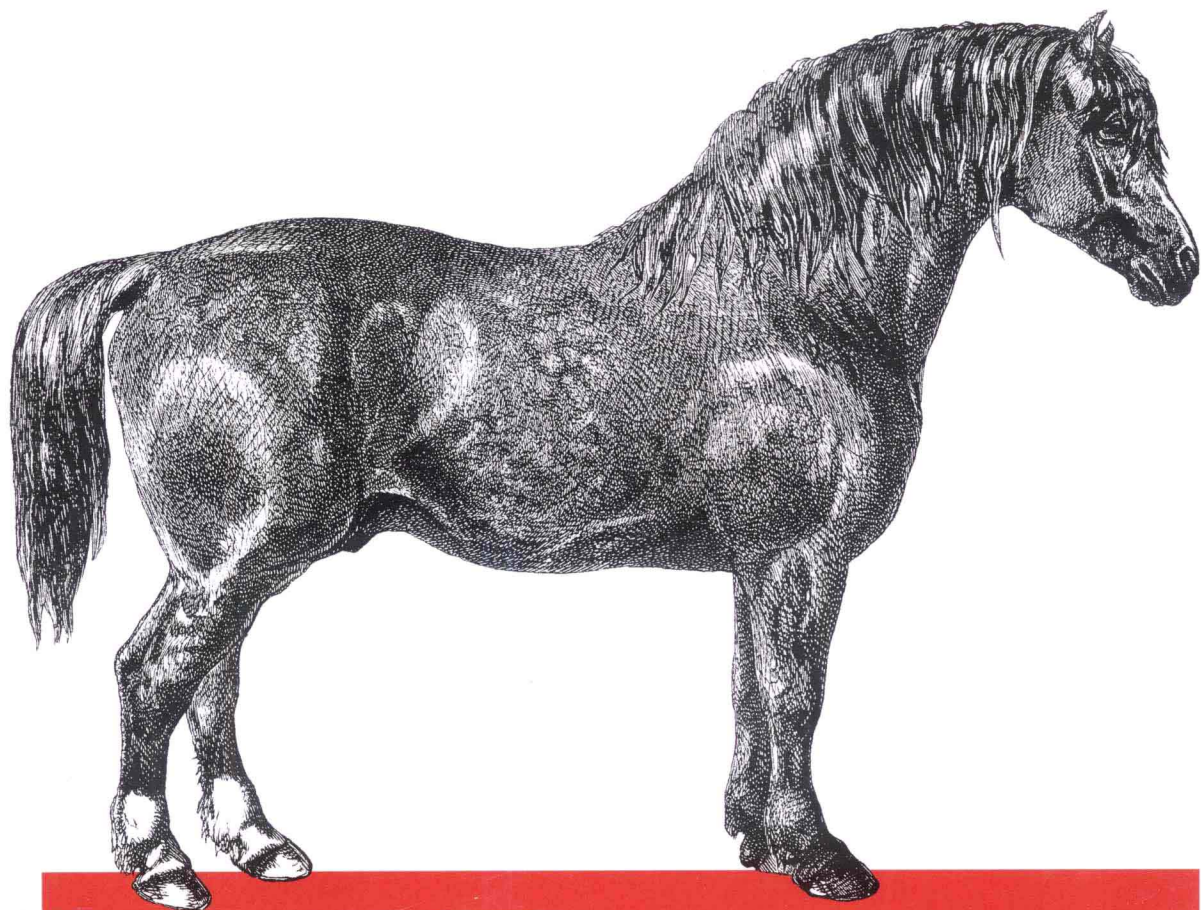


“十二五”

国家重点图书出版规划项目

HBase: The Definitive Guide



HBase

权威指南

[美] Lars George 著
代志远 刘佳 蒋杰 译

O'REILLY®



人民邮电出版社
POSTS & TELECOM PRESS

HBase 权威指南

[美] *Lars George* 著

代志远 刘 佳 蒋 杰 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'Reilly Media, Inc. 授权人民邮电出版社出版

