接或间接地触及了地球上几乎每个人的生活。

和它之前的BigTable一样，Apache HBase已经在大数据生态系统中处于独特的地位：它在一个不变的分式文件系统中支持了一种可更新的、无限扩展的数据集存储。因此，它构建了纯文件存储和OLTP / OLAP数据库之间的桥梁。

HBase无处不在：Facebook、苹果、Salesforce.com、Adobe、雅虎、Bloomberg、华为、Gap和许多其他公司都在使用它。谷歌采用了HBase API为其提供公共云Bigtable，这也是HBase受欢迎的一个证据。

尽管HBase无处不在，但它也不是即插即用。分布式系统很难，诸如分区容忍性、一致性和可用性等在每个讨论区都变成了大家热议的话题，不久之后更深奥的术语如热点(hotspotting)和数据加盐(salting)存储也出现了。成百上千台机器的水平扩展都需要经历痛苦的数据平衡过程，这些平衡使得我们更加难以有效地使用这些系统，HBase也不例外。

在混迹于HBase和Hadoop社区的这段岁月里，我已获得了这些挑战的第一手经验。必须仔细设计和构建这些用例，以发挥HBase的优势。

这本书由两位内部人士撰写，他们一直在现场做支持。本书是一个非常有帮助的指南，它详细介绍了如何基于HBase架构来设计应用程序，并能够支持数千台机器的集群扩展。

如果你正在构建或计划构建高扩展性及高可靠性的新应用程序，那么这本书就是为你设计的。Jean-Marc 和Kevin已经看到这一切：使用案例，人们犯的错误以及单服务器系统的设计臆想。最重要的是，他们知道哪些是运行正常的，如何解决那些运行不正常的情况，以及如何理解这里面的本质。

—— Lars Hofhansl,

HBase的提交者和成员

项目管理委员会

Apache基金会成员Salesforce.com副总裁

前言

随着Hadoop越来越受到很多人的欢迎，其生态系统也充满活力，包括广泛使用的工具，如Hive、Spark、Impala及HBase。这本书着重于工具Apache HBase，它构建于Hadoop分布式文件系统（HDFS）之上，具有可扩展性、容错性、低延迟等特征。HBase整合了Hadoop的水平扩展能力和实时数据服务两方面的优势。在规模方面，HBase允许从单个集群中进行每秒钟数百万次的读写操作，同时仍可保持Hadoop所有应用的可用。HBase迅速普及，现在已经为世界上一些较大的Hadoop集群部署提供了支持，如Apple、Salesforce.com及Facebook。

然而，开始使用HBase是一个艰巨的任务。虽然有许多可以帮助开发人员入手的资源（包括邮件列表，联机手册和Javadocs），但使用Apache HBase构建、设计和部署真正生产应用程序的信息相当有限。这正是写作本书的原因。

本书的目的是为了真正生产应用的HBase部署。虽然本书中讨论的每个用例已经部署并投入生产，但这并不意味着没有改进的余地，或者即使你不需要修改你的特定任务，但它确实展示了事情的实际情况。

这本书还包括了非常有用的故障处理内容（第三部分）。我们的目标是帮助你避免常见的部署错误。第三部分还提供了经常被我们忽略的性能调优内容，如垃圾回收和区域分配。

本书的读者对象？

本书主要针对那些架构师及开发人员而设计，希望他们能更好地理解大数据应用程序的部署。在这之前，你应该具备基本的Hadoop知识，包括所需组件的设置以及成

功安装过Hadoop集群，我们不会在Hadoop的配置或NodeManager功能上花费时间。阅读本书的架构师不需要有一个完整的Java知识，但必须充分了解部署章节的内容。这本书涵盖多个垂直用例，希望能够协助各个企业和初创公司。

架构师可以品读以细节为导向的用例章节，其中包括各个组件的部署方式以及它们是如何被集成在一起的。在开发章节，开发人员可快速查看详细代码示例，这样可以加快生产部署。部署章节提供了对特定API的深入了解以及相应的性能提升技巧，这样可大大减少我们的故障处理时间。那些对大数据好奇的人将会发现架构和部署章节都很有用，并且还可以深入了解HBase生态系统以及HBase的部署细节。

本书结构

本书分为三个部分：第一部分，HBase的介绍，涵盖的主题有：HBase是什么；HBase生态系统是什么样的以及如何部署它。第二部分，涵盖了具体用例，是本书的核心。我们希望这是你最常参考的部分，因为它包含了对你有用的提示和技巧。最后，第三部分讨论了故障处理，你应该经常参阅这一部分。我们希望这将是第二个最常参考的部分（以积极的态度，而不是被动的）。本部分提供了关于region数量控制、垃圾收集调优和避免热点等方面的内容。

额外的资源

这本书没有涵盖HBase的内部设计。我的好朋友Lars George通过《HBase权威指南》（O'Reilly出版）将HBase的内部设计带到了一个全新的水平。我们建议把他的书当作我们的启蒙书。这将有助于你更好地理解一些术语，而我们只是用一两段注释来解释它们。