人 民 邮 电 出 版 社
北 京

图书在版编目 (C I P) 数据

HBase权威指南 / (美) 乔治 (George, L.) 著 ; 代志远, 刘佳, 蒋杰译. -- 北京 : 人民邮电出版社, 2013.10

ISBN 978-7-115-31889-3

I. ①H… II. ①乔… ②代… ③刘… ④蒋… III. ①计算机网络—信息存贮—指南 IV. ①TP393-62

中国版本图书馆CIP数据核字(2013)第105049号

版权声明

Copyright ©2011 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2011. Authorized translation of the English edition, 2011 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.

本书中文简体版由 **O'Reilly Media, Inc.** 授权人民邮电出版社出版。未经出版者书面许可, 对本书的任何部分不得以任何方式复制或抄袭。

版权所有, 侵权必究。

-
- ◆ 著 [美] Lars George
 - 译 代志远 刘佳 蒋杰
 - 责任编辑 杨海玲
 - 执行编辑 赵越
 - 责任印制 程彦红 杨林杰
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
 - 邮编 100061 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 三河市海波印务有限公司印刷
 - ◆ 开本: 787×1000 1/16
 - 印张: 31.5
 - 字数: 646千字 2013年10月第1版
 - 印数: 1-4000册 2013年10月河北第1次印刷



著作权合同登记号 图字: 01-2011-7480 号

定价: 89.00 元

读者服务热线: (010)67132692 印装质量热线: (010)67129223
反盗版热线: (010)67171154

内容提要

本书探讨了如何通过使用与 HBase 高度集成的 Hadoop 将 HBase 的可伸缩性变得简单；把大型数据集分布到相对廉价的商业服务器集群中；使用本地 Java 客户端，或者通过提供了 REST、Avro 和 Thrift 应用编程接口的网关服务器来访问 HBase；了解 HBase 架构的细节，包括存储格式、预写日志、后台进程等；在 HBase 中集成 MapReduce 框架；了解如何调节集群、设计模式、拷贝表、导入批量数据和删除节点等。

本书适合使用 HBase 进行数据库开发的高级数据库研发人员阅读。

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

感谢我的妻子 Katja, 感谢我的女儿 Laura,
以及我的儿子 Leon。我爱你们!

推荐序

近年来，新兴的互联网服务领域，以及电信、金融和交通等各传统行业出现了数据资产的爆炸性增长，这些数据资产的类型以非结构化和半结构化为主，如何低成本且高效率地存储和处理 PB 甚至 EB 量级的数据成为了极大的挑战。

Google 公司提出的 MapReduce 编程框架、GFS 文件系统和 BigTable 存储系统成为了大数据处理技术的开拓者和领导者，而源于这三项技术的 Apache Hadoop 等开源项目则成为了大数据处理技术的事实标准，迅速推广至国内外各大互联网企业，成为了 PB 量级大数据处理的成熟技术和系统。面对不同的应用需求，基于 Hadoop 的数据处理工具也应运而生，例如，Hive、Pig 等已能够很好地解决大规模数据的离线式批量处理问题。但是，Hadoop HDFS 适合于存储非结构化数据，且受限于 Hadoop MapReduce 编程框架的高延迟数据处理机制，使得 Hadoop 无法满足大规模数据实时处理应用的需求。

传统的信息系统和 Web 应用大多采用 LAMP 架构构建，并使用关系型数据库存储、组织和管理结构化或半结构化数据。通用的关系型数据库无法很好地应对在数据规模剧增时导致的系统扩展性和性能问题。因此，业界出现了一类面向半结构化数据存储和处理的高可扩展、低写入/查询延迟的系统，例如，键值存储系统、文档存储系统和类 BigTable 存储系统等，这些特性各异的系统也可统称为 NoSQL 系统。Apache HBase 就是其中已迈向实用的成熟系统之一。HBase 之所以能成为迈向实用的成熟系统，一是核心思想来源于 Google 的 BigTable，二是有 Apache 及 Hadoop 开源社区的支撑，三是有诸如 Facebook、淘宝和支付宝等互联网公司的应用实践，保证了 HBase 系统的稳定性和可用性。目前，作为关系型数据库的有益补充，HBase 已成功应用于互联网服务领域和传统行业的众多在线式数据分析处理系统中。

本书涉及 HBase 使用和开发过程中的各方面内容，章节组织由浅入深，内容阐述细致入微并且贴近实际，可以作为参考书以方便读者在开发过程中随时查阅。本书译者之一刘佳向 HBase 开源社区提交过多项错误修复和新功能，参与过多项 HBase 有关的大数据分析系统研发项目，积累了丰富的 HBase 系统开发经验。我相信本书对于 HBase 使用者和开发者来说，都是及时和不可或缺的。

查礼

于中科院计算所

2013 年 7 月

译者序

随着历史数据的积累和数据量的高速增长，海量数据领域越来越被重视，且该领域涌现出了非常多的新技术。技术的发展和时间的沉淀使得 HBase 开始被大家广泛认可，成为海量数据在线存储领域的首选。

本书是讲述 HBase 相关技术的第一本图书，也是著名图书出版商 O'Reilly 出版发行的 HBase 权威书籍。

本书从架构、开发、应用和运维等多个角度描述了 HBase，深入介绍了 HBase 内核的原理和机制以及社区的发展方向，并提供了应用层面的多种示例和源代码。本书为每个用例和知识点提供了丰富的解释和注意要点，使用户可以由浅入深地了解原理并深度使用其功能，并且体现了在 HBase 教学方面的最新进展和最高水平。

本书的成功离不开 Lars George 的努力。在 HBase 还处于萌芽时期时，Lars George 就开始投入了大量的精力，从修复 HBase 中的问题到优化性能，推广 HBase 并编写 HBase 可用性文档，他是 HBase 领域里大师级的人物。而这本《HBase 权威指南》花费了 Lars George 许多的时间和精力。

阅读本书后，我们不得不承认这本大师级的著作很好地应对了社区中 HBase 发展所面临的挑战。不得不说的是，本书著作和翻译经历的时间较长，而社区中 HBase 发展速度较快，许多版本已经发行，许多问题也得以修复，因此，本书最终落地后会与最新 HBase 版本的功能特性有少许描述性出入，还望广大读者见谅。

在翻译过程中，我们深刻地发现国外技术领域的专业性，深深地被世界级的高水平技术所震撼。我们由衷地希望本书中文版的出版能够推动国内 HBase 教学、使用和发展。本书译者代志远在翻译期间就职于阿里巴巴，译者刘佳是中科院计算所研究生，现为普泽天玑技术总监，译者蒋杰在腾讯担任数据与运营支撑平台副总经理。

感谢人民邮电出版社的编辑，他们为保证本书质量付出了大量的努力。

本书中概念和术语较多，许多概念和术语尚无公认的中文译法，加之译者水平有限，译文中若有不妥之处，恳请读者批评指正。

代志远

2013 年 7 月

序

HBase 的故事开始于 2006 年，当时旧金山的 Powerset 创业公司试图建立一个网页的自然语言搜索引擎，但他们构建索引时涉及一个复杂的过程，比标准的分词索引结果集大了两个数量级。他们曾经使用 Amazon Web Service 存储索引，但是爬虫抓取过程中的负荷主要集中在这一方面。（叮铃铃，叮铃铃“您好！这里是 AWS，无论您正在运行什么，请停止运行！”）他们恰好在寻求解决方案，而此时 Google 的 BigTable 论文发表了。

Powerset 公司的工程负责人 Chad Walters 此时发表了如下的言论：

与 Google 基于 GFS (Google File System) 构建的 BigTable 一样，在 Hadoop 的分布式文件系统 (HDFS) 基础上构建一个开源系统是一个非常不错的主意：（1）这套架构是成熟的并且可拓展；（2）我们可以直接利用 Hadoop 的 HDFS；（3）我们可以扩大 Hadoop 生态系统的影响力。