虽然我们花了不少时间讨论部署HBase的细节，但本书不会涵盖部署HBase的大部分理论。Nick Dimiduk和Amandeep Khurana的《HBase in Action》（Manning出版）涵盖了部署HBase的实用性。虽然他们的书不太关注整个应用程序的开发，但在生产实践上做了较详细的介绍。

排版约定

本书使用了下述排版约定：

斜体 (Italic)

表示新术语、URL、电子邮件地址、文件名和文件扩展名。

等宽字体 (Constant width)

用于程序列表, 以及用于引用程序元素的段落内, 如变量或函数名、数据库、数据类型、环境变量、语句和关键字。

等宽粗体 (Constant width Bold)

显示应由用户逐字键入的命令或其他文本。

等宽斜体 (Constant width italic)

显示应由用户提供的值或由上下文确定的值替换的文本。



表示提示或建议。



表示一般注意事项。



表示提醒或警告。

使用代码示例

补充材料 (代码示例, 练习等) 可从<https://github.com/ArchitectingHBase/examples> 下载。

这本书是为了帮助你完成你的工作。一般来说, 如果本书提供了代码实例, 你可以在程序和文档中使用它。除非复制代码的重要部分, 否则不需要与我们联系。例如, 编写使用本书中多个代码块的程序不需要许可。出售或分发CD-ROM的例子来自O'Reilly的书籍需要许可。引用这本书来回答一个问题和引用示例代码不需要许可。将本书中的大量示例代码写入产品文档中, 则需要获得许可。

我们推荐使用书属性的引用方式，但并不一定完全要求这样。书属性的引用方式通常包括标题，作者，出版商和ISBN。例如：“Architecting HBase Applications, Jean-Marc Spaggiari and Kevin O’Dell (O’Reilly), Copyright 2016 Jean-Marc Spaggiari and Kevin O’Dell, 978-1-491-91581-3。”如果你觉得所使用的代码示例超出了上面的许可范围，可以随时通过邮件permissions@oreilly.com与我们联系。

Safari在线图书

Safari Books Online (<http://safaribooksonline.com>) 是应需而变的数字图书馆，它同时以图书和视频的形式出版世界顶级技术和商务作家的专业作品。

Safari Books Online是技术专家、软件开发人员、Web设计师、商务人士和创意人士开展调研、解决问题、学习和认证培训的第一手资料来源。

Safari Books Online为企业、政府部门、教育机构和个人提供了多种套餐和价格。

订阅者可以在一个完全可搜索的全文数据库中访问上千种图书、培训视频和正式出版之前的书稿。这些内容由以下出版社提供：O’Reilly Media、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、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、Course Technology等。关于Safari Books Online的更多信息，请访问我们的网站：<http://www.safaribooksonline.com>。

如何联系我们

请将有关本书的意见和问题反馈给出版商：

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

本书有一个Web页面，我们列出了勘误表、示例和其他相关内容信息。可以通过<http://bit.ly/architecting-hbase-applications>访问此页面。

要评论本书或咨询技术问题，请发送电子邮件至bookquestions@oreilly.com。

想了解更多关于我们的书籍，课程，会议和新闻的信息，请访问我们的网站<http://www.oreilly.com>。

在Facebook上找到我们：<http://facebook.com/oreilly>

在Twitter上关注我们：<http://twitter.com/oreillymedia>

在YouTube上观看我们：<http://www.youtube.com/oreillymedia>

致谢

Kevin和Jean Marc要感谢每一位让这本书成功出版的人，感谢他们所有的辛苦工作：我们的编辑Marie Beaugureau；在O'Reilly Media的卓越员工；撰写序的Lars Hofhansl；我们的主要审稿人Nate Neff、Suzanne McIntosh、Jeff “Jeffrey” Holoman、Prateek Rungta、Jon Hsieh、holoman、Sean Busbey和Nicolae Popa；以及给予我们无限支持和指导的各位作者：Ben Spivey、Joey Echeverria、Ted Malaska、Gwen Shapira、Lars George、Eric Sammer、Amandeep Khurana、Tom White。我们还要感谢Linden Hillenbrand、Eric Driscoll、Ron Beck、Paul Beduhn、Matt Jackson、Ryan Blue、Aaron “ATM” Meyers、Dave Shuman、Ryan Bosshart（谢谢你的鼎力相助）、Jean-Daniel “JD” Cryans、St. Ack、Elliot Clark、Harsh J Chouraria、Amy O'Connor、Patrick Angeles和Alex Moundalexis。最后，我们要感谢Cloudera和Rocana的所有人对我们的支持、建议和鼓励。

来自Kevin

感谢我最好的朋友和兄弟的全力支持和鼓励：Matthew “Kabuki” Langrehr、Scott Hopkins、Paul Bernier、Zack Myers、Matthew Ring、Brian Clay、Chris Holt、Cole Sillivant、Viktor “Shrek” Skowronek、Kyle Prawdzik和船长Matt Jones。我还要感谢我的朋友、同事、合作伙伴以及一路帮助过我的客户：Ron Kent、John Lynch、Brian Burton、Mark Schnegelberger、David Hackett、Sekou McKissick、Scott Burkey、David Rahm、Steve Williams、Nick Preztak、Steve “Totty” Totman、Brock Noland、Josh “Nooga” Patterson、Shawn Dolley、Stephen Fritz、Richard