BigTable 论文发表后，在社区中，人们一次又一次地讨论基于 Hadoop 构建类 BigTable 系统的可行性。在 2007 年年初，Mike Cafarella 出乎意料地在 Hadoop 的问题跟踪系统中上传了一个包含 30 多个 Java 文件的 tar 包：“我实现了一个类 BigTable 架构的存储系统 demo，叫做 HBase，虽然它还不完善，但是它已经做好准备让用户进行实验和检查了。” Mike 与 Doug Cutting 在 Nutch（一个开源搜索引擎）项目中长期共事，Doug Cutting 在 Nutch 中实现了一个类似于 Google 分布式文件系统的项目来管理磁盘，因此 Nutch 中构建的索引存储可以不仅仅存储在一台机器中（Nutch 分布式文件系统最后发展成为 HDFS）。

Powerset 公司的 Jim Kellerman 增加了测试用例并填补了其他空白，使得 HBase 可以作为 Hadoop 的一部分代码进行提交。Doug Cutting 在 2007 年 4 月 3 日完成了 HBase 的第一次代码提交，代码提交到了 Hadoop 工程根目录的 *contrib* 子目录中。HBase 的第一个版本在 2007 年 10 月作为 Hadoop 0.15.0 的一部分发布了。

没过多久，本书作者 Lars 开始在 #hbase IRC 交流频道出现。当时 Lars 面临大数据的问题，并且尝试用 HBase 来解决这个问题。经过一番苦苦的摸索，Lars 成为了 Powerset 之外的 HBase 的第一个用户。我清楚地记得，Lars 当时记录了他在 WorldLingo 公司的生产集群的问题反馈清单，Lars 当时在这家公司担任 CTO。清单展示了他们的生产集群中 HBase 的 10 个版本（从 Hadoop 0.15.1 到 HBase 0.20），每个版本的集群都有将近 40 台机器。

在这些年来所有为 HBase 做出贡献的人中，具有史诗般意义的就是 Lars，因为他写了这

本书。Lars 一直在为 HBase 贡献文档，HBase 想要被更好地使用和推广，就需要有良好的文档。每个人都同意 Lars 的想法，并且能够专注地投入编程工作中，因此 Lars 在工作和欧洲旅行期间开始编写如何使用 HBase 的文档和架构描述，并承担起了 HBase 非官方的欧洲大使职责。Lars 在其关于 HBase 的博客 (<http://www.larsgeorage.com>) 中记录了 HBase 的工作原理，并在关键阶段推动了 HBase 社区的发展（一篇重要的博客文章解释了 HBase 依赖 Ivy 进行编译是个非常棒的主意）。

在微软公司赞助 HBase 的时期，HBase 也发生了非常有趣的事情。Powerset 在 2008 年 7 月被微软收购，在此期间其员工不允许贡献代码，因为微软法务部门需要审核 HBase 代码库并查看 HBase 与 SQLServer 的关系，直到一个月后才宣布重新贡献代码给社区（我是微软的一名员工，全职为 Apache 开源项目工作）。之后 Facebook 也开始使用 HBase，用于存储海量的邮件信息或点击信息，后来 Yahoo 部署了 1000 台 HBase 集群用于定位微软 Bing 的爬虫快照。同期非运行在 HDFS 上的 MapR 系统也仍处在开发中。

我很清楚，社区和 HBase 的发展得利于一群 HBase 的核心 committer 的辛苦努力。一些核心开发成员，如 Todd Lipcon、Gary Helmling 和 Nicolas Spiegelberg，已经付出了多年的努力，没有他们我们无法走到今天这一步，HBase 目前已经从一个分支代码成长为了一个独立存储项目。Jonathan Gray 冒险将其初创的 streamy.com 网站基于 HBase 进行建设，Andrew Purtell 在趋势科技组建了一只 HBase 团队，Ryan Rawson 得到了 StumbleUpon 的赞助，这是 HBase 在 Powerset、微软之后获得的最主要的赞助，并且还发掘了一个非常厉害的 commiter——John-Daniel Cryans，而当时 Cryans 还只是一个繁忙的学生。之后 Lars 不断地修复缺陷，并撰写文档。因此，Lars 是撰写第一本关键的 HBase 书籍的最佳人选，也让所有人都可以了解 HBase。

——Michael Stack, HBase 项目管理人

前言

你阅读本书的理由可能有很多。可能是因为听说了 Hadoop，并了解到它能够在合理的时间范围内处理 PB 级的数据，在研读 Hadoop 的过程中发现了一个处理随机读写的系统，它叫做 HBase。或者将其称为目前流行的一种新的数据存储架构，传统数据库解决大数据问题时成本更高，更适合的技术范围是 NoSQL。

无论你是如何来到这里的，我都希望你能够了解并学习如何在企业或组织中使用 HBase 解决海量数据问题。你可能有关系型数据库的背景，但更希望去研究这个“列式存储”系统；也许你听说 HBase 能够不费力地进行线性拓展，并且有足够的理由成为下一代网络系统。

在 2007 年年底，我曾面临百万级的文档存储需求，并且需要满足容错和可扩展等要求。我拥有丰富的 MySQL 数据库经验，并使用这种数据库来存储数据，最终服务于我的网站的用户。MySQL 可以在运行于一台服务器的同时，拥有另一台备份服务器，其无法应对如此海量数据的处理，于是我只好寻找其他可用的存储数据库。

我的口头禅是：“Google 是如何解决这类问题的？”后来我接触了 Hadoop，在短暂使用 Hadoop 之后，我面临随机读写的问题——但是这个问题已经得以解决：2006 年 Google 发表了 BigTable^① 论文，Hadoop 开发者拥有了 BigTable 的开源实现，并称其为 HBase。这就是解决我的问题的答案，所以这一切看起来顺理成章……

如今，我已经不再回忆自己刚开始接触 Hadoop 和 HBase 的日子有多艰难了。我希望可以从今天开始使用 HBase，HBase 目前已经成熟，接近 1.0 版本，并且目前已经有大量知名企业在使用，如 Facebook、Adobe、Twitter、Yahoo!、趋势科技和 StumbleUpon (<http://wiki.apache.org/hadoop/HBase/PoweredBy>)。我的集群是第一个生产集群（迄今为止），到目前也遇到了许多有趣的问题。

如预期所料，HBase 从 0.1x 版本开始成为社区项目，我有幸为这个项目贡献代码，并最终被要求成为全职的 committer。

过去几年我从其他开发者身上学到了许多知识，并且一直在努力地学习。我的信念是，我们还远没有达到这个技术的顶峰，而这个技术也会随着时间的推移不断地成长和演变。让我们用这本书对整个 HBase 开发者社区致以敬意，我的写作目标不仅仅是覆盖 HBase 的工作机制，而且还要为用户提供如何将这一技术用到自己的使用场景中。

① 见链接 <http://labs.google.com/papers/bigtable-osdi06.pdf>。

我强烈地感觉到你来到这里的原因是打算使用 HBase 解决你遇到的问题。现在让我们来解开谜底。

基本信息

在我们开始之前，以下是一些基本介绍。

HBase 版本

在写这本书时，HBase 社区已经决定发布 0.92.0 版本，社区主干的代码已经开发完成 (<http://svn.apache.org/viewvc/hbase/trunk/>)，之前刚刚发布了 0.91.0- SNAPSHOT 版本。

我们很难跟上开发的步伐，因为这本书有一个截止日期——在 0.92.0 版本发布之前。因此，它只能记录到一个特定的修订版 1130916 (<http://svn.apache.org/viewvc/hbase/trunk/?pathrev=1130916>) 为止。如果你发现书中描述的和 HBase 提供的之间似乎并不匹配，你可以使用以上的版本与最新版本比对近期的修改。

我们正尽一切努力更新本网站 (<http://www.hbasebook.com>) 上的 JDiff (比较不同软件版本的工具) 文档。使用它可以快速看到不同版本间有什么变化。

编译示例程序

有关这本书的示例代码可以在 GitHub 中 (<http://github.com/larsgeorge/hbase-book>) 找到更详细的信息。为了简洁，本书只提供了部分代码片段，即重要代码，尽量避免出现重复的样板代码。

每个例子的名字都能与链接库中的文件名匹配上，因此很容易就能找到这些示例。每章都有独立目录，呈树形结构，更利于用户查找。例如，正在读第 3 章，你可以到对应目录下查找第 3 章完整的示例源代码。

许多示例中所示特性都使用了内部辅助类来协助运行。例如，使用了 HBaseHelper 类来建立测试环境和收集测试结果。你可以根据具体情况来修改测试代码，或植入错误数据以观察示例中所示特性的行为。

编译示例代码需要借助以下辅助的命令行工具。

Java

HBase 是使用 Java 语言实现的系统，因此必然需要 Java 环境。2.2.3 节的“Java”描述了应该如何安装 Java 环境。

Git

示例源代码是通过 GitHub 来提供托管服务，因此我们还需要支持 Git——一个分布

式版本协作控制系统，Linux 内核开发最初就依托此系统做版本控制。^①开源社区提供了非常多的 Git 客户端二进制安装包。

此外，用户还可以通过 GitHub 提供的下载链接（<https://github.com/larsgeorge/hbase-book/archives/master>）下载文件的静态快照。

Maven

书中提供的代码编译过程是通过 Apache Maven^②完成的。它使用 Project Object Model (POM) 来描述编译的过程，其中包括有哪些代码可以编译以及哪些不需要编译。因此读者需要首先下载 Maven 安装库并在本机上进行安装。

一旦读者按照上述信息安装好所需要的工具，就可以按照以下命令开始编译项目。

```
~$ cd /tmp
/tmp$ git clone git://github.com/larsgeorge/hbase-book.git
Initialized empty Git repository in /tmp/hbase-book/.git/
remote: Counting objects: 420, done.
remote: Compressing objects: 100% (252/252), done.
remote: Total 420 (delta 159), reused 144 (delta 58)
Receiving objects: 100% (420/420), 70.87 KiB, done.
Resolving deltas: 100% (159/159), done.
/tmp$ cd hbase-book/
/tmp/hbase-book$ mvn package
[INFO] Scanning for projects...
[INFO] Reactor build order:
[INFO]   HBase Book
[INFO]   HBase Book Chapter 3
[INFO]   HBase Book Chapter 4
[INFO]   HBase Book Chapter 5
[INFO]   HBase Book Chapter 6
[INFO]   HBase Book Chapter 11
[INFO]   HBase URL Shortener
[INFO] -----
[INFO] Building HBase Book
[INFO]   task-segment: [package]
[INFO] -----
[INFO] [site:attach-descriptor {execution: default-attach-descriptor}]
[INFO] -----
[INFO] Building HBase Book Chapter 3
[INFO]   task-segment: [package]
[INFO] -----
[INFO] [resources:resources {execution: default-resources}]
...
[INFO] -----
[INFO] Reactor Summary:
```

① 见网址 <http://git-scm.com/>。

② 见网址 <http://maven.apache.org/>。

```

[INFO] -----
[INFO] HBase Book ..... SUCCESS [1.601s]
[INFO] HBase Book Chapter 3 ..... SUCCESS [3.233s]
[INFO] HBase Book Chapter 4 ..... SUCCESS [0.589s]
[INFO] HBase Book Chapter 5 ..... SUCCESS [0.162s]
[INFO] HBase Book Chapter 6 ..... SUCCESS [1.354s]
[INFO] HBase Book Chapter 11 ..... SUCCESS [0.271s]
[INFO] HBase URL Shortener ..... SUCCESS [4.910s]
[INFO] -----
[INFO] -----
[INFO] BUILD SUCCESSFUL
[INFO] -----
[INFO] Total time: 12 seconds
[INFO] Finished at: Mon Jun 20 17:08:30 CEST 2011
[INFO] Final Memory: 35M/81M
[INFO] -----

```

这段信息意味着有依赖关系的库已经下载到了本地目录，源代码也随后编译成功了。在每个子目录的 *target* 文件夹下都留下了编译后的 JAR 文件，即每章的源代码和 *.class* 文件归档：

```

/tmp/hbase-book$ ls -l ch04/target/
total 152
drwxr-xr-x 48 larsgeorge wheel 1632 Apr 15 10:31 classes
drwxr-xr-x 3 larsgeorge wheel 102 Apr 15 10:31 generated-sources
-rw-r--r-- 1 larsgeorge wheel 75754 Apr 15 10:31 hbase-book-ch04-1.0.jar
drwxr-xr-x 3 larsgeorge wheel 102 Apr 15 10:31 maven-archiver

```

上述场景中，*hbase-book-ch04-1.0.jar* 文件包含了第 4 章的示例代码。我们可以使用如下命令进行测试：

```

/tmp/hbase-book$ cd ch04/
/tmp/hbase-book/ch04$ bin/run.sh client.PutExample
/tmp/hbase-book/ch04$ bin/run.sh client.GetExample
Value: val1

```

bin/run.sh 已经配置了所需的 Java classpath，并添加了依赖的 JAR 文件。

Hush: HBase URL 简写

通过了解 HBase 提供的功能来了解 HBase 是一个好办法。本书使用了具有代表性的表的集合作为例子，并包含具体的数据，这样可以很容易地理解经过操作后数据状态前后改变的过程。读者可以执行每一个示例并查阅结果，该结果应该与书中提供的结果匹配。修改示例是进一步探讨和研究 HBase 的好方法，借助辅助类可以进行更有效的验证。

在前期的学习过程中，了解系统所有的功能是非常重要的一个环节，而这本书恰好提供了一个现实中的例子来展示 HBase 的大部分功能。同时这个例子也被用于与其他存储数据库进行对比，例如，与传统的基于 RDBMS 的系统进行对比。

这个应用的名称叫做 Hush——HBase URL Shortener (HBase URL 简写)。互联网提供了非常多的服务,每个服务都可以通过 URL 来访问。例如,你提交了一个网页,但你得到了一个返回的短网址。例如, Twitter 每次只运行并发送最多 140 个字符,但 URL 最长可达 4096 字节,因此不得不将 URL 简化到 20 字节以下,以节省更多的空间给实际内容。

例如,这是由 Google 地图引用 California 的 Sebastopol 的 URL:

```
http://maps.google.com/maps?f=q&source=s_q&hl=en&geocode=&q=Sebastopol, \
+CA,+United+States&aq=0&sl=47.85931,10.85165&sspn=0.93616,1.345825&ie=UTF8& \
hq=&hnear=Sebastopol,+Sonoma,+California&z=14
```

经过 Hush 模式,其可以转化成如下 URL:

```
http://hush.li/1337
```

显然,这个 URL 更短,更有利于复制到电子邮件,或通过限制字数的媒体(如 Twitter 和 SMS)发送。

但是,这种服务并不仅仅是一张大的查询表。诚然,非常多的用户希望能够映射短网址到完整网址,但需求并不仅仅如此。用户更多的是想了解短网址究竟被使用了多少次?因此短网址就需要保留一个计数器,用于在用户点击网址的时候进行统计。

更高级的功能是,用户可以使用固定的域名或定制的短网址 ID 而不是自动生成的地址,就像上面例子描述的那样。用户必须能够登录短网址跟踪,并查看日、周、月报表。

所有的这些都在 Hush 模式中实现了,用户可以很容易地编译和运行。它使用了 HBase 的大多数功能,在适当的时机我们会讨论这些问题。

用户可以自己创建账号并使用 Hush,这也是一个了解如何从已有系统导入数据的好例子。本书使用了网络上提供的数据集: Delicious RSS feed。这里面有少量集合是可以随意下载的。

使用场景: Hush

留意本书中从 Hush 角度解释的功能。很多地方使用了 Hush 的示例代码,但是它仅仅保持在非常简单的展示这一层面。Hush 作为一个使用场景更多的是应用在生产系统中。

Hush 没有华丽的 UI,只有朴实的功能,Hush 在经过负载均衡后达到数千的 TPS 请求完全没有困难。

有关 Hush 如何运行的代码片段已经在本书中进行了描述,完整的代码可以从 Git 库中下载并独立运行、调整和学习有关它的一切!

运行 Hush

使用示例代码编译和运行 Hush 非常简单,一旦用户可以克隆或下载,就执行以下命令:

```
$ mvn package
```

编译整个工程后，读者就可以使用以下命令来运行 Hush:

```
$ hush/bin/start-hush.sh
=====
Starting Hush...
=====
INFO [main] (HushMain.java:57) - Initializing HBase
INFO [main] (HushMain.java:60) - Creating/updating HBase schema
...
INFO [main] (HushMain.java:90) - Web server setup.
INFO [main] (HushMain.java:111) - Configuring security.
INFO [main] (Slf4jLog.java:55) - jetty-7.3.1.v20110307
INFO [main] (Slf4jLog.java:55) - started ...
INFO [main] (Slf4jLog.java:55) - Started SelectChannelConnector@0.0.0. 0:8080
```

看到控制台中的最后一行输出后，读者可以直接通过浏览器访问 <http://localhost:8080> 进入 Hush 服务器进行处理。

由于数据已经存储在了 HBase 中，因此用户可以放心地执行 Ctrl+C 来停止 Hush 的启动脚本。

排版约定

本书使用如下排版约定。

- 斜体：URL、文件名、文件后缀和 Unix 命令。
- 等宽代码字体：用于程序清单、功能变量、环境变量、数据类型、描述和关键字等。
- 加粗的等宽代码字体：展示命令或其他应该由用户输入的文本，也用于代码清单中强调的部分。
- 斜的等宽代码字体：展示应该使用用户所提供的值或根据上下文确定的值来替换的文本。



这个图表示一个提示、建议或一般的注意。



这个图表示警告或需要谨慎处理。

代码示例

本书的目标是帮助你完成工作。一般而言，你可以在自己的程序和文档中使用本书中的代码，如果你要复制的不是核心代码，则无须取得我们的许可。例如，你可以在程序中使用本书中的多个代码块，无须获取我们许可。但是，要销售或分发来源于 O'Reilly 图书中的示例的光盘则需要取得我们的许可。通过引用本书中的示例代码来回答问题时，不需要事先获得我们的许可。但是，如果你的产品文档中融合了本书中的大量示例代码，则需要取得我们的许可。

在引用本书中的代码示例时，如果能列出本书的属性信息是最好不过了。属性信息通常包括书名、作者、出版社和 ISBN。例如：“*HBase: The Definitive Guide by Lars George* (O'Reilly). Copyright 2011 Lars George, 978-1-449-39610-7”。

在使用书中的代码时，如果不确定是否属于合理使用，或是否超出了我们的许可，请通过 permissions@oreilly.com 与我们联系。

我们的联系方式

如果你想就本书发表评论或有任何疑问，敬请联系出版社。

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

我们还为本书建立了一个网页，其中包含了勘误表、示例和其他额外的信息。你可以通过如下地址访问该网页：

<http://www.oreilly.com/catalog/9781449396107>

作者还为本书创建了一个网站：

<http://hbasebook.com/>

关于本书的技术性问题或建议，请发邮件到：

bookquestions@oreilly.com

欢迎登录我们的网站 (<http://www.oreilly.com>)，查看更多我们的书籍、课程、会议和